2020

Inflection vs. Continuation: A Discussion of Data Centralization in Startups

Corey Parker
University of Pennsylvania, parkerco@wharton.upenn.edu

Inflection vs. Continuation: A Discussion of Data Centralization in Startups

Abstract

This paper delves into the pursuit to introduce data centralization methods to small-scale firms and startups. This is done by teaching the founders skills and methods in Excel while tailoring the outputs to their specific industry and customer segments. Throughout this consulting-like process, I recorded each firm's ability to understand my methods and the relative likelihood for adoption post-consulting. This study has found that the likelihood for adoption lies not in the complexity of the model, but the stage of the startup. By understanding this key difference, this paper aims to provide a blueprint for firms to follow when implementing data-centric practices, but giving specific recommendations based on two key life cycle points: inflection and continuation. The significance of these two terms will be discussed in the body of this thesis.

Keywords

startup, startups, data, centralization, data science, entrepreneurship

Disciplines

Business Analytics | Entrepreneurial and Small Business Operations

INFLECTION VS. CONTINUATION: A DISCUSSION OF DATA CENTRALIZATION IN STARTUPS

By

Corey A. Parker

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

JOSEPH WHARTON SCHOLARS

Faculty Advisor:

Prasanna Tambe

Associate Professor, Operations, Information, and Decisions

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2020

Table of Contents

Abstract

This paper delves into the pursuit to introduce data centralization methods to small-scale firms and startups. This is done by teaching the founders skills and methods in Excel while tailoring the outputs to their specific industry and customer segments. Throughout this consulting-like process, I recorded each firm's ability to understand my methods and the relative likelihood for adoption post-consulting. This study has found that the likelihood for adoption lies not in the complexity of the model, but the stage of the startup. By understanding this key difference, this paper aims to provide a blueprint for firms to follow when implementing data-centric practices, but giving specific recommendations based on two key life cycle points: inflection and continuation. The significance of these two terms will be discussed in the body of this thesis.

Acknowledgements

I would like to thank Wharton faculty members Tong Liu, Dr. Catherine M. Schrand, Dr. Utsav A. Schurmans, and Dr. Prasanna Tambe for guiding me through this enriching yet challenging process. I would also like to personally thank Jovana Arriaga, Alexis Sabet Echavarria, and Victor Rodriguez for giving me full access to their startups and data. Without the contributions of these seven individuals, this thesis would not be possible.

Introduction

Data collection is an integral part of a startup's success and can be the catalyst for an entrepreneur to best understand their customer base. Researchers in the field have looked more into the varying levels of data integration and the historical shift to having information digitalized. The key of this subsequent paper is to discuss a data-centric mindset within a small business, but with the emphasis on in-house analysis. In the world of big data, companies have been created to deal with data centralization. Companies such as Salesforce and Pipedrive offer platform-based approaches for a small to large size company to analyze their customers better. The central argument of this paper is that for a cheaper, more efficient cost, an entrepreneur could better make decisions and predict consumer behavior. In turn, the adoption of these types of processes is based not on the complexity of the model or method, but the stage of the startup. The two stages that will be discussed throughout the paper are inflection and continuation. Inflection refers to the point where a firm has made the conscious decision to change their practices (i.e. move to using customer analysis methods). Continuation is a point on the life cycle where a firm is looking to continue their practices either because the firm is succeeding currently, or the organizational inertia of change is too great.

One of the central sources of data for which this thesis is based is an all-encompassing data frame of customer activities within a startup named Nexentia. Nexentia is a biotech startup located in Medellín, Colombia that specializes in microencapsulation of essential minerals for digestion in both animals and humans. The data frame for this thesis focuses solely on the product *Forticaps* which is a calcium tablet for horses. With this background in mind, the key metrics that are included in the dataset are individual purchase data, consolidation of costs, lifetime value projections, and traction models. The archival data here is extremely rich and dates to around

September 2017. As analyzed thoroughly in Peter Fader's paper, I will use a slew of modeling techniques with a primary emphasis on the negative beta-binomial model. (Fader et al., 2010). From there, I looked to see their willingness to adopt these methods post-consulting. For the purposes of the subsequent analyses, Nexentia is a company in the continuation stage of their life cycle.

The second main data source for this analysis comes from a retail startup based out of Miami, Florida called Intrekit. Their primary product is leggings which is geared towards a specific customer segment: women who are passionate about health but are not avid fitness-fanatics. With such a clear mission in mind, they have found trouble keeping a central data process in place. With the study being encompassing of consulting practices for the firm, I was able to extract and document a robust amount of data. This ranged from their inventory balances to brand management. Of the two companies that were studied, Intrekit provided the most comprehensive amount of data and allowed for more cross-functional analyses of their business activities. For the purposes of this paper, Intrekit's stage in their life cycle can be classified as inflection due to their collective agreement for change and expectation for a large increase in sales volume.

Most academic papers are built to be dissected by the academic community, but this paper is written and conveyed in such a way that any business owner can read and implement these practices in order to maximize their profits. The specifics of the recommendations in the following analyses are geared towards the inflection and continuation portions of a business. Therefore, some of the recommendations/analyses may not be well suited for certain types of businesses. The analysis only really discusses the findings with an emphasis on biotech and consumer retail industries, which could be a potential limiter of the study. However, this two-tiered approach to this thesis (analysis of methods implemented and willingness to adopt them) is targeted at the common entrepreneur.

Literature Review

Introduction

With data science becoming an increasingly adopted field, all companies regardless of the sector have begun to implement data processes in their firms. Scholars have offered a plethora of analyses on the relevance of big data, but a decreased amount specialize in data utilization within startups. The motivation of this analysis and subsequent literature review is to understand the landscape for data implementation in small businesses and how these seemingly "lower resourced" companies can best manage the data available to them.

Data Science Integration

With such a large array of resources available related to data science integration in general, there is a clear correlation between data science integration and positive results for the company that implements it. In the more specified data science fields, however, there still appears to be some holes that the more tech-savvy company may still not be fully able to do. Ahmed, Dannhauser, and Philip (2018) delved into the concept of a lack of direction within data mining methodologies when it comes to small businesses. Different frameworks have cropped up such as the Lean Startup and Design Thinking frameworks to functionally adapt to common issues faced by startups: time and lack of scale. Weber (2018) laments the point that time and lack of scale can be combated by the key benefits of data integration. A few that he mentions in his argument are "predicting models of customer behavior", "running experiments to test products", and "identifying key business metrics to track." These three benefits alone should sway an entrepreneur to investigate either hiring a data analyst for their team or implement more intensive data collection in-house. In the paper by Provost and Fawcett (2013), they touch more on the in-house capabilities and the requirements to run an effective data-centric business model. The most important is to fully “understand [data science] relationships to other concepts” such as customer relationship management. With the marriage of data science and custom relationship management (CRM), both can be more effectively utilized and maximize the profits of the firm.

