IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

CHARLES OKEOGHENE ONOKOHWOMO

APRIL 21ST, 2024

# OUTLINE

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

# EXECUTIVE SUMMARY

This project aimed to develop a Machine Learning Model to predict SpaceX's first-stage rocket launch reusability. Through thorough Data Analysis, including Exploratory Data Analysis and Model Building, we evaluated Logistic Regression, SVM, Decision Tree, and KNN models.

Our analysis revealed that the Decision Tree Model outperformed others, demonstrating superior accuracy on test data. Techniques such as train-test splitting and hyperparameter tuning were employed to optimize model performance.

Confusion matrices provided insights into model performance, with the Decision Tree Model emerging as the most promising. Our findings were summarized in a concise bar chart, illustrating the relative performance of each model.

In conclusion, this project successfully developed a predictive model for SpaceX's rocket launch reusability, with the decision tree model exhibiting exceptional performance.

# INTRODUCTION

## PROJECT BACKGROUND AND CONTEXT

In the era of commercial space exploration, several companies are striving to make space travel more accessible. Among these contenders, including Virgin Galactic, Blue Origin, Rocket Lab, and SpaceX, SpaceX stands out as the most successful. One pivotal factor contributing to SpaceX's success is the remarkable affordability of its rocket launches. For instance, while a SpaceX Falcon 9 launch costs approximately $62 million, launches from other providers can exceed $165 million each, underscoring a significant disparity in pricing.

However, SpaceX's ability to offer such competitive pricing is not merely fortuitous. A key element driving down costs is its innovative approach to reusing the first stages of its rocket launches. By refurbishing and relaunching these stages, SpaceX has effectively revolutionized the economics of space travel, enabling more frequent and cost-effective missions.

# PROBLEM STATEMENT

**IN THE WAKE OF SPACEX DOMINANCE, A NEW PLAYER HAS ENTERED THE SCENE:**

- **SPACEY**, with aspirations to challenge **SPACEX** market leadership. To advance this goal, we are tasked with a crucial mission: to determine the pricing dynamics of **SPACEX** launches and establish correlations between launch prices and the probability of first-stage rocket reuse.

- Recognizing the first-stage reuse is a pivotal factor in **SPACEX** success, our objective is to develop a machine learning model capable of addressing this critical inquiry. By doing so, we aim to provide valuable insights that will inform strategic decisions within the burgeoning commercial space industry.

Section 1

# Methodology

# METHODOLOGY

**Executive Summary**

**Data collection methodology:**

Describe how data was collected

**Perform data wrangling**

Describe how data was processed

**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

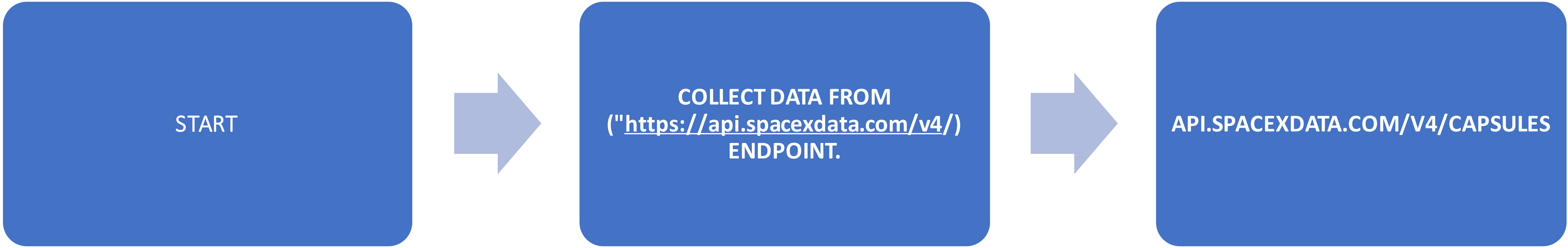
**The data sets were collected through systematic extraction from reliable sources such as official SpaceX records and publicly available databases, Here is a Breakdown:**

- **Identification of Data Sources**: The first step involved identifying reputable sources of SpaceX launch data, including official SpaceX databases, government agencies, and public repositories.
- **Data Gathering**: Once the sources were identified, data gathering procedures were implemented to retrieve relevant information on SpaceX launches. **This involved accessing online databases, scraping web pages, and utilizing APIs to extract structured data.**
- **Data Cleaning and Preprocessing**: After gathering the raw data, thorough cleaning and preprocessing steps were undertaken to ensure data quality and consistency. This included handling missing values, removing duplicates, and standardizing data formats.
- **Data Integration**: Multiple datasets from different sources were integrated and merged to create a comprehensive dataset. This involved aligning data fields, resolving inconsistencies, and consolidating redundant information.
- **Quality Assurance**: Quality assurance measures were applied to validate the accuracy and integrity of the collected data.
- **Documentation**: Detailed documentation of the data collection process was maintained, documenting sources, extraction methods, preprocessing steps, and any transformations applied to the data.

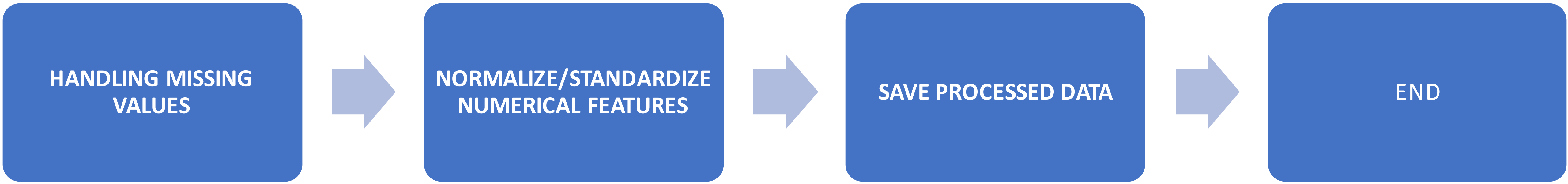
# DATA COLLECTION – SPACEX API

**THIS IS A FLOW CHART OF THE DATA COLLECTION PROCESS THROUGH SPACE X REST API**

[GitHub Url ( Data Collection - SPACE X API)]()




API.SPACEXDATA.COM/v4/CORES
API.SPACEXDATA.com/v4/LAUNCHES/PAST
CHECK RESPONSE FOR VALIDITY : 200
PARSE JSON RESPONSE
EXTRACT RELEVANT DATA
DATA CLEANING AND TRANSFORMATION




[GitHub Url (Data Collection - Space X API)]

# DATA COLLECTION - WEBSCRAPING

**WEBSCRAPING PROCESS USING KEY PHRASES AS FLOW-CHARTS:**

- **Start**
- **Perform HTTP GET request** to retrieve the SpaceX Falcon 9 launch page HTML.
- **Create BeautifulSoup object** from the HTML response.
- **Find all tables** on the webpage.
- **Iterate through tables** to find the Falcon 9 launch table.
- **Extract column names** from the table header.
- **Iterate through table rows** to extract launch data.
- **Parse each row** to extract relevant launch information (Flight No., Date, Time, Version Booster, Launch Site, Payload, Payload mass, Orbit, Customer, Launch outcome, Booster landing).
- **Store extracted data** in a suitable data structure (e.g., dictionary, list of dictionaries).
- **End** ------------

GitHub Url (Webscrapping Data)

# DATA WRANGLING

## DESCRIPTION OF HOW THE DATA WAS PROCESSED

- **Initial Exploration**: We started by exploring the dataset to understand its structure and contents. This involved checking for missing values, identifying the columns relevant to our analysis, and understanding the meaning of different variables.

- **Extracting Relevant Information**: We focused on extracting relevant information from the dataset, such as the number of launches for each site, the occurrence of each orbit, and the mission outcomes of the orbits. This step helped us gain insights into the distribution of launches and mission outcomes.

- **Creating a Landing Outcome Label**: We created a new variable called `landing_class` based on the outcome of each launch. This variable represents the classification of whether the first stage of the rocket landed successfully or not.

- **Assigning Labels to Launch Outcomes**: Using the outcomes of each launch, we assigned labels to represent whether the first stage landed successfully or not. This involved iterating through the outcomes and assigning a value of zero for unsuccessful landings and a value of one for successful landings.

- **Verification and Validation**: Finally, we verified the results by checking the first few rows of the newly created 'Class' column to ensure that the labels were assigned correctly.

**Overall, the data wrangling process involved extracting, transforming, and preparing the data to be suitable for further analysis, particularly for training machine learning models to predict the success of future rocket launches.**

GitHub Url (Data Wrangling)

# EDA WITH DATA VISUALIZATION

## SUMMARY OF THE CHARTS THAT WHERE PLOTTED, ACCOMPANIED BY REASONS FOR THEIR USAGE

- **FlightNumber vs. LaunchSite (Scatter Plot):** This chart was used to visualize the relationship between the flight number and the launch site. It helps in understanding if there's any pattern or correlation between the two variables.

- **Payload vs. LaunchSite (Scatter Plot):** This chart was used to observe the relationship between the payload mass and the launch site. It helps in identifying any potential patterns or trends in payload mass across different launch sites.

- **Success Rate of Each Orbit Type (Bar Chart)**: This chart was used to visualize the success rate of each orbit type. It provides a clear comparison of success rates for different types of orbits.

- **FlightNumber vs. Orbit Type (Scatter Plot)**: This chart was used to explore the relationship between the flight number and the orbit type. It helps in understanding if there's any correlation between the two variables over time.

- **Payload vs. Orbit Type (Scatter Plot):** This chart was used to examine the relationship between the payload mass and the orbit type. It helps in identifying any patterns or trends in payload mass across different orbit types.

- **Launch Success Yearly Trend (Line Chart):** This chart was used to visualize the trend of launch success rates over the years. It helps in understanding how the success rate has evolved over time and if there are any noticeable trends or patterns.

GitHub Url (EDA With Data Visualization)

# EDA with SQL

**SUMMARY OF THE SQL QUERIES CARRIED OUT**

- **Task 1:** Displayed the names of the unique launch sites in the space mission.
- **Task 2:** Displayed 5 records where launch sites begin with the string 'CCA'.
- **Task 3:** Displayed the total payload mass carried by boosters launched by NASA (CRS).
- **Task 4:** Displayed the average payload mass carried by booster version F9 v1.1.
- **Task 5:** Listed the date when the first successful landing outcome in ground pad was achieved.
- **Task 6:** Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.

- **Task 7:** Listed the total number of successful and failure mission outcomes.
- **Task 8:** Listed the names of the booster_versions which have carried the maximum payload mass. Used a subquery.
- **Task 9:** Listed the records displaying the month names, failure landing outcomes in drone ship, booster versions, and launch sites for the months in the year 2015.
- **Task 10:** Ranked 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.

GitHub Url (EDA With SQL)

# BUILD AN INTERACTIVE MAP WITH FOLIUM

## SUMMARY ON THE MAP OBJECTS (MARKERS, CIRCLES, LINES) -

- **Markers**: Markers were employed to pinpoint specific locations of interest on the map, such as launch sites, coastlines, and proximities like cities, railways, and highways. They served as visual identifiers for these locations, aiding in their recognition and interpretation.

- **Circles**: Circles were used to highlight areas of interest, such as the vicinity around launch sites or specific coordinates. They helped visualize spatial relationships and provided a clear indication of the extent or radius of certain features, such as the coverage area around a launch site or the proximity to a coastline.

- **Lines**: PolyLines were drawn to connect different points on the map, such as the launch site to proximities like coastlines, cities, railways, and highways. These lines visually represented spatial connections or distances between locations and allowed for the assessment of proximity and spatial relationships.

## EXPLANATION FOR ADDING OBJECTS -

- **Markers**: Markers were added to identify and label key locations relevant to the analysis, such as launch sites, coastlines, and proximities. They provided visual cues for specific data points and allowed for easy identification and interpretation of locations and attributes on the map.

- **Circles**: Circles were used to highlight and delineate specific areas of interest, such as the coverage area around launch sites or the proximity to certain features like coastlines. They helped visualize spatial extents and provided context for assessing distances and spatial relationships.

