IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

David Paul Ong
22 May 2025

OUTLINE

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

Executive Summary

01. Methodology

- The data was collected using SpaceX's API and with web scraping.
- The data was wrangled and EDA was conducted using SQL and chart visualizations.
- A geomap was used to visualize launchpad locations alongside the failure rates of each.
- An interactive dashboard was created .

02 Results

There are correlations between variables such as payload mass, orbit target, launch site to the success of a SpaceX rocket launch. A Decision Tree model was created having high accuracy in predicting a launches' success.

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, much of the savings is because SpaceX can reuse the first stage of a launch.

Therefore, the ability to accurately predict whether a first stage will successfully land is very useful. Such a predictive capability allows for a more precise determination of launch costs and informing strategic decisions

Methodology

Methodology

1. Data Collection

SpaceX APIs and web scraping techniques.

2. Data Wrangling

Dataset cleaning and labeling

3. EDA

SQL and data charts

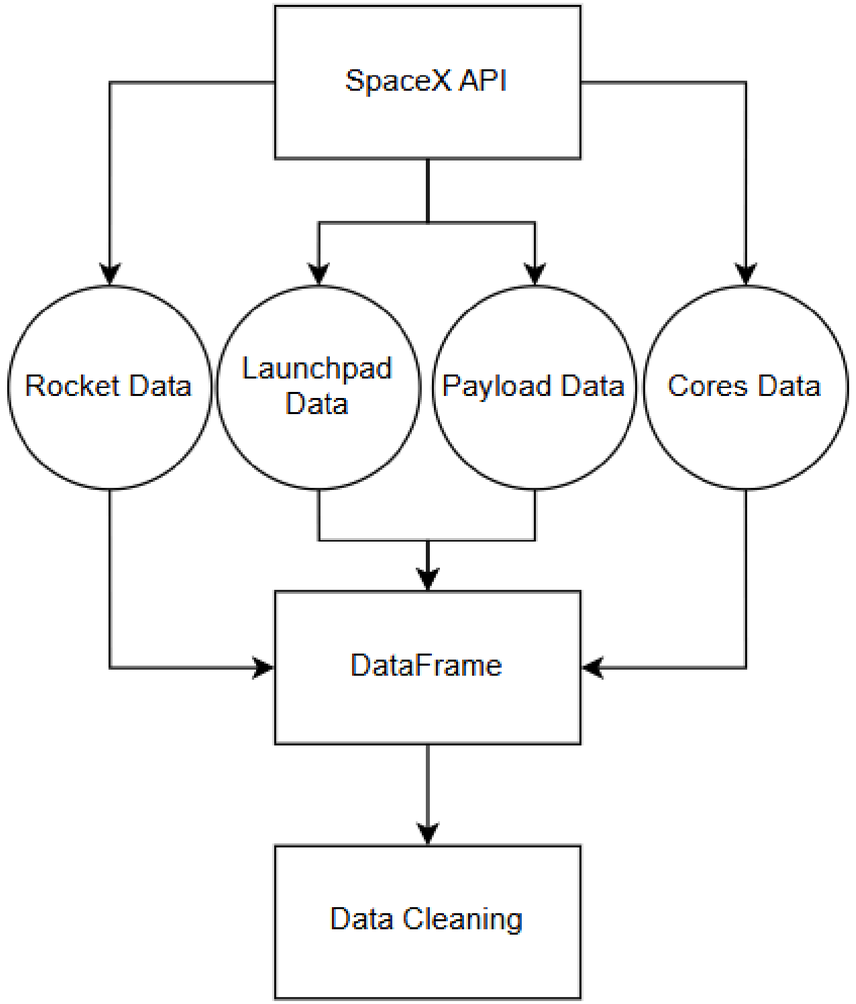
4. Interactive Visualizations

Interactable Geomaps and dashboards

5. Predictive Modelling

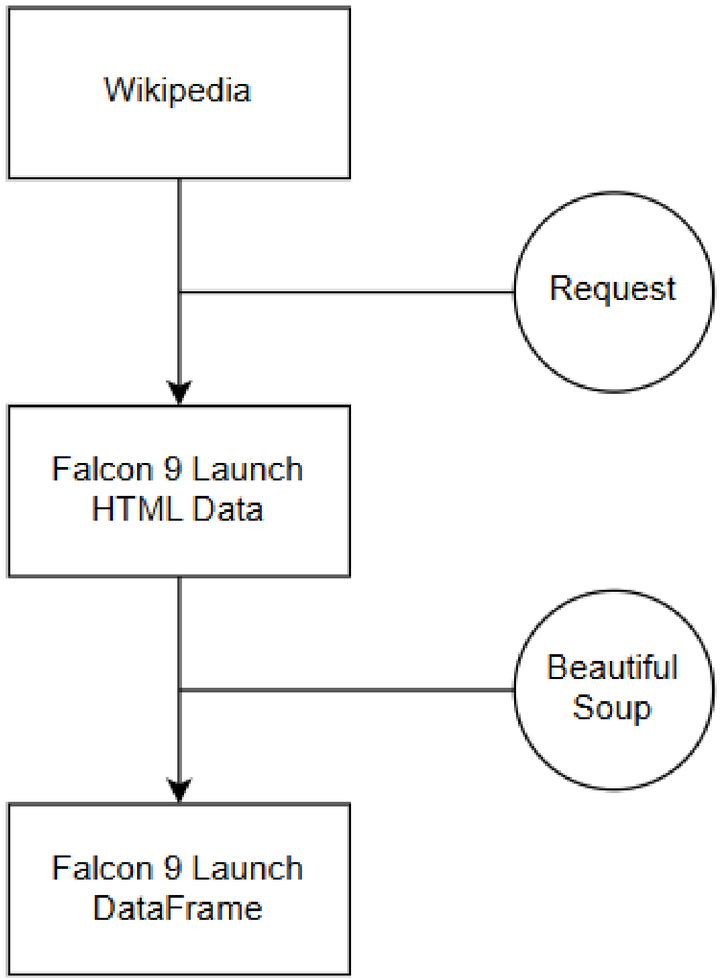
Training and tuning machine learning models

