IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adrian P. Bustamante
January 20, 2025

Outline

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

Executive Summary

- Summary of methodology
  - In this data science project we intend to predict if the Falcon 9 first stage will land successfully. To achieve this goal we: collected and scrapped the data, performed some exploratory data analysis and feature engineering, built an interactive map with folium and an interactive dashboard. Finally we implemented some machine learning models to make some predictions.
- Summary of all results
  - Certain orbits have high success rate but it is due to the small number of flights visiting those orbits.
  - Heavy payloads show larger successful landings for most types of orbits, however there are a small number of flights with heavy payloads.
  - A decision tree model gave the more accurate predictions.
  - Jupyter notebook with the whole project can be found at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/Capstone-project.ipynb

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully.
- 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.

Section 1

Methodology

Methodology

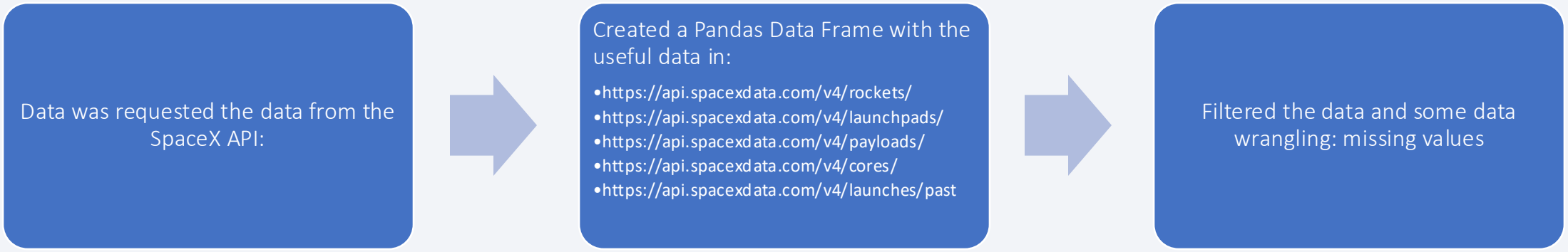
Executive Summary

- Data collection methodology:
  - Request from SpaceX API and Web scraping
- Perform data wrangling
  - Handling missing values and quantifying categorical data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune, evaluate classification models

Data Collection

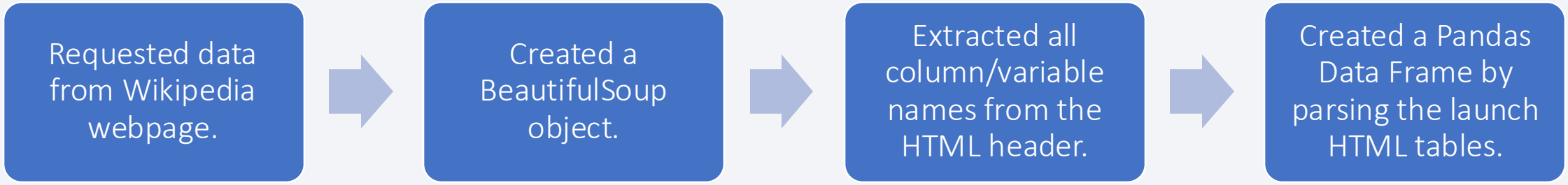
- We requested data from SpaceX API
  - Filter the data we consider useful
- We perform web scraping to obtain extra information of the launches

Data Collection – SpaceX API



- Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/spacex-data-collection-api.ipynb

Data Collection - Scraping



- Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/webscraping.ipynb

Data Wrangling

- Computed number of launches on each site.
- Calculated the number of occurrences of each orbit.
- Computed the mission outcomes for each orbit occurrence.
- Created a new variable 'Class':
  - Class = 0, first stage did not land successfully.
  - Class =1, first stage landed successfully.

- Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/spacex-data-wrangling.ipynb

EDA with Data Visualization

- Some Exploratory Data Analysis was performed, using the libraries Matplotlib and Seaborn, to explore the relation between some of the variables. The following plots were considered:
  - Flight number vs Payload Mass (Scatter plot)
  - Flight number vs Launch Site (Scatter plot)
  - Payload Mass vs Launch Site (Scatter plot)
  - Flight number vs Orbit (Scatter plot)
  - Payload Mass vs Orbit (Scatter plot)
- Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/eda-dataviz.ipynb

EDA with SQL

Using sqlite we performed the following queries:

- %sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not null
- %sql select distinct Launch_Site from SPACEXTABLE
- %sql select sum(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer like '%NASA%'
- %sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Booster_Version like '%F9 v1.1%'
- %sql select min(Date), Landing_Outcome from SPACEXTABLE WHERE Landing_Outcome like '%ground pad%'
- %sql select Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ from SPACEXTABLE where ((Landing_Outcome like '%drone ship%') and (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) and (Landing_Outcome like '%Success%'))
- %sql select Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
- %sql SELECT Date, STRFTIME('%m',Date), STRFTIME('%Y',Date), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE WH ERE (Landing_Outcome like '%Failure%' and Landing_Outcome like '%drone ship%' and STRFTIME('%Y', Date) = '2015')
- %sql select Landing_Outcome, count(Landing_Outcome) from (select Landing_Outcome from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20') group by Landing_Outcome order by count(Landing_Outcome) desc

Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/eda-sql-sqlite.ipynb

Build an Interactive Map with Folium

- Marked all launch sites on a map
- Marked the success/failed launches for each site on the map
- Calculated distance between launch sites to their proximities
- Jupyter Notebook available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/launch-site-location.ipynb

Build a Dashboard with Plotly Dash

- Built a dashboard showing pie charts showing the success rate for each launch site.
- The dashboard also includes plots showing the correlation between Payload Mass and Success for all launch sites.
- Python code available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/spacex_dash_app.py

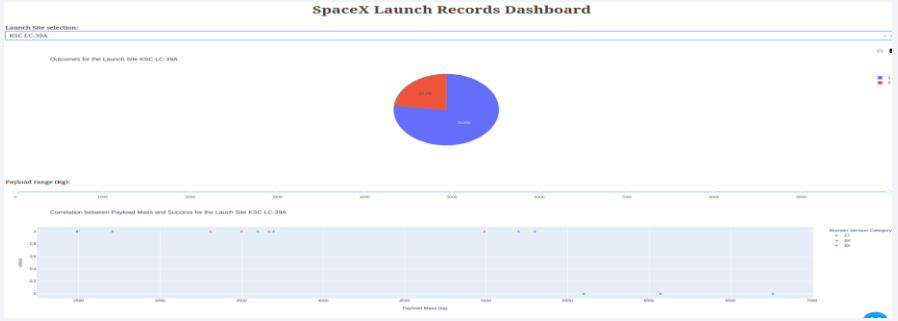
Predictive Analysis (Classification)

- We used scikit-learn to build, fit, evaluate and cross validate well known machine learning classification models:
  - Logistic Regression, Support Vector Machine (SVM), Decision Tree classifier, and k-nearest neighbors (KNN)
- The process to build and cross validate the classification models is as follows:

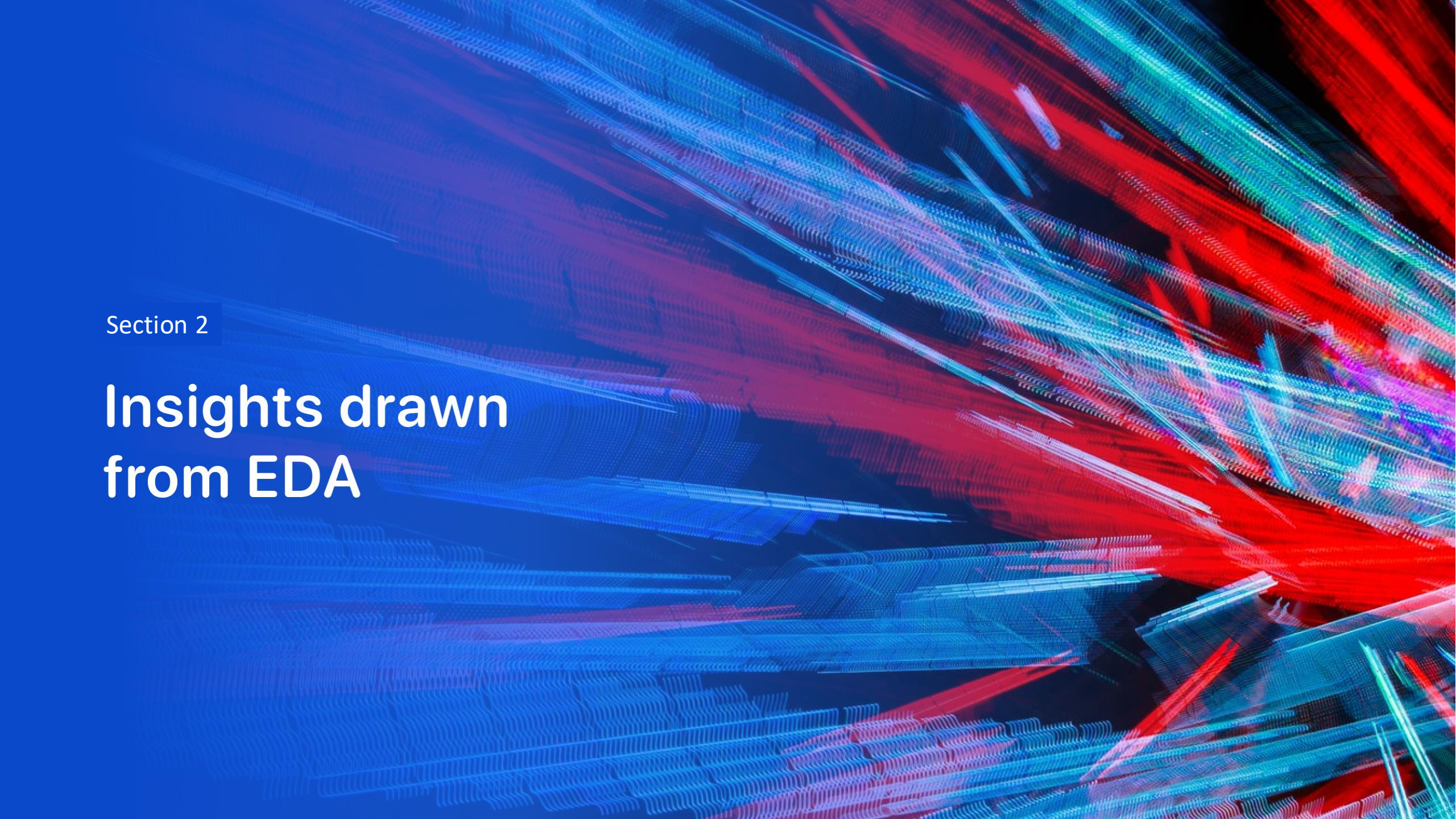
- Python code available at https://github.com/adrian-pbustamante/Space-X-Falcon-9-first-stage-landing-prediction/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Results

