IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrea Galvão
August 12th, 2025

Outline

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

Executive Summary

Methodologies:

- Data Collection: API (/v4/launches/past) + Web scraping for launch/landing data
- Data Wrangling: Cleaned nulls, created binary outcome, normalized payload features
- EDA: SQL + visual analysis to uncover patterns by site, payload, and booster
- Stock Data: Extracted and visualized market trends for SpaceX context
- Interactive Dash: Built a dashboard (Plotly Dash) with site/payload filters and real-time plots
- Machine Learning: Trained and tuned Logistic Regression, SVM, Decision Tree, KNN using GridSearchCV

Key Results:

- Launch Success Rates: ~77% at KSC & VAFB, ~60% at CCAFS
- Payload Correlation: Higher payloads often linked with specific boosters and outcomes
- Dash Insights: Users can interactively explore success rates by site and payload range
- Best Model: Logistic Regression achieved highest accuracy after tuning

Introduction

Background & Context:

The commercial space industry is accelerating, driven by innovators like SpaceX with reusable rockets and cost-efficient launch models. New entrants like Space Y must navigate a competitive landscape and make strategic, data-driven decisions.

Project Goal:

Leverage public SpaceX launch data to extract insights and develop tools that support business intelligence in aerospace.

Key Questions:

1. What factors influence the cost of a SpaceX rocket launch?
2. Can we predict whether the Falcon 9 first stage will land successfully?
3. How can dashboards and visual analytics drive better decisions for new players like Space Y?

Section 1

Methodology

Methodology

Executive Summary

Data Collection:
Launch data was collected using the **SpaceX API** (/v4/launches/past) with Python, and **web scraping (BeautifulSoup)** was used to enrich data with additional details like landing outcomes.

Data Wrangling:
- Cleaned missing values and standardized data types
- Created new features (e.g., binary Outcome)
- Filtered key columns (Flight Number, Launch Site, Payload Mass)
- Aggregated data for trend analysis
-

Exploratory Data Analysis (EDA):
Used Matplotlib, Seaborn, and SQL to uncover trends in payload success, launch site performance, and orbit outcomes.

Interactive Visual Analytics:
- Folium map to visualize global launch sites and outcomes
- Plotly Dash dashboard for dynamic filtering and success rate analysis

Predictive Modeling:
Built classification models (Logistic Regression, SVM, Decision Tree, KNN) to predict booster landing success based on payload, orbit, and site.
Model Tuning & Evaluation:
- Used Grid Search + cross-validation
- Evaluated with accuracy, F1-score, confusion matrix
- Decision Tree and Logistic Regression performed best

Data Collection - API (Methodology Section)

Launch data for SpaceX Falcon 9 missions was collected directly from the SpaceX public REST API. The goal was to build a clean and structured dataset for analysis and modeling.

The process began by sending a GET request to the SpaceX API to retrieve all past launch data in JSON format. This data included rich and nested information such as rocket configurations, payload details, launch sites, and landing outcomes.

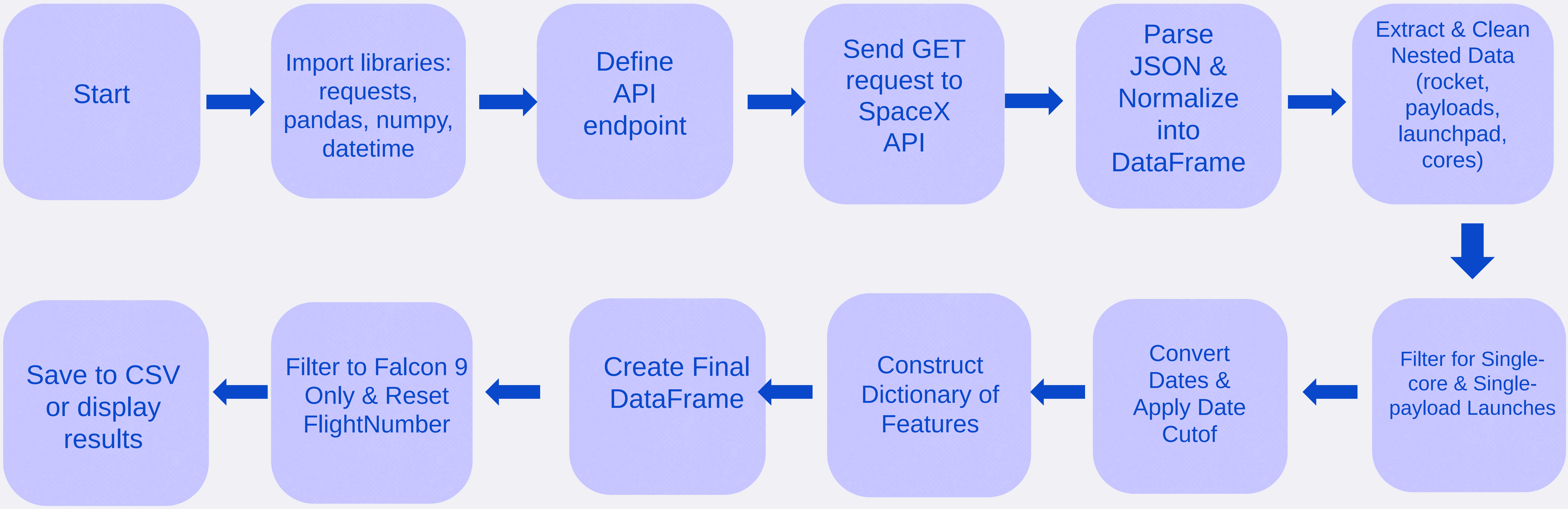
The JSON data was then normalized into a flat table using pandas, and filtered to include only relevant launches:

- Only Falcon 9 missions were kept.
- Launches with a single core and a single payload were selected to simplify analysis.
- Key fields such as Booster Version, Launch Site, Payload Mass, Orbit, Landing Type, and Mission Outcome were extracted from nested fields.

The result was a well-structured pandas DataFrame with one row per launch, which was saved for further data analysis and modeling tasks later in the project.

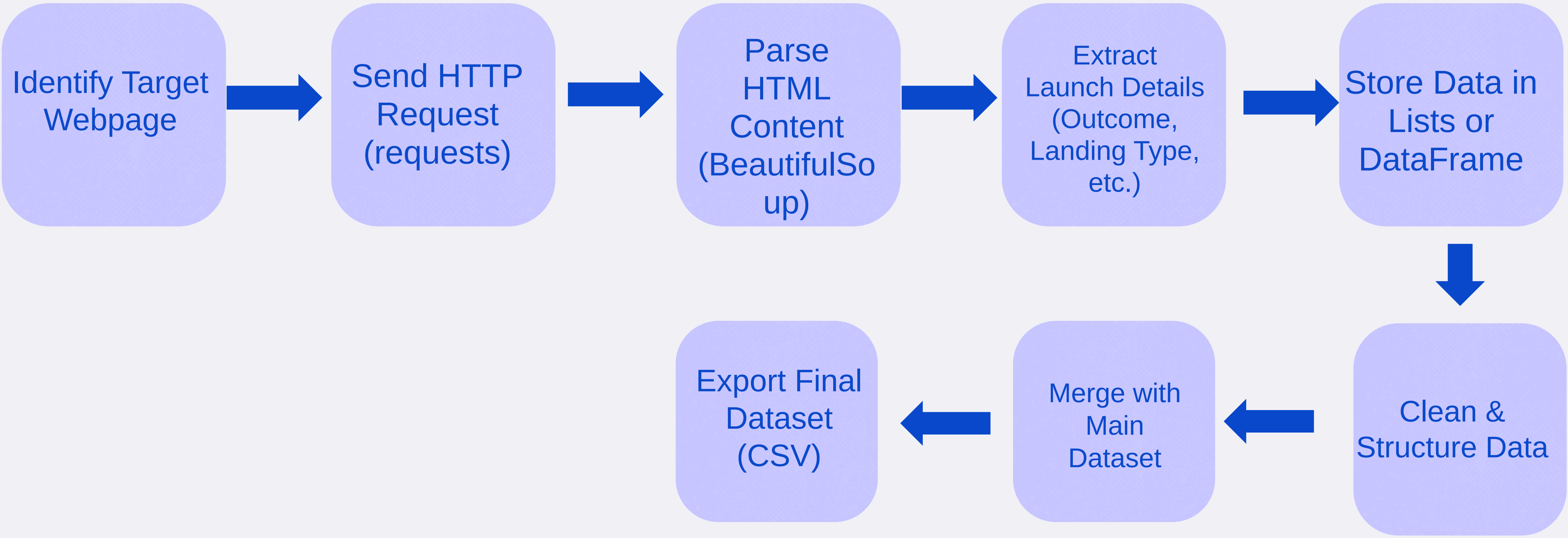
DATA COLLECTION API - GITHUB LINK

Data Collection - API (Methodology Section)



Data Collection - Web Scraping

SpaceX Falcon 9 launch data was collected by scraping tables from Wikipedia using BeautifulSoup. The HTML tables were parsed to extract key details such as launch outcomes (success/failure), landing types, launch sites, and payload information. This scraped dataset was later combined with data retrieved from a SpaceX API to create a richer, more complete dataset for analysis and modeling.



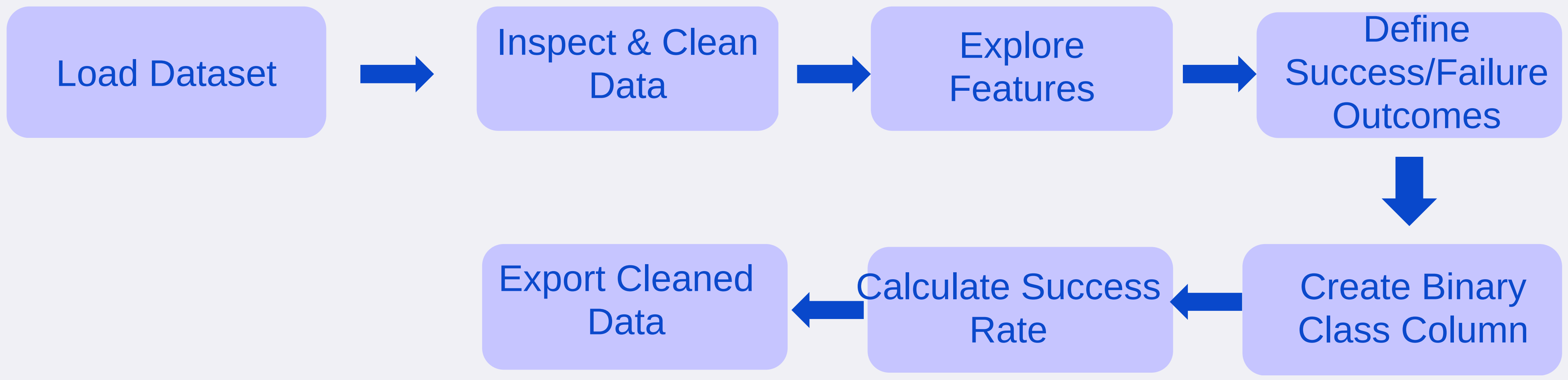
DATA COLLECTION - WEB SCRAPING GITHUB LINK

Data Wrangling

In the Data Wrangling stage, the collected SpaceX launch data was cleaned and transformed to prepare it for modeling. The dataset, built from both the API and web scraping, was loaded and inspected for missing values and inconsistencies. Key categorical and numerical features were explored, and mission outcomes were converted into binary labels to indicate success (1) or failure (0). This structured and labeled data was then saved for use in machine learning tasks.

