IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

A. Genin
April 2025

Outline

- Executive Summary
- Introduction
- Methodology
  - EDA
- Results
  - Improvements
- Conclusions

Executive Summary

In this capstone we have examined the success rates of SpaceX first stage rockets landing. This feature of the SpaceX rockets is a major factor in their pricing structure compared to other rocket manufacturers. Through the use of the SpaceX API and several open source libraries the following analysis was done showing their success rates based on different orbit types as well as looking at how they are improving and where they may still have an inability to recover the first stage. Several proposals for future research were made should an organization be interested in purchasing one of these rockets as well as a model with 83.33% accuracy score was created to predict the circumstances of first stage landing success.

Introduction

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of $62 million dollars; other providers cost upward of $165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine if the cost savings are true. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

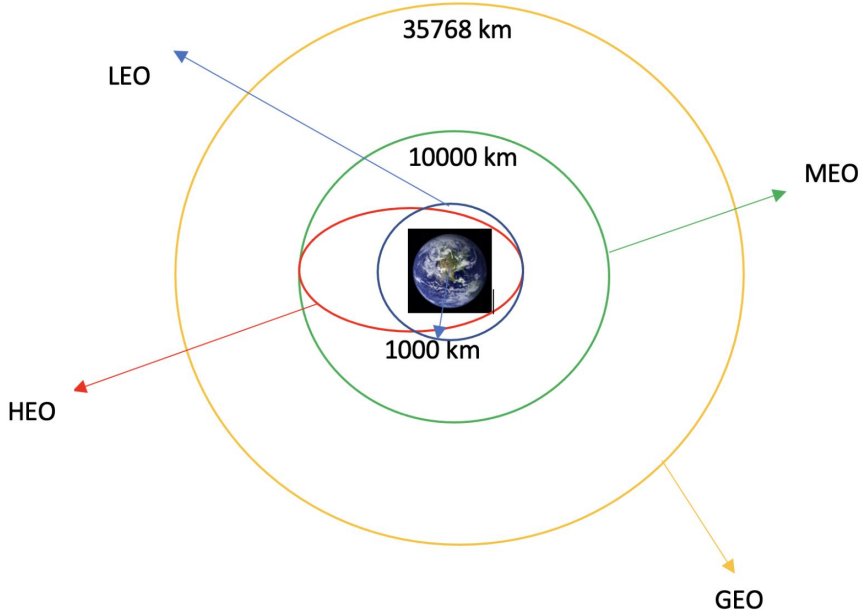
Methodology

Data Sourcing

The data was compiled via the SpaceX API and allows us to examine variables like first stage success rates, payloads used, launch locations, and the order of the launches. Both SQL and Python was used to format the data in effective storage methods.

Types of Orbit

- LEO: Low Earth orbit
- VLEO: Very Low Earth Orbit
- MEO: Medium Earth Orbit
- HEO: High Earth Orbit
- SSO: Sun Synchronous Orbit
- ISS: International Space Station
- GEO: Circular Geosynchronous Orbit



