IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rahul

13th March, 2025

Outline

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

Executive Summary

- Overview of the project goals and key findings.
- Summary of methodologies used.
- Main results and insights gained.
- Final conclusions and recommendations.

Introduction

- Project Background: The commercial space age is here, with companies making space travel more affordable. SpaceX has significantly reduced costs by reusing the Falcon 9 first stage. If we can determine whether the first stage will land successfully, we can estimate the cost of a launch, which is crucial for competitors like "Space Y.“
- Problem Statement: Predicting the successful landing of the Falcon 9 first stage to estimate launch costs.

Section 1

Methodology

Methodology

Executive Summary

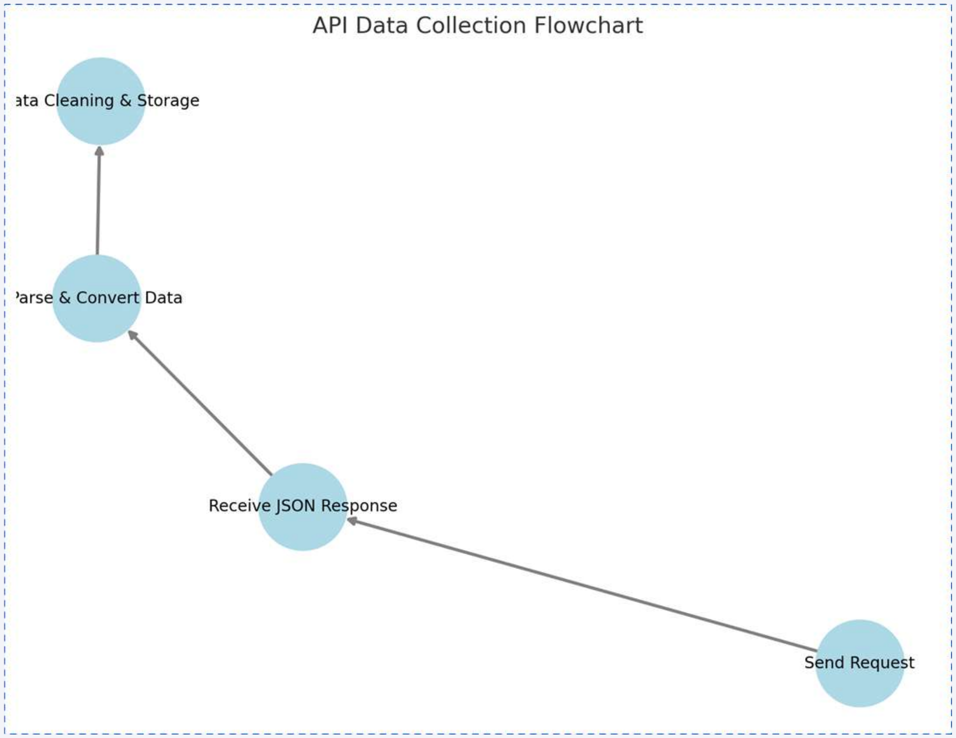
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data was gathered from SpaceX API and web scraping techniques.
- API calls were used to obtain Falcon 9 launch data, including payload, orbit type, and landing outcomes.
- Web scraping was performed to collect additional data on launch sites and success rates.
- API Data Collection Flowchart
    1. Send Request → Query SpaceX API for launch data.
    2. Receive JSON Response → Extract launch details (payload, launch site, landing outcome).
    3. Parse & Convert Data → Transform JSON into a structured Pandas DataFrame.
    4. Data Cleaning & Storage → Handle missing values, filter necessary fields, and store in a local database.

Data Collection – SpaceX API

- Send Request → Query SpaceX API for launch data.
- Receive JSON Response → Extract launch details (payload, launch site, landing outcome).
- Parse & Convert Data → Transform JSON into a structured Pandas DataFrame.
- Data Cleaning & Storage → Handle missing values, filter necessary fields, and store in a local database.



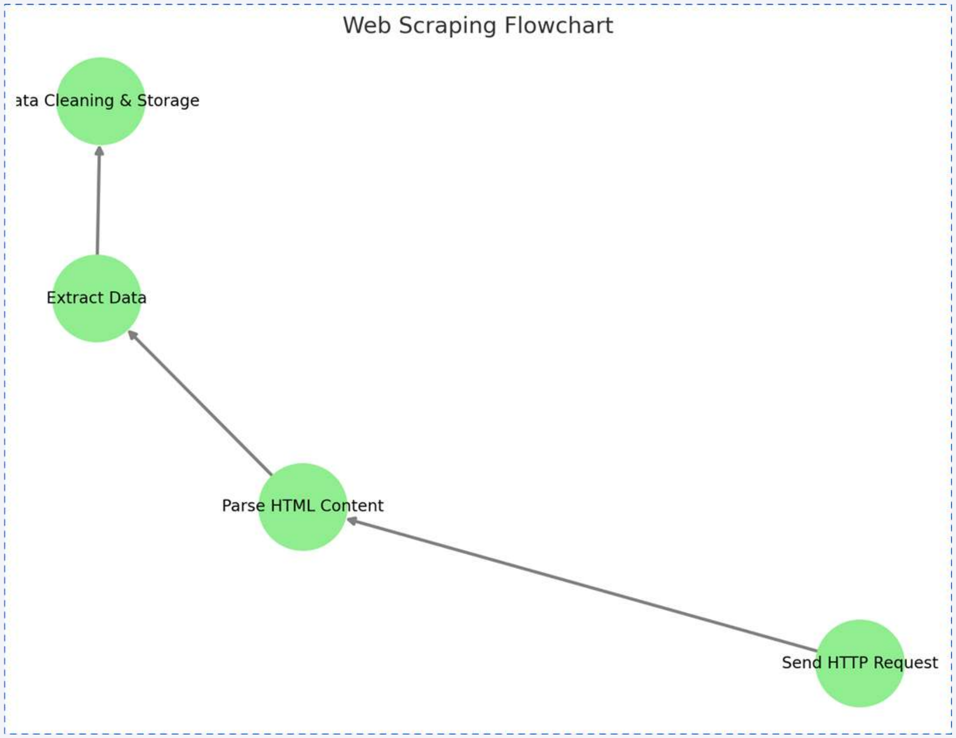
Data Collection - Scraping

Web Scraping Process

- Used BeautifulSoup and Selenium to scrape SpaceX launch data.
- Extracted data such as mission names, launch sites, dates, and outcomes.
- Cleaned and structured the extracted data for further analysis.

