IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ryan Rachmad Hidayat
1 February 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA) with Data Visualization
  - Building an Interactive Map with Folium
  - Building a Dashboards with Plotly Dash
  - Predictive Analysis
- Summary of all results
  - EDA Result
  - Interactive Analytics
  - Predictive Analytics

Introduction

- Project background and context
  - With businesses like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX lowering the cost of space travel, the commercial space industry is expanding.
  - Among SpaceX's many achievements are human missions, satellite internet, and spacecraft sent to the International Space Station (ISS).
  - In part because they recycle the first stage, SpaceX rockets are less expensive than those of their rivals.
- Problems you want to find answers
  - Find out if SpaceX plans to reuse the Falcon 9 rocket's first stage. Cost is affected by this.
  - Compile SpaceX-related data and produce dashboards for the group.
  - Using open data, train a machine learning model to determine whether SpaceX will reuse the first phase.
  - Ascertain the launch expenses for a new rocket business that intends to take on SpaceX.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - Dropping unnecessary columns
  - One Hot Encoding for classification 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
  - How to build, tune, evaluate classification models

Data Collection

- The dataset is taken from the following APIs:
  - From https://api.spacexdata.com/v4/rockets/ we would like to learn the booster name.
  - From https://api.spacexdata.com/v4/launchpads/ we would like to know the name of the launch site being used, the logitude, and the latitude.
  - From https://api.spacexdata.com/v4/payloads/ we would like to learn the mass of the payload and the orbit that it is going to.
  - From https://api.spacexdata.com/v4/cores/ we would like to learn the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, wheter the core is reused, wheter legs were used, the landing pad used, the block of the core which is a number used to seperate version of cores, the number of times this specific core has been reused, and the serial of the core.
  - Requesting rocket launch data from SpaceX API with the https://api.spacexdata.com/v4/launches/past
- Apart from that, data obtained through scraping from the following Wikipedia URL:
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

1. Find and review the SpaceX API documentation to understand what data is available and how to request it.

2. Import the Python requests module to make API calls.

3. Make a GET request to the endpoint you want data from

4. Convert the JSON response to a Python dictionary

5. Explore the dictionary to understand the structure and extract the specific data you need.

6. Optionally clean or transform the data into the structure you want to work with/analyze.

7. Load the data into a Pandas dataframe for easier manipulation and analysis.

7. Use the data to build visualizations, train models, etc.

[Code](#)

Data Collection - Scraping

1. Identify the website and pages you want to scrape.

2. Use the library to download the page content with Requests.

3. Parse the HTML content using the library APIs to extract the data with BeautifulSoup.

4. Traverse the parsed HTML using methods like find(), find_all() to isolate the data elements.

5. Repeat extraction process for additional pages if needed by looping over URLs.

6. Create Dictionary and Add data to keys

7. Create dataframe from dictionary and Export to file CSV.

Code

Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully; for example:
    - True Ocean means the mission outcome was successfully landed to a specific region of the ocean.
    - False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean.
    - True RTLS means the mission outcome was successfully landed to a ground pad.
    - False RTLS means the mission outcome was unsuccessfully landed to a ground pad.
    - True ASDS means the mission outcome was successfully landed on a drone ship.
    - False ASDS means the mission outcome was unsuccessfully landed on a drone ship.
- In this lab we will mainly convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful. here are the steps:
    - Import data into Python environment (pandas, numpy etc.).
    - Explore data - check for null values, data types, summary stats etc. Look for data quality issues.
    - Calculate the number of launches on each site, number & occurrence of each orbit, and number & occurence of mission outcome of the orbits.
    - Create a landing outcome label from Outcome column.
    - Export it to a CSV.

Code

EDA with Data Visualization

- Scatter Plot: Represents individual data points on a two-dimensional graph.
  - FlightNumber vs. PayloadMass
  - FlightNumber vs LaunchSite
  - LaunchSite Vs PayloadMass
  - Orbit Vs FlightNumber
  - Orbit Vs PayloadMass
- Bar Plot: Displays categorical data with rectangular bars, where the length of each bar corresponds to the value it represents.
  - Success rate vs. Orbit
- Line Plot: Connects data points with straight lines, emphasizing the trend or pattern in the data.
  - Success rate vs. Year

Code

Build an Interactive Map with Folium

- Mark all launch sites on a map.
  - We first need to create a folium Map object, with an initial center location to be NASA Johnson Space Center at Houston, Texas.
- Mark the success/failed launches for each site on the map
  - Create markers for all launch records. If a launch was successful (class=1), then we use a green marker and if a launch was failed, we use a red marker (class=0)
- Calculate the distances between a launch site to its proximities
  - Let's first add a MousePosition on the map to get coordinate for a mouse over a point on the map. As such, while you are exploring the map, you can easily find the coordinates of any points of interests. You can draw a line betwee a launch site to its closest city, railway, highway, etc)

Code

Build a Dashboard with Plotly Dash

- Dropdown list to enable Launch Site selection.
- Pie chart to show the total successful launches count for all sites.
- Slider to select payload range.
- Scatter chart to show the correlation between payload and launch success.

Code

Predictive Analysis (Classification)

- This section compares four classification algorithms: Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors with the following steps:
  - Import Libraries.
  - Load the dataframe.
  - Create a NumPy array from the column Class in data, by applying the method to_numpy() then assign it to the variable Y.
  - Standardize the data in X then reassign it to the variable X using the transform.
  - Use the function train_test_split to split the data X and Y into training and test data.
  - Create all algorithms object then create a GridSearchCV object, Fit the object to find the best parameters from the dictionary parameters.
  - Calculate the accuracy on the test data using the method score.
  - Find the method performs best.

