# *SpaceX Falcon 9 First Stage Landing Prediction*

# Optimizing Falcon 9 Launch Outcomes with Data Science

**Alexander Johnfernandez**

**Date : 27th April 2025**

# Outline

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

# Executive Summary

- **Summary of Methodologies:**
  - Data was collected using a RESTful API and web scraping techniques.
  - Preprocessing and data cleaning were performed to prepare the dataset.
  - Exploratory Data Analysis (EDA) was conducted to identify patterns and insights.
  - Machine learning models (Logistic Regression, SVM, Decision Trees, KNN) were applied to Falcon 9 first-stage landing success.

- **Summary of All Results:**
  - All models (Logistic Regression, SVM, KNN, Decision Trees) achieved an accuracy of 0.83, i equal performance across all models.
  - No single model outperformed the others, meaning the dataset does not favor a specific

# Introduction

- This project aims to predict the success of Falcon 9's first-stage landing.
- SpaceX offers Falcon 9 rockets at a cost of $62 million, while other companies charge upwards each launch.
- A significant portion of SpaceX's savings comes from the ability to reuse the first stage, allowir lower prices compared to competitors.

**Problems You Want to Find Answers:**

- Can we predict the success of the Falcon 9 first-stage landing?
- What factors influence the success of the first-stage landing?
- How can this prediction be used to improve pricing strategies for competing companies?

# Section 1
# Methodology

# Methodology

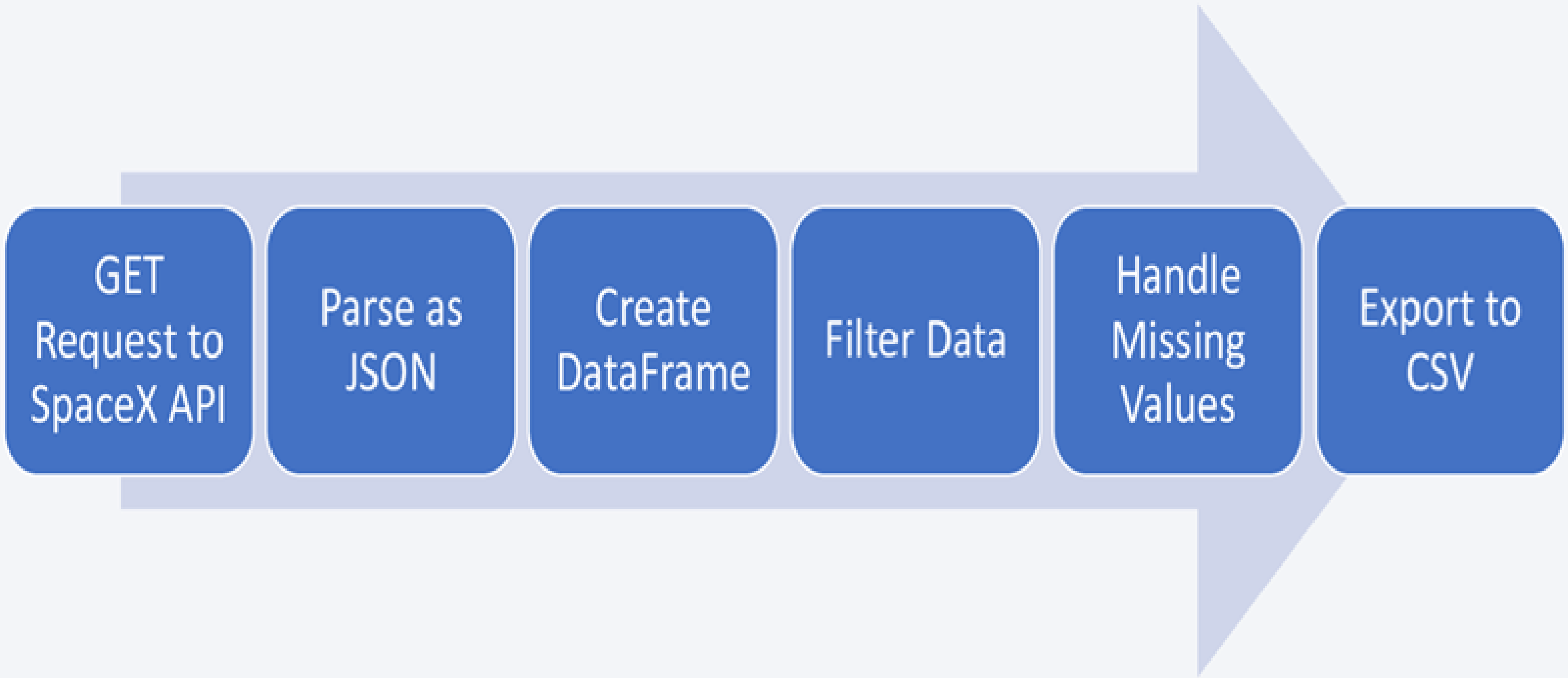
## Executive Summary

- Data collection methodology:
    - Fetched Falcon 9 and Falcon Heavy launch data using web scraping and SpaceX API.
- Perform data wrangling
    - Perform exploratory data analysis (EDA) and determine appropriate training labels.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

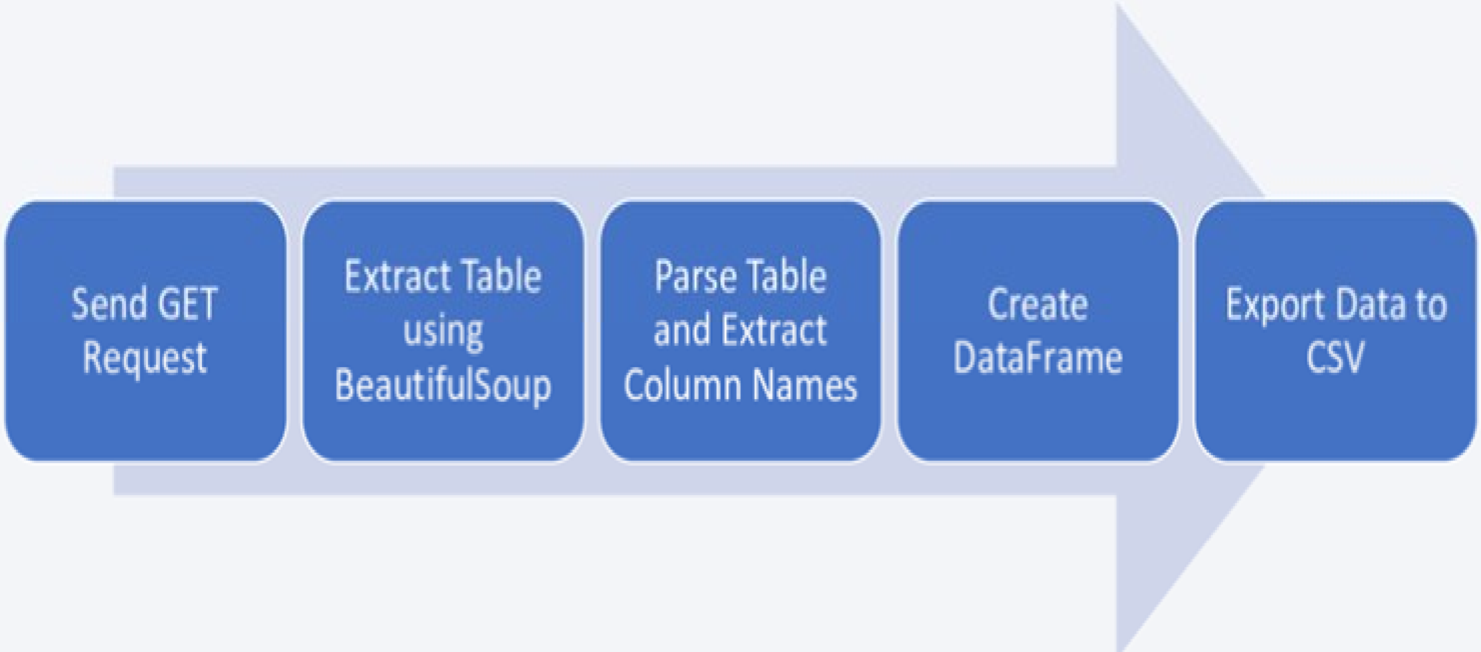
# Data Collection

In this step, we collect data on Falcon 9 rocket launches using two methods: the **SpaceX API** and
API provides launch data in JSON format, while Web Scraping allows us to extract structured dat
on relevant web pages for further analysis.

