SPACE RACE USING DATA

TABLE OF CONTENTS

- Executive Summary
  - Project Background
  - Research Questions
- Introduction
  - Objectives
  - Importance of Study
- Data Collection
  - API Data Collection
  - Web Scraping Methods
  - Data Sources and Transformation

TABLE OF CONTENTS

- Methodology
    - Data Wrangling
        - Handling Missing Data
        - Exploratory Data Analysis (EDA)
        - Feature Engineering
    - Data Visualization
        - Types of Graphs Used
        - Key Graphs and Insights
    - SQL Queries
        - Query Descriptions and Results
    - Interactive Visual Analytics
        - Folium Maps
        - Plotly Dashboards
    - Predictive Analysis
        - Modeling Techniques
        - Model Evaluation and Selection

TABLE OF CONTENTS

- **Results**
  - **Key Findings and Insights**
    - **Flight Number vs. Launch Site**
    - **Flight Number vs. Payload Mass**
    - **Success Rate per Orbit**
    - **Launch Success Rate Yearly**
  - **SQL Results**
  - **Interactive Visual Analytics Insights**
    - **Geographical Distribution**
    - **Proximity to Infrastructure**
- **Conclusion**
  - Summary of Findings
  - Implications for SpaceX Launch Strategies
  - Recommendations for Future Research
- **Appendix**

EXECUTIVE SUMMARY

- **Project Goal**: Predicting the success of Falcon 9 first stage landings to determine launch costs.
- **Methods Used**: Data collection via API and web scraping, EDA, interactive visual analytics, predictive modeling.
- **Key Findings**: Success rate increases with flight number, payload mass affects success differently across orbits.

INTRODUCTION

Project Background

SpaceX's Falcon 9 launches and their cost-efficiency due to reusable first stages.

Research Questions

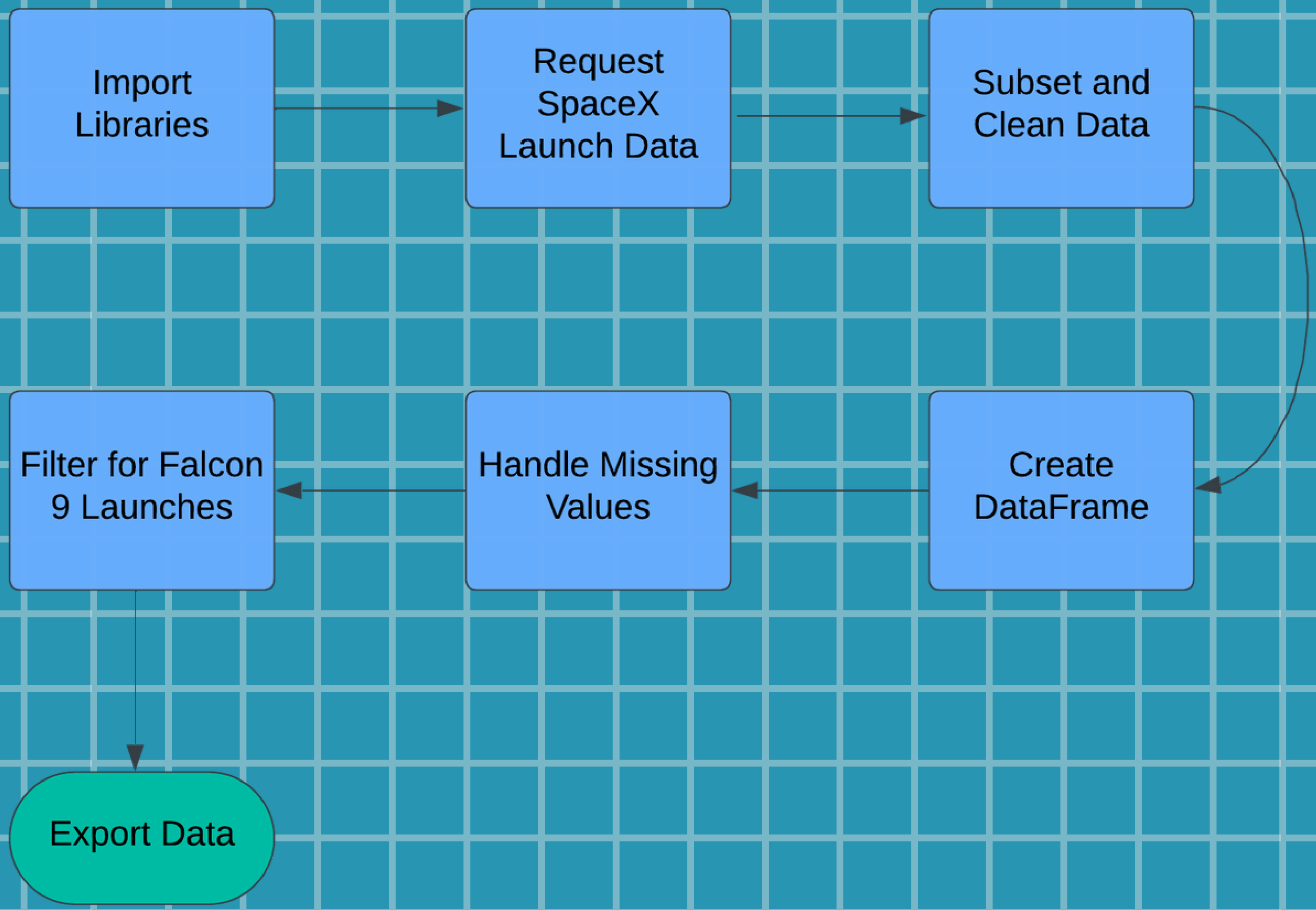
Characteristics of successful landings, effects of rocket variables, conditions for best landing success rate.

METHODOLOGY 01

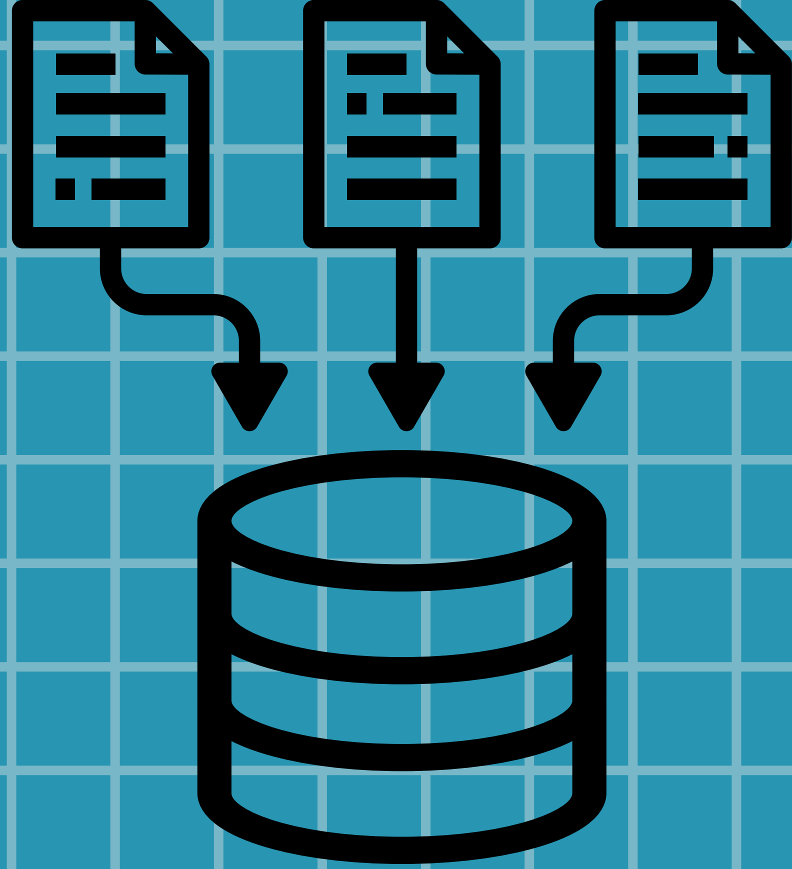
Data Collection

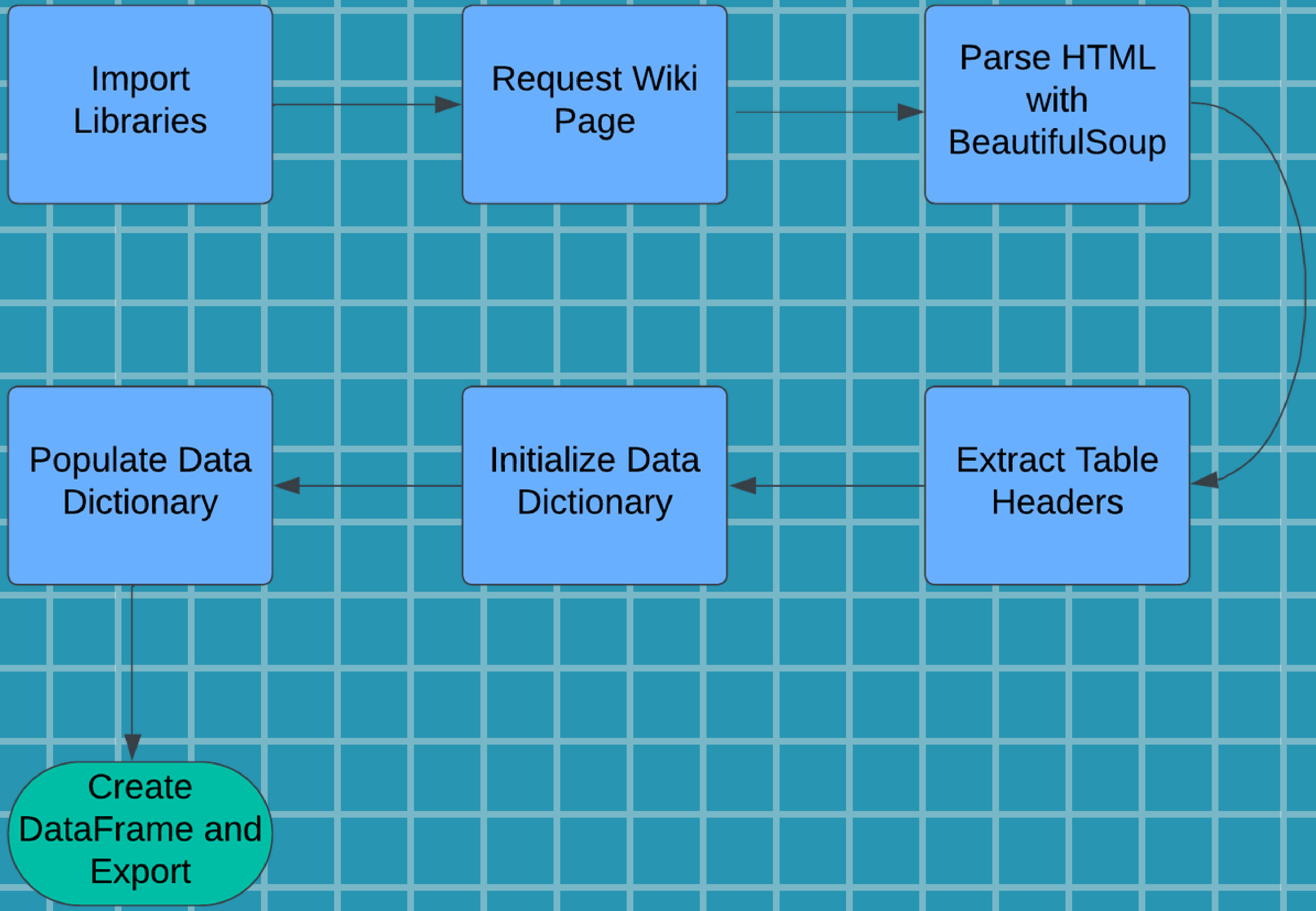
- **API**: Collected data on rockets, launches, and payloads from SpaceX REST API.
- **Web Scraping**: Scraped Wikipedia for launch and landing data.

DATA COLLECTION API



DATA COLLECTION
WEB SCRAPPING





METHODOLOGY 02

Transformations:

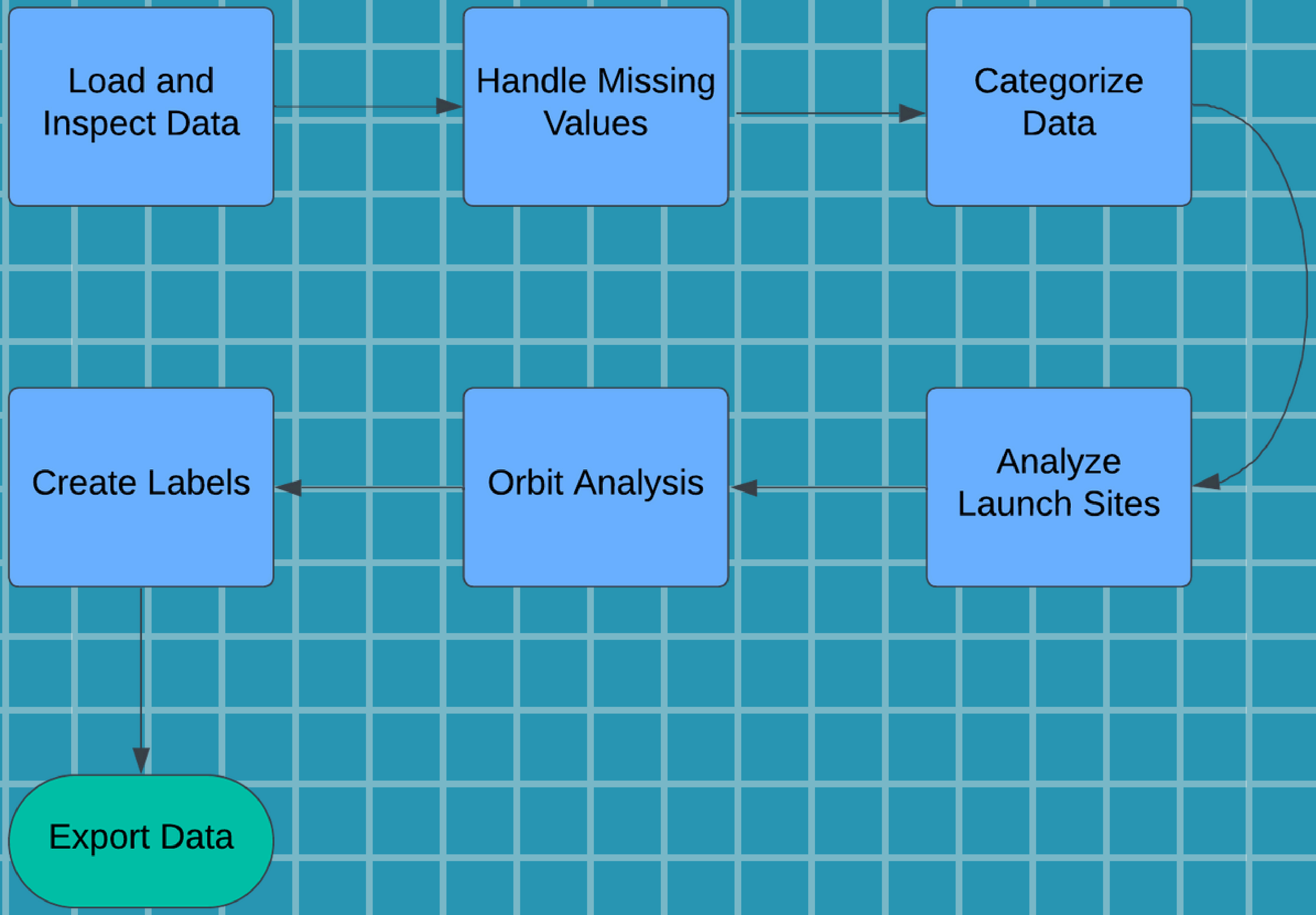
- Converting string variables to categorical, calculating launch site statistics.

WHY DATA WRANGLING?

Objective: Data wrangling prepares raw data for analysis by cleaning and transforming it into a format suitable for modeling. For our Space X Falcon 9 landing prediction project, this involves:

- **Handling Missing Data**: Identifying and addressing missing values to ensure completeness and accuracy.
- **Exploratory Data Analysis (EDA)**: Exploring patterns and summarizing key statistics to understand the dataset's structure and distribution.
- **Feature Engineering**: Converting raw outcomes into binary labels for successful and unsuccessful landings, which is crucial for training predictive models.

DATA WRANGLING PROCESS OVERVIEW

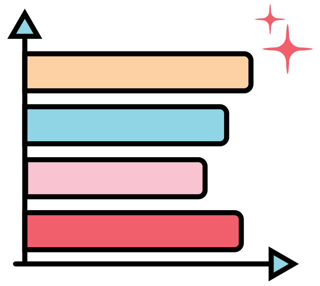


METHODOLOGY 03

EDA with Data Visualization

- **Graphs:** Scatter plots, bar graphs, line graphs to show relationships between variables.

