IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

By
ADIJAT TITILAYO ATANDA
October 2nd,2024

Outline

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

Executive Summary

Objective:
Analyze SpaceX launch data to assess success and identify critical performance factors.

Methodologies:
•Data collection, cleaning, and processing
•Exploratory Data Analysis (EDA) with visualizations and SQL
•Development of interactive dashboards using Folium and Plotly Dash
•Predictive analysis utilizing logistic regression, SVM, decision trees, and KNN

Results:
•Identified key predictors of launch success
•Achieved over 90% accuracy with the top classification model
•Provided insights to improve future mission planning and safety

Introduction

Overview of SpaceX:
Founded by Elon Musk in 2002, SpaceX seeks to transform space travel through the use of reusable rockets, drastically reducing transportation costs and paving the way for the future colonization of Mars.

Data Focus:
This project analyzes historical launch data to uncover trends and factors influencing launch success, with the goal of optimizing future missions and improving resource allocation.

Objective:
To assess SpaceX's launch success through data analysis, offering insights to improve mission planning and safety evaluations.

Section 1

Methodology

Methodology

- Data collection methodology: SpaceX launch data was sourced from an open dataset. Loaded using pandas for analysis
- Perform data wrangling: Cleaned data by handling missing values, filtering columns, and standardizing formats for analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built classification models (Logistic Regression, SVM, Decision Tree, KNN) to predict launch success
  - Applied **GridSearchCV** for hyperparameter tuning. Evaluated models with accuracy scores

Data Collection

- **Describe how data sets were collected.**

Data was collected via API/Wikipedia, it was then transformed and exported, the data was saved to a csv format

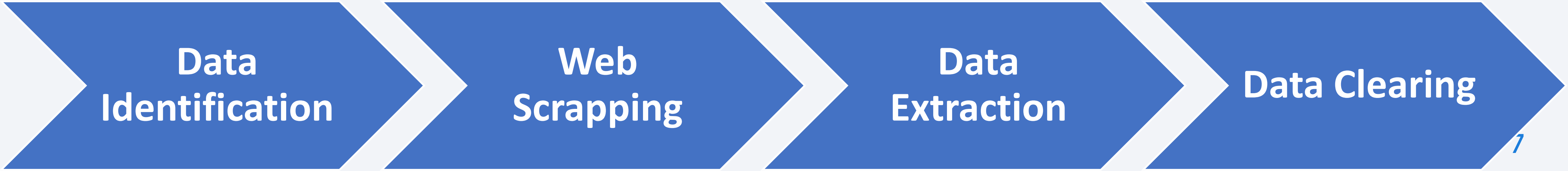
- You need to present your data collection process use key phrases and flowcharts

Step 1: API/Wikipedia data collection

Step 2: Web scrapping to get the data needed.

Step 3: Parsing and extracting dataset using BeautifulSoup

Step 4: Data cleaning and preparation



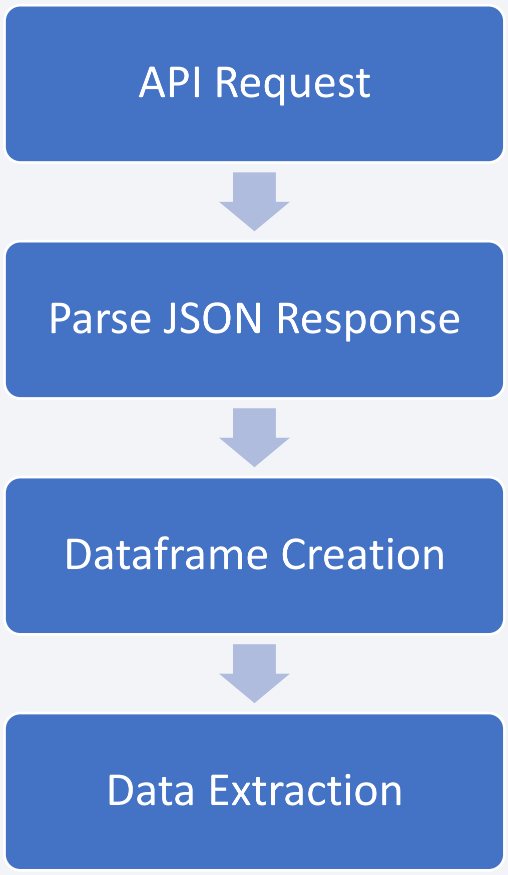
Data Collection – SpaceX API

- **Data Source**: SpaceX Launch Data via **SpaceX REST API**

Collection Process:

- **Step 1:** Made **API calls** to retrieve launch data.
- **Step 2:** Parsed JSON response into **pandas** DataFrame.
- **Step 3:** Extracted key fields such as Rocket, Payloads, Launchpads, and Cores.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/Data_Collection_API.ipynb



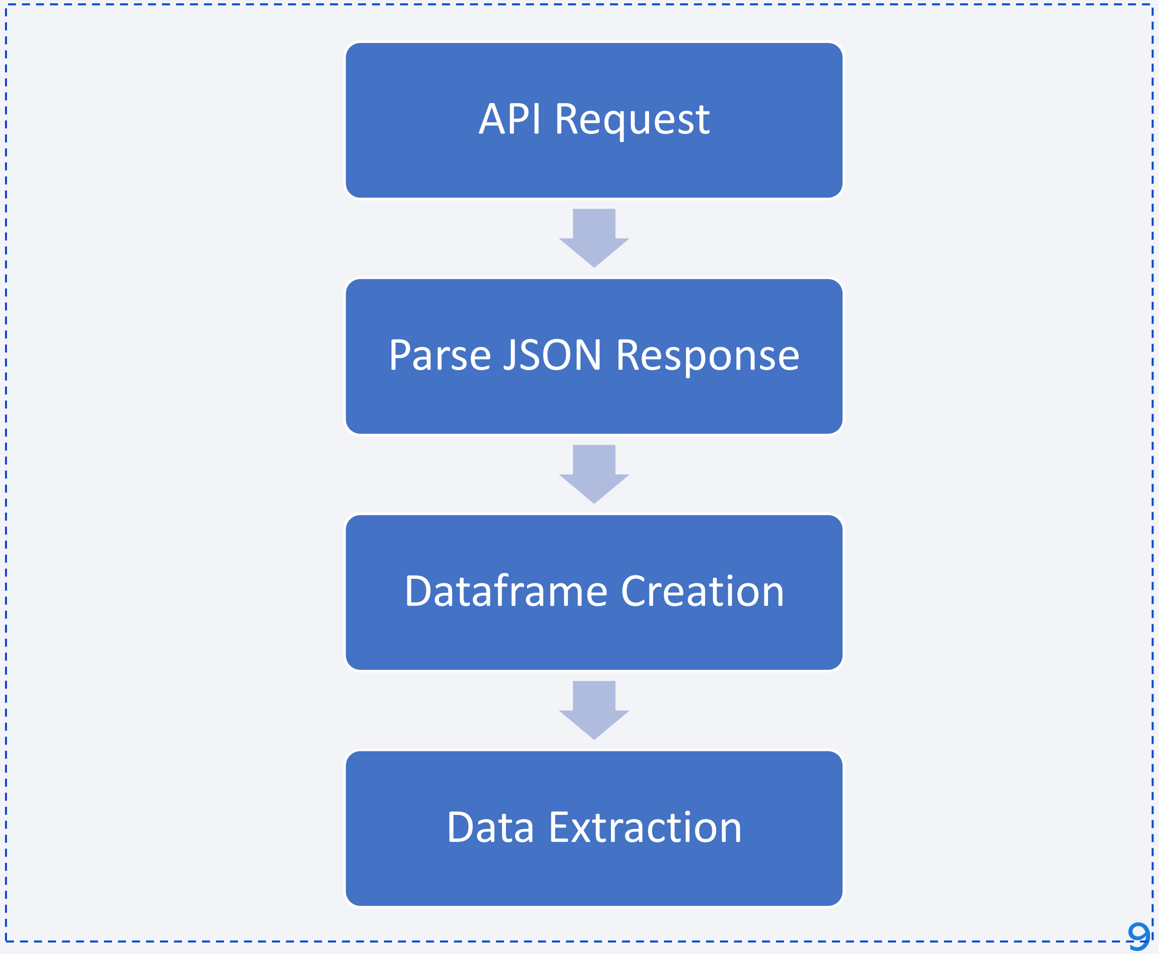
Data Collection - Scraping

Data Source:
Scraped historical SpaceX data from SpaceX website using **BeautifulSoup** and **requests** libraries.

