Cheaper Space Missions

Arian Abdipour

https://github.com/Arian-Abdi

7/30/2023

Table of Contents

Executive Summary

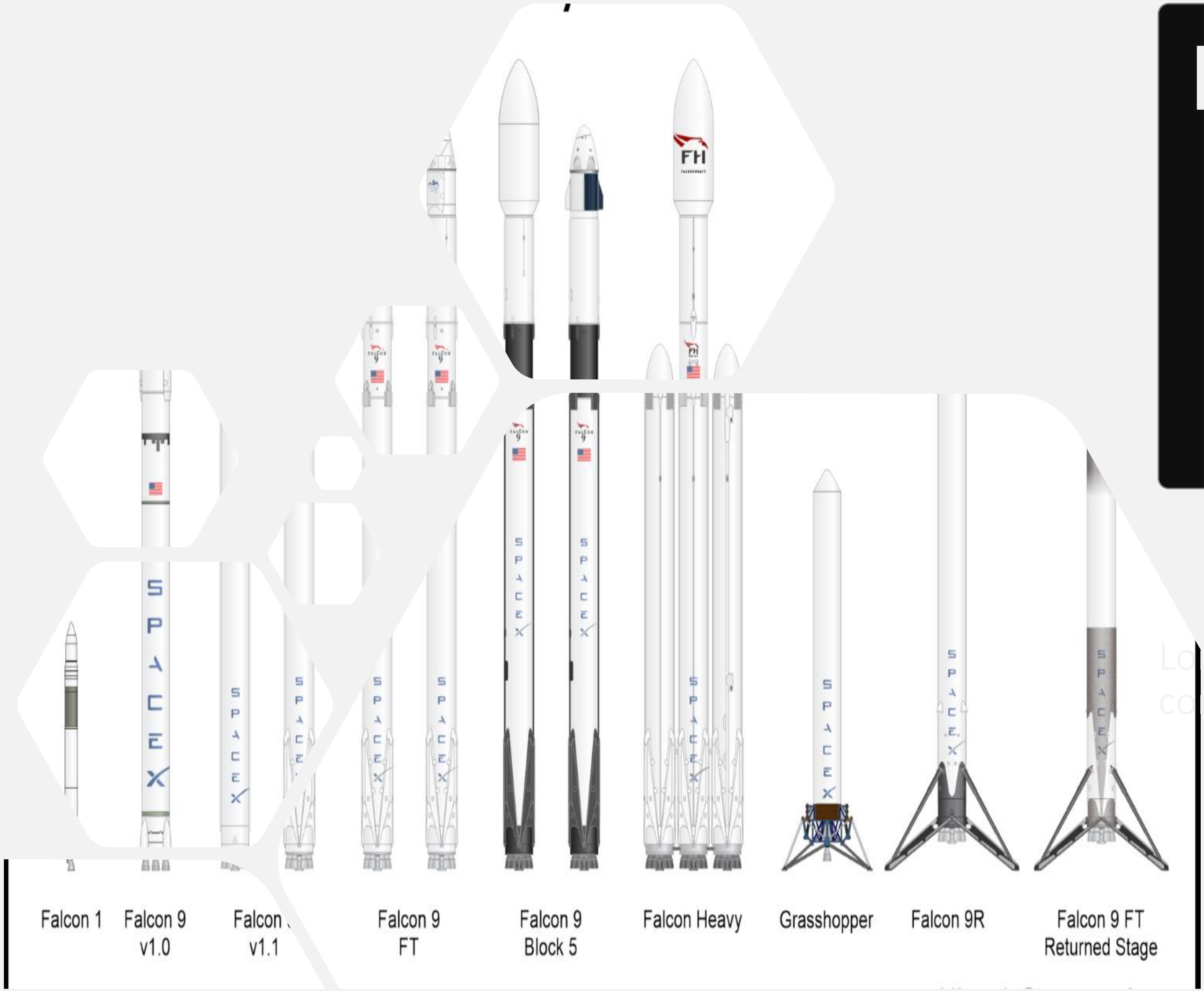
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- SpaceY is biggest competitor of SpaceX in Commercial Space Missions. We have been tasked by them to train a ML model that can predict reusability of stage 1 of their rockets.
- Data was collected through SpaceX API and by web scraping from Wikipedia; EDA(exploratory data analysis) including data wrangling, data visualization and interactive data analytics was performed on collected data.
- Through publicly available sources we were able to collect valuable data. EDA that was performed on collected data gave us insight into the best features needed to predict a successful landing. Finally the machine learning prediction showed the best model to predict important characteristics to achieve our goal.

Introduction

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore if SpaceY can determine if the first stage will land, they can determine the cost of a launch.

