IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Daniel Yahya
7th July, 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies

  Data was analyzed using the following methodologies:

  - Data Collection using web scraping and SpaceX API
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics
  - Machine Learning Prediction.

- Summary of all results

  - Exploratory Data Analysis (EDA) results
  - Interactive maps and dashboard
  - Machine Learning Prediction results

Introduction

- Project background and context

  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. The aim of this project is to predict if the Falcon 9 first stage will land successfully. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This project is to provide Space Y with useful information to compete against Space X.

- Problems you want to find answers

  1. Determine the price of each launch?
  2. Determine if Space X would reuse the first stage?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Data was collected from two (2) sources:

  - Space X API
  - Web Scraping from Wikipedia page titled List of Falcon 9 and Falcon Heavy launches

- Perform data wrangling
  - After summarizing and analyzing the collected data, it was supplemented by producing a landing outcome label based on the outcome data.
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data acquired up to this point was normalized, separated into training and test data sets, and assessed by four distinct classification models, with the accuracy of each model evaluated using different parameters.

Data Collection

- Data sets were collected from:
  - ❑ Space X API

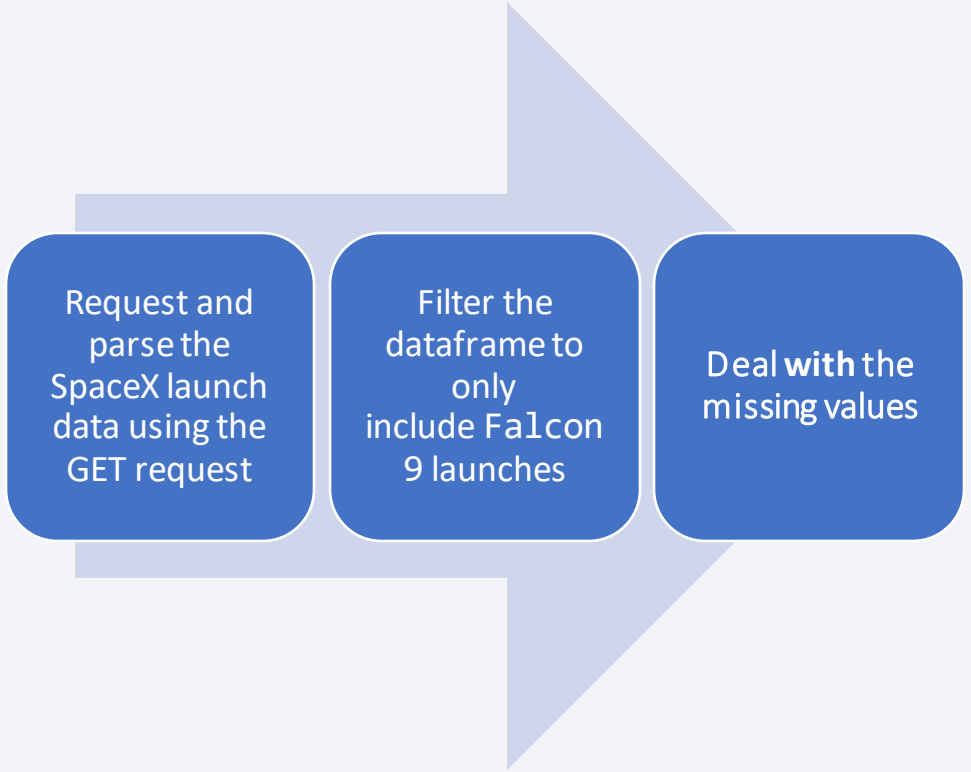
    https://api.spacexdata.com/v4/rockets/ and
  - ❑ web scraping from Wikipedia

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

Data Collection – SpaceX API

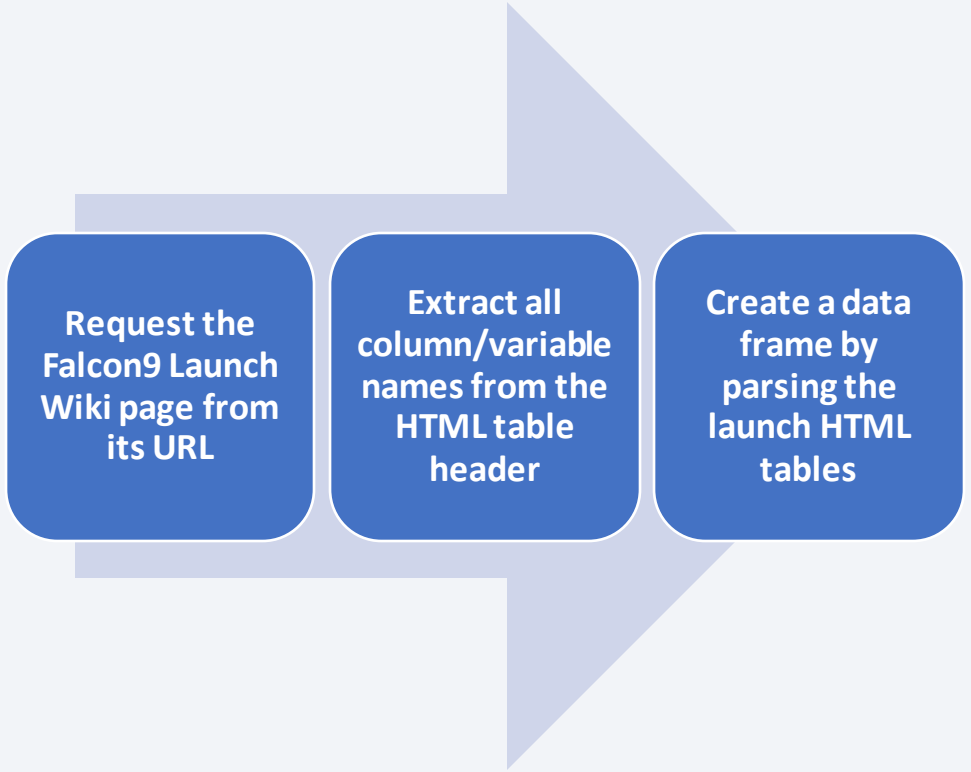
- Data were requested from SpaceX API and some basic data wrangling and formatting was performed on the requested data.

- github link



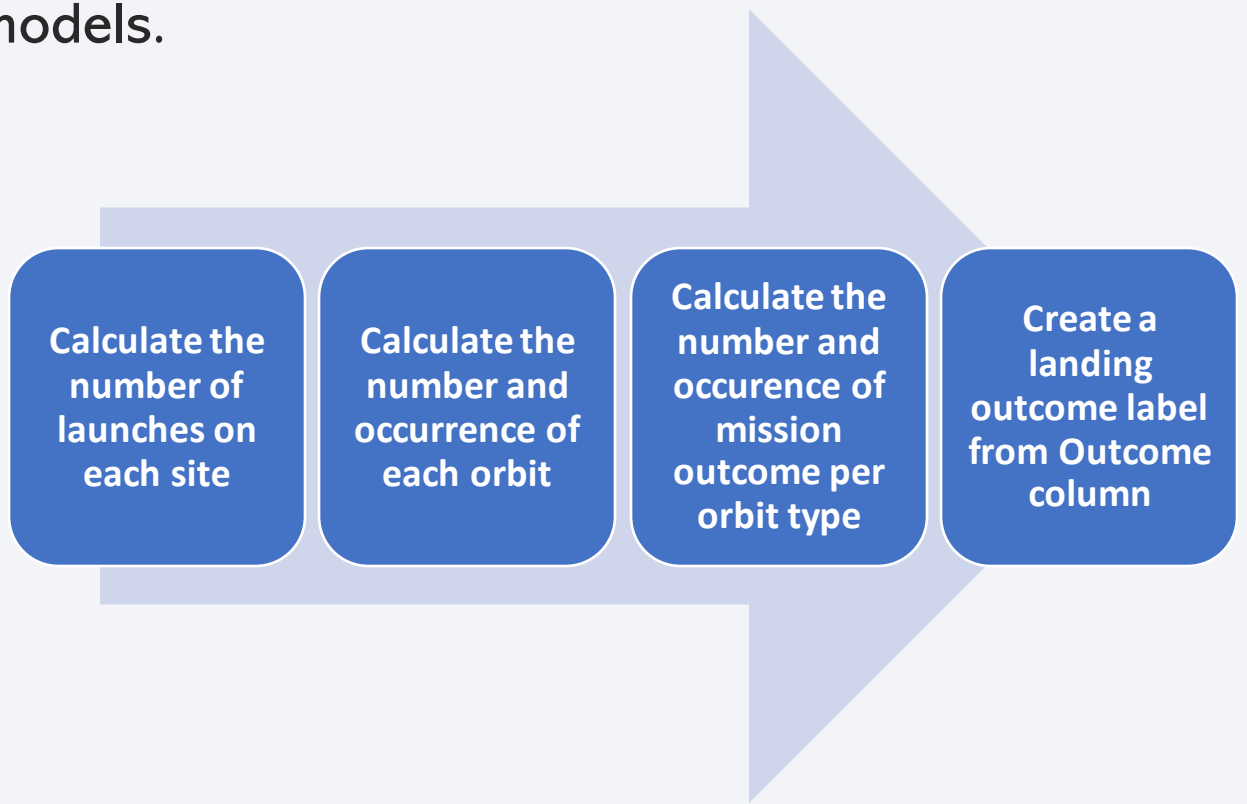
Data Collection - Scraping

- Web scrap Falcon 9 launch records with BeautifulSoup:
- Falcon 9 launch records HTML table was extracted from Wikipedia
- The table was parsed and converted into a Pandas data frame
- [github link]



Data Wrangling

- Some exploratory data analysis was performed to determine label for training supervised models.



