IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Purushottam Phuyal
Feb 21,2025

Outline

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

Executive Summary

- **Summary of methodologies**
- Collected and processed SpaceX launch data via API and web scraping.
- Conducted Exploratory Data Analysis (EDA) to uncover key trends.
- Applied machine learning models to predict Falcon 9 first-stage landings.
- Achieved a model accuracy of 83% in predicting landing outcomes.
- Developed a machine learning model with an 83% accuracy in predicting Falcon 9 landing success

Introduction

- Project background and context
- Problems you want to find answers

- SpaceX reuses Falcon 9 first-stage boosters to cut launch costs.
- Not all landings are successful—understanding key factors is crucial.

- Key research questions:
- What influences landing success?
- Can we predict successful landings using machine learning?
- How do payload, launch site, and booster type affect outcomes?

Section 1

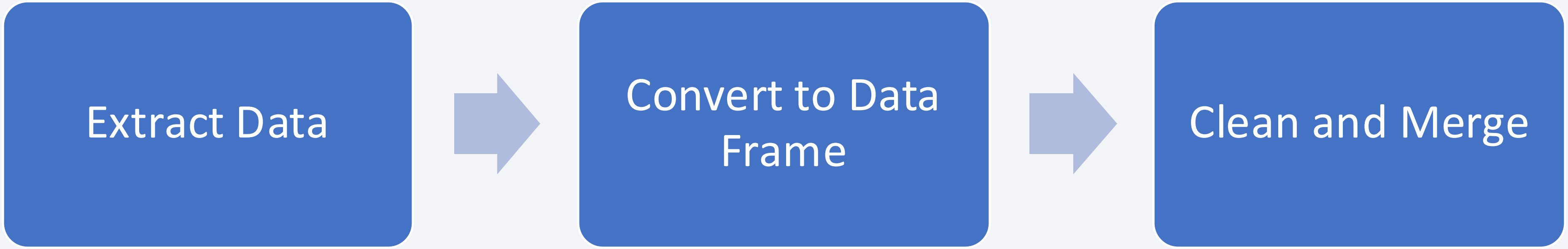
Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API



- [Click here to view the Juypternotebook]

Data Collection - Scraping

Extract Data → Convert HTML Table to Data Frame → Clean and Merge

- [Click here to view the Juypter Notebook]

Data Wrangling

Cleaning and Processing Data

Feature Engineering

Merging and Structuring Data

Tools Used:

Python Pandas NumPy Juypternotebook

- [Click here to view the Jupyter Notebook]

EDA with Data Visualization

Exploratory Data Analysis (EDA)

Identify trends, Patterns and outliers in SpaceX Launch Data

Feature Exploration & Distribution

Bar Charts-Success vs Failure Rates& Launch Sites

Pie Charts-Proportion of Successful Landing Sides

Correlation Analysis

Scatter Plots-Payload Mass vs Landing Success

Box Plots-Payload Distribution across Success/Failure

Advances Insights

Heatmaps-Correlation between launch Parameters

Interactive Visualization for better understanding

Click here to view the Juypter NoteBook

EDA with SQL

Data Retrieval with SQL Queries

Analyzing Launch Site performance

Payload Mass and Landing Success

SQL –Based Data Aggregation and Visualization

[Click here to View the Juypter Notebook]

Build an Interactive Map with Folium

Map objects and Purpose

Markers: Plotted SpaceX Launch Sites to identify key location

Circles:Higlighted Surrounding regions for better Visualization

Lines: Drew Connection Between Launch Sites and Landing Zones

Tools Used:Python,Flium,Pandas,Jupyter Notebook

Click here to view the Jupyter Notebook

Build a Dashboard with Plotly Dash

Bar Charts-Helps identify the best –performance launch sites

Pie Chart-Provides an overview of landing success distribution

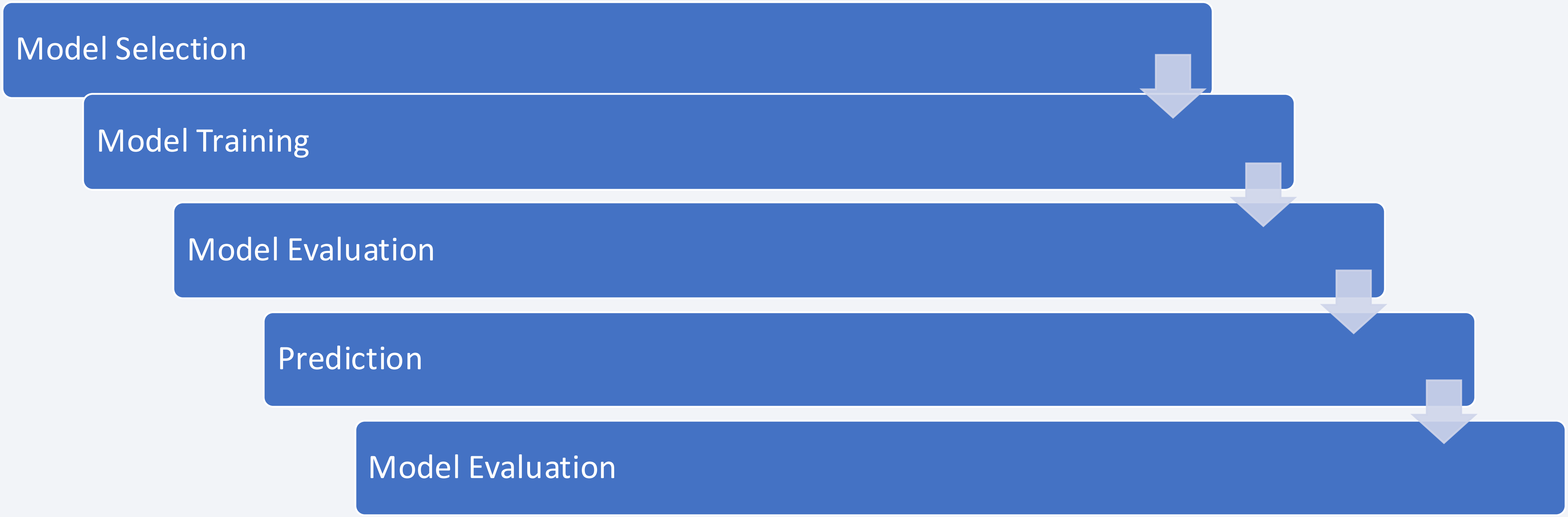
Scatter Plot-Explores how payload mass influences landing outcomes

Dropdown filters-Enhances users interactivity by allowing site/year selection

Sliders-Enables users to analyze payload effect dynamically

Click here to view the Juypter Notebook

Predictive Analysis (Classification)