Methodology



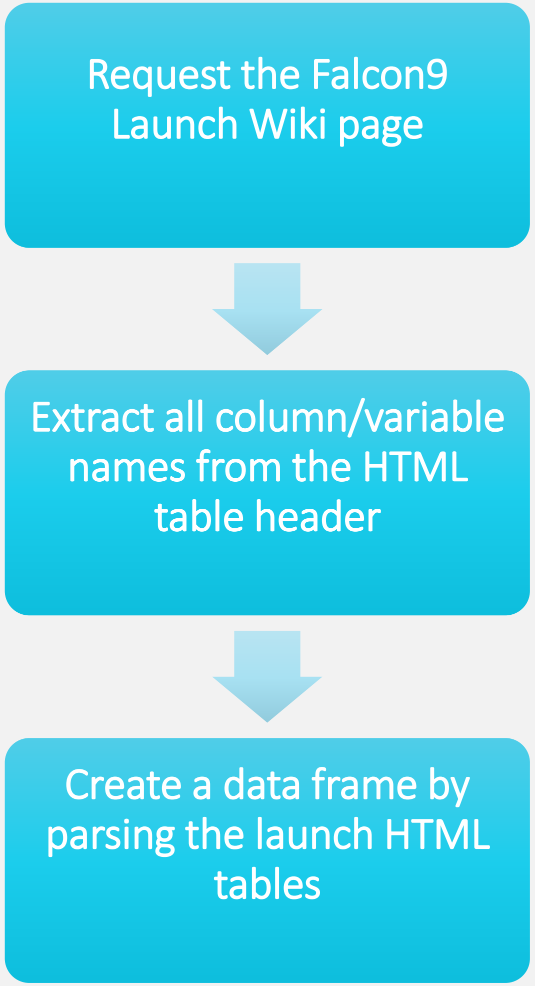
Methodology

- SpaceX offers a public API from where data can be obtained and then used.

- Source code: https://github.com/Arian-Abdi/applied-datascience-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Methodology

- Data from SpaceX launches can also be obtained from Wikipedia



- Source code: https://github.com/Arian-Abdi/applied-datascience-capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.

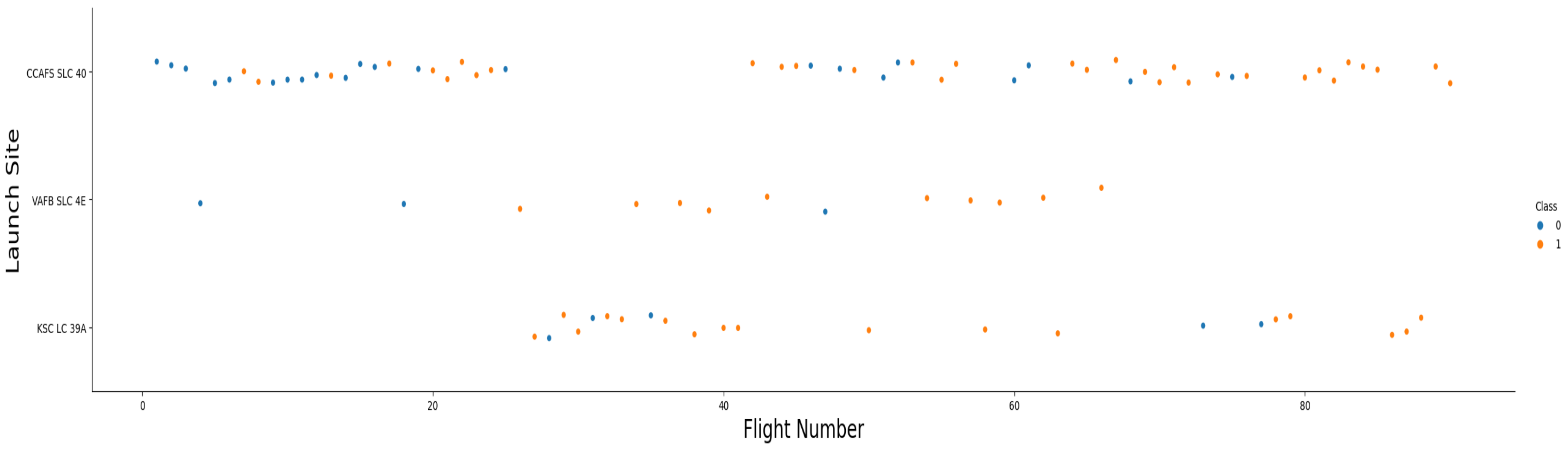
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.

- Finally, the landing outcome label was created from Outcome column.

