SpaceX- Falcon 9 Advantage

Christopher Chapman

03/21/2023

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Methodology for Data Collection
- Methodology for Exploratory Data Analysis (EDA) in SQL and Visual Representations
- Results From Exploratory Data Analysis
    - Visual processing using matplotlib and seaborn
    - EDA from SQL queries
- Plotly Dashboard Presentation
- Folium Results Walkthrough
- Predictive Analysis and Model Results
- Conclusion

INTRODUCTION

- SpaceX has solidified itself as a power in the private sector for low earth orbit and other extraplanetary needs.
- By understanding their strategic advantage that has allowed them to rise to the top for government contracts SpaceY can become a better competitor and close the gap for the growing extraplanetary mission market.
  - The first major advantage of SpaceX when compared to competitors is the lower expense of their rockets due to innovation.
  - By better understanding the success rate of their innovations we can find open niches that SpaceY could occupy.

DATA COLLECTION METHODOLOGY

- Data was collected by webscraping SpaceX's wikipedia page as well as by utilizing the SpaceX REST API on mission success.
- Data was then processed to generate a high level understanding of the intricacies in the data and relevant areas to focus.
- By understanding that SpaceX's first major advantage is the cost of their launches and that this lower cost is associated with a recoverable module, success of recovery directly impacts cost to SpaceX.
- Machine Learning algorithms were then implemented to predict the successful recovery from test data.
    - These insights can then be used to find conditions unfavorable to the successful recovery of the target module.

EXPLORATORY DATA ANALYSIS METHODOLGY

- For this section trends are inspected through database queries and visual representations to explore trends in the following variables as they pertain to success of retrieval.
  - Payload Mass- Does payload mass affect success of retrieval and if so which masses give the worst chance of success.
  - Launch Site- Of the three launch sites used for Falcon 9 launches is there an advantage to any site?
  - Mission Orbit- Of the different target orbits are their different success rates for each one?
  - Landing Pad- Of the different pads used by SpaceX, is there an advantage and use cases that can be observed?
  - Success Over Time- How has SpaceX retrieval success changed over time?

Predictive Analysis Methodology

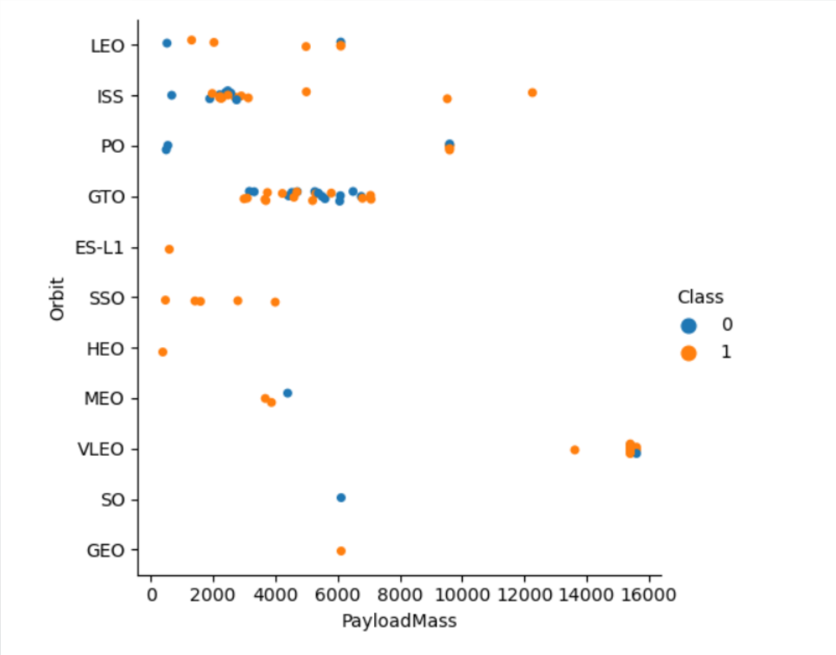
- After understanding the intricacies of the data gathered machine learning models were then used alongside gridsearch methods to find the ideal model for predicting outcomes based on the parameters provided.
- K-Nearest-Neighbors, Logistic Regression, SVM, and Decision Tree Classification were fitted using the training data following a 80/20 split. The models were then qualified using the test set of the data.
- These trained models can be used in the future to find specific launches where SpaceX does not have a comparative advantage due to the inability to recover the rocket module.

RESULTS

Exploratory Data Analysis - FINDINGS & IMPLICATIONS

Mission Orbit, Payload Mass and Outcome

- Different orbits have different success rates
  - Very low earth orbit has a high rate with the heaviest payloads
  - Sun-Synchronous Orbit has a 100% success rate
  - Geostationary Transfer Orbit has very mixed success as well across all weights
  - International Space Station Orbit has mixed success at low payload mass and higher success with heavier payloads.



Exploratory Data Analysis- SQL

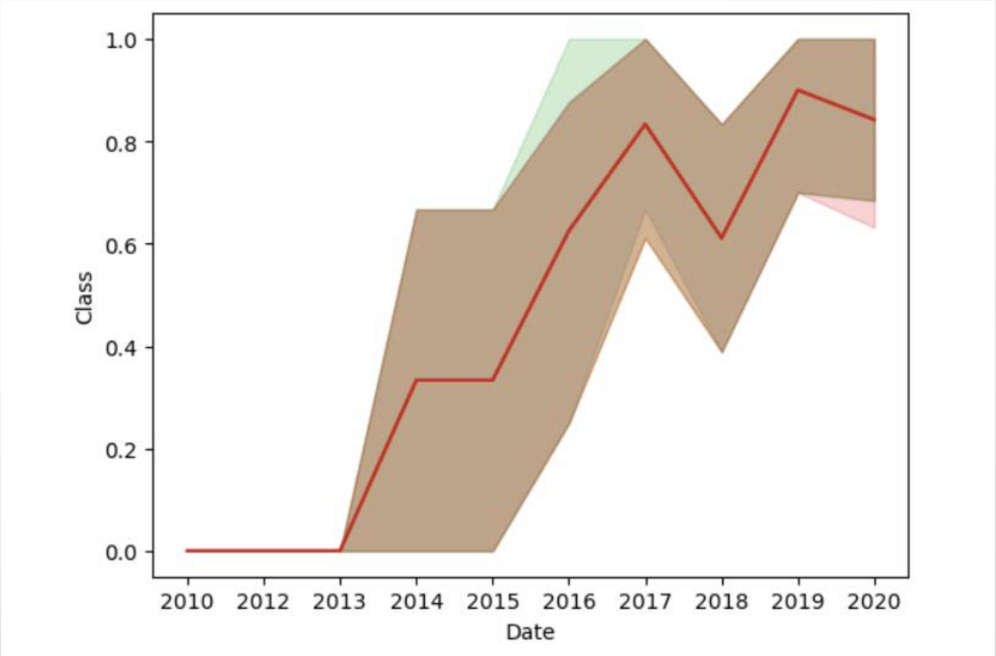
- Queries
  - Average payload mass from F9 Boosters – 2534
  - Success and failure mission counts – 99 Success 1 Failure
  - Booster versions carrying the largest payload
  - Top 10 outcomes- Shown right

From the queries the data can be processed and answers can be found for the general shape of data. From the final query, no attempt is shows as the most common outcome.
This shows that the data is going to be skewed more towards failure than would really be the case. From this we can decide how to handle cases where no attempt was made.

LANDING_OUTCOME	COUNTS
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

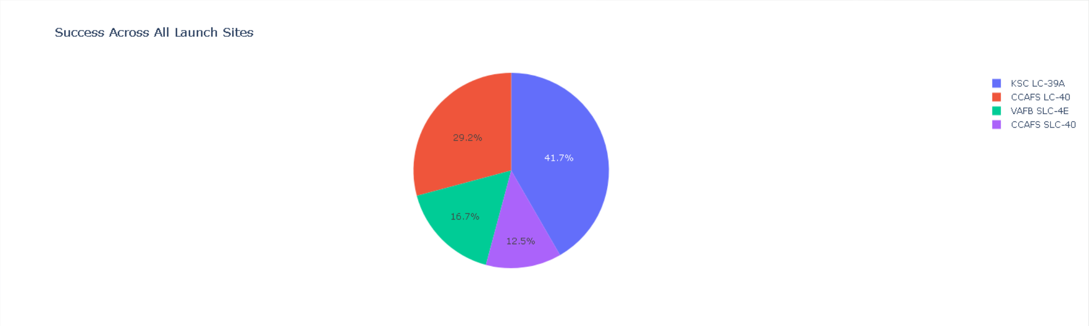
Exploratory Data Analysis - FINDINGS & IMPLICATIONS

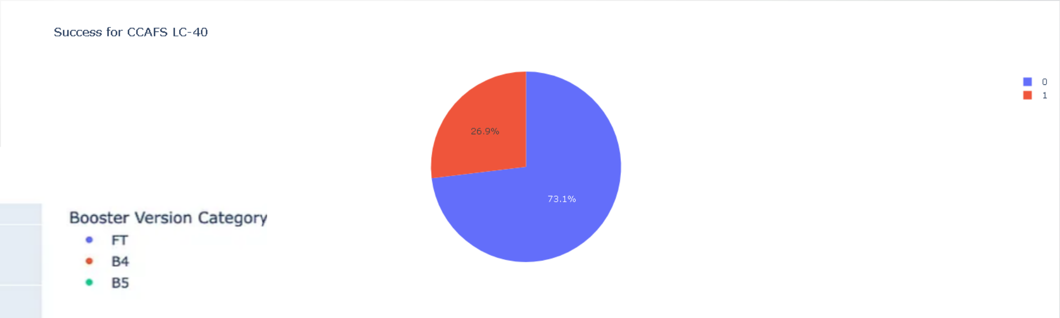
- Successful retrieval over time
  - This plot demonstrates the average success for a given year.
  - SpaceX is shown to greatly increase their retrieval success, closing in on a nearly 90% success rate. This shows that much of our data for failed retrievals comes from earlier launches.
  - This data also shows that SpaceX is implementing improvements to successfully lower launch costs making them more competitive.

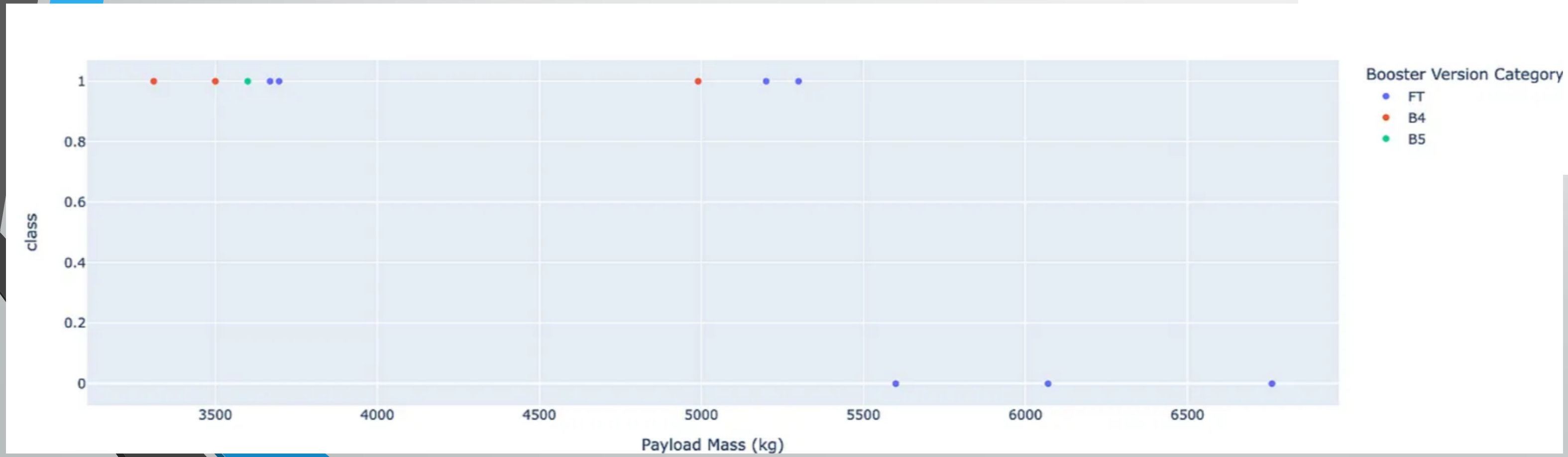


Plotly Dashboard

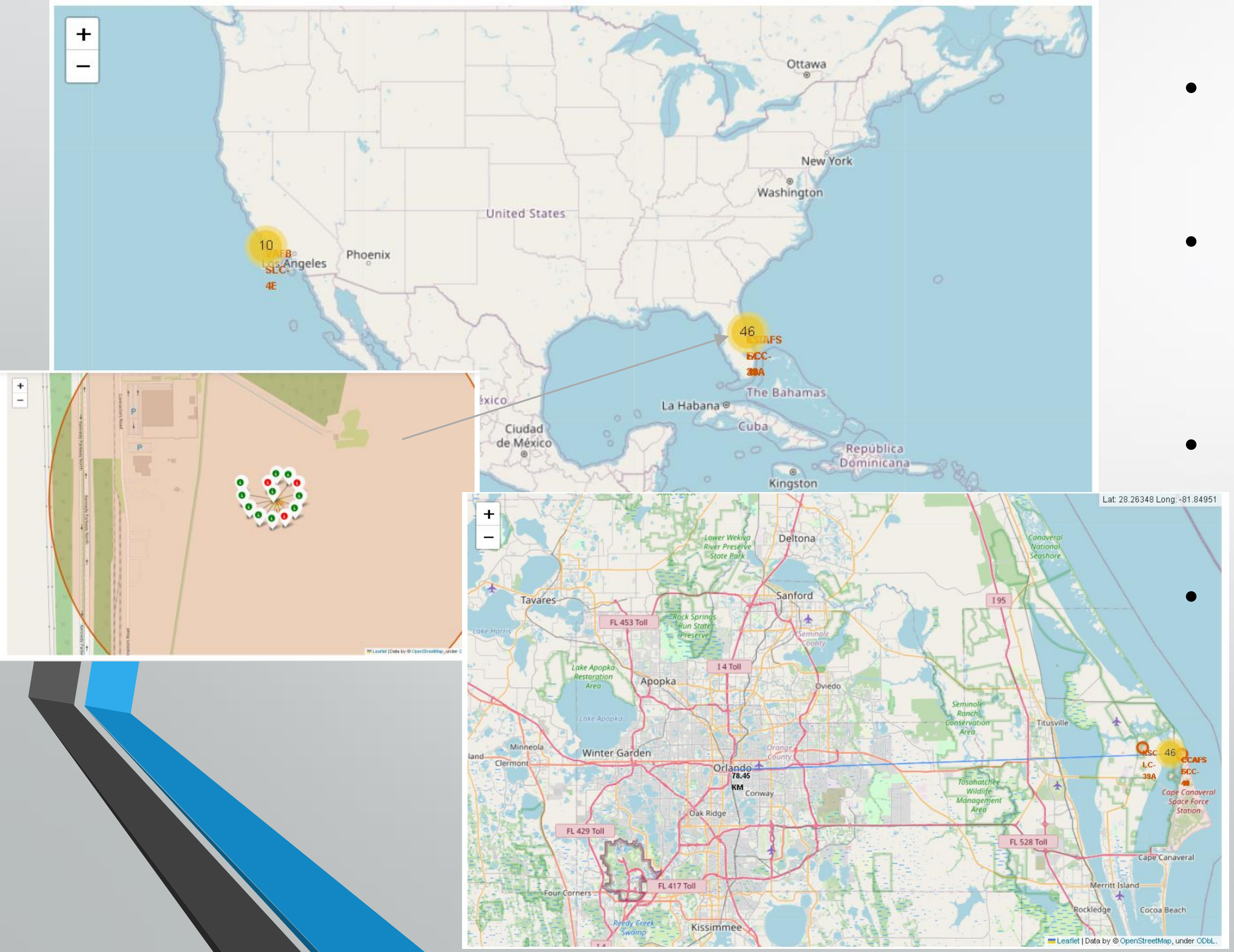
Plotly Dash used to create an interactive dashboard that allows for a presentable and quickly sortable view of the data.
From this dashboard quick insights can be gained as to each sites success as well as classification of success/failure when correlated to booster version and payload mass.