Code

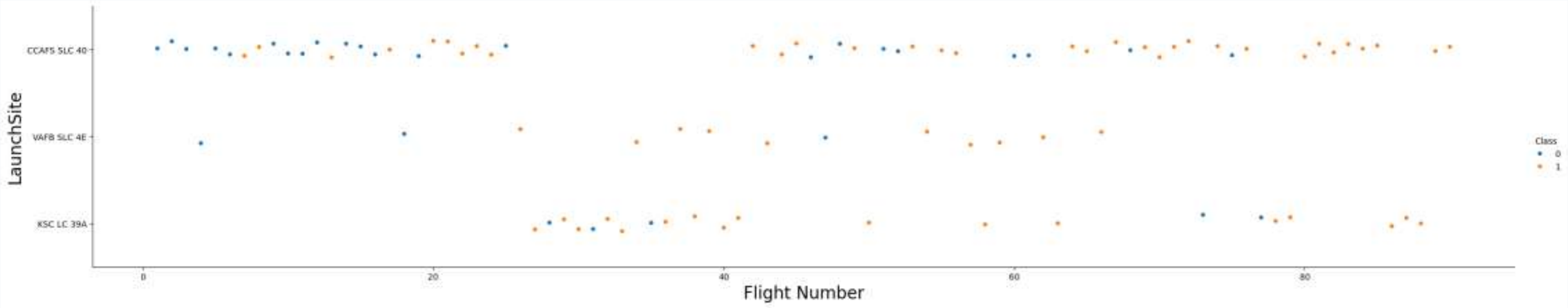
Results

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

Section 2

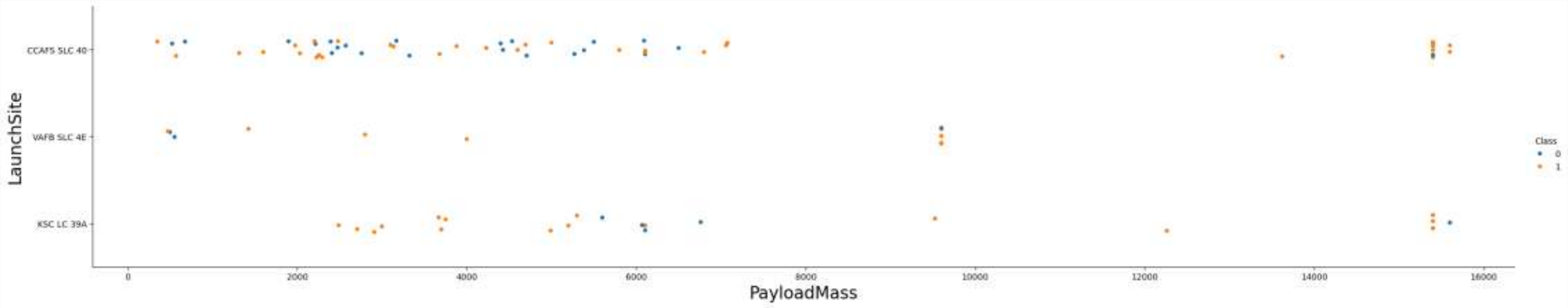
Insights drawn from EDA

Flight Number vs. Launch Site



- It can be seen that most flights were at the CCAFS SLC-40 launchsite and the fewest were at VAFB SLC 4E.
- From this statement it can be concluded that Launchsite CCAFS SLC-40 is the best at the moment.