Web Scraping Flowchart

1. Send HTTP Request → Access SpaceX website.
2. Parse HTML Content → Extract relevant tables and text.
3. Extract Data → Retrieve mission details, launch dates, and outcomes.
4. Data Cleaning & Storage → Format data, handle missing values, and store it in a DataFrame.



Data Wrangling

- Handled missing values and standardized data formats.
- Performed feature engineering to extract relevant information.
- Encoded categorical variables and normalized numerical values.

Data Wrangling Flowchart

1. Handle Missing Values → Identify and fill/remove missing data.
2. Convert Data Types → Ensure numerical and categorical consistency.
3. Feature Engineering → Create new meaningful features from raw data.
4. Normalization & Encoding → Scale numerical values and encode categorical variables.
5. Store Cleaned Data → Save processed data for modeling and analysis.

EDA with Data Visualization

- Visualized relationships between payload mass, launch sites, and success rates.
- Used SQL queries to extract key insights from structured data.
- Interactive visual analytics with Folium and Plotly Dash to explore launch patterns.
- EDA Plots Used:
  - Scatter Plot: Payload vs. Launch Site.
  - Bar Chart: Success Rate by Orbit Type.
  - Scatter Plot: Flight Number vs. Launch Site.
  - Line Chart: Yearly Average Success Rate.
  - Pie Chart: Proportion of Successful and Failed Landings.

EDA with SQL

SQL Queries Performed

- Find unique launch sites: Identified all distinct launch sites in the dataset.
- Retrieve launches from a specific site: Filtered launches that began with 'CCA'.
- Calculate total payload by NASA boosters: Summed the total payload mass carried by NASA boosters.
- Find average payload for specific booster version: Computed the mean payload mass for Falcon 9 v1.1.
- Find first successful landing date on ground pad: Retrieved the earliest successful landing outcome on a ground pad.
- Filter successful drone ship landings with specific payload range: Found boosters that successfully landed on drone ships and carried payloads between 4000 and 6000 kg.
- Count successful and failed mission outcomes: Aggregated and compared the number of successful vs. failed landings.
- Find boosters carrying maximum payload: Identified booster names that carried the heaviest payloads.
- Retrieve failed drone ship landings in 2015: Filtered records of failed drone ship landings in 2015 along with their booster versions and launch sites.
- Rank landing outcomes within a specific time range: Ordered the frequency of landing outcomes between 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

Map Objects in Folium

- Markers: Added markers to indicate the locations of launch sites.
- Circles: Used circles to represent impact zones of landings with varying sizes based on payload mass.
- Lines: Plotted lines connecting launch sites to landing zones to visualize flight trajectories.
- Color Coding: Applied color differentiation to distinguish successful vs. failed landings.
- Pop-ups: Implemented pop-ups displaying details about each launch upon clicking a marker.

Purpose of These Objects:

- Helped visualize key launch and landing locations.
- Provided an interactive way to explore mission data.
- Allowed for quick comparison of successful and failed landings.
- Enhanced data storytelling and analysis.

Build a Dashboard with Plotly Dash

- Pie Chart (Plotly Dash)Displays the proportion of successful vs. failed landings for all launch sites. Helps users quickly see SpaceX's landing success rate.
- Scatter Plot with Payload Slider (Plotly Dash)Plots Payload vs. Launch Outcome, allowing users to analyze how payload size affects landing success.
- Interactive Slider: Filters payload weight dynamically, helping users observe trends for different weight categories.

Predictive Analysis (Classification)

- Trained classification models: Logistic Regression, Support Vector Machines (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
- Hyperparameter tuning using GridSearchCV.
- Model evaluation using accuracy scores and confusion matrices.

Results

Exploratory Data Analysis Results

- Found key trends in payload capacity and launch site success rates.
- Scatter plots and bar charts revealed insights into Falcon 9 landing success.

