Winning Space Race with Data Science

CAMILLE FARRELL

AUGUST 3, 2025

Outline

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

Executive Summary

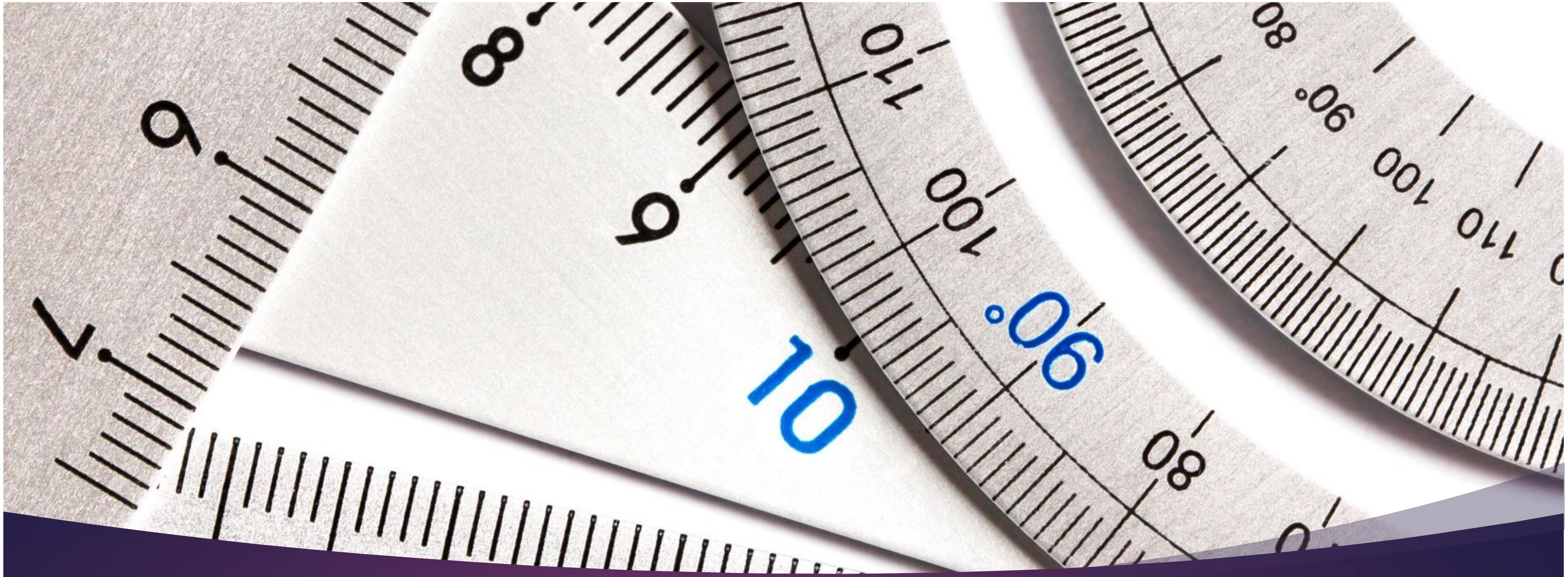
- SpaceX launch data was collected using the **SpaceX REST API** and **web scraping** (BeautifulSoup).
- **Exploratory data analysis** included interactive **Folium maps** showing launch outcomes by site, and **Plotly dashboards** for payload success rates and feature correlations.
- Four classification models were built—**Logistic Regression, SVM, Decision Tree, and KNN**—and fine-tuned using **GridSearchCV** with 10-fold cross-validation.
- The **Decision Tree** achieved the **highest CV accuracy**, though **test accuracy was consistent (~83%)** across all models.
- Confusion matrices highlighted false positives as a key area for improvement.
- Key predictors of launch success included **payload mass** and **launch site**.

Introduction

Our competitor SpaceX is unique in the industry where it can recover and reuse one of its rockets (Falcon9) if it successfully lands.

Our company would like to compete with SpaceX and we'd like to determine the price of each launch. Using SpaceX's Falcon9 data, we're able to estimate the price of each launch along with determining if SpaceX can reuse the first stage.

This presentation is the first step of a further study that analyses Falcon9's launch data and success rate in landing, leveraging various data sources, data visualization, and regression analysis.



Section 1 - Methodology

Methodology

Data Collection (SpaceX REST API and web scrapping wiki pages)

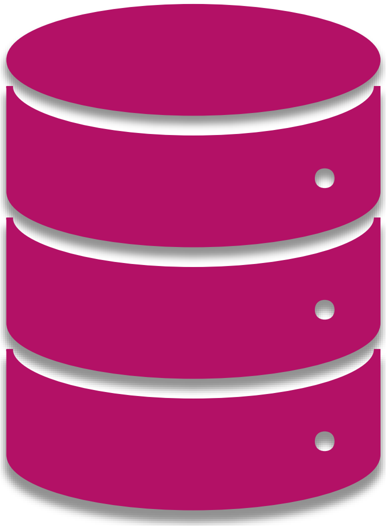
Perform Data Wrangling (data cleaning, missing variables treatment)

Perform Exploratory Data Analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium of launch sites, success rates, along with Plotly Dash for interactive visualization

Perform predictive analysis using classification models (Logistic regression, SVM, Decision Tree, and KNN)

Data Collection



- **1. SpaceX Data Collection via REST API**
  - Utilizes SpaceX's public REST API to programmatically retrieve structured launch data.
  - Extracts key information such as launch dates, sites, payloads, and outcomes.
  - Data is cleaned and organized into a DataFrame for analysis.
  - Enables automated, reproducible, and up-to-date data retrieval.
- **2. SpaceX Web Scraping with BeautifulSoup**
  - Scrapes additional unstructured data directly from the SpaceX website.
  - Parses HTML content to extract supplementary information not available via the API.
  - Employs BeautifulSoup library to navigate and extract relevant webpage elements.
  - Complements API data to enrich the dataset for comprehensive analysis.

Data Collection - SpaceX

This process depicts the collection of SpaceX launch data by querying their public REST API. The raw JSON response is retrieved, parsed, and transformed into a clean, structured dataset containing key launch details. This enables efficient and reproducible data extraction for further analysis.

Define API Endpoint URL
Set the SpaceX API URL to fetch launch data (https://api.spacexdata.com/v3/launches).

Send GET Request to API
Use requests.get() to request launch data from SpaceX's API.

Check API Response
Verify that the API call was successful (HTTP status code 200).

Parse JSON Data
Convert the JSON response into a Python list/dictionary.

Normalize JSON Data to Tabular Format
Use pandas.json_normalize() to flatten nested JSON into a DataFrame.

Select Relevant Columns
Extract key features such as launch date, mission name, launch site, rocket details, payload info, and launch success.

Data Cleaning & Formatting
Handle missing values, format dates, and rename columns for clarity.

Save or Prepare Data for Analysis
Store the cleaned data in a DataFrame for downstream analysis or export it as CSV.

Data Collection – Web Scrapping

This process collects additional SpaceX data by scraping information directly from their website. It uses web scraping techniques with BeautifulSoup to extract and structure relevant details not accessible through the API, enriching the overall dataset.

Import Libraries	• Load libraries such as requests for HTTP calls and BeautifulSoup for HTML parsing.
Define Target URL	• Specify the SpaceX webpage URL containing the desired data.
Send HTTP GET Request	• Retrieve the raw HTML content from the webpage.
Parse HTML Content	• Use BeautifulSoup to parse and navigate the HTML DOM.
Locate Data Elements	• Identify and extract specific HTML elements (e.g., tables, divs, spans) containing target data.
Extract Text and Attributes	• Clean and extract the textual content or attributes from the HTML elements.
Organize Data into Structured Format	• Compile extracted data into lists or DataFrames for easier analysis.
Data Cleaning	• Perform necessary cleaning such as trimming whitespace and handling missing values.
Prepare Data for Analysis	• Finalize the dataset for merging with API data or direct analysis.

Data Wrangling

This process focuses on cleaning, transforming, and preparing the SpaceX launch dataset for analysis. It applies various data wrangling techniques to handle missing data, convert formats, and enrich the dataset with calculated fields, ensuring a clean and consistent foundation for further exploration and modeling.

Load Data: Import the raw SpaceX launch data collected from previous steps (API and/or web scraping) into a pandas DataFrame.

Inspect Dataset: Perform initial exploration such as viewing summary statistics, data types, and checking for missing or inconsistent values.

Handle Missing Data: Identify missing or null values and apply appropriate techniques, such as filling with default values or dropping irrelevant rows.

Convert Data Types: Standardize data types for columns (e.g., convert date strings to datetime objects, numerical fields to numeric types) for proper analysis.

Rename Columns: Rename columns for clarity and consistency across the dataset.

Create New Features: Engineer new variables such as calculating launch success/failure flags, extracting year/month from dates, or grouping launch sites.

Filter and Subset Data: Remove irrelevant columns or filter rows to focus on specific subsets (e.g., only successful launches or a certain time period).

Handle Categorical Data: Process categorical variables by encoding or categorizing launch outcomes, rocket types, or payload categories.

Validate Cleaned Data: Re-check data quality, verify transformations, and ensure data integrity before saving or using for analysis.

Save Cleaned Dataset: Export the cleaned and wrangled data for visualization, modeling, or reporting.

Exploratory Data Analysis with Data Visualization

Below is summary of charts that were produced to provide additional insights

Scatter	Scatter plot of launch outcome, flight Number, and pay Load Mass (kg) – to see the correlation between outcome and flight number and pay load mass
Scatter	Scatter of plot of launch outcome, flight number and launch site – to visualize the correlation between success rate, and launch site
Scatter	Scatter plot of launch outcome, launch site, and pay load mass – to visualize the correlation between success rate, launch site, along with payload mass
Type	Bar chart of success rate and orbit type – to visualize relationship between success rate and orbit type
Scatter	Scatter plot of launch outcome, orbit, and flight number – to visualize the relationship between Orbit, Flight number and overlayed with outcome
Scatter	Scatter plot of launch outcome, orbit, and pay load mass – to visualize how success rates vary for different combinations of orbit and pay load mass.
Line	Line chart of average success rate yearly trend - to determine how the average success rate has varied over the years

Exploratory Data Analysis with SQL – SQL queries performed

Below is a list of SQL queries that were performed:

- Retrieve a list of distinct launch sites with no filters applied.
- List launch sites that begin with "CCA", limited to 5 results by filtering for launch site names starting with 'CCA'.
- Calculate the total payload mass for missions where the customer is NASA (CRS).
- Compute the average payload mass for missions using the F9 v1.0 booster version.
- Find missions launched from 'CCAFS LC-40' with payloads over 4000 kg that successfully landed on a drone ship.
- (Similar to above) List mission details with the same filters: 'CCAFS LC-40', payloads over 4000 kg, and successful drone ship landings.
- (Similar to above) Another variation of the same filtered mission query.
- Find the maximum payload mass among missions with successful drone ship landings.
- Retrieve booster version and landing outcome for missions using the F9 B4 booster version through a join query.
- Count the number of successful drone ship landings, grouped by booster version.

Building an Interactive Map with Folium

This notebook visualizes SpaceX launch site locations and their proximity to nearby facilities using interactive Folium maps. It enriches spatial understanding by overlaying custom map elements to analyze geography-based launch factors.

