IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mark Joseph Bantayan
June 29, 2025

Outline

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

Executive Summary

Summary of Methodologies

- Acquired data via API and web scraping
- Cleaned and structured data using Pandas and SQL
- Performed EDA using visualization tools and SQL queries
- Built interactive visual analytics using Folium and Plotly Dash
- Applied and fine-tuned classification models with GridSearchCV
- Evaluated models using accuracy scores and confusion matrices

Summary of Results

- **Logistic Regression**: 83.33% test accuracy
- **Support Vector Machine (RBF kernel)**: 83.33% test accuracy
- **Decision Tree**: 83.33% test accuracy (best: entropy, log2, depth=4)
- **K-Nearest Neighbors**: 83.33% test accuracy
- **Best Validation Accuracy**: Decision Tree (87.5%)
- **Most Interpretable Model**: Logistic Regression

Introduction

Project background and context

This project aims to analyze and predict the outcomes of SpaceX Falcon 9 rocket launches. The data-driven approach helps understand key factors influencing successful landings, which is critical for reducing costs and improving the reusability of rockets.

Problems We Want to Answer

- What launch characteristics most influence a successful landing?
- Can we accurately predict the success of a launch landing using historical data?
- Which classification model performs best for this prediction task?

Section 1

Methodology

Methodology

Executive Summary

- Collected SpaceX Falcon 9 launch data via API and web scraping
- Cleaned and wrangled data for consistency and usability
- Processed data to engineer relevant features
- Performed EDA using visualizations and SQL queries
- Created interactive visualizations using Folium and Plotly Dash
- Built classification models to predict landing success
- Tuned models with GridSearchCV and evaluated performance with accuracy and confusion matrices

Data Collection

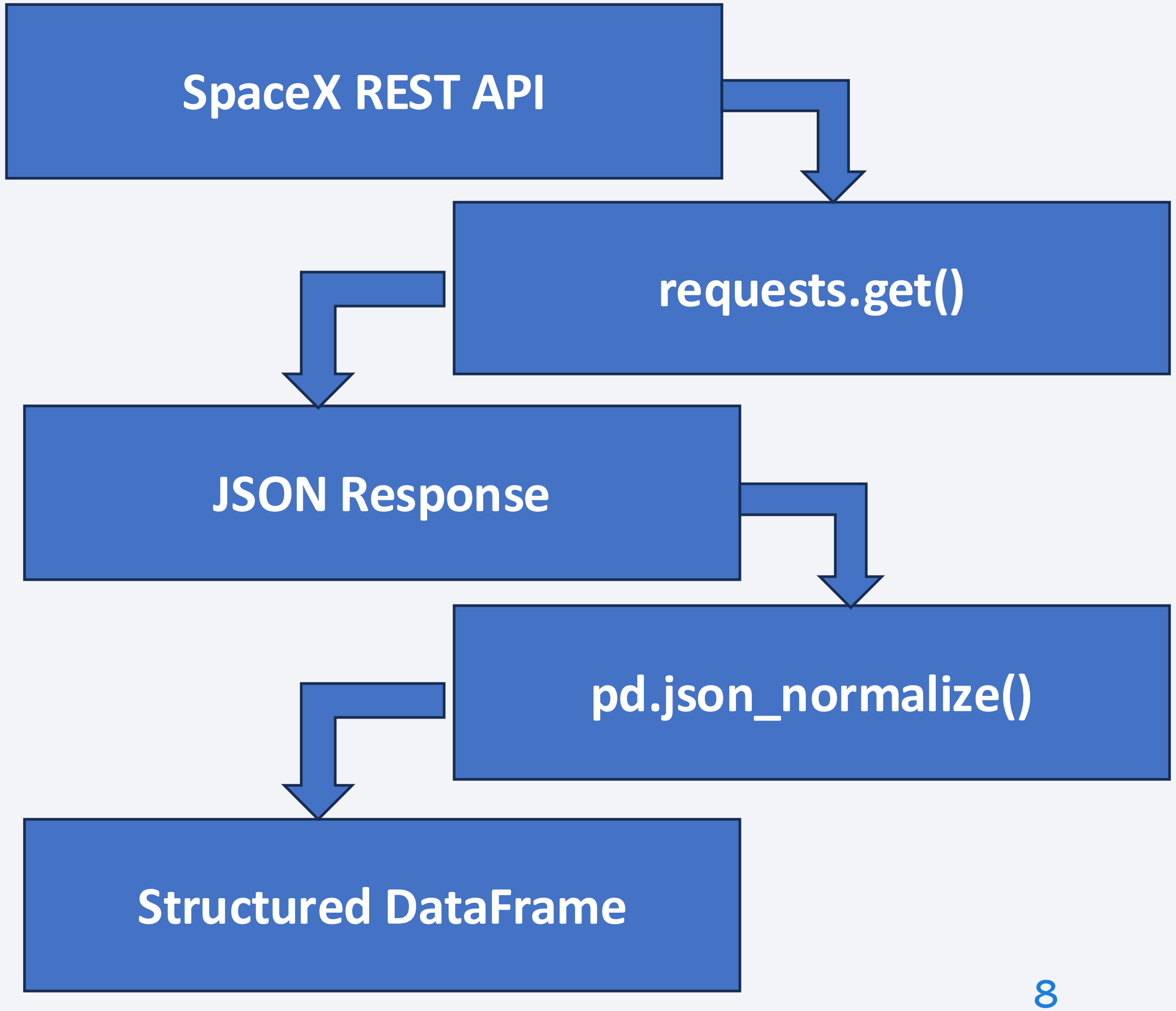
Data Collection Process

- **API Retrieval**:
  Used SpaceX Launch Data API to collect structured data on launch events (launch dates, payload mass, launch site, etc.)
- **Web Scraping**:
  Scraped SpaceX official website for additional features not available via API (booster versions, landing outcomes)
- **Tools Used**:
  Python, `requests`, `BeautifulSoup`, `pandas`, `json`, `re` for extraction and structuring
- **Data Storage**:
  Stored collected data in Pandas DataFrames for further cleaning and wrangling

Data Collection – SpaceX API

SpaceX API Data Collection Overview

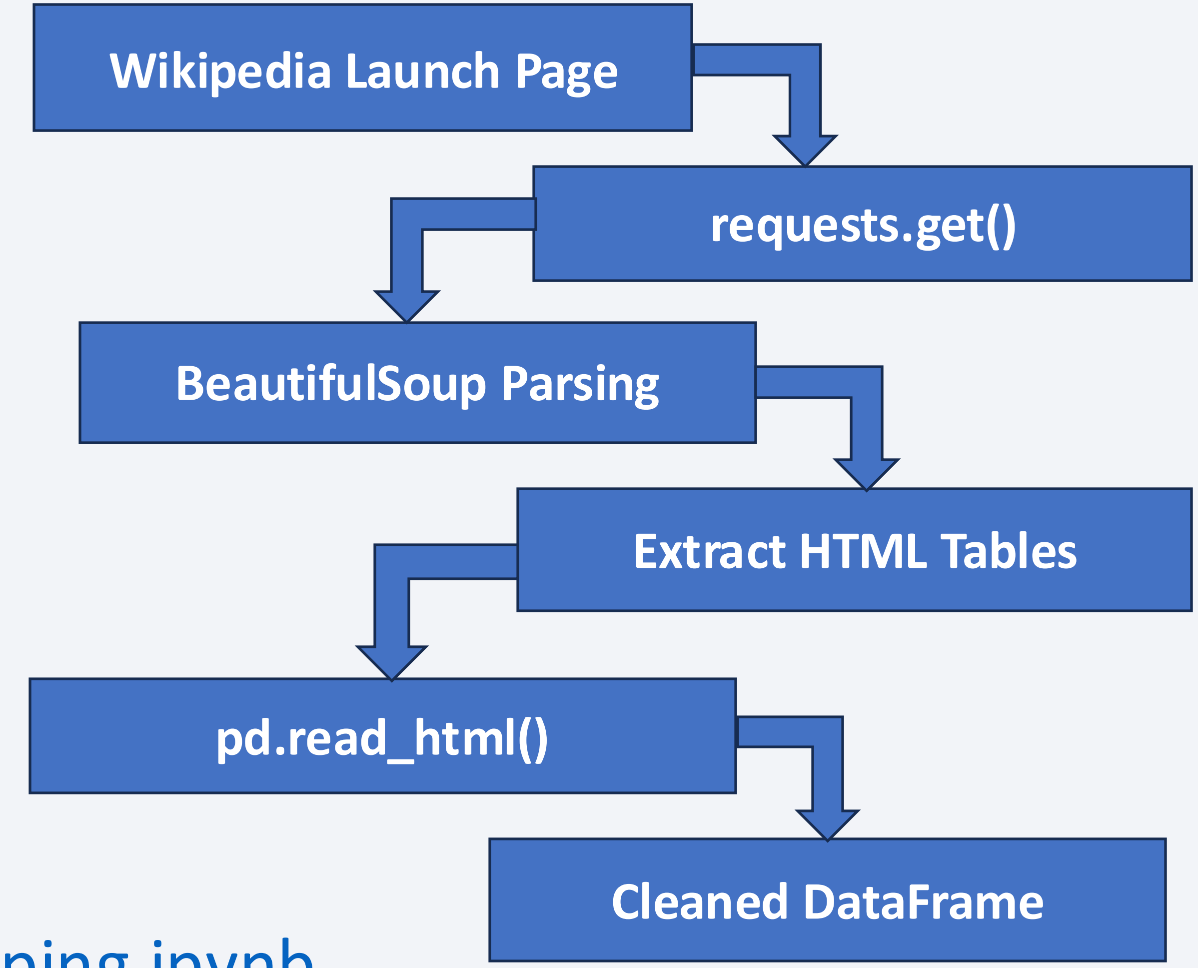
- Accessed **SpaceX Launch Data API** using `requests.get()`
- Extracted data in **JSON** format
- Parsed and normalized JSON with `pandas.json_normalize()`
- Focused on fields such as: `rocket_name`, `payload_mass`, `launch_site`, `landing_outcome`
- Stored as DataFrame for wrangling and analysis
- GitHub link: [001 - jupyter-labs-spacex-data-collection-api.ipynb]