- **Lines**: Lines were drawn to represent spatial connections and distances between different points on the map, such as the launch site to proximities like coastlines, cities, railways, and highways. They facilitated the assessment of proximity and spatial relationships, allowing for the analysis of distances and spatial patterns between locations.

GitHub Url (LaunchSite Location Analysis With Folium )

# BUILD A DASHBOARD WITH PLOTLY DASH

## SUMMARY OF THE PLOTS/GRAPHS AND INTERACTIONS ADDED TO THE DASHBOARD:

- **Dropdown Menu:** Added a dropdown menu to select launch sites, including an option for selecting all sites.
- **Pie Chart:** Displayed the total successful launches count for all sites. When a specific launch site is selected, it shows the success vs. failed counts for that site.
- **Range Slider:** Added a slider to select payload range. This allows users to filter launches based on the payload mass range.
- **Scatter Plot:** Created a scatter plot to show the correlation between payload mass and launch success. It dynamically updates based on the selected launch site and payload range.

## EXPLANATION FOR ADDING PLOTS AND INTERACTIONS -

- **Dropdown Menu:** Enables users to select specific launch sites or view data for all sites. This provides flexibility in analyzing launches from individual sites or collectively.

- **Pie Chart:** Offers a visual representation of the overall success rates across different launch sites. It allows users to quickly identify which sites have higher success rates.

- **Range Slider:** Allows users to filter launches based on the payload mass range, providing insights into how payload mass affects launch success.

- **Scatter Plot:** Provides a detailed view of the correlation between payload mass and launch success for each selected launch site. Users can visually analyze how payload mass influences launch outcomes.

GitHub Url (Building Dashboard With Plotly Dash)

# PREDICTIVE ANALYSIS (CLASSIFICATION)

**SUMMARY OF THE PROCESS INVOLVED IN BUILDING, EVALUATING, IMPROVING AND FINDING THE BEST PERFORMING CLASSIFICATION MODEL:**

- **Data Preparation:** The process began with exploratory data analysis (EDA) to understand the dataset's characteristics and distributions. We created the target variable by converting the 'Class' column into a NumPy array, which was assigned to the variable Y, ensuring it remained a Pandas series.

- **Data Standardization:** The features in the dataset were standardized to ensure they had a mean of 0 and a standard deviation of 1, which helps in improving the performance of some machine learning algorithms.

- **Train-Test Split:** The dataset was split into training and test sets using the **`train_test_split`** function, with 20% of the data allocated for testing. The split data was assigned to variables **`X_train`**, **`X_test`**, **`Y_train`**, and **`Y_test`**.

- **Model Training and Hyperparameter Tuning:** Three different Classification Algorithms were used: **Logistic Regression**, **Support Vector Machine (SVM)**, and **Decision trees**. **GridSearchCV** was employed to perform hyperparameter tuning for each algorithm, allowing us to find the optimal parameters. Additionally, **K-Nearest Neighbors (KNN)** algorithm was used with its respective hyperparameters.

- **Model Evaluation:** The performance of each model was evaluated using Accuracy metrics on the test data, calculated using the `score` method. **Confusion matrices** were generated to visualize the model's performance in terms of **True positives, True negatives, False positives, and False negatives.**

- **Comparison and Selection:** After evaluating each model's performance, the model with the highest accuracy on the test data was identified as the best performing model. The best model was selected based on its **Accuracy score** and performance in the **Confusion matrix**( Decision Tree : 0.9444).

GitHub Url (Machine Learning - Predictive Analysis)

# RESULTS

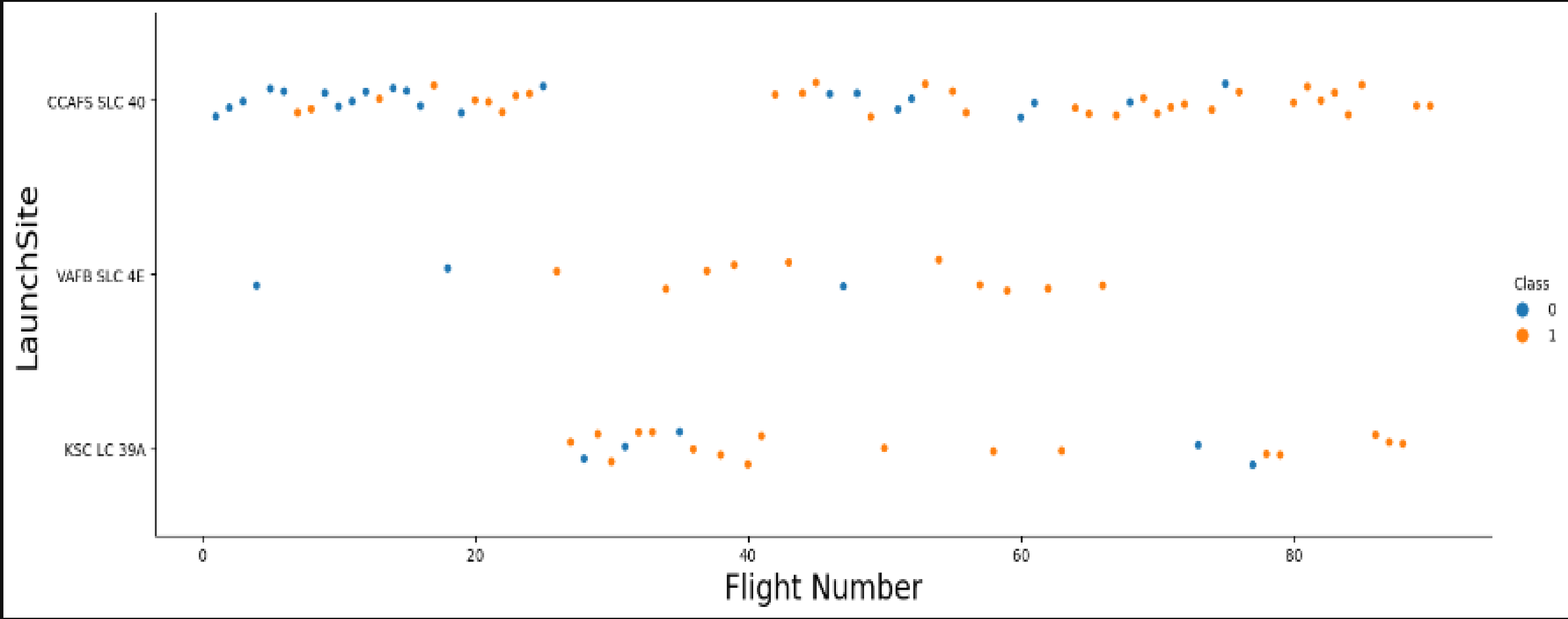
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

# Insights drawn from EDA

# FLIGHT NUMBER VS. LAUNCH SITE

## SCATTER PLOT OF FLIGHT NUMBER VS LAUNCH SITE



## RELATIONSHIP BETWEEN THE FLIGHT NUMBER VS LAUNCH SITE

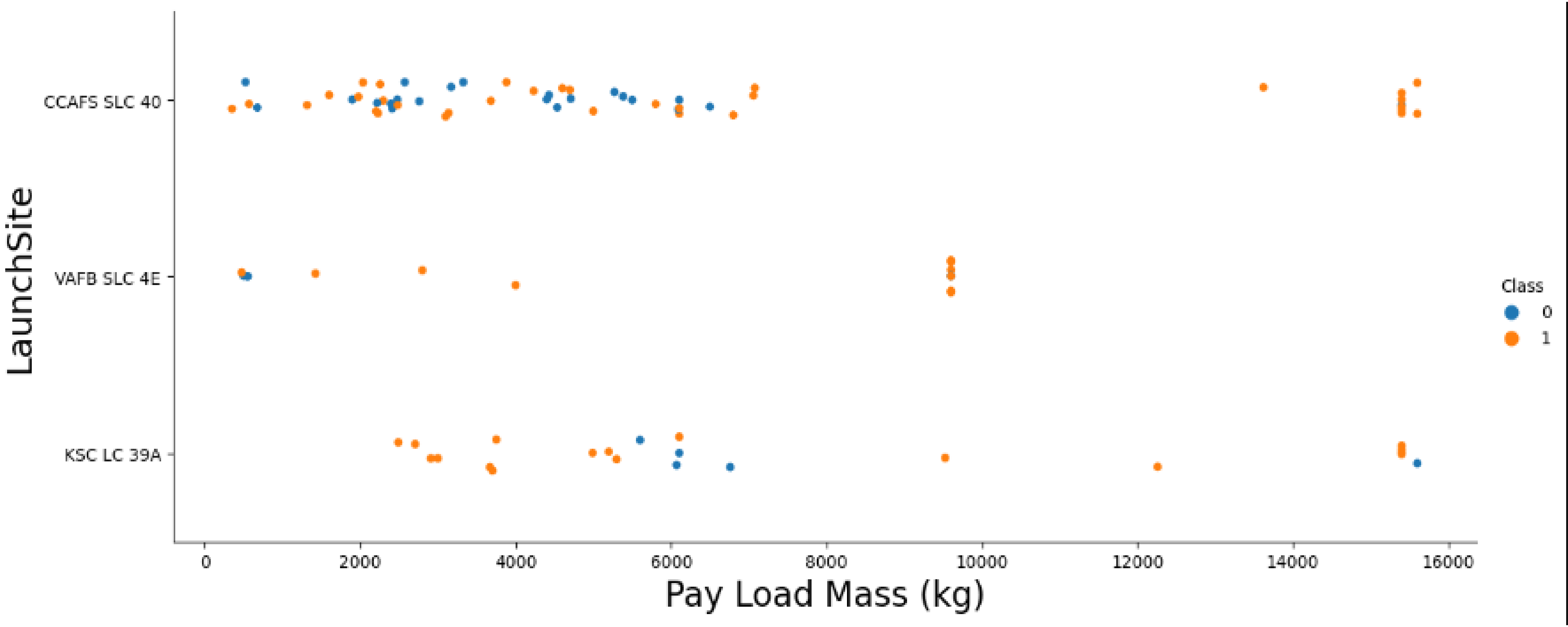
**From the screenshot provided**, it's evident that there is a correlation between the Flight Number and the Probability of a Successful First-Stage Rocket Landing at each Launch site.

**While the data may not be flawless, a comparison of successful and unsuccessful landings reveals a clear trend:** as the flight number increases, the likelihood of a successful landing also increases. Although there are instances of unsuccessful landings, they are outweighed by successful ones as the flight number rises.

Notably, **Launch Site VAFBSLC 4E** demonstrates a consistent level of successful launches across increased flight numbers.

