Winning Space Race with Data Science

Vasmi Sudarshini PN
07/07/2022

Outline

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

Executive Summary

- Data collection methodology:
  - Data was collected via APIs and Webscrapping.
- Perform data wrangling
  - Data wrangling was done to find the perfect 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
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Find the method that performs best using test data

Introduction

- Space X 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 Space X 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 space X for a rocket launch.

Section 1

Methodology

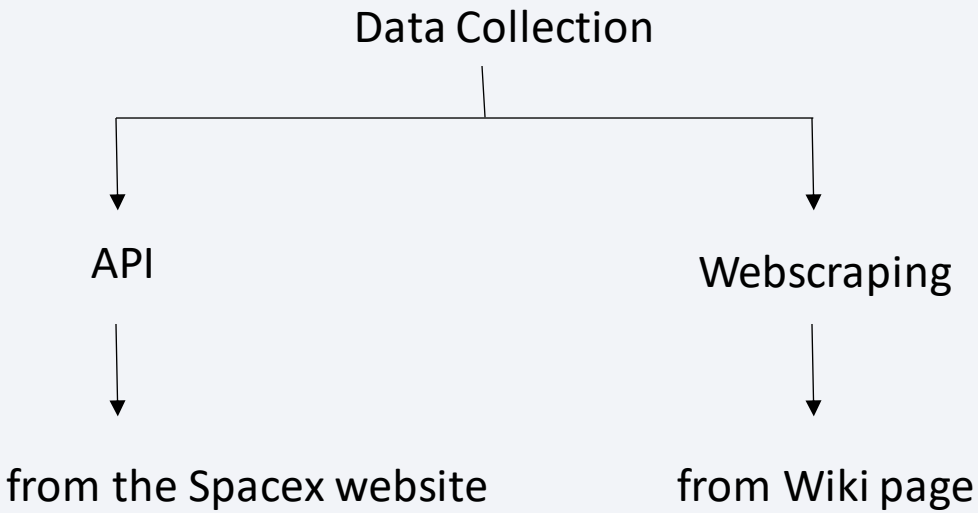
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via APIs and Webscrapping.
- Perform data wrangling
  - Data wrangling was done to find the perfect 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
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Find the method that performs best using test data

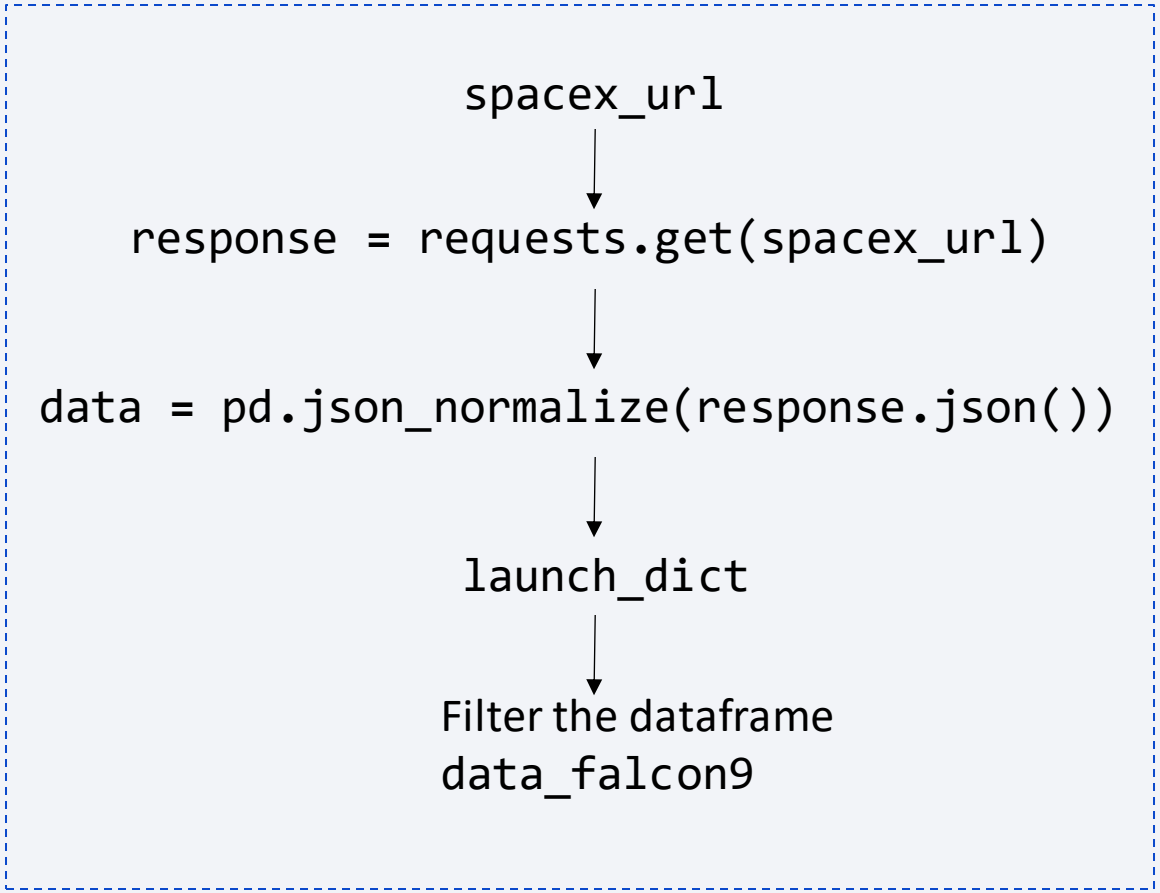
Data Collection

- Data was collected through 2 ways:
  - from the Spacex website directly via **APIs** and
  - from Wiki page via **Webscraping**.
- Data was collected and the wrangled for further EDA (Exploratory Data Analysis).



