IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Raaid Yousuf>
<19-6-2025>

Outline

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

Executive Summary

- 🛠 **Methodologies Used:**
- **Data Collection:** APIs & cloud CSVs for launch, location, and outcome data
- **Exploratory Data Analysis (EDA):** Trends, correlations, and visual summaries
- **Feature Engineering:** One-hot encoding, scaling, handling missing values
- **Modeling:** Logistic Regression, SVM, Decision Tree, KNN with GridSearchCV
- **Dashboard:** Interactive data exploration with Plotly Dash

- 📈 **Summary of Results:**
- Built & tuned 4 models with 10-fold cross-validation
- **Best accuracy:** ~88.88% from **Decision Tree Classifier**
- Developed **interactive dashboard** for real-time analytics

Introduction

- 🎯 **Project Title:**
- Analyzing the Impact of Recession on Falcon 9 Launch Outcomes

- 🌐 **Project Background:**
- SpaceX's Falcon 9 is a reusable rocket aimed at reducing launch costs.
- Launch success and reusability are critical to mission efficiency and sustainability.

- ❓ **Problems We Wanted to Solve:**
- What factors affect the success of a Falcon 9 rocket landing?
- Can we **predict landing success** using historical launch data?

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

Two primary methods used

- 🚀 **REST API** (SpaceX Launch Data)
- 🌐 **Web Scraping** (Wikipedia: Falcon 9 Launch Table)

- **Goal:**

  Collect structured launch data for Falcon 9 missions

- **Tools Used:**

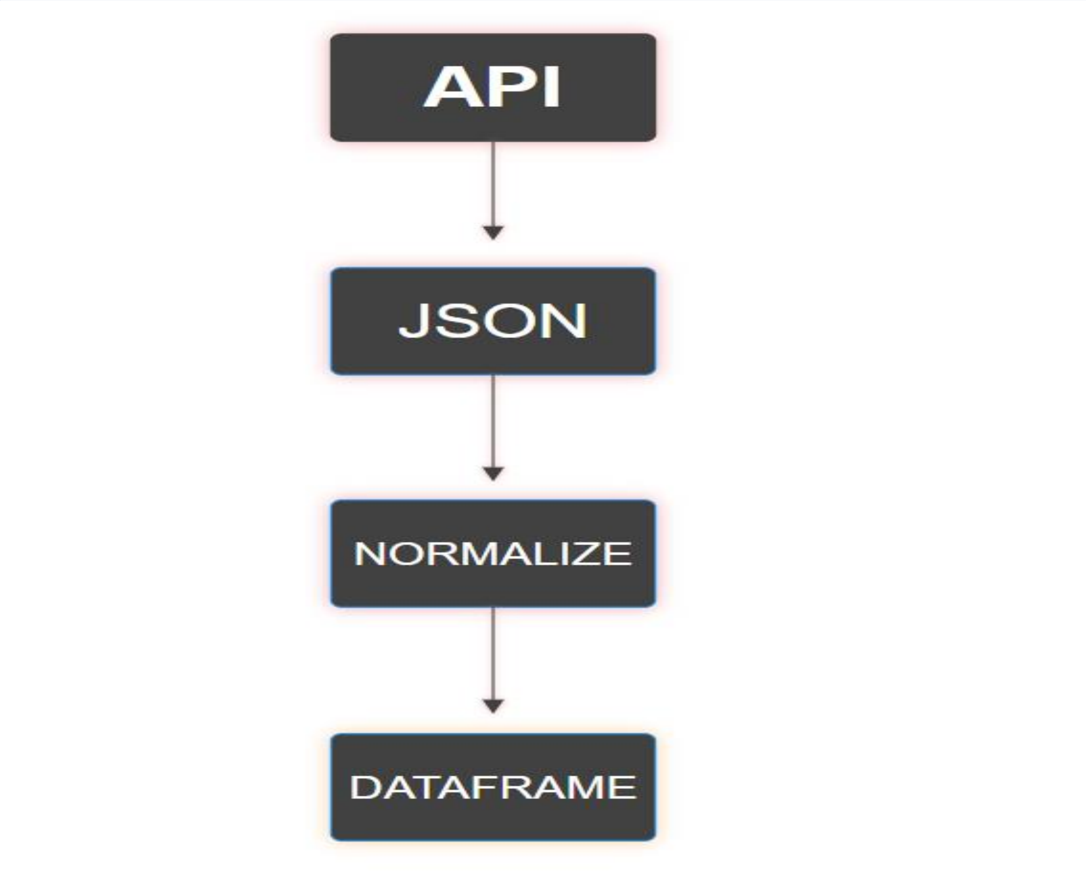
  requests, pandas, BeautifulSoup, json

Data Collection – SpaceX API

•Used https://api.spacexdata.com/v4/launches for launch data

•Retrieved:

1. Launch date & time
2. Rocket ID (mapped to Falcon 9)
3. Payload mass
4. Landing pad info
5. Launch success/failure



Data Collection-SpaceX API

Total launches fetched: **X**
(e.g., 195 Falcon 9 launches in lab dataset)

Extracted & cleaned key fields (e.g., rocket_name, landingPad)

Handled nested fields like cores[0]['landing_pad'] using .apply()

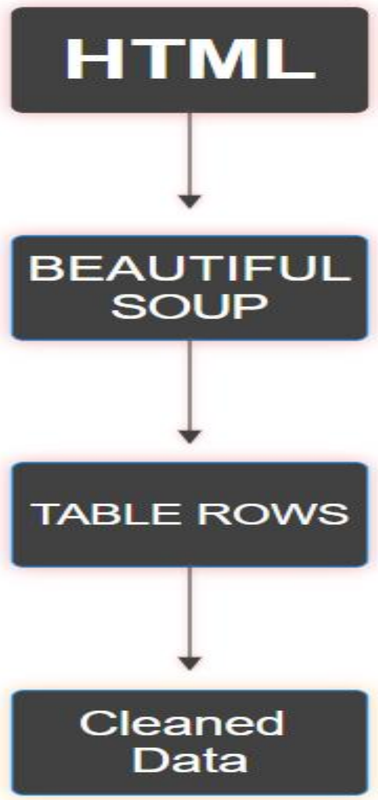
Data Collection - Scraping

•**Target:** Wikipedia page
List of Falcon 9 and Falcon Heavy launches

•Used requests.get() to fetch static URL snapshot

•Parsed HTML using Beautiful Soup

•Focused on 3rd table: wiki table plain row headers collapsible



Web Scraping- Outcome

Content:

Extracted columns:

- Flight No., Date, Time, Booster Version, Launch Site
- Payload, Orbit, Outcome, Booster Landing

Cleaned:

- Text with .strip(), Unicode normalization
- Skipped rows with missing or malformed entries

Data Wrangling- Preparing SpaceX launch Data

•Loaded SpaceX Falcon 9 launch dataset from a public cloud CSV file using pandas.
•Explored key columns including LaunchSite, Orbit, and Outcome.
•Used .value_counts() and enumerate() to analyze:
•Number of launches per site
•Occurrences of each mission outcome
•Identified and grouped landing outcomes into:
•**Successful Landings** (e.g., True ASDS, True RTLS)
•**Unsuccessful Landings** (e.g., False Ocean, None None)

Data Wrangling- Generating Classification Labels

Created a new binary feature Class:

1 → Successful landing

0 → Failed/No landing

Mapped each outcome from Outcome column to Class using conditional logic.

Calculated **overall landing success rate** using df["Class"].mean().

Exported the processed dataset as dataset_part_2.csv for use in the classification model.

