Space technology's revolutionization

Ommair Ishaque
ommair@udel.edu

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of methodologies

- Data Collection through API
- Data Collection with Web Scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Data Visualization
- Interactive Visual Analytics with Folium
- Machine Learning Prediction

GitHub Notebook

- Data Collection through API

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

- Data Collection with Web Scraping

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-webscraping.ipynb

- Data Wrangling

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

- Exploratory Data Analysis with SQL

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-eda-sql-edx_sqllite.ipynb

- Exploratory Data Analysis with Data Visualization

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

- Interactive Visual Analytics with Folium

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/Dashboard_Ploty_dash.ipynb

- Machine Learning Prediction

https://github.com/ommair/Data-Science-and-Machine-Learning-Capstone-Project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Introduction

SpaceX, a revolutionary company, offers Falcon 9 rocket launches for 62 million dollars, while other providers charge up to 165 million dollars. SpaceX's brilliant idea to reuse the rocket's first stage by re-landing it saved most of this. Repetition will lower the price. As a data scientist at a startup competing with SpaceX, this project aims to create a machine learning pipeline to predict the first stage landing outcome.

This project determines the best rocket launch bid against SpaceX.

Main problem statement:

- One of the issues was pinpointing all of the variables that could affect the success of the landing.
- The interplay between the various factors and their impact on the final result.
- Third, the optimal scenario required to maximize the chances of a safe landing.

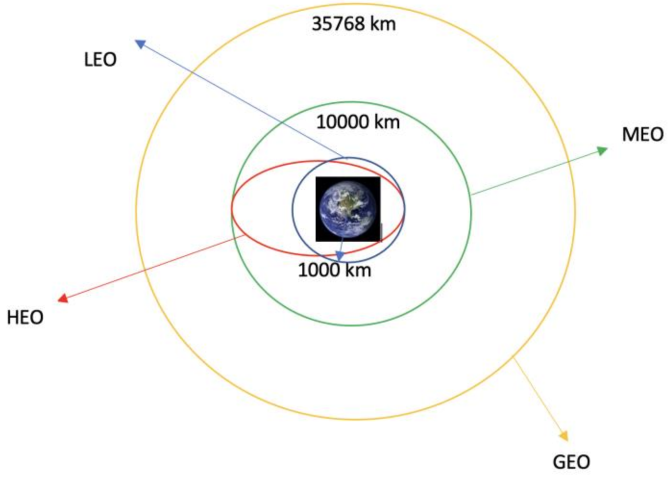
Methodolgy

- ✓ Methods of data collection:
  - ➢ Information was gathered via the SpaceX REST API and Wikipedia web scraping.
- ✓ Gather and organize data
  - ➢ One-hot encoding was used for categorical features in the data processing.
- ✓ Explore data visually and through SQL to perform exploratory analysis (EDA).
- ✓ Analyze data visually and interactively with Folium and Plotly Dash.
- ✓ Use classification models for predictive analysis.
  - ➢ How to construct, tweak, and assess classifier performance

Data Collection

- For the purpose of answering pertinent questions and evaluating outcomes, data collection is the process of gathering and measuring information on specific variables within a preexisting system. The dataset was obtained by using the REST API and web scraping to collect information from Wikipedia.
- The get request serves as the entry point for REST API. The Json content of the response was then decoded and a pandas dataframe was created from it with the help of the json_normalize() function. After that, we made sure there were no blanks in the data and filled them in where necessary.
- BeautifulSoup will be used to scrape the web for the launch records as an HTML table, whereupon the table will be parsed and converted into a pandas dataframe.

