IBM Applied Data Science. Capstone Project

John Doe

Skills Network

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Key Findings

Technology Advancements
Success rate improved from ~0% (2010-2013) to 80%+ (2019-2020) due to technological progress.
Launch Site Specialization
Some excel with high-mass payloads, others succeed more with low-mass missions.
Orbit-Based Success
Orbits like ES-L1, GEO, HEO, and SSO have perfect success rates; GTO has ~50%, showing varying risk by orbit.
Geospatial Insights
Launch sites near highways (logistics), coastlines (safe trajectories), and away from towns (public safety).

Model Results

- Models consistently achieve ~84% accuracy, indicating strong predictive capability.
- Opportunity exists to improve further via advanced techniques.

INTRODUCTION

SpaceX offers Falcon 9 rocket launches at a cost of $62 million, significantly lower than other providers, whose prices often exceed $165 million. Much of SpaceX's cost advantage comes from its ability to reuse the rocket's first stage. By predicting whether the first stage will successfully land, we can better estimate the true cost of a launch. This information would be valuable for other companies considering competitive bids against SpaceX for rocket launches.

The objective of this project is to develop a machine learning pipeline capable of predicting the likelihood of a successful landing of the Falcon 9's first stage.

Key Questions and Objectives:

- What factors influence the likelihood of a successful rocket landing?
- How do various factors interact to impact the success rate of these landings?
- What operational conditions are necessary to ensure a reliable and successful landing program?
- This predictive model could be a valuable tool for stakeholders seeking to optimize launch costs, enhance landing reliability, or enter the competitive space launch market.

METHODOLOGY

Data Sources: The data for this project was gathered through multiple methods:

- SpaceX's API was used to obtain specific launch data.
- Additional details were collected via web scraping from Wikipedia.

Data Preparation:

- Data wrangling techniques were applied to clean and structure the data.
- One-hot encoding was used to convert categorical features into a format suitable for analysis and modeling.

Exploratory Data Analysis (EDA):

- EDA was conducted through both data visualization and SQL queries to identify patterns, correlations, and key insights.
- Visualization tools were employed to explore the relationships between variables and better understand factors affecting successful landings.

Interactive Visual Analytics:

- Folium maps and Plotly Dash were used to create interactive visualizations, allowing for deeper insights and user-driven exploration of the data.

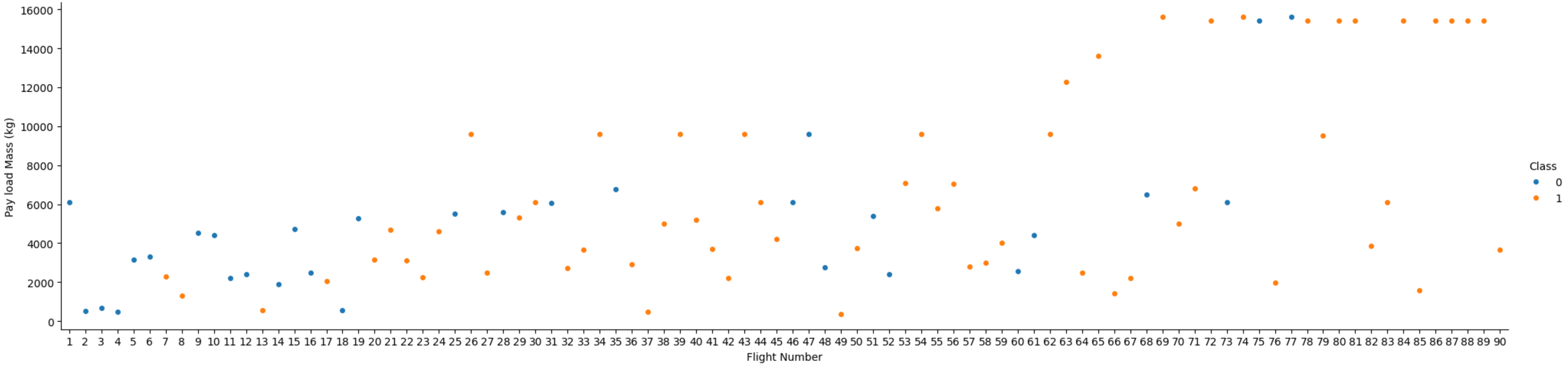
Predictive Analysis:

- Building, tuning, and evaluating classification models to optimize performance.
- Using various metrics to assess model accuracy and ensure reliability in predictions.

