IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Juan Ignacio Padin
07/25/2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

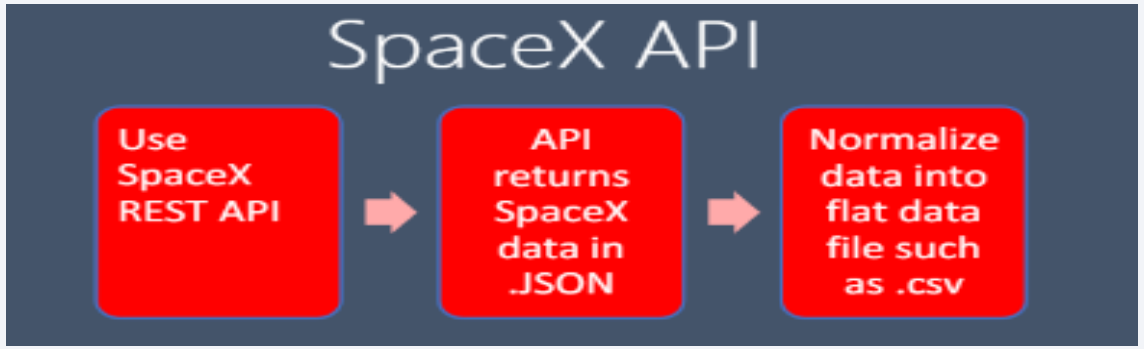
- Data collection methodology:
  - Webscrapping
  - Space X API
- Perform data wrangling
  - Transforming data in order to apply to ML
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Web Scraping
- APIs.
- Request library

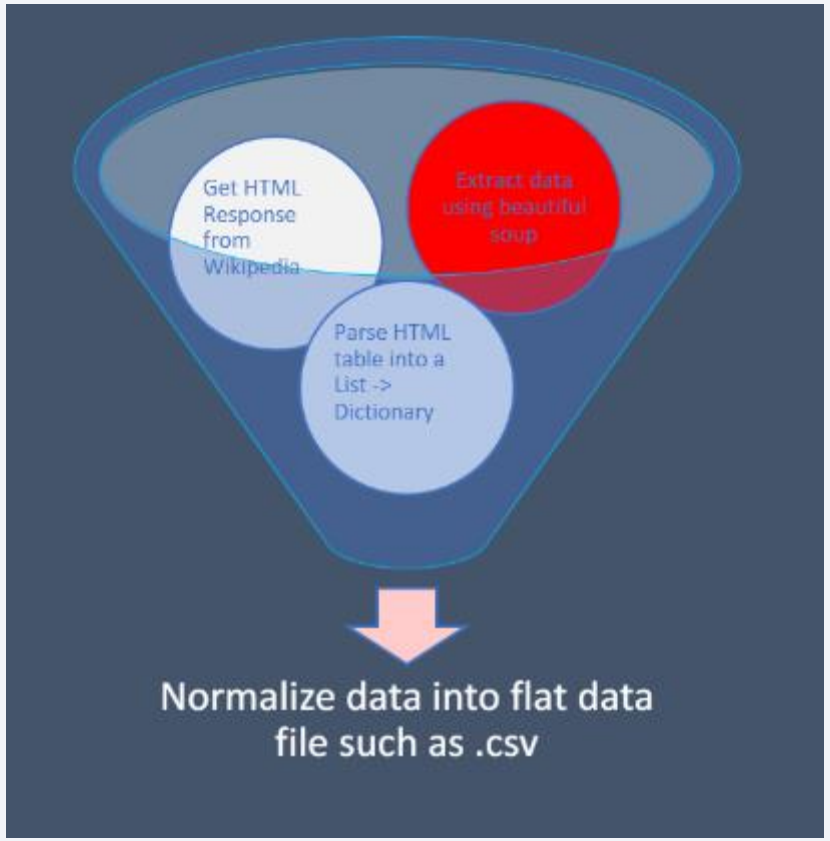
Data Collection – SpaceX API

- https://github.com/juanipadin/DataScience/tree/main/SpaceX



Data Collection - Scraping

- https://github.com/juanipadin/DataScience/blob/main/SpaceX/labs-jupyter-spacex-Data%20wrangling.ipynb



Data Wrangling

- Within the dataset, there are several instances where the booster did not achieve a successful landing. These unsuccessful landings occurred for various reasons. For instance, in cases marked as "True Ocean," the mission outcome indicates a successful landing in a specific region of the ocean, while "False Ocean" represents an unsuccessful landing in that region. Similarly, "True RTLS" signifies a successful landing on a ground pad, whereas "False RTLS" indicates an unsuccessful landing on a ground pad. Additionally, "True ASDS" denotes a successful landing on a drone ship, while "False ASDS" indicates an unsuccessful landing on a drone ship.
- To simplify the outcomes for training purposes, we convert them into Training Labels, where a value of 1 denotes a successful booster landing, and a value of 0 represents an unsuccessful landing.

EDA with Data Visualization

For example of some questions we were asked about the data we needed information about. Which we are using SQL queries to get the answers in the dataset :

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC’

EDA with SQL

For example of some questions we were asked about the data we needed information about.
Which we are using SQL queries to get the answers in the dataset :

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC’

Build an Interactive Map with Folium

To visualize the Launch Data into an interactive map. We took the Latitude and Longitude Coordinates at each launch site and added a Circle Marker around each launch site with a label of the name of the launch site.

