Winning Space Race with Data Science

Femi Olawale
25th April 2023

Outline

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

Executive Summary

- The goal of this project is to predict if the first stage of the SpaceX Falcon 9 rocket launch will land successfully. A successful landing will help determine the cost of a launch.
- Key factors such as payload mass, orbit type, booster type, and launch site influence the success rate of landing.
- The geographical location and proximities of launch sites to coastline, cities and the equator were also examined as an indicator of landing success rate.
- Overall, the result of the predictive analysis show that the first stage of Falcon 9 rocket will land successfully with 94% certainty.
- Further optimization or fine-tunning of key parameters such as booster type and payload mass may further increase the chances of successful landing.

Introduction

- SpaceX has gained worldwide attention for a series of historic milestones.

- It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

- By establishing that the first stage of the Falcon 9 launch will land successfully, this will provide useful information in determining the cost of a launch. It can also serve as a reference for another company that wants to bid against SpaceX for rocket launch.

Section 1

Methodology

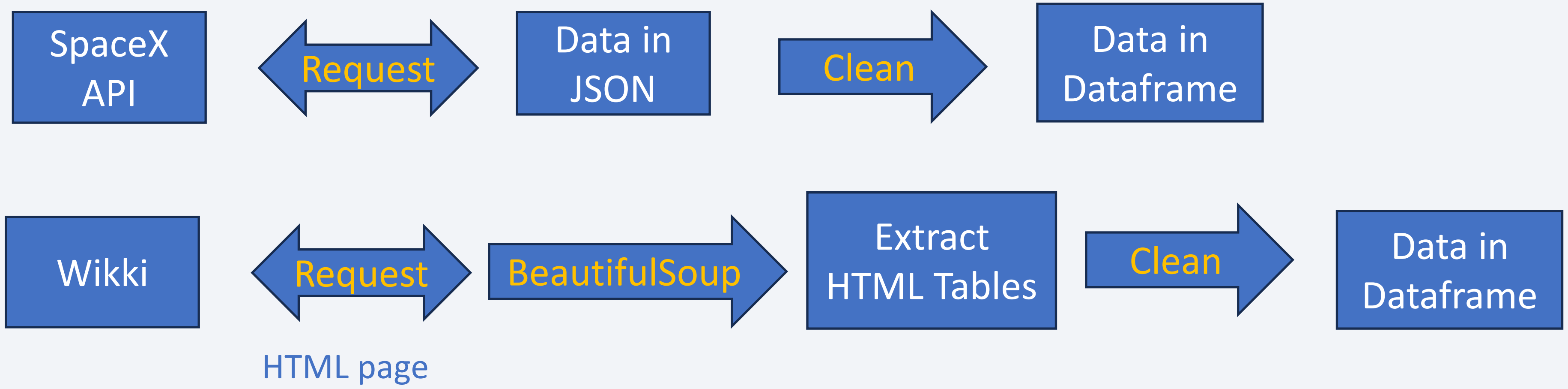
Methodology

Executive Summary

- Data collection methodology:
  - The rocket launch dataset was collected by making a request to the SpaceX API. Falcon 9 launch records were also web-scrapped from Wikipedia using BeautifulSoup.
- Perform data wrangling
  - Collected data were converted to Pandas data frame for cleaning and exploratory analysis and to determine what would be the label for training supervised machine learning models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Cleaned data was standardized and split into training and test data.

Data Collection

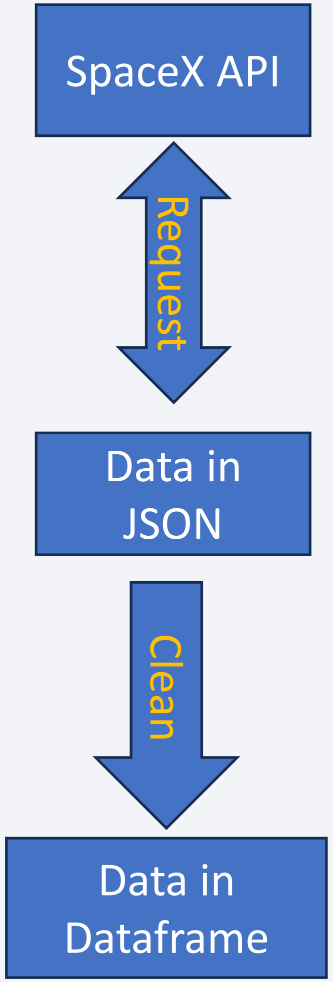
- The rocket launch dataset in JSON format was collected by making a get request to the SpaceX API. List of Falcon 9 and Falcon heavy launches in HTML tables format were also scraped from Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

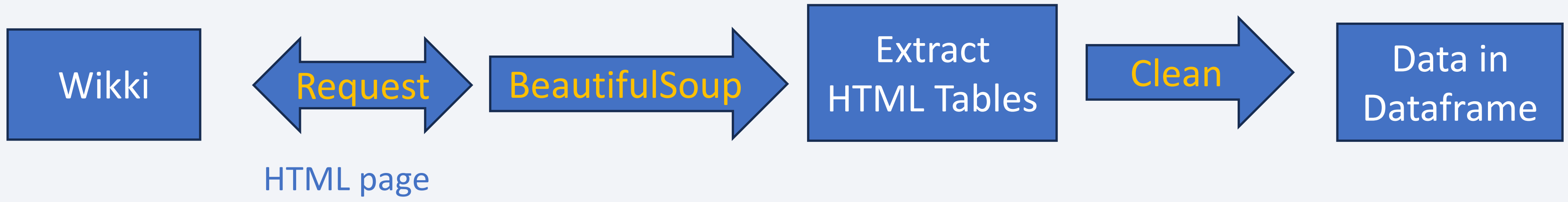
- The collected SpaceX dataset in JSON format was converted to a Data frame using Pandas for easy processing.

- Link to the GitHub URL of the completed SpaceX API calls notebook: https://github.com/olafem/Demo-IBM-/blob/main/SpaceX_API.ipynb



Data Collection - Scraping

- Falcon 9 launch records stored in HTML tables on Wikki are extracted using get request
- To access the HTML tables a BeautifulSoup object was created and the table converted to Pandas data frame.
- Link to the GitHub URL of the completed webscrapping: https://github.com/olafem/Demo-IBM-/blob/main/SpaceX_BeautifulSoup.ipynb



Data Wrangling

- Collected data were converted to Pandas data frame for cleaning and exploratory analysis and to determine what would be the label for training supervised machine learning models.
- GitHub URL: https://github.com/olafem/Demo-IBM-/blob/main/SpaceX_Data_wrangling.ipynb

- Determined the number of launches on each site
- Determined the number and occurrence of each orbit
- Determined the number and occurrence of landing outcomes per orbit type
- Created a landing outcome label

EDA with Data Visualization

- Categorical plots were made to visualize the relationship between variables such as flight number, payload mass, orbit type, and launch site; and how they affect landing outcomes.
- A barchart showing the success rate of each orbit
- A line plot of how success rate trended from 2010 to 2020.

- GitHub URL: https://github.com/olafem/Demo-IBM-/blob/main/EDA_Data_Visualization.ipynb

EDA with SQL

The following queries was conducted with SQL:

- Names of unique launch sites in the space mission
- 5 records where launch sites begin with CCA
- Total payload mass carried by boosters launched by NASA
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was achieved
- Booster types which have success in drone ship and with payload mass between 4000 and 6000 kg
- Total number of successful and failure mission outcomes
- Booster versions with the maximum payload mass
- Records of failure landing outcome in drone ship showing the booster versions, launch site, month and year.
- Ranking the number of successful landing outcomes between 04-06-2010 and 20-03-2017 in descending order.

