SpaceX Capstone Project Report

AbdulMujeeb Leasu

Table Of Content

This Presentation goes through the SpaceX Capstone Project. Displaying findings in an explainable way.

- Data Collection and Data Wrangling Methodology
- Predictive Analysis Methodology
- Visualizations
- Interactive Folium Map
- Explanatory Data Analysis
- Dash Dashboard using Plotly
- Predictive Analysis (Machine Learning)
- Extra Insights

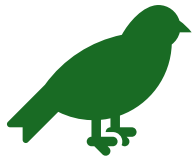
Data Collection

Using Github to gain data from the:

https://github.com/r-spacex/SpaceX-API

Using Pandas to Convert JSON files into a readable DataFrame.

From the Onset difficulties in loading the actual data. Using Webscraping seems to be harder and gives unreliable results.



Preprocessing the data.

Removing the Booster Version ‘Falcon 1’ as we only need ‘Falcon 9’.

Replacing Null values for Payload Mass with the average value.

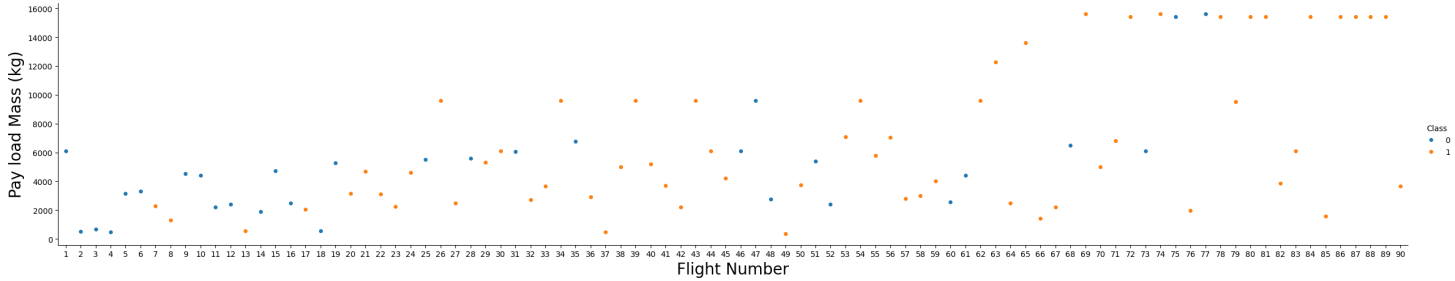
Data Wrangling

- I used Data Wrangling to set up the Data for further analysis
  - Organising the outcome values by creating a ‘Class’ Series. This Class Series contains values for different types of Landing Outcomes where the bad outcomes are considered 'None ASDS', 'False RTLS', 'False Ocean', 'False ASDS', 'True RTLS’.
  - I Obtained the percentage of successful lands which yielded a percentage of 66%
  - Saving the data into a new CSV for further Analysis