TYPES OF GRAPHS USED



Bar Graphs: Used to compare groups

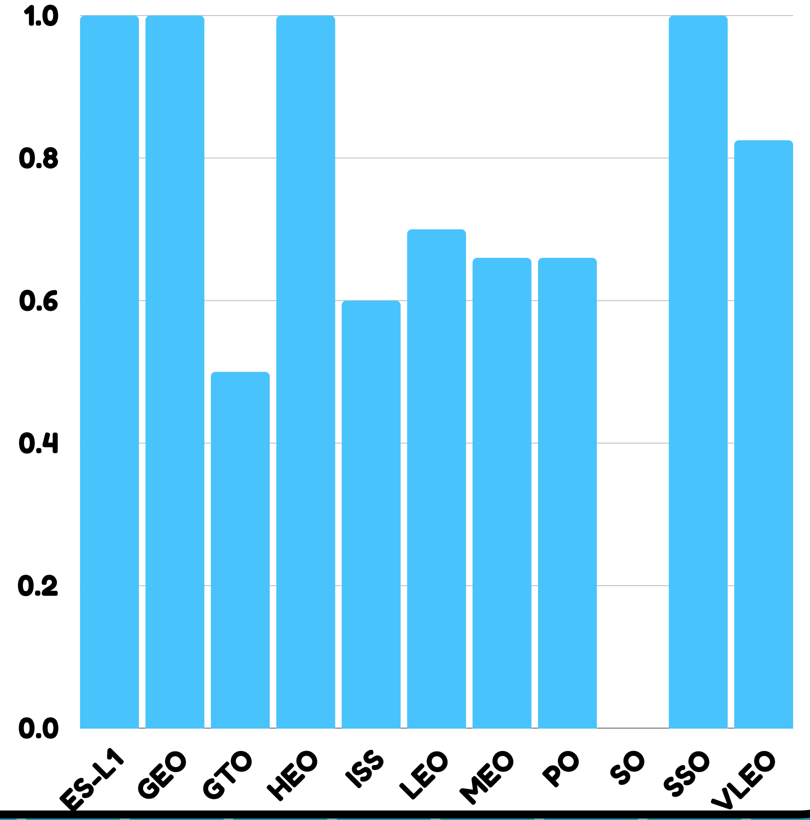
Line Graphs: Used to show changes over time

Scatter Plot: Used to show relationship between variables

BAR GRAPHS

Success Rate by Orbit Type

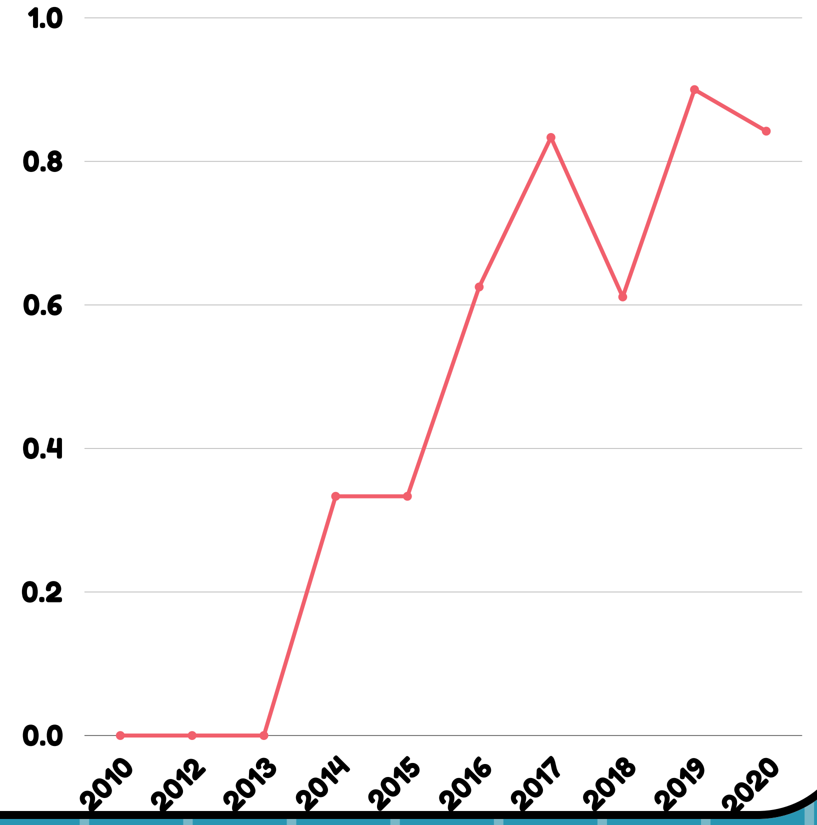
- To evaluate and compare the success rates across different orbit types. This bar chart shows which orbit types have higher success rates, providing insights into how orbit choice affects mission outcomes.

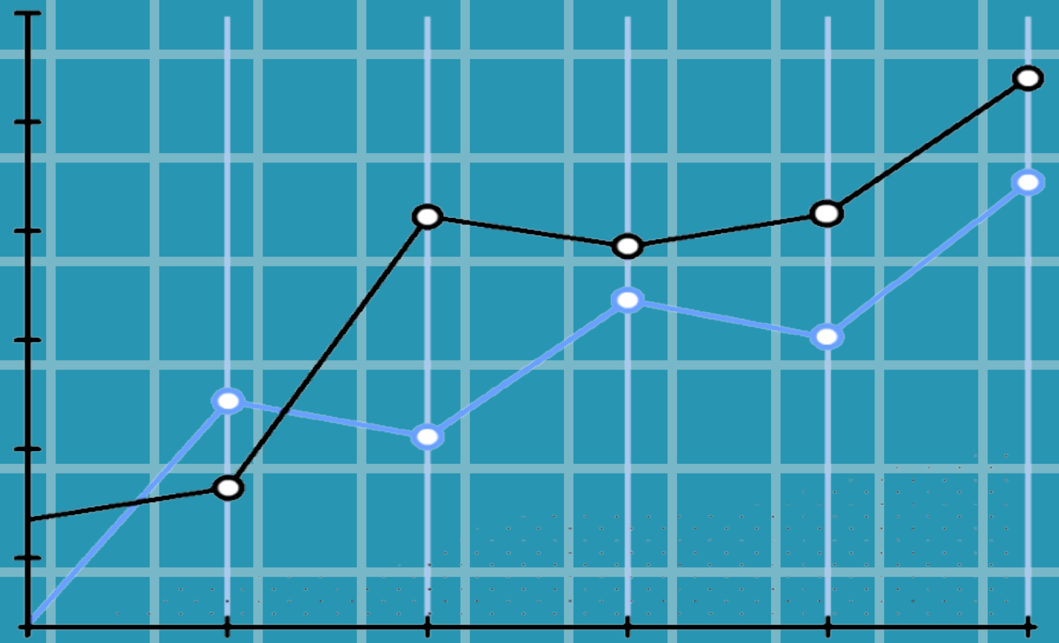


LINE GRAPH

Success Rate by Year

- To analyze the trend in success rates over the years. This line chart helps identify any long-term improvements or declines in landing success, showing how SpaceX's performance has evolved from year to year.





SCATTER PLOT

Flight Number vs. Payload Mass

- To visualize how the flight number and payload mass affect the landing outcome. This plot helps identify if there's a pattern in the success rate based on flight history and payload size. It shows that as the flight number increases, the likelihood of a successful landing improves. Payload mass also affects the outcome, with heavier payloads still showing a reasonable success rate.

SCATTER PLOT

Flight Number vs. Launch Site

- To explore the distribution of launches across different sites and their success rates. This plot highlights the frequency of launches at each site and how it correlates with success, helping to understand site-specific performance.

SCATTER PLOT

Payload Mass vs. Launch Site

- To examine the relationship between payload mass and launch sites. This visualization reveals whether certain sites handle specific payload sizes better, and helps in identifying if some sites are more likely to launch heavier payloads.

SCATTER PLOT

Flight Number vs. Orbit Type

- To examine the relationship between payload mass and launch sites. This visualization reveals whether certain sites handle specific payload sizes better, and helps in identifying if some sites are more likely to launch heavier payloads.

SCATTER PLOT

Payload Mass vs. Orbit Type

- To explore how payload mass affects success rates across different orbits. It helps in understanding if heavier payloads impact the success rate in certain orbits, providing a visual correlation between mass and orbit type.

METHODOLOGY 04

EDA with SQL

- **SQL:** Extract insights using SQL queries.

QUERIES PERFORMED

- Display the names of the unique launch sites.
- Display 5 records where launch sites begin with the string 'CCA'.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display the average payload mass carried by booster version F9 v1.1.
- List the date when the first successful landing outcome on a ground pad was achieved.
- List the names of the boosters that have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass using a subquery.
- List the records displaying the month names, failure landing outcomes in drone ship, booster versions, and launch sites for the months in the year 2015.
- Rank the count of landing outcomes between the dates 2010-06-04 and 2017-03-20, in descending order.

METHODOLOGY 04

Interactive Visual Analytics:

- **Folium Maps:** Interactive maps to show launch sites and success rates.
- **Plotly Dashboards:** Interactive dashboards to visualize data dynamically.

FOLIUM MAPS

1. **Marked All Launch Sites on a Map:**
   - Created a map centered on NASA Johnson Space Center.
   - Added circles and markers for each launch site using their latitude and longitude coordinates.
2. **Marked Success/Failed Launches for Each Site:**
   - Used a MarkerCluster to visualize the success (green) and failure (red) of launches.
   - Added markers to the map for each launch record based on their success or failure status.
3. **Calculated Distances Between Launch Sites and Proximities:**
   - Added a MousePosition plugin to get coordinates of points of interest.
   - Calculated distances from launch sites to nearby coastlines, highways, etc.
   - Drew PolyLine lines between launch sites and their closest points of interest to visualize distances.

PLOTLY DASHBOARDS

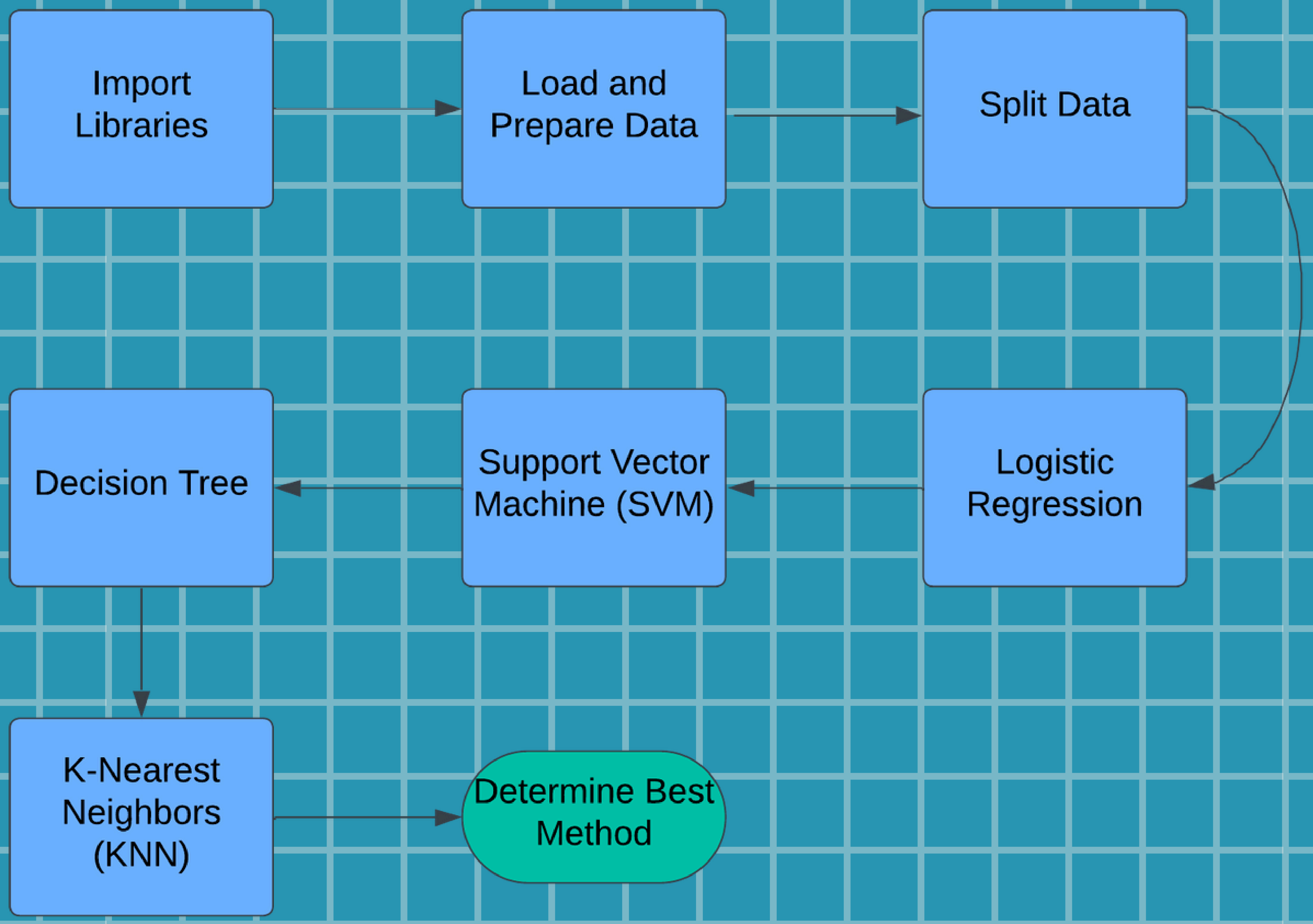
1. **Dropdown for Launch Site Selection:**
   - Added a dropdown menu to select a specific launch site or view data for all sites.
2. **Pie Chart for Launch Success:**
   - Created a pie chart to show the distribution of successful vs. failed launches.
   - If a specific launch site is selected, the pie chart displays success vs. failure counts for that site.
3. **Slider for Payload Range:**
   - Added a slider to select a range of payload masses for filtering the data.
4. **Scatter Chart for Payload vs. Launch Success:**
   - Created a scatter plot to show the correlation between payload mass and launch success.
   - The scatter plot is updated based on the selected launch site and payload range.

METHODOLOGY 04

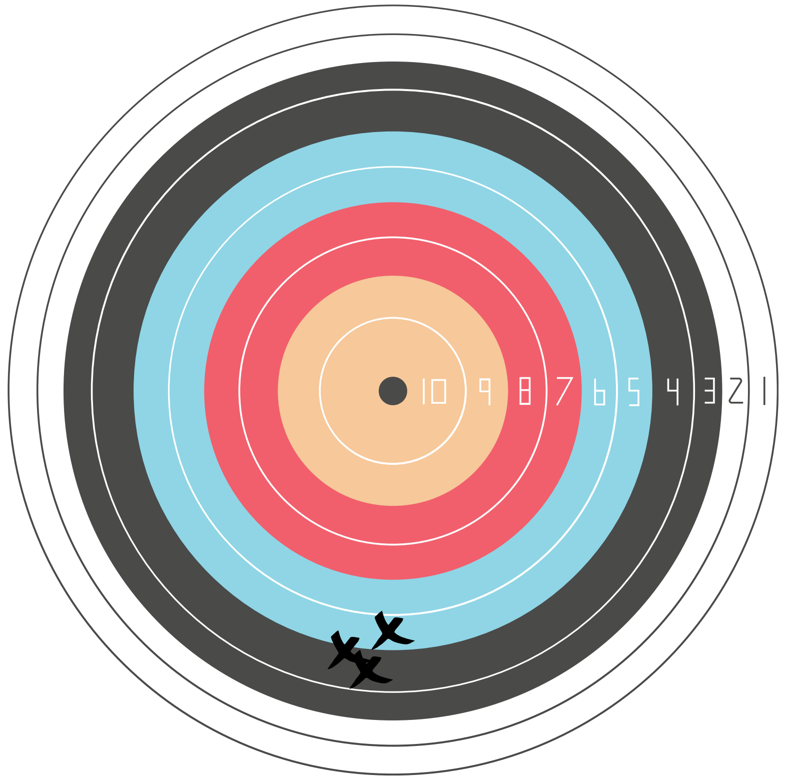
Predictive Analysis:

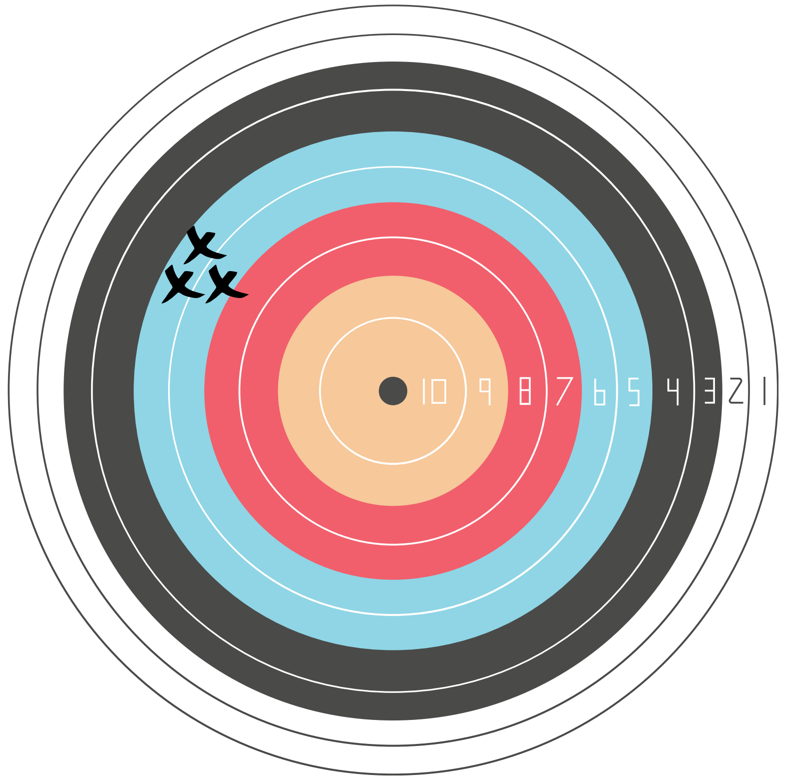
- Modeling**:** Building, tuning, and evaluating classification models.

