IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name - Fahad Ahmed
Date - 2nd July 2025
Contact Info – abc@gmail.com

Outline

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

Executive Summary

🔧 Methodology Overview (Summary of methodologies)

- **Data Preparation**: Combined and cleaned datasets (dataset_part_1.csv, dataset_part_2.csv, spacex_launch_geo.csv). Engineered binary target: Class (1 = Success, 0 = Failure).
- **Exploratory Analysis**: Charts on Flight Number vs Payload, Launch Sites, Orbits. Tools used: pandas, matplotlib, seaborn, numpy.
- **Feature Engineering**: One-hot encoded categorical features like Orbit, LaunchSite.
- **Geospatial Analysis (Folium)**: Launch site clustering, distance to coast, rail, highway (Haversine).
- **Interactive Dashboard (Dash)**: Features site filter, payload range slider, dynamic Pie & Scatter plots.
- **Predictive Modeling:** Test Accuracy: Logistic Regression: 83.33%, SVM: 83.33%, KNN: 83.33%, Decision Tree: 88.89%

🔍 Key Insights (Summary of all results)

- **Most Launches**: CCAFS SLC 40
- **Highest Success Rate**: KSC LC 39A
- **Overall Success Rate**: ~67%
- **Optimal Orbits**: LEO, ISS, PO
- **Payload Effect**: Heavier payloads succeed more (LEO/ISS)
- **Closest Coastline**: CCAFS (~1.27 km)
- **Top Boosters**: B5 & FT versions

🛠️ Tools Used

- Python
- pandas
- seaborn
- matplotlib
- folium
- plotly
- Dash

Introduction

- Project background and context

SpaceX's **Falcon 9** rocket is a two-stage reusable launch system that has revolutionized commercial spaceflight. With its mission to reduce space travel costs through reusability, SpaceX has conducted hundreds of launches, each with varying payloads, orbits, and boosters.
This project analyzes **historical Falcon 9 launch data** to uncover key factors behind mission success, evaluate site-specific performance, and build models that can help **predict the success of future launches**.

- Problems you want to find answers

1. **Which launch site** has the most successful missions?
2. **What payload mass range** yields the highest success rate?
3. **Do orbit type, payload mass, or booster version** significantly affect outcomes?
4. **Can we accurately predict** a launch's success using mission features?

Section 1

Methodology

Methodology

Executive Summary

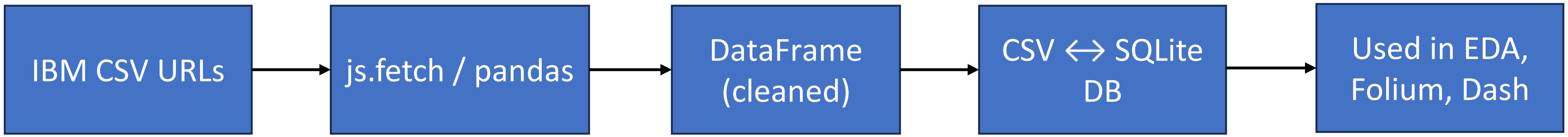
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using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

Data Collection

Data Collection Summary

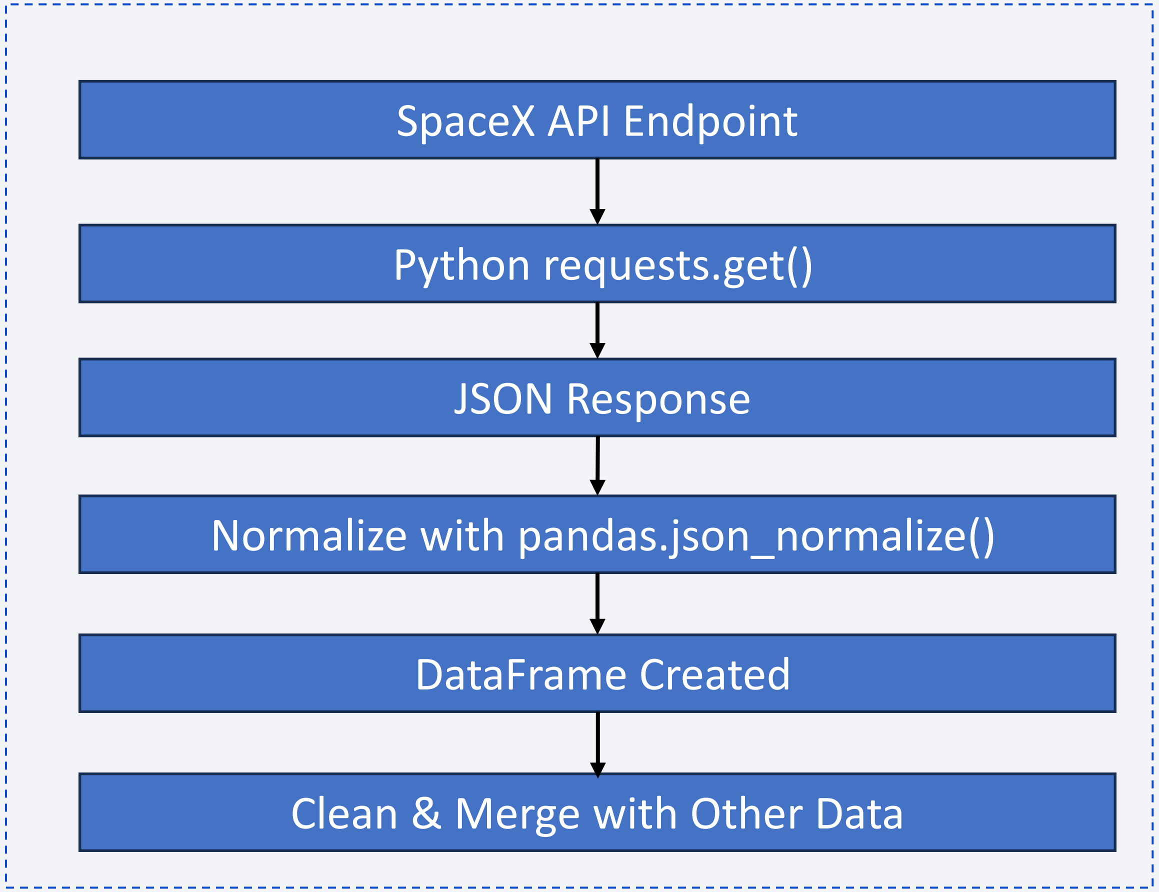
- **Source:** IBM Cloud-hosted CSV files (e.g., dataset_part_1.csv, spacex_launch_geo.csv)
- **Tools Used:** pandas, js.fetch, sqlite3
- **Method:**
  - CSVs fetched via read_csv() or fetch in Theia environment
  - Loaded into Pandas DataFrames for cleaning
  - Saved locally and imported into SQLite for SQL queries
  - Geo data (spacex_launch_geo.csv) used for Folium mapping

Flowchart (Condensed)



Data Collection – SpaceX API

- Data Collection – SpaceX API
  - **Source:** SpaceX REST API (https://api.spacexdata.com/v4/launches)
  - **Access Method:** requests library in Python.
  - **Purpose:** To retrieve real-time launch data including payload, launch site, rocket type, and landing outcome
  - **Processing:** JSON data → Pandas DataFrame → Cleaned and merged with other sources
- GitHub URL- https://github.com/FahadAhmed-8/space-y-/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb

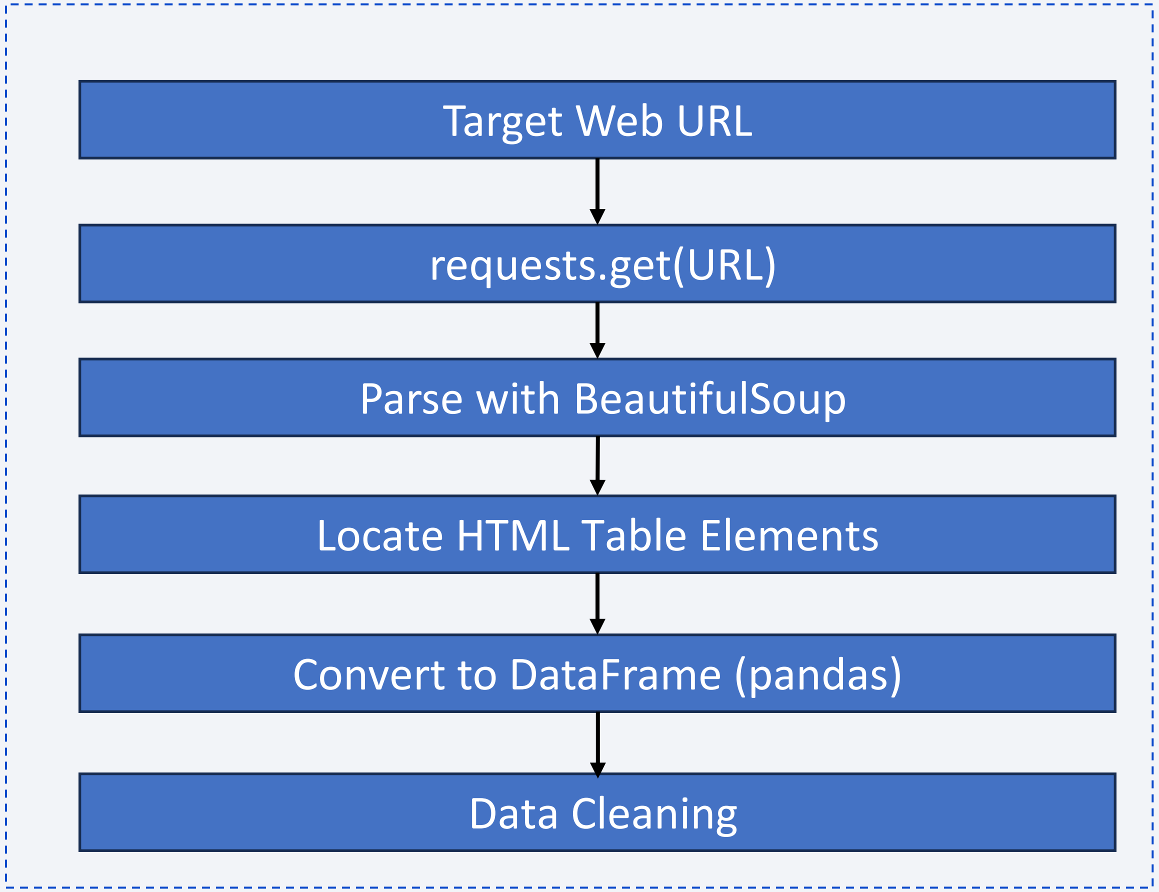


Data Collection - Scraping

- **Data Collection – Web Scraping**

  •**Source:** SpaceX Launch Data web pages
  •**Tools Used:** requests, BeautifulSoup, pandas
  •**Objective:** Extract structured tables (e.g., launch history, payload, mission outcome)
  •**Processing:** Parsed HTML → Extracted table → Converted to DataFrame → Cleaned for analysis

- **GitHub URL -**
  https://github.com/FahadAhmed-8/space-y-/blob/main/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

- **Objective:** Clean and prepare raw data for analysis
- **Steps Performed:**
  - Converted **Date** column to datetime format
  - Extracted **Season** from Date
  - Handled **missing values** (NaN in columns like LandingPad)
  - Created **binary classification** label from Outcome
  - Renamed and standardized column names
  - One-hot encoded categorical features (e.g., **Orbit, LaunchSite**)

- **GitHub URL -** https://github.com/FahadAhmed-8/space-y-/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Chart Type	Purpose
Scatter Plot (Flight # vs Payload)	To observe how launch success varies with flight experience and payload mass
Scatter Plot (Flight # vs Launch Site)	To compare performance trends across launch sites over time
Scatter Plot (Payload vs Orbit)	To study correlation between payload mass and orbit types
Pie Chart (Launch Success by Site)	To visualize proportion of successful launches per site
Line Chart (Year vs Success Rate)	To analyze yearly success trend of missions
Map with Markers (Folium)	To geospatially analyze launch outcomes and site proximities
Interactive Scatter (Plotly Dash)	To explore payload-success patterns across booster types and launch sites

- GitHub URL - https://github.com/FahadAhmed-8/space-y-/blob/main/edadataviz.ipynb

EDA with SQL

SQL Queries Summary

- **Launch Site Analysis**
  - Retrieved **unique launch site names** using SELECT DISTINCT
  - Counted **number of launches per site** with GROUP BY Launch_Site
- **Orbit Type Analysis**
  - Calculated **number and occurrences of each orbit type** using GROUP BY Orbit
- **Mission Outcome Analysis**
  - Counted **success/failure mission outcomes** for each **launch site and orbit type**
- **Payload Mass Insights**
  - Retrieved **minimum and maximum payload mass** using MIN() and MAX() functions
  - Summed up total payloads carried per booster version with SUM() and GROUP BY
- **Filtered Search Queries**
  - Displayed **launch records** using filters like launch site and mission year (WHERE Launch_Site = ..., substr(Date,0,5)='2015')
- **Ranking and Aggregation**
  - Ranked **landing outcomes by frequency** between two dates with GROUP BY and ORDER BY COUNT(*) DESC

- GitHub URL - https://github.com/FahadAhmed-8/space-y-/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Folium Map Summary

- **Markers**:
  - Added for each **launch site** and individual **launch outcomes** (green for success, red for failure)
  - Helped visualize launch results and locations
- **Circles**:
  - Highlighted each **launch site area** for visibility
- **Marker Clusters**:
  - Grouped overlapping markers to avoid clutter
- **Mouse Position Plugin**:
  - Showed real-time coordinates to locate nearby features
- **Distance Markers & Lines**:
  - Marked **coastlines/highways**, showed **distances** from launch sites
- **Polylines** visualized spatial proximity

Why These Were Added

- To provide **intuitive visual cues** about launch locations and success patterns
- To analyze **geographical proximity** to key features
- To support **interactive exploration** and location-based insights

- GitHub URL - https://github.com/FahadAhmed-8/space-y-/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

Dashboard Plots and Interactions Summary

1. **Dropdown (Launch Site Selection):**

   Enables filtering data by site to analyze launch performance per location.

2. **Pie Chart (Launch Success Rate):**

   Visualizes success vs. failure launches; helps compare outcomes across sites.

3. **Range Slider (Payload Selection):**

   Allows interactive selection of payload range to study its impact on outcomes.

4. **Scatter Plot (Payload vs. Launch Outcome):**

   Displays how payload mass relates to success, color-coded by booster version for deeper insight.

Why added: These interactive visualizations help users explore SpaceX launch data, uncover success patterns, and analyze performance by site, payload, and booster.