EDA with Data Visualization

Objective:
To explore SpaceX launch data and identify patterns or relationships that influence mission success.
Key Visualizations Used:
•**Flight Number vs Launch Site:**
Strip plot to observe how launch success varies over time across different sites.
•**Payload Mass vs Launch Site:**
To analyze whether heavier payloads affect mission success at specific locations.
•**Success Rate by Orbit Type:**
Bar chart showing which orbit types have higher launch success rates.
•**Flight Number vs Orbit & Payload vs Orbit:**
Revealed how orbit selection correlates with mission order and payload weight.
•**Yearly Success Trend:**
Line chart tracking improvements or dips in success rate over time.

Why These Charts ?

Categorical relationships (e.g., Orbit, Launch Site) were visualized using **strip and bar plots** to effectively show group-wise differences.

Numerical trends over time were explored with a **line plot** to understand progress in launch success.

The visualizations supported **feature selection** and **hypothesis building** for machine learning modelling in the next phase.

EDA with SQL

Key SQL EDA Activities

•**Data Cleaning**: Removed null dates to ensure accurate temporal analysis.

•**Launch Site Analysis**:

- Retrieved unique launch sites.
- Filtered missions starting with 'CCA' to study specific sites.

•**Payload Insights**:

- Aggregated payload mass for NASA (CRS) missions.
- Found booster versions that carried **maximum payload** using subqueries.

•**Mission Outcome Trends**:

- Counted total **successful and failed** missions.
- Analyzed success/failure across different launch sites and orbit types.

•**Temporal Exploration**:

- Identified the **first successful ground pad landing**.
- Ranked landing outcomes between specific date ranges.

Build an Interactive Map with Folium

Objective:
Visualize SpaceX launch sites and their proximities using interactive maps.

Key Map Features Added:

•**Markers:** Indicate individual launch outcomes with color-coded icons (Success, Failure).

•**Circles:** Represent launch site locations clearly on zoomed-in views.

• **Distance Lines (Polylines):** Show distances from launch sites to:

- Coastlines
- Railways
- Highways
- Cities

•**Mouse Position Tool:** Enabled live coordinate capture for precise proximity mapping.

•**Div Icon Labels:** Display distance annotations between launch sites and surrounding features.

Purpose of Map Objects

Why These Objects Were Used:

✅ **Markers & Circles:**
To clearly visualize the **exact location** of each SpaceX launch site and distinguish successful vs. failed missions.

✅ **Polylines & Labels:**
To **analyze proximity** and assess strategic placement of launch sites relative to transport, cities, and safety zones.

✅ **Mouse Position Tool:**
Allowed dynamic coordinate retrieval, making it easier to measure real-world distances interactively.

✅ **Insight Derived:**
Launch sites are purposefully located near **railways, highways, and coastlines**, but are **distant from cities** for safety.

Build a Dashboard with Plotly Dash

•**Objective**: Built an interactive dashboard to explore SpaceX launch data dynamically.

•**Key Components**:

•**Launch Site Dropdown**

Allows selection of *All Sites* or individual launch sites to filter results.

•**Success Pie Chart**

Shows overall launch success distribution or success vs. failure for a selected site.

•**Payload Range Slider**

Filters data by payload mass (0–10,000 kg) to analyze correlation with mission outcome.

•**Success vs. Payload Scatter Plot**

Visualizes the relationship between payload mass and launch outcome, color-coded by *Booster Version*.

Predictive Analysis (Classification)

- ## Objective:
- To build and evaluate classification models that predict whether a Falcon 9 rocket launch will result in a successful landing.

- ## Process Summary:
- **Data Preparation:** Standardized numerical features, applied one-hot encoding for categorical data
- **Target Variable:** Class (1 = Landed Successfully, 0 = Did Not Land)
- **Train-Test Split:** 80% training, 20% testing using train_test_split

- ## Models Built:
- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree Classifier
- K-Nearest Neighbors (KNN)

- ## Model Selection:
- Applied **GridSearchCV (cv=10)** for hyperparameter tuning
- Evaluated using **accuracy score** and **confusion matrix**
- Compared models on **validation** and **test data accuracy**

Model Development Workflow

```
Raw Dataset
     ↓
Feature Selection & Engineering
     ↓
Standardization & One-Hot Encoding
     ↓
Train-Test Split (80-20)
     ↓
Model Building:
   • Logistic Regression
   • SVM
   • Decision Tree
   • KNN
     ↓
Hyperparameter Tuning (GridSearchCV)
     ↓
Model Evaluation (Accuracy, Confusion Matrix)
     ↓
Best Model Selection
```

Results

📊 **EDA Findings:**

•Landing success is influenced by **launch site**, **booster version**, and **orbit type**

•Boosters reused multiple times tend to land successfully more often

📟 **Interactive Dashboard Highlights:**

•Pie chart of launch success by site

•Scatter plot of payload vs. outcome

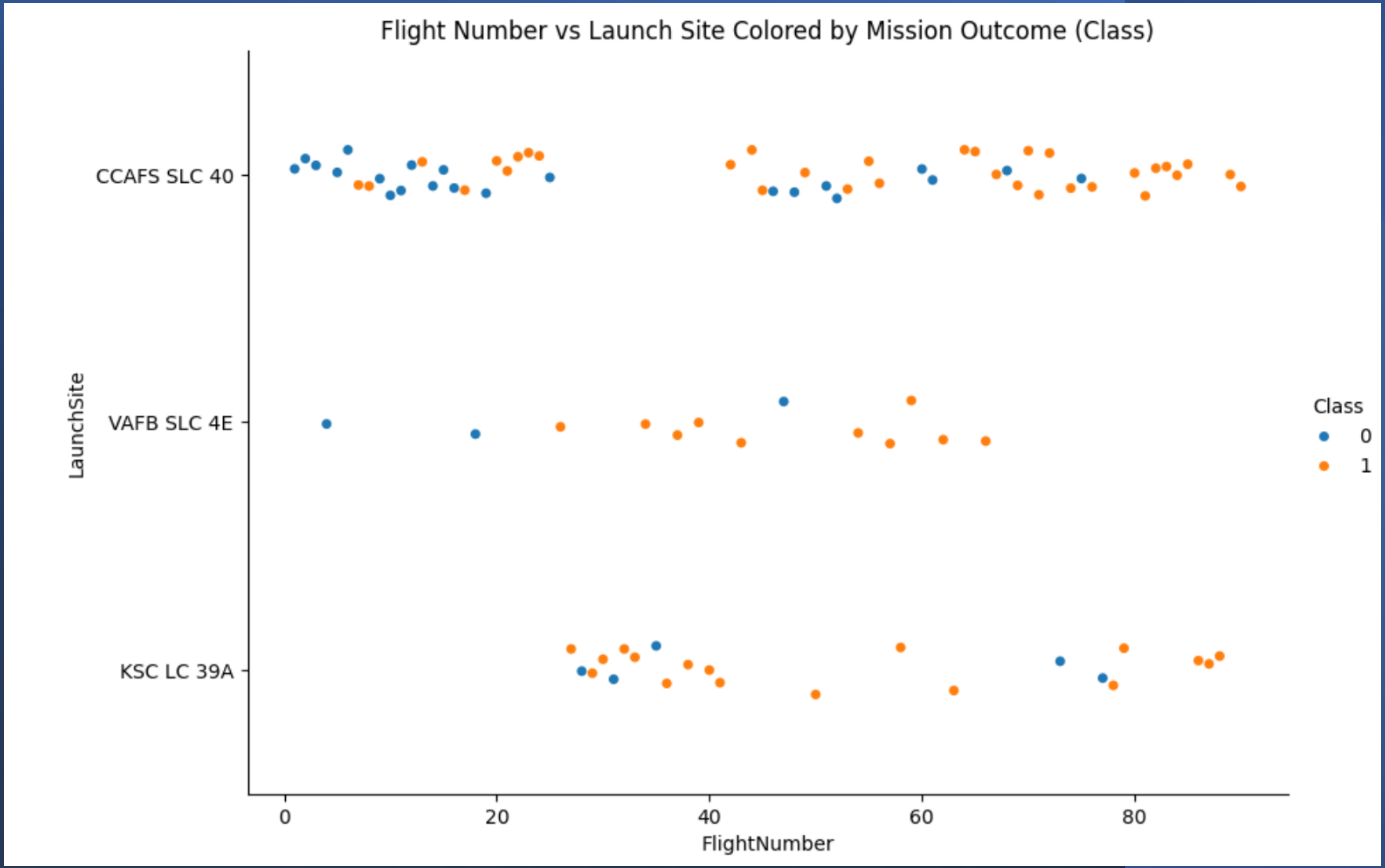
•Payload slider to filter and analyze patterns

