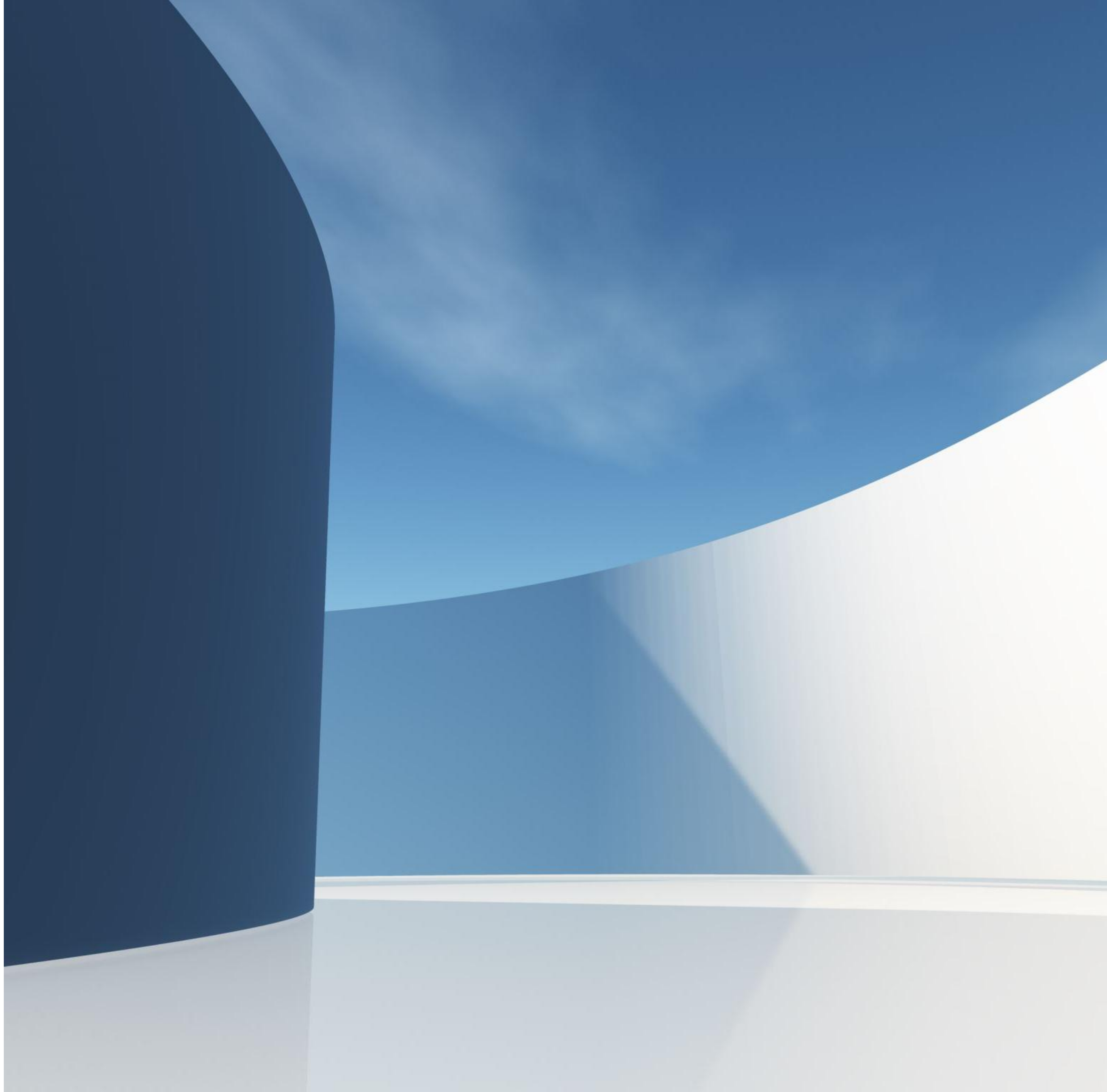
Title:
Predicting SpaceX Launch Success Using Data Science
Subtitle:
By Abdul Wahab
IBM Applied Data Science Capstone
June 2025



Executive Summary

This project analyzes SpaceX launch data to discover patterns behind successful rocket landings. Using API data, web scraping, and SQL, we cleaned and explored the dataset. We then created visualizations and built a predictive model. The main aim is to help predict whether a future Falcon 9 launch will land successfully.

Introduction

SpaceX’s Falcon 9 is a reusable rocket that lowers the cost of space access.

Understanding what contributes to a successful launch is critical.

This project explores SpaceX’s launch data to derive insights and build predictive models.

Data Collection & Wrangling

Data sources: SpaceX API, web scraping, and datasets provided in the course.

Cleaned and transformed data to handle missing values and standardized formats.

Combined datasets: Launch records, launch site info, booster versions, and payload data.

Exploratory Data Analysis (EDA) - Part 1

Explored distributions of payload mass, launch success, orbit types, and booster versions.
Identified key trends between payload size and mission outcome.

EDA & Visual Analytics
Used Matplotlib and Seaborn to visualize:
Success rate vs. launch site
Payload mass vs. launch outcome
Booster version impact on mission result

EDA with SQL

Used SQL to:

- Count successful launches by site
- Retrieve missions with highest payloads
- Join launch and booster data for analysis

Example query:

```
SELECT LaunchSite, COUNT(*) FROM Launches
WHERE Outcome='Success' GROUP BY LaunchSite;
```

Interactive Map with Folium

Built an interactive Folium map showing:

Launch site locations

Mission outcomes by site

Included tooltips with launch info for user exploration.

Plotly Dash App

Developed a Dash dashboard showing:
Pie charts of launch success by site
Scatter plots of payload vs. success
Interactive filters for launch site and payload range.

Predictive Analysis Methodology

Features used: Payload mass, orbit, booster version, and launch site.
Preprocessing: One-hot encoding, normalization.
Models: Logistic Regression, SVM, Random Forest, KNN.

Predictive Analysis Results

Evaluated models using accuracy, precision, and F1-score.

Best model: Random Forest (accuracy ≈ 85%)

Insights:

Payload mass and booster version were strong predictors.

Some launch sites consistently performed better.

Conclusion

Machine learning can help SpaceX assess launch risk in advance.

Random Forest proved most effective among tested models.

Future work: include weather data and refine model features.

Creativity & Innovation

Added custom visuals and color schemes for clarity.
Explained key metrics in simple terms for better interpretation.
Provided extra analysis using feature importance plots and correlation matrices.

Innovative Insights

Found that success rates improve significantly with newer booster versions.

Launch sites near equator tend to have slightly higher success rates.

Recommended focusing future launches from high-success sites using optimal payload ranges.