IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohammed Naheemat Temitope

3rd October, 2024

Outline

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

Executive Summary

- **Objective:** Analyze SpaceX launch data to evaluate success and identify key performance factors.

- **Methodologies:**
    - Data collection cleaning and processing
    - Exploratory Data Analysis (EDA) using visualization and SQL
    - Interactive dashboards with Folium and Plotly Dash
    - Predictive analysis using logistic regression, SVM, decision trees, and KNN

- **Results:**
    - Identified significant predictors of launch success
    - Achieved over 90% accuracy with the best classification model
    - Insights to enhance future mission planning and safety

Introduction

- **Overview of SpaceX:** Founded in 2002 by Elon Musk, SpaceX aims to revolutionize space travel with reusable rockets, significantly lowering transportation costs and enabling future colonization of Mars.

- **Data Focus:** This project analyzes historical launch data to identify trends and factors affecting launch success, aiding in optimizing future missions and resource allocation.

- **Objective:** To evaluate SpaceX launch success using data analysis, providing insights to enhance mission planning and safety assessment.

Section 1

Methodology

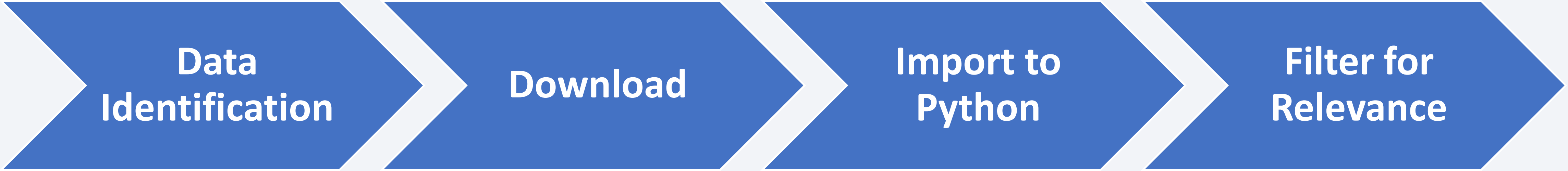
Methodology

- **Data Collection:**
  SpaceX launch data sourced from an open dataset. Loaded using pandas for analysis.
- **Data Wrangling and Processing:**
  Cleaned data by handling missing values, filtering columns, and standardizing formats for analysis.
- **EDA and SQL Analysis:**
  Used visualizations and SQL queries to explore trends in launch success, rocket types and launch pads.
- **Interactive Visual Analytics:**
  Created geographic maps with **Folium** and interactive dashboards using **Plotly Dash**.
- **Predictive Analysis:**
  Built classification models (Logistic Regression, SVM, Decision Tree, KNN) to predict launch success.
- **Model Tuning and Evaluation:**
  Applied **GridSearchCV** for hyperparameter tuning. Evaluated models with accuracy scores; top model achieved over 90% accuracy

Data Collection

- **Data Source:**
  SpaceX launch data collected from publicly available datasets.

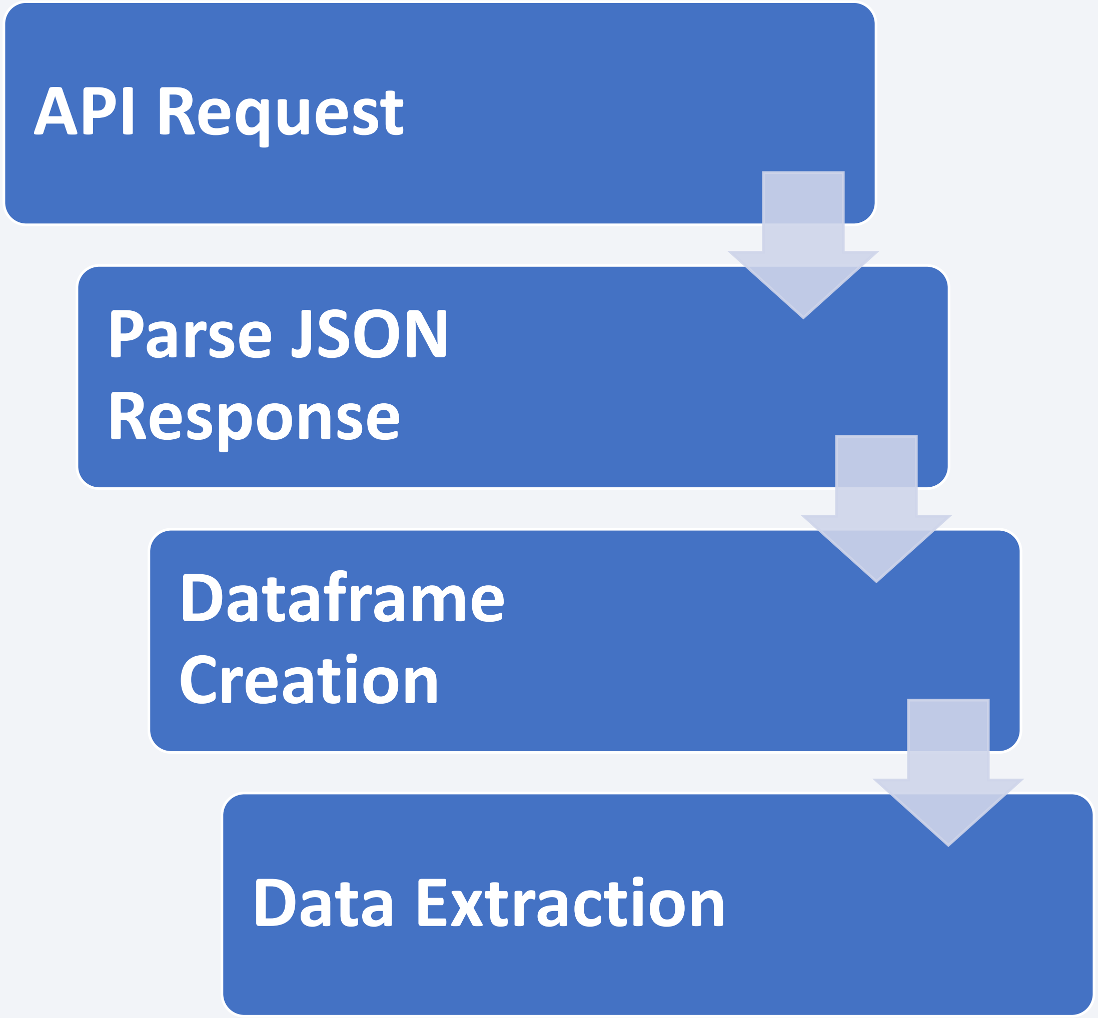
- **Collection Process:**
  Step 1: Identified and downloaded dataset from API and CSV files.
  Step 2: Imported into Python using pandas library for processing.
  Step 3: Filtered relevant columns: Rocket, Payloads, Launchpads, Cores



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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/Data_Collection_API.ipynb



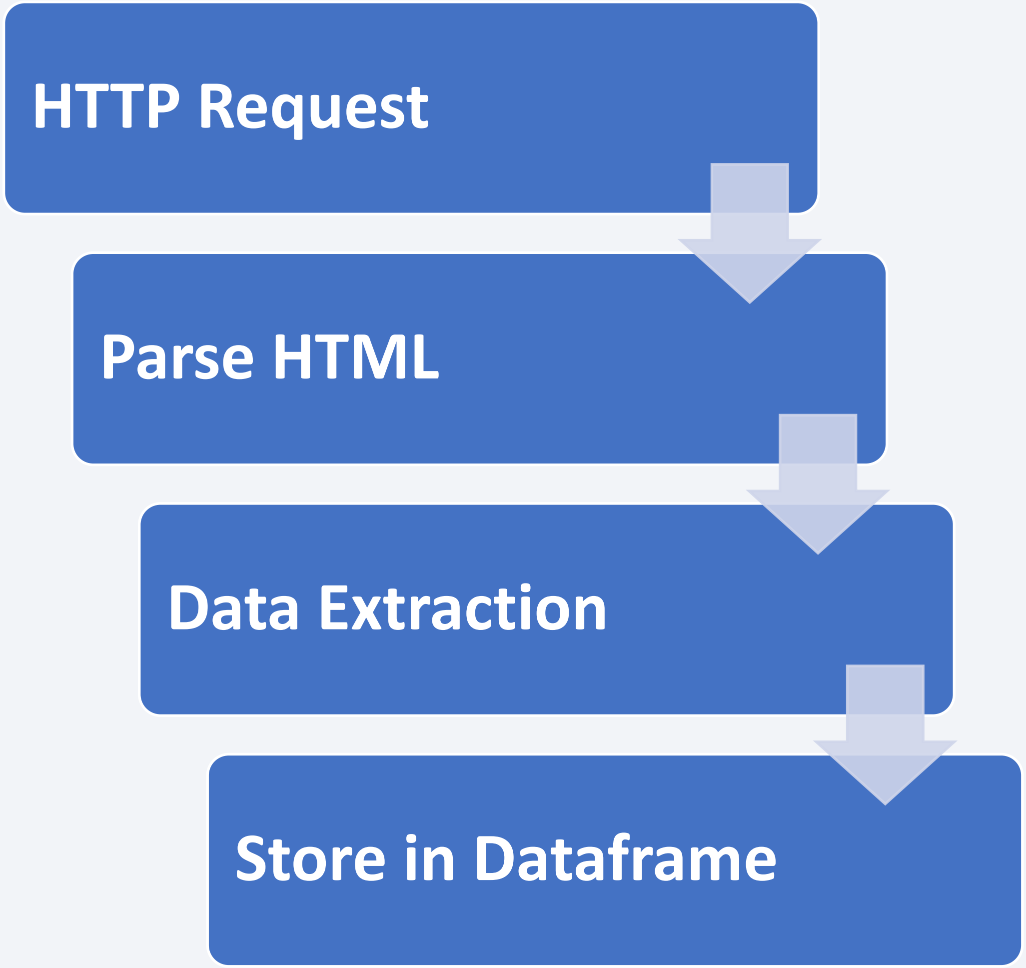
Data Collection - Scraping

- **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.

