IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Brandon W.
07/18/2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Exploratory Data Analysis (Visualization & SQL)
  - Interactive Analytics (Folium & Plotly Dash)
  - Predictive Analysis
- Conclusion
- Appendix

Executive Summary

- Goal: To predict the success of SpaceX Falcon 9 first-stage landings to determine launch costs and identify key factors influencing mission outcomes.
- Methodology:
    1. Data Collection: Gathered historical launch data using the SpaceX REST API and by web scraping a Wikipedia page.
    2. Data Wrangling & EDA: Cleaned the data, performed feature engineering, and used data visualization and SQL to uncover initial insights.
    3. Interactive Analytics: Developed an interactive map with Folium and a data dashboard with Plotly Dash for dynamic data exploration.
    4. Predictive Modeling: Built and evaluated four classification models (Logistic Regression, SVM, Decision Tree, K-Nearest Neighbors) to find the best predictor of landing success.

Introduction

Project Background:
SpaceX has revolutionized the space industry by reusing the Falcon 9 rocket's first stage, drastically reducing launch costs from over $165M to around $62M. The ability to predict if the first stage will land successfully is crucial for determining the final cost of a launch.

Key Questions:
1. What factors, such as payload mass, launch site, and orbit, have the greatest impact on landing success?
2. How has the success rate of first-stage landings evolved over the years?
3. Which machine learning model provides the most accurate predictions for landing success based on the available data?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API: Programmatically requested historical launch data.
  - Web Scraping: Extracted Falcon 9 launch records from a Wikipedia page using BeautifulSoup.
- Perform data wrangling
  - Handled missing values, particularly in the PayloadMass column, by imputing the mean.
  - Filtered the dataset to include only Falcon 9 launches.
  - Created the target variable Class by converting landing outcomes into binary labels (1 for success, 0 for failure).
  - Applied one-hot encoding to categorical features like Orbit, LaunchSite, and Serial to prepare the data for machine learning.

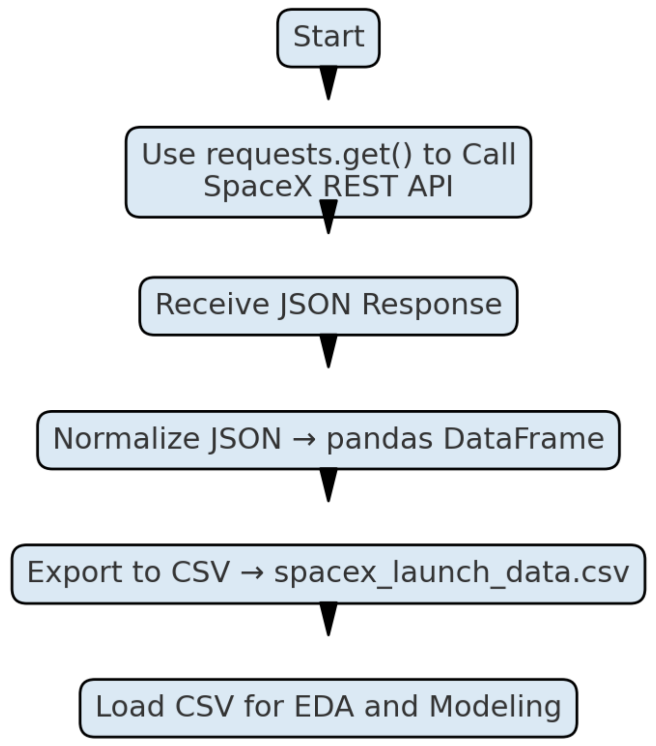
Data Collection

- The data for this project was collected through multiple methods:
  - REST API calls to SpaceX's official API to obtain structured launch data including mission outcome, booster version, payload mass, and orbit.
  - Web scraping from Wikipedia using BeautifulSoup to obtain additional historical SpaceX launch records.
  - CSV file loading for preprocessed data, such as payload and booster performance metrics.
- These methods allowed for comprehensive data acquisition, combining real-time, web-scraped, and static resources to create a robust dataset.

REST API Call to SpaceX API → Web Scraping with BeautifulSoup from Wikipedia → Data Saved to CSV / DataFrames → Combine & Clean Data Sets → Final Merged Dataset Ready for EDA & Modeling

Data Collection – SpaceX API

- The SpaceX REST API (https://api.spacexdata.com/v4/launches) was used to retrieve launch data.
- We used Python’s requests library to perform API GET calls and fetch structured JSON data.
- The JSON data was normalized and converted into pandas DataFrames.
- This data was stored locally as CSV files and used throughout the project for analysis and modeling.



https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/jupyter-labs-spacex-data-collection-api_complete.ipynb

Data Collection - Scraping

- Web Scraping Overview
- Target: Wikipedia page listing Falcon 9 and Falcon Heavy launches.
  - Sent an HTTP GET request using Python's requests library.Parsed HTML using BeautifulSoup to locate the table containing mission records.
  - Extracted launch details such as flight number, date, site, and outcome.
  - Cleaned and transformed extracted rows into a pandas DataFrame.Exported to CSV for use in EDA and modeling stages.

