IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- The process of prediction of Spacex Falcon9 land went through data collection, data wrangling, visualization and Machine learning algorithms.

- The SpaceX is gradually succeeding in its mission. The success of mission is dependent on booster versions, payload mass, launch site and some other factors which will be touched upon on in the presentation.

Introduction

- In this final project, successful landing of Falcon 9 first stage will be evaluated. 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 we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- Various factors affecting the landing should be evaluated such as payload mass, launch sites, orbits in order to determine the successful landing. Machine learning algorithms needs to be developed for this task.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is gathered from SpaceX REST API and Wikipedia Pages through Web Scraping
- Perform data wrangling
  - Wrangling data using an API, sampling data and dealing with Null values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The machine learning model will use public information to determine if SpaceX reuse first stage

Data Collection

Data which are public information gathered from different sources:

SpaceX REST API: The API shares information about launches, payload, landing specification and outcomes.

Web scraping of Wikipedia pages which has data about Falcon 9 launch data

Data Collection – SpaceX API

Data collection with SpaceX REST

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

SpaceX Rest API for getting past launches data

```
url = "https://api.spacexdata.com/v4/launches/past"
```

Obtaining the launch data using request library

```
response = requests.get(url)
response.json()
```

Converting Json objects to Dataframe

```
data = pd.json_normalize(response.json())
```

Data Collection - Scraping

Web scraping process

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Performing an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_
response = requests.get(static_url).text
```

Create a BeautifulSoup object from the HTML response

```
soup = BeautifulSoup(response, 'html5lib')
```

Finding the tables, creating list of column names and assigning to dictionary

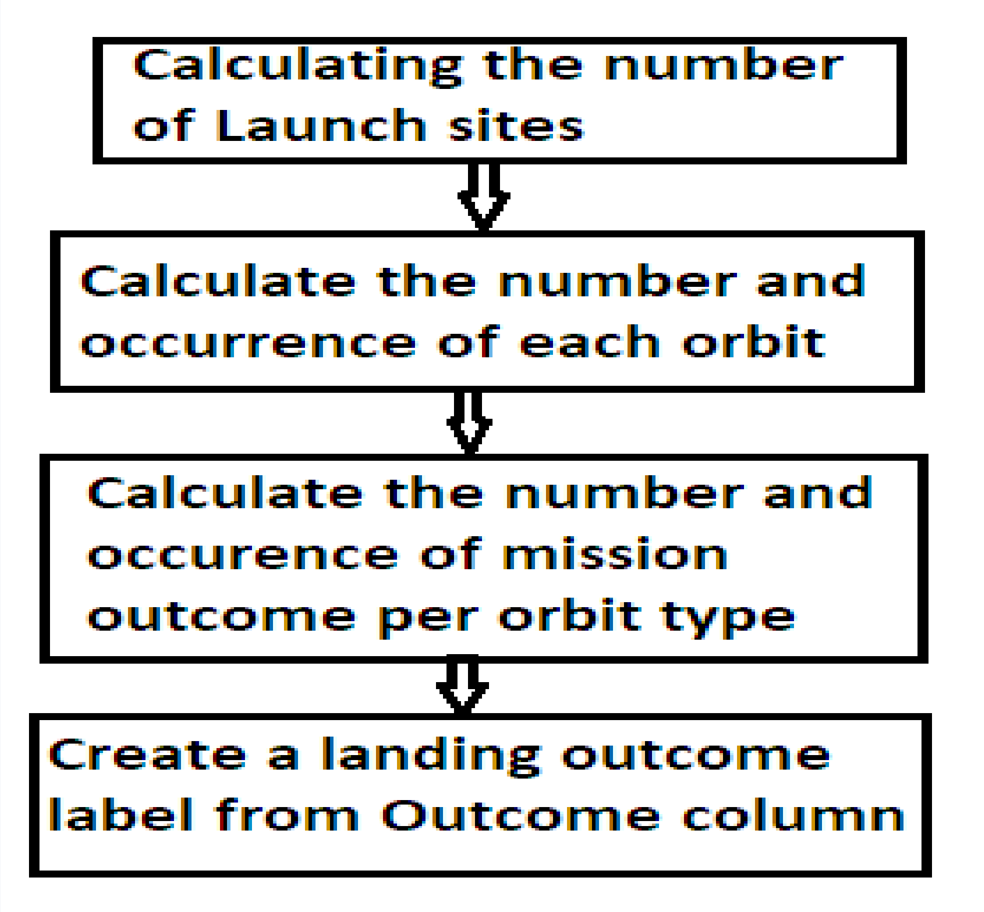
```
html_tables = soup.find_all('table')
first_table = soup.find_all('th')
launch_dict= dict.fromkeys(column_names)
```

Creating DataFrame from dictionary