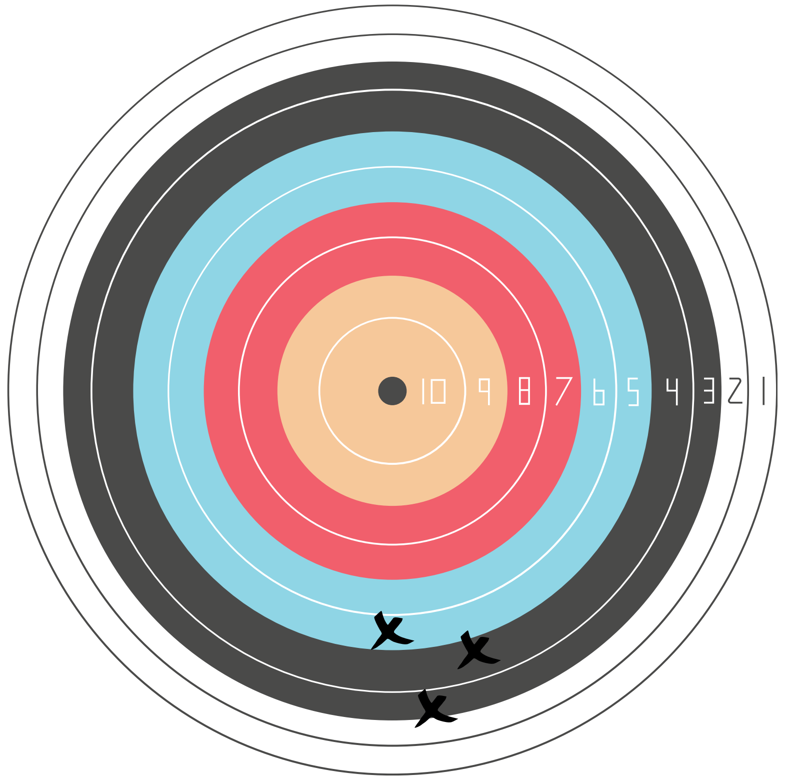
PREDICTIVE ANALYSIS



RESULTS

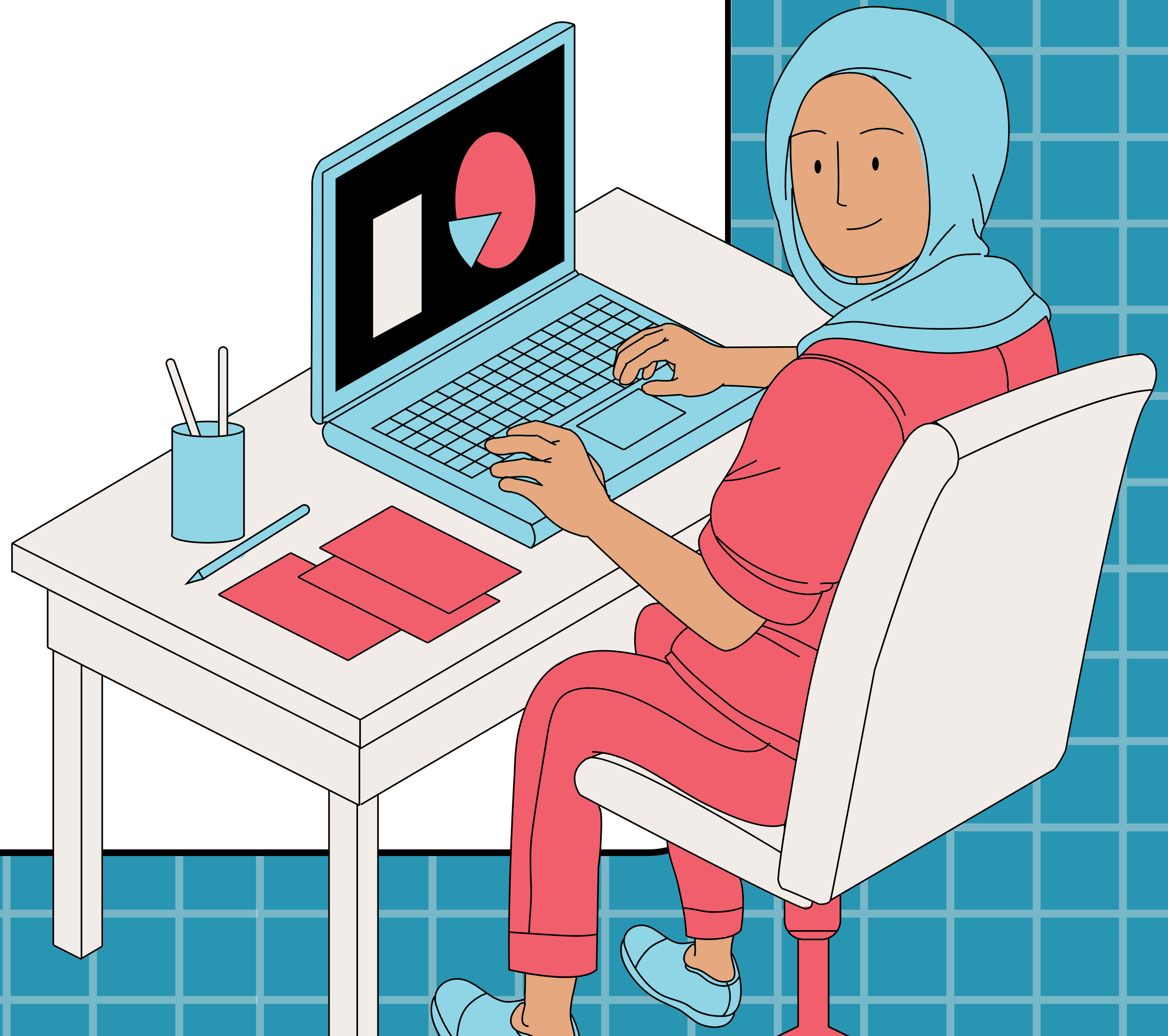




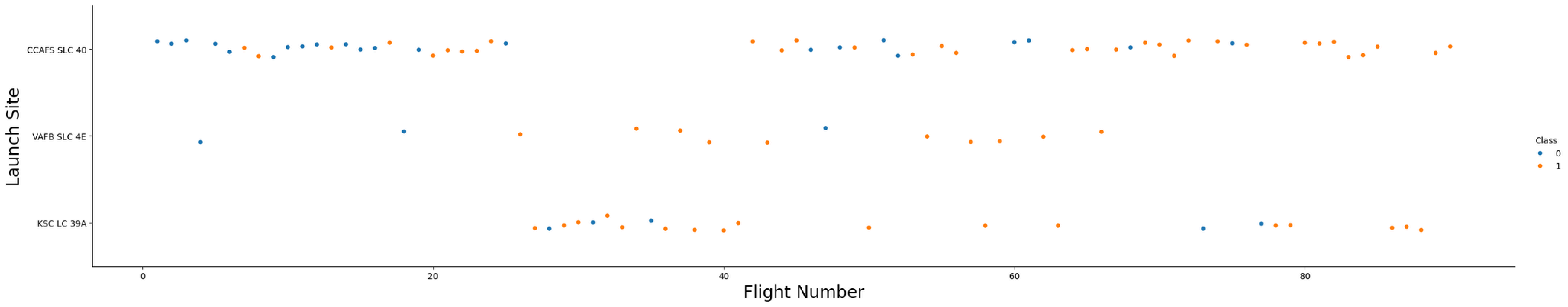


VISUALIZATIONS

In this section, we will explore a series of visualizations designed to uncover insights and patterns within our dataset. This section DOESN'T include all the graphs plotted during the investigation.



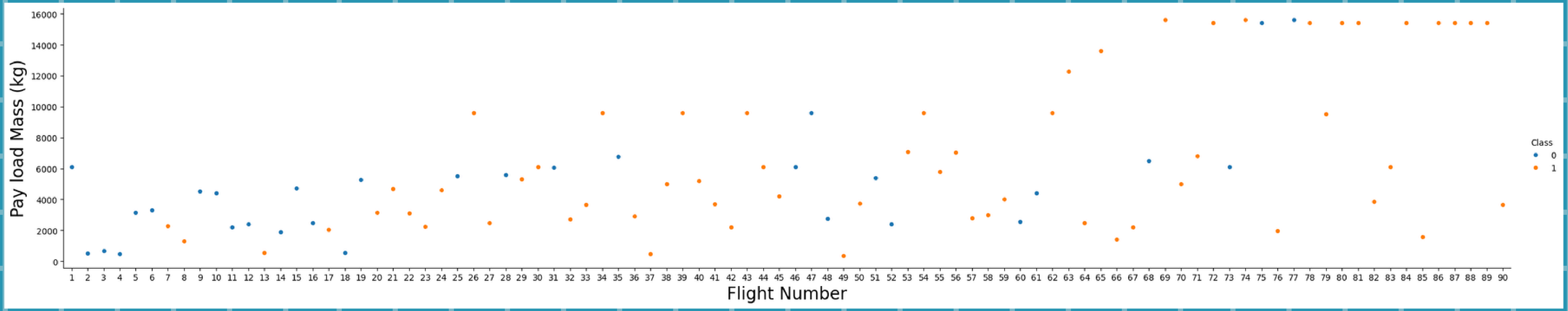
FLIGHT NUMBER VS. LAUNCH SITE



Insight:

The scatter plot reveals a clear trend of increasing landing success rates with higher flight numbers across different launch sites, highlighting the advancements and learning within SpaceX's launch program over time. The CCAFS SLC 40 and KSC LC 39A sites show particularly strong trends towards higher success rates in later flights.

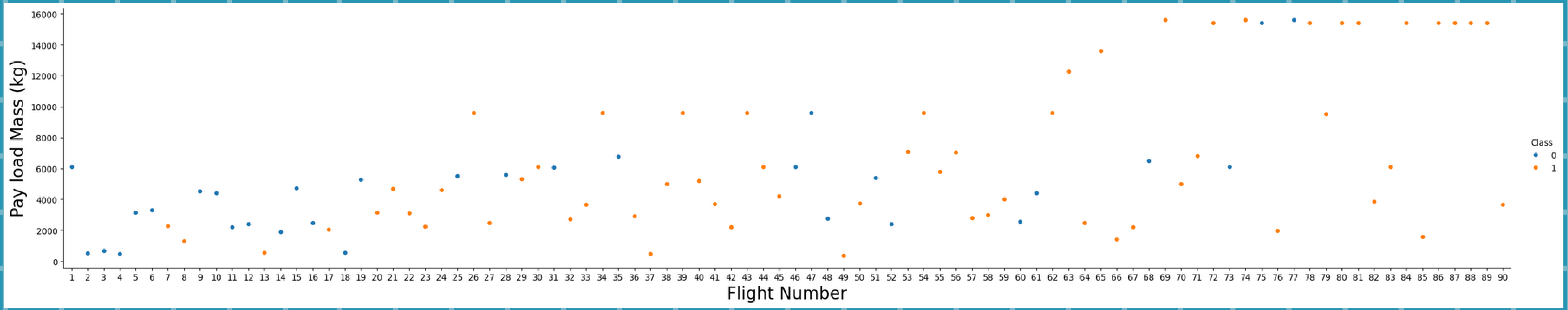
FLIGHT NUMBER VS. PAY LOAD MASS



Insight:

The scatter plot reveals a clear trend of increasing landing success rates with higher flight numbers across different launch sites, highlighting the advancements and learning within SpaceX's launch program over time. The CCAFS SLC 40 and KSC LC 39A sites show particularly strong trends towards higher success rates in later flights. Additionally, the data indicates that higher payload mass is generally associated with higher success rates, and KSC LC 39A has demonstrated a perfect success rate for payloads under 5500 kg.

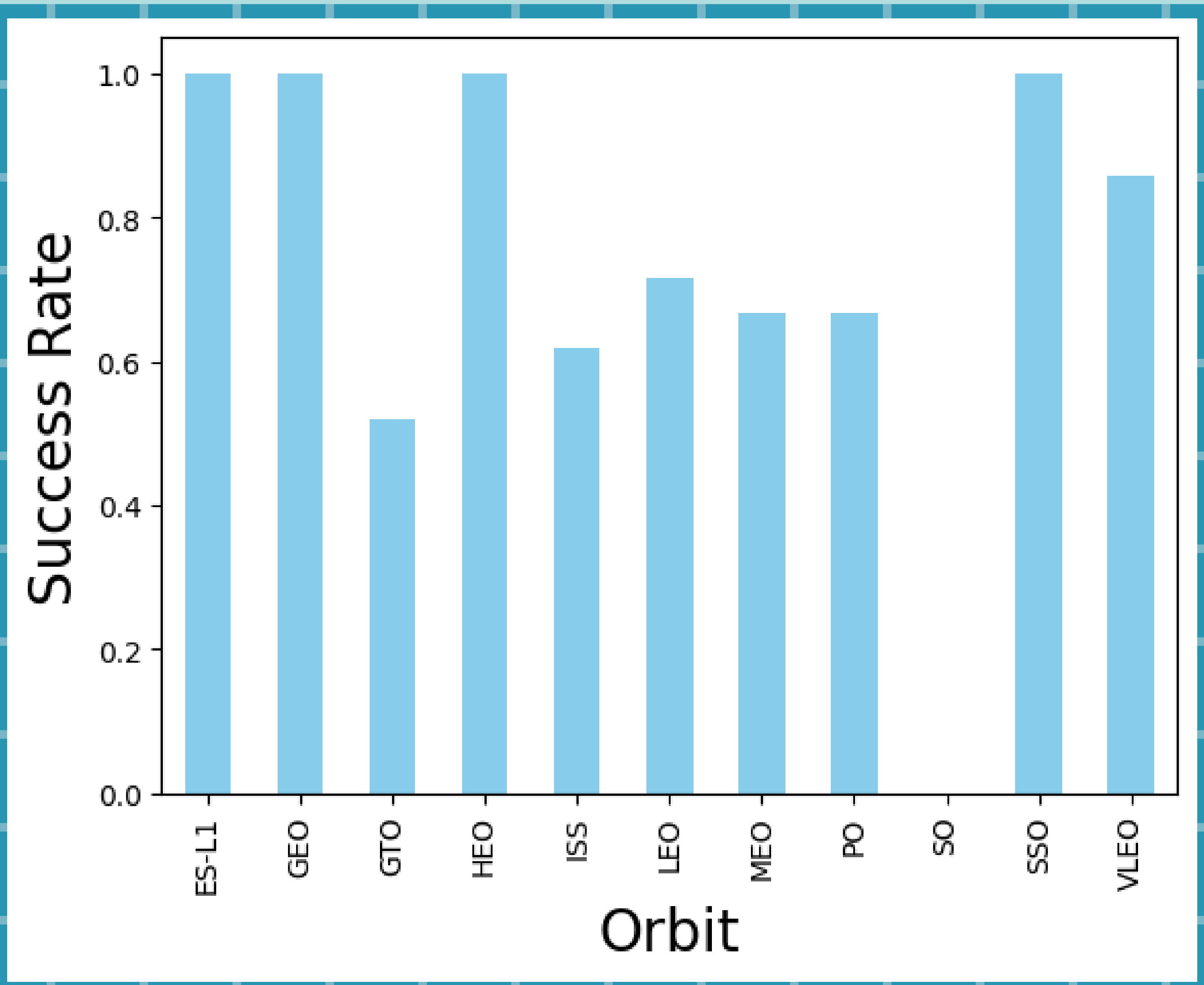
FLIGHT NUMBER VS. PAY LOAD MASS



Insight:

The scatter plot reveals a clear trend of increasing landing success rates with higher flight numbers across different launch sites, highlighting the advancements and learning within SpaceX's launch program over time. The CCAFS SLC 40 and KSC LC 39A sites show particularly strong trends towards higher success rates in later flights. Additionally, the data indicates that higher payload mass is generally associated with higher success rates, and KSC LC 39A has demonstrated a perfect success rate for payloads under 5500 kg.

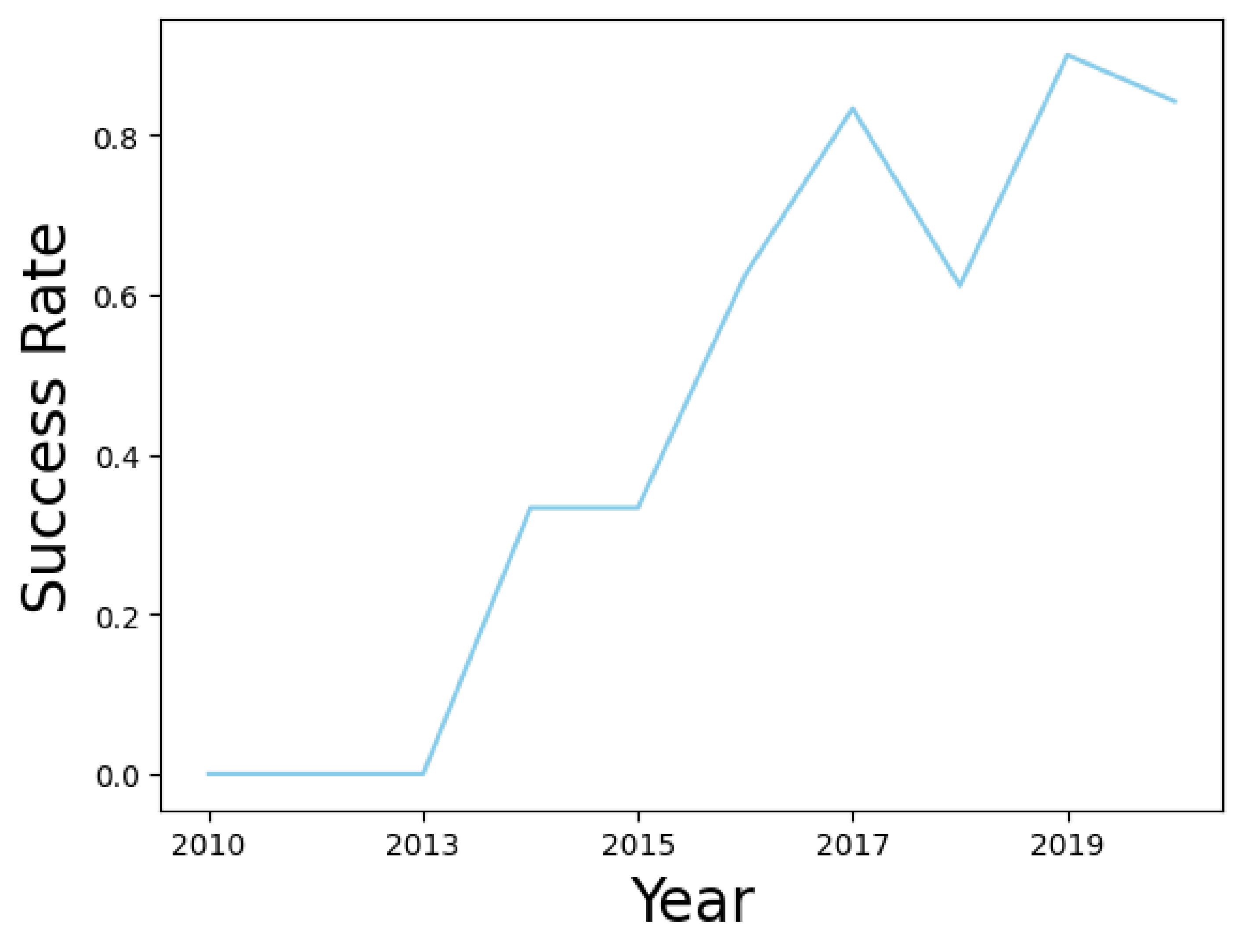
SUCCESS RATE PER ORBIT



Recommendations

- **For High Reliability:**
  - Prefer orbits with a 100% success rate (ES-L1, GEO, HEO, SSO) for mission-critical launches. These orbits have demonstrated consistent success and are likely to provide reliable outcomes.
- **For Caution:**
  - Be cautious with orbits with a 0% success rate (SO). Investigate the causes of past failures and consider risk mitigation strategies if planning a mission for this orbit.
- **For Moderate Success:**
  - For orbits with success rates between 50% and 85%, ensure thorough preparation and analysis to increase the likelihood of mission success. Consider implementing best practices and learning from past missions to enhance reliability.

