SpaceX

Re-use the first stage of Launch

Re-land the Rocket to be Use on net Mission

Muhammad Hammad
27-Apr-23

Contents

Introduction

SpaceX is a revolutionary company who has disrupt the space industry by offering a rocket launches specifically Falcon 9 as low as 62 million dollars; while other providers cost upward of 165 million dollar each. Most of this saving thanks to SpaceX Outstanding idea to reuse the first stage of the launch by re-lands the rocket to be used on the next mission. Repeating this process will make the price down even further. As a data scientist of a startup rivaling SpaceX, the goal of this project is to create the machine learning pipeline to predict the landing outcome of the first stage in the future. This project is crucial in identifying the right price to bid against SpaceX for a rocket Launch.

Basic Problems

Identifying all the factors that influence the landing outcomes

The Relationship among the Variables and how it affects outcomes

The Best Condition needed to Increase the probability of successful landing.

# Methodology

Data Collection

Data Was Collected using SpaceX REST API and web scrapping form Wikipedia.

For REST API, we send request to the server, then, we get the response content as Json and we convert it in to pandas dataframe using json_nomalize(). We than clean the data, checked for missing value and fill with whatever needed.

For web scrapping, we will use the BeautifulSoup to extract the launch records as HTML Tables, and then we convert the tables in to pandas DataFrame for further analysis.

Data Wrangling

Data wrangling is the process of cleaning and unifying messy and complex data into simple and readable form for Exploratory Data Analysis.
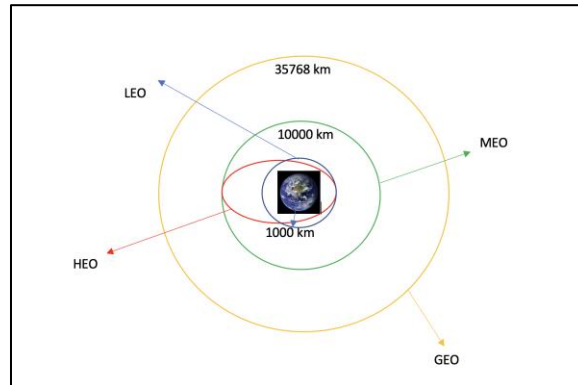

Data Visualization

We start by using scatter graph to find the relationship among the attributes such as:

- Payload and flight Number.
- Flight Number and Launch Site.
- Payload and launch Site.
- Flight Number and Orbit Type.
- Orbit type and Payload Mass.

Scatter plot shows the dependence between these attribute among them. Once the relation is determine in the graph. It is very easy to see which attribute effecting the most to the success of the landing outcome.

