Capstone Final

T'Keyah Marshall

June 2024

Outline

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

Executive Summary

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

Introduction

Rocket company Space Y would like to compete with SpaceX founded by Billionaire industrialist Allon Musk.

Objective is to determine the price of each launch by gathering information about Space X and creating dashboards for Space Y team.

Findings will also determine if SpaceX will reuse the first stage by training a machine learning model using public information to provided by SpaceX.

Section 1

Methodology

Methodology

Executive Summary

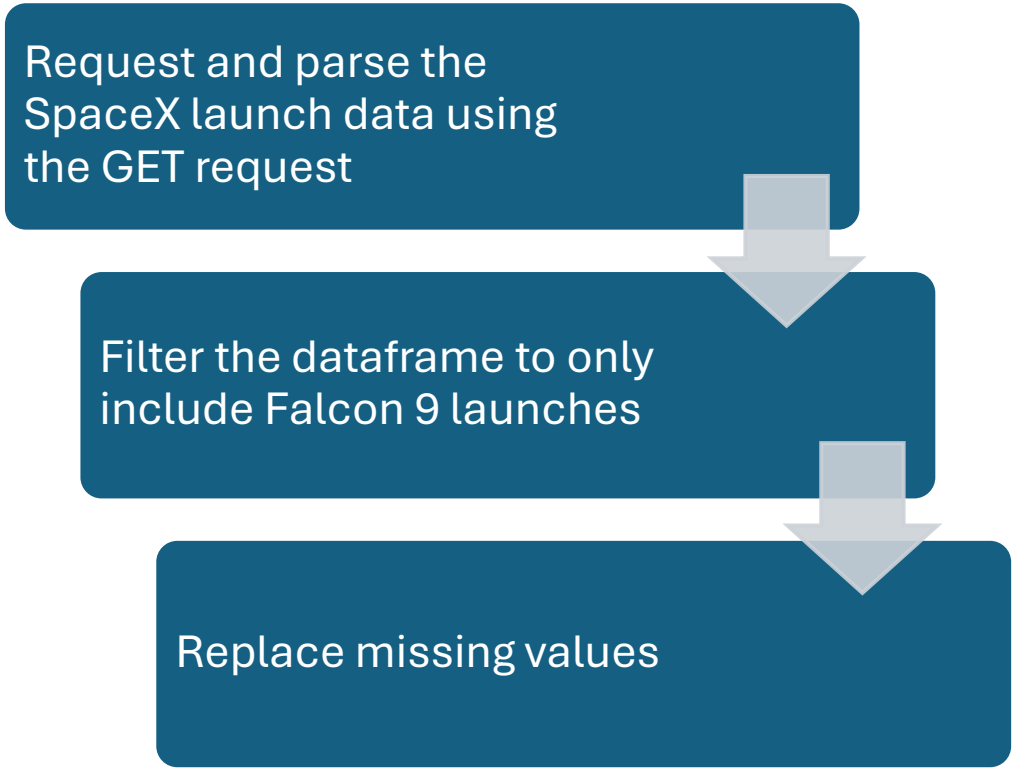
- Data collection methodology:
  - Request to the SpaceX API
  - Clean the requested data
  - Extract a Falcon 9 launch records HTML table from Wikipedia
  - Parse the table and convert it into a Pandas data frame
- Perform data wrangling
  - Exploratory Data Analysis
  - Determine Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

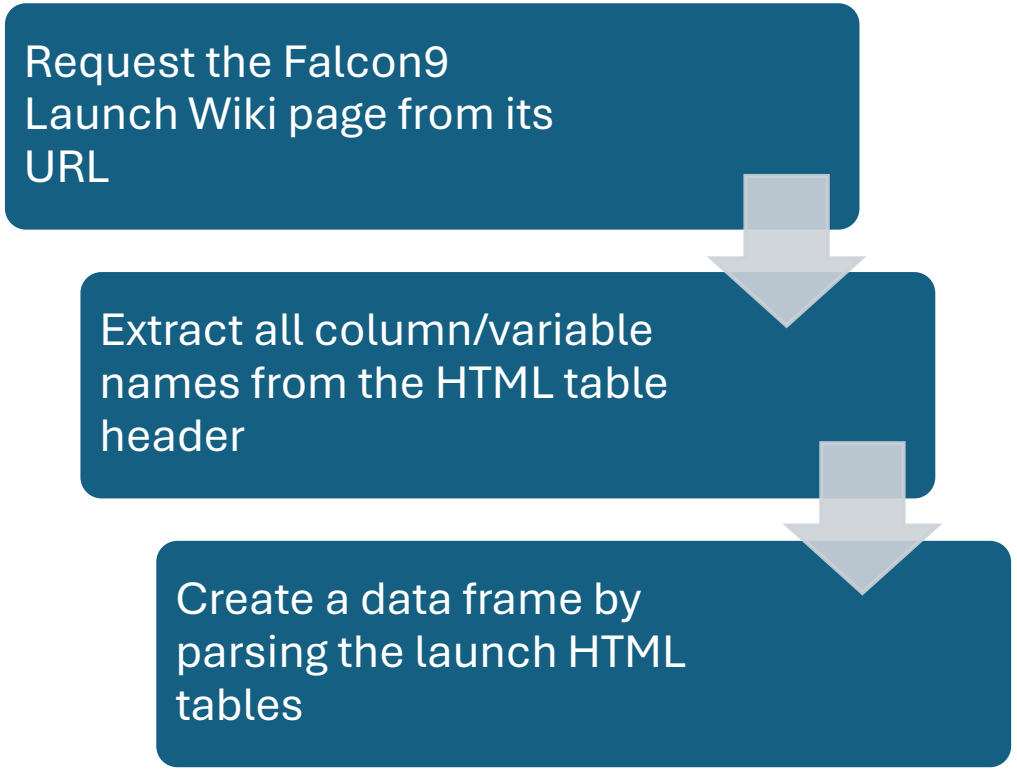
Data Collection – SpaceX API

- Define a series of helper functions that will help us use the API to extract information using identification numbers in the launch data.
- SpaceX Falcon 9 first stage Landing Prediction-Collecting the data



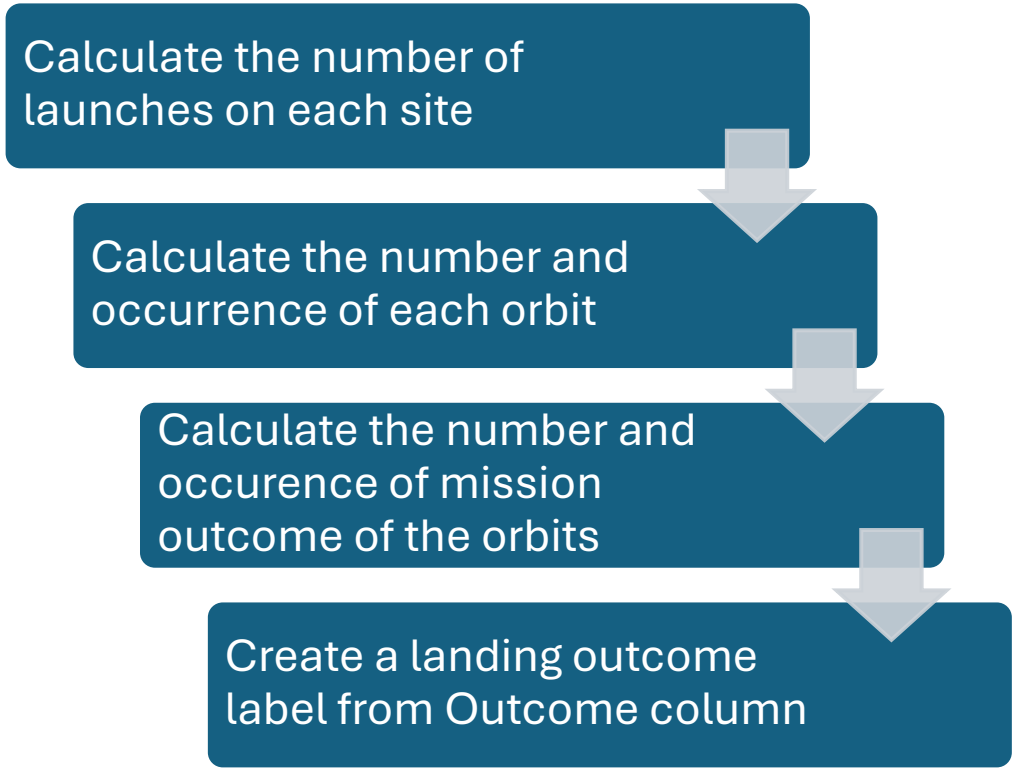
Data Collection - Scraping

- Falcon 9 historical launch records from a Wikipedia page
- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia



Data Wrangling

- Exploratory Data Analysis (EDA) to find patterns in the data and determine what would be the label for training supervised models.
- Space X Falcon 9 First Stage Landing Prediction – Data Wrangling



EDA with Data Visualization

- Scatter plot charts observe if there is any relationship between launch sites and their payload mass
- Bar charts visually check if there are any relationship between success rate and orbit type
- Line chart to to get the average launch success trend over years of launches
- SpaceX Falcon 9 First Stage Landing Prediction – Data Visualization