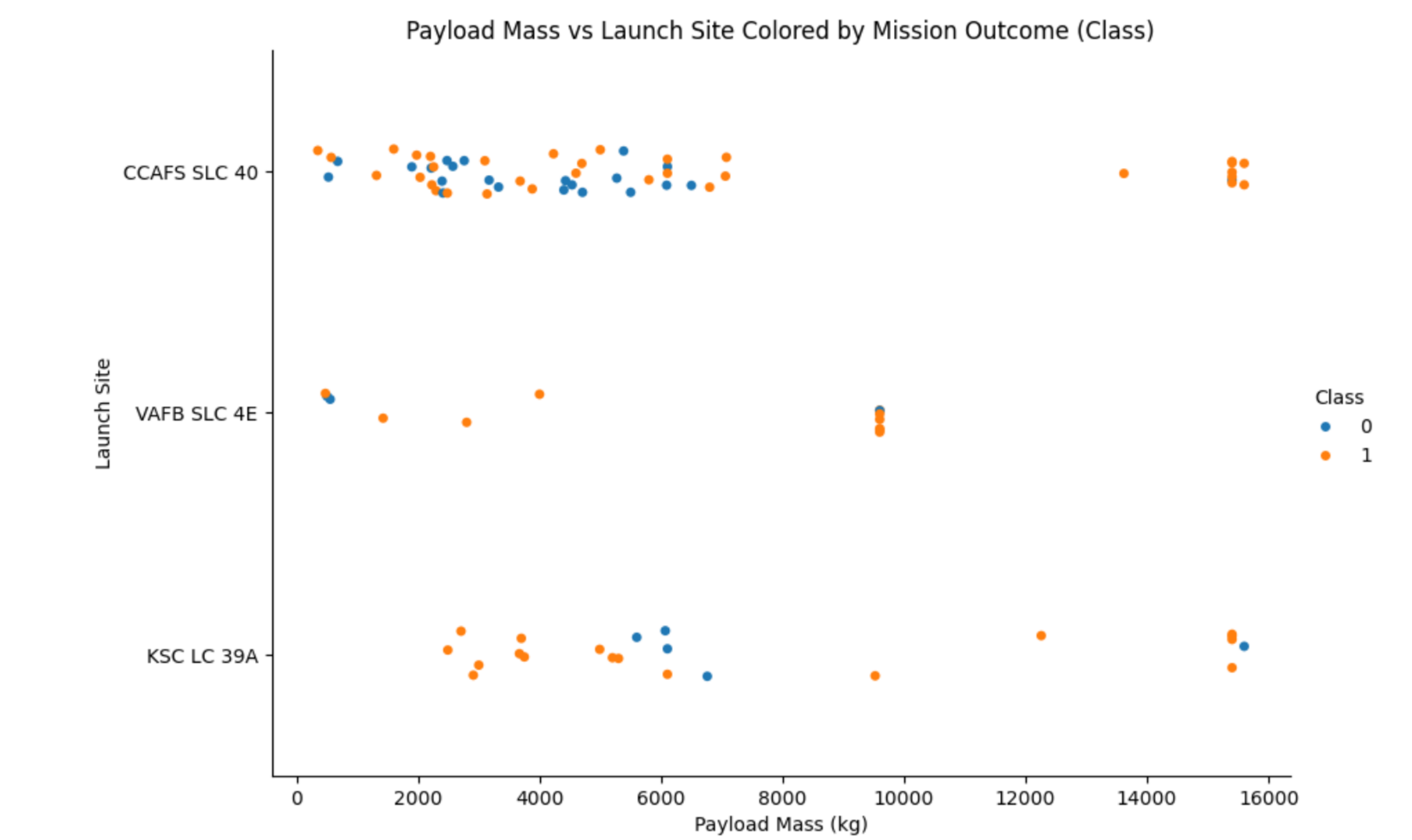
🤖 **Predictive Model Accuracy**:

Model	Test Accuracy
Logistic Regression	86.66%
SVM	88.88%
Decision Tree	**88.88%** ✅
KNN	84.44%

