Decoding the Skies

Insights from SpaceX Rocket Data Science

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- The following methodologies were used to analyze the data
  - Data collection using Web scraping and SpaceX API
  - Exploratory data analysis using Data Visualization, Data Wrangling, interactive Visual analytics.
  - Various Machine Learning Algorithms
- This data science study examines the landing of rockets by SpaceX, focusing on the analysis of key factors that contribute to successful landings. By leveraging extensive data collected from numerous SpaceX rocket launches, This study aims to provide valuable insights into the challenges, trends, and improvements associated with rocket landings.

Introduction

- The main objective of this study is to find the trends and factors which affect landings of rockets.
- Question: **How can we use the previous data to predict if a rocket and land Safely?**
- Desired Answers:
  - The best way to estimate the total cost for launches, by predicting successful landing of the first stage of rockets
  - Where is the best place to land and launch ?

Methodology

Methodology

- Data collection Methodology:
  - Data from Space C was obtained from 2 sources:
    Space C APi- https://api.spacexdata.com/v4/rockets/
    Web Scraping-
    https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
- Perform Data Wrangling
  - Collected Data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features

Notebooks:

https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/Data-Collection-API.ipynb
https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/Data-Collection-With-Web-Scrapping.ipynb
https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/Data-Wrangling.ipynb

Methodology

- Perform exploratory data analysis using Visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Data that was collected until this step were normalized, divide into training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters

Notebooks:

https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/EDA.ipynb
https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/EDA-With-SQL.ipynb
https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/DashBoard.ipynb

Predictive Analysis

- One of the primary goals of predictive analysis is to uncover hidden insights and make informed decisions based on the predicted outcomes. It can be applied across various industries and sectors, such as finance, healthcare, marketing, manufacturing, and more, to optimize operations, improve customer experiences, mitigate risks, and drive strategic planning.
- Created a interactive dashboard using Dash.

  https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/Predictive-Analysis.ipynb

  https://github.com/Gyaneshwar-Seemala/Data-Science-Capstone/blob/master/spacex_dash_app.py

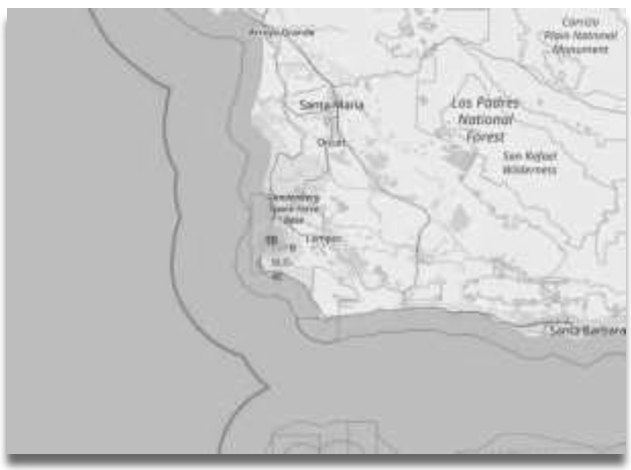
Results

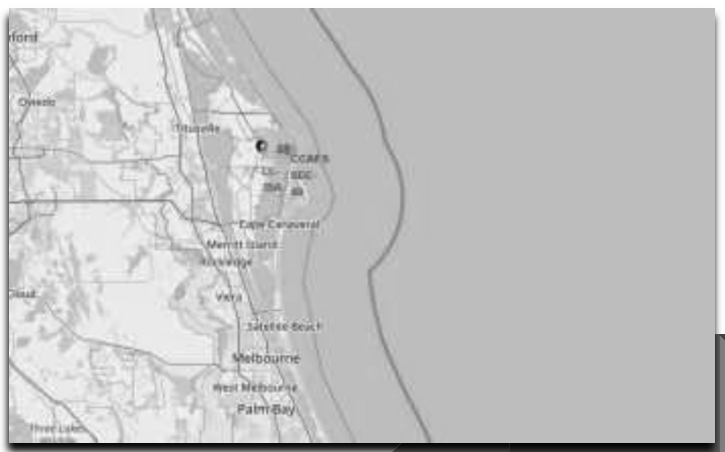
Exploratory data analysis results

- Space X uses 4 different launch sites
- The first launched were done to space X itself and NASA
- The average payload of F9 v1.1 booster is 2928 kg
- The first success landing outcome happened in 2015 five years after the first launch
- Many falcon 9 booster versions were successful at landing in drone ships having payload above the average
- Almost 100% of mission outcomes were successful
- Two booster versions failed at landing in frone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
- The number of landing outcomes became as better as years passed

Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east coast launch sites





Results

- Predictive analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuray overy 83% and accuracy for test data over 91%