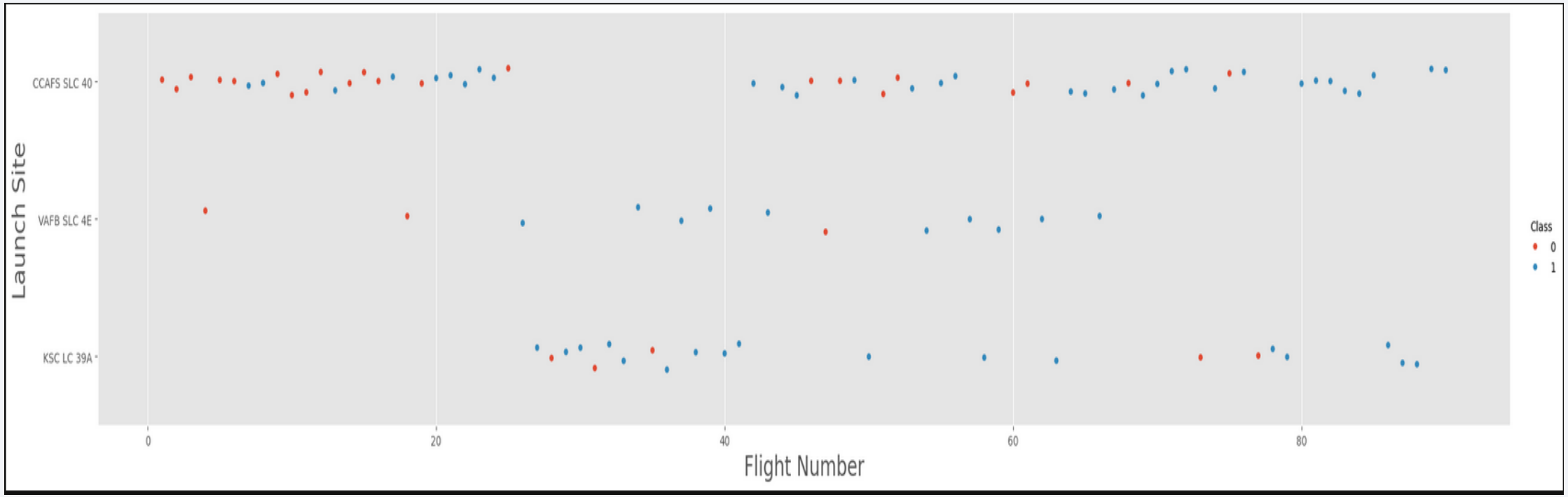
- Exploratory data analysis results
  - Certain orbits have high success rate but it is due to the small number of flights visiting those orbits.
  - Heavy payloads show larger successful landings for most types of orbits, however there are a small number of flights with heavy payloads.
- Built a Dashboard showing the success rate per launch site and the relation between the Payload Mass and Booster Version



- Predictive analysis results
  - A decision tree with the following parameters seems to be the best model to make predictionstuned hpyerparameters :(best parameters) `{'criterion': 'entropy', 'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 10, 'splitter': 'best'}`

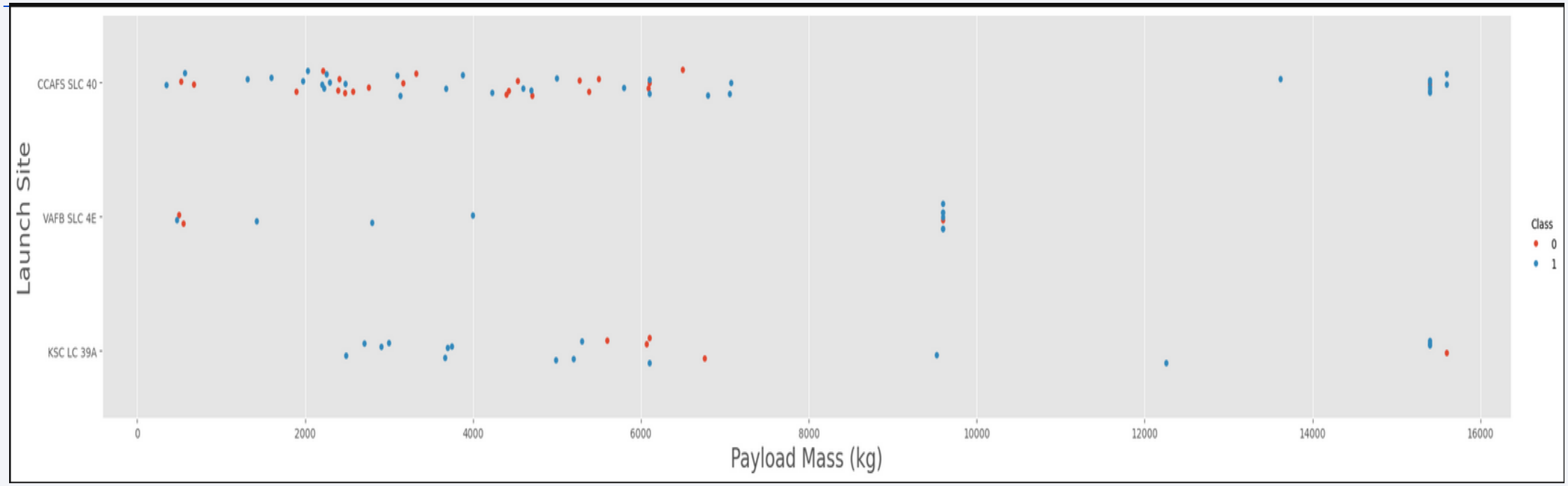

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



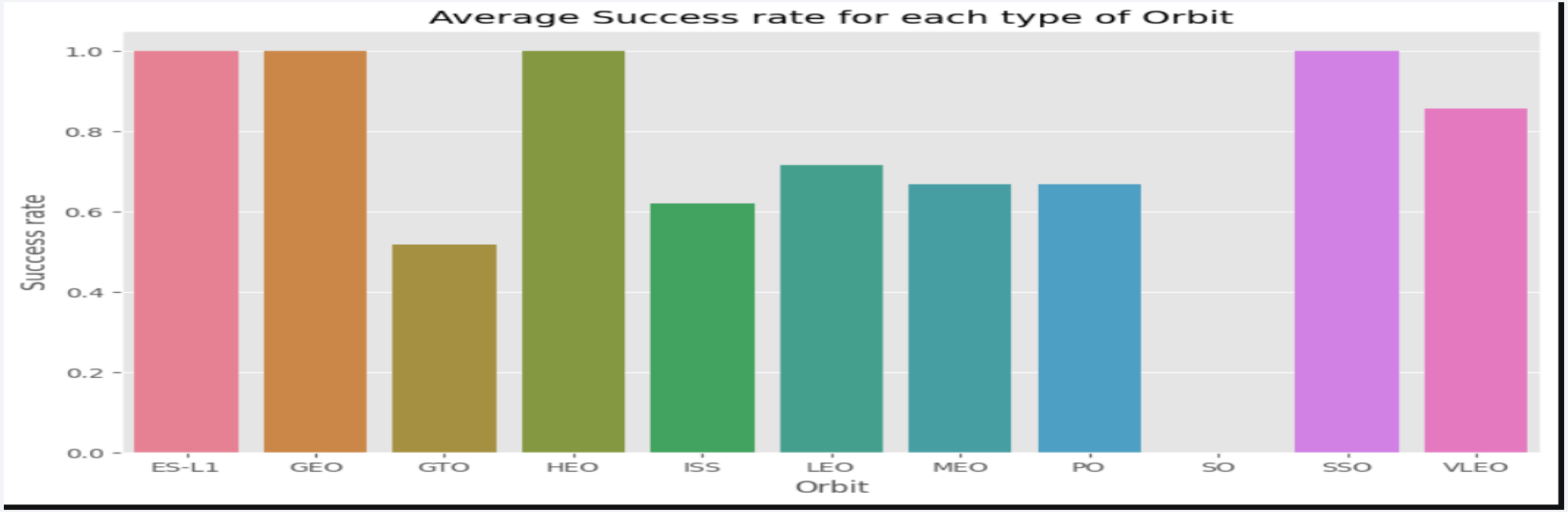
Plot of Flight number vs Launch site. Success rates for sites KSC LC 39A and VAFB SLC 4E seem to be larger, though there are less flights launched from these sites.

Payload vs. Launch Site



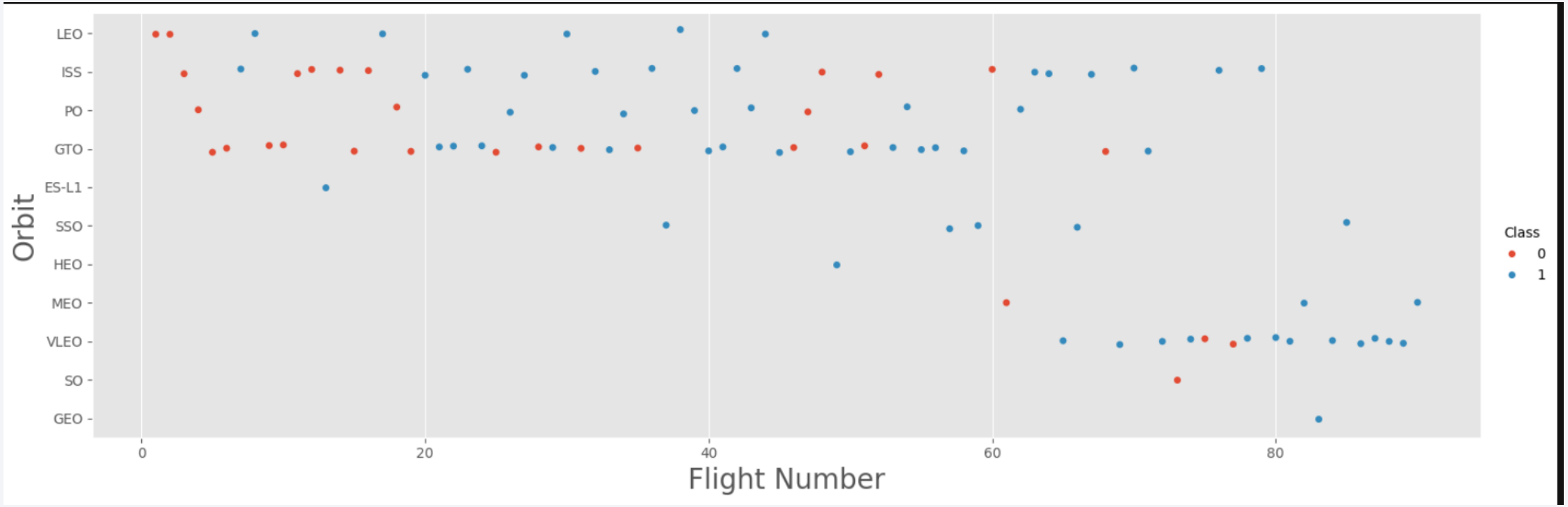
Scatter plot of Payload Mass vs Launch Site. The plot suggest that the success rates is higher for large payloads.

