



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

Juan David Pereira Bermudez  
January 2025



# Table of contents:

---

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

# Executive Summary

---

- Summary of methodologies
- Summary of all results

# Introduction

---

The space sector has experienced exponential growth in recent decades, driven by increasing demand for launch services and cost reduction. SpaceX, as a pioneer in rocket reusability, has revolutionized the industry by significantly reducing the cost of accessing space.

This project is part of the ongoing quest to improve the efficiency and reliability of space launches. By predicting the success of Falcon 9 first stage rocket landings, operations can be optimized, costs reduced and more informed decisions about future missions can be made.

# Objectives

---

The main objectives of this project are:

- Identify the key factors that influence the success of Falcon 9 rocket first stage landings.
- Develop an accurate and robust prediction model to predict the success of future launches.
- Interpret model results to gain valuable insights into launch operations.
- Evaluate the model's ability to generalize to new data and its potential for use in space industry decision making.



Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Request to the SpaceX API
  - Web scraping records from Wikipedia
- Perform data wrangling
  - The data was filtered to keep only the data corresponding to the launch of Falcon 9, then a missing data treatment was performed in which 5 missing data in the PayloadMass variable were replaced by the value of the mean of this variable.
- Perform exploratory data analysis (EDA) using visualization and SQL

# Methodology

---

## Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash

An interactive dashboard was developed that provides an in-depth exploration of space launch data. In the dashboard, data can be filtered by launch site and payload mass range, allowing the success rate of each site to be dynamically visualized.

- 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

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

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

# 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



# 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

# 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

# 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

# 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

# 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 base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

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 background is a deep blue gradient.

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!

