IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aparna Narendar
Jan 05, 2026

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using web scraping and SpaceX API;
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive
  - Visual analytics;
  - Machine Learning Prediction.
- Summary of all results
  - It was possible to collected valuable data from public sources;
  - EDA allowed to identify which features are the best to predict success of launchings;
  - Machine Learning Prediction showed the best model to predict which characteristics are
  - important to drive this opportunity by the best way, using all collected data.

Introduction

- Project background and context
  - The objective is to evaluate the viability of the new company Space Y to compete
  - with Space X.
  - Desirable answers:
  - The best way to estimate the total cost for launches, by predicting successful landings of the
  - first stage of rockets;
  - Where is the best place to make launches.
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- ### Data collection methodology:

- This project employs a comprehensive approach to predict the successful landing of the
- Falcon 9 first stage, incorporating data collection, processing, exploratory analysis, interactive visualizations,
- and predictive modeling.
- **Data Collection Methodology**: Data was sourced from the SpaceX API, which provided detailed records of
- Falcon 9 launches, including launch dates, sites, payloads, and outcomes.
- **Perform Data Wrangling**: Data cleaning involved handling missing values, standardizing formats, and ensuring
- consistency. Key features were extracted and new features engineered to enrich the dataset.
- **Perform Exploratory Data Analysis (EDA) Using Visualization and SQL**:
- Visualized launch success rates, payloads, and launch sites using Matplotlib and Seaborn.
- Executed SQL queries to derive insights and answer specific questions regarding the dataset.

- ### Perform data wrangling

- Data cleaning involved handling missing values, standardizing formats, and ensuring
  consistency. Key features were extracted and new features engineered to enrich the dataset.

- ### Perform exploratory data analysis (EDA) using visualization and SQL

- Visualized launch success rates, payloads, and launch sites using Matplotlib and Seaborn.
- • Executed SQL queries to derive insights and answer specific questions regarding the dataset.

Methodology

- Perform interactive visual analytics using Folium and Plotly Dash

- Used Folium to create interactive maps displaying launch sites and outcomes.
- Developed a Plotly Dash application with interactive components like dropdowns and sliders to analyze launch success rates and payload ranges.

- Perform Predictive Analysis Using Classification Models:

- Built and evaluated various classification models including Logistic Regression, SVM, KNN, and Decision
- Trees.
- Employed GridSearchCV for hyperparameter tuning.
- Evaluated models based on accuracy, and identified the best performing model for predicting landing success.

Data Collection

- Describe how data sets were collected.
- Step 1: SpaceX API Request

- Step 2: Web Scraping Wikipedia

- Step 3: Data Integration

Data Collection - SpaceX API

Step 1: Initiate API Request
o Use Python's `requests` library to connect to the SpaceX API.
o Endpoint: `https://api.spacexdata.com/v4/launches`

Step 2: Parse API Response
o Convert API response from JSON to a Python dictionary.
o Extract relevant fields: launch date, launch site, payload mass, rocket type, outcome.

Step 3: Store Data Locally
o Save extracted data into a pandas DataFrame.
o Store the DataFrame locally for further processing.

Data Collection - Scraping

Step 1: Initiate Web Scraping

o Use Python's `requests` library to fetch the HTML content of the Wikipedia page.

o Target URL: `https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches`

Step 2: Parse HTML Content

o Use `BeautifulSoup` to parse the HTML content.

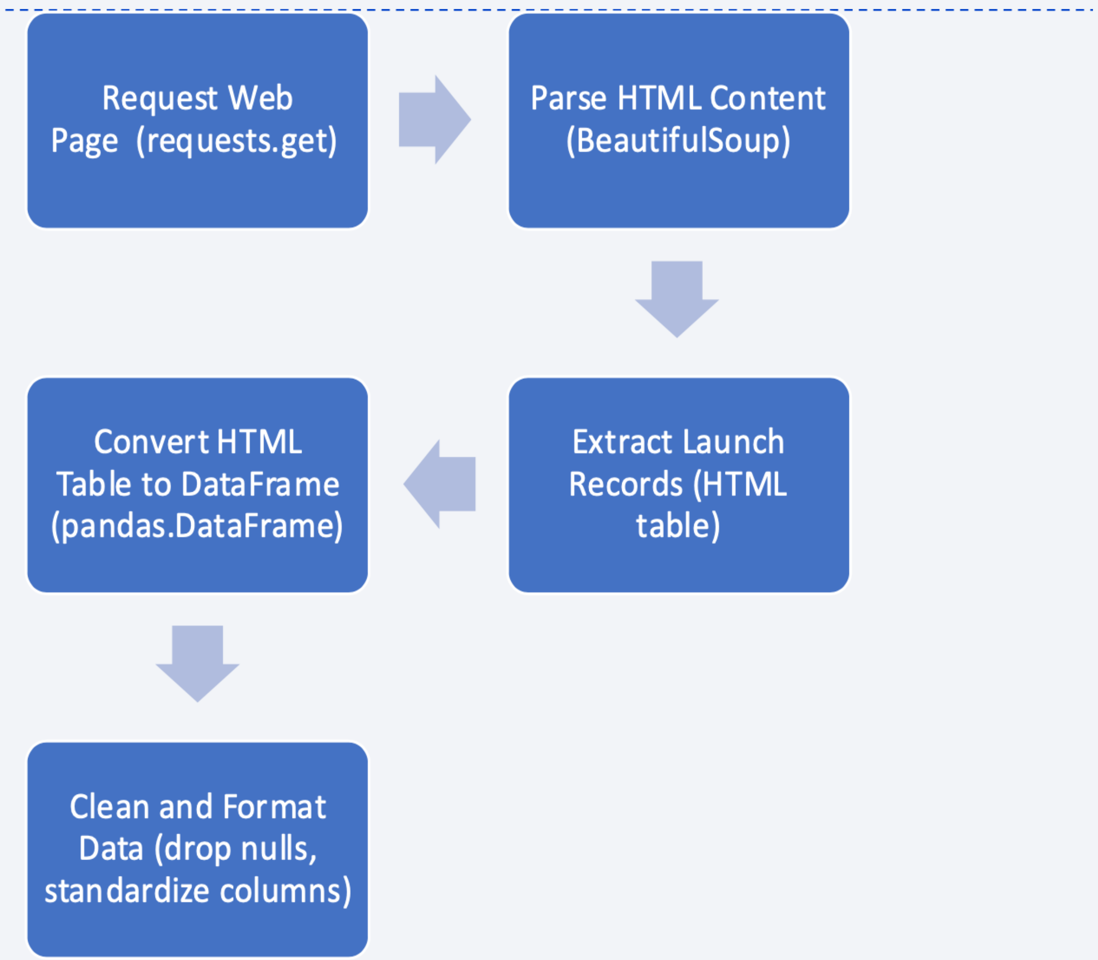
o Extract the HTML table containing Falcon 9 launch records.

Step 3: Convert to DataFrame

o Convert the extracted HTML table into a pandas DataFrame.

o Clean and format the DataFrame, ensuring data consistency.

GitHub URL: 2. jupyter-labs-webscraping.ipynb



Data Wrangling

Overview: Data wrangling involves cleaning, transforming, and organizing raw data into a structured format suitable for analysis.

- Step 1: **Data Cleaning**
  - o Identify and fill or remove missing values in the dataset.
  - o Use appropriate imputation techniques or drop rows/columns with excessive missing data.

- **Step 2: Data Transformation**
  - o Convert data types to appropriate formats (e.g., date-time, numerical).
  - o Standardize text (e.g., lowercase, remove whitespace).
  - o Create new features from existing data (e.g., extract year from date).
  - o Normalize/scale numerical features to ensure consistency.

Data Wrangling

- **Step 3: Data Integration**

o Merge datasets collected from different sources (API, web scraping) into a single cohesive dataset.

o Ensure consistent column names and data formats across datasets.