Scraping Process:

- **Step 1:** Sent **HTTP requests** to target web pages.
- **Step 2:** Parsed **HTML content** with BeautifulSoup.
- **Step 3:** Extracted relevant data such as mission names, launch dates, and rocket details into a **pandas DataFrame**.

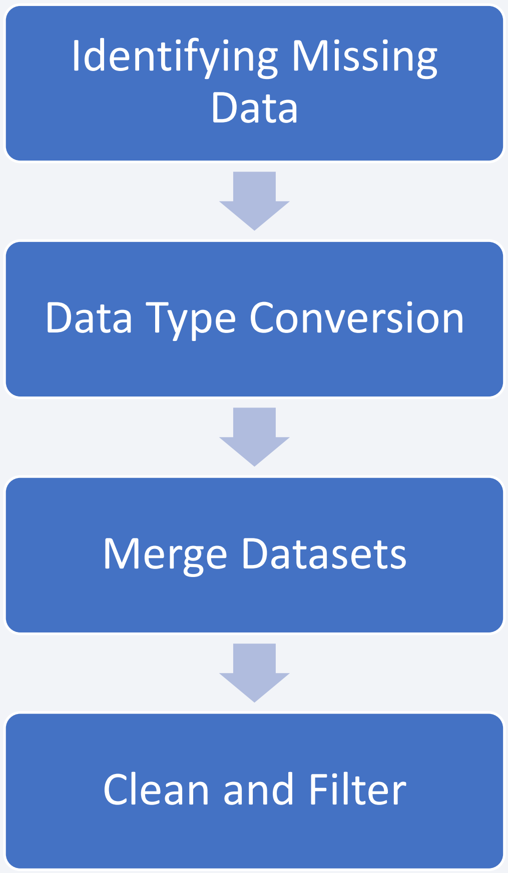
https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/Data_Collection_WebScraping.ipynb



Data Wrangling

- **Data Processing Overview:**
- **Step 1:** Identified and handled **missing values** using pandas.
- **Step 2:** Performed **data type conversions** for numerical and datetime fields.
- **Step 3:** Cleaned, filtered, and **merged datasets** (API & scraped data) to create a unified dataset for analysis.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/Data_Wrangling.ipynb



EDA with Data Visualization

- **Charts Plotted:**
  - **Bar Charts:** To visualize rocket launch success rates across years.
  - **Pie Charts:** To show the distribution of launch outcomes (success vs failure).
  - **Scatter Plots:** To analyze the relationship between payload mass and launch success.
  - **Line Plots:** To track trends in launch frequency over time.

- **Purpose of Charts:**
  - **Bar & Pie Charts:** Highlighted categorical data for success analysis.
  - **Scatter & Line Plots:** Examined trends and relationships over time.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/EDA_Visualizations.ipynb

EDA with SQL

- **SQL Queries Performed:**

  SELECT: Extracted data on successful rocket launches.

  GROUP BY: Analyzed launch outcomes by year and launchpad.

  JOIN: Combined datasets for rocket performance and payload analysis.

  ORDER BY: Ranked launches by success rate and payload mass.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/EDA_SQL.ipynb

Build an Interactive Map with Folium

- **Map Objects Created:**

  Markers: Indicated launch sites for better visibility of locations.

  Circles: Represented launch zones to visualize operational areas.

  Polylines: Illustrated rocket trajectories to demonstrate flight paths.

- **Rationale for Objects:**

  Markers provide quick identification of launch sites.

  Circles help visualize operational impact zones.

  Polylines enhance understanding of launch trajectories.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/Folium_Visualization.ipynb

Build a Dashboard with Plotly Dash

- **Dashboard Components:**

  Pie Chart: Displays the percentage of launches per site.

  Scatter Plot: Visualizes the relationship between payload and launch outcome.

- **Rationale for Components:**

  Pie Chart provides quick insights into launch site distributions.

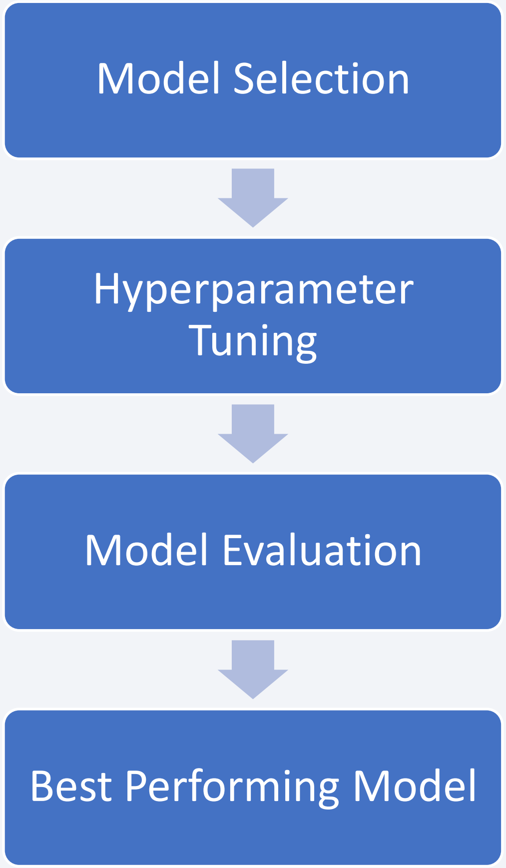
  Scatter Plot helps analyze the correlation between payload sizes and launch success rate

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/dashboard.py

Predictive Analysis (Classification)

- **Model Development Process:**
  - **Model Selection**: Chose classifiers (Logistic Regression, SVM, Decision Tree, KNN).
  - **Hyperparameter Tuning**: Used GridSearchCV for optimal parameters.
  - **Model Evaluation**: Employed accuracy, precision, recall, and F1-score metrics.
- **Best Performing Model**: Identified Decision Tree with the highest accuracy after tuning.

https://github.com/Myalayo/Myalayo-Cousera-Course-Webfala/blob/main/Machine_Learning_Prediction_Analysis.ipynb