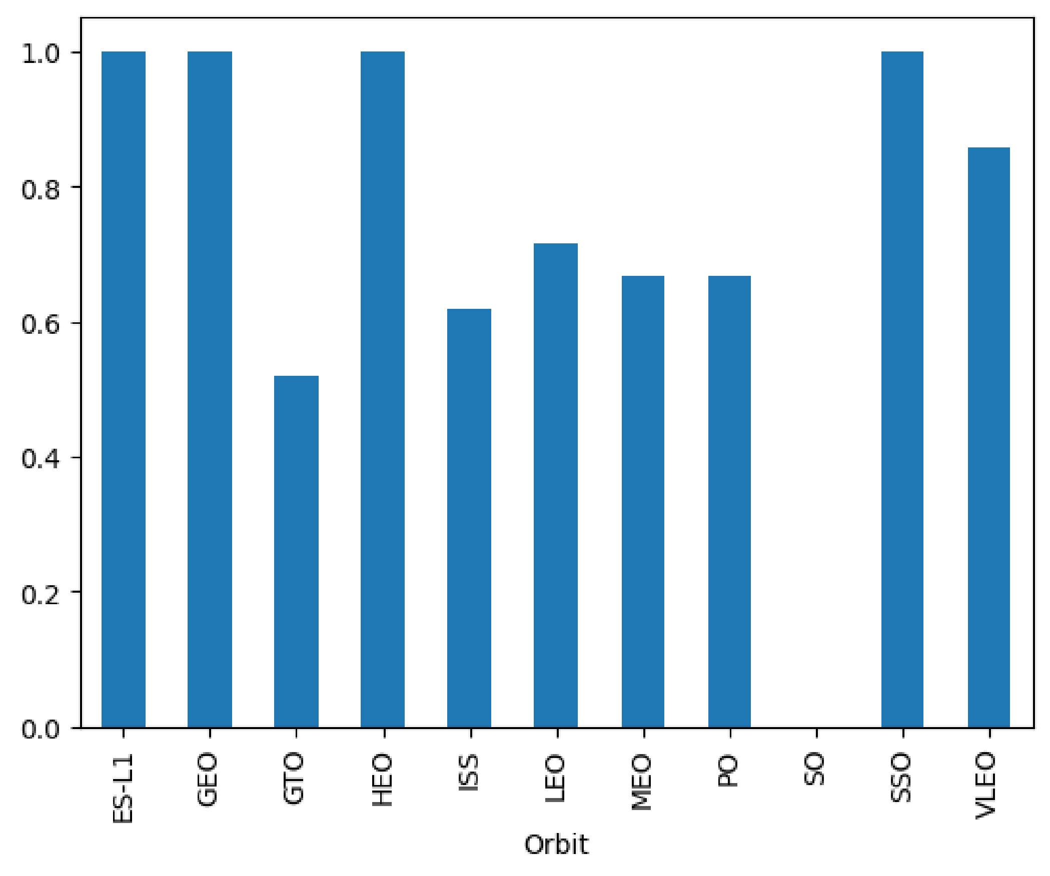
Payload vs. Launch Site



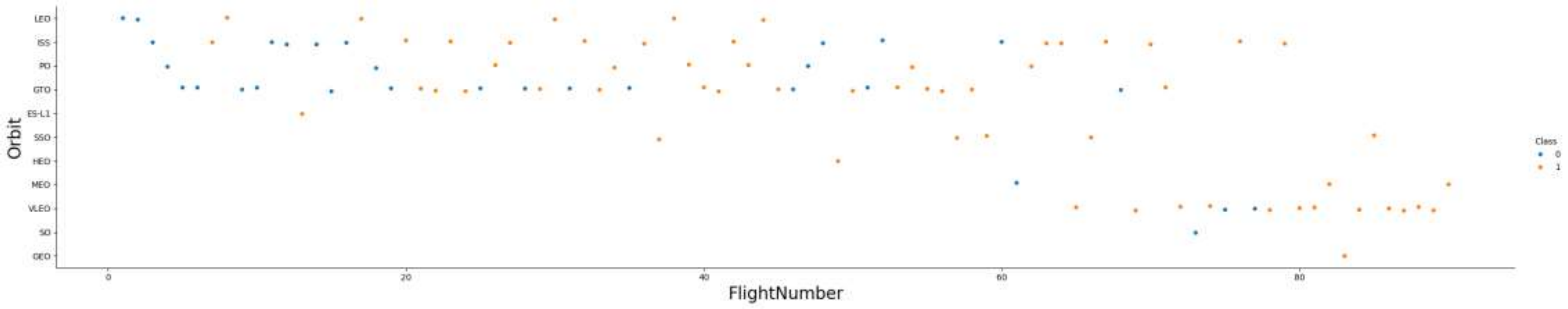
- Payloadmass over 8000 kg has a high success rate.
- Payload mass over 12000 kg can only be seen on the CCAFS SLC-40 and KSC LC-39A Launchsite

Success Rate vs. Orbit Type

- It can be seen that ES-L1, GEO, HEO, SSO are the orbits that have the highest success rate, namely approximately 100%.
- Meanwhile, SO orbit has the lowest success rate, namely 0%.

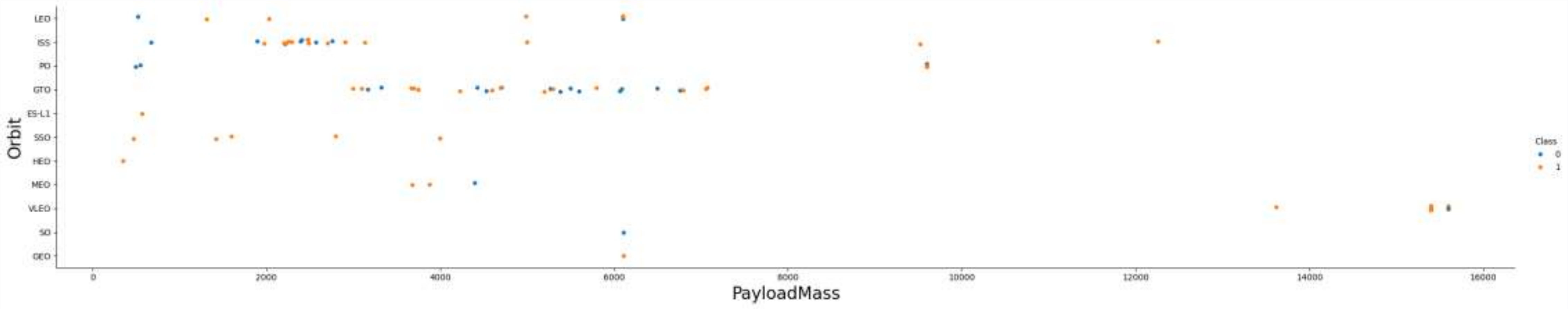


Flight Number vs. Orbit Type



- It can be seen that the success rate of LEO orbits increases with the number of flights.
- Meanwhile, VLEO orbit can be said to be a new business opportunity, this is due to the recent increase in frequency.