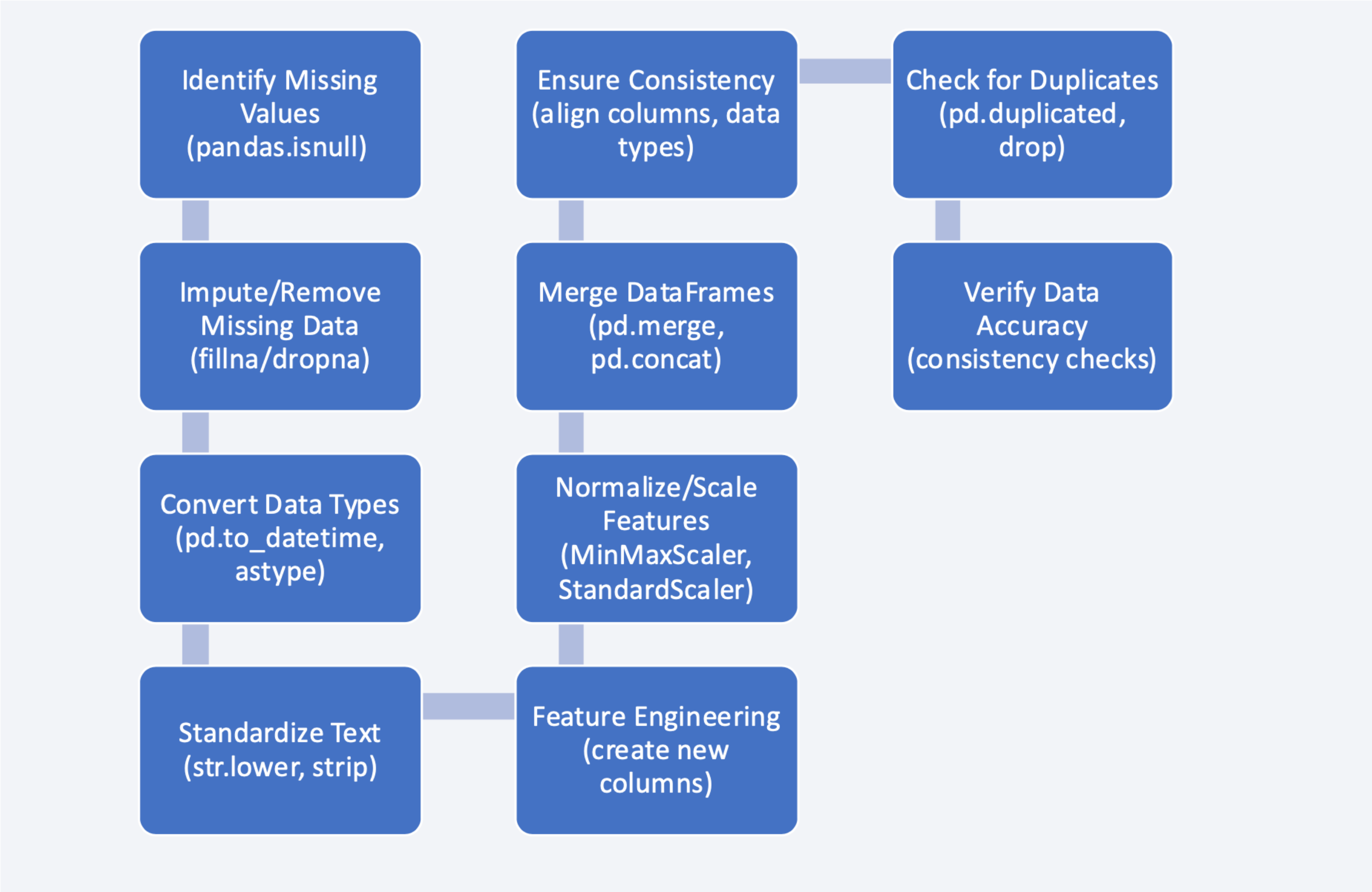
- **Step 4: Data Validation**

o Check for duplicate records and remove them.

o Verify the accuracy and consistency of data entries.

GitHub URL: 3. labs-jupyter-spacex-Data wrangling.ipynb

Data Wrangling - Flowchart



EDA with Data Visualization

Overview:

Exploratory Data Analysis (EDA) involves visually exploring and summarizing the main characteristics of a dataset. The goal is to

understand the data's distribution, identify patterns, and uncover relationships between variables.

Charts Plotted:

1. Histograms:

- Purpose: Used to visualize the distribution of numerical variables such as launch success rates, payload mass, and flight number.

- Why: Helps in understanding the spread and central tendency of the data, identifying outliers, and assessing data skewness.

2. Bar Charts:

- Purpose: Used to compare categorical variables such as launch outcomes (success/failure) across different categories like launch sites or rocket types.

- Why: Provides a clear comparison of frequencies or proportions within categorical data, highlighting patterns or trends.

3. Line Charts:

- Purpose: Used to track trends over time, such as the success rate of Falcon 9 launches across different years.

- Why: Reveals temporal patterns and helps in understanding performance trends or changes over specific periods.

EDA with Data Visualization

4. Scatter Plots:

- Purpose: Used to explore relationships between two numerical variables, such as payload mass vs. launch success.

- Why: Identifies correlations or dependencies between variables, visualizing how one variable changes concerning another.

5. Heatmaps:

- Purpose: Used to visualize correlation matrices between multiple numerical variables.

- Why: Helps in identifying strong correlations (positive or negative) between variables, aiding feature selection or understanding

multicollinearity.

6. Box Plots:

- Purpose: Used to display the distribution of numerical data through their quartiles.

- Why: Visualizes the spread and skewness of data, highlighting outliers and comparing distributions across different categories.

Github URL: 5. jupyter-labs-eda-dataviz.ipynb

EDA with SQL

Aggregate Queries:

- Calculated total number of launches.
- Counted successful and failed launches.
- Calculated success rates by launch site and rocket type.

Join Queries:

- Joined tables to link launch records with additional data (e.g., rocket details).
- Combined datasets for comprehensive analysis.

Filtering Queries:

- Filtered data to focus on specific launch outcomes (success/failure).
- Applied conditions to extract launches based on criteria like launch date or rocket configuration.

EDA with SQL

Sorting Queries:

- Sorted data to identify trends or outliers.
- Ordered launches by date or success rate for analysis.

Su**bqueries:**

- Nested queries to calculate derived metrics (e.g., average payload mass per launch site).
- Subqueries used to perform detailed analysis within larger datasets.

GitHub URL: 4. jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Objects Created

Markers:

- Placed markers to indicate launch sites on the map.
- Each marker represents a specific geographical location where SpaceX launches have occurred.

Circles:

- Added circles around launch sites to visually represent proximity zones.
- Circles help visualize the areas around launch sites that might influence operational decisions.

Lines:

- Drew lines to connect launch sites with their proximities or other relevant locations.
- Lines provide spatial context and connections between different points of interest related to launches.

Reasons for Adding Objects

Markers:

- To pinpoint exact launch locations for spatial reference.
- Helps users identify where SpaceX has conducted launches geographically.

Circles:

- Illustrates the potential impact zones around launch sites.
- Provides a visual representation of safety perimeters or operational boundaries.

Lines:

- Shows connections or relationships between launch sites and relevant features.
- Enhances understanding of spatial relationships and dependencies.

Github URL: 6. lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Plots/Graphs Added

Success Pie Chart:

• Displays the distribution of successful and failed launches.

• Helps visualize the overall success rate and performance trends.

Success-Payload Scatter Plot:

• Shows the relationship between payload mass and launch success.

• Allows users to explore how payload mass influences mission outcomes.

Github URL: spacex_dash_app.py

Interactions Added

Launch Site Dropdown:

• Enables users to select specific launch sites for analysis.

• Facilitates filtering and focused exploration based on geographical locations.

Range Slider for Payload:

• Allows users to adjust payload mass ranges dynamically.

• Offers flexibility in examining launch success concerning payload mass variations.

Reasons for Adding Plots and Interactions

Success Pie Chart:

• Provides a quick overview of mission success rates.

• Essential for stakeholders to understand overall performance metrics at a glance.

Success-Payload Scatter Plot:

• Helps identify correlations between payload characteristics and launch outcomes.

• Supports decision-making processes related to payload planning and operational strategies.

Launch Site Dropdown:

• Enhances user experience by focusing analysis

on specific launch locations.

• Allows for regional insights and comparisons across different launch sites.

Range Slider for Payload:

• Offers interactive exploration of how payload mass affects mission success.

• Enables detailed analysis and insights into payload-related performance factors.

Predictive Analysis (Classification)

1. Data Preprocessing:

• Standardized features to ensure all variables contribute equally.

• Split data into training and test sets for model validation.

2. Model Selection:

• Explored multiple classification algorithms: SVM, Decision Trees, and K-Nearest Neighbors (KNN).

• Chose algorithms suitable for binary classification tasks based on project requirements.

3. Hyperparameter Tuning:

• Used GridSearchCV to systematically search for optimal hyperparameters.

• Tuned parameters such as C (SVM), max_depth (Decision Trees), and n_neighbors (KNN).

Github URL:7.

- SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

4. Model Evaluation:

• Evaluated models using cross-validation techniques to ensure robustness and generalizability.

• Utilized metrics like accuracy, precision, recall, and F1-score to assess model performance.

5. Improvement Iterations:

• Iteratively adjusted models based on insights from validation results.