We assigned the dataframe launch_outcomes(failures, successes) to classes 0 and 1 with Green and Red markers on the map in a MarkerCluster()

Using Haversine's formula we calculated the distance from the Launch Site to various landmarks to find various trends about what is around the Launch Site to measure patterns. Lines are drawn on the map to measure distance to landmarks

Predictive Analysis (Classification)

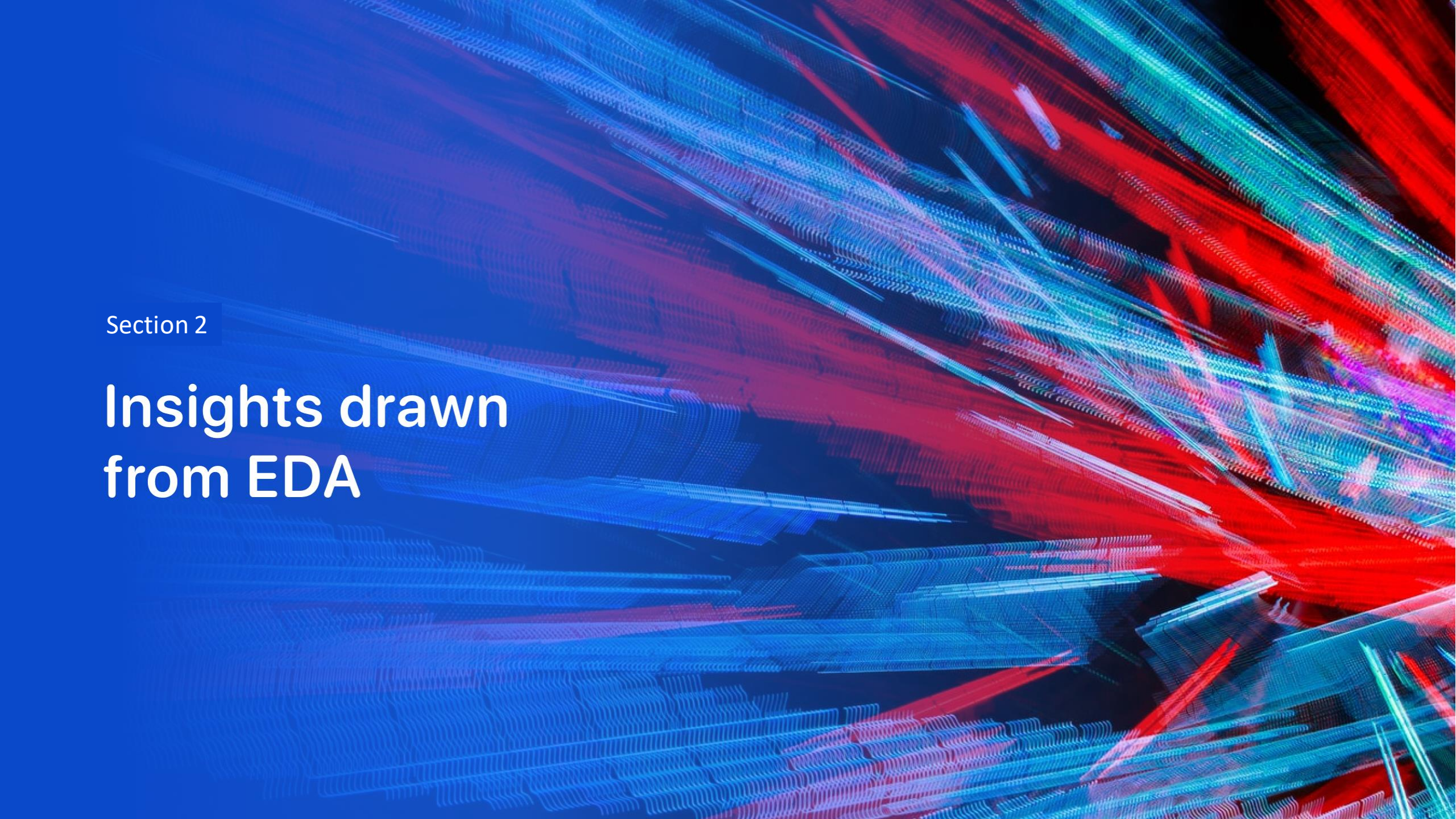
- Load our dataset into NumPy and Pandas
- Transform Data
- Split our data into training and test data sets
- Check how many test samples we have
- Decide which type of machine learning algorithms we want to use
- Set our parameters and algorithms to GridSearchCV

Results

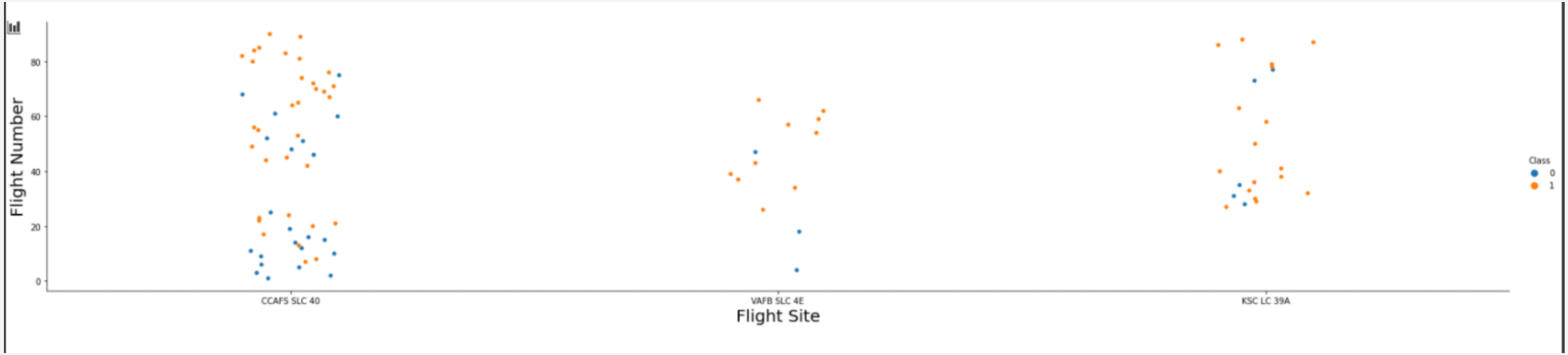
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

