IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

ASHINZE IFECHUKWUDE ANTHONY
NOVEMBER, 2024

Outline

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

Executive Summary

- This report presents an analysis aimed at predicting the success rate of SpaceX Falcon 9's first-stage booster landings. Using data retrieved from SpaceX's REST API and web scraping Falcon 9 launch records from Wikipedia, we applied extensive data cleaning, exploratory data analysis (EDA), and predictive modeling to uncover insights. A Plotly Dash application was developed for interactive data exploration, and various machine learning algorithms were employed to establish the model with the highest predictive performance.
- Key findings indicate that certain orbit types and payload weights significantly impact landing success rates. Additionally, as flight numbers increased over time, success rates improved. Launch sites, orbit types, and payload masses were examined for their correlation with successful landing outcomes, providing insights for future mission planning and reliability predictions.

Introduction

- SpaceX’s Falcon 9 rocket series has revolutionized reusable rocketry by successfully landing and reusing first-stage boosters. This analysis aims to identify the factors that influence the success rate of these landings, assess the significance of variables like payload mass, orbit type, and launch site, and predict future landing outcomes. By investigating historical launch data, this report provides insights into patterns and the technological evolution of Falcon 9's landing success.
- SpaceX lists Falcon 9 rocket launches on its website at a cost of $62 million, while other providers charge over $165 million per launch. A large part of these savings comes from SpaceX’s ability to reuse the first stage of the rocket. Therefore, knowing whether the first stage will successfully land allows us to estimate the launch cost.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API: to retrieve Falcon 9 launch data programmatically using Python's requests library.
  - Web Scraping from Wikipedia: Falcon 9 and Falcon Heavy historical launch records were scraped to gather additional contextual information
- Perform data wrangling
  - Missing data was handled by using column means where appropriate, and irrelevant fields were removed.
  - The mission outcome was transformed into binary training labels (1 for successful landings and 0 for unsuccessful ones)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Feature engineering, data standardization, data splitting, model implementation, model evaluation.

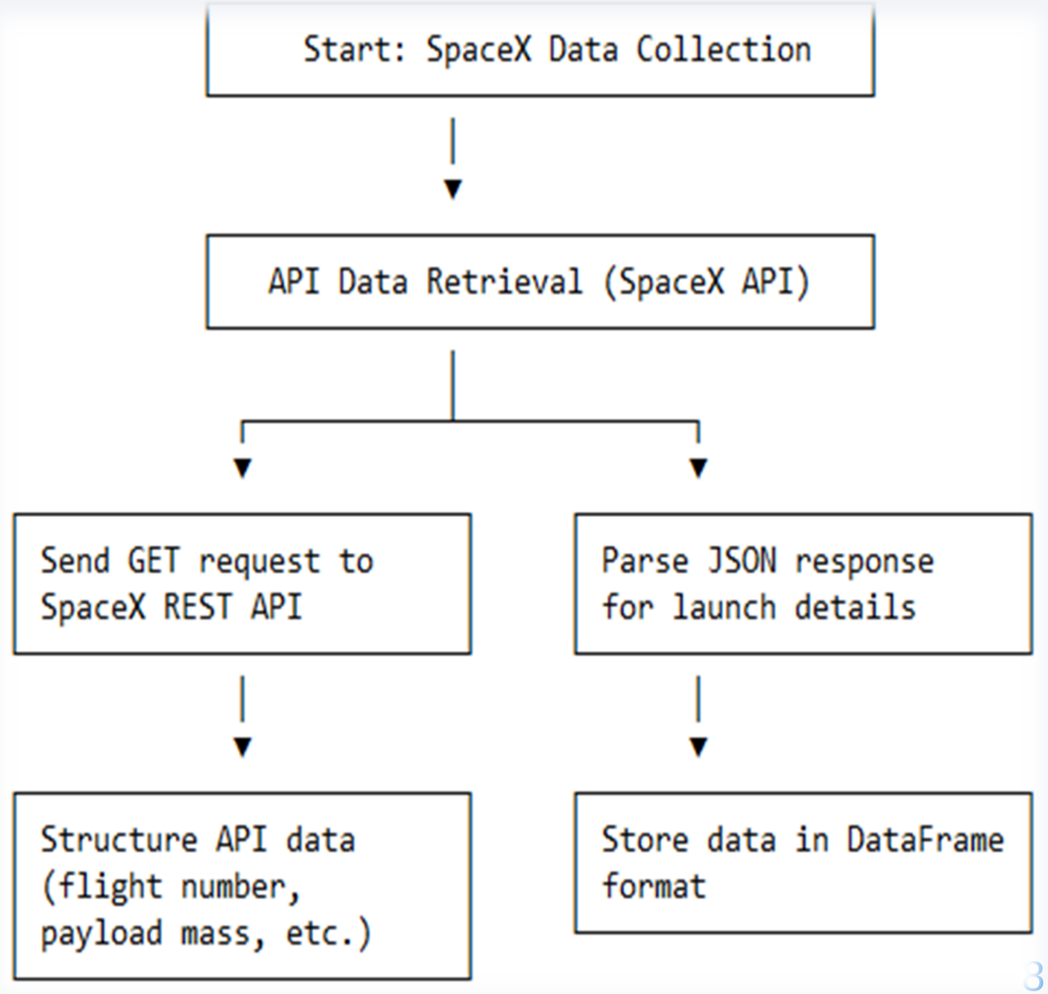
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