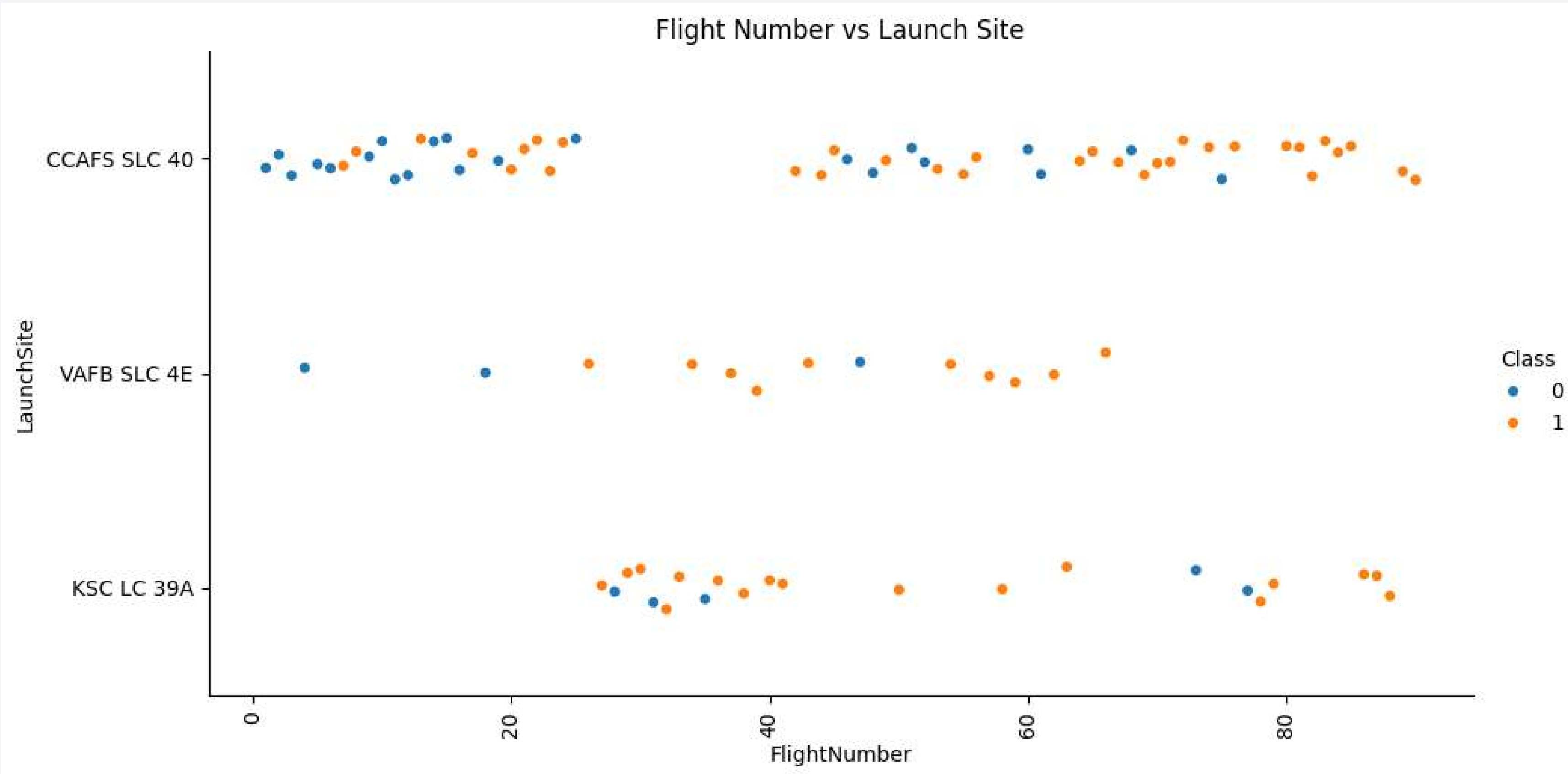
Model Performance Comparison

- Logistic Regression: Accuracy = 83.3%
- SVM: Accuracy = 83.3%
- Decision Tree: Accuracy = 77.7%
- KNN: Accuracy = 83.3%

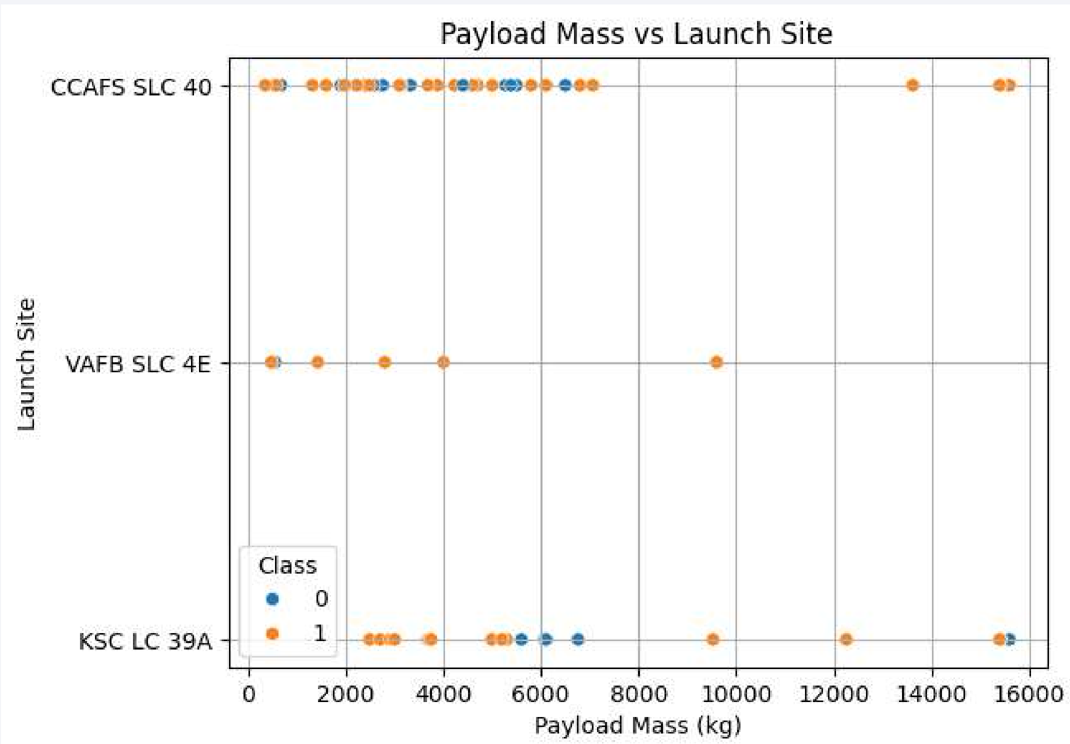
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

