MAY 2025

SPACEX DATA SCIENCE PROJECT

Predicting Falcon 9 Landing Success

Presented by Hanifah Gladis Amalia

EXECUTIVE SUMMARY

Objective:

The objective of this project is to predict the success of Falcon 9's first stage landing. Successful landings are crucial for reducing costs and achieving SpaceX's mission of reusable rockets.

Methods:

We performed data collection from SpaceX API, web scraping, data wrangling, exploratory data analysis (EDA), interactive visualization, and predictive modeling using machine learning algorithms.

Results:

Logistic Regression demonstrated the highest accuracy in predicting landing success (83.33%). Interactive dashboards were built to visualize the key metrics and launch outcomes.

INTRODUCTION

Background:
SpaceX aims to make space travel more affordable by achieving reusable rockets through successful landings of Falcon 9's first stage.

Problem Statement:
Identifying the factors that significantly influence the success of Falcon 9's first stage landing.

Make it interactive
Analyze the data from past launches to build a predictive model that determines the landing success.

DATA COLLECTION AND WRANGLING

Data Sources:

- SpaceX API: Collected launch data including flight number, payload mass, orbit, landing outcome, and more.
- Web Scraping: Gathered additional data about launch sites and geographical information.
- SQL Database: Stored structured data for efficient querying and analysis.

Data Cleaning

- Addressed missing values and inconsistent data types.
- Standardized variable formats.
- Removed duplicates and irrelevant columns.

EDA AND INTERACTIVE VISUAL ANALYTICS – PART 1

Exploratory Data Analysis (EDA):

- Distribution of launch success rate based on orbits.
- Analysis of payload mass distribution and its correlation with landing success.
- Visualized with histograms, scatter plots, and box plots.
- Tools: Pandas, Seaborn

Key Insights:

- Orbits LEO and ISS have a higher success rate.
- Heavier payloads tend to have a lower success rate.

EDA AND INTERACTIVE VISUAL ANALYTICS – PART 2

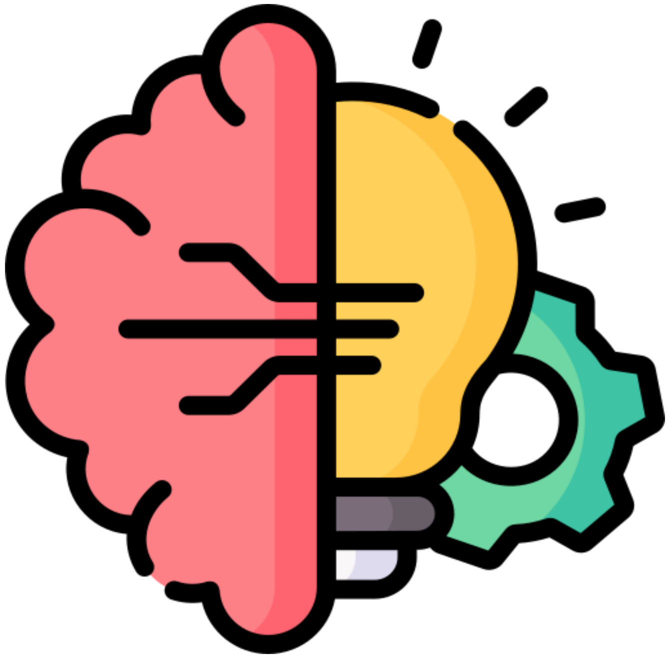
Interactive Visualization:

- Dashboard built using Plotly Dash.
- Features:
    - Filter by launch site and orbit type.
    - Compare success rate across different configurations.

Key Insights:

- VAFB SLC 4E has the lowest success rate.
- Interactive map

PREDICTIVE ANALYSIS - METHODOLOGY



Machine Learning Models Used:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree Classifier
- K-Nearest Neighbors (KNN)

Hyperparameter Tuning:

- Used GridSearchCV for optimal parameter selection.
- Cross-validated using 10-fold CV to ensure robustness.

PREDICTIVE ANALYSIS - MODEL EVALUATION

Model Performance

- Logistic Regression: 83.33% accuracy (best model)
- SVM: 83.33% accuracy
- Decision Tree: 66.67% accuracy
- KNN: 83.33% accuracy

Conclusion

- Logistic Regression, SVM, and KNN are the best-performing models with the same accuracy.
- Decision Tree has the lowest accuracy and may require further optimization.