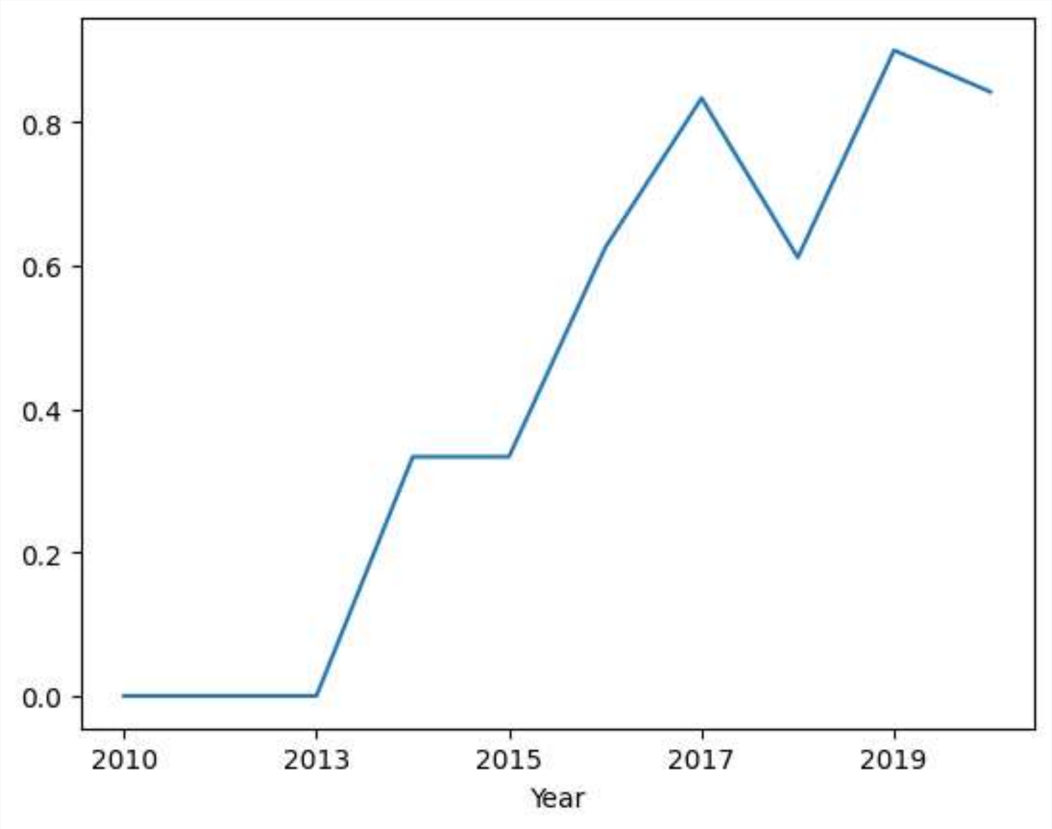
Payload vs. Orbit Type



- The weight of the payload may have an influence on the launch success rate in a particular orbit.
  - Heavier payloads increase success rates in LEO orbits.
- There are few launches to SO and GEO orbits.

Launch Success Yearly Trend

- The success rate began to increase in 2013 and continued until 2019, although it experienced a decline in 2018.
- Apart from that, it can be said that the first three years were a period of adjustment and improvement in technology so that there were no significant improvements