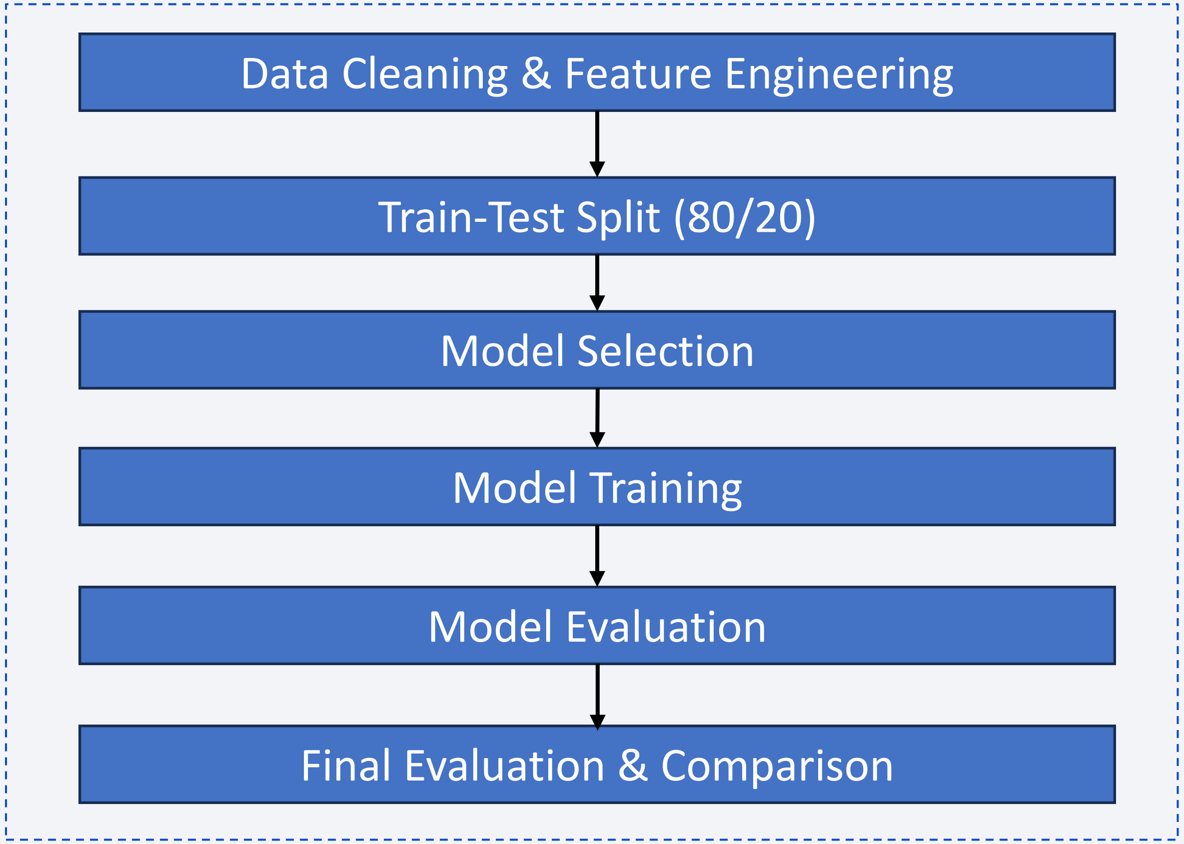
GitHub URL - https://github.com/FahadAhmed-8/space-y-/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

Key Actions & Insights

- **Feature Preparation:** Used one-hot encoding on categorical columns (Orbit, LaunchSite, etc.)
- **Model Evaluation Metric:** Accuracy
- **Tools Used:** sklearn, StandardScaler, GridSearchCV
- **Best Model: Decision Tree** with **88.89% accuracy** on test set
- **Improvement:** Hyperparameter tuning improved Decision Tree over default settings

- GitHub URL - https://github.com/FahadAhmed-8/space-y-/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

Exploratory Data Analysis (EDA) Results

•**Flight Number vs. Success:**
Higher flight numbers showed improved success rates, indicating growing reliability over time.
• **Payload Mass vs. Outcome:**
Heavier payloads (especially in Polar, ISS, and LEO orbits) still achieved successful landings.
•**Orbit vs. Success Rate:**
LEO and ISS had higher success rates compared to GTO, where outcomes were more variable.
•**Launch Site Analysis:**
Some launch sites (e.g., CCAFS SLC 40) had more frequent and successful launches.
•**Yearly Trend:**
Success rates increased steadily from 2010 to 2020, reflecting SpaceX's improvements.

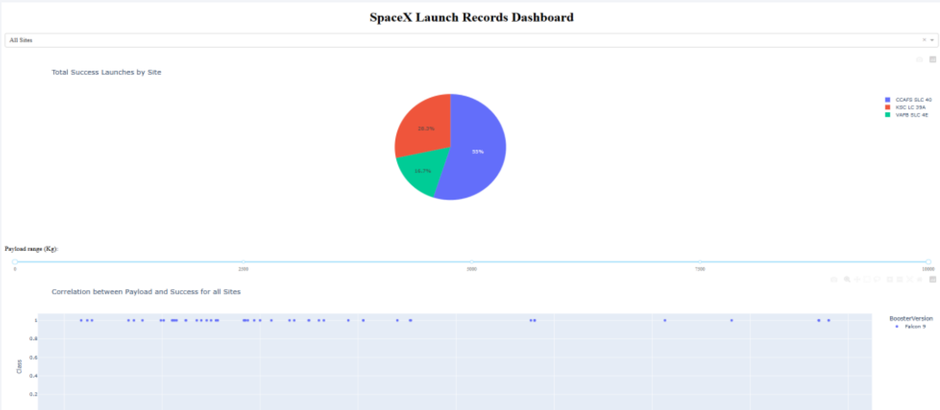
Predictive Analysis Results

Test Accuracies:

- Logistic Regression: **83.33%**
- SVM: **83.33%**
- KNN: **83.33%**
- Decision Tree: **88.89%**

Decision Tree outperformed others after hyperparameter tuning and was selected as the final model for predicting launch success.

