

ROCKET SCIENCE THROUGH DATA

- Ander Nebreda
- 27/02/2024



INDEX

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

EXECUTIVE SUMMARY

- Methodology of the project:
- Summary of the results:

INTRODUCTION

- The objective of this project is to analyze SpaceX public data, so we can predict the feasibility of the competitor company SpaceY.
- For that purpose, the project cover all major steps:
- Scraping the data from the available websites and obtaining visual insights from the data.
- Creating predictive models using ML techniques, which will help us predicting the successfull landing of the first stage rockets.

DATA COLLECTION

- Data collected from SpaceX API:

The data is extracted using requests, and then is filtered for Falcon 9 rockets only.

After it, we deal with missing values.

Data source: <https://api.spacexdata.com/v4/launches/past>

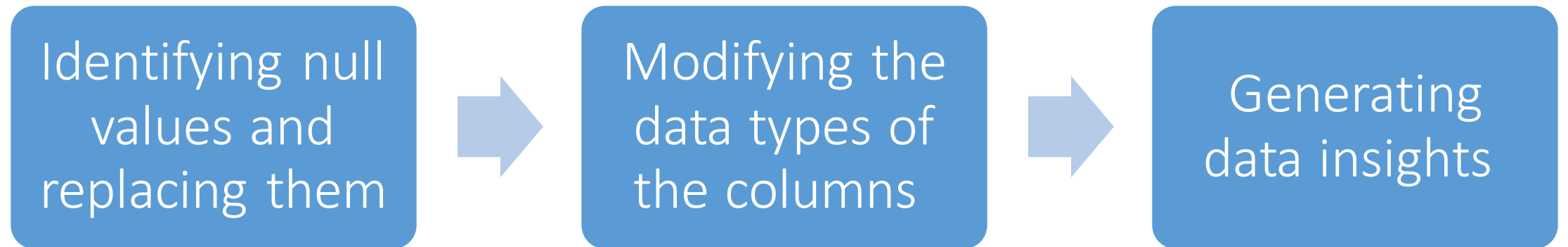
- Data collected from Wikipedia:

In this process, the data is scraped from a Wikipedia using BeautifulSoup.

Data

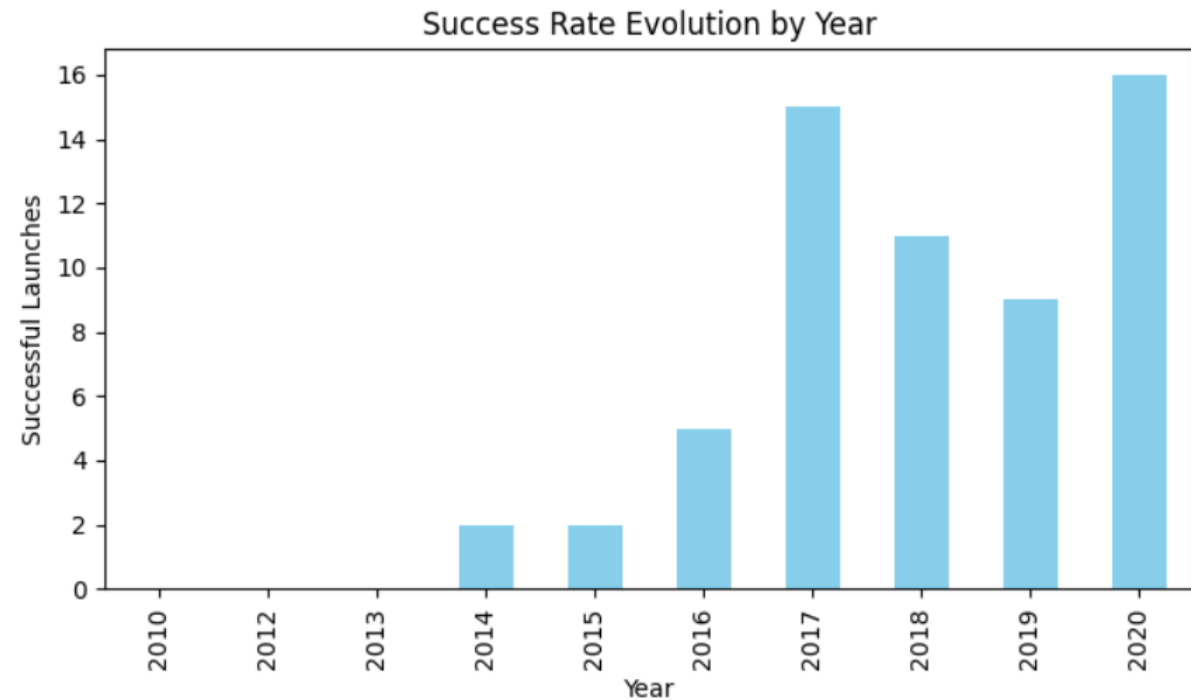
source: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

DATA WRANGLING



EDA VISUAL ANALYSIS

For the sake of a better understanding of the data, in this stage of the process an exploratory visual analysis of the data is carried out.