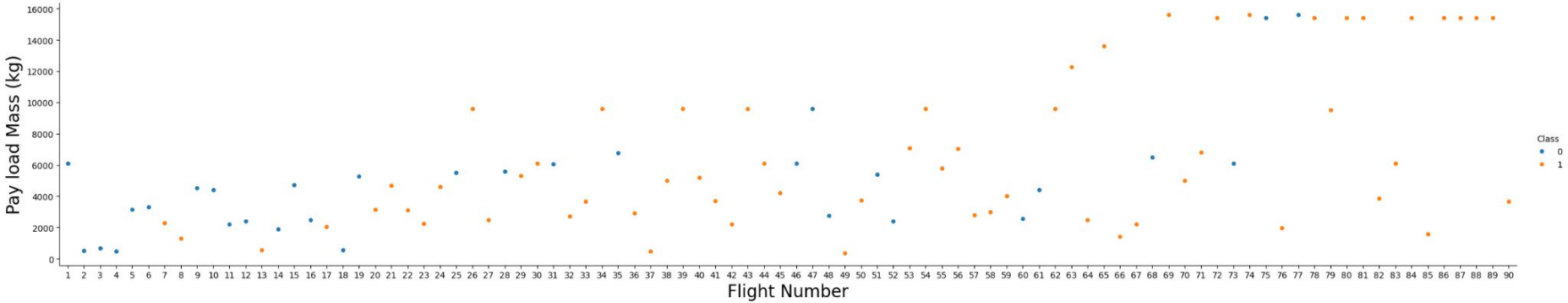
Transformation and Visualization

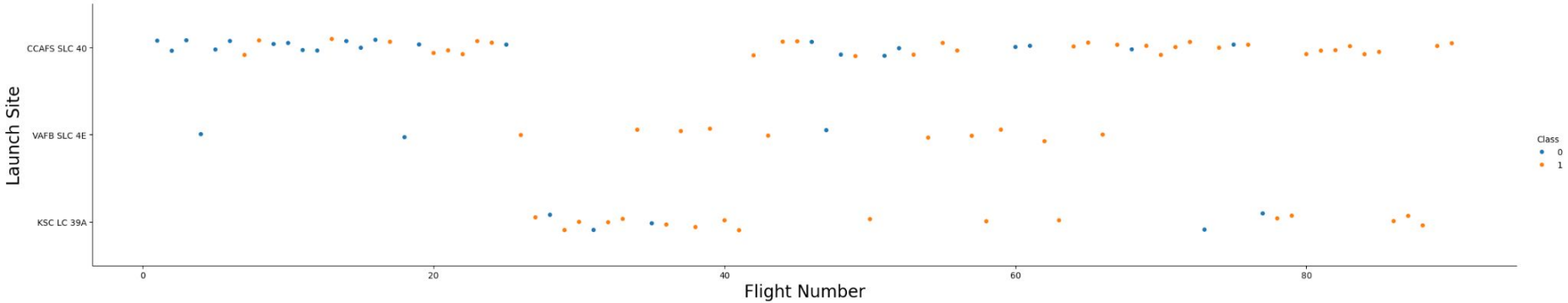
This data was transformed with a handful of open source libraries, including pandas and one hot encoder. These allowed for numerous data wrangling solutions that allowed for the final analysis and visualization.

Using the transformed data, visualizations created using scikit learn for machine learning, seaborn for graphing, and folium for interactive geographic visualizations.