[Send HTTP GET request to Wikipedia launch data page]

↓

[Parse HTML content using BeautifulSoup]

↓

[Extract table rows with launch details]

↓

[Convert extracted data into pandas DataFrame]

↓

[Clean & Save to CSV for further analysis]

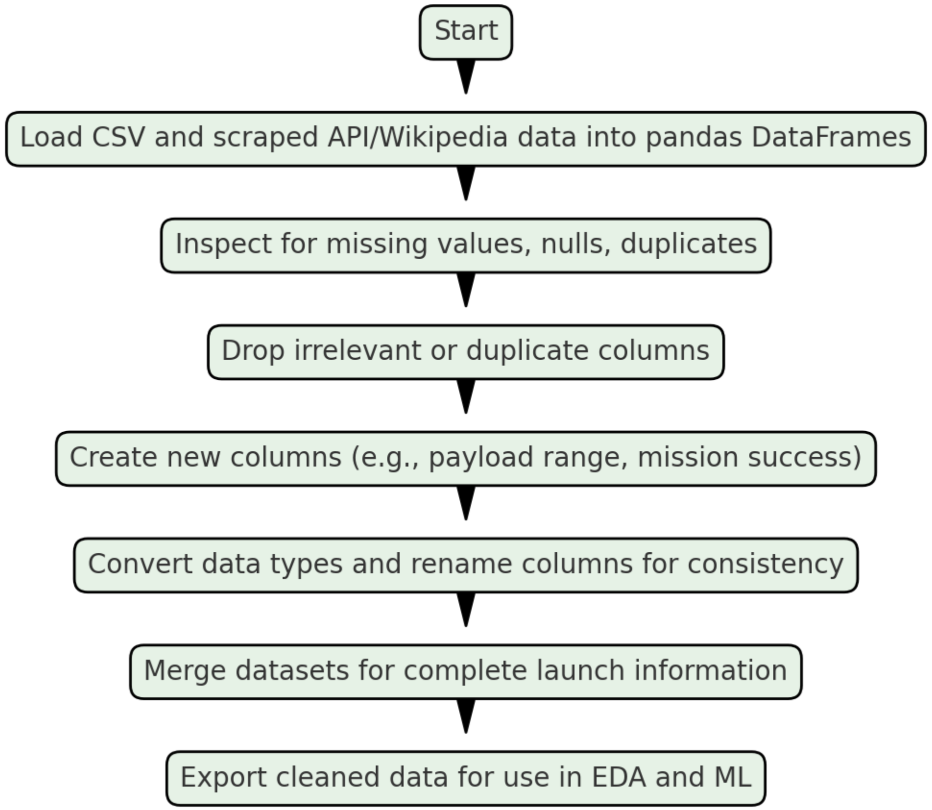
b scraping here

https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/jupyter-labs-webscraping_complete.ipynb

Data Wrangling

Data Wrangling Process

- Loaded launch datasets collected via SpaceX API and Wikipedia scraping.
- Used pandas to inspect for null values, duplicated entries, and inconsistent column names.
- Dropped unnecessary columns and handled missing data appropriately.
- Engineered new features such as:
  - **Launch success flag** (binary)
  - **Payload categories**
  - **Normalized site and booster names**
- Merged cleaned datasets into a master DataFrame for analysis.
- Final data was saved to CSV and used in EDA and machine learning models.



https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/labs-jupyter-spacex-Data%20wrangling_complete.ipynb

EDA with Data Visualization

Charts Created & Purpose:

•**Flight Number vs. Launch Site (Scatter Plot):**
Showed trends in launch success over time across sites.
•**Payload Mass vs. Launch Site (Scatter Plot):**
Compared payload capacities across different launch locations.
•**Success Rate by Orbit Type (Bar Chart):**
Identified which orbit types had higher mission success.
•**Flight Number vs. Orbit Type (Scatter Plot):**
Tracked how orbit types evolved with SpaceX's mission growth.
•**Payload vs. Orbit Type (Scatter Plot):**
Assessed payload weight distribution by orbit category.
•**Yearly Success Trend (Line Chart):**
Illustrated SpaceX's improvement in reliability over years.

https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/eda-dataviz_complete.ipynb

EDA with SQL

Query	Purpose
Distinct Launch Sites	Identify all unique SpaceX launch locations
Launch Sites Starting with 'CCA'	Filter missions from Cape Canaveral
Total Payload by NASA	Sum payload mass where customer is NASA
Average Payload – F9 v1.1	Compute average payload for specific booster
First Successful Ground Pad Landing	Get first date of successful ground landing
Successful Drone Ship Landings (4k–6k kg)	Find heavy payloads with successful drone landings
Total Mission Outcomes	Count all successful and failed missions
Max Payload Booster	Identify boosters that carried max payload
Failed Drone Ship Landings – 2015	List failed landings by site and booster in 2015
Ranked Landing Outcomes (2010–2017)	Rank outcome types by frequency over a time window

https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/jupyter-labs-eda-sql-coursera_complete.ipynb

