IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Joshua Murphy
4/4/2025

Outline

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

Executive Summary

- Space X can consistently launch falcon-9 rockets at prices much lower than competitors due to their ability to recover the first stage of many of the rockets. Using readily available data to compare which Space X launches were recovered and which weren't accurate predictions can be made regarding whether a Space X rocket stage 1 will be successfully recovered.

Introduction

- Understanding Which factors lead to successful or unsuccessful rocket recoveries is pivotal as it can better help our company predict how much a particular Space X launch will cost.

- By exploring Space X launch data we hope to learn which characteristics (payload mass, rocket launch site, date of launch, etc.) have the greatest effect on whether a Space X rocket will be successfully recovered.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - We used a GET request to acquire our launch data from the Space X API, then we filtered out any data that wasn't related to the falcon-9 launches
- Data wrangling
    - Once collected, we cleaned the data set removing any unnecessary null values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - We standardized our data, split it into training and testing sets, used GridSearchCV to find the best hyperparameters, and calculated the accuracies of multiple classification models to find the best one.

Data Collection

- To acquire our data we sent a GET request to the Space X API to request the required launch data. Afterwards it was filtered only to include launches of the falcon-9 rocket. We then replaced all null values for rocket payload mass with mean values so that the only remaining missing values belonged to the landing pad column of our data set.

Data Collection – SpaceX API

- Link to data collection Jupyter Notebook: https://github.com/MurphyJ474/jupyter-labs-spacex-data-collection-api

Set the URL of the required data to the variable “static_json_url”.

↓

Use the request.get function on created variable and make it a new variable “response”.

↓

Decode the response variable by using .json() and set the result to a new variable “res_json”.

↓

Turn the json data into a pandas dataframe using the pd.json_normalize() function.

Data Collection - Scraping

- Link to web scraping Jupyter Notebook: https://github.com/MurphyJ474/jupyter-labs-webscraping

Set the URL of the required data to the variable “static_url” and use request.get function.

↓

Create beautiful soup object from the content of the previously created variable.

↓

Search through the object for the table of launch records and iterate through <th> elements to create a list of column headers.

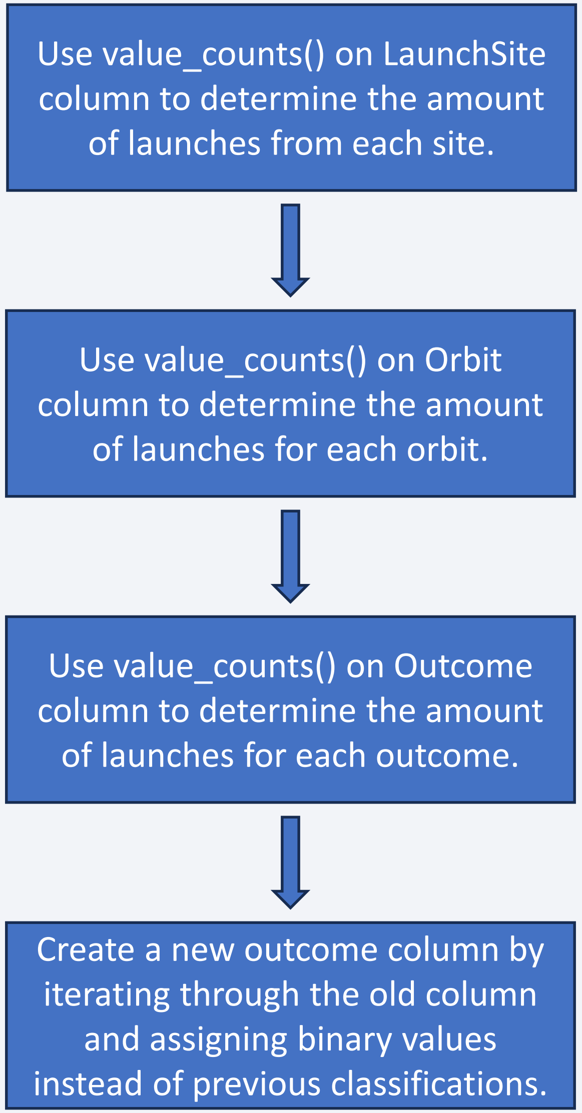
↓

Create a dictionary with column names as keys and iterate through the soup object to fill the dictionary for use as a dataframe.

Data Wrangling

- To process the data we explored which sites, orbits, and mission outcomes were most and least likely by finding the number of occurrences for each. Afterwards, we added a new column to the dataframe that determined whether the outcome for a given launch for bad (0) or good (1).

- Link to data wrangling related notebooks: https://github.com/MurphyJ474/labs-jupyter-spacex-Data-wrangling



EDA with Data Visualization

- To visualize the data we used a variety of charts to compare factors that might be correlated to the outcome of a launch. The charts we explored include the following: flight number vs. launch site, mass vs. launch site, success rate vs. orbit type, flight number vs. orbit, mass vs. orbit, and success rate vs. year
- Link to EDA with Data Visualization Jupyter Notebook: https://github.com/MurphyJ474/edadataviz

EDA with SQL

- Queries performed
  - 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 succesful 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 all the booster_versions that have carried the maximum payload mass. Use a subquery.
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months 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.
- Link to EDA with SQL Jupyter Notebook: https://github.com/MurphyJ474/jupyter-labs-eda-sql-coursera_sqllite

Build an Interactive Map with Folium

- On our map we provided circles around the launch sites, markers denoting successful and failed launches from each site, and a line to the nearest shore for the CCAFS SLC launch site.

- These objects were added to help us better visualize what type of area SpaceX choses for its sites and how see how close features like railroads and coastlines are from them.