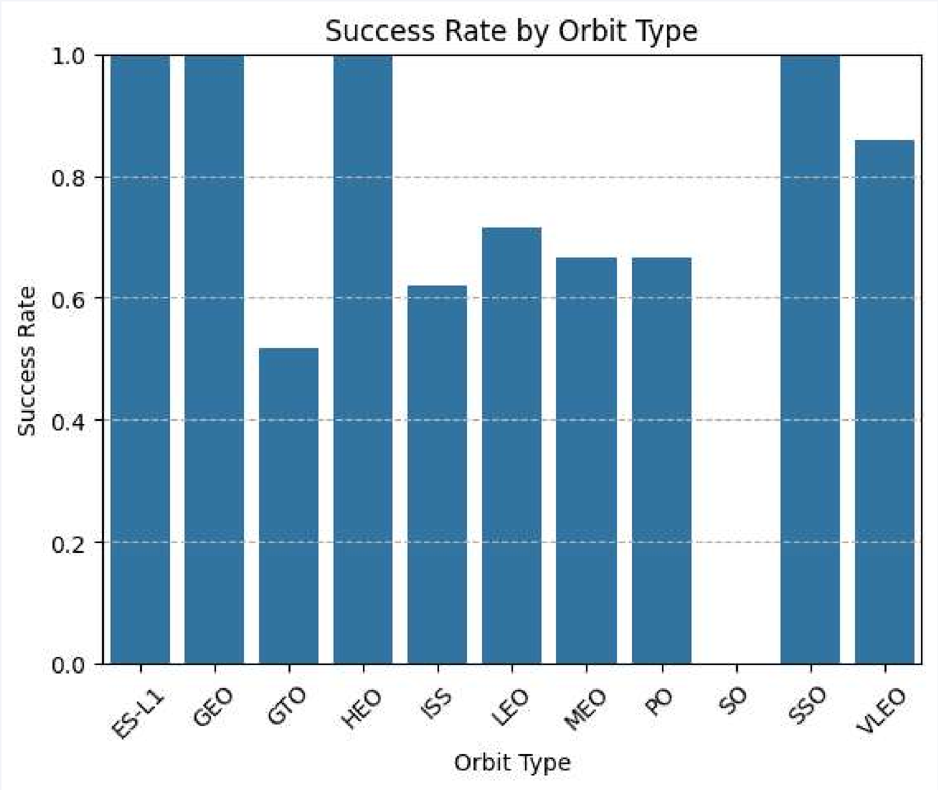


Payload vs. Launch Site

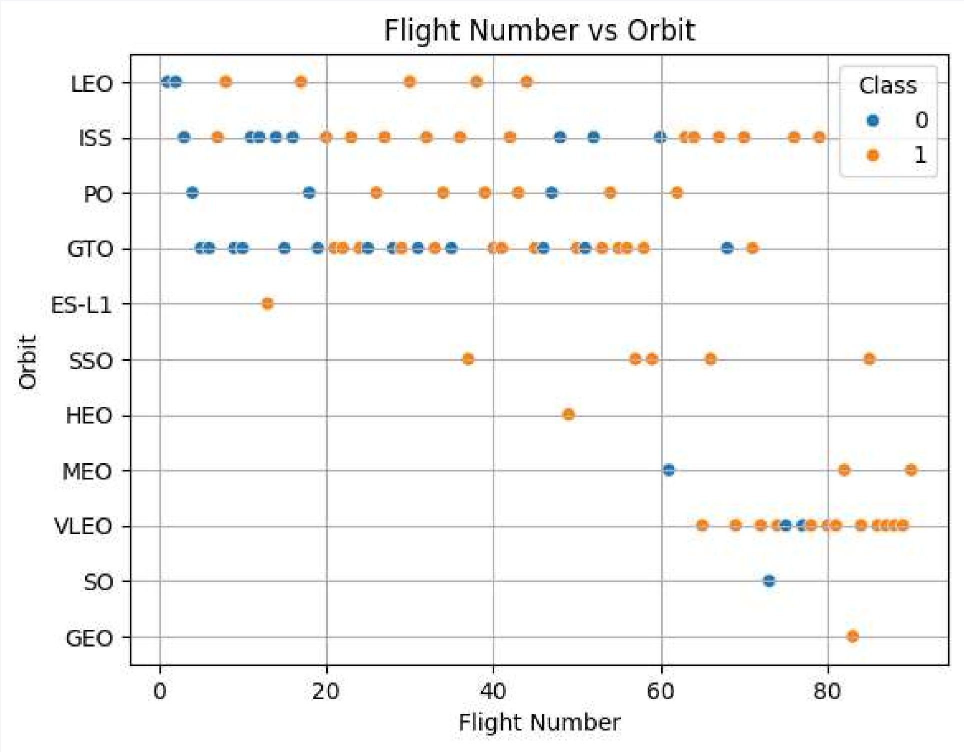


VAFB-SLC launch site, there are no rockets launched for a heavy payload mass(greater than 10000)

