IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Renan Cardoso S.
16/06/2025

Outline

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

Executive Summary

This project analyzes SpaceX Falcon 9 launch data to support SpaceY's cost estimation and mission planning. Data was collected using the SpaceX REST API and web scraping, followed by cleaning, feature engineering, and exploratory data analysis. Key patterns were identified, including launch site performance, payload mass impact, and booster version improvements.

Interactive dashboards were developed to visualize launch success rates and payload correlations. Predictive models were built using logistic regression, decision trees, and random forests. The random forest model achieved the best performance in predicting first-stage recovery, providing actionable insights to optimize launch costs.

Introduction

Project Background and Context

The commercial space industry is rapidly evolving, driven by private companies that significantly reduce launch costs through technological innovations such as reusable rockets. SpaceX, one of the industry leaders, has demonstrated the financial advantage of reusing the Falcon 9 first stage, which substantially lowers the cost per launch compared to traditional expendable rockets. SpaceY, as a new entrant in this competitive market, aims to leverage data science to better understand the cost drivers of rocket launches and predict the likelihood of first-stage recovery, providing a competitive advantage in planning and pricing future missions.

Problems You Want to Find Answers

- What factors influence the success of Falcon 9 first-stage recovery?
- Can machine learning models accurately predict the probability of first-stage reuse based on mission parameters?
- How do launch site, payload mass, and booster version affect launch outcomes and cost estimation?
- How can data-driven insights support operational decision-making and risk assessment for future launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

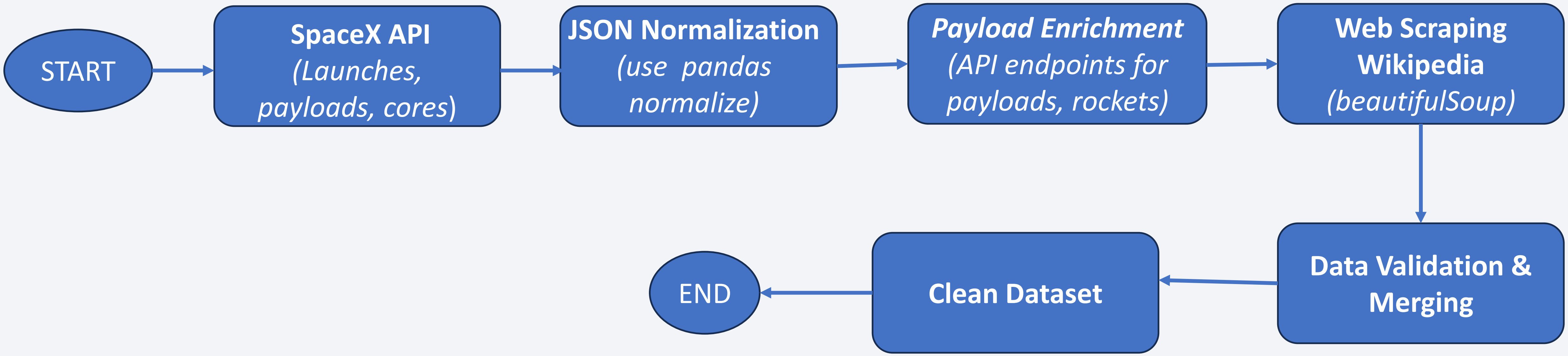
Data Collection

The data collected from two sources below that was merged, cleaned, and transformed into a unified dataset ready for exploratory data analysis and predictive modeling.

1. **SpaceX REST API**
   - ✓ The primary dataset was retrieved from the public SpaceX API which provides historical launch data.
   - ✓ Using Python's requests library, GET requests were performed to extract information on launches, rockets, payloads, cores, and landing pads, this data was normalized into tabular format using JSON normalized.
   - ✓ Additional API endpoints were called to enrich launch records with rocket version details, payload mass, orbit types, and landing outcomes.
2. **Web Scraping**
   - ✓ Supplemental launch data was extracted via web scraping from public Wikipedia pages using the BeautifulSoup library.
   - ✓ HTML tables containing Falcon 9 launch records were parsed and converted into DataFrames.
   - ✓ Scraped data was cross-referenced with API data to fill missing fields and validate consistency.

Data Collection

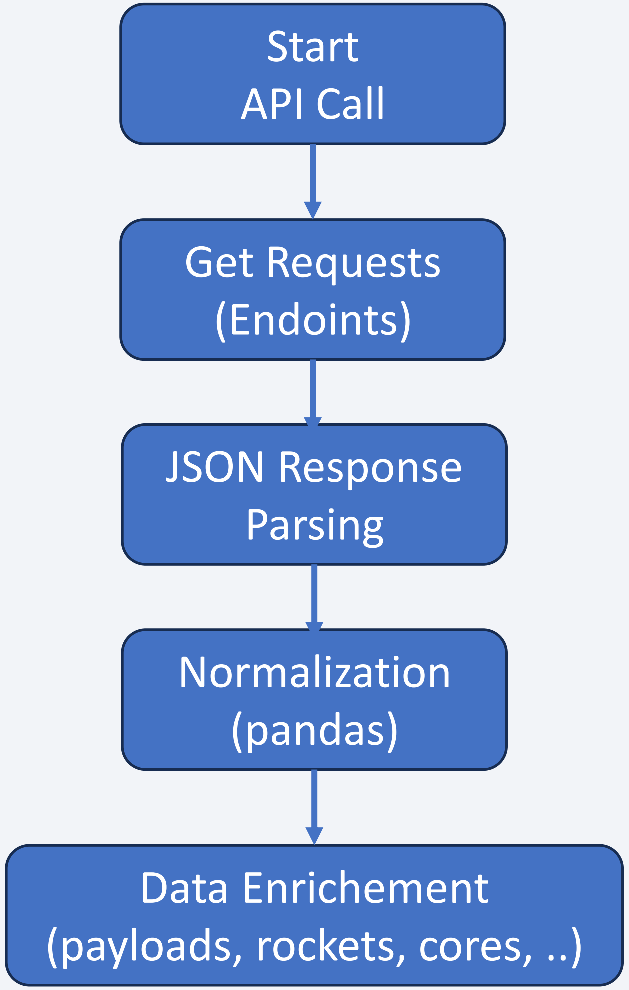
The Data Collection Process



Data Collection – SpaceX API

SpaceX API Data Collection Process

- ✓ Data collected using SpaceX REST API: https://api.spacexdata.com/v4/launches/past
- ✓ Python requests library used to perform GET requests. Retrieved JSON data was normalized using **pandas.json_normalize.**
- ✓ Additional endpoints used for data enrichment:
  - ✓ /rockets (rocket details)
  - ✓ /payloads (payload mass, orbit, customer info)
  - ✓ /cores (booster reuse info)
  - ✓ /launchpads (launch site info)
- ✓ Final dataset combined launch, payload, rocket, core, and landing information. Cleaned data prepared for further analysis.

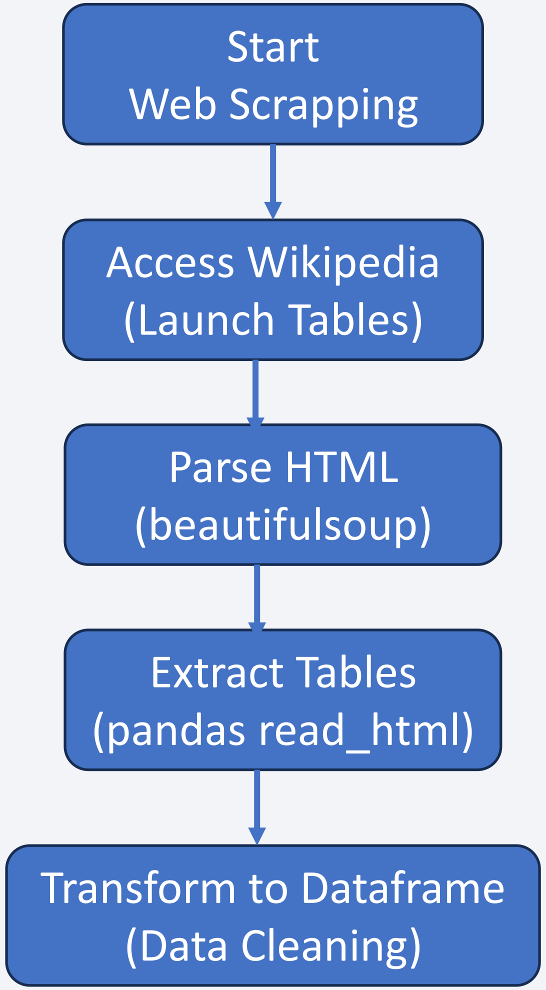


GitHub URL: *https://github.com/reynancs/spacey-capstone-project/blob/main/spacex_data_collection_api_v2.ipynb*

Data Collection - Scraping

Web Scraping Data Collection Process

- ✓ Additional launch data collected via web scraping from public Wikipedia pages.
- ✓ Python BeautifulSoup library used to parse HTML content.
- ✓ HTML tables containing Falcon 9 historical launches extracted and converted to pandas DataFrames.
- ✓ Columns parsed: (Flight Number; Date; Booster Vision; Payload Mass; Orbit; Launch Site; Outcome (landing success or failure)
- ✓ Scraped data merged with API dataset to validate, enrich, and complete missing records.

