IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Syed Zain Ul Haque

8/4/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
  - Data Collections
  - Data Wrangling
  - Exploratory Data Analysis EDA
  - Data Visualizations
  - Interactive Analytics
  - Machine Learning Predictions Classification
- Summary of all results
  - EDA results
  - Interactive Analytics
  - Classification Predictions

Introduction

- Objectives
  - To determine the cost of a launch and whether will it be able to land successfully. So, that an alternate company can ably bid against SpaceX for a rocket launch.
- Outcomes:
  - Determine the launch locations for the rocket.
  - Determine the optimal payload mass for the rocket.
  - Determine the orbit type for the rocket.
  - Determine the landing type for the rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The SpaceX data collected from two sources:
    - From SpaceX API
    - From Web Scraping
- Perform data wrangling
  - Data is cleaned and categorized the landing outcome based on the outcomes of the landing of SpaceX rockets.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

- Perform predictive analysis using classification models
  - Data was normalized, and then Data was split for testing and training for the classification models using Scikit Learn. The 20% was allocated for the test. Training and Testing data were used to evaluate the accuracy of different models. The model was tuned with a combination of different parameters.

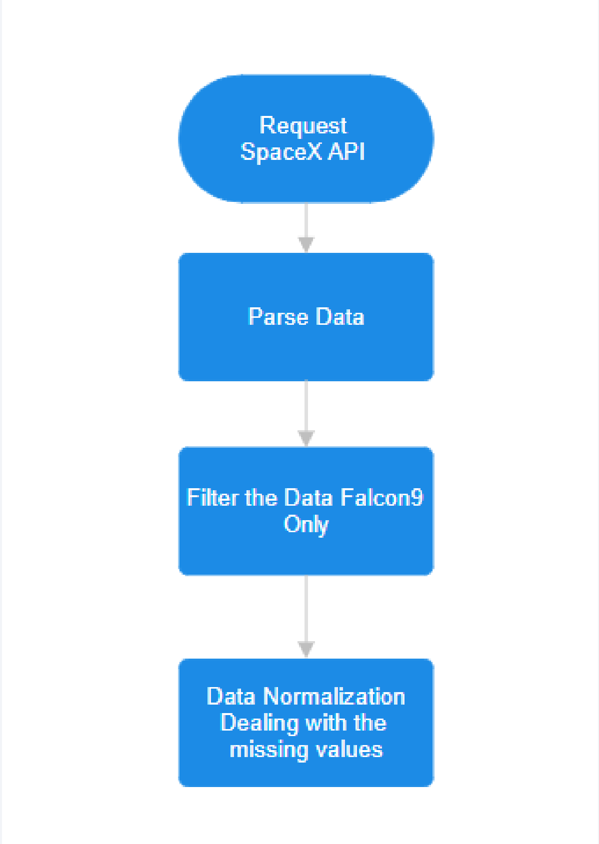
Data Collection

- Data sets were collection from two sources.
  - SpaceX API: https://api.spacexdata.com/v4/rockets/
  - Wikipedia: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

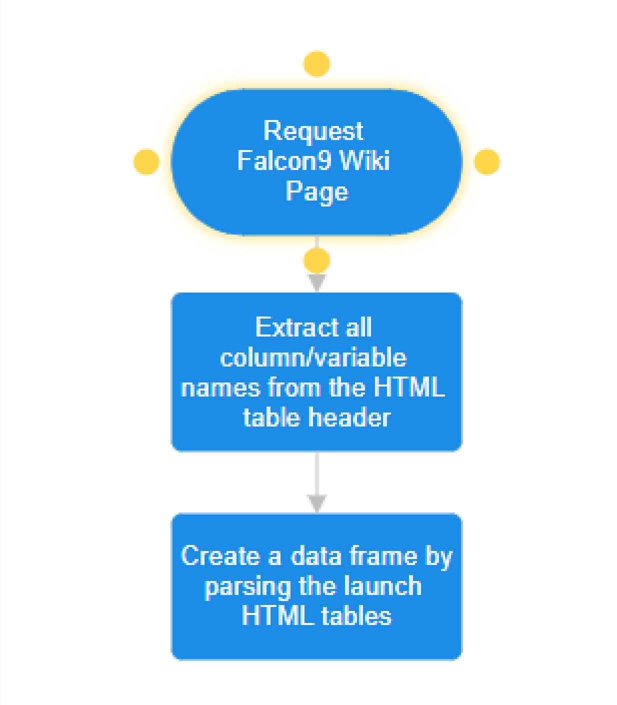
- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/364c94cfb3532269a02a0e6e4c55719af7610b7e/Module%201/jupyter_labs_spacex_data_collection_api.ipynb



Data Collection - Scraping

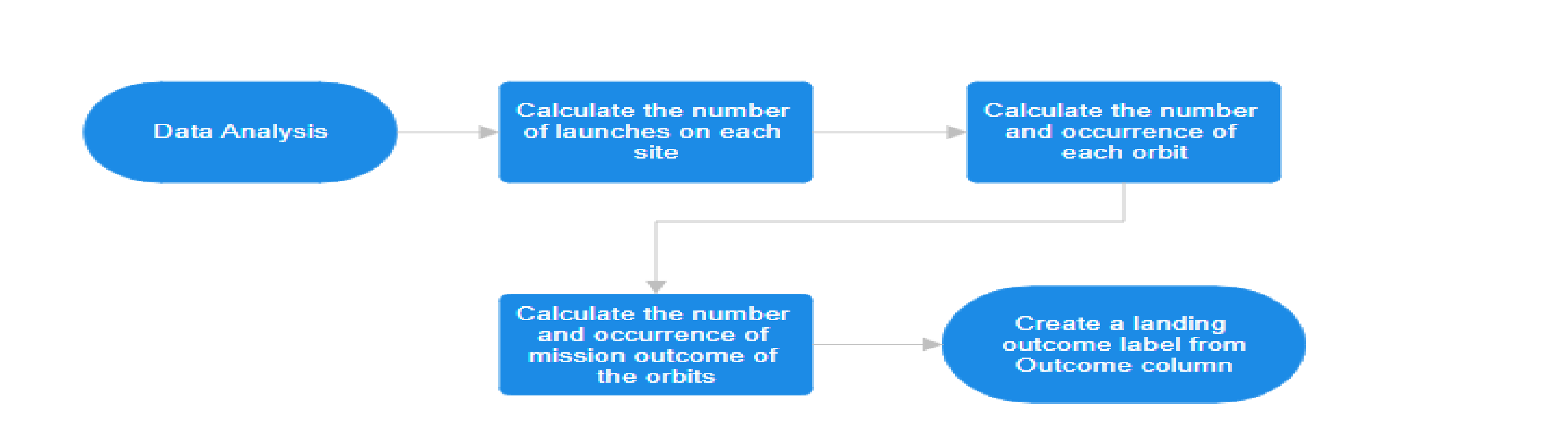
- Present your web scraping process using key phrases and flowcharts

- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%201/jupyter_labs_webscraping.ipynb



Data Wrangling

- Processing Steps:
  - First, we calculate the number of launches of each site, orbits, and their outcomes.
  - Create a landing outcome label from the Outcomes.



- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%201/labs_jupyter_spacex_Data_wrangling.ipynb

EDA with Data Visualization

- Scatter plot, bar plot, line plot were used.
    - Flight Number Vs Payload Mass, Flight Number Vs Launch Site, Payload Mass Vs Launch Site, Orbit Type Vs Landing Outcome, Flight Number Vs Orbit Type, Payload Mass Vs Orbit Type
- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%202/Jupyter_Labs_EDA_DataVIZ.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA’
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%202/Jupyter_Labs_EDA_SQL_Coursera_SQLLite.ipynb

Build an Interactive Map with Folium

- Markers: All launch sites on a map.
- Circles: The success/failed launches for each site on the map.
- Lines: The distances between a launch site to its proximities

- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%203/Lab_Jupyter_Launch_Site_Location.ipynb

Build a Dashboard with Plotly Dash

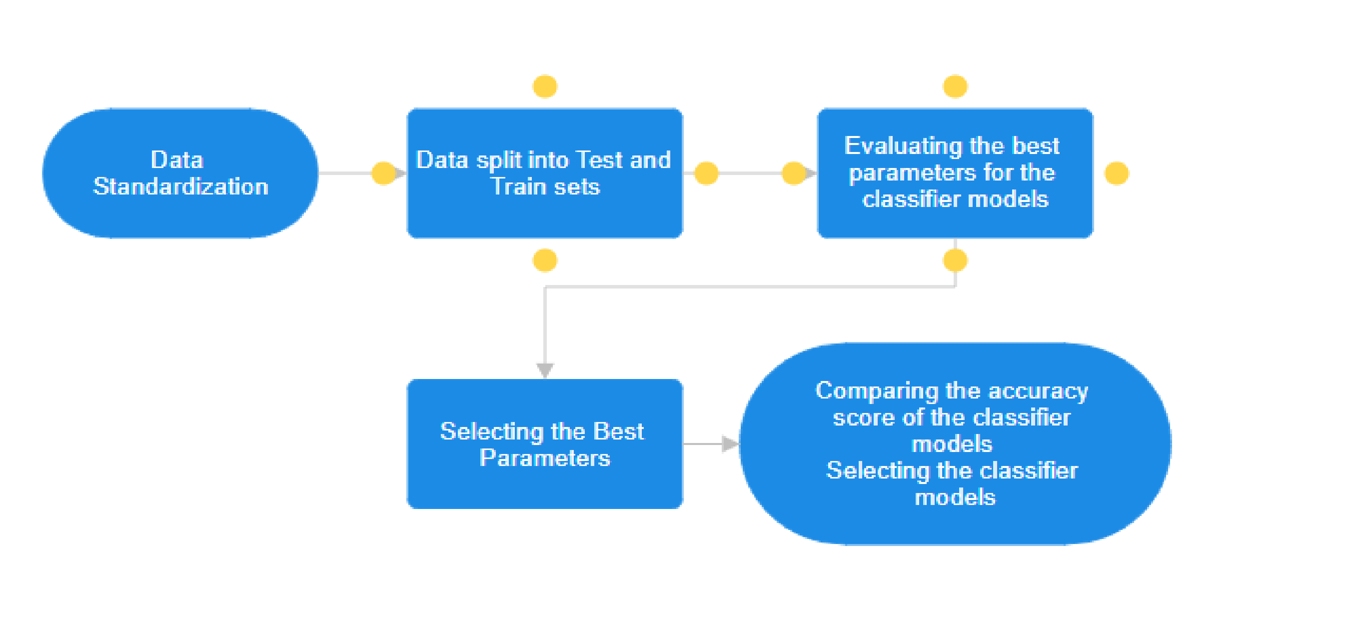
- Two plots/graphs were used:
  - Pie Chart
  - Scatter plot
- Two interactions were used:
  - Dropdown menu
  - Payload range

- It were used in combination to quickly visualized and interactive with the data.

- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%203/spacex_dash_app.py

Predictive Analysis (Classification)

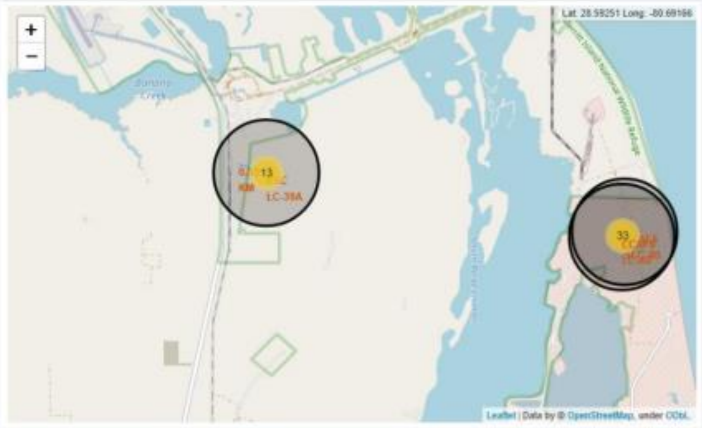
- Four classifier models were used:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree Classifier
  - k-nearest neighbors



- Source: https://github.com/Syed-Zain-Ul-Haque/Applied-Data-Science-Capstone/blob/1741df2ef1a9e0644b221ea67d60591e3cddcd03/Module%204/SpaceX_Machine_Learning_Prediction_Part_5_jupyterlite.ipynb

Results

- Exploratory data analysis results
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.
- Interactive analytics demo in screenshots
  - Launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around