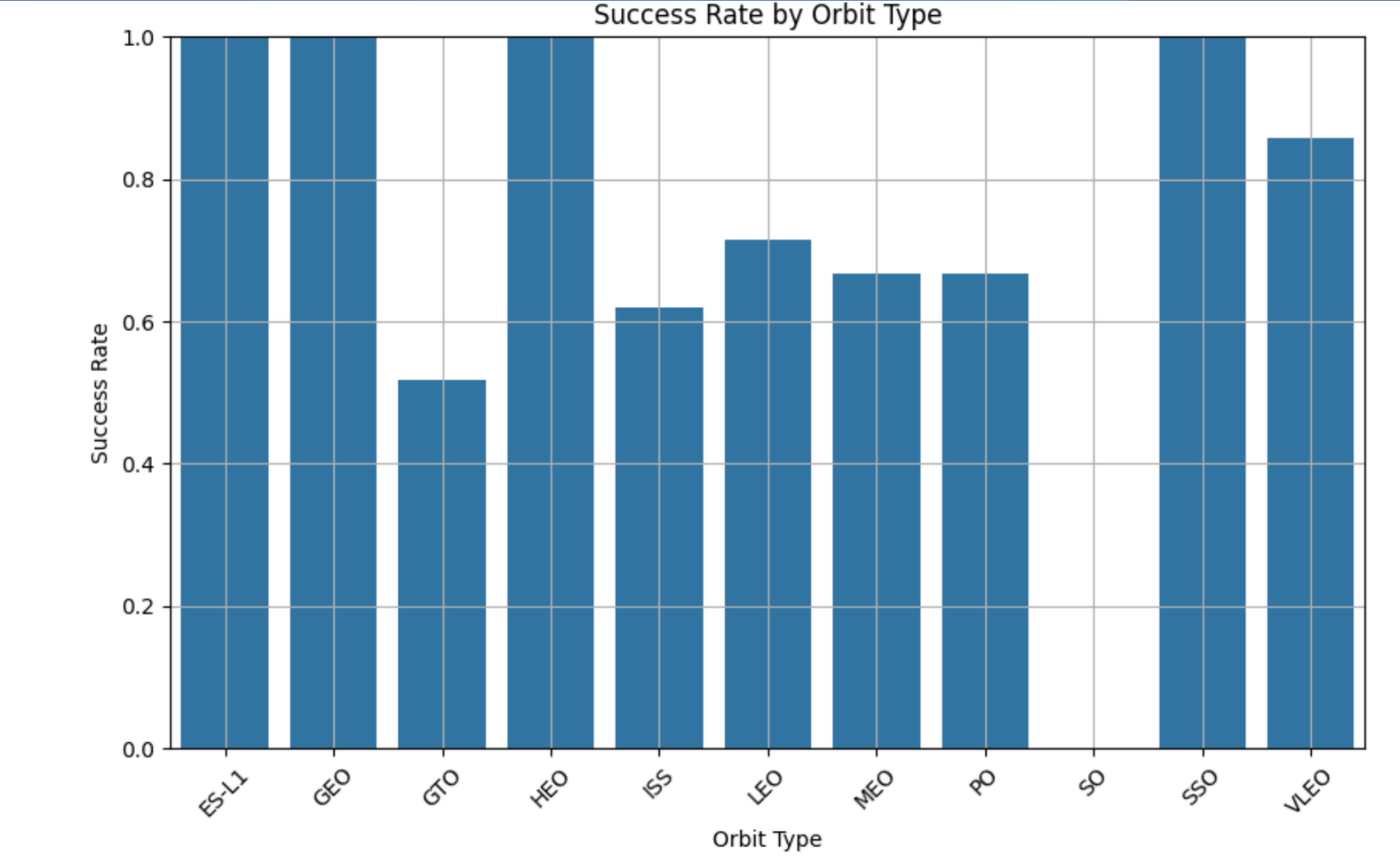
Section 2

Insights drawn from EDA

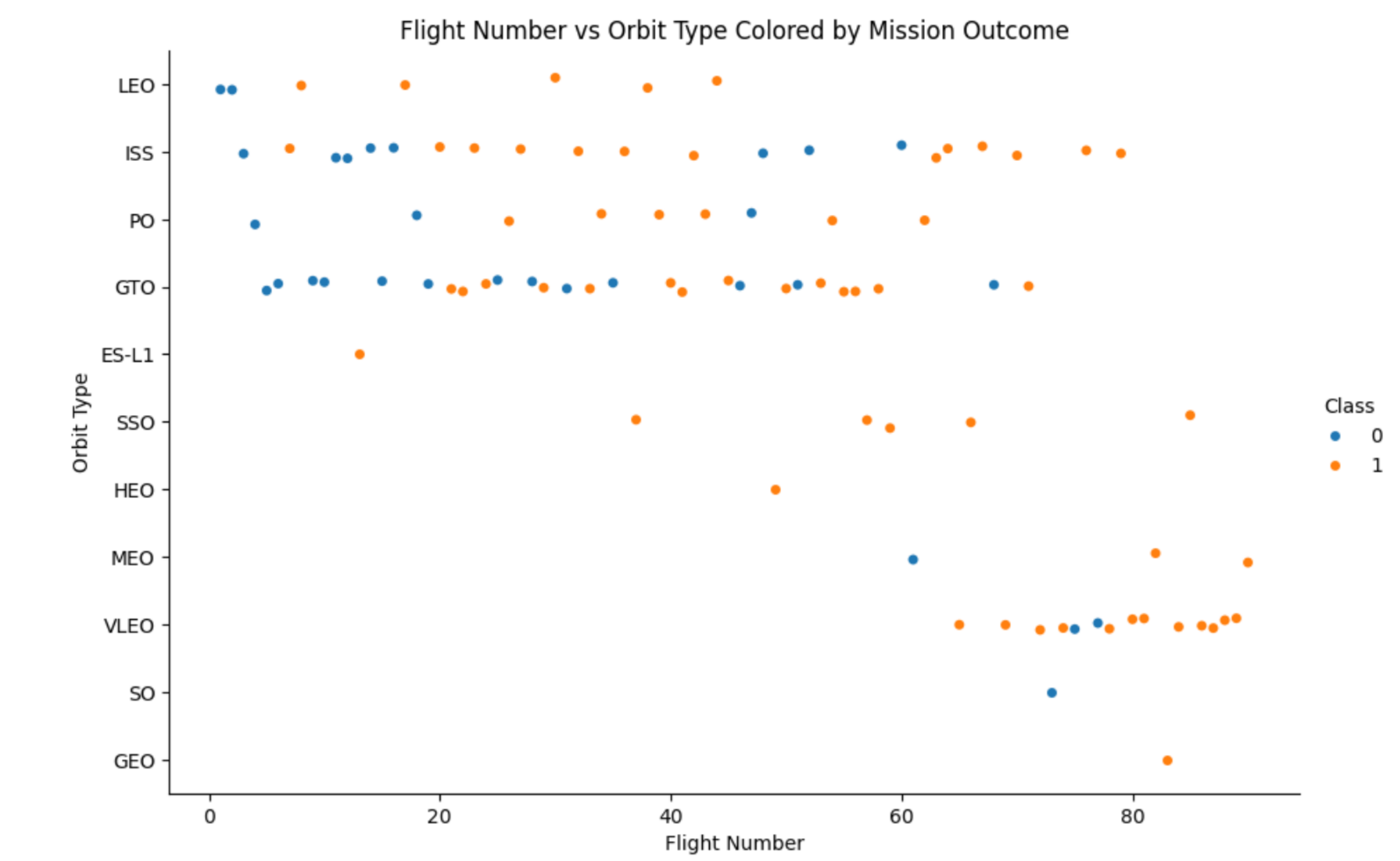

Flight Number vs Launch Site Colored by Mission Outcome (Class)
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
LaunchSite
0
20
40
60
80
FlightNumber
Class
0
1

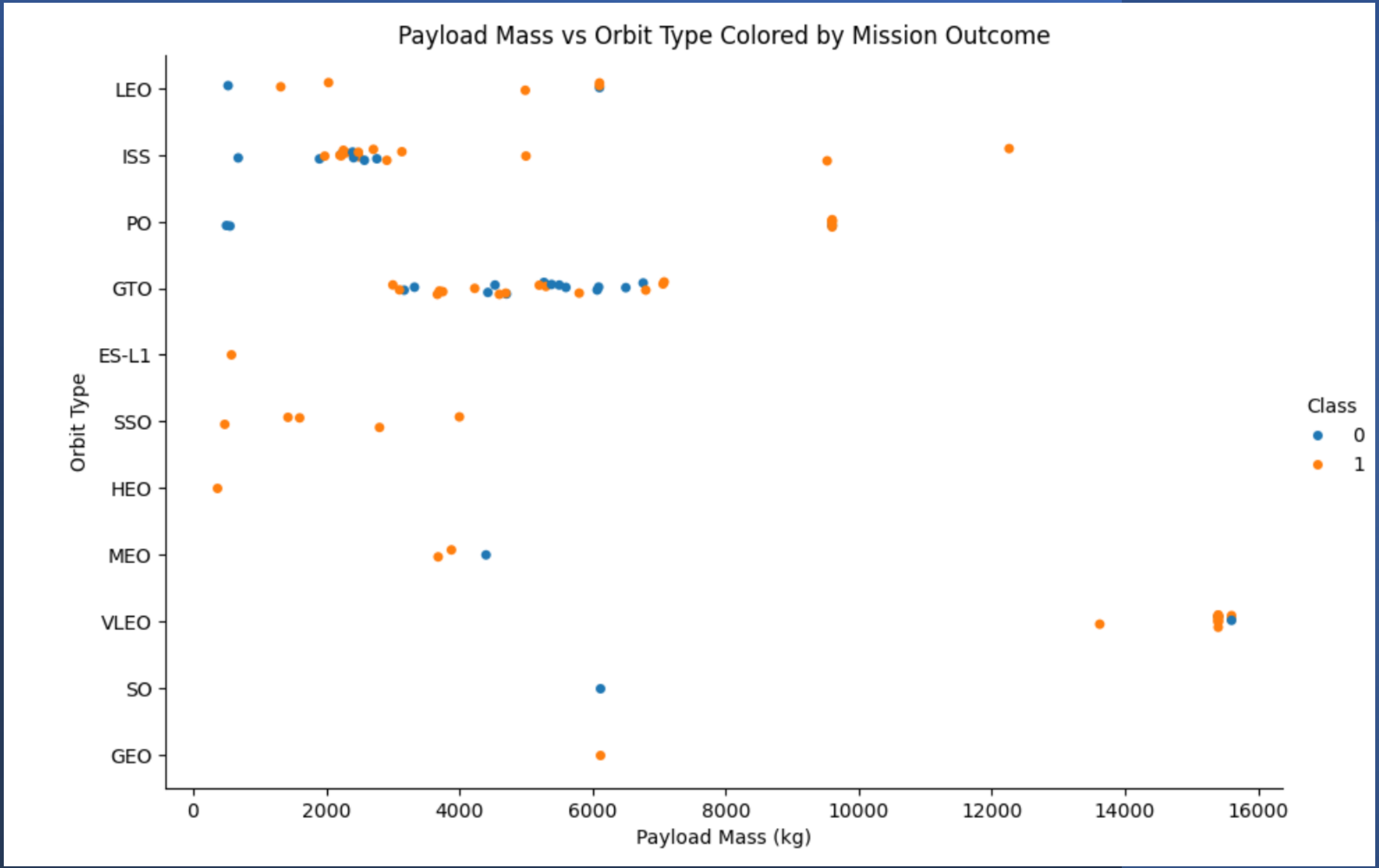


Payload Mass vs Launch Site Colored by Mission Outcome (Class)
Launch Site
CCAFS SLC 40
VAFB SLC 4E
KSC LC 39A
Payload Mass (kg)
0
2000
4000
6000
8000
10000
12000
14000
16000
Class
0
1

