School of Professional Studies   Master's Papers

2018

Audubon Data Project Final Report

Askhat Beygenov
Clark University

Valinur Kutlambetov
Clark University

Shrikant Patel
Clark University

Phoebe Roberts
Clark University

Ulfat Sayyed
Clark University

See next page for additional authors

Author

Askhat Beygenov, Valinur Kutlambetov, Shrikant Patel, Phoebe Roberts, Ulfat Sayyed, and Shriram Sivaraman

Course	Capstone Practicum
Professor	Richard Aroian
Students	Askhat Beygenov, Valinur Kutlambetov, Shrikant Patel, Phoebe Roberts, Ulfat Sayyed, Shriram Sivaraman
Client	Mass Audobon – Shaping the Future of Your Community program
Title	Audubon Data Project Final Report

Acknowledgements

Table of Contents

Executive Summary

The Audubon Data Project was initiated as a Clark University Capstone project. The project's client, Mass Audubon's Shaping the Future of Your Community program, had identified a need to improve their data management methods and make better use of their data. The Capstone team, composed of Clark University graduate students, met with the client regularly to review the current state of the data and potential improvements to be made. The process began with a data review. During the review we worked with the client to explicitly define the purposes and requirements of the data, the current process for updating and using the data, and the ways that different types of records were related to one another. After the review, we were able to identify the issues in the current system which we would seek to resolve. These included data integrity issues such as ensuring crucial items (such as a town name) were always included when entering data, and data structure issues such as having a relatively user-friendly way to express relationships and update records that were part of a relationship.

We developed a database using MS Access in order to enable the client to store their growing set of data securely and continue collecting it for future use, as well as providing data analysis reports. The decision to use Access was partly based on the

client's lack of a budget for the project. We needed to ensure any solution we provided would be usable without purchasing new software. Access is also highly customizable and is SQL-based, meaning that the client would have the option to either make enhancements or move their data to another database system later if desired. The database provided several advantages over the original spreadsheet. It consolidated the client's list of contacts, which were spread out across several different sheets, into one table in order to allow them to easily identify and merge duplicate records. It provided a way to easily update and find information regarding workshop attendance where there had been a complex cross-referencing system in the original sheets. It also incorporated a set of data analytics tools which allowed the client to explore and compare engagement levels between different towns and locations, to identify areas of potential improvement and growth, and to highlight successes.

Finally we provided the client with steps to accessing a free online visualization tool in order to create more interesting and engaging charts than are available by default through MS Office programs. We designed several queries in the database which aggregated the engagement metrics from the client's data, providing a breakdown of the workshop attendance statistics and engagement levels (based on points metrics) for each municipality and region. We also provided a set of instructions for setting up their visualizations. The client can display these interactive visualizations to stakeholders to illustrate their impact in different parts of the state, or to make an argument for additional funds needed to hold workshops or otherwise boost engagement in an underserved area.

Introduction

The Audubon Data Project is a Clark University Capstone project. The Capstone program groups graduate degree candidates together and pairs each group with a client organization that has a need for help within the group's area of expertise. Our project team consisted of five MSIT graduate students and one MPA student. The client was Mass Audubon Society's Shaping the Future of Your Community Program, represented by Ariel Maiorano, Stefanie Covino, and Christina Wiseman.

The client collects data throughout the year which reflects engagement in their program through several metrics they've determined are the most meaningful. We worked with the client to enable them to make better use of that data by putting the tools in place for them to do so.

Project Context

The program frequently conducts workshops across different towns in Massachusetts and attendees from all over the state attend these workshops. They keep track of the attendees, workshop details, technical assistance and external information requests etc., and want to maintain a scoring pattern. In addition to these details, it is essential for them to segment and manage data in an efficient manner with a stable database. Additionally, they want to visualize data to gain insights and determine trends and patterns across towns to discern their overall performance and analyze their engagement levels. Hence the scope of the project is to focus in delivering a data management platform and a data visualization solution to enable them to make

data driven decisions efficiently and effectively, to evaluate their standpoint and to focus on future planning.

Problem Statement

The Shaping program has collected data regarding the program's various types of community engagement efforts in a spreadsheet per the advise of a consultant, a company called Root Cause. The program coordinator (Stefanie Covino) has requested a review of how the data is collected, stored, and analyzed as she believes the program isn't making the most effective possible use of this data. She has also requested analysis tools to be implemented in order to allow a lay user (i.e. someone with no IT experience) to pull data analysis reports regarding engagement in different areas.

Purpose of the Project

The project was initiated to put in place a data integrity and analysis solution and recommendations for the future. The project team agreed to review the data, recommend updates as needed, and implement a solution for continued use by the client based on the current data and the client's needs.

Significance of the Project

The client's stake in the project is the ability to better analyze their available data and improve their data management practices at a low cost. The project team's stake is real-world experience which will help them in their professional lives, along with fulfilling the requirements of their degree. Additional benefits for both may include the potential for networking.

Trends in the Industry

During the beginning phase of our project, we researched trends in the industries of data management and data visualization software as used by nonprofit organizations. We also reviewed the consultant report that was provided to the Shaping program and established the need for our project. This section outlines what we found in our research.

Literature on Similar Solutions

Data Visualization Solutions

Ghosh, S., Datta, A., Tan, K., Choi, H., & Stegle, O. (2018). SLIDE–a web-based tool for interactive visualization of large-scale–omics data. *Bioinformatics*, *1*, 3.

SLIDE is a web-based Java application designed for visualizing large and complex quantitative data sets such as those used in molecular biology research. It was developed by Soumita Ghosh, Abhik Datta , Kaisen Tan and Hyungwon Choi from the Institute of Molecular and Cell Biology, Singapore. It includes drill-down and search capabilities for pinpointing specific items of interest once a larger scale data summary has been reviewed. It is also customizable to an extent, allowing users to save information that is of interest for later, such as specific genes they are tracking. While our project is focused on a much smaller-scale data set, the strategies used for designing this and similar tools are useful to consider for the client's future needs, should they scale up their operation.

Weissgerber, T. L., Savic, M., Winham, S. J., Stanisavljevic, D., Garovic, V. D., & Milic, N. M. (2017). Data visualization, bar naked: A free tool for creating interactive graphics. *Journal of Biological Chemistry*, jbc-000147.

Interactive Dot Plot is a free web-based tool presented by Tracey L. Weissgerber, Marko Savic, Stacey J. Winham, Dejana Stanisavljevic, Vesna D. Garovic, Natasa M. Milic from the Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN, USA, Department of Medical Statistics and Informatics, Medical Faculty, University of Belgrade, Serbia and Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA. This tool allows users to create an interactive bar chart. Users without programming experience can start to create interactive graphics with different features. Data can be presented in scatter plots, box plots and violin plots. The authors describe the value of this tool in that static figures can be presented in more meaningful interactive forms. Moreover, they conclude that interactive content can improve clarity of publications. This type of visualization tool can be useful for our project as an example of a tool for users without programming experience. This approach allows users to visualize and work with data more easily.

Tsoi, K. K., Chan, F. C., Hirai, H. W., Leung, G. K., Kuo, Y. H., Tai, S., & Meng, H. M. (2017). Data visualization on global trends on cancer incidence an application of IBM Watson Analytics.

IBM Watson Analytics - visual analytics platform presented by Kelvin KF Tsoi, Felix CH Chan, Hoyee W Hirai, Gary KS Leung, Yong-Hong Kuo, Samson TAI, Helen ML Meng, Stanley Ho Big from Data Decision Analytics Research Centre Chinese University of Hong Kong, IBM China/Hong Kong Limited. The authors used this tool to show trends of falling global cancer based on cancer data collected by World Health Organization (WHO). This visual analytics tool requires a large amount of computing power. Cloud computing helped the authors to perform this challenging task. Moreover, cloud technology is a contemporary approach for big data processing, storage and software delivery, moving from traditional desktop computers and local servers to remote servers. This study could be useful to our project by providing insights into the cloud platform approach that would allow our client to process data in large volumes.

Luo, W., Pant, G., Bhavnasi, Y. K., Blanchard Jr, S. G., & Brouwer, C. (2017). Pathview Web: user friendly pathway visualization and data integration. *Nucleic acids research*, *45*(W1), W501-W508.

Pathview is a web-based application which provides easy interactive visualized access and generates high quality, hyperlinked graphs presented by Weijun Luo, Gaurav Pant, Yeshvant K. Bhavnasi, Steven G. Blanchard and Cory Brouwer from Department of Bioinformatics and Genomics, UNC Charlotte, North Carolina Research Campus and Department of Computer Science, UNC Charlotte. With this application users without special training or computing resources have full access to high quality data visualization, integration and analytics. Pathview can process different types of data. In this visualization tool data mapping and integration function are interactive. The tool also records data analysis historical information such as input data, analysis settings and results and saves this historical information on a server. It allows users to review, repeat and share their analyses with other users. This feature enables collective research. For our project we can take the idea of web-based applications and making all data interactive and hyperlinked with a saved history of analysis actions and without the requirement for special IT training.

Data Management Solutions

Gupta, H., Xu, Z., & Ramachandran, U. (2018, July). DataFog: Towards a Holistic Data Management Platform for the IoT Age at the Network Edge. In *USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18), Boston, MA*.

Data Fog is a data-management platform developed by Harshit Gupta, Zhuangdi Xu and Umakishore Ramachandran from Georgia Institute of Technology. This platform is oriented for the Internet of Things (IoT) infrastructure. IoT applications

have to process different kinds of data and this data should have a high-speed

feedback requirements between sensors and actuators. Data Fog performs data

placement and data compression based on relevance and priority that allows low

storage capacity. Data Fog can save storage space without taking extra

data. For example, the air temperature data during short periods of time

fluctuates within a small range. In this situation use of a full 64-bit integer is not

required. For our project the most important aspect is data-compression. These

techniques could be applied for future projects to improve the client's data

management process.

Stanford, N. J., Wolstencroft, K., Golebiewski, M., Kania, R., Juty, N., Tomlinson, C., &

Mueller, W. (2015). The evolution of standards and data management practices in

systems biology. *Molecular systems biology*, *11*(12), 851.

For modeling and predicting biological processes data have to be integrated. For

exchange and reuse, data should be formatted and described in standard ways.

Natalie J Stanford, Katherine Wolstencroft, Martin Golebiewski, Renate Kania,

from Manchester Institute of Biotechnology, The University of Manchester,

Manchester, UK and *S*chool of Computer Science, University of Manchester,

Manchester, UK have conducted a survey that covered areas of standards of

usage, data storage, sharing, reusability of data and results. In this the most

important aspect for our project is data storage and data sharing. To be more

successful, researchers need to exchange experimental data, computer code

and models between colleagues inside their institute in collaboration with others.

Most researchers still store their work on local hard disc. It creates additional barriers and difficulties for sharing with collaborators. Common repositories allows users to work with data in remote mode. Contemporary platforms can provide users space for data with remote access and with more security options for data backup and sharing. For our project the idea of using shared/cloud-based platforms to distribute data across the community can be useful.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*.

Data availability is very important today for reuse of the data. Infrastructure which can automatically find the data should be on a high level. Representatives from different departments like financial, academic, industry and scholarly publishers have come together to design a set of principles that can increase the reusability of their data. This group of researchers have concluded that good data management should include correct collection and correct storage of the data. This allows a variety of types of users to easily identify and analyze the data. Researchers have developed four fundamental principles: findability, accessibility, interoperability and reusability. Also, they have introduced examples of important data repositories. This is a good example of Data Management best practices. Based on this study, we can incorporate some of these principals when organizing the client's data. We can also recommend the

client use these principles to make data citable, discoverable and available for reuse for present and future projects.

Musick, B. S., Robb, S. L., Burns, D. S., Stegenga, K., Yan, M., McCorkle, K. J., & Haase, J. E. (2011). Development and use of a web-based data management system for a randomizedclinical trial of adolescents and young adults. Computers, informatics, nursing: CIN, 29(6), 337.

Data entry is an important part of Data Management. Modern technology can provide support to make this process go smoothly, without errors. In this study researchers have described Data Management processes where all data entry is supported by a customized and secured data entry system. In this case the data input process works like a quality assurance mechanism. This study shows us that data collection processes can be provided under computer control with maximum integrity. Moreover, oversight can be achieved by automated e-mail messages when a data entry system is in active mode. This example can be useful for us in terms of standardization of the data entry process from different users and ensuring integrity of the data via a computerized data management system.

Within Context of Consultant Review

The client had previously employed consulting company Root Cause to review their data management strategy. Root Cause suggested that the client use reporting tools to review and analyze their data. Using these tools, the coordinator can assess the program's performance based on the recommended indicators. All data analysis

techniques should provide a basic understanding of how to use the data to assess current performance and make improvements.

Data Storage and Update Process

We first focused on how data were collected and stored, based on interviews with the program coordinator. The program coordinator stores the data on Google Drive in order to make it available to multiple users. To make updates in spreadsheets users download the data on local computer and then upload updated spreadsheet. When we made our updates, we recommended that the data be stored using SharePoint / OneDrive instead. This allows for benefits such as greater security, built-in data backup, tracking of updates, and check-out / check-in functionality to avoid a user accidentally duplicating or erasing others' work.

Data Review

Our team reviewed the data collected by the client. Data was presented in excel files. The data was several times cleaned up to avoid errors before implementing analysis tools. After cleaning, additional spreadsheets were added to arrange data in a more convenient form to work with it.

Defining the Purpose of the Data

The client's goal is to reach as many people as possible to participate in trainings in different regions of the state. Root Cause suggested making some baselines and setting up a new target for each new year. Progress relating to these goals can be identified based on indicators in the visualization tools provided by our group, which are based on the data.

Indicators: Weighting, Scores and Points.

To analyze data our team worked with Program Coordinator to select four main indicators (Workshop Attendance, External Info, Workshops, Tech Assists) which would give us the best sense of organization's overall progress towards current strategic goals and which the client had enough data to accurately measure. The client had previously assigned point weights to each of these activities.

The points were assigned based on locations. For example, if a workshop was held in Natick, Natick would receive five engagement points. If someone from Worcester attended that workshop, Worcester would also receive one engagement point. Overall, points were assigned as follows:

- Town hosts workshop: 5 points

- Resident of town attends a workshop (and signs in, enabling them to be identified): 1 point

- Resident of town makes a request for assistance: 3 points

- Town works with the program on a long-term tech assist project: 10 points

Multi-Purpose Database:

We organized the data in tables that the client would be able to manage using an Access database, and incorporated the scoring system into the database. We used a system of multiples which would allow the client to later change the point weighting if desired.

Along with keeping track of points for data analysis purpose, the database acts as storage for records such as stakeholders' contact information, details about past

workshops, descriptions of assistance requests and tech assist projects, and cross-referencing information on workshop attendance.

Data Visualization:

To analyze the data, Root Cause suggested using visualizations such as graphs and charts. To this end we set up several visualizations in the form of a Workbook using free software Tableau Public. These visualizations are based on total aggregate points from the four activities mentioned above. They are categorized based on location (via towns and regions) and can be used to help the client discover trends in the data and display successes and areas of opportunity to advantage when working with funders, partners, and internal stakeholders.

Methods Used

This section outlines the steps we took to fulfill the client's need for data review, recommendations, and implementation of certain recommendations. Our team initially planned on making recommendations only; however, the client had a clear need for implementation of a tool that would be immediately usable to pull reports for data analysis. We modified our initial plan to accommodate this need for the client.

Project Design

The project had the following phases:

Project Set Up:

The purpose of this phase was to establish structure around the project. The phase included assigning roles and determining scope. To document what was done,

we drafted the Project Charter, Risk Assessment, Work Breakdown Structure, and other items relevant to project planning. We also performed and wrote up client research prior to the initial meeting with the client, held a kickoff meeting with the client, and each project team member completed an initial peer assessment survey.

Needs Assessment:

The purpose of this phase was to ensure understanding of the client's needs for the end state of the project. While the project charter codified these needs as project requirements, a more in-depth analysis was used to ensure we continued on a track that would ensure our deliverables could add value for the client as well as meeting their baseline requirements.

Design Phase:

The purpose of this phase was to design the database structure and reports that would be delivered to the client. Initial design, client testing and review, and documentation of the design were all included in this phase.

Implementation Phase:

The purpose of this phase was to create the items which were designed in the previous phase. It included several iterative rounds of build and test, with review by the client, to ensure the project was headed in the direction the client expected. Two separate sides of implementation were database management and visualization solutions.

Finalize Solution Phase:

This phase included final walk-through and delivery of the database structure, help file, data dictionary, and all other supplemental documentation as well as import of the client's data to the extent practical, and instructions for any remaining import needed.

Documentation Phase:

This phase was focused on documenting the solutions provided for the purpose of use by the client and by any subsequent IT professionals. A help file was created as well as various documentation to be used by IT professionals, in order to facilitate any later improvements made to the solutions.

Formal Report and Deliverables Phase:

This phase was related to the Capstone requirements as opposed to the deliverables promised to the client in the Project Charter. The Capstone Report was also delivered to the client, but contained specific content to fulfill academic requirements which might be of more interest to Clark University than to the client. This phase included finalizing this report, prepping and delivery of the presentation, delivery of this report to our Clark faculty advisor and to the client, and each team member completing the final peer evaluation.

Resources Used

The team used both tangible and intangible resources to complete the project. The client also expended resources to assist with the project, and many of the resources used belonged to Clark University Though the project did not have a

monetary budget, the resources listed here will give the reader some idea of the project's cost to all stakeholders involved.

Time:

We met as a team between 1.5 and 2 hours per week, each week. In some instances we had a meeting with the client in addition to our team meeting, which would generally be another 1.5 to 2 hours. Sometimes the client meetings were combined with the team meetings. In addition, project team members each spent an average of 10 hours per week on project-related tasks. Client representatives spent an average of 1 hour per week, outside of meetings, on project-related tasks.

Space:

For at least four meetings with the client, Clark University conference rooms were reserved and used for about two hours each. For the majority of project team meetings, team members met in the Goddard Library Academic Commons at Clark.

Computing Resources:

The project team used a combination of Clark computer labs (including printing facilities) and their own personal desktop, laptop, and/or tablet computers. One project team member borrowed a roommate's computer in order to be able to use MS Access software which was not compatible with his Macintosh (similar compatibility issues arose with several team members). Proprietary software such as MS Office programs and Tableau were used. The team focused on the use of software that the client would also have access to, namely MS Office suite programs.

Skill Sets:

The team drew on their education at Clark, and in some cases prior education (e.g. Bachelor's degree specializations and skills learned during employment), to complete the project tasks. The MSIT students were able to use their knowledge regarding database system best practices to design and implement a viable and useful database for the client. The MPA student used organizational communication and project management concepts to help plan the team's actions and lead the project, as well as using some limited background in IT to help understand the nature of the work being done by the rest of the project team.

Ethical Concerns

Data Privacy:

Because the client's data contained personally identifiable information for individuals who attended their workshops and served as contacts in various contexts, they requested that the team sign a privacy agreement prior to accessing the data. This was done during the initial phase of the project in order to ensure the team understood that they needed to prevent any leaking of data, either intentional or unintentional, by maintaining the level of security the client had established for the data.

Plagiarism:

The project manager reviewed all written products submitted to guard against plagiarism or academic dishonesty. She ensured guidelines were clear regarding what was acceptable or not when citing work from another source.

Conflicts of Interest:

The project charter included a provision indicating that any project team member with a conflict of interest (for example, someone holding a stake in the potential purchase of a piece of software being recommended) should disclose this conflict so that the client could take it into account.

Needs Assessment

Baseline Needs

While the client had some end goals that were explicit for the project, we determined there were also some implicit assumptions that spelled out the client's fundamental baseline needs, without which any solution we provided would not be of use.

Cost

The client was a small program within the Mass Audubon non-profit organization. They did not have any budget resources allocated to the project, meaning we needed to provide a zero-cost solution. This had several implications:

- A zero-cost solution meant using the client's existing tools to implement, and / or using tools which can be obtained at no cost and are compatible with their existing hardware.

- Trade-offs would be needed relating to capability and ease of use for any solutions provided.

- For future recommended expansions that required a one-time purchase or a subscription service fee, we needed to provide a justification of benefits over the no-cost options and a cost comparison.

Usability

We use the term "usability" to encompass two categories of barriers that the solutions provided would need to span.

- First, it means compatibility with current systems owned by the client, and for data to easily be exported from storage for analysis.

- Second, it means any solutions provided needed a user friendly interface with which a non-technical person would be able to easily manipulate the data.

Value-Added Needs

The client had several items that they listed in their initial project requirements, which we summarized within the following two categories:

Data Management

The client needed to be able to better manage their data. They were looking for recommendations as well as an interim solution for data storage and use that would make their data more manageable.

Data Analysis and/or Visualization

Alongside the data management solution, the client wanted more options for analyzing, visualizing, and using their data.

Data Review

The client's data consisted of an Excel sheet that housed several tables, one table per sheet, containing related data. The client wanted to analyze the data they collected in order to measure different types of public engagement for their program. They wanted to be able to measure engagement for different locations and be able to filter for the most current data as needed. Activities that measured engagement were each recorded in separate sheets of the Excel workbook. Specifically, activities used for measurement were workshop attendance, workshop hosting, external information or assistance requests, and tech assistance requests. The client was interested in how much the residents of various towns in surrounding areas were engaged through these various different actions. They had assigned a points value to the different types of activities, as follows:

- Workshop Attended by resident of town: 1 point.

- Workshop Hosted by town: 5 points.

- External info / assistance request received from resident of town: 3 points.

- Tech Assist project with town: 10 points.

The client was also interested in determining overall exposure (via attendance numbers) to specific items covered in certain workshops, such as the MAPPR tool and climate change information.

Knowledge Transfer

Meetings with the client were used to go over their data and gain a good understanding of how they used it and what it signified, as well as what changes they

would like to see. Meetings were documented via meeting minutes and the project charter in order to ensure correct understanding of the topics discussed.

Identifying Goals

The project's aim in importing data to a new format had several end goals.

Enforce Data Integrity

The data review revealed inconsistencies in the data which could be resolved by storing it in a database management system that would enforce entry of certain fields in every record and would eliminate duplication of records as much as possible.

Allow Easy Access

The original data storage was a simple Excel sheet which was relatively easy for non-technical users to understand. If we wanted to implement a more sophisticated data management solution, we needed to ensure that end users would still not need any technical knowledge to access the data as they had before, which would require building a user interface to help them manage their data.

Provide Analytical Tools

One of the items at the top of the client's wish list was a way to easily pull relevant information out of the data. This was the most complex goal of the project and was achieved through the steps outlined below under Implementation.

Parallel Goal - Recommendations for the Future

The team recognized that though they could upgrade the program's data management and analysis tools, a more nimble solution might be desired in the future.

Microsoft's native visualizations are functional but primitive, and more attractive and interesting visualization tools are available. MS Access can be used in conjunction with a website or shared file, but it is not cloud-based and doesn't adapt to remote user account functionality easily.

To help the program prepare to scale up their data management when desired in the future, the team did a comparison analysis of some likely data management and visualization tools which would provide advantages such as cloud server storage, remote access, more visual appeal, and easier compatibility with data analysis and exploration tools of various types.

Identifying and Resolving Issues

A variety of issues were identified while reviewing the data which needed to be resolved in order to provide a worthwhile solution.

Data integrity issues

Completeness and consistency were of concern when reviewing the data. For example, we noted early in the process that most of the client's desired data analysis focused on location - specifically most records were relevant to the town they were associated with - however, the town field was blank for a significant number of records. The major data integrity issues identified were as follows:

- **Town name missing or misspelled**.

- **Dates missing or in an inconsistent format**. Along with location, timeframe was most likely to be of use for the client's data analysis needs, and so dates needed to be described in a way that allowed the client to select a date range and receive relevant results.

- **Multiple large and seemingly unused data fields to the right of the main record**. A large number of seemingly superfluous fields had to be accounted for which would take up space in a database and slow processing down, as well as being unwieldy when designing input forms. We reviewed these with the client to determine which ones were relevant to the data they wanted to store, and which were unused and could be removed

- **Records listed for Workshop Speakers didn't indicate what workshop they spoke at**. This prevented us from giving the towns credit for their engagement when doing data analysis.

The client took actions to resolve the bulk of these issues by updating or normalizing the questionable records, and advised us which fields could be removed. We also went through the data to remove duplicates, and similar basic data cleaning tasks. The Workshop Speakers table, being rather small and incomplete, was taken out of the mix for the purpose of data analysis based on the client's decision.

Data structure issues

In order to develop a working database that could be used for both storage and meaningful data analysis, several issues needed to be addressed in the client's existing flat-file data structure.