# PAYLOAD VS. LAUNCH SITE

## SCATTER PLOT OF PAYLOAD VS LAUNCH SITE

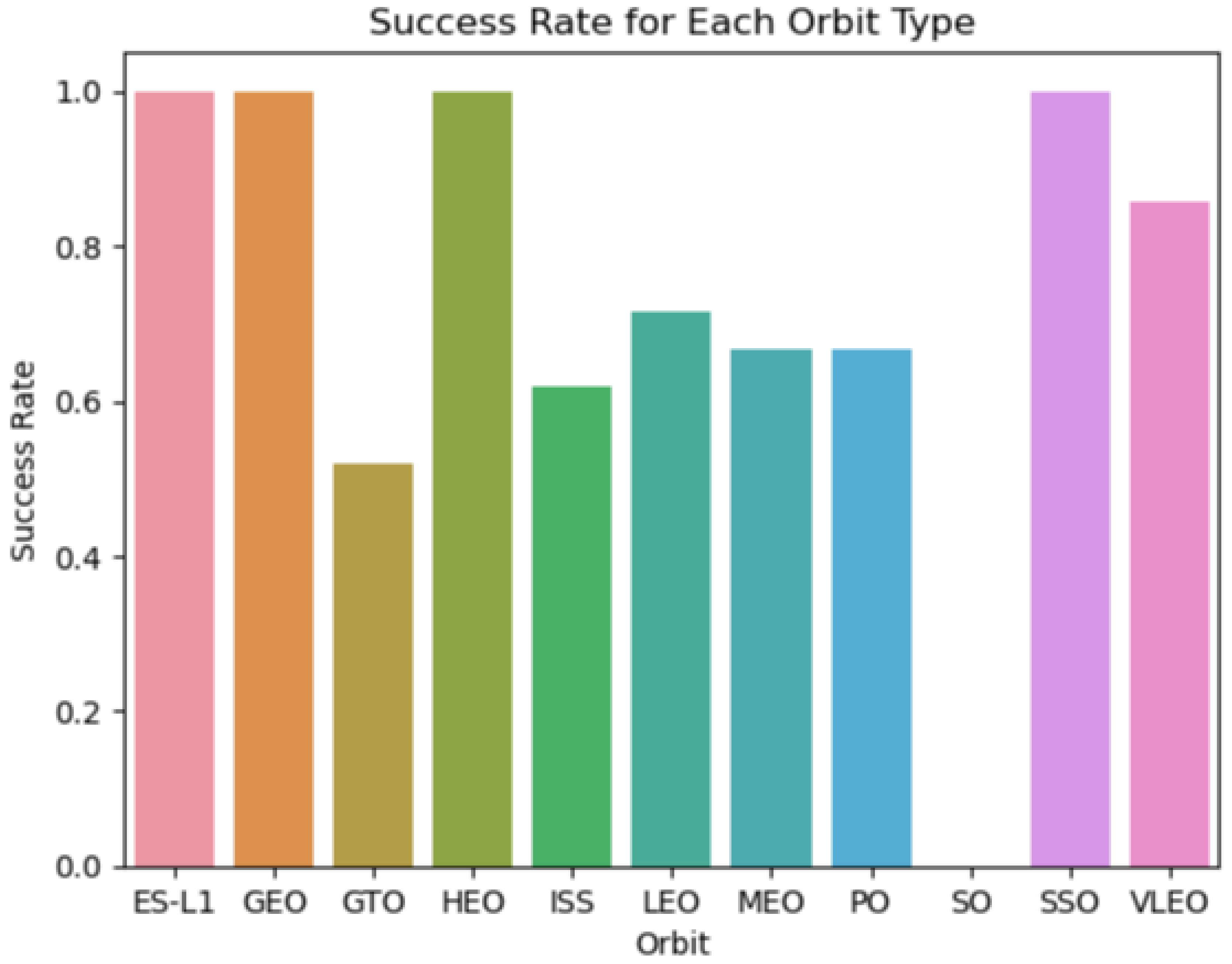


## RELATIONSHIP BETWEEN THE PAYLOAD VS LAUNCH SITE

Upon observing the scatter plot depicting Payload versus Launch Site, notable trends emerge. The **VAFB SLC 4E** and the **KSC LC 39A** stands out for their consistency in **Successive Launches of Payload** delivered within the **Range of 2,000kg to 16,000kg**. While occasional failures occur, the overall success rate outweighs these instances.

Remarkably, the **VAFB SLC 4E** exhibits exceptional performance within the **Payload Mass Range of 2,000kg to 10,000kg, with zero failures recorded**. Additionally, it's worth noting that no rockets are launched for **Heavy Payload Masses Exceeding 10,000kg** at the **VAFB-SLC Launch Site.**

An argument can be made for the effectiveness of the **CCAFS SLC 40**, particularly within the **Payload Mass Range of 14,000kg to 16,000kg.**



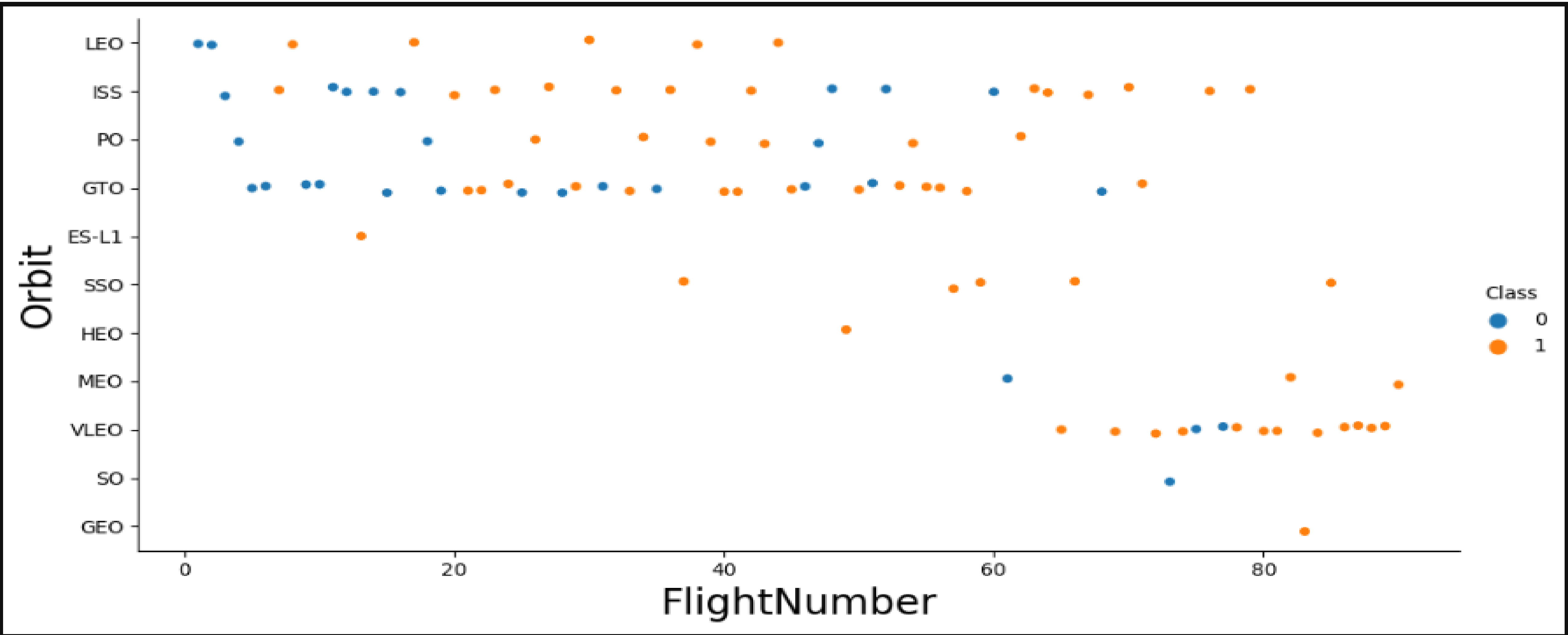

# SUCCESS RATE VS. ORBIT TYPE

**BAR CHART REPSENTING THE SUCCESS RATE OF EACH ORBIT TYPE -**

**From observation its apparent that the ES-L1, GEO, HEO AND SSO Orbits have the Highest Success Rate whilst the GTO has the Lowest Success Rate.**

# FLIGHT NUMBER VS ORBIT TYPE

**SCATTER PLOT OF FLIGHT NUMBER VS ORBIT TYPE**



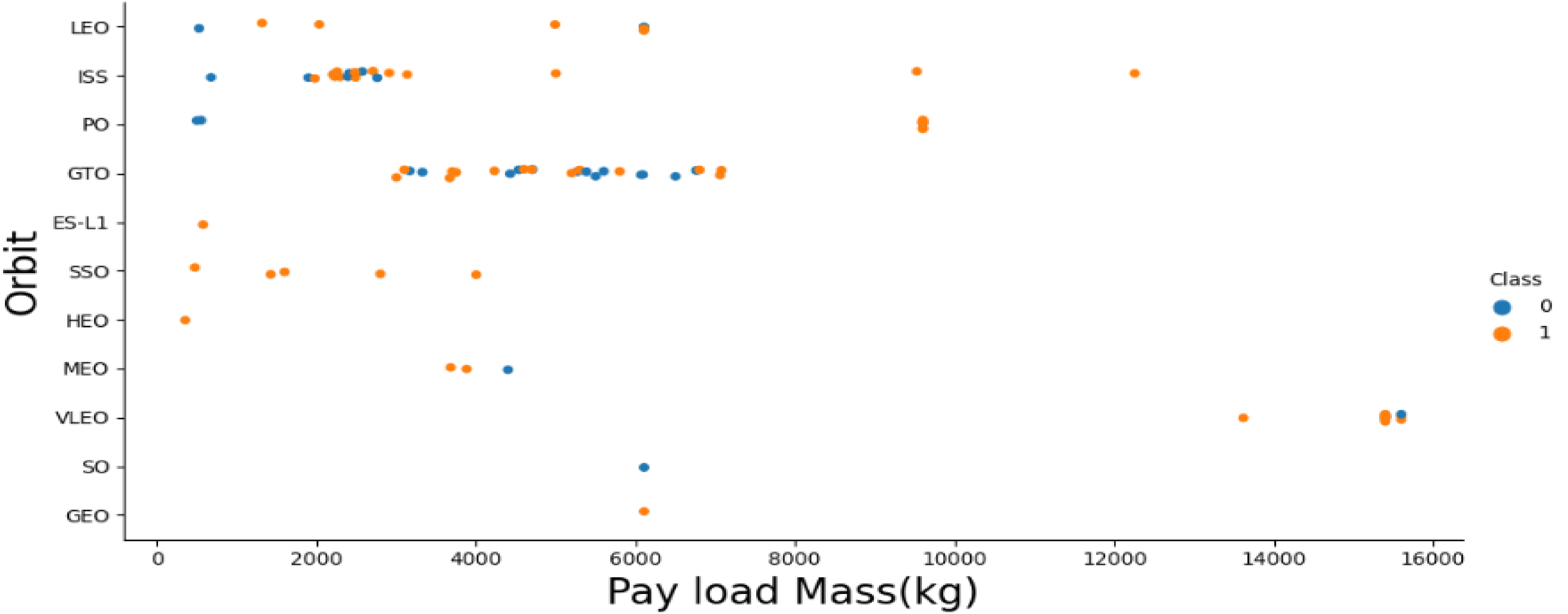
## RELATIONSHIP BETWEEN THE FLIGHT NUMBER VS ORBIT TYPE

Upon observation, a clear correlation emerges between **Flight Number and Orbit Type.** Specifically, the **Low Earth Orbit (LEO)** exhibits a direct relationship with **Flight Number**, with each successive flight demonstrating a higher probability of successful landings. In contrast, the **Geostationary Transfer Orbit (GTO)** shows no discernible correlation with **Flight Number.**

Similarly, the **Sun-Synchronous Orbit (SSO)** demonstrates a high degree of success relative to **Flight Number**, akin to the **LEO orbit.**

# PAYLOAD VS. ORBIT TYPE

**SCATTER PLOT OF PAYLOAD VS ORBIT TYPE**



## RELATIONSHIP BETWEEN THE PAYLOAD VS ORBIT TYPE

Upon observation, a clear correlation emerges between **Payload and Orbit Type**.