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

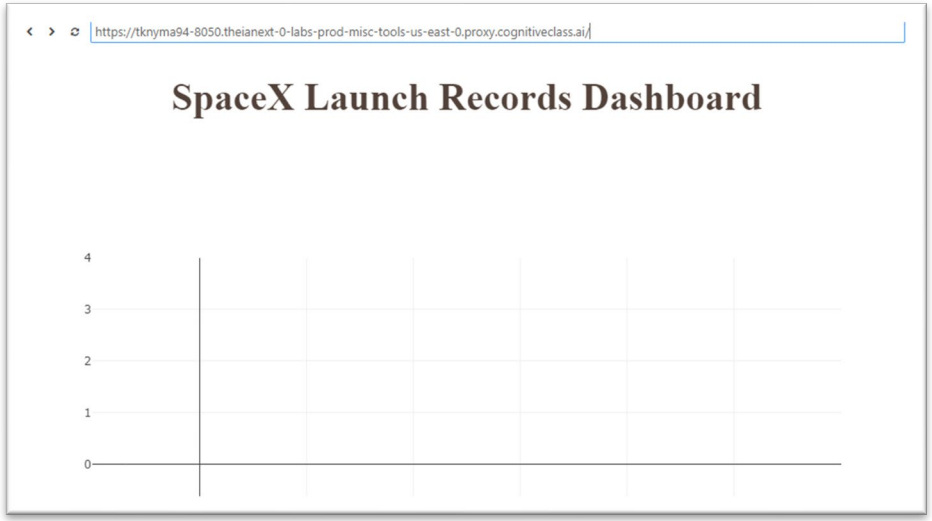
- SQL Notebook for Peer Assignment

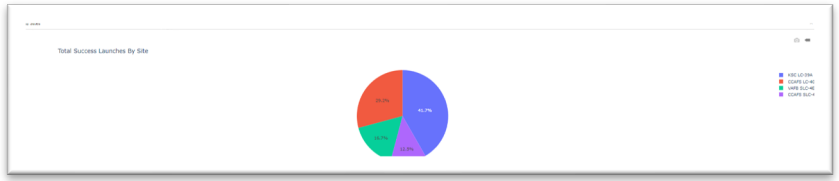
Build an Interactive Map with Folium

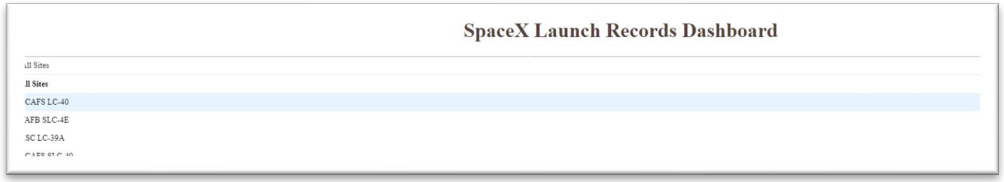
- Lines calculate the distance between the coastline point and the launch site
- Markers for all launch records, if a launch was successful, we use a green marker and if a launch was failed, we use a red marker
- Circle for each launch site in data frame launch sites, enhance the map by adding the launch outcomes for each site, and see which sites have high success rates
- Launch Sites Locations Analysis with Folium

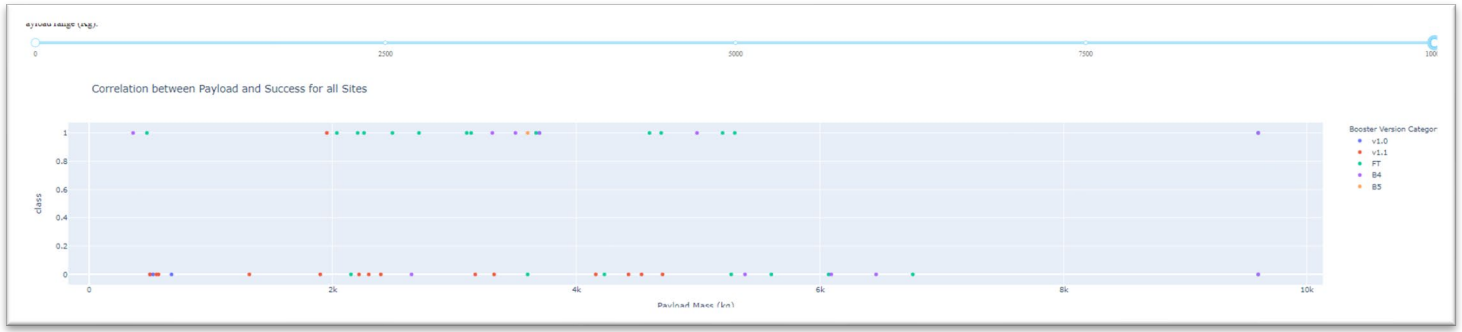
Build a Dashboard with Plotly Dash

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose







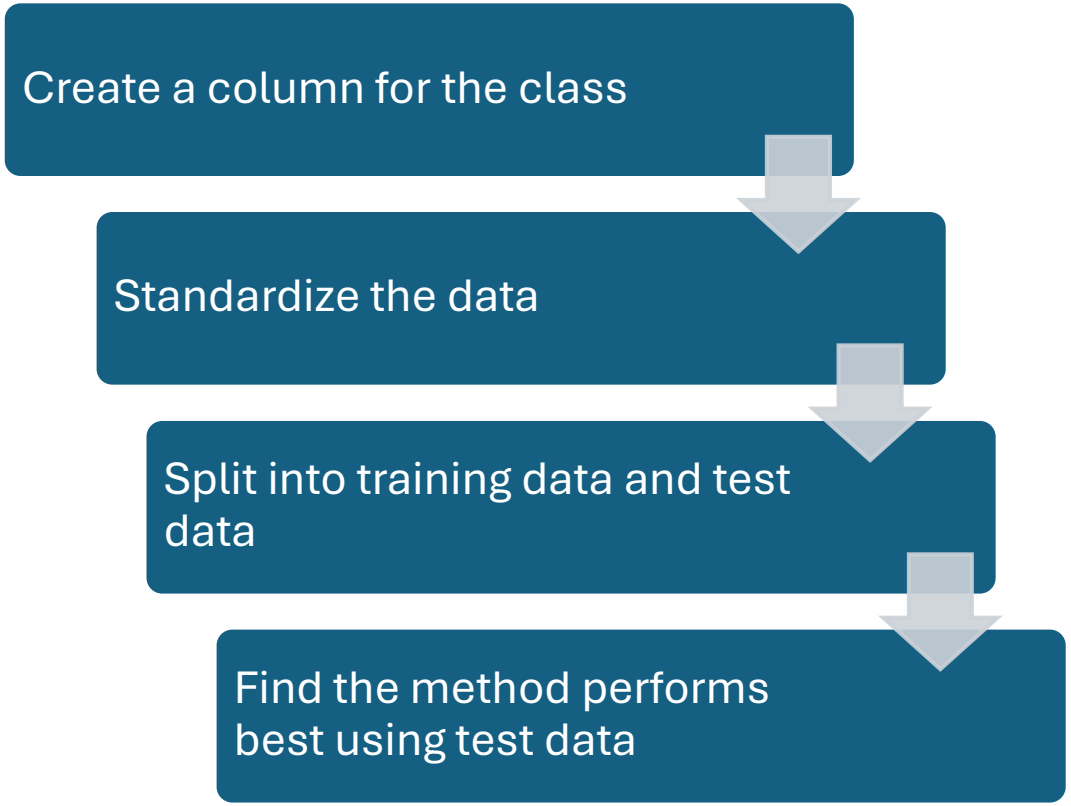


Predictive Analysis (Classification)

- Predictive Analysis Lab

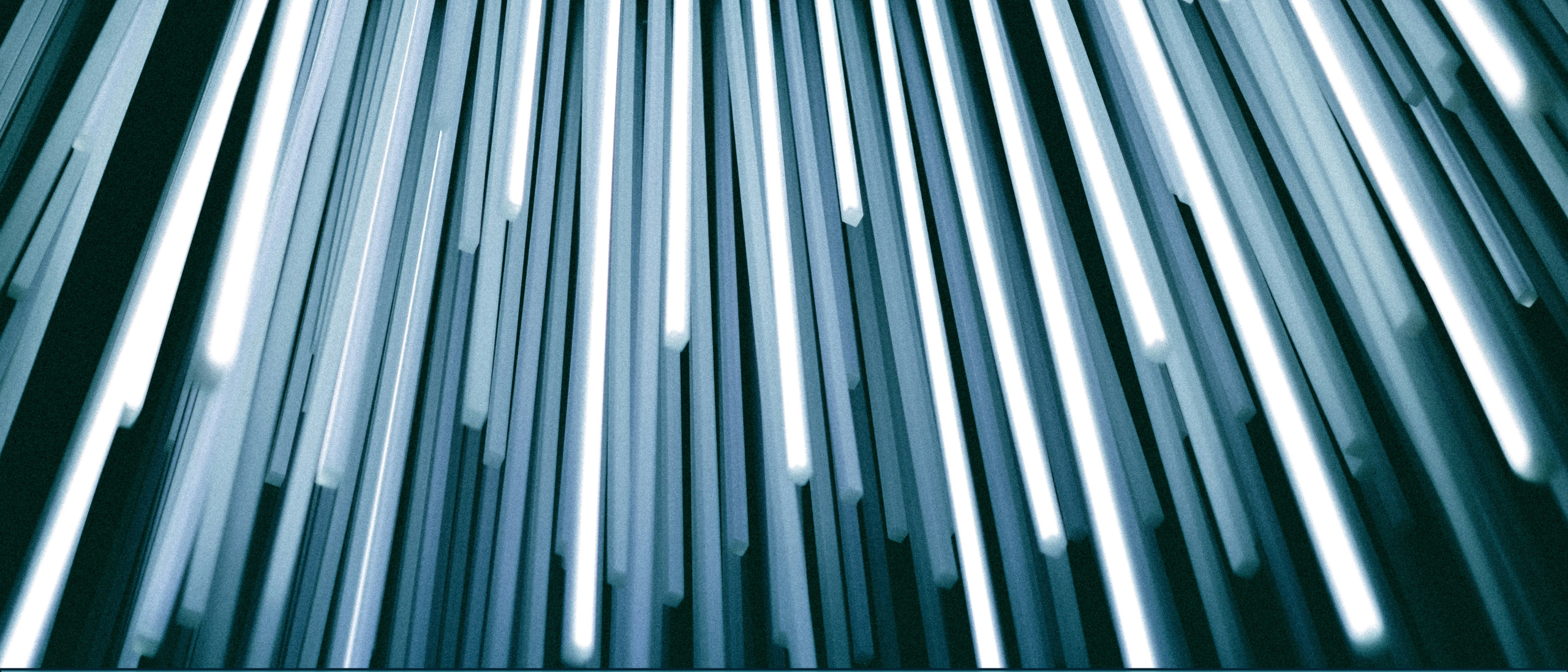
Out[68]:

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333



Results

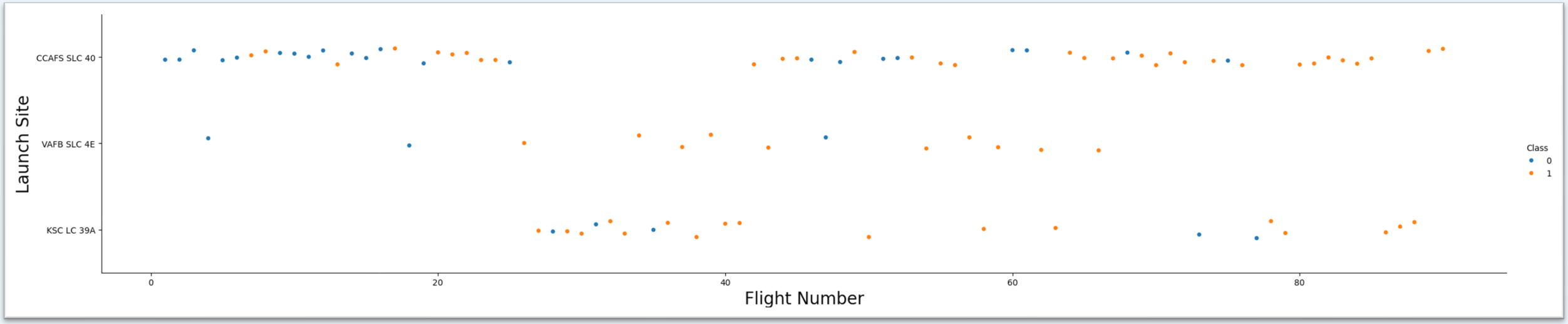
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



