Space Launch Success Prediction

Capstone Project

Mo Faizanur Rahman

Table of Content

Executive Summary

Problem :

- Rocket launch failures lead to huge financial losses and damaged reputation.
- Private space companies face tight competition and pressure to improve reliability.
- Predicting launch outcomes is critical for cost savings and mission planning.

Objective :

- Use data science to predict whether a launch will succeed.
- Identify key influencing factors like payload mass, orbit type, and site.
- Compare and evaluate different ML models to maximize accuracy.

Outcome :

- Found payload mass, orbit type, and launch site to be most impactful features.
- Built ML models, with Decision Tree reaching ~88% accuracy.
- Delivered insights to reduce risk and optimize launch strategies.

Introduction

The New Era of Space Exploration :

- Over the past two decades, private companies (SpaceX, Blue Origin, Rocket Lab, ISRO’s Antrix/NSIL) have transformed space from a government-only domain into a commercial frontier.
- Rocket launches remain high-risk, high-cost: a single failure can mean millions lost, delays in satellite networks, and reputational damage.
- With growing demand for satellite constellations, interplanetary missions, and even space tourism, predictability and reliability are more important than ever.

Literature Review

- Publicly available datasets provide historical launch data, but most studies analyze them descriptively, not predictively.
- Previous research often focused on orbital dynamics or engineering simulations rather than data-driven prediction.
- Limited work applies machine learning to systematically predict launch outcomes across different sites, payloads, and orbits.
- Our project addresses this gap by combining exploratory data analysis (EDA) with predictive modeling.

Methodology

Data Source:

SpaceX APIs open data

Approach :

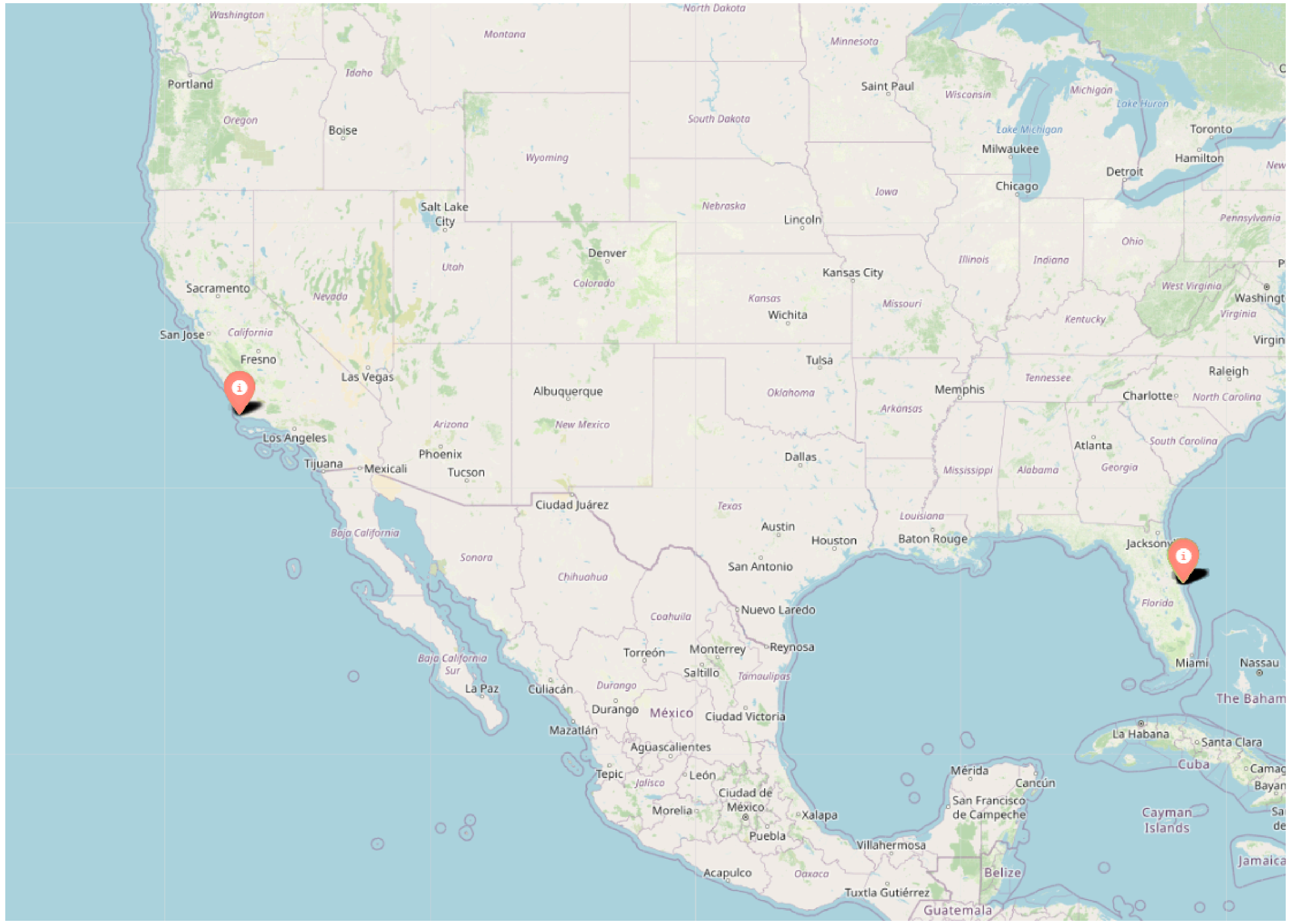
- Data Wrangling – cleaning, handling missing values, engineering new features.
- Exploratory Data Analysis (EDA) – uncovering trends in payload, orbit, site, and outcomes.
- Feature Selection – identifying key drivers of success.
- Model Development – Logistic Regression, Decision Tree, KNN, SVM.
- Evaluation – cross-validation and accuracy comparison.

