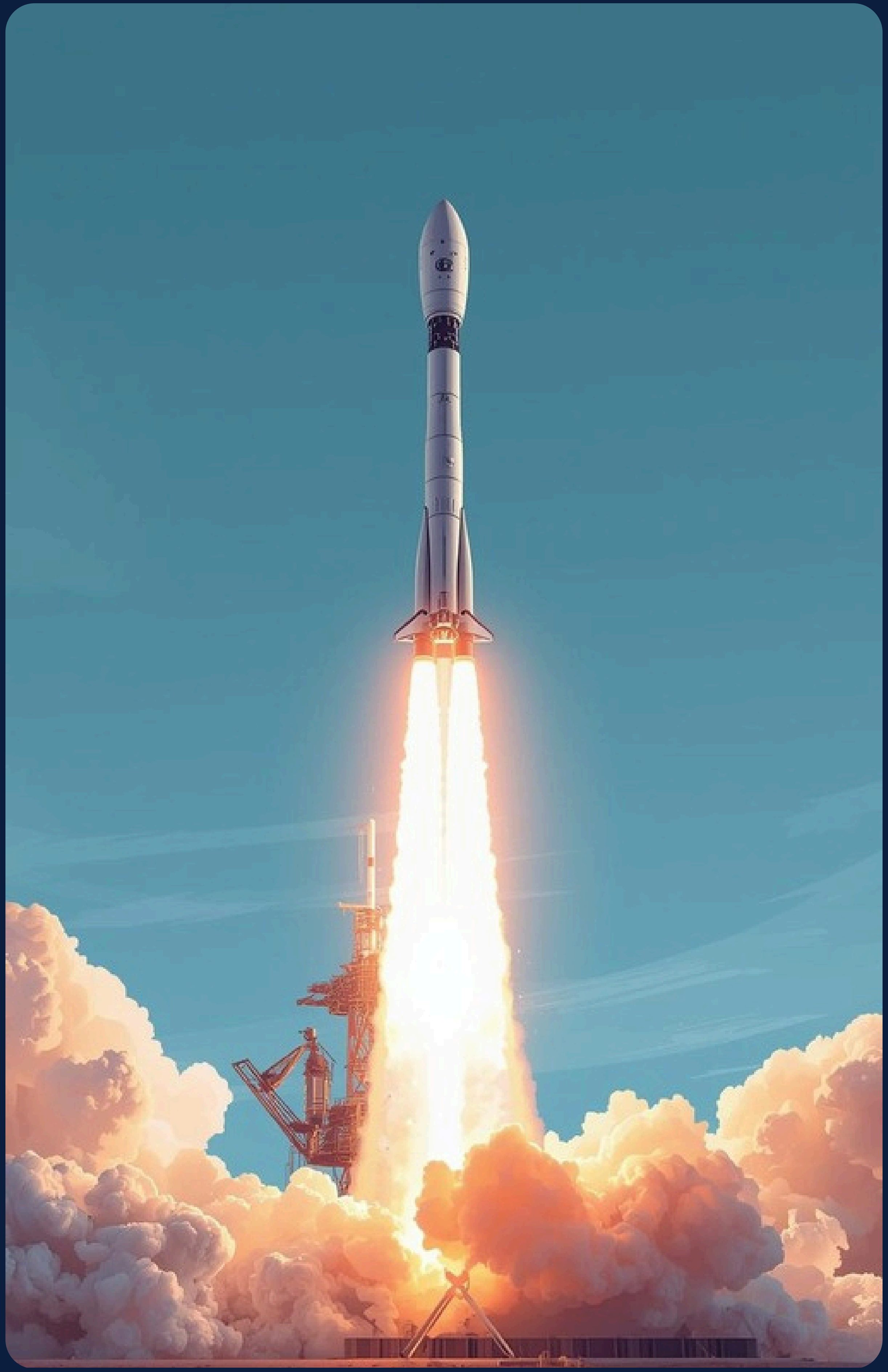
SpaceX Falcon 9 Prediction

Data Science Capstone Project

Merve EROL

GitHub Link:https://github.com/Elrevmore/SpaceX-Falcon9-Prediction-IBM-Captone-Project-



Executive Summary

Overview of Project Objectives and Value

- Objective: Predict whether the SpaceX Falcon 9 first stage will land successfully using historical launch data.
- Business Value:SpaceX advertises Falcon 9 launches at $62M vs. $165M+ for competitors—cost savings depend on first-stage reuse. Predicting landing success enables cost estimation and competitive bidding.
- Approach:Data wrangling → EDA (visualization + SQL) → Interactive mapping (Folium) → Plotly Dash dashboard → Machine learning classification (SVM, Decision Tree, Logistic Regression).