Metadata, definitions, and data dictionary:

Data had clear definitions and meanings for those using it every day, but how it should be used was unclear for outside IT workers looking at it for the first time. One of our actions in preparing the data for import was to create a comprehensive data dictionary that defined data types and meanings in order to streamline future data

management efforts for the client. There were also several tables that were implicit in the current structure but that needed to be spelled out. These included the points allocation for each activity as well as towns and regions.

Relationships and how best to define:
Towns and regions were separated out into a reference table to ensure consistency and easy analysis based on towns. People whose names and contact information were listed under various different records were centralized into one table so that their data wouldn't be duplicated in different locations. This ensured that updating a person's contact information in one place resulted in an update everywhere.

A many-to-many cross-reference relationship existed between any person attending a workshop and the record for the workshop itself, which was reflected in a rather unwieldy way in the Excel file. The table listing Workshop Attendees included a column for each workshop (over 200 of which were recorded) in order to cross-reference the attendee with each workshop they attended. Updating the table required one to scroll far to the right, losing sight of the attendee's name, and search for the workshop name in the column headers. This relationship would be more easily adapted to a database environment using a join table.

Expressing score metrics
Although the prior consultant had defined the recommended metrics (the points system) for the client, and the client was interested in using those metrics, the data itself didn't reflect any points values. This information needed to be defined explicitly in the data structure so that it could be reflected for the client to analyze and visualize local engagement levels through those metrics.

Database Design

We determined that a relational database design would best express the client's data, dividing it into a structure based on objects that could easily be cross-referenced for streamlined data entry with minimal errors, and for pulling reports that would contain meaningful statistics.

### High Level Relational Database Concept
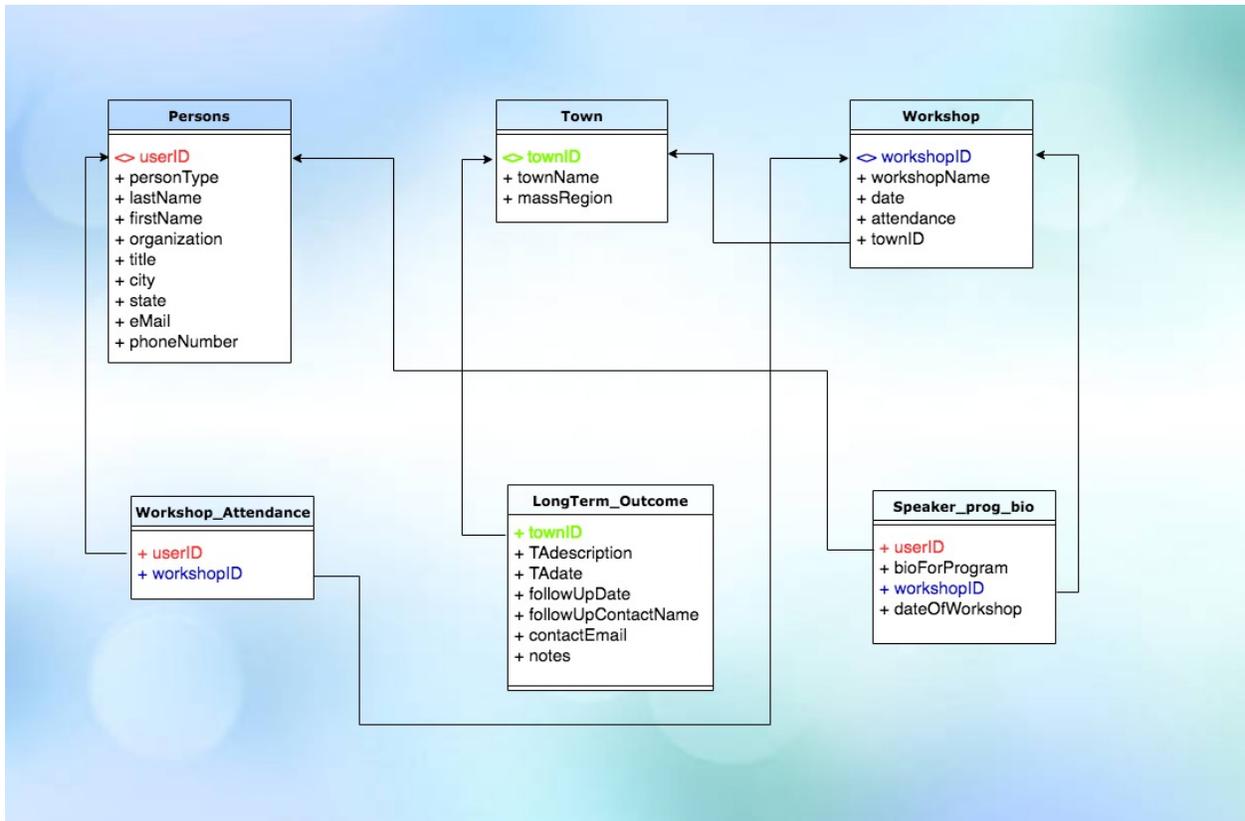
Several changes were made from the original data structure expressed in the Excel spreadsheet, to promote less redundancy and more streamlining of information, and to ensure that the database was robust enough to stand up to modifications and expansions to the design as needed.

- **Person table incorporates all people in the same place**: Since a person might conceivably attend a workshop and then later speak at a future workshop, entering the person's data twice in two different tables would be a waste of time and potentially create errors in the system. Instead, once the person was registered as a speaker they could be flagged as also a speaker, cross-referenced to the workshop where they were speaking, and a bio created for the program.

- **Cross reference table for workshop attendees**: For the reasons outlined above in the data review, a cross reference table was created for each instance where someone attended a workshop.

- **Incorporation of Scores**: The client had already created a scoring model, allocating a certain number of points to indicate a municipality's level of engagement based on how many workshops hosted, how many residents attended a workshop, how many information requests came from the town, and how many tech assistance projects were completed in the town. The original spreadsheets did not explicitly include points values. We incorporated these by adding a new Score column to each table of information that would affect the overall engagement score. This enabled us to create location based visualizations for the client to compare the level of engagement across different regions and towns.

Finalized Relational Database Design

person
- personID INT
- person_type INT
- f_name VARCHAR(45)
- l_name VARCHAR(45)
- organization VARCHAR(255)
- title VARCHAR(255)
- townID INT
- state VARCHAR(45)
- phone_number VARCHAR(45)
- town_townID INT
- extinfo_personID INT
- workshop_attendance_wa_mappingID INT
- workshop_speakers_ws_mappingID INT

Indexes
- PRIMARY
- personID_UNIQUE
- fk_person_town_idx
- fk_person_extinfo1_idx
- fk_person_workshop_attendance1_idx
- fk_person_workshop_speakers1_idx

workshop_attendance
- personID INT
- workshopID INT
- wsa_score INT
- workshop_attendancecol VARCHAR(45)
- workshop_attendancecol1 VARCHAR(45)
- wa_mappingID INT
- workshop_workshopID INT
- workshop_workshop_attendance_personID INT
- workshop_workshop_attendance_workshopID VARCHAR(45)

Indexes

workshop_speakers
- personID INT
- workshopID INT
- ws_score INT
- ws_mappingID INT
- workshop_workshopID INT
- workshop_workshop_attendance_personID INT
- workshop_workshop_attendance_workshopID VARCH...
- 1 more...

Indexes

extinfo
- personID INT
- request VARCHAR(1000)
- date VARCHAR(45)
- ei_score INT

Indexes

workshop
- workshopID INT
- workshop_name VARCHAR(45)
- date VARCHAR(45)
- attendance VARCHAR(45)
- townID INT
- workshop_attendance_personID INT
- workshop_attendance_workshopID VARCHAR(45)
- workshopscore_workshopID INT

Indexes

town
- townID INT
- town VARCHAR(45)
- region VARCHAR(45)
- longterm_outcome_lto_mappingID INT
- workshop_workshopID INT
- workshop_workshop_attendance_personID INT
- workshop_workshop_attendance_workshopID VARCHAR(45)
- workshop_workshopscore_workshopID INT

Indexes

workshopscore
- workshopID INT
- workshop_score VARCHAR(45)

Indexes

longterm_outcome
- townID INT
- date VARCHAR(45)
- description VARCHAR(45)
- followup_dates VARCHAR(45)
- followup_name VARCHAR(45)
- email VARCHAR(45)
- note VARCHAR(255)
- lto_score INT
- lto_mappingID INT

Indexes

Peer review and attempted implementation of the initial high-level design revealed some issues, listed below:

- Person records included a PersonType field which was the subject of debate. It denoted whether the person would be a member of the group workshop attendees, external assistance requesters, or long-term (tech assist) requesters. The issue was whether or how the table would point in the other direction, to

records that were related in the one-to-many relationship. One person could have more than one type and be cross-referenced by more than one table (for example, workshop attendees and also tech assist). This was considered important from the perspective of building input and data viewing forms, since each person would need to be located if adding them in a new capacity.

To resolve this, we did the following:

- When adding a person, they would always be added in some capacity (not just an unconnected person going into the database for no reason) so they could be added from the tables pointing to the Person table rather than only from the Person table itself, so as long as the other tables referenced the Person table and there was a way to check that data to see if the person already existed, we would be capable of updating or adding to the Person table.

- Subform functionality in Access can be used to display the two linked records when viewing and updating, so that duplicate Person records would not need to be created.

- PersonType was left as a field in the database in case it could be of use later. It was structured as a drop-down to make it easy to input.

Implementation

After the above process of data definition, cleaning, and normalizing, the data was transitioned into the proposed database structure and manually updated to include foreign and primary keys. It was then imported from Excel to Access, where the

relational structure was defined. Forms were made for viewing, updating, inputting and deleting data records. The database structure was then used to construct data analysis reports for the client.

Data Import

The bulk of the data import process involved preparing the data to be imported to Access by changing it to the structure that had been designed and adding any missing fields such as keys which would be needed in the database format. The steps to that process are outlined here.

Data Definition

The team worked with the client to build a data dictionary that accurately represented the fields in each table of the database, which was not only needed in preparation for data import and building of analysis tools, but will also help the client's IT professionals or consultants they hire with any additional improvements to the database.

Data Cleaning and Normalizing

Not all data in the tables was consistent, and there were duplicates and incomplete records which needed to be merged or updated. Many changes were made in partnership with the client to ensure that the data could be imported into a database environment.

Normalizing the Structure

In order for the Excel data to easily be imported to Access, the data first needed to be reorganized in the same tables and fields that would be used in Access. This

included primary and foreign keys, which needed to be manually generated so that they would be ready for use after importing and the data could easily be related between tables.

Data Management Forms and Interface

While the client had previously worked directly in Excel, simply adding another line for each record, the Access database needed a more structured approach to avoid the client needing technical know-how to update their data in the relational structure. Forms were built in Access for data input, update, and general viewing. The process was somewhat complicated by the relational structure, which required the use of sub-forms. Searchable drop-down menus were used in instances where a table's data needed to be related to another table.

Data Analysis Reports (Native)

A few data analysis reports were added to the Access database to allow the client to easily pull some information without going to another tool. For more advanced visualizations, see the section on external data visualization below.

Workshop Reports

These reports pull data from the Workshops table only. They provide the following information for each calendar year:

- Number of workshops and total attendance for each year.

- Average attendance per workshop for each year.

- Number of workshops / participants exposed to MAPPR, and to climate change discussions.

- Number of workshops held in partnership, and the total attendance of these workshops.

External Assistance Request Report: This report provides totals for External Assistance / Info Requests received each year.

Tech Assist Report: This report provides totals for Tech Assistance projects initiated each year.

Engagement Summary Report: This report tallies up the engagement points scored by each town and breaks them down by type of engagement activity. Towns are grouped by region.

Data Visualization (External)

We explored several solutions for data visualization, i.e. graphs, charts, and other visual representations of key data points. Our challenge was to balance the client's need for a no-cost solution with their need for a user-friendly solution that could be updated and used by a layperson with no IT skills. The main options we considered are listed in the table below:

Solution	Pros	Cons	Decision
Tableau (proprietary software)	Interactive and advanced customizable visualizations, high presentation value.	Does not meet client's no-cost criteria, involves at least some learning curve.	Included in recommendations for the future.

Excel graphs	Available at no cost since client already has MS Office	Easy to break, due to the fact that data would either have to be kept in Excel, where integrity is not enforced, or visualizations would have to work in tandem with Access export.	Mentioned in this report with its limitations, so the client is aware of potential issues and can work with them if implementing in future.
R Shiny program	Open source (free) platform. Attractive interactive visualizations, high presentation value.	Requires a learning curve, and original program written by one of our team members, so no support in place if it malfunctions.	Scrapped due to lack of user support for client's future needs.
Tableau Public	Free program with support provided by company, attractive interactive visualizations / high presentation value.	Some learning curve / instructions required to get the data out of Access and into the viz program. Specialized query needed to be developed in Access.	Adopted as best compromise for no-cost, user-friendly solution which still adds value and is sustainable / supported.

As outlined above, we settled on Tableau Public as the best solution to fit the client's visualization needs. We took the following steps to maximize the value of this solution for the client:

1. **Created queries in Access that would pull the most useful data for Tableau Public** to parse for visualization - this ensured that the visualization would present meaningful data and would not arbitrarily present data in unintended ways which might not relate to the client's needs, for example measuring something that the client had no interest in measuring.

2. **Designed workbooks (data visualization set-ups) in Tableau Public** tailored to analyzing the client's data. Unfortunately we weren't able to directly share

these workbooks due to the limitations of the free version of the software;
however, the client would be able to duplicate our steps in the free version.

3. **Create a step-by-step instruction set** for doing one-time set-up to configure the
parameters needed to create the visualizations we'd designed - this was done to
ease the learning curve that was the major drawback of either version of
Tableau.

4. **Provide instructions for exporting data** from Access into Excel using the query
we had created, and loading the data into the Tableau Public workbook,
whenever the client wishes to update the visualizations with more current data.
This addressed the other usability issue with Tableau Public, which is the need
for import of data formatted correctly in order to provide meaningful
visualizations. As with the workbooks, the client had the option invest more time
in the future to create additional specialized queries so that they can use Tableau
Public to visualize more of their data.

Documentation

Several types of documentation were required to ensure that no information or
knowledge would be lost at the end of the project. For IT and data projects especially,
documentation is crucial to the sustainability of any solution provided, as it allows future
professionals to expand and improve on any items that were designed by someone no
longer working with the organization, and provides users with instructions on set-up and
use of any non-intuitive portions of the solution.

User Help Documentation

A basic help manual was provided to assist with use of the interface designed in MS Access, and with updating records manually if needed.

Documentation for IT Professionals

The relational structure, data dictionary definitions, and test plan were provided to the client for their records, so that future professionals who were asked to make improvements or expansions to the database would not need to start from scratch.

Formal Report and Project Documentation

This report consists of a comprehensive accounting of the project's structure and methodology. It includes all project documents and other related items in the Appendices section. Though the client may not ever read the entire report, it is provided as a reference and a clarification of the project team's methods, in order to remove as much ambiguity as possible relating to the "hows" and "whys" of the project.

Recommendations

Included in the Summary Conclusion section of this report is a set of recommendations for use of the solution provided as well as potential upgrades to the client's data management and visualization capabilities. These recommendations fill in some of the gaps for actions we believe would be beneficial to the client but couldn't be done due to the limited time, scope, and budget of the project. We believe they would be beneficial for the client in the future if they have resources available to implement one or more of these recommended solutions.

Results and Reflection

The team's goal, along with fulfilling the client's need for a better data management and analysis solution, was to learn from the experience and to use the skills gained in our pursuit of our degrees at Clark University. To that end, we included our thoughts regarding the project in this section, and assessed what we were able to learn from the process. This will help us get the most value out of the project for our own future benefit, and that of anyone who works with us on similar projects. We hope that relating what we have learned may also be of use to the client, the University, and future students in similar situations.

Askhat

This project gave me a chance to apply my four years' experience in the software development industry and take a part in different activities including the literature review, data analysis, java code review and quality assurance.

In the literature review I have mostly been researching information related to data management systems, their advantages, disadvantages and applicability to our system. I spent most of the time on data analysis, because the data was unprocessed and some of the information was missing. Thanks to my team, who helped me to understand the requirements in more detail. As I have java coding experience, I thought of building an application using this programming language, but I ran into issues with visualization, so we decided to split the data management and visualization parts.

My biggest contribution in the project was quality assurance. During the semester I did regression testing of different versions of the data management system. Every feature we introduced required additional testing. For documenting test plan and test results I used TestRail solution which is one of the most popular in IT projects. I thoroughly tested our application and I am completely confident that it is ready to be used in real conditions.

I appreciate the support I received from my team, and I am really happy I had a chance to be a part of an international group of highly skilled professionals.

Valinur

After the project was completed, I became more confident in the use of database systems. My database skills, which we developed together, includes how to draw and use relationship diagram. I had understood on how to create flowchart that illustrates how persons, workshops, towns relate to each other within Audubon system. Now I can more understand relations between databases in terms logic and business rules. This step, in software engineering is initial to identify requirements for an information systems project. Later, in testing stage, understanding of relationship diagram helped me to analyze database and to find and fix problems.

At the software testing stage, I took part in real and deep software quality testing activity. It was very interesting to test software based on our databases and debug system in real time. I was excited to do so. I have also learned that how to create Audubon Data Project Test Plan with test units and test cases. All testing activities allow me to gain great experience which I can implement in my future career path.

During literature review and analyzing industry trends for visualization software I have studied many visualization tools which implementing in different science. Now I'm more confident in that how to select and implement visualization tool to present data in more meaningful forms and how make data interactive. For our project we have selected Tableau due to many advantages for users and I think it was a correct choice.

At the end of my review, I would like to mention the main experience I gained during the work on the capstone project – is team work. In this team I was lucky to work with unique people who love their work and are responsible for their words and actions. During the whole time of the project we discussed a lot and always came to a common opinion. I think the ability to hear everyone's opinion and find a common solution to complex problems is the main experience that I have earned here.

Shrikant

It's always best to practice and go through your notes before stepping on stage. The Mass Audubon project was that practice, the 'being thorough' with all technical and people skills we have before stepping into the real world of actual work. Apart from sharpening my skills, I also had the pleasure of working side by side with five incredibly skilled professionals, motivating professors, and a very engaging client.

My role in this team was to understand the data provided by the client and develop the database structure that can help data to flow easily in Microsoft Access and provide favorable reports to the client. Our first and biggest challenge was understanding the data. Since the data was really huge, it took 2-3 months and 4-5 meetings with the client to understand it. During this period we not only understood the

data well but also bonded with the client. That helped the team to set the deliverables clearly to the client.

After understanding the data, my approach to managing the data was to segregate the data according to real-world entities. For example, all the data about a person should be in the Person table, all the data about workshops should be in the Workshop table and all the data about towns should be in the Town table. This simple approach was game-changing as this helped us to develop reports in the later part of the course.

After the forms were placed in MS Access for data entry, my next challenge was to develop SQL queries for the report. Even though I have work experience with SQL, my knowledge was limited since it was less SQL oriented. However, this project was more SQL focused which opened doors for me to explore SQL more.

Conclusively, the Mass Audubon project helped us to embrace our maximum potential as graduate students. Personally, it helped me not only to be a skilled professional but also to be a passionate IT manager.

Phoebe

This project gave me a chance to stretch my project management skills with a team that was very dedicated and very diverse. I was challenged with bringing the team to convergence with understanding various aspects of the project, as well as coordinating that understanding with the client. One of the biggest challenges was breaking down incorrect assumptions - my own, as well as those of the client and the rest of the team.

If I had the project to do over again, I would want to take more time at the start of the process working with the team to understand their expectations for what they would do during the project. I would want to spend less time initially with rushing to propose solutions, and find a way to explicitly define the possibilities for a solution so that everyone was on the same page for understanding what the solution would entail. I made assumptions about the steps that would go into completing the project, because of what I felt was a lack of sufficient time with the team to hash this out together. It would have been more helpful to keep discussions on track with the whole team present, and to decide together what steps would be needed for completing the project.

Having said this, I am still thinking about how I would improve my technique for managing such a process. The great work ethic and dedication of many members of the team manifested in a desire to forge ahead with their vision of how the project should go - often leaving others behind and diverging in separate directions, leading to wasted work. Converging those visions via direct conversation between six people of diverse backgrounds is easier said than done. This directs my attention to one of my continuing weaknesses in the workplace as well as in a project setting, which is my lack of ability to share my vision.

One of my limitations in the past has been my tendency to shrug off certain leadership concepts. The concept of a shared vision is one of those that I have often ignored, dismissing that and similar phrases as essentially buzzwords, used emptily by corporate suits to sound intelligent. Unfortunately, I do not think I am wrong in this; however, I think there is value in a *true* shared vision - and in many of the other "soft"

concepts that are associated with leadership. The issue is in tying these buzzwords to the solid meanings on the ground that they were meant to describe. This is something I am still working on, but one of my biggest lessons of this project was its importance.

Ulfat

The journey with Mass Audubon data project has been very exciting to me. This project genuinely gave me the real life experience to professionalism. Being a fresher, it was like my first real time project other than the projects which I already did in my undergrad program. By the help of this project, I got to know that there are many responsibilities and formalities which we need to take care of before starting with the project. When our team signed an agreement on the privacy of data for Mass Audubon, that time I realised that the capstone is really big. I will be interacting with the real-time data of the real-time company. Things were no more dummy, different formalities such as project charter, weekly reports, work distribution and many such things made the project more interesting.

I was fortunate to have such an amazing team with the diverse skill set and a well experienced faculty advisor. Working in a team made me realize my strength and weakness. I felt that I was confident enough to explain my point and keep the clear image of the outcomes of my work but my weakness is I was not patient. I wanted things to work according to my pace which discouraged me a little. Later, talking with my faculty advisor and teammates helped me realise that this is how the structure flows in any project, it was just me being inpatient.

After researching and filtering out various aspects, I introduced the solution of using MS Access which best suited for the requirements of the client. I contributed for the user Interface of MS Access and designed various forms and generated multiple reports so that user can interact directly with the data without going through the tables everytime. I tried designing it as easy as possible so that it will not be complex for not only for client but also for any new employee they hire. I remember our client saying "I want tables to talk with each other" and this is how "We made them talk" by connecting and relating the content of data, defining perfect keys for them with taking special care of data integrity. This project not only fortified my researching, database management, analytical skills but also made me discover my own hidden skills. It showed me my strengths and weaknesses. So, now I know that in what areas I should focus on. I again want to mention, this was only possible because of the good explanation of data given by the clients, understanding and friendly nature of our faculty advisor and co operative, helping and excellent teammates.

Shriram

Mass Audubon data project has provided me a great opportunity to perform market research in search of optimum products for data management and data visualization platform. The rich data set prompted me to create different types of visualization catering to specific set of needs. The data set is a right mix of categorical and numerical metrics, which makes it easy to segment data in a different way and provides a clear pathway for plotting.

During my work in this project, I invested a lot of time researching for possible alternatives in the market for managing data from a cloud domain. Snowflake, Google BigQuery and Amazon RedShift are some of the pioneering cloud database systems that offers wide range of functionalities in hosting data. From the list, I researched the usage and working mechanism of snowflake cloud system. It is easy to install and maintain this system and it offers SQL enabled features to load and analyze data. It has a user-friendly web interface, which are separated by tabs where each tab has its own set of unique functionalities like worksheet, database and warehouses. Hence, from the knowledge I gained through this research, I would strongly recommend Snowflake for future data management projects.

For data visualization, there are products like Tableau, R Shiny and data studio that offers a diversified form of visualization. From the lot, I would like to recommend Tableau due to its ease in use. With the data set, I was able to create scatter plots, bar graphs, bubble chart etc., where each visualization conveyed a unique insight. As the data set is large, the viz platform assists to gain insights and leverage desired outcome. Queries on highest & lowest performing towns/regions, highest & lowest attendance count per towns/regions can be arrived at faster pace and measures can be taken immediately to improve the awareness in the future workshops. From the bubble chart (heat map), the attendance distribution and concentration levels can be sorted out and possible ways can be figured out to conduct workshops in the unexplored areas.

This project has assisted me to work with a great team consisting of six members where each of them excelled in their respective areas. Everyone had their valuable

inputs for the better outcome and assisted in developing MS Access database, design ER model, quality testing, documentation and Tableau data analysis. I have strengthened my skill sets in the data analysis through this project and will apply this experience to my future projects.

End State of Project

We are providing the following deliverables to the client.

- **An Access file containing the majority of the client's data, imported from the original Excel structure**. Not all data was imported due to time restrictions. When changing data structure, specifically when creating relationships between tables, some data must be entered by hand rather than importing. Below, in the Getting Started section, we outline items that may need to be re-entered in the database.