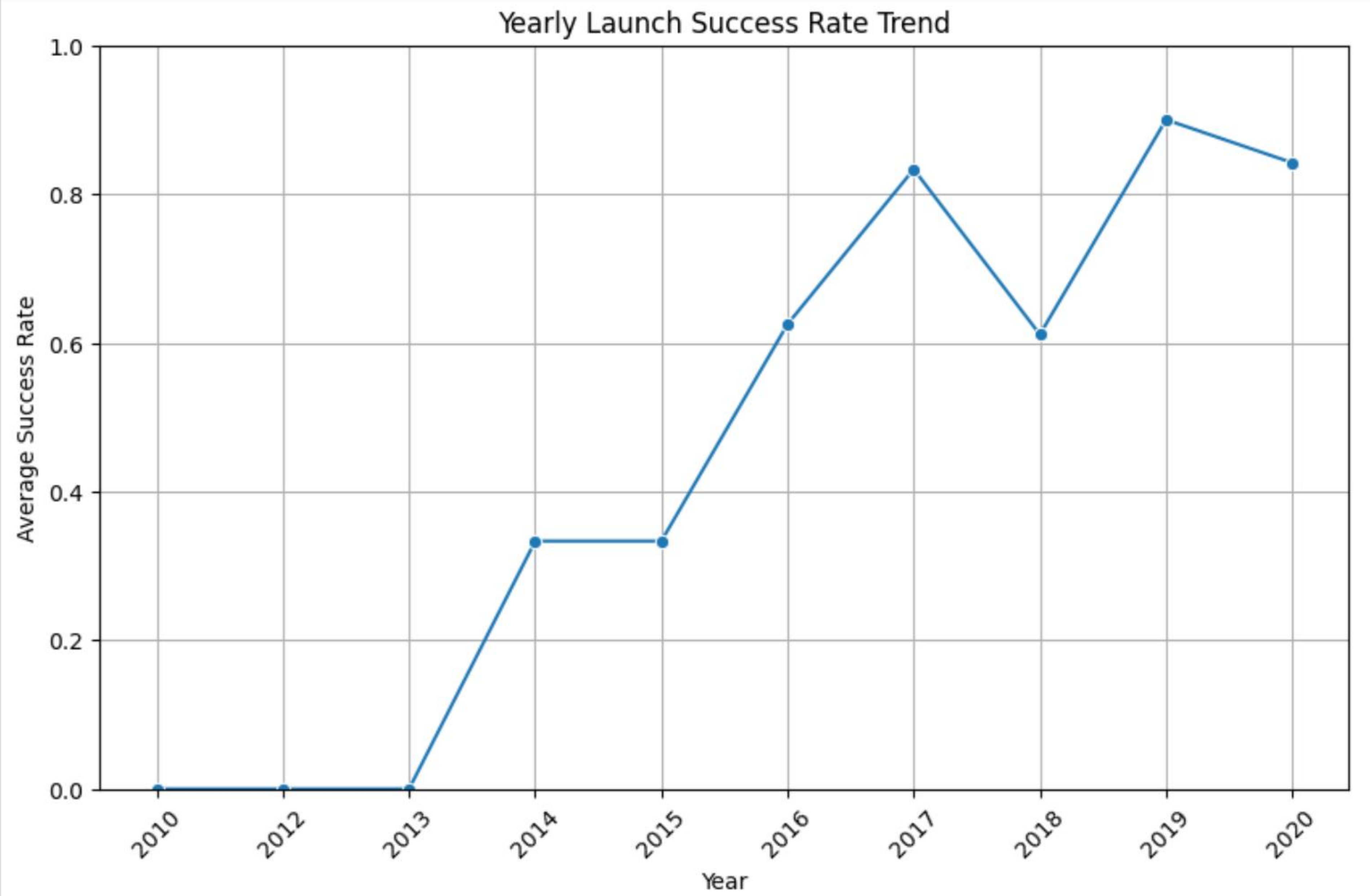


Success Rate by Orbit Type
Success Rate
1.0
0.8
0.6
0.4
0.2
0.0
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit Type


Flight Number vs Orbit Type Colored by Mission Outcome
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
Orbit Type
0
20
40
60
80
Flight Number
Class
0
1



Payload Mass vs Orbit Type Colored by Mission Outcome
Orbit Type
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
0
2000
4000
6000
8000
10000
12000
14000
16000
Payload Mass (kg)
Class
0
1