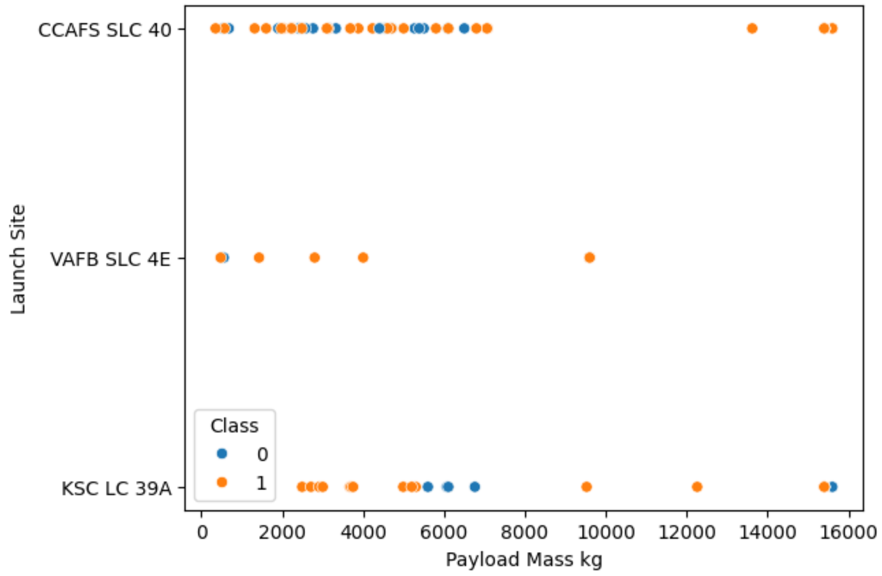
RESULTS

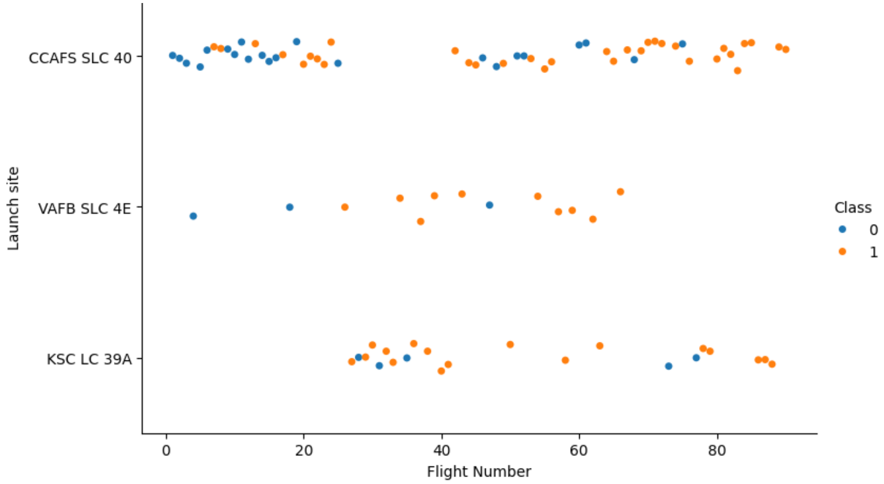
In the analysed dataset, we observe a clear trend of rapid advancements in launch technology. This progress is reflected in two key areas:

- **Higher Success Rates per Flight:** Over time, the success rate of launches has increased, demonstrating improvements in reliability and precision.
- **Increase in Payload Capacity**: The data also shows a steady increase in the payload mass that rockets can carry. This increase points to enhancements in rocket design and technology, allowing for larger and more complex missions.