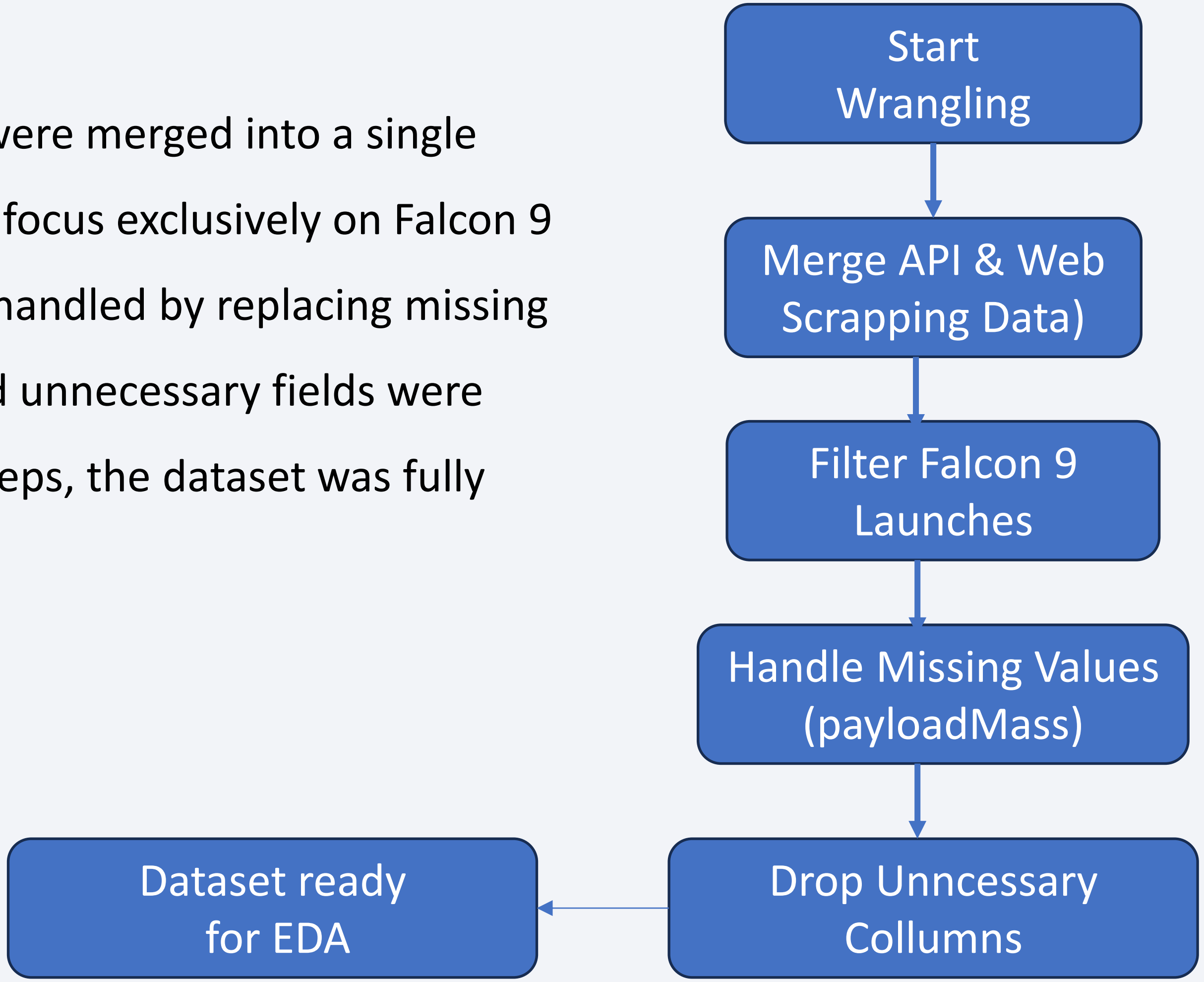


GitHub URL: *https://github.com/reynancs/spacey-capstone-project/blob/main/jupyter_labs_webscraping.ipynb*

Data Wrangling

Data Wrangling Process:

The datasets collected from the API and web scraping were merged into a single DataFrame. Non-Falcon 9 launches were filtered out to focus exclusively on Falcon 9 missions. Null values in the PayloadMass column were handled by replacing missing entries with the column mean. Redundant columns and unnecessary fields were dropped to simplify the dataset. After these cleaning steps, the dataset was fully prepared for exploratory data analysis and modeling.



GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/labs_jupyter_spacex_Data_wrangling_v2.ipynb

EDA with Data Visualization

During the exploratory data analysis phase, several visualizations were generated to explore key relationships and trends in the Falcon 9 launch dataset. The charts were grouped by type of visualization based on their objectives:

i. **Scatter and Categorical Point Plots**

These visualizations were used to analyze relationships between numerical and categorical variables, allowing identification of patterns associated with launch outcomes. (*FlightNumber vs PayloadMass; FlightNumber vs Orbit Type).*

ii. **Bar Plot**

Used to evaluate proportions and categorical success rates *(Success Rate by Orbit Type).*

ii. **Line Plot**

Used to analyze temporal evolution of mission success *(Launch Success Yearly Trend).*

These visualizations provided insights into how flight experience, payload mass, launch site, orbit type, and time evolution affect Falcon 9 first stage recovery outcomes. The results guided feature selection for subsequent machine learning modeling.

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/jupyter_labs_eda_dataviz_v2.ipynb

EDA with SQL

SQL Queries Summarized that was performed:

- **EDA 1: Queried the unique launch site names.**

SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

- **EDA 2: Retrieved 5 records where the launch site names start with 'CCA'.**

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

- **EDA 3: Calculated the total payload mass carried by boosters launched for customer 'NASA (CRS)'.**

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

- **EDA 4: Calculated the average payload mass for booster version 'F9 v1.1'.**

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

- **EDA 5: Retrieved the earliest date of a successful ground pad landing.**

SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/jupyter_labs_eda_sql_coursera_sqllite.ipynb

EDA with SQL

- **EDA 6: Retrieved the earliest date of a successful ground pad landing.**

SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)’;

- **EDA 7: Counted the number of missions grouped by mission outcome (success/failure).**

%sql SELECT Mission_Outcome AS 'Mission Outcome', COUNT(*) AS "Value" FROM SPACEXTABLE GROUP BY Mission_Outcome

- **EDA 8: Listed booster versions that carried the maximum payload mass using a subquery.**

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

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/jupyter_labs_eda_sql_coursera_sqllite.ipynb

EDA with SQL

EDA 9: Retrieved month names, booster versions, launch site and failure landing outcomes on drone ship during 2015.

SELECT CASE substr(Date, 6, 2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March' WHEN '04' THEN 'April' WHEN '05' THEN 'May' WHEN '06' THEN 'June' WHEN '07' THEN 'July' WHEN '08' THEN 'August'
WHEN '09' THEN 'September' WHEN '10' THEN 'October' WHEN '11' THEN 'November' WHEN '12' THEN 'December'
END AS Month_Name, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome LIKE 'Failure%'   AND Landing_Outcome LIKE '%drone ship%'

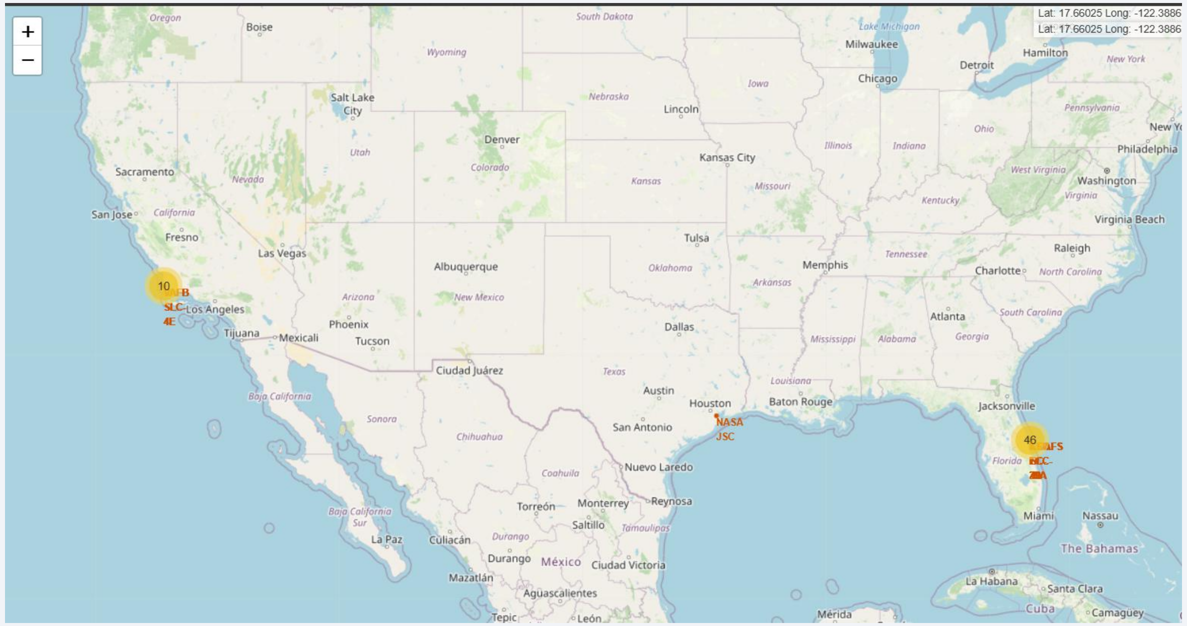
EDA 10: Ranked landing outcomes (success/failure) between two dates.