Data Collection – SpaceX API

- Made a Get Request to the SpaceX API
- Cleaned the requested data
- Data Collection API



Data Collection - Scraping

- Webscraping using **BeautifulSoup**
- Extracted Falcon 9 launch records HTML table from Wikipedia
- Parsed the table and did convert it into a Pandas data frame
- Data Collection Webscraping

```
static_url (wiki page)
        ↓
r = requests.get(static_url).text
        ↓
soup = BeautifulSoup(r, 'html')
        ↓
html_tables = soup.find_all('table')
        ↓
Create a data frame by parsing the launch
HTML tables
```

Data Wrangling

- Performed Exploratory Data Analysis (EDA) to find patterns in the data and determine what would be the label for training supervised models.
- Did convert the outcomes into Training Labels with 1 means the booster successfully landed and 0 means it was unsuccessful.
- Data Wrangling

Calculate the number of launches on each site
`df['LaunchSite'].value_counts()`

Calculate the number and occurrence of each orbit
`df['Orbit'].value_counts()`

Calculate the number and occurence of mission outcome per orbit type
`landing_outcomes = df['Outcome'].value_counts()`

Create a landing outcome label from Outcome column `df['Class']`

EDA with Data Visualization

- Charts plotted:
  - FlightNumber vs. PayloadMass
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Orbit vs Success rate
  - FlightNumber vs Orbit type
  - Payload vs Orbit type
  - Launch success yearly trend

- The charts were plotted to find how different variables affect the success rate of the launch.
- With the help of the charts we will select the features that will be used in success prediction.
- Data Visualization

EDA with SQL

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites began with the string 'CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- Listed the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Ranked the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order
- SQL

Build an Interactive Map with Folium

- Created **folium markers and circles** to locate the launch sites in the map with a text specifying the name of the specific launchsite.
- Created **marker cluster** to simplify the map by clustering the markers of the same coordinate.
- Created **lines** to show the distance between a launch site and it's close proximities.
- Interactive Map with Folium

Build a Dashboard with Plotly Dash

- A **dropdown menu** showing a list of Launch sites was created to enable the user to select a particular launch site and view the success rate.
- A **pie chart** showing the success rate of all launch site was created to visualize the success rate on the whole as well as separately for each launch site.
- A **slider** to select a payload range was created to visualize the success rate for different payload range.
- A **scatter plot** to show the relationship between different payloads and their respective success rates.
- Dashboard-Application-with-Plotly-Dash

Predictive Analysis (Classification)

- Created a column for the class
- Standardized the data
- Split into training data and test data
- Found best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Found the method that performs best using test data.
- Machine Learning Prediction

Created dependent variable and standardized the data

↓

Train Test split

↓

Created LogReg, SVM and Classification Trees models

↓

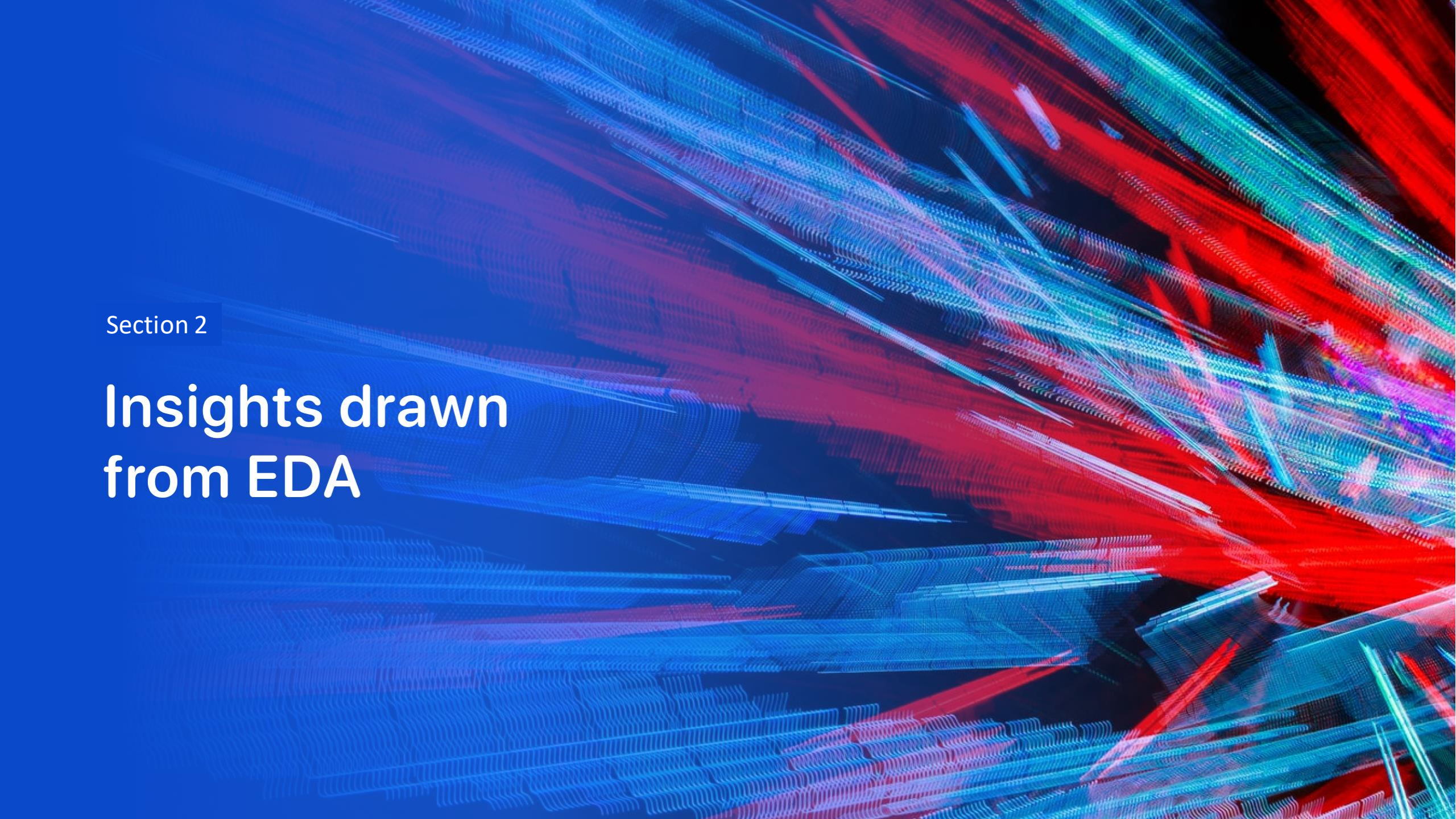
Found the best score and the perfect model for prediction