- Link to Folium Jupyter Notebook: https://github.com/MurphyJ474/lab_jupyter_launch_site_location

Build a Dashboard with Plotly Dash

- In our interactive dashboard we've added a pie chart of the different landing sites and their success rates as well as a scatter plot of the payload mass vs. launch outcome for a given range.

- These plots were included to help us explore how much correlation there was between the mass of a launch or the site it was launched from and the chance of it being recovered successfully.

Predictive Analysis (Classification)

- First we standardized our dataframe, then we split it into testing and training sets. Grid search was used to find the best hyperparameters before the accuracy of the created model was tested. This process was done for a variety of model types.
- Link to Predictive Analysis Jupyter Notebook: https://github.com/MurphyJ474/SpaceX_Machine-Learning-Prediction_Part_5

Model was standardized using StandardScaler() and fit_transform()

↓

The data was split using the train_test_split() function

↓

GridSearchCV() was used alongside unique parameter dictionaries to find the best hyperparameters for the models

↓

The score() method was used on the testing set to assess model accuracy

Results

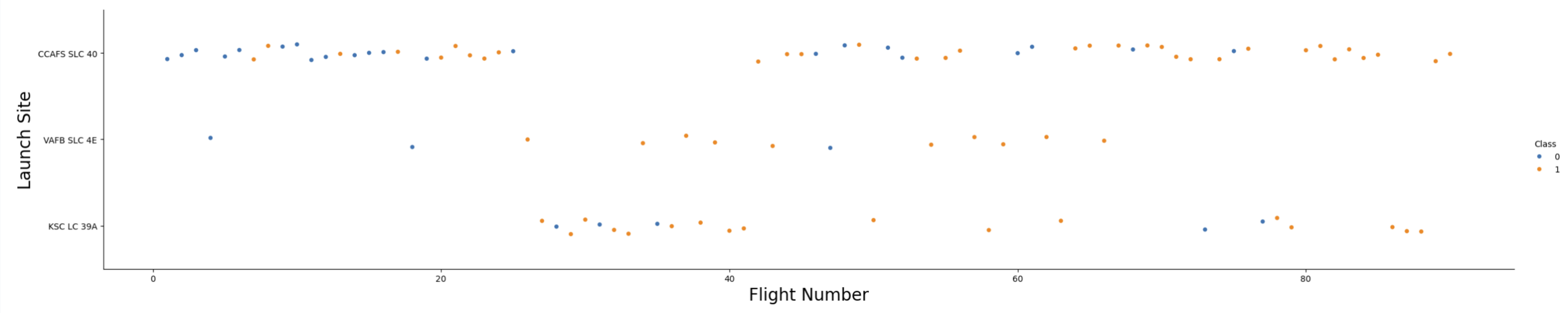
- The exploratory data analysis results show that as flight number increases launch sites change, success rate notably increases, and common orbits drastically change.

- Predictive analysis results show that the decision tree classification method seems to be slightly more accurate than the other classification methods.

Section 2

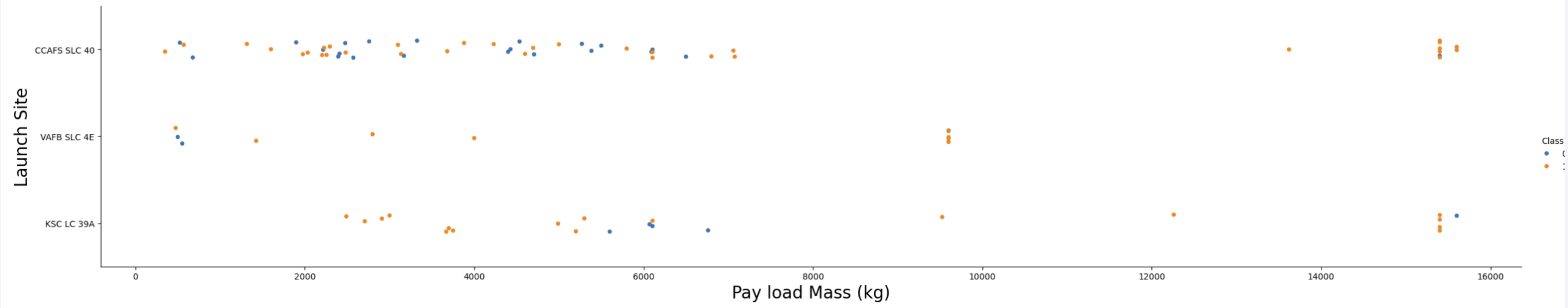
Insights drawn from EDA

Flight Number vs. Launch Site



- Initial flights with low flight number are mostly from CCAFS SLC 40 launch site

