Winning Space Race with Data Science

Michael Quigley
9/10/2021

Outline

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

Executive Summary

- Multiple models were tested to determine which could best predict landing outcome
- Successful landings proved easier to predict than failed landings

Introduction

- The success of SpaceX can be attributed to its reuse of their first stage rockets
- If we can predict whether the first stage landing succeeds we can significantly decrease costs

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected through SpaceX API and web scraping Wikipedia
- Perform data wrangling
  - Categorical landing data converted into binary success/failure
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Performed Grid Search with Logistic regression, SVM, Decision Tree, and KNN to find best fit for data

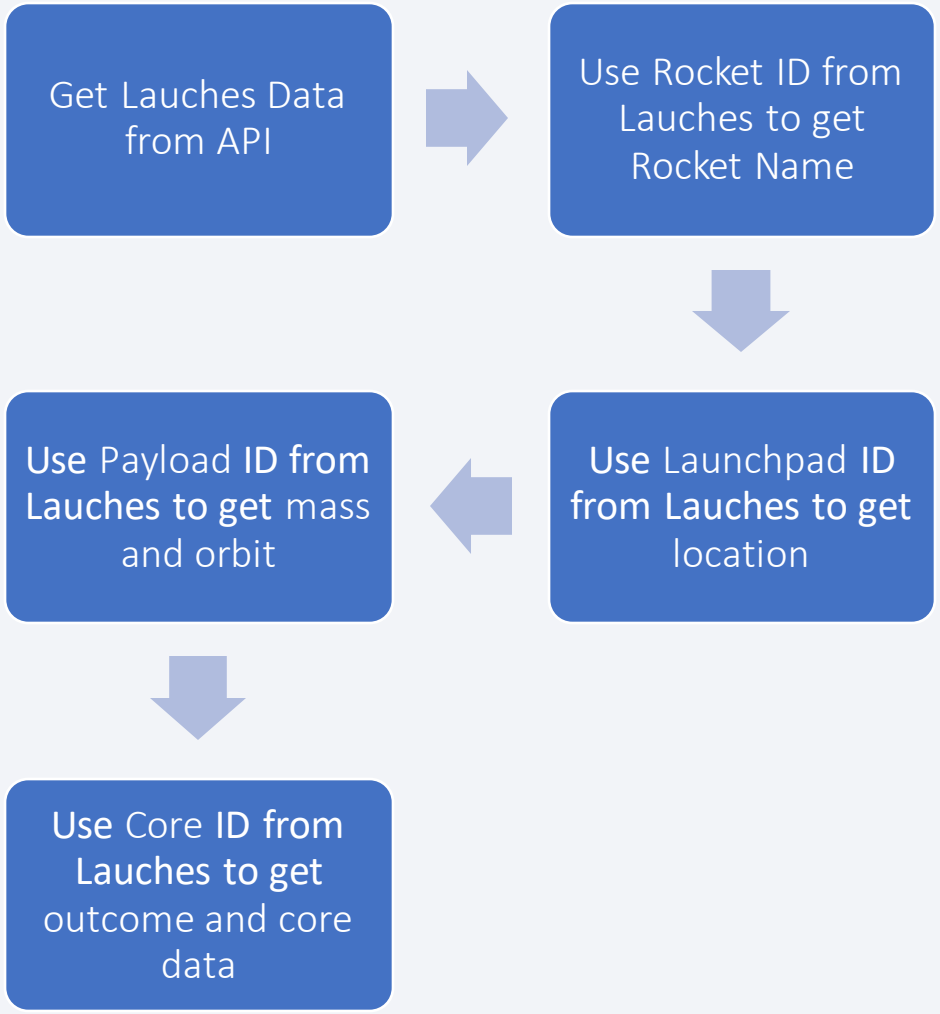
Data Collection

- Data was coloected through the SpaceX API and through Wikipedia.
- API calls were made for each launch to get more specific information.

Data Collection – SpaceX API

- An API call is made to find all launches. Subsequent calls are made using information from the first call.

- https://github.com/MQuigley2/Coursera_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data on SpaceX launches collected via webscraping from Wikipedia

- https://github.com/MQuigley2/Coursera_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Landing Outcome was converted from categorical data into binary success/failure column renamed Class. This new column will be used as the outcome for machine learning analysis.
- https://github.com/MQuigley2/Coursera_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter plots were made comparing different combinations of Flight Number, Paload Mass, Launch SIte and Orbit to see how these variables affect the success of a landing.
- A bar graph was also made to visualize the success rate of different orbits
- A line graph was made to show average success rate for each year
- https://github.com/MQuigley2/Coursera_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Used sql to get aggregate data and group by various labels
- https://github.com/MQuigley2/Coursera_Capstone/blob/main/eda.ipynb

Build an Interactive Map with Folium

- Lauches and lauch sites were visualized on map with folium
- This allows us to understand logistic issues relating to different launch site
- https://github.com/MQuigley2/Coursera_Capstone/blob/main/folium.ipynb