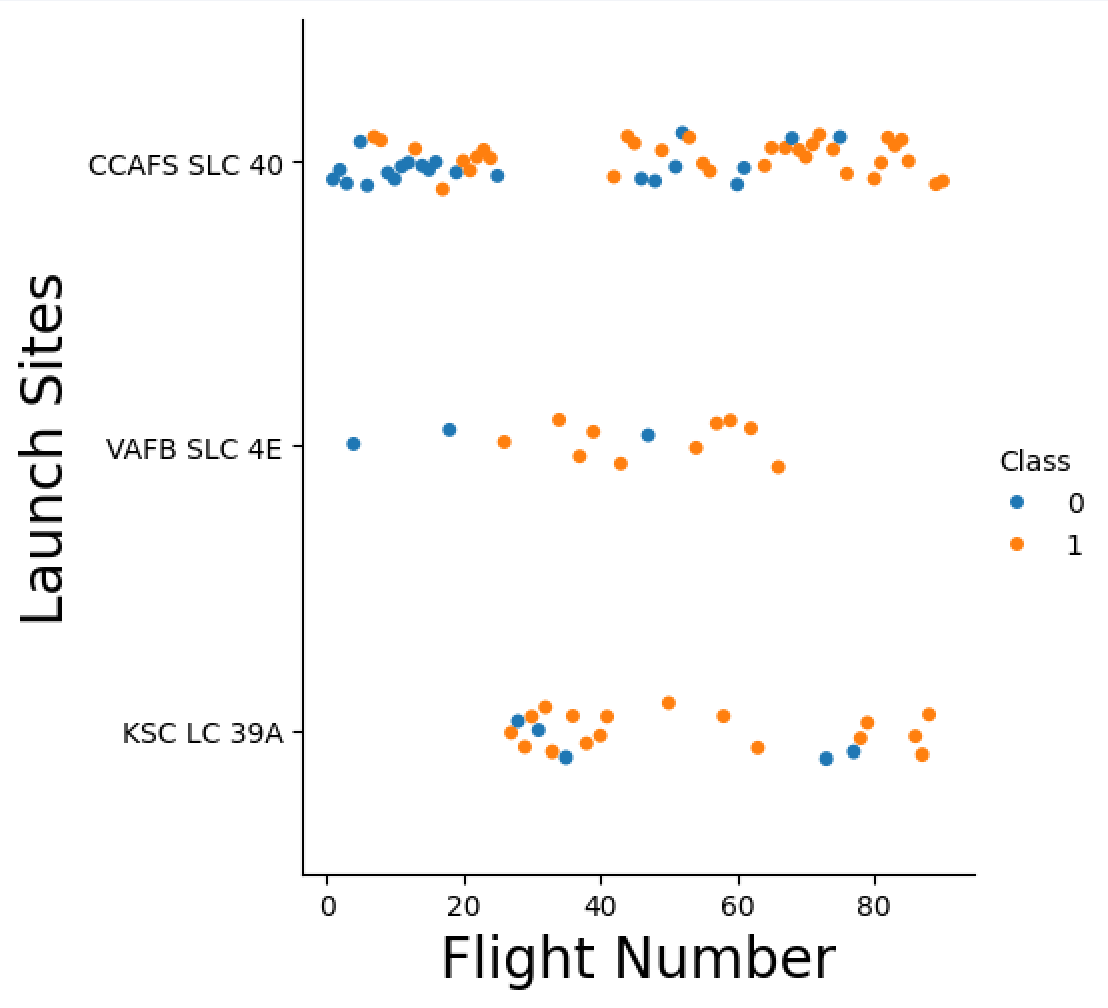


- Predictive analysis results
  - Accuracy for Logistic Regression: :83.33
  - Accuracy for Support Vector Machine: 83.33
  - Accuracy for Decision Tree: 83.33
  - Accuracy for K Nearest Neighbors: 83.33
  - All the four method have the same accuracy.

