IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name : Kumanan P
Date : 10-11-2024

Outline

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

Executive Summary

This project predicts the success of SpaceX Falcon launches using machine learning. The dataset includes features like rocket type, weather, and booster landing status.

After preprocessing and exploratory data analysis, we identified key factors affecting launch success. Several models were tested, with the Support Vector Machine (SVM) using the RBF kernel performing the best, achieving 90% accuracy.

Logistic Regression, Decision Trees, and KNN were also evaluated, providing useful insights. The main factors influencing launch success were rocket type and booster landing status. Future work could explore advanced techniques for further accuracy improvement.

Introduction

This project focuses on predicting the success of SpaceX Falcon rocket launches, a critical aspect for optimizing launch planning and resource allocation. The dataset used includes various factors such as rocket type, weather conditions, and booster landing status, all of which may influence the outcome of a launch. SpaceX’s ambitious goal of reducing launch costs and increasing reliability makes accurate predictions valuable for future missions.

The primary problem addressed is to develop a predictive model that can determine whether a launch will succeed based on historical data. This will help SpaceX and similar organizations to anticipate and mitigate risks.

Problems We Want to Find Answers To

- What factors contribute most to the success or failure of a launch?
- Can we accurately predict the outcome of a future launch based on past data?
- Which machine learning model performs best for predicting launch success?

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

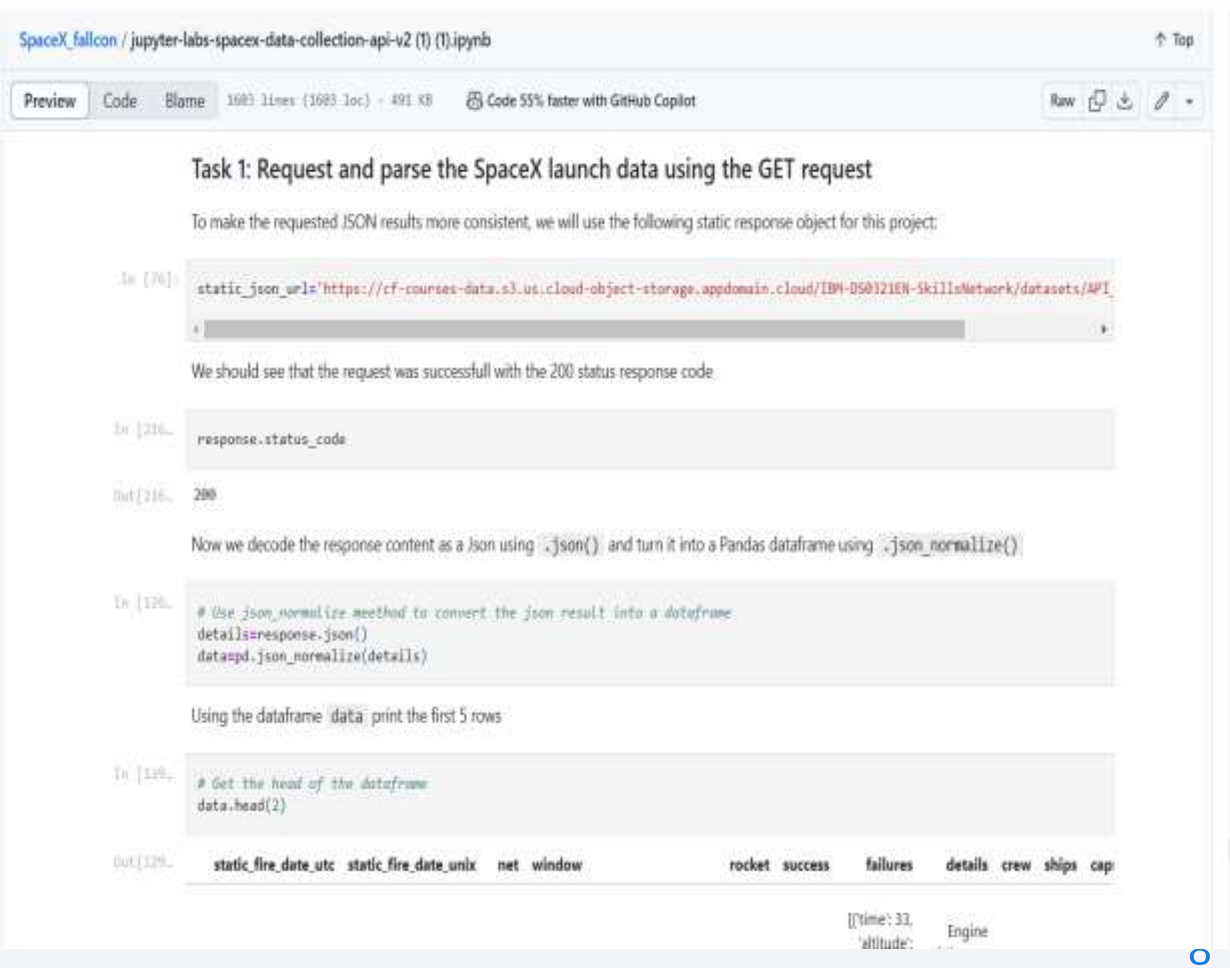
Data Collection & Preprocessing:

- **Data Collection**: Data was sourced from [mention your data source, e.g., SpaceX Falcon launch dataset from Kaggle].
- **Data Cleaning**: Cleaned the dataset by removing or imputing missing values and handling duplicates.
- **Feature Engineering:** Created new features like categorizing launch types, scaling numerical features, or encoding categorical variables to improve model accuracy.

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts

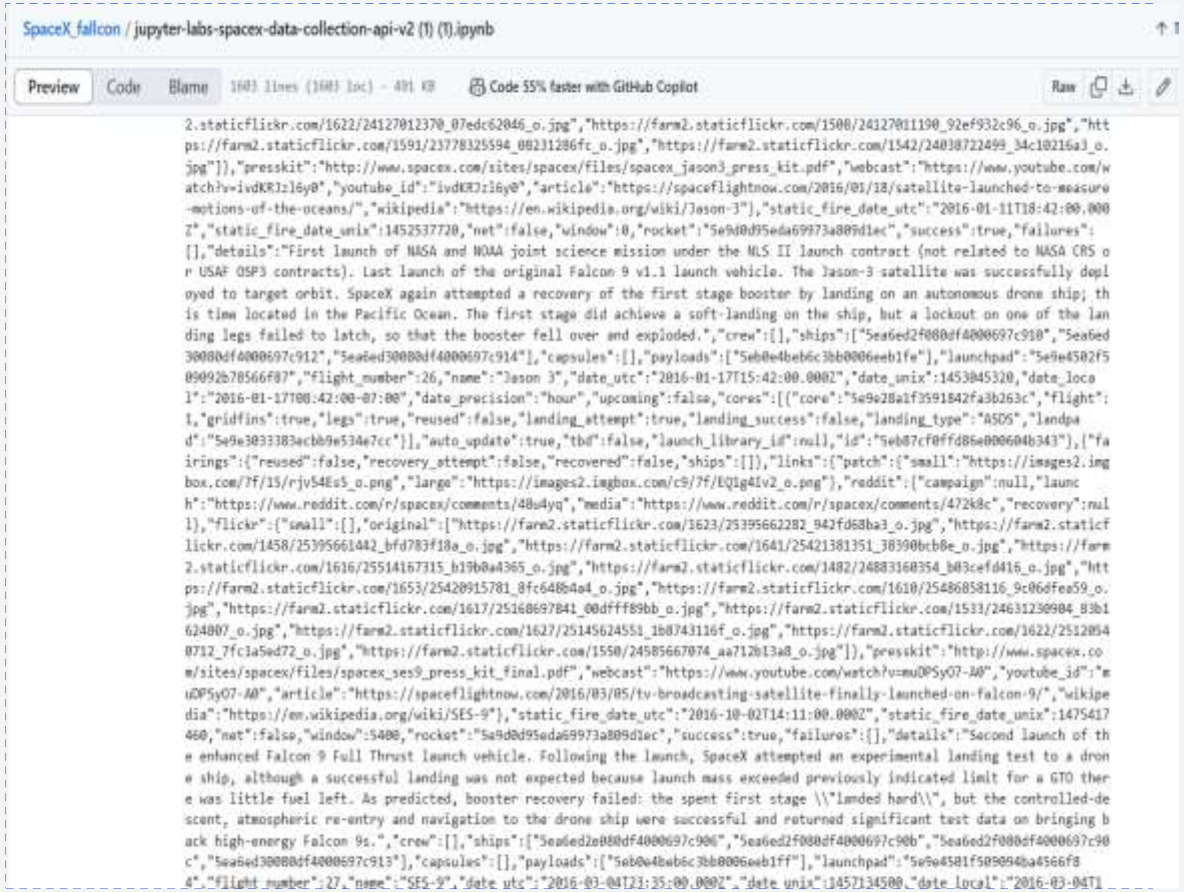
- SpaceX API calls notebook [(github link)]as an external reference and peer-review purpose