**Heavy Payloads**

exhibit a **Higher Rate Of Successful Or Positive Landings** For **Polar(PO), Low Earth Orbit (LEO)**, and

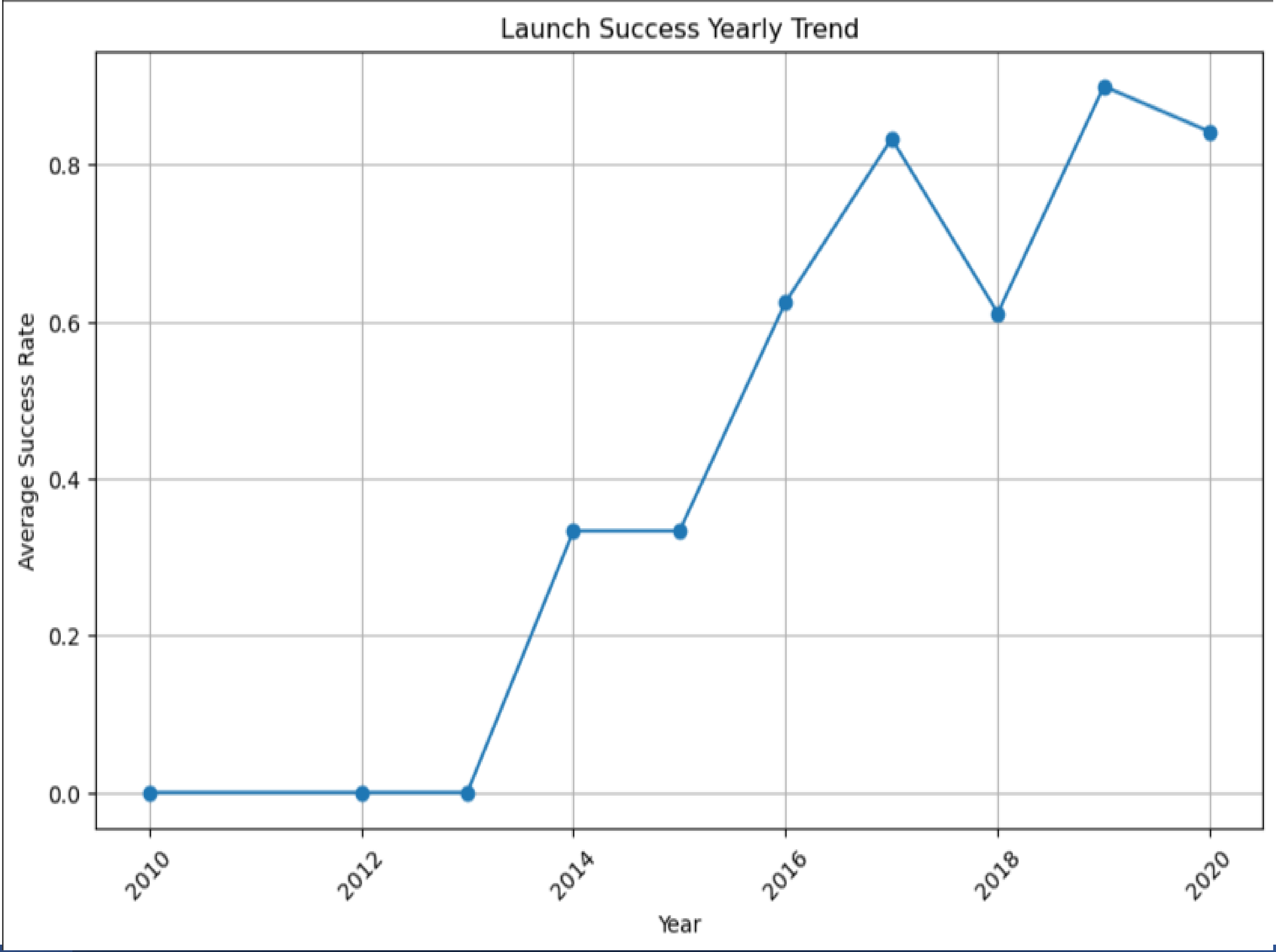
**International Space Station (ISS) Orbits.**

However, distinguishing this correlation is challenging for **Geostationary Transfer Orbit (GTO)**, as both **Positive Landing Rates and Negative Landings (Unsuccessful Missions)** are present.

# LAUNCH SUCCESS YEARLY TREND

**LINE CHART REPRESENTING THE YEARLY AVERAGE SUCCESS RATE :**

**You can observe that the Success Rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.**



# ALL LAUNCH SITE NAMES

**THIS QUERY DISPLAYED THE UNIQUE NAMES OF THE LAUNCH SITES IN THE SPACE MISSION:**

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

| LAUNCH_SITE |
|---|
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |

# LAUNCH SITE NAMES BEGIN WITH 'CCA'

**THIS QUERY DISPLAYED 5 RECORDS OF LAUNCH SITES WHICH BEGAN WITH `CCA`:**

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

| 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 |

# TOTAL PAYLOAD MASS

**THIS QUERY CALCULATED THE TOTAL PAYLOAD CARRIED BY BOOSTERS FROM NASA:**

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

| TOTAL_PAYLOAD_MASS |
|---|
| 48213 |

# AVERAGE PAYLOAD MASS BY F9 v1.1

**THIS QUERY CALCULATED THE AVERAGE PAYLOAD MASS CARRIED BY BOOSTER VERSION F9 v1.1:**

- **%sql** SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

| AVERAGE_PAYLOAD_MASS |
| --- |
| **2928.4** |

# FIRST SUCCESSFUL GROUND LANDING DATE

**THIS QUERY FOUND THE DATES OF THE FIRST SUCCESSFUL LANDING OUTCOME ON GROUND PAD:**

- **%sql** SELECT MIN(Date) AS First_Successful_Landing_On_Ground_Pad FROM SPACEXTABLE WHERE Mission_Outcome = 'Success' AND Landing_Outcome = 'Success (ground pad)';

| FIRST_SUCCESSFUL_LANDING_ON_GROUND_PAD |
|---|
| 2015-12-22 |

# SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

**THIS QUERY LISTED THE NAMES OF BOOSTERS WHICH HAVE SUCCESSFULLY LANDED ON DRONE SHIP AND HAD PAYLAD MASS GREATER THAN 4000 BUT LESS THAN 6000:**

- `sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Mission_Outcome = 'Success' AND Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;`

| BOOSTER_VERSION |
| --- |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

# TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

**THIS QUERY CALCULATED THE TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES**

- `%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome;`

| MISSION_OUTCOME | COUNT(*) |
|---|---|
| FAILURE (IN FLIGHT) | 1 |
| SUCCESS | 98 |
| SUCCESS | 1 |
| SUCCESS (PAYLOAD STATUS UNCLEAR) | 1 |

# BOOSTERS CARRIED MAXIMUM PAYLOAD

## THIS QUERY LISTED THE NAMES OF THE BOOSTER WHICH HAVE CARRIED THE MAXIMUM PAYLOAD MASS

- `%sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);`

| BOOSTER_VERSION | |
|---|---|
| F9 B5 B1048.4 | F9 B5 B1049.5 |
| F9 B5 B1049.4 | F9 B5 B1060.2 |
| F9 B5 B1051.3 | F9 B5 B1058.3 |
| F9 B5 B1056.4 | F9 B5 B1051.6 |
| F9 B5 B1048.5 | F9 B5 B1060.3 |
| F9 B5 B1051.4 | F9 B5 B1049.7 |

# 2015 LAUNCH RECORDS

**THIS QUERY LISTED THE FAILED LANDING_OUTCOMES IN DRONE SHIP, THEIR BOOSTER VERSIONS, AND LAUNCH SITE NAMES FOR YEAR 2015 :**

- `%sql SELECT substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome LIKE '%Failure%' AND Landing_Outcome LIKE '%Drone Ship%';`

**RESULT**

| 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 2010-06-40 AND 2017-03-20

**AN SQL Query Which Ranks 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 :**

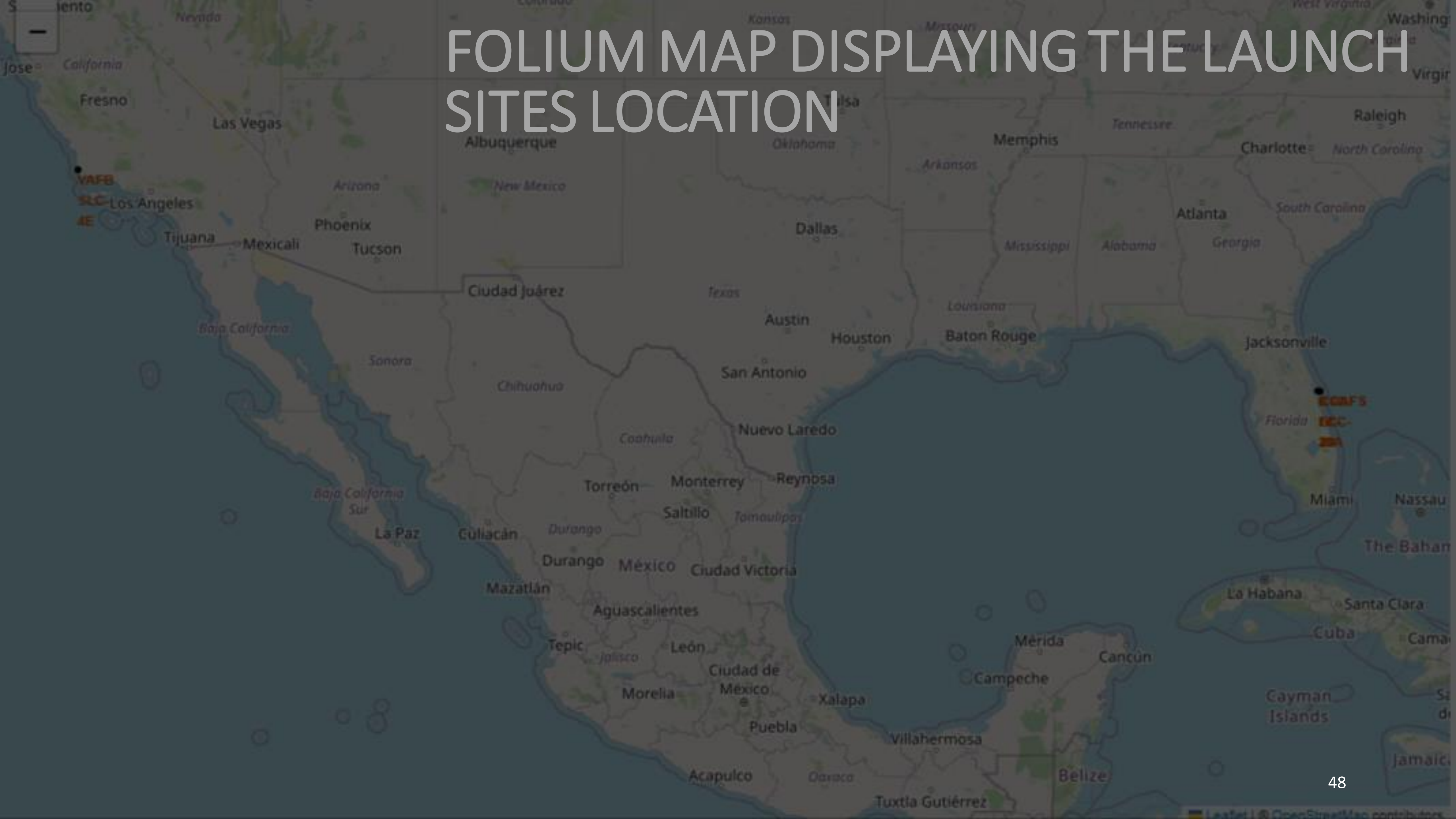
- **%sql** SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' AND (Landing_Outcome LIKE '%Failure (drone ship)%' OR Landing_Outcome LIKE '%Success (ground pad)%') GROUP BY Landing_Outcome ORDER BY Count DESC;

| LANDING_OUTCOME | COUNT |
| --- | --- |
| FAILURE (DRONE SHIP) | 5 |
| SUCCESS(GROUND PAD) | 3 |

Section 3

# Launch Sites Proximities Analysis

# FOLIUM MAP DISPLAYING THE LAUNCH SITES LOCATION



## NOTEWORTHY ELEMENTS/FINDINGS PRESENT ON THE MAP

- P**roximity to the Equator line:** Not all launch sites are in close proximity to the Equator line. Some launch sites, such as Cape Canaveral and Kennedy Space Center, are located closer to the Equator compared to others like Vandenberg Air Force Base. However, none of the launch sites are directly on the Equator line.

- **Proximity to the coast:** Most of the launch sites are in close proximity to the coast. For example, Cape Canaveral and Kennedy Space Center are located along the eastern coast of the United States, while Vandenberg Air Force Base is situated on the western coast.

These findings are consistent with the geographical considerations for launch sites. Launch sites are often situated close to the Equator to take advantage of the Earth's rotational speed, which provides an additional velocity boost to rockets launched eastward. Additionally, proximity to the coast allows for safer launch trajectories over open water and provides logistical advantages for transporting and assembling rocket components.