Data Collection - SpaceX API



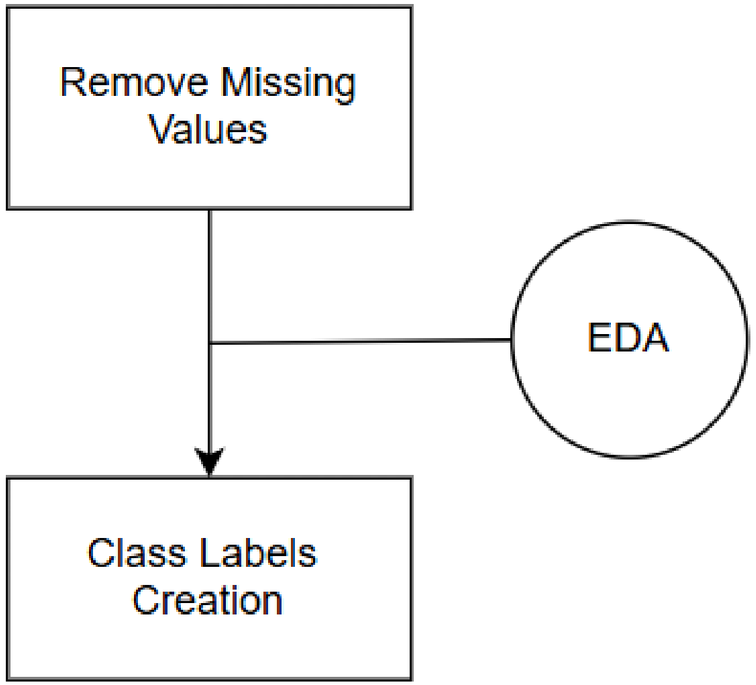
- Four API calls were made, each retrieving data from SpaceX launches.
- They were combined into a single clean DataFrame which will be used for further processing
- Github

Data Collection - Scraping



- The List of Falcon 9 and Falcon Heavy launches Wikipedia page was scraped
- The HTML data was formatted into a Pandas DataFrame with the help of the BeautifulSoup Python library.
- Github

Data Wrangling



- Missing values were identified and removed
- A binary class label feature was created with the help of EDA. The labels of the class being whether a launch was successful or not.
- Github

EDA with Data Visualization

- Cat plots, bar charts, scatter plots, and line charts are used to analyze the relationship between variables in the dataset. These visualizations help identify patterns, trends, and correlations.
  - Cat plots and Scatter plots are useful for visualizing relationships between categorical and numerical variables.
  - Bar charts are effective for comparing quantities across different categories.
  - Line charts are ideal for showing trends over time or continuous sequences.

- Github

Interactive Map with Folium

- In this section, an interactive map was built using the Folium library to visually explore SpaceX launch sites. Key features of the map include:
  - Launch Site Markers: Displayed the geographical locations of the four main SpaceX launch sites on a world map using marker pins.
  - Zoom & Interactivity: Enabled zoom functionality to closely inspect each launch site.
  - Launch Outcomes: Visual indicators on the map to show whether the launches from a site were successful or not.
  - Proximity Lines: Drew lines from each launch site to nearby features such as railways, highways, and coastlines.

- Github

Interactive Dashboard with Dash

- An interactive dashboard using Dash was created, showing the relationships of several variables in a more visually engaging way :
  - Pie Charts: A pie chart showing the success-to-failure ratio of each launch site or the success rate ratio of all launch sites
  - Scatter plot showing the relationship between the Payload of a launch to its success. The range of the payload mass is customizable
  - Showing the boosters used in each launch based on the success or failure of the launch

- Github

Predictive Analysis (Classification)

- Machine learning models were trained to predict the success or failure of a SpaceX rocket launch. The models trained are:
  - Logisstic Regression:
  - SVM
  - Decision Tree
  - KNN

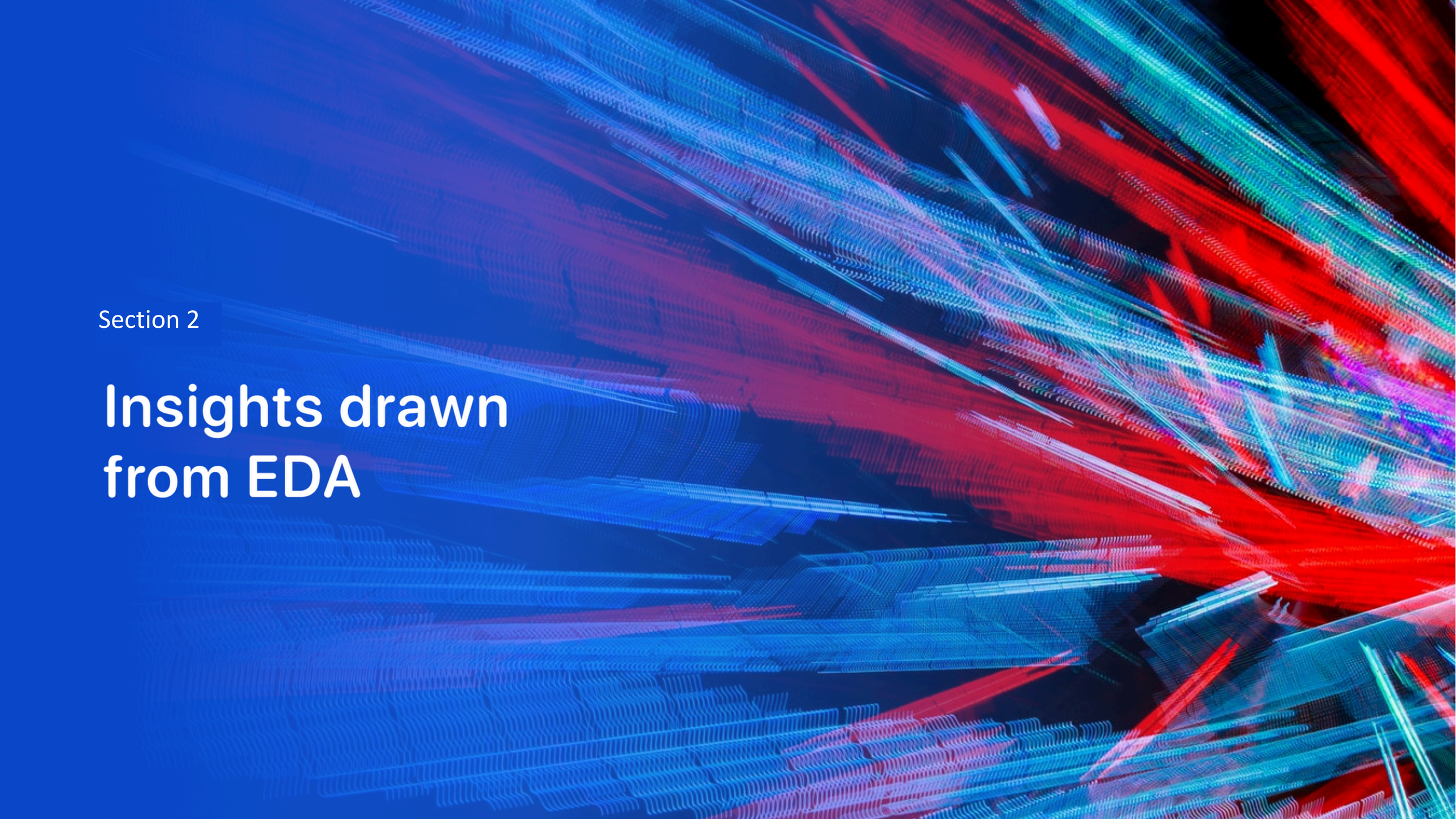
- A grid search was performed for all models using 5-fold cross-validation.

