INTEGRATED ORGANIZATIONAL MACHINE LEARNING FOR AVIATION FLIGHT DATA

MICHAEL J. PRITCHARD, PAUL THOMAS, ERIC WEBB, JON MARTIN, & AUSTIN WALDEN

NATIONAL TRAINING AIRCRAFT SYMPOSIUM

OCTOBER, 2022

KANSAS STATE UNIVERSITY | Salina Aerospace Campus

SITUATION

Major challenges face many flight organizations:

1. Integration and automation of data collection frameworks
2. Data feature engineering, cleanup and preparation
3. Operationalizing embedded machine learning frameworks

CHALLENGES

While integration and automation of data collection efforts within many organizations is quite mature…

…there are special challenges for flight-based organizations (i.e., the automatic and efficient transmission of aircraft flight data to centralized analytical data processing systems).

OPPORTUNITY

- Constraints for implementing classical machine learning methods (i.e., clustering, classification, or prediction)
- This magnifies design challenges for novel 'prescriptive-based' architectures

Our research is focused on a design pattern for:

a) The integration and automation of data collection for…

b) …an organizationally embedded ensemble machine learning method

APPLIED RESEARCH QUESTIONS

1. Identify challenges associated with the integration and automation of fleet data collection frameworks
2. Determine feature engineering, cleanup and preparation processes
3. Operationalizing embedded machine learning frameworks

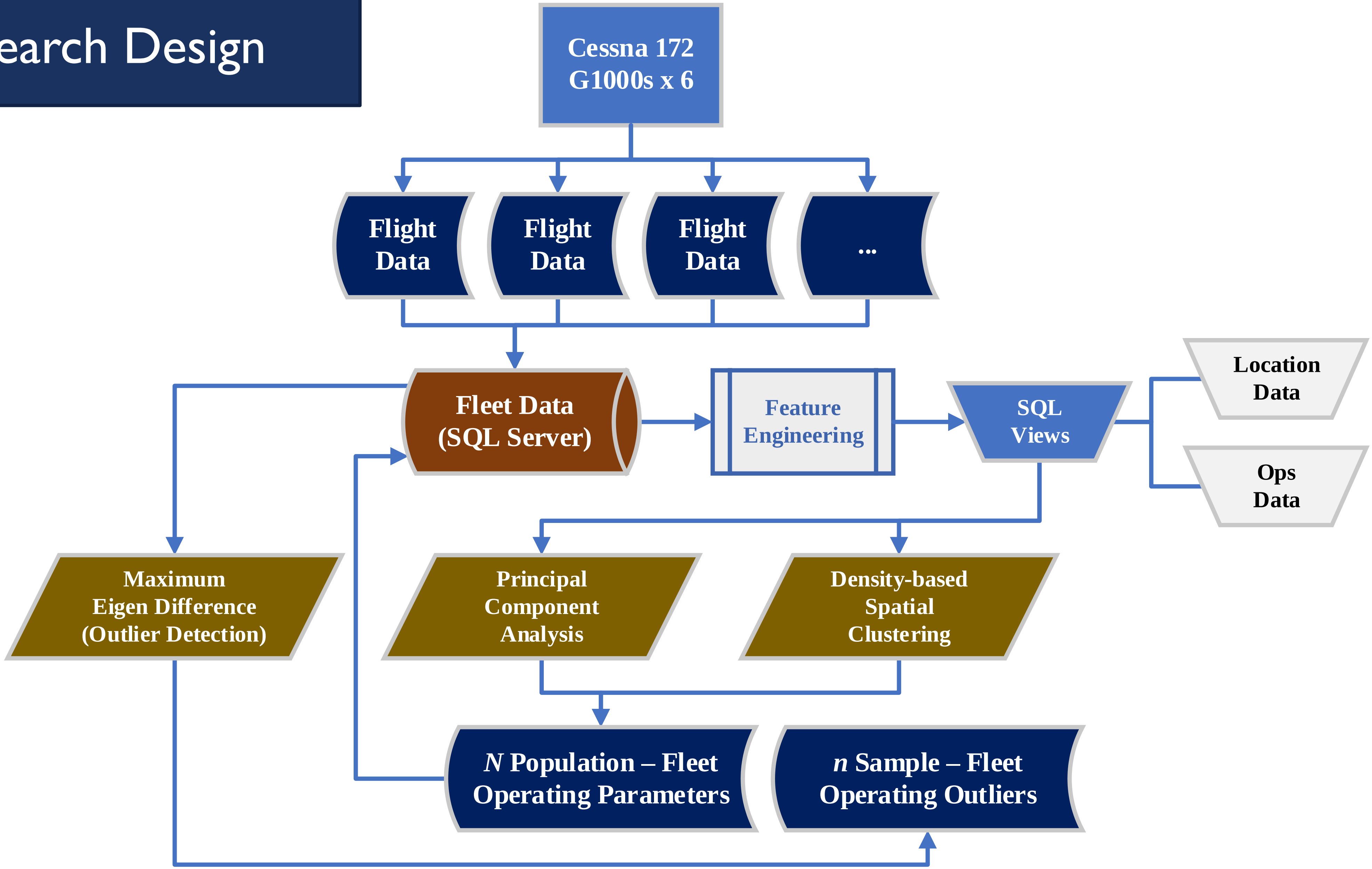
BACKGROUND RESEARCH, PART I

- **Airplane monitoring systems** have been around for several decades…
  - (Taylor, 1969; Milligan, Zhou, and Wilkerson, 1995).
- …data from sensors for **location, structure**, **engine**, and **cabin environment**…
  - (Gao et al., 2018).
- …monitoring systems are ***wired*** and ***wireless; and are used to enhance and predict*** maintenance…
  - (Zelenika et al., 2020).

BACKGROUND RESEARCH, PART II

- **Prevalence of monitoring systems** and the prompt analysis of data from collected from fleets can allow for more timely and effective maintenance activities which will **reduce aircraft downtime** while also **reducing operational costs** arising from maintenance (Dupuy, Wesely, and Jenkins, 2011).
- There has been a trend for **applying statistical techniques to data collected from fleets** of commercial aircraft **to identify aircraft anomalies** or abnormal trends (Gorinesky, Matthews, and Martin, 2012; Sumathi et al., 2017).