Click here for Juypter Notebook

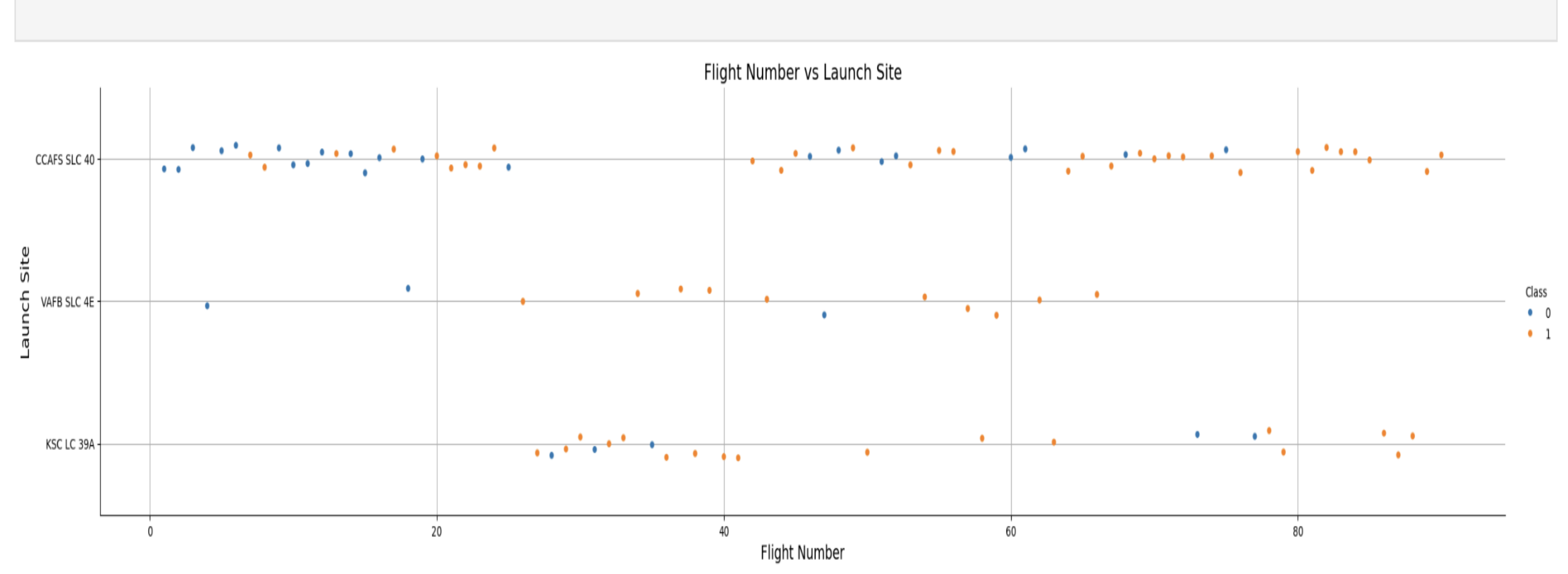
Results

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

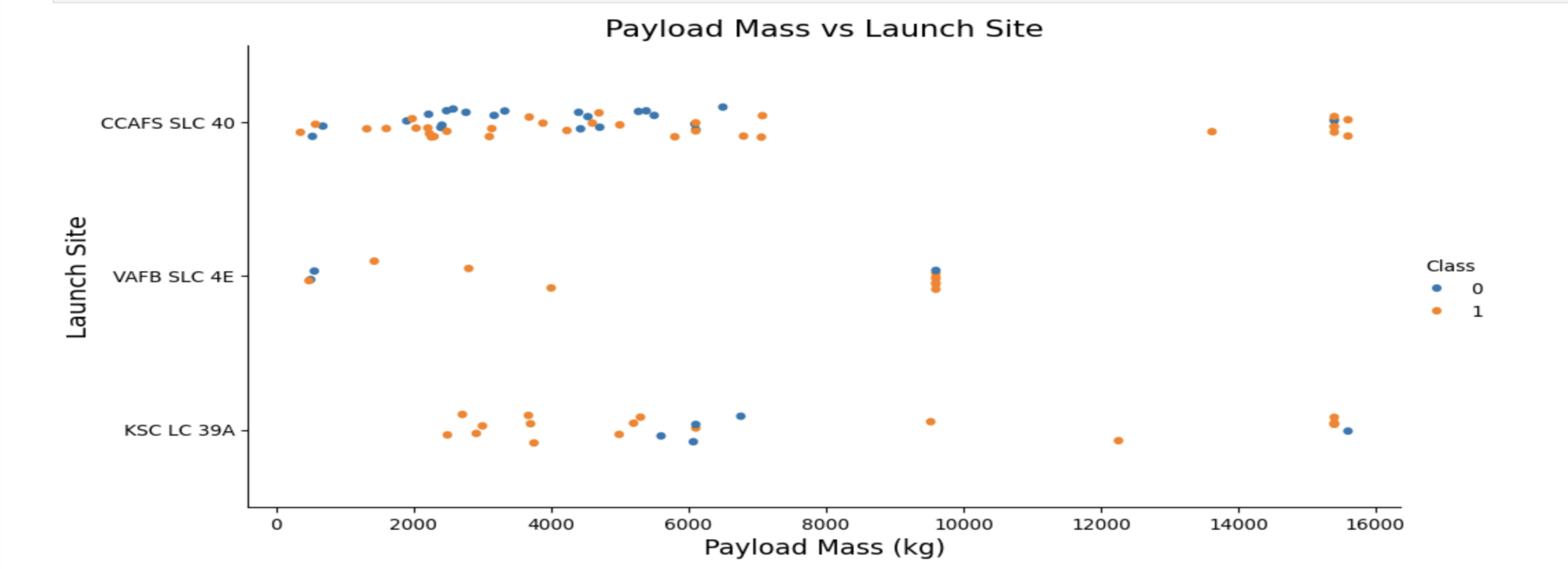
Section 2

Insights drawn from EDA

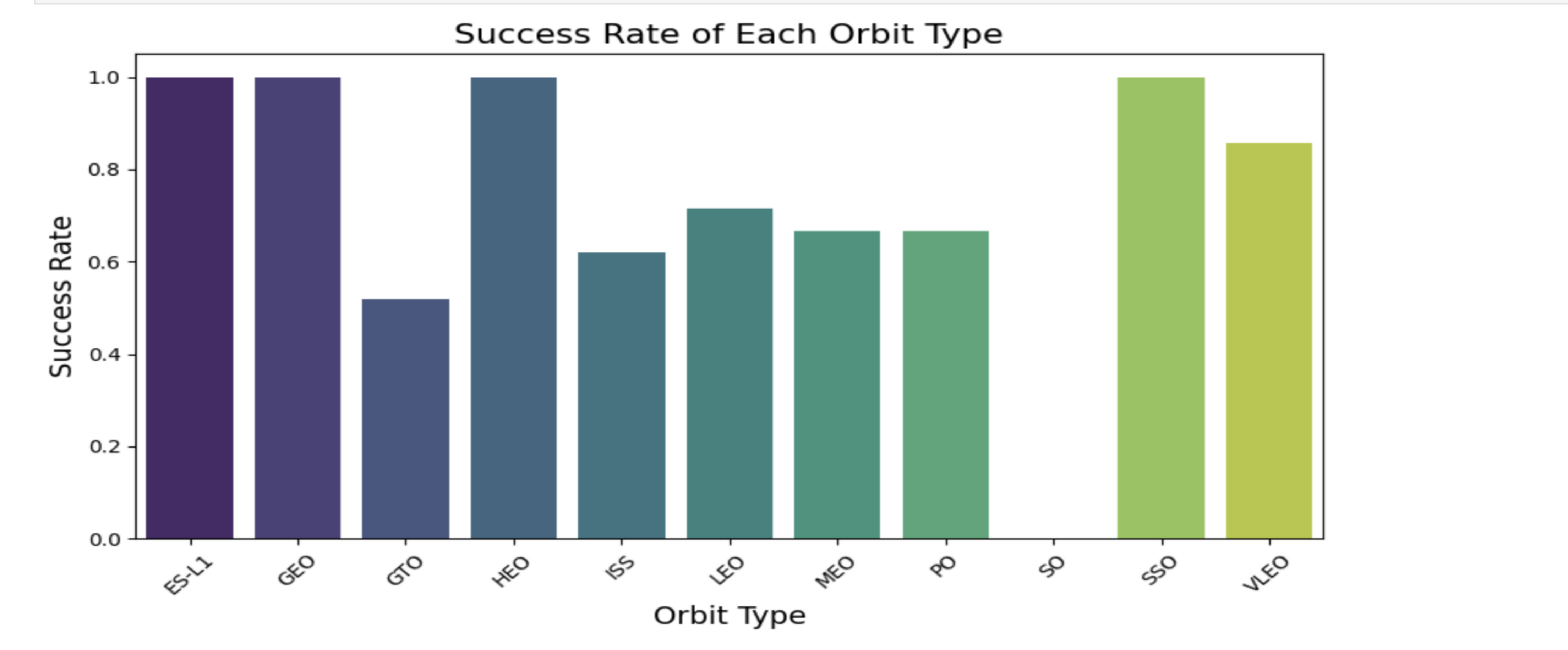
Flight Number vs. Launch Site



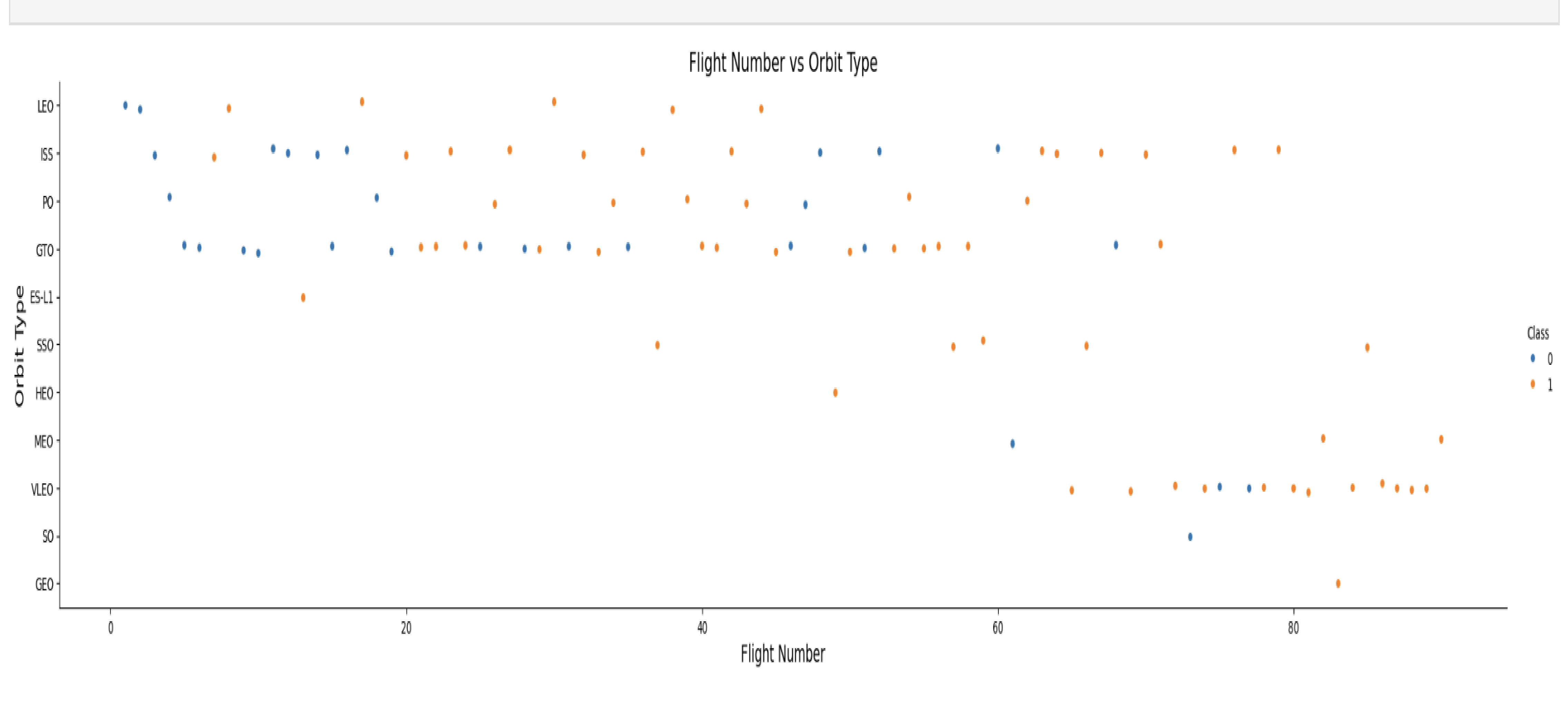
Payload vs. Launch Site



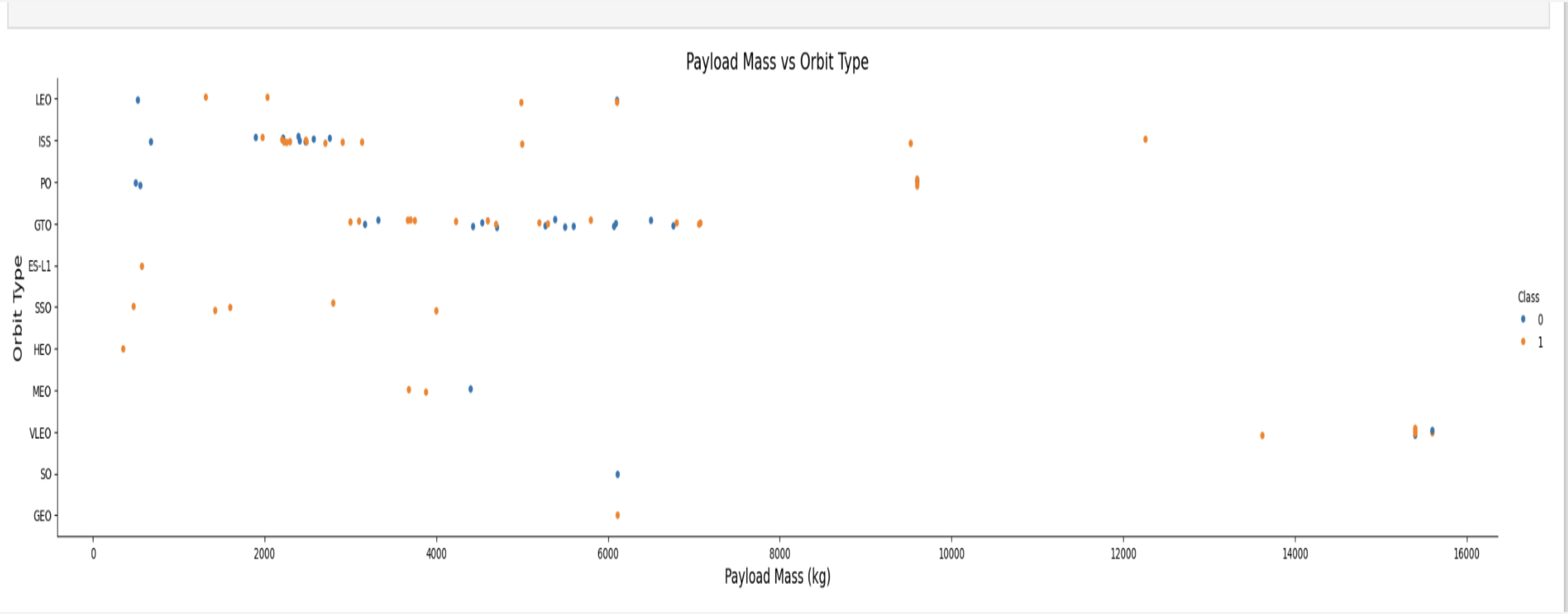
Success Rate vs. Orbit Type



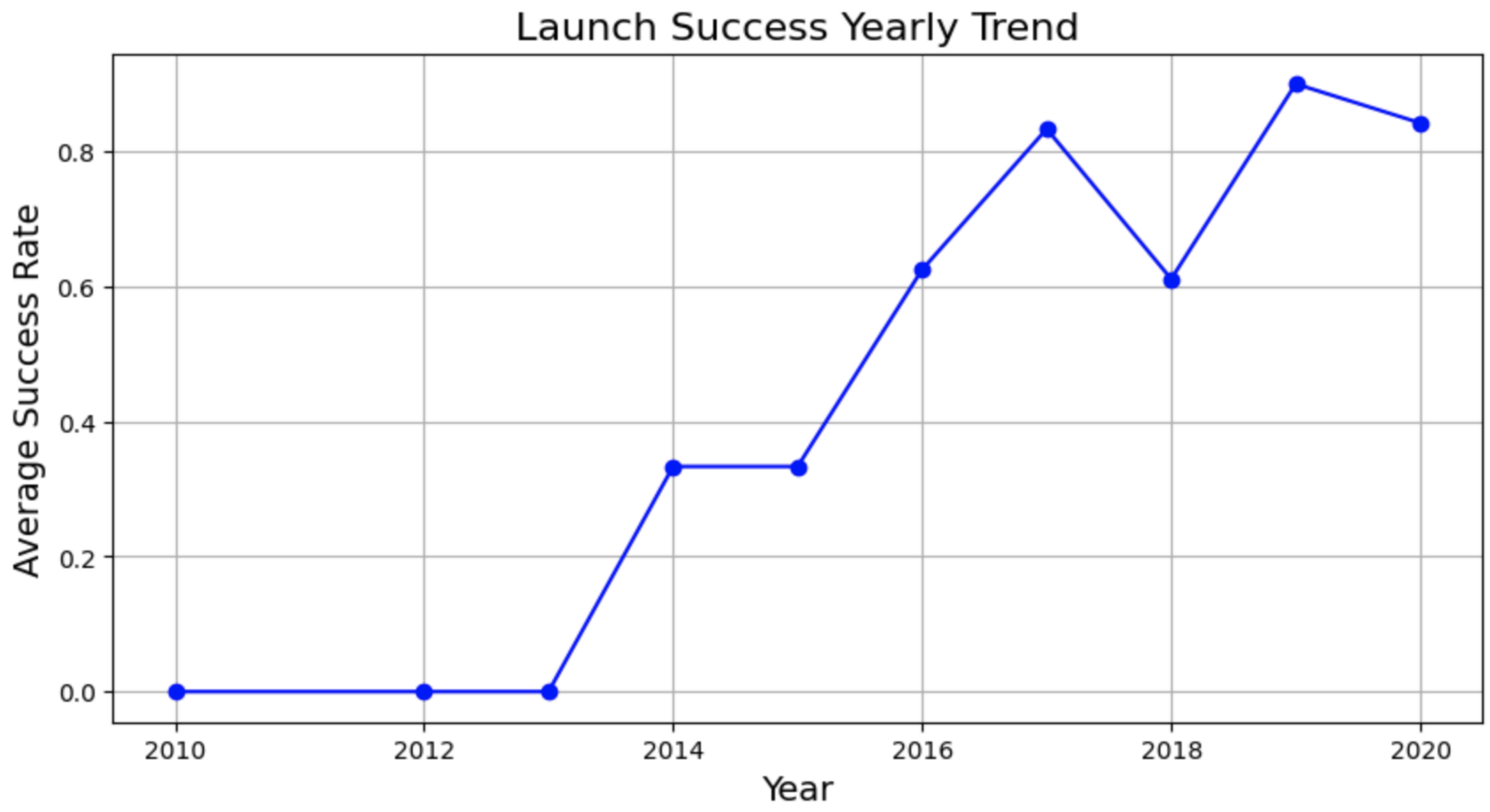
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Launch Sites:

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

The query identified four unique launch sites from the dataset. These sites are key locations for SpaceX launches. The Cape Canaveral Air Force Station (CCAFS) and Kennedy Space Center (KSC) are among the most