- **A modified copy of the client's Excel file which the client currently uses for data storage**. The major purpose of this item is to provide easy references for updating the Access sheet. The formatting of the original Excel sheets made it unusually difficult to input cross-reference data regarding workshop attendees, so the team modified the sheet for their use to improve the formatting, and is sharing this tool with the client to make their data entry process go more smoothly.

- **Steps to replicate the Tableau Public workbook we designed with customized data visualizations**. This will allow the client to import data into a workbook and view a user-friendly and attractive data visualization,

as well as giving them the potential to expand and reconfigure their visualization workbook in the future.

- **Help documentation for data management and data visualization solutions**. Two PDF's are being provided to the client with instructions on how to use the Access interface for the database we've built, and how to use and update the Tableau Public workbook.

- **An electronic copy of this report**. This report includes our future recommendations for the client, as well as more detailed aspects of the project which may be of use to the client or to professionals that they may engage in the future to make updates to the solutions we've built for them.

Salient Issues and Lessons Learned

Project scope assumptions: At the outset, there was a clear discrepancy between two potential directions the project could take, one being purely a written recommendation for use of specific software tools or data storage methods, in the vein of what a consulting firm might do; the other being the implementation of an actual solution usable by the client and the import of the client's data into that solution for their immediate use at the end of the project. The client determined that the latter would best serve their needs. **Lesson learned**: shared vision and cohesion needs to be established early in the process of building the project team, and the team should not rush to propose a solution before ensuring they have the correct understanding of the project.

Timeframes required for work: Cleaning and normalizing the data took longer than initially assumed by the client and project manager due to several inconsistencies and some missing data which had to be handled manually and due to the limitations of MS Access in importing data from Excel. **Lesson learned**: breaking down required steps with the subject matter experts in the team, and consulting with them to ensure they agree with any proposed schedule, is crucial to a successful project.

Assumptions regarding platform functionality: Project team members working on the Access file were unfamiliar with many aspects of the database application's functionality, and needed additional research time to determine how to do actions such as pulling data from multiple tables and using subforms when updating data with a one-to-many relationship. **Lesson learned**: subject matter experts should be open to learning new aspects of their field of expertise. Particularly in the IT field, there is a need to adapt knowledge to the current tools available, not to mention keeping up with the constant evolution of technology in the field.

Data structuring decisions: Project team members had differences of opinion regarding how to restructure the client's data. There was a concern that the client would be confused by a relational database structure and would need to adapt to inputting information in a different way than they had been used to previously. This was resolved by creating a front end which displayed records logically from the client's perspective, making the transition smoother. **Lesson learned**: in many cases, team productivity hinges less on good ideas or hard work, and more on knowing the right questions to ask

in order to bring out team members' assumptions and adjust them as needed, so that the team shares a vision of a complete and specific common goal.

Summary Conclusion

Throughout our project, we have kept our focus squarely on the client's needs and how best to fulfill them. Our biggest challenges were in ensuring our full understanding of the client's needs, the meanings and current status of their data, and the nature of the tools that were available to us for helping them. We've learned a lot about adapting our expectations and assumptions to those of an external stakeholder (the client), as well as working with software tools and systems that we are unfamiliar with in roles of mixed consulting and implementation. We are grateful for the opportunity we've received to learn valuable real-world skills, and we believe that in return for this experience, we've been able to provide a solution that adds value for the client. We will carry what we've learned forward into our future careers. Likewise, we believe the solution we've provided for the client will provide a path toward building a data collection that can help them to identify their strengths and potential improvement areas, highlight their successes, and work more effectively with partner organizations and other engaged stakeholders.

Recommendations for Use of Solution

Getting Started

While our goal has been to ensure that the solution has very few needs for set-up and is immediately usable, there are some steps we recommend the client take to ensure they get the most out of the solution.

Storage and Backup

As mentioned previously, we recommend using OneDrive / SharePoint for storage of the Access database. This allows users in multiple locations to access it securely and allows users to check in and out the file so that they do not accidentally overwrite someone else's work. It also protects the database from loss due to damage or other issues with a computer hard drive, if it were stored on an individual desktop computer.

We recommend creating a designated folder for making frequent backups of the database, labeled with the date the file was saved. This will allow you to restore your data without too much loss if anything happens to the file.

Review and Update

The team prioritized making a user-friendly interface and ensuring data could be stored and updated in the structure; however, some data may still need to be updated in the database as time did not allow us to check all the data for correctness, remove duplicates, etc. The following steps should be taken to ensure the integrity of your data:

1. **Update workshop attendance in the Attendance table**. The Excel file containing the Workshop Attendees sheet currently has all the cross-

reference information, in the columns to the right of attendees' personal information. The Attendance table has been designed as two simple drop-down menus which can be updated to associate a person with a workshop. It has been partially updated, but full cross-reference information was not input due to time restrictions. This process is easy but tedious.

2. **Check the Person table for duplicates**. Due to the fact that most records in the Person table include workshop attendance cross-references, we tried to avoid deleting any lines even if they appeared to be duplicates, as more information would need to be merged together across another table (the Attendance table). If you need to delete a duplicate, ensure that the remaining record contains all the relevant cross-reference information.

3. **Hide ribbons and left hand menu**. For general day-to-day use, the point and click interface should be sufficient for viewing records, pulling reports, and updating the database. As a rule, it is not recommended for end users to interact with the tables themselves when using an Access database. Instead, it's best to do everything through forms and dialogues in order to ensure no data is corrupted.

Day to Day Use

The database interface is designed to be relatively intuitive; however, for instances where a user may need direction, a help file has been provided. We recommend storing the help file alongside the database in the same SharePoint folder.

For users to know: Ensure that users are aware they should not be accessing the inner workings of the database - such as tables or queries - other than through the form interface, and to request help from a more technical person if they need to do something they can't find a way of doing with the forms provided.

Ensure users are aware of SharePoint dynamics which require check out / check in of files in order to avoid synchronization mishaps.

Tableau Visualizations

Installation

1. Go to https://public.tableau.com/en-us/s/download

2. Type your Email ID and click the "download the app" option

3. Once after the file download, click the file and install the software in your respective system.

4. Once after installation, the Tableau Public Book1 page will pop-up

5. This step is of utmost importance, where you will have to connect your Tableau with the data-set. From the left display pane, click on the "Microsoft excel" and find the location of your local file where the data set is stored.

6. After connecting to the data source, the above pane gets generated. Connections menu displays all the possible connections (data sources) which were connected to the Tableau public workbook. The sheets menu displays all the worksheets which are available in the respective excel data source. For representation purpose, I am using the aggregated data set(External info, workshop summary and Long term outcome) and going to create different types of visualizations.

7. Drag "ultimate" sheet from the left pane and drop in the "drag sheets here" section.

Now the data associated with the ultimate sheet gets displayed below the source. Now

"Go to worksheet" icon gets highlighted close to the sheet1. Click on the sheet1 menu.

8. The data from the worksheet is automatically categorized by Tableau as

Dimensions and Measures. In general, Dimensions are the non-numerical quantities

under which the plots can be categorized. For example, state, district, town or region

are some of the example of Dimension. Whereas Measures are the numerical quantities

under which we can view the aggregated and non-aggregated level of information.

9. From the dimensions pane, drag and drop "town" to the columns section. Similarly,

from the measures drag and drop the "score" to the rows section.

10. Now, drag and drop the "regions" from the dimensions pane to the "color" pane

under the "Marks" section. Now the legends will appear towards the right section of the

Sheet1 with all the regions color coded. The colored legends acts as a filter where

single clicking on any one of the region will isolate its associated town from the overall

plot.

11. Click on the "label" icon from the and enable "show mark labels" to display all the

score values of the towns.

12. There is an alternative way to display filters in the sheet. Drag and drop "regions"

from the dimensions pane to the filters pane (adjacent to dimensions pane). The pop-up

list displays all the values from the region. Now click on apply and ok. Now the "region"

icon gets populated in the filters section.

13. Right click on the region icon in the filters section and click on "show filter" drop the list. This will display a filter on the right side of the plot. Now, the regions can be unchecked and specific set or individual region shall be isolated and viewed from the plot.

14. From the dropdown from the Marks panel, you can select different types of visualization such as Bar, Line, Area, Circle, Scatter, Pie etc.

Steps to Create a Bubble chart

15. From the dimensions pane, click on the down arrow from the "town" icon, from the dropdown hover around "geographic role" and click on "city" from the list. After this step the icon symbol gets changed from "abc" to a globe icon. This step emphasizes a change in the data type from a character to a geographic entity.

16. Now, drag and drop the town icon from the dimensions pane to the sheet area. After this step, the map of the Massachusetts is automatically created in the background with bubble markers locating certain towns from the state.

17. After this step, there will be some "unknown" values that gets populated towards the bottom-right end of the plot. Click on the unknown icon.

18. Now, click on the Edit locations from the pop-up. Now from the new pop-up, click on state/province drop down from the display menu. Choose the fixed button and from the state, search and selected "Massachusetts".

19. Then click OK, Now more bubble data points gets populated on the Massachusetts map.

20. From the measures panel, drag and drop "score" on the sheet. Now the bubble chart shapes get realigned based on the volume of scores across different towns around the state.

21. From the dimensions pane, drag and drop the "region" icon to the color box in the Marks panel. This step distinguishes towns based on Region under a color-coded pattern. Also the legends gets generated towards the right of the sheet using which specific regions can be clicked, isolated and analyzed.

How to open a workbook using Tableau Public

The workbooks which are created using Tableau public gets saved in the Tableau public webpage https://public.tableau.com/en-us/s/. It is vital to go to this webpage and to create a user account for your works to be shared and saved. Once after login, the workbook shall be downloaded from Tableau public and can be saved in your local system. The local workbook file can be opened anytime and updates shall be made. The file extension will be in .twbx format.

Important Information

Tableau public offers a wide range of plots, where each type of plot conveys its intended meaning. As the product is public, it is not possible to save the work in a workbook. Unlike Tableau desktop, a licensed product which can save the reports and dashboards under a single file (workbook), Tableau public could not accomplish the same. Instead, if we click on the save icon, the reports get saved in the public domain. Public domain is a portal offered by Tableau, where any Tableau public enables user can access reports created by any other user. Hence, before saving the reports make

sure that the confidential attendees information such as Name, Mobile number, Address is not included anywhere in the visualization.

Future Product Improvement

There may be a point when you decide to engage an IT professional to make improvements to your existing database structure. Here are some ideas for what they may be able to do:

- **Additional Reports**: There is a multitude of possibilities, given sufficient time, for coding items such as basic visualizations and other helpful formatting into reports, or for creating more complex queries which may provide new revelations from your existing data.

- **User Experience Refinement**: Keep a log of issues experienced by employees in your organization when using the database; for example, if users regularly have a problem finding one of the buttons on the input forms provided, this may indicate the position or look of the button should be changed to make the interface more intuitive. These items can be provided to a database management or IT professional when engaging them in the process of improving the interface.

- **Linking to SharePoint**: While you will likely be storing your database file on SharePoint, Access also has the capability of interacting with SharePoint on a more sophisticated level, via publishing to SharePoint or linking to SharePoint lists. This requires some amount of preparation and planning, and you should consult your organization's IT professionals before attempting to do this; however, it is good to be aware that the capability exists. Advantages provided

by this include easier collaboration and syncing (no need to check in / out the entire database file when multiple people may be updating it), built-in data backup and alerts to track changes made.

Recommendations for the Future

Objective:

To create a data management solution and a visualization platform for the client (Mass Audobon) data and derive useful information and leverage meaningful insights.

Tools:

- Data Management Solution – Snowflake Cloud Datawarehouse
- Data Visualization Tool – Tableau Software

Data Management tools:

Snowflake Cloud Datawarehouse: Snowflake has the capability to store and access huge data sets. It is a SQL enabled data warehousing platform, where we can upload data directly from the excel sheets. It can handle multiple users at the same time and process customized requests. With Snowflake, we can create table structures in accordance with the structure of records(columns) present in the excel sheet. We can create different table structures catering to the nature of the data present in sheets such as workshop summary stats, workshop attendees, workshop speakers etc., Analytical tables can also be created having the aggregated level of information on workshop towns and attendee's information. Using SQL queries, we can retrieve data at faster pace and download information in varied formats.

Cost – Standard Version – $40 per month and $2 per hour(access)

Snowflake Architecture
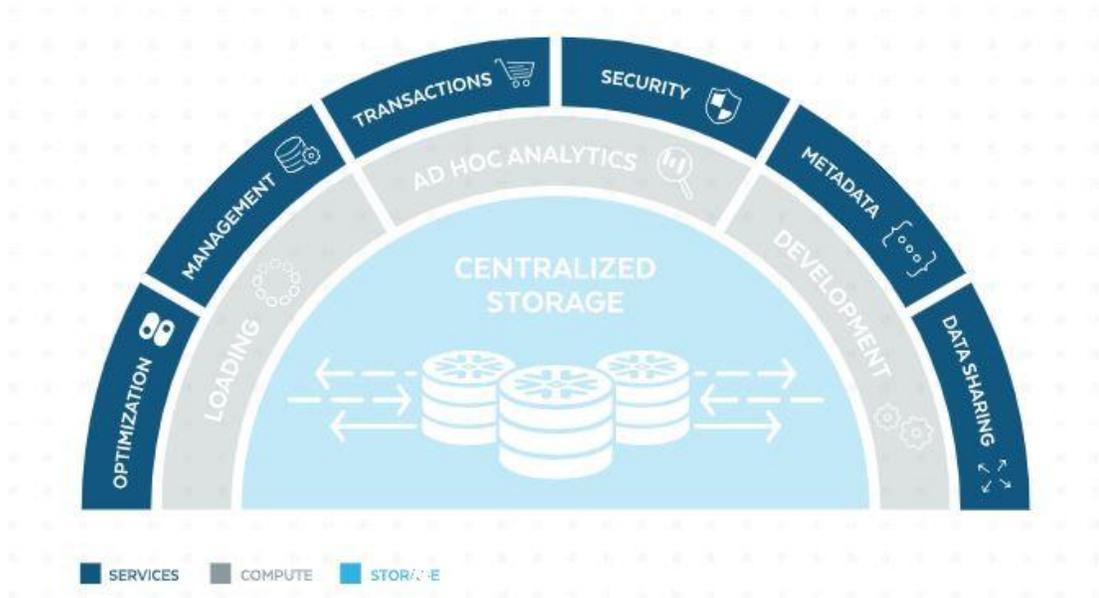
Figure 1. Screenshot taken by Author about the Snowflake architecture retrieved on October 29th 2018 from, https://www.snowflake.com/product/architecture/

Amazon Redshift: Redshift is a cloud-based representation of a traditional data warehouse. Redshift requires computing resources to be provisioned and set up in the form of clusters, which contain a collection of one or more nodes. Each node has its own CPU, storage, and RAM. A leader node compiles queries and transfers them to compute nodes, which execute the queries. On each node, data is stored in chunks, called slices. Redshift uses a columnar storage, meaning each block of data contains values from a single column across a number of rows, instead of a single row with values from multiple columns. Redshift uses an MPP architecture, breaking up large data sets into chunks which are assigned to slices within each node. Queries perform

faster because the compute nodes process queries in each slice simultaneously. The

Leader Node aggregates the results and returns them to the client application. Client

applications, such as BI and analytics tools, can directly connect to Redshift using open

source PostgreSQL JDBC and ODBC drivers. Analysts can thus perform their tasks

directly on the Redshift data. Redshift can load only structured data. It is possible to

load data to Redshift using pre-integrated systems including Amazon S3 and

DynamoDB, by pushing data from any on-premise host with SSH connectivity, or by

integrating other data sources using the Redshift API.

Cost – Standard Version - $1000 per year

Redshift Architecture
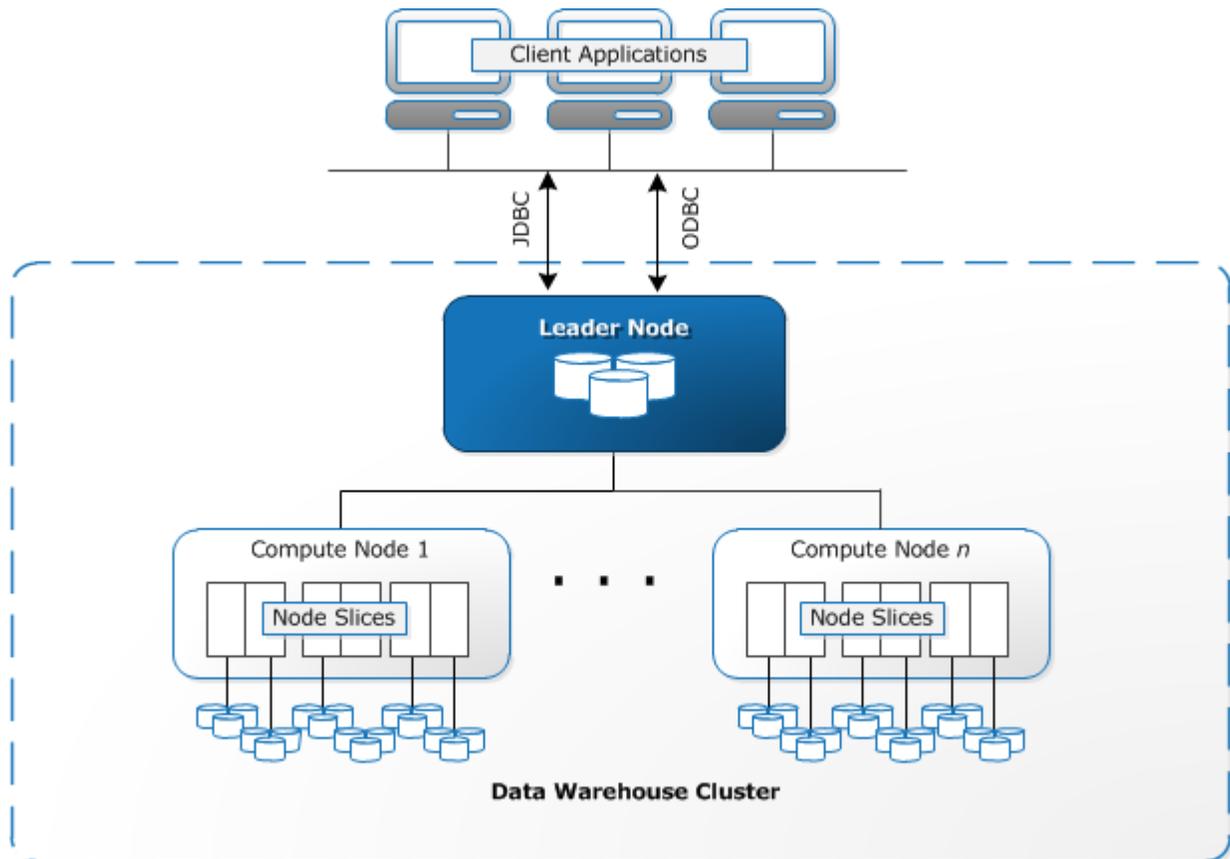
Figure 2. Screenshot by Author about the Redshift architecture retrieved on 29th 2018

from, https://panoply.io/data-warehouse-guide/data-warehouse-architecture-traditional-vs-cloud/

Google BigQuery: BigQuery's architecture is serverless, meaning Google dynamically

manages the allocation of machine resources. All resource management decisions are,

therefore, hidden from the user. BigQuery lets us load data from Google Cloud Storage

and other readable data sources. The alternative option is to stream data, which allows

developers to add data to the data warehouse in real-time, row-by-row, as it becomes

available. BigQuery uses a query execution engine named Dremel, which can scan

billions of rows of data in just a few seconds. Dremel uses massively parallel querying

to scan data in the underlying Colossus file management system. Colossus distributes

files into chunks of 64 megabytes among many computing resources named nodes,

which are grouped into clusters. Dremel uses a columnar data structure, like Redshift. A

tree architecture dispatches queries among thousands of machines in seconds.

Cost – Standard Version – less than $1000 per year

BigQuery Architecture
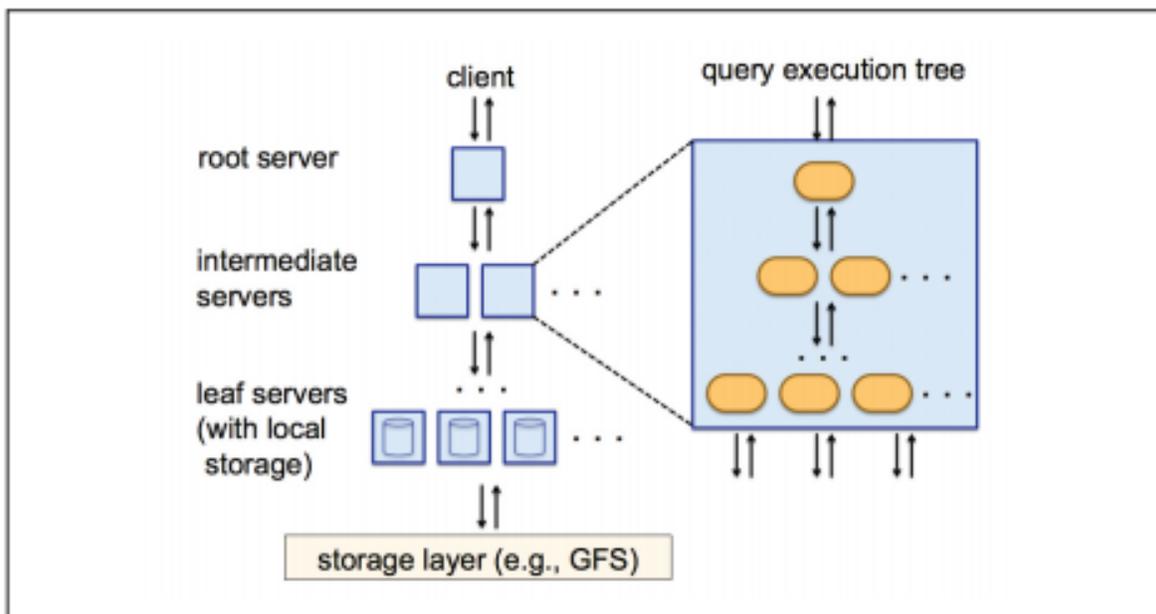
Tree architecture of Dremel

Figure 3. Screenshot by Author about the Bigquery architecture retrieved on October

29th 2018 from, https://panoply.io/data-warehouse-guide/data-warehouse-architecture-

traditional-vs-cloud/

Data Visualization platform

Tableau: Tableau, a data visualization platform, can pictorially represent data in meaningful ways. Tableau can create heat maps, bar graphs, line chart, scatter plots etc, where each plotting mechanism caters to specific needs. Tableau has an easy user interface where we can drag and drop fields to create customized visualizations. It has multiple functionalities where we can aggregated metrics, create customized fields, apply filters, differentiate based on colors etc,. One of the objectives proposed by the client is to create a heat map and Tableau has an excellent embedded functionality which creates heat map based on different towns and counties related to every state in US.

Cost – Differs according to the product purchase, $1000 for desktop version, $4000 per year for 5 users Tableau online solution

A peek into various types of plots using Tableau

Figure 4. Screenshot by Author about various visualization plots using Tableau retrieved on Oct 29th 2018 from, https://www.homeagency.co.uk/blog/data/battle-of-the-dashboards/

R Shiny: R Shiny is a web-based open source application product from R programming language. R also has an in-built functionality to create flex dashboards for business and customers. Like Tableau, R Shiny can also be used to create heat maps, bar graphs, line chart, scatter plots etc. To build R Shiny apps, developers need to have a basic understanding of R programming language. Unlike Tableau, RShiny doesn't have drag and drop features to create the required kind of visualization. But, it provides an excellent feature where the visualizations can be viewed along with the associated data.

Reactive plots can also be developed where selecting a specific set of data points from the plot displays its associated data from the table.