Success Rate vs. Orbit Type



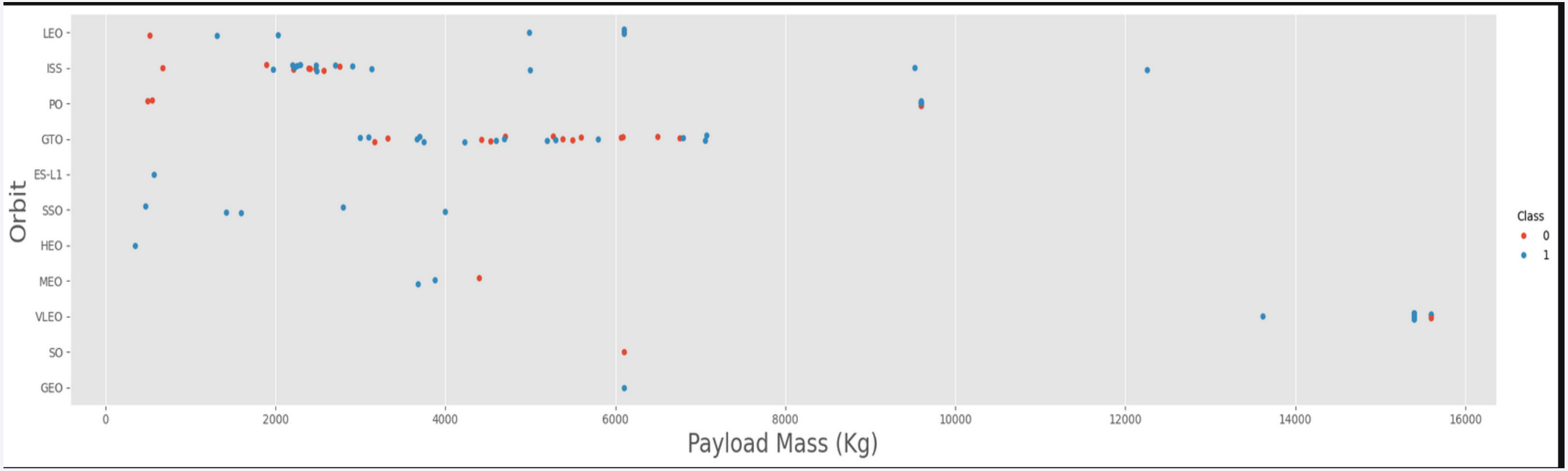
Success rate for orbits ES-L1, GEO, HEO, and SSO its quite high.

Flight Number vs. Orbit Type



Scatter plot of Flight number vs Orbit. The success rate for the ES-L1, GEO, HEO, and SSO orbits is due the small number of flights visiting those orbits.