frequently used launch sites. The Vandenberg Air Force Base (VAFB) supports polar orbit launches, which require different launch trajectories.

Launch Site Names Begin with 'CCA'

- Five records where launch sites begin with `CCA`
- Date & Time
- .Booster Version:
- Payload & Mass: .
- Orbit: .
- Customer:. Mission Outcome & Landing Outcome:
- The query output shows the first five records where the Launch Site starts with "CCA".

Total Payload Mass

The query **sums the "Payload Mass"** column for records where the **Customer** column contains 'NASA (CRS)'.However, the result returned **0.0**, indicating: Either NASA (CRS) launches in the dataset had **zero payload mass**, or

There are **no matching records** under "Customer" containing 'NASA (CRS)'.

Average Payload Mass by F9 v1.1

It uses the **AVG("Payload Mass")** function to compute the mean payload mass from the **SPACEXTABLE**.

The **WHERE clause** filters the records to include only those where **Booster Version = 'F9 v1.1'**.

The result is **0.0**, indicating:

- There might be no records in the dataset for **F9 v1.1**.
- The payload mass values for these records could be **missing or zero**.

First Successful Ground Landing Date

The query retrieves the **earliest date** when a successful landing on a ground pad was achieved.

The **MIN("Date")** function is used to find the earliest landing date.

The **WHERE clause** filters records to only include landings with the outcome **'Success (ground pad)'**

The result, **2015-12-22**, indicates the first recorded **successful ground pad landing** in the dataset.

