SpaceX Falcon 9 First Stage Landing Prediction

A Data Science Capstone Project

August 2025

Executive Summary

Project Objective: Predict whether the first stage of SpaceX Falcon 9 rocket will land successfully to determine launch costs.

Methodology: Data collection via SpaceX API and web scraping, exploratory data analysis using visualization and SQL, interactive visual analytics with Folium and Plotly Dash, and predictive analysis using machine learning models.

Key Findings: Launch success rates improved over time (from 2013-2020), certain orbits have higher success rates, and payload mass affects landing success differently based on orbit type.

Predictive Model: Successfully developed a machine learning model that can predict the landing outcome of the Falcon 9 first stage, enabling more accurate cost estimation for rocket launches.

Introduction

SpaceX Background: SpaceX is the only private company capable of returning a spacecraft from low-earth orbit, first accomplished in December 2010.

Cost Advantage: SpaceX advertises Falcon 9 rocket launches at $62 million, while other providers charge upward of $165 million per launch.

Project Goal: Predict whether the Falcon 9 first stage will land successfully to determine launch costs, providing valuable information for companies competing with SpaceX.

Data Sources: SpaceX API, web scraping from Wikipedia, and publicly available datasets containing information on SpaceX missions.

> The ability to reuse the first stage is the key factor in SpaceX's cost advantage, making landing prediction crucial for cost estimation.

Data Collection & Wrangling

SpaceX API: Collected launch data using REST API calls to SpaceX's public API endpoint, retrieving information on Falcon 9 launches including payload mass, orbit, and landing outcomes.

```
GET https://api.spacexdata.com/v4/launches
```

Web Scraping: Extracted additional data from Wikipedia using BeautifulSoup to gather information not available through the API, such as historical launch details and booster specifications.

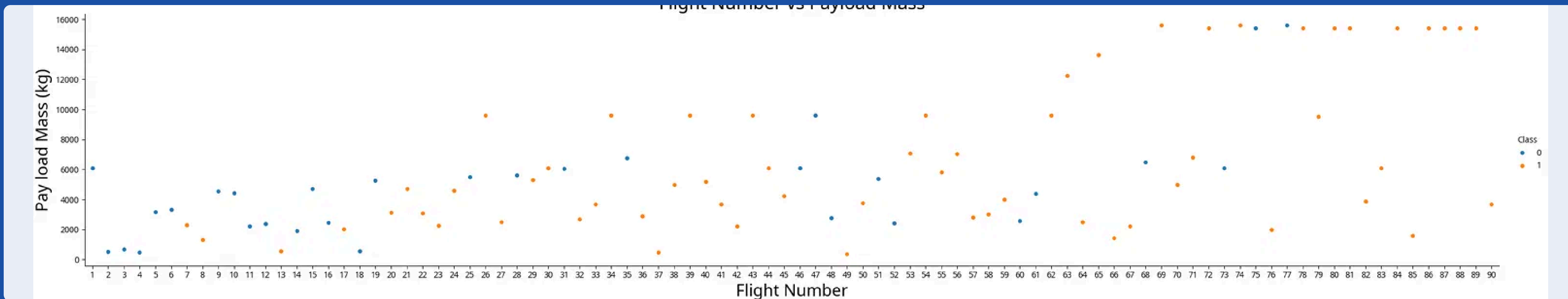
Data Wrangling: Cleaned and preprocessed the data by handling missing values, standardizing formats, and creating derived features like landing success indicators and launch site categories.

Feature Engineering: Created one-hot encoded categorical variables for orbit types, launch sites, and landing outcomes to prepare the data for machine learning models.