- **Map Initialization**
  A base Folium map is created and centered over the continental United States using a default tile layer.
- **Launch Site Markers**
  Custom markers (with popup labels) are placed at known SpaceX launch site coordinates to identify each location visually.
- **Launch Site Circles**
  Circles are drawn around each launch site to emphasize its area and make locations easier to see at different zoom levels.
- **Proximity Markers**
  Additional markers are added for nearby facilities such as highways, railways, and coastlines using manually input coordinates. These represent logistical proximity features relevant to launch site suitability.
- **Popups and Labels**
  Each marker includes a popup label describing what the location represents (e.g., "Launch Site: CCAFS LC-40" or "Nearest Railway").
- **Distance Lines (Optional Section)**
  Polylines (straight lines) may be drawn between launch sites and nearby facilities to visually show distance or directional relationships.

Building a Dashboard with Plotly Dash

- We developed an interactive dashboard using Plotly Dash to visualize and explore SpaceX launch performance. It incorporates dynamic charts and user-driven controls to support deeper insights into launch outcomes by site and payload.
- **Summary of Plots, Graphs, and Interactions**
  - **Dropdown Filter: Launch Site Selector**
    Enables the user to filter the dashboard by individual launch site or view all sites combined.
  - **Pie Chart: Total Success Launches by Site**
    - When "All Sites" is selected: displays the share of successful launches across all launch sites.
    - When a specific site is selected: shows the count of successful vs. failed launches at that site.
  - **Payload Range Slider**
    Allows users to select a payload mass range (in kg) to filter the scatter plot dynamically.
  - **Scatter Plot: Payload vs. Launch Outcome**
    Displays the correlation between payload mass and launch success/failure.
    - Points are color-coded by success status.
    - Hover data reveals payload and booster information.
    - Updates based on both the selected site and payload range.

Predictive Analysis - Classification

This notebook applies supervised machine learning to predict the success of SpaceX Falcon 9 first-stage landings. The goal is to develop a classification model using historical launch data and engineered features to assess the likelihood of a successful landing based on mission parameters.

The **Decision Tree classifier** yielded the highest cross-validation accuracy.
All models performed similarly on the test set, but the Decision Tree showed better generalization and interpretability, making it the preferred model for predicting landing success.

Data Prep: cleaned the data, extracted features and target, standardized features

Data Split: Training (80%) and testing (20%) sets with fixed random seed

Modeling: Built logistic regression, SVM, Decision Tree, and KNN models. Used GridSearchCV (10-fold CV) for hyperparameter tuning

Evaluation: Compared models using CV accuracy, test accuracy, confusion matrices, and classification reports

Results: Models showed similar strong performance on this dataset

Model Selection: Tree model seemed to yield higher best score, although the margin between other alternatives was minimal

Next steps: Explore ensemble methods and feature engineering for further improvements.

Results and Insights

Data Exploration & Visualization

Majority of **successful launches** occurred at **CCAFS LC-40** and **KSC LC-39A**.

Payload mass ranged widely but showed no strong linear relationship with success.

Coastal proximity and access to **infrastructure (rail, roads)** observed via Folium maps suggest logistical considerations in site selection.

Interactive Dashboard Insights

Launch success rates vary by site, with some sites (e.g., VAFB) showing higher failure rates.

The **scatter plot** indicates slightly **higher success rates for medium-weight payloads** (2000–8000 kg), though not conclusively.

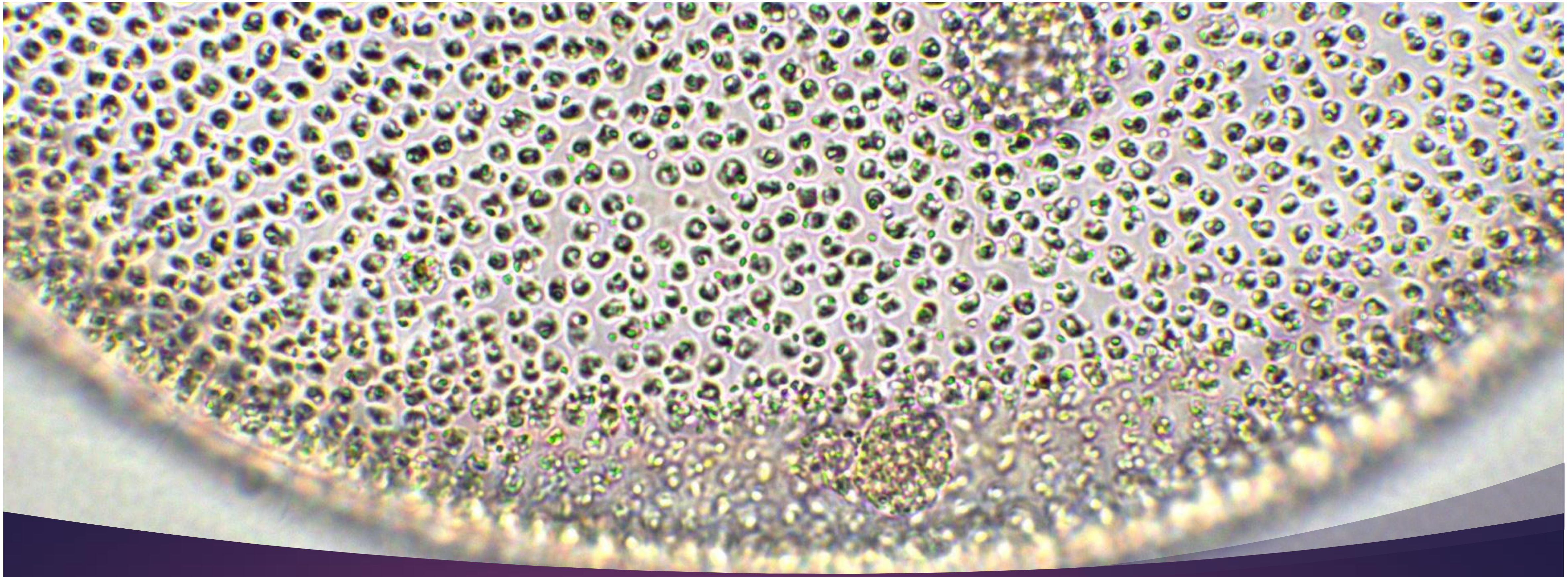
Dashboard filters allowed exploration by **site and payload**, revealing nuanced patterns.

Machine Learning Results

Decision Tree model had the **best cross-validation score**, indicating strong predictive potential.

No model vastly outperformed others on the test set, suggesting **complexity of predicting launch success**.

Key predictive features included **launch site**, **payload mass**, and **booster version category**.



Section 2 – Insights drawn from EDA

SQL Queries – Launch Site Names

- Below are the list of all launch site names, along with the SQL query used to generate the results:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

- SQL Query:

```
%%sql
SELECT DISTINCT LAUNCH_SITE
FROM SPACEXTBL;
```

SQL Queries – Records with Launch Sites that begin with "CCA"

Below are the list of 5 records where launch sites begin with "CCA", along with the SQL query used to generate the results

SQL Query:

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

Date	Time (UTC)	Booster Ver.	Launch Site	PayLoad	PayLoad Mass	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

SQL Queries – Total Payload Mass from NASA

- The total payload carried by boosters from NASA is **45,596KG**.
- The query used to generate that result is the following and it calculates the total payload mass (in kilograms) carried by SpaceX missions where the customer is "NASA (CRS)."

- SQL Query:

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

SQL Queries – Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 340.4.
- The following SQL query calculates the average payload mass for all records in the table where the `Booster_Version` starts with the string `'F9 v1.0'`.

- SQL Query:

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) as
AVGPAYLOADMASS
FROM SPACEXTBL
WHERE Booster_Version like 'F9 v1.0%';
```

SQL Queries – First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad is 2015-12-22.
- The following SQL query retrieves the earliest date from the table where the Landing_Outcome is 'Success (ground pad)'`.

- SQL Query:

```
%%sql
Select MIN(Date) as FIRST_DATE
from SPACEXTBL
where Landing_Outcome = 'Success (ground
pad)';
```

SQL Queries – Successful Drone Ship Landing with Payload between 4000 and 6000 KG

- Below is a list of the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:
- F9 FT B1021.1
- F9 FT B1022
- F9 FT B1023.1
- F9 FT B1026
- F9 FT B1029.1
- F9 FT B1021.2
- F9 FT B1029.2
- F9 FT B1036.1
- F9 FT B1038.1
- F9 B4 B1041.1
- F9 FT B1031.2
- F9 B4 B1042.1
- F9 B4 B1045.1
- F9 B5 B1046.1

The following SQL query retrieves a list of unique booster versions from the `SPACEXTBL` table that meet two specific criteria:
1. **Landing Outcome:** The booster must have successfully landed on a drone ship (`'Success (drone ship)'`).
2. **Payload Mass**: The payload mass carried by the booster must be greater than 4000 kg and less than 6000 kg.

The `DISTINCT` keyword ensures that only unique booster versions are returned, avoiding duplicates in the result.

```
%%sql
SELECT DISTINCT
BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING_OUTCOME =
'Success (drone ship)'
and
4000<PAYLOAD_MASS__KG_<6000;
```

SQL Queries – Total Number of Successful and Failure Mission Outcomes

- Below is a total number of successful and failure mission outcomes

Mission Outcome	Total Number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

This SQL query retrieves the count of each unique mission outcome from the `SPACEXTBL` table.

- `MISSION_OUTCOME`: Represents the result of each mission (e.g., success or failure).
- `COUNT(MISSION_OUTCOME)`: Counts how many times each mission outcome appears in the table.
- `GROUP BY MISSION_OUTCOME`: Groups the data by each unique mission outcome to calculate the count for each group

```
%%sql
SELECT MISSION_OUTCOME,
COUNT(MISSION_OUTCOME) as TOTAL_NUMBER
from SPACEXTBL
group by MISSION_OUTCOME;
```

SQL Queries – Booster Carried Maximum Payload

- Here are the names of the boosters which have carried the maximum payload mass:

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

- The following query retrieves the distinct booster versions that carried the maximum payload mass from the `SPACEXTBL` table.
  - The inner query (`SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL`) finds the maximum payload mass in the table.
  - The outer query filters rows where the payload mass matches this maximum value and selects the distinct booster versions associated with those rows.

```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
SELECT MAX(PAYLOAD_MASS__KG_)
FROM SPACEXTBL);
```

SQL Queries – 2015 Failed Launch Records

- Below are the failed landing outcomes in drone ship, their booster versions, and launch site names that occurred in 2015:

Landing Outcome	Booster Version	Launch Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- The following SQL query retrieves specific columns (`LANDING_OUTCOME`, `BOOSTER_VERSION`, and `LAUNCH_SITE`) from the `SPACEXTBL` table. It filters the results to include only rows where:

- The `LANDING_OUTCOME` is `'Failure (drone ship)'`.
- The year extracted from the `DATE` column is `'2015'` (using the `strftime('%Y', DATE)` function to extract the year).