Results

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

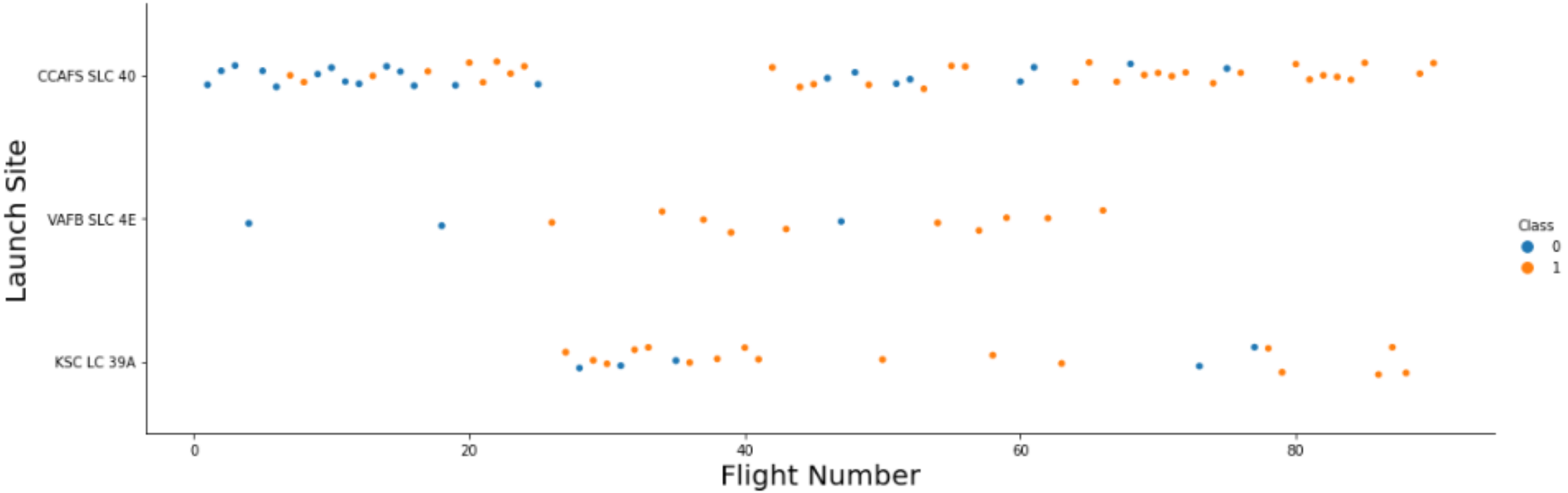
Section 2

Insights drawn from EDA



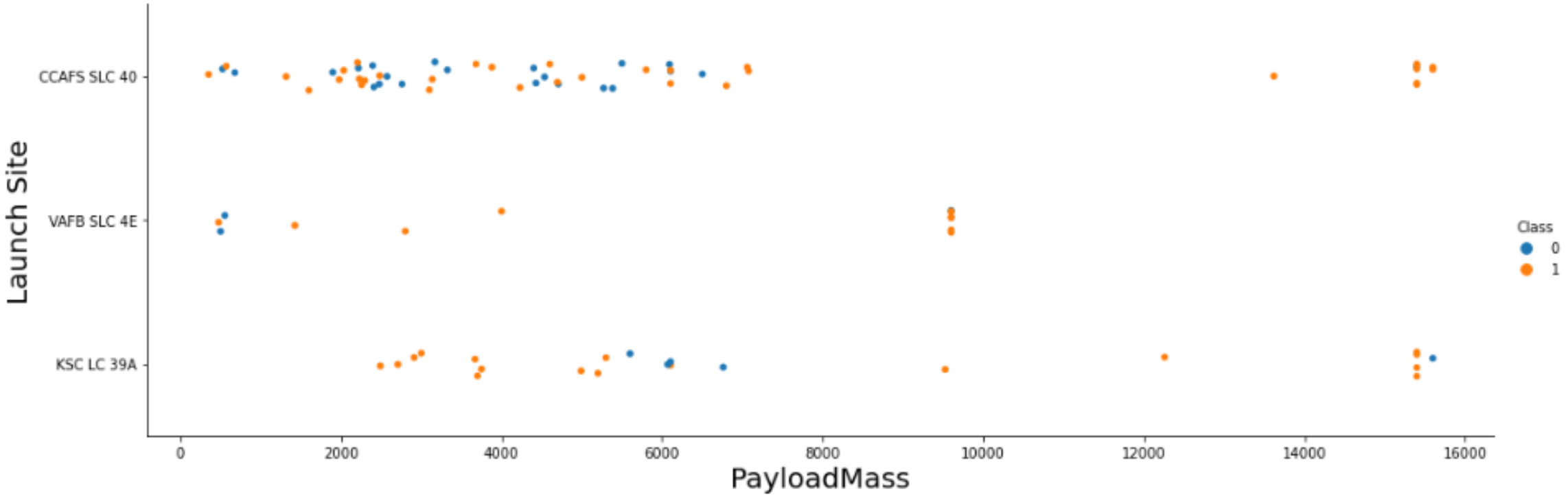
Flight Number vs. Launch Site

- As the Flight Number increases, the success rate increases.
- Success rate for VAFB-SLC launch site is greater than other launch sites.