```
df=pd.DataFrame(launch_dict)
```

Data Wrangling

Data wrangling process



https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Scatter point charts, bar plot and line chart plotted in order to see the following results:

- Relationship between landing and payload mass
- Relationship between landing and launch sites
- Relationship between launch sites and payload mass
- Relationship between success rate and orbit types
- Relationship between Payload and Orbit type
- To get the success rate by year

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

The following SQL queries performed on dataset:

- Found the maximum and average payload carried by boosters
- Found successful landing outcome on ground pad and drone ship
- Found the number of both successful and failed landings
- Found the landing outcomes within specific time intervals

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

For better understanding and visualization of the launch sites, markers, circles and lines created using folium for the following purposes:

- Marking all launch sites on a map
- Marking both successful and failed launches for a site
- Measure the distance between launch sites

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

The Dashboard has a pie chart and scatter point chart along with dropdown list and range slider to interact with them.

The plots and interactions added:

To find out the largest successful launches

To find out the highest launch success rate

To find out payload range(s) has the highest launch success rate

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX_lanuch_records_dashboard.ipynb

Predictive Analysis (Classification)

The data transformed and split into training and testing.

Different machine learning models were built and tune different hyperparameters using GridSearchCV.

Accuracy as the metric for our model used, improved the model using feature engineering and algorithm tuning.