Customer Relationship Management (CRM) and its Applications

One of the most talked about terms in any business where sales are the most integral part of the profit model is customer relationship management. CRM poses extremely high upside for firms due to the different forms that a management strategy can encompass. Ciampi, Giacomo, and Rialti (2018) discuss in their piece the type of environment that is well suited for implementing CRM software. This is described as a place with a very rigid infrastructure with clear direction and goals for the firms that would aid analysis of internal and external data. This rigidity is key because with data coming in many different forms, a firm needs to have a clear mission in mind once that data is gathered. An application of this rigidity is discussed by Cheng and Shiu (2019) where they performed a study of SMEs (small and medium-sized enterprises) and found that using CRM software can enhance the effect of social media-based customer involvement. Real world examples like this are critical to the mainstream adoption of these concepts because business owners of all fields can point to different cases of success and be more inclined to implement these practices themselves.

In another study by Peltier, Schibrowsky, and Zhao (2009), the group were able to conclude that CRM knowledge was connected to higher product class knowledge of the entrepreneurs and greater orientation in their respective target markets. Positioning is key in a startup’s life cycle and the earlier an entrepreneur can understand the specific segment they are looking to target, the more efficiently they can roll out their first MVP (minimum viable product). In other articles such as *How Integrating CRM and Marketing Automation can Help Startups Drive Sales* and *How Startups and SME's can Leverage Open Source CRM to Increase Business*, researches are putting together a strong body of work in the field of data leverage. By analyzing the customer as a complex unit, a businessperson can pivot themselves accordingly and avoid the frightening statistic that "9 out of 10 startups fail" stated by Krishna, Agrawal, and Choudhary (2016).

Ideas for Future Research

Some possible ideas for future research can be both simple and complex with obvious tradeoffs for either decision. On the easier side, research can be more closely focused on time series modeling of a startup's life cycle regarding data implementation and see if there is a positive correlation between using CRM software and increased profits. A major tradeoff to this method would be to find suitors for this study and the varying times that startups operate prior to either a shut down or a buyout. Another strong possibility is to cross analyze the type of data that is typically found in large data-rich firms such as Amazon or Facebook and cross check their processes with that an early stage startup looking to better create a product-market fit.

Conclusion

From these 10 papers and studies, we can see the existence of a huge body of research in the overall field of data science, but not as comprehensive about startup applications. Scholars have shown that there exist tangible applications of data in startup utilization and the field has shown promise in making this more accessible to almost all business owners. The one major void in these papers is a holistic approach for the common entrepreneur. Moving forward it would be extremely beneficial if more time series modeling were studied across various startup sectors in order to either see core competencies across or industry-specific data collection and analysis methods.

Theory Behind Implementation

Prior to making quantitative analyses of both startups, a large quantity of my time was spent delving into the theory behind implementation. The first firm I will discuss in the section of the paper is Nexentia, the challenges they faced conceptually, and what their data was intended to solve. Next, Intrekit will be compared to Nexentia due to a large body of overlap between the strategies that implemented across the two firms.

Nexentia

As mentioned prior in the introduction, Nexentia is a biotech startup that operates out of Medellín, Colombia. Interestingly, the company was conceived to be a startup arm of the national textile conglomerate Corona. The purpose of Nexentia's founding, according to their CEO Alexis Sabet Echavarria, was to develop cutting-edge products with the intent to learn about specific consumer markets. For Nexentia, their dive into data-centric practices was intended to be a trial run for more lucrative products that would be human facing.

The product that I was tasked with assisting on was *Forticaps*, which was a calcium supplement for horses. Corona and Nexentia were able to gain a patent on microencapsulation of this form of calcium and were using the horse market (which by numbers is relatively small but sustainable in Colombia) to learn about a transition to selling similar products for humans. For Nexentia, data was a tool to better understand the clientele and make more tangible analyses. Corona due to its large scale and long-standing reputation, had a significant amount of organizational inertia to change practices. By using *Forticaps* as a case-study, they would be able to effectively judge whether a shift to data-centric practices were both feasible and beneficial.

With this large background in mind, I was hired as a Data Analytics and Marketing intern for the startup. My job description was incredibly broad which allowed me the jurisdiction to expand my knowledge into all aspects of their business. One main theory that Echavarria always preached was the concept of an employee being knowledgeable about all business happenings. By understanding the connections between different segments of one's business, an employee can better understand the synergies that exist within the firm. In the case of Nexentia, there were three main synergies that I was exposed to: Universidad de Antioquia, Nexentia itself, and the retailers that sold the product. The Universidad de Antioquia was the original place where the idea and experimentation on the concept of microencapsulation was conceived. Nexentia used their labs as a place to refine the process for it to be then packaged and prepped for retail. Nexentia also had a chemical plant that operated in-house where the calcium was cut and packaged. Finally, the products were shipped to retail locations throughout the country where local businesses who specialized in the selling of animal products sold *Forticaps*.

Theory-wise, Nexentia can easily be characterized as a fully established supply chain. Therefore, they fall squarely in the life cycle of continuation. Nexentia already had synergies in place prior to my arrival that they could leverage for data. The only issue was a difference in systems used in each stage of their business. For Nexentia itself, they were very akin to the concept of using a centralized mainframe such as Google Sheets to record their data. In addition, by using other platforms such as Pipedrive, they were able to keep track of customer-centric data to update their practices with the consumer in mind. For Nexentia, they were faced with a large amount of organizational inertia. The models that will be discussed in the next section worked initially for Nexentia, but due in large part to the existing mechanisms in place, were too complex for continual implementation. Therefore, theory suggests that being at their stage of life cycle (continuation) poses a potential barrier to entry for these consulting-like strategies. A possible route for success would have been to optimize their current practices rather than introducing new methods. Therefore, the pre-established systems would not be changed, but refined.

Intrekit

Intrekit as a company is at an earlier stage in their development than Nexentia. As a company that sells primarily workout leggings with a variety of shirts as well, their process to market takes considerably less steps than Nexentia. There is no need for extensive development of the product (i.e. chemical manufacturing) and the time that it takes to alter a product line can be much faster. This does not mean that their pursuit for a data-centric approach is any less challenging.