Payload vs. Launch Site

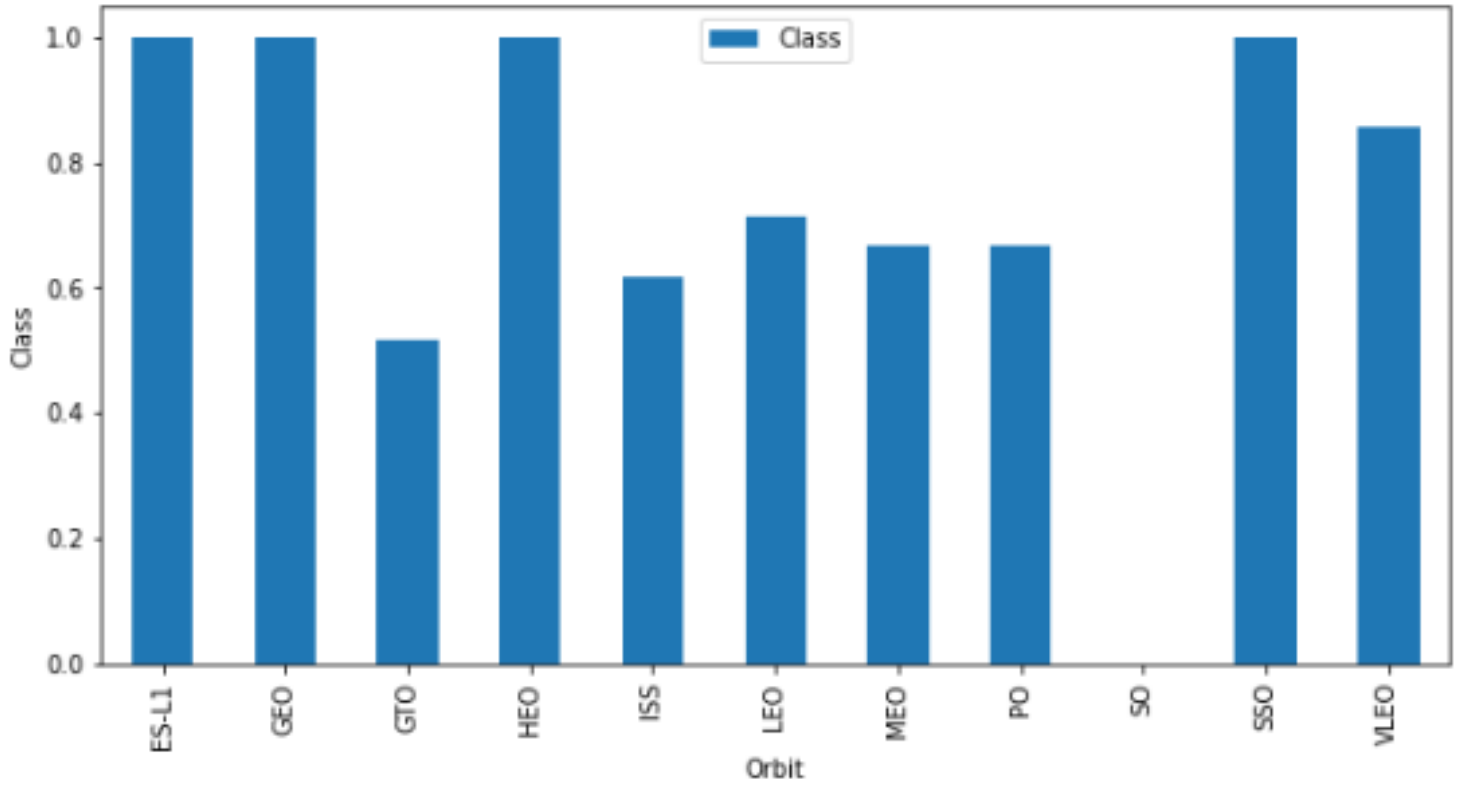
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- As the payload mass increases, the success rate shows promising results.



Success Rate vs. Orbit Type

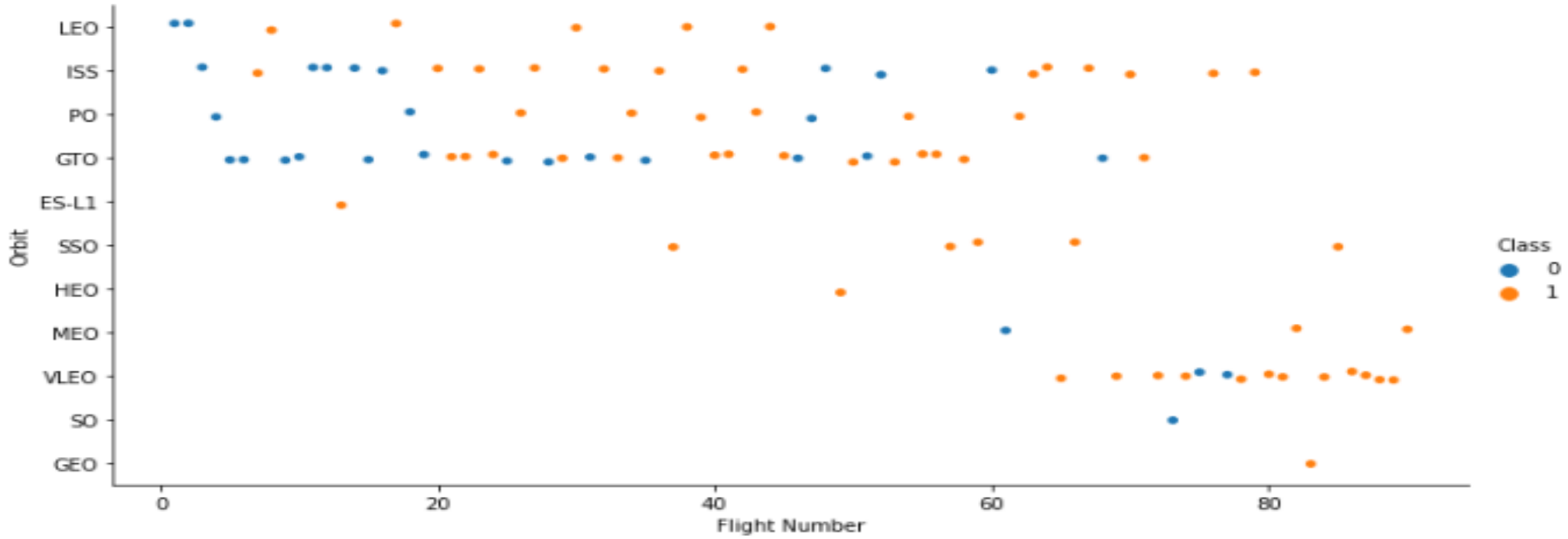
Success rate is higher for 4 orbits:

- ES-L1
- GEO
- HEO
- SSO



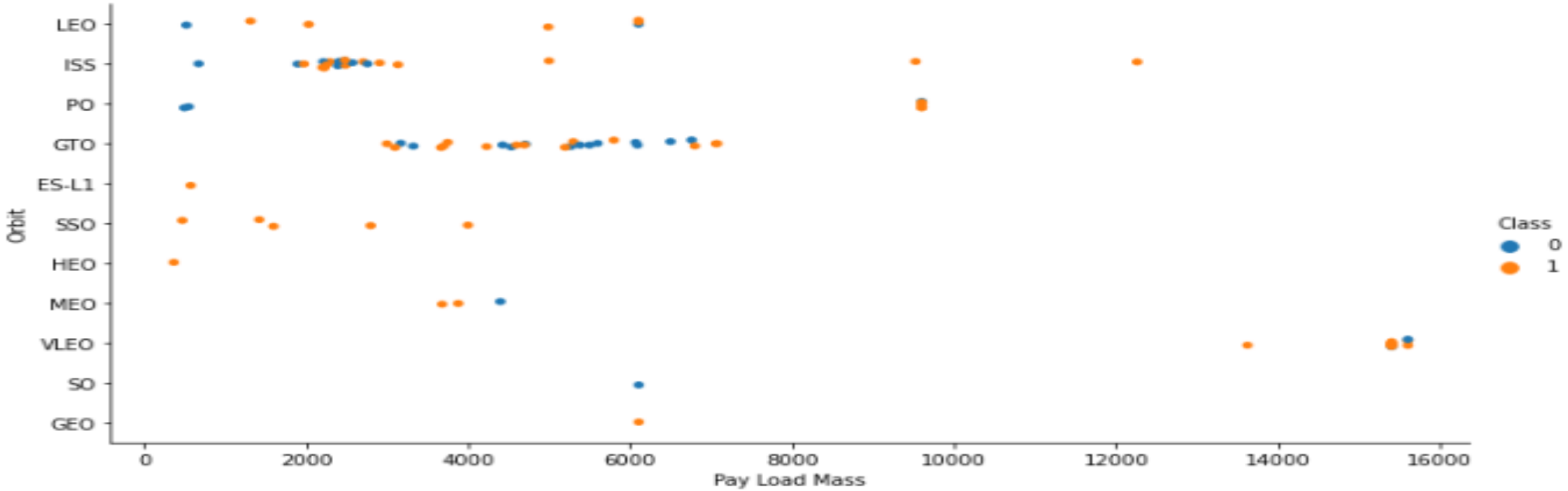
Flight Number vs. Orbit Type

- You should see that in the LEO orbit the Success appears related to the number of flights.
- There seems to be no relationship between flight number when in GTO orbit.



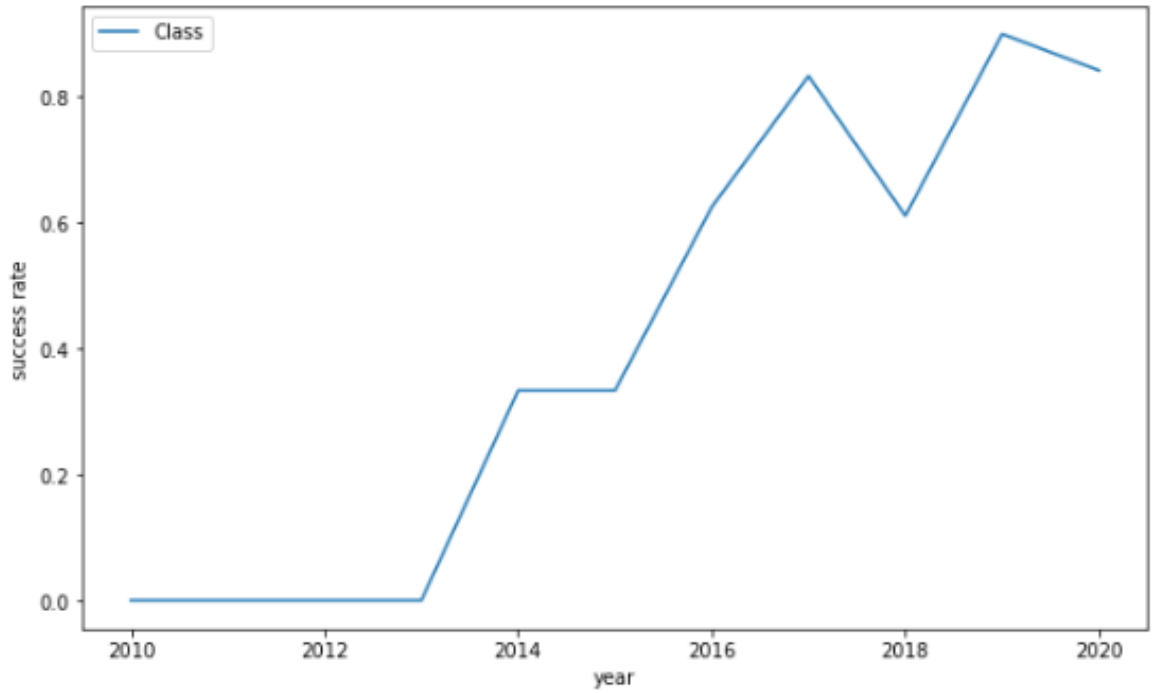
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- For GTO we cannot distinguish this well as both positive landing rate and negative landing are both here.
- For SSO, there is complete positive landing rate with less payload mass.