```
%%sql
SELECT LANDING_OUTCOME,
BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Failure
(drone ship)'
AND strftime('%Y', DATE) = '2015';
```

SQL Queries – Ranking Landing Outcome between June 4th 2010 and March 20th, 2017

- Table below shows the landing outcomes between the specified dates:

Landing Outcome	Total Number
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

- The following query retrieves the count of each unique LANDING_OUTCOME for SpaceX missions that occurred between the dates '2010-06-04' and '2017-03-20'. It groups the results by LANDING_OUTCOME, calculates the total count for each outcome, and orders the results in descending order of the count.

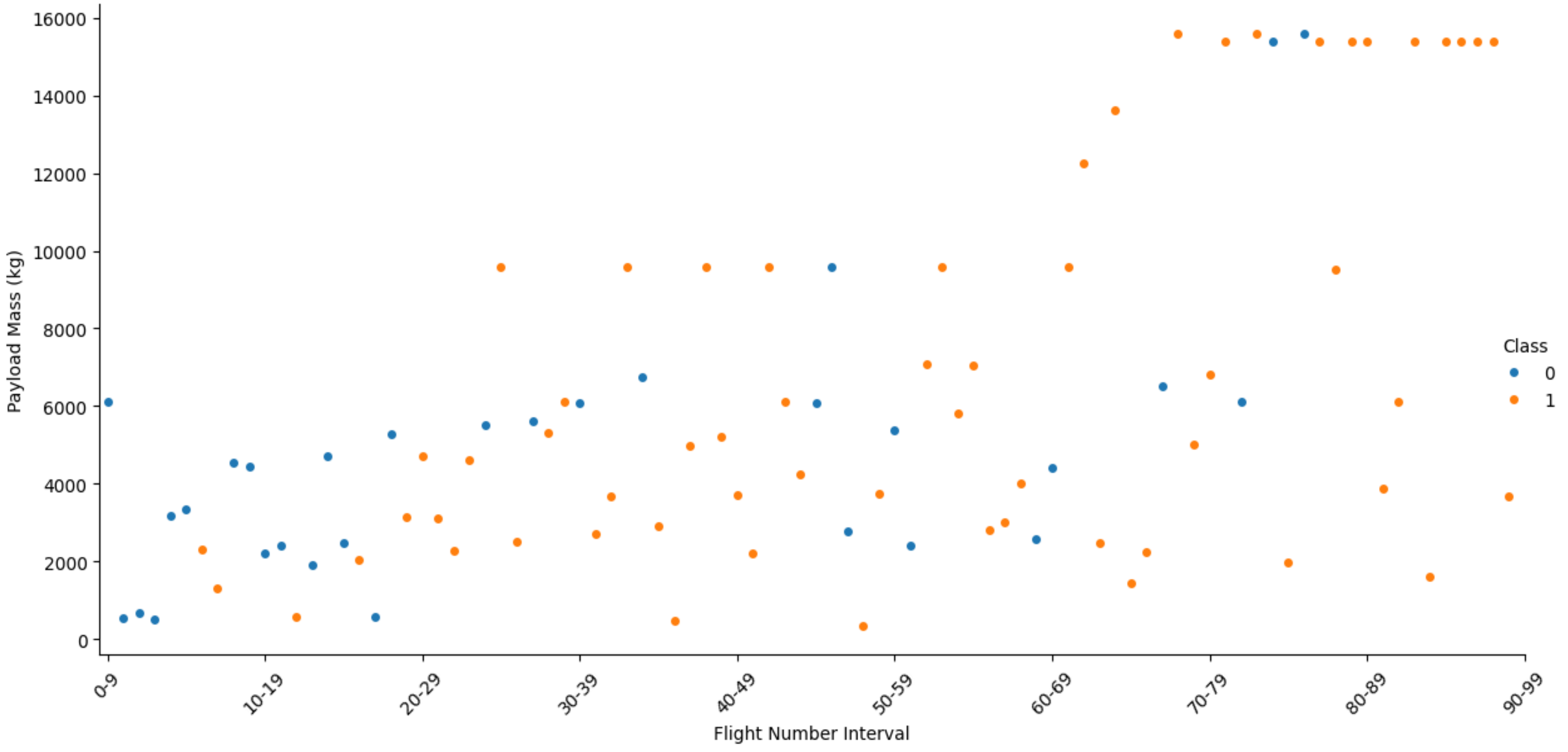
```
%%sql
SELECT LANDING_OUTCOME,
COUNT(LANDING_OUTCOME) AS
TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND
'2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

Visualization – Flight Number vs. Payload Mass

This plot visualizes the relationship between **payload mass** and **flight number**, with launch **success (Class = 1)** and **failure (Class = 0)** differentiated by color. Each point represents a SpaceX launch, allowing us to observe how **payload mass varied over time** and whether it influenced launch outcomes.

Key Findings

- **Payload mass varies widely across flights**, indicating a broad range of mission types and payload capacities over time.
- Earlier flights tend to have **lower payload masses**, suggesting initial testing and smaller payloads in early launches.
- From mid to later flights, there is a noticeable increase in the frequency of **higher payload masses**, reflecting SpaceX's growing capability and confidence with heavier payloads.
- There is **no clear direct correlation between payload mass and launch success**, as successful and failed launches appear across the entire payload range.
- Payloads clustered in the mid-range (around 2000–8000 kg) show a dense distribution of both successes and failures, indicating complexity in this payload segment.



Visualization – Flight Number vs. Launch Site

This plot displays the **launch success (Class = 1) and failure (Class = 0)** outcomes across different **launch sites**, plotted against **Flight Number intervals**. Each point represents a specific flight from a specific site, color-coded by outcome.

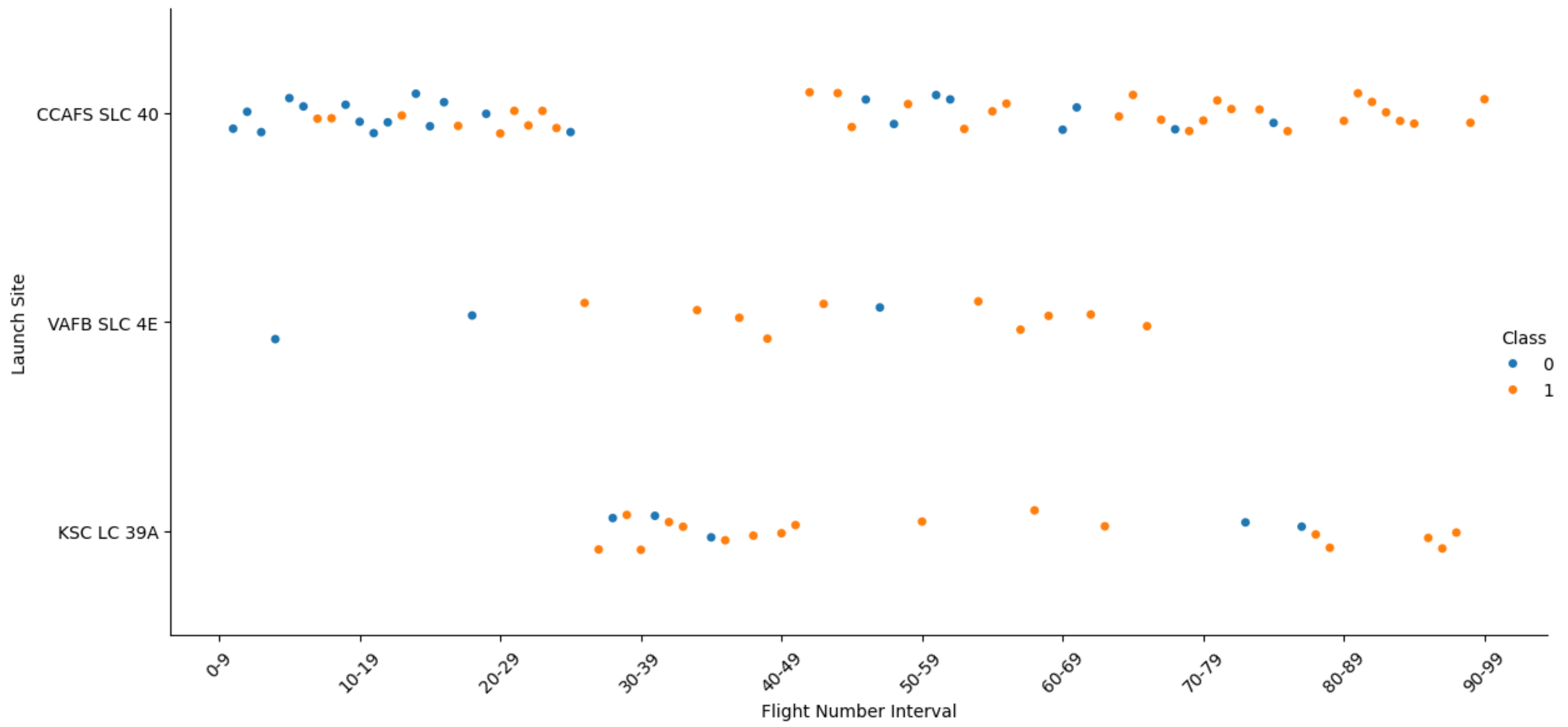
Key Findings

- **CCAFS SLC 40**:
This launch site has a mix of successful and unsuccessful launches.
As the flight number increases, the success rate improves, suggesting that SpaceX's experience and improvements over time have positively impacted launch outcomes.

- **VAFB SLC 4E**:
This site has fewer launches compared to CCAFS SLC 40.
The launch site shows increased success rate, especially amongst higher flight numbers

- **KSC LC 39A**:
This site shows a higher success rate even for lower flight numbers.
It suggests that this site might be used for more critical or well-prepared missions, leading to better outcomes.

- Overall, the plot indicates that SpaceX's success rate improves with experience (higher flight numbers) and that launch site characteristics or mission types may influence outcomes.