LAUNCH SUCCESS RATE YEARLY

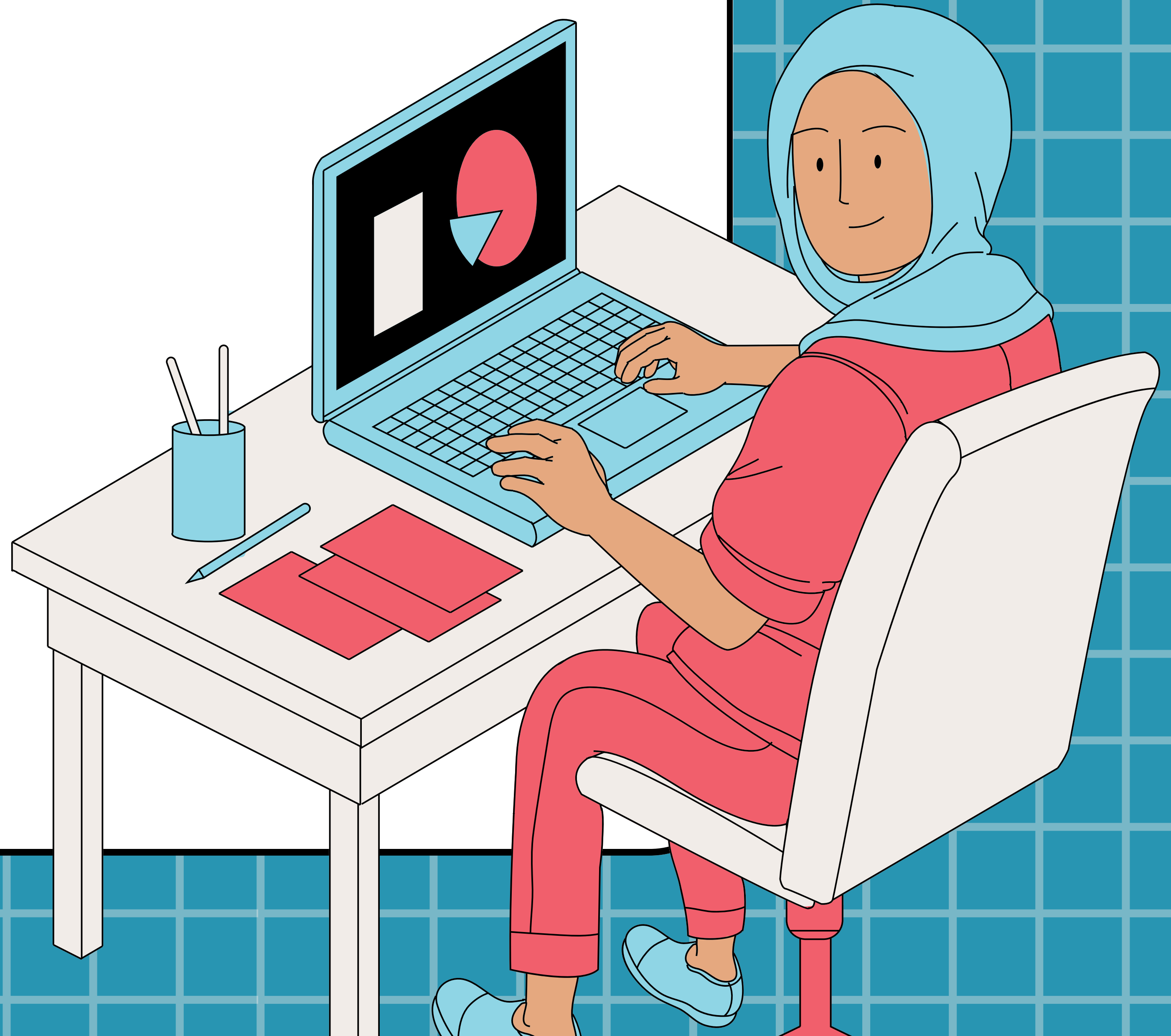


Insight:

- Since 2013, the success rate of launches has consistently improved each year.
- The increasing success rate indicates significant improvements in technology, engineering, and operational procedures. This could be due to advancements in rocket design, more rigorous testing, or enhanced launch strategies.

SQL RESULTS

In this section,



SQL 01

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

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

Explanation
Displaying the names of the unique launch sites in the space mission

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

✓ 0.0s

* 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

Explanation
Displaying 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

Average
2928.4

Explanation
Displaying the average payload mass carried by booster version F9 v1.1

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as Total FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

Total
45596

Explanation
Total payload mass carried by boosters launched by NASA (CRS)

SQL 02

```
%sql SELECT MIN(DATE) AS First_Success FROM SPACEXTABLE WHERE Landing_Outcome ='Success (ground pad)'
✓ 0.0s
* sqlite:///my_data1.db
Done.
```

First_Success
2015-12-22

Explanation
Showing the date when the first successful landing outcome in ground pad was achieved.

```
%sql SELECT COUNT(Landing_Outcome) AS Total FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success%' OR Landing_Outcome LIKE 'Failure%'
✓ 0.0s
* sqlite:///my_data1.db
Done.
```

Total
71

Explanation
Total number of successful and failure mission outcomes

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
✓ 0.0s
* sqlite:///my_data1.db
Done.
```

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

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

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
✓ 0.0s
* 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

Explanation
Listing the names of the booster_versions which have carried the maximum payload mass.

SQL 03

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

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

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

Explanation

Displaying the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year2015.

