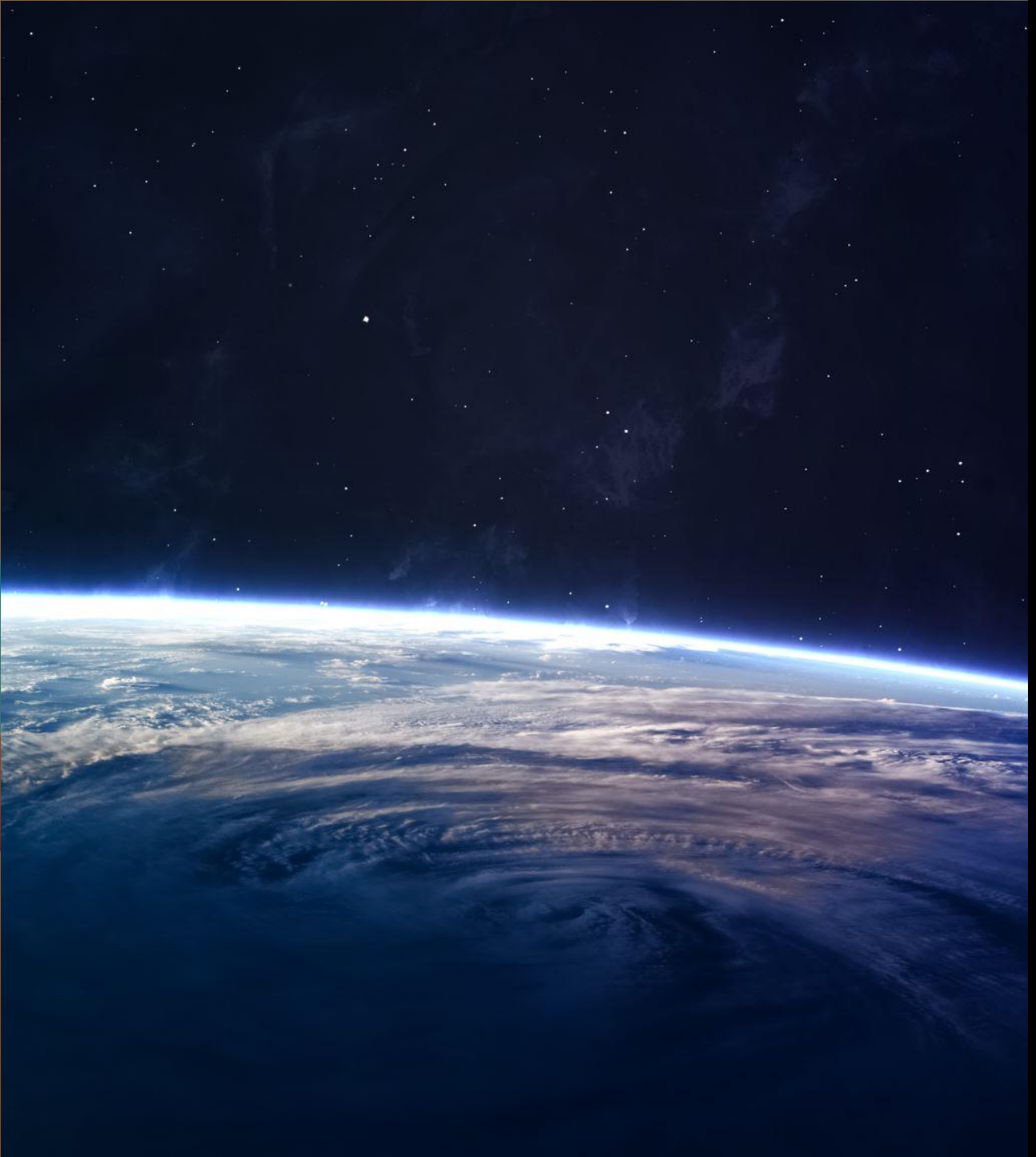
# SpaceX Falcon-9 First Stage Landing Prediction

Presented by Carmen Giambra



# Outline

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

# Executive Summary

- The objective of this project is to predict if the Falcon 9 first stage will land successfully. Space X is able to provide lower cost rocket launches due to their ability to reuse the first stage rocket if it landed properly. By predicting whether or not the first stage will land successfully, we are able to determine the cost of the launch.
- Techniques such as web scrapping and data wrangling were used to collect and organize the necessary data. EDA by Visualization was then utilized to analyze patterns and trends. Lastly, several machine learning models were built to determine if the first stage rocket launches would be successful through predictive modeling.

# Introduction

SpaceX advertises Falcon 9 rocket launches with a cost of $62M while competitors cost upwards of $165M.

Space Y is a competitor founded by Allon Musk that would like to compete with SpaceX.

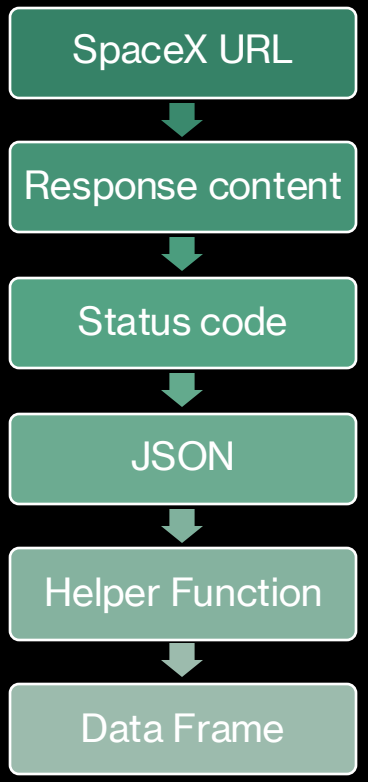
Information available about mission parameters such as payload, orbit and customer can be used to predict the success of the launch.

