IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Waseem Akram Hans>
<16/05/2023>

Outline

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

Executive Summary

Summary of methodologies

- Data gathering using the SpaceX API and web scraping;
- Exploratory Data Analysis (EDA), which includes interactive visual analytics and data visualisation;
- Prediction using machine learning.

Summary of all results

- Valuable data was collected from public sources.
- Exploratory Data Analysis (EDA) identified the most predictive features for launch success.
- Machine Learning Prediction determined the optimal model for predicting the key characteristics that drive this opportunity effectively, leveraging the collected data.

Introduction

- To assess the potential of Space Y as a rival to Space X, we need to:

  - Find the most accurate way to calculate the launch costs, by factoring in the success rate of first-stage rocket landings;

  - Identify the optimal location for launching rockets

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- We collected data from Space X using two methods:
- Accessing the Space X API (https://api.spacexdata.com/v4/rockets/)
- Scraping the web page (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - We created a label for the landing outcome by summarizing and analyzing the outcome data and other features.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We normalized the data and split it into training and test sets. We then tested four different classification models and measured their accuracy with various parameter settings.

Data Collection

We used web scraping techniques to collect data from Space X API (https://api.spacexdata.com/v4/rockets/) and Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches).

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- We accessed a public API provided by SpaceX to get and use data;

• We followed the flowchart beside and saved the data.

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- • We also got data from SpaceX launches from Wikipedia;
- • We downloaded the data following the flowchart and stored it.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Place your flowchart of web scraping here

Data Wrangling

- We started with some Exploratory Data Analysis (EDA) on the dataset.
- Next, we computed the summaries of launches per site, orbits per launch and mission outcomes per orbit type.
- Lastly, we created the landing outcome label from the Outcome column.

- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- We used scatterplots and barplots to explore the data and visualize how features are related:

- Payload Mass vs Flight Number, Launch Site vs Flight Number, Launch Site vs Payload Mass, Orbit vs Flight Number, Payload vs Orbit

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- We performed these SQL queries:
- The unique launch sites in the space mission;
- The top 5 launch sites starting with 'CCA’;
- The total payload mass of boosters launched by NASA (CRS);
- The average payload mass of booster version F9 v1.1;
- The date of the first successful landing outcome in ground pad;
- The boosters that succeeded in drone ship and had payload mass between 4000 and 6000 kg;
- The number of successful and failure mission outcomes;
- The booster versions that carried the maximum payload mass;
- The failed landing outcomes in drone ship, their booster versions, and launch sites for in 2015; and The rank of the landing outcomes (such as Failure (drone ship) or Success (ground pad)) between 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

We used Folium Maps with these features:

- Markers to show points like launch sites;
- Circles to highlight areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters to group events in each coordinate, like launches in a launch site; and
- Lines to measure distances between two coordinates.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- We used these graphs and plots to visualize data
- • Launches by site as percentage

• Payload range

• This helped us to quickly analyze how payloads and launch sites are related, and to find the best place to launch based on payloads

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- We compared four classification models: logistic regression, support vector machine, decision tree and k nearest neighbors.

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

- We found these results from exploratory data analysis:
  - Space X has 4 launch sites;
  - The first launches were for Space X and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
  - The first successful landing outcome was in 2015, five years after the first launch;
- Many Falcon 9 booster versions landed successfully in drone ships with payload above the average;
- Almost all mission outcomes were successful;
  - Two booster versions failed to land in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The landing outcomes improved over time.

Results

- We used interactive analytics to find that launch sites are usually in safe places, near the sea, and have good logistics around.
- Most launches occur at east coast launch sites

Results

- We used interactive analytics to find that launch sites are usually in safe places, near the sea, and have good logistics around.
- Most launches occur at east coast launch sites

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- The plot above shows that the best launch site now is CCAF5 SLC 40, where most of recent launches succeeded;
- The second best is VAFB SLC 4E and the third best is KSC LC 39A;
- It also shows that the success rate increased over time.

Payload vs. Launch Site

- Payloads over 9,000kg (about a school bus) have a high success rate;
- Payloads over 12,000kg seem to be only possible at CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The highest success rates are for orbits:
- ES-L1 ,GEO HEO , and SSO.
- Next are: VLEO (over 80%), and LFO (over 70%).

