IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Joseph
3/27/2023

Outline

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

Executive Summary

- Data related to SpaceX launches were collected, wrangled, and visualized to generate an understanding.
- Specific insights were drawn from the data and visualizations.
- The data was used to create models that were able to accurately predict launch outcome about 83% of the time

Introduction

- SpaceX is launching and trying to land rockets.
- We would like to figure out if we can determine if the rockets will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via web scrapping using python libraries
- Perform data wrangling
  - The data was examined and classified as a success or 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
  - The data was split into train an test segments, and were tested with a variety of models using different parameters to find the best model.

Data Collection

Steps:

1. Using python library "requests" data was collected from SpaceX API
2. The data was applied to a data frame for analysis
3. Related data was collected from Wikipedia
4. Using "BeautifulSoup" the data was extracted from html to a data frame

Data Collection – SpaceX API

- Data was collected using SpaceX API calls and prepared into a usable data frame.

- Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab1%20SpacexAPI.ipynb

SpaceX -> calls -> data frame

Data Collection - Scraping

- Using web scrapping, data was collected from Wikipedia as html and useful data was extracted to a data frame using beautifulsoup

- Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab1.1%20Web%20Scraping.ipynb

Wikipedia -> html -> soup -> data frame

Data Wrangling

- Data was examined for data types, value counts, and means
- Data was made more useful by classifying landing outcomes

Examine Data -> Classify Data

Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab2%20EDA.ipynb

EDA with Data Visualization

- Scatter plots with hue identifying the class were used to the effects of multiple variables on the classification(success or failure).
- A bar chart with the success rate for each orbit identified which orbits have high success rates.
- A line plot of year vs success rate shows the general trend for the success rate.

- Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab4%20jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries performed to get:

- Unique launch sites
- 5 data entries for launch site starting with 'CCA'
- Total mass carried by boosters launched by NASA (CRS)
- Average payload carried by booster version F9 v1.1
- Date of first successful landing outcome in ground pad
- Booster names for success in drone ship with payloads in a specified range
- Number of successful and failed missions
- Booster version that carried the maximum payload
- Failed landing outcome, booster, and launch site from 2015
- Rank of count for landing outcome in specified date range

Notebook: https://github.com/j123212321/Capstone/blob/main/Lab3%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

Objects added:

- Circle markers for launch sites, to know where launches have happened
- Markers for successful and failed launches, to see the classification of each location
- Lines from the coastline, city, railway, and highway to launch site, to see locations of points of interest in relation to the launch sites.

- Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab5%20lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Configurable pie chart showing number of successful launches for each launch site
- Configurable scatter plot of Class vs Payload with each booster version labeled

- Code For This Dashboard:

https://github.com/j123212321/Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Imported and prepared the data for classifaction
- Split the data into independent and dependent variables and split into training and test data
- Used training data to train a variety of models on the data
- Used testing data to evaluate the models

File -> Data Frame -> Split -> Models -> Evaluation

- Code For This Notebook:

https://github.com/j123212321/Capstone/blob/main/Lab6%20SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

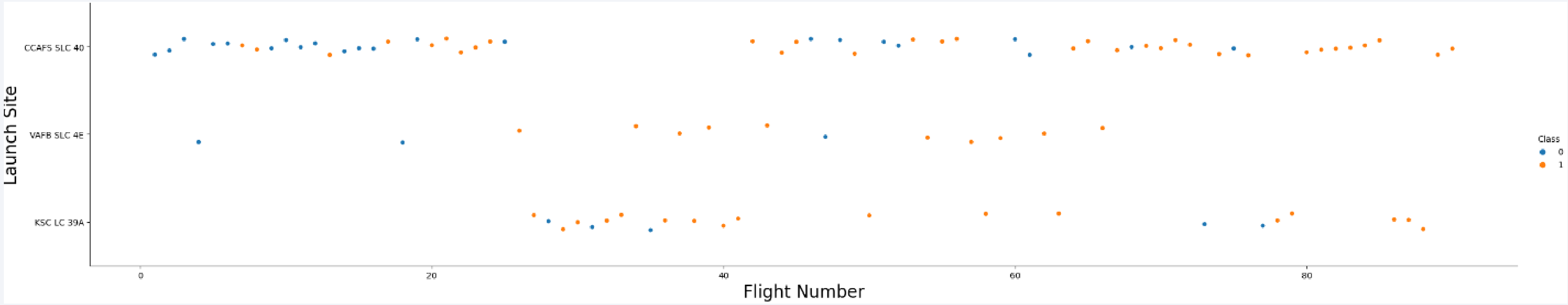
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Scatter plot of Flight Number vs. Launch Site

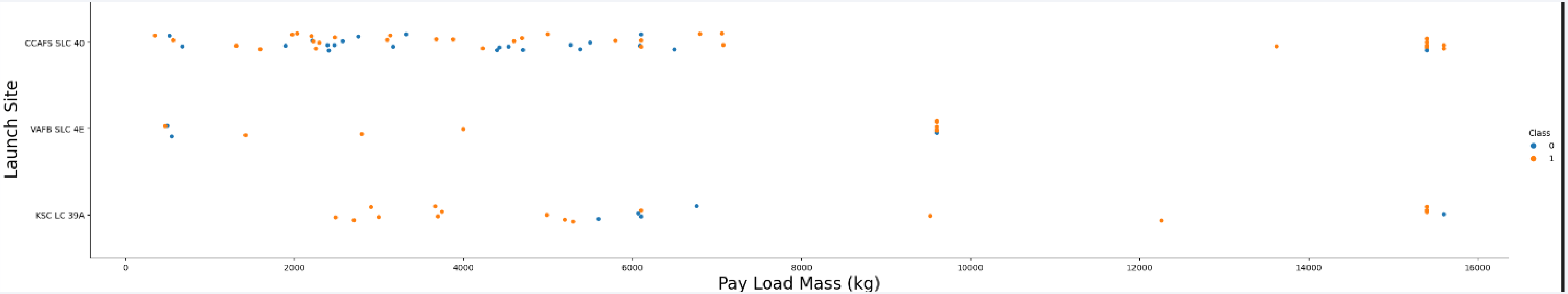


Explanation:

- Looking at only Flight Number and Launch Sites together does not show an obvious pattern for predicting whether the landing will be a success or not. Although, there are some trends such as higher flight numbers have a higher success rate, and certain launch sites also having a higher success rate.

Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site



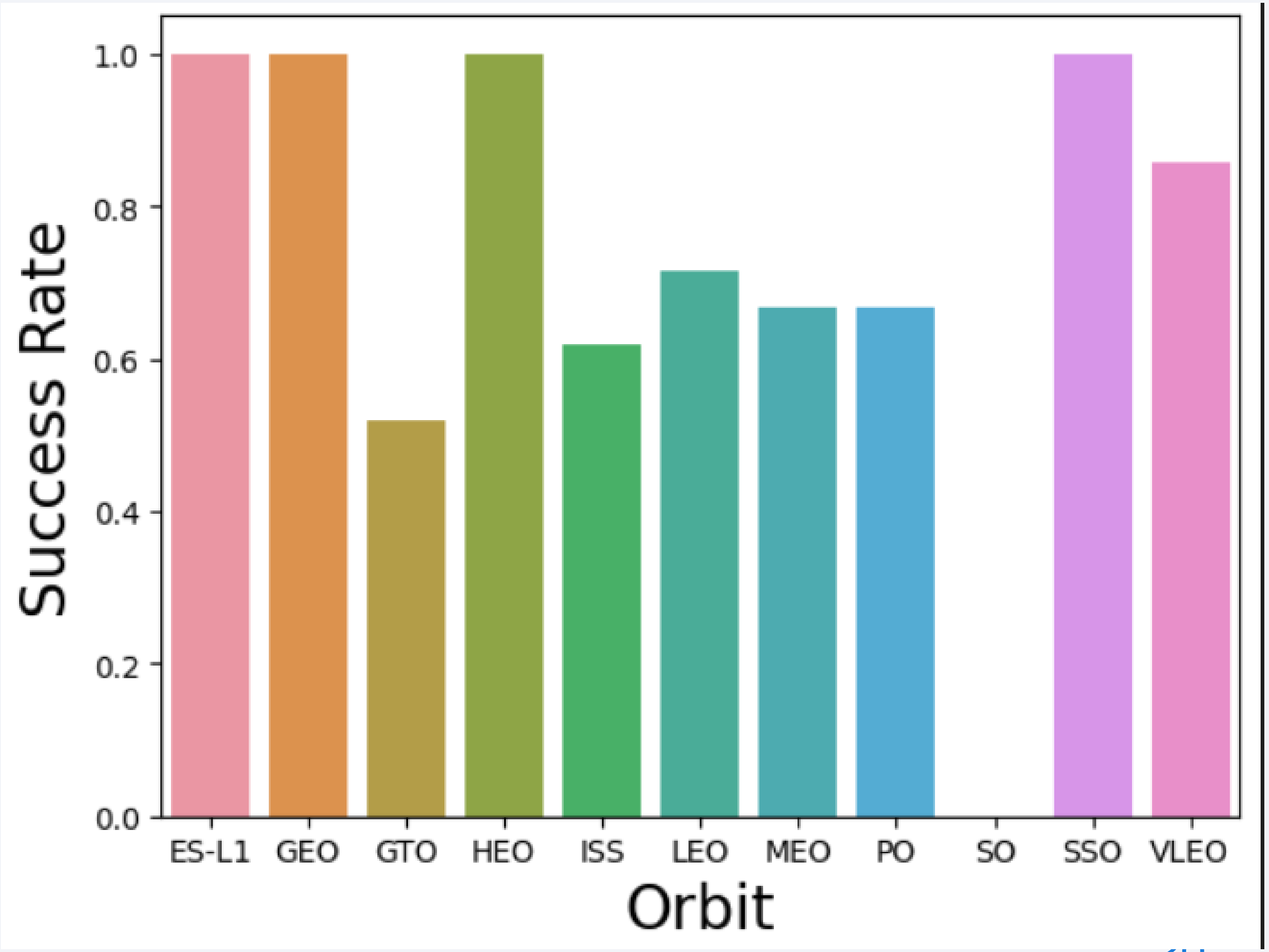
Explanation:

- Payload and Launch Site do not have a clear obvious pattern to predict the classification of landing outcome, but there are some noticeable areas for successful landing outcomes(i.e. KSC LC 39A less than 4000kg)

