IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ziyu Jiang

September 15, 2024

Outline

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

Executive Summary

The objective of this project was to predict the success of the Falcon 9 first-stage landing, which is a critical component in reducing launch costs by reusing boosters. The project involved various stages, including data collection, data wrangling, exploratory data analysis(EDA), data visualization and machine learning model development.

Methodology Summary:

- Data Collection: Data was collected from the SpaceX API and through web scraping. The API provided essential details about each launch, while web scraping supplemented the dataset with additional information.
- Data Wrangling: The data was cleaned and processed to ensure it was suitable for analysis. This included handling missing values and feature engineering, such as creating the `Class` variable to represent landing success.
- Exploratory Data Analysis (EDA): EDA was conducted using SQL queries and visualizations to identify key patterns in the data, such as the relationship between payload mass and landing success. Interactive maps were created using Folium, and an interactive dashboard was built using Plotly Dash.

Executive Summary (Continued)

- Predictive Modeling : Four machine learning models were developed, Logistic Regression, SVM, Decision Tree, and k-Nearest Neighbors (kNN). Each model was tuned using GridSearchCV, and 10-fold cross-validation was used to ensure robust model performance. The models were then evaluated on test data.

Results Summary:

- **Model Performance:**
  - **Decision Tree** was the best-performing model with a **cross-validation score of 87.68%** and a test accuracy of **83.33%**.
  - Other models, including **Logistic Regression**, **SVM**, and **kNN**, all achieved similar test accuracies of **83.33%**, but their cross-validation scores were slightly lower (Logistic Regression: 84.64%, SVM: 84.82%, kNN: 84.82%).
- **Data Visualization:** The analysis revealed key trends, such as the correlation between payload mass and landing success, and the varying success rates across launch sites. These insights were further explored with interactive visualizations in the dashboard.

Introduction

SpaceX has revolutionized the space industry by significantly reducing the cost of launching payloads into space, primarily through its ability to reuse the Falcon 9 rocket's first stage. Successfully landing the first stage after launch is critical to reusability, and predicting whether a landing will be successful can help optimize launch operations and reduce costs further. This project leverages historical data on SpaceX launches to build machine learning models that predict the success of Falcon 9 landings.

The central questions addressed by this project are:

1. Can we accurately predict the success of a Falcon 9 first-stage landing using historical launch data?

2. Which machine learning model provides the best prediction accuracy for landing success?

3. What are the key factors that influence the success of Falcon 9 landings, such as payload mass, launch site, and booster version?

This project aims to develop and evaluate machine learning models to answer these questions, providing insights that can help improve SpaceX's operational efficiency and predict the likelihood of future successful landings.

Section 1

Methodology

Methodology

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

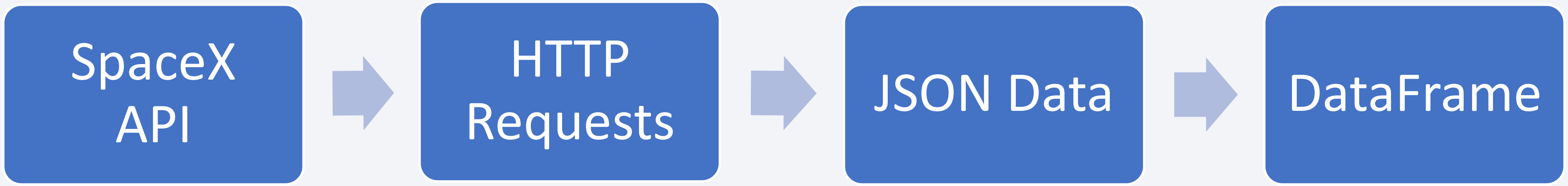
Data Collection-SpaceX API

The data collection process involved gathering information from multiple sources to ensure the dataset had all the relevant features for predicting the success of Falcon 9 landings. The primary data collection methods were:

SpaceX API:

SpaceX provides an open API that allows users to retrieve detailed information on every launch. This API was used to collect core data such as: **Launch Dates, Launch Sites**, **Booster Versions**, **Payload Mass**, **Orbit Type**, and **Landing Outcomes (Success or Failure)**

The data was pulled using HTTP requests, and the API returned the information in JSON format, which was then processed into a structured dataframe for further analysis (See the Jupyter notebook).

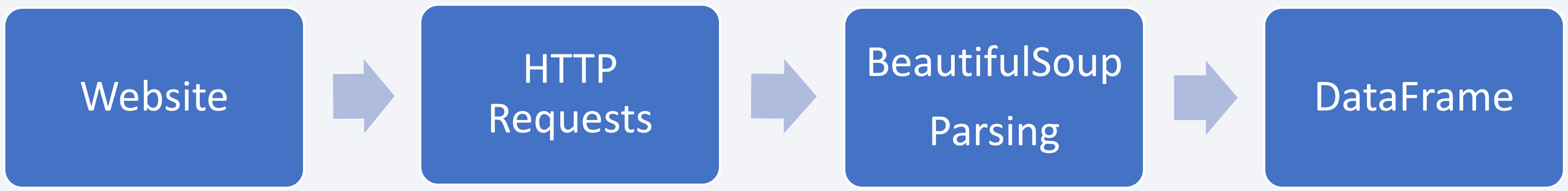


Data Collection - Scraping

Web Scraping:

In addition to the SpaceX API, web scraping was used to gather supplementary data, including more granular information such as the condition of the rocket boosters post-landing.

BeautifulSoup and **requests** libraries in Python were utilized to scrape data from the SpaceX website and other relevant sources. The scraped data was processed, cleaned, and integrated into the main dataset (See the Jupyter Notebook).



Data Wrangling-Data Cleaning

The data wrangling process involved cleaning, structuring, and transforming the collected data to ensure its readiness for analysis and machine learning modeling. Here is an overview of the key steps involved:

Data Cleaning:

Handling Missing Values: Missing values in key columns, such as payload mass or landing outcomes, were either filled with appropriate substitutes (e.g., median for numerical values) or removed if insufficient information was available.

Data Type Conversion: Certain fields, such as dates and boolean indicators (e.g., success or failure), were converted into the appropriate data types (e.g., datetime, integer) to facilitate analysis.

Data Wrangling-Feature Engineering

Feature Engineering:

Creating Target Variable (Class): A new column was created to represent the success (1) or failure (0) of the Falcon 9 first-stage landing. This binary target variable was essential for building the classification models.

Encoding Categorical Variables: Features such as **launch site**, **orbit type**, and **booster version** were encoded as numerical values using one-hot encoding, allowing machine learning models to process them effectively.

Data Wrangling- Data Normalization/Standardization

Data Normalization/Standardization:

Scaling: For models that are sensitive to the magnitude of feature values (e.g., SVM), the numerical features were standardized using **StandardScaler**. This ensured that features like payload mass and flight number were on the same scale.

(See the Jupyter notebook)

EDA with Data Visualization

During the exploratory data analysis (EDA) phase, several visualizations were created to uncover patterns and relationships in the SpaceX launch data. Each chart served a specific purpose in understanding the key features that impact the success of Falcon 9 landings(See the Jupyter Notebook).

Scatter Plot: Payload Mass vs. Launch Outcome:

- **Purpose**: This plot was used to visualize the relationship between the payload mass of the rockets and their landing outcomes (success or failure). It helped determine whether heavier payloads were more or less likely to result in successful landings.
- **Insight**: The chart revealed no strong linear correlation between payload mass and landing success, but it helped identify payload ranges that showed a tendency towards successful outcomes.

EDA with Data Visualization-continued

Bar Chart: Launch Success Rate by Orbit Type:

- **Purpose**: This bar chart showed the success rates for each orbit type. It was used to identify which orbit types had the highest or lowest rates of successful landings.
- **Insight**: The analysis revealed that certain orbit types consistently performed better than others, making them more reliable for successful landings.

Line Chart: Launch Success Yearly Trend:

- **Purpose**: The line chart was plotted to track the yearly trend of Falcon 9 landing success rates over time.
- **Insight**: The chart reveals a clear upward trend in landing success rates, particularly starting from 2013. This steady increase in successful landings continues through to 2020, reflecting the improvements in SpaceX's rocket technology and reusability efforts over time.

EDA with SQL

- Display the names of the unique launch sites.
- Display 5 records where launch sites begin with the string 'CCA’.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display the average payload mass carried by booster version F9 v1.1.
- List the date when the first successful landing outcome on a ground pad was achieved.
- List the names of boosters that successfully landed on drone ships and carried payloads between 4000 and 6000 kg.
- List the total number of successful and failed mission outcomes.
- List the names of booster versions that carried the maximum payload mass.
- List the records displaying month names, failure landing outcomes on drone ships, booster versions, and launch sites for the months in 2015.
- Rank the count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order.