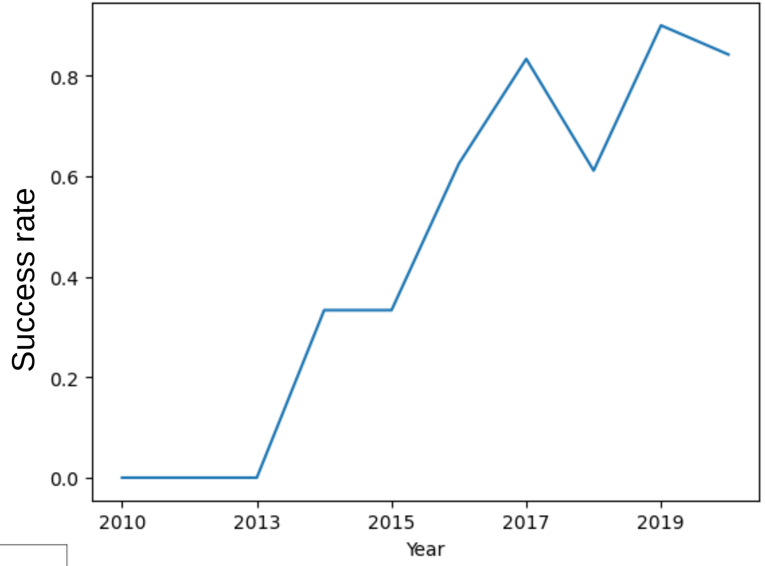
Data Wrangling

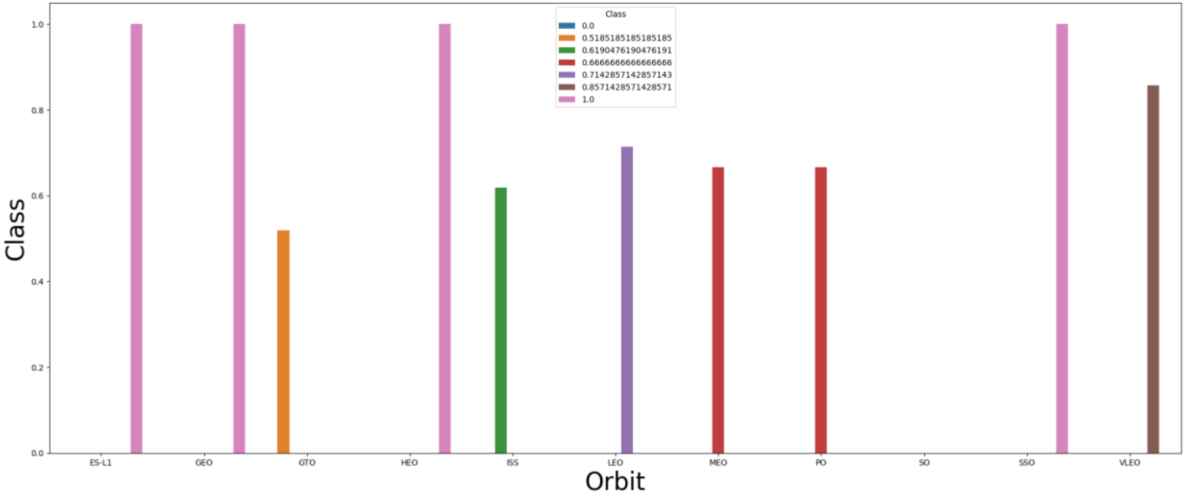


- To facilitate access and Exploratory Data Analysis (EDA), "data wrangling" is the process of organizing and standardizing disparate data sets.
- The number of launches from each site and the frequency with which each orbit type is visited will be determined.
- The information in the outcome column is then used to generate a landing outcome label. This will facilitate additional study, visualization, and machine learning efforts. The final step is to save the data as a CSV file.

EDA with Data Visualization

We plotted charts showing the correlation between launch location, payload, orbit, success rate, flight number and payload mass, as well as the yearly trend in launch success.