Build an Interactive Map with Folium

Map Objects Added

1. **Markers:**
   - Added for each of the four SpaceX launch sites (e.g., CCAFS, KSC, VAFB) using their coordinates.
   - Purpose: Identify exact launch site locations on the global map.
2. **Circle Markers:**
   - Colored and sized based on success rate or activity at each site.
   - Purpose: Quickly visualize mission density or outcome intensity.
3. **Popups:**
   - Custom HTML popups show detailed info (site name, coordinates, proximity measurements).
   - Purpose: Provide interactive, on-click site insights.
4. **Lines:**
   - Drawn from each launch site to the nearest infrastructure points (e.g., highways, coasts, rail).
   - Purpose: Show proximity to logistical elements affecting operations.
5. **Distance Labels:**
   - Calculated and displayed distances from each site to key landmarks.
   - Purpose: Assist with analyzing accessibility and location-based advantages.

https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/lab_jupyter_launch_site_location_complete.ipynb

Build a Dashboard with Plotly Dash

Plots & Interactions Added:

•**Launch Site Dropdown (Dropdown Filter):**

- •Allows users select a specific site or view all sites
- •Enables filtering of visualizations dynamically based on launch location.

•**Pie Chart (Success Metrics):**

- For **All Sites**: shows proportion of successful launches by site.
- For **Single Site**: shows Success vs. Failure.
  - Helps compare launch success rates across locations interactively.

•**Payload Range Slider (RangeSlider):**

- •Allows users to filter data based on payload weight (in kilograms).
- •Provides custom control over the weight spectrum being analyzed.

•**Scatter Plot (Payload vs. Success):**

- •Displays payload mass vs. mission success, color-coded by Booster Version Category.
- •Helps explore how different payloads and boosters influence success outcomes.

Build a Dashboard with Plotly Dash (continued)

Why These Features Were Added:

1. To provide interactivity and exploration for users analyzing SpaceX data.
2. To visualize key performance metrics (success/failure) across multiple filters.
3. To make trends easier to interpret, such as payload influence or site effectiveness.

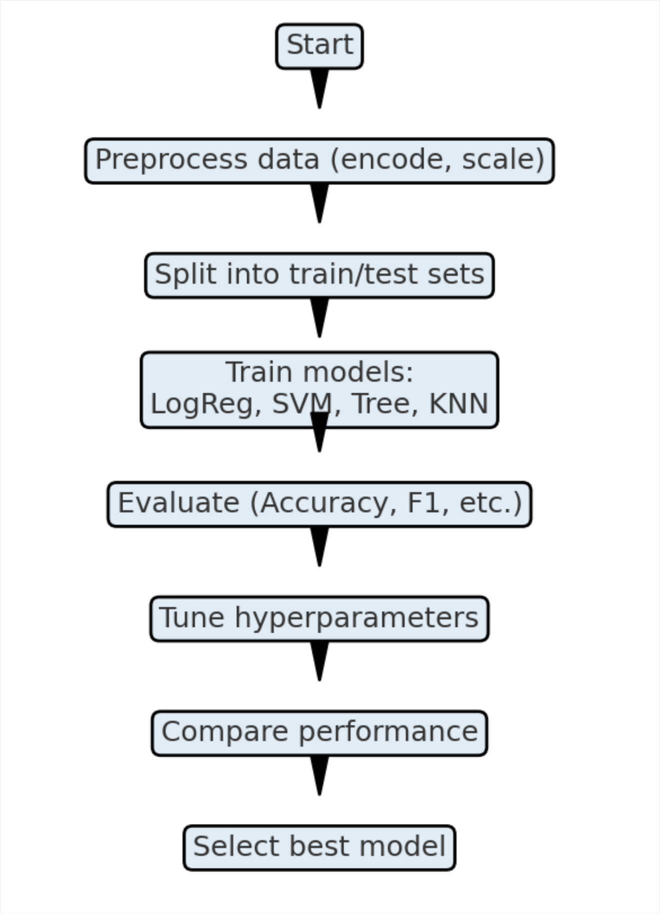
Built using Dash, Plotly Express, and pandas for live, responsive charts.

https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/spacex-dash-app_Complete.py

Predictive Analysis (Classification)

Summary of Model Development:

- **Data Preprocessing:**
  - Encoded categorical variables and normalized features.
- **Train-Test Split:**
  - Divided the data into training (80%) and testing (20%) sets.
- **Model Training:**
  - Trained and compared the following classifiers:
    - Logistic Regression
    - Support Vector Machine (SVM)
    - Decision Tree
    - K-Nearest Neighbors (KNN)
- **Model Evaluation:**
  - Assessed using Accuracy, Precision, Recall, and F1 Score.
- **Hyperparameter Tuning:**
  - Tuned max_depth (Decision Tree) and C (SVM, Logistic) for better performance.
- **Best Model Selection:**
  - Chose the model with the highest overall metrics on test data.