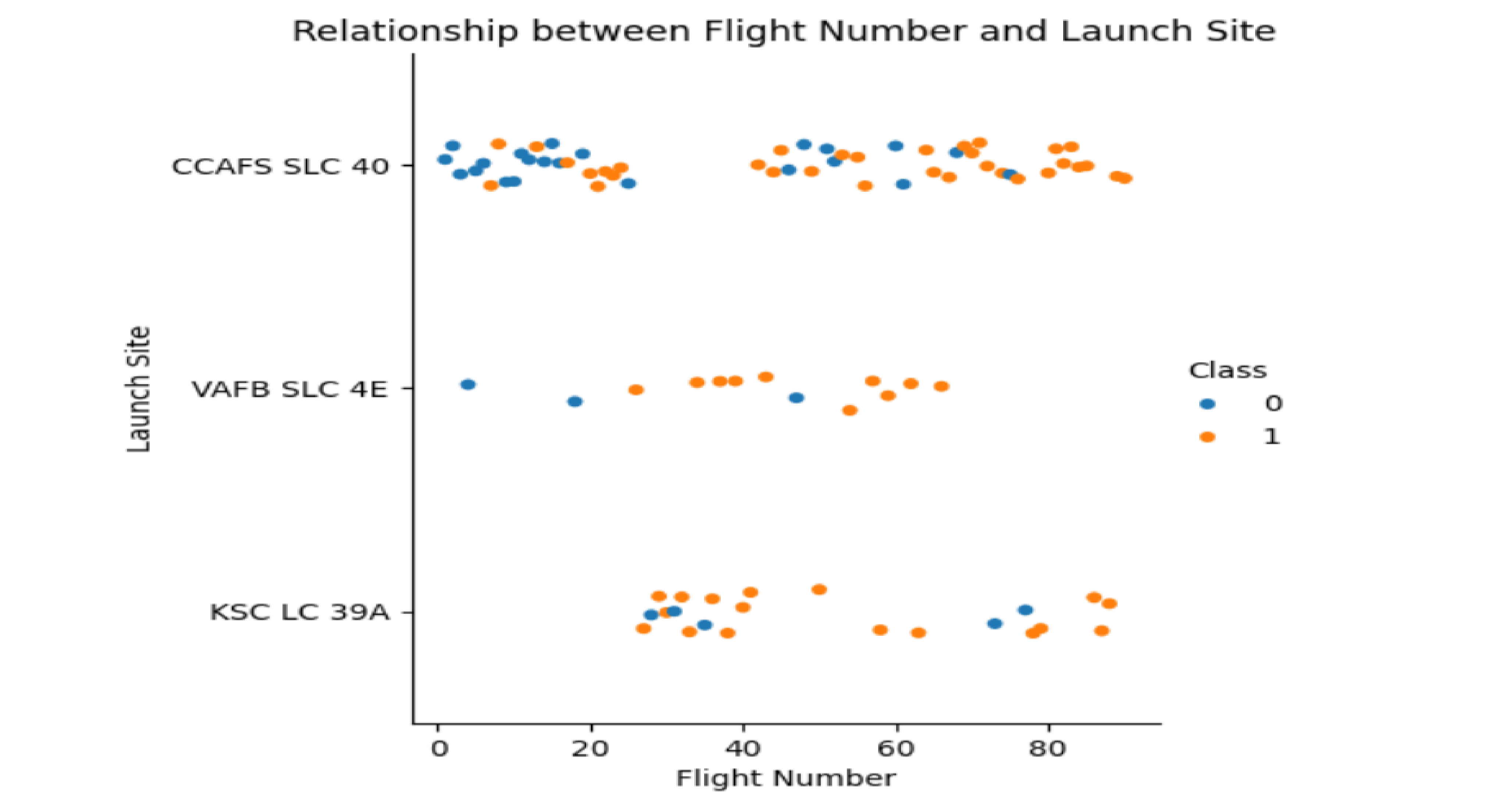
Results

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

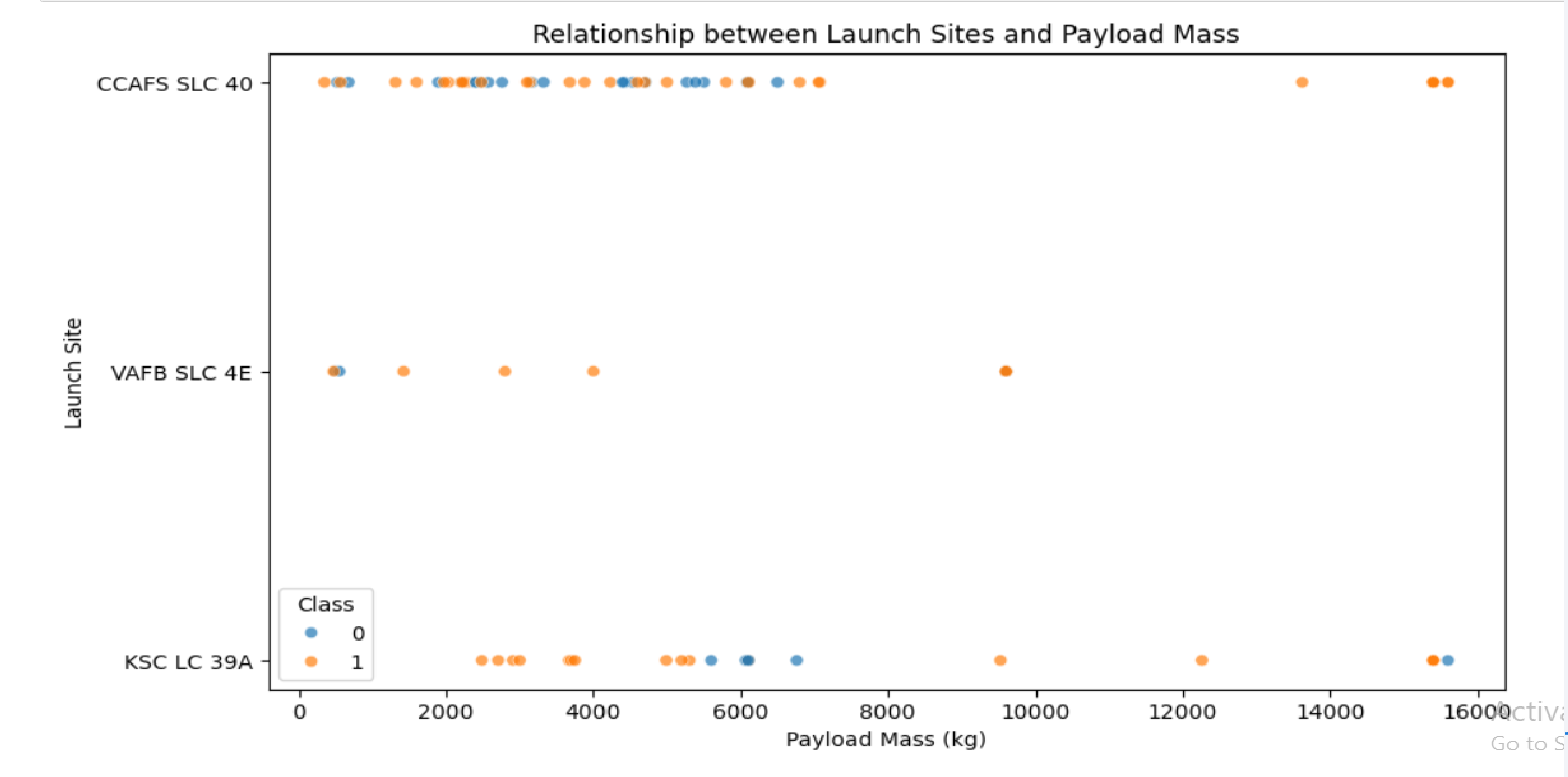
Section 2

Insights drawn from EDA

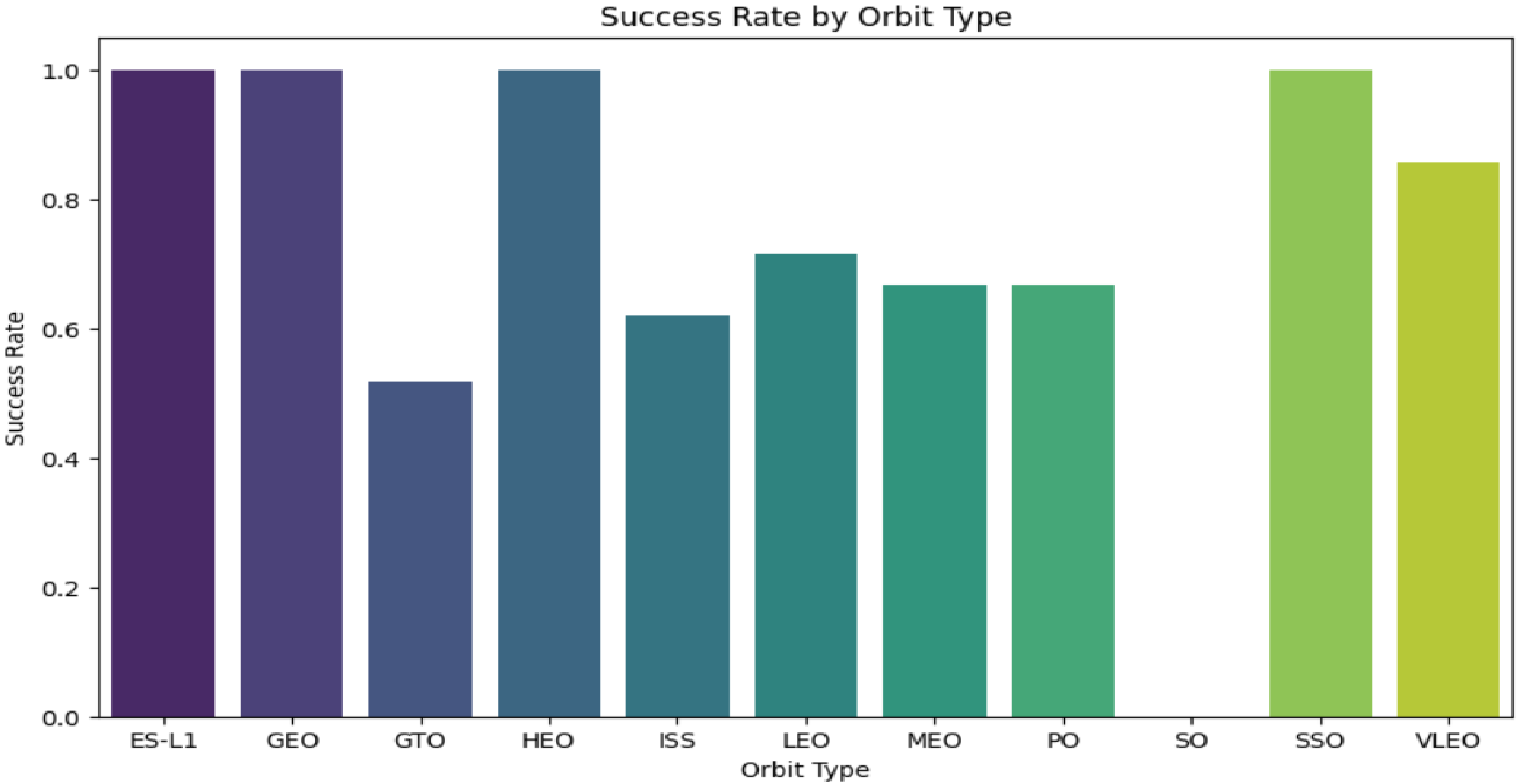
Flight Number vs. Launch Site



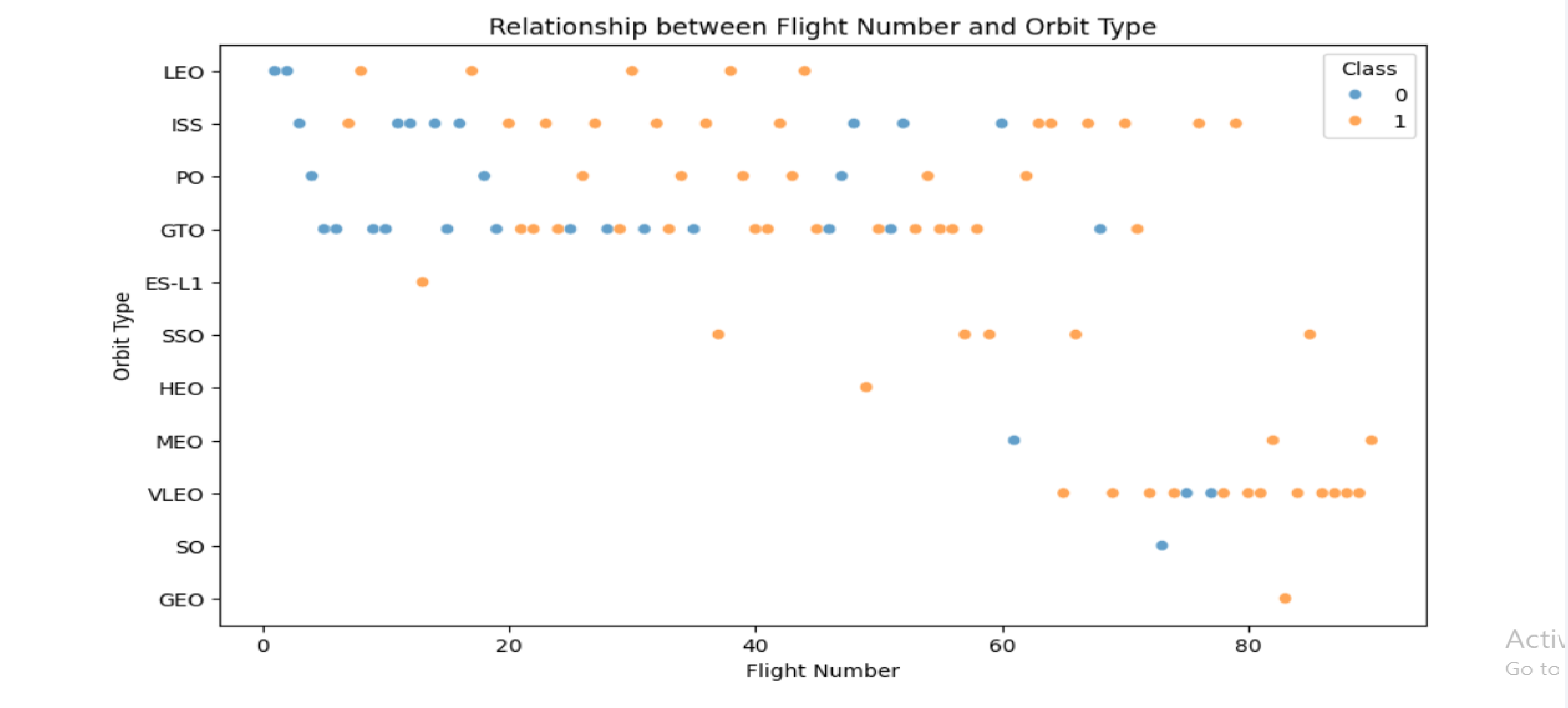
Payload vs. Launch Site