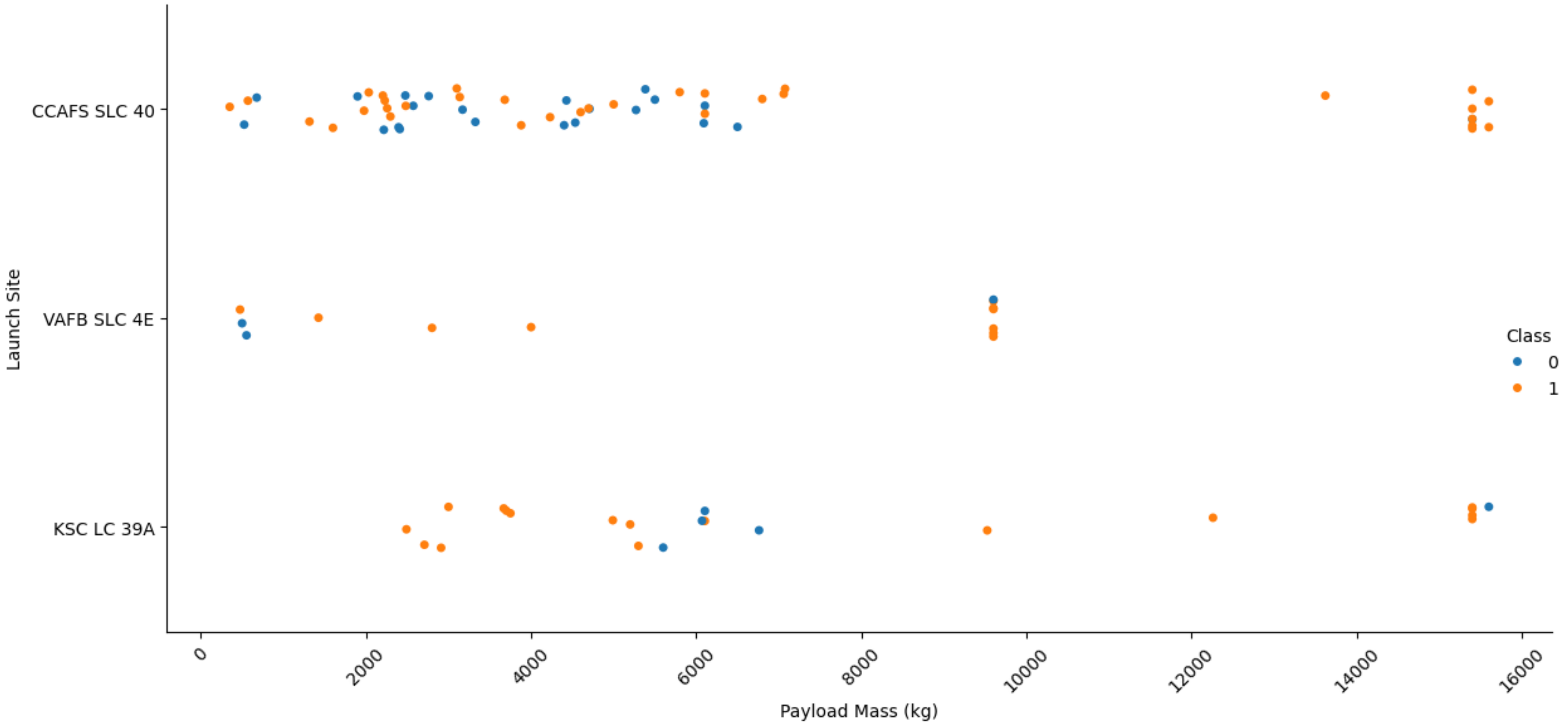


Visualization – Payload vs. Launch Site

This chart shows the distribution of **payload masses** for SpaceX launches across different **launch sites**. Launch outcomes are color-coded by **success (Class = 1)** or **failure (Class = 0)**, allowing us to see how payload mass varies by site and how it relates to launch success.

Key Findings

- **KSC LC-39A** and **CCAFS LC-40** handle the widest range of payload masses, including some of the heaviest launches.
- **VAFB SLC-4E** generally manages lighter payloads, with fewer total launches.
- Both successful and failed launches occur across nearly all payload masses, indicating no strict payload threshold for success or failure.
- The spread of payload masses at each site reflects the diversity of mission profiles and vehicle capabilities.

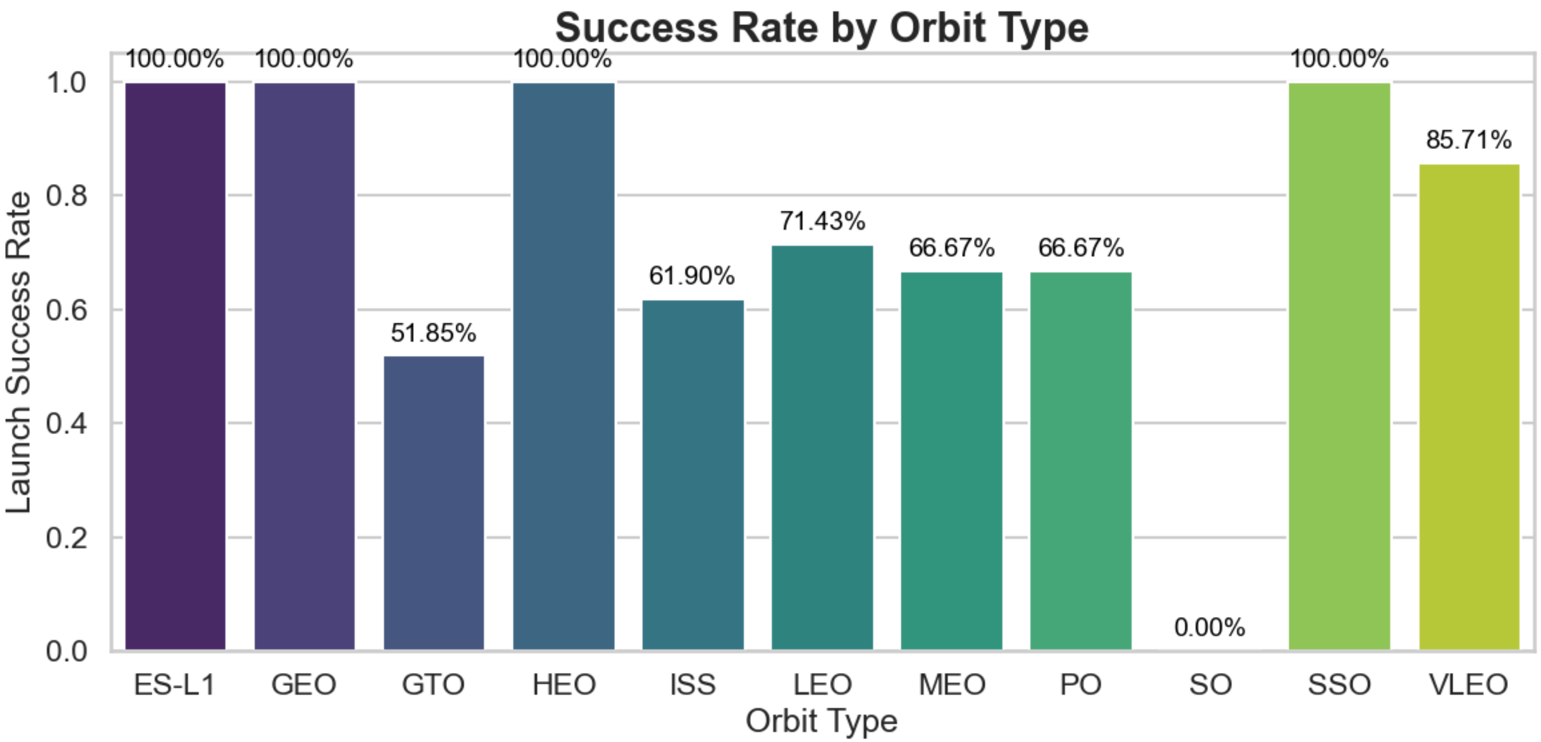


Visualization – Success Rate vs Orbit Type

This bar chart displays the **average launch success rate** for SpaceX missions grouped by **orbit type**. It helps identify which orbits have historically seen higher or lower success rates.

Key Findings

- **LEO (Low Earth Orbit)** and **GTO (Geostationary Transfer Orbit)** have the highest success rates, reflecting their frequent use and mission maturity.
- Orbits like **SSO (Sun-Synchronous Orbit)** and **Polar** show slightly lower success rates, potentially due to more challenging mission profiles.
- Understanding orbit-specific success trends can guide risk assessment and operational planning for future launches.

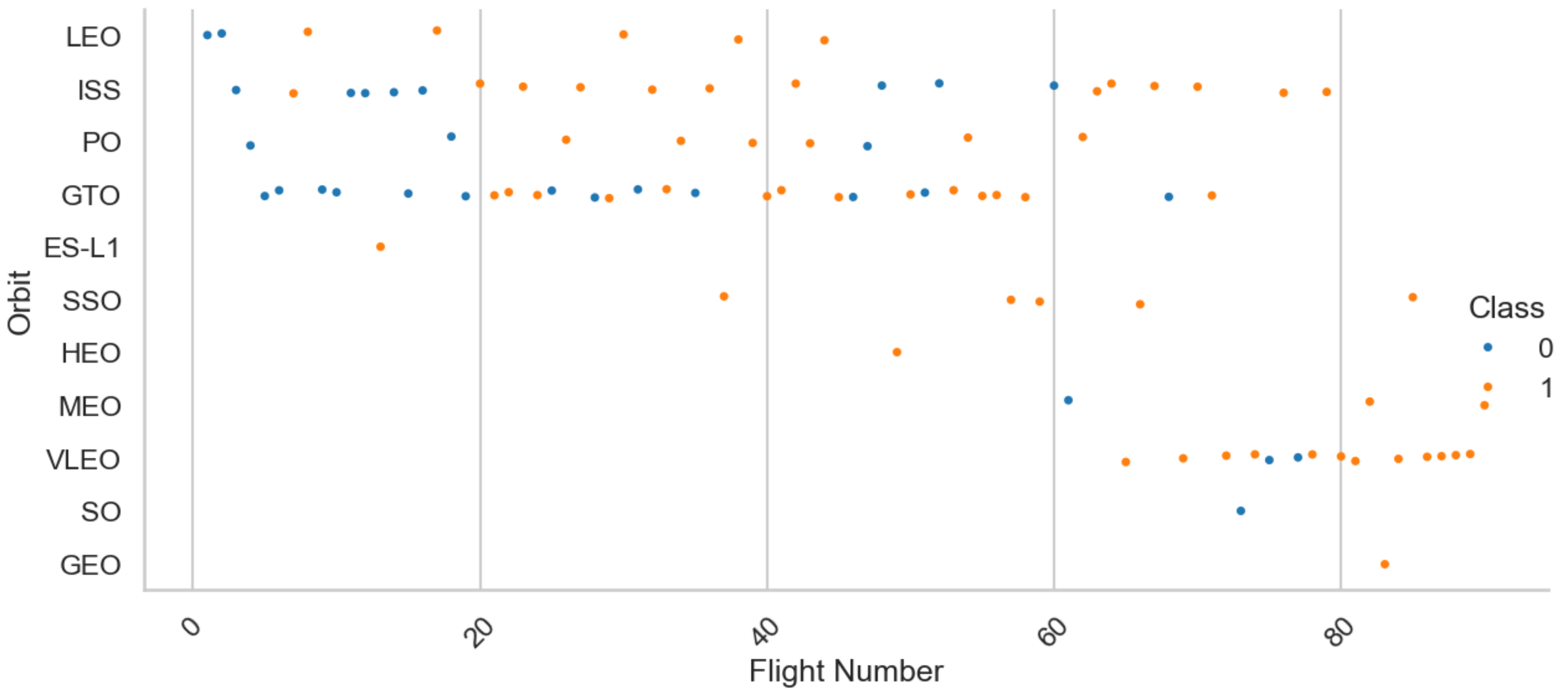


Visualization – Flight Number vs Orbit Type

This plot visualizes SpaceX launches over time (by flight number), showing the orbit types for each mission. Launch outcomes are indicated by color, allowing observation of success and failure distribution across different orbits throughout the program.

Key Findings

- **Low Earth Orbit (LEO)** missions dominate early flight numbers and show a high success rate.
- **Geostationary Transfer Orbit (GTO)** missions appear more frequently as the program matures, also showing mostly successful launches.
- Orbits like **Polar** and **Sun-Synchronous Orbit (SSO)** have fewer launches with mixed success outcomes, indicating more specialized and challenging missions.
- The plot highlights how SpaceX expanded mission diversity over time while maintaining generally strong success rates across orbits.