- GitHub URL EDA with SQL notebook: https://github.com/olafem/Demo-IBM-/blob/main/EDA_with_SQL.ipynb

Build an Interactive Map with Folium

- The Folium map objects Circle and Marker are used to add a highlighted circle area and a text label on a specific coordinate respectively.
- The MarkerCluster object was used to create markers for all launch records indicating if the launch was successful(green) or failed(red).
- The Polyline object was used to draw lines and calculate distances between a launch site to its proximities, while the MousePosition to get coordinate for a mouse over a point on the map.
- GitHub URL: https://github.com/olafem/Demo-IBM-/blob/main/Interactive_map_Folium.ipynb

Build a Dashboard with Plotly Dash

- A Launch Site Drop-down Input Component was added to interact with a pie chart via a callback function which displays the success count for all launch sites and success launches for each site.

- A Range Slider was included to select payload to interact with a scatter plot via callback function displaying the payloads, booster versions and their success rates.

- GitHub URL of Plotly Dash lab: https://github.com/olafem/Demo-IBM-/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- EDA was conducted to determine Training Labels by:
  - Creating a column for the class
  - Standardizing the data using StandardScalar object
  - Spliting the standardized data into training data and test data using the function train_test_split

- The training data is divided into validation data, a second set was used for training data; then the models are trained and hyperparameters are selected using the function GridSearchCV.
- Classification models used for the prediction include: logistic regression, KNN, decision trees, and support vector machine(SVM).
- The GridSearchCV object is output for each classification model. Also, to make the best prediction the best parameter and accuracy on validation data is displayed for each model.
- Finally the accuracy on the test data was calculated for each model and the best classification model was selected.
- GitHub URL of predictive analysis lab: https://github.com/olafem/Demo-IBM-/blob/main/SpaceX_ML_prediction.ipynb