Research Design



Data Capture

Microsoft® SQL Server®

n = 65,525 (flight log entries)

Data Framework

Flight Logs

SQLQuery9.sql...RS\mjp001 (64))* SQLQuery8.sql - not connected* SQLQuery6.sql - not co

```sql
USE [FleetData]
GO

/****** Object:  View [dbo].[v_flight_log]    Script Date: 10/24/2022 4:06:3

alter view [dbo].[v_flight_log]
as
select '164459_KSLN' as Location, * from [dbo].[log_170118_164459_KSLN]
union
select '130128_KSLN' as Location, * from [dbo].[log_170130_130128_KSLN]
union
select '155300_KMKC' as Location, * from [dbo].[log_170130_155300_KMKC]
union
select '155300_KMKC' as Location, * from [dbo].[log_180308_143550_KFOE]
union
select '155300_KMKC' as Location, * from [dbo].[log_180504_185459_KHSD]
union
select '155300_KMKC' as Location, * from [dbo].[log_180919_002527_KJEF]
union
select '155300_KMKC' as Location, * from [dbo].[log_190914_001545_KPYX]
union
select '155300_KMKC' as Location, * from [dbo].[log_191025_185233_KSLN]
union
```

121 %

Messages

```
Commands completed successfully.

Completion time: 2022-10-24T16:06:53.3805350-05:00
```

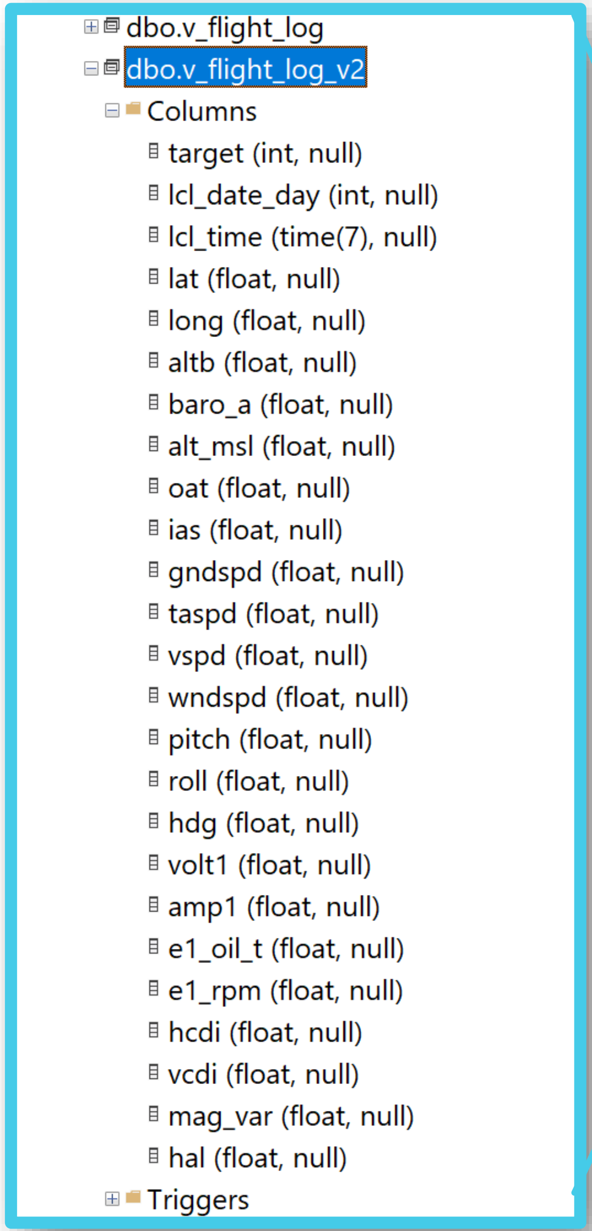
SQLQuery9.sql...RS\mjp001 (64))* SQLQuery8.sql...RS\mjp001 (68))* SQLQuery6.sql