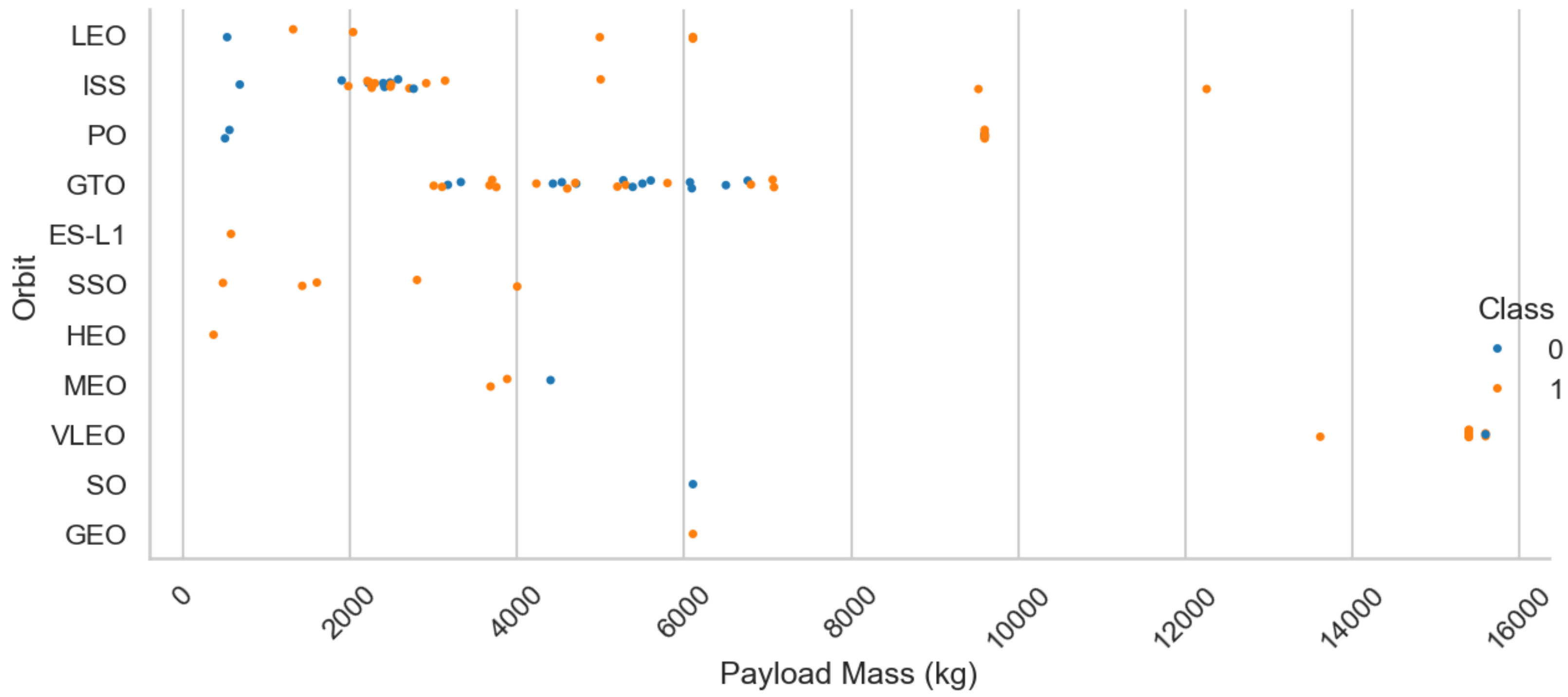


Visualization – Payload vs Orbit Type

This plot illustrates the relationship between **payload mass** and **orbit type** for SpaceX launches, with launch outcomes color-coded by success or failure. It highlights how payload sizes vary across different orbits and their corresponding mission success rates.

Key Insights

- **Low Earth Orbit (LEO)** missions cover a wide range of payload masses, from light to heavy, with a majority being successful.
- **Geostationary Transfer Orbit (GTO)** missions tend to have heavier payloads and maintain a high success rate.
- Orbits such as **Sun-Synchronous Orbit (SSO)** and **Polar** generally feature lighter payloads and show a mix of successful and failed launches.
- The diversity of payload masses across orbits reflects SpaceX's versatility in handling varied mission profiles.

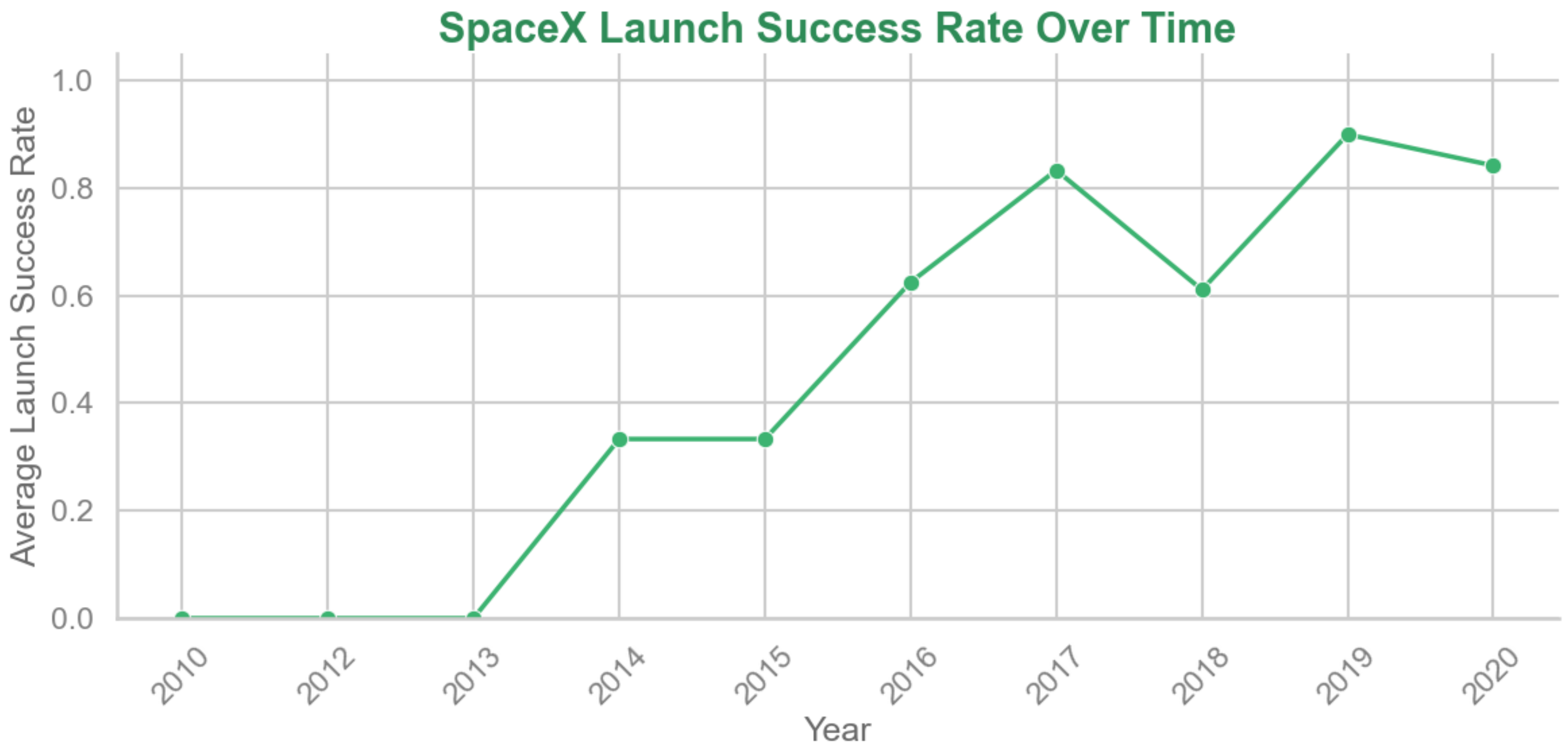


Visualization – Launch Success Yearly Trend

This line chart shows the **average yearly success rate** of SpaceX launches, illustrating trends in launch reliability over time. It provides a clear view of how launch outcomes have evolved as the program has matured.

Key Insights

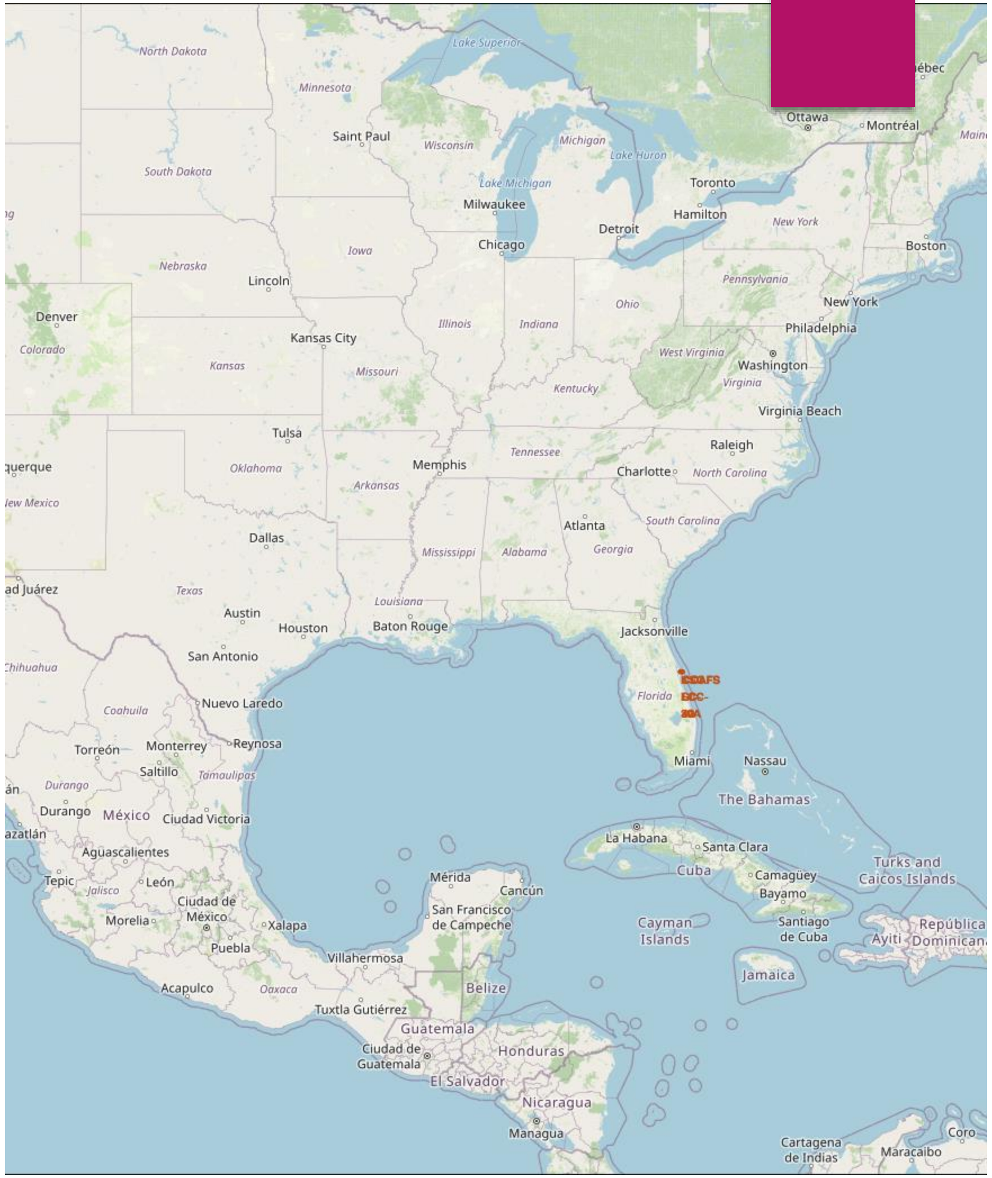
- The overall **launch success rate has improved steadily** from early years to recent ones.
- Initial years showed more variability and lower success rates, reflecting program development and testing phases.
- From around 2015 onward, success rates consistently approach or exceed **90%**, indicating significant operational maturity and reliability.
- The upward trend underscores SpaceX's growing experience and capability in achieving successful missions.



