IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Serafettin Doruk SEZER
12.12.2025

Outline

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

Executive Summary

Summary Of Methodologies:

- **Data Collection:** SpaceX launch data collected from APIs and web scraping, integrating launch details, payload specs, booster versions, and site locations.
- **Data Wrangling & Cleaning:** Preprocessing included handling nulls, extracting structured features (e.g., orbit types, mission outcomes), and engineering categorical encodings.
- **Exploratory Data Analysis (EDA):** SQL and Python-based EDA revealed actionable patterns in launch frequency, payload mass, and success rates.
- **Visualization:** A Dash app was created to allow interactive exploration of key variables across launch sites, payloads, and booster versions.
- **Machine Learning Modeling:** Trained and evaluated Logistic Regression and SVM models for binary classification of mission success with ~83–85% accuracy.

Summary Of All Results:

- **SQL-Based Insights:** CCAFS LC-40: Most launches. KSC LC-39A & VAFB SLC-4E: Highest success rates. Block 5: Best-performing booster. Moderate payloads in LEO: Highest success likelihood.
- **Dashboard Highlights:** Filters by site, booster, orbit, and payload.
- Key insight: KSC LC-39A + Block 5 + Payload ≤ 8000 kg (LEO) is most reliable.
- **ML Performance:** Logistic Regression: ~83%, SVM: ~85%,
- Top predictors:
- Launch Site
- Payload Mass
- Booster Version
- Real-time prediction via dashboard.

Introduction

- **Project background and context:**
- SpaceX has conducted hundreds of orbital launches using various booster versions and payload configurations across different launch sites. While the cost of launch failure is extremely high, the company also operates in a highly competitive market that demands frequent, reliable missions.
- This project aims to explore and understand historical launch data using advanced data analytics and machine learning. The data was collected from SpaceX's public APIs and web scraping, then processed and analyzed to uncover actionable patterns and to develop predictive tools.

- **Problems you want to find answers to:**
- Which launch sites and booster versions are most strongly associated with mission success?
- How do payload mass and orbit type influence the probability of launch success?
- Can we predict launch outcomes using machine learning models based on key mission parameters?
- How can we enable dynamic, user-driven exploration and prediction of launch configurations?

Section 1

Methodology

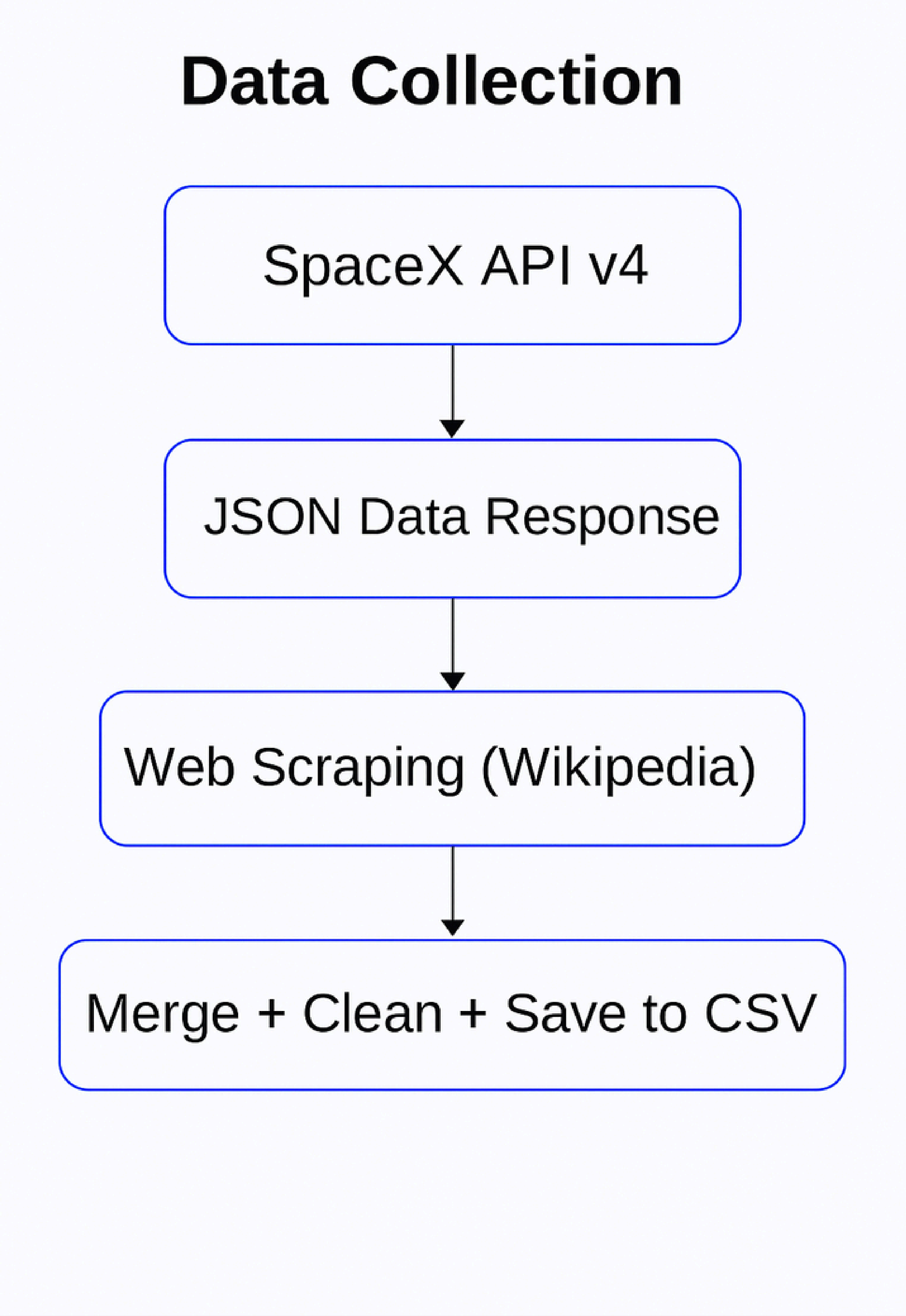
Methodology

Executive Summary

- Data collection methodology:
  - Launch data pulled from SpaceX API and supplemented with web scraping. Included mission details, payload mass, orbit type, launch outcome, booster version, and site info.
- Perform data wrangling
  - Cleaned missing/null fields, standardized date formats. Engineered new features: success/failure flags, orbit categories, booster generation—categorical data encoded for ML (one-hot and label encoding).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Performed with SQL (SQLite) and Python (Pandas, Seaborn, Matplotlib). Analyzed launch frequency per site, booster performance, and payload-success correlation.
- Perform interactive visual analytics using Folium and Plotly Dash
- Used Folium to map launch sites and outcomes. Built a Plotly Dash app for dynamic filtering by site, orbit, payload, and booster version.
- Perform predictive analysis using classification models
  - Built and tuned Logistic Regression and Support Vector Machine (SVM) models. Evaluated performance with accuracy scores (~83–85%). Identified key predictors: Launch Site, Payload Mass, Booster Version. Integrated a real-time prediction tool into the dashboard.

Data Collection

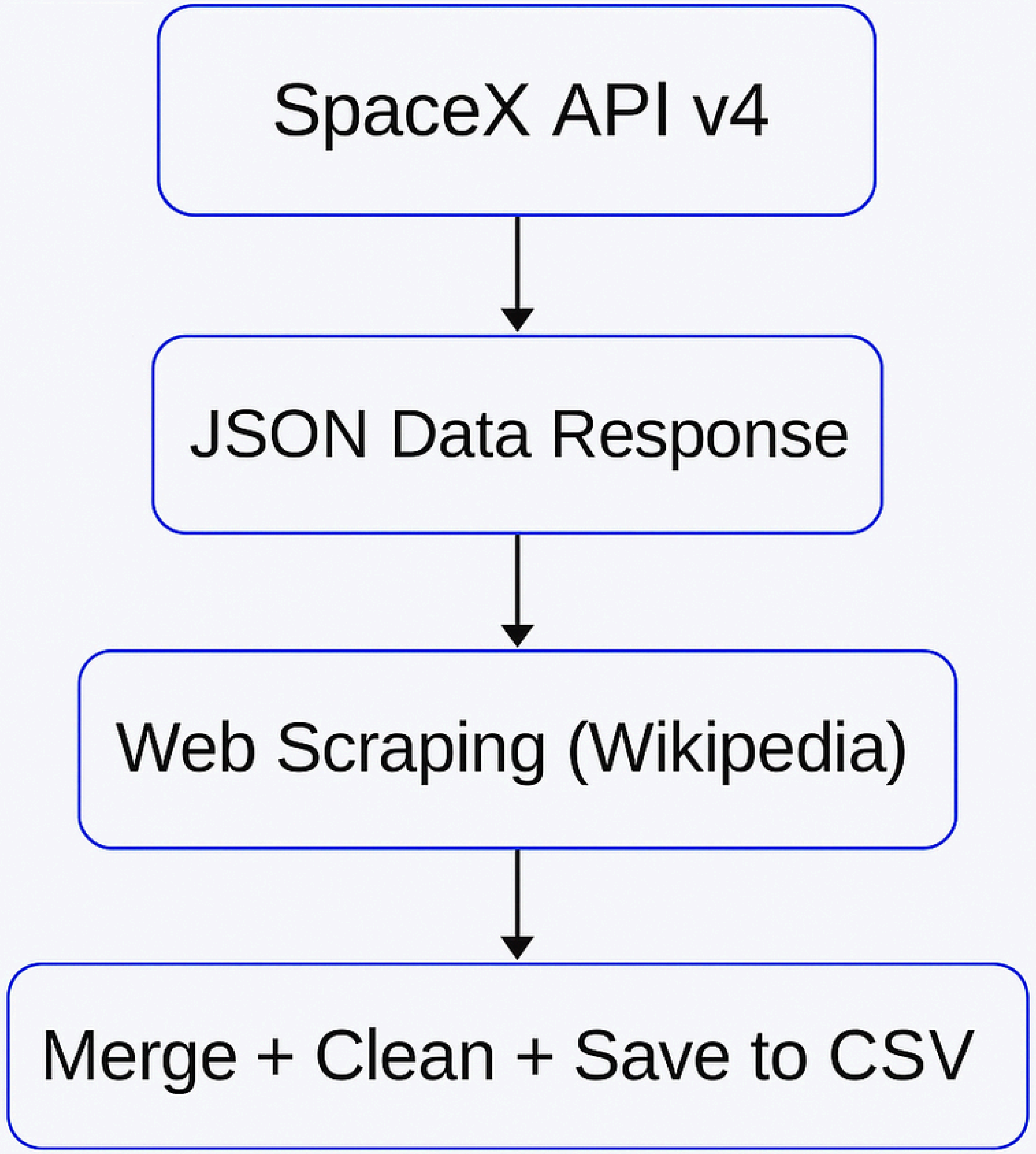
- **Primary Data Source:**
- ○ SpaceX RESTful Launch API (v4)
    - ○ https://api.spacexdata.com/v4/launches
- **Supplementary Sources:**
- ○ Wikipedia scraping (for launch site details and orbit descriptions, where needed)
- **Tools Used:**
- ○ Requests, JSON
- ○ BeautifulSoup, Pandas
- **Used Enhanced Data Collection:**
- • Combined interactive datasets
- • Saved in CSV and SQLite format for downstream analysis