- github link

EDA with Data Visualization

- Scatterplot was used to Visualize the relationship between Flight Number and Launch Site, Payload and Launch Site, Flight Number and Orbit type, Payload and Orbit type
- Bar plot was used to Visualize the relationship between success rate of each orbit type
- Line plot was used to Visualize the launch success yearly trend
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

- github link

EDA with SQL

SQL queries performed were:

- *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 acheived.*
- *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 failed landing_outcomes in drone ship, their booster versions, and launch site names for 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

github link

Build an Interactive Map with Folium

- Markers, circles, lines and marker cluster were added to the folium maps.
  - Markers shows points like launch sites, e.g NASA Johnson Space Center.
  - Marker clusters shows multiple and different information for a specific coordinate.
  - Circles shows areas around highlighted coordinates
  - Lines are used to show the distance between two coordinates.

github link

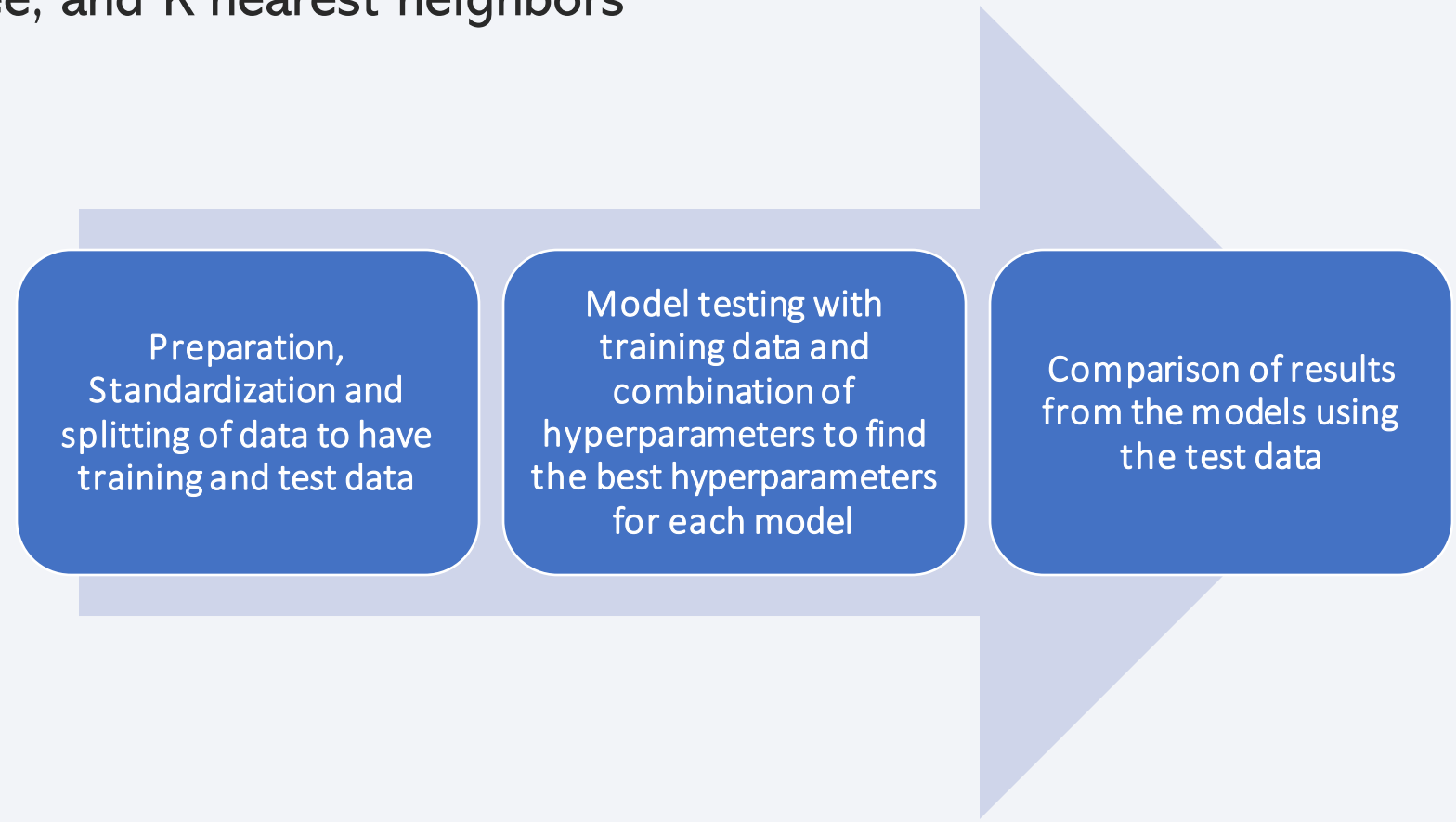
Build a Dashboard with Plotly Dash

- Drop-down, pie chart, scatter plot, range slider were added to the dashboard
  - Drop-down were added for interactive selection of launch site
  - Pie chart were added to show the total success for the selected launch site
  - Range slider were added for interactive payload mass selection
  - Scatter plot were added to show the relationship between Payload and Success of the selected launch site

github link

Predictive Analysis (Classification)

- Evaluation of four models – logistic regression, support vector machine (SVM), decision tree, and K nearest neighbors



