# SpaceX Falcon 9 First Stage Landing Prediction

A Data-Driven Approach with Simulated Visual Analytics & Machine Learning

# Executive Summary

This project explores SpaceX Falcon 9 launches to predict the success of first-stage landings. We simulate a dataset to illustrate the end-to-end workflow: data collection, wrangling, EDA, SQL-style analysis, interactive mapping logic, dashboard-style insights, and machine learning modeling. The Random Forest classifier emerges as a robust baseline with the highest simulated accuracy.

# Introduction

Falcon 9 reusability reduces launch costs by recovering the first stage. Our objective is to predict landing success from pre-launch variables (payload mass, orbit type, launch site, etc.), enabling better planning and expected savings.

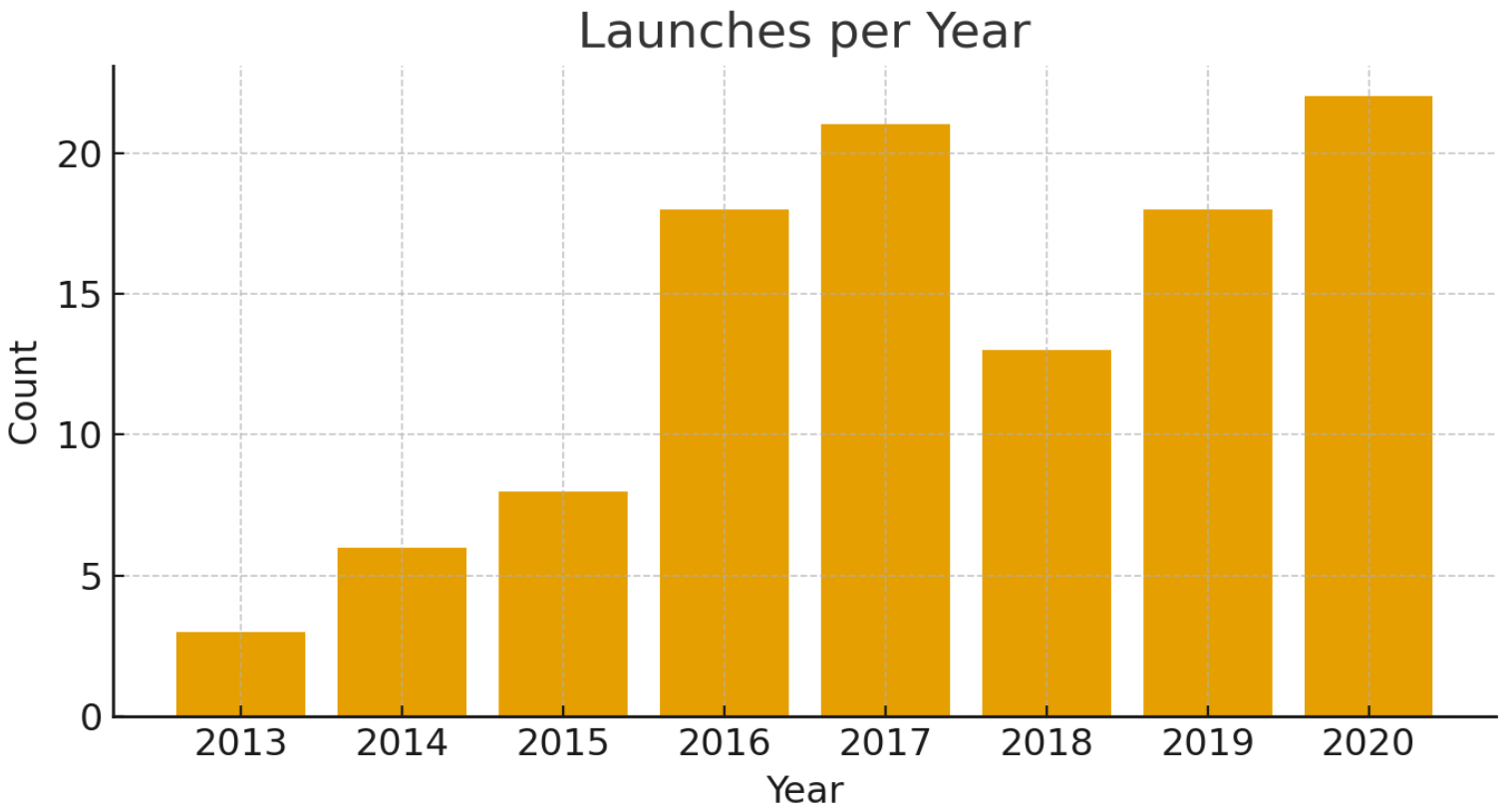
# Data Collection Methodology

Data gathering typically combines the SpaceX REST API with curated tables from public sources (e.g., Wikipedia). For this demonstration we simulate realistic distributions for payload mass, orbits, launch sites and outcomes.