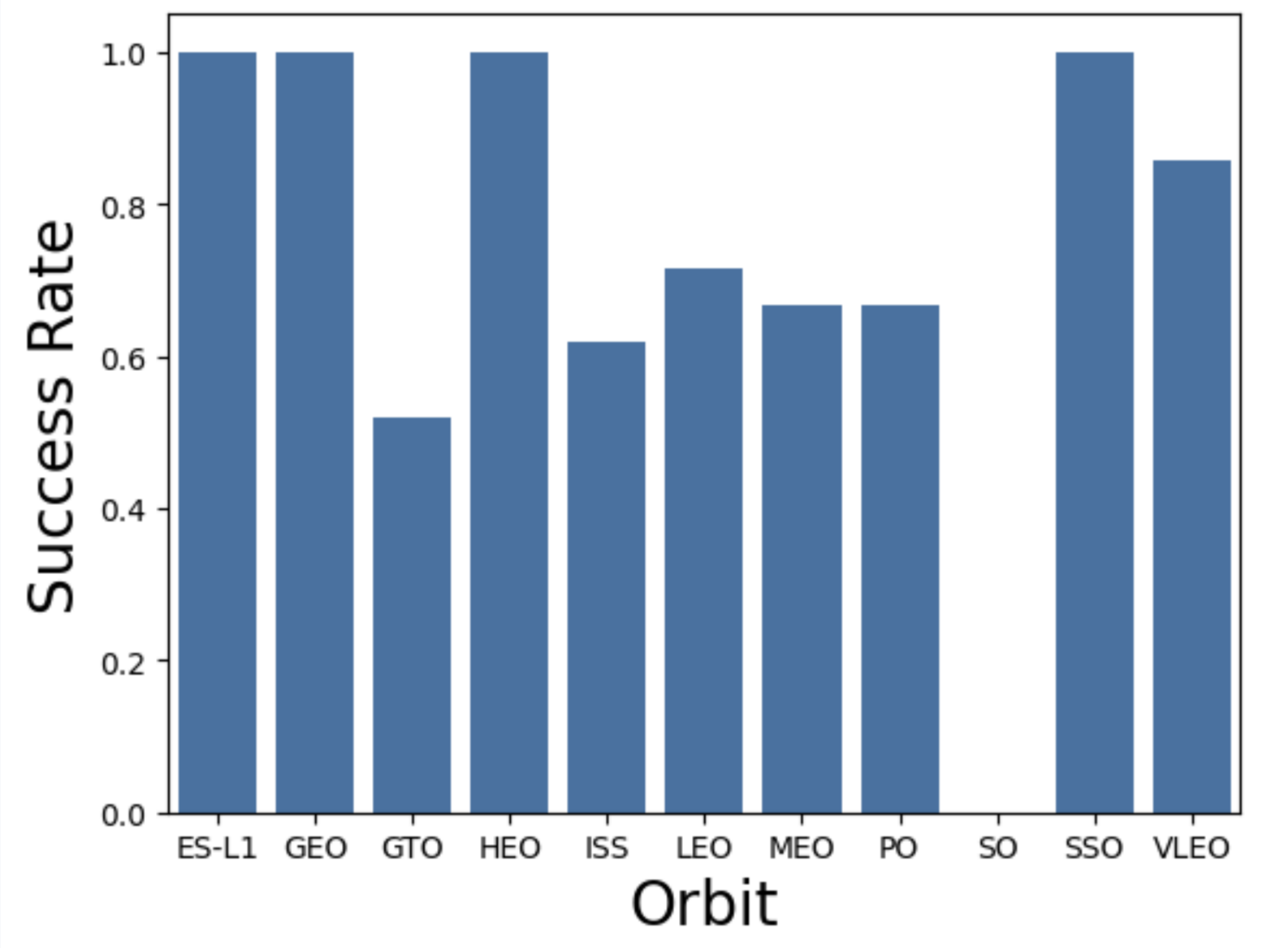
Payload vs. Launch Site



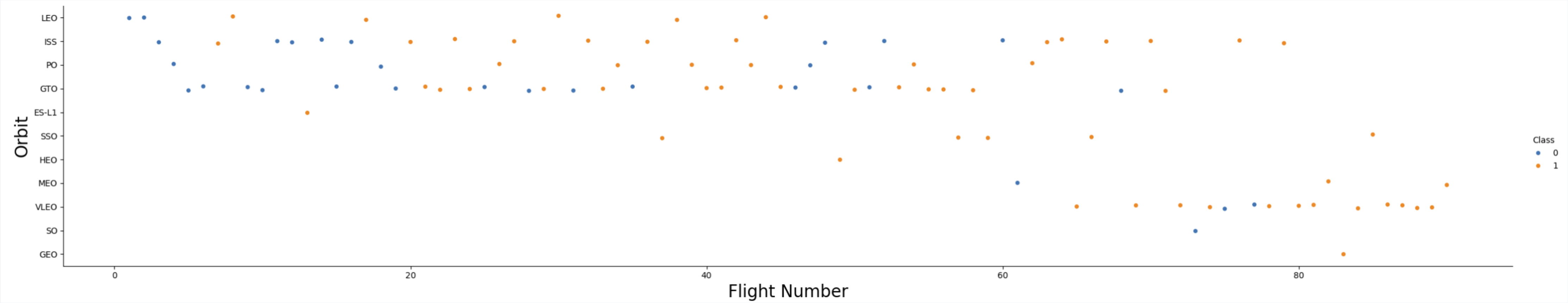
- The VAFB SLC 4E launch site seems only to be used for lighter payloads

Success Rate vs. Orbit Type

- The success rate seems to depend heavily upon which orbit the mission is aiming for.

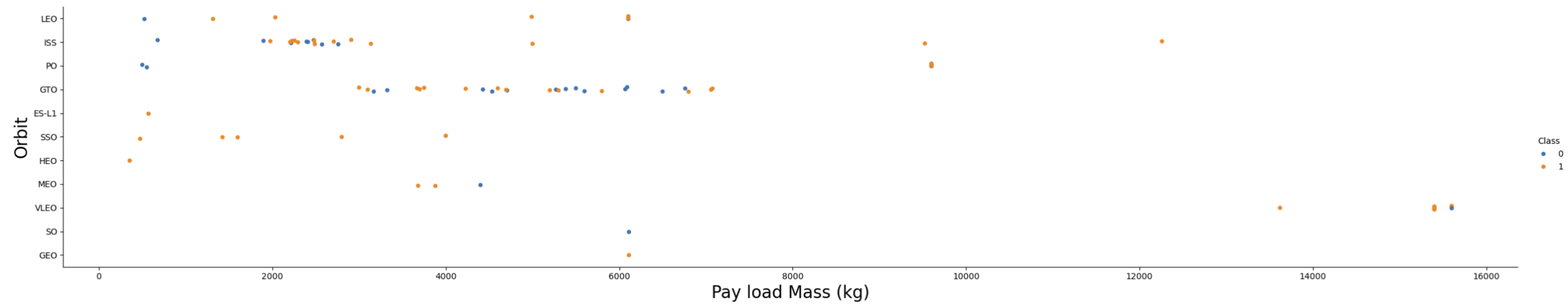


Flight Number vs. Orbit Type



- As the flight number grows larger the common orbits for launches begins to change.