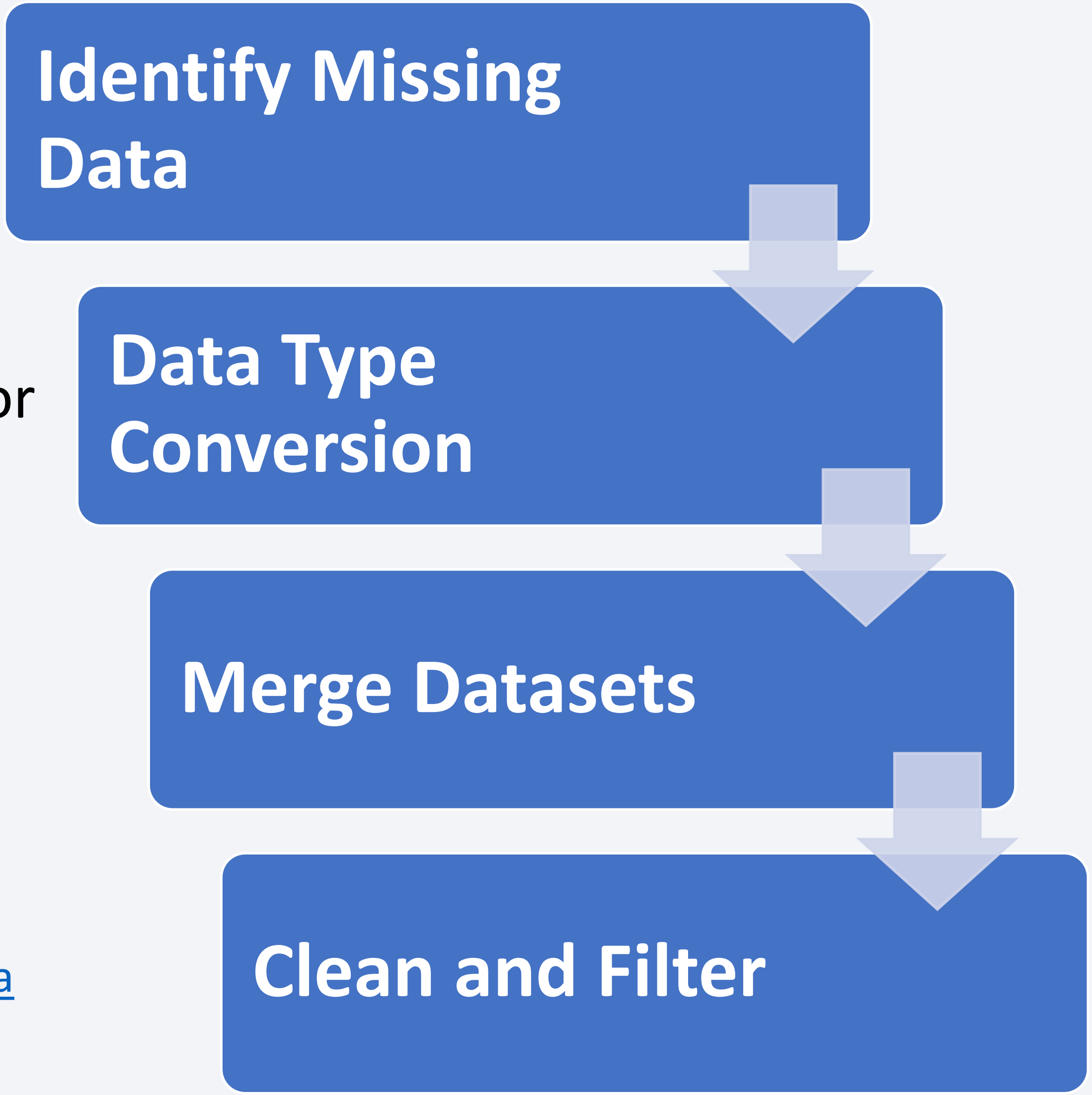
- **Step 3:** Extracted relevant data such as mission names, launch dates, and rocket details into a **pandas DataFrame**.



https://github.com/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/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

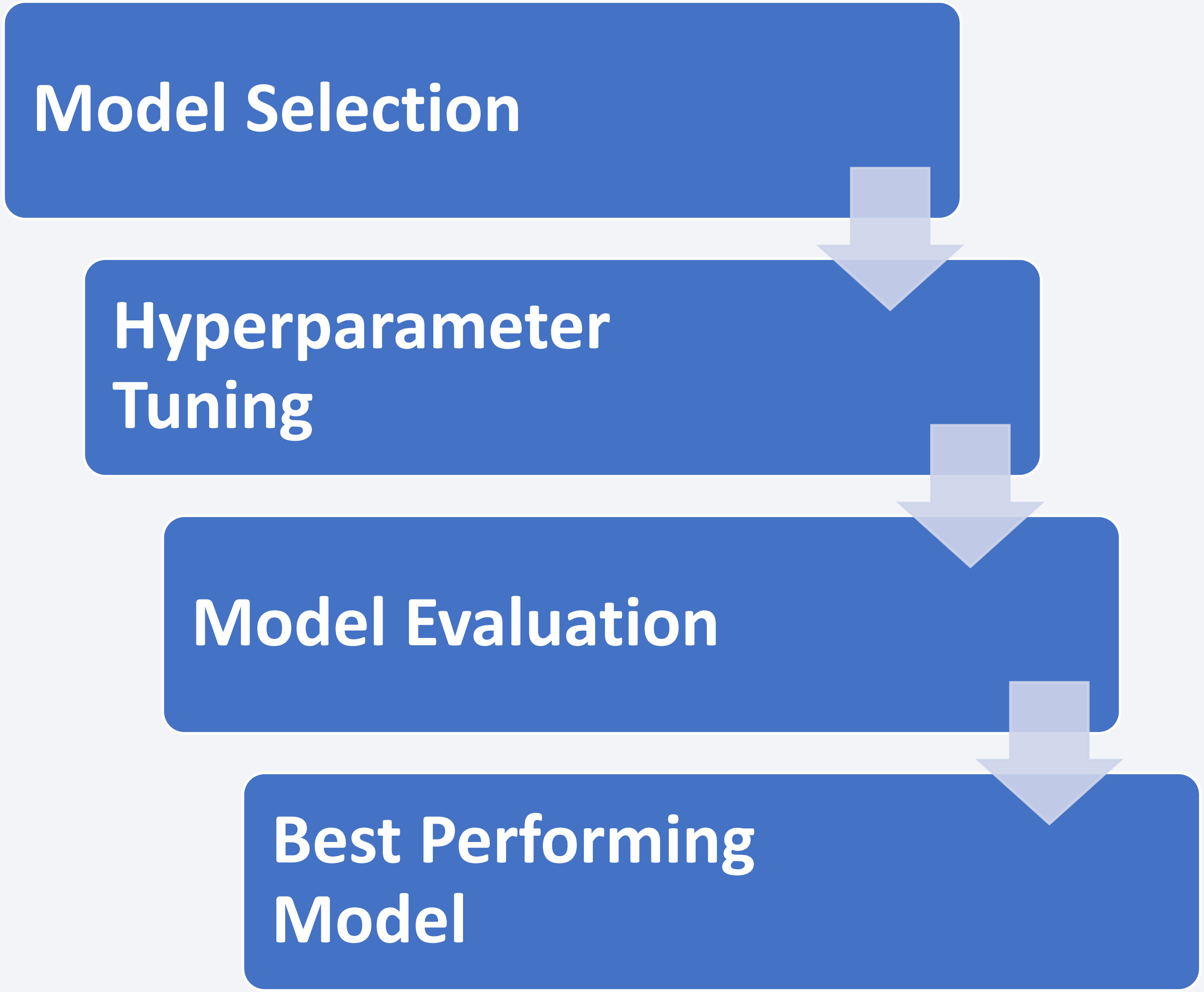
- https://github.com/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/dashboard.py

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/Mynaheemah/Naat/blob/main/Coursera%20Final%20Assignment/Machine_Learning_Prediction_Analysis.ipynb