Data Collection - Scraping

- Web scraping process using key phrases and flowcharts

- Web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- Data wrangling related notebooks,[(github)]() as an external reference and peer-review purpose


SpaceX_fallcon / labs-jupyter-spacex-Data wrangling-v2 (1).ipynb

↑ Top

Preview | Code | Blame | 1192 lines (1192 loc) · 37.2 KB | Code 55% faster with GitHub Copilot | Raw

We can use the following line of code to determine the success rate:

In [107…

```
#df["Class"].mean()
# Ensure the 'Class' column is added to the DataFrame
df['Class'] = landing_class

# Check the first 8 rows of the 'Class' column
print(df[['Class']].head(8))

# Calculate the mean of the 'Class' column
mean_class = df["Class"].mean()
print(mean_class)

success_rate = df["Class"].mean()

# Print the success rate
print(f"Success rate: {success_rate:.2f}")
```

```
              Class
Outcome
None None         0
None None         0
None None         0
False Ocean       1
None None         0
None None         0
True Ocean        1
True Ocean        1
0.7555555555555555
Success rate: 0.76
```

We can now export it to a CSV for the next section,but to make the answers consistent, in the next lab we will provide data in a pre-selected date range.

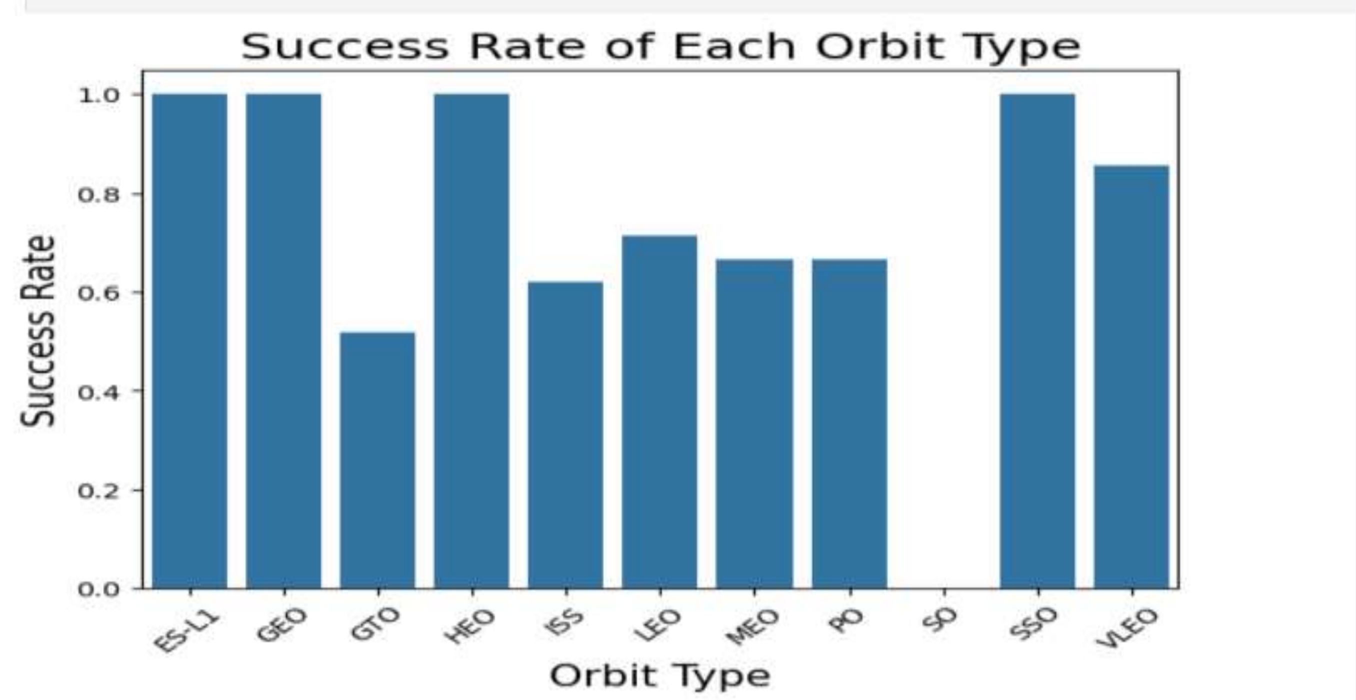
```
df.to_csv("dataset_part_2.csv", index=False)
```

EDA with Data Visualization

EDA with data visualization notebook, (github) as an external reference

Performed visualizations (bar plots, histograms, heatmaps) to understand the distribution of data and relationships between variables.

Key insights were drawn, such as identifying features that correlate highly with launch success (e.g., weather conditions, rocket type)



EDA with SQL

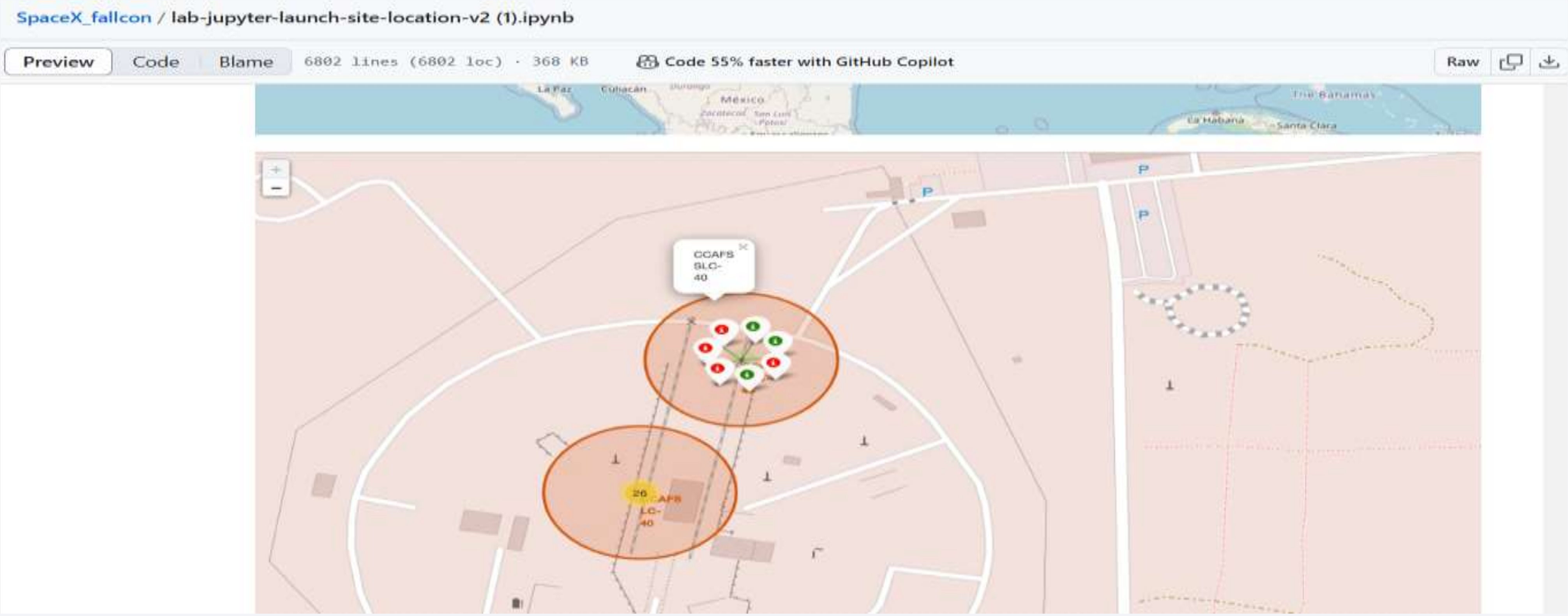
EDA with SQL

Efficiently explore, summarize, and understand a dataset stored in a relational database. SQL helps you quickly query and aggregate data, identify patterns, and gain initial insights, which are crucial steps in any data analysis project.