Success Rate vs. Orbit Type

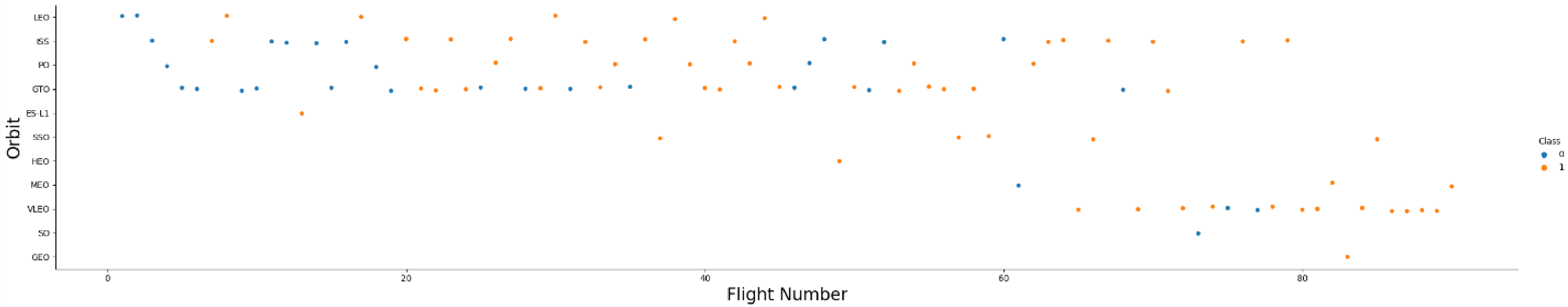
Explanation:

- The bar chart shows that some orbits have a high success rate



Flight Number vs. Orbit Type

Scatter plot of Flight number vs. Orbit type



Explanation:

- The plot shows that some orbits' success rates could be related to the flight number.

Payload vs. Orbit Type

Scatter plot of payload vs. orbit type

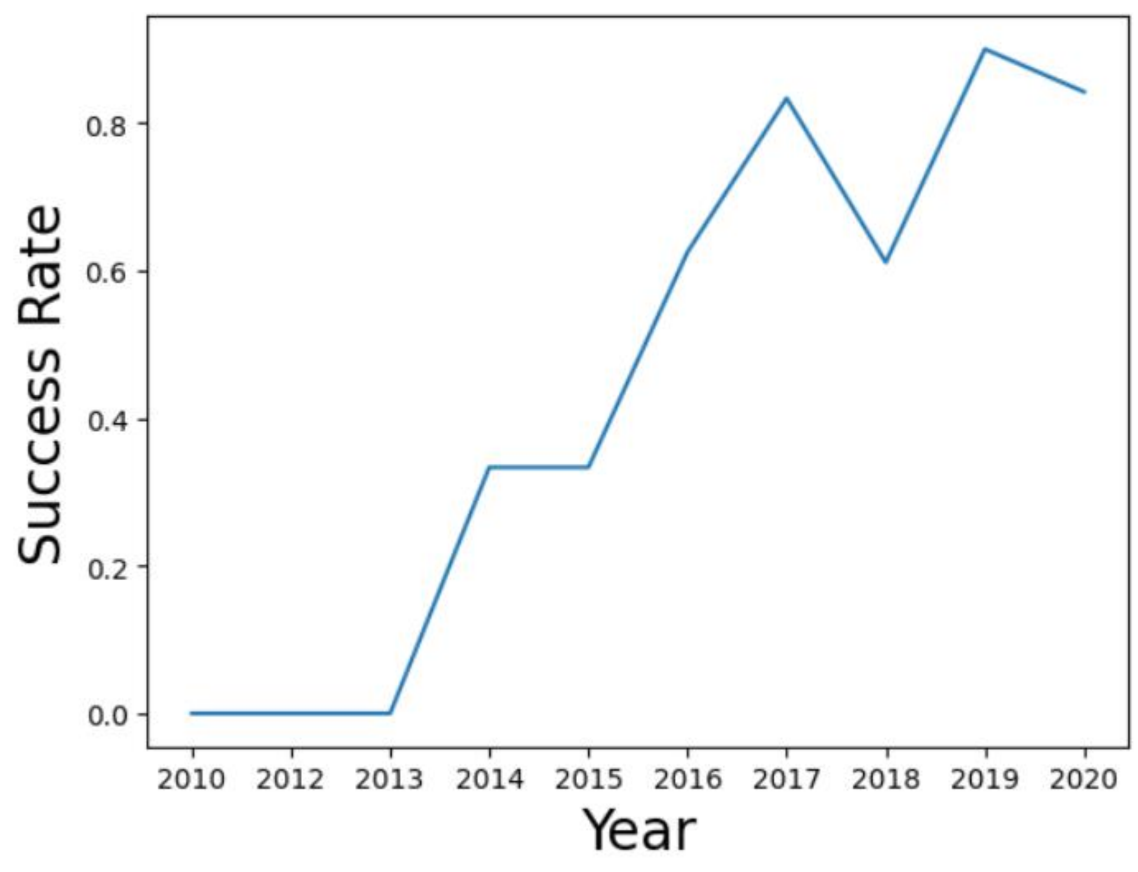
Explanation:

- The plot shows that payload mass is related to the success rate for some of the orbits

Launch Success Yearly Trend

Explanation:

- The success rate has generally been increasing since 2013.



All Launch Site Names

- Unique Launch Sites

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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

Total Payload Mass

- Total payload carried by boosters from NASA in kg

45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 in kg

2534

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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 successful and failure mission outcomes

1
101

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

payload_mass__kg_	booster_version
15600	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	F9 B5 B1051.4
15600	F9 B5 B1049.5
15600	F9 B5 B1060.2
15600	F9 B5 B1058.3
15600	F9 B5 B1051.6
15600	F9 B5 B1060.3
15600	F9 B5 B1049.7