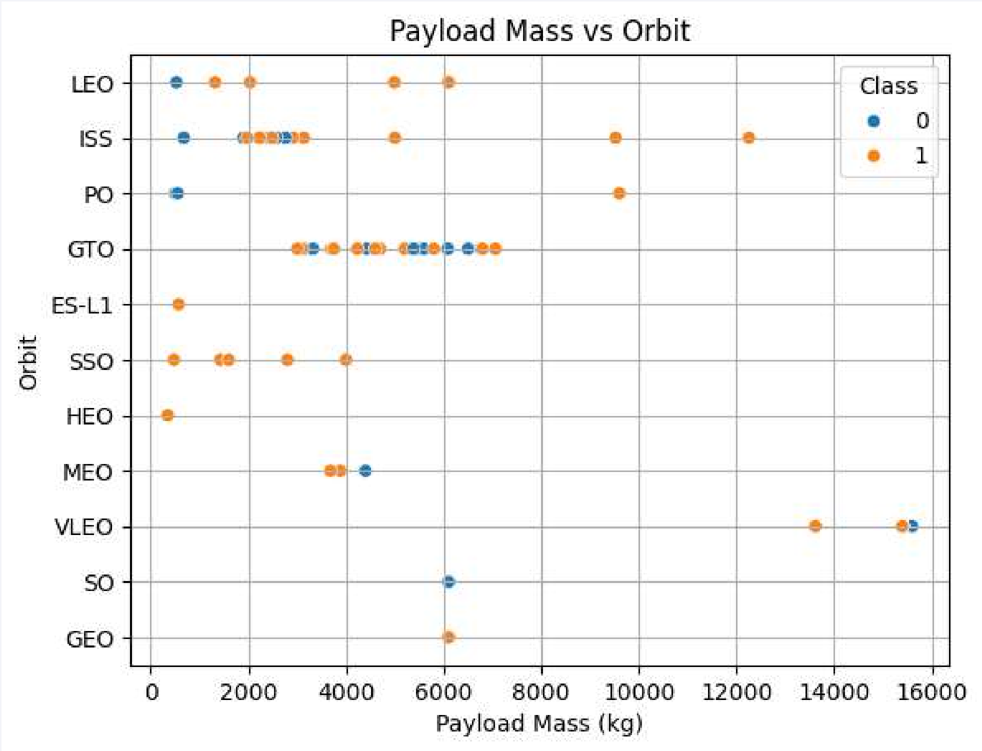
Success Rate vs. Orbit Type



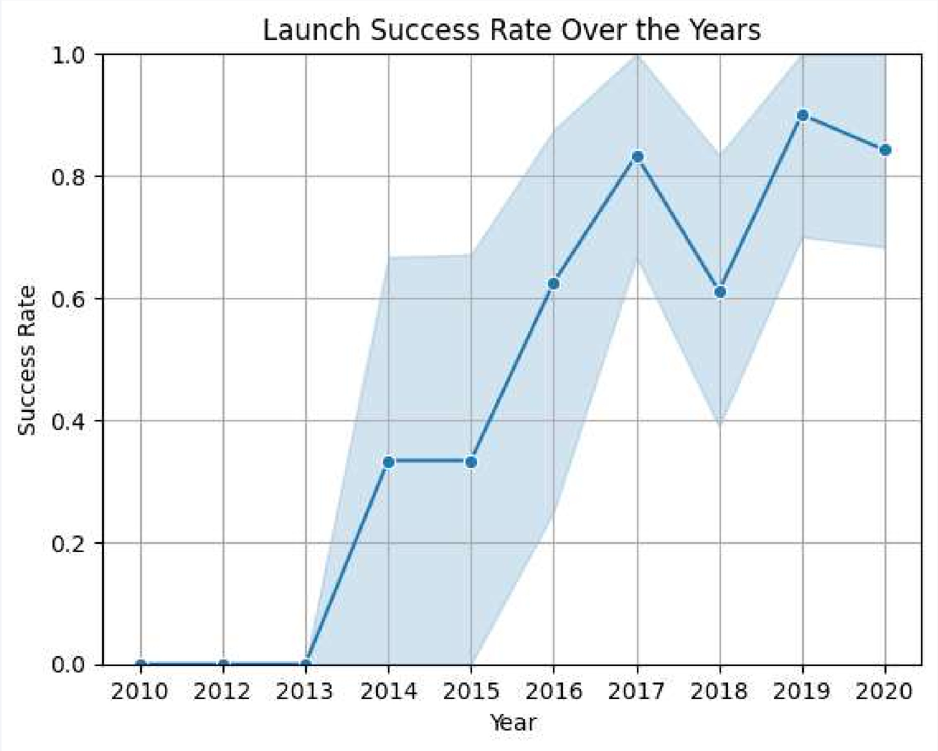
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

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

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

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

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

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

2015 Launch Records

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

- Present your query result with a short explanation here

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

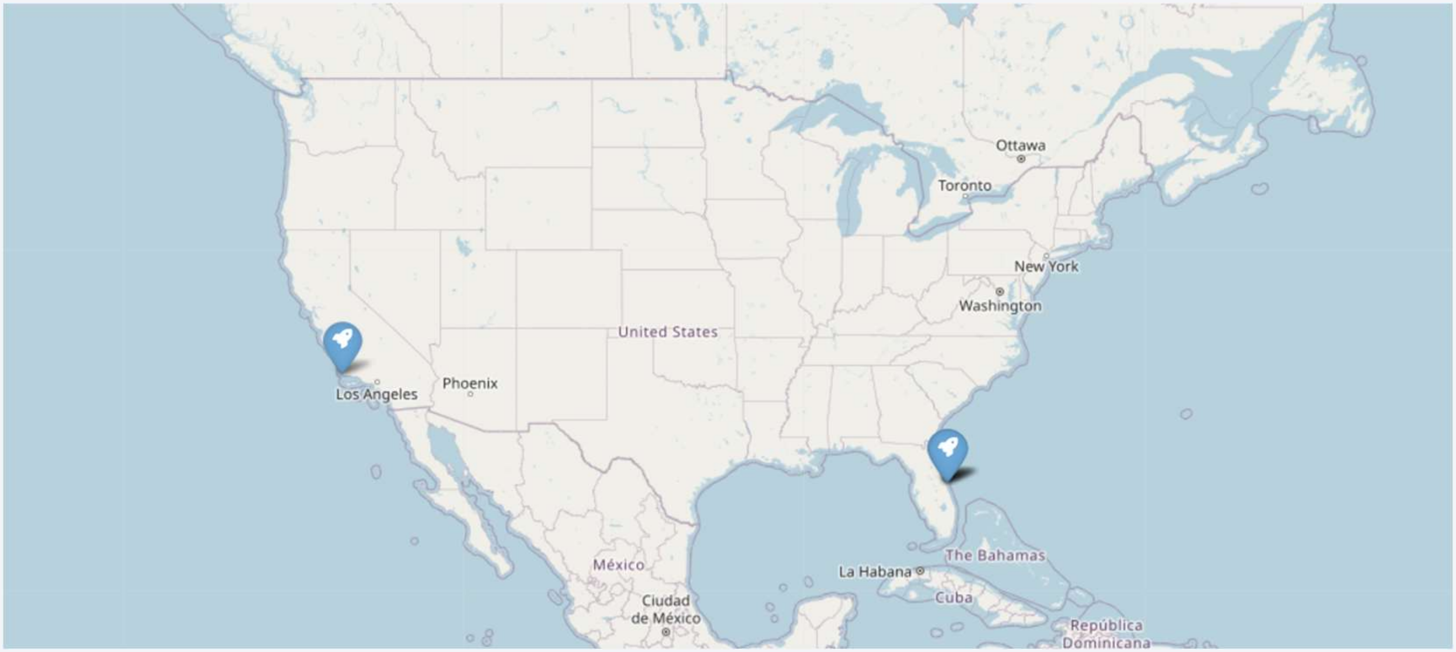
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