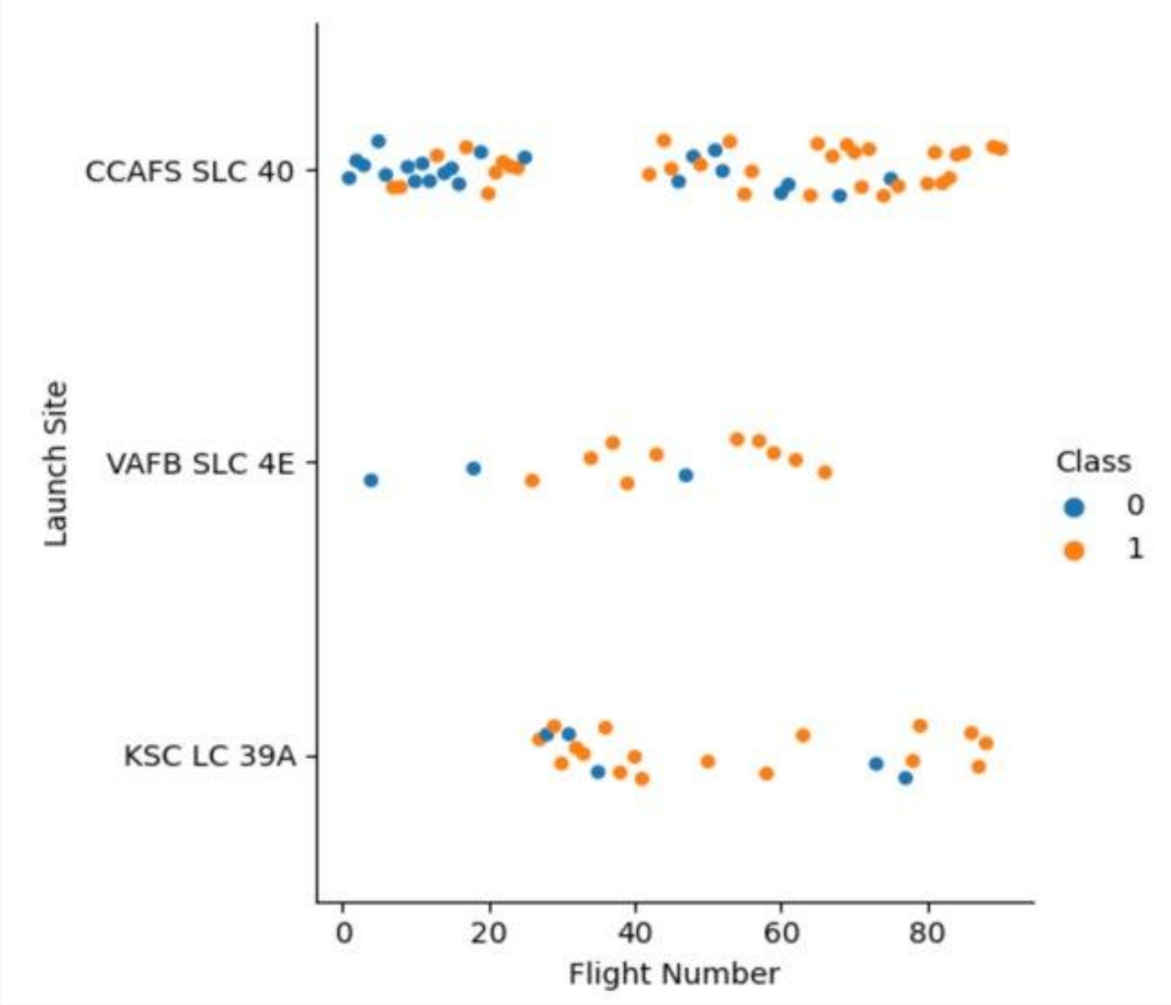
Results

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

Section 2

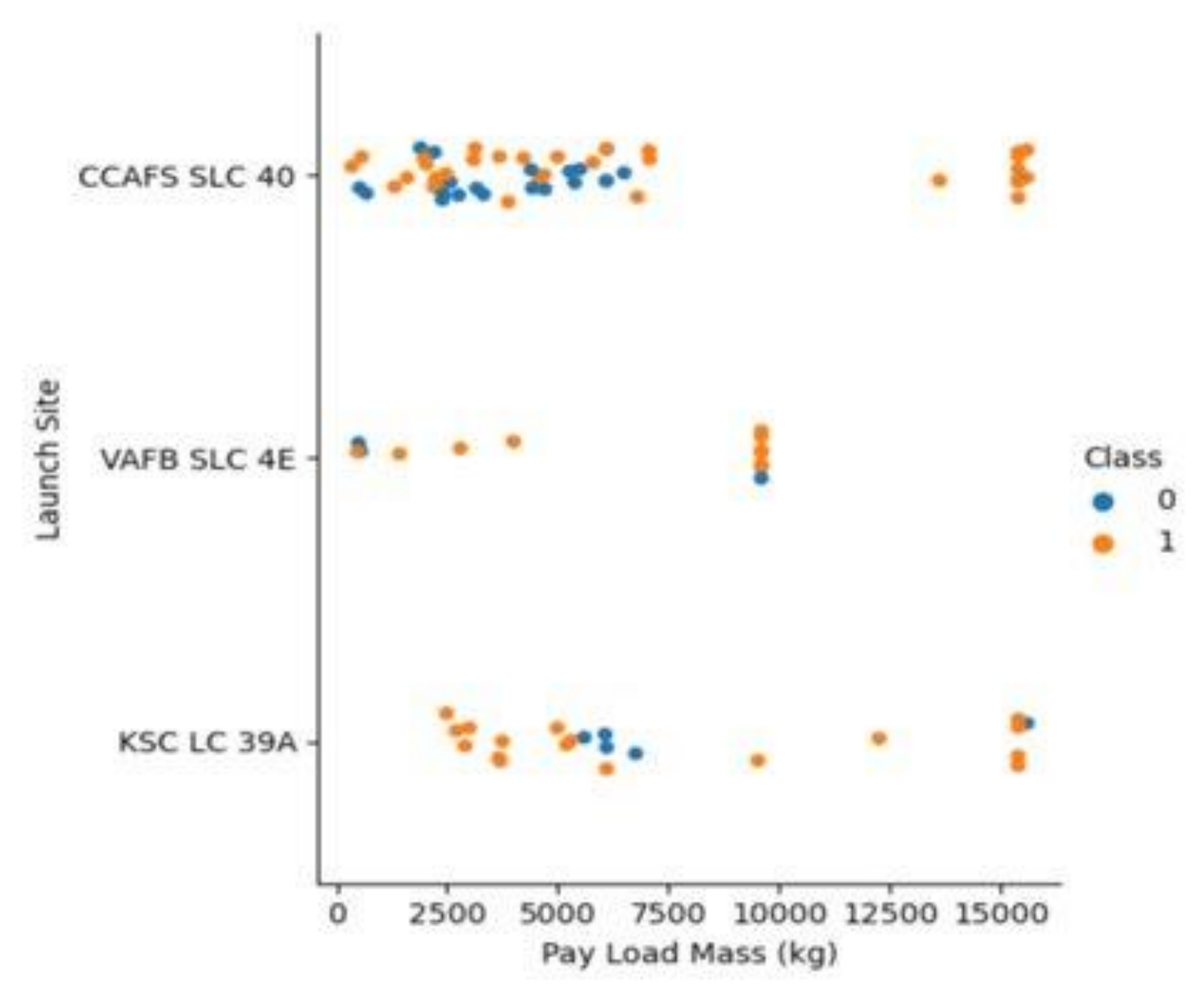
Insights drawn from EDA

Flight Number vs. Launch Site



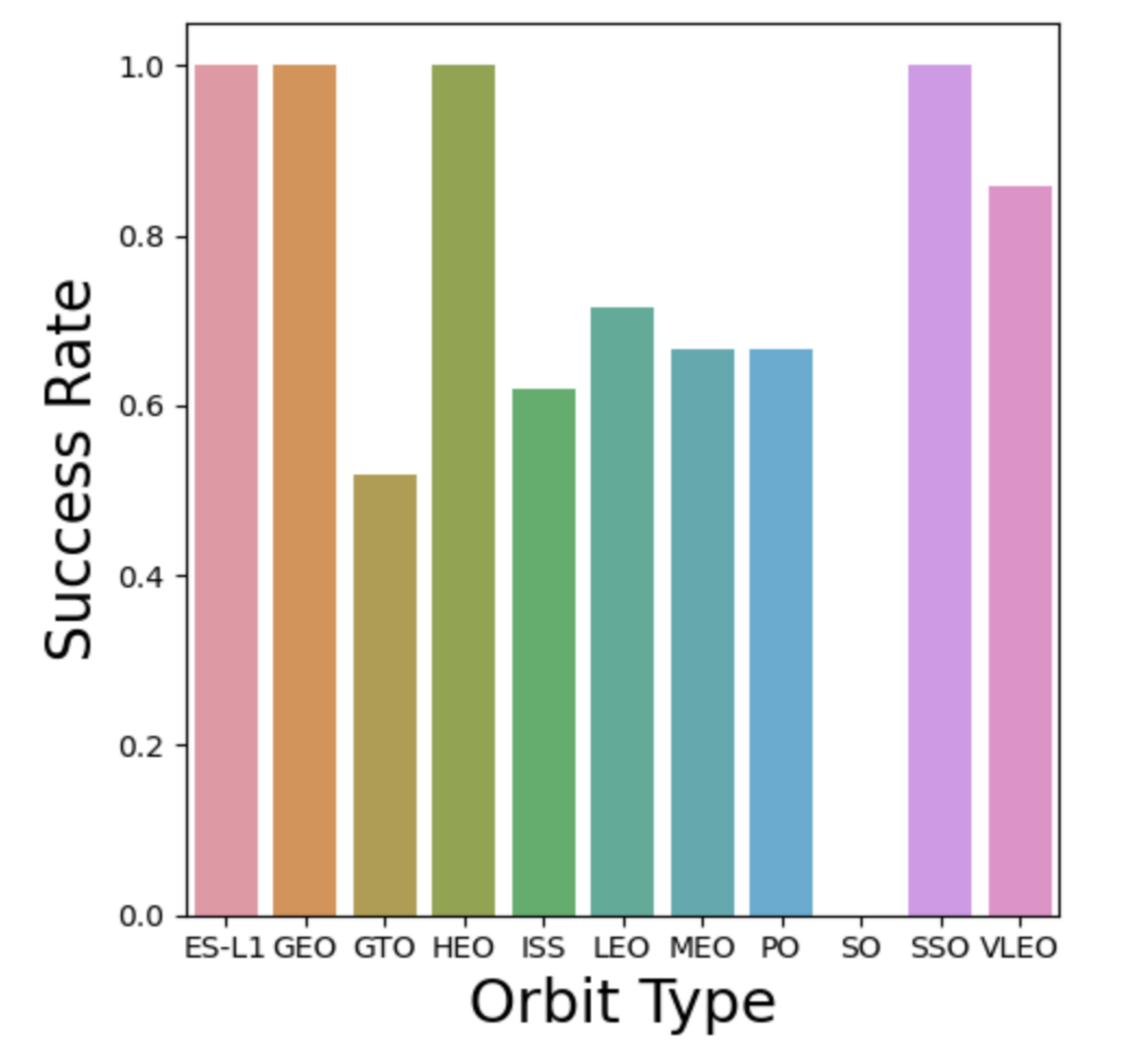
As the flight number increases the success rate for all launch sites increases.

Payload vs. Launch Site



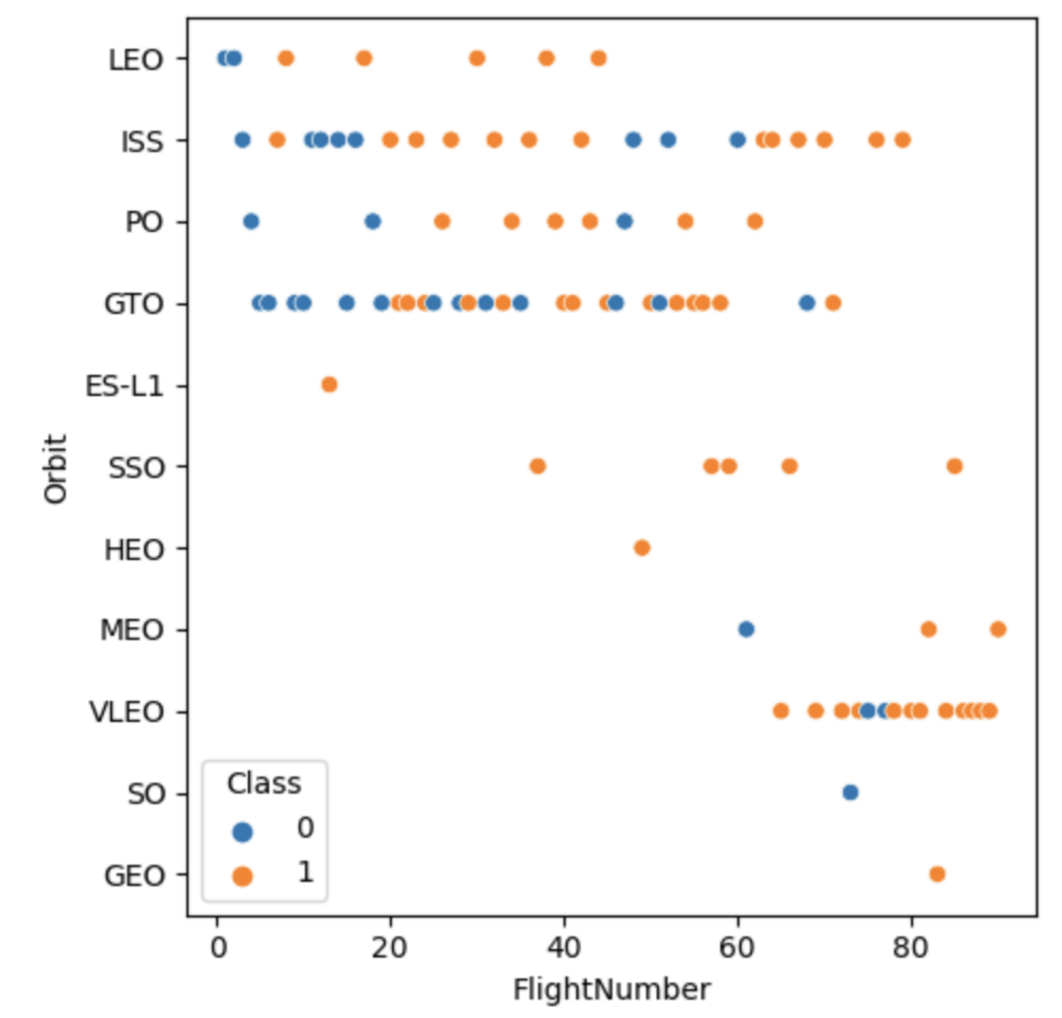
- No rockets launched for heavy payload mass (greater than 10,000kg) for the VAFB-SLC Launch site
- Success rate is high for payload mass at or above 10,000kg where applicable for all sites.