- Github

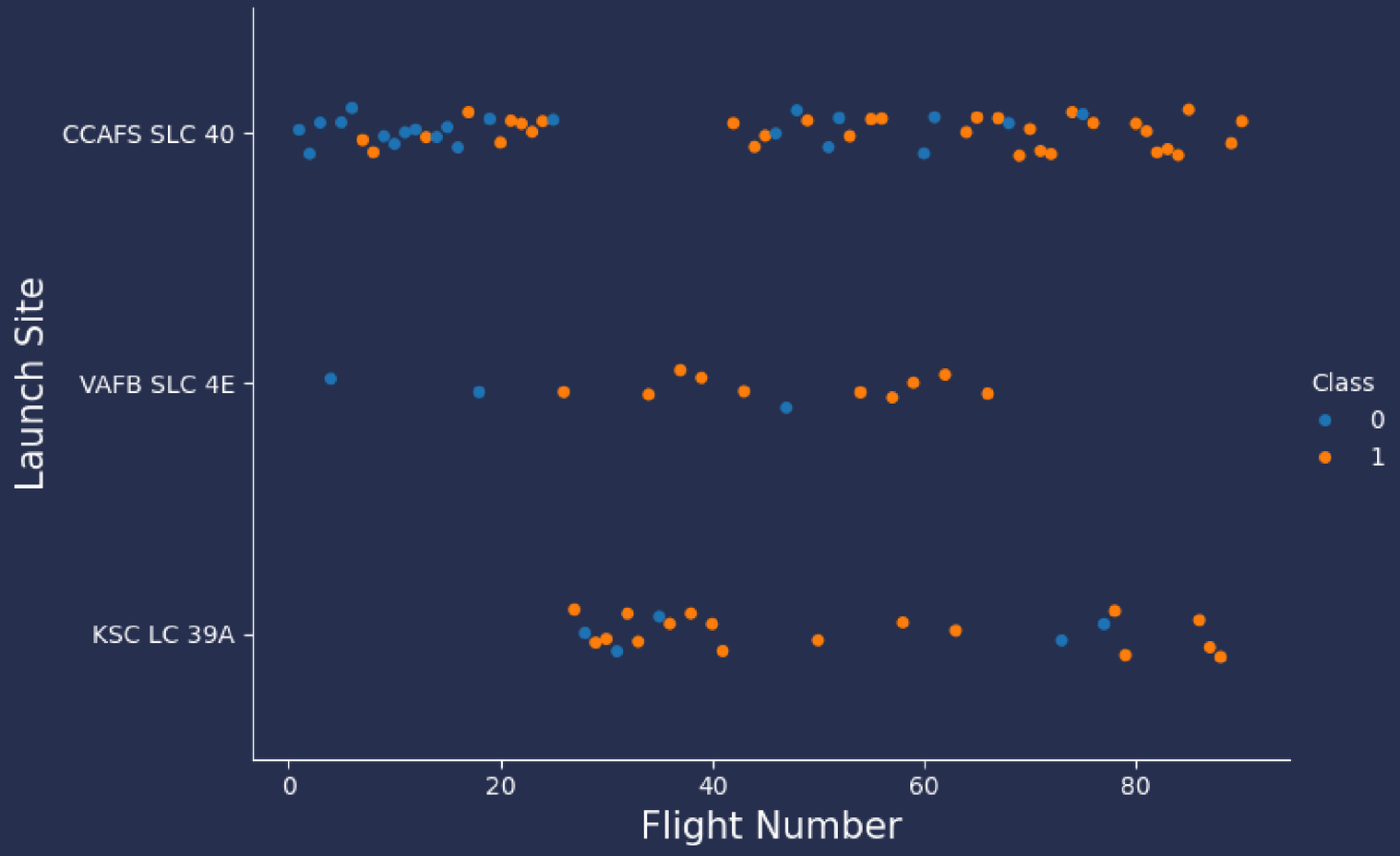
Results

Section 2

Insights drawn from EDA

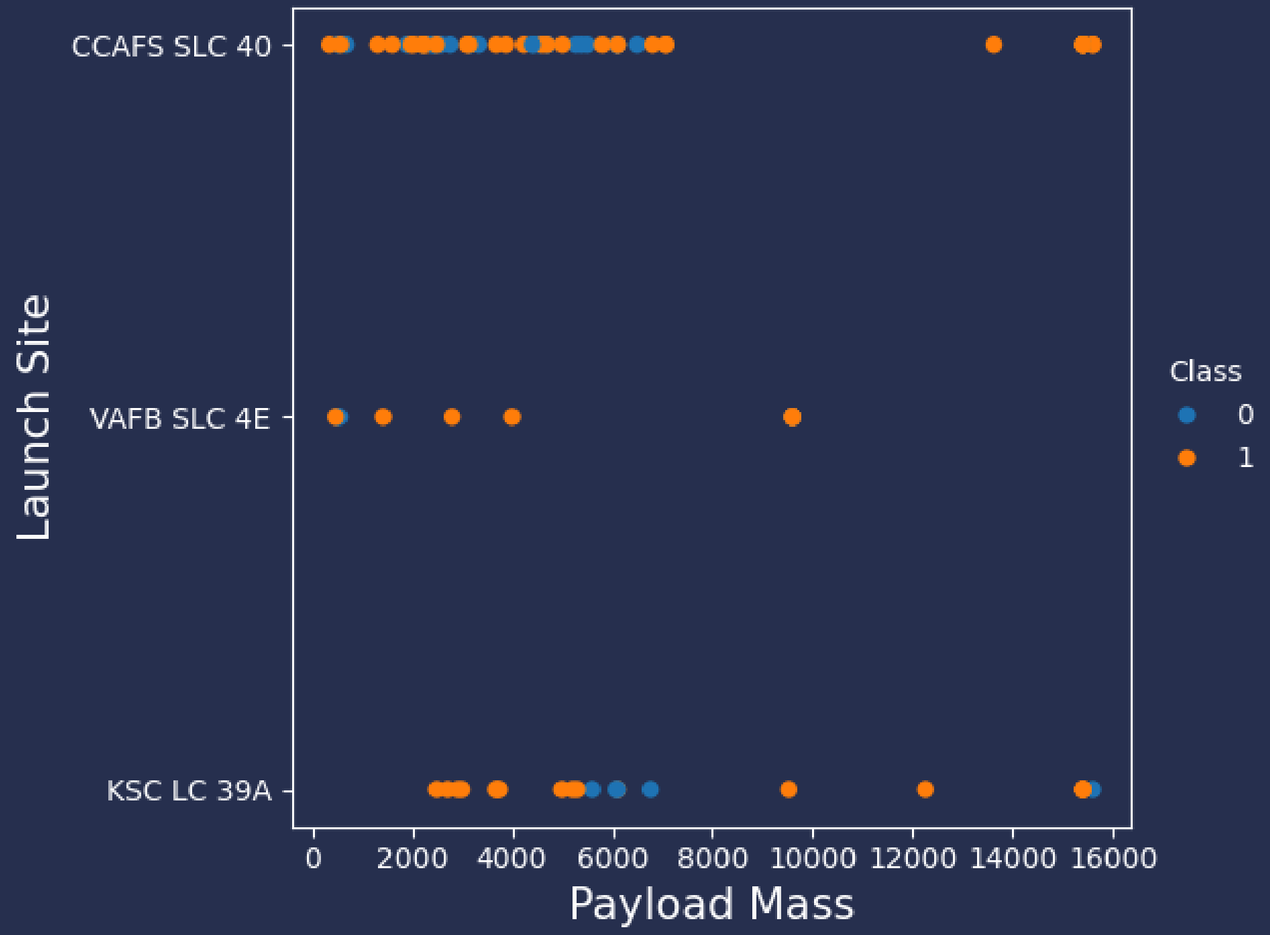


Flight Number vs. Launch Site



There are more launch failures in earlier flight numbers, particularly in CCAFS SLC 40.