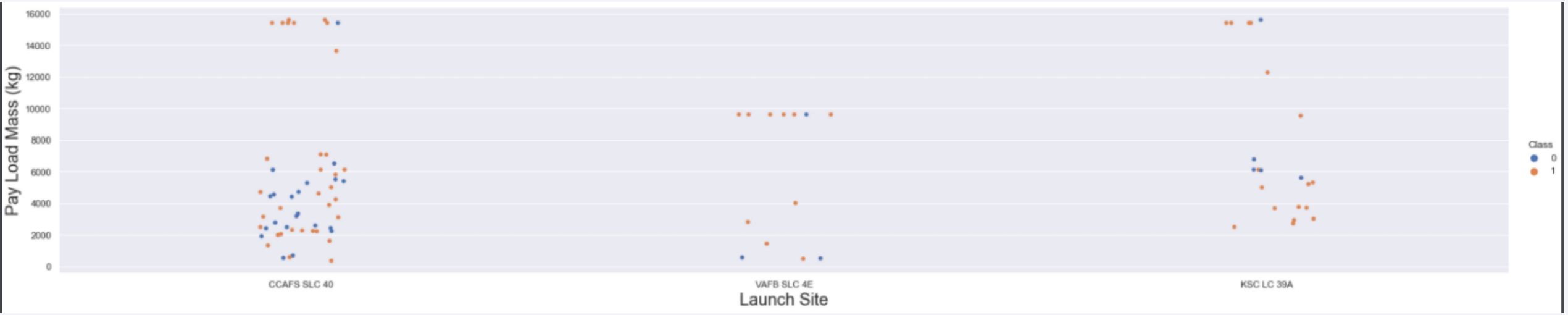
Insights drawn from EDA



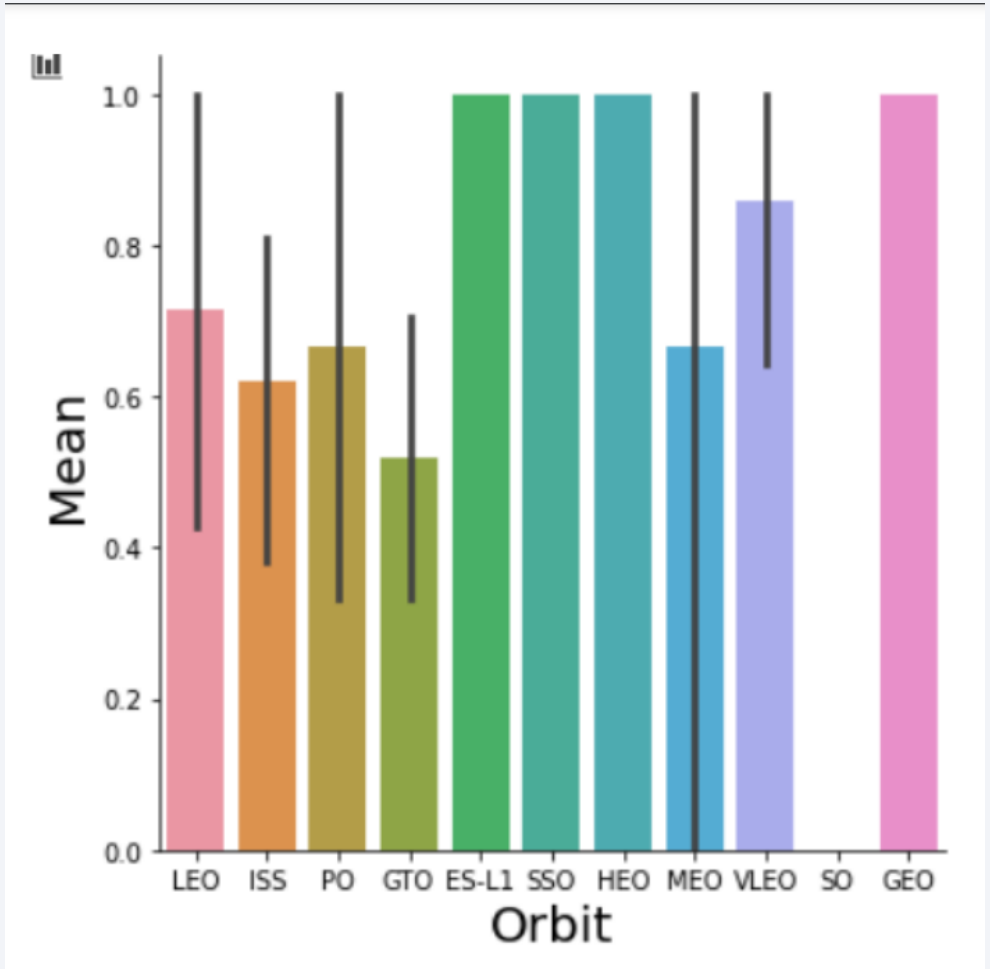
Flight Number vs. Launch Site



