Data Science and Machine Learning Capstone Project

- **Space X Falcon 9 First Stage Landing Prediction**
- **Juan Mauricio Padilla Villarreal**

Summary

- Introduction
- Data Collection
- Data Wrangling
- Exploratory Data Analysis Visual Analytics
- Predictive Analysis
- Exploratory Data Analysis Visualization
- Exploratory Data Analysis with SQL
- Follium Maps
- Plotly Dash
- Predictive Analysis
- Conclusion

Introduction

- SpaceX, has been revolutionizing the spaceflight industry with its cutting-edge technologies and innovative solutions. The company has achieved several milestones over the years, including the first privately-funded liquid-propellant rocket to reach orbit, the first privately-funded spacecraft to dock with the International Space Station, and the first privately-funded spacecraft to send humans to space.
- In this research exercise , we will explore and quantify the advancements and the facts involved by SpaceX in the field of spaceflight and how they have improved the performance of their rockets under a statistical and data science scope.
- Through this research, we hope to provide a comprehensive overview of SpaceX's contributions to the field of spaceflight and showcase the power of data science and statistical analysis in evaluating the performance of space technologies.

Data Collection

The Data was collected from the web using the Wikipedia´s SpaceX information page.

Then the data was organized into data frames using json.

Specifically the data was taken from HTML table and adapt to a data frame.

Checked for missing values.

Space X API

- Used the SpaceX API to get data.
- Wrangling and formatting was needed.

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_l
```

Next, request the HTML page from the above URL and get a `response` object

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
# use requests.get() method with the provided static_url
# assign the response to a object
data  = requests.get(static_url).text
```

Create a `BeautifulSoup` object from the HTML `response`

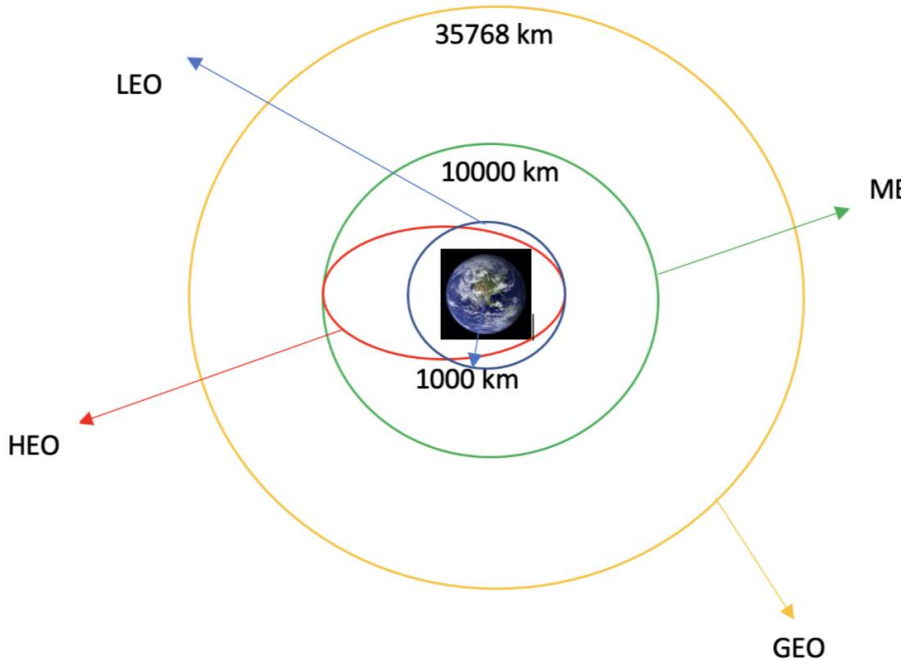
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
import requests
from bs4 import BeautifulSoup

url = "https://www.spacex.com/launches/"
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')