2015 Launch Records

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

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

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

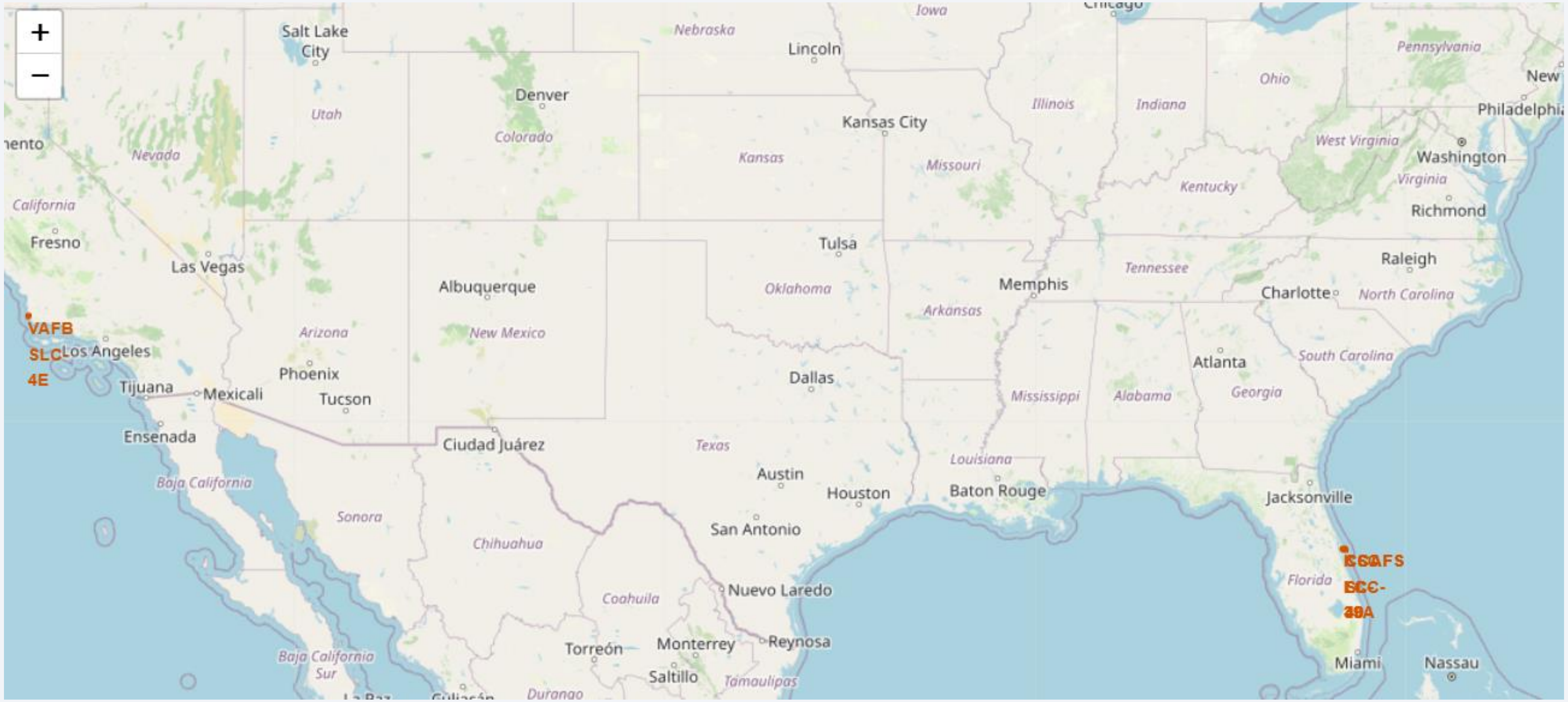
- Ranking of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	2
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 3