Flight Number vs. Orbit Type

- Success rate increased over time for all orbits;
- VLEO orbit offers a new business opportunity, as its frequency rose recently.

Payload vs. Orbit Type

- Payload and success rate are unrelated for orbit GTO;
- ISS orbit has the largest payload range and a high success rate;
- SO and GEO orbits have few launches.

Launch Success Yearly Trend

- Success rate rose from 2013 to 2020;
- The first three years were a time of adjustment and technology improvement.

All Launch Site Names

- • Data shows four launch sites: Launch Site CCAFS LC-40 CCAFS SLC-40 KSC LC-39A VAFB SLC-4E
- They are unique “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

5 Cape Canaveral launches with CCA sites:

Total Payload Mass

- Total payload by NASA boosters: Total Payload (kg) 111.268
- Total payload is the sum of all payloads with ‘CRS’ codes for NASA.

Average Payload Mass by F9 v1.1

- Average payload mass by booster version F9 v1.1: Avg Payload (kg) 2.928
- Data filtered by booster version and average payload mass calculated as 2,928 kg.

First Successful Ground Landing Date

- First successful landing on ground pad: Min Date 2015-12-22
- Data filtered by successful landing on ground pad and minimum date value is 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters with successful drone ship landing and payload mass between 4000 and 6000 Booster Version F9 FT B1021.2 , F9 FT B1031.2 , F9 FT B1022 , F9 FT B1026
- These 4 are the distinct booster versions with the filters.

Total Number of Successful and Failure Mission Outcomes

- Mission outcomes grouped and counted for each group as shown below. Total Number of Mission Outcomes by Success and Failure Mission Outcome Occurrences Success 99 Success (payload status unclear) 1 Failure (in flight) 1 .

Boosters Carried Maximum Payload

- Boosters with maximum payload mass
- These boosters have the highest payload mass in the dataset.

Boosters with Maximum Payload Booster Version (…) F9 B5 B1048.4 ,F9 B5 B1048.5 , F9 B5 B1049.4 , F9 B5 B1049.5 , F9 B5 B1049.7 , F9 B5 B1051.3 Booster Version , F9 B5 B1051.4 , F9 B5 B1051.6 , F9 B5 B1056.4 , F9 B5 B1058.3 , F9 B5 B1060.2 , F9 B5 B1060.3

2015 Launch Records

- Drone ship landing failures, booster versions, and launch sites in 2015
- These are the only two cases. 2015 Launch Records Booster Version Launch Site F9 v1.1 B1012 CCAFS LC-40 F9 v1.1 B1015 CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranking of landing outcomes from 2010-06-04 to 2017-03-20:

• Data shows “No attempt” must be considered. Rank Landing Outcomes from 2010-06-04 to 2017-03-20 Landing Outcome Occurrences No attempt 10 Failure (drone ship) 5 Success (drone ship) 5 Controlled (ocean) 3 Success (ground pad) 3 Failure (parachute) 2 Uncontrolled (ocean) 2 Precluded (drone ship) 1

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

- Launch sites are close to sea for safety, but near roads and railroads.

<Folium Map Screenshot 2>

- KSC LC-39A launch site launch outcomes
- Green for success and red for failure.

<Folium Map Screenshot 3>

- Launch site KSC LC-39A has good logistics, close to railroad and road and away from populated areas.

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Launch site seems to be a key factor for mission success.

<Dashboard Screenshot 2>

- This site has a success rate of 76.9% for launches.

<Dashboard Screenshot 3>

- FT boosters and payloads ≤6,000kg have the highest success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The plot shows the accuracies of four classification models:
- Decision Tree Classifier has the highest accuracy, above 87%

Confusion Matrix

- The confusion matrix shows that Decision Tree Classifier has high accuracy, with many true positives and negatives and few false ones.

Conclusions

- The main findings are:
- KSC LC-39A is the best launch site
- Launches over 7,000kg have lower risk
- Landing success improves over time with better processes and rockets
- Decision Tree Classifier can predict landing success and increase profits.

Appendix

Thank you!