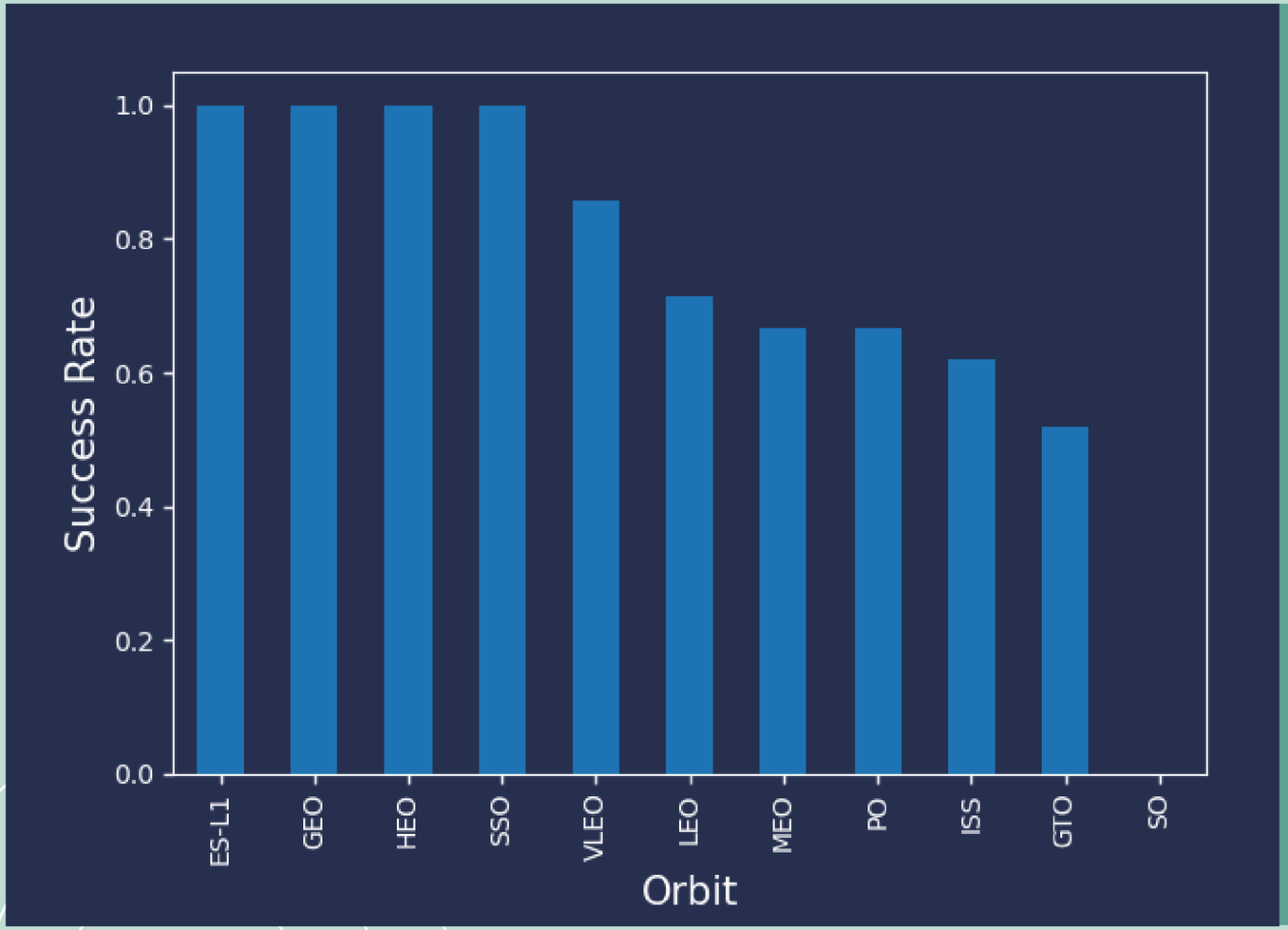
Payload vs. Launch Site



The scatter plot shows that for the VAFB-SLC launchsite there are no rockets launched for a heavy payload, none exceeding 10000 KGs.