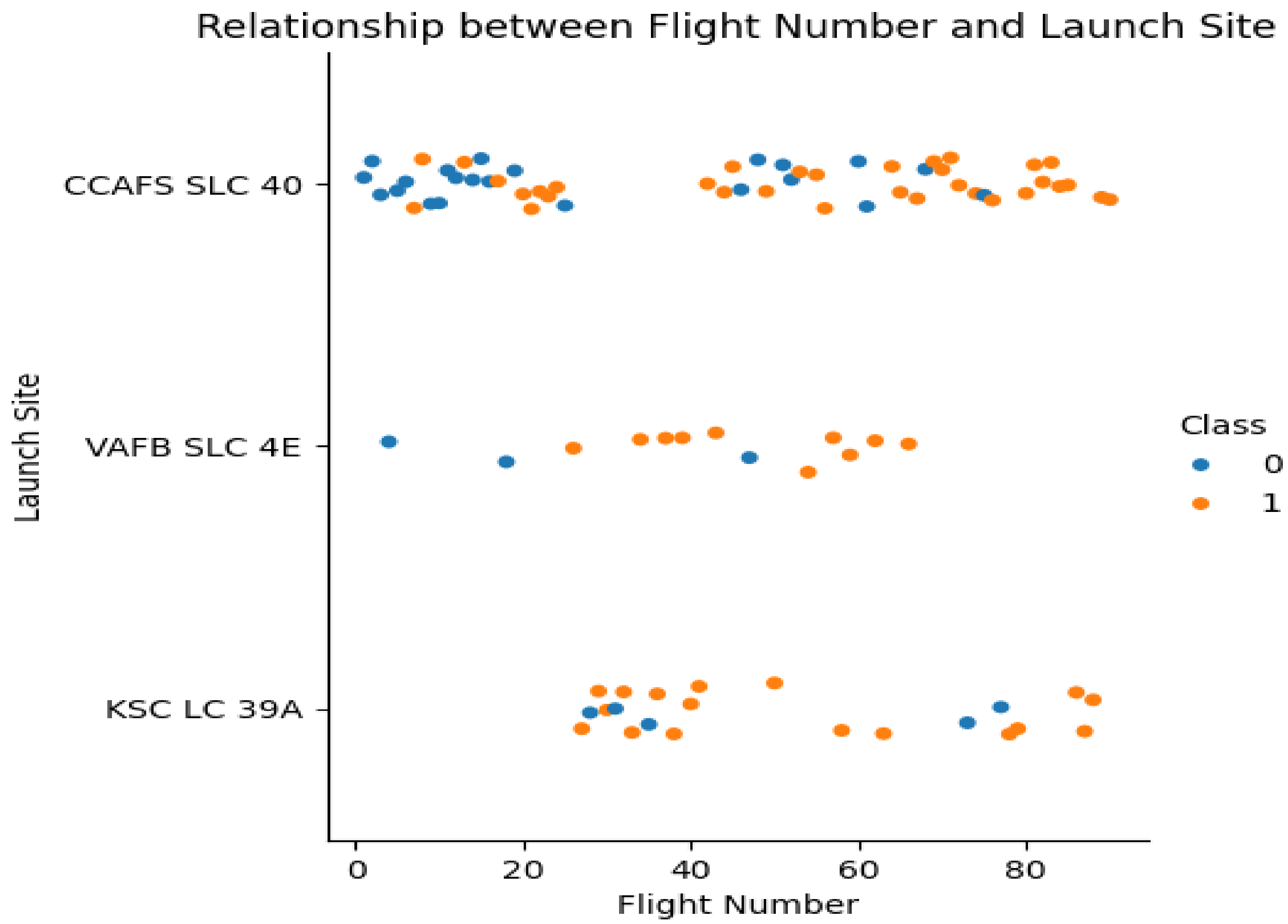
Results

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

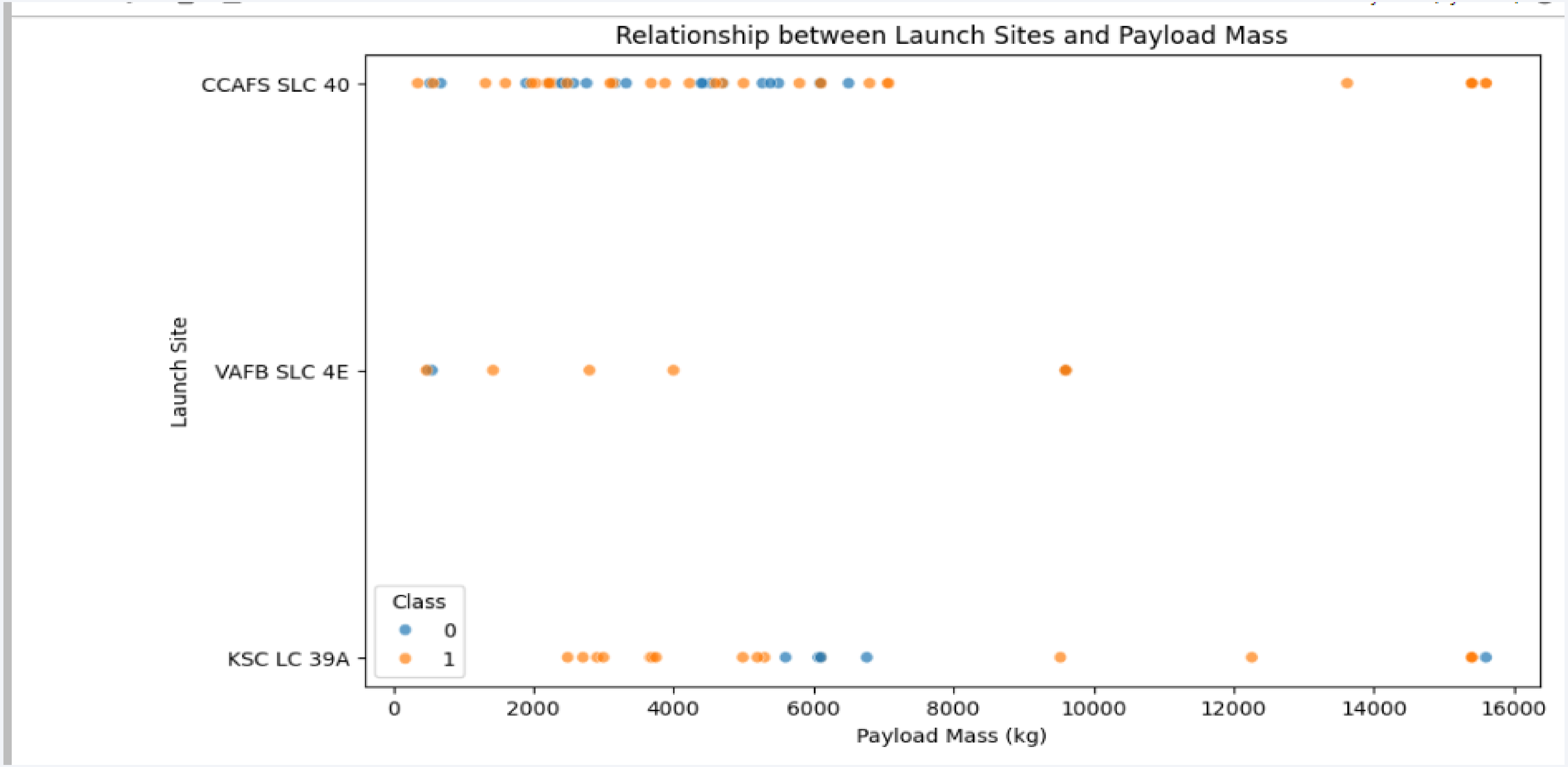
Section 2

Insights drawn from EDA

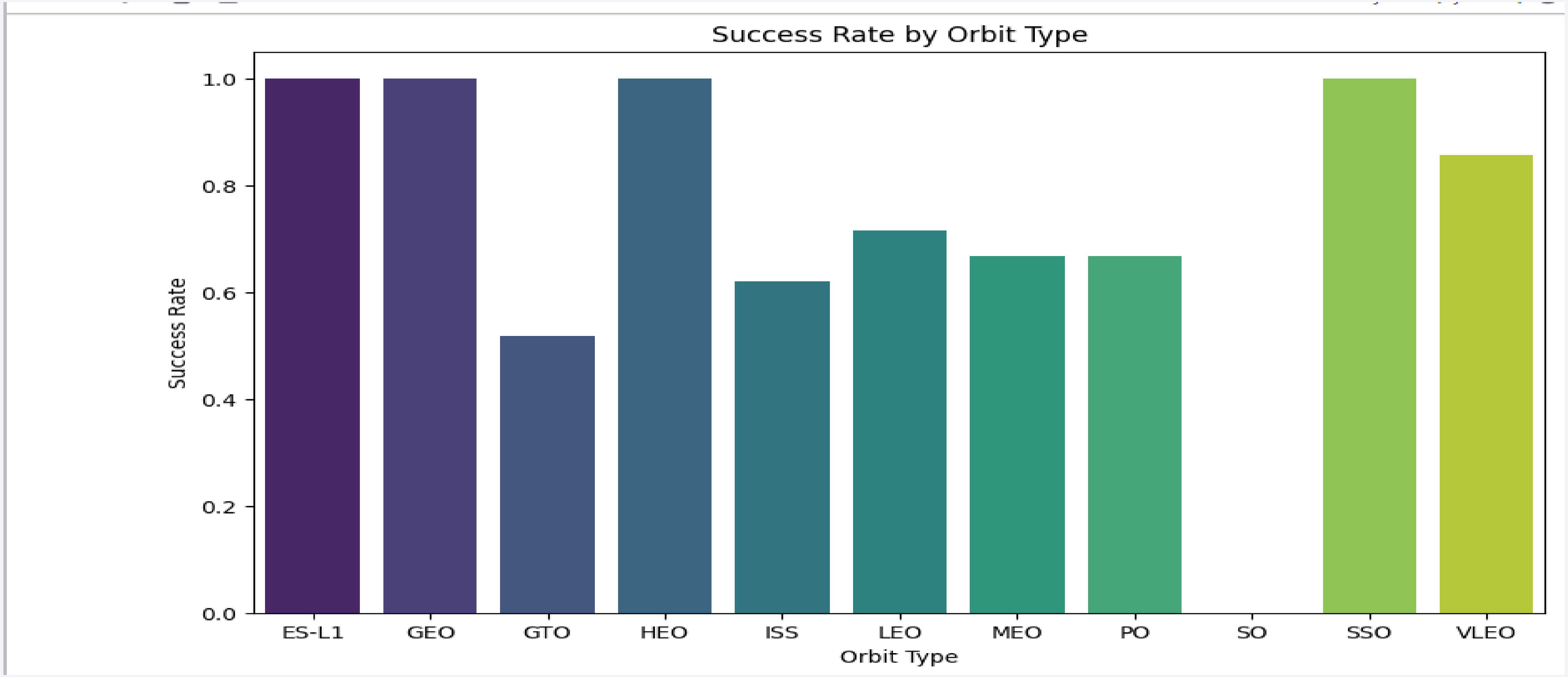
Flight Number vs. Launch Site



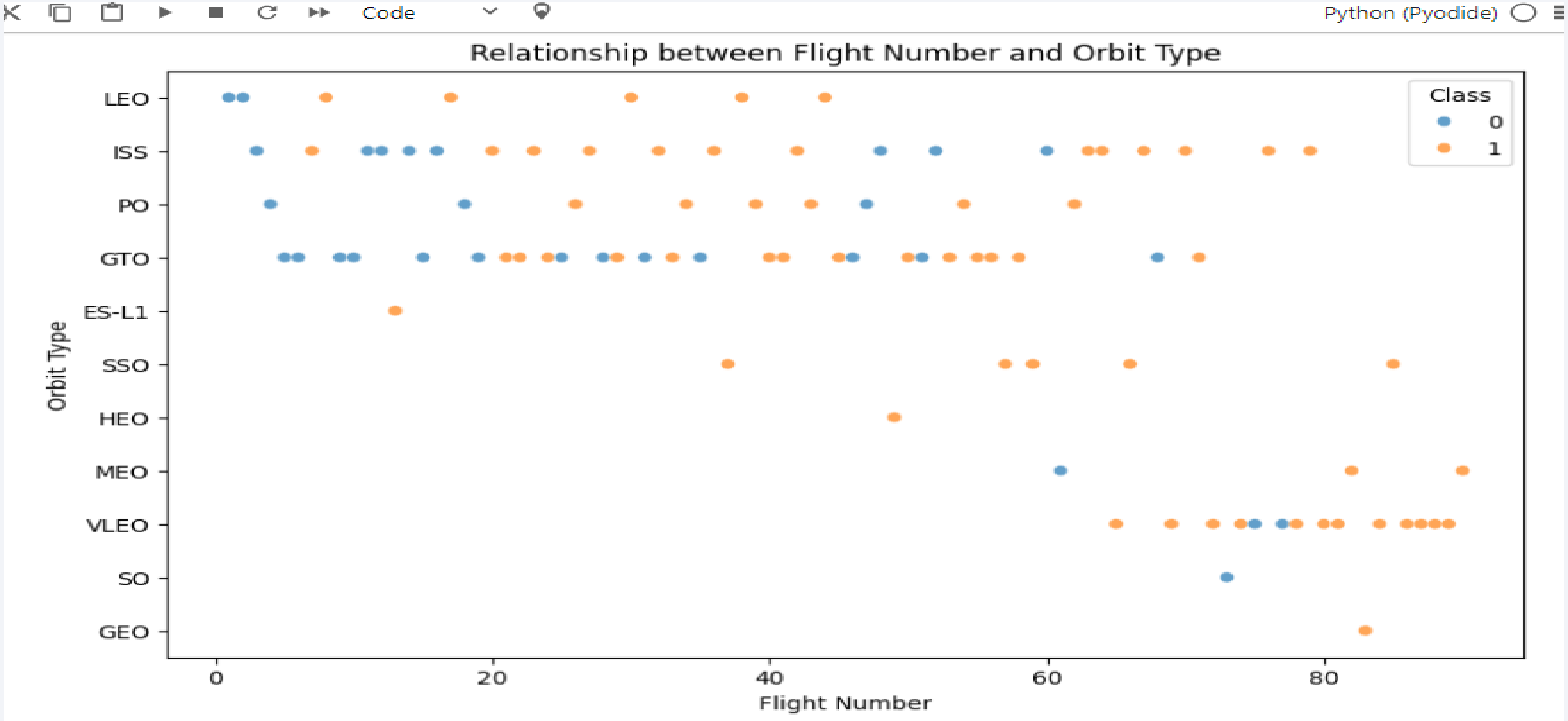
Payload vs. Launch Site



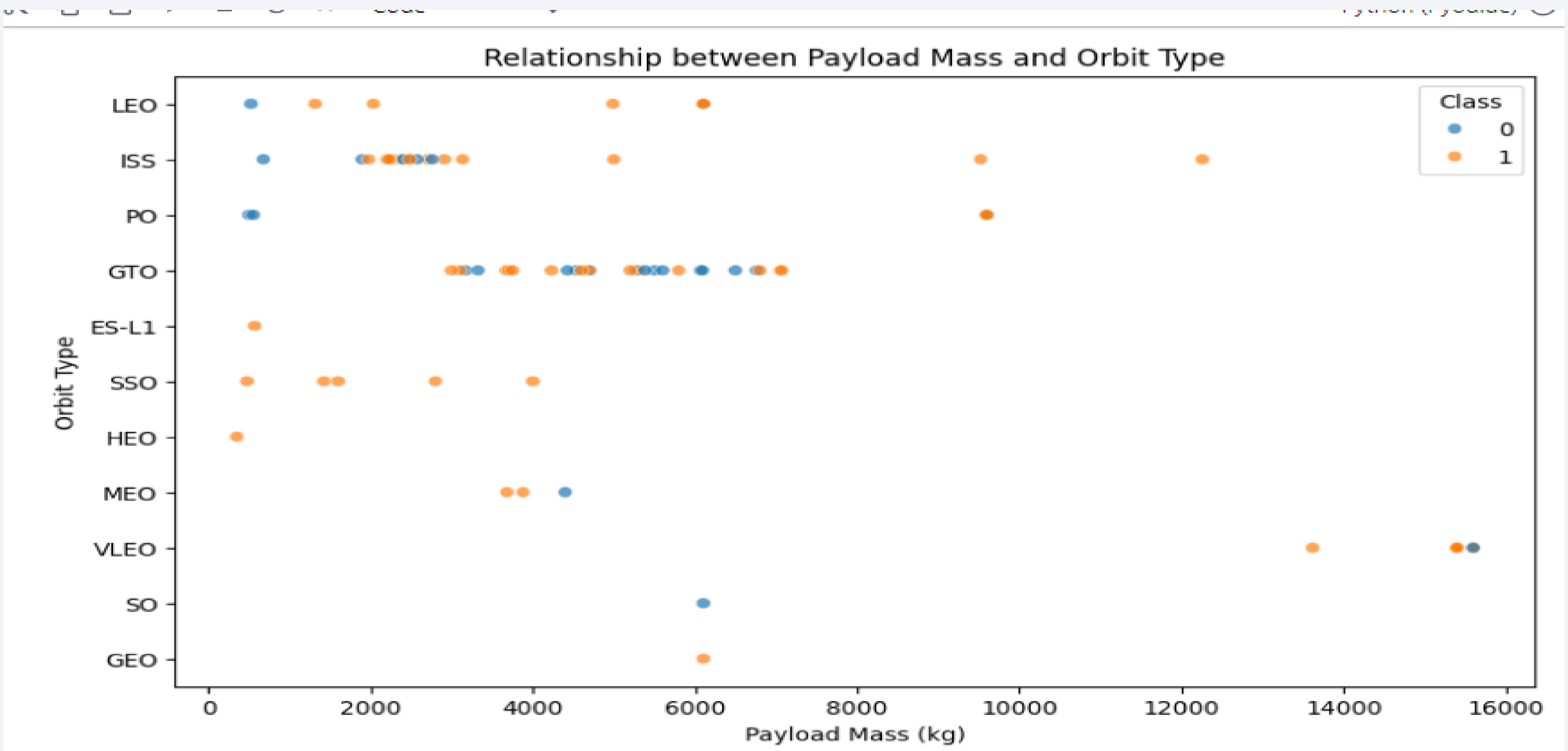
Success Rate vs. Orbit Type



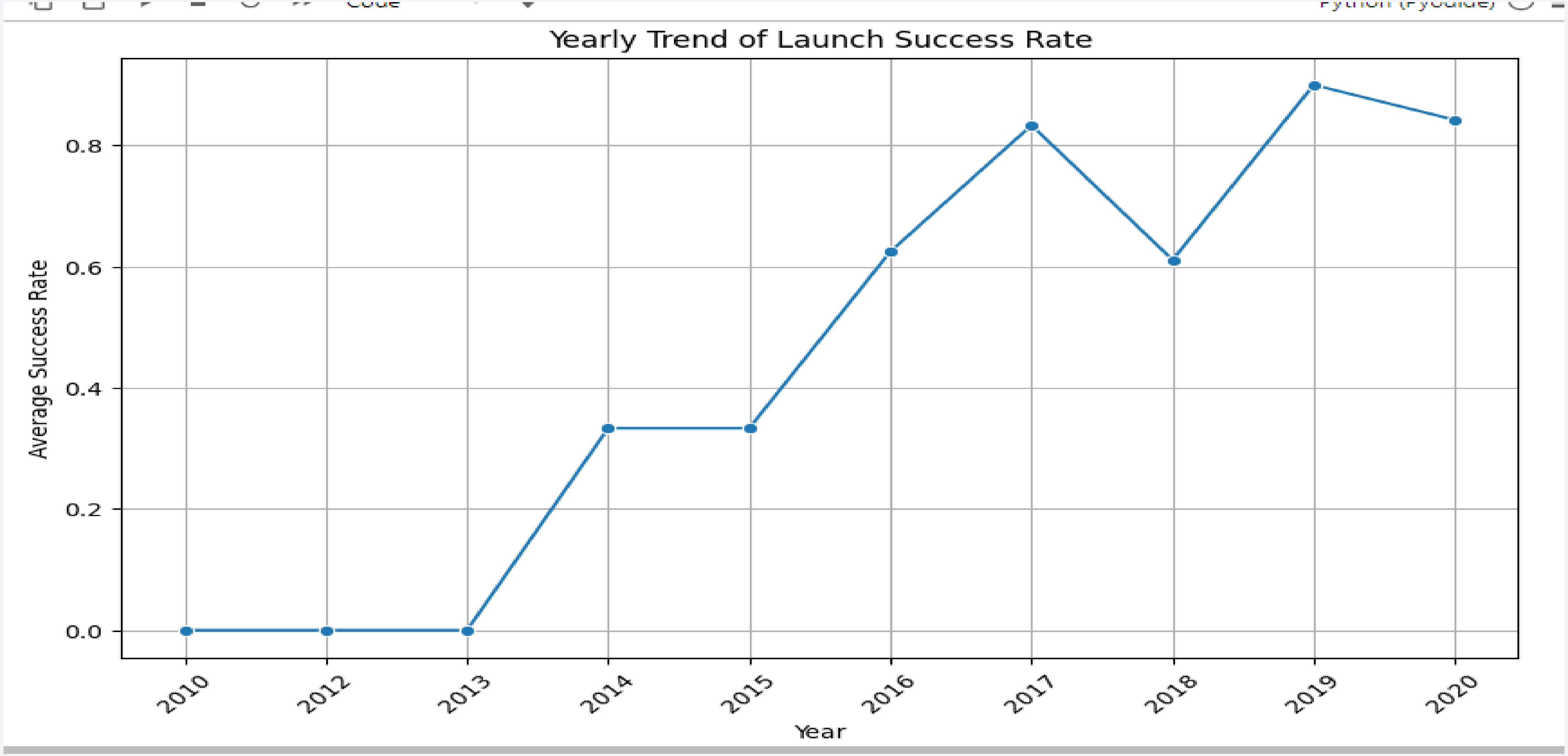
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

[13]:

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_Ou
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (par
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 (par