Steps Performed:

- Load the combined dataset into a pandas DataFrame
- Identify and handle missing or inconsistent data
- Explore key features (e.g., Launch Site, Orbit, Payload Mass)
- Define successful vs. failed landing outcomes
- Create a binary Class column: 1 = success, 0 = failure
- Calculate landing success rate
- Export the cleaned dataset (dataset_part_2.csv) for further analysis



[DATA WRANGLING - GITHUB LINK](#)

EDA with Data Visualization

1. Scatter Plots

- Flight Number vs Launch Site: Visualizes how flight numbers are distributed across launch sites and identifies patterns of success or failure by location.
- Flight Number vs Payload Mass: Explores the relationship between flight number and payload mass to see if heavier payloads influence launch outcomes.
- Payload Mass vs Launch Site: Investigates how payload mass varies among different launch sites.
- Orbit Type vs Flight Number: Analyzes which orbit types were used across flights over time.
- Payload Mass vs Orbit Type: Evaluates if certain orbit types are associated with heavier or lighter payloads.

2. Bar Chart

Orbit Type vs Success Rate: Shows average success rates by orbit type, allowing comparison between different orbits.

3. Line Chart

Yearly Average Success Rate: Displays how the landing success rate evolves over the years, highlighting progress and trends.

Why these charts?

Scatter plots are great for visualizing continuous variable relationships and spotting trends or clusters.
Bar charts help compare categorical variables like orbit types against success rates.
Line charts show temporal trends clearly, ideal for tracking performance over time.

EDA - DATA VISUALIZATION GITHUB LINK

EDA with SQL

- Displayed the list of tables and explored the structure of each.
- Retrieved distinct launch sites and landing outcomes.
- Selected sample launch records from specific launch sites (e.g. those starting with "CCA").
- Calculated total payload mass for missions by NASA (CRS).
- Computed average payload mass for booster version “F9 v1.1”.
- Identified the earliest successful ground-pad landing.
- Filtered boosters that landed on drone ships with payloads between 4000–6000 kg.
- Counted occurrences of each mission outcome.
- Found booster versions with the maximum payload mass.
- Queried failed drone-ship landings in 2015.
- Ranked landing outcomes by frequency within a time range.

EDA WITH SQL - GITHUB LINK

Build an Interactive Map with Folium

Map: Created a base Folium map centered on launch sites. To provide an interactive view centered on launch areas

Circle markers: Highlighted each launch site’s location. To clearly mark and identify launch sites

Text labels: Added names of launch sites using DivIcon

Colored markers: Green for successful launches, red for failed ones. To visually differentiate successful vs. failed launches

MarkerCluster: Grouped nearby launch records to reduce clutter.To manage overlapping data points and reduce clutter

MousePosition plugin: Enabled coordinate display on mouse hover. To easily capture coordinates for nearby features (e.g., coast, roads)

Lines (PolyLine): Drew lines between launch sites and reference points (e.g., coast)

Distance markers: Labeled distances between launch sites and nearby features. To visualize and measure proximity between launch sites and key locations

Purpose: Enhance interactivity and visually explore spatial relationships, launch outcomes, and proximity to infrastructure.

FOLIUM MAP - GITHUB LINK

Build a Dashboard with Plotly Dash

Dashboard Plots & Interactions

- Launch Site Dropdown: Filter data by specific site or view all sites
- Success Pie Chart: Shows success rates overall or per selected site
- Payload Range Slider: Select payload mass range to filter launches
- Payload vs. Success Scatter Plot: Visualizes correlation between payload and launch outcome, colored by booster type

Purpose:

- Enable interactive exploration of launch performance
- Highlight success vs. failure distributions by site
- Reveal how payload size and booster version impact mission outcomes
- Provide user-friendly tools for focused data analysis

DASHBOARD WITH PLOTLY DASH - GITHUB LINK

Predictive Analysis (Classification)

Model Development Process:

- **Data Preprocessing**

Loaded and cleaned the dataset.
Encoded categorical variables.
Split data into training and testing sets.

- **Model Training**

Trained multiple classification models:
Logistic Regression
K-Nearest Neighbors (KNN)
Decision Tree
Support Vector Machine (SVM)

- **Model Evaluation**

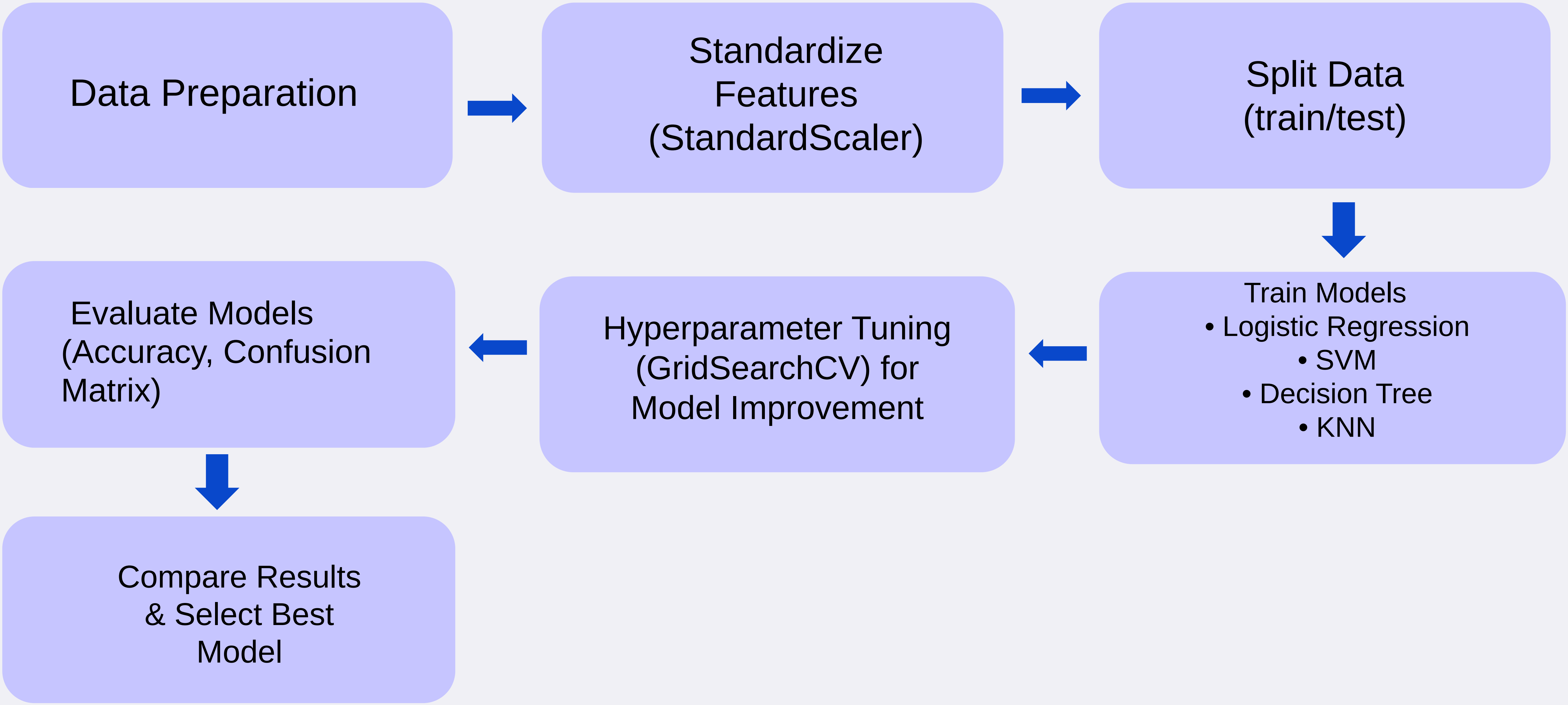
Assessed models using accuracy scores.
Generated confusion matrices for each model.

- **Model Improvement**

Applied hyperparameter tuning using GridSearchCV.
Optimized models for better performance.

- **Best Model Selection**

Compared performance metrics.
Selected the model with the highest accuracy.



PREDICTIVE ANALYSIS - GITHUB LINK

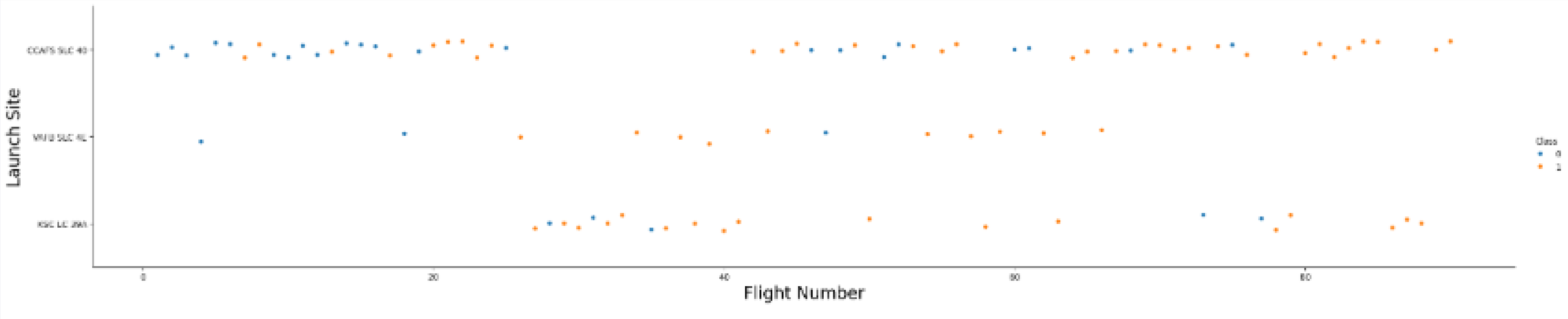
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