SELECT Landing_Outcome, COUNT(*) AS Outcome_Count, RANK() OVER (ORDER BY COUNT(*) DESC) AS Outcome_Rank FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Rank;

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/jupyter_labs_eda_sql_coursera_sqllite.ipynb

Build an Interactive Map with Folium

In this section, several map objects were created to visualize the SpaceX launch sites using Folium. Circles and labeled markers were added for each site to highlight geographic locations and site names. A reference point for NASA Johnson Space Center was also included to center the initial map. Finally, a Marker Cluster was created where each individual launch was plotted using colored markers — green for successful launches and red for failures — providing an interactive view of launch outcomes per site.



GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/lab_jupyter_launch_site_location_v2.ipynb

Build a Dashboard with Plotly Dash

An interactive dashboard was developed using Plotly Dash to enable real-time analysis of SpaceX launch data.

Plots and Interactions Added:

❖ **Launch Site Drop-down Menu**

The first interactive element implemented, allowing users to select either all launch sites or a specific one. This selection directly filters the displayed data, providing flexibility to analyze success rates globally or per site.

❖ **Pie Chart**

Added to to visualize launch success. When "All Sites" is selected, the pie chart displays the total number of successful launches per launch site. If a specific launch site is selected, the chart shows the proportion of successful and failed launches for that site, offering quick insights into site performance.

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/spacex_dash_app.py

Build a Dashboard with Plotly Dash

❖ **Payload Range Slider**

To analyze the relationship between payload mass and launch outcomes, a Payload Range Slider was incorporated. This slider allows users to define a specific payload range, dynamically filtering the dataset for further analysis. The slider provides the ability to isolate and examine payload intervals that may impact success rates.

❖ **Scatter Plot**

Finally, a Scatter Plot was included to visualize the correlation between payload mass and launch success across booster versions. The plot updates dynamically based on the selected launch site and payload range, with booster versions represented by different colors. This enables comprehensive multivariate analysis, revealing how payload weight, booster version, and launch site collectively influence mission outcomes.

GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

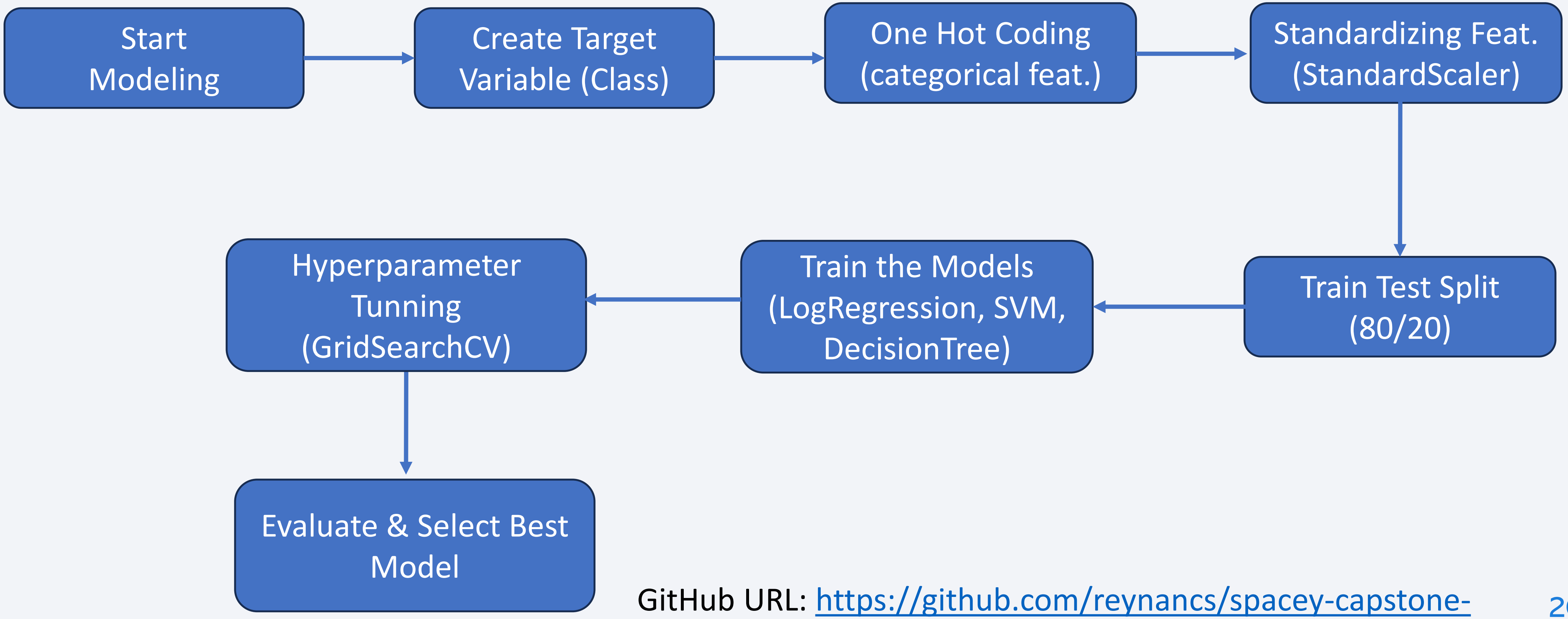
First, a binary target column (Class) was created to represent landing outcome: 1 for successful landing, 0 for failure. The dataset was then prepared by converting categorical variables using **one-hot encoding** and standardizing the numerical features using **StandardScaler** to ensure model stability and performance.
The dataset was split into training and testing sets (80% training, 20% testing) to evaluate model generalization.

Three classification algorithms were applied:

- **Logistic Regression;**
- **Support Vector Machine (SVM);**
- **Decision Tree Classifier.**
- For each model, hyperparameter tuning was performed using **GridSearchCV.** Models were evaluated using the **test dataset accuracy score**. After evaluation, the **Decision Tree Classifier** demonstrated the best performance, achieving the highest test accuracy among the tested models.

Predictive Analysis (Classification)

The Modeling Flowchart:



GitHub URL: https://github.com/reynancs/spacey-capstone-project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5_v1.ipynb

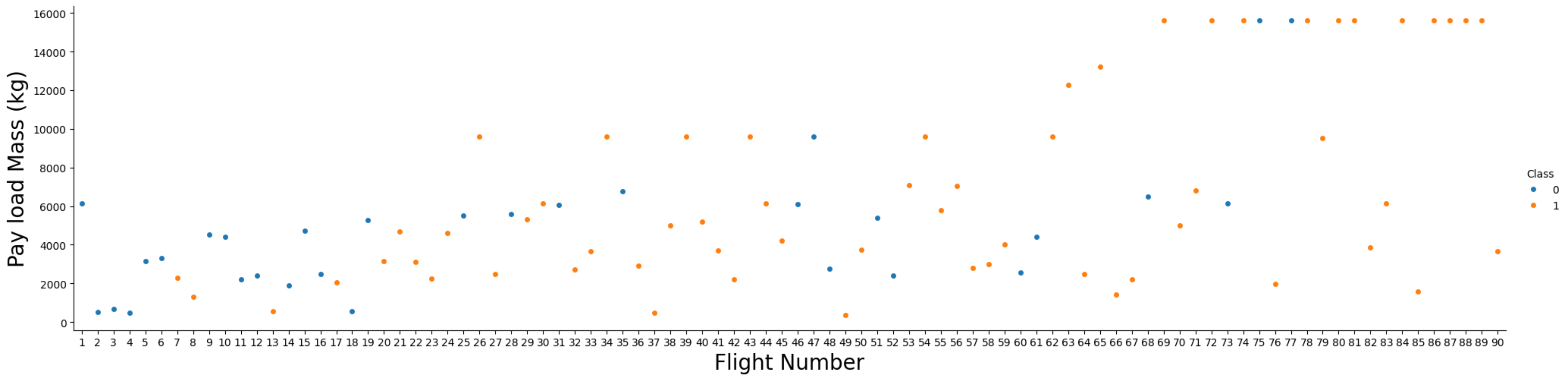
Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