print(soup.prettify())
```

```
<!DOCTYPE html>
<!--
```

Data Wrangling

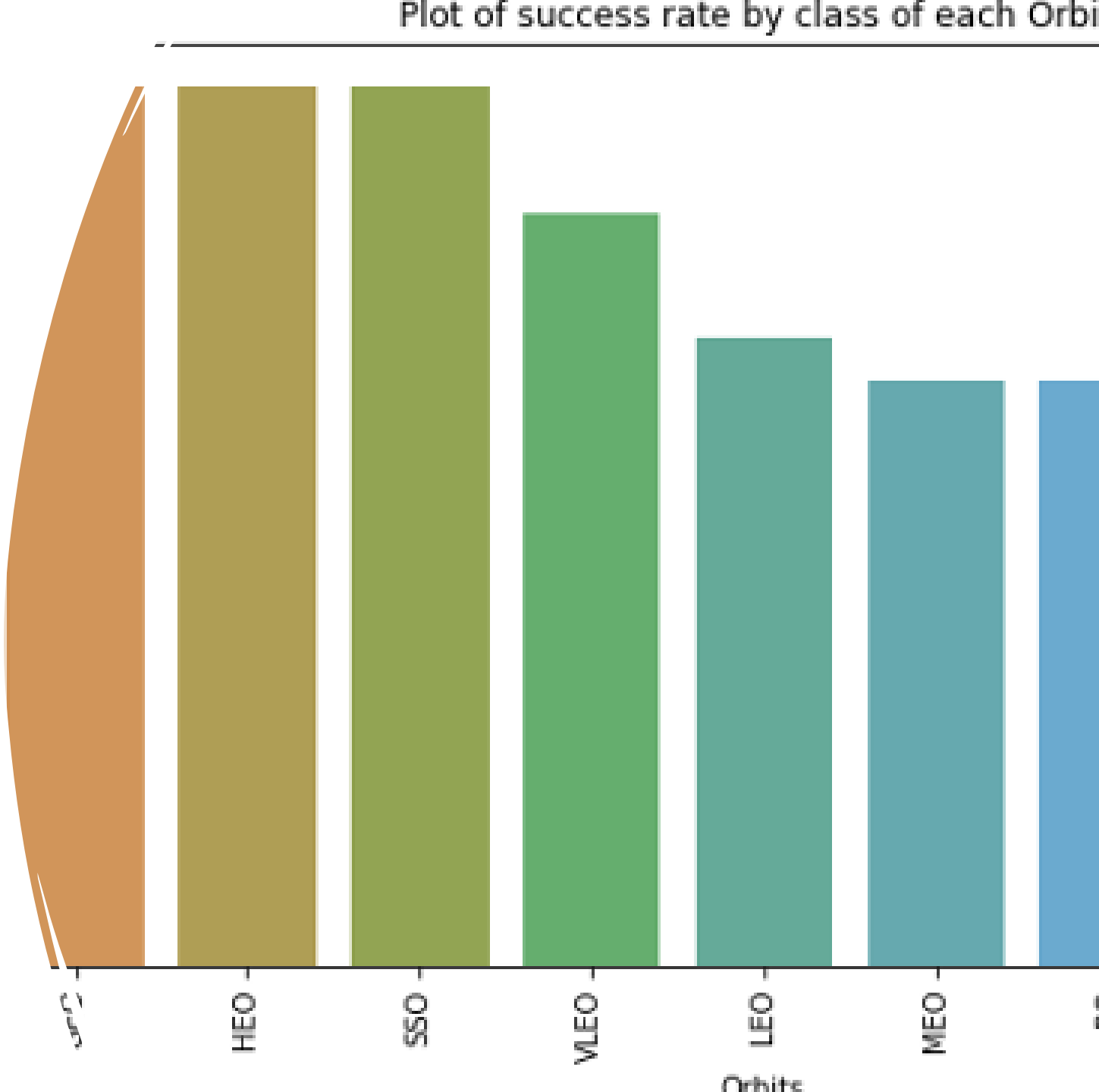
- Was performed some Exploratory Data Analysis and determine the training labes.

- Determinated the number of launches from each site and checked the outcome of each space flight and the landing success rate.



EDA Data Visualization

- Many variables were analyzed to obtain a tendency in the outcome related to each factor



EDA with SQL

SpaceX dataset was loaded into a PosgreSQL database

Were obtain the data of the failed landings in the drone ship.

Total successful and failure results.

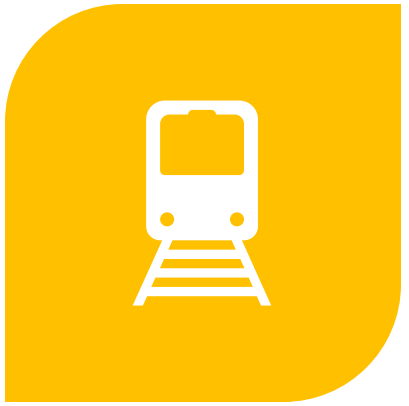
Average payload mass carried.

Unique launch sites.

Build a Folium Map

ADDED LAUNCH SITES TO THE MAP.

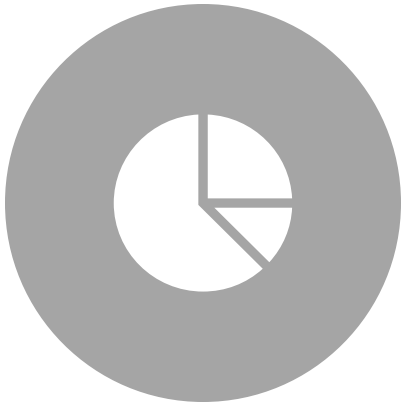
ASSIGNED LABELS TO EACH LAUNCH SITE.



CALCULATED DISTANCES FROM THE SITES TO SOME FEATURES AS RAILROADS ETC.

Ploty Dash

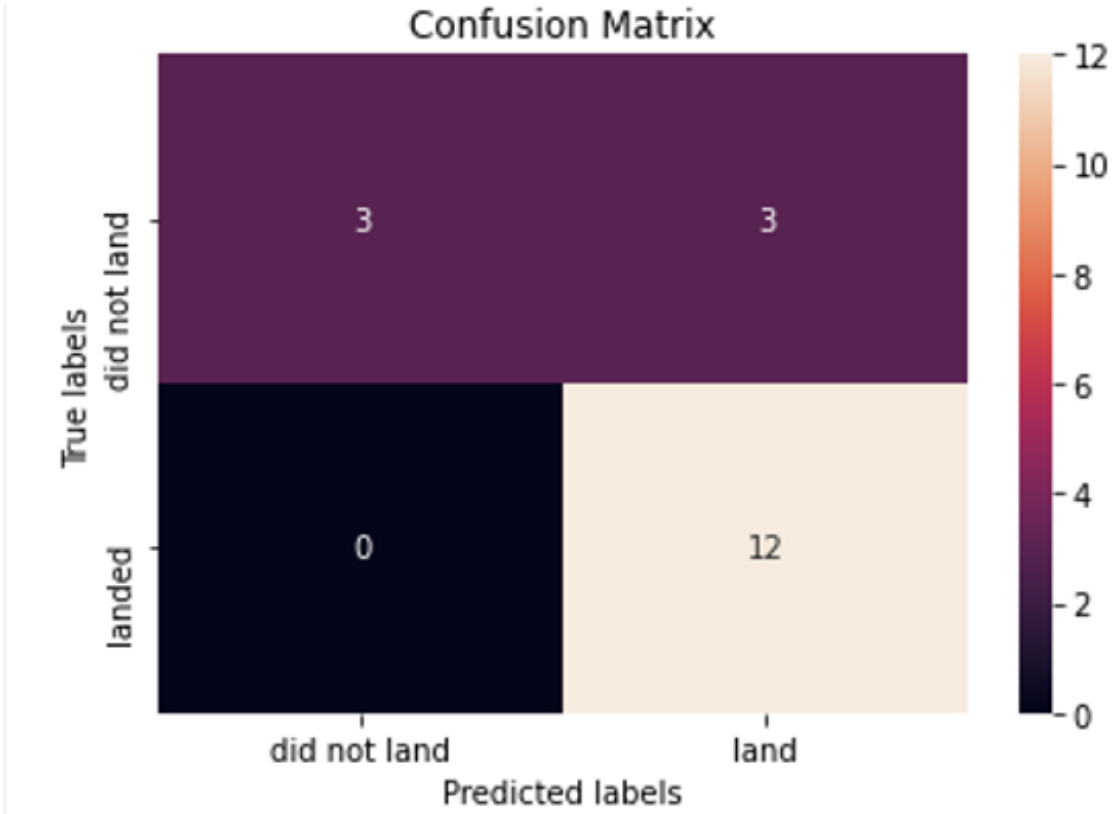
INTERACTIVE DASHBOARD WITH PLOTLY DASH



PIE CHARTS WITH LAUNCH OUTCOMES WERE PLOTED.

PLOTTED SCATTER GRAPH SHOWING THE RELATIONSHIP WITH OUTCOME AND PAYLOAD MASS (KG) FOR THE DIFFERENT BOOSTER VERSION.

Predictive Analysis



- Overall, the confusion matrix is a useful tool for understanding how well a classification model is performing, particularly in cases where the cost of false positives and false negatives differs significantly.

Conclusion

In conclusion, the Data Science and Machine Learning Capstone Project has been a valuable learning experience for our team. We have gained hands-on experience in data cleaning, exploratory data analysis, feature engineering, model selection, and evaluation.

Our project aimed to solve a real-world problem using machine learning techniques. We followed a rigorous methodology and experimented with various models to find the one that best fits our problem.

We presented our findings in a clear and concise manner using visualizations and metrics. We also discussed the limitations of our approach and suggested future work that could improve our model's performance.

Overall, this project has helped us develop a strong foundation in data science and machine learning, and we are confident that the skills and knowledge we have gained will be beneficial in our future endeavors.