Build a Dashboard with Plotly Dash

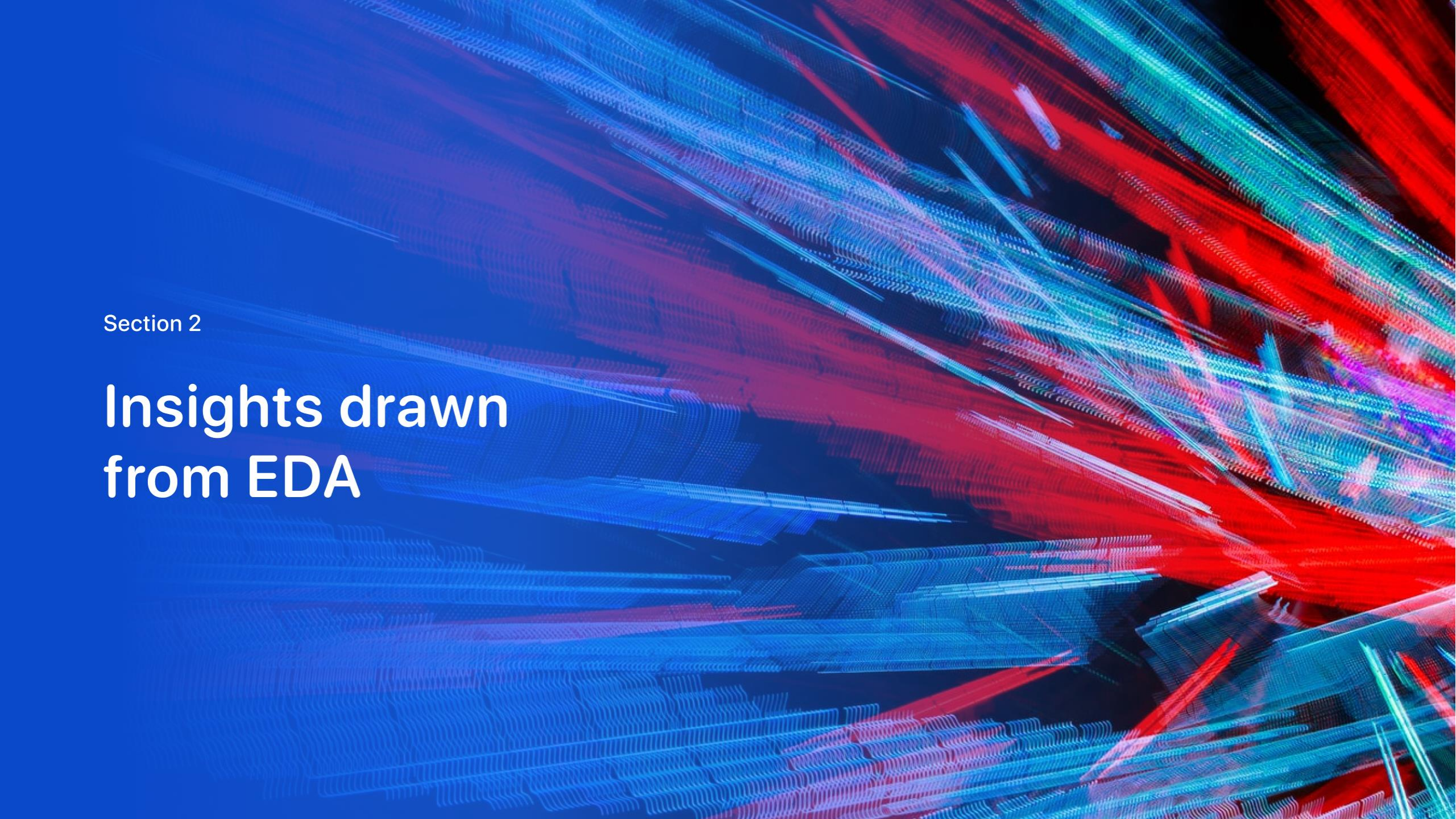
- Used Plotly dash to create dashboard displaying pie chart of success rate at different launch sites and scatter plot displaying success against payload mass
- The dashboard allows you to easily view different chunks of data without needing to make a new plot every time
- https://github.com/MQuigley2/Coursera_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

- The most significant trend found in exploratory data analysis was the massive increase in success rate with respect to time between 2013 and 2020
- We were able to accurately predict succesful landings, but failed landings proved more difficult to predict