❖ API Data Retrieval

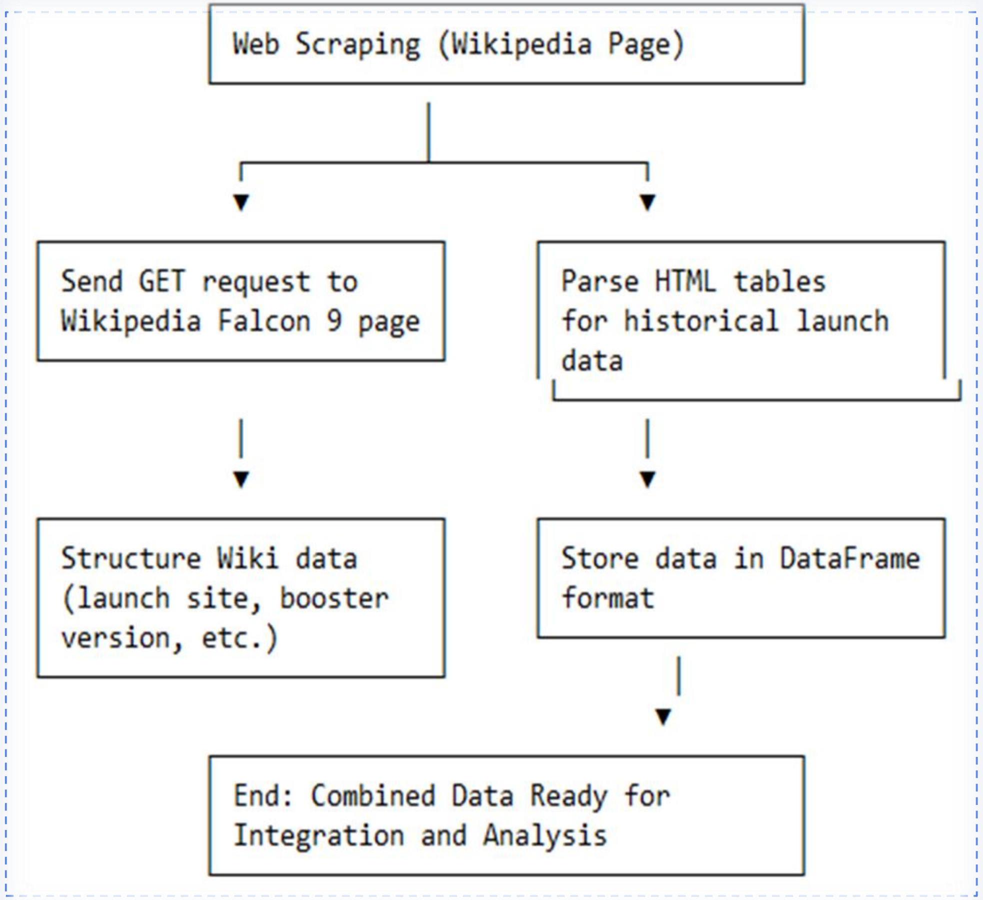
- Source: SpaceX REST API
- Tool: Python requests library
- Purpose: Retrieve detailed historical data for Falcon 9 launches.
- Steps:
    - Send GET request to SpaceX REST API endpoint.
    - Parse JSON response for fields like launch date, flight number, payload mass, orbit type, and landing outcome.
    - Store results in a structured format (DataFrame).
- https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

❖ Web Scraping

- Source: Wikipedia - "List of Falcon 9 and Falcon Heavy launches“
- Tool: BeautifulSoup, Python requests
- Purpose: Supplement API data with additional contextual information.
- Steps:
  - Send GET request to Wikipedia page.
  - Parse HTML to locate tables containing historical launch details.
  - Extract fields such as launch site, mission outcome, and booster version.
  - Store results in a structured format (DataFrame).



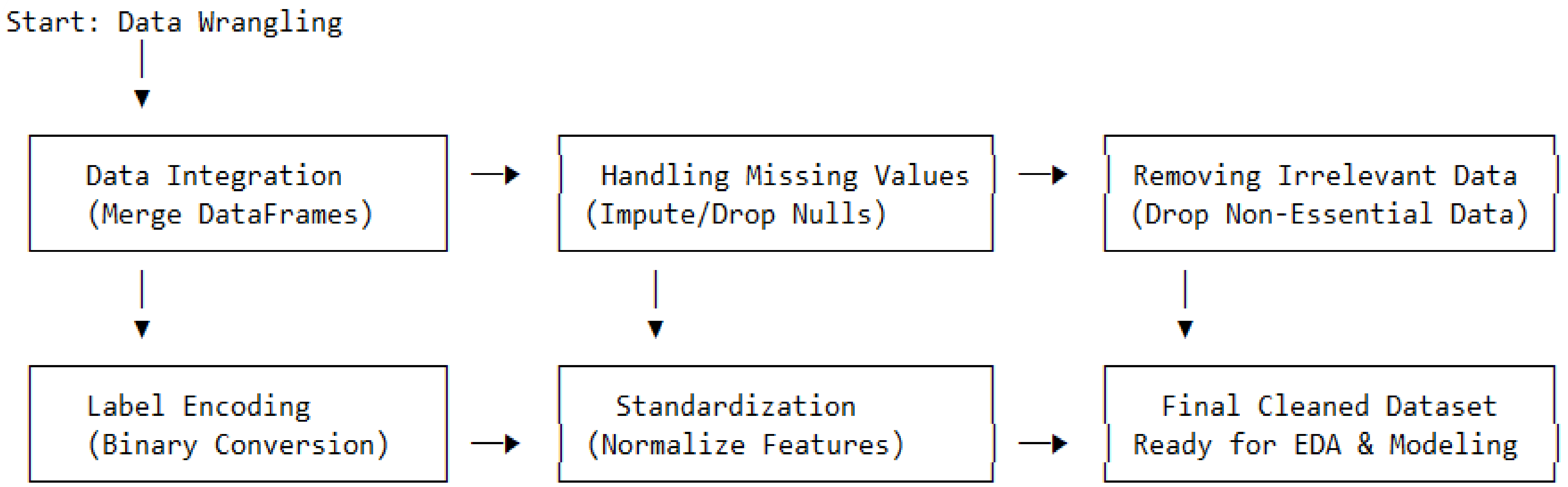
❖ https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Data Integration: Merged SpaceX API and Wikipedia data by matching fields like launch dates and flight numbers for a unified dataset.
- Handling Missing Values: Imputed missing values with column means and removed excessive nulls, minimizing data gaps.
- Removing Irrelevant Data: Dropped non-essential columns to focus on key factors like payload, orbit, and outcome.
- Label Encoding: Converted mission outcomes to binary labels (1 for success, 0 for failure) for model training.
- Standardization: Normalized numerical fields, producing a dataset ready for analysis and modeling.