Data Collection - Scraping

Web Scraping Process Overview

- Used `requests` to retrieve the HTML content about SpaceX Falcon 9 launches on the **Wikipedia page** listing
- Parsed the HTML using `BeautifulSoup`
- Extracted **launch data tables** with mission details and outcomes
- Converted tables into **pandas DataFrames** for further analysis



GitHub link: [002 - jupyter-labs-webscraping.ipynb]

Data Wrangling

Data Processing & Wrangling

- Combined data from **API** and **web scraping**
- Renamed columns for clarity and consistency
- Handled missing values and standardized labels
- Converted datatypes for analysis (e.g., dates, floats)
- Created new features like **Launch Outcome**, **Payload Mass**, etc.
- Saved cleaned data to new DataFrames for downstream use

GitHub link: 003 - labs-jupyter-spacex-Data wrangling.ipynb

EDA with Data Visualization

Charts Plotted & Their Purpose

- **Flight Number vs Payload Mass** (scatter with hue=Class)
  Visualized how launch count and payload weight affected landing success—higher flight number (experience) improved outcomes, even under heavier payloads.
- **Flight Number vs Launch Site (**catplot)
  Compared launch progression across sites (e.g., CCAFS vs VAFB) to spot site-specific trends in success over time.
- **Payload Mass vs Launch Site** (catplot)
  Checked if certain sites handled heavier payloads more reliably and how payload weight varied by location.
- **Success Rate by Orbit Type** (bar chart)
  Illustrated success/failure rates for each orbit (LEO, GTO, ISS, etc.) to understand orbit complexity vs. landing outcome.
- **Flight Number vs Orbit Type** (scatter/catplot)
  Examined how booster experience impacted missions to different orbits—did more flights help with challenging orbits?
- **Payload Mass vs Orbit Type** (scatter)
  Explored the relationship between payload mass and orbit type to see if heavy payloads correlated with specific orbits and success.
- **Annual Success Trend** (line chart)
  Showed year-over-year average success rate, highlighting improving reliability as SpaceX matured operations.

GitHub link: [IBM-Data-Scientist-Capstone-Proj /005 - edadatavizz.ipynb]

EDA with SQL

SQL Query Summary

- Retrieved unique launch sites
- Filtered launches by site prefix
- Calculated total payload for NASA (CRS)
- Averaged payload for specific boosters
- Identified first successful ground landing
- Found boosters with 4,000–6,000 kg drone ship successes
- Counted landing outcomes (success/failure)
- Retrieved booster with highest payload
- Analyzed 2015 drone ship failures by month
- Ranked outcomes between 2010–2017 by count

GitHub link: 004 - jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Folium Map Summary

- **Markers**: Plotted all launch sites
- **Popups**: Displayed site names on click
- **Circles**: Highlighted proximity areas around sites
- **Lines**: Drew launch site-to-orbit path (simulated)
- **Custom Icons**: Made site markers visually distinct

GitHub link: [006 - lab_jupyter_launch_site_location.ipynb]

Build a Dashboard with Plotly Dash

Dashboard Visualizations and Interactions

- Added interactive dropdown menu to select launch site
- Plotted pie chart showing successful vs failed launches per site
- Plotted scatter plot showing payload vs success, with booster version color-coded
- Enabled dynamic filtering to update graphs based on selected site and payload range

Purpose:
These visual elements allow users to explore launch outcomes by site and payload, identify patterns, and understand relationships between features and success rates.

GitHub (Plotly Dash App): spacex-dash-app.py

Predictive Analysis (Classification)

Model Development Summary

- Built four classification models: Logistic Regression, SVM, Decision Tree, and KNN
- Applied **GridSearchCV** with 10-fold cross-validation to tune hyperparameters
- Evaluated models using test accuracy and confusion matrices
- Identified best model based on validation and test scores
- **Logistic Regression, SVM, and KNN** all achieved top test accuracy of **83.33%**

Purpose:

To predict launch outcomes accurately and determine the most reliable model.

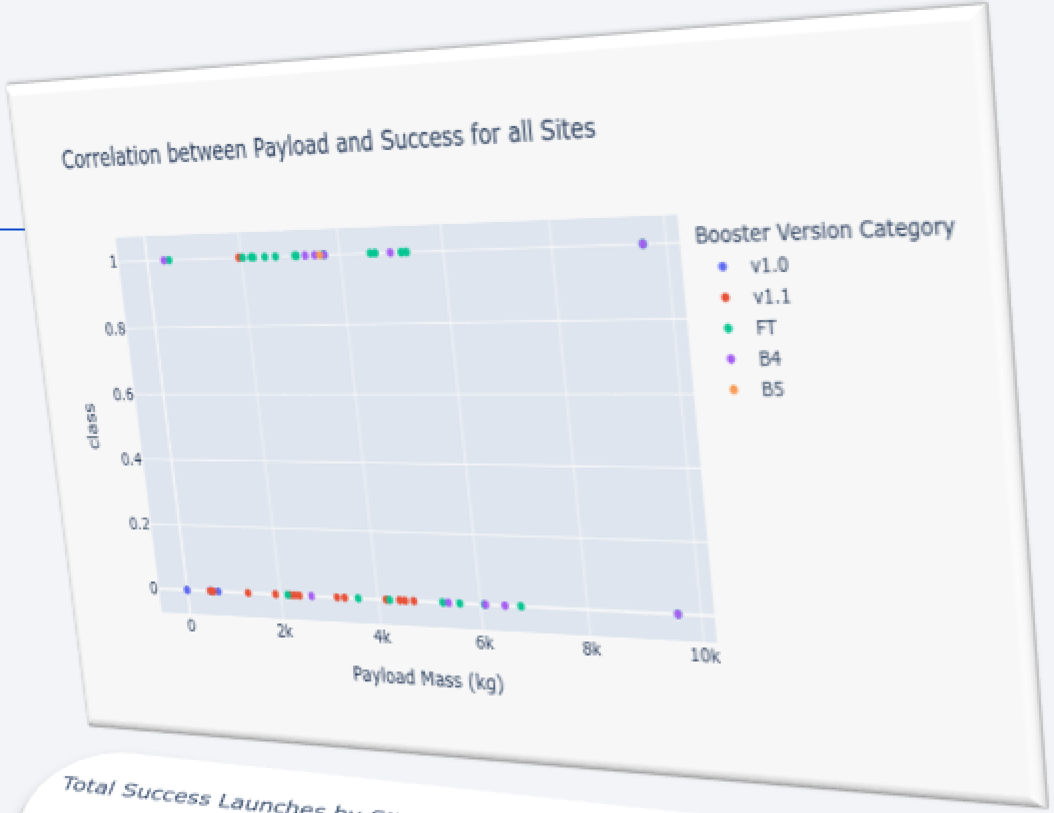
GitHub (Predictive Analysis Notebook): 008 - SpaceX_Machine Learning Prediction_Part_5.ipynb

Results

Exploratory & Predictive Analysis Results

- **EDA Highlights:**
  - Launch success rates by site
  - Correlation between payload mass and success
  - Booster version class impact on outcomes

- **Predictive Analysis Results:**
  - Best models (LogReg, SVM, KNN): **83.33% accuracy**
  - Decision Tree: **83.33%**, with best precision on class 1
  - Models tuned with GridSearchCV and evaluated on test set