<Folium Map Screenshot 2>

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

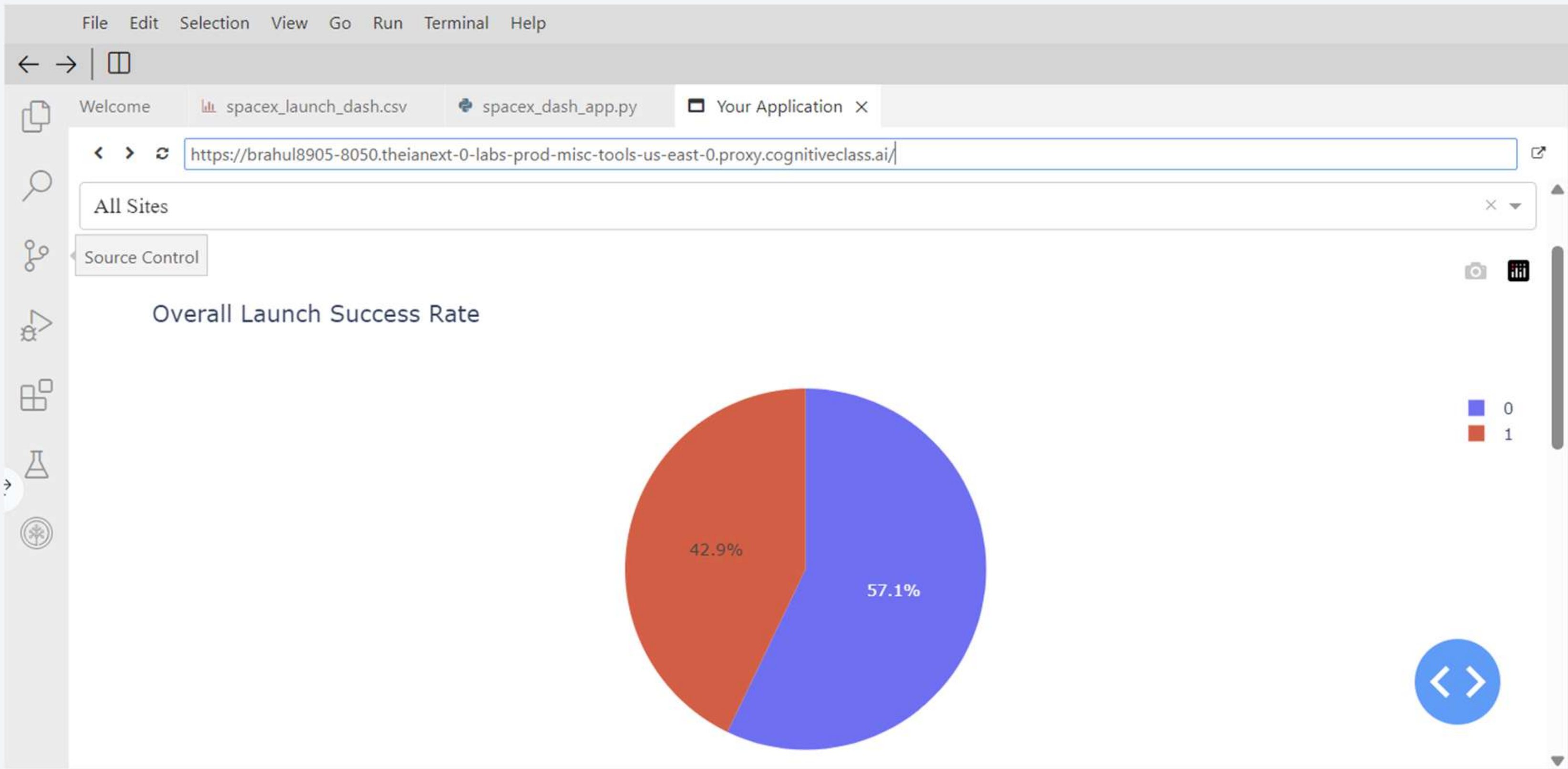
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