```
%sql SELECT Landing_Outcome, Count(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC
```

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

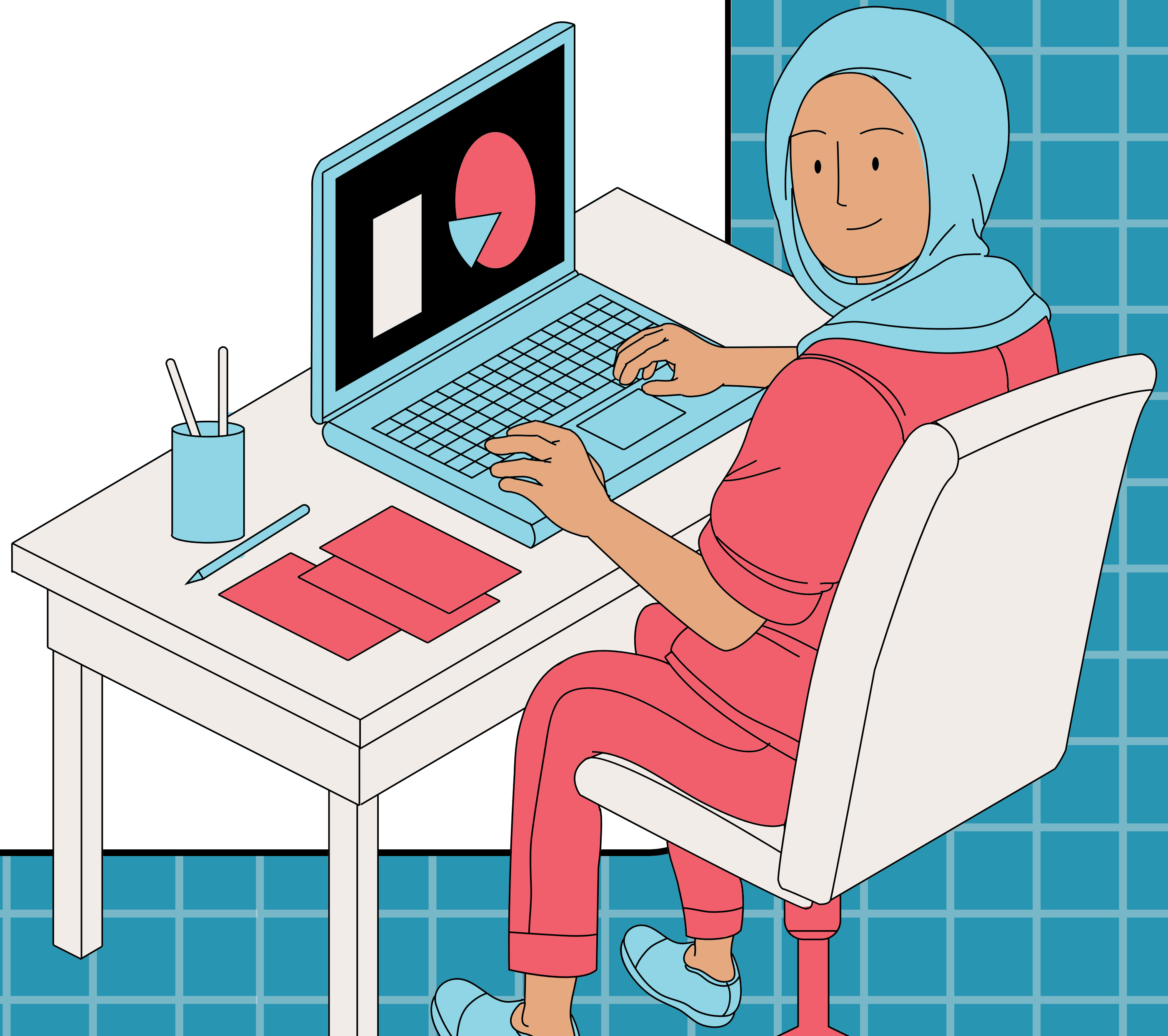
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Explanation

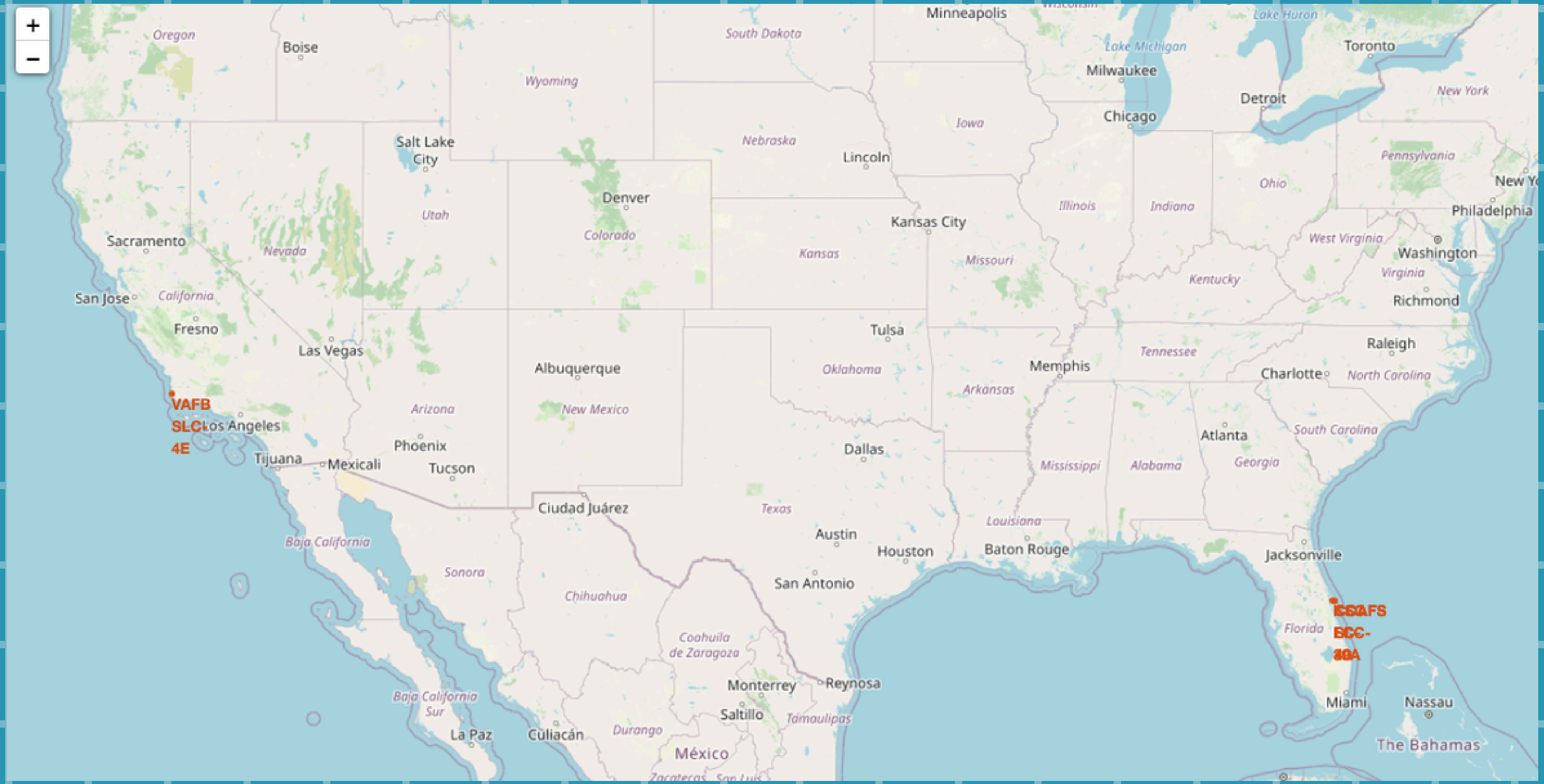
Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

INTERACTIVE VISUAL ANALYTICS

In this section,



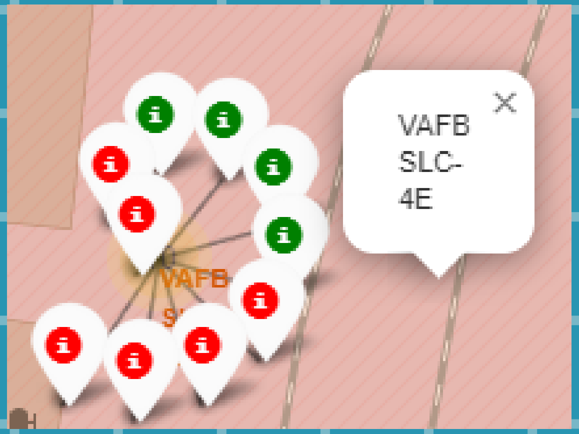
FOLIUM 01

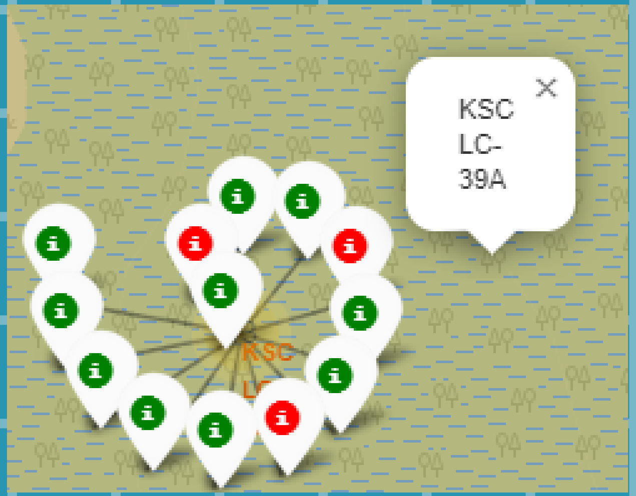


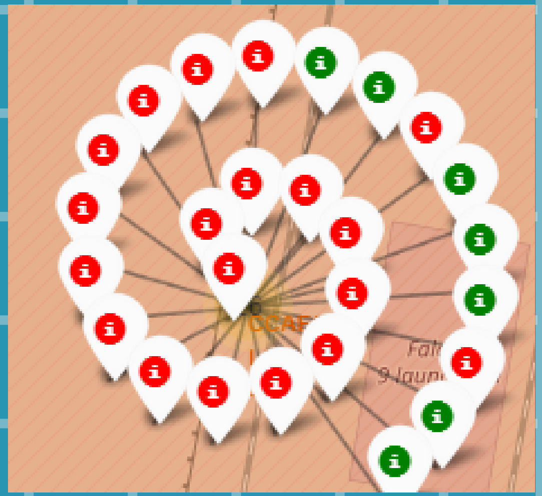
Insights

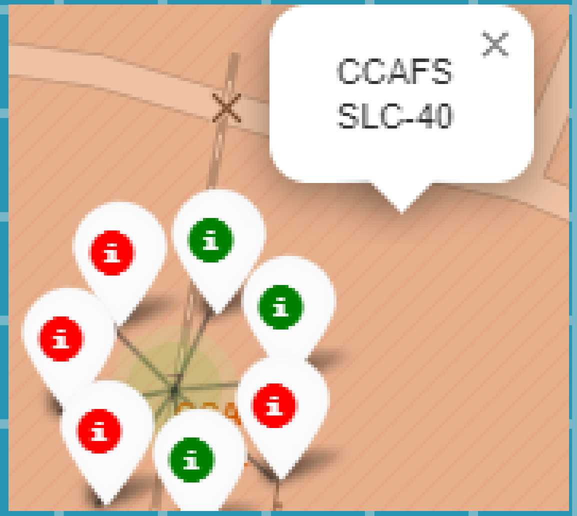
- **Geographical Distribution:**
  - The launch sites are clustered in two main regions: the West Coast (California) and the East Coast (Florida).
  - California launch sites: VAFB (Vandenberg Air Force Base), SLC, 4E.
  - Florida launch sites: LC-39A (Launch Complex 39A), SLC-40 (Space Launch Complex 40), and SLC-41 (Space Launch Complex 41).
- **Proximity to Coastlines:**
  - All the launch sites are located near coastlines. This is strategic as launching over the ocean minimizes the risk to populated areas in case of launch failures and allows for a safe splashdown of rocket stages or components.
- **Proximity to Major Cities:**
  - The Florida launch sites are relatively close to major cities such as Miami and Orlando, which might be beneficial for logistics and workforce accessibility.
  - The California launch sites are near Los Angeles, another major city that can support the logistics and workforce needs.
- **Proximity to Infrastructure:**
  - Launch sites are likely chosen based on their proximity to critical infrastructure such as highways and railways, ensuring that transportation of rockets and supplies is efficient.
  - Vandenberg Air Force Base and the Florida sites are both located near major transportation routes.

FOLIUM 02









Insights

- **VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4E):**
  - The majority of launches are successful (green markers).
  - There are a few failures (red markers), but successes significantly outnumber failures.
- **KSC LC-39A (Kennedy Space Center Launch Complex 39A):**
  - **Predominantly successful launches (green markers).**
  - **Very few failures (red markers), indicating a high success rate.**

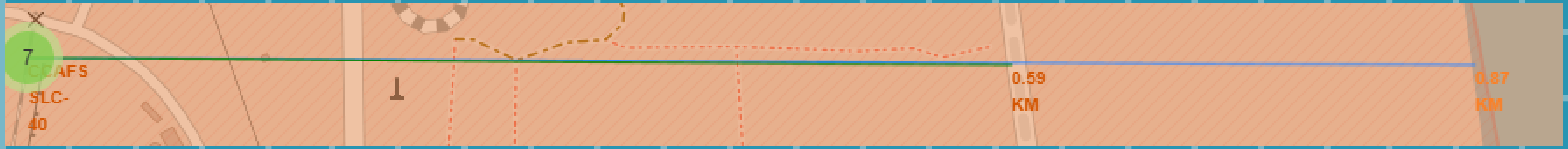
Insights

- **CCAFS LC-40 (Cape Canaveral Air Force Station Space Launch Complex 40):**
  - There is a mix of successful and failed launches, with a notable number of failures (red markers).
  - However, successful launches (green markers) are also present, indicating a moderate success rate.
- **CCAFS SLC-40 (Cape Canaveral Air Force Station Space Launch Complex 40):**
  - Similar to VAFB SLC-4E, there are both successes (green markers) and failures (red markers).
  - The ratio of successful to failed launches seems balanced, suggesting room for improvement in launch outcomes.

Overall Insights:

- Kennedy Space Center Launch Complex 39A stands out with the highest success rate, suggesting this site is particularly reliable for launches.
- Vandenberg Air Force Base Space Launch Complex 4E also shows a strong success rate, although it has a few failures.
- Cape Canaveral Air Force Station Space Launch Complex 40 has a more mixed record with a relatively higher number of failures, indicating that this site may have faced more challenges or higher-risk missions.

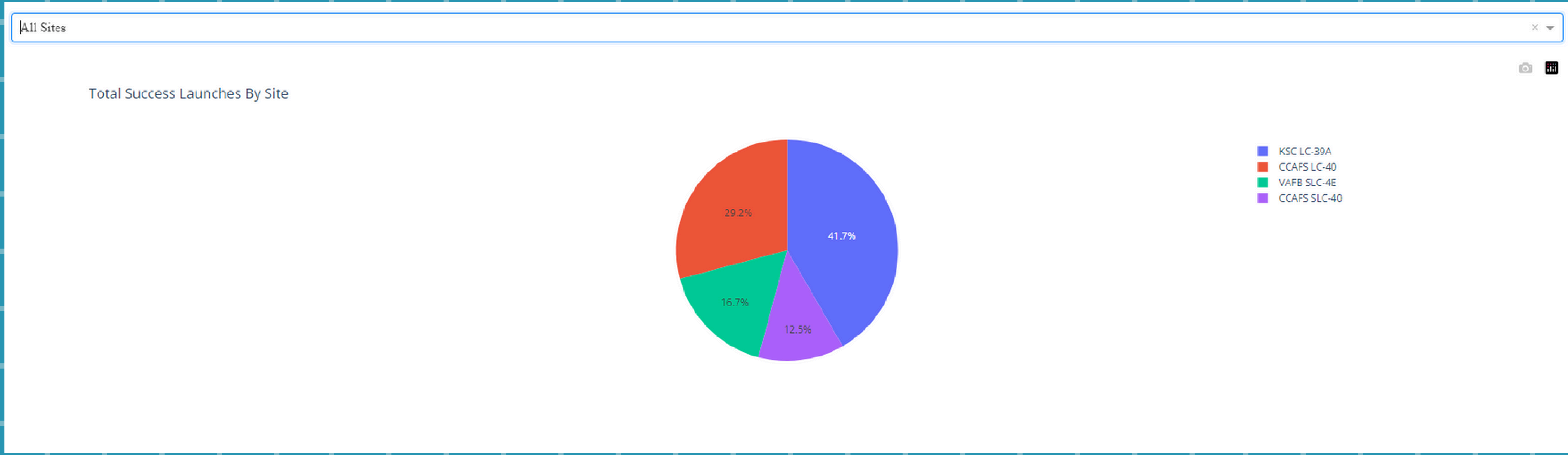
FOLIUM 03



Findings:

1. **Proximity to Highway**
    - **Distance:** The launch point is located 0.59 km from the nearest highway.
    - **Implications:**
        - **Safety and Regulations:** This proximity ensures quick emergency response times and compliance with safety regulations.
        - **Logistics and Accessibility:** Easy access for transportation and logistics, facilitating the movement of equipment and personnel.
2. **Proximity to the Coast**
    - **Distance:** The launch point is 0.87 km from the coast.
    - **Implications:**
        - **Environmental Impact:** Monitoring the potential impact on marine life and coastal ecosystems is crucial.
        - **Noise and Vibration:** Assessing the effects of noise and vibrations on coastal areas to mitigate any negative impacts.

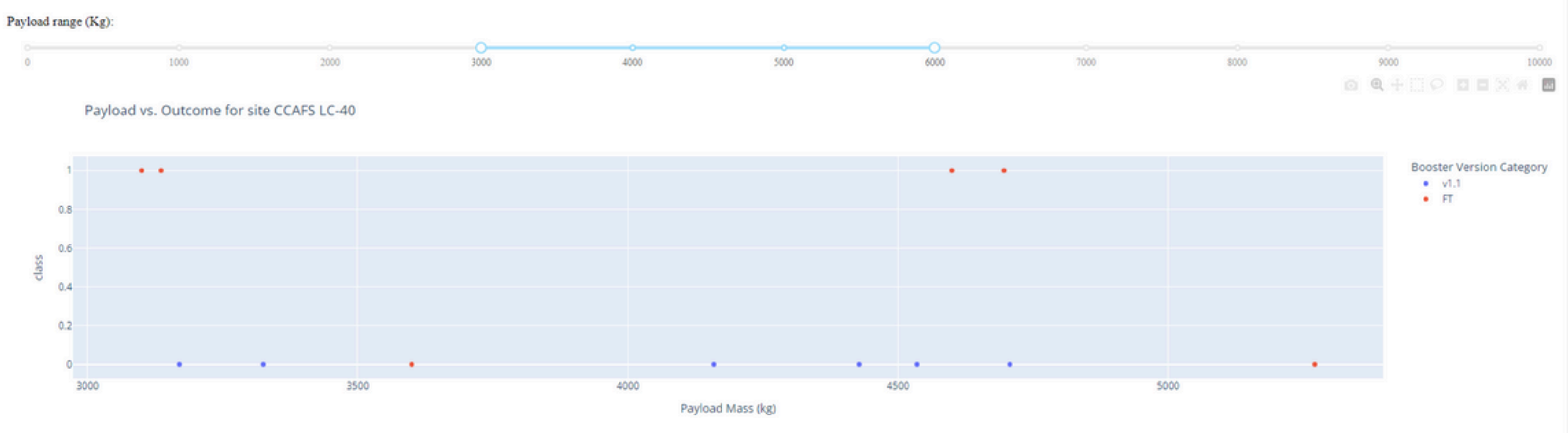
PLOTLY 01



Findings:

1. **Dominant Launch Site:** The site KSC LC-39A has the highest success rate, accounting for 41.7% of the total successful launches. This indicates that KSC LC-39A is the most reliable or frequently used site for successful launches.
2. **Other Sites:** The remaining sites have significantly lower success rates:
   - CCAFS LC-40: 29.2%
   - VAFB SLC-4E: 16.7%
   - CCAFS SLC-40: 12.5%
3. **Performance Disparity:** There is a disparity in performance among the launch sites. KSC LC-39A and CCAFS LC-40 success rate are substantially higher than the others, suggesting that resources or conditions at CKD-347 might be more favorable for successful launches.
