
# Winning Space Race with Data Science

Julian Camilo Garcia
2023-11-19

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wragling
  - Data Analysis
  - Data Viz
  - Machine Learning Model
- Summary of all results
  - Space race data was extracted by public space launches resources
  - EDA allowed to understand data and know which features could be treated with feature engineering
  - iteration through different model showed which model is the best option

# Introduction

- Project background and context
  - Evaluate the market environment for a new company in the market race
- Problems you want to find answers
  - Principal insights related with space race and its competitors
  - Predict successful landings of the first stage of rockets

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - WebScraping (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
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

# Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

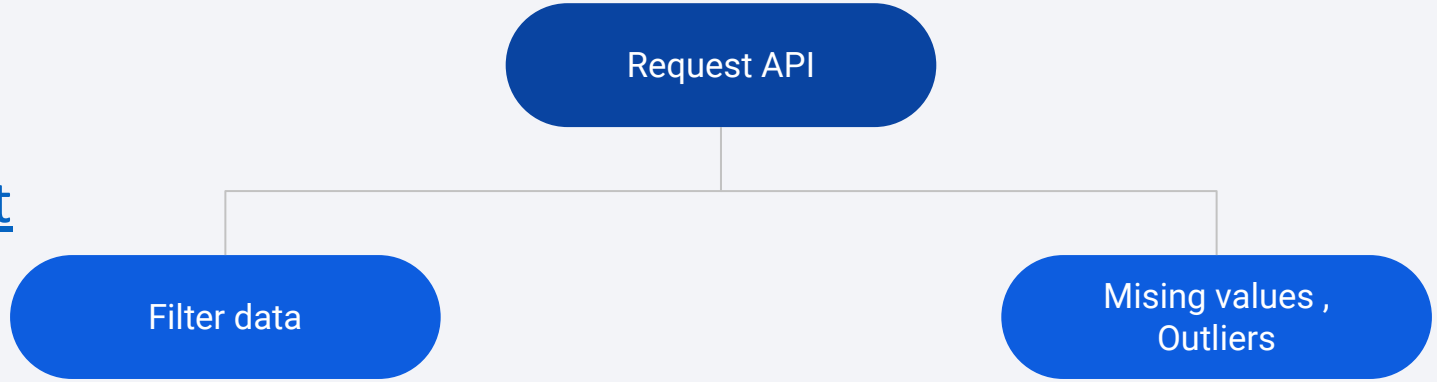
# Data Collection – SpaceX API

- SpaceX offers a public API from

  where data can be obtained and then used;

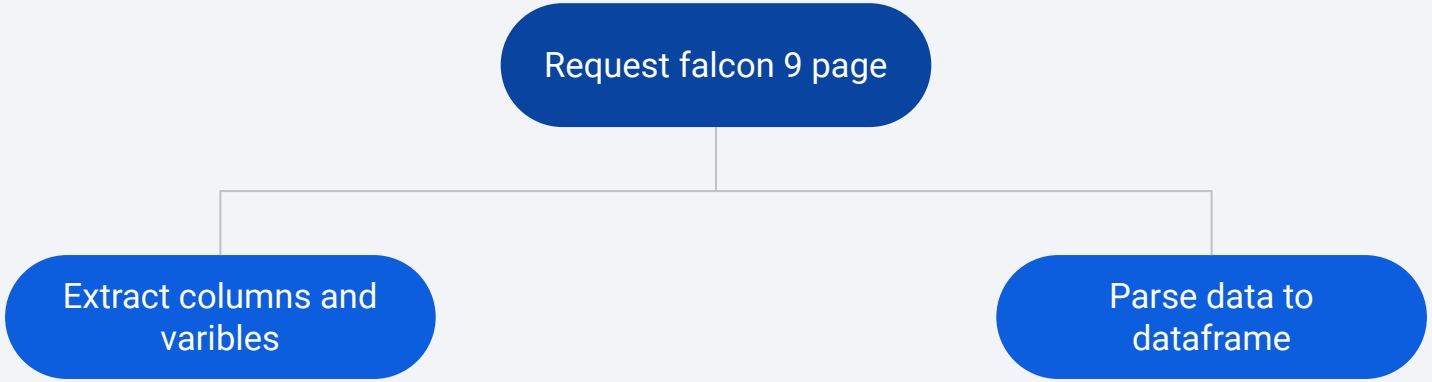
  This API was used according to the flowchart beside and then data is persisted.

https://github.com/JuliXnCXm/datasicenci-capstone



# Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.




- https://github.com/JuliXnCXm/datasicenci-capstone

# Data Wrangling

• Initially some Exploratory Data Analysis (EDA) was performed on the dataset.

• Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.

• Finally, the landing outcome label was created from Outcome column.

# EDA with Data Visualization

- https://github.com/JuliXnCXm/datasicenci-capstone

# EDA with SQL

- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between

4000 and 6000 kg;

- Total number of successful and failure mission outcomes;

# Build an Interactive Map with Folium

Markers, circles, lines and marker clusters were used with Folium Maps

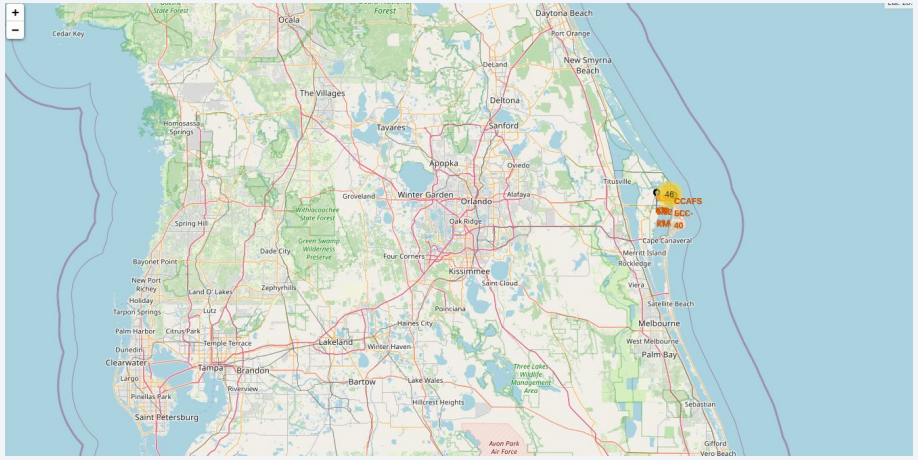
• Markers indicate points like launch sites;

• Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;

• Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and

• Lines are used to indicate distances between two coordinates.

• https://github.com/JuliXnCXm/datasicenci-capstone



# Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data

• Percentage of launches by site

• Payload range

• This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

- https://github.com/JuliXnCXm/datasicenci-capstone

# Predictive Analysis (Classification)

Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

- Data Prep and normalize
- build the model
- evaluate the model

- https://github.com/JuliXnCXm/datasicenci-capstone

# Thank you!