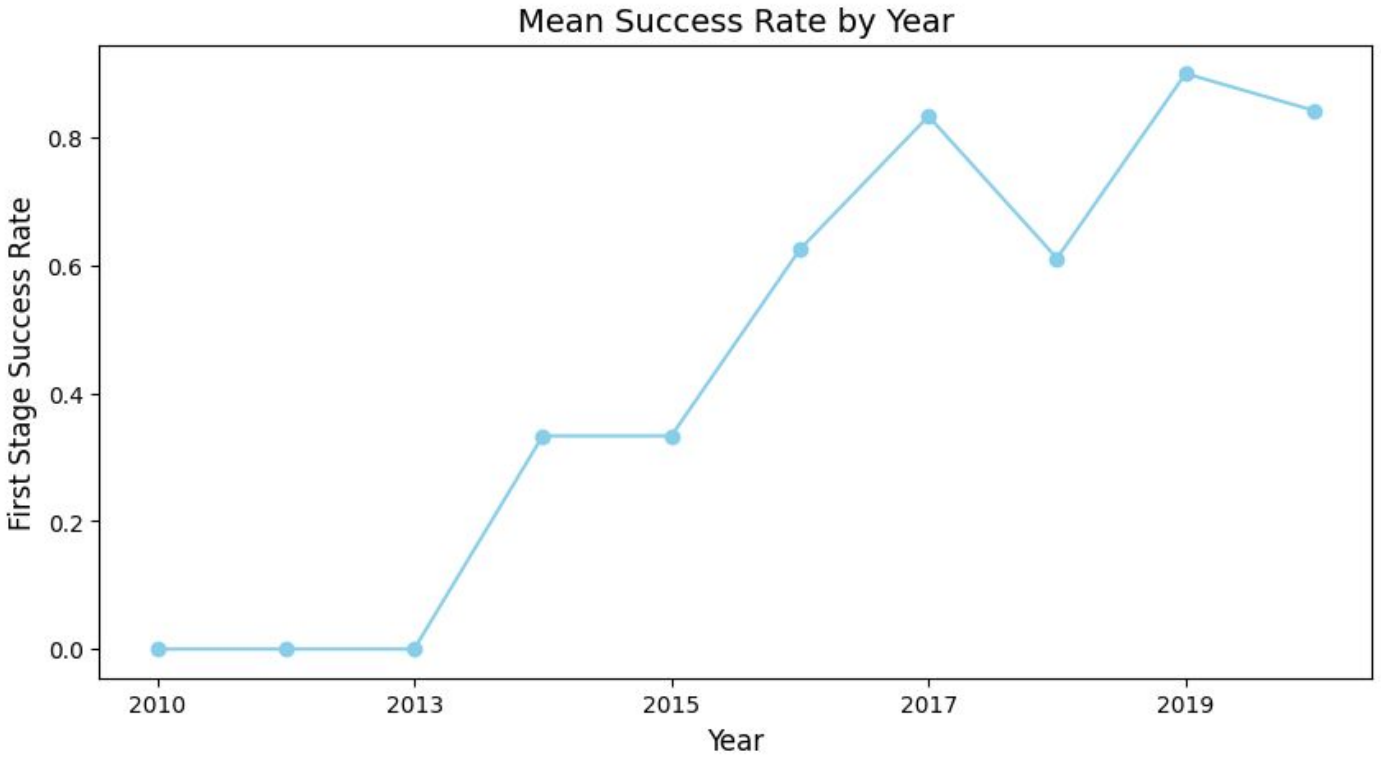
Exploratory Data Analysis

Flight Progression EDA

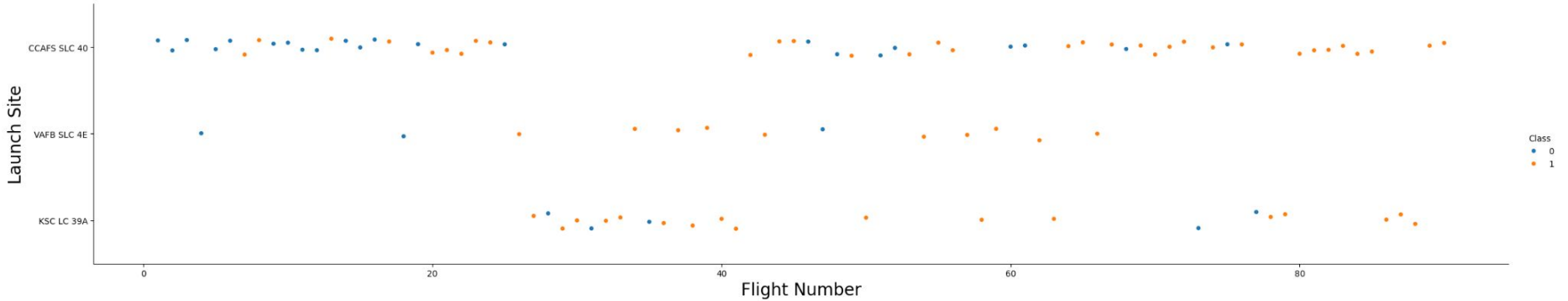


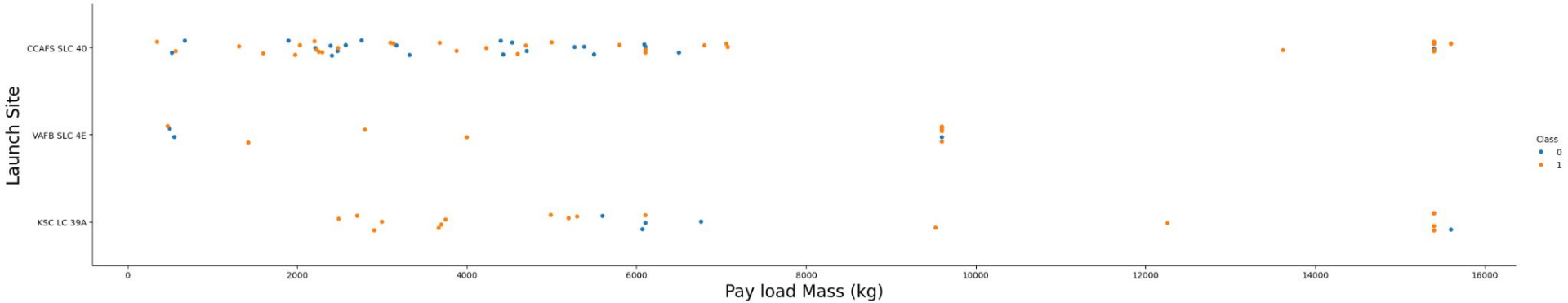