# Methodology

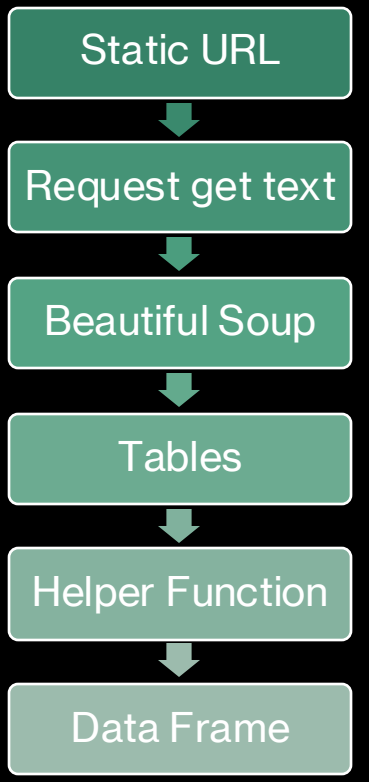
- Data collection through SpaceX API and web scrapping
- Data wrangling by using pandas and cleaning missing values
- Perform exploratory data analysis (EDA) using SQL and Visualization
- Perform interactive visual analytics using Folium and Plotly Dash
- Build classification models to execute predictive analysis

# Data Collection

## Via SpaceX APIs




## Via Web Scrapping



# Data Wrangling

Data collected from the SpaceX APIs and web scrapping were cleaned and organized under the following steps:

1. Data imported and percentage of missing values found
2. Verified the data type of each column (int64, object, float64, and bool)
3. Value counts for various launch sites analyzed
4. Created new "class" feature to assign value "0" for failed launch outcome and "1" for successful launch outcome
5. Calculated the percentage of successful outcomes

# EDA through SQL

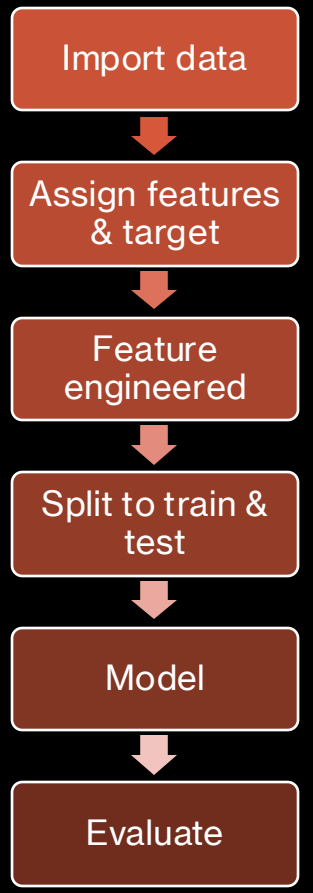
- Used SQL to list the total number of successful and failed mission outcomes
- Analyzed which booster versions carried the maximum payload value
- Sorted and listed launch data by month, location, booster version, etc.
- Ranked the count of launding outcomes per specific date ranges

# EDA through Visualization

- Used Visualization to compare relationships between success rate and orbit type
- Visualized relationships between launch sites, payload, flight number, etc.
- Visualized the launch success yearly trend

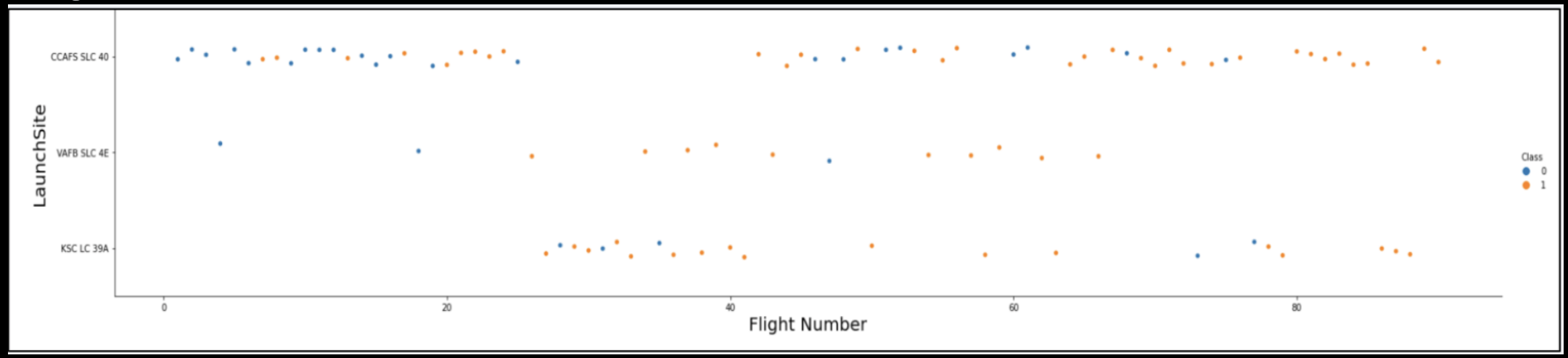
# Classification with Predictive Analysis

- Cleaned data was imported then assgined variable X for features and variable Y for the target. Features were scaled and split with 20% assigned to testing.
- Hyperparameter "GridSearchCV" created to optimize the best performing parameters.
- The training set was fitted while the test set was evaluated. A decision tree classifier model proved to be the most successful with a success rate of 88.88%.