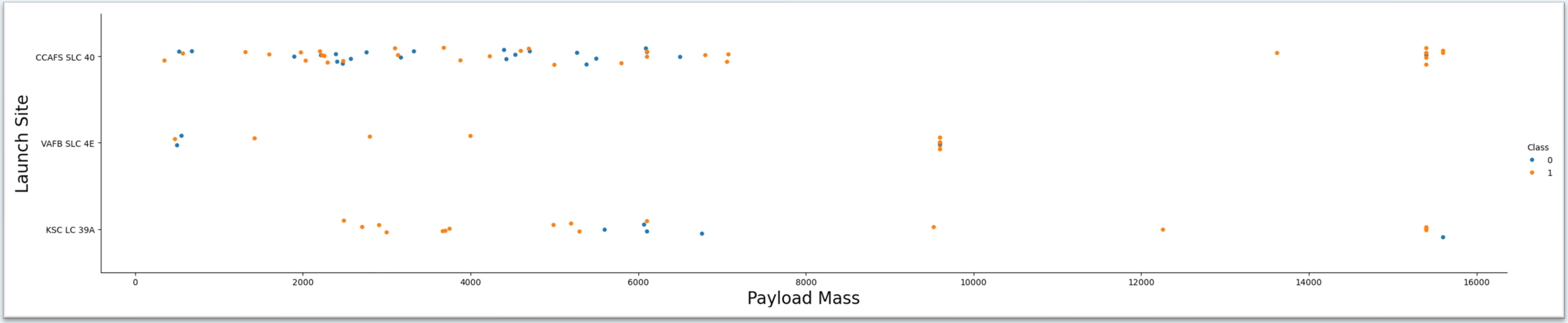
Section 2

EDA Insights

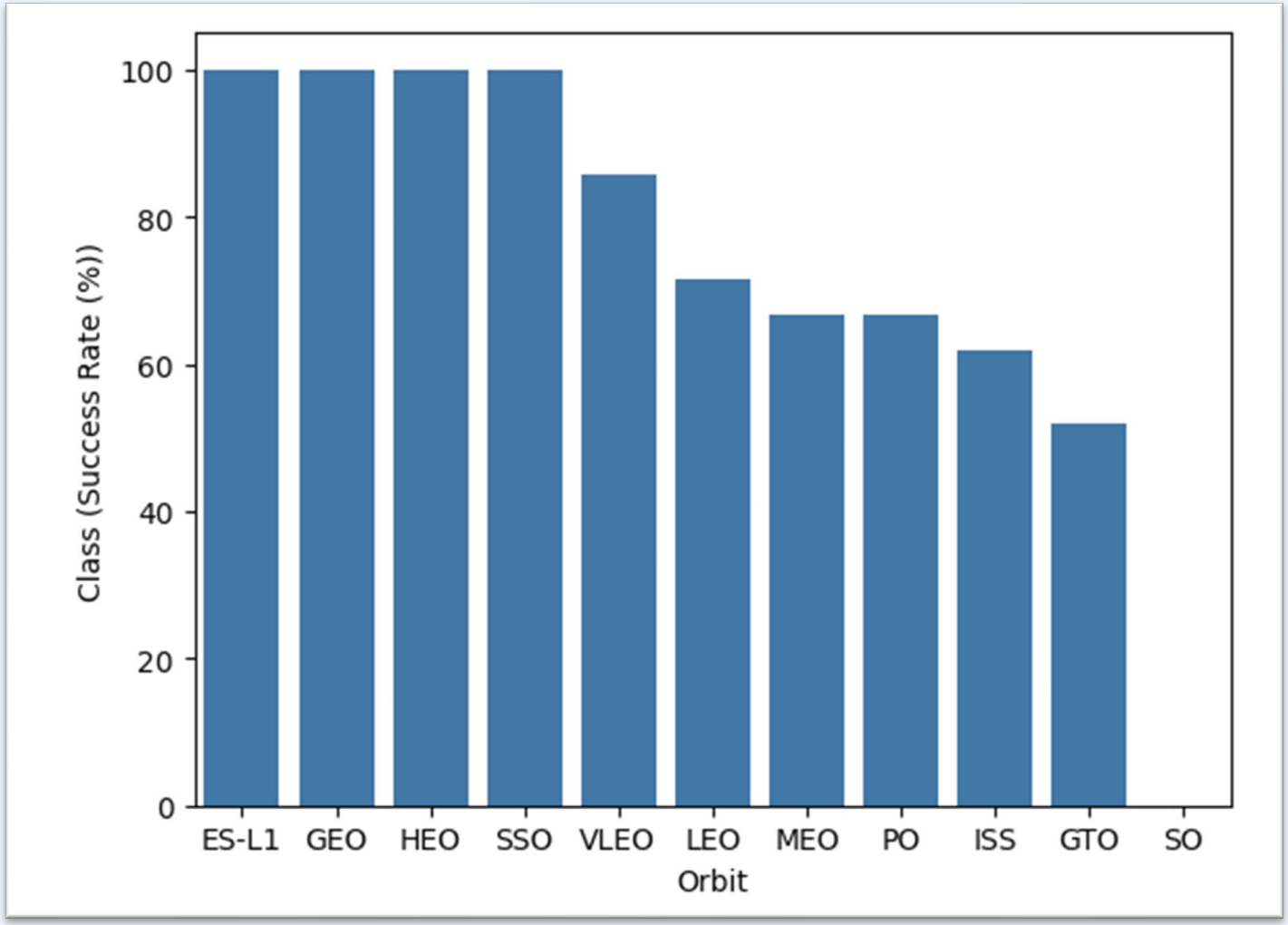
Flight Number vs. Launch Site



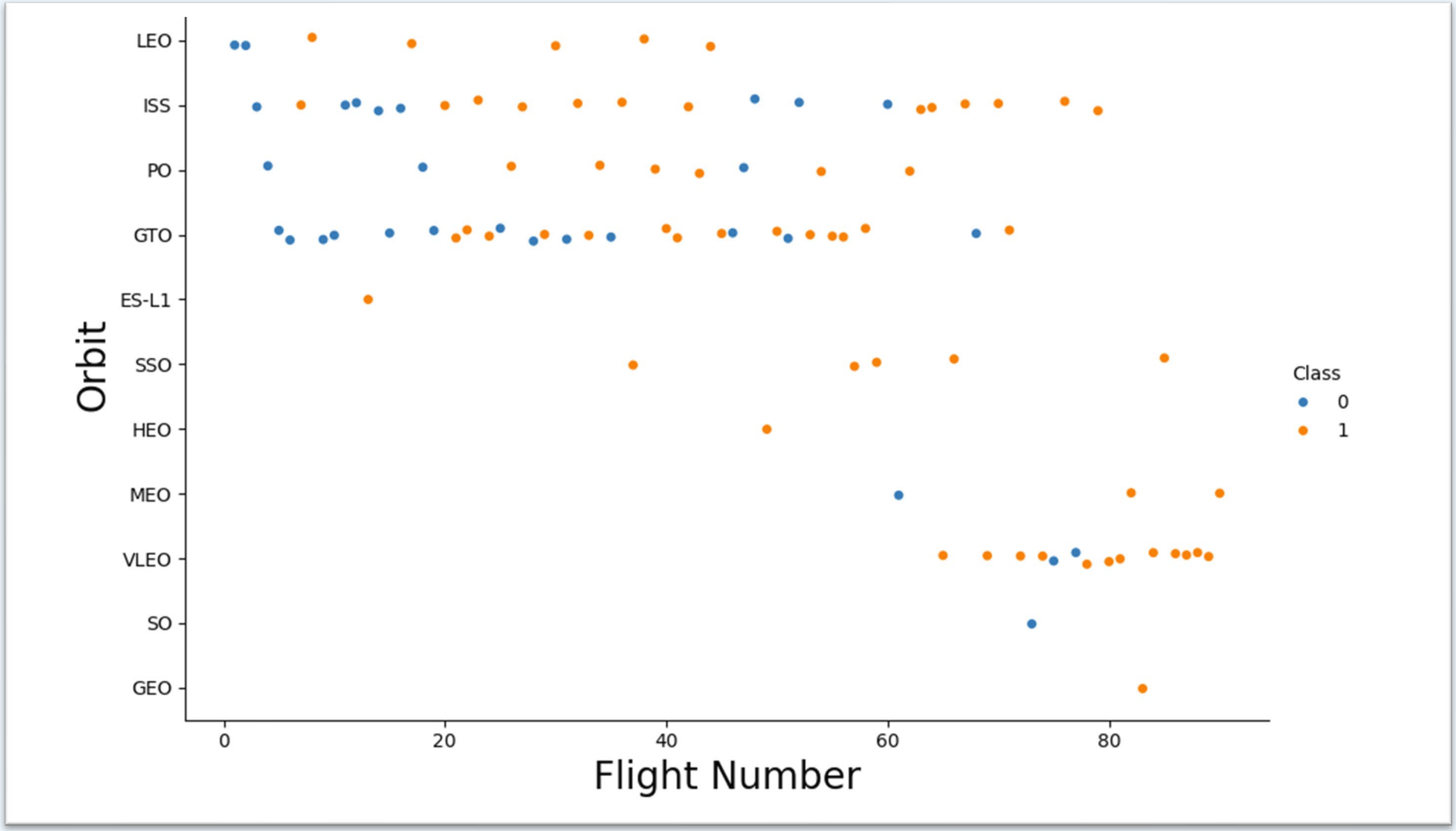
Payload vs. Launch Site



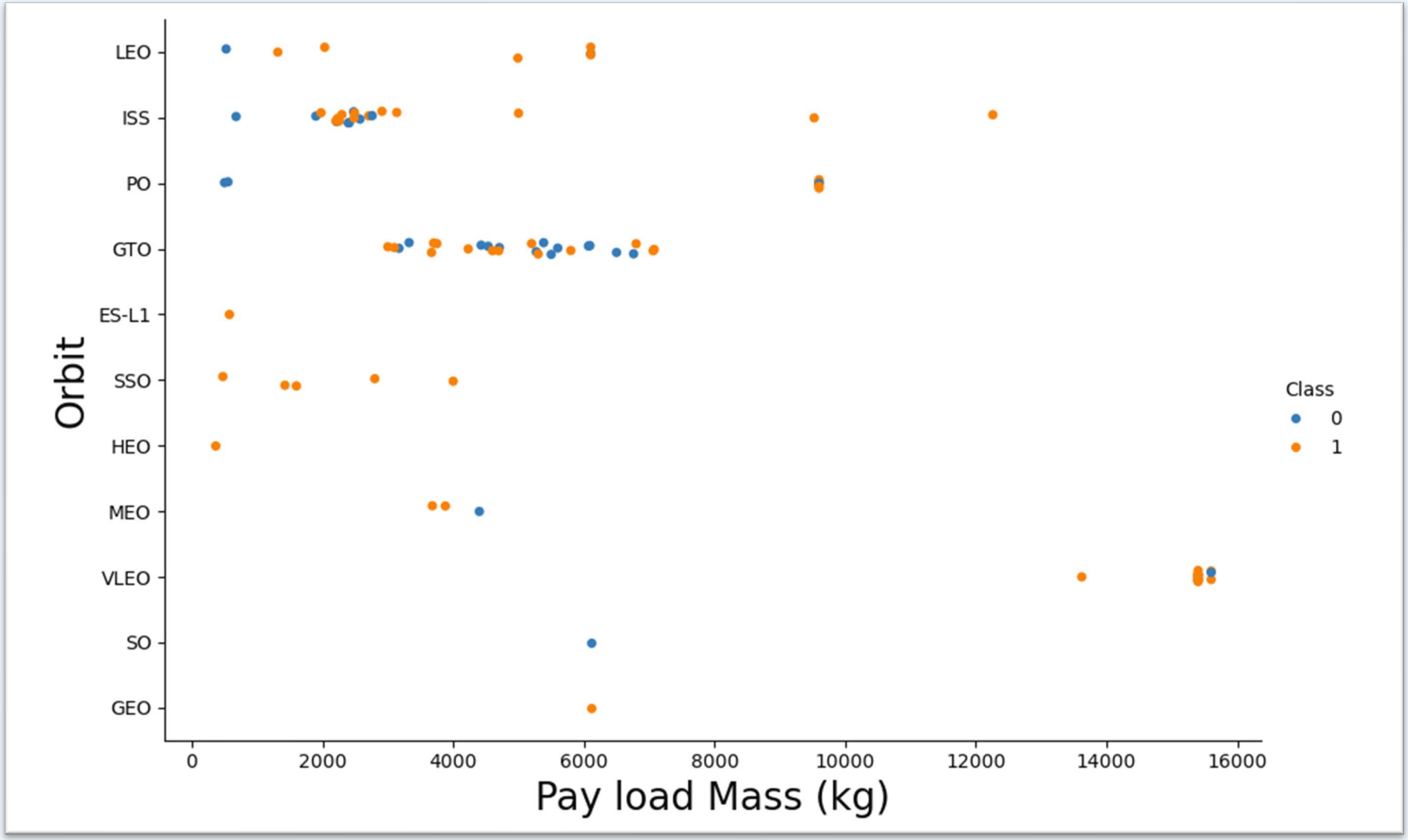
Success Rate vs. Orbit Type



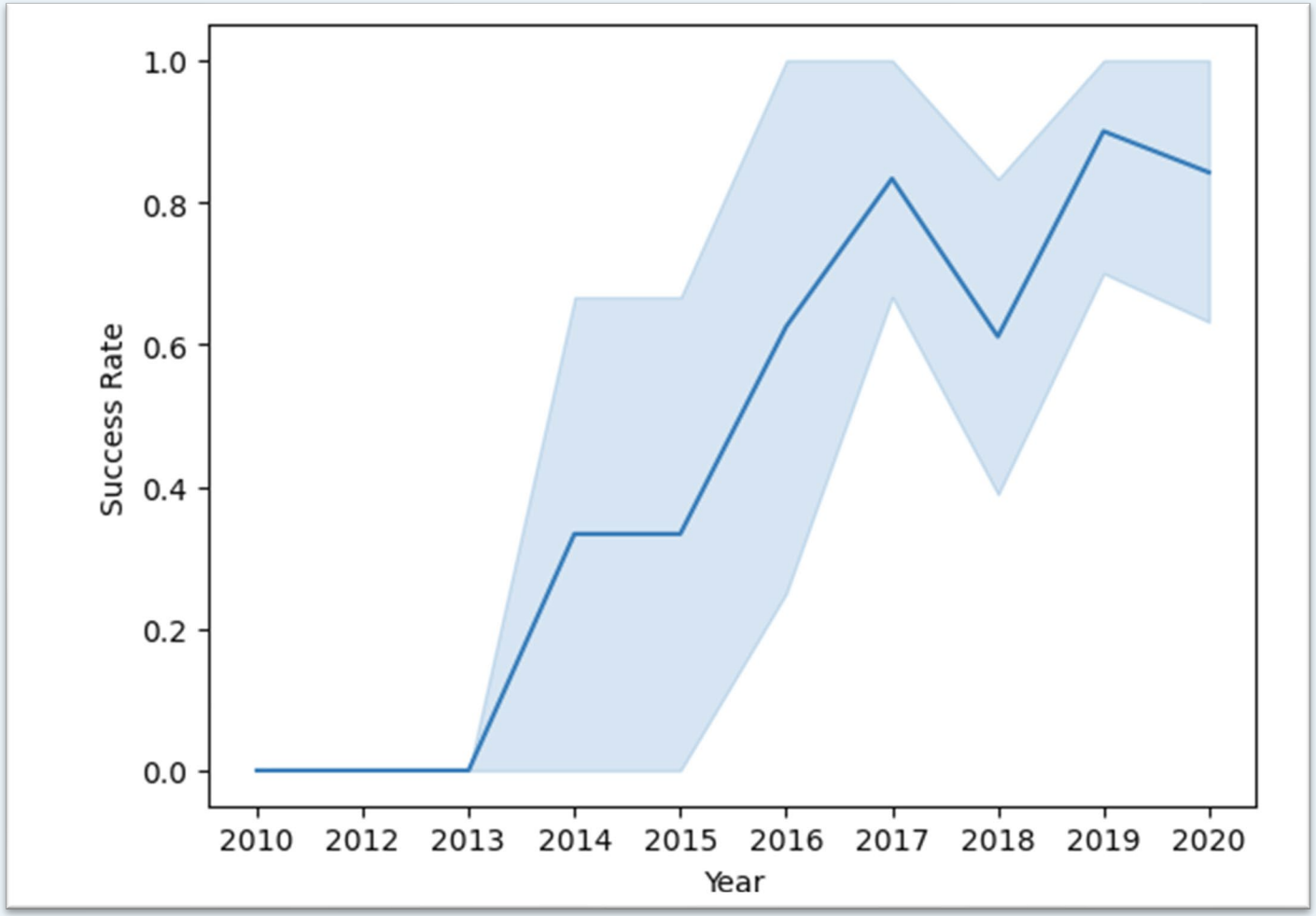
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