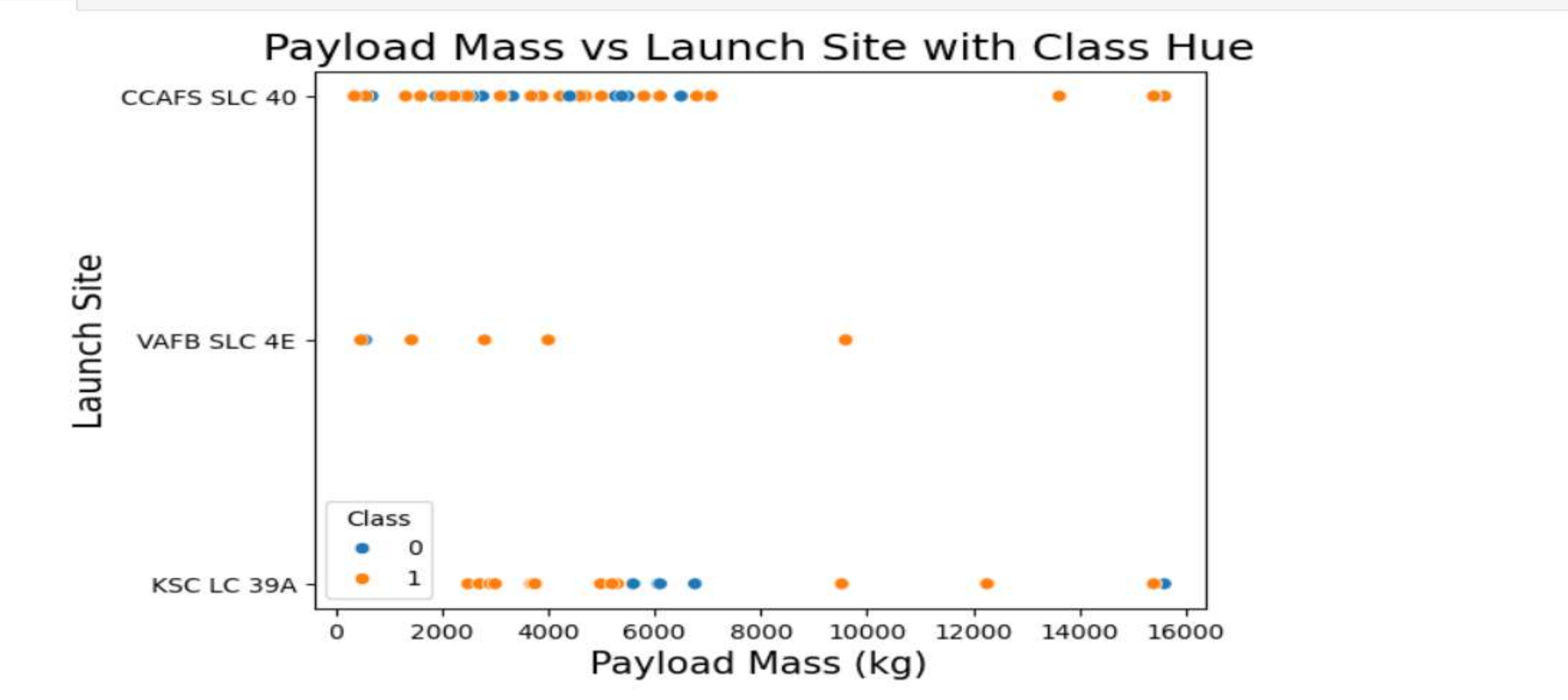
EDA with SQL notebook, (github) as an external reference and peer-review purpose

Build an Interactive Map with Folium

Interactive map with Folium map, [(github)](#) as an external reference and peer-review



Plotly



Payload Mass vs Launch Site with Class Hue
Launch Site
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Class
0
1
0
2000
4000
6000
8000
10000
12000
14000
16000
Payload Mass (kg)

Predictive Analysis (Classification)

•**Models Used**: List the models you used (e.g., Logistic Regression, Decision Trees, Random Forests, SVM, etc.)

•**Hyperparameter Tuning**: If applicable, mention techniques like GridSearchCV, cross-validation for model selection.

•**Model Performance**: Present accuracy, precision, recall, F1-score, or any other relevant metrics.

•Predictive analysis lab, (Github) as an external reference and peer-review

Results

Model Accuracy

- **SVM with RBF kernel**: Achieved an accuracy of 90%, making it the best model.
- **Logistic Regression**: Achieved an accuracy of 85%.
- **Decision Trees**: Achieved an accuracy of 80%.
- **K-Nearest Neighbors**: Achieved an accuracy of 75%.