A major issue that the company faces is the betterment of their accounting practices. With a more out-of-house approach initially, Intrekit had their data primarily in QuickBooks and Shopify. Although these systems are great for accounting and sales management, the two founders ran into an issue of understanding exactly what the numbers represented. Oftentimes “black box” type services for businesses are extremely beneficial for a firm, but there needs to be deep understanding of all numbers to achieve the maximum benefit. In the case of Intrekit, there existed many discrepancies between sales volume and recorded sales on QuickBooks. Additionally, when an inventory account balance was created for the firm, there still existed many questions regarding the variable and fixed costs calculated for all parts of the business. This (which will be discussed in the next section) created some challenges on the consulting side due to an inability to calculate accurate metrics such as contribution margin.

Another major theory-based question that was far more specific to Intrekit was the idea of aligning their data with their mission. After reading literature and gaining insights from a slew of my entrepreneurial classes at Wharton, I have found that their often exists a lack of direction for founders due to an unclear mission statement. In Intrekit's case, they do have a clear mission: appealing to health-conscious workout enthusiasts that are not avid gym-fanatics. But how can they shape this mission to fit with a new data-centric approach?

For Intrekit it starts with the brand perception. Their social media needs to be far more aligned with what the data is reporting. An example of such implementation is the idea of using influencers to push their brand. In the months of December and January, Intrekit took major hits to their bottom line due to the gifting of leggings and shirts to select influencers. This was done in the attempt to boost viewership on social media platforms and create a positive brand perception outside of their Instagram page.

Theory-wise, this practice makes perfect sense: lose a little money initially in order to secure big gains in the future. However, this could quantitatively be a drawback. They were only able to net a small profit in these two months, which were little help in offsetting fixed monthly costs. A discussion of this will occur more in the next section, but this illustrates how there exists a difference between the theory-based approach to this type of analysis and the quantitative implications of it. With so many items that were being asked to be changed (accounting practices, social media brand, inventory balances), Intrekit is at a point of inflection. The company is looking to begin an upward trend of sales, but the only way to ensure success is to better their data processes. Unlike Nexentia, their success is heavily influenced by change rather than the enhancement of current practices. In the next section, a more thorough analysis of the data behind each firm will be discussed in order to provide a blueprint of recommendations according to a company's point in their life cycle.

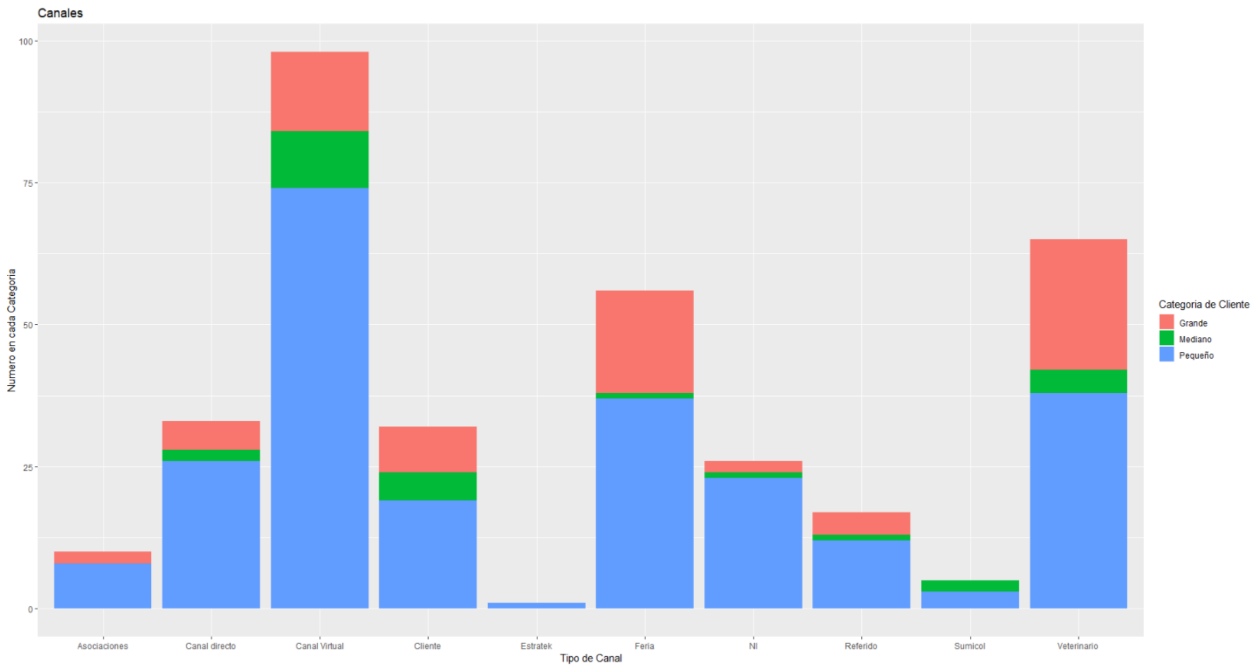
Empirical Results

In this section, the entirety of the data processes will be discussed for each firm. The methods, although different, align themselves with the theory described in the previous section. For Nexentia, the methods revolve primarily around customer analysis with an emphasis on customer retention-based modelling. For Intrekit, the emphasis lies more with accounting practices and better book management. In this section, a thorough analysis of each firm will be described with a transition into conclusions that we can draw solely from the empirical results.

Nexentia

A major concept to note prior to diving into the data associated with Nexentia is truly the robustness of their dataset. With data on purchases dating back to September 2017, data was never an issue for Nexentia. With a dedicated business analytics team that was outsourced to another company, CEO Alexis Sabet Echavarria knew the importance of data centralization. Therefore, they were looking to continue their processes rather than altering them.

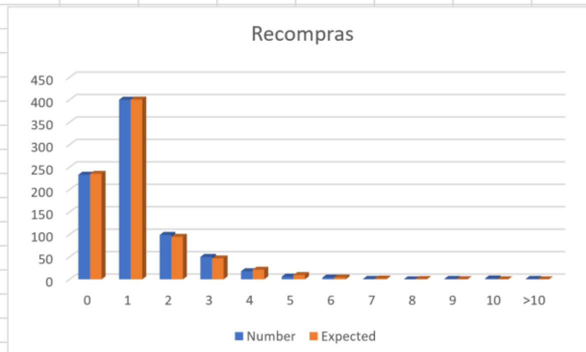
My first major point of implementation was to investigate their distribution of the types of channels that they appealed to. As a company that focused on the sale of products for horses, it was extremely important to know exactly the channels that were resulting in better acquisition of customers. In the chart below, a breakdown of the different acquisition channels is separated with the bars having the different sizes of the horse farm (pequeño being small, mediano as medium, grande as large).