Success Rate vs. Orbit Type



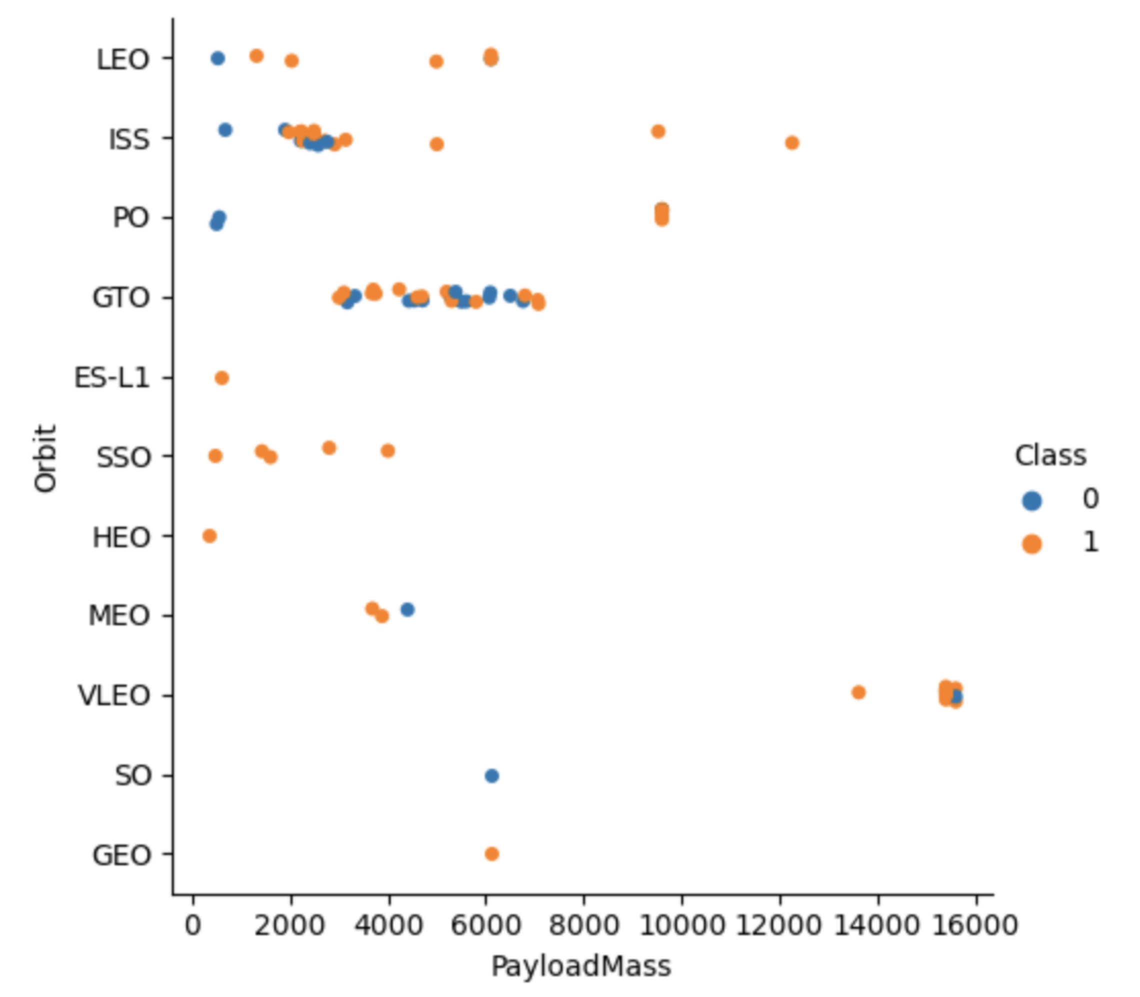
Orbits ES-L1, GEO, HEO, and SSO have high success rate

Flight Number vs. Orbit Type



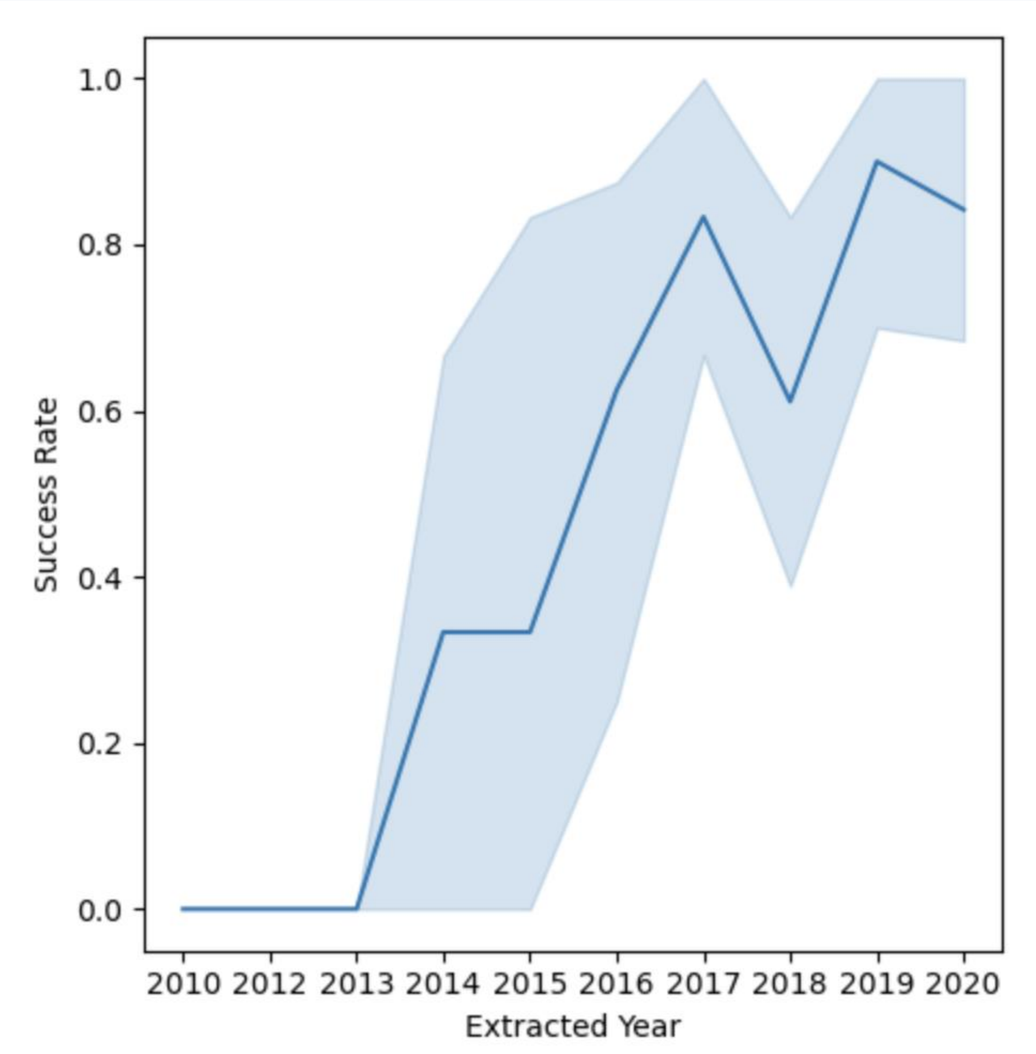
- For LEO success rate increases with increase in number of flight
- No relationship exists between flight number in GTO orbit
- Landing for SSO is successful, though flight number is at the higher end

