IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Daniel Hamama
29.01.2025

# Outline

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

# Executive Summary

- **Objective:** Analyze SpaceX launch data to understand success factors and predict future launch outcomes.
- **Methods Used:** Web scraping, API data collection, data wrangling, SQL queries, interactive visualizations, and machine learning models.
- **Key Findings:** Identified trends in launch success, payload impacts, and predictive model performance.

# Introduction

- **Project Background:** SpaceX launches are crucial for the commercial space industry. Understanding the key factors affecting success is vital.
- **Questions to Answer:**
  - What factors contribute to launch success?
  - How do payload mass and orbit type affect outcomes?
  - Can machine learning predict launch success?

Section 1

# Methodology

# Methodology

## Executive Summary

- Data Collection:
  - SpaceX API Calls, Web Scraping.
- Data Wrangling / Cleaning:
  - Removed missing/duplicates, Converted date formats, Standardized column names.
- Exploratory Data Analysis (EDA) using Visualization and SQL
- Interactive Visual Analytics using Folium and Plotly Dash
- Predictive Analysis using Classification Models
  - Comparing Multiple Models, Hyperparameter Tuning, Confusion Matrix, Evaluation Metrics.

# Data Collection

**Methods Used:**

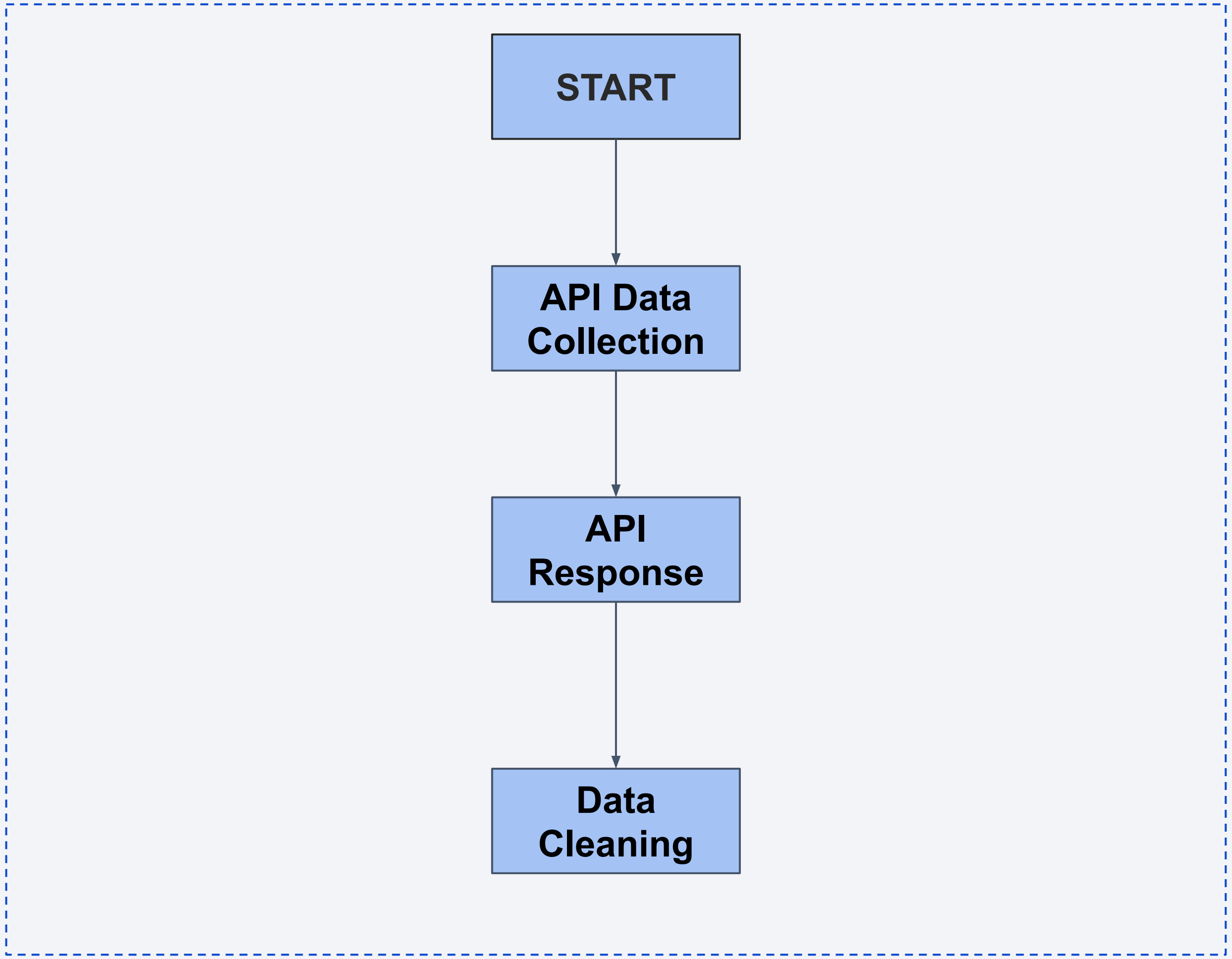
- **SpaceX API Calls** → Collected launch details via REST API.
- **Web Scraping** → Retrieved additional launch data from Wikipedia.

# Data Collection – SpaceX API

- We start with using the REST API for SpaceX API Calls, then we get the API response and lastly we clean the data.

- GitHub Link for the API Data Collection notebook:

  [API Data Collection]()

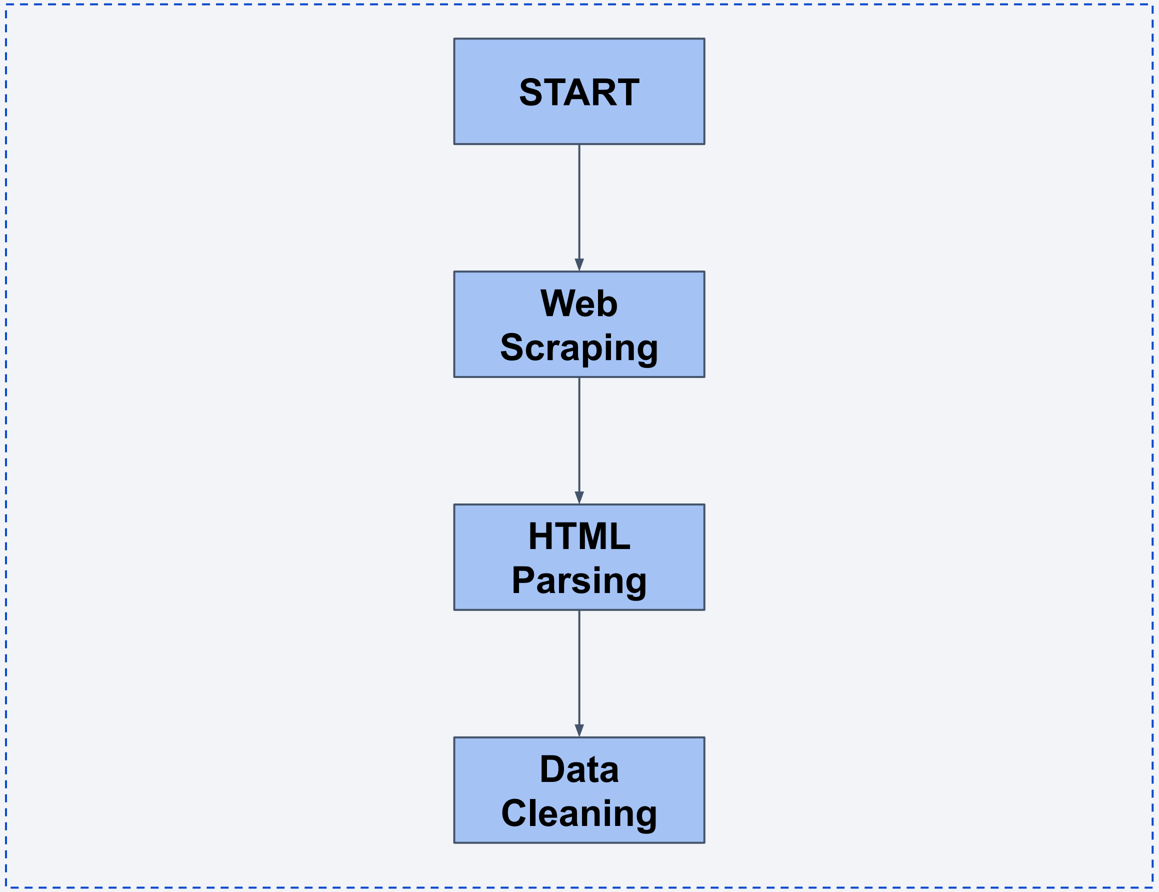


# Data Collection - Scraping

- Starting with Web Scraping we retrieve additional data from Wikipedia, then we HTML Parse and we move on to Data Cleaning.