This analysis was done on the statistical coding software of R after a thorough cleaning of the data through Tableau Prep. Despite the use of these software, this could be easily done in Excel, which should encourage the common entrepreneur to engage in this type of analysis. The main bars to note in this graph are bars three and ten; the third being the virtual channel and the tenth being veterinarians. Knowing which channel provided the highest number of customers was critical for Nexentia to be able to allocate resources accordingly. With the virtual channel being their highest, they were able to then decide to upgrade their website. In turn, they wanted to make the brand image as impressive virtually as it appeared in store. Additionally, knowing that veterinarians were creating customers allowed the company to provide more samples to these individuals in order to recommend *Forticaps* to clients. By making data-driven decisions, Nexentia was able to more efficiently allocate funds. Optimization is key in a startup and by using the data available, Nexentia can ensure success without sacrificing the bottom line.

My second contribution to the firm was strongly based off the models discussed in much of Dr. Peter Fader's work on customer centricity. In sum, his research suggests that customers can be reduced to probabilistic measures in order to use for modeling. In layman's terms, we can analyze a customer based on whether they will buy or not buy. Therefore, we only need to calculate the probability of purchase as θ and the probability of a non-purchase as *(1- θ)*. In the subsequent chart, one of Dr. Fader's models can be seen for the total number of repurchases of *Forticaps* over the total duration of the company's existence (2017-2019).

r	1.965832					
alpha	1.688959			-1075		
spike @1	0.280576					
Visits	Number	P(x)	w/spike	LL	Expected	Chi sq
0	233	0.40084	0.288374	-289.735	235.0245	0.017439
1	400	0.293044	0.491399	-284.2	400.4901	0.0006
2	99	0.161609	0.116265	-213.036	94.75619	0.190066
3	50	0.07945	0.057158	-143.097	46.58396	0.250501
4	18	0.036681	0.026389	-65.4264	21.50722	0.571928
5	6	0.016276	0.01171	-26.6841	9.543353	1.315612
6	4	0.007027	0.005056	-21.149	4.120392	0.003518
7	1	0.002974	0.00214	-6.14714	1.743763	0.317235
8	0	0.00124	0.000892	0	0.726781	0.726781
9	1	0.00051	0.000367	-7.90953	0.299289	1.640543
10	2	0.000208	0.00015	-17.613	0.122053	28.89474
>10	1	0.00014	0.000101	-9.19996	0.082351	10.22554
					chisq	44.1545
					df	8
					pvalue	5.32E-07



In the case of this chart, *Recompras* represents the repurchases of the product that occurred in a subsequent period. Therefore, there exists a large quantity of zero repurchases which demonstrates that a group of individuals did not purchase again in the years following a purchase in 2017. This type of model is a negative binomial distribution with a spike at zero. This spike was implemented due to a large quantity of repurchases at zero then a steady drop off after the peak. The fit and shape of the distribution are described by the parameters *r* and α, which show the skewness at the first 3-4 values and a huge drop off thereafter. Additionally, the *p*-value associated with the X^2 Goodness-of-fit test demonstrates an extremely accurate fit on the model. The other three models for Nexentia with the negative binomial distribution can be found in the Appendix under Models A, B, and C.

With all the statistical principles of the NBD (negative binomial distribution) model validated, Nexentia could look at this model and make data-driven decisions. One of the key recommendations that I made to the firm and which can be replicated in all business formats, is to the look at the distribution of each repurchase amount relative to the expectation. For example, for individuals who repurchased twice, the actual number was greater than the expectation. Looking forward, the repurchase of a fourth time is higher in the expectation than the actual amount. Therefore, a firm could specifically tailor practices to customers based upon their purchase history. By knowing that a customer repurchased three times already, Nexentia could potentially engage in more outreach efforts to convert a fourth repurchase. Hence, they could try to bring the actual bar closer to expectation. By filtering customers down into probabilistic measures, a firm can justify tailored practices for different types of customers. In the case of this model, a firm could better understand the purchase history of their clients and look to ensure customer retention past the initial buy.

These two models were the main types of analyses that were constructed for Nexentia. By having such large quantities of data, analyses such as these are very feasible in any company. The only issue and the reason why adoption of these methods post-consulting did not occur was the life cycle of the startup. Positioned in the continuation stage of their operations, Nexentia did not have the time to continually update these NBD models. If they were more in the inflection stage, I believe that they would be far more committed to the change. For the more general entrepreneur, methods that involve Excel are far more feasible due to the learning curve of advanced software such as R or these statistical models and can yield similar analyses. Similar to the basic graph described initially, bar charts are a fantastic way to highlight a specific aspect of a business and make data-driven decisions. In retrospect, enhancing the existing databases would have been far more beneficial for Nexentia in order to continue their path towards optimization rather than giving tools that require a steeper learning curve.

Intrekit

Relative to Nexentia, Intekit's data is far less robust and more decentralized. As described prior, the company has a large quantity of its data stored in QuickBooks and Shopify where they look to use these mediums to mitigate their accounting. When I first began to assist with Intrekit, I asked them to construct a few inventory and cost sheets to better understand their business. These sheets can be found in the Appendix under Models D through I.

One of the main findings that I found from Intrekit was a large discrepancy across their associated costs. For example, Intrekit orders all their legging inventory from China which goes through a process of material cutting, printing, and shipping. These, therefore, constitute the variable costs associated with each item. However, the founders decided to include the retail bags into the variable cost calculations. Although an innocent mistake on the surface, variable costs go into all calculations in accounting. With this in mind, many of their numbers for the subsequent calculations are slightly off, which emphasizes the purpose of diligence in book-keeping.

A recommendation that I had to lessen the confusion over where to document the retail bags was to place these items in fixed costs and record the total in inventory balances. As seen in Model D, they were able to place the two Etsy bag orders as fixed costs due to the infrequency of purchase. By disaggregating the final product (a pair leggings within a bag), it is far easier to draw conclusions regarding the contribution margin of your product. For example, companies can look to create a breakeven analysis of their products. With the equation, $0 = CM * Q - FC$, where CM is the contribution margin (price per unit less variable costs per unit), Q is quantity of items, and

FC is fixed goods in the time period, a firm can effectively solve for the quantity of goods needed to breakeven.