In [10]:
```
%sql select DISTINCT Launch_Site   from SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

Out[10]:

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

Launch Site Names Begin with 'CCA'

In [11]:
```
%sql select * from SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Out[11]:

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

Display the total payload mass carried by boosters launched by NASA (CRS)

In [12]:
```
%sql select COUNT(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[12]:

COUNT(PAYLOAD_MASS__KG_)
20

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

In [13]:
```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[13]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
In [14]: %sql select MIN(Date) from SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)' LIMIT 1
 * sqlite:///my_data1.db
Done.
Out[14]: MIN(Date)
         2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

In [15]:
```
%sql select Booster_Version from SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND
```

```
 * sqlite:///my_data1.db
Done.
```

Out[15]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

```
In [16]: %sql select COUNT(Landing_Outcome) from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
Out[16]: COUNT(Landing_Outcome)
                              101
```

Boosters Carried Maximum Payload

```
In [17]: %sql select DISTINCT Booster_Version from SPACEXTABLE WHERE (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
 * sqlite:///my_data1.db
Done.
```

Out[17]:

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014
F9 v1.1 B1015
F9 v1.1 B1016
F9 v1.1 B1018
F9 FT B1019
F9 v1.1 B1017
F9 FT B1020
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1024
F9 FT B1025.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1031.1
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 FT B1034
F9 FT B1035.1
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1037
F9 B4 B1039.1
F9 FT B1038.1
F9 B4 B1040.1
F9 B4 B1041.1
F9 FT B1031.2

2015 Launch Records

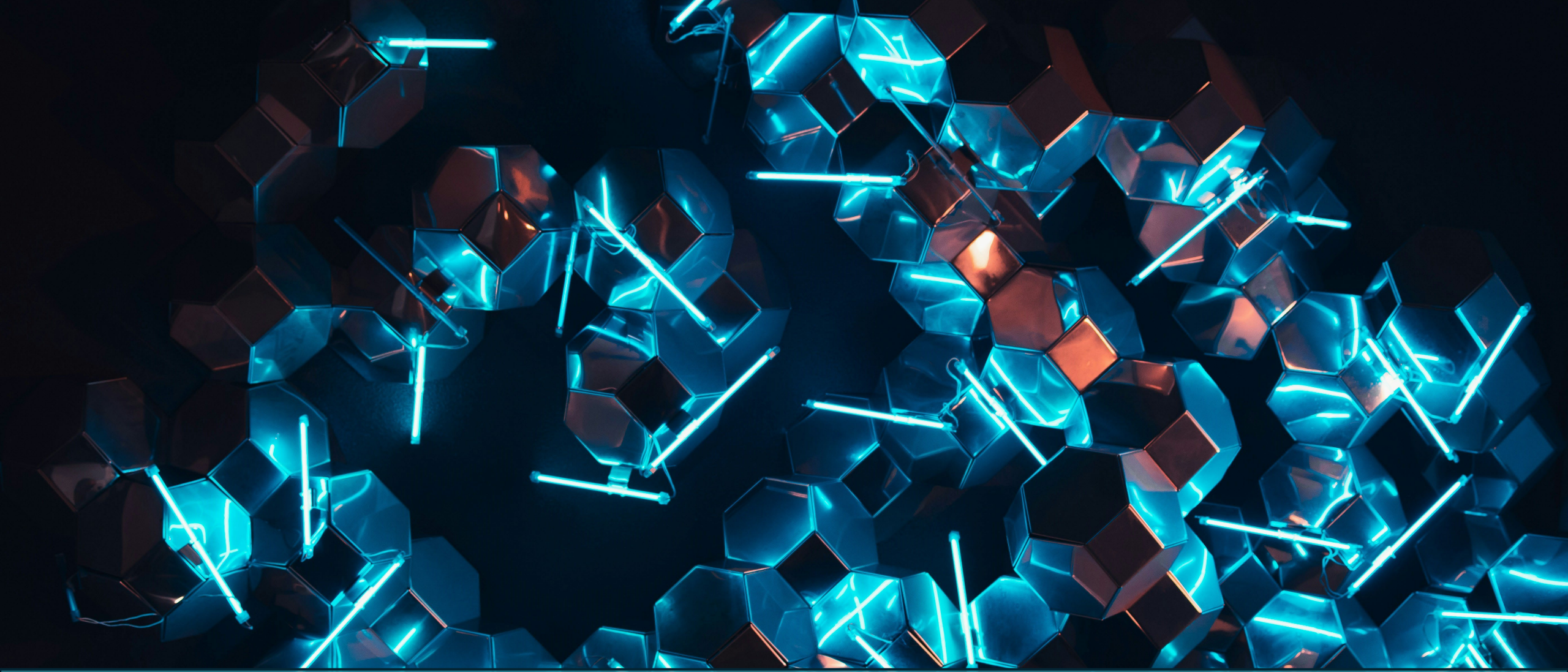
```
In [29]: %sql select substr(Date, 6,2) as month, substr(Date, 0,5) as year, Landing_Outcome, Booster_Version, Launch_Site from SPACEX
 * sqlite:///my_data1.db
Done.
```

Out[29]:

month	year	Landing_Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

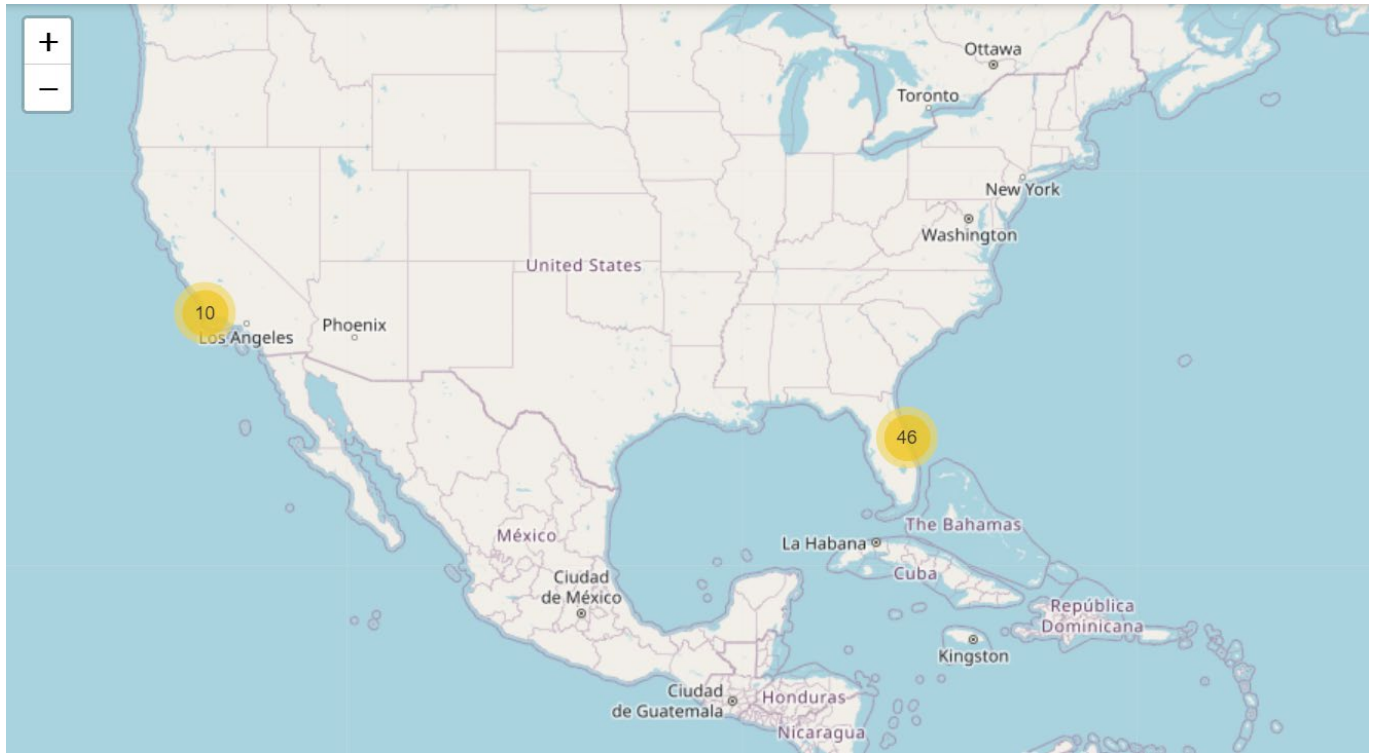
```
In [31]: %sql select Landing_Outcome from SPACEXTABLE GROUP BY Landing_Outcome WHERE
```