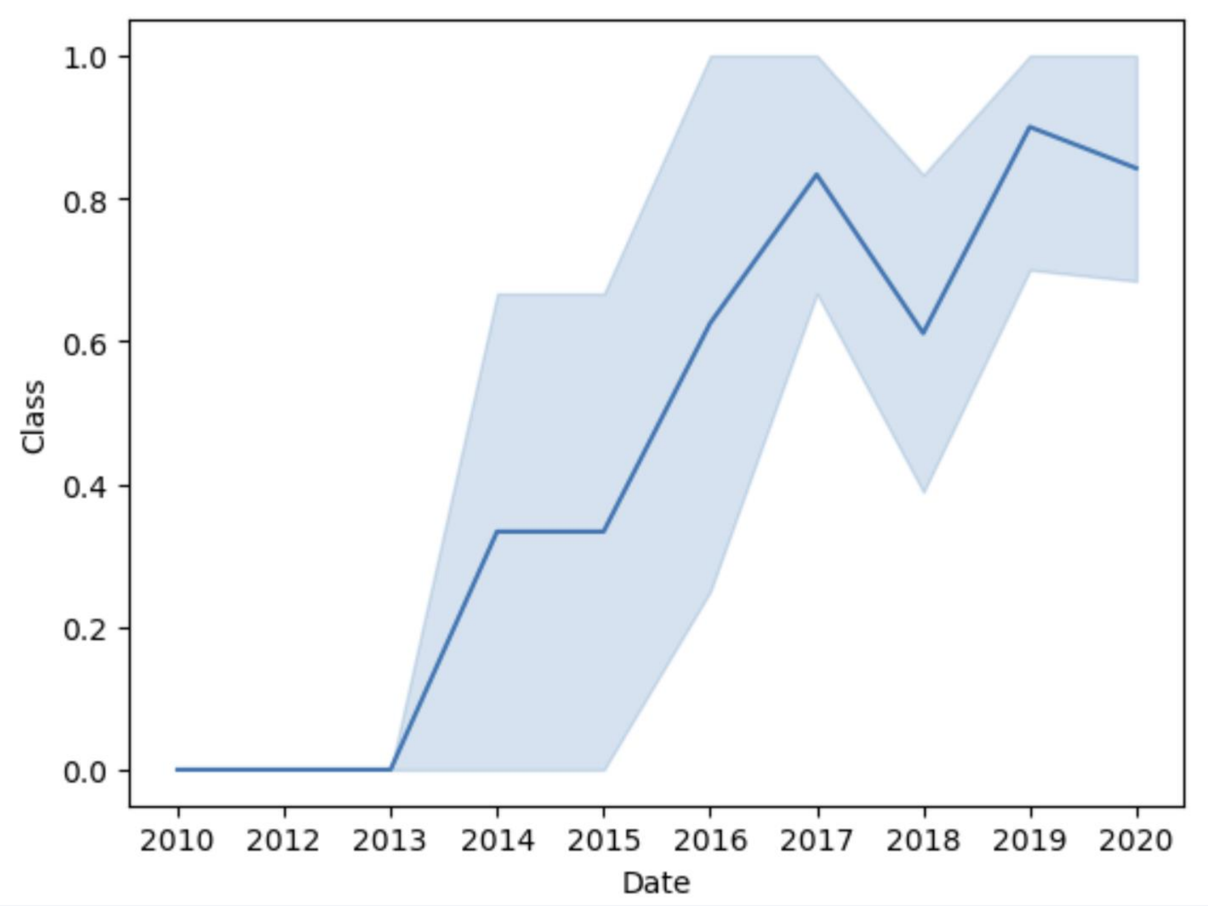
Payload vs. Orbit Type



- The heaviest payloads seem to use orbits not frequently used by lighter payloads, however there is a large variation of orbits for most payloads at a lower weight.

Launch Success Yearly Trend

- As the year increases the success rate of launches continues to increase.



All Launch Site Names

- Here are the names of all the sites with launches listed in the dataframe.

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

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- The first 5 records beginning with ‘CCA’ all belong to the launch site CCAFS LC-40

Total Payload Mass

- We calculated the total payload mass of boosters launched by NASA (CRS)

TOTAL_MASS_CRS
45596

Average Payload Mass by F9 v1.1

- The calculated number for average payload mass carried by the F9 v1.1 booster version.

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

- The first successful ground landing was found to be in 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters in the given list were the only versions to successfully land payloads that were between the weights 4000 and 6000 kg.

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

Total Number of Successful and Failure Mission Outcomes

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

- Despite the payload often being unrecovered the above graph shows that the mission failure rate is extremely low.

Boosters Carried Maximum Payload

- A list of the booster versions that carried the heaviest payload recorded in the database.

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- This graph shows the month names for launches that ended in dropship failure landings during the year 2015.

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

- This list shows the landing outcomes in descending order of most common to least common between the dates 2010-06-04 and 2017-03-20.

