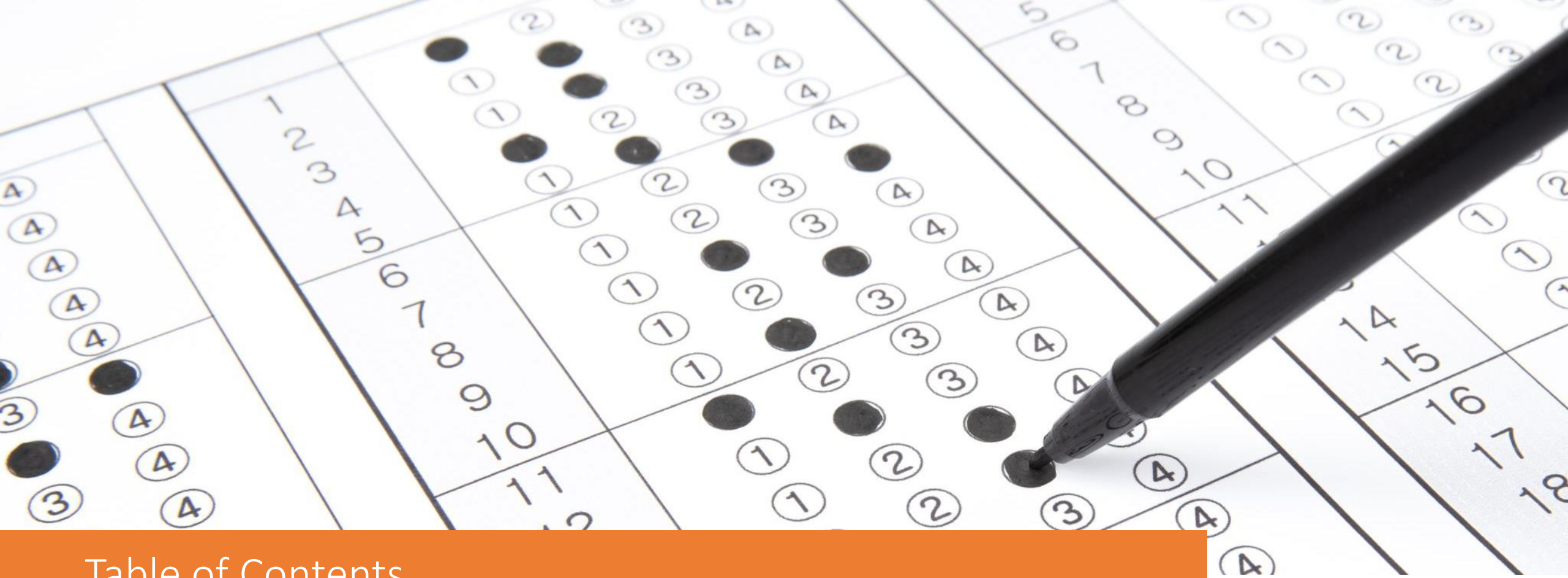
# Applied Data Science Capstone

Dafni R.



# Table of Contents

# Executive Summary

- **Summary of methodologies:**
- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

- **Summary of all results:**
- Exploratory Data Analysis result
- Interactive analytics demo in screenshots
- Predictive analysis results

# Introduction

In this capstone we completed a lot of labs related to data analysis and visualization. In order to analyse the Falcon 9 data and determine if the Falcon 9 first stage would successfully land, we gathered data using a RESTful API and web scraping.

The data was also transformed into a dataframe, and after that, some data wrangling was done. Later, using Plotly Dash and Folium, we created a dashboard to interactively evaluate launch records and a map to assess the closeness of launch sites.

Lastly, by comparing the performance of Hyperparameter for SVM, Classifier Trees, and Logistic Regression, we utilised machine learning to predict whether the first stage of the Falcon 9 will safely land.