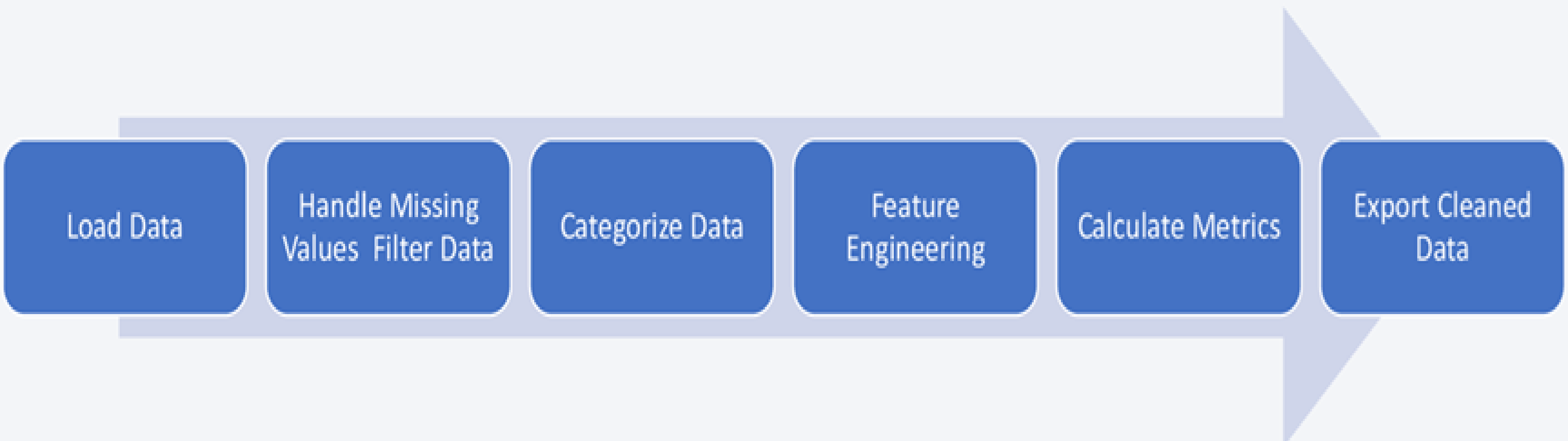
# Data Collection – SpaceX API



# Data Collection - Scraping



# Data Wrangling



# EDA with Data Visualization

- **Flight Number vs. Launch Site (Scatter Plot):**
  - Purpose: To explore the distribution of launches across different launch sites.
- **Payload Mass vs. Launch Site (Scatter Plot):**
  - Purpose: To analyze how payload mass correlates with launch sites.
- **Success Rate by Orbit Type (Bar Chart):**
  - Purpose: To compare success rates across different orbit types.
- **Flight Number vs. Orbit Type (Scatter Plot):**
  - Purpose: To examine the relationship between flight experience and success rates in different orbits.
- **Payload Mass vs. Orbit Type (Scatter Plot):**
  - Purpose: To understand the impact of payload mass on launch success by orbit type.
- **Yearly Launch Success Trend (Line Chart):**
  - Purpose: To track the success trend of launches over the years.

# EDA with SQL

- **Unique Launch Sites:** Find names of unique launch sites.
- **Launch Sites Starting with 'CCA':** Retrieve 5 records where launch sites begin with 'C
- **Payload by NASA Boosters:** Calculate total payload carried by NASA boosters.
- **Payload for F9 v1.1:** Calculate average payload mass for F9 v1.1 booster version.
- **First Successful Landing Date:** List date of the first successful landing on ground pad.
- **Boosters Landing on Drone Ship:** List boosters that successfully landed on drone shi

between 4000 and 6000 kg

# EDA with SQL (Continued)

- **Mission Outcomes Count:** Count total successful and failed mission outcomes.
- **Max Payload Booster Versions:** List booster versions with maximum payload mass.
- **Failure Landing Outcomes in 2015:** Retrieve records of failed drone ship landings in 2015.
- **Landing Outcomes Ranking (2010-06-04 to 2017-03-20):** Rank landing outcomes within the specified date range.

# Build an Interactive Map with Folium

- **Markers:** Accurately identify launch site locations, helping to visualize their geographical positions.
- **Circles:** Show the influence areas around the sites, highlighting nearby infrastructure like coastlines and highways.
- **Lines:** Display the distances between launch sites and key proximities, making spatial relationships clearer.
- **Popups:** Provide additional information on launch success or failure, enhancing interactivity with the data.

# Build a Dashboard with Plotly Dash

📊 **Pie Chart:** Displays the launch success rate for each site to compare performance.

📉 **Scatter Plot:** Analyzes the relationship between payload and launch success rate.

🔽 **Dropdown List:** Allows selecting a launch site for individual analysis.

📏 **Range Slider:** Enables selecting a payload range to study its impact on success.

**Importance of These Interactions:**

- Identify the most successful launch sites.
- Analyze the impact of payload on success.
- Evaluate Falcon 9 booster performance.
- Provide a flexible and interactive analysis experience.

# Predictive Analysis (Classification)

The process began with data preprocessing, which involved creating a class label column and standardizing the features. The dataset was split into 80% for training and 20% for testing. Multiple models, including SVM, KNN, Classification Trees, and Logistic Regression, were trained. Hyperparameters were tuned using GridSearchCV. Upon evaluation, all models achieved similar performance, with an accuracy of 0.83.