Flight Progression EDA

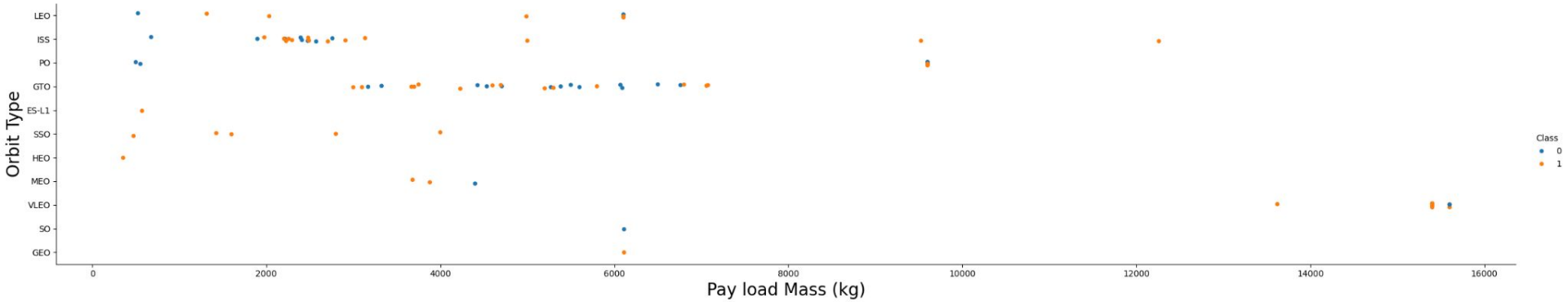


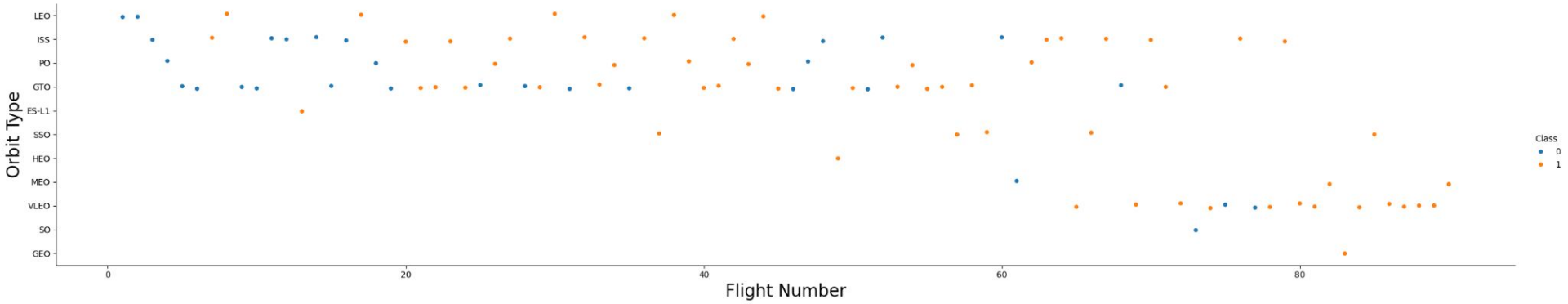
Launch Site EDA