# LAUNCH OUTCOMES AT CCAFS SLC-40 SITE

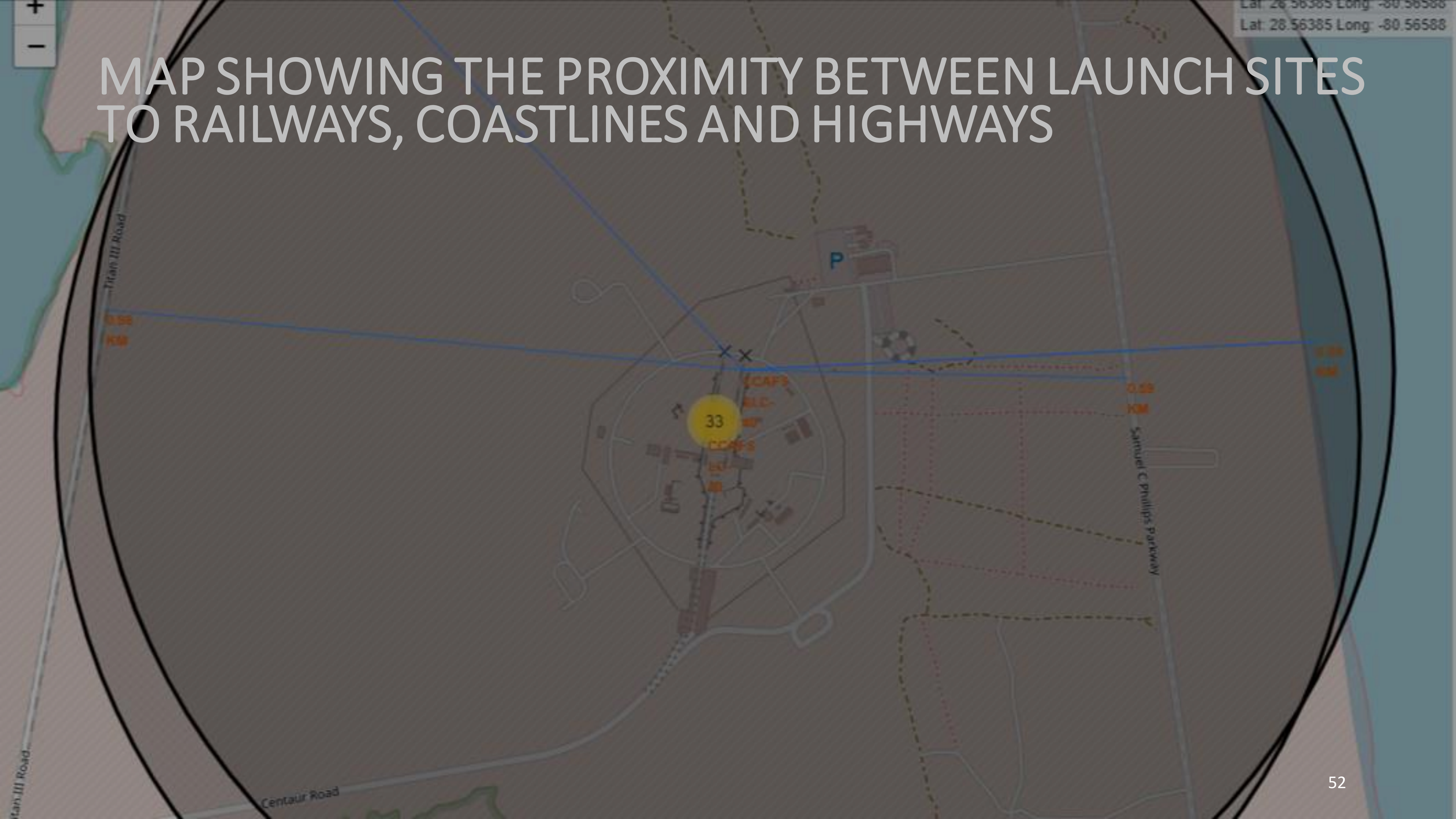
## NOTEWORTHY ELEMENTS/FINDINGS PRESENT ON THE MAP

The screenshot provided consist of the launch outcomes at Cape Canaveral center (CCAFS SLC-40), The Red Marker depicts Failure whilst the Green Marker suggest Successful Landing.

A total of 33 Rocket Launches where executed off the Launch Site :

- **POINT 1** – 26 Launches (19 Failures and 7 Success)
- **POINT 2** – 7 Launches (4 Failures and 3 Success)

# MAP SHOWING THE PROXIMITY BETWEEN LAUNCH SITES TO RAILWAYS, COASTLINES AND HIGHWAYS



## NOTEWORTHY ELEMENTS/FINDINGS PRESENT ON THE MAP

The screenshot provided shows the Proximity of Cape Canaveral center (CCAFS SLC-40) to Railways, Highways and Coastline, the distance was measured in Kilometers. Also their latitude and longitude where visualized alongside(Top right corner).

I am very much aware of how strenuous it might be on the eyes to get the accurate distance from the screenshot, therefore I will provide the entire details of the finding below:

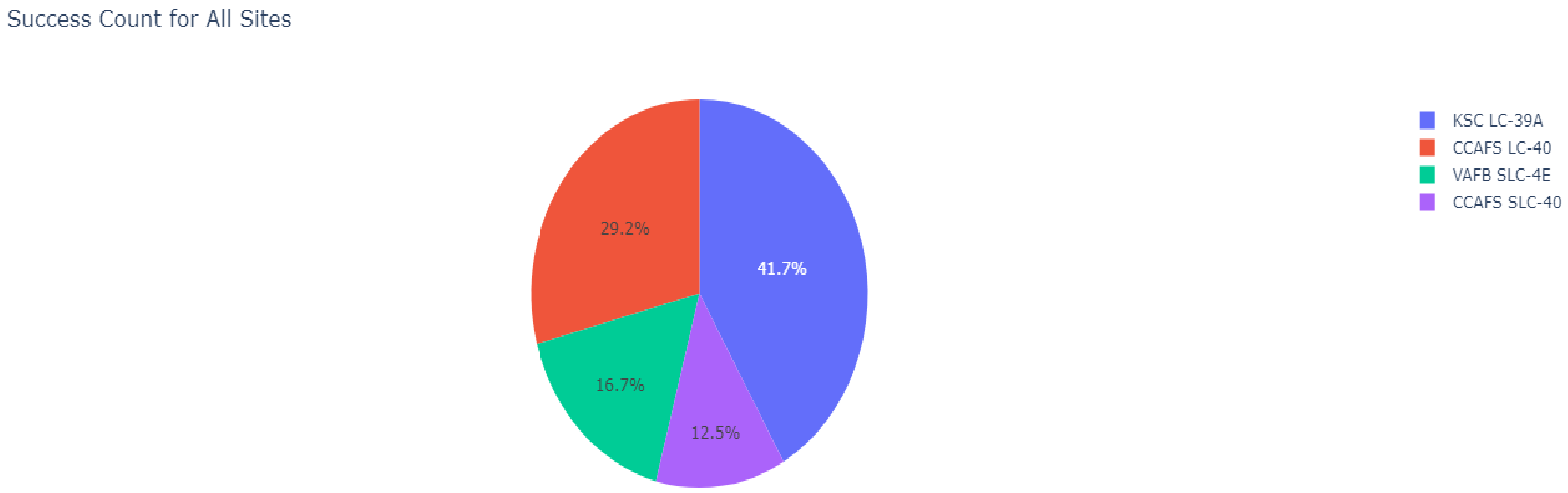
- **COASTLINE DISTANCE** – 0.88KM (Lat:28.56361, Lon:-80.5678)
- **RAILWAY DISTANCE** – 0.98KM (Lat:28.56403, Lon:-80.58679)
- **HIGHWAY DISTANCE** – 0.59KM (Lat:28.56309, Lon:-80.57076)

From the above we can clearly see that they all are in close proximity to the Space station but the Highway showed to be the closest out of the three. There are a few reasons to why Launch sites are built with this in mind, its usually encompasses; Safety, Security, Infrastructure and Accessibility.

One major observation is that the Launch site was situated very far from the City, This is done for Safety Reasons.

Section 4

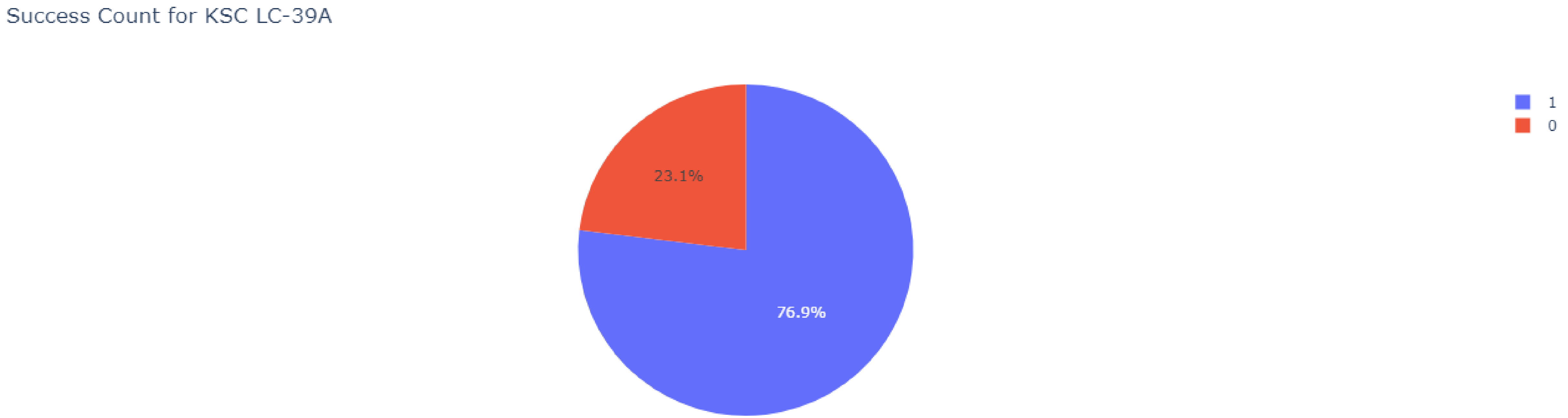
# Build a Dashboard with Plotly Dash




# PIE CHART DEPICTING THE RATIO FOR EACH SITES SUCCESSFUL LAUNCHES

**EACH LAUNCH SITES SUCCESS RATE -**

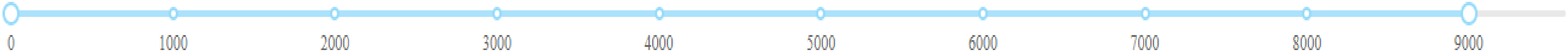
- **KSC LC-39A : 41.7%**
- **CCAFS LC-40 : 29.2%**
- **VAFB SLC-4E : 16.7%**
- **CCAFS SLC-40 : 12.5%**