4. **Focus Areas:** For improving overall success rates, it might be beneficial to analyze what makes KSC LC-39A and CCAFS LC-40 so successful and apply those insights to the other sites.

PLOTLY 02

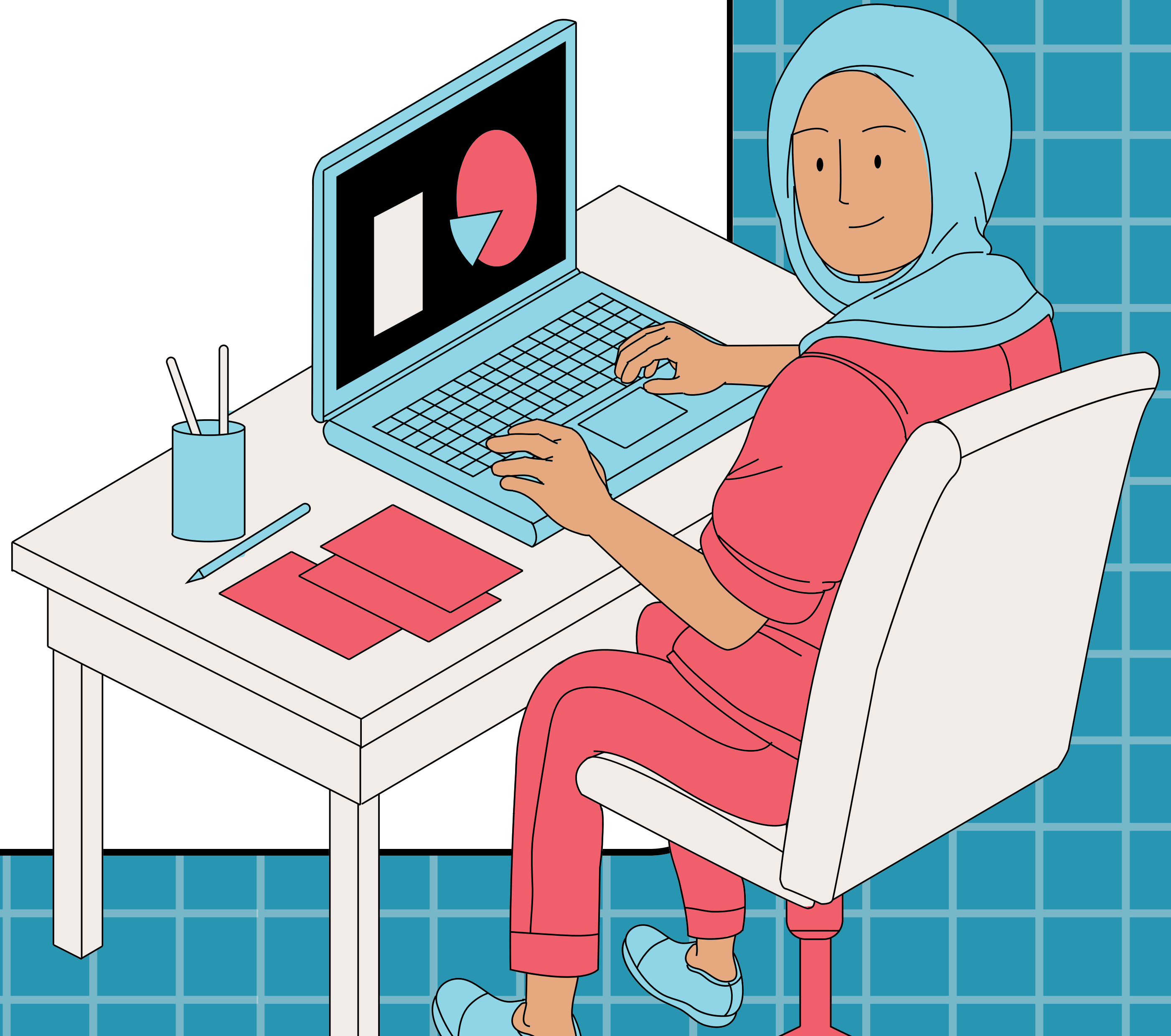


Findings:

1. **The charts show that payloads between 1800 and 3200 kg have the highest success rate.**
2. **The FT (Full Thrust) boosters are the one with the highest success rate**

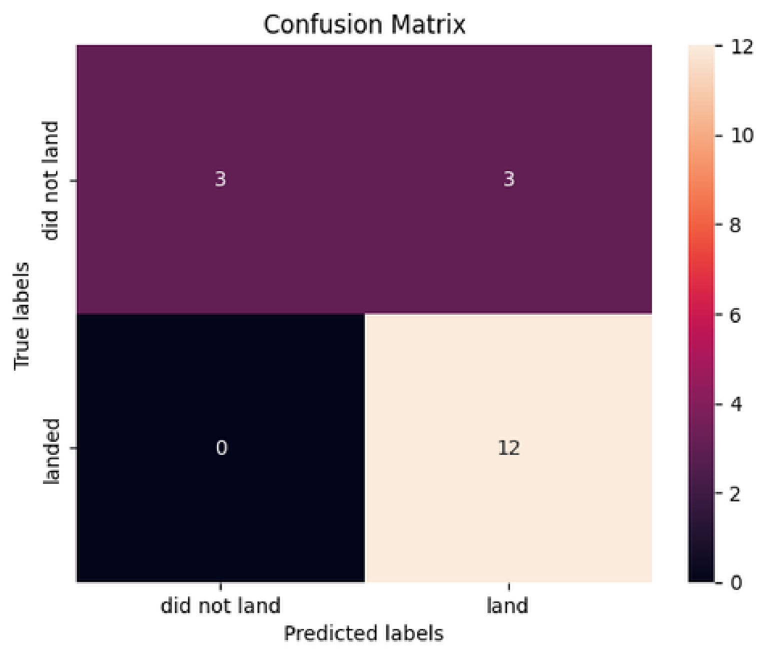
PREDICTIVE ANALYSIS:

In this section,

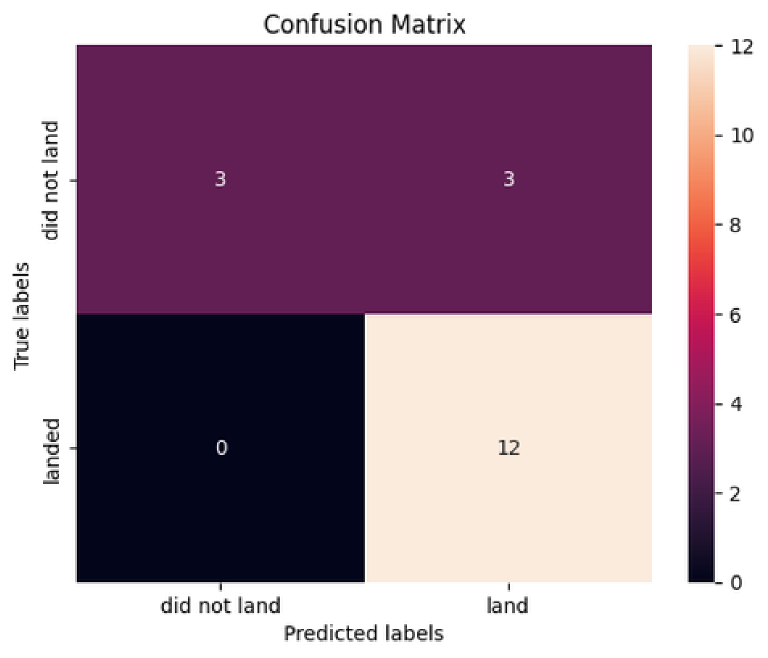


CONFUSION MATRIXES

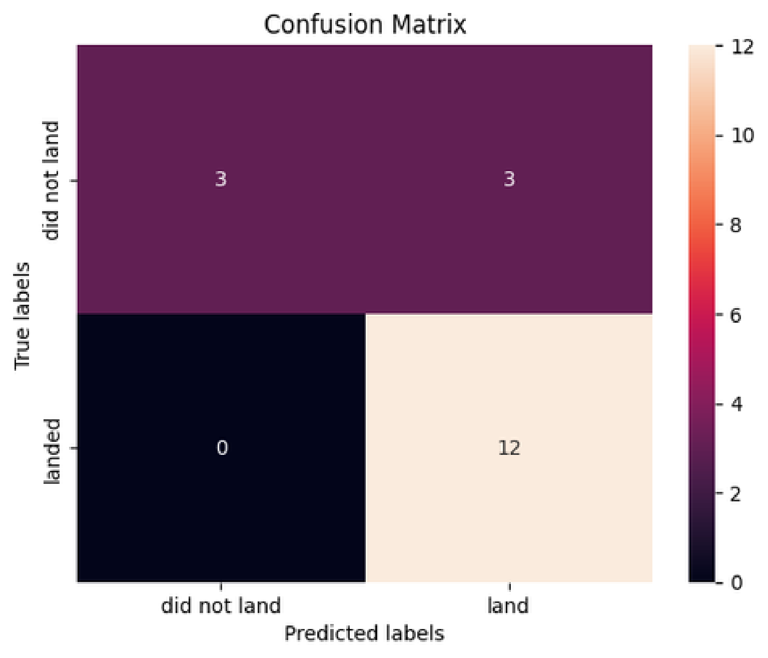
LOGISTIC REGRESION



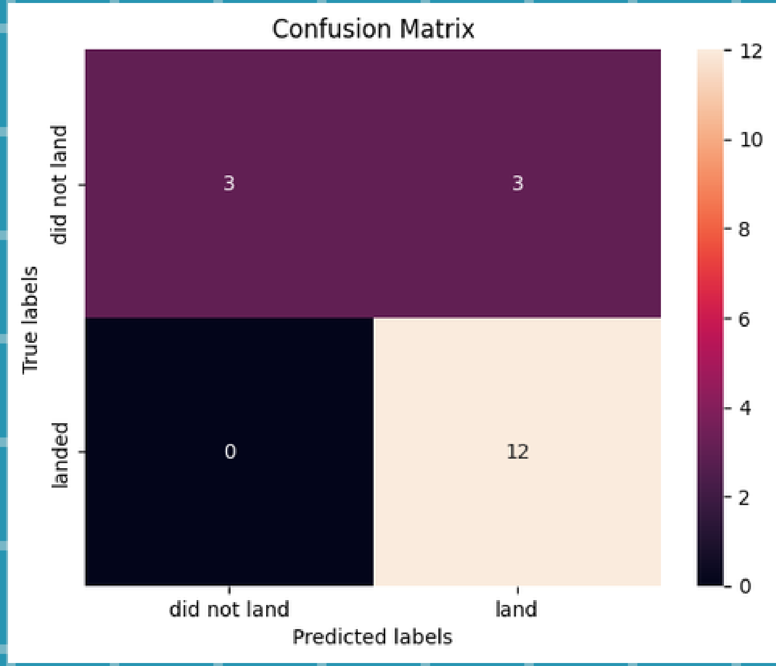
SVM



DECISION TREE



KNN



Insights:

1. **Consistency Across Models:**
   - All models have produced identical confusion matrices, indicating a uniform performance in classifying the landing outcomes.
2. **High Accuracy in Landing Prediction:**
   - The models correctly identified all instances where the first stage successfully landed (12 out of 12).
3. **Moderate Accuracy in Non-Landing Prediction:**
   - For instances where the first stage did not land, the models correctly identified 3 out of 6 cases, resulting in 3 false positives (cases where the model predicted a landing, but there was none).

BEST METHOD

	Validation Accuracy	Test Accuracy
Logistic Regression	0.84	0.83
Support Vector Machine	0.84	0.83
Decision Tree	0.875	0.83
K Nearest Neighbors	0.84	0.83

Based on the provided data, the best model is the Decision Tree. Here's why:

- **Validation Accuracy:** The Decision Tree has the highest validation accuracy at 0.87, compared to the other models.
- **Test Accuracy:** Although the test accuracy for all models is the same (0.83), the Decision Tree's superior validation accuracy suggests it might perform better on unseen data as well.

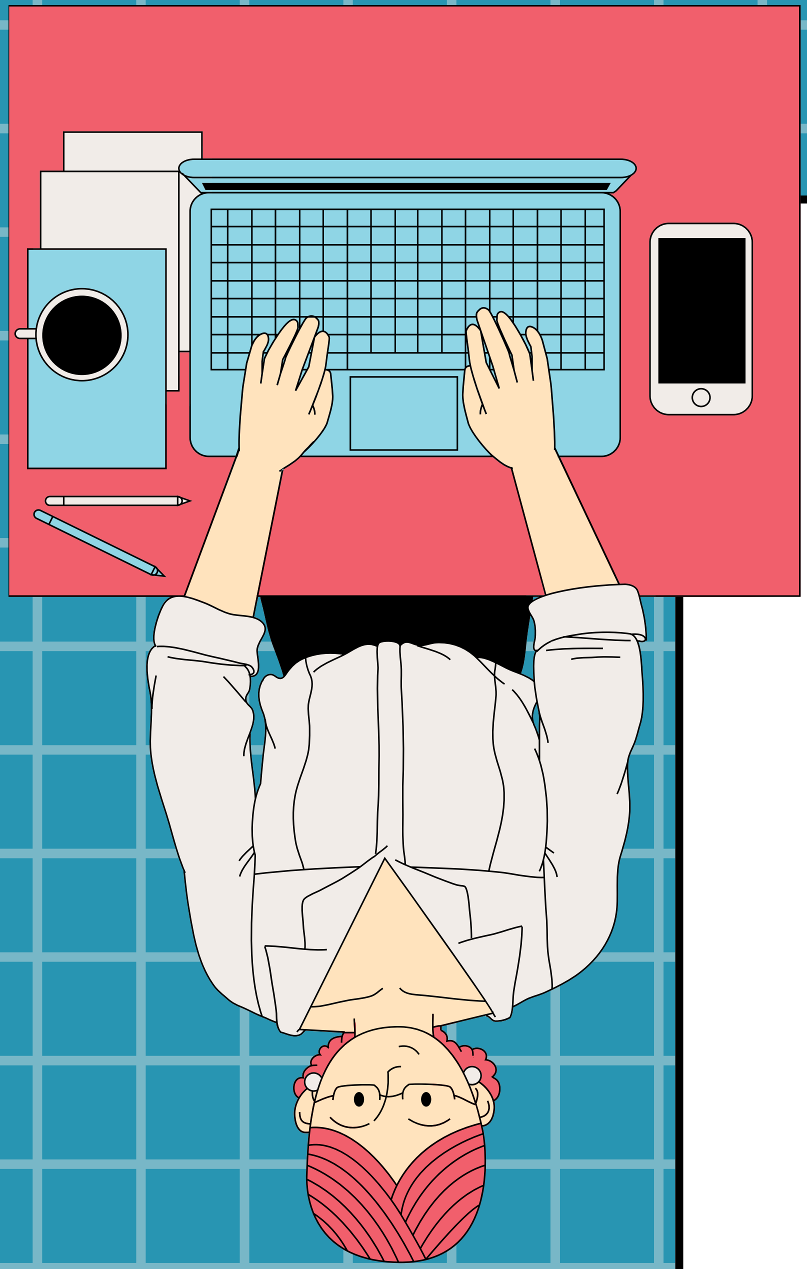
Even though the test accuracy does not differentiate between the models, the higher validation accuracy of the Decision Tree indicates it may be the most effective model among those listed.

CONCLUSION

In conclusion, the analysis of Falcon 9 launch data has provided valuable insights into the factors that influence the success of first stage landings. By leveraging API data collection, web scraping, and rigorous data wrangling processes, we were able to clean and transform the raw data into a usable format for in-depth analysis. The Exploratory Data Analysis (EDA) using various visualization techniques helped uncover significant patterns and correlations, such as the impact of flight number and payload mass on landing success rates.

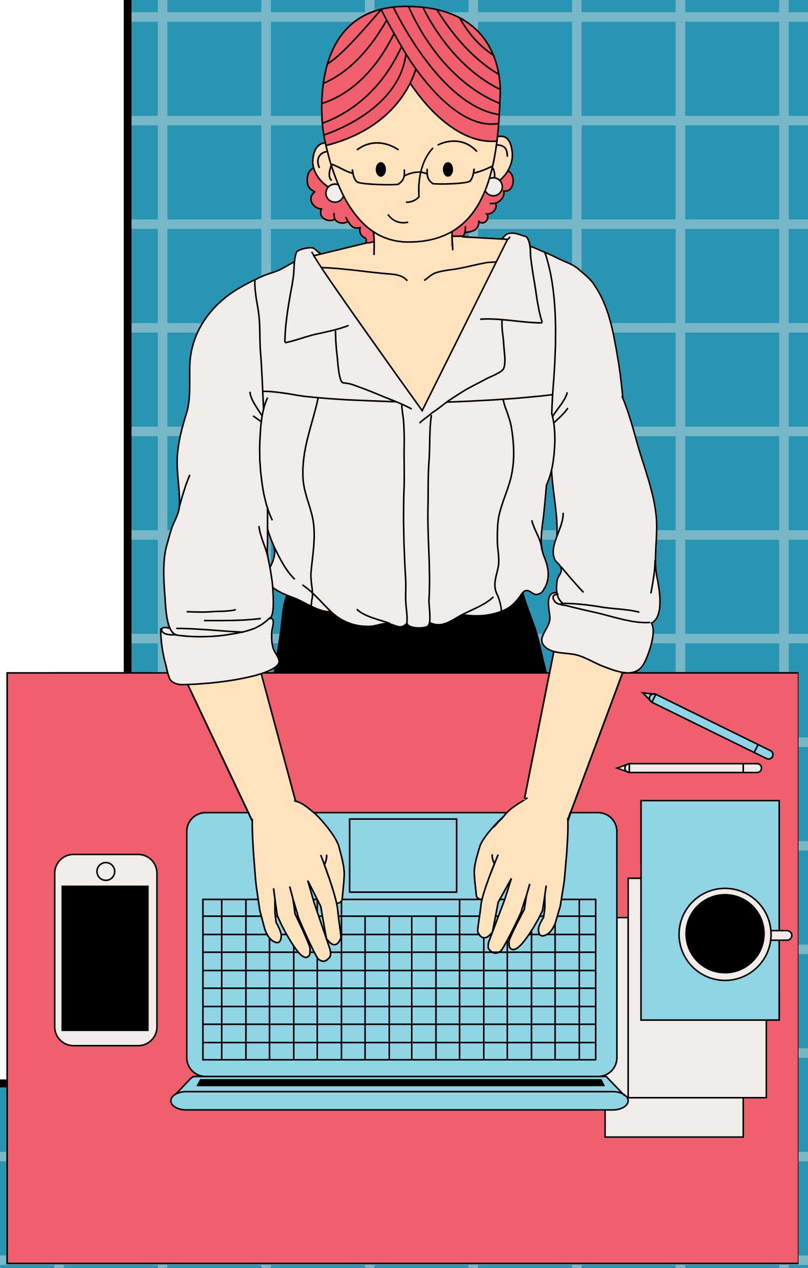
The interactive visual analytics further enhanced our understanding by allowing dynamic exploration of the data, revealing trends and site-specific performance metrics. The predictive modeling phase demonstrated that the Decision Tree model, with its superior validation accuracy, is the most effective method (for this dataset) for predicting landing outcomes among the models tested.

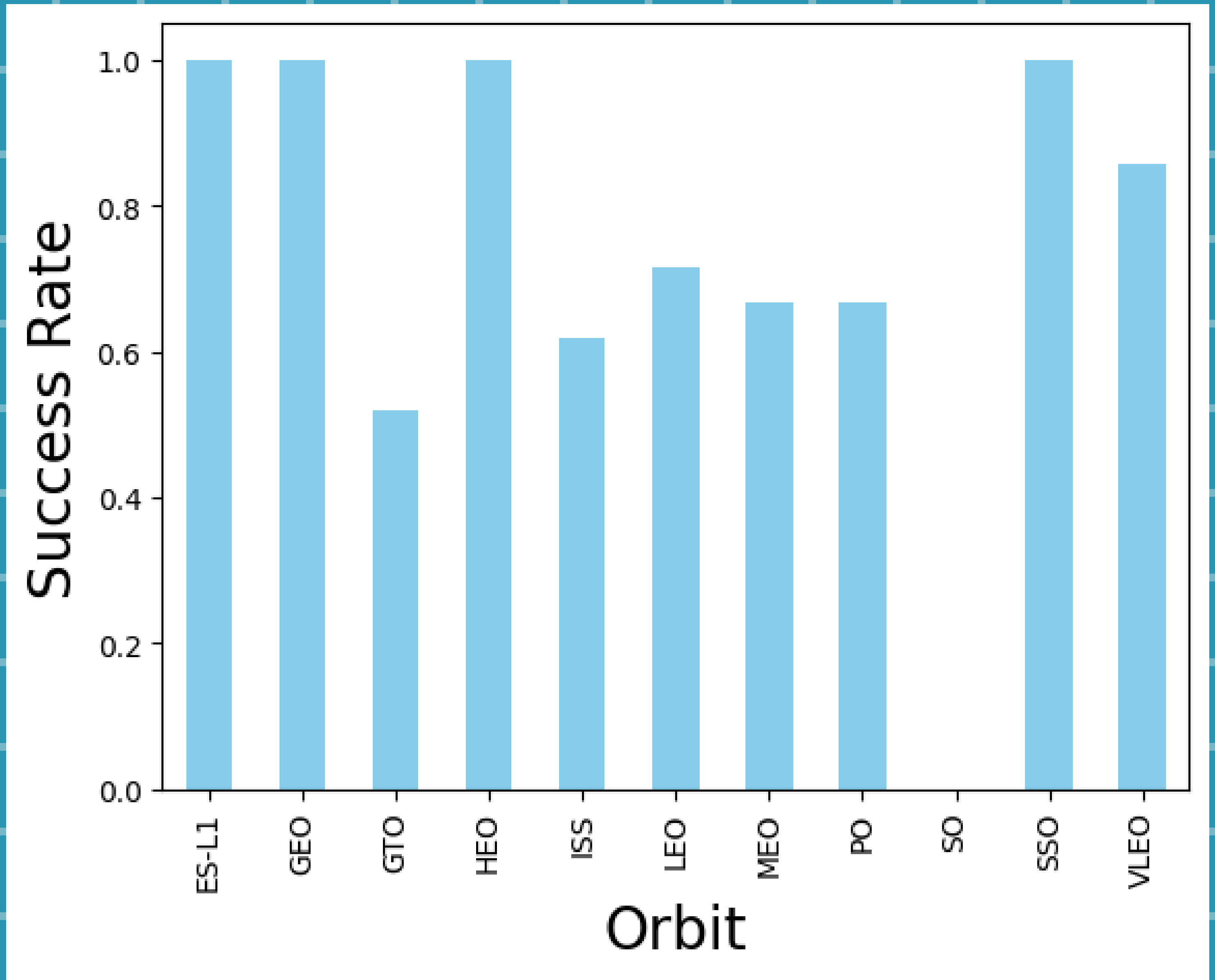
Overall, this project not only highlights the importance of data collection and analysis in understanding and improving space launch operations but also provides a framework for future research and optimization in this field. The findings can help SpaceX and other organizations enhance their launch strategies, reduce costs, and increase the reliability of their missions.

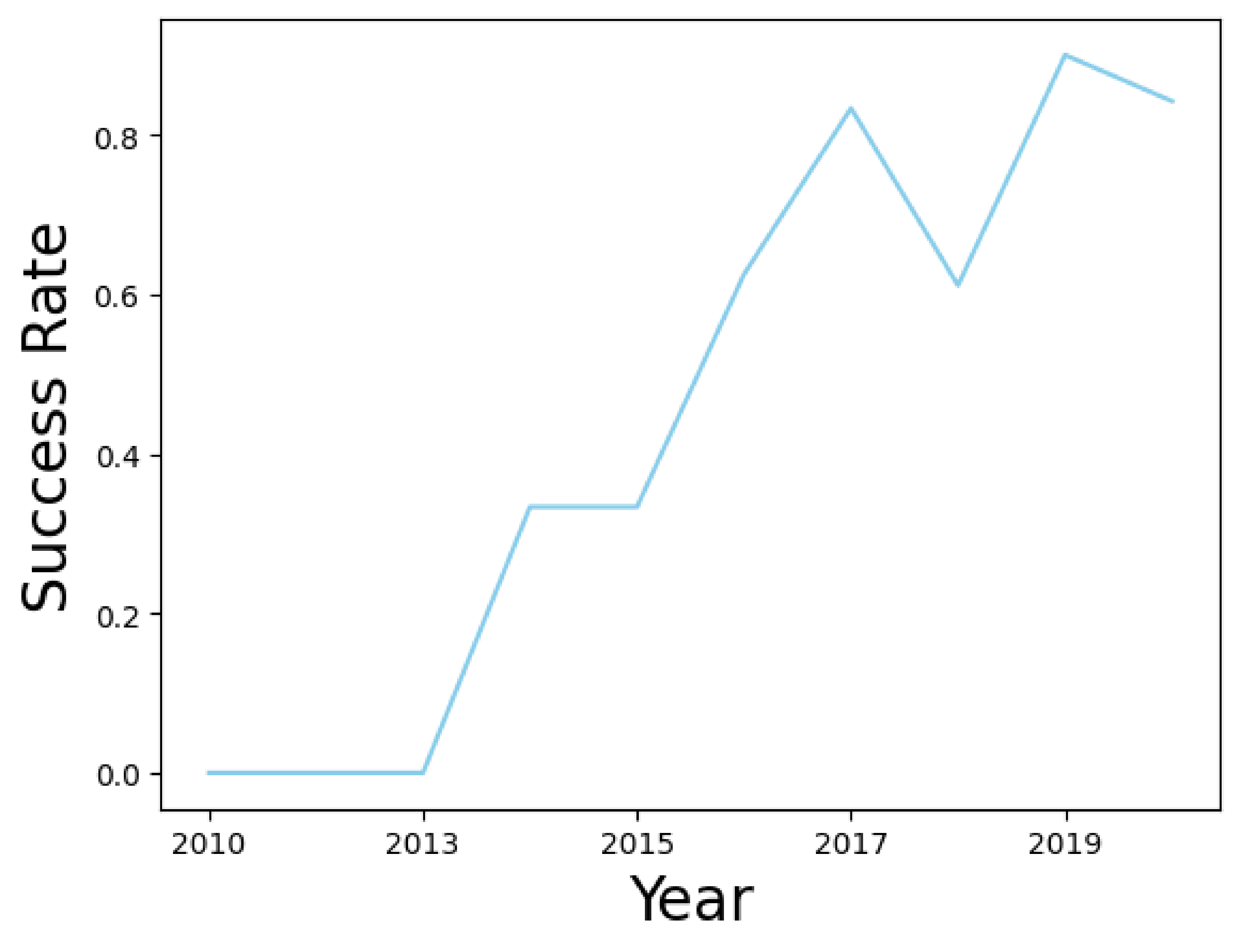


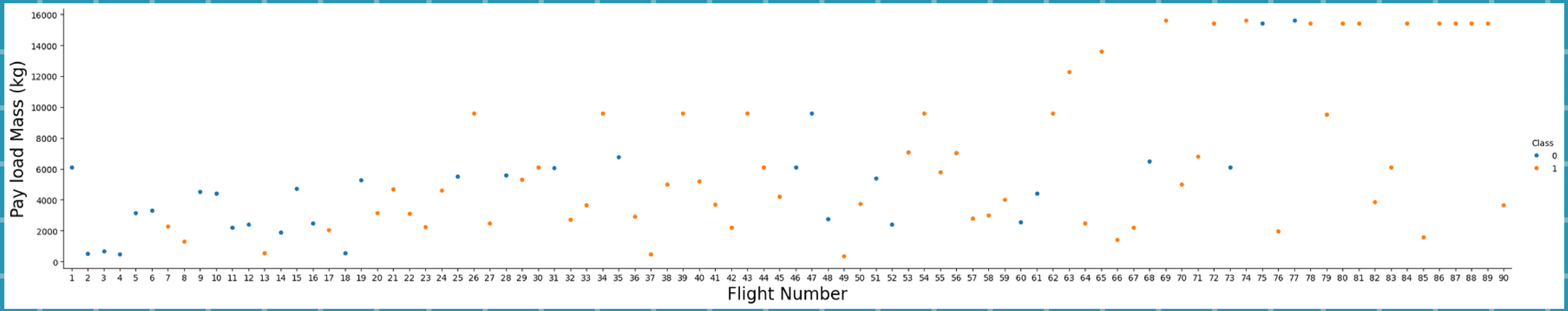
APPENDIX SECTION

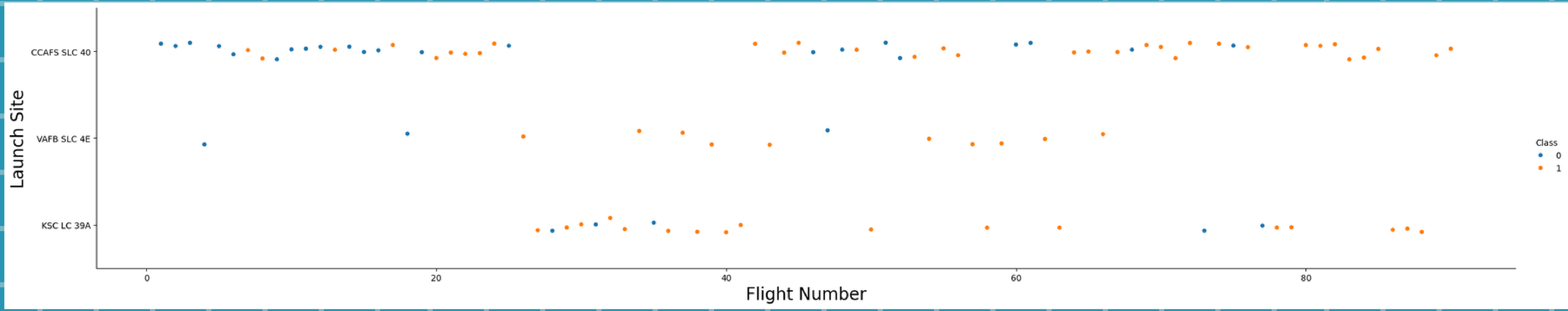
You will find all the charts plotted during the investigation

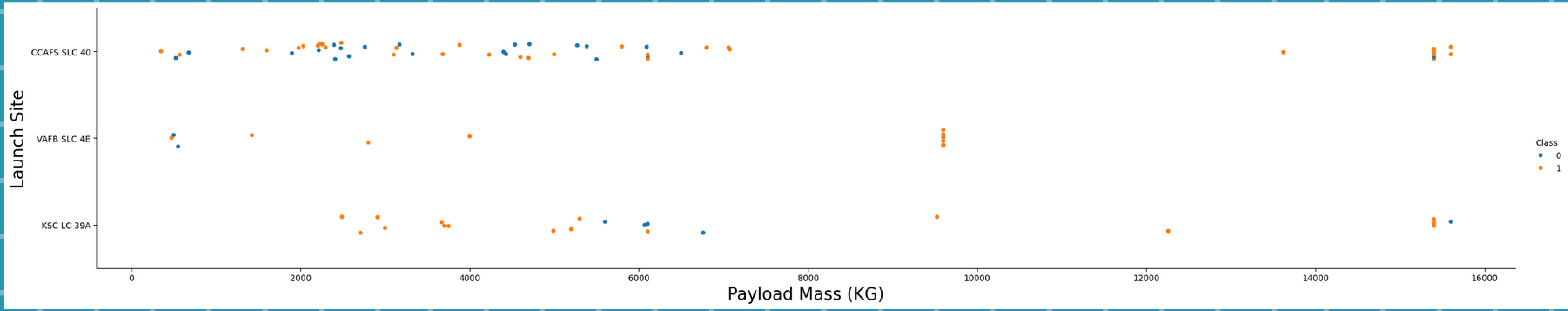


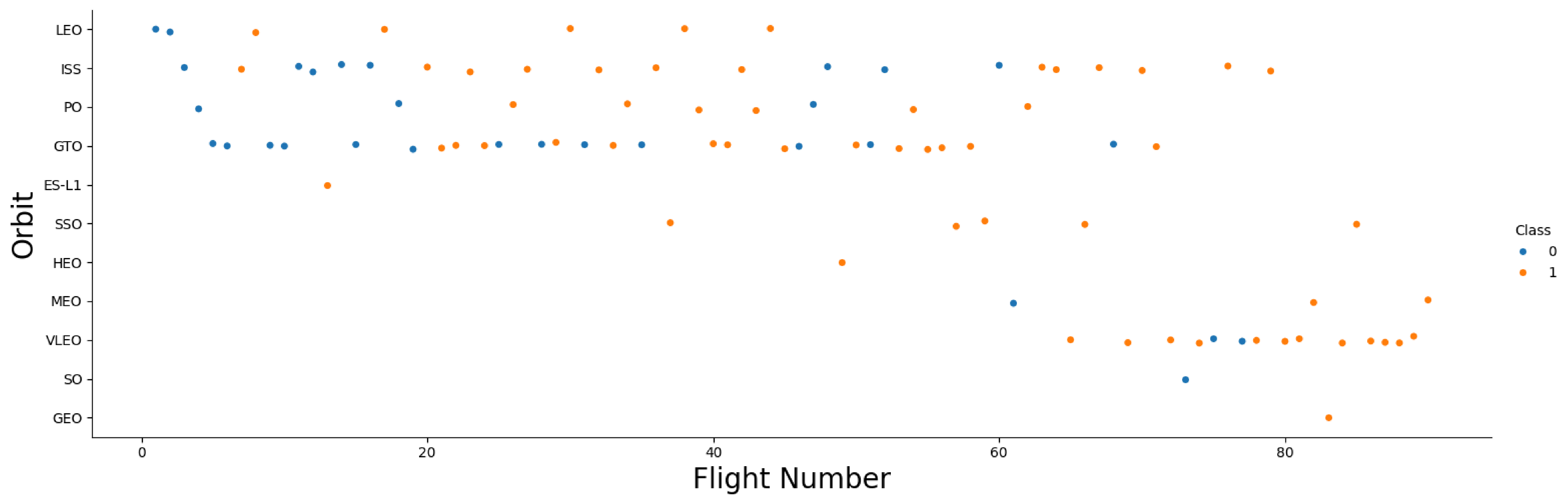

1.0
0.8
0.6
0.4
0.2
0.0
Success Rate
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit


0.8
0.6
0.4
0.2
0.0
Success Rate
2010
2013
2015
2017
2019
Year

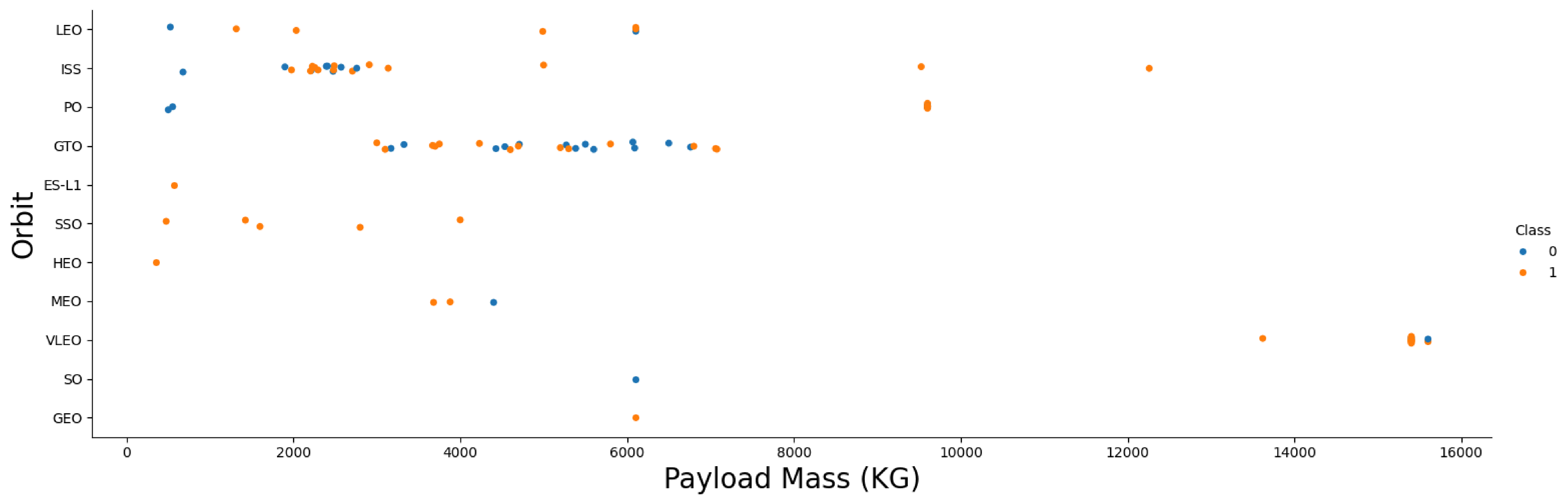

Pay load Mass (kg)
Flight Number
Class
0
1


Launch Site
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Flight Number
Class
0
1


Launch Site
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Payload Mass (KG)
Class
0
1



LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
Orbit
0
20
40
60
80
Flight Number
Class
0
1



LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
Orbit
0
2000
4000
6000
8000
10000
12000
14000
16000
Payload Mass (KG)
Class
0
1

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

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

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```
✓ 0.0s

* 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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as Total FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

Total
45596

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

Average
2928.4