Key Outcome: Built an end-to-end pipeline from raw data to deployable predictive models, with interactive visualizations for stakeholder insights.

Introduction to Falcon 9

- Context: SpaceX has achieved historic milestones in reusable rocket technology. The Falcon 9 first stage can land on ground pads (RTLS), drone ships (ASDS), or ocean (planned failures).
- Problem: Determining landing success in advance supports launch cost estimation and strategic planning.
- Dataset: SpaceX launch records including payload mass, orbit type, launch site, booster version, and mission outcomes.
- Deliverables:Data wrangling, EDA with visualizations and SQL, interactive Folium maps, Plotly Dash dashboard, and classification models for landing prediction.

Data Collection and Data Wrangling Methodology

Data Sources:

- SpaceX API and IBM Cloud-hosted datasets (Spacex.csv, spacex_launch_geo.csv, dataset_part_2.csv)
- Payload and launch metadata from official SpaceX mission records

Wrangling Steps:

1- Load & merge:** Import CSV datasets using pandas; combine API and static data where needed.
2- Outcome encoding: Convert mission outcomes (True/False Ocean, True/False RTLS, True/False ASDS) into binary labels: 1 = successful landing, 0 = unsuccessful.
3- Feature engineering:Create derived features (e.g., payload mass bins, orbit categories) for modeling.
4- Cleaning:Handle missing values, standardize column names, and validate data types.
5- Export: Save wrangled datasets for EDA, SQL, mapping, and ML pipelines.

Tools: pandas, numpy

EDA and Interactive Visual Analytics Methodology

Exploratory Data Analysis:

- Summary statistics and distributions for numeric/categorical variables
- Correlation analysis between features and landing success
- Group-by analyses (launch site, orbit, booster version, payload mass)

Visual Analytics:

- Matplotlib & Seaborn: Bar charts, scatter plots, heatmaps, box plots
- Folium: Interactive maps with launch sites, success/failure markers, and proximity analysis
- Plotly Dash: Interactive dashboard for filtering and exploring launch records

Methodology:

1. Define questions (e.g., which launch site has highest success rate?)
2. Select appropriate visualizations
3. Iterate based on findings
4. Document insights for stakeholders

Predictive Analysis Methodology

Goal: Classify whether the Falcon 9 first stage will land successfully (binary classification).

Pipeline:

1. Features: Payload mass, orbit type, launch site, booster version, flight number, etc.
2. Preprocessing: Standardization (StandardScaler), train-test split (80/20)
3. Models: Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors
4. Hyperparameter tuning: GridSearchCV for optimal parameters
5. Evaluation: Accuracy, precision, recall, F1-score, confusion matrix
6. Selection: Choose best-performing model based on test metrics

Tools: scikit-learn (preprocessing, model selection, evaluation)

EDA with Visualization Results

Key Findings:

- Launch success over time:** Success rate improved significantly after 2017; early missions had more failures.
- Launch site: CCAFS SLC-40 and KSC LC-39A show higher success rates; VAFB SLC-4E has fewer launches but consistent performance.
- Payload mass: Heavier payloads (e.g., GTO missions) correlate with different landing strategies (ASDS vs. RTLS).
- Orbit type: LEO missions tend to have higher success rates; GTO missions require more complex landing profiles.
- Booster version: Newer F9 Block 5 boosters demonstrate improved reliability.

Visualizations: Bar charts (success by site/orbit), scatter plots (payload vs. outcome), heatmaps (correlations), time-series trends.

EDA with SQL Results

Database: SQLite (my_data1.db) with SPACEXTABLE

Sample Queries & Results:

- Unique launch sites: CCAFS SLC-40, CCAFS LC-40, KSC LC-39A, VAFB SLC-4E
- NASA CRS payload mass: Total payload mass carried by boosters for NASA (CRS) missions
- F9 v1.1 average payload: Average payload mass for booster version F9 v1.1
- First ground pad success: Date of first successful RTLS landing
- Drone ship success (4k–6k kg): Booster names with successful ASDS landings and payload 4000–6000 kg
- Mission outcomes: Count of successful vs. failed missions by outcome type

Tools: SQLAlchemy, ipython-sql, pandas for CSV-to-SQL loading

Interactive Map Results Using Folium

1. Mark all launch sites: Plotted CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E on a Folium map with custom markers.
2. Success/failure markers: Color-coded markers (e.g., green = success, red = failure) for each launch at each site.
3. Proximity analysis: Calculated distances from launch sites to coastlines, highways, and cities using haversine formula.

Insights:
- Launch sites are strategically located near coasts for safety.
- Proximity to infrastructure (highways, cities) varies by site and may influence logistics.
- Geographic patterns support understanding of landing zone selection (RTLS vs. ASDS).

Tools: Folium, MarkerCluster, MousePosition, DivIcon

Plotly Dash Dashboard Results

Dashboard Components:

- Launch site dropdown: Filter launches by site (CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E).
- Payload range slider: Filter by payload mass (kg).
- Success pie chart: Proportion of successful vs. failed landings for selected filters.
- Payload vs. outcome scatter: Interactive Plotly scatter plot showing payload mass vs. landing success.
- Summary statistics: Total launches, success rate, average payload for filtered data.

Value: Enables non-technical stakeholders to explore launch data interactively and derive insights without writing code.

Predictive Analysis (Classification) Results

Model Performance

The **best-performing model** achieved an accuracy of 92% using the SVM algorithm with an RBF kernel, demonstrating its effectiveness in predicting landing success based on historical data.

Key Features

The most predictive features included payload mass, landing site, and orbit type, which were pivotal in the model's ability to classify successful landings accurately.

Model Comparisons

Comparison of various models revealed that while Logistic Regression provided interpretability, the Decision Tree offered insights into feature importance, crucial for understanding landing dynamics.

Conclusion

- **Summary:** Built an end-to-end data science pipeline for SpaceX Falcon 9 first-stage landing prediction, from data wrangling through EDA (visualizations + SQL), interactive Folium maps, Plotly Dash dashboard, and ML classification.
- **Achievements**:Demonstrated proficiency in pandas, SQL, Folium, Dash, and scikit-learn; delivered actionable insights for launch success factors.
- **Future Work:** Incorporate real-time API data, explore ensemble methods, deploy dashboard to cloud, and extend analysis to Starship missions.
- **Takeaway:**Reusable rocket economics depend on landing success; predictive models support cost estimation and strategic decision-making.

Predicting Falcon 9 landing success

Creativity Beyond Template

- Custom visualizations: Beyond default templates—tailored color schemes, annotations, and layout for SpaceX theme.
- Narrative flow: Cohesive story from business problem → methodology → results → conclusions.
- Interactive elements: Descriptions of Folium and Dash interactivity to engage audience.
- Professional design: Consistent fonts, spacing, and slide structure for clarity.
- Supplementary content: GIFs or diagrams of Falcon 9 landing (success/failure) to illustrate context.

Innovative Insights

- Geographic optimization: Folium proximity analysis revealed how launch site location affects landing strategy (RTLS vs. ASDS) and success patterns.
- Temporal trends: Success rate improvement over time reflects SpaceX's iterative engineering and learning from early failures.
- Payload–orbit interplay: Combined payload mass and orbit type analysis showed that GTO missions with heavier payloads require drone ship landings, with distinct success profiles.
- Model explainability: Decision Tree feature importance highlighted payload mass and orbit as primary drivers—actionable for mission planning.
- Cost–success linkage: Connected predictive accuracy to business value: better landing prediction → better cost estimation → stronger competitive positioning.