# Data collection and data wrangling methodology

We used web scraping in these experiments to get historical Falcon 9 launch data from a Wikipedia article titled "List of Falcon 9 and Falcon Heavy launches." Also, we used exploratory data analysis (EDA) to identify trends in the data and choose the label for supervised model training.

# EDA and interactive visual analytics methodology

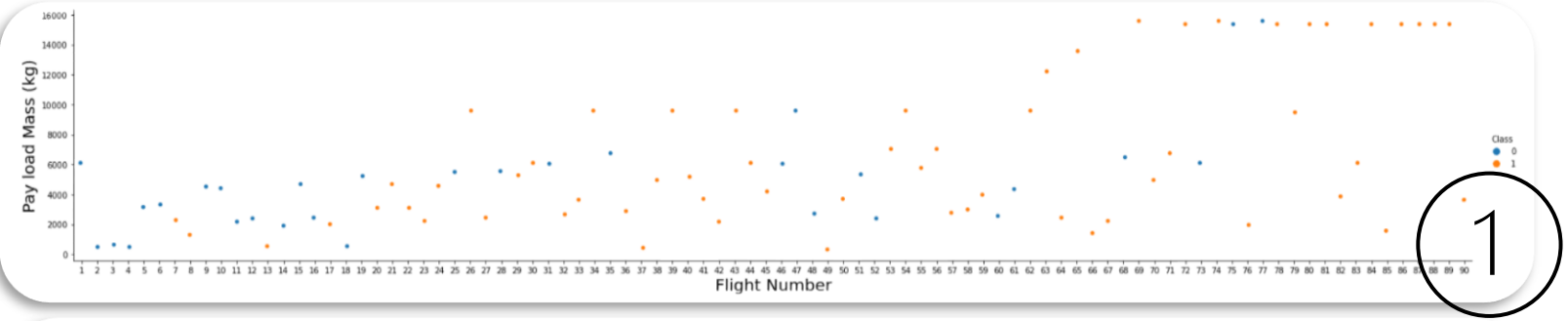
These exercises got us familiar with the Spacex DataSet, put the dataset into the appropriate table in a Db2 database, and ran SQL queries to respond to assignment questions.

We eventually predicted if the Falcon 9’s first will land successfully.

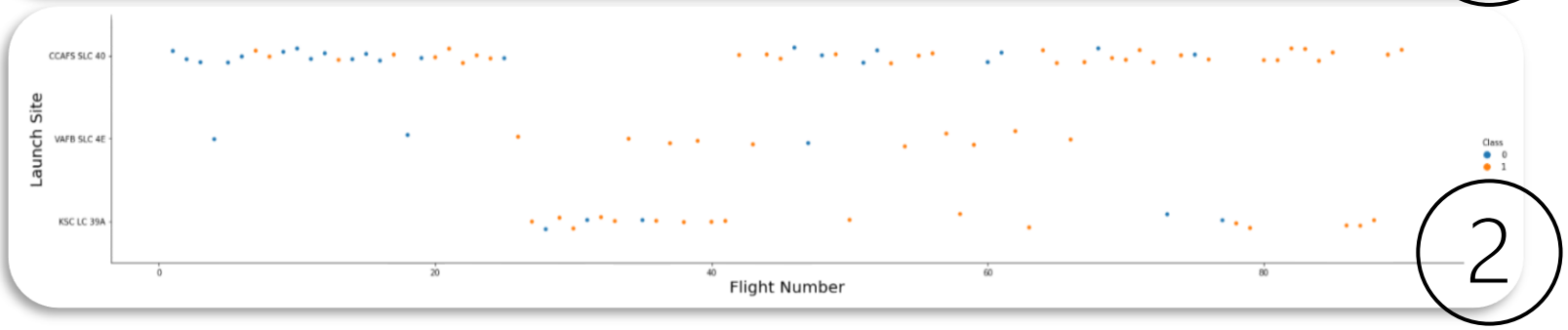
# Predictive analysis methodology

On its website, Space X promotes Falcon 9 rocket launches for 62 million dollars; other suppliers charge upwards of 165 million dollars for each launch. A large portion of the savings is due to Space X's ability to reuse the first stage. Hence, if we can figure out if the first stage will land, we can figure out how much a launch will cost. If another business wishes to submit a proposal for a rocket launch against space X, they can utilise this information. Using the data from the previous labs, you will build a machine learning pipeline in this experiment to forecast whether the first stage will land.

