IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmad Kurtubi
November 2024

OUTLINE

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

EXECUTIVE SUMMARY

METHODOLOGY:

DATA COLLECTION

- SPACEX REST API AND WEB SCRAPING,

DATA PROCESSING AND CLEANING

- PYTHON (PANDAS)

DATA EXTRACTION

- SQL QUERIES

DATA VISUALIZATION

- PYTHON (MATPLOTLIB, SEABORN, AND FOLIUM)

Key Results:

Launch Success Rate:

- Analyzed a total of 90 SpaceX (Falcon 9) launches, revealing an overall success rate of **67%%**, with **60 successful launches** and **30 failures** across all missions.

Launch Sites:

CCAFS SLC 40 was identified as the primary launch site, accounting for **61% of total launches**. It demonstrated a high success rate of **60%** compared to other sites.

Payload Analysis:

The average payload capacity for the Falcon 9 was calculated at **6,105 kg**, with the highest payload reaching **15,600 kg** for geostationary missions.

Orbit Types:

- LEO, ISS, PO, GTO, and ES-L1 have a high success rate, with very few unsuccessful launches.

Trend Analysis:

The yearly average success rate showed a significant upward trend, indicating improvements in launch reliability over the years, with the most successful year being **2019**.

Conclusion:

The analysis highlighted SpaceX's effectiveness in launching payloads with high reliability, showcasing advancements in their technology and operations. These findings underscore the importance of continuous monitoring and enhancement of launch strategies to maintain and improve success rates.

INTRODUCTION

Project Background and Context:

- SpaceX has revolutionized the aerospace industry with its advancements in reusable rocket technology, reducing the cost of space missions.
- By analyzing SpaceX’s historical launch data, we can gain insights into the performance, reliability, and success rates of different launch missions and boosters.
- This project leverages data science techniques to evaluate SpaceX’s performance metrics, mission outcomes, and launch site effectiveness, contributing to a deeper understanding of factors influencing mission success.

Problems to Find Answers:

- What factors affect the success rate of SpaceX launches?
- Which launch sites and booster versions have the highest success rates?
- How do payload mass and orbit types influence mission outcomes?
- What are the trends in launch success rates over time?
- How does SpaceX's performance compare across different orbit types?

Section 1

Methodology

METHODOLOGY

Data Sources:

- Data was collected from publicly available SpaceX API and additional online datasets, including information on launch dates, sites, payloads, rocket specifications, and mission outcomes

Data Wrangling:

- Data was cleaned to handle missing values, inconsistencies, and errors. Unnecessary columns were removed, and categorical data was encoded for analysis.

Exploratory Data Analysis (EDA) - Visualization and SQL Analysis:

- Used SQL queries to summarize data on launch sites, payload mass, record of landing and mission outcomes. Visualizations were generated to explore correlations, distributions, and patterns.

Interactive Visual Analytics:

- **Mapping Launch Sites with Folium:** Visualized launch site locations and success rates using interactive maps.
- **Interactive Dashboards with Plotly Dash:** Created dashboards for exploring mission outcomes by launch site, booster version, and payload type.

Predictive Analysis:

- **Classification Modeling:** Developed classification models (e.g., Logistic Regression, Decision Trees) to predict mission success based on payload, booster version, and other features.
- **Model Tuning and Evaluation:** Performed hyperparameter tuning using Grid Search. Evaluated model accuracy, precision, and recall to select the best model for predicting mission success.

Data Collection

Data Sources:

- **SpaceX API:** Collected data on past missions, launch sites, rocket types, payloads, and mission outcomes.

Collection Process:

- **API Requests:** Data extraction using Python scripts to pull json data from the SpaceX API.
- **Data Aggregation:** Combined data from different sources to create a unified dataset.
- **Data Cleaning:** Removed duplicates, filled missing values, and standardized formats for columns like dates, payload weights, and rocket types.

Flowchart:

API Request ⟶ Data Aggregation ⟶ Data Cleaning ⟶ Unified Dataset

Data Collection – SpaceX API

- **SpaceX REST API Calls**: Used Python's requests library to interact with the SpaceX REST API.
- **Endpoints**: Focused on endpoints for launches, rockets, launchpads, and payloads to gather comprehensive mission data.

- **Key API Calls:**
    1. **Core Data**: /v4/core – Retrieved cores data such as gridfins, legs, serial, etc
    2. **Rocket Data**: /v4/rockets – Gathered information on rocket types, specifications, and engine details.
    3. **Launchpad Data**: /v4/launchpads – Collected geographic locations of the launch sites (latitude and longitude).
    4. **Payload Data**: /v4/payloads – Extracted data on payload mass and orbit
    5. **Past Launch Data**: v4/launches/past – Data on past Spacex launch record

- **Data Collection Process:**
    1. **Initiate REST Calls**: Used Python scripts to access SpaceX API endpoints.
    2. **Extract & Store Data**: Parsed JSON responses and stored in a structured format.
    3. **Data Cleaning**: Removed duplicates and standardized data formats.
    4. **Integrated Dataset**: Combined results from all endpoints into a single dataset for analysis.

Flowchart:

Initiate REST Calls ⟶ Extract JSON Responses ⟶ Parse & Store Data ⟶ Data Cleaning ⟶ Integrated Dataset

Data Collection - Scraping

- **Python Libraries**: Used BeautifulSoup and requests to scrape SpaceX-related data from spacex website
- **Targeted Website**: Scraped data from publicly available SpaceX site.

- **Key Web Scraping Steps:**
    1. **Send HTTP Request**: Accessed SpaceX target website pages with the requests library.
    2. **Parse HTML Content**: Used BeautifulSoup to identify and locate specific HTML tags containing relevant information.
    3. **Extract Data**: Focused on launch details, news articles, and mission information.
    4. **Data Cleaning**: Standardized fields such as date formats and removed duplicates.
    5. **Store Data**: Saved cleaned data in structured formats (CSV or database) for further analysis.

- **Data Collection Process:**
    1. **Send HTTP Request** ⟶ Download HTML page content from the SpaceX site.
    2. **HTML Parsing** ⟶ Use BeautifulSoup to locate specific tags (, , etc.) containing launch and mission data.
    3. **Data Extraction** ⟶ Extract text data on launch names, dates, success statuses, and mission descriptions.
    4. **Data Cleaning & Storage** ⟶ Format, clean, and store data for analysis in CSV or database.

Flowchart:

HTTP Request ⟶ HTML Parsing ⟶ Data Extraction ⟶ Data Cleaning ⟶ Store Data

Data Wrangling

- **Objective:** Clean and transform raw SpaceX data for analysis, including both API-collected data and web-scraped data.

- **Data Integration:** Combined data from multiple sources (SpaceX API and web scraping results) into a single dataset.

- **Data Quality Checks:** Ensured consistency in data types, resolved missing values, and removed duplicates.

Key Data Wrangling Steps:

Data Cleaning:
- Removed any duplicate entries across API and web-scraped data.
- Handled missing values by filling, dropping, or estimating based on relevant criteria.

Data Transformation:
- Standardized formats for dates, numeric fields, and categorical data (e.g., converting launch dates to a consistent format).
- Applied data type conversions for numerical analysis (e.g., string to integer or float where applicable).

Feature Engineering:
- Created new columns "Launch Outcome" to differentiate successful outcome (green) with failed ones (red)
- Extracted key metrics from text fields for further analysis.

Data Structuring:
- Organized cleaned data into a structured format (e.g., CSV, DataFrame) for use in SQL, EDA, and model training.

Data Wrangling Process Flowchart:

Data Cleaning ⟶ Data Transformation ⟶ Feature Engineering ⟶ Data Structuring

EDA with Data Visualization

Launch Sites Location and its Proximity to Coastline, Highway and Railway (Map)

- **Purpose**: Visualize the geographic distribution of launch sites along with the success rate of launches at each location.
- **Insights**: Helps determine if certain sites have higher success rates due to geographic or operational factors

Payload vs. Launch Site (Scatterplot)

- **Purpose**: Examine the relationship between payload mass, launch sites and launch success
- **Insights**: Determines if heavier payloads correlate with a lower or higher success rate

Flight Number vs. Launch Site (Scatterplot)

- **Purpose**: Examine which sites have frequent number of launch and what the success rates are.
- **Insights**: Determines if certain launch site is more prone to mission success and which ones are not

Success Rate vs. Orbit Type (Scatterplot)

- **Purpose**: Compare the success rate across different orbit types to identify any correlation between orbit types and launch outcomes.
- **Insights**: Useful for identifying which orbit are more prone to success

Flight Number vs. Orbit Type (Scatterplot)

- **Purpose**: Determine which orbits have higher number of mission and what the success rate are.
- **Insights** : Useful in indicating which orbit is having higher success rate

Payload vs. Orbit Type

- **Purpose**: Examine the relationship between payload mass, launch sites and launch success
- **Insights**: Determines if heavier payloads correlate with a lower or higher success rate, aiding in understanding payload limitations

Launch Success Yearly Trend

- **Purpose**: Illustrate the trend of successful launches over years
- **Insights**: Shows how launch success has changed over time, which can reveal rising quality of launches

Visualizations Summary: Each visualization was selected to provide a unique perspective on factors affecting SpaceX's launch operations, including success trends, geographical insights, orbit, payload impact, and overall reliability.