❖ https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Data Wrangling – Flow Chart



EDA with Data Visualization

- Scatter and Bar Charts: Visualized relationships between flight number, payload mass, launch site, orbit type, and success rates to identify patterns in launch success based on mission parameters.
- Line Plot: Illustrated success rate trends by year, highlighting improvements in landing success over time.
- These charts were chosen to visualize key relationships in the data, such as how launch success correlates with payload mass, orbit, and flight number, and to illustrate trends in SpaceX’s progress with reusable booster technology over time.

- https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/edadataviz.ipynb

EDA with SQL

- Unique Launch Sites
- Launch Sites Starting with 'CCA‘
- Total Payload Mass by NASA (CRS)
- Average Payload Mass for F9 v1.1
- First Successful Ground Pad Landing
- Boosters with Drone Ship Success (4000–6000 kg)
- Mission Outcomes (Success/Failure Counts)
- Booster Versions with Max Payload
- 2015 Drone Ship Failures
- Landing Outcomes Rank (2010–2017)

- https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers: Placed at each launch site to indicate location.
- Circles: Added to represent successful and failed launches, color-coded for clarity.
- Lines: Drawn to measure distances between launch sites and nearest cities, offering proximity insights.

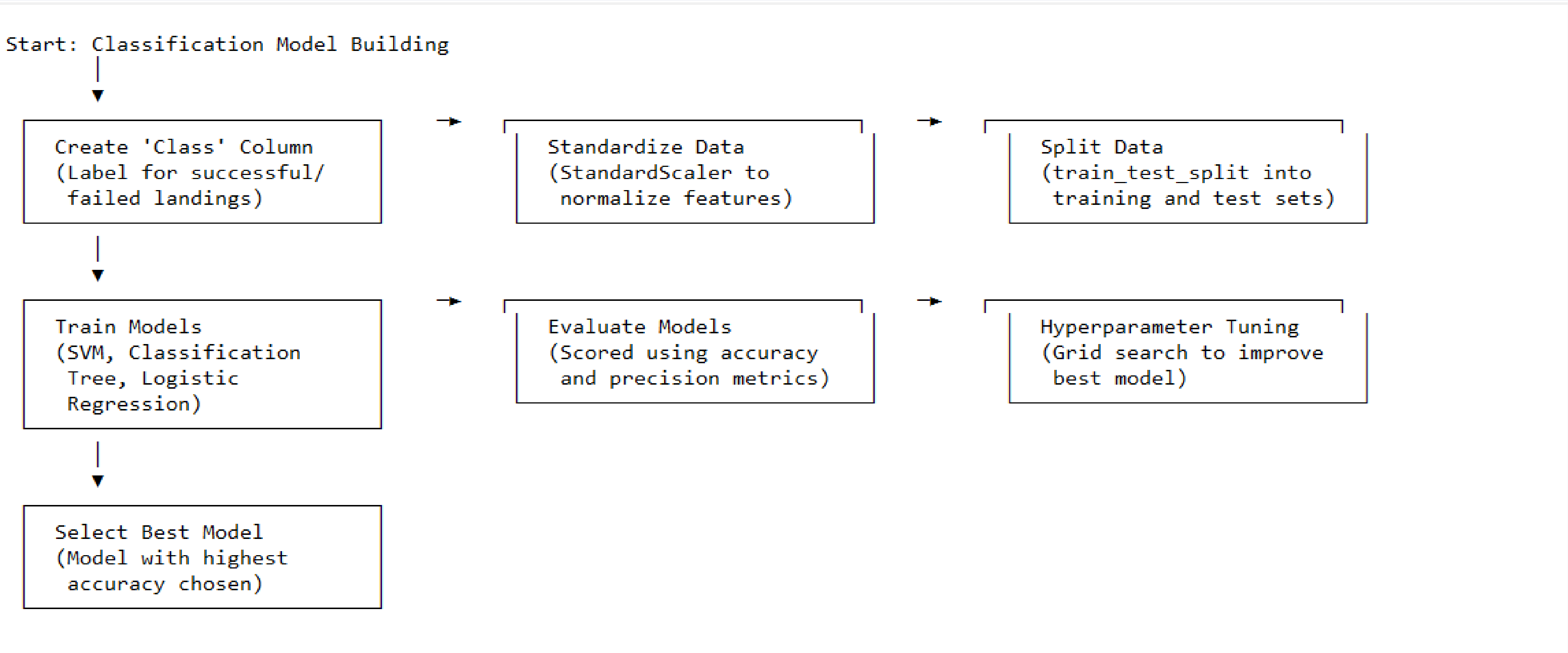
- https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie Chart: Displays mission success rates, allowing quick success/failure comparison.
- Scatter Plot: Shows relationships between payload and launch success with interactive filters for orbit type and launch site.
- Dropdown and Slider Interactions: Users can adjust filters (e.g., launch site, payload range) to dynamically update visualizations.

- https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/spacex_dash_app.txt

Predictive Analysis (Classification)



https://github.com/AIAshinze/SPACEX-first-stage-success-analysis/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