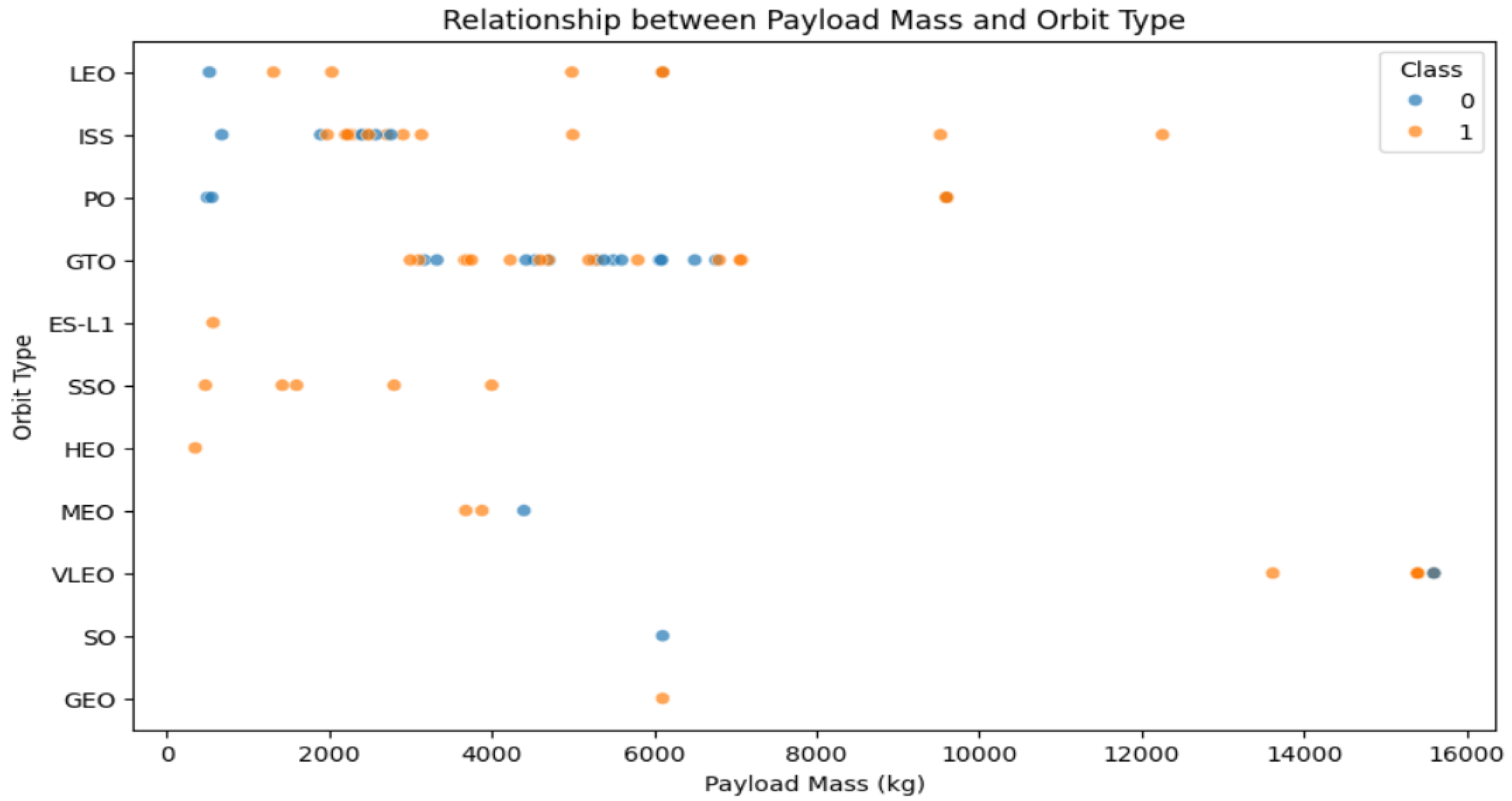
Success Rate vs. Orbit Type



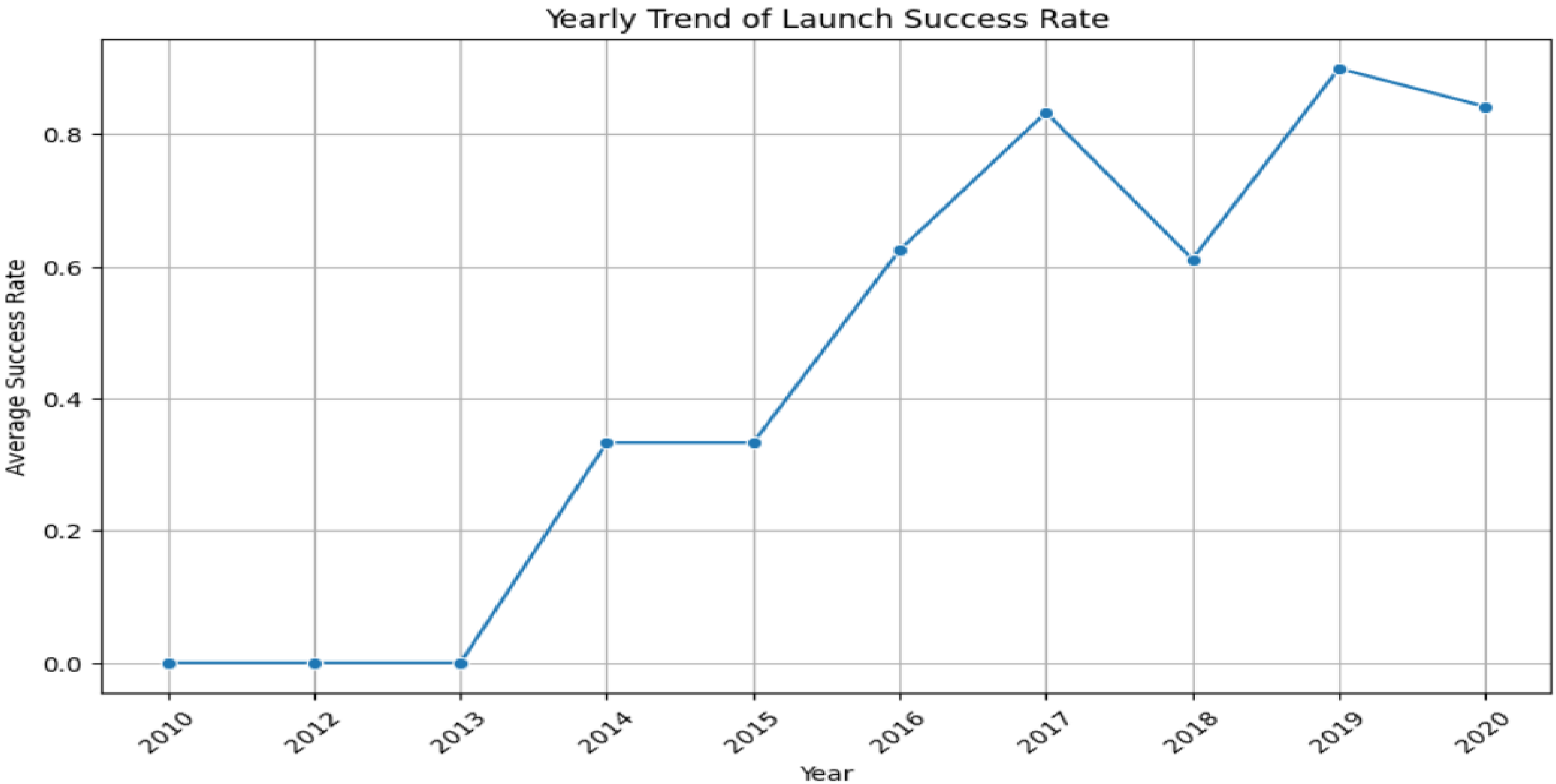
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

[12]:

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcom
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 attemp