Successful Drone Ship Landing with Payload between 4000 and 6000

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

This analysis helps identify **successful reusable boosters** that handled moderate payload masses while landing on drone ships, supporting cost-efficient space missions.

Total Number of Successful and Failure Mission Outcomes

Success: **38 missions**

Failure: **3 missions**

This analysis provides a **comprehensive view** of mission outcomes, useful for evaluating **reusability strategies** and improving landing techniques.

Boosters Carried Maximum Payload

- The **booster versions that carried the highest payloads** include:
  - **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**

  Falcon 9 Block 5 boosters consistently carried the maximum payload mass, highlighting their reliability for heavy missions.

2015 Launch Records

Failed Landing Outcomes in Drone Ship (2015)

Landing Outcome: Failure (drone ship)

Booster Versions: F9 v1.1 B1012, F9 v1.1 B1015

Launch Site: CCAFS LC-40

Insight: In 2015, two booster versions (B1012 & B1015) launched from CCAFS LC-40 failed to land on a drone ship, indicating early challenges in achieving successful drone ship landings.

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

Ranking: 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

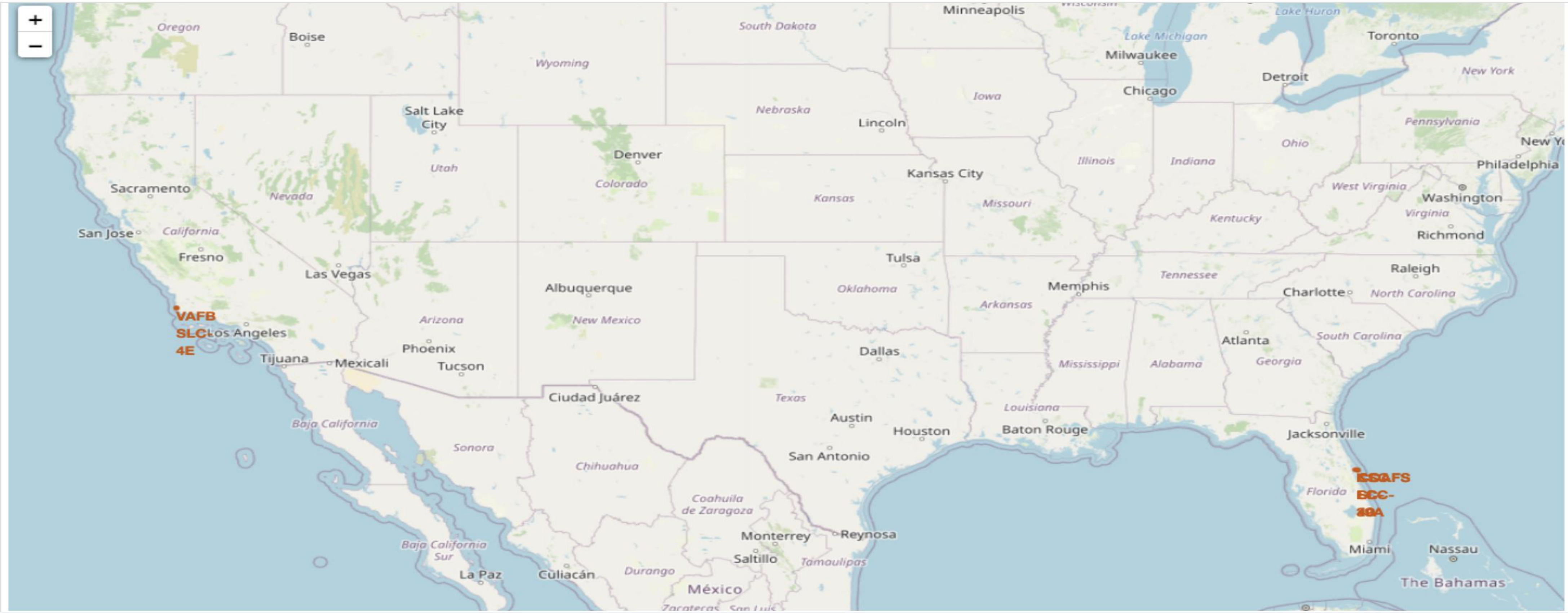
Insight: The highest count of landing outcomes was "No Attempt," showing a significant number of missions without a landing attempt, followed by an equal number of successes and failures on drone ships.