Section 3

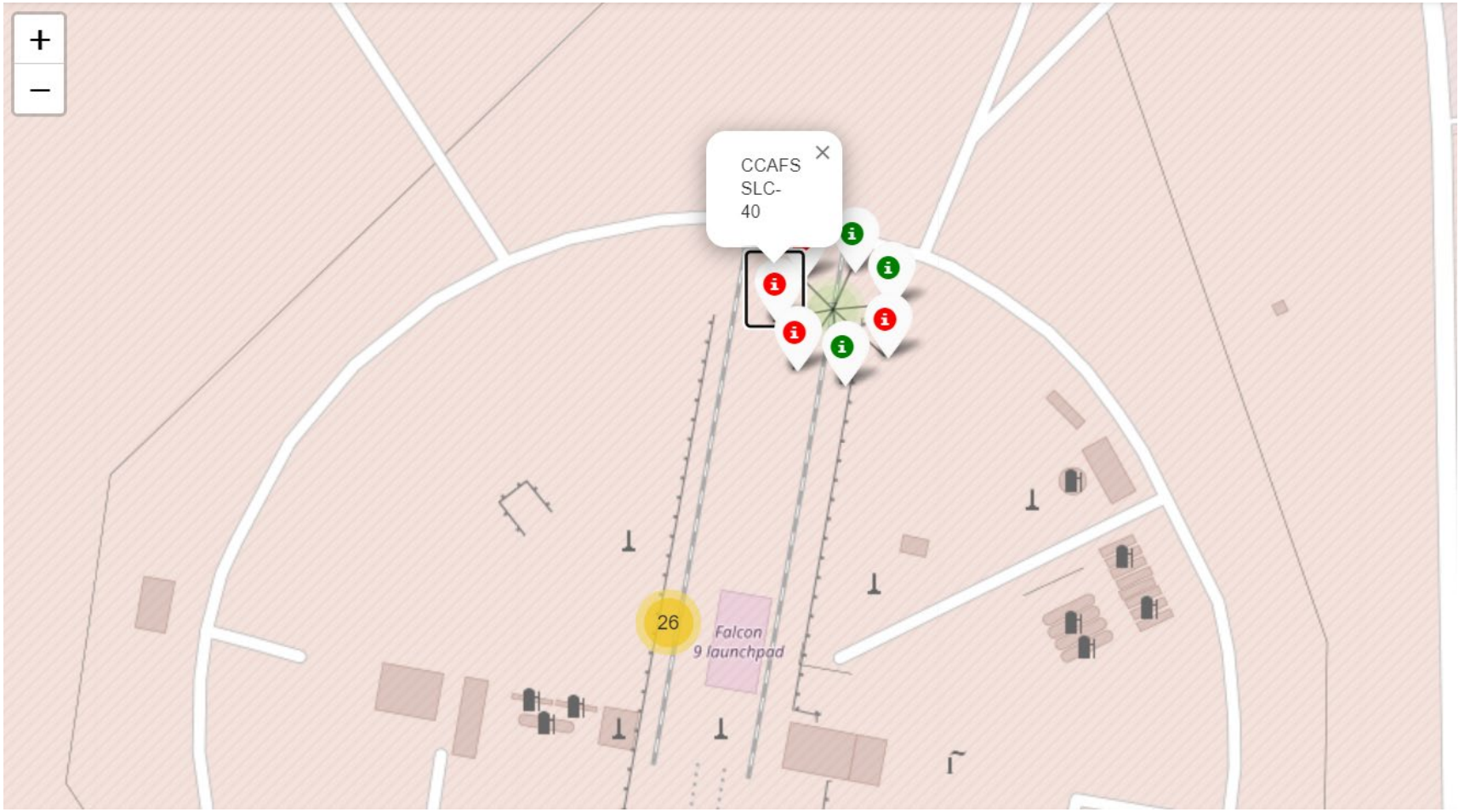
Launch Sites Proximities Analysis

Launch Site Locations



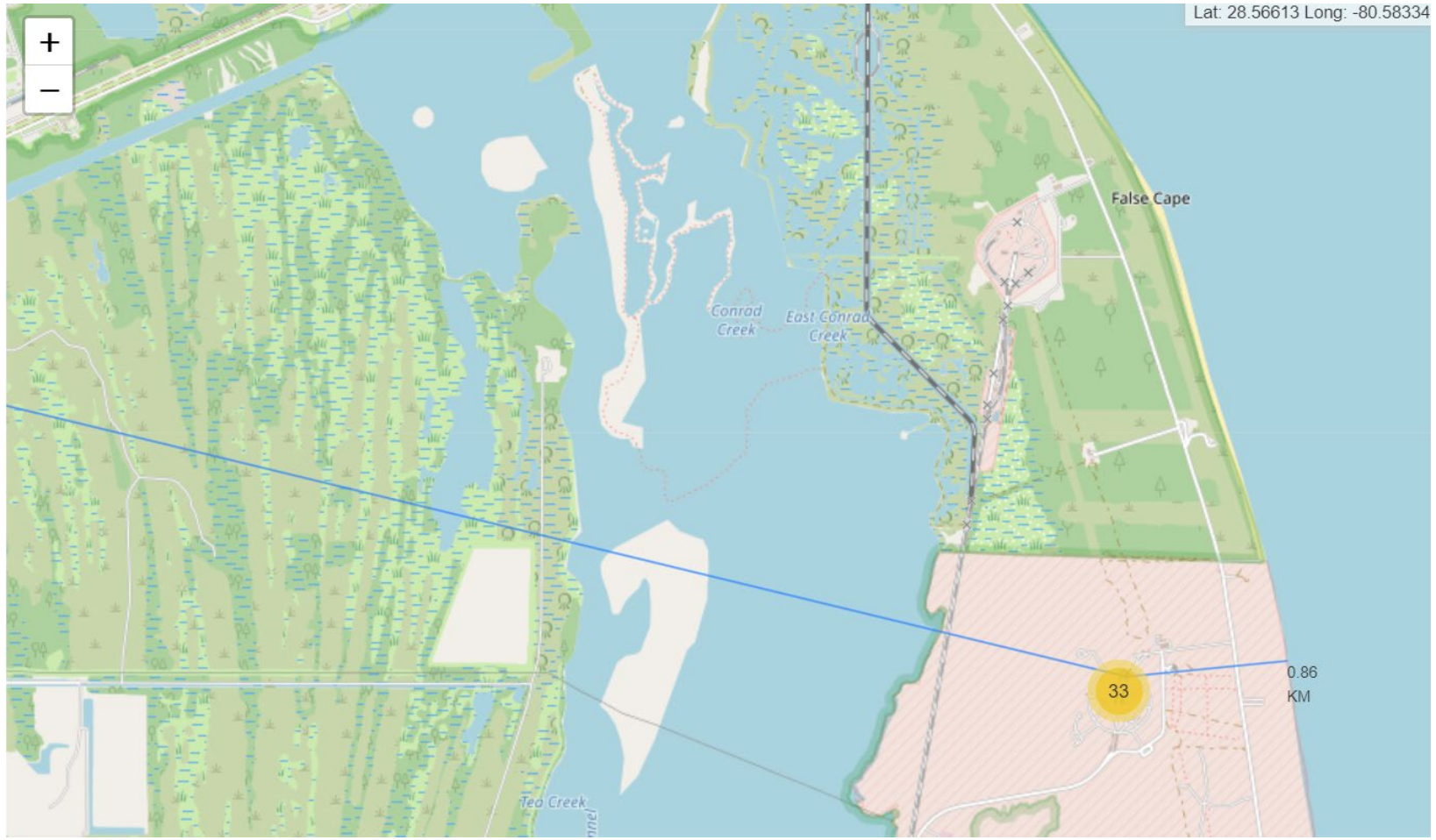
Launch only happens in one of the four launch sites, which means many launch records will have the exact same coordinate. Marker clusters can be a good way to simplify a map containing many markers having the same coordinate.