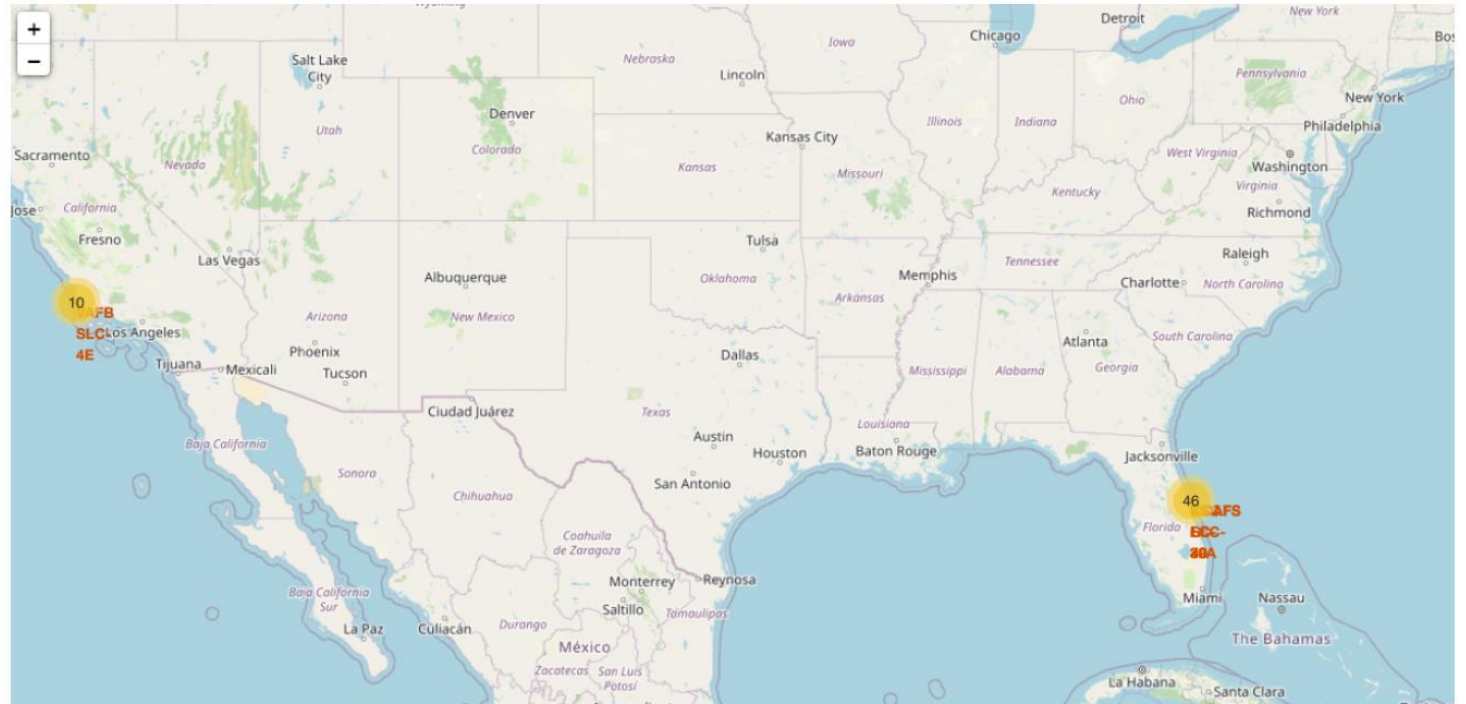
EDA SQL

In this stage were addressed different data extraction tasks from 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

MAP WITH FOLIUM

- In this section, data related to rocket launches is downloaded, and then is used to mark in a map with the help of Folium the launching sites by their success.



BUILDING A DASHBOARD WITH PLOTLY DASH

Creating a dashboard allow us to better visualize the data.

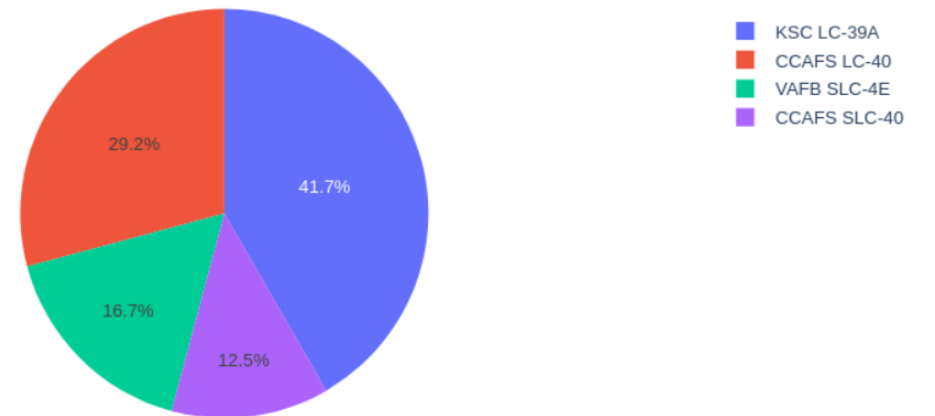
Among the different data visualized tgere are:

- Percent of launches by site.
- Payload mass.

All Sites



Total Success Launches By Site



PREDICTIVE ANALYSIS

In the final stage of the project, different machine learning models are used to predict the success rate of future rocket launches.

Those models used can be divided as:

- Supervised models: Logistic Regression, SVM and Decision Trees.
- Unsupervised models: K-Nearest.

RESULTS

- Space X uses 4 different launch sites
- The first launches were done to Space X and Nasa
- The average payload of F9 v1.1 booster is 2928 kg
- The first successful landing outcome happened in 2015. Five years after the first launch.
- Almost 100% of mission outcomes were successful
- Two booster versions failed at landing in droneships in 2015
- The success rate improved each year.

RESULTS

On the results obtained from the predictive analysis:

- After normalizing the data and using GridSearch to achieve the parameters that maximizes the prediction in the train set, I've concluded that every models has the same prediction score when tested against the data.

CONCLUSION

This capstone project embodies the whole Data Science course, and throughout it, all the learnt skills have been tested.