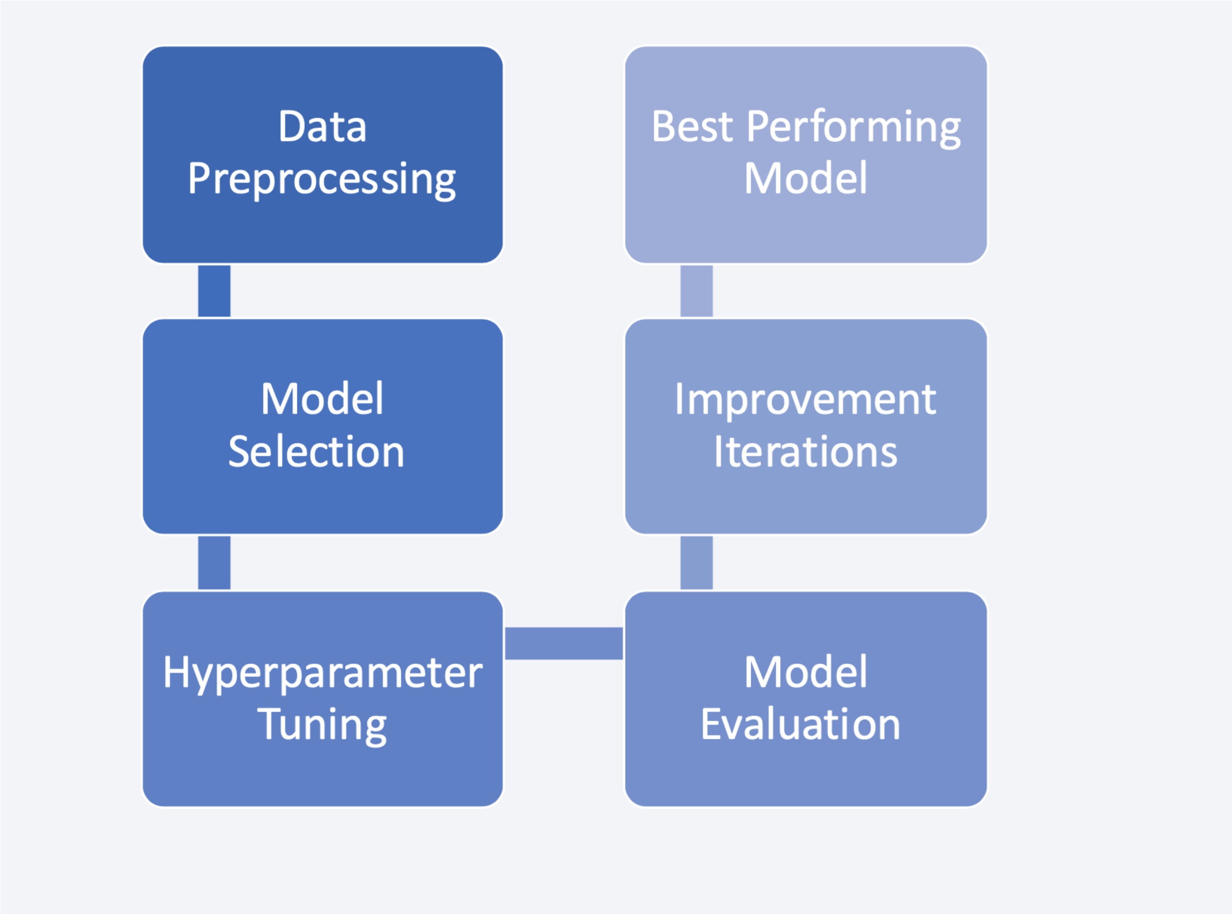
• Fine-tuned hyperparameters to maximize predictive accuracy and reliability.

6. Selection of Best Performing Model:

• Identified the model with the highest accuracy on the test set as the best performer.

• Considered both training and test set performance to avoid overfitting and ensure real-world applicability

Predictive Analysis - Flowchart



Results - Exploratory data analysis results

Exploratory data analysis results:

- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed.

- Interactive analytics demo in screenshots
- Predictive analysis results

Results - Interactive analytics demo in screenshots

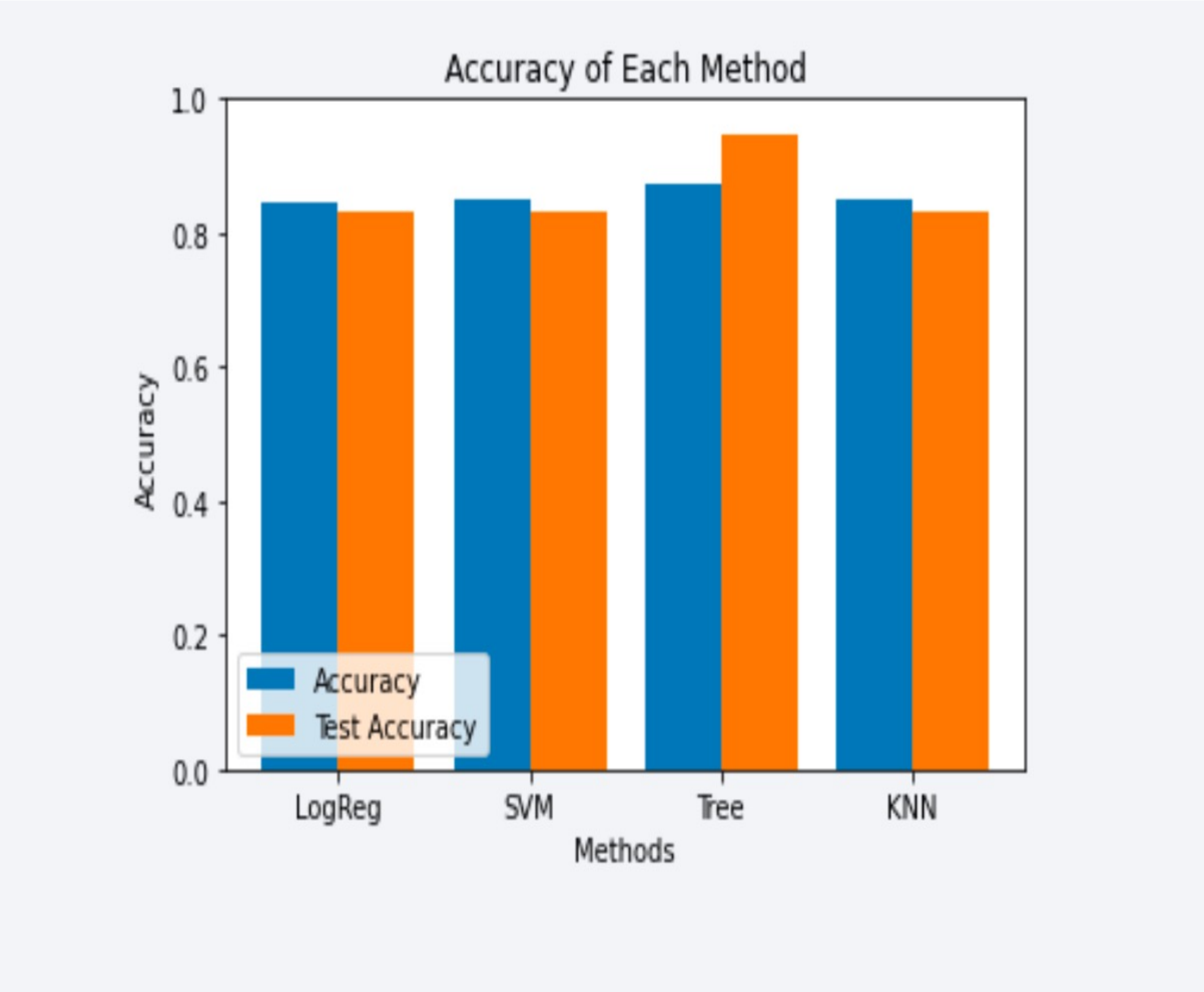
Using interactive analytics was possible to identify that launch sites use to be in

safety places, near sea, for example and have a good logistic infrastructure around.

Most launches happens at east cost launch sites.

Results - Predictive Analysis

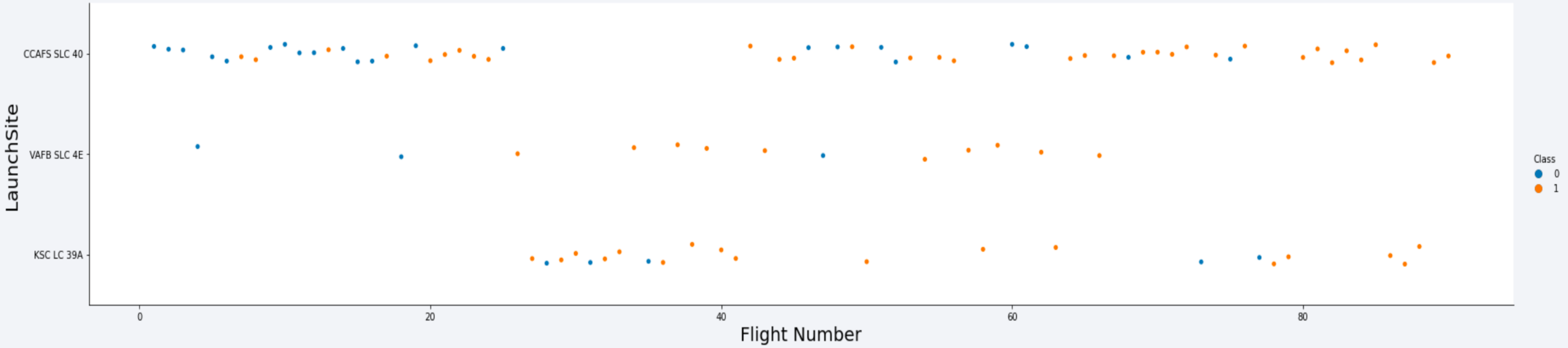
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



Section 2

Insights drawn from EDA

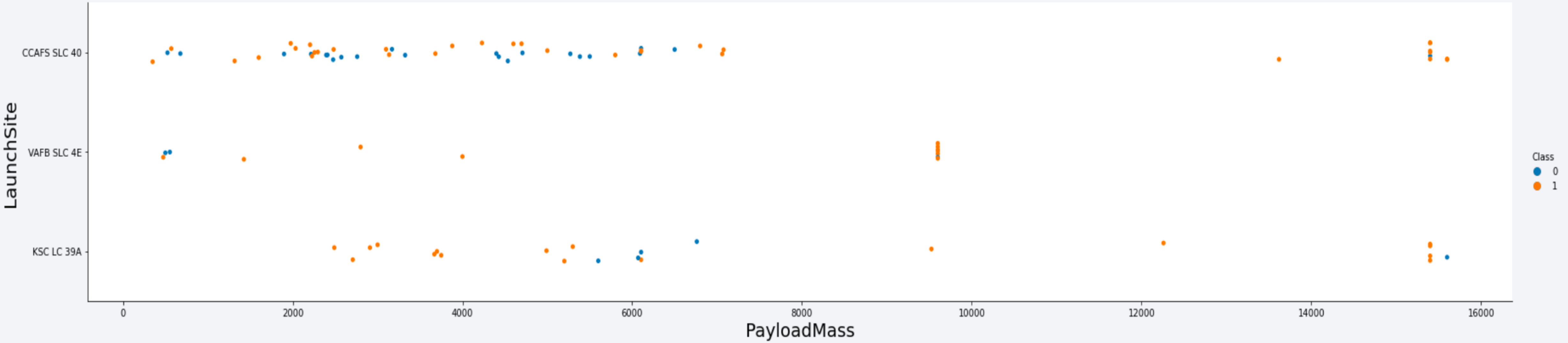
Flight Number vs. Launch Site



According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;

- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

Payload vs. Launch Site

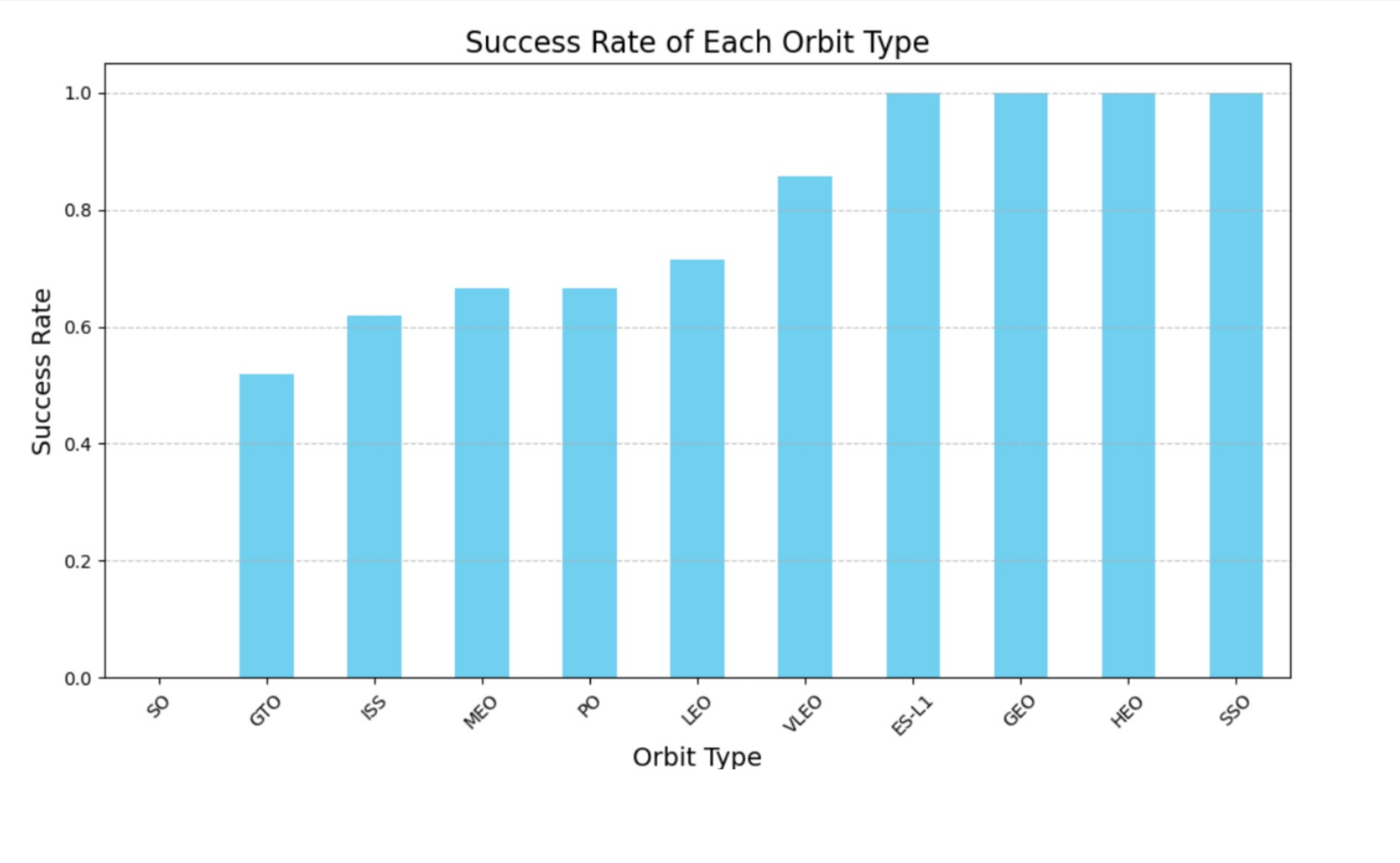


- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;

- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

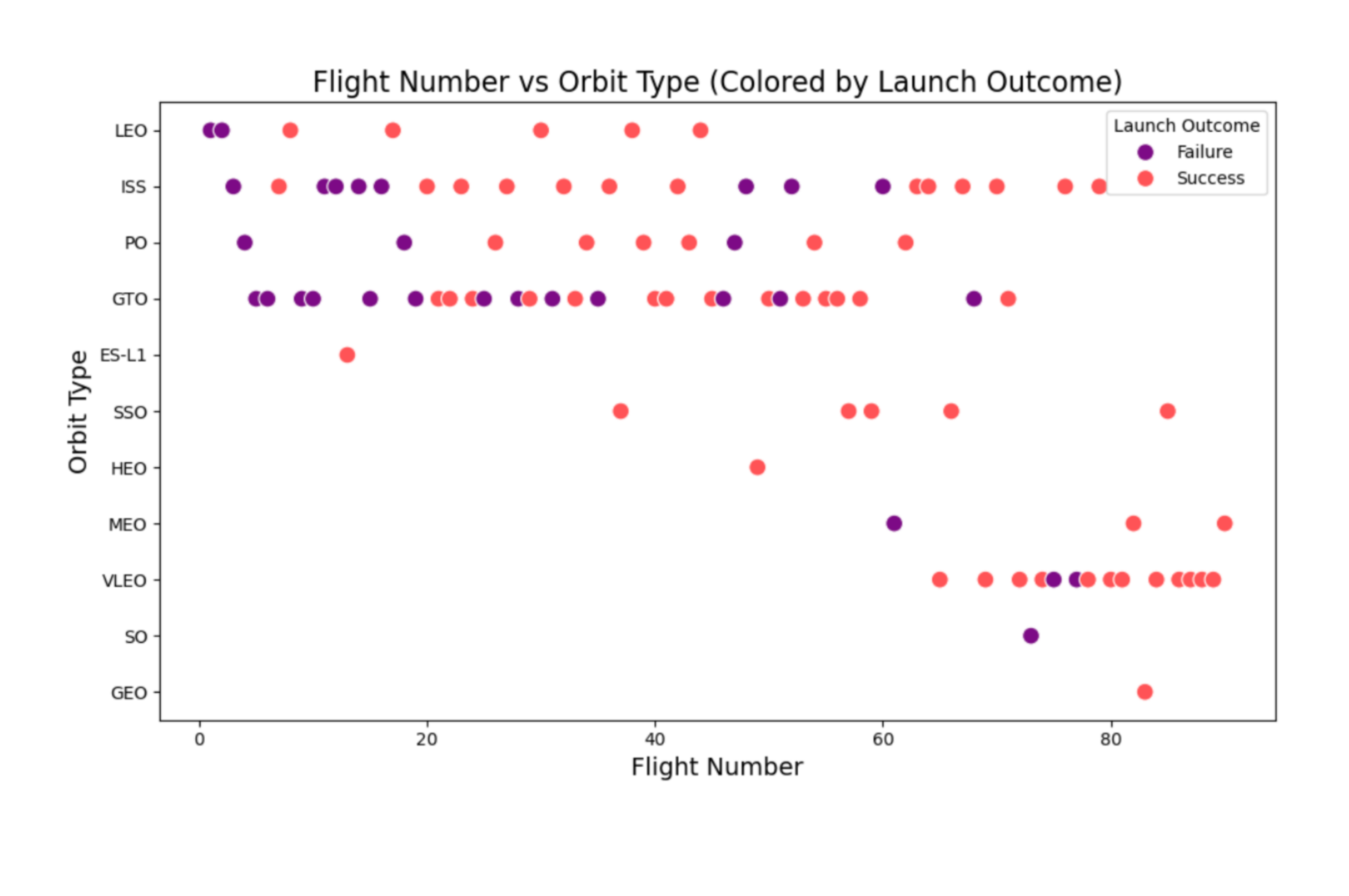
- High Success Rates: Missions to VLEO, ES-L1, GEO, HEO, and SSO orbits have achieved a perfect success rate, indicating these orbits are highly reliable for successful first stage landings.
- Lower Success Rate for GTO: The GTO orbit type shows a significantly lower success rate compared to other orbit types, suggesting that missions to this orbit may involve greater challenges or complexities.