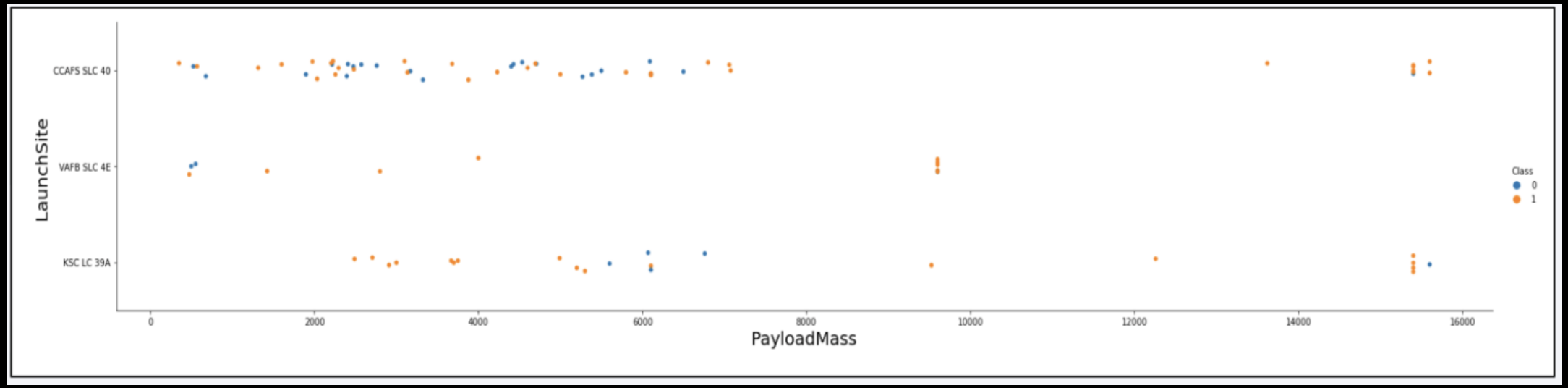


# EDA Results with Visualization

## Flight Number vs. Launch Site

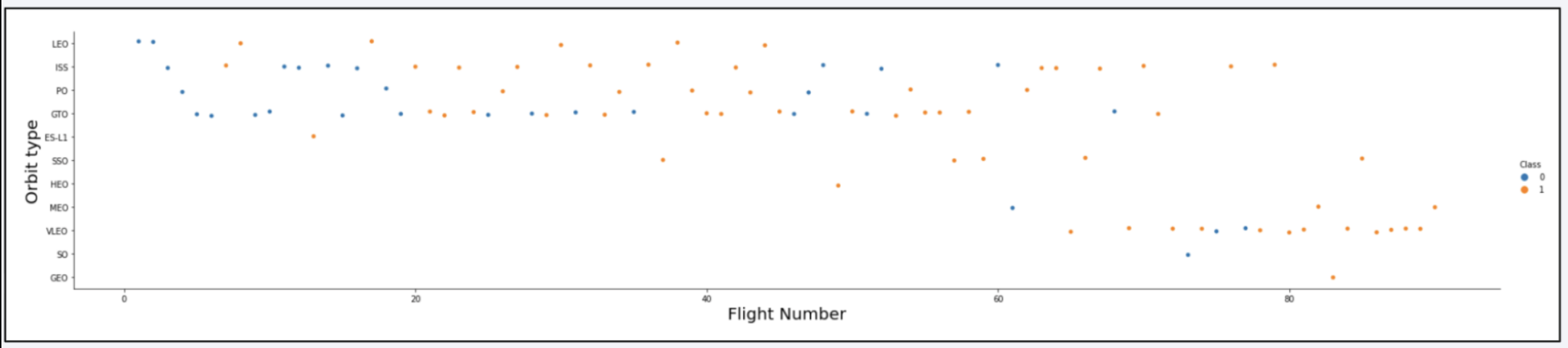



## Payload vs. Launch Site

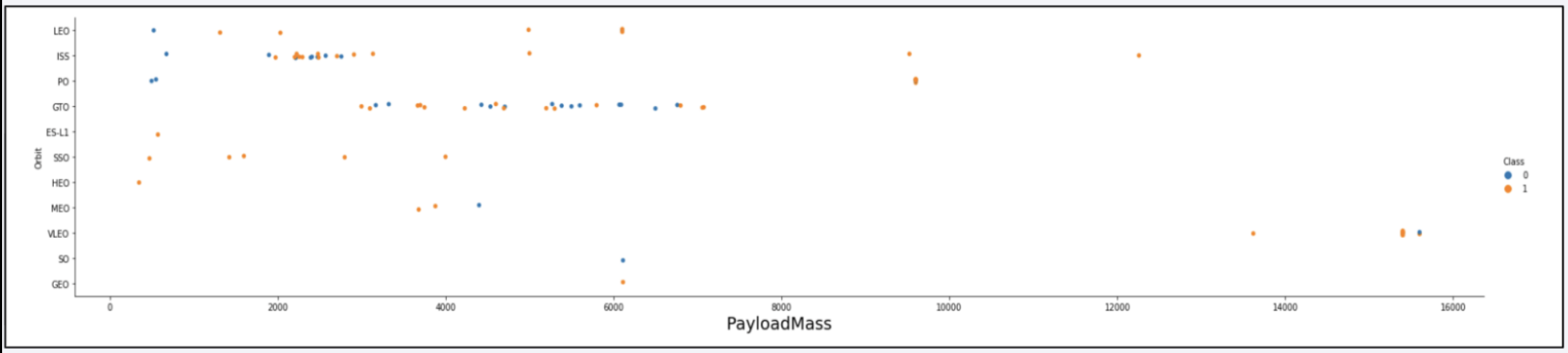


# EDA Results with Visualization

## Flight Number vs. Orbit Type

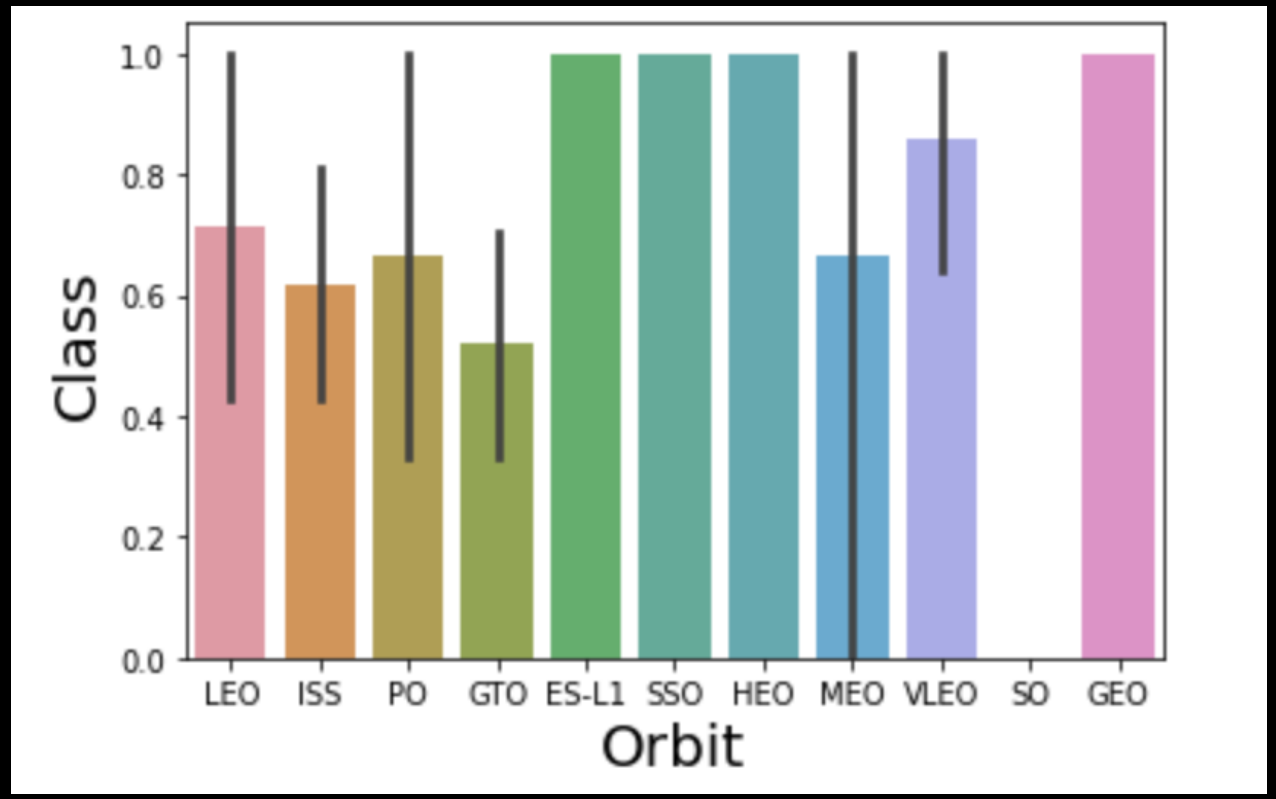