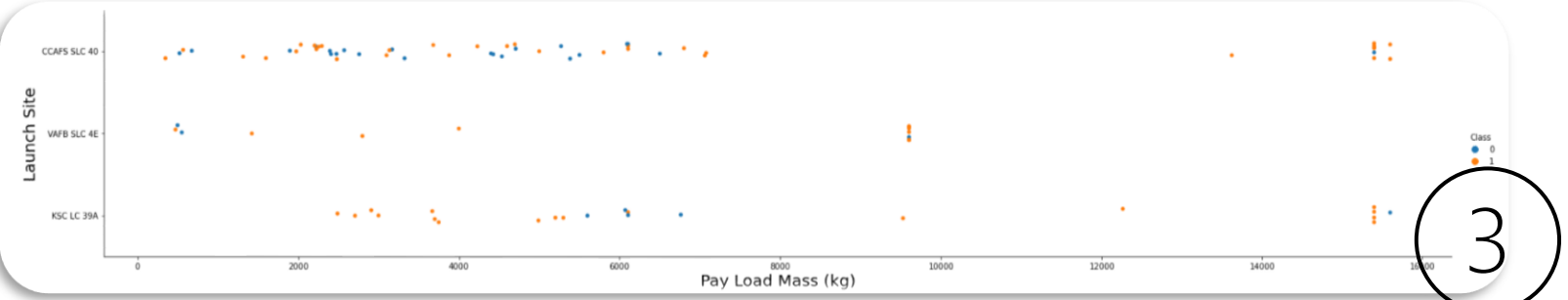
# EDA with visualization results




1

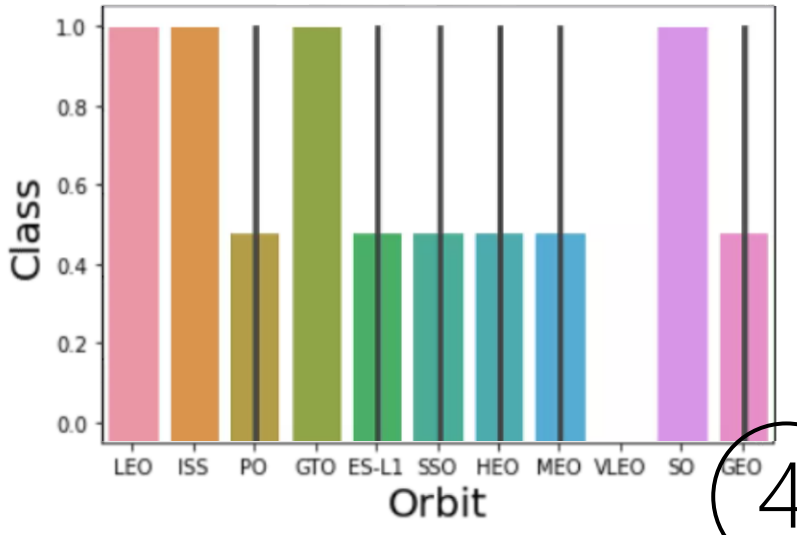



2




3

Graphs for Pay Load Mass vs Flight Number (1), Launch Site vs Flight Number (2) and Launch Site vs Pay Load Mass (3).
Finally Class vs Orbit (4).