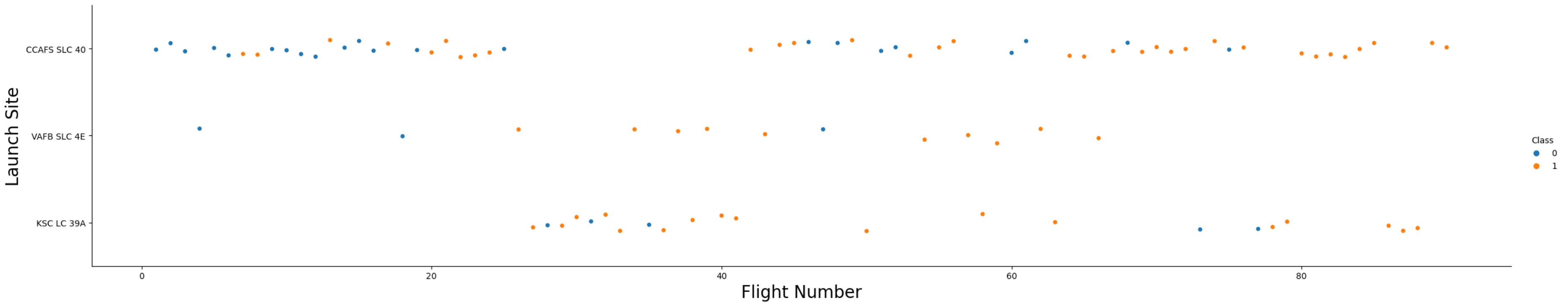




This showed a rising trend from 2013 to 2020. If this trend persists. Success will increase until 1/100%.