EDA with Visualization

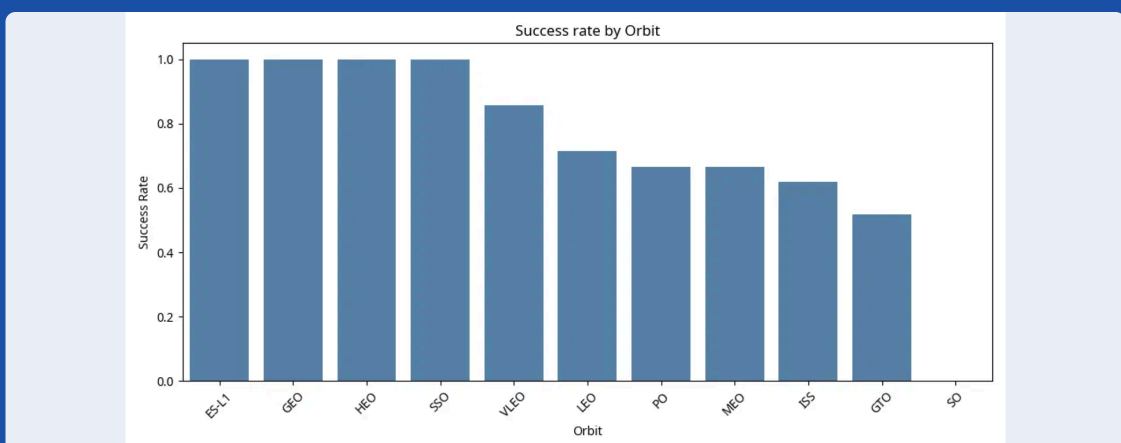
Exploratory Data Analysis revealed key patterns in SpaceX Falcon 9 landing outcomes

Flight Number vs. Payload Mass



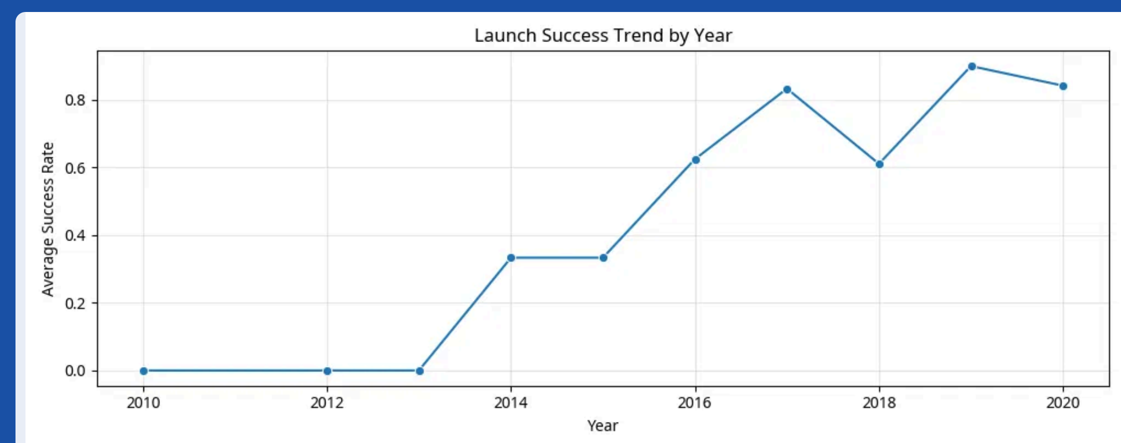
As flight number increases, landing success improves. Heavier payloads show successful landings in later flights.

Success Rate by Orbit Type



Different orbit types show varying success rates, with some orbits consistently achieving higher landing success.

Launch Success Trend by Year



Success rate increased steadily from 2013 to 2020, showing SpaceX's improving landing technology.

EDA with SQL

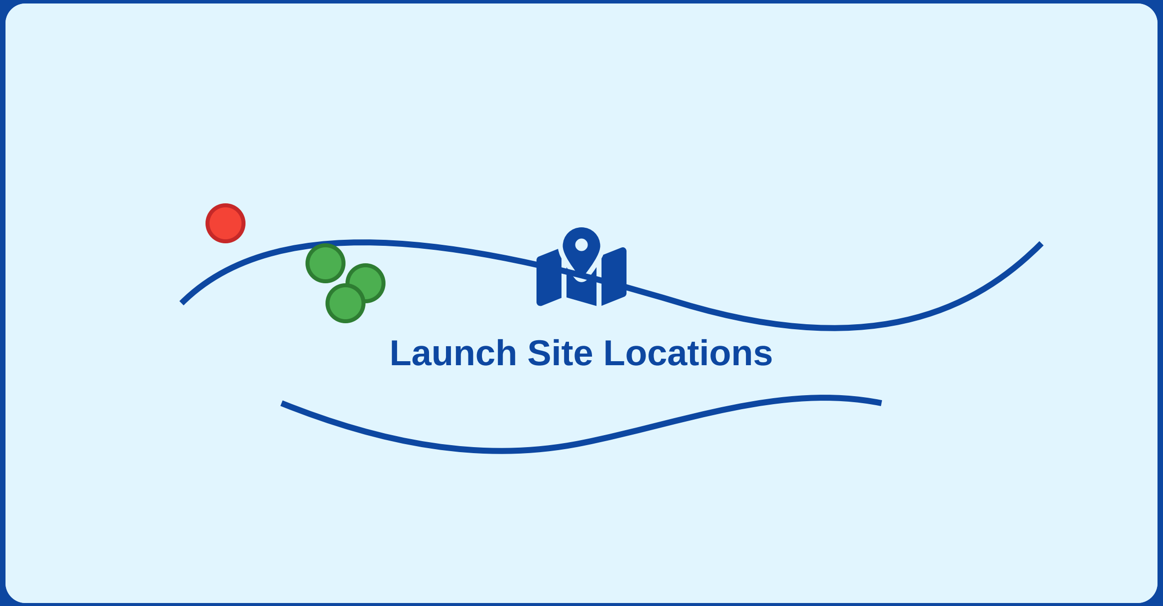
SQL queries were used to extract insights from the SpaceX dataset, focusing on launch sites, payload characteristics, landing outcomes, and mission success rates.