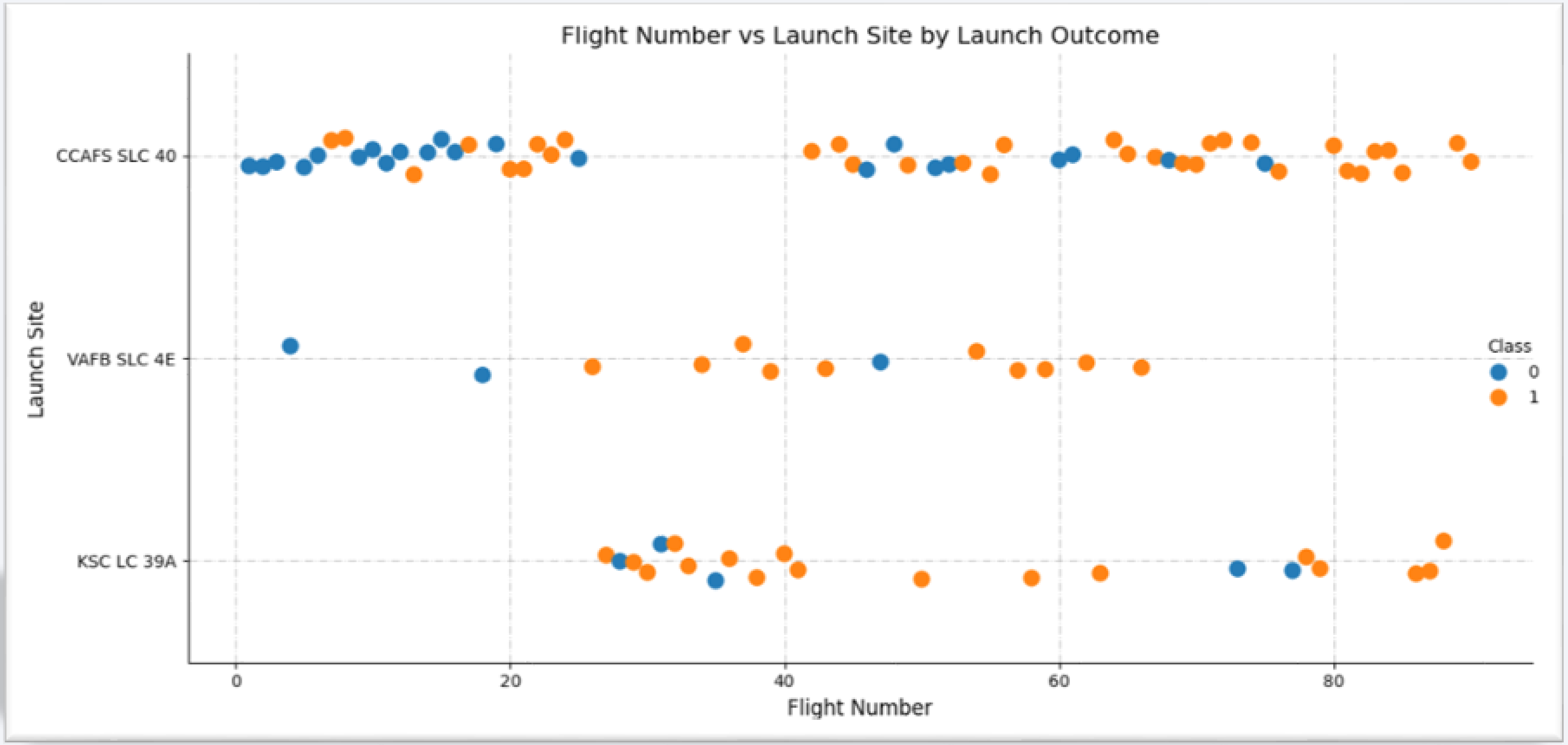
Section 2

Insights drawn from EDA

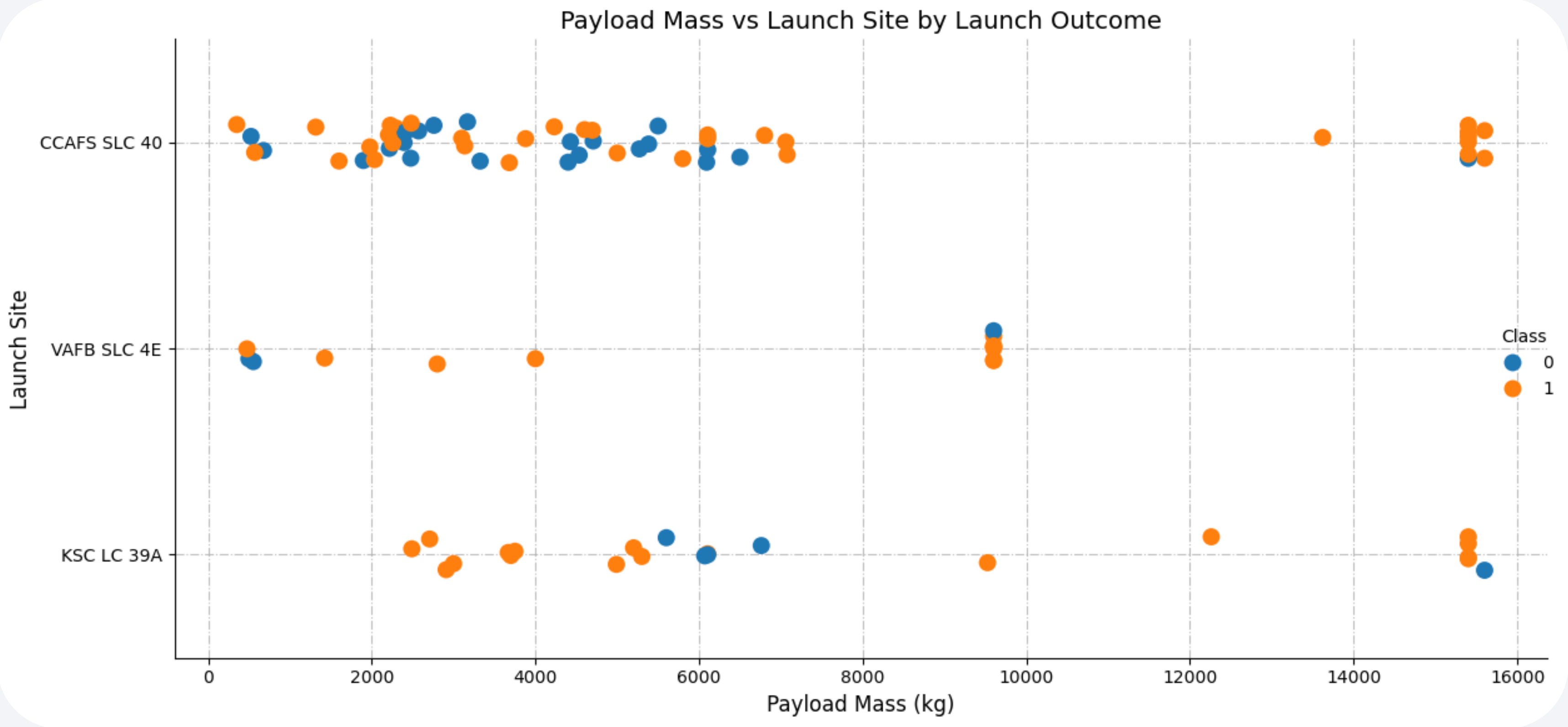
Flight Number vs. Launch Site

Purpose: Examine launch success over time across launch sites.

Insight: CCAFS SLC 40 hosted the most launches; success (orange) increases with later flight numbers, especially at this site suggesting reliability improvement.



Payload vs. Launch Site



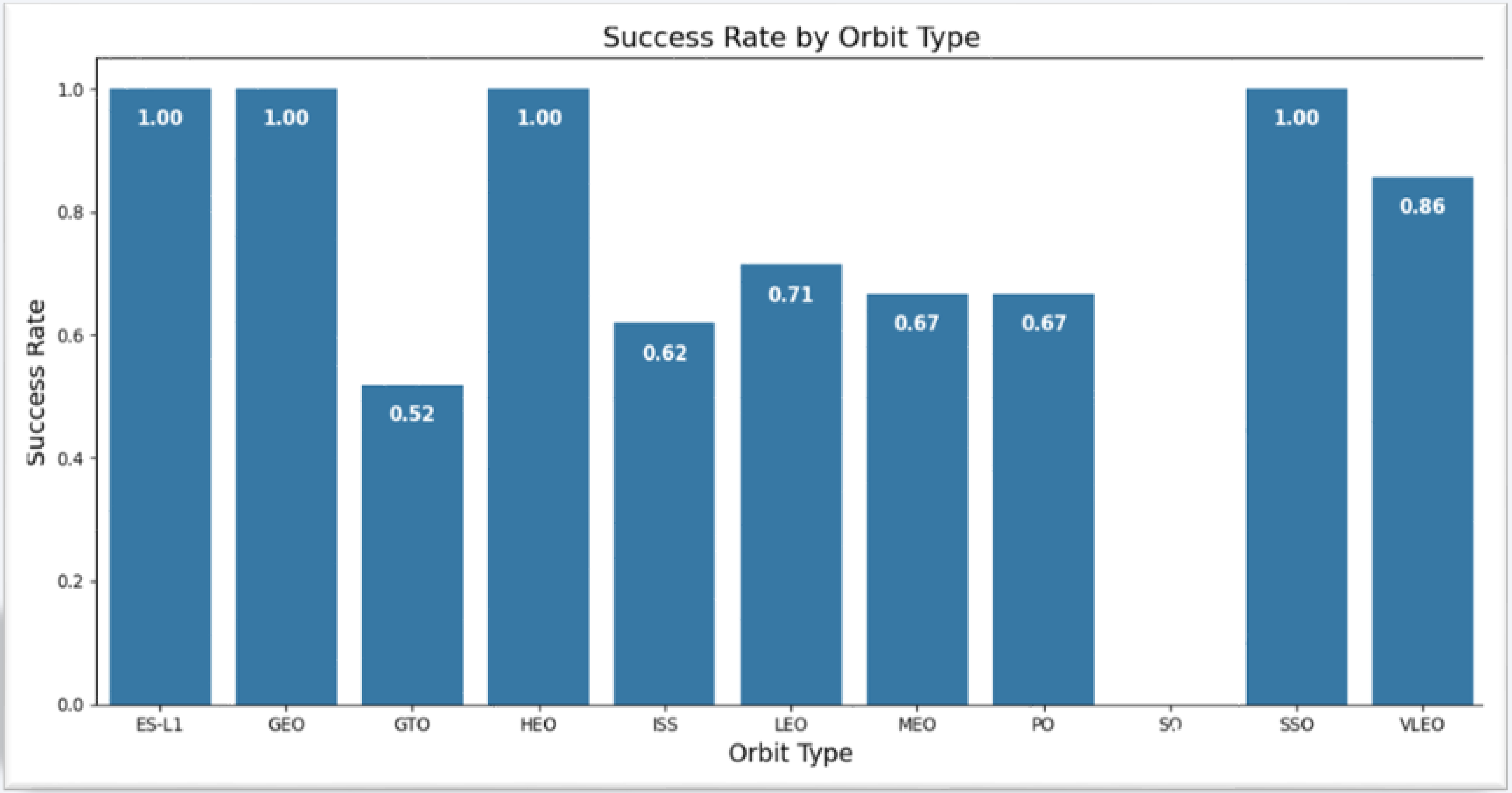
Purpose: Analyze how launch site and payload weight affect outcomes.

Insight: High-mass launches are concentrated at **CCAFS SLC 40**, with a solid success rate, implying this site handles heavier missions more reliably.