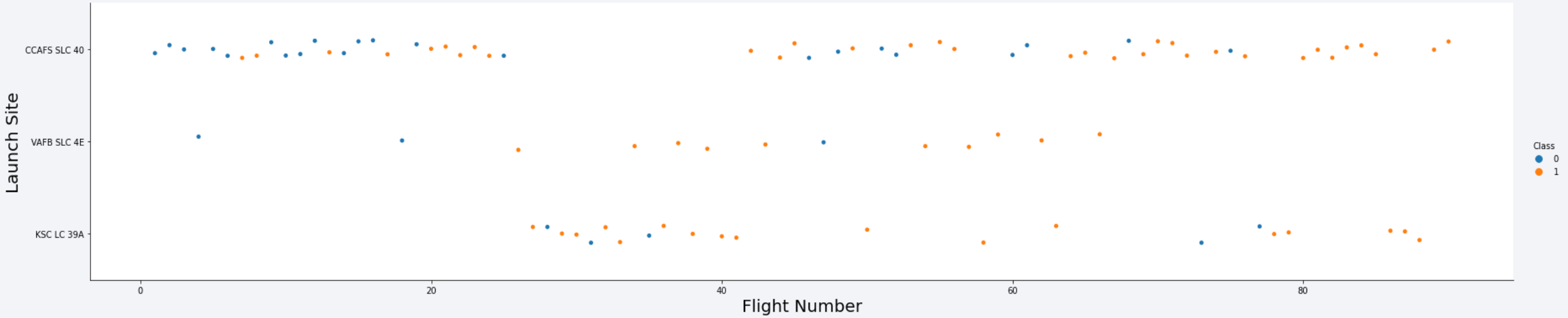


Section 2

Insights drawn from EDA

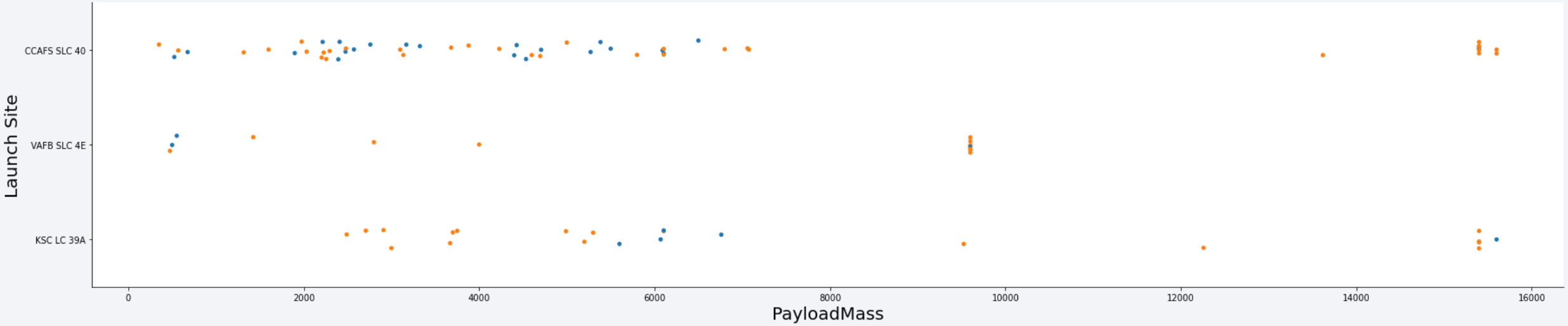
Flight Number vs. Launch Site

- Here the Launch SIte is plotted against the flight number with color denoting success



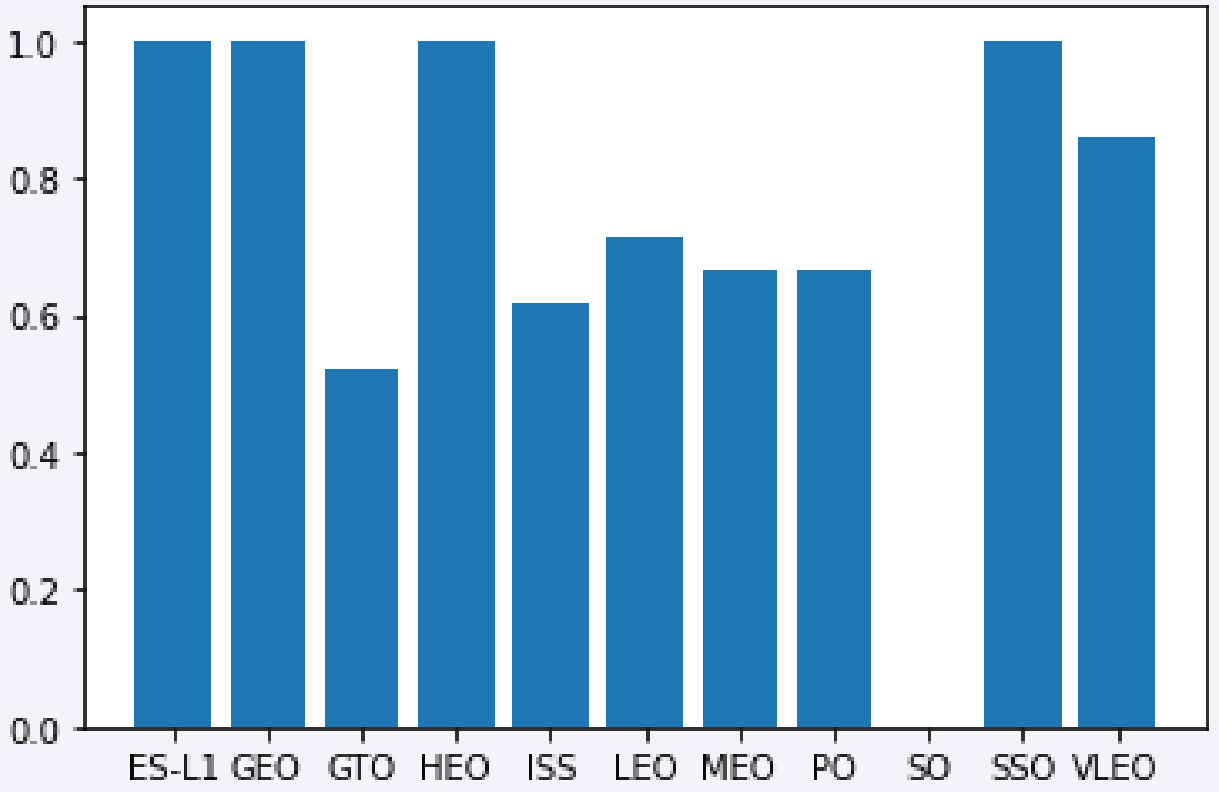
Payload vs. Launch Site

- Here the Launch Site is plotted against the payload mass with color denoting success