Visualizations

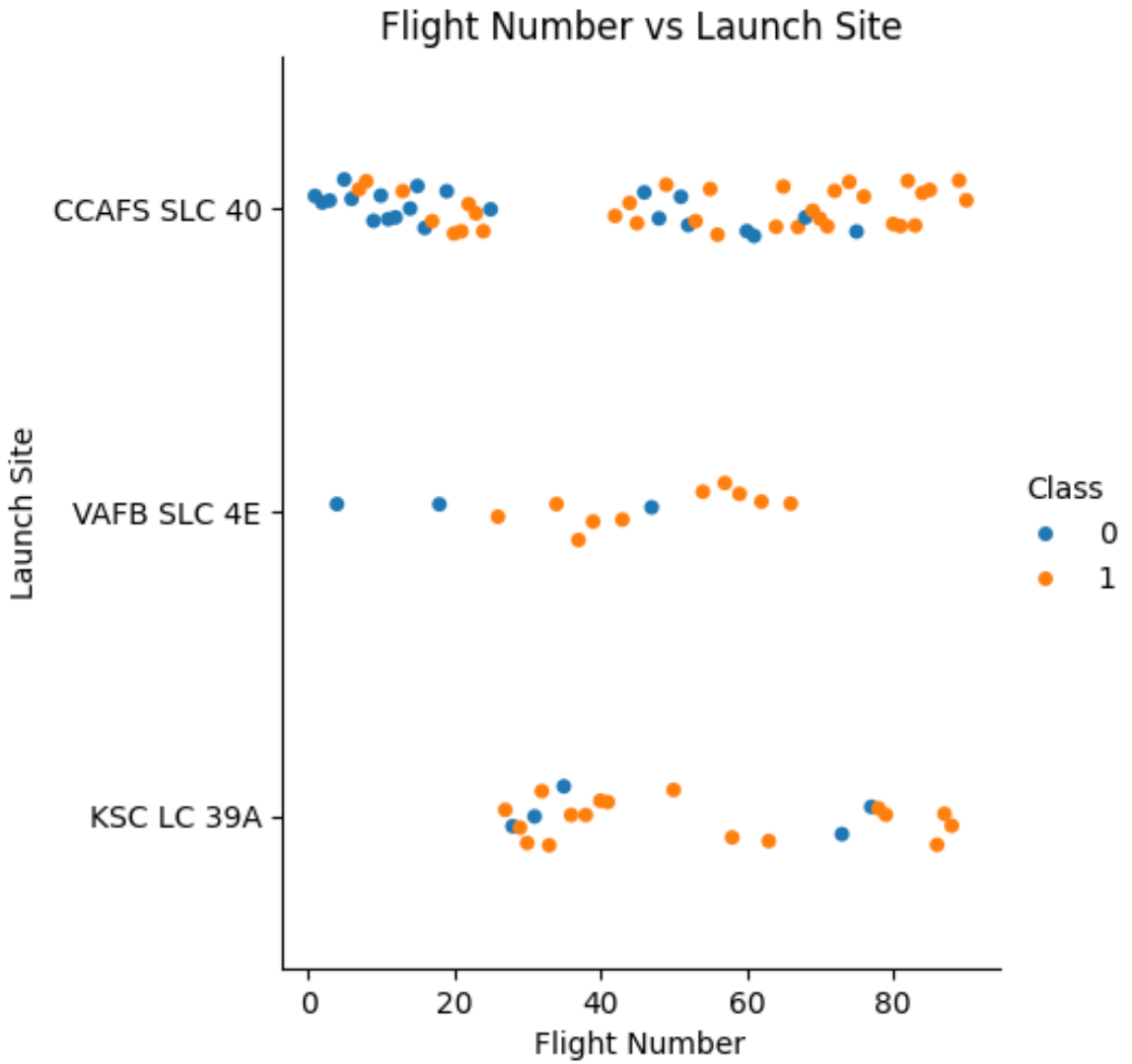
Payload Mass Vs. Flight Number

- There seems to be a slight bit of correlation between The payload Mass and Flight Number
  - This may be due to an increasing amount of confidence and successes as Payload mass is
    - Higher for later flight numbers.
  - We can also see a greater number of successes as Flight Number increases as there are more
    - Data points of Class 1.



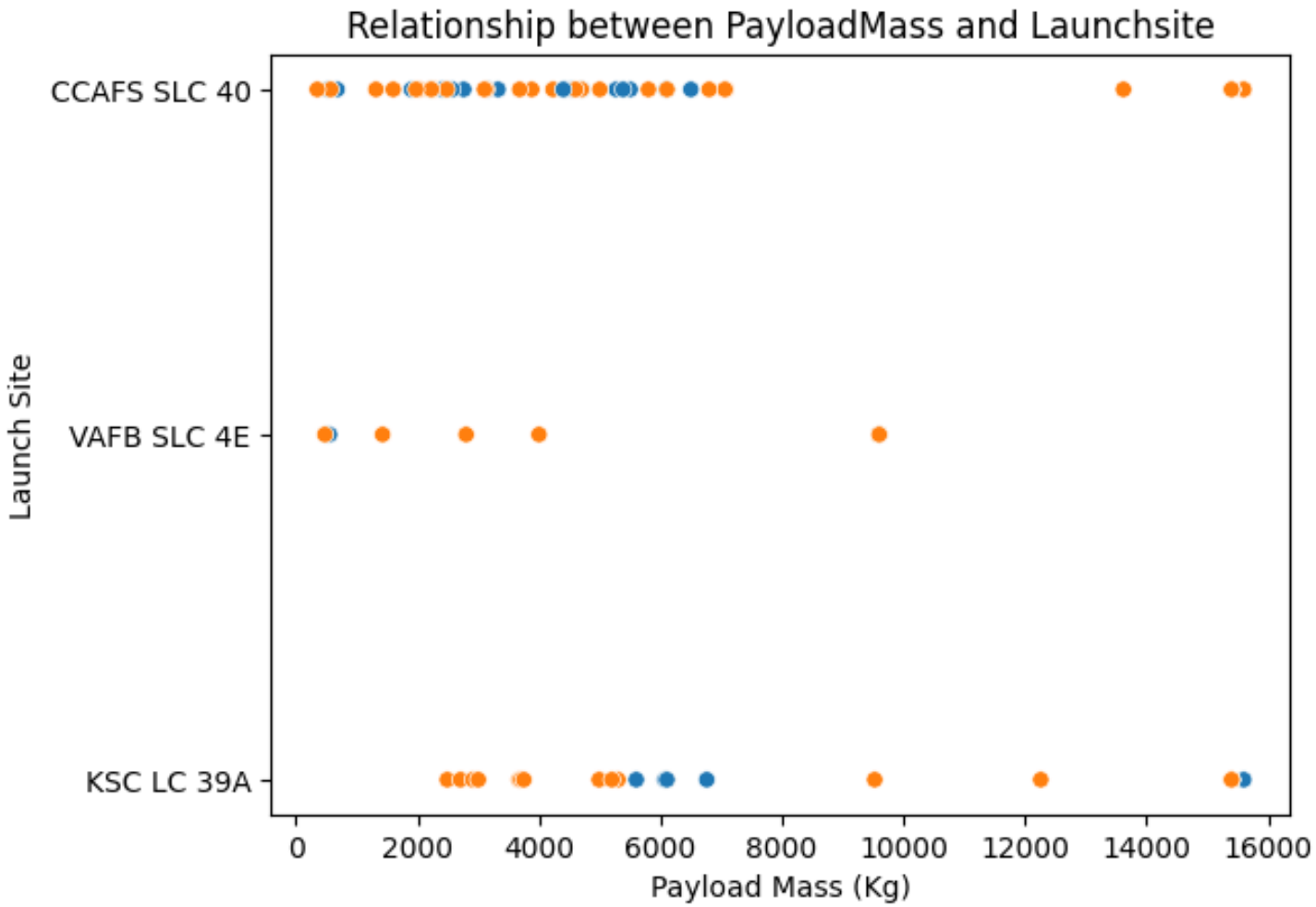
Flight Number Vs Launch Site

- CCAFS SLC-40 Launch Site has the
- highest number of flights. This seems
- to be the main launch site
- VAFB SLC 4E has the highest success
- to non success ratio
- KSC LC 39A. Clearly shows that the flights
- Switch to this launch site temporarily
- as there a gap between CCAFS SLC 40