Launch Success Yearly Trend

- Sucess rate since 2013 kept increasing till 2020
- There is a fall in success rate in 2018.



All Launch Site Names

- There are 4 unique launch sites.
- The Unique Launch site names are:
  - **CCAFS LC-40**
  - **VAFB SLC-4E**
  - **KSC LC-39A**
  - **CCAFS SLC-40**

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA` are:

[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload carried by boosters from NASA is **45596 kg**.
- The result is arrived by adding (using the function sum())the payload mass for NASA(CRS) in SQL.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is **2534.67 kg**.
- The result is arrived by finding mean (using the function AVG()) for the booster version F9 v1.1 in SQL.

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad

01/05/2017

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which had successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- **F9 FT B1022**
- **F9 FT B1026**
- **F9 FT B1021.2**
- **F9 FT B1031.2**

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes are:

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

Boosters Carried Maximum Payload

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600

2015 Launch Records

The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

substr(Date,4,2)	Booster_Version	Landing _Outcome	Launch_Site
01	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

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

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, in descending order

count("Landing_Outcome")	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
20	07-08-2018	05:18:00	F9 B5 B1046.2	CCAFS SLC-40	Merah Putih	5800	GTO	Telkom Indonesia	Success	Success


Section 3
Launch Sites
Proximities Analysis

All Launch sites

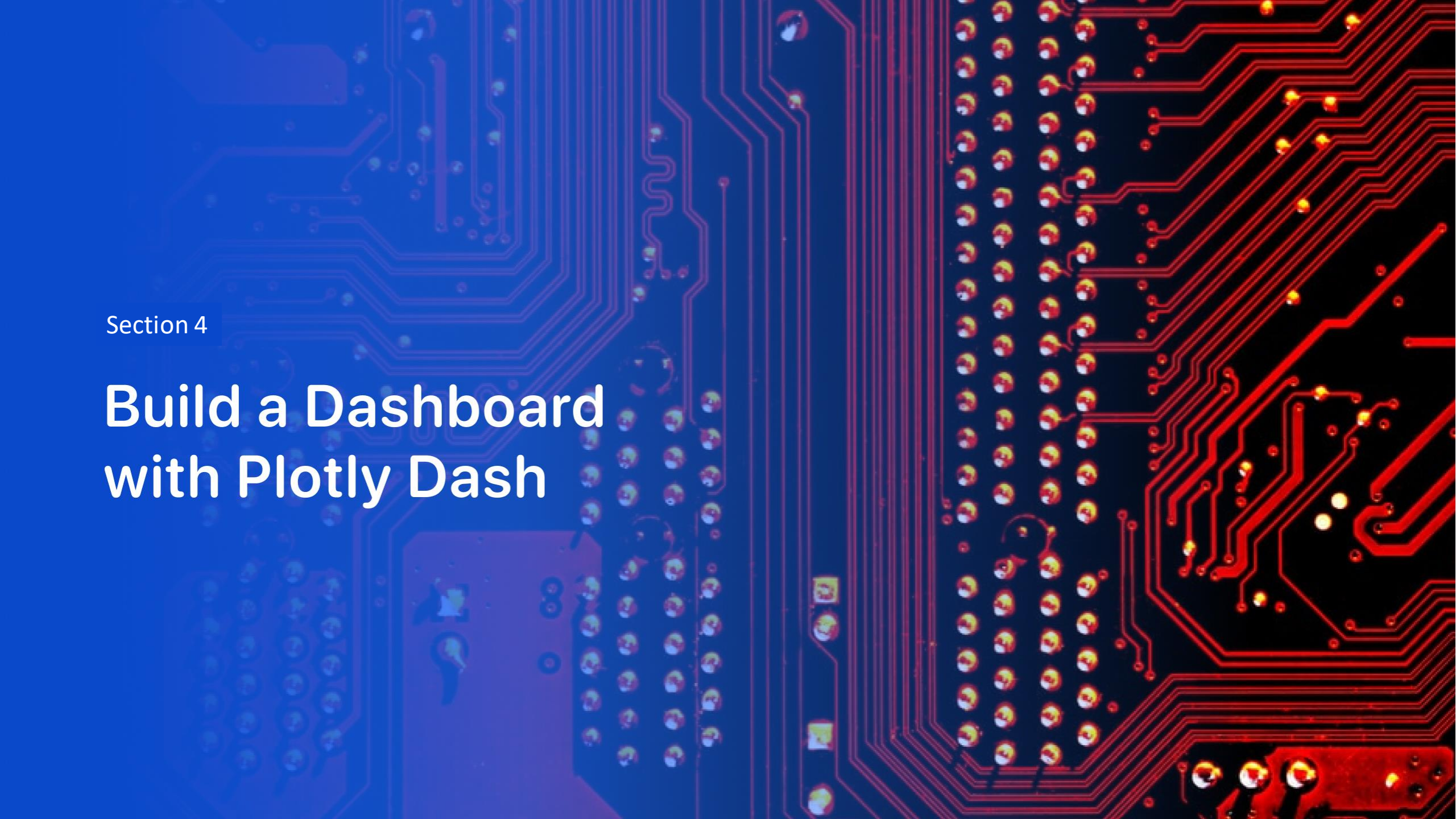
4 launch sites are marked in the map.

Success/Failed launches for each site

- Green marker represents successful launch.
- Red marker represents failed launch.

