IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexander
2025-03-31

Outline

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

Executive Summary

This report presents an analysis of Falcon 9 first-stage landing success using historical launch data. The primary objective of this study was to build a predictive model to estimate whether a Falcon 9's first stage would land successfully, with a focus on identifying key factors influencing the outcome.

- Summary of methodologies

  Data Cleaning, Exploratory Data Analysis (EDA), Machine Learning Models Development, Model Evaluation.

- Summary of all results

  The results indicated that certain variables, such as launch site and payload, were highly influential in determining landing success. Models such **as LogisticRegression** and **KNeighborsClassifier** showed the best performance.

Introduction

The ability to predict whether the Falcon 9 first stage will land successfully can influence not only SpaceX's cost structures but also the competitive dynamics in the space launch market. If the success prediction can be made accurately, it would allow SpaceX, and potentially other companies, to better estimate costs and refine pricing models.

The main objective of this project is trying to determine if the first stage will land. This information can be used if an alternate company wants to bid against space X for a rocket launch. In this lab, you will create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Request to the SpaceX API to collect previous data
- Perform data wrangling
  - Data cleaning method to handle missing data, data formation, etc.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Select models, train the models and evaluate the models.

Data Collection

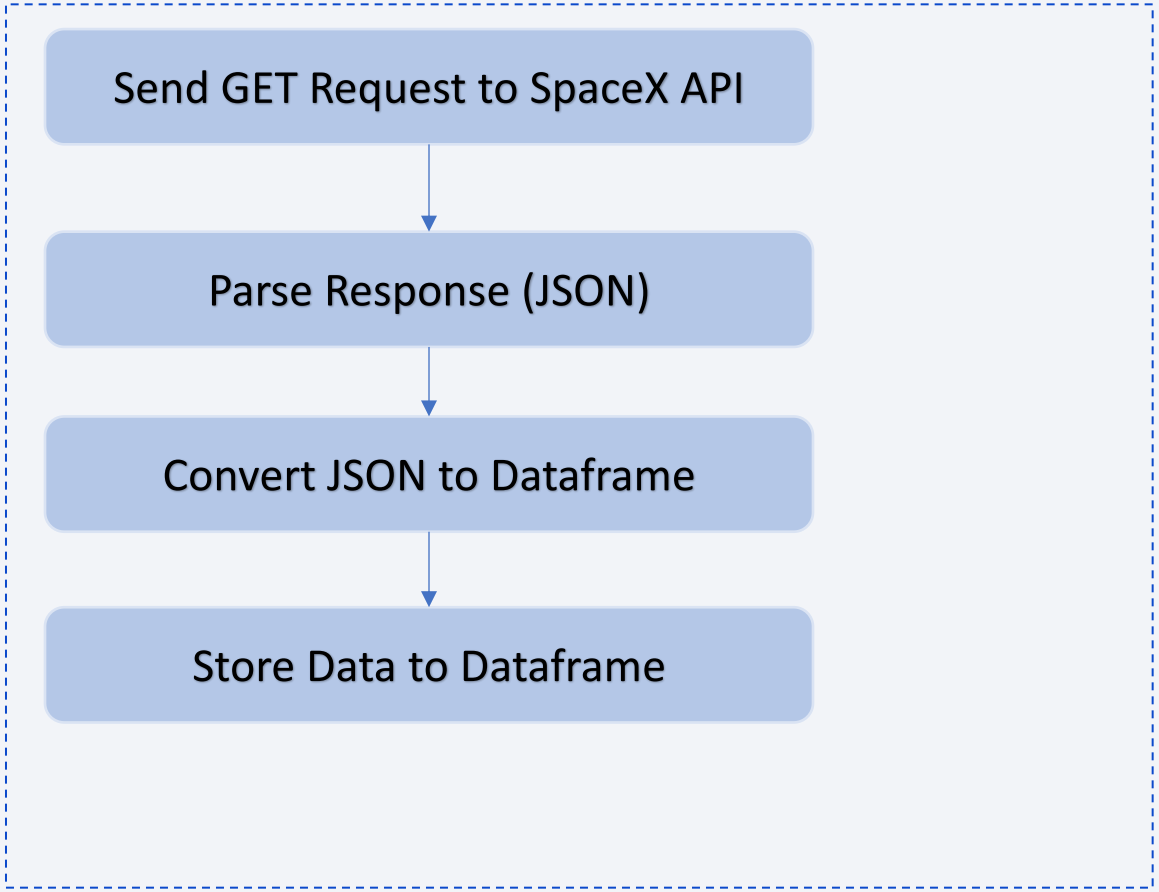
By using the **SpaceX API**, the data collection process is automated and efficient, enabling us to gather historical launch data about SpaceX missions.

Data Collection – SpaceX API

- Use request to the SpaceX API to get the JSON file data, then convert it to pandas dataframe.

- GitHub URL of the completed SpaceX API calls notebook:

  https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/jupyter-labs-spacex-data-collection-api.ipynb