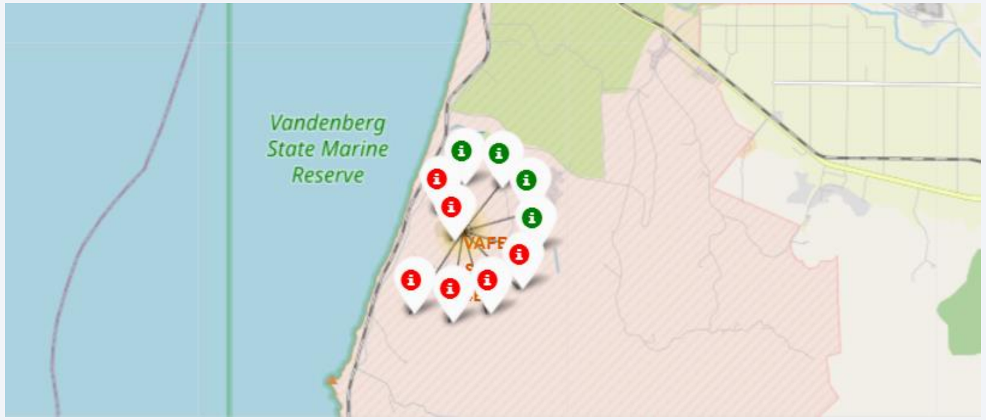
Launch Sites Proximities Analysis

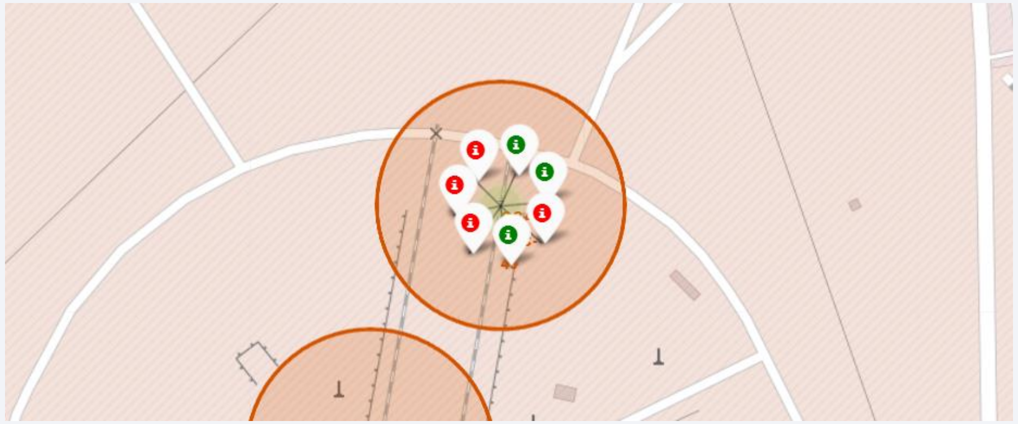
Map of Launch Sites

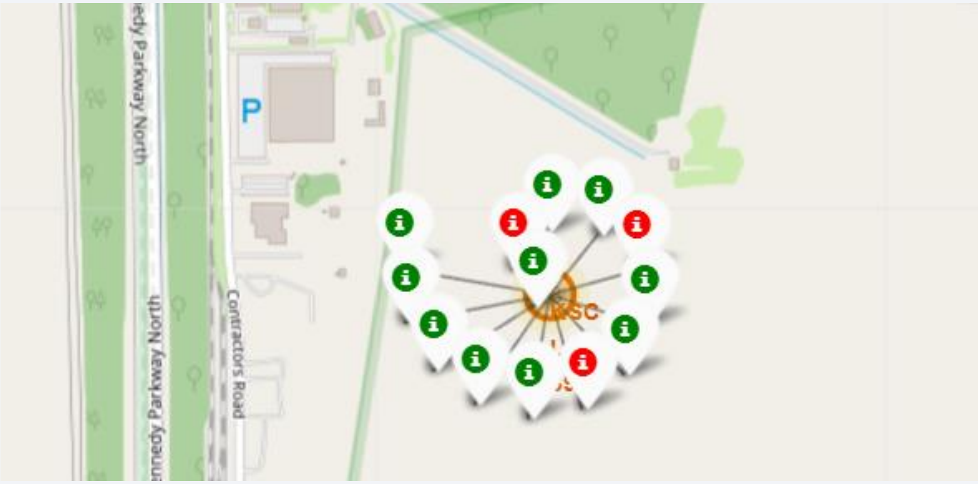


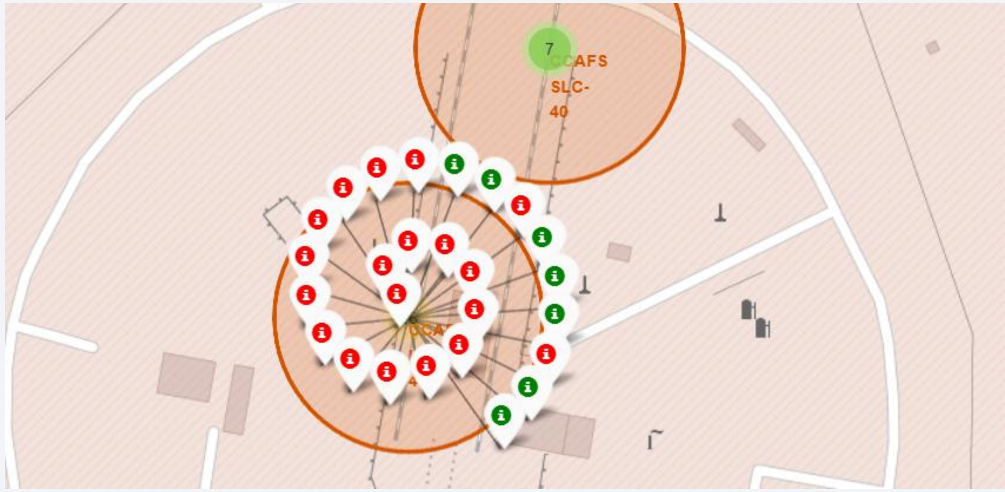
- There is a launch site in California, and multiple launch sites in Florida