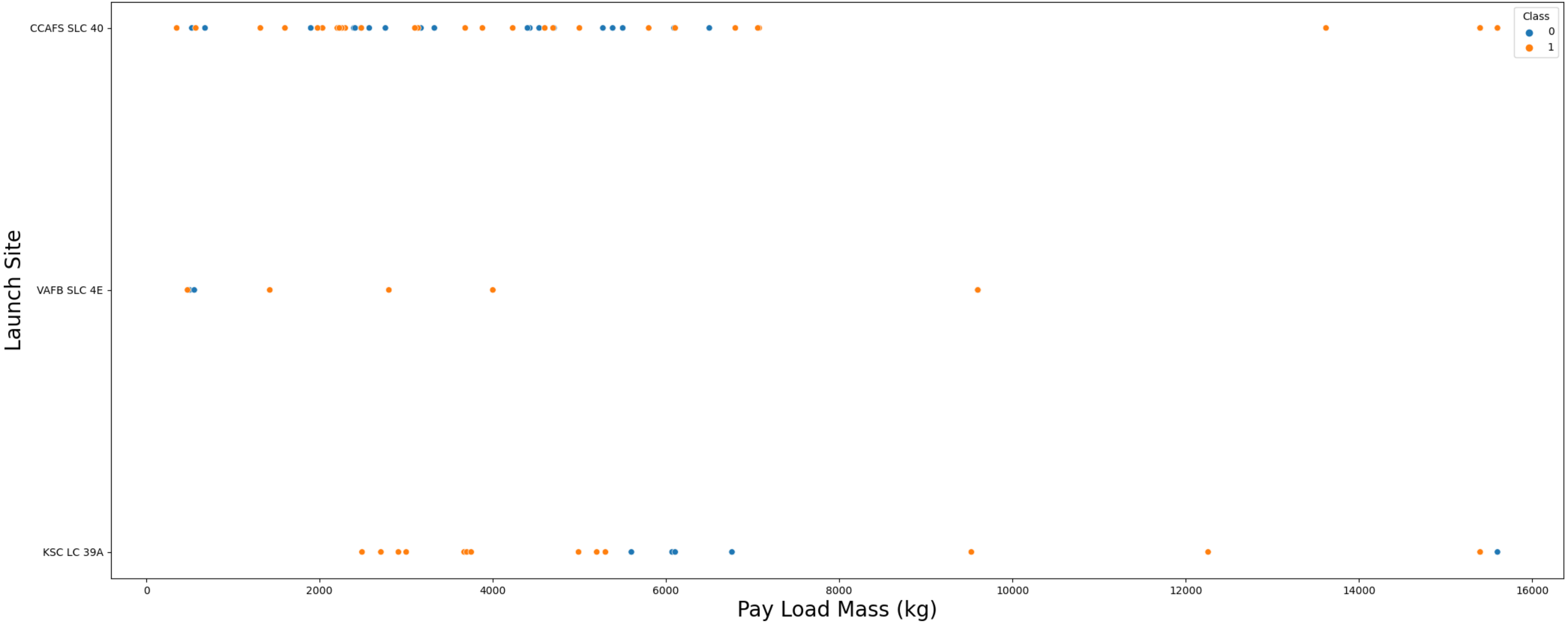
Scatter Plots

Launch Site vs Flight Number



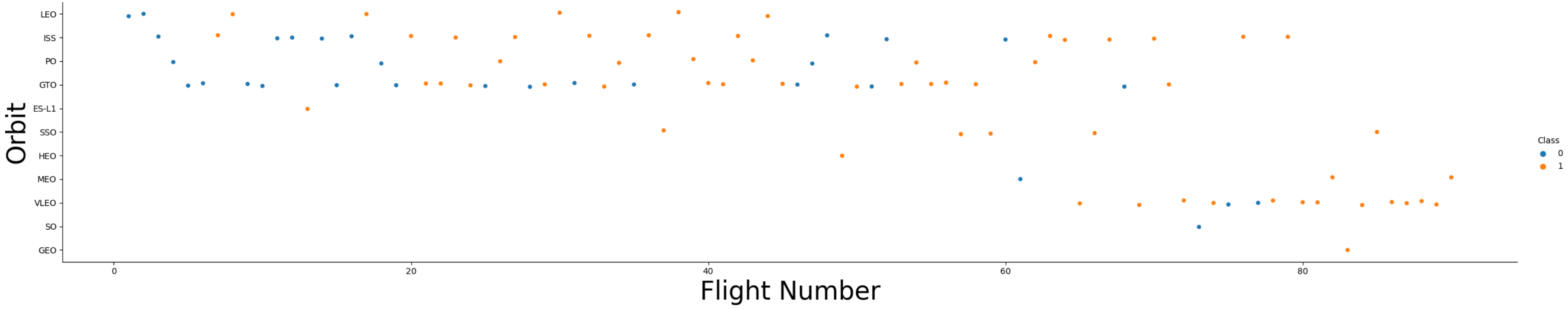
This success rate increases with distance from the launch site, as shown by the scatter plot. The CCAFS SLC40 site, however, exhibits the weakest evidence of this trend.

Launch Site vs Pay Load Mass



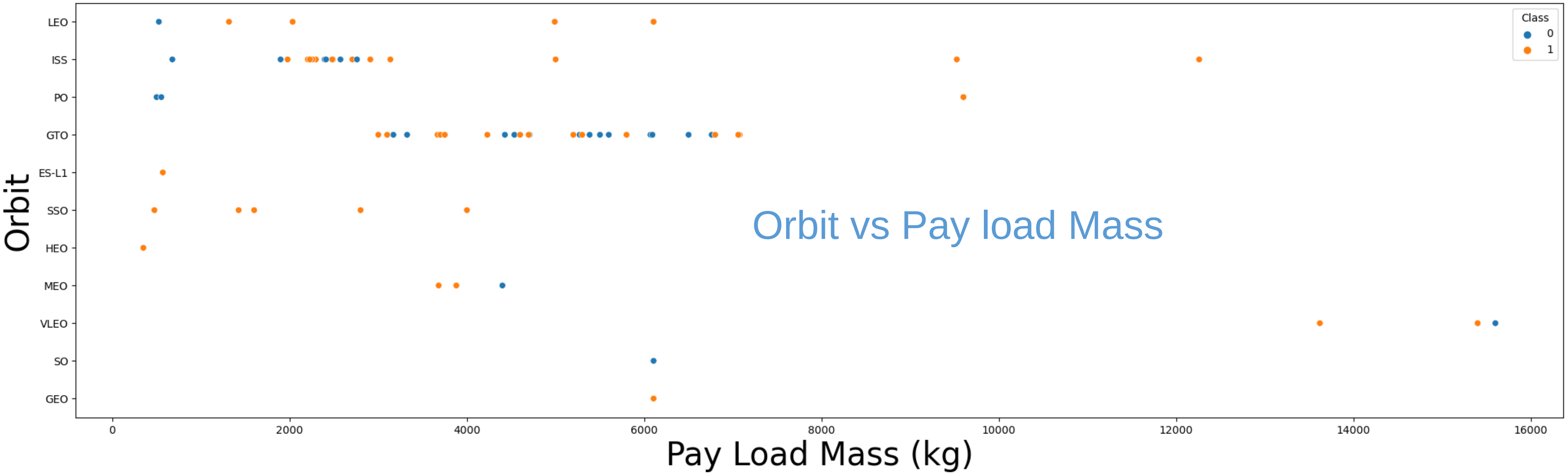
This scatter plot demonstrates that the success rate dramatically improves once the pay load mass is greater than 7,000kg. However, there is no observable correlation between launch location and payload size in terms of success.