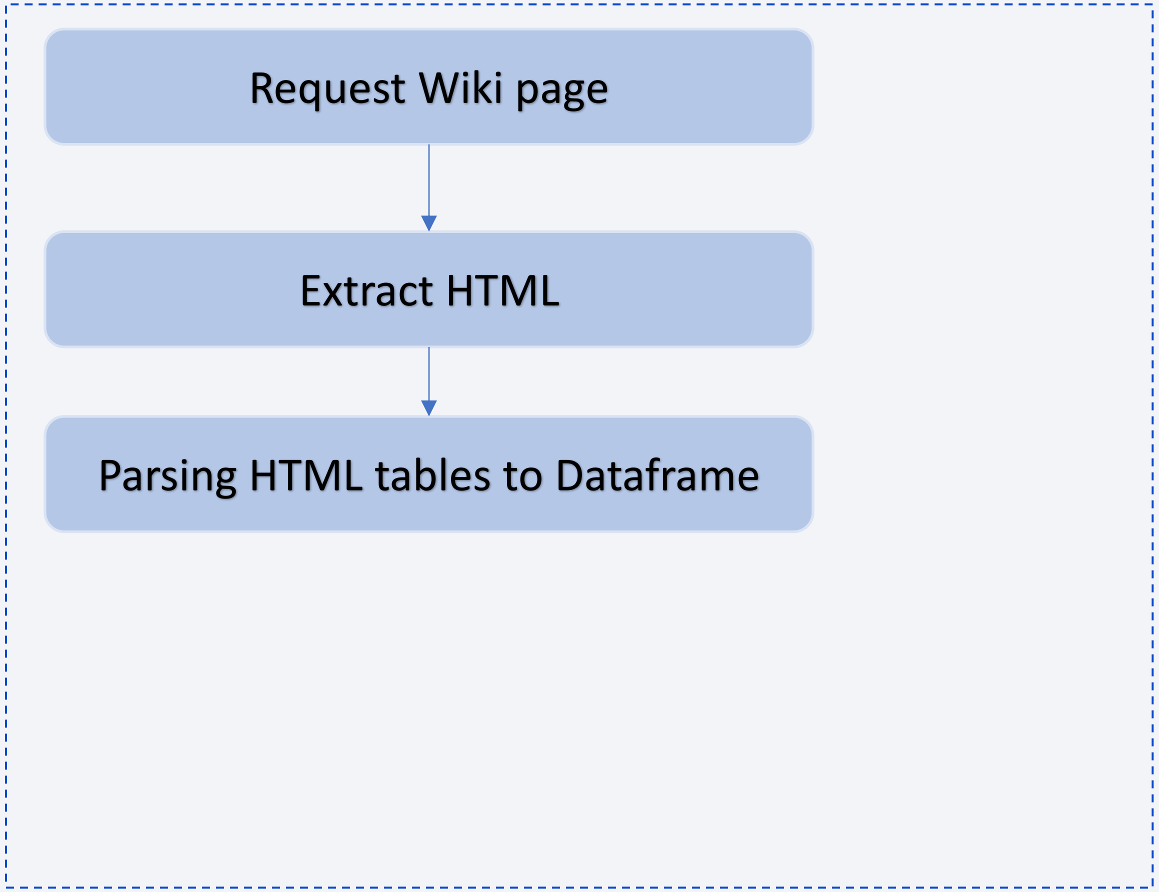


Data Collection - Scraping

- Request Falcon 9 wiki page, then extract the HTML data to dataframe

- GitHub URL of the completed web scraping notebook:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Data wrangling process involves cleaning, transforming, and structuring raw data into a usable format for analysis. First is data loading, second is handling missing data, then is data transformation.

- GitHub URL of completed data wrangling related notebooks:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- **Catplot. It allows me to visualize patterns across categories and understand the spread of data within each category.**
- **Barplot. It gives a clear visual representation of comparisons between categorical variables.**
- **Scatterplot. It can visualize how one variable affects another.**
- **Line chart. It is great for visualizing time-dependent patterns and trends**

- GitHub URL of completed EDA with data visualization notebook:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/edadataviz.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Read data into a DataFrame.
- Count rows.
- Select limit rows.
- Select sum, average, min by specific category.
- Use sub select.
- Use substr
- GitHub URL of completed EDA with SQL notebook:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Added the following map objects using Folium
- Markers. Clearly identify launch locations on the map.
- Circles. Shows the area of launch sites.
- Lines. Helps analyze how far to a closest city, railway, highway, etc.

- GitHub URL of your completed interactive map with Folium map:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Added the following plots/graphs a dashboard
- Pie Chart. Easily visualizes the success rate of SpaceX's sites.
- Scatter Chart. Shows how payload weight affects landing success.

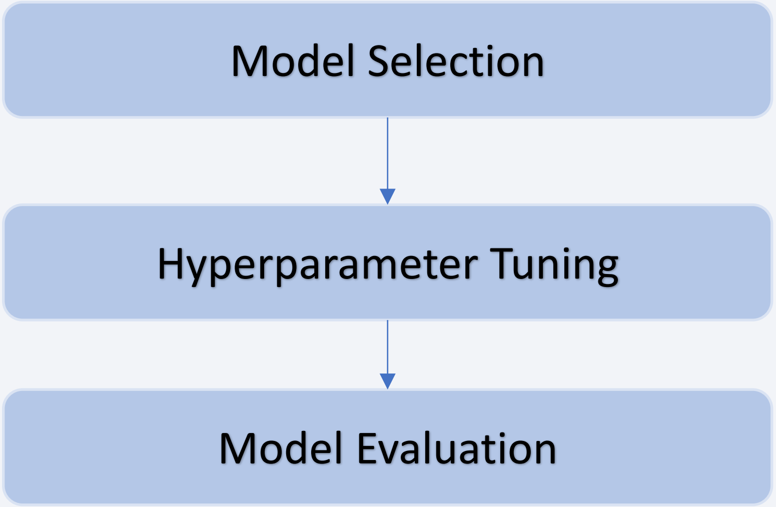
- GitHub URL of your completed Plotly Dash lab:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/spacex-dash-app.py

https://github.com/AlexTank99/IBM_Data_Science/blob/2929b35d6779164e8711dfe8f35b6f6dd7c1e11d/Applied%20Data%20Science%20Capstone/spacex-dash-app.png

Predictive Analysis (Classification)

- Following steps to impletement predictive analysis.
  - Model Selection & Baseline Evaluation. Tested multiple classification models to find the best-performing one:
  - Hyperparameter Tuning (Model Optimization). Used GridSearchCV to optimize parameters:
  - Model Evaluation. Get best accuracy model.



- GitHub URL of your completed predictive analysis lab:

https://github.com/AlexTank99/IBM_Data_Science/blob/2929b35d6779164e8711dfe8f35b6f6dd7c1e11d/Applied%20Data%20Science%20Capstone/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

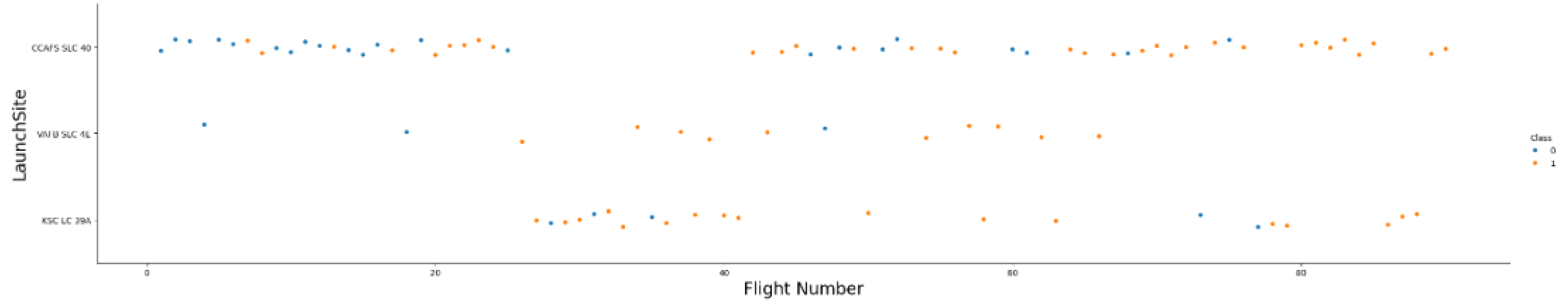
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

