Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

Executive Summary

- Objective: Determine Falcon 9 first stage landing success to predict launch costs.

- Methodology: Utilize a standard Data Science pipeline including:

- Data Collection: Aggregate historical launch records.

- Data Processing: Clean and structure data for analysis.

- Feature Engineering: Identify key predictors of landing outcomes.

- Model Selection: Test various algorithms for best predictive accuracy.

- Validation: Employ cross-validation techniques to ensure model robustness.

- Impact: Enable cost estimation for SpaceX and potential competitors.

- Outcome: Aid in competitive bidding strategy for rocket launch services.

Introduction

- Project background and context

- In this capstone, we will 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.

Data Collection – SpaceX API

- The data is collected via the request to the SpaceX API

- GitHub URL

- https://github.com/JulioWangUS/DS-jupyter/blob/main/1-spacex-data-collection-api.ipynb

Data Wrangling

- Performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident.

- Converted those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

- URL

- https://github.com/JulioWangUS/DS-jupyter/blob/main/2-webscraping.ipynb

- https://github.com/JulioWangUS/DS-jupyter/blob/main/3-spacex-Data%20wrangling.ipynb

EDA and interactive visual analytics methodology

- performed EDA using a database.

- checked if the data can be used to automatically determine if the Falcon 9's first stage will land. Some attributes can be used to determine if the first stage can be reused.

- used these features with machine learning to automatically predict if the first stage can land successfully

- determined what attributes are correlated with successful landings. The categorical variables will be converted using one hot encoding, preparing the data for a machine learning model that will predict if the first stage will successfully land.

EDA with Data Visualization

- Plotted scatter, bar and line charts, and conducted Feature Engineering to explore the relationships between the features and outcome.

- Add the GitHub URL

- https://github.com/JulioWangUS/DS-jupyter/blob/main/5-eda-dataviz.ipynb

EDA with SQL

- **Display the names of the unique launch sites in the space mission.**

- **Display 5 records where launch sites begin with the string 'CCA'**

- **Display the total payload mass carried by boosters launched by NASA (CRS)**

- **Display average payload mass carried by booster version F9 v1.1**

- **List the date when the first succesful landing outcome in ground pad was acheived.**

- **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**

- **List the total number of successful and failure mission outcomes**

- **List the names of the booster_versions which have carried the maximum payload mass. Use a subquery**