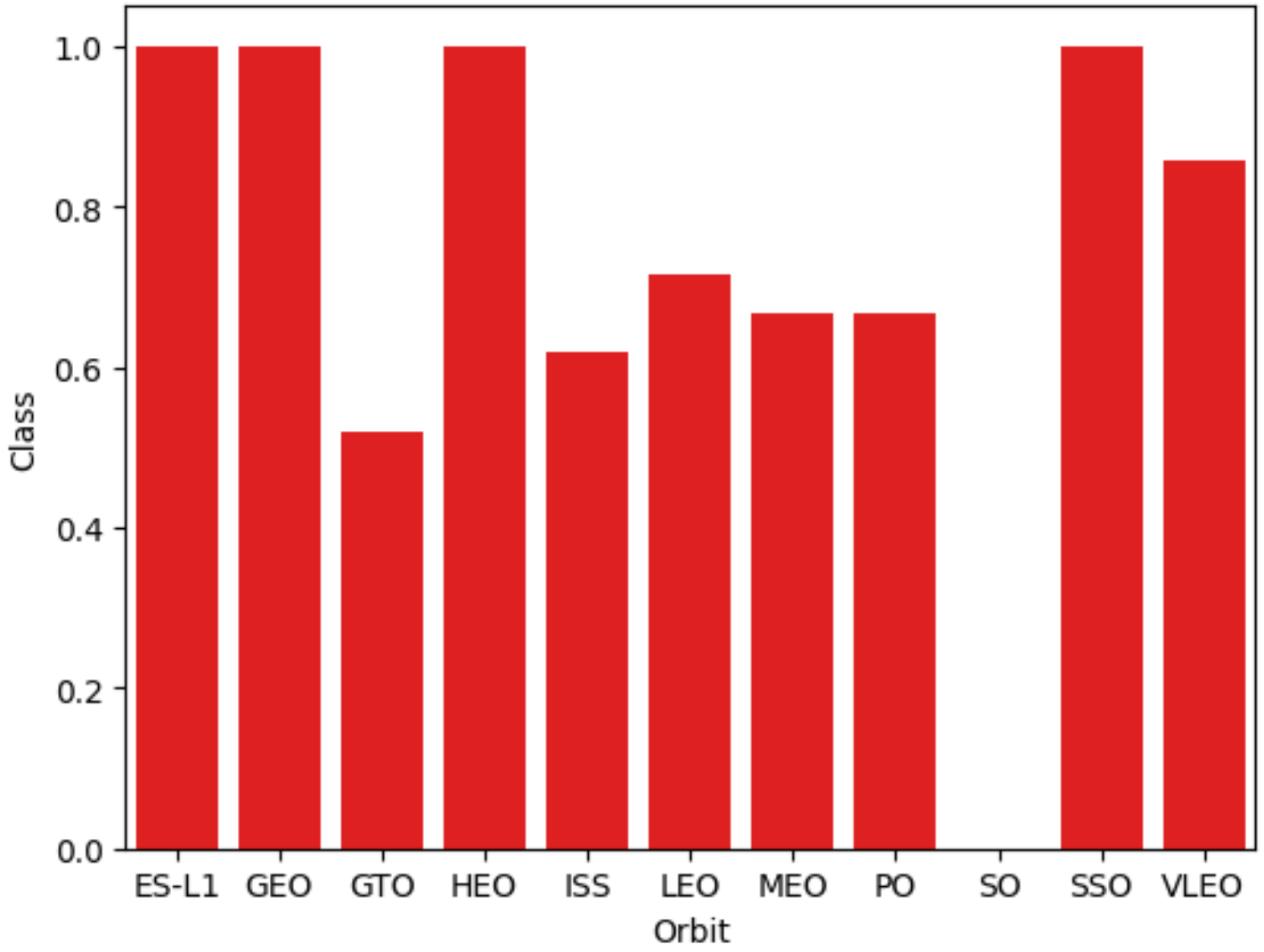
Payload Mass vs LaunchSite

- SO orbit level clearly has no good outcomes
- While ES-L1, GEO, HEO, SSO have a success rate of 100%