Map of Successes/Failures



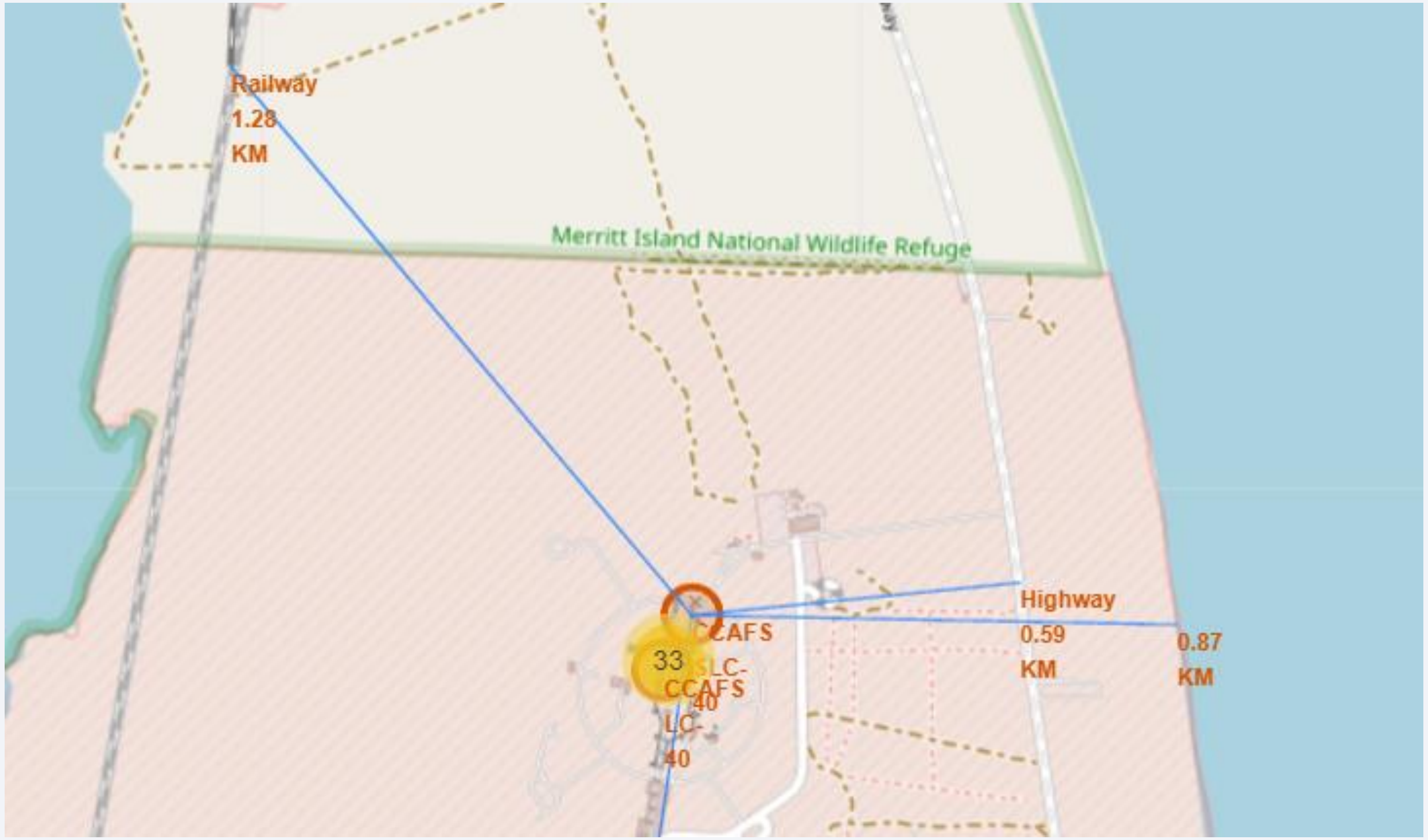






- Different launch sites have different numbers of successes and failures

Launch Site Distances From Other Locations

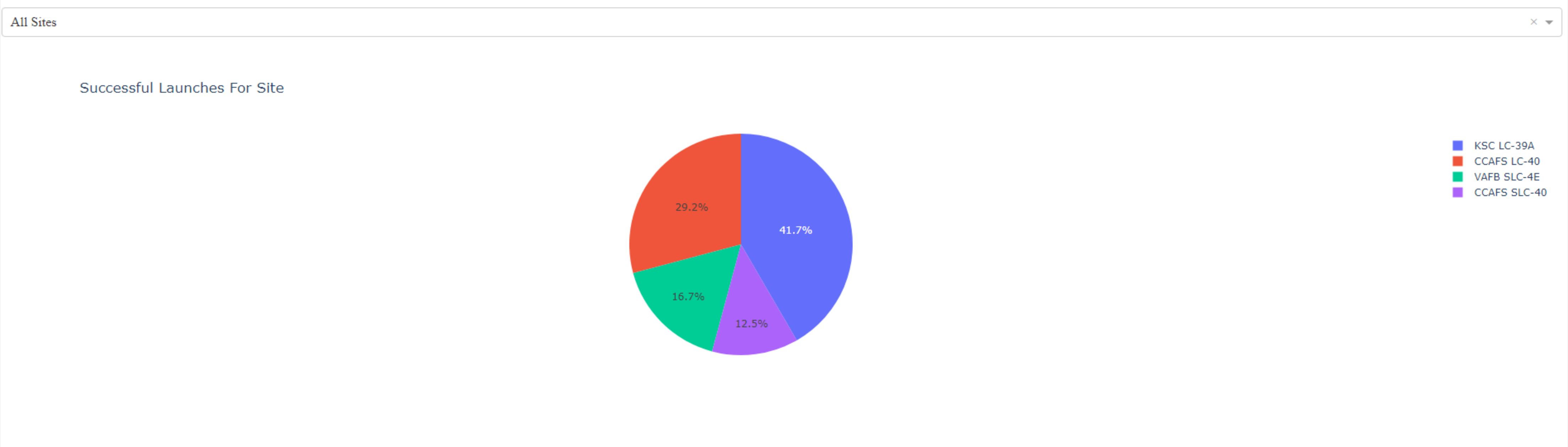


- The Launch Site is relatively close to a railway, a highway, and a coastline. The closest city is out of view of the screenshot but is 19.2 km away.