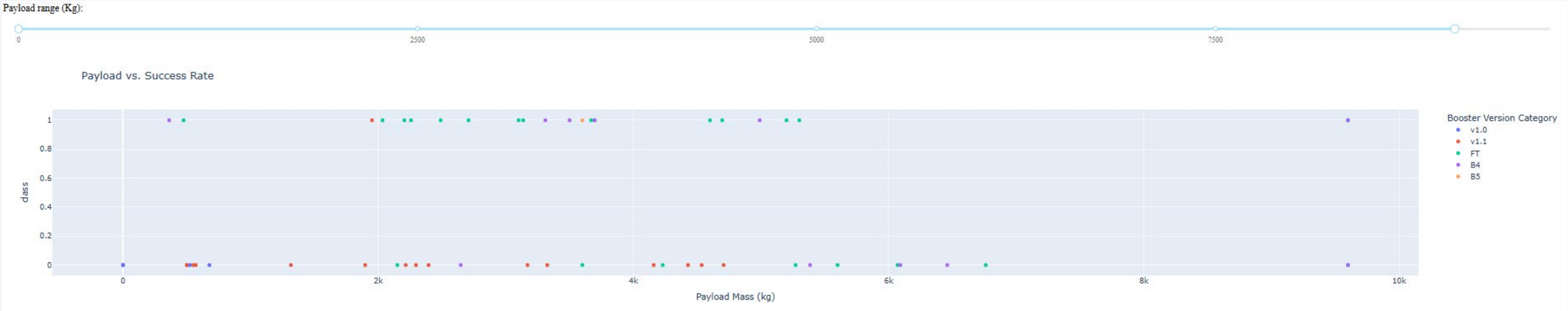


https://github.com/Weintraub-B/Portfolio/blob/main/IBM-data-science-projects/Project5_Capstone/Completed-Notebooks-PythonFiles/SpaceX_Machine%20Learning%20Prediction_Part_5_complete.ipynb

Results – Data Analysis Result

- **Exploratory Data Analysis (EDA) Results**
  - Found that payload mass, booster version, and launch site significantly influenced launch outcomes.
  - Used heatmaps and correlation matrices to detect feature relationships.
  - Observed imbalance in class labels — mostly successful launches — requiring careful evaluation during classification.
  - Created scatter plots and box plots to visualize payload vs. success and site vs. mission outcome.

