IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alyza Paige Ng
January 27, 2024

Outline

I. Executive Summary

II. Introduction

III. Methodology

IV. Results

V. Conclusion

VI. Appendix

Executive Summary

Objective: Assess SpaceY's potential to compete with SpaceX by predicting the success of Falcon 9's first-stage landings using data science methodologies.

Methodology Overview:

1. Data Collection: Utilized SpaceX's public API to gather launch data.
2. Data Wrangling: Filter, cleaned, and one hot encoded values and variables appropriately.
3. Exploratory Data Analysis (EDA): Conducted EDA using SQL queries and data visualizations.
4. Interactive Visual Analytics: Developed maps with Folium and dashboards with Plotly Dash.
5. Predictive Modeling: Built and evaluated classification models.

Executive Summary

Objective: Assess SpaceY's potential to compete with SpaceX by predicting the success of Falcon 9's first-stage landings using data science methodologies.

Results Overview:

1. <u>Insights from EDA:</u>
    - Identified factors influencing landing success, such as payload mass, launch site, and orbit type.
    - Observed an increasing trend in successful landings over recent years.
2. <u>Predictive Analysis:</u>
    - Decision Tree Classifier achieved the highest *training* accuracy (90%) in predicting landing success.
    - Model validation confirmed reliability across models with *test* data accuracy about 83%.

Introduction

Background and Context:

- The space race began in 1957, with countries competing to expand human capabilities beyond Earth.
- Traditional space missions are costly, averaging $165 million per launch.
- SpaceX revolutionized space exploration by reducing launch costs to $60 million, leveraging advanced reusable rocket technologies.

Problem Statements:

- *What attributes influence the success of rocket launches, particularly the first-stage landings?*
- *How can data science be used to analyze and model key factors like payload mass, launch locations, and orbits?*
- *Can SpaceY replicate or improve SpaceX's strategies to achieve competitive success?*

Section 1

Methodology

I. Methodology: Overview

Step 1: Data Collection

- Collected data through SpaceX REST API and web scraping from Wikipedia pages.

Step 2: Data Wrangling

- Processed data using Pandas and NumPy.
- Techniques included OneHot Encoding, removing unnecessary columns, and normalizing and standardizing data.

Step 3: Exploratory Data Analysis (EDA):

- Conducted visualizations using Seaborn and Matplotlib.
- Utilized SQL for in-depth data queries.

Step 4: Interactive Visual Analytics:

- Created interactive maps using Folium.
- Built dynamic dashboards with Plotly Dash.

Step 5: Predictive Modeling:

- Split data into training and testing sets.
- Tuned hyperparameters using Grid Search to identify the best model.
- Adopted the best algorithm for deployment.

I. Methodology: Data Collection

Data Sources:

- **SpaceX API:** An open-source REST API providing comprehensive data on SpaceX launches, rockets, capsules, and more.
- **Wikipedia:** A free online encyclopedia offering detailed articles on various topics, including SpaceX missions and launch histories.

Data Collection Process:

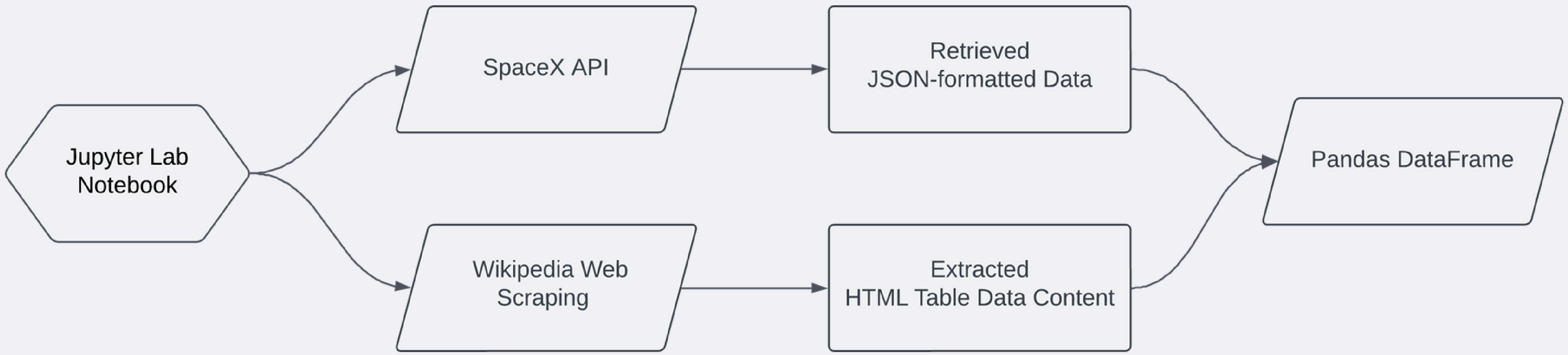


Figure 1.1: Data Collection Process

I. Methodology: Data Collection, SpaceX API

Accessing SpaceX API:

- Sent HTTP GET requests to retrieve JSON-formatted data on past and upcoming launches.
- Parsed the JSON responses to extract: flight numbers, mission names, launch dates, rocket types, payload details, and landing outcomes.
- Converted the extracted data into a structured Pandas DataFrame.

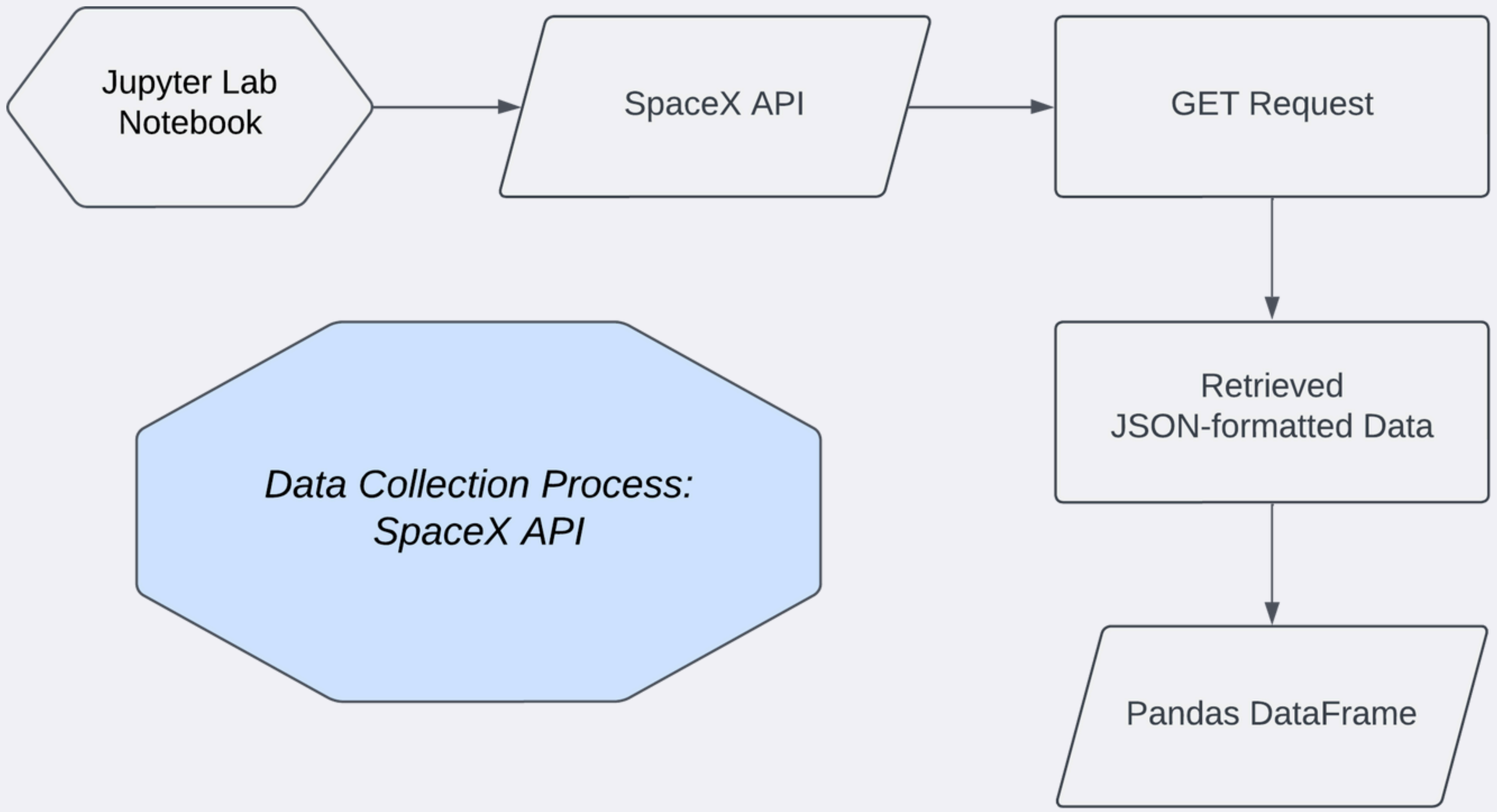


Figure 1.2: Data Collection Process: SpaceX API

I. Methodology: Data Collection, Wikipedia Web Scraping

Wikipedia Web Scraping:

- Identified target Wikipedia pages containing tables of Falcon 9 and Falcon Heavy launch records.
- Utilized Python's BeautifulSoup library to parse the HTML content of pages.
- Extracted table data, focusing on columns like launch dates, payloads, launch sites, and mission outcomes.
- Cleaned and structured the scraped data into Pandas DataFrames.

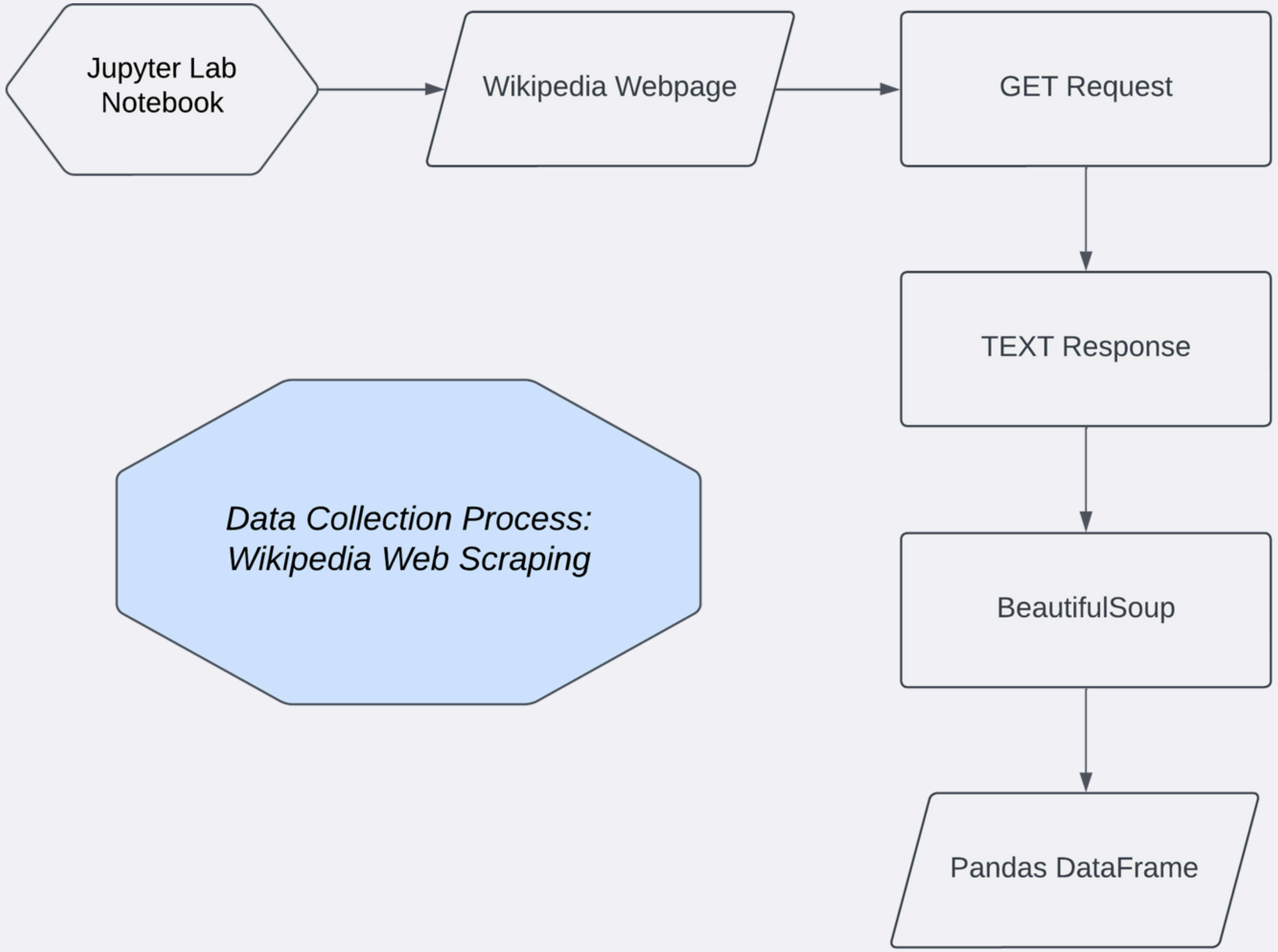


Figure 1.3: Data Collection Process: Wikipedia Web Scraping

I. Methodology: Data Wrangling

Purpose: To prepare the collected datasets for exploratory data analysis and machine learning by ensuring data quality and consistency.