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No a
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No a
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No a

Total Payload Mass

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum(PAYLOAD_MASS__KG_) as TotalPayloadMass from SPACEXTABLE where Customer = 'NASA (CRS)';
```

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

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

Task 4

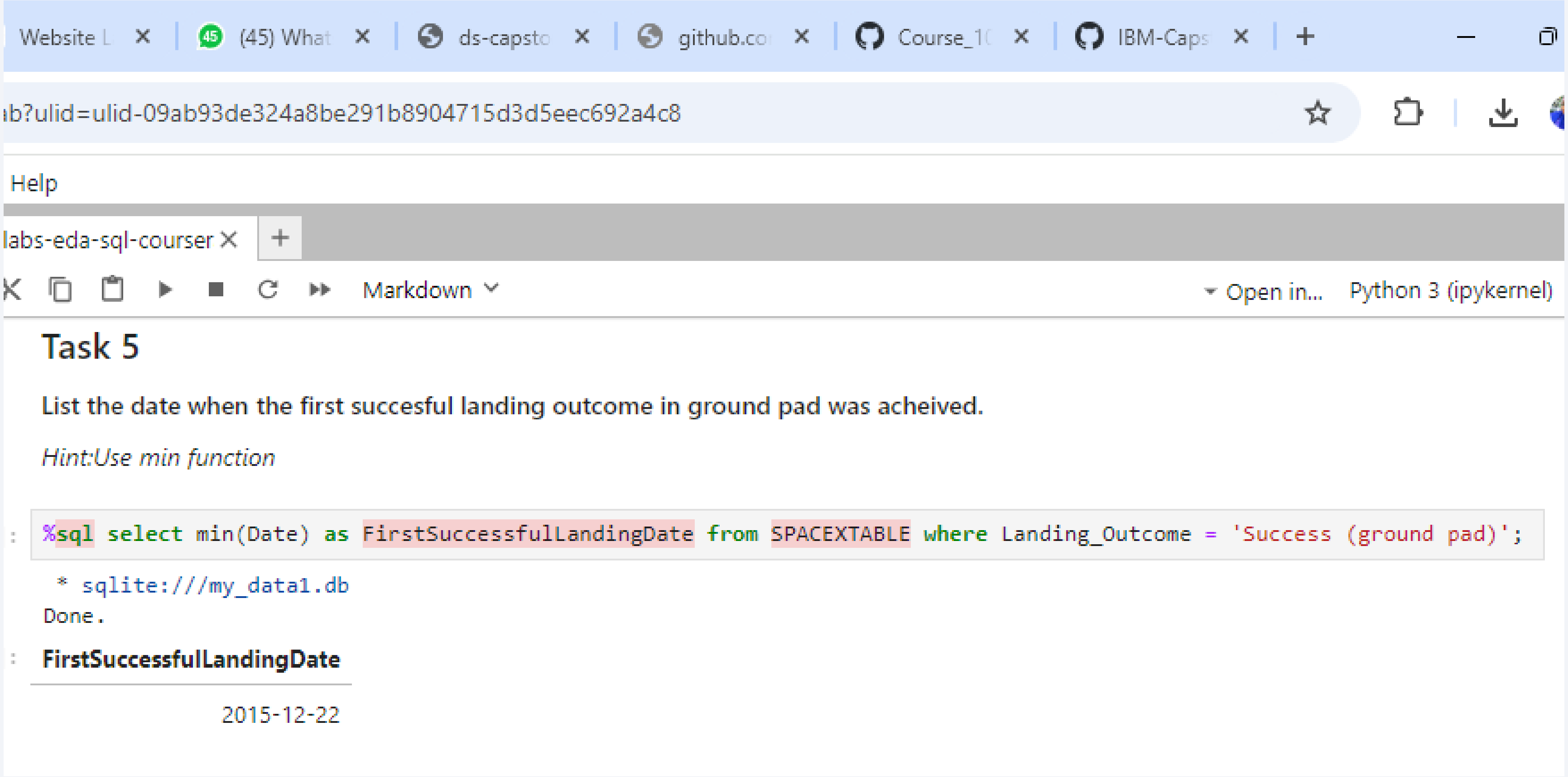
Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) as AveragePayloadMass from SPACEXTABLE where Booster_Version = 'F9 v1.1';
```

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

AveragePayloadMass
2928.4

First Successful Ground Landing Date



Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select distinct (Booster_Version) from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MAS
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, count(*) as Total from SPACEXTABLE 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

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%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

2015 Launch Records

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

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

```
 * 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