## Payload vs. Orbit Type



# EDA Results with Visualization

Success Rate vs. Orbit Type

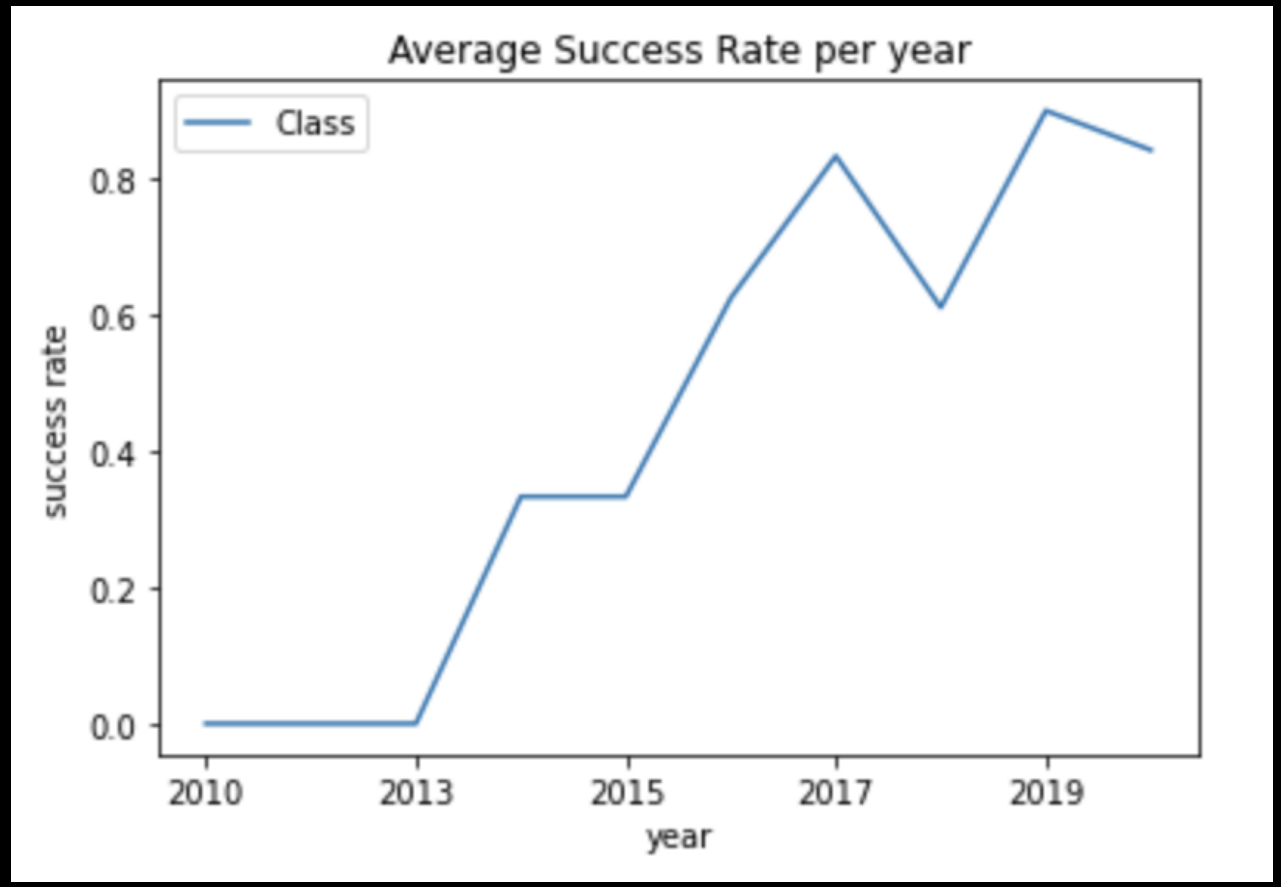



Findings:

Orbit types ES-L1, SSO, HEO and GEO have a success rate of 100%

# EDA Results with Visualization

Launch Success Yearly Trend




Findings:

Success rate continued increasing from 2013 until 2020.

# EDA Results with SQL

Unique Launch Sites:

- CCAFS LC-40 – Cape Canaveral, Florida
- VAFB SLC-4E – Vandenberg Space Force Base, California
- KSC LC-39A – Kennedy Space Launch Center, Florida

First successful ground pad landing outcome was on December 22nd, 2015:

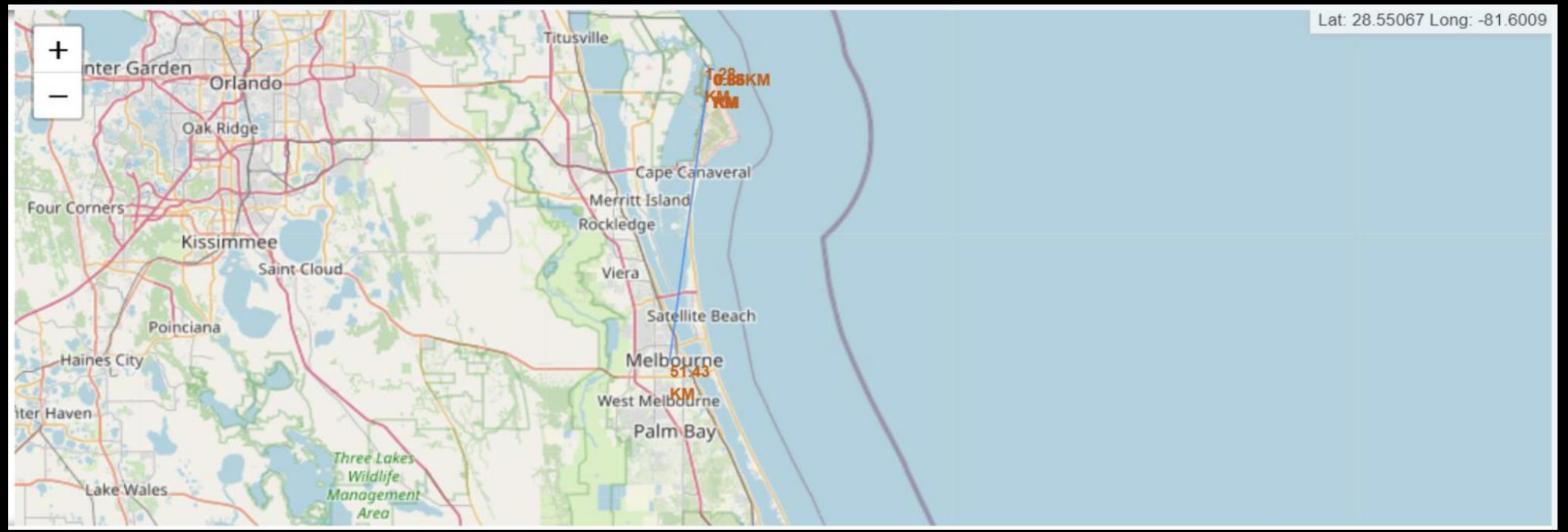
| Mission_Outcome | Date_First_Succ_Land |
|---|---|
| Success | 2015-12-22 |

