SPACEX FALCON 9 LAUNCH COST OPTIMIZATION AND REUSABILITY PREDICTION

A DATA SCIENCE–DRIVEN APPROACH USING PUBLIC SPACEX DATA

CAPSTONE PROJECT | IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Presented by
Durga Vani Moka

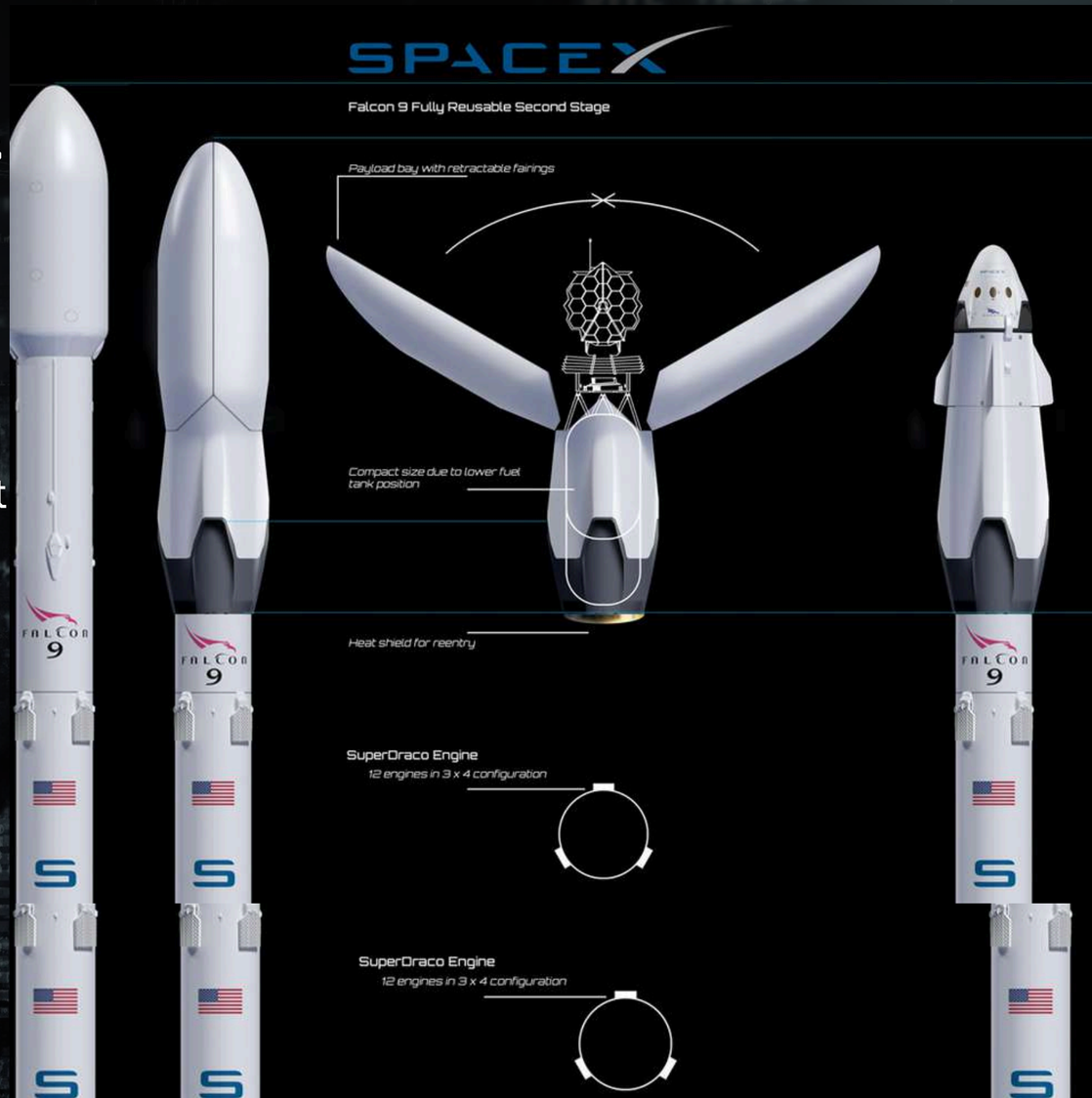
EXECUTIVE SUMMARY

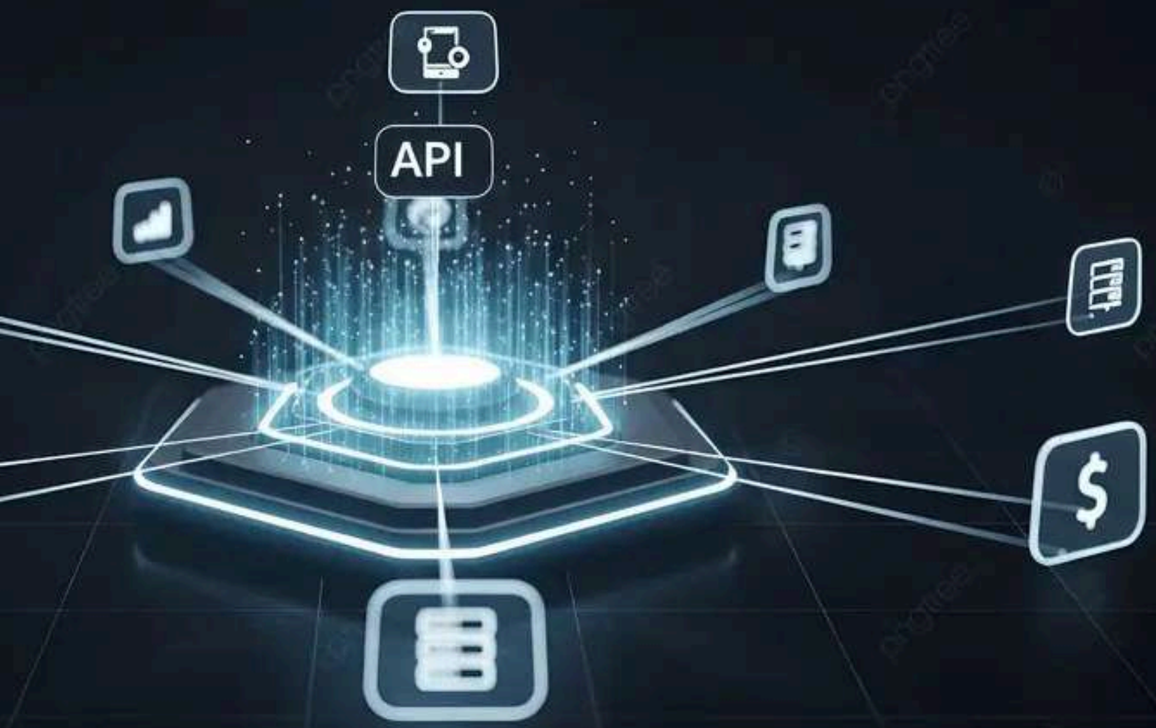
- This project analyzes SpaceX Falcon 9 launch data to understand launch cost factors and first-stage reusability.
- Public data from SpaceX APIs and web sources was collected, cleaned, and analyzed.
- Exploratory data analysis, SQL analysis, and interactive visualizations were performed.
- A machine learning classification model was built to predict first-stage reuse.
- Insights were delivered through dashboards and interactive maps to support data-driven decision-making.

INTRODUCTION

- SpaceX’s Falcon 9 achieves significantly lower launch costs by reusing the first-stage booster.
- Falcon 9 launches cost approximately $62 million, compared to $165+ million for other providers.
- The success of first-stage landing directly impacts launch cost and reusability.
- This project focuses on predicting whether the Falcon 9 first stage will land successfully using historical data.
- Accurate prediction enables indirect estimation of launch cost and risk, supporting competitive bidding decisions.

Objective: Predict first-stage landing success to estimate launch cost impact.





DATA COLLECTION & DATA WRANGLING METHODOLOGY

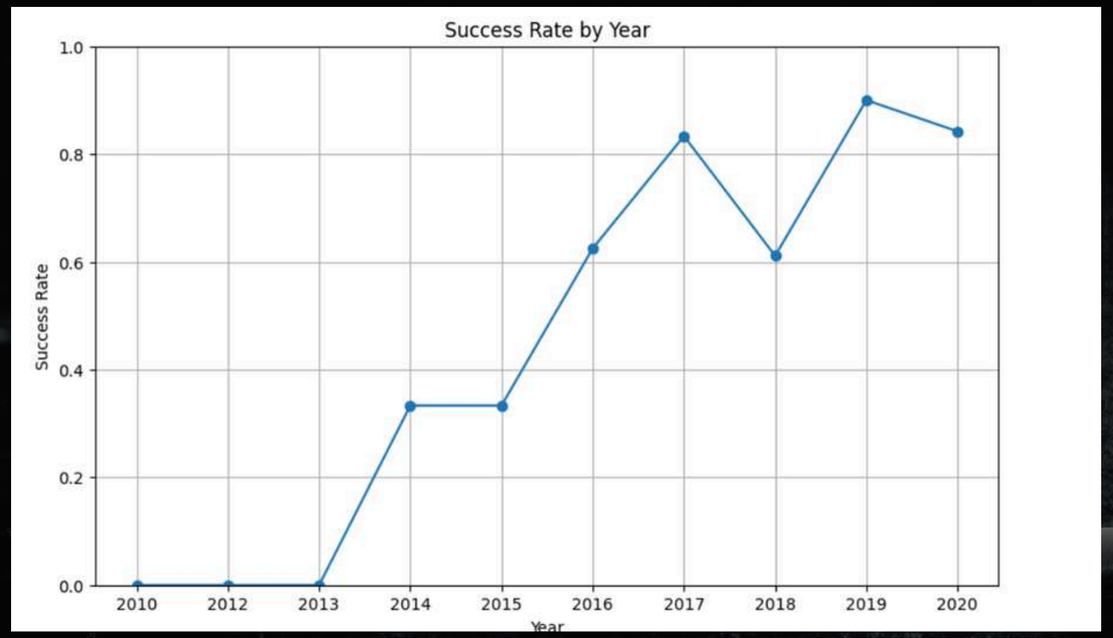
- Collected Falcon 9 launch data from SpaceX REST APIs and Wikipedia.
- Integrated multiple sources into a single dataset.
- Cleaned and prepared data by handling missing values and standardizing features.
- Created the target variable for first-stage landing success.

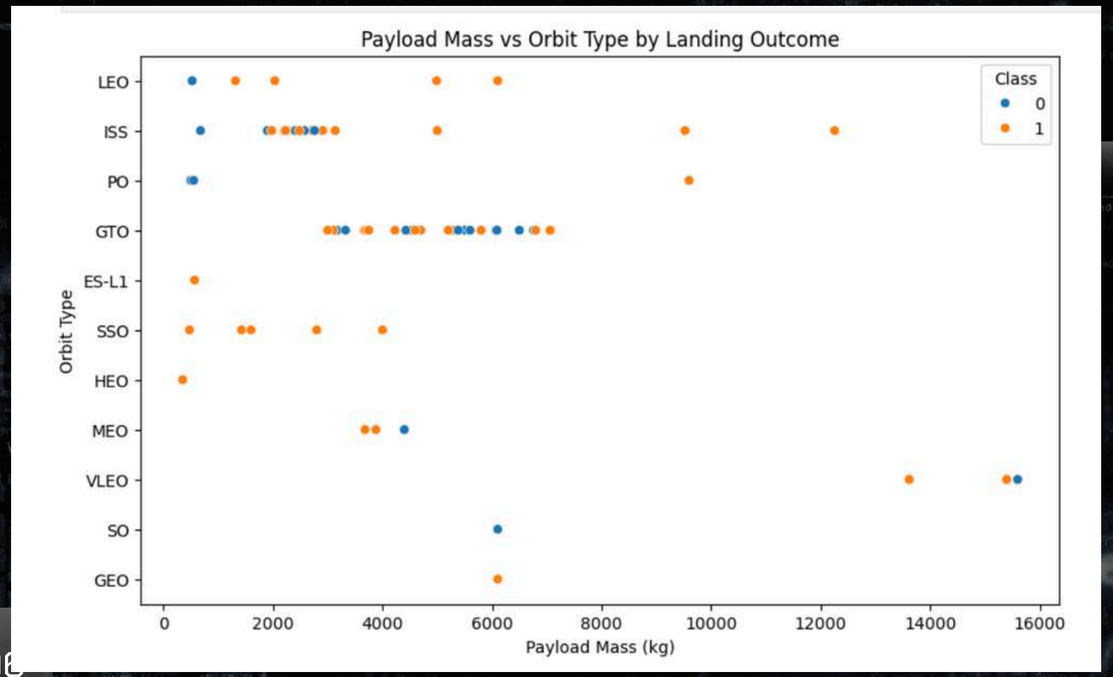
EXPLORATORY DATA ANALYSIS (EDA) – METHODOLOGY

- Performed exploratory data analysis (EDA) to identify patterns related to first-stage landing success.
- Analyzed relationships between payload mass, orbit type, launch site, and flight number.
- Used statistical plots and interactive visualizations to compare successful vs unsuccessful landings.
- Built interactive charts and dashboards to enable dynamic exploration of launch outcomes.

PREDICTIVE ANALYSIS METHODOLOGY

- Defined the prediction problem as a binary classification task.
- Selected relevant features related to launch conditions and mission characteristics.
- Split the data into training and testing sets.
- Trained multiple machine learning classification models to predict first-stage landing success.
- Evaluated model performance using accuracy and classification metrics.





EDA WITH VISUALIZATION RESULTS

Key Insights

- Landing success rate improves over time, indicating increased operational experience and reuse capability.
- Payload mass and orbit type influence landing outcomes, with certain orbits showing higher success rates.
- More complex missions (higher payloads and specific orbits) show greater variability in landing success.

EDA WITH SQL RESULTS

Key Insights

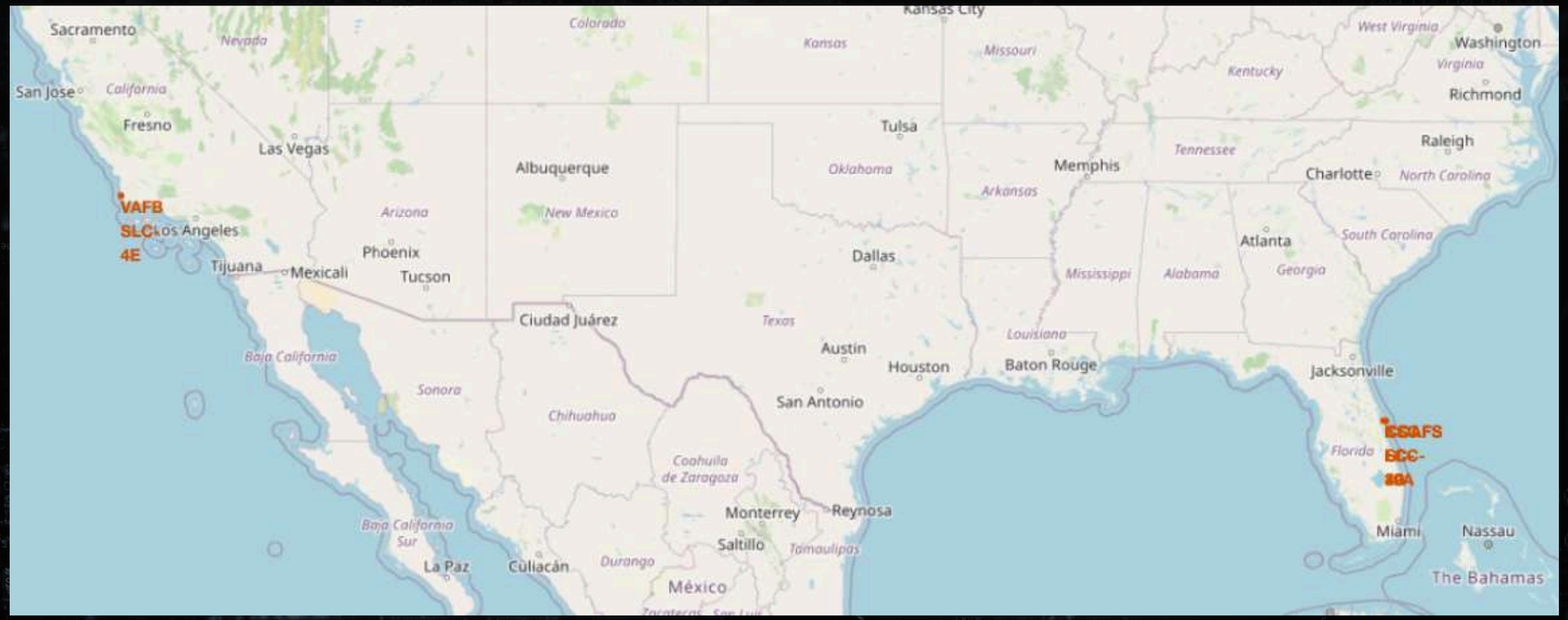
- SQL analysis shows a strong dominance of successful missions over failures.
- Landing outcome type (drone ship or ground pad) plays a critical role in booster reusability.
- These results highlight imbalanced classes, which is important for predictive modeling.

- NASA CRS missions carried a total payload mass of 45,596 kg, highlighting their significant contribution to Falcon 9 launch activity.

Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Customer	total_payload_mass
NASA (CRS)	45596

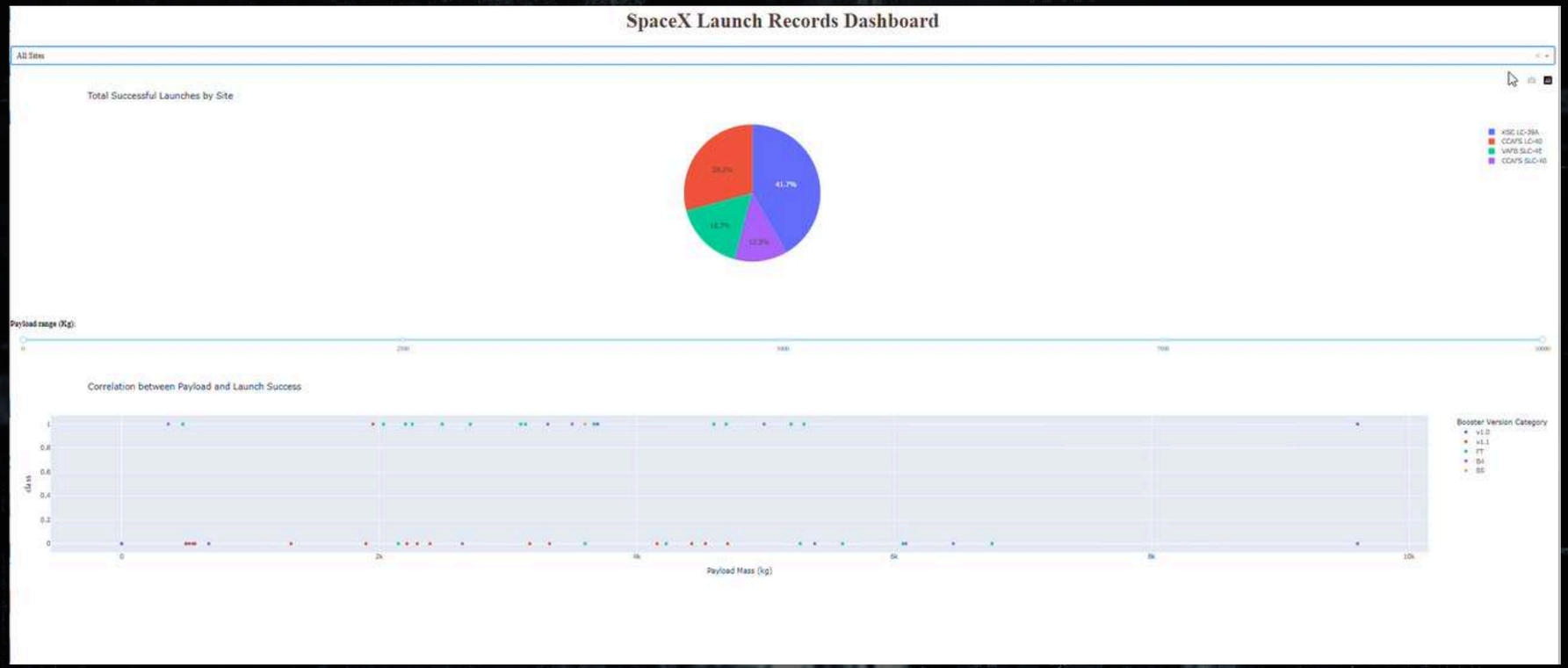
INTERACTIVE MAP RESULTS (FOLIUM)



- Visualized Falcon 9 launch site locations across the United States.
- Launch activity is concentrated at CCAFS LC-40, KSC LC-39A, and VAFB SLC-4E.
- Coastal launch locations support recovery operations and landing strategies.
- Geographic distribution influences mission planning and first-stage reusabilities.

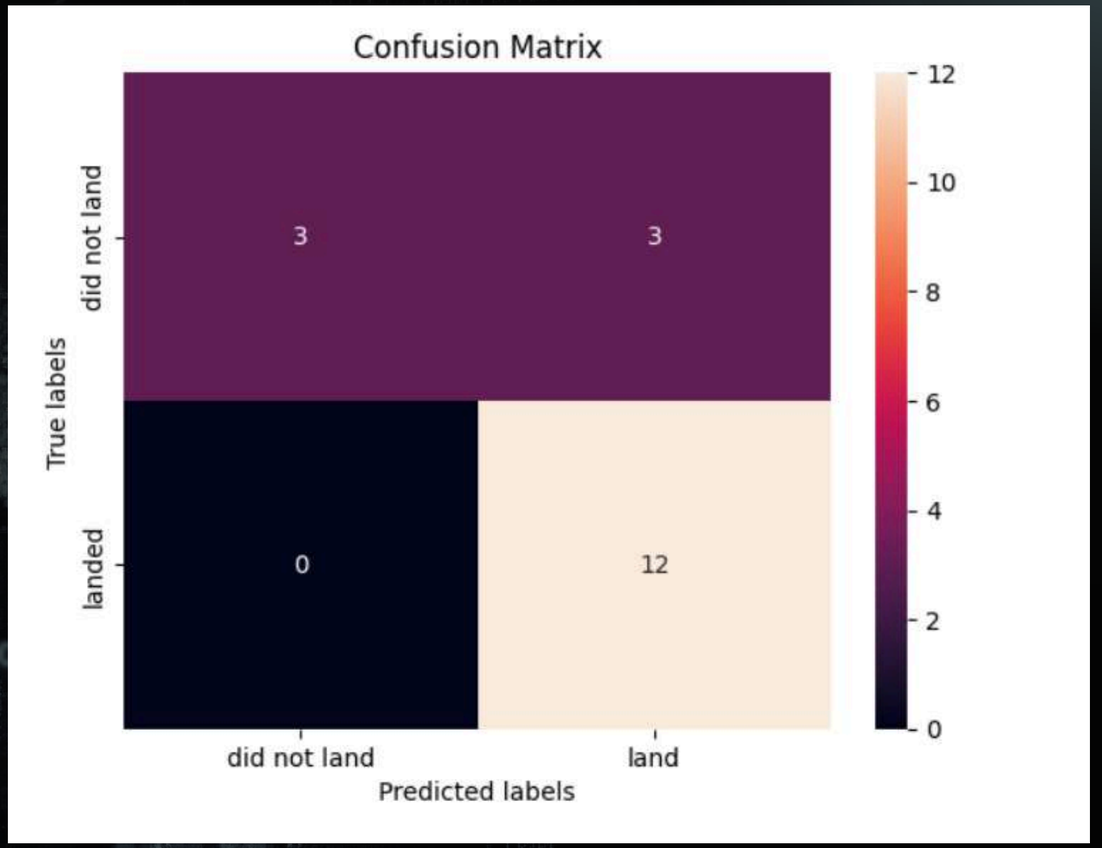
INTERACTIVE DASHBOARD BUILT WITH PLOTLY DASH.

- Built an interactive dashboard using Plotly Dash to analyze Falcon 9 launch data.
- Visualizes successful launches by site and payload mass vs launch success.
- Provides interactive filters for launch site and payload range.
- Enables stakeholders to explore patterns affecting first-stage landing success.



PREDICTIVE ANALYSIS (CLASSIFICATION) RESULTS

- Multiple classification models (Logistic Regression, SVM, KNN, Decision Tree) were trained and evaluated.
- All models achieved a similar accuracy of approximately 83% on the test data.
- Models showed high recall for successful landings, indicating strong ability to predict reuse.
- Lower recall for failure cases highlights the impact of class imbalance and limited data.



The model predicts successful landings more reliably than failures.

CONCLUSION

- This project demonstrates how public SpaceX launch data can be transformed into actionable insights.
- Analysis shows that first-stage reusability is predictable using historical mission characteristics.
- The predictive model achieves ~83% accuracy, enabling risk-aware launch cost estimation.
- Interactive dashboards and maps allow stakeholders to explore launch performance and reuse patterns without technical expertise.
- These insights support competitive bidding, cost optimization, and strategic planning for commercial launch providers.

THANK YOU

Thank you for your time and attention.
This presentation demonstrated a data-driven approach to predicting Falcon 9 first-stage reusability and estimating launch cost impact.
I welcome any questions, feedback, or discussion on the analysis and results.

Name: Durga Vani Moka
Project: IBM Data Science Professional Certificate – Capstone
Focus: Launch Cost Estimation & Reusability Prediction