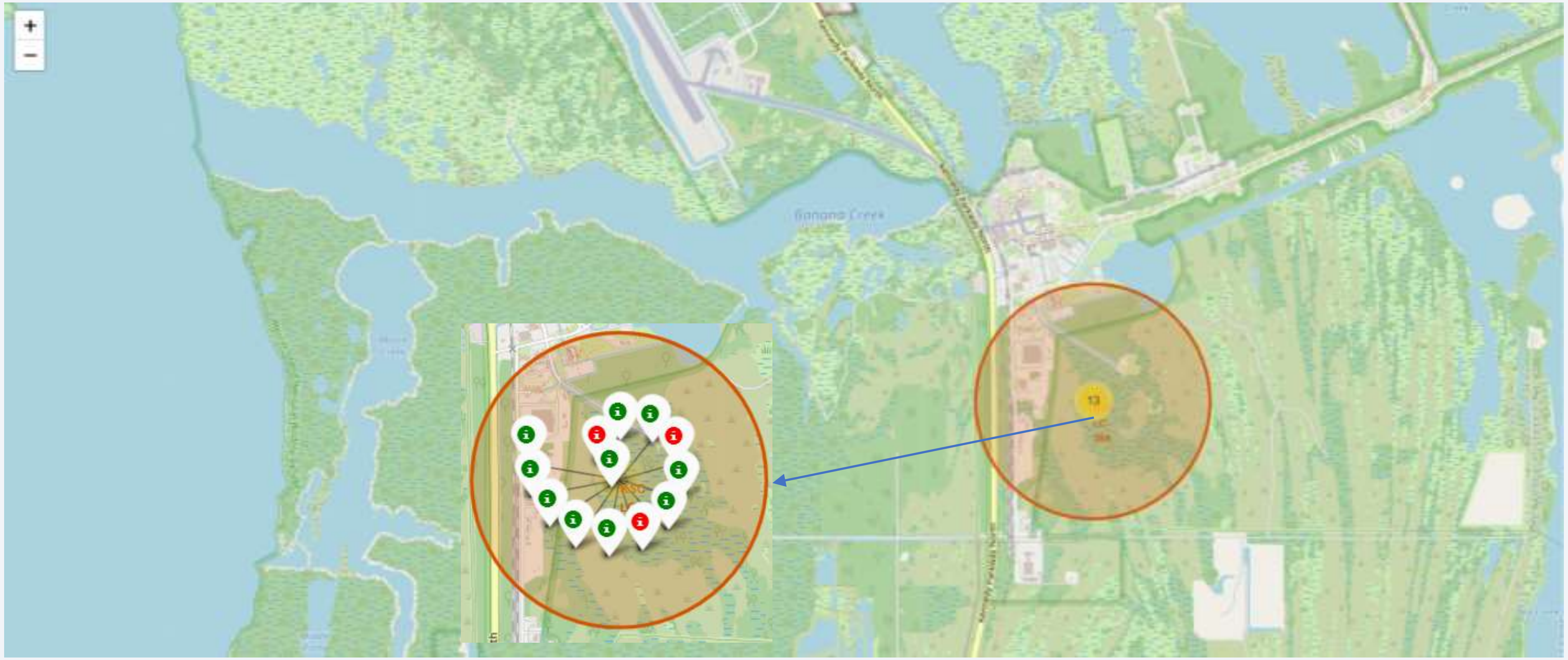
Section 3

Launch Sites Proximities Analysis

All Launch Sites

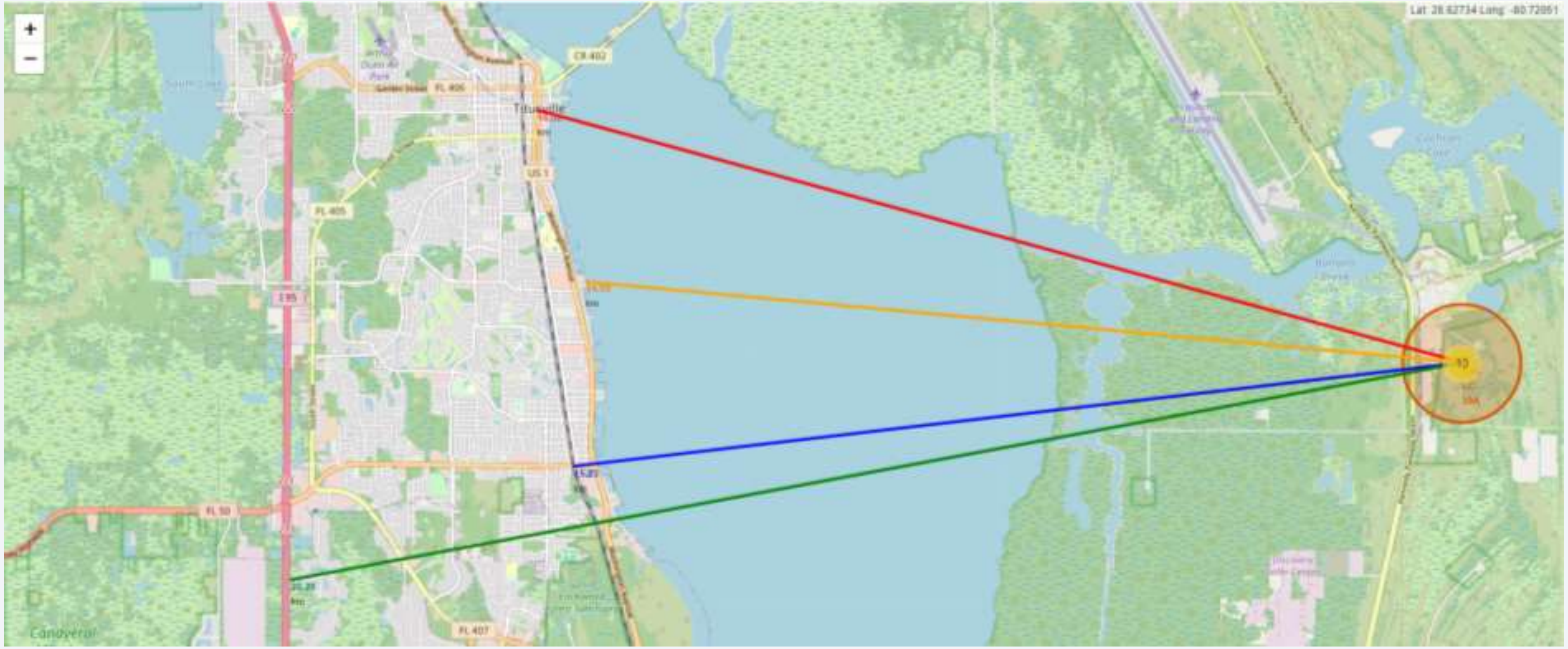
It can be seen that the Spacex launch site is located near the sea or perhaps on the beach to be safe.

Color Labeled Markers



- Example of launch results from the KSC LC-39A launch site.
- The Green marker indicates a successful launch while the red marker indicates a failed launch. So it can be seen that the KSC LC-39A has a high success rate.