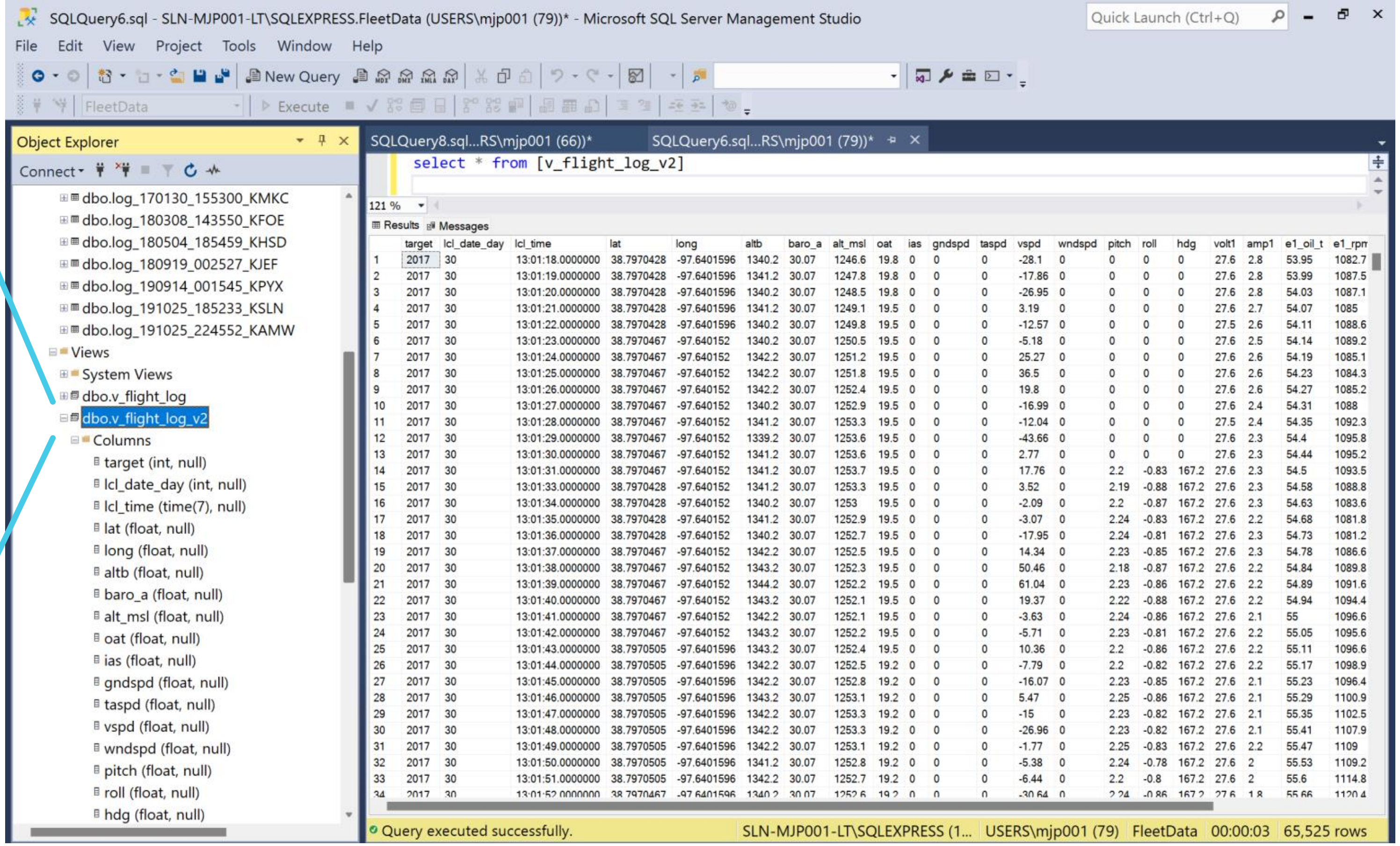
```sql
USE [FleetData]
GO

/****** Object:  View [dbo].[v_flight_log_v2]    Script Date: 10/24/202
alter view [dbo].[v_flight_log_v2]
as
SELECT
-- isnull logic on everything
convert(int, isnull(year([Lcl_Date]), 0)         ) as target,
convert(int, isnull(day([Lcl_Date]), 0)          ) as lcl_date_day,
convert(time, isnull([Lcl_Time], 0)              ) as lcl_time,
convert(float, isnull([Latitude], 0)             ) as lat,
convert(float, isnull([Longitude], 0)            ) as long,
convert(float, isnull([AltB], 0)                 ) as altb,
convert(float, isnull([BaroA], 0)                ) as baro_a,
convert(float, isnull([AltMSL], 0)               ) as alt_msl,
convert(float, isnull([OAT], 0)                  ) as oat,
convert(float, isnull([IAS], 0)                  ) as ias,
convert(float, isnull([GndSpd], 0)               ) as gndspd,
convert(float, isnull([TAS], 0)                  ) as taspd,
convert(float, isnull([VSpd], 0)                 ) as vspd,
convert(float, isnull([WndSpd], 0)               ) as wndspd,
convert(float, isnull([Pitch], 0)                ) as pitch,
convert(float, isnull([Roll], 0)                 ) as roll,
convert(float, isnull([HDG], 0)                  ) as hdg,
convert(float, isnull([volt1], 0)                ) as volt1,
convert(float, isnull([amp1], 0)                 ) as amp1,
convert(float, isnull([E1_OilT], 0)              ) as e1_oil_t,
convert(float, isnull([E1_RPM], 0)               ) as e1_rpm,
convert(float, isnull([HCDI], 0)                 ) as hcdi,
```