Source Code: https://github.com/Arian-Abdi/applied-datascience-capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and bar plots were used to visualize the relationship between pair of features:
  - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass,
    Orbit and Flight Number, Payload and Orbit



Source Code: https://github.com/Arian-Abdi/applied-datascience-capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

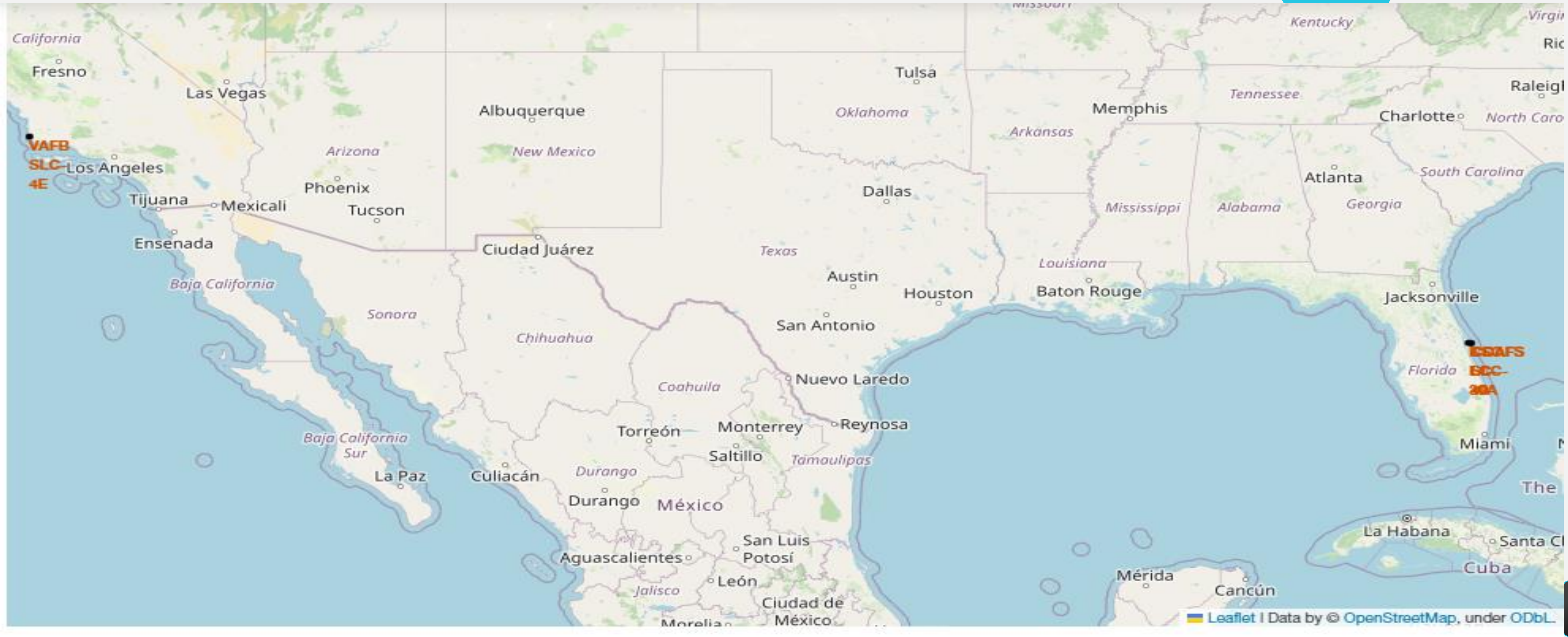
- The following SQL queries were performed:

  - Names of the unique launch sites in the space mission;
  - Top 5 launch sites whose name begin with the string 'CCA';
  - Total payload mass carried by boosters launched by NASA (CRS);
  - Average payload mass carried by booster version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  - Total number of successful and failure mission outcomes;
  - Names of the booster versions which have carried the maximum payload mass;
    Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
  - Rank of the count of landing outcomes (such as Failure)