Orbit vs Flight Number



While there is no correlation between flight number and success rate for GTO orbits, the scatter plot shows a positive correlation between flight number and success rate for LEO orbits. A single-occurrence orbit should also be left out of the above statement because more data is required.

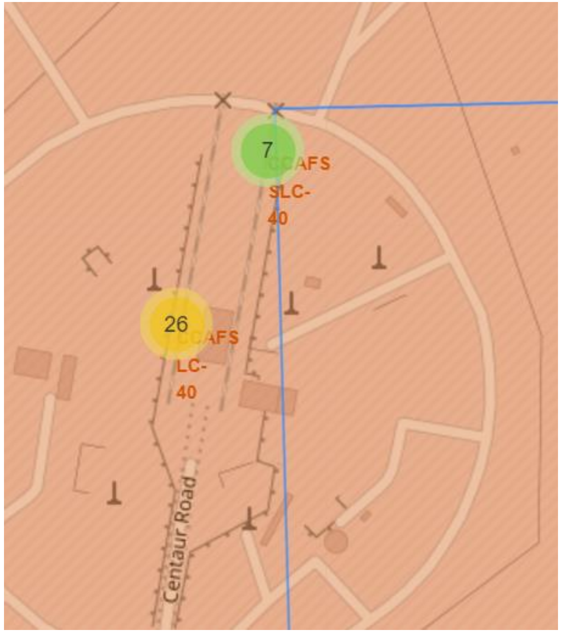
Orbit vs Pay load Mass

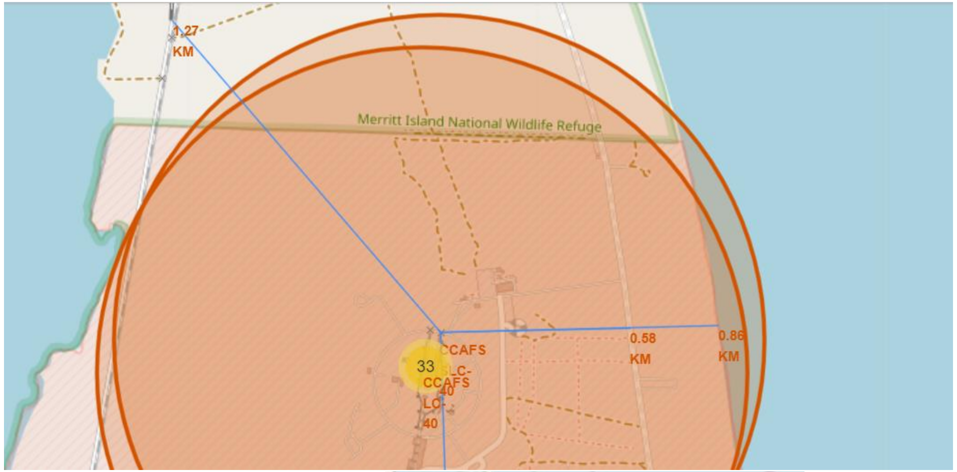


There is a beneficial effect on LEO, ISS, and P0 orbit when a heavier payload is used. However, it has a deleterious effect on medium- and high-Earth orbit. No apparent connection is shown between the two in the GTO orbit. Meanwhile, more data are required for SO, GEO, and HEO orbits before any patterns or trends can be discerned.