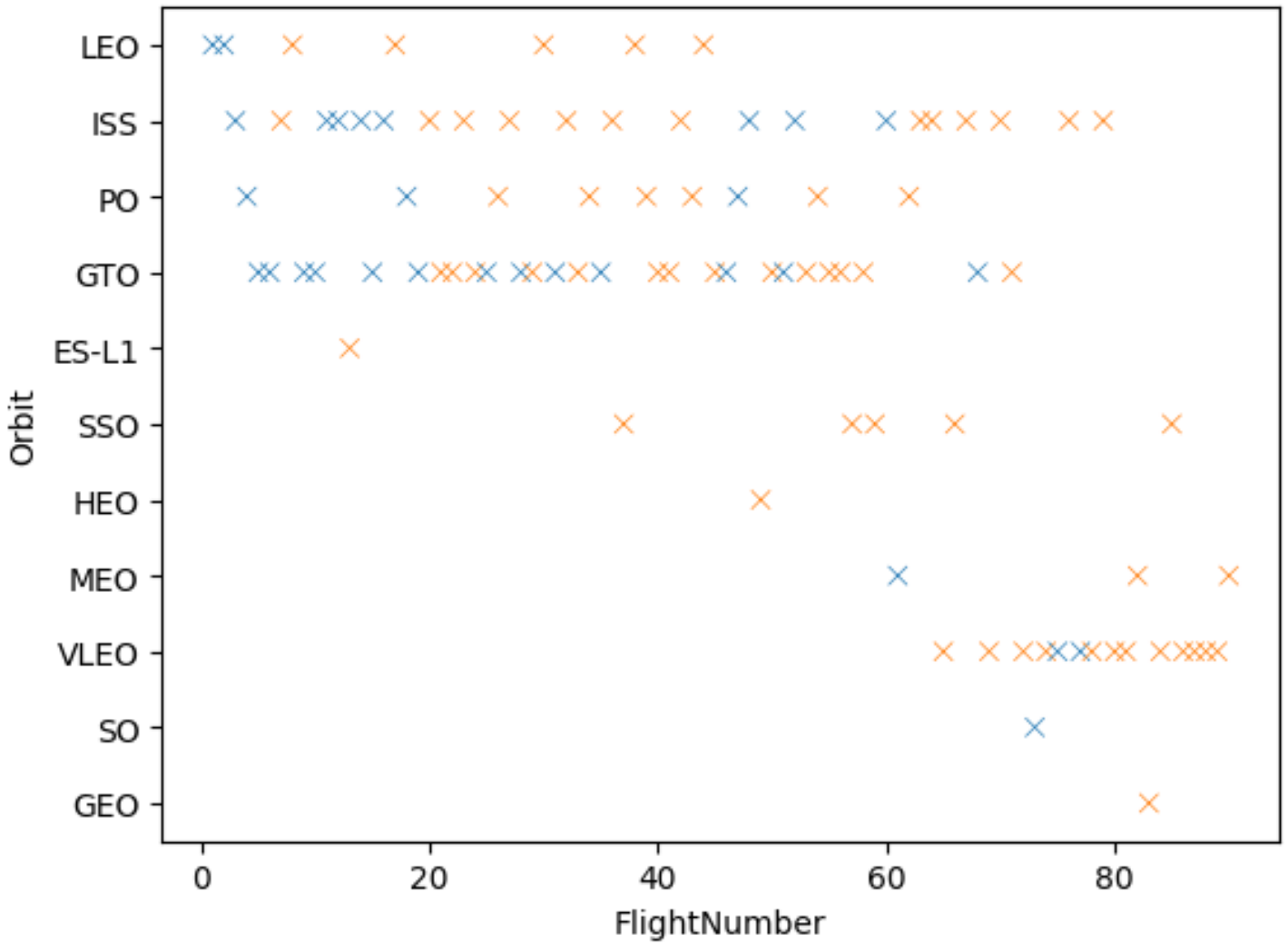
Payload Mass vs LaunchSite

- At CCAFS SLC 40. a smaller payload mass is used for most of the launches as there is a clear left skew in the data.
- At KSC LC 39A a higher average payload mass is used as oppose to CCAFS
- The plot shows that the success rate is over 50%



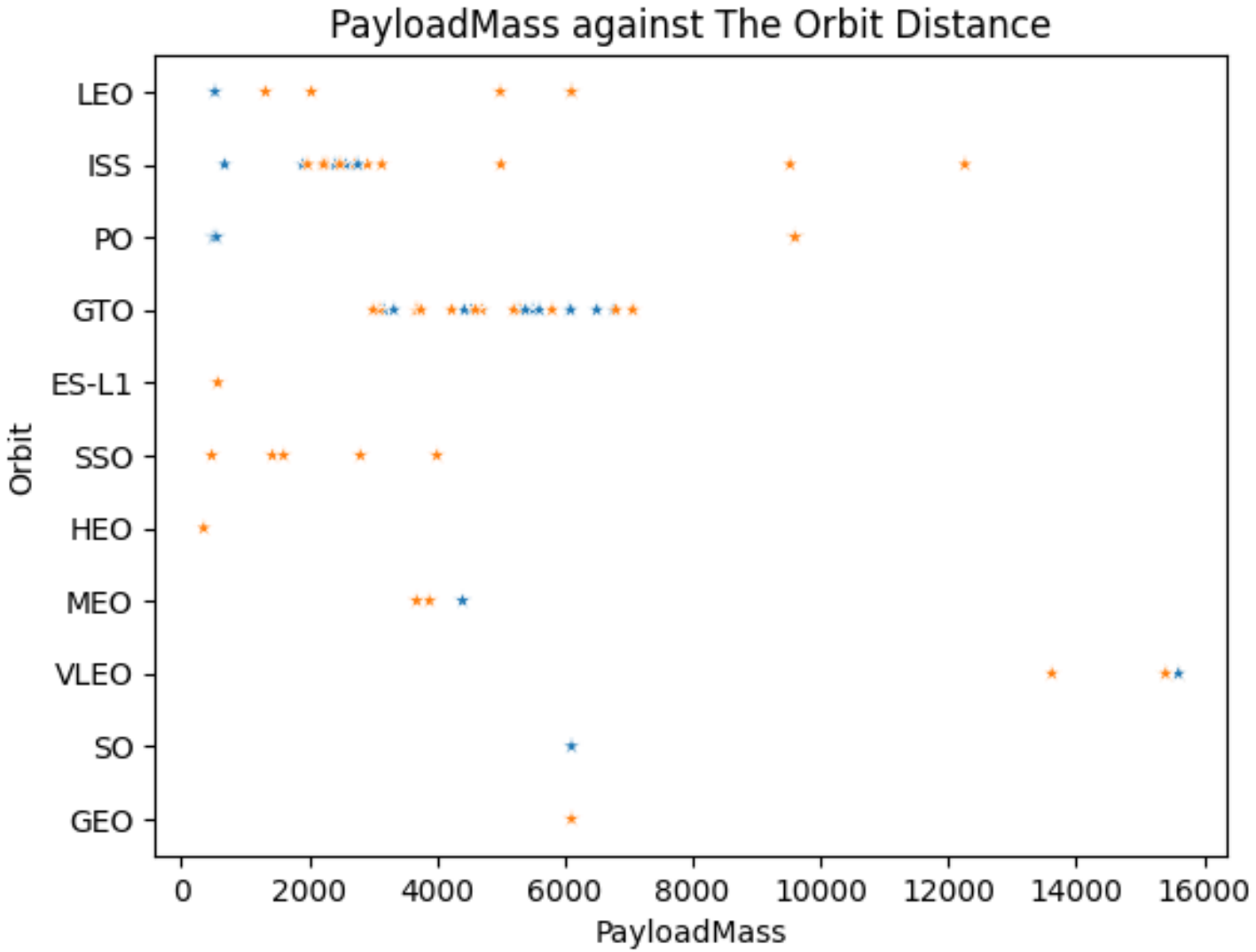
Orbit Distance Vs FlightNumber

- There seems to be a greater orbit distance
- as the flight number increased. This is probably
- due to an increase in confidence. There is also
- a greater amount of variables classed as
- a success.



Payload Mass vs Orbit Distance

- Greater increase in Payload Mass for further
- Orbit distances. GTO has constant values of Payload
- Mass compared to other orbit distances.