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

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

2015 Launch Records

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

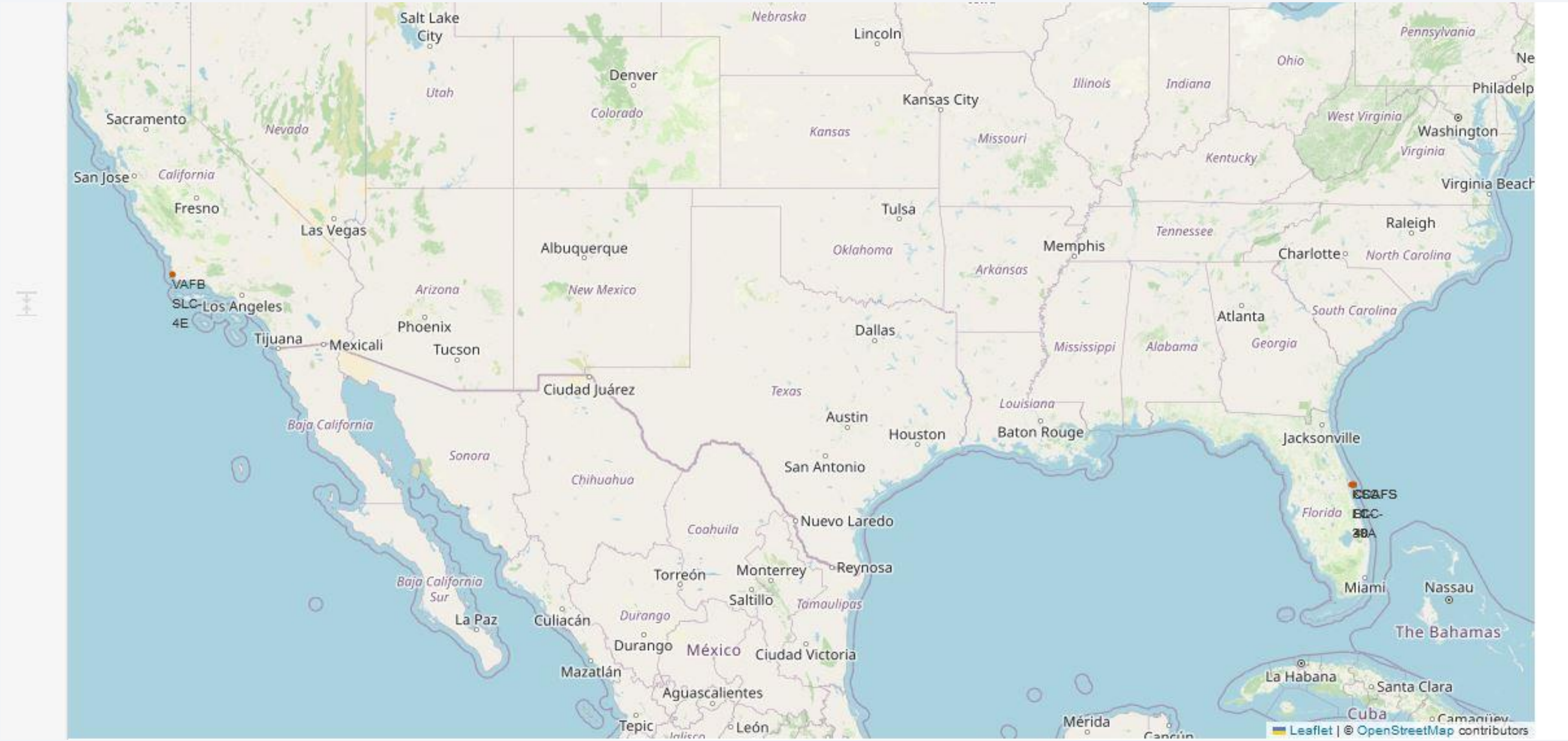
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Date	count(Landing_Outcome)
None	0