RESULTS



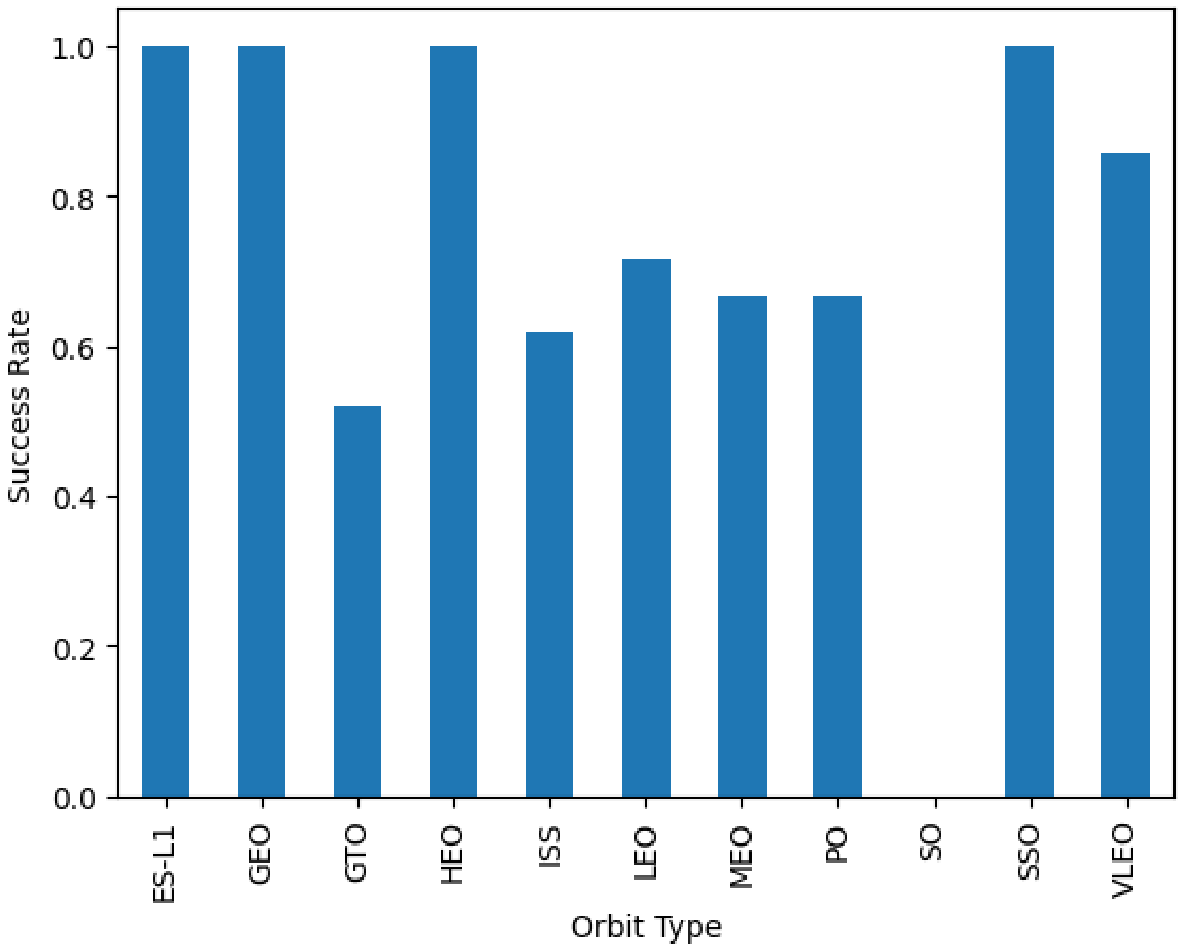


Analysing the data by launch site reveals that certain sites have specialized capabilities:

- **High-Mass Payload Launches**: Some launch sites are better equipped and more frequently used for launching heavier payloads, indicating a focus on missions requiring robust capacity and technology.
- **Low-Mass Payload Success**: Other sites have a higher success rate with lower-weight payloads, suggesting a specialization in missions that prioritize precision and reliability for smaller payloads.

This trend suggests that launch sites develop unique strengths, likely due to differences in infrastructure, geographic location, and mission types. Recognizing these site-specific specializations can aid in strategic planning, ensuring that launches are matched to the most suitable site for their payload and mission requirements.