4

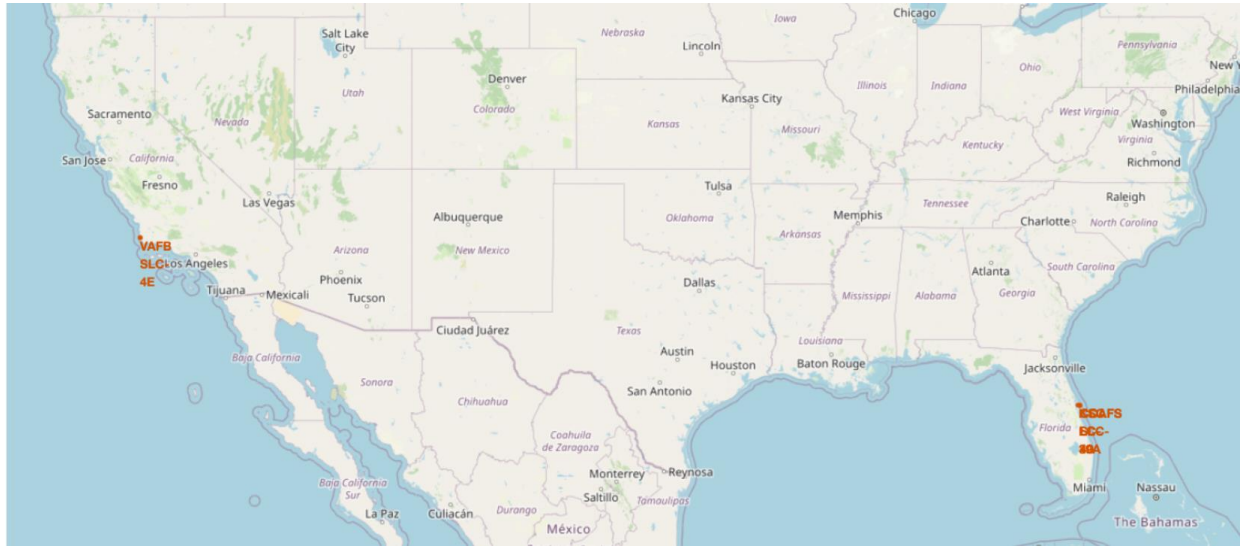
# EDA with SQL results

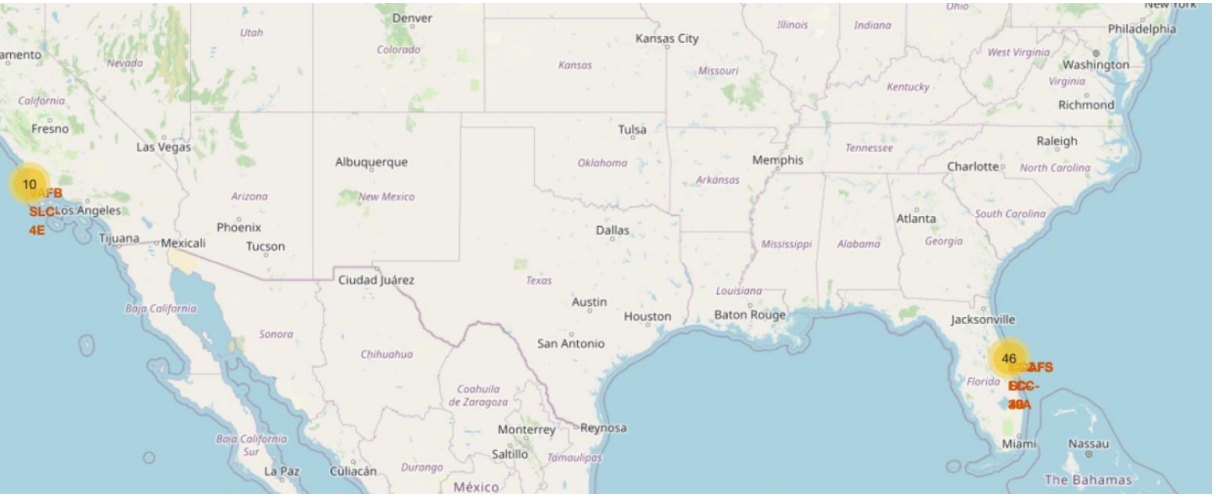
- Displaying 5 records where launch sites begin with the string 'CCA'.

