Rocket Landing Prediction

- Data Science and Machine Learning Project

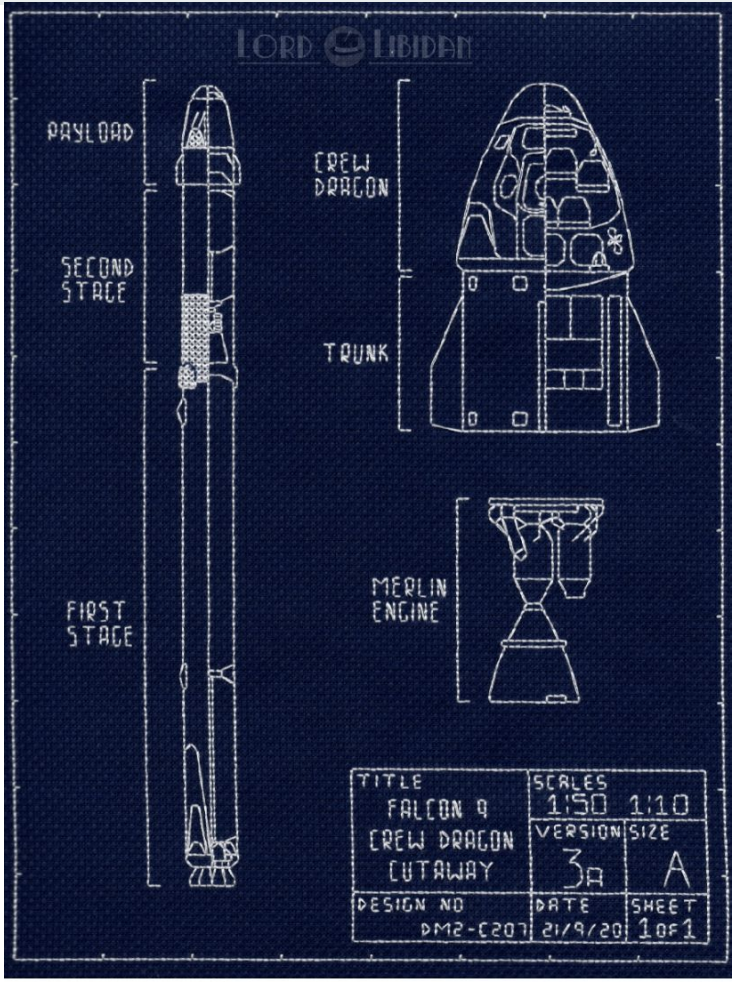
Andriy Fedorych

September 23, 2022

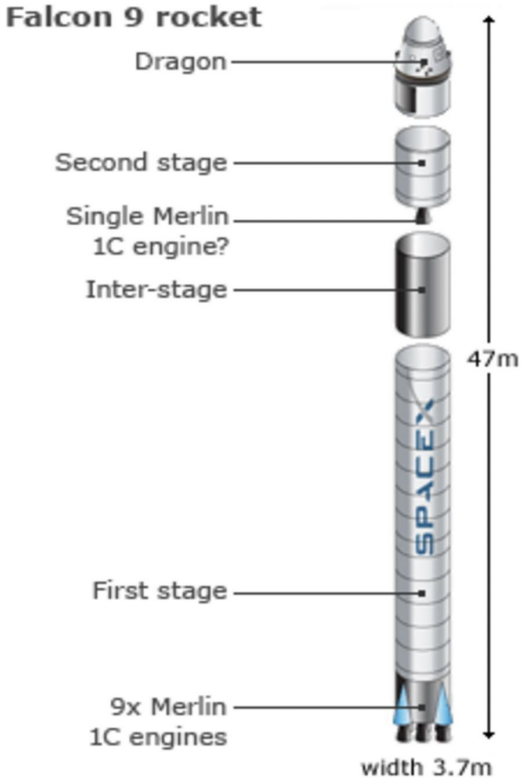
Table of Contents

Executive Summary

1. Summary of methodologies
    - Data collection
    - Data wrangling
    - EDA with data visualization
    - EDA with SQL
    - Building an interactive map with Folium
    - Building a Dashboard with Plotly Dash
    - Predictive analysis (Classification)
2. Summary of all results
    - Exploratory data analysis results
    - Interactive analytics demo in screenshots
    - Predictive analysis results



Business problem



We all know that launching rockets is extremely expensive, and therefore landing the first stage of a rocket, which is the most expensive due to its dimensions, is a significant problem, predicting the success of landing is a significant problem for space technologies