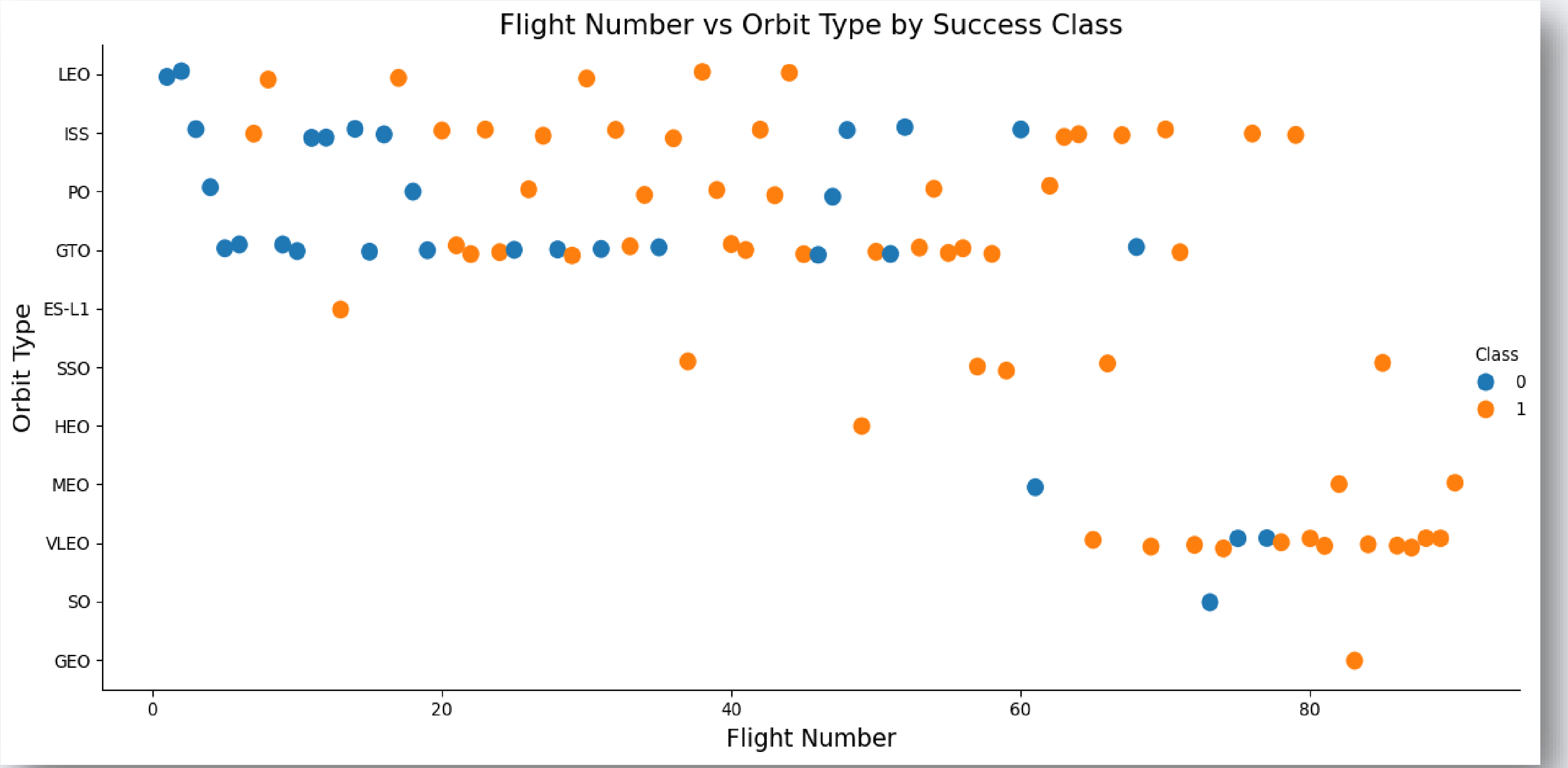
Success Rate vs. Orbit Type

Purpose: To compare SpaceX launch success rates across different orbit destinations.

Insight: Orbits like **ES-L1, GEO, HEO, and SSO** show a **100% success rate**, while others such as **GTO, ISS, LEO, MEO** have lower rates, indicating more variability in mission outcomes depending on orbit complexity.



Flight Number vs. Orbit Type



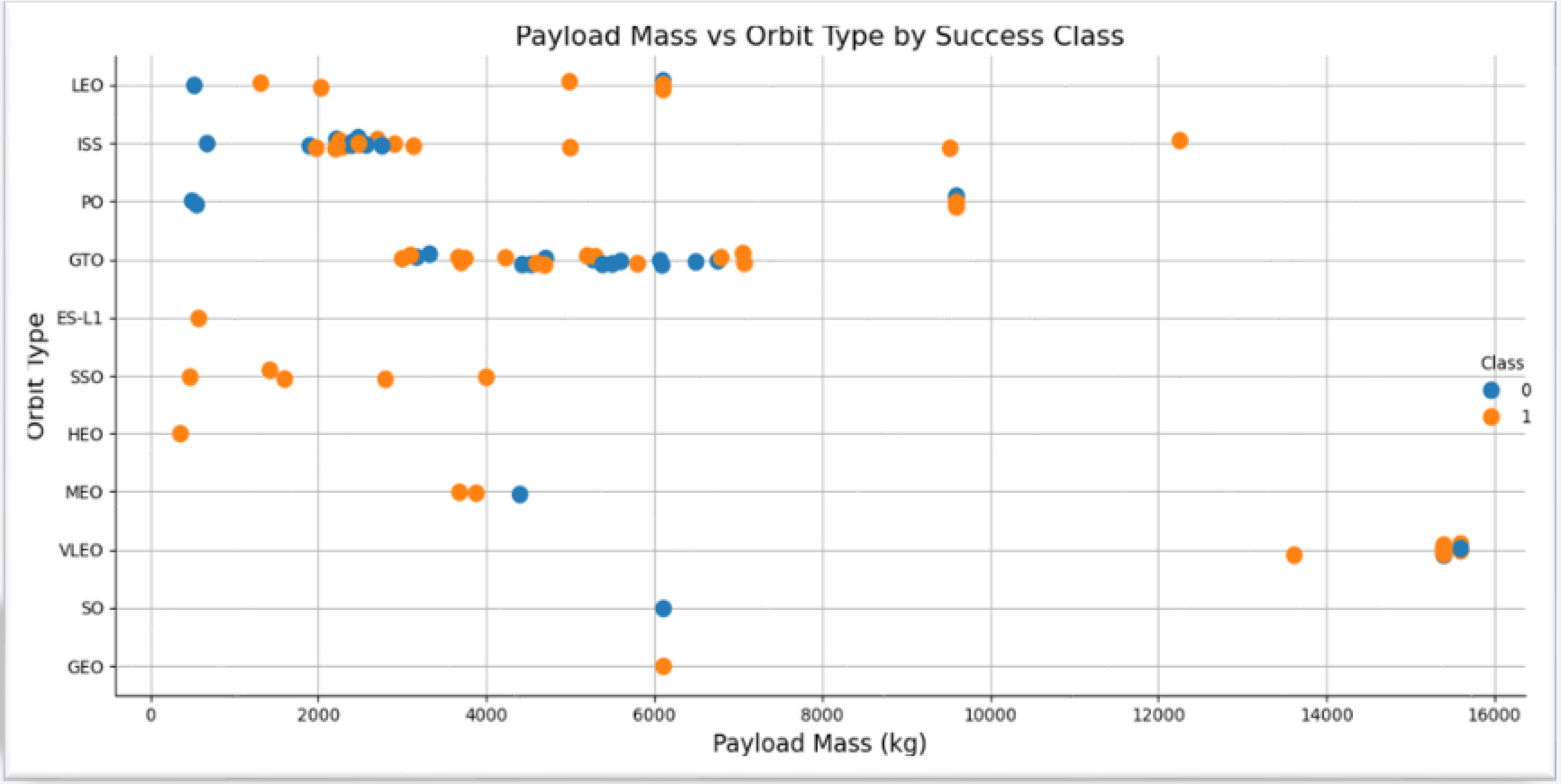
Purpose: Track which orbits were targeted over time and how successful those missions were.

Insight: Certain orbits like LEO and GTO are used consistently across flights. Earlier failures (blue) are more frequent, indicating progress in mission handling.

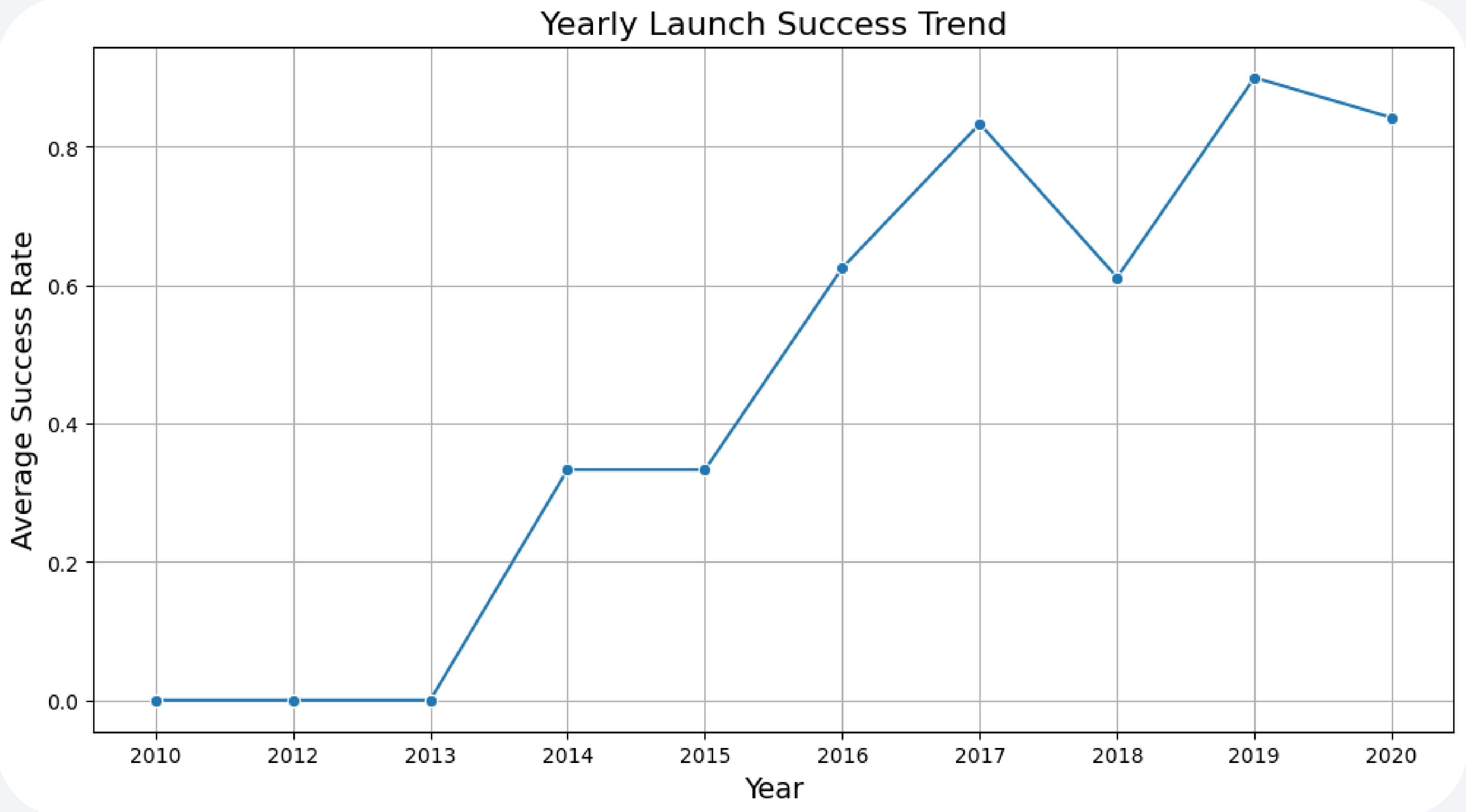
Payload vs. Orbit Type

Purpose: Visualize how payload mass and orbit destination relate to launch success.