Results – Analytics Demo



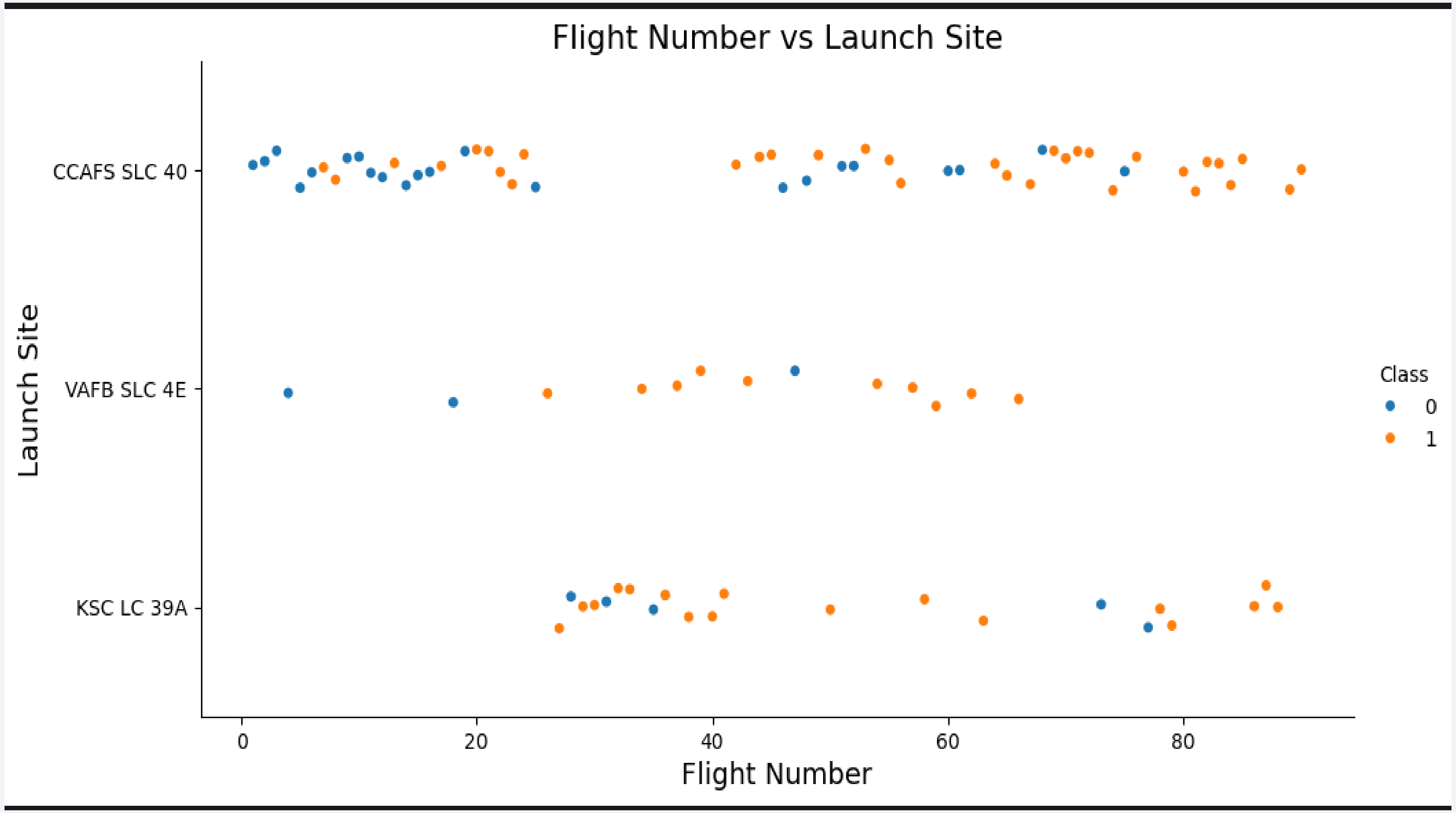
Results – Predictive Analysis Results

- Models trained: Logistic Regression, Support Vector Machine (SVM), Decision Tree, KNN.
- Evaluation metrics used: Accuracy, Precision, Recall, F1 Score.
- Best Model: Decision Tree with tuned depth (max_depth=3) yielded highest accuracy and F1-score.
- Used confusion matrix to analyze false positives/negatives and validate model quality.