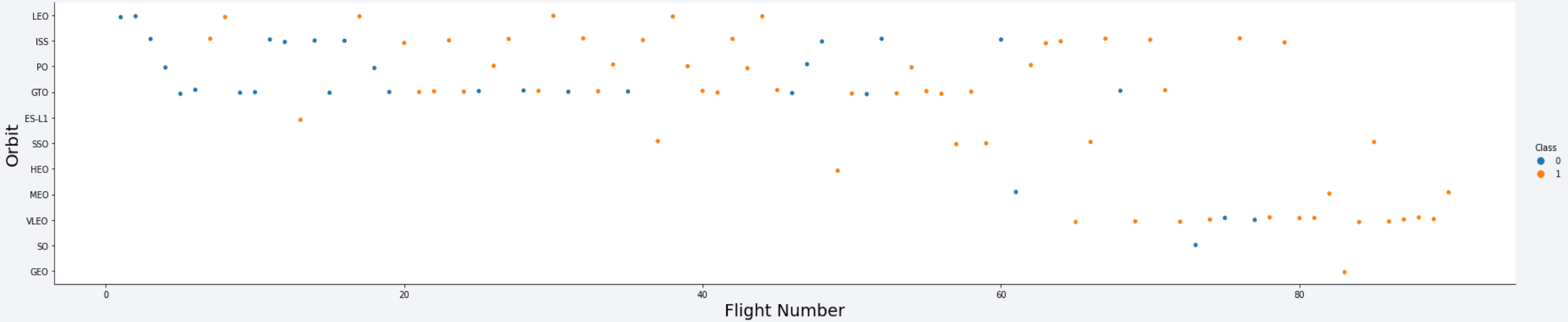
Success Rate vs. Orbit Type

- Here is the average success rate for each orbit type



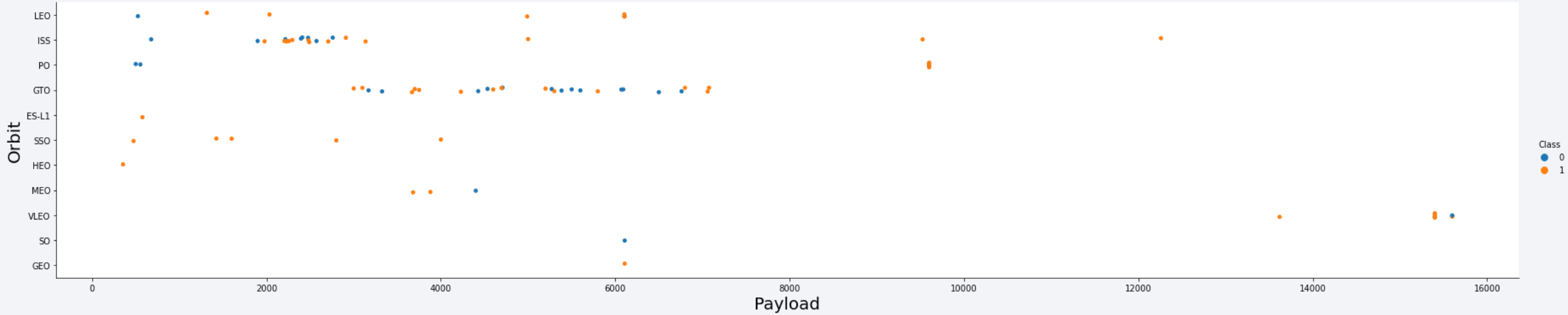
Flight Number vs. Orbit Type

- Here the orbit type is plotted against the flight number with color denoting success



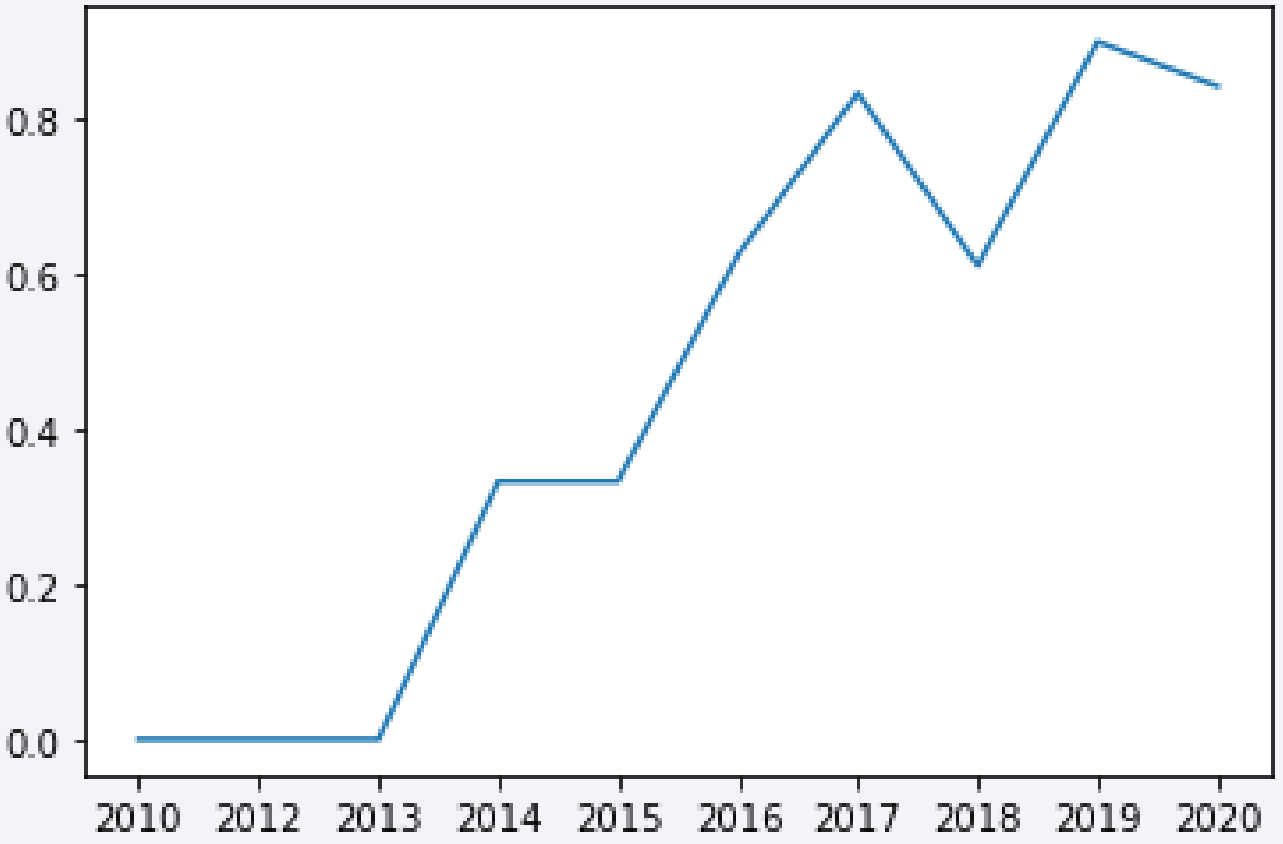
Payload vs. Orbit Type

- Here the orbit type is plotted against the payloadr with color denoting success



Launch Success Yearly Trend

- Here the average success rate is plotted against the year.
- This shows that the success rate has increased significantly from 2013 to 2020



All Launch Site Names

- Here are the unique launch sites used by SpaceX

- SELECT UNIQUE launch_site

  FROM SPACEXTBL;

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

Launch Site Names Begin with 'CCA'