Logistics and Safety

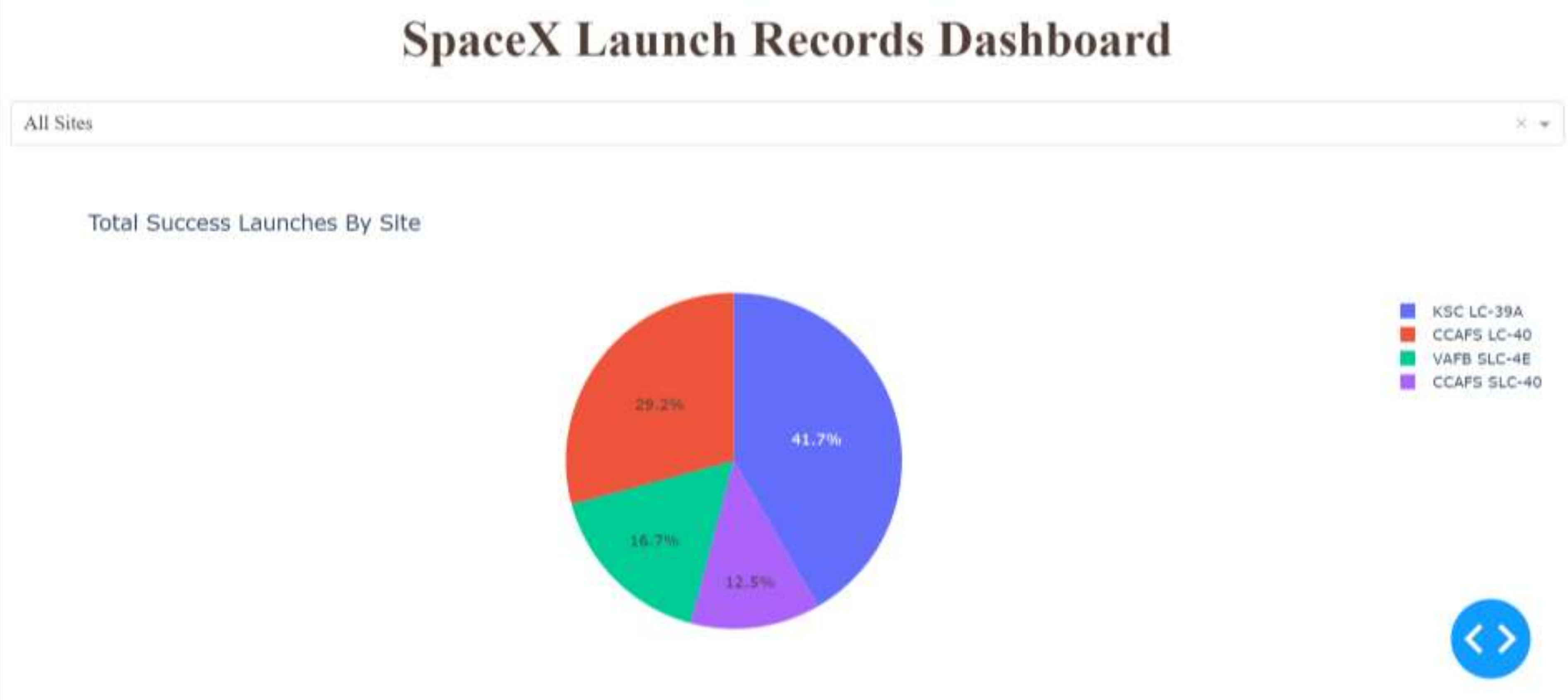


It can be seen that KSC LC-39A is far from railways, highways, coastlines, and maintains a certain distance from the city.