Section 4

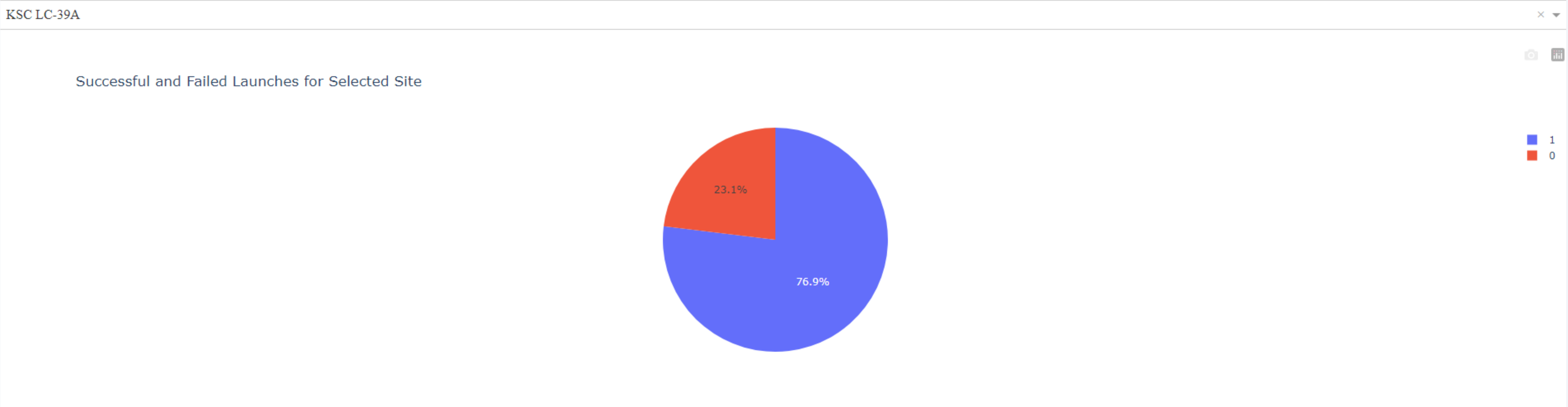
Build a Dashboard with Plotly Dash

Successful Launches For Each Site



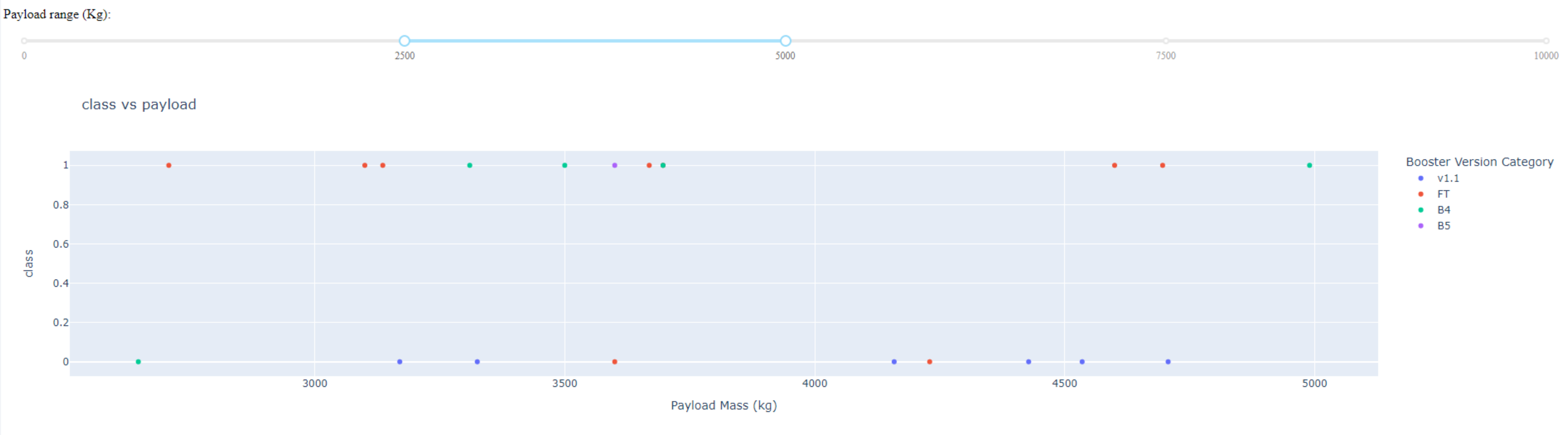
- Different launch sites have different number of successful launches, with KSC LC-39A having the most successful launches.

Successful and Failed Launches for KSC LC-39A



- KSC LC-39A had the highest success rate (76.9%) out of all the launch sites.

Class vs Payload for Selected Range



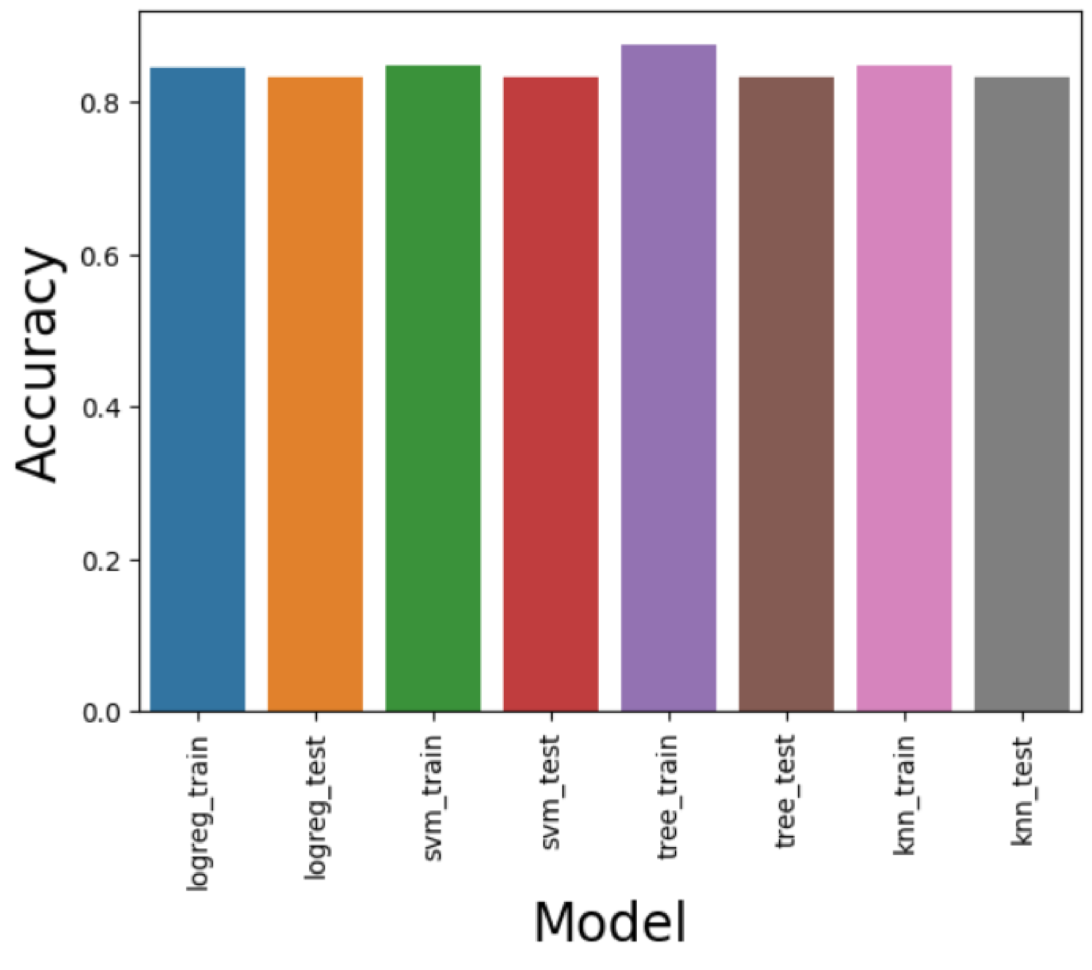
- In the selected payload range of 2500kg to 5000kg all of the v1.1 boosters failed, the almost all B4 boosters succeeded, the one B5 boosted succeeded, and the FT boosters were mostly successful.