# Results

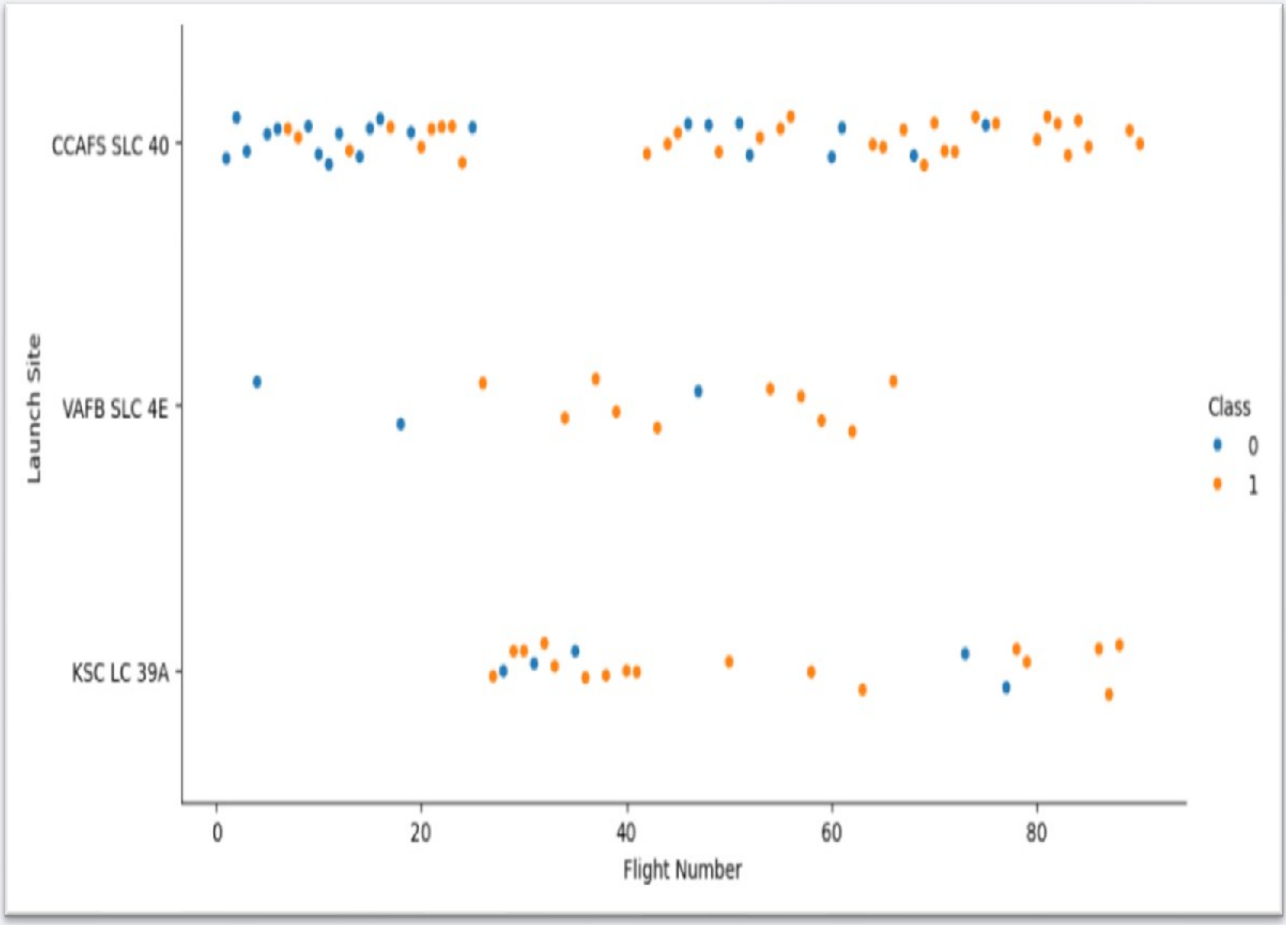
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Skills Network