RESULTS



When analysing launch success rates by orbit type, we can observe distinct trends in reliability:

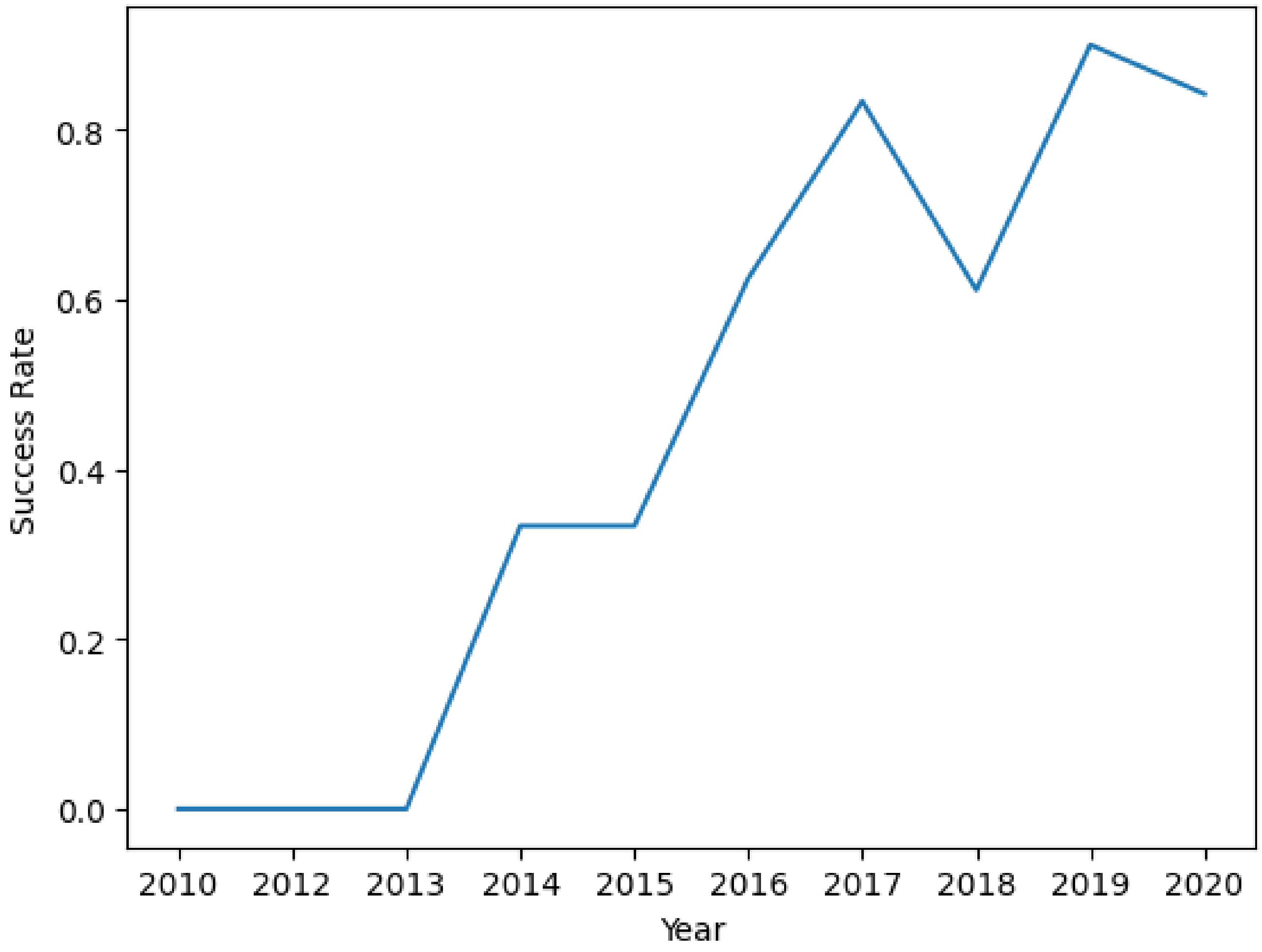
- **High Success Orbits**: Certain orbits, such as ES-L1, GEO, HEO, and SSO, show a perfect success rate (1.0). This high reliability suggests that launches to these orbits are either better understood, have more established procedures, or benefit from specialized technology that enhances success rates.
- **Lower Success Orbits**: In contrast, launches to orbits like GTO (Geostationary Transfer Orbit) display a much lower success rate, closer to 50%. This may be due to the technical challenges associated with these orbits, which often involve complex manoeuvres and higher energy requirements, leading to a greater risk of mission failure.