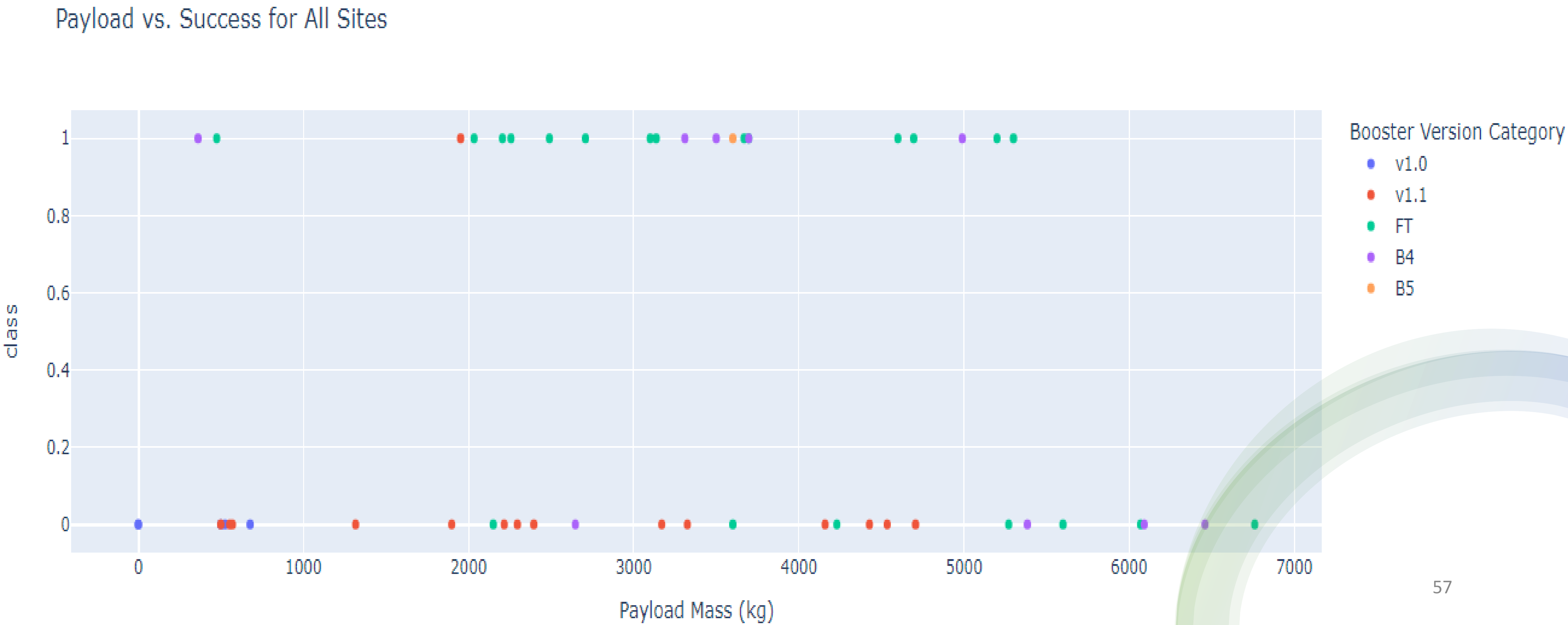



# PIECHART OF THE LAUNCH SITE WITH THE HIGHEST SUCCESS RATE

**FROM THE SCREENSHOT WE CAN OBSERVE THAT THE KSC LC-39A LAUNCH SITE HAS THE HIGHEST SUCCESS RATE.**

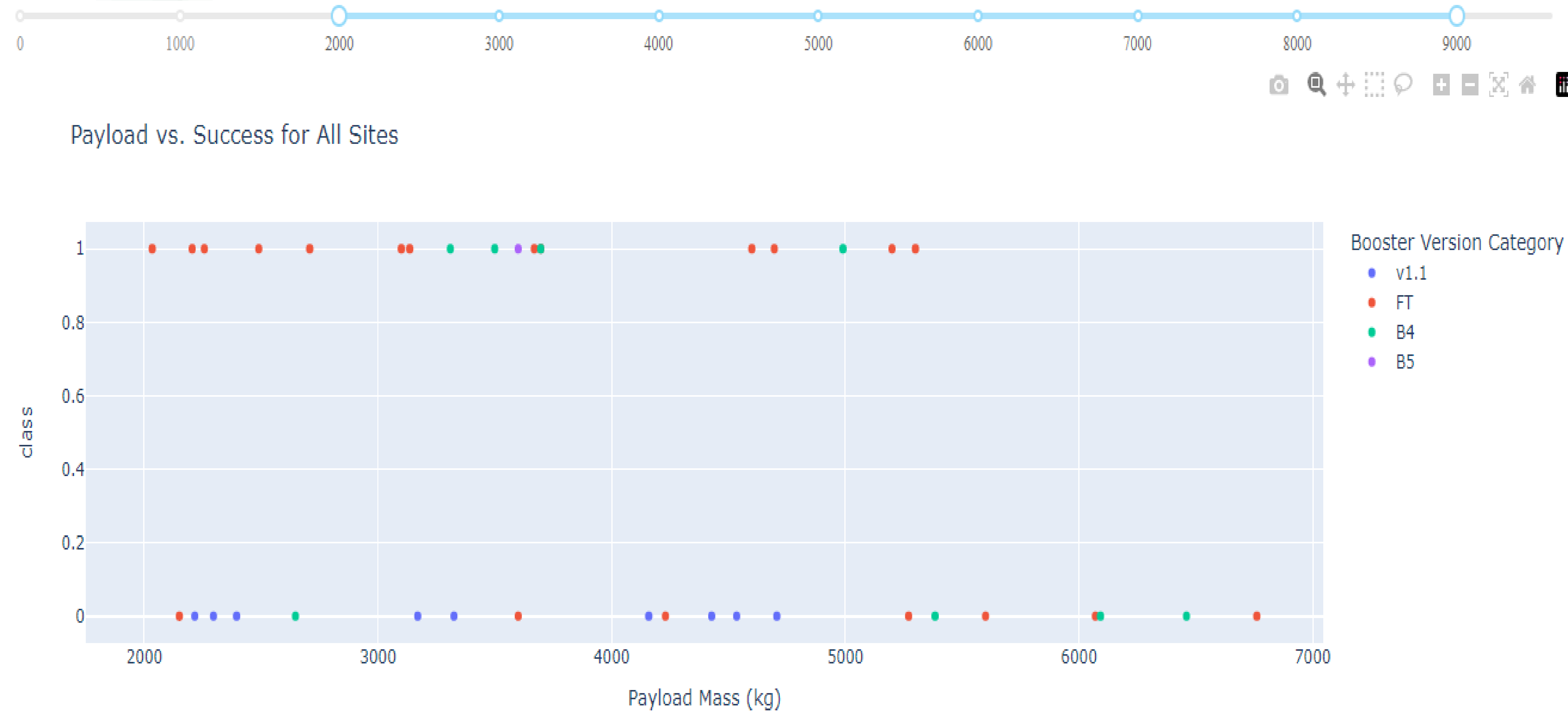
# SCATTER PLOT OF PAYLOAD VS OUTCOMES FOR ALL BOOSTER VERSION



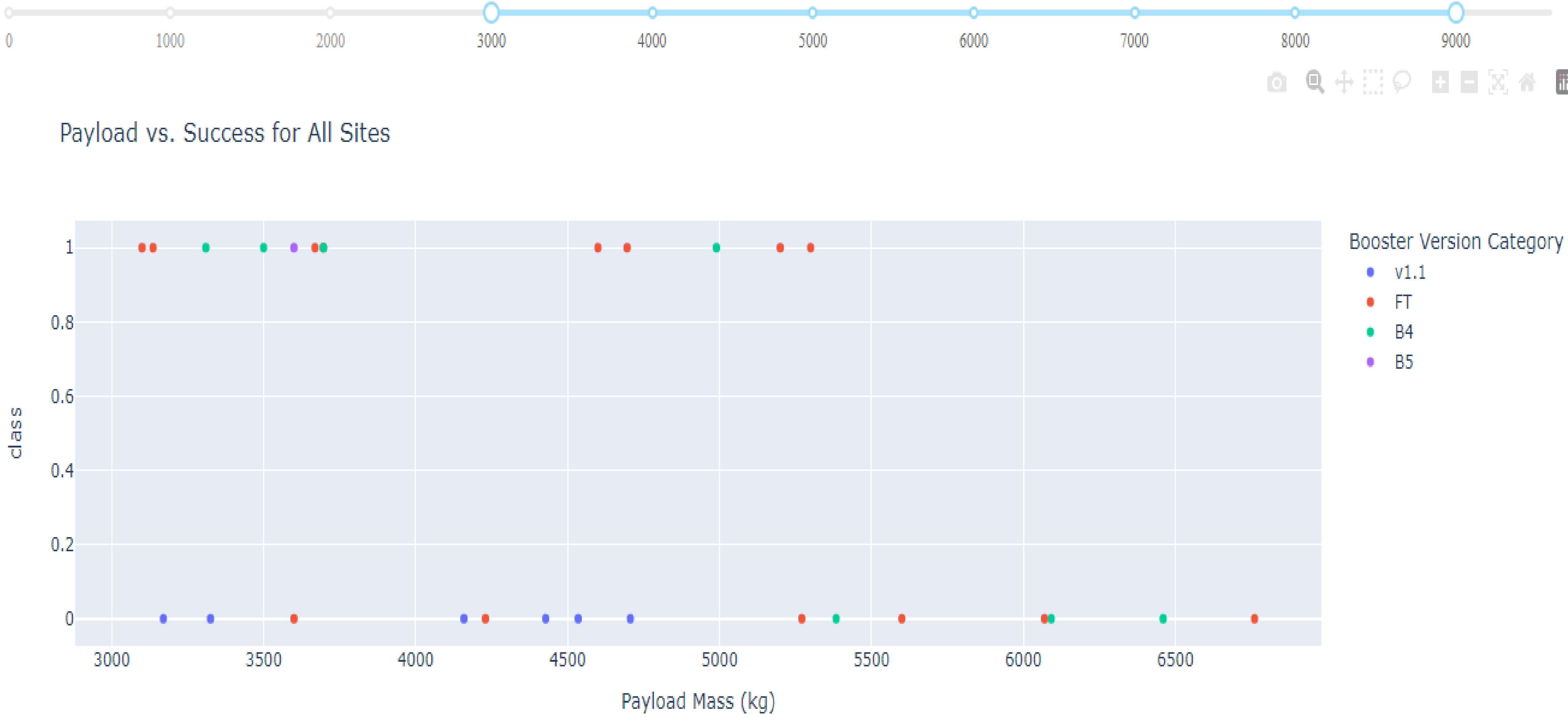



# SCATTER PLOT OF PAYLOAD RANGE (1000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION

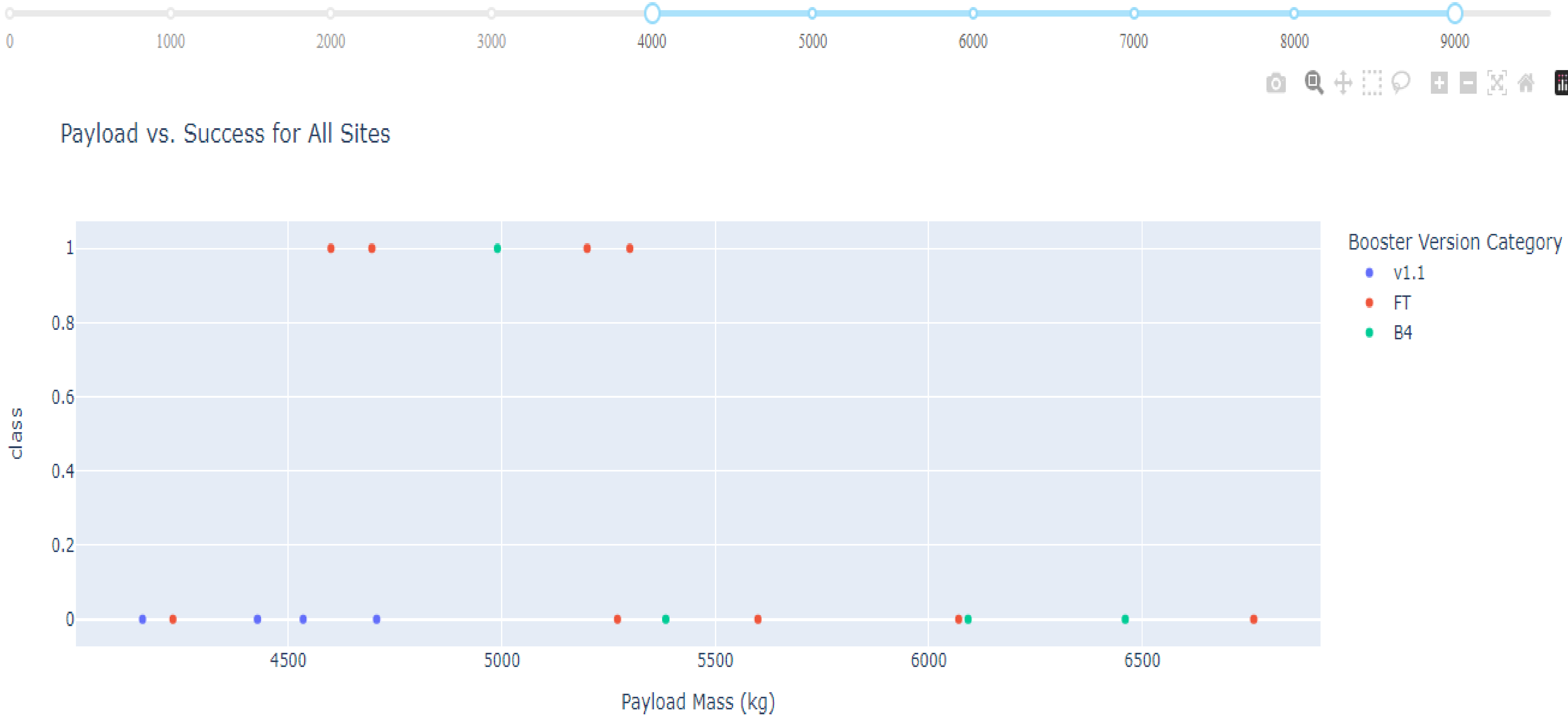
# SCATTER PLOT OF PAYLOAD RANGE (2000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