Payload vs. Orbit Type



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

Launch Success Yearly Trend



- No success in landing from 2010 to 2013.
- Landing success rate increases from 2013 to 2020

All Launch Site Names

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

List of unique names of the Launch sites used in the space mission by SpaceX

Launch Site Names Begin with 'CCA'

CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

5 records where launch sites begin with the string 'CCA'

Total Payload Mass

Payloadmass
48213

Total payload mass carried by all the boosters launched by NASA (CRS)

Average Payload Mass by F9 v1.1

Average_Payload_Mass
2534.6666666666665

Average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

min(Date)
01-05-2017

Date when the first successful landing outcome in ground pad was achieved.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster version with payload greater than 4000kg but less than 6000kg and Landed successfully in drone ship

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	count(*)
Failure	1
Success	100

Boosters Carried Maximum Payload

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

Booster versions that carried a maximum payload mass

2015 Launch Records

Failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015

Month	Year	Failure_Landing_Outcomes	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

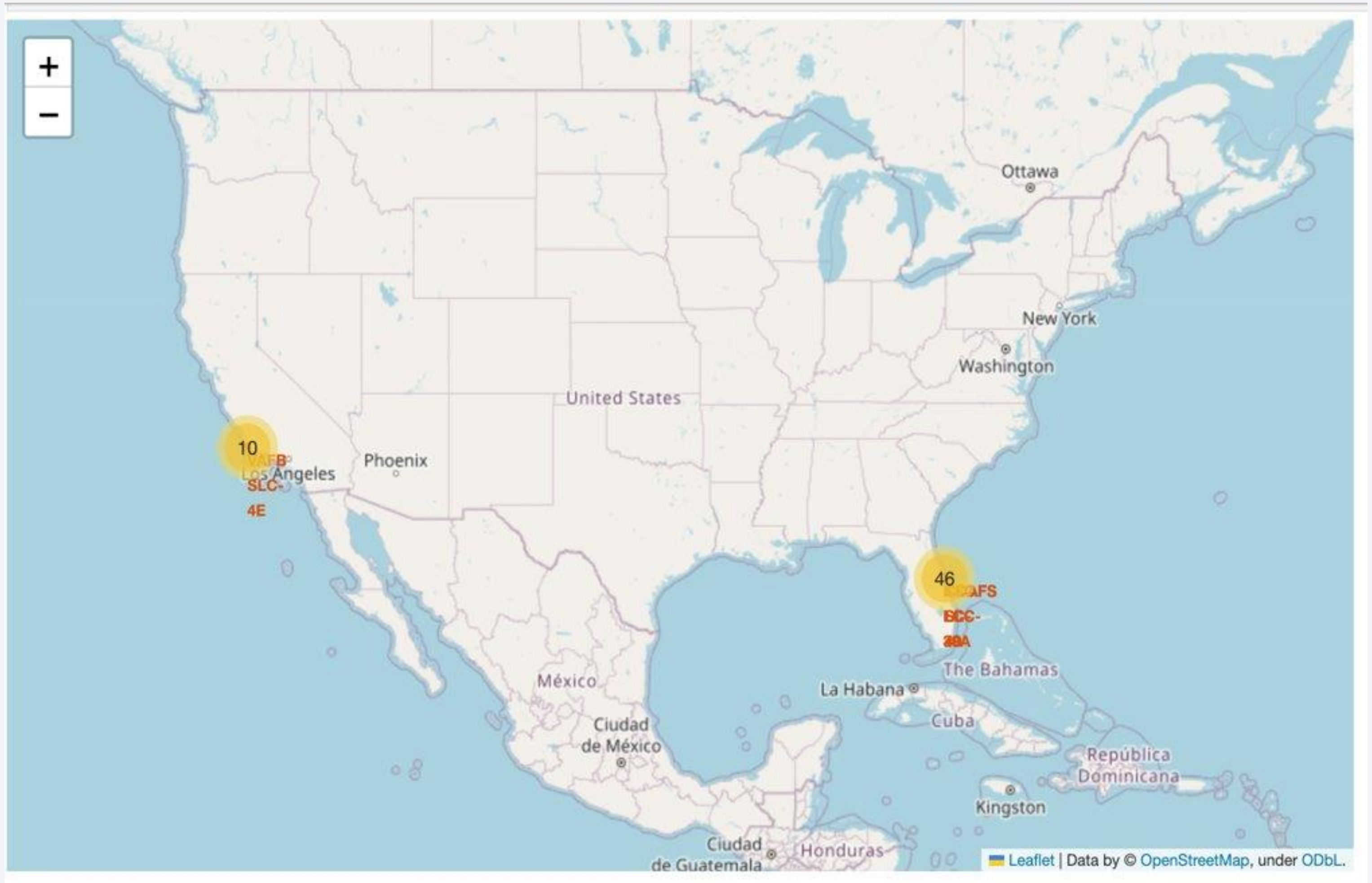
Rank of Successful landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Successful_Landing_Outcomes	count(*)
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