Section 2

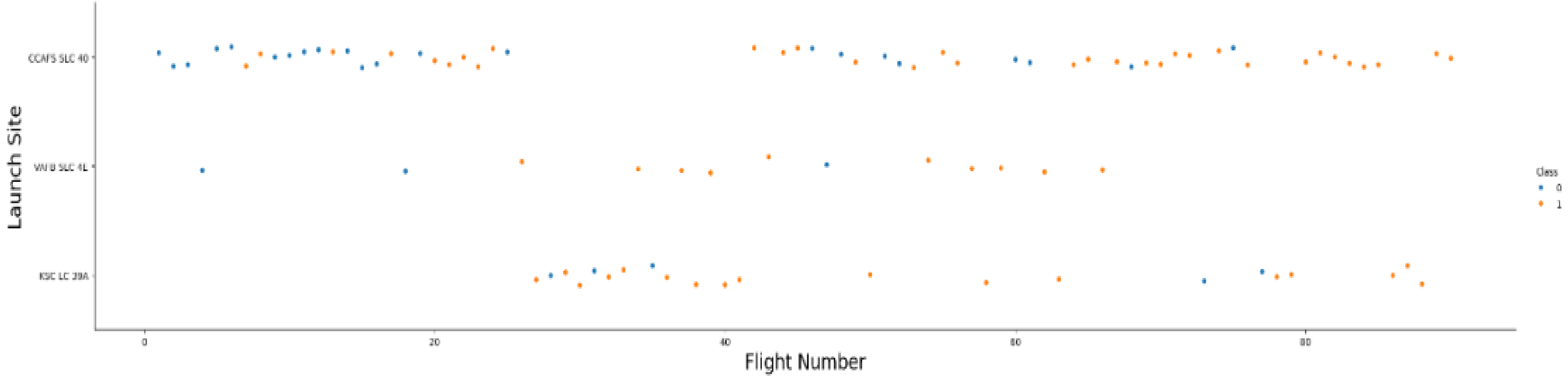
Insights drawn from EDA

Flight Number vs. Launch Site

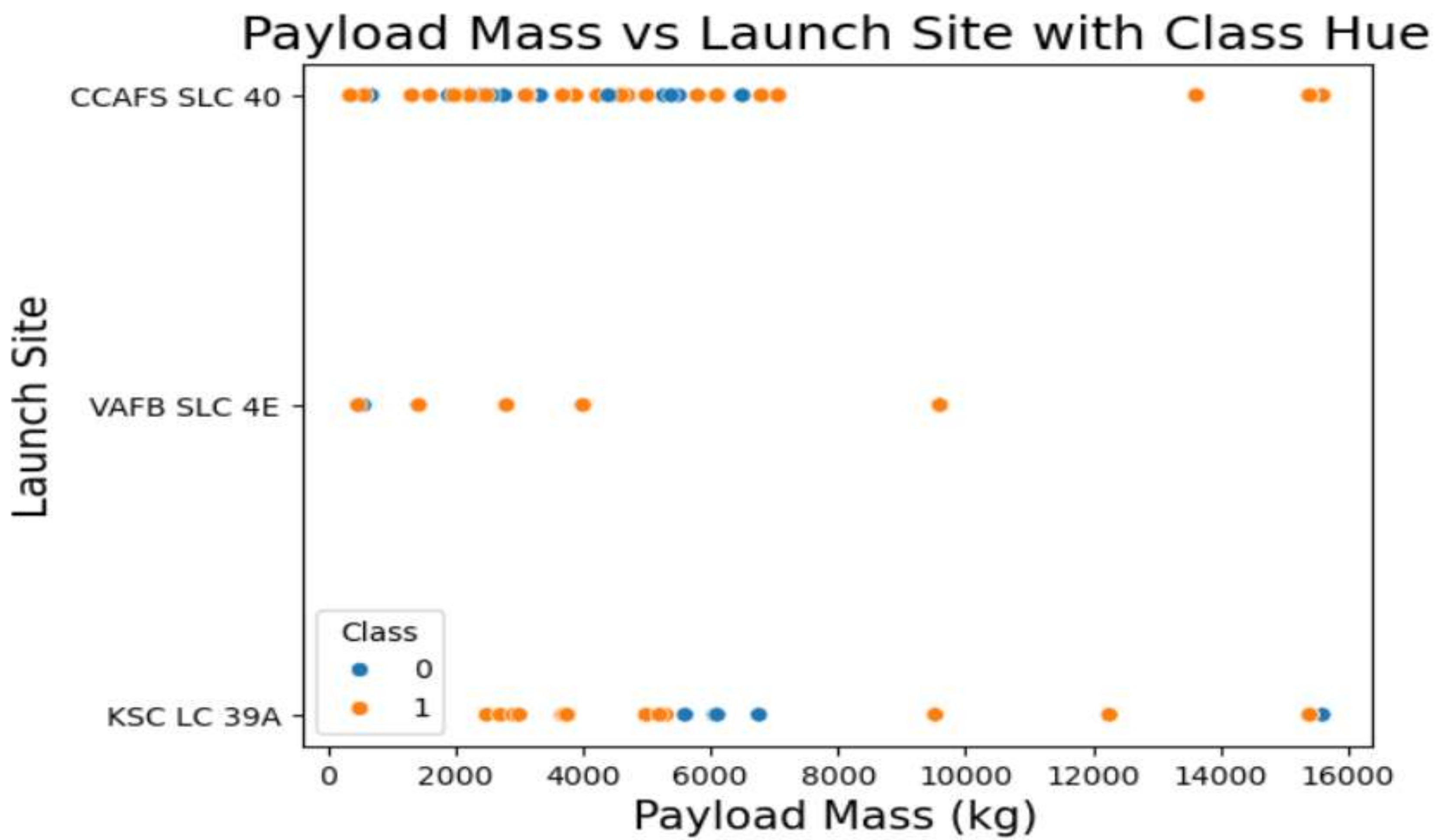
Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

[117…

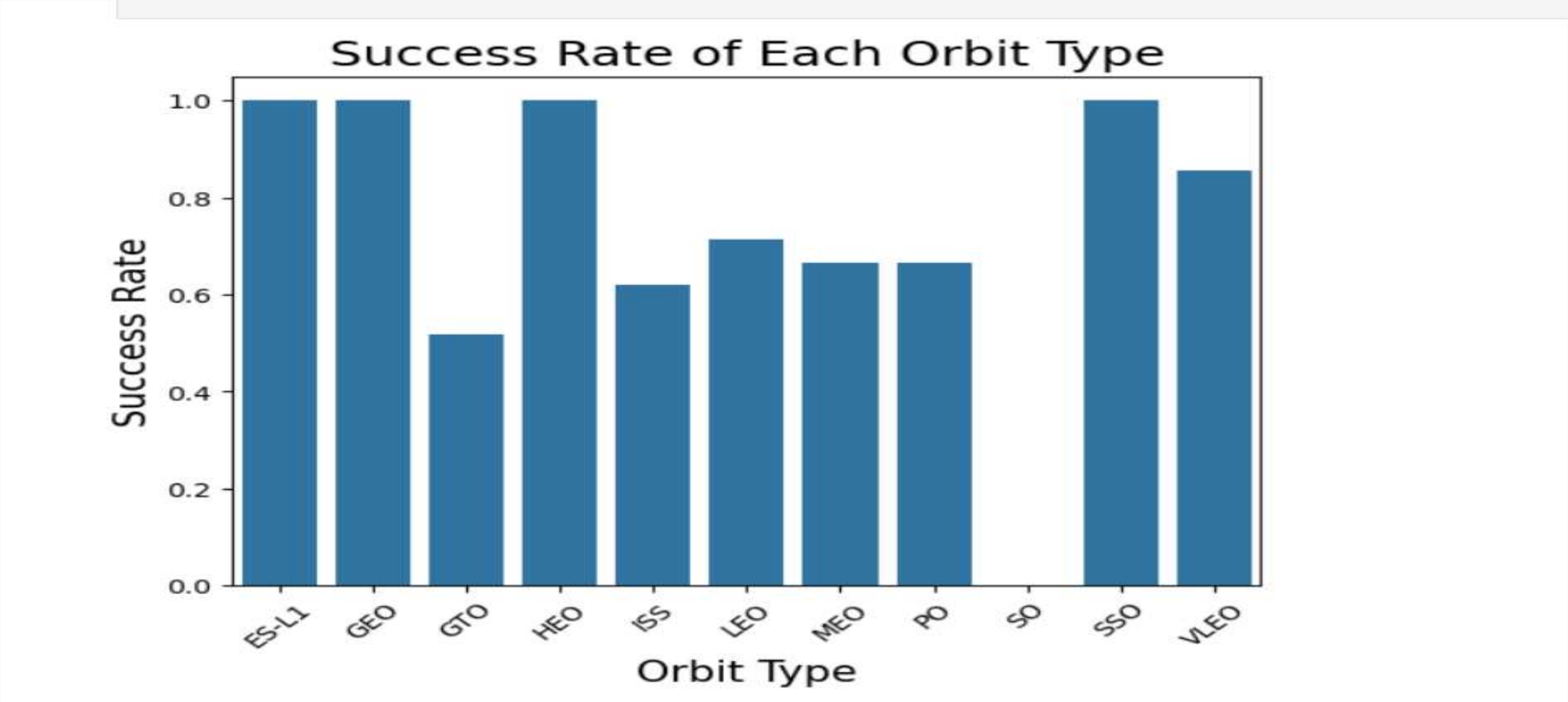
```
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect=5)
plt.xlabel("Flight Number", fontsize=20)
plt.ylabel("Launch Site", fontsize=20)
plt.show()
```



