SPACEX Falcon 9

A study to assess Falcon 9 efficiency by reusing first stage

- Introduction
- Data collection
- Data wrangling
- Exploratory data analysis using SQL
- Exploratory data analysis using pandas and matplotlib
- Visualization with Folium
- Dashboard with Ploty dash
- Machine learning to determine successful landing
- Conclustion

Introduction

In this study, I am trying to 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 the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. This prediction is done using all the techniques like data collection, data wrangling, visualization, machine learning etc, that was studied during the entire course

Data Collection

Data collection was done in two methods

1. Data collection using API :

Data collection was done by sending get request to the SpaceX API and then the data was cleaned for further use

2. Data collection using web-scraping:

Web scraping was done using Beautiful soap package to collect Falcon 9 historical launch records from a Wikipedia page titled “List of Falcon 9 and Falcon Heavy launches”

Data Wrangling

In the data set, there are several different cases where the booster did not land successfully and sometimes it did, these outcomes were converted into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful using exploratory data analysis

Exploratory data analysis using pandas and matplotlib

- Exploratory Data Analysis and Feature Engineering were performed using Pandas and Matplotlib to prepare data for downstream analysis.
- This was mainly done to find the correlation between launch sites and success rates

Visualization with Folium

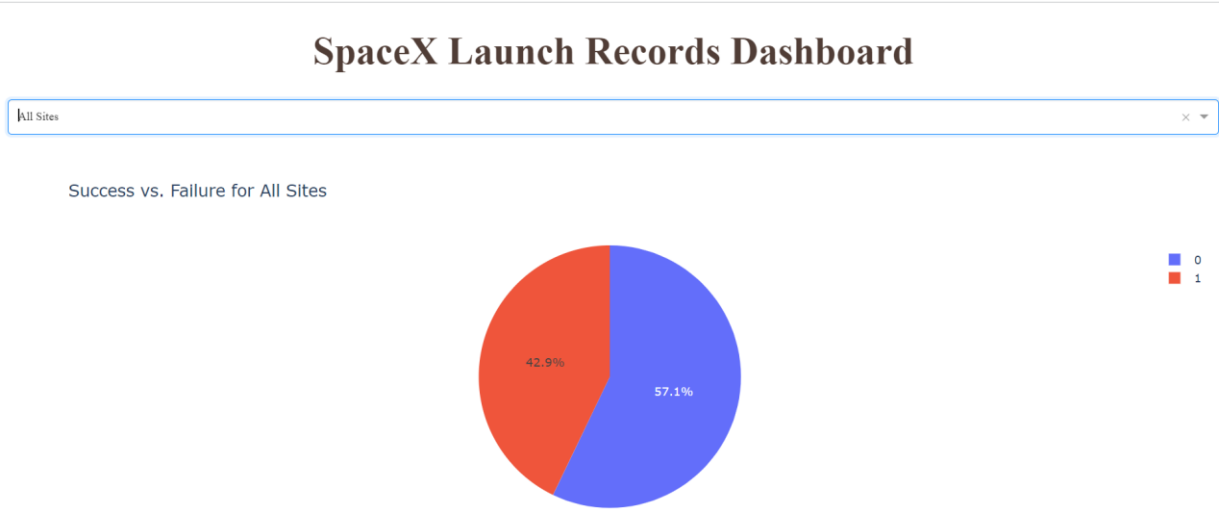
The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site. More interactive visual analytics using Folium were performed to:

1. Mark all launch sites on a map
2. Mark the success/failed launches for each site on the map
3. Calculate the distances between a launch site to its proximities

This was done to find some geographical patterns about launch sites.

Dashboard with Ploty dash

Plotly Dash application was created for users to perform interactive visual analytics on SpaceX launch data in real-time. This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart



Machine learning to determine successful landing

The final step was creating a machine learning pipeline to predict if the first stage will land given the data from the preceding studies.

Exploratory Data Analysis was performed to determine Training Labels by:

1. create a column for the class
2. Standardize the data
3. Split into training data and test data

Finally best Hyperparameter for SVM, Classification Trees and Logistic Regression were calculated and the method that performs best using test data was determined

Conclusion

- From the dashboard it was clear that Lauch site “KSC LC-39A” had the most successful launches with 76.9%

- F9 Booster version FT has the highest launch success rate with payload mass of 5300kg