EDA with SQL

Launch Success Rate by Year

- Query to calculate the yearly success rate of launches.
- Helps identify trends in launch success over time.

Launch Success Rate by Launch Site

- Query to determine success rates at each launch site.
- Useful for evaluating performance variations across locations.

Rocket Type Success Analysis

- Query to group launches by rocket type and calculate their respective success rates.
- Provides insight into which rocket types are more reliable.

Average Payload Mass for Successful Launches

- Query to find the average payload mass for successful launches.
- Helps analyze if payload mass affects launch success.

Launch Outcome Distribution

- Query to count the total number of successful vs. unsuccessful launches.
- Gives a quick view of the overall success rate of SpaceX launches.

Build an Interactive Map with Folium

Launch Site Markers

- Added markers for each SpaceX launch site.
- Allows users to quickly locate and identify each launch site's location on the map.

Circle Markers for Launch Site Proximity

- Added circles around each launch site with a specified radius.
- Visually highlights the area surrounding each launch site, showing the proximity region for safety analysis.

Popups with Launch Site Details

- Enabled popup windows with additional information (e.g., site name,.

Color-Coded Markers for Launch Outcomes

- Green for successful launch, red for failed one.

Build a Dashboard with Plotly Dash

Plots and Graphs Included

- **Payload vs. Launch Outcome (Scatter Plot)**
  - **Purpose:** To visualize the relationship between payload weight and launch outcomes, highlighting any correlations between payload size and success.
- **Launches Site by Success Rate (Pie Chart)**
  - **Purpose:** To illustrate the proportion of successful launches by the launch sites.

Interactions Added

Dynamic Dropdown Menus:

- **Functionality:** Users can select specific Lauch site to see the breakdown in terms of success rate and payload

Predictive Analysis (Classification)

Model Development Process Overview

The predictive analysis for classification in the SpaceX project involved a systematic approach to building, evaluating, improving, and identifying the best-performing classification model. The following summarizes the key steps in this process:

Data Preparation:
- Data Cleaning: Handled missing values and outliers.
- Feature Selection: Identified and selected relevant features (e.g., payload, rocket type).

Data Splitting:
- Train-Test Split: Divided the dataset into training (80%) and testing (20%) subsets to evaluate model performance.

Model Selection:
- Algorithm Choice: Evaluated various classification algorithms (e.g., Logistic Regression, Decision Trees, Random Forest, Support Vector Machine).

Model Training:
- Fit Models: Trained each model using the training dataset.

Model Evaluation:
- Performance Metrics: Assessed models using accuracy, precision, recall, F1-score, and confusion matrix.
- Cross-Validation: Employed k-fold cross-validation to ensure robustness.

Model Improvement:
- Hyperparameter Tuning: Used Grid Search or Random Search to optimize hyperparameters.
- Feature Engineering: Experimented with creating new features or modifying existing ones.

Best Model Selection:
- Final Comparison: Compared model performance metrics to identify the top-performing model.
- Model Validation: Validated the selection

EDA (Exploratory Data Analysis)

1. Exploratory Data Analysis (EDA) Results

- **Launch Success Rate:** Visualization of success rates over time, showing an upward trend in successful launches.
- **Payload Distribution:** Analysis of payload weights indicating that heavier payloads tend to have varied success rates.
- **Rocket Type Performance:** Insights into which rocket types (e.g., Falcon 9, Falcon Heavy) have higher success rates, highlighting the Falcon 9 as the most reliable option.
- **Launch Frequency:** Visualization of launch counts by year, revealing an increase in launch activity over the years.

2. Interactive Analytics Demo

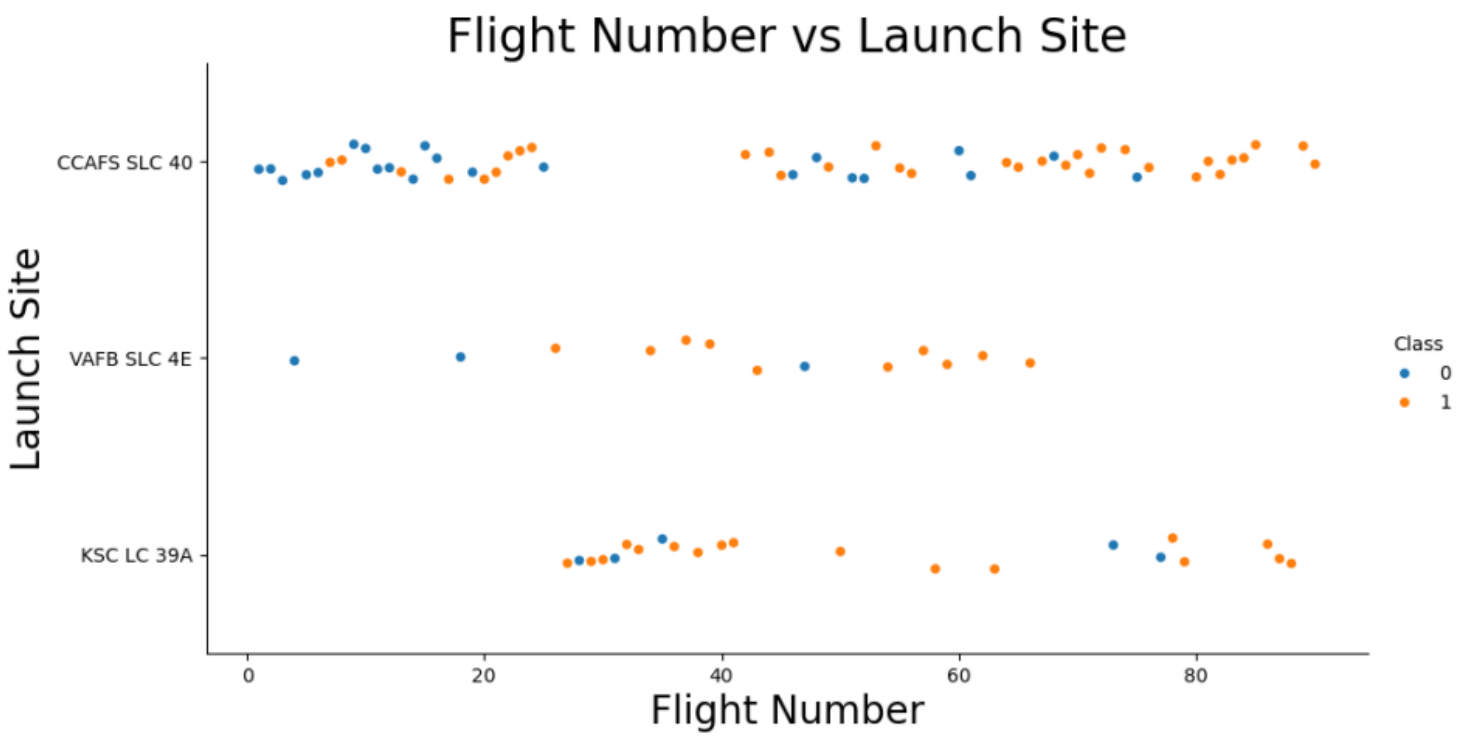
- **Launch Success Rate Graph:** A bar chart showing success rates with interactivity to filter by year and rocket type.
- **Payload vs. Launch Outcome:** A scatter plot with hover tooltips providing additional details on each launch.
- **Geographical Map:** A map visualization showing the locations of launch sites

The EDA results provided valuable insights into SpaceX's launch performance and trends

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Overall Launch Distribution:

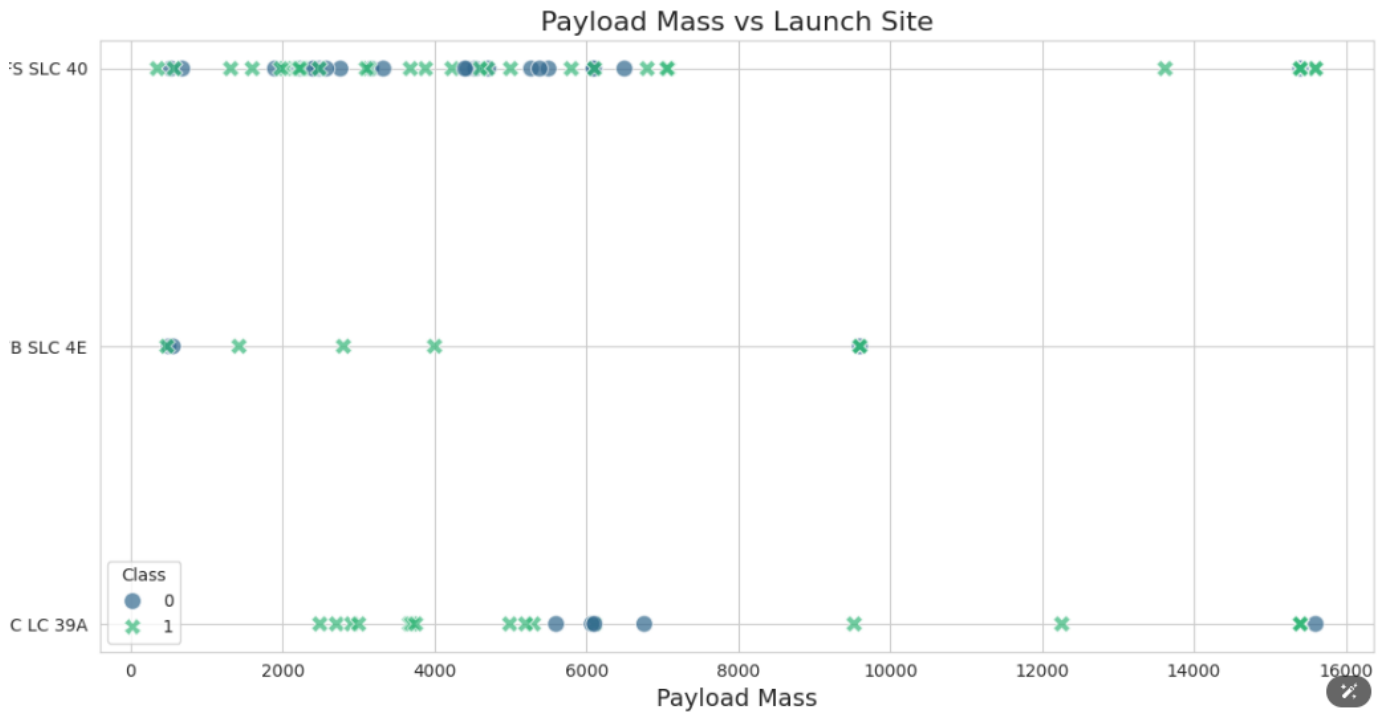
- The majority of launches have occurred from CCAFS SLC 40, followed by KSC LC 39A, and then VAFB SLC 4E.

Launch Site Usage Over Time:

- CCAFS SLC 40 was the primary launch site for the early flights.
- As the flight number increased, KSC LC 39A became more frequently used.
- VAFB SLC 4E has been used less consistently, with some clusters of launches and longer gaps in between.

Potential Clustering:

- There appear to be clusters of launches at specific flight numbers within each launch site, suggesting potential grouping of missions with similar objectives or payloads.



PAYLOAD VS. LAUNCH SITE

Overall Observations:

- **CCAFS SLC 40** is the most frequently used launch site, with a high number of launches across a wide range of payload masses.
- **VAFB-SLC 4E** does not launch rockets with heavy payload mass (greater than 10000).
- **KSC LC 39A** has a mix of launches with both low and high payload masses.

Success Rates:

- **CCAFS SLC 40** shows a higher concentration of successful launches across all payload masses, suggesting a strong track record.
- **VAFB SLC 4E** has a mix of successful and unsuccessful launches, but the successful ones tend to be associated with higher payload masses.
- **KSC LC 39A** has a lower success rate overall, with unsuccessful launches distributed across different payload mass ranges.

Potential Insights:

- **CCAFS SLC 40** appears to be a reliable launch site for a wide range of payloads, making it a preferred choice for SpaceX.
- **VAFB SLC 4E** might be better suited for heavier payloads, although its success rate is less consistent.
- **KSC LC 39A** shows a more varied performance, with both successful and unsuccessful launches across different payload categories.

Success Rate vs. Orbit Type

Overall Success Rate:

- The majority of orbit types have a high success rate, with many exceeding 90%.

Orbit Types with High Success Rates:

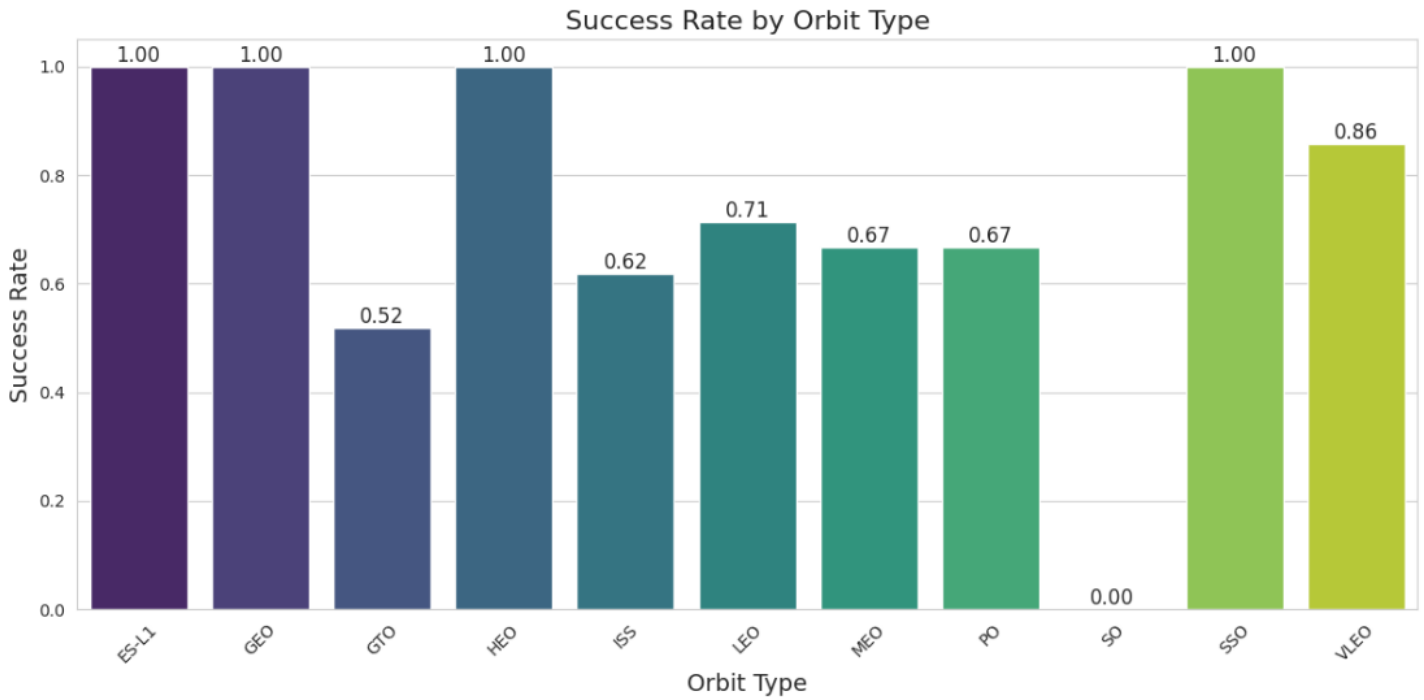
- **ES-L1, GEO, GTO, HEO, SSO, VLEO** all have perfect success rates of 100%.
- **LEO** has a success rate of 98.6%.

Orbit Types with Lower Success Rates:

- **MEO** has a success rate of 71%.
- **PO** has a success rate of 67%.