Out[5]:

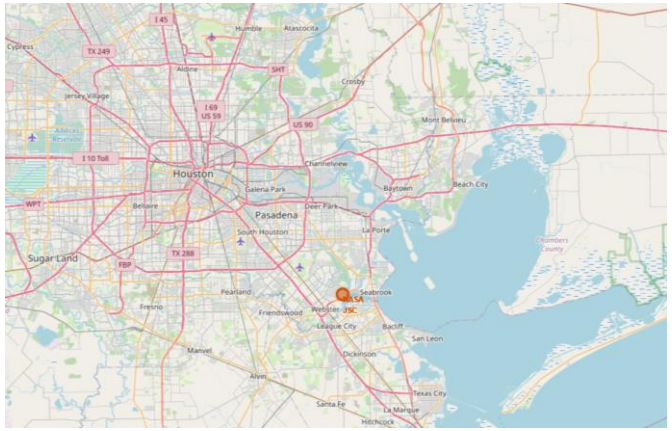
| DATE | time_utc_ | booster_version | launch_site | payload | payload_mass_kg_ | orbit | customer | mission_outcome | landing_outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | 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 |

# Interactive map with Folium results



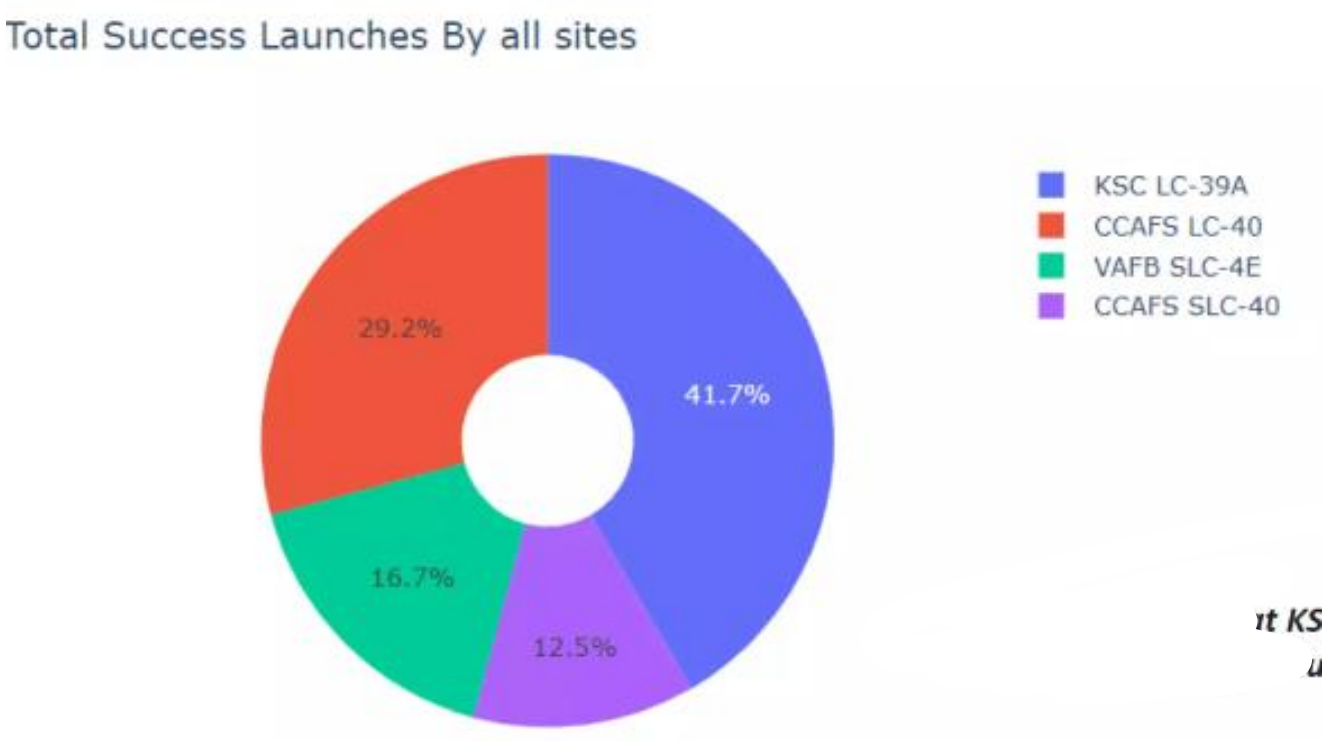

Some of the results of the interactive map are shown here.



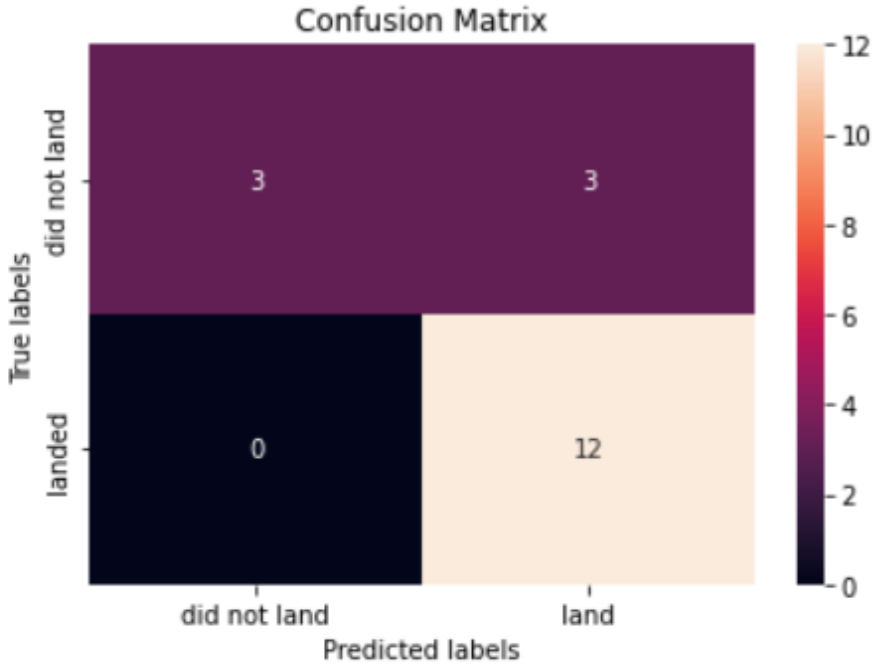

# Dashboard with Plotly Dash

Added:

- a drop-down menu to allow for Launch Site selection.
- a pie chart to display the number of successful launches across all sites and the number of successful versus unsuccessful launches for the site, if a specific Launch Site was chosen,
- a slider to select Payload range
- and a scatter chart to display the relationship between Payload and Launch Success.



# Predictive analysis (classification) results




We can see that logistic regression can discriminate between the various groups by looking at the confusion matrix. We can observe that false positives are the main issue.

| | LogReg | SVM | Tree | KNN |
|---|---|---|---|---|
| **Jaccard_Score** | 0.800000 | 0.800000 | 0.800000 | 0.800000 |
| **F1_Score** | 0.888889 | 0.888889 | 0.888889 | 0.888889 |
| **Accuracy** | 0.833333 | 0.833333 | 0.833333 | 0.833333 |

# Conclusion

The most effective method for this dataset is the decision tree model. Lower payload mass rockets perform better than higher payload mass launches. The majority of launch locations are near to the equator, and every site is within a short distance of the shore. With time, launches have a higher success percentage. Among all the launch locations, KSC LC-39A has the best success rate. The success rate for the ES-L1, GEO, HEO, and SSO orbits is 100%.