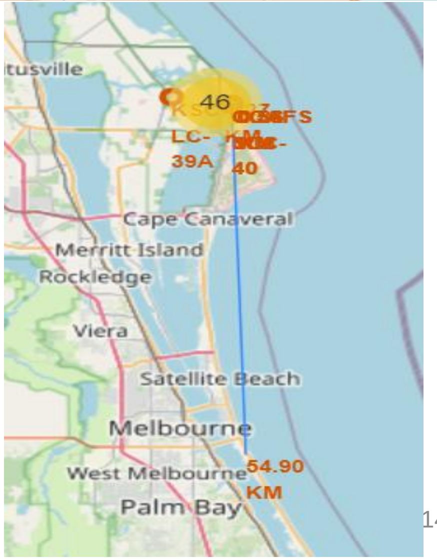
SpaceX Launch Sites

Florida and California are home to SpaceX's launch pads. VAFB SLC 4E is the only launch site in California; the others are in Florida.





CCAFS SLC-40 Site Data



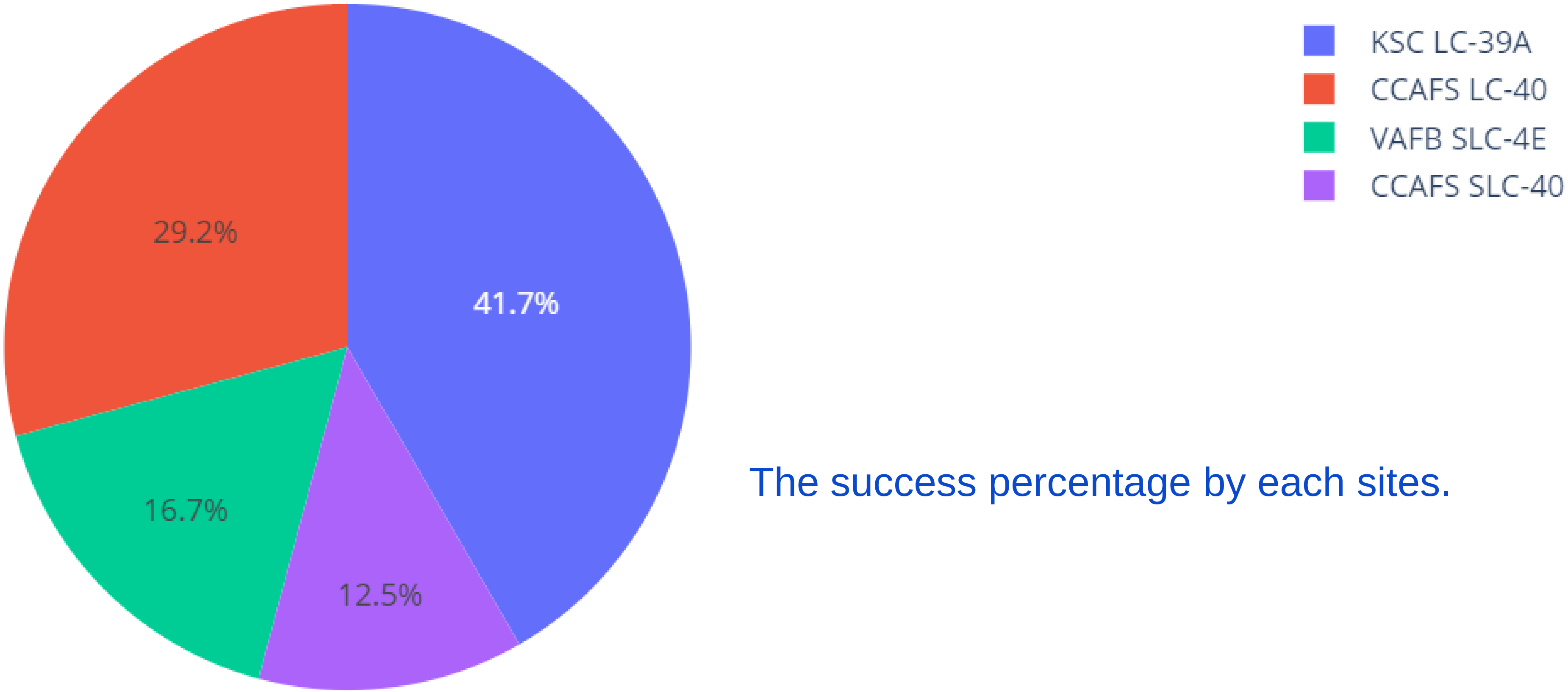