Section 5

Predictive Analysis (Classification)

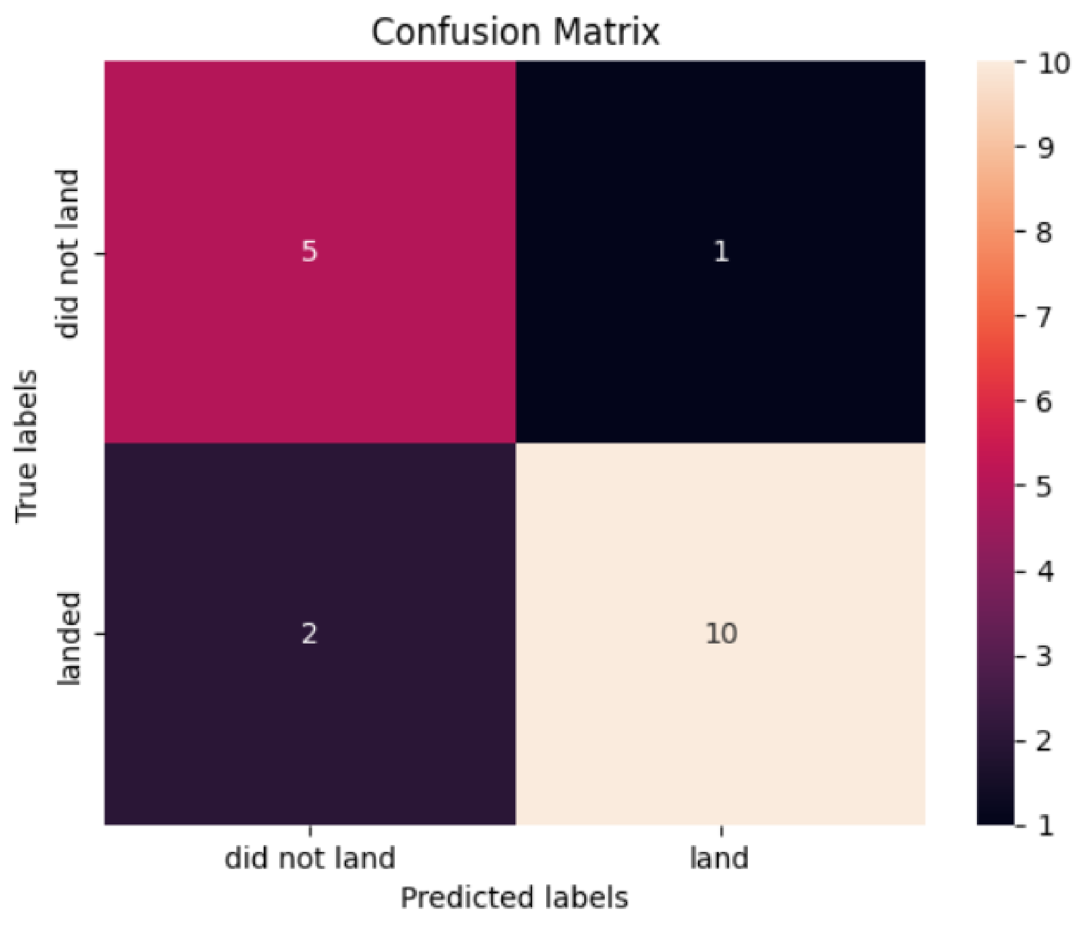
Classification Accuracy

- The decision tree classifier had the highest accuracy on the training data. All models had the same accuracy on the test data.



Confusion Matrix For Decision Tree

- The decision tree classifier corrected predicted 10 landings and 5 not landings. The classifier wrongly predicted that 2 would not land, and that 1 would land. The resulting accuracy is approximately 83%.



Conclusions

- Based on given data, launch data could be used to predict launch outcome with an accuracy close to 83%
- This finding demonstrates that collecting and analyzing data related SpaceX launches could be useful

Appendix

All code can be found at:

https://github.com/j123212321/Capstone

Thank you!