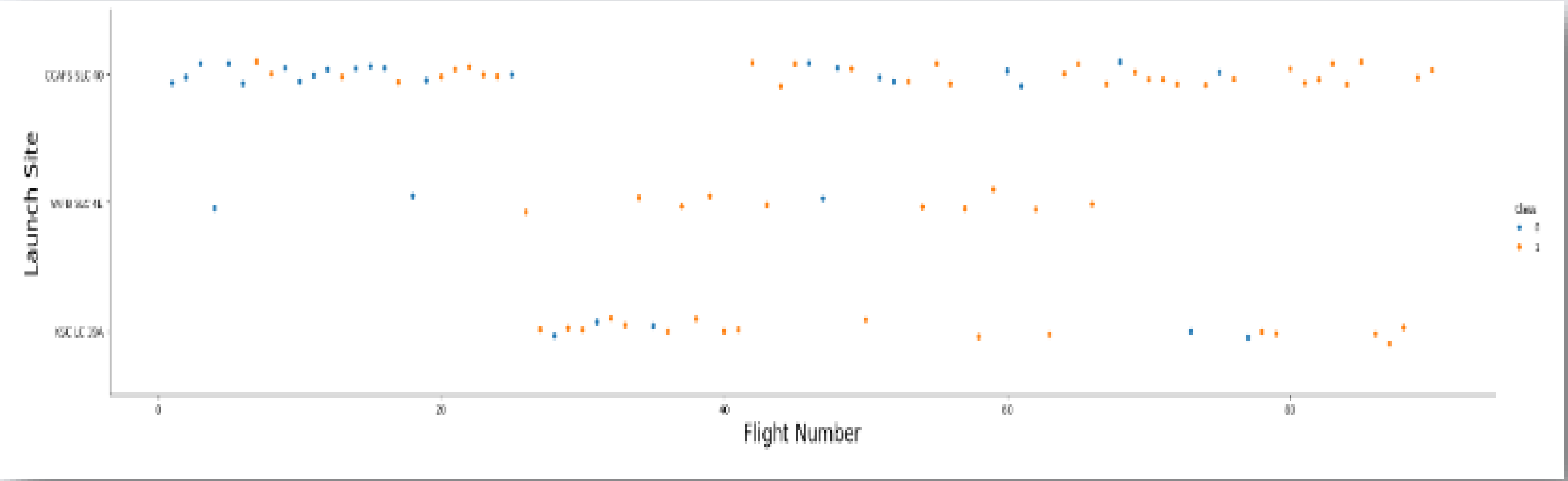
Results

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

Section 2

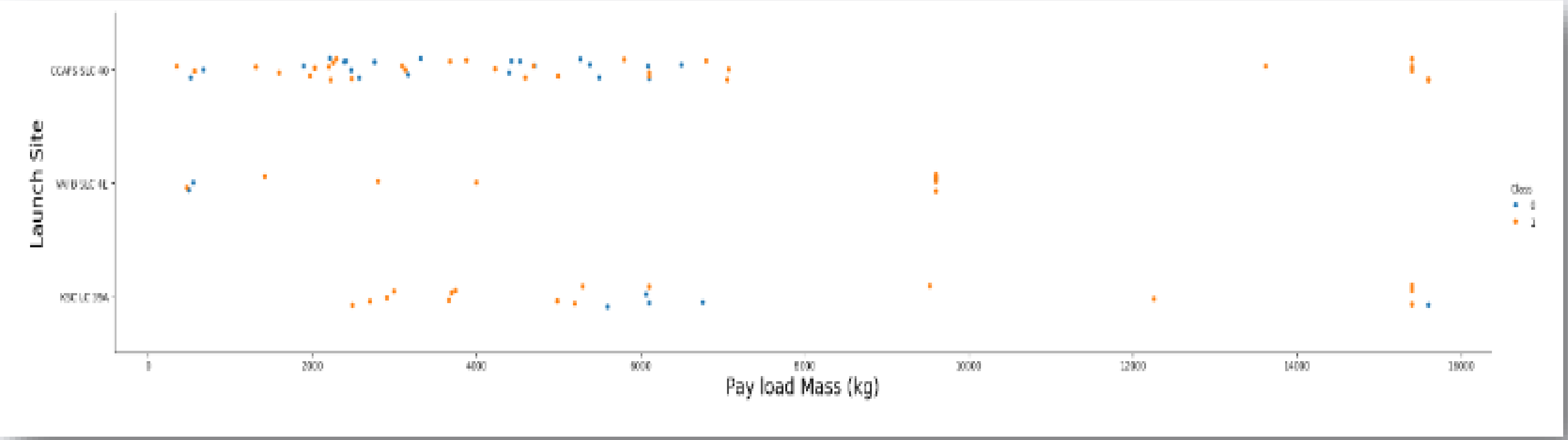
Insights drawn from EDA

Flight Number vs. Launch Site



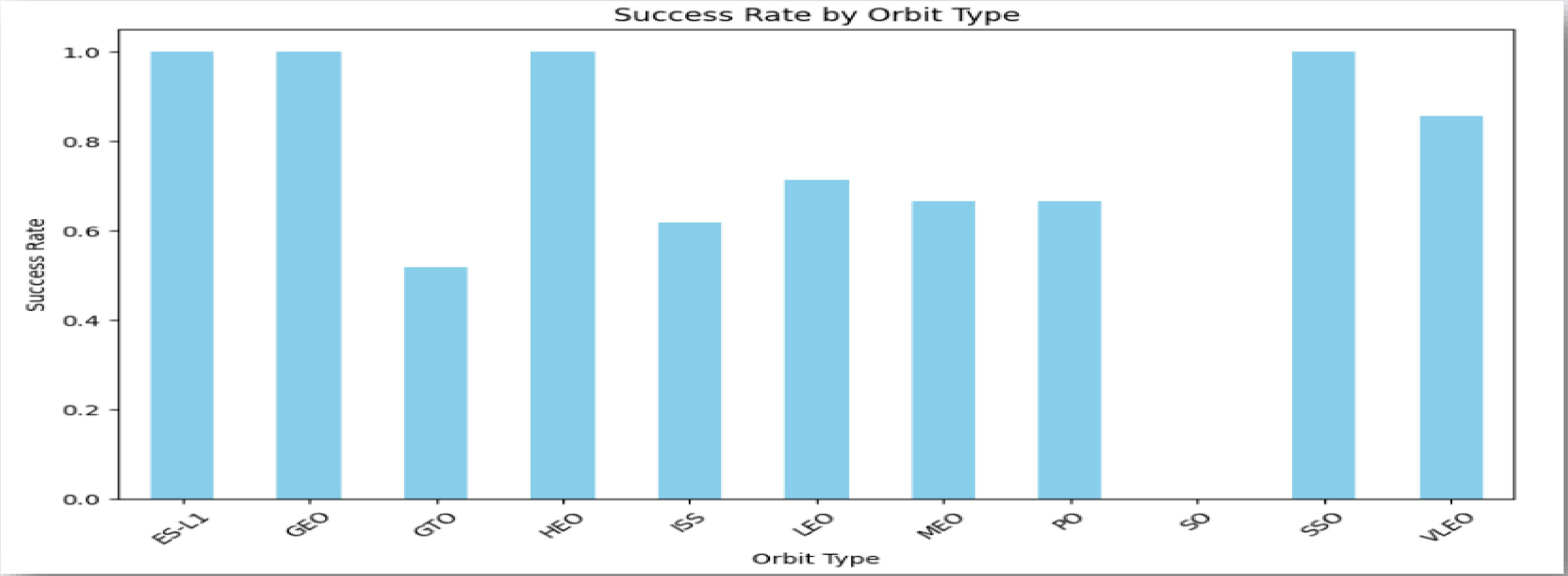
Shows the distribution of launches across various sites, highlighting site usage frequency as flight numbers increase. CCAFS SLC-40, recorded the highest flight numbers.