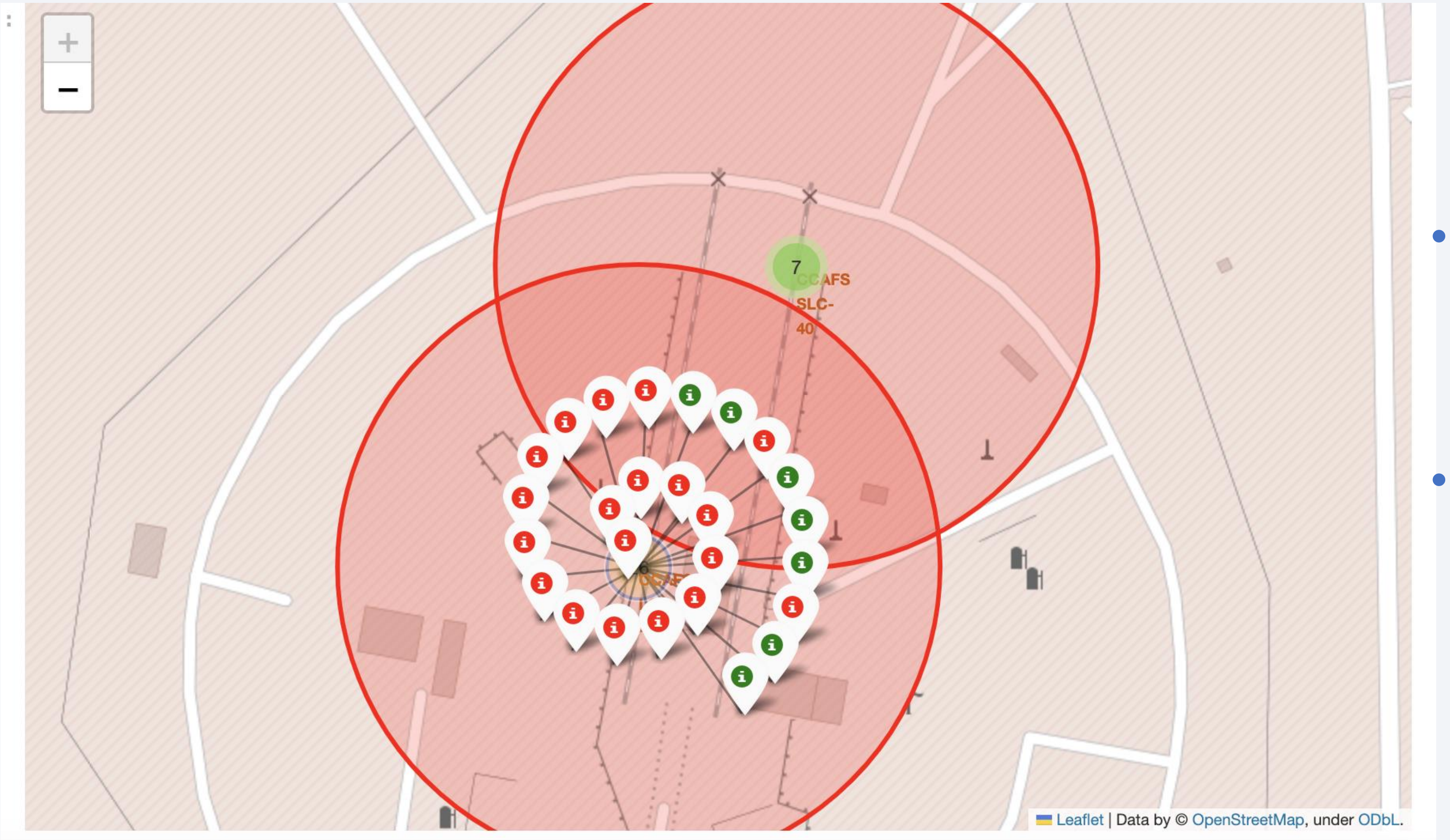
Launch Sites Proximities Analysis

Launch Sites Location on Folium Map



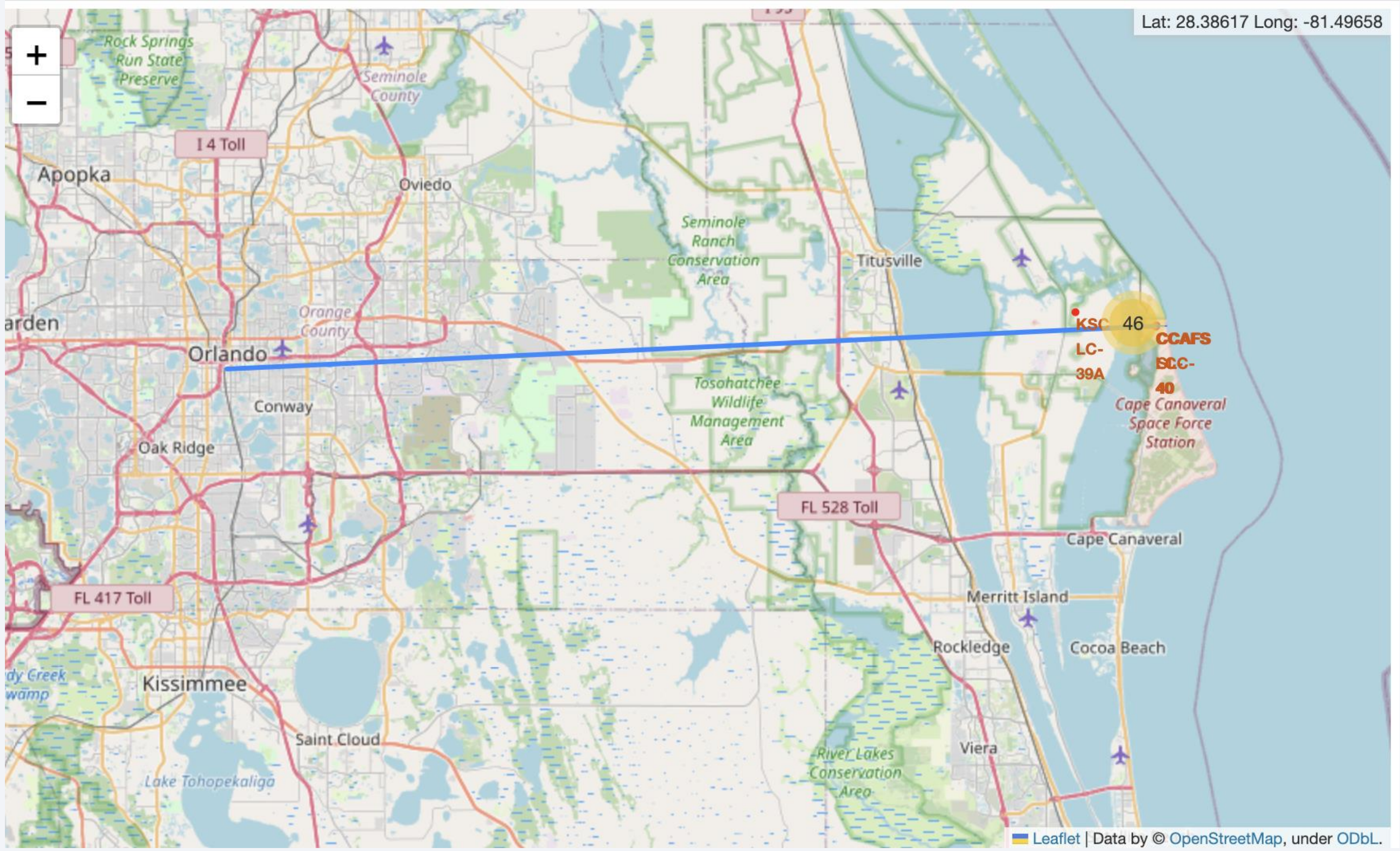
Launch sites in close proximity to coastline.

Launch Outcomes For a Site Displayed



- 26 Launch outcomes displayed for site CCAFS-LC-40 using MarkerCluster object
- Green colour indicate success and red colour indicate failure