See the Jupyter notebook

Build an Interactive Map with Folium

Summary of Map Objects Created and Added to a Folium Map (See the Jupyter notebook):

- **Markers**:
  - **Launch Sites**: A marker was added for each SpaceX launch site. These markers included labels showing the names of the launch sites.
  - **Outcome Markers**: For each launch, a marker was added based on the outcome. A **green marker** represented a successful launch, while a **red marker** indicated a failed launch.
- **Circles**:
  - **Launch Site Circles**: A circle with a radius of 1000 meters was added around each launch site.
- **PolyLines**:
  - **Lines to Points of Interest (POI)**: PolyLines were drawn from the launch site to points of interest such as the coastline, nearby cities, highways, and railways.
- **Distance Markers**:
  - For each point of interest, a marker was added showing the distance from the launch site.

Build an Interactive Map with Folium-Continued

Why These Objects Were Added:

- **Markers and Circles**: These provide a clear, labeled representation of each launch site and nearby features. Circles around launch sites help users understand the site's physical extent on the map.
- **Outcome Markers**: Color-coded markers (green for success, red for failure) allow users to immediately see which launches were successful and which failed, directly on the map. This visualization can highlight performance trends at specific launch sites.
- **PolyLines**: The lines connecting launch sites to nearby infrastructure like coastlines, railways, and highways visually convey proximity, helping users understand the logistical and environmental context of each launch site.
- **Distance Markers**: Showing the distance to nearby features can help in strategic planning, safety analysis, and evaluating the convenience of launch site locations in relation to transportation routes and urban areas.

Build a Dashboard with Plotly Dash

Plots/Graphs and Interactions Added to the Dashboard (See the Python Program)

- **Pie Chart: Launch Success Rates by Launch Site**
  - **Plot**: A pie chart displaying the proportion of successful launches by all launch sites.
- **Scatter Plot: Correlation Between Payload Mass and Launch Success**
  - **Plot**: A scatter plot that shows the relationship between payload mass and launch outcome (success or failure).
- **Range Slider: Filter by Payload Mass**
  - **Plot**: This interactive range slider component allows users to select a range of payload masses.
- **Dropdown Menu: Filter by Launch Site**
  - **Plot**: A dropdown menu component allows users to select a launch site or choose "All Sites."

Build a Dashboard with Plotly Dash-continued

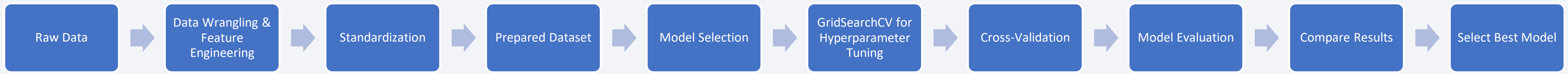
Why These Plots and Interactions Were Added

- **User-Friendly Data Exploration**: The combination of pie charts, scatter plots, dropdowns, and sliders makes the dashboard highly interactive, allowing users to filter and explore the data based on their interests and analytical needs. This interactivity enhances the user experience and provides actionable insights based on user selections.
- **Visualizing Relationships and Performance**: Each chart highlights a key aspect of the data, such as success rates at different sites, the relationship between payload mass and launch outcomes, or the overall success distribution. These visualizations make it easier for users to understand complex relationships within the data.
- **Customization and Exploration**: The sliders and dropdowns offer users control over the dashboard's display, enabling them to view data that is most relevant to their specific inquiries. This feature is crucial for a real-time data exploration tool, as it empowers users to dive into specific areas without needing to manually process the data.

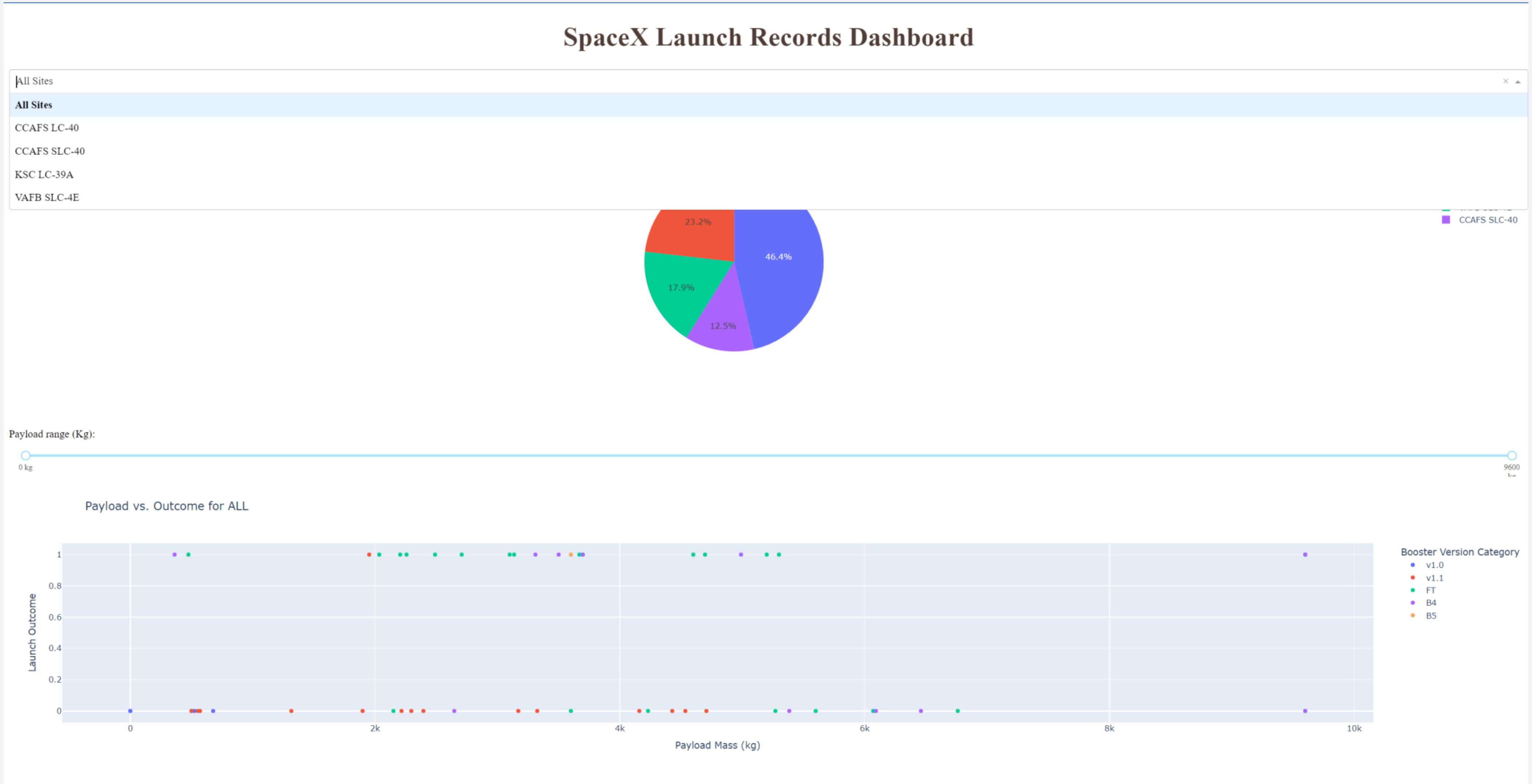
These plots and interactions collectively provide a comprehensive view of SpaceX’s launch performance, helping users to understand the impact of launch sites, payloads, and other factors on the success of Falcon 9 missions.

Predictive Analysis (Classification)

The process of building, evaluating, and improving a classification model to predict Falcon 9 landing success involved several key stages. It began with data preparation, where the data was collected through APIs and web scraping, and then cleaned and transformed. Key steps in this stage included handling missing values, encoding categorical variables (e.g., launch site, booster version), and creating a target variable to represent the success or failure of each landing. Standardization was also applied to numerical features like payload mass to ensure uniformity in model training. After the dataset was ready, the next stage was model selection and hyperparameter tuning. Four machine learning models were selected for experimentation: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (kNN). Each model was tuned using GridSearchCV with 10-fold cross-validation, allowing for the identification of the best hyperparameters for each model. This stage ensured that the models were optimized for accuracy, and cross-validation provided a robust method for assessing the models' performance across different subsets of the data. The models were then evaluated based on cross-validation scores and their test accuracy. The cross-validation scores provided an indication of each model's stability and performance across multiple folds, while the test accuracy demonstrated how well each model generalized to unseen data. This comparison enabled the identification of the best This structured approach ensured that the classification model was rigorously built, evaluated, and optimized, leading to the selection of the most suitable model for predicting Falcon 9 landing outcomes.-performing model based on its ability to predict landing success consistently. (See the Jupyter notebook)



Results



SpaceX Launch Records Dashboard
All Sites
All Sites
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E
23.2%
46.4%
17.9%
12.5%
CCAFS SLC-40
Payload range (Kg):
0 kg
9600
Payload vs. Outcome for ALL
Launch Outcome
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