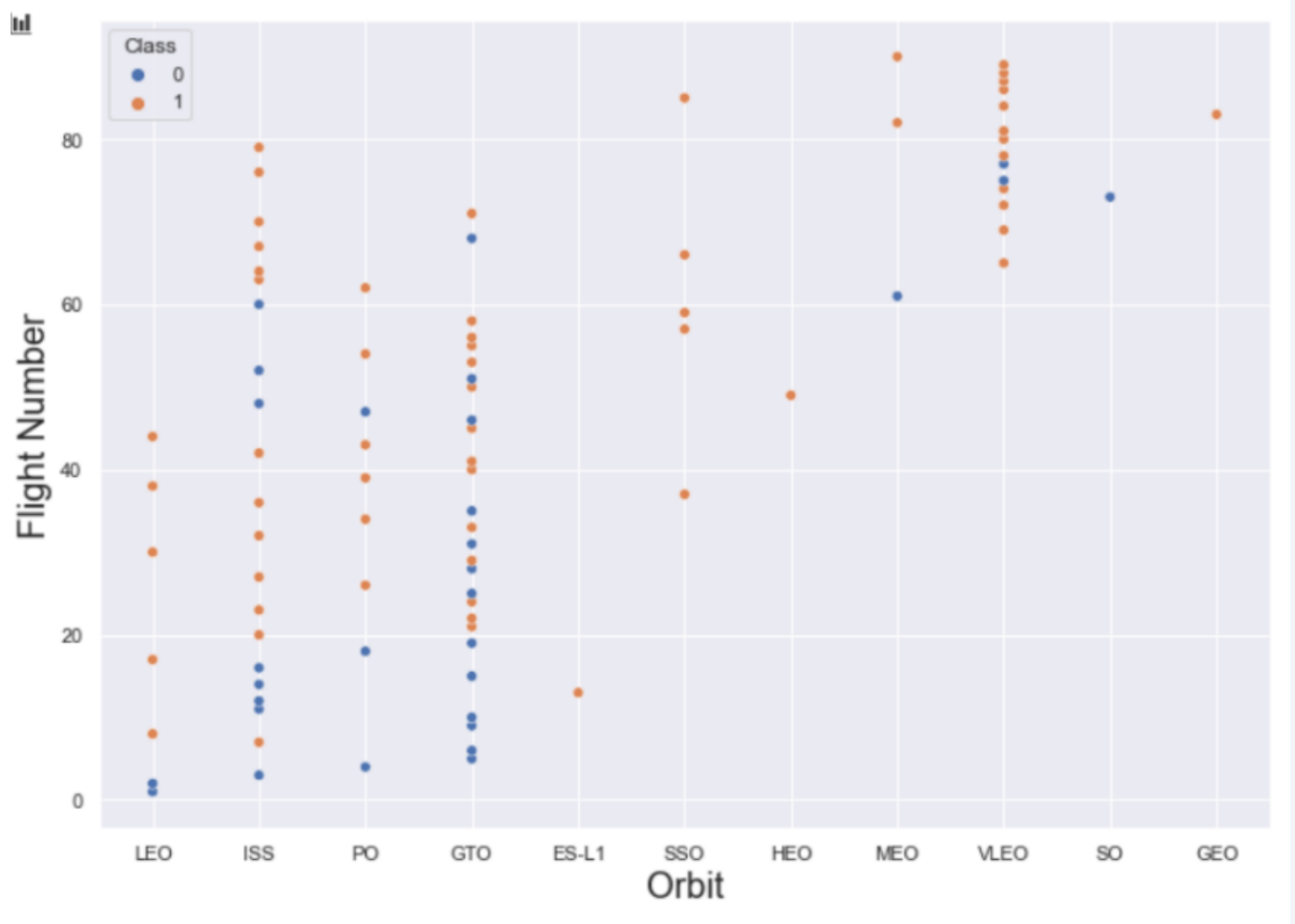
Payload vs. Launch Site



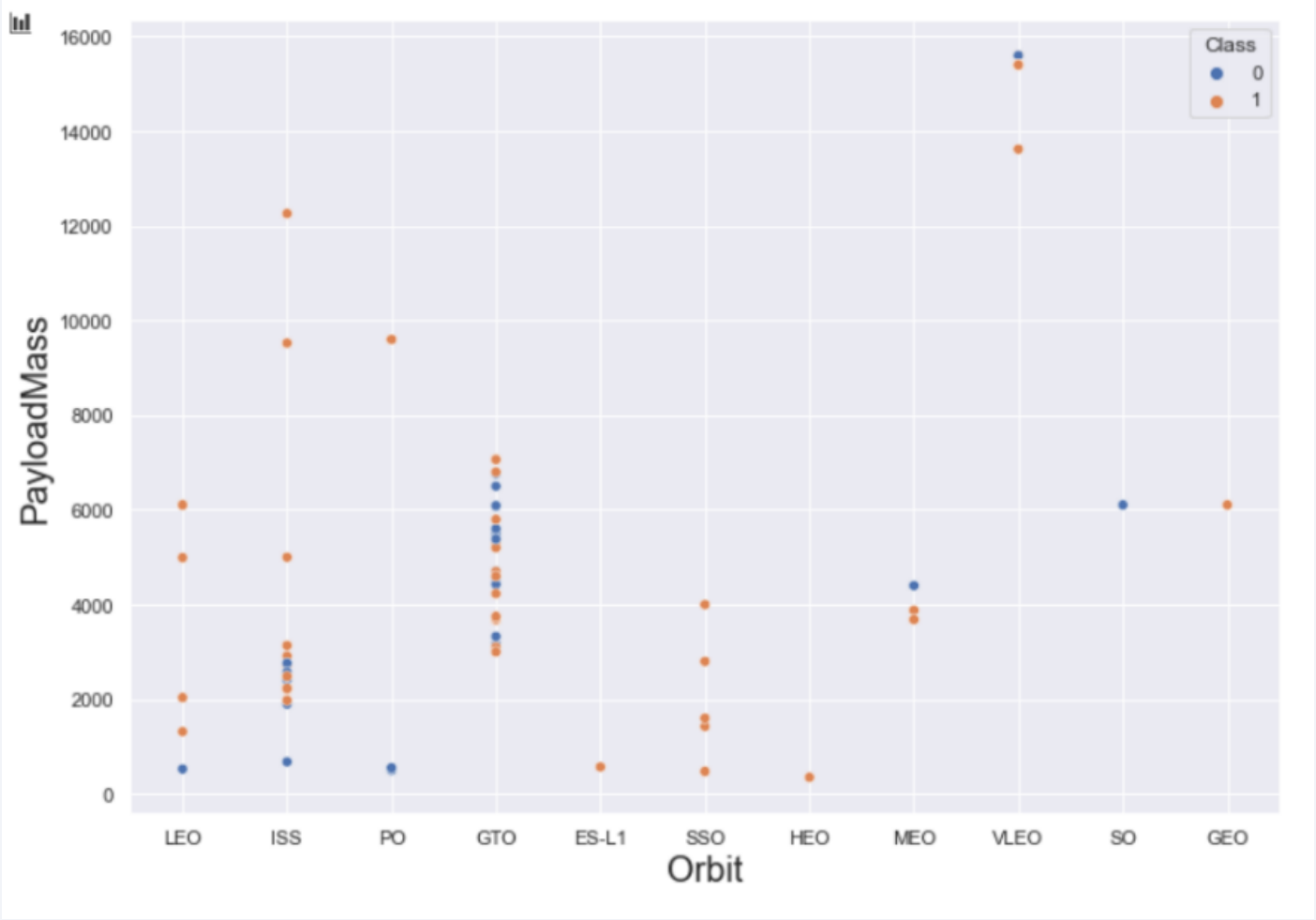