Payload vs. Launch Site



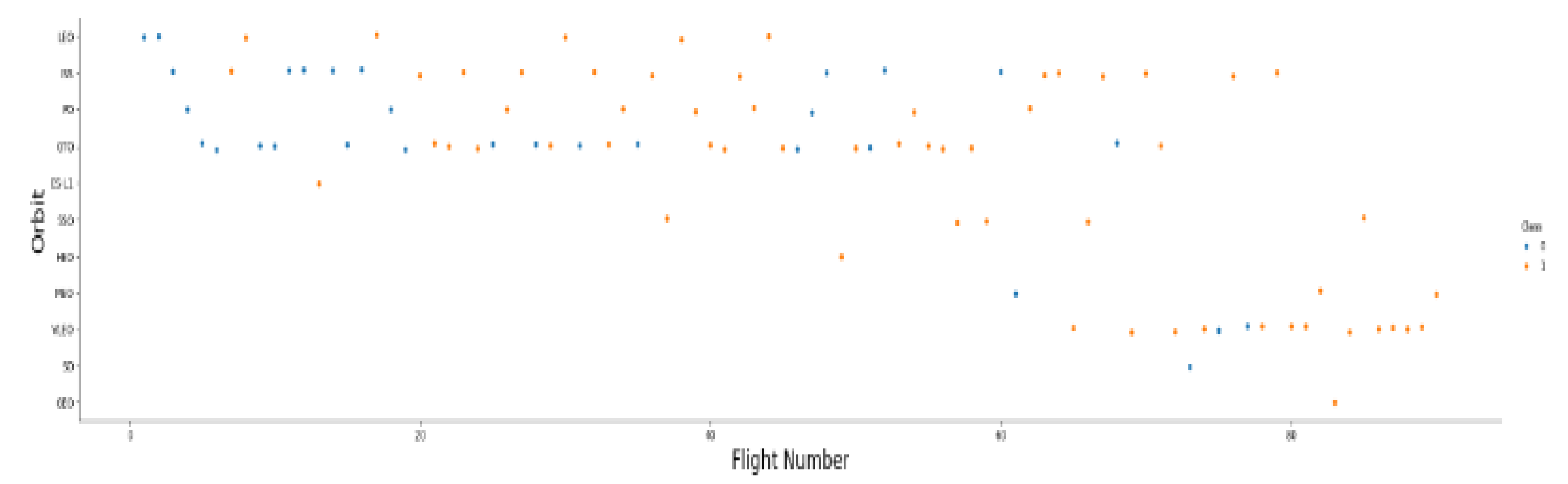
Examines payload mass distribution across launch sites, identifying any launch sites specialized in handling larger payloads. VAFB-SLC site did not launch rockets with payloads exceeding 10,000 kg.

Success Rate vs. Orbit Type



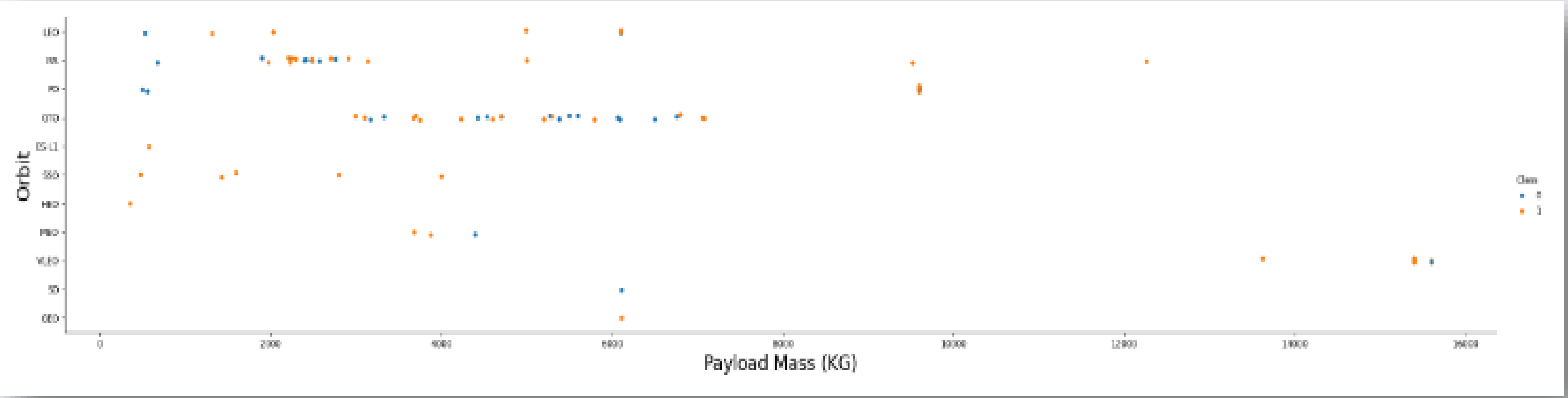
Shows success rate across different orbit types, helping identify which orbits are most reliable for successful landings. Orbits like LEO (Low Earth Orbit), GEO (Geostationary Orbit), and SSO (Sun-Synchronous Orbit) recorded the highest success rates.

Flight Number vs. Orbit Type



Analyzes the relationship between orbit types and flight numbers, useful for assessing the evolution of mission types over time. Success rates in the orbit types appear to correlate with the number of flights, suggesting increased reliability over time.

Payload vs. Orbit Type



Demonstrates the relationship between payload, orbit type, and landing success, identifying which orbits and payload sizes correlate with successful landings. Increased success rates have been recorded with heavy payloads in Polar, LEO, and ISS orbits, while GTO displayed mixed outcomes.