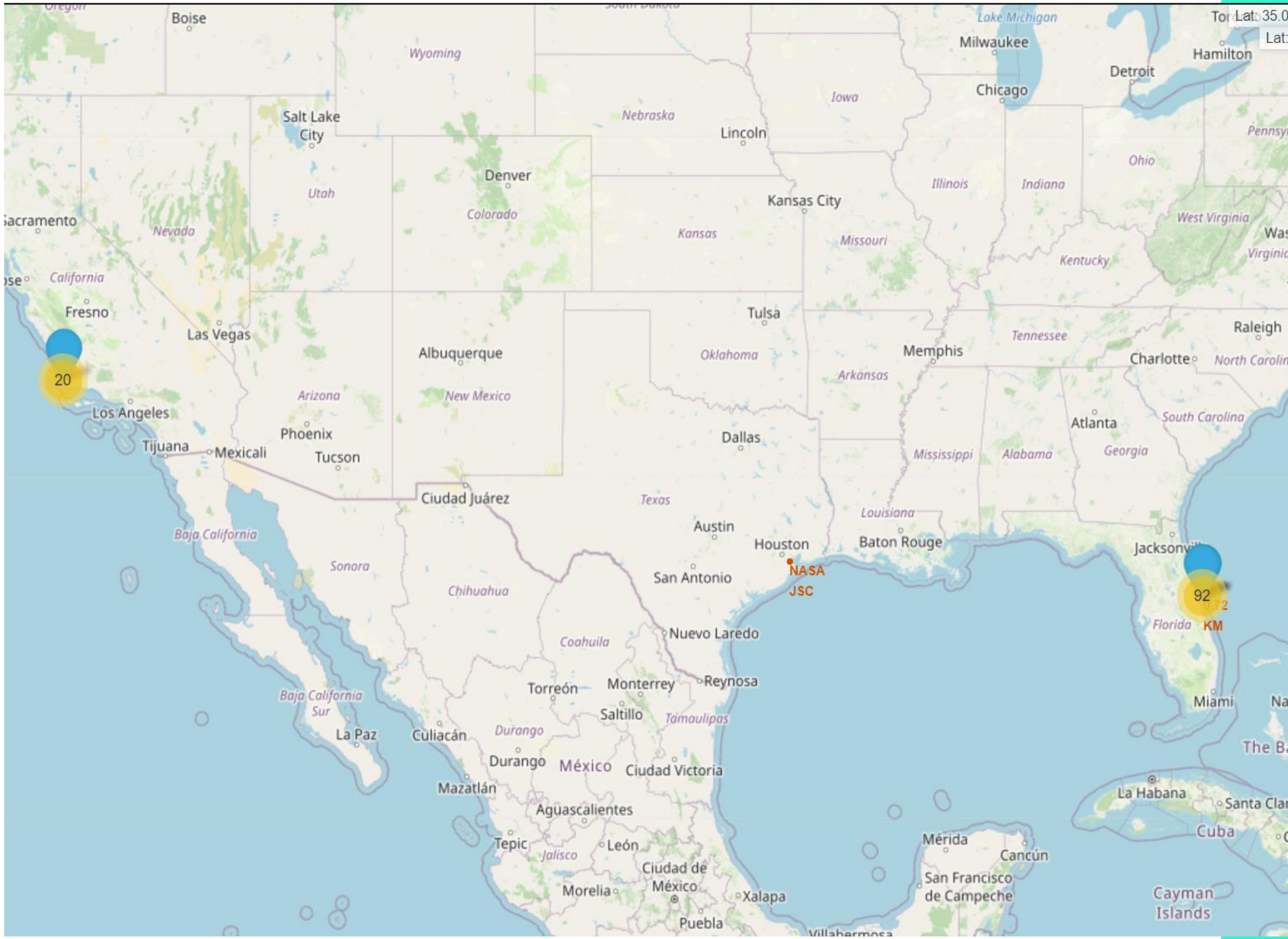
RESULTS

Folium Map Visualization

- Created a map displaying launch sites and successful landing locations.
- PolyLine drawn to show the path between launch sites and landing coordinates.