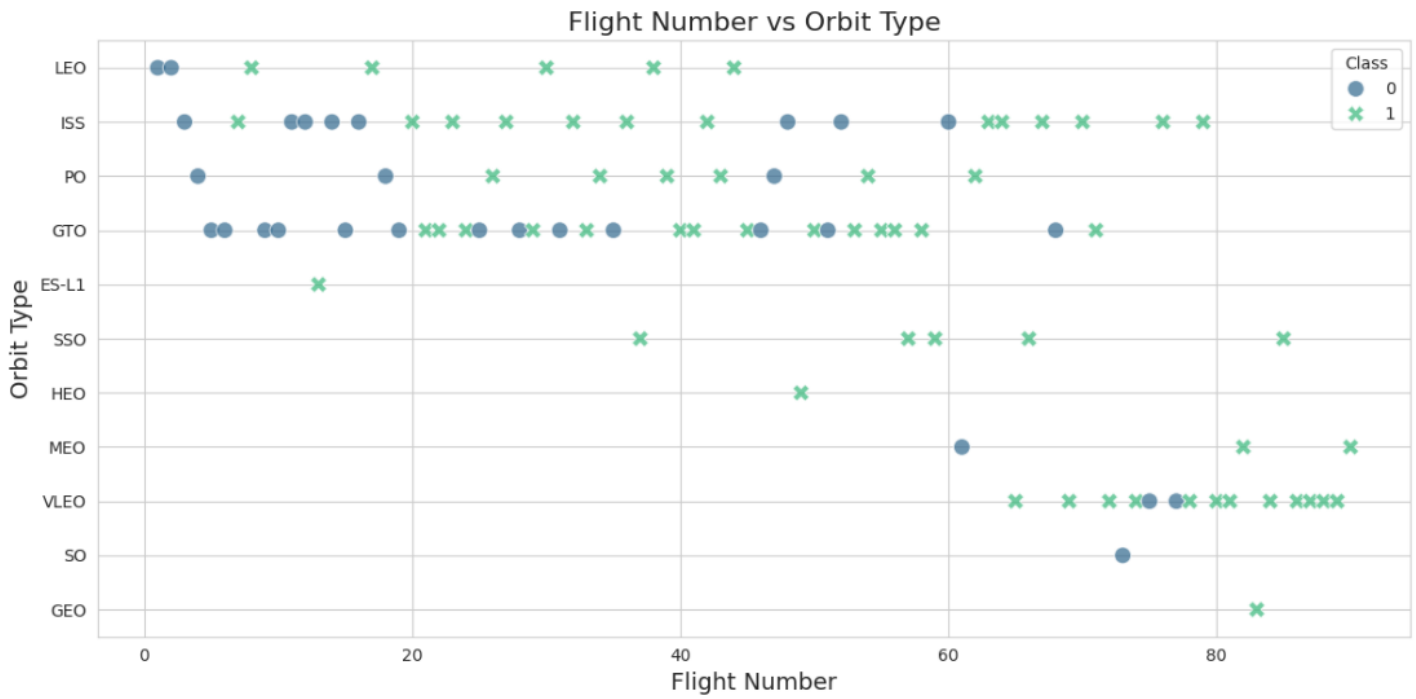
Potential Explanations for Lower Success Rates:

- **MEO and PO:** These orbit types might have specific technical challenges or require more complex maneuvers, which could contribute to a slightly lower success rate.



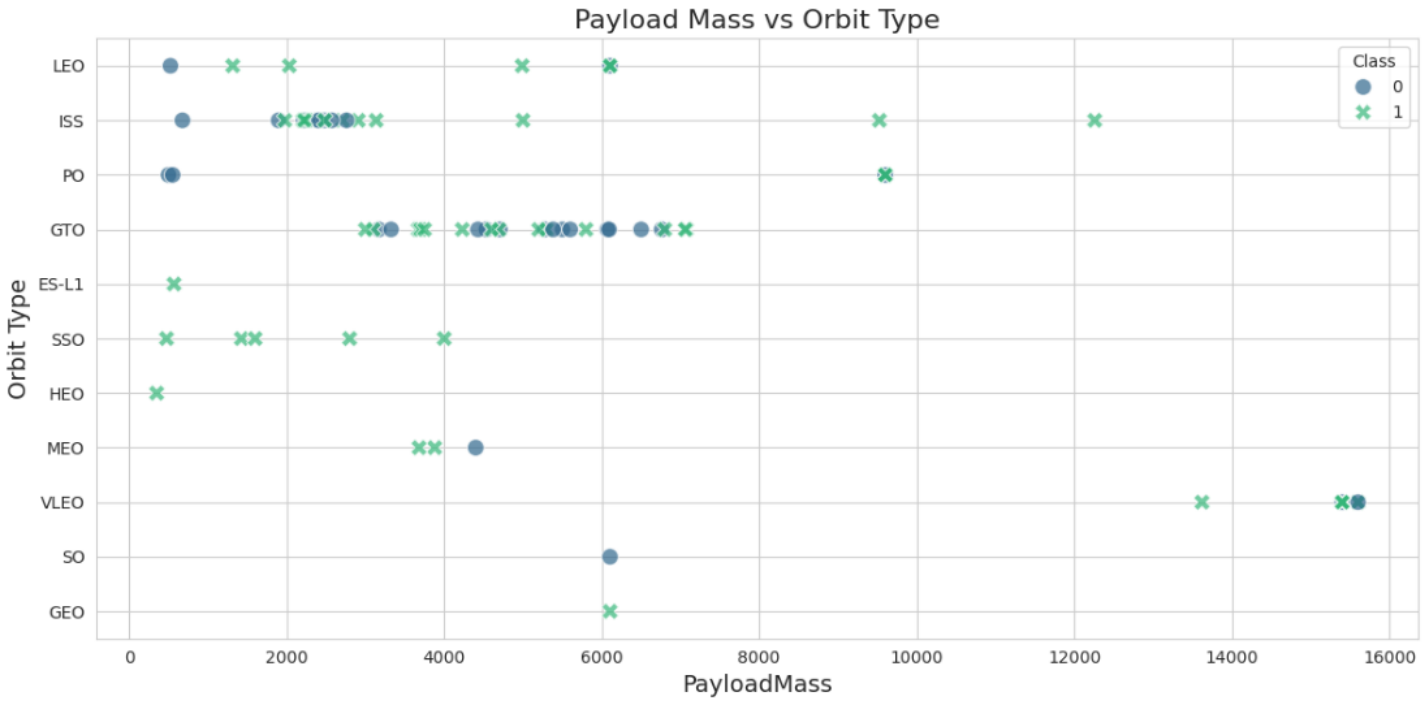
Flight Number vs. Orbit Type

in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

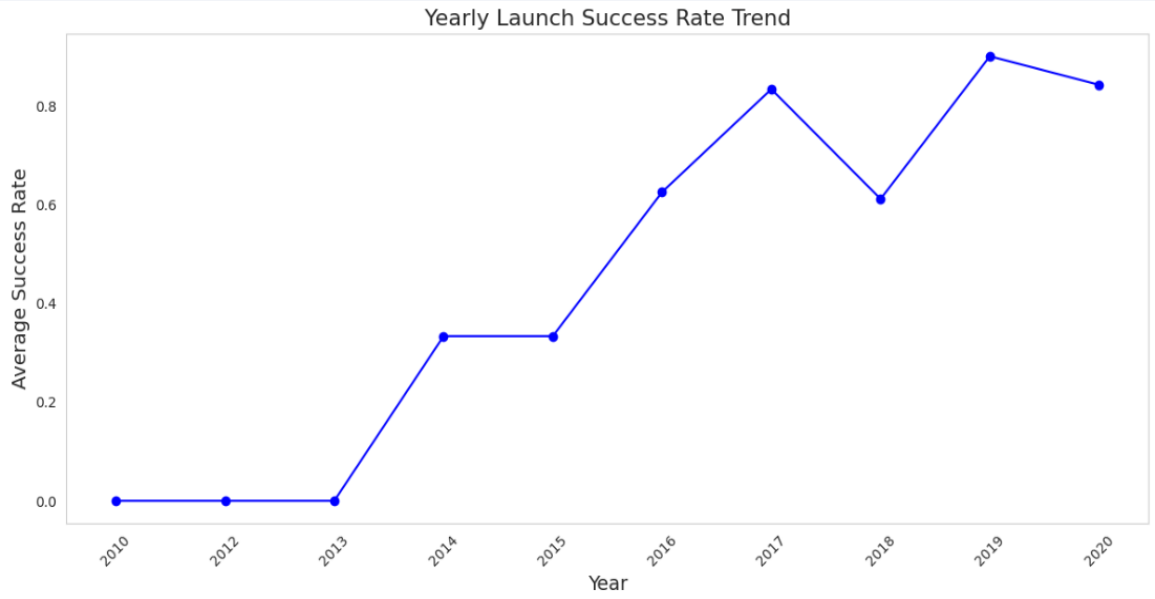


Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend



The success rate since 2013 kept increasing till 2020

All Launch Site Names

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to get unique launch sites
query = "SELECT DISTINCT Launch_Site FROM SPACEXTBL;"
unique_launch_sites = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the results
print(unique_launch_sites)
```