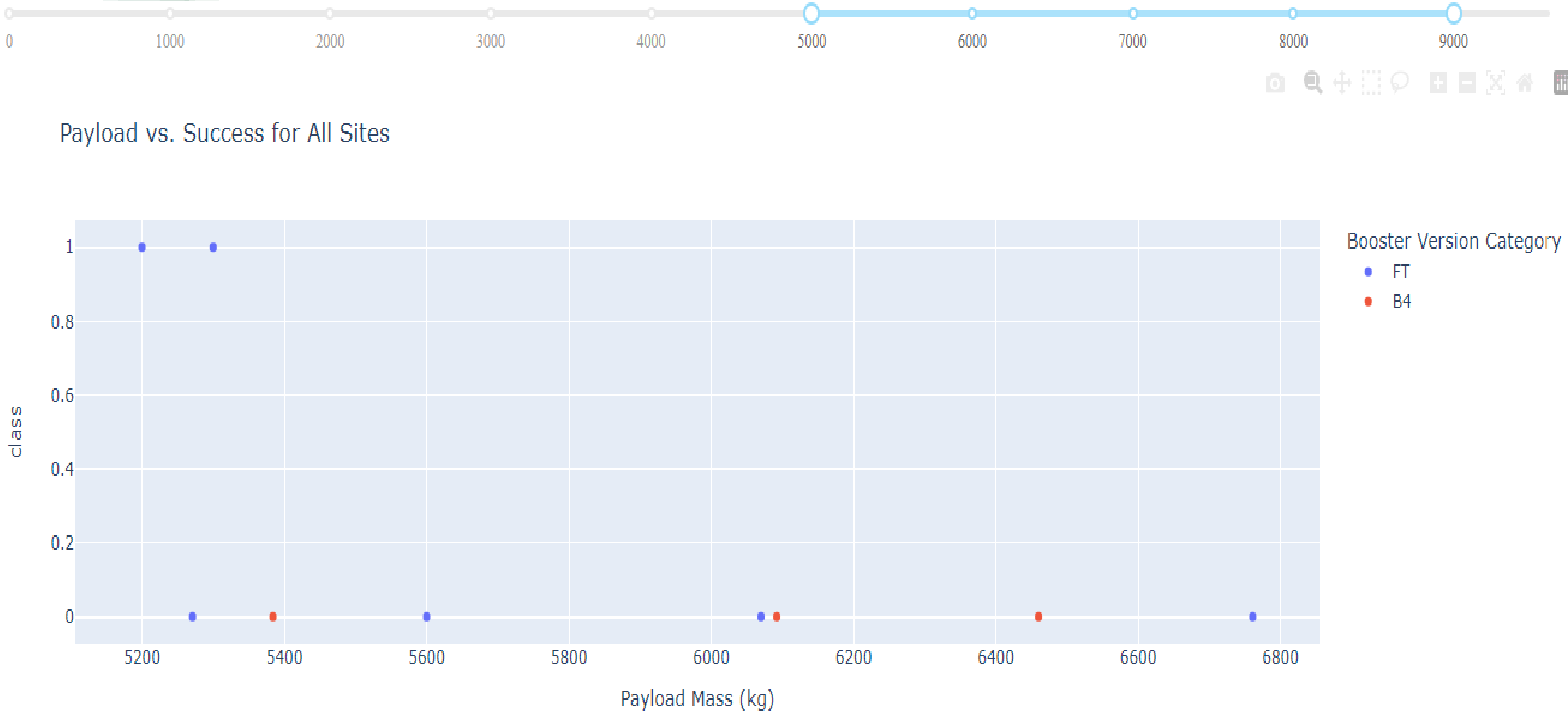
# SCATTER PLOT OF PAYLOAD RANGE (3000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



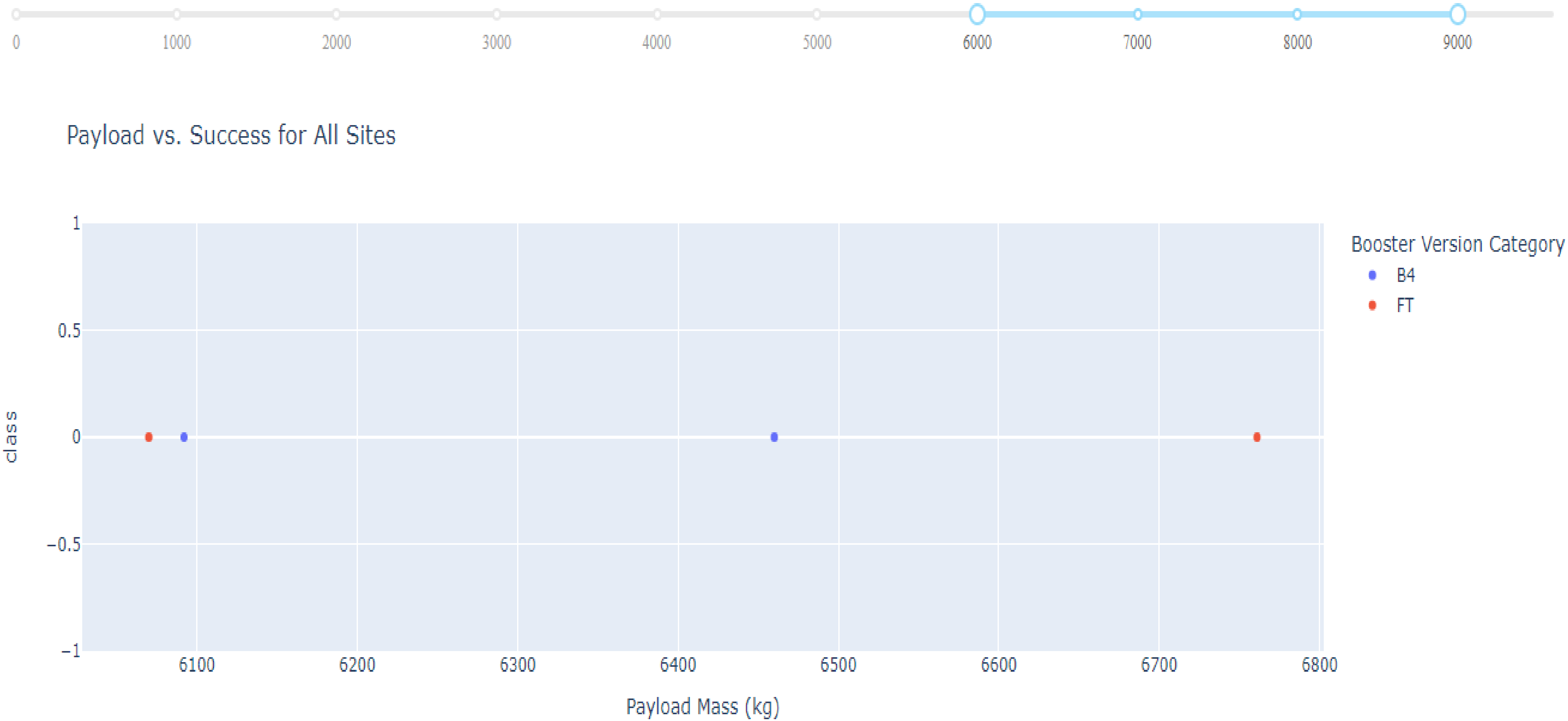
# SCATTER PLOT OF PAYLOAD RANGE (4000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



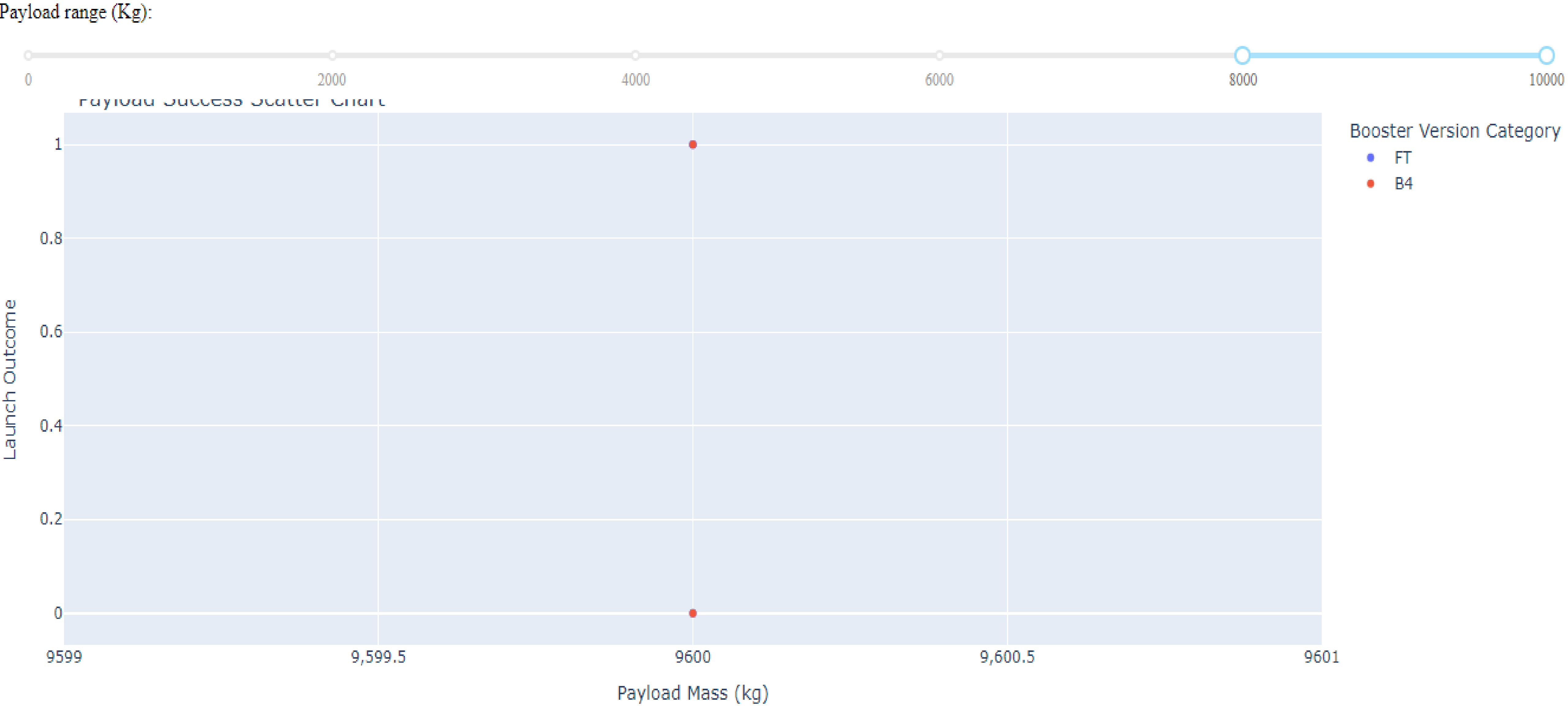
# SCATTER PLOT OF PAYLOAD RANGE (5000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



# SCATTER PLOT OF PAYLOAD RANGE (6000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



# SCATTER PLOT OF PAYLOAD RANGE (8000KG – 10,000KG) VS OUTCOMES FOR ALL BOOSTER VERSION



## NOTEWORTHY ELEMENTS/FINDINGS PRESENT ON THE MAP

**First thing First, I will begin by explaining the key labels which were present in the screenshots I provided:**

- **The First Depiction we have is the Payload Range Slider:** The function of this feature on the dashboard is basically to give the user the ability to navigate between different payload ranges and access the outcomes in respect to the Booster Version, The range was divided between 0kg all the way to 10,000kg, by doing this it allowed all outcomes to be recorded and accounted for.