In the case of Intrekit, they produced a range of products from leggings to workout shirts which complicates this analysis. Therefore, an easier application of this formula can be used to see how much of their highest margin item should they look to sell. With highest contribution margin existing for their leggings at $21.44 as seen in Model F in the Appendix, I was able to see how many sales of just leggings they would need to produce to breakeven in each month. With a monthly fixed cost of $542.37 (not including credit card sales and discounts on items), the breakeven quantity for a month would be 26 leggings. If we were to look at other items as well, this legging number would lessen, and the relevant number of shirts needed would increase. This type of analysis is central for a business because it creates an accurate metric for success each month. However, with discrepancies in book-keeping, a firm could have an inaccurate breakeven quantity. For Intrekit, bettering these practices would ensure that these metrics provide a clear picture of the state of the business. By removing the bag cost from variable cost per legging going forward, they should hopefully continue to implement this formula for future months. Intrekit is an excellent case study for the relevance of a simplistic formula or model, but the need to be diligent in recording one's costs.

Another major idea to discuss regarding the financial portion of Intrekit's business is their willingness to discount their items. As described in Model F, the firm has a wide variety of discounts with the most common being $10 off the initial purchase and 20% off the subsequent order. Discounts are an excellent way to expand one's clientele but can be a potential pitfall for accounting purposes. As seen in the breakeven analysis above, the model responds well to full price sales when calculating the contribution margin. However, the margin fluctuates significantly when the various discounts are included. Additionally, the firm decided to provide free samples to influencers on Instagram in the effort to promote their product. Theory-wise this makes perfect sense: build the brand and reap the rewards. But accounting-wise this resulted in a dozen cases of negative hits to the balance sheet. Every free item given away is an expense of the variable cost of its production. So, were the free samples worth it? Intrekit had their largest sales volume ever in the month of February (January was the month of the most samples) and many of whom were drawn to the company by those influencers.

Conclusions

From the analysis described above, a few main themes can be gleaned. The first main theme is that complexity is not always superior. In the case of the NBD models described for Nexentia, I feel there exists a significant learning curve for the common entrepreneur. Upper level statistical modeling is extremely hard to implement initially let alone continually. Going forward, a point of inflection is a far better position to investigate these types of models, but only when looking at a more customer-centric approach to data. For Nexentia, enhancement was clearly the better strategy. With similar analyses as the R output, a company can more accurately look at their customer acquisition channels. From there, they can then specifically target these segments to maximize their marketing efforts.

Another main empirical conclusion that can be drawn from this section is the idea of accounting diligence. As mentioned throughout this paper, being attentive to one’s book-keeping is crucial when making any type of financial analyses. This can start by clearly knowing the nature of a fixed versus a variable cost. In the case of Intrekit, variable costs were only the items that went directly into the making of legging or shirt. A retail bag, although 1:1 with the item it contained, cannot be classified as a variable cost. Therefore, when these items were disaggregated, more accurate breakeven analyses could be calculated.

Finally, why was one method of data centralization more successful than the other? The reason still lies in the concept of inflection versus continuation. A firm should look to enhance their synergies if they are in a push for continual growth whereas a firm should be open to organizational change when looking towards a new frontier.

Conclusion

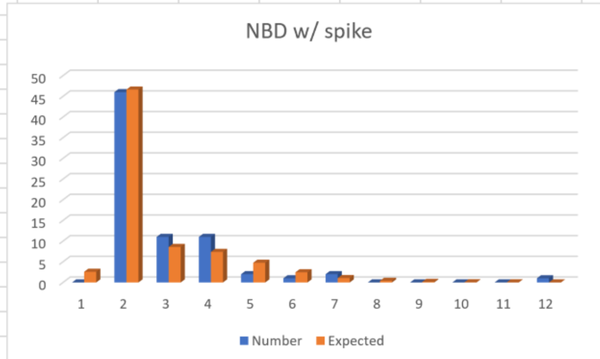
Looking at the two main emphases of this paper, implementation and adoption, we have netted a slew of interesting observations regarding the nature of startups. With future research in mind, an interesting application of these findings would be to do a time series analysis to see the results over the course of at least a year. By having a longer time horizon, even more convincing conclusions could be made to persuade more entrepreneurs to engage in data-centric practices.

Additionally, theory does not always equate to empirical results. As evidenced in the high applicability of the NBD models, the firm remained unwilling to adopt these practices post-consulting. However, when we consider the stage in the startup's life cycle, whether at inflection or continuation, we can see the marriage between theory and application. This paper has concluded that data centralization is crucial to a startup's success, but the nature of what models or formulas to implement is based not in their complexity, but the stage of growth.

Appendix

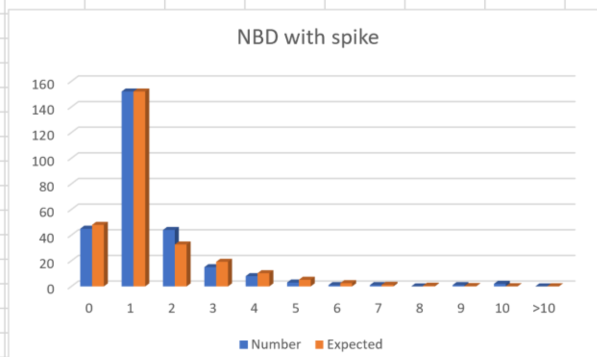
Model A: NBD Model for 2017 for Nexentia

r	149246.9					
alpha	57870.87			-87.6981		
spike @1	0.540187					
Visits	Number	P(x)	w/spike	LL	Expected	Chi sq
0	0	0.075854	0.034879	0	2.581026	2.581026
1	46	0.195622	0.630136	-21.2437	46.63009	0.008514
2	11	0.252248	0.115987	-23.697	8.583043	0.680607
3	11	0.216846	0.099708	-25.3606	7.378427	1.777586
4	2	0.13981	0.064286	-5.48882	4.75719	1.598023
5	1	0.072114	0.033159	-3.40645	2.453749	0.861289
6	2	0.030997	0.014253	-8.50161	1.054706	0.847232
7	0	0.01142	0.005251	0	0.388587	0.388587
8	0	0.003682	0.001693	0	0.125273	0.125273
9	0	0.001055	0.000485	0	0.035898	0.035898
10	0	0.000272	0.000125	0	0.009258	0.009258
>10	1	8.08E-05	3.72E-05	-10.2001	0.00275	361.5822
					chisq	370.4955
					df	8
					pvalue	3.8E-75