Launch Sites and Proximities

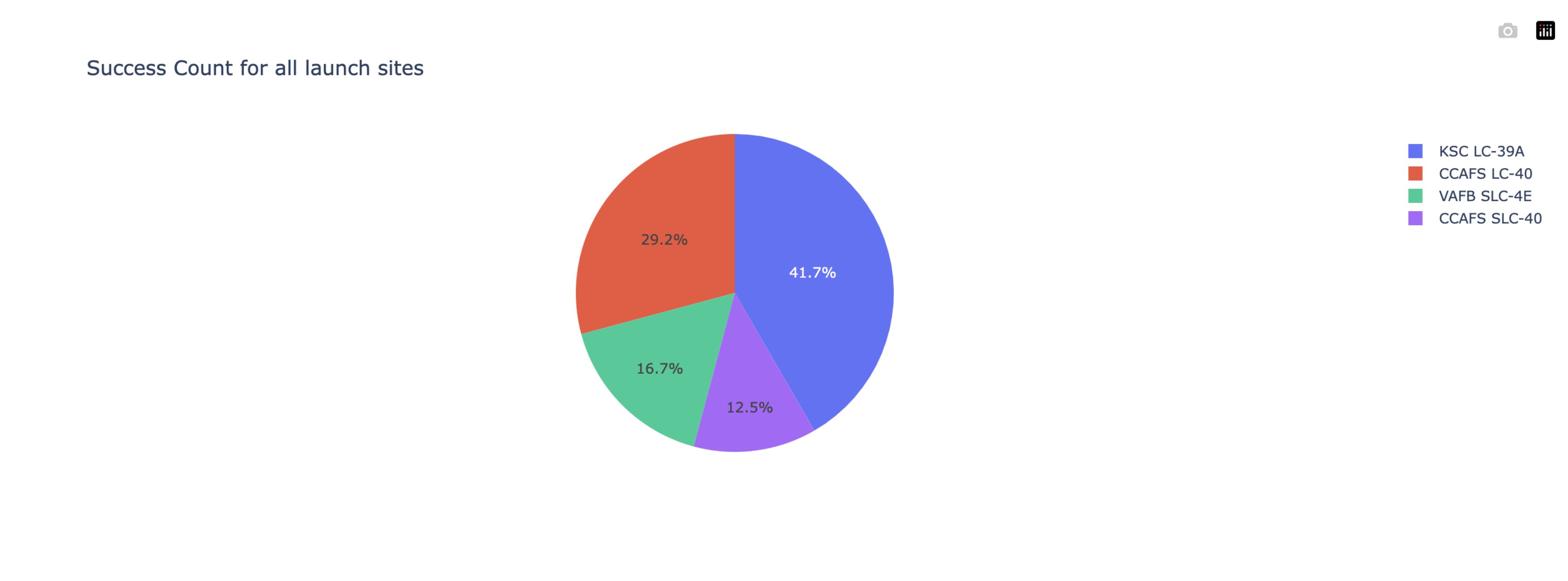


- Launch sites keep certain distances away from cities.
- Distance between Orlando and CCAFS-SLC-40 is 78.56KM.

Section 4

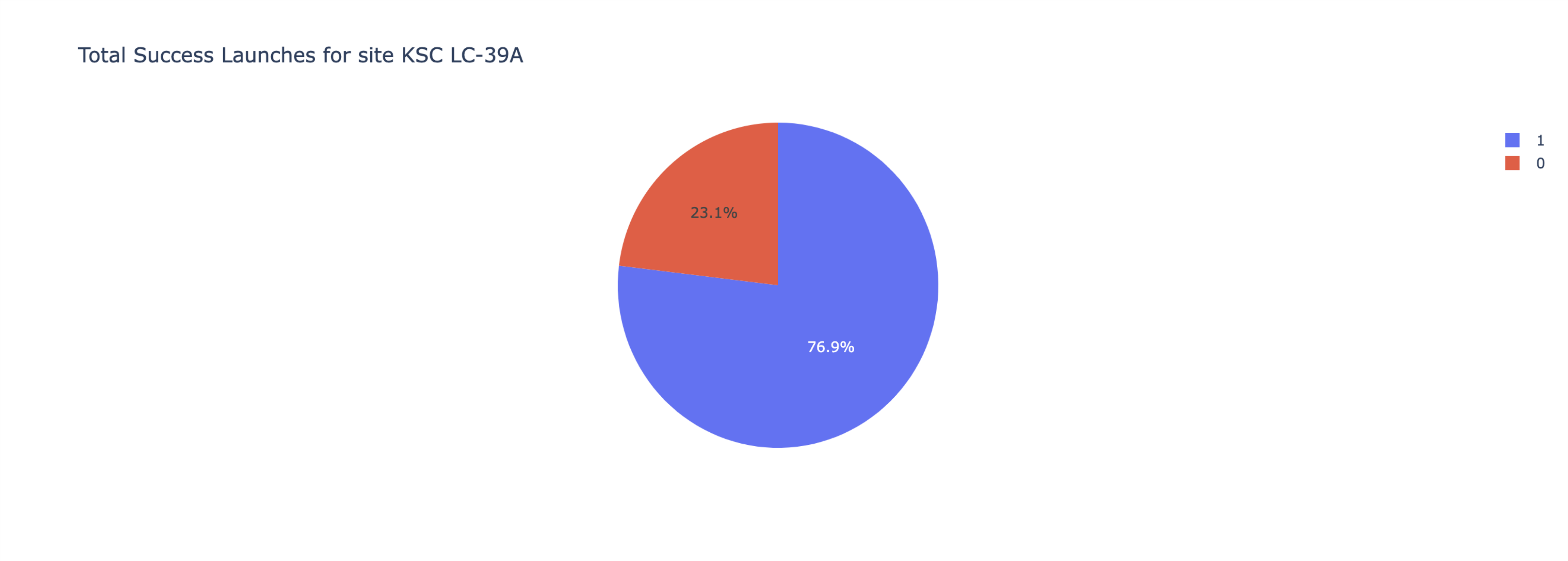
Build a Dashboard with Plotly Dash

Launch Success Count for all sites



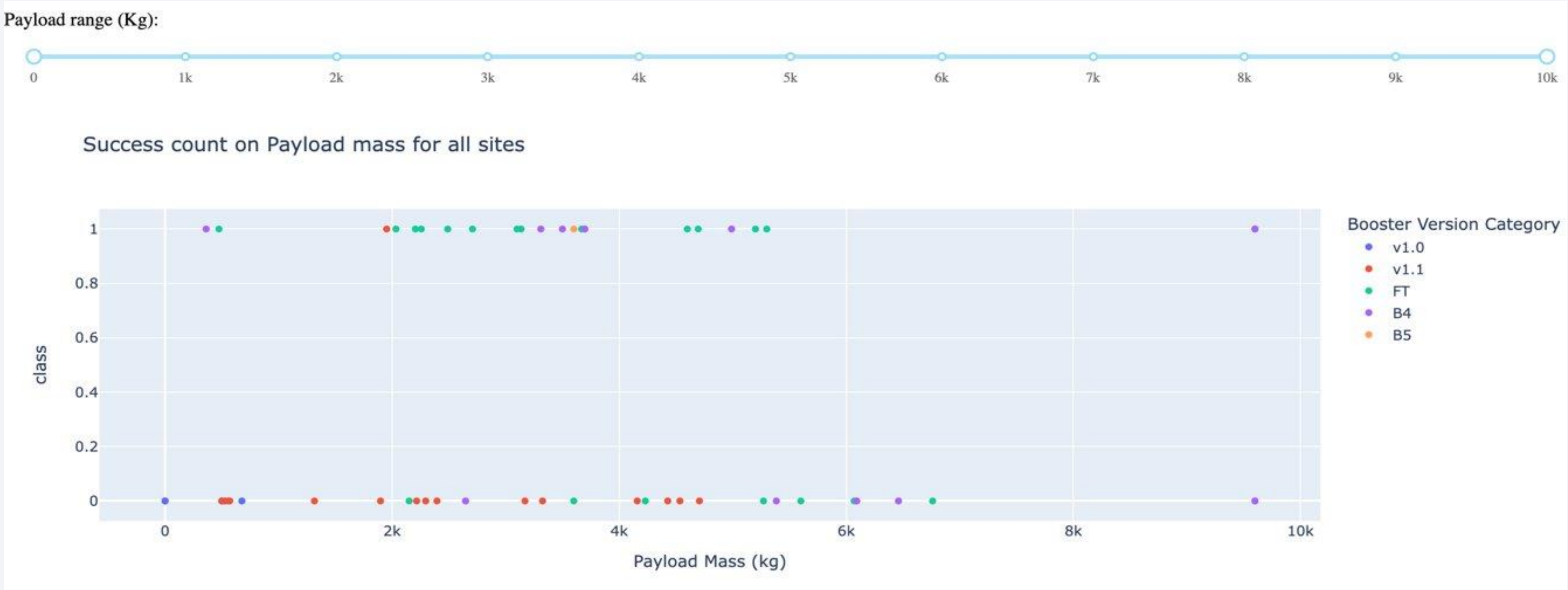
Site KSC-LC-39A has the largest successful count