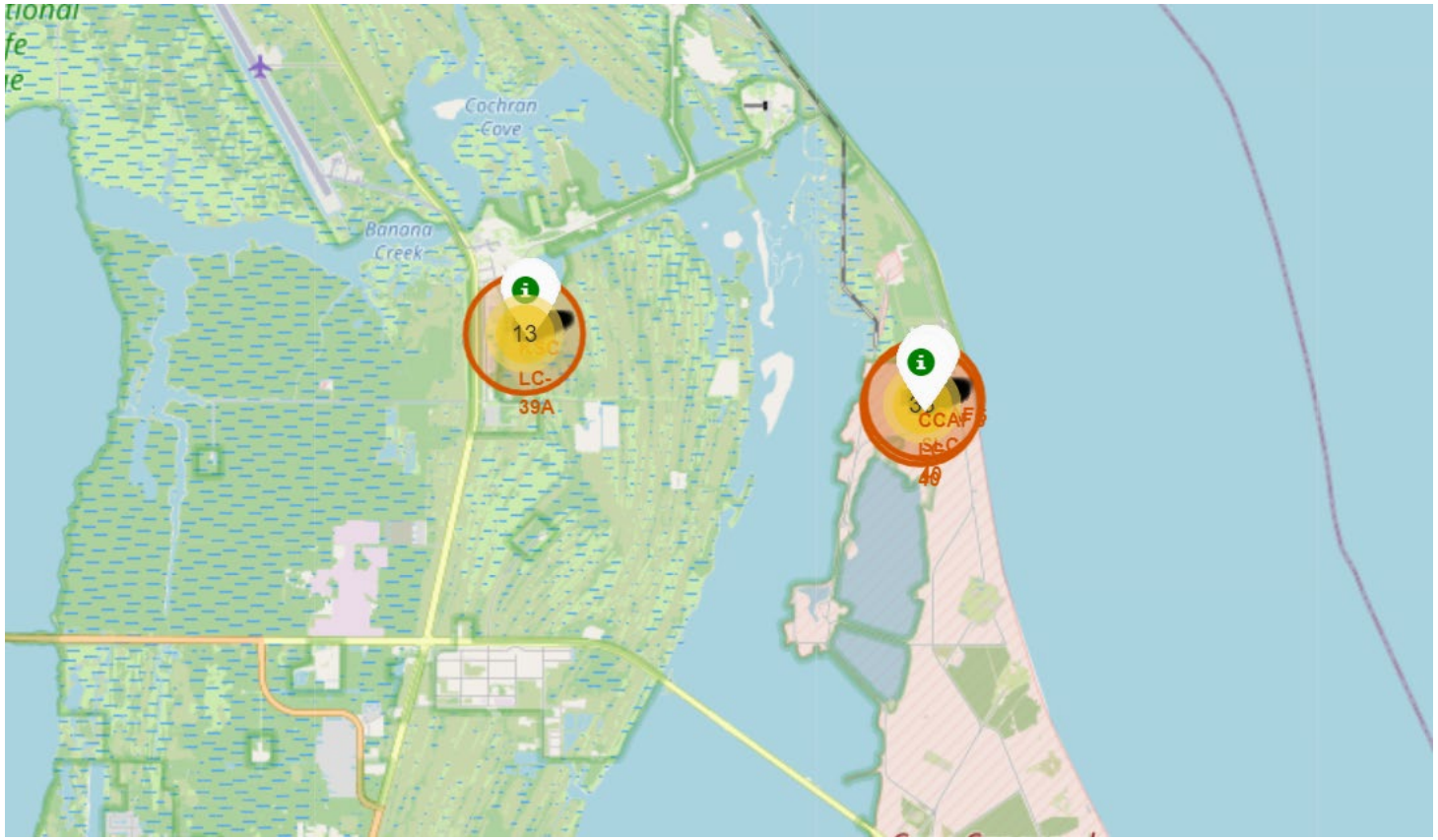
Folium Maps

Geographical locations and number of Launches per LaunchSite

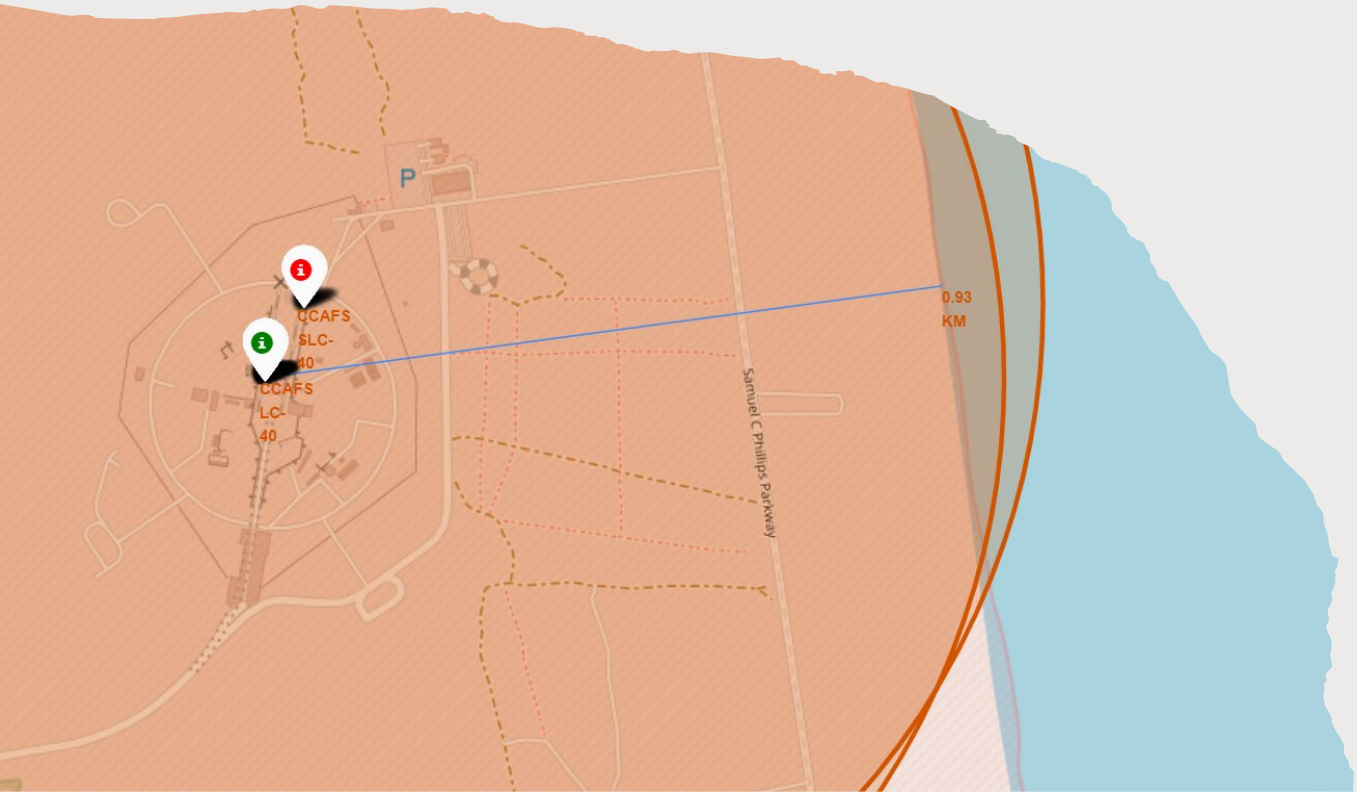
- More Launches occurred in Florida (56) than in California (10).