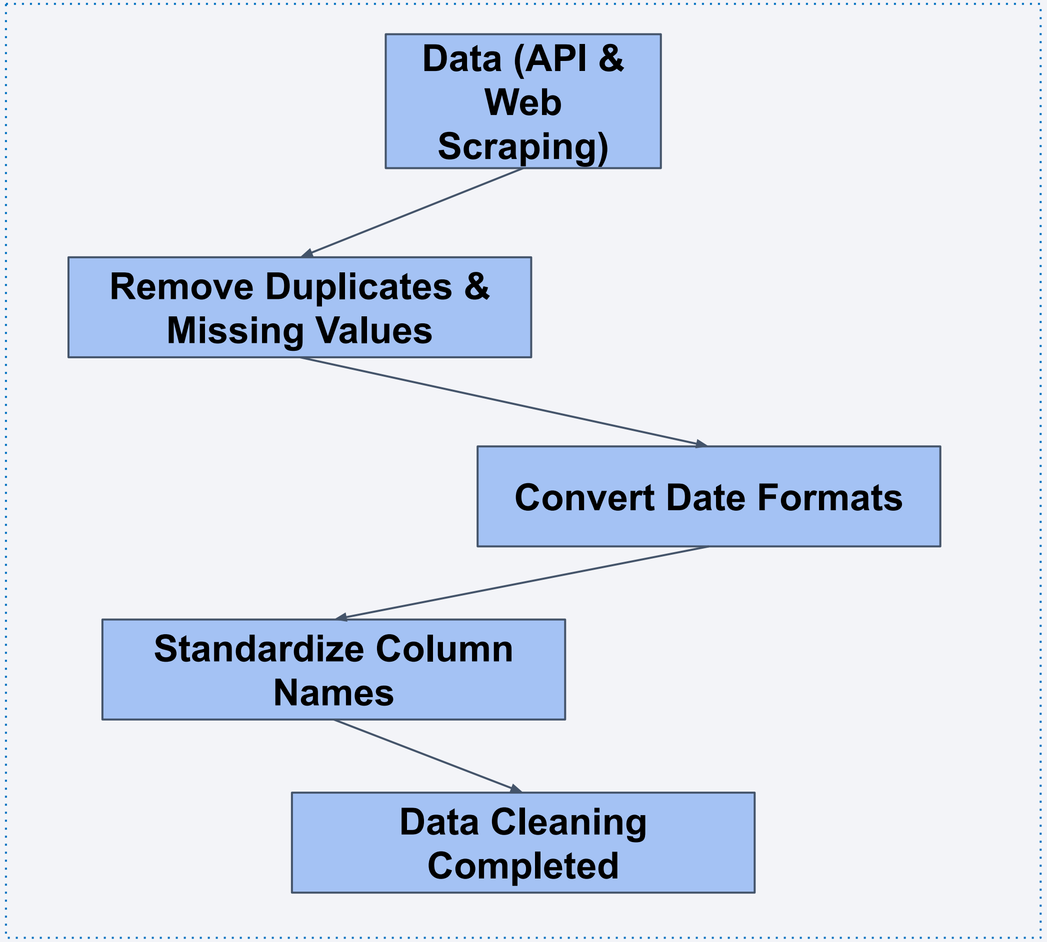
- GitHub Link for the Web Scraping notebook:

  [Web Scraping]



# Data Wrangling & Cleaning

- **Tasks Performed:**
  - Removed missing/duplicate records.
  - Converted date formats.
  - Standardized column names.
- **Key Outcome:** A clean dataset ready for exploratory analysis.
- **GitHub Links:** Data Wrangling & Cleaning



# EDA with Data Visualization

- **Data Visualization:**
  - **Scatter Plots:** Flight Number vs. Launch Site, Payload vs. Launch Site.
  - **Bar Chart:** Success rate for each orbit.
  - **Line Chart:** Launch success trend over time.
- **GitHub Links:** EDA Data Visualization

# EDA with SQL

- **SQL Queries & Findings:**
  - Unique launch sites identified.
  - Total payload mass carried by NASA boosters.
  - Average payload mass for Falcon 9.
  - Success rate by orbit type.
- **GitHub Links: SQL EDA**

# Build an Interactive Map with Folium

**Folium Map:**

- Displayed launch site locations with markers.
- Proximities to railways, highways, coastlines.
- Color-coded success/failure outcomes.