Section 4

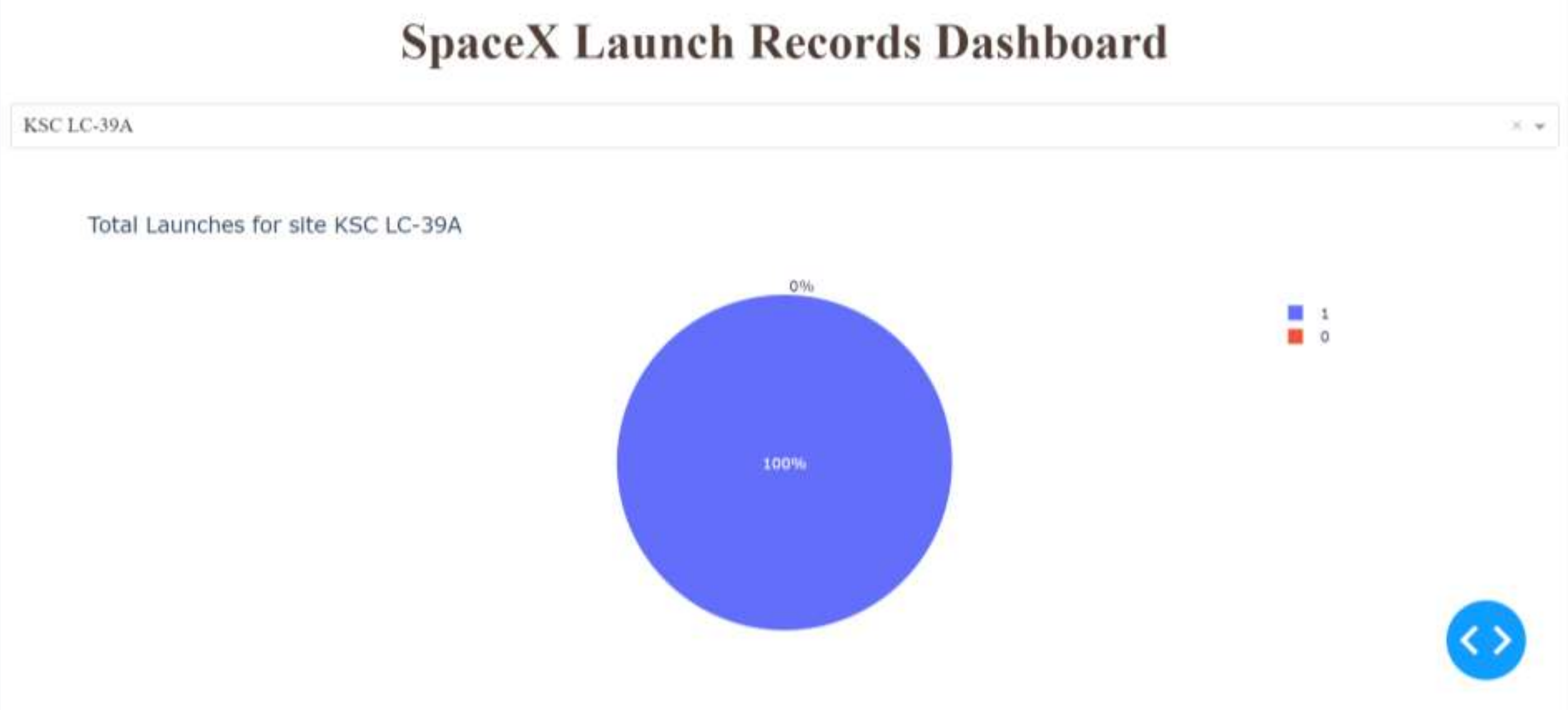
Build a Dashboard with Plotly Dash

Successful Launches by Site



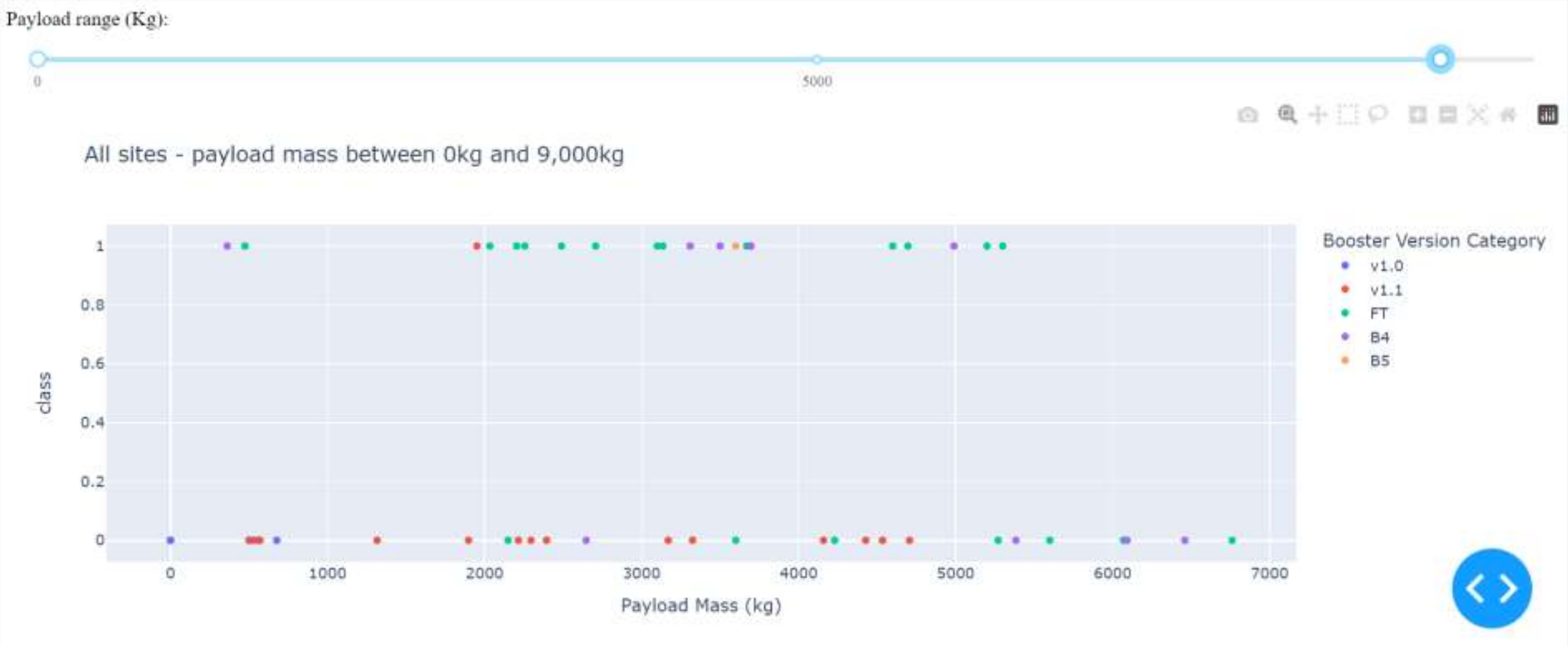
It can be seen that the KSC LC-39A launch site has the highest launch success rate.

Successful Launches by Site



It can be seen that the KSC LC-39A launch site has a success rate of 100%.

Payload vs. Launch Outcome



Payloads under 6,000kg and FT boosters are the most successful combination.