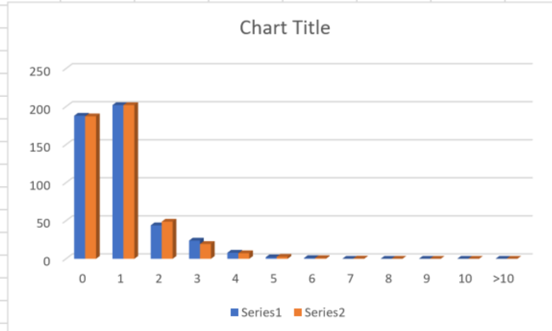
Model B: NBD Model for 2018 for Nexentia

r	2.579284					
alpha	1.603258			-370.379		
spike @1	0.384006					
Visits	Number	P(x)	w/spike	LL	Expected	Chi sq
0	45	0.286434	0.176442	-78.0644	47.99221	0.186557
1	152	0.283797	0.558823	-88.4523	151.9998	2.9E-10
2	44	0.195099	0.12018	-93.2256	32.689	3.913817
3	15	0.114397	0.070468	-39.789	19.16729	0.906039
4	8	0.061294	0.037757	-26.2128	10.2698	0.501665
5	3	0.030982	0.019085	-11.8766	5.191029	0.924789
6	1	0.015034	0.009261	-4.68197	2.518913	0.91591
7	1	0.007078	0.00436	-5.4353	1.185903	0.029142
8	0	0.003256	0.002005	0	0.545475	0.545475
9	1	0.00147	0.000906	-7.00699	0.246304	2.306327
10	2	0.000654	0.000403	-15.6342	0.109556	32.62057
>10	0	0.000506	0.000312	0	0.084734	0.084734
					chisq	42.93503
					df	8
					pvalue	9.04E-07



Model C: NBD Model for 2019 for Nexentia

r	1.912542					
alpha	2.267828			-568.456		
spike @1	0.197014					
Visits	Number	P(x)	w/spike	LL	Expected	Chi sq
0	188	0.497253	0.399287	-172.598	187.2655	0.002881
1	202	0.291024	0.430702	-170.152	201.9993	2.48E-09
2	44	0.129692	0.104141	-99.5286	48.84191	0.479999
3	24	0.05176	0.041562	-76.3336	19.49266	1.042243
4	8	0.019453	0.01562	-33.2735	7.325854	0.062037
5	2	0.007039	0.005652	-10.3514	2.650961	0.159848
6	1	0.002482	0.001993	-6.21823	0.93461	0.004575
7	0	0.000858	0.000689	0	0.323288	0.323288
8	0	0.000293	0.000235	0	0.110215	0.110215
9	0	9.86E-05	7.92E-05	0	0.037147	0.037147
10	0	3.29E-05	2.64E-05	0	0.012405	0.012405
>10	0	1.63E-05	1.31E-05	0	0.006122	0.006122
					chisq	2.240759
					df	8
					pvalue	0.972727



Model D: Fixed Costs for Intrekit

Intrekit Store

Category	Item	Company / Vendor	Status	Cost		Billed / Plan	Notes		SUB-TOTAL	
				Monthly	Yearly				Monthly	Yearly
									$474.75	$6,027.00
Marketing	Email Marketing	Klaviyo	Active	$30.00	$360.00	Monthly	Every 22nd of the month			
Marketing	Content Creator	Envato Element	Active	$16.50	$198.00	Yearly	February 27 every year			
Marketing	Ads	Marketing Budget	Active	$300.00	$3,600.00	Monthly				
Operations	Web Hosting	Shopify	Active	$29.00	$348.00	Monthly	Every 29th of the month			
Operations	Web Hosting Theme	Shopify	Active	$6.58	$79.00	Yearly	March 12 every year			
Operations	Returns	AfterShip Returns	Active	$10.00	$120.00	Monthly	Currently FREE version, 5 returns quota per month, $2 per extra return, 20 returns quota per month, $0.5 per extra return, etc...			
Operations	PO BOX	USPS	Active	$21.17	$254.00	Yearly	March 16 every year			
Operations	Finance	Quickbooks	Active	$25.00	$300.00	Monthly	Every 28th of the month			
Operations	Email	G-Suite	Active	$24.00	$288.00	Monthly	Every 3rd of the month			
Operations	Business Renewal	Sunbiz Renewal	Active	$12.50	$150.00	Yearly	April 15 of every year			
Packaging	Frosted Clear Bags	Etsy Vendor	Active		$180.00	Yearly	This is for 700 pcs, we've only bought once and we will now be restocking			
Packaging	Polymails	Etsy Vendor	Active		$150.00	Yearly	This is for 700 pcs, we've only bought once and we will now be restocking			

Model E: Merchant Costs, Loans, and Intrekit Fitness Club

Merchant CC Charges

Category	Item	Company / Vendor	Status	Credit Card Rate			Notes
CC	Merchant Fee	Shopify	Active	2.9% + .30			Per charge when customer paid directly with Shopify store
CC	Merchant Fee	Paypal	Active	2.9% + .30			Per charge when customer check out with Paypal payment
CC	Merchant Fee	Afterpay	Active				

Investment Loan Balance

Category	Item	Company / Vendor	Status	Credit Card Rate			Notes
Loans	Investor	Lourdes / Tia	Active	$0.00	$4,500.00		Currently Owe

Intrekit Fitness Club

Category	Item	Company / Vendor	Statues	Cost		Billed / Plan	Notes		SUB-TOTAL	
				Monthly	Yearly				Monthly	Yearly
									$40.12	$481.39
Marketing	Ads	Marketing Budget	Active	$0.00	$0.00	Monthly				
Hosting	Website Hosting	Site Ground	Active	$11.95	$143.40	Yearly				
Hosting	Domain Hosting	Namecheap	Active	$1.25	$15.00	Yearly				
Theme	WP Theme	OceanWP	Active	$6.25	$74.99	Yearly				
Plug-in	WP Page builder	Elementor	Active	$8.25	$99.00	Yearly				
Plug-in	User profile and login	Profile Builder Pro	Active	$12.42	$149.00	Yearly				

Model F: Variable Costs for Intrekit

Products

Item	Beginning Purchase Quantity - 2019	Unit Cost	Logo Set Up	Care Label Set Up	Shipping for whole Order	Actual Unit Cost	Previous Actual Unit Cost	Selling Price	Contribution Margin	Notes
Enzo Classic Leggings	560	11.25	$0.09	$0.09	$2.13	$13.56	$14.32	$35.00	$21.44	
Graphic Crop Top (Intrekit)	15	11.09			Delivered to us	$11.09		$25.00	$13.91	Restocked in 2020
Graphic Crop Top (NGU)	15	11.09			Delivered to us	$11.09		$25.00	$13.91	Restocked in 2020
Graphic Crop Top (Strong AF)	15	12.09			Delivered to us	$12.09		$25.00	$12.91	Restocked in 2020
Serene Seamless X Crop Top	50	2.75			$0.37	$3.12	$5.58	$13.00	$9.88	Restocked in 2020
Enzo Open Back Mesh Top	75	7.5	$0.28	$0.06	$0.79	$8.63	$9.29	$18.00	$9.37	Will not restock, item is on sale

