IBM Developer
SKILLS NETWORK

SpaceX Analytics

Presentation by Faizaan Al Faisal

March 1st, 2024

Overview

01

Executive Summary

The following methodologies were used to analyze data:

- Data Collection using web scraping and SpaceX API;
- Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
- Machine Learning Prediction.

Summary of all results:

- It was possible to collected valuable data from publicly available sources;
- EDA allowed to identify which features are the best to predict success of launchings;
- Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way

02

Introduction

01

The objective is to evaluate the viability of a new company, SpaceY, to compete with SpaceX

02

Using SpaceX’s publicly available data, SpaceY wants to forecast costs and success

03

Various features like launch sites, payload masses, types of boosters, etc., can determine these factors

04

Desirable answers:

- Estimate cost of successful landings
- Locations to launch from
- Specifications of rocket

SECTION 1
DATA GATHERING

03

Data Collection

Using SpaceX API

Data was collected through utilizing SpaceX's own API for their publicly available launch data.

URL: SpaceX's API

Using Webscraping

Additional Data was collected through publicly maintained records on Wikipedia.

URL: Wikipedia Falcon 9 Launches

04

Data Wrangling

01 Exploratory Data Analysis (EDA):

- Identify patterns in Space X data
- Determine training labels for supervised models

02 Data Analysis:

- Calculate launches per site
- Analyze orbit occurrences and mission outcomes

03 Outcome Labeling:

- Create a landing outcome label
- Classify launches as successful or unsuccessful

IBM Developer
SKILLS NETWORK

SECTION 2
EXPLORATORY DATA ANALYSIS

05

Exploratory Data Analysis

Data Visualization

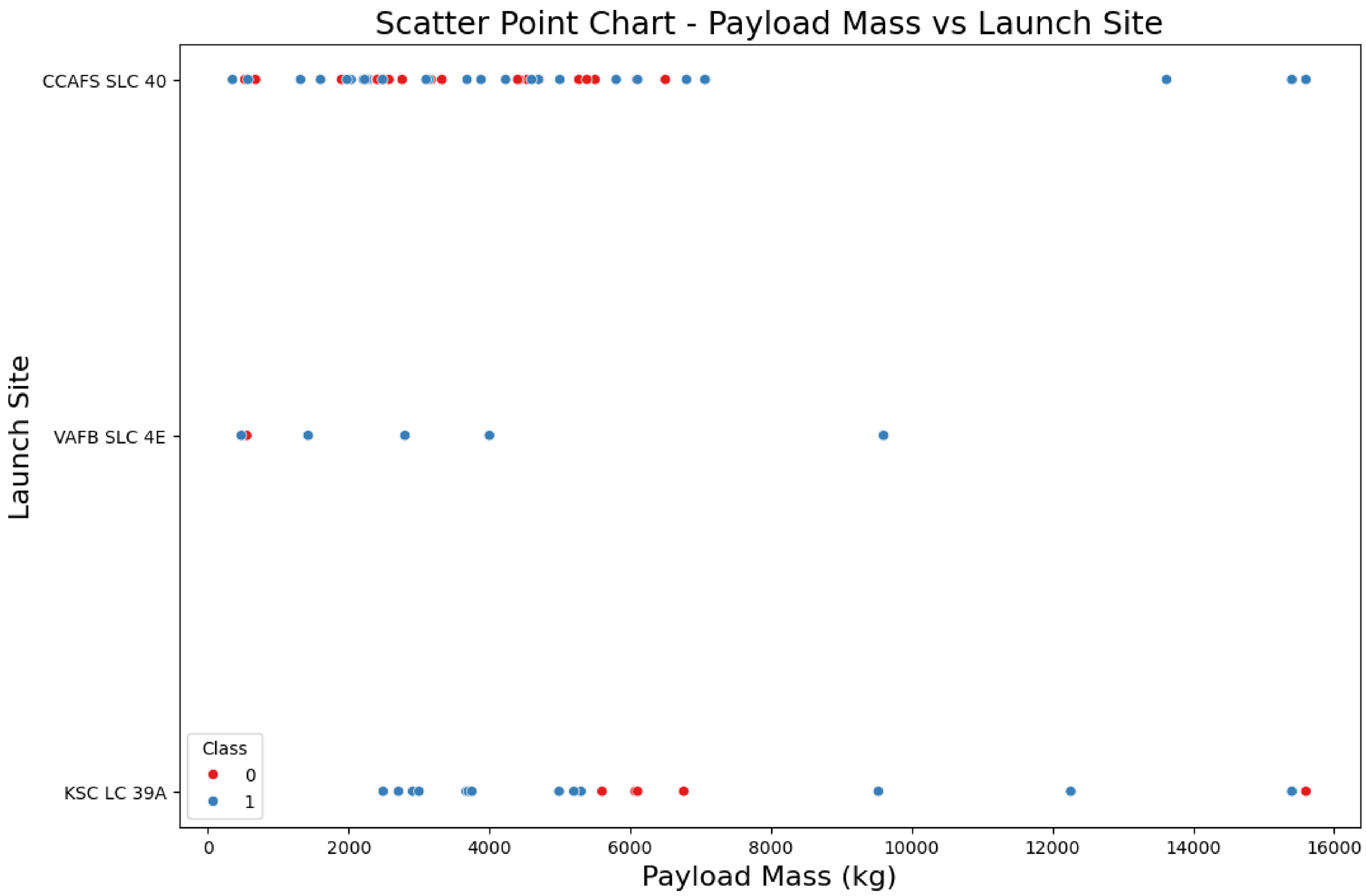
Visualizing the data provides key insights into the trends of different launch factors and variables, such as payloads, boosters utilized, etc.