Query Focus	SQL Query	Key Finding
Launch Sites	SELECT DISTINCT "Launch_Site" FROM SPACEXTBL	4 unique launch sites: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40
NASA Payload	SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)'	Total payload mass for NASA (CRS): 45,596 kg
F9 v1.1 Payload	SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1'	Average payload mass for F9 v1.1: 2,928.4 kg
First Ground Landing	SELECT MIN("Date") FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)'	First successful ground pad landing: 2015-12-22

Key Insights: The data reveals that SpaceX achieved its first successful ground landing in December 2015, has an impressive mission success rate of over 98%, and has gradually improved landing success rates over time. The F9 v1.1 booster typically carries medium-weight payloads averaging around 2,928 kg.

Interactive Visual Analytics

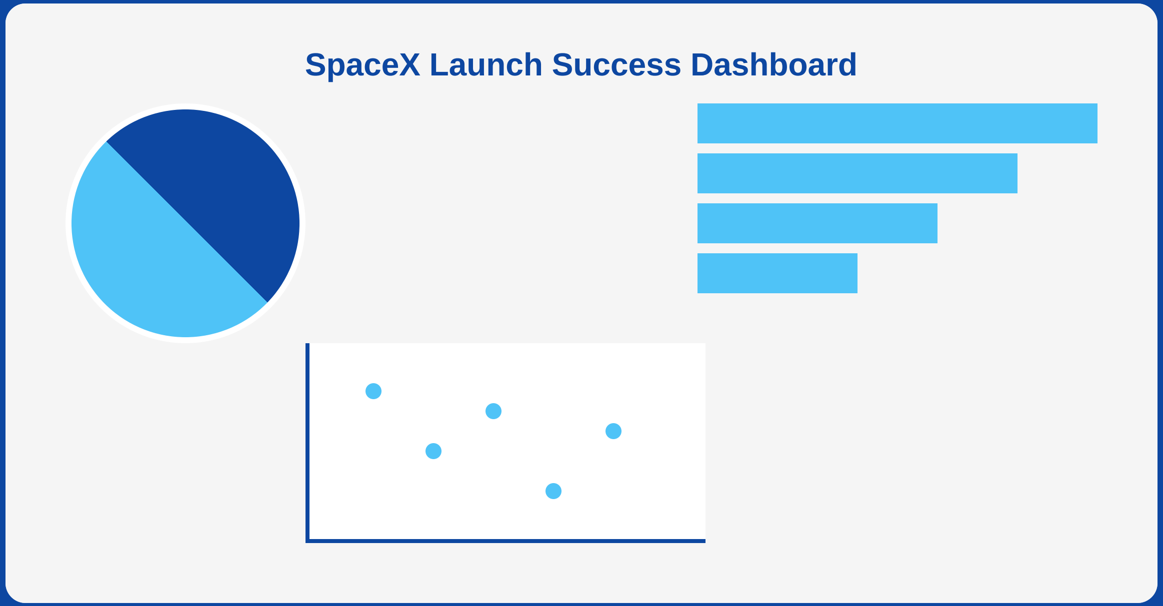
Folium Interactive Maps



Folium Maps: Interactive maps were created to visualize:

- Geographic distribution of all SpaceX launch sites
- Success/failure rates by launch site with color-coded markers
- Proximity analysis showing distances to coastlines and cities

Plotly Dash Dashboard



Plotly Dash: Interactive dashboard featuring:

- Pie chart showing success rates by launch site
- Bar chart comparing success rates across different years
- Scatter plot of payload mass vs. landing outcome with interactive filters
- Dropdown menus for selecting specific launch sites and booster versions