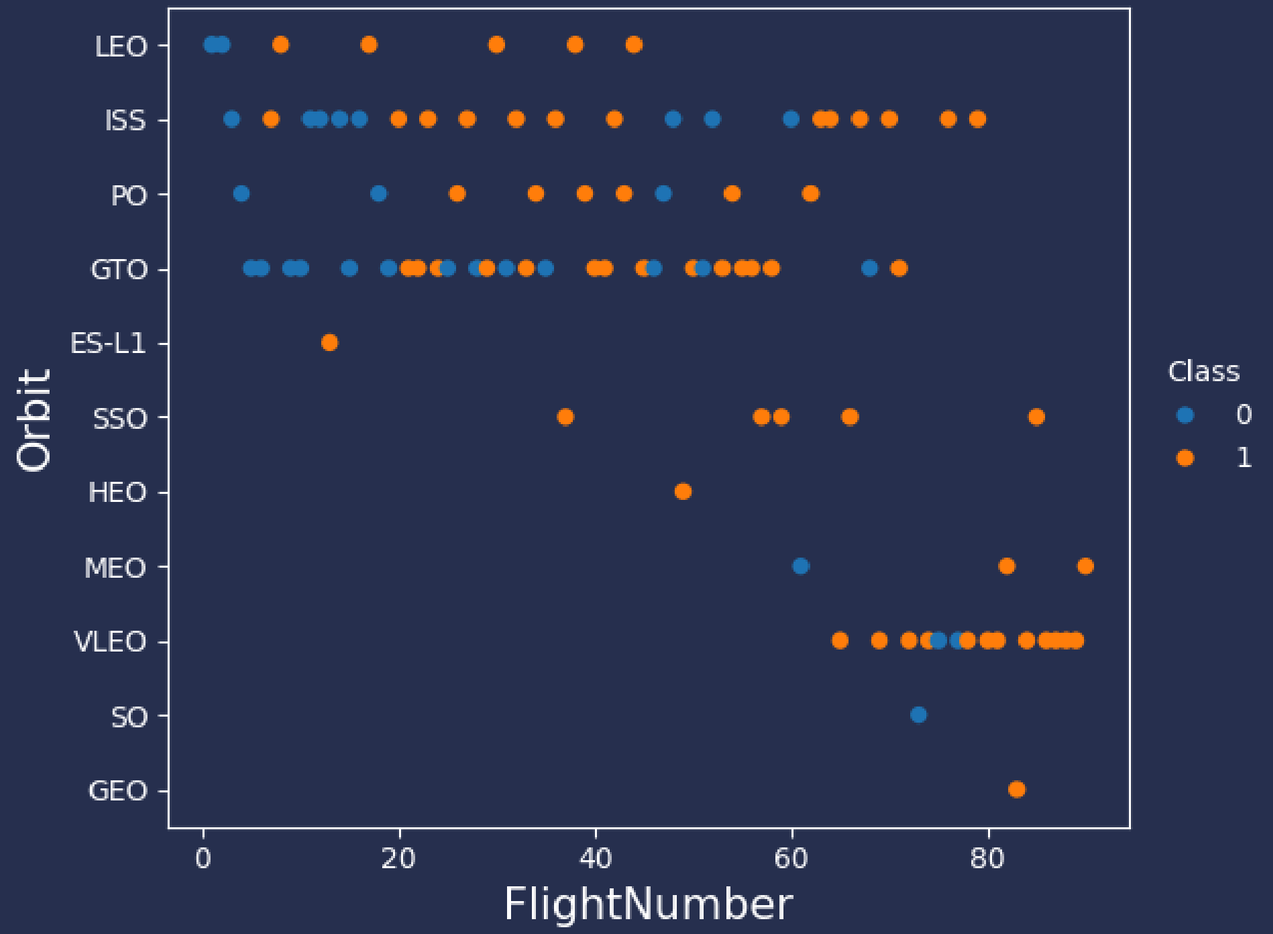
Most launches with heavy payloads tend to be a success, with one exception at KSC LC 39A.

Success Rate vs. Orbit Type



Launches to orbit type ES-L1, GEO, HEO and SSO have never failed. In contrast, launches to SO have never succeded