Centralized Data View

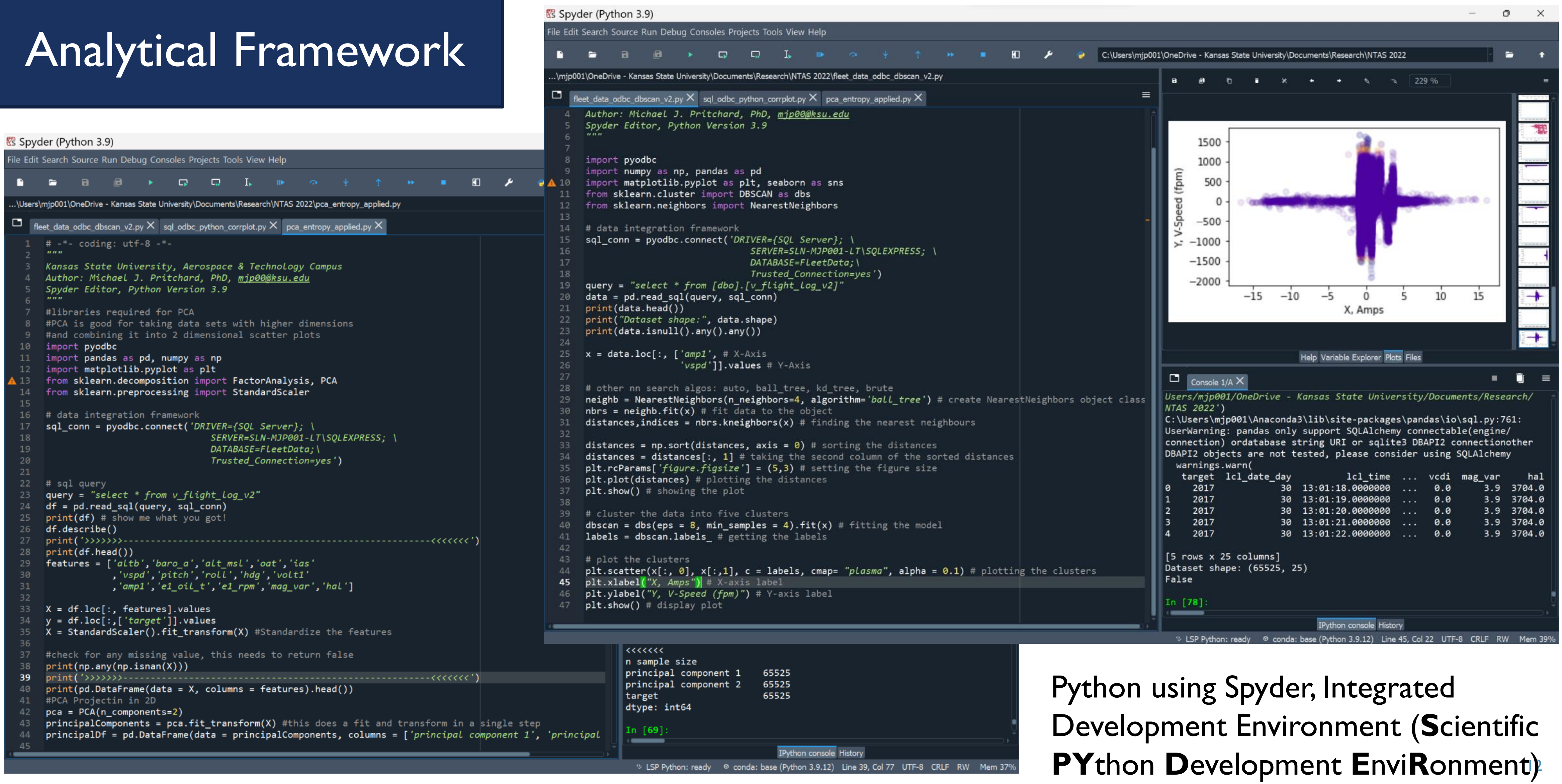
Centralized Data View





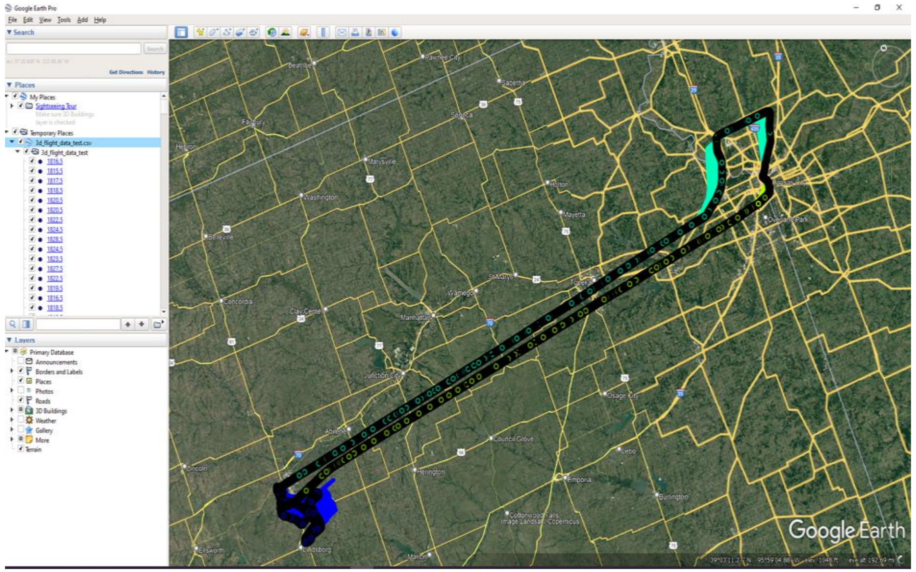
G1000 Data Definitions: Fala, N. (2019). *Data-driven safety feedback as part of debrief for General Aviation pilots* (Doctoral dissertation, Purdue University Graduate School).

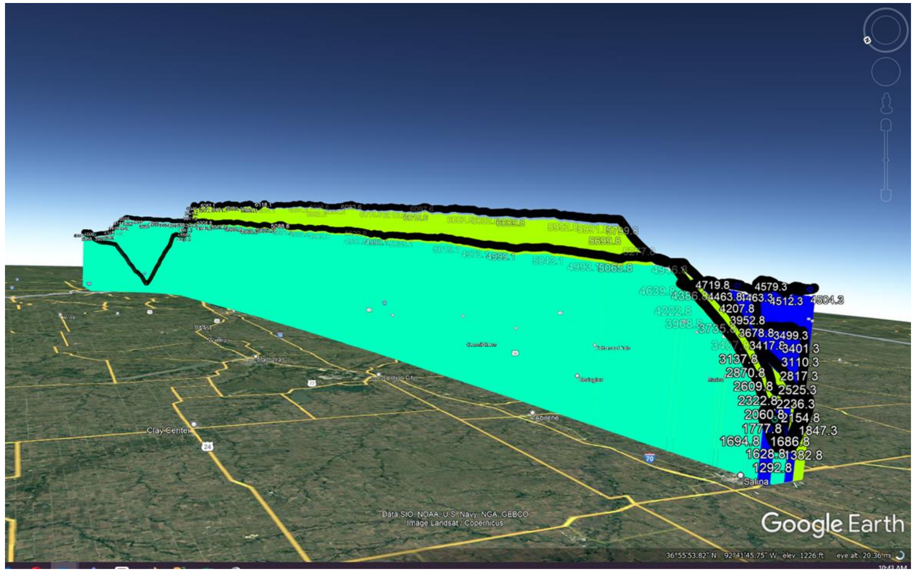
Analytical Framework



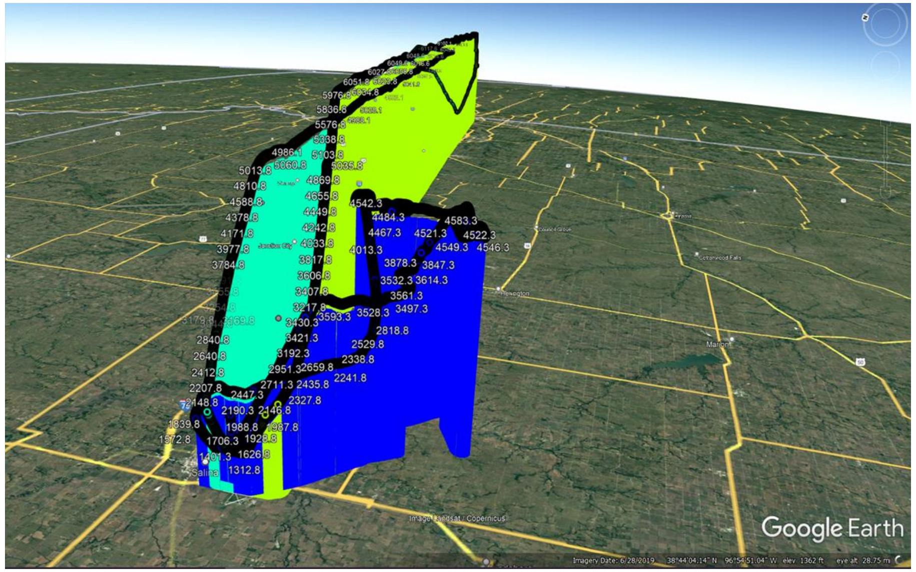
Python using Spyder, Integrated Development Environment (**S**cientific **PY**thon **D**evelopment **E**nvi**R**onment)