Github link

Results

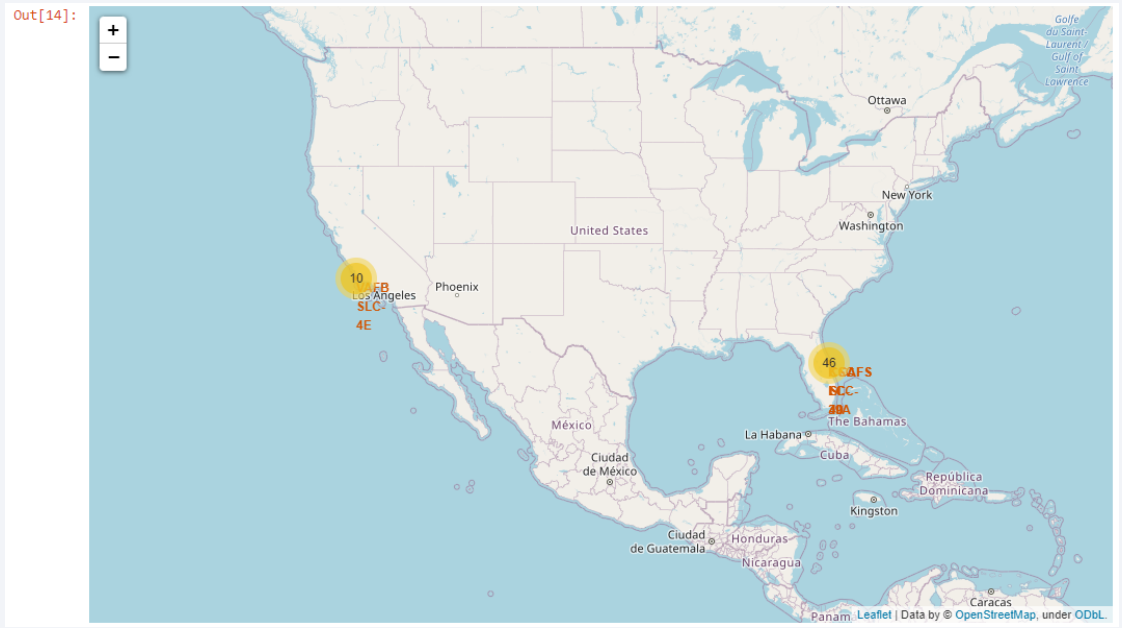
Exploratory data analysis results

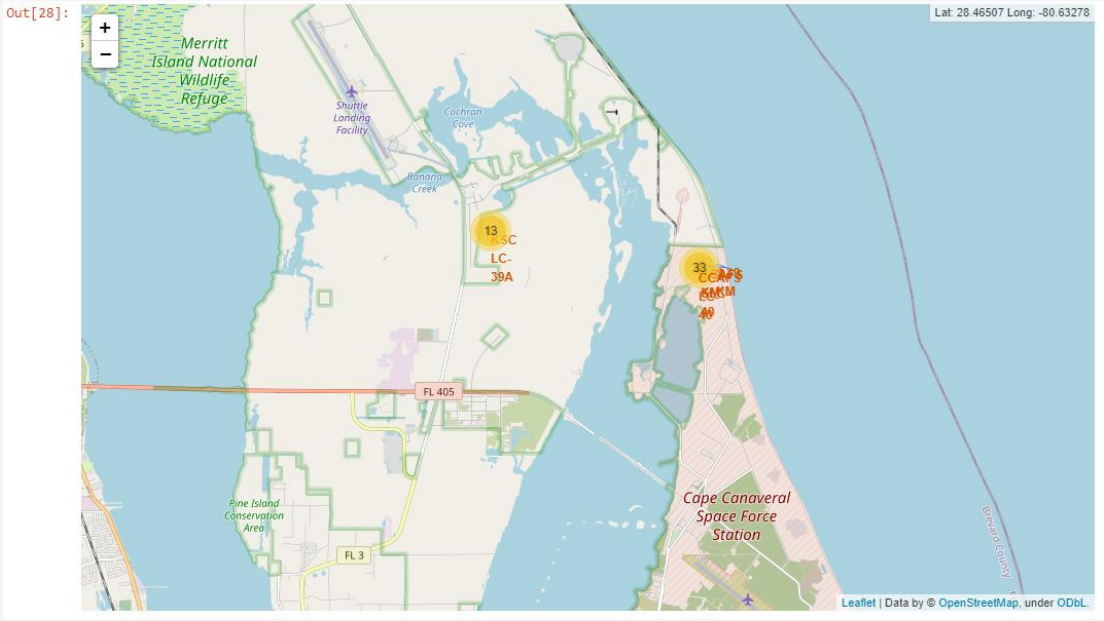
- A landing outcome label was established for each launch
  - landing_class = 0 if bad_outcome
  - landing_class = 1 otherwise

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

Results

- **Interactive analytics** was used to find some geographical patterns about launch sites
- Most launch site are close to the coastline

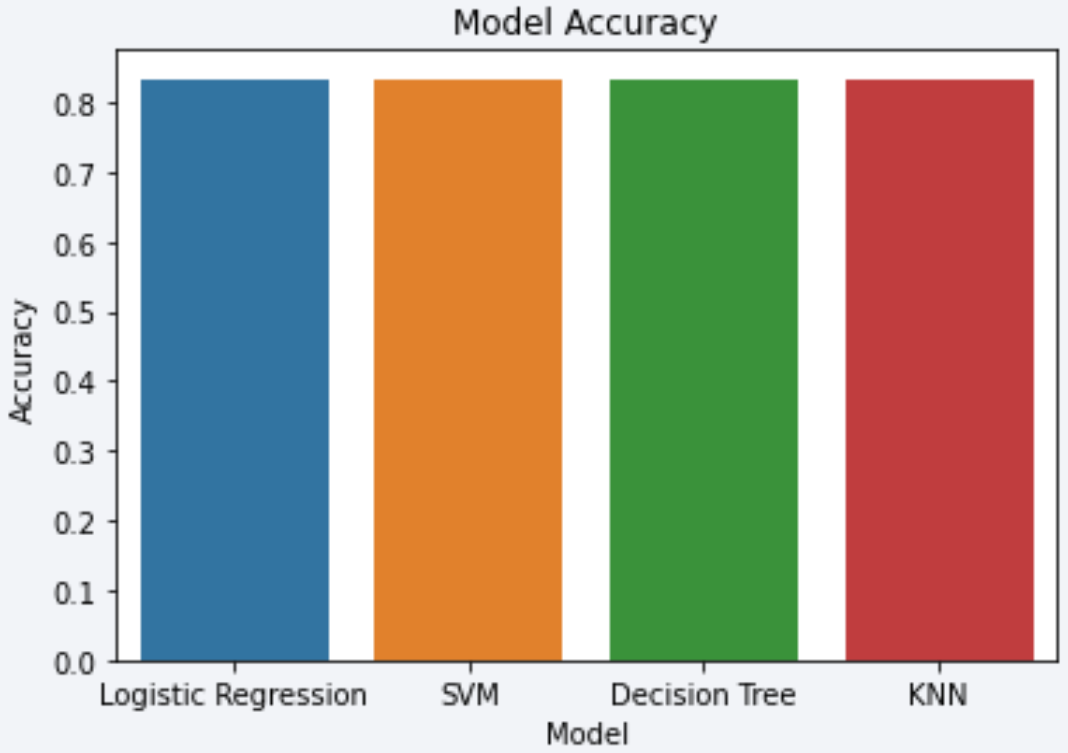




Results

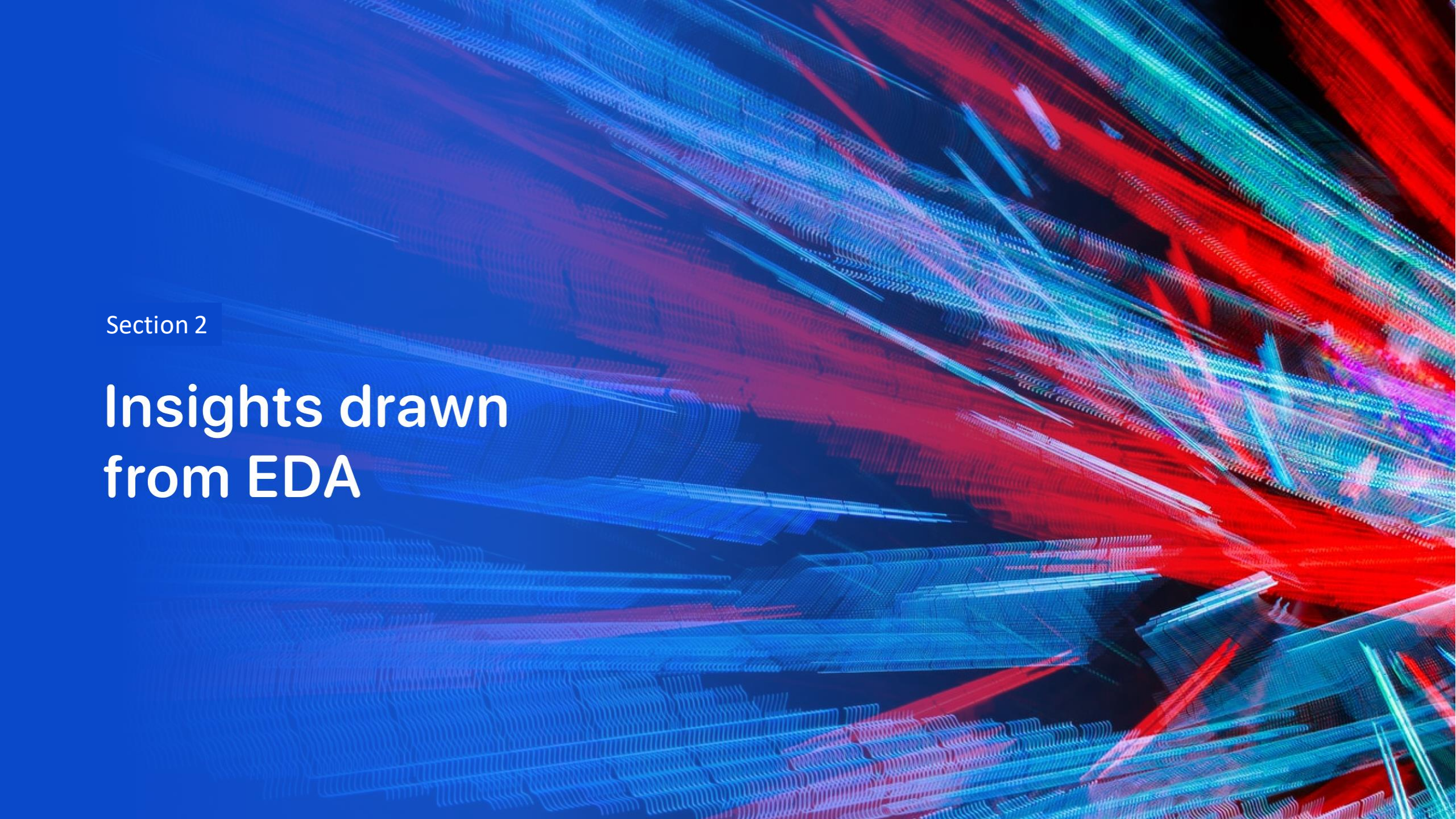
Predictive analysis results

- The result showed that the models' accuracy are similar based on the training and test data.