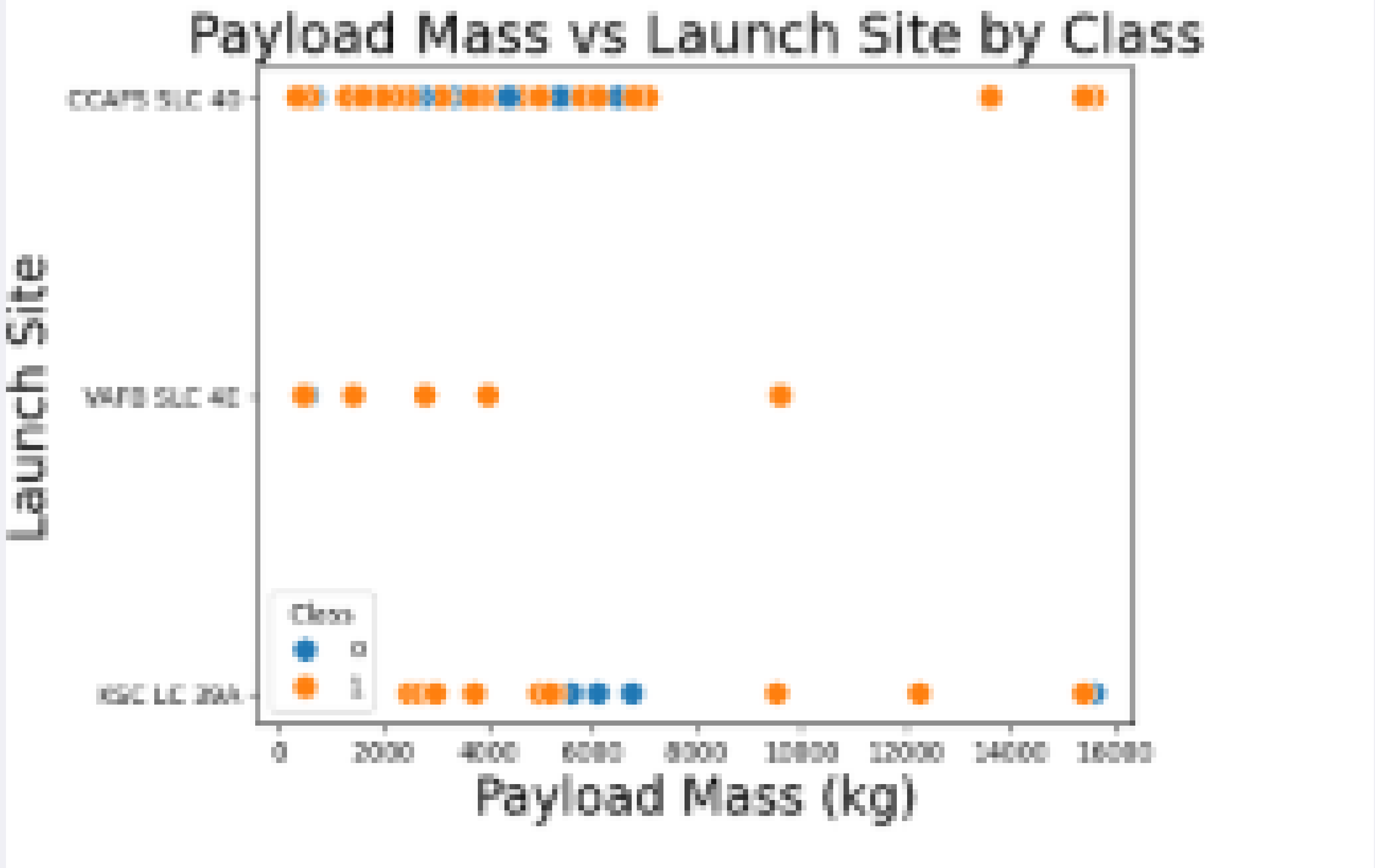
What it shows:
This scatter plot visualizes each SpaceX flight by its flight number (x-axis) and launch site (y-axis), with points color-coded by launch success (Class).
Key insights:
The spread of flight numbers along different launch sites (CCAFS SLC 40, VAFB SLC 4E, KSC LC 39A) shows the timeline and frequency of launches from each location.
The color differentiation between successful (Class 1) and unsuccessful (Class 0) launches quickly reveals success trends per site over time.
We observe that some launch sites have a dense cluster of flights indicating more frequent launch activity.
The plot helps identify if certain launch sites might correlate with higher or lower success rates visually.
Why it matters:
Understanding the relationship between flight number (launch order/time) and launch sites helps assess how location impacts mission success and provides valuable insights for launch planning and optimization.

Payload vs. Launch Site

What it shows:
This scatter plot illustrates the relationship between the payload mass (x-axis) and the launch site (y-axis), with points color-coded by launch outcome (Class) to indicate success or failure.

Key insights:
Payload masses vary significantly across different launch sites (CCAFS SLC 40, VAFB SLC 4E, KSC LC 39A), showing the diversity of mission payload sizes at each location.
Most successful launches (orange points) are clustered within certain payload mass ranges, reflecting typical payload capacities for successful missions.
Some launch sites display wider payload ranges, suggesting flexibility in launch capabilities, while others are more concentrated.
The color coding visually emphasizes the correlation between payload size, site, and mission success.

Why it matters:
Understanding how payload mass correlates with launch site and success rate helps optimize planning and decision-making in mission design and site utilization.



Success Rate vs. Orbit Type

What it shows:

This bar chart displays the average success rate for each orbit type, calculated as the mean of the launch success indicator (Class). The x-axis represents different orbit categories, and the y-axis shows the corresponding success rate.

Key insights:

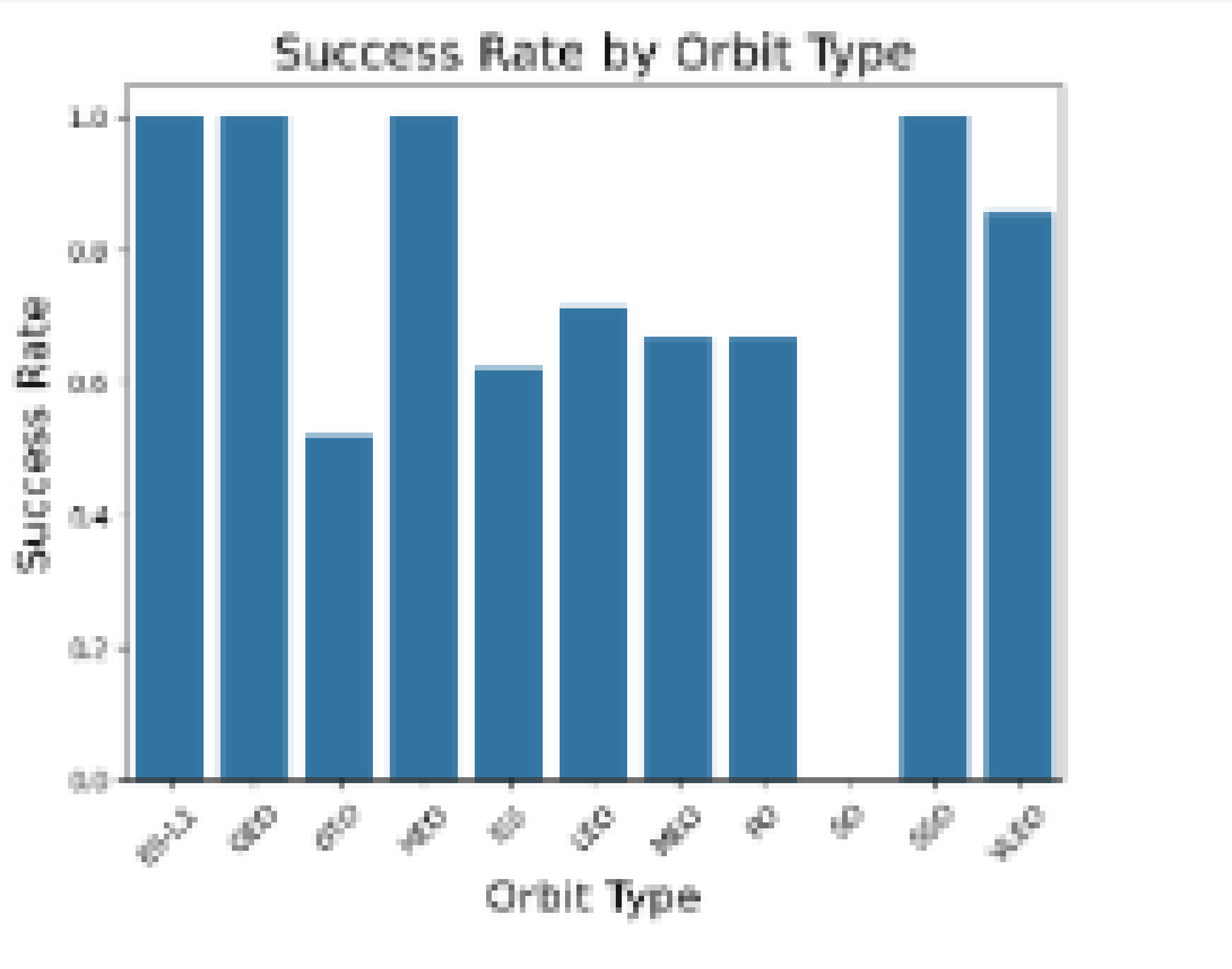
Certain orbit types, such as GTO and LEO, have notably high success rates near or at 100%, indicating reliable launch outcomes for these missions.

Other orbit types like ISS and PO show comparatively lower success rates, reflecting higher risk or more complex mission profiles.

The variations in success rates across orbits highlight different technical challenges and mission complexities depending on the orbit targeted.