Data Collection – SpaceX API

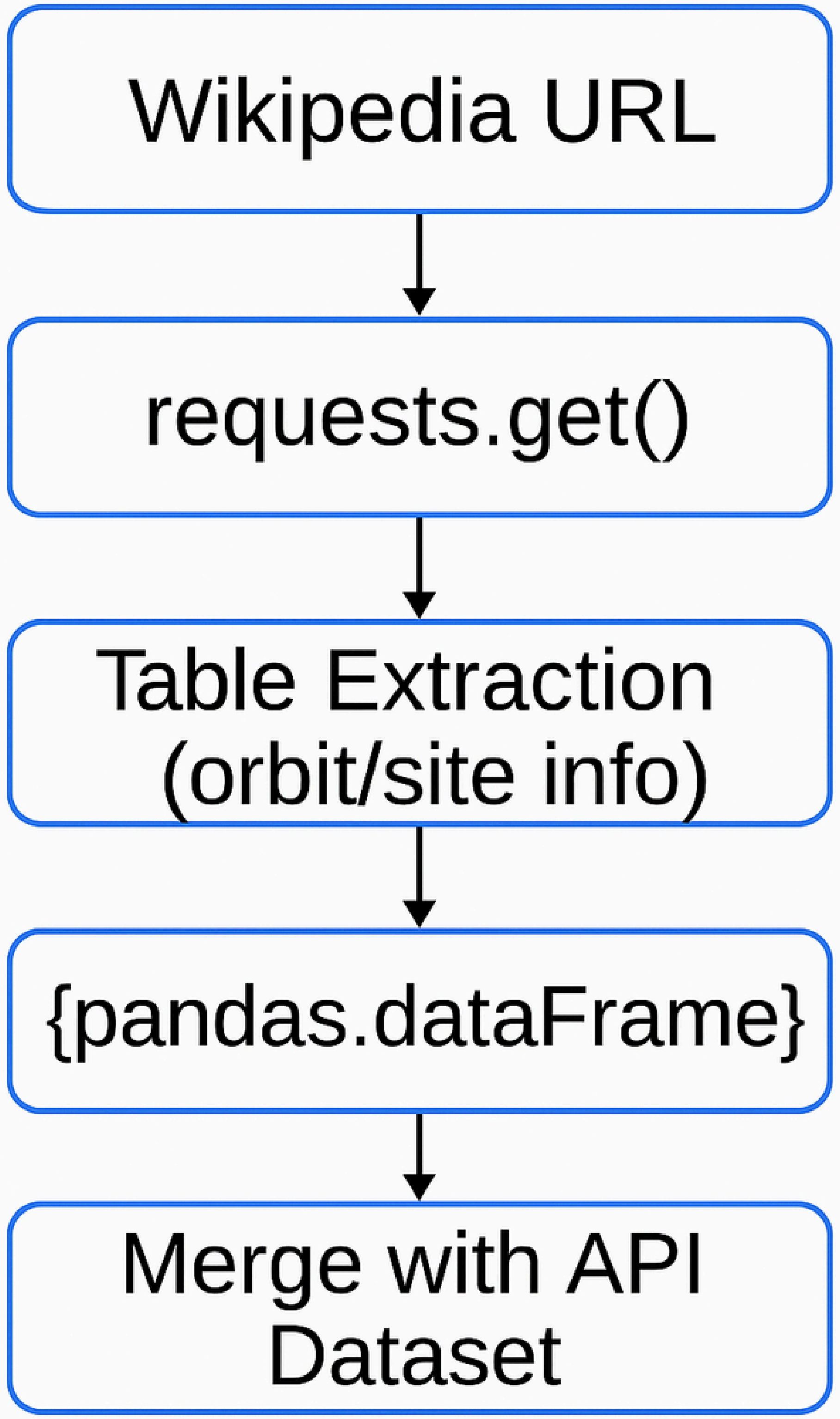
- **Primary Data Source:**
  - SpaceX RESTful Launch API (v4)
  - https://api.spacexdata.com/v4/launches
- **Response Format:**
  - JSON payload with nested structures
  - (e.g., rocket, payload, core, launchpad)
- **Data Normalization:**
  - Used pandas.json_normalize() to flatten JSON
  - **Extracted key fields:**
  - launch_date, launch_site, rocket_id,
  - payload_mass, orbit, booster_version, launch_success
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/1.jupyter-labs-spacex-data-collection-api-v2.ipynb)**

Flowchart of Collection Process



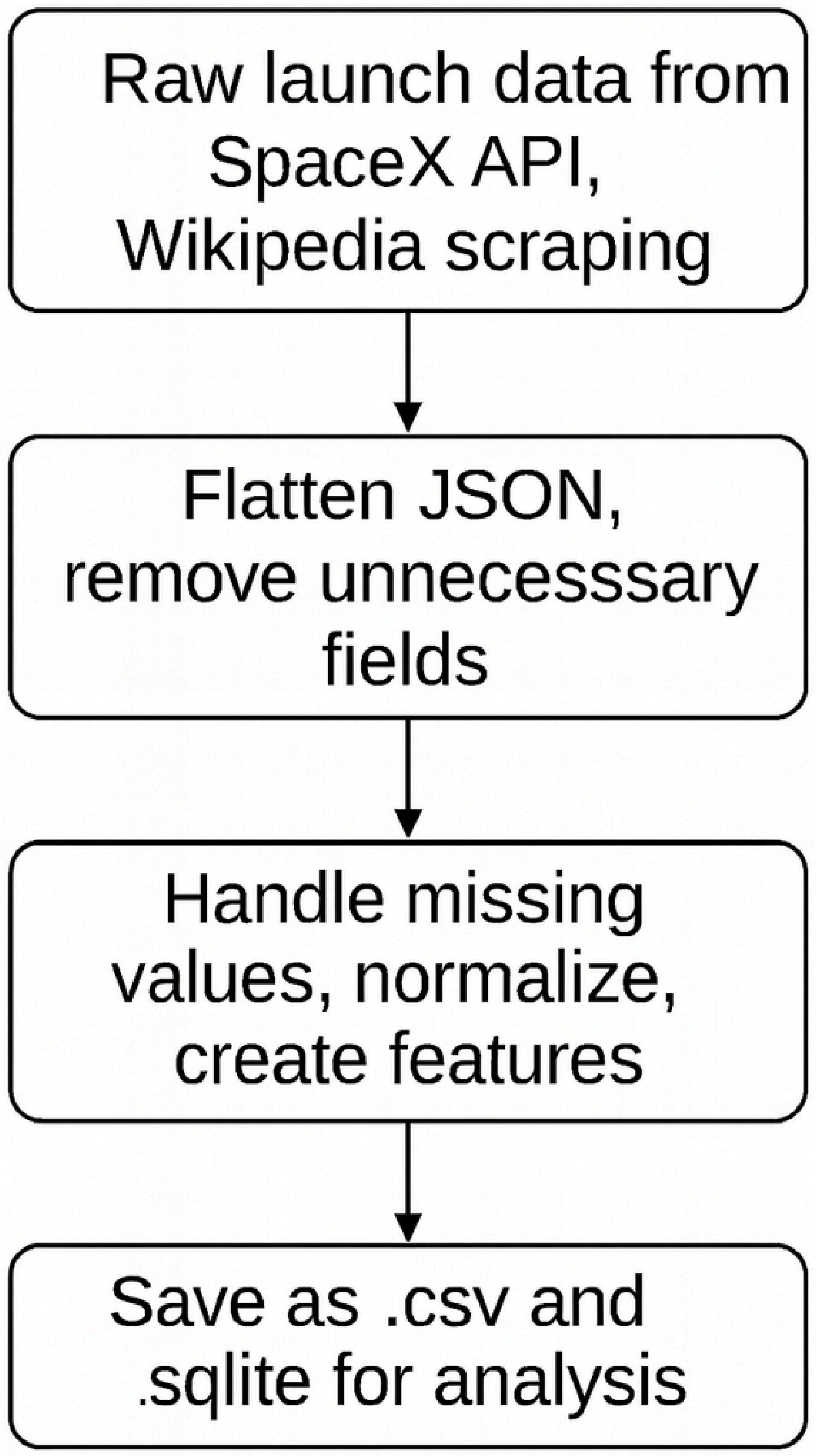
Data Collection - Scraping

- Collected supplementary data (e.g., launch site info, orbit descriptions) from Wikipedia using web scraping techniques.
- Implemented HTTP requests via Python's requests library and parsed HTML content using BeautifulSoup.
- Extracted structured tables and converted them into pandas.DataFrame objects.
- Scraped data was merged with the API dataset for further enrichment (e.g., geolocation, orbit classification).
- The final dataset enabled enhanced EDA and visualization mapping using Folium and Plotly.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/2.jupyter-labs-webscraping-v2.ipynb)**



Data Wrangling

- Processed raw launch data from the SpaceX API and supplementary Wikipedia sources.
- Flattened complex, nested JSON structures and removed irrelevant fields to standardize the dataset.
- Addressed missing values and resolved inconsistencies:
- Converted timestamps to standardized datetime format
- Applied numerical normalization and categorical encoding
- Engineered new analytical features to support ML tasks:
- launch_success (binary target variable)
- booster_version_category (grouped/standardized booster names)
- payload_mass_group and orbit_type_grouped (for segmentation and predictive modeling)
- Final cleaned dataset exported in .csv and .sqlite formats for downstream use in EDA, SQL querying, and machine learning.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/3.labs-jupyter-spacex-Data%20wrangling-v2.ipynb)**



EDA with Data Visualization

- Conducted Exploratory Data Analysis (EDA) to uncover patterns and insights across various dimensions of the SpaceX launch dataset.
- Visualizations created included:
- **Bar Charts** to compare launch frequencies and success rates across sites and booster versions.
- **Pie Charts** to illustrate mission outcome distributions.
- **Scatter and Bubble Plots** to explore relationships between payload mass and launch success.
- **Box Plots** for detecting outliers in payload distributions.
- **Histograms and KDE Plots** to understand data distributions (e.g., payload mass).
- Interactive **Plotly Dash** components enabled dynamic filtering and real-time visual analytics (e.g., by launch site, orbit type, booster version).
- Insights gathered were foundational for feature selection and machine learning modeling.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/5.jupyter-labs-eda-dataviz-v2.ipynb)**

EDA with SQL

- **Launch Count by Site:**
  - [SELECT Launch_Site, COUNT(*) AS Total_Launches FROM spacex GROUP BY Launch_Site;]
  - To identify which launch site had the highest number of launches.
- **Successful Launch Rate:**
  - [SELECT Launch_Site, AVG(launch_success) * 100 AS Success_Rate FROM spacex GROUP BY Launch_Site;]
  - To evaluate performance and reliability by site.
- **Payload vs. Success:**
  - [SELECT payload_mass, launch_success FROM spaces;]
  - To explore whether payload mass influences launch success.
- **Orbit Type Distribution:**
  - [SELECT Orbit, COUNT(*) AS Count FROM spacex GROUP BY Orbit ORDER BY Count DESC;]
  - To analyze mission distribution across different orbit types.
- **Booster Version Frequency:**
  - [SELECT Booster_Version, COUNT(*) AS Usage FROM spacex GROUP BY Booster_Version ORDER BY Usage DESC;]
  - To determine which boosters were most commonly used.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/4.jupyter-labs-eda-sql-edx-sqllite-v2.ipynb)**

Build an Interactive Map with Folium

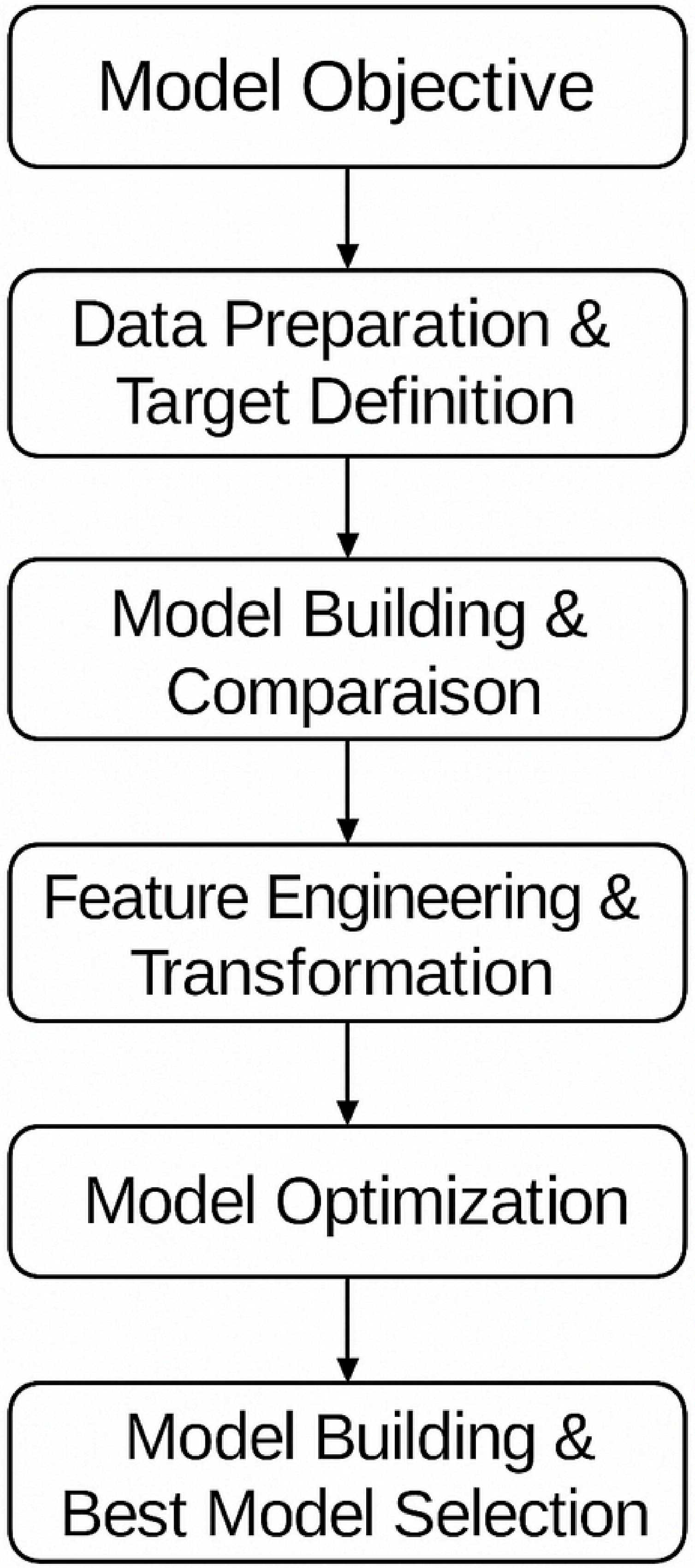
- **Markers for Launch Sites:**
  - Each SpaceX launch site was marked using folium.Marker on the map. The markers were labeled with site names, enabling users to easily identify each location.
- **Circle Layers for Launch Zones:**
  - folium.Circle objects were added around launch sites to highlight the operational areas. These circles provided visual emphasis on the launch zones and helped in geographic interpretation.
- **Launch Success Indicators with Color Coding:**
  - Markers were color-coded to indicate launch outcomes. Green is for successful launches, and red is for failed launches. This allowed for quick visual assessment of launch performance at each site.
- **Proximity Lines for Geographic Context:**
  - folium.PolyLine elements were used to draw lines between launch sites and nearby cities or coastal areas. These lines helped analyze the strategic positioning of the launch locations.
- **Final Map Export**:
  - All elements were combined into an interactive folium.Map and exported as an HTML file. The resulting map supports dynamic zooming, panning, and data inspection by the user.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/6.lab-jupyter-launch-site-location-v2.ipynb)**