Launch Site with the Highest Launch Success Ratio



Site KSC-LC-39A also has the highest launch success rate of 76.9%

Payload vs. Launch Outcome

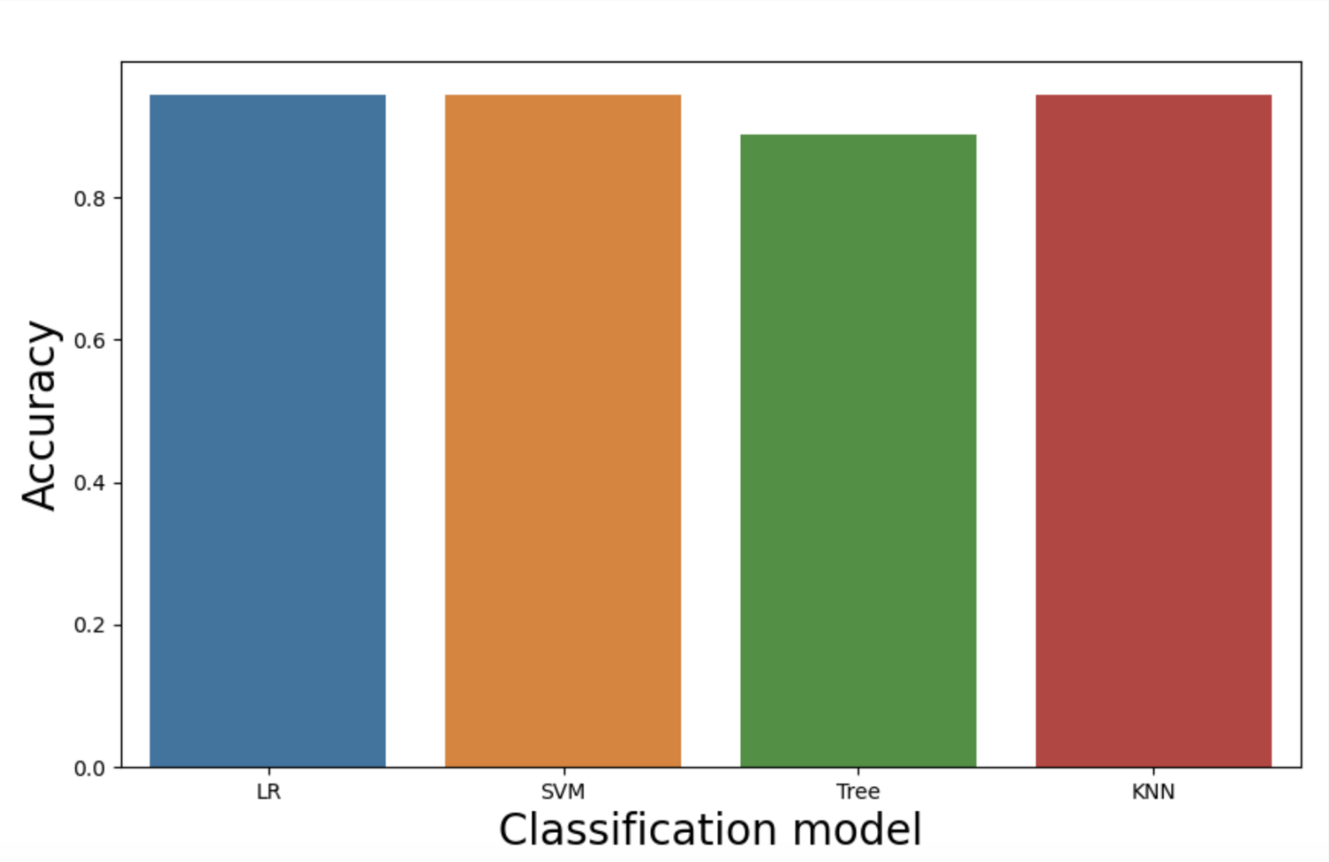


- The highest launch success rates occur in the payload range of 2000-6000kg
- The lowest launch success rates occur in the payload range of 500 - 7000kg
- F9 Booster version FT has the highest launch success rate, while v1.1 has the lowest launch success rate

Section 5

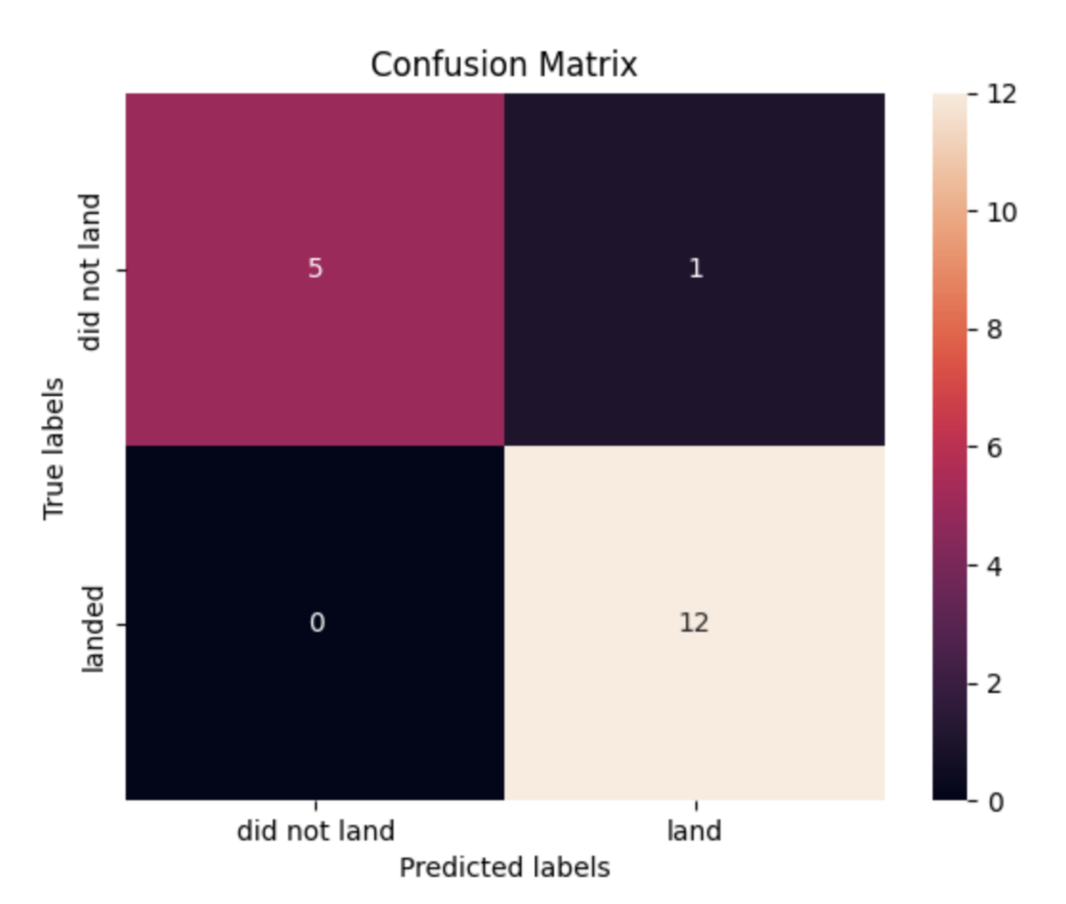
Predictive Analysis (Classification)

Classification Accuracy



LR, SVM and KNN all have similar accuracy of 94%

Confusion Matrix



- The model was able to predict correctly all of the 12 successful landings with no errors.
- While only one incorrect prediction was made for unsuccessful landings but predicted 5 correctly

Conclusions

- In conclusion, we were able to build a model that predicts the landing outcome of the SpaceX Falcon 9 rocket launch, with an accuracy of 94%.
- This work also reveals that with certain booster type and payload and 100% success rate can be achieved.
- The information would serve as a bedrock when estimating the cost of a rocket launch.

Insights

- This work reveals that certain booster types and payload will guarantee a100% success rate. This can help with budget planning for the next mission.

Thank you!