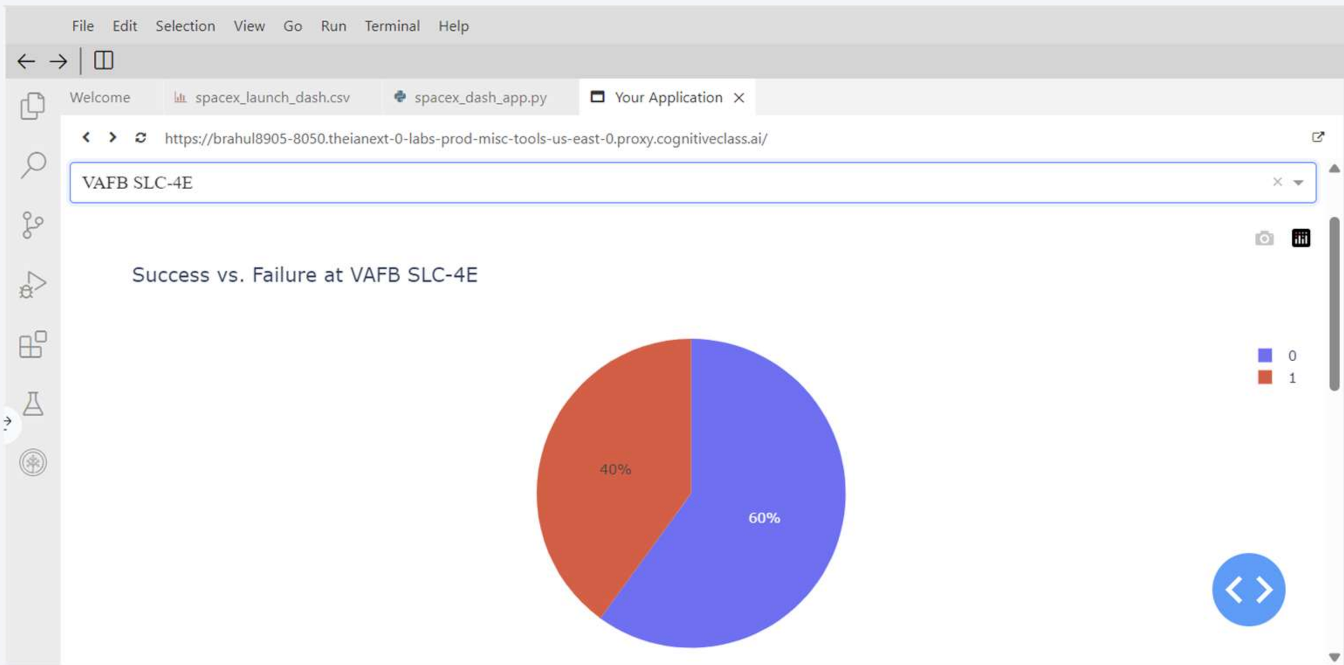
Section 4

Build a Dashboard with Plotly Dash

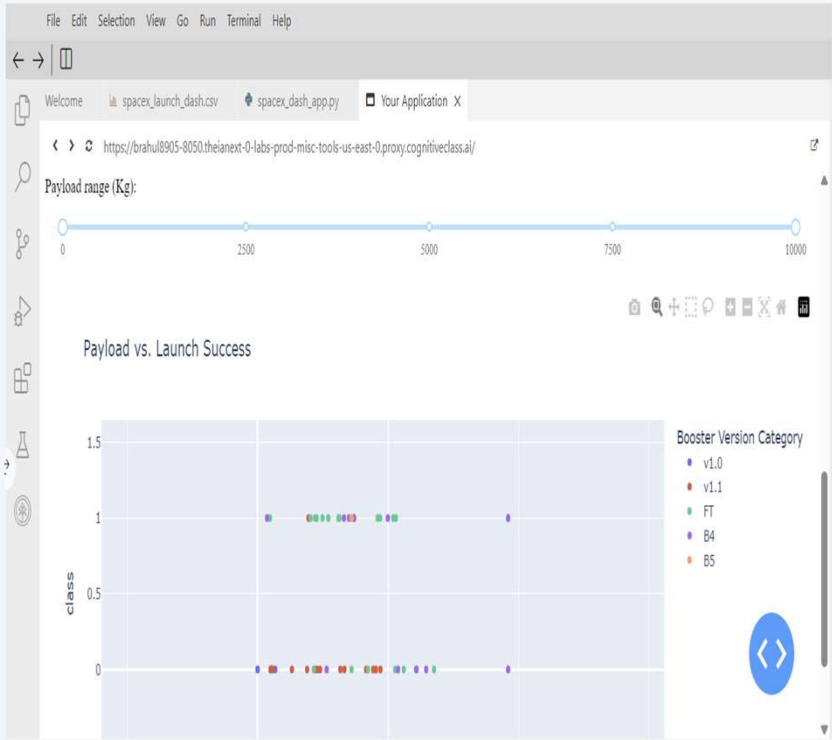
<Dashboard Screenshot 1>

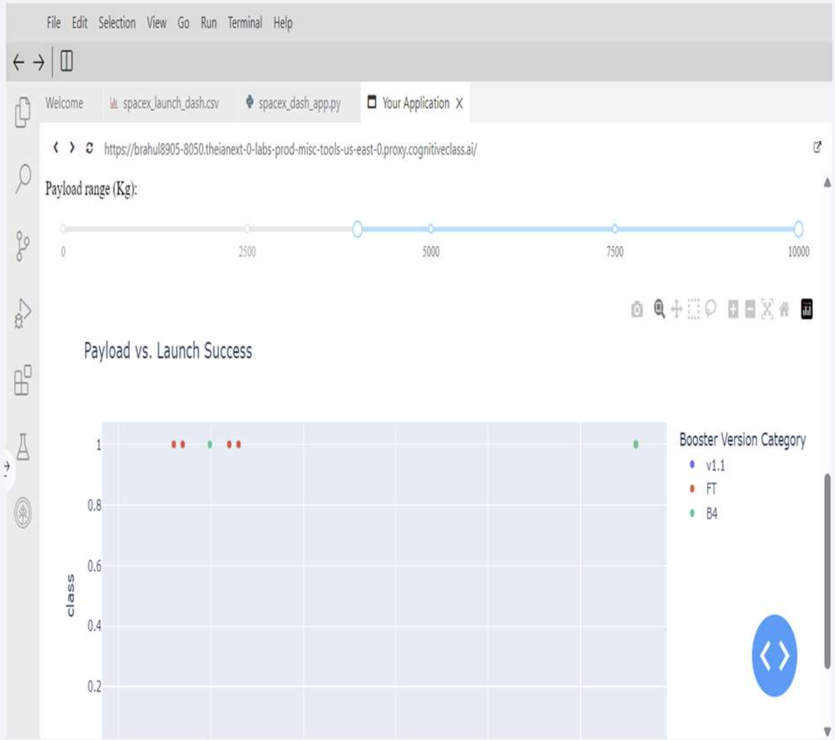


<Dashboard Screenshot 2>



<Dashboard Screenshot 3>

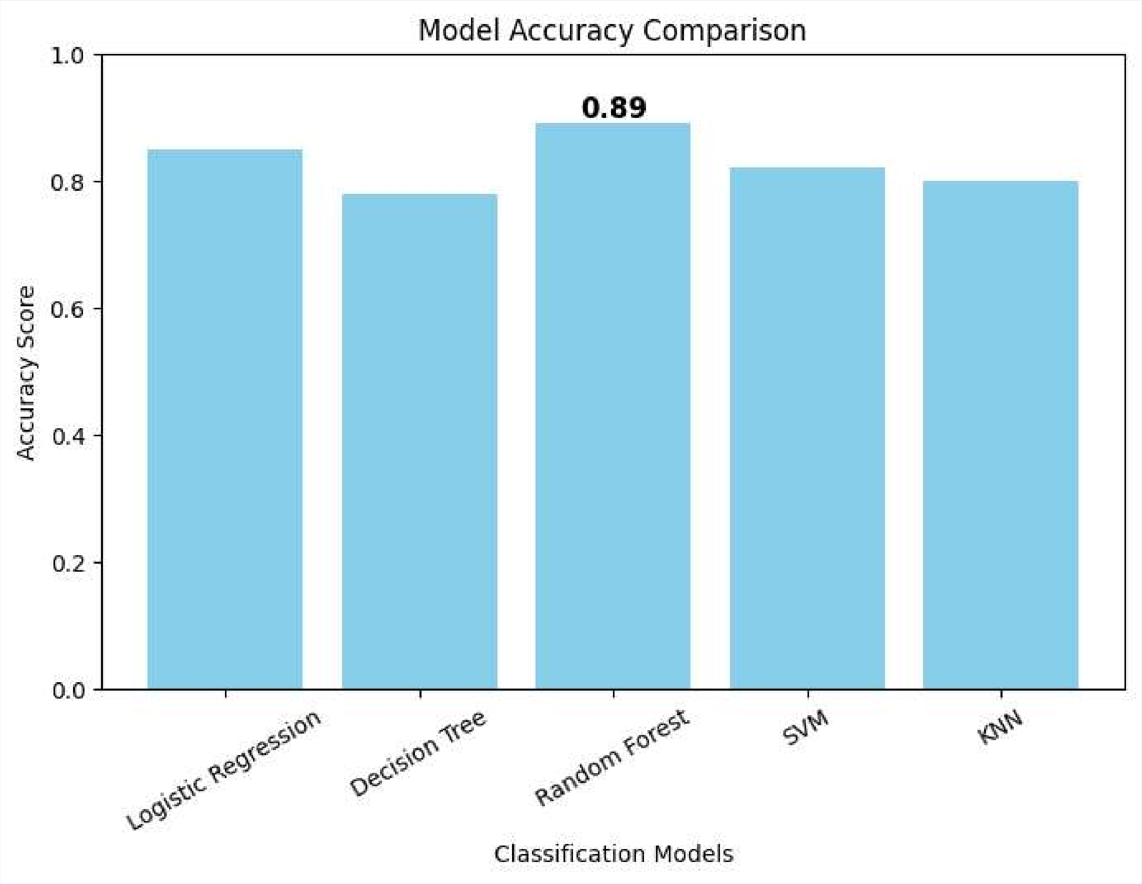