Finally, the best performing classification model.

https://github.com/SuhrabAria/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb

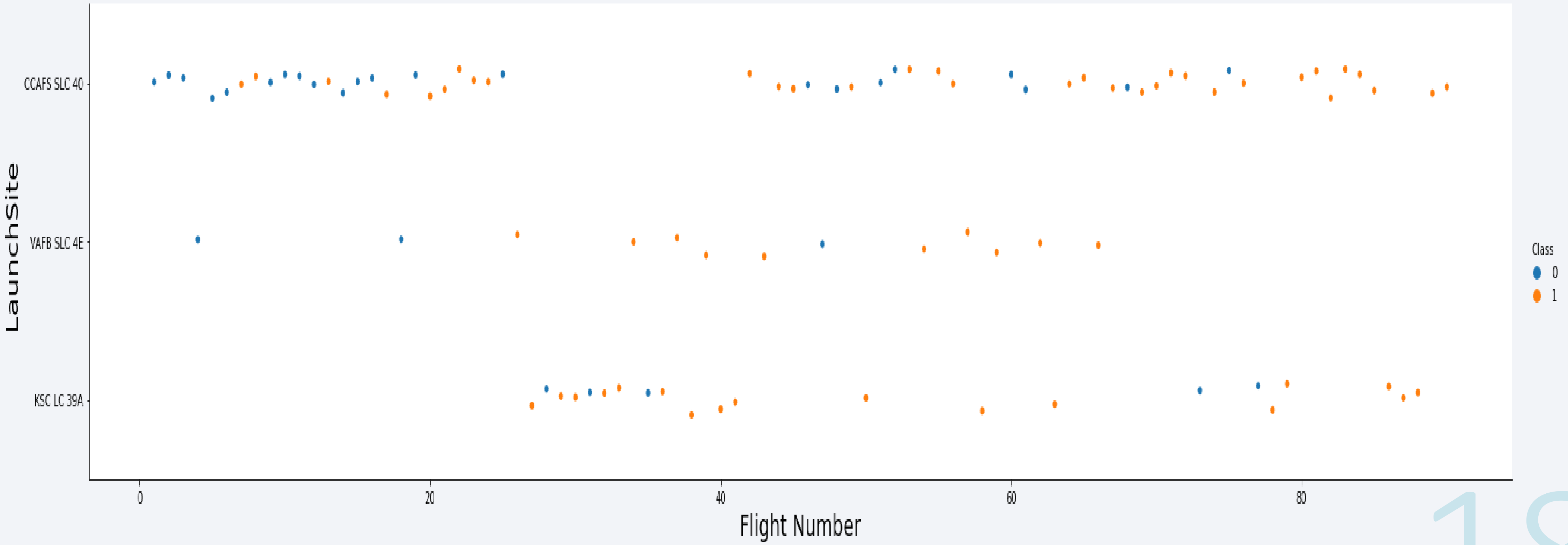
Results

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

Section 2

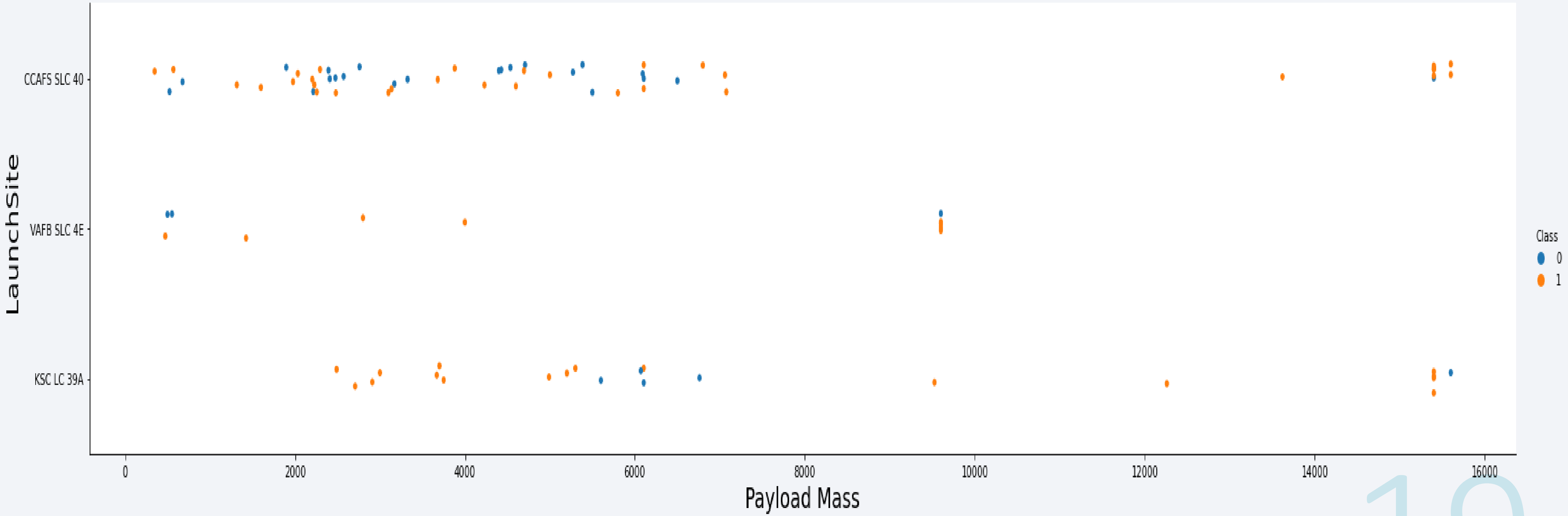
Insights drawn from EDA

Flight Number vs. Launch Site



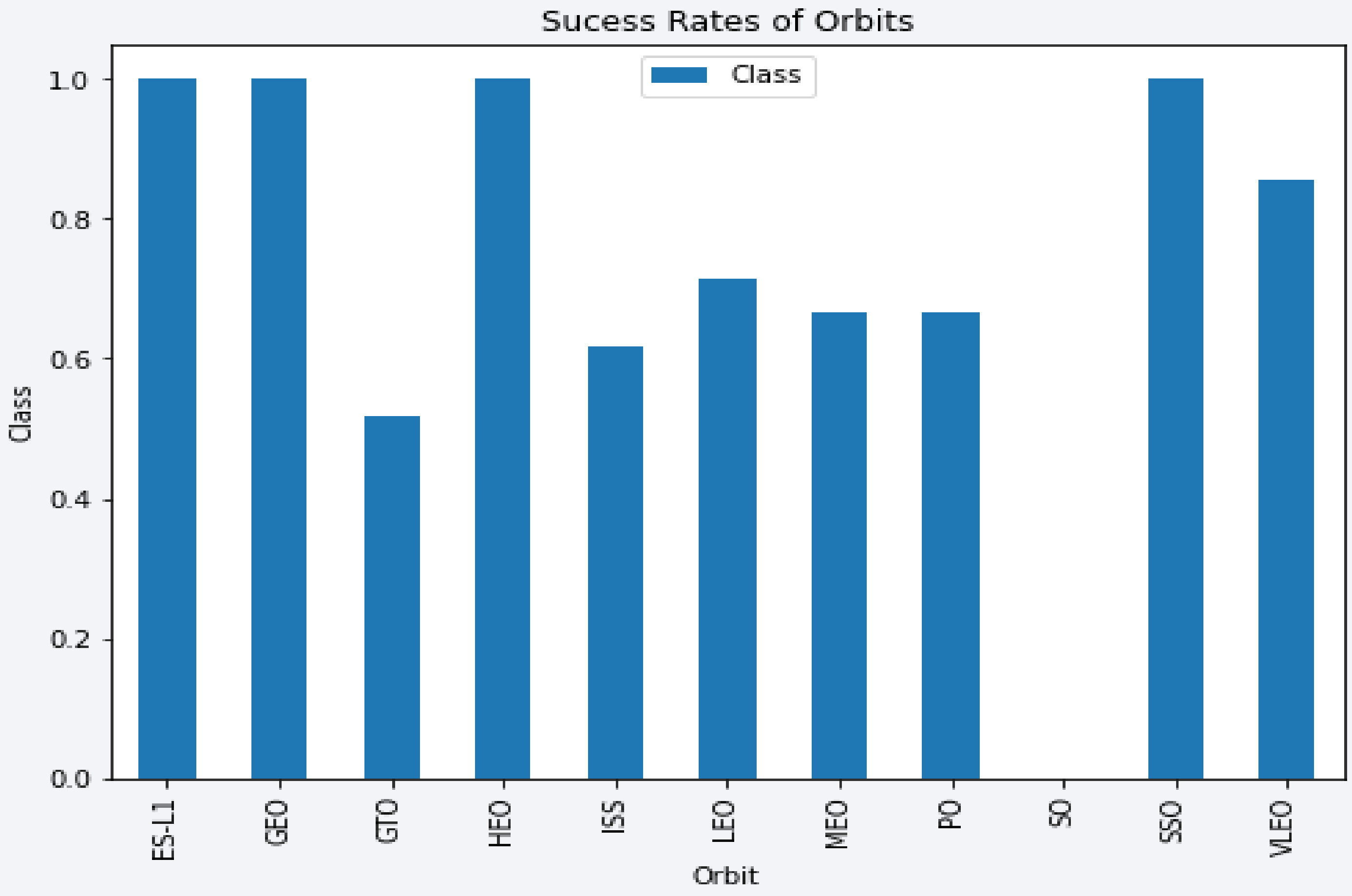
Payload vs. Launch Site

The chart shows for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)



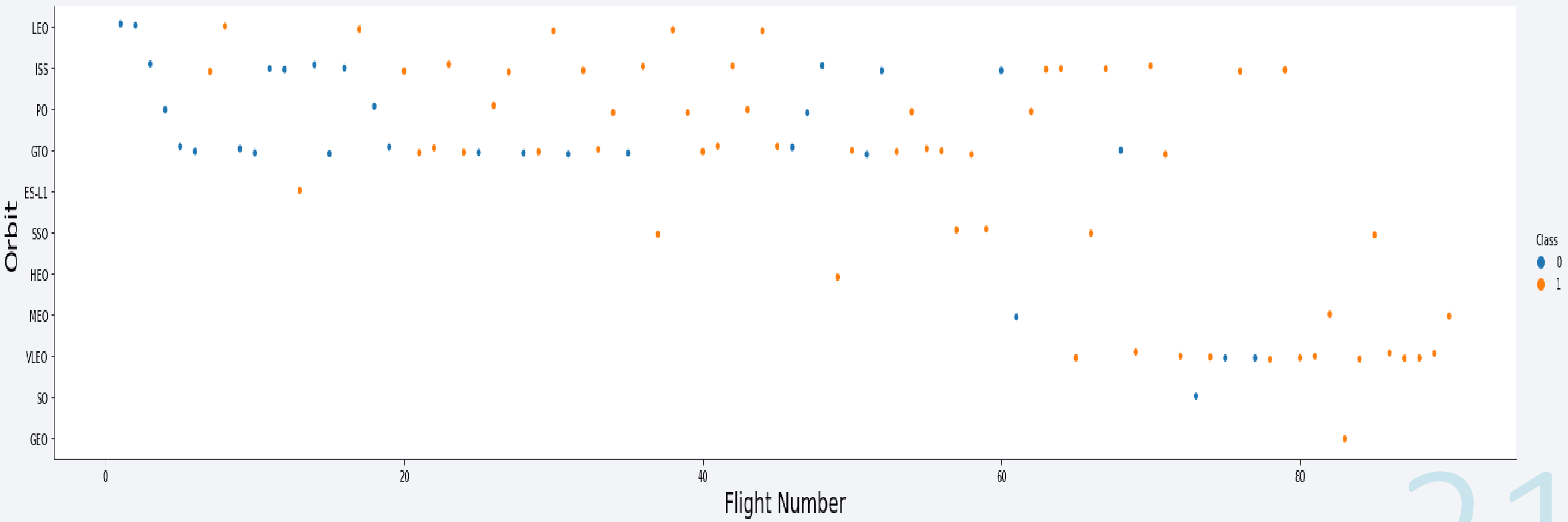
Success Rate vs. Orbit Type

The bar chart shows that ES-L1, GEO, HEO and VLEO orbits have highest sucess rate.



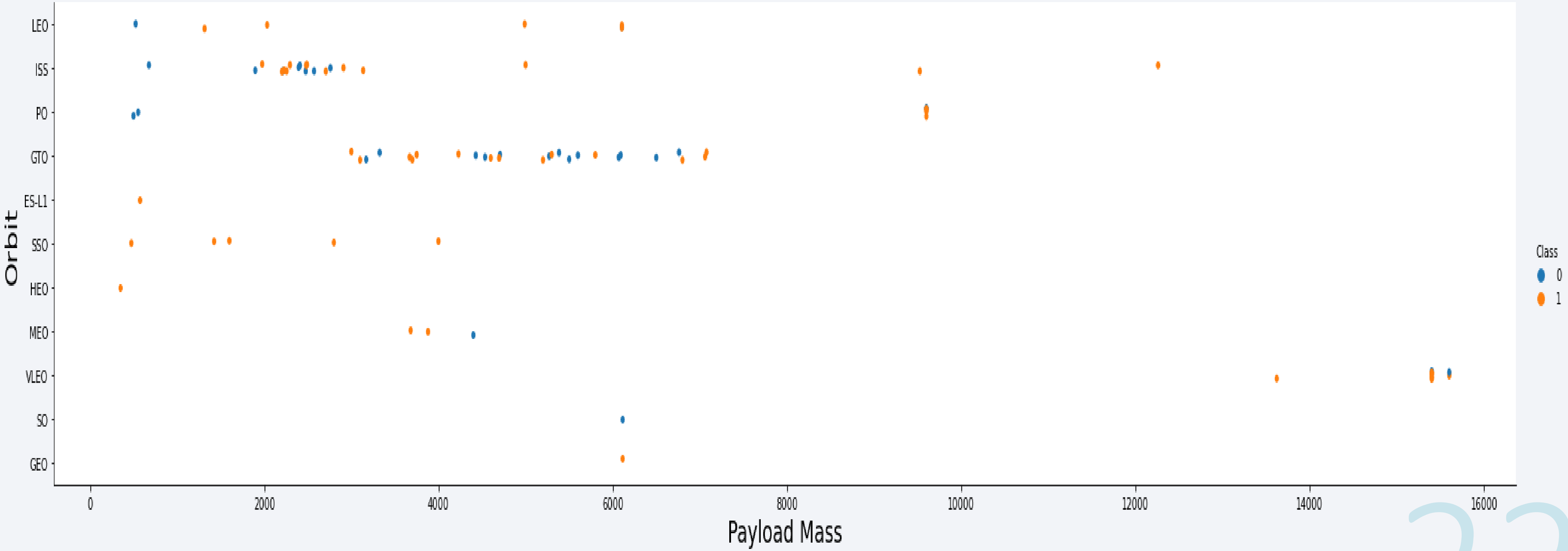
Flight Number vs. Orbit Type

The scatter point shows that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



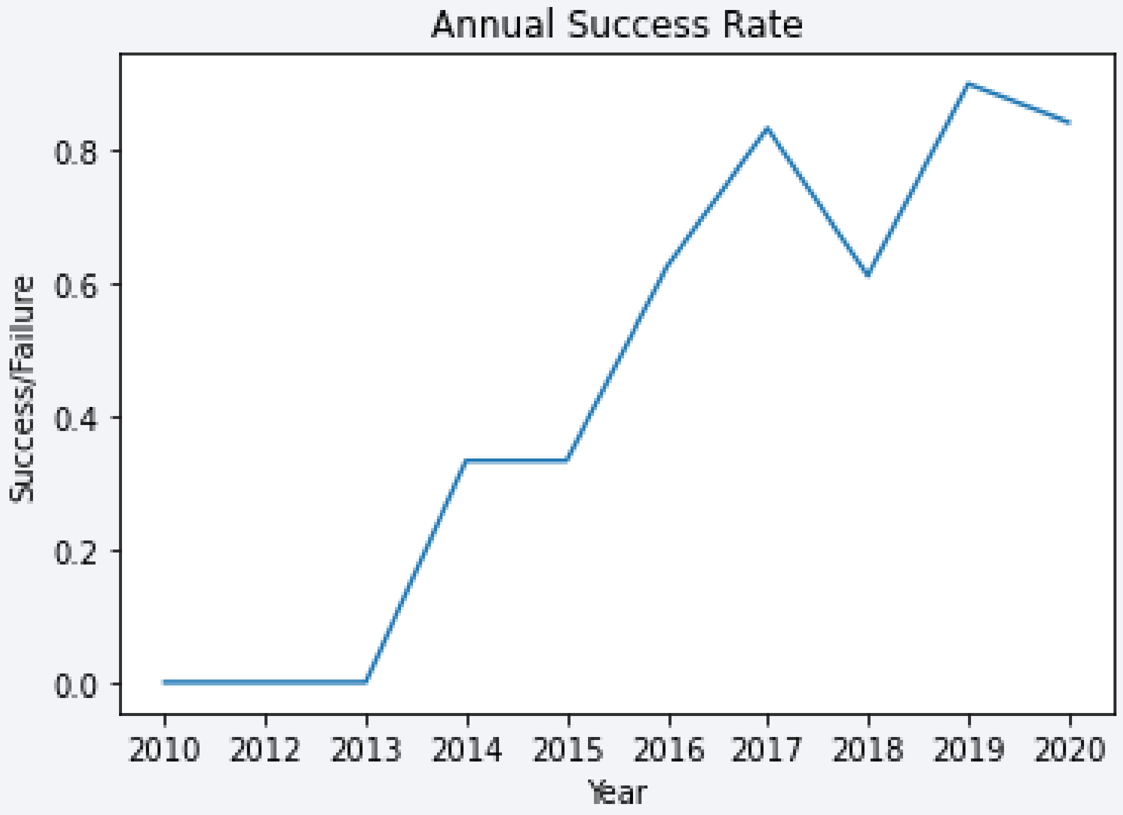
Payload vs. Orbit Type