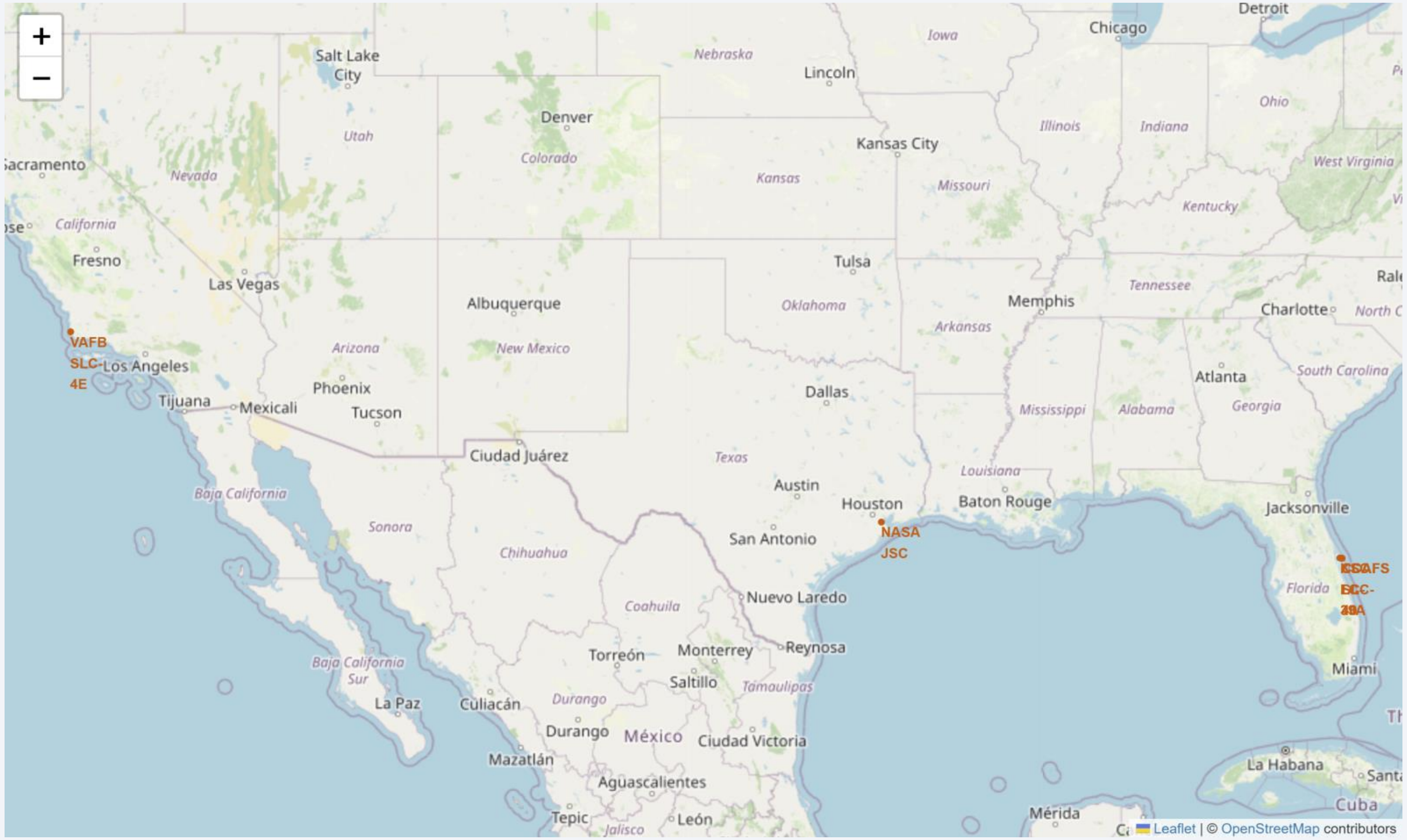
COUNT("Landing_Outcome")	Landing_Outcome
10	No attempt
5	Success (drone ship)
5	Failure (drone ship)
3	Success (ground pad)
3	Controlled (ocean)
2	Uncontrolled (ocean)
2	Failure (parachute)
1	Precluded (drone ship)