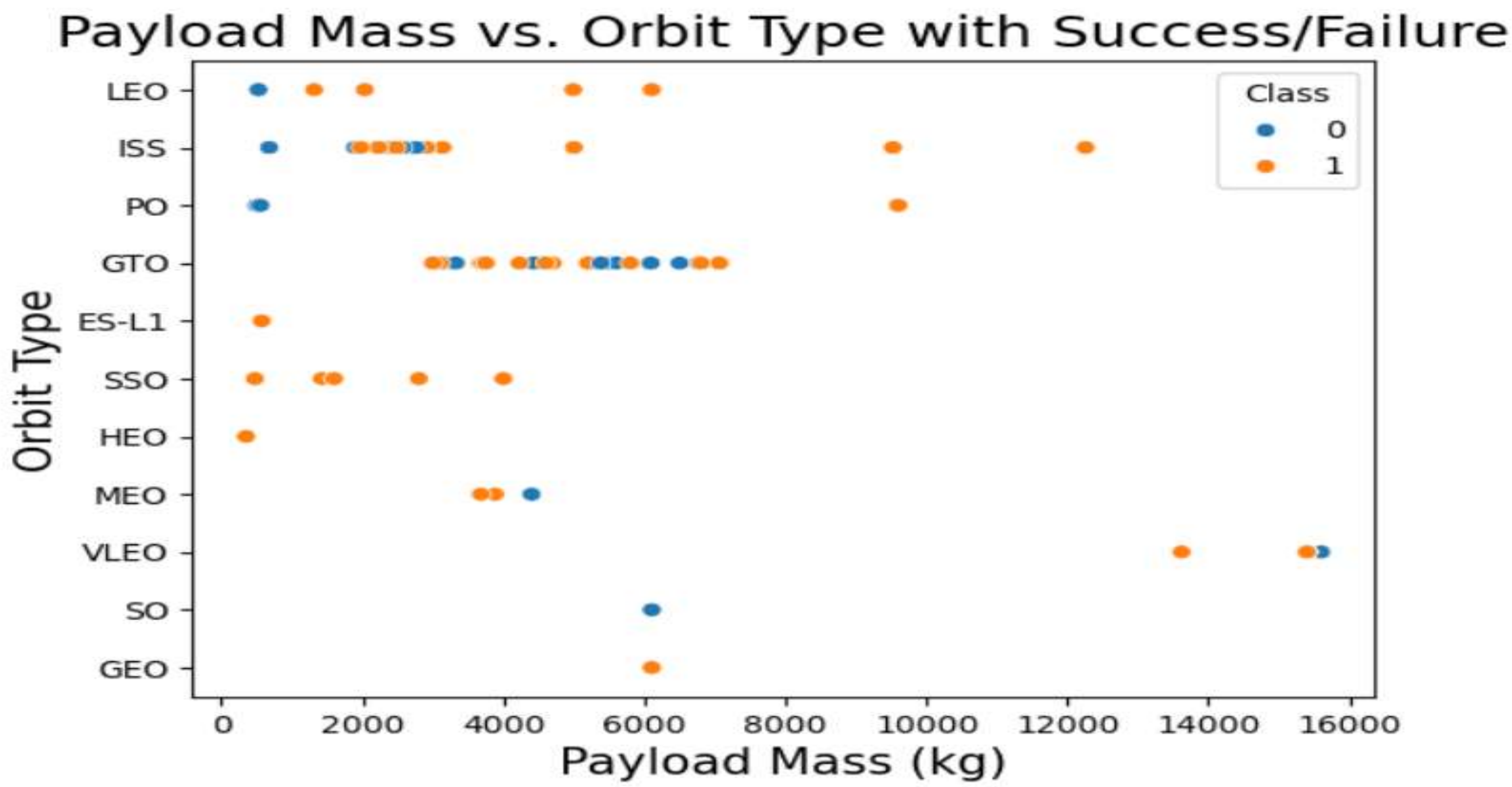
Payload vs. Launch Site



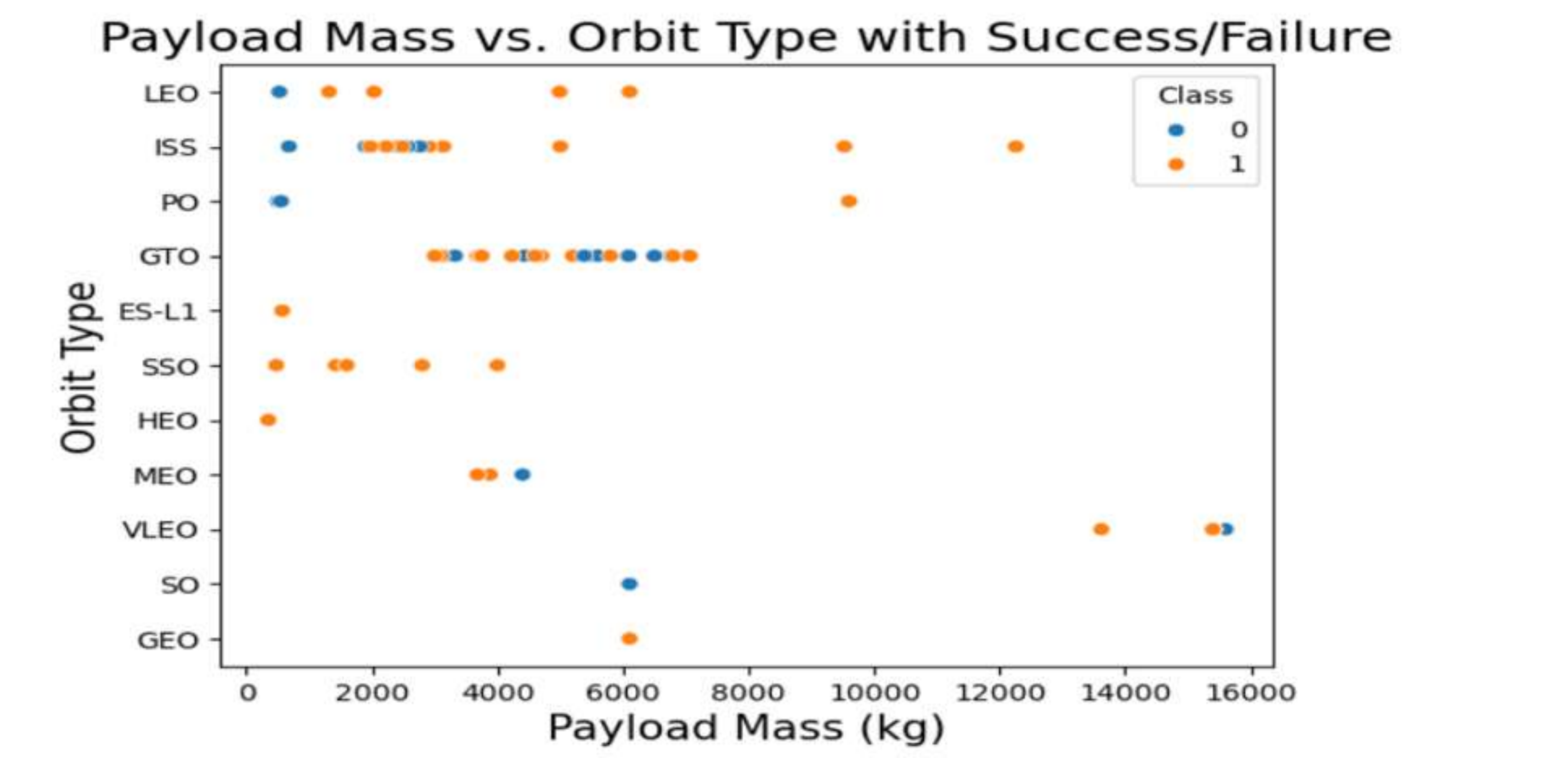
Success Rate vs. Orbit Type



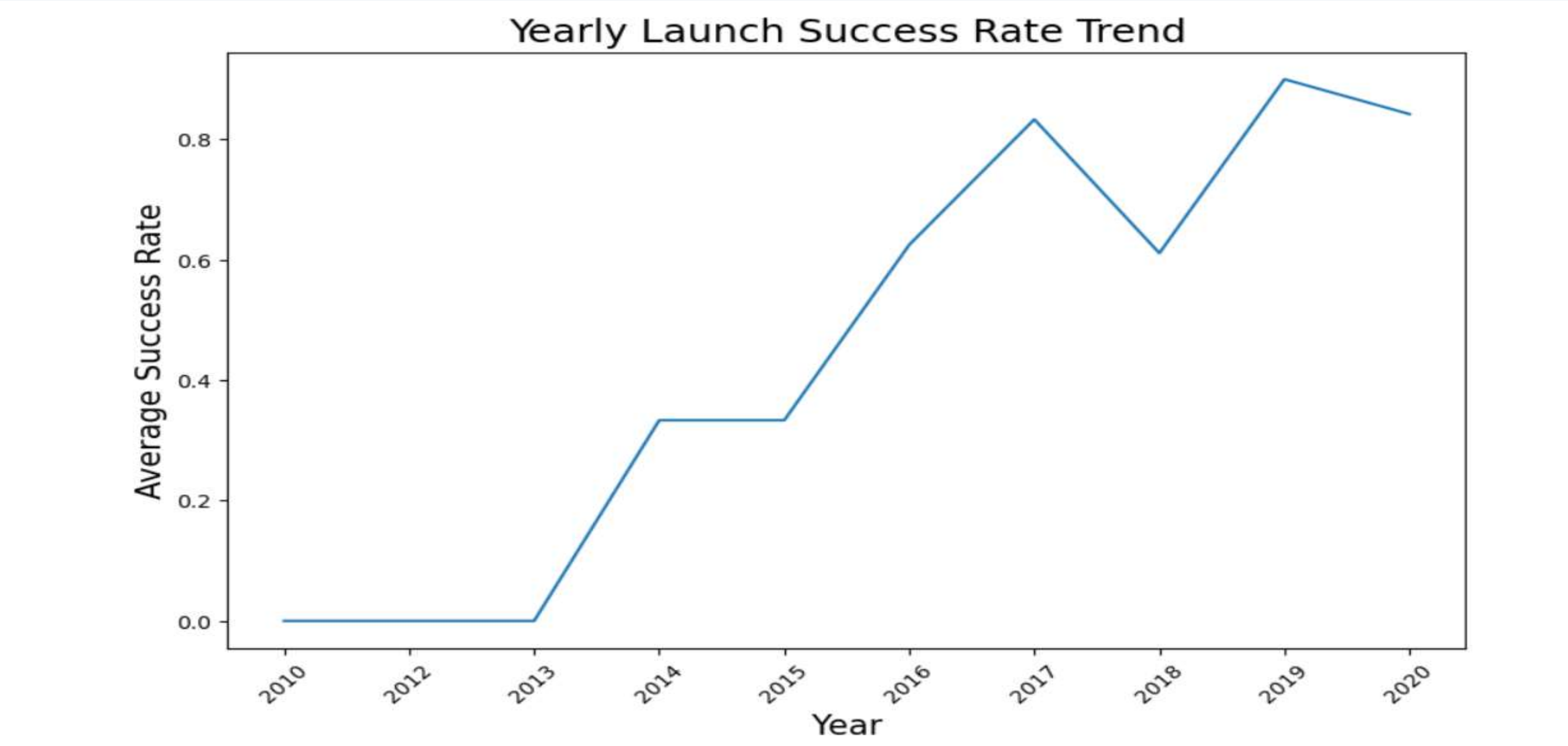
Payload vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

Launch Site Names Begin with 'CCA'

CCAFS LC-40

Cape Canaveral Space Force Station Launch Complex 40 (CCAFS LC-40)

SpaceX's key launch sites, located in Florida on the U.S. Space Force’s Cape Canaveral Space Force Station. It primarily supports launches of Falcon 9 rockets, including missions for satellite deployments, International Space Station (ISS) resupply, and commercial payloads. LC-40 has seen a high success rate, contributing to SpaceX’s reputation for reliable launches, though it has also experienced notable failures, such as the AMOS-6 explosion in 2016 during pre-launch testing. Despite occasional setbacks, LC-40 remains one of SpaceX's most frequently used and successful launch pads.

Total Payload Mass

- SpaceX’s NASA Commercial Resupply Service (CRS) missions have carried a total payload mass of 45,596 kg to the International Space Station (ISS) via Falcon 9 rockets launched from CCAFS LC-40 and other SpaceX sites

```
%%sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass
FROM "SPACEXTABLE"
WHERE "Customer" = 'NASA (CRS)';
```

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

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

The Falcon 9 v1.1 booster version, an early variant of SpaceX's Falcon 9 rocket, achieved an average payload mass of 2,928.4 kg per launch. This version was an upgraded model introduced in 2013, designed to carry larger payloads and provide improved performance over the original Falcon 9. It was primarily used for launching satellites and cargo to the International Space Station, as well as commercial payloads. Falcon 9 v1.1 featured extended fuel tanks, enhanced engines, and a more efficient structure, contributing to its reliable performance during numerous missions. This booster was pivotal in SpaceX’s early development of reusable r

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

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

The first successful ground pad landing by SpaceX occurred on **2015-12-22**
This landmark event took place after the launch of the Falcon 9 rocket, marking a major milestone in SpaceX's goal of achieving reusable rocket technology. This landing demonstrated that boosters could be returned to Earth safely, paving the way for future cost-effective launches. The success of this landing reinforced SpaceX’s progress toward reducing the costs of space travel and opened up new possibilities for sustainable space exploration and satellite deployment.

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The following Falcon 9 boosters successfully landed on a drone ship and carried payloads between 4,000 kg and 6,000 kg: F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2.

These boosters were part of SpaceX's upgraded Falcon 9 Full Thrust (FT) version, designed for higher performance and reusability. Drone ship landings are essential for missions that require higher velocity, as they allow for safe recovery of the booster at sea. These specific missions demonstrated SpaceX's ability to recover and reuse boosters, even on high-energy missions with substantial payloads.

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful mission outcomes for SpaceX missions is approximately 98%, showcasing a high success rate in achieving mission objectives. SpaceX has managed to maintain this strong track record through rigorous testing, continuous improvements, and advancements in rocket technology. The remaining 2% accounts for mission failures, which include technical anomalies and unforeseen challenges. Despite these few setbacks, SpaceX’s focus on reusability and reliability has driven a remarkable success rate, establishing the company as a leader in commercial spaceflight and reusable launch technology.

Boosters Carried Maximum Payload

The following Falcon 9 Block 5 boosters carried the maximum payload mass on their missions

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

These Block 5 boosters represent the most advanced and powerful iteration of the Falcon 9, designed for higher payload capacity and reusability. Their ability to carry heavier payloads efficiently is crucial for missions with demanding requirements, underscoring SpaceX’s commitment to high-capacity, reusable launch solutions.

2015 Launch Records

The following records show the failed drone ship landings for SpaceX missions in 2015, along with their respective booster versions and launch sites

In 2015, SpaceX experienced drone ship landing failures in January and April. The missions launched from Cape Canaveral Air Force Station (CCAFS) Launch Complex 40 using Falcon 9 v1.1 boosters (B1012 and B1015). These early landing attempts were part of SpaceX's development efforts in reusable rocket technology, providing valuable data to improve future landings and contributing to SpaceX’s success in later missions.

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

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

The Landing outcomes for SpaceX missions between 2010-06-04 and 2017-03-20

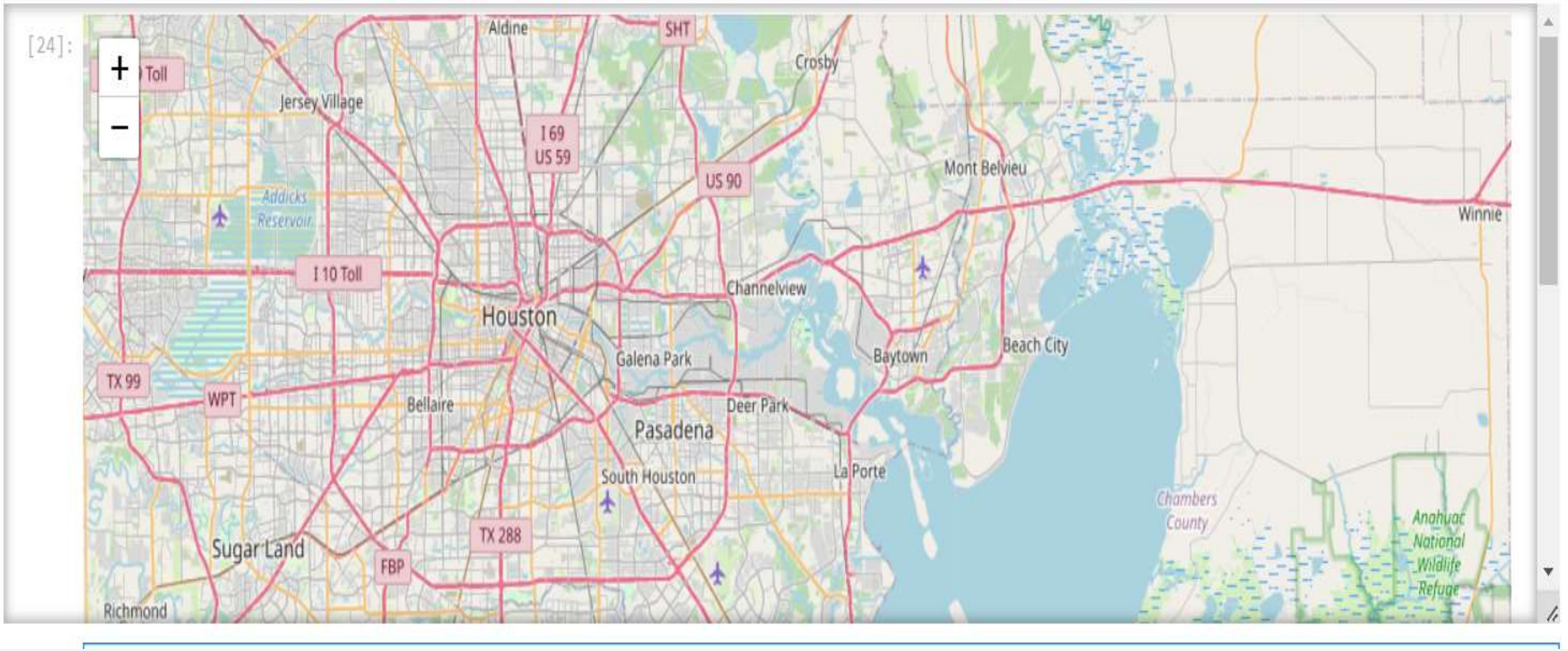
During this period, "No attempt" landings occurred most frequently, likely reflecting early missions before reusability became a primary focus. Drone ship successes and failures both appear next, highlighting the iterative progress SpaceX made in mastering drone ship landings. Other outcomes like ground pad successes and controlled ocean landings also illustrate SpaceX’s varied landing approaches as they refined their reusability strategies.

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

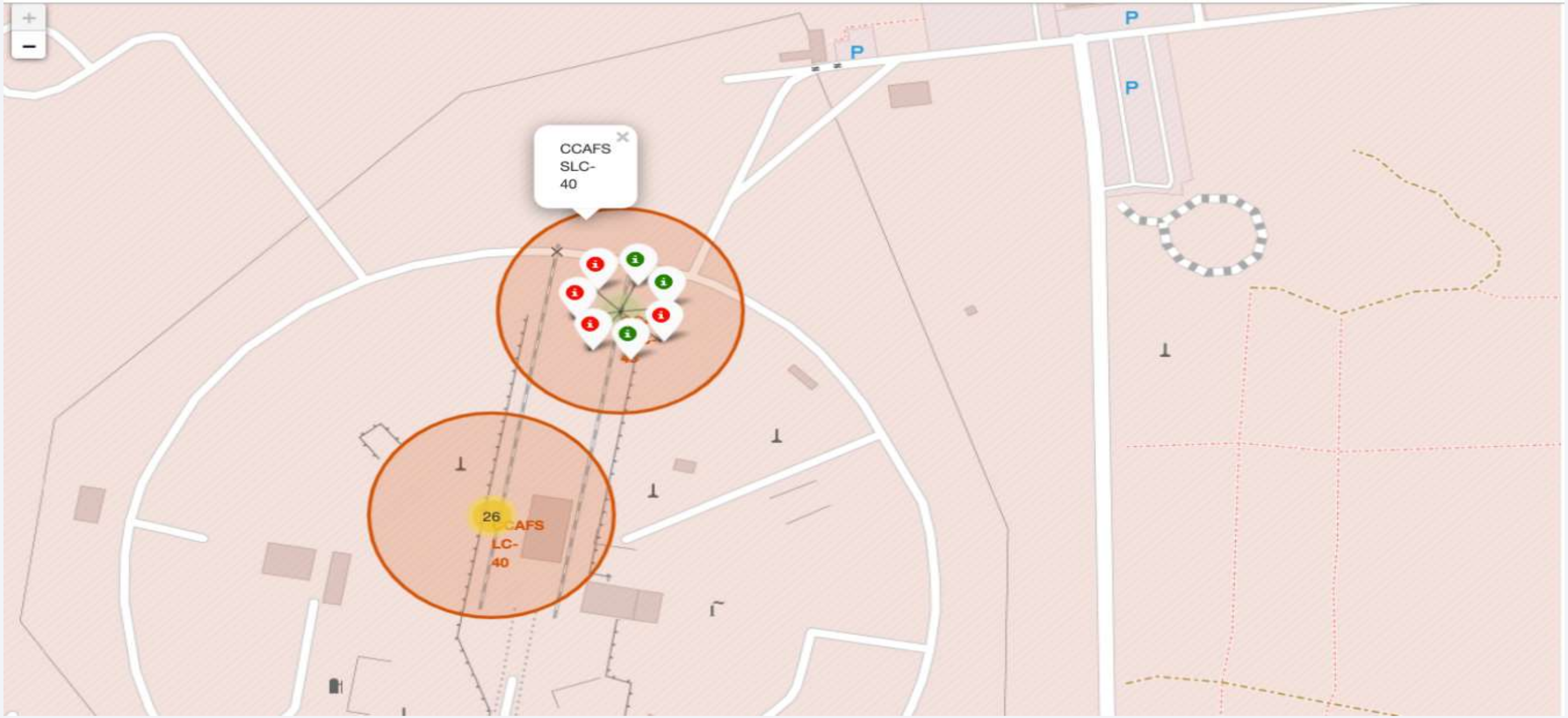
Launch Sites Proximities Analysis

Folium Map Screenshot 1



Folium Map Screenshot 2

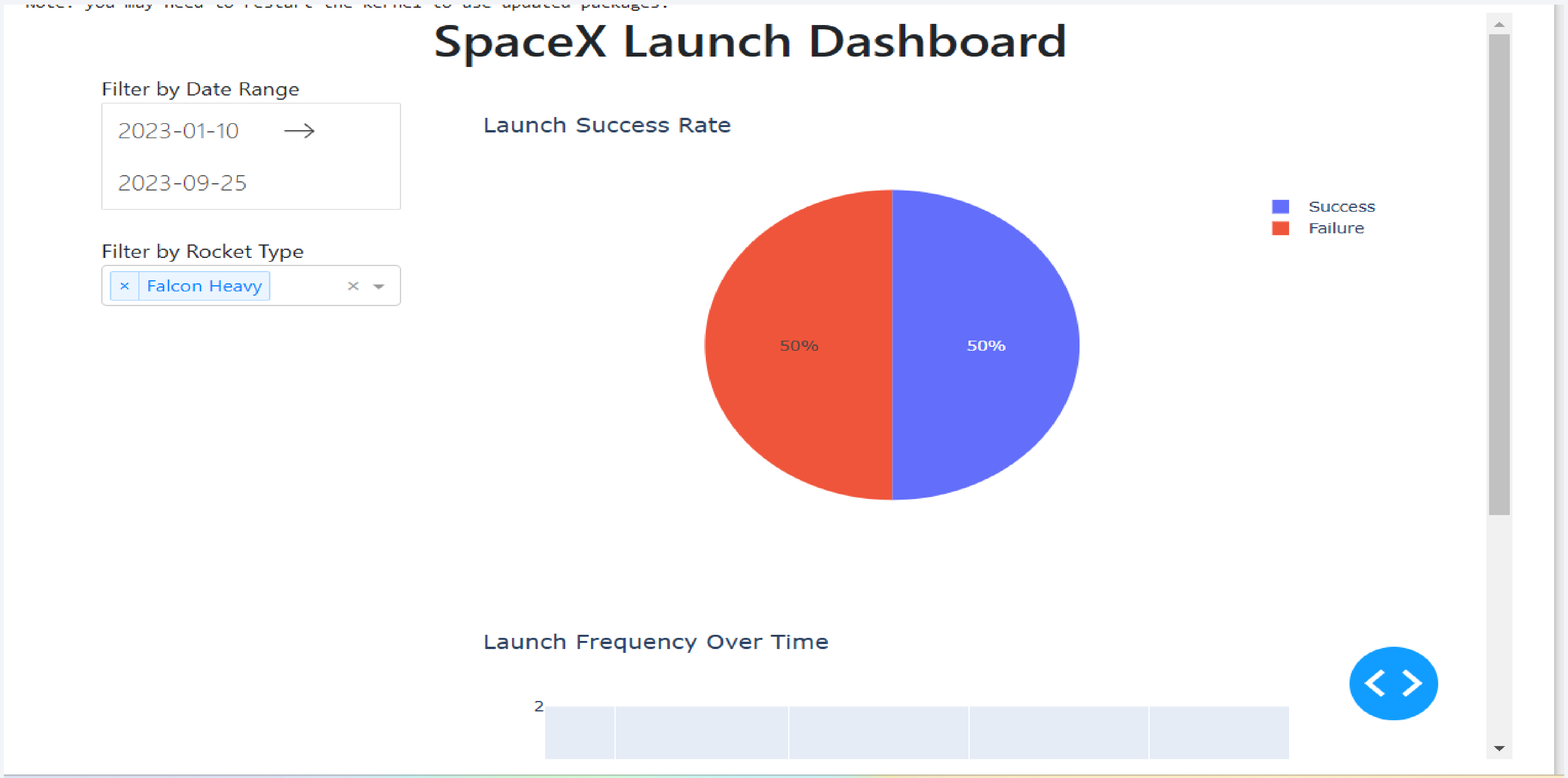
Folium Map Screenshot 3



Section 4

Build a Dashboard with Plotly Dash

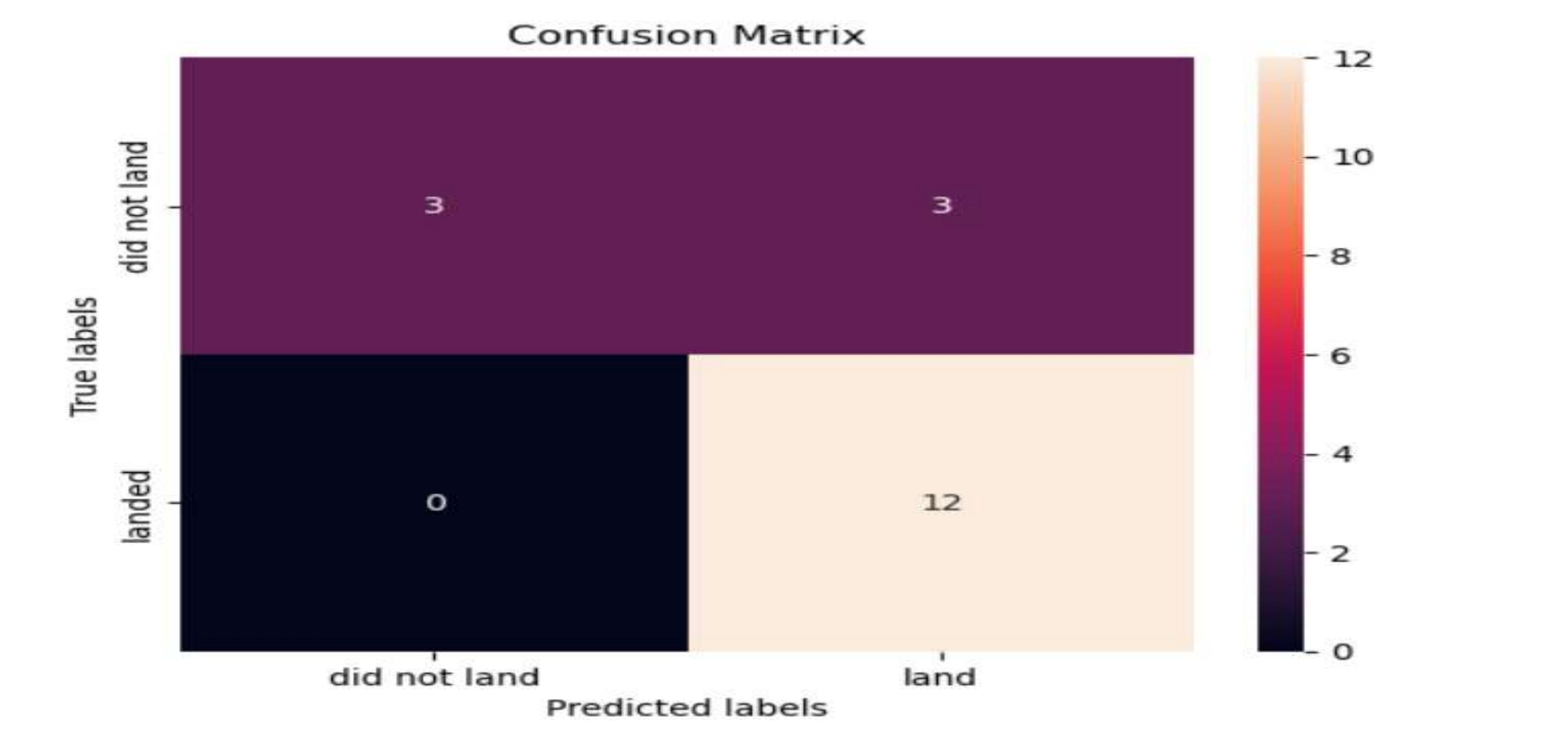
Dashboard Screenshot 1



Section 5

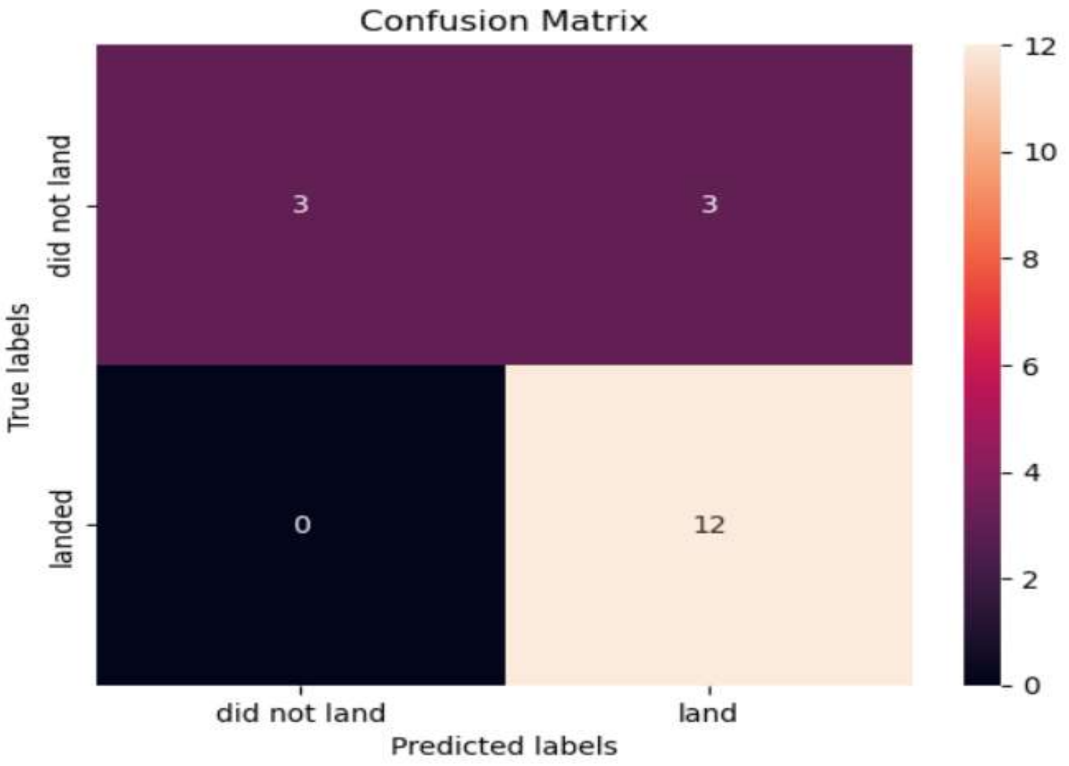
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

[58]:
```
yhat=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- Best-performing model: **Logistic Regression Accuracy**
- Key Insights: Rocket type and booster landing significantly influence the success of a launch.
- Future Work: Explore additional features, refine models, and consider more advanced techniques like deep learning for prediction

Appendix

- Gut hub link below
- https://github.com/Garykumanan/SpaceX_fallcon/tree/main

Thank you!