Section 3

Launch Sites Proximities Analysis

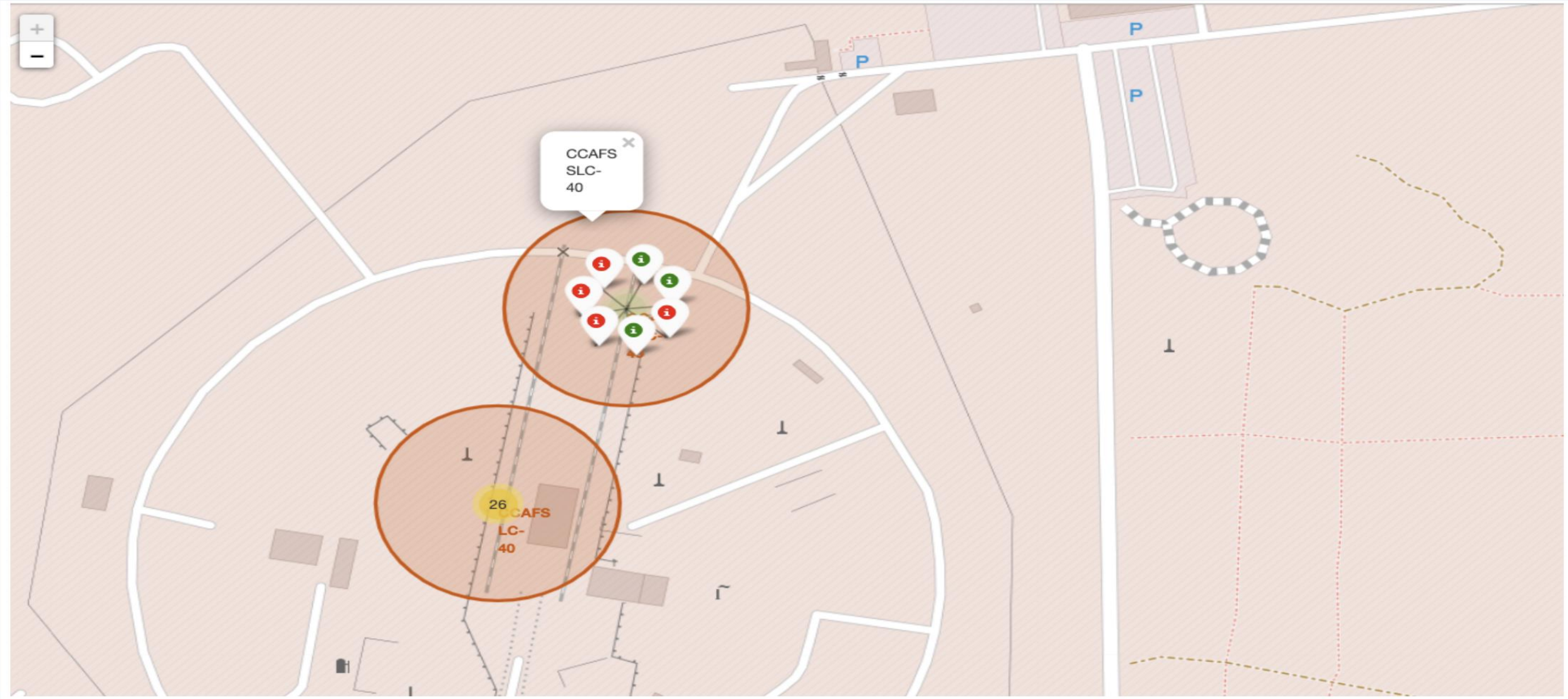
<Folium Map Screenshot 1>

SpaceX Launch Sites in the United States

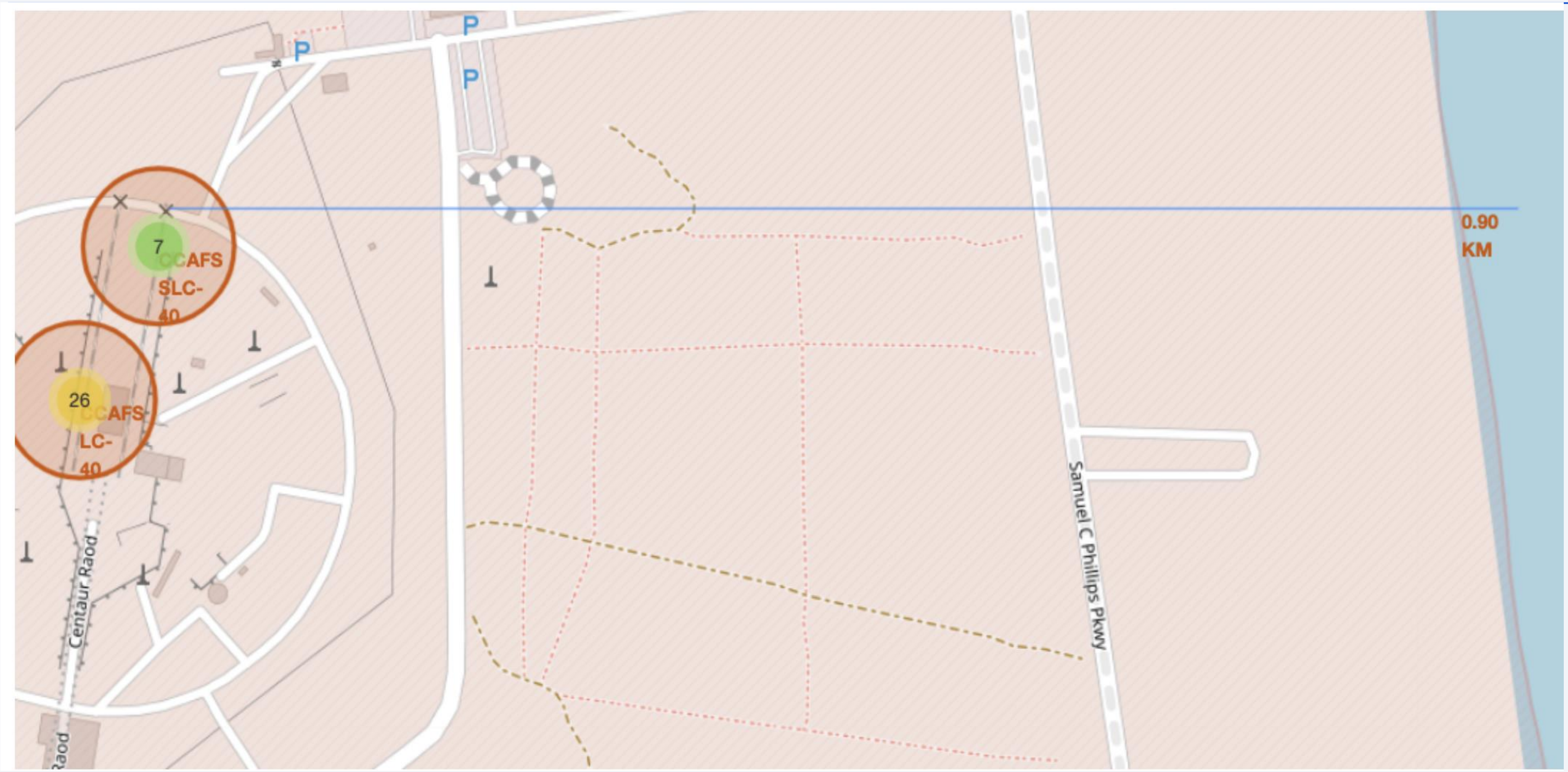


Most SpaceX launch sites are located near coastal regions in **Florida and California**, ensuring safety and efficiency for rocket launches.