Total number of successful and failed mission outcomes:

| Mission_Outcome | Total (Success or failure) |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

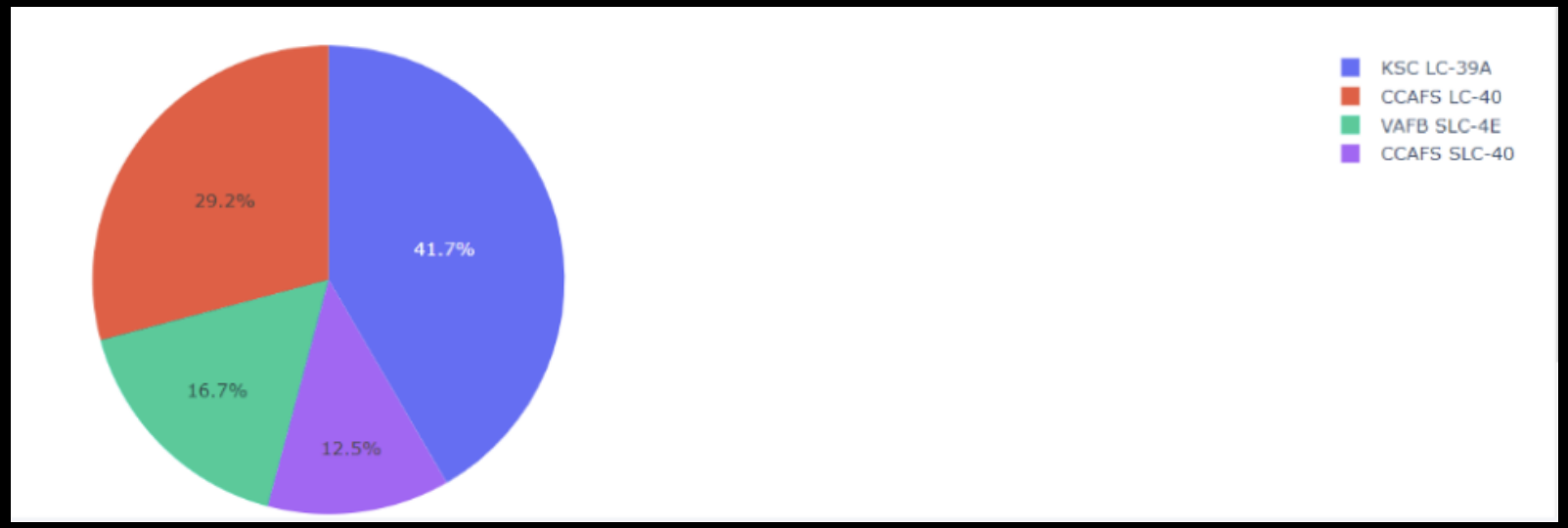
# Folium Map Proximities

Launch site distance proximity to railway, highway and coastline:

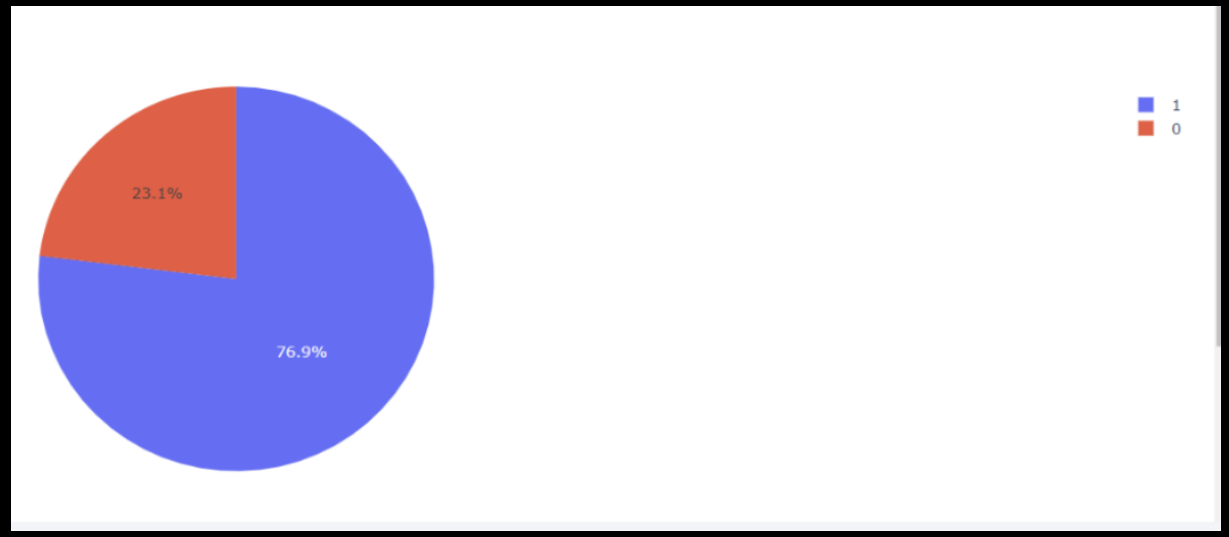


# Plotly Dashboard

Launch sites pie chart:

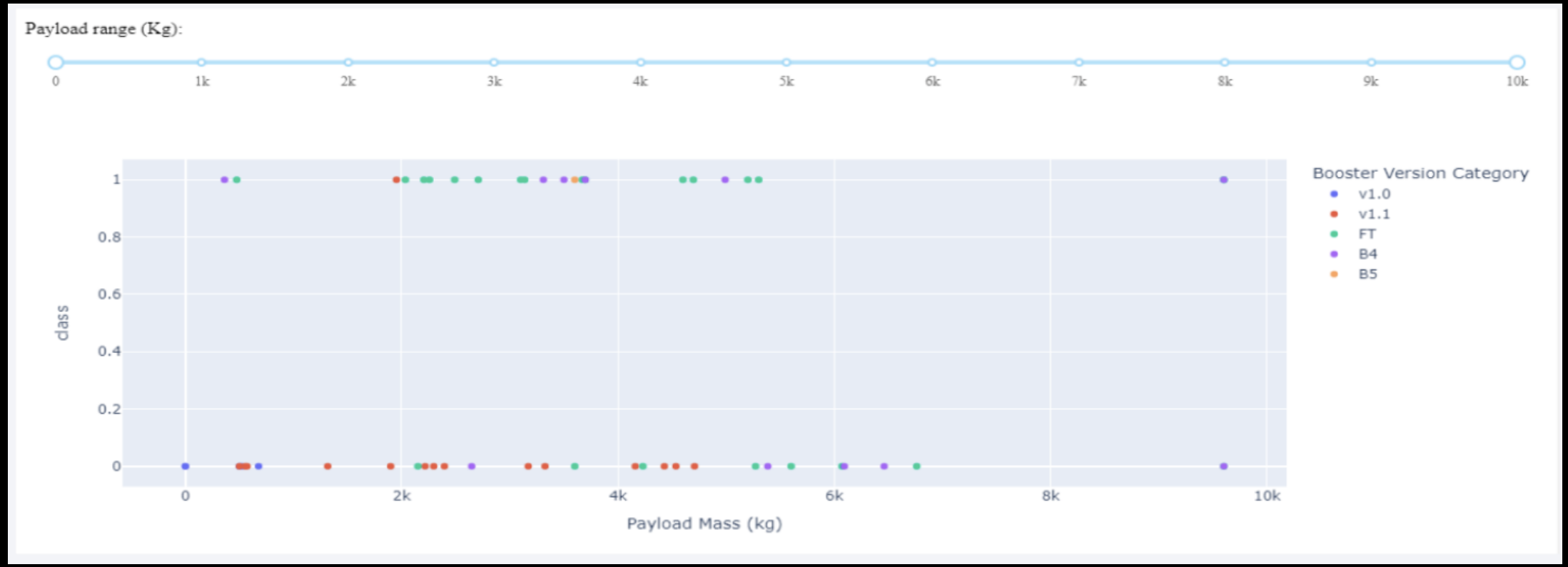



Launch sites KSC LC-39A success rate:



# Plotly Dashboard

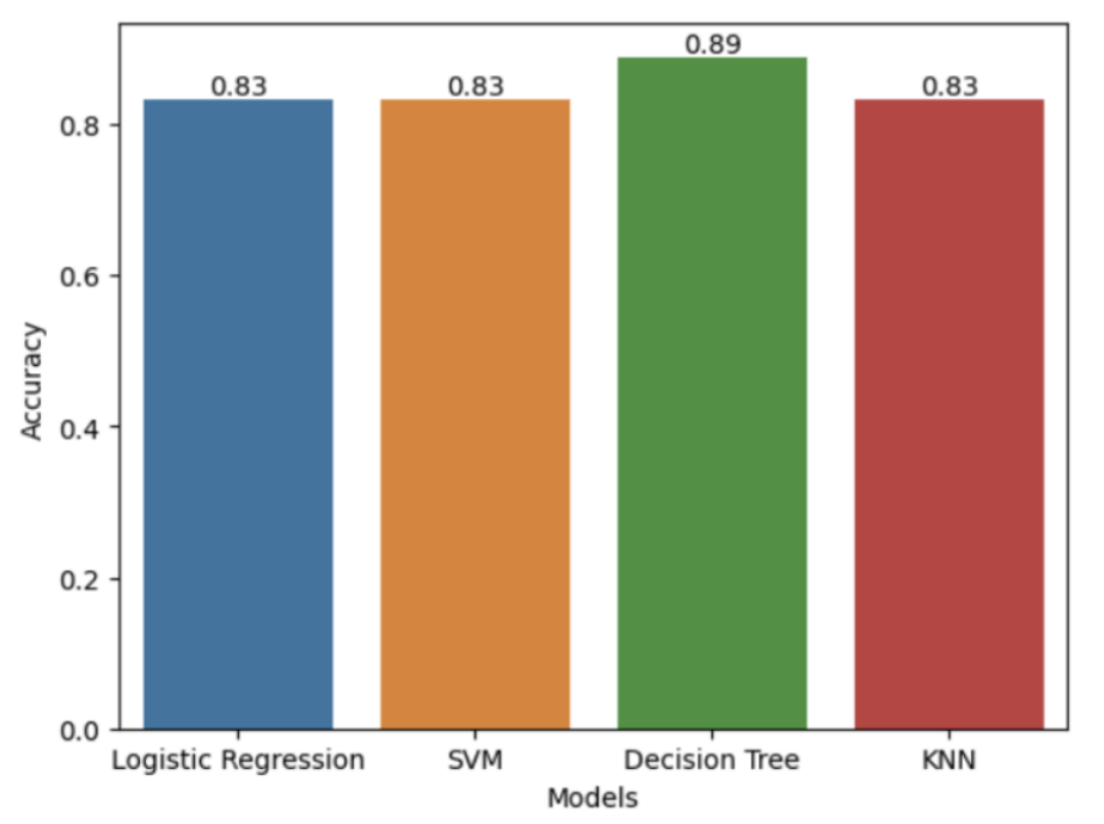
Payload vs. Launch Outcome:




Findings: As payload mass grows so does the success rate for booster version FT, while the opposite is true for booster version v1.1.

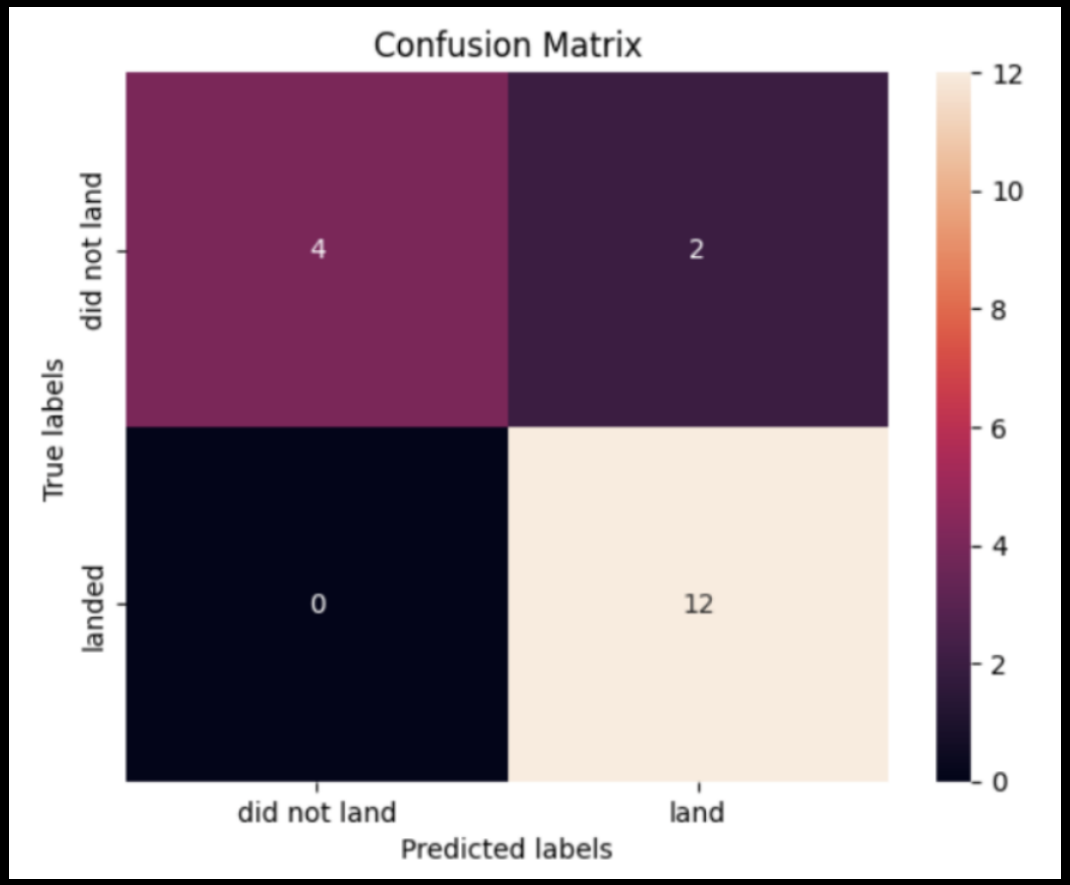
# Classification Accuracy

Classfication models accuracies:




Findings: The decision tree classification model was the most accurate with an 88.88% accuracy.

# Confusion Matrix




Findings: The decision tree classification model also had 2 false positives and no false negative with a precision accuracy of 85.71%

# Conclusion

Orbit types ES-L1, SSO, HEO, and GEO have a 100% landing success rate.

The Falcon 9 landing success rate has been increasing since 2010 until 2020.

The decision tree classification model can predict landing success rate with an accuracy of 88.88%

# Appendix

Python code notebooks, SQL queries, charts, and data sets used in presentation can be found on my GitHub.