Cost – Open source (free) desktop, RStudio commercial $995 per user, RStudio connect $14,995 per year (upto 20 users)

A peek into a specific type of plot using RShiny
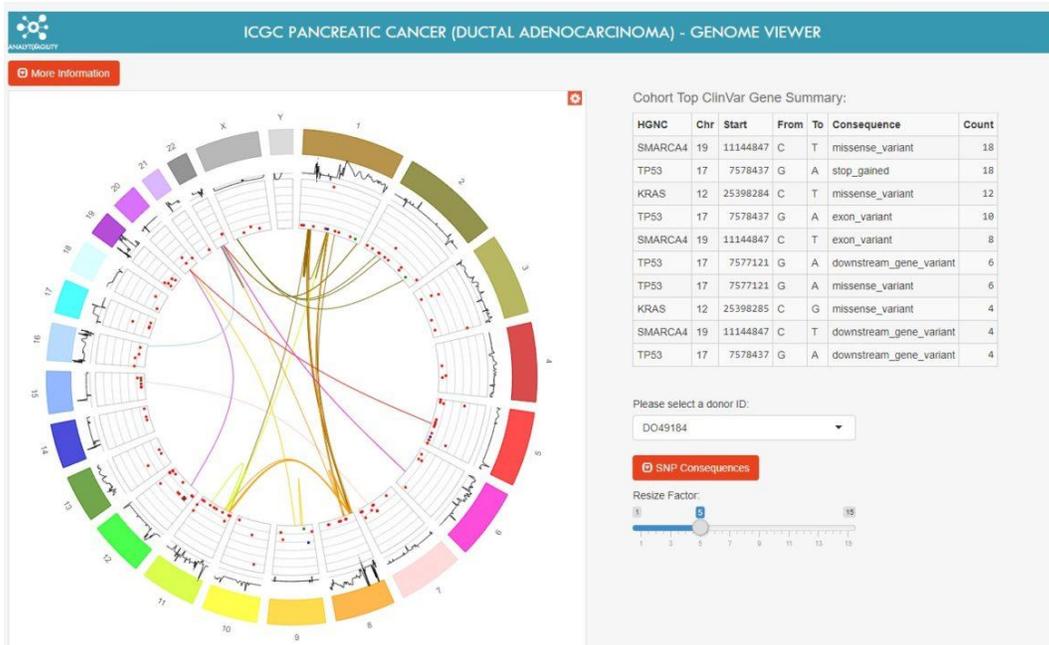
Figure 5.Screenshot by Author about data visualization plot using R Shiny retrieved on Oct 29th 2018 from, https://www.homeagency.co.uk/blog/data/battle-of-the-dashboards/

Types of plots using RShiny
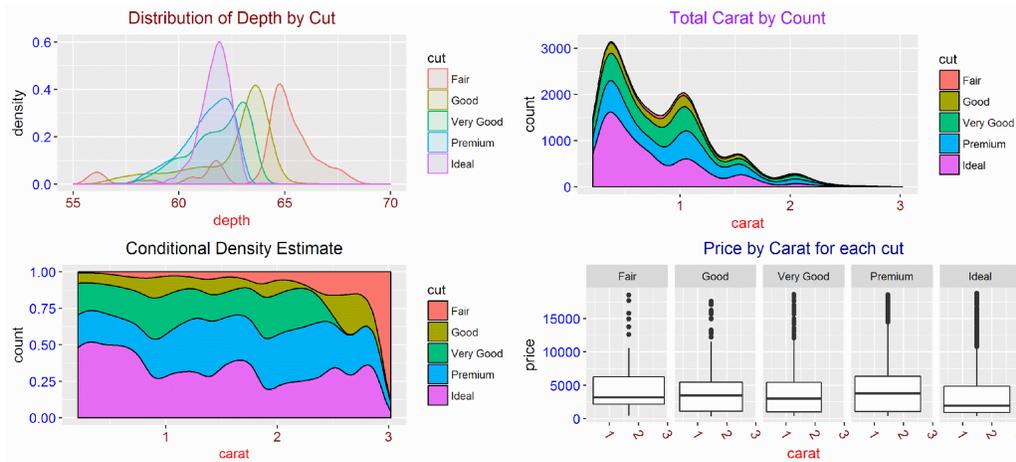
Figure 6. Screenshot by Author about various visualization plots using R Shiny retrieved on Oct 29th 2018 from, https://www.homeagency.co.uk/blog/data/battle-of-the-dashboards/

Data Studio: Data Studio, part of Google Marketing Platform and closely integrated with Google Cloud, allows you to easily access data from Google Analytics, Google Ads, Display & Video 360, Search Ads 360, YouTube Analytics, Google Sheets, Google BigQuery and over 500 more data sources, both Google and non-Google, to visualize and interactively explore data. It allows you to easily share your insights with anyone in your organization. And beyond just sharing, Data Studio offers seamless real-time collaboration with others who are located around the globe.

Cost – Open source (Free)

A Peek into types of plots using Data Studio
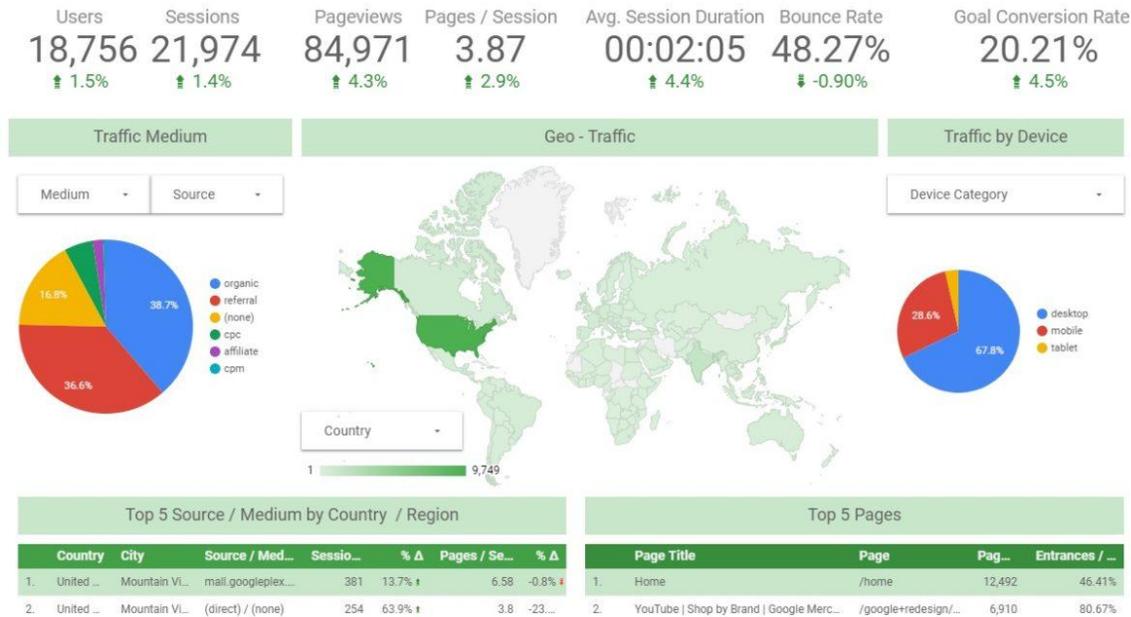
Figure 7. Screenshot by Author about various visualization plots using Data Studio retrieved on October 29th 2018 from, https://www.homeagency.co.uk/blog/data/battle-of-the-dashboards/

Comparison between the functionalities of Tableau, RShiny and Data Studio

1. **CHART AND GRAPH TYPES**

- Tableau does relatively well here, offering a wide range of chart types to meet pretty much all dashboarding needs.

- Shiny excels in this category, with a wealth of packages from the past 20 years, including the incredible ggplot2 and plot.ly to make pretty much any visualization imaginable.

- Data Studio has the least variation here, offering standard chart types, but it is important to remember that most charts used in dashboards fit into this category.

These include bar or stacked bar charts, time series charts, pie charts, maps, scorecards, scatter, bullet and area charts.

2. DATA CONNECTORS

- Tableau again does reasonably well, offering many built-in connections including many Google products such as Google Analytics and BigQuery. However, it does need third-party tools to connect to AdWords and YouTube Analytics.

- Shiny can connect to just about any data source by using either a premade package or custom API calls.

- Data Studio can of course connect to all Google products as well as MySQL and PostgreSQL. Other data sources may have to be routed through BigQuery or Google Sheets, but this is easily done in most cases.

3. EASE OF USE

- Tableau is simple to pick up, offering a simple drag and drop format, and a relatively natural workflow.

- Shiny requires knowledge of the R programming language which is cumbersome at first, and perhaps not very intuitive without prior knowledge of similar language. Once learnt however, it becomes extremely powerful.

- Data Studio has a natural workflow which is easy to pick up from the off – a good indicator is how rarely a Google search is needed when attempting something new.

4. DATA MANIPULATION AND ANALYSIS

- Tableau has basic data manipulation functions allowing users to pivot columns, create groups and filter at source. It also offers basic analysis features.

- Shiny has arguably the best data manipulation and analysis tools. It is incredibly powerful for data manipulation or "wrangling" – the process of getting data into the correct form for analysis. It also offers many statistical modelling and advanced forecasting packages.

- Data Studio has no built-in features for data manipulation or analysis, and thus data must be wrangled elsewhere such as BigQuery, and analysis then completed by exporting the data.

5. SPEED

- All three of these products offer high speed and efficient processing of data, but for larger amounts of data, none can compete with the super-fast packages R offers such as "DT (datatable)".

6. REPRODUCIBILITY AND REPEATABILITY

- In both Data Studio and Tableau, once a dashboard is decommissioned or deleted, it would have to be built again from scratch.

- In R's Shiny, since it is purely text based all that would be needed to be done is to re-run the script, which massively helps with repeatability and scalability. For

instance, if two dashboards share a similar setup, the script may simply be

tweaked to accommodate the changes.

Steps to create a plot using Tableau

This tutorial walks you through some of the most common tasks you might

perform when creating maps in Tableau. You'll learn how to connect to and join

geographic data; format that data in Tableau; create location hierarchies; build and

present a basic map view; and apply key mapping features along the way.

Step 1: Connect to your geographic data

Geographic data comes in many shapes and formats. When you open Tableau

Desktop, the start page shows you the connectors available in the left Connect pane.

These are how you will connect to your data.

You can work with geographic data by connecting to spatial files, or you can

connect to location data stored in spreadsheets, text files, or on a server.

Spatial files, such as a shapefile or geoJSON file, contain actual geometries

(points, lines, or polygons), whereas text files or spreadsheets contain point locations in

latitude and longitude coordinates, or named locations that, when brought into Tableau,

connect to the Tableau geocoding
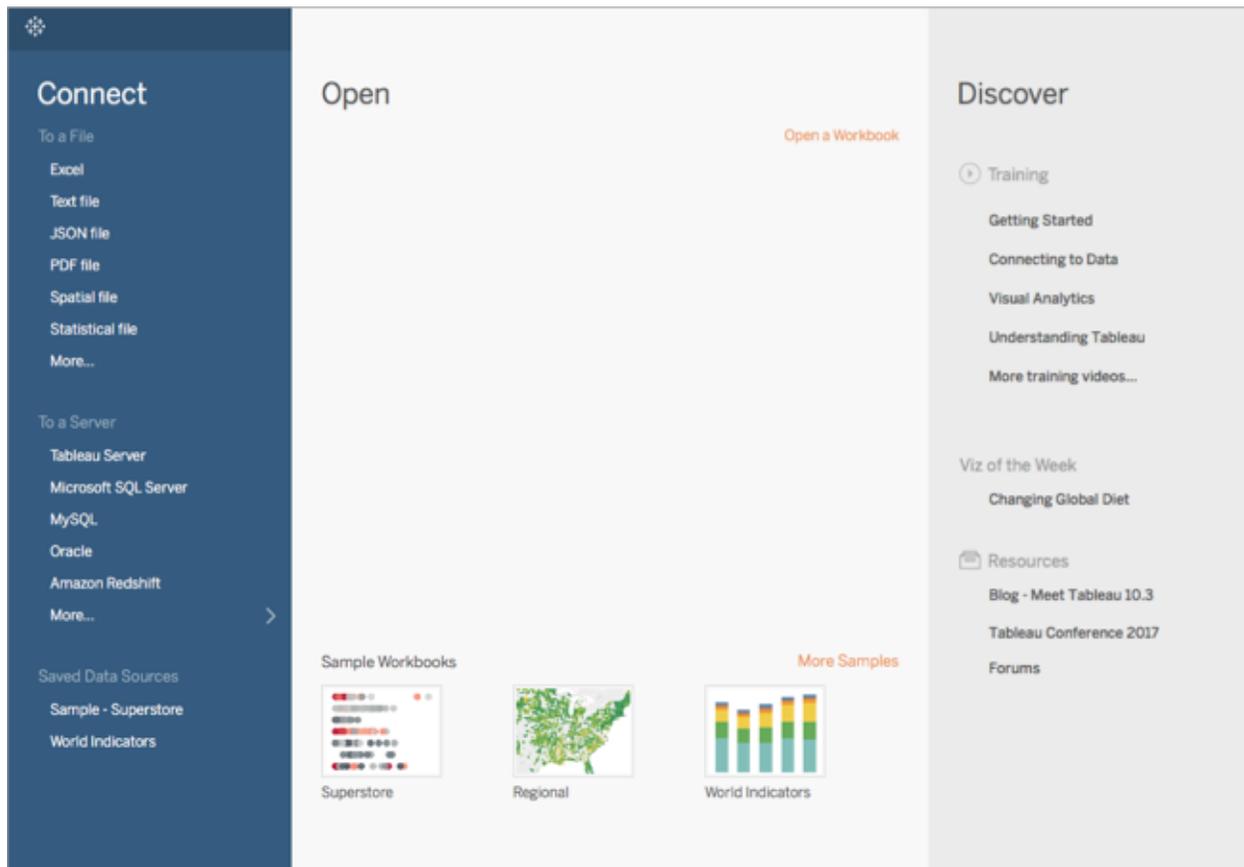
Figure 8. Screenshot by Author about connecting to the source file in Tableau retrieved

on 29th October 2018 from, https://onlinehelp.tableau.com/current/pro/desktop/en-

us/buildexamples_maps.html

For this tutorial, you are going to connect to an Excel file that comes with

Tableau Desktop. It contains location names that Tableau can geocode. When you build

a map view, the location names reference the geometries stored in the Tableau Map

Service based on the geographic role you assign to the field. You'll learn more about

geographic roles later in this tutorial.

1. Open Tableau Desktop.

2. In the Connect pane, click Excel.

3. Navigate to Documents > My Tableau Repository > Data Sources, and then open the Shaping Master Document file.

Once you connect to the data source, your screen will look like this:
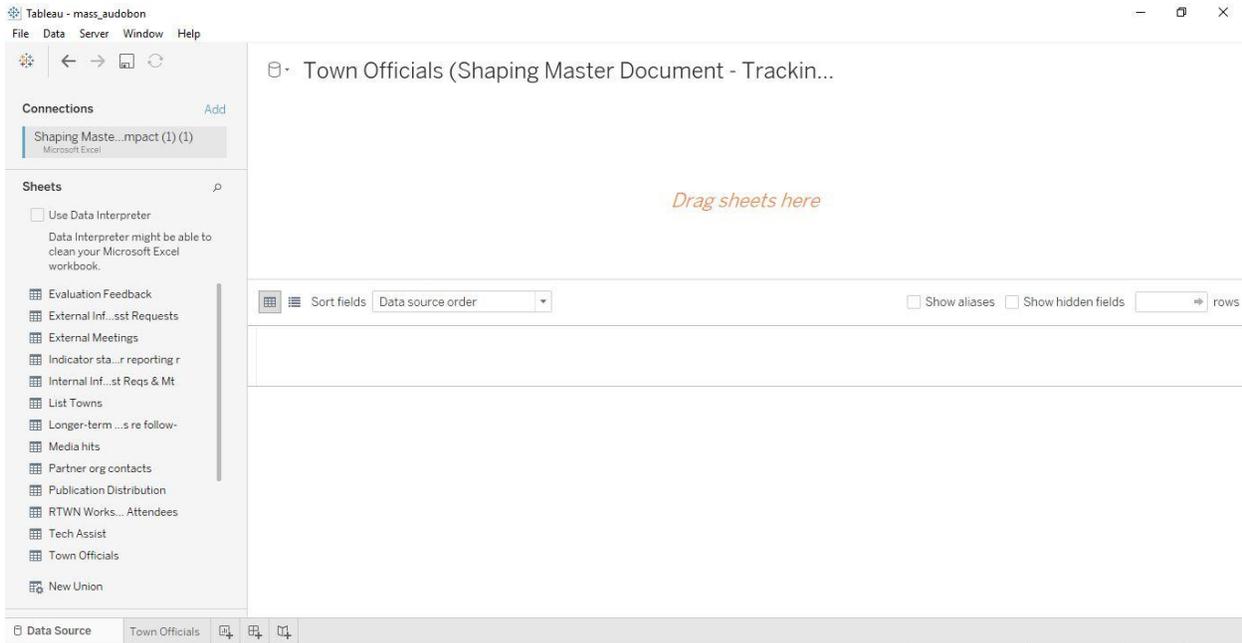
Figure 9. Screenshot by Author about connecting to the data source in Tableau retrieved on 29th October 2018 from,

Step 2: Drag sheets to the work pane area

From the list of sheets displayed in the left section of the pane, drag Workshop_Attendees sheet from left pane and drop it on the work pane. After dropping the sheet, the data table gets generated and displayed in the pane below to the work area. In similar fashion, any sheet from the Master Shaping Document shall be pulled and dropped into the work area. Tableau also has an interesting feature to join two or

more sheets together. If two sheets are dragged and dropped into the work area, then

the sheets are joined together with a common field (ID or Name or email ID or Mobile

Number or Area etc,.). Then the data table that comes below the work pane has all the

fields from both the sheets. When you connect to geographic data, Tableau also

assigns data roles to all your columns. A column can be a dimension or measure. In

most cases, your latitude and longitude columns should be measures. For special

cases, such as if you want to plot every location in your data source on a map without

the ability to drill up or down a level of detail (such as from City to State), they can be

dimensions. A great example of this is a point distribution map. The rest of your

geographic data should be dimensions.
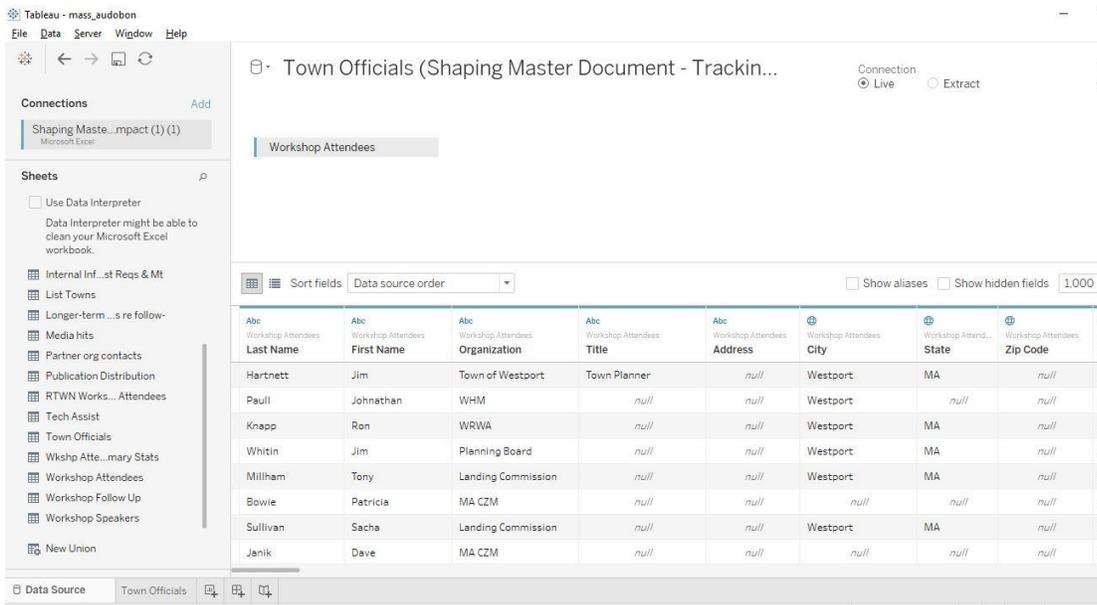
Figure 10. Screenshot by Author about exporting excel sheet to the work pane area

retrieved on 29th October 2018 from,

https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_maps.html

Step 3: Working with Worksheets

On the Data Source page, click Sheet 1.

Your workspace updates to look like this:
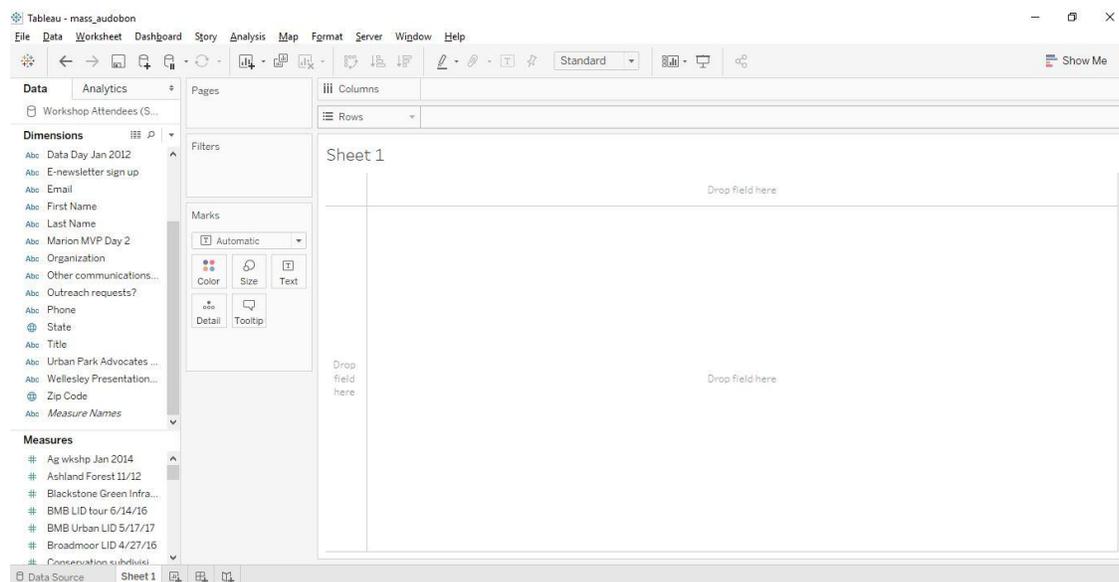
Figure 11. Screenshot by Author about experience with worksheet in Tableau retrieved

on 29th October 2018 from, https://onlinehelp.tableau.com/current/pro/desktop/en-

us/buildexamples_maps.html

This is called a worksheet, and it is where you will build your map. On the left-

side of the screen is the Data pane. All the columns in your data source are listed as

fields in this pane. For example, State and City. These fields contain all the raw data in

your columns. Note that Tableau has generated a Latitude and Longitude field (Latitude

(generated) and Longitude (generated)) below the Measures pane. This is because that data sheet has geographic level of information such as State, City and Zip Code.

The fields in the data pane are divided into measures and dimensions. The fields placed in the Dimensions section of the Data pane are often categorical data, such as Date and ID, while the fields placed in the Measures section of the Data pane are often quantitative data, such as Quantity.

Step 5: Creating a geographical hierarchy

Now that you are in the worksheet space, you can create geographic hierarchies. This is not required to create a map view but creating a geographic hierarchy will allow you to quickly drill into the levels of geographic detail your data contains, in the order you specify.

To create a geographic hierarchy:

1. In the Data pane, right-click the geographic field, State, and then select Hierarchy > Create Hierarchy.

2. In the Create Hierarchy dialog box that opens, give the hierarchy a name, such as Geo_data, and then click OK. At the bottom of the Dimensions section, the Mapping Items hierarchy is created with the State field.

3. In the Data pane, drag the City field to the hierarchy and place it below the State field.

4. Repeat step 3 for the zip code field as well.
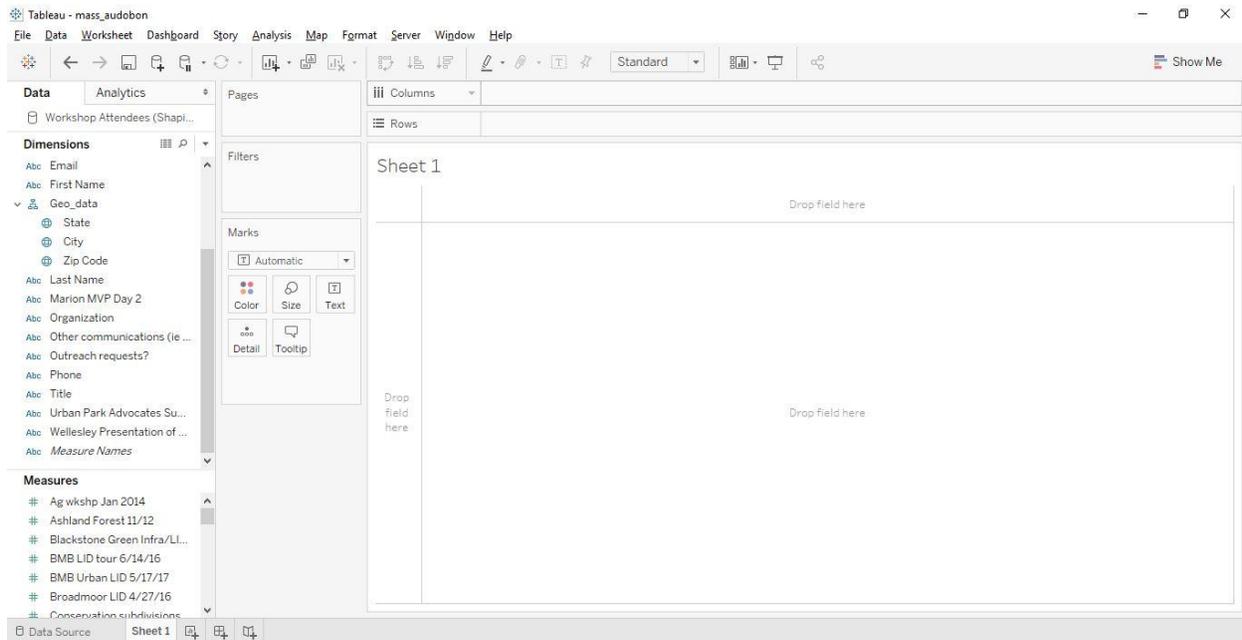
Figure 12. Screenshot by Author about creating a geographical hierarchy in Tableau

retrieved on 29th October 2018 from,

https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_maps.html .