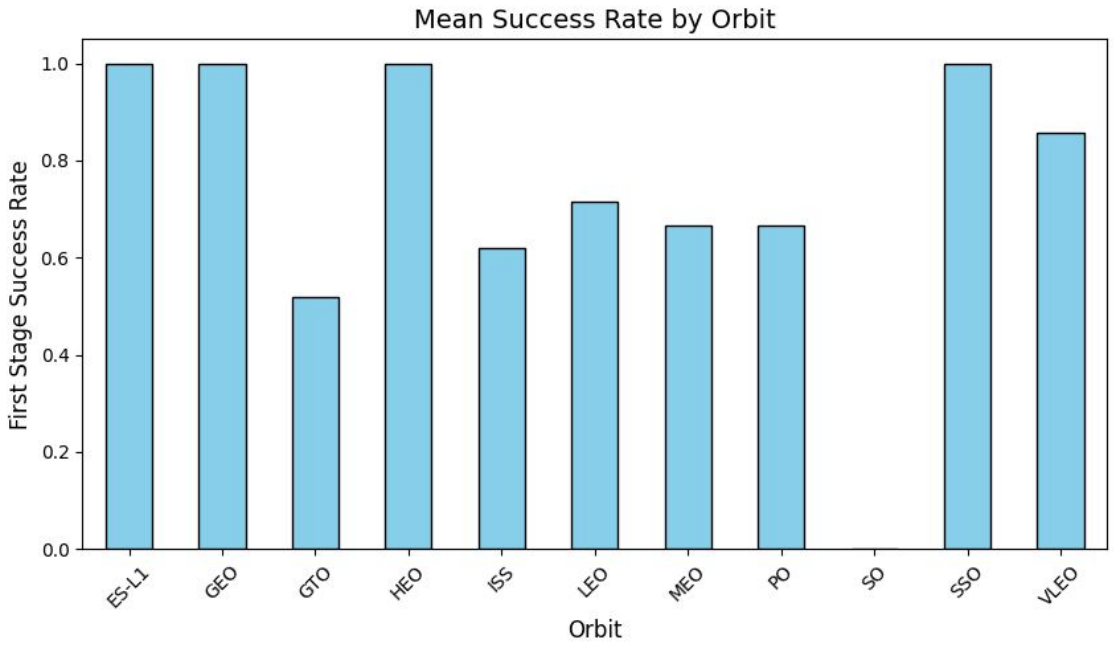
Orbit Type EDA





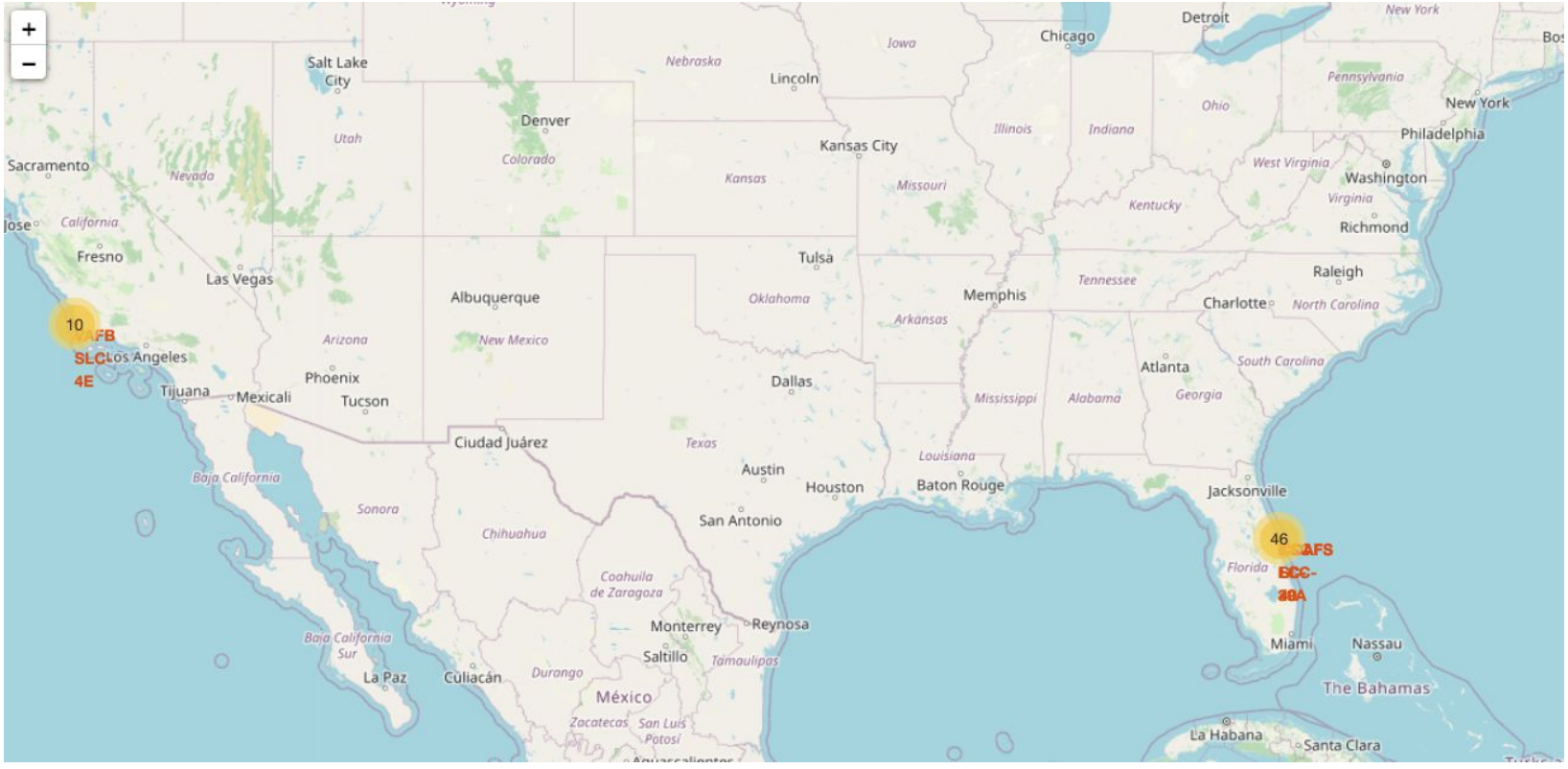
Orbit Type Success Rate

From here we can see that the rockets in question have high first stage landing success rates overall with near perfect success for several orbit types. A consumer could reasonably read this and see potential savings.