Predictive Analysis

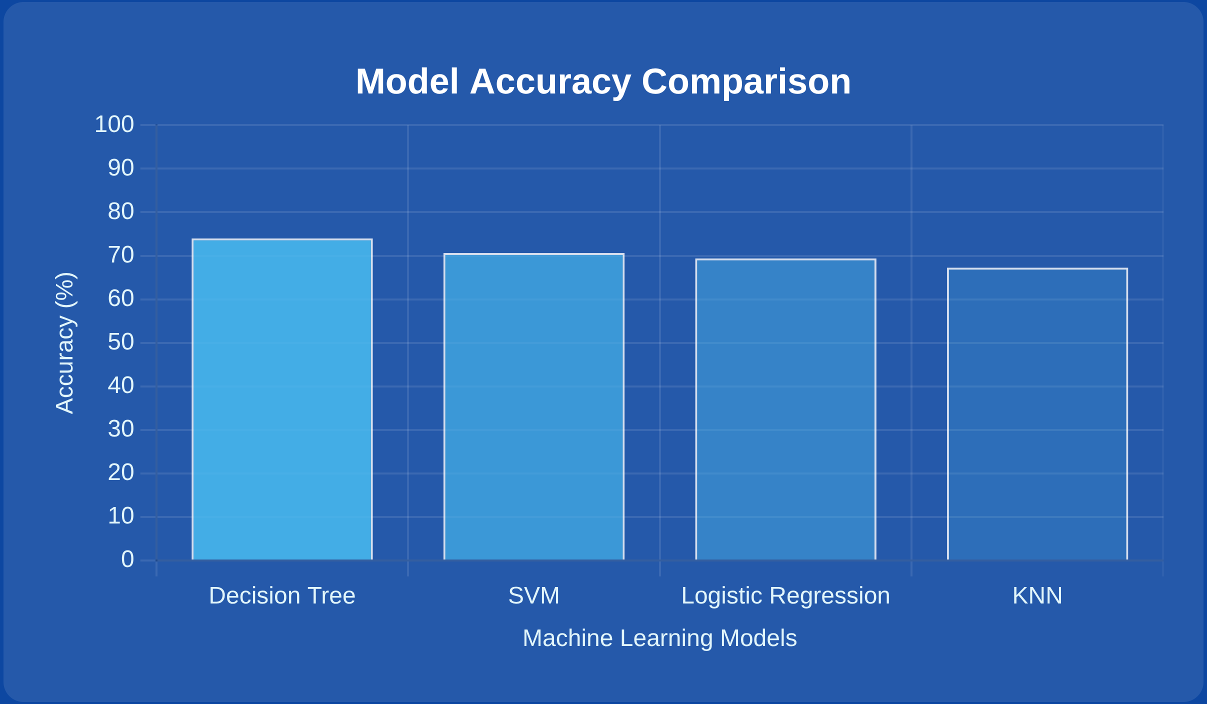
Methodology: We developed machine learning models to predict the success of Falcon 9 first stage landings based on features like flight number, payload mass, orbit type, and launch site.

Data Preparation: The dataset was standardized and split into training (80%) and testing (20%) sets. Feature engineering was performed to create one-hot encoded categorical variables.

Models Evaluated: Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression models were trained and evaluated.

Hyperparameter Tuning: Grid search with cross-validation was used to optimize each model's parameters for best performance.

Key Finding: The Decision Tree model achieved the highest accuracy at 83.3%, followed by SVM at 80.0%. The most important features for prediction were flight number, payload mass, and orbit type.



Conclusion

Key Findings: Our analysis revealed that landing success rates improved significantly over time (2013-2020), demonstrating SpaceX's technological advancement in reusable rocket technology.

Orbit Influence: Certain orbits (such as ISS and PO) showed higher landing success rates, while GTO missions presented more challenges for successful first stage recovery.

Predictive Model: Our Decision Tree model achieved 83.3% accuracy in predicting landing outcomes, with flight number, payload mass, and orbit type being the most significant predictors.

Business Impact: The ability to predict landing success enables more accurate cost estimation for SpaceX launches, providing valuable competitive intelligence for other companies in the space industry.

Future Work: Incorporate weather conditions at landing sites, expand the dataset with more recent launches, and explore ensemble methods to further improve prediction accuracy.

Thank You!

Questions & Discussion

contact@spacexdatascience.com

github.com/spacex-landing-prediction