Data Wrangling Stages:

1. Loading the Collected Dataset

2. Identifying and calculating the percentage of the missing values in each attribute.

3. Identifying which columns are numerical and categorical.

4. Calculating the number of launches on each site.

5. Calculating the number and occurrence of each orbit.

6. Creating a landing outcome label from Outcome column.

7. Determining the success rate of returning the first stage of the rocket.

I. Methodology: EDA and Visualization

Purpose: To uncover patterns and insights by examining relationships between features and the target variable.

Tools Utilized:

1. Matplotlib: A library for creating static, animated, and interactive visualizations in Python.
2. Seaborn: A library based on Matplotlib, providing a high-level interface for drawing statistical graphics.

Key Visualizations:

Flight Number vs. Launch Site: Analyzed flight number distributions across launch sites to assess site utilization over time.

Payload vs. Launch Site: Examined payload mass variations by launch site to gain insights into sites' capacity and mission types.

Payload vs. Orbit Type: Explored the payload mass and orbit type to understand mission profiles and payload requirements.

Success Rate by Orbit Type: Calculated and visualized mission success rates for each orbit type, with probabilities.

Flight Number vs. Orbit Type: Investigated correlations on flight numbers and orbit types for trends and launch strategies over time.

Annual Launch Success Trends: Plotted yearly trends of launch successes to evaluate technological efficiency over time.

I. Methodology: EDA with SQL

Purpose: To perform exploratory data analysis by executing SQL queries on the collected dataset, gaining structured insights into the launch data.

SQL Query Higlights:

Unique Launch Sites:
Identified distinct launch sites used in space missions.

Launch Sites Starting with 'CCA':
Retrieved five records of launch sites beginning with the prefix 'CCA' for pattern analysis.

Payload Mass by NASA (CRS):
Calculated the total payload mass carried by boosters in NASA's CRS missions.

First Successful Ground Pad Landing:
Extracted the date of the first successful landing outcome on a ground pad.

Successful Drone Ship Landings:
Listed boosters that succeeded on a drone ship with payloads between 4000 and 6000 kg.

Mission Outcomes:
Counted the total number of successful and failed mission outcomes for statistical overview.

Maximum Payload Booster Versions:
Identified booster versions with the highest payload mass using a subquery.

Failed Drone Ship Landings in 2015:
Detailed failed drone ship landings in 2015, including booster versions and launch site names.

Landing Outcomes Ranking:
Ranked landing outcomes by count within the date range 2010-06-04 to 2017-03-20 in descending order.

I. Methodology: Interactive Visual Analytics with Folium

Implementation:

1. Drawing Launch Site Circles

- Plotted circles on four distinct Falcon 9 rocket launch sites, shown in Table 1.1.

2. Adding Success/Failure Markers

- Placed markers at each launch site to indicate the outcome of the first-stage rocket return.
- Utilized color-coded markers:
  - Green for successful landings and red for failed landings.

3. Calculating and Visualizing Distances (CCAFS LC-40)

- Computed distances from CCAFS LC-40 to:
  - The nearest: city, coastline, and highway.

Table 1.1: Launch Site Coordinates

Launch Site	Latitude	Longitude
CCAFS LC-40	28.5623	-80.5773
CCAFS SLC-40	28.5632	-80.5768
KSC LC-39A	28.5733	-80.6469
VAFB SLC-4E	34.6328	-120.6107

I. Methodology: Interactive Visual Analytics with Plotly Dash

Implementation:

1. Launch Site Selection Dropdown
   - Incorporated a dropdown menu enabling users to select from launch sites.

2. Success Count Pie Chart
   - Added a pie chart to display the total count of successful launches across all sites.

3. Payload Range Slider
   - Implemented a slider allowing users to filter payload mass within a range of 0 to 10,000 kg.

4. Payload vs. Launch Success Scatter Plot
   - Introduced a scatter plot illustrating the correlation between payload mass and launch success.

I. Methodology: Predictive Modelling

Objective: To develop and evaluate machine learning models capable of predicting the success of SpaceX's first-stage rocket landings.

Stages of Predictive Modeling:

1. Import Required Libraries
   - Pandas and Numpy for data manipulation.
   - Scikit-Learn for machine learning models and evaluation metrics.

2. Loading and Standardizing Cleaned Data
   - Imported the preprocessed dataset prepared during the data wrangling phase.

3. Splitting the Data
   - 80% for training and 20% for testing.

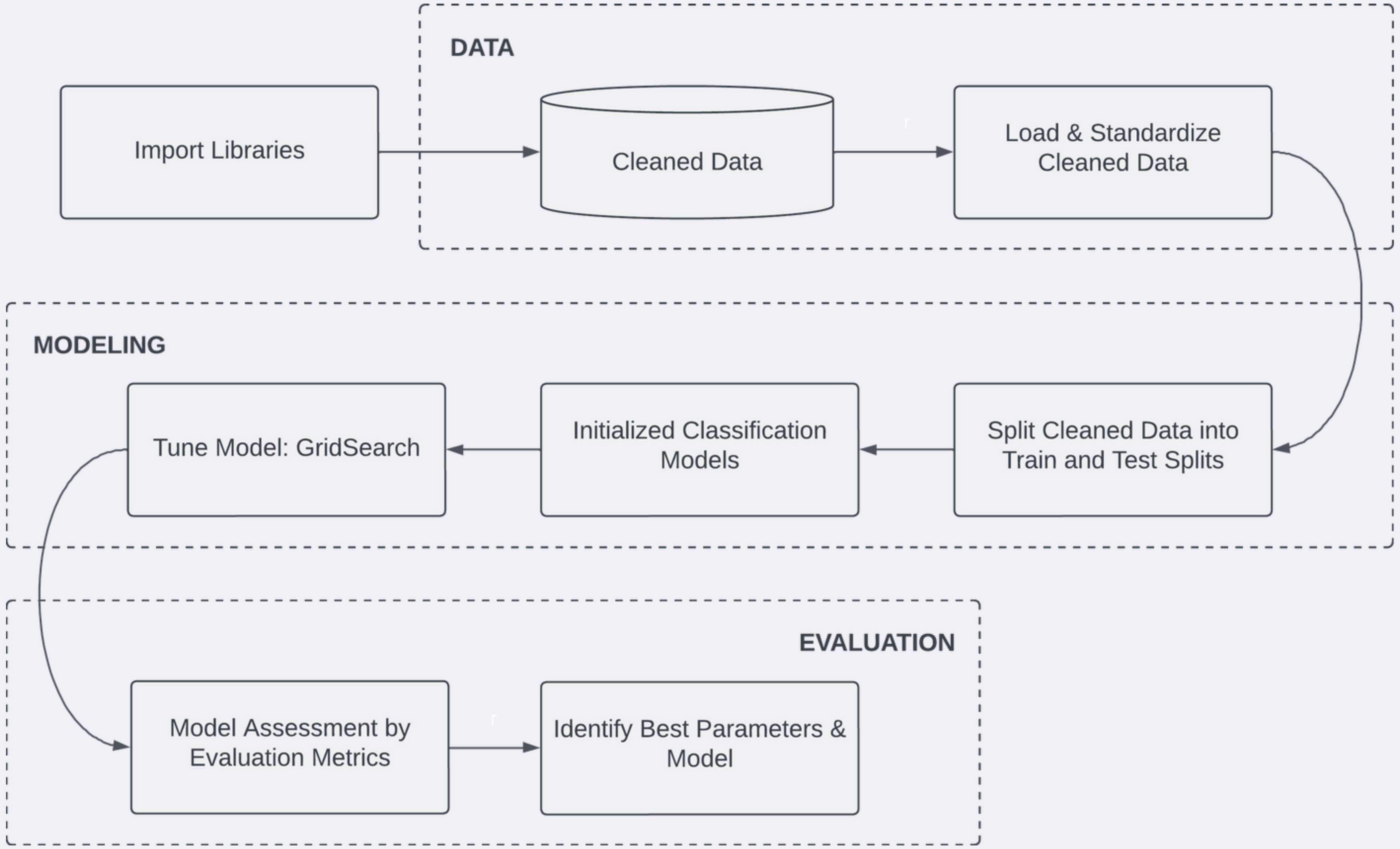


Figure 1.4: Modeling Pipeline

I. Methodology: Predictive Modelling

Objective: To develop and evaluate machine learning models capable of predicting the success of SpaceX's first-stage rocket landings.

Stages of Predictive Modeling:

5. Initialize Classification Models
   - Four distinct models selected:
     - Logistic Regression
     - Support Vector Machine
     - Decision Tree
     - K-Nearest Neighbors