With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS, but for GTO it is the opposite.



Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020



All Launch Site Names

There have been four unique launch sites

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

Launch Site Names Begin with 'CCA'

Five records where launch sites begin with `CCA`

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

The Total payload carried by boosters from NASA is 619967 Kilogram.

Payload_mass
619967

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 was 2928.4 Kilogram.

Payload_mass
2928.4

First Successful Ground Landing Date

End of 2015 marked with first successful landing outcome on ground pad.

Date	landing_outcome
22-12-2015	Success (ground pad)
18-07-2016	Success (ground pad)
19-02-2017	Success (ground pad)
01-05-2017	Success (ground pad)
03-06-2017	Success (ground pad)
14-08-2017	Success (ground pad)
07-09-2017	Success (ground pad)
15-12-2017	Success (ground pad)
08-01-2018	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

There had been four boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version	payload_mass_kg	landing_outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

There have been a total number of 101 successful and failure mission

Mission_Outcome	COUNT(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Booster F9 v1.0 B0003 carried the maximum payload mass as shown below

Booster_Version	MAX(Payload_mass_kg)
F9 B5 B1048.4	15600

2015 Launch Records

The following query result shows list of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Date	Booster_Version	Launch_Site	landing_outcome
10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

The following result shows landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

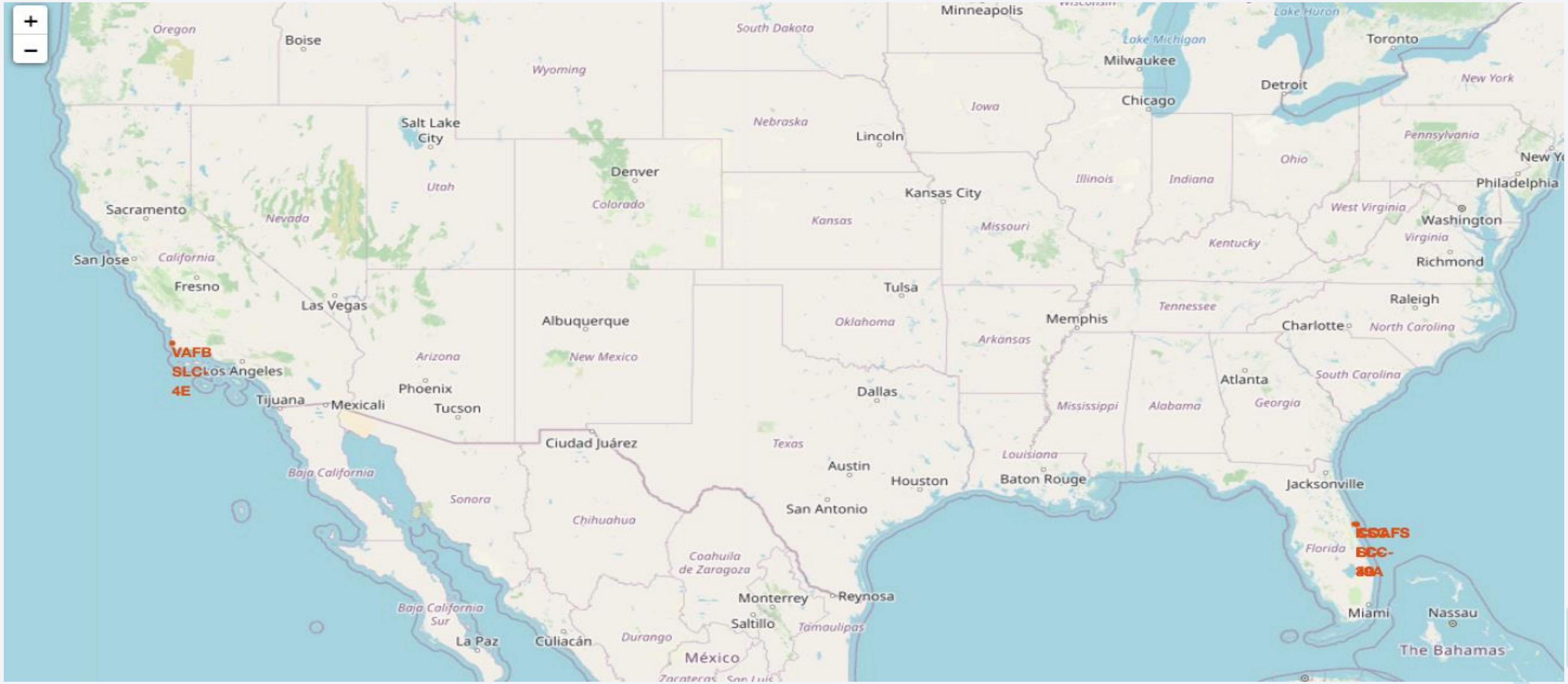
Date	landing_outcome	COUNT(Landing_outcome)
04-06-2010	Failure (parachute)	57

Section 3