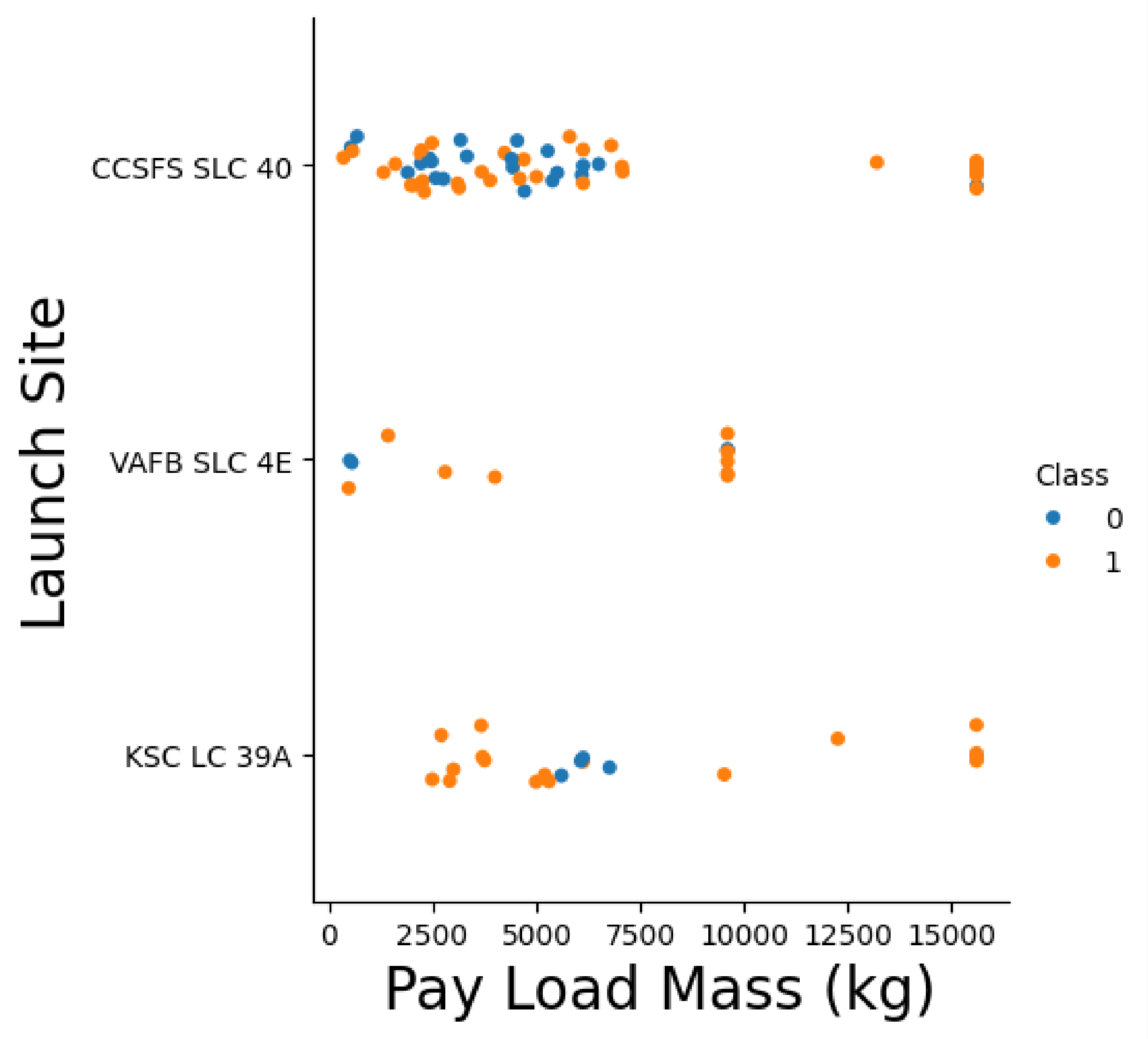


The plot shows that as Flight Number increases, the success rate improves across all launch sites, reflecting the learning curve effect. CCAFS SLC 40 has the highest number of launches with consistent success in later flights. KSC LC 39A and VAFB SLC 4E show fewer launches but follow a similar trend of increased success with experience.

Payload vs. Launch Site

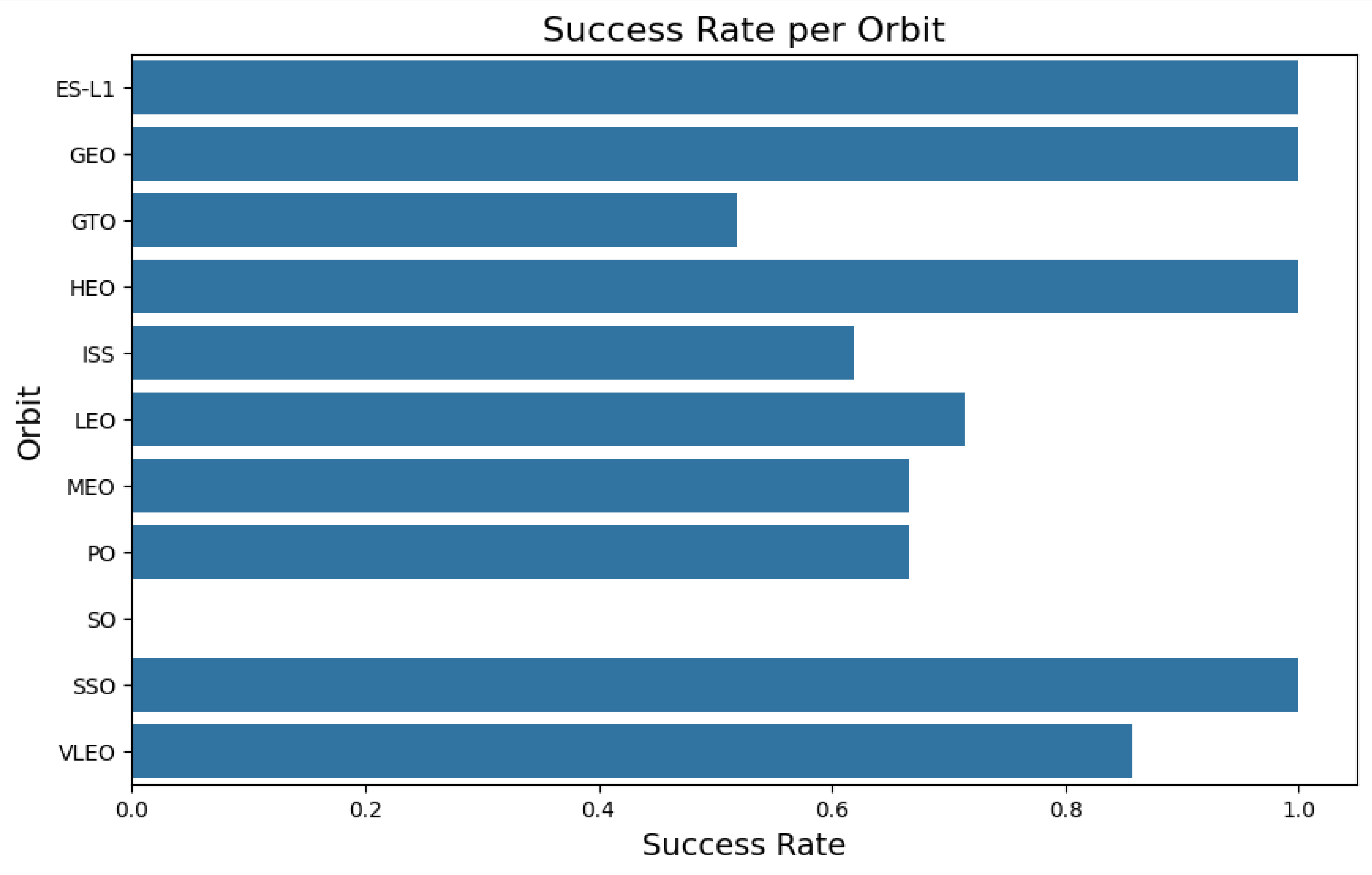
The plot shows that CCAFS SLC 40 and KSC LC 39A handled a wide range of payload masses, including heavy payloads above 10,000 kg.

In contrast, VAFB SLC 4E had no launches with heavy payloads, operating only with lighter missions.

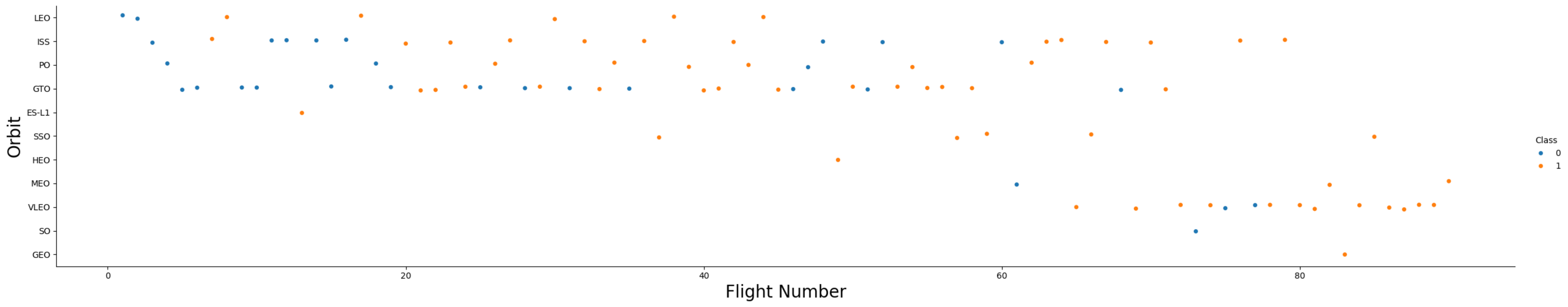


Success Rate vs. Orbit Type

The bar chart shows that most orbits achieved high success rates, especially ES-L1, GEO, HEO, SSO, and PO with near or full 100% success. In contrast, GTO exhibits a notably lower success rate, indicating higher recovery challenges for missions targeting this orbit