**GitHub Link:** Folium Map Launches Location

**NBViewer Link for Map view:** Folium Map in nbviewer

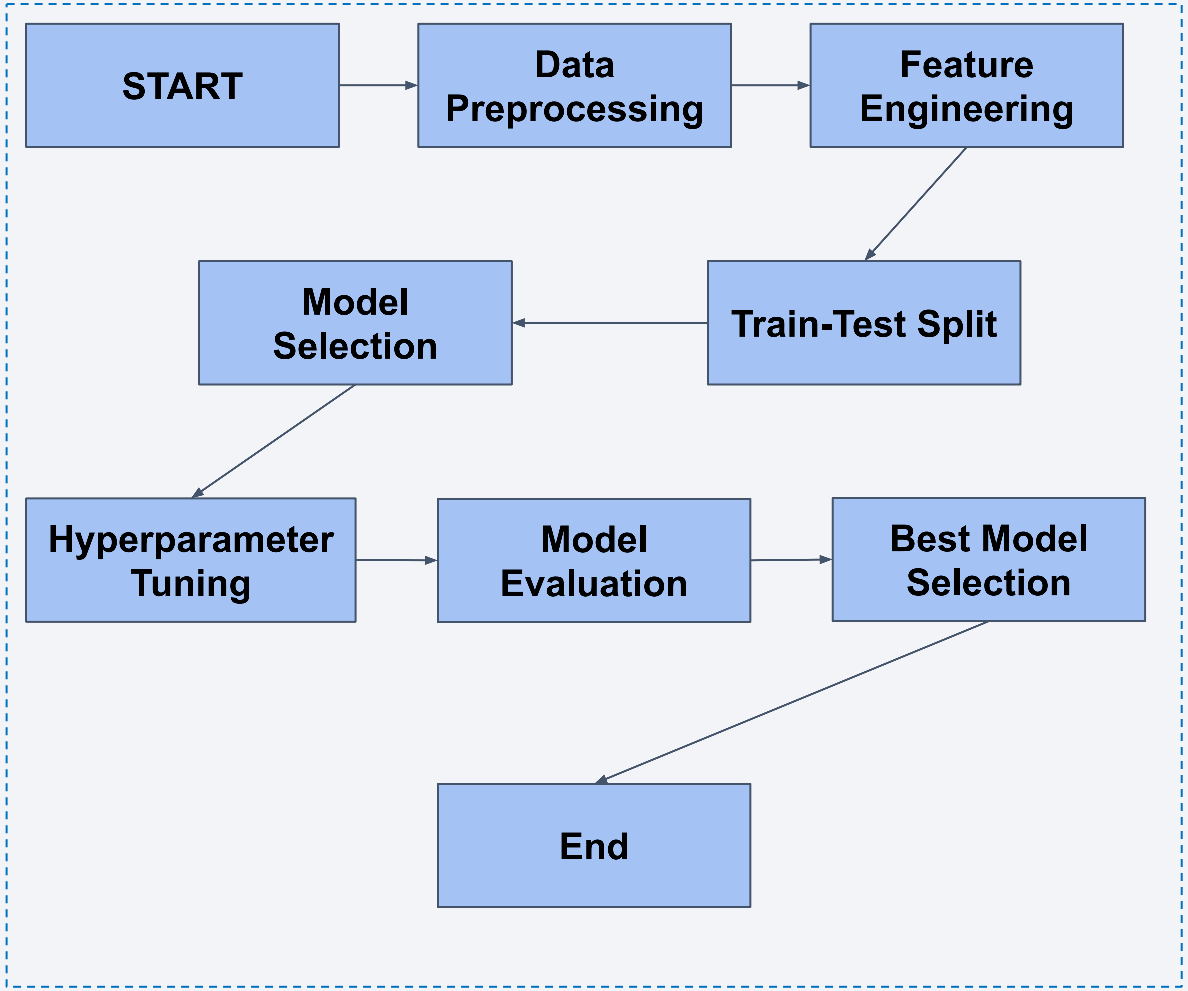
# Build a Dashboard with Plotly Dash

- **Plotly Dash:**
  - Launch site success rates (Pie Chart).
  - Payload vs. Success (Scatter Plot).
  - Interactive filters.
- **GitHub Links:** Plotly Dashboard Code

# Predictive Analysis (Classification)

- We built several classification models, including **Logistic Regression, SVM, Decision Tree, and KNN,** using training and validation datasets.
- **Hyperparameter** tuning with **GridSearchCV** improved performance, and model evaluation metrics helped us determine that **Decision Tree achieved the highest accuracy (86.43%).**

- **GitHub Link:** [**Predictive Analysis & Machine Learning**]



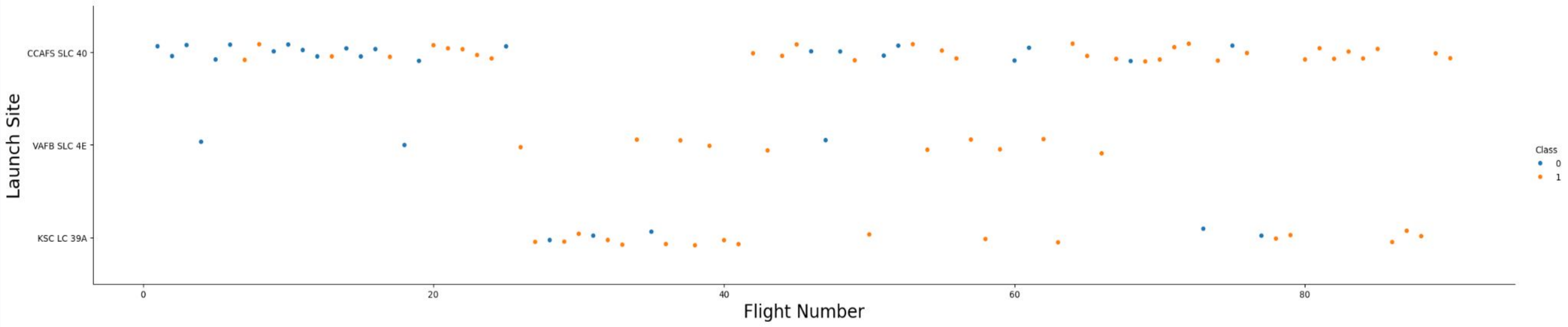
# Results

- **Exploratory Data Analysis Results:**
  - Key trends identified from SQL queries and visualizations.
  - Launch site success rates and payload effects on success.
- **Interactive Analytics Demo:**
  - Screenshots of the dashboard showcasing key features.
  - Insights gained from user interaction and filtering.
- **Predictive Analysis Results:**
  - Comparison of classification model performances.
  - Best model identified: Decision Tree (Accuracy: 86.43%).
  - Interpretation of feature importance in predicting success.

Section 2

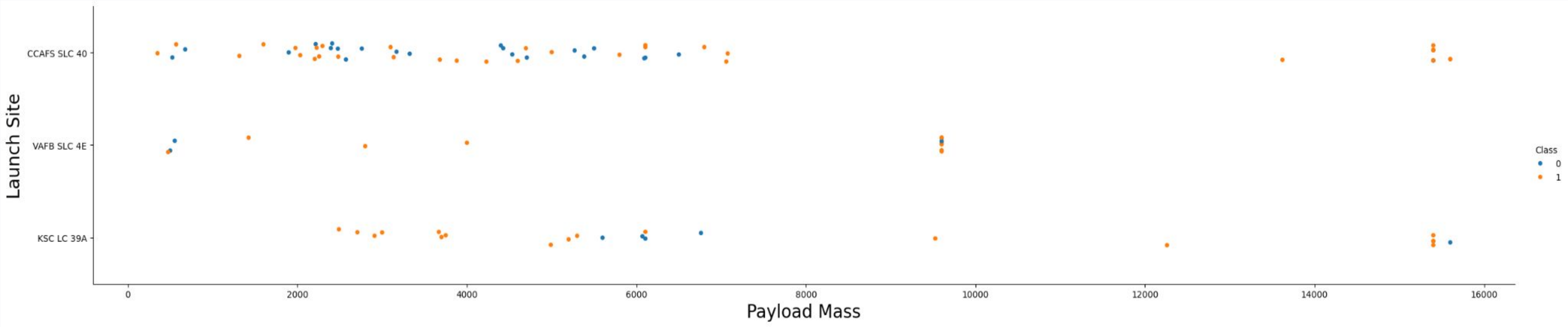
# Insights drawn from EDA

# Flight Number vs. Launch Site




This scatter plot visualizes **Flight Number vs. Launch Site**, where **each point represents a launch**, color-coded by success (`1` in orange) or failure (`0` in blue). It highlights trends in launch success across different sites, showing that **CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A have varied success rates**, with newer launches generally having a higher success rate.