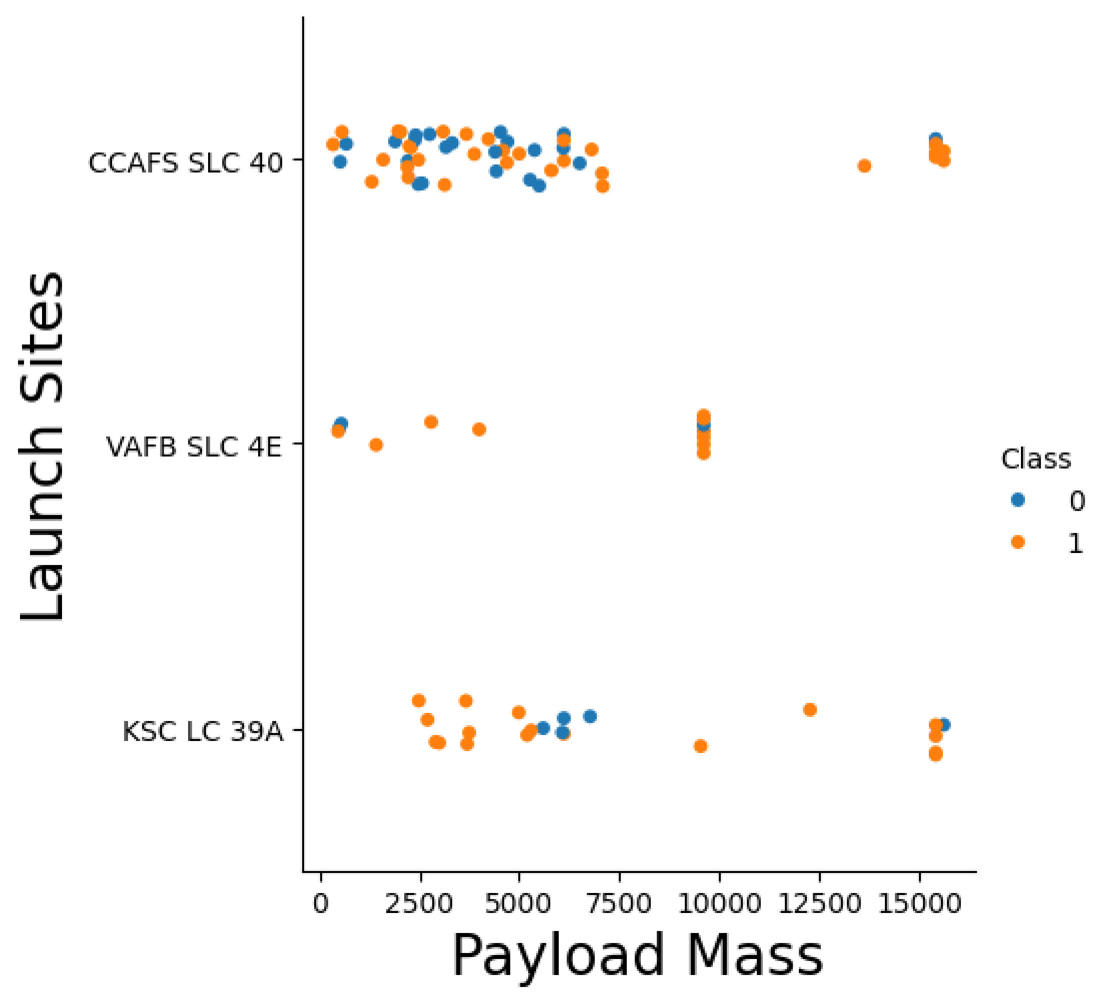
Section 2

Insights drawn from EDA

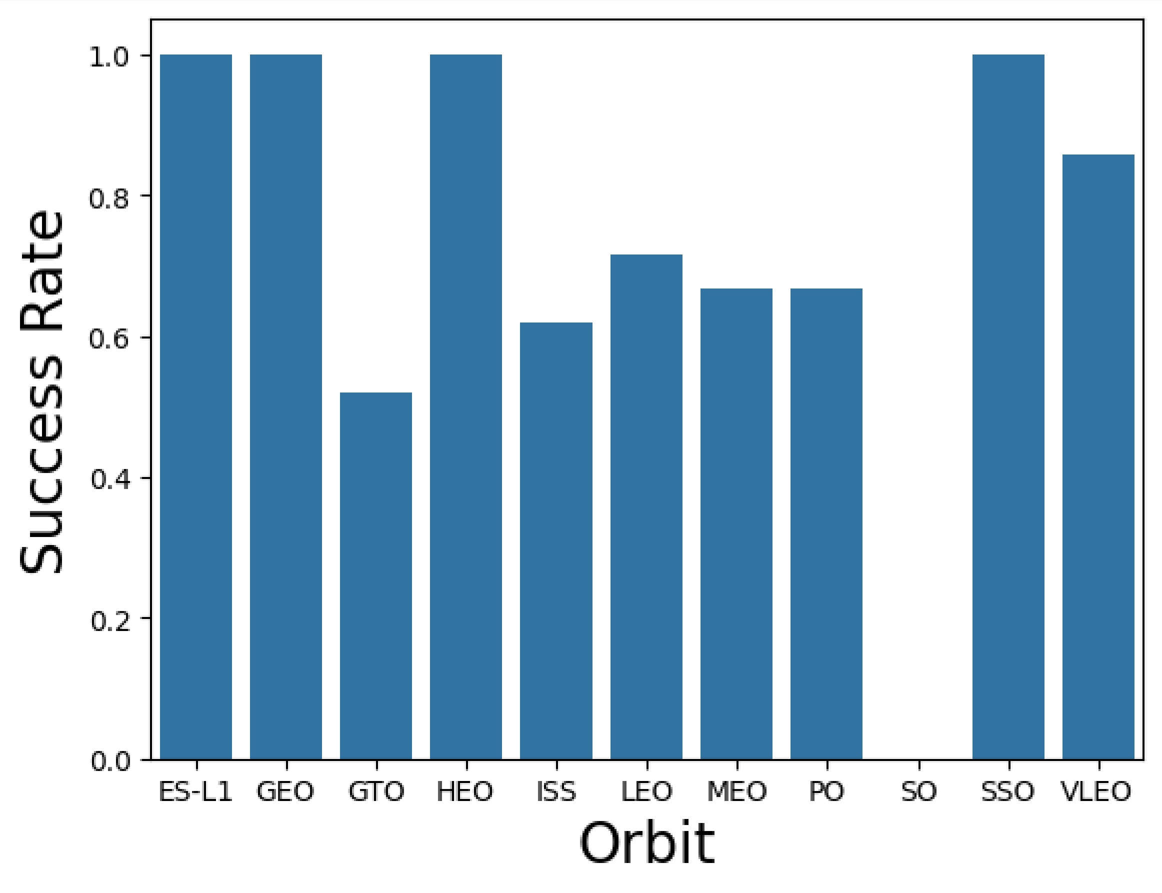
Flight Number vs. Launch Site



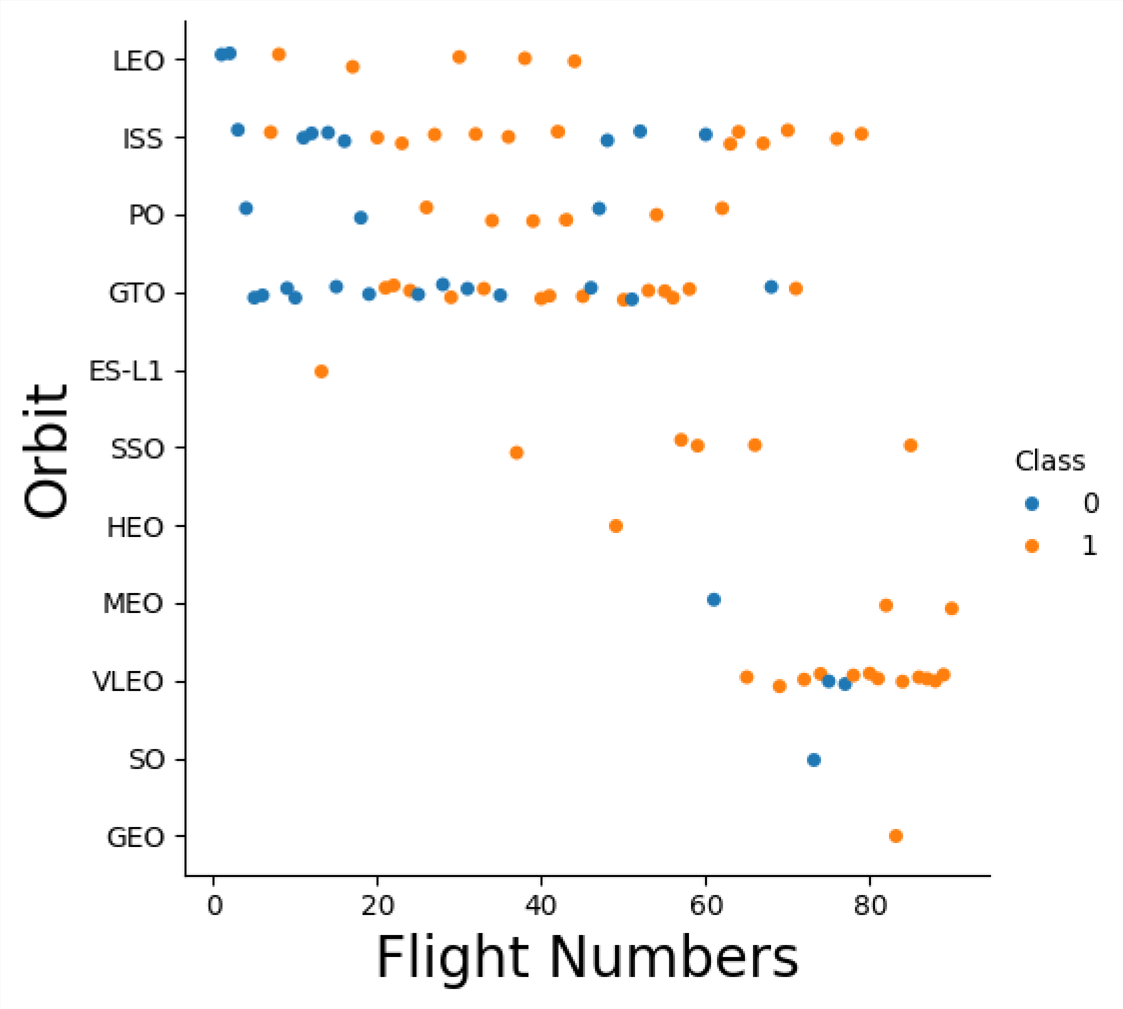
Payload vs. Launch Site



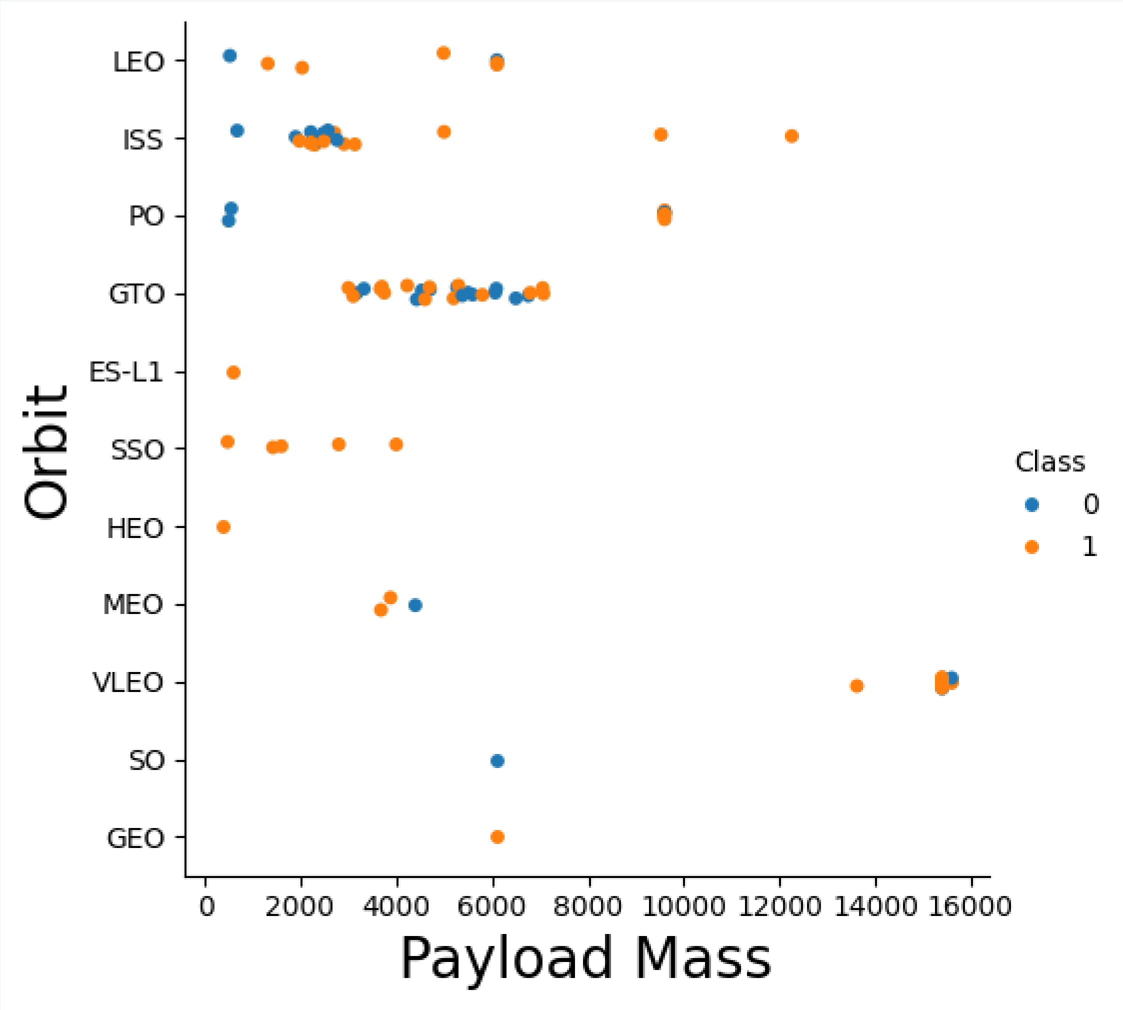
Success Rate vs. Orbit Type



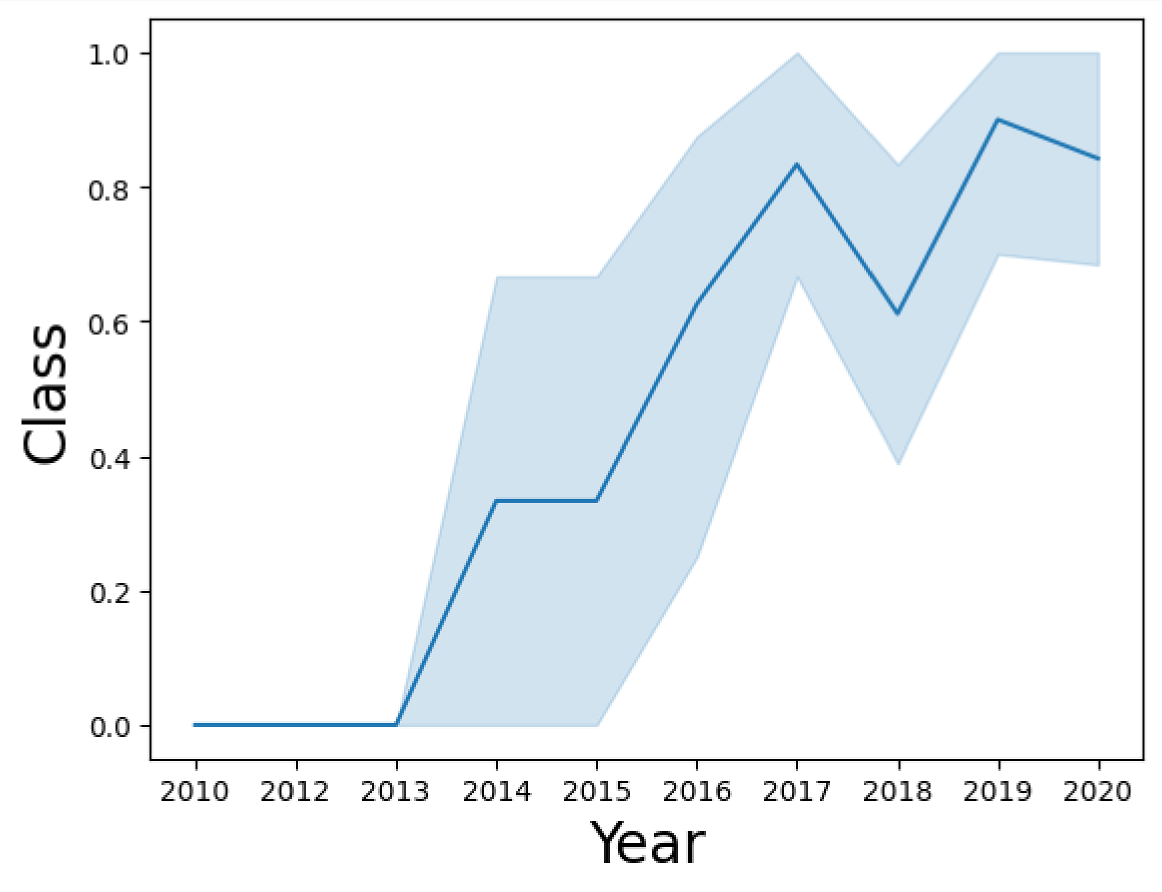
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- `%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;`

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- `%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- ```