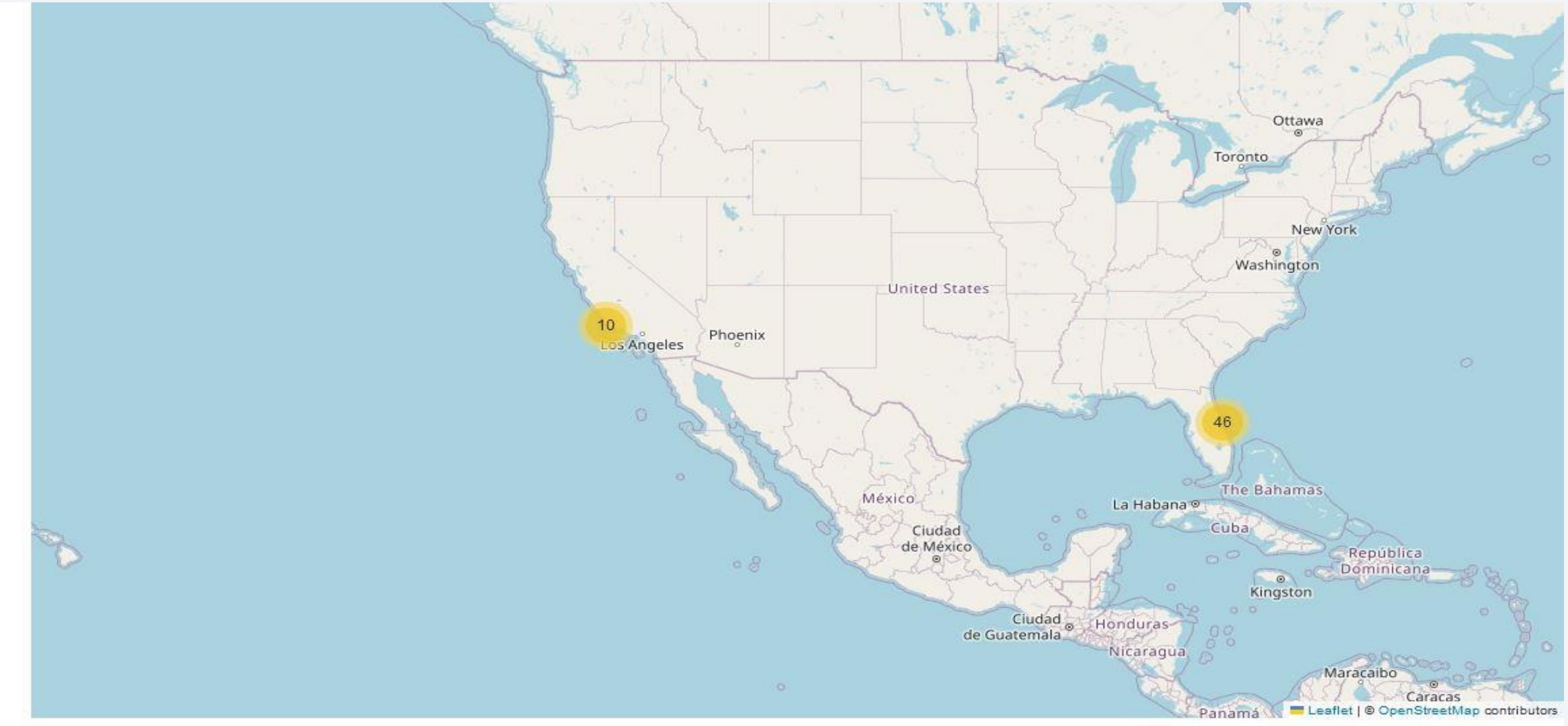
Section 3

Launch Sites Proximities Analysis

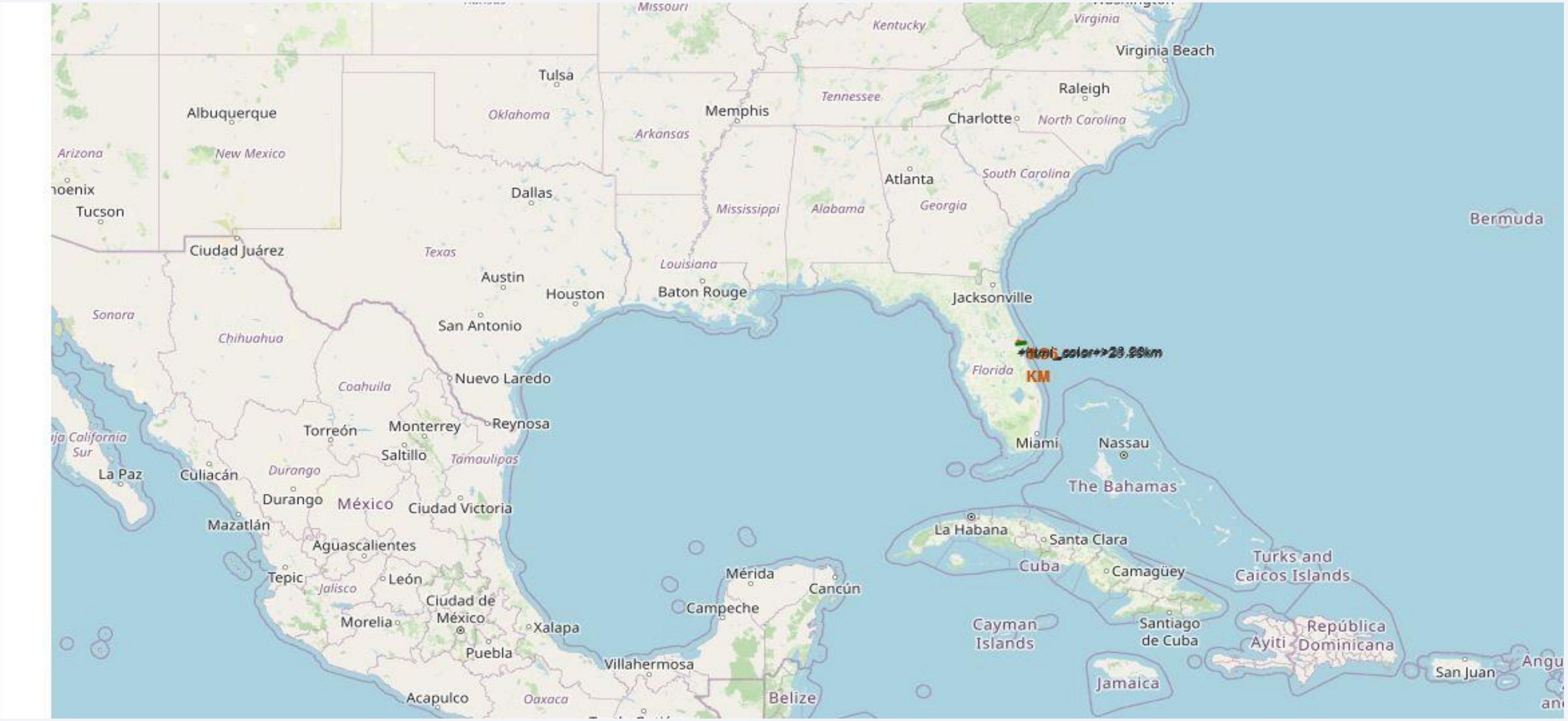
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