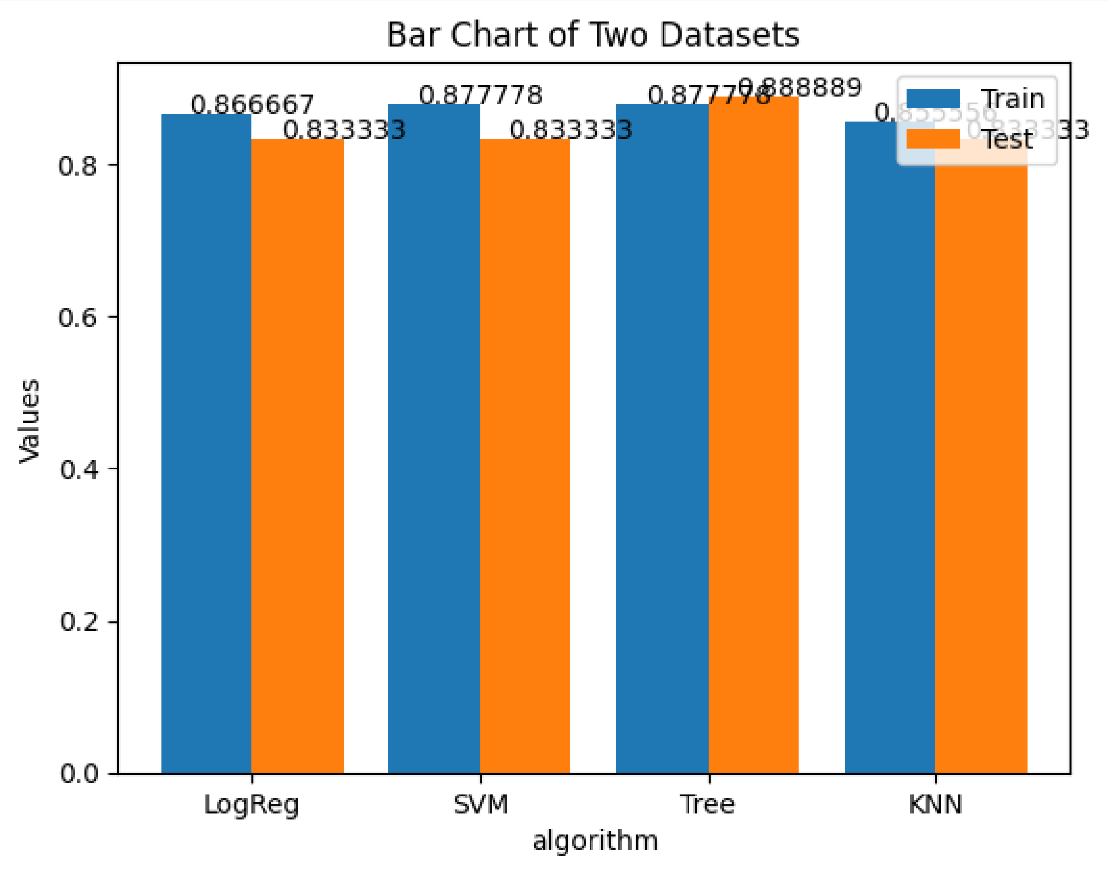
Section 5

Predictive Analysis (Classification)

Classification Accuracy

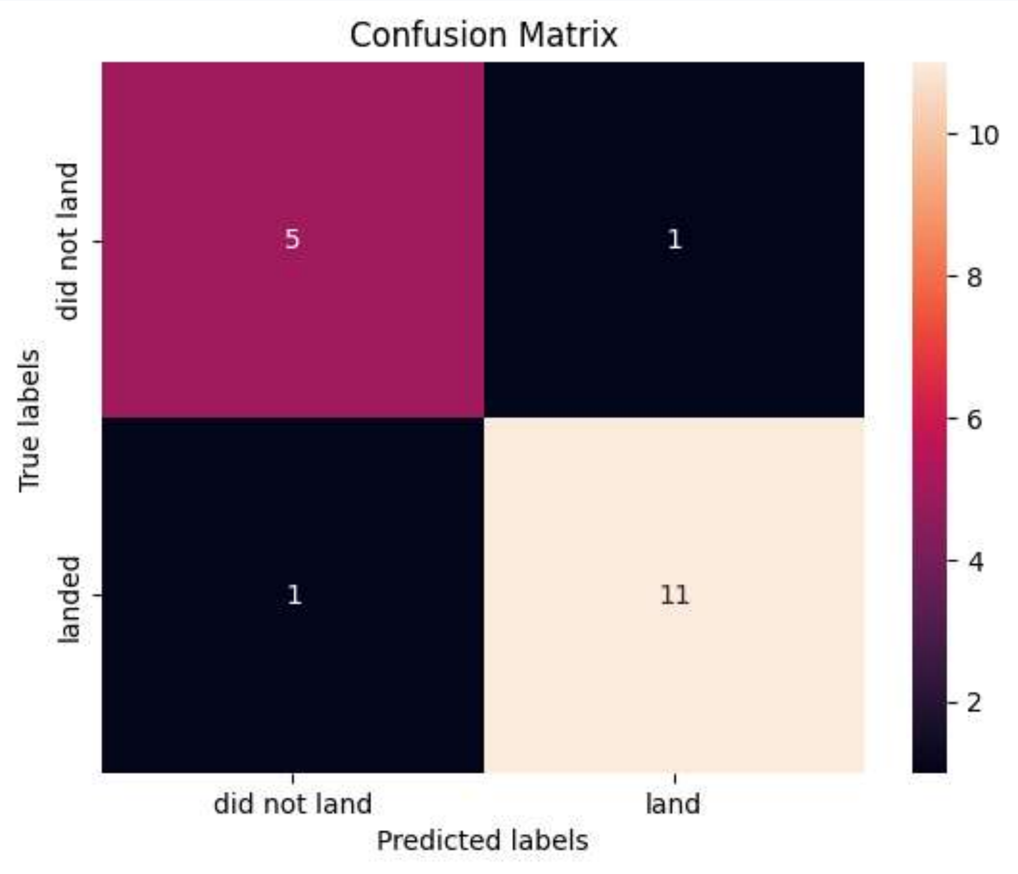
It can be seen that the decision tree has the highest accuracy, namely 88,89% for test and 87,78% for train with the tuned hpyerparameters (best parameters) being as follows:

- criterion: gini
- max_depth: 6
- max_features: auto
- min_samples_leaf: 2
- min_samples_split: 10
- splitter: random



Confusion Matrix

The confusion matrix of Decision Tree Classifier proves its accuracy by showing the number of true positives compared to false ones.



Conclusions

- It can be seen that most flights were at the CCAFS SLC-40 launchsite and the fewest were at VAFB SLC 4E. From this statement it can be concluded that Launchsite CCAFS SLC-40 is the best at the moment.
- ES-L1, GEO, HEO, SSO are the orbits that have the highest success rate (100%).
- The success rate began to increase in 2013 and continued until 2019, although it experienced a decline in 2018.
- The Spacex launch site is located near the sea or perhaps on the beach to be safe.
- The KSC LC-39A launch site has the highest launch success rate.
- The decision tree has the highest accuracy, namely 88,89% for test and 87,78% for train

Thank you!