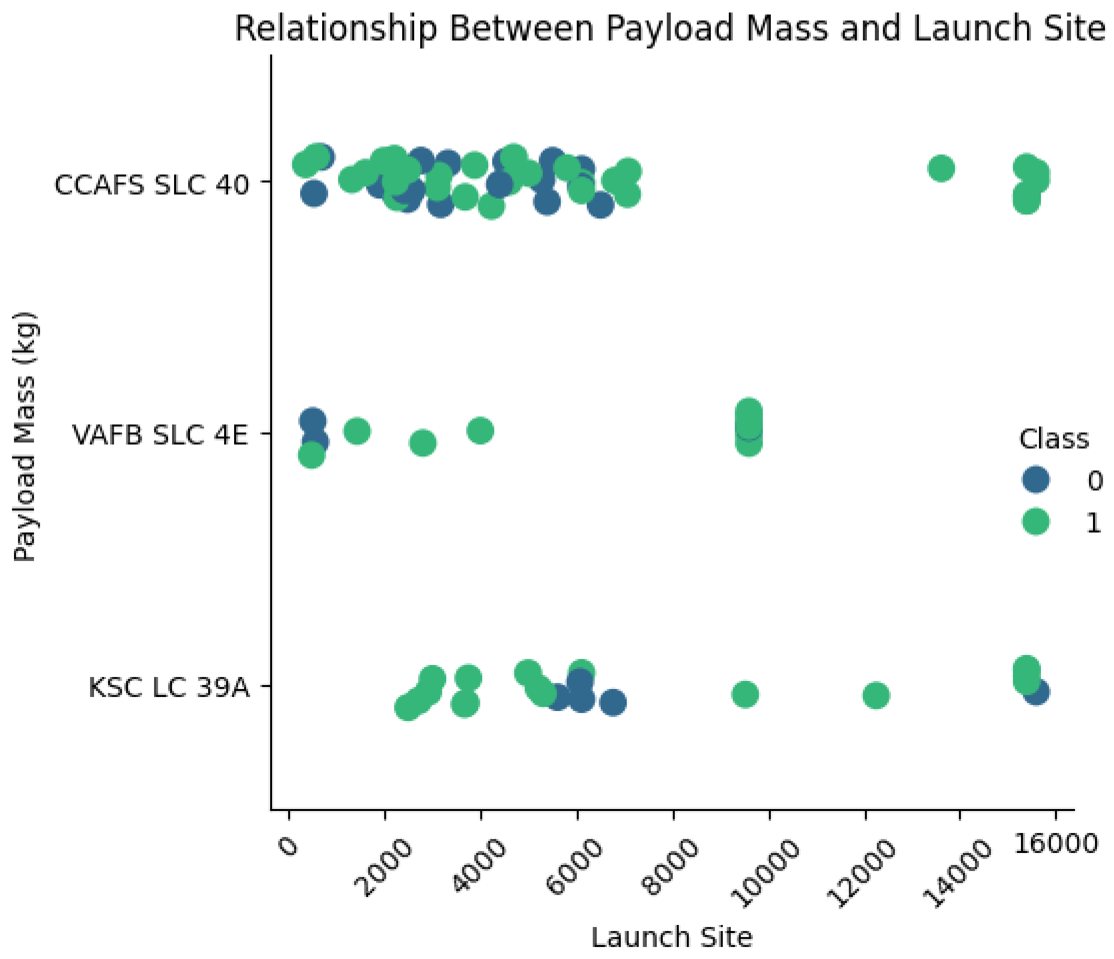
Flight Number vs. Launch Site



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

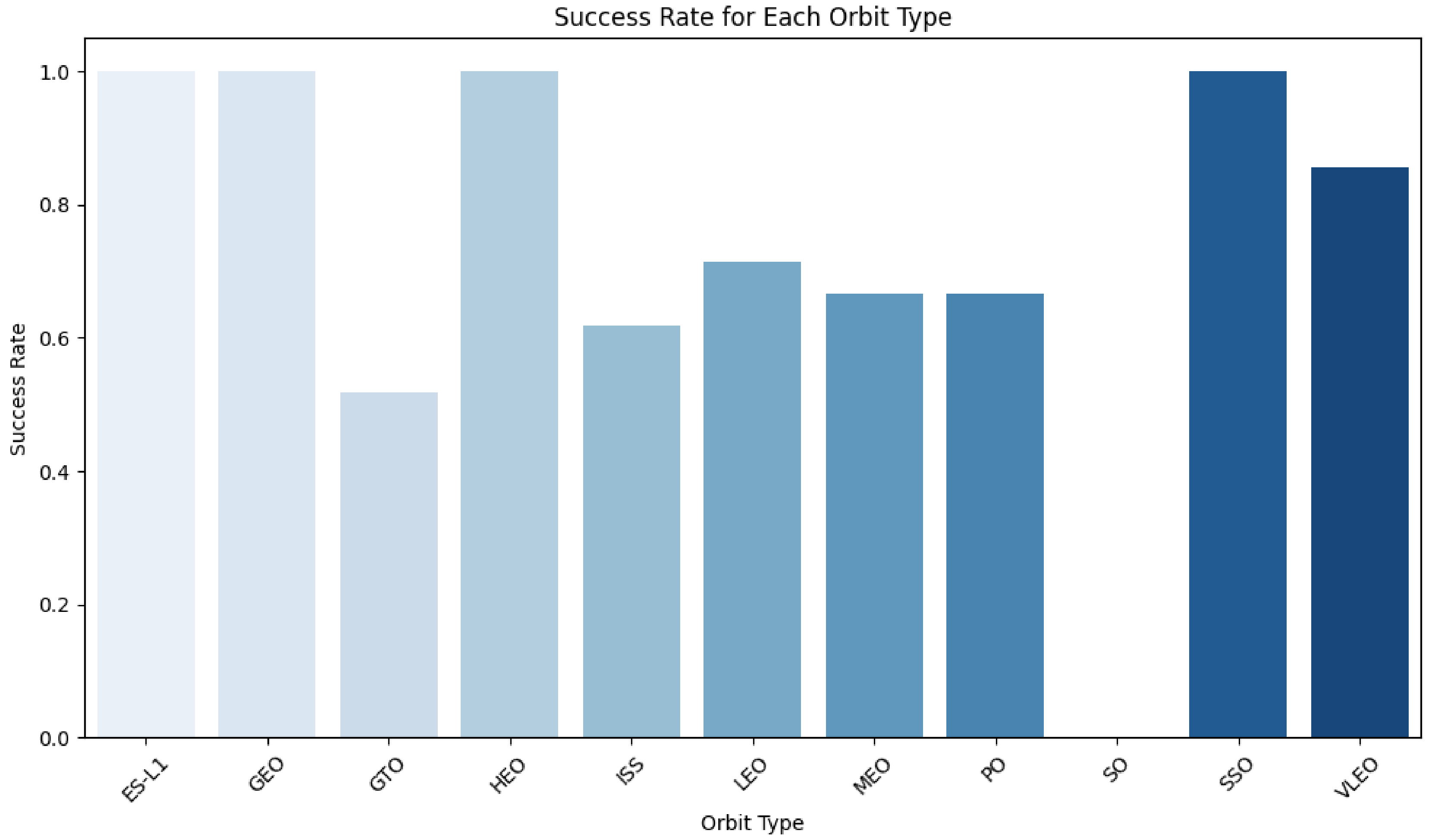
```
Launch Site "CCAFS SLC 40" flights and success rate increased over flight, "VAFB SLC 4E" flights decreased.
```