SQL

Utilizing SQL allows for more objective quantitative analysis of the data acquired. This can confirm the exact values

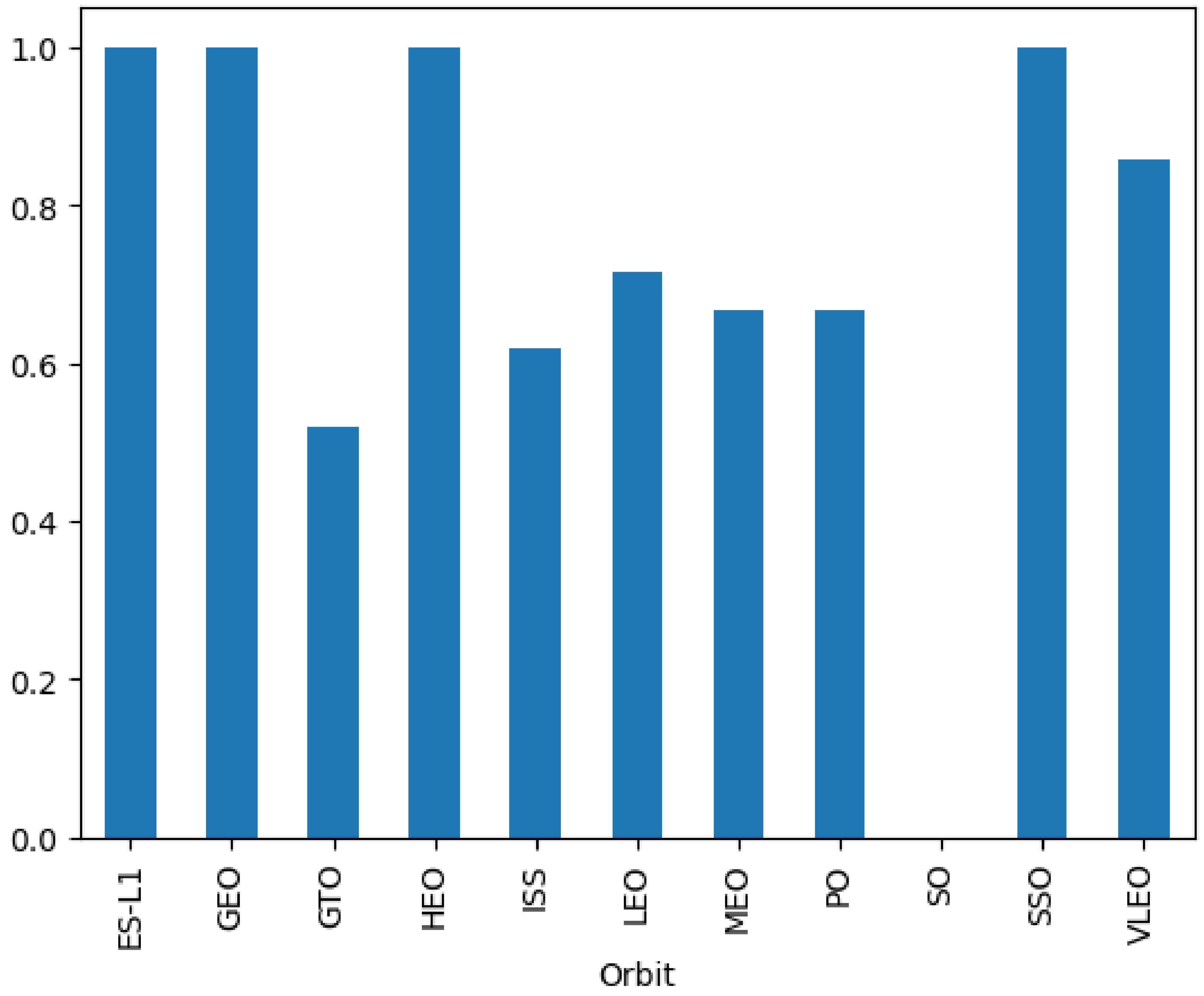
06

Data Visualization

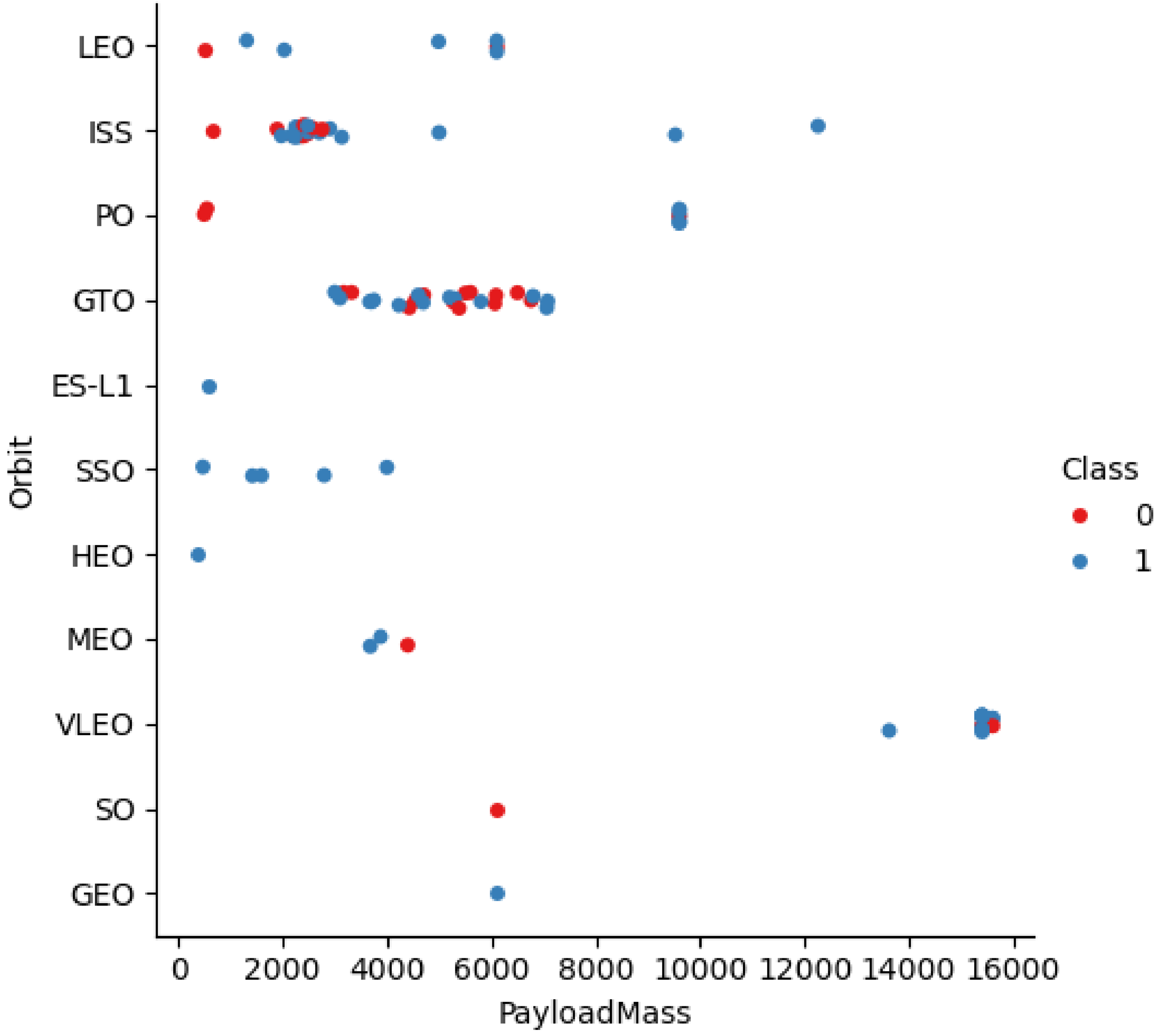


07

Data Visualization