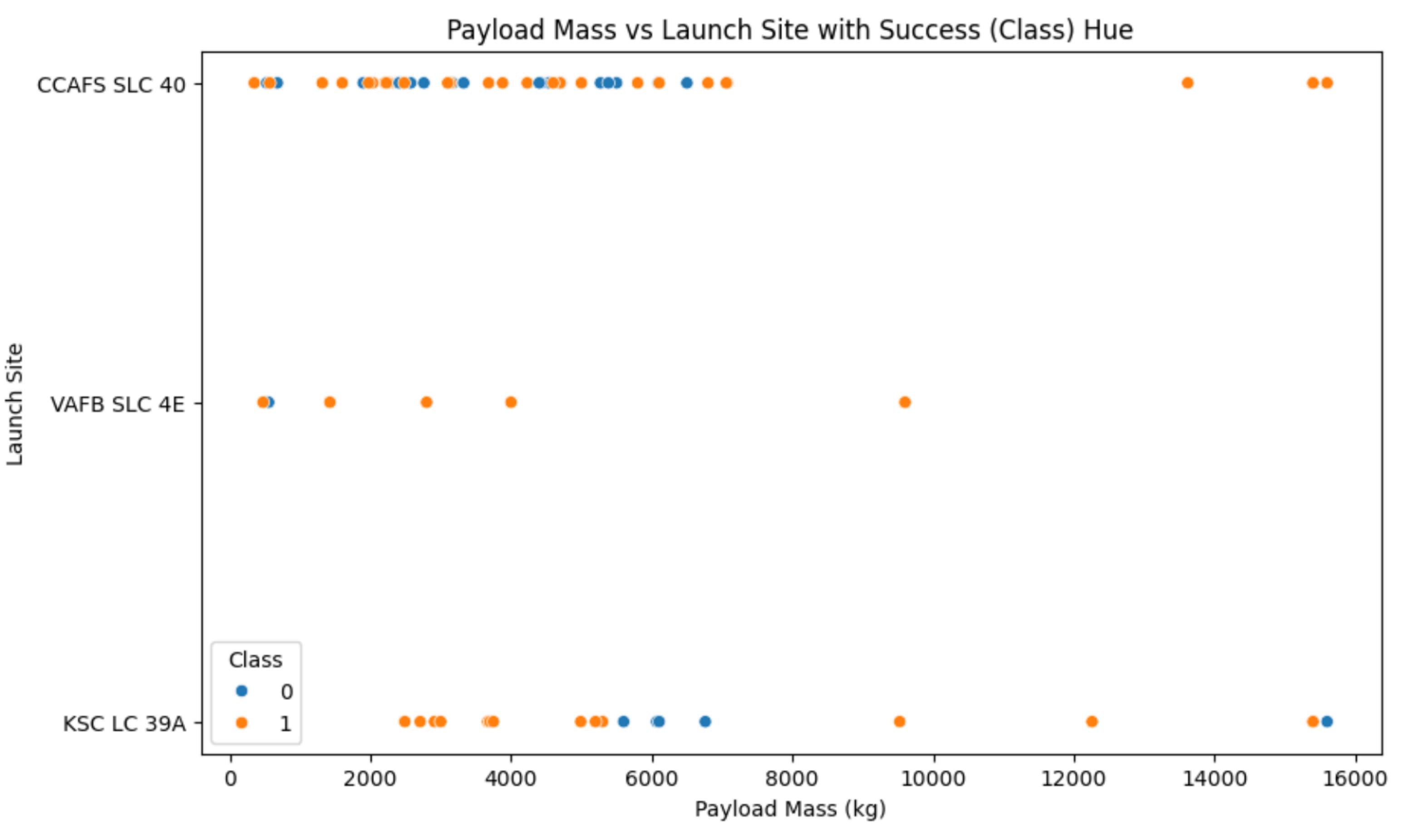
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

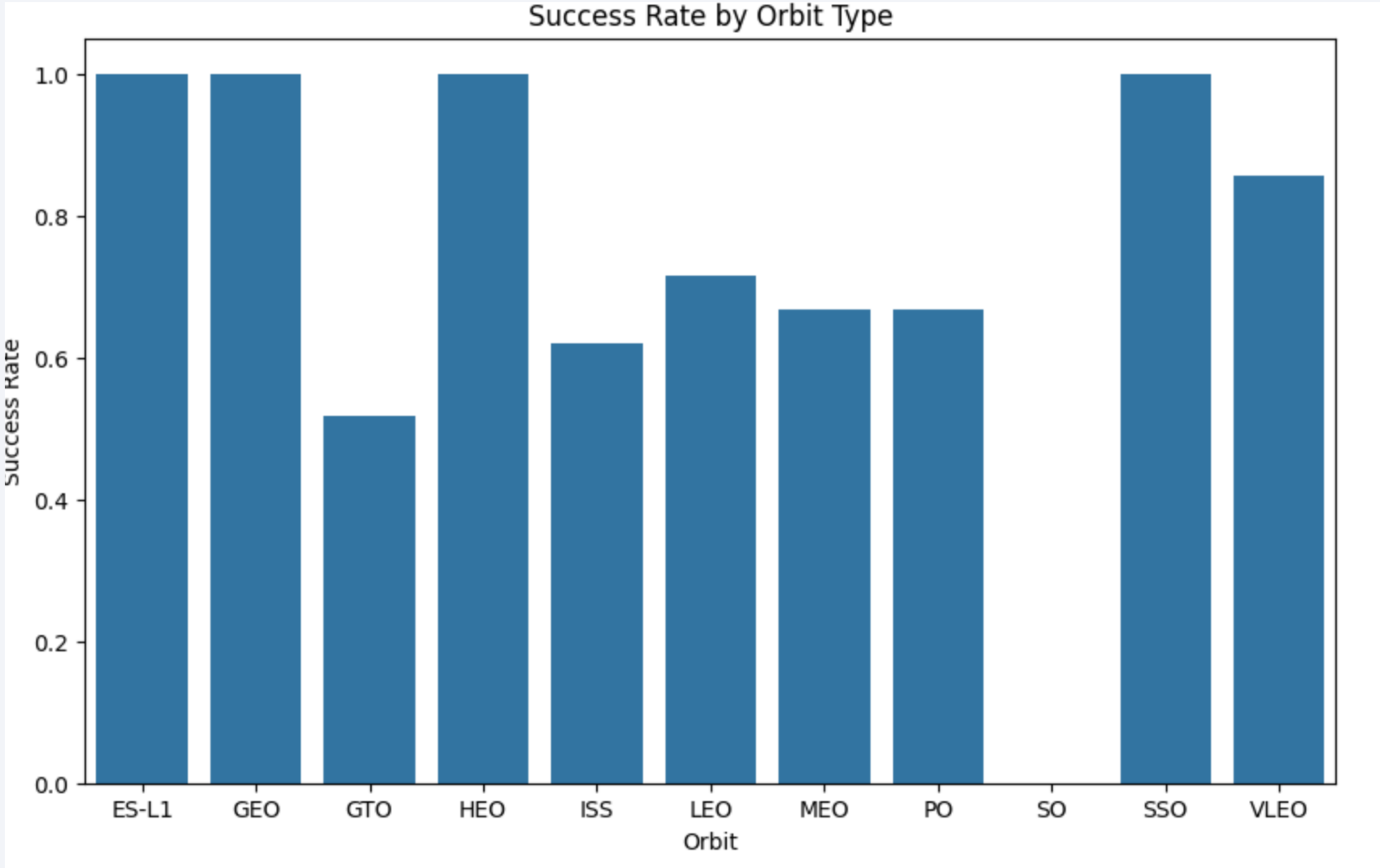
This scatter plot highlights the improvement in landing success over time (represented by flight number), particularly at the CCAFS SLC 40 site, which showed more early failures compared to VAFB SLC 4E and KSC LC 39A. It also indicates that VAFB SLC 4E and KSC LC 39A had a stronger performance with fewer failures overall, possibly due to optimized conditions or better preparedness at these sites.