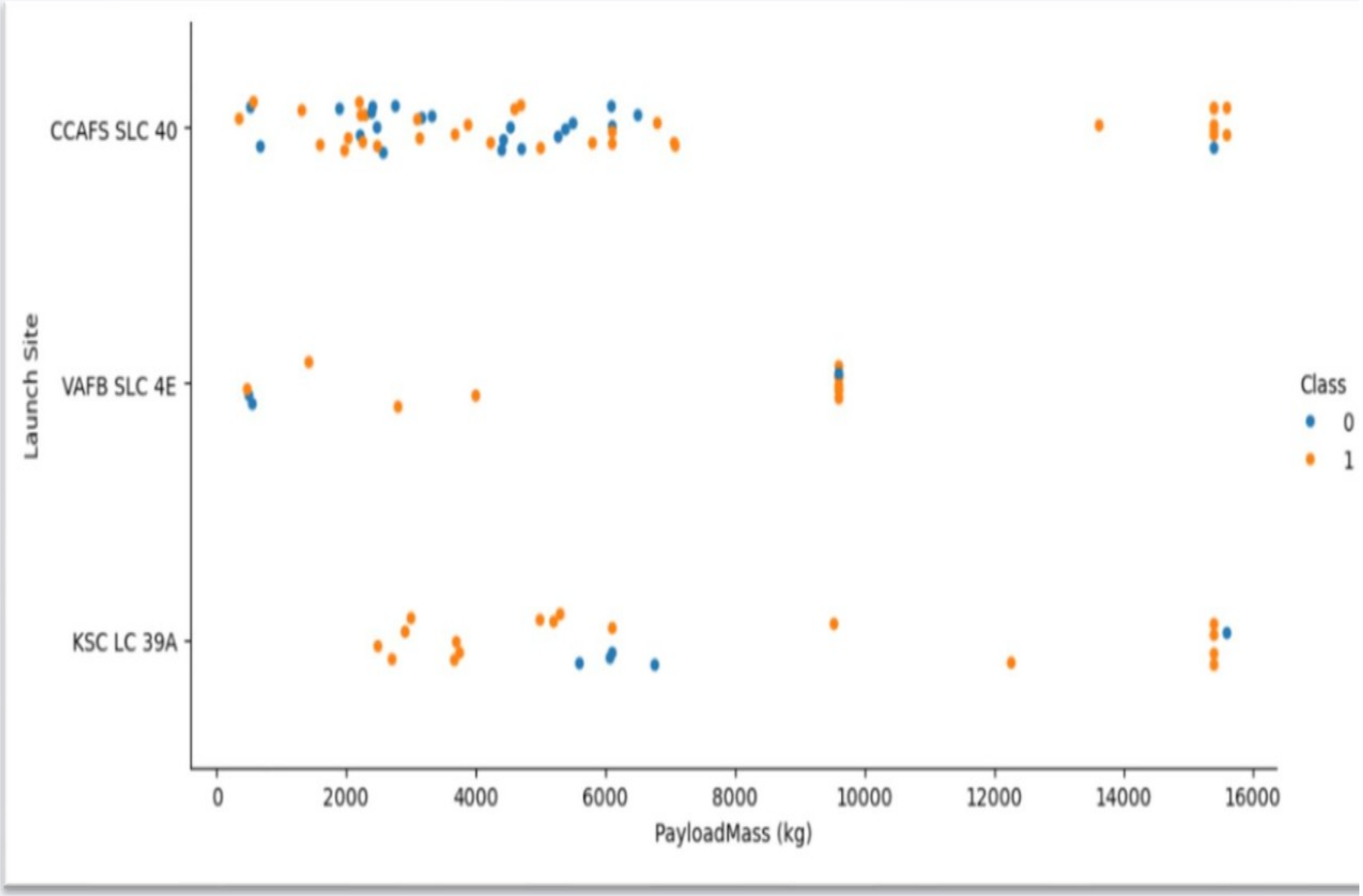
# Section 2

# Insight drawn from EDA

# Flight Number vs. Launch Site

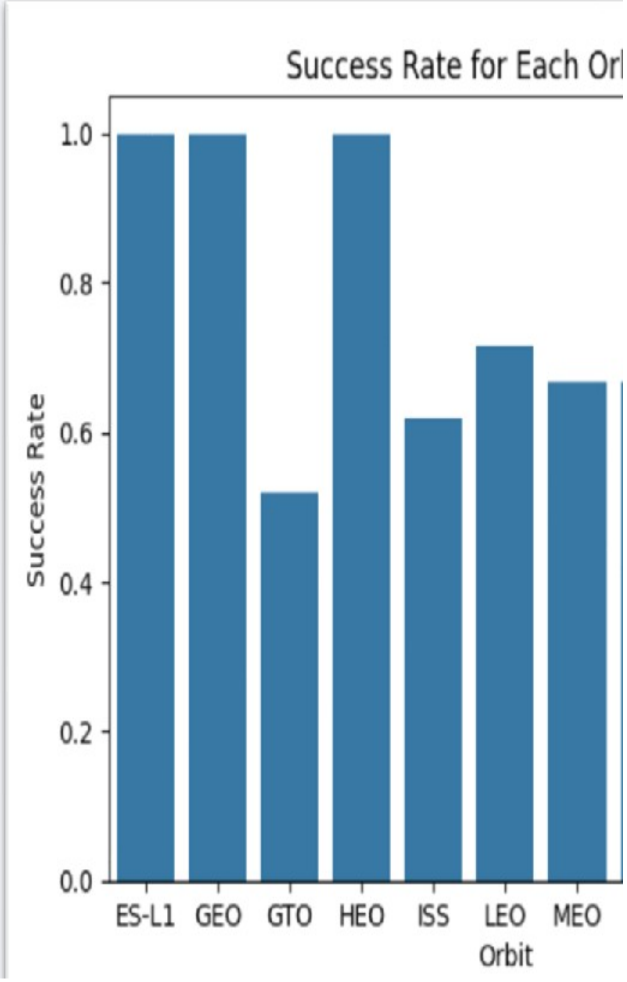


# Payload vs. Launch Site

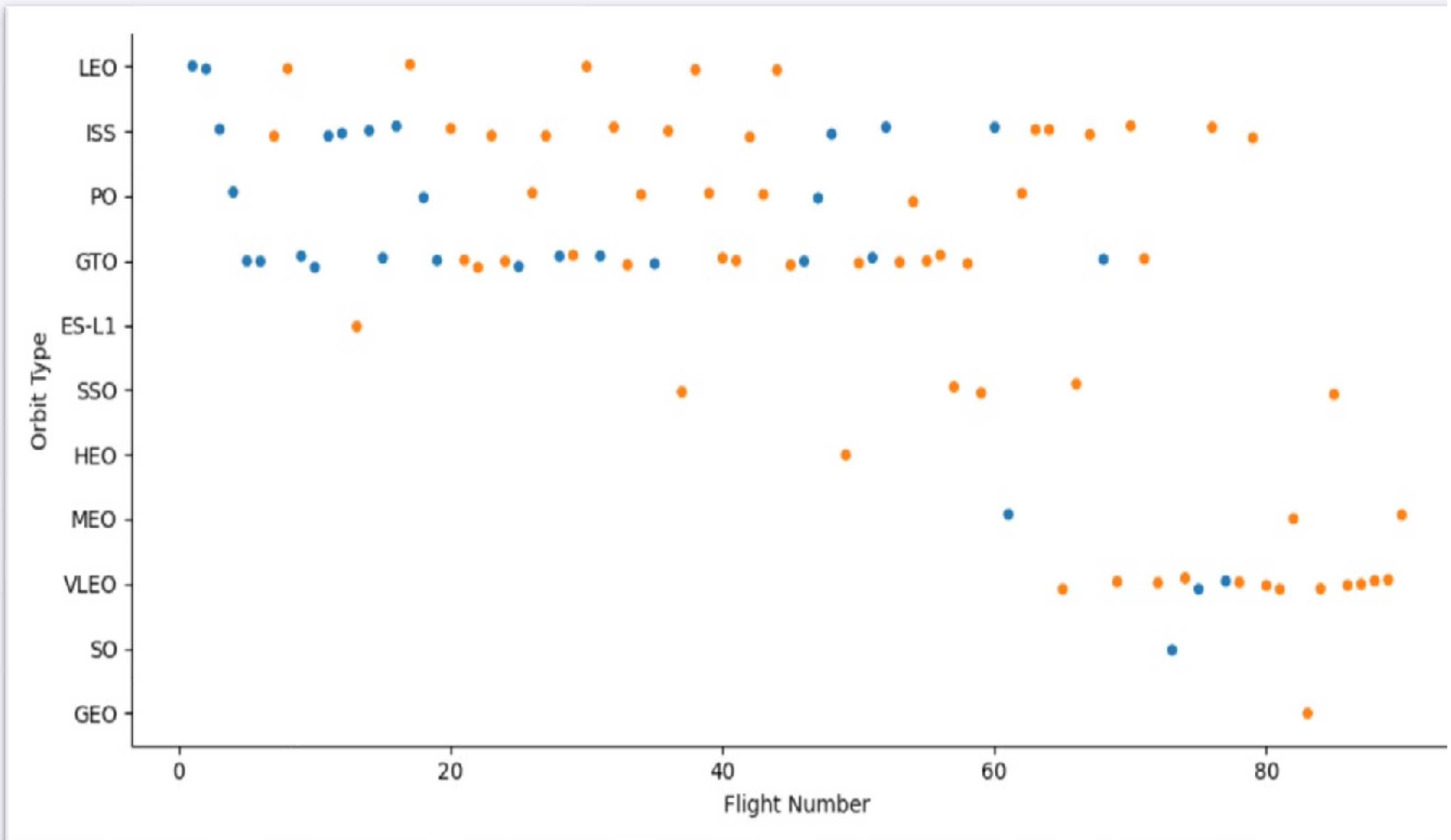


# Success Rate vs. Orbit Type

The bar chart shows that the **ES-L1, GEO, HEO,** and **SSO** orbits have the highest success rates in launches, indicating a correlation between the orbit type and launch success.