6. Hyperparameter Tuning with GridSearch
   - Employed Grid Search to identify optimal hyperparameters for each model.

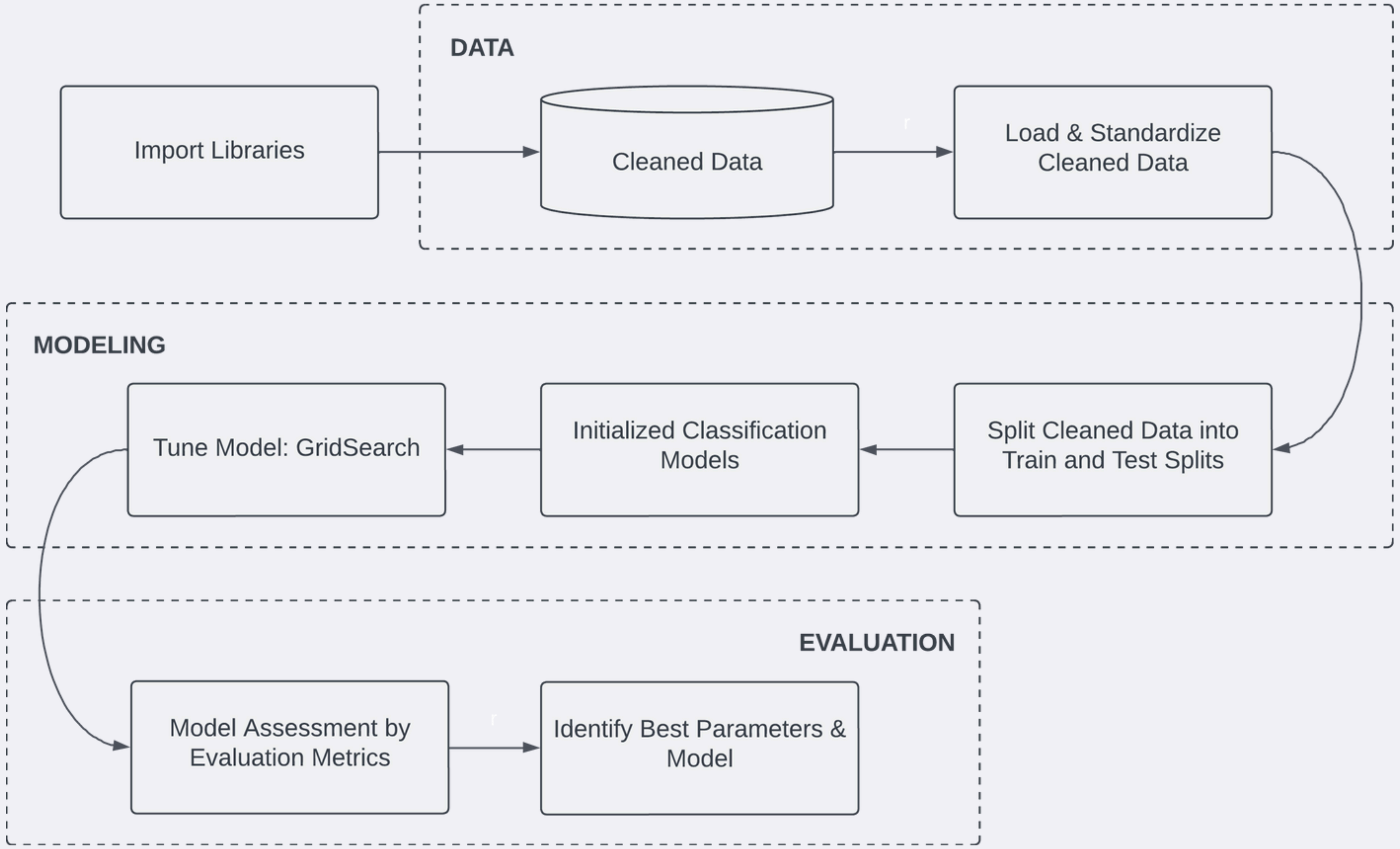


Figure 1.4: Modeling Pipeline

I. Methodology: Predictive Modelling

Objective: To develop and evaluate machine learning models capable of predicting the success of SpaceX's first-stage rocket landings.

Stages of Predictive Modeling:

7. Model Evaluation
   - Assessed models using:
     - Confusion Matrix: Evaluated true vs. predicted classifications.
     - F1-Score: Measured balance between precision and recall.
     - Jaccard Score: Compared similarity on predicted and actual labels.

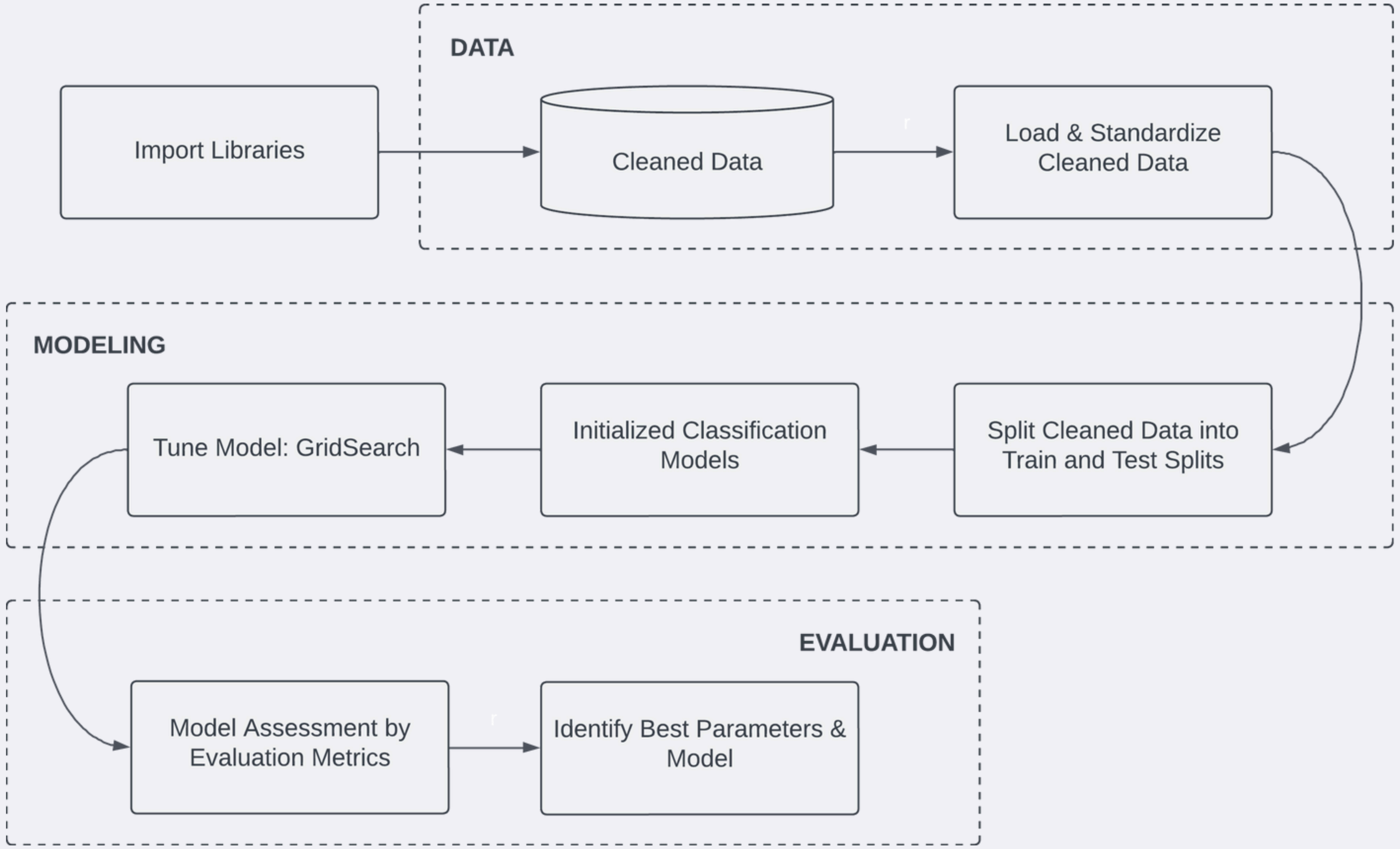


Figure 1.4: Modeling Pipeline

I. Methodology: Results

I. Exploratory Data Analysis Results

II. Interactive Analytics Screenshots

III. Predictive Analysis Results

Section 2

Insights drawn from EDA

II. EDA Insights: Launch Site and Flight Number

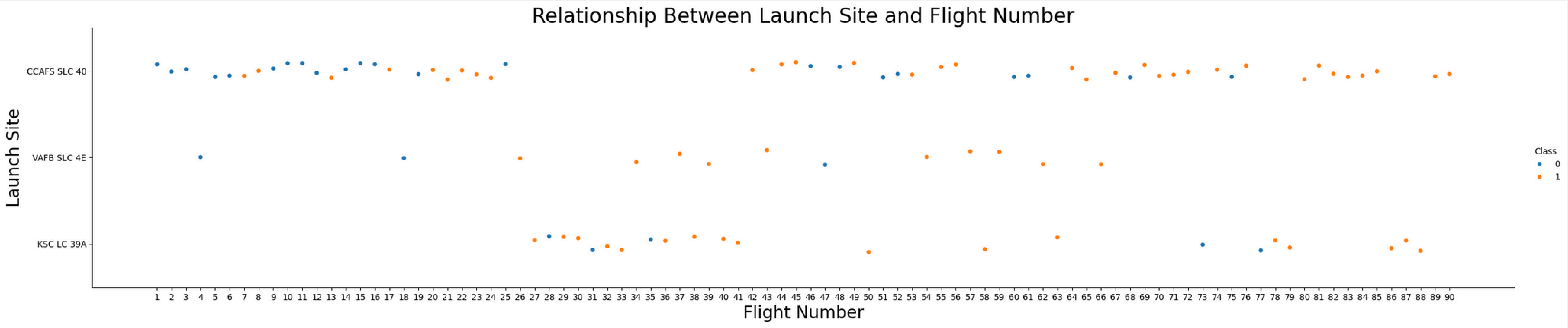


Figure 2.1: Relationship Between Launch Site and Flight Number

CCAFS SLC 40

- Most frequently used launch site with 55 trials.
- Achieved 60% success rate: 33 successes and 22 failures.
- Higher number of trials indicates operational preference for this site.

VAFB SLC 4E

- Least frequently used launch site with 13 trials.
- Achieved the highest success rate of 77%: 10 successes and 3 failures.
- Indicates efficiency despite fewer trials.

KSC LC 39A

- Moderately used launch site with 22 trials.
- Achieved a success rate of 77%: 17 successes and 5 failures.
- Suggests reliable performance for its operational frequency.

II. EDA Insights: Launch Site and Payload

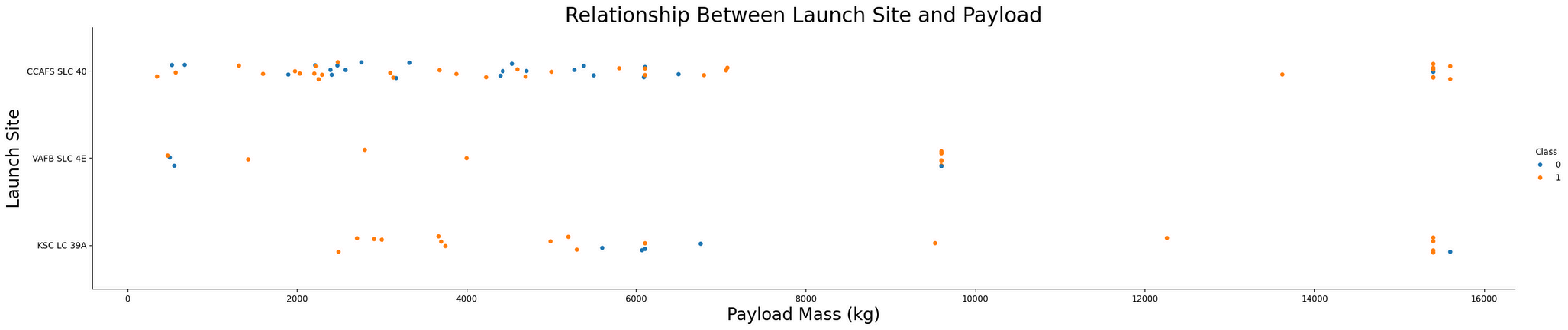


Figure 2.2: Relationship Between Launch Site and Payload

Payload Mass and Success Relationship:

- The scatter plot indicates no significant correlation between payload mass and the success of first-stage returns.
- Both successful and failed trials are distributed across various payload ranges.
  - It suggests that payload mass is not a decisive factor.