SQL-based Analysis

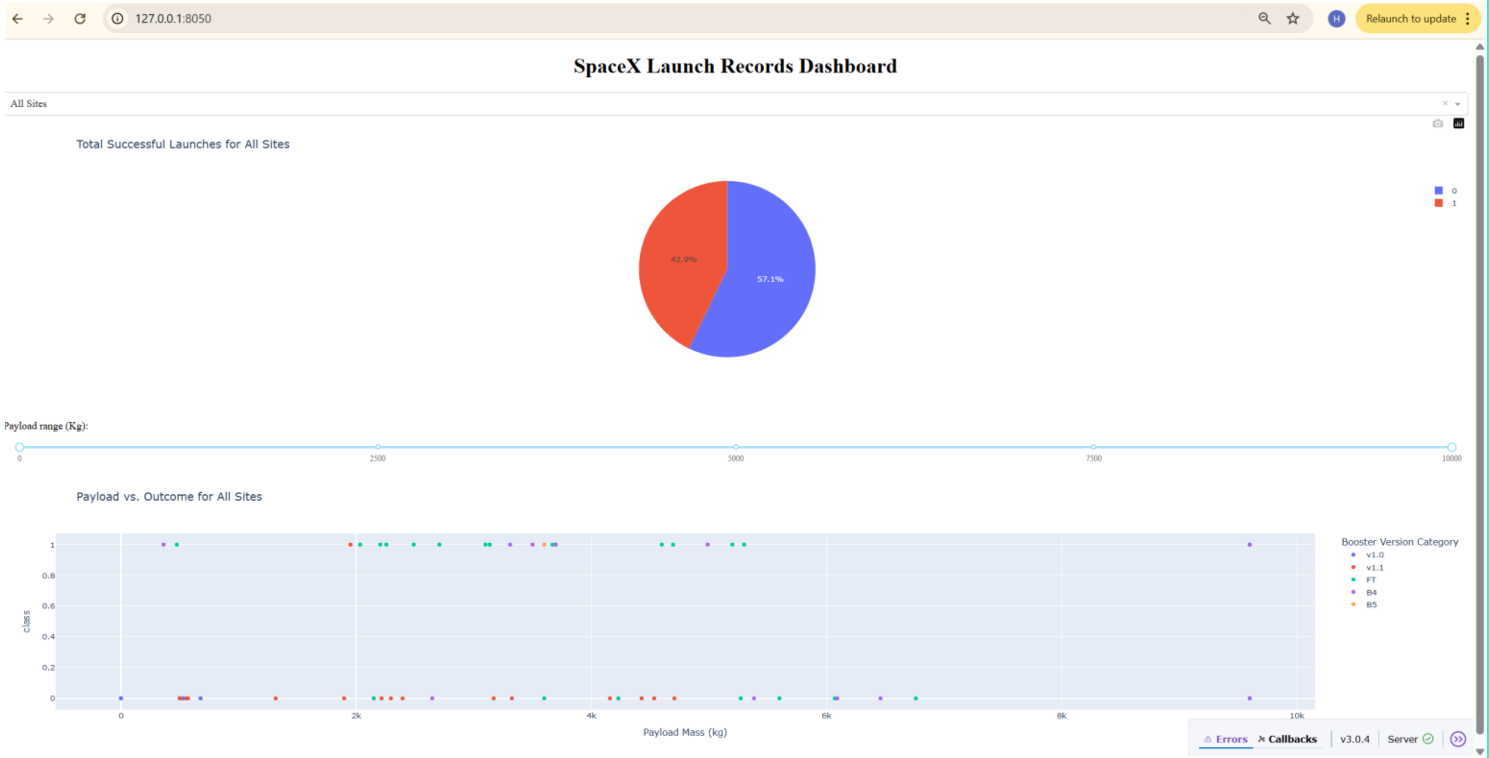
- Queried success rate by launch site.
- Analyzed the impact of payload mass and orbit type using SQL queries.
- Visualized results using Matplotlib.