Payload vs. Launch Site



Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

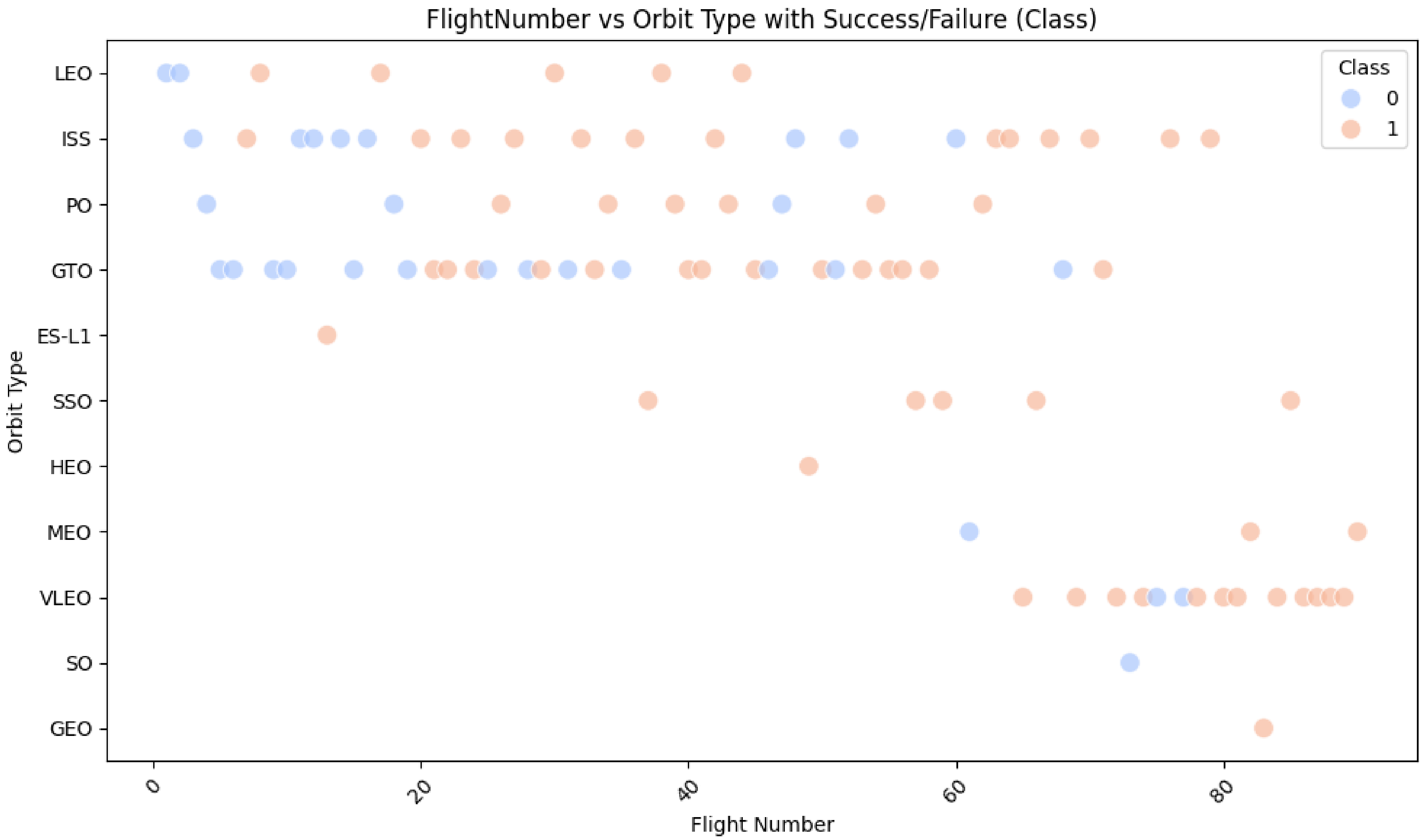
Success Rate vs. Orbit Type



Analyze the plotted bar chart to identify which orbits have the highest success rates.

```
ES-L1, GEO, HEO and SSO have the highest success rates.
```

Flight Number vs. Orbit Type



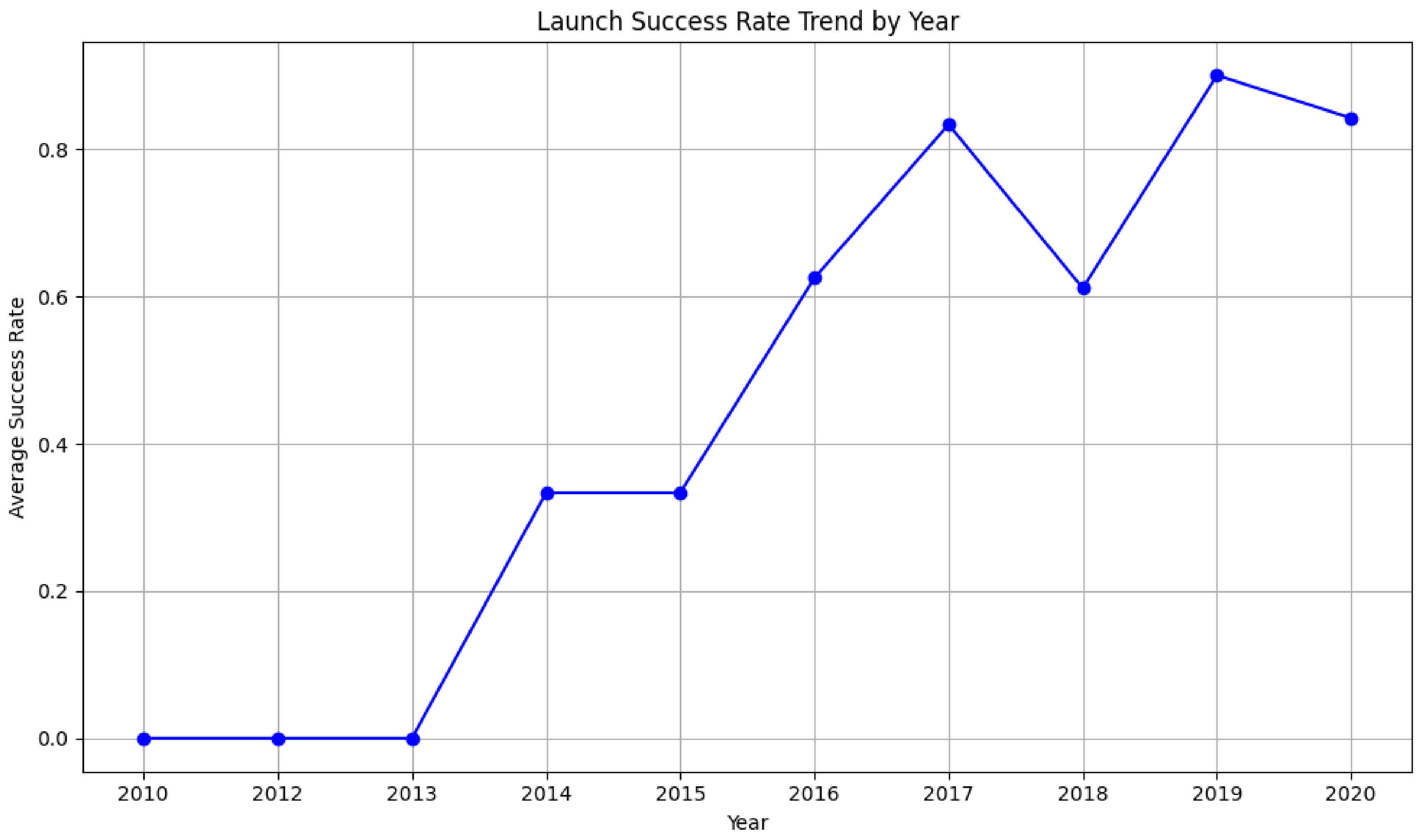
You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
```