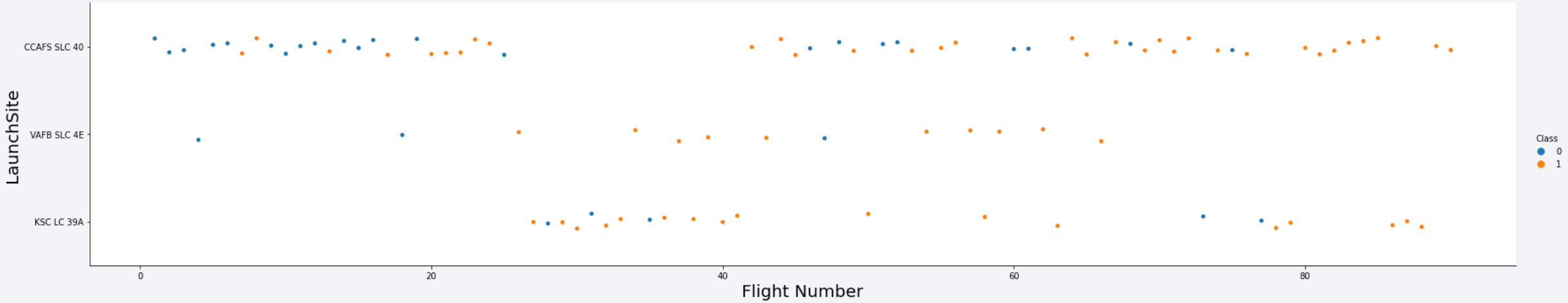


Section 2

Insights drawn from EDA

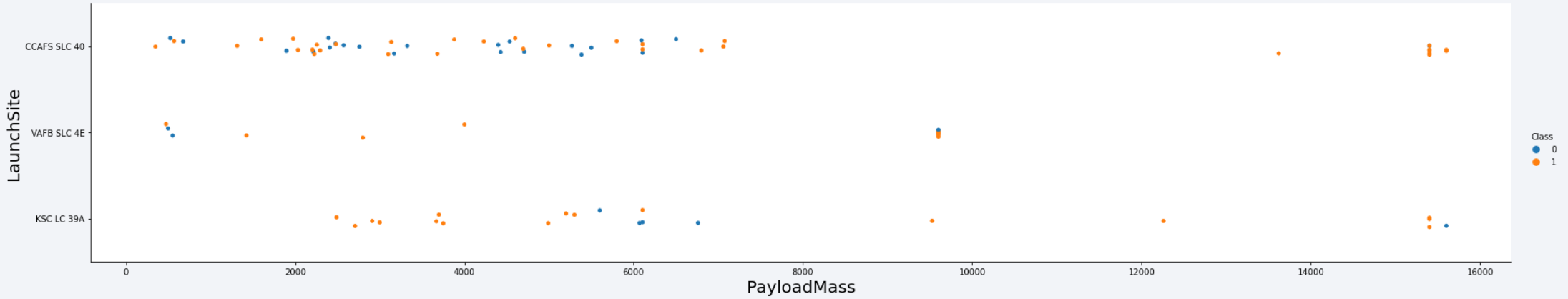


Flight Number vs. Launch Site



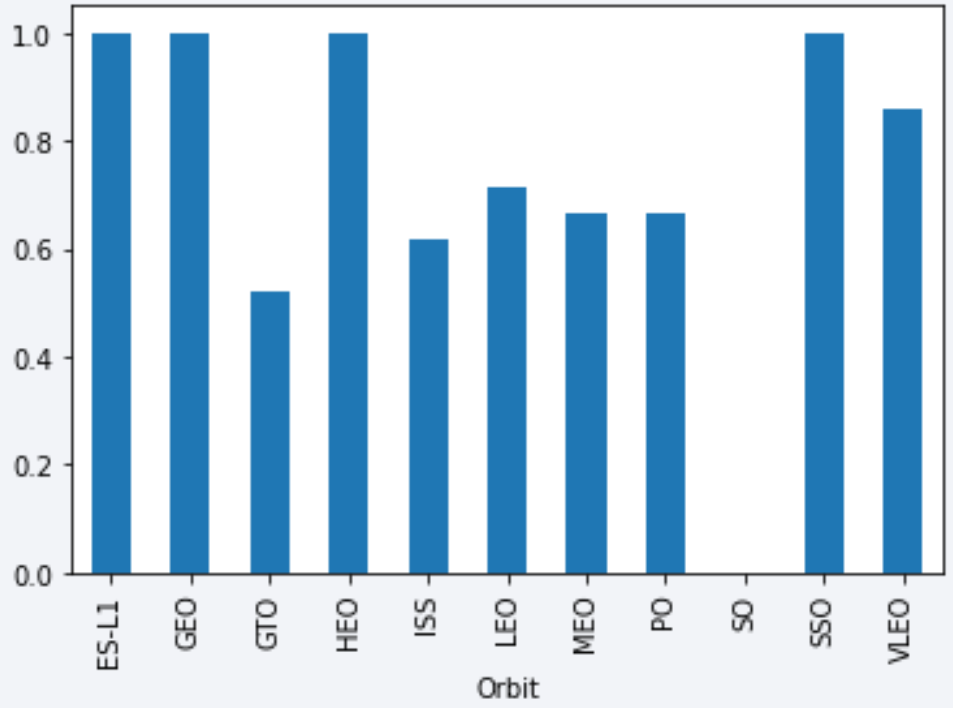
- From the result, each site has an increasing success rate

Payload vs. Launch Site



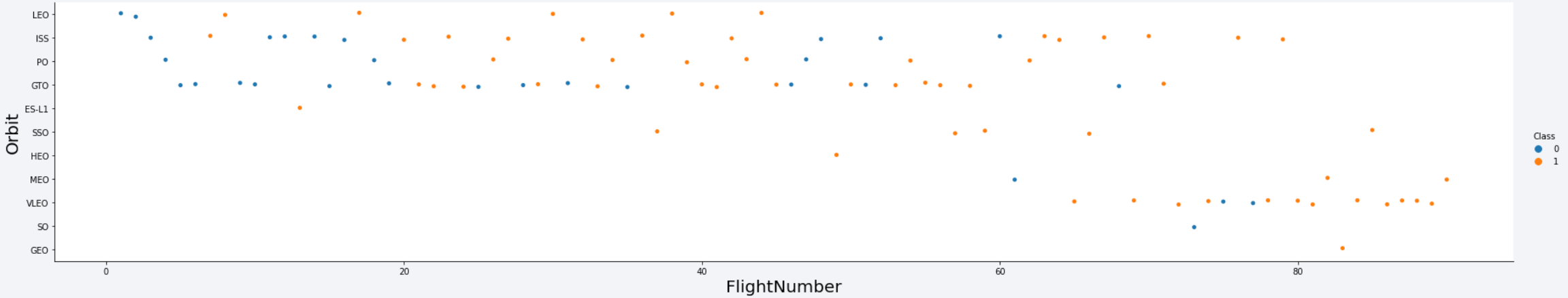
- For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- In addition, depending on the launch site, significant payload mass may be a factor in a successful landing.