RESULTS

It's noteworthy that the overall launch success rate has shown significant improvement over time:

- **Early Years (2010-2013):** In this period, the success rate was relatively low, closer to 0%, reflecting the challenges and risks in early-stage launches as new technologies and procedures were still being developed and tested.
- **Recent Years (2019-2020):** By 2019-2020, the success rate has increased dramatically, reaching 80% or higher.

This upward trend underscores how rapidly the industry is evolving, benefiting from continuous innovation and refinement in rocket and launch technology.



RESULTS

Additionally, the data collected in this study was analyzed based on geographic location, allowing for insights into how regional factors impact launch outcomes.

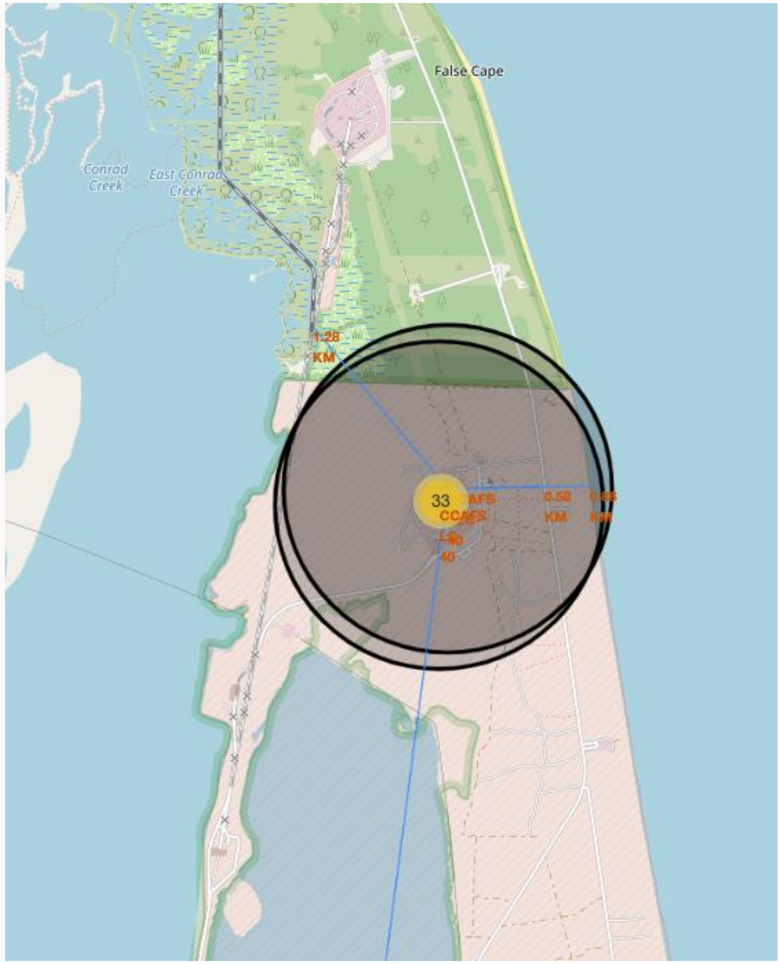
Certain geographic areas may specialize in particular types of missions or payloads due to factors like infrastructure, proximity to the equator (which can influence fuel efficiency and launch trajectory), and regulatory environments.

Some launch sites demonstrate consistently higher success rates, possibly due to favorable weather patterns, advanced facilities, or optimized operational practices tailored to local conditions.

RESULTS

As part of the geospatial analysis, we examined the proximity of launch sites to key geographic features, such as highways, coastlines, and towns. This analysis provides insights into how these factors may influence launch site selection, accessibility, and operational success:

Proximity to Highways: Being close to major highways can facilitate the transportation of heavy equipment, rocket stages, and personnel, improving logistical efficiency.
Coastline Proximity: Many launch sites are located near coastlines to allow rockets to launch over open water. This minimizes risk to populated areas, making launches safer and more feasible.
Distance from Towns: Launch sites are often positioned away from densely populated areas to reduce risks associated with potential launch failures or debris.