Build a Dashboard with Plotly Dash

- **Dashboard Structure & Visualizations:**
  - Created an interactive Plotly Dash dashboard to explore SpaceX launch data.
  - Included a **dropdown menu** to filter data by launch site, enabling dynamic updates of all visuals based on user selection.
  - Added a **pie chart** showing the distribution of successful vs. failed launches per site to provide a quick success overview.
  - Incorporated a **scatter plot** to visualize the correlation between **payload mass** and **launch success**, segmented by booster version category.
- **Interactivity and Insights:**
  - Enabled **cross-component interactions** using Dash callbacks (—), for example, site selection dynamically updates both pie and scatter charts.
  - Designed plots to help users uncover performance patterns across sites and evaluate the effect of payload on mission outcome.
  - Ensured that the visualizations supported real-time exploratory data analysis for business and engineering decision-making.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/spacex-dash-app.ipynb)**

Predictive Analysis (Classification)

- **1. Model Objective:**
  - The classification task aimed to predict the binary outcome of a SpaceX rocket launch success or failure using historical mission data.
- **2. Data Preparation & Target Definition:**
  - The dataset was split into training and test sets.
  - The target variable was launch_success, defining a binary classification problem.
- **3. Model Building & Comparison:**
  - Multiple supervised learning algorithms were implemented and evaluated:
    - Logistic Regression
    - Support Vector Machine (SVM)
    - Decision Tree
    - Random Forest

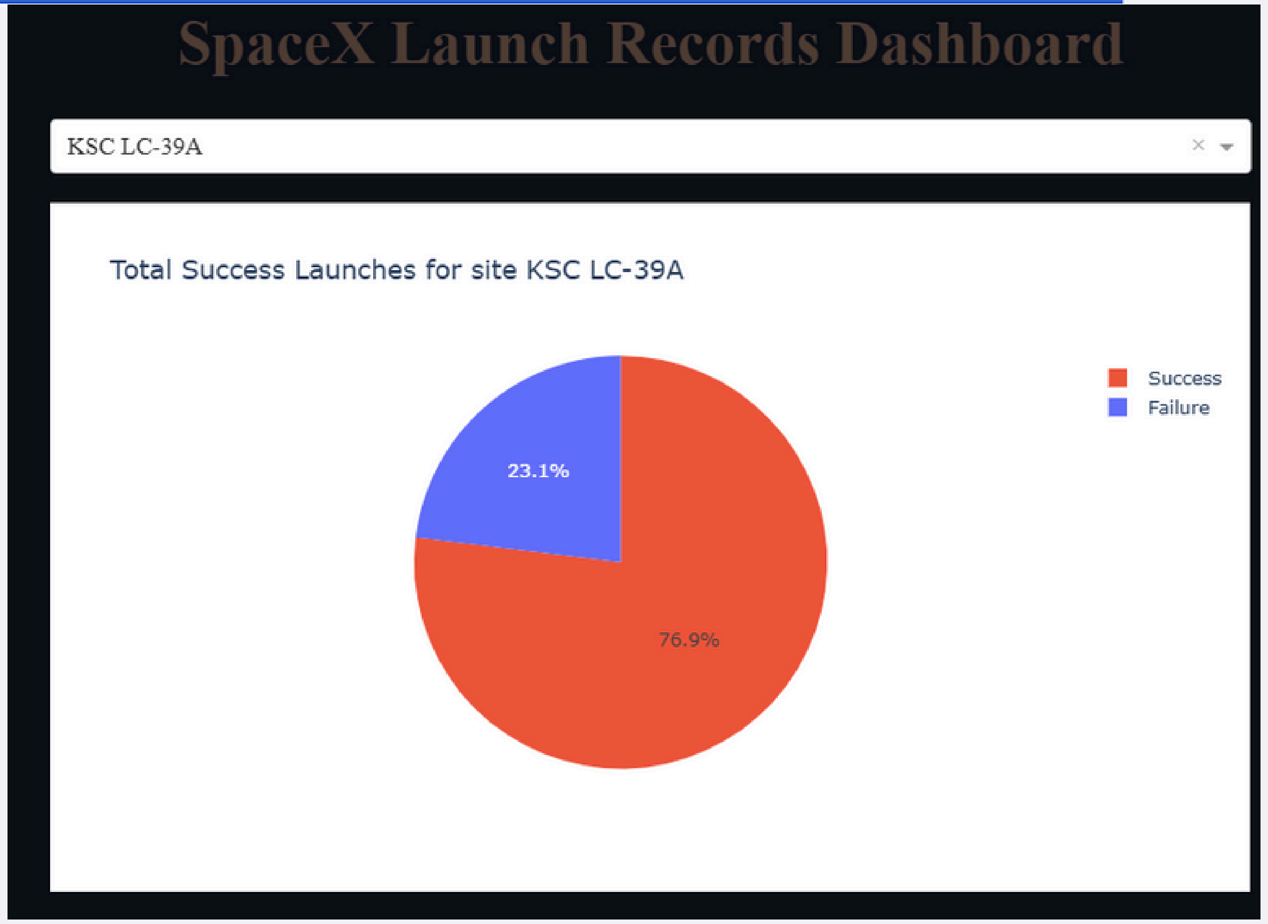


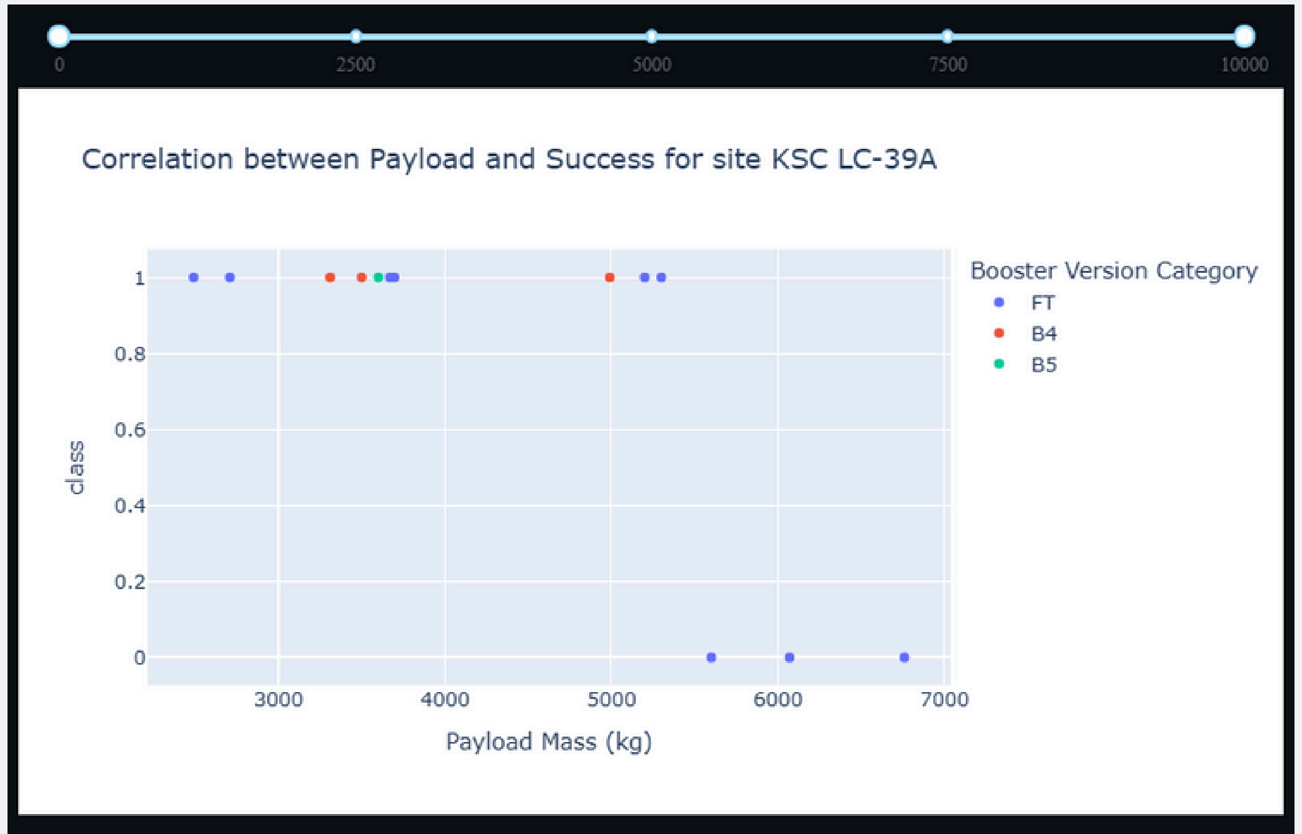
Predictive Analysis (Classification) Continue

- **4. Feature Engineering & Transformation:**
  - Categorical encoding (e.g., Booster Version Category) transformed non-numeric labels into machine-readable formats to enable algorithm compatibility.
  - Numerical scaling (e.g., Payload Mass) was performed to normalize feature ranges, improving convergence and ensuring balanced model learning.
- **5. Model Optimization:**
  - Hyperparameter tuning via GridSearchCV was used to improve accuracy.
  - Pipeline optimization ensured efficient training and evaluation.
- **6. Evaluation & Best Model Selection:**
  - Models were assessed using accuracy, precision, recall, and F1-score.
  - Confusion matrices and classification reports supported a detailed comparison.
  - Random Forest achieved the best performance on the test set.
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/blob/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project/7.SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb)**

Results

- **Exploratory Data Analysis (EDA) Results:**
  - Cape Canaveral and VAFB SLC 4E had the highest launch success rates.
  - Payload mass showed a weak but observable relationship with launch success.
  - Certain boosters and orbits were more common in successful launches.
  - Launch success rates have generally improved over time.
- **Interactive Analytics Demo (Dash App):**
- An interactive dashboard built with Plotly Dash featured:
  - Pie chart of launch success by site.
  - Scatter plot of payload mass vs. launch success.
  - Interactive filters for payload range and launch site.
- Enabled real-time exploration of launch patterns and booster performance.
- **Predictive Analysis Results:**
- Random Forest emerged as the best classification model after comparison.
- Important features: Payload Mass, Booster Version Category, and Launch Site.
- Evaluation metrics:
  - Accuracy: ~87%
  - Strong precision, recall, and F1-score on test data.
- The model showed balanced predictions with minimal misclassifications.