Payload vs. Launch Site



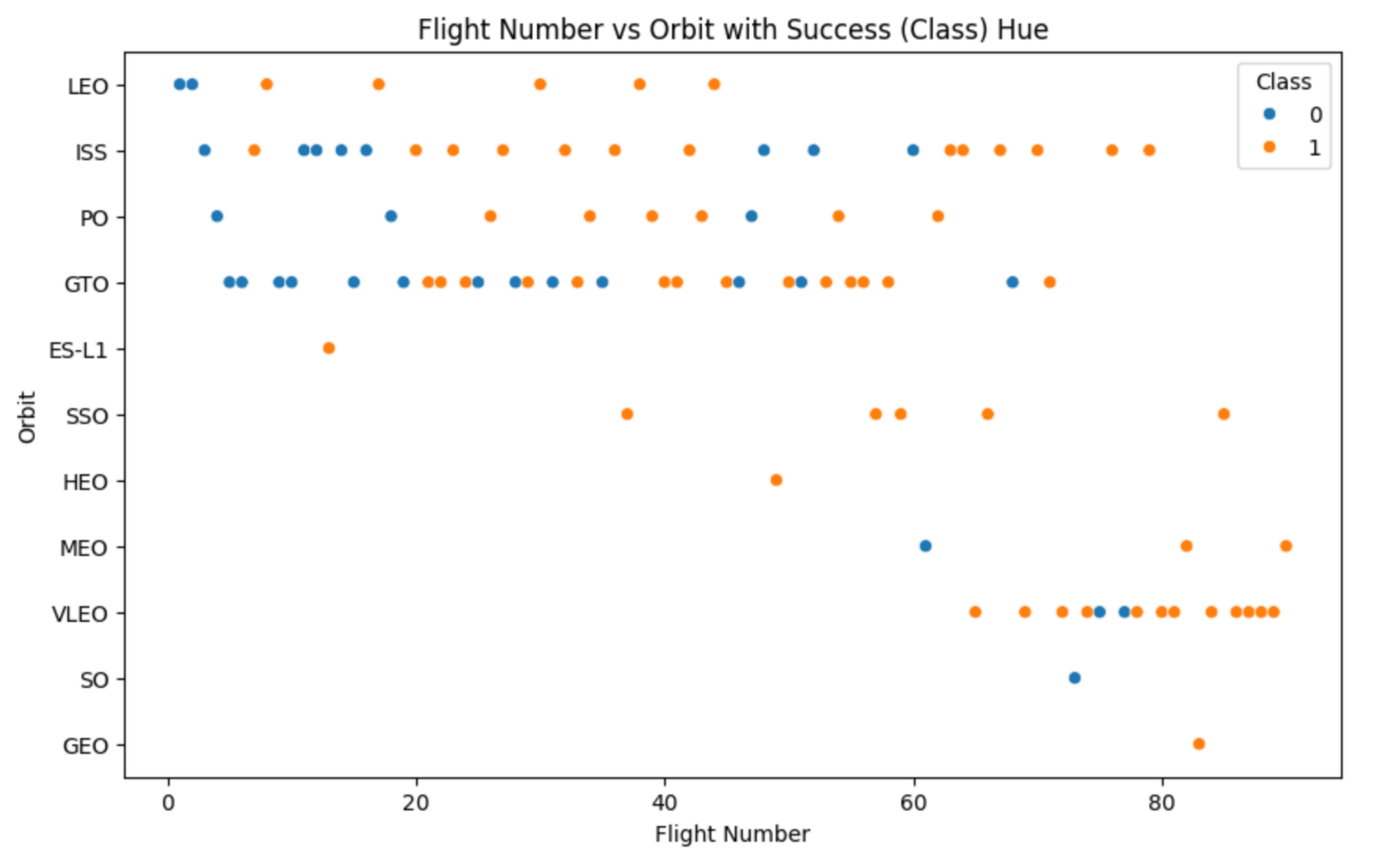
This scatter plot demonstrates that the relationship between payload mass and landing success is not straightforward, as both successful and failed landings occur across a wide range of payloads. However, sites like **VAFB SLC 4E** and **KSC LC 39A** show higher success rates, especially at larger payload masses. This suggests that other factors, possibly related to site conditions or mission specifics, play a significant role in landing outcomes.

Success Rate vs. Orbit Type



This bar chart highlights the varying success rates of SpaceX launches across different orbit types, suggesting that certain orbits, such as **GEO**, **SSO**, and **ISS**, are more reliable, while **GTO** missions present higher challenges, leading to more frequent failures. This insight can help focus efforts on improving mission outcomes for specific orbit types with lower success rates.

Flight Number vs. Orbit Type



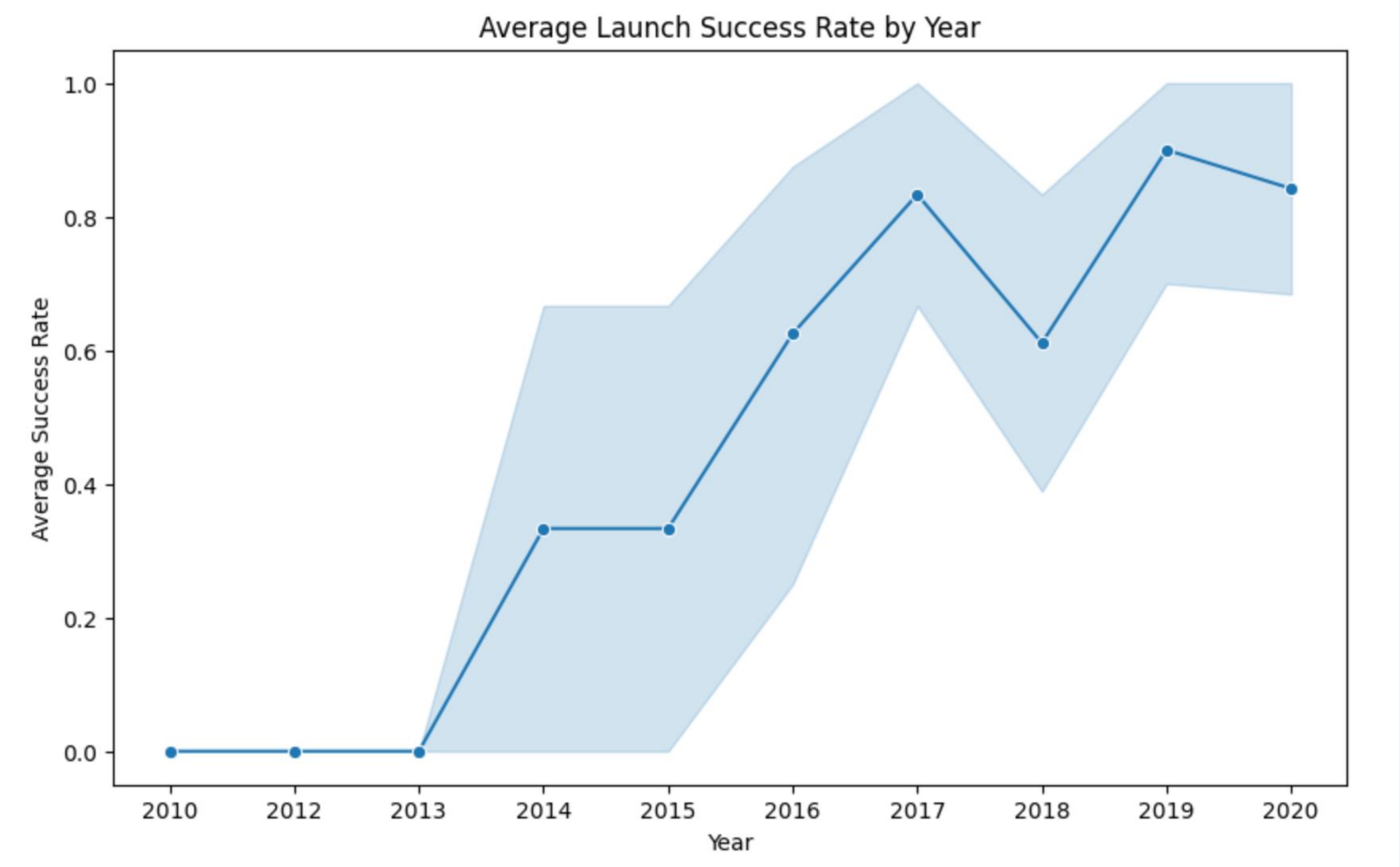
The plot shows a clear trend of increased success as SpaceX gained more experience across various orbits, especially for **LEO**, **ISS**, and **VLEO** missions. However, certain orbits like **GTO** continue to present challenges, with frequent failures occurring even at higher flight numbers. This scatter plot underscores the complexity of different orbit types and the technological advancements SpaceX made over time to improve landing outcomes.

Payload vs. Orbit Type

The scatter plot shows that payload mass does not strongly predict landing success across all orbit types, as successful and failed landings occur for a wide range of payloads. However, orbits like **ISS**, **SSO**, and **PO** show a higher success rate for most payload masses, while **GTO** presents more challenges, especially for larger payloads. This suggests that the complexity of certain orbits plays a significant role in the success of SpaceX's landing attempts, beyond just payload mass.

Launch Success Yearly Trend

The chart demonstrates SpaceX's progress over the years in achieving higher success rates for rocket landings. Starting with a low success rate in the early years, the company made rapid advancements, reaching nearly 100% success in some years. The minor fluctuations observed in **2017** and **2020** indicate occasional setbacks, but the overall trend is one of consistent improvement, showcasing the effectiveness of SpaceX's reusability and landing strategies.



All Launch Site Names

SQL: SELECT DISTINCT Launch_Site FROM spacextable;

Result:
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

This query retrieves the unique launch site names from the dataset, allowing you to see all the locations where Falcon 9 rockets were launched. Knowing the launch sites is essential for further analysis, such as determining which sites are more successful or frequently used.

Launch Site Names Begin with 'CCA'

SQL: SELECT * FROM spacextable WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
Result:
Booster Version: F9 v1.0 (various versions)
Payload Mass: Varies from 0 to 677 kg
Mission Outcome: Success for all 5 records
Launch Site: CCAFS LC-40
Landing Outcome: Mixed (Failures and No Attempts)

This query filters the records to display only those launches that occurred at sites beginning with "CCA." It provides the first five records from the dataset where the launch sites start with "CCA", which includes locations like **CCAFS LC-40** and **CCAFS SLC-40**. This is useful for understanding the activity at Cape Canaveral Air Force Station (CCAFS), one of SpaceX’s key launch sites.