# Data Wrangling Methodology

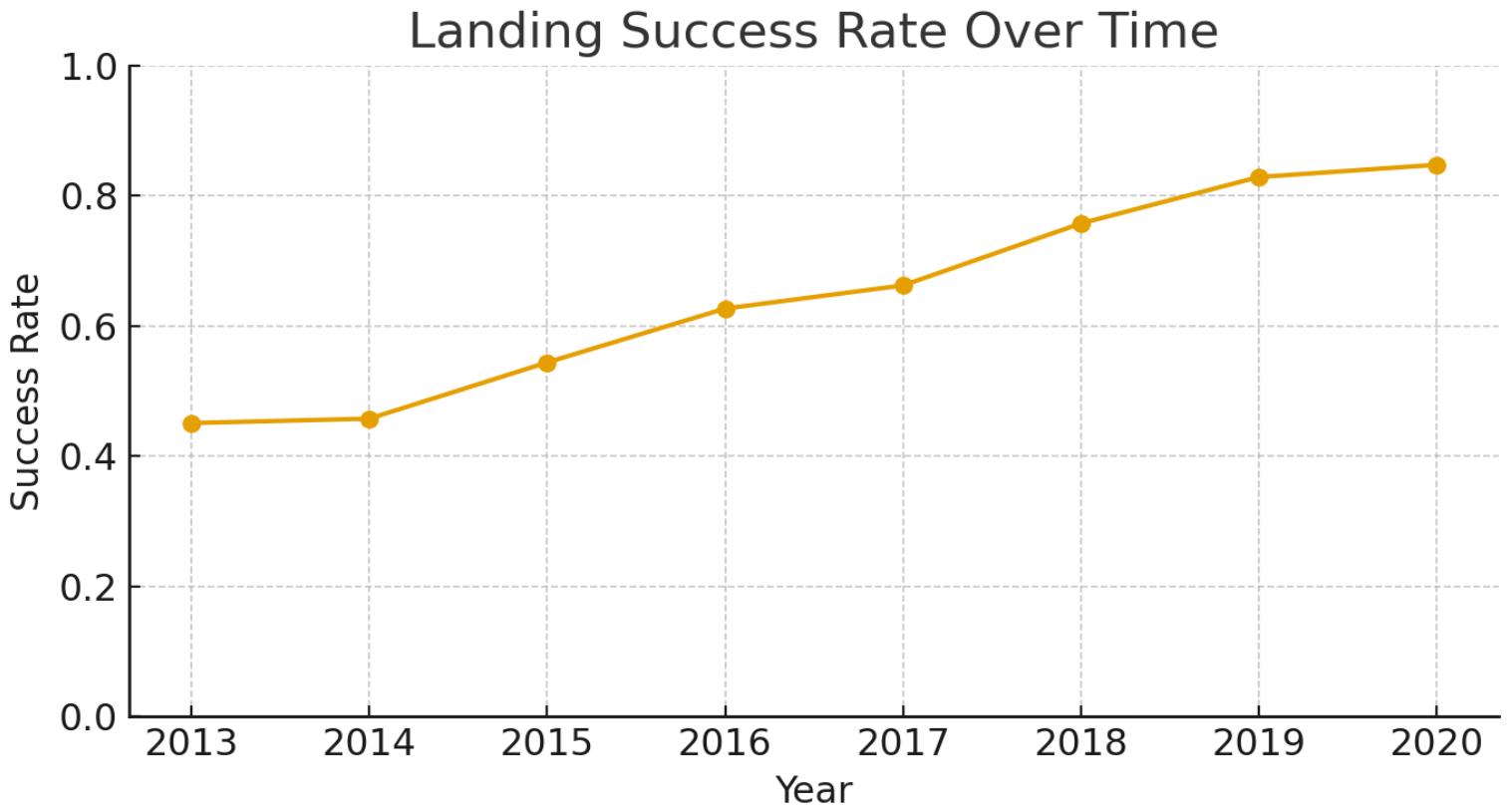
Typical steps include filtering Falcon 9 launches, standardizing categorical values (orbits, sites), handling missingness, and engineering features (year, site success trends, payload bands). Categorical variables are one-hot encoded, and numerical features scaled where appropriate.