Section 2

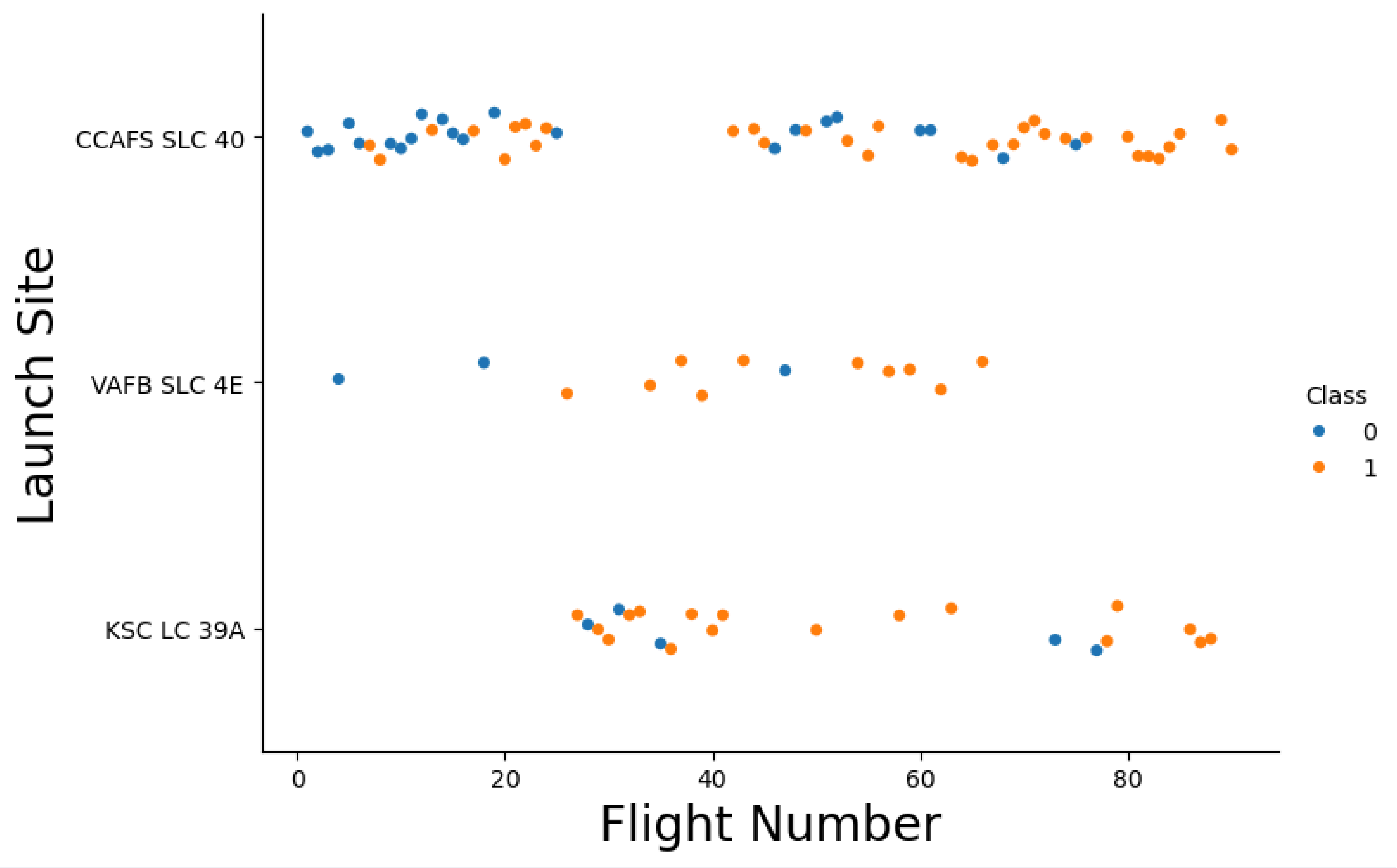
Insights drawn from EDA

Flight Number vs. Launch Site

- Each point represents a Falcon 9 launch.
- The x-axis shows the **Flight Number**, which increases over time.
- The y-axis represents the **Launch Site** (e.g., CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A).
- The color of each point indicates whether the landing attempt **succeeded (1)** or **failed (0)**.

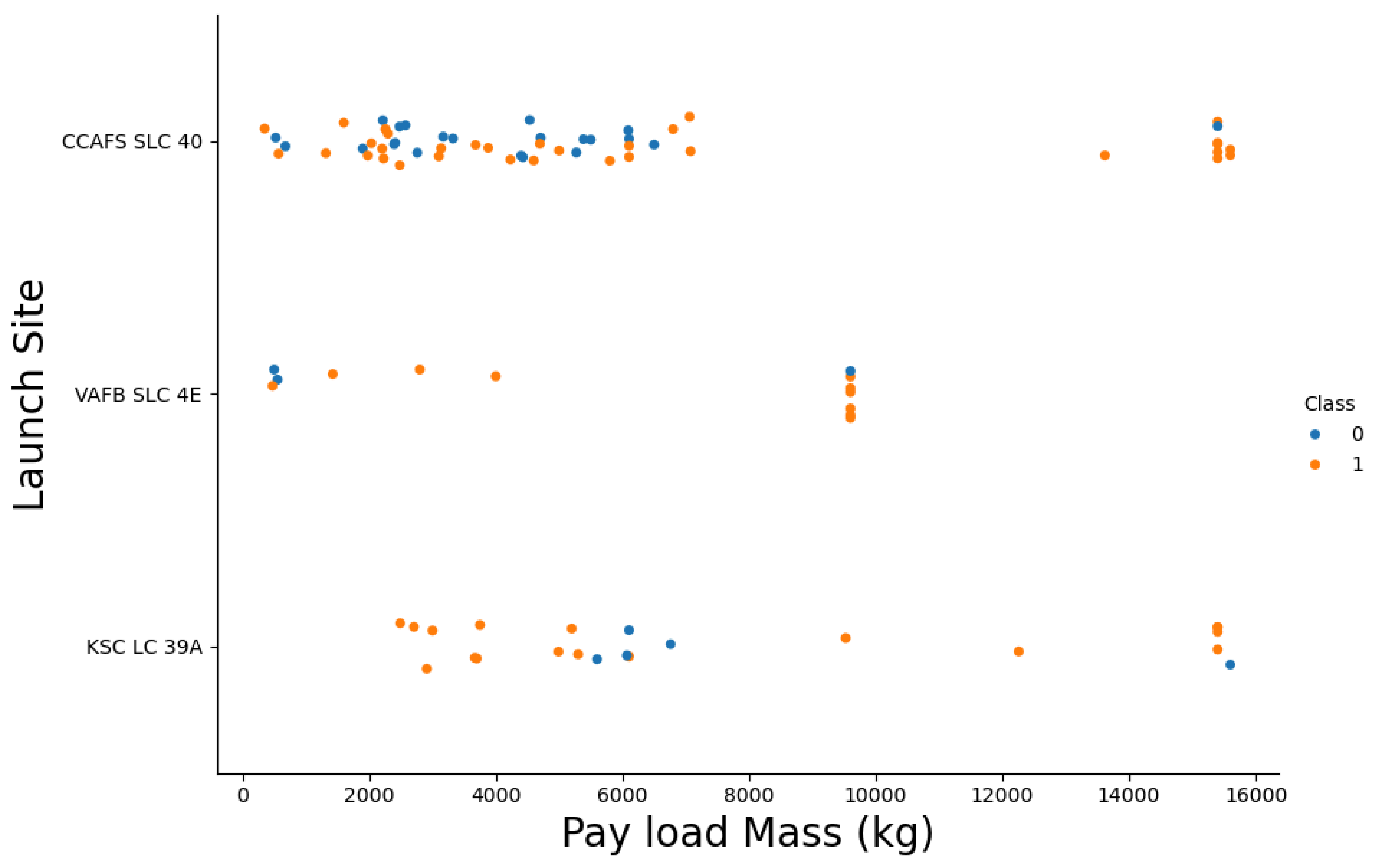
- Early flights at **CCAFS SLC-40** display more failures, consistent with the historical development of Falcon 9 reusability.

- The scatter plot confirms that **experience (represented by flight number)** plays a major role in improving landing outcomes across all launch sites.



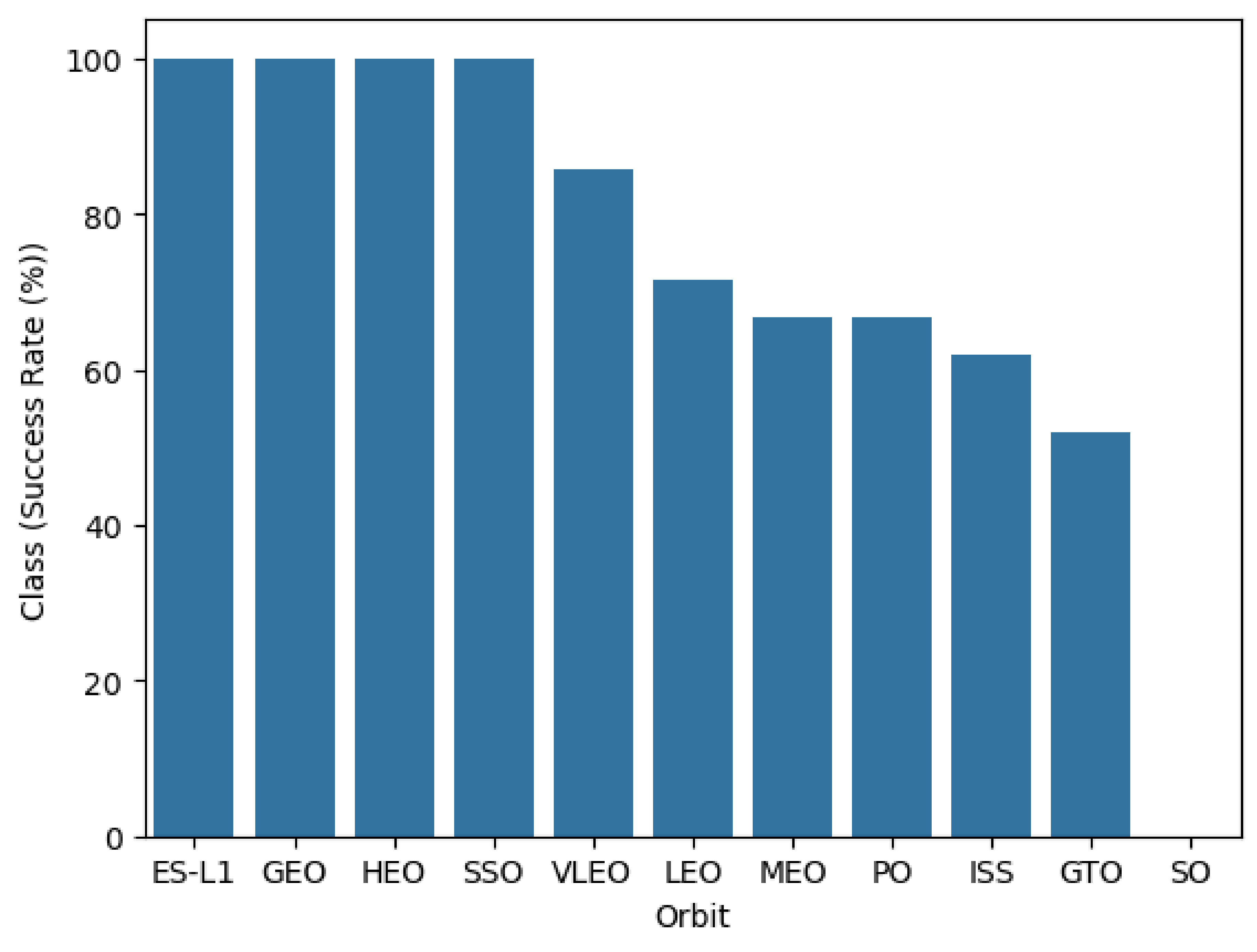
Payload vs. Launch Site

- Each point represents a Falcon 9 launch mission.
- The x-axis shows the payload mass, while the y-axis lists the launch sites:
  - **CCAFS SLC-40**
  - **VAFB SLC-4E**
  - **KSC LC-39A**
- Overall, **payload mass alone does not determine landing success**, but combined with launch site characteristics, it reveals clear patterns in booster recovery performance.