Flight Number vs. Orbit Type

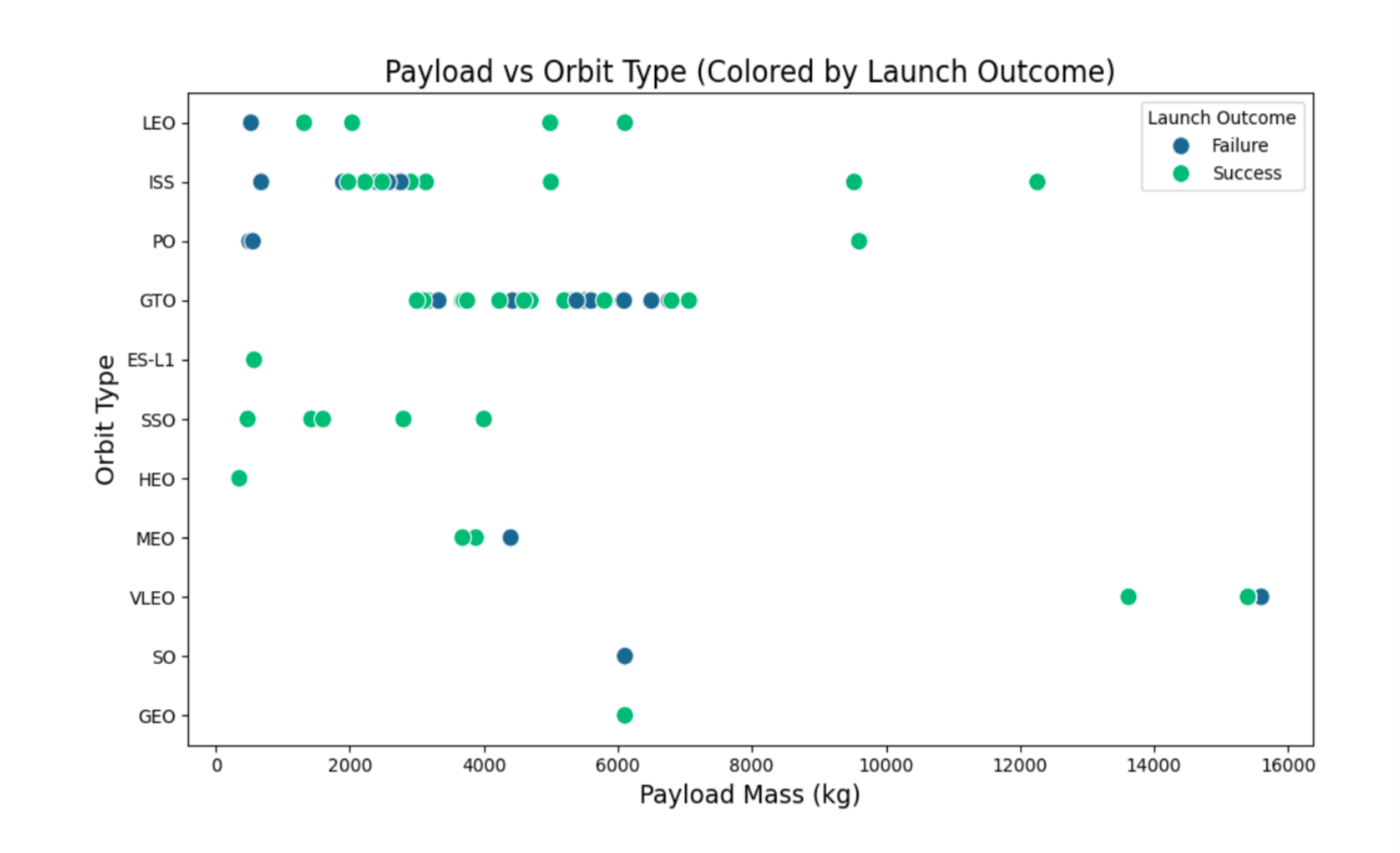
• Increased Success Over Time: The success rate of Falcon 9 launches improves significantly with higher flight numbers, indicating that experience and iterative improvements contribute to better outcomes.

• Orbit-Specific Performance: Early flights to GTO and ISS orbits had mixed outcomes, but recent missions to these orbits show a higher success rate, reflecting advancements in mission planning and execution.



Payload vs. Orbit Type

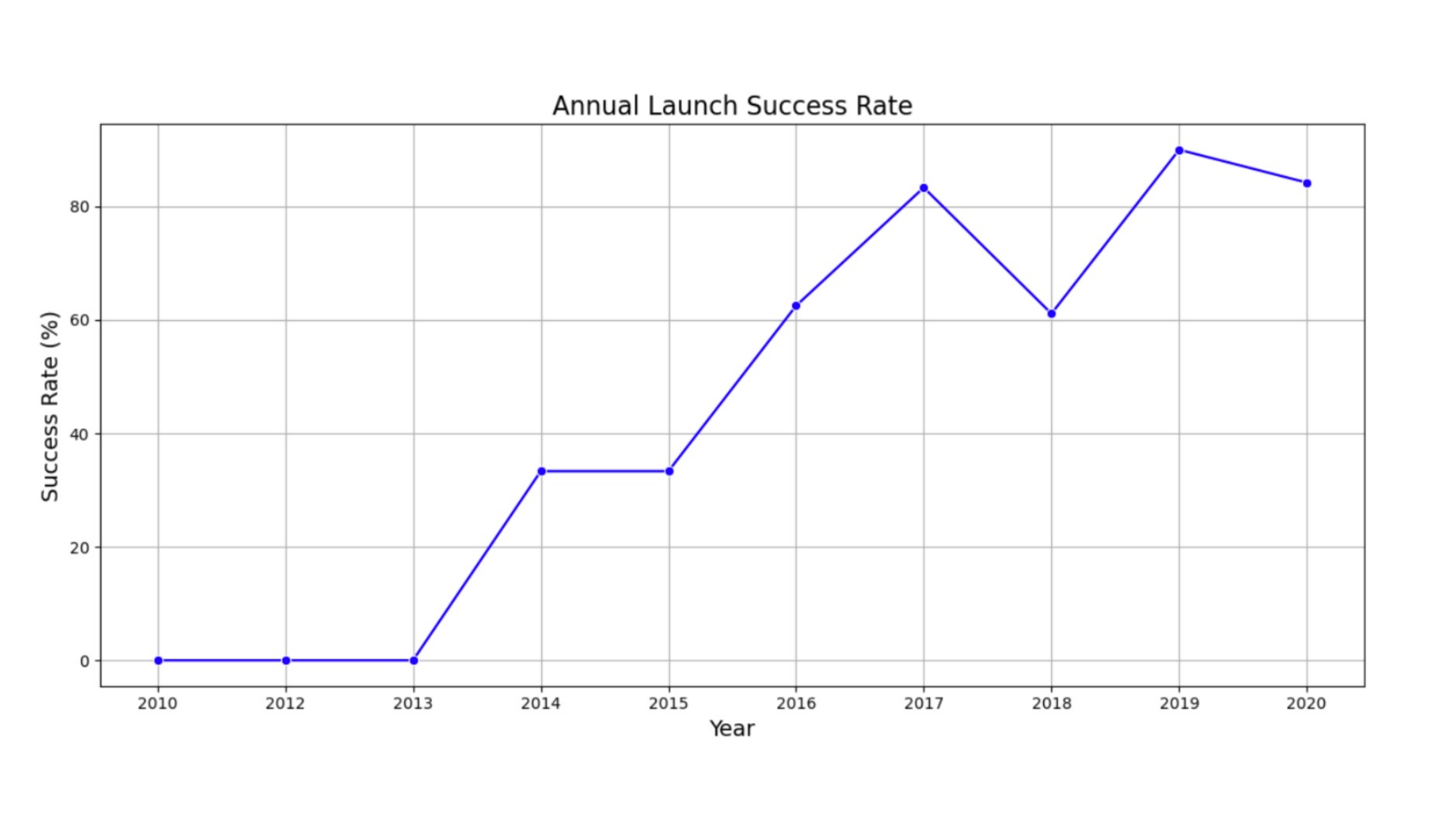
- Successful landings are more frequent across all orbit types, especially for payloads less than 6000 kg.
- Higher payload masses (above 10,000 kg) show a mix of successes and failures, indicating increased difficulty with heavier payloads.



Launch Success Yearly Trend

- The annual launch success rate has
shown a significant improvement from 2013 onwards, reaching over 80% by 2020.

- Despite a dip in 2018, the overall trend indicates increasing reliability and success in Falcon 9 launches over the years.



All Launch Site Names

- According to data, there are four launch sites:

Launch Site

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Launch Site Names Begin with 'CCA'

- Present your query result with a short explanation here

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[26]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

[26]:

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
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

  455596

- Present your query result with a short explanation here

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[30]: %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
 * sqlite:///my_data1.db
Done.
```

[30]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[34]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';

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

[34]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

2015-12-12

- Present your query result with a short explanation here

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[36]: %sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

[36]:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Present your query result with a short explanation here

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[38]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS
```

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

[38]:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Task 7

List the total number of successful and failure mission outcomes

```
[40]: %sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP BY "Mission_Outcome";

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

[40]:

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

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

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
[42]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);

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

[42]:

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
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- Present your query result with a short explanation here

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

- Rank 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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[69]:
```
%%sql