Section 2

Insights drawn from EDA

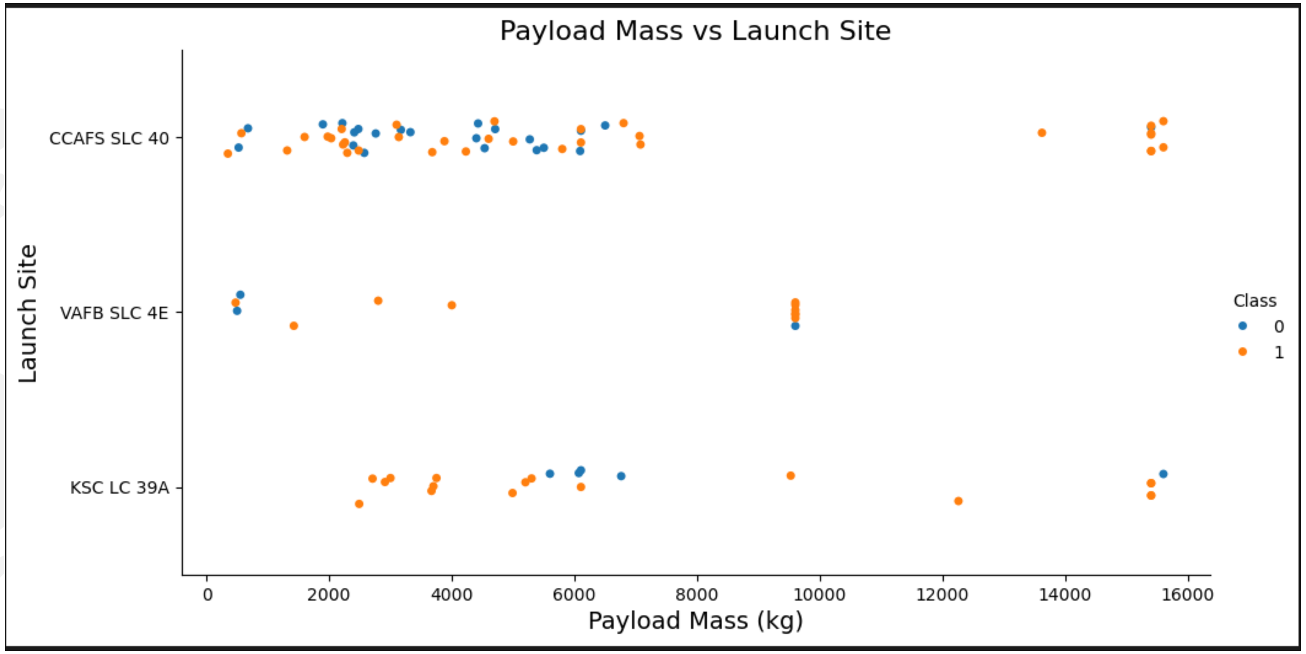
Flight Number vs. Launch Site



- **Flight Number vs. Launch Site – Scatter Plot**
- **What the Plot Shows:**
- Each dot represents a SpaceX launch.
- **X-axis:** Flight number (chronological order).
- **Y-axis:** Launch site (e.g., CCAFS SLC 40, VAFB SLC 4E, KSC LC 39A).
- **Color-coded by outcome:**
  - Orange = Successful launch
  - Blue = Failed launch
- **Key Insights:**
- CCAFS SLC 40 has the highest number of launches, showing improved success over time.
- KSC LC 39A shows a mix of early failures and later consistent successes.
- VAFB SLC 4E has fewer launches with scattered results.
- **Why This Matters:**
- Highlights site performance over time.
- Helps identify reliability trends tied to specific launch locations.

Payload vs. Launch Site

- **Purpose of Plot**:
Visualize how payload mass affects mission outcomes at different SpaceX launch sites.
    - **X-axis**:
    Payload Mass (kg)
    - **Y-axis**:
    Launch Sites
    (CCAFS SLC 40, VAFB SLC 4E, KSC LC 39A)
- **Color Legend**:
    - **Orange (1)**: Successful Launch
    - **Blue (0)**: Failed Launch
- **Key Observations**:
    - Heavier payloads (above 10,000 kg) are mostly launched from **KSC LC 39A** and **CCAFS SLC 40**.
    - **KSC LC 39A** shows consistently high success rates regardless of payload size.
    - **VAFB SLC 4E** has a more limited range of payloads, with moderate success.
- **Conclusion**:
    - This visualization supports identifying which launch sites handle heavier payloads more successfully and highlights the relationship between launch site and mission outcome.