Three main question

1. **Find a correlation between different rocket parameters and successful landing rate**
2. **Separate the best ratios that will guarantee a successful landing and will use as few resources as possible**
3. **Train a model to predict the landing of a launched rocket**

The main steps and phases of the research

Methodology

1. Data Collection
2. Data Wrangling
3. Exploratory Data Analysis
4. Data Visualization
5. Model Development
6. Create a data-driven insights for problem

Data Collection

Web Scraping

- Using BeautifulSoup, data on rocket launches was obtained from Wikipedia
- Extracted all column/variable names from the HTML table header
- Parsed the table and converted it into a Pandas DataFrame

API

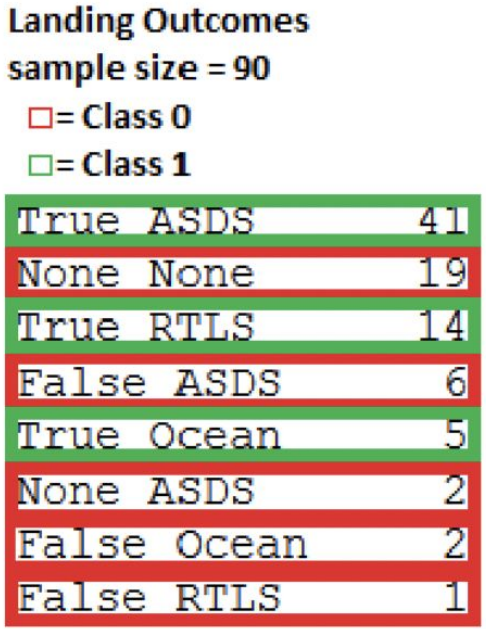
- Acquired historical launch data from Open Source REST API for SpaceX
- Requested and parsed the launch data using the GET request
- Filtered the DataFrame
- Replaced missing values

Orbit	Customer	Launch outcome	Booster landing
LEO	SpaceX	Success	Success (drone ship)
d astronomical observations.[14]			
Sub-orbital[17]	NASA (CTS)[18]	Success	No attempt
m (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[22] First flight of a			
LEO	SpaceX	Success	Success (drone ship)
LEO	SpaceX	Success	Failure (drone ship)
ring the second stage engine twice. The first stage booster failed to land on the drone ship.[26] due to incorrect wind			
LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the			
LEO	SpaceX	Success	Failure (drone ship)
e first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first dual cleaning fluid trapped inside a sensor.[35]			
LEO	SpaceX	Success	Success (drone ship)
2019).[38]			
LEO (ISS)	NASA (CCDev)	Success	Success (drone ship)
live stream was peaked at 4.1 million viewers, while NASA estimated roughly 10 million people watched on various			
LEO	SpaceX	Success	Success (drone ship)

Data Wrangling

- Fully processed datasets, replaced all lost values
- As well as converted values for further use in model training

On the right in the photo you see a new classification of rocket landing statuses



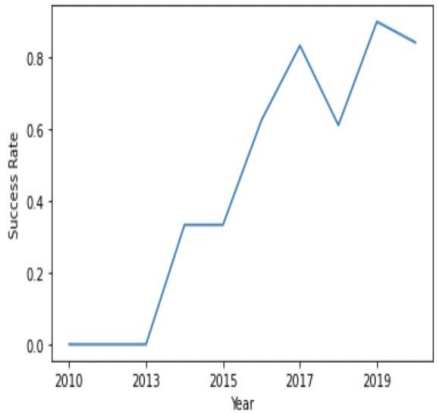
Exploratory Data Analysis

- Read the dataset into a Pandas dataframe
- Used Matplotlib and Seaborn visualization libraries to plot
- Build different types of chart

PayloadMass

In [68…
```
year=[]
def Extract_year(date):
    for i in df["Date"]:
        year.append(i.split("-")[0])
    return year
```

In [69…
```
df.groupby(Extract_year(df['Date'])).mean()['Class'].plot(kind='line', xlabel='Year', ylabel='Success Rate')
plt.show()
```



Scatter chart - shows how much one variable is affected by another. This plot is generally composed of large data bodies.