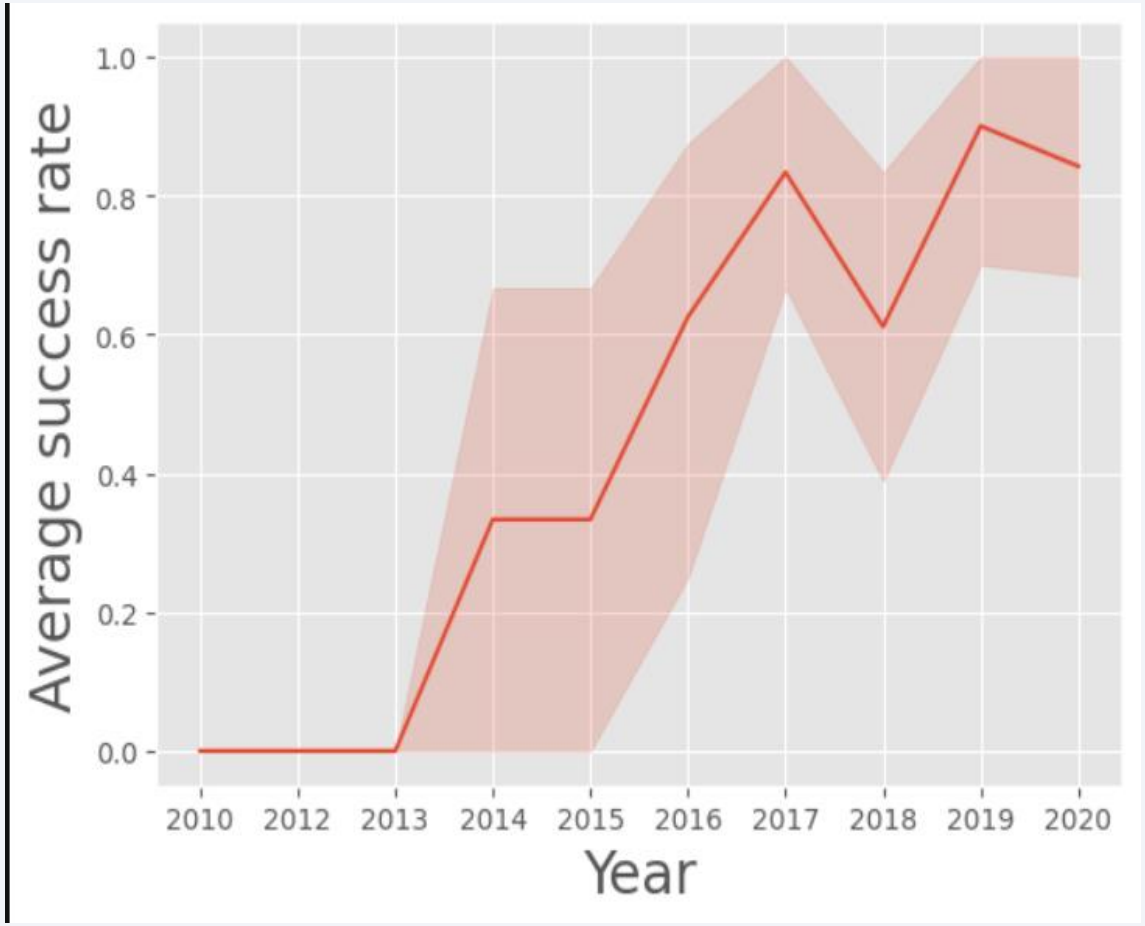
Payload vs. Orbit Type



The plot seems to indicate that the success rate is higher for heavy payloads, at least for the orbits VLEO, ISS, LEO.

Launch Success Yearly Trend

The plot shows that the success rate has increased through the years.



All Launch Site Names

- Find the names of the unique launch sites using sqlite

%sql select distinct Launch_Site from SPACEXTABLE

- The query above shows all the unique launch sites from the table SPACEXTABLE

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA` using sqlite

%sql select * from SPACEXTABLE where Launch_Site like '%CCA%' limit 5

The query above finds the first 5 entries on the table SPACEXTABLE such that the name of the Launch_Site contains the charaters 'CCA'

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql select sum(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer like
'%NASA%'
```

- The query above sums all the payload mass entries from the table SAPCEXTABLE where the names of the Customer contains the characters 'NASA'

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Booster_Version like '%F9 v1.1%'
```

- The query above computes the average payload mass (kg) of all the entries on the table SAPCEXTABLE where the Booster_Version contains the characters 'F9 v1.1'

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql select min(Date), Landing_Outcome from SPACEXTABLE WHERE
Landing_Outcome like '%ground pad%'
```

- The query above selects the first date and the Landing_Outcome form the SPACEXTABLE where the Landing_Outcome contains the characters 'ground pad'

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql select Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ from SPACEXTABLE where ((Landing_Outcome like '%drone ship%') and (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) and (Landing_Outcome like '%Success%'))
```

- The query above shows the Booster_Version, Landing Outcome and Payload mass of the entries where Landing_Outcome contains the characters 'drone ship', the Payload Mass is between 4000 and 6000 (kg), and the Landing_Outcome is successful.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

- %sql select Mission_Outcome,COUNT(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome

- The query above shows the number of each type of Mission Outcome from SPACEXTABLE

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- The query above shows the Booster_Version and the Payload_Mass from SPACEXTABLE for the entries such that the Payload Mass is the largest.

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT Date, STRFTIME('%m',Date), STRFTIME('%Y',Date), Landing_Outcome,
   Booster_Version, Launch_Site from SPACEXTABLE WHERE (Landing_Outcome like
   '%Failure%' and Landing_Outcome like '%drone ship%' and STRFTIME('%Y', Date) =
   '2015')
```

- The query above shows the date, month, year, Landing Outcome, Booster_Version, Launch_Site from the SPACEXTABLE for the entries where the Landing_Outcome contains the characters 'Failure' and 'drone ship'; and the year of the date is 2015

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

- 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 select Landing_Outcome, count(Landing_Outcome) from (select
Landing_Outcome from SPACEXTABLE where Date between '2010-06-04' and
'2017-03-20' ) group by Landing_Outcome order by count(Landing_Outcome) desc
```

- The query above shows the count of each type of Landing_Outcome that occurred between 2010-06-04 and 2017-03-20.