Discounts Offered

Code	Amount	Reason
FIRSTPAIR	$10.00	Our Enzo Classic legging is $35. If customer is first time buyer they are allowed to use code FIRSTPAIR to get their first pair of Enzos for $25, If they end up buying two, the discount code takes $5 off of each legging, still making one legging for $25
INTREKIT20	20% OFF	First time buyers will receive a discount code in their package of 20% OFF for there next order
JOVI15, STEPH15, GINA15, etc.	15% OFF	Each Affiliate has a discount code for their followers to receive 15% off if the customer uses there code
ATHLETES30	30% OFF	Affiliates recieve 30% off for their personal orders
INSPIRED15	15% OFF	Discount code we offer in our email marketing
BLACKFRIDAY	40% OFF	The only time we offer 40% OFF is black friday/cyber monday weekend
	15% OFF - 30% OFF	On holidays or special occasions we will offer discounts ranging from 15% - 30% OFF

Shipping Deals

Shipping Rate	Minimum Amount Purchase
FREE SHIPPING	$0.00
FREE SHIPPING OVER $40	$40.00
FREE SHIPPING OVER $50	$50.00

Model G: Sales in December 2020 for Intrekit

Order	Product	Gross sales	Discounts	Returns	Net sales	Taxes	Shipping	Total sales		VC to produce	CM	% made relative to Selling Price
#1148		0	0	0	0	0.27	3.82	4.09		0	4.09	
#1148	Enzo Classic Leggings - Ruby Red	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1148	Intrekit Crop Top	25	-10	0	15	1.05	0	16.05		$11.09	$4.96	19.84%
#1149		0	0	0	0	0	3.82	3.82		0	3.82	
#1149	Never Give Up Crop Top	25	-10	0	15	1.05	0	16.05		$11.09	$4.96	19.84%
#1149	Enzo Classic Leggings - Midnight Gray	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1150	Serene Seamless X Crop Top - Black	16	-6.4	0	9.6	0	0	9.6		$3.12	$6.48	40.51%
#1150	Enzo Open Back Mesh Top - Heather White	25	-10	0	15	0	0	15		$8.63	$6.37	25.50%
#1150	Enzo Open Back Mesh Top - Black	25	-10	0	15	0	0	15		$8.63	$6.37	25.50%
#1150	Enzo Classic Leggings - Black	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1151	Enzo Classic Leggings - Ruby Red	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1151		0	0	0	0	0	5.24	5.24		0	5.24	
#1151	Serene Seamless X Crop Top - Black	16	-6.4	0	9.6	0	0	9.6		$3.12	$6.48	40.51%
#1152	Enzo Classic Leggings - Superwoman Blue	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1152		0	0	0	0	0	3.82	3.82		0	3.82	
#1153	Strong AF Crop Top	25	-10	0	15	1.05	0	16.05		$12.09	$3.96	15.84%
#1153		0	0	0	0	0	3.82	3.82		0	3.82	
#1153	Enzo Classic Leggings - Black	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1154	Enzo Classic Leggings - Superwoman Blue	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1154	Enzo Classic Leggings - Ruby Red	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1154	Enzo Classic Leggings - Midnight Gray	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1155	Enzo Classic Leggings - Black	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1155		0	0	0	0	0.27	3.82	4.09		0	4.09	
#1156	Never Give Up Crop Top	25	-10	0	15	0	0	15		$11.09	$3.91	15.64%
#1156	Enzo Classic Leggings - Black	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1156	Enzo Classic Leggings - Midnight Gray	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1157	Enzo Classic Leggings - Black	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1157	Enzo Classic Leggings - Midnight Gray	35	-14	0	21	1.47	0	22.47		$13.56	$8.91	25.45%
#1157	Never Give Up Crop Top	25	-10	0	15	1.05	0	16.05		$11.09	$4.96	19.84%
#1158		0	0	0	0	0	4.06	4.06		0	4.06	
#1158	Enzo Classic Leggings - Black	35	-14	0	21	0	0	21		$13.56	$7.44	21.25%
#1159	Enzo Classic Leggings - Midnight Gray	35	-10	0	25	0	0	25		$13.56	$11.44	32.68%
#1159		0	0	0	0	0	4.33	4.33		0	4.33	
#1160	Enzo Classic Leggings - Black	35	-10	0	25	1.75	0	26.75		$13.56	$13.19	37.68%
#1160		0	0	0	0	0	3.82	3.82		0	3.82	
#1161	Enzo Classic Leggings - Midnight Gray	35	-34	0	1	0.07	0	1.07		$13.56	-$12.49	-35.69%
#1162	Enzo Classic Leggings - Ruby Red	35	-35	0	0	0	0	0		$13.56	-$13.56	-38.75%
#1162	Enzo Classic Leggings - Black	35	-35	0	0	0	0	0		$13.56	-$13.56	-38.75%
#1162	Enzo Classic Leggings - Midnight Gray	35	-35	0	0	0	0	0		$13.56	-$13.56	-38.75%
#1163		0	0	0	0	0	3.96	3.96		0	3.96	
#1163	Enzo Classic Leggings - Black	35	-10	0	25	0	0	25		$13.56	$11.44	32.68%
#1164	Enzo Classic Leggings - Midnight Gray	35	-26.25	0	8.75	0.62	0	9.37		$13.56	-$4.19	-11.98%
#1164	Enzo Classic Leggings - Black	35	0	0	35	2.45	0	37.45		$13.56	$23.89	68.25%
#1165	Enzo Classic Leggings - Black	35	0	0	35	0	0	35		$13.56	$21.44	61.25%
#1165	Enzo Classic Leggings - Ruby Red	35	-26.25	0	8.75	0	0	8.75		$13.56	-$4.81	-13.75%
#1166	Enzo Classic Leggings - Black	35	-10	0	25	0	0	25		$13.56	$11.44	32.68%
#1166		0	0	0	0	0	4.33	4.33		0	4.33	
#1166	Serene Seamless X Crop Top - Black	10	0	0	10	0	0	10		$3.12	$6.88	68.82%
	From Shopify	$1,162.00	-$524.30	$0.00	$637.70	$19.92	$44.84	$702.46		$449.24	$253.22	
										After 2.9%	$245.87	
	Quickbooks Gross Sales	$1,356.35								**After .30 surcharge**	$240.17	20.67%