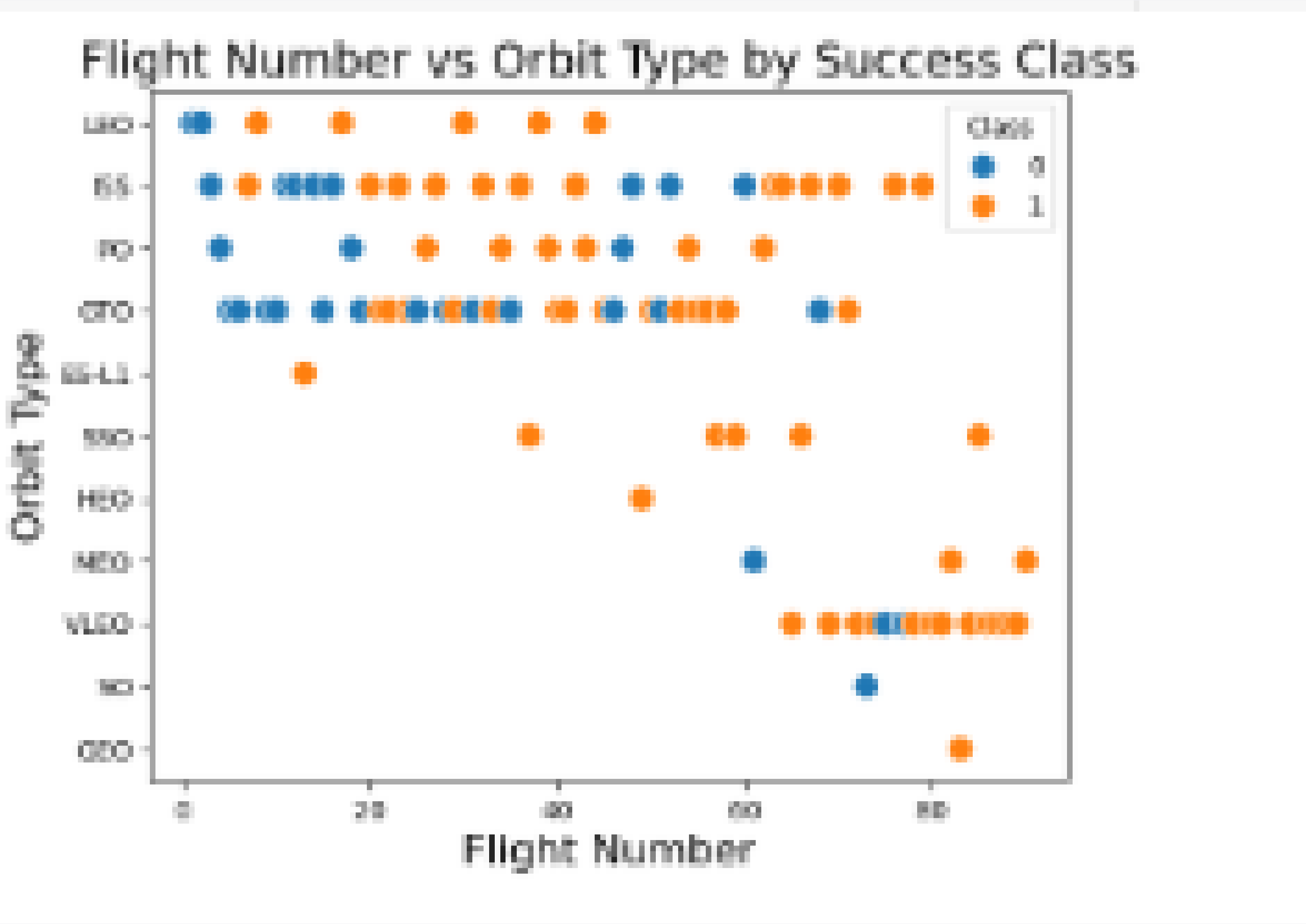
Why it matters:

Understanding success rates by orbit type provides essential insights for risk assessment, mission planning, and prioritizing improvements in launch strategies for different orbital destinations.



Flight Number vs. Orbit Type

What it shows:
This plot charts each flight's success (blue for success, orange for failure) against its chronological flight number and the type of orbit achieved.
Key insights:
Early flights (low numbers) had a higher proportion of failures (orange dots). As the flight number increases, there is a clear trend towards consistent success (more blue dots). The plot shows improved reliability over time and a successful expansion into various orbit types, like GTO and LEO, with high success rates.
Why it matters:
It demonstrates the historical improvement in launch reliability and the system's growing capability to achieve different orbits dependably.



Payload vs. Orbit Type

What it shows:

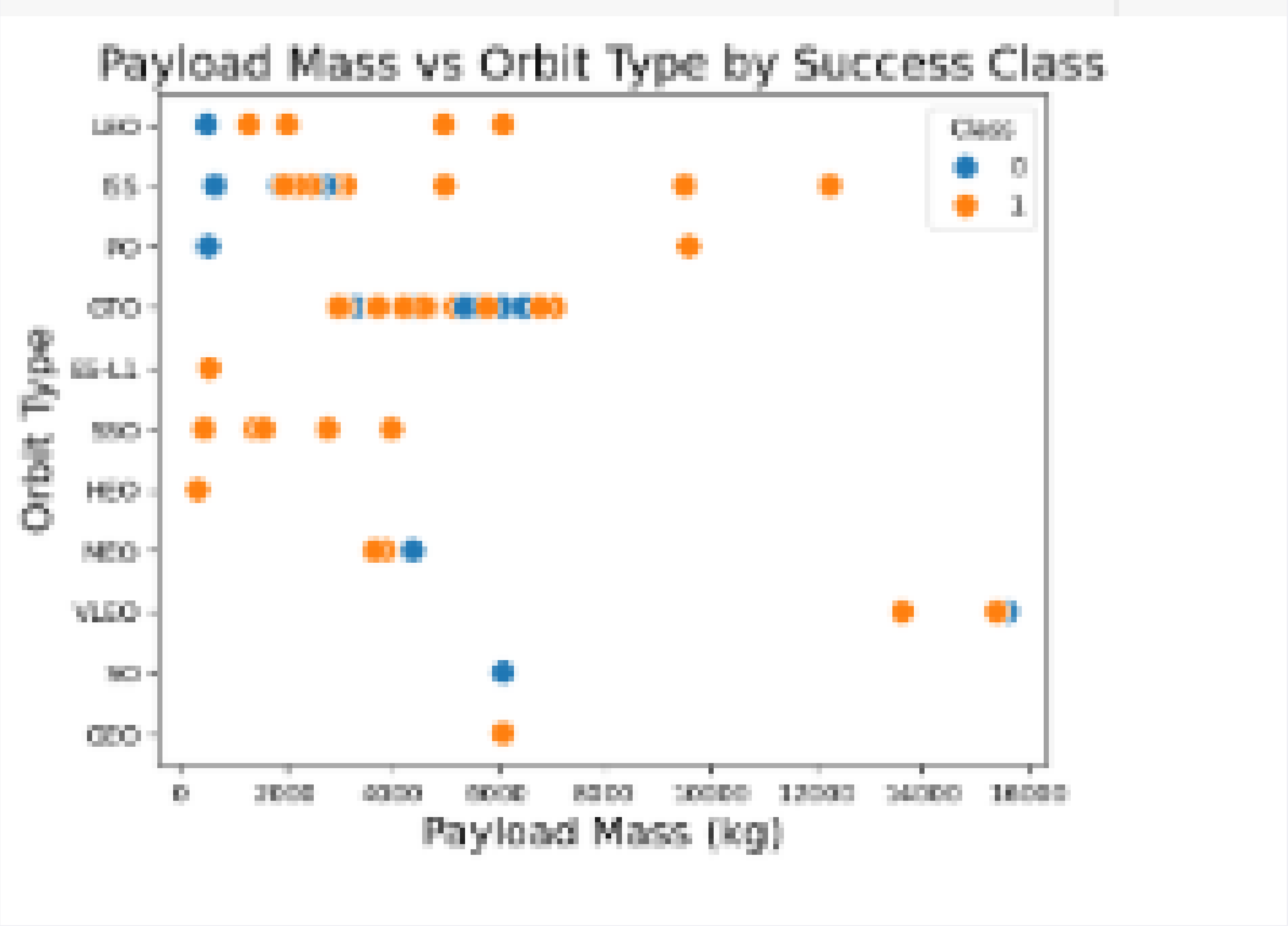
This plot visualizes the relationship between the mass of the payload (x-axis) and the type of orbit achieved (y-axis) for each mission. Each point represents a launch, colored to show success (blue) or failure (orange).

Key insights:

There is no strong correlation between payload mass and success rate; both successful and failed missions occurred across the full range of payload masses. Some orbit types, like GTO and VLEO, consistently accommodate heavier payloads. Successful missions (blue points) are distributed across various payload masses and orbit types, demonstrating the system's flexibility.

Why it matters:

This analysis confirms the launch system's capability to successfully carry a wide range of payload masses, which is crucial for attracting a diverse customer base with different mission requirements. It also shows that payload mass itself is not a primary factor in determining a mission's success.



Launch Success Yearly Trend

What it shows:

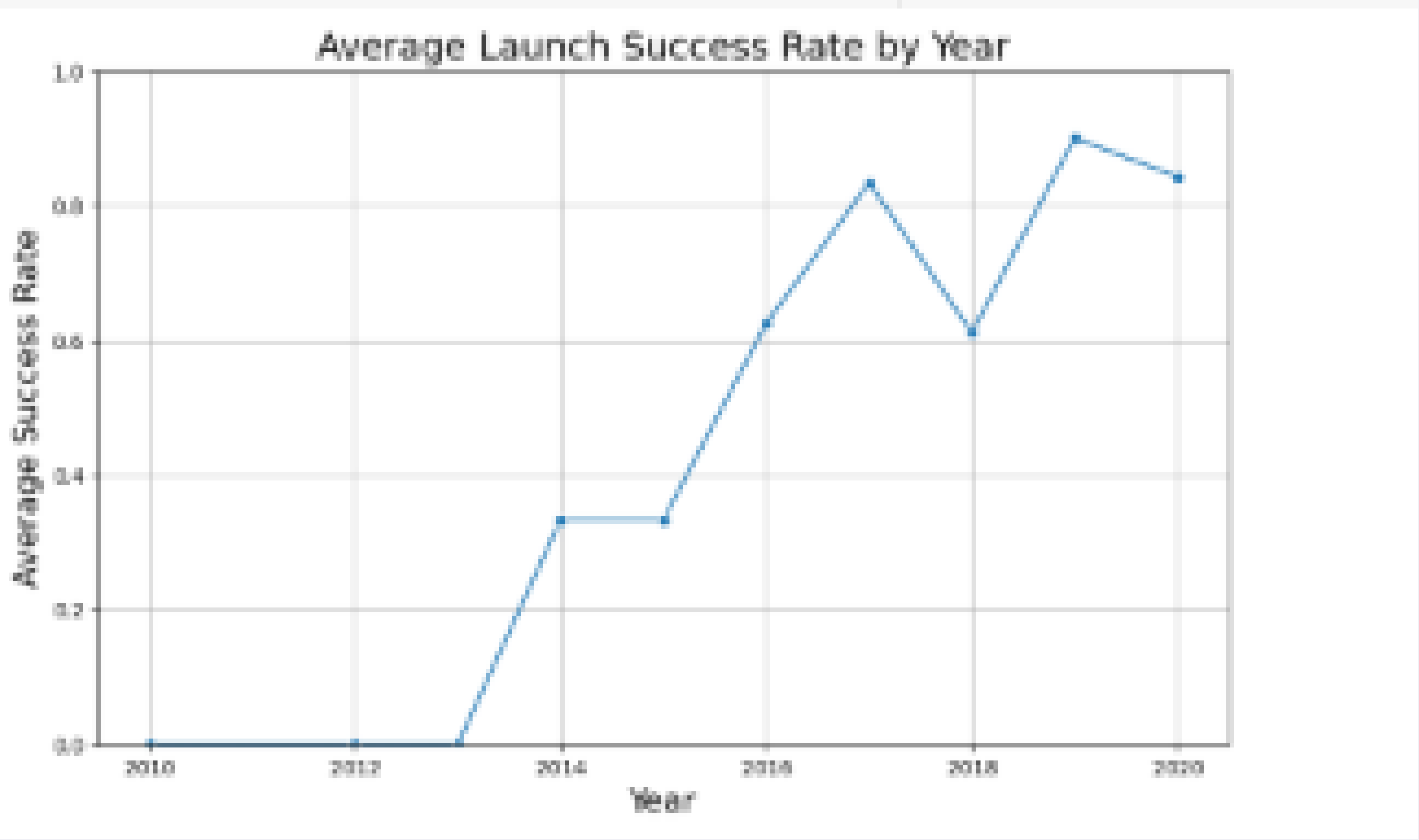
This line chart tracks the average launch success rate over time, with the year on the x-axis and the success rate (from 0 to 1) on the y-axis.

