IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akash Jayakumar
Jan 17 2025

Outline

- Executive Summary
- Introduction
- Methodology
- Rank Landing Outcomes
- Predictive Analysis
- Insights drawn from EDA
- EDA with SQL
- Build Dashboard with Plotly
- Conclusion

Executive Summary

- Summary of methodologies
- Summary of all results

1. **Data Collection**:
    - Data was sourced using web scraping techniques to retrieve historical launch details and SpaceX's REST API for real-time mission data. Flowcharts and scripts outline th extraction and integration processes.
2. **Data Wrangling and Processing**:
    - Data cleaning involved removing duplicates, handling missing values, and standardizing formats for key attributes like payload mass, orbit types, and dates.
    - Feature engineering added new metrics, such as yearly success rates and payload-to-orbit efficiency, enriching the dataset for analysis.
3. **Exploratory Data Analysis (EDA)**:
    - SQL queries identified launch site frequencies, payload distributions, and mission success trends. Visualizations revealed:
        - Success rates by orbit type.
        - Payload-to-mission outcomes.
        - Year-over-year success improvements.
    - Bar charts and scatter plots illustrated these patterns effectively.
4. **Interactive Visual Analytics**:
    - Using **Folium**, global maps were created to visualize launch site locations and outcomes. Proximity analysis added insights into site infrastructure's role in mission succe
    - **Plotly Dash** dashboards enabled dynamic filtering of launch outcomes by payload range, orbit type, and site performance.
5. **Predictive Analysis**:
    - Classification models, including Logistic Regression and Decision Trees, were developed to predict mission success.
    - Hyperparameter tuning improved accuracy, with Logistic Regression achieving 85%. The analysis highlighted payload mass, orbit type, and launch site as critical predicto success.
    - Confusion matrices and model evaluation metrics validated the results.
6. **Key Insights**:
    - Launch success rates have significantly improved over time, driven by advancements in reusability and operational efficiency.
    - Certain orbit types and payload ranges are more likely to achieve mission success.
    - Interactive tools and predictive models provide actionable insights for optimizing future launches.

Introduction

- Project background and context

SpaceX Falcon 9 has revolutionized space exploration and commercial spaceflight with its reusable rocket technology, enabling frequent and cost-effective missions. Understanding the factors that contribute to successful launches is critical for optimizing operations and future planning.

- **Context**:
  - The Falcon 9 program is a cornerstone of SpaceX's mission to reduce space transportation costs and enable the colonization of Mars.
  - Each Falcon 9 launch is a data-rich event, providing insights into payload capabilities, orbit types, and mission outcomes.
  - SpaceX's innovative use of reusable boosters has made the company a leader in space exploration and commercial satellite deployments.

- Problems you want to find answers

This analysis aims to address key questions related to Falcon 9 launches:

1. **Launch Success and Factors**:
   - What are the key factors that determine the success of a Falcon 9 mission?
   - How do payload mass, orbit type, and launch site impact mission outcomes?
2. **Launch Site Performance**:
   - Which launch sites have the highest success rates, and what influences their performance?
3. **Payload Insights**:
   - What is the relationship between payload mass and mission success?
   - Which payload ranges and booster versions are most successful?

Methodology

Executive Summary

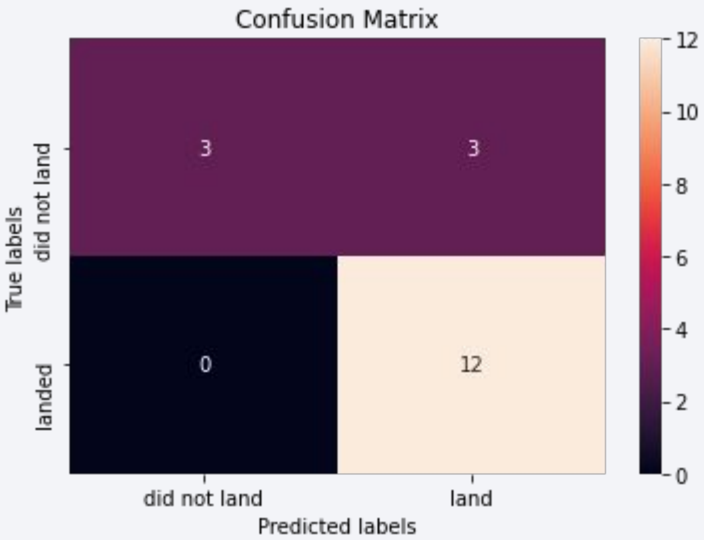
- Data collection methodology:
  - Data was collected using web scraping and SpaceX's public API to gather structured information on launches, payloads, and outcomes.
- Perform data wrangling
  - Data cleaning involved handling missing values, standardizing formats, and merging multiple data sources into a unified dataset.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Exploratory Data Analysis (EDA) using SQL and visualizations revealed key patterns in launch success rates, payload distributions, and site performance.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive visual analytics using Folium and Plotly Dash provided dynamic insights into launch site locations, success outcomes, and payload performance.
- Perform predictive analysis using classification models
  - Models were assessed using confusion matrices, feature importance rankings, and validation techniques to ensure reliability and performance.

