SpaceX Launch Analysis Capstone Project

- Name: Stefan McIntyre
- Date: July 2025

Executive Summary

- Objective: Analyze historical SpaceX launch data.
- Approach: Combined API calls.
- Outcome: Identified launch trends.
- Tools Used:
  - Python
  - Pandas
  - SQL
  - Folium

Introduction

- SpaceX revolutionized private aerospace.
- Questions:
  - What factors influence success?
  - Which sites are most reliable?
  - Can we predict outcomes?

Data Collection Methodology

- Sources: SpaceX REST API + Wikipedia
- Flow:
  - GET requests to API
  - Parse JSON
  - Scrape Wikipedia
  - Merge data
- GitHub: API Collection notebook

Data Wrangling

- Processed launch data:
  - Converted types, imputed values
  - Encoded categories
- Flow:
  - Inspect
  - Clean
  - Engineer features
  - Export clean data

EDA – Visual Analysis

- Plots:
  - Flight vs Launch Site
  - Payload vs Launch Site
  - Orbit Success Rate
  - Launch Trends
- GitHub: EDA Visuals

EDA – SQL Queries

- Queries:
  - Unique sites
  - NASA payload
  - Avg payload by booster
  - Landing outcomes
- GitHub: SQL Notebook

Interactive Maps – Folium

- Created:
  - Site markers
  - Outcome colors
  - Distance overlays
- GitHub: Folium Notebook

Interactive Dashboards – Plotly Dash

- Elements:
  - Site success pie chart
  - Payload vs Outcome
  - GitHub: Dash App

Predictive Analytics (Classification)

- Models: LR, SVM, RF, DT
- Best: Decision Tree (83%)
- Metrics:
  - Precision: 0.85
  - Recall: 0.82
  - Confusion Matrix: 180 TP, 10 FN

Results & Conclusion

- Top Sites: CCAFS SLC 40, KSC LC 39A
- Payload trends
- Decision Tree performed best
- Future: Add weather, mission type

Appendix

- Code Snippets
- Extra Charts
- GitHub: Full repo