Key insights:

The success rate remained at 0% in the initial years (2010-2013), representing early development and learning phases. A significant improvement is seen starting in 2014, with the success rate rising sharply and remaining high. The success rate shows a generally upward trend over the years, demonstrating increased reliability and maturity of the launch system.

Why it matters:

This chart provides a clear and powerful story of the company's journey from initial challenges to becoming a highly reliable launch provider. It is a key metric for showing progress, building customer trust, and forecasting future performance.



All Launch Site Names

What it shows:
This query retrieves the distinct names of all launch sites used in the dataset. The result shows that missions were launched from four unique locations.
Result:
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
Why it matters:
Understanding the unique launch sites is essential for analyzing geographic and operational factors in mission success. It allows for further analysis, such as comparing success rates across different sites or understanding which sites were used for specific types of missions.

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

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

Launch Site Names Begin with 'CCA'

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

What it shows:

This query selects the first five entries from the dataset that were launched from a site with a name containing 'CCAFS'. The output provides key details for each of these early missions, including the date, booster version, payload, and the outcome of the mission and landing.

Why it matters:

This type of query is useful for performing site-specific analysis. By examining these early launches from CCAFS LC-40, we can see the initial missions, their outcomes, and the progression of the company's capabilities. This can help in understanding the history and early performance of specific launch operations.

Total Payload Mass

What it shows:
This query calculates the total payload mass carried for all missions with the customer name 'NASA (CRS)'. The result is a single value, representing the cumulative mass in kilograms.
Result:
45598 kg
Why it matters: Calculating the total payload mass for a specific customer provides a clear measure of the volume of work performed for them. This metric is important for understanding key customer relationships, revenue analysis, and the scale of the missions they have contracted.

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

 * sqlite:///my_data1.db
Done.
SUM(PAYLOAD_MASS__KG_)
45596
```

Average Payload Mass by F9 v1.1

What it shows:
This query computes the average payload mass for all launches that used the 'F9 v1.1' booster version. The result provides a single numerical value representing the average mass in kilograms.
Result: 2928.4 kg
Why it matters:
This calculation is useful for assessing the typical performance and capability of a specific booster version. Understanding the average payload for different booster types helps in mission planning and provides insight into the evolution of the vehicle's capacity.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

 * sqlite:///my_data1.db
Done.
AVG(PAYLOAD_MASS__KG_)
2928.4
```

First Successful Ground Landing Date

What it shows:
This query retrieves the earliest date from the dataset for a mission that had a successful landing on a ground pad.
Result: 2015-12-22
Why it matters: This date marks a significant milestone: the first time a booster was successfully landed on a solid ground pad after launch. This achievement was a crucial step towards reusability and a major breakthrough in aerospace technology, leading to significant cost reductions in space access.

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

 * sqlite:///my_data1.db
Done.
MIN(Date)
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

What it shows:
This query identifies the specific booster versions that successfully landed on a drone ship while carrying a payload mass between 4000 kg and 6000 kg. The result is a list of the booster version names.
Result:
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2
Why it matters:
This analysis helps to understand which specific versions of the booster were capable of a complex maneuver (drone ship landing) with a significant payload. This demonstrates the performance capabilities and reliability of these particular booster versions, which is valuable for engineering and operational planning.

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

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

What it shows: This query summarizes all mission outcomes. By combining the counts for all success-related categories, we get a total number of successful missions.
Result:
Total Successful Missions: 100
Why it matters: This calculation provides a single, clear metric for the overall success of the launch program. It shows the high level of reliability achieved by the company, which is a key factor for customers and stakeholders in assessing performance.

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

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

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

Boosters Carried Maximum Payload

What it shows:
This query identifies all booster versions that have carried the single largest payload mass recorded in the dataset. The result lists the names of these boosters and their corresponding payload mass.
Result:
F9 B5 B1048.4 - 15800 kg
F9 B5 B1049.4 - 15800 kg
F9 B5 B1051.3 - 15800 kg
F9 B5 B1056.4 - 15800 kg
F9 B5 B1048.5 - 15800 kg
F9 B5 B1051.4 - 15800 kg
F9 B5 B1049.5 - 15800 kg
F9 B5 B1060.2 - 15800 kg
F9 B5 B1058.3 - 15800 kg
F9 B5 B1051.6 - 15800 kg
F9 B5 B1060.3 - 15800 kg
F9 B5 B1049.7 - 15800 kg
Why it matters:
This analysis pinpoints the specific booster versions that have demonstrated the highest lift capacity. This information is crucial for understanding the performance limits and heavy-lift capabilities of the launch system, which is a key selling point for commercial and government customers.