```
%sql select Landing_Outcome, COUNT(*) AS Outcome_Count from SPACEXTABLE where Date BETWEEN '2010-06-04' AND '2017-03-2
```

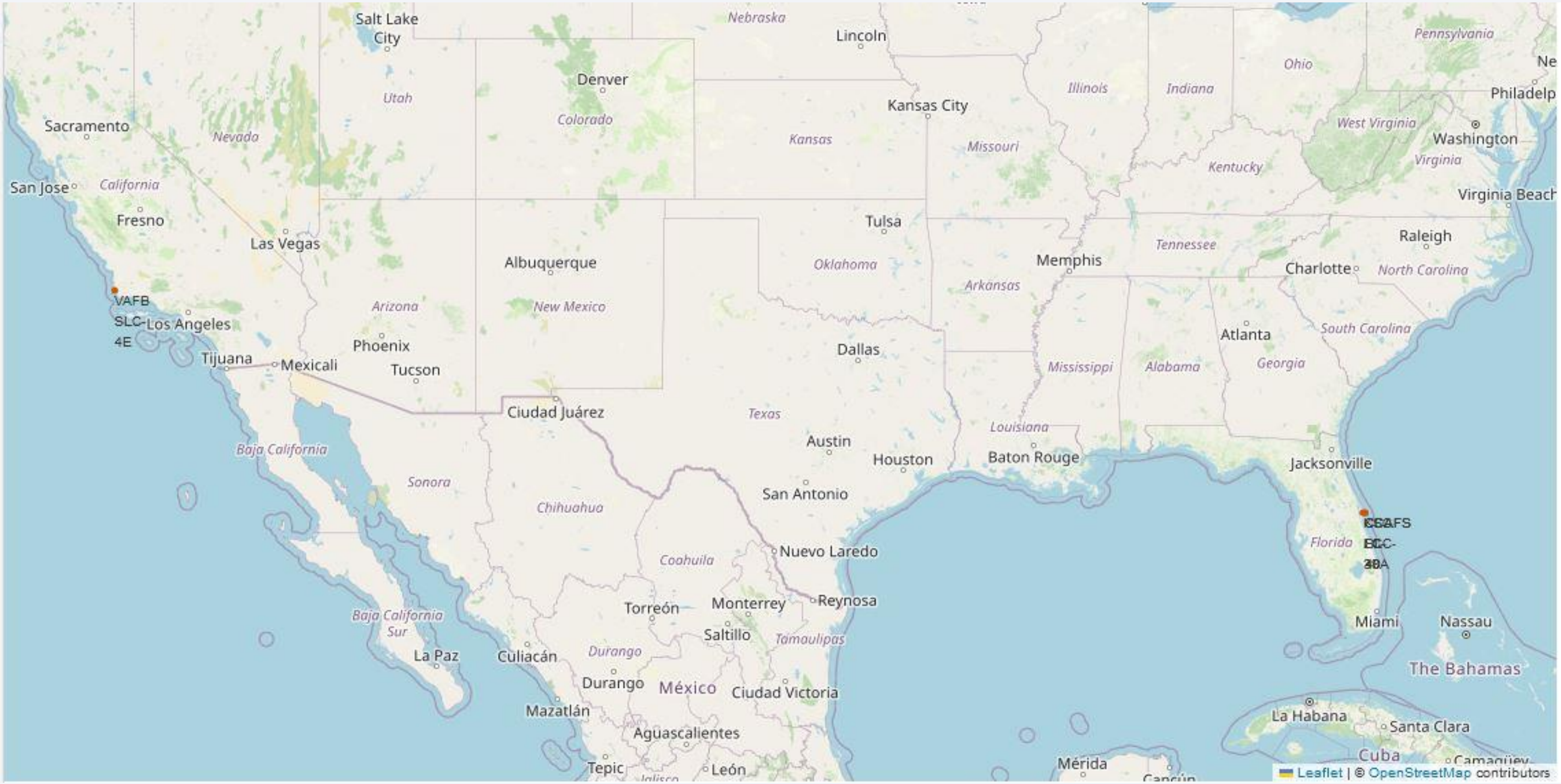
```
 * sqlite:///my_data1.db
Done.
```

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

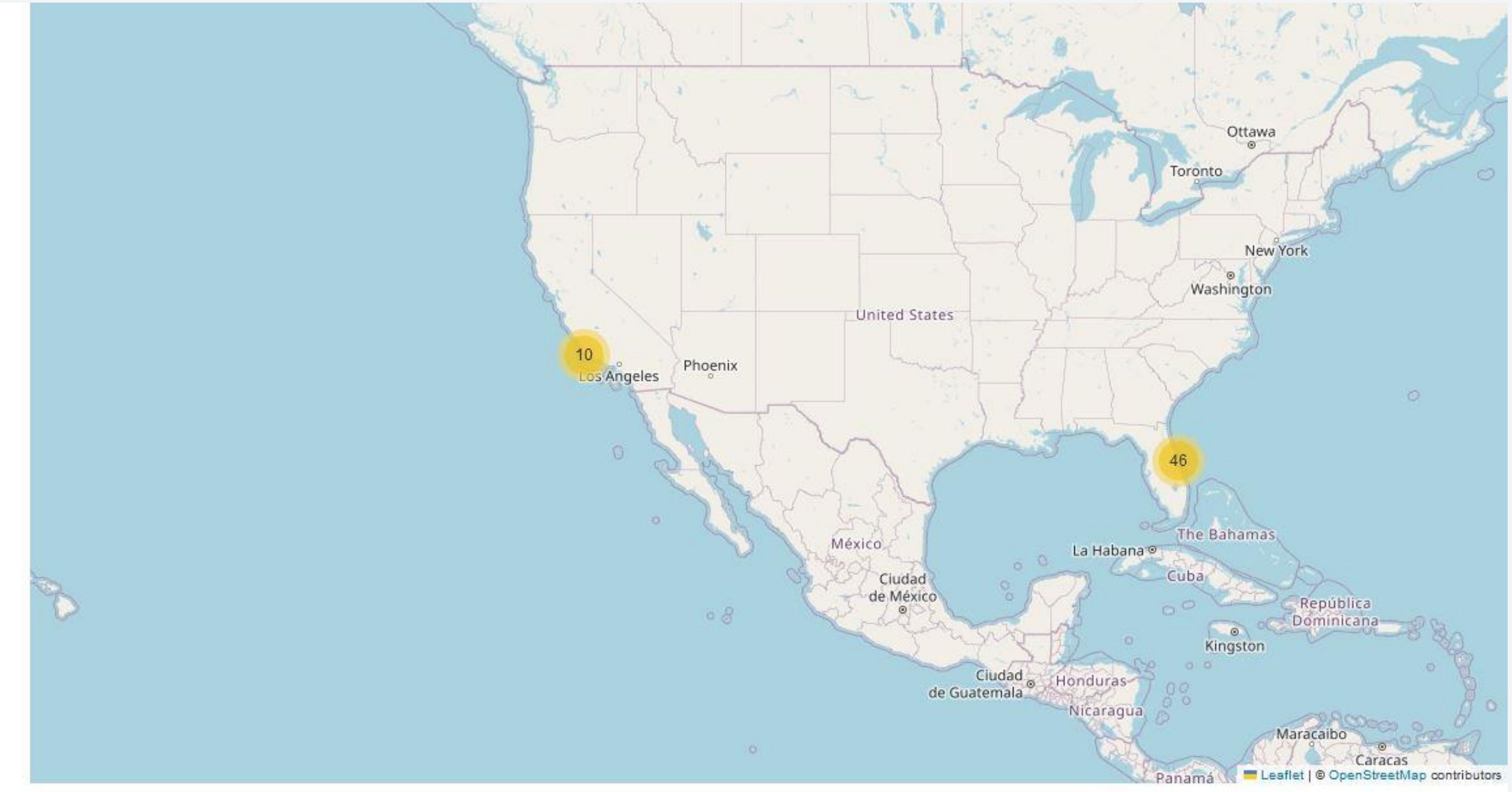
Section 3

Launch Sites Proximities Analysis

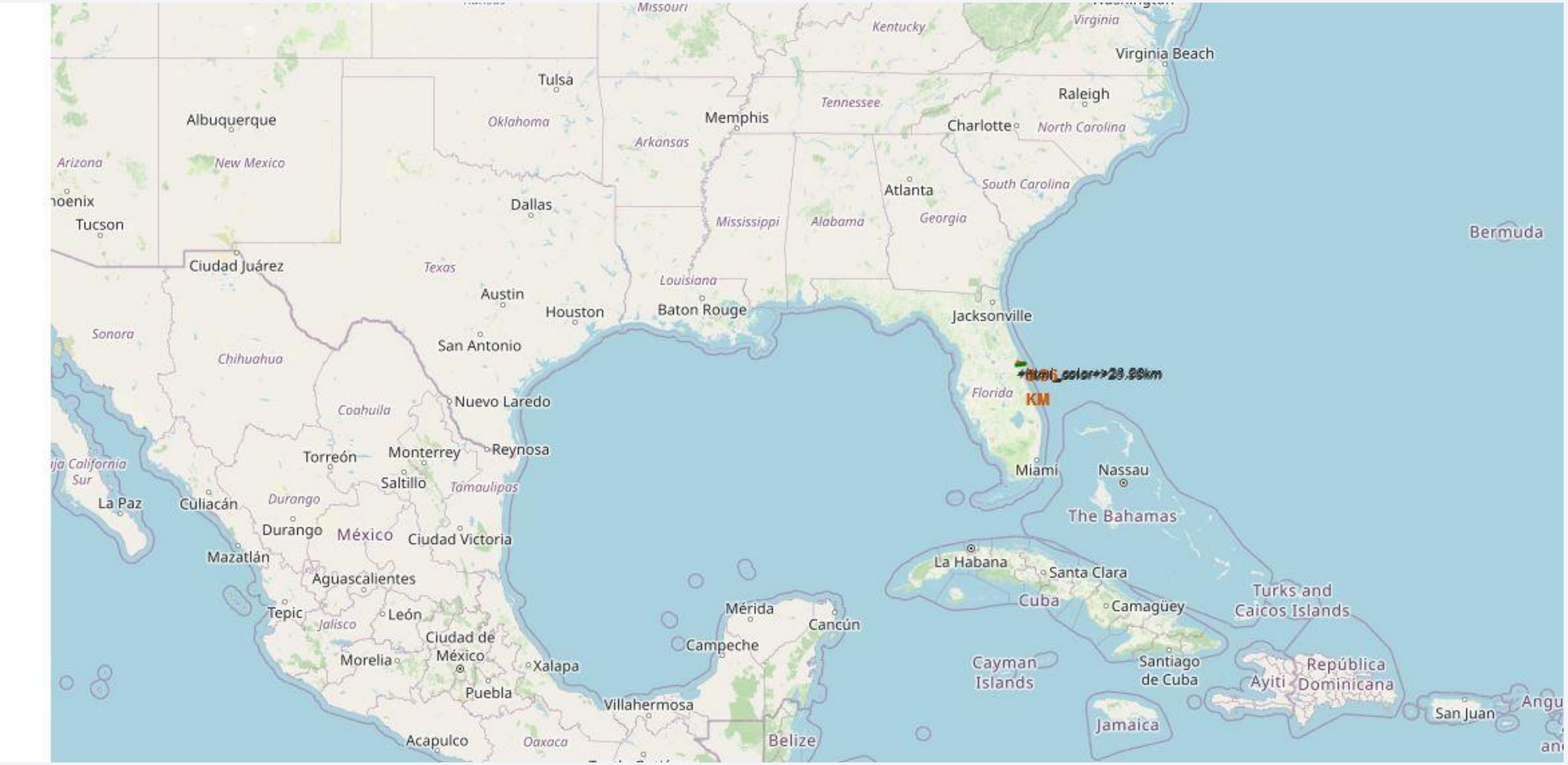
SpaceX Launch Sites' Location Markers on Global Map



Launch Outcomes Depiction on Global Map



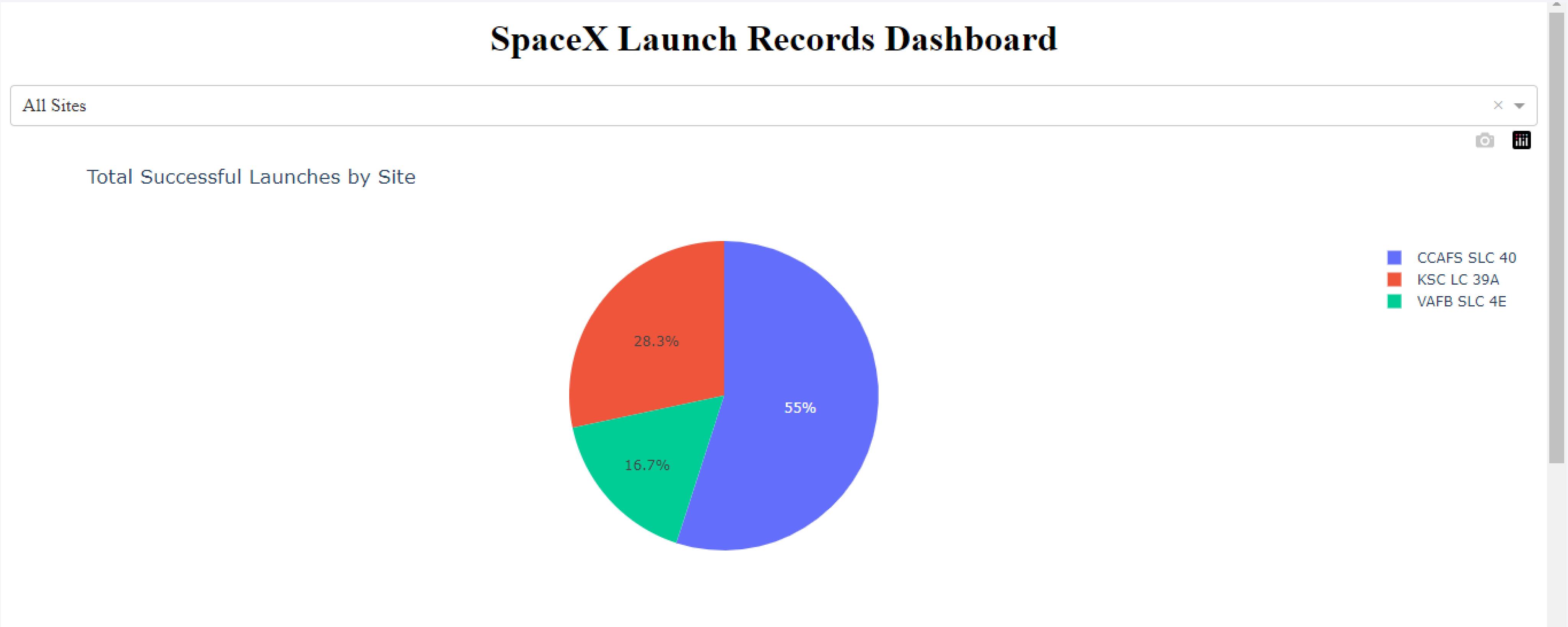
Proximity Analysis of Selected Launch Site on Global Map