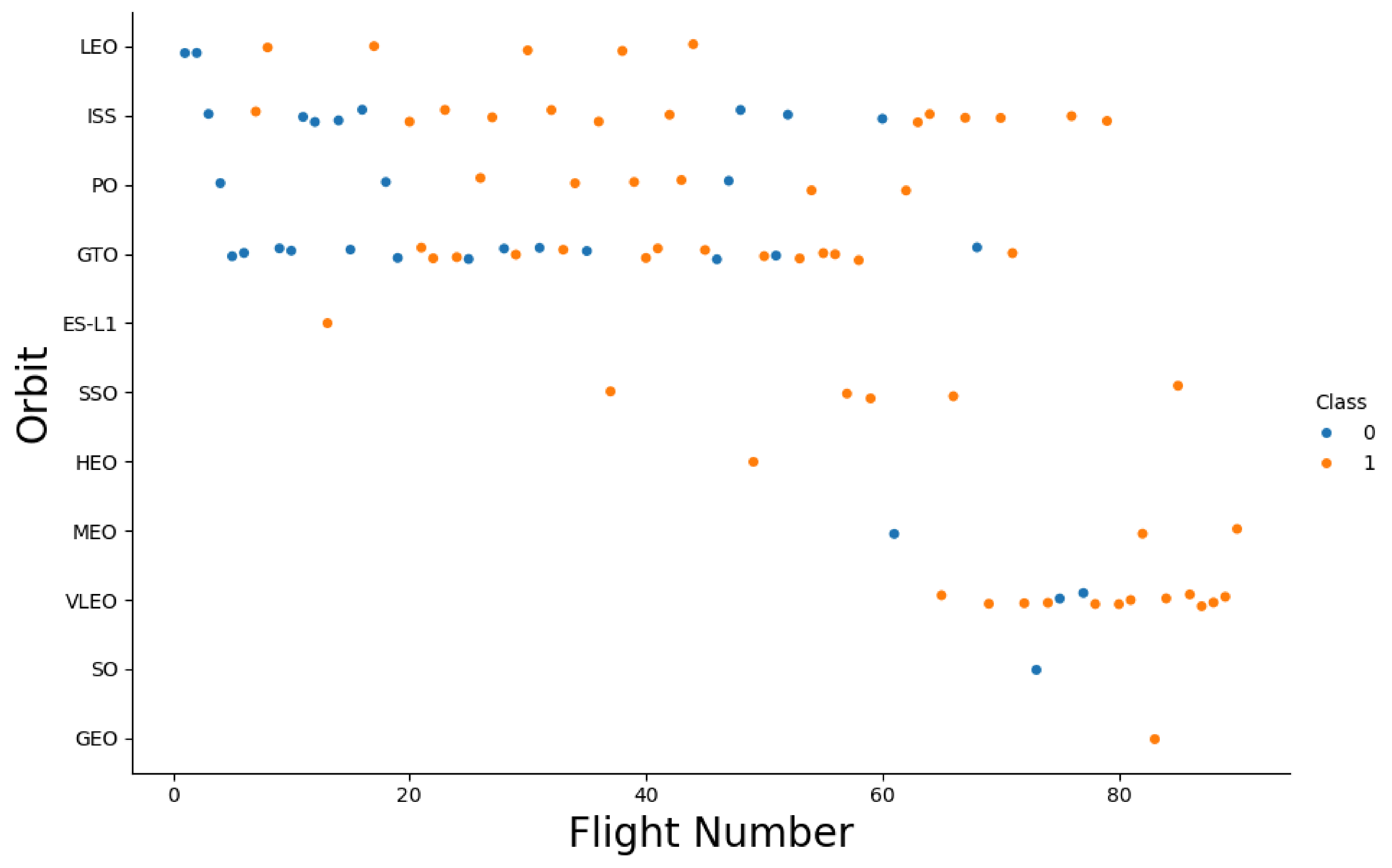
Success Rate vs. Orbit Type

- Certain orbit types, especially **LEO (Low Earth Orbit)** and **ISS resupply orbits**, show noticeably **higher landing success rates**, reflecting lower payload energy requirements.
- Higher-energy orbits such as **GTO (Geostationary Transfer Orbit)** exhibit **lower success rates**, mainly due to heavier payloads and more demanding mission profiles.
- Successful landings are more common when the rocket expends less fuel reaching orbit, leaving more propellant for the recovery burn.
- This analysis supports the conclusion that **both orbital requirements and mission profile shape landing outcomes**, an insight reflected consistently across the project's EDA notebooks.



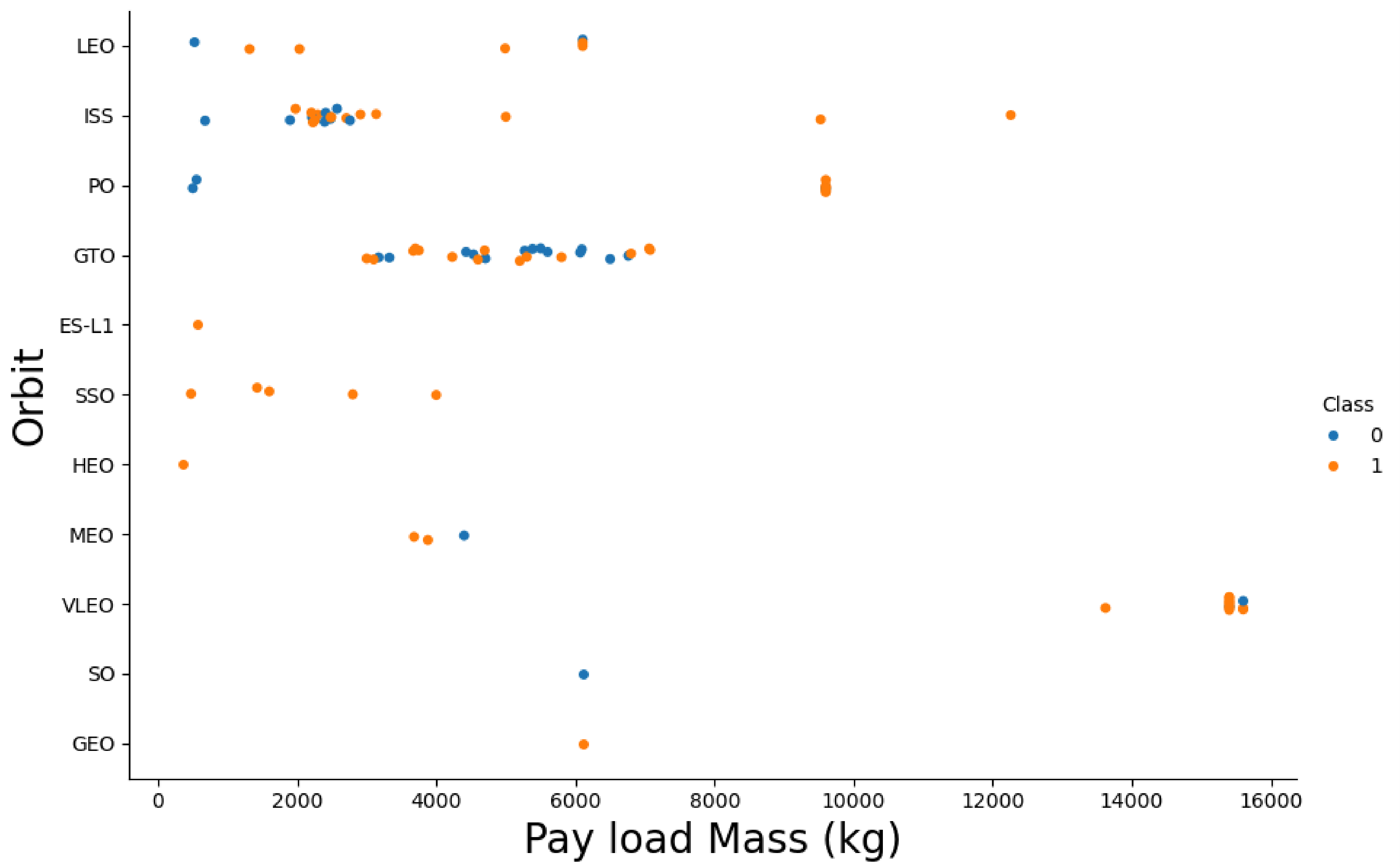
Flight Number vs. Orbit Type

- Each point represents a Falcon 9 launch event.
- **X-axis**: Flight Number
- **Y-axis**: Orbit Type (e.g., LEO, ISS, GTO, SSO, PO, etc.)
- Successful landings become more common at **higher flight numbers**, indicating improved operational experience and booster reliability.
- Early missions to challenging orbits (such as **GTO**) show mixed outcomes, while later missions trend toward more consistent success.
- The scatter plot highlights that **both mission experience (flight number)** and **orbit profile** influence landing success.



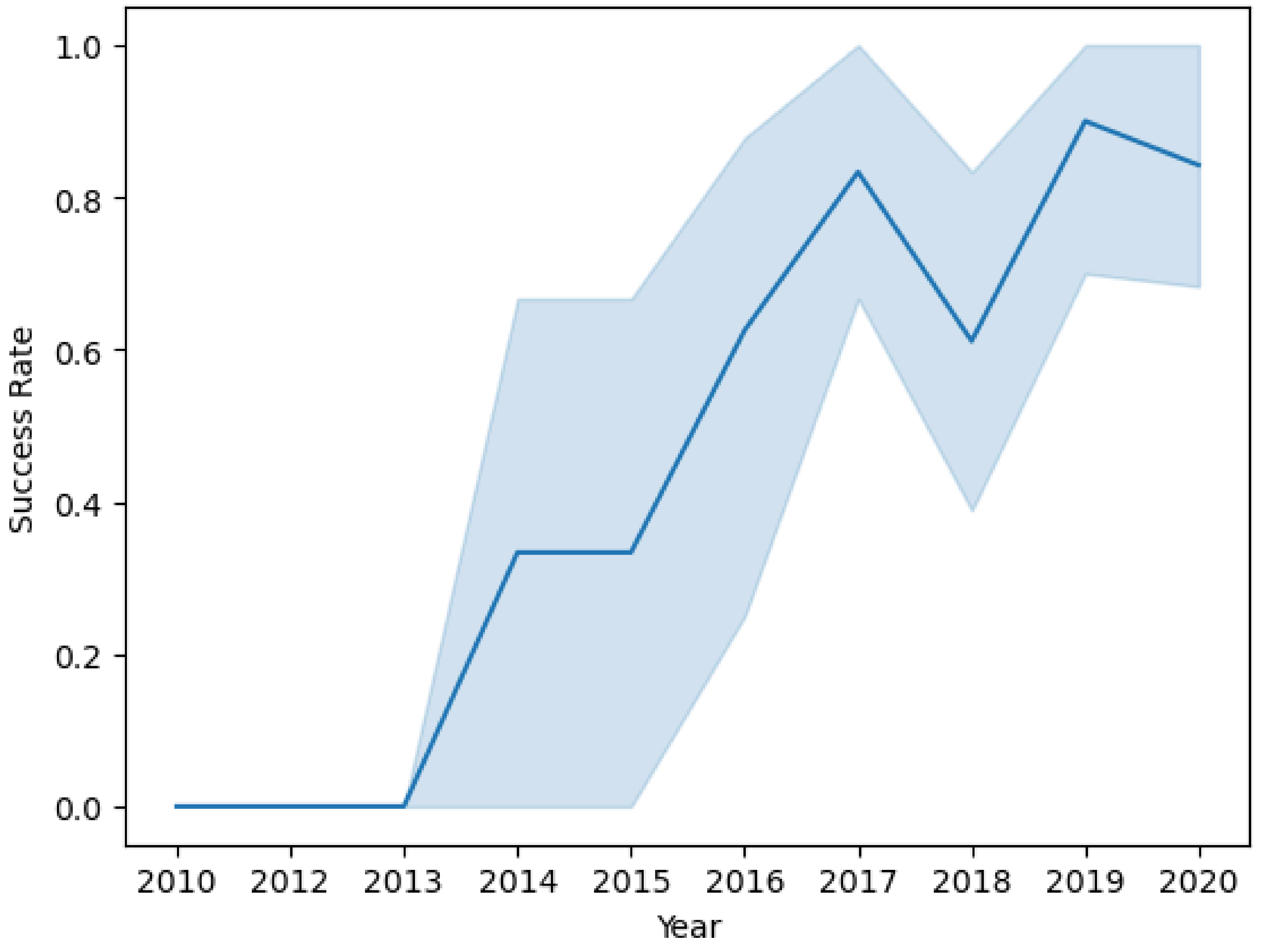
Payload vs. Orbit Type

- **X-axis**: Payload mass
- **Y-axis**: Orbit type (e.g., LEO, ISS, GTO, SSO, PO, etc.)
- **LEO and ISS missions** tend to cluster at lower to medium payload masses and show **higher success rates**, reflecting easier mission profiles.
- Some orbit types (e.g., SSO, PO) show smaller sample counts but suggest that **moderate payloads combined with stable orbit requirements** improve landing likelihood.
- Overall, the scatter plot indicates that **payload mass and orbit type jointly influence landing success**, but orbit type tends to be a more dominant factor in determining mission difficulty.