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

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

What it shows:
This query identifies all missions in the year 2015 that attempted a landing on a drone ship and failed. It lists the month of the attempt, the specific booster version used, and the launch site.
Result:
Month: 01, Landing_Outcome: Failure (drone ship), Booster_Version: F9 v1.1 B1012, Launch_Site: CCAFS LC-40
Month: 04, Landing_Outcome: Failure (drone ship), Booster_Version: F9 v1.1 B1015, Launch_Site: CCAFS LC-40
Why it matters:
This analysis is valuable for understanding specific failures during a critical developmental period. By pinpointing the exact missions, booster versions, and launch sites involved in these early drone ship landing failures, engineers can identify patterns, technical challenges, and improvements needed to achieve future success.

```
%%sql SELECT substr(Date, 6, 2) AS Month,
             "Landing_Outcome",
             Booster_Version,
             Launch_Site
      FROM SPACEXTABLE
      WHERE "Landing_Outcome" = 'Failure (drone ship)'
        AND substr(Date, 1, 4) = '2015';
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

What it shows:
This query ranks the various landing outcomes by the number of occurrences between June 4, 2010, and March 20, 2017. The results are ordered from the most frequent outcome to the least frequent.

Result:
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

Why it matters:
This analysis provides a clear overview of landing outcomes during a critical period of development. It shows that "No attempt" was the most common outcome, as reusability was not yet a standard practice. The equal count of "Success (drone ship)" and "Failure (drone ship)" highlights the experimental nature of this technology during this phase, while the smaller number of successful ground pad landings indicates this was a more challenging or less common objective.

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

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

Landing_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

Section 3

Launch Sites Proximities Analysis

A Map of Launch Sites

This map shows the locations of several launch sites, with markers indicating their positions.

Important Elements and Findings:

Launch Sites: Two launch sites are marked on the map.

VAFB (Vandenberg Air Force Base): Located on the west coast of the United States. It's associated with 3LC and 4E, which may be launch complexes or missions.

CCAFS (Cape Canaveral Air Force Station): Located on the east coast of the United States, in Florida. It's associated with BDC and 40A, which may also be l**aunch complexes or missions.**

Location: The map is a global projection, showing the marked sites in relation to continents and oceans. Both sites are situated within the United States.

Findings: The map clearly visualizes the geographical placement of these two major U.S. launch sites.

Launch Site Proximity Analysis:

Equator Proximity: The launch sites are not in close proximity to the Equator. Both are located in the Northern Hemisphere, significantly north of the Equator. Launch sites are often located closer to the Equator to take advantage of the Earth's rotational speed, which provides a boost to rockets launched eastward.