Total Payload Mass

SQL: SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass

FROM spacextable

WHERE Customer LIKE '%NASA (CRS)%’;

Result: Total Payload Mass: 48,213 kg

This query calculates the total payload mass carried by boosters launched for NASA's **Commercial Resupply Services (CRS)** missions. The query filters the records to include only those where the customer is NASA (CRS), then sums the payload mass of all such launches.

Average Payload Mass by F9 v1.1

SQL: SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass

FROM spacextable

WHERE Booster_Version = 'F9 v1.1';

Result: Average Payload Mass: 2,928.4 kg

This query calculates the average payload mass carried by launches using the **F9 v1.1** booster version. The AVG function is applied to the PAYLOAD_MASS__KG_ column to compute the mean mass, and the results are filtered to only include records where the booster version is F9 v1.1.

First Successful Ground Landing Date

SQL: SELECT MIN(Date) AS First_Successful_Landing

FROM spacextable

WHERE Landing_Outcome = 'Success (ground pad)’;

Result: First Successful Ground Pad Landing: December 22, 2015

This query identifies the date of the first successful landing on a ground pad by using the MIN function to retrieve the earliest date where the landing outcome was 'Success (ground pad)'. This milestone is important for SpaceX, as it marks the beginning of their ability to land rockets reliably on ground-based pads, a key feature for rocket reusability.

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL: SELECT Booster_Version

FROM spacextable

WHERE Landing_Outcome = 'Success (drone ship)'

AND PAYLOAD_MASS__KG_ > 4000

AND PAYLOAD_MASS__KG_ < 6000;

Result:

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

F9 FT B1022

This query retrieves the names of the **booster versions** that successfully landed on a drone ship and carried a payload mass between 4000 and 6000 kg. By filtering based on landing outcome and payload mass, the query focuses on specific launches that meet both criteria.

Total Number of Successful and Failure Mission Outcomes

SQL: SELECT Mission_Outcome, COUNT(*) as Outcome_Count

FROM spacextable

GROUP BY Mission_Outcome;

Result:

Mission Outcome Counts:

- Failure (in flight): 1
- Success: 98
- Success: 1 (appears to be a duplication, likely due to data inconsistency)
- Success (payload status unclear): 1

This query calculates the total number of missions with each type of outcome by grouping the data by the Mission_Outcome column and counting the occurrences of each outcome. It provides an overview of how many missions resulted in success, failure, or unclear outcomes.

Boosters Carried Maximum Payload

SQL: SELECT Booster_Version
FROM spacextable
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM spacextable);

Result:

Booster Versions:

- F9 B5 B1048.4
- F9 B5 B1049.4
- F9 B5 B1051.3
- F9 B5 B1056.4
- F9 B5 B1048.5
- F9 B5 B1051.4
- F9 B5 B1049.5
- F9 B5 B1060.2
- F9 B5 B1058.3
- F9 B5 B1051.6
- F9 B5 B1060.3
- F9 B5 B1049.7

This query identifies the **booster versions** that carried the heaviest payload by using a subquery to find the maximum payload mass in the dataset, and then selecting the booster versions that match this value. This approach ensures that only the rockets with the largest payloads are listed.

2015 Launch Records

SQL: SELECT substr(Date, 6, 2) AS Month, Booster_Version, Launch_Site, Landing_Outcome

FROM spacextable

WHERE Landing_Outcome = 'Failure (drone ship)'

AND substr(Date, 0, 5) = '2015';

Result:

Month: January

- Booster Version: F9 v1.1 B1012
- Launch Site: CCAFS LC-40
- Landing Outcome: Failure (drone ship)

Month: April

- Booster Version: F9 v1.1 B1015
- Launch Site: CCAFS LC-40
- Landing Outcome: Failure (drone ship)

This query retrieves the records of launches in **2015** where the **landing on a drone ship** failed. It also extracts the month from the launch date using substr(Date, 6, 2) to display it separately and filters the data by year (2015) using substr(Date, 0, 5) = '2015'. The output includes the booster version, launch site, and the failure outcome.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
SQL: SELECT Landing_Outcome, COUNT(*) as Outcome_Count

FROM spacextable

WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'