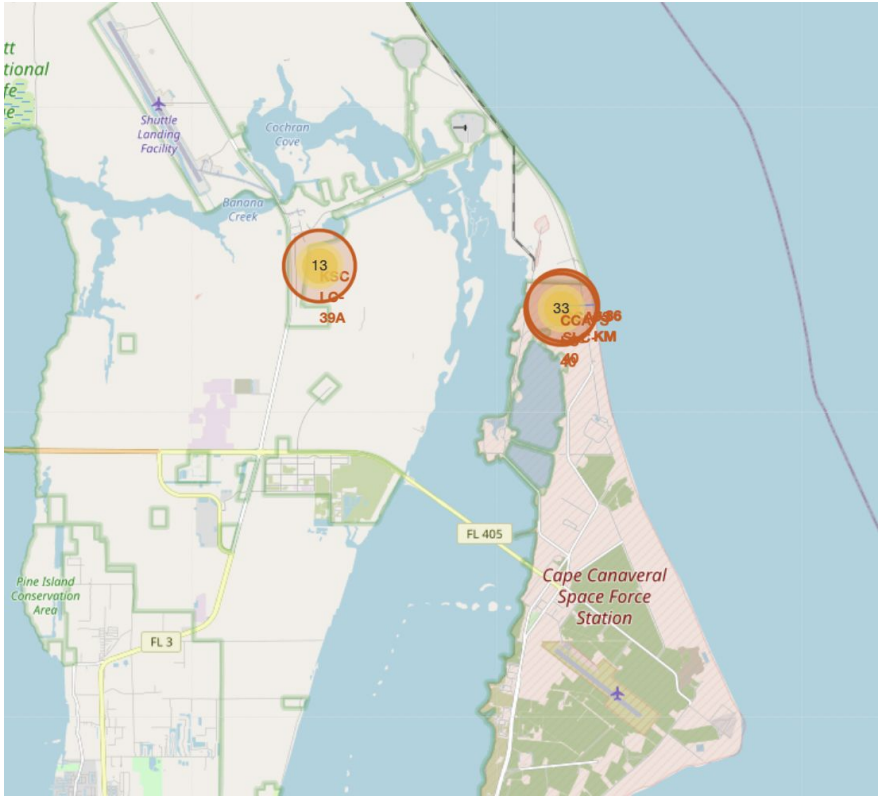
Bar chart makes it easy to compare datasets between multiple groups at a glance.

Line chart shows data variables and trends very clearly and helps predict the results of
data that has not yet been recorded.

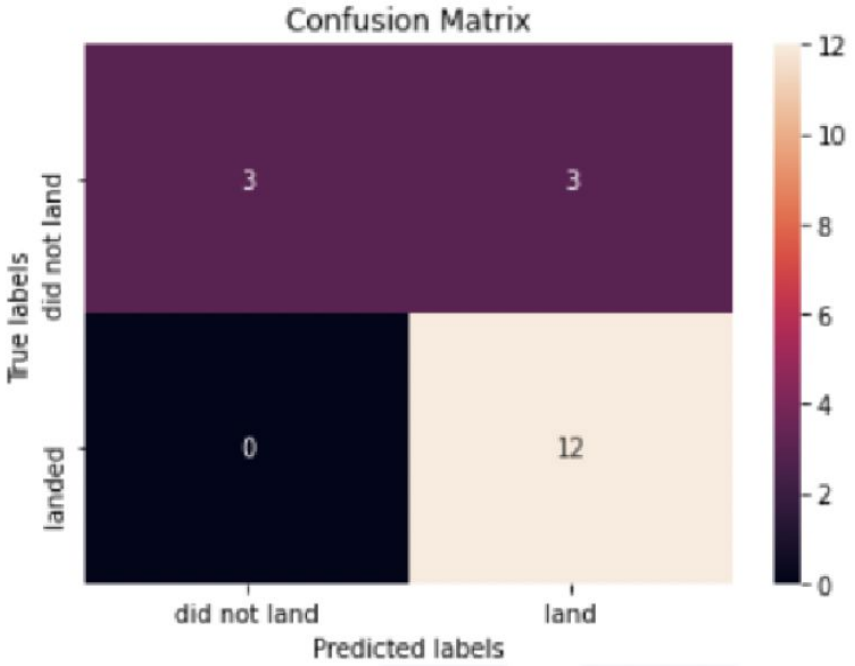
Data Visualization

- **Launch Records Dashboard -** *Used Python interactive dashboarding library called Plotly Dash to enable to explore and manipulate data in an interactive and real-time*

- **Launch Sites Location Analysis -** *Used Python interactive mapping library called Folium*



Model Development



- Find best Hyperparameter for SVM,Classification Trees and Logistic Regression
- Perform exploratory Data Analysis and determine Training Labels

Results

- **The most optimal parameters for building rockets have been found**
- Came to the conclusion that **the heavier the first stage** of the rocket, the **greater the chance of its landing**
- Converted all types of values and trained a model that can predict the probability of successful landings with an **accuracy** of **83.3%**

The vast majority of insights are not presented in the introductory presentation, for this view Jupyter notebooks