Launch Success Yearly Trend

Tracks the trend of success rates over time, revealing improvements in landing success with increasing launch experience. Landing success rates have consistently risen since 2013, reaching a peak in 2020.

All Launch Site Names

- Unique Launch Sites: Retrieved unique launch sites, including CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40.
- Helps identify site distribution.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landi
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failu
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	Failu
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

Filtered 5 records with sites beginning 'CCA'. Used for site-specific analysis.

Total Payload Mass

- Total Payload Mass by NASA (CRS): Calculated 48,213 kg.
- Shows total payload handled by NASA missions.

Average Payload Mass by F9 v1.1

- Average Payload Mass for F9 v1.1: Found 2,928.4 kg.
- Highlights payload capacity for specific booster versions.

First Successful Ground Landing Date

- First Successful Ground Pad Landing: Date was 2015-12-22.
- Marks milestone in landing success.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters with Drone Ship Success (4000–6000 kg): Identified F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2.
- Analyzes booster success based on payload.

Total Number of Successful and Failure Mission Outcomes

- Mission Outcomes (Success/Failure Counts): Found 98 successes, 1 in-flight failure.
- Quantifies overall success and reliability.

Boosters Carried Maximum Payload

- Booster Versions with Max Payload: Includes F9 B5 B1048.4, F9 B5 B1051.6, F9 B5 B1060.3, etc.
- Highlights top-performing boosters.

2015 Launch Records

- In January 2015, booster version F9 v1.1 B1012 failed to land successfully in drone ship, n CCAFS LC-40 launch site.
- In March 2015, booster version F9 v1.1 B101 failed to land successfully in drone ship, on CCAFS LC-40 launch site.
- This Identifies setbacks (mission failure) recorded in 2015

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

LANDING OUTCOME	OUTCOME COUNT
NO ATTEMPT	10
SUCCESS (DRONE SHIP)	5
FAILURE (DRONE SHIP)	5
SUCCESS (GROUND PAD)	3
CONTROLLED (OCEAN)	3
UNCONTROLLED (OCEAN)	2
FAILURE (PARACHUTE)	2
PRECLUDED (DRONE SHIP)	1

Ranked outcomes with No attempt (10), Success (drone ship) (5), etc. Prioritizes landing success types.

Section 3

Launch Sites Proximities Analysis

Map - Launch Site Location

Markers placed at each launch site to indicate location on a folium map

Map - Launch Outcome

Folium circles added to represent successful and failed launches, color-coded for clarity.

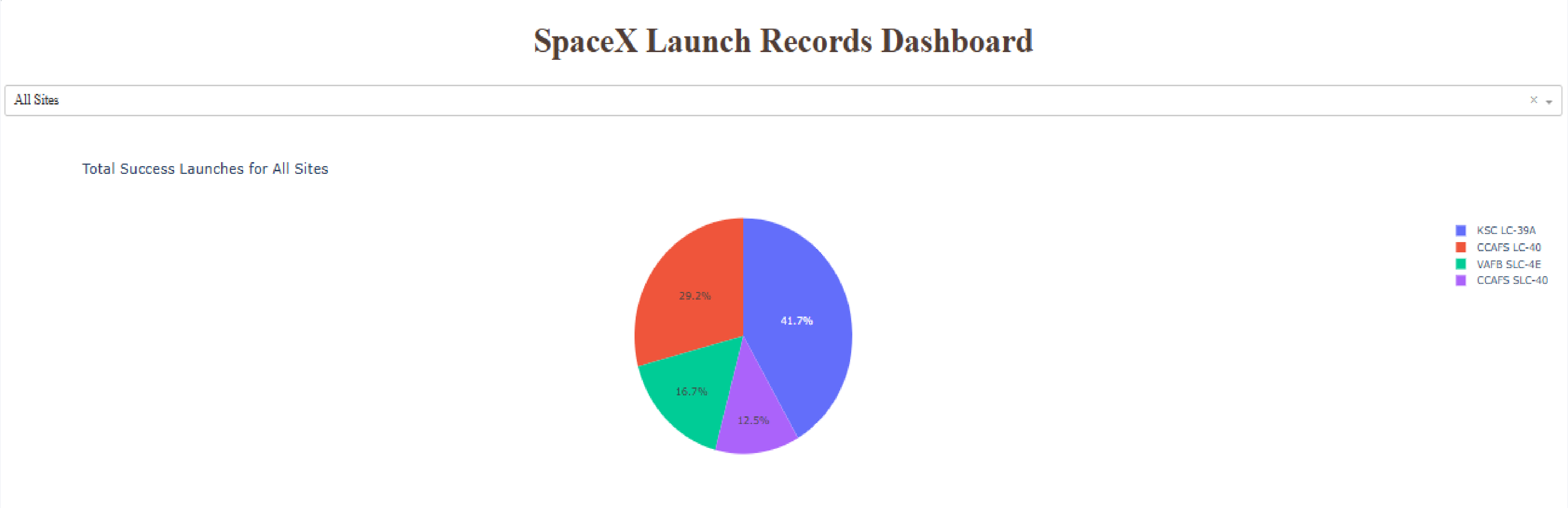
Launch Site Proximity

Folium lines drawn to measure distances between launch sites and nearest cities, offering proximity insights.