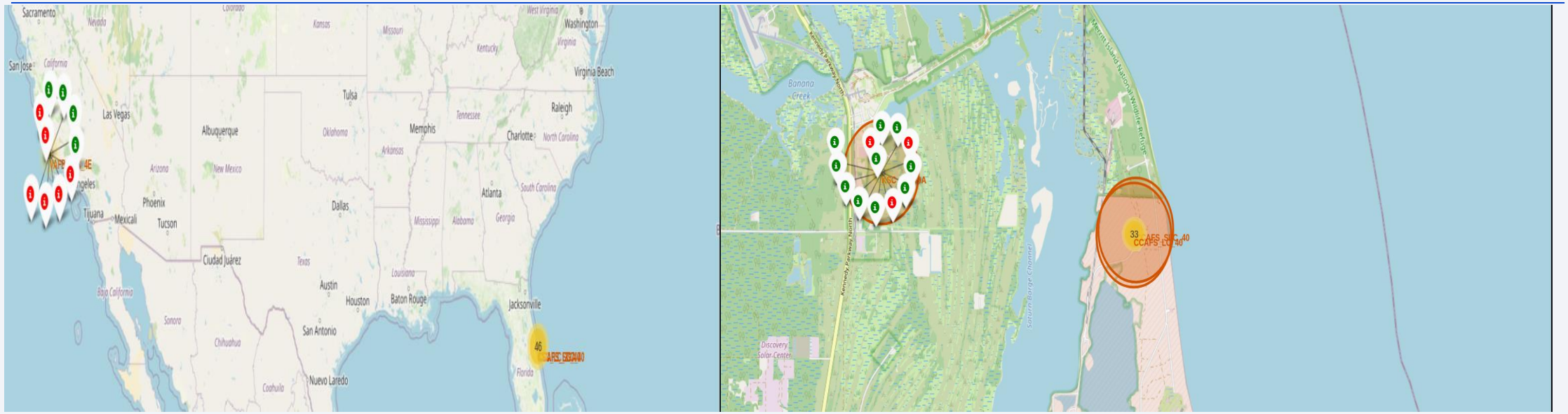
Section 3

Launch Sites Proximities Analysis

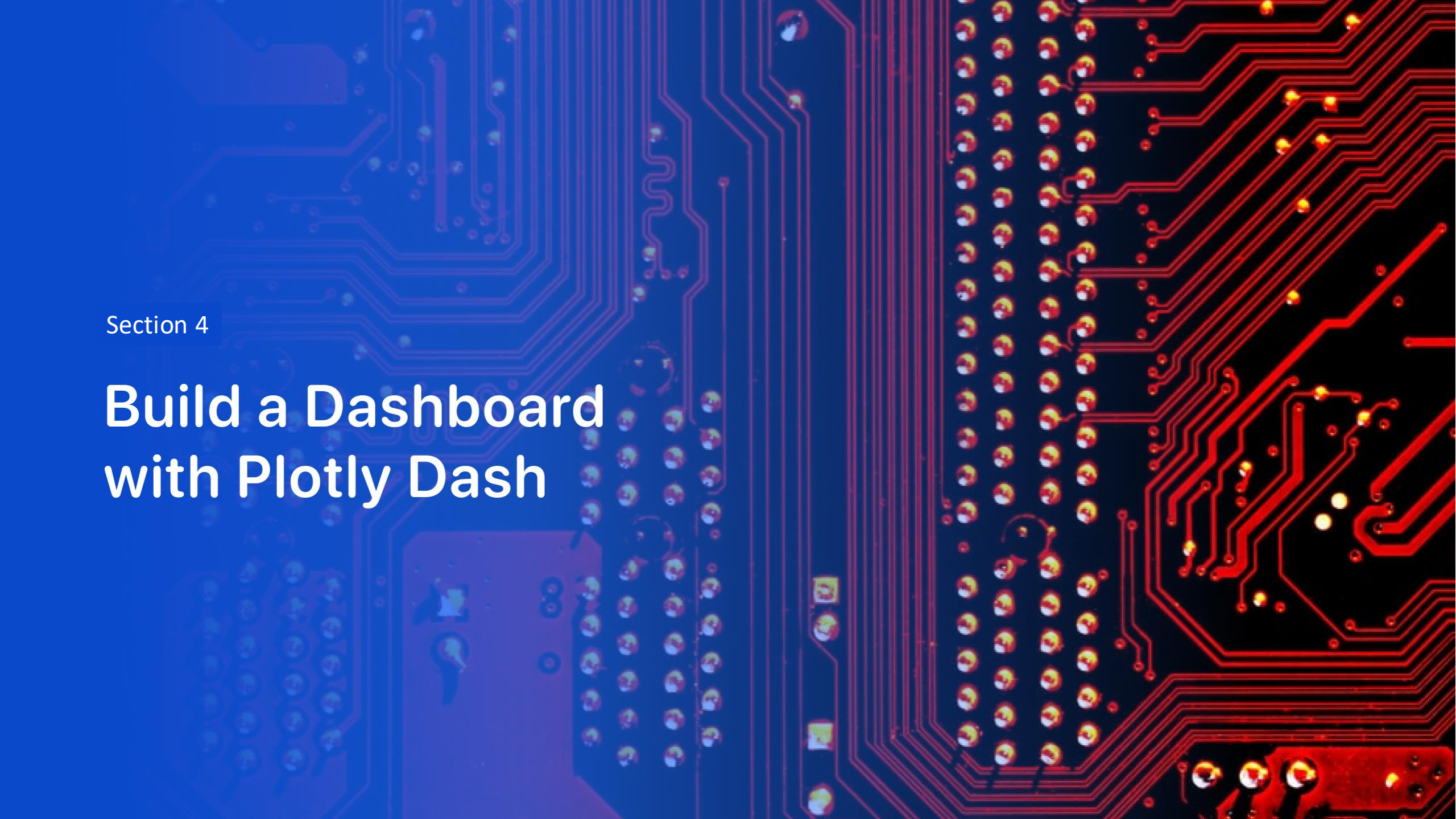
Folium map launch sites locations

Screenshots of the folium map generated with markers on all Launch Sites locations.