Goal :

Build a reproducible pipeline that moves from data to insights to prediction.

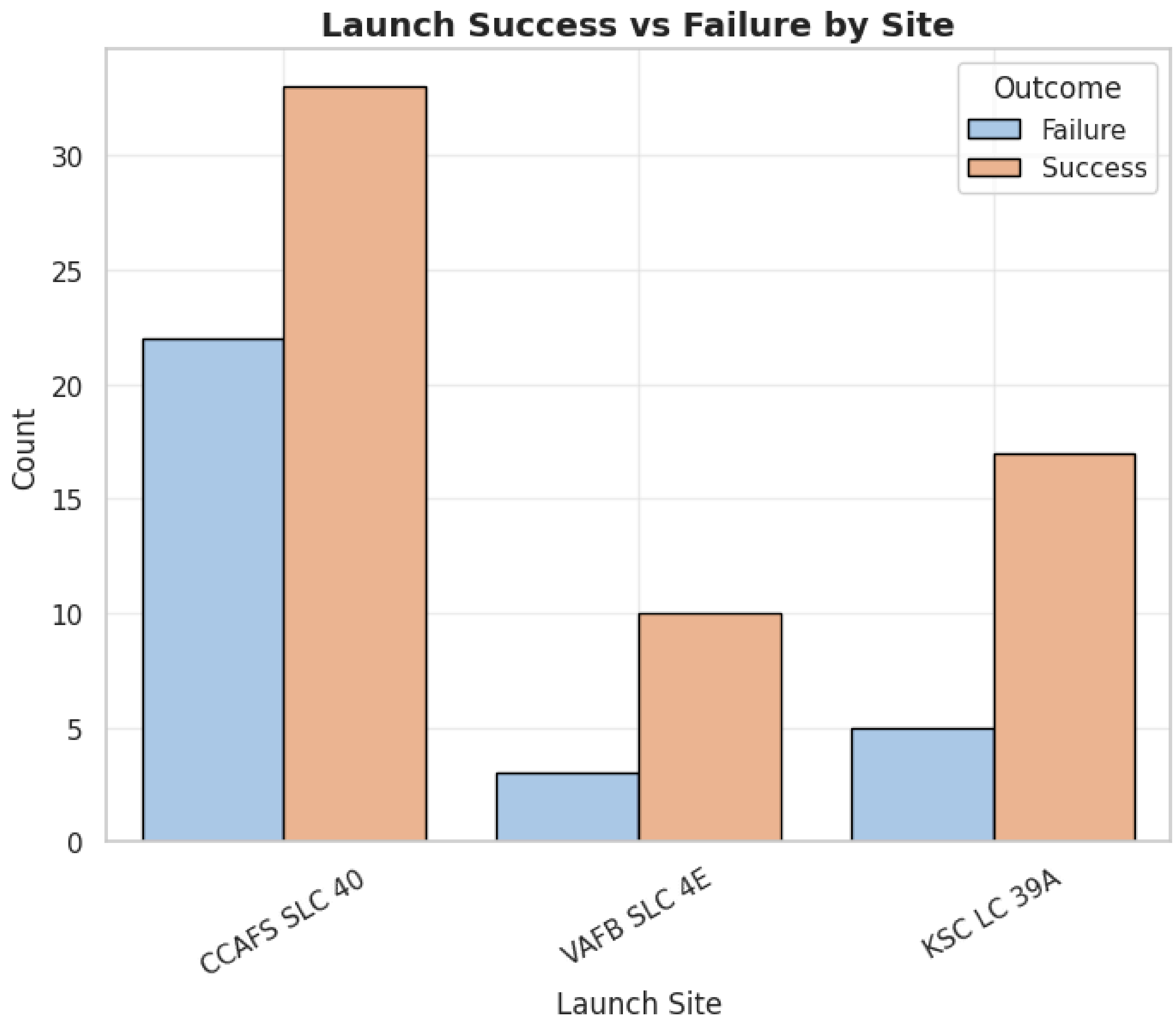
Launch Sites

- Rocket launches occur across multiple global sites, each with unique geographical and technical factors.
- Launch location plays a critical role in determining success rates — from favorable weather conditions to proximity to the equator (which provides natural velocity boosts).
- Insight : Certain sites consistently perform better than others, giving companies a competitive edge.