08 Data Visualization



09

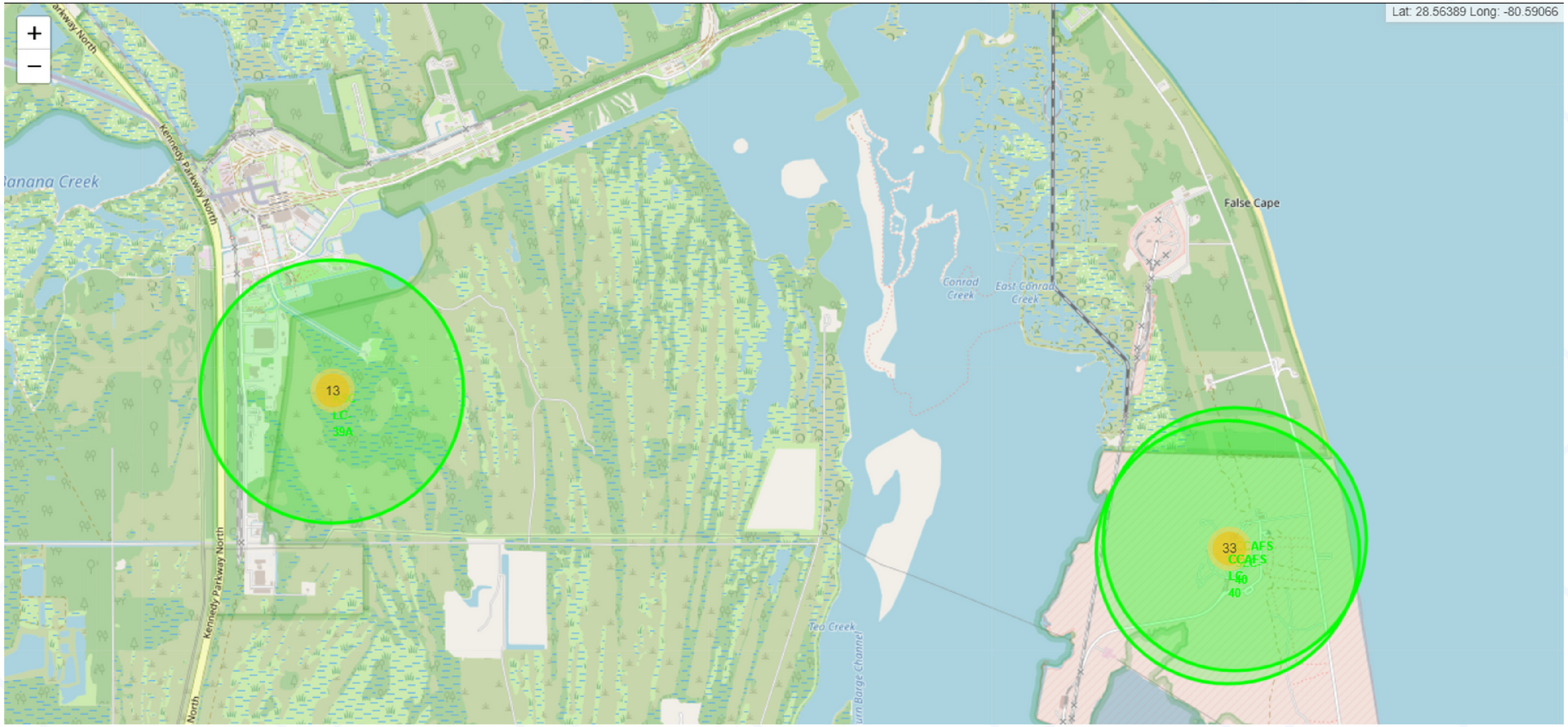
SQL Results

The following SQL queries were performed:

- Names of the unique launch sites;
- 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;