- **List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.**

- **Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.**

- Add the GitHub URL

- https://github.com/JulioWangUS/DS-jupyter/blob/main/4-eda-sql.ipynb

Build an Interactive Map with Folium

- Marked all launch sites on a map

- Marked the success/failed launches for each site on the map

- Calculated the distances between a launch site to its proximities

- Add the GitHub URL
- https://github.com/JulioWangUS/DS-jupyter/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- 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.
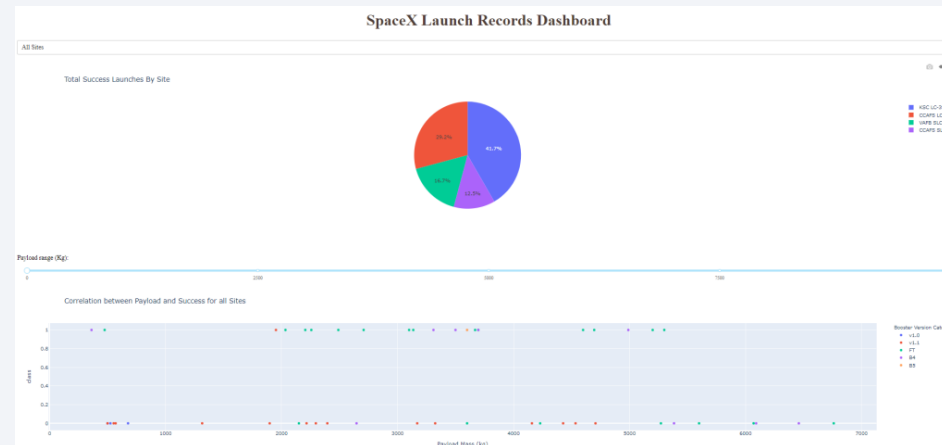
- Add the GitHub URL
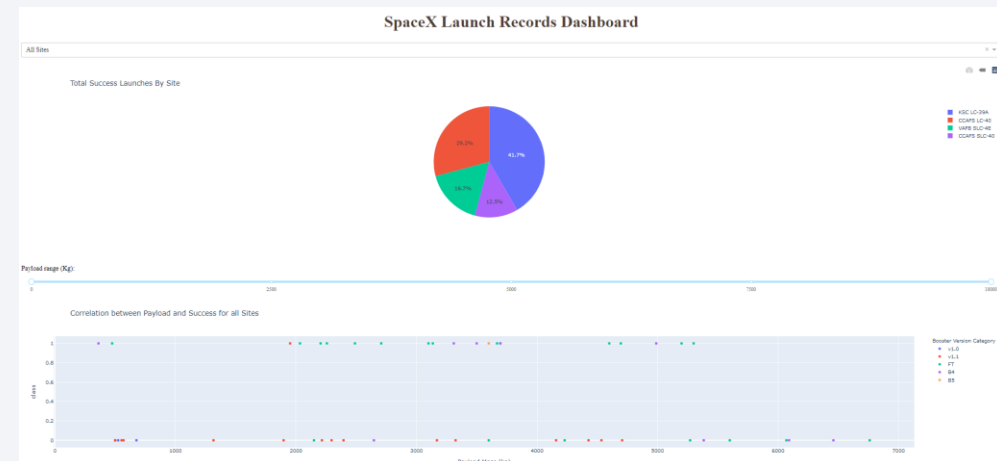- https://github.com/JulioWangUS/DS-jupyter/blob/main/6_spacex_dash_app.py
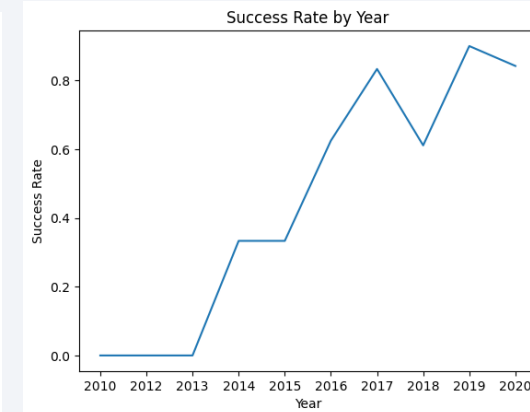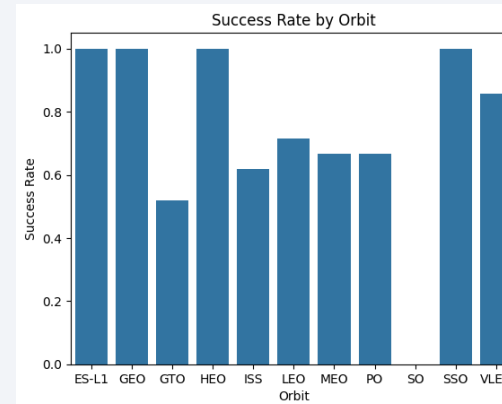
Predictive Analysis (Classification)

- Perform exploratory Data Analysis and determine Training Labels

    - create a column for the class

    - Standardize the data

    - Split into training data and test dataYou need present your model development process using key phrases and flowchart

- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
    - Find the method performs best using test data

- Add the GitHub URL

- https://github.com/JulioWangUS/DS-jupyter/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results

- Interactive analytics demo in screenshots

- Predictive analysis results
  - Whole sample size: 90
  - Test size: 18
  - Best SVM kernel: 'sigmoid'
  - Highest accuracy by decision tree via GridSearch: 0.83

Thank you!