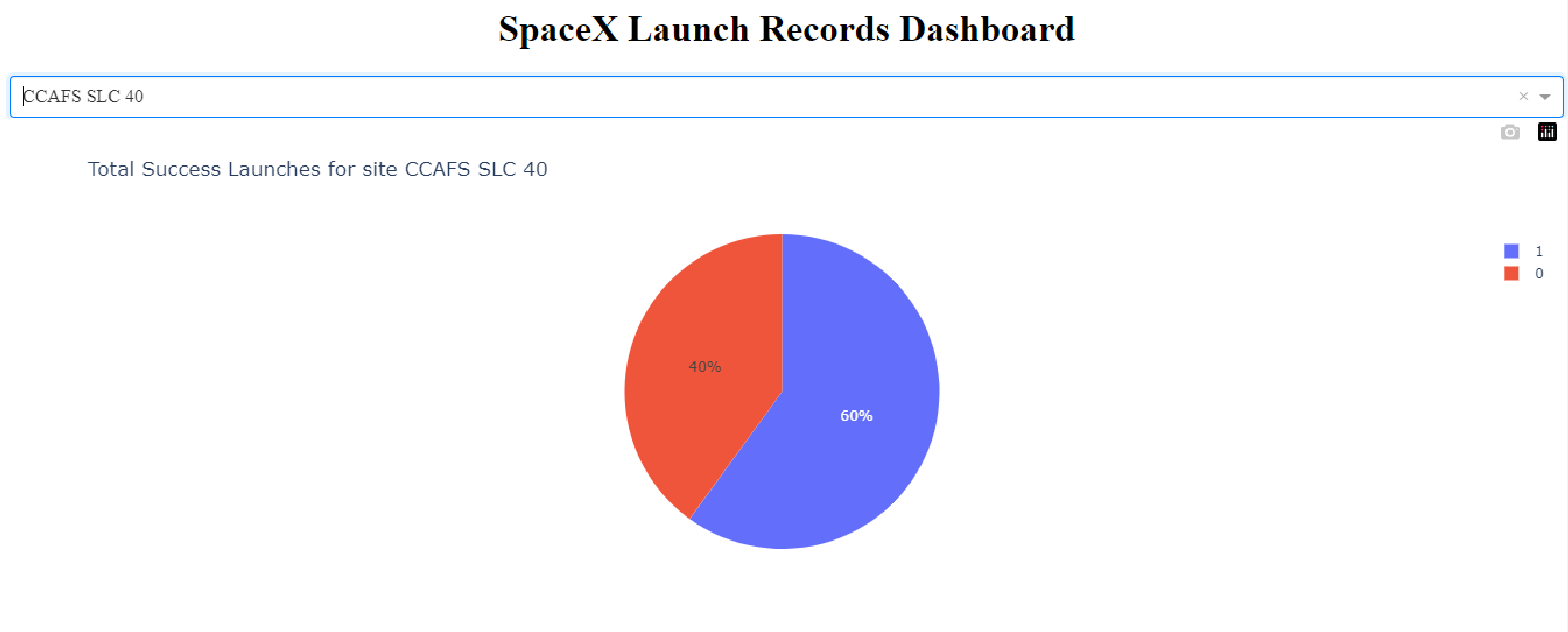
Section 4

Build a Dashboard with Plotly Dash

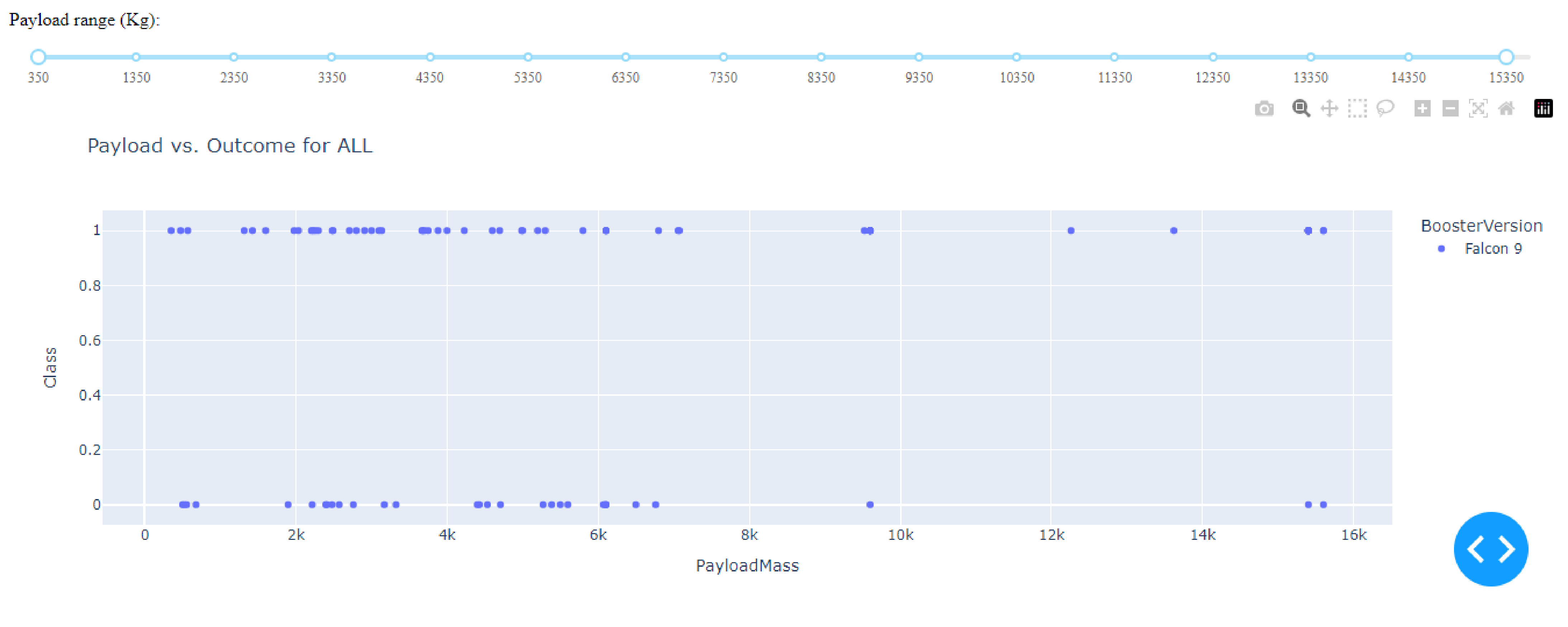
Total Successful Launches by Site



Total Success Launches for Site CCAFS SLC 40



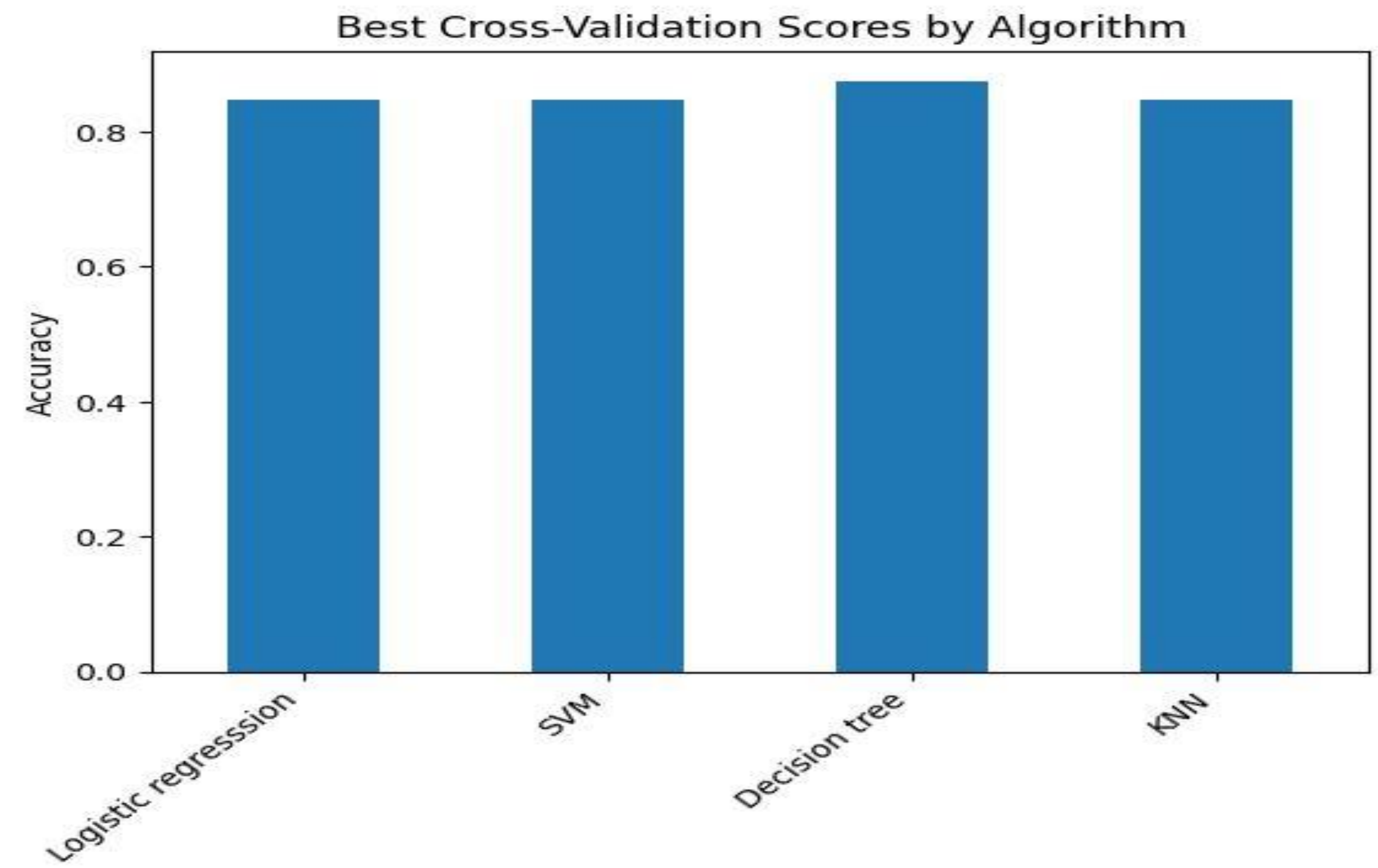
PayloadMass VS Outome for All Sites



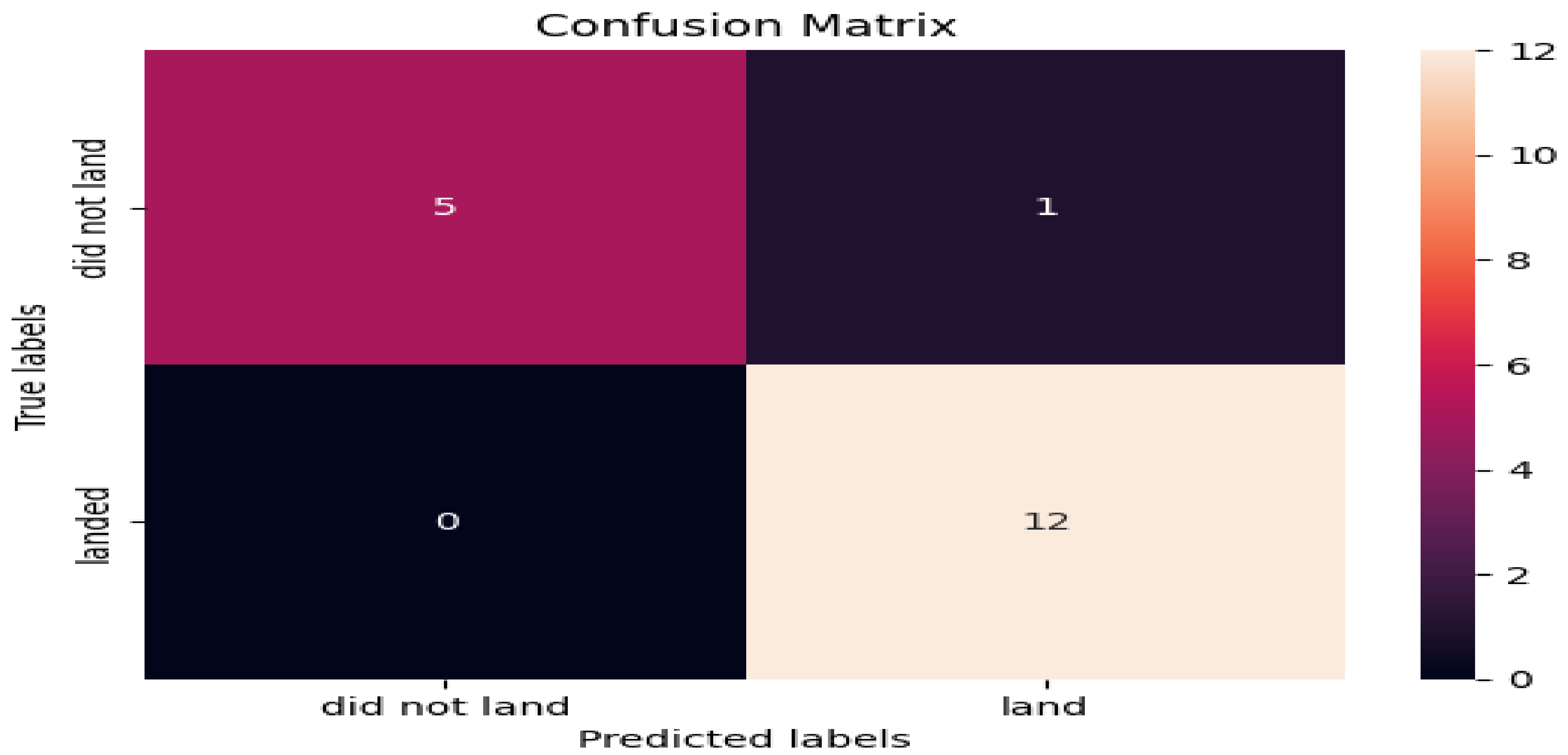
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix of Decision Tree Classifier Model



Conclusions

- SpaceX primarily uses a few key launch sites, which are optimized for frequent and successful launches.
- Analysis shows a significant increase in the number of launches over time, indicating SpaceX's operational scalability and growth.
- Factors such as flight number, payload, and core configurations influence the success of launches and overall mission reliability.
- The **Decision Tree Classifier** proved to be the best-performing model in the predictive analysis, delivering the highest accuracy in forecasting future launch outcomes.
- Visual analytics and SQL queries provided actionable insights for SpaceX's operational strategies, further aiding future mission planning and 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!