```
    Launch_Site
0   CCAFS LC-40
1   VAFB SLC-4E
2    KSC LC-39A
3  CCAFS SLC-40
```

There are 4 Launch Site in SpaceX :

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

Launch Site Names Begin with 'CCA'

```
# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to get records where Launch_Site begins with 'CCA'
query = "SELECT distinct Launch_Site FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%';"
launch_sites_CCA = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the first 5 records
print(launch_sites_CCA.head(5))
```

```
    Launch_Site
0   CCAFS LC-40
1  CCAFS SLC-40
```

Launch Sites starting with 'CCA' :

1. CCAFS LC-40
2. CCAFS SLC-40

Total Payload Mass

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to calculate total payload mass for NASA (CRS) launches
query = """
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass
FROM SPACEXTBL
WHERE Customer = 'NASA (CRS)';
"""
total_payload_mass = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(total_payload_mass)
```

```
   Total_Payload_Mass
0               45596
```

Calculate the total payload carried by boosters from NASA

Total Payload Mass is 45596 kg

Average Payload Mass by F9 v1.1

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to calculate the average payload mass for booster version F9 v1.1
query = """
SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass
FROM SPACEXTBL
WHERE Booster_Version = 'F9 v1.1';
"""
average_payload_mass = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(average_payload_mass)
```

```
   Average_Payload_Mass
0                2928.4
```

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

Average Payload Mass is 2928.4 kg

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to find the date of the first successful landing outcome on a ground pad
query = """
SELECT min(Date) AS First_Successful_Ground_Landing_Date, Landing_Outcome
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success';
"""

first_successful_landing_date = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(first_successful_landing_date)
```

```
  First_Successful_Ground_Landing_Date Landing_Outcome
0                           2018-07-22         Success
```

First Successful Ground Landing Date

THE FIRST SUCCESSFUL LANDING OUTCOME DATE WAS 22 JULY 2018

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to find boosters with successful drone ship landings and specified payload mass
query = """
SELECT DISTINCT Booster_Version
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success'
AND PAYLOAD_MASS__KG_ > 4000
AND PAYLOAD_MASS__KG_ < 6000;
"""
boosters = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(boosters)
```

```
  Booster_Version
0   F9 B5 B1046.2
1   F9 B5 B1047.2
2   F9 B5 B1048.3
3  F9 B5 B1051.2
4    F9 B5B1060.1
5  F9 B5 B1058.2
6    F9 B5B1062.1
```

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

1. F9 B5 B1046.2
2. F9 B5 B1047.2
3. F9 B5 B1048.3
4. F9 B5 B1051.2
5. F9 B5B1060.1
6. F9 B5 B1058.2
7. F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to count landing outcomes in the specified date range
query = """
SELECT
    Mission_Outcome,
    COUNT(Mission_Outcome) AS Outcome_Count
FROM SPACEXTBL
GROUP BY Mission_Outcome;
"""

# Execute the query and load the result into a DataFrame
mission_outcomes_class = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(mission_outcomes_class)
```

```
                    Mission_Outcome  Outcome_Count
0               Failure (in flight)              1
1                           Success             98
2                           Success              1
3  Success (payload status unclear)              1
```

Number of Successful Mission Outcome = 100

Number of Failure one : 1

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass:

1. F9 B5 B1048.4
2. F9 B5 B1049.4
3. F9 B5 B1051.3
4. F9 B5 B1056.4
5. F9 B5 B1048.5
6. F9 B5 B1051.4
7. F9 B5 B1049.5
8. F9 B5 B1060.2
9. F9 B5 B1058.3
10. F9 B5 B1051.6
11. F9 B5 B1060.3
12. F9 B5 B1049.7

```python
import sqlite3
import pandas as pd

# Connect to the SQLite database
conn = sqlite3.connect('my_data1.db')

# Run the SQL query to find booster versions with maximum payload mass using a subquery
query = """
SELECT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL
);
"""

# Execute the query and load the result into a DataFrame
max_payload_boosters = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(max_payload_boosters)
```

```
   Booster_Version
0    F9 B5 B1048.4
1    F9 B5 B1049.4
2    F9 B5 B1051.3
3    F9 B5 B1056.4
4    F9 B5 B1048.5
5    F9 B5 B1051.4
6    F9 B5 B1049.5
7   F9 B5 B1060.2
8   F9 B5 B1058.3
9    F9 B5 B1051.6
10   F9 B5 B1060.3
11  F9 B5 B1049.7
```

2015 Launch Records

List of failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

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

```
# Run the SQL query to get records for failed drone ship landings in 2015
query = """
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) = '05' THEN 'May'
        WHEN substr(Date, 6, 2) = '06' THEN 'June'
        WHEN substr(Date, 6, 2) = '07' THEN 'July'
        WHEN substr(Date, 6, 2) = '08' THEN 'August'
        WHEN substr(Date, 6, 2) = '09' THEN 'September'
        WHEN substr(Date, 6, 2) = '10' THEN 'October'
        WHEN substr(Date, 6, 2) = '11' THEN 'November'
        WHEN substr(Date, 6, 2) = '12' THEN 'December'
    END AS Month,
    Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTBL
WHERE substr(Date, 0, 5) = '2015'
AND Landing_Outcome LIKE '%Failure%';
"""

# AND Landing_Outcome LIKE 'Failure';
#"""

# Execute the query and load the result into a DataFrame
failed_landings_2015 = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(failed_landings_2015)
```

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

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

No	Landing Outcome	Number of Occurrence
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	2
7	Precluded (drone ship)	1

```python
# Run the SQL query to count landing outcomes in the specified date range
query = """
SELECT
    Landing_Outcome,
    COUNT(*) AS Outcome_Count
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Outcome_Count DESC;
"""

# Execute the query and load the result into a DataFrame
landing_outcomes_ranked = pd.read_sql_query(query, conn)

# Close the connection
conn.close()

# Display the result
print(landing_outcomes_ranked)
```