```
%sql SELECT MIN(DATE) AS First_Success FROM SPACEXTABLE WHERE Landing_Outcome ='Success (ground pad)'
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

First_Success
2015-12-22

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

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

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```
✓ 0.0s

* 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

```
%sql SELECT COUNT(Landing_Outcome) AS Total FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success%' OR Landing_Outcome LIKE 'Failure%'
```
✓ 0.0s

* sqlite:///my_data1.db
Done.

Total
71

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

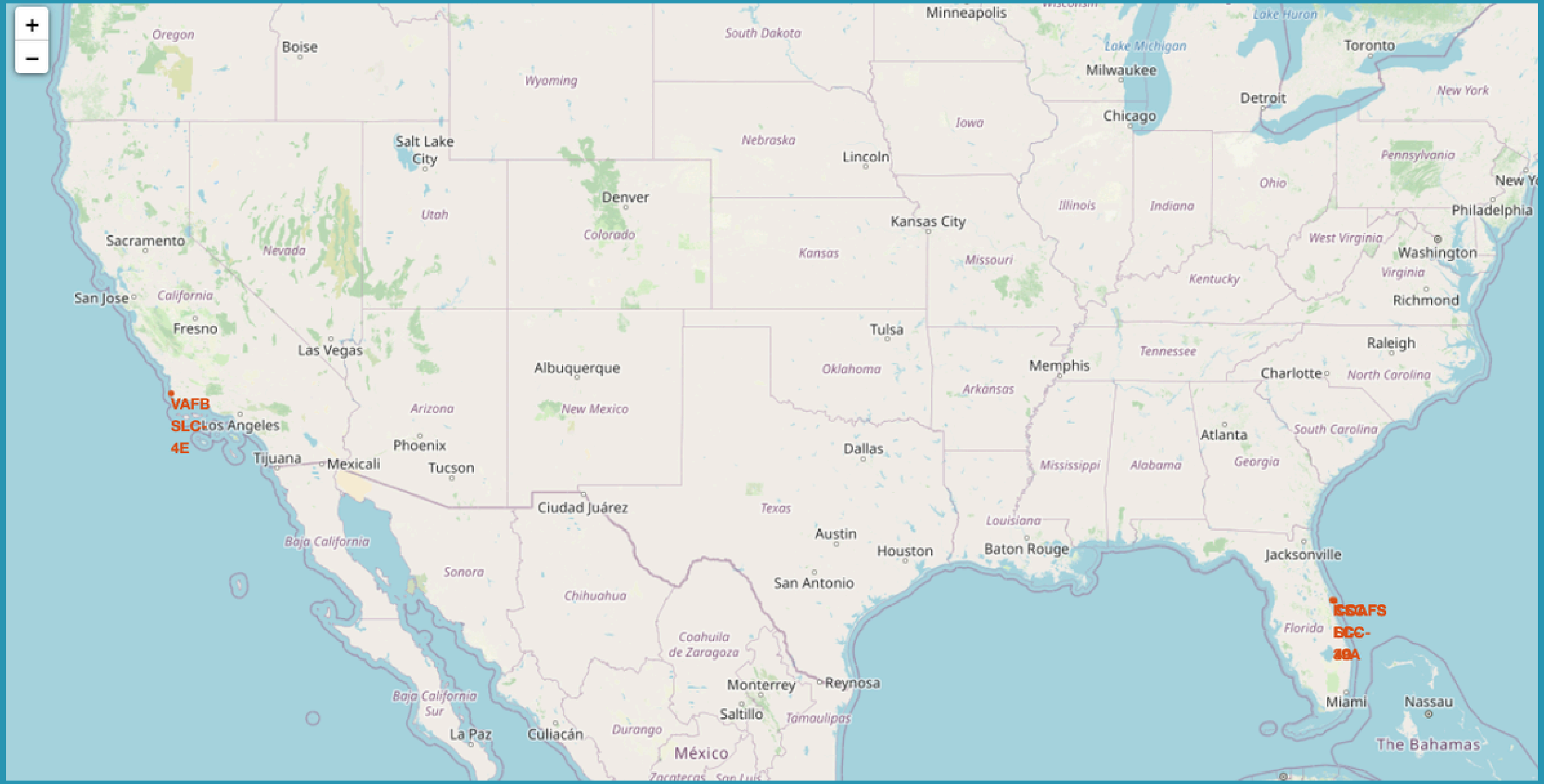
* sqlite:///my_data1.db
Done.

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

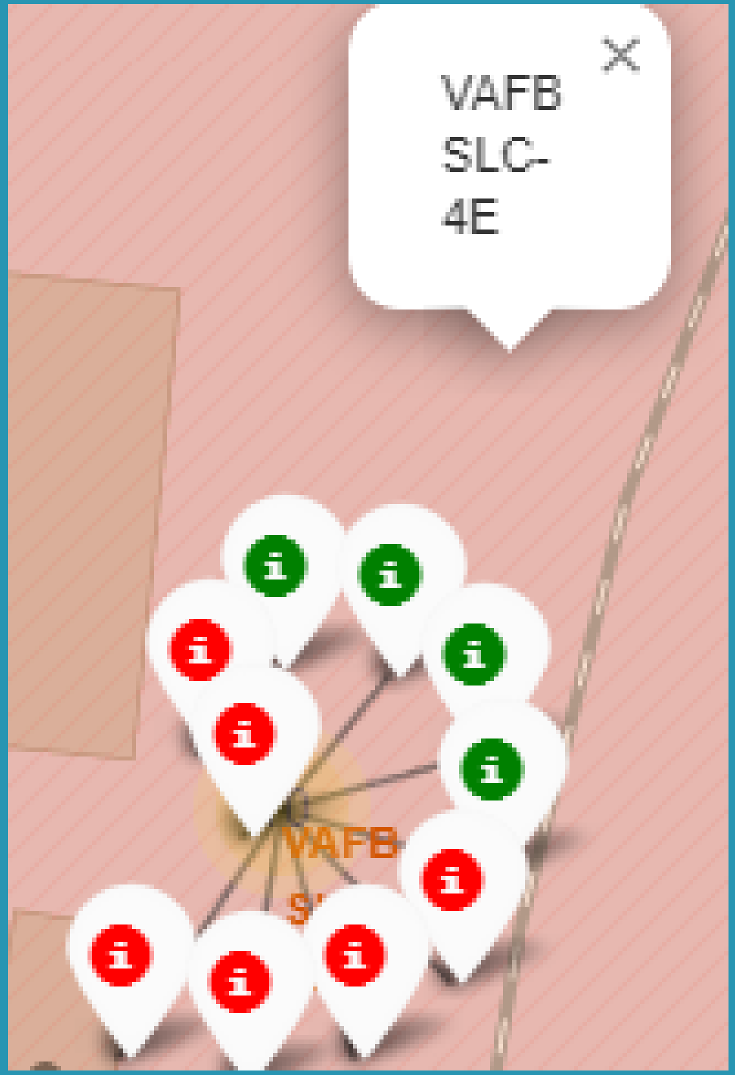
```
%sql SELECT Landing_Outcome, Count(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC
```

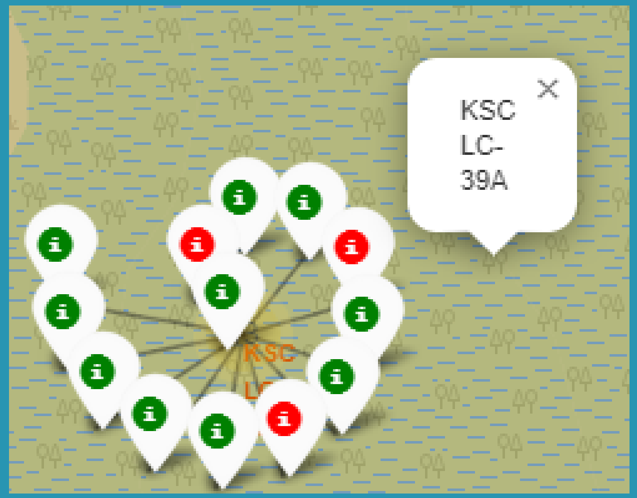
* sqlite:///my_data1.db
Done.

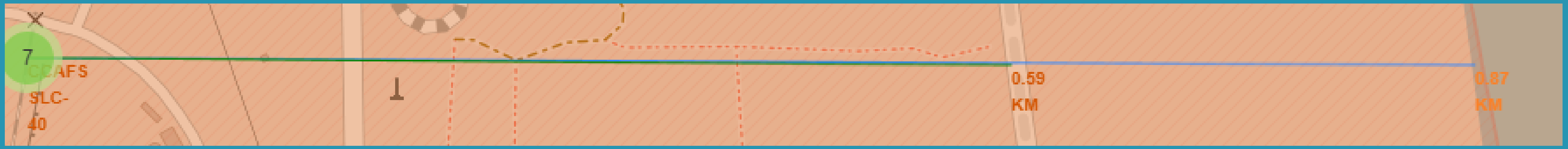
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

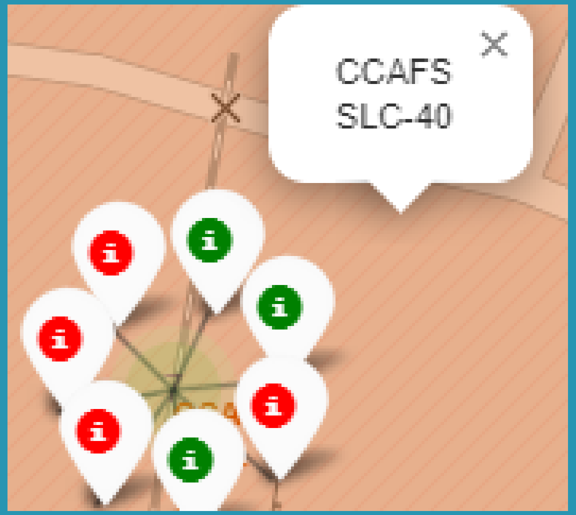

+
−
Oregon
Boise
Wyoming
South Dakota
Minneapolis
Lake Michigan
Lake Huron
Toronto
Milwaukee
Detroit
New York
Chicago
Salt Lake City
Nebraska
Iowa
Lincoln
Pennsylvania
Ohio
Denver
New York
Philadelphia
Sacramento
Nevada
Utah
Colorado
Kansas City
Illinois
Indiana
West Virginia
Washington
Kansas
Missouri
Virginia
Kentucky
San Jose
California
Richmond
Fresno
Tulsa
Raleigh
Las Vegas
Tennessee
Albuquerque
Oklahoma
Memphis
Charlotte
North Carolina
Arkansas
VAFB SLC-4E
Los Angeles
Arizona
New Mexico
South Carolina
Atlanta
Phoenix
Dallas
Tijuana
Mexicali
Tucson
Mississippi
Alabama
Georgia
Ciudad Juárez
Texas
Louisiana
Austin
Baton Rouge
Jacksonville
Houston
Baja California
Sonora
Chihuahua
San Antonio
Florida
Coahuila de Zaragoza
Monterrey
Reynosa
Miami
Nassau
Baja California Sur
Saltillo
Tamaulipas
La Paz
Culiacán
Durango
México
The Bahamas

