Title: Predicting SpaceX Launch Success Using Data Science

Subtitle: *IBM Applied Data Science Capstone Project*

Presented by: *Abdul Wahab*

Date: *June 2025*

Executive Summary

Goal: Analyze SpaceX Falcon 9 launches to identify factors that influence success.
Techniques Used: EDA, SQL, Visualization, Machine Learning (Logistic Regression, SVM, Random Forest, etc.).
Outcome: Built a model to predict launch success and derived strategic insights.

Introduction

- **Why SpaceX?** It represents modern space exploration and commercial launch innovation.
- **Problem:** Launch success is not guaranteed—understanding patterns is valuable.
- **Objective:** Use data science to analyze historical launch data and build a predictive model.

Data Collection

Sources
SpaceX launch data from SpaceX API and Kaggle
Launch site coordinates for Folium visualization
Tools: Python, Pandas, requests, BeautifulSoup.

Data Wrangling

Cleaned null values
Formatted date/time fields
Created binary target column for success/failure
Joined coordinates with site names for map visualization

EDA with SQL – Methodology

Imported data into SQLite
Performed SQL queries to:
Count launches per year
Find success rates
Explore launch site performance

EDA with SQL – Results

Example result:
In 2020: 24 launches, 22 successes → 91.6% success
SQL output visualized in bar graphs

Interactive Map (Folium)

Used folium to plot launch site locations
Color-coded circles: green = success, red = failure
Popup tooltips for site info
Easy comparison between sites

EDA with Python – Methodology

Used pandas, matplotlib, seaborn
Explored relationships between:
Payload mass
Booster version
Orbit type
Launch site

EDA with Python – Results

Heavier payloads had lower success rates
Booster version B5 showed higher success
Certain orbits had consistently higher success rates

Interactive Dashboard (Plotly Dash)

Dropdown to select launch site
Pie chart displays success/failure ratio
Built with Dash, dcc, plotly.express

Predictive Analysis – Methodology

Features:
Year
Payload mass
Booster version
Orbit
Models used:
Logistic Regression
SVM
K-Nearest Neighbors
Random Forest

Predictive Analysis – Results

Random Forest gave the best accuracy: ~90%

Feature importance:

Booster Version

Payload Mass

Orbit Type

Conclusion

Machine learning can predict launch success with high accuracy
Feature analysis helps understand key factors
Insights can assist future mission planning

Extra Insights

Booster B5 had the highest success rates
KNN model was weakest among all
Potential to extend project with real-time launch forecasts

Acknowledgments & Tools Used

Tools: Python, Pandas, SQLite, Folium, Plotly Dash, Scikit-learn

Thanks to: IBM, Coursera, SpaceX data contributors