  %sql SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total Payload Mass' FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
  ```

Total Payload Mass
45596
  ```

Average Payload Mass by F9 v1.1

- ```
  %sql SELECT AVG(PAYLOAD_MASS__KG_) AS 'Average Payload Mass' FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%'
  ```

Average Payload Mass
2534.6666666666665
  ```

First Successful Ground Landing Date

- ```
  %sql SELECT MIN(Date) AS 'First Successful Landing Outcome' FROM SPACEXTABLE
  WHERE Landing_Outcome = 'Success (ground pad)’
  ```

First Successful Landing Outcome
2015-12-22
  ```

Successful Drone Ship Landing with Payload between 4000 and 6000

- ```
  %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
  ```

Booster_Version

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2
  ```

Total Number of Successful and Failure Mission Outcomes

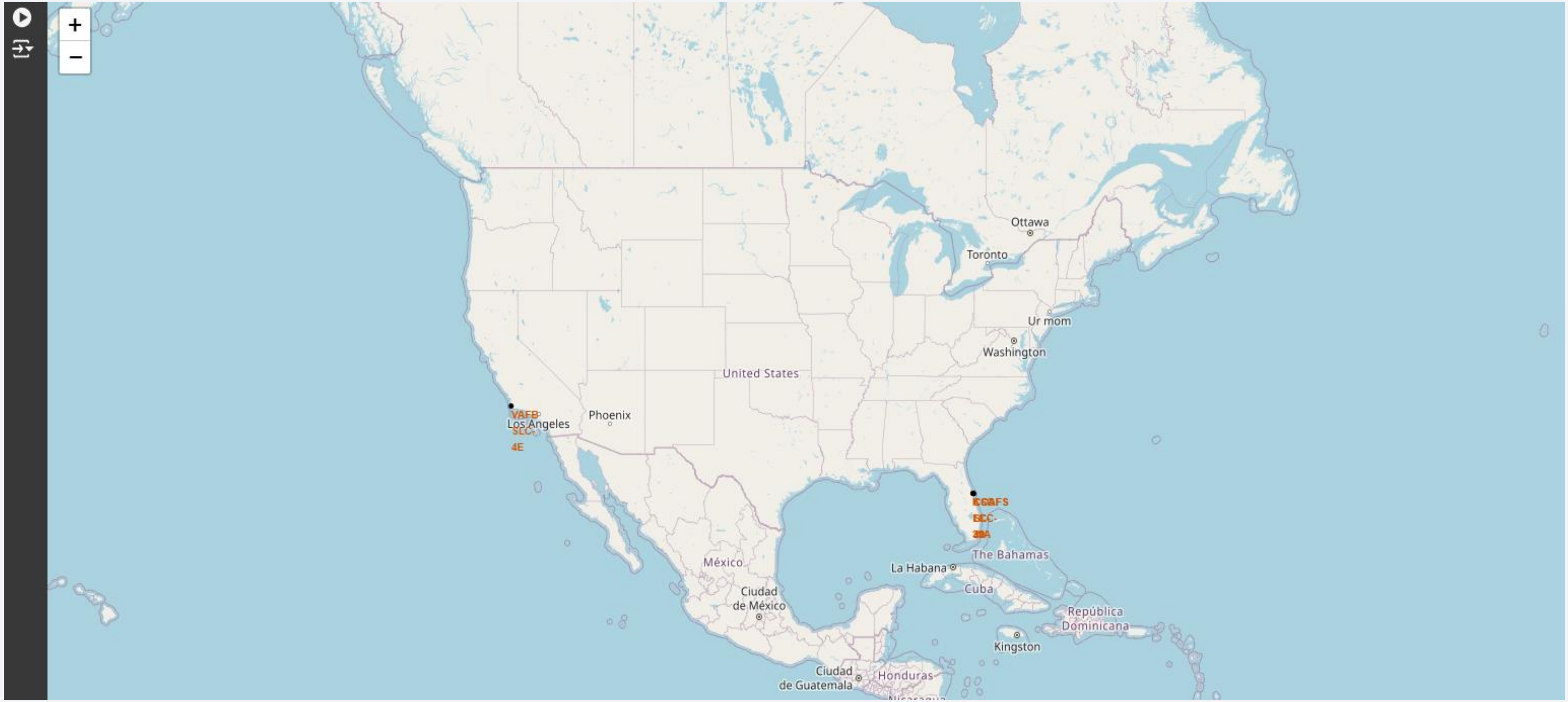
- ```
  %sql SELECT Mission_Outcome, COUNT(*) AS 'Total Mission Outcomes' FROM SPACEXTABLE GROUP BY Mission_Outcome
  ```

Mission_Outcome	Total Mission Outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1
  ```