Success Rate vs. Orbit Type



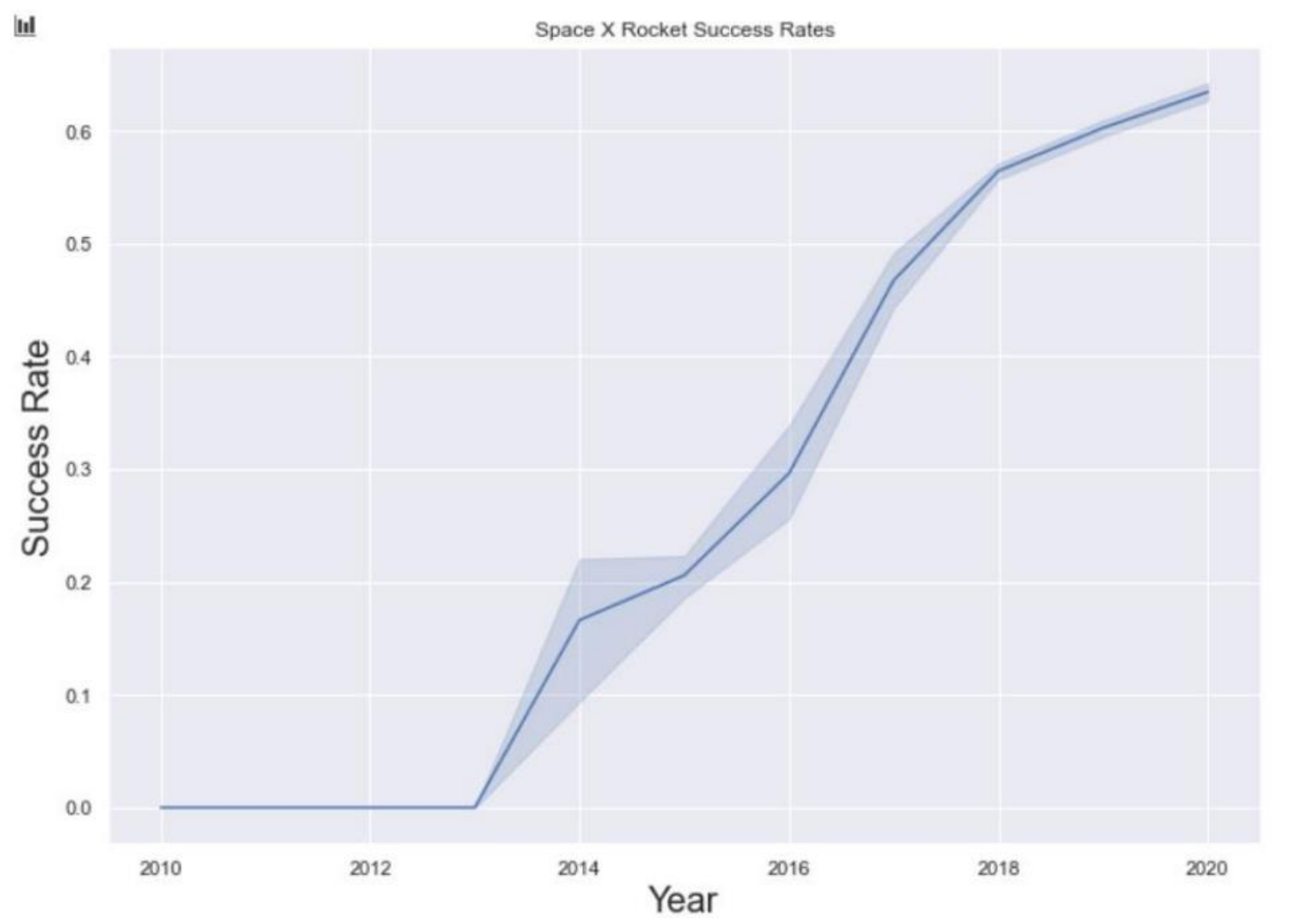
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend




Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

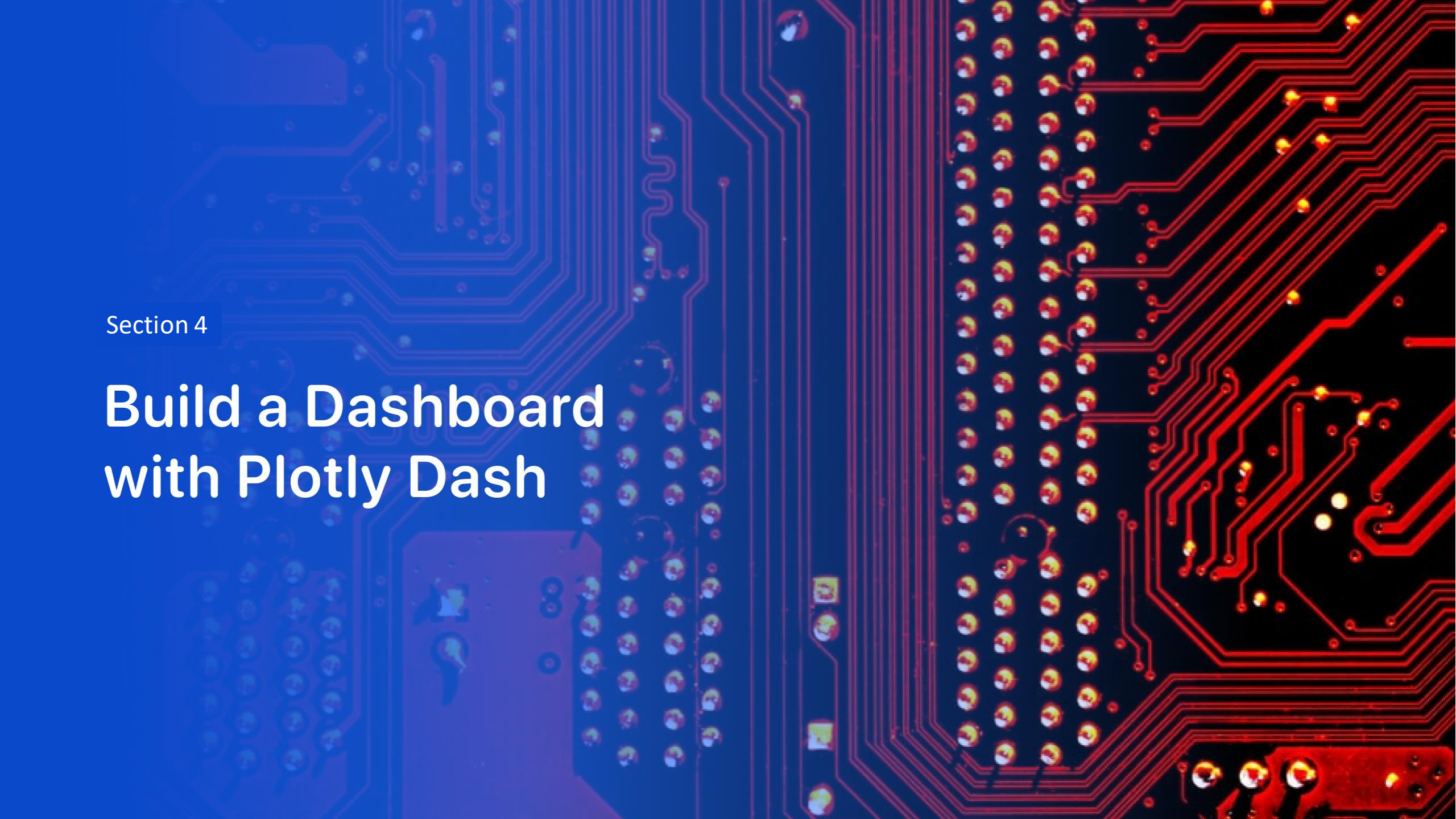
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

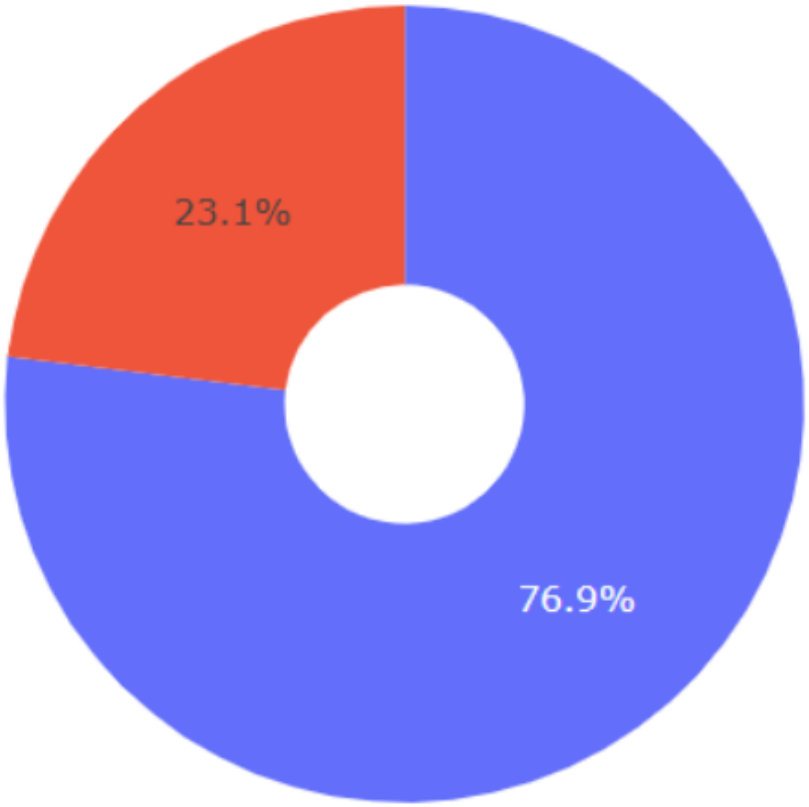
- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