Success Rate vs. Orbit Type



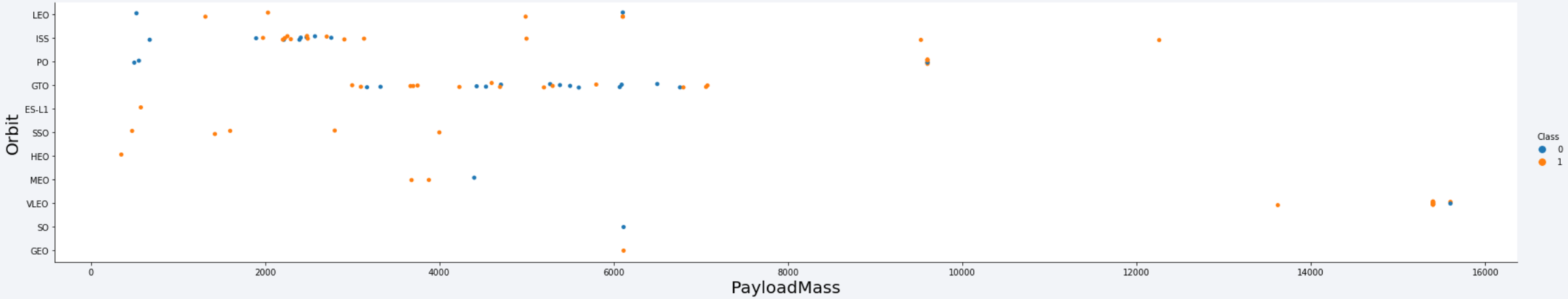
- This plot highlights the success rate of each orbit.
- ES-L1, GEO, HEO, SSO have the highest success rate

Flight Number vs. Orbit Type



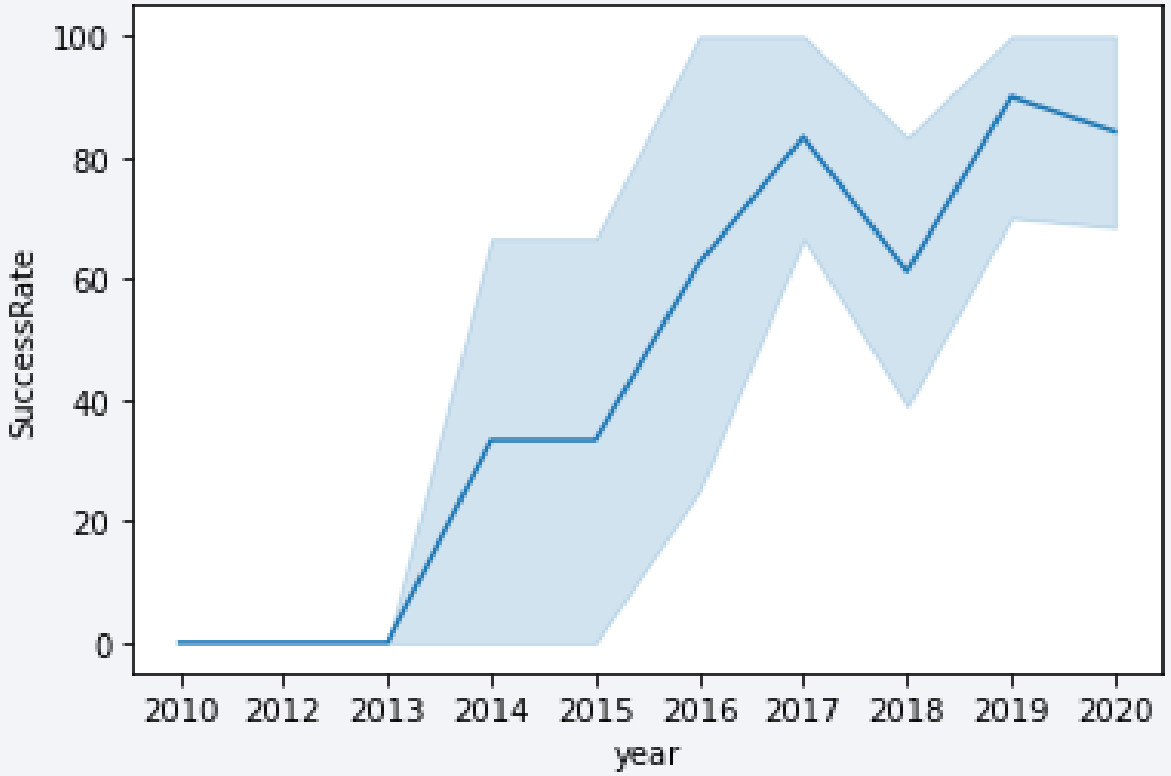
- For the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However, for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there.

Launch Success Yearly Trend



- Sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

- Result was obtained by displaying the names of the unique launch sites in the space mission

Launch Site Names Begin with 'CCA'

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Result was obtained by displaying 5 records where launch sites begin with the string 'CCA'

Total Payload Mass

total_payload (kg)
111268

- The result above shows the total payload mass carried by boosters launched by NASA (CRS)

Average Payload Mass by F9 v1.1

average_payload_mass (kg)
2928

- The result above shows the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

first_successful_landing_gp
2015-12-22

- The first successful landing outcome in ground pad was achieved on 22nd December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- The result above shows the list of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

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

- The result above shows that the total number of successful and failure mission outcomes are 100 and 1, respectively.

Boosters Carried Maximum Payload

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

- The result shows the names of the booster which have carried the maximum payload mass

2015 Launch Records

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

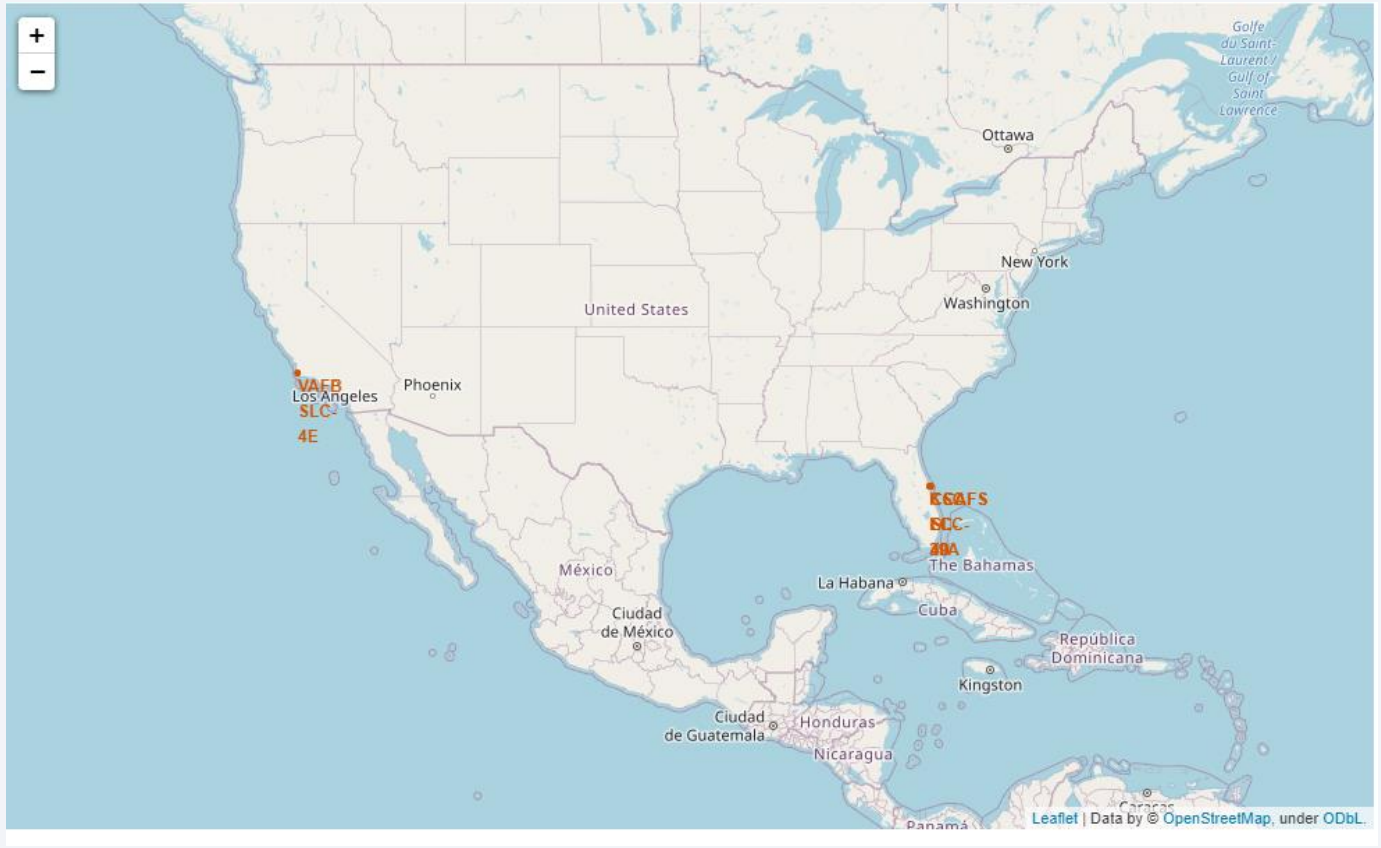
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

landing__outcome	qty
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

- The result above shows the rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