PLOTLY DASH DASHBOARD

Overview

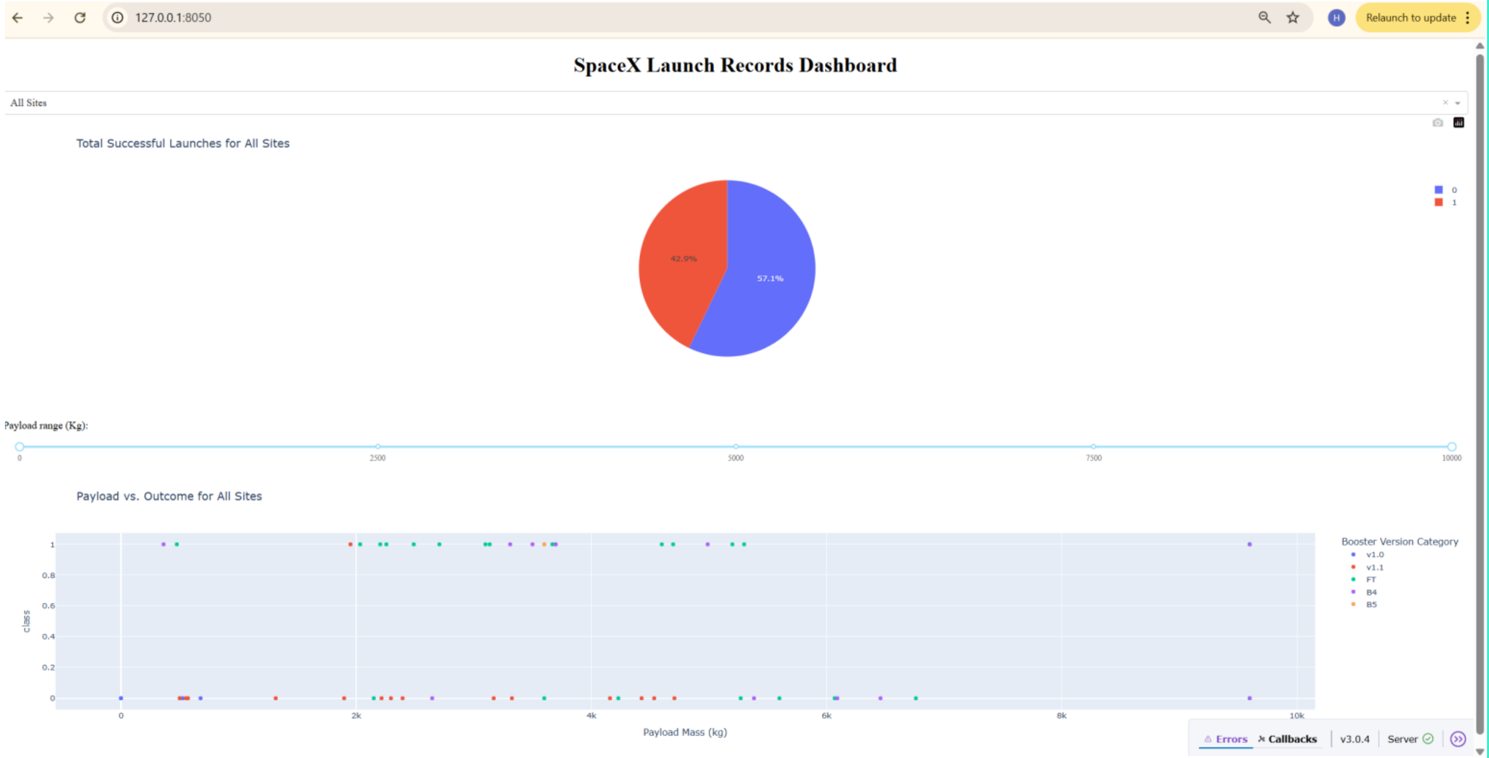
- An interactive dashboard was built using Plotly Dash to visualize SpaceX launch data.
- The dashboard enables users to filter by launch site and adjust payload range dynamically.



PLOTLY DASH DASHBOARD

Key Features

- Pie Chart: Shows the proportion of successful vs failed launches across all sites.
- Payload Range Slider: Allows selection of payload mass range (0 to 10,000 kg).
- Scatter Plot: Displays the relationship between payload mass and launch outcome, differentiated by booster version.



INSIGHTS

- Higher payloads do not always correlate with mission failure or success.
- Booster version plays a role in outcome variability.
- Most successful launches are associated with newer booster versions (e.g., B5, FT).

CONCLUSION

Summary

- Logistic Regression proved to be the most reliable model with 83.33% accuracy.
- The interactive dashboards enhanced data interpretation.
- Maps and spatial analysis provided insights into the geographical patterns of successful landings.

Future Work:

- Incorporate more launch data to improve model accuracy.
- Explore deep learning models for better prediction.
- Integrate real-time data streaming from SpaceX API.

THANK YOU