Marker Clusters/Launch Success Rates

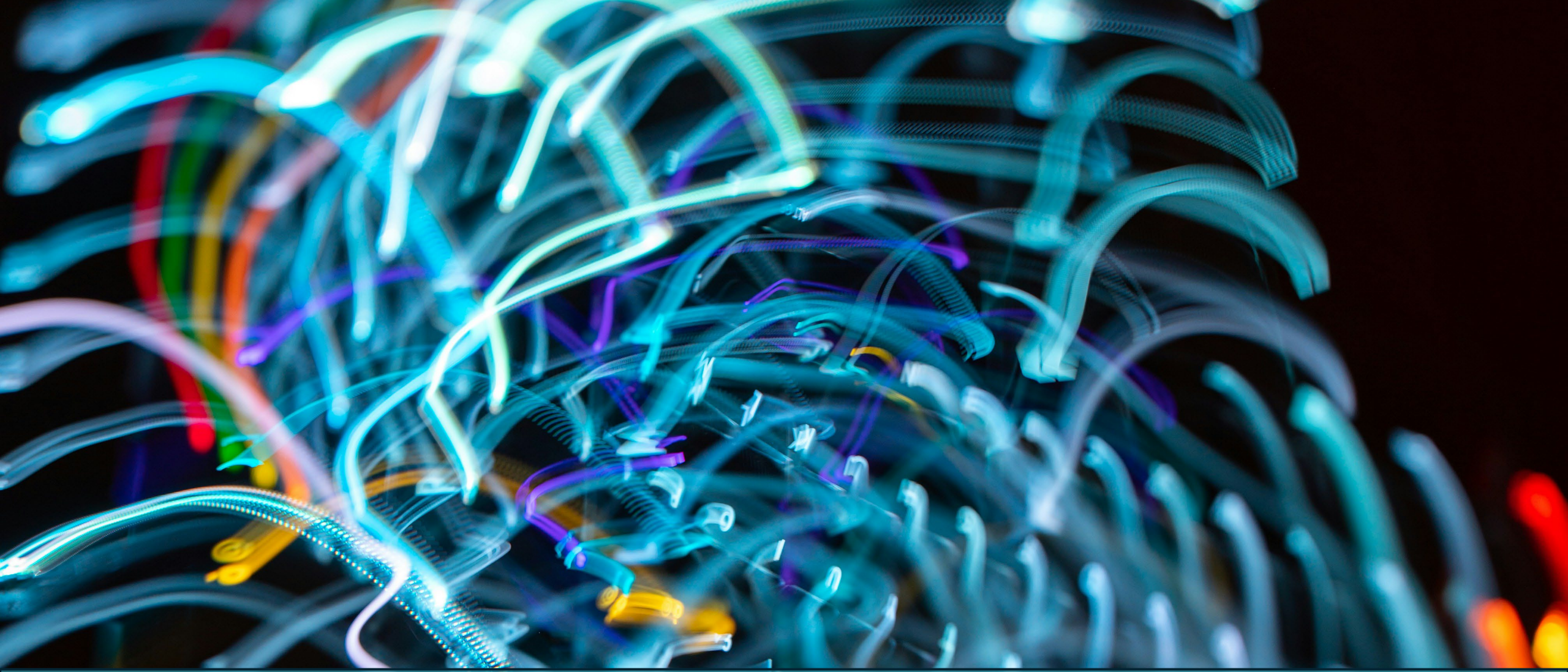


From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

Launch Site Proximities



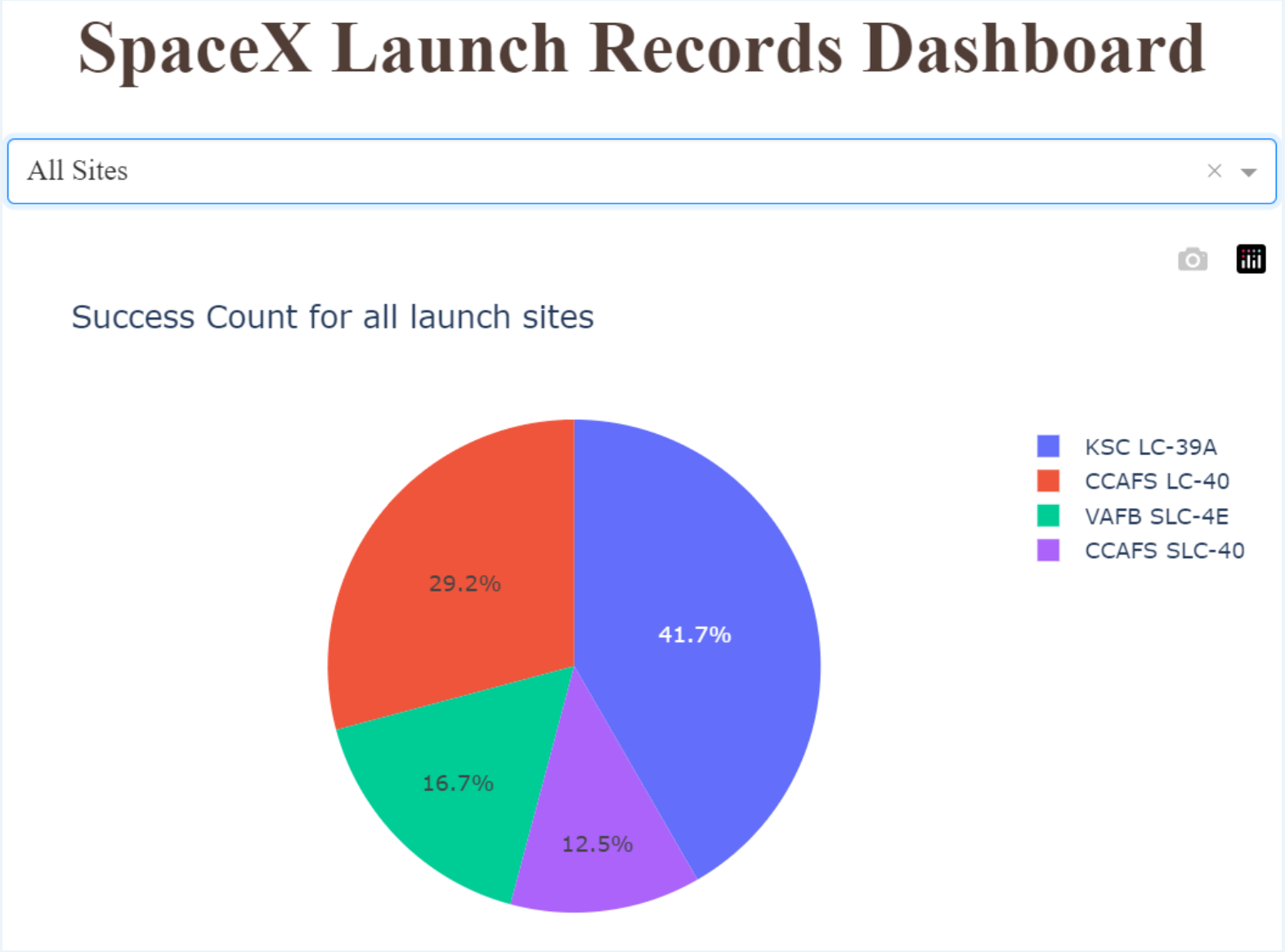
Zoom in to a launch site and explore its proximity to see if you can easily find any railway, highway, coastline, etc.