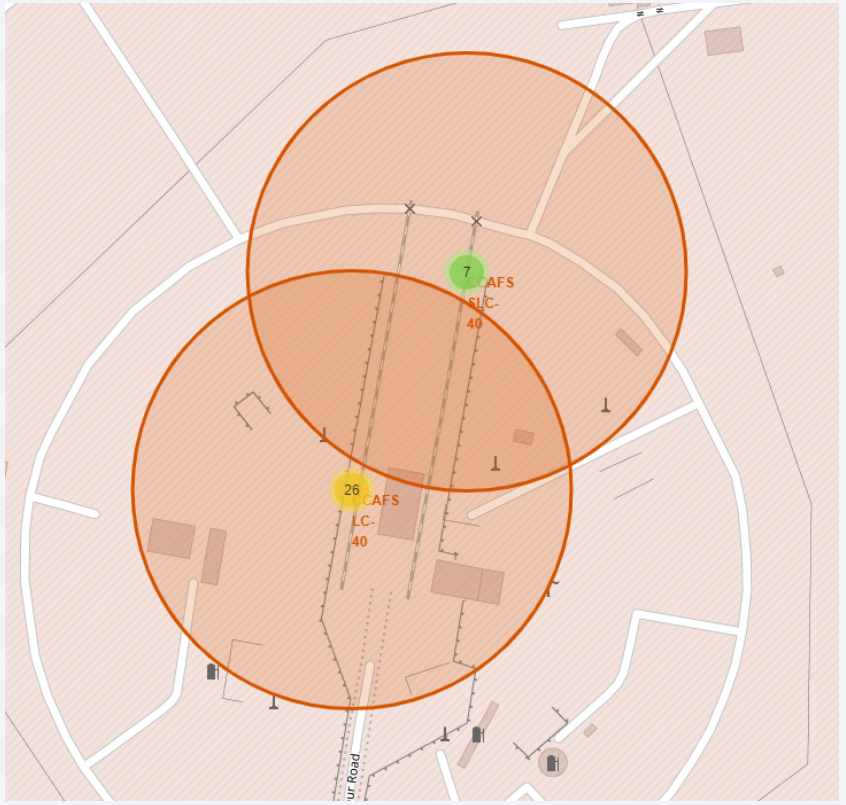

Section 3
Launch Sites
Proximities Analysis

All Launch Sites



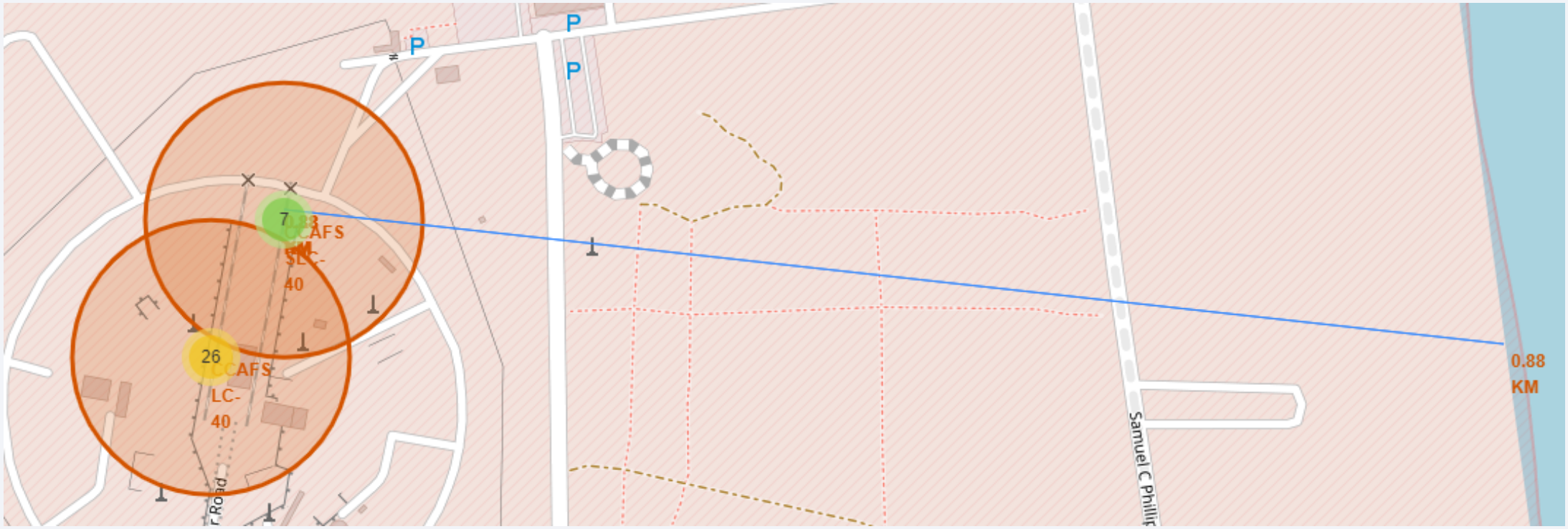
- All launch sites in very close proximity to the coast.

Site Launch Outcomes

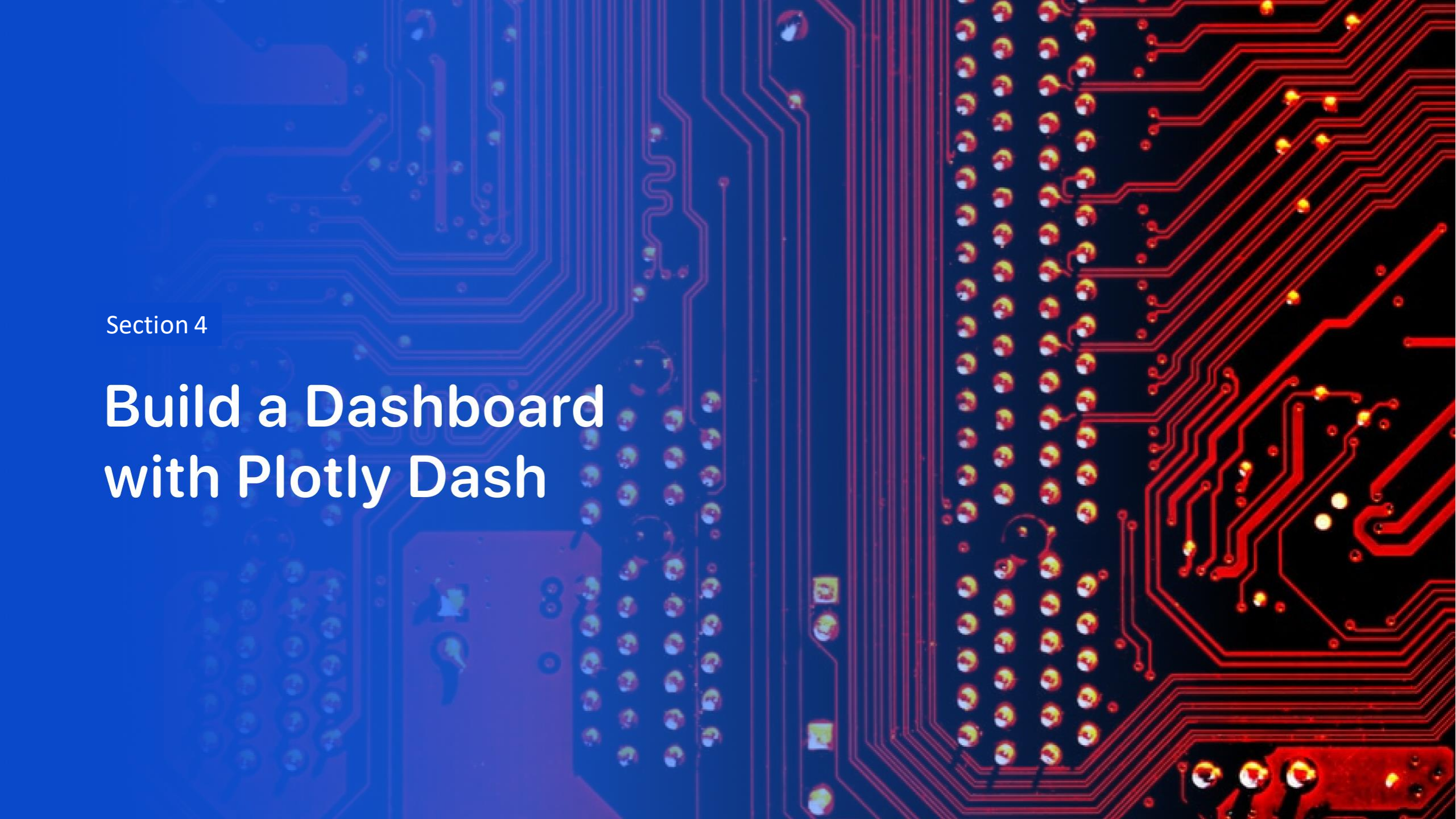


- Green markers indicate successful launch outcome.