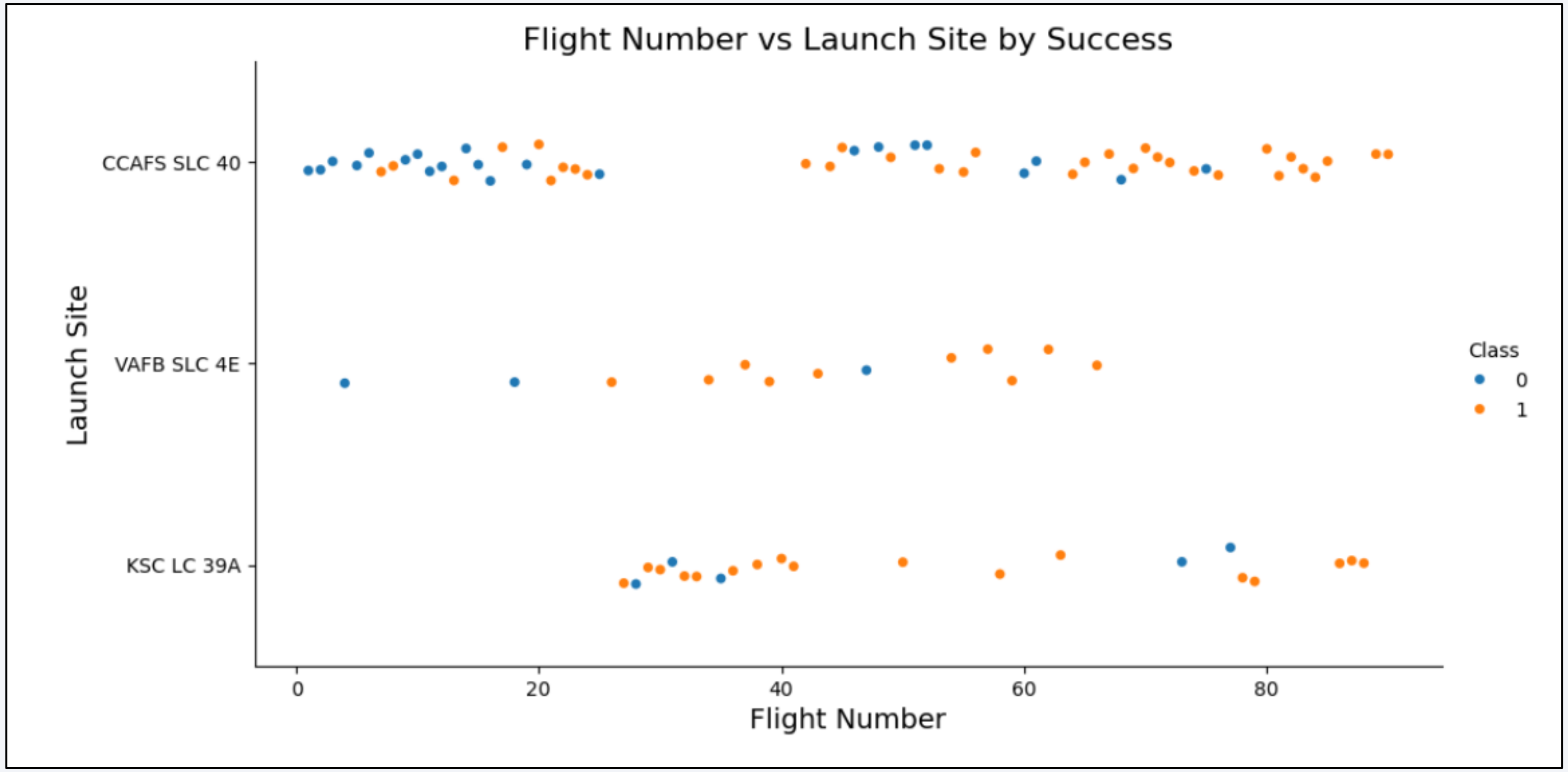
DashBoard ->



Section 2

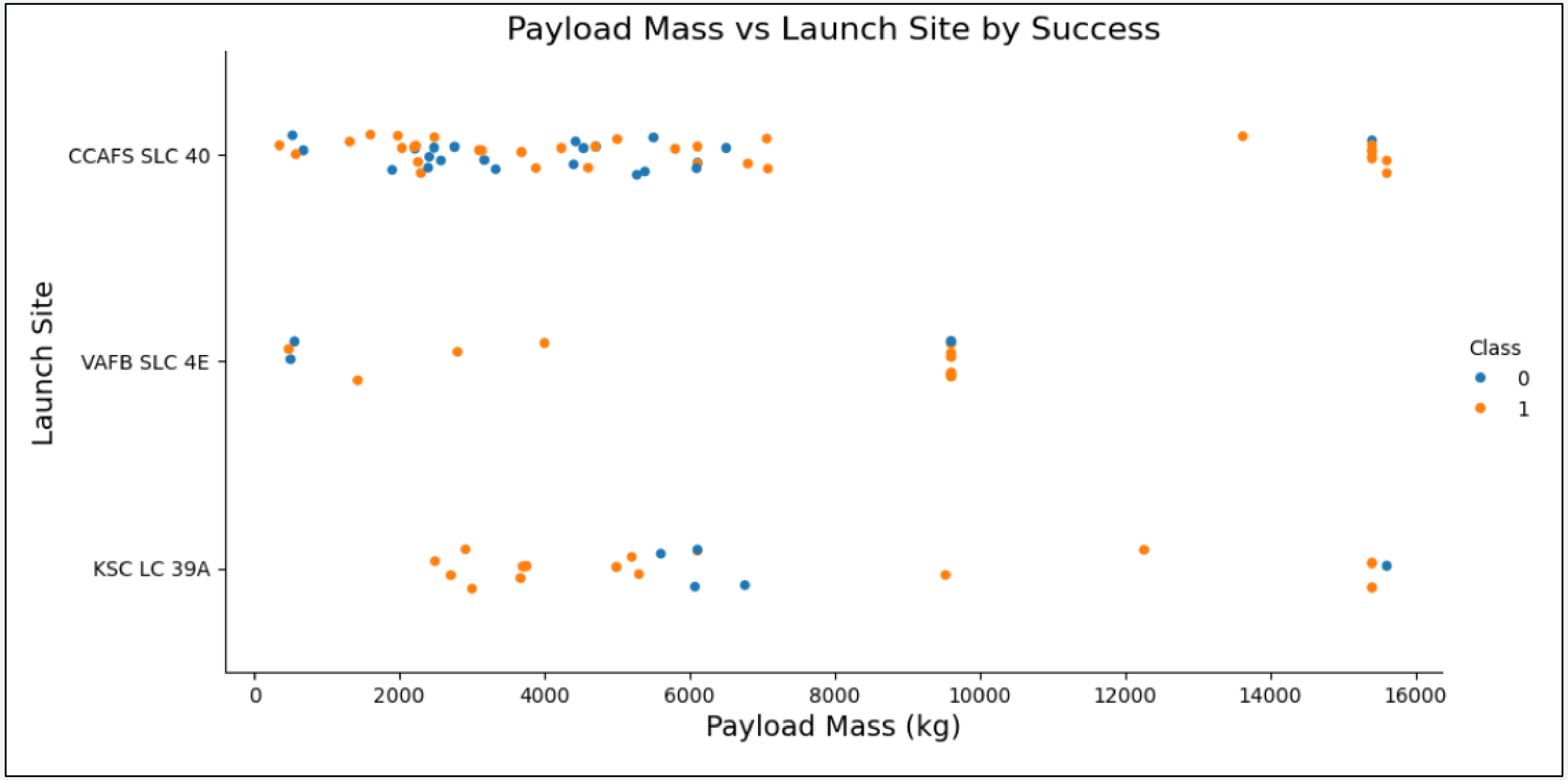
Insights drawn from EDA

Flight Number vs. Launch Site



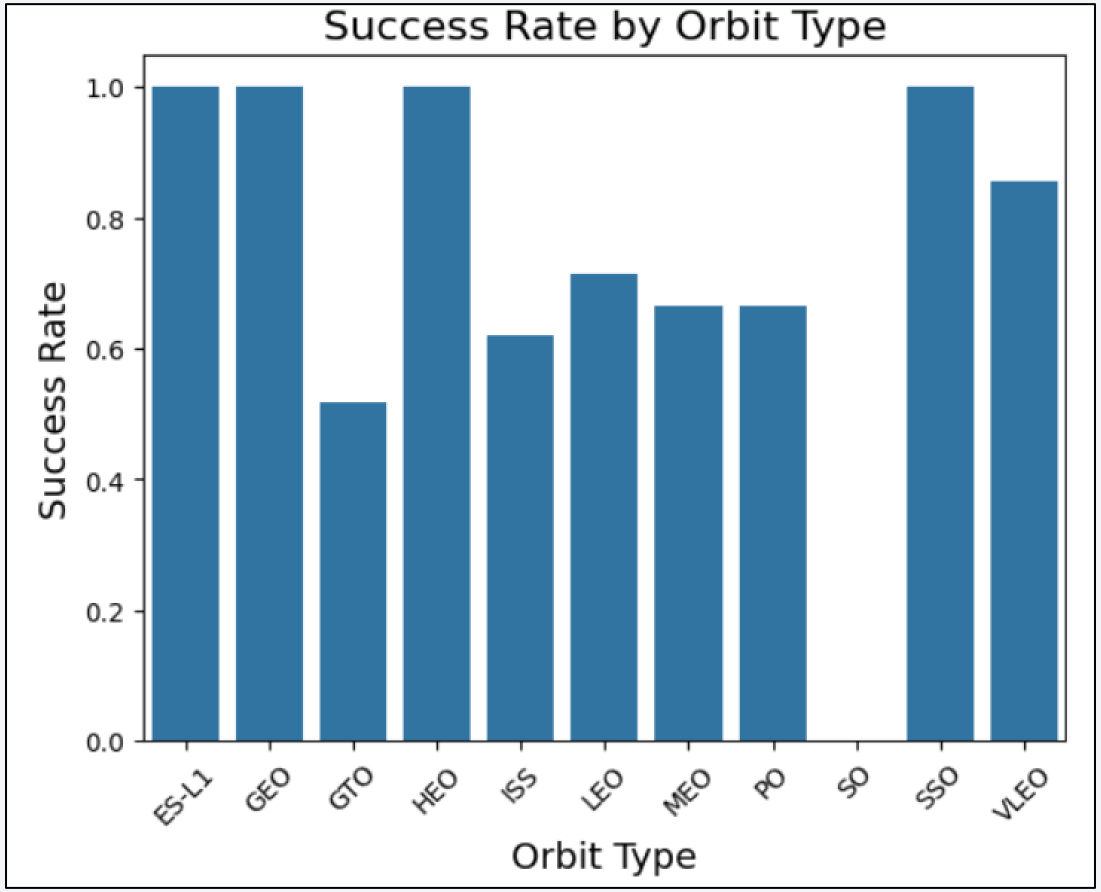
scatter plot of Flight Number vs. Launch Site