- Here are the first 5 records with lauch site beginning with 'CCA'`
- Some columns are omited to save space

```
SELECT DATE,launch_site,payload_mass__kg_,orbit,mission_outcome,landing__outcome
FROM SPACEXTBL
WHERE launch_site LIKE 'CCA%'
LIMIT 5;
```

Click to add text

DATE	launch_site	payload_mass__kg_	orbit	mission_outcome	landing__outcome
2010-06-04	CCAFS LC-40	0	LEO	Success	Failure (parachute)
2010-12-08	CCAFS LC-40	0	LEO (ISS)	Success	Failure (parachute)
2012-05-22	CCAFS LC-40	525	LEO (ISS)	Success	No attempt
2012-10-08	CCAFS LC-40	500	LEO (ISS)	Success	No attempt
2013-03-01	CCAFS LC-40	677	LEO (ISS)	Success	No attempt

Total Payload Mass

- Here is the total payload carried by boosters from NASA (CRS)

```
SELECT SUM(payload_mass__kg_) AS Total_Payload
FROM SPACEXTBL
WHERE customer='NASA (CRS)';
```

total_payload
45596

Average Payload Mass by F9 v1.1

- Here the average payload mass is determined for the F9 v1.1

```
SELECT AVG(payload_mass__kg_) AS Total
FROM SPACEXTBL
WHERE booster_version LIKE 'F9 v1.1%';
```

average
2534

First Successful Ground Landing Date

- Here is a query to find the ealiest succesful landing on a ground pad

```
SELECT MIN(DATE) AS earliest_success
FROM SPACEXTBL
WHERE landing__outcome='Success (ground pad)';
```

earliest_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Here are all the booster versions which have succesfully landed on a drone ship with a payload between 4000 and 6000 kg

```
SELECT booster_version
FROM SPACEXTBL
WHERE landing__outcome='Success (drone ship)'
AND payload_mass__kg_ BETWEEN 4000 AND 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

- Total number of missions by outcome

```
SELECT mission_outcome, COUNT(*) as Count
FROM SPACEXTBL
GROUP BY mission_outcome;
```

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Here is a list of all boosters that have carried the maximum payload

```
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE payload_mass__kg_=(SELECT MAX(payload_mass__kg_)
FROM SPACEXTBL);
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- Here are the failed drone ship landing in 2015 with their booster version and launch site

```
SELECT booster_version, launch_site,landing__outcome
FROM SPACEXTBL
WHERE landing__outcome='Failure (drone ship)'
AND YEAR(DATE)='2015';
```

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- Here are the landing outcomes ranked by frequency between 2010-06-04 and 2017-03-20

- SELECT landing__outcome, COUNT(*) as "Count"
- FROM SPACEXTBL
- WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
- GROUP BY landing__outcome
- ORDER BY Count DESC;