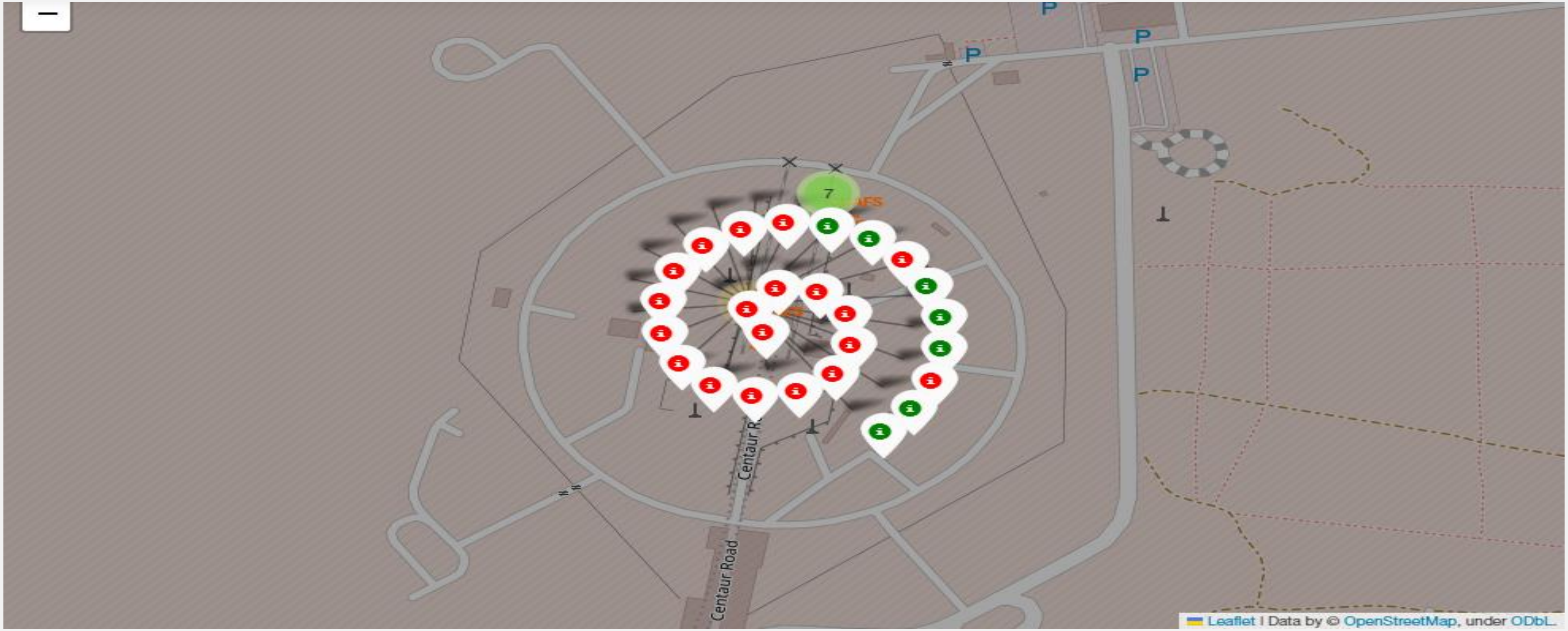
Source Code: https://github.com/Arian-Abdi/applied-datascience-capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Interactive map with Folium

