

Applied Data Science Project



© IBM Corporation. All rights reserved.



EXECUTIVE SUMMARY



- Objective: Predict SpaceX Falcon 9 first-stage landing success.
- Data sources: SpaceX API, web scraping, SQL datasets.
- Methods: EDA, SQL analysis, visualization, ML classification
- Key result: Falcon 9 Block 5 and KSC LC-39A show highest success rates
- Business impact: Improved mission planning and cost optimization



INTRODUCTION



- SpaceX aims to reduce launch costs through booster reuse
- Predicting landing success improves operational efficiency
- This project analyzes historical launch data to identify success drivers
- Audience: SpaceX management and mission planners



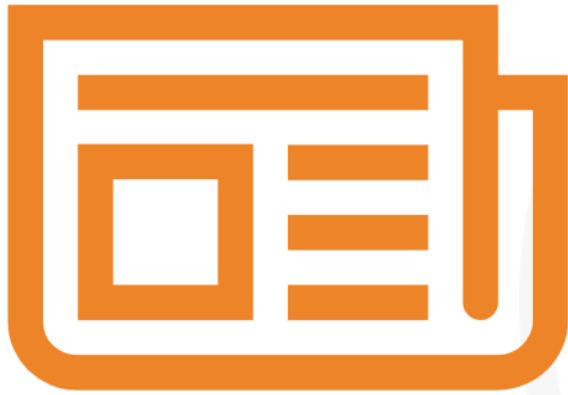
DATA COLLECTION METHODOLOGY



- SpaceX REST API (launch, rocket, payload data)
- Web scraping from Wikipedia
- CSV datasets:
 - dataset_part_1.csv
 - dataset_part_3.csv
 - spacex_web_scraped.csv



DATA WRANGLING METHODOLOGY



- Removed missing and inconsistent values
- Converted categorical variables
- Created target variable: Class (1 = Success, 0 = Failure)
- Merged multiple datasets

EXPLORATORY DATA ANALYSIS (EDA) OVERVIEW



- Distribution of launches by site
- Payload mass vs success
- Booster version impact
- Temporal trends

LAUNCH SUCCESS BY SITE



- KSC LC-39A has highest success count
- CCAFS follows closely



PAYLOAD MASS VS SUCCESS



- Payloads between 4000–6000 kg show higher success
- Very low payloads show lower success



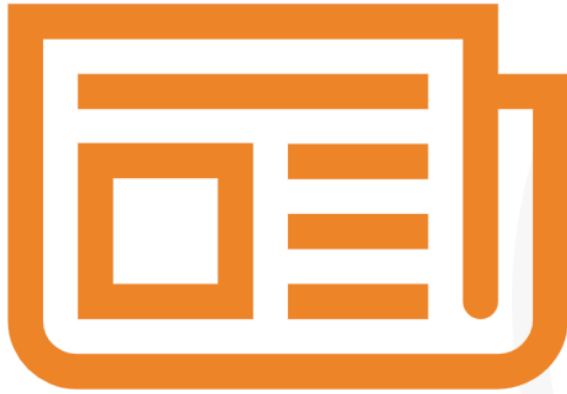
BOOSTER VERSION ANALYSIS



- Falcon 9 Block 5 has highest reliability
- Earlier versions show more failures



YEARLY TRENDS & IMPROVEMENTS



- Success rate increases over time
- Indicates learning curve and technology maturity



SQL ANALYSIS – LAUNCH SUCCESS



- SQL used to aggregate success counts
- Grouped by:
 - Launch site
 - Booster version
 - Payload range



SQL INSIGHTS



- KSC LC-39A highest success rate
- Block 5 dominates success
- Medium payloads outperform extremes



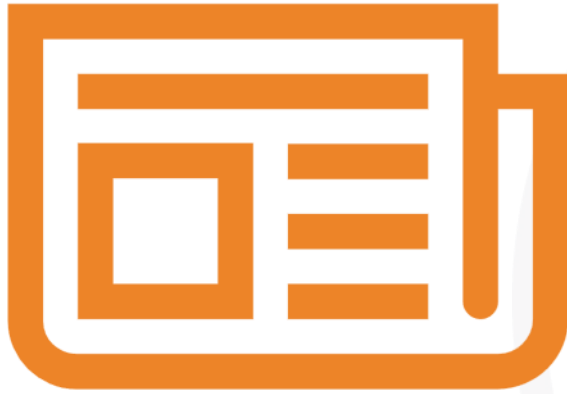
LAUNCH SITE LOCATION ANALYSIS



- All launch sites near coastlines
- Proximity to highways, cities, railways
- Improves logistics and safety



MACHINE LEARNING METHODOLOGY



- Models used:
 - Logistic Regression
 - SVM
 - Decision Tree
 - KNN
- Evaluation metrics:
 - Accuracy
 - Confusion Matrix

ML MODEL PERFORMANCE



- Decision Tree performed best
- Key features:
 - Payload mass
 - Booster version
 - Launch site



RECOMMENDATIONS & INNOVATION



- Prioritize Block 5 boosters
- Optimize payload mass ranges
- Use predictive model pre-launch
- Innovation:
 - Combined SQL + ML + geospatial analysis
 - End-to-end analytics pipeline



CONCLUSION



- Falcon 9 Block 5 is most reliable
- Launch site and payload significantly impact success
- ML models can guide mission planning