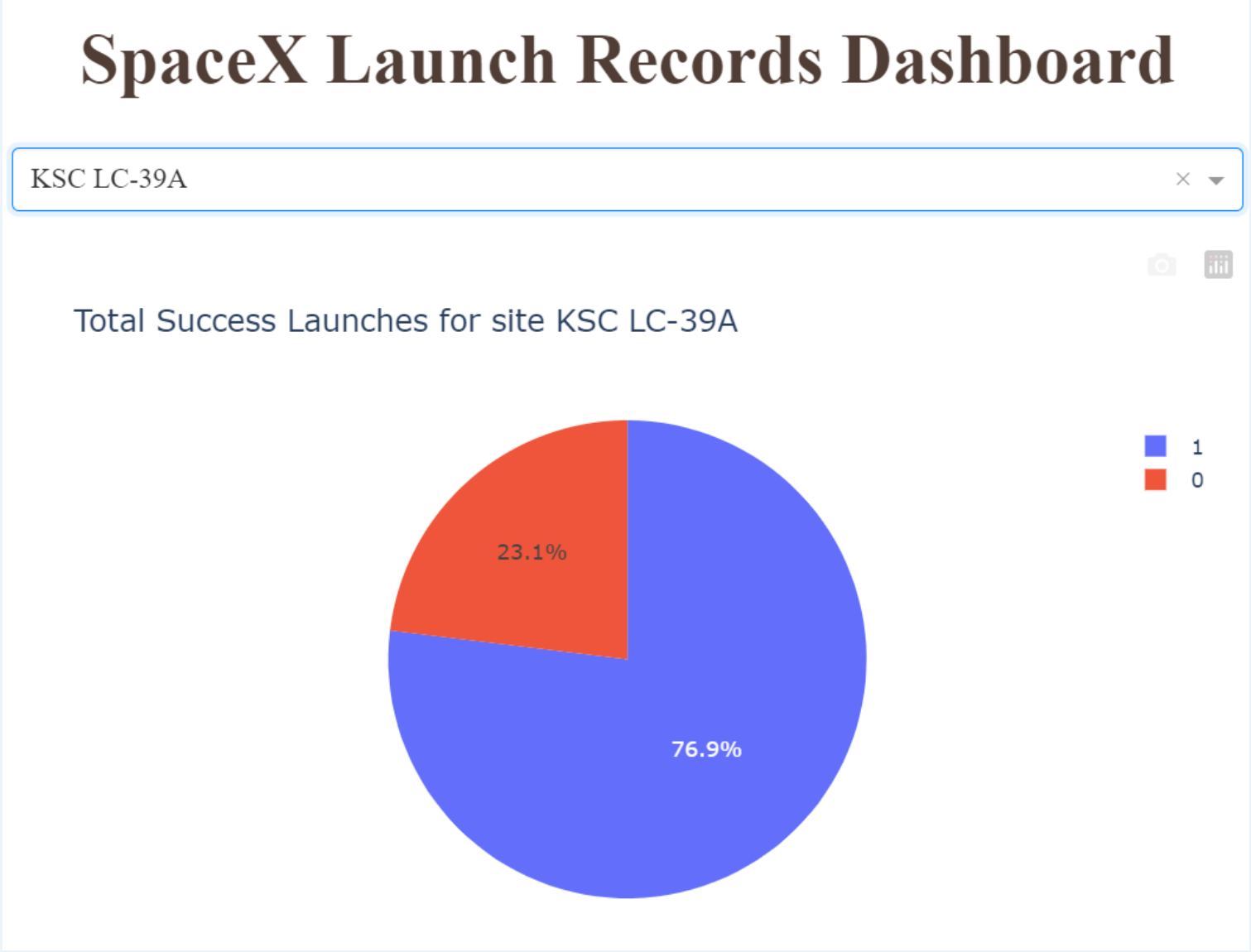
Section 4

Build a Dashboard with Plotly Dash

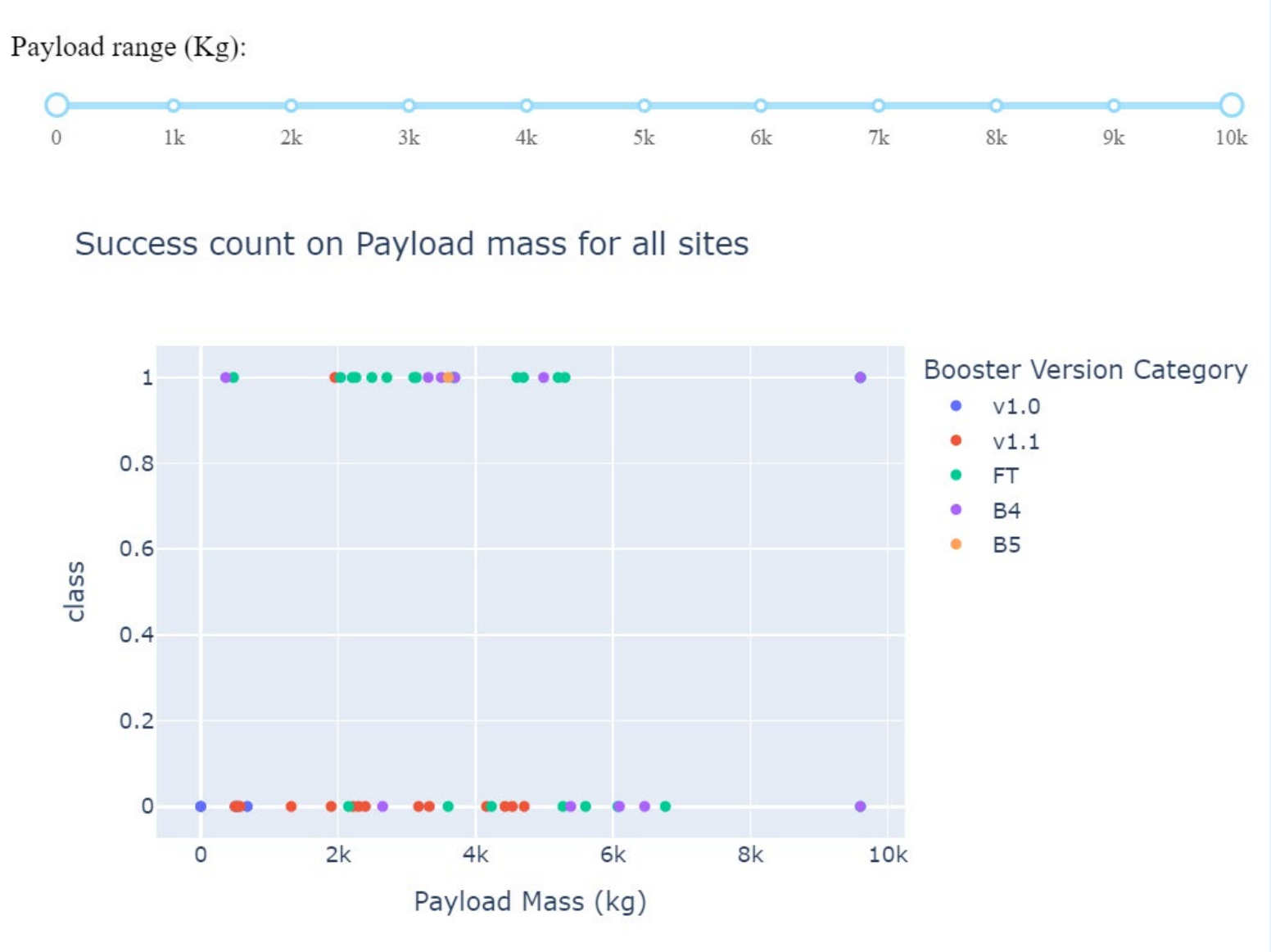
Successful Launch Count



Highest Launch Success Ratio



Payload vs Launch Outcome



Section 5

Predictive Analysis (Classification)

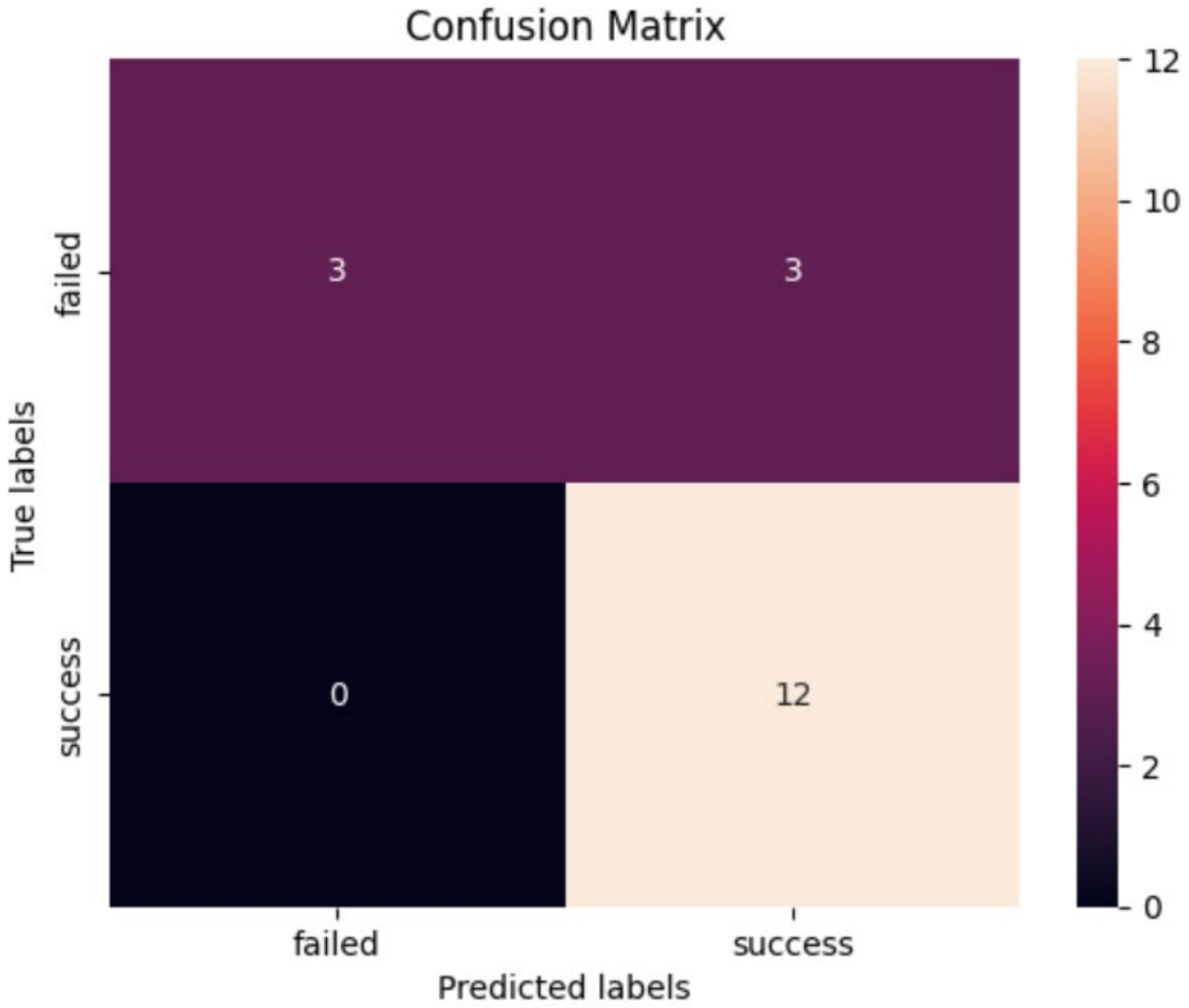
Classification Accuracy

```
[50]  print("Logistic Regression test data accuracy :",logreg_cv.score(X_test, Y_test))
                                                                                    Python
```

```
...   Logistic Regression test data accuracy : 0.8333333333333334
```

Confusion Matrix

...



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

Appendix

- https://github.com/tkeyahnyree/capstonerepo

Thank you