- The map show all launch sites

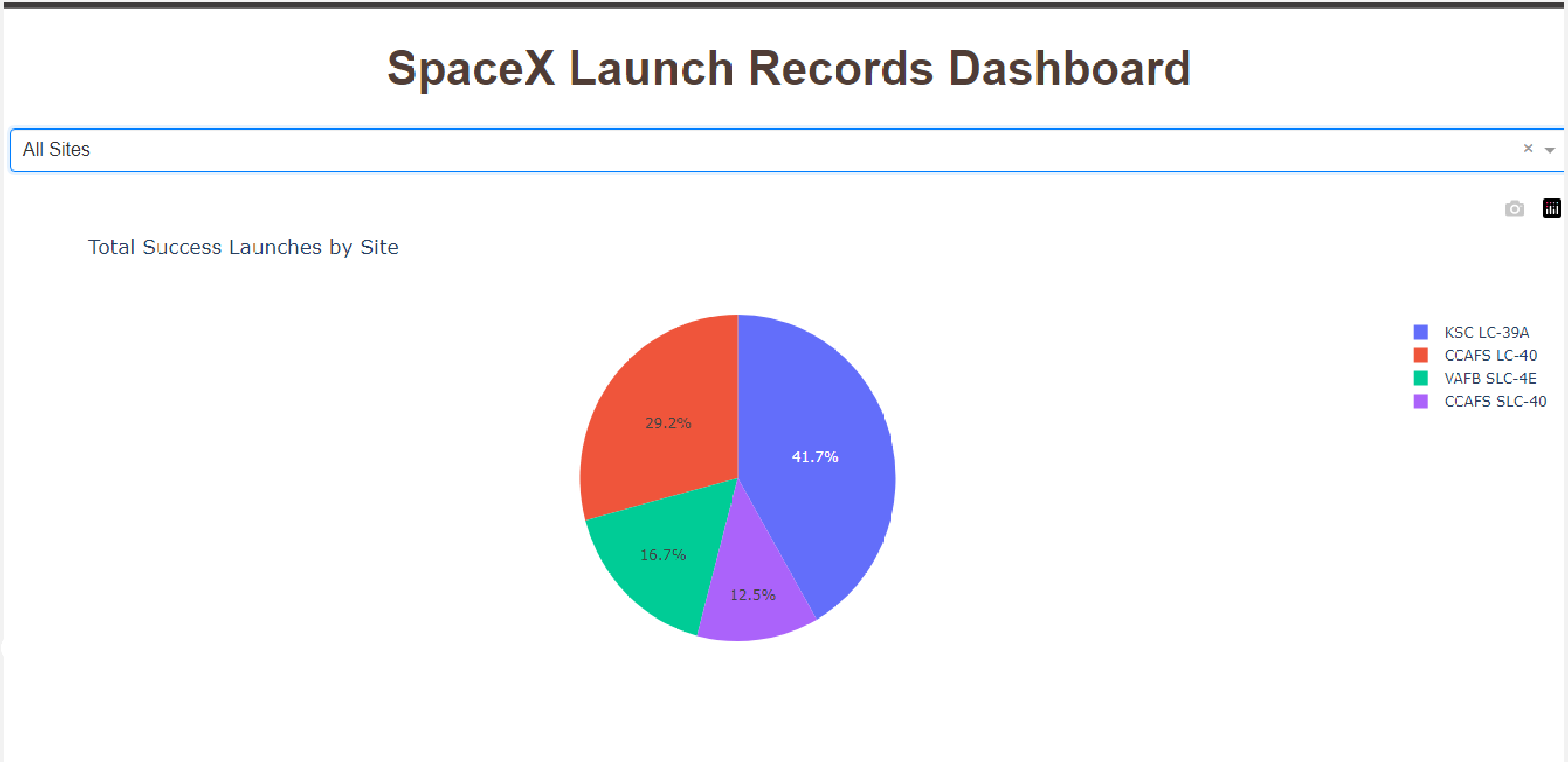


- Clusters on Folium map can be clicked on to display each successful landing (green icon) and failed landing (red icon).

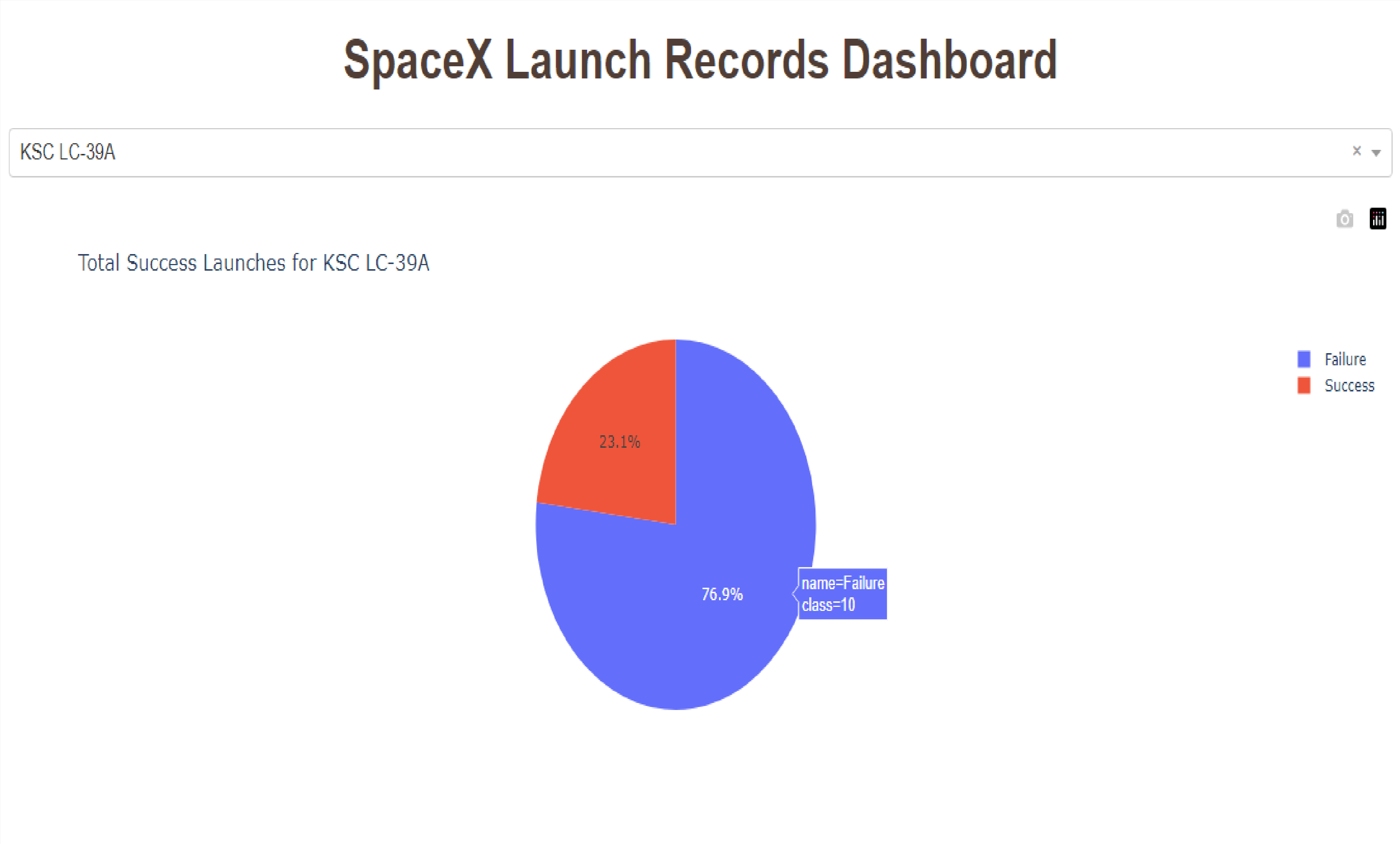


Dashboard with Plotly Dash

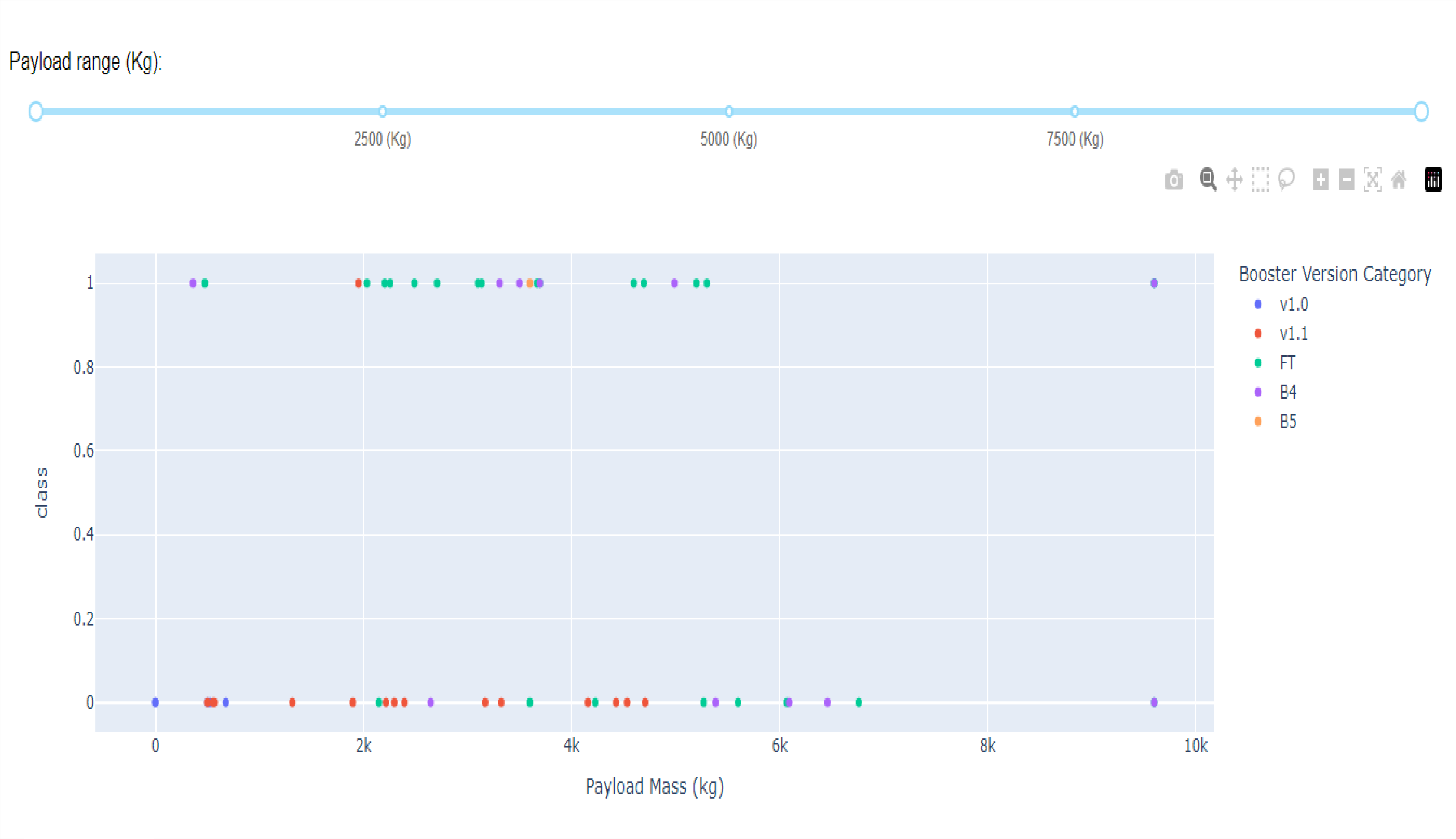
- The place from where launches are done seems to be a very important factor of success of missions.