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

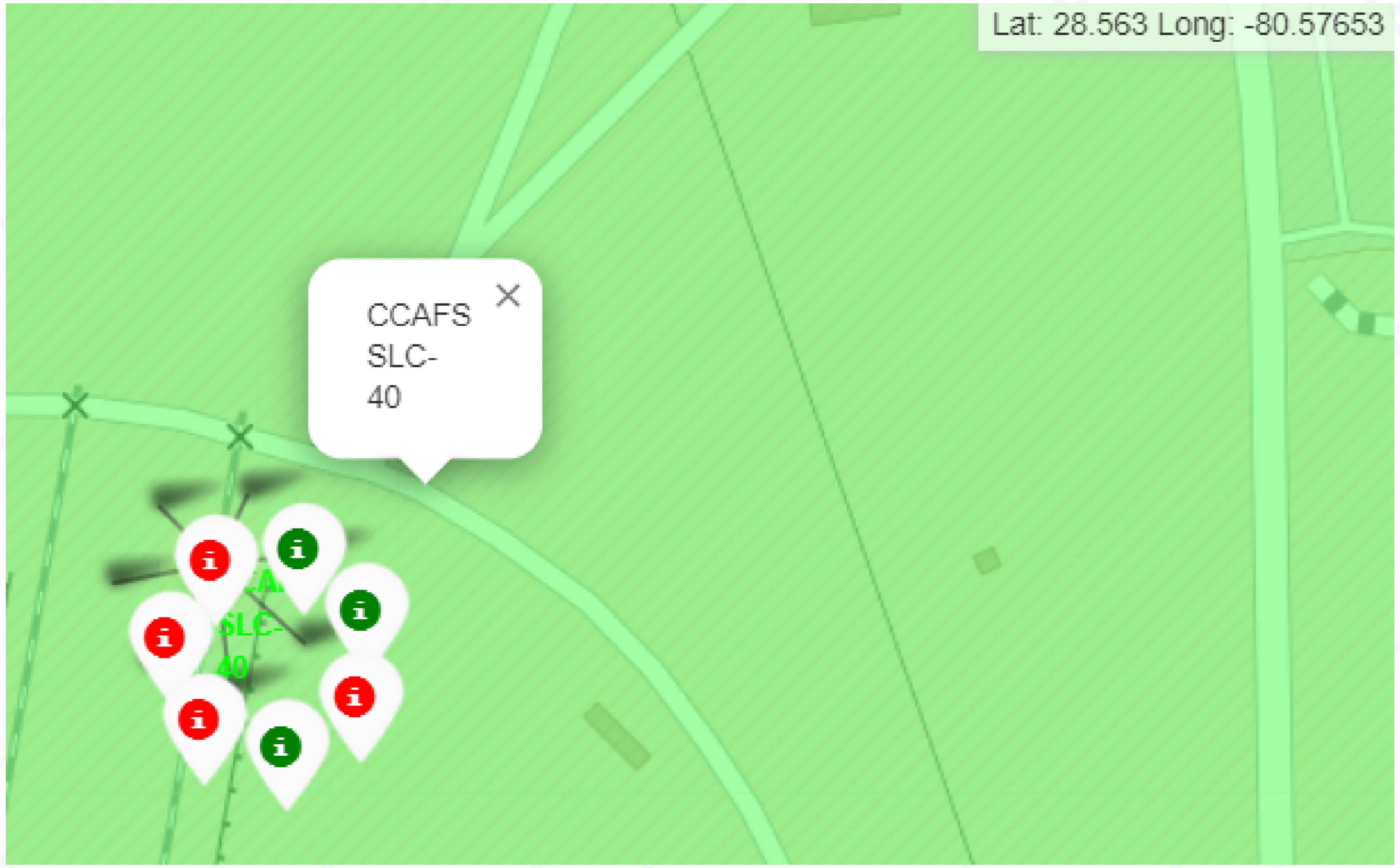
10 Map Building - Launch Sites

Folium was used to create interactive maps, which could be used to explore the various launch sites and successes/fails at each place.