Florida LaunchSites

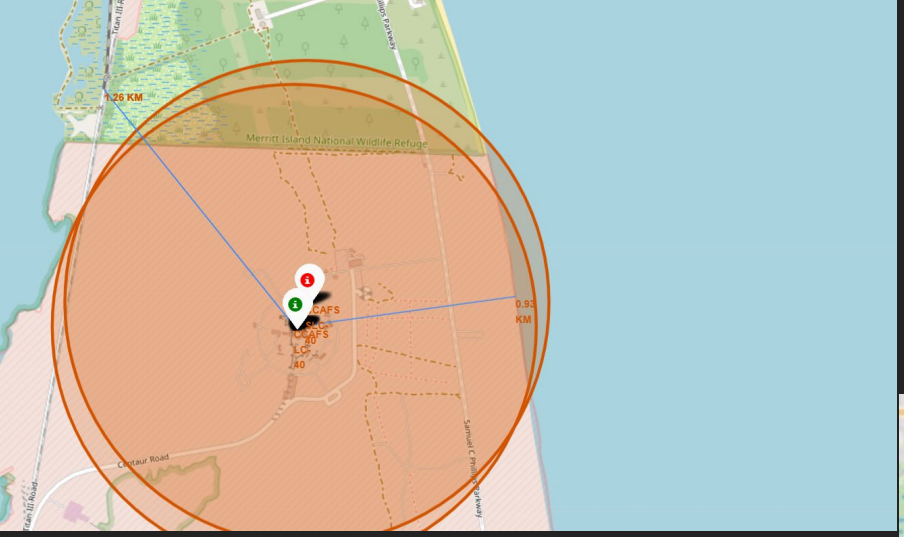
- 3 Launch Sites are in close proximity to each other. Showing that CCAFS LC-40 having the most Launches overall



Patterns obtained in selected locations

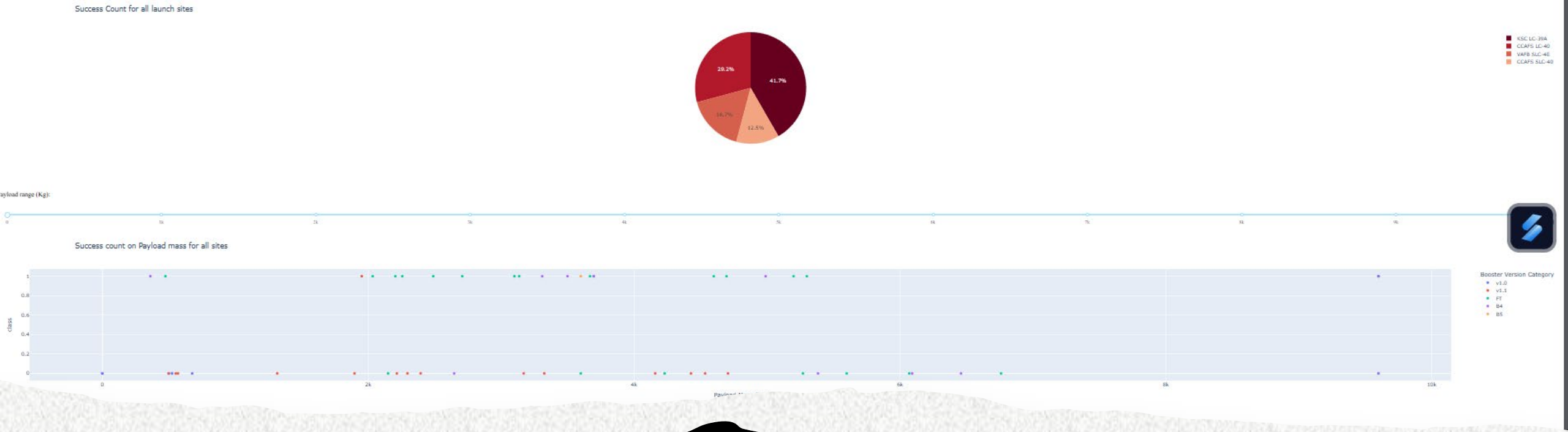


- The map Prior showed that the Launch Sites where in Florida and California. These are located near the coastline.
  - This is due to the fact that the Stage 1 parts rockets returns down for relanding. The possibility of an unsuccessful rocket landing still exist so it is preferred to launch the rockets near uninhabited areas.
- Seen 0.93km proximity with the coastline.
- Any other relationships…?



