

# Predicting SpaceX Falcon 9 First Stage Landing Success

- Data Science Capstone Project
- Bikash Kushwaha
- Date : 6/19/2025

# Executive Summary

- Goal: Predict Falcon 9 first-stage landing success.
- Methods: Data collection via SpaceX API & Wikipedia, data wrangling, EDA, interactive visualizations, ML.
- Outcome: Identified key factors influencing landing success, built accurate models.

# Introduction

- Brief on SpaceX and reusable rockets.
- Importance of predicting landing outcomes.
- Overview: Data collection → EDA →  
Interactive Dashboards → Predictive Analysis.

# Data Collection & Wrangling Methodology

- Data from SpaceX API & Falcon 9 Wiki page.
- Normalized JSON into pandas DataFrames.
- Cleaned missing values, transformed features, merged datasets.

# EDA & Interactive Visual Analytics Methodology

- Used pandas, SQL, matplotlib, seaborn for EDA.
- Created Folium maps for launch sites.
- Built Plotly Dash dashboards for interactive analysis.

# Predictive Analysis Methodology

- Applied Logistic Regression, Decision Tree, SVM.
- Performed hyperparameter tuning.
- Evaluated with accuracy, precision, recall, confusion matrices.

# EDA Results: Launch Sites

- Bar chart: Launches per site.
- Pie chart: Success vs fail per site.
- Insight: CCAFS SLC 40 and KSC LC 39A had the most launches.

# EDA Results: Payload Mass Impact

- Scatterplot: Payload Mass vs Outcome.
- Observation: Payload 2000–4000 kg had better success rates.



# EDA Results: Orbits

- Bar chart: Launches by orbit type.
- Insight: GTO & LEO most common; success varied by orbit.

# EDA with SQL Results

- SQL queries for:
  - - Total launches
  - - Success by site
  - - Launches per year
- Results shown as SQL snippets and table outputs.

# Interactive Map with Folium

- Map showing launch sites with outcome markers.
- Insight: Coastal proximity influences landing success.

# Plotly Dash Dashboard

- Screenshot:
- - Dropdown for launch sites
- - Slider for payload
- - Dynamic pie & scatter charts
- Insight: Enabled interactive EDA.

# Predictive Analysis Results

- Decision Tree: 83.33%
- Logistic Regression: 80%
- SVM (RBF): 83.33%
- Confusion matrix for best model.

# Conclusion

- Decision Tree & SVM best performers.
- Payload Mass & Launch Site key predictors.
- Recommendation: Adjust operations based on predictions.
- Future work: Deep learning models, expand data.

# Creativity & Insights

- Added trend analysis.
- Correlation heatmaps.
- Proposed operational recommendations.
- Visual: Success rate over years, Correlation matrix.