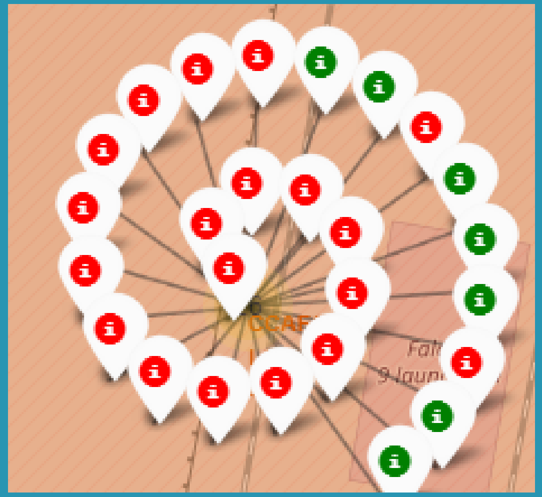

VAFB SLC-4E
VAFB

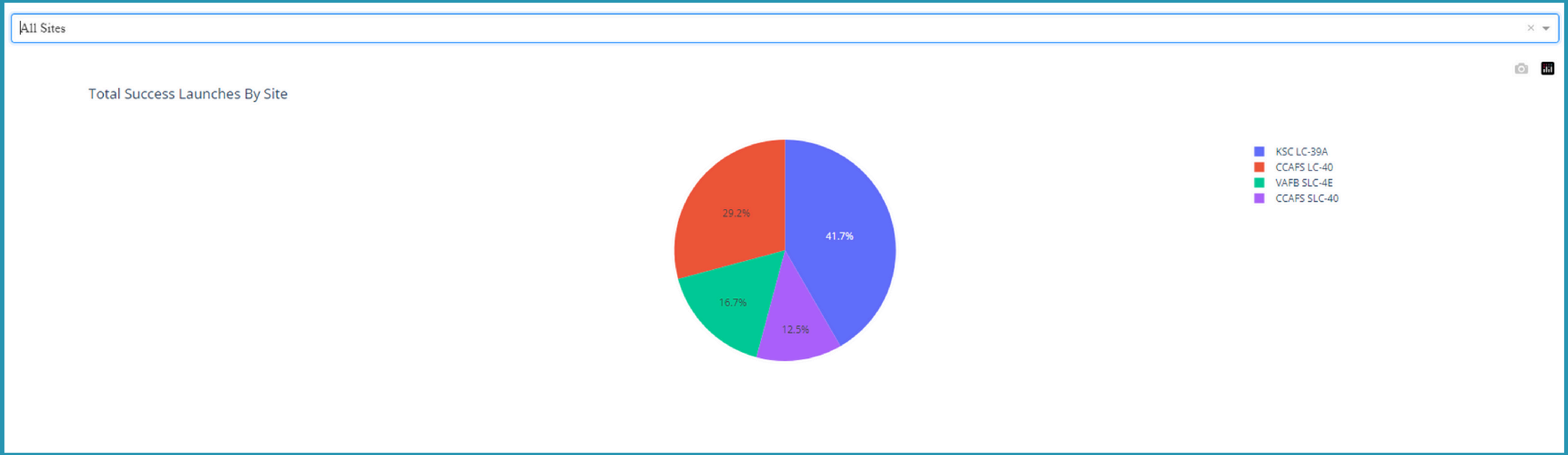

VAFB SLC-4E
VAFB


KSC LC-39A

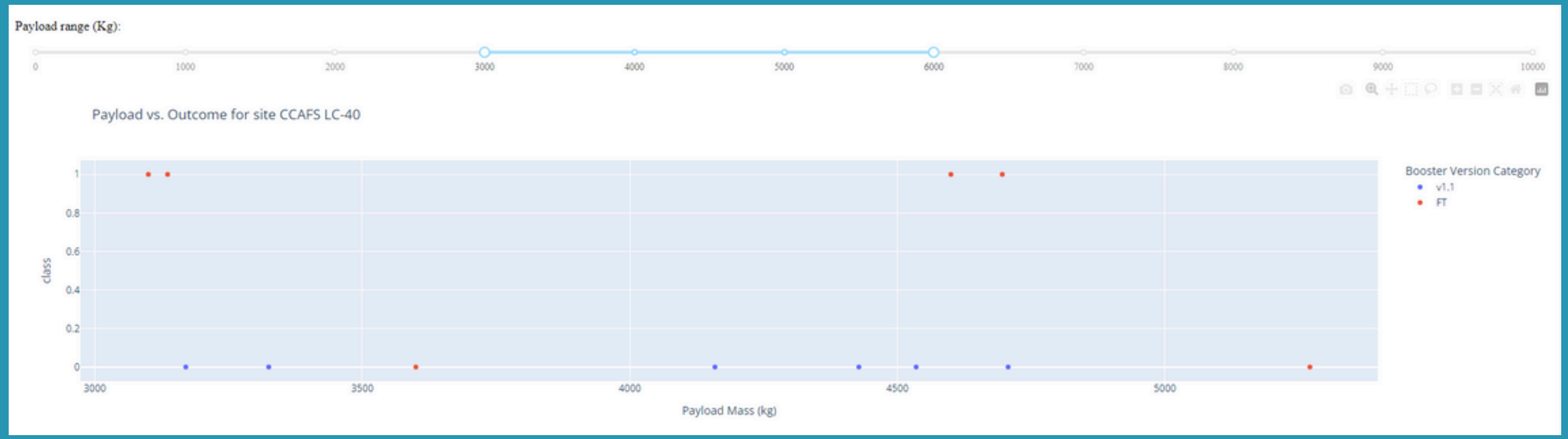

7
CCAFS SLC-40
0.59 KM
0.87 KM

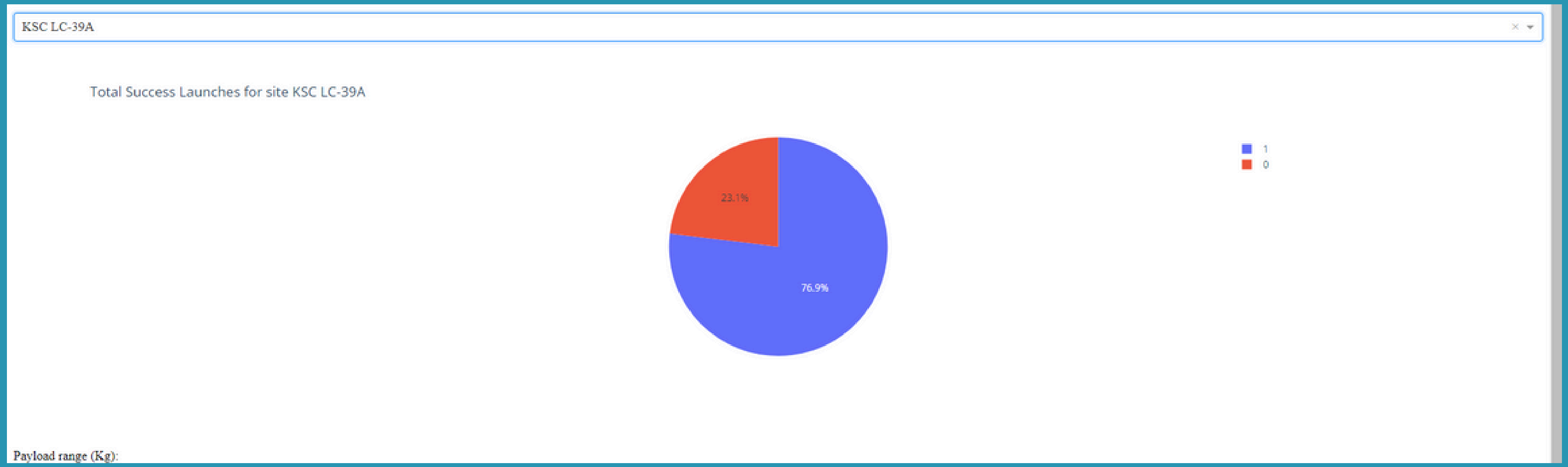

CCAFS SLC-40

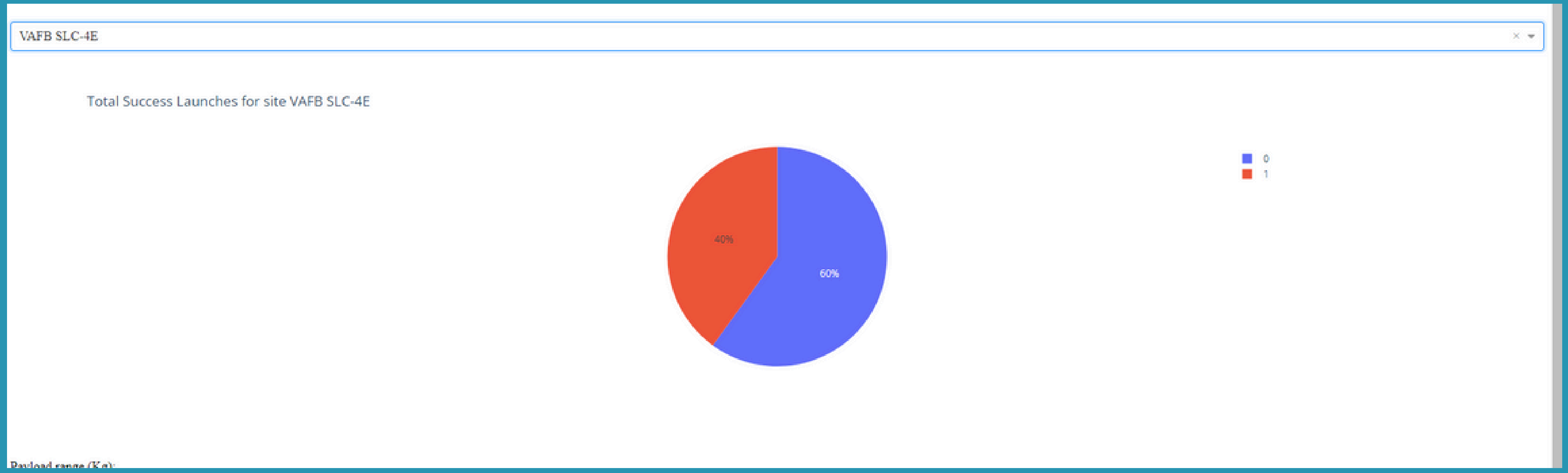


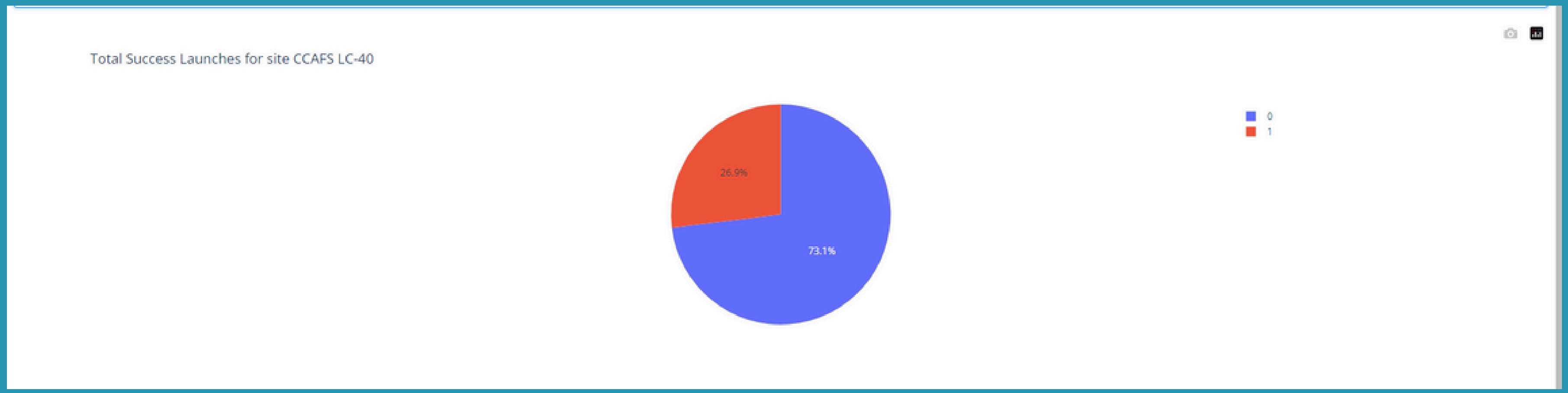

All Sites
Total Success Launches By Site
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40

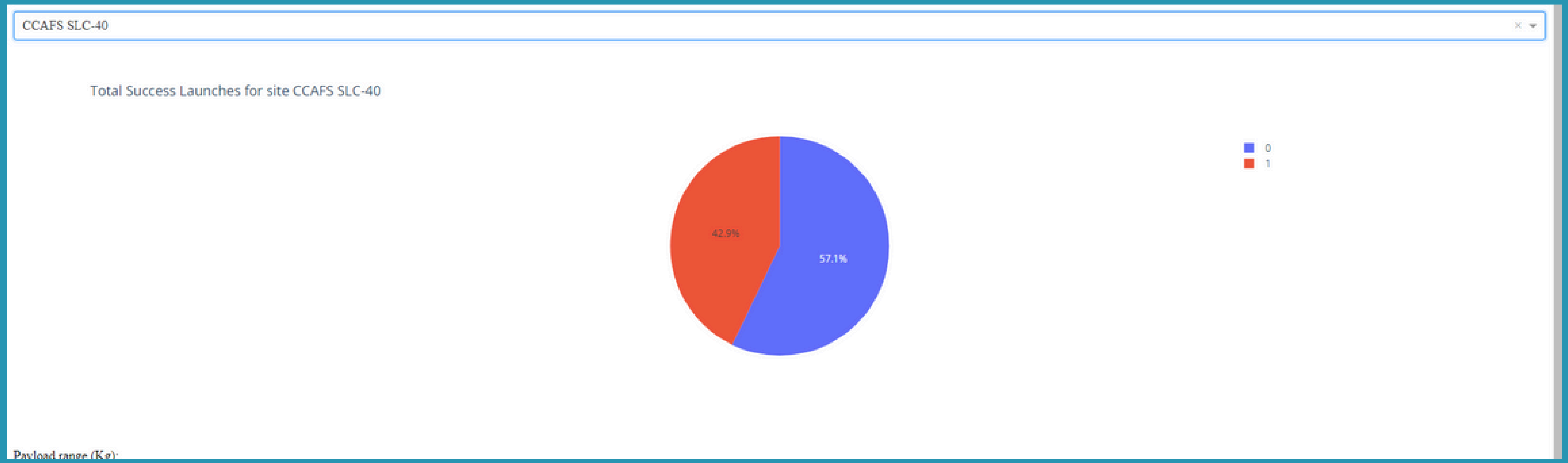

Payload range (Kg):
Payload vs. Outcome for site CCAFS LC-40
Booster Version Category
v1.0
v1.1
FT
class
Payload Mass (kg)

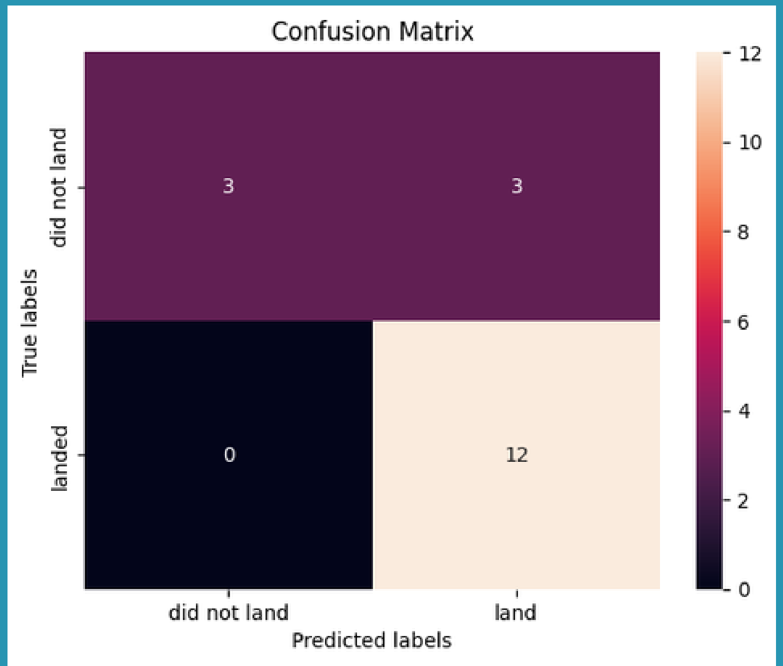

Payload range (Kg):
Payload vs. Outcome for site CCAFS LC-40
Booster Version Category
v1.1
FT
class
Payload Mass (kg)

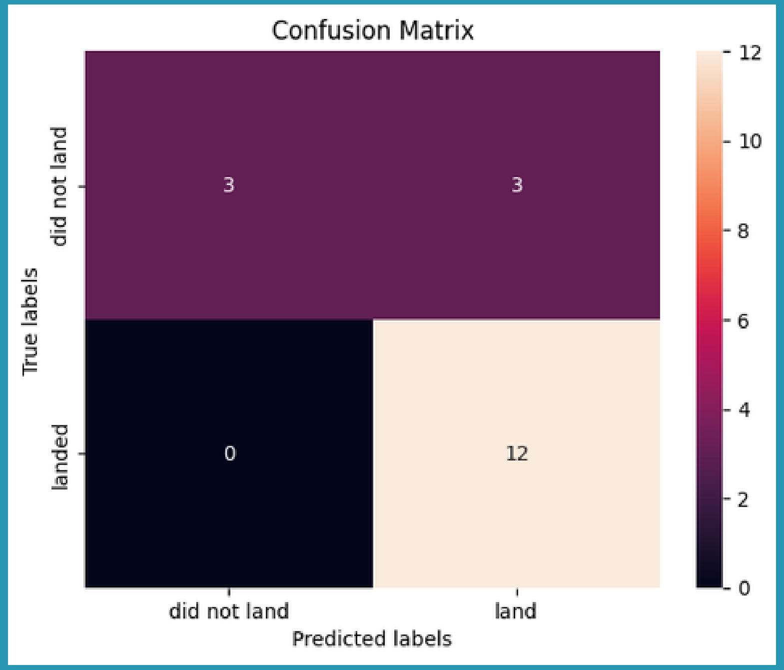

KSC LC-39A
Total Success Launches for site KSC LC-39A
23.1%
76.9%
1
0
Payload range (Kg):

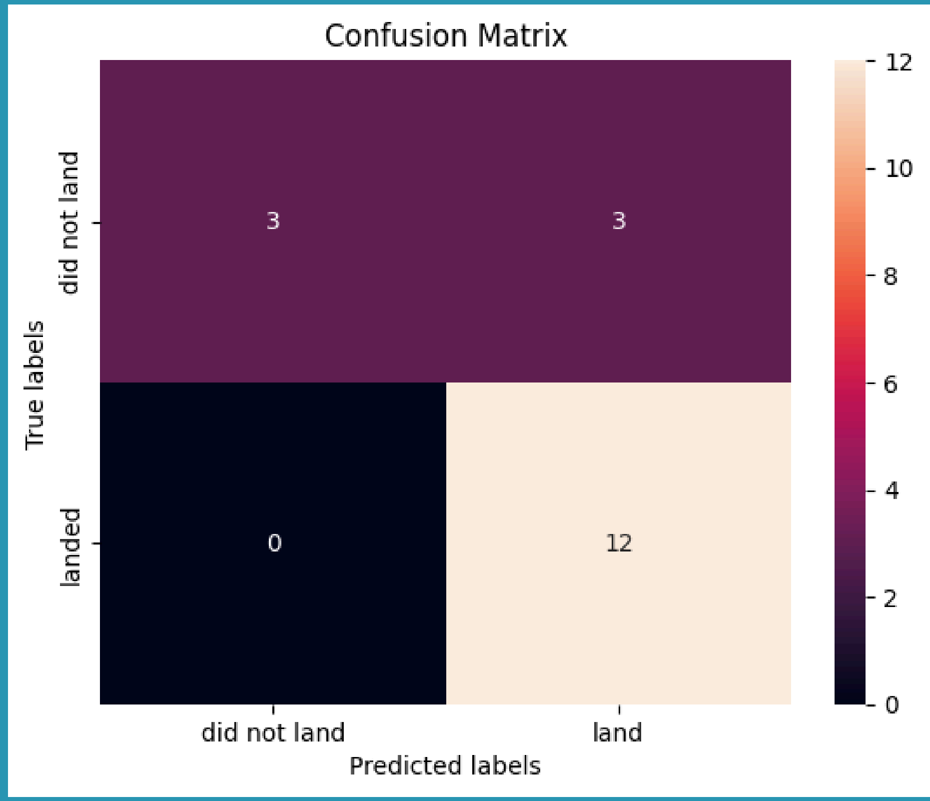

VAFB SLC-4E
Total Success Launches for site VAFB SLC-4E
40%
60%
0
1

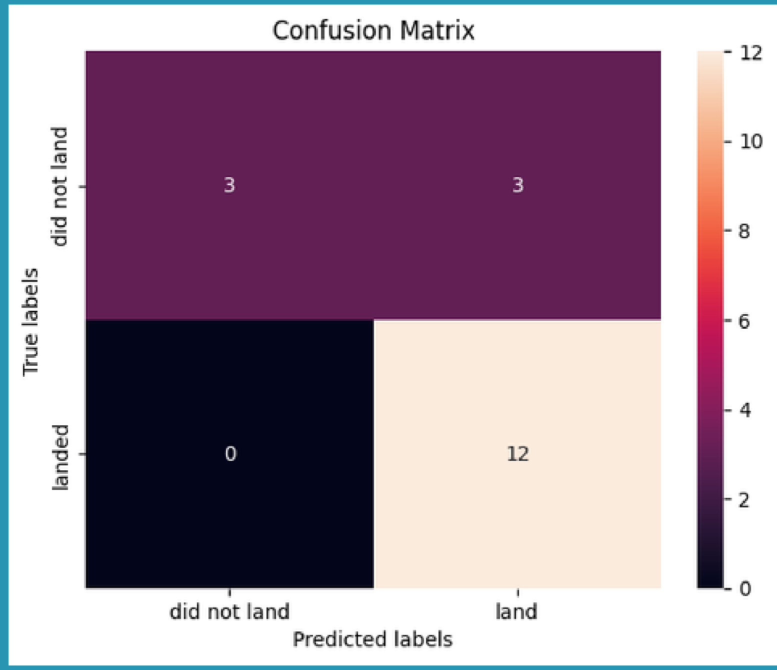

Total Success Launches for site CCAFS LC-40
26.9%
73.1%
0
1


CCAFS SLC-40
Total Success Launches for site CCAFS SLC-40
42.9%
57.1%
0
1


Confusion Matrix
did not land
landed
True labels
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0


Confusion Matrix
did not land
landed
True labels
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0


Confusion Matrix
did not land
landed
True labels
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0


Confusion Matrix
did not land
landed
True labels
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0