Flight Number vs. Orbit Type

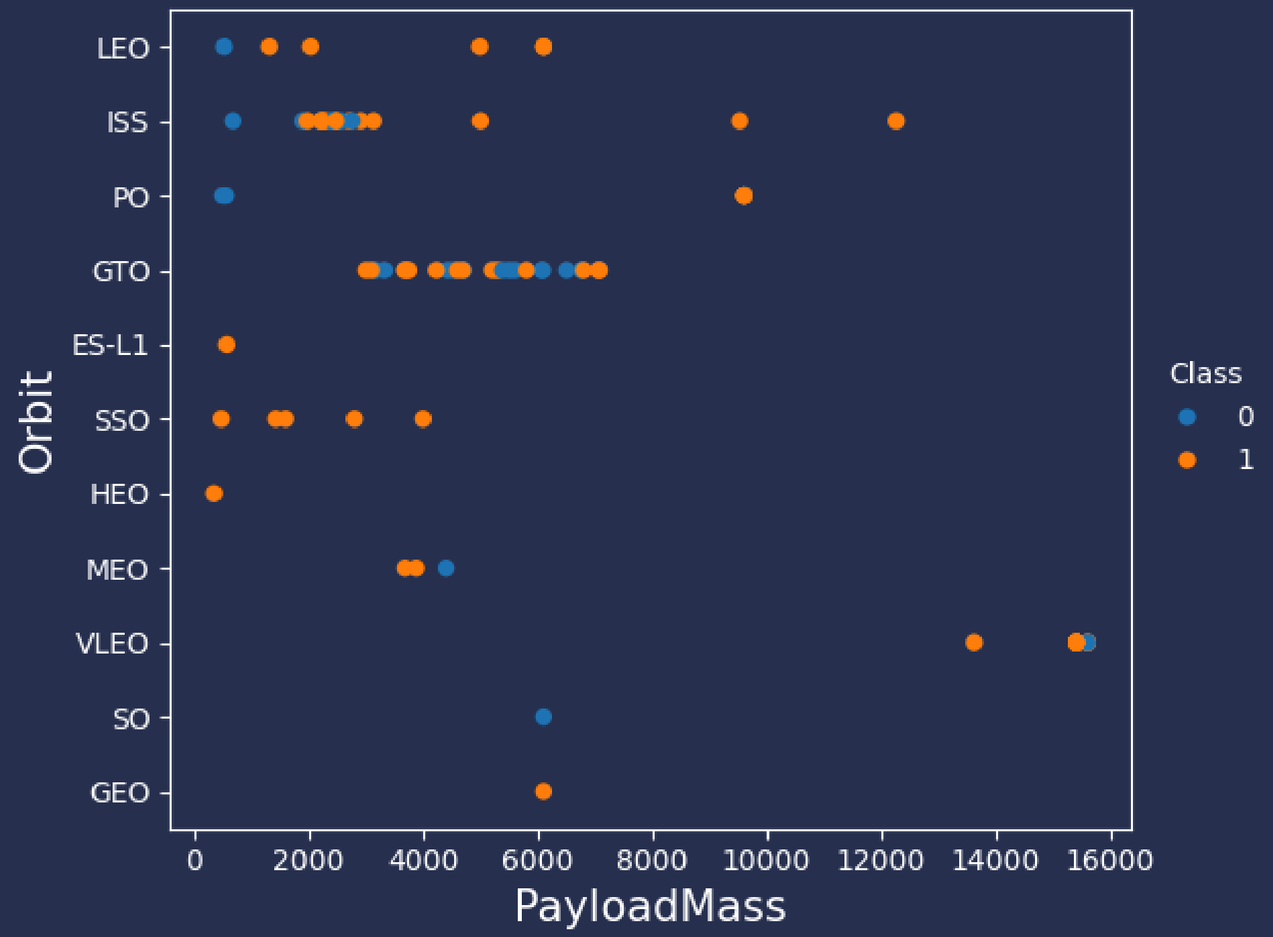


You can observe that in the LEO orbit, success seems to be related to the number of flights.

The ISS and PO orbits seem to have a somewhat similar alternating pattern at certain flight number sections.

In the GTO orbit, there appears to be no relationship between flight number and success.

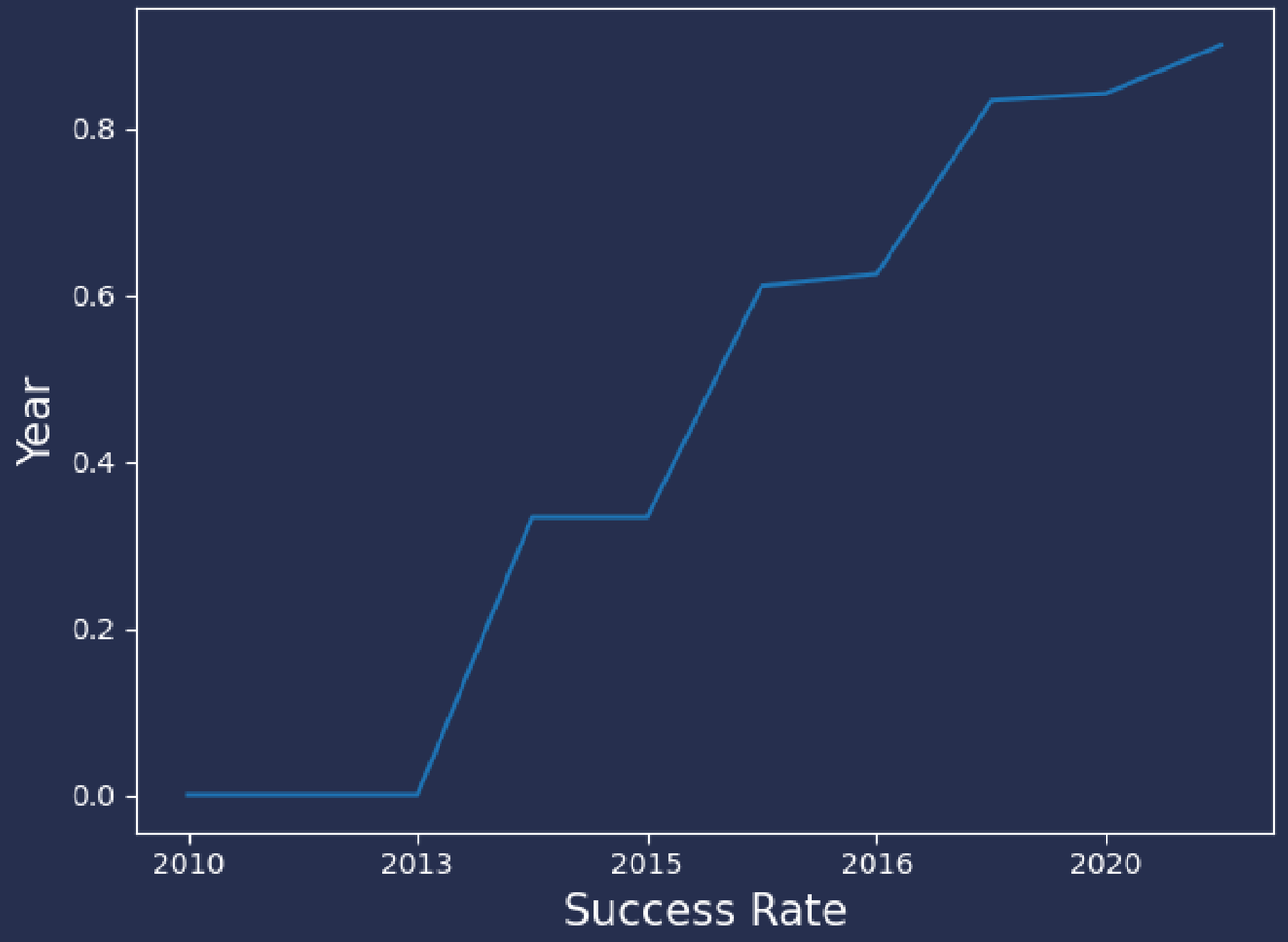
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