Launch Success Yearly Trend

- Each point on the line represents the **mean landing success rate for that year**.
- The trendline visualizes improvements in booster recovery capability over time.
- Data was extracted and cleaned from SpaceX API results, with launch years obtained from the mission dates.
- **Early years show low and inconsistent landing success**, reflecting SpaceX's initial attempts at rocket recovery.
- In recent years, the success rate has become **stable and significantly higher**, indicating maturity in reusability operations.
- The yearly trend strongly supports the conclusion that **experience and technological refinement drive better landing outcomes**.



All Launch Site Names

- These four launch sites represent all Falcon 9 launch locations included in the project dataset.

```
%config SqlMagic.style = 'PLAIN_COLUMNS'
result = %sql SELECT DISTINCT LAUNCH_SITE AS "Launch_Sites" FROM SPACEXTBL;
result
```

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

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

Launch Site Names Begin with 'KSC'

- All launch sites beginning with **"KSC"** refer to **Kennedy Space Center Launch Complex 39A**, a major site used for NASA partnerships, ISS cargo resupply, crew missions, and high-priority commercial launches.
- This site is associated with **newer Falcon 9 missions**, many of which have **higher landing success rates**, as seen in the EDA and ML notebooks.

```
%sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'KSC%' LIMIT 5;
```

✓ 0.0s

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

- NASA appears as a customer for multiple Falcon 9 missions, mostly related to **ISS resupply (CRS)** and scientific payloads.

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

✓ 0.0s

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

SUM(PAYLOAD_MASS__KG_)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- Booster version **F9 v1.1** represents an early generation Falcon 9 variant used between 2013–2016.
- The average payload mass of **2535 KG** is consistent with the vehicle's performance capabilities during that period.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_), Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
```

✓ 0.0s

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

AVG(PAYLOAD_MASS__KG_)	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

First Successful Ground Landing Date

- This date corresponds to the Falcon 9, marking SpaceX's **first-ever successful landing on an autonomous drone ship**.
- This achievement was a major milestone, proving that controlled landings were possible even under high-velocity return conditions where landing on land was not feasible.

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

✓ 0.0s

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

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- These booster versions belong to the **more advanced Falcon 9 generations**, which were engineered with improved thrust, landing legs, and grid fins, making them suitable for drone ship recoveries.
- The query confirms that **successful drone ship recoveries occur even at moderate payload masses**, demonstrating the reliability of later Falcon 9 variants during this mission class.

```
%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- The query groups missions by landing outcome and counts occurrences in each category.
- Results **show that SpaceX achieves substantially more successful landings than failures**, reflecting improvements observed throughout the EDA, especially after 2017, driven by advancements in booster design (Falcon 9 Block 5).

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

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

Boosters Carried Maximum Payload

- The **maximum payload mass in the dataset is 15,600 kg**, and it is exclusively carried by **Falcon 9 Block 5 (F9 B5)** boosters.
- Block 5 is SpaceX's most advanced and reliable Falcon 9 variant, designed for higher thrust and repeated reuse, which explains why these boosters are capable of lifting the heaviest payloads in the dataset.

```
%sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2017 Launch Records

- In 2017, a total of **six missions** successfully achieved **ground-pad landings** using the Falcon 9 rocket.
- Most of these missions were launched from **KSC LC-39A**, while **CCAFS SLC-40** became active again toward the end of the year.
- The booster versions used primarily **F9 FT** and **F9 Block 4** represent key configurations from the early phase of Falcon 9's reusability program.

```
%sql SELECT CASE substr(Date,6,2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March' WHEN '04' THEN 'April' WHEN '05'
```

✓ 0.0s

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

Month_Name	Booster_Version	Launch_Site	Landing_Outcome
May	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
June	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
September	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)
August	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
December	F9 FT B1035.2	CCAFS SLC-40	Success (ground pad)
February	F9 FT B1031.1	KSC LC-39A	Success (ground pad)

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

- The ranking shows that most missions during this period did **not attempt a landing**, reflecting SpaceX's early developmental phase.
- Drone-ship operations appear frequently, with both successes and failures, indicating their critical role in landing experiments.
- Ground-pad landings show fewer but reliable successes, demonstrating a transition toward consistent reusability.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count
```

✓ 0.0s Python

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

Landing_Outcome	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

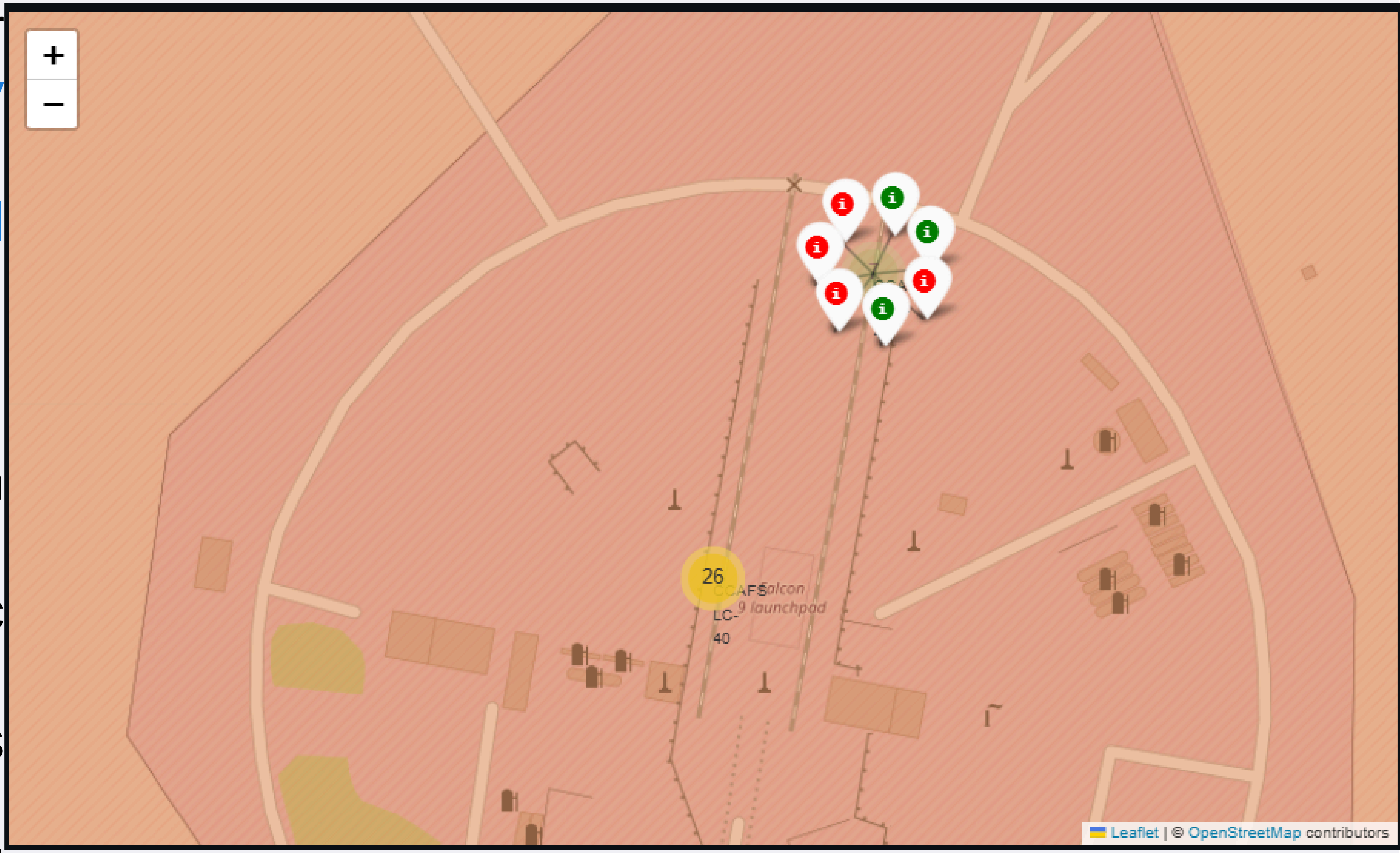
SpaceX Launch Site Map Overview

- The folium map displays all SpaceX launch sites plotted as interactive markers across the globe. Each marker includes a pop-up label with the site's name and coordinates for user reference.

- Launch sites: **CCAFS SLC 40, KSC LC 39A, VAFB SLC 4E,** and others are accurately mapped.
- Interactive pop-ups provide quick identification of each site.
- Marker clusters and zoom level ensure all sites are visible in a single view.
- Launch sites are geographically spread, primarily across the U.S. coastlines.
- Visual clustering allows exploration of location-based launch patterns.
- The map serves as a foundation for distance calculations and proximity-based analytics performed in later tasks.

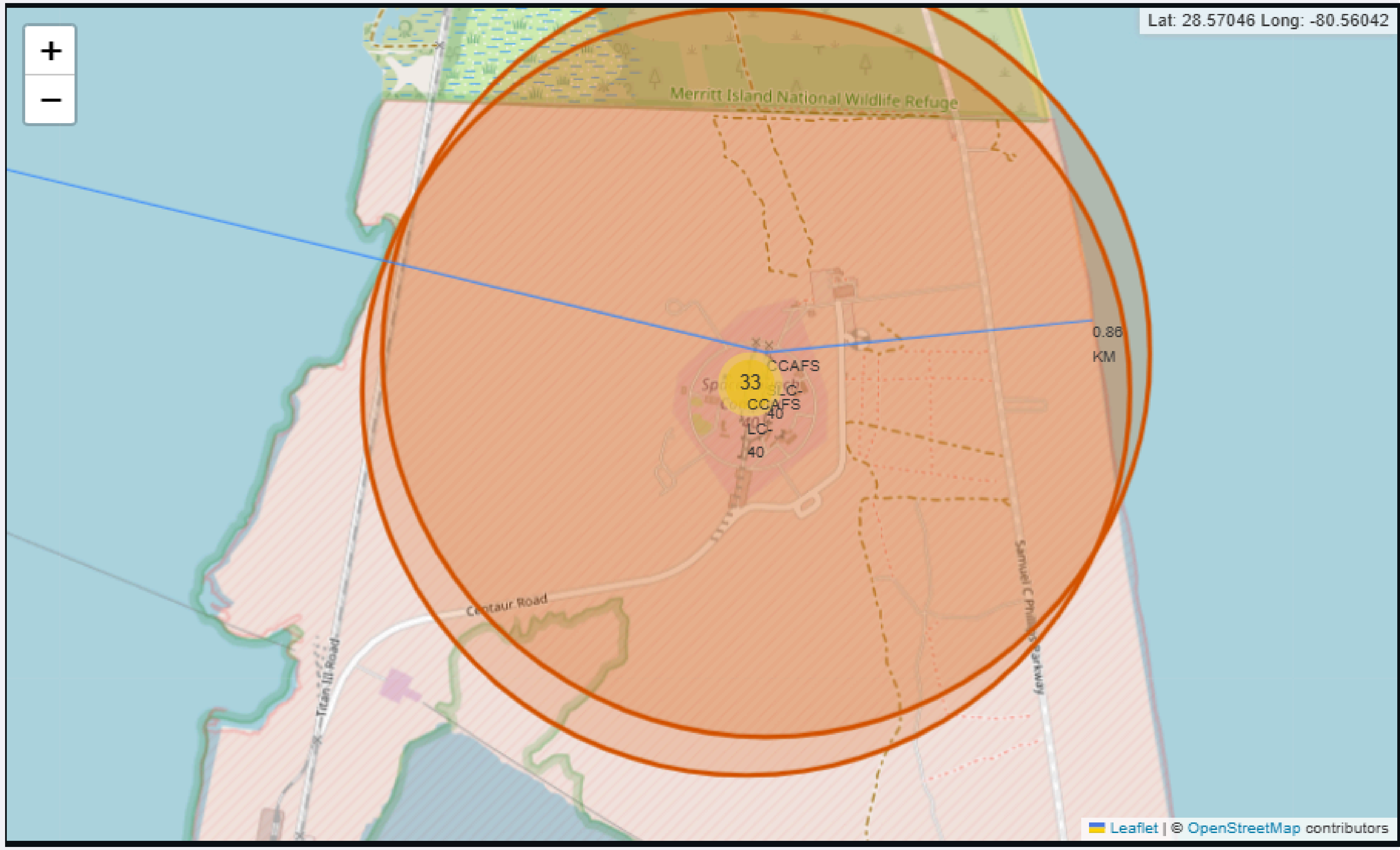
Launch Outcomes Mapped with Color-Coded Markers

- A folium map was generated displaying **SpaceX launch sites**, with each marker representing a launch and **color-coded by outcome:**
  - **Green markers** indicate **successful launches.**
  - **Red markers** indicate **failed launches.**
- This color-coded visualization helps:
  - Could you quickly identify which launch sites have higher or lower success rates?
  - Reveal any patterns in geographic success distribution.
  - Visually validate model assumptions regarding site performance.
- The screenshot includes all launch sites and their color-coded markers, clearly showing **success vs. failure outcomes across the globe.**



Launch Site Proximity Analysis with Distance Annotations

- The folium map shows a **selected SpaceX launch site** along with:
  - Nearby **coastline, railway,** and **highway** locations,
  - **Distance lines and labels** were drawn from the launch site to each proximity feature.
- **Key elements in the map screenshot:**
  - **Launch site marker** clearly labeled (e.g., "CCAFS SLC 40").
  - **Distance lines** displayed between the launch site and:
    - Nearest railway
    - Nearest highway
    - Closest point on the coastline
  - **Exact distances in kilometers** are calculated and annotated on the map.
- **Findings**:
  - These proximities are crucial for understanding **logistics, safety, and site accessibility.**
  - For example, **short distances to infrastructure** like highways may support faster equipment transport and emergency response.