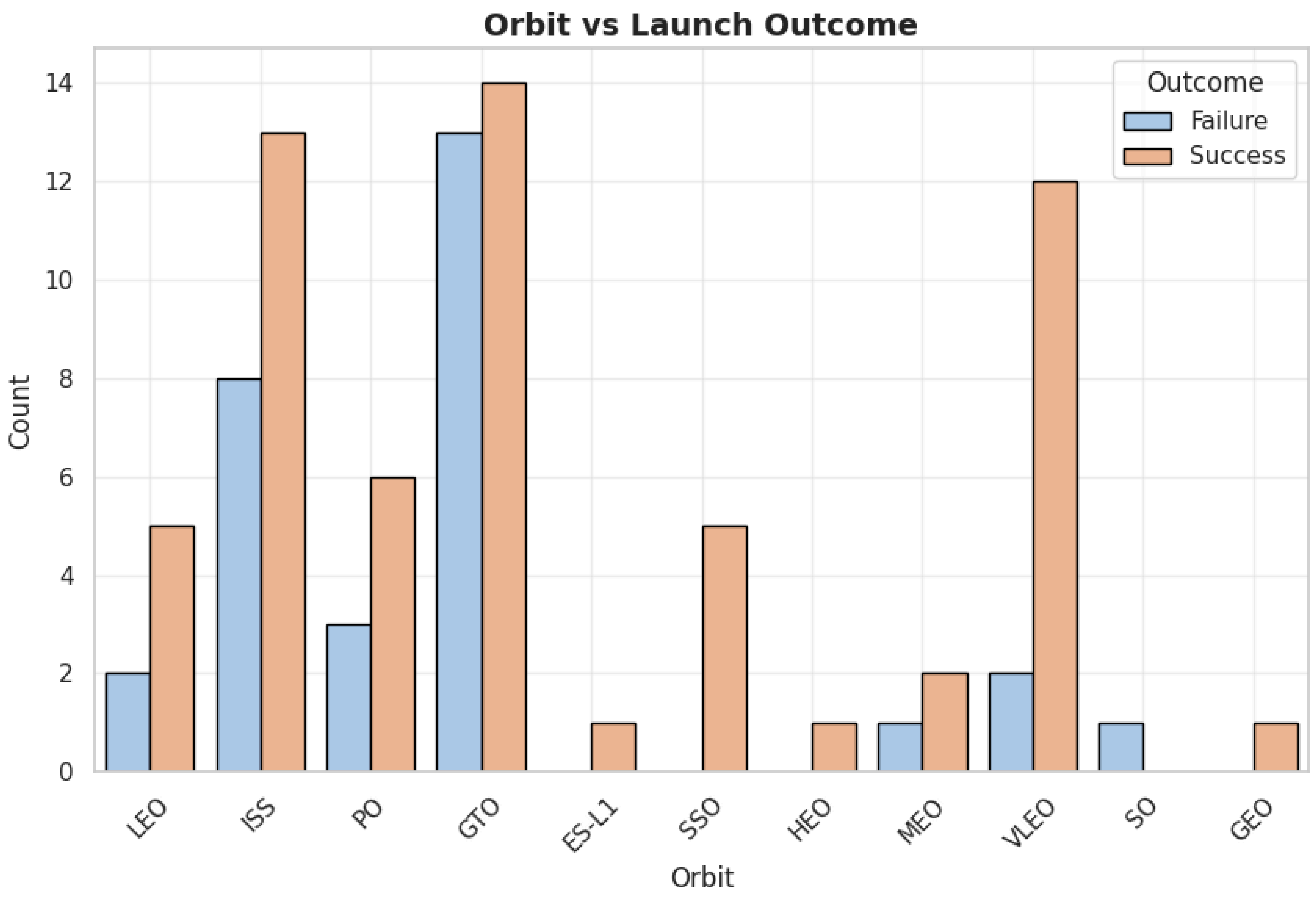
Launch Success by Site

- Different launch sites show different reliability levels.
- Sites with higher infrastructure investment and more flight history tend to have better success records.
- Business implication : Site selection is not just logistical, but also strategic for maximizing launch success.



