Winning Space Race with Data Science

PRANAV LENIN
31/3/2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

This project is about SpaceX rocket landing prediction. This will be predicting the answer of : Will this SpaceX rocket land successfully or not?

Questions to be answered :

I. What Data is needed for this prediction?

II. Where will the data will be collected from?

III. What features are required for the prediction?

IV. Which algorithms are best suited for this problem?

V. What evaluation metric should I use?

VI. How will I handle overfitting or underfitting?

VII. How often should the model be updated?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected form "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv"
- Perform data wrangling
  - The data was encoded and refined by dropping the unwanted columns.
- 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

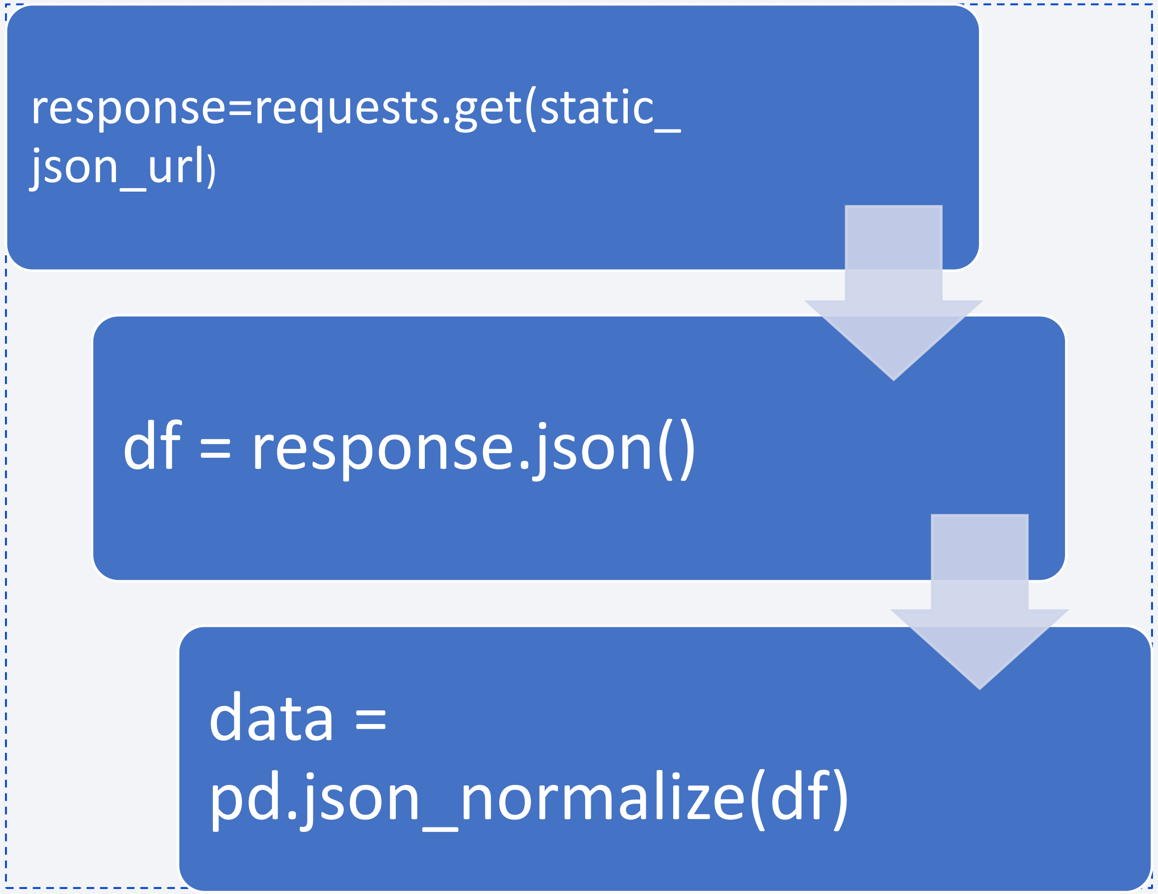
The data was collected from the **IBM** database about the rocket launches of **SPACEX.** It is a **JSON file format** which is then converted into data frame by **pandas**.

```
response=requests.get(static_json_url)
df = response.json()
data = pd.json_normalize(df)
```

Data Collection – SpaceX API

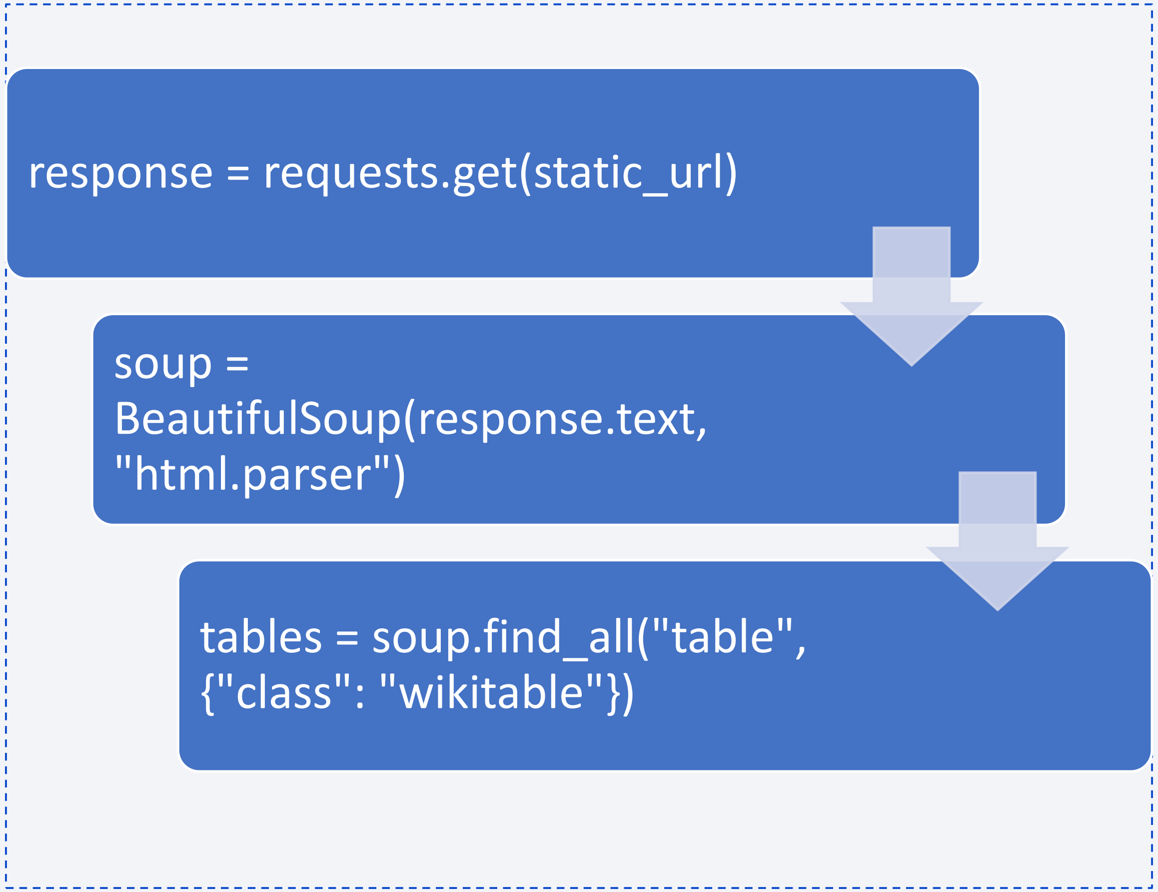
- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- I Web Scraped from :- “https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922”



Data Wrangling

The SpaceX rocket launch data was collected and cleaned by handling missing values, removing duplicates, and fixing data types. Categorical variables like LaunchSite, Orbit, and LandingPad were encoded using pd.get_dummies(), while numerical features like PayloadMass(kg) were scaled. The dataset was split into training and testing sets, ensuring no data leakage. Finally, the processed data was used to train machine learning models to predict the success of rocket landings. Git hub link:'
https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-d8f550da53a199b98550928754ae733955db18ec'

EDA with Data Visualization

•**Bar Chart** – Used to compare the success rate of launches across different launch sites.
•**Scatter Plot** – Plotted **FlightNumber vs. PayloadMass**(kg) to analyze the impact of payload on landing success.
•**Catplot** – Displayed **FlightNumber vs. LaunchSite** with hue as Class to visualize the distribution of successful and failed landings.
•**Pie Chart** – Showed the proportion of successful vs. failed launches for an overall success rate analysis.
•**Box Plot** – Used to detect outliers and analyze the distribution of **PayloadMass(kg)** across different orbits.
•**GITHUB LINK**:" https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-3950bf2f61f89c02c297b52a2ef57061beb5053d"

Build an Interactive Map with Folium

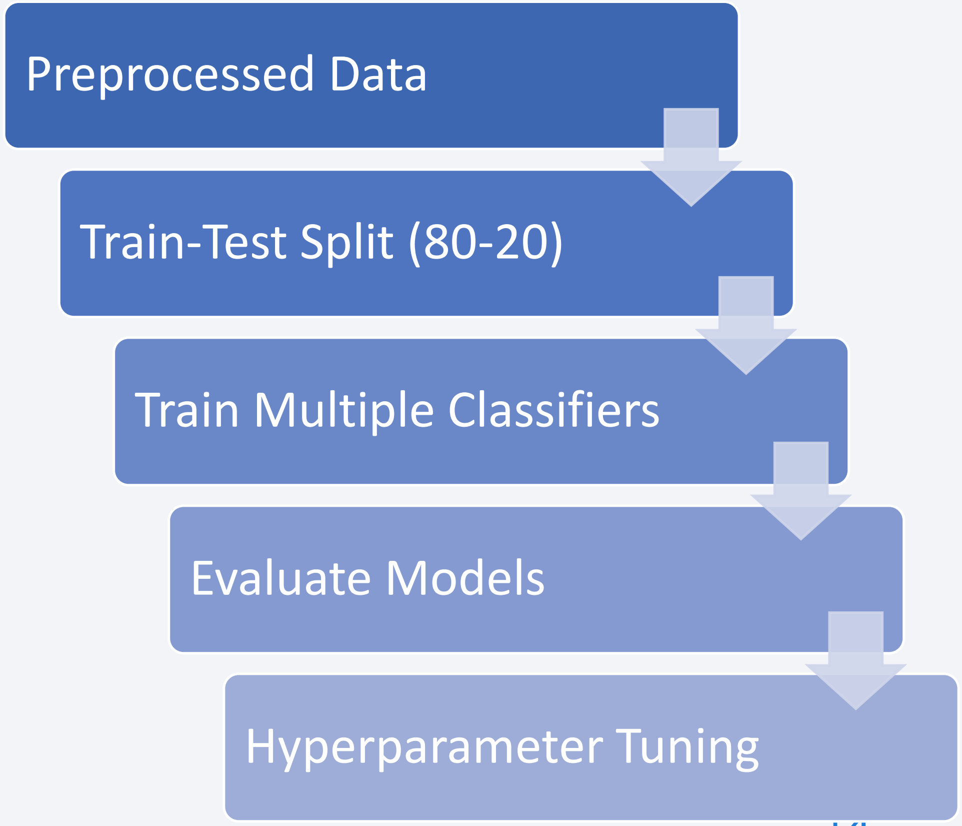
- **Folium Map Objects and Their Purpose**
- **Markers** – Placed at each SpaceX launch site to pinpoint their exact locations.
- **Circles** – Added around launch sites to highlight their geographical areas.
- **Lines (Polylines)** – Drawn to show distances between launch sites and key locations (e.g., headquarters or landing zones).
- **Popups** – Included additional details about each launch site when clicked.
- These objects were added to provide an **interactive visualization** of SpaceX launch sites, helping to analyze their locations and surroundings effectively.
- GITHUB LINK-" https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulidd8f550da53a199b98550928754ae733955db18ec"

Build a Dashboard with Plotly Dash

- **Bar Chart** – Displays the success count of launches for each launch site to compare performance.
- **Scatter Plot** – Shows the relationship between **PayloadMass(kg) and success rate,** helping to analyze the impact of payload size.
- **Pie Chart** – Visualizes the proportion of successful vs. failed landings for overall mission success analysis.
- **Line Chart** – Represents trends in launch success over time.

Predictive Analysis (Classification)

•**Data Preprocessing** – Cleaned missing values, encoded categorical variables (Orbit, LaunchSite), and scaled numerical features (**PayloadMass(kg)**).
•**Model Selection** – Tested multiple classification models, including **Logistic Regression, Support Vector Machine (SVM), Decision Tree, and Random Forest**.
•**Evaluation** – Used accuracy, precision, recall, and F1-score to compare model performance.
•**Hyperparameter Tuning** – Applied **GridSearchCV** and **RandomizedSearchCV** to optimize parameters for better accuracy.
•**Best Model Selection** – Found that **SVM with an RBF kernel** performed best on the validation dataset



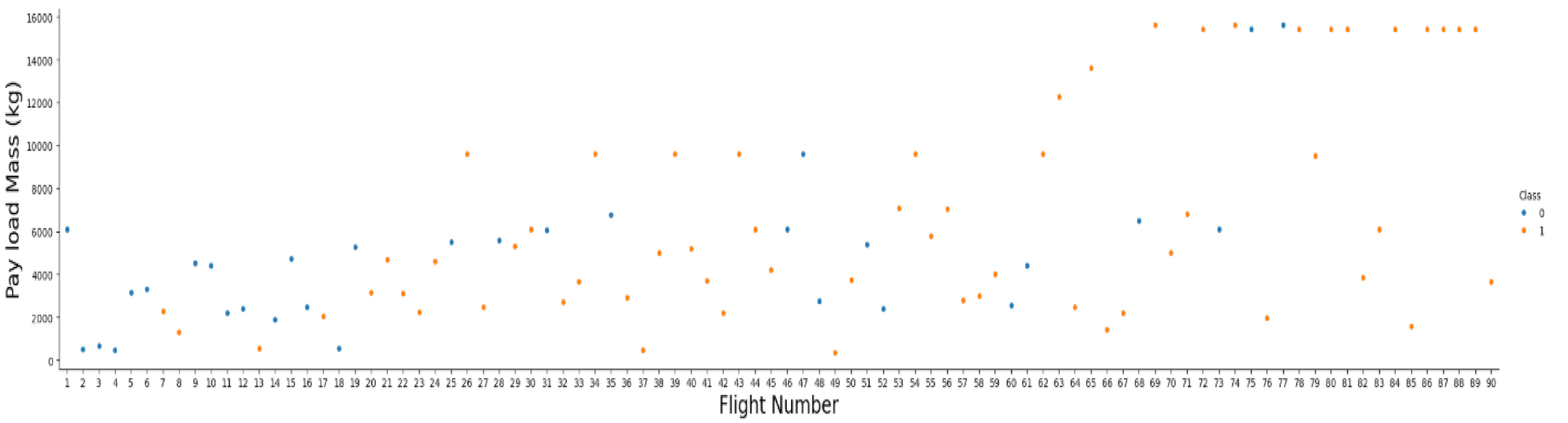
Results

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

Section 2

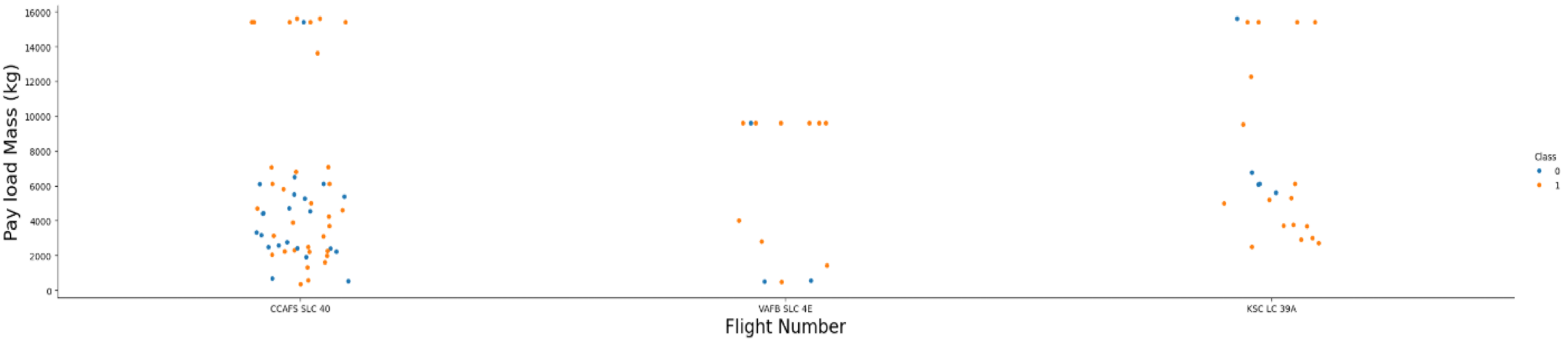
Insights drawn from EDA

Flight Number vs. Launch Site



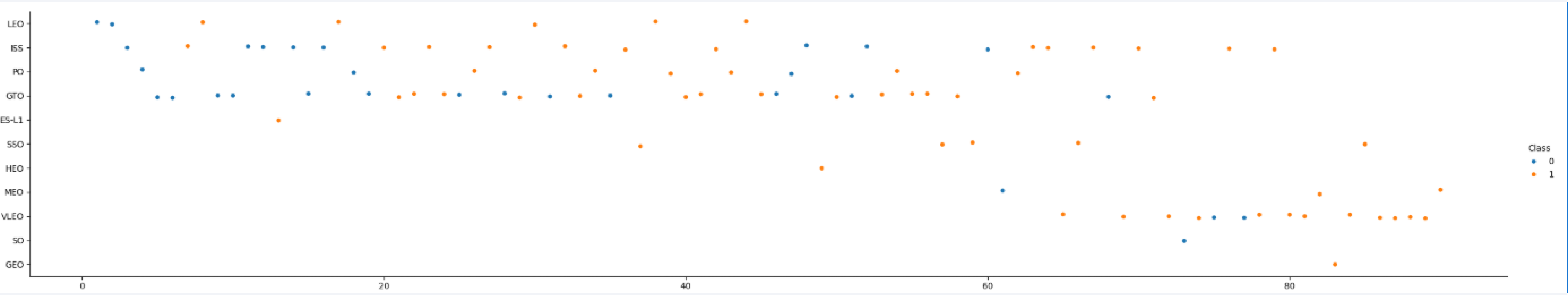
Payload vs. Launch Site

```
sns.catplot(y="PayloadMass", x="LaunchSite", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```

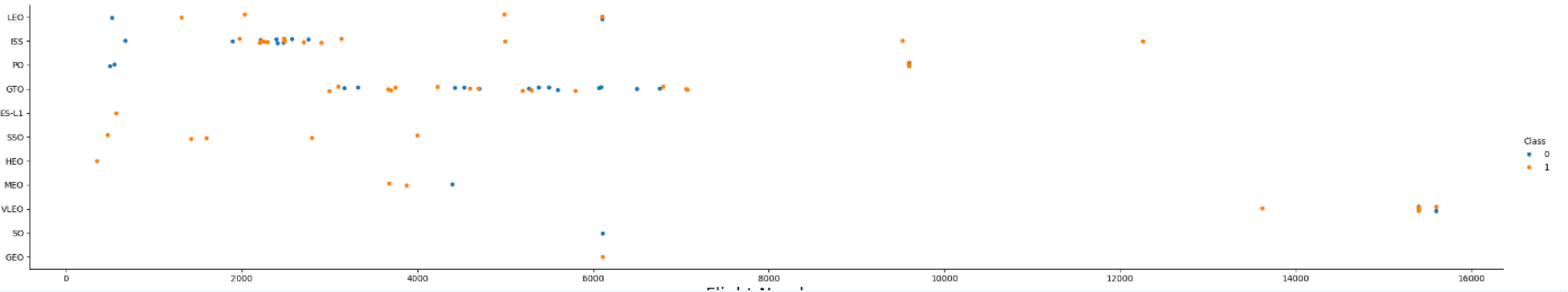


Next, let's drill down to each site visualize its detailed launch records.

Flight Number vs. Orbit Type



Payload vs. Orbit Type



All Launch Site Names

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

```
Totally, there are four launch sites named CCAFS LC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40.
```

Launch Site Names Begin with 'CCA'

```
('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')
```

Above are the total records where launch sites begin with the string "CCA".

Total Payload Mass

```
Total Payload Mass carried by NASA (CRS): 45596 kg
```

```
As showed above, the total payload mass is 45596 kg.
```

Average Payload Mass by F9 v1.1

```
Average Payload Mass carried by F9 v1.1: 2534.67 kg
```

As showed above, the average payload mass is 2534.67 kg.

First Successful Ground Landing Date

```
First successful landing on ground pad: 2015-12-22
```

```
First successful landing outcome on ground pad is 2015-12-22.
```

Successful Drone Ship Landing with Payload between 4000 and 6000

```
Boosters with successful drone ship landing and payload between 4000 and 6000 kg:
F9 FT B1022
F9 FT B1026
F9 FT  B1021.2
F9 FT  B1031.2
```

Boosters with successful drone ship landing and had payload mass greater than 4000 but less than 6000 kg as above.

Total Number of Successful and Failure Mission Outcomes

```
Mission Outcomes Count:
Failure (in flight): 1
Success: 98
Success : 1
Success (payload status unclear): 1
```

```
The total number of successful and failure mission outcomes are 98 and 1.
```

Boosters Carried Maximum Payload

```
Booster Versions that carried the maximum payload mass:
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
```

Above are the names of the booster which have carried the maximum payload mass.

2015 Launch Records

```
('January', 'F9 v1.1 B1012', 'CCAFS LC-40')
('April', 'F9 v1.1 B1015', 'CCAFS LC-40')
```

Above are the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.

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

```
('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)
```

Above are the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

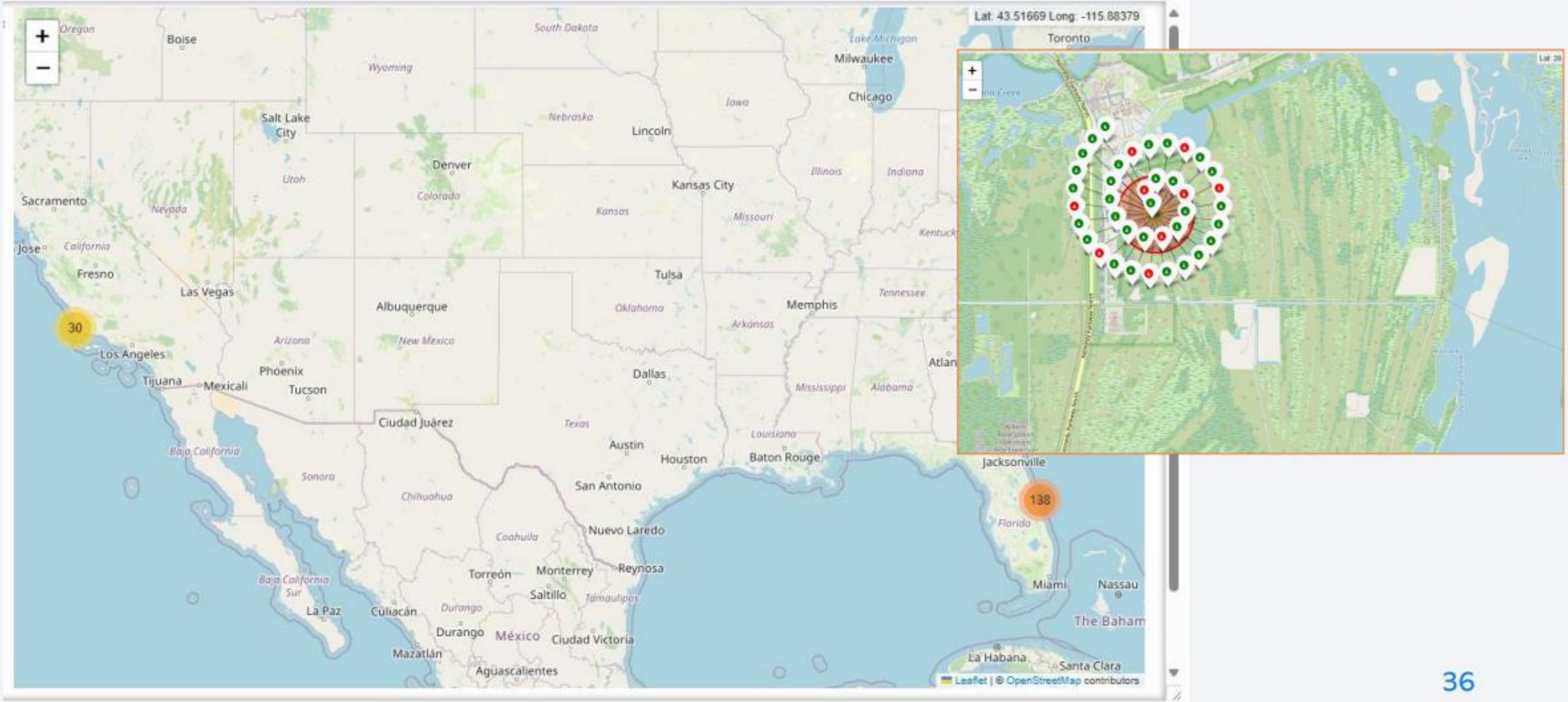
Section 3

Launch Sites Proximities Analysis

Launch Sites on Map

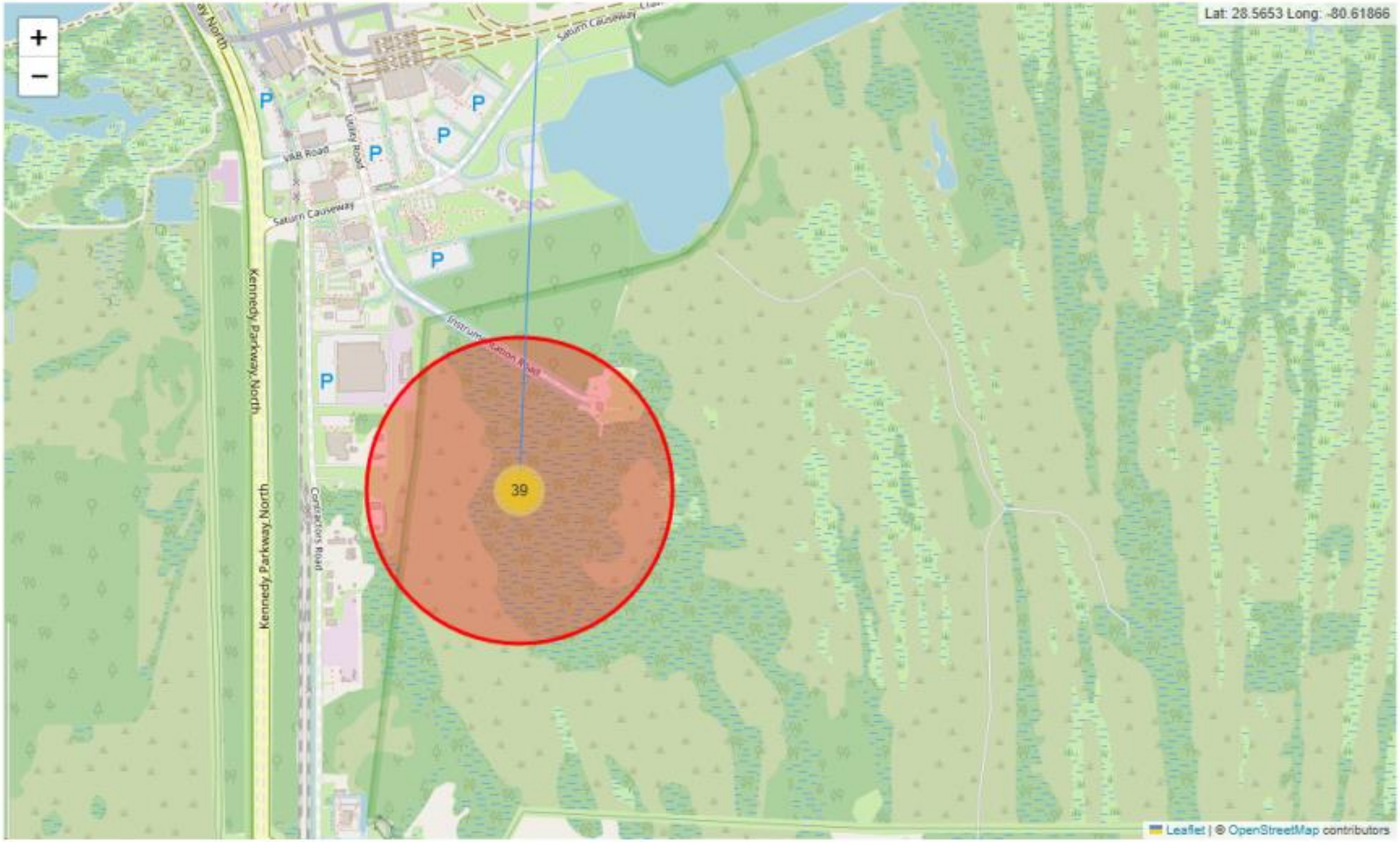
Total 4 launch sites marked as above.

Color-labeled Launch Outcomes on Map



Above is the color-labeled launch outcomes on the map.

Launch Site with Proximities

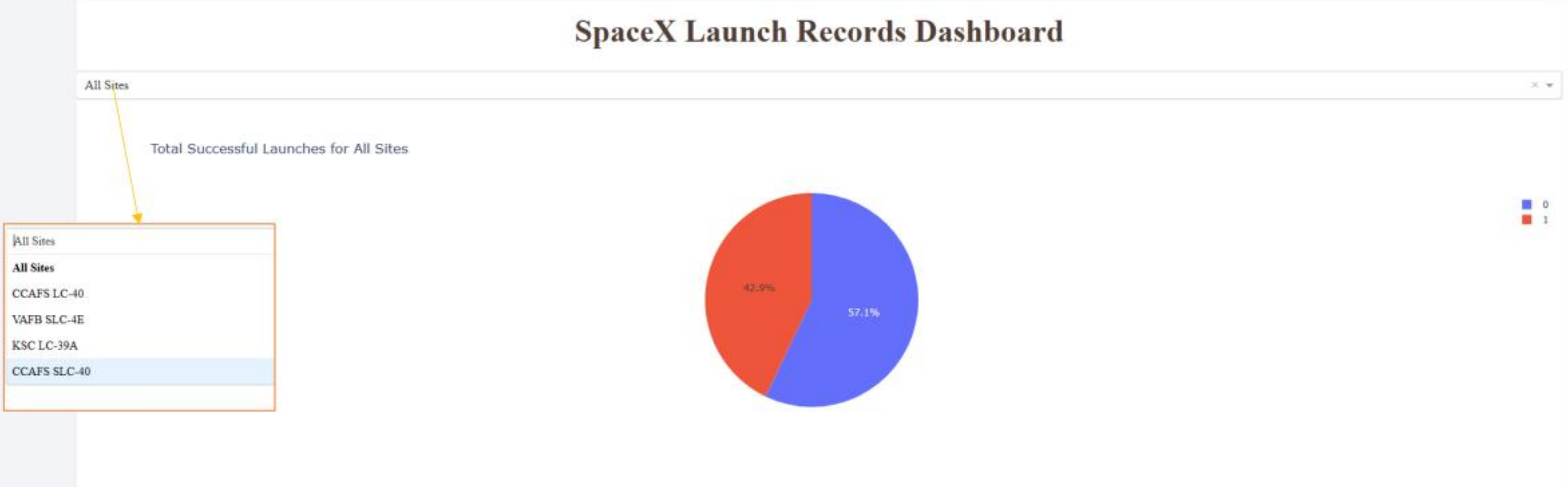


Above is a show of launch site to its proximities such as railway.

Section 4

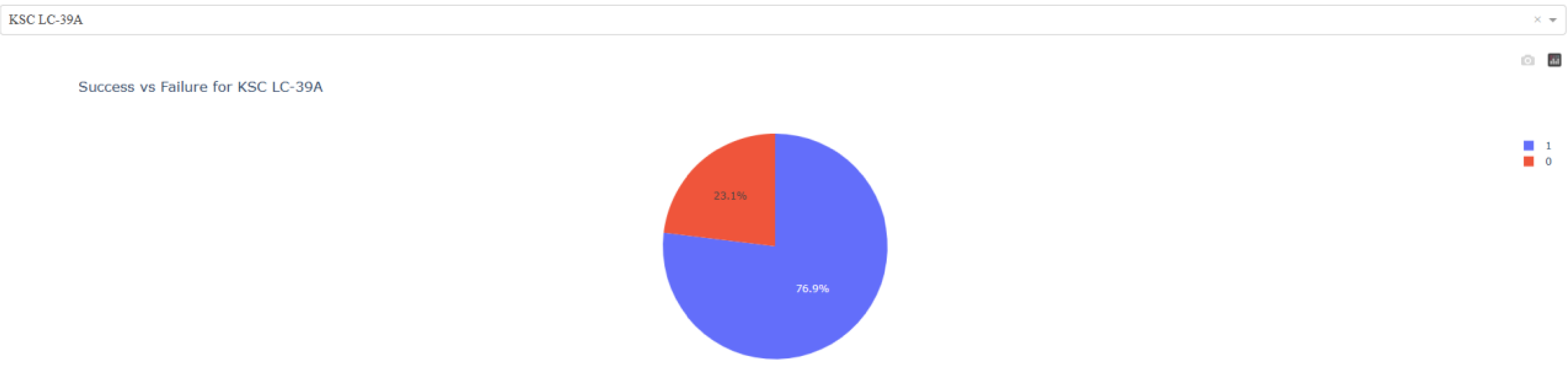
Build a Dashboard with Plotly Dash

Successful Launches of All Sites



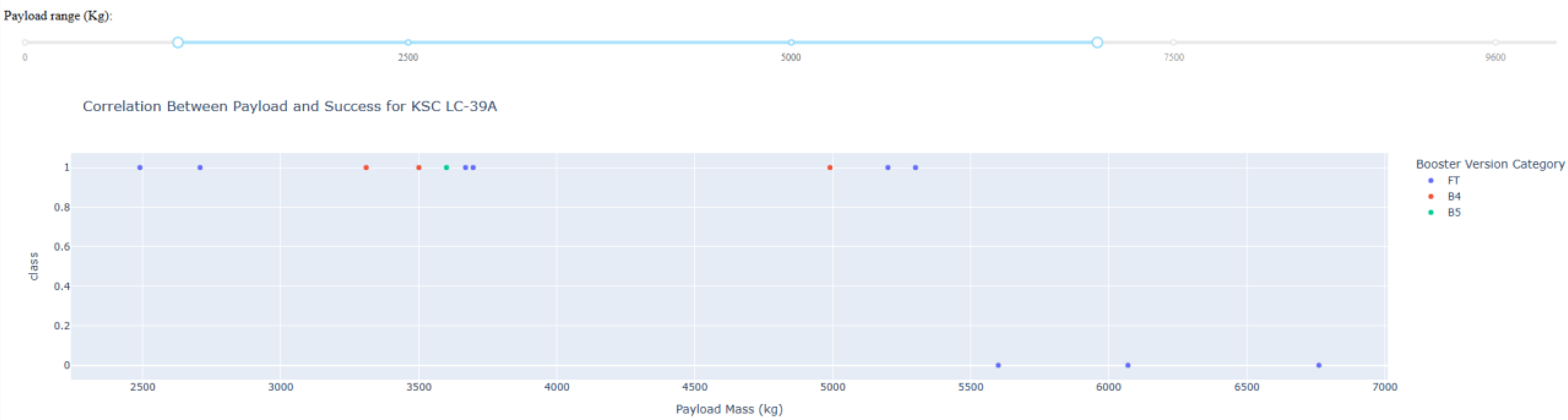
- As above showed, launch sites with success ratio by pie char.

Highest Launch Success Ratio



- As above showed, highest launch sites with success ratio by pie char.

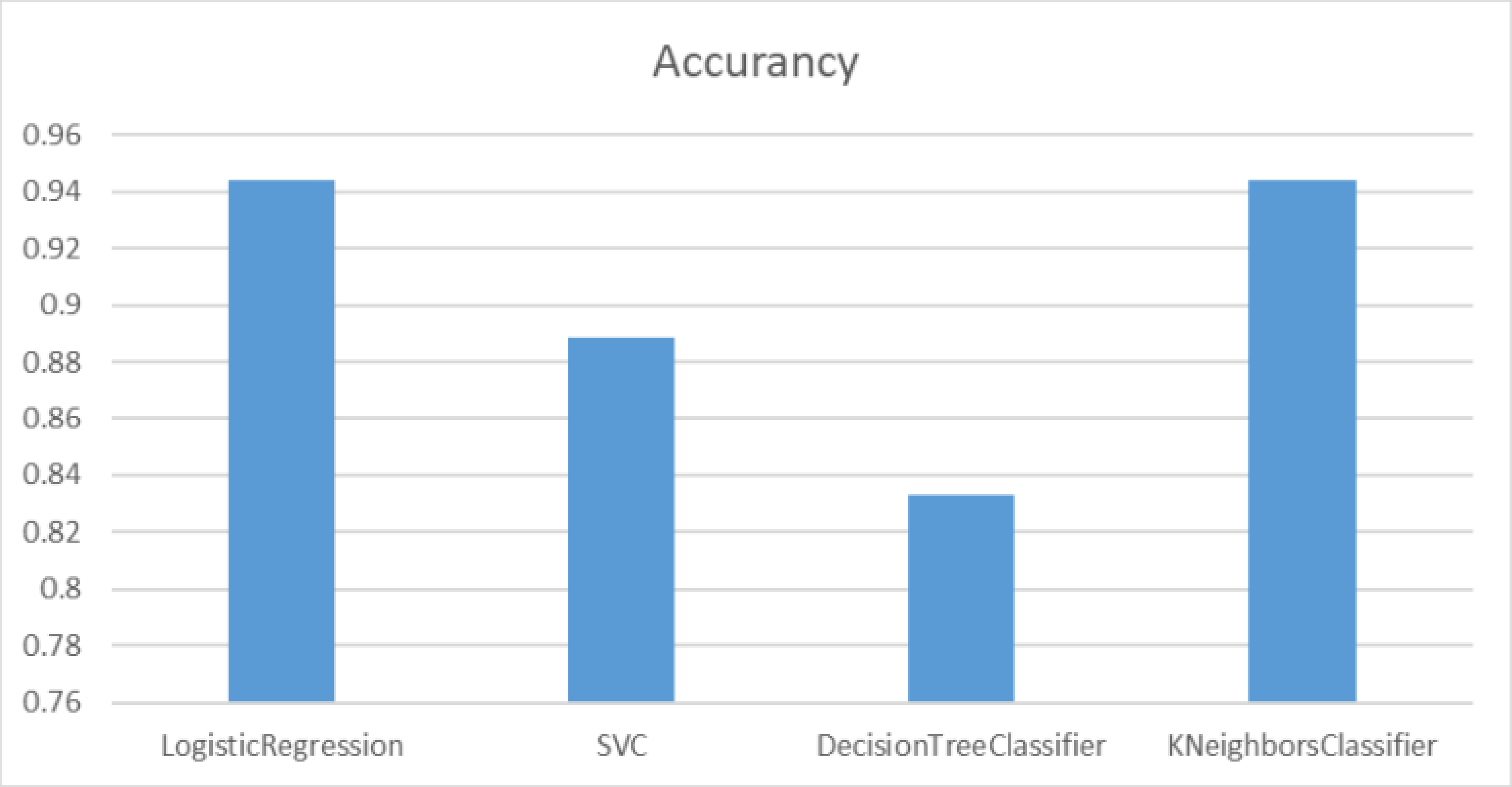
Payload vs. Launch Outcome



Section 5

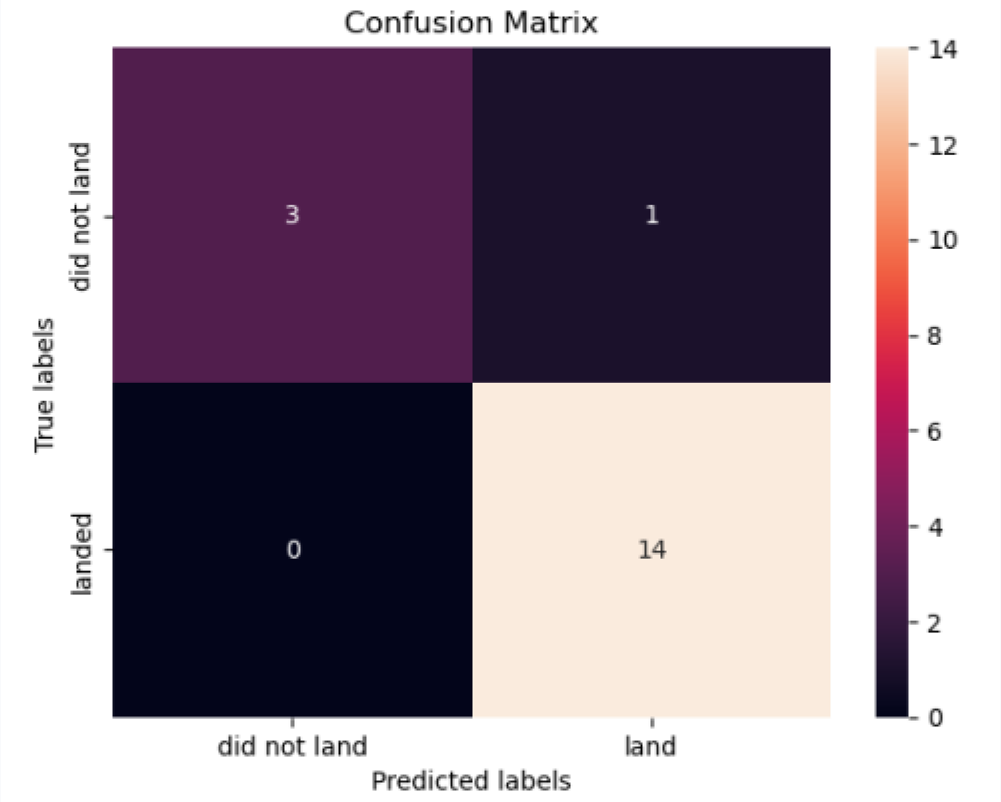
Predictive Analysis (Classification)

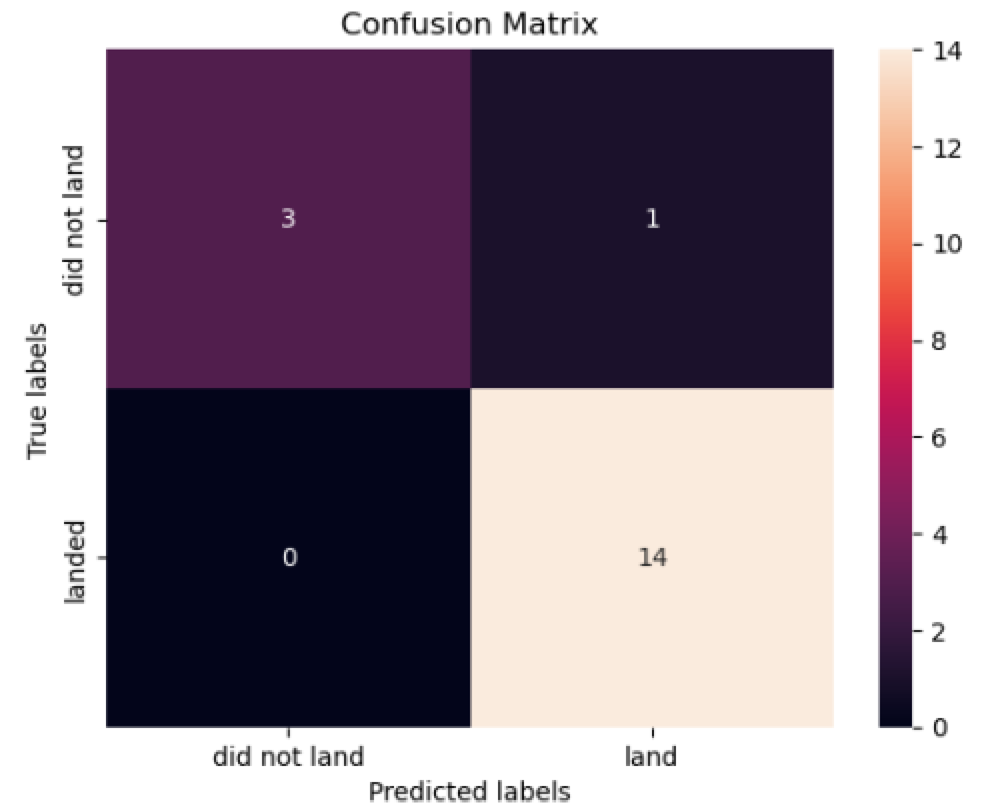
Classification Accuracy



- Logistic Regression and KNeighborsClassifier have the highest accuracy.

Confusion Matrix





Overview:

True Postive - 14 (True label is landed, Predicted label is also landed)

False Postive - 3 (True label is not landed, Predicted label is landed)

Conclusions

- Logistic Regression and KNeighborsClassifier have the highest accuracy.
- Logistic Regression and KNeighborsClassifier, predictive results:

  True Postive - 14 (True label is landed, Predicted label is also landed)

  False Postive - 3 (True label is not landed, Predicted label is landed)

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!