Sucess rate of launches kept increasing since 2013 until 2020

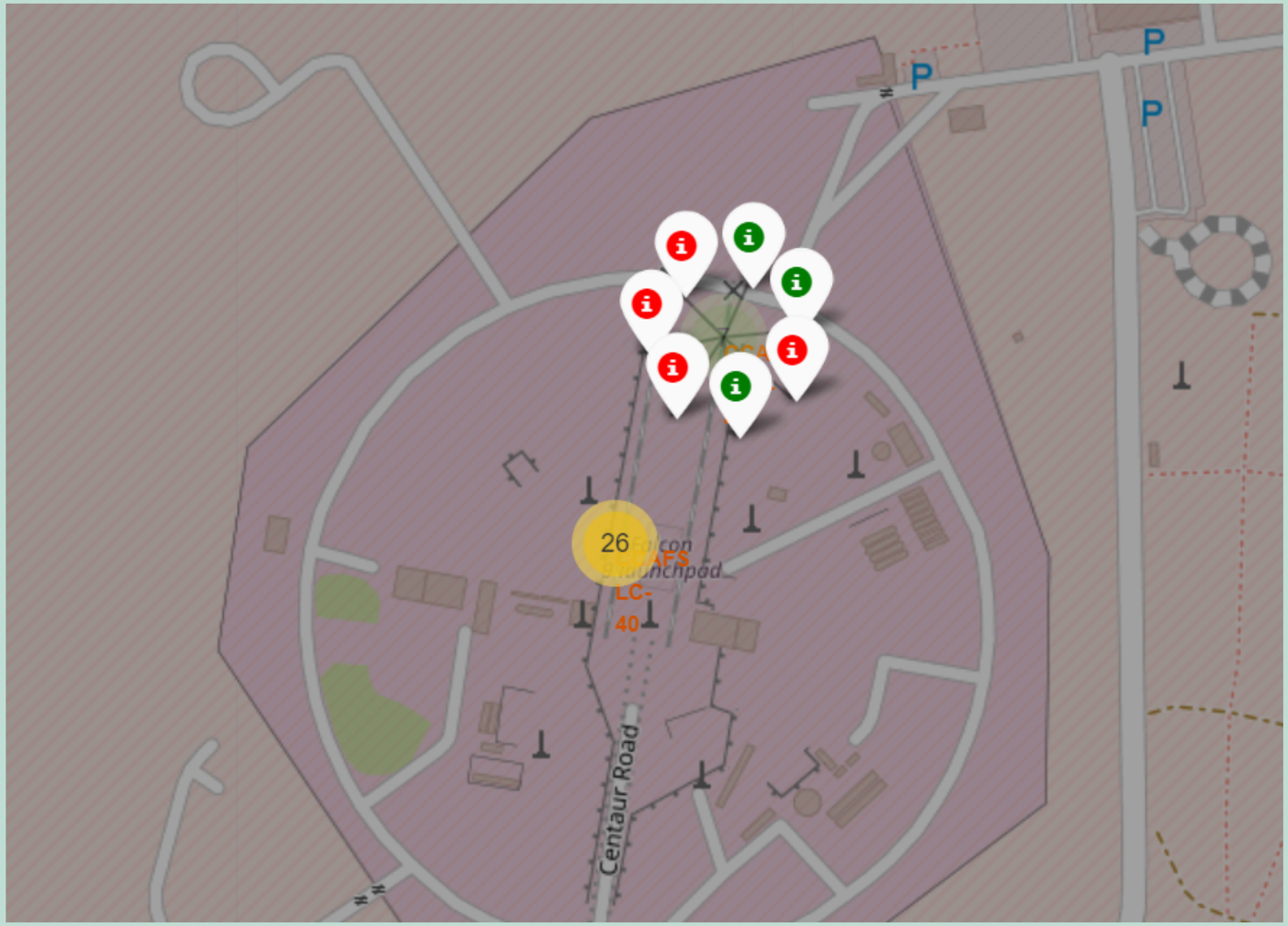
Section 3

Launch Sites Proximities Analysis

SpaceX Launch Site Locations

These are the locations of the four SpaceX launch sites shown in a geomap

Labelled Launch Outcomes

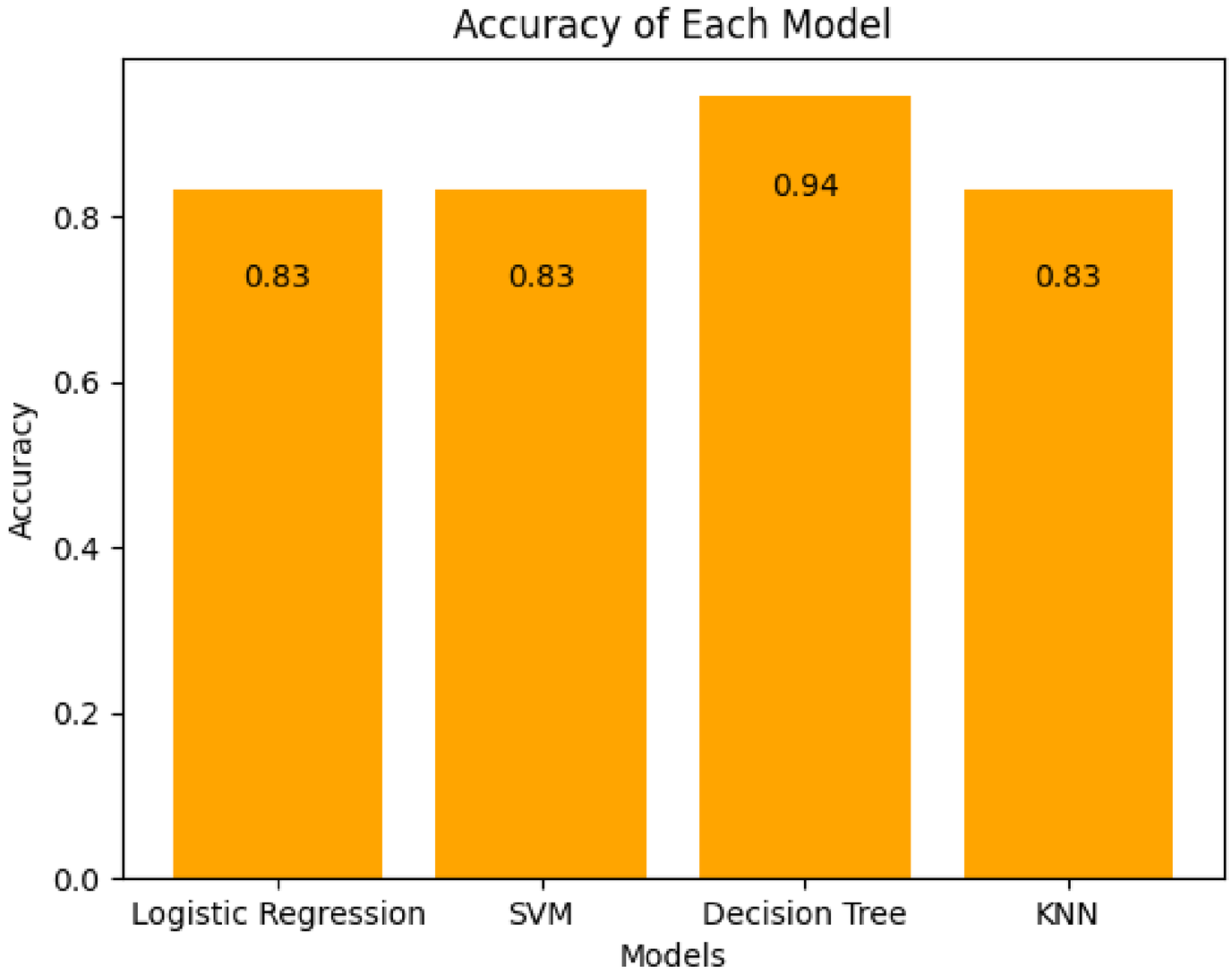


You can zoom into each launch site and see the success or failure outcomes of launches from that site

Section 5

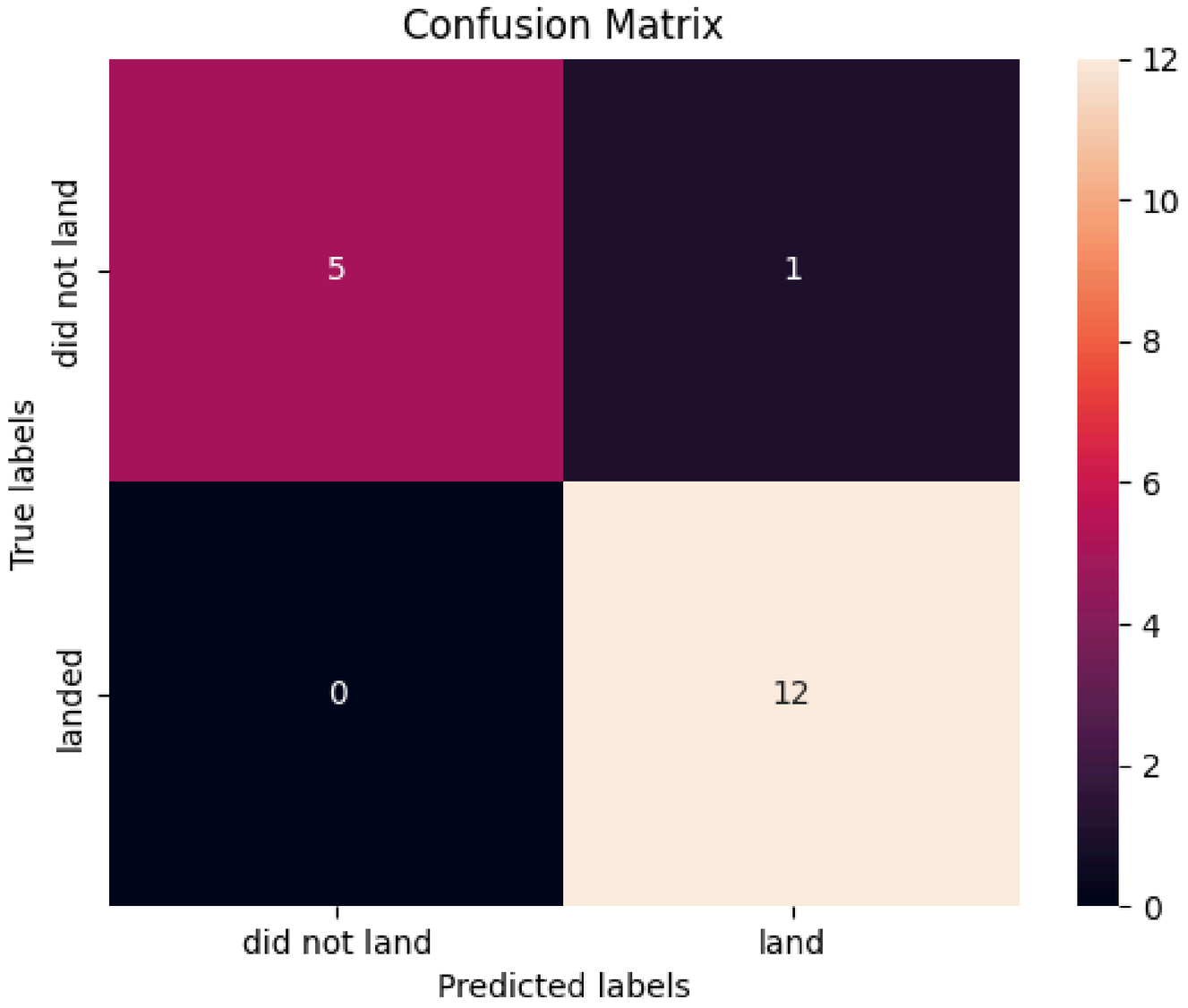
Predictive Analysis (Classification)

Classification Accuracy



The best model by accuracy is the Decision Tree Model

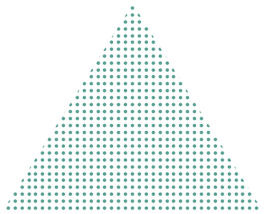
Confusion Matrix



The confusion matrix of the Decision Tree Model shows that the model only misclassified one instance.

Conclusions

- There are correlations between variables such as payload mass, orbit target, launch site to the success of a SpaceX rocket launch.
- A machine learning model was created and was found to be good at predicting the likelihood of a successful launch based on these features.
- The Decision Tree model demonstrated high accuracy predicting the test data.



Thank you!