<Folium Map Screenshot 2>



<Folium Map Screenshot 3>

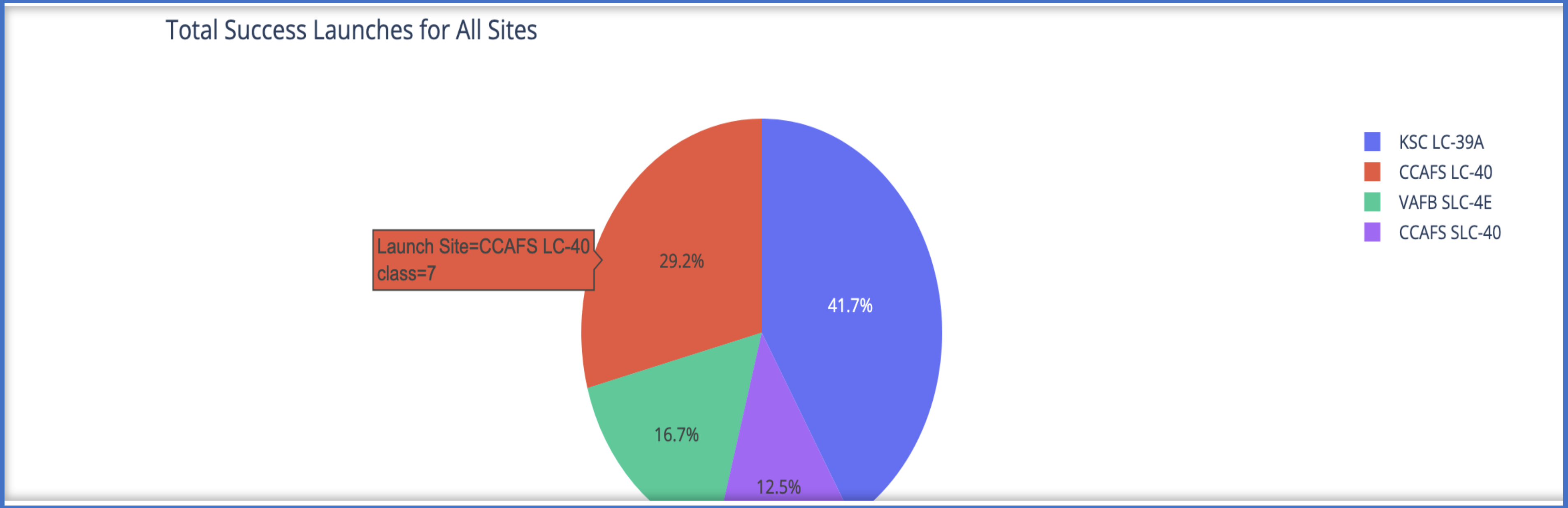


Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

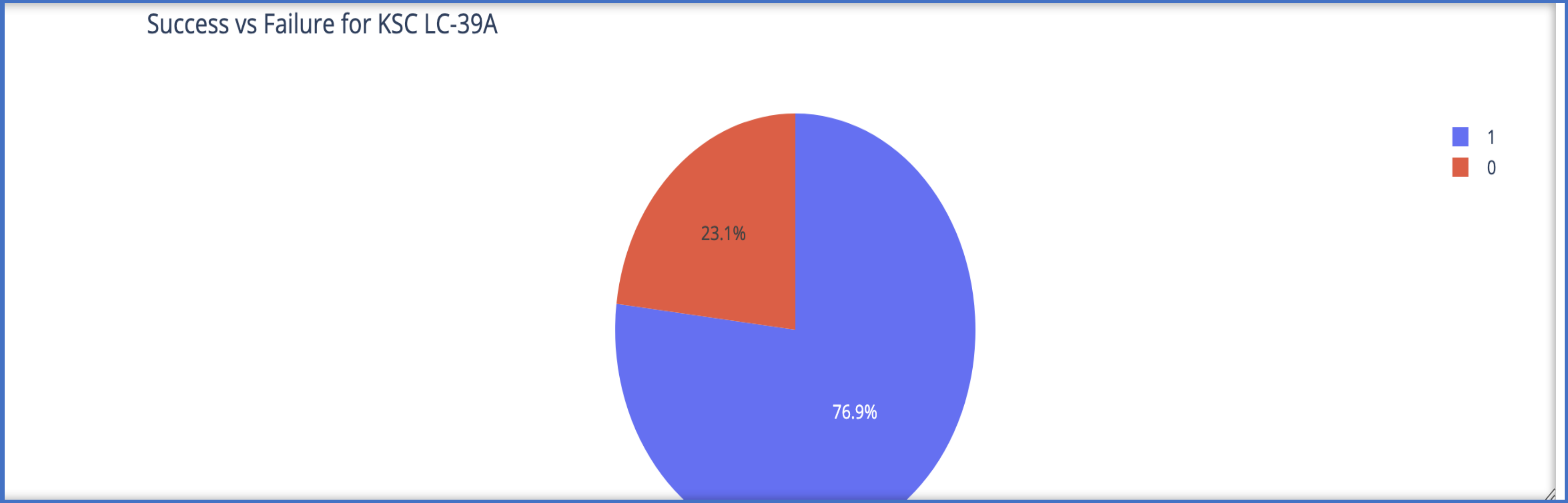
- Success Rate of SpaceX Launches by Site



The pie chart illustrates the distribution of successful SpaceX launches across different sites, with KSC LC-39A having the highest success rate (41.7%), followed by CCAFS LC-40 (29.2%), VAFB SLC-4E (16.7%), and CCAFS SLC-40 (12.5%), indicating varying launch performance across locations.

<Dashboard Screenshot 2>

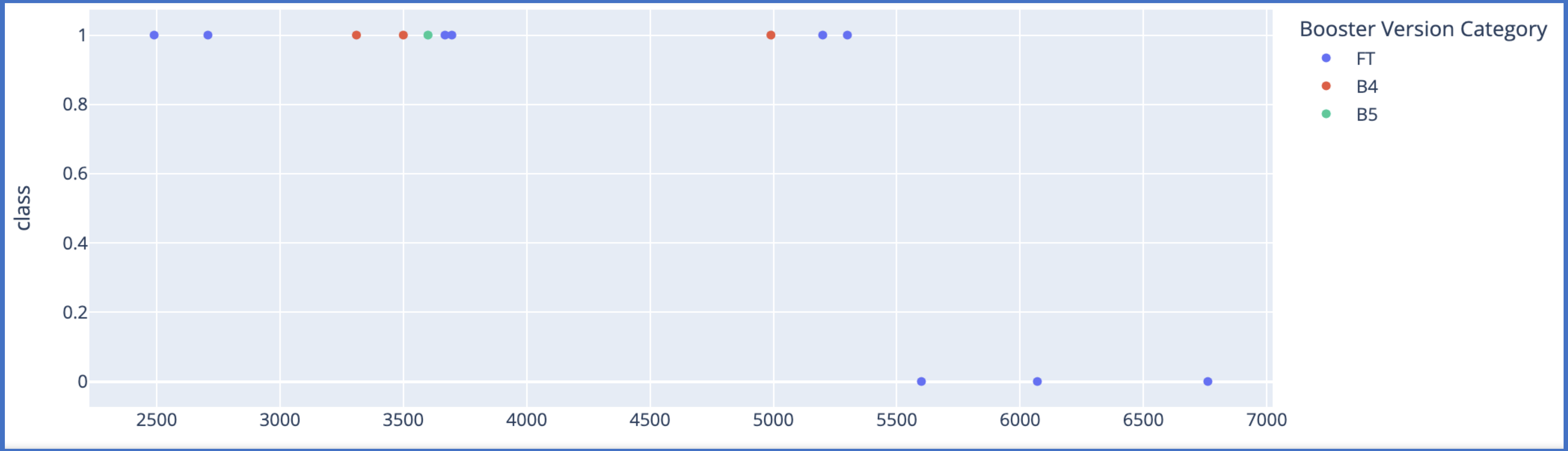
- Highest Lauch Success rate launch site



- The pie chart displays the distribution of successful launches across all sites, with **KSC LC-39A having the highest success rate (41.7%)**, followed by **CCAFS LC-40 (29.2%)**, **VAFB SLC-4E (16.7%)**, and **CCAFS SLC-40 (12.5%)**.