Section 3 – Rocket Launch Sites Proximities Analysis

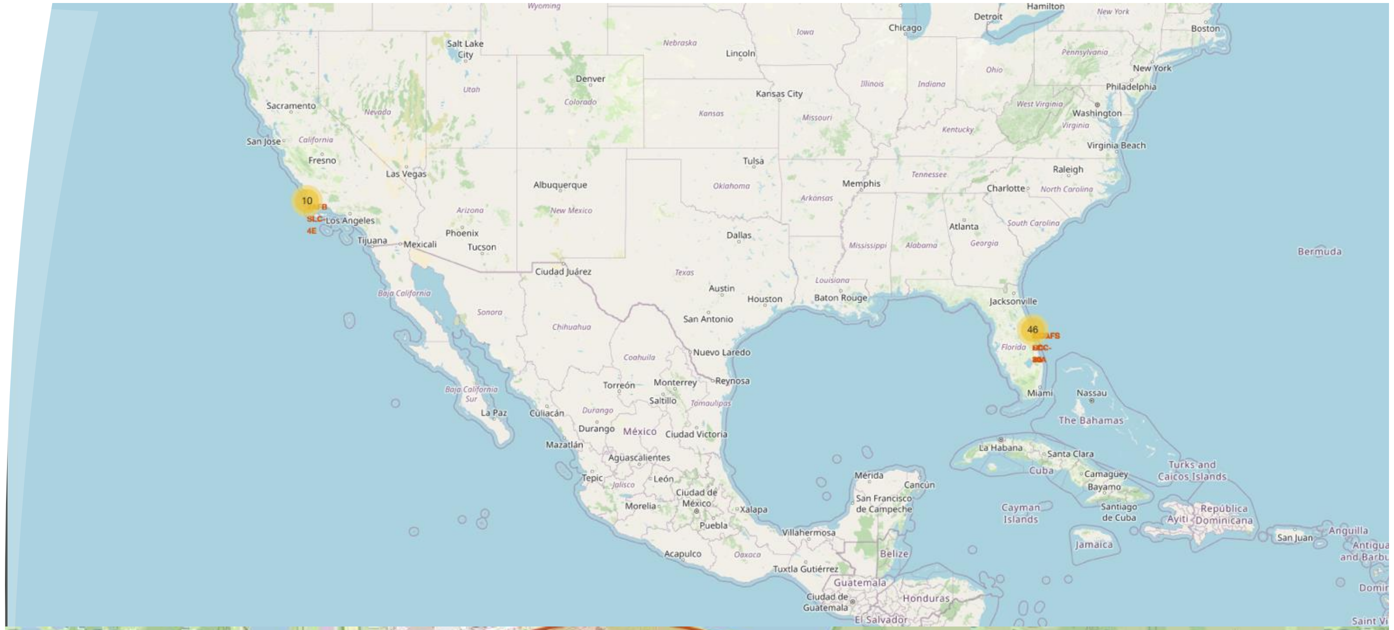
Launch Sites Locations

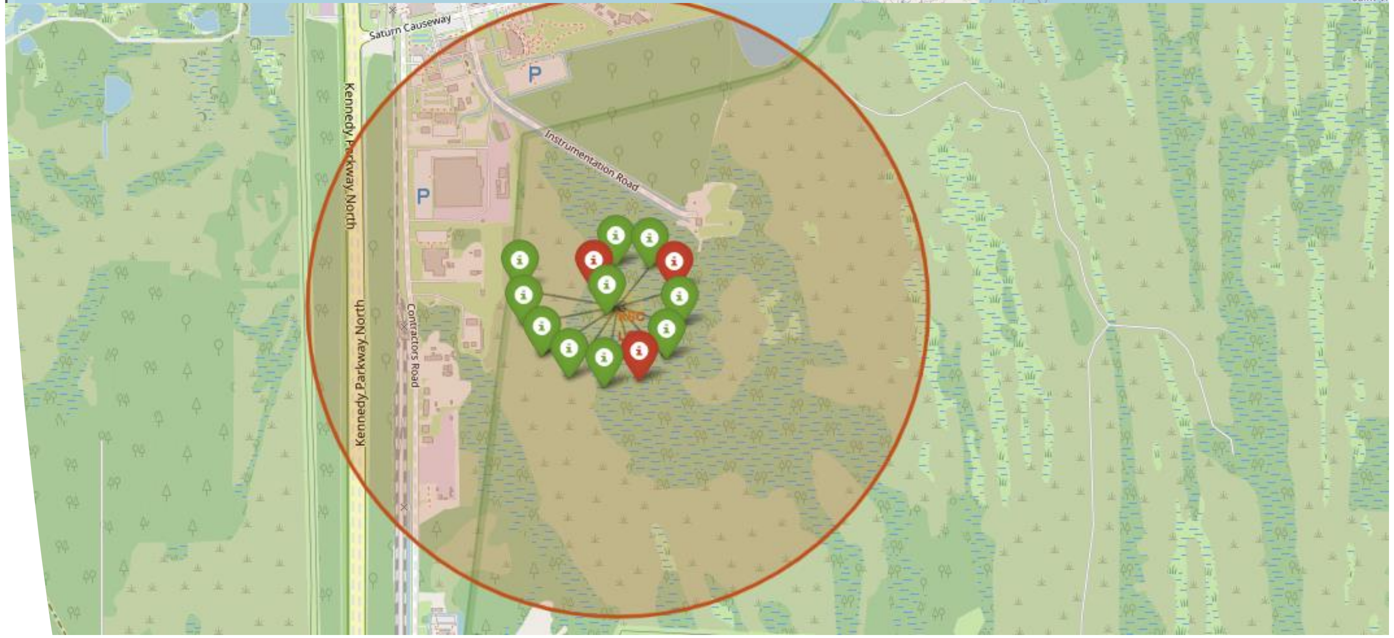
- Below are some of the key elements of the Folium Map
  - **Markers** placed at the four major SpaceX launch sites:
    - KSC LC-39A (Florida)
    - CCAFS SLC-40 (Florida)
    - VAFB SLC-4E (California)
    - SpaceX LC-4 (Texas)
  - **Popups** or labels may indicate launch site names for quick reference.
  - The **zoom level** is adjusted to capture all sites on a single world map view.
- **Key Findings:**
  - Launch sites are **strategically located** along the U.S. coastline, likely for safety and trajectory advantages over oceans.
  - **Two sites are clustered in Florida**, highlighting the region's importance for missions.
  - Visualizing locations globally shows **geographic distribution** and reinforces SpaceX's preference for U.S.-based launches.

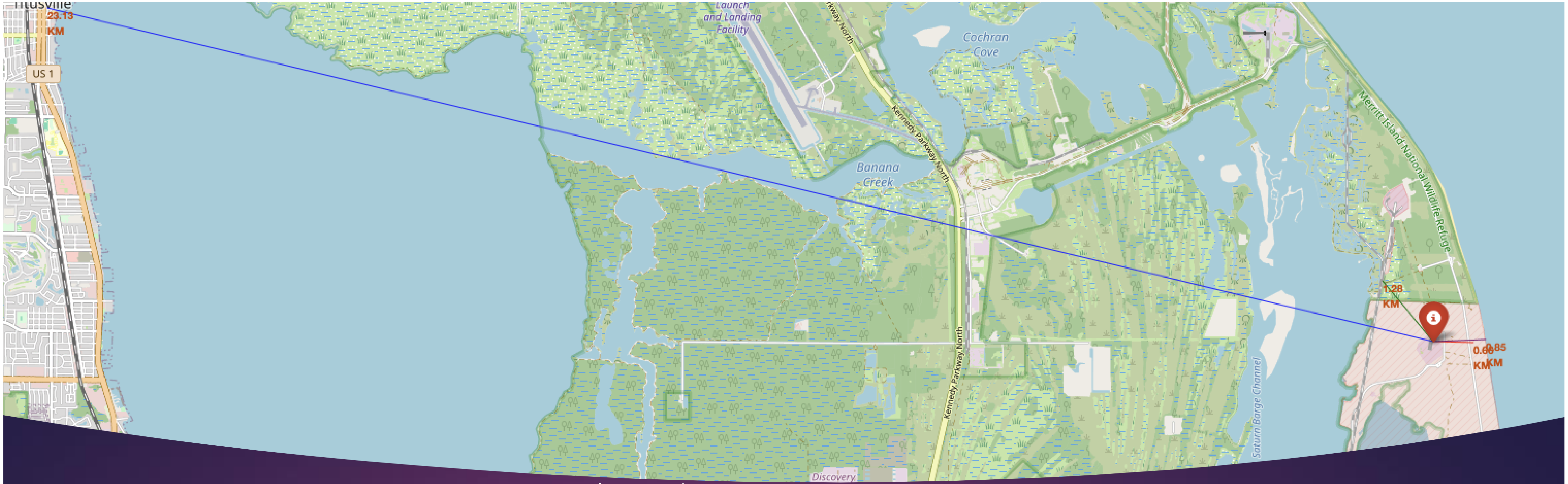


Launch Outcomes by Location

- Key Elements to Note:
  - Color-coded markers represent launch outcomes (Green = Success, Red = Failure)
  - Each marker is placed at the corresponding launch site's coordinates for each recorded launch
- Key Findings:
  - Most launches are successful, indicated by the dominance of green markers
  - Most launches occurred at CCAFS and KSC, indicating they are SpaceX's primary operational hubs.
  - Success rates were highest at these main sites, with failures more common in early missions or at less-used sites like VAFB.
  - Launch sites are geographically clustered, especially in Florida, offering logistical flexibility for different mission types.





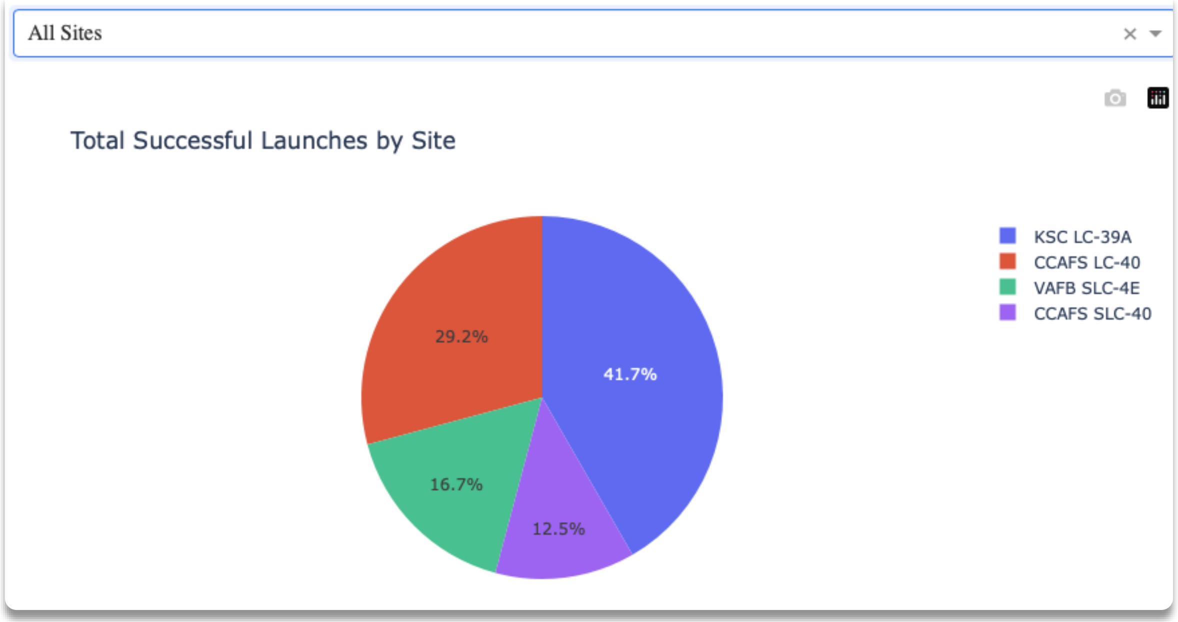


Distance from Launch Site to Key Points of Interests

- Key Map Elements:
  - **Launch Site Marker**: Labeled (e.g., LC-40)
  - **Distance Lines**: To nearest **railway**, **highway**, and **coastline**
  - **Popups**: Show distance in km/meters
  - **Context**: Nearby infrastructure and terrain for spatial analysis
- Key Insights
  - Launch sites are close to key infrastructure, aiding transport and logistics
  - Example: Highway (0.66 km), Railway (1.28 km), Water (0.85 km), City (23.13 km)
  - Coastal placement reduces risk and supports safe ocean-bound launches

Section 4 – Build a Dashboard with Plotly Dash

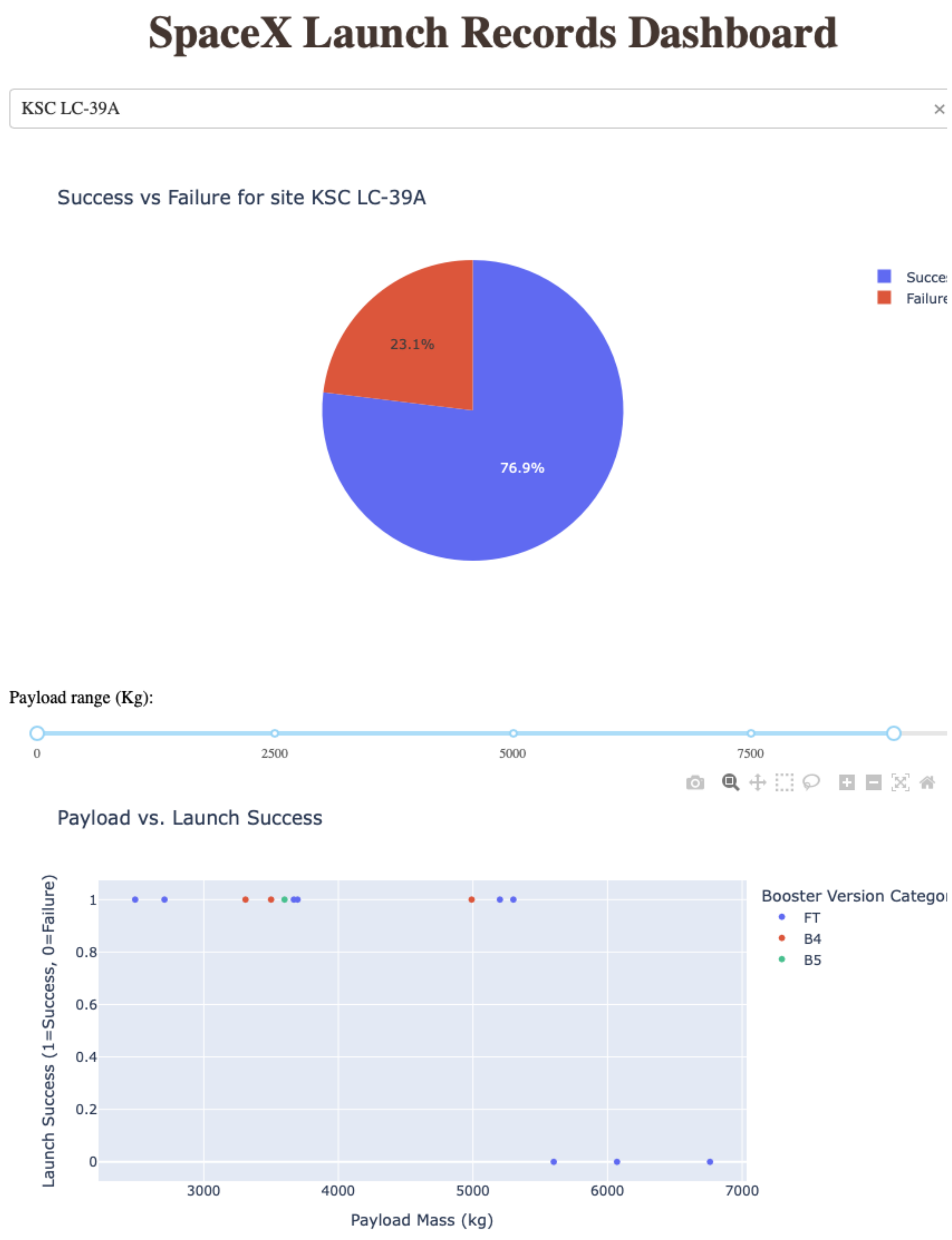
Distribution of Successful Launches by Site



- **Visual Overview**: Pie chart shows total successful launches by site.
- **Proportional Slices**: Each slice size reflects the **number of successes** at a specific launch site.
- **Top Performer**: **KSC LC-39A** has the most successful launches — likely the most reliable or heavily used.
- **Quick Comparison**: Enables fast visual comparison of performance across launch sites.
- **Interactive Dashboard**: Allows deeper exploration — filter by site to view **success vs. failure breakdowns**.

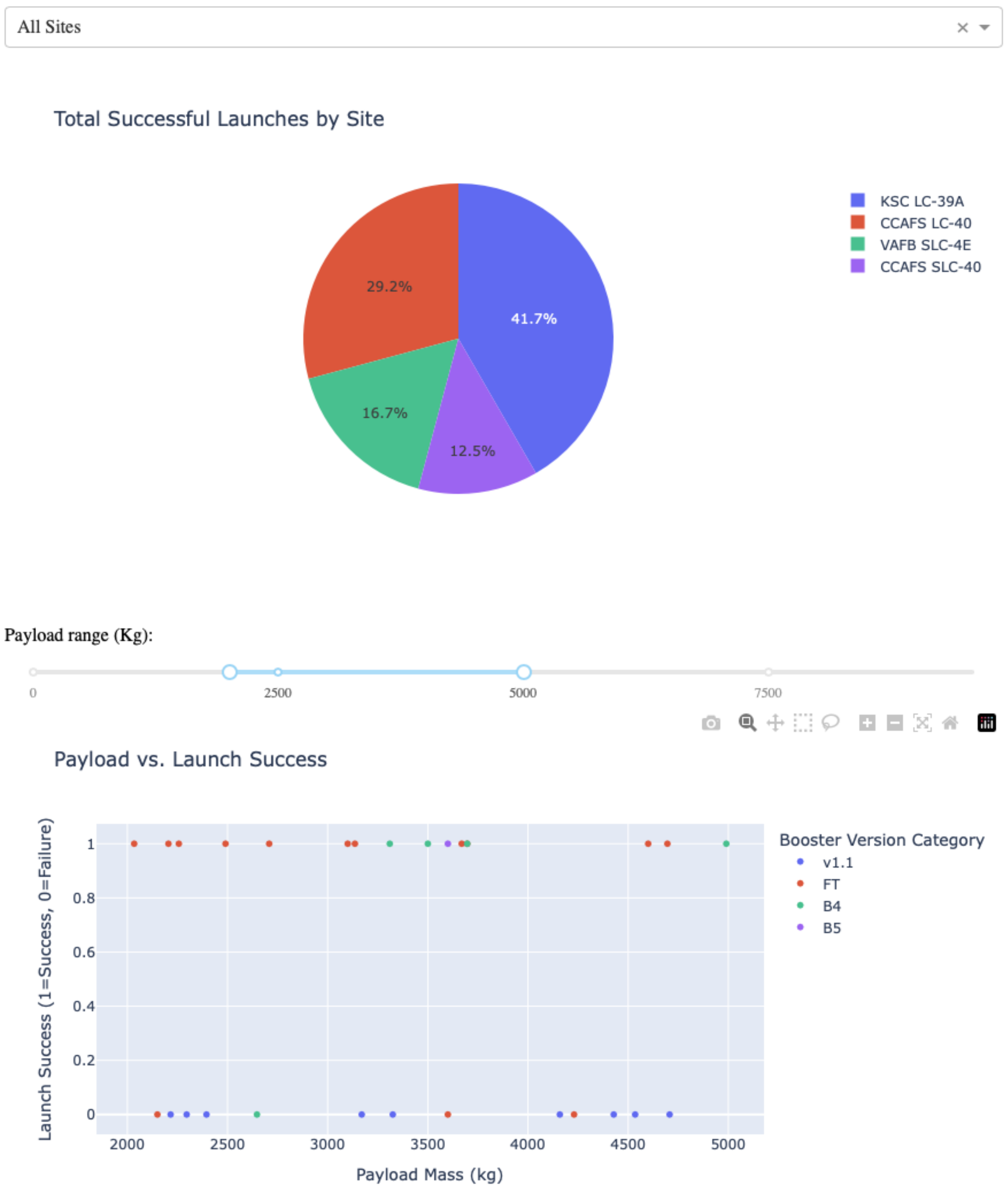
Insights – KSC LC-39A

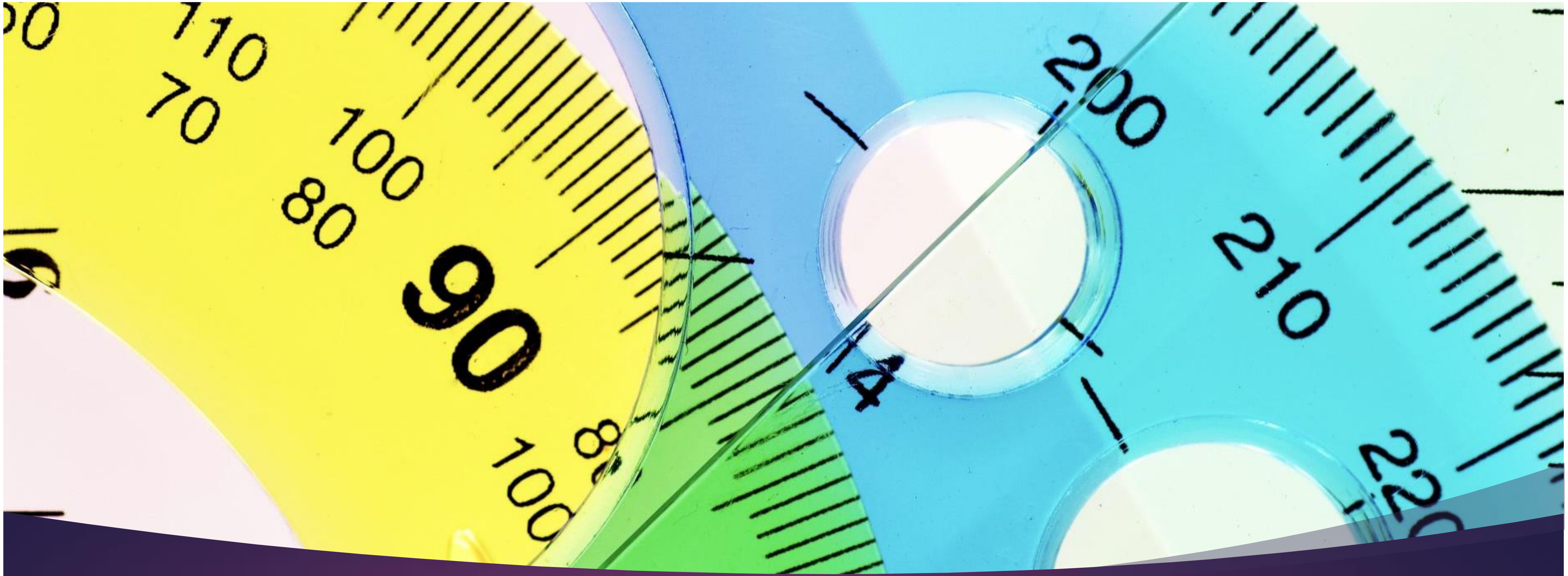
- **Success vs Failure**
  - **High success rate**: KSC LC-39A shows **a strong majority of successful launches**, with only **a small number of failures.**
  - Indicates **high reliability** of operations at this site — consistent with its status as a primary launch pad.
- **Payload Mass vs. Launch Success (by Booster Version)**
  - **Successes are more common** across a wide **range of payload masses**, especially in the **medium to high payload**range (e.g., 5,000–10,000 kg).
  - Certain **booster versions (e.g., FT or Block 5)** demonstrate **higher success rates**, especially with heavier payloads.
  - **Failures are rare** and tend to occur with **older booster versions** or **lighter payloads**, suggesting improvements over time.