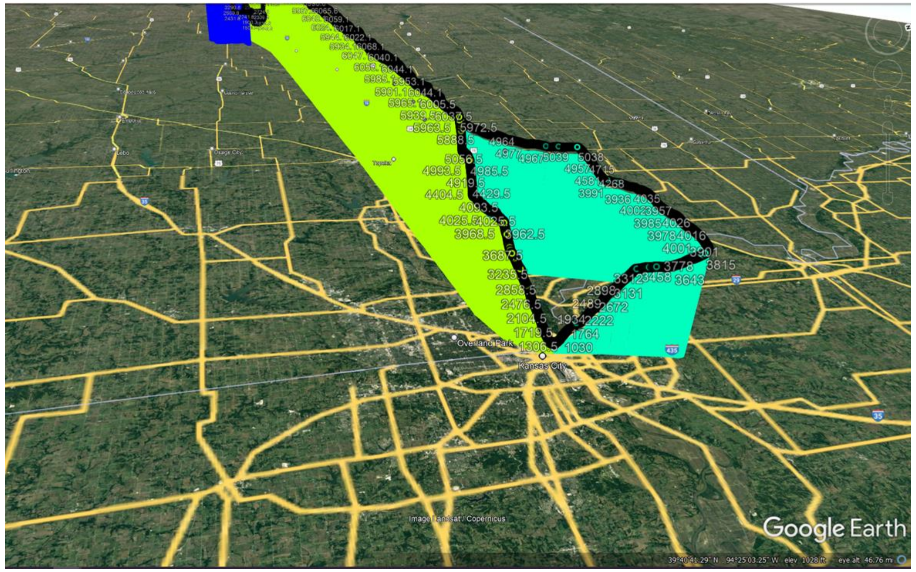
Flight Data Analysis: “*Location Data*” Verification





- **Local Training (Blue)**
- **Solo Flight (Cyan/Lime)**



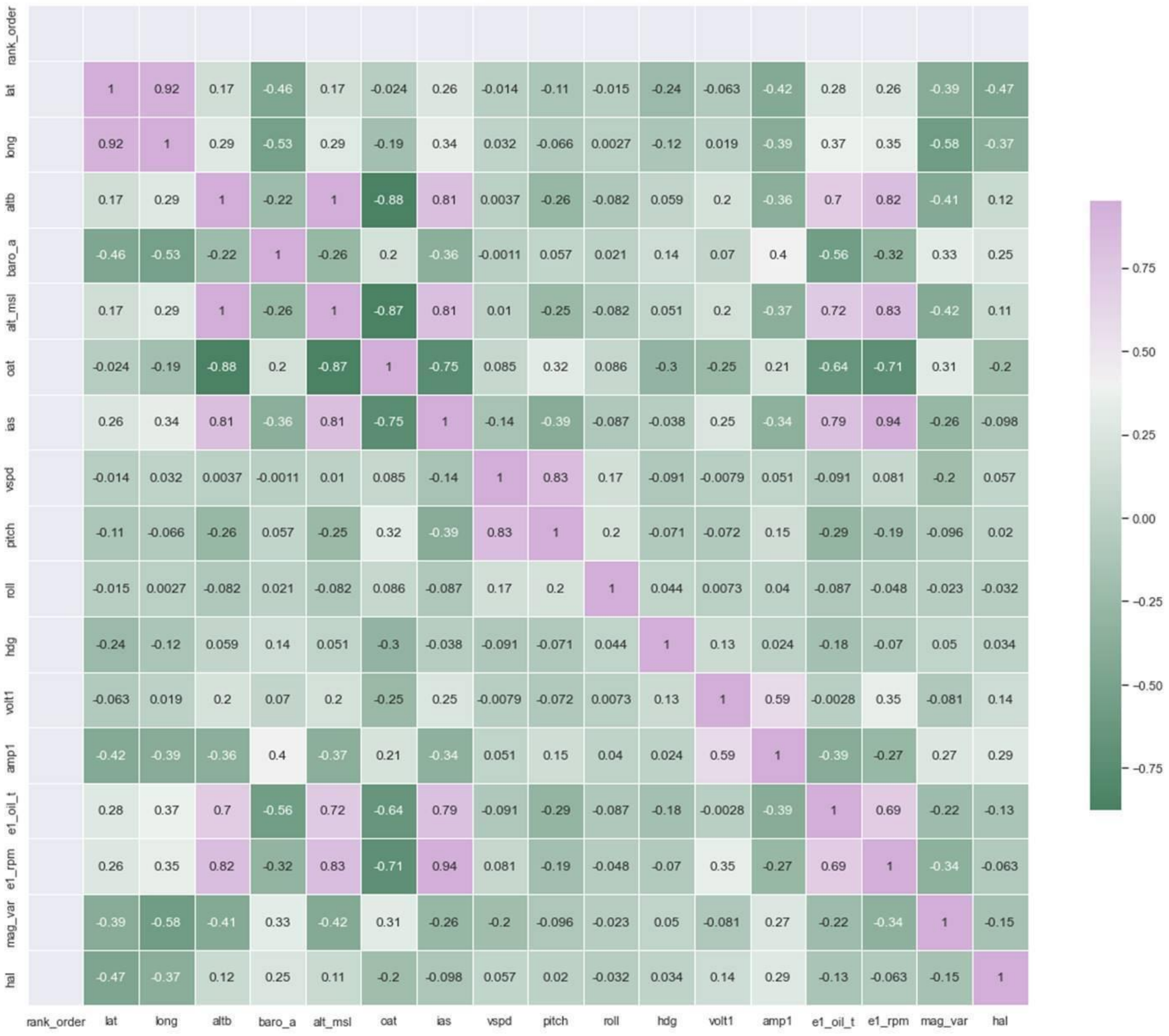


Flight Data Analysis: “*Ops Data*” Verification

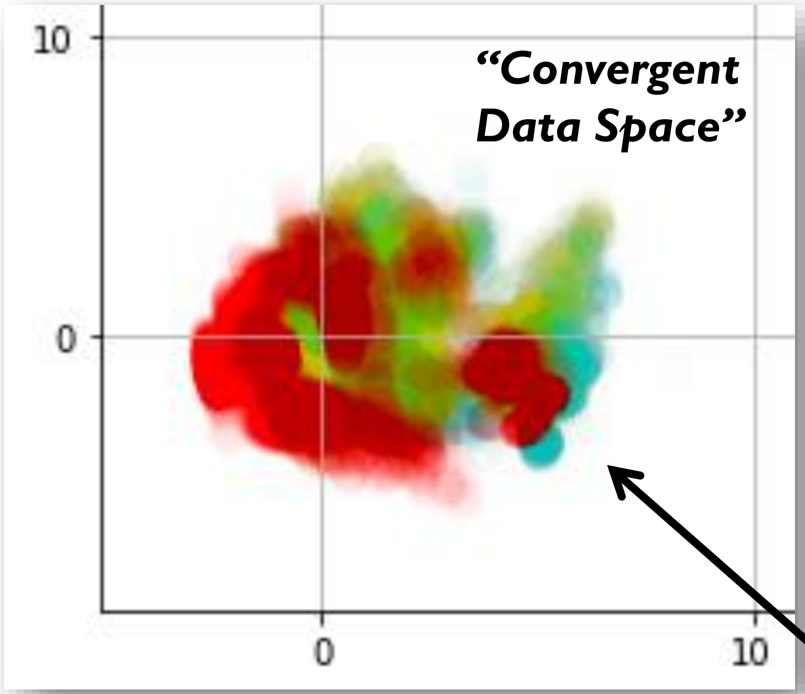
AltB Positive Correlation:
- E1_Oil_T
- E1_RPM
- IAS

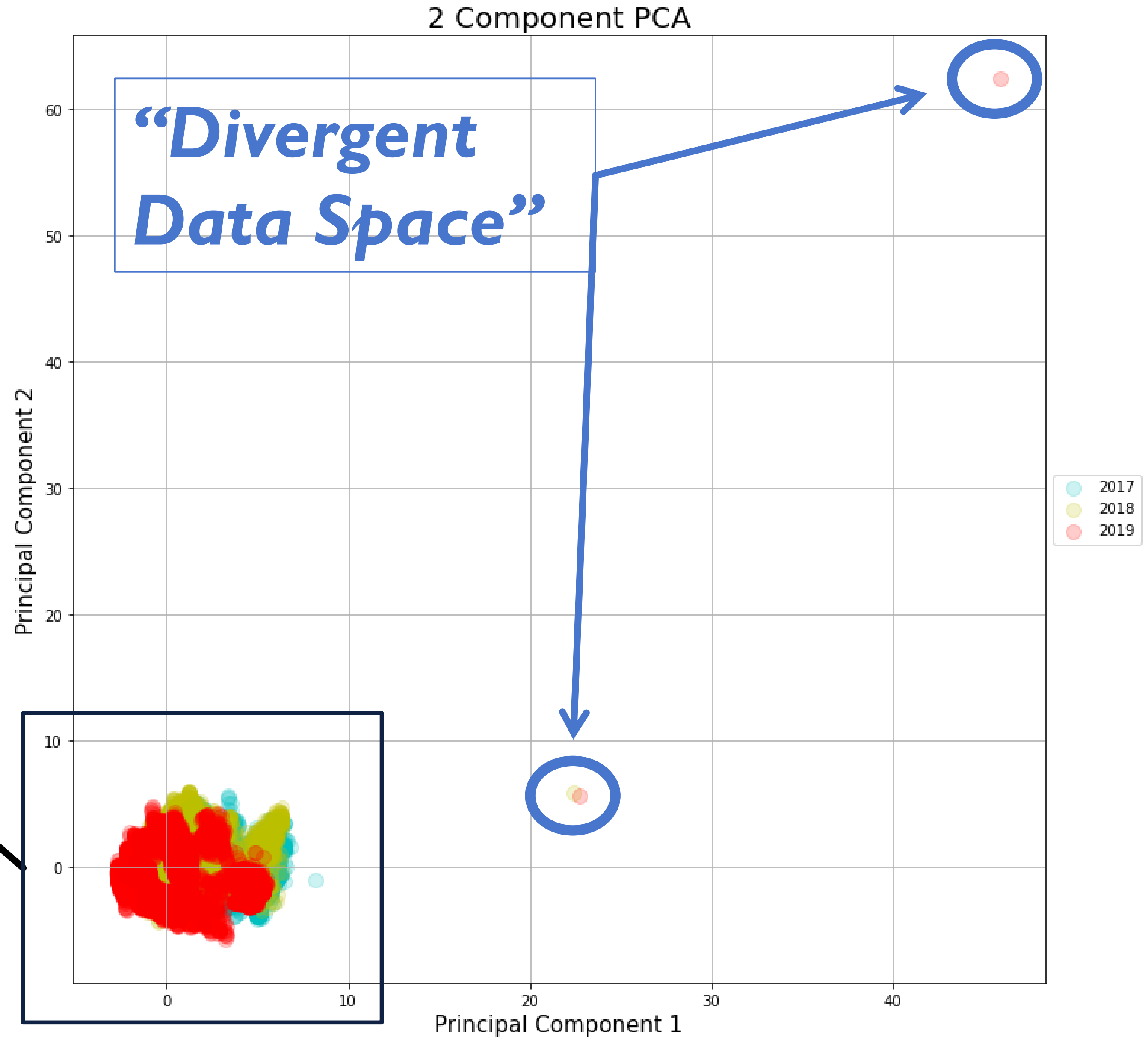
Alt_MSL Positive Correlation:
- E1_Oil_T
- E1_RPM
- IAS
- (Note: Alt_MSL is more strongly correlated than AltB)