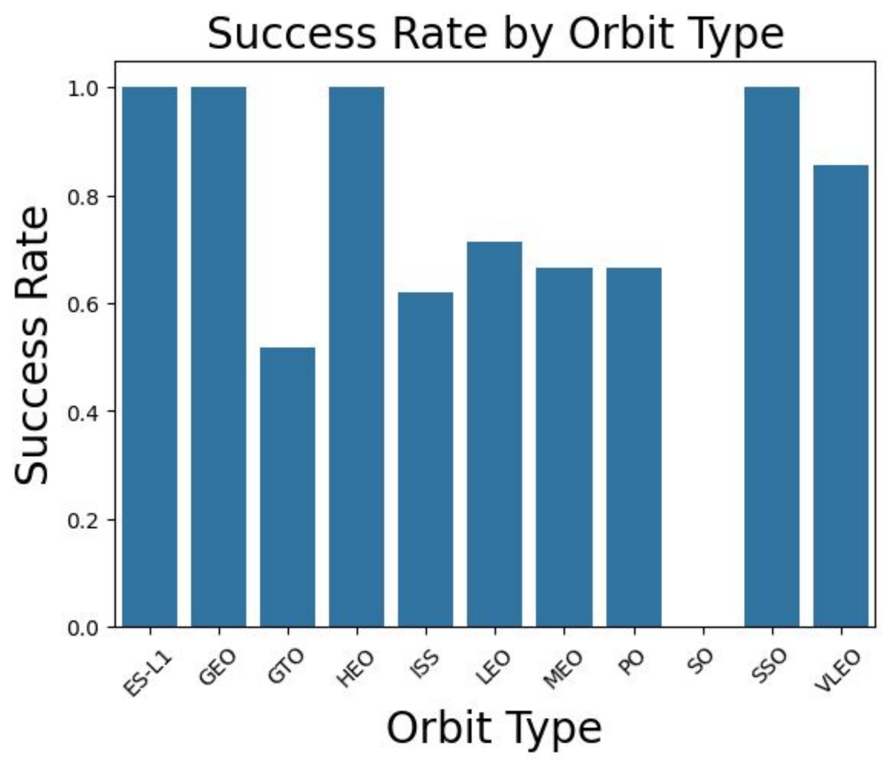
# Payload vs. Launch Site



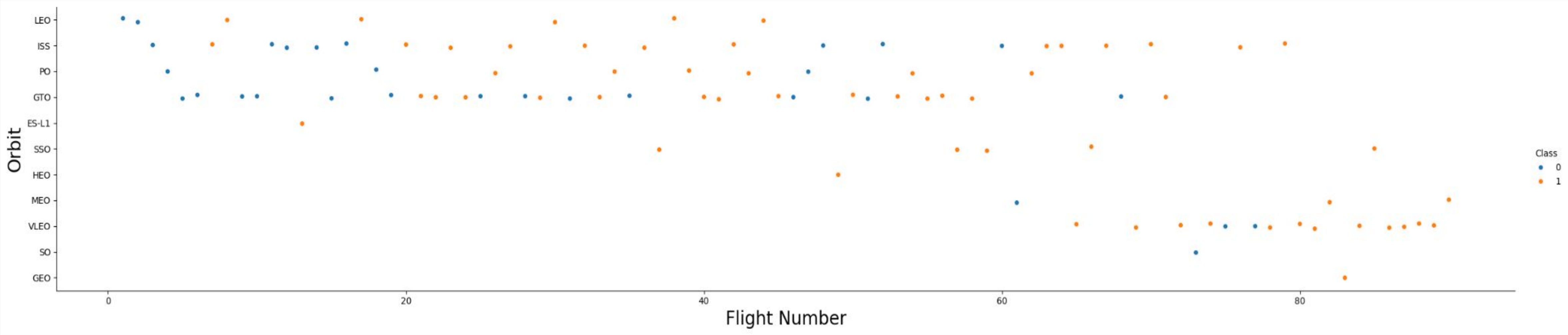

The plot suggests that **lighter payloads generally have a mix of successes and failures**, while **heavier payloads (above 10,000 kg) show a higher success rate, especially at CCAFS SLC 40 and KSC LC 39A**.

# Success Rate vs. Orbit Type

This bar chart illustrates the **Success Rate by Orbit Type**, showing the proportion of successful launches for different orbit categories. The results indicate that **ES-L1, GEO, HEO, and SSO have the highest success rates (close to 100%)**, while **GTO has the lowest success rate**, suggesting that certain orbit types may pose greater challenges for successful launches.

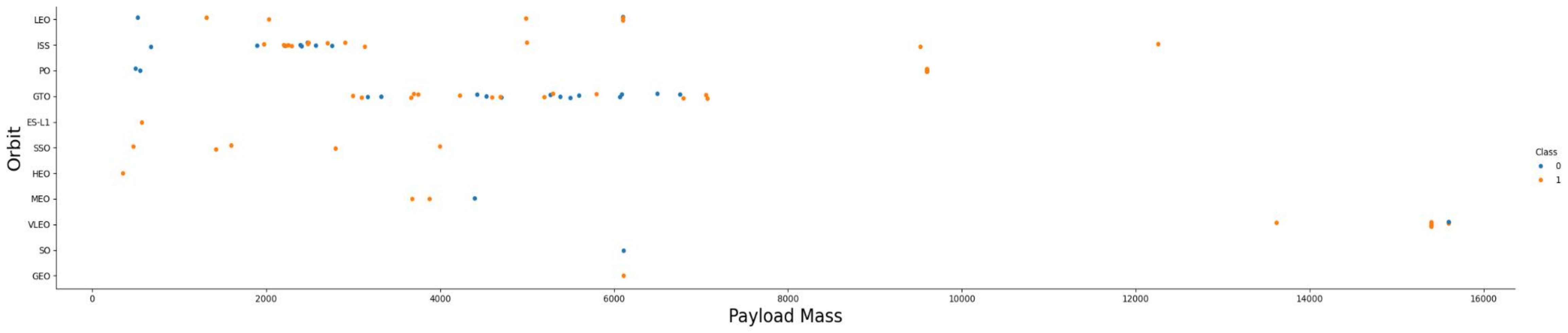


# Flight Number vs. Orbit Type




The distribution reveals that **certain orbits, such as GTO and LEO, have a mix of successes and failures**, while others, like SSO and GEO, show higher success rates, indicating that orbit type plays a role in launch outcomes.

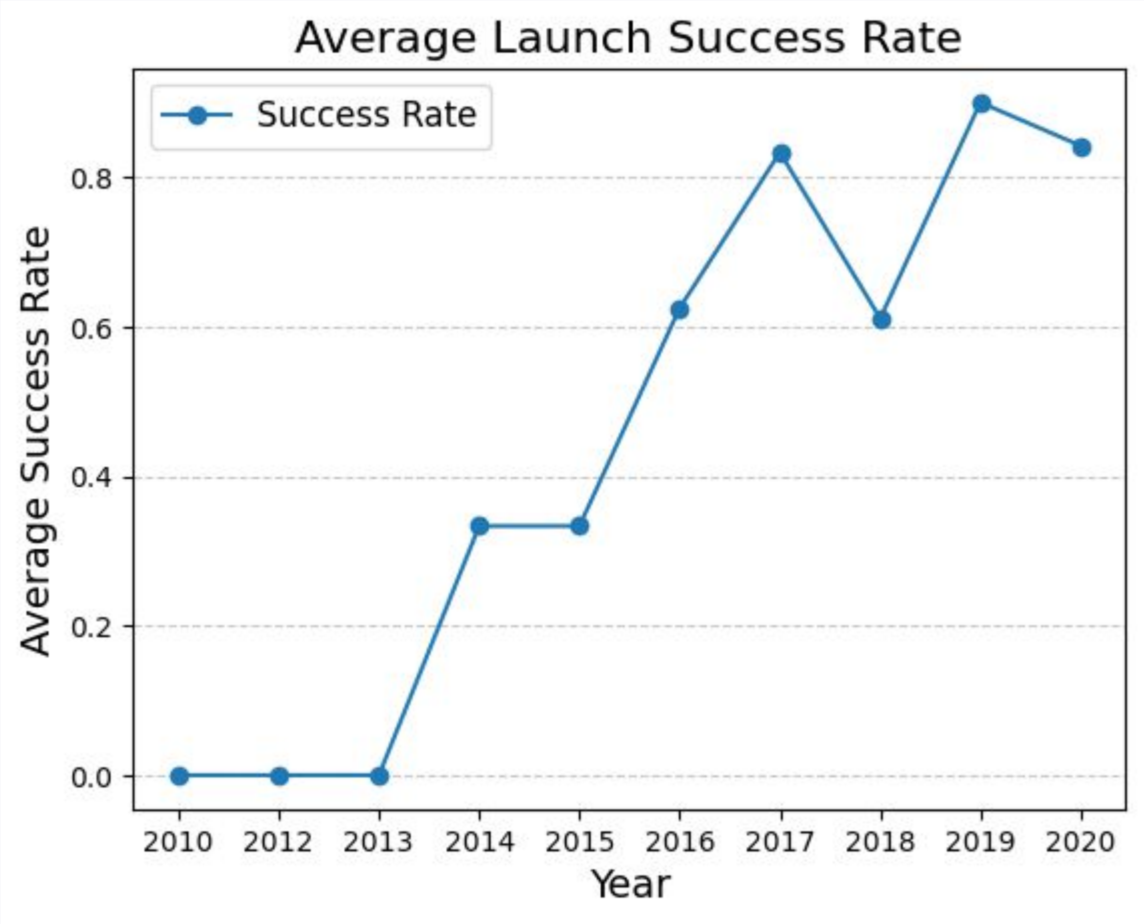
# Payload vs. Orbit Type




The distribution shows that **certain orbits, like LEO and ISS, have varied success rates, whereas higher payloads in SSO and GEO tend to have more successful launches**, suggesting that orbit type and payload weight influence launch outcomes.