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

Section 3

Launch Sites Proximities Analysis

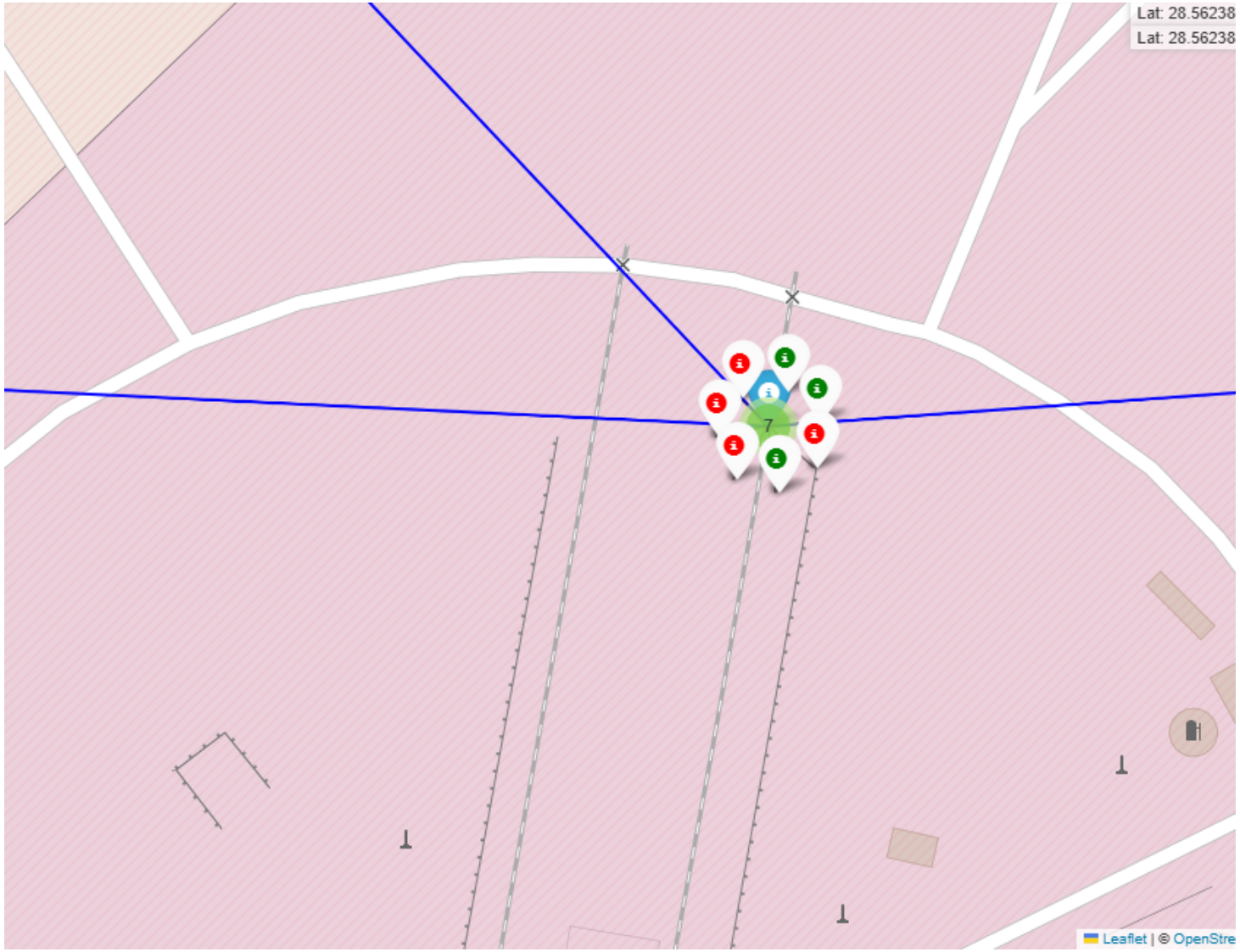
Locations of Launch Sites

Geographic Distribution:

- **East Coast:** Cape Canaveral Space Force Station (CCAFS) and Kennedy Space Center (KSC) are located on the east coast of Florida, providing convenient access to the Atlantic Ocean for launches into various orbits.
- **West Coast:** Vandenberg Space Force Base (VAFB) is situated on the west coast of California, providing access to the Pacific Ocean for launches into polar and sun-synchronous orbits.

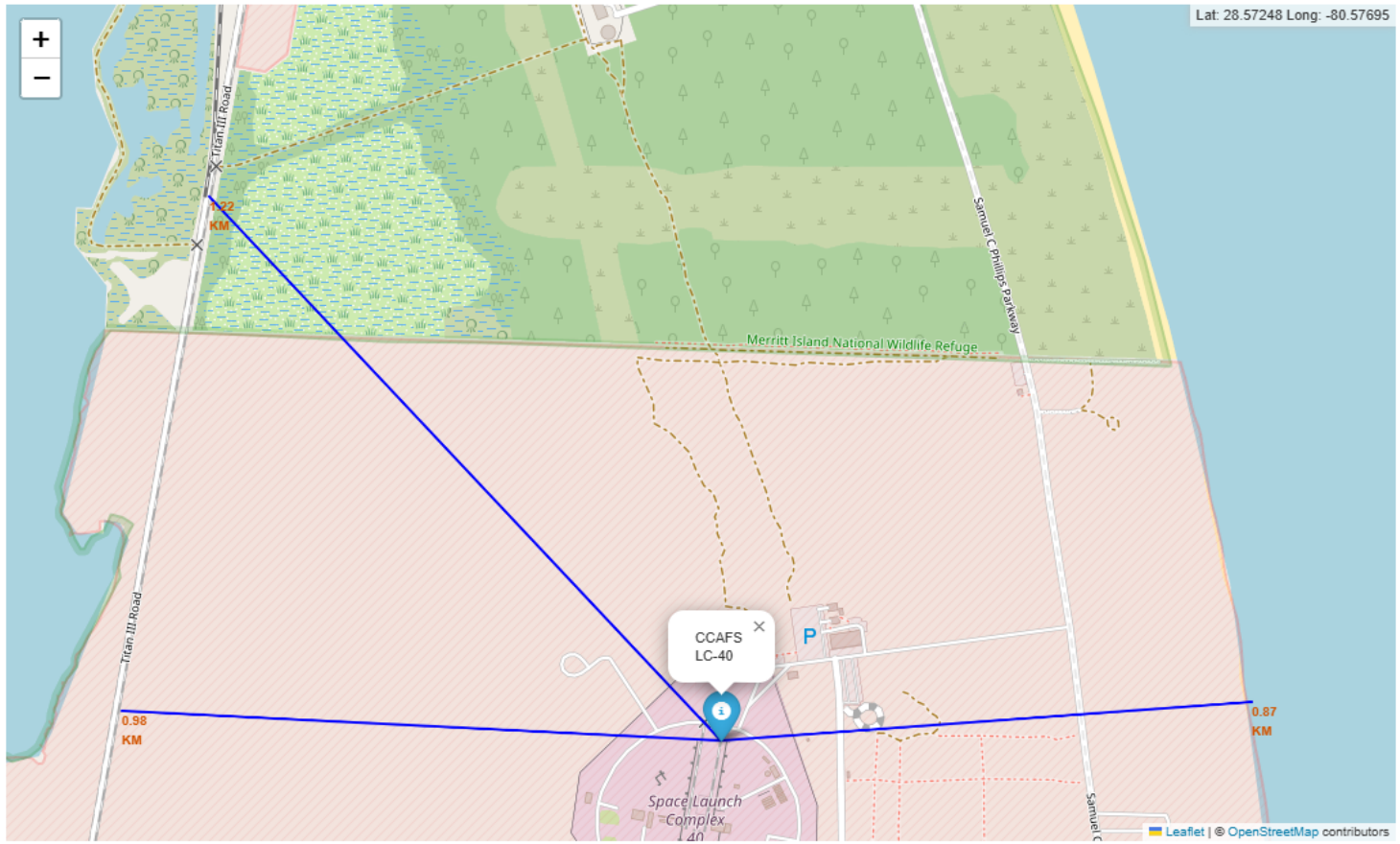
Potential Insights:

- **Diverse Launch Capabilities:** The strategic placement of launch sites on both coasts allows SpaceX to launch a wide range of missions, including those requiring specific orbital inclinations or launch windows.
- **Proximity to Facilities:** The proximity of these launch sites to SpaceX's manufacturing and testing facilities in California and Texas facilitates efficient operations and logistics.



Map Location of Successful / Unsuccessful Outcomes

- **Successful outcomes are represented by green markers, while the unsuccessful ones are represented by red ones.**



Proximity of Lauch Sites to Coastline, Highway and Railway

Key Observations:

- The map highlights the location of CCAFS LC-40, which is situated on the eastern coast of Florida, near the Atlantic Ocean.
- The blue lines likely represent the trajectory of a rocket launched from this pad.
- The presence of the "Space Launch Complex" label further confirms the location of the launch site.

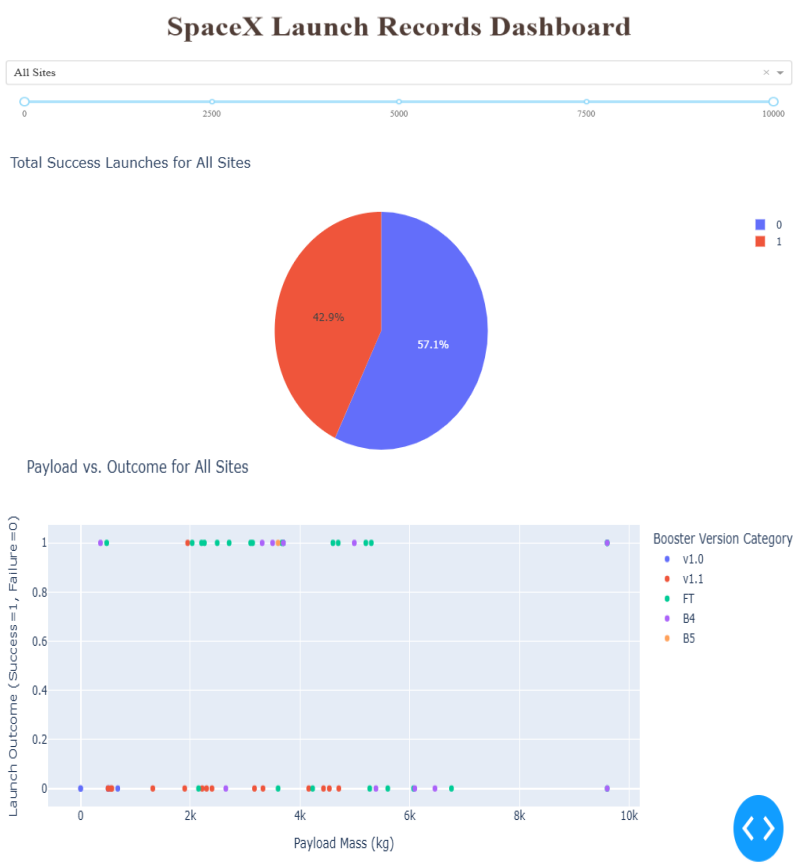
Potential Insights:

- **Strategic Location:** The location of CCAFS LC-40 on the eastern coast provides easy access to the Atlantic Ocean, making it suitable for launches into various orbits, including geostationary orbit (GEO) and geosynchronous transfer orbit (GTO).