Section 3

Launch Sites Proximities Analysis

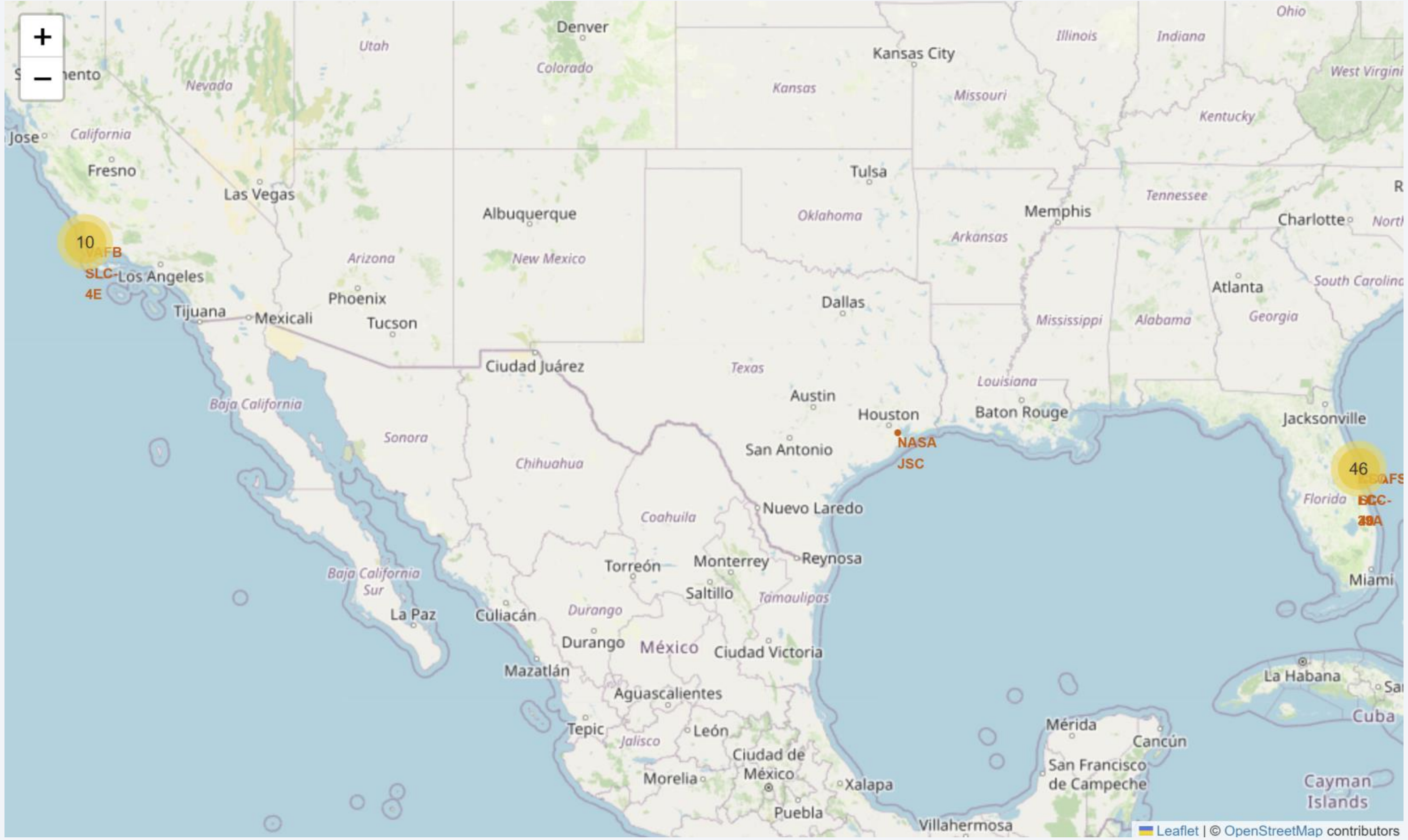
Rocket Launch Sites

- The rocket launch sites in the database are all located on the southern side of the United States near the coast.



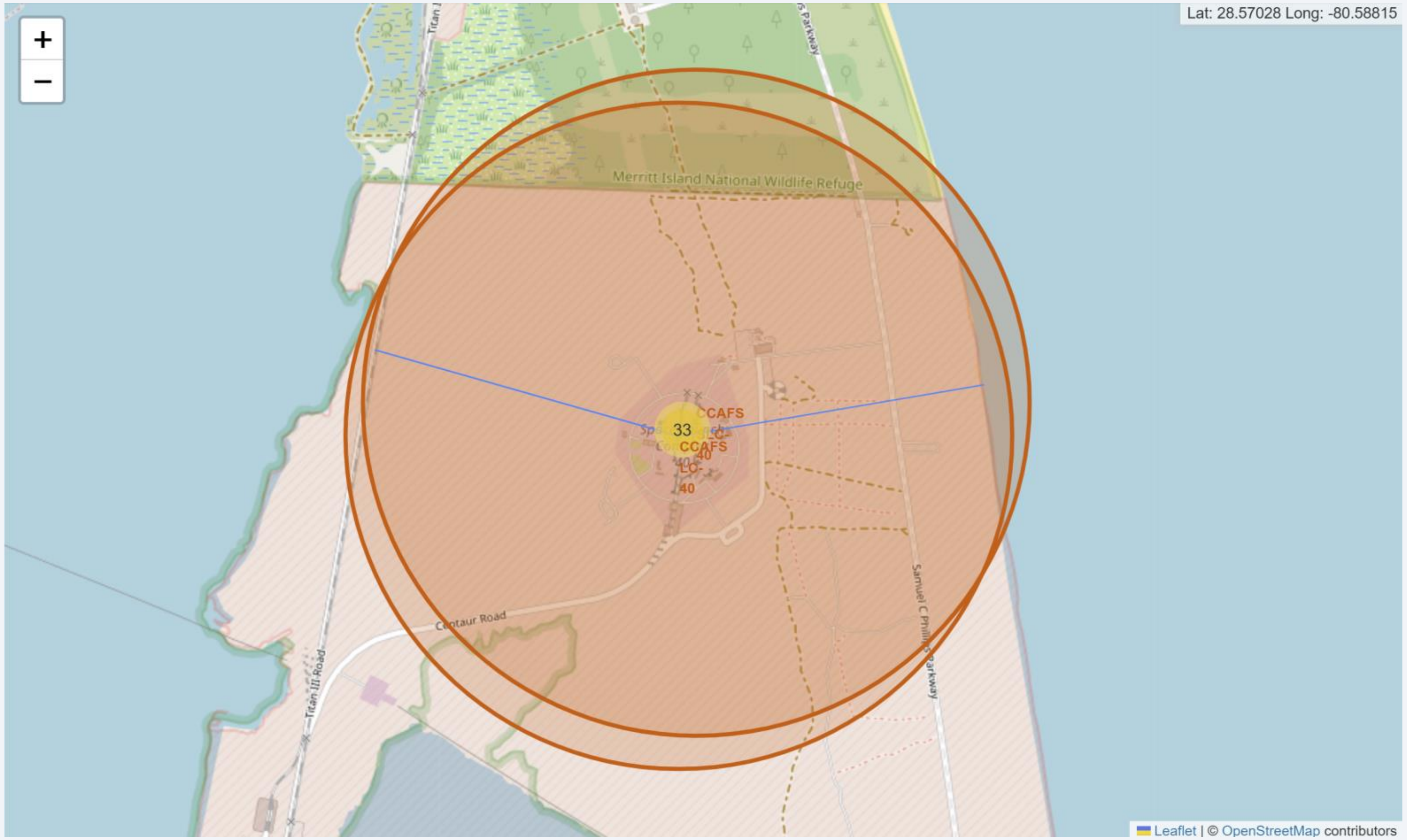
Number of Launches from each site

- The map was modified to have markers denoting each launch.