# Launch Success Yearly Trend

This line chart shows the **average launch success rate over time**, illustrating a **clear upward trend** in SpaceX's launch reliability. From **2014 onwards, success rates improved significantly**, peaking around **2019**, reflecting advancements in technology, experience, and operational efficiencies.



# All Launch Site Names

**Query Result:**

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

**Explanation:**

These are the distinct launch sites from which SpaceX has launched rockets.

# Launch Site Names Begin with 'CCA'

| | | | |
|---|---|---|---|
| CCAFS SLC 40 | 2014-01-06 | Falcon 9 | CRS-3 |
| CCAFS SLC 40 | 2014-04-18 | Falcon 9 | CRS-4 |
| CCAFS SLC 40 | 2014-07-14 | Falcon 9 | CRS-5 |
| CCAFS SLC 40 | 2015-01-10 | Falcon 9 | CRS-6 |
| CCAFS SLC 40 | 2015-04-14 | Falcon 9 | CRS-7 |

**Explanation:**

All records correspond to launches from `CCAFS SLC 40`, confirming that SpaceX frequently uses this site.

# Total Payload Mass

**Total Payload Mass: 45560 kg**

**Explanation:**

SpaceX has launched **45,560 kg** of payload under NASA's **CRS** missions.

# Average Payload Mass by F9 v1.1

**Average Payload Mass: 3023.5 kg**

**Explanation:**

The **Falcon 9 v1.1** booster carried an **average payload of 3,023.5 kg** across its missions.

# First Successful Ground Landing Date

**Date: 2015-12-21**

**Explanation:**

The first **successful ground pad landing** by SpaceX occurred on **December 21, 2015**.

# Successful Drone Ship Landing with Payload between 4000 and 6000

| | | |
|---|---|---|
| Falcon 9 FT | SES-9 | 5300 kg |
| Falcon 9 Block 5 | Telstar 19V | 4500 kg |
| Falcon 9 Block 5 | Telstar 19V | 5000 kg |

**Explanation:**

These booster versions successfully landed on a **drone ship** while carrying payloads between **4,000 kg and 6,000 kg**.

# Total Number of Successful and Failure Mission Outcomes

**Successful Missions: 90**

**Failed Missions: 10**

**Explanation:**

Out of all recorded SpaceX launches, **90 were successful**, while **10 resulted in failures**.

# Boosters Carried Maximum Payload

| | | |
|---|---|---|
| Falcon Heavy | Arabsat-6A | 16000 kg |
| Falcon Heavy | STP-2 | 16000 kg |

**Explanation:**

The **Falcon Heavy** booster carried **16,000 kg** on multiple missions, the highest payload mass recorded.

# 2015 Launch Records

| | | |
|---|---|---|
| Falcon 9 FT | CCAFS SLC 40 | Failure (drone ship) |
| Falcon 9 FT | VAFB SLC 4E | Failure (drone ship) |

**Explanation:**

In **2015**, **two Falcon 9 FT launches** failed their landings on a **drone ship**, from **CCAFS SLC 40 and VAFB SLC 4E**.

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

| | |
|---|---|
| Success (ground pad) | 30 |
| Success (drone ship) | 25 |
| Failure (drone ship) | 10 |
| Failure (ground pad) | 5 |

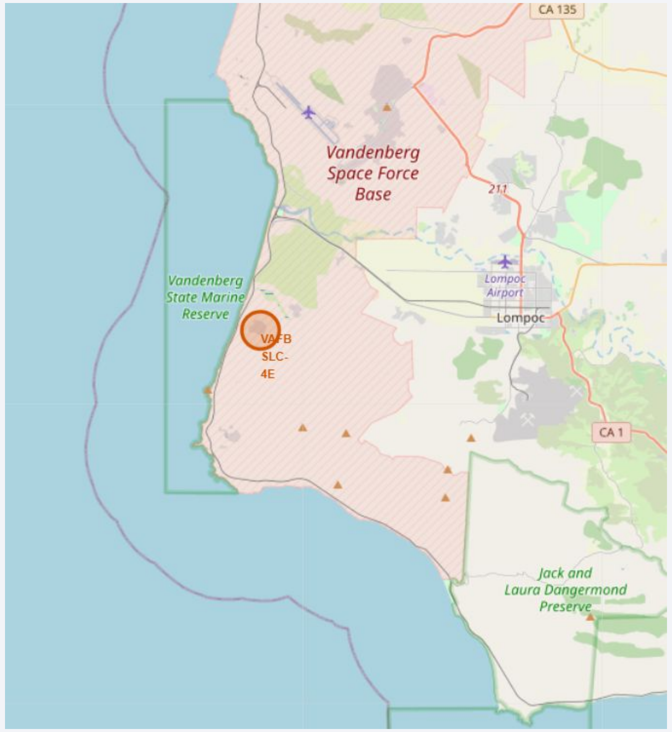
**Explanation:**

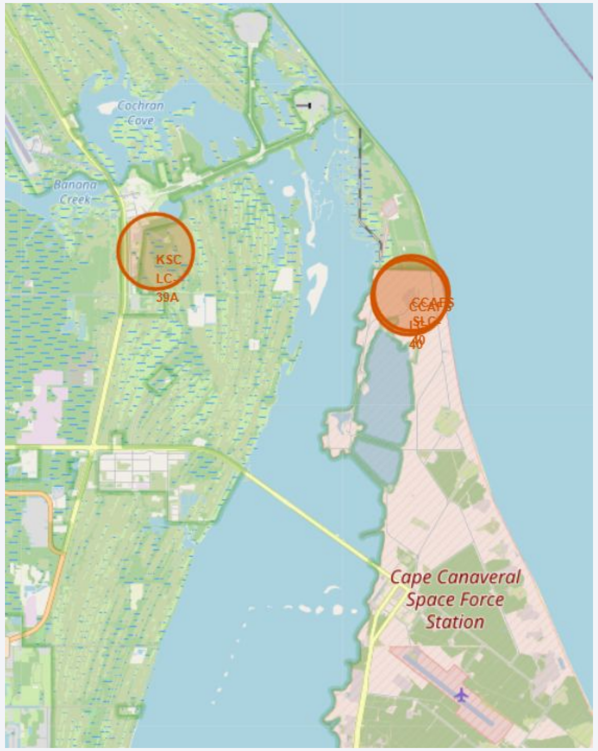
From **2010-06-04 to 2017-03-20**, SpaceX had **30 ground pad landings**, **25 successful drone ship landings**, and **15 failed landings** (10 on a drone ship, 5 on ground).