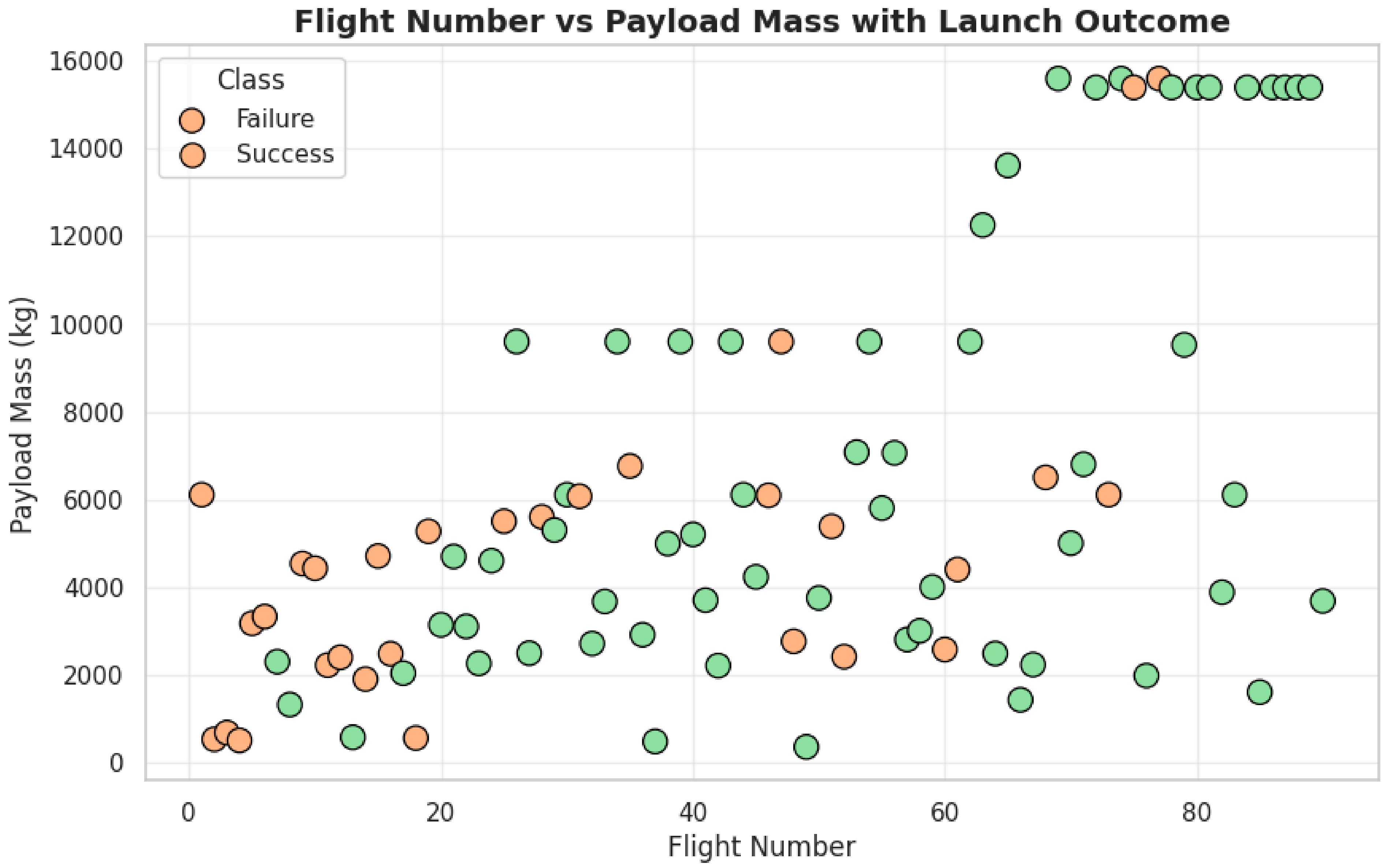
Orbit Type vs Success

- The type of orbit targeted plays a major role in launch difficulty and reliability.
- Orbits like LEO (Low Earth Orbit) and GTO (Geostationary Transfer Orbit) show different success probabilities.
- Complex orbits may demand higher precision, which can increase the chance of mission failure.
- Orbit selection directly influences mission risk.



Flight Num vs Success

- Experience matters: as the number of launches (flight number) increases, the probability of success improves.
- Payload matters: very heavy payloads reduce success probability due to engineering constraints.
- Combined Effect : Companies gain confidence and efficiency with more flights, but must balance payload mass carefully.