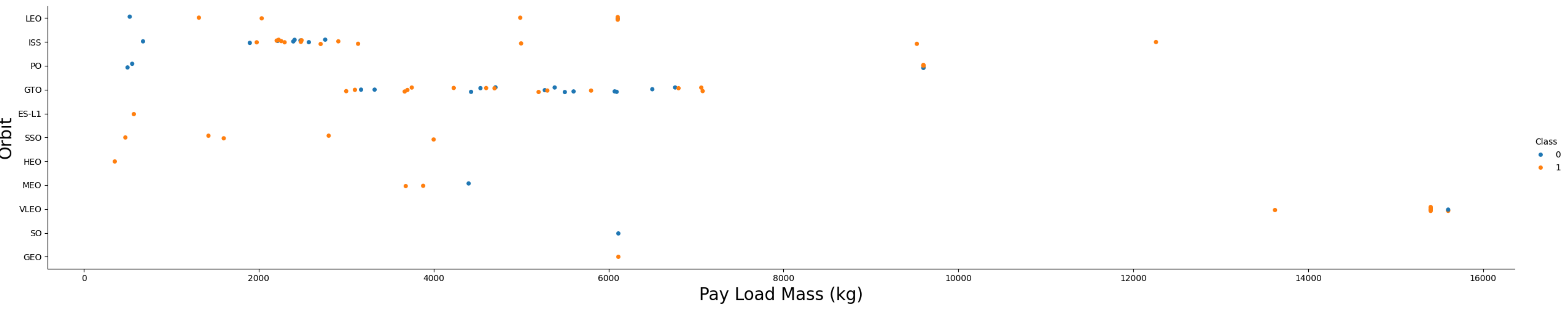


Flight Number vs. Orbit Type



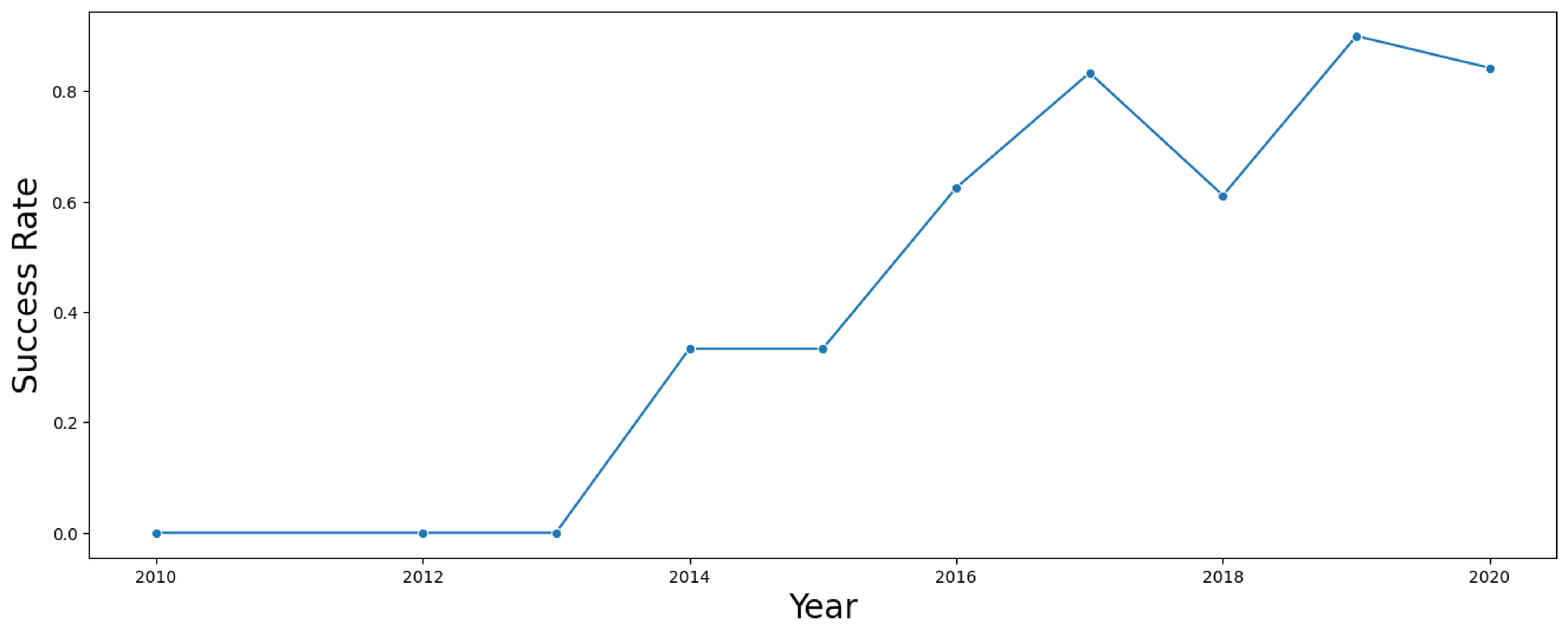
For the LEO orbit, success rates improve with higher flight numbers, indicating experience positively impacts outcomes. In contrast, for GTO orbit, there’s no clear relationship between flight experience and success, suggesting GTO remains technically more challenging regardless of accumulated launches.

Payload vs. Orbit Type



With heavier payloads, orbits like Polar, LEO, and ISS show higher success rates. For GTO, both successes and failures are observed across the payload range, making it harder to identify a clear pattern between payload mass and landing outcome for this orbit.

Launch Success Yearly Trend



The success rate consistently improved from 2013 to 2017, with some fluctuations after 2015. The learning curve is evident, as operational experience led to steady performance gains, reaching high success levels by 2019.

All Launch Site Names

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

 * sqlite:///my_data1.db
Done.
 Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
```

The query retrieved the distinct launch sites used in the missions. Four unique launch sites were identified: **CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40** (with duplicate naming due to slight variations). These sites represent the primary locations for SpaceX Falcon 9 launches.

Launch Site Names Begin with 'CCA'

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

The query filtered launch sites that start with 'CCA', returning 5 records from **CCAFS LC-40**. These records include early Falcon 9 missions, with payloads mostly targeting the LEO orbit and missions for NASA and SpaceX. Some early missions show no landing attempt or parachute failures

Total Payload Mass

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

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

The query calculated the total payload mass carried for NASA (CRS) missions, resulting **in 45,596 kg**. This reflects the cumulative payload transported by SpaceX under NASA's Commercial Resupply Services (CRS) program.

Average Payload Mass by F9 v1.1

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

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

AVG(PAYLOAD_MASS__KG_)
2928.4

The query calculated the average payload mass carried by the F9 v1.1 booster version, resulting **in 2,928.4 kg**.

This provides insight into the typical payload capacity managed by this early Falcon 9 version.

First Successful Ground Landing Date

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

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

he query identified that the first successful ground pad landing occurred on **2015-12-22.** This milestone marked a significant achievement in SpaceX's landing and reusability capabilities.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[ ] %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

 * sqlite:///my_data1.db
Done.
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2
```

The query returned booster versions that successfully landed on drone ships carrying payloads between 4000 kg and 6000 kg. The boosters identified include:**F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2.**

Total Number of Successful and Failure Mission Outcomes

```
[ ] %sql SELECT Mission_Outcome AS 'Mission Outcome', COUNT(*) AS "Value" FROM SPACEXTABLE GROUP BY Mission_Outcome

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

Mission Outcome	Value
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

The query aggregated mission outcomes, revealing:

- 98 Successful missions;
- 1 Failure (in flight)
- 1 Success (payload status unclear)
- 1 additional Success entry (likely duplicate or data inconsistency)

This indicates a high overall success rate for SpaceX missions.

Boosters Carried Maximum Payload

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

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

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

The query identified booster versions that carried the maximum payload mass. All boosters listed belong to the F9 B5 series, including variants such as **B1048.1, B1049.4, B1051.3, B1056.4, B1058.3, B1060.9**, among others. These boosters represent SpaceX's most advanced Falcon 9 Block 5 configuration, optimized for high payload capacity.

2015 Launch Records

The query retrieved failure landings on drone ships during 2015. Two failures were identified:

- **January:** Booster F9 v1.1 B1012 at CCAFS LC-40
- **April**: Booster F9 v1.1 B1015 at CCAFS LC-10

This highlights early challenges with drone ship landings in the initial Falcon 9 missions.