```
Totally, there are four launch sites named CCAFS LC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40.
```

Launch Site Names Begin with 'CCA'

```
('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')
```

Above are the total records where launch sites begin with the string "CCA".

Total Payload Mass

```
Total Payload Mass carried by NASA (CRS): 45596 kg
```

```
As showed above, the total payload mass is 45596 kg.
```

Average Payload Mass by F9 v1.1

```
Average Payload Mass carried by F9 v1.1: 2534.67 kg
```

As showed above, the average payload mass is 2534.67 kg.

First Successful Ground Landing Date

```
First successful landing on ground pad: 2015-12-22
```

```
First successful landing outcome on ground pad is 2015-12-22.
```

Successful Drone Ship Landing with Payload between 4000 and 6000

```
Boosters with successful drone ship landing and payload between 4000 and 6000 kg:
F9 FT B1022
F9 FT B1026
F9 FT  B1021.2
F9 FT  B1031.2
```

Boosters with successful drone ship landing and had payload mass greater than 4000 but less than 6000 kg as above.

Total Number of Successful and Failure Mission Outcomes

```
Mission Outcomes Count:
Failure (in flight): 1
Success: 98
Success : 1
Success (payload status unclear): 1
```

```
The total number of successful and failure mission outcomes are 98 and 1.
```

Boosters Carried Maximum Payload

```
Booster Versions that carried the maximum payload mass:
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7
```

Above are the names of the booster which have carried the maximum payload mass.

2015 Launch Records