Section 4

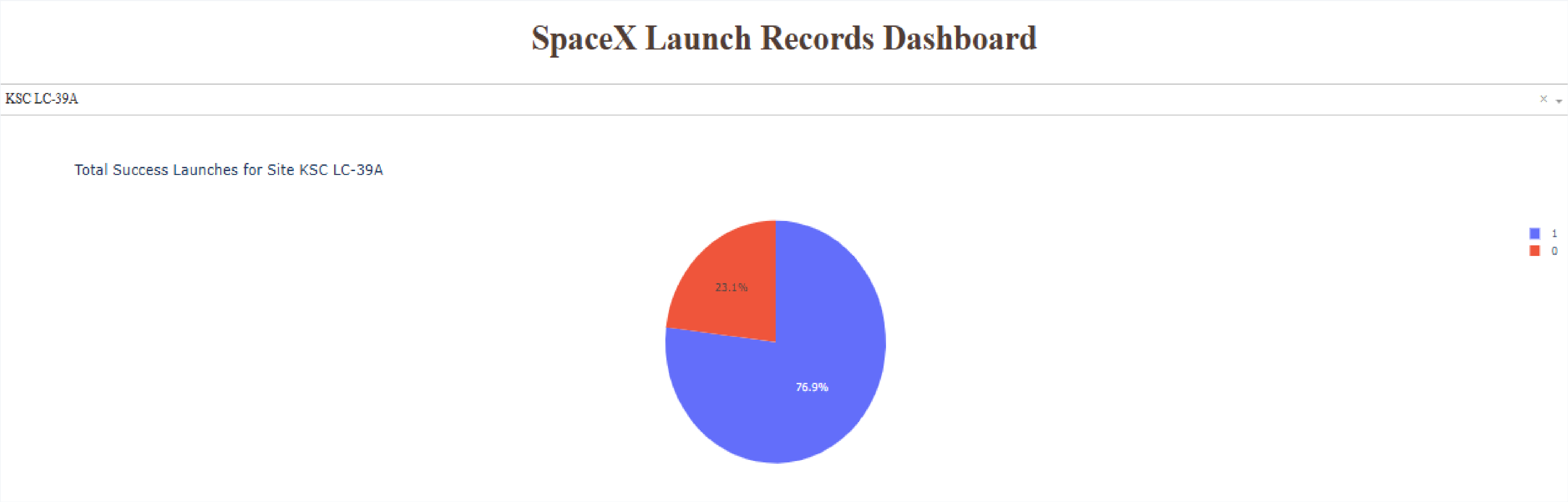
Build a Dashboard with Plotly Dash

Dashboard – Launch Success Piechart



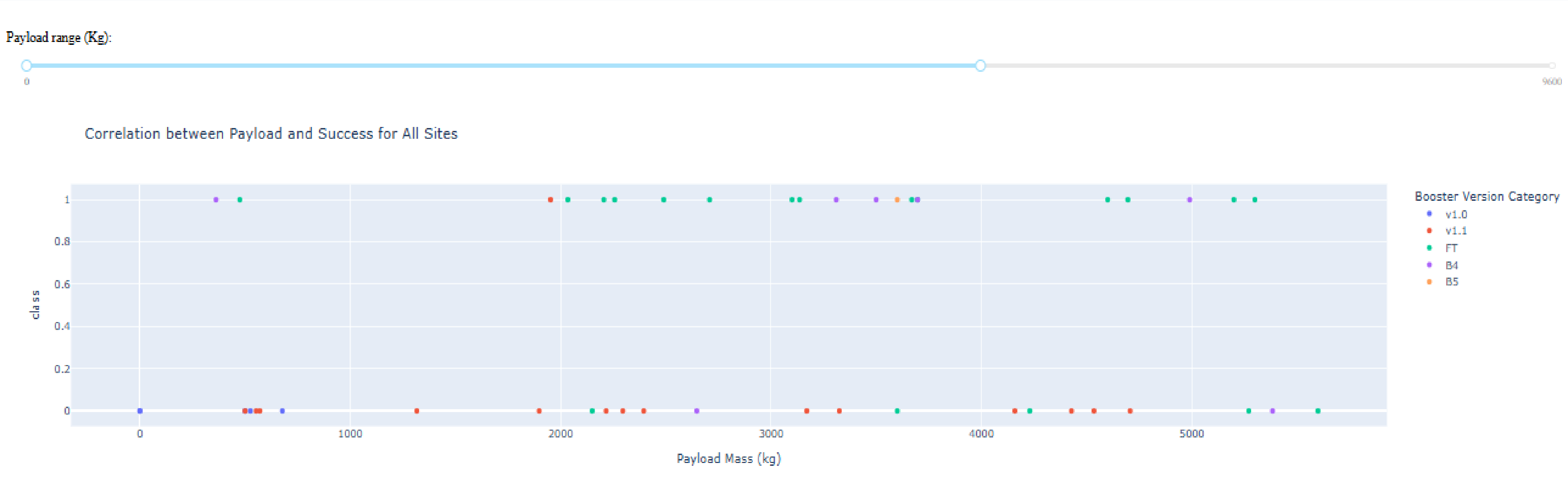
Dashboard Pie Chart displays mission success rates, allowing quick success/failure comparison.

Dashboard – Launch Site Success Ratio



Pie Chart displays mission success rates for launch site KSC LC-39A, the launch site recorded the highest mission success rate

Dashboard – Payload VS Launch Outcome slider

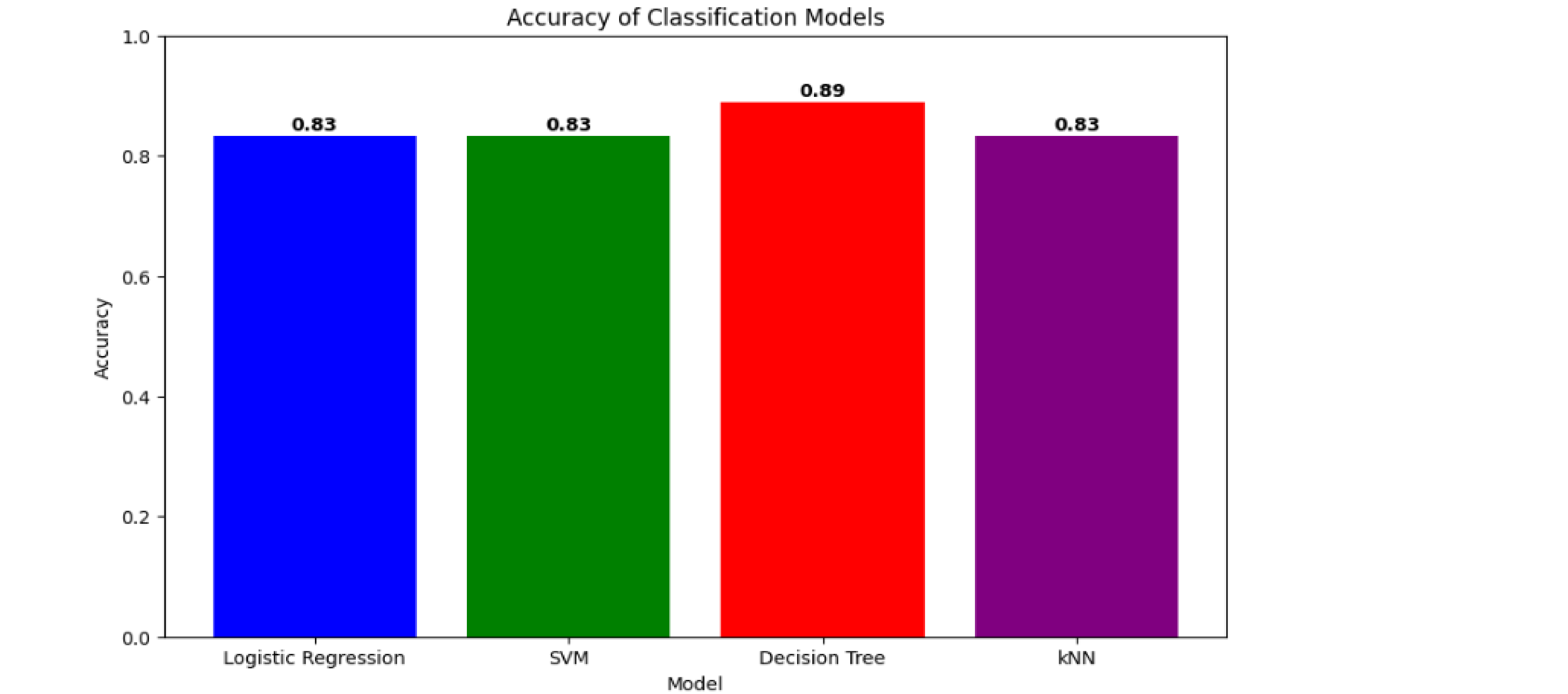


A Scatter Plot shows relationships between payload and launch success with interactive filters for orbit type and launch site. Users can adjust filters (e.g., launch site, payload range) to dynamically update visualizations using Slider Interactions.

Section 5

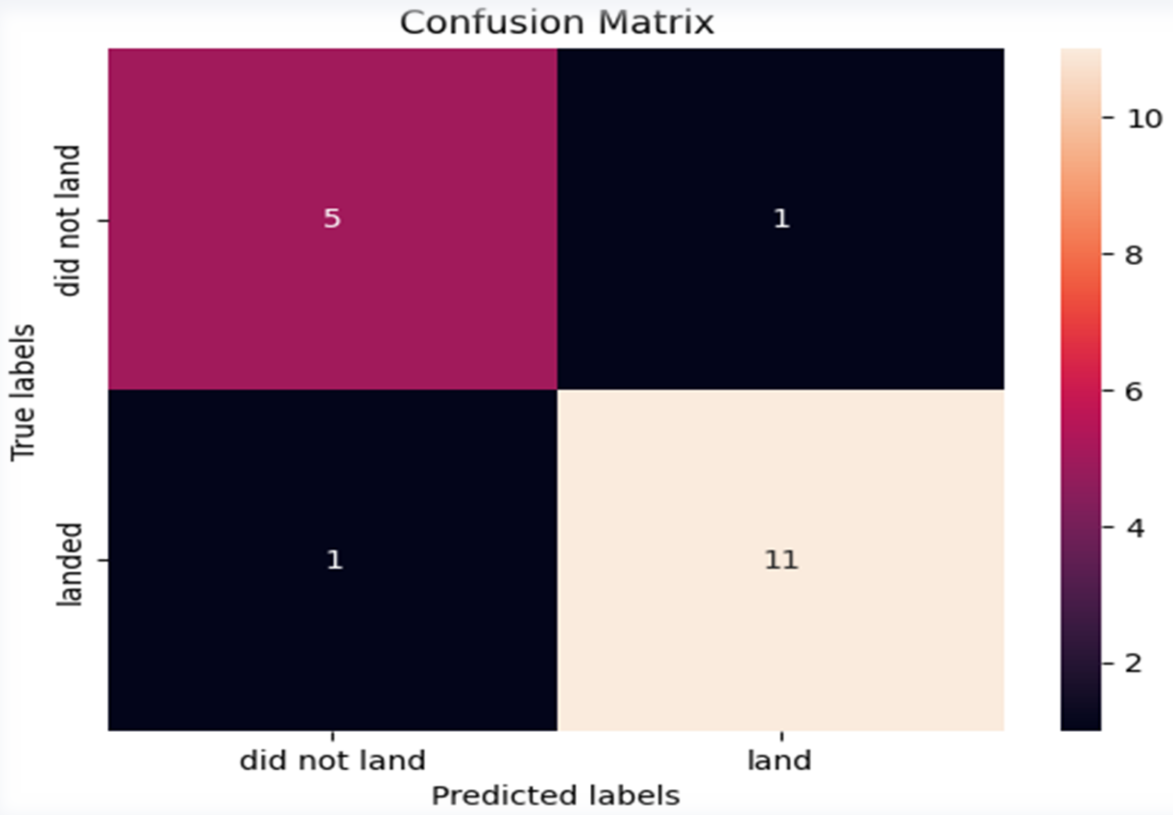
Predictive Analysis (Classification)

Classification Accuracy



The decision tree model emerged as the classification model with the best accuracy

Confusion Matrix



The decision tree confusion matrix indicates that the model correctly predicted 5 instances where the booster did not land and 11 instances where it successfully landed, with 2 misclassifications (1 false positive and 1 false negative).

Conclusions

- Success rates are influenced by orbit type, payload mass, flight numbers, and launch site.
- Analysis shows increasing success rates over time, especially for heavy payloads in LEO and Polar orbits.
- The success-orbit relationship suggests certain missions are better suited for reusable rocket technology, aligning with SpaceX's goals.
- Folium maps and SQL queries provided detailed insights into mission data.
- Machine learning models enabled predictive analysis of mission outcomes.
- The Plotly Dash application allows stakeholders to interactively visualize data, supporting real-time decision-making for future missions.

Appendix

Link to project Github repository main branch:

[AIAshinze/SPACEX-first-stage-success-analysis: This project will try to predict if a Falcon 9 first stage launch will land successfully. (github.com)]

Thank you!