GROUP BY Landing_Outcome

ORDER BY Outcome_Count DESC;
```

Result:

- No attempt: 10 occurrences
- Success (drone ship): 5 occurrences
- Failure (drone ship): 5 occurrences
- Success (ground pad): 3 occurrences
- Controlled (ocean): 3 occurrences
- Uncontrolled (ocean): 2 occurrences
- Failure (parachute): 2 occurrences
- Precluded (drone ship): 1 occurrence

This query ranks the different landing outcomes (e.g., "Success (drone ship)", "Failure (ground pad)") in descending order of occurrence between the specified date range of **2010-06-04** and **2017-03-20**. The query groups the results by Landing_Outcome, counts the occurrences of each outcome, and orders them from most frequent to least frequent.

Section 3

Launch Sites Proximities Analysis

Folium Map that Shows All Launch Sites

This map visualization shows the locations of key SpaceX launch sites across the United States.

Launch Sites:

VAFB SLC-4E (Vandenberg Air Force Base):

Located on the west coast near Los Angeles, California, this site is prominently marked on the map.

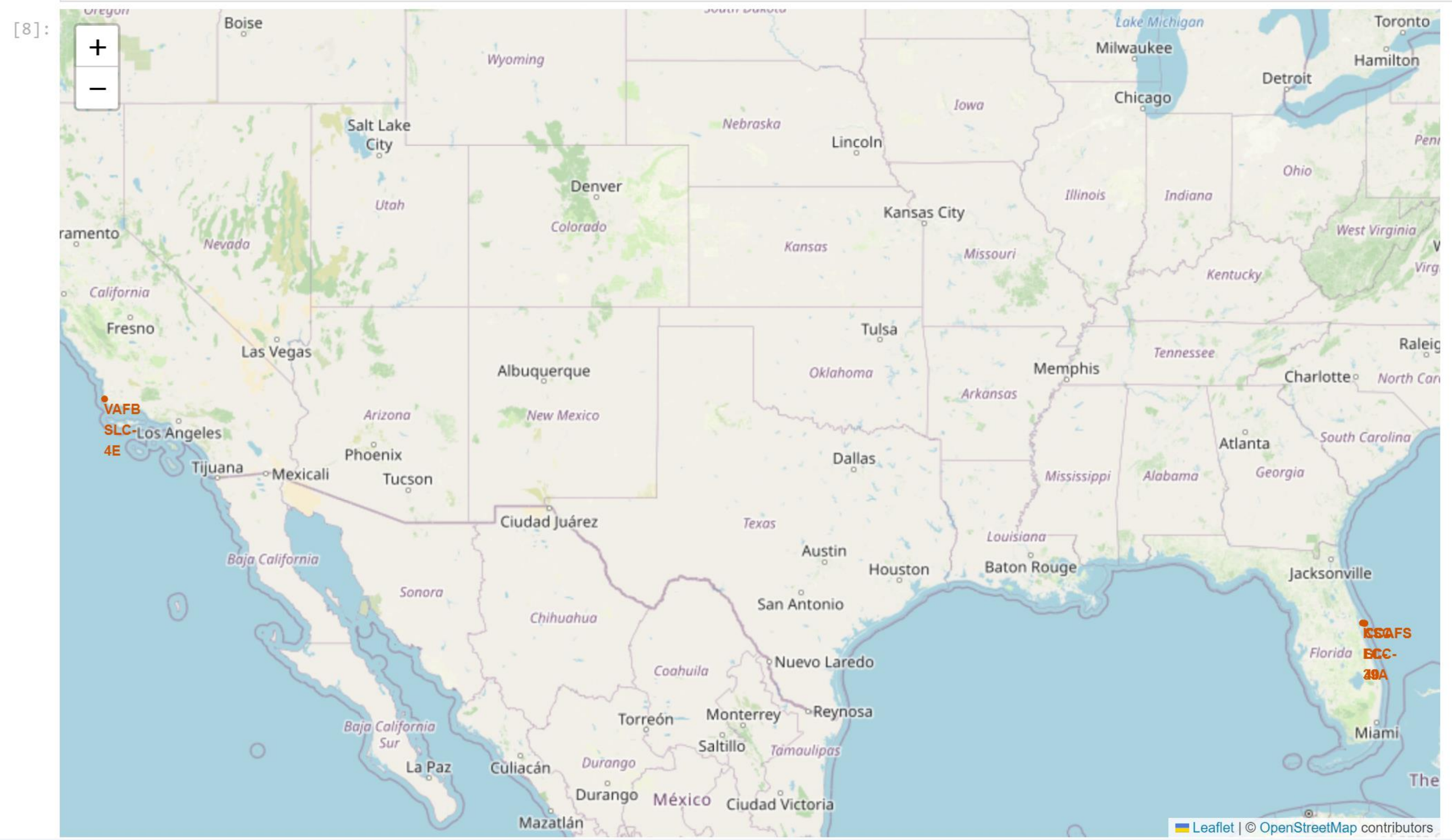
This site is typically used for polar orbit launches (e.g., Sun-Synchronous Orbits) as its location provides an ideal trajectory for launching satellites into polar orbits.

CCAFS LC-40 (Cape Canaveral Air Force Station) and **KSC LC-39A (Kennedy Space Center)**:

Located in Florida, these two sites are some of the most important launchpads for SpaceX.

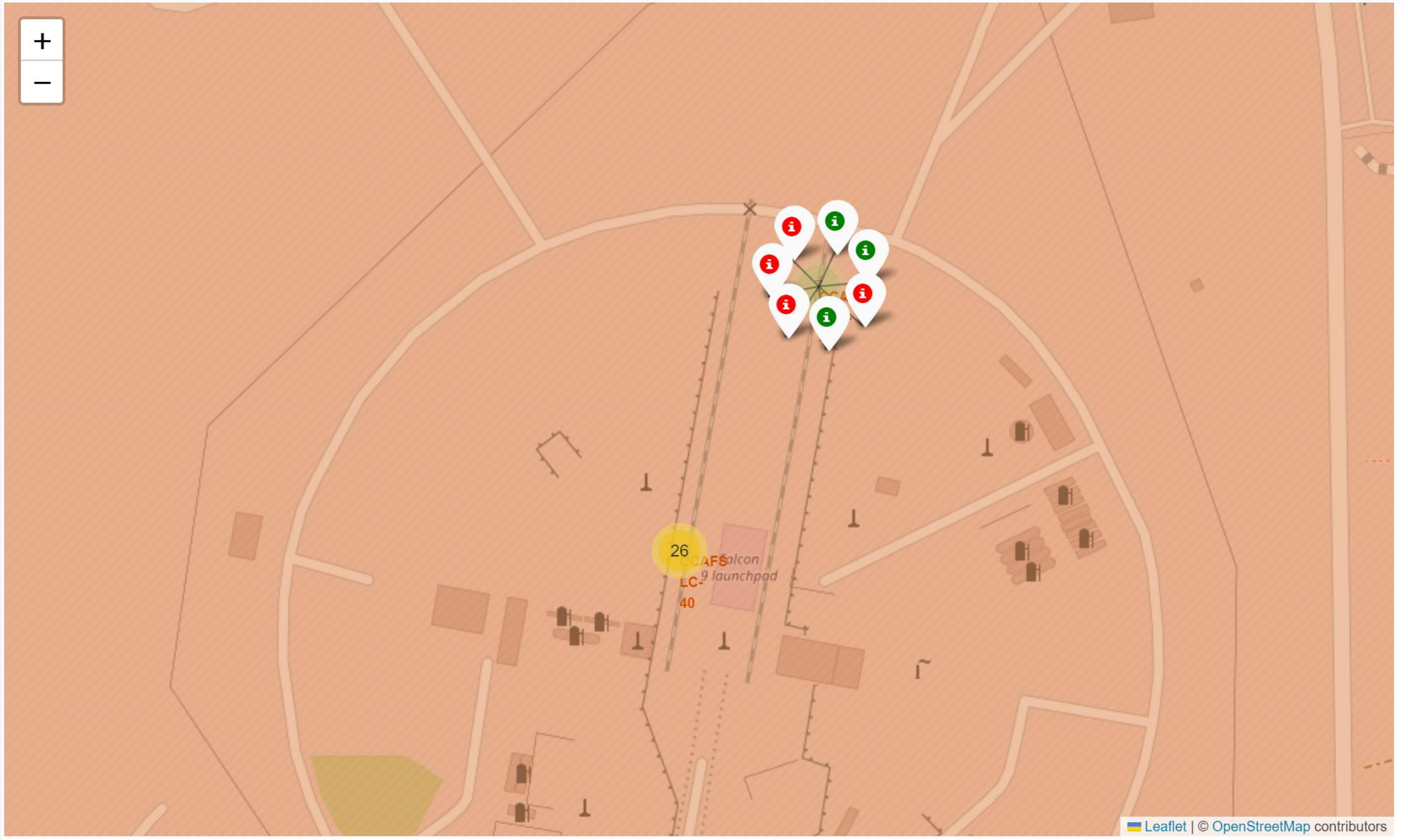
LC-39A is especially significant as it has been used for numerous historical missions, including NASA missions and commercial satellite launches.

LC-40 is similarly important for geostationary and other launches.



Folium Map that shows color-labled launch outcomes

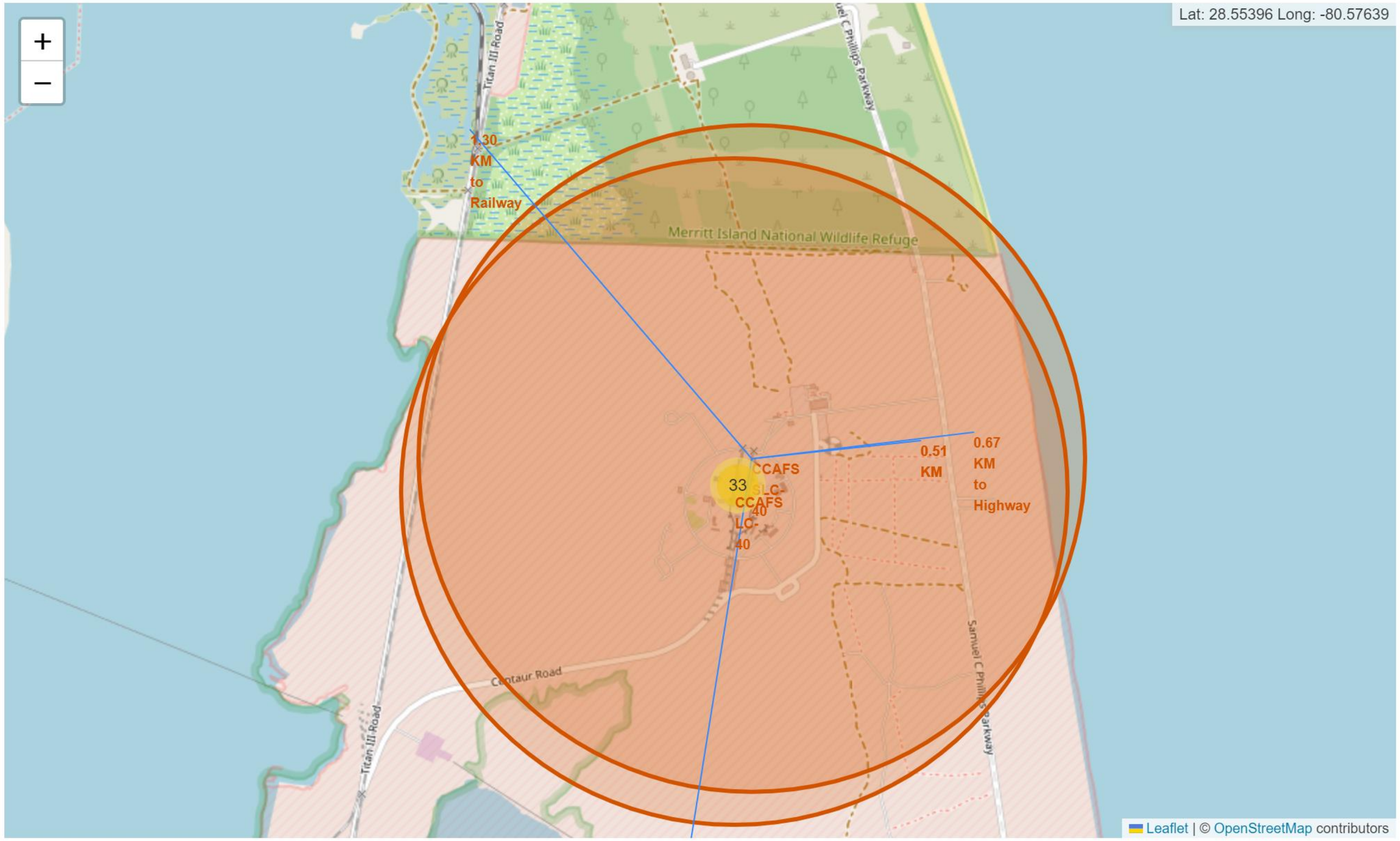
This map provides detailed insights into SpaceX's launch sites. The interactive features and detailed markers allow users to explore each mission's success or failure, contributing to a better understanding of SpaceX's performance at each launch site.



Folium Map that shows launch site's proximities