```
('January', 'F9 v1.1 B1012', 'CCAFS LC-40')
('April', 'F9 v1.1 B1015', 'CCAFS LC-40')
```

Above are the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.

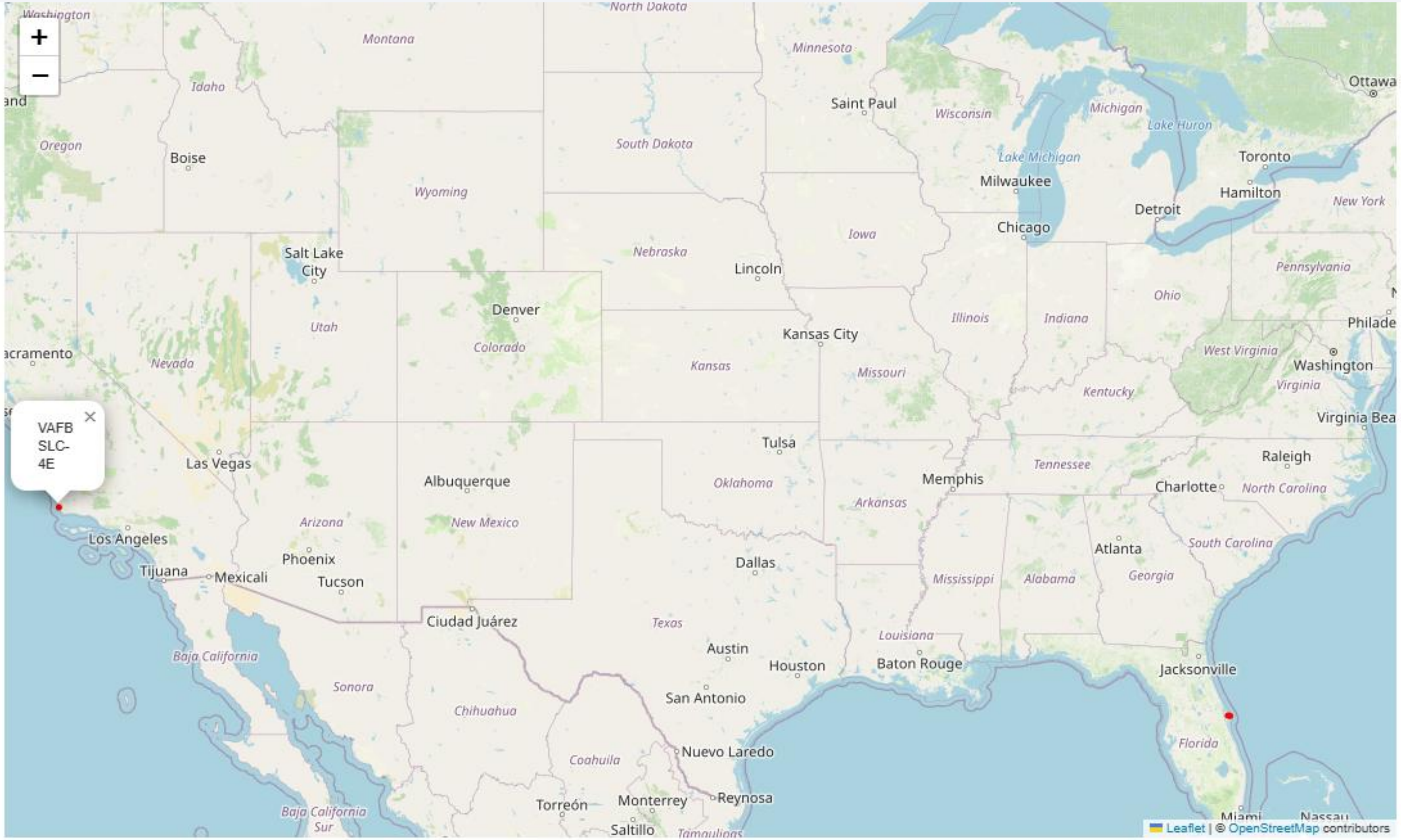
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

Above are 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.

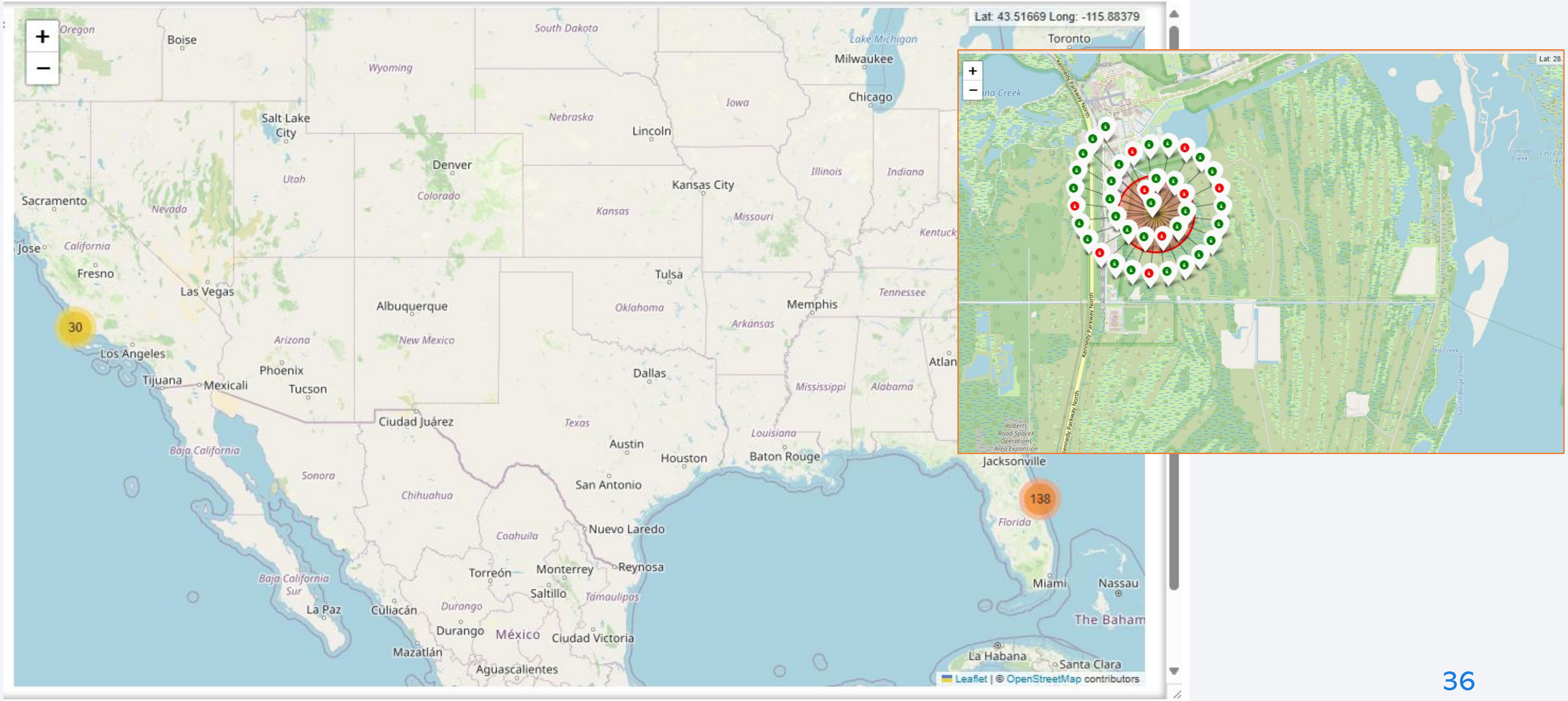

Section 3
Launch Sites
Proximities Analysis

Launch Sites on Map



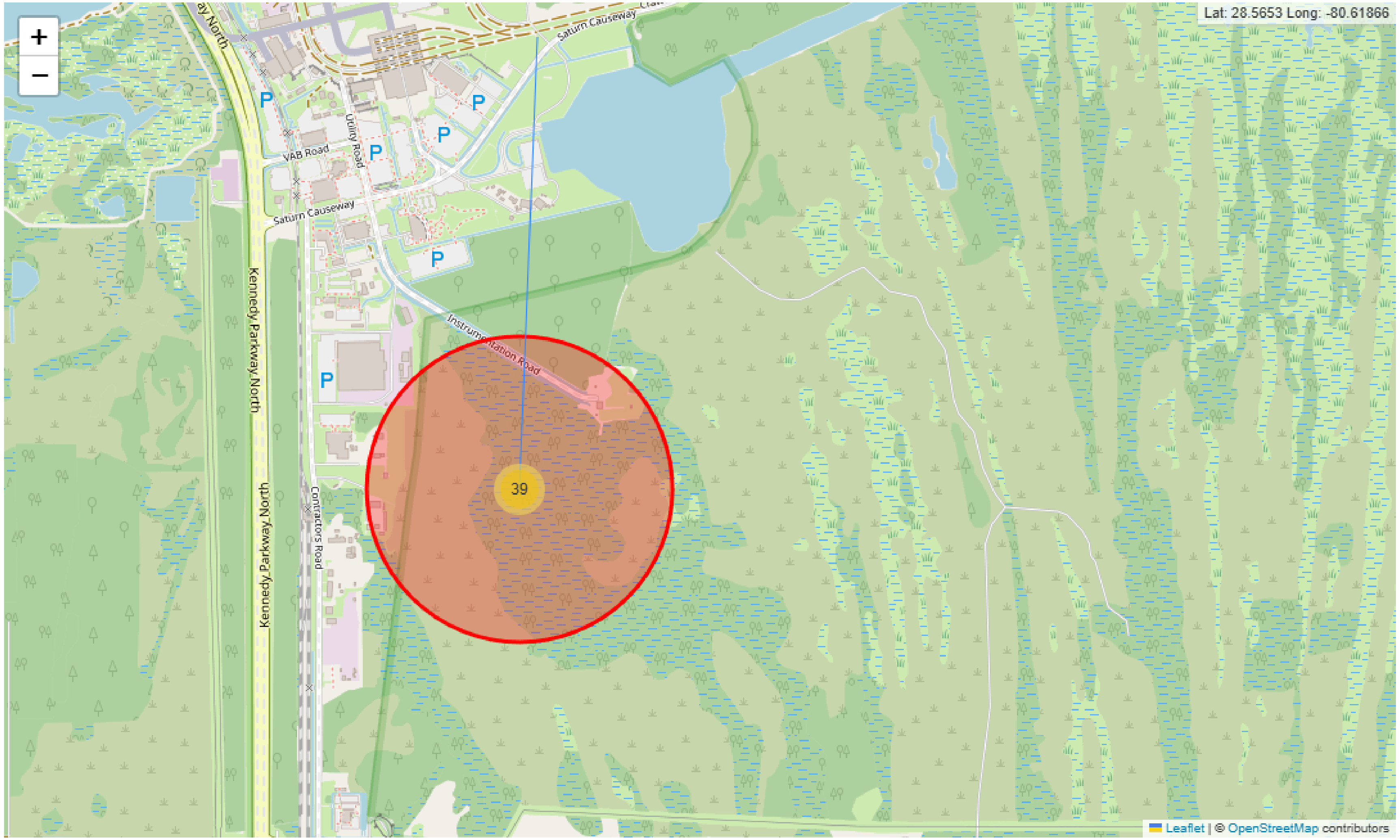