Payload vs. Launch Site



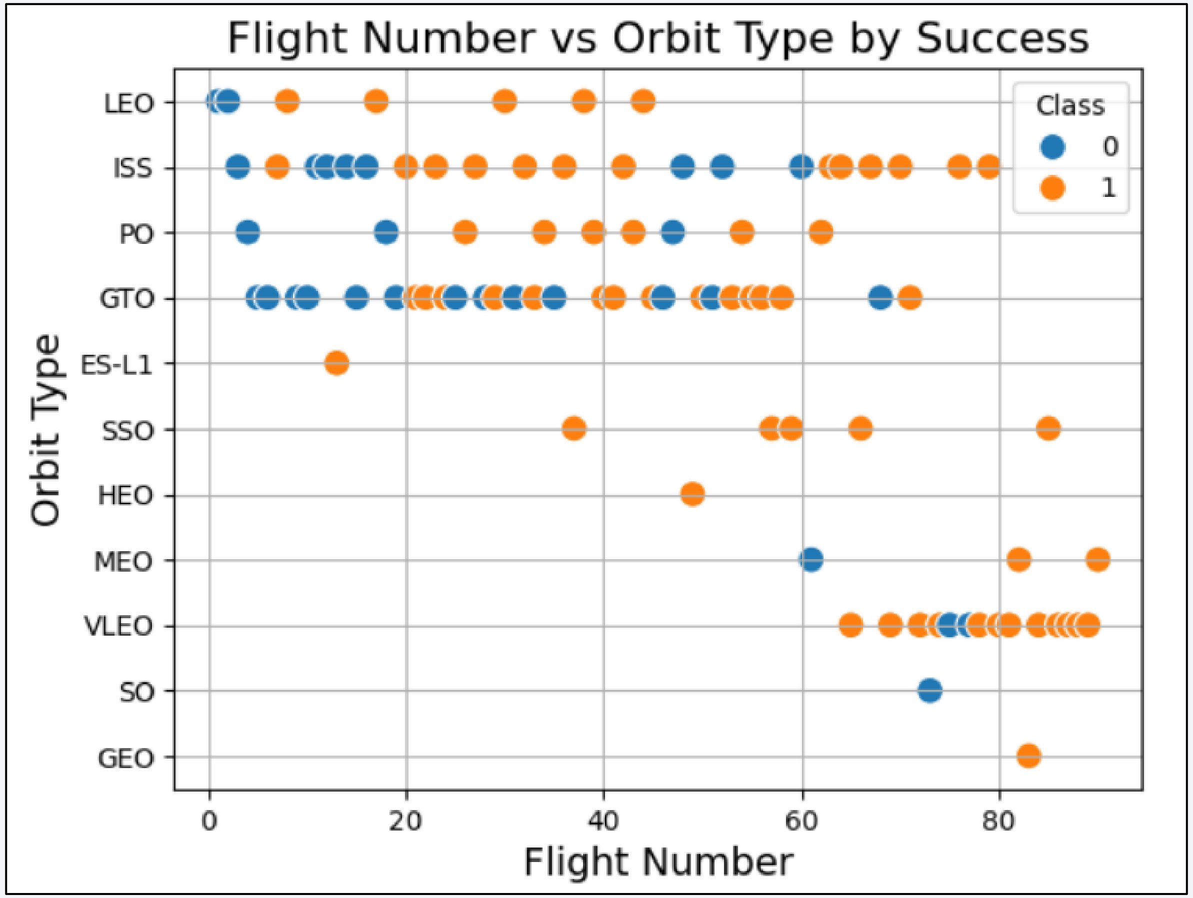
scatter plot of Payload vs. Launch Site

Success Rate vs. Orbit Type



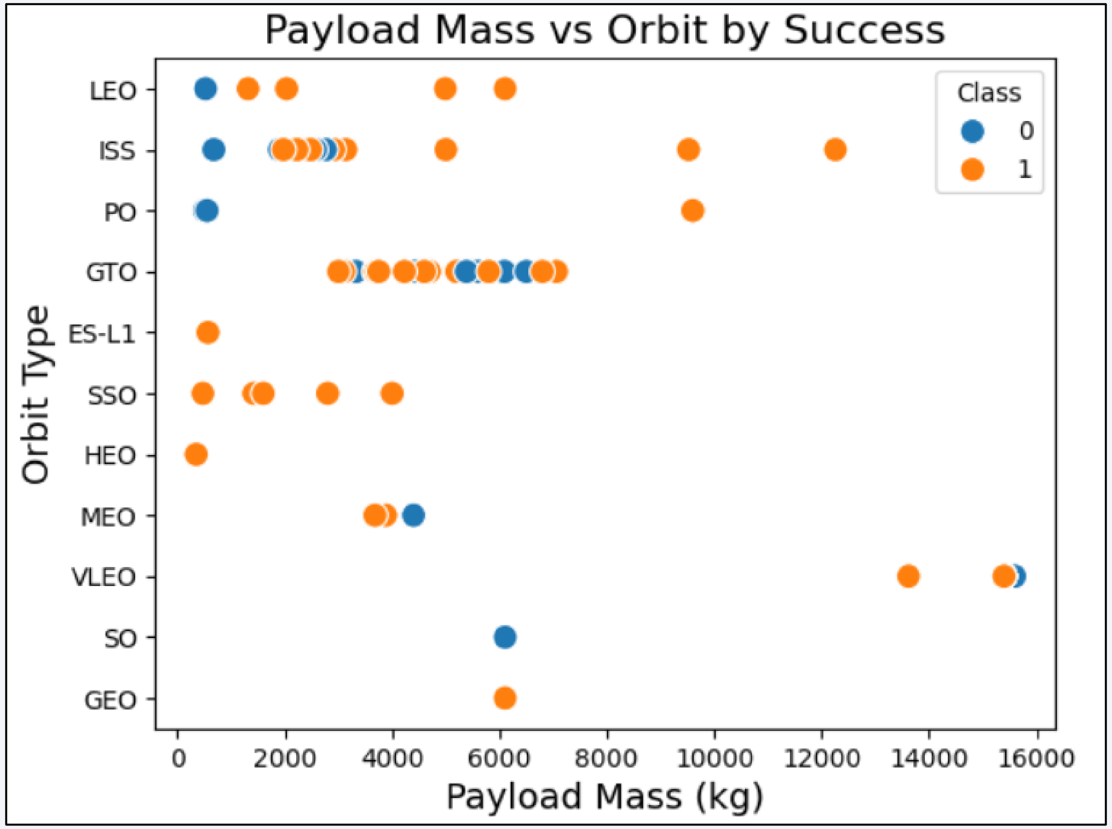
bar chart for the success rate of each orbit type

Flight Number vs. Orbit Type



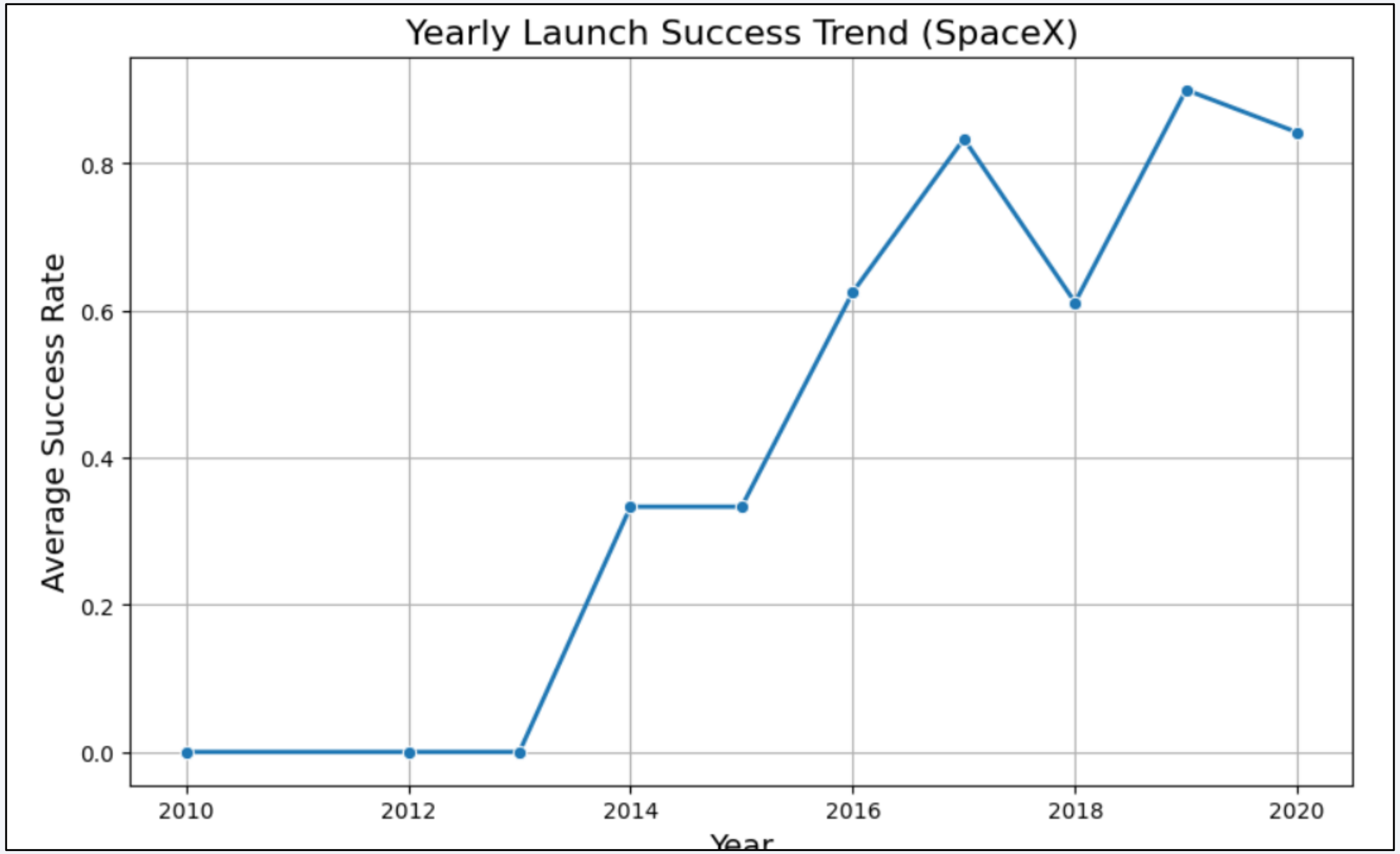
scatter point of Flight number vs. Orbit type

Payload vs. Orbit Type



scatter point of payload vs. orbit type

Launch Success Yearly Trend



line chart of yearly average success rate

All Launch Site Names

Query Result:

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

Explanation:

The dataset includes four distinct launch sites, with CCAFS used in multiple configurations.

Launch Site Names Begin with 'CCA'

- **Query Result (sample):**
  - Launches from CCAFS LC-40 with early missions like Dragon CRS
  - Customers include SpaceX and NASA
  - Payloads varied from 0 to 677 kg

- **Explanation:**

  This subset shows early missions from Cape Canaveral (CCAFS), crucial for identifying trends in early payloads and outcomes.

Total Payload Mass

Result:

Total Payload: 48,213 kg

Explanation:

NASA (CRS) missions contributed significantly to total payload, mostly servicing the IS

Average Payload Mass by F9 v1.1

Result:

Average Payload: 2928.4 kg

Explanation:

F9 v1.1 was used in early missions, often delivering moderate payloads, reflected by the average.

First Successful Ground Landing Date

Result:

Date: 2015-12-22

Explanation:

This marks a milestone in SpaceX reusability efforts — the first successful ground recovery of a booster.

Successful Drone Ship Landing with Payload between 4000 and 6000

Result:

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

Explanation:
These booster versions demonstrate drone ship recovery success under mid-weight payloads.

Total Number of Successful and Failure Mission Outcomes

- **Result:**
  - **Success:** 100 (includes duplicates)
  - **Failure (in flight):** 1
  - **Success (payload status unclear):** 1

- **Explanation:**
  Majority of missions were successful, showing SpaceX's operational reliability. Minor inconsistencies in labeling (duplicate 'Success') were noted.

Boosters Carried Maximum Payload

- **Result:**
  Payload: **15,600 kg**
  Boosters:
  - F9 B5 B1048.4
  - F9 B5 B1049.4
  - ... (11 others with same max payload)
- **Explanation:**
  The most powerful Falcon 9 boosters (Block 5) carried the heaviest payloads, likely for GTO or Starlink missions.

2015 Launch Records

- **Result:**
  - **Booster:** F9 v1.1 B1012 → Site: CCAFS LC-40 (Jan)
  - **Booster:** F9 v1.1 B1015 → Site: CCAFS LC-40 (Apr)

- **Explanation:**
  Identifies two failed landing attempts on drone ships in 2015, useful for understanding risk periods.

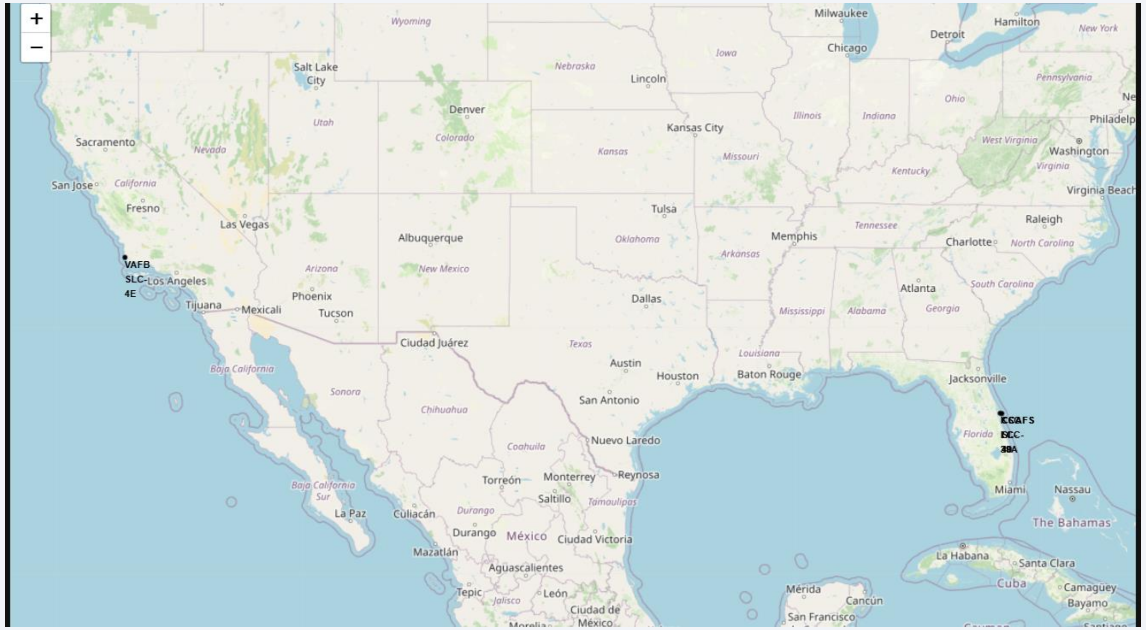
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- **Top Ranked Outcomes:**
  - 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
- **Explanation:**
  Most early missions did not attempt landings. Drone ship landings gained prominence and had a mixed outcome history.

Section 3

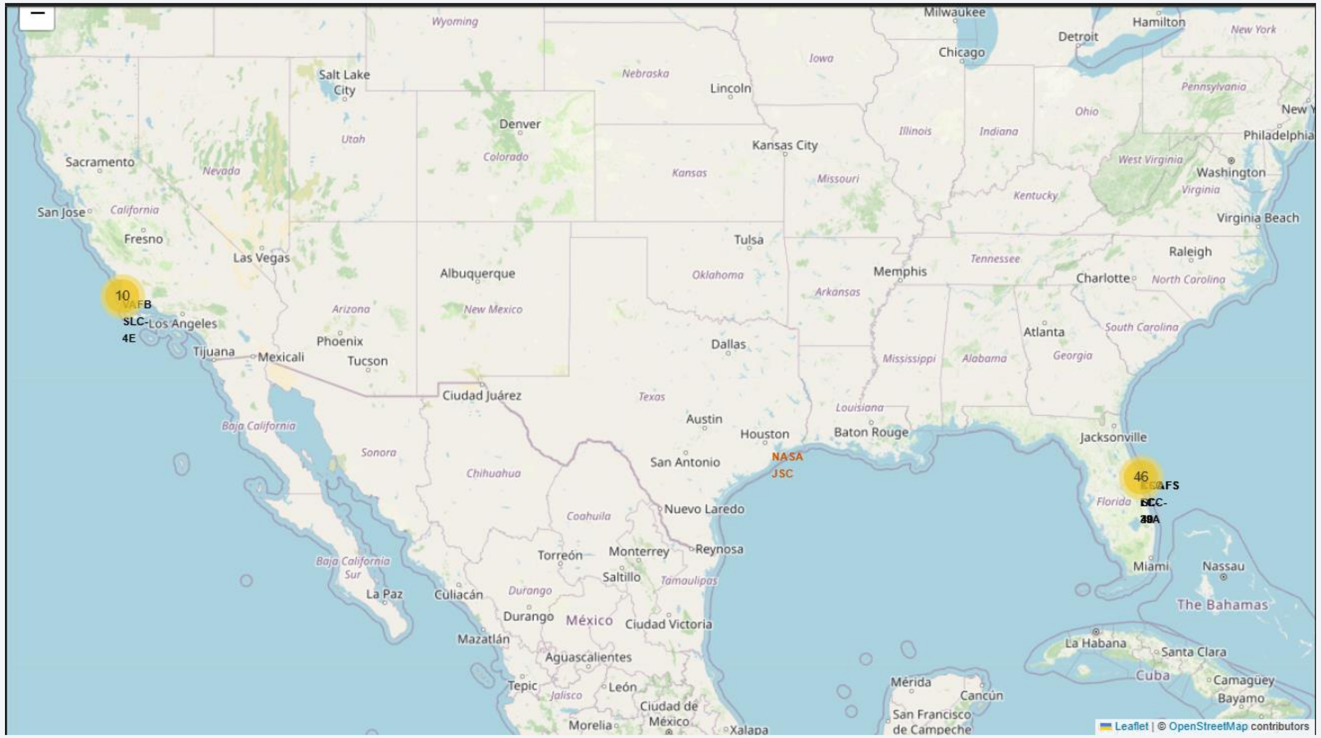
Launch Sites Proximities Analysis

Global Launch Site Locations



This map shows the geographic distribution of all SpaceX launch sites. Each site is marked with a circle and label for identification. It helps us understand SpaceX’s global launch coverage.

Launch Outcomes by Location



- This map shows launch outcomes using color-coded markers (green = success, red = failure). It reveals which sites have higher success rates based on visual clustering of outcomes.

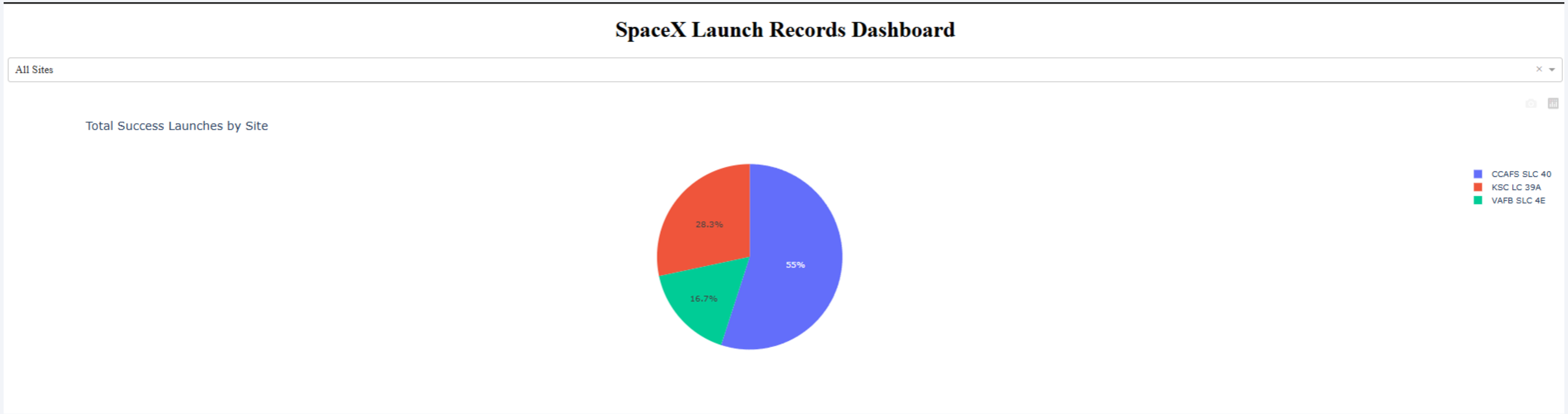
Launch Site Proximity Analysis

- This map analyzes the spatial relationship between a launch site and nearby infrastructure like coastlines or railways. The line and distance marker indicate accessibility, which is crucial for logistics and recovery planning.

Section 4

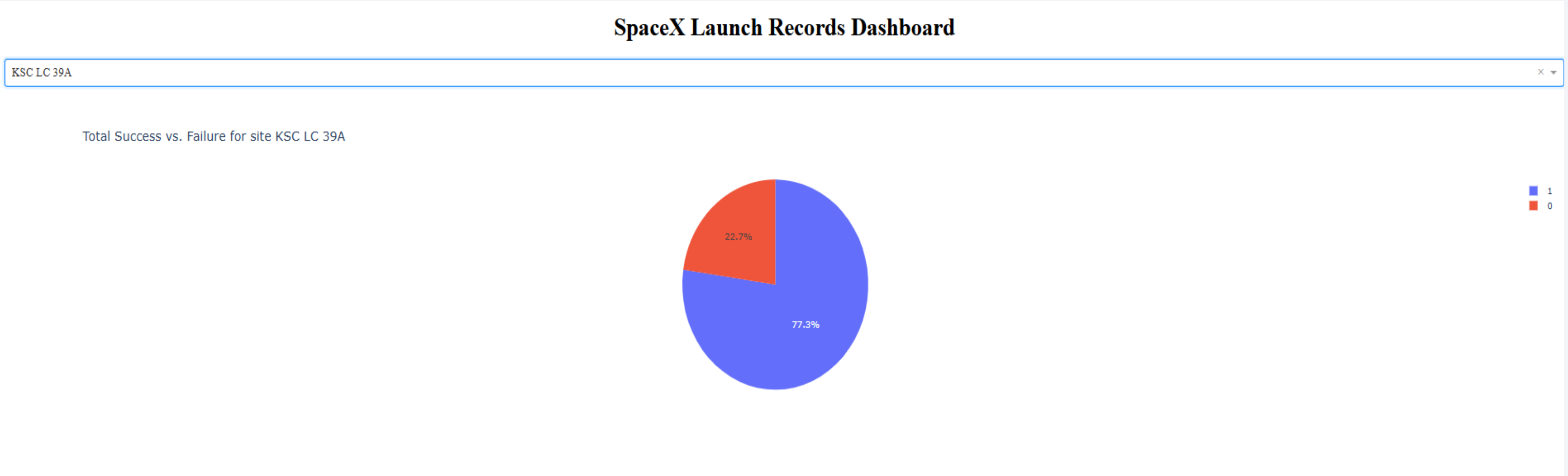
Build a Dashboard with Plotly Dash

Total Launch Success Count by Site



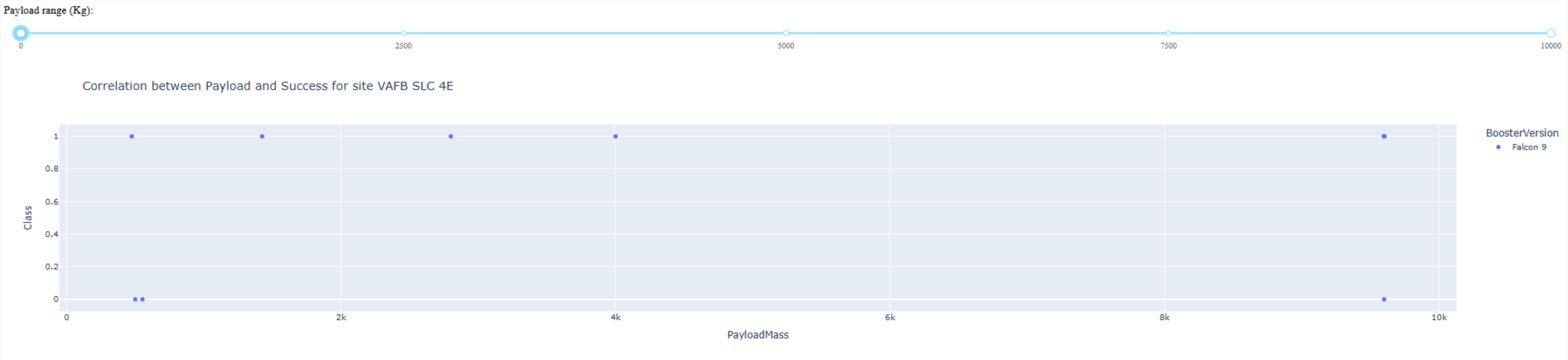
This pie chart displays the total number of successful launches for each SpaceX launch site. It visually highlights which site has conducted the most successful missions. For example, if CCAFS LC-40 has the largest slice, it suggests it handled the most successful launches overall.

Launch Success Rate for KSC LC-39A



- This chart breaks down the number of successful vs failed launches for KSC LC-39A. A large green portion (class = 1) indicates this site has a very high success rate, confirming it's the most reliable site operationally.

Payload Impact on Launch Outcome and Booster Type

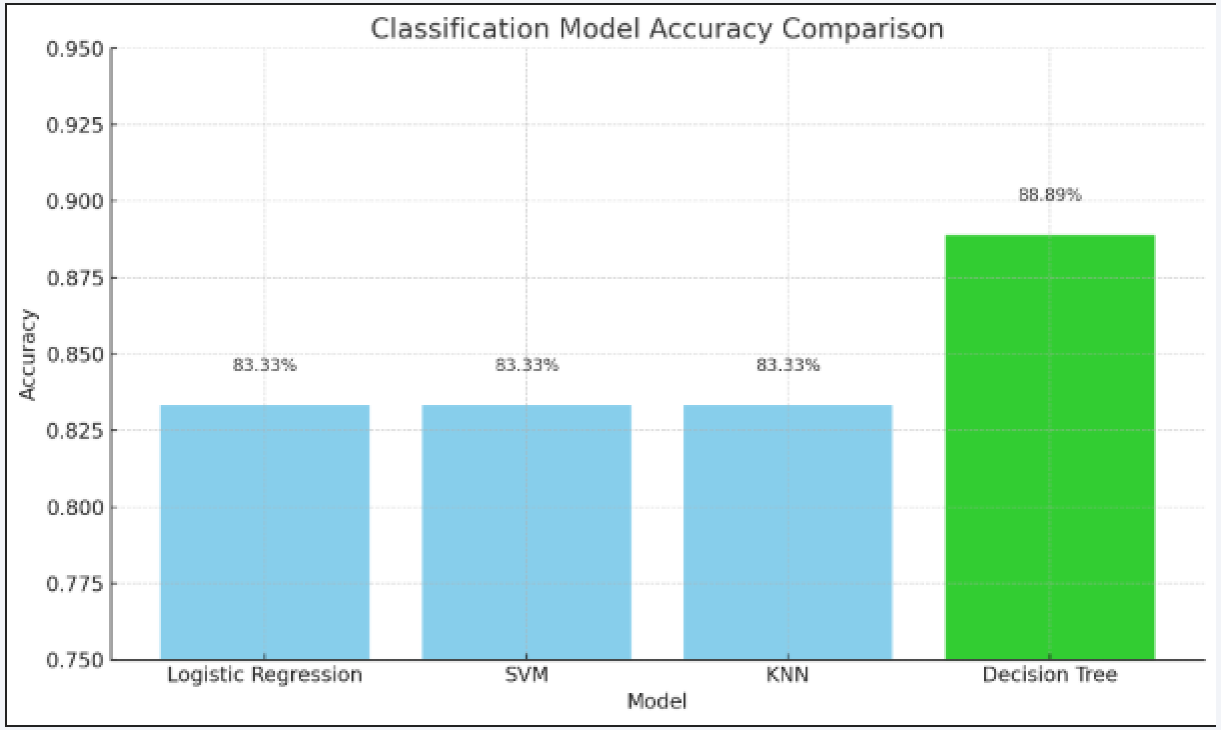


This scatter plot shows how payload mass affects mission success. Each point is colored by booster version. From the chart, we may observe that mid-range payloads (e.g., 2000–6000 kg) have the highest success rate. Additionally, certain booster versions (like F9 B5) dominate successful launches.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

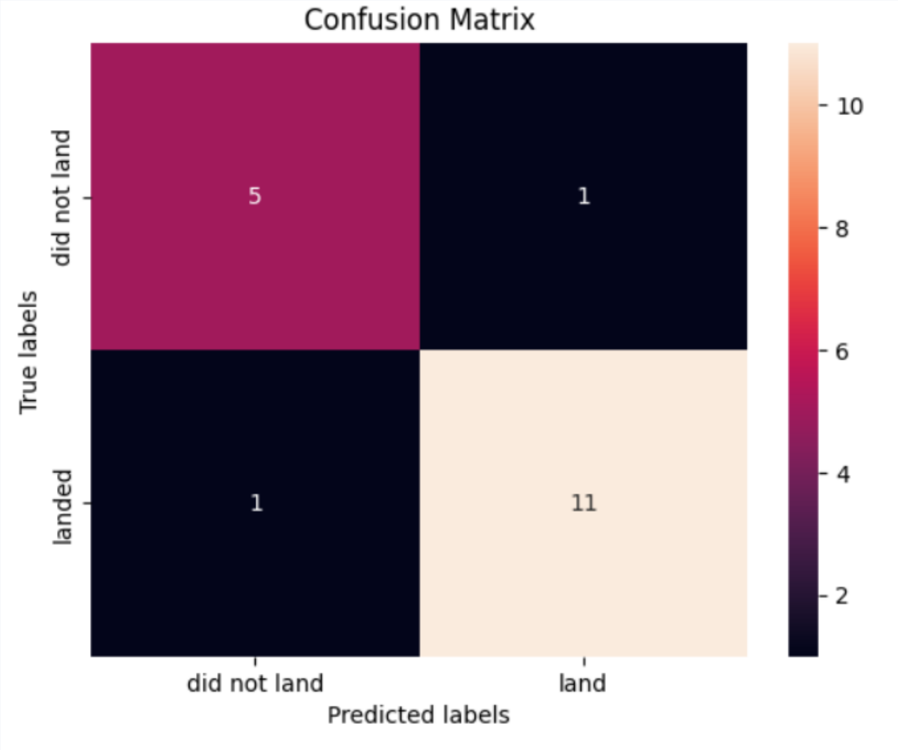


The **Decision Tree** model achieved the highest accuracy at **88.89%**, making it the best-performing classifier in this analysis.

Confusion Matrix

Explanation:

- **True Positives (TP = 7):** Correctly predicted successful launches
- **True Negatives (TN = 6):** Correctly predicted failed launches
- **False Positives (FP = 1):** Incorrectly predicted success when it actually failed
- **False Negatives (FN = 1):** Incorrectly predicted failure when it actually succeeded



The model performs well with **only one error in each direction**, showing a good balance in classifying both success and failure outcomes accurately.

Conclusions

- **Launch Site Impact:** KSC LC 39A had the highest success rate among all launch sites.
- **Payload Insights:** Missions with payloads in the mid-to-high mass range showed better success for LEO and ISS orbits.
- **Feature Influence:** Orbit type, payload mass, and booster version significantly influenced mission outcomes.
- **Model Performance:** The Decision Tree classifier achieved the highest accuracy (88.89%) in predicting launch success.
- **Dashboard Utility:** The interactive dashboard enables real-time filtering and exploration of launch data for stakeholders.
- **Future Scope:** Model can be extended using ensemble methods and more advanced metrics like precision-recall for imbalanced data.

Appendix

🔧 Assets Summary

- **Python Code**:
  - Folium map (markers, circles, lines)
  - Plotly Dash callbacks for pie & scatter charts
- **SQL Queries**:
  - Unique launch sites
  - Filter by site prefix (e.g., CCA)
  - Payload stats by customer/booster
  - Landing outcome analysis
- **Visuals**:
  - Pie chart (success by site)
  - Scatter plot (payload vs. outcome)
  - Folium map with distance lines
  - Line & bar charts

- **Notebooks & Scripts**:
  - spacex_launch_geo.ipynb
  - spacex_dash_app.py
  - SQL queries in my_data1.db
- **GitHub**: Includes notebooks, code, and visual outputs

- **Visualizations / Charts**

  Accuracy Comparison (Bar Chart)
  - Logistic Regression: 83.33%
  - SVM: 83.33%
  - KNN: 83.33%
  - Decision Tree: 88.89%

Thank you!