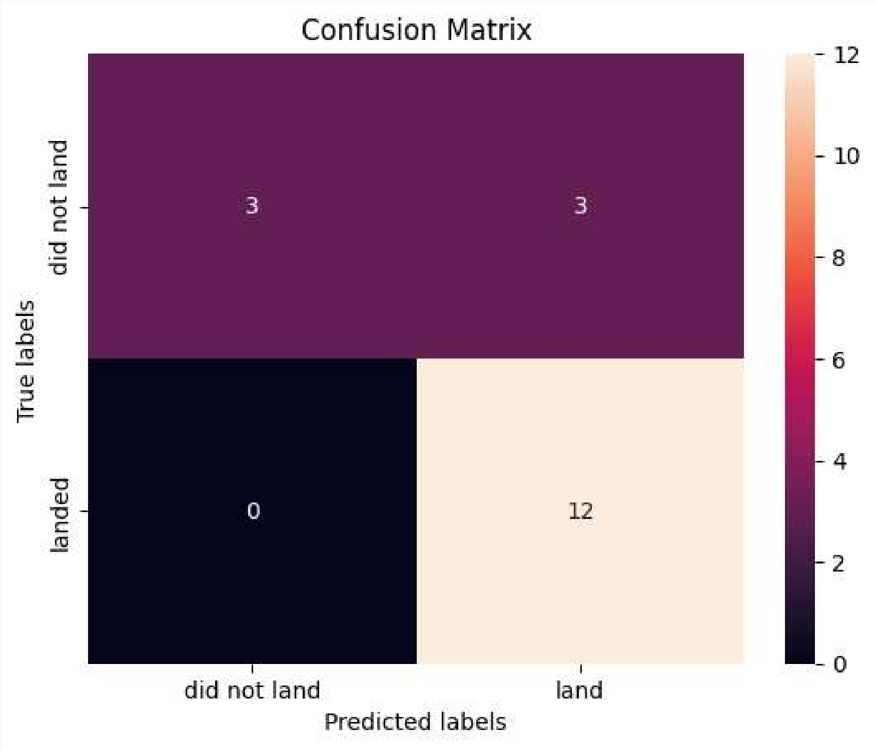
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- Successfully predicted the likelihood of Falcon 9 first-stage landings.
- Identified key factors influencing successful landings.
- Found the best-performing model for prediction.


Thank you!