```
[ ] %%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' THEN 'May'
            WHEN '06' THEN 'June'
            WHEN '07' THEN 'July'
            WHEN '08' THEN 'August'
            WHEN '09' THEN 'September'
            WHEN '10' THEN 'October'
            WHEN '11' THEN 'November'
            WHEN '12' THEN 'December'
        END AS Month_Name,
        Landing_Outcome,
        Booster_Version,
        Launch_Site
    FROM
        SPACEXTABLE
    WHERE
        substr(Date, 0, 5) = '2015'
        AND Landing_Outcome LIKE 'Failure%'
        AND Landing_Outcome LIKE '%drone ship%'
```

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

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

The query ranked landing outcomes from 2010-06-04 to 2017-03-20.

- **No attempt** was the most frequent outcome (10 times).
- **Success (drone ship)** occurred 5 times, followed by Failure (drone ship) and Success (ground pad), both with 4 occurrences.
- **Controlled landings** and parachute failures were less frequent.

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

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

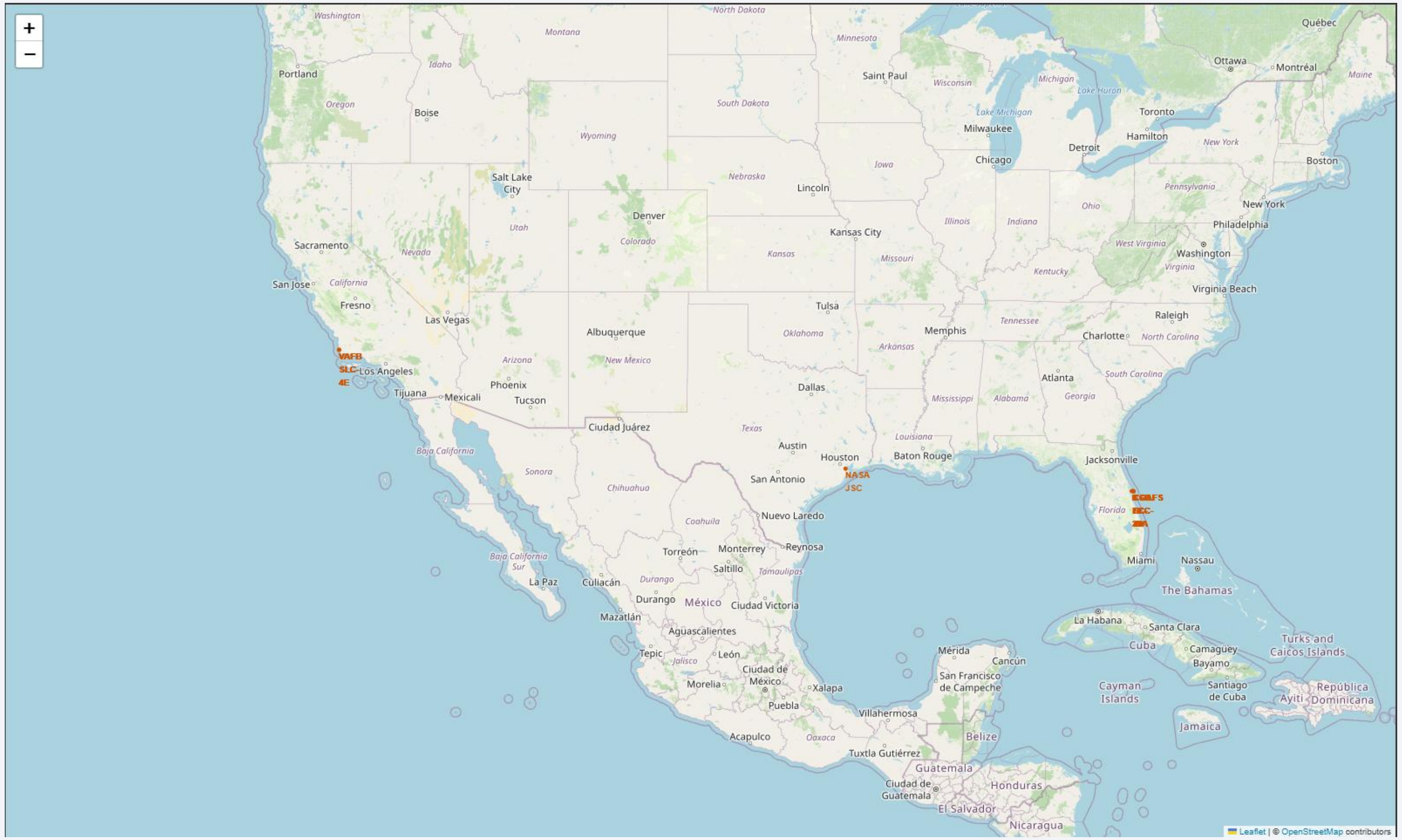
Landing_Outcome	Outcome_Count	Outcome_Rank
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	4
Controlled (ocean)	3	4
Uncontrolled (ocean)	2	6
Failure (parachute)	2	6
Precluded (drone ship)	1	8

Section 3

Launch Sites Proximities Analysis

Global Map with Launch Sites

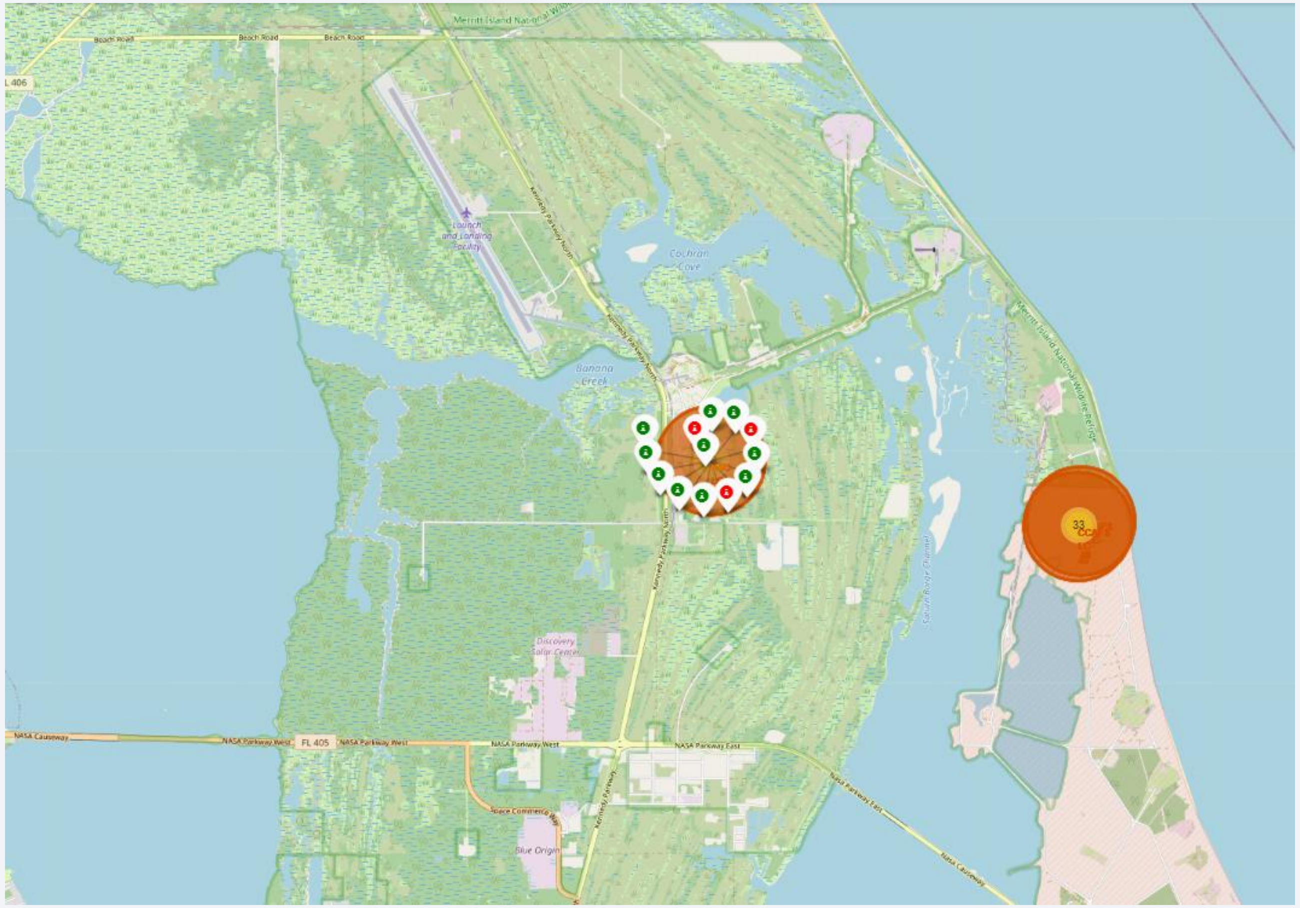
- ✓ A global Folium map was generated and a screenshot was taken showing all SpaceX launch sites marked.
- ✓ All launch sites are located **near coastlines**, allowing for safe overwater flight paths and booster recovery operations.
- ✓ All locations are clearly indicated with circles and labeled markers, providing a global view of SpaceX's operational launch facilities.



Markers the success/failed launches