Insights – Payload Between 2500 and 5000

- **Success vs Failure**
  - **KSC LC-39A and CCAFS SLC-40** account for the majority of successful launches in this payload band.
- **Booster Version Category vs Launch Outcome**
  - **Booster versions like Block 5 and FT** perform particularly well in this payload range, with **near-perfect success rates**.



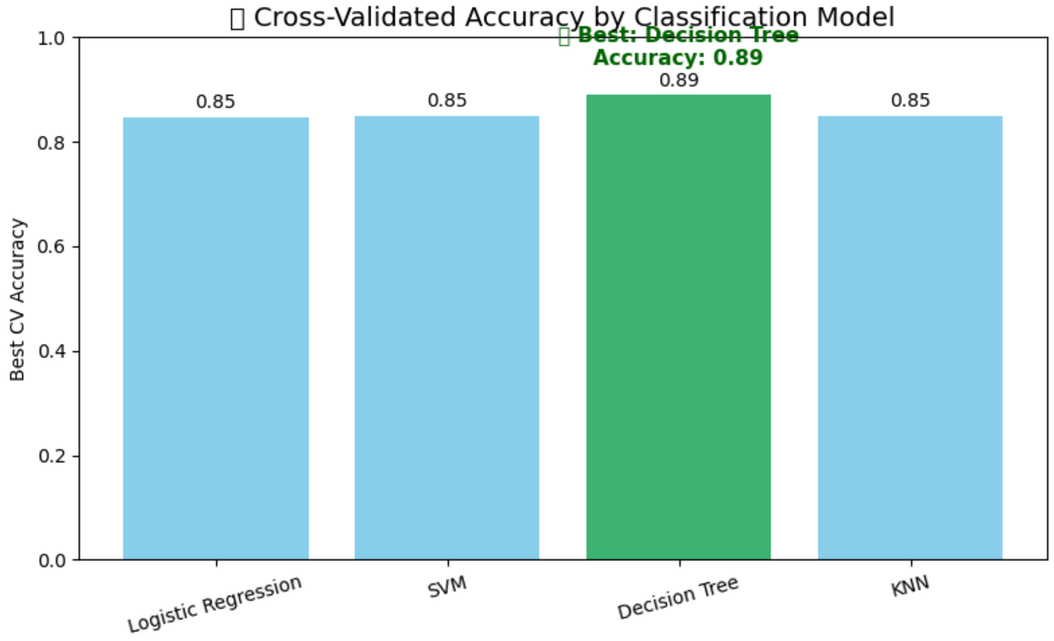


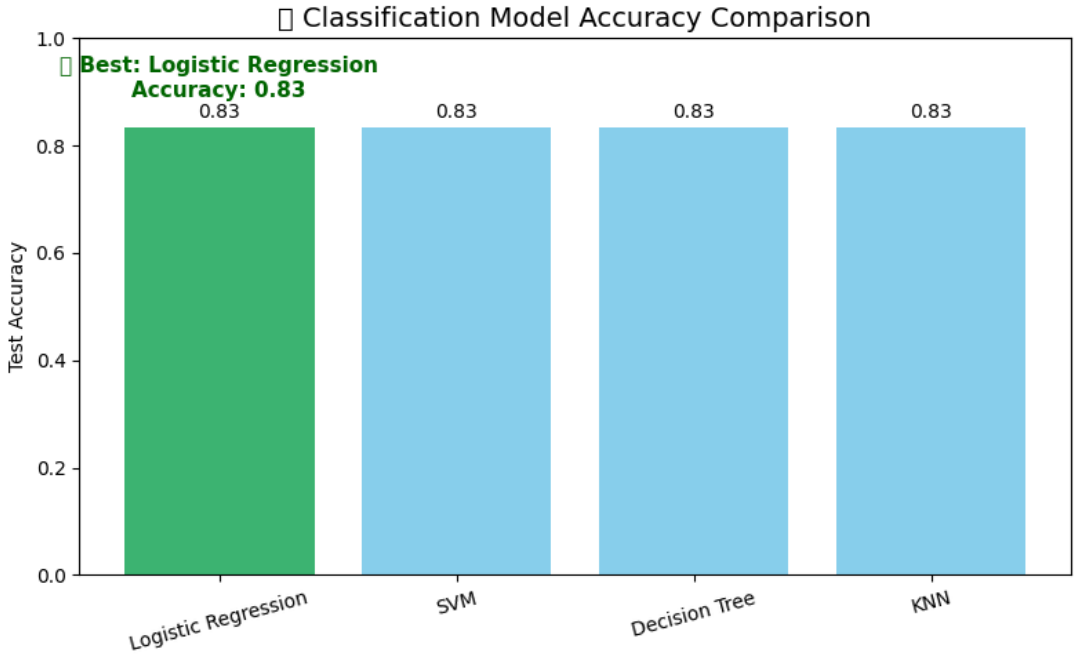
Section 5 – Predictive Analysis (Classification

Classification Accuracy

Following our modeling process – we've gathered the following insights on the accuracy of the models:

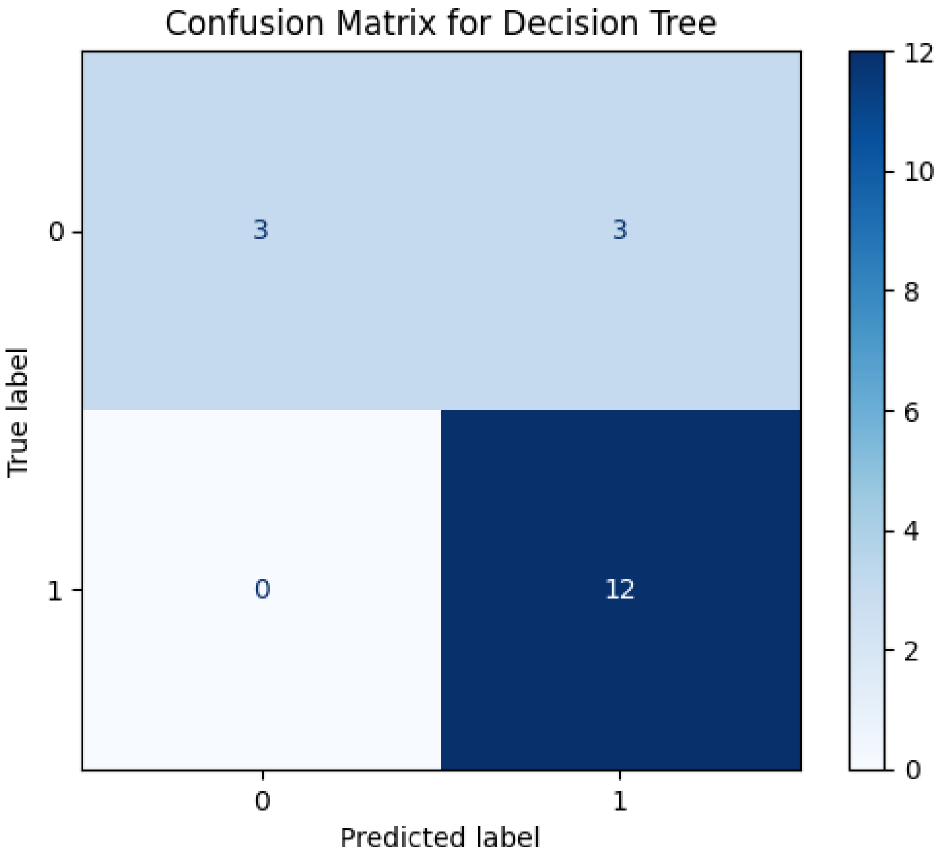
- The test accuracy of all four classification models was visualized. All models performed identically due to relatively low sample size. We compared models using 10-fold cross-validation.
  The **highest cross-validated accuracy** was achieved by **Decision Tree** at **89%.**
- The charts below shows the test accuracy and best cross-validation accuracy (best_score_) for each classifier.





Confusion Matrix – Decision Tree

- A confusion matrix is a table that summarizes the performance of a classification model by showing the counts of true positives, true negatives, false positives, and false negatives, helping to evaluate how well the model distinguishes between classes. T
- **True Positives & True Negatives** (diagonal values) indicate correct predictions.
- **False Positives / False Negatives** (off-diagonal) show misclassification
- Most errors came from classifying X as Y (e.g., false positives).
- Overall, the model shows strong classification accuracy with relativel low misclassification rates.



Conclusion

- Conducted thorough **data collection and exploratory analysis** using SQL and visualizations to identify key launch trends.
- Performed **launch site proximity analysis**, revealing strategic advantages for logistics and safety.
- Developed an **interactive Plotly dashboard** for dynamic insights on launch outcomes by site, payload, and booster type.
- Applied **regression analysis** to predict factors influencing launch success.
- Together, these efforts provide a strong foundation for **data-driven decisions** and operational improvements in future SpaceX missions.

Appendix A – GitHub Links

- Below are links to GitHub Repository underlying the insights brought forward in this presentation:
  - Repository: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025
    - Data Collection – REST API: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/1_jupyter-labs-spacex-data-collection-api-v2.ipynb
    - Data Collection – Webscrapping: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/2_jupyter-labs-webscraping.ipynb
    - Data Wrangling: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/3_labs-jupyter-spacex-Data%20wrangling-v2.ipynb
    - EDA – SQL: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb
    - EDA – Visualization: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/5_jupyter-labs-eda-dataviz-v2.ipynb
    - Launch Sites Location: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/6_lab-jupyter-launch-site-location-v2.ipynb
    - Plotly Dashboard Python Code: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/7_labsjupyter_plotlydash.ipynb
    - Regression/Predictive Analysis: https://github.com/camfarrell25/IBM-Capstone-Project-CF2025/blob/main/8_SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb