IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Mohammad Meysam Arjomand>
<13.01.2025>

Outline

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

Executive Summary

This project focused on analyzing and predicting the success of Space-X Falcon 9 first-stage landings using historical launch data. By leveraging the Space-X REST API and web scraping techniques, we gathered detailed information on past launches, including rocket types, launch sites, payloads, and landing outcomes.

The data was cleaned and processed, followed by exploratory data analysis (EDA) to uncover patterns such as success rates for different orbit types and the relationship between launch sites and payload mass. We used machine learning models (Logistic Regression, Decision Tree, KNN, and SVM) to predict whether the first stage would land successfully, achieving a consistent accuracy of 83.33%.

Additionally, we developed an interactive dashboard with Plotly Dash and Folium, allowing real-time exploration of the data, including interactive maps for launch site locations and trends over time.

The project provides valuable insights into Space-X's operations and helps in identifying factors influencing landing success, contributing to the improvement of future launch strategies and decision-making.

Introduction

- This project aims to predict the success of Falcon 9 first-stage landings using historical launch data. By analyzing factors such as mission conditions, rocket parameters, and launch sites, the goal is to develop a predictive model that can help Space-X optimize their operations and improve the accuracy of landing predictions. The project involves exploring various machine learning techniques to identify patterns and trends in the data that affect the success rate of landings
- Problems

Can we predict the success or failure of a landing using available data?

Which features (e.g., rocket conditions, mission specifics) have the most impact on landing success?

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

1. Space-X REST API:

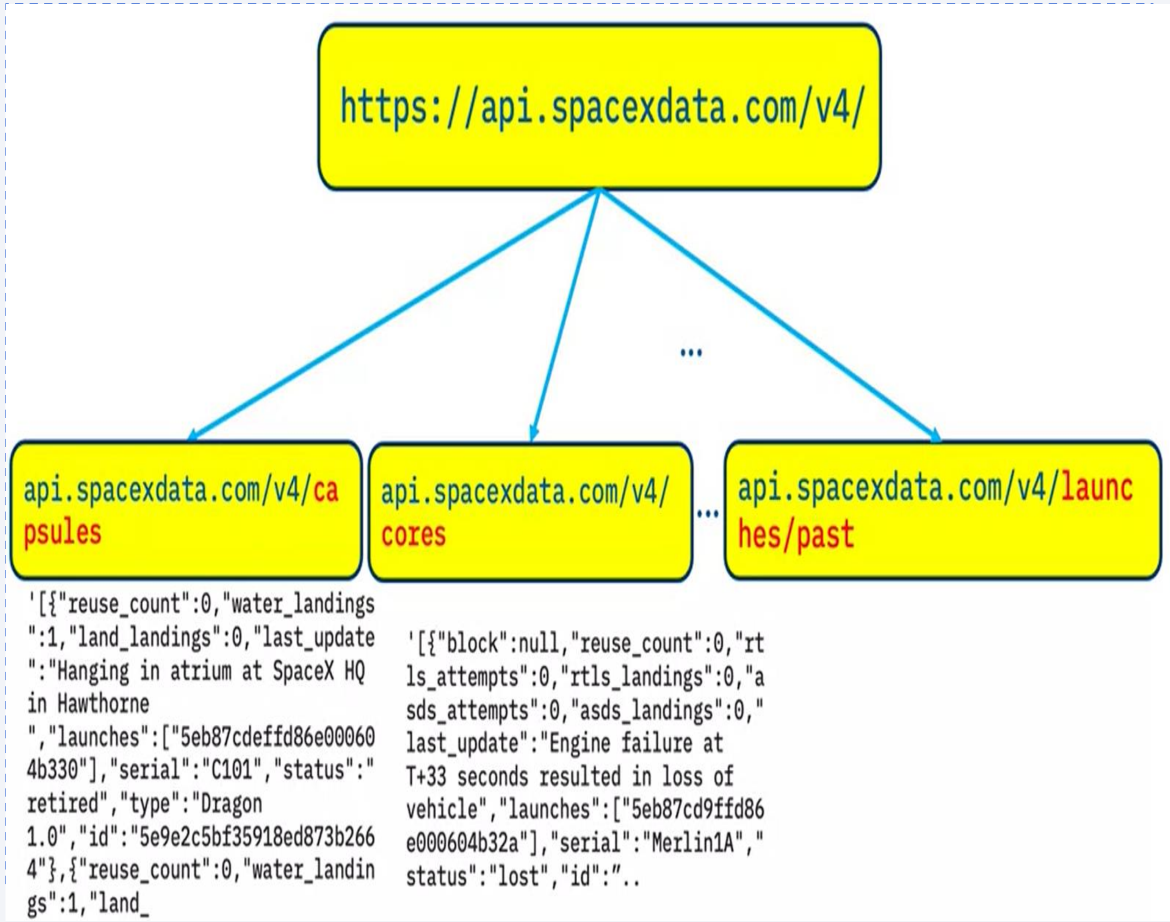
- Data on past rocket launches was retrieved from the endpoint **api.spacexdata.com/v4/launches/past**
- The JSON data includes information about rockets, launch and landing parameters, as well as results.

2. Web-Scraping

- Using **BeautifulSoup**, HTML tables of Falcon 9 launches were extracted and stored in a **Dataframe**.

Data Collection – SpaceX API

- This API will provide us with data about launches, including information about the rocket used, launch specifications, landing specifications, and the landing outcome.
- There are various endpoint APIs; we will work with **api.spacexdata.com/v4/launches/past**.
- [LINK]



Data Collection – Scraping Process

- Extracted Falcon 9 launch records from Wikipedia using BeautifulSoup.
- Parsed the HTML table to retrieve rows and columns.
- Converted the extracted data into Validated ana Pandas DataFrame.
- Validated and cleaned the data for further analysis.
- [LINK]

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Wrangling

- Performed exploratory data analysis to identify patterns and determine the labels for training supervised models.

- Defined training labels and converted them into binary values (1 and 0).

- Handling of Null Values:

  Null values were identified and handled to ensure data quality.

- LINK

EDA with Data Visualization

- Perform Exploratory Data Analysis to uncover patterns and relationships in the data.
- Prepare data through feature engineering to improve model performance and
- analysis.

- LINK

EDA with SQL

- Understand the Space-X dataset and its structure.
- Load the dataset into the corresponding table in a Db2 database.
- Execute SQL queries to answer key assignment questions and extract insights.

- LINK

Build an Interactive Map with Folium

1. Markers:

- Added to pinpoint the exact locations of launch sites.
- Help users visually identify key sites on the map.

2. Circles:

    - Used to represent proximities or safety zones around launch sites.
    - Highlighted the area of influence and potential impact zones.

3. Lines:

    - Added to show connections between launch sites and other key locations (e.g., landing zones).
    - Helped visualize relationships or distances between points.

Reason for Adding These Objects:
These objects enhance the map's interactivity and provide a clear, visual representation of spatial relationships and patterns, aiding in the exploration of optimal launch site locations.

- LINK

Build a Dashboard with Plotly Dash

1. Bar Charts:
   - Provide a clear comparison of performance across different orbits.

2. Scatter Plots:
   - Visualize the relationship between launch sites and payload mass.

3. Line Charts:
   - Show the yearly trend of launch success rates.

Reason for Adding These Elements:
These visualizations provide insights into critical aspects of the data, such as the effectiveness of different orbit types, site-specific payload trends, and overall success trends over time. They enable users to explore patterns and draw meaningful conclusions dynamically.

Predictive Analysis (Classification)

1.Data Preparation:

- Split the dataset into training and testing sets to evaluate model performance.

2.Model Building:

- Built and trained four different classification models:

Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, K-Nearest Neighbors (K-NN)

3.Model Evaluation:

- Used a Confusion Matrix to evaluate the accuracy and performance of each model.
- LINK

Results

1.Exploratory data analysis results :

- Bar chart: Success rate by orbit type.
- Scatter plot: Launch site vs. payload mass.
- Line chart: Yearly launch success trends.

2.Interactive analysis results :

- Dashboard: Interactive maps and real-time data exploration with Folium and Plotly Dash.

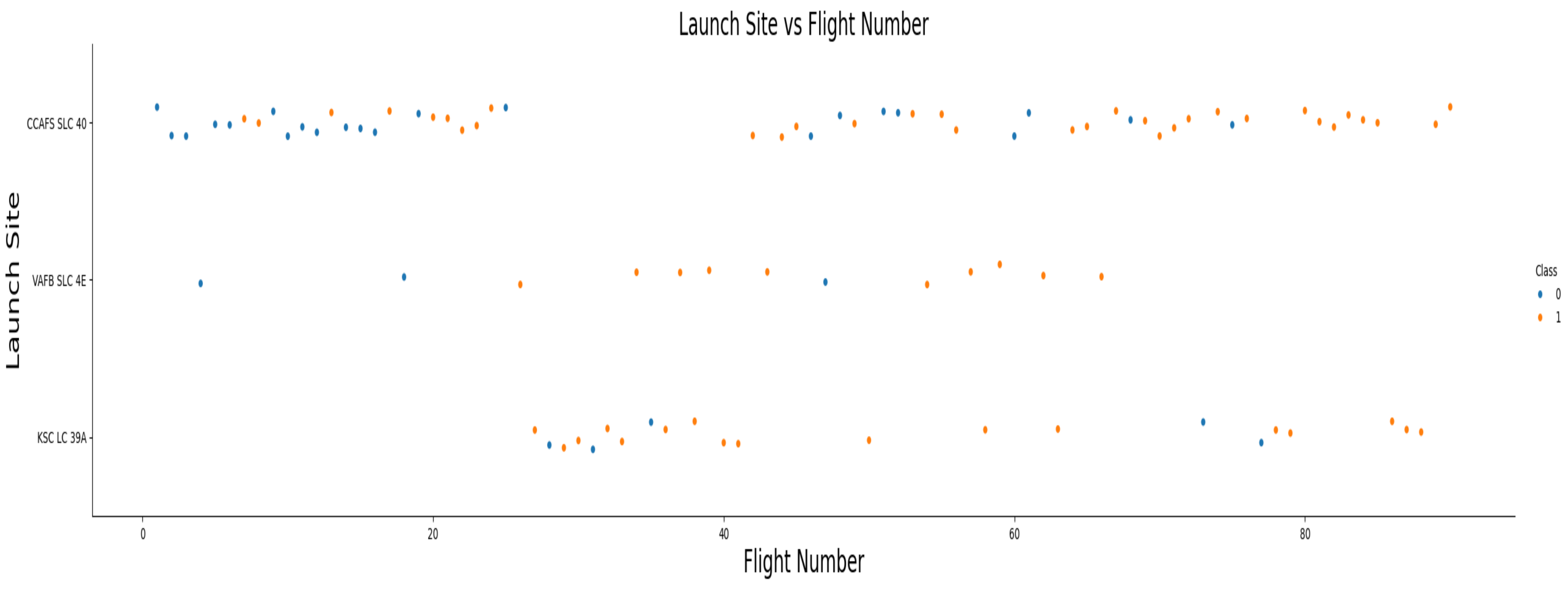
3.Predictive analysis results :

- All models (Logistic Regression, Decision Tree, KNN, SVM) achieved an accuracy of **83.33%**.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

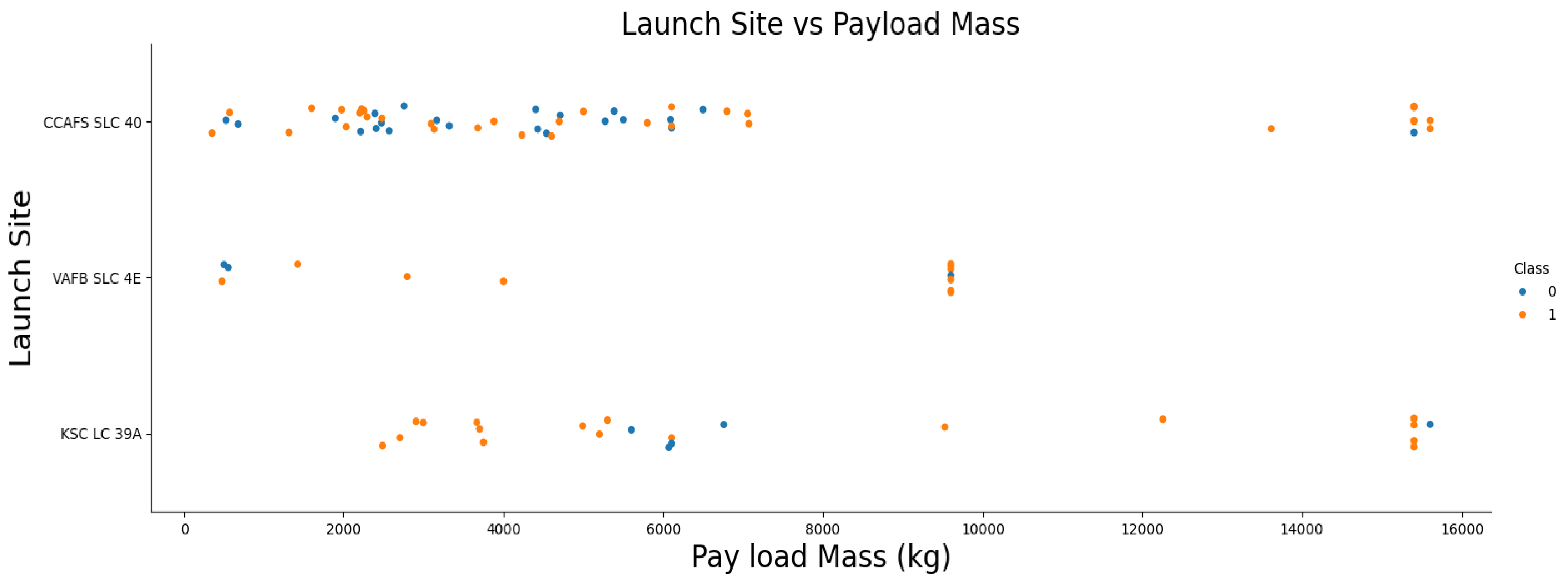


Relationship between flight number and launch site

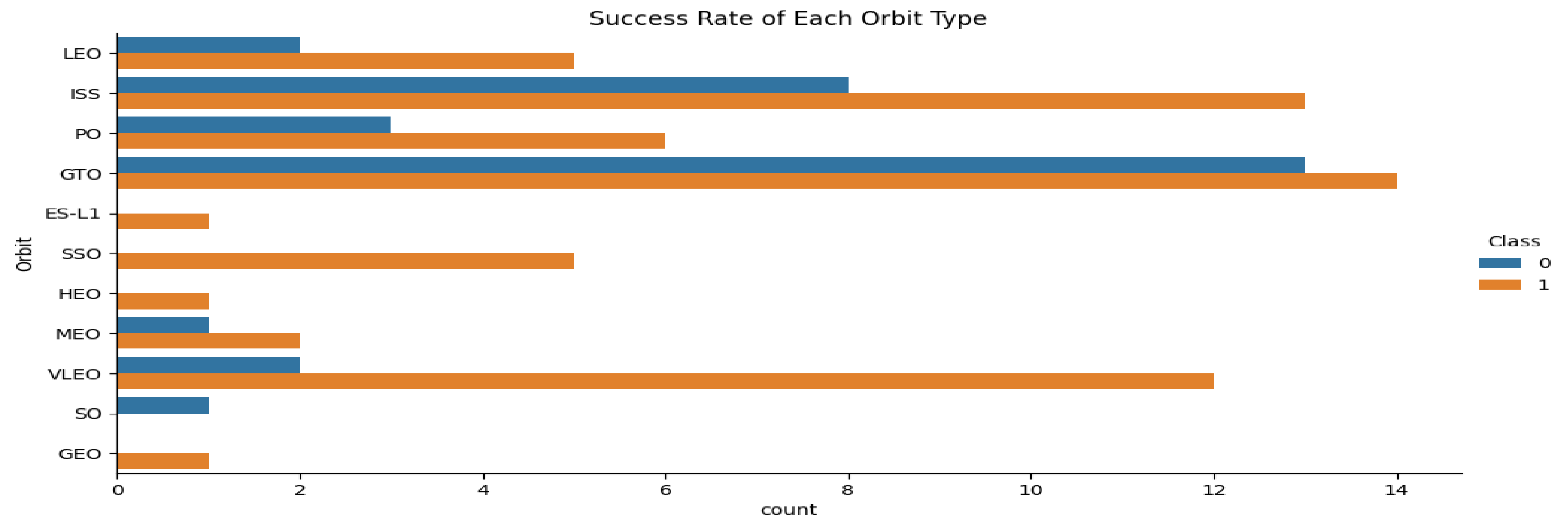
Class 0 = Unsuccessful landing of the Falcon 9 first stage.

Class 1 = Successful landing of the Falcon 9 first stage.

Payload vs. Launch Site

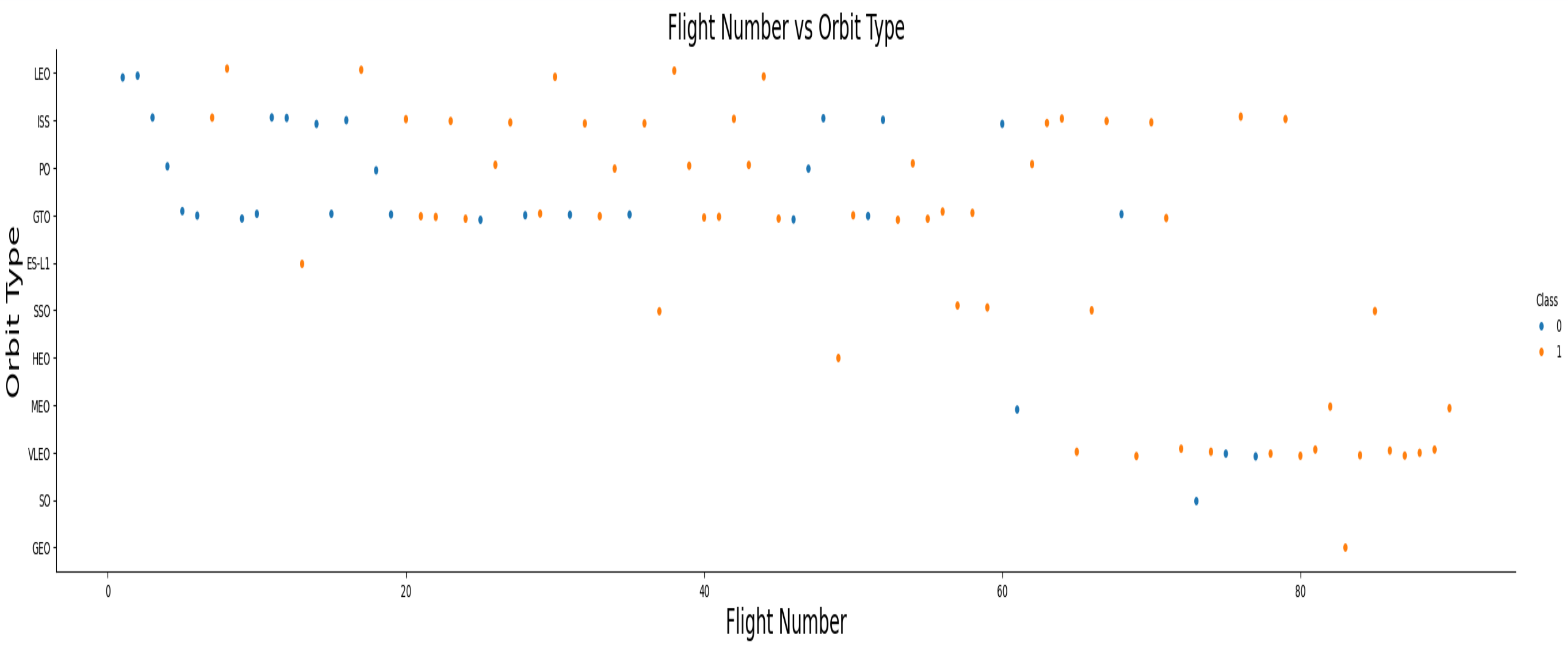


Success Rate vs. Orbit Type

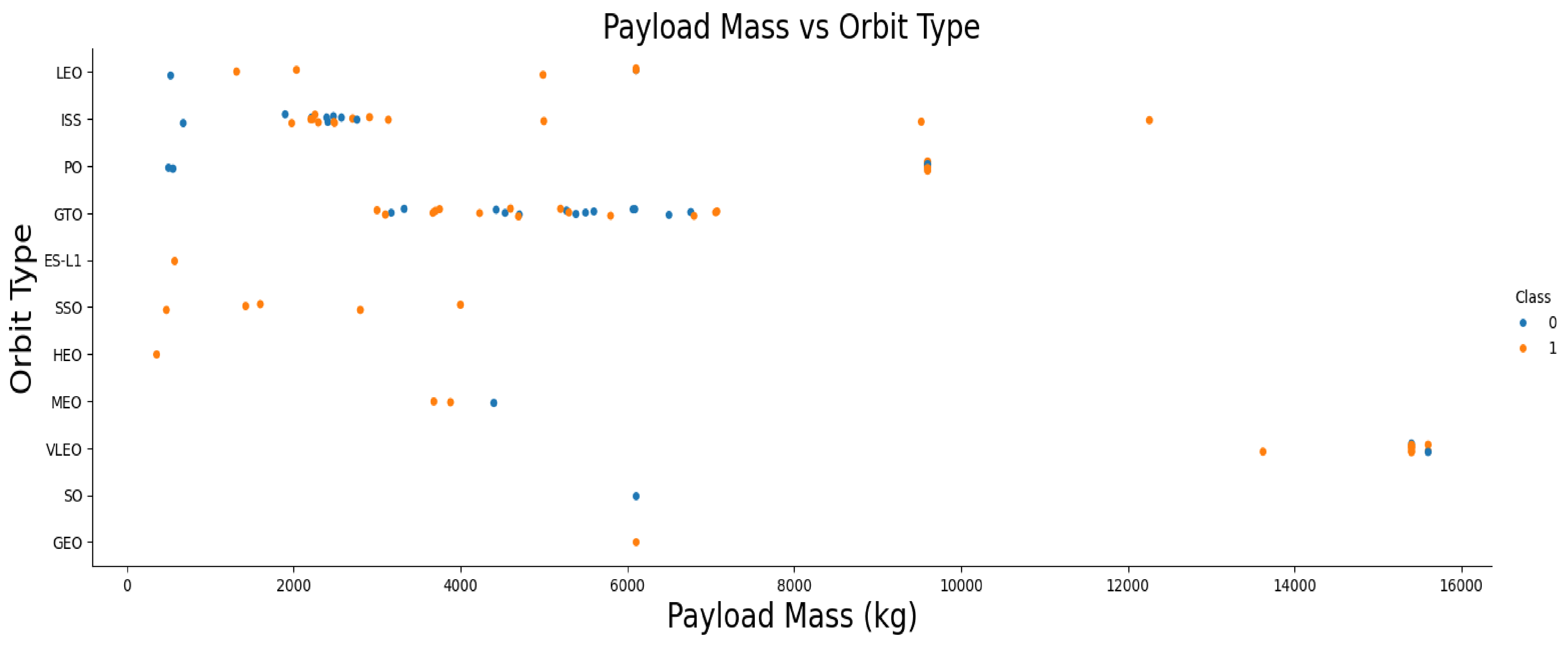


- Most failures occurred in GTO
- In GEO, SSO, and HEO, all attempts were successful.

Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend

- Until 2013, there were no successful landings.

All Launch Site Names

- The names of the unique launch sites

```
# Display the name of unique launch sites in the space mission
%sql select distinct(Launch_Site) from SPACEXTABLE;
```

[10]

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

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
#Display 5 records where launch sites begin with the string 'CCA'
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5;
```

* sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
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

Total Payload Mass

- The total payload carried by boosters from NASA is 45596 KG

```
[16]
# Display the total payload mass carried by boosters launched by NASA (CRS)
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer='NASA (CRS)';
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 KG

```
# Display average payload mass carried by booster version F9 v1.1
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version='F9 v1.1';
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

```
# List the date when the first successful landing outcome in ground pad was achieved
%sql select min(Date) from SPACEXTABLE where Landing_Outcome='Success (ground pad)';
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000;
```

```
 * 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

- The total number of successful and failure mission outcomes

```
[24] %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome='Success';

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

count(Mission_Outcome)
98

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

```
%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

- List the failed Landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select strftime('%m',Date) as Month,Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where strftime('%Y',Date)='2015' and Landing_Outcome='Failure (drone ship)';
```

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

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

- Rank 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

```
%sql select Landing_Outcome, count(Landing_Outcome) as Count from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by Count desc;
```

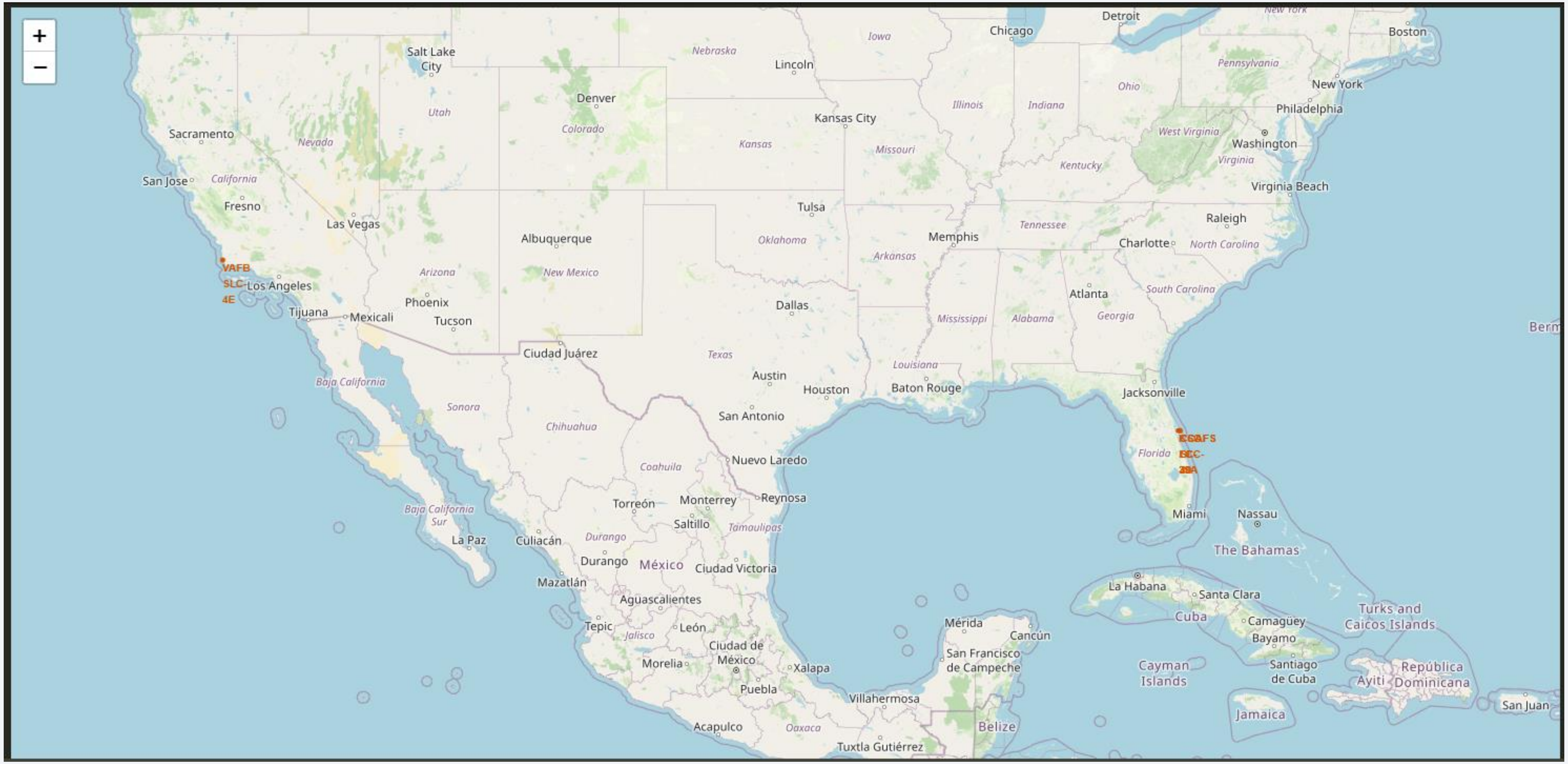
```
 * sqlite:///my_data1.db
Done.
```

Landing_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

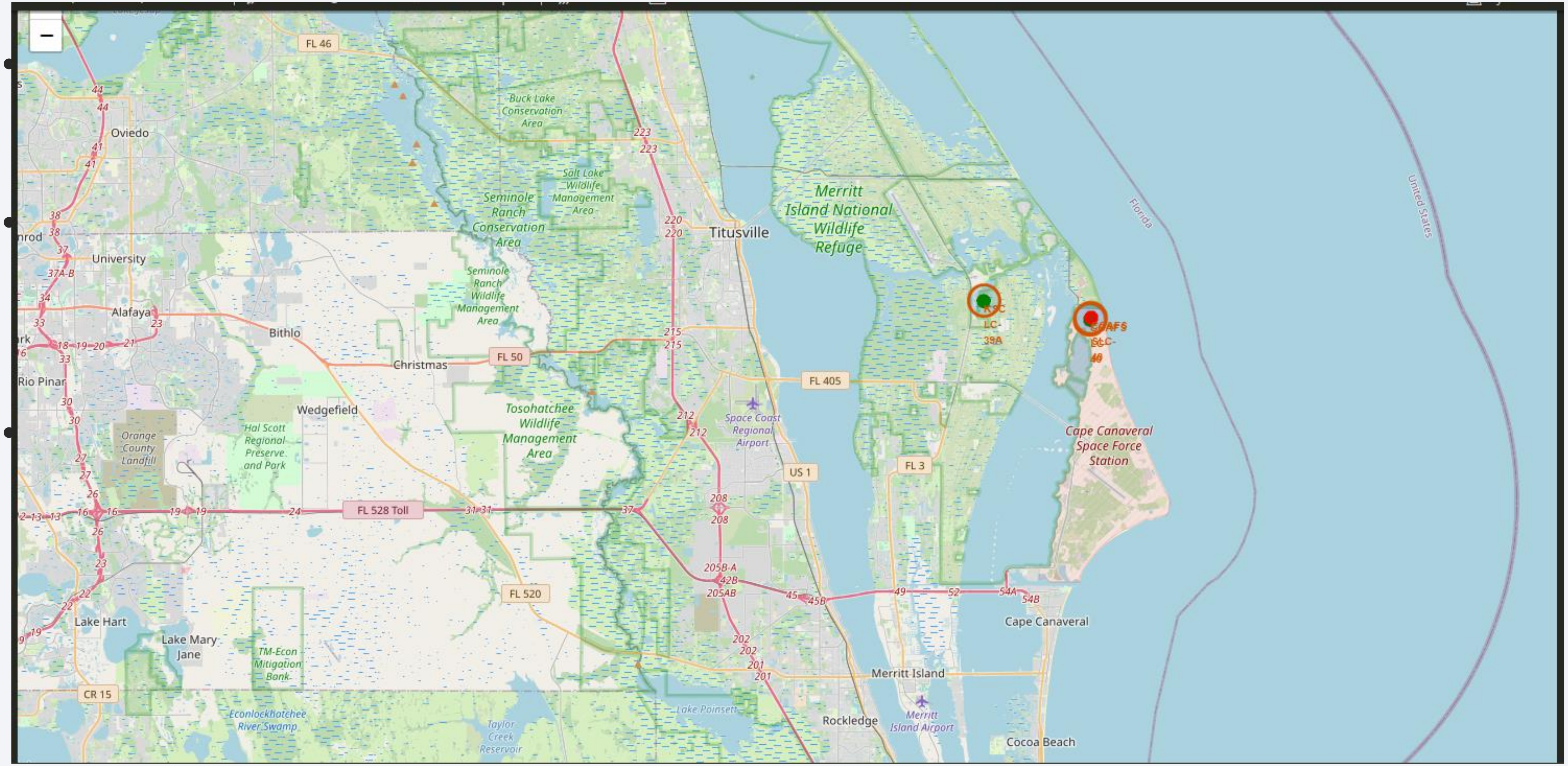
Section 3

Launch Sites Proximities Analysis

Interactive Visual Analytics with Folium



Success/failed launches for each site on the map

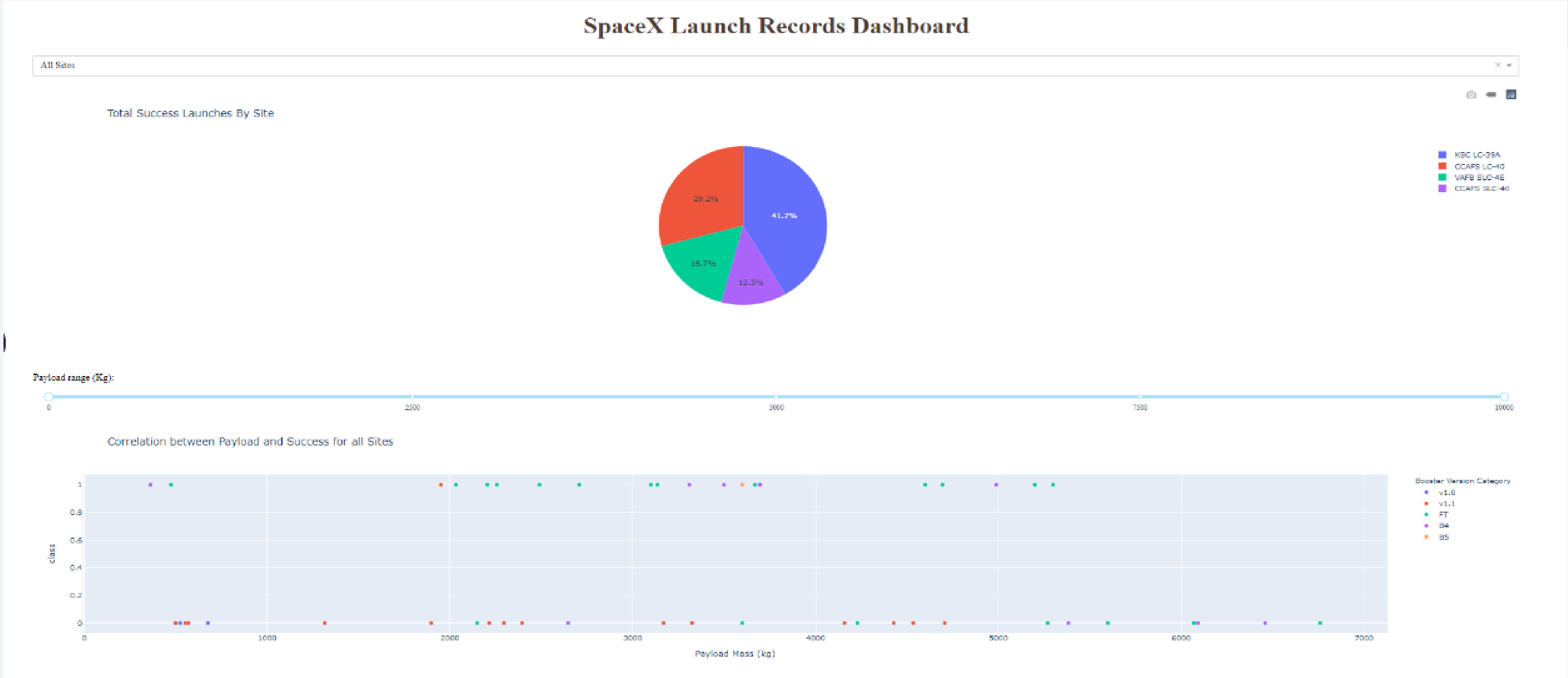


The distances between a launch site to its proximities

Section 4

Build a Dashboard with Plotly Dash

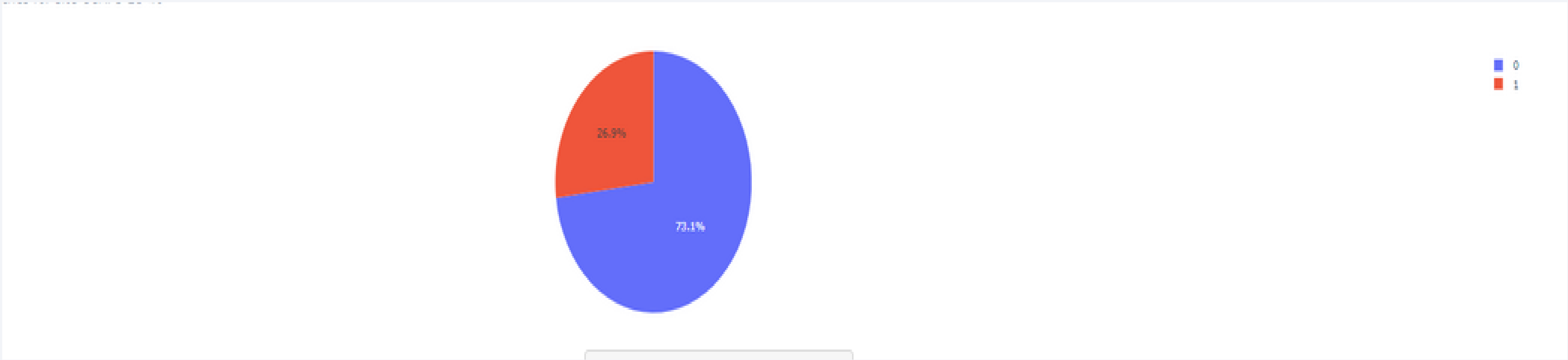
Interactive Dashboard with Ploty-Dash



This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.

Total Success Launches for site CCAFS LC-40

- The screenshot of the pie-chart for the launch site with highest launch success ratio



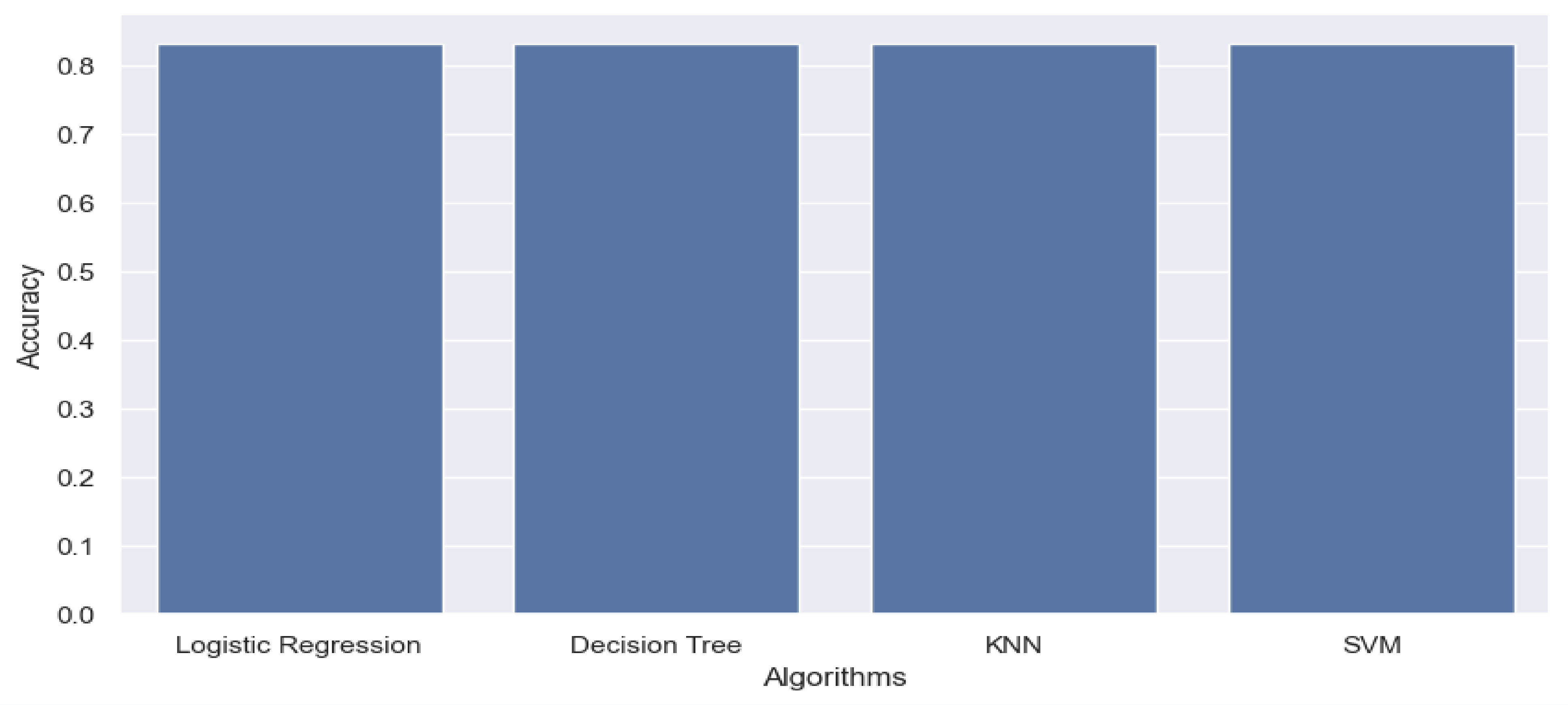
<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

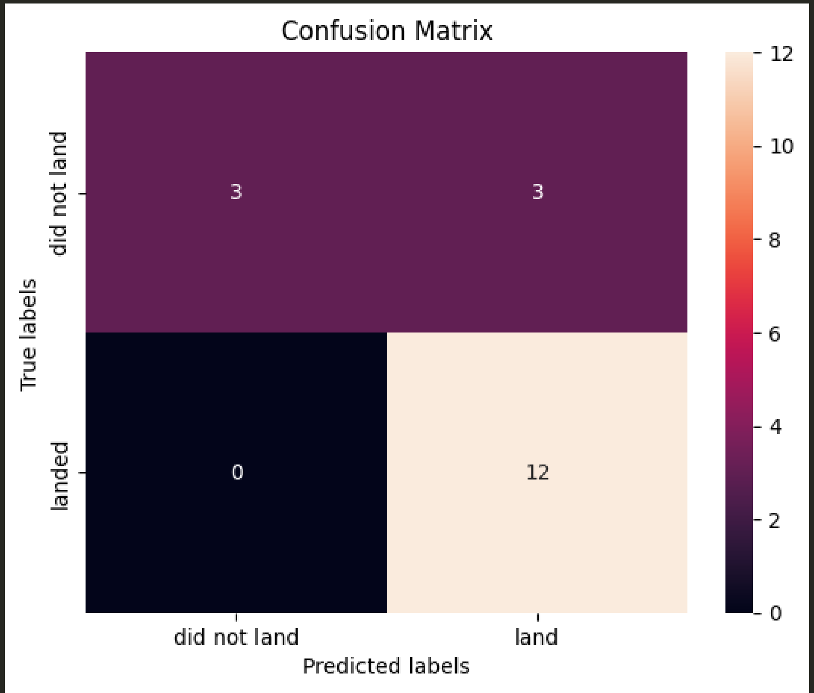
Predictive Analysis (Classification)

Classification Accuracy



- Find which model has the highest classification accuracy

Confusion Matrix

- 

Overview:

True Postive - 12 (True label is landed, Predicted label is also landed)

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

Conclusions

- The analysis provided valuable insights into the success rates of Space-X launches across different orbit types and launch sites.
- Interactive dashboards enabled real-time exploration of data, enhancing the understanding of patterns and trends.
- Predictive models (Logistic Regression, Decision Tree, KNN, SVM) demonstrated similar performance with an accuracy of 83.33%.
- The findings can help optimize launch site selection and improve future landing predictions for Falcon 9.

Appendix

- API – LINK
- Web Scraping – LINK
- Data Wrangling – LINK
- Data-Set, SQL queries – LINK
- Exploring and Preparing Data - LINK
- Interactive Visual Analytics with Folium – LINK
- Machine Learning Prediction - LINK

Thank you!