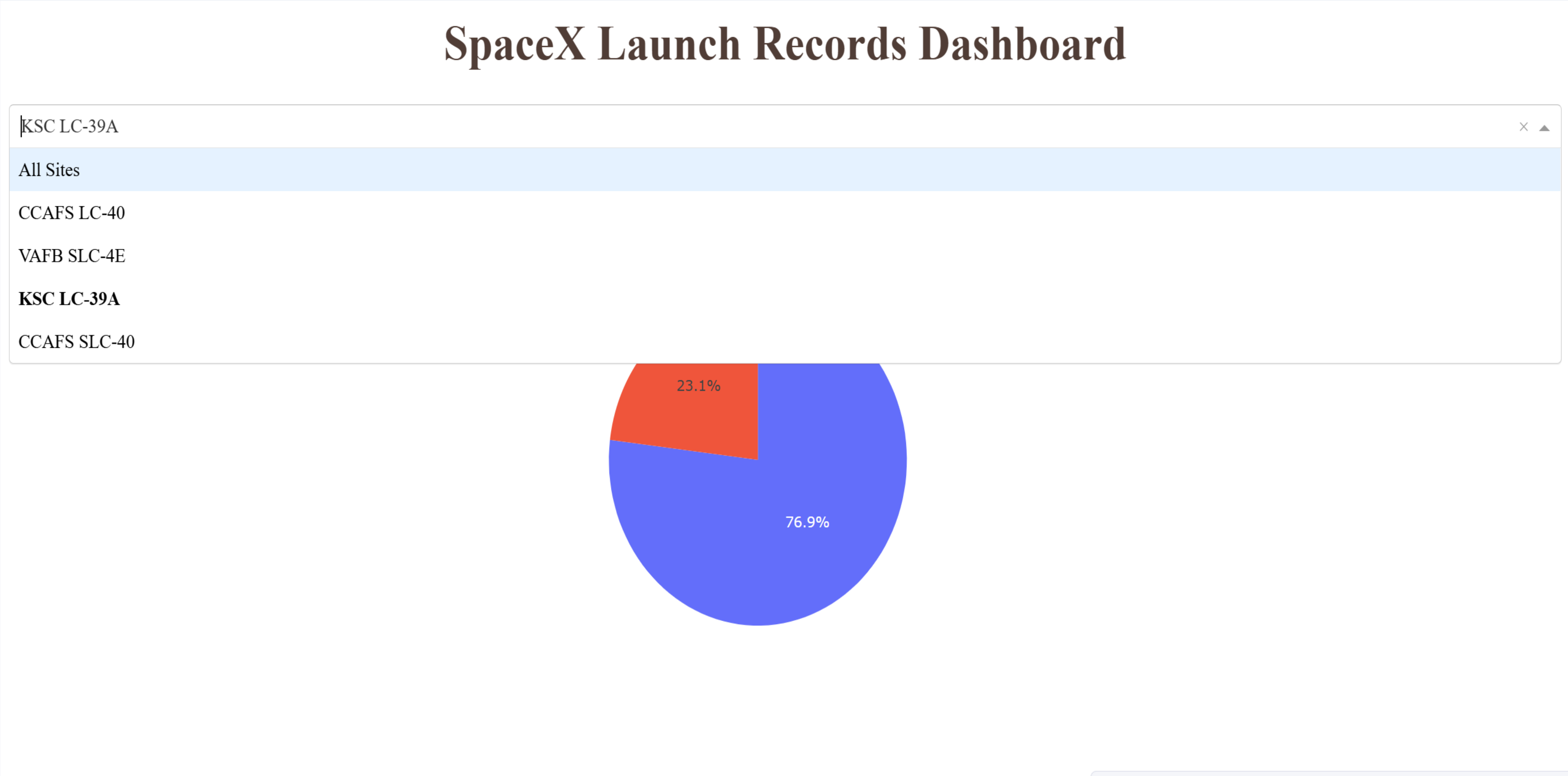


Yearly Launch Success Rate Trend
Average Success Rate
1.0
0.8
0.6
0.4
0.2
0.0
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year

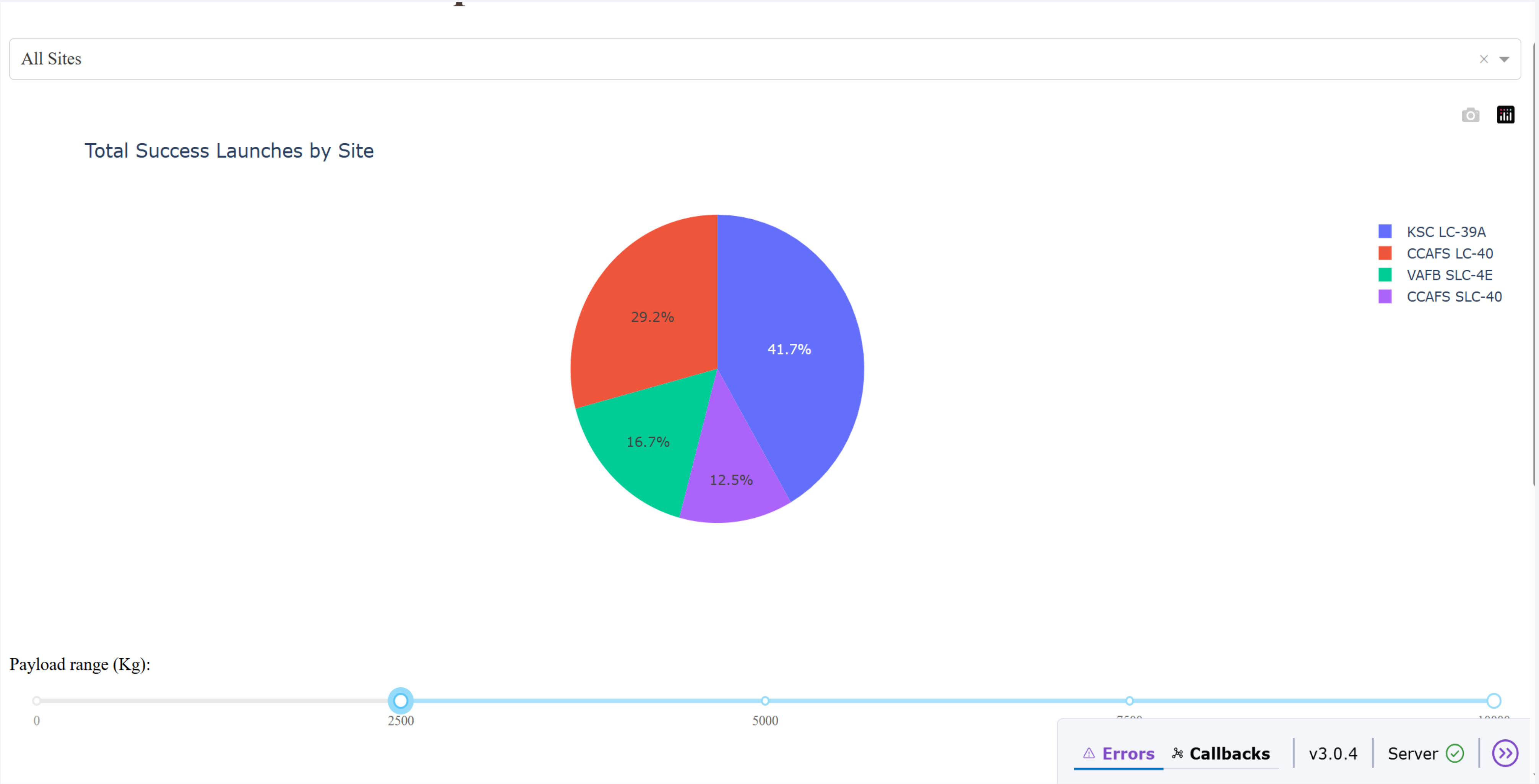
Section 4

Build a Dashboard with Plotly Dash

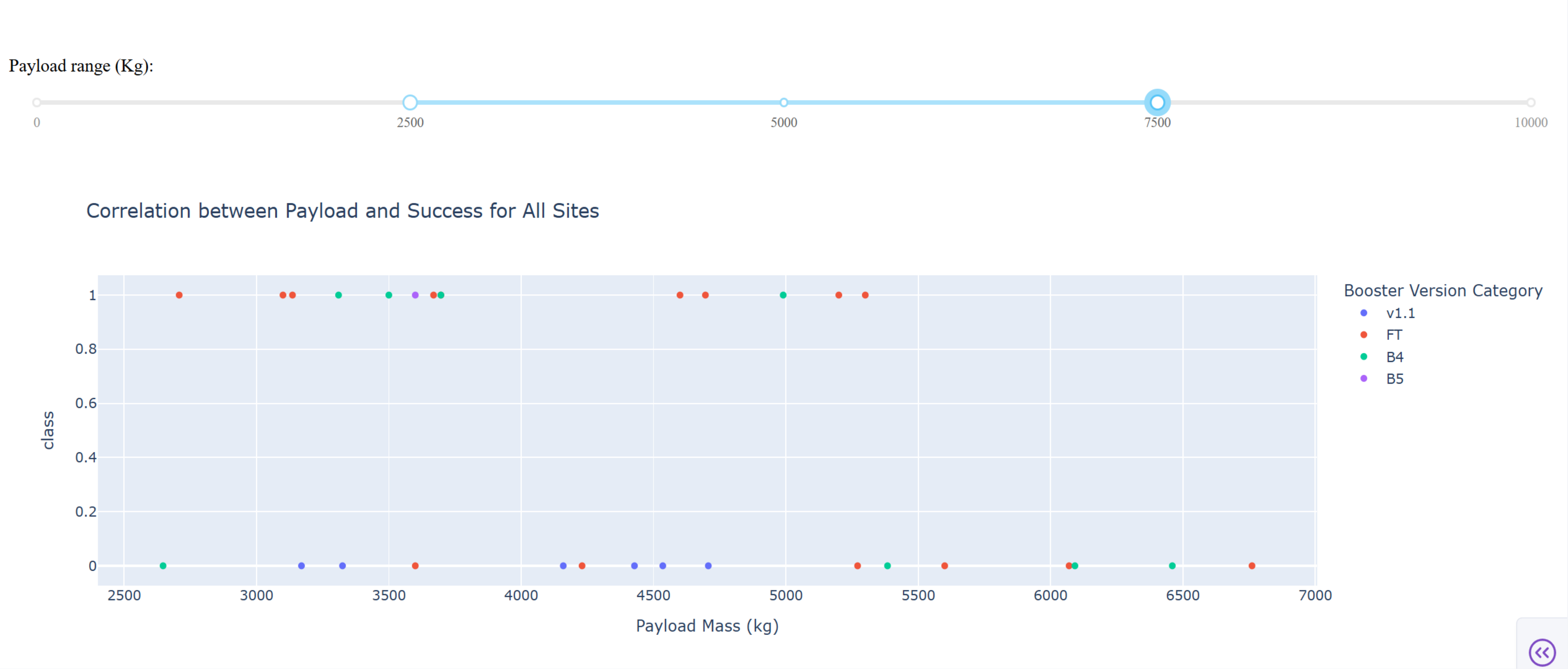
SpaceX Launch dashboard



Highest Success Ratio



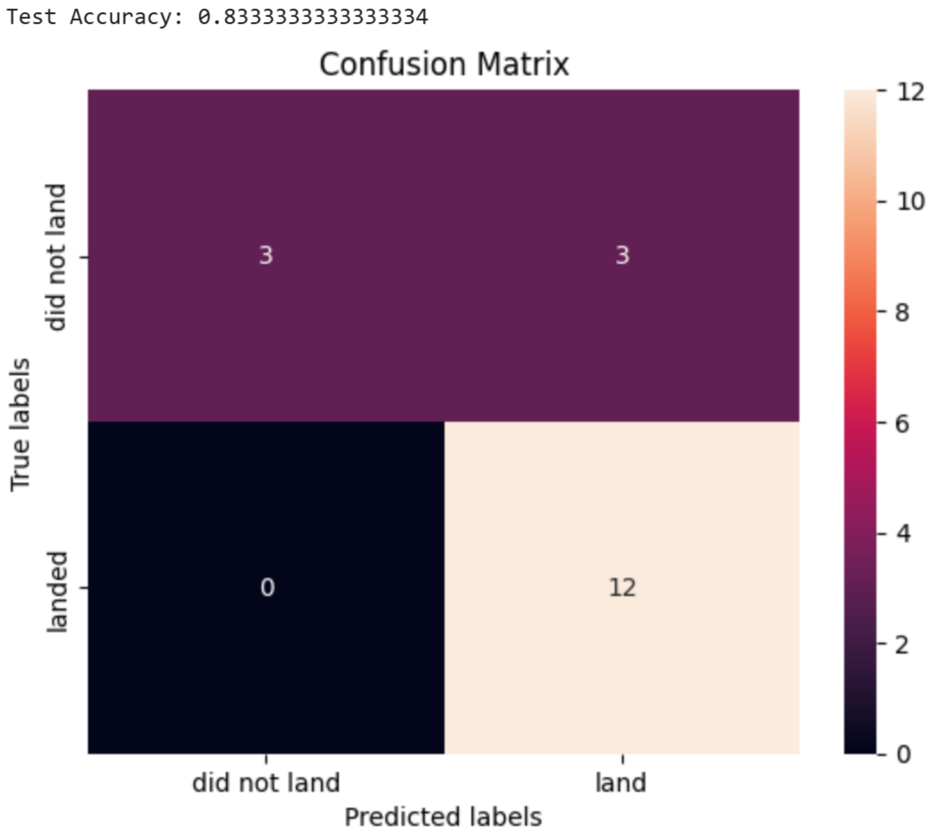
Scatter Plot Dashboard



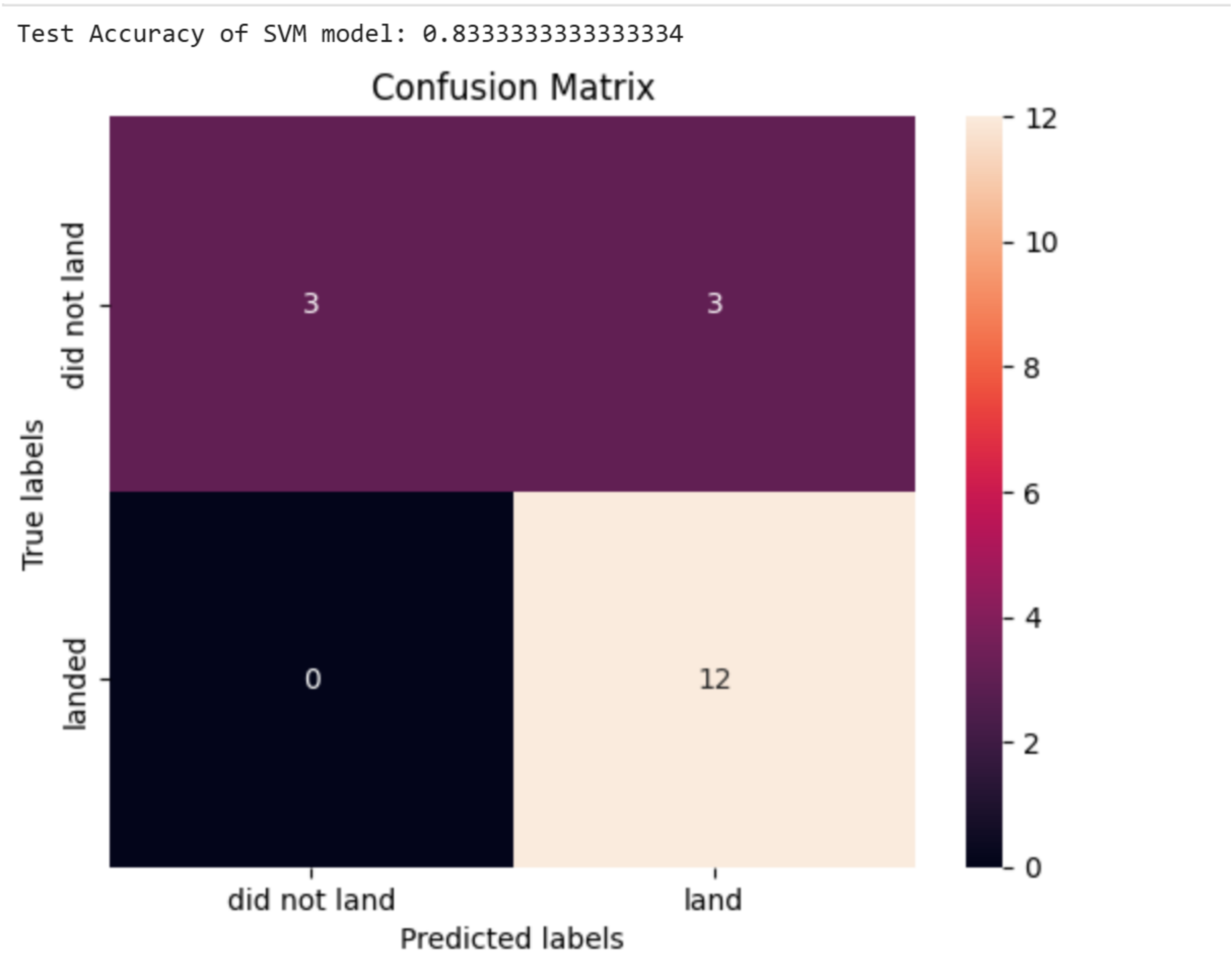
Section 5

Predictive Analysis (Classification)

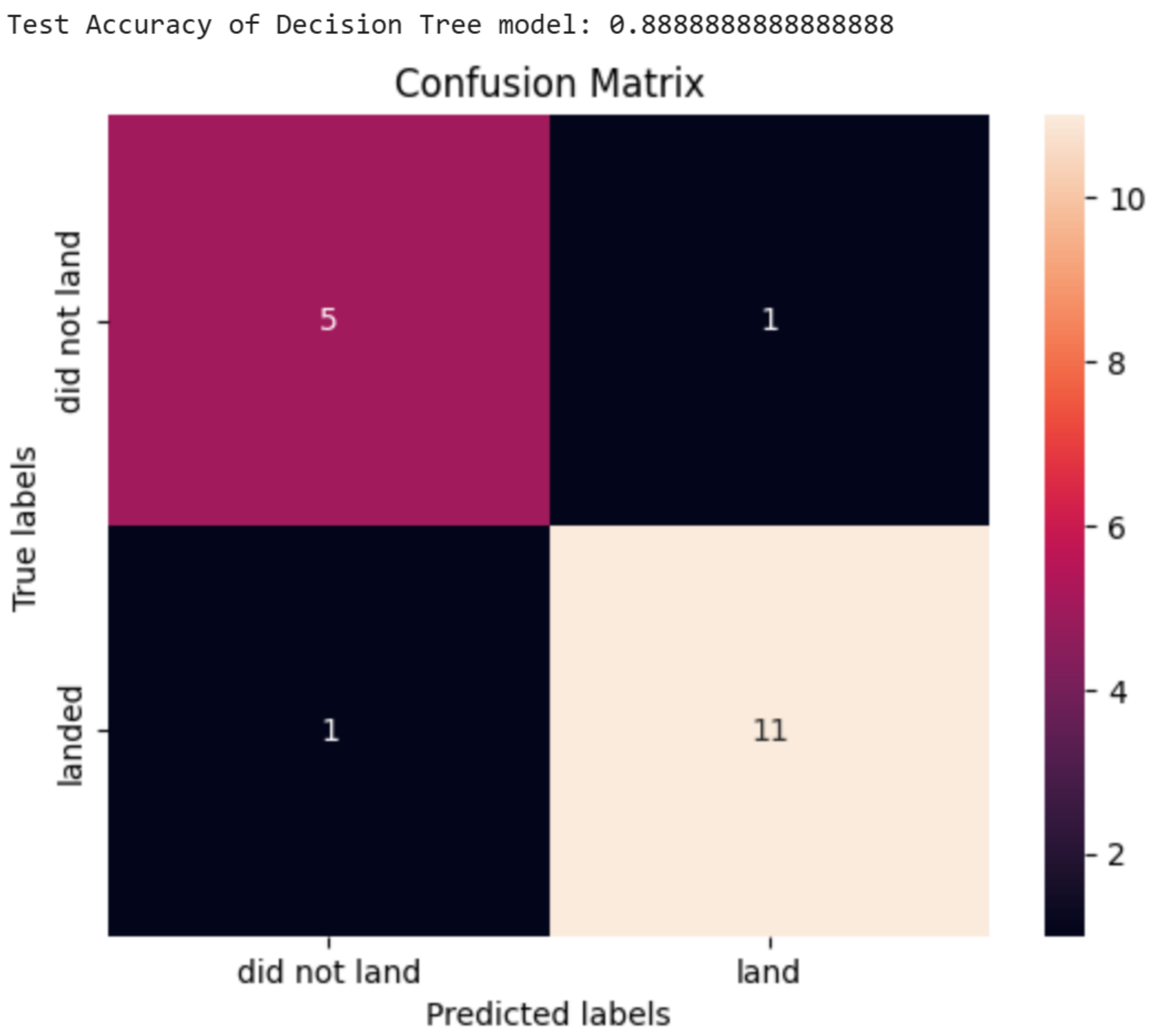
Logistic Regression Confusion Matrix



SVM Confusion Matrix

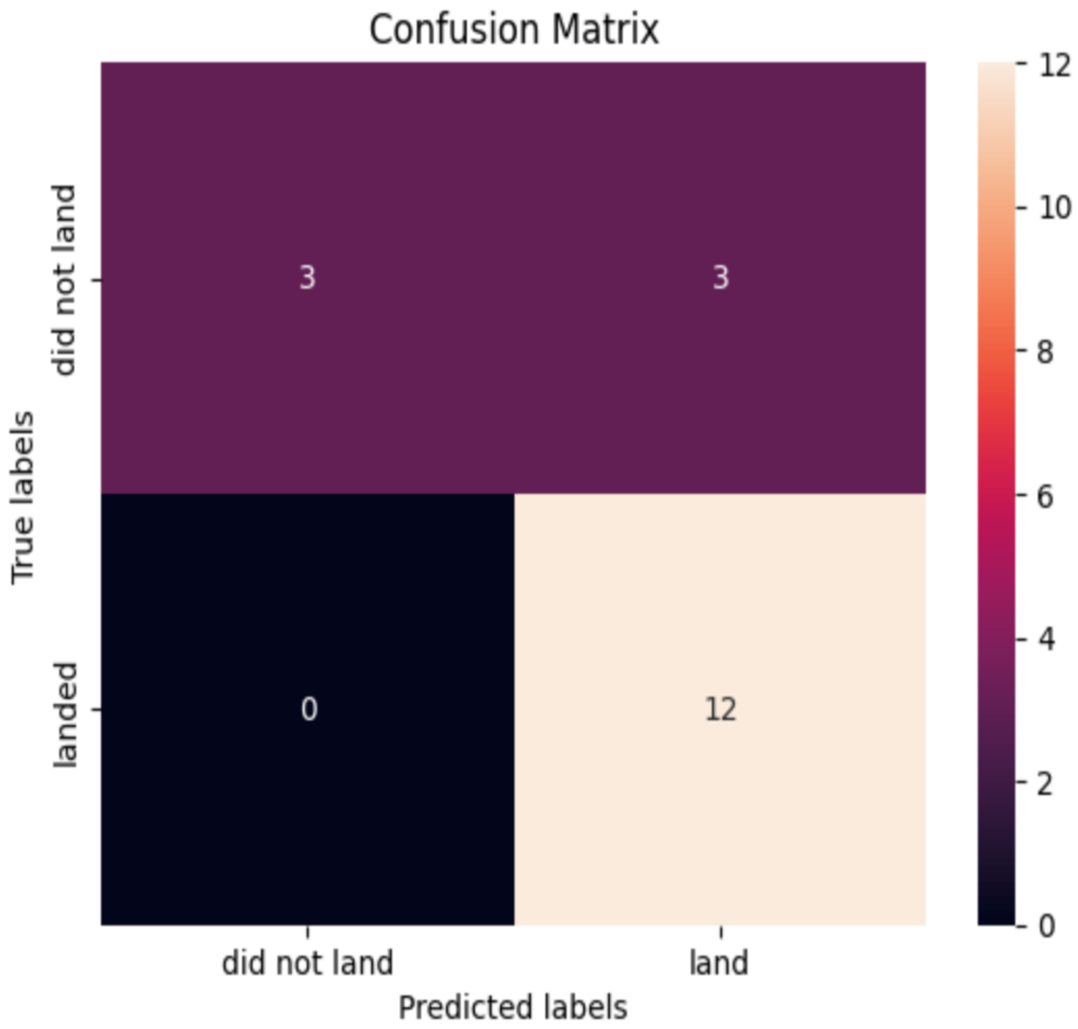


Decision Tree Confusion Matrix



KNN Model Confusion Matrix

Test Accuracy of KNN model: 0.8333333333333334



Conclusions

✅ Key Takeaways:

- Successfully predicted rocket landing outcomes using machine learning
- Decision Tree performed best with ~88.88% test accuracy
- Dashboard allows interactive insights into launch patterns
- Model can assist mission planning and risk management for future launches

Thank you!