Distance to Closest Coastline –0.86 km

Distance to Closest Highway –0.58 km

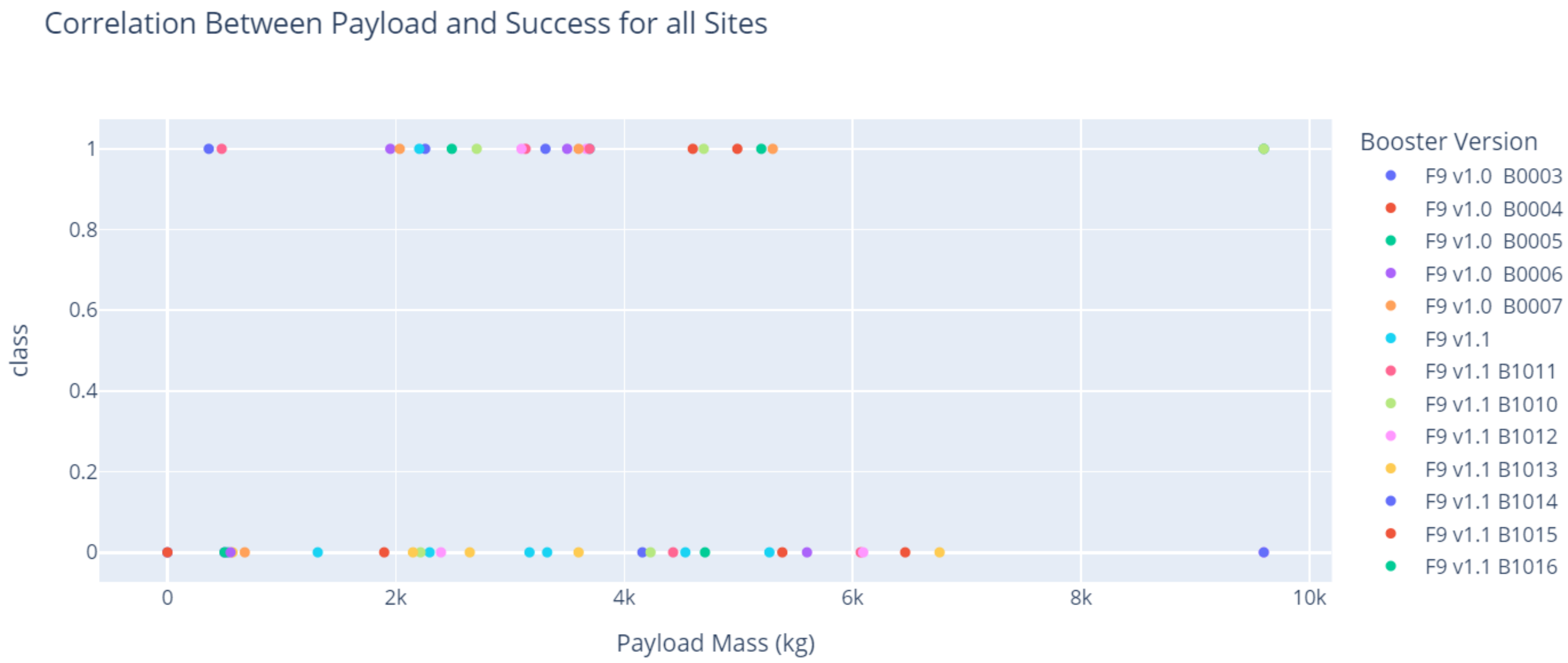
Distance to Closest Railway–1.27 km

Distance to Closest City –54.9 km

Dash board with Ploty Dash



The success percentage by each sites.