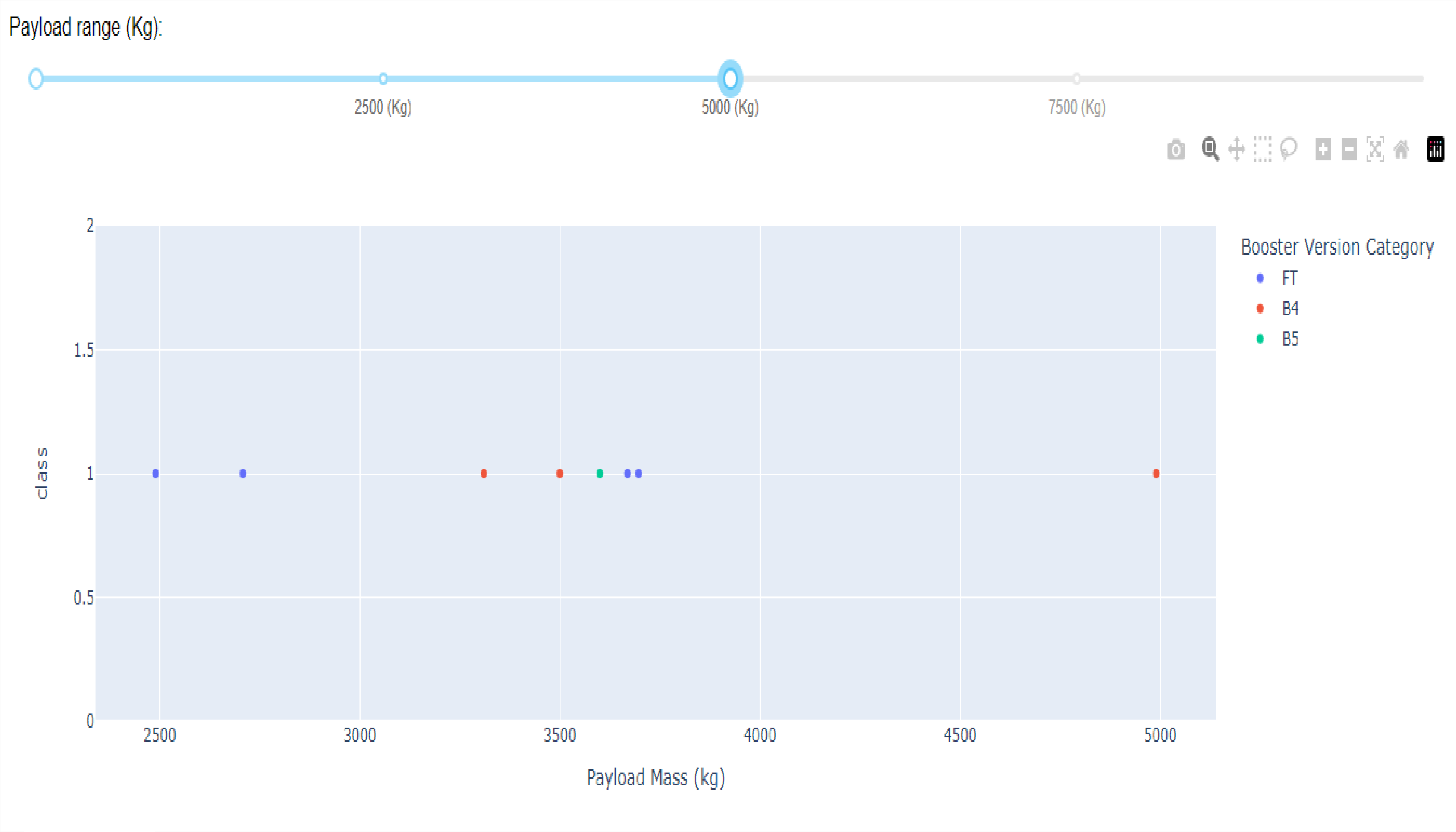
Launch Success Ratio for KSC LC-39A



Payload vs. Launch Outcome



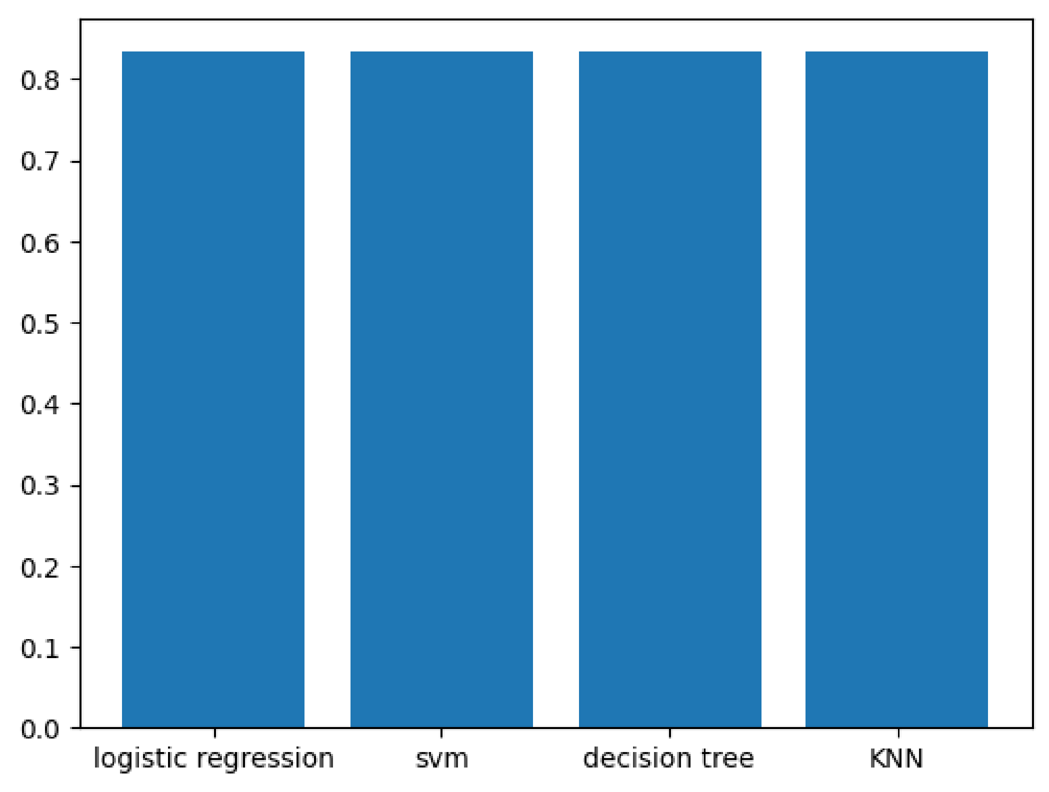
Payload vs. Launch Outcome



Predictive Analysis (Classification)

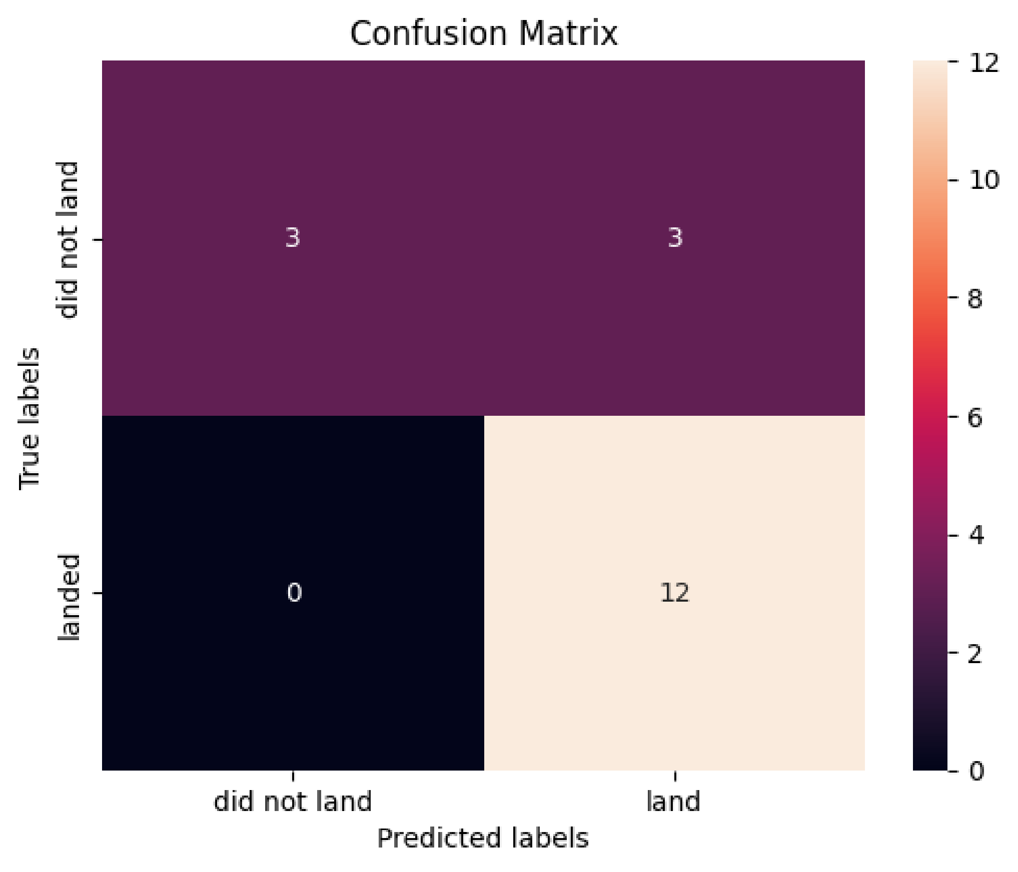
Classification Accuracy

- All models had the same accuracy on the test set at 83.33% accuracy. It should be noted that sample size was small and only 18.
- We are likely to need more data to determine the accuracy.



Confusion Matrix

- Since all models performed the same for the test set, the confusion matrix is the same across all models. The models predicted 12 successful landings when the true label was successful landing.
- The models predicted 3 unsuccessful landings when the true label was unsuccessful landing.
- The models predicted 3 successful landings when the true label was unsuccessful landings (false positives). Our models over predict successful landings.



Conclusion

Conclusion

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Many thanks to instructor team for their efforts.