IAS Positive Correlation:
- E1_Oil_T
- E1_RPM
- AltB



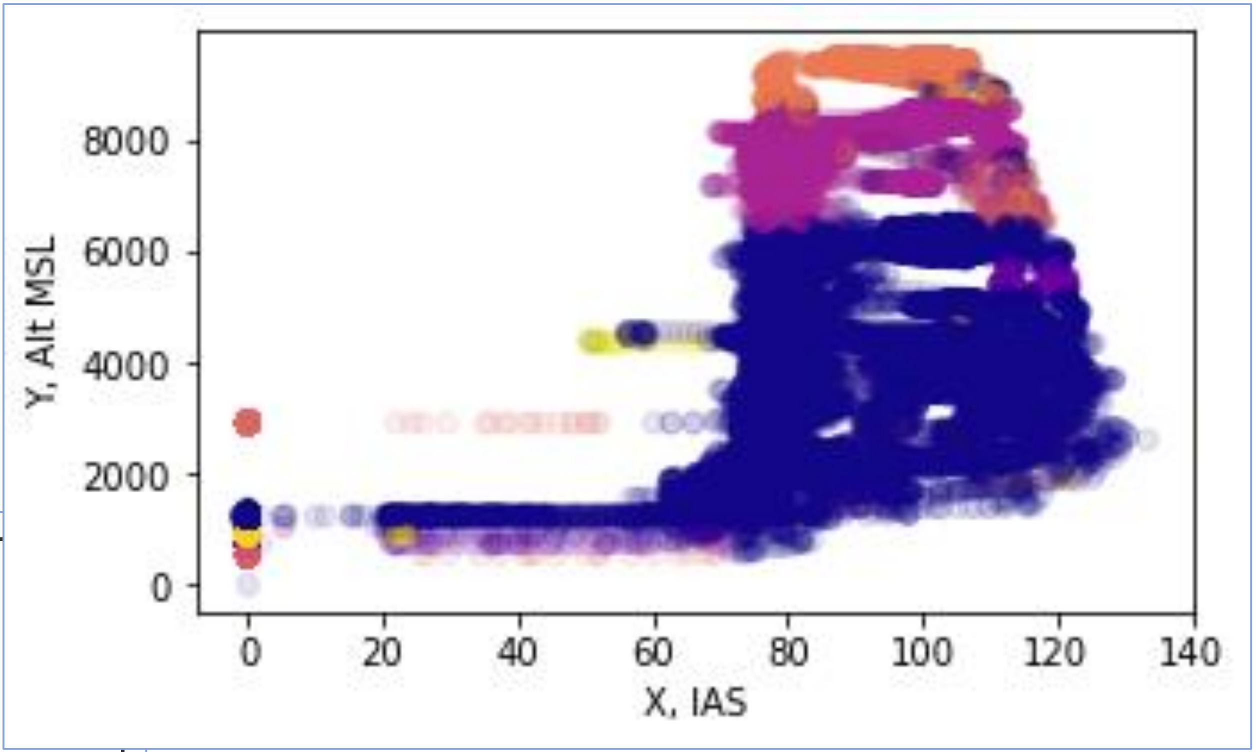
Principal Component Analysis (PCA)

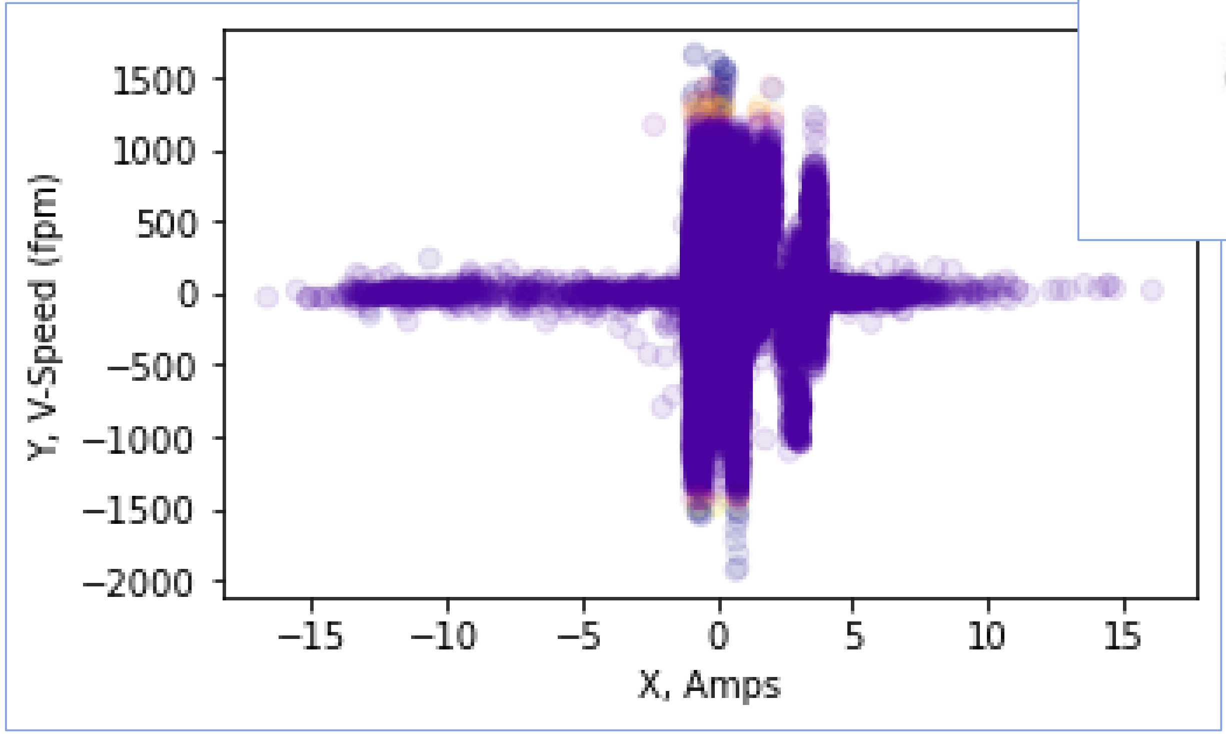




Macro-Level: PCA for Many Data Features at Once

Density-based Spectral Clustering (DBSCAN)