- **Next, The Booster Version Category :** The Booster version which where utilized in this Screenshot were depicted with different individual colors in order to ease identification. The Following Booster Version were employed in the Launches on All sites; v1.0, v1.1, FT, B4, B5.

- **Scatter Plot:** The Scatter Plot was utilized to visually depict the Launch Outcomes for all Booster versions which were engaged in every launch sites.

- **The Y- axis consist of a label 'CLASS';** This Lable basically involves Two Categorical variables, **1** and **0**. **1** means Successful Landing and **0** means Failure.

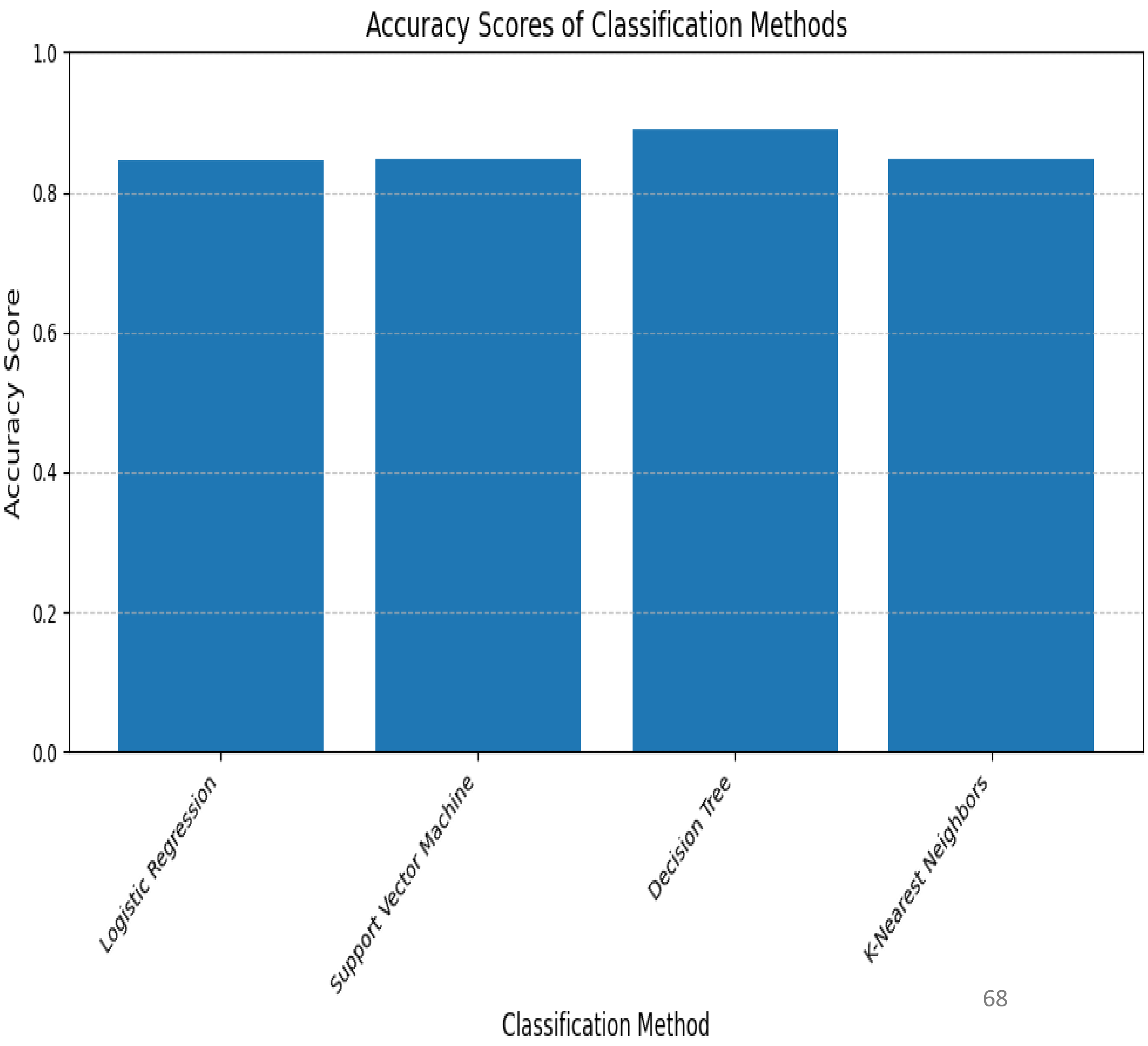
- **The X-axis consist of the Payload mass in kg;** the payload changes in respect to the Payload Range slider inputs.

After visually exploring each Payload Ranges on the Scatter Plot with the help of the Payload Slider, we can draw a straight conclusion that the **FT Booster Version** has the Highest Success Rate from all the Categories, this is a truism regardless of the Payload Mass being utilized.

Section 5

# Predictive Analysis (Classification)

# CLASSIFICATION ACCURACY

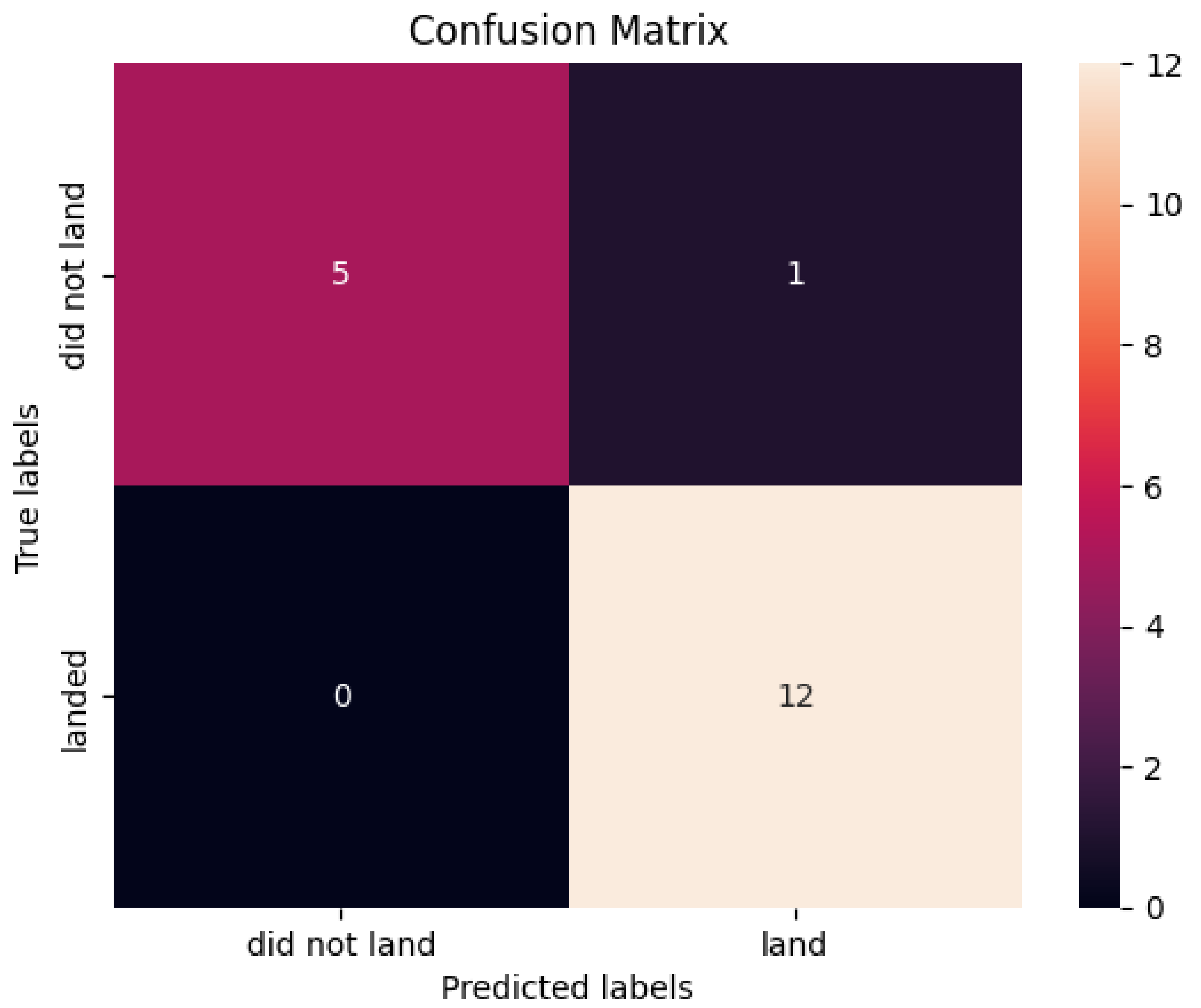


# CONFUSION MATRIX

THE DECISION TREE SHOWED TO HAVE THE BEST PERFORMING CONFUSION MATRIX, THE ONLY SIGNIFICANT PROBLEM:

- FALSE NEGATIVE – DID NOT LAND COLUMN HAS A FALSE NEGATVE OF -1, THE ACTUAL RESULT IS 6 WHILST THE PREDICTED LABEL OUTPUTS 5.

- FALSE POSITIVE – LANDED COLUMN HAS A FALSE POSITIVE OF +1, THE ACTUAL RESULT IS 12 WHILST THE PREDICTED LABEL PRODCES 13.



# CONCLUSIONS

**The entire concept behind this project was to develop a Machine Learning Model which would be capable of Predicting the Re-usability of Space X – First Stage rocket launch, I am proud to say that this objective was met and exceeded expectation. Highlighted below in Four clear points are the concluded objectives which were met alongside further details for clear readability;**

- **Model Performance Comparison**: Through rigorous analysis and evaluation using machine learning models such as Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors, we have determined the performance of each model in predicting the success or failure of SpaceX launches. By comparing each of these accuracy scores, we identified the most effective model for this Classification task which turned out to be the Decision Tree!

- **Optimal Hyperparameters Selection**: Utilizing techniques like GridSearchCV, we fine-tuned the hyperparameters of each model to optimize their performance. This process allowed us to identify the best combination of parameters that maximized accuracy, ensuring that our models were well-optimized for the given dataset.

- **Insights into Predictive Power**: By analyzing confusion matrices and accuracy scores, we gained insights into the predictive power of each model. We observed how well each model classified successful and unsuccessful launches, providing valuable information for stakeholders and decision-makers in SPACE Y. Onca again the Decision Tree proved to provide the Nearest Accurate matric.
- **Practical Implications and Future Directions**: The findings from this analysis have practical implications for SPACE Y and the broader aerospace community. By understanding which models perform best in predicting launch outcomes, stakeholders can make more informed decisions about mission planning, risk assessment, and resource allocation. Furthermore, this study opens avenues for future research, such as exploring additional features or incorporating more advanced machine learning techniques to further improve predictive accuracy and robustness.

**In summary, this comprehensive workflow has provided valuable insights into the predictive capabilities of various machine learning models for SpaceX launch success prediction. These findings have significant implications for decision-making processes within the aerospace industry and lay the groundwork for further advancements in predictive modeling and analysis.**

**I would like to end on this Highnote, Outplacing SPACEX dominance is no longer a monumental action, due to the insights derived from this Project/data we can clearly see the steps which must be taken and the actions needed to be employed. Success is no longer a dream but now a reality for SPACE Y.**

# Appendix

## IF YOU ARE HAVING A HARD TIME CLICKING ANY OF THE LINKS, BELOW ARE ALL THE GITHUB URL FOR YOUR EASE;

- **DATA COLLECTION - API**
- **DATA COLLECTION - WEBSCRAPPING**

- **DATA WRANGLING**
- **DATA VISUALIZATION**
- **DATA EDA - SQL**

- **LAUNCH SITE VISUALIZATION (FOLIUM MAP)**
- BUILDING WITH PLOTLY DASHBOARD
- MACHINE LEARNING MODEL

# Thank you!