landing__outcome	Count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

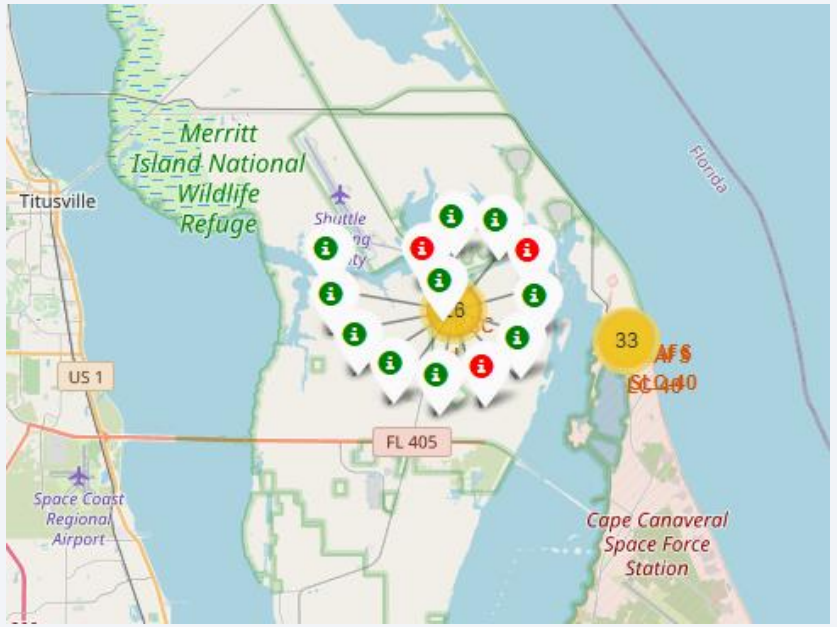
Locations of Launch Sites

- Here is a map showing all SpaceX launch sites

- Launch sites can be found in Florida and Southern California

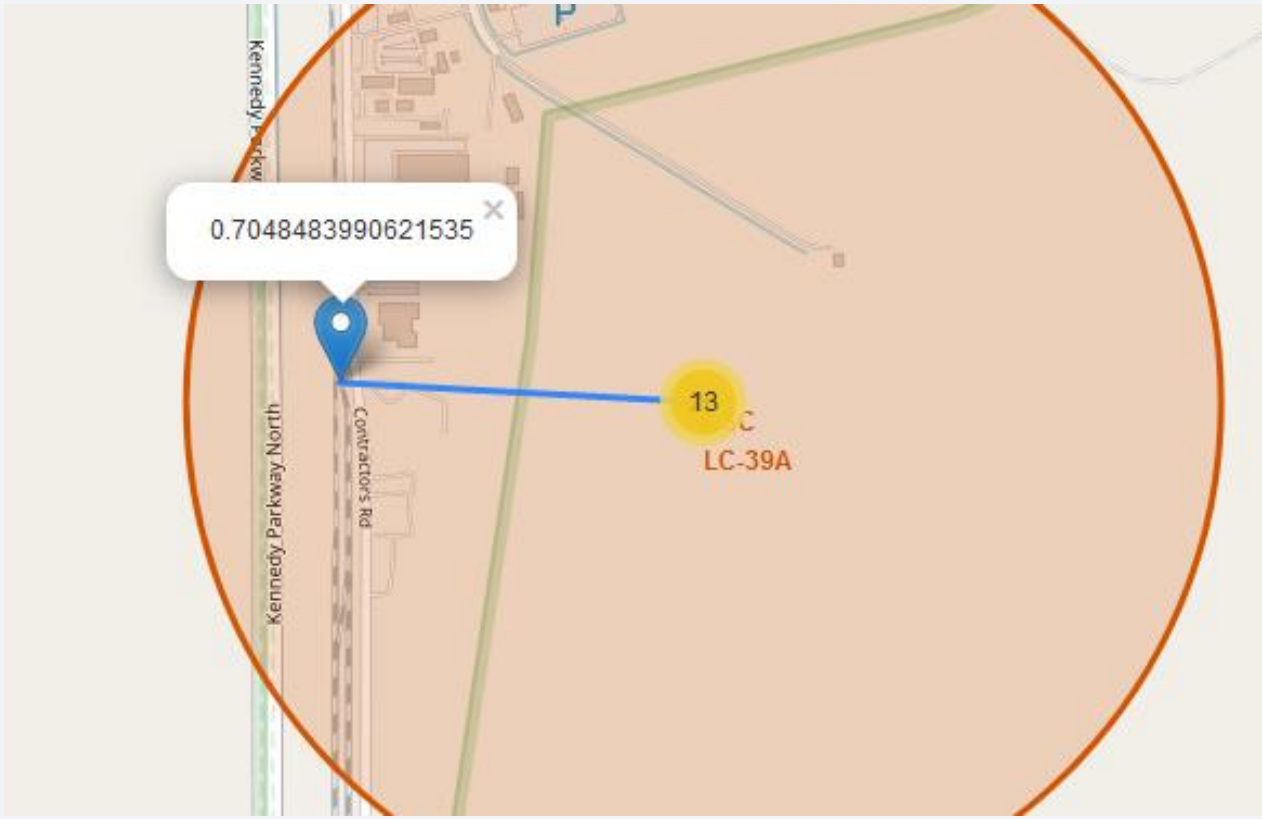
Success and Failure by Site

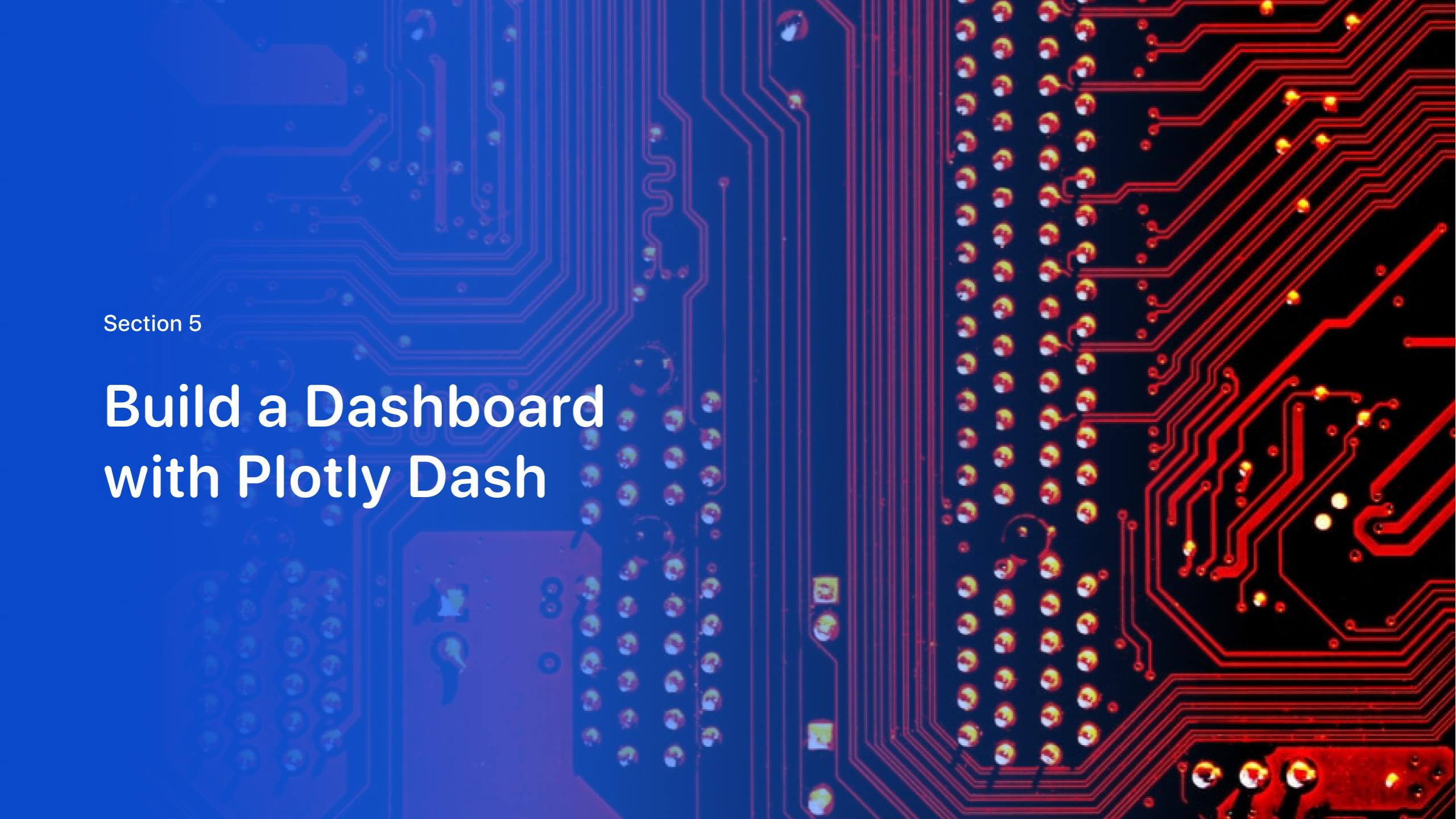
Here the succcesses and failures at each site are displayed as green and red icons on the map



