



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vicente Soler
14/11/2024



Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix
- Disclaimer
 - As you can see, I wasn't able to complete everything, but I've done all the tasks and now I am out of time (asking for a pass should be straightforward) but with all the challenges of flooding, creating visually appealing charts is the least of our concerns. (check news about Comunidad Valenciana and you will understand). Thanks in advance.
 - Added links to the tasks so you can check they are done.
 - <https://github.com/visolerga/IBMCapstone.git>

Executive Summary

- **Methodology**

- Data Collection: Gather data using APIs or web scraping techniques.
- Data Wrangling: Clean and structure the data, ensuring it is in the proper format.
- Exploratory Data Analysis (EDA): Analyze data to ensure correctness and identify patterns.
- Data Analysis: Conduct in-depth analysis using appropriate statistical or machine learning techniques.
- Result Extraction and Presentation: Present key findings clearly and compellingly.

- **Results**

- Rocket first-stage reuse significantly reduces costs.
- SpaceX is succeeding with first-stage rocket reuses.
- FT Booster has the best performance.
- Kennedy Space Center (KSC LC-39A) has the highest launch success rate.

Introduction

- Project background and context
 - First-stage rocket reuse is crucial for reducing costs.
 - Investigating the relationship between boosters, launch sites, and success rates.
- Problems you want to find answers
 - How does first-stage reuse impact launch costs?
 - Which booster has the highest success rate?
 - Which launch site has the best success rate?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
 - Describe how data was collected
- Perform data wrangling
 - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- 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
- <https://github.com/visolerga/IBMCapstone/blob/bf146689678e474e3c74fd72b1db580ca2e34b8c/jupyter-labs-spacex-data-collection-api-v2.ipynb>

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
- <https://github.com/visolerga/IBMCaps tone/blob/bf146689678e474e3c74fd72b1db580ca2e34b8c/jupyter-labs-webscraping.ipynb>

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
- <https://github.com/visolerga/IBMCapstone/blob/bf146689678e474e3c74fd72b1db580ca2e34b8c/labs-jupyter-spacex-Data%20wrangling-v2.ipynb>

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
- <https://github.com/visolerga/IBMCapstone/blob/bf146689678e474e3c74fd72b1db580ca2e34b8c/jupyter-labs-eda-dataviz-v2.ipynb>

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
- https://github.com/visolerga/IBMCapstone/blob/bf146689678e474e3c74fd72b1db580ca2e34b8c/jupyter-labs-eda-sql-coursera_sqlite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
- https://github.com/visolerga/IBMCapstone/blob/02c56e9f3b8bcc12de348f79adb820e65c7cb01f/my_map.html
- <https://github.com/visolerga/IBMCapstone/blob/02c56e9f3b8bcc12de348f79adb820e65c7cb01f/lab-jupyter-launch-site-location-v2.ipynb>

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
- https://github.com/visolerga/IBMCapstone/blob/02c56e9f3b8bcc12de348f79adb820e65c7cb01f/InteractiveDashboard/spacex_dash_app.py

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose
- <https://github.com/visolerga/IBMCapstone/blob/02c56e9f3b8bcc12de348f79adb820e65c7cb01f/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb>

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

The background of the slide is an abstract composition. It features a dark blue field on the left side, which transitions into a complex pattern of diagonal streaks in shades of blue, red, and teal on the right. These streaks have a textured, almost woven appearance. Overlaid on this pattern is a faint, light blue grid that recedes into the distance, creating a sense of depth and perspective.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

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

- 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
- Present your query result with a short explanation here

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The image is a composite of a solid blue background on the left and a satellite image of Earth on the right. The Earth's surface is dark blue, with numerous bright yellow and orange lights representing cities and urban areas. The lights are concentrated in the lower right portion of the image, following the curve of the Earth's horizon. The overall composition suggests a global or space-related theme.

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- ...

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!