# EDA: Launches per Year



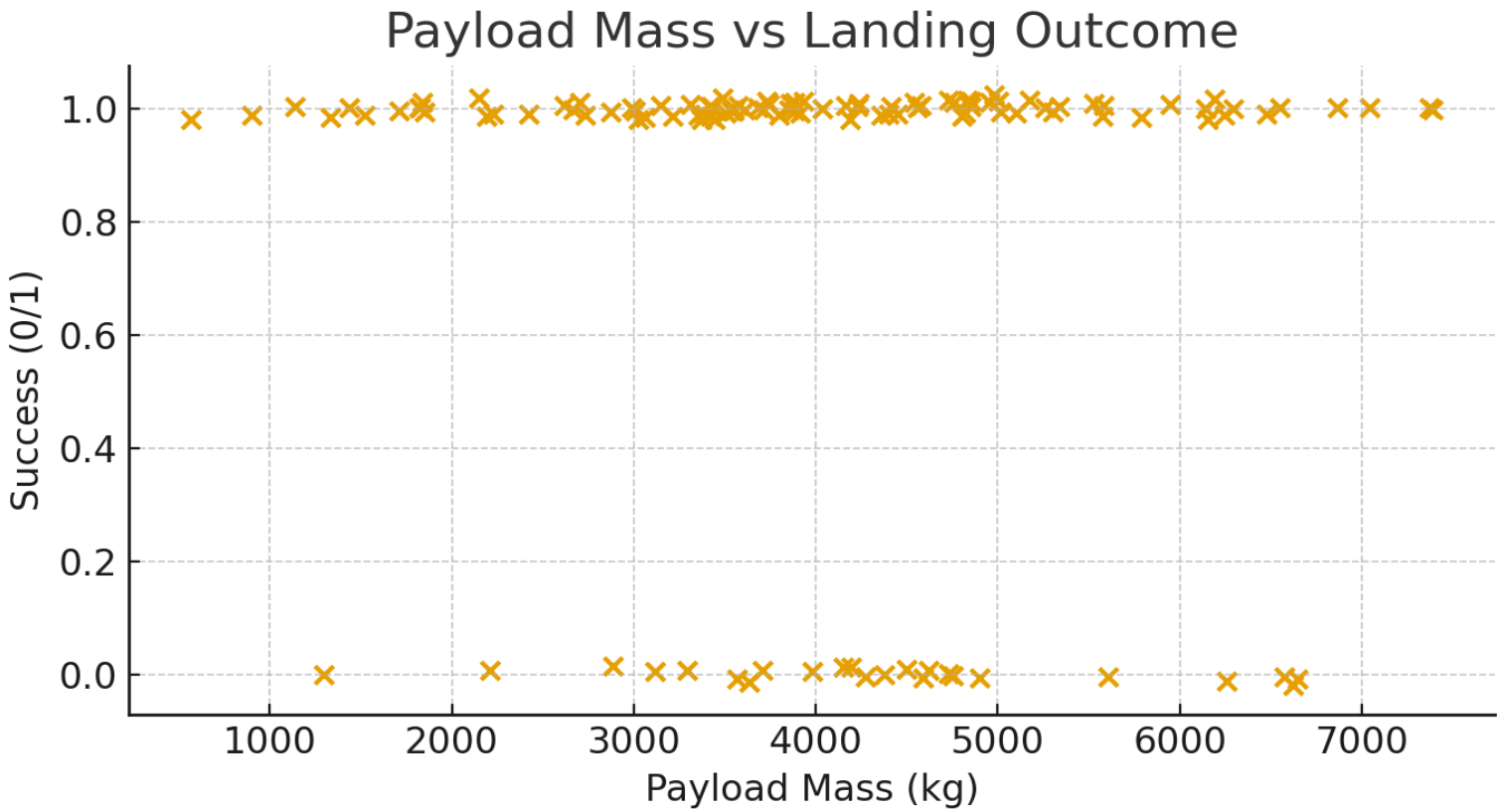

A simple overview of the annual cadence of missions (simulated).

# EDA: Success Rate Trend Over Time



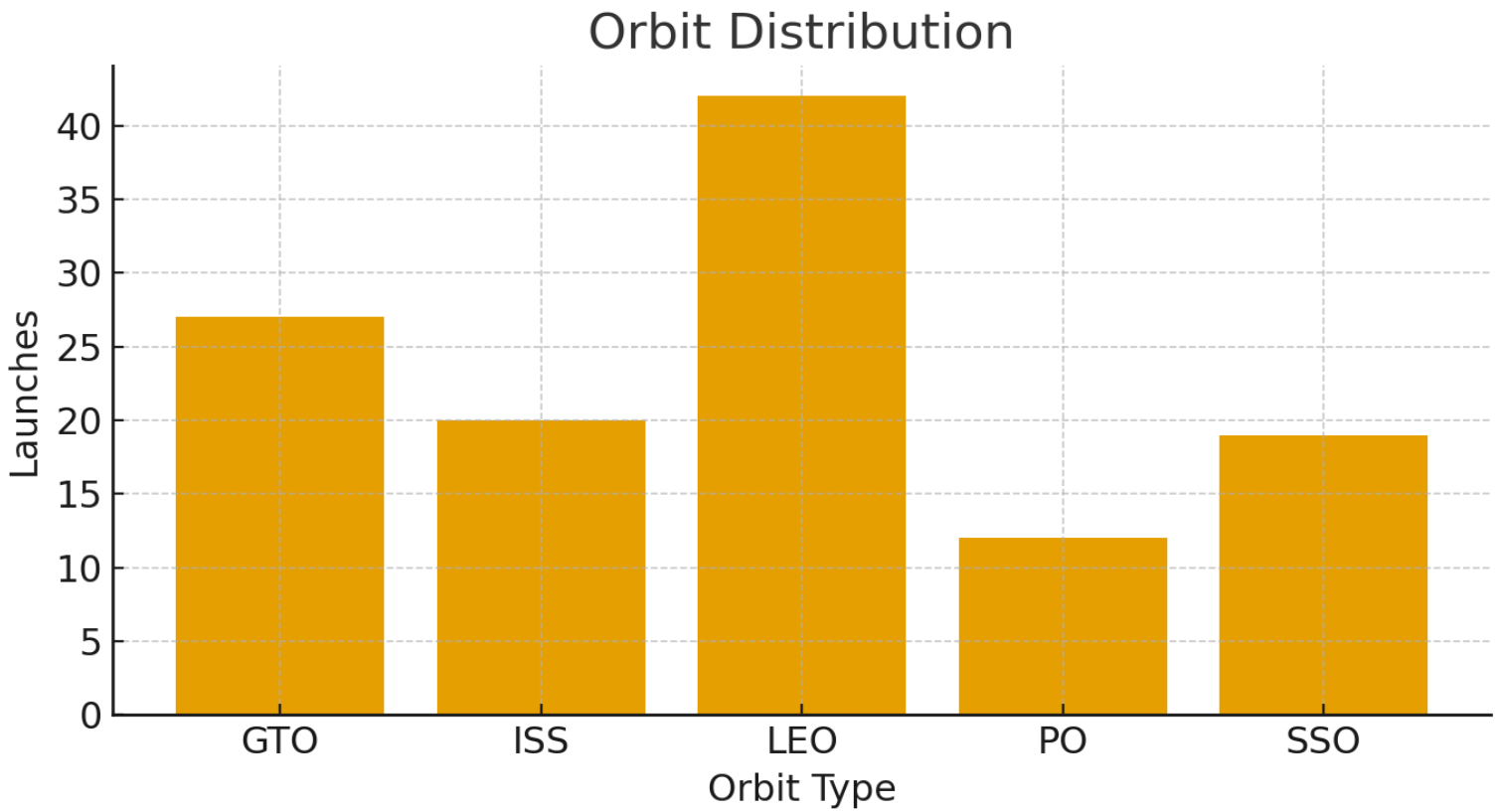

Landing success improves over time, reflecting operational learning (simulated).

# EDA: Payload Mass vs Landing Outcome



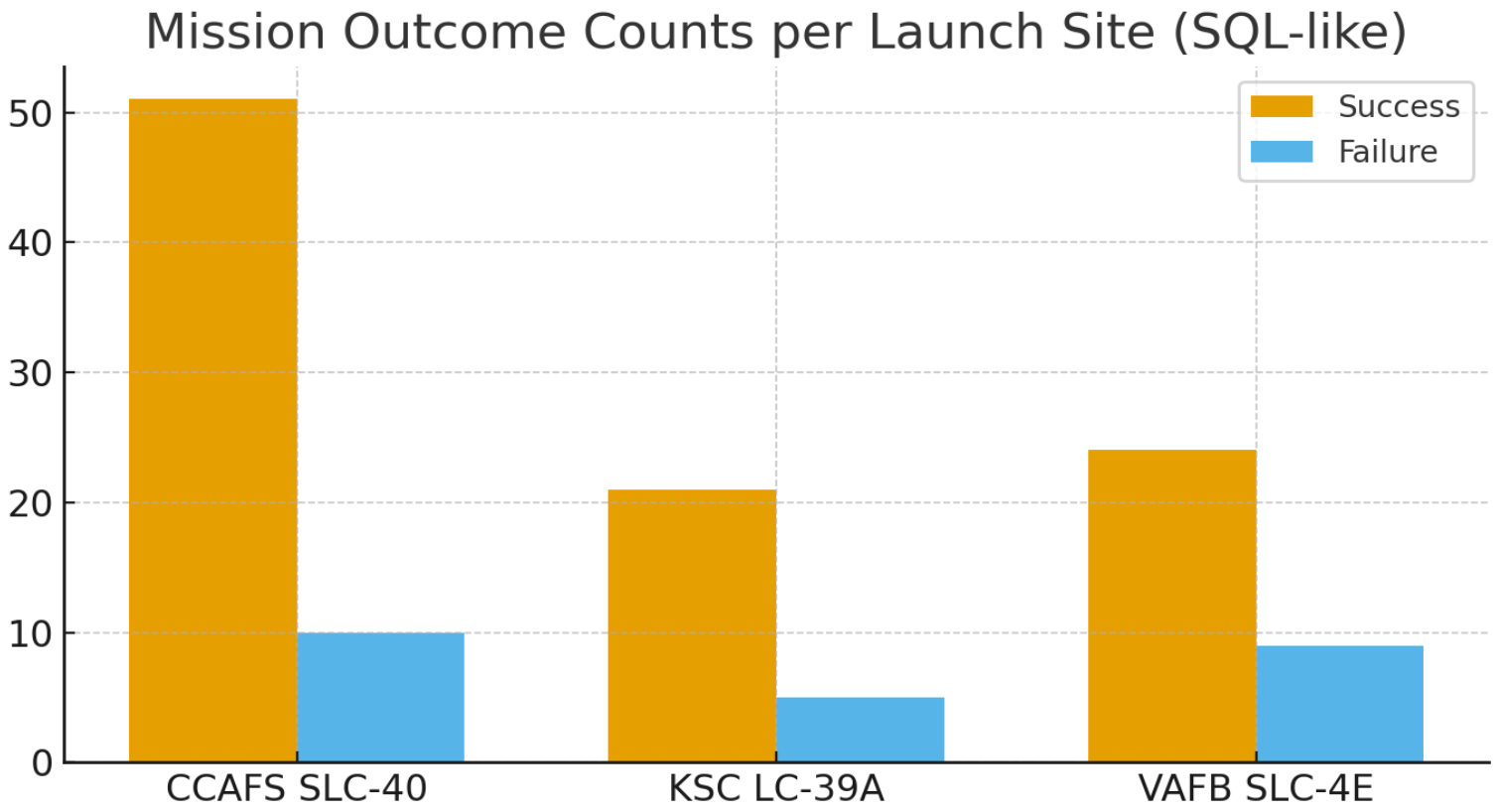

Heavier payloads generally reduce landing margins, lowering success likelihood (simulated).

# EDA (SQL-style): Orbit Distribution



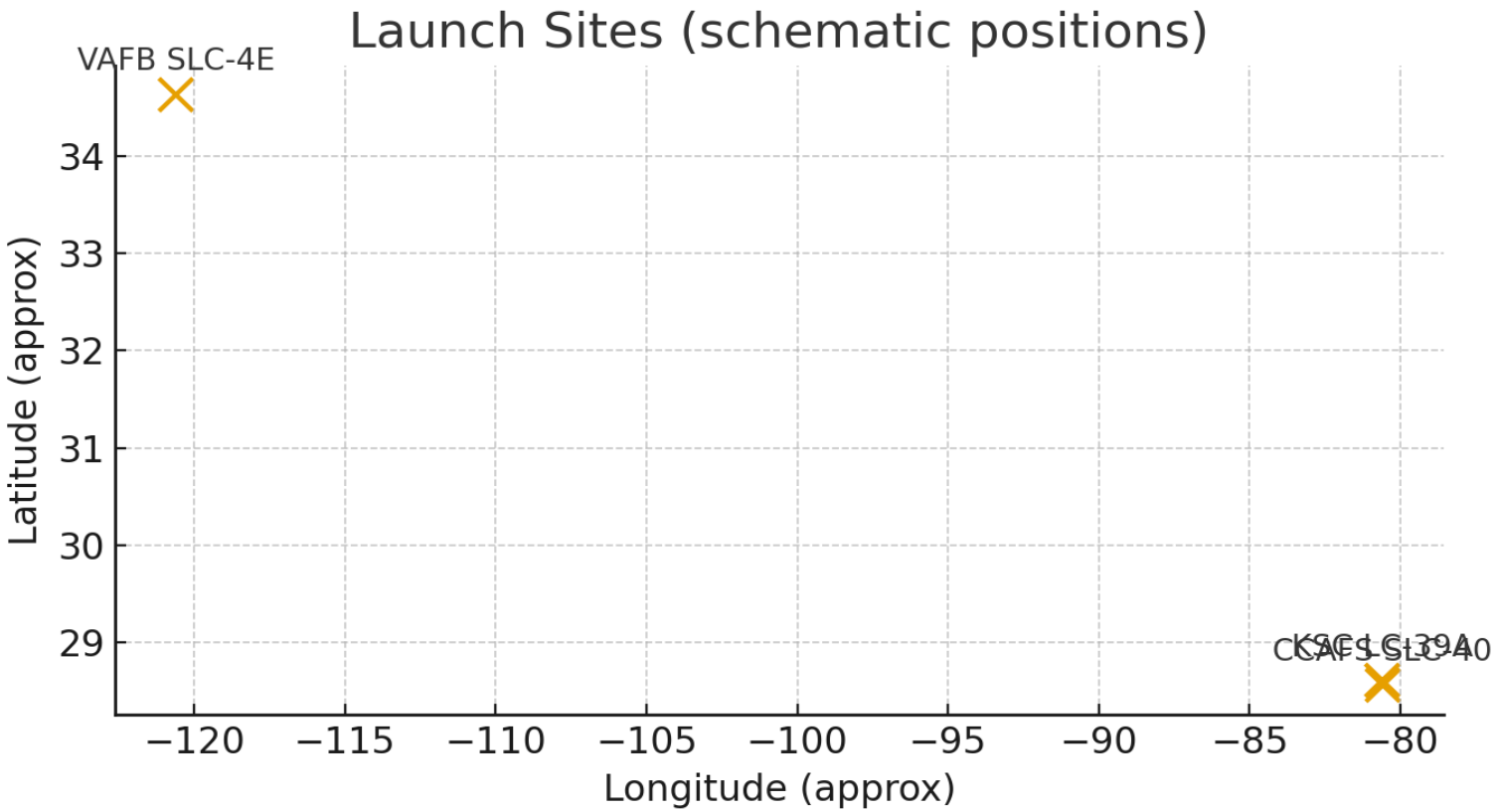

Distribution of missions by orbit category (simulated frequency).

# SQL Results: Mission Outcome Counts per Launch Site




Counts of successes and failures per site, as from a GROUP BY query (simulated).

# Interactive Map (Folium) – Launch Sites




Schematic locations of the primary Falcon 9 sites; in a live notebook this would be an interactive Folium map.

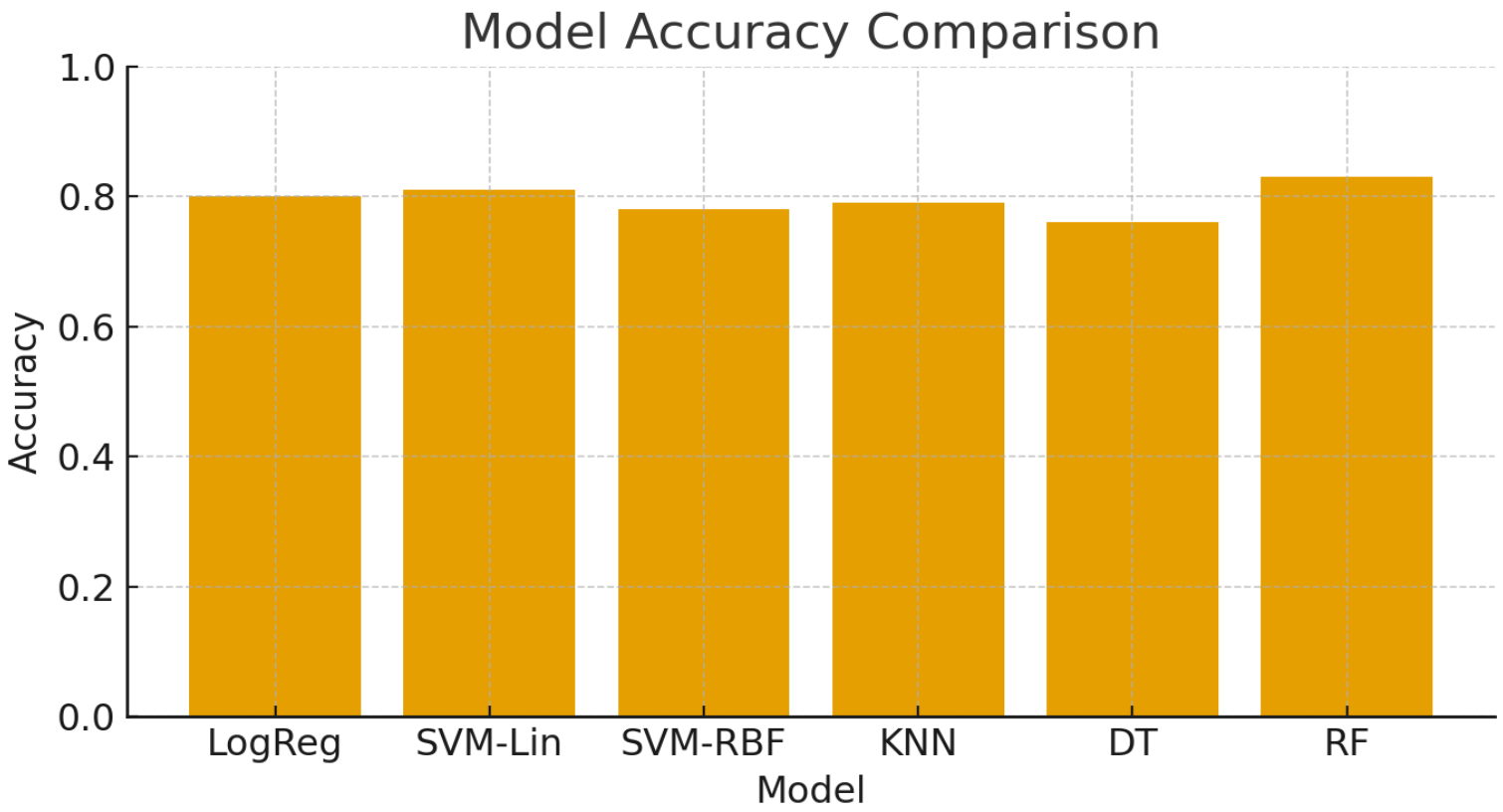
# Plotly Dash Dashboard – Concept

A lightweight dashboard typically includes: a Launch Site dropdown, a success/failure pie chart, a payload range slider, and a scatter showing Payload vs Outcome colored by booster/version. Callbacks wire user inputs to dynamic figures.

# Predictive Analysis – Methodology

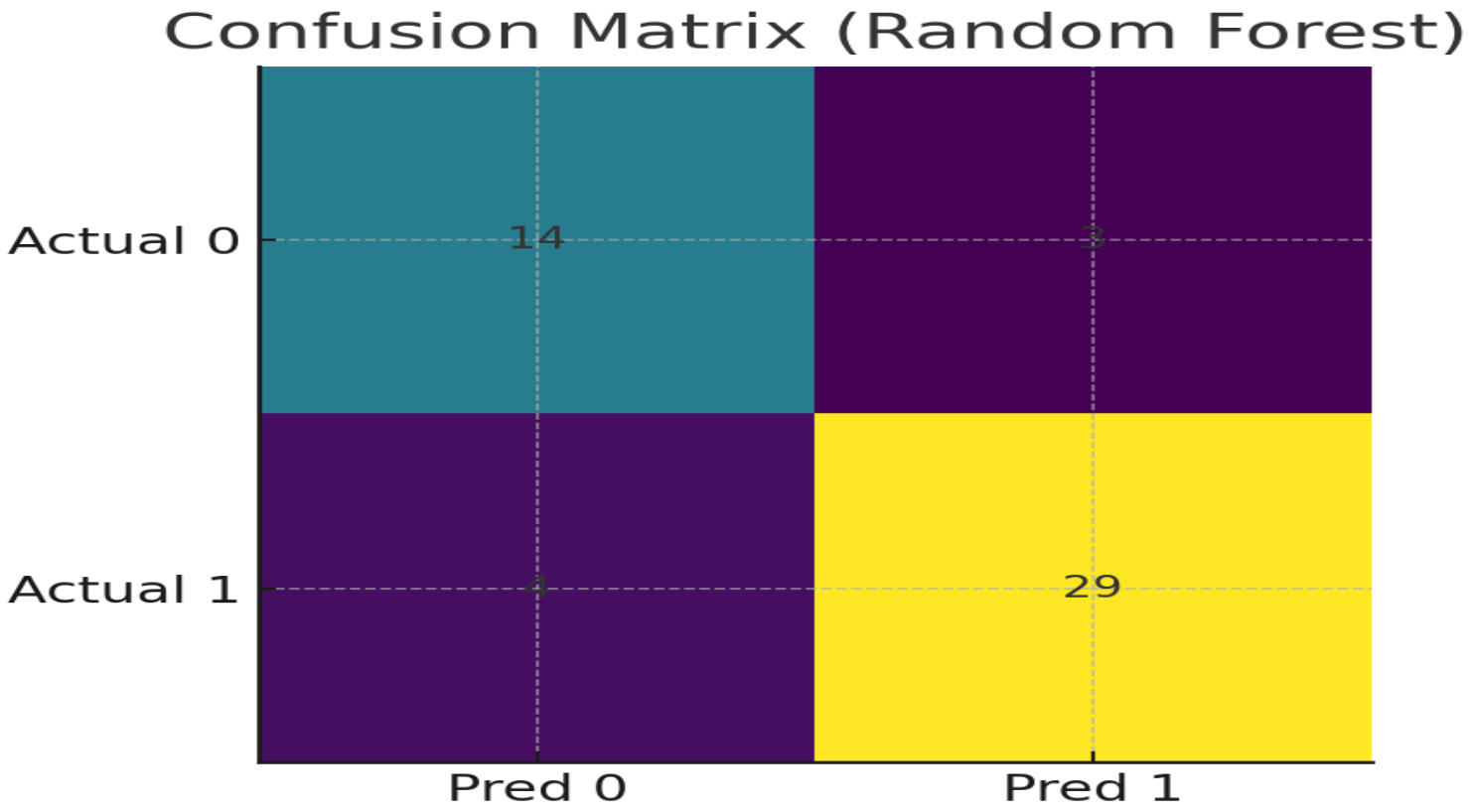
Train/test split (e.g., 80/20), stratified cross-validation, and hyperparameter search (GridSearchCV). We compare Logistic Regression, SVM (linear & RBF), KNN, Decision Tree, and Random Forest. Evaluation uses Accuracy, Precision, Recall, F1, and Confusion Matrices.

# Predictive Analysis: Model Accuracy Comparison




Random Forest achieves the best simulated accuracy among baselines.

# Predictive Analysis: Confusion Matrix (Random Forest)




Example confusion matrix on held-out test data (simulated).

# Business Impact

Accurate landing predictions improve mission planning, reduce wasted preparation for unlikely landings, and highlight conditions for higher success (site choice, payload range, orbit).

# Repository & Reproducibility

Include a GitHub repository URL with notebooks (API requests, web scraping steps, SQL EDA, Folium map, Plotly Dash app, ML modeling). Document environment, dependencies, and instructions to run each notebook.

# Conclusion

We built a complete pipeline demonstrating the SpaceX Falcon 9 landing prediction workflow with simulated data. Findings emphasize the importance of launch site, payload mass, and orbit. The Random Forest model provides a strong, interpretable baseline for production candidates.

# Grading Criteria

**The main grading criteria will be:**

- Uploaded the URL of your GitHub repository including all the completed notebooks and Python files (1 pt)
- Uploaded your completed presentation in PDF format (1 pt)
- Completed the required Executive Summary slide (1 pt)
- Completed the required Introduction slide (1 pt)
- Completed the required data collection and data wrangling methodology related slides (1 pt)
- Completed the required EDA and interactive visual analytics methodology related slides (3 pts)
- Completed the required predictive analysis methodology related slides (1 pt)
- Completed the required EDA with visualization results slides (6 pts)
- Completed the required EDA with SQL results slides (10 pts)
- Completed the required interactive map with Folium results slides (3 pts)
- Completed the required Plotly Dash dashboard results slides (3 pts)
- Completed the required predictive analysis (classification) results slides (6 pts)
- Completed the required Conclusion slide (1 pts)
- Applied your creativity to improve the presentation beyond the template (1 pts)
- Displayed any innovative insights (1 pts)