Section 4

Build a Dashboard with Plotly Dash

Successful Launches (All Sites)



Overall Success Rate:

- The dashboard indicates that the overall success rate for all SpaceX launches is 57.1%. This means that approximately 57% of SpaceX launches have been successful.

Payload vs. Outcome for All Sites:

- The scatter plot shows the relationship between payload mass and launch outcome (success or failure) for different booster versions.
- We can observe that the majority of successful launches (represented by blue dots) have payload masses ranging from 2,000 kg to 10,000 kg.
- There are a few unsuccessful launches (red dots) scattered across different payload mass ranges.

Booster Version Category:

- The scatter plot also includes a legend indicating the booster version category for each data point.
- This suggests that different booster versions might have varying success rates or be better suited for specific payload ranges.

Potential Insights:

- SpaceX has achieved a relatively high success rate of 57.1% for all its launches.
- The success rate appears to be higher for launches with payload masses between 2,000 kg and 10,000 kg.
- Different booster versions might have varying performance characteristics, influencing their success rates and suitability for different payload ranges.

Launch Site with Highest Success Ratio

Total Success Launches:

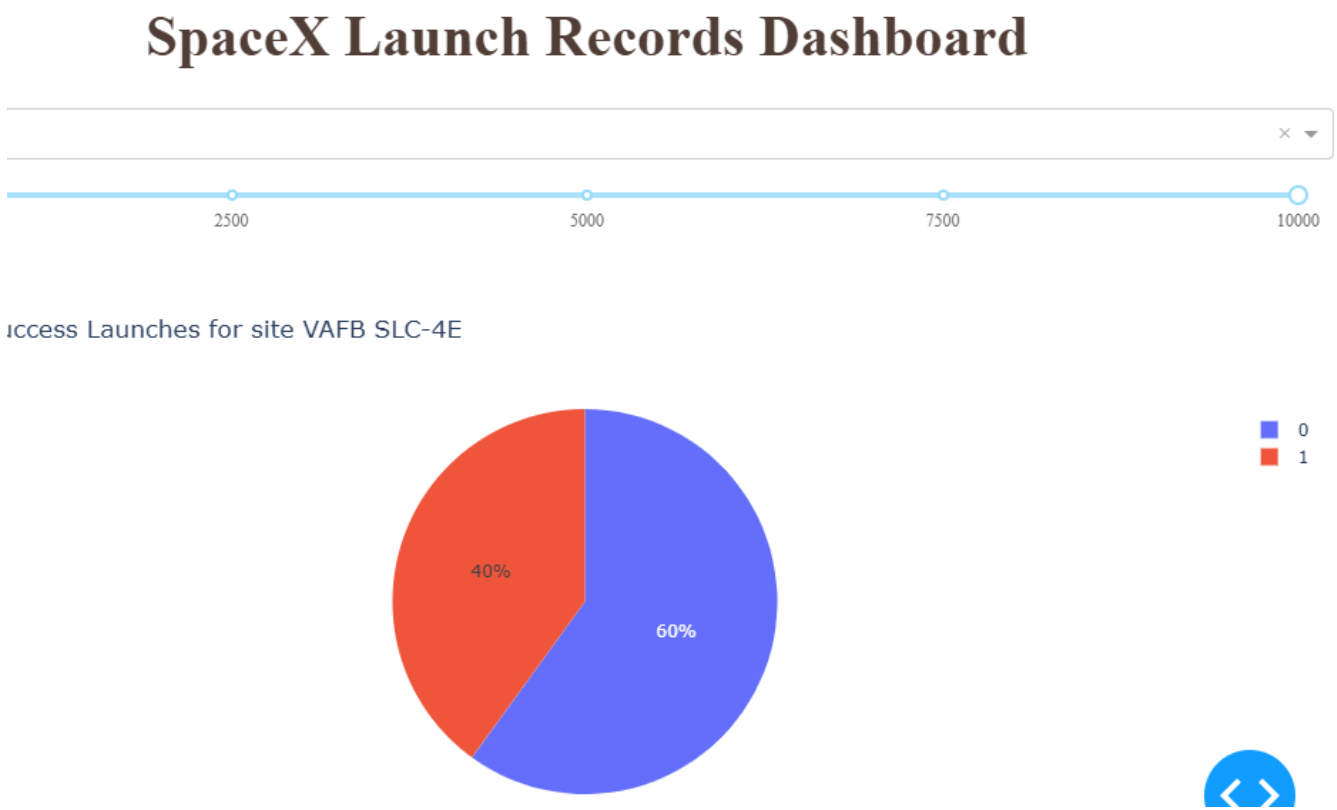
- The dashboard indicates that there have been a total of 10 launches from VAFB SLC-4E.
- Out of these 10 launches, 6 have been successful, resulting in a success rate of 60%.

Key Observations:

- VAFB SLC-4E has a relatively low number of launches compared to other SpaceX launch sites.
- Despite the limited number of launches, the success rate of 60% is respectable.

Potential Insights:

- VAFB SLC-4E might be used for specific types of missions or payloads that require its unique characteristics.
- The lower number of launches could be due to various factors, such as limited availability, specific mission requirements, or logistical constraints.



Payload vs. Launch Outcome

Overall Observations:

- The graph shows the relationship between payload mass and launch outcome (success or failure) for different booster versions.
- The majority of successful launches (represented by blue dots) have payload masses ranging from 2,000 kg to 10,000 kg.
- There are a few unsuccessful launches (red dots) scattered across different payload mass ranges.

Booster Version Category:

- The graph includes a legend indicating the booster version category for each data point.
- This suggests that different booster versions might have varying success rates or be better suited for specific payload ranges.

Potential Insights:

- SpaceX has achieved a relatively high success rate for launches with payload masses between 2,000 kg and 10,000 kg.
- Different booster versions might have varying performance characteristics, influencing their success rates and suitability for different payload ranges.

Section 5

Predictive Analysis (Classification)