Zoomed in Map of the CCAFS Sites

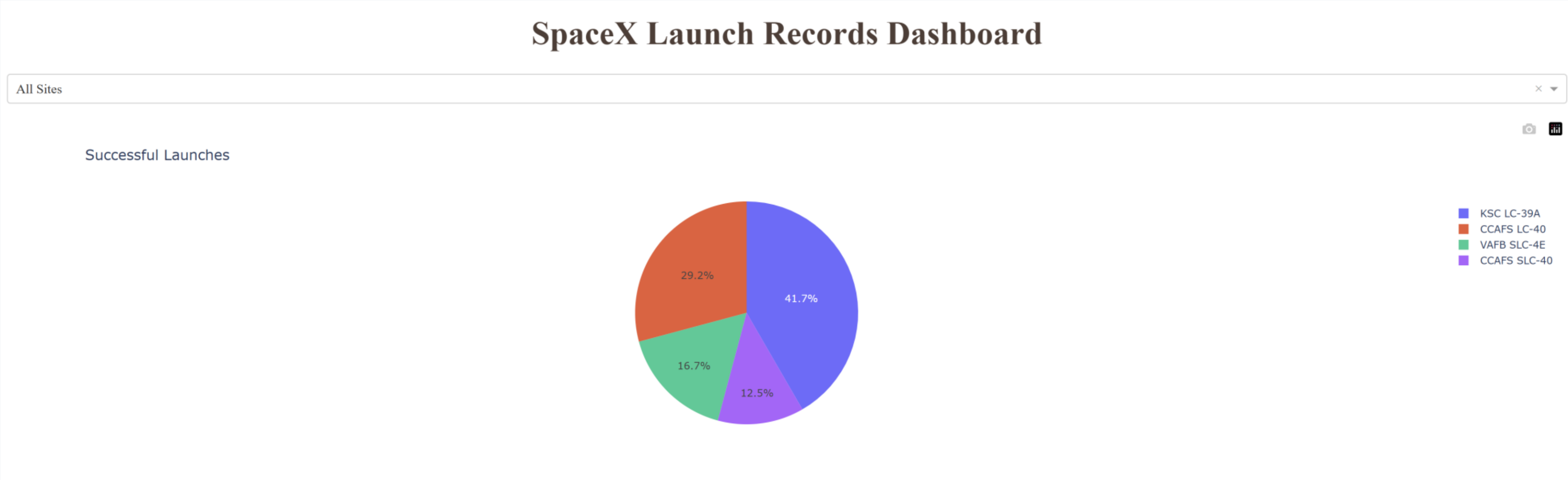
- This map shows the distance from the launch site to the coastline as well as its distance to the nearest railroad with blue lines.
- Both launch sites in the picture are in very close proximity to both.



Section 4

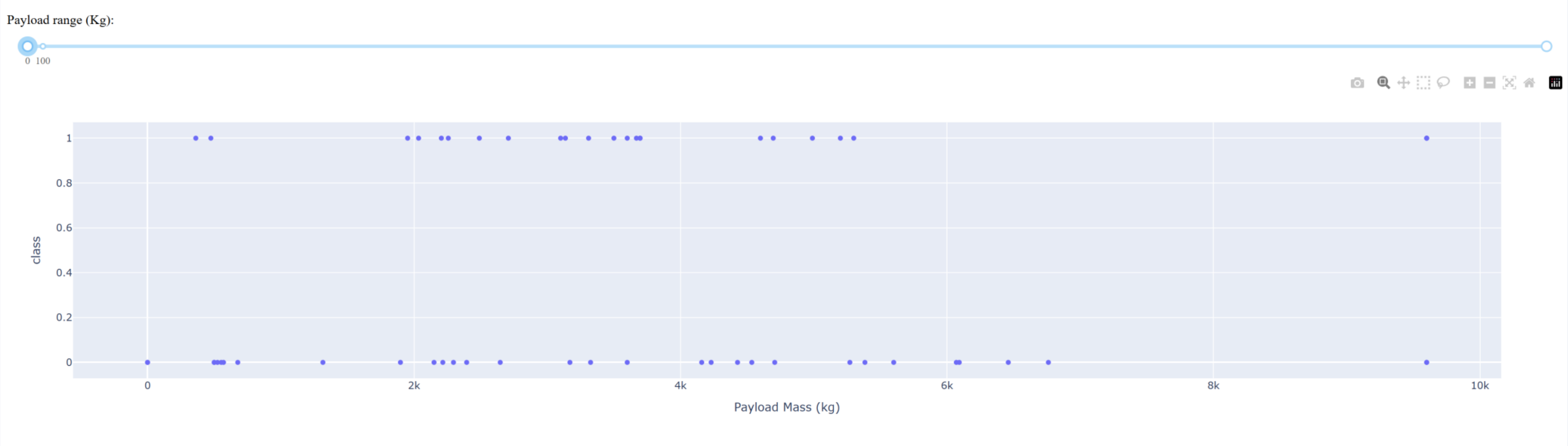
Build a Dashboard with Plotly Dash

Successful Launches Pie Chart



- This pie chart shows that the KSC LC-39A launch site has the greatest number of successful launches.