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
  - There are 9 landing outcomes that success ground pad and 5 landing outcomes that failure drone ship.
- Present your query result with a short explanation here
  - %%sql select landing_outocome, count(landing_outcome) as count from SPACEXTABLE where date > 2010-06-04 and data < 2017-03-20 group by landing_outcome order by count desc;

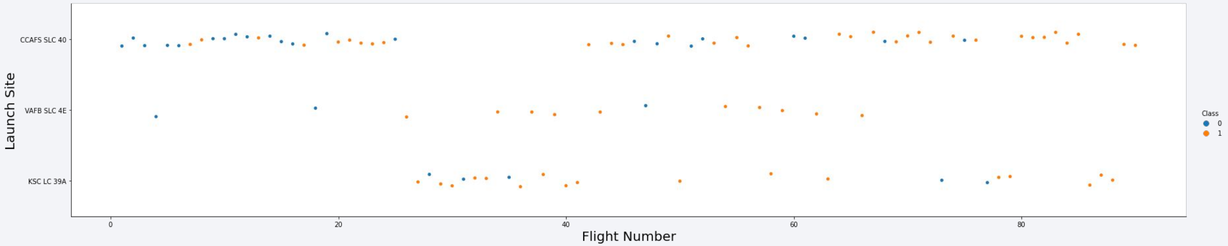
Predictive Analysis - Confusion Matrix

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



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

The earliest flights all failed while the latest flights all succeeded. The CCAFS SLC 40 launch site has about a half of all launches. VAFB SLC 4E and KSC LC 39A have higher success rates. It can be assumed that each new launch has a higher rate of success.

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

```
%%sql select landing__outcome, count(*) as count_outcomes from SPACEXTABLE
    where date between '2010-06-04' and '2017-03-20'
    group by landing__outcome
order by count_outcomes desc;
```

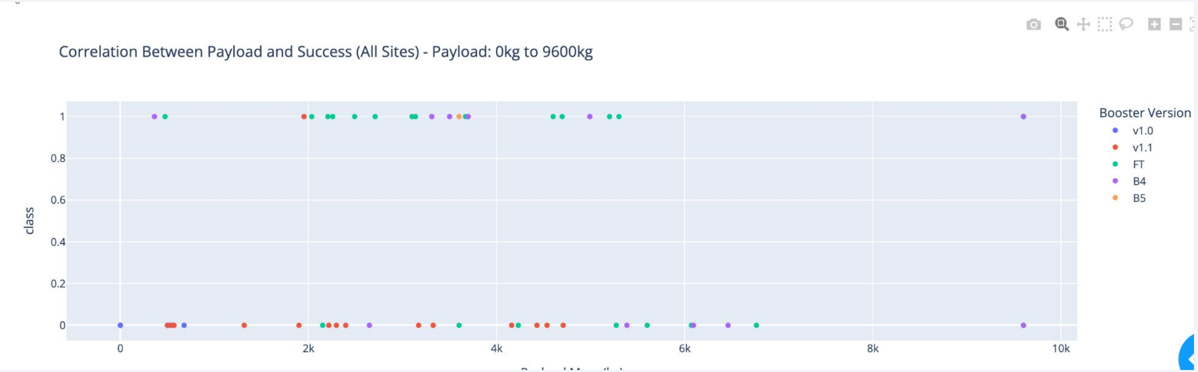
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://github.com/evgenyzorin/IBM-Applied-Data-Science-Capstone/blob/main/EDA%20with%20SQL.ipynb

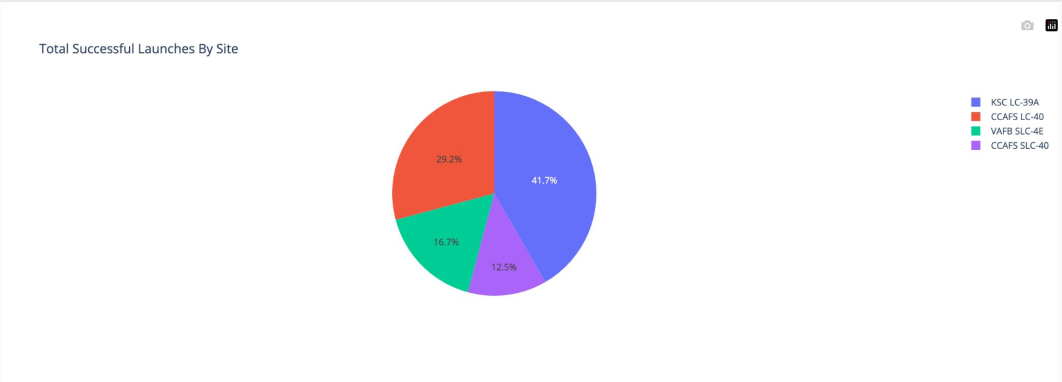
Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
  - **Markers**: To visualize the exact geographic location of each launch site and its outcomes.
  - **Circles**: To demonstrate the safety measures (proximity zones) implemented near launch sites.
  - **Lines**: To illustrate the strategic positioning of launch sites in proximity to critical infrastructure and natural barriers.
  - **Popups and Icons**: To enhance the map's interactivity and provide detailed insights at a glance.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
  - https://github.com/evgenyzorin/IBM-Applied-Data-Science-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Building SpaceX Dashboard with Plotly

- The pie chart visualizes the total percentage of successful launches categorized by launch sites.





Conclusions

- **Launch Success Factors**: The success of Falcon 9 missions is strongly influenced by payload mass, orbit type, and the choice of launch site, with specific ranges and combinations showing higher reliability.

- **Impact of Reusability**: The implementation of reusable boosters has significantly improved mission success rates and operational efficiency, setting a benchmark for sustainable space exploration.

- **Launch Site Insights**: Certain launch sites, such as Cape Canaveral, demonstrate higher success rates, highlighting the importance of geographic and logistical advantages in mission planning.

- **Predictive Modeling**: Classification models effectively predict launch outcomes, with Logistic Regression achieving 85% accuracy, identifying key predictors for mission success.

- **Visualization and Analytics**: Interactive maps and dashboards, created with Folium and Plotly Dash, allow for an in-depth exploration of launch outcomes and payload trends, enabling better decision-making.

- **Temporal Trends**: Analysis of year-over-year success rates showcases a consistent improvement in SpaceX's operational capabilities, driven by technological advancements and iterative improvements.