<Dashboard Screenshot 3>

Payload vs. Launch Outcome scatter plot Show

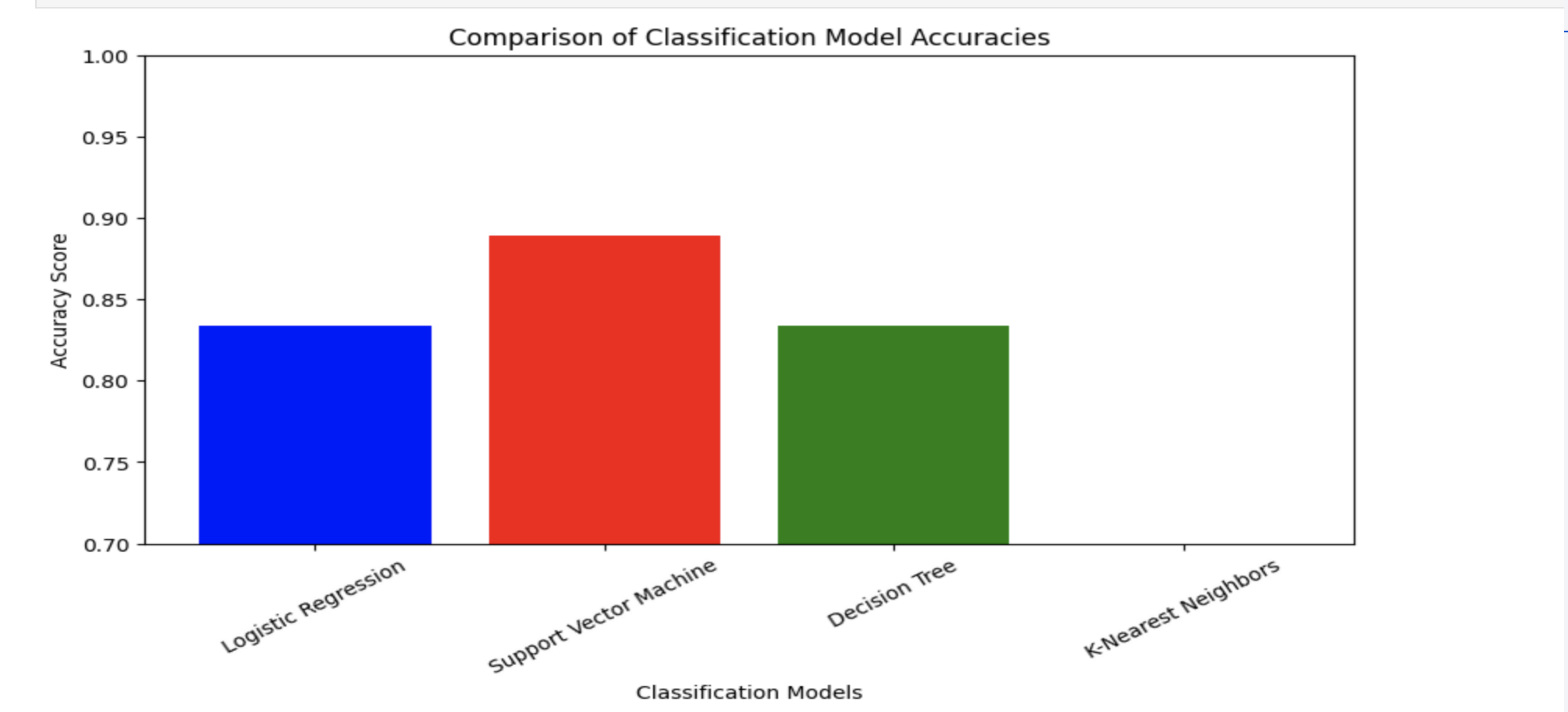


The scatter plot shows that **all payload launches at KSC LC-39A were successful (class=1), with FT booster version having the highest frequency of successes, followed by B4 and B5.**

Section 5

Predictive Analysis (Classification)

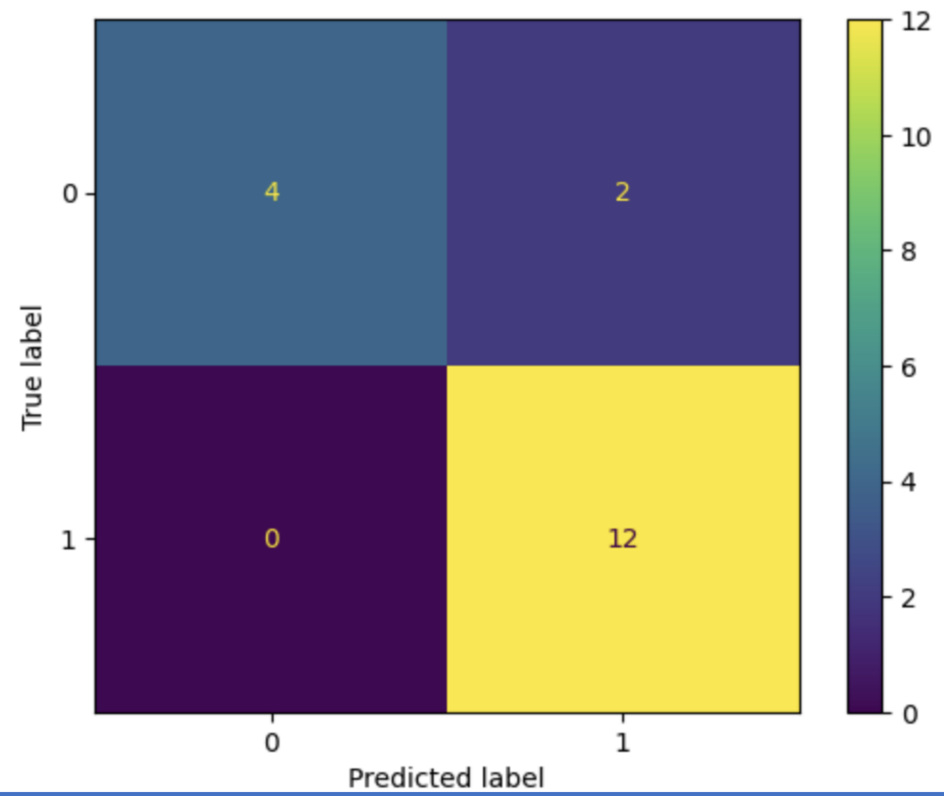
Classification Accuracy



The **Support Vector Machine (SVM)** model has the **highest classification accuracy**, as shown in the bar chart.

Confusion Matrix

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1311c8290>
```



The **Support Vector Machine (SVM)** model has a **high classification accuracy** with **no false negatives (FN = 0),** meaning it correctly predicted all positive cases, but it misclassified **2 negative cases as positive (FP = 2).**

Conclusions

1. **Machine Learning Model Performance:** The Support Vector Machine (SVM) model achieved the highest classification accuracy (88.89%) in predicting Falcon 9 landing outcomes.
2. **Launch Site Insights:** KSC LC-39A had the highest launch success rate (41.7%), making it the most reliable site.
3. **Payload vs. Success Rate:** All payload launches at KSC LC-39A were successful, with the FT booster version contributing the most to successful missions.
4. **Landing Outcome Trends:** The highest count of landing outcomes was "No Attempt," followed by equal numbers of successes and failures on drone ships.
5. **Reusable Boosters:** Falcon 9 Block 5 boosters consistently carried the highest payloads, reinforcing their reliability in heavy missions.
6. **Failed Drone Ship Landings in 2015:** Two booster versions (B1012 & B1015) from CCAFS LC-40 failed to land on a drone ship, highlighting early challenges in landing technology.
7. **Data Collection and Analysis:** API and web scraping techniques were used to collect SpaceX launch data, which was analyzed using SQL, visualizations, and predictive models.

Appendix

Python scripts for data processing, exploratory data analysis (EDA), and machine learning model training. SQL queries used for database analysis. Charts and visualizations, including bar charts, pie charts, and scatter plots, showcasing key insights.Jupyter Notebook outputs from data wrangling, predictive analysis, and classification model evaluation.Confusion matrix and performance metrics for the best classification model (SVM).Interactive maps (Folium) and dashboards (Plotly Dash) displaying SpaceX launch site analytics.

Thank you!