Boosters Carried Maximum Payload

- `%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)`

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

-
```
%sql SELECT substr(Date, 6,2) AS "Month", Landing_Outcome, Booster_Version,
Launch_Site FROM SPACEXTABLE WHERE substr(Date, 0,5) = "2015" AND
Landing_Outcome = "Failure (drone ship)"
```

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- ```
  %sql SELECT Landing_Outcome, COUNT(*) AS 'Total Outcomes' FROM SPACEXTABLE
  WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome
  ORDER BY 'Total Outcomes' DESC
  ```

Landing_Outcome	Total Outcomes
Uncontrolled (ocean)	2
Success (ground pad)	3
Success (drone ship)	5
Precluded (drone ship)	1
No attempt	10
Failure (parachute)	2
Failure (drone ship)	5
Controlled (ocean)	3
  ```

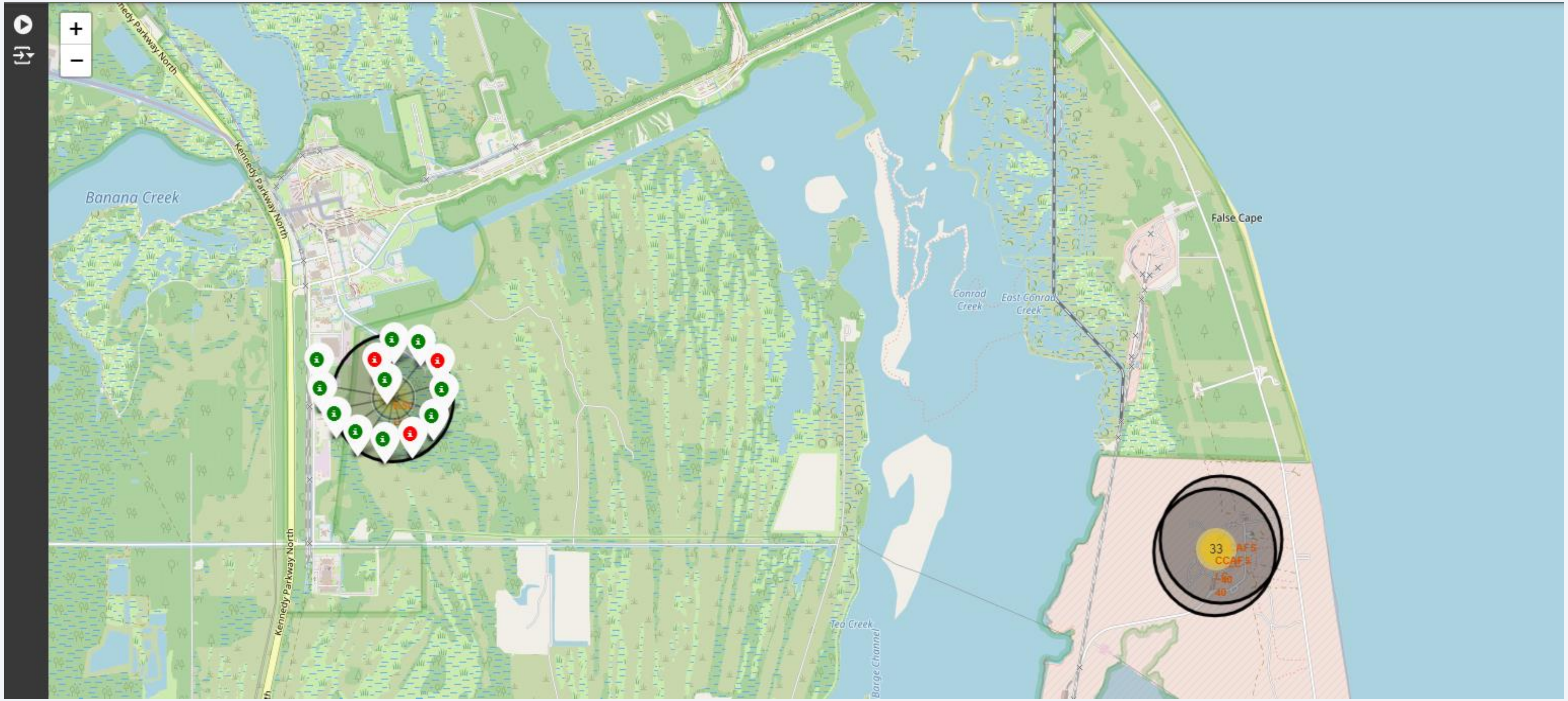
Section 3

Launch Sites Proximities Analysis

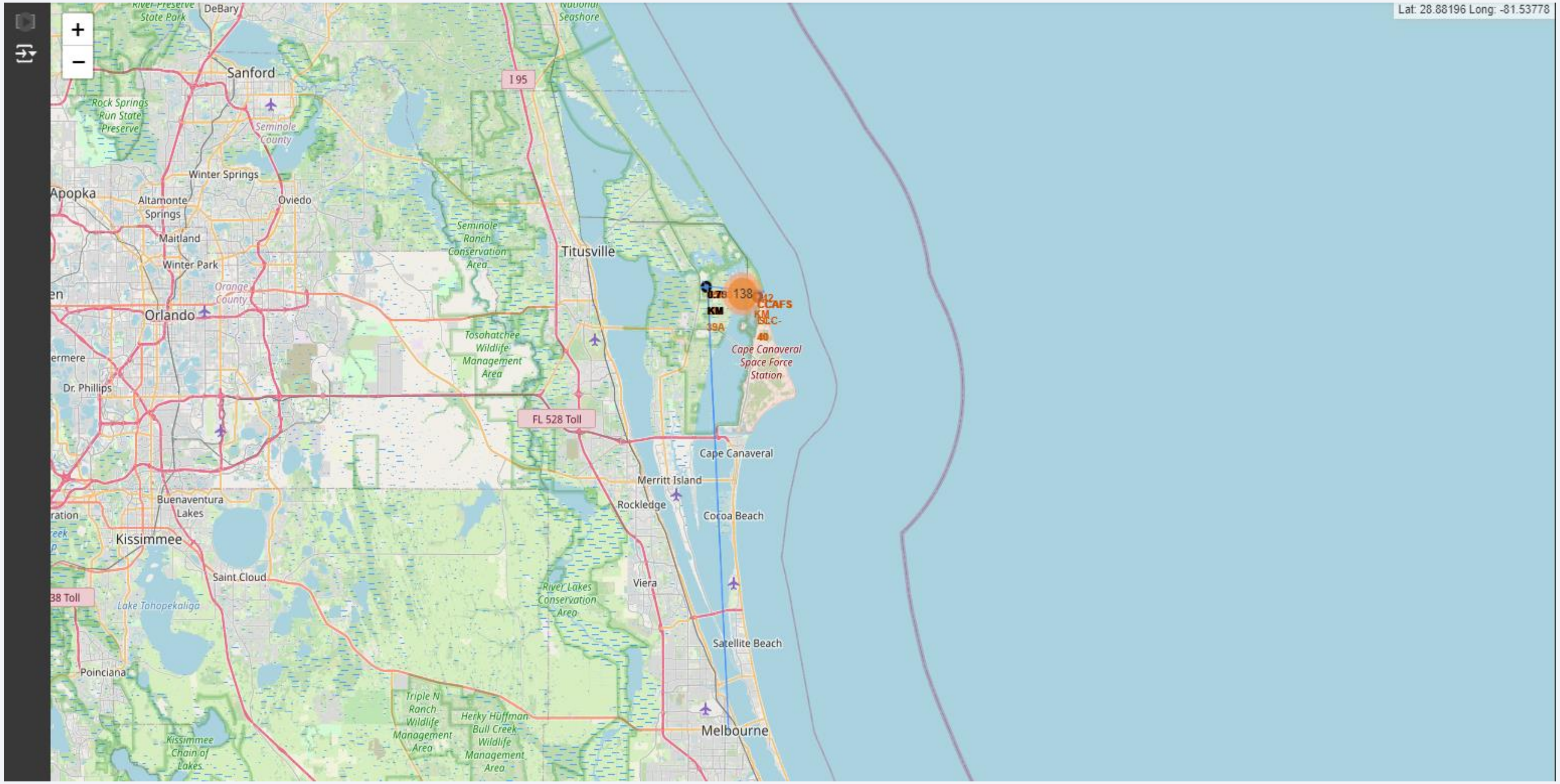
All launch sites on a map



The success/failed launches for each site on the map



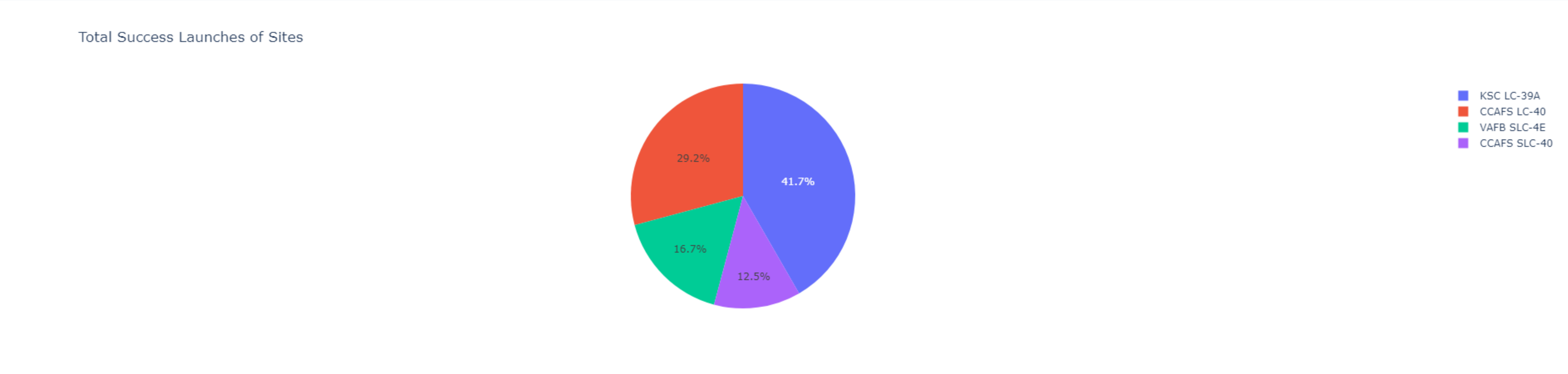
The distances between a launch site to its proximities