Launch Site Proximity to Rail Line

- Here the launch site KSC LC-39A is displayed on the map with its proximity to the nearest rail line



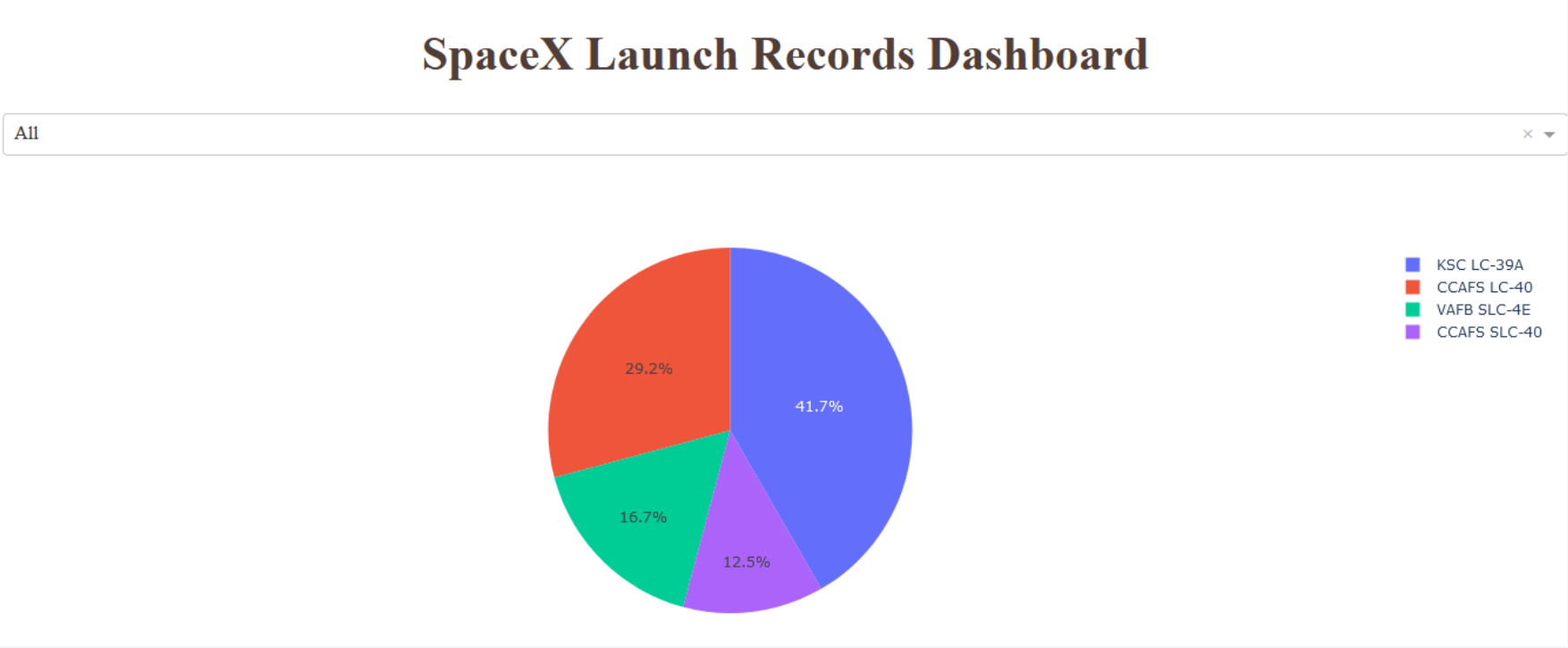


Section 5

Build a Dashboard with Plotly Dash

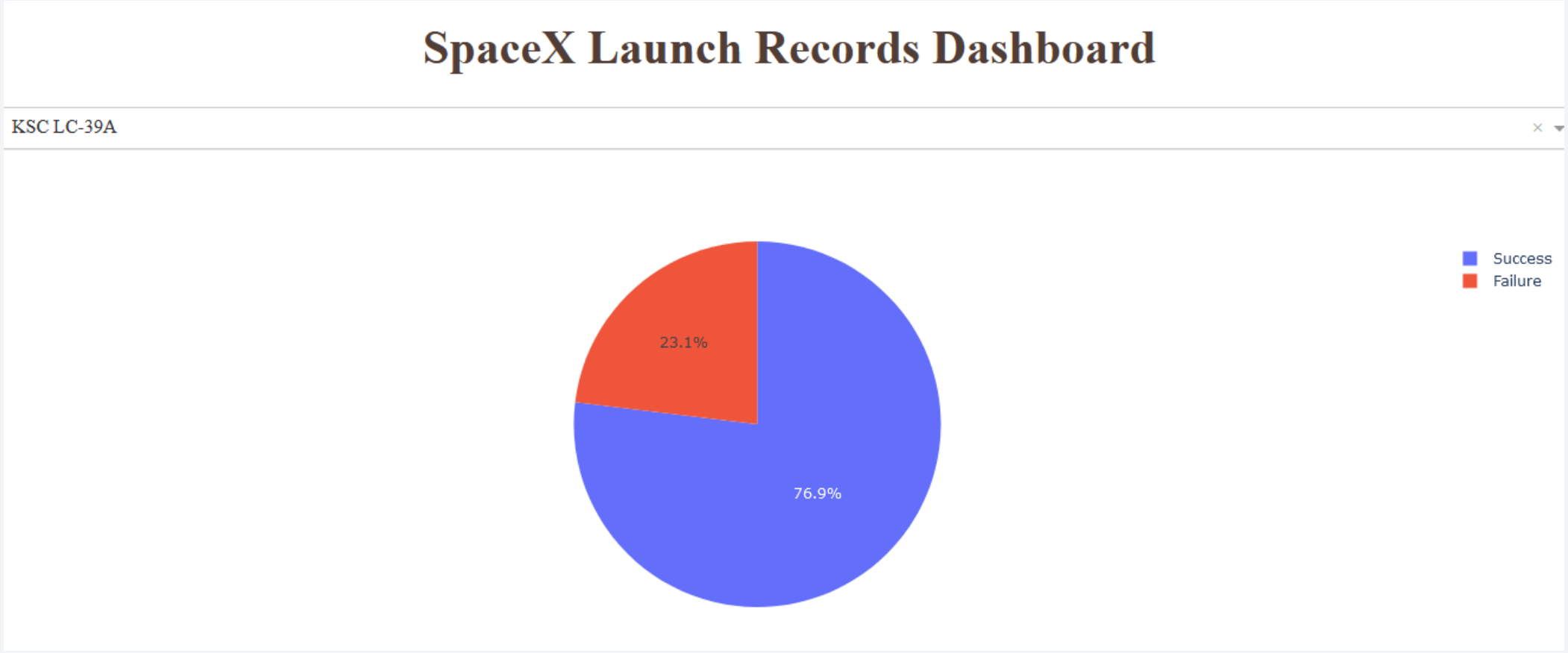
Total Successes by Site

- Here the total successes by site are shown in a pie chart



Success Rate at KSC LC-39A

- Here the successes and failures are shown for the most succesful site (KSC LC-39A)



Success vs. Payload Range

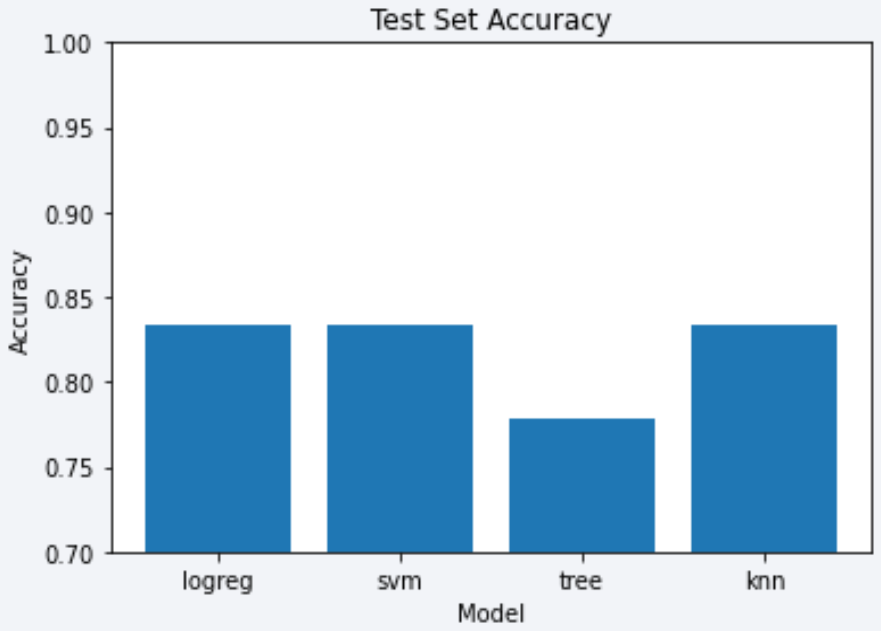
- Here the success rate is plotted against the payload mass. Color is used to show the booster version
- The slider allows for only a specific range of payload mass to be shown a a given time

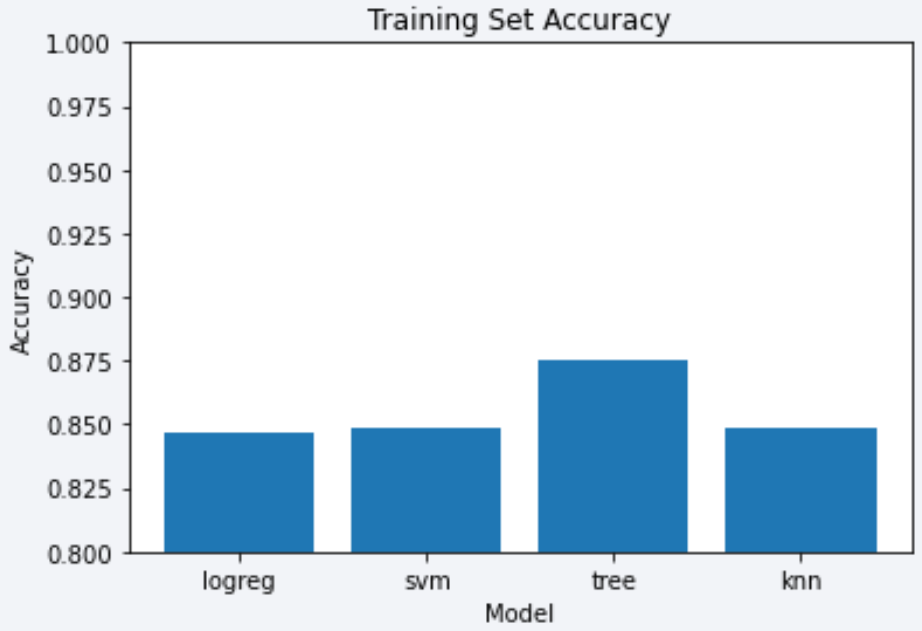
Section 6

Predictive Analysis (Classification)

Classification Accuracy

- Here the training and test accuracies are plotted for each model.
- Due to small sample size, Logistic regression, svm and knn model all had the same test accuracy of .833





Confusion Matrix

- The confusion matrix shows the number of true and false predictions relative to their actual values.
- The model was able to predict every actual landing, but also predicted that half of failed landings were succesful
- Results are the same for each of Logistic regression, KNN and SVM Models

Conclusions

- Our model shows that while the successfuls landings could be easily predicted, failed landings posed more of a challenge
- This indicates that factors unknown to the model may cause some landings to fail
- More data needs to be collected in order to determine what factors besides those tracked could affect landing success
- Weather data may be one factor which our analysis did not take into account

Appendix

- Wikipedia data collected here:

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Thank you!