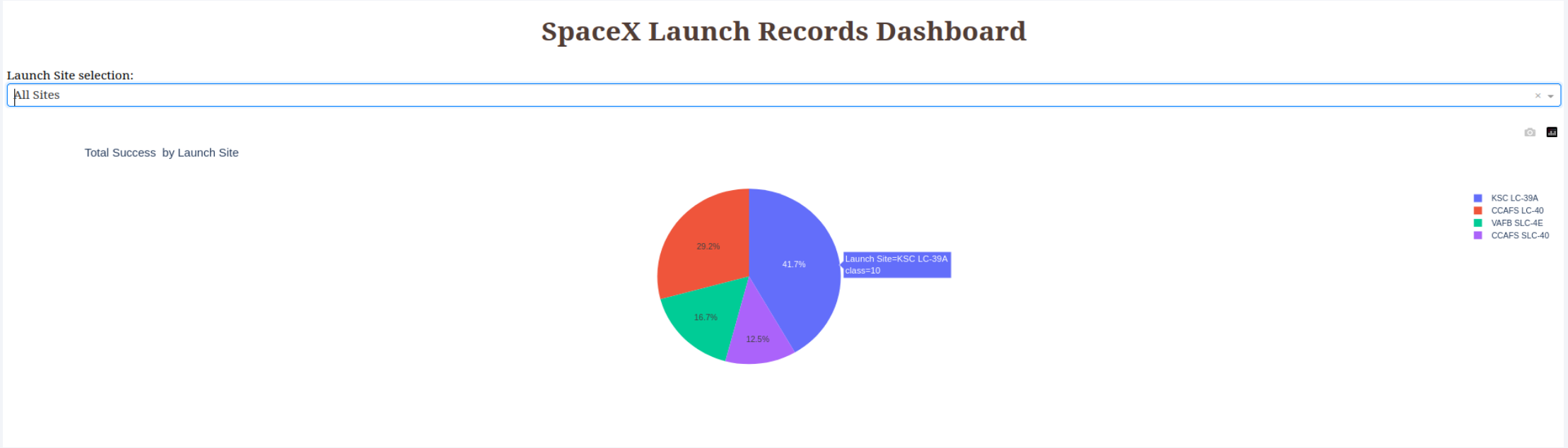
Folium map with cluster markers



We built in folium an interactive map showing the outcome of each launch at each launch site.

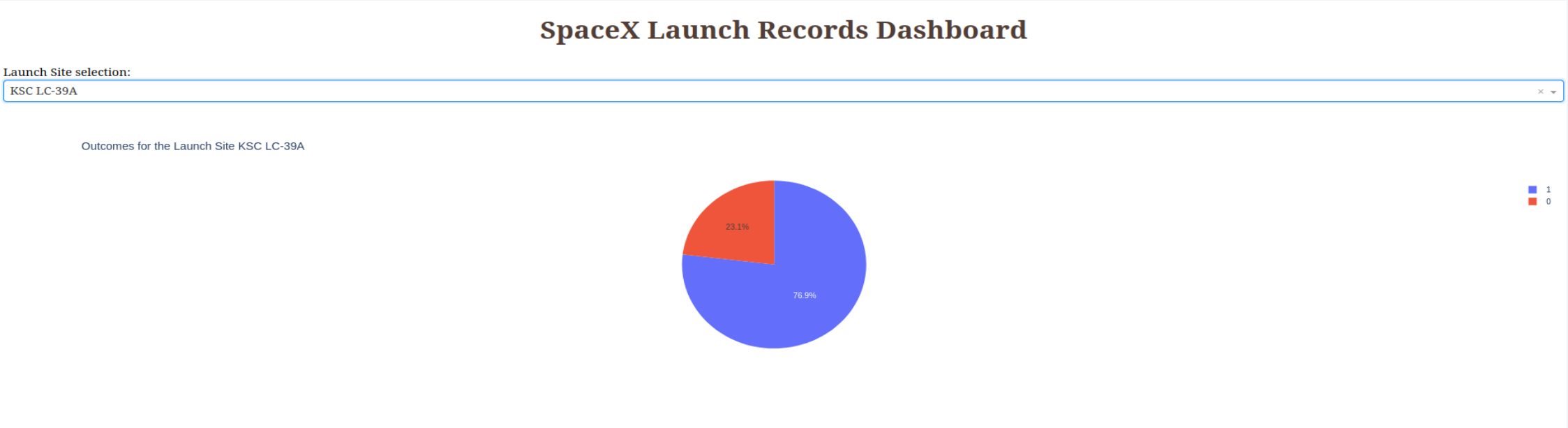

Section 4
Build a Dashboard with Plotly Dash

Dashboard Launch Success count



An interactive dashboard was built. In the picture we can see and pie chart with the success rate of all Launch sites.

Dashboard Launch Site with largest success ratio



The interactive dashboard shows a pie chart with the percentage of fail/success outcomes for the launch site with the highest success ration.

Dashboard: Payload vs Launch Outcome

The dasboard includes a scatter plot showin the correlation between Payload Mass and success for the Launch Site selected. In the interactive dashboard the Payload Mass can be adjusted.

Thank you!