Success Rate vs. Orbit Type

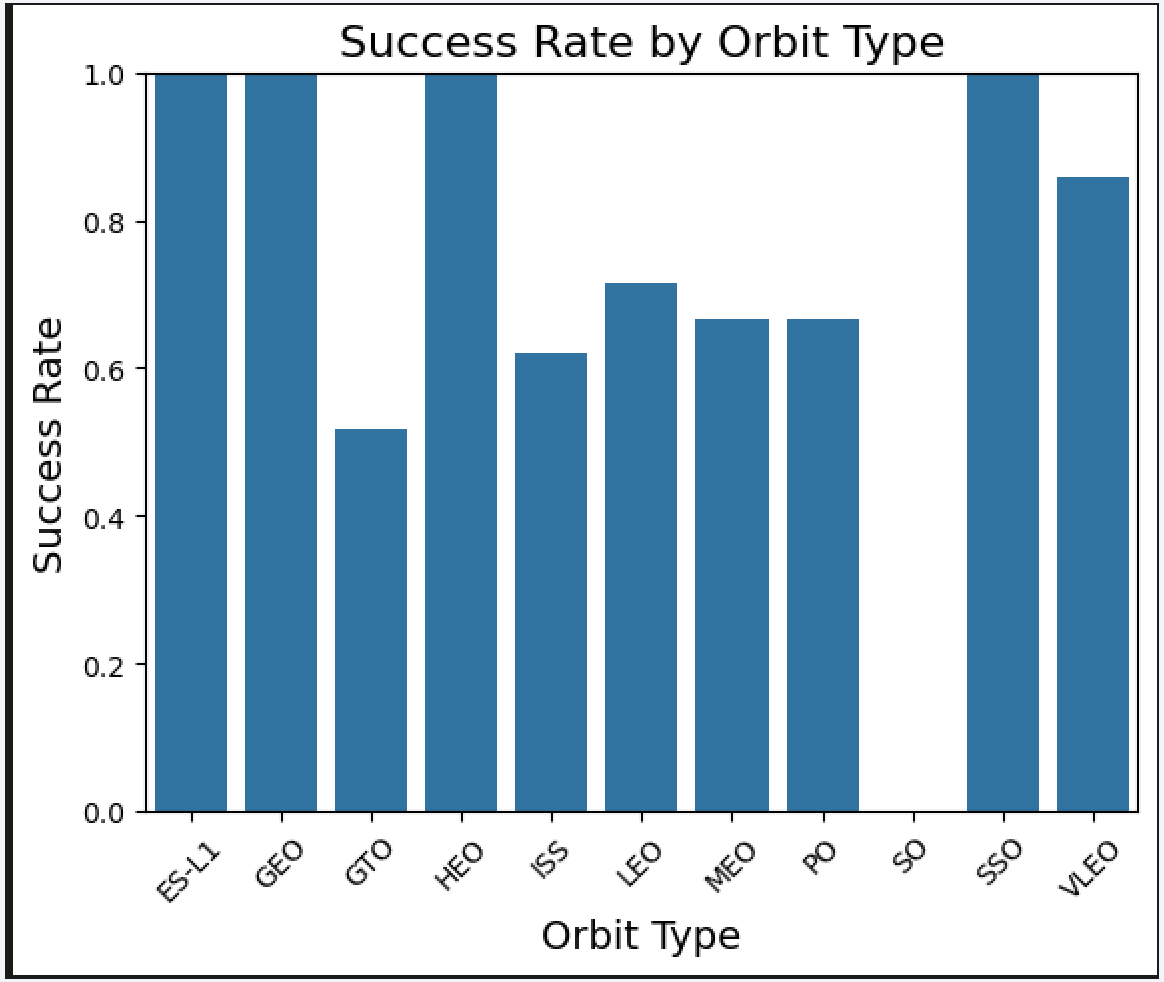
- **Purpose of Chart:**

Analyze how SpaceX launch success varies by orbital destination.

- X-axis: Orbit Types (e.g., LEO, GTO, ISS, SSO)
- Y-axis: Launch Success Rate (0 = failure, 1 = success)

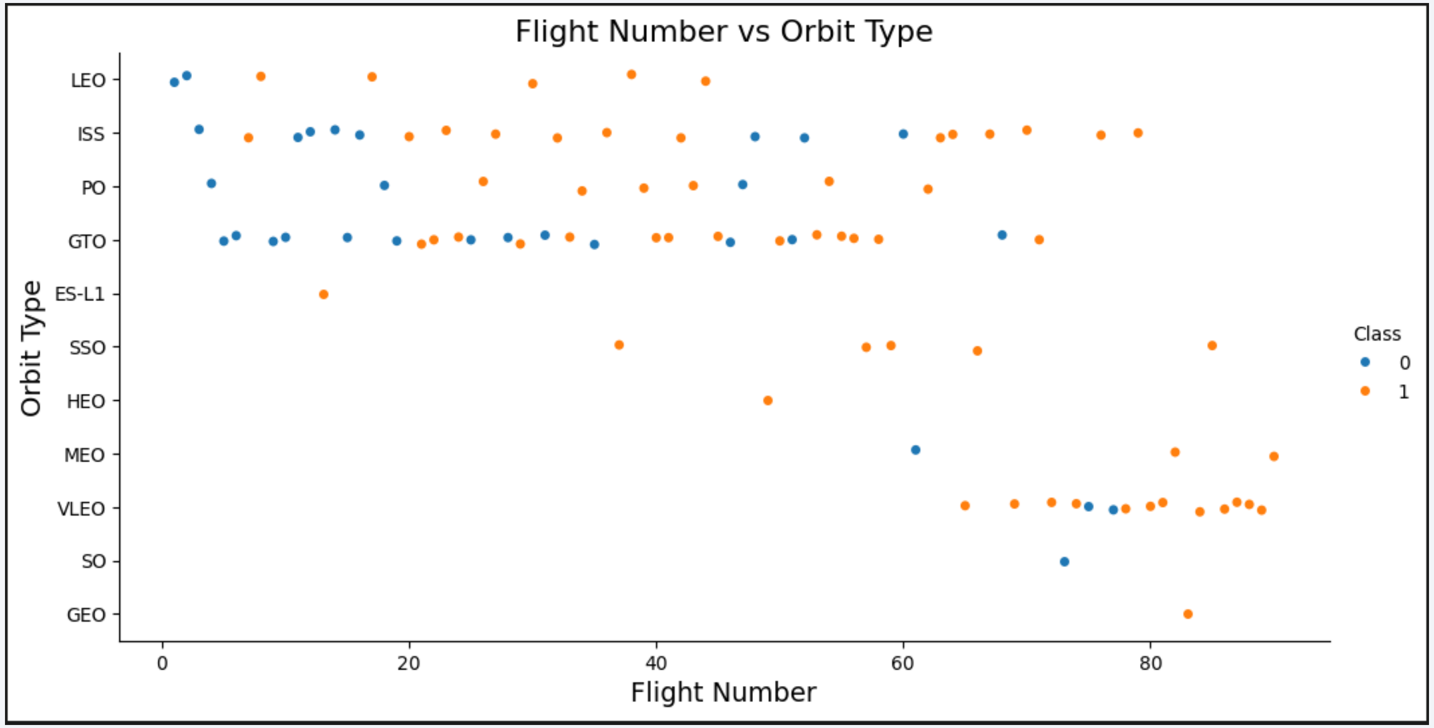
- Key Observations:
  - GEO, HEO, SSO, and ES-L1 orbits show a 100% success rate, indicating highly reliable mission performance.
  - GTO (Geostationary Transfer Orbit) has the lowest success rate, around 53%, highlighting greater mission risk or complexity.
  - Most other orbits (e.g., LEO, MEO, ISS) maintain moderate success rates between 65% and 85%.