11 Map Building - Launch Sites

Markers were added to explore the individual launch sites



SECTION 3
DASHBOARDS

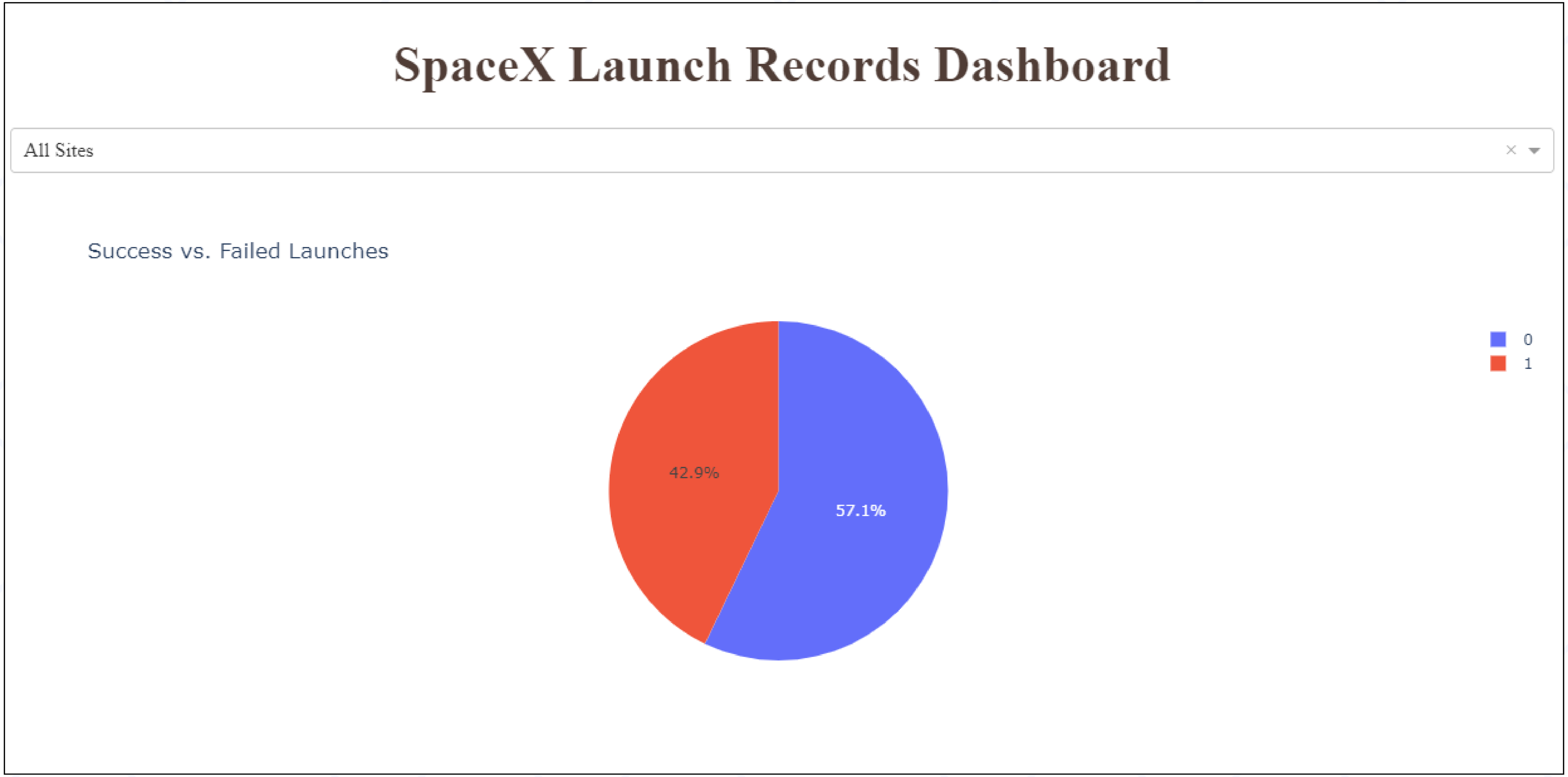
12

Interactive Dashboard

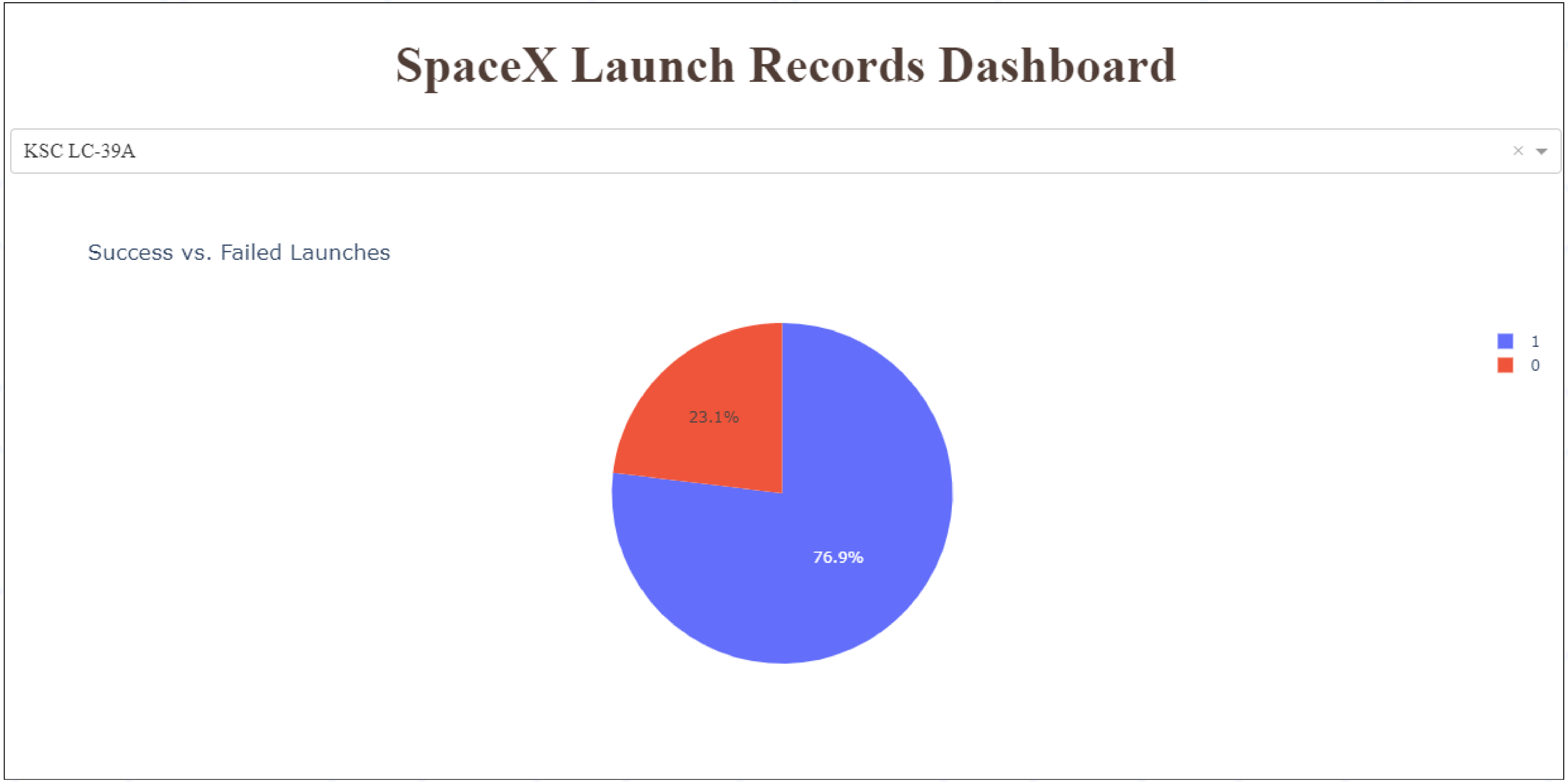
An interactive dashboard was built in Plotly (Python). This dashboard can be used to explore the data by customizing the fields and ranges selected that it could interact with. Possible interactions included:

- Success Rate by modifying the payload masses of the range
- Success Rate by modifying the launch site results