The map shows the KSC-LC-39A launch site with clustered markers representing individual launch outcomes. Green markers indicate successful landings (class=1), and red markers represent failures (class=0). The marker cluster effectively handles overlapping coordinates due to multiple launches occurring at the same site.

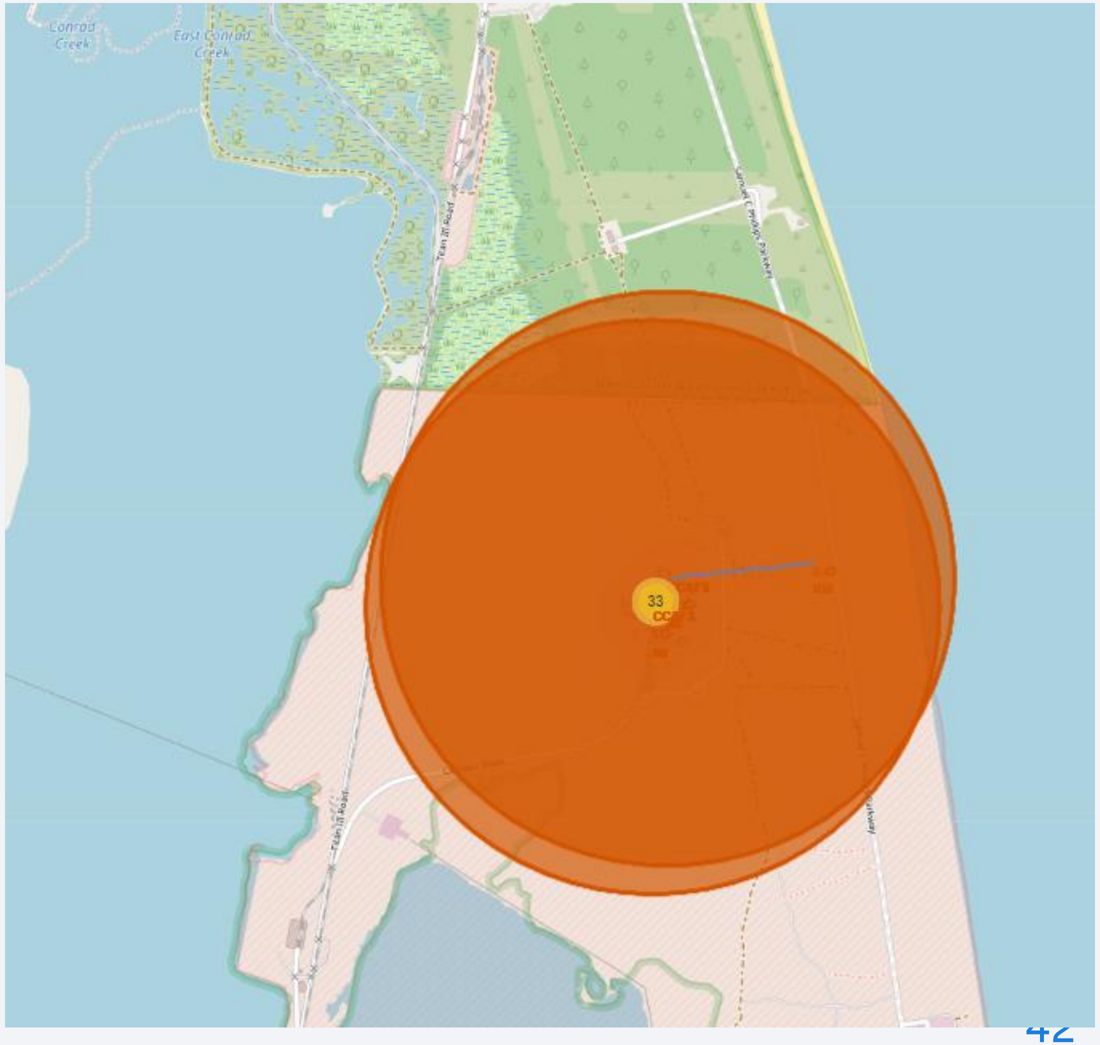
The visualization highlights that **most launches from KSC-LC-39A resulted in success**, as indicated by the larger number of green markers.



Launch Site Proximity Analysis

Using the Folium map, distances from the CCAFS LC-40 launch site to nearby infrastructures were calculated. The launch site is:

- Very close to the coastline, at approximately **0.49 km**, enabling safe rocket return paths over the ocean.
- In proximity to highways and railways, allowing efficient ground transportation of rocket stages and equipment.
- Located at a safe distance from populated cities, reducing risks in case of launch failures or anomalies.

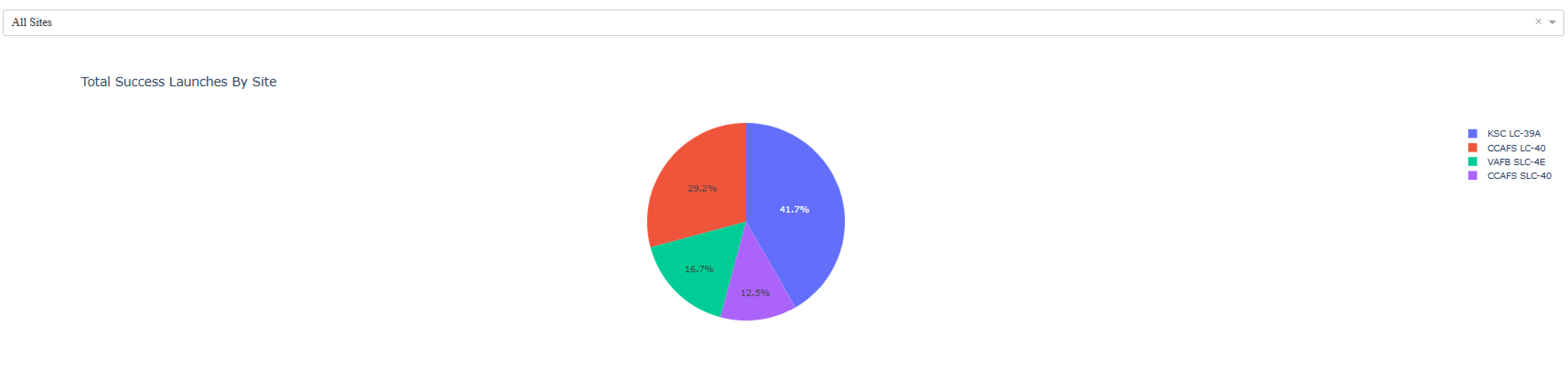


The proximity to multiple infrastructures highlights that site selection balances both operational logistics and safety considerations.

Section 4

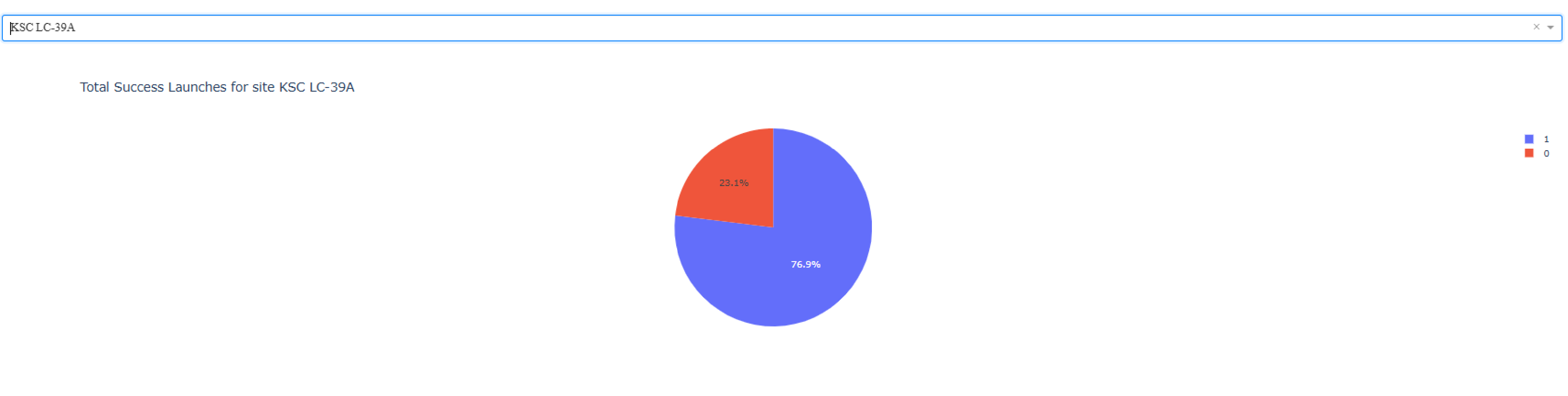
Build a Dashboard with Plotly Dash

Launch Success Count for All Sites



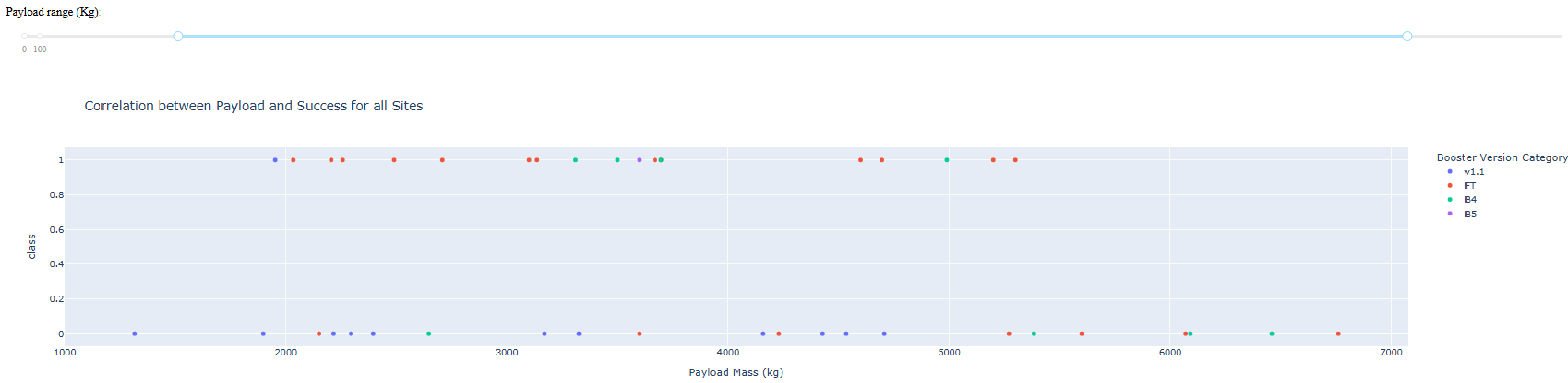
The chart highlights that most successful launches are concentrated in Florida launch sites (KSC and CCAFS), confirming them as SpaceX's primary operational hubs.

Launch Sites with Highest Launch Success Ratio



The pie chart shows the success ratio for KSC LC-39A, which demonstrates the highest launch success rate among all sites. This confirms KSC LC-39A as one of the most reliable sites for SpaceX operations, reflecting both technical maturity and operational efficiency.

Payload vs. Launch Outcome for All Sites

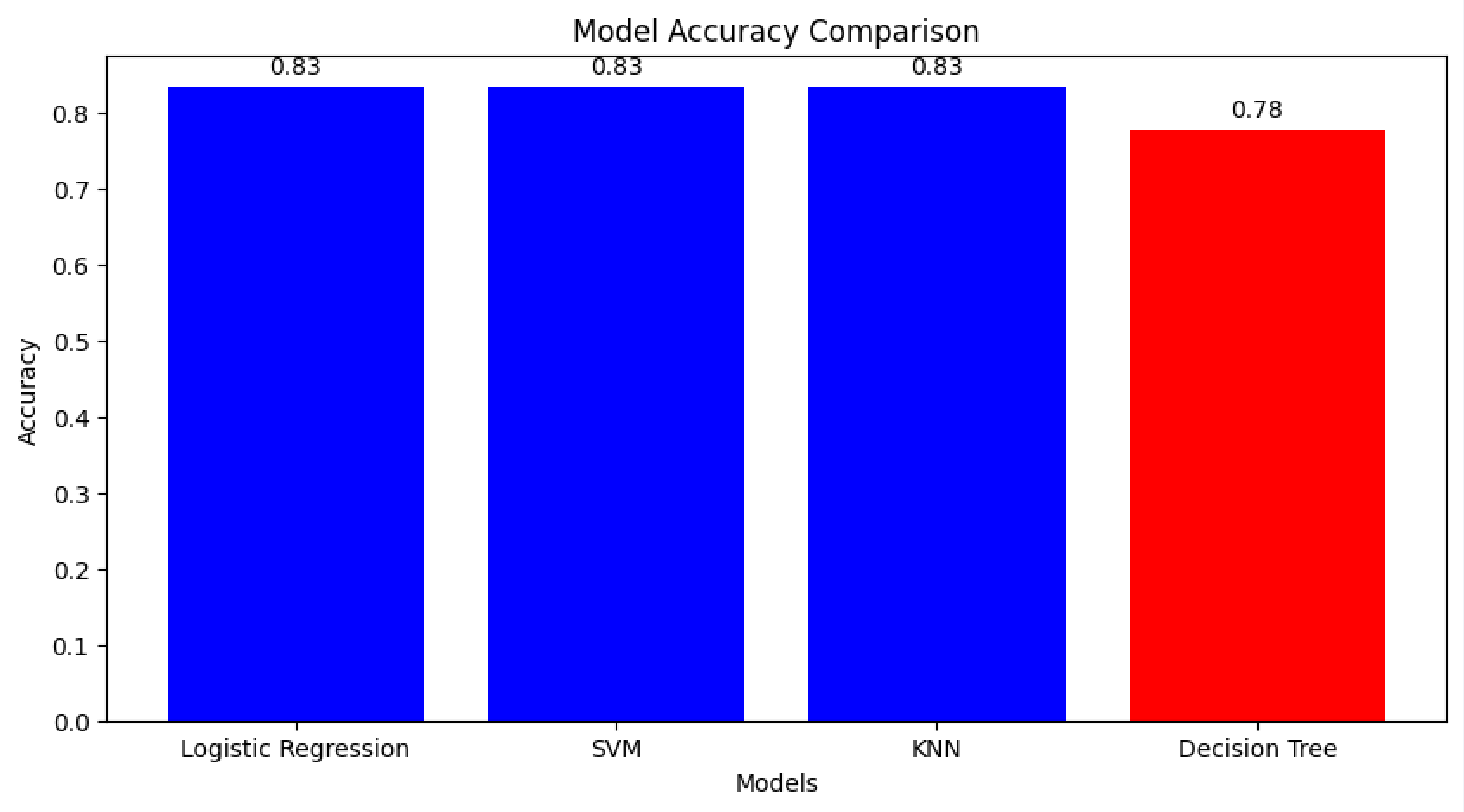


- ✓ Successful landings (class=1) occur across a wide range of payloads, with higher success concentration **between 2000 kg and 6000 kg.**
- ✓ Failure points (class=0) are scattered but tend to appear more frequently at lower payloads or extreme ranges.