This map provides a detailed view of the proximity between launch sites and their surrounding infrastructure, highlighting its strategic location near key transport routes while maintaining necessary safety distances from the launchpad. Understanding these distances is crucial for logistical planning and maintaining safety during launch operations.

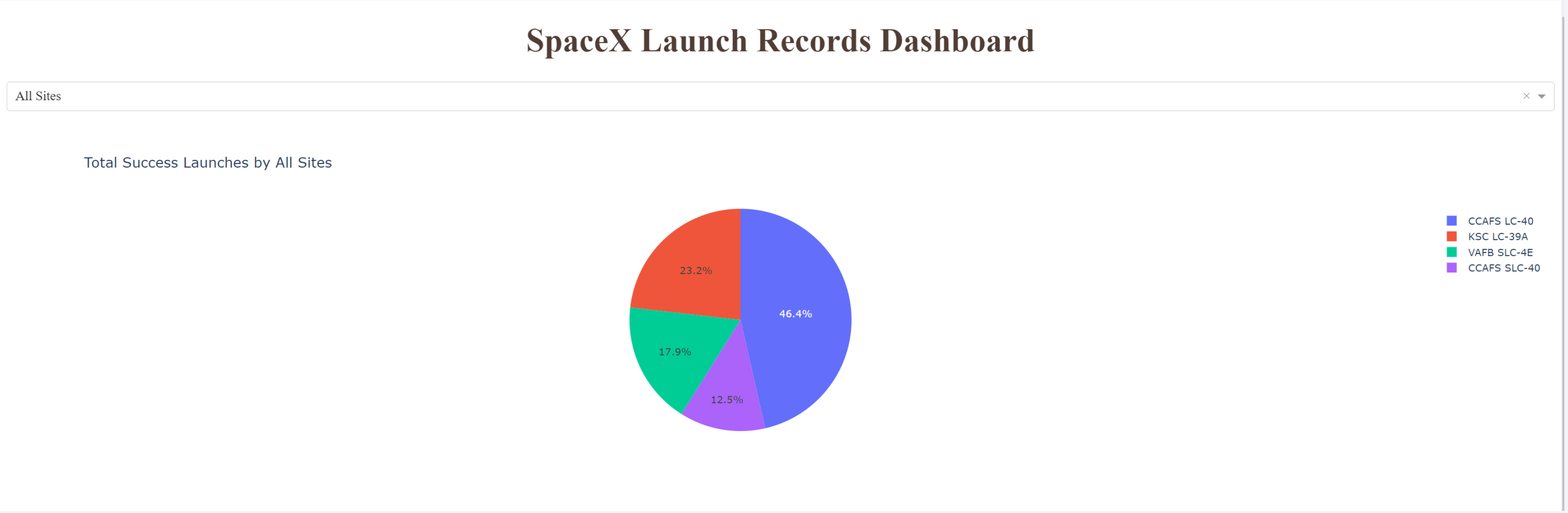
[22]:



Section 4

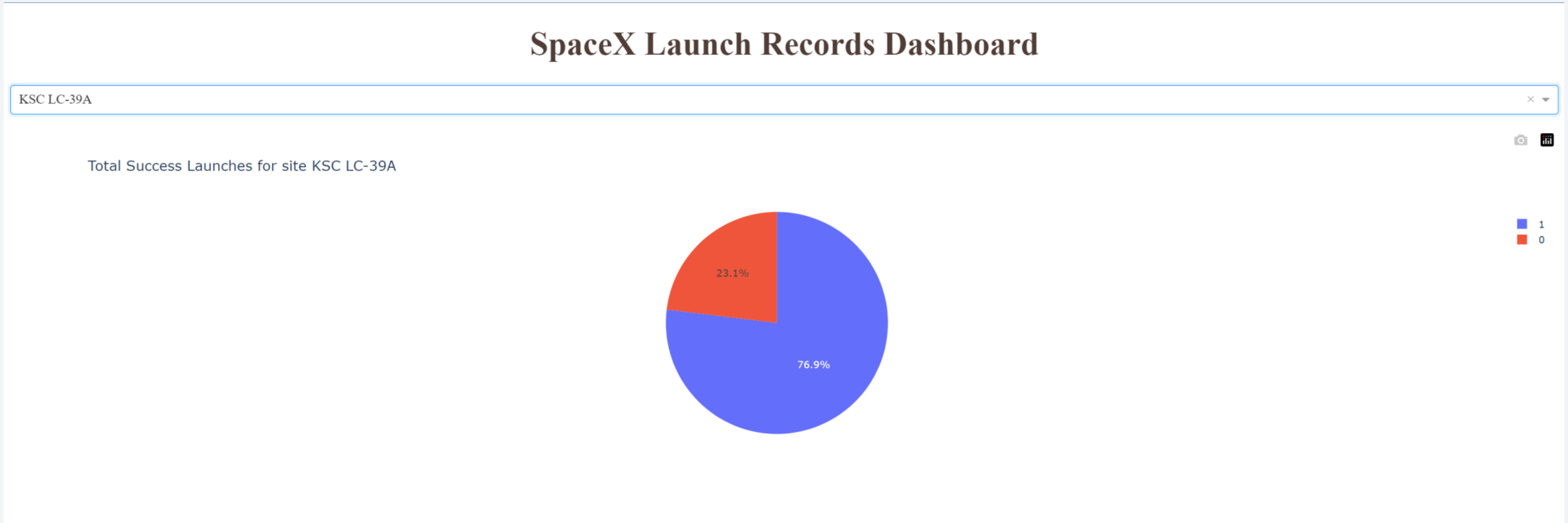
Build a Dashboard with Plotly Dash

Dashboard that shows launch success count for all sites



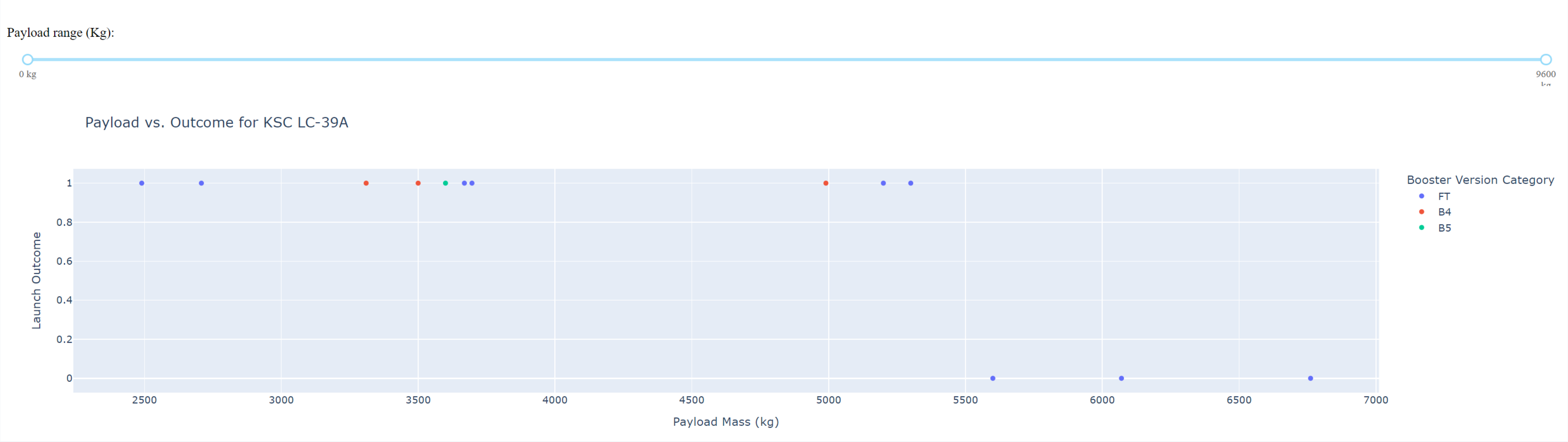
This dashboard provides an easy-to-understand visualization of SpaceX's successful launches across multiple sites. **CCAFS LC-40** emerges as the most prominent launch site, followed by **KSC LC-39A**, emphasizing the importance of these two locations in SpaceX's operations. The interactive dropdown menu allows users to explore the data further by filtering specific sites, making this dashboard a valuable tool for analyzing SpaceX's launch performance.

launch site with highest launch success ratio



This pie chart highlights the strong performance of KSC LC-39A, with the highest success rate of 76.9%. The high success rate makes KSC LC-39A a key site for SpaceX's most reliable and successful launches.

Payload vs. Launch Outcome Scatter Plot for All Sites



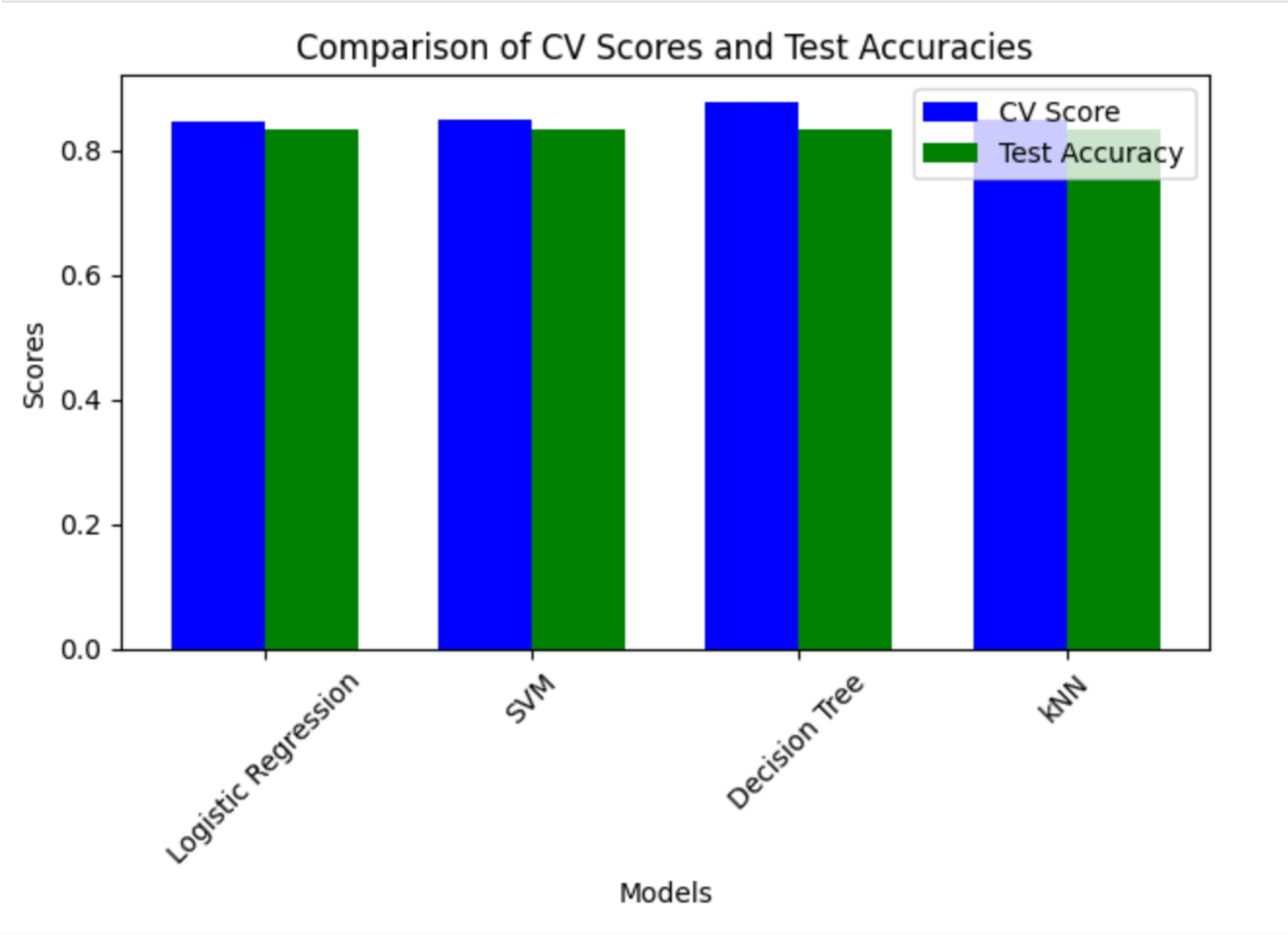
This scatter plot visualizes the relationship between payload mass and launch outcome for **KSC LC-39A**, showing that both payload mass and booster version affect the likelihood of success. **Block 5 (B5)** boosters are associated with more successful launches, and higher payload masses also tend to be more successful, especially with newer booster versions. The interactive **payload range slider** allows users to filter the data for more focused insights.

Section 5

Predictive Analysis (Classification)

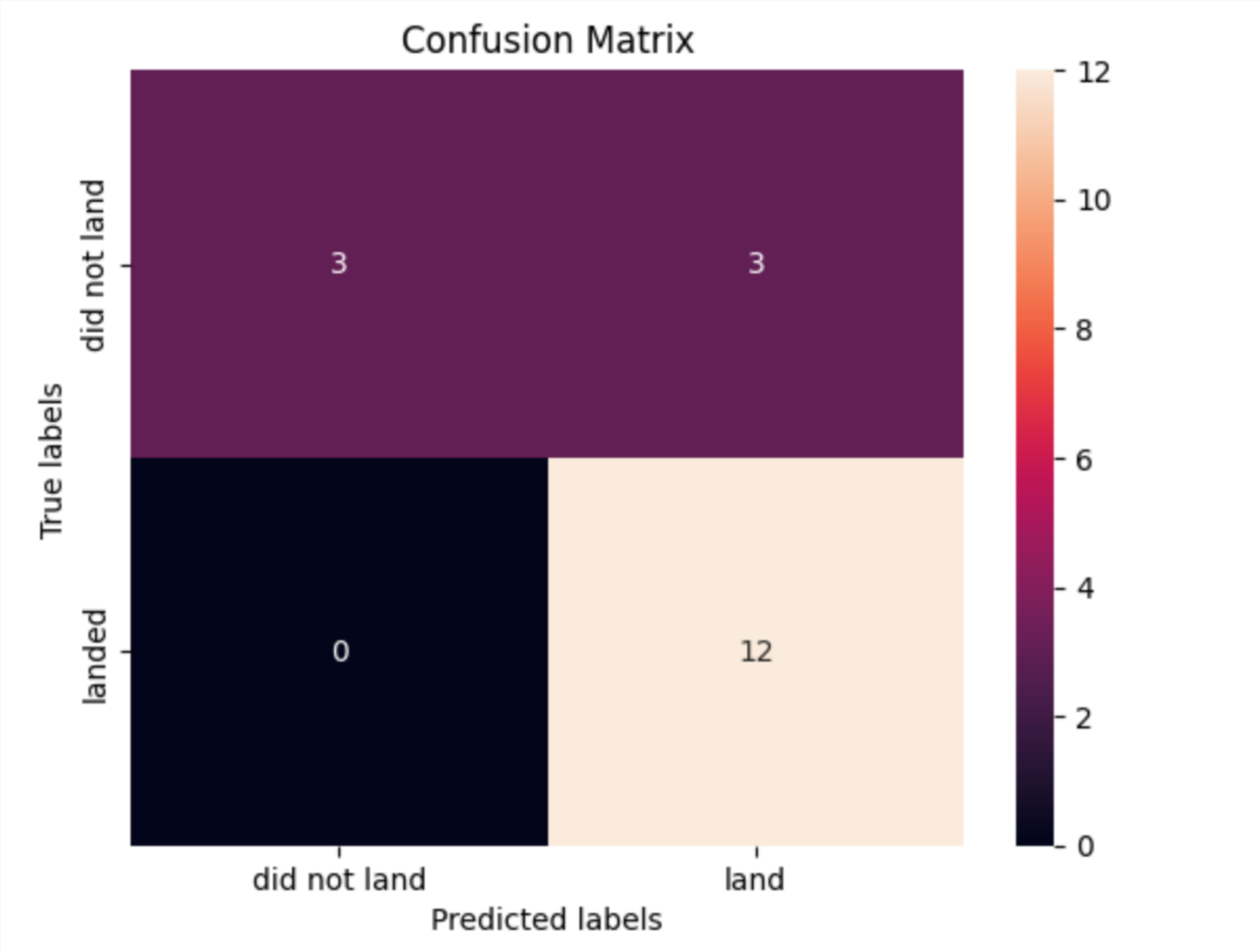
Classification Accuracy

While all models have the same **Test Accuracy**, the **Decision Tree** model stands out with the highest **CV Score**, indicating that it may generalize better based on cross-validation.