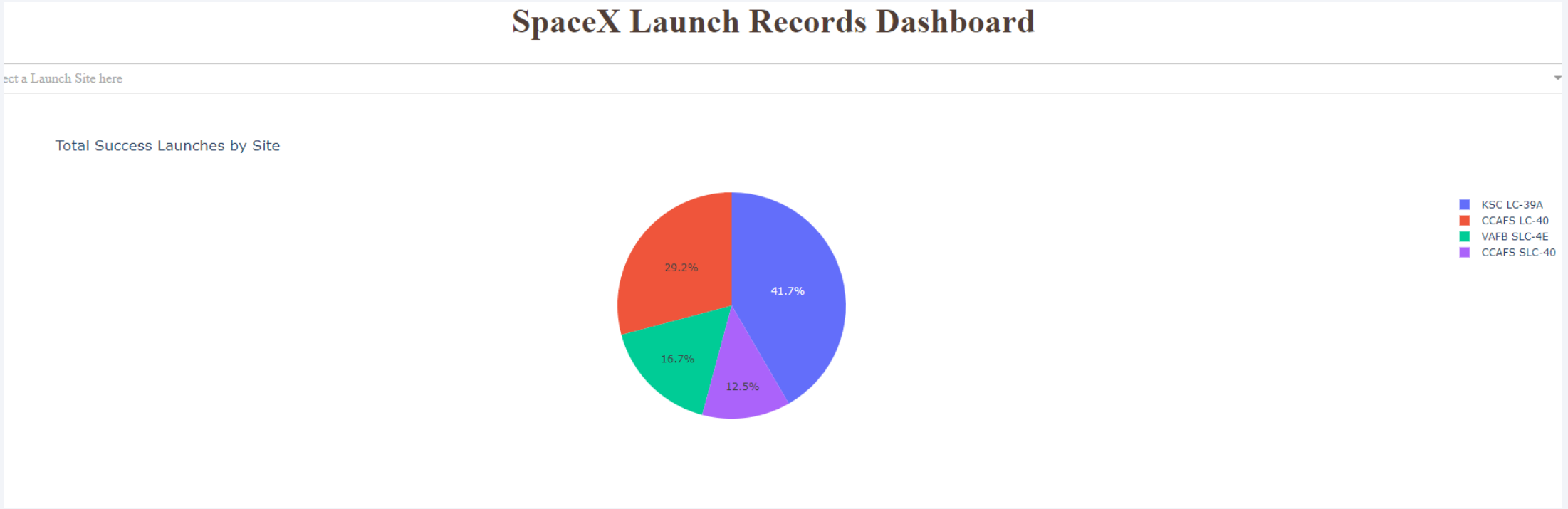
Proximity to Coastline



- CCAFS SLC-40 is close to the coastline with approximate distance of 0.88 km

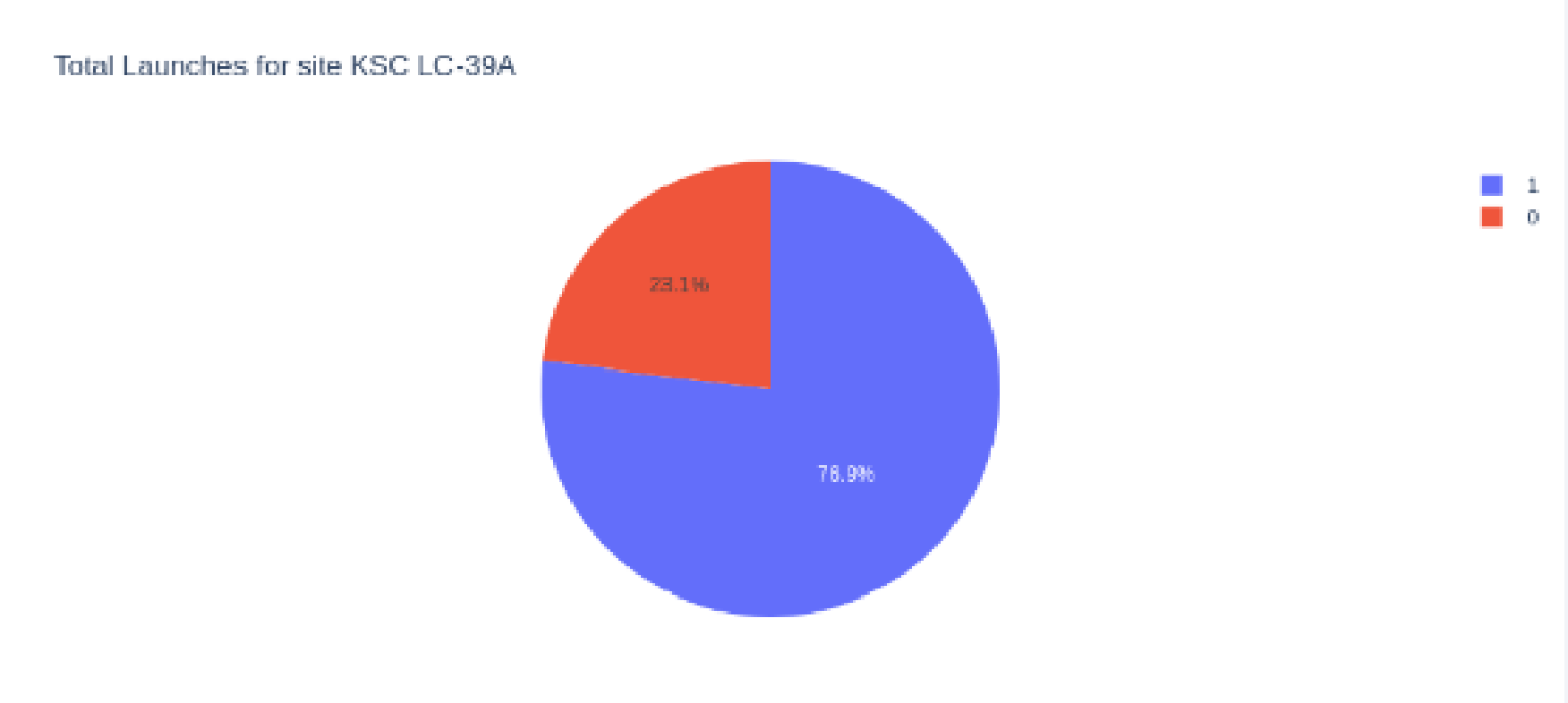

Section 4
Build a Dashboard with Plotly Dash

Launch success ratio for all sites



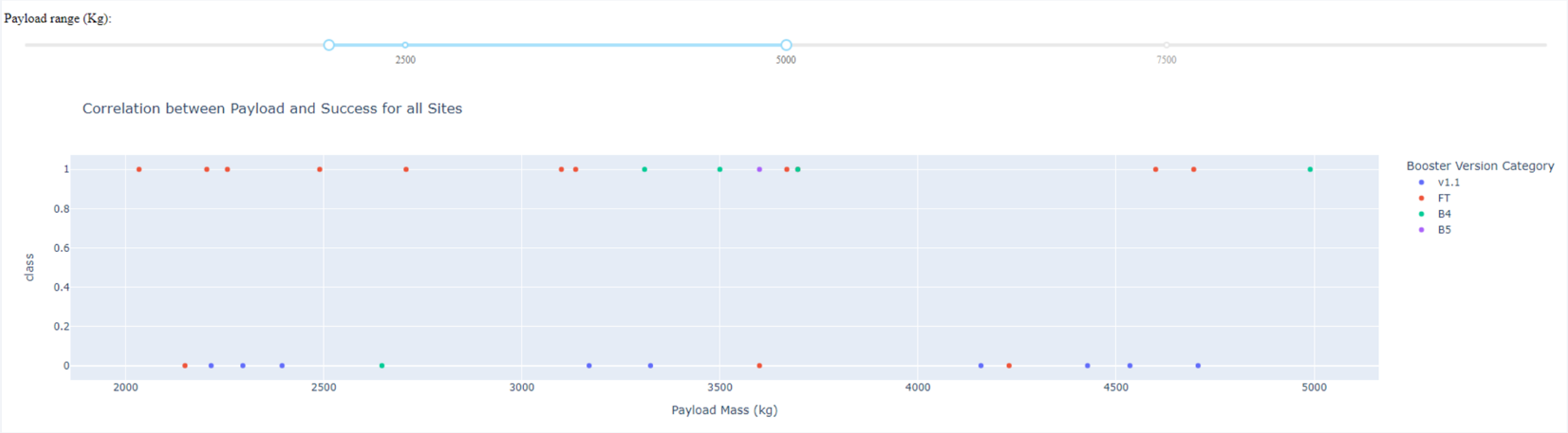
- KSC LC – 39A has the highest success count