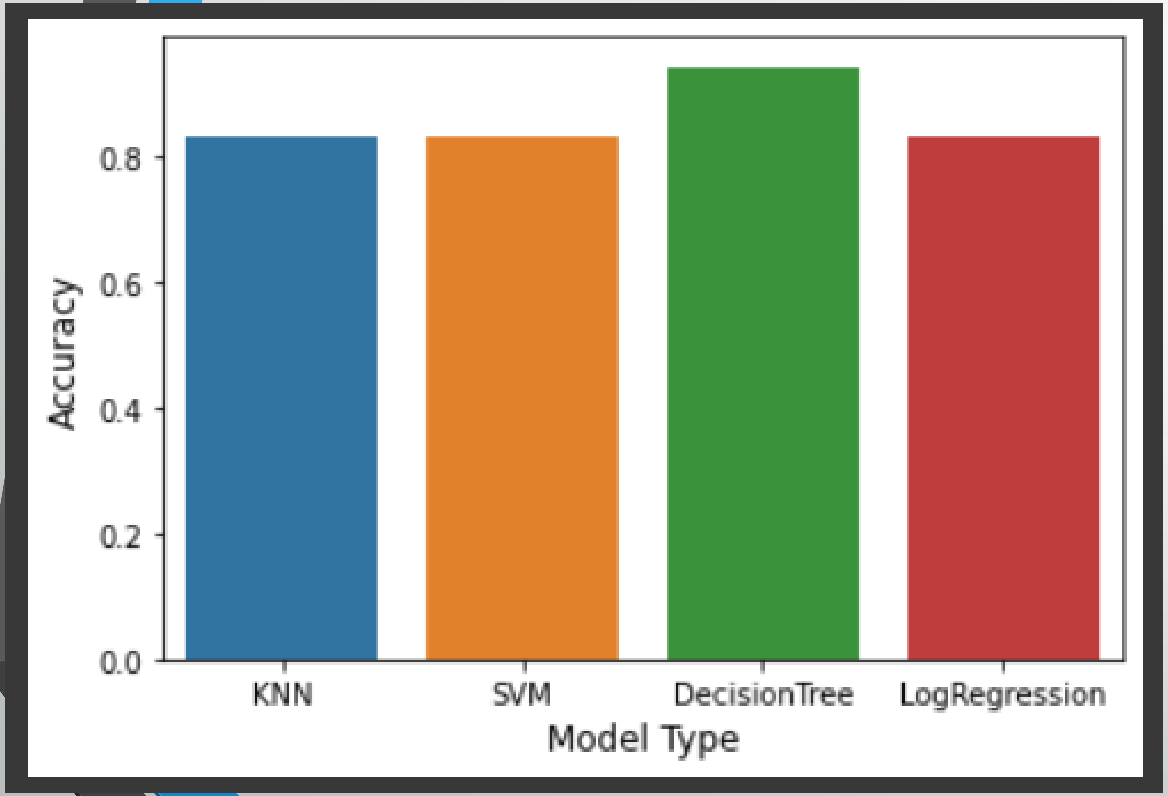


Folium Visualization of Sites

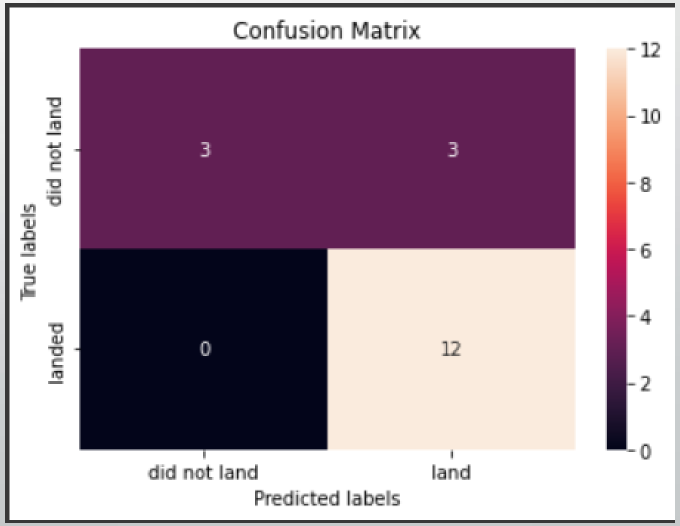


- Folium was implemented to visualize the launch site locations to better understand relation of launch site to surrounding areas.
- Marker clusters were added to allow each launch's class to be observed when zoomed in to give a visual representation of success and failure to retrieve the module at each site.
- Distance was measured to Orlando from the CCAFS site to show that the clustered sites are close to a major metropolitan area.
- All sites can also be observed to be directly linked to an ocean.

Predictive Analysis Results



- From the dataset models were fitted and scores which left the best performing model to be the decision tree classifier. Which had the confusion matrix provided.



CONCLUSION

- From the exploratory data analysis the data collected is shown to be a good representation of the history of SpaceX recovery but also skews the data to make the odds of a successful launch appear lower that are likely to be the current case.
- This flaw of the collected data will also have an impact on the models that were fitted with the data and could misdirect the decisions on an approachable niche for competition.
- From the findings, more data should be collected in the future and models should be fit with the most recent data to better inform management of the current outlook of SpaceX launches.

GitHub PROJECT LINK

https://github.com/cnoah713/Data-Science