After getting the hint of relationship we use more visualization tools such as bar charts and line graphs. Bar graph is one of the easy ways to interpret the relationship. We will use bar graph to find which orbit has high probability to success. Line chart is use to see launch success yearly trend.
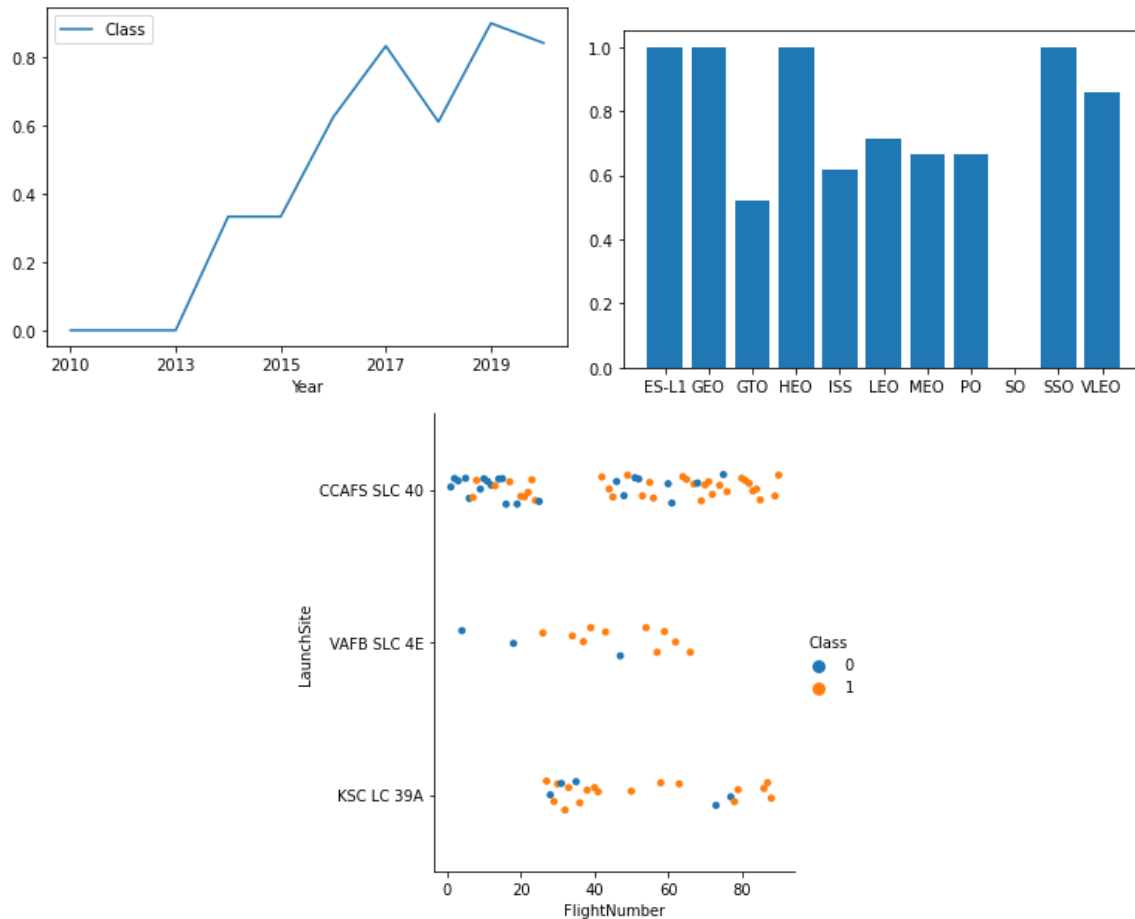
Exploratory Data Analysis (EDA) with SQL