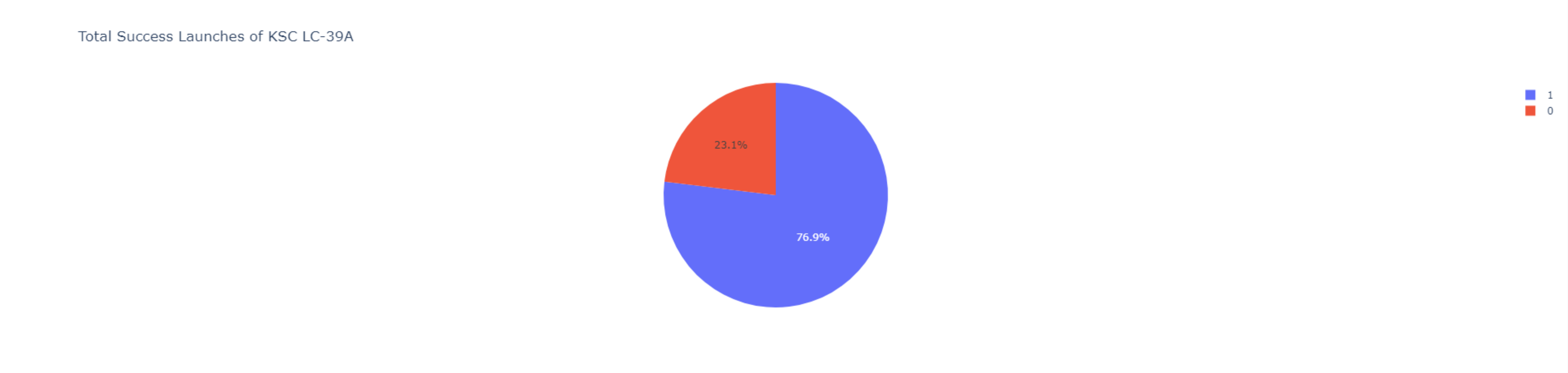
Section 4

Build a Dashboard with Plotly Dash

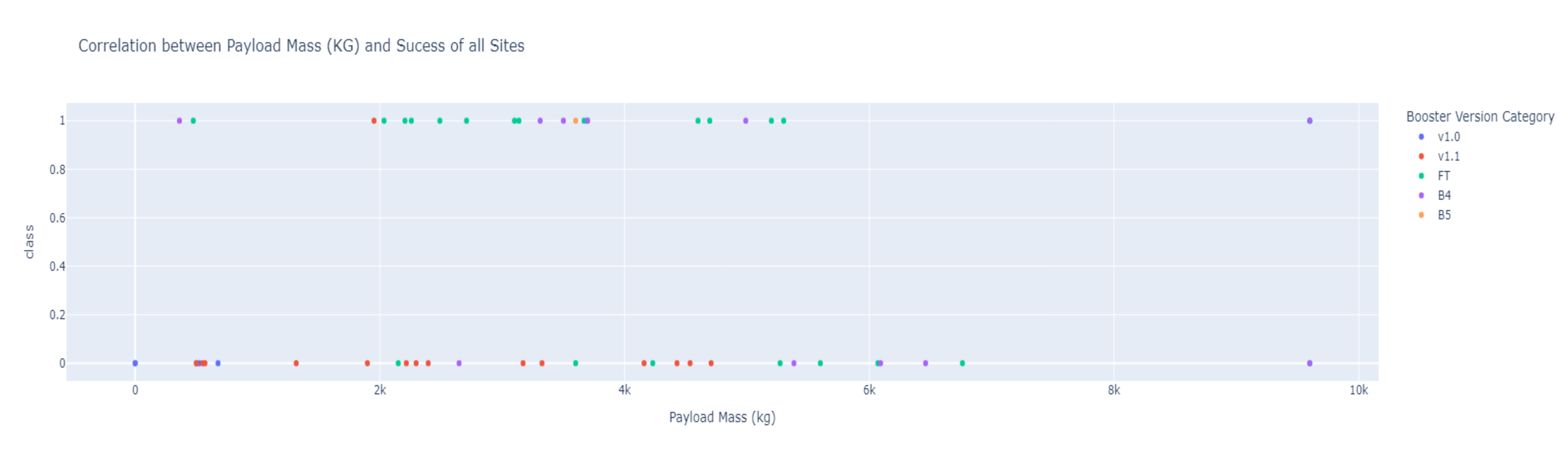
Total Success Launches of Sites



Highest Success Launches of a Site



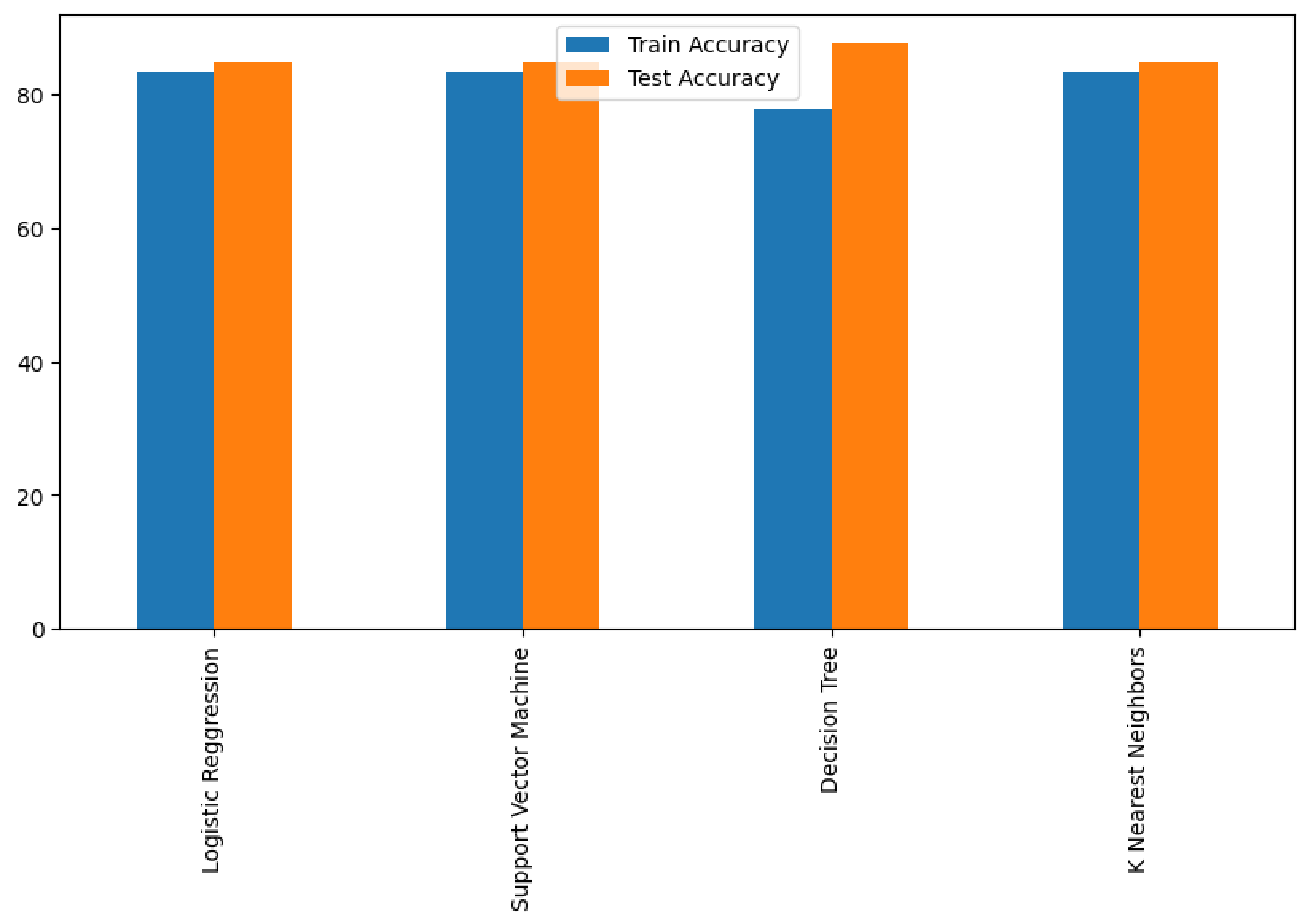
Correlation between Payload and Outcome of Site



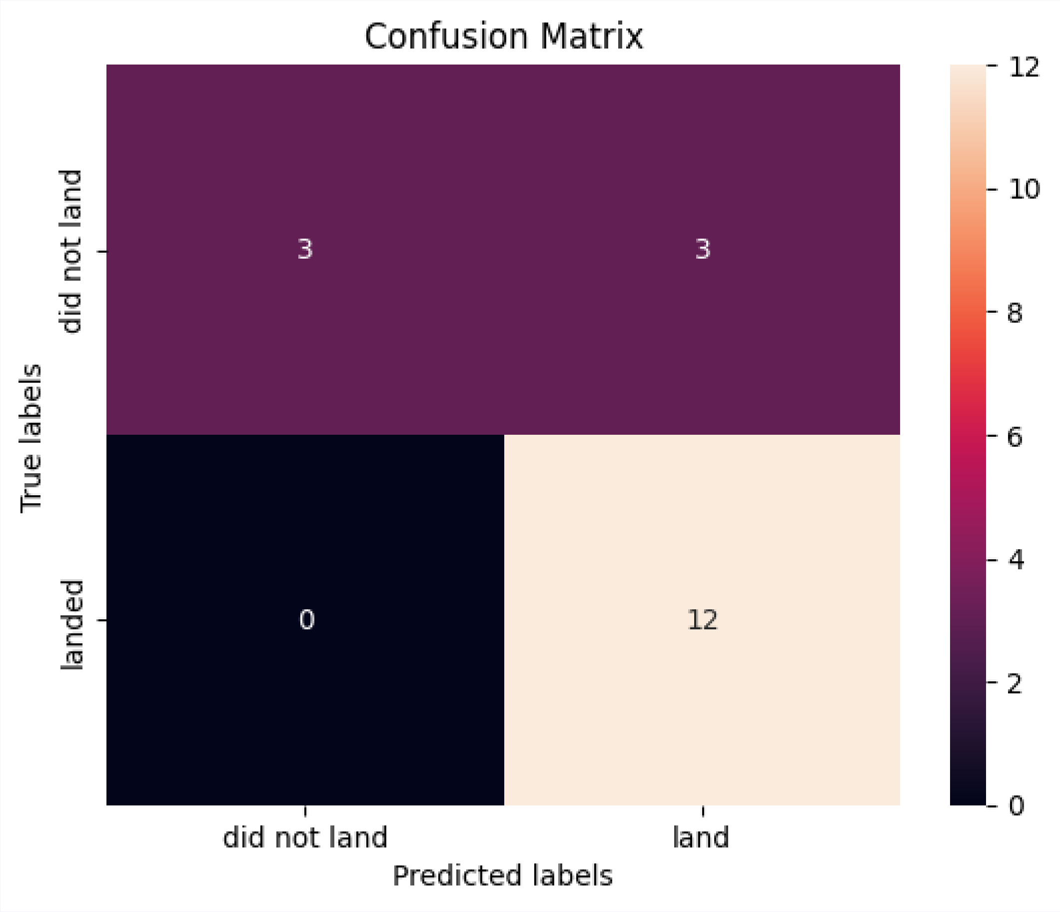
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- 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.
- Logistic Regression can be used to predict successful landings.

Thank you!