SpaceX Falcon 9 Launch Data Analysis and Prediction

IBM Data Science Capstone Project

2025

Executive Summary

- This project analyzes SpaceX Falcon 9 launch data to determine the factors influencing successful landings. Through exploratory data analysis, visualization, and machine learning, we built a predictive model to estimate launch success probability. Results show that payload mass, orbit type, and launch site are key predictors of success.

Introduction

- SpaceX's Falcon 9 rocket has revolutionized the space industry through reusable rockets. The main goal of this project is to predict the likelihood of a successful first-stage landing. We use publicly available SpaceX data from the company API and web scraping from Wikipedia.

Data Collection Methodology

- Data was collected from multiple sources:
- - SpaceX REST API for launch details
- - Wikipedia for supplementary launch information
- - Data compiled into pandas DataFrames for further analysis

Data Wrangling

- Data cleaning steps included:
- - Handling missing values and inconsistent formats
- - Converting categorical data into numerical form using get_dummies()
- - Extracting year, payload mass, and orbit for analysis
- - Creating binary target variable ‘Class’ for success/failure

Exploratory Data Analysis (Visual)

- We visualized relationships between key features and the success outcome.
- - Scatter plots showing Payload Mass vs Launch Site
- - Catplots for Flight Number vs Launch Site colored by success
- - Trends of launch success over the years

Exploratory Data Analysis (SQL)

- Using SQL queries integrated in Jupyter notebooks:
- - Retrieved average success rate by launch site
- - Identified the orbit types with the highest success probability
- - Filtered missions by year and payload range

Interactive Map with Folium

- Created an interactive Folium map displaying launch sites with success/failure markers.
- - MarkerCluster used to group close coordinates
- - Popups show launch site names and outcomes
- - Geospatial patterns reveal success concentration at Cape Canaveral

Interactive Dashboard with Plotly Dash

- Developed a Plotly Dash app integrating multiple components:
- - Dropdown menu for selecting launch site
- - Pie charts for success rate per site
- - Scatter plots showing Payload vs Launch success correlation
- - Dynamic updates through callback functions

Predictive Analysis Methodology

- Applied machine learning algorithms to classify launch success:
- - Logistic Regression
- - Support Vector Machines (SVM)
- - Decision Tree Classifier
- Models trained and evaluated using accuracy and confusion matrix metrics.

Predictive Analysis Results

- The Decision Tree model achieved the highest accuracy (~94%).
- Feature importance showed Payload Mass and Orbit type as strongest predictors.
- Predictions aligned closely with real outcomes, validating model robustness.

Conclusions

- The project demonstrates how data science can predict SpaceX launch outcomes.
- Reusable rockets have higher success probability over time.
- Our final model effectively predicts successful landings with high accuracy.
- Future work may involve integrating real-time launch telemetry.

Creativity and Insights

- Innovations introduced:
- - Combined geospatial and predictive analytics
- - Custom dashboards with interactive filters
- - Discovered correlation between payload range and orbit type success

References

- Data Sources:
- - SpaceX API
- - Wikipedia
- - IBM Data Science Capstone Materials

- Tools:
- - Python (Pandas, Matplotlib, Seaborn, Plotly, Folium)
- - SQL, Scikit-learn, Jupyter Notebook, Dash