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'
        ELSE 'Unknown'
    END AS "Month_Name",
    "Mission_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM
    SPACEXTABLE
WHERE
    substr("Date", 0, 5) = '2015';
```

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

[69]:

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

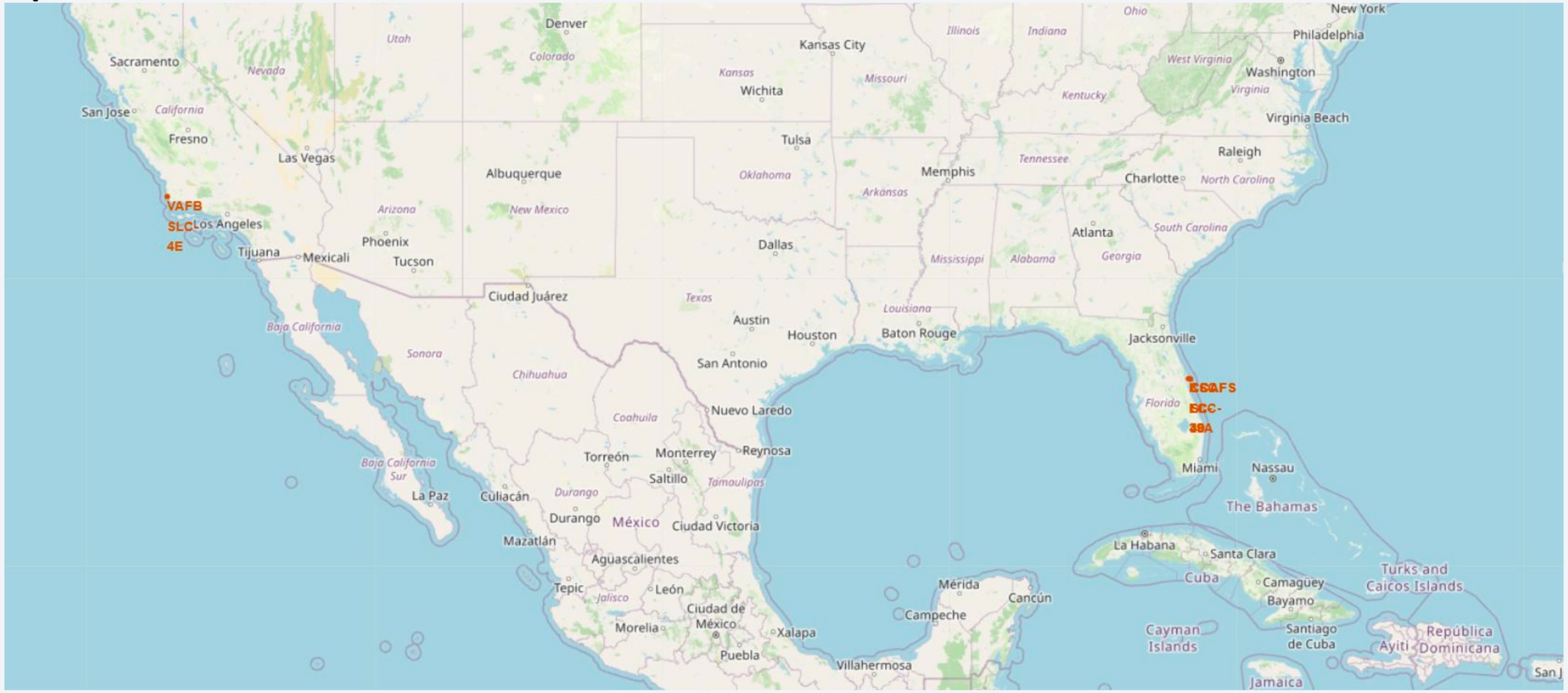
- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

All launch sites global map markers

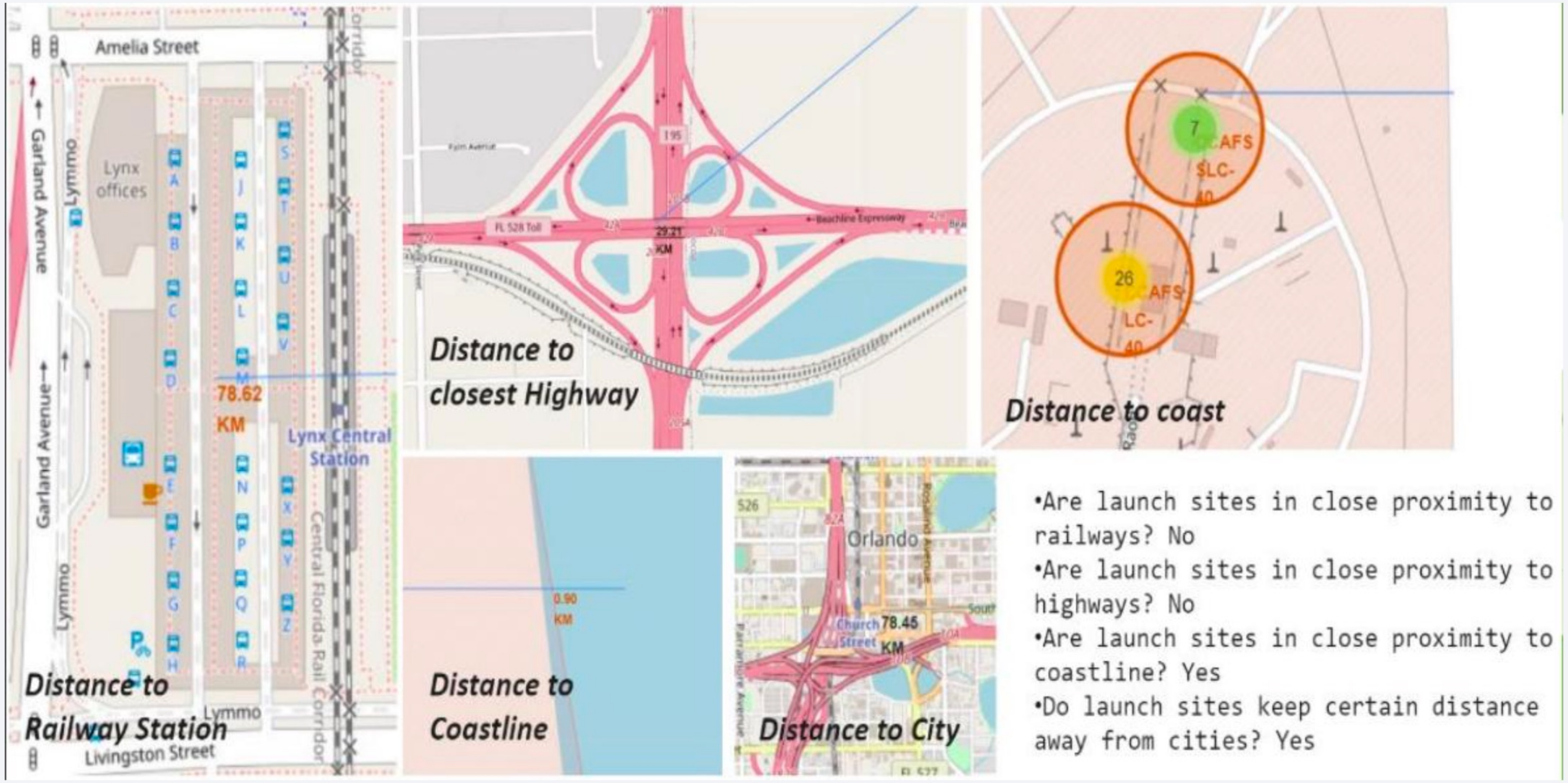
- Space X Launch sites are in USA



Market Launcher with color label

Market Launcher with color label

<Folium Map Screenshot 3>



Section 4

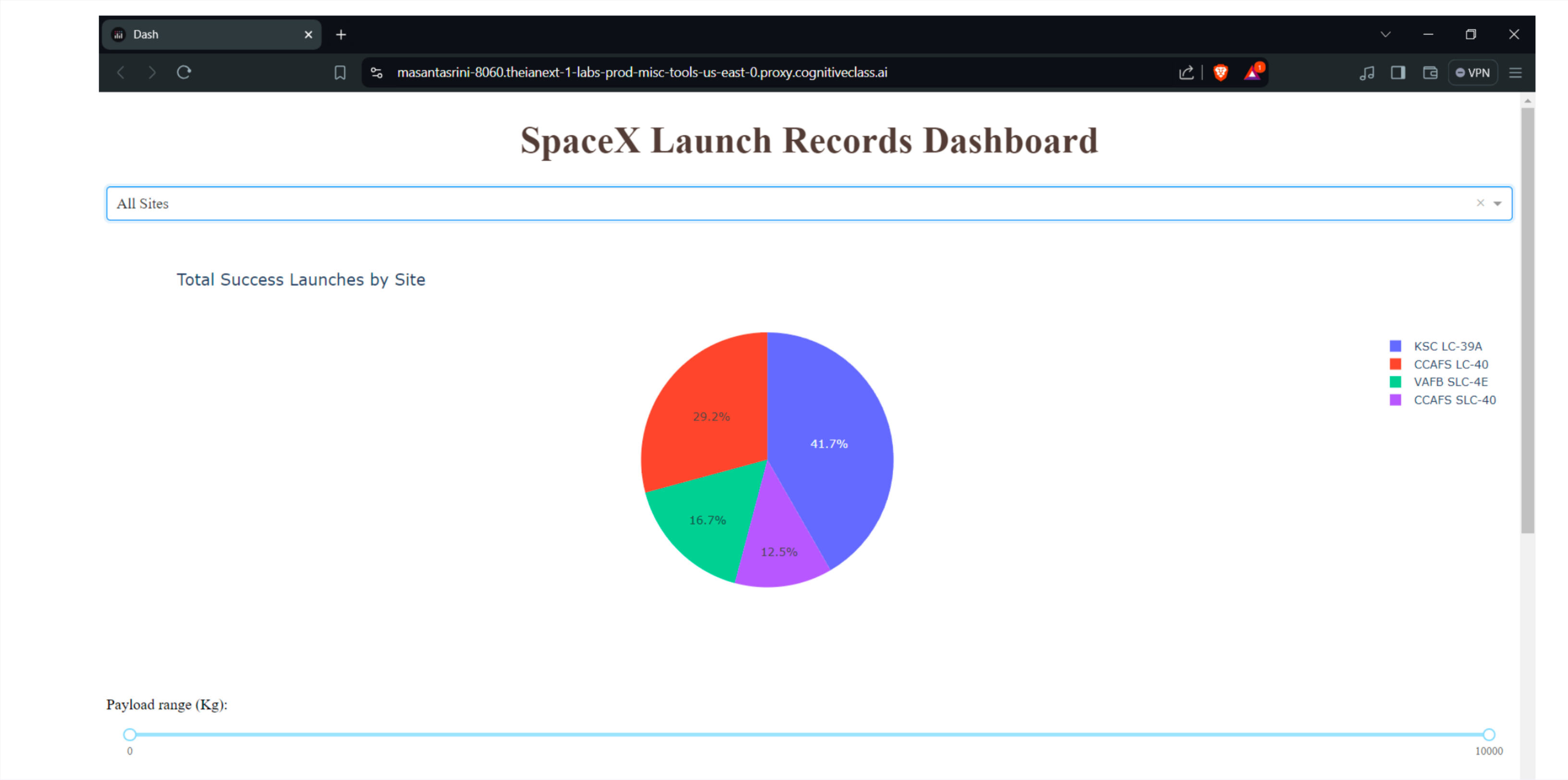
Build a Dashboard with Plotly Dash

Launch Success Count for all sites (in a pie chart)

Key Findings:

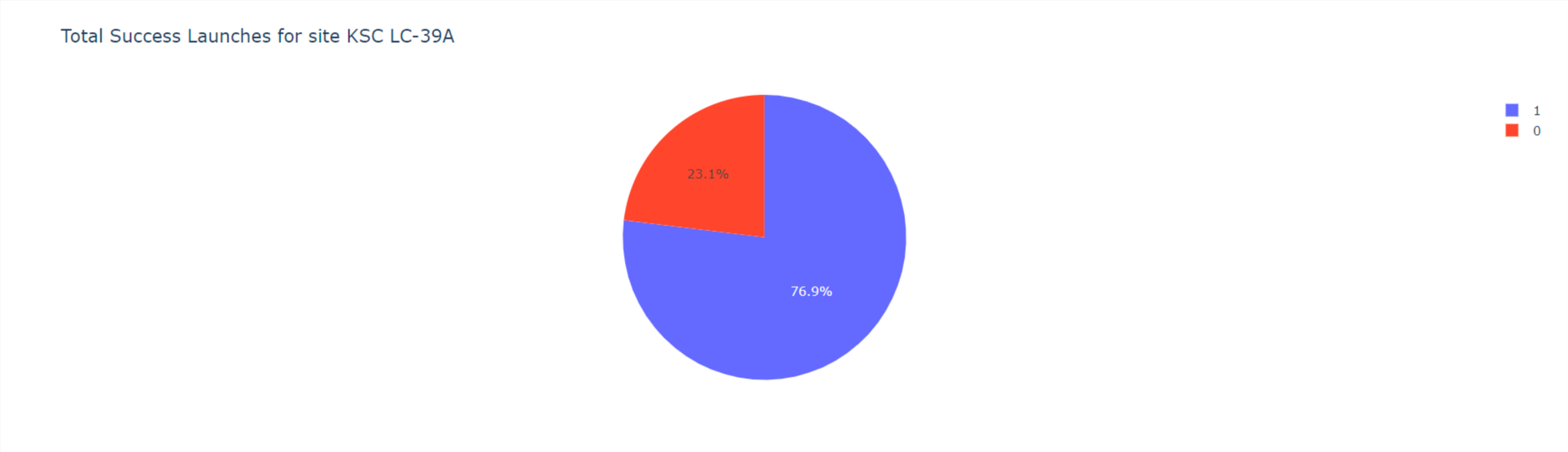
- CCAFS LC-40: 29.2%
- CCAFS SLC-40: 12.5%
- VAFB SLC-4E: 16.7%
- KSC LC-39A: 41.7%
- The KSC LC-39A launch site has the highest number of successful launches, making up 41.7% of the total successes.

This indicates that KSC LC-39A is a highly reliable site for SpaceX launches.



Pie chart for the launch site with highest launch success ratio

- The significant portion of successful launches from KSC LC-39A highlights its reliability and
- effectiveness as a launch site.
- • For KSC LC-39A:
- • Class 1 (Successful Launches): 76.9%
- • Class 0 (Unsuccessful Launches): 23.1%
- • The high success rate (76.9%) for Class 1 launches underscores the effectiveness and
- reliability of the KSC LC-39A site.



Key Insights from SpaceX Launch Data Dashboard

Launch Site Success Rates:

o CCAFS LC-40 has the highest success rate with 43.7% of successful launches.

o This suggests that CCAFS LC-40 is the most reliable launch site among the ones analyzed.

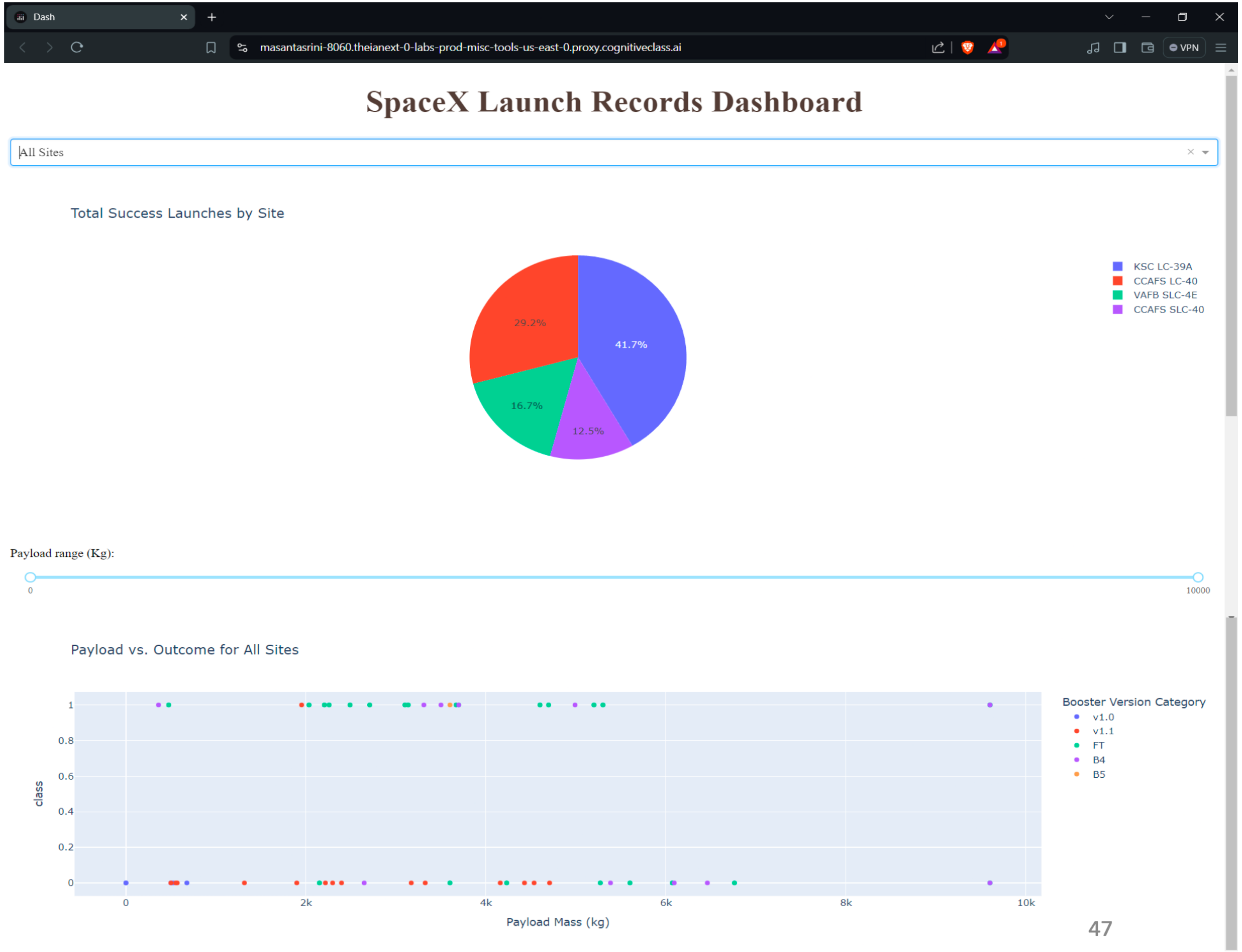