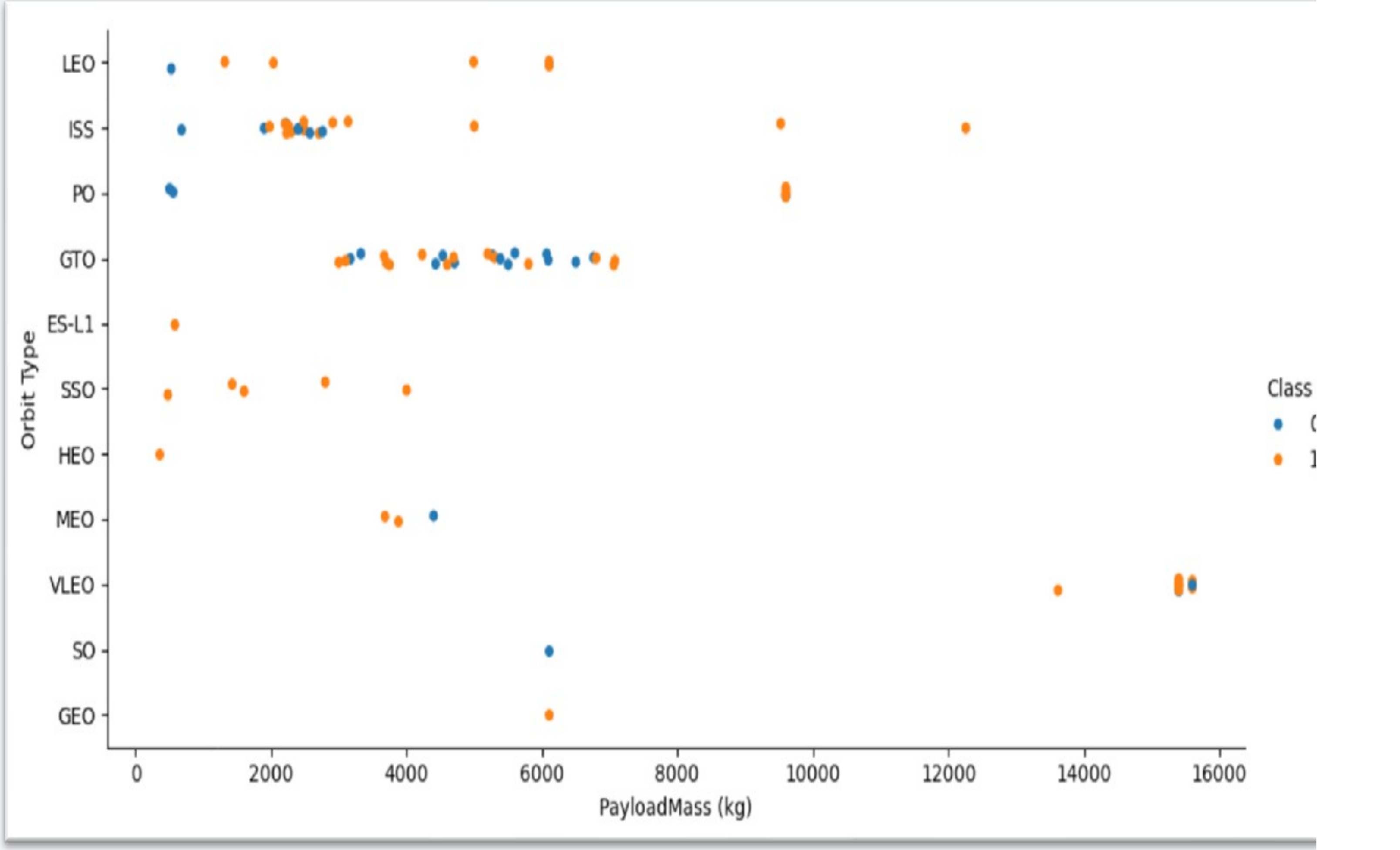


# Flight Number vs. Orbit Type



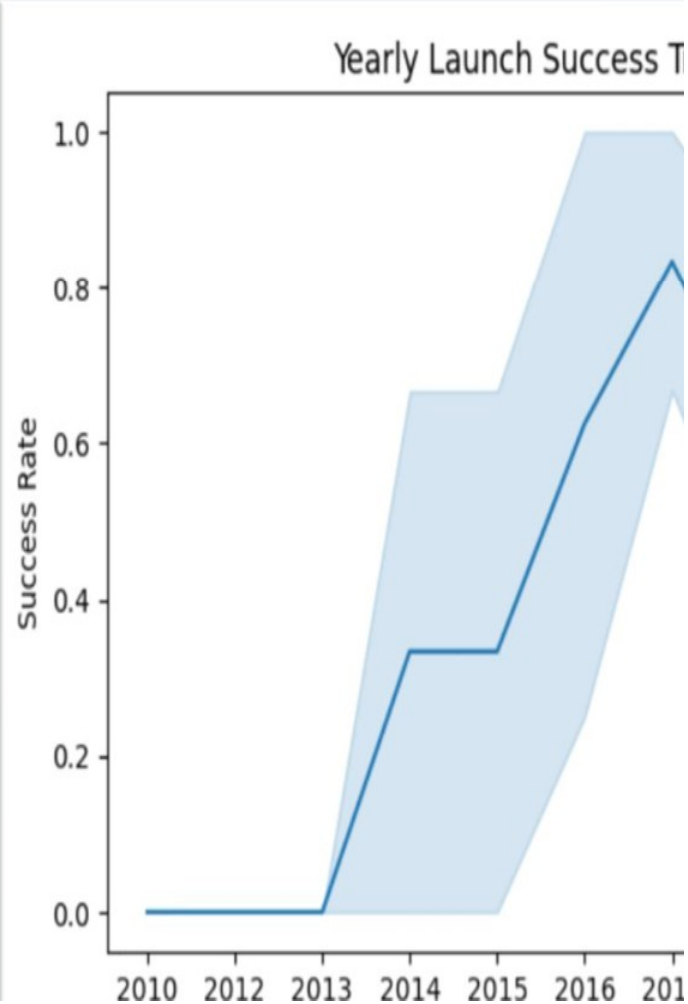

Scatter plot showing the relationship between flight number and orbit type. In the **LEO orbit**, succes

# Payload vs. Orbit Type



# Launch Success Yearly Trend

The line chart shows a continuous increase in the **success rate** from 2013 to 2020, reflecting a noticeable improvement in launch performance over time.



# All Launch Site Names

This SQL query retrieves the **unique launch site names** from the SPACEXTBL table and displays a
used by SpaceX for rocket launches.

```
%sql select distinct LAUNCH_SITE from SPACEXTBL;
```

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

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

# Launch Site Names Begin with 'CCA'

This SQL query retrieves 5 records from the SPACEXTBL table where the launch sites begin with '(
is used with the **'CCA%'** pattern to filter the records, and the **'LIMIT 5'** clause ensures only 5 resul

```
%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5;
```

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

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS_KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcom |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachut |
| 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 (parachut |
| | | | | Dragon | | | | | |

# Total Payload Mass

This SQL query calculates the **total** payload mass carried by boosters from NASA. The SUM functi
the PAYLOAD_MASS__KG_ values for all records where the CUSTOMER field contains **'NASA'** and
provides the total payload mass in kg.

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER like 'NASA%CRS%';
```

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

| SUM(PAYLOAD_MASS__KG_) |
| --- |
| 48213 |

# Average Payload Mass by F9 v1.1

This SQL query calculates the **average** payload mass for the F9 v1.1 booster by using the AVG fur
PAYLOAD_MASS__KG_ column, filtering for rows where the BOOSTER_VERSION is **'F9 v1.1'**. The

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION ='F9 v1.1';
```

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

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

# First Successful Ground Landing Date

This SQL query retrieves the earliest date when a successful landing outcome on the ground by using the MIN function on the DATE column, filtered by the LANDING_OUTCOME field for **pad)'.** The result shows the first successful landing date for this outcome.

```
%sql select MIN(DATE) from SPACEXTBL where LANDING_OUTCOME ='Success (ground pad)';
```

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

| MIN(DATE) |
|---|
| 2015-12-22 |

# Successful Drone Ship Landing with Payload between 4000

This SQL query retrieves the names of boosters that successfully landed on a drone ship with a payload mass
and 6000 kilograms. It filters for **LANDING_OUTCOME = "Success (drone ship)"** and **PAYLOAD_MASS__KG_**
specified range, then displays the matching booster versions.

```
[ ] %sql select BOOSTER_VERSION from SPACEXTBL where LANDING_OUTCOME is 'Success (drone ship)' and PAYLOAD_MASS__KG_>4000 and PAYLOAD_MASS__KG_<6000;
```

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

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

# Total Number of Successful and Failure Mission Outc

This SQL query calculates the total number of successful and failed missions from the **SPACEXTBL** table. It c
missions by filtering the **MISSION_OUTCOME** column for 'Success' and failed missions by excluding 'Succes

```
%sql select count(MISSION_OUTCOME) as Success from SPACEXTBL where MISSION_OUTCOME = 'Success';
```

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

| Success |
|---|
| 98 |

```
%sql select count(MISSION_OUTCOME) as Failure from SPACEXTBL where MISSION_OUTCOME != 'Success';
```

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

# Boosters Carried Maximum Payload

This SQL query lists the names of the booster versions that carried the **maximum** payload mass subquery to first find the maximum payload mass (MAX(PAYLOAD_MASS__KG_)) and then selec versions that match this maximum payload mass.

```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

```
 * 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 |

# 2015 Launch Records

This SQL query retrieves records showing the month, booster version, launch site, and landing outcome for 2
failures. The **substr(DATE, 6, 2)** function extracts the month, and **substr(DATE, 0, 5) = '2015'** filters for the ye
records to show only failures on the drone ship.

```
%sql select substr(DATE, 6, 2) as Month, BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTC
from SPACEXTBL
where substr(DATE, 0, 5) = '2015'
and LANDING_OUTCOME = 'Failure (drone ship)';
```

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

| Month | Booster_Version | Launch_Site | Landing_Outcome |
|---|---|---|---|
| 01 | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |

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

This SQL query ranks landing outcomes (e.g., 'Failure (drone ship)', 'Success (ground pad)') between 2010-06-(
It groups records by **LANDING_OUTCOME**, counts occurrences, and sorts the results in descending order by c

```
%%sql select LANDING_OUTCOME, count(LANDING_OUTCOME) as Event_Count from SPACEXTBL
where DATE between '2010-06-04' and '2017-03-20'
group by LANDING_OUTCOME
order by Event_Count desc;
```

```
* sqlite:///my_data1.db
)one.
```

| Landing_Outcome | Event_Count |
|---|---|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |

# Section 3
# Launch site proximity Analysis

# Launch Sites Locations

# Launch Outcomes: Success vs Failure



**KSCLC-39A** shows the highest success rate, while CCAFS SLC-40 has a relatively lower success rate, as indica

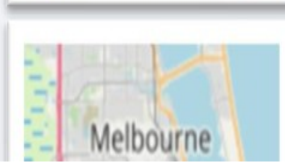
# Launch Site Proximity Distances

**Proximity Analysis of Launch Sites**

- Distance to Closest Coastline: 0.856 km
- Distance to Space Launch Complex 37: 3.15 km
- Distance to NASA Railroad: 1.23 km

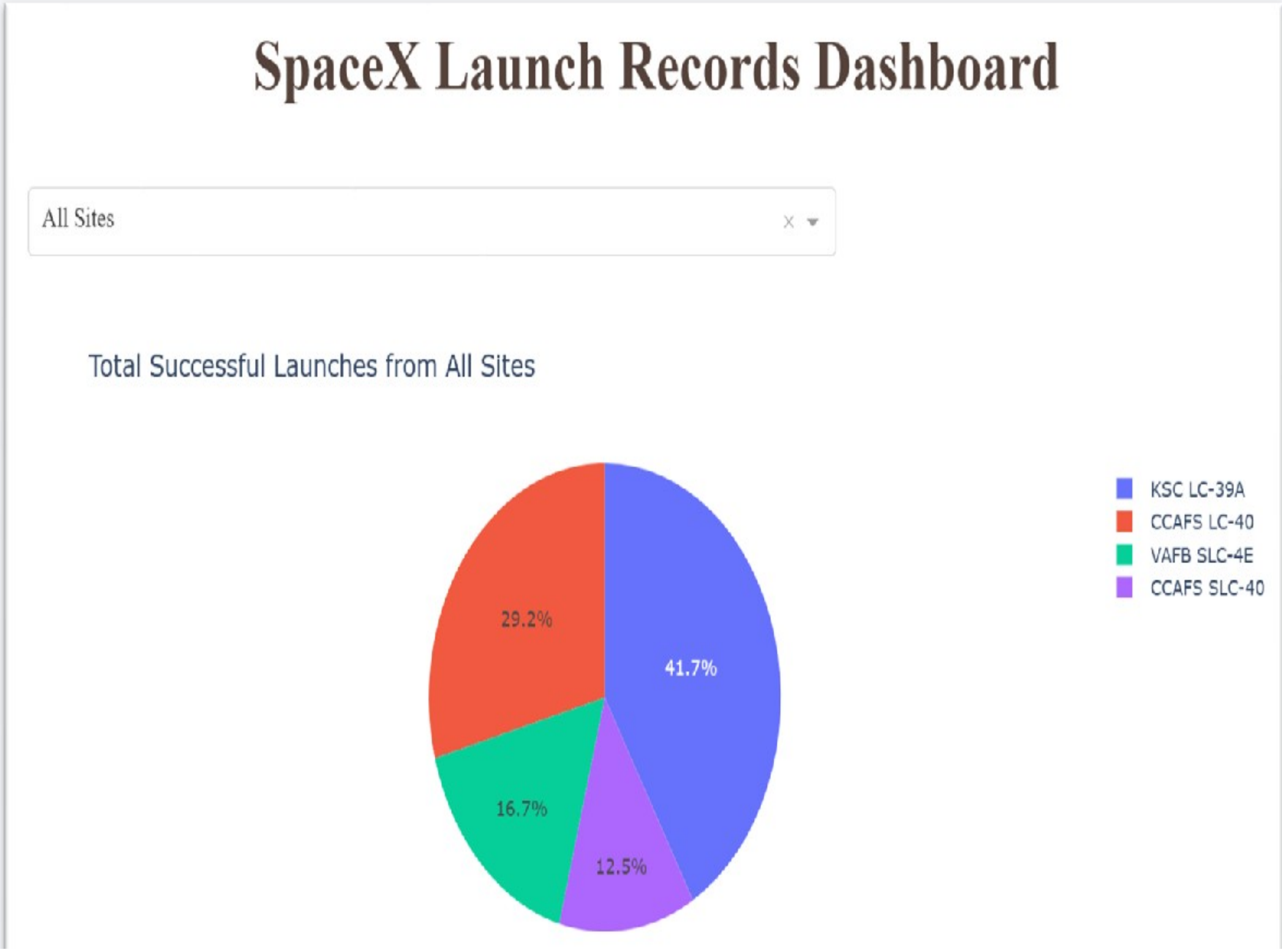
**Key Insights**

- Launch Sites: Located near coastlines to ensure optimal launch trajectories.
- Distance from Cities: Strategically distanced from cities and control centers for safety.
- Railways/Highways: Likely private, designed for