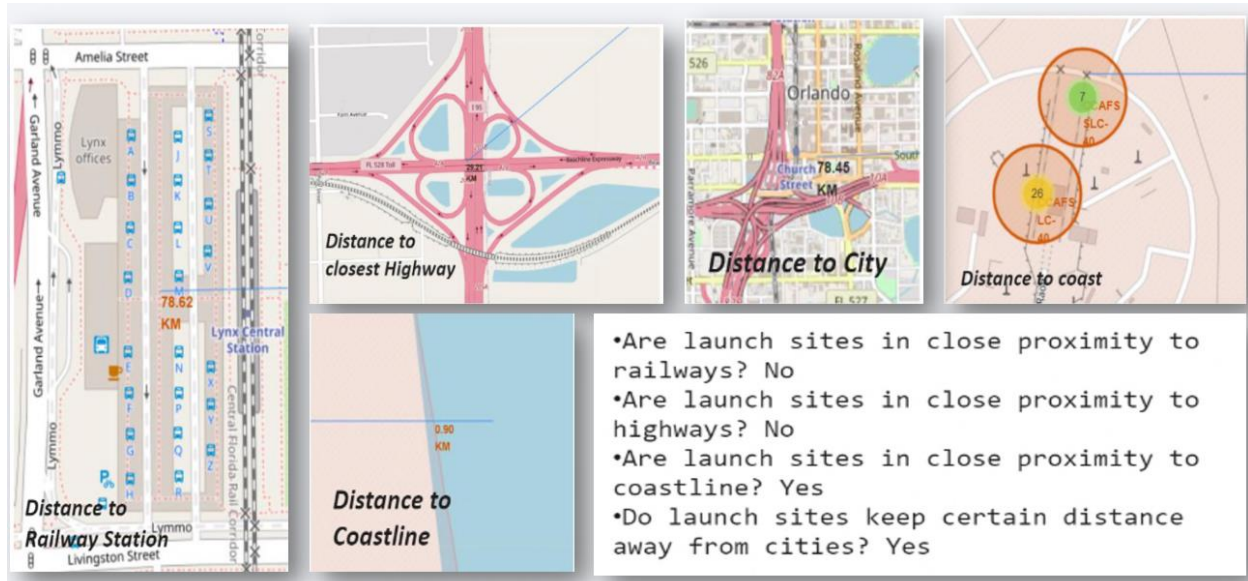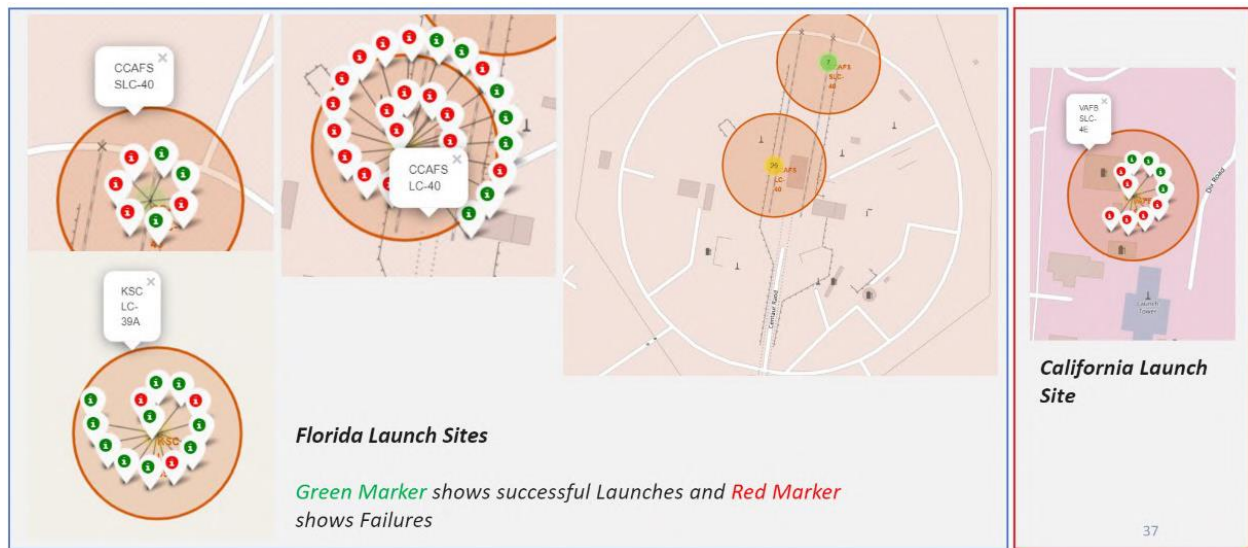We perform many queries of SQL to get better understand the dataset. Example

- Displaying the name of launch sites
- Displaying the total payload mass carried by the booster version F9 v1.1.
- Displaying the average payload mass carried by the booster version F9 v1.1.
- Listing the date when the first successful landing outcome in ground pad was archived.
- Listing the name of boosters which have success in drone ship and payload mass greater than 4000 but lower than 6000.
- Listing of total number of successful and failure mission outcomes.
- Listing the name of boosters which carry the maximum payload masses.

Interactive Map with Folium

To visualize the launch data in to interactive map, we took coordinates at each site and add a circle marker around each site with the label name of launch site.

We assigned launch outcome (Success and failure) with green and red color.

We use Haversine's formula to calculate the distance between launch site and various landmarks such as railways, highways, and coastline and nearby cities.

Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

Distance to closest Highway

Distance to City

Distance to coast

Distance to Railway Station

Distance to Coastline

•Are launch sites in close proximity to railways? No
•Are launch sites in close proximity to highways? No
•Are launch sites in close proximity to coastline? Yes
•Do launch sites keep certain distance away from cities? Yes

Dashboard with Plotly Dash

We build an interactive dashboard with Plotly Dash which allowing the user to play around with the data as they need. We plat a pay chart showing total launches by certain sites. And we plot a scatter graph shows the relation between outcomes and Payload masses (Kg) for different booster version.
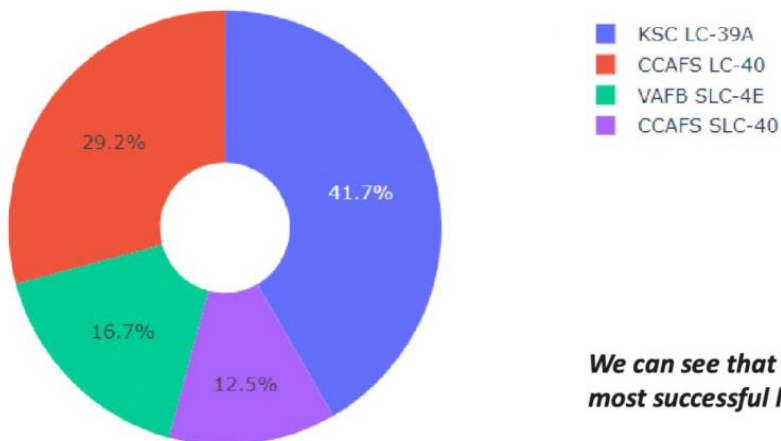
We can see that KSC LC-39A had the most successful launches from all the sites
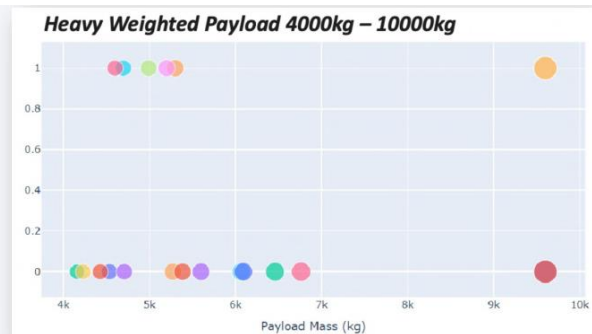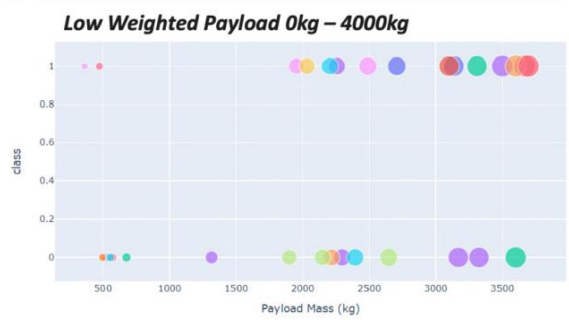
Results

The payload mass is greater than 7000 Kg, the probability of success rate is highly increase. However there is no clear pattern to say that Launch site is dependent on the payload mass for Success rate.

Heavier payload has positive impact on LEO, ISS and P0 orbit. However, it has negative impact on MEO and VLEO orbit. GTO orbit seem to depict no relation between the attributes. Meanwhile, again, SO, GEO and HEO orbit need more dataset to see any pattern or trend.

Increasing trend from the year 2013 until 2020. If this trend continue for the next year onward. The success rate will steadily increase until reaching 1/100% success rate.

Conclusion

- The Tree Classifier Algorithm is the best machine learning approach for this data.
- The low weighted payload (less than 4000 Kg) performs better than heavy weighted payloads.
- The Success rate of SpaceX launches increasing with time, which it eventually perfect the launches in future.
- KSC LC-39A has more successful launches rate i.e. 76.9%.
- SSO orbit has more successful rate; 100% and more then one occurrence.