Results

Launch Sites with Folium

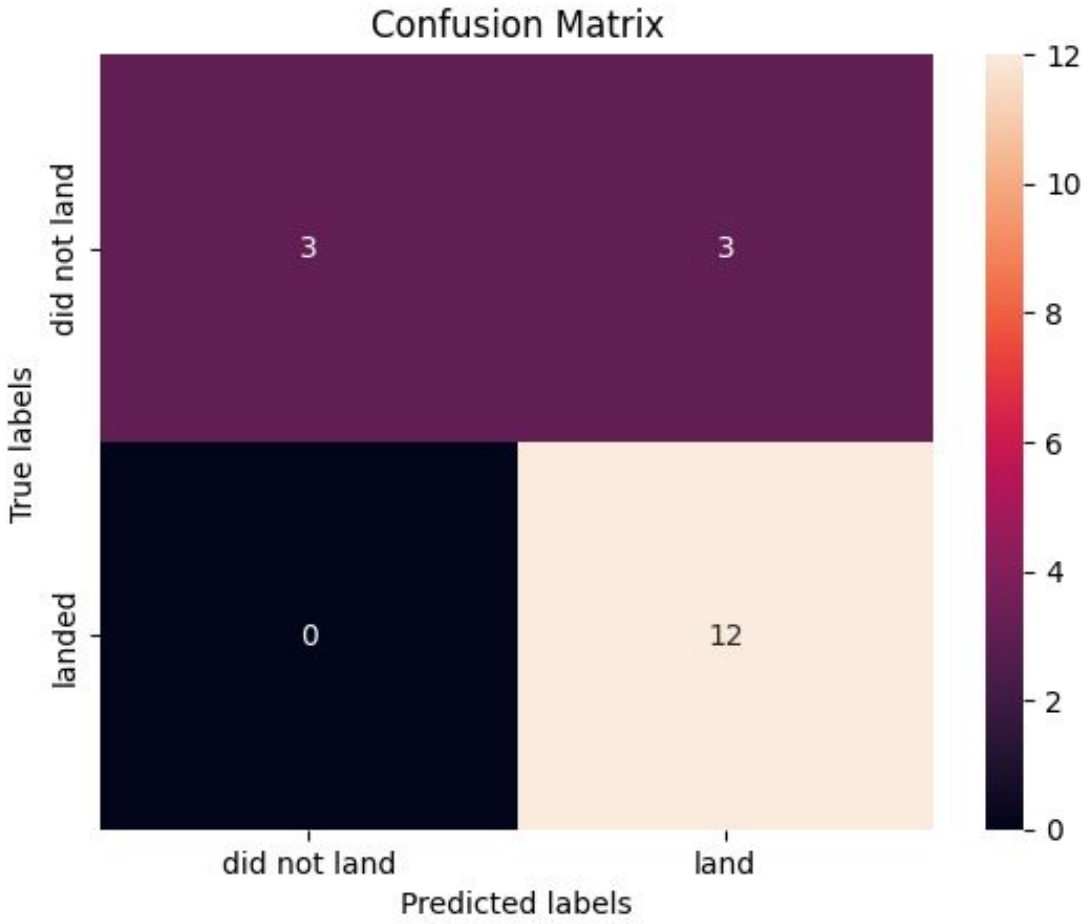


Launch Sites with Folium

By using Folium we can visualize the launch sites as clusters of launches and exam they specific success rate. In this case we see this Florida based SpaceX launch site on a folium map.

Predictive Model Performance

No model improved better than the following confusion matrix, which had a score of .8333 repeating. This was achieved with both the SVM algorithm and the Logistic regression algorithm. It should be noted that some algorithms had class imbalance issues in there **inability to correctly predict true did not land instances**.



Improvements

- As time progresses a larger n value for rocket launches will invariably improve predictive model accuracy and hopefully performance.
- Accounting for landing types used could also be an important factor to determine whether the landing was a success.
- We could use the dates and location data to try and examine if there were a correlation between negative outcomes and weather.
- We could also look at the later failures and see if they could help make a model of the amount of reuses a first stage could see. That information could improve a customer’s ability to determine how much the SpaceX savings truly are.

Conclusions

Conclusions

We have examined the success rates of the SpaceX first stage rockets and in doing so verified the increasing success rate in their landings. The results show that the cost savings do in fact exist with progress continuing over time. We also can see that the launch sites appear to be relatively specialized. Although certain orbit types do not have the success rates one may want to see if they were looking to invest, the near perfect success rates for other desired orbit types also exist.