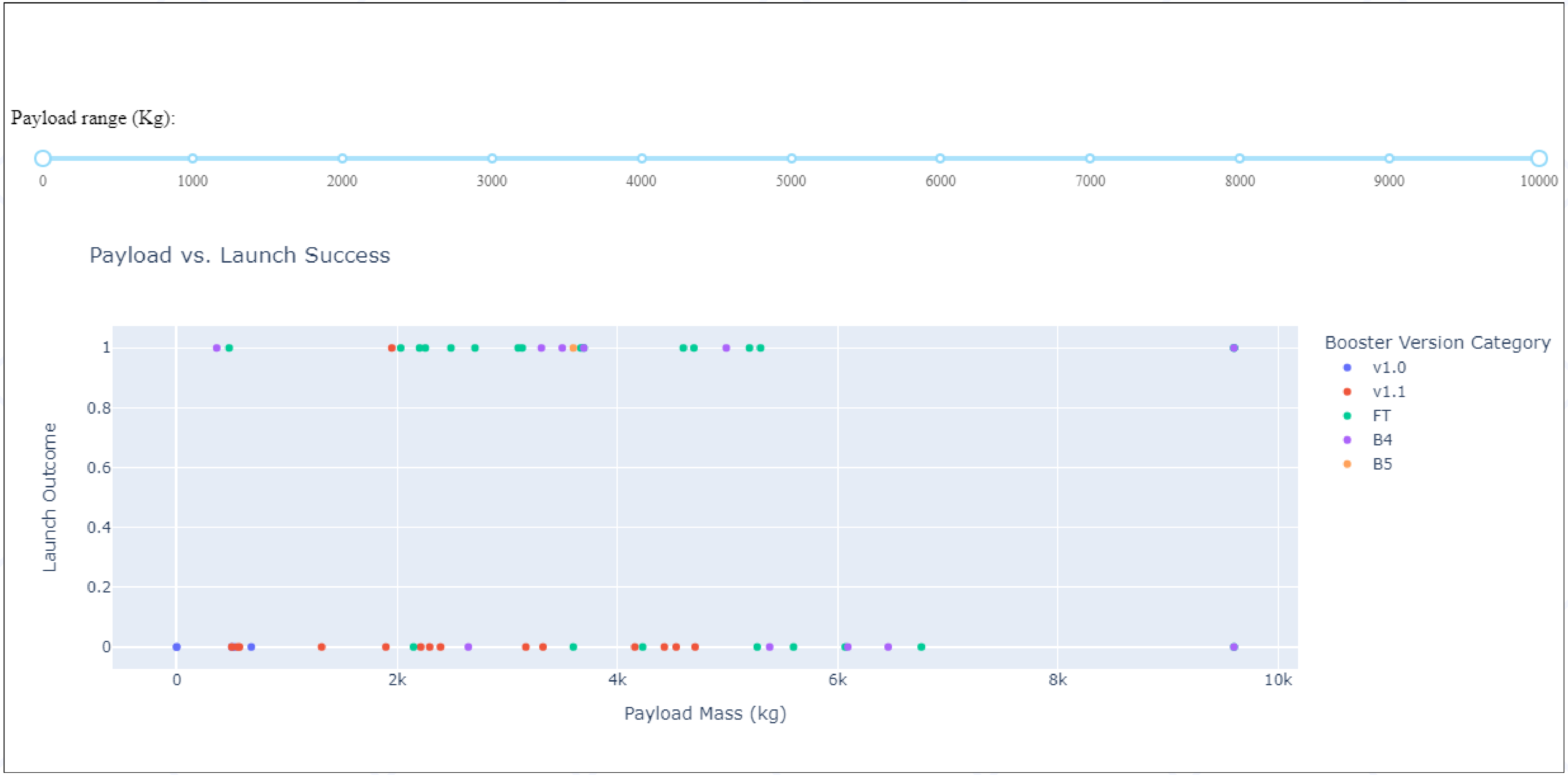
13 Interactive Dashboard - Launch



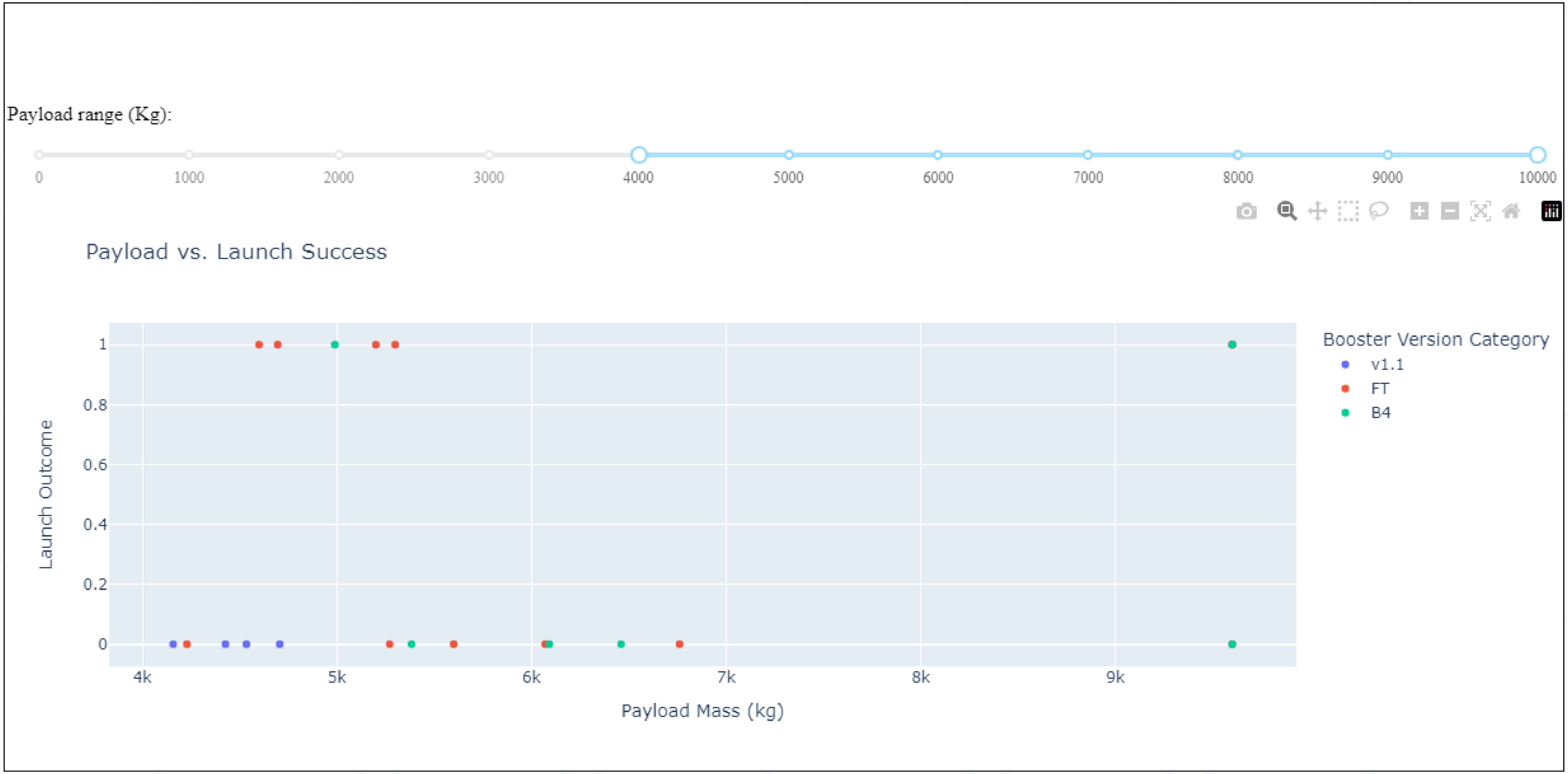
14 Interactive Dashboard - Launch



15 Interactive Dashboard - Payloads



16 Interactive Dashboard - Payloads



SECTION 4

PREDICTIVE ANALYSIS

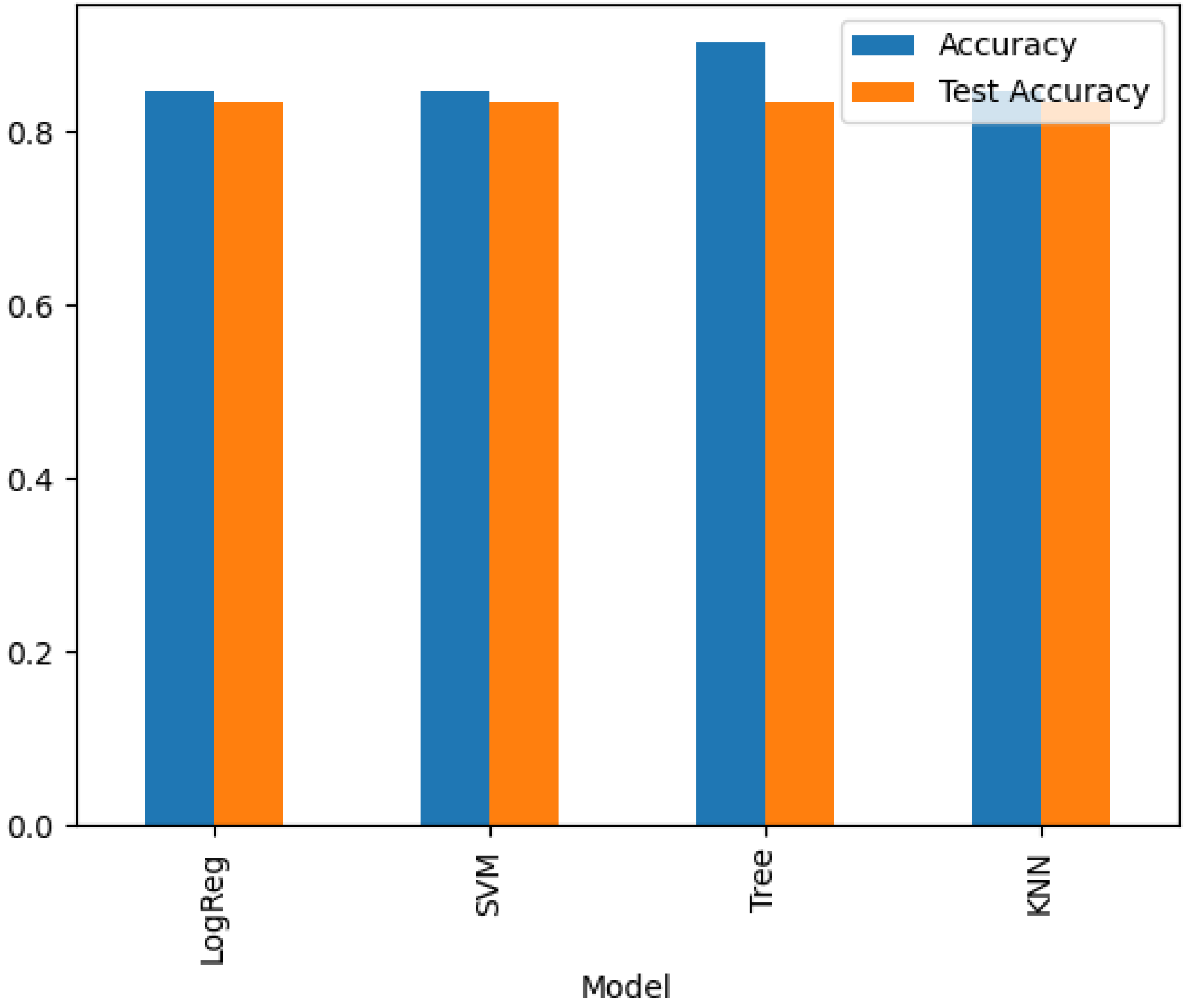
Models Chosen

- KNN
- Logistic Regression
- Decision Tree
- SVMs

18

Model Results

Performance was more or less the same for each model tried, however **Decision Trees** provided a consistently *slightly* better output than the rest.



SECTION 5

INSIGHTS & CONCLUSION

Conclusion

01 Different data sources were analyzed, refining conclusions along the process;

02 The best launch site is KSC LC-39A;

03 Launches above 7,000kg are less risky;

04 Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets

05 Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank You