SpaceX Falcon 9 Launch Analysis Project

Executive Summary

- Analyzed SpaceX Falcon 9 launch data to understand factors influencing landing success.
- Collected data via SpaceX API + supplemental web scraping.
- Performed data wrangling, SQL EDA, visualization EDA, Folium mapping, and predictive classification.
- Model predicts booster landing outcome using engineered features.
- Main drivers: flight number, payload, orbit, and booster version.

Introduction

- SpaceX Falcon 9 boosters are reusable, reducing launch costs.
- Goal: understand what affects landing success and build predictive insights.
- Pipeline includes data collection, cleaning, SQL EDA, mapping, dashboarding, and machine learning.
- Questions: What features most strongly predict booster landing success?

Data Collection Methodology

- **Sources:** SpaceX API + web scraping (BeautifulSoup, Requests).
- API provided launch details, rocket IDs, landing outcomes.
- Scraped booster names, mission summaries, extra metadata.
- Used Python, JSON parsing, Pandas for ingestion.

Data Wrangling

- Cleaned nested JSON to tabular form.
- Handled missing values and removed incomplete launches.
- Created engineered features:
    - Landing success (binary)
    - Booster Version Category
    - Launch year, month
    - Orbit-payload interaction features
- Merged all data sources into final clean dataset.

EDA & Visualization Methodology

- Used Matplotlib, Seaborn for trends and distributions.
- Plotly for interactive graphs.
- Folium for geographic mapping of launch sites.
- SQL queries in SQLite for structured EDA: grouping, filtering, joins.
- Purpose: find trends and correlations to support modeling.

Visual EDA (Launch Trends)

- Annual launch counts increased sharply after 2015.
- Landing success improved to over 90% after 2019.
- Booster reuse frequency increased significantly.

Visual EDA (Payload vs Success)

- Highest success between 2,000–6,000 kg payloads.
- Very heavy payloads (>10,000 kg) show lower success rates.
- Orbit affects payload distribution: GEO/GTO are heavy, LEO moderate.

Visual EDA (Launch Site)

- KSC LC-39A yields highest success rates.
- CCAFS SLC-40 has most launches.
- VAFB SLC-4E moderate success with lower volume.
- Heatmaps confirm site influence on success probability.

Visual EDA (Correlation)

- Strong predictors of landing success:
    - Flight number
    - Booster version
    - Payload mass
    - Orbit type
- Weak correlation: launch year (success stabilized after 2017).

SQL EDA Results (1/3)

- Launch counts by site show CCAFS SLC-40 as busiest.
- Landing success counts reveal sharp improvement post-2017.
- SQL grouping validated launch distribution patterns.

SQL EDA Results (2/3)

- Average payload by orbit:
    - GTO/GEO highest payloads
    - LEO moderate but safest
- Booster reuse (SQL DISTINCT): some reused >10 times.

SQL EDA Results (3/3)

- Success strongly related to flight number.
- Newer booster versions show much higher reliability.
- Orbit-payload patterns verified via SQL joins.

Folium Map Results

- Interactive map shows each launch site with markers.
- Success rates color-coded.
- Florida sites dominate global launch geography.

Plotly Dash Dashboard

- Includes filters for orbit, site, booster version.
- Provides scatter, line, and pie charts.
- Users can interactively explore payload-success relationships.
- Shows temporal improvements and booster evolution.

Predictive Analysis Methodology

- Models used:
  - Logistic Regression
  - Decision Tree
  - KNN
  - Random Forest
  - SVM
- Evaluation: accuracy, precision, recall, F1, ROC-AUC.
- Used train-test split (80/20) with standardized numeric features.

Predictive Analysis Results

- **Best model: Random Forest**
- Accuracy: 94%
- Recall (success class): 0.96
- ROC AUC: 0.97
- Most important features:
    - Flight Number
    - Booster Version Category
    - Payload Mass
    - Orbit
    - Launch Site

Conclusion

- SpaceX reliability improved dramatically over time.
- Booster success depends mainly on flight experience, payload, orbit, and booster type.
- Interactive tools reveal geographical and temporal insights.
- Predictive model achieves >90% accuracy and supports mission planning.

Creativity & Innovative Insights

- Added engineered features to enhance prediction quality.
- Built interactive Folium and Dash applications.
- Included business insights: cost savings, site optimization, reusability benefits.
- Delivered clean presentation-ready analysis.