```
Total 4 launch sites marked as above.
```

Color-labeled Launch Outcomes on Map



Above is the color-labeled launch outcomes on the map.

Launch Site with Proximities

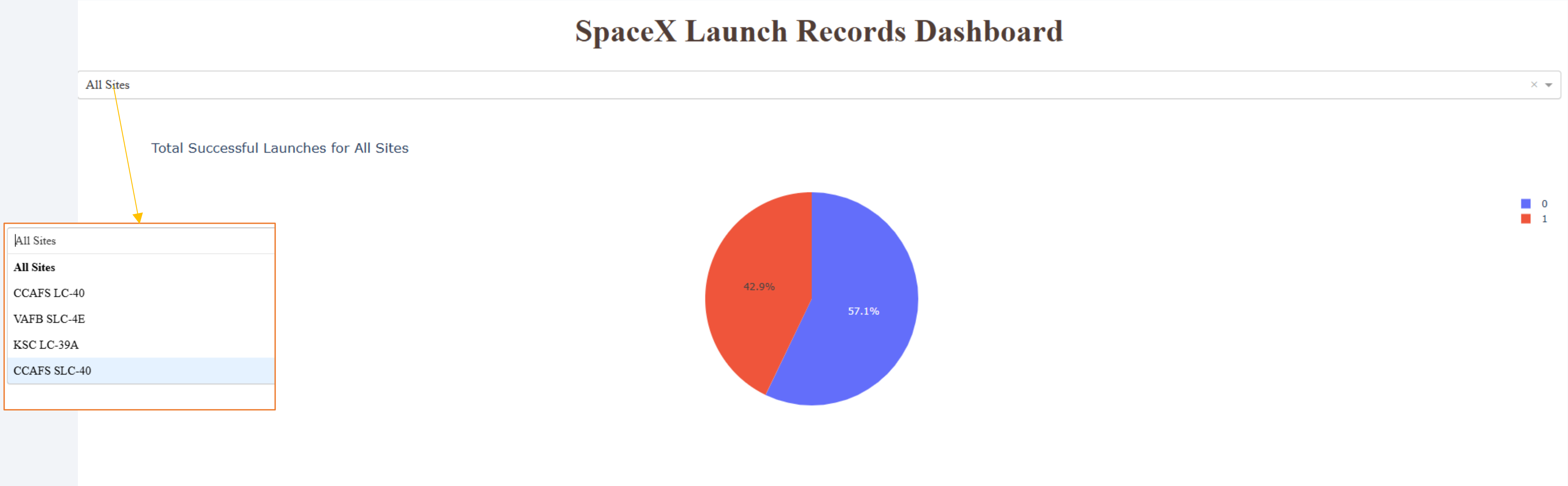


Above is a show of launch site to its proximities such as railway.

Section 4

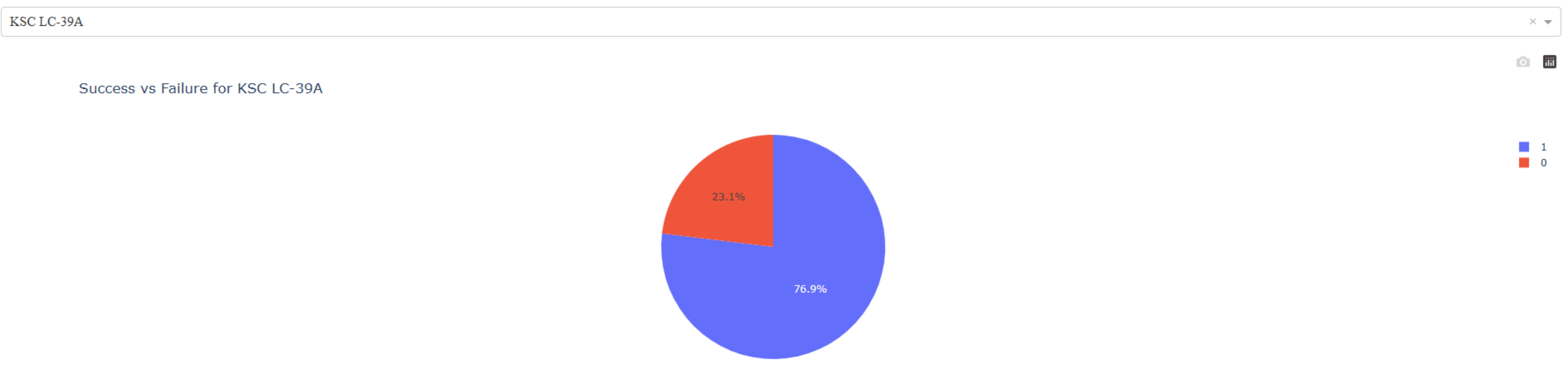
Build a Dashboard with Plotly Dash

Successful Launches of All Sites



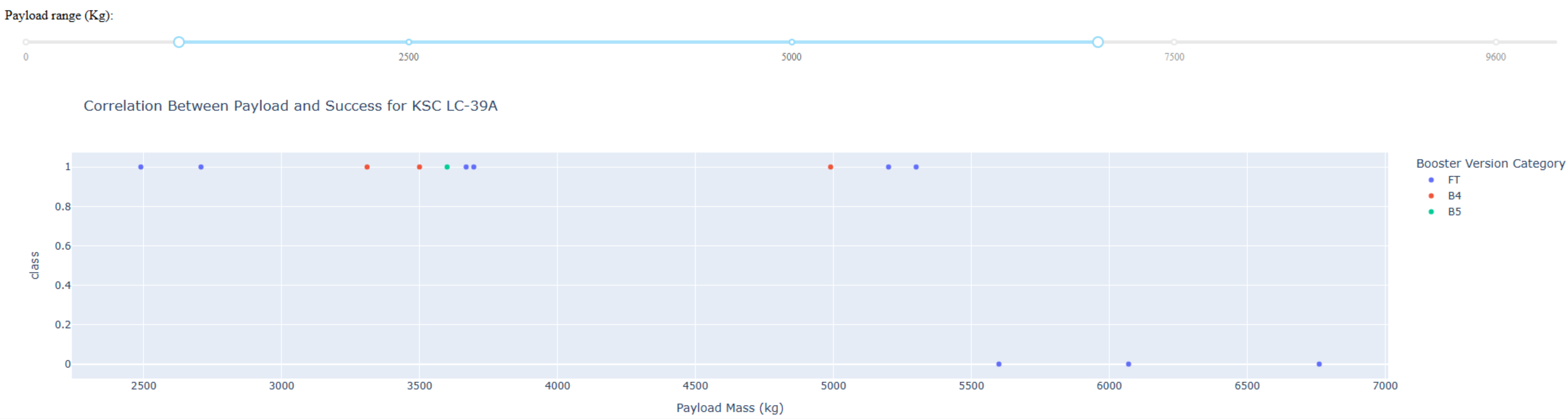
- As above showed, launch sites with success ratio by pie char.

Highest Launch Success Ratio



- As above showed, highest launch sites with success ratio by pie char.

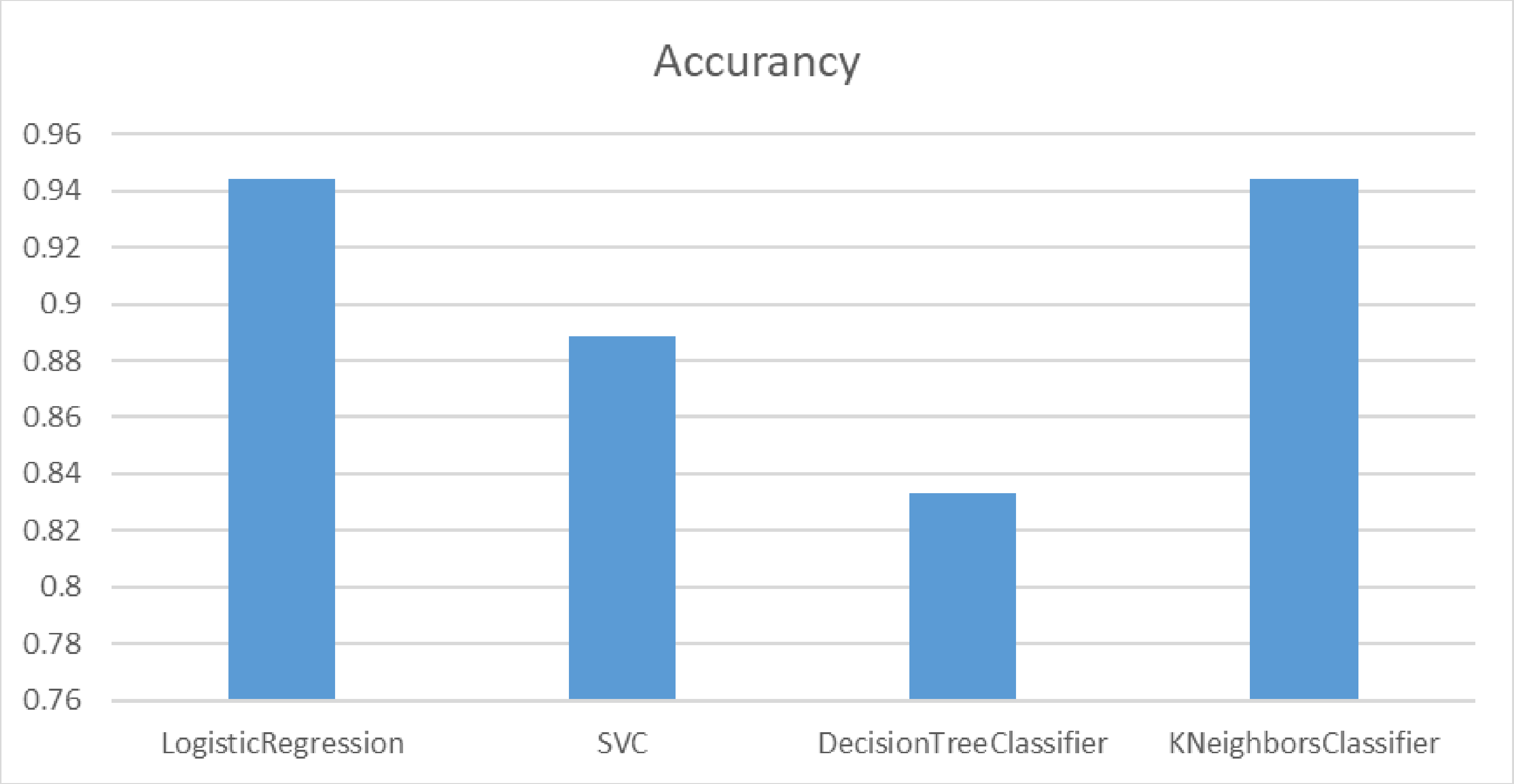
Payload vs. Launch Outcome



Section 5

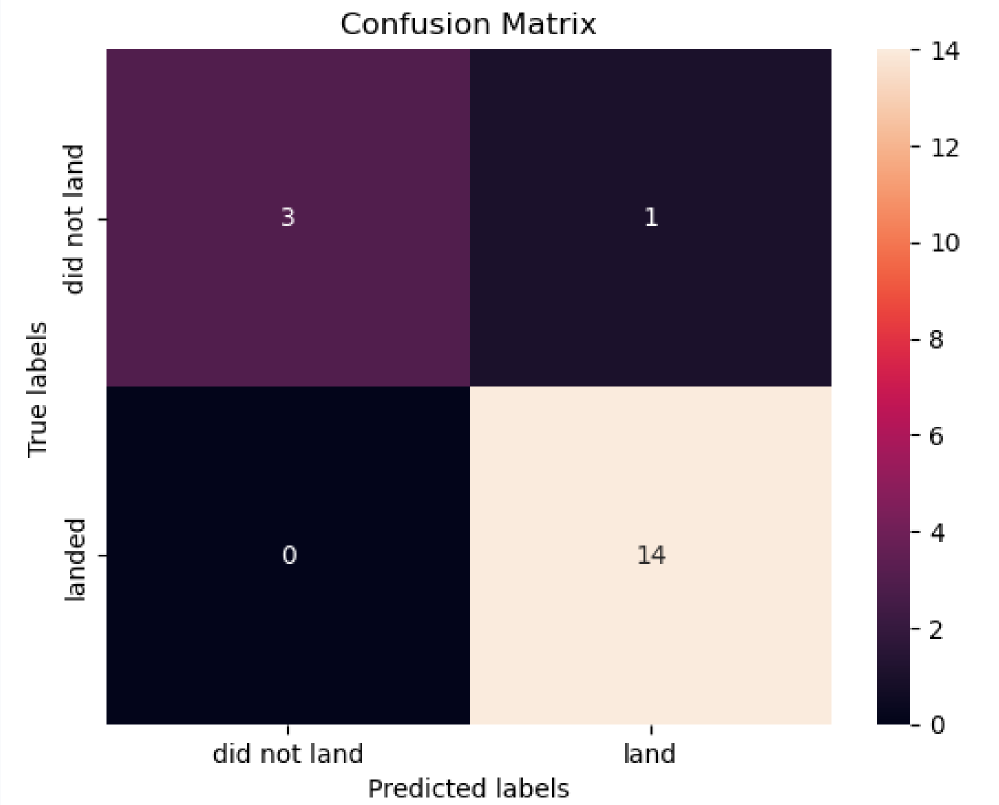
Predictive Analysis (Classification)

Classification Accuracy



- Logistic Regression and KNeighborsClassifier have the highest accuracy.

Confusion Matrix



Overview:

True Postive - 14 (True label is landed, Predicted label is also landed)

False Postive - 3 (True label is not landed, Predicted label is landed)

Conclusions

- Logistic Regression and KNeighborsClassifier have the highest accuracy.
- Logistic Regression and KNeighborsClassifier, predictive results:

  True Postive - 14 (True label is landed, Predicted label is also landed)

  False Postive - 3 (True label is not landed, Predicted label is landed)

Appendix

- Python code for dash app:

https://github.com/AlexTank99/IBM_Data_Science/blob/7853161368c89fa44c6f379bee3072b14b0a7f52/Applied%20Data%20Science%20Capstone/spacex-dash-app.py

- Bar-chart for model comparation:

https://github.com/AlexTank99/IBM_Data_Science/blob/c76ec45196d1f9ff4d177b415d4d33c4ca2cf663/Applied%20Data%20Science%20Capstone/ds-capstone-template-coursera%20bar-chart.xlsx

Thank you!