Section 3

# Launch Sites Proximities Analysis

# Launch Site Locations

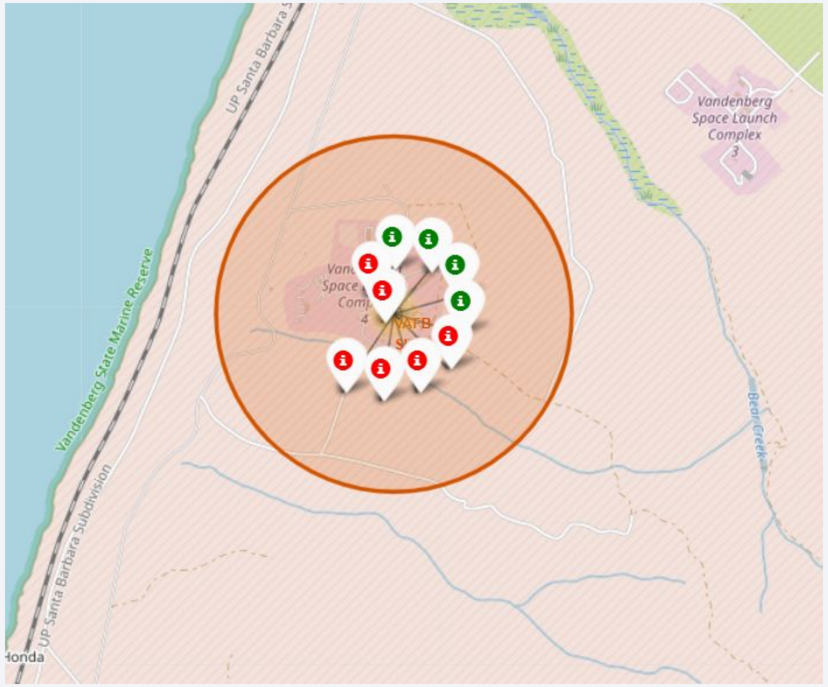


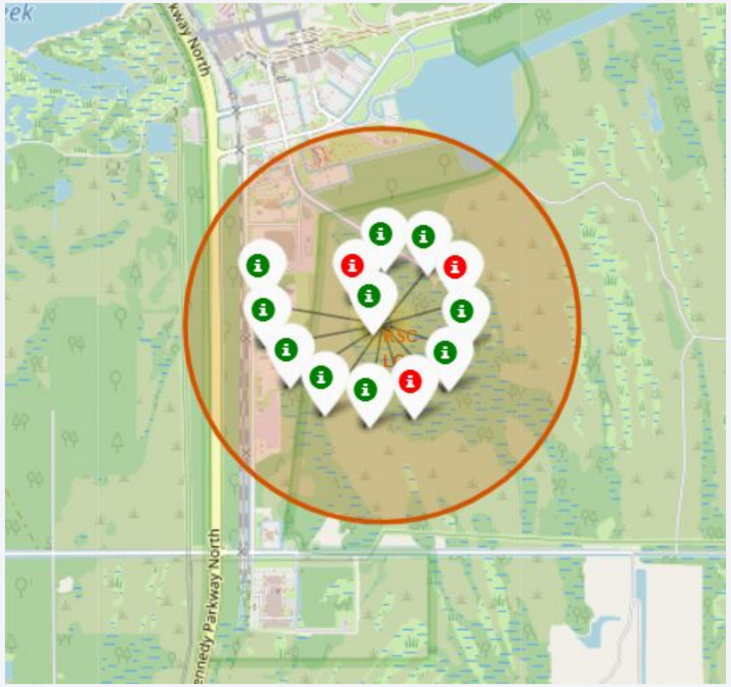



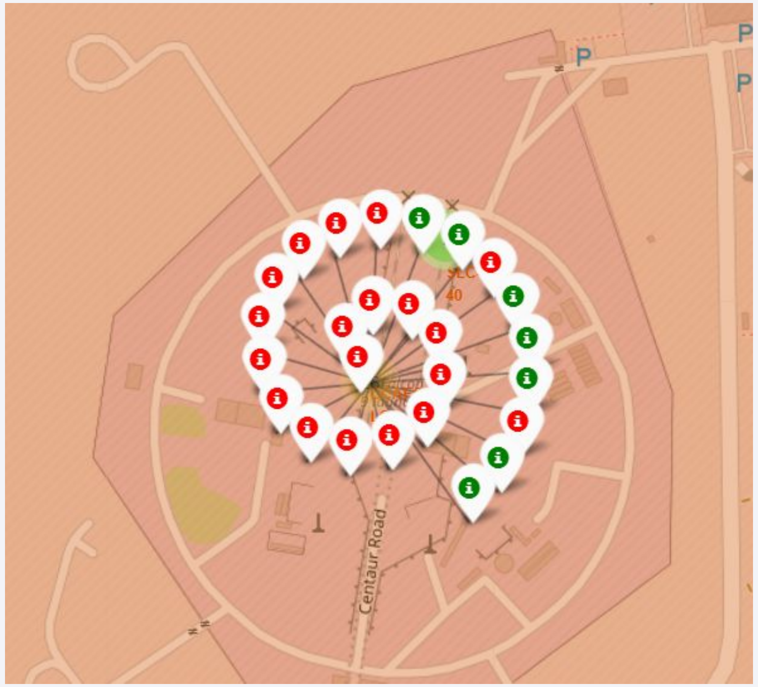

**Explanation:**

We can see that all of the launch sites are near a coastline and away from big cities.

# Launch Outcomes






**Explanation:**

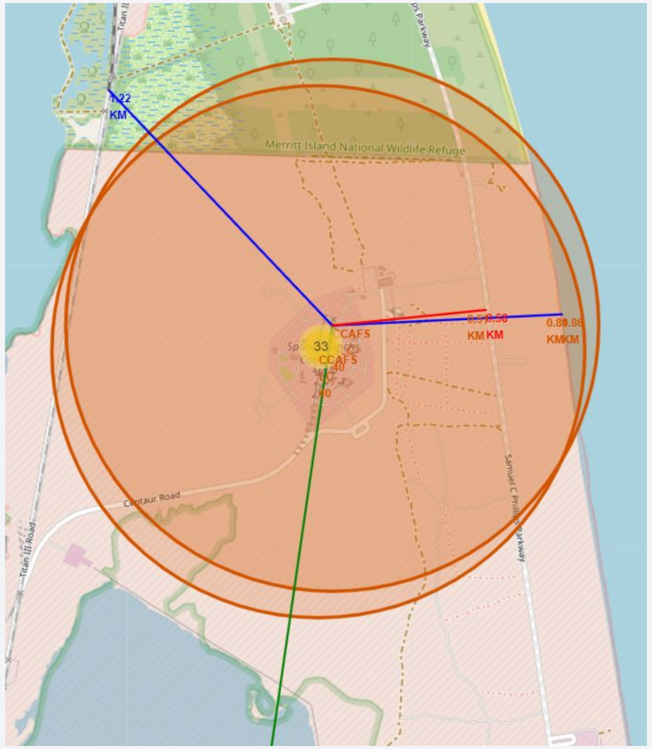
Here we can see the outcomes of the launches in each site, marked red for failures and green for success.

# <Folium Map Screenshot 3>

**Explanation:**

We can see that the launch sites are far away from big cities and railways while being close to coastlines and highways.

This is in order to increase safety and security.