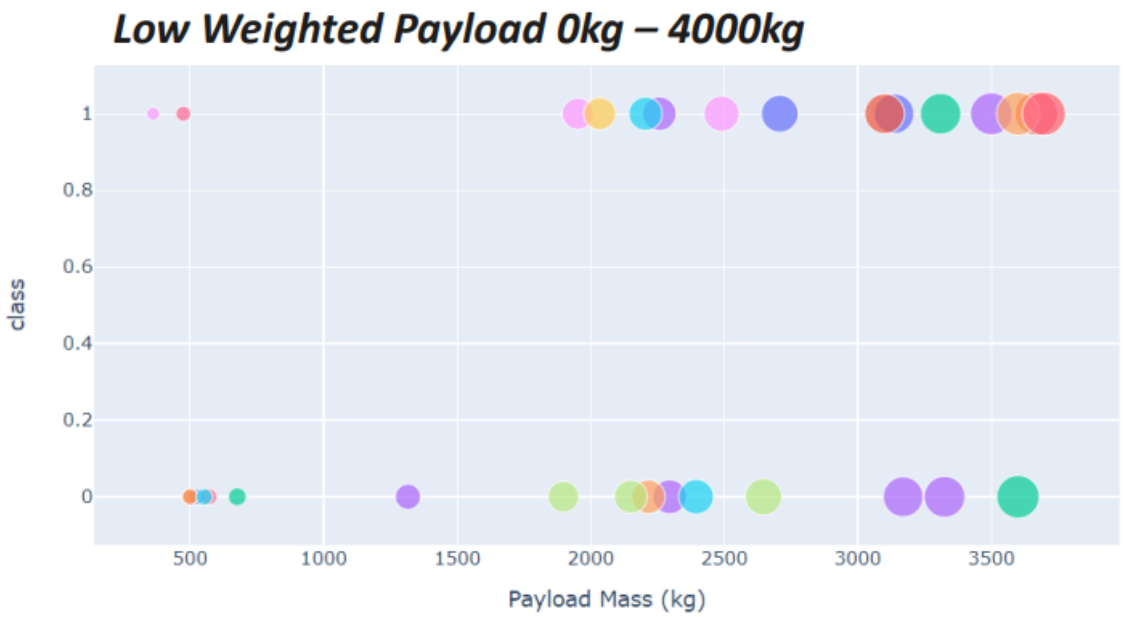
- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