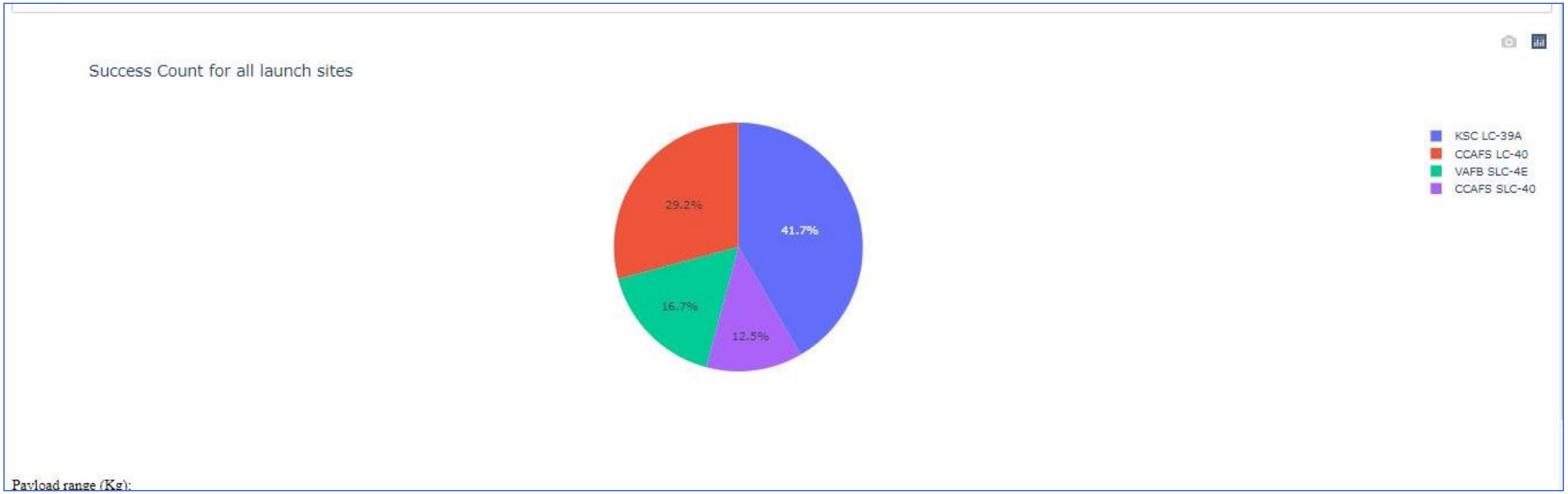
Launch site and its proximities

- The launch site shown in the map is CCAFS – LC40.
- Nearest coast is at 0.86km.
- Nearest city is at 0.86km.
- Nearest railway line is at 1.29km.
- Nearest highway is at 0.58 km.


Section 4
Build a Dashboard with Plotly Dash

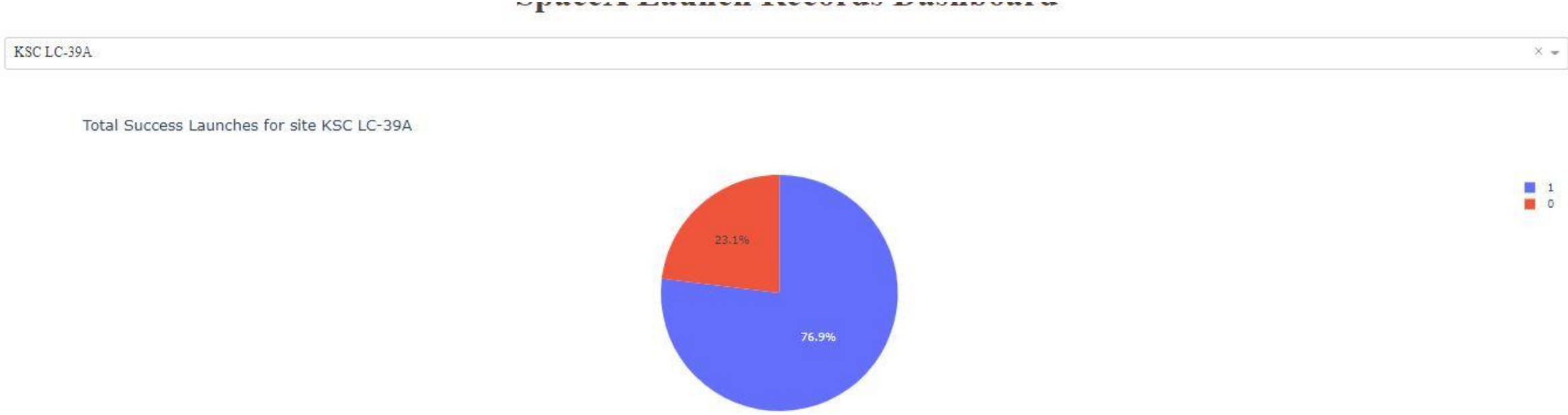
Success count for all sites

- Launchsite **KSC LC-39A** has the **highest success count of 41.7%.**
- Launchsite **CCAFS SLC-40** has the **lowest success count of 12.5%**.



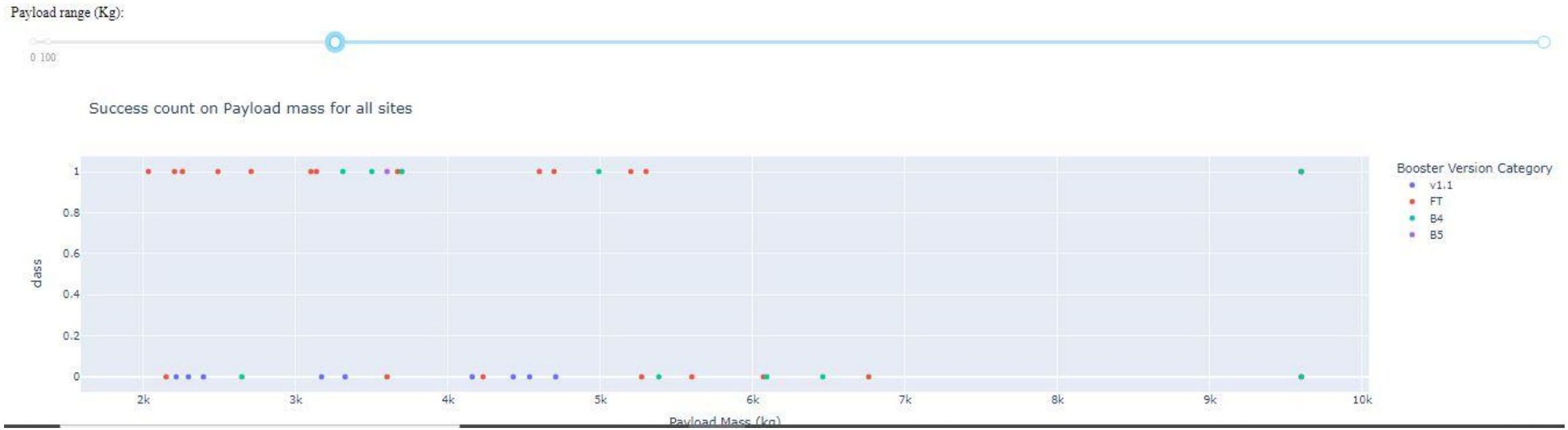
Launch site with highest launch success ratio