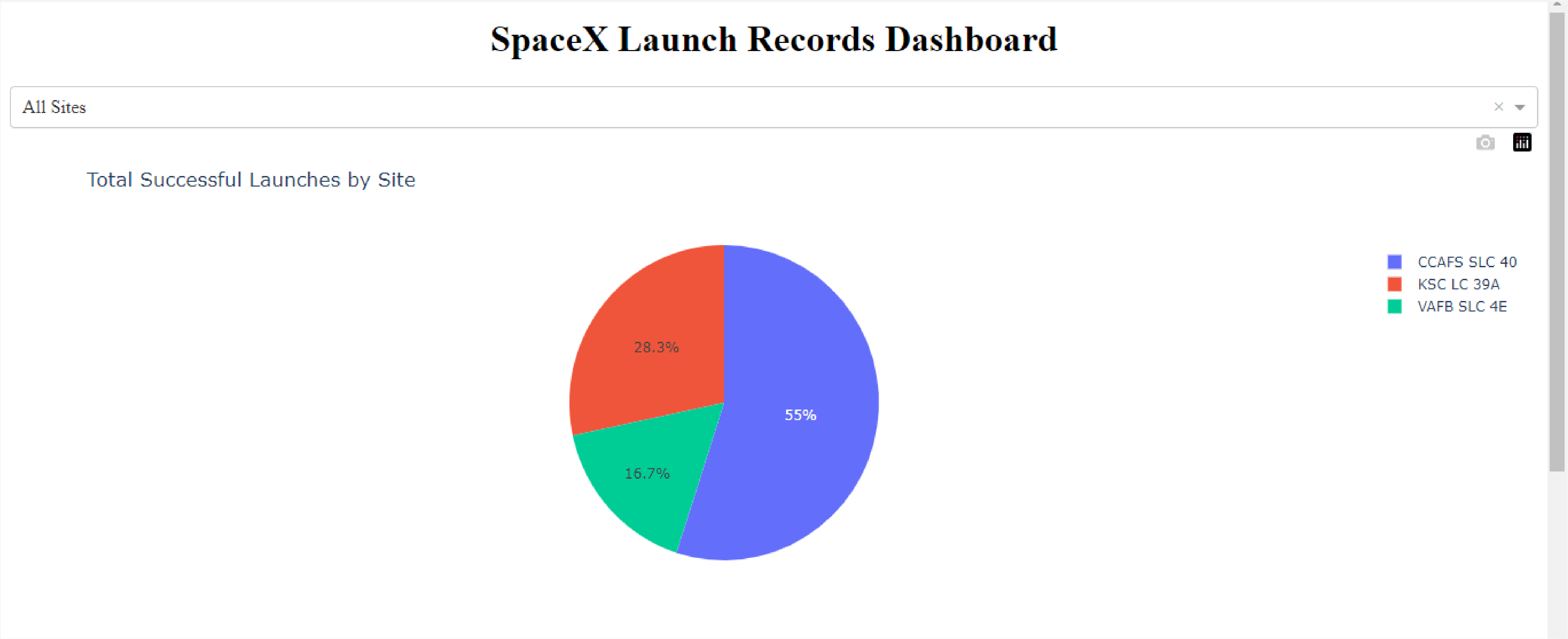
<Folium Map Screenshot 3>



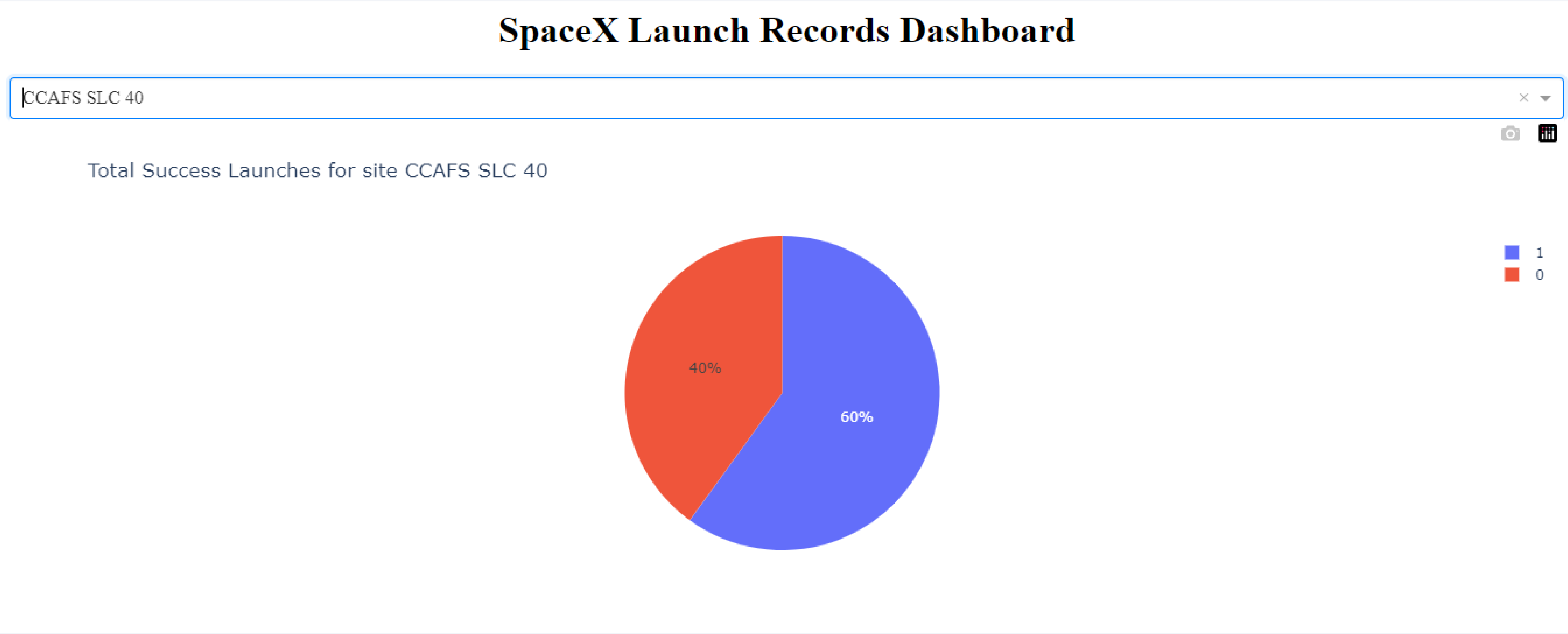
Section 4

Build a Dashboard with Plotly Dash

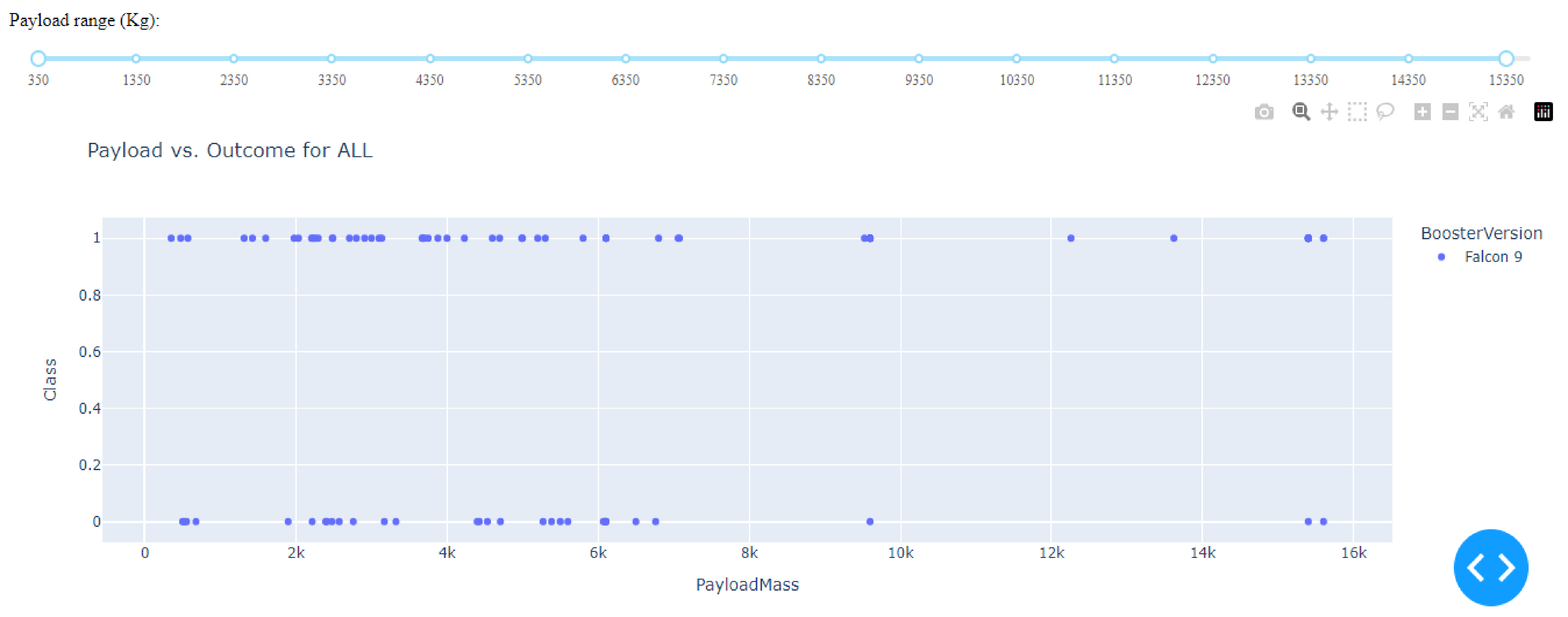
<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