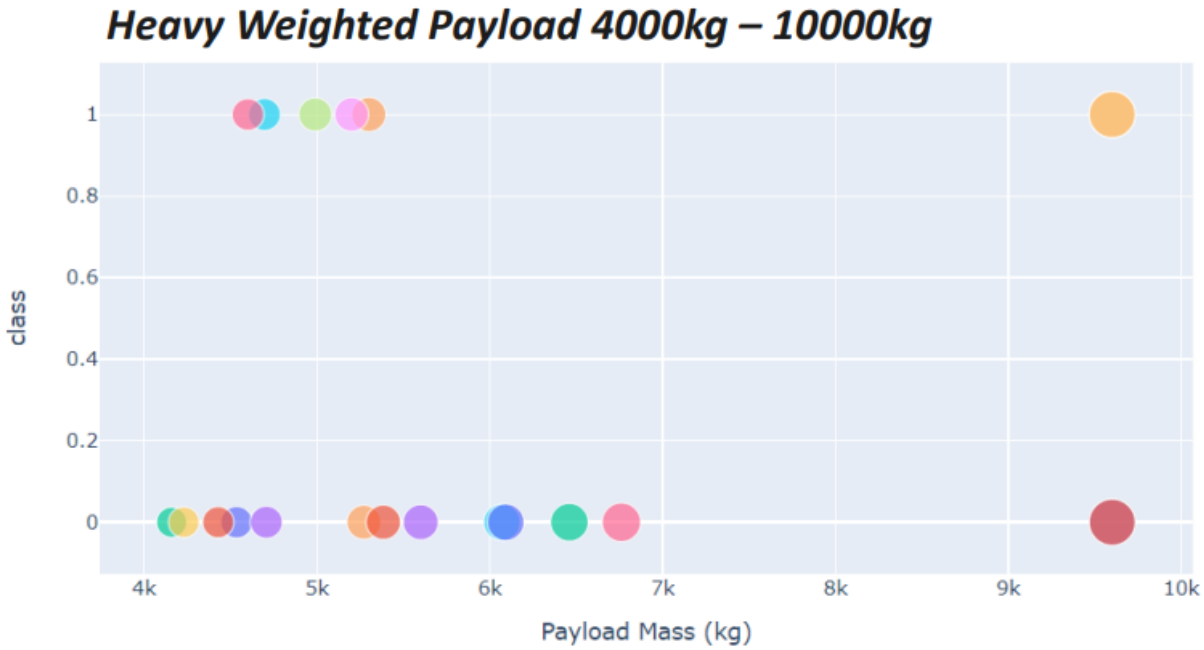

Section 4
Build a Dashboard with Plotly Dash

Launch site with highest launch success ratio



Payload vs. Launch Outcome scatter plot for all sites

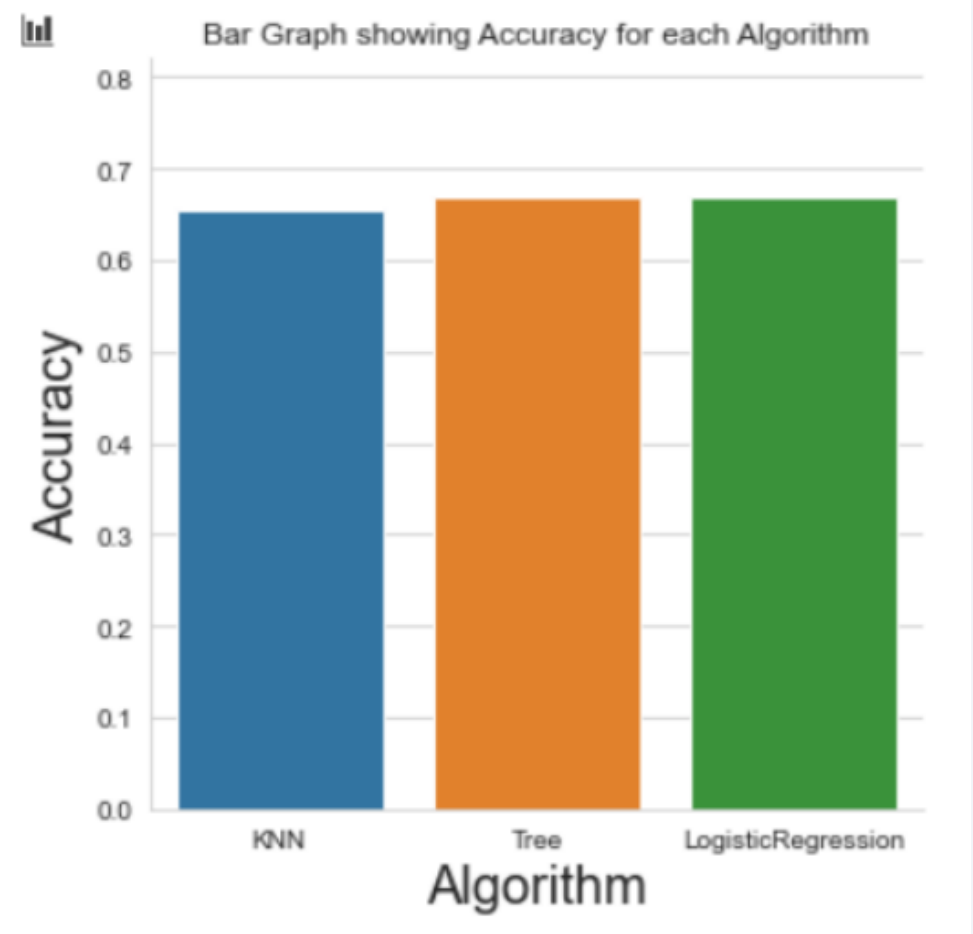




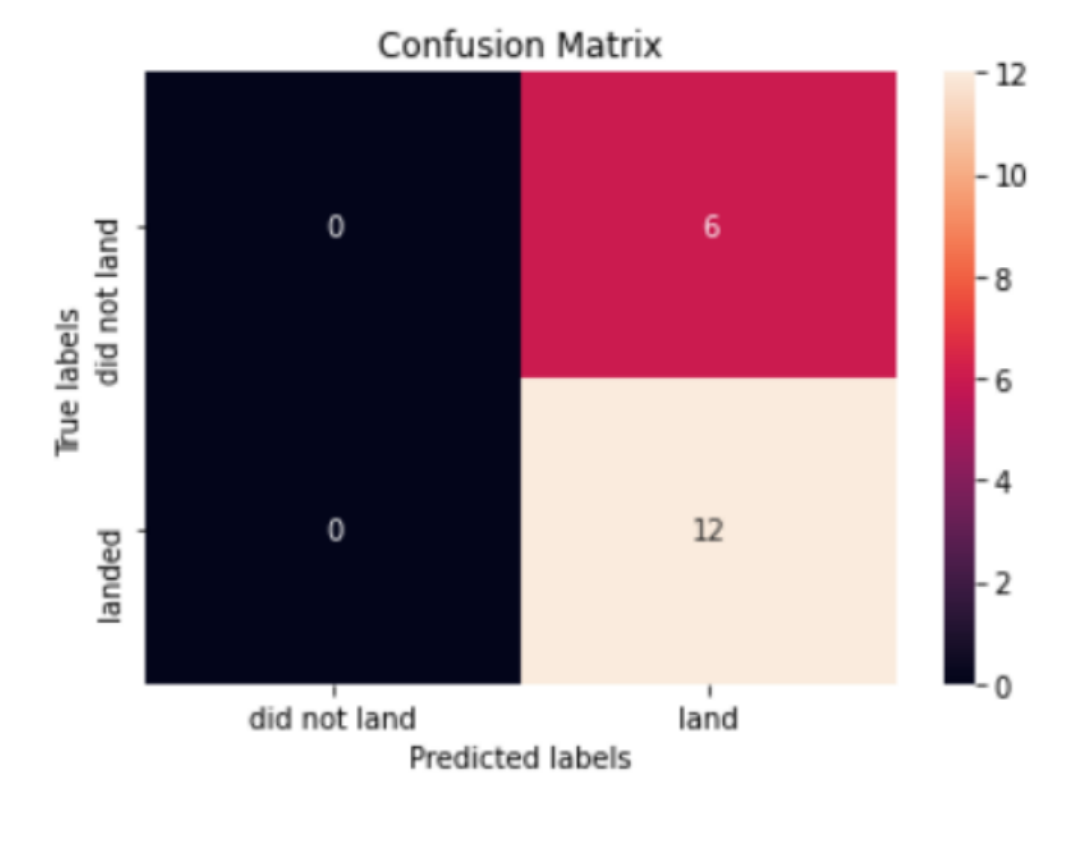
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

It appears that you have made some observations and conclusions about the dataset based on your analysis. However, it's important to note that the statements provided are subjective and should be carefully evaluated before drawing definitive conclusions. Let's rephrase the statements while maintaining their essence:

- The Tree Classifier Algorithm shows promising results for Machine Learning with this dataset.
- The performance of low-weighted payloads seems to outperform that of heavier payloads.
- There appears to be a positive correlation between the success rates of SpaceX launches and the time they spend perfecting their launches over the years.
- KSC LC-39A stands out as the launch site with the highest number of successful launches compared to other sites.
- Launches into specific orbits like GEO, HEO, SSO, and ES-L1 demonstrate the best success rates.
- Remember, the validity of these observations depends on the accuracy of the data and the thoroughness of the analysis. It's essential to conduct further investigations and possibly employ statistical tests to substantiate these claims.

Thank you!