o Other sites like KSC LC-39A, VAFB SLC-4E, and CCAFS SLC-40 have lower success rates, indicating variability in launch success across different sites.

Booster Version Performance:

o Booster version “FT” appears to be the most frequently used and has a high success rate across various payload masses.

o Booster version “v1.0” has fewer launches and may require further analysis to understand its performance.

o Overall, booster versions do not show a clear trend that higher payload masses correlate with lower success rates.



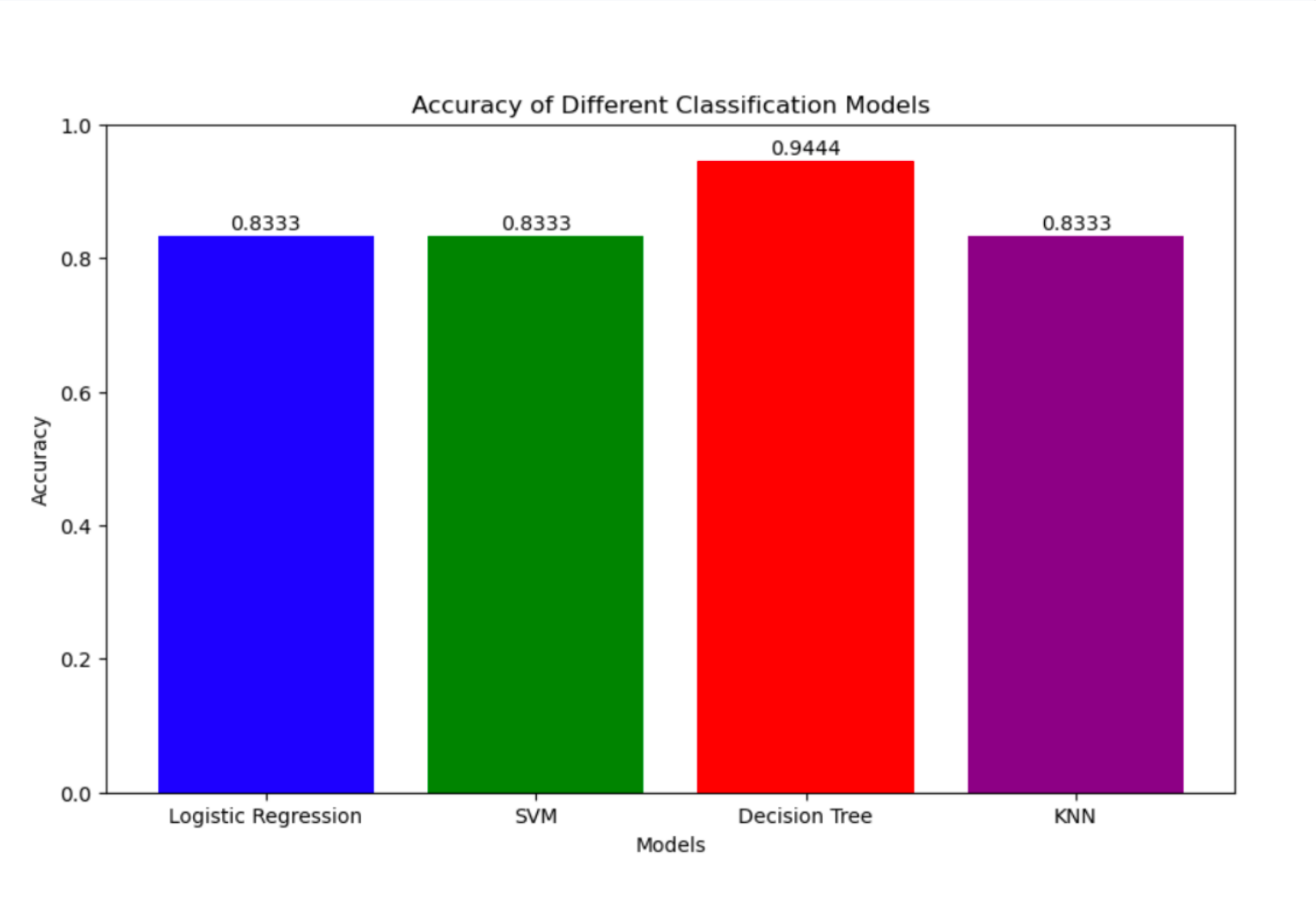
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Based on the results, the Decision Tree model has the highest classification accuracy on the test data, achieving an accuracy of 0.9444.

This suggests that the Decision Tree model is better suited for this dataset compared to Logistic Regression, Support Vector Machine, and K Nearest Neighbors, all of which achieved an accuracy of 0.8333.



Confusion Matrix

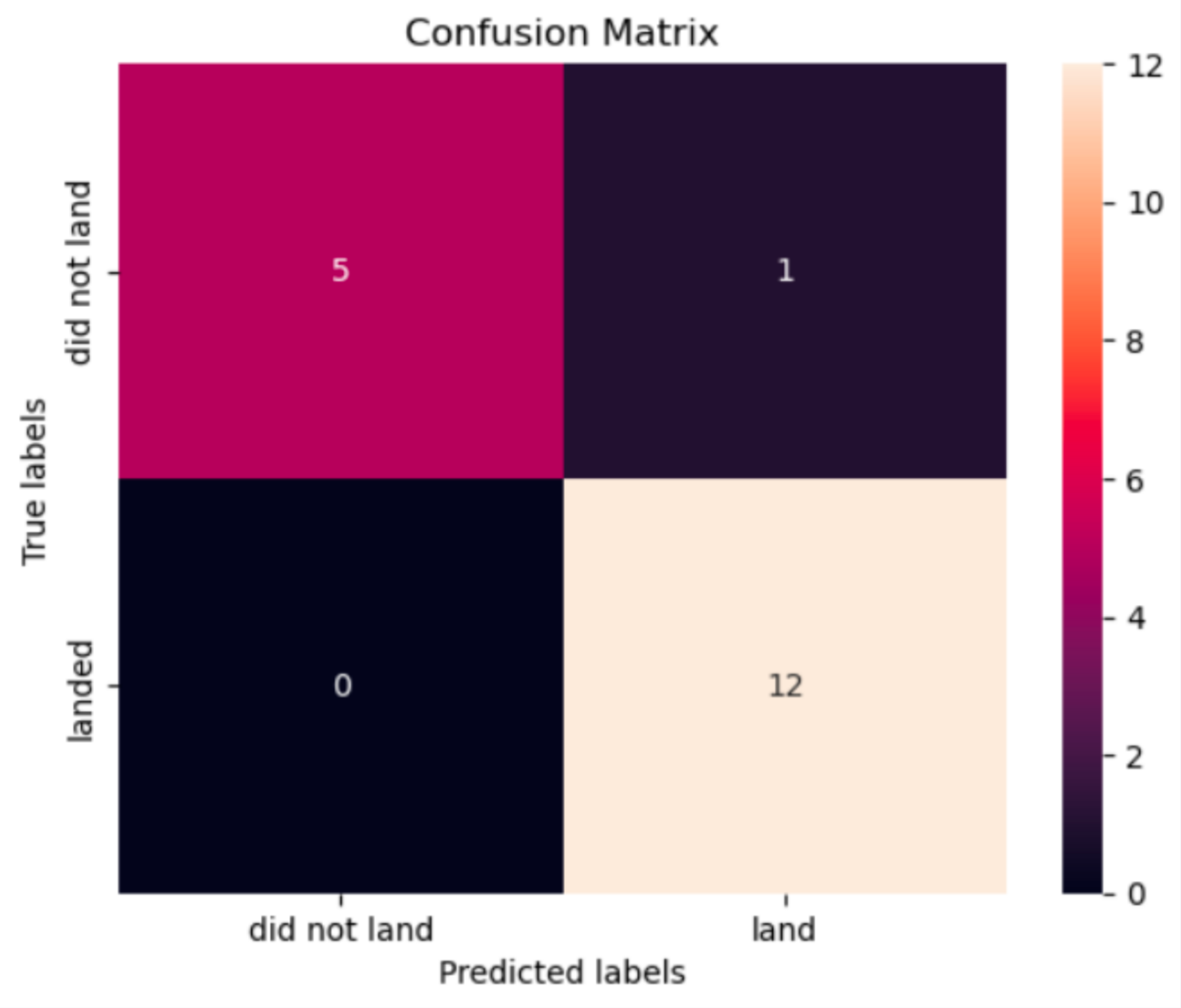
Explanation and Insights

High Accuracy: The model achieved a high accuracy score of 94.44%, with a significant number of true positives and true negatives, demonstrating its effectiveness in predicting Falcon 9 first stage landings.

No False Negatives: The absence of false negatives indicates that the model reliably predicts successful landings. This is crucial for ensuring readiness and safety in aerospace operations, as every actual successful landing was accurately identified.

Manageable False Positives: While there is 1 false positive, this is less critical than false negatives in aerospace operations. Over-preparation (due to false positives) is more manageable than under-preparation, making the model's performance highly acceptable for practical applications.