Insight: Launches to orbits like LEO, ISS, and PO cover a wide mass range; high-mass launches still succeed, indicating no clear mass threshold for failure.



Launch Success Yearly Trend



Purpose: To show how SpaceX's average launch success rate has evolved over time.

Insight: There's a **clear upward trend** in success rate starting from around 2013, peaking in **2019**, suggesting increasing reliability and maturity of SpaceX's launch operations over the years.

All Launch Site Names

- **KSC LC-39A**: Kennedy Space Center Launch Complex 39A, originally built for Apollo missions, is now SpaceX's key pad for launching crewed and heavy-lift missions.

- **CCAFS SLC-40**: Often used interchangeably with LC-40, this Florida-based site supports a wide range of Falcon 9 missions and was refurbished by SpaceX for regular use.

Display the names of the unique launch sites in the space mission

```
In [10]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

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

Out[10]:

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

- **CCAFS LC-40**: Cape Canaveral Air Force Station Launch Complex 40 in Florida is one of SpaceX's primary sites for launching commercial and government payloads to orbit.

- **VAFB SLC-4E**: Vandenberg Air Force Base Space Launch Complex 4E in California is used for polar orbit launches, especially for Earth observation satellites.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

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

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

[12]:

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

Cape Canaveral Air Force Station (CCA) is highly significant SpaceX's launch history, querying about it allow us to focus on one of its most frequently used and historically important sites to better understand mission performance and outcomes tied to this location.

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

[23]:
```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") AS TotalPayloadMass
FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA (CRS)%';
```

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

[23]:

TotalPayloadMass
48213

- Knowing the payloads carried by boosters from NASA is crucial for evaluating the technical demands and strategic importance of SpaceX missions, as these payloads often represent high-priority scientific equipment, ISS resupply cargo, or experimental technologies that reflect both NASA's trust in SpaceX and the mission's broader contribution to space exploration.

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
[24]: %%sql
SELECT AVG("PAYLOAD_MASS__KG_") AS AveragePayload
FROM SPACEXTABLE
WHERE "Booster_Version" LIKE '%F9 v1.1%';

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

[24]:

AveragePayload
2534.6666666666665

Knowing the **average payload mass carried by booster version F9 v1.1** is important because it helps assess the typical mission capacity and operational efficiency of that specific booster configuration. It offers insights into:

- **The maturity and reliability** of F9 v1.1 in delivering moderate-weight missions.
- **Performance trends** across booster versions, aiding comparison with newer models like F9 Full Thrust or F9 Block 5.
- **Mission planning implications**, especially for payload mass expectations in future launches or contracts involving similar hardware.

First Successful Ground Landing Date

```
[25]: %%sql
SELECT MIN("Date") AS FirstGroundSuccess
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';

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

[25]:

FirstGroundSuccess
2015-12-22

The first successful ground landing marks a shift in space travel, moving from expendable first stage, to a reusable first stage. This is

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[22]:
```
%%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.
```

[22]:

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

Booster versions that successfully landed on a drone ship while carrying medium-to-heavy payloads (4000–6000 kg) are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2 highlighting SpaceX's capability to recover rockets even under challenging mission loads.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

[26]:
```
%%sql
SELECT "Mission_Outcome", COUNT(*) AS Count
FROM SPACEXTABLE
GROUP BY "Mission_Outcome";
```

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

[26]:

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

This result shows that the vast majority of SpaceX missions were successful, with only one confirmed in-flight failure and one mission with an unclear payload status—demonstrating a high reliability rate with over 98% confirmed mission success.

Boosters Carried Maximum Payload

This query identifies all SpaceX booster versions that carried the **heaviest recorded payload of 15,600 kg**. All of them are **Falcon 9 Block 5 (F9 B5)** boosters, demonstrating that this upgraded version consistently handled the most demanding missions in terms of payload mass.

Implications:

- It confirms the **payload capacity and operational reliability** of the Falcon 9 Block 5, SpaceX's most advanced and reusable booster.
- It highlights the **standardization and repeatability** of heavy-lift missions using different refurbished boosters.
- This capacity is crucial for launching **large satellites, crewed missions, or multiple payloads**, showcasing SpaceX's scalability in commercial and government missions.

[27]:
```
%%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.
```

[27]:

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

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

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

[28]:

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

This query shows that **in 2015**, there were **two failed drone ship landings**, both involving **F9 v1.1 boosters** launched from **CCAFS LC-40**, occurring in **January and April**.

Implications:

- These early failures highlight the **experimental phase** of SpaceX's drone ship recovery efforts.
- The use of **F9 v1.1**, an earlier booster version, suggests **technological limitations** at that time compared to later versions like F9 FT or F9 B5.
- These attempts laid the groundwork for **future success** in booster reusability, reflecting SpaceX’s **trial-and-error innovation strategy**.

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

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

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

[20]:

Landing_Outcome	OutcomeCount
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

This query summarizes the **count of each type of landing outcome** in the SpaceX dataset, a snapshot of the **success rates** during booster recovery.

Why this is important:

- helps **evaluate the effectiveness** of different recovery methods
- **high number of successful landings** on drone ships and ground pads demonstrate **progress toward reliable reusability**, a key goal in reducing launch costs.
- The presence of **failures and "no attempts"** reveals phases of **testing, risk mitigation, or mission constraints**, offering insights into the **evolution of SpaceX's recovery strategy**.

Section 3

Launch Sites Proximities Analysis

Global Locations of Launch Sites

Important Elements and Findings:

The map plots **four SpaceX launch sites**:

- **CCAFS LC-40**
- **CCAFS SLC-40**
- **KSC LC-39A**
- **VAFB SLC-4E**
  using their corresponding latitude and longitude coordinates.

A **distinct marker and circle** highlight the location of **NASA Johnson Space Center (JSC)** in Houston, Texas — representing the administrative and mission control hub.

This visualization provides a **spatial perspective** on where major launches occur and their **geographic relation to NASA JSC**.

Color-coded Launch Sites

Key Findings:

The **folium map displays launch site markers color-coded by mission outcome**:

- 🟢 **Green** = Successful landings
- 🔴 **Red** = Failed landings

The **KSC LC-39A** site (left) shows a higher cluster of **successful launches** (green markers), indicating strong reliability.

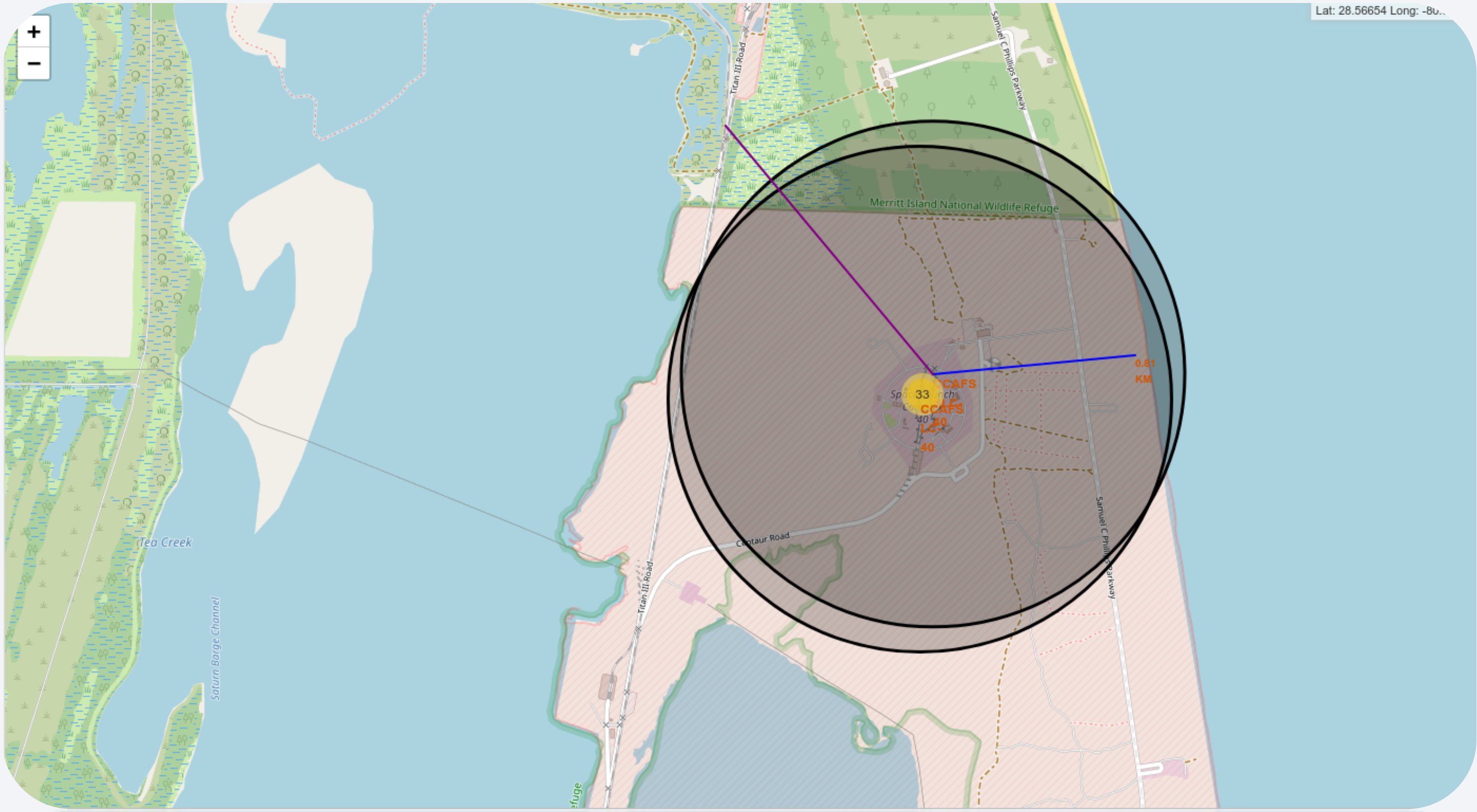
CCAFS LC-40 (right) has a **mixed outcome** distribution (green, red, orange), showing both successes and failures.