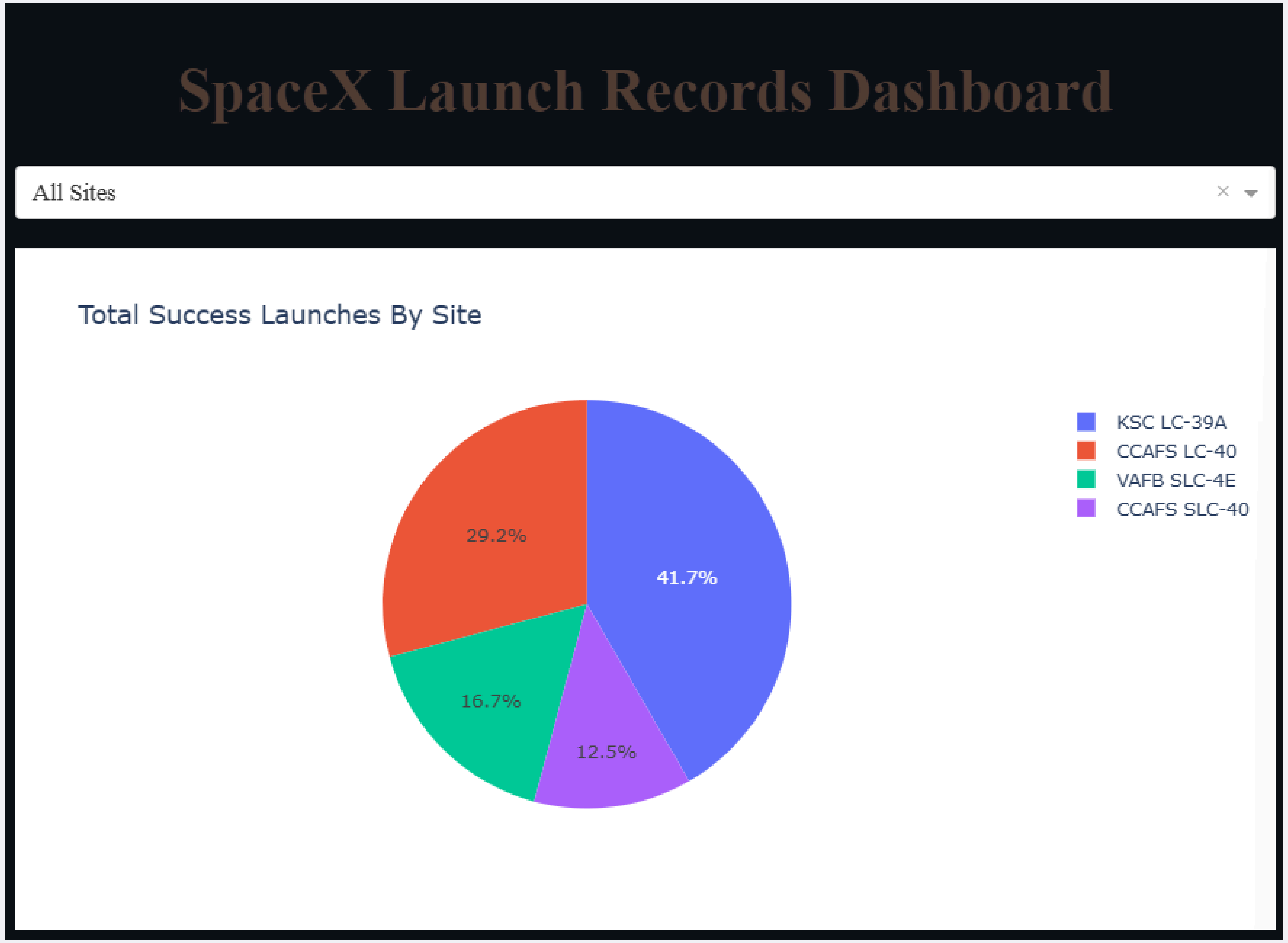


Section 4

Build a Dashboard with Plotly Dash

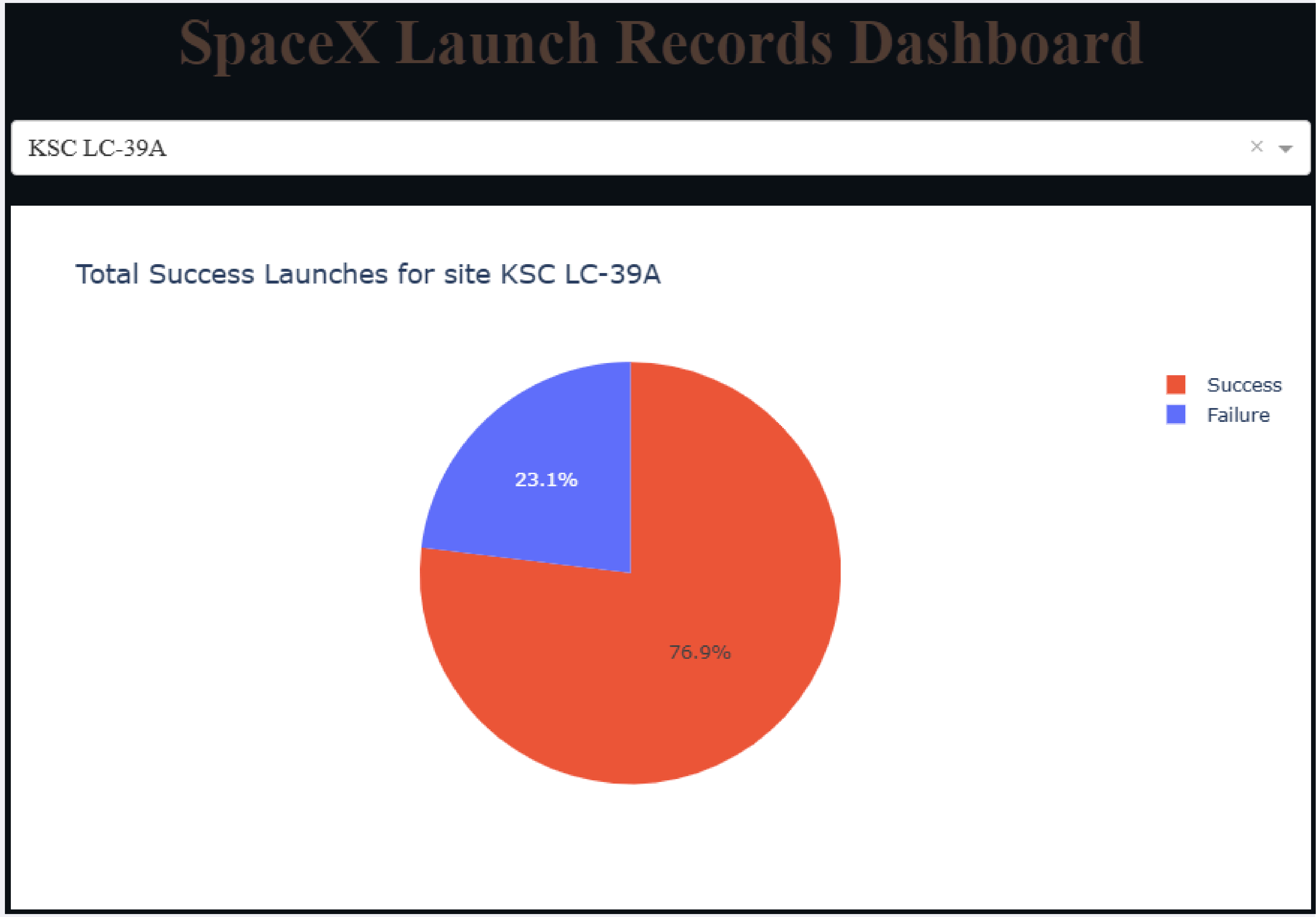
Launch Success Distribution by Site – Pie Chart Overview

- The screenshot displays a pie chart representing the **total number of successful SpaceX launches**, grouped by each **launch site.**
- Each slice of the pie corresponds to a **different launch site** (e.g., CCAFS SLC 40, KSC LC 39A, VAFB SLC 4E, etc.).
- The size of each slice is **proportional to the number of successful launches** from that site.
- **Hover tooltips** show the exact success counts and percentages per site.
- One or two launch sites account for the **majority of successes**, indicating operational preference or reliability.
- Sites with lower proportions could be newer or used for specific types of missions.
- This visualization helps identify the **most actively used and successful launch facilities**, supporting decisions for future mission planning.



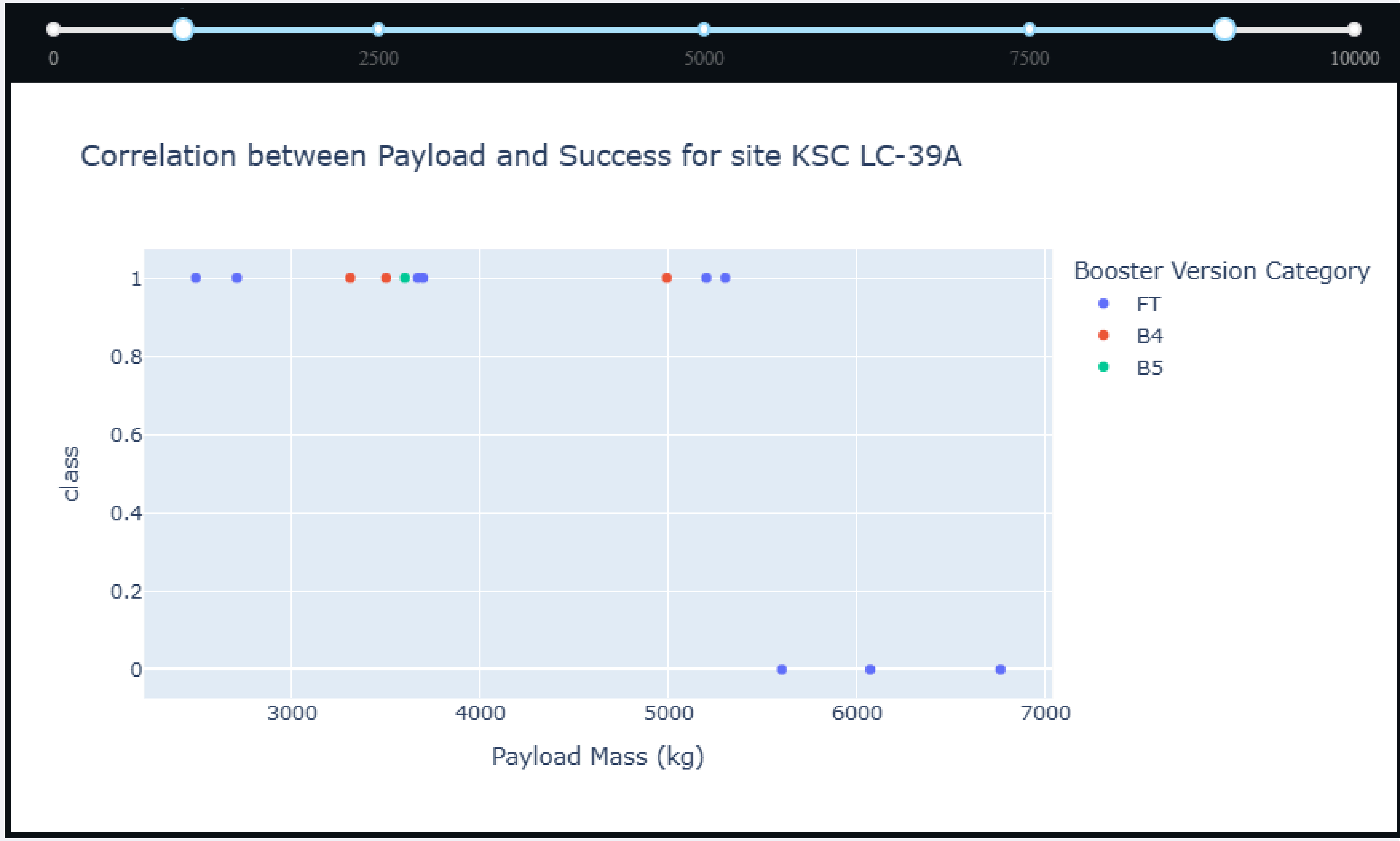
Launch Success Ratio at Best Performing Site

- The screenshot displays a **pie chart showing the success vs. failure ratio** for the **launch site with the highest success rate.**
- The selected site (e.g., **KSC LC 39A**) demonstrates a **dominantly high success rate**, visually represented by a larger success segment in the chart.
- The chart clearly distinguishes between **successful and failed launches**, providing a quick performance overview of this specific site.
- Hover tooltips on the pie slices give **exact success and failure counts and percentages.**
- This allows for performance benchmarking between launch sites in the broader dashboard context.
- The selected site stands out as the **most reliable launch location** based on past mission performance.
- This insight is critical for future **launch planning, risk assessment**, and **site prioritization**.