Model Accuracy Comparison

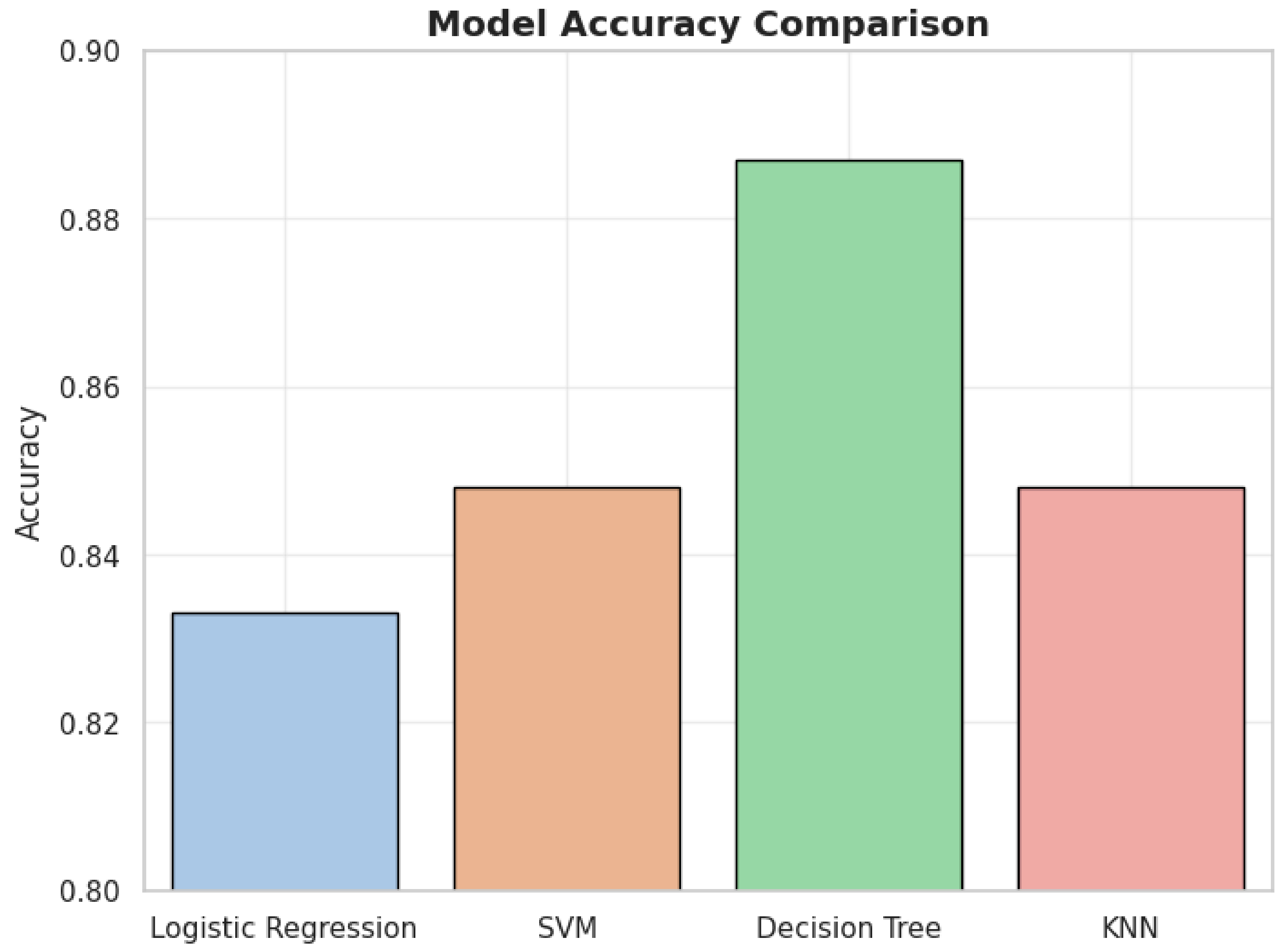
We tested multiple ML models: Logistic Regression, SVM, KNN, Decision Tree.

Result

- Logistic Regression: Simple but limited.
- KNN: Performed moderately.
- SVM: Strong but less interpretable.
- Decision Tree: Best performer (›88% accuracy).

Conclusion

Decision Trees provide both accuracy and interpretability — the most practical model for this use case.



Discussion

Key Insight :

- Launch site, payload mass, and orbit type are the top predictors.
- Experience (flight number) increases success rate.

Implications :

- Companies can plan missions more strategically, reducing financial and reputational risks.
- Predictive models support data-driven decision-making in aerospace operations.

Conclusion

Machine learning can predict launch success with high accuracy (~88%).
Decision Tree outperformed other models in balancing interpretability and accuracy.

Business Value :

- Optimized launch planning.
- Improved risk management.
- Cost savings through data-driven insights.

Future Work :

- Include weather and real-time telemetry data.
- Explore ensemble models and deep learning for higher accuracy.

References :

SpaceX API (Open data)

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
https://api.spacexdata.com/v4/rockets/{rocket_id}
https://api.spacexdata.com/v4/launchpads/{launchpad_id}
https://api.spacexdata.com/v4/payloads/{payload_id}
https://api.spacexdata.com/v4/cores/{core_id}
https://api.spacexdata.com/v4/launches/past
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_3.csv

Acknowledgments

- Course Instructors: Joseph, Yan, Azim
- IBM Data Science Capstone Team
- Open-source data science community