Launch Sites Proximities Analysis

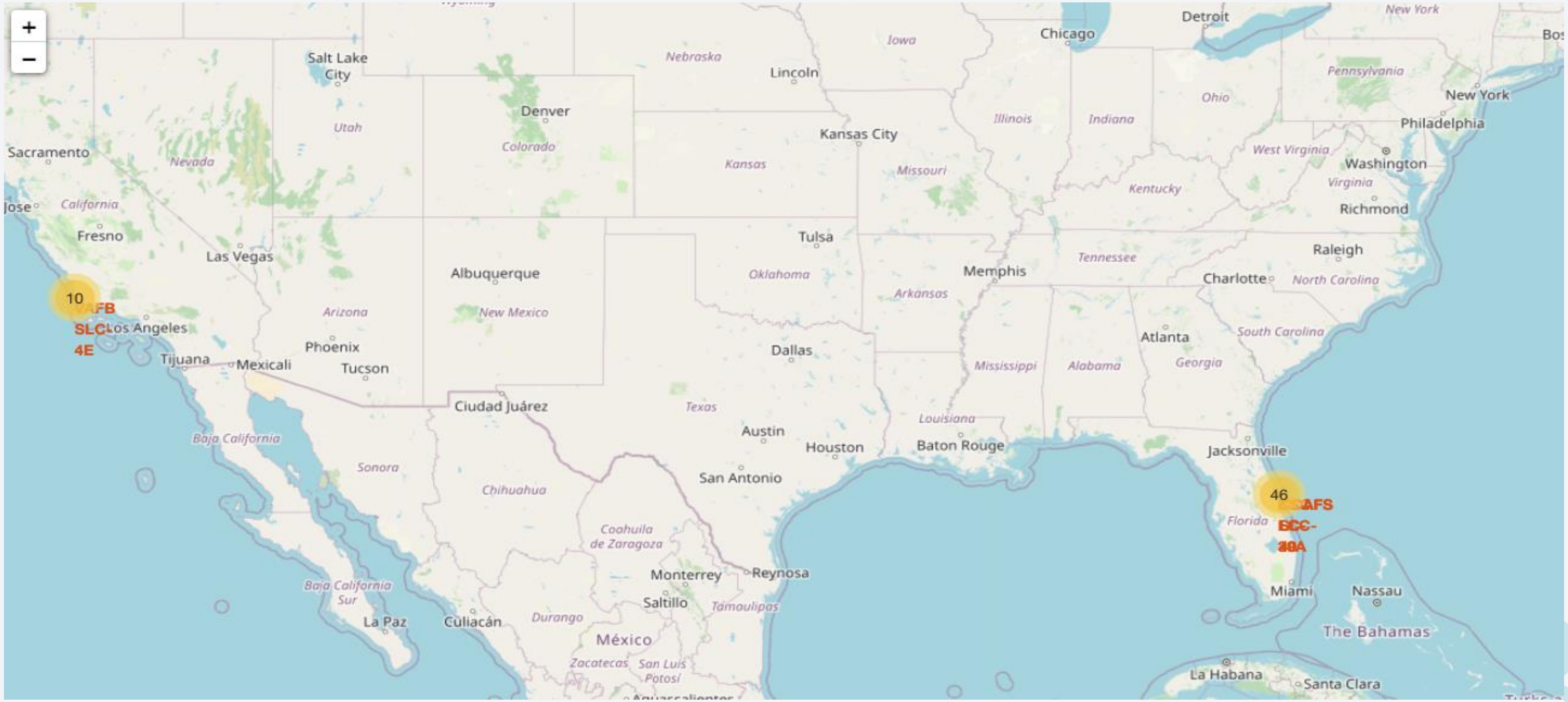
Launch Sites

The following map shows the launch sites.



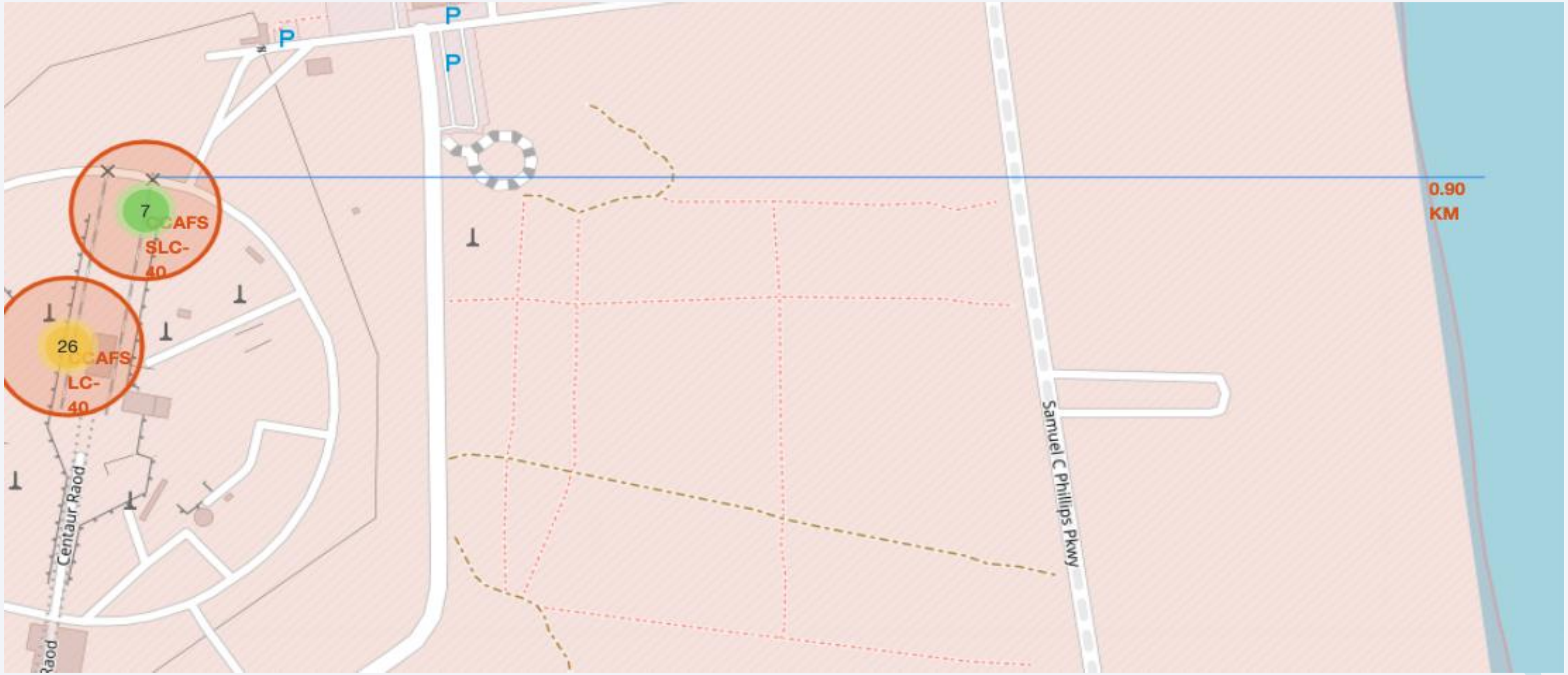
Launch Sites with High Success Rate

The following map shows launch sites which have relatively high success rates



Launch Site Proximities

The following map shows a launch site distance with costal area and a highway closer than coast

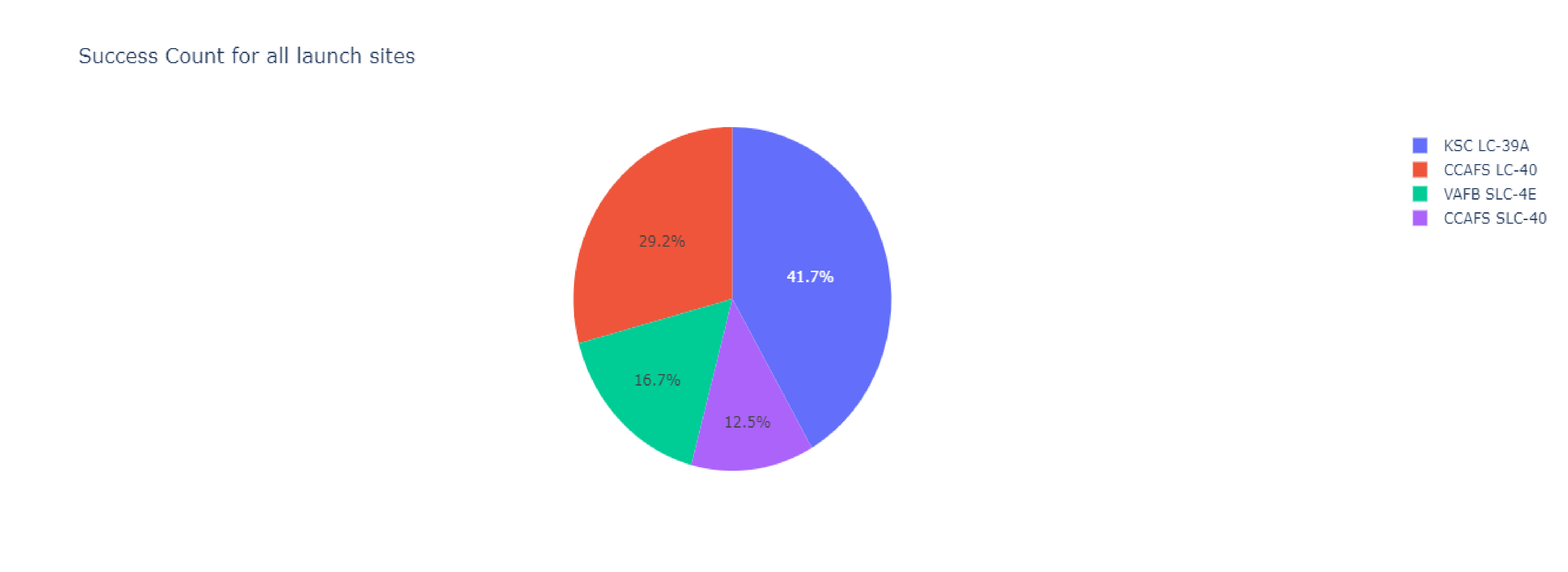


Section 4

Build a Dashboard with Plotly Dash

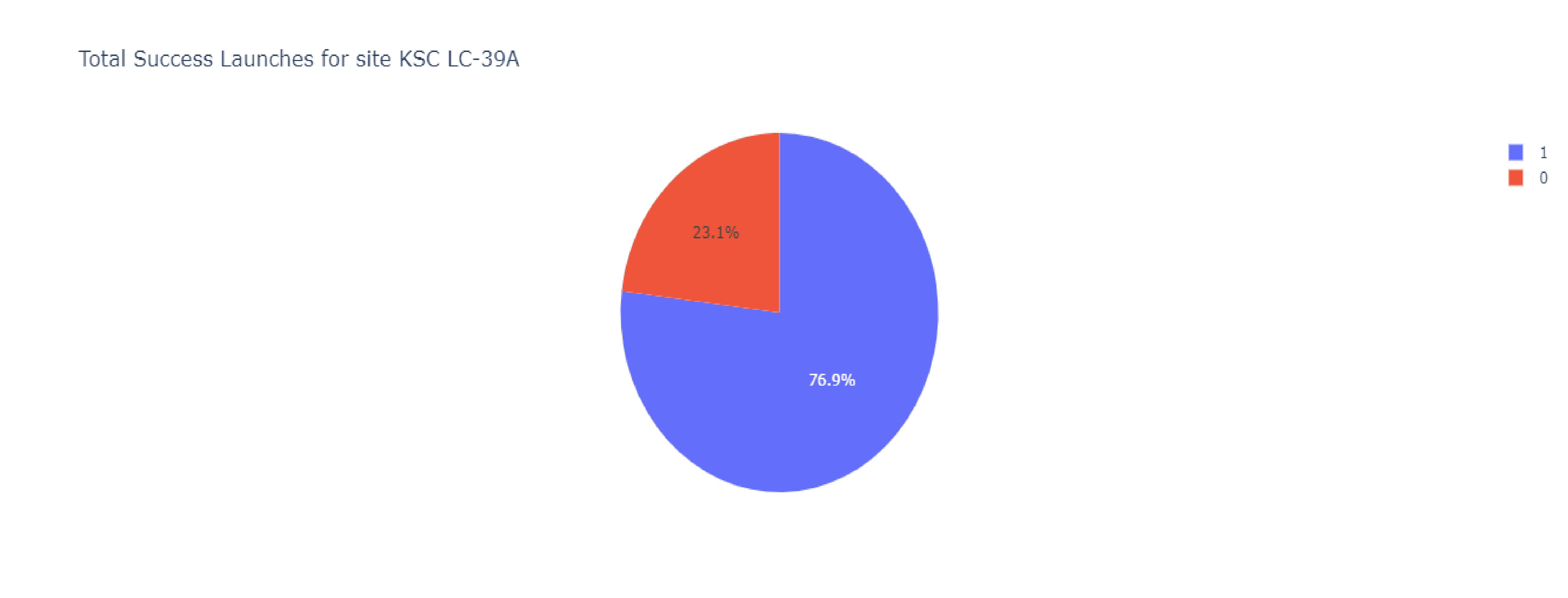
Success Rates of All Launch sites

The pie chart shows the success rates of all launch sites.



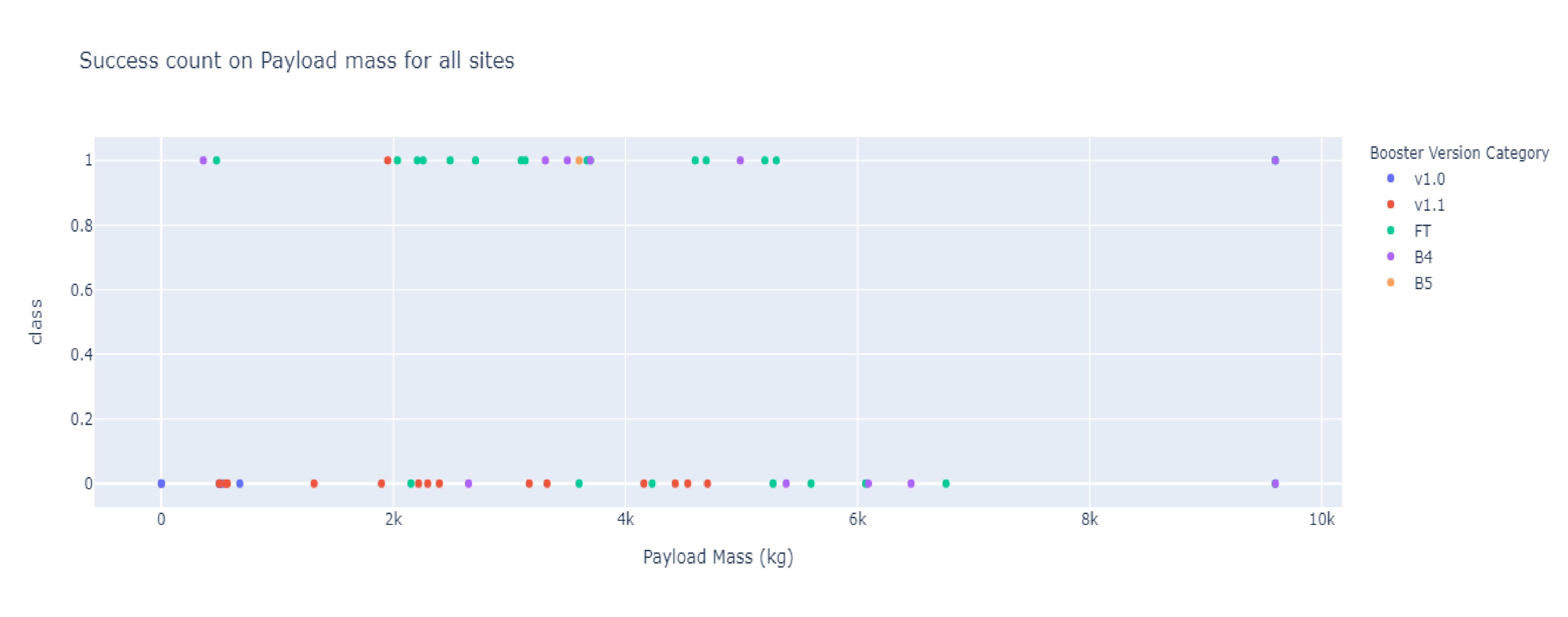
Highest Success Ration

As shown below KSC LC-39A launch sites highest success ratio.



Success Count on Payload mass for all sites

Booster version FT has the largest success rate according to below scatter point



Section 5

Predictive Analysis (Classification)

Classification Accuracy

The accuracy of all algorithms generate the same result as shown below

```
print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print( 'Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))
```

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8333333333333334
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Confusion Matrix

Since Logistic Regression, SVM, K nearest Neighbor and decision tree has the same accuracy, it is concluded that all mentioned models perform well.

Conclusions

The launch success rate shows an upward trend from 2013 onward. It is worth mentioning that number of launches has also increased.

From launch sites, KSC LC-39A has the highest success rate which was 41.7 percent.

Considering the payload mass, BT and F4 booster versions had highest success rates.

Logistic Regression, SVM, K nearest Neighbor and decision tree are best model to predict the success as their accuracy was the same.

Thank you!