Other

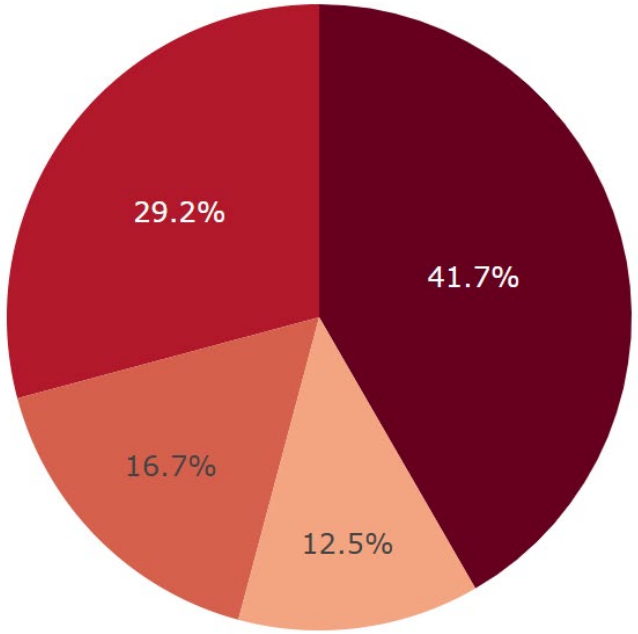
- Cities are usually located a far distance away. The closest city away from a Launch Site is 16.41km
  - This is again due to wanting a launch site away from inhabited areas.
  - This is also due to the amount of noise created per rocket launch
- A highway is located in close proximity to the launch sites.
  - Used for transporting Rockets etc.
- Rail Tracks are also nearby.
  - Also Used for transportation of materials and rockets.



Dash Dashboard

- Using an Active Dashboard to display data. This presents the success rates in each site.

All sites Success rate

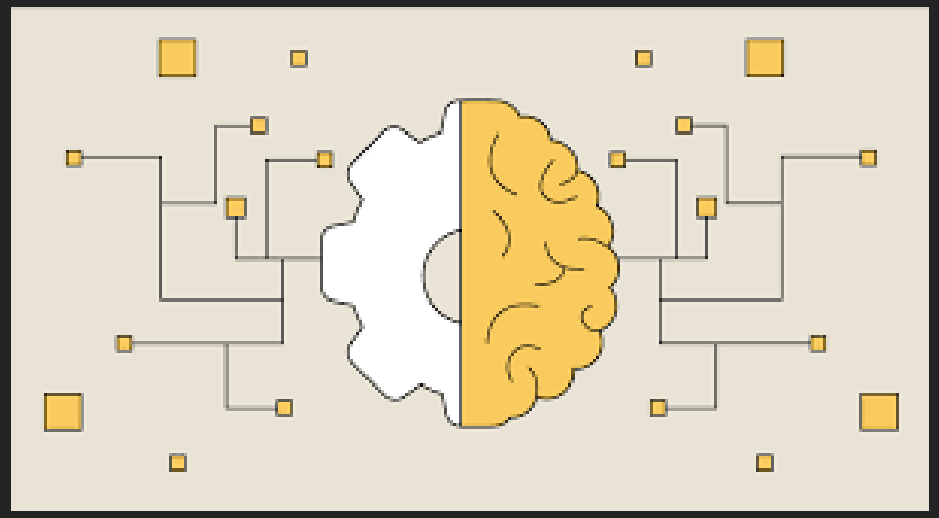


KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40

Machine Learning

Explanation

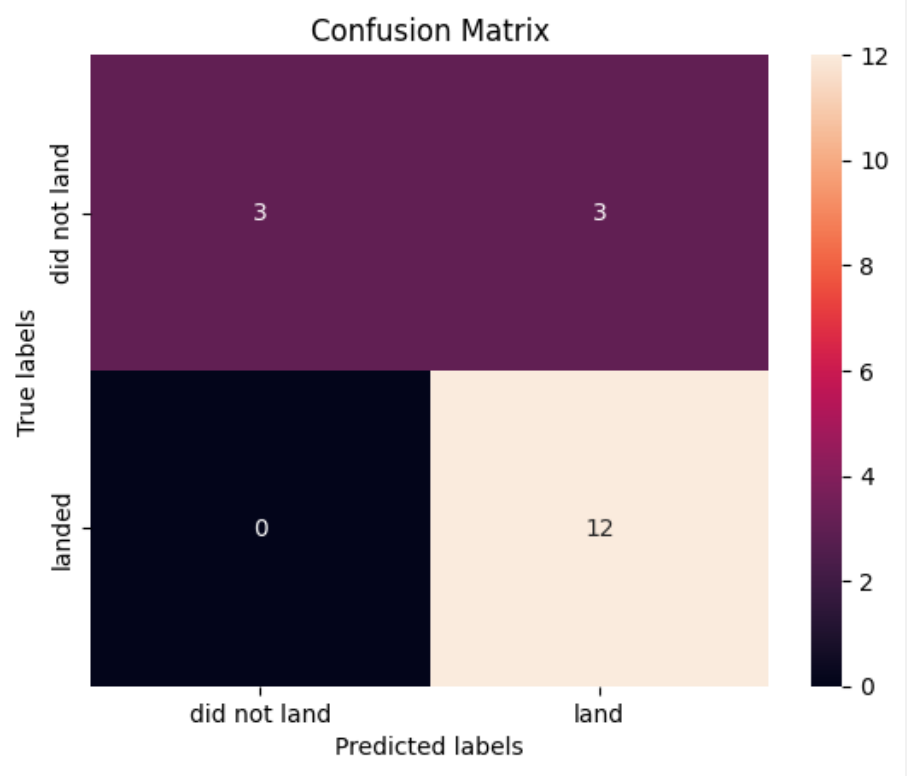
- Machine learning models are used to predict new pieces of data.
  - I use multiple models such as Logistic Regression, Decision Trees etc.
  - Split the data up into Training and Testing data then review the model accuracy to see if the model is suitable.