Success Rate by Payload Range Scatterplot



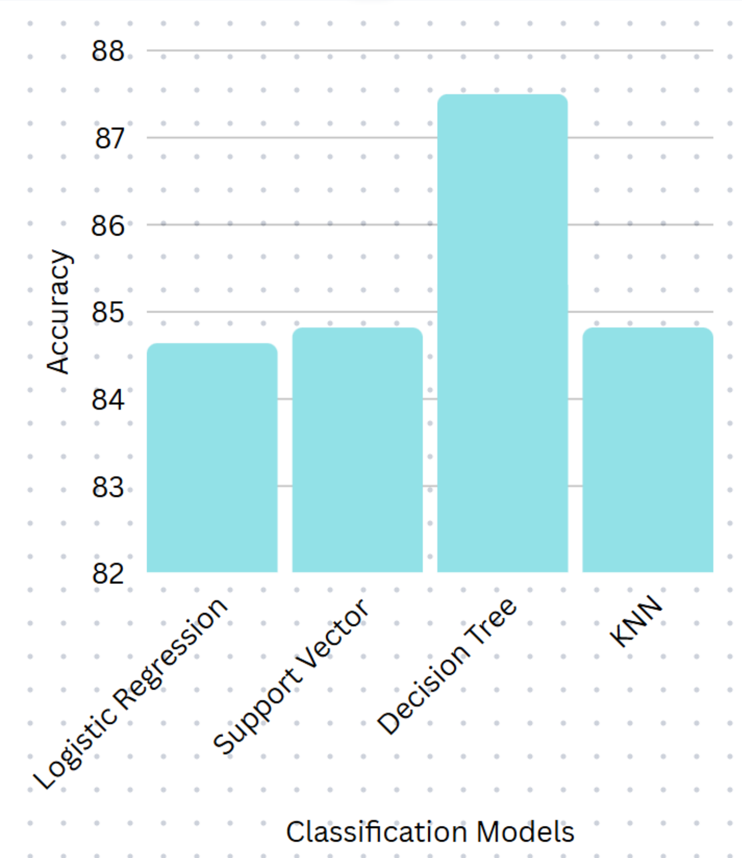
- Payloads in between 2000 and 6000 kg seem to have the highest success rate.

Section 5

Predictive Analysis (Classification)

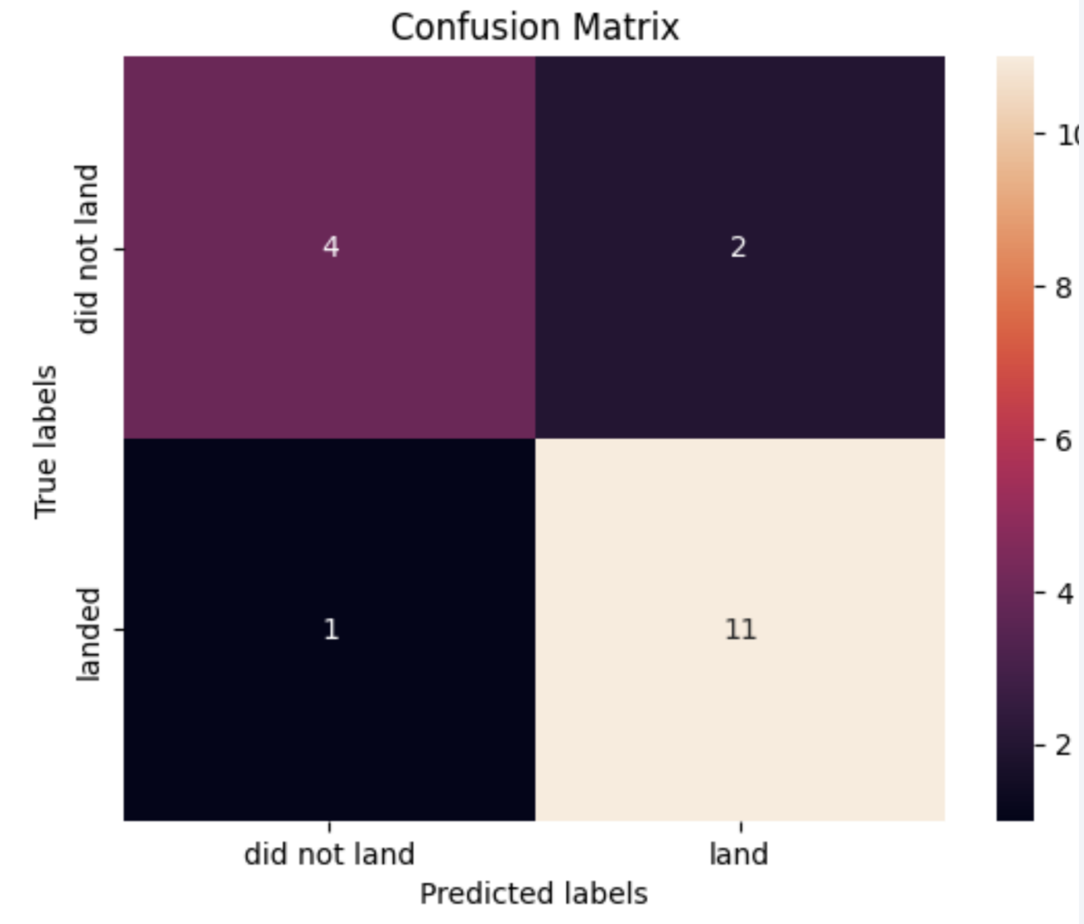
Classification Accuracy

- The Decision tree classification model was found to have the highest accuracy when attempting to predict landing outcomes.



Confusion Matrix

- The confusion matrix of the decision tree model shows that the model predicted 2 false positives and 1 false negative.



Conclusions

- There is a lot of correlation between flight number and payload mass with the success rate of SpaceX recovering its rockets
- As time proceeds rockets seem to have a higher chance of being recovered and rockets within the weight range of 2000 and 6000 kg seem to also have a higher chance of being recovered.
- Orbit type also plays a big role in whether or not a rocket is likely to be successfully recovered, however orbit types also changed drastically with time so the two categories are heavily tied to one another.
- The decision tree prediction model was found to be the best fit for predicting whether or not a SpaceX rocket was likely to be retrieved successfully.

Thank you!