- Launch site with **highest launch success ratio** is **KSC LC-39A.**
- Success ratio is **76.9%.**



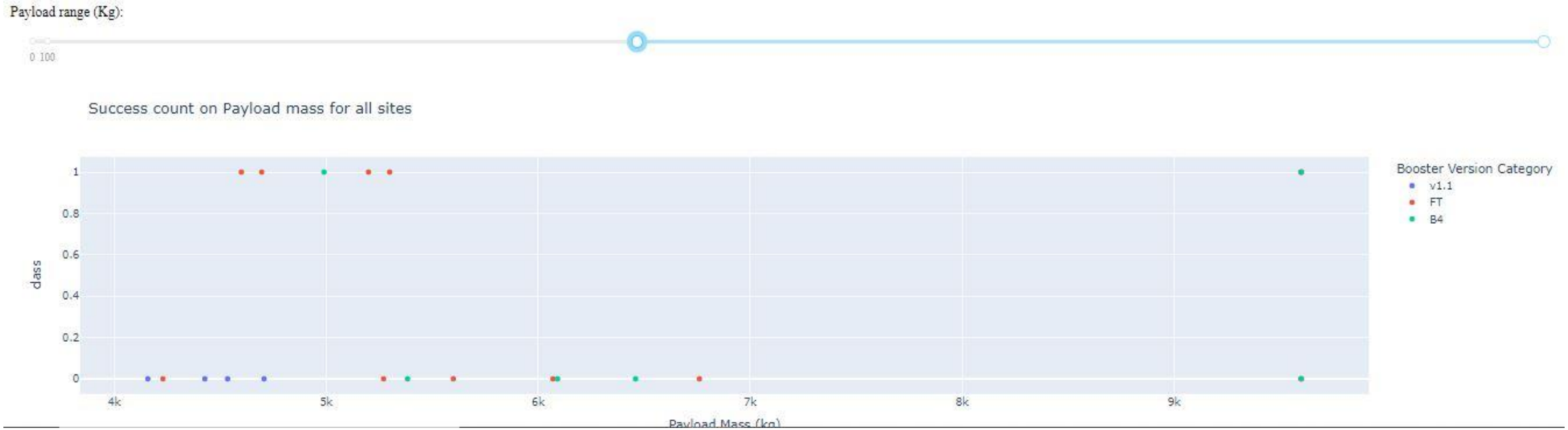
Payload vs. Launch Outcome scatter plot for all sites

The payload range(s) that has the **highest** launch success rate is **3000-4000 kg.**



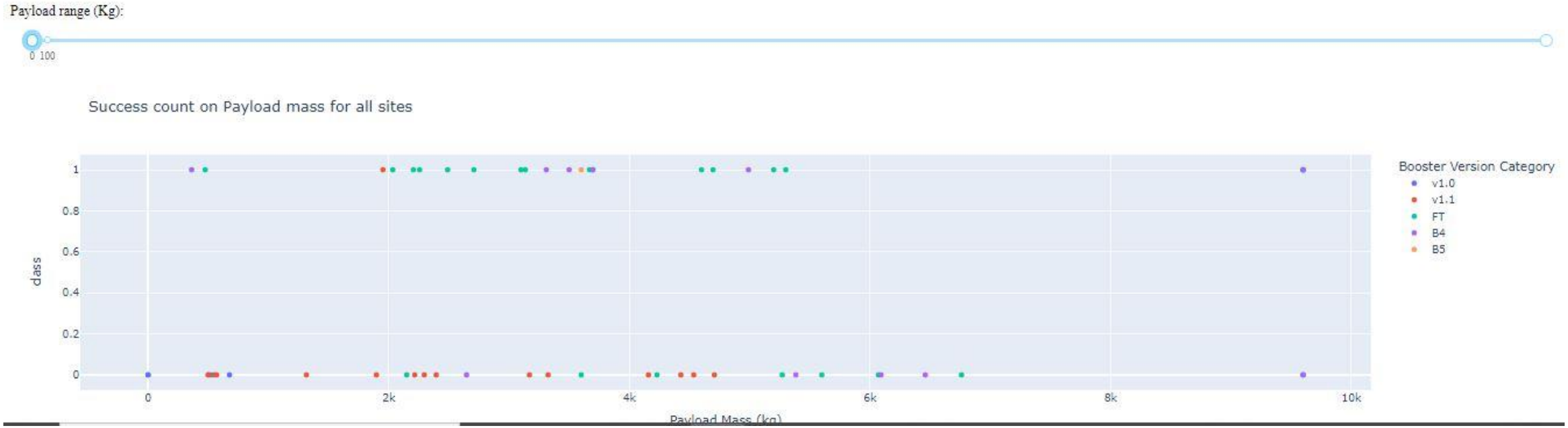
Payload vs. Launch Outcome scatter plot for all sites

The payload range(s) that has the **lowest** launch success rate is **9000-10000 kg.**



Payload vs. Launch Outcome scatter plot for all sites

The **F9 Booster version** (v1.0, v1.1, FT, B4, B5, etc.) that has the **highest** launch success rate is **FT.**



Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Appendix

- Webscraping: List_of_Falcon_9_and_Falcon_Heavy_launches
- API data collection: https://api.spacexdata.com/v4/launches/past

Thank you!