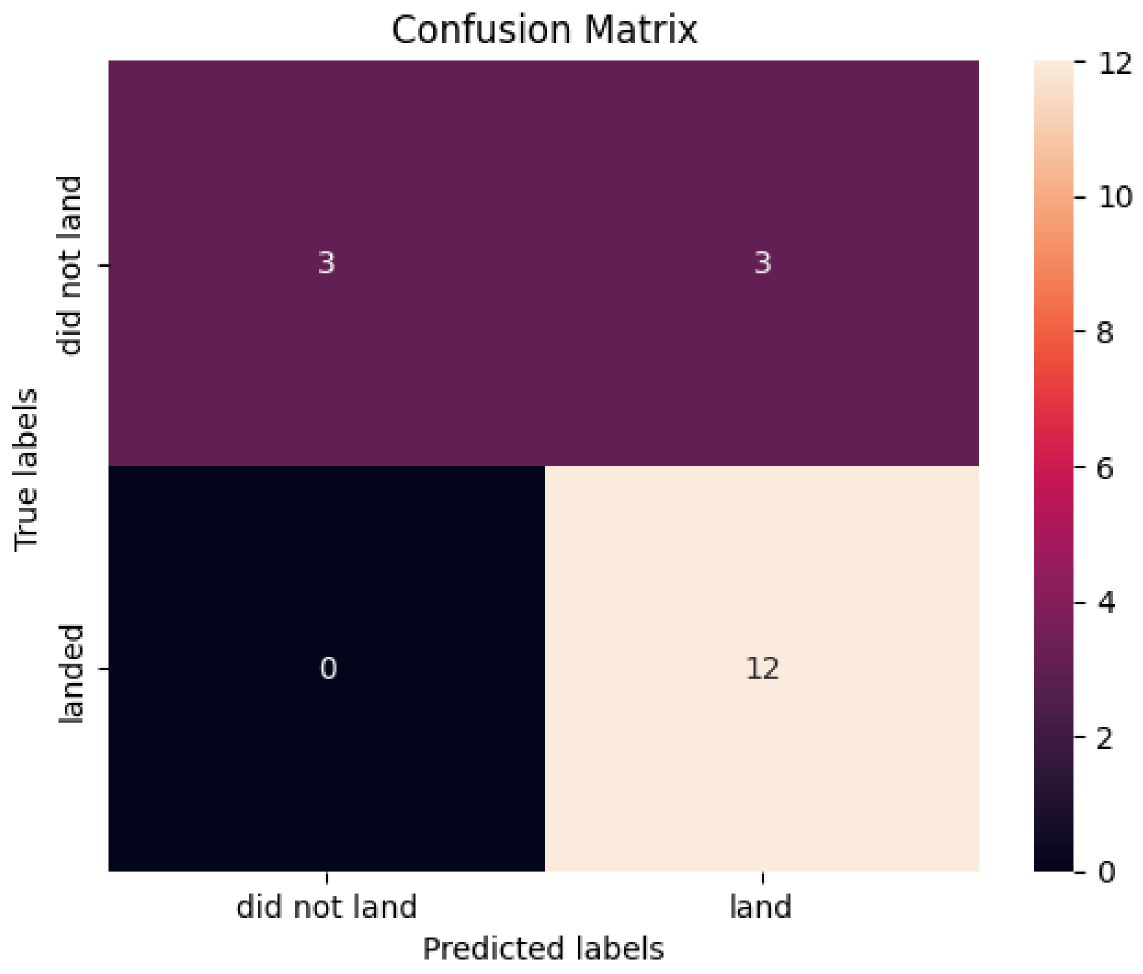
RESULTS

By applying machine learning techniques, we discovered that different models consistently achieved a high accuracy, approximately 84%. This result suggests that:

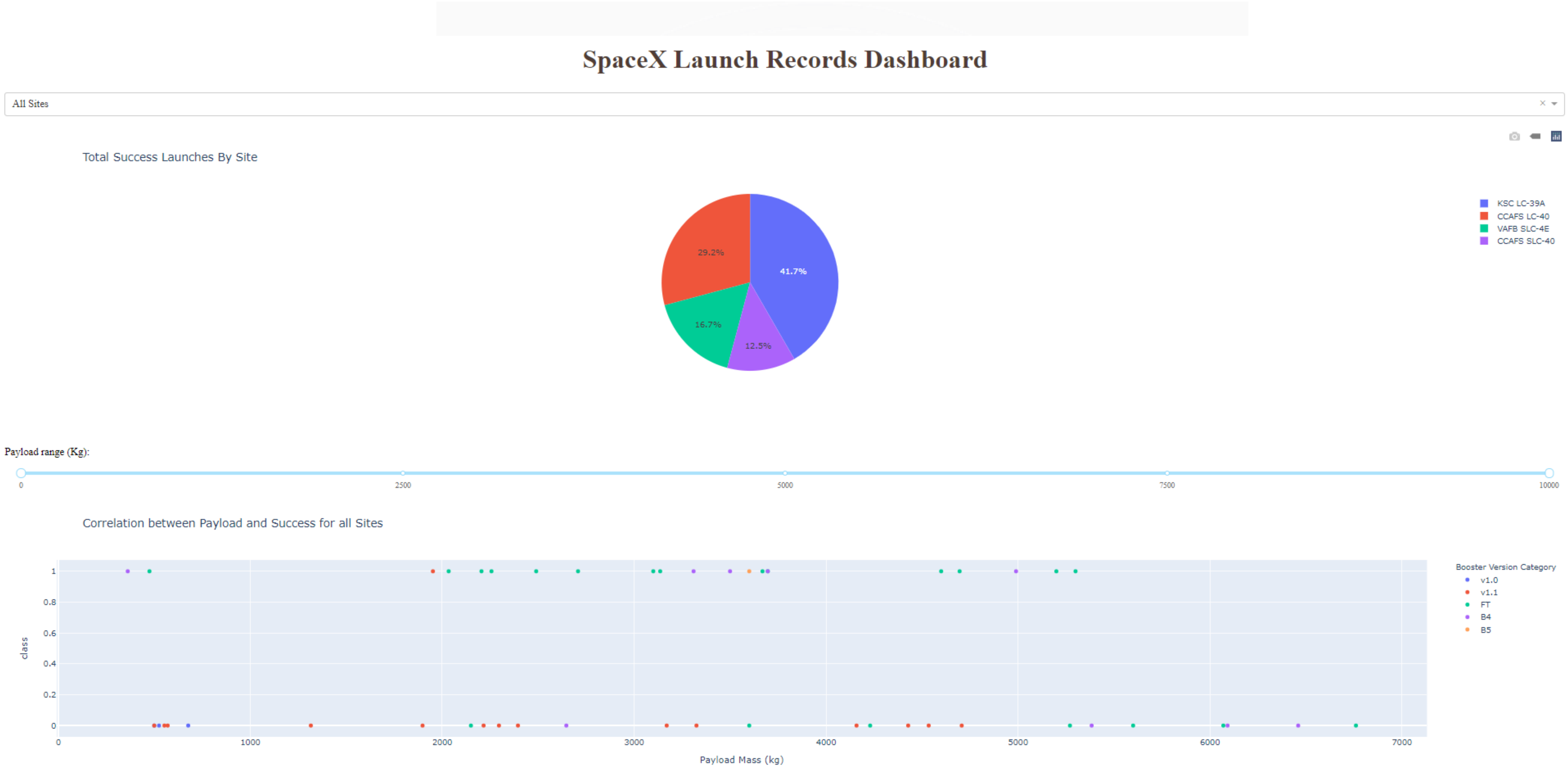
Strong Predictive Power: The models are effectively capturing key factors related to launch success, indicating that the selected features are relevant and informative for predicting outcomes.

Reliable Model Performance: The consistency of accuracy across different models suggests that the prediction is robust and that the models are generalizing well to the data.

Potential for Optimization: Although 84% is a strong accuracy, there may still be room for fine-tuning the models or exploring advanced techniques to push accuracy even higher.



RESULTS


SpaceX Launch Records Dashboard
All Sites
Total Success Launches By Site
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
0
2500
5000
7500
10000
Correlation between Payload and Success for all Sites
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5