Launch Site distances to the Coastlines, Railroads, Cities

- **Closest city**: Titusville, Florida is the nearest city to both CCAFS LC-40 and KSC LC-39A, providing essential infrastructure and human resources for SpaceX operations.

- **Closest railroad**: The Florida East Coast Railway runs near the Kennedy Space Center, enabling heavy cargo and rocket transport by rail.

- **Closest coastline**: The Atlantic Ocean coastline is immediately adjacent to the launch pads, allowing safe eastward launches and easy maritime recovery of boosters.



Section 4

Build a Dashboard with Plotly Dash

Plotly DASH Dashboard

Launch Site Dropdown Menu

- Allow users filter data by specific launch sites or view all.
- Enables comparative analysis of launch performance across different sites.

Pie Chart

- Show launch success distribution.
- If "All Sites" is selected: display total successes by launch site.
- If one site is selected: compares success vs failures for that site.

Payload Range Slider

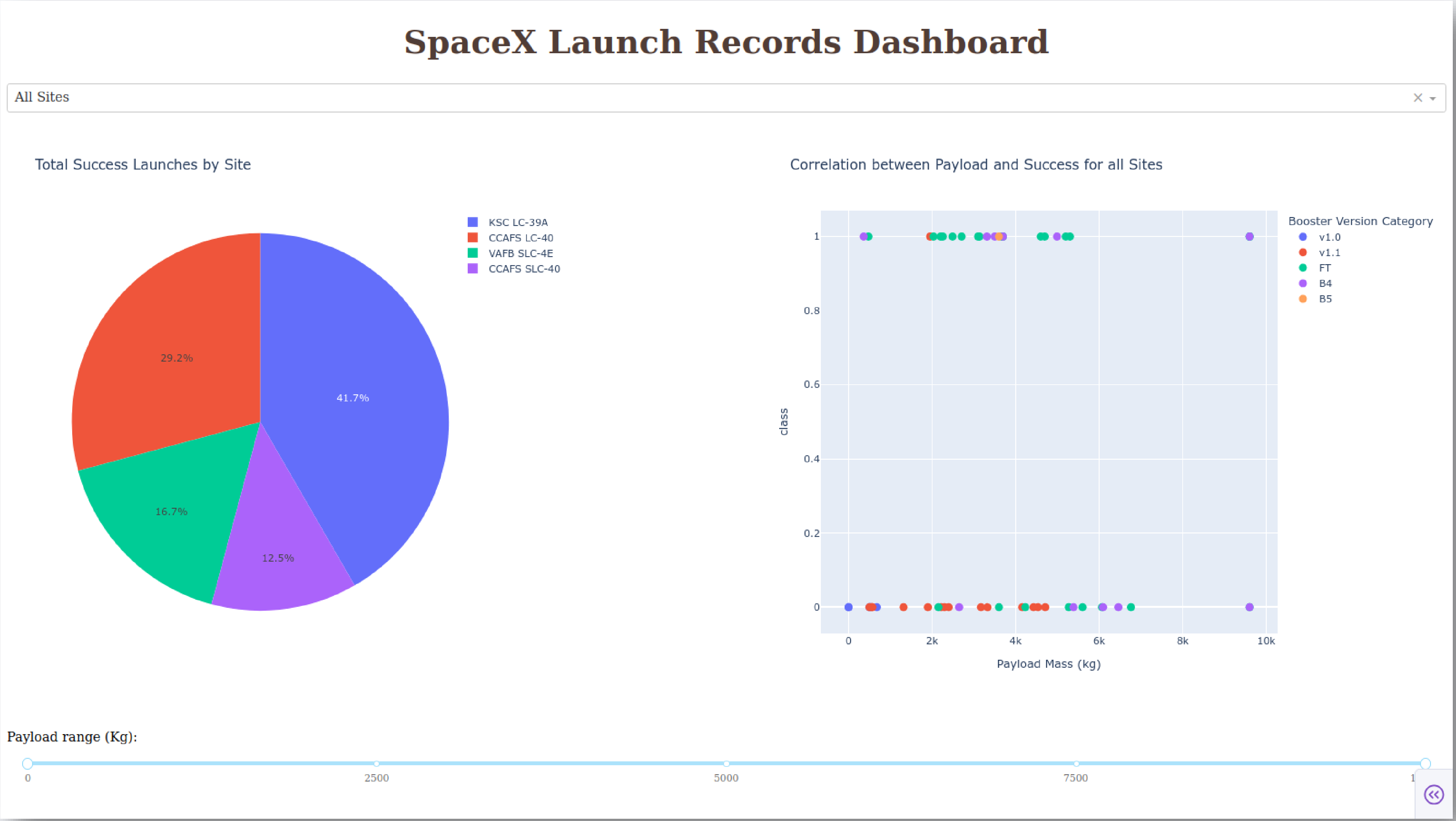
- Filters launches by payload mass (kg).
- Helps explore how payload weight affects launch success.

Scatter Plot

- Shows correlation between payload mass and launch success.
- Uses color to distinguish booster version categories.
- Updates dynamically based on dropdown and slider selections.

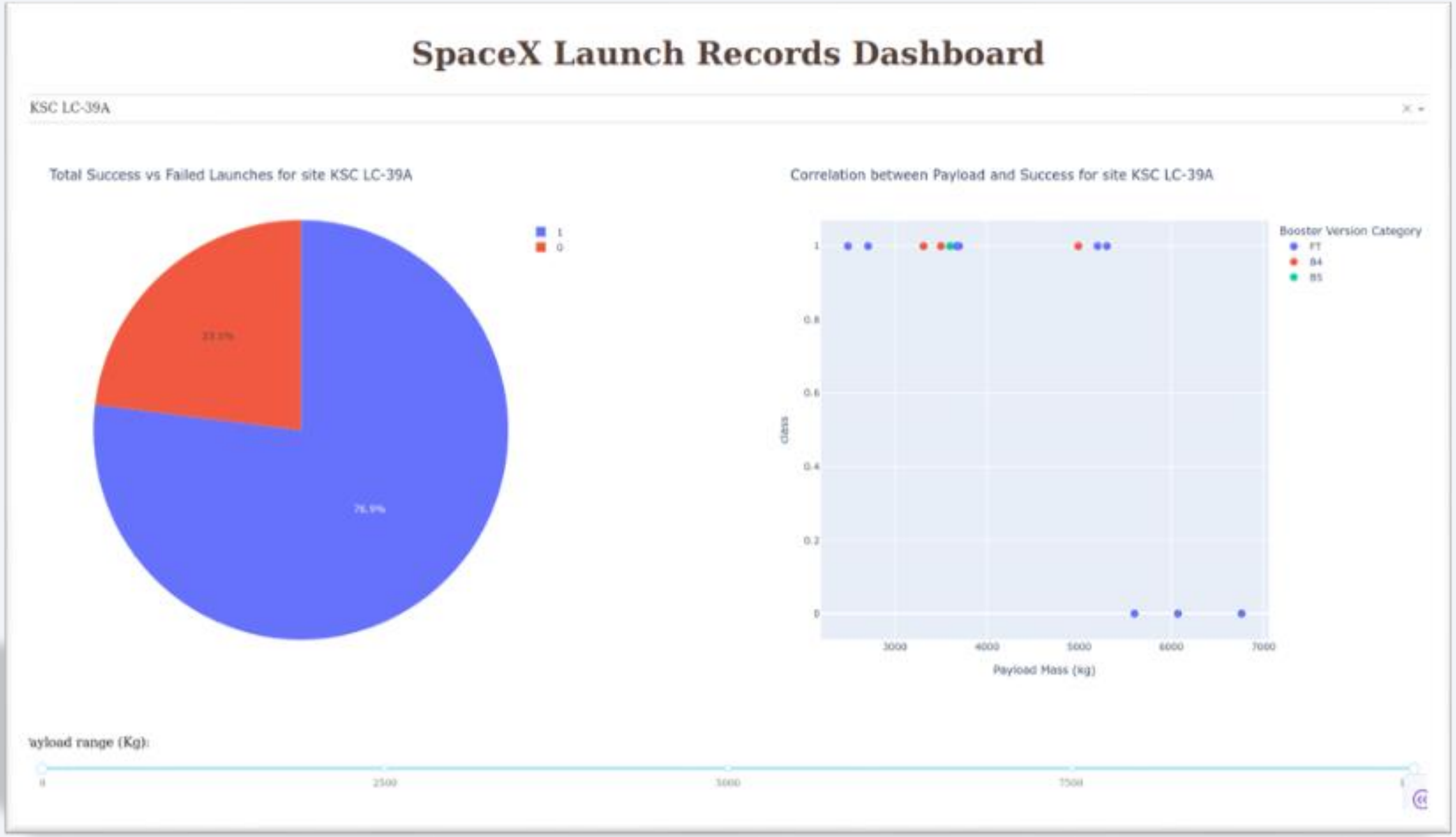
Responsive Layout

- Pie chart and scatter plot are displayed **side-by-side**, improving readability and user interaction.


SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
Correlation between Payload and Success for all Sites
Booster Version Category
v1.0
v1.1
FT
B4
B5
class
1
0.8
0.6
0.4
0.2
0
0
2k
4k
6k
8k
10k
Payload Mass (kg)
Payload range (Kg):
0
2500
5000
7500

Highest Launch Success

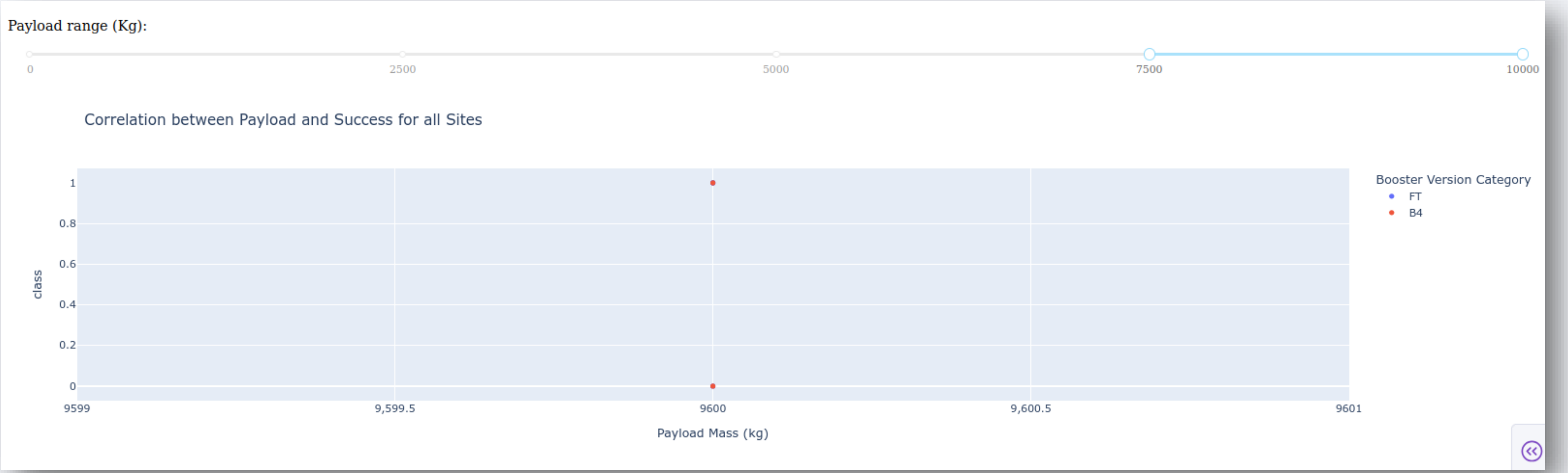
- The site with the highest successful launch is KSC LC-39A
- It comprises 41.7% of total successful launches (prev page)
- Successful launches in this site accounts to 76.9% and failed is 23.1%
- All booster types across all payload ranges contributed to successful launch while FT boosters correlated to most failed launches.



Payload Mass vs Successful Launch

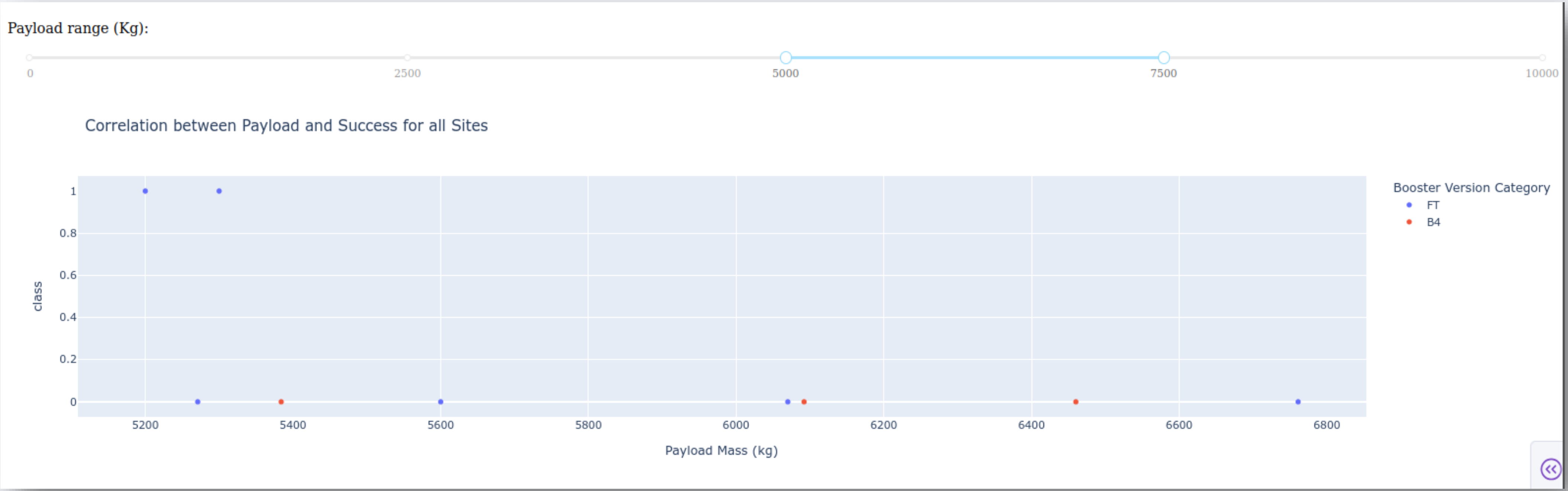
- This diagram shows the Payload mass vs Success for all sites, segmented by booster types represented by different dot colors. The x-axis shows the mass, the y-axis shows 0 for fail and 1 for success.
- The payload range selector can show the different masses of the payload and will show which boosters succeeded and failed for each mass in the range.

Payload Mass vs Successful Launch



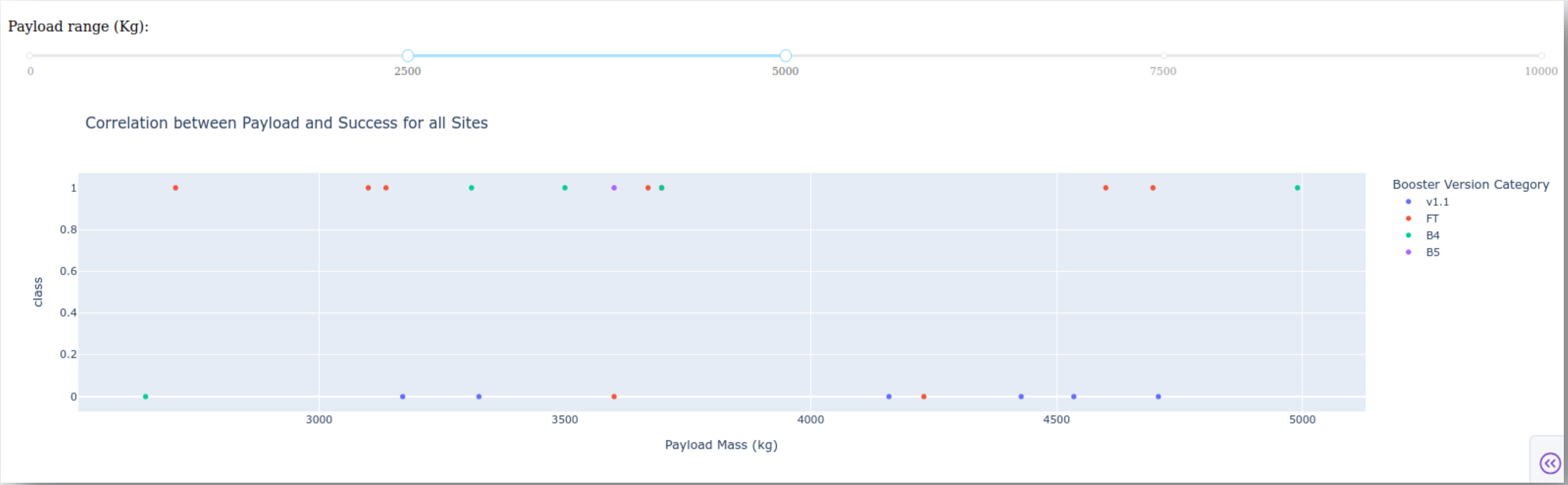
- For the payload range 7500 to 1000 Kg there are only two points, meaning there were two launches.
- The booster used for each launch is the B4 booster., carrying a payload of 9600 Kg.
- There were 1 failed launch and 1 successful launch.

Payload Mass vs Successful Launch



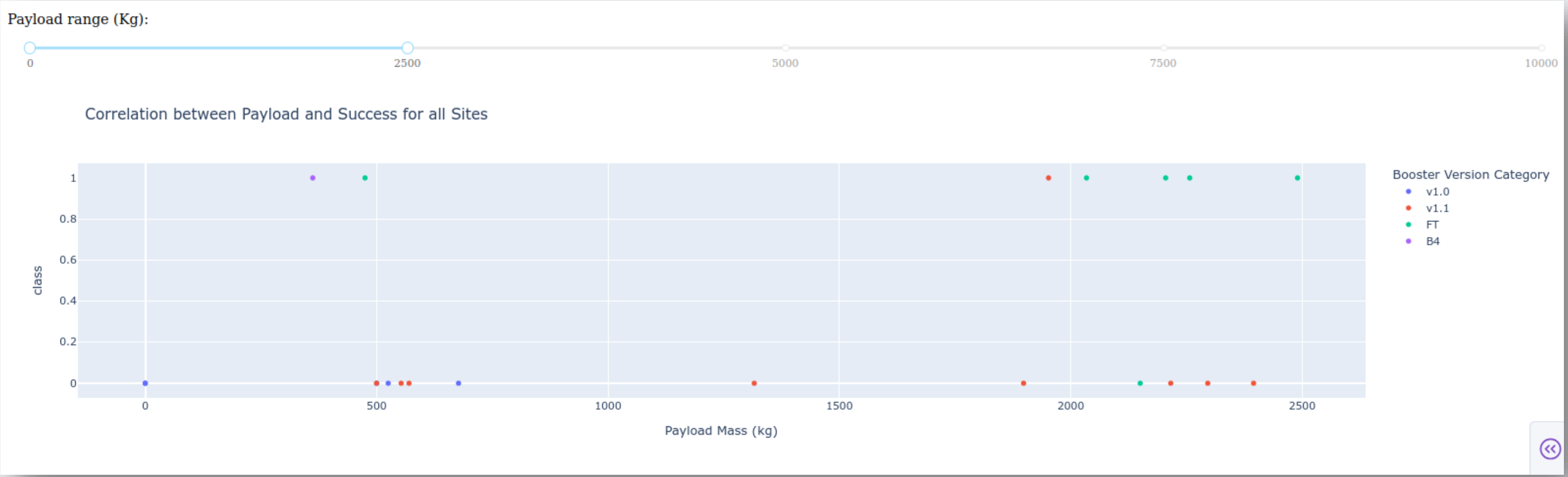
- For the payload range 5000 to 7500 Kg there are 9 points for 9 different launches.
- The booster used for the launches are FT for 6 launches and B4 for 3 launches.
- There are only two successful launches both carrying below 5400 Kg (one carrying 5200 Kg) both of which are FT boosters. The failed launches are distributed across the range with more fails above 5600.
- The failed launches are composed of 4 FT boosters and 3 B4 boosters.