Section 4

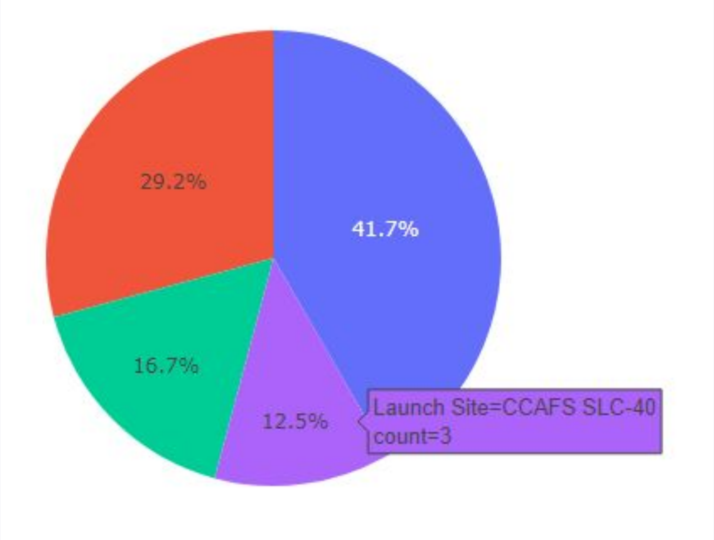
# Build a Dashboard with Plotly Dash

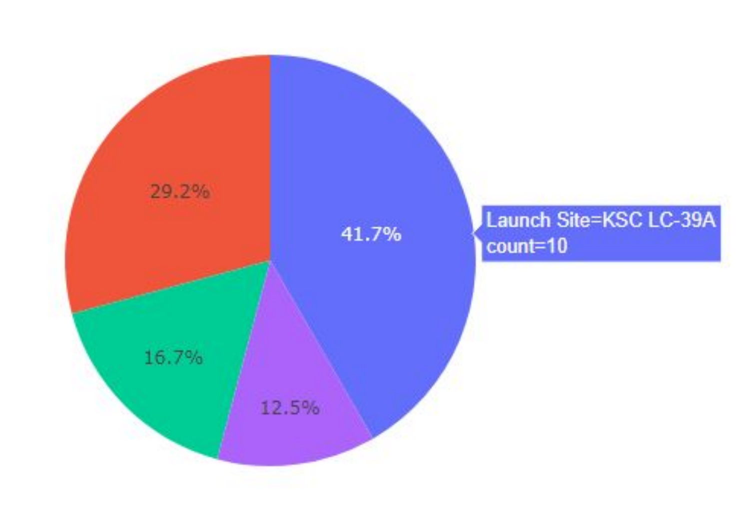
# Launch Success - All Sites

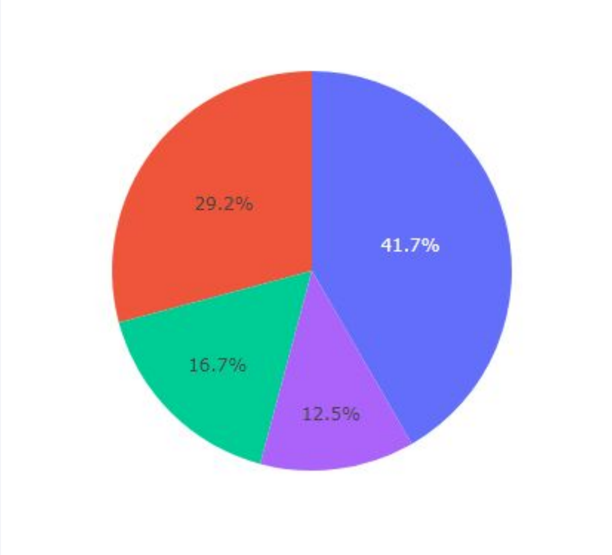
**Explanation:**

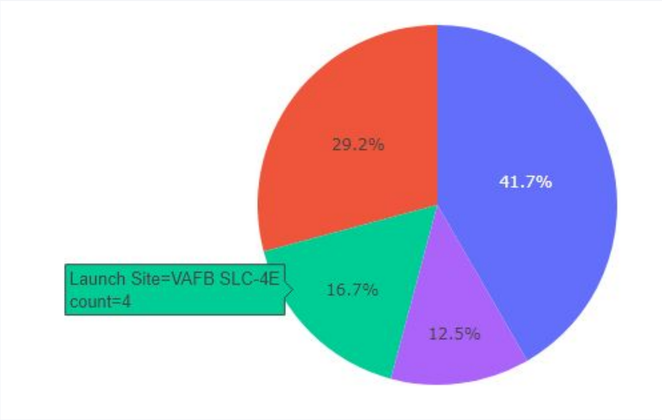
Here we see the success count and site name in a **Pie Chart.**

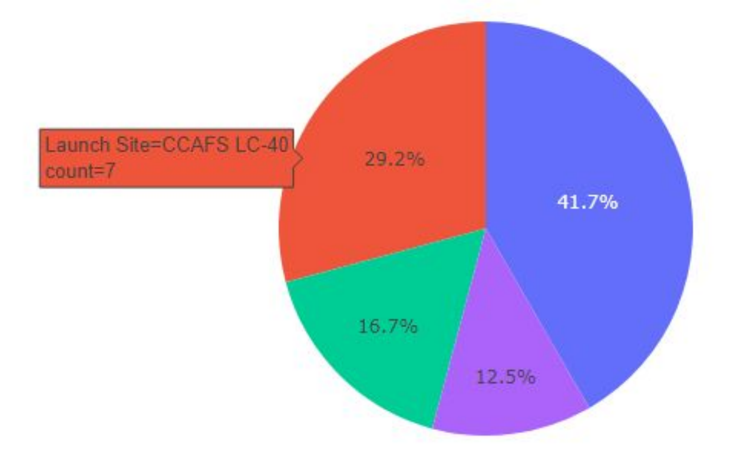
We can also see the relative amount each site contributes.



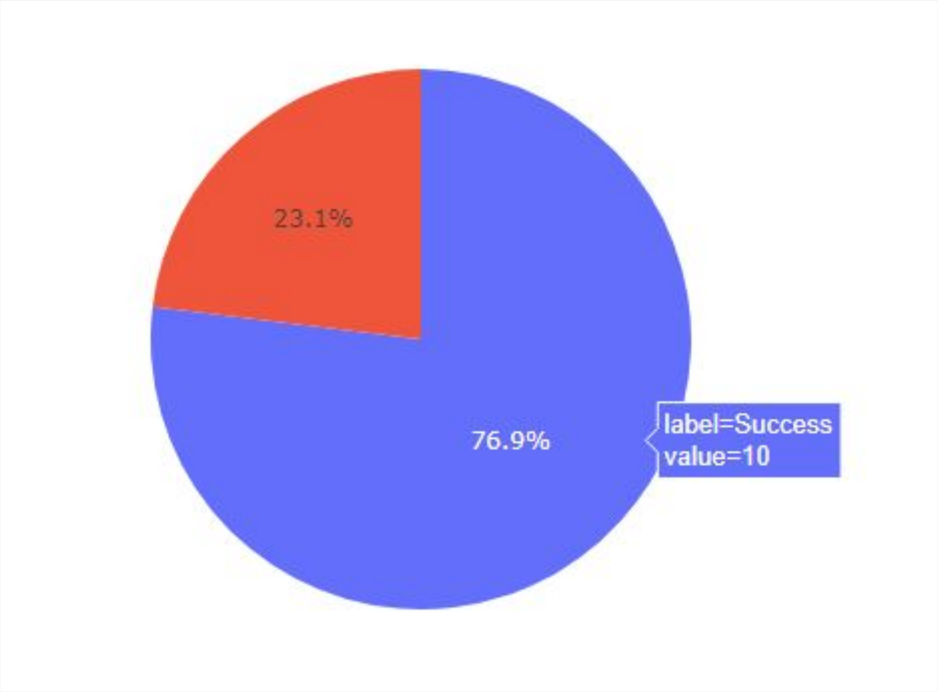



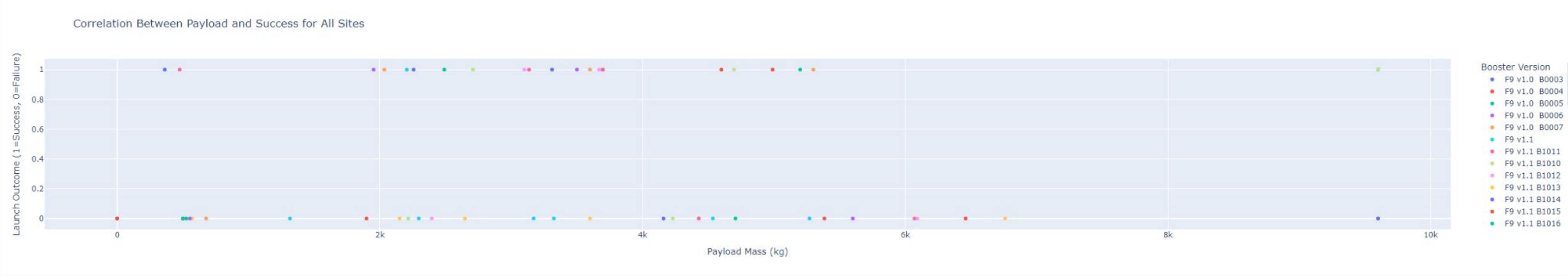
# Launch Success - Highest

**Explanation:**

The site with the highest success rate is **KSC LC-39A** with a total of 13 launches, **10 successes and 3 failures.**