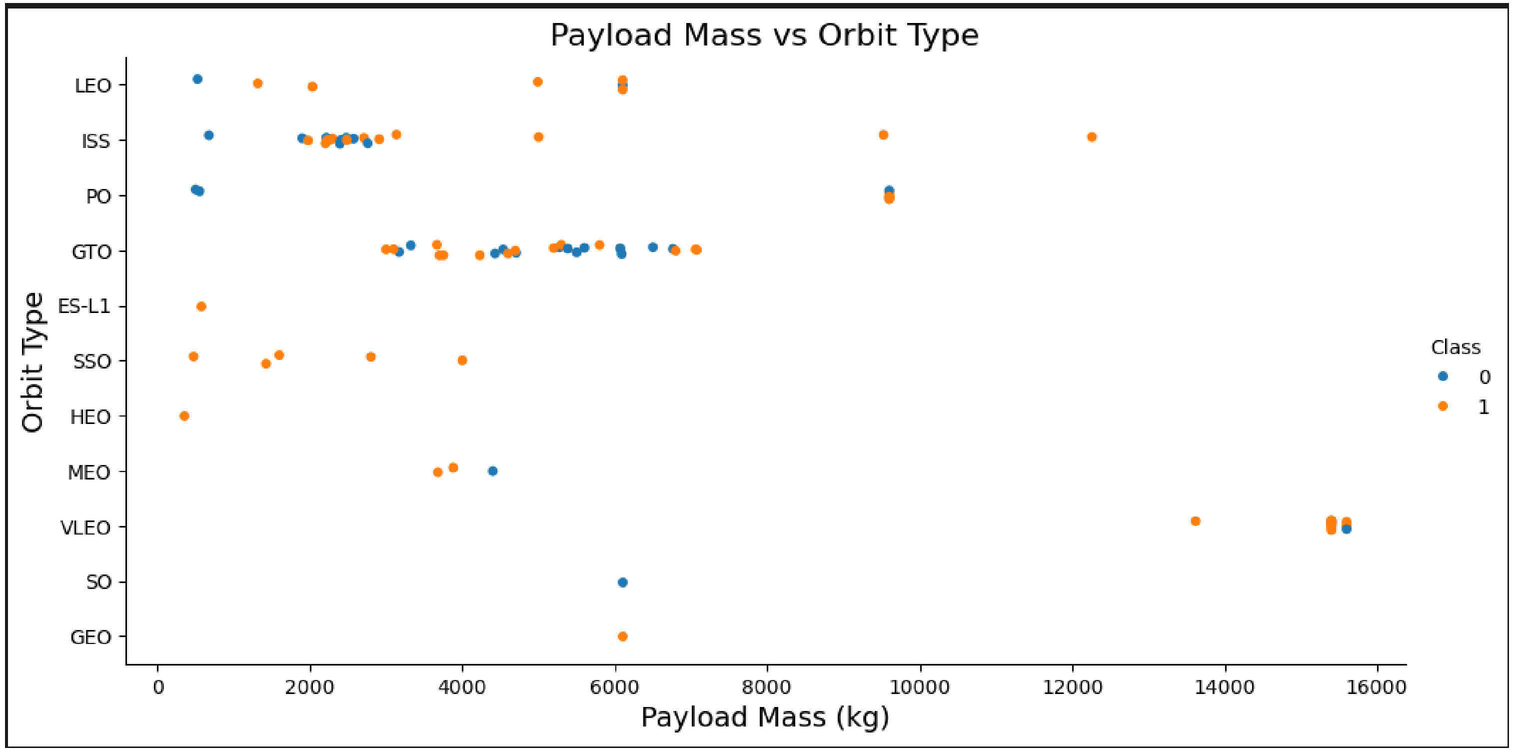
- Conclusion:
  - SpaceX has demonstrated strong performance for missions targeting higher Earth orbits and sun-synchronous orbits. GTO missions may require further optimization or carry higher technical risks.



Flight Number vs. Orbit Type

- **Objective**:
  Visualize how mission outcomes (success/failure) vary across different orbit types over time.
- **X-axis**:
  Flight Number (chronological order of launches)
- **Y-axis**:
  Orbit Type (e.g., LEO, GTO, ISS, SSO)
- **Data Points**:
  - **Orange Dots (1)**: Successful launches
  - **Blue Dots (0)**: Failed launches
- **Key Insights**:
  - **GTO** and **VLEO** show a mix of successes and failures, suggesting higher technical risk or complexity.
  - Later flight numbers (higher on X-axis) show more **consistent success** across all orbit types.
  - **LEO**, **ISS**, and **SSO** maintain high success rates, especially in recent missions.
- **Conclusion**:
  SpaceX's success rate improves over time, especially for complex orbits, indicating growing launch reliability and operational maturity.