Logistic Regression Model

- Using a Logistic Regression to obtain the Accuracy

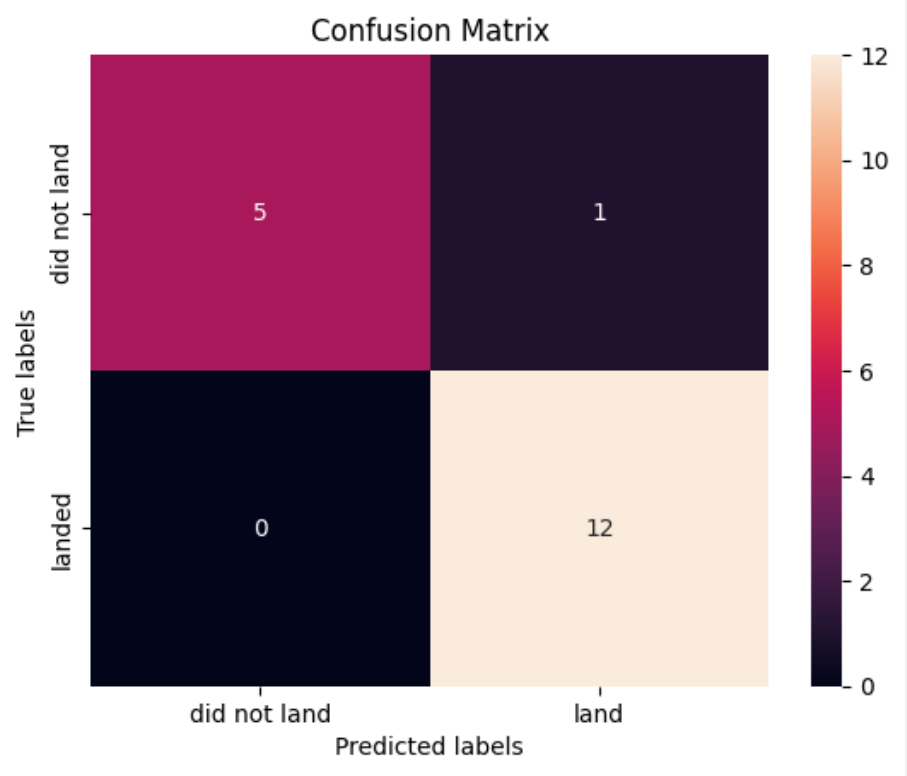
- Test Accuracy Obtained: 83%



Decision Tree Model

- Using Decision Tree to obtain the Accuracy

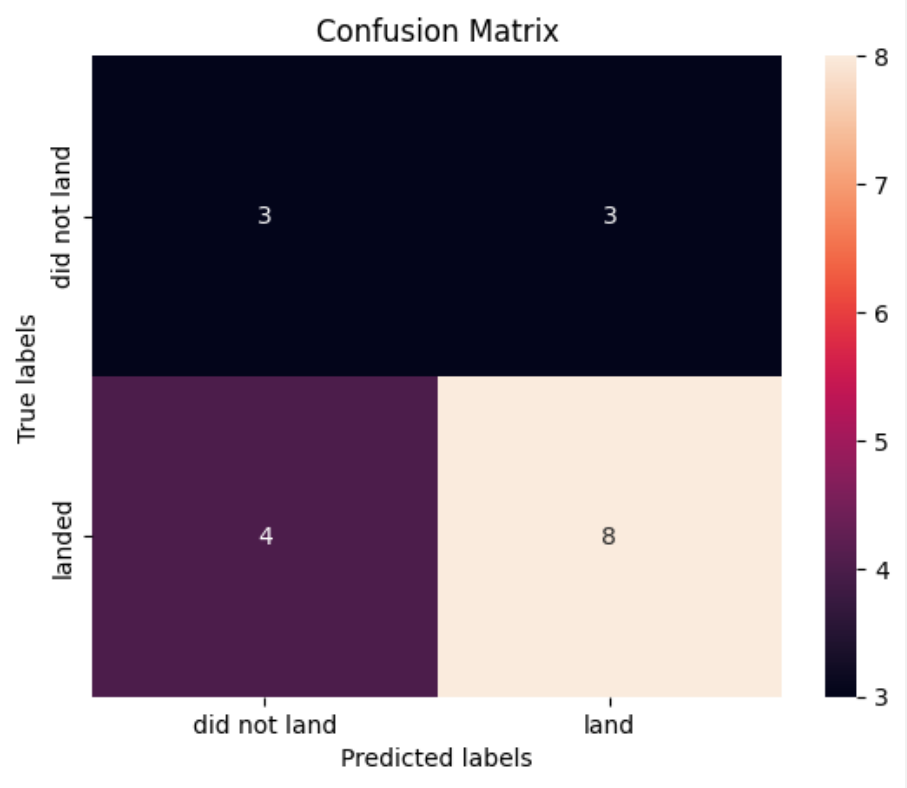
- Test Accuracy Obtained: 94%



K-Nearest-Neighbors Model

- Using KNN to obtain the Accuracy

- Test Accuracy Obtained: 61%



Machine Learning Conclusion

- The Decision Tree Model Clearly has the highest Accuracy at 94%. This model should be used for predicting outcomes for future flights.
- KNN should definitely not be used as the test accuracy is 61% this is very low!
- Logistic Regression Model can possibly be used to predict with an accuracy of 83% but should not be used for very large datasets.

Conclusion

- In Conclusion the Analysis shows some decent insights. There is a clear upwards trend in successes over the years as SpaceX used new and innovative methods to land there rockets. This has clearly had a profound effect on the price of space exploration as now the expensive Stage 1 of the rocket can be reused.
- The Decision tree model can also be used to predict further rocket launches to further increase the success rate.

Credits

- IBM Skills Network. This project was done for the Data Science Capstone Project Course on Coursera.
- Me Ofcourse ☺
- Thank You for going through the course.