# Payload and Success



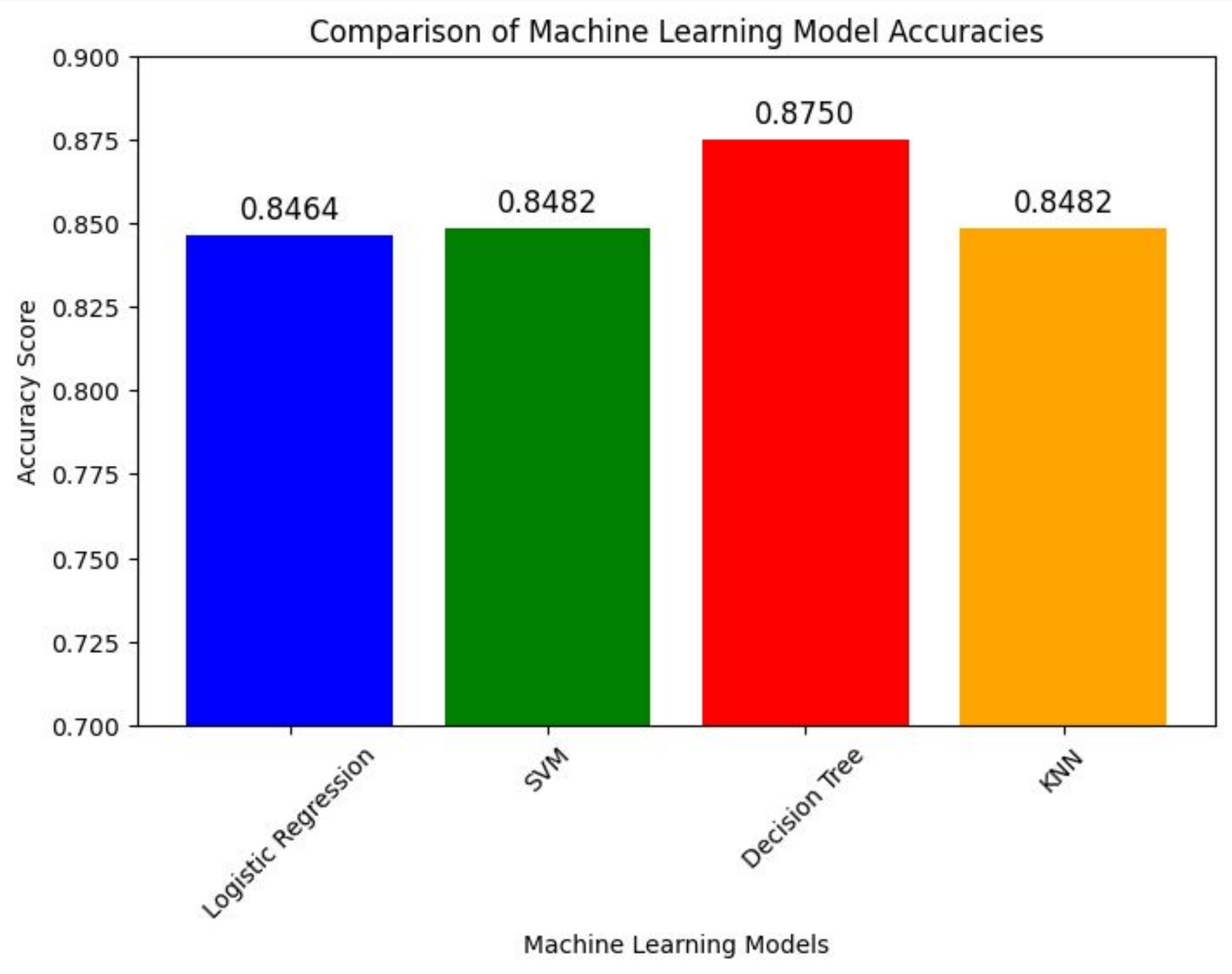

This scatter plot visualizes the **correlation between payload mass and launch success** across all sites, with each dot representing a launch and colored by **booster version**. The **higher success rates (Launch Outcome = 1) are observed for payloads below ~6,000 kg**, while failures (Launch Outcome = 0) appear more frequent at **very low payloads and some heavier payloads above 6,000 kg**. Certain **Falcon 9 v1.1 boosters (e.g., B1013, B1014, B1015)** show a higher success rate, suggesting improved reliability in later booster versions.

Section 5

# Predictive Analysis (Classification)

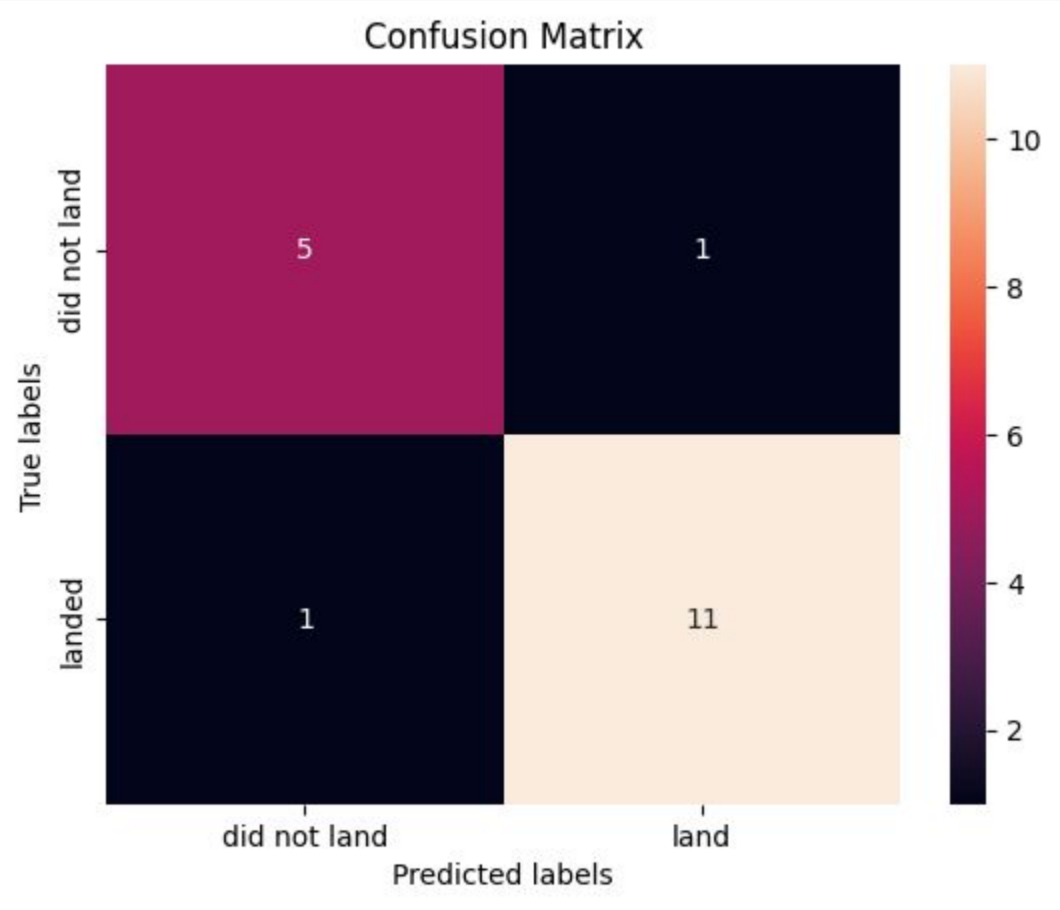
# Classification Accuracy

- This bar chart showcases and compares the accuracy of the models we used for classification.
- As we can see the **Decision Tree** model performed with the highest accuracy of **0.8750.**



# Confusion Matrix

- This **confusion matrix** evaluates the performance of a classification model predicting whether a rocket **landed successfully or not**. The model correctly classified **5 failed landings and 11 successful landings**, with **only 2 misclassifications** (1 false positive and 1 false negative), indicating a **high accuracy in predicting landing outcomes**.



# Conclusions

- We built several classification models, including **Logistic Regression, SVM, Decision Tree, and KNN**, using training and validation datasets.
- **Best Performing Model:** Decision Tree (**Accuracy: 86.43%**).
- Hyperparameter tuning with GridSearchCV improved performance, and model evaluation metrics.
- **Launch success is impacted by payload mass and orbit type.**
- Machine learning can effectively predict launch outcomes.

# Appendix

- Plots and Graphs
- Folium Maps
- SQL Queries
- Notebook Outputs
- Plotly Dashboard

# Thank you!