Launch success ratio for KSC LC - 39A



- Successful launch is 76.9% while failure is 23.1%

Payload vs. Launch Outcome for all sites



- FT boosters have the largest success rate under the payload range selected above

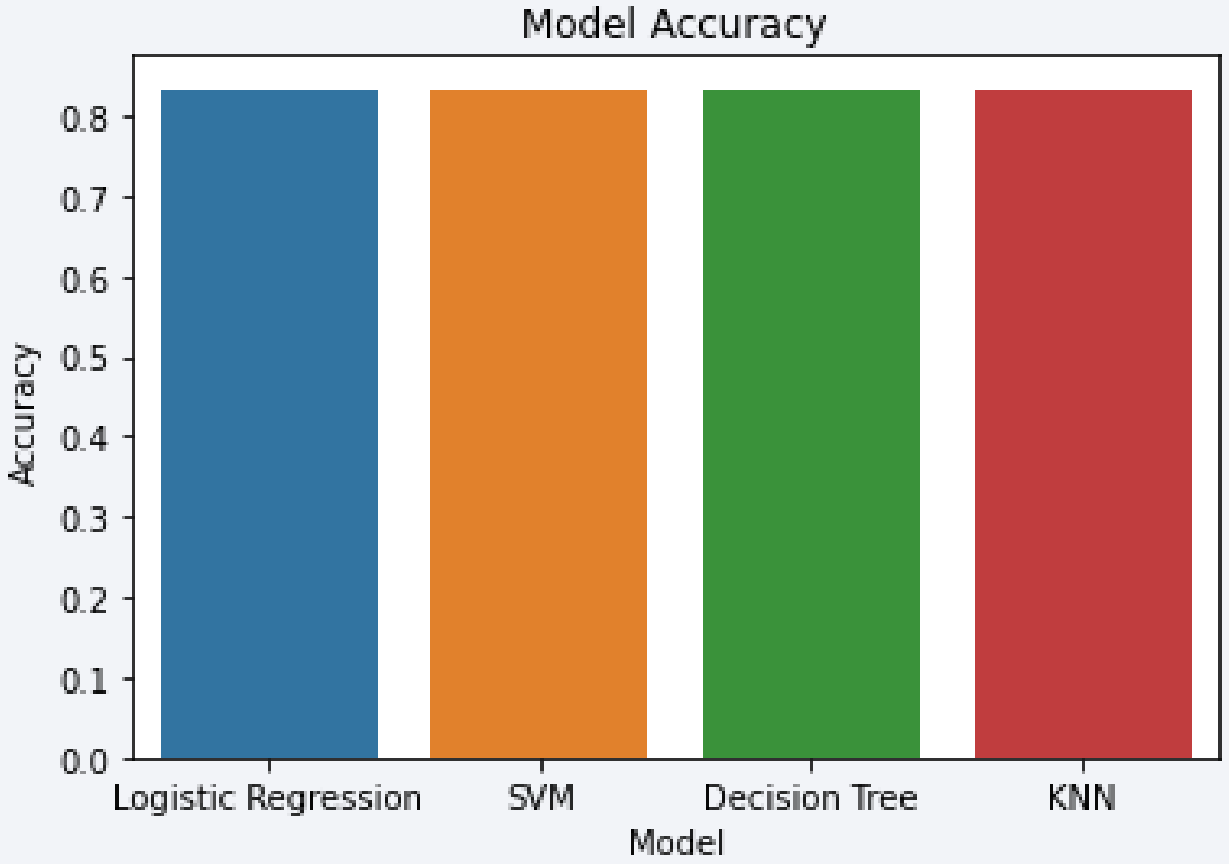
Payload vs. Launch Outcome for all sites

- FT boosters have the largest success rate under the payload range selected above and it appears that the higher the payload the lower the success rate.

Section 5

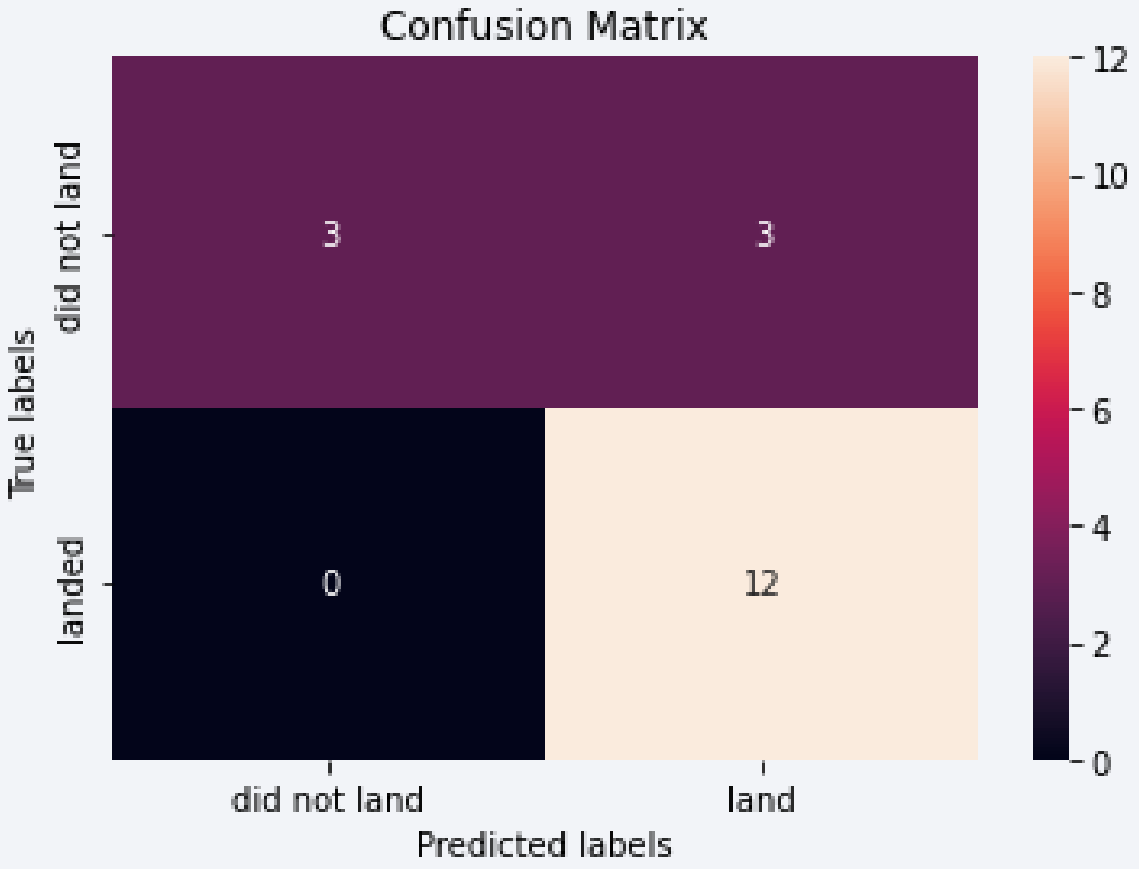
Predictive Analysis (Classification)

Classification Accuracy



- All the model have similar classification test accuracy of 83.33%.
- But decision tree is taken as the best model because it has the highest train accuracy

Confusion Matrix



- The confusion matrix of the decision tree model is shown which is similar to the other models with perfect prediction for true positive.
- A problem of the model is false positive.

Conclusions

- Several factors influence the missions' success, including the launch site, orbit, and, most importantly, the number of prior launches.
- ES-L1, GEO, HEO, SSO are the orbits with the highest success rate
- Analysis of various data sources shows that best launch site is KSC LC - 39A
- The payload mass may or may not be a factor in determining a mission's success depending on the orbits. In general, light weighted payloads outperform high weighted payloads.
- Even though the test accuracy for all the models utilized is the same, The Decision Tree Algorithm is taken as the best model for this dataset. Because the Decision Tree Algorithm has a higher train accuracy

Thank you!