Payload Mass vs Successful Launch



- For the payload range 2500 to 5000 Kg there are 20 points for 20 different launches, segmented into 4 colors for different boosters.
- The booster used for the launches are V1.1 for 6 launches, FT for 8 launches, B4 for 4 launches and B5 for 1 launch
- There are 6 successful FT launches 4 of which carried 3669 Kg and below and 2 carried 4600 and 4696 Kg respectively. Of the 4 successful launches of B4, 3 carried 3696.65 Kg and below and 1 carried 4990 Kg. There is 1 B5 success carrying 3600 Kg.
- There are 9 failed launches composed of 1 B4, 6 for V1.1, 2 for FT. Majority of successes (8 launches) carried below 4000 Kg and majority of fails (5 launches) carried payload above it.

Payload Mass vs Successful Launch



- In the payload range of 0 to 2500 Kg there are 7 successful and 12 failed launches, a total of 19 launches in all classified into 4 booster types.
- 1 V1.1 carrying 1952 Kg payload was a success and 8 fails with the lowest payload at 500 Kg and highest at 2395 Kg. Launches made with FT are 5 successes with 1 carrying 475 kg and 4 carrying above 2034 Kg and 1 fail at 2150 kg. Solitary B4 launch succeed at 475 Kg. All 3 launches of V1.1 failed including 1 with 0 Kg, and 2 below 600 Kg.
- Majority of fails (7 out of 12 launches) occurred below 1500 Kg and majority of succees (5 out of 7 launches) above 1500 Kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
print("Logistic Regression Test Accuracy:", logreg_cv.score(X_test, Y_test))
print("SVM Test Accuracy:", svm_cv.score(X_test, Y_test))
print("Decision Tree Test Accuracy:", tree_cv.score(X_test, Y_test))
print("KNN Test Accuracy:", knn_cv.score(X_test, Y_test))
```

```
Logistic Regression Test Accuracy: 0.8333333333333334
SVM Test Accuracy: 0.8333333333333334
Decision Tree Test Accuracy: 0.8333333333333334
KNN Test Accuracy: 0.8333333333333334
```

Final Model Comparison and Conclusion

After evaluating four classification models—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—on the test dataset, we obtained the following accuracies:

    Logistic Regression: 83.33%

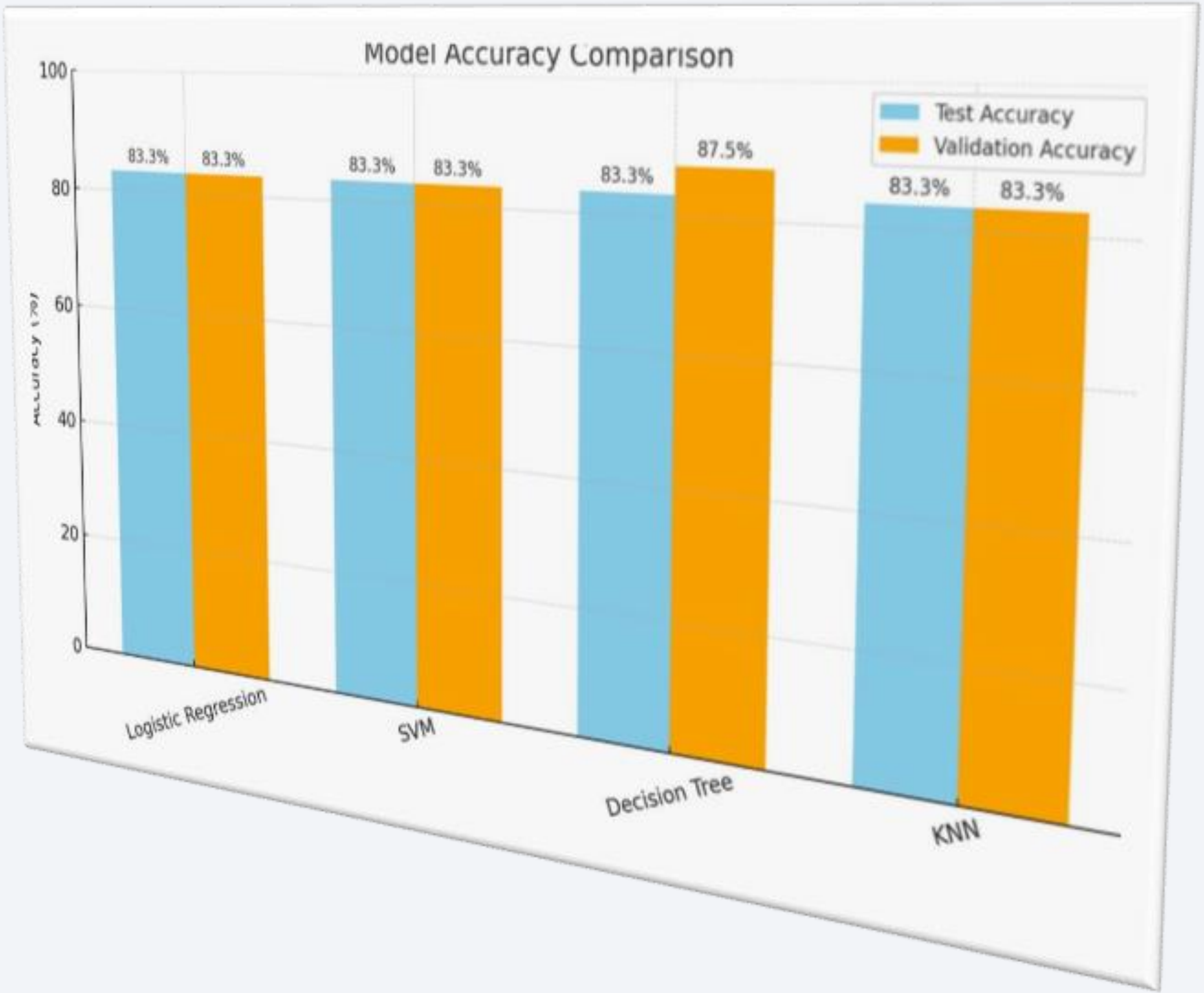
    SVM: 83.33%

    Decision Tree: 83.33%

    KNN: 83.33%

All four models achieved identical test accuracy scores of 83.33%, indicating consistent performance in classifying the test data. However, during cross-validation, the **Decision Tree** model achieved the highest validation accuracy at 87.5%, suggesting stronger generalization during model tuning.

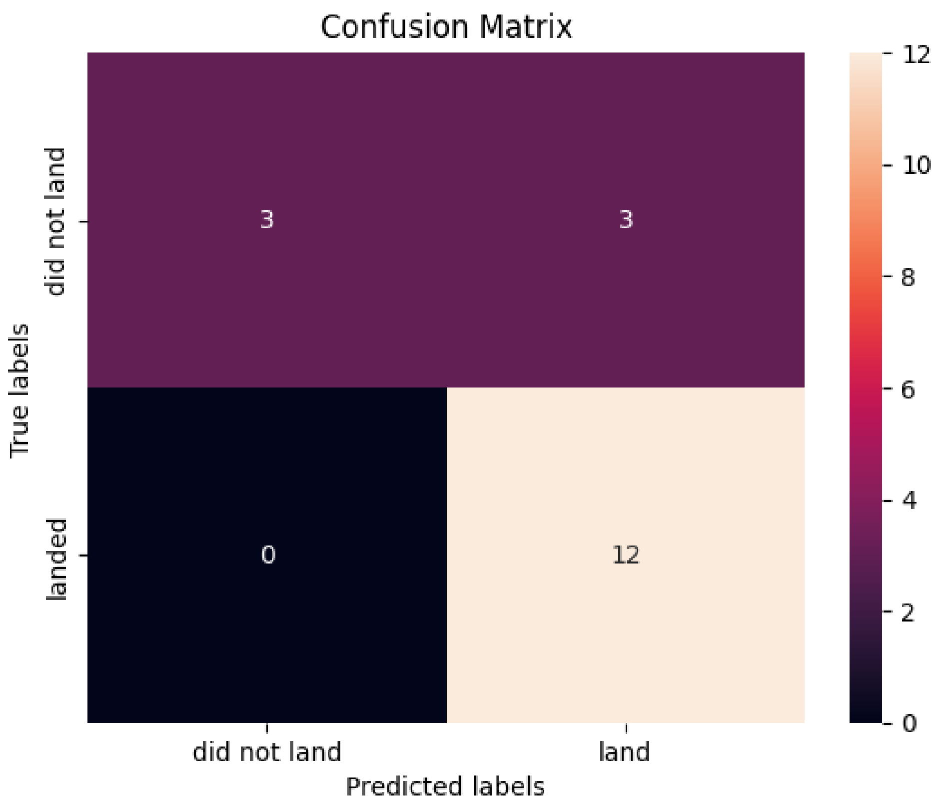
Given these results, all models are viable options for this classification task. If interpretability and speed are key considerations, **Logistic Regression** remains a practical choice. However, the **Decision Tree** may offer a slight edge when prioritizing validation performance and decision rule transparency.



Confusion Matrix

In [40]:

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



- The model **correctly predicted all 12 successful launches** (True Positives).
- It **misclassified 3 failed launches as successes** (False Positives).
- There were **no False Negatives** — the model never predicted a failure when the actual outcome was success.
- This shows the model is highly **sensitive to successful launches** (perfect recall for class = 1), though there's **some trade-off in precision** due to the false positives.

Conclusions

- **Launch Site Performance**: *KSC LC-39A* emerged as the top-performing site, contributing to the highest number of successful launches, with a success ratio of **76.9%**, highlighting it as a reliable hub for SpaceX missions.
- **Payload Insights**: Launches carrying **payloads above 1500 Kg** had a noticeably higher success rate, especially when using boosters like **FT and B5**, suggesting that SpaceX's technology handles heavier payloads more effectively than lighter ones.
- **Booster Evaluation**: While most booster types contributed to successes, **FT boosters** were linked to a larger share of failures in the mid-weight payload range (2500–7500 Kg), pointing to potential areas for further engineering analysis or version upgrades.
- **Predictive Modeling**: Our Decision Tree model achieved the highest validation accuracy at **87.5%**, correctly identifying all successful launches, making it a useful tool for **forecasting launch outcomes based on mission parameters**.

Appendix

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

In [20]:

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

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

Out[20]:

Landing_Outcome	OutcomeCount
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

Appendix

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

In [27]:

```
%%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.
```

Out[27]:

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

Appendix

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

In [28]:

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

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

Out[28]:

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

Thank you!