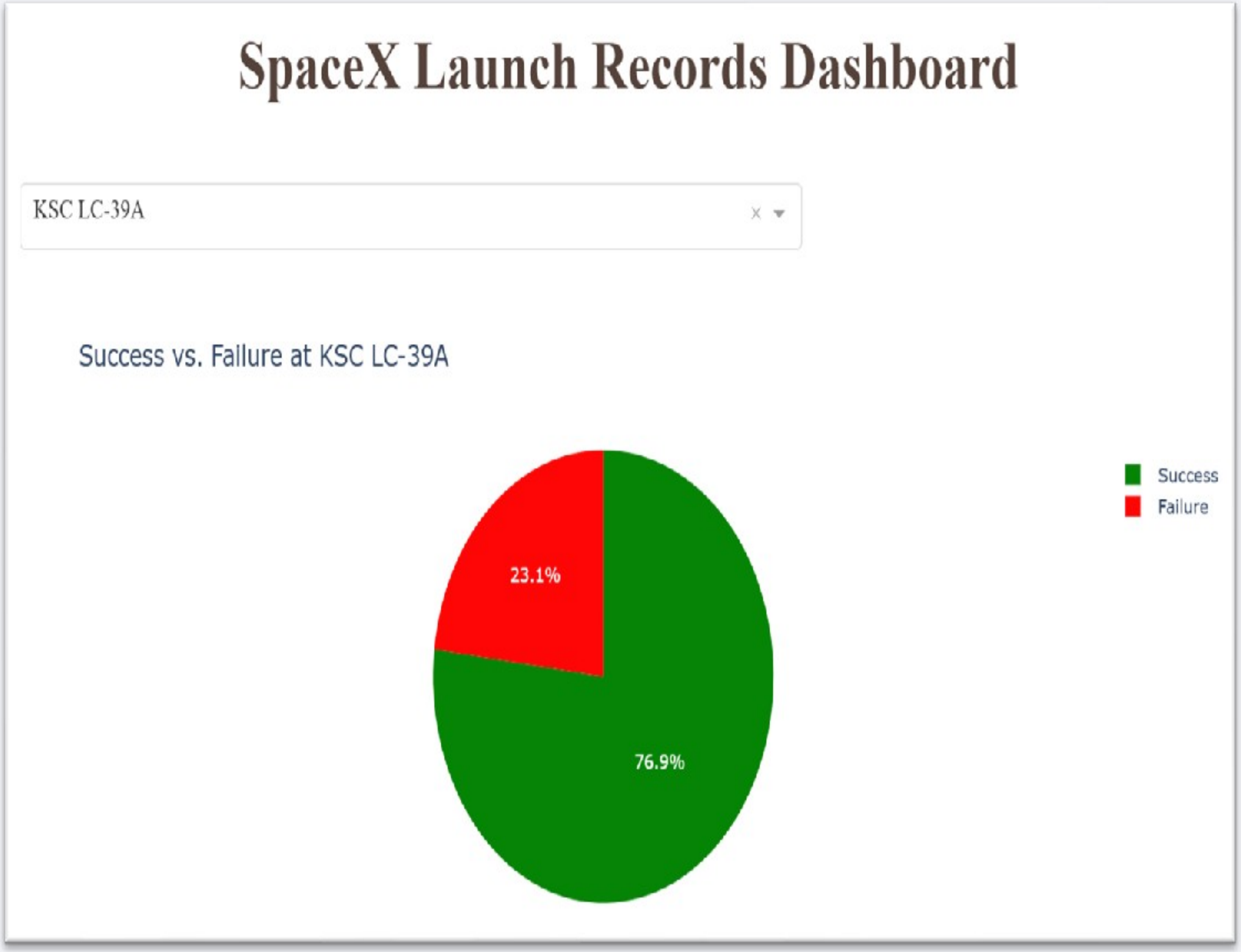



# Section 4
# Build a Dashboard With Plotly Dash

# Pie Chart for Launch Success Count for All Si



# Pie Chart for Launch Site with Highest Launch Succes



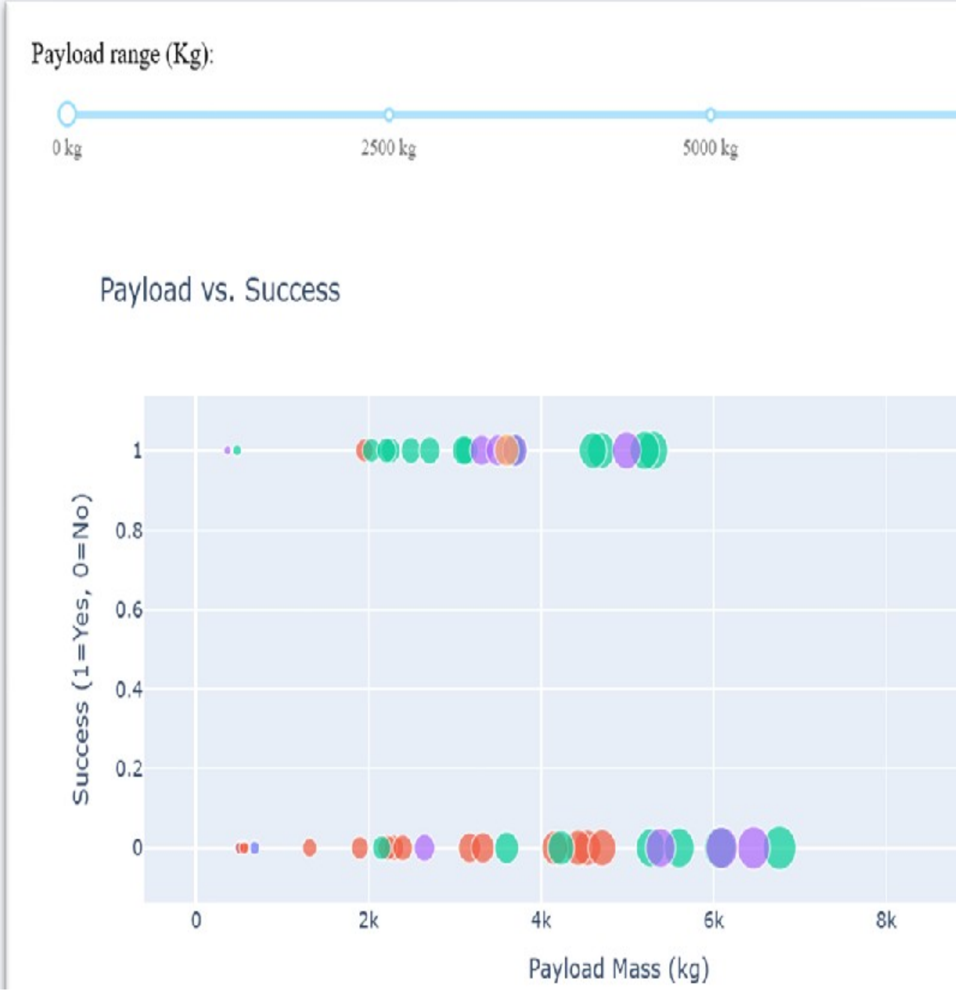
# Payload vs. Launch Outcome Scatter Plot for All

- **Highest Success Rates:**

  FT and B4 are the most successful in launches.

- **Payload Success:**
  - **2000-6000 kg:** Highest success rates.
  - **Less than 2000 kg:** Higher failure rates with older boosters.
  - **More than 6000 kg:** Newer boosters achieve higher success.

- **Key Insight:**
  - FT and B4 are the most reliable for launching

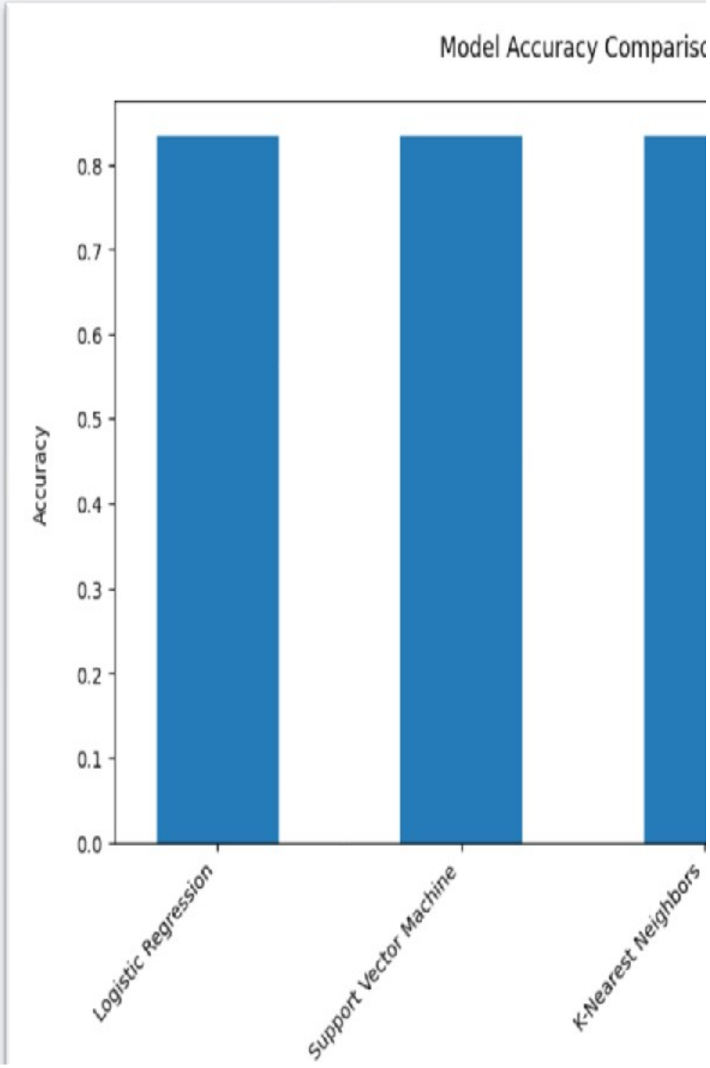


Skills Network

# Section 5
# Predictive Analysis (Classification)

# Classification Accuracy

- The bar chart shows the accuracy of four models: Logistic Regression, SVM, KNN, and Decision Tree Classifier.
- All models achieved the same accuracy of **0.83**, which was rounded to 0.8 in the chart, indicating no significant difference in performance.

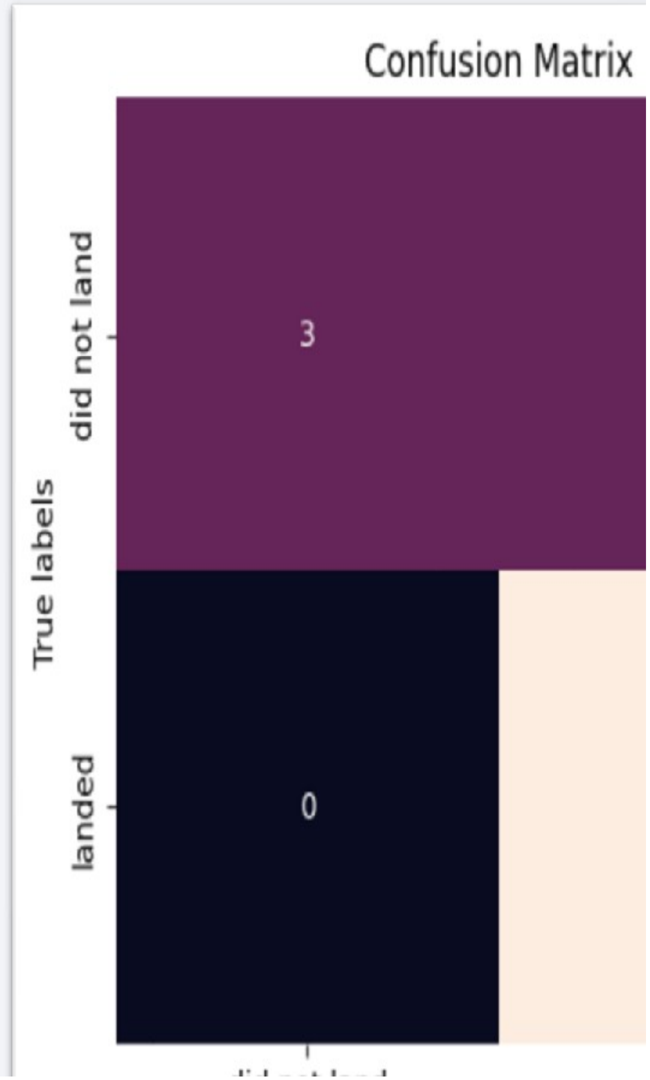


# Confusion Matrix

After examining the confusion matrix, it is clear that the models effectively distinguish between the classes. However, the main issue identified is **false positives**.

- **True Positives (12):** Correct predictions of "landed."
- **False Positives (3):** Incorrect predictions of "landed" when the true label is "not landed."

Overall, the performance is good, but the focus should be on reducing **false positives** for further model improvement



# Conclusions

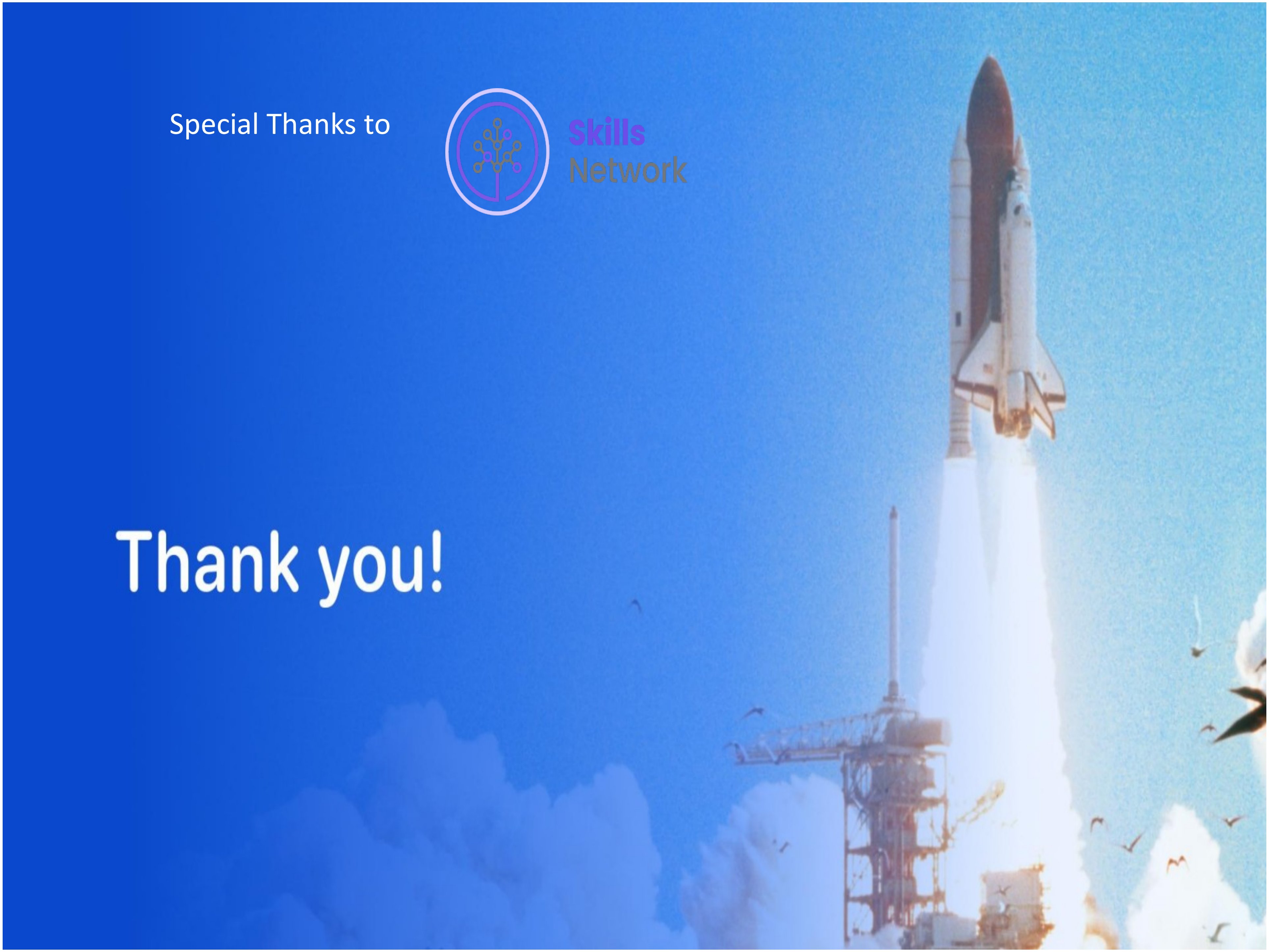
- **Models Used & Accuracy:**
  - Logistic Regression, SVM, Random Forest, and KNN were used to predict the successful la
    the first stage of the Falcon 9 rocket.
  - All models achieved an accuracy of 0.83, showing equal performance.
- **Factors Influencing Success:**
  - Location: KSC LC-39A showed the highest success rate (76.9%).
  - Payload Weight: Payloads between 2000 and 6000 kg showed higher success rates,

# Conclusions (Continued)

- **Importance of Results:**
  - These findings are crucial for developing pricing strategies for competing space companie
    reducing costs.
  - Future Improvements:
  - Models can be enhanced through further analysis or feature engineering.
- **Final Outcome:**
  - Successfully predicting the landing can improve operational efficiency, reduce risks, lowe
    enhance competitiveness in the space market.

# Appendix

- **Python Code Snippets**: Available in the GitHub repository.
- **SQL Queries**: Available in the GitHub repository.
- **Charts and Plots**: Available in the GitHub repository.


Special Thanks to
Skills
Network
Thank you!