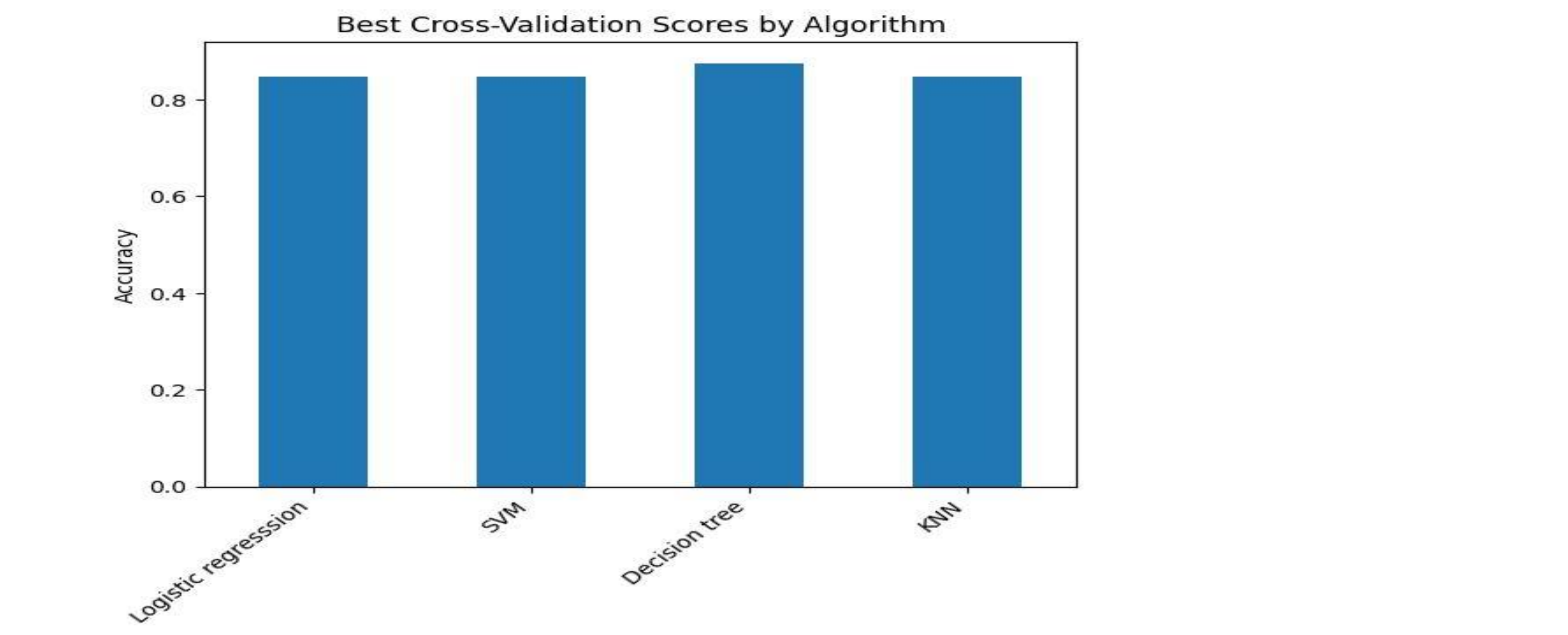
<Dashboard Screenshot 3>



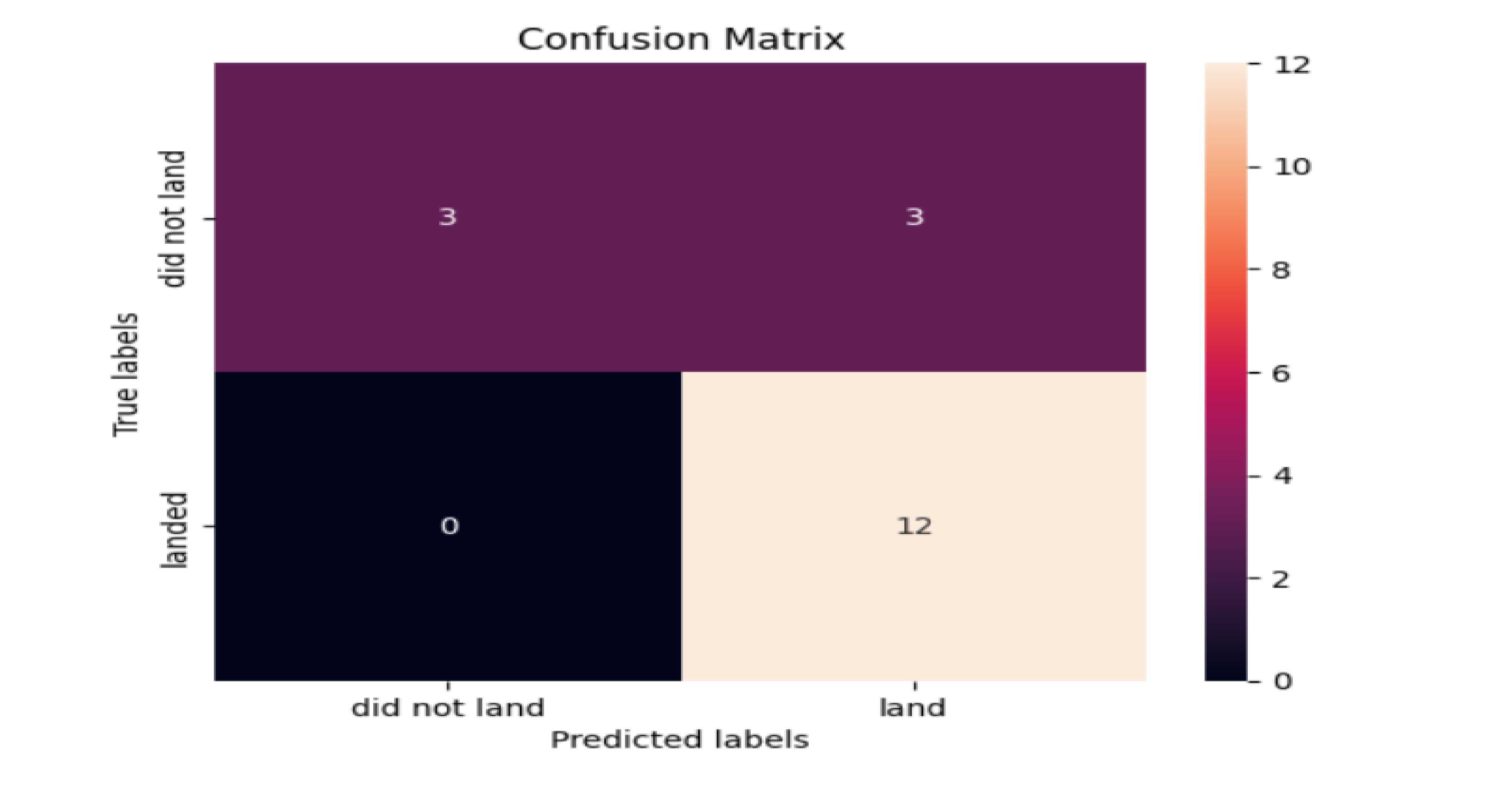
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

SpaceX primarily utilizes a select number of key launch sites that are optimized for frequent and successful missions.

The analysis reveals a substantial rise in the number of launches over time, demonstrating SpaceX's operational scalability and expansion

The Decision Tree Classifier emerged as the top-performing model in the predictive analysis, providing the highest accuracy in predicting future launch outcomes.

Visual analytics and SQL queries offered valuable insights into SpaceX's operational strategies, contributing to improved mission planning and enhancing future success rates.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!