Step 6: Building a basic map

1. From the data pane(dimensions), drag and drop the city icon to the sheet area.

    Now you can see all the all the cities being spotted from the United States map.

    In the search bar in the sheet area, type "Massachusetts" and select the state.

    Now the Massachusetts map is generated with the data points locating the cities.

2. Below the map, there will be an icon displaying the unknown cities. Click on the

    unknown and select "Massachusetts" state from the state dropdown menu.

3. Drag the "number of records" from the measures pane and drop it in the size icon

    adjacent to the left side of the sheet. Now click on the size icon and increase the

    size of the scattered data points.

4. Now, drag and drop the city icon from the dimension and drop it on the color icon adjacent to the size icon. After this step, a basic map is created with the concentrated set of data points focusing on the total number of participants from different cities.

The map should like this
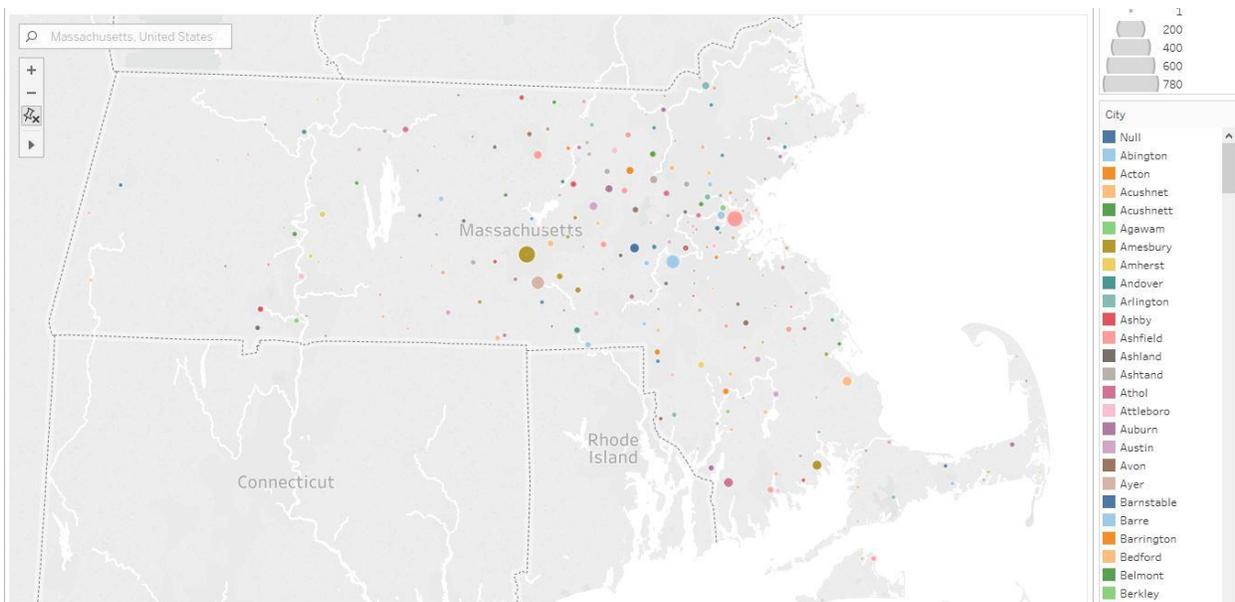
Figure 13. Screenshot by Author about creating a bubble map in Tableau retrieved on 29th October 2018 from, https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_maps.html

By following these steps, we can create a bubble map with the density of participants from different cities. This visual representation can aid in determining the best performing cities from the underlying data.

Steps to load data using Snowflake

Step 1: Creating table structure

Even before using snowflake, you must create a table structure which resonates the same structure of the records exists in the excel sheet.

Consider the below mentioned is the structure of records present in excel sheet

First Name	Last Name	Attendee ID	Email ID	Town
John	Oliver	1001	joliv@gmail.com	Natick
Pepper	Potts	1002	peppot@gmail.com	Malden
Peter	Parker	1003	petpar@gmail.com	Canton
Bruce	Wayne	1004	bruway@gmail.com	Bedford

Note: The attendee name and other information are made-up, and it doesn't reflect original attendee information from the Shaping Master Document

Based on the above information, a table structure should be created using SQL language.

create or replace TABLE MASS_AUDOBON (

firstName VARCHAR(50),

lastName VARCHAR(50),

attendeeID NUMBER(38),

emailID VARCHAR(50),

Town STRING

);

Step 2: Creating warehouse and database

Once after logging into snowflake solution, click on warehouses icon on the panel on top. Click on create icon from the section.
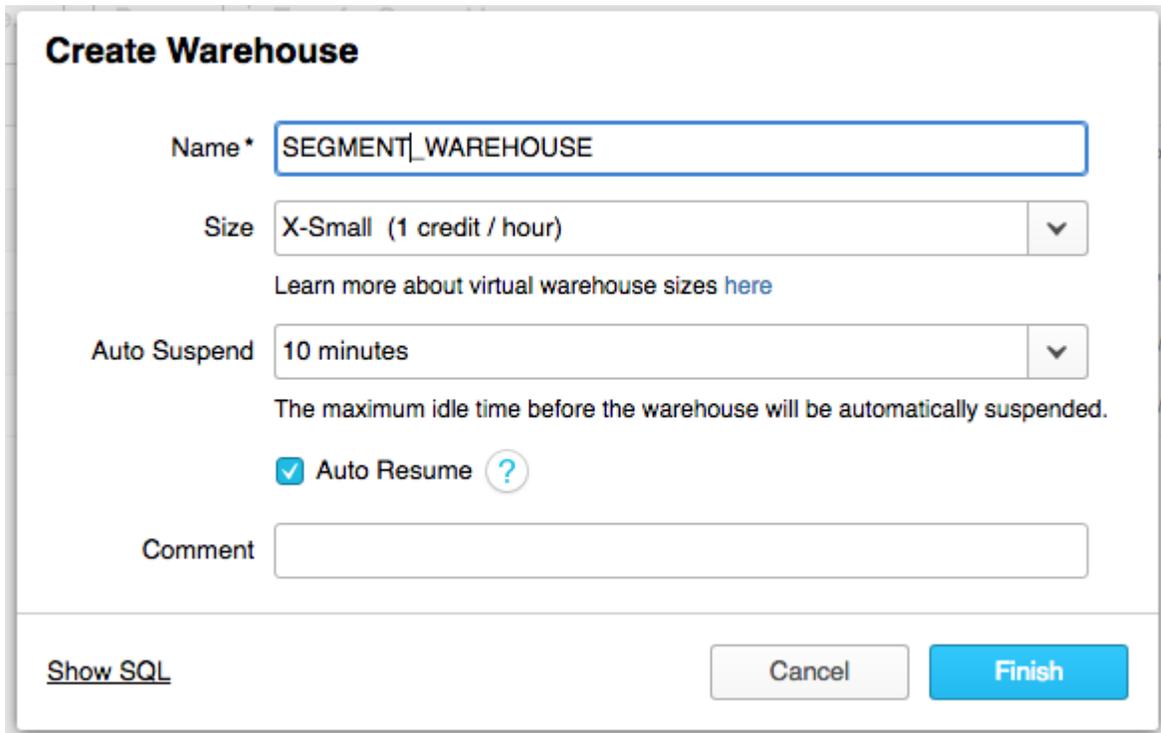
Figure 14. Snapshot by Author about creating a warehouse using Snowflake retrieved on 29th October 2018 from, https://segment.com/docs/destinations/snowflake/

Provide warehouse name, size, auto-suspend (billed every hour) and comment (if needed) on the display box. After furnishing this information, a warehouse is created to house Mass Audobon records. Now, you can click on database icon from the top panel and click on create icon.
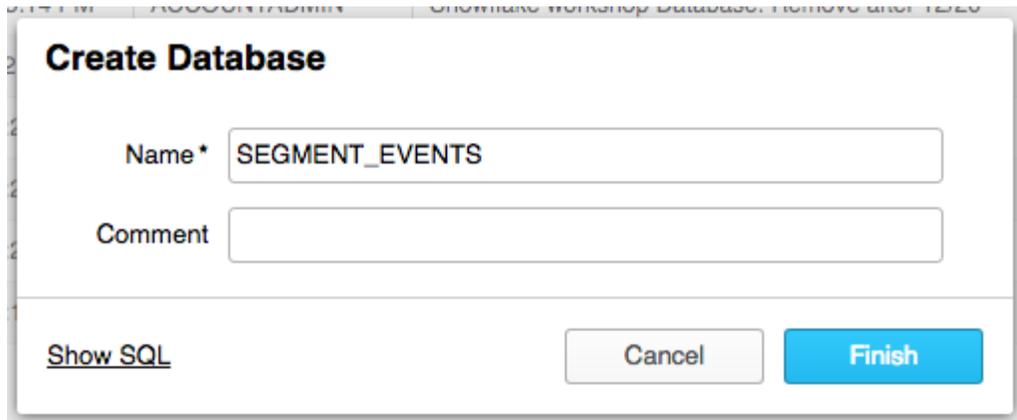
Figure 15. Snapshot by Author about creating a database using Snowflake retrieved on 29th October 2018 from, https://segment.com/docs/destinations/snowflake/

You shall provide a name for the database and click on finish. This step completes both the warehouse and database creation.

Step 3: Create table in database

Now, click on the worksheet icon from the top panel. It opens up an empty work area. In that space, copy and paste the table structure from step 1 and execute the query. This step finalizes the table structure creation.
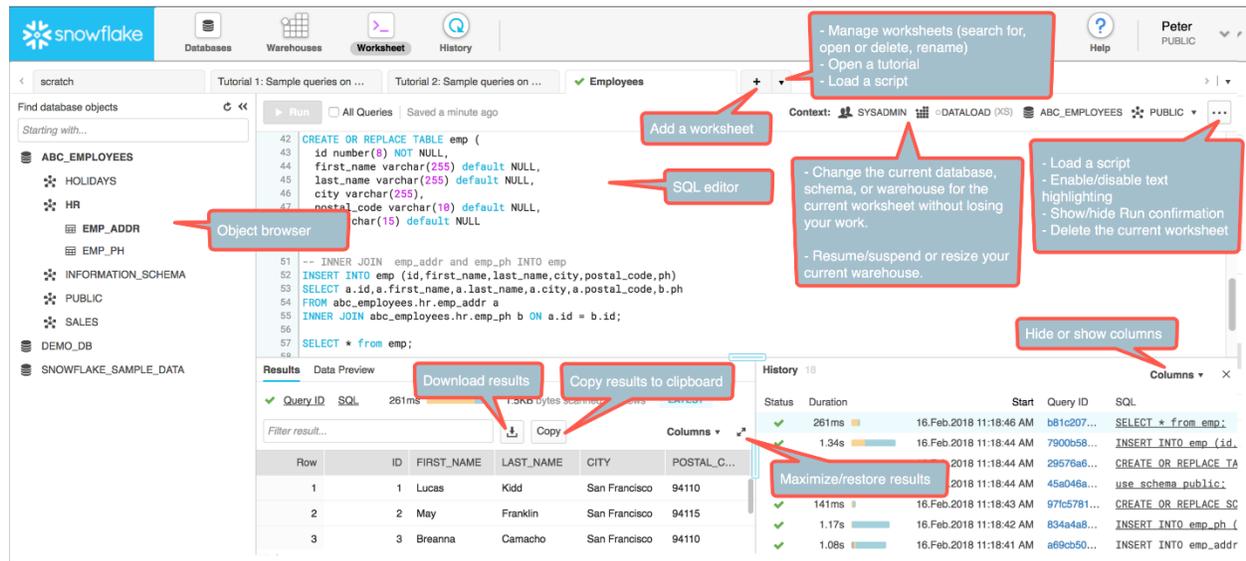
Figure 16. Snapshot by Author about creating a table using Snowflake retrieved on 29th October 2018 from, https://docs.snowflake.net/manuals/user-guide/snowflake-manager.html

Step 4: Loading data from excel sheets

Click on the database icon from the top panel and click the database name which was created in Step 2. Inside the database, you shall find the name of the table which was created in step 3. Now, on the top right corner in the worksheet, click on the load data icon. Now a pop-up screen opens with a provision to locate the file location in the computer. Browse and locate the file(excel sheet with data) and click on the next button.
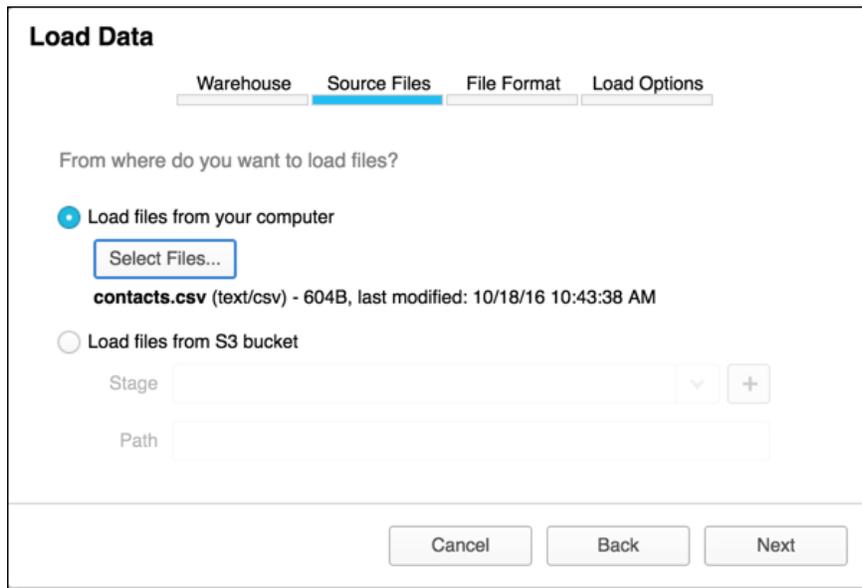
Figure 17. Snapshot by Author about process of loading data files using Snowflake

retrieved on 29th October 2018 from, https://docs.snowflake.net/manuals/user-

guide/data-load-web-ui.html

In the next step, you have to select the file format under which the data needs to

be loaded in the table. You have to create a new file format catering to the format of the

data present in the excel sheet. Input all the required fields in the file format section and

click next.

Figure 18. Snapshot by Author about the process of creating a file format using

Snowflake retrieved on 29th October 2018 from,

https://docs.snowflake.net/manuals/user-guide/data-load-web-ui.html

Now, you have to choose from a list of inputs on what operation to perform if the

load experiences an error. Click on the required option from the list and click on load.

Now the data is loaded from the excel sheet to the database.
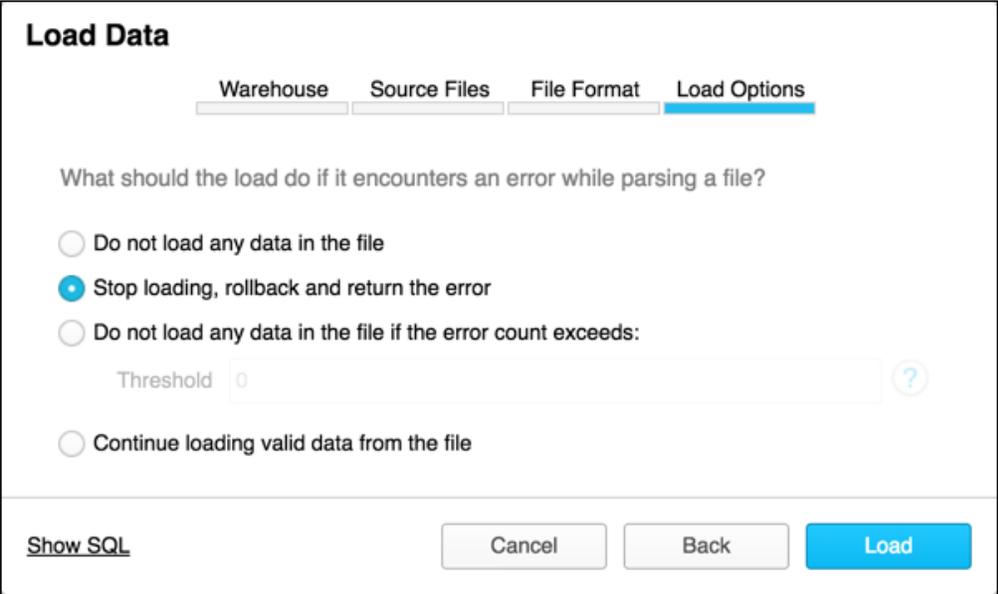
Figure 19. Snapshot by Author about the process of different load options using

Snowflake retrieved on 29th October 2018 from,

https://docs.snowflake.net/manuals/user-guide/data-load-web-ui.html

Step 5: Data Management and Analysis

Snowflake data warehouse is based on SQL coding. With snowflake, we can

perform cross referential analysis between different set of tables having a common

column such as Name, ID, Mobile number etc., All the SQL queries can be made to run

inside the worksheets. After each query is executed, its associated data gets displayed

below the workspace area.
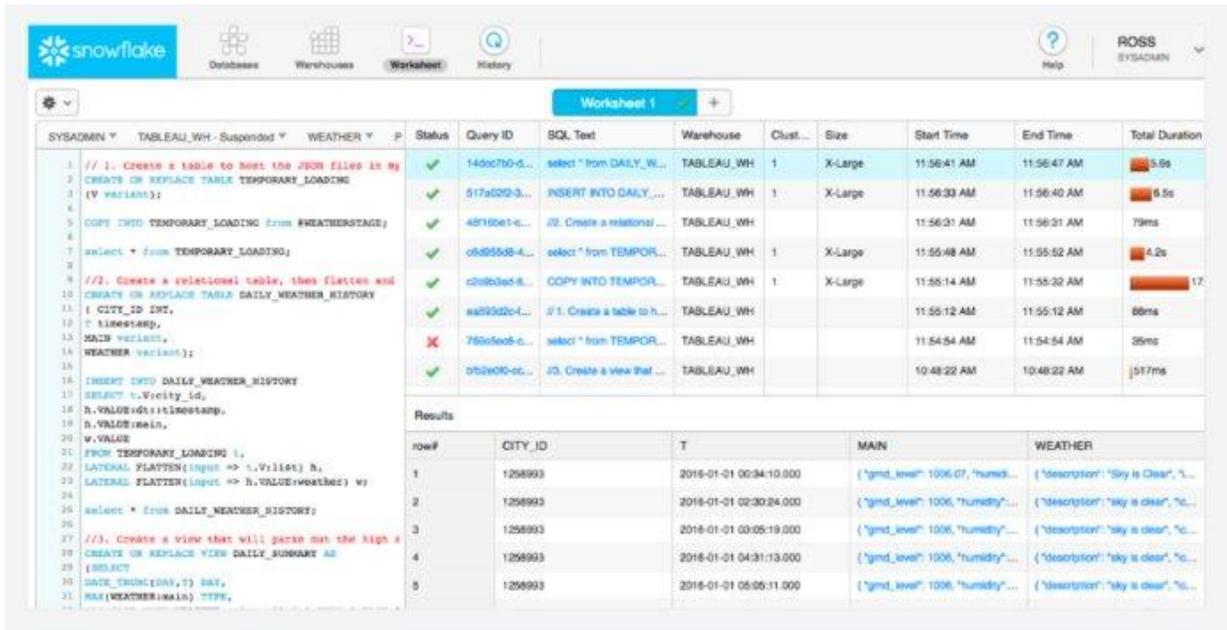
Figure 20. Snapshot by Author about SQL querying using Snowflake retrieved on 29th

October 2018 from, https://www.cio.com/article/3250703/data-management/snowflake-

now-a-unicorn-eyes-global-growth-for-cloud-data-warehouse.html

Pros and Cons of purchasing Tableau and Snowflake

<u>**Pros**</u>:

1. **Snowflake –** With growing size of data, I would recommend installing snowflake

   solution as it is an easy way to manage and access data. The ultimate benefit of

   using snowflake when compared to other database is because the whole

   architecture resides in the cloud. Multiple users have the ability anytime from all

   around the world and it has excellent scalability features.

   - Server hosted and maintained by Snowflake

   - Easy to use and maintain tables and databases

   - Fast retrieval of data and accessing information

- Easy to perform cross referential analysis

- Ability to create multiple analytical tables for analysis

- Easy to upload data onto cloud

2. **Tableau –** Tableau has an easy drag and drop functionality which makes it the best visualization platform. It is easy to learn, and it has some excellent analytical features from which we can derive decisions at faster pace. The whole product is interactive in nature and we can spot data points at ease.

- User interactive visualizations

- Ability to create different types of visualizations such as heat maps, bar graph, pie chart etc,

- Easy to share the report with others

- Quick to derive meaningful insights

- Excellent User Interface

Cons:

1. **Snowflake**

- Basic knowledge of SQL required

- Cost per month increases based on usage

2. **Tableau**

- Based on product package, the product cost increases

- Basic knowledge of using Tableau

- For tableau online, the total cost for 5 tableau users can cost around $3300 per year.

Snapshots

Figure 21. Snapshot taken by Author from Tableau desktop on 29th October 2019 which

shows the density map based on the volume of attendees from different towns attending

workshops held by Mass Audubon.

Town Officials
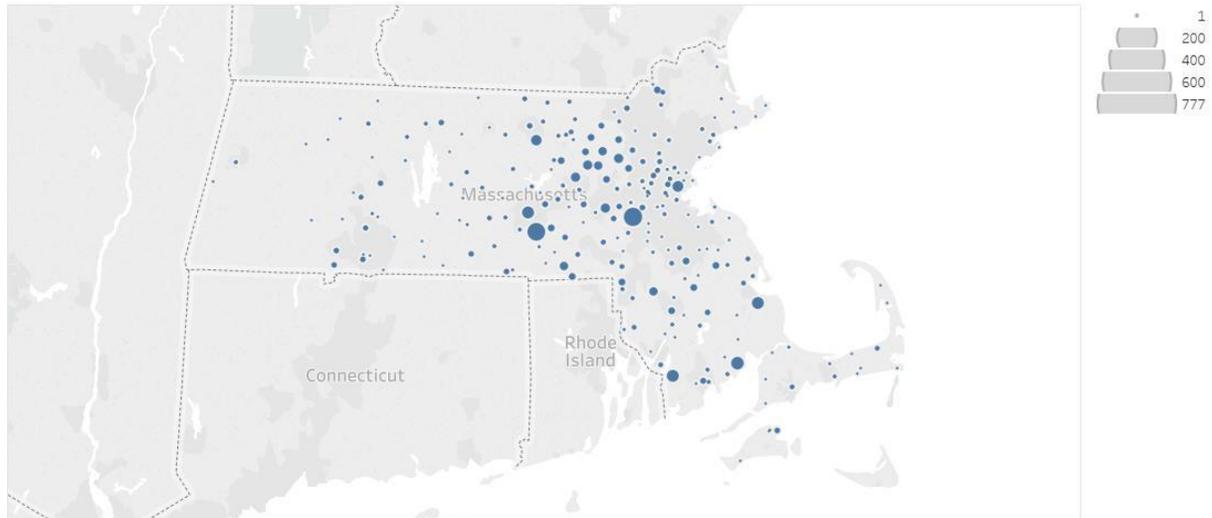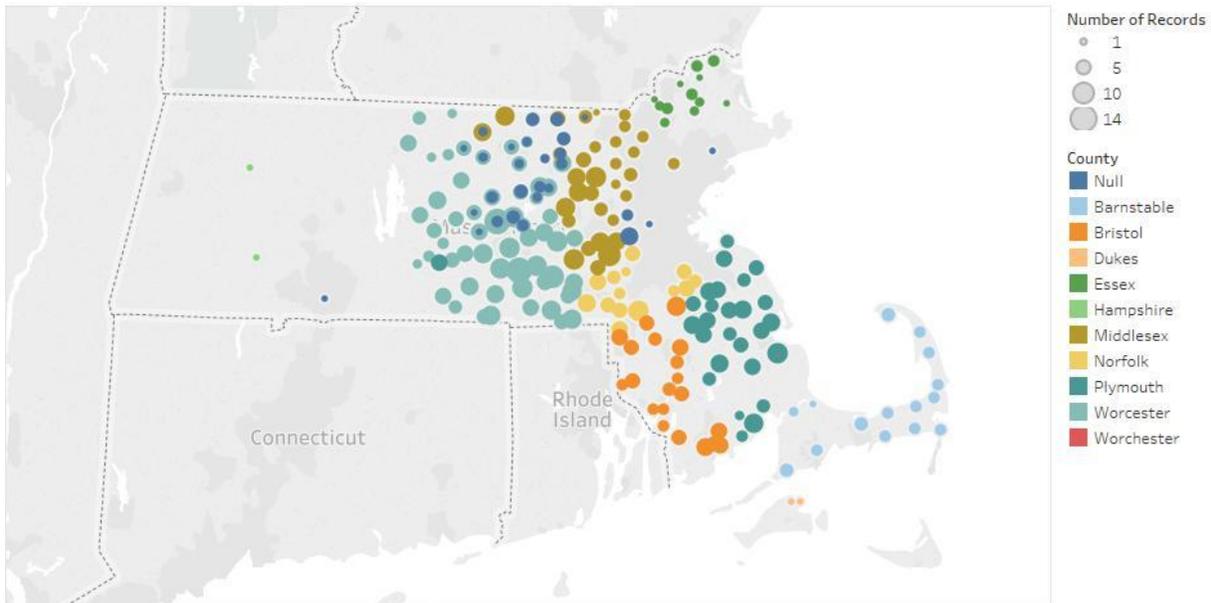
Map based on Longitude (generated) and Latitude (generated). Color shows details about County. Size shows sum of Number of Records. Details are shown for Town.