Coast Proximity: Yes, both launch sites are in very close proximity to the coast.

VAFB is on the Pacific coast of California.

CCAFS is on the Atlantic coast of Florida. This coastal location is crucial for safety, as rockets launched eastward over the ocean can drop spent stages and debris into the water, away from populated areas.

Launch Success Rate Analysis

This map visualizes the location of two important launch complexes, with color-coded markers indicating the success rate of their launches.

Launch Locations: The map shows two major locations in the United States:

VAFB (Vandenberg Air Force Base): Located on the west coast of California.

CCAFS (Cape Canaveral Air Force Station): Located on the east coast of Florida.

Success Rate: The markers on the maps are color-coded to indicate the outcome of the launches:

Green: Represents a successful launch.

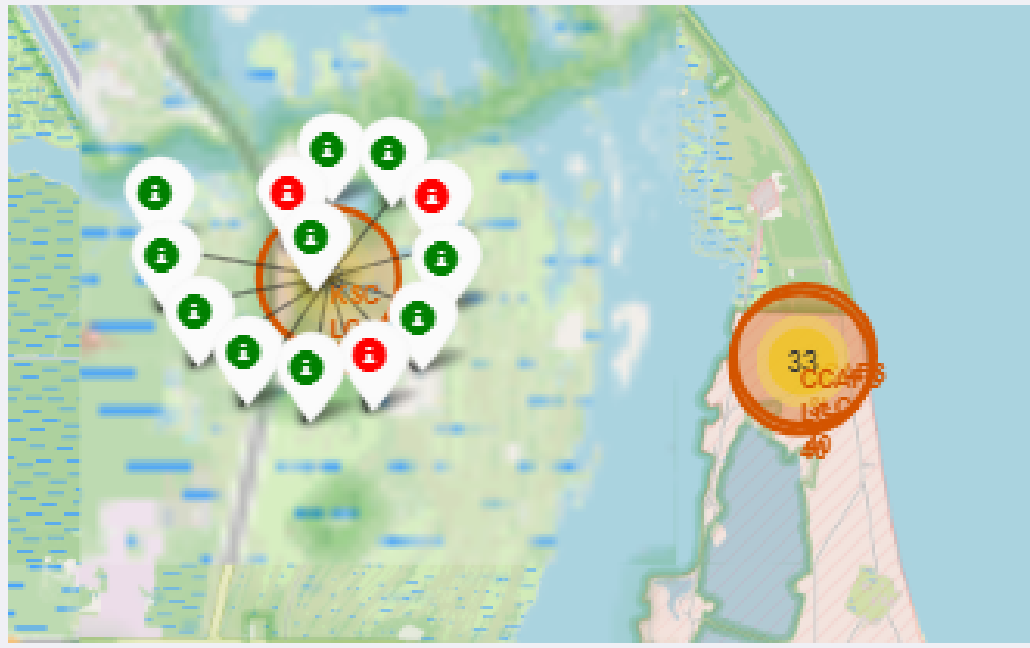
Red: Represents an unsuccessful launch.

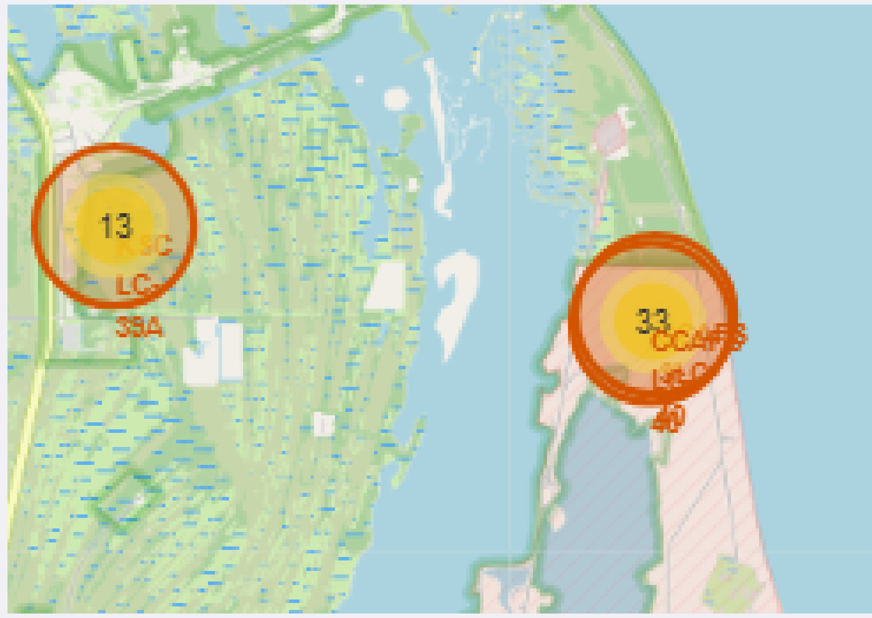
Findings:

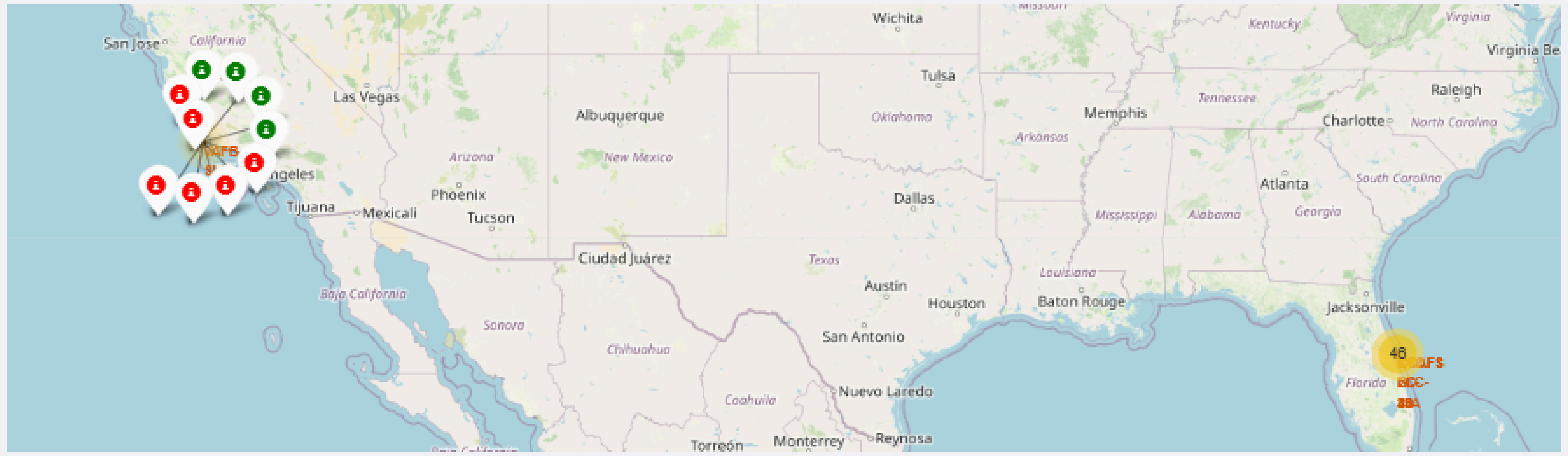
Both VAFB and CCAFS have a mix of green and red markers, indicating that both sites have a history of successful and unsuccessful launches.

The marker clusters at each location show the high concentration of launch activities in these areas.

The proximity to the coast is crucial for safety, allowing spent rocket stages to fall safely over the ocean.







Launch Site Proximity Analysis

This map displays a selected launch site with calculated distances to nearby key features such as the coastline, railways, and highways.

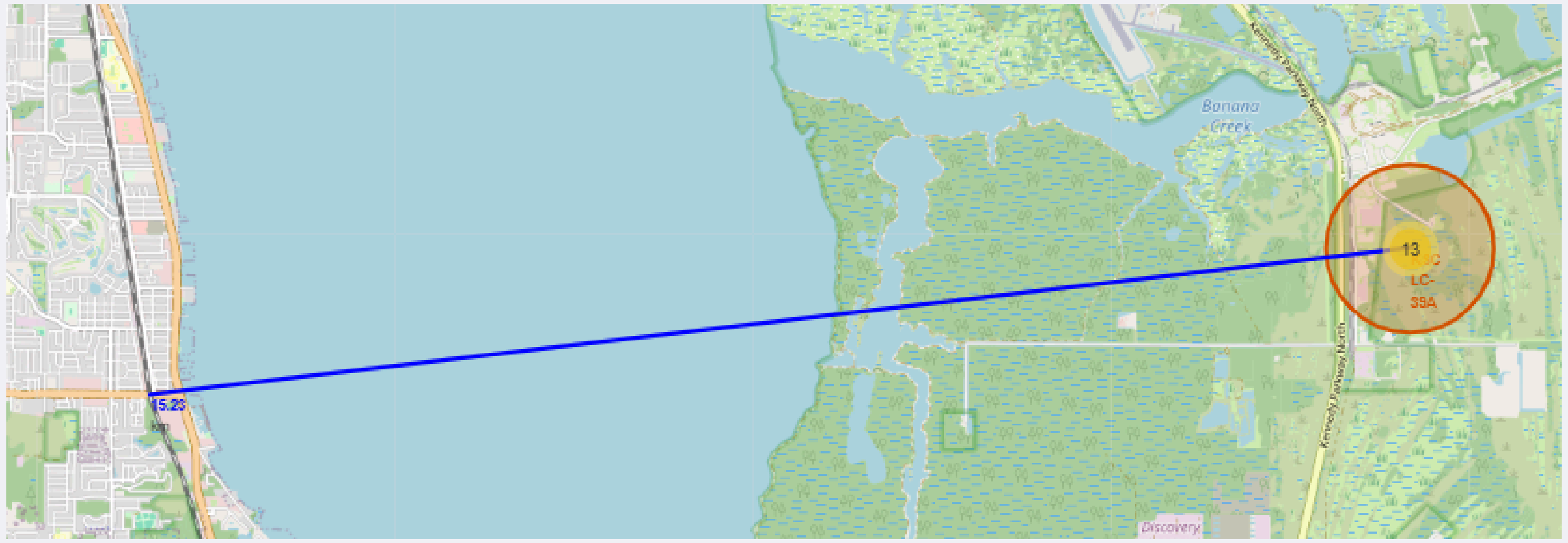
Important Elements and Findings:

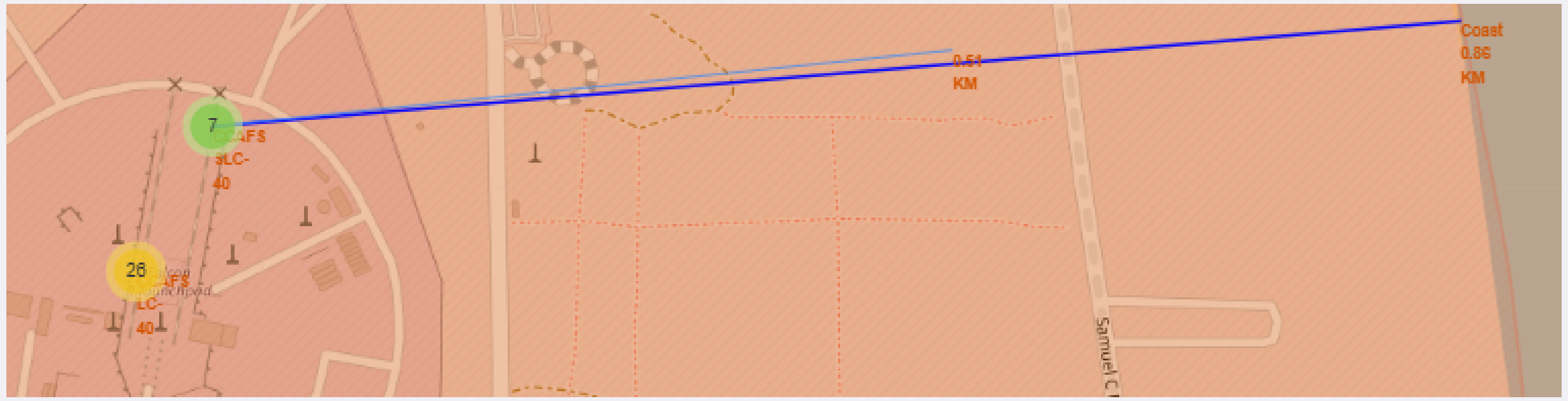
Proximity to Coastline: The screenshot shows that the launch site is located a short distance from the coastline, with a measured distance of 0.86 KM. This is important for safety, as rockets are typically launched eastward over the ocean.

Proximity to Railways: A railway line is also shown near the launch site, with a measured distance of 0.51 KM. This suggests that launch sites have close access to railways for logistical purposes.

Proximity to Highways: The map indicates that a launch site is easily accessible by highways. One screenshot shows a highway close to the launch site, with a measured distance of 5.29 KM.

Distance from Cities: The launch site is located a certain distance away from cities. The screenshot with the 5.29 KM distance likely demonstrates this. This finding suggests that launch sites are kept at a safe distance from populated areas.

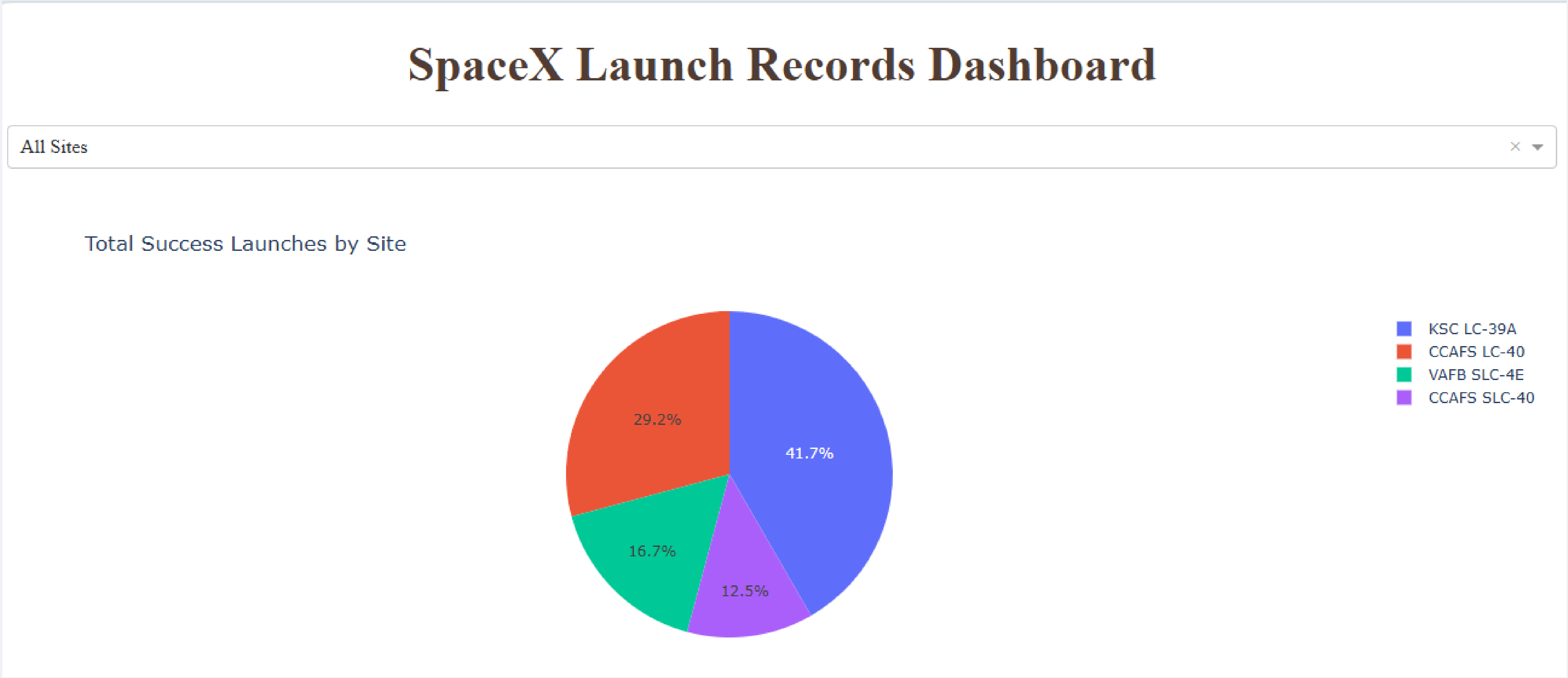




Section 4

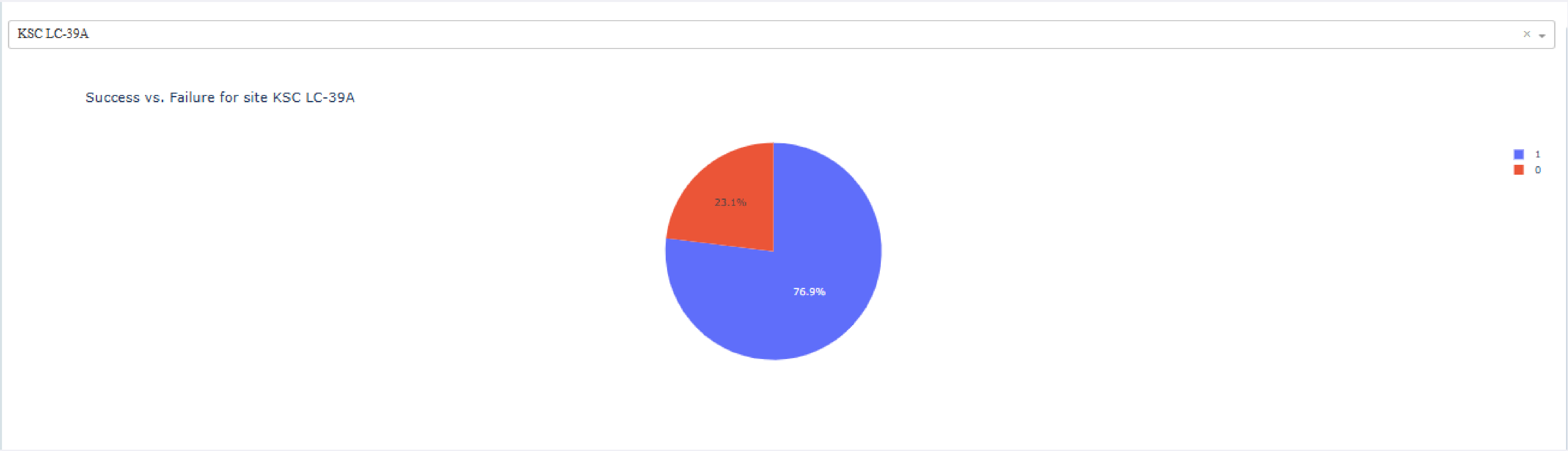
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard



This dashboard uses a pie chart to show the distribution of successful SpaceX launches by site. The chart indicates that KSC LC-39A had the largest percentage of successful launches (41.7%).

SpaceX Launch Success Rate for KSC LC-39A



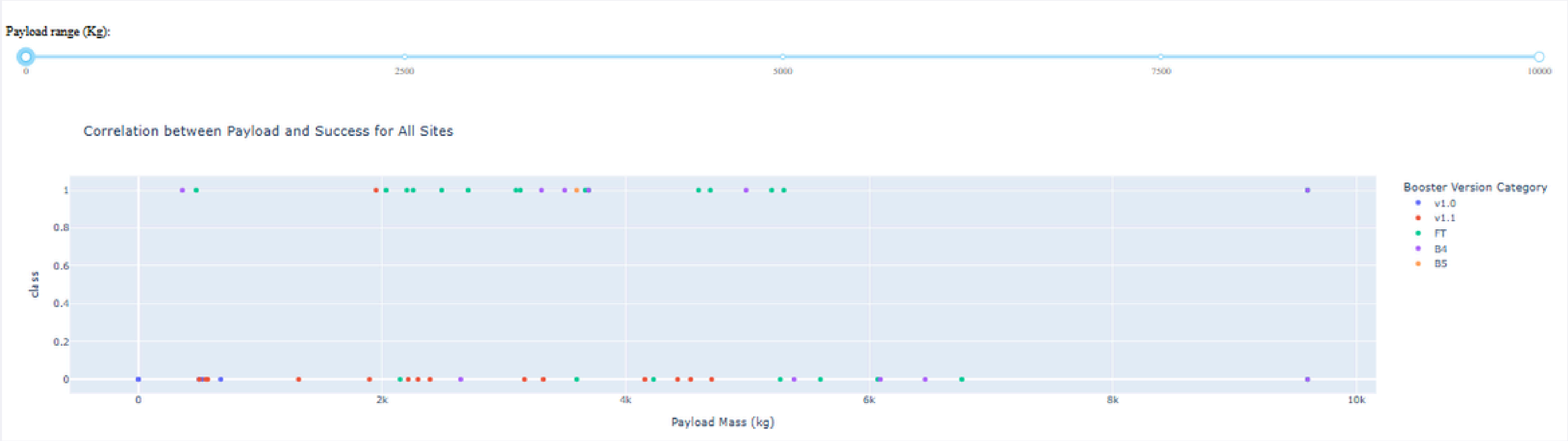
- The pie chart shows the success (76.9%) and failure (23.1%) rate for the KSC LC-39A launch site.
- Conclusion: KSC LC-39A has a high launch success rate of approximately 77%.

