

SpaceX Launch Data Analysis & Prediction

Your Name

IBM Data Science Capstone Project

2025

Executive Summary

- Objective: Analyze SpaceX Falcon 9 launch data to predict landing success.
- Tools: Python, Pandas, SQL, Folium, Plotly Dash, Scikit-learn.
- Key Insights:
 - - Launch success rates improving over time.
 - - Payload mass affects landing success probability.
 - - Models achieve $\approx 82 - 80\%$ accuracy predicting

Introduction

- Problem: Rocket reusability is critical for reducing costs in space exploration.
- Goal: Explore launch data → visualize trends → build predictive model.
- Approach: Combine EDA, SQL, geospatial mapping, and machine learning.

Data Collection & Wrangling

- Source: SpaceX API, web scraping, dataset provided.
- Cleaning: Removed null values, standardized site names.
- Feature Engineering:
 - - Encoded categorical variables.
 - - Created binary target variable: Landing Success.


EDA & Visualization Methodology

- Python (Matplotlib, Seaborn) for distributions and trends.
- SQL for data queries and aggregations.
- Folium for launch site geospatial mapping.
- Plotly Dash for interactive dashboards.


Predictive Analysis Methodology

- Models: Logistic Regression, Decision Tree, SVM.
- Evaluation Metrics: Accuracy, Confusion Matrix, ROC Curve.
- Goal: Predict landing success based on payload, site, and booster version.


EDA Results – Visualizations

- Payload mass distribution (Insert chart).
- Launch success rates by site (Insert bar chart).
- Correlation heatmap of features (Insert heatmap).
-  Replace placeholders with your charts.


EDA Results – SQL Queries

- Sample Query:
- `SELECT LaunchSite, COUNT(*) FROM SpaceX
WHERE Success = 1 GROUP BY LaunchSite;`
- Insights:
 - - KSC and CCAFS have higher success rates.
 - - Some sites are used less frequently.
-  Insert screenshot of SQL query + output table.


EDA Results – Folium Map

- Interactive Folium map of launch sites.
- Insights: Distribution of launches across US East/West coasts.
-  Insert Folium map screenshot.

Plotly Dash Dashboard Results

- Interactive features:
 - - Payload slider for success correlation.
 - - Dropdown for selecting launch site.
- Insight: Payload range strongly influences landing success.
-  Insert dashboard screenshot.

Predictive Analysis Results

- Logistic Regression accuracy: ~83%.
- Decision Tree accuracy: ~89%.
- Confusion Matrix → shows model performance.
- ROC Curve → indicates good model separation.
-  Insert evaluation visuals.

Conclusion

- SpaceX has significantly improved launch success rates.
- Payload and launch site are strong predictors of success.
- Predictive models achieve ~85–90% accuracy.
- Future scope: Incorporate newer launch data and advanced ML models.

Creativity & Innovation

- Built interactive dashboards and maps beyond static charts.
- Unique insight: Heavier payloads correlate with lower landing success.
- Presentation styled with storytelling focus.