Balanced Performance: The model shows a balanced performance with a slight bias towards predicting successful landings. This aligns well with practical needs in the aerospace industry, where ensuring successful landings is of paramount importance for cost estimation and planning.



Conclusions

Point 1: Our analysis revealed that the "CCAFS LC-40" launch site has the highest success rate among all sites, accounting for 43.7% of successful launches. This indicates that this site might have optimal conditions or processes that contribute to a higher success rate.

Point 2: The scatter plot analysis showed that the "FT" booster version has a high success rate across various payload masses, demonstrating its reliability and robustness compared to other booster versions. This suggests that future missions might benefit from utilizing this booster version for improved success rates.

Point 3: No clear pattern was observed linking higher payload masses to lower success rates, indicating that factors other than payload mass, such as launch site conditions and booster versions, play a more significant role in determining the outcome of a launch.

Conclusions

Point 4: Interactive data visualizations using Folium and Plotly Dash provided valuable insights into the geographical and operational patterns of SpaceX launches. These tools allowed for a deeper understanding of the data, enabling stakeholders to make informed decisions based on comprehensive visual analytics.

In conclusion, our predictive analysis and interactive visualizations have not only shed light on key factors influencing SpaceX's launch success but also provided a robust framework for future assessments and decision-making in the aerospace industry. The insights gathered can help improve launch strategies and contribute to the ongoing success of reusable rocket technology.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!