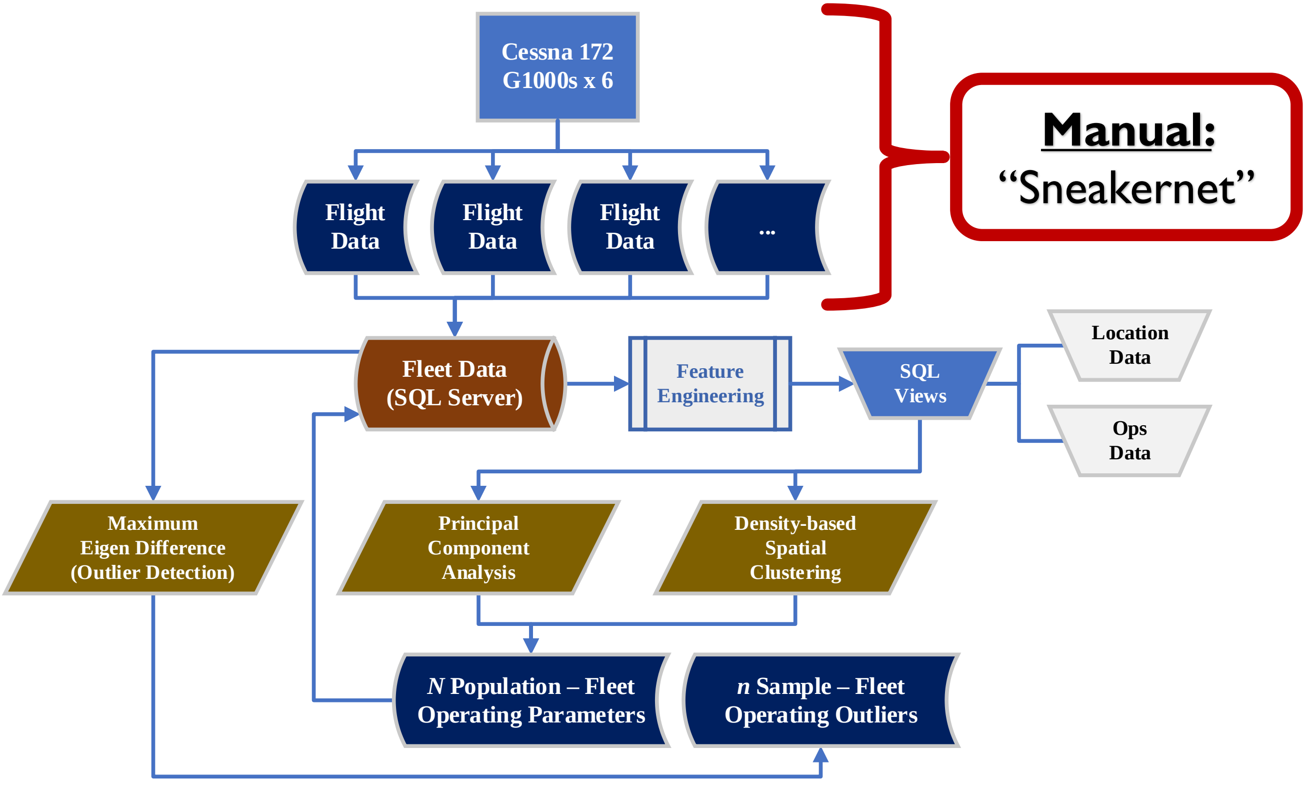




Micro-Level: DBSCAN for Isolated Use Cases

LIMITATIONS

- Operating Parameters (Limited to Cessna 172)
- Aircraft within a Training Environment
- Portions of Data Automation Systems are Manual (Sneakernet)
- Scalar Eigenvalue Integration



LESSONS LEARNED

- Use a prototype method, spread your initial data capture across a few aircraft
- Data feature engineering was the most time consuming (reworking and massaging the data)
- Leverage SQL Server Views for your Analytical Data Structure (i.e., SQL Server Views are Virtual Table structures)
- Null Logic Implementation (i.e., ISNULL() function required on all data features)
- Python made for an excellent backend system-level programing language
- PCA: High Level Feature Analysis | DBSCAN: Low Level Feature Analysis

CONCLUSIONS….AND…

- **Conclusions**
  - G1000 Flight Data is remarkably structured and easy to work with…
  - …organizational Machine Learning for Fleet Data is **Very Doable** ☺
  - **Machine Learning Model** development was easier than…
  - …automating the upfront **"sneakernet" system**.
  - For larger datasets consider using Indexed Views to improve analytical data performance.

….FUTURE DIRECTIONS

- **Areas for Future Research**
  - Design Science – Wireless transmission of fleet data via flight line access point system.
    - Future testing of onboard microcontroller/communication structure
    - Isolating the Scalar Eigenvalues at time interval level (not just at the day level)
  - Additional Categorical Data Capture
    - Maintenance Records
    - Type of Training Flight Performed (IFR, VFR, MVFR, LIFR, etc…)
  - SQL Server Engine Optimization
    - Support for Inline Data Predictions (i.e., native T-SQL)

THANK YOU!

Michael J. Pritchard, PHD
Associate Dean for Research & Graduate Studies
Assistant Professor, Machine Learning & Autonomous Systems
mjp001@ksu.edu | (785) 979-0706

SELECTED REFERENCES

- Aziz, N. (2012). Detection of Outliers in Multivariate Data: A Method Based on Influence Eigen. In *Proceedings of the World Congress on Engineering 2012* (Vol. 1).
- Cartocci, N., Costante, G., Napolitano, M. R., Valigi, P., Crocetti, F., & Fravolini, M. L. (2020, September). PCA methods and evidence based filtering for robust aircraft sensor fault diagnosis. In *2020 28th Mediterranean Conference on Control and Automation (MED)* (pp. 550-555). IEEE.
- Dufrenois, F., & Noyer, J. C. (2015, July). Generalized eigenvalue proximal support vector machines for outlier description. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-9). IEEE.
- Fala, N. (2019). Data-driven safety feedback as part of debrief for General Aviation pilots (Doctoral dissertation, Purdue University Graduate School).
- Jajo, N. K., & Matawie, K. M. (2009, July). Eigenvalues Application in Robust outlier Detection. In *24th International Workshop on Statistical Modelling* (p. 182).
- Li, L., Gariel, M., Hansman, R. J., & Palacios, R. (2011, October). Anomaly detection in onboard-recorded flight data using cluster analysis. In *2011 IEEE/AIAA 30th Digital Avionics Systems Conference* (pp. 4A4-1). IEEE.
- Puranik, T. G., & Mavris, D. N. (2018). Anomaly detection in general-aviation operations using energy metrics and flight-data records. *Journal of Aerospace Information Systems*, *15*(1), 22-36.
- Reddy, K. K., Sarkar, S., Venugopalan, V., & Giering, M. (2016). Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach. In *Annual Conference of the PHM Society* (Vol. 8, No. 1).
- Sheridan, K., Puranik, T. G., Mangortey, E., Pinon-Fischer, O. J., Kirby, M., & Mavris, D. N. (2020). An application of dbscan clustering for flight anomaly detection during the approach phase. In *AIAA Scitech 2020 Forum* (p. 1851).
- Smart, E., Brown, D., & Denman, J. (2012). A two-phase method of detecting abnormalities in aircraft flight data and ranking their impact on individual flights. *IEEE Transactions on Intelligent Transportation Systems*, *13*(3), 1253-1265.
- Wang, Y., Wang, X., & Wang, X. L. (2016, July). A spectral clustering based outlier detection technique. In *International Conference on Machine Learning and Data Mining in Pattern Recognition* (pp. 15-27). Springer, Cham.