- ✓ Booster versions **FT, B4**, and **B5** dominate successful missions, indicating improved reliability in newer boosters.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

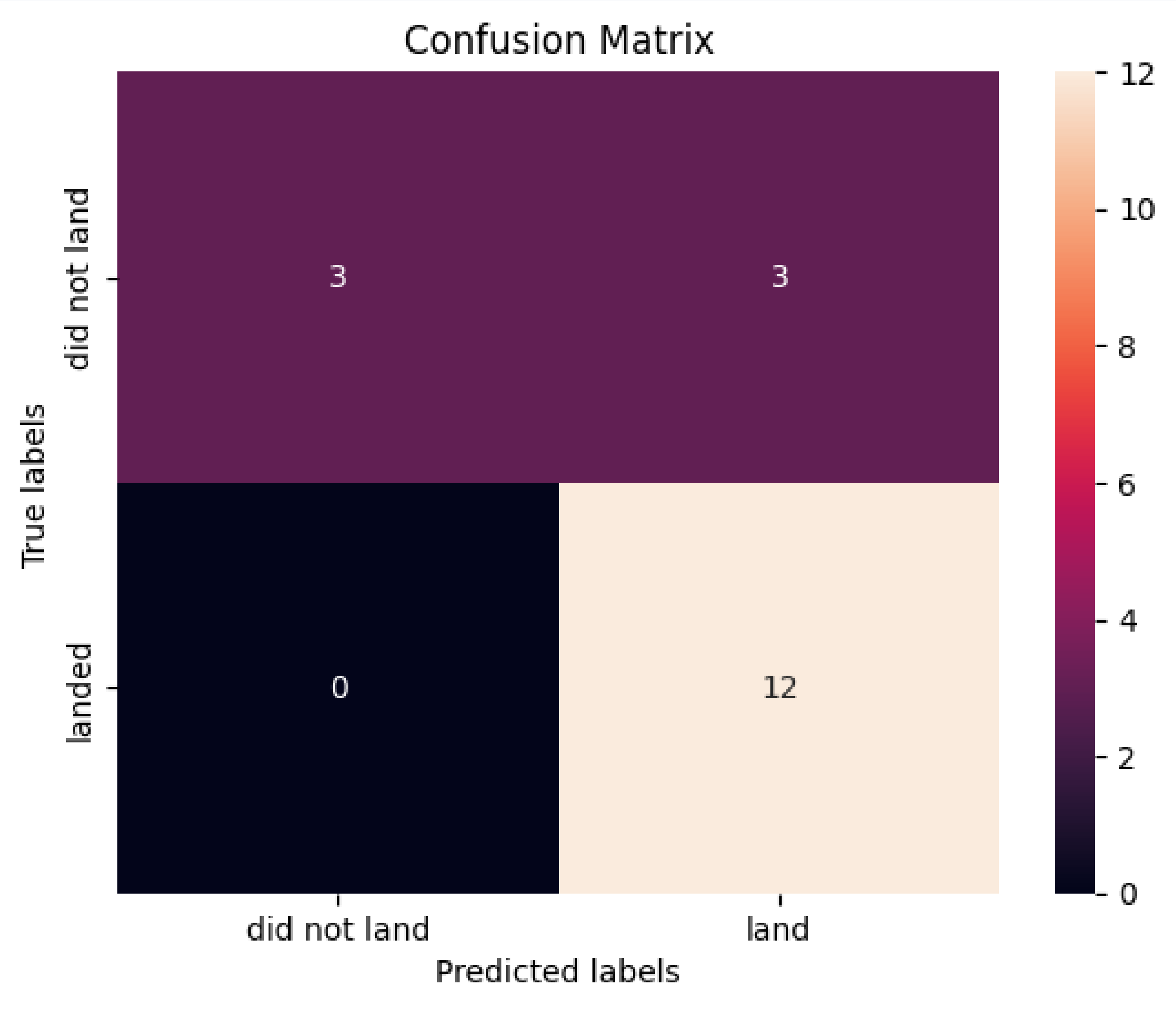


- ✓ The Logistic Regression, SVM, and KNN models continue to show the highest classification accuracy at 83%.
- ✓ The **Decision Tree accuracy dropped to 78%**, confirming that it no longer has the best overall accuracy;

Confusion Matrix

The best model choose was Logistic Regression, because it never misses a successful landing and balances safety with reasonable prediction performance.

- ✓ It achieves perfect sensitivity for detecting successful landings (FN = 0), which is critical in the rocket landing domain;
- ✓ Although false positives exist, from an operational risk perspective, missing a successful landing (false negative) is much more costly.



Conclusions

- ✓ Successfully collected, cleaned, and integrated SpaceX Falcon 9 launch data from multiple sources, including API, web scraping, and SQL databases.
- ✓ Performed comprehensive exploratory data analysis (EDA) revealing key factors affecting landing success: flight number, payload mass, launch site, and orbit type.
- ✓ Visualized data using interactive dashboards and folium maps, identifying that most launches occur near coastlines and in favorable geographic locations for orbital insertion.
- ✓ Identified KSC LC-39A as the launch site with the highest success ratio (76.9%).
- ✓ Conducted predictive modeling using multiple machine learning algorithms: Logistic Regression, SVM, KNN, and Decision Tree.
- ✓ Logistic Regression, SVM, and KNN models achieved highest classification accuracy (83%) with zero false negatives — correctly predicting all successful landings.
- ✓ Confusion matrix analysis indicated that avoiding false negatives is critical in rocket landings, where failing to predict a successful landing may result in lost reusability opportunities.

Thank you!