Figure 22. Snapshot taken by Author from Tableau desktop on 29th October 2019 which shows the heat map of town officials who are related to Mass Audubon.

Steps to purchase Snowflake and Tableau

Snowflake

The process of creating a cloud server account is easy for snowflake.

Step 1: goto https://www.snowflake.com/

Step 2: Click on "start for free" icon on the top right corner of the page

Step 3: It redirects to the 30-day trial account page. You shall give all the necessary details in the section and create an account with snowflake.

Tableau

Step 1: Go to https://www.tableau.com/products/desktop

Step 2: Install the 14-day free-trial version from the Tableau website.

Step 3: Once the trial is over, the software will prompt to get the licensed version.

Step 4: Click on the link and finish the installation after payment.

Appendices

Client Briefing - Mass Audubon Society

Prepared based on research and brainstorming by Phoebe Roberts, Askhat Beygenov, Ulfat Sayyed, Valinur Kutlambetov, Shrikant Patel and Shriram Sivaraman.

This document contains information that may be helpful for us when we meet with Stefanie Covino of Mass. Audubon Society.

Objectives of our Capstone team
- To analyze available data collected by Mass Audubon.

- To analyze the client's data collection process.

- To produce recommendations per the client's specifications to help them improve data management and effective data use.

Our observations from initial research
The Client

- Mission, vision, and core values center around love of nature, and conservation of ecological systems (Audubon). The website also expresses a culture which values adventure and exploration.

- Many of the client's programs interact with local communities, for example school districts, to educate people on what Audubon does, why it's valuable, and how people can help (Audubon).

- Mass Audubon was founded in the late 19th century (Eliot, 1890). It was founded before and independent of the National Audubon Society.

- Mass Audubon Society is named after John James Audubon (Eliot, 1890), an ornithologist who also drew and painted birds.

- Stefanie Covino, the client representative we will be meeting with, is the coordinator for the Future of Your Community Program.

The Data

- The client has a professional and user-friendly website, which lays out very clearly what they value and what they do. It also makes it easy for visitors to donate.

- The website pages covering the accomplishments of Mass Audubon are text-dense and do not appear to be dynamic or interactive.

- Images on the website do not appear to always directly tie into the sections of the website being visited. Some seem to be stock photos.

Industry trends

- For pursuits such as grant funding, partnering, and sponsorship, meaningful statistics highlighting impact are essential for grabbing their attention. This will give them tangible evidence that an organization asking for funding can produce meaningful results. Ensuring the right metrics are used is also vital to framing the data correctly to highlight impact (Johnson, 2017).

- Companies such as EcoWest have created data visualization specifically tailored to land conservation and environmental preservation (EcoWest). Interactive visualization tools exist which are geared toward the area of conservation and environment, like that developed by the USA National Phenology Network (LCCN, 2016).

- Gamification - the practice of enhancing otherwise boring tasks by electronically monitoring achievements and providing reward feedback - is increasingly being used for non-profit efforts such as citizen science programs in order to engage and attract more of the public, particularly younger people (Pugh, 2015).

- Mobile technology is a method recommended for engaging potential donors and participants (Haddad, 2017). Examples of this might be printing QR codes on signs located at Audubon's public properties, or creating a mobile app which would allow visitors to log their visits, get

more information, or make a donation. It can also be tied in with social media technology for an even bigger impact (Piccinini, 2017).

Questions for the Client
- Is the scope of this project specific to the Future of Your Community Program, or to a larger context within Mass Audubon?

- Which data will we be reviewing? Please outline how we will access the data set and some information on the context of the data.

- Can we have a walk-through of the data collection, storage, and management process as it currently stands?

- Relating to the initial client request, please elaborate on the fourth bullet point, which requests recommendations for engaging business and developer communities. Is this related to software development or real estate / land development? How does this tie into the data management plan - e.g., are you looking to find new ways to use your data for persuasive presentations, are you looking to engage businesses in helping to collect data, or some other purpose?

- What target audience(s) would you like us to have in mind when making recommendations related to use of data for awareness, outreach or promotion? Public, partnering organizations, donors, or someone else?

- We would like to know about the sensitivity of data involved for management and visualization. Brief required about Data Privacy and Governance policies?

- Please give a brief summary of the types of resources that would be available to implement any recommendations we make. What is available in terms of person-hours, budget, physical office space and hardware, server space, etc.?

- What software do you currently use for data collection and storage? Do you have programs developed in-house, licensed software, SAAS with remote storage, etc. and what are the restrictions on use and quantity (such as how many user licenses)?

- What organizations do you compete with for resources or to attract donors?

References:

Audubon, M. Education & Community Outreach.   Retrieved from

https://www.massaudubon.org/our-conservation-work/education-community-outreach

Audubon, M. Mission & Values.   Retrieved from https://www.massaudubon.org/about-

us/mission-values

EcoWest. What is EcoWest?   Retrieved from http://ecowest.org/about/what-is-ecowest/

Eliot, C. (1890). Letter to the Editor. *Garden and Forest, 3*(106).

Haddad, D. (2017, 12/01/2017). Nonprofit Trends for 2018 and Beyond. *NonProfit PRO*.

Johnson, S. (2017, 11/17/2017). Why is Data Important: Nonprofit Data Management. Online

Retrieved from https://www.socialsolutions.com/blog/why-is-data-important-nonprofit-data-

management/

LCCN. (2016, 6/27/2016). Phenology Visualization Tool.   Retrieved from

https://lccnetwork.org/resource/phenology-visualization-tool

Piccinini, P. (2017, 12/04/2017). Mobile Giving Trends Your Nonprofit Needs to Know. *Classy*.

Pugh, J. (2015). How Nonprofits Can Increase Engagement Through Gamification. Retrieved

from https://www.shareprogress.org/blog/nonprofit-gamification/

School of Professional Studies

Project Charter
Audubon Data Project

Table of Contents

V11

Glossary of Terms in a Project Charter

This glossary defines key terms used in this document. Although some of the terms will have slightly different definitions outside of this project, this glossary defines the meaning within this initiative.

Assumption – An item taken to be factual even though that fact has not been confirmed. Wherever possible the accuracy of assumptions is validated during the project

Constraint – An unchangeable condition that impacts the project.

Contingency – An activity, budget or time period that is held in reserve in order to minimise the impact that a risk has on the project if that risk is realised

Major Stakeholder – One of the key interested parties and decision makers in the project.

Mitigation – An activity that is undertaken to minimise the impact and /or the likelihood of occurrence of an adverse risk or to maximise the impact and /or the likelihood of occurrence of a positive risk

Project Charter – This document. The document that authorises the project and sets out the framework for what is to be done and how it is to be managed.

Project Manager – The person responsible for the management of the execution of all work items.

Required End State – The definition of what constitutes a completed project.

Risk – An uncertainty that may impact the project in either a positive or negative manner if it occurs.

Scope – The sum of the changes to be made in order to achieve the Required End State.

1 Project Overview

1.1 Introduction

This charter covers the Audubon Data Project, which is undertaken by six Clark University graduate students to deliver a data review and a set of recommendations regarding data management and data use for the Building the Future of Your Community Program.

The client – the program coordinator – stands to obtain useful recommendations and knowledge to go forward with a better data management practice for her program. The students, in turn, stand to fulfill their Capstone degree requirement by providing an effective solution for the client.

1.2 Major Stakeholders

The key stakeholders for the project are Stefanie Covino, the Program Coordinator for Mass Audubon's Building the Future of Your Community program, and Christina Wiseman, the Advocacy Associate for the same program. Additional stakeholders are the Clark University students: Phoebe Roberts, who will be acting as the project manager, and additional students who will be acting as Information Technology Subject Matter Experts (SMEs) with various specializations. These are Askhat Beygenov, Valinur Kutlambetov, Shrikant Patel, Ulfat Sayyed, and Shriram Sivaraman. Another key stakeholder is Richard Aroian, the faculty advisor at Clark University.

1.3 Document Purpose

The project charter exists to ensure understanding and agreement of the following important parameters regarding the project, as outlined in the sections below:

- *The desired end state.* This includes deliverables and intangible results of the project which should be present at the end.
- *The project scope.* This defines what will be included in the work completed by the Clark project team.
- *Assumptions, constraints and risks.* This defines certain parameters that restrict the project and factors that affect the work of the project.
- *Communications and structure.* This defines what communications should be expected, and with what frequency, by each of the project stakeholders from the others.

IX

      ○ *Roles and responsibilities.* This ensures agreement and understanding of each stakeholder's role in the project and what the role entails.

This document is not valid unless signed by all stakeholders named above, indicating agreement with all the contents of the charter.

2   Project End State and Scope

2.1   Required End State

At the end of a successful project the Clark project team will have delivered a solution to the client which includes documentation of the team's findings on reviewing the client's data, a small-scale data analysis and integrity solution built on top of the client's current data structure, and documentation for use and maintenance of that solution. As an additional option if appropriate, the report would include a brief summary of recommendations which may be of interest to the client regarding future data management practices and data use strategies on a larger scale, including potential costs and requirements. The project team will have presented this solution to the client in person with opportunities for clarification, questions, and follow-up requests and will have provided the client with a physical copy of all documentation leading to the recommendations.

2.2   Project Scope

Work Area	In Scope	Out of Scope
Data Review	Review and analysis of the data in the Excel file shared with the team by the client on 9/18/2018 and modified subsequently based on conversations with the client. Review of processes and metrics outlined in the consultant report shared with the team by the client on the same date.	Review of any other data owned by the client.
Data Analysis and Data Integrity Tools	A small-scale data analysis and integrity solution built on top of the client's current data structure which would be immediately usable by a layperson with no IT knowledge, and documentation	A solution which would incur a significant cost to the client in terms of money or other resources, including

X

	for use and maintenance of that solution. This may include forms for data search with appropriate parameters and data entry.	time spent training interns or employees on IT skills.
Data Management Recommendations	A brief summary of some basic best practice recommendations on how to more efficiently manage the data reviewed. These may include reorganization of data, import / export into a different format, or use of a different data management software program. Any recommendation would include approximate cost and considerations regarding client's current setup.	A full working solution or software tool for data management.
Data Use Recommendations	Recommendations for Information Technology solutions which may allow the client to use available data to highlight successes and make future strategic decisions for outreach. These may include demos or prototypes to show the potential ways in which such a recommended solution could be used.	A full working solution or software tool for data visualization or analysis.

2.2.1 Change Management

If a change to the scope of the project is required, the following information should be provided to all stakeholders named above for review, by the party requesting the change:

- o An overview of the requested change
- o An explanation of the reason for the request
- o An analysis of the impact of the change (time, cost and other)
- o A recommendation to the major stakeholders for the course of action that should be taken

All stakeholders must sign off on any requested change to scope before it can be implemented.

X1

3 Assumptions

The following will be assumed to be the case for the duration of this project:

- o The client is able to provide all information required by the project team in order to deliver the desired end state.
- o The project team will not request any information that is unnecessary for the project.
- o The project manager will be the first contact for all stakeholders regarding any issues that arise with the project.
- o The faculty advisor will intervene if there are any major issues with the project which the project manager isn't able to resolve.
- o All parties are acting in good faith, and do not have competing interests with the project or any incentive to recommend doing business with a specific vendor.

4 Constraints

The following boundaries constrain the project

- o Timeline: the desired end state of the project should be completed by no later than December 6, 2018, as the project team consists of graduate students whose semester ends in mid-December and many of whom plan to complete their degree at that time.
- o Resources: the project will not have a budget, and will be conducted by students who will pay their own expenses, if any apply. Students' time spent on the project is constrained by other classes, work, and personal commitments. This creates a need for flexibility and may result in changes to meeting times or meeting attendance.
- o Scope: due to the time and resource constraints of the project, the scope of the project is limited as outlined in Section 2 above.

5 Risks

Risk Identified	Mitigation Strategies
Database Access and User Account Issues	Multifaceted circular communication between the project team and the client, and documentation of the iterative process.

Visualization Tools Access and User Account Issues	Multifaceted circular communication between the project team and the client, and documentation of the iterative process.
Compromise of Data Privacy	Signed agreement provided by the project team to the client, to indicate understanding of the sensitive nature of data being provided by the client. Restrictions on how the project team will use and share the data.
Failure to Meet Project Charter Provisions	Project structure for role clarity, clear communication, and frequent documented discussion and follow-up regarding task completion.
Team Cohesion Issues	Frequent team communication through a variety of media to reinforce important information.
Client Representative Unable to Attend Crucial Meetings	Robust coordination with client, including determining meeting times as early as possible.
Scope creep	Clear definition of scope in this charter; change management process outlined in this charter.
Team Member Unavailable to Complete a Crucial Task	Each team member has a partner or a backup person to take over tasks for them if they are unavailable, and each team member is responsible for notifying the project manager or the entire team of their absence as soon as possible.
Bad Weather	The team has agreed to use Skype to videoconference if inclement weather prevents a planned in-person meeting.

6 Communication Strategy

The project communication plan will include the following:

- *Official project documents as required.* Documents such as this charter will be reviewed and approved by stakeholders and other relevant individuals and will be made available to all interested parties once approved.
- *Formal milestone reports.* Project stakeholders will be notified of the following milestones:
  - *Completion of project set-up (this is defined as initial client meeting, risk assessment, project plan, and project charter).*
  - *Completion of initial review (this includes general industry review of data management for non-profit outreach, and review of the client's data and current process).*
  - *Finalization of solution and documentation (this is defined as when the team has created a working solution for the client, including documentation for continued use and maintenance, based on several iterations of feedback with the client).*

- o *Delivery of final report (this is defined as the delivery of the formal report and presentation to the client, and marks the end of the project).*
- *Monthly status report.* A summary of information about the project for distribution to all project stakeholders. This will be provided in soft copy office document format towards the end of each month, and discussed at the next meeting with the client. The contents will include:
  - o Traffic light reporting against major milestones, and for overall project.
  - o Summary of actions planned in reporting period and actions completed in reporting period, with explanation of variances.
  - o Actions planned for upcoming reporting period.
  - o Current issues.
  - o Current risks.
  - o Other information as needed.
- *Weekly status meeting.* A meeting between the project manager, core team members and key contributors to in progress work packages/phases. Agenda to be included in the previous meeting's minutes, and minutes to be provided afterward to invitees and other interested stakeholders. Depending on the work underway at any given time, the client or advisor may be present at some of these meetings.
- *Informal communications.* Staff responsible for all in-progress and soon-to-start tasks will be in regular communication with the project manager and with one another in order to identify and address any concerns as soon as they appear. Additionally, all staff members will be encouraged to communicate any problems as soon as they appear. The project manager will focus on the creating an environment that encourages open, two-way communication focused on problem resolution.

- o *Delivery of final report (this is defined as the delivery of the formal report and presentation to the client, and marks the end of the project).*
- *Monthly status report.* A summary of information about the project for distribution to all project stakeholders. This will be provided in soft copy office document format towards the end of each month, and discussed at the next meeting with the client. The contents will include:
  - o Traffic light reporting against major milestones, and for overall project.
  - o Summary of actions planned in reporting period and actions completed in reporting period, with explanation of variances.
  - o Actions planned for upcoming reporting period.
  - o Current issues.
  - o Current risks.
  - o Other information as needed.
- *Weekly status meeting.* A meeting between the project manager, core team members and key contributors to in progress work packages/phases. Agenda to be included in the previous meeting's minutes, and minutes to be provided afterward to invitees and other interested stakeholders. Depending on the work underway at any given time, the client or advisor may be present at some of these meetings.
- *Informal communications.* Staff responsible for all in-progress and soon-to-start tasks will be in regular communication with the project manager and with one another in order to identify and address any concerns as soon as they appear. Additionally, all staff members will be encouraged to communicate any problems as soon as they appear. The project manager will focus on the creating an environment that encourages open, two-way communication focused on problem resolution.

7 Project Structure
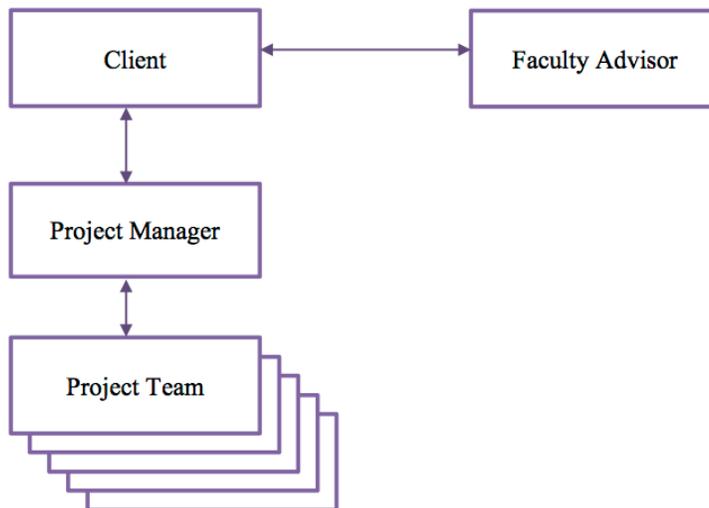
8 Stakeholder Commitments

o Agreement of all stakeholders is required before making any changes to the terms laid out in this charter.
o All decisions, signoffs, etc. requiring agreement must be made within three business days of submission. Similarly, if a stakeholder is unavailable to provide approval this must be communicated within three days and the stakeholder must name an alternate or agree to allow the other stakeholders to make decisions on their behalf.

9 Roles & Responsibilities/RASCI Chart

Responsible – The role(s) that is expected to complete the work
Accountable – The role that is expected to ensure that the work is completed (escalation point)
Sign-Off – The role(s) that is expected to approve the work
Consulted – The role(s) that is consulted on/contributes to the completion of the work
Informed – The role(s) that receives the output of the work and/or receives status reports on the progress of the work

	Roles / Responsibilities			
	Project Manager	Faculty Advisor	Project Team	Client
Project charter including end state and scope	A,R,S	C,I,S	I,S	I,S
Project management and control	A,R	C	C	C
Project communication	A,R	I	C, I	C, I
Project planning	A,R	C	C	C
Resource allocation	A,R		C	I
Problem identification and analysis	A	I	R	C
Problem resolution	A,R	C,I	C	C

10 Measures of Success

This section of the project charter should detailed measurements that will indicate that the project is a success. The following table provides examples of measures of success that teams can decide are appropriate for their projects.

Project Performance Dimensions by Project Success Factor	
Project Outcomes	**Measure of Success**

Agreements	Project is finished on time
	Project conforms to desired end state in this charter
Client Success	Client is satisfied
	Client uses the deliverables
	Client's needs were addressed by the deliverables
Project Team	Team members experience personal growth
	Team members are satisfied with the work done
	Team is loyal to project

11 Stakeholder Sign-off

This project charter has been signed off by the following stakeholders:

Stefanie Covino   Coordinator, Shaping Program   10/16/18
Print Name   Title   Date

Signature

Christina Wiseman   Advocacy Associate   10/16/18
Print Name   Title   Date

Signature

Project Charter   School of Professional Studies   Page 12 of 14

Rich Avani   Project Director   10/18/1
Print Name   Title   Date

Signature

Project Team Members:

Phoebe Roberts   Project Mgr   10-18-18
Print Name   Title   Date

Signature

Askhat Beygenov   ITSME   10-18-18
Print Name   Title   Date

Signature

ULFAT SAYYED   IT SME   10/18/2018
Print Name   Title   Date
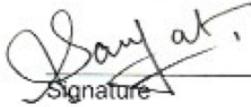
Signature

Volinur Kutlambetor   IT SME   10/18/2018
Print Name   Title   Date

Signature

SHRIKANT PATEL   IT SME   10/18/2018
Print Name   Title   Date

Signature

SHRIRAM SIVARAMAN   IT SME   10/25/2018
Print Name   Title   Date

Signature

xx
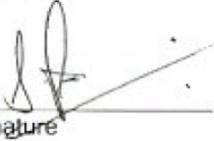
WORKSHOP

Field Name	Data Type	Data Format	Field Size	Description	Example
Workshop_ID (Primary key)	Integer	NNN	Double	Unique identifier and primary key. Must be manually input when adding workshops. Check last record's ID and add +1	101
Workshop_Title	Text	String	128	Name of the Workshop	Framingham LID workshop
Attendance	Integer	NNNN	4	Total number of people who attended the workshop.	170
Workshop_Date	Date	MM/DD/YYYY	10	Date when the workshop was held.	07/10/2017
Town_ID (Foreign key)	Text	String	16	A unique identifier that is tied to the Town table. When inputting data, a drop-down will display to allow users to select the town from the related Town table.	DIG
Score	Short Text	String	255	Allocates points to the town that held the workshop (the one referenced by Town_ID) if applicable. Should be zero if it is a regional / statewide workshop or if other exceptions apply. Should be 5 otherwise.	5
Held in Partnership	Y/N	Checkbox	1	Indicates that the workshop was held in partnership with another organization or program.	Y
Internal Workship / Training	Y/N	Checkbox	1	Indicates an internal workshop which would not generate engagement data.	Y
MAPPR Shared	Y/N	Checkbox	1	Indicates if the MAPPR program was discussed in this workshop.	Y
Climate Change	Y/N	Checkbox	1	Indicates if climate change was discussed at the workshop.	Y
Statewide or regional	Y/N	Checkbox	1	Indicates a statewide or regional workshop which might generate engagement metrics via attendance, but the city hosting it would not receive engagement points for doing so.	Y

REQUEST

Field Name	Data	Data	Field	Description	Example

	Type	Format	Size		
Person_ID (Foreign key)	Number	NNNN	Double	A unique identifier that is tied to the Person table. When inputting data, a drop-down will display to allow users to select the appropriate requestor from the related Person table. If they're not in the database yet, users will need to click the button to add a person, then reopen the Request form.	2001
Date	Date	MM/DD/ YYYY	10	Date the request was received by the Shaping program.	03/02/2017
Request	Long Text	String	2^30-1	Content of the request / what is being requested.	Info on historical groundwater levels vs present day
Score	Number	N	Double	Allocates points to the town the request came from (the one referenced in the requestor's record in the Person table under Town_ID). Should generally be 3.	3
R_MappingID (Required)	Number	NNN	Double	Unique identifier. Must be manually input when adding requests. Check last record's ID and add +1	121

PERSON

Field Name	Data Type	Data Format	Field Size	Description	Example
Personid (Primary key)	Number	NNNN	Double	Unique identifier and primary key. Must be manually input when adding people. Check last record's ID and add +1	2021
person_type	Short Text	String	Double	Indicates whether Shaping program's first contact with the individual was due to workshop attendance (P1), request for assistance (P3), or participation in tech assist project (P10). Data entry form has a drop-down menu with the three options.	P1
last_name	Short Text	String	16	Individual's last name	Smith
first_name	Short Text	String	16	Individual's first name	John
Organization	Short Text	String	50	Organization represented by the individual, if any.	NDAA
Title	Short Text	String	50	Position in the organization	Selectman
Address	Short Text	String	50	Street address	142 South St

townid	Short Text	String	Double	(Residual – not needed as Town name cross-references the Town table instead. Was previously used as a foreign key. Left in the db to avoid bugs but does not need to be updated for new records).	ABI
town (Foreign key)	Short Text	String	16	Town where the person lives. Also acts as a unique identifier to cross-reference the Town table and establish region and other information needed for data analysis.	Worcester
State	Short Text	String	16	State where the person lives. Input form includes a drop-down with all US state abbreviations. It defaults to MA.	MA
zipcode	Integer	NNNNN	5	Zip Code	02762
Email	Short Text	String	50	Email	email@gmail.com
Phone	Short Text	NNNN NNNN NN	10	Phone number	123-456-7890
score	Number	N	Double	Allocates engagement points to the town cross-referenced by the "town" column based on the person's activities. Should generally be 1 for P1, 3 for P3, and 10 for P10.	1

TA

Field Name	Data Type	Data Format	Field Size	Description	Example
Person_ID (Foreign Key)	Number	NNNN	Double	A unique identifier that is tied to the Person table. When inputting data, a drop-down will display to allow users to select the appropriate requestor from the related Person table. If they're not in the database yet, users will need to click the button to add a person, then reopen the Request form.	2001
Town (Foreign Key)	Short Text	String	16	Town benefited by the tech assist project. Also acts as a unique identifier to cross-reference the Town table and establish region and other information needed for data analysis.	Worcester
Town_ID	Short Text	String	255	(Residual – not needed as Town name cross-references the Town table instead. Was previously used as a foreign key. Left in the db to avoid bugs but does not need to be updated for new records).	WOR

Start Date	Date/Time	MM/DD/YYYY	10	Date the project was initiated.	01/01/2017
TA Description	Short Text	String	255	Nature of TA project / what is being done.	LID Bylaw Review
End Date	Date/Time	MM/DD/YYYY	10	Date the project was finished.	06/06/2018
Contact Last Name	Short Text	String	255	Optional field to designate an additional contact (not the cross-reference person) - last name.	Smith
Contact First Name	Short Text	String	255	Optional field to designate an additional contact (not the cross-reference person) - last name.	John
Email	Short Text	String	255	Optional field to designate an additional contact (not the cross-reference person) - email.	someone@example.com
Notes	Short Text	String	255	Additional information regarding the tech assist project.	Provider through RTWN

TOWN

Field Name	Data Type	Data Format	Field Size	Description	Example
Town_ID	Short Text	String	255	Unique identifier still used by one table (Workshop table), though primary key was switched to Town field for all other related tables.	WOR
Town (Primary Key)	Short Text	String	255	Name of the town – used as the primary key and unique identifier for town records. TA table, Request table, and Person table all reference this field	Worcester
Region	Short Text	String	255	Region where the town is located	Central

ATTENDEE

Field Name	Data Type	Data Format	Field Size	Description	Example
Person_ID (Foreign Key)	Number	NNNN	Double	References a record in the Person table to indicate that person's attendance at a specific workshop. Makes up a composite key along with Workshop_ID.	2001
Workshop_ID(Foreign Key)	Number	NNNN	Double	References a record in the Workshop table to indicate that someone from the Person table attended that workshop. Makes up a composite key along with	132

				Person_ID.	

Test Plan

Project Name	Audubon Data Project
Product Release Version	1.0
Quality Assurance Team	Askhat Beygenov
Valinur Kutlambetov

Test Plan
Document Version: 1.0

Test designer: Askhat Beygenov
Test executor: Valinur Kutlambetov

Document prepared by: Askhat Beygenov

DEFINITON

A TEST PLAN is a document describing software testing scope and activities. It is the basis for formally testing any software/product in a project.

Audubon Data Project Test Plan is a document describing the scope, approach, and schedule of intended test activities. It identifies amongst others test items, the features to be tested, the testing tasks, the test environment, the test design techniques and any risks requiring contingency planning. It is a record of the test planning process.

FEATURES TO BE TESTED

Audubon Data Project Test Scope includes testing for next features (Test Units):

- Add a new person from any form that references people, with all info that was previously in the spreadsheet, page through person records, update and delete person.

AUDUBON DATA PROJECT FINAL REPORT

- Add a new workshop and all related information that was in the spreadsheet, page through workshop records, update workshop.
- Add new tech assist info, relate it to a person or add a new person from there, and include all info that was previously in the spreadsheet, page through tech assist records, update tech assist record.
- Add new external assistance request info, relate it to a person or add a new person from there, and include all info that was previously in the spreadsheet, page through external assistance records, update records.

APPROACH
To perform project testing only functional and regression testing are needed.

TEST UNIT PASS/FAIL CRITERIA
If actual Test Cases results are equal to expected results then Test Unit can be considered as Passed. If not, then it is Failed.

TEST DELIVERABLES
Test Report on TestReail web-service.

ENVIRONMENTAL NEEDS
Only Windows PC is required.

SCHEDULE
Regression testing should be performed before each new release.

KNOWN ISSUES
Access search fields are case sensitive.

TEST PLAN CONTEXT

- Verify that system allows to add new person who is not connected to workshop
- Verify that system allows to delete person who is not connected to workshop
- Verify that system allows to add new person who is connected to workshop
- Verify that system doesn't allow to delete person who is connected to workshop
- Verify that system allows to update person information
- Verify that system allows to update person attendance information
- Verify that system allows to add new workshop which is not connected to person
- Verify that system allows to add new workshop which is connected to person
- Verify that system allows to update workshop information
- Verify that system allows to add new tech assist
- Verify that system allows to update tech assist
- Verify that system allows to add new external assistance request
- Verify that system allows to update external assistance request

TEST UNIT 1
PERSON

TEST CASE #1
Verify that system allows to add new person who is not connected to workshop
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Last button to remember last taken person ID	
3. Click New button	Empty form is opened
4. Enter next fields: - Personid = last taken person ID + 1 - Other fields could be blank	
5. Click Add button	Form is refreshed. All fields are blank
6. Go to person table	
7. Scroll to the last record	Person with all entered data is displayed

TEST CASE #2
Verify that system allows to delete person who is not connected to workshop
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Find button	Find and Replace popup is displayed
3. Enter search parameter: ex. testLast	
4. Click Find Next button	Person is found and displayed
5. Close search form and click Delete button	"You are about to delete 1 record(s)" popup is displayed.
6. Click Yes button	Form is refreshed. All fields are blank
7. Go to person table	
8. Scroll to the last record	Person is deleted.

TEST CASE #3
Verify that system allows to add new person who is connected to workshop
Preconditions:
Main menu form is opened

AUDUBON DATA PROJECT FINAL REPORT

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Last button to remember last taken person ID	
3. Click New button	Empty form is opened
4. Enter next fields: - Personid = last taken person ID + 1 - Other fields could be blank	
5. In Attendee subform select person and workshop	
6. Click Add button	Form is refreshed. All fields are blank
7. Go to person table	
8. Scroll to the last record	Person with all entered data is displayed
9. Go to Attendee table	
10. Find person	Person and workshop IDs are displayed

TEST CASE #4
Verify that system doesn't allow to delete person who is connected to workshop
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Find button	Find and Replace popup is displayed
3. Enter search parameter: ex. testLast	
4. Click Find Next button	Person is found and displayed
5. Close search form and click Delete button	"The record cannot be deleted or changed because table 'Attendee' includes related records." popup is displayed.
6. Click Yes button	Form is refreshed. All fields are blank
7. Go to person table	
8. Scroll to the last record	Person is deleted.

TEST CASE #5
Verify that system allows to update person information
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Find button	
3. Find a person who you want to update	
4. Click Find Next button	Person is found and displayed
5. Close search form	
6. Update any fields	
7. Click Add button	Form is refreshed. All fields are blank

8. Go to person table	
9. Scroll to the last record	Person with updated data is displayed

TEST CASE #6
Verify that system allows to update person attendance information
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2. Click Find button	
3. Find a person who is connected to workshop	
4. Click Find Next button	Person is found and displayed
5. Close search form	
6. Update workshop information in subform	
7. Click Add button	Form is refreshed. All fields are blank
6. Go to Attendee table	
7. Find person	Person and new workshop IDs are displayed

TEST UNIT 2
WORKSHOP

TEST CASE #7
Verify that system allows to add new workshop which is not connected to person
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Workshop button	Workshop form is opened. By default first record is displayed
2. Click Last and Next buttons to open empty form	Empty Workshop form is opened
3. Fill in fields	
4. Click Add Record button	Form is refreshed. All fields are blank
5. Go to Worksop table	
6. Scroll to the last record	Added workshop is displayed

TEST CASE #8
Verify that system allows to add new workshop which is connected to person
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Workshop button	Workshop form is opened. By default first record is displayed
2. Click Last and Next buttons to open empty	Empty Workshop form is opened

form	
3. Fill in fields	
4. Add person in person subform	
5. Click Add Record button	Form is refreshed. All fields are blank
6. Go to Worksop table	
7. Scroll to the last record	Added workshop is displayed
8. Go to Attendee table	Person linked to workshop is shown

TEST CASE #9
Verify that system allows to update workshop information
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Workshop button	Workshop form is opened. By default first record is displayed
2. Click Find button	
3. Enter search parameter	
4. Click Find Next button	Workshop is found and displayed
5. Close search form	
6. Update any field	
7. Click Add Record button	Form is refreshed. All fields are blank
8. Go to Workshop table and find workshop	Updated workshop is displayed

TEST UNIT 3
TECH ASSIST

TEST CASE #10
Verify that system allows to add new tech assist
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Tech Assist button	Tech_assist form is opened. By default first record is displayed
2. Click Add New Contact button	personForm is opened
3. Click Last button to remember last taken person ID	
4. Click New button	Empty form is opened
5. Enter next fields: - Personid = last taken person ID + 1 - Person_type = P10 - Other fields could be blank	
6. Click Add button	Form is refreshed. All fields are blank
7. Go to person table	
8. Scroll to the last record	Person with type P10 is displayed
9. Go back to tect_assist form, click Last	Empty form is opened

Record and Next record buttons to open empty form	
10. In the Contact field enter first name of TA	Person is found
11. Fill in other fields	
12. Click Add Record button	
13. Go to TA table	New TA person is displayed

TEST CASE #11
Verify that system allows to update tech assist
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage Tech Assist button	Tech_assist form is opened. By default first record is displayed
2. Click Find button	
3. Enter search parameter	
4. Click Find Next button	Person is found and displayed
5. Close search form	
6. Update any field	
7. Click Add Record button	Form is refreshed. All fields are blank
8. Go to TA table and find person	Updated information is displayed

TEST UNIT 4
EXTERNAL REQUEST

TEST CASE #12
Verify that system allows to add new external assistance request
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage External Assistance Request button	Extenral_info form is opened. By default first record is displayed
2. Click Add New Person button	personForm is opened
3. Click Last button to remember last taken person ID	
4. Click New button	Empty form is opened
5. Enter next fields: - Personid = last taken person ID + 1 - Person_type = P3 - Other fields could be blank	
6. Click Add button	Form is refreshed. All fields are blank
7. Go to person table	
8. Scroll to the last record	Person with type P10 is displayed
9. Go back to external_info form, click Last Record and Next record buttons to open empty form	Empty form is opened
10. In the Requestor field enter first name of TA	Person is found
11. Fill in other fields	

12. Click Add Record button	
13. Go to Request table	New external assistance request is displayed

TEST CASE #13
Verify that system allows to update external assistance request
Preconditions:
Main menu form is opened

Steps	Expected result
1. Click Manage External Assistance Request button	Extenral_info form is opened. By default first record is displayed
2. Click Find button	
3. Enter search parameter	
4. Click Find Next button	Person is found and displayed
5. Close search form	
6. Update any field	
7. Click Add Record button	Form is refreshed. All fields are blank
8. Go to Request table and find person	Updated information is displayed

TEST REPORT
List of created test cases:

Test case details:

AUDUBON DATA PROJECT FINAL REPORT

| Overview | Todo | Milestones | Test Runs & Results | Test Cases | Reports | | Administration |

C1044 Verify that system allows to add new person who is not connected to workshop

◄ ► 🖨 ✏ Edit

Other Webshops > Audubon Data Project > Test Unit 1

Type	Priority	Estimate	References
Other	Medium	None	None

Automation Type
None

Preconditions

Main menu form is opened

Steps

	Step	Expected Result
1	Click Manage Person/Attendance button	personForm is opened. By default first record is displayed
2	Click Last button to remember last taken person ID	
3	Click New button	Empty form is opened
4	Enter next fields: - Personid = last taken person ID + 1 - Other fields could be blank	
5	Click Add button	Form is refreshed. All fields are blank
6	Go to person table	
7	Scroll to the last record	Person with all entered data is displayed

In section Test Unit 1.

Details

Tests & Results

Defects

History

No attachments.

People & Dates

| Created | Askhat
12/1/2018 6:29 AM |
| Updated | Askhat
12/1/2018 7:01 AM |

Test case execution:

Add Result ✖

Status *
Passed ▼
Set the test status (*passed*, *failed* etc.).

Assign To
Me ▼
Assign to another team member.

Comment

Describe your test result.

Version

The version you tested against.

Elapsed Start

How long the test took.

Steps P B U R F

Defects Push Add

A list of IDs in your bug tracker.

1	Click Manage Person/Attendance button	Passed ▼
	Expected Result	
	personForm is opened. By default first record is displayed	
	▸ Enter actual result	

| 2 | Click Last button to remember last taken person ID | Passed ▼ |
| | ▸ Enter actual result | |

3	Click New button	Passed ▼
	Expected Result	
	Empty form is opened	
	▸ Enter actual result	

✔ Add Result   ✖ Cancel

Test Run execution:

AUDUBON DATA PROJECT FINAL REPORT

Final Score of test results.

AUDUBON DATA PROJECT FINAL REPORT

HELP FILE

Main menu

Manage Your Data: The main menu has a section on the left which links you to forms for searching or updating the tables in your database.

Analyze Your Data: This section contains two buttons.
1. Generate Data. This button will generate two Excel files for use in uploading to the Tableau Public website for visualization. It automatically drops the files in your Documents folder.
2. View Reports. This button will open the Reports menu.

Manage Person / Attendance form
Adding new person

From the Person form you can do the following:
- **View Records**: You can page through existing records or jump to first / last records in the table using the buttons at the bottom.
- **Update Records**: If you make a change to the record you are viewing, this will automatically update the database.
- **Search Records**: A Find dialogue box will display if you click "Find"
- **Add Records**: Click "Last" and note the Personid at the top. Click "Add" to get a blank form for entering a new record, and add +1 to the Personid from the last record to create the Personid for the new record.
- **Update Workshop Attendance**: Use the drop-downs in the "Attendee" sub-form to add new workshops that the person attended. This will update the Attendee table.

AUDUBON DATA PROJECT FINAL REPORT

To find person click Find button and enter search parameter.

Note that the Find dialogue gives you several options for how to search, including whether to check the current field or all fields.

Delete Error:

If you click Delete for a person connected to a workshop, then a pop-up is shown that you can't delete record because table 'Attendee' includes related records. In this case, you may need to manually access the Attendee table and delete all records associated with the person first. Use caution when accessing the tables directly outside of the interface.

Tech Assist:

From the Tech Assist form you can do the following:
- **View Records**: You can page through existing records or jump to first / last records in the table using the buttons at the bottom.
- **Update Records**: If you make a change to the record you are viewing, this will automatically update the database.
- **Search Records**: A Find dialogue box will display if you click "Find"
- **Add Records**: Click "Last" and note the TE_mappingid. Click "Add Record" to get a blank form for entering a new record, and add +1 to the ID from the last record to create the ID for the new record.
- **Add New Contact**: If the contact person for your tech assist isn't in the drop-down, you will need to click this button to jump to the Person table and add them. You may need to refresh your database or close / reopen the form before they appear in the drop-down.

AUDUBON DATA PROJECT FINAL REPORT

To add external assistance request first you need to add person with type P3. After that you can go to external_request form and add this person.

External Info / Assistance Request:

From the External Assistance form you can do the following:
- **View Records**: You can page through existing records or jump to first / last records in the table using the buttons at the bottom.
- **Update Records**: If you make a change to the record you are viewing, this will automatically update the database.
- **Search Records**: A Find dialogue box will display if you click "Find Record"
- **Add Records**: Click "Last" and note the R_MappingID. Click "Add Record" to get a blank form for entering a new record, and add +1 to the ID from the last record to create the ID for the new record.
- **Add New Person**: If the requestor's name isn't in the drop-down, you will need to click this button to jump to the Person table and add them. You may need to refresh your database or close / reopen the form before they appear in the drop-down.

Workshop form.

From the Workshop form you can do the following:
- **View Records**: You can page through existing records or jump to first / last records in the table using the buttons at the bottom.
- **Update Records**: If you make a change to the record you are viewing, this will automatically update the database.
- **Search Records**: A Find dialogue box will display if you click "Find Record"
- **Add Records**: Click "Last" and note the Workshop_ID. Click "Add Record" to get a blank form for entering a new record, and add +1 to the ID from the last record to create the ID for the new record.
- **Add New Attendees**: Use the drop-downs in the Person sub-form to add new attendees to the workshop. This will update the Attendee table.
- Note that you will not be able to delete a workshop that has attendees listed. If you need to do so, you will first need to manually access the Attendee table and delete all records associated with the workshop. Use caution when accessing the tables directly outside of the interface.

Reports Menu

The Reports Menu contains two sections:
- **Workshop Reports** contains all reports related to the Workshop table. The reports will show you statistics as described. The bottom button also pulls from the Attendee and Person tables to identify those people in the database who have attended more than one workshop.
- **Other Reports** contains some basic yearly statistics reports from the other tables, and also contains an Engagement Summary report which adds up the engagement score per town, grouped by region, based on all activities which generate points.
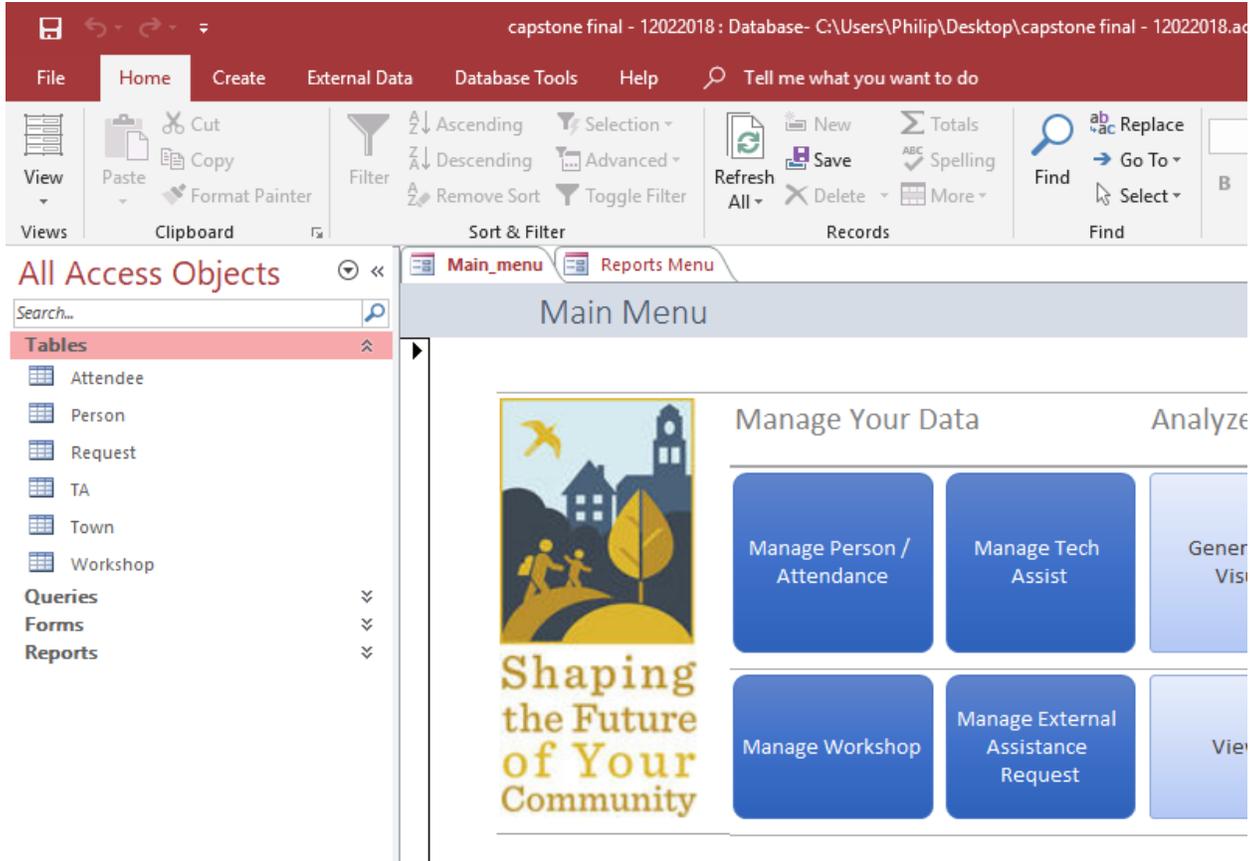
We recommend using the interface forms for day-to-day use of your database; however, it's possible to go directly into the tables if needed. Use caution when doing so. Depending on your knowledge of Access and of database management systems, you may want to work with an IT professional when accessing the tables or any other database elements outside of the form interface provided.

Monthly Project Updates

Audubon Data Project

Monthly Status Update

September 30, 2018

Questions? Feel free to contact Phoebe Roberts
Phoebe.Michels@gmail.com ● 603-484-0080

Completed Items

Kickoff Meeting Held ✔

Project Team Roles Assigned ✔

Risk Assessment ✔

In Progress

Project Charter
- Discussed project scope ✔
- Prepped draft charter ✔
- Next Steps: Finalize and sign.

Initial Review
- Researched industry trends ✔
- Reviewed client data ✔
- Next Steps: Literature review, data descriptive write-up.

Coming Up

Build Round 1

Let's get started!
The Round 1 build will involve creating an example of a working solution. This will be shared with the client at the next meeting to get feedback. Our testers will also try it out and provide internal feedback.

Where are we now?

	9/6/2018	9/13/2018	9/20/2018	9/27/2018	10/4/2018	10/11/2018	10/18/2018	10/25/2018	11/1/2018
Initial client meeting (prep / hold)									
Risk assessment (create a risk plan for project)									
WBS - timeline and assign roles									
Project charter: signed agreement with client of what project will be									
Client's data review: look at data, draw up relational structure, begin design									
Current trends literature review - nonprofit data mgt									
Sept. report									
Peer eval 1									
Build round 1: design a prototype solution and write out specifications.									
Test round 1: recc for next round									
Oct. report									
Doc & report: description / changes									
Updates per reccs in last round									
Test round 2: rec									

Project Status at a glance: ONE DELAY (see Issues section)

Issues

Current: One

Issue:
Scope not determined.

Impact:
- Delay to start of project work.

Proposed Resolution:
1. Request choice from client between small-scale / short-term versus a robust / long-term solution.
2. Adapt our project work based on client decision.

Resolved in September: None

Risk Outlook

Access Issues

Risks:
If we don't ensure proper assessment of the client's resources, we may propose a solution for which client's infrastructure doesn't meet requirements for access and login.

What we are doing:

Planning an iterative schedule

Creating a written log of analysis

Documenting discussions in team meeting minutes

Data Privacy

Risks:
If we don't ensure proper communication about data privacy and correct precautions, the client's data could be compromised.

What we are doing:

Signed and submitted data privacy agreements to the client.

Discussed rules at meeting regarding import / export of data and how to download and store safely.

Communication Issues

Risks:
Misunderstandings between team and client or lack of cohesion within the team could lead to delays or incorrect work delivered.

What we are doing:

Robust and frequent team communication.

Frequent updates to team and client.

Frequent sanity checks to confirm our assumptions.

Frequent cycles of build / test / demo to incorporate feedback in our work.

Audubon Data Project

Monthly Status Update

October 27, 2018

Questions? Feel free to contact Phoebe Roberts
Phoebe.Michels@gmail.com ● 603-484-0080

Completed Items

Project Charter Finalized / Signed ✔

Data Dictionary Drafted ✔

Database Design Completed ✔

In Progress

Implementation
1. Data Cleaning
2. Import to Access
3. Build Structure.

Client Feedback
- Received clarifications ✔
- Received list of reports ✔
- Next Steps: Incorporate feedback in implementation.

Coming Up

User Interface

Reports and Forms
Once the data structure is in Access, we will begin connecting up the data input forms and data analysis reports. Some of these forms and reports are already designed but may need to be modified to fit the structure.

Where are we now?

	9/13/2018	9/20/2018	9/27/2018	10/4/2018	10/11/2018	10/18/2018	10/25/2018	11/1/2018	11/8/2018	11/15/2018	11/22/2018	11/29/2018	12/5/2018

- ng (prep / hold)
- Risk assessment (create a risk plan for project)
- WBS - timeline and assign roles
- Project charter: signed agreement with client of what project will be (delayed but now resolved - see issues log)
- Client's data review: look at data, draw up relational structure, begin design
- Current trends literature review - nonprofit data mgt (delayed due to corrections requested by PM)
- Sept. report (delayed due to wrong email but now done)
- Peer eval 1 (delayed due to dependency but now done by most team members)
- Design phase: create db structure design, prototype forms, and identify data reorganization needed.
- Oct. report
- Implementation phase: clean data, import data, create structure, create reports.
- Updates per reccs in last round
- Client feedback cycles as needed
- Final build: finalize solution
- Write report: compile all sections from prior writing
- Nov. report
- Finalize report and make .ppt
- Peer eval 2
- Present project

Project Status at a glance: ON TRACK

Issues

Current: none

Resolved in October: Three

Issue	Impact	Resolution
Project Scope	Delay to starting work / confusion	Determined we'll build a solution rather than just recommending.
Data Integrity Issues	Problems creating data structure	Client provided clarifications and updates needed.
Client Feedback Needs	Info needed regularly from client	PM working with client to get additional info needed in real time.

Risk Outlook

Scope Issues

Risks:

If we don't clarify scope with the client, we may not be on the same page as to what we're delivering.

What we are doing:

Codified project scope in charter document

Providing demo examples of what we plan to deliver

Requesting frequent feedback to confirm our own understanding and the client's

Data Integrity

Risks:

If we don't build relationships into the data structure, we will have duplicate information which can easily become inconsistent (example: updated in Table A but not in Table B).

What we are doing:

Creating a relational structure to tie all possible related items together so when one updates, its counterparts auto-update.

Reviewing our understanding of the data regularly with the client.

Timing Issues

Risks:

If we don't set a time by which we need information, we may be left without enough time to act on the information. We are now working on timeframes less than a week.

What we are doing:

Team is coordinating internally regarding what days they are free to get their tasks done.

When info is needed from the client, it's filtered through the PM, who requests the info along with a target date when it is urgent.

PM continues to update the WBS Task List as project tasks change.

Audubon Data Project

Monthly Status Update

November 24, 2018

Questions? Feel free to contact Phoebe Roberts
Phoebe.Michels@gmail.com ☎ 603-484-0080

Completed Items

Data Cleaned / Normalized ✓

Data Structure Created ✓

Data Imported to Access ✓

Data Mgt Forms Created ✓

In Progress

Analysis Implementation
- Creating Access Reports
- Building an Import Query
- Creating Tableau Workbooks

Documentation
- Creating Help Document
- Creating Tableau Import Instructions
- Designing Presentation
- Finalizing Formal Project Report

Coming Up

Delivery and Presentation

Delivery
We plan to deliver an Access file containing your data and a built-in interface, a customized Tableau Public workbook, and a set of documentation and recommendations to help with immediate everyday use and future improvements.

Presentation
We'll present and demo the solutions we are delivering to you on December 5, 2018.

Where are we now?

| 10/25/2018 | 11/1/2018 | 11/8/2018 | 11/15/2018 | 1/22/2018 | 11/29/2018 | 12/5/2018 |

tion needed.
Oct. report
Implementation phase: clean data, import data, create structure, create reports.
Updates per reccs in last round
Client feedback cycles as needed
Final build: finalize solution
Write report: compile all sections from prior writing
Nov. report
Finalize report and make .ppt
Peer eval 2
Present project

Project Status at a glance: **ON TRACK**

Issues

Current: none

Resolved in November: Three

Issue	Impact	Resolution
Database Structure Decision	Delay to implementing database / confusion	Determined we'll use relational structure to optimize data integrity.
MS Access learning curve	Delay to implementing interface / confusion	Research and self-directed learning on how to create Access forms.
Visualization Decision	Delay to implementing visualization	Determined we'll use Tableau Public with a pre-built Access query.

Risk Outlook

Structure Issues

Risks:

If we don't build a data structure that reflects the correct relationships, we may make the client's job of compiling and using their data more difficult.

What we are doing:

Implementing a pre-designed structure

Testing as we go to ensure structure integrity and client usability

Using Access data integrity checks to reinforce structure.

Learning Curve

Risks:

If we don't understand the capabilities of our tools, we may not use them to the best advantage of the client.

What we are doing:

Educating ourselves on capabilities of MS Access and linking it to our existing data management expertise.

Implementing solutions that work with the built-in capabilities of Access.

Timing Issues

Risks:

If we don't develop a visualization import query timely, we would not be able to deliver this portion of the project.

What we are doing:

Team is working together on developing the query to meet clear specifications.

To avoid overcomplicating, we are focusing the query on only the most essential data metrics for the purpose of visualization.

For other aspects of the data, we are designing Access reports to ensure that information is also available even if not visualized.

Risk Management Plan

Audubon Data Project

Introduction

Purpose of This Document

Our aim is to create a comprehensive risk plan which provides for responses for the highest-impact and/or highest-probability risks which are likely to affect our project. To this end, we have held a risk brainstorming session involving the entire project team.

The material from our brainstorming session is synthesized in this document. Further, a specific response to each of our priority risks has been outlined.

Our Top-Priority Risks

Below is a list of our top-priority risks in the format of "If [**Risk**] Then [**Impact**]" in priority order, including Probability/Impact (or **P/I**) ratings of H/M/L – for example, a high-probability and low-impact risk would have a rating of **HL**.

If (risk)	Then (impact)	P/I Rating	Priority
Internal Risks (built-in checks on the process of creating project deliverables)			
Database access and user role creation information is unclear	The project deliverables would be unusable by the client.	HH	1
Visualization tools access and user role creation	The project deliverables would be unusable by the client.	HH	2

information is unclear			
The team does not abide by nondisclosure and data privacy agreements	The client's data would be exposed to vulnerability.	MH	3
The team does not deliver what is promised in the project charter	The project would not be considered successful, and the client could have a bad impression of Clark as well as the individual project team members.	MH	4
The team has a low level of cohesion	There could be duplicated work, delays, or substandard end product.	MH	5
A client representative is unable to attend required meetings	Client approval and feedback would not be available to ensure we are going the right direction with our project work.	LH	6
Scope creep occurs	We would create something the client did not ask for in addition to project deliverables.	ML	7
External Risks (mitigation of unforeseen circumstances)			
A team member has a personal issue come up (family emergency etc)	The team member would be unable to complete a key project deliverable which could have a domino effect on timeframes.	MH	8
Bad weather keeps the team from being able to attend meetings	Time would be lost or communication would be incomplete regarding required actions during the timeframe of the	MM	9

weather issue.		

Internal Risks – The Prime Focus of our Risk Management Plan

Because we are completing a project which relies on strong communication between a variety of stakeholders with different backgrounds and skill sets, who generally have not interacted previously, we are focusing almost entirely on **internal risks**.

Our experience from previous projects and in this project so far has told us that while unforeseen events such as weather issues may come up and should be provided for, these issues do not require targeting as a high-priority risk. As a result we feel that internal elements such as cohesion, communication, clarification, and feedback structure should be our main focus.

Access and User Roles:

For a solution designed to be used by a client without IT familiarity, our own clear understanding of the client's access, sharing, and updating capability is essential to recommending or implementing a successful solution which can be used by the client. Clarity should be top priority here, which is why these are the first two items on the list of internal risks.

Reputation:

We need to remember in our planning that not only is there an immediate affect on the project inherent in these type of internal risks, but an affect on our reputation as well – both individually and as representatives of Clark University to an outside organization. We should also keep in mind that in some cases, a negative risk such as a miscommunication within the team might even be turned to a positive risk if we are able to be frank regarding the misunderstanding, quickly and effectively clarify it, and move on. The client may well remember that we were prepared for some of the stickier aspects of completing an IT project, and even spread their positive experience of working with Clark by word of mouth.

Relationships:

Our personnel relations - not only our management of our relationship with client representatives, but also our own internal team building process, is another item on which will be critical to our success. If a relationship breaks down between these important stakeholders, it will present a significant barrier to producing the deliverables requested by the client.

Checks, not Fixes:

Ideally, our internal risks will have been provided for as part of the initial planning process. The portion of our risk plan dealing with these risks is an additional check to ensure they have been addressed properly and give opportunities to reopen the conversation regarding any internal variance and our relations with sponsors and personnel.

External Risks – Basic Precautions

Though we are pinpointing internal risks as our highest priority, we recognize that unforeseen events can happen that are out of our control. We cannot ignore the negative impact our project would suffer if these elements of our plan are not executed correctly.

Action Items

Purpose – Mapping to the Project Plan

The table shown on the following pages provides both a summary of our risk response strategies and a guideline for mapping those risk response strategies onto our project plan and making the appropriate team members accountable for managing the risks which fall within their area.

With a few exceptions, risk responses generally do not happen when the risk becomes an issue, but are put in place well before that point. A perfect example is responding to the risk of potential liability or damages by purchasing insurance. The insurance must be purchased before any actual accident or liability occurs, and cannot be purchased in reaction to the risk after it becomes an issue – thus most of our responses, similar to buying insurance, need to be *proactive* rather than *reactive*, and thus will be built into our plan.

To that end, this portion of our risk plan serves as a guide to how our project plan should be

modified, or where our lower-level planning documents (such as task lists) need to take into account

specific steps for risk mitigation, avoidance, and transference as well as the optimization of positive

risks.

On reviewing the information shown in our action items, it is clear there are several specific tools we will be using to manage most of our top-priority risks.

Risk Toolbox

- Communication

    ◦ Frequent check-ins with all stakeholders (meetings, email check-ins, reports).

    ◦ Documentation of meetings and other relevant discussions (minutes).

    ◦ Informal / incidental communication method that works for everyone (WhatsApp).

    ◦ Shared file repository (Google Drive).

    ◦ Mindfulness of potential need for translation of profession related terms (e.g. IT-related vs. nonprofit-related vs. project management terms).

    ◦ Status tracking via dynamic WBS document.

- Monitoring and Accountability

    ◦ Role clarity to avoid confusion on task assignment.

    ◦ WBS deadlines and assignments clear and regularly referenced.

    ◦ Multiple iterations with client feedback to ensure we're meeting their needs.

- Technology

    ◦ Google Drive folders for storing project work in a shared location.

    ◦ WhatsApp for quick informal notifications and questions.

    ◦ Email for more formal communications.

Next Steps

Regular Updates

We will consult this risk plan when we meet as a team to verify that all risk management actions are being taken as outlined, and to make changes to the document itself as required, such as if are made aware of a new and high-priority risk, or a risk is closed out because its timeframe has

AUDUBON DATA PROJECT FINAL REPORT
passed.

The Issues Log

While not a part of the risk plan, the issues log has an important relationship to the risk plan

and the two should be considered together to get a proper picture of the risk planning landscape. As

certain risks become issues, they will be recorded in the issues log along with our responses.

Information collected in the issues log will be added to the project team members' body of knowledge

in order to improve our risk planning process in our future careers.

Issues Log

| Date Discovered / Identified	Description of Issue	Possible Contributing Factors	Resolution	Date Resolved	Lesson(s) Learned
9/27/2018	Scope has not been decided yet and therefore project charter can't be finalized timely by 9/27 per goal set in WBS.	Client needs usability and low costs which may result in a tradeoff in terms of robustness and ability to implement a true data management solution. We are still at the point of determining whether a solution can be implemented or if we should make a more long-term recommendation with a road map.	Reviewed the issue with the client by providing two alternate project charters, going over the differences in a meeting, and alloing the client to choose their path forward.	10/16/2018	Ensure clarity in the area of client resources and what they are picturing, specifically, as a solution. Ensure clarity of terms ("data management" may mean different levels and scales of structure to different people). Be ready to change our direction if what we are considering is not in line with what the client needs - don't get attached to a solution before communicating it clearly and checking for approval.
10/18/2018	Pattern identified where project team members each have different understanding of what's needed from the client, and what the client has agreed to deliver.	Project manager hasn't included expected actions by the client in the WBS in order identify what will be delivered when, what may be missing and need to be added to the task list, and hold the client to deadlines needed to keep the project on track.	Added client to the WBS, negotiated deadlines with the client for when they would provide us with required information.	10/25/2018	Consider up front what expectations should be set for client participation in the project. Include these in writing - ideally in the charter but also share WBS with client and keep them up to date along with the team, so that actions needed from them to move the project forward, and deadlines for those actions, are clearly

					understood by everyone.
10/21/2018	Several team members don't have the ability to easily access a computer that can use MS Access, as they have Macintosh computers which are not compatible with the software.	Assumption wasn't identified regarding what software and hardware we each have access to.	PM checked in with team to determine who has access to what, and assigned tasks accordingly going forward. Those without a PC at home were able to use Clark computer labs but might have time / location restrictions so were given more written work and less database work. Shrikant also set up a Windows Virtual Machine on his Mac so that he was eventually able to use Access.	10/25/2018	Issue a questionnaire to new project team members to determine not only their access to various types of software but also potentially other items such as planned vacation days, skill set and strengths / areas where they'll need help.
11/1/2018	Team members disagree on how best to implement the database structure.	Issues between data integrity versus easy implementation. Also issues between ease of use and robustness.	We discussed the pros and cons and determined that providing a database solution makes more sense if the tables reflect real-world entities, and that the client's data as shown in their spreadsheet is inconsistent since, for example, one person might be represented separately in many spots.	11/8/2018	The reason we went so long without realizing this was an issue was that we had different assumptions regarding what the database would look like. We might have benefitted from spending more time discussing the relational database structure design that Shrikant originally created and ensuring commitment and understanding from everyone regarding how it would work.
11/8/2018	There is a learning curve issue with the functionality of MS Access, where the team is mostly not familiar with the platform though they are experts in database management.	MS Access is designed so that a layperson can create and manage a database, so it has point-and-click interfaces for many things which an IT specialist might prefer to access more directly using code or SQL queries. It's somewhat clunky from a programming perspective and the MSIT program likely doesn't cover it in depth, as using Access is more like being an end user for a complex software program than doing actual database	PM knew more about Access than most others in the group, so she worked with them to determine what was possible to do using Access, and demonstrated what types of functionality were available in order to connect her own software use expertise with the technical expertise of the team.	11/15/2018	All IT fields require some amount of adaptability. It's recommended that anyone in an IT-related field recognize the parameters of their work will be set by their employer or client's budget limitations, and that they may not be able to use a platform they are familiar with to do the things that need to be done. Adopting a posture of constant purpose-based research and learning will ensure they can add value for

		management tasks as a programmer.			any prospective client regardless of the client's specific needs and restrictions.

PowerPoint Slides

AUDUBON DATA PROJECT

5 December 2018

Agenda

Overview	Data Management Platform
Challenges	Unit Testing
Goals and Purpose	Data Visualization
Project Flowchart	Recommendations
Defining a Scoring Pattern	Implications
Data Design and Restructure	Acknowledgements

2

Overview

Mass Audubon Society's Shaping the Future of Your Community Program, represented by Stefanie Covino, Ariel Maiorano and Christina Wiseman.

- Conducts workshops across different towns
- An initiative to create awareness among residents
- Track attendee level details, workshop summary, technical assistance and external information requests
- Manage and maintain data using excel sheets

3

Overview - About the Client

Client's Goals:

- Promote sustainable planning and development in Massachusetts.
- Engage communities via outreach and education:
  - Workshops
  - Technical Assistance
  - Special requests from the public

4

Overview - About the Data

The Client's data consists of:

- Workshops grouped by fiscal year.
- Workshop attendees including contact information and workshops attended.
- Tech assist projects including contact information / town.
- External requests including requestor.
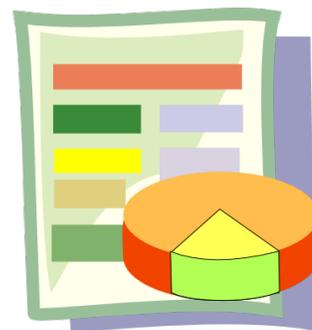
5

Overview - Preliminary Research

Data Management Research

- **Standardized data entry**
- **Searchable solutions**
- Cloud based solutions
- Efficient use of computing resources

Data Visualization Research

- **User friendly solutions**
- Cloud based solutions
- Saving / tracking
- Potential for scaling up later.

6

Challenges

- Understanding the data flow
- Unstructured data and redundancy issues
- Fixating on a scoring pattern

- Workshop summary attendance stats
- External info or asst requests
- Workshop attendees
- Longer term outcomes re-follow

7

Goals and purpose of the project

Goals

- To streamline the scoring pattern
- To structure the unstructured data
- To create a data management platform
- To deliver a data visualization product

Purpose

- To make right decisions at faster pace
- To evaluate current standpoint
- To focus on future planning

8

Project flowchart

9
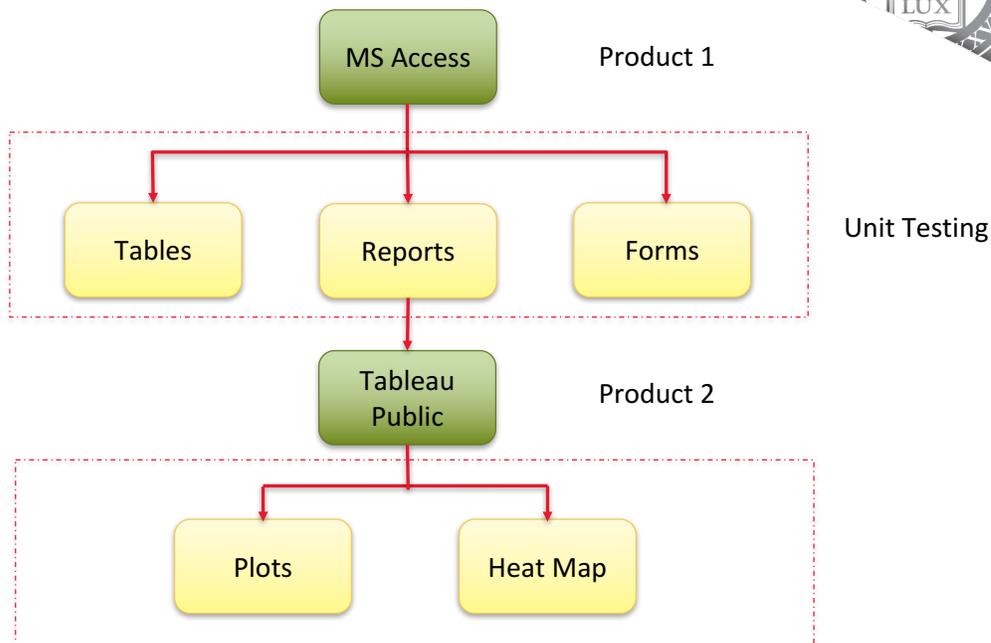
Defining a scoring pattern

Dimension	Scores
Attendees	1
External info requests	3
Workshop	5
Technical assistance	10

✦ Towns with highest and lowest engagement levels?

✦ Towns with highest and lowest attendance volume?

✦ Region-wise comparison for scores and attendance?

10

Data design and restructure

Unstructured data

Structured data

11
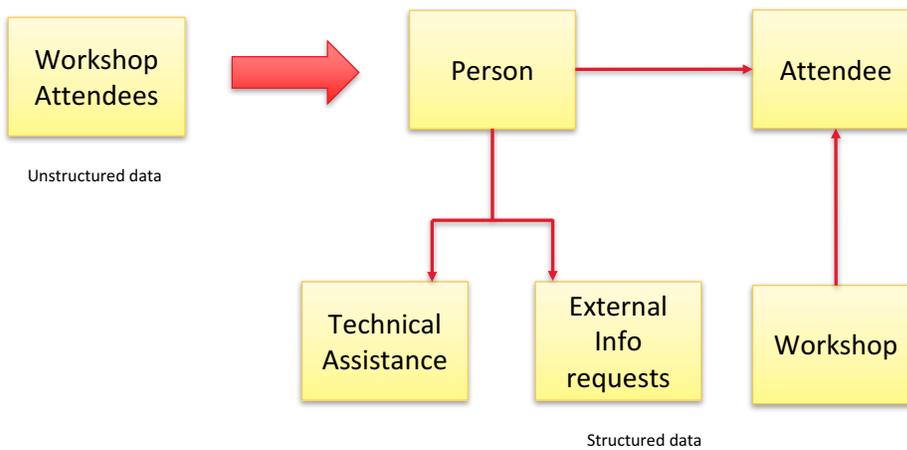
Data Management Platform
MS Access - Tables

Data Management Platform
MS Access - Forms

Data Management Platform
MS Access – Data Input

14

Data Management Platform
MS Access Reports

15

Unit Testing

# Test Results

Data Visualization
Tableau Public

Summary Bubble Chart

Map based on Longitude (generated) and Latitude (generated). Color shows details about Region. Size shows sum of Score. Details are shown for Town and Region.

18

Data Visualization
Tableau Public

Summary Score per Town

Sum of Score for each Town. Color shows details about Region. The view is filtered on sum of Score, which includes values greater than or equal to 37.

19

Data Visualization
Tableau Public

Summary Score per Region
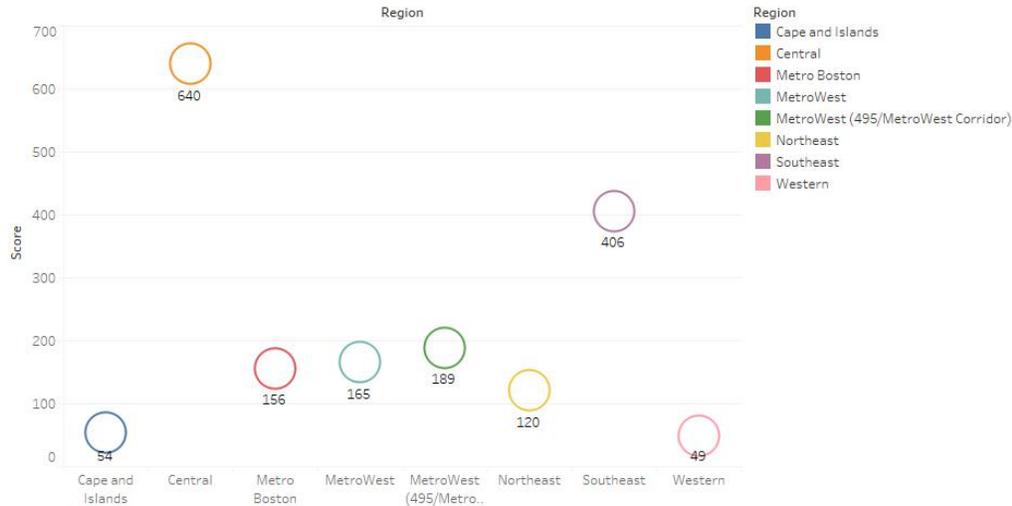
Sum of Score for each Region. Color shows details about Region.

20

Recommendations

Cloud database server – Snowflake, BigQuery and Redshift

Pros
- Effective data management and data security
- Perform cross referential analysis
- Faster data access and scalability
- Data not local

Cons
- SQL knowledge
- High cost

21

Recommendations

Data Visualization Platform – Tableau desktop professional, Data Studio, R Shiny

Pros

- User interactive visualizations
- Easy report sharing and maintenance
- Excellent user interface
- Quick decision making
- Derive meaningful insights

Cons

- High cost
- Basic product knowledge

22

Implications for our professional development

The project team's experience aligned with real-world scenarios

- Team work
- Time Management
- Product delivery
- Expanding skill sets
- Exposure to professional atmosphere

23

Acknowledgements

Mass Audubon Coordinators:

- Ariel Maiorano
- Christina Wiseman
- Stefanie Covino

Faculty Advisor:

- Richard Aroian

Clark University Capstone Team:

- Phoebe Roberts
- Ulfat Sayyed
- Askhat Beygenov
- Valinur Kutlambetov
- Shrikant Patel
- Shriram Sivaraman

24

THANK YOU FOR YOUR ATTENTION

ANY QUESTIONS?

CLARK UNIVERSITY

CHALLENGE CONVENTION. CHANGE OUR WORLD.