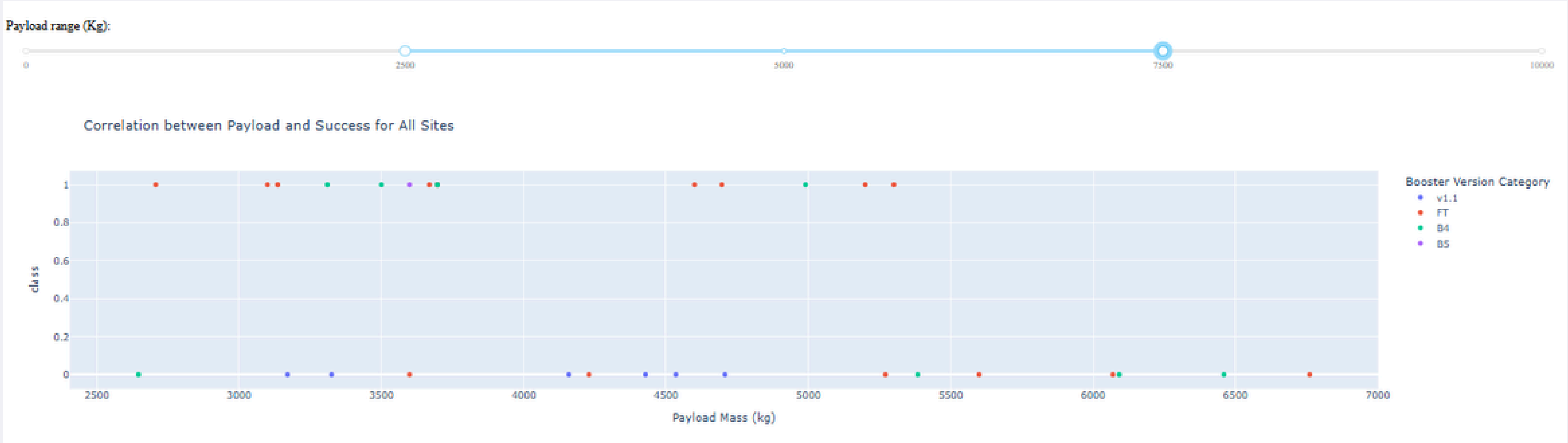
Correlation between Payload Mass and Launch Outcome

- The charts show the relationship between payload mass (X-axis), booster version (dot color), and launch outcome (Y-axis, where 1 = success and 0 = failure).

Key Conclusions

- Versions with Highest Success Rate: The FT (green) and B5 (purple) booster versions have an almost perfect success rate.
- Versions with Lower Success Rate: Older versions like v1.0 (blue) and v1.1 (red) have a higher number of failures.
- Payload: Payload mass is not the main factor for success; the most correlated factor is the booster version.





Section 5

Predictive Analysis (Classification)

Classification Accuracy

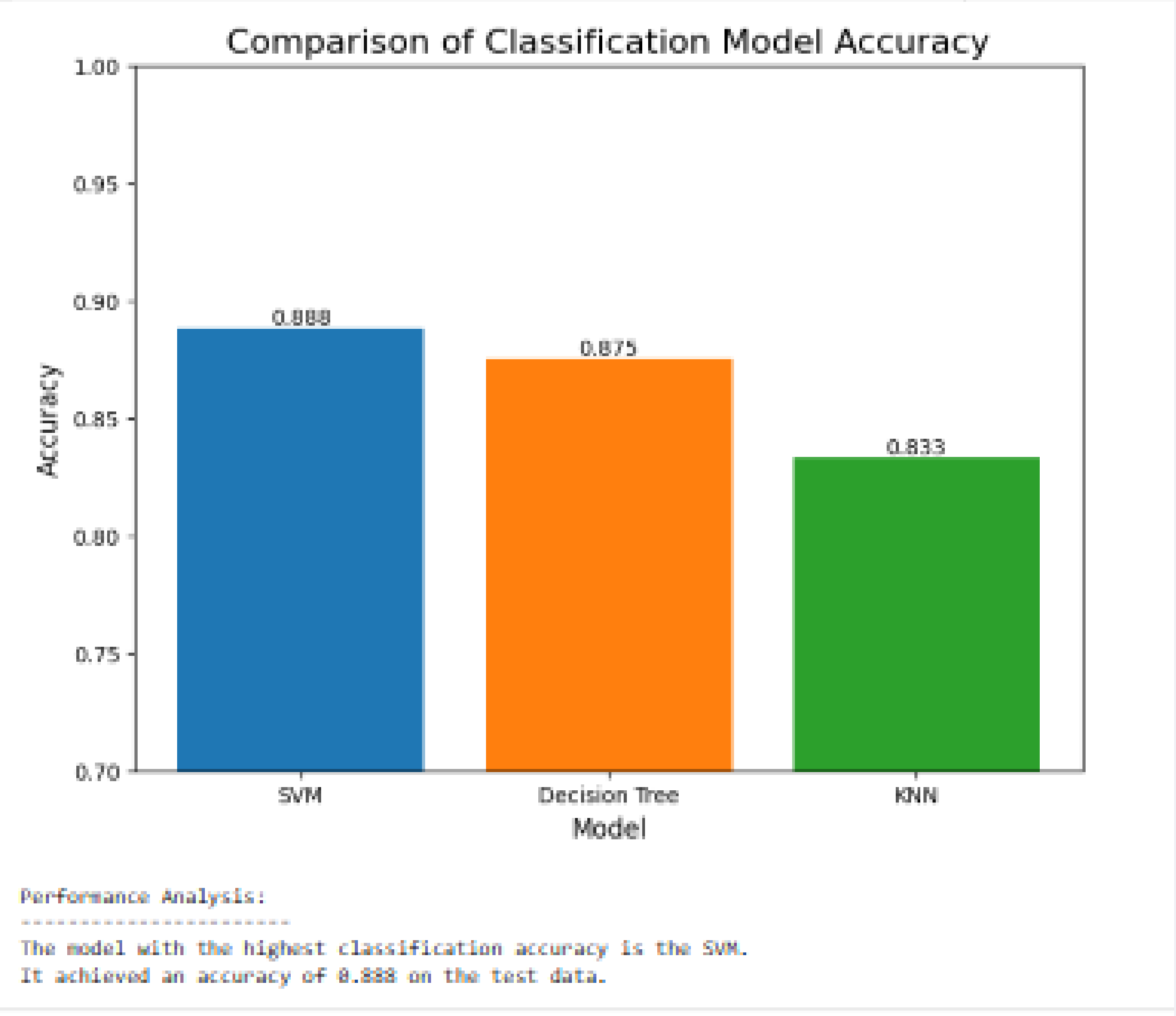
Model Performance Comparison: The bar chart visually compares the test accuracy of three different classification models: SVM, Decision Tree, and KNN.

Best Performing Model: The SVM model has the highest accuracy at 0.888. This is the tallest bar on the chart.

Second Best Model: The Decision Tree model is the second-best performer with an accuracy of 0.875.

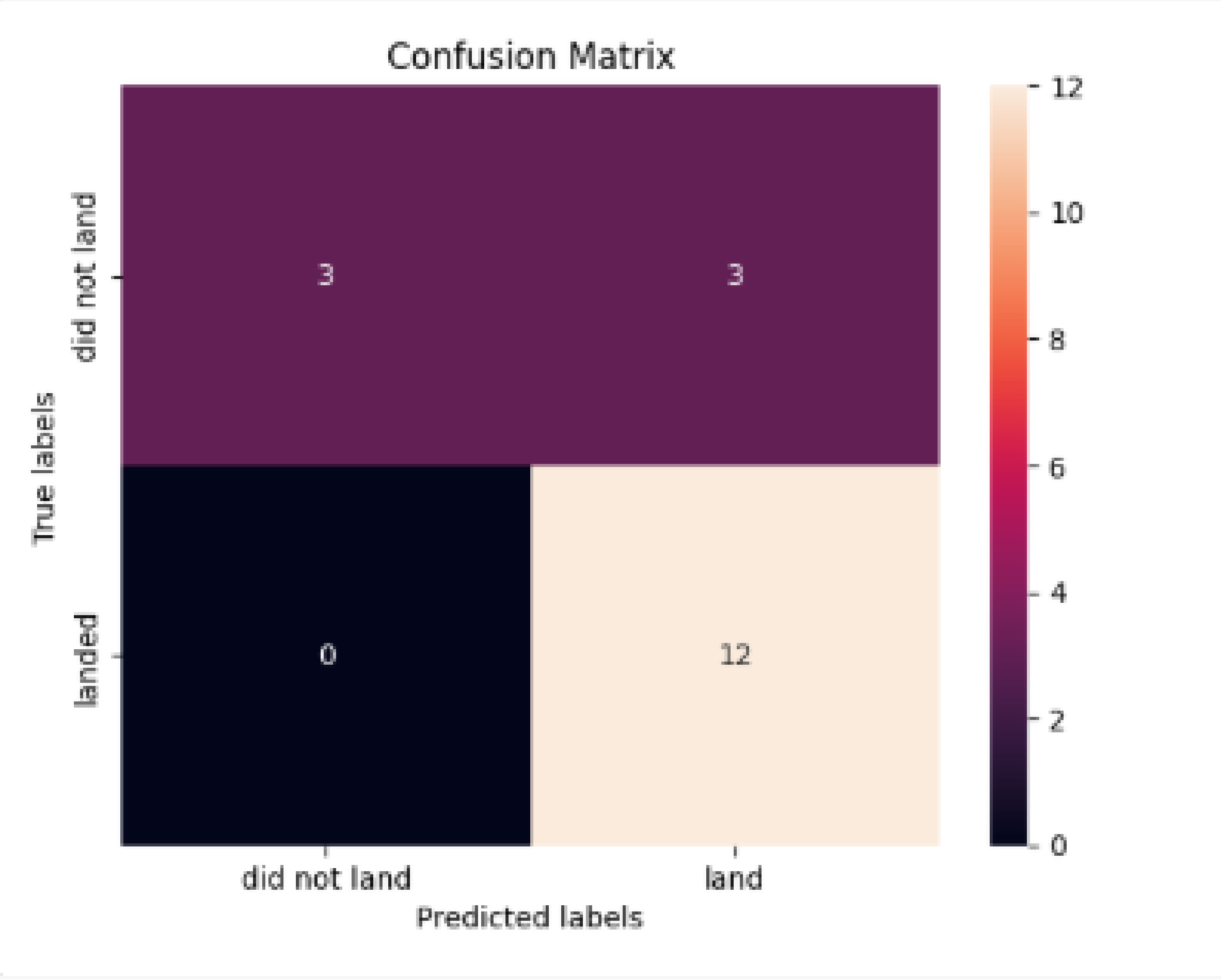
Lowest Performing Model: The KNN model has the lowest accuracy among the three at 0.833.

Conclusion: The analysis shows that for this specific dataset and the chosen hyperparameters, the SVM model is the most effective at classifying the data, followed closely by the Decision Tree model. The KNN model performed the least well.



Confusion Matrix

- The confusion matrix gives a more detailed look at the model's performance beyond a simple accuracy score.
- Key Insights:
- True Positives (12): The model correctly predicted 12 successful landings. This is a crucial metric, showing the model is excellent at identifying positive outcomes.
- True Negatives (3): The model correctly predicted 3 unsuccessful landings.
- False Positives (3): The model incorrectly predicted a landing 3 times when the rocket actually did not land. This is a "Type I Error."
- False Negatives (0): The model did not incorrectly predict any unsuccessful landings when the rocket actually landed. This is a "Type II Error," and a value of 0 is excellent, as it means the model is highly reliable in predicting successful landings.
- Conclusion:
- The SVM model performs very well, particularly in avoiding false negatives. It successfully identified all successful landings, which is critical for a task like this.



Conclusions

- The analysis of public SpaceX launch data provided valuable insights to support strategic decisions for new aerospace players.
- The data shows a clear upward trend in mission success rate, especially since 2014, demonstrating the increased reliability of the launch system over time.
- There is no strong correlation between payload mass and launch success; instead, the booster version is a more significant factor.
- The trained machine learning models successfully predict booster landing outcomes. The Support Vector Machine (SVM) model performed best with an accuracy of 88.8%.
- The SVM model proved to be highly reliable, correctly identifying all successful landings (no false negatives), which is crucial for operational planning

Thank you!