The data clearly show that the success rate increases as the payload weight decreases.

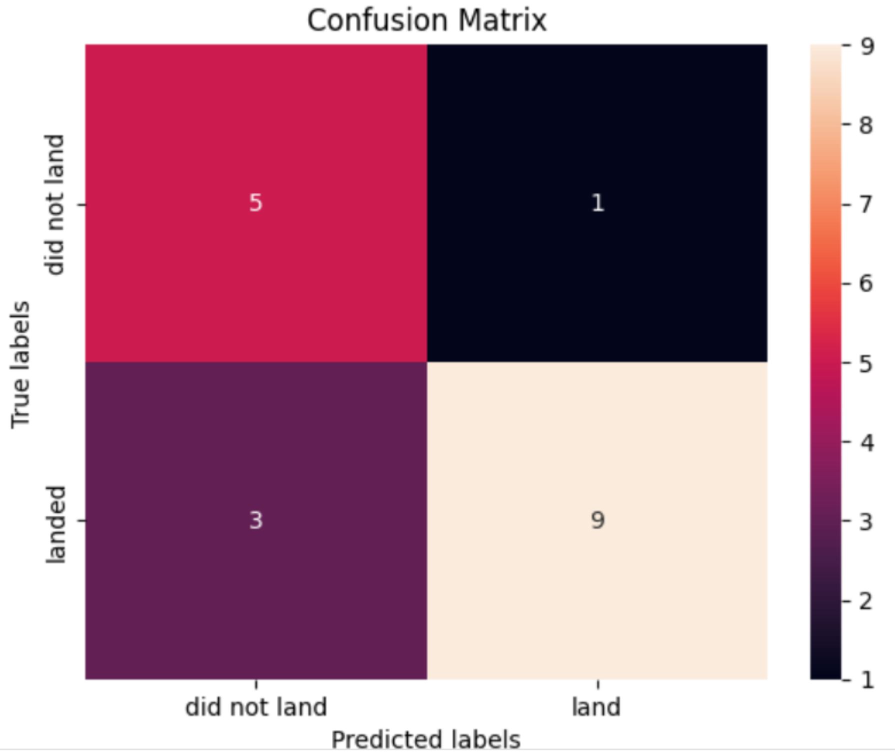
Predictive Analysis

	Model	Accuracy	Score
0	LogisticRegression()	0.8464285714285713	0.8333333333333334
1	SVC()	0.8482142857142856	0.8333333333333334
2	DecisionTreeClassifier()	0.8892857142857145	0.7777777777777778
3	KNeighborsClassifier()	0.8482142857142858	0.8333333333333334

We can see that the decision tree classifier has the highest classification accuracy of 0.889

Confusion Matrix

A clear separation between classes is demonstrated by the confusion matrix for the decision tree classifier. False positives are a major issue. That is, the classifier mistakenly interpreted a landing attempt that failed as a successful one.



Conclusion

From this, we learn that:

- The Tree Classifier Algorithm is the optimal Machine Learning strategy for this dataset.
- Lighter payloads (defined as 4,000 kilograms or less) fared better than their heavier counterparts.
- The success rate of SpaceX launches has increased steadily since 2013, and this trend is expected to continue until 2020, when the company will have perfected its launch procedures.
- Among launch sites, KSC LC-39A has the highest success rate of 76.9%.
- The success rate for SSO orbits is both high (100%) and frequent (multiple occurrences).