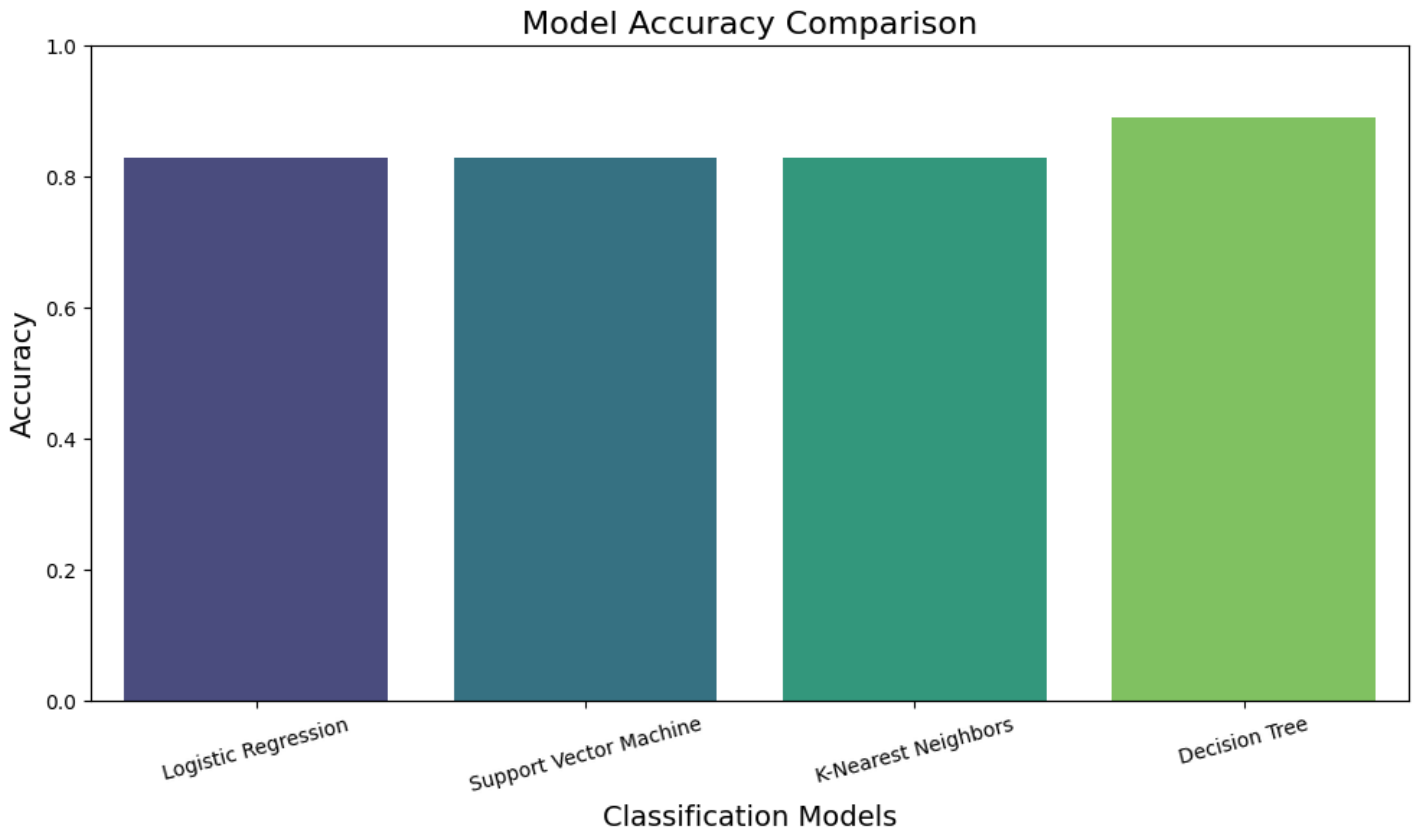
Accuracy of the Classification Models

Overall Comparison:

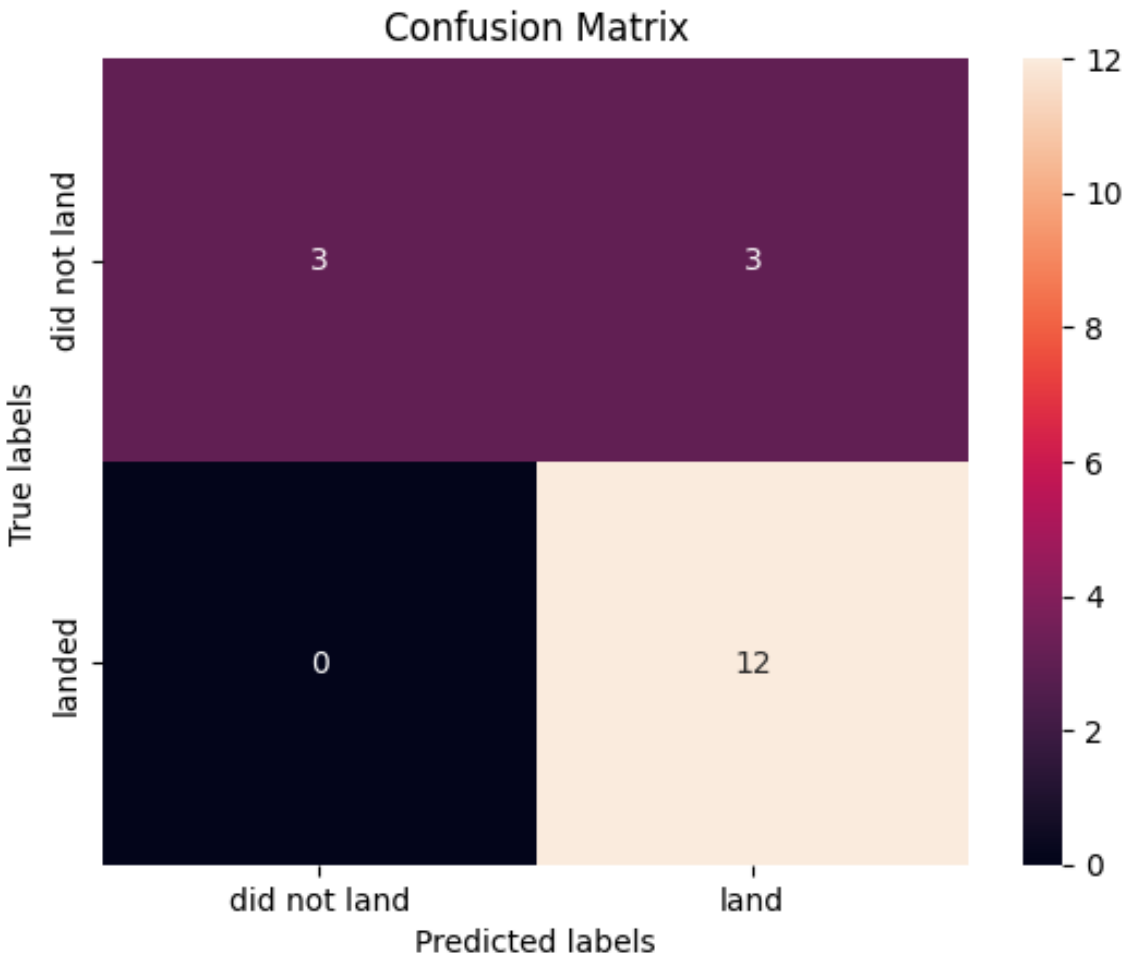
- **The Decision Tree model** has the highest accuracy among the four, with an accuracy of **0.89.**
- Logistic Regression and Support Vector Machine models have similar accuracy levels, both around 0.83.
- K-Nearest Neighbors model has the lowest accuracy among the four, with an accuracy of 0.82.

Potential Explanations:

- The Decision Tree model might be particularly well-suited for the specific dataset and classification problem being addressed.
- The complexity of the decision boundaries in the dataset could be better captured by the Decision Tree model compared to the other models.
- The hyperparameters of the Decision Tree model might have been tuned effectively to achieve this high accuracy.




```
The model with the highest classification accuracy is: Decision Tree with an accuracy of 0.89
```



Evaluation of the Model (Confusion Matrix)

From this, we can calculate some performance metrics:

- **Accuracy:** (TP + TN) / (TP + TN + FP + FN) = (12 + 3) / (12 + 3 + 3 + 0) = 0.83 or 83%
- **Precision:** TP / (TP + FP) = 12 / (12 + 3) = 0.8 or 80%
- **Recall:** TP / (TP + FN) = 12 / (12 + 0) = 1.0 or 100%
- **F1-Score:** 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.8 * 1.0) / (0.8 + 1.0) = 0.89 or 89%

Interpretation:

- The model has an overall accuracy of 83%, meaning it correctly predicted the landing outcome for 83% of the cases.
- The precision of 80% indicates that when the model predicts a landing, it is correct 80% of the time.
- The recall of 100% means that the model correctly identifies all actual landing events.
- The F1-score of 89% provides a balanced measure of precision and recall.

Overall, the model performs well in predicting rocket landings, with a good balanced F1-score.

Conclusions

SpaceX Launch Success:

- The majority of SpaceX launches have been successful, especially in recent years.
- Different launch sites have varying success rates, with some sites showing higher success rates than others.
- The success of a launch is influenced by factors like payload mass, orbit type, and booster version.

Launch Site Performance:

- CCAFS SLC 40 is the most frequently used launch site and has a relatively high success rate.
- VAFB SLC 4E has a lower number of launches but a respectable success rate.

Payload and Orbit Success:

- SpaceX has demonstrated strong capabilities in launching missions to LEO, ISS, PO, GTO, and ES-L1.
- SSO and HEO launches might be more challenging due to specific orbital requirements or payload complexities.
- MEO, VLEO, SO, and GEO launches appear to have faced more difficulties.

Model Performance:

- The Decision Tree model outperformed Logistic Regression, Support Vector Machine, and K-Nearest Neighbors in terms of accuracy for a specific classification task.
- The confusion matrix shows that the model has a high accuracy and perfect recall in predicting rocket landings.

Innovative Insights

Predictive Analytics

- **Feature Importance Analysis**: Present an analysis to deep dive which variables (e.g., weather conditions, type of rocket, payload) most significantly impact launch success rates.

Geographic Analysis

- **Optimal Launch Sites**: Analyze geographic data to identify potential new launch sites based on proximity to critical infrastructure (e.g., population density, coastlines, military bases) and assess their viability for future launches.

Technological Innovations

- **Impact of Technology**: Discuss the role of emerging technologies (e.g., reusable rocket technology, AI in operations) in enhancing launch efficiency and success rates.

Economic Analysis

- **Cost-Benefit Analysis**: Conduct an economic analysis to weigh the costs of failed launches against the potential revenue generated from successful missions.

Appendices

Python codes, SQL queries, charts, Notebook outputs, and data sets that created during this project are in the following github link.

Thank you!