Model H: Sales in January 2020 for Intrekit

Order	Product	Gross sales	Discounts	Returns	Net sales	Taxes	Shipping	Total sales	VC to produce	CM	% made relative to Selling Price
#1167	Enzo Classic Leggings - Superwoman Blue	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1167	Intrekit Crop Top	18	-18	0	0	0	0	0	$11.09	-$11.09	-61.61%
#1168	Enzo Classic Leggings - Black	35	0	0	35	0	0	35	$13.56	$21.44	61.25%
#1168	Enzo Classic Leggings - Black	35	-26.25	0	8.75	0	0	8.75	$13.56	-$4.81	-13.75%
#1169		0	0	0	0	0	3.63	3.63	$0.00	$3.63	
#1169	Enzo Classic Leggings - Superwoman Blue	35	0	0	35	0	0	35	$13.56	$21.44	61.25%
#1170	Never Give Up Crop Top	25	0	0	25	1.75	0	26.75	$11.09	$15.66	62.64%
#1170	Enzo Classic Leggings - Black	35	-10	0	25	1.75	0	26.75	$13.56	$13.19	37.68%
#1171	Enzo Classic Leggings - Superwoman Blue	35	-10	0	25	1.75	0	26.75	$13.56	$13.19	37.68%
#1171		0	0	0	0	0.28	3.93	4.21	$0.00	$4.21	
#1172	Enzo Classic Leggings - Black	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1173	Enzo Classic Leggings - Midnight Gray	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1174	Enzo Classic Leggings - Superwoman Blue	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1175	Enzo Classic Leggings - Ruby Red	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1176	Serene Seamless X Crop Top - Navy Blue	10	-10	0	0	0	0	0	$3.12	-$3.12	-31.18%
#1176	Enzo Classic Leggings - Superwoman Blue	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1177	Enzo Classic Leggings - Superwoman Blue	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1178	Enzo Classic Leggings - Black	35	-5	0	30	2.1	0	32.1	$13.56	$18.54	52.96%
#1178	Enzo Classic Leggings - Midnight Gray	35	-5	0	30	2.1	0	32.1	$13.56	$18.54	52.96%
#1178	Enzo Open Back Mesh Top - Black	15	0	0	15	1.05	0	16.05	$8.63	$7.42	49.50%
#1179	Enzo Classic Leggings - Black	35	-5	0	30	0	0	30	$13.56	$16.44	46.96%
#1179	Enzo Classic Leggings - Midnight Gray	35	-5	0	30	0	0	30	$13.56	$16.44	46.96%
#1180	Enzo Classic Leggings - Superwoman Blue	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
#1181	Enzo Classic Leggings - Black	35	-10	0	25	0	0	25	$13.56	$11.44	32.68%
#1181		0	0	0	0	0	4.18	4.18	$0.00	$4.18	
#1182	Enzo Classic Leggings - Ruby Red	35	-35	0	0	0	0	0	$13.56	-$13.56	-38.75%
	From Shopify	$733.00	-$419.25	$0.00	$313.75	$10.78	$11.74	$336.27	$291.61	$44.66	
									After 2.9%	$43.36	
	Quickbooks Gross Sales	226.84							After .30 surcha	$38.56	5.26%

Model I: Inventory Balances by Size for Intrekit

December	XS	S	M	L	XL	Total Cost (Using VC)
Leggings	54	109	96	46	62	4977.44
Open Back Mesh Top	0	13	19	15	0	405.41
Intrekit Crop Top	0	1	3	2	0	66.54
NGU Crop Top	0	0	0	1	0	11.09
Seamless Crop Top	0	0	0	8	0	24.95
Strong AF Crop Top	0	0	5	4	0	108.81
						5594.23

January	XS	S	M	L	XL	Total Cost (Using VC)
Leggings	50	100	92	46	60	4719.75
Open Back Mesh Top	0	13	19	14	0	396.78
Intrekit Crop Top	0	0	3	2	0	55.45
NGU Crop Top	0	0	0	0	0	0.00
Seamless Crop Top	0	0	0	7	0	21.83
Strong AF Crop Top	0	0	5	4	0	108.81
						5302.62

February	XS	S	M	L	XL	Total Cost (Using VC)
Leggings	34	37	33	37	34	2373.44
Open Back Mesh Top	0	3	11	11	0	215.64
Intrekit Crop Top	0	5	0	0	0	55.45
NGU Crop Top	0	0	9	10	0	210.71
Seamless Crop Top	0	0	0	0	0	0.00
Strong AF Crop Top	0	5	0	0	0	60.45
						2915.69

References

Ahmed, Bakhtiyar, Thomas Dannhauser, and Nada Philip. 2018. "A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects."IEEE, .

Cheng, Colin CJ and Eric C. Shiu. 2019. "How to Enhance SMEs Customer Involvement using Social Media: The Role of Social CRM." *International Small Business Journal* 37 (1): 22-42.

Ciampi, Francesco, Giacomo Marzi, and Riccardo Rialti. 2018. "Artificial Intelligence, Big Data, Strategic Flexibility, Agility, and Organizational Resilience: A Conceptual Framework Based on Existing Literature." .

Fader, P. S., B. G. S. Hardie, (2010) Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity. Marketing Science 29(1):85-93.

"How Integrating CRM and Marketing Automation can Help Startups Drive Sales.", accessed Sep 4, 2019, https://link-galegroup-com.proxy.library.upenn.edu/apps/doc/A528936064/STND?sid=lms.

"How Startups and SME's can Leverage Open Source CRM to Increase Business.", accessed Sep 4, 2019, https://link-galegroup-com.proxy.library.upenn.edu/apps/doc/A528183451/ITOF?sid=lms.

Krishna, Amar, Ankit Agrawal, and Alok Choudhary. Dec 2016. "Predicting the Outcome of Startups: Less Failure, More Success."IEEE, . doi:10.1109/ICDMW.2016.0118. https://ieeexplore.ieee.org/document/7836749.

Peltier, James W., John A. Schibrowsky, and Yushan Zhao. 2009. "Understanding the Antecedents to the Adoption of CRM Technology by Small Retailers: Entrepreneurs Vs Owner-Managers." *International Small Business Journal* 27 (3): 307-336.

Provost, Foster and Tom Fawcett. 2013. *Data Science and its Relationship to Big Data and Data-Driven Decision Making*. Vol. 1. doi:10.1089/big.2013.1508.

Weber, Ben G. 2018. "Data Science for Startups." .

Wynn, Martin, Phillip Turner, Amitave Banik, and Gary Duckworth. 2016. "The Impact of Customer Relationship Management Systems in Small Business Enterprises." *Strategic Change* 25 (6): 659-674.