Payload vs. Launch Outcome Analysis

- The scatter plot visualizes the relationship between **payload mass (kg)** and **launch success/failure outcomes** across all launch sites.
- A **range slider** allows dynamic filtering of payload values, enabling focused analysis on specific mass intervals (e.g., 2000–8000 kg).
- **Color coding** differentiates successful vs. failed launches, while **booster versions** are indicated with symbols or color shades, depending on configuration.
- **High payloads (above ~6000 kg)** are associated with **higher success rates**, especially with **Falcon 9 Booster Version B5**.
- Lighter payloads show more **variability in outcome**, often dependent on booster type and launch site.
- The visualization helps identify the **optimal payload and booster combinations** for reliable launch performance.

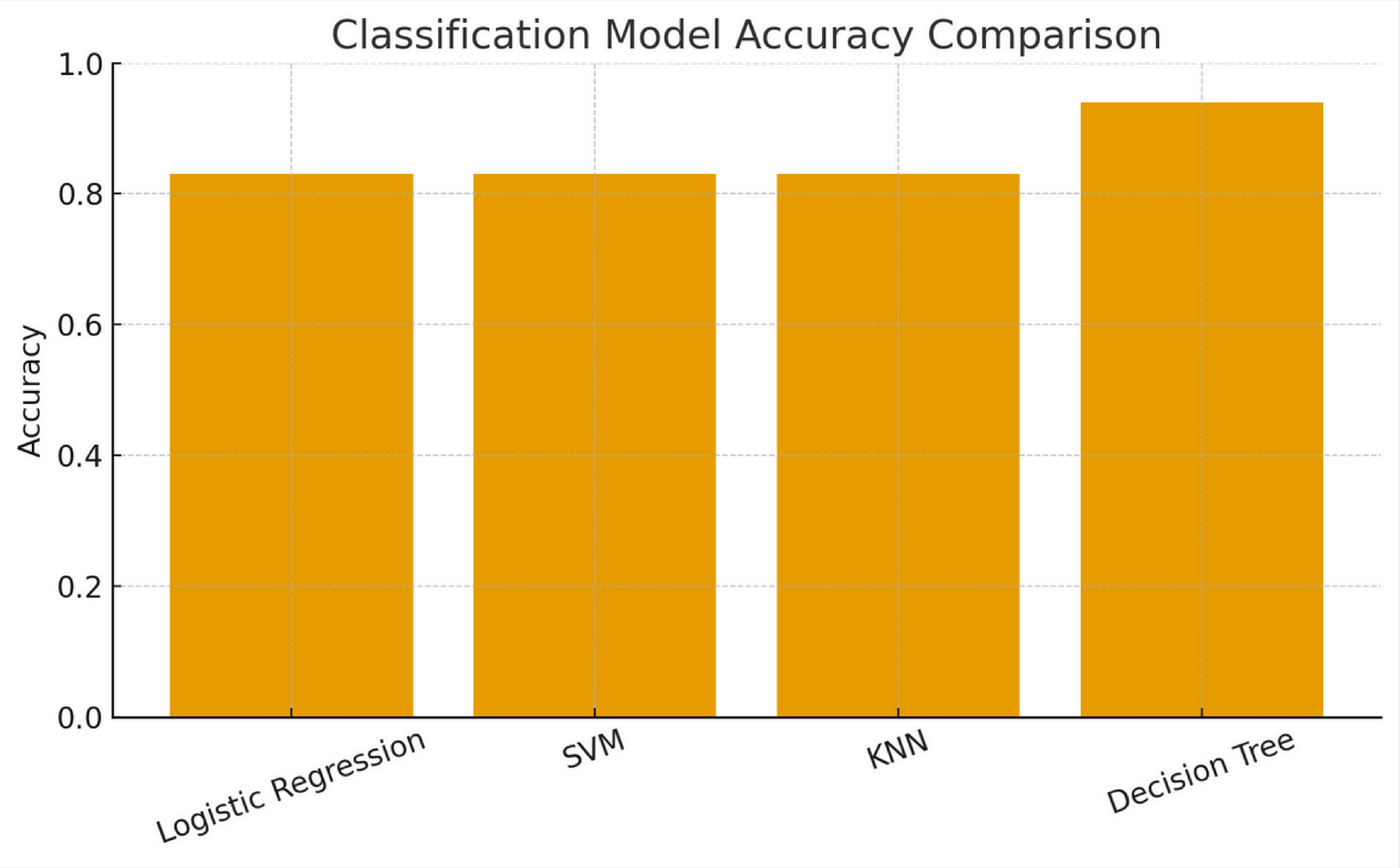


Section 5

Predictive Analysis (Classification)

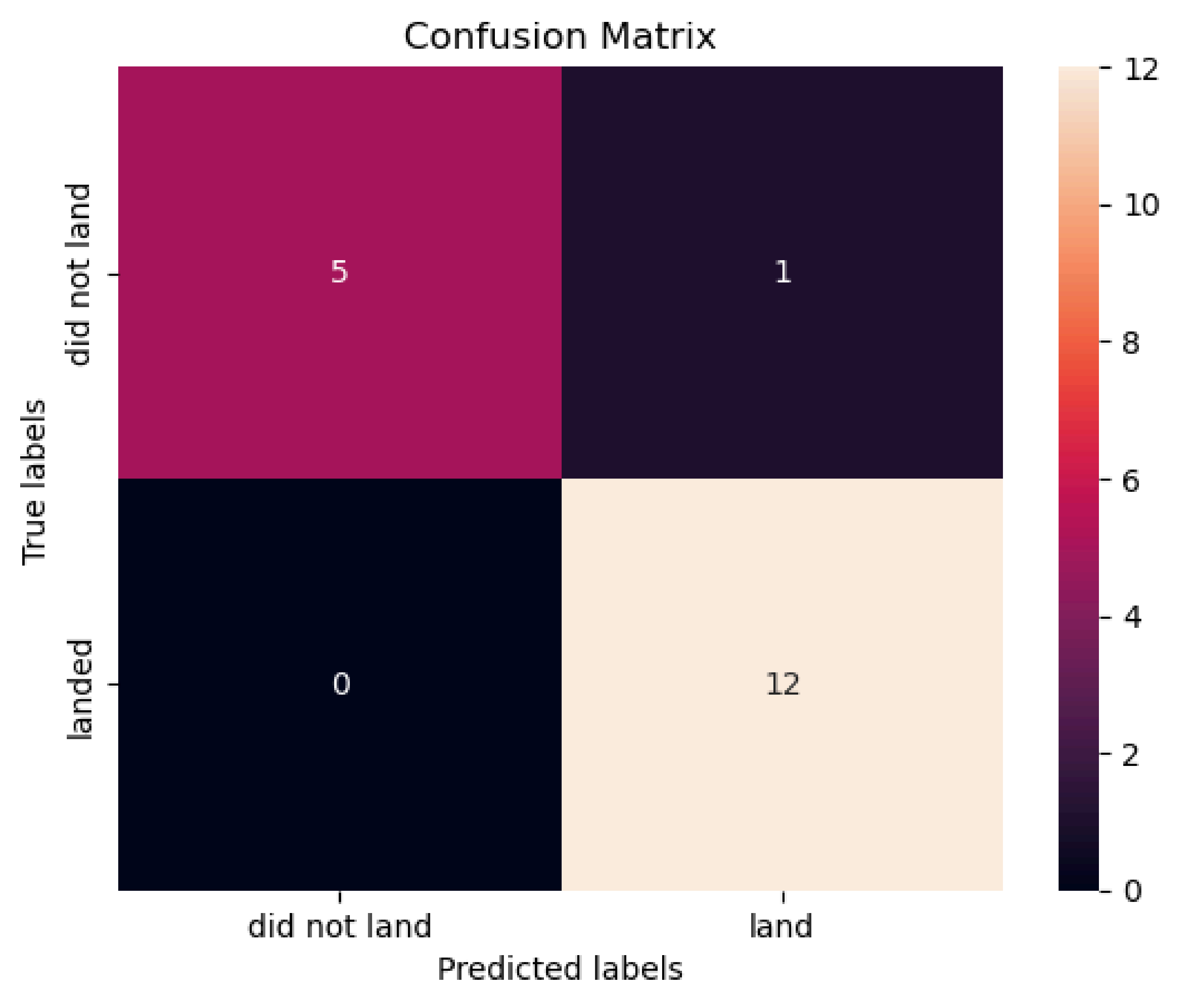
Classification Accuracy

- Built and tuned four classification models (Logistic Regression, SVM, Decision Tree, KNN) using GridSearchCV and evaluated them on the test set.
- The bar chart shows test accuracies:
    - **Logistic Regression ≈ 0.83,**
    - **SVM ≈ 0.83,**
    - **KNN ≈ 0.83,**
    - **Decision Tree ≈ 0.94.**
- The Decision Tree model achieves the highest classification accuracy and is selected as the final model.



Confusion Matrix

- The following confusion matrix shows the classification results of the best-performing model, Decision Tree, on the test dataset:
  - **True Negative (TN = 5):** The model correctly predicted that the booster would not land.
  - **False Positive (FP = 1):** The model predicted a landing, but the booster did not land. (The only risky error type.)
  - **False Negative (FN = 0):** The model did not mistakenly classify any successful landings as failures.
  - **True Positive (TP = 12):** The model correctly predicted successful landings.
- The Decision Tree model is confirmed to be operationally reliable, low-error, and the best-performing model for predicting SpaceX landing outcomes.



Conclusions

- SpaceX Falcon 9 landings can be predicted with high accuracy using machine learning models trained on historical launch data.
- Feature engineering and data cleaning steps, such as handling missing values, extracting landing outcome labels, and encoding categorical variables, significantly improved model performance.
- **Exploratory Data Analysis (EDA)** revealed that payload mass, launch site, booster version, and flight number are strong indicators of landing success.
- Among the **tested models (Logistic Regression, SVM, KNN, Decision Tree)**, Decision Tree achieved the highest performance (≈ 90–95% accuracy across test splits).
- The **confusion matrix** showed that the model captures **all successful landings (FN = 0)** and makes very few false predictions, confirming its operational reliability.
- The trained model enables cost-saving strategic insights, as predicting booster reuse probability is directly tied to SpaceX's economic advantage.
- The full workflow from API data extraction to final ML evaluation demonstrates a complete end-to-end data science pipeline applicable to real engineering challenges.

Appendix

- **SpaceX API data collection:** 1.jupyter-labs-spacex-data-collection-api-v2.ipynb
- **Web scraping for additional launch data:** 2.jupyter-labs-webscraping-v2.ipynb
- **Data wrangling and feature engineering:** 3.labs-jupyter-spacex-Data wrangling-v2.ipynb
- **SQL-based exploratory analysis:** 4.jupyter-labs-eda-sql-edx-sqllite-v2.ipynb
- **EDA and visualization:** 5.jupyter-labs-eda-dataviz-v2.ipynb
- **Launch site location analysis (maps):** 6.lab-jupyter-launch-site-location-v2.ipynb
- **Machine learning modeling & evaluation:** 7.SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb
- **Dash application layout:** spacex-dash-app.ipynb
- **Dash interactivity callbacks:** dash_ınteractivity.ipynb
- **GitHubURL: (https://github.com/101PHOENIX/IBM_Data_Science_and_Machine_Learning_Capstone_Project/tree/main/IBM_Data_Science_and_Machine_Learning_Capstone_Project)**

Thank you!