II. EDA Insights: Success Rate by Orbit Type

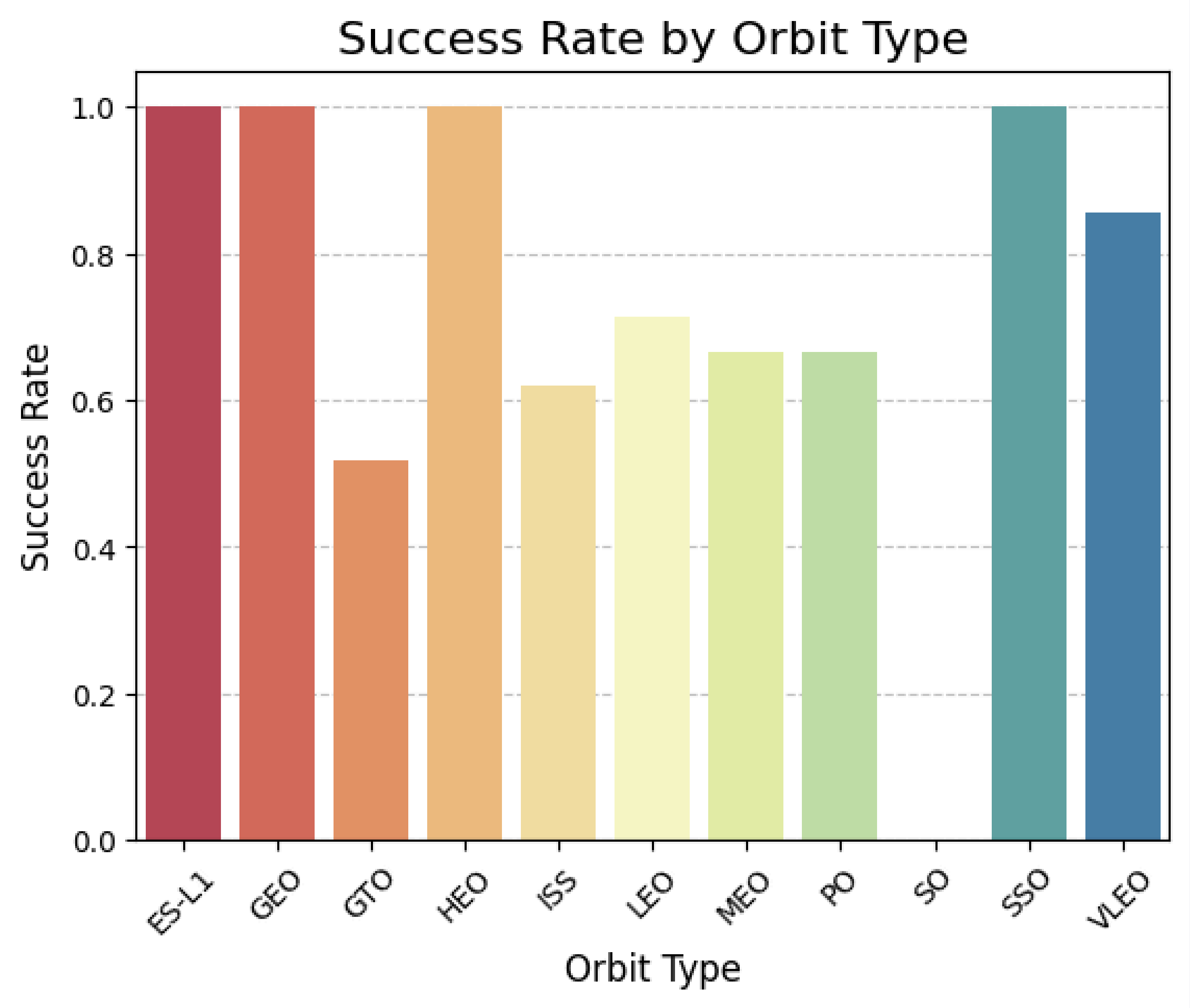


Figure 2.3: Success Rate by Orbit Type

High Success Orbits:

- Orbits such as ES-L1, GEO, HEO, SSO, and VLEO exhibit the highest success rates, reaching nearly or is 100%.

Low Success Orbits:

- GTO (Geostationary Transfer Orbit) shows the lowest success rate among the orbits analyzed.
- SO (Sun-synchronous Orbit) exhibits no success rate at all.

Moderate Success Orbits:

- Orbits such as ISS, LEO, MEO, and PO exhibit moderate success rates.
    - It suggests room for improvement in operational efficiency.

II. EDA Insights: Orbit Type by Flight Number

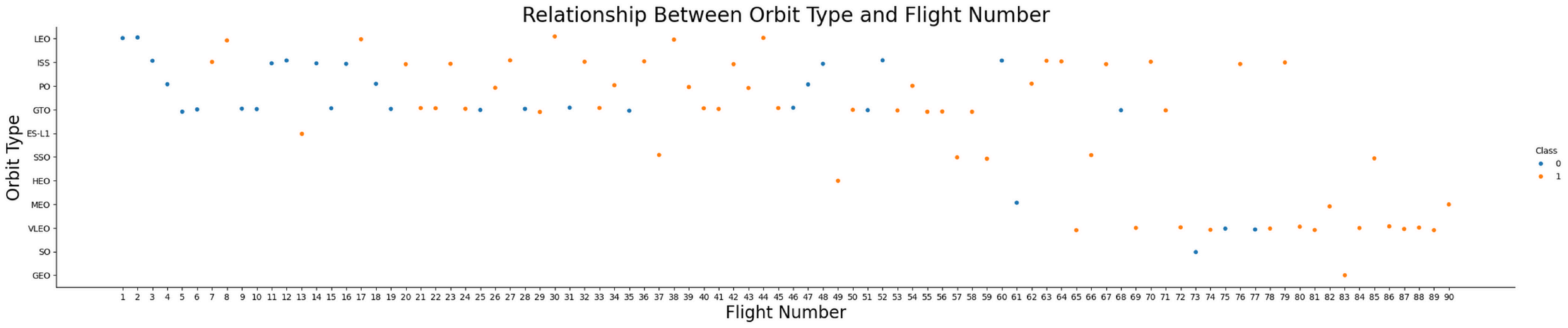


Figure 2.3: Relationship Between Orbit Type and Flight Number

General Observations:

- Orbit-specific dynamics heavily influence success rates, with experience playing a role in some orbits (e.g., LEO) but not in others (e.g., GTO).
- Continuous monitoring and orbit-specific adaptations are crucial for improving success rates.

II. EDA Insights: Orbit Type by Payload

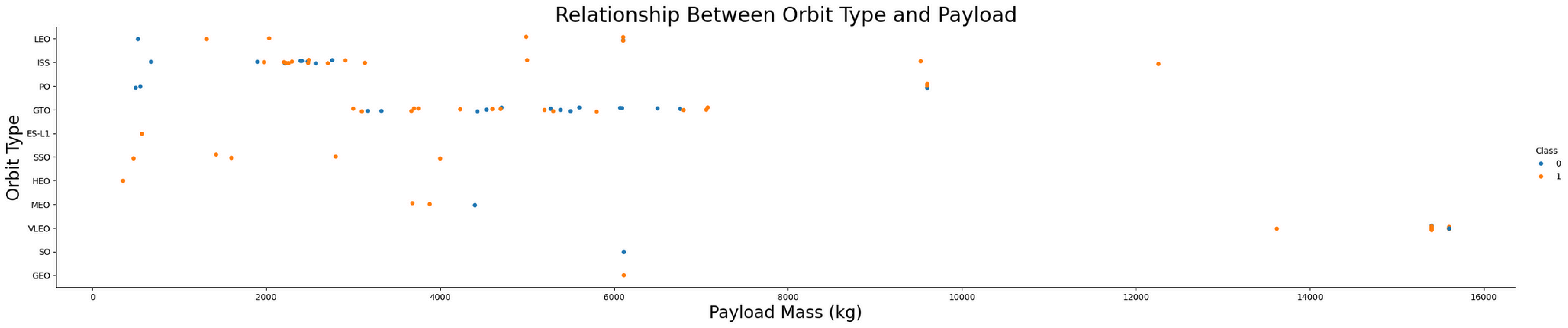


Figure 2.4: Relationship Between Orbit Type and Payload

Heavy Payloads and Orbit Outcomes:

- <u>GTO Orbit:</u> Heavy payloads negatively impact success rates in GTO, suggesting challenges in achieving successful first-stage returns for high-mass missions in this orbit.
- <u>LEO and Polar (ISS) Orbits:</u> Heavy payloads are positively correlated with success, indicating that these orbits can accommodate larger payloads effectively.

II. EDA Insights: Launch Success Yearly Trend

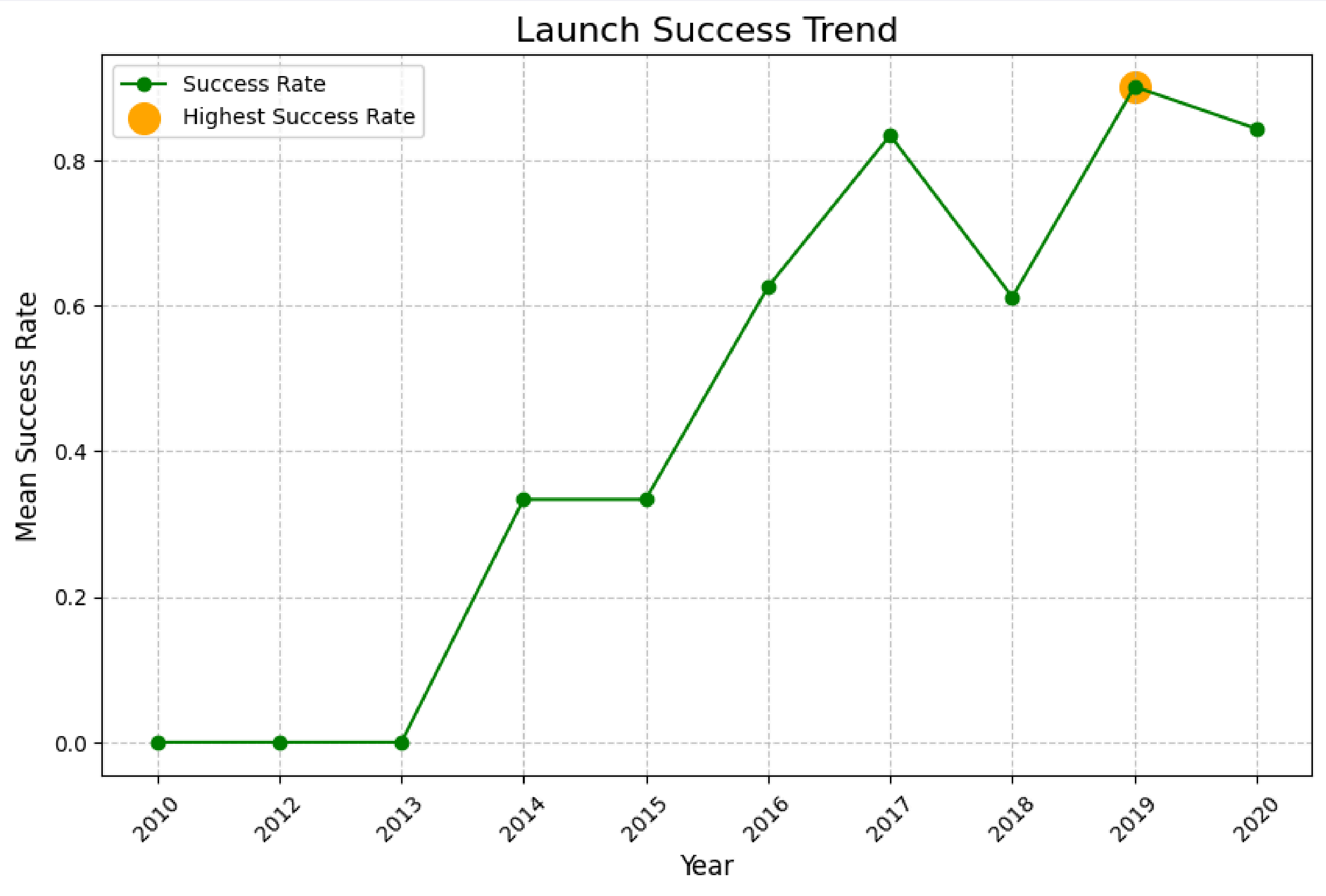


Figure 2.5: Launch Success Yearly Trend

Overall Trend:

- The success rate of launches has consistently increased since 2013, reflecting advancements in technology and operations.

Peak Performance:

- The highest success rate was observed in 2019, marking a culmination of consistent improvements.

Stability in Recent Years:

- Success rates remained high from 2018 to 2020, indicating sustained reliability in launch operations.

II. EDA Insights: All Launch Site Names

SQL Query:

- The SQL query retrieves all unique values from the Launch_Site column in the SPACEXTABLE.

Output:

- The query returned the following unique launch sites:
    - *CCAFS LC-40*
    - *VAFB SLC-4E*
    - *KSC LC-39A*
    - *CCAFS SLC-40*

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
 * sqlite:///my_data1.db
Done.
[25]:
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Figure 2.6: All Launch Site Names

II. EDA Insights: Launch Site Names Beginning with ‘CCA’

[26]:

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

[26]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)

Figure 2.7: All Launch Names Beginning with ‘CCA’

SQL Query:

- The query filters the SPACEXTABLE to return records where the Launch_Site column starts with "CCA."
- The LIKE 'CCA%' clause specifies that only launch site names beginning with "CCA" are included.
- The LIMIT 5 clause restricts the output to the first five matching records.

Output:

- The query returned launch site names beginning with the “CCA” string, along with its corresponding attributes.

II. EDA Insights: Total Payload Mass

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE
"Customer" = 'NASA (CRS)';
```

Query 2.1: Total Payload Mass Query

```
[43]:
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (
 * sqlite:///my_data1.db
Done.
[43]:
SUM(PAYLOAD_MASS__KG_)
                 45596
```

Figure 2.8: Total Payload Mass

SQL Query:

- The query calculates the total payload mass in kilograms for missions where the customer is 'NASA (CRS).'
- The SUM("PAYLOAD_MASS__KG_") function adds all values in the PAYLOAD_MASS__KG_ column that match the condition specified in the WHERE clause.

Output:

- The result, 45,596 kg., represents the total payload mass carried for NASA's CRS missions.

II. EDA Insights: 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';
```

Query 2.2: Average Payload Mass Query

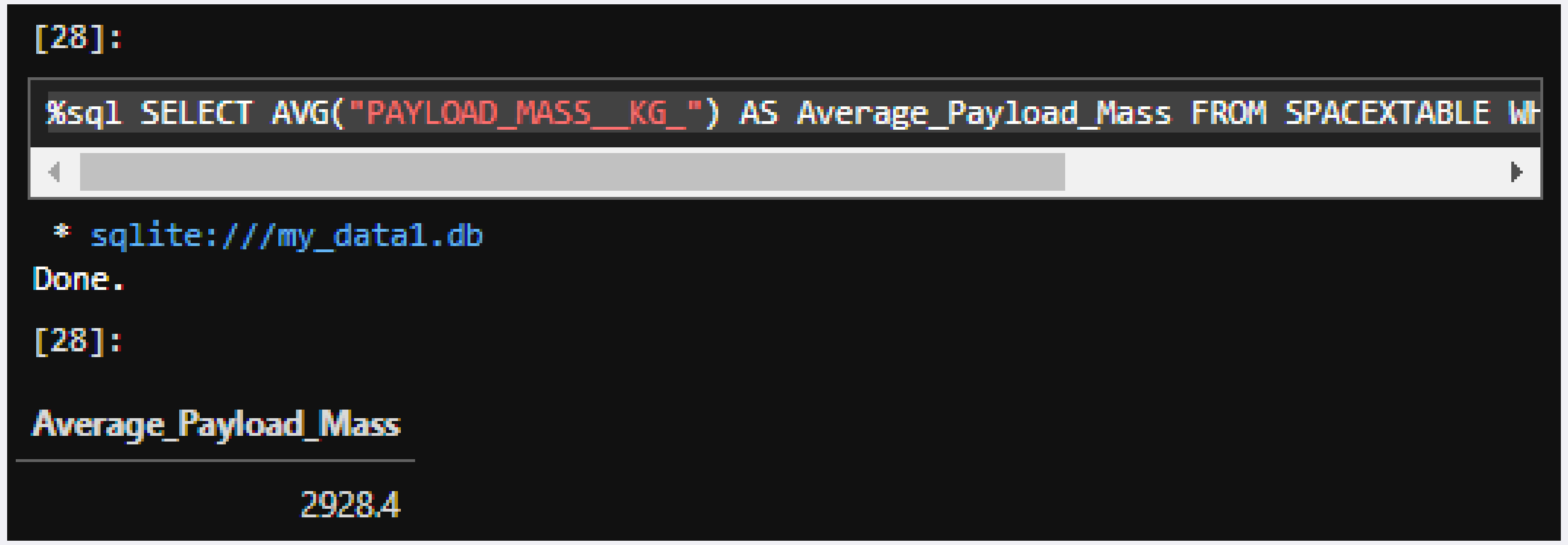


Figure 2.9: Average Payload Mass by F9 v1.1

SQL Query:

- The query calculates the average payload mass in kilograms for launches conducted with the Falcon 9 v1.1 rocket version.
- The AVG("PAYLOAD_MASS__KG_") function computes the mean of the payload masses from the PAYLOAD_MASS__KG_ column.

Output:

- The result, 2928.4 kg., represents the average payload mass carried by Falcon 9 v1.1 missions.

II. EDA Insights: First Successful Ground Landing Date

```
%sql SELECT MIN("Date") AS First_Successful_Landing FROM
SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

Query 2.3: First Successful Ground Landing Date Query

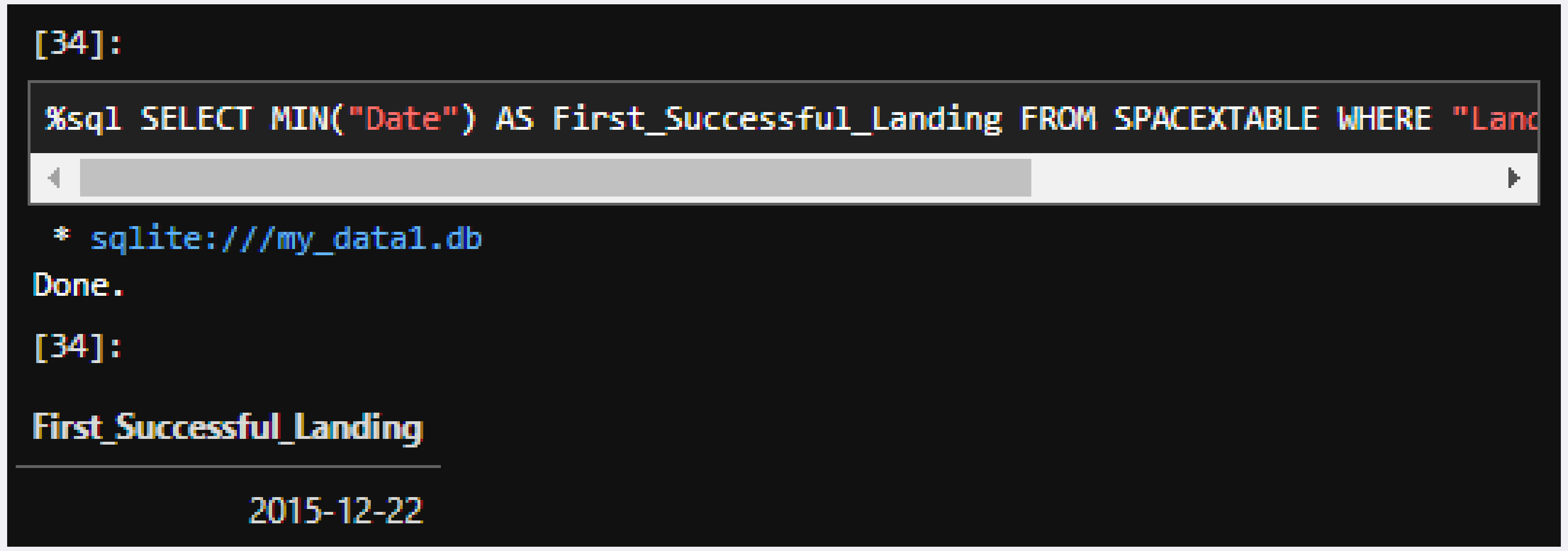


Figure 2.10: First Successful Ground Landing Date

SQL Query:

- The query identifies the earliest successful ground pad landing date by using the MIN("Date") function.
- The WHERE "Landing_Outcome" = 'Success (ground pad)' clause filters records to include only successful landings on ground pads.

Output:

- The result, 2015-12-22 or 22nd of December 2015, marks the first successful landing on a ground pad.

II. EDA Insights: Successful Drone Ship Landing (4000 kg. to 6000 kg.)

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE
"Landing_Outcome" = 'Success (drone ship)' AND
"PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000;
```

Query 2.4: Successful Drone Ship Landing Query

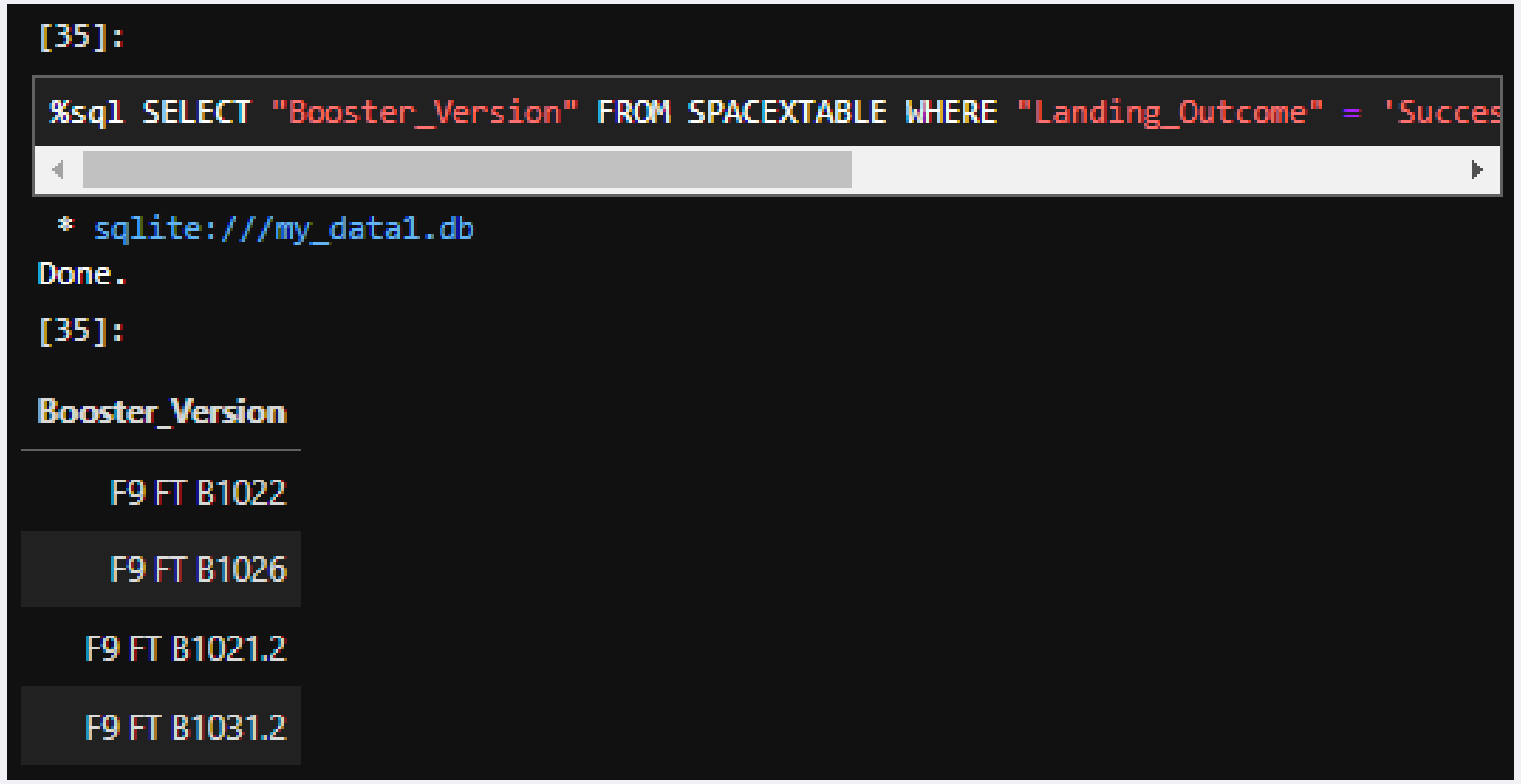


Figure 2.11: Successful Drone Ship Landing (4000 kg. to 6000 kg.)

SQL Query:

- The WHERE "Landing_Outcome" = 'Success (drone ship)' clause identifies landings classified as successful on a drone ship.
- The AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000 condition further narrows the selection to payloads within the specified range.

Output:

- The booster versions that meet these criteria are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

II. EDA Insights: Total Successful and Failure Mission Outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS numbers FROM
SPACEXTABLE GROUP BY "Mission_Outcome";
```

Query 2.5: Total Successful and Failure Outcomes Query

```
[37]:
%sql SELECT "Mission_Outcome", COUNT(*) AS numbers FROM SPACEXTABLE GROUP BY "Mi
 * sqlite:///my_data1.db
Done.
[37]:
```

Mission_Outcome	numbers
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Figure 2.12: Total Successful and Failure Mission Outcomes

SQL Query:

- The query groups the data by Mission_Outcome and counts the number of occurrences for each outcome using the COUNT(*) function.

Output:

- Mission Outcomes:
  - Failure (in flight): 1 occurrence.
  - Success: 98 occurrences.
  - Success (payload status unclear): 1 occurrence.

II. EDA Insights: Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE
"PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM
SPACEXTABLE);
```

Query 2.6: Boosters Carried Maximum Payload Query

Booster_Version
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

Figure 2.13: Boosters Carried Maximum Payload

SQL Query:

- The query retrieves the Booster_Version that carried the maximum payload mass by using a nested query:
  - The inner query SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE identifies the maximum payload mass.
  - The outer query matches this payload mass to the respective booster versions.

Output:

- The results, shown in Figure 2.13, return the booster versions carrying the maximum payload.

II. EDA Insights: 2015 Launch Records

```
%sql SELECT "Booster_Version", "Launch_Site" FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Failure (drone ship)' AND
strftime('%Y', "Date") = '2015';
```

Query 2.7: 2015 Launch Records Query

```
[41]:
%sql SELECT "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Out
 * sqlite:///my_data1.db
Done.
[41]:
```

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Figure 2.14: 2015 Launch Records

SQL Query:

- The query filters launch records from the year 2015 where the Landing_Outcome was a "Failure (drone ship)."
- The condition extracts the year from the Date column to focus on 2015 launches.

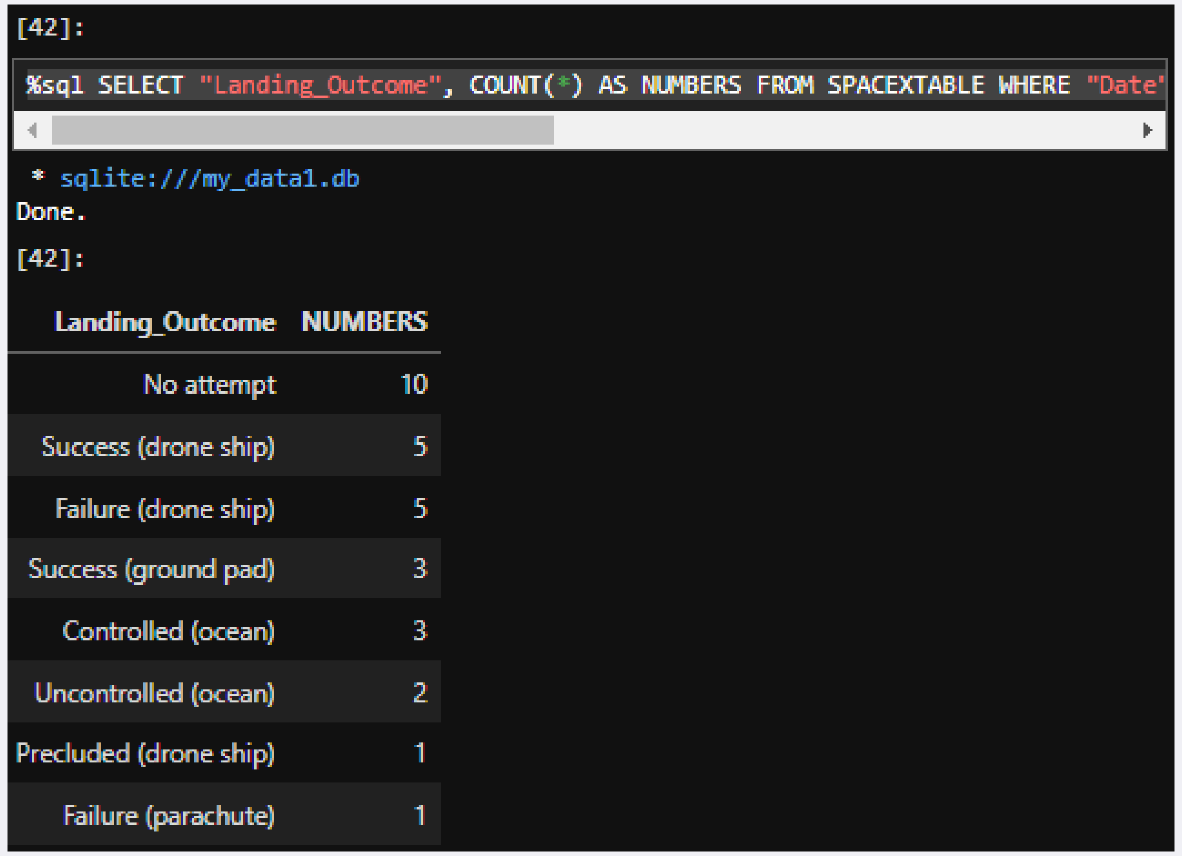
Output:

- Booster Versions:
  - F9 v1.1 B1012
  - F9 v1.1 B1015
- Launch Site:
  - Both launches occurred at CCAFS LC-40.

II. EDA Insights: Rank Landing Outcomes (2010-06-04 to 2017-03-20)

```
%sql SELECT "Landing_Outcome", COUNT(*) AS NUMBERS FROM
SPACEXTABLE WHERE "Date" > '2010-06-04' AND "Date" < '2017-03-
20' GROUP BY "Landing_Outcome" ORDER BY NUMBERS DESC;
```

Query 2.8: Rank Landing Outcomes Query



[42]:

```
%sql SELECT "Landing_Outcome", COUNT(*) AS NUMBERS FROM SPACEXTABLE WHERE "Date"
```

* sqlite:///my_data1.db
Done.

[42]:

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

Figure 2.15: Rank Landing Outcomes (2010-06-04 to 2017-03-20)

SQL Query:

- The query retrieves the count of each outcome for launches between June 4, 2010, and March 20, 2017.
- Records are grouped by Landing_Outcome, and the COUNT(*) function determines the number of occurrences for each outcome.

Output:

- The landing outcomes and their counts during the specified period is shown in Figure 2.15.

Section 3

Launch Sites Proximities Analysis

III. Folium Map: Launch Sites

Launch sites are distributed in two states: *California* and *Florida.*

Figure 3.1: United States Launch Sites Overview

Figure 3.2: VAFB SLC-4E

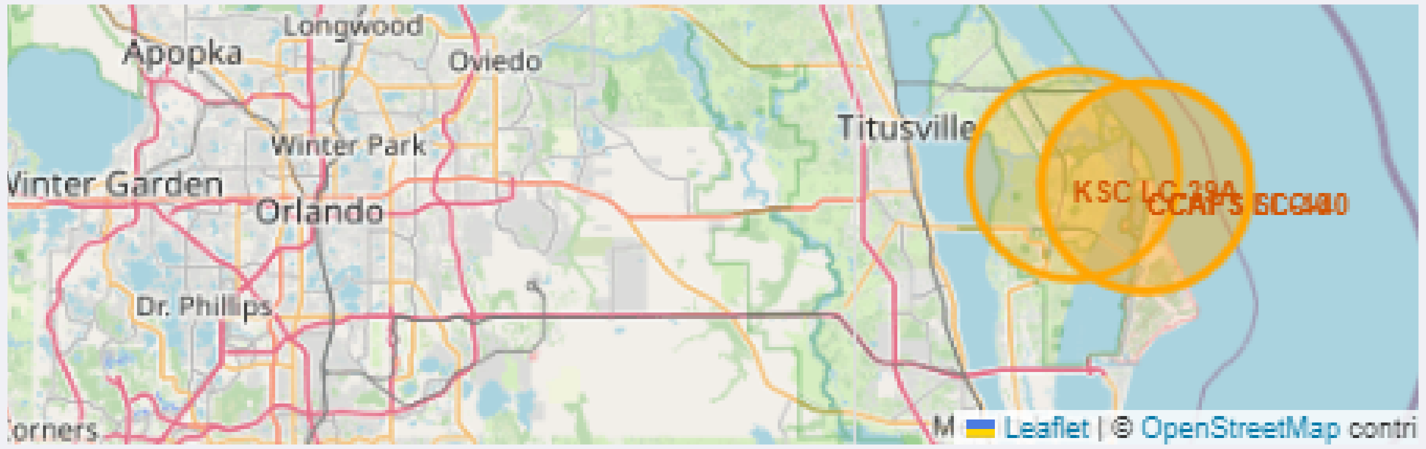


Figure 3.3: CCAFS LC-40 & CCAFS SLC-40

III. Folium Map: Sites Launch Rate

Launch Rate Markers:

- Color-labeled markers illustrate launch sites’ rate.
  - *Green* markers mean a successful return.
  - *Red* makers mean a failed return.

Sample Site Launch Rate:

- Figure 3.4. illustrates a sample for CCAFS SLC-40.
  - The site has 7 total launch attempts.
  - 3 were successful, whereas 4 failed.

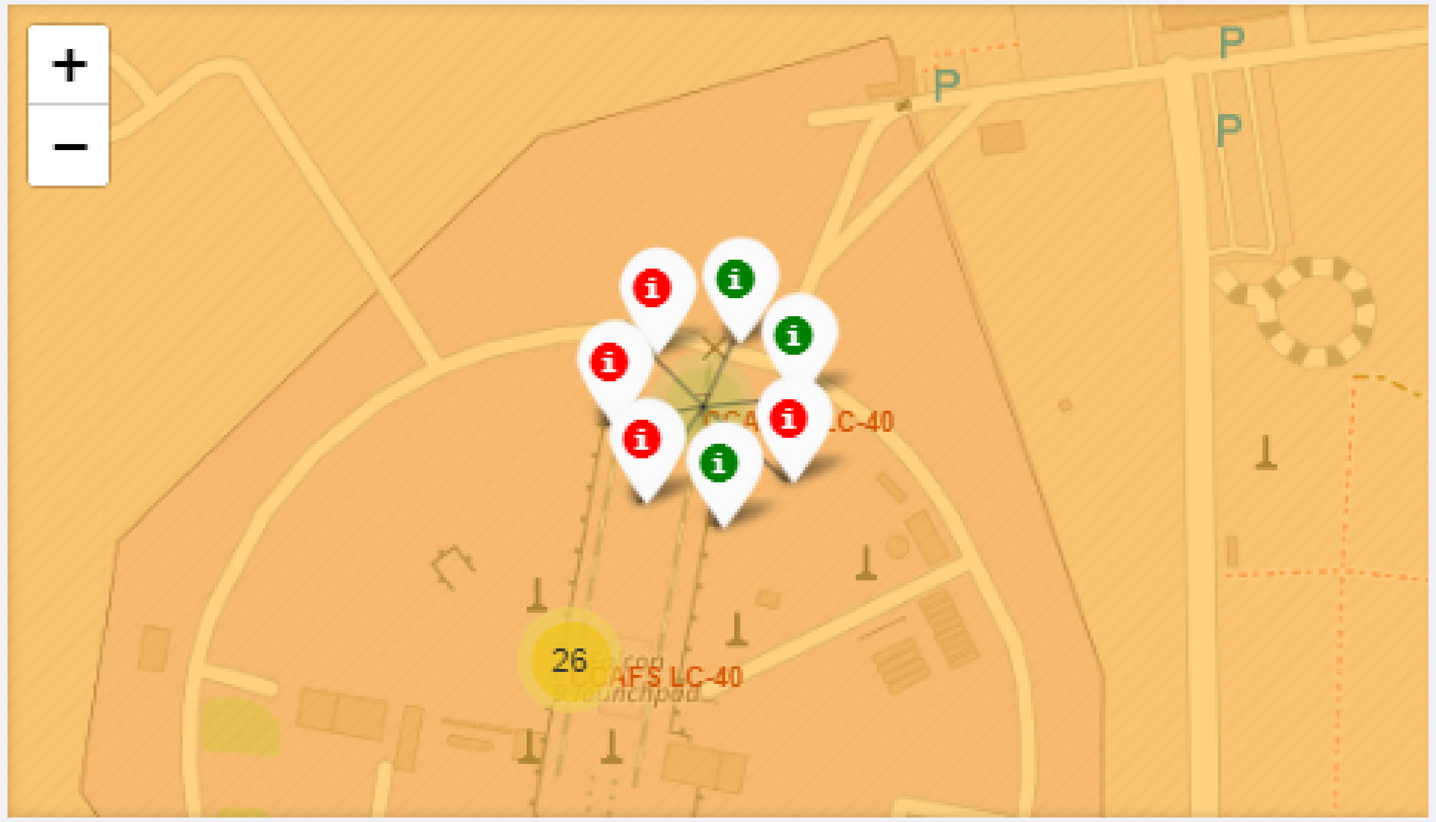


Figure 3.4: CCAFS SLC-40 Markers

III. Folium Map: Closest Proximities to CCAFS SLC-40

Table 3.1: Location Coordinates

Location	Latitude	Longitude
Melbourne City *Closest City*	28.1050	-80.6453
Samuel C Phillips Pkwy *Closest Highway*	28.5633	-80/5709
Titan III Railroad *Closest Railroad*	28.5721	-80.5853

Table 3.2: Location Distance

Location	Distance (in kilometers)
Closest City	**51.46 km.**
Closest Highway	**0.58 km.**
Closest Railrod	**1.27 km.**
Closest Coastline	**0.86 km**

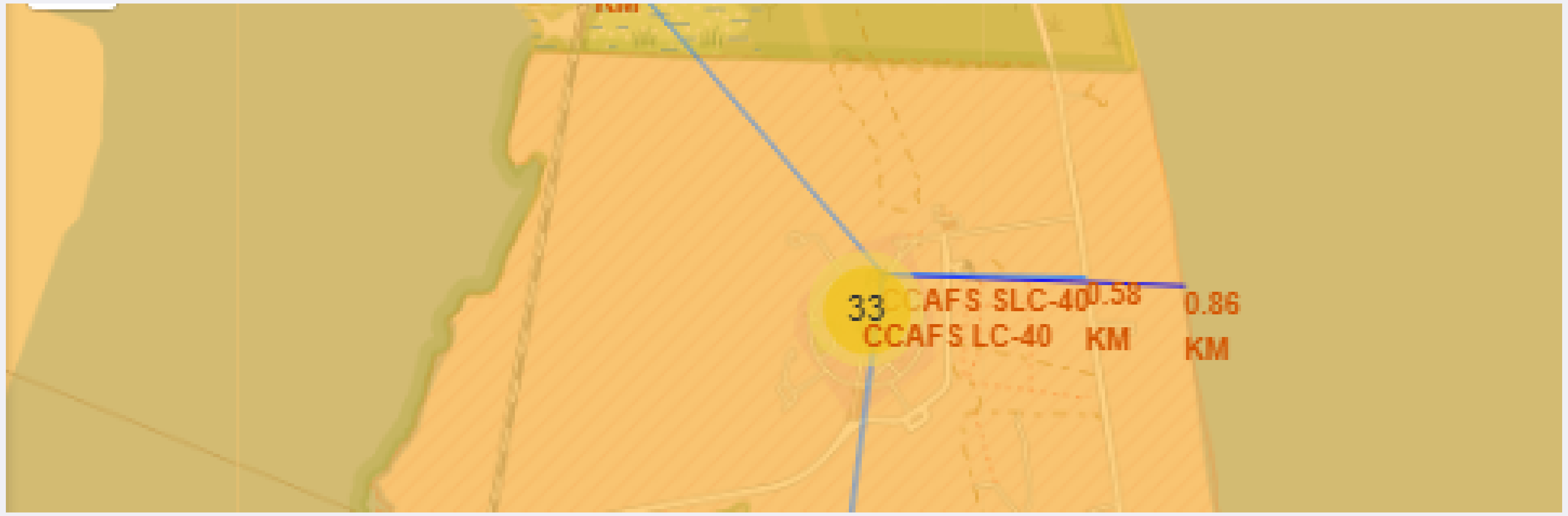


Figure 3.4: CCAFS SLC-40 Polylines

Key Observation:

- The strategic location of CCAFS LC-40 reflects careful planning to balance accessibility, safety, and operational efficiency.

Section 4

Build a Dashboard with Plotly Dash

IV. Plotly Dash: Launch Site Success Rate

Overview: The pie chart visualizes the proportion of successful first-stage returns for each launch site.

Key Observations:

- KSC LC-39A
    - The best-performing site with a success rate of 41.7%.
    - Indicates high reliability and operational efficiency.
- CCAFS SLC-40
    - The least successful site, contributing only 12.5% of total successes.
    - Suggests potential operational or geographical challenges

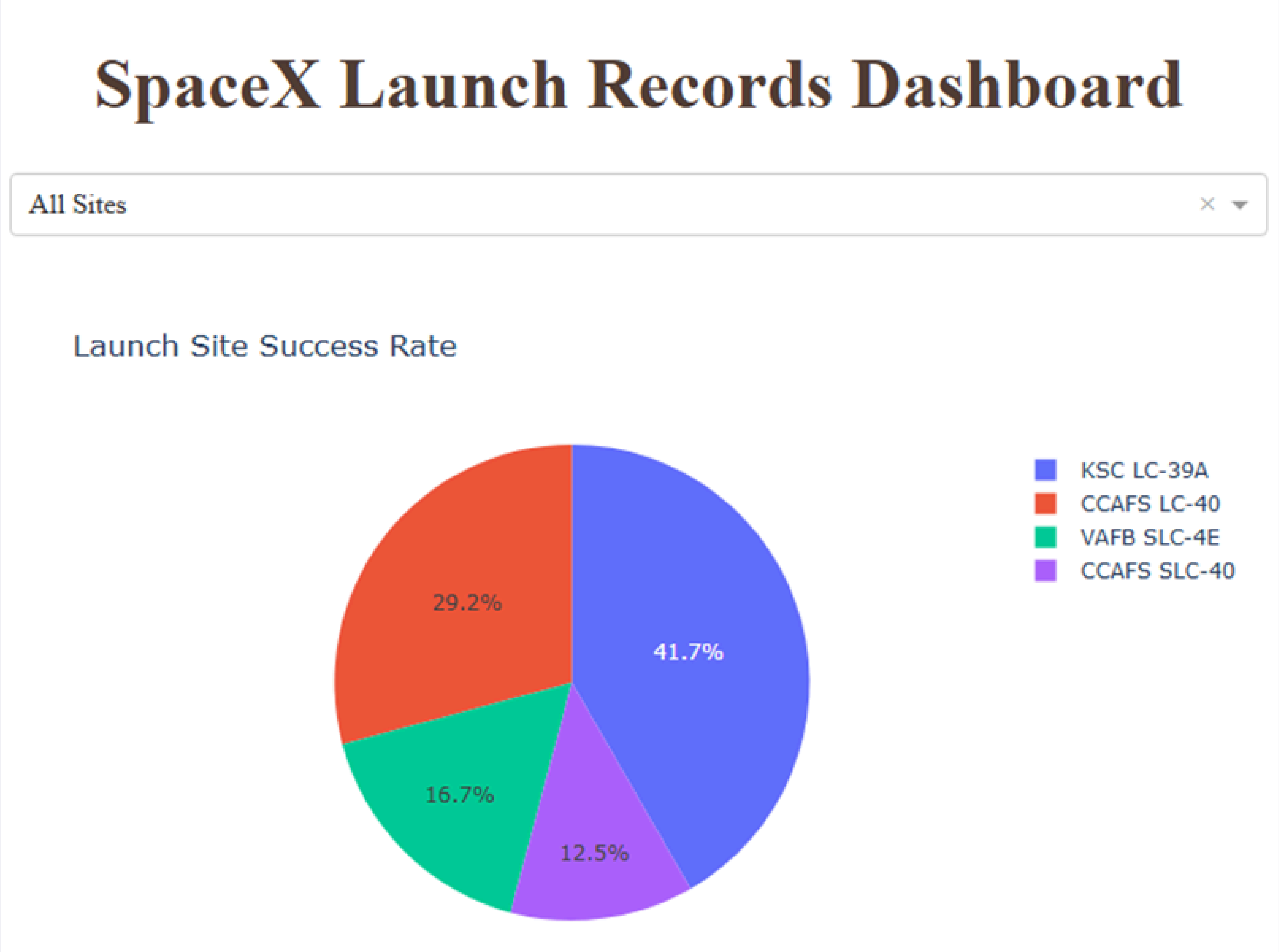


Figure 4.1: Launch Site Success Rate

IV. Plotly Dash: Best Launch Success Ratio

Overview: The pie chart represents the success and failure rates for launches at KSC LC-39A, the best-performing site.

Key Observations:

- KSC LC-39A shows the highest success ratio among all sites.
  - Improvements could further minimize the failure rate and maximize performance.

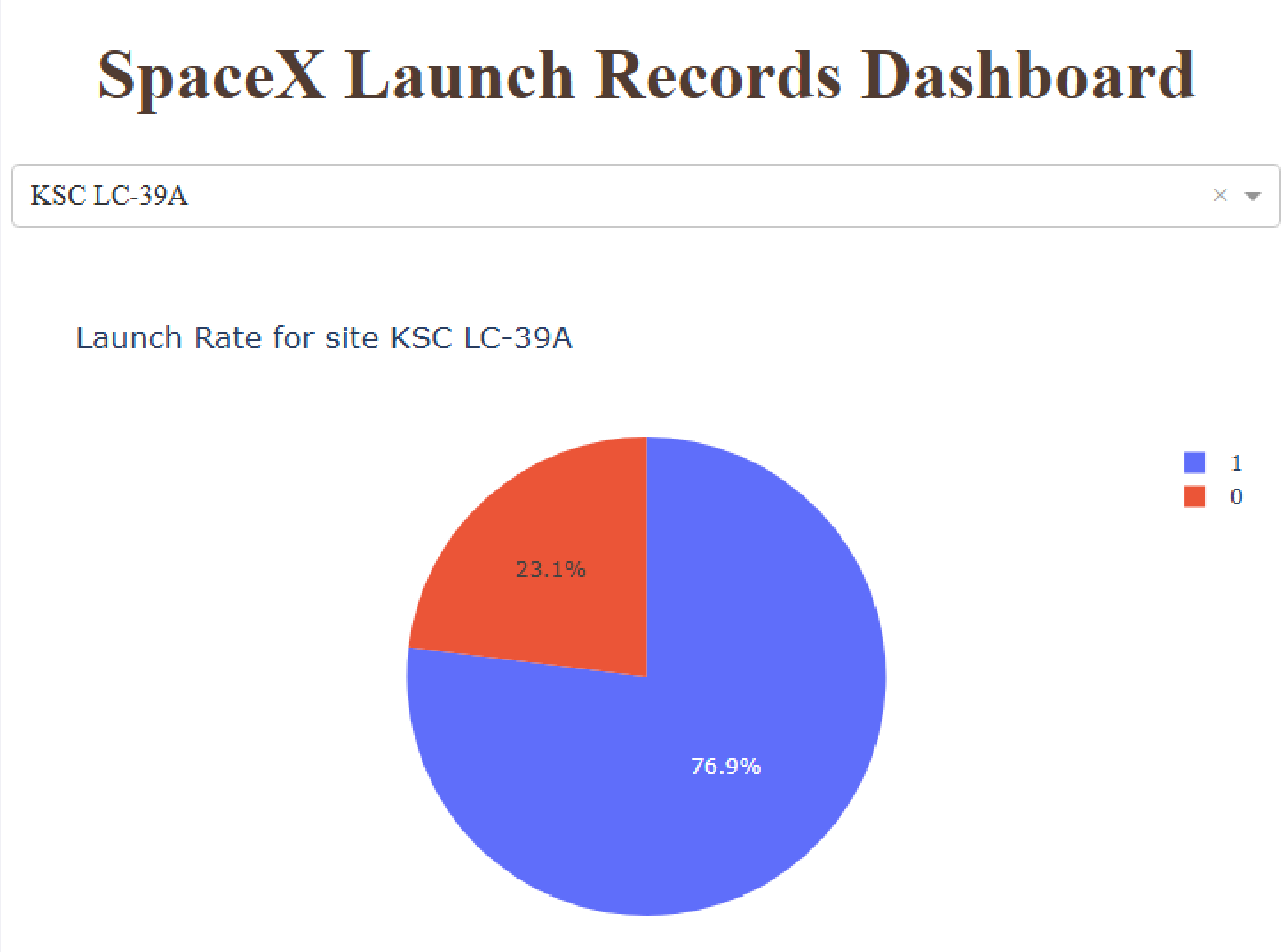


Figure 4.2: Best Launch Success Ratio

IV. Plotly Dash: Payload vs. Outcome

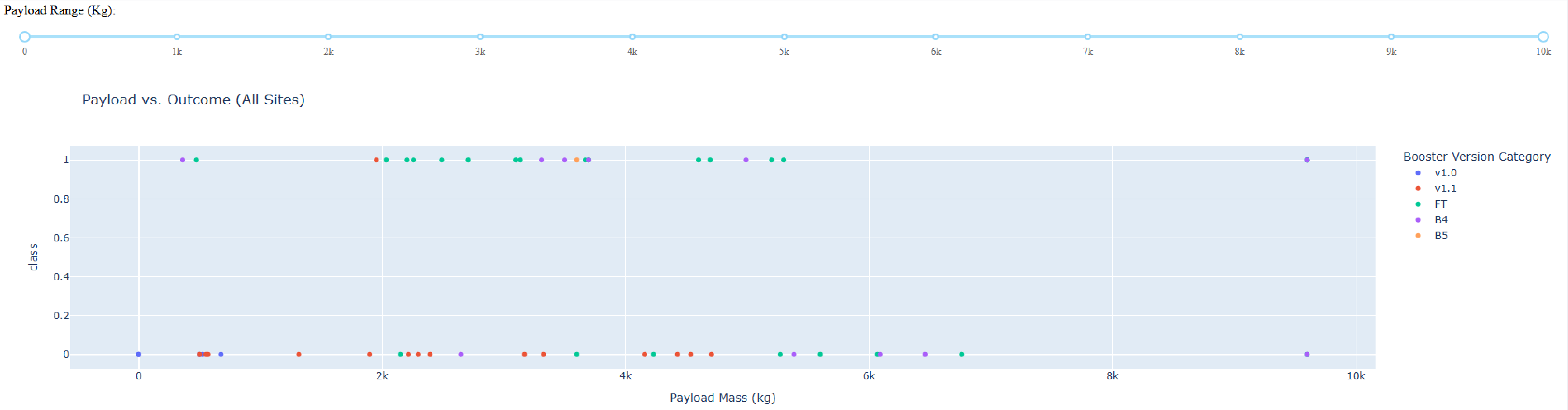


Figure 4.3: Payload vs. Outcome

Key Observation:

- Success probability decreases with higher payloads.
  - Indicates the need for booster-specific strategies to optimize performance for heavier missions.

Section 5

Predictive Analysis (Classification)

V. Predictive Analysis: Classification Test Accuracy

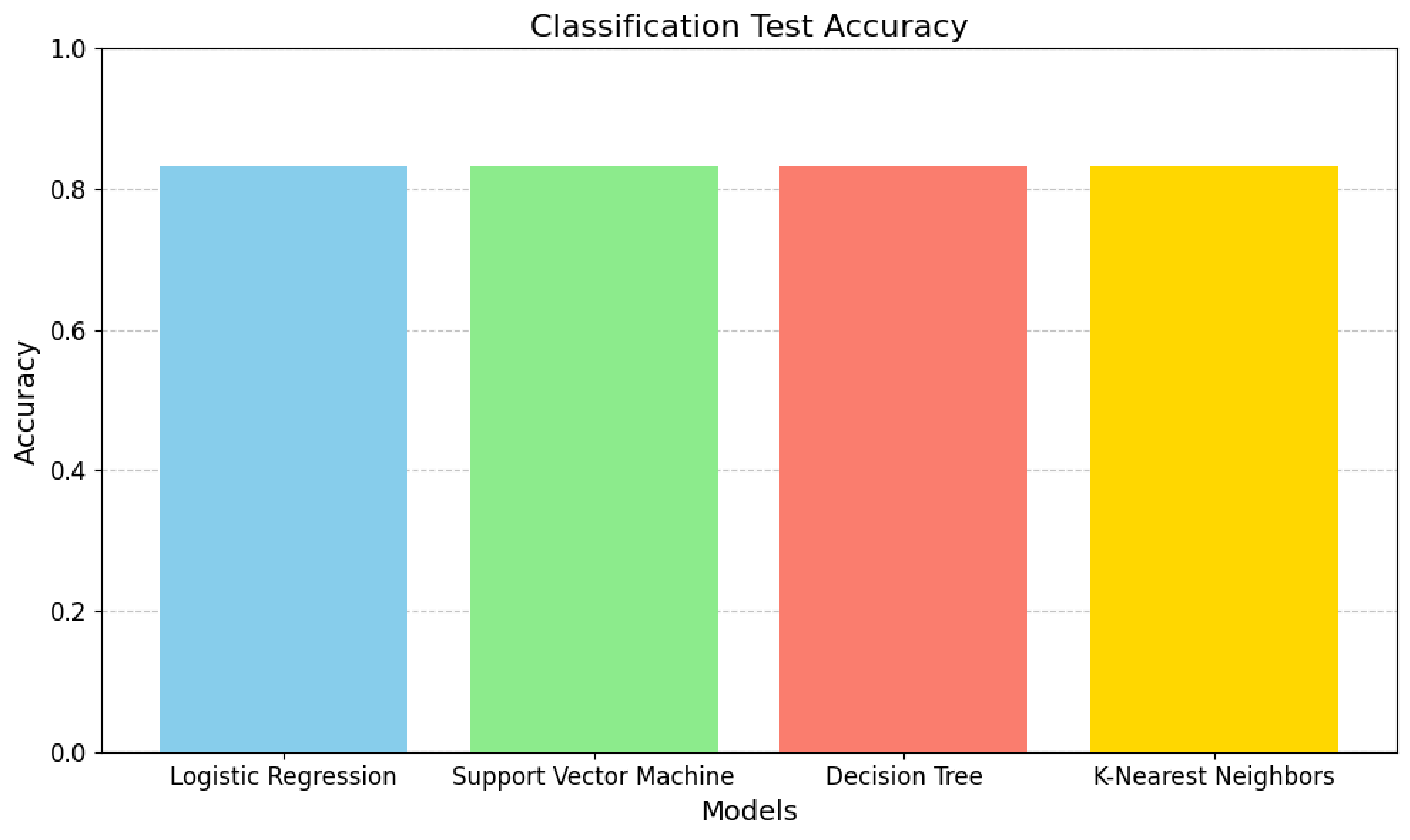


Figure 5.1: Classification Test Accuracy Across Models

Key Observation:

- All models showed identical testing accuracy.
  - Reflects consistent performance across different models.

V. Predictive Analysis: Confusion Matrix, Support Vector Machine

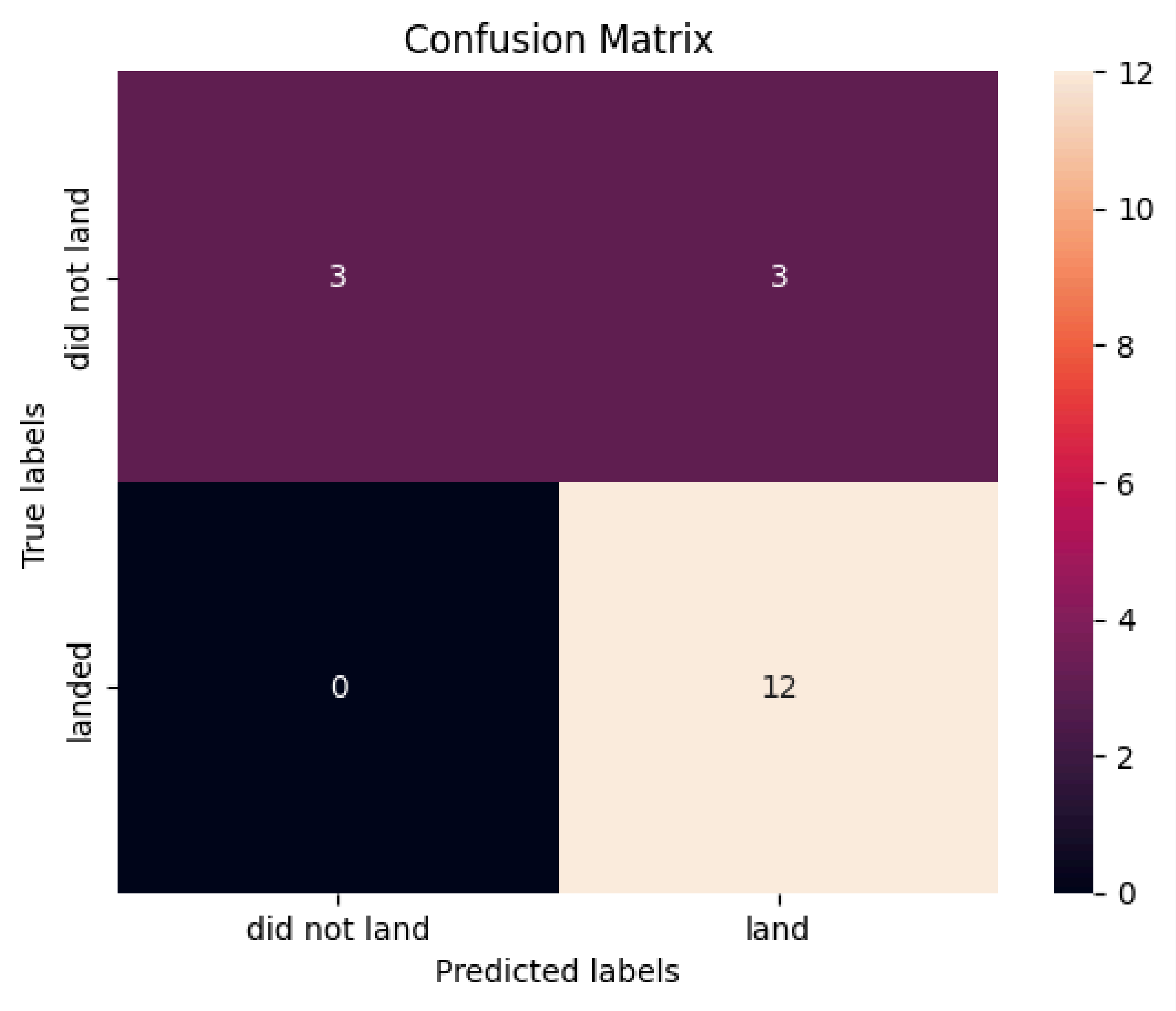


Figure 5.1: SVM Confusion Matrix

Key Observation:

- The SVM model demonstrated strong predictive accuracy for successful landings, but had some struggle identifying unsuccessful landings.
  - Adjusting the decision threshold or incorporating additional features may enhance performance.

V. Predictive Analysis: Conclusions

Overview: *Four* conclusions were derived in this analysis.

1. Significance of First-Stage Recovery

The successful recovery and reuse of Falcon 9's first stage have substantially reduced launch costs. It places SpaceX as a leader in cost-effective and efficient space missions.

2. Influential Factors on Landing Success

Multiple factors, including launch site proximity to infrastructure, payload mass, orbit type, and booster version. These evidently influence the success rate of first-stage landings.

V. Predictive Analysis: Conclusions

Overview: *Four* conclusions were derived in this analysis.

3. Strategic Launch Site Selection

paceX's selection of launch sites near highways, railways, and coastlines has optimized logistics and transportation. It contributes to efficiency and cost-reduction.

4. Model Performance Consistency

Predictive models, including Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors, showed consistent accuracy in forecasting landing outcomes.

Appendix

SpaceX. (n.d.). *SpaceX API*. Retrieved December 2024, from https://github.com/r-spacex/SpaceX-API

Wikipedia contributors. (n.d.). *List of Falcon 9 and Falcon Heavy launches. In Wikipedia, The Free Encyclopedia*. Retrieved December 2024, from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Thank you!