Confusion Matrix

The model performs well with more accurate predictions of landings than failures, but the presence of **3 false positives** shows room for improvement in identifying when a rocket will fail to land.



Conclusions

Key Takeaways:

- Through EDA, we explored relationships between features like launch site, payload mass, and launch outcome. We found that KSC LC-39A had a significantly higher success rate than other sites, which could inform future launch site selection.
- Several machine learning models, including Logistic Regression, SVM, Decision Tree, and kNN, were trained to predict launch success based on features like payload mass, booster version, and orbit type. Among these, the Decision Tree model performed best, with a CV score of 87.68%, outperforming others in cross-validation, though the test accuracy remained the same across all models (83.33%).
- Key features like payload mass and booster version played an important role in predicting launch success. The analysis showed that newer booster versions like Block 5 (B5) were more reliable than earlier versions. The transition from earlier booster versions (like Block 4) to the Block 5 variant improved success rates significantly. This suggests that engineering improvements over time have had a tangible impact on SpaceX's launch success.

Recommendations:

- **Further Model Tuning**: Given that the Decision Tree performed best in cross-validation, further hyperparameter tuning and feature engineering could improve predictive accuracy.
- **Boosters and Launch Success**: SpaceX should continue focusing on upgrading booster versions like Block 5, which showed higher success rates.
- **Launch Site Optimization**: The higher success rates at KSC LC-39A suggest that SpaceX might consider increasing launches from this site while reviewing why CCAFS SLC-40 had a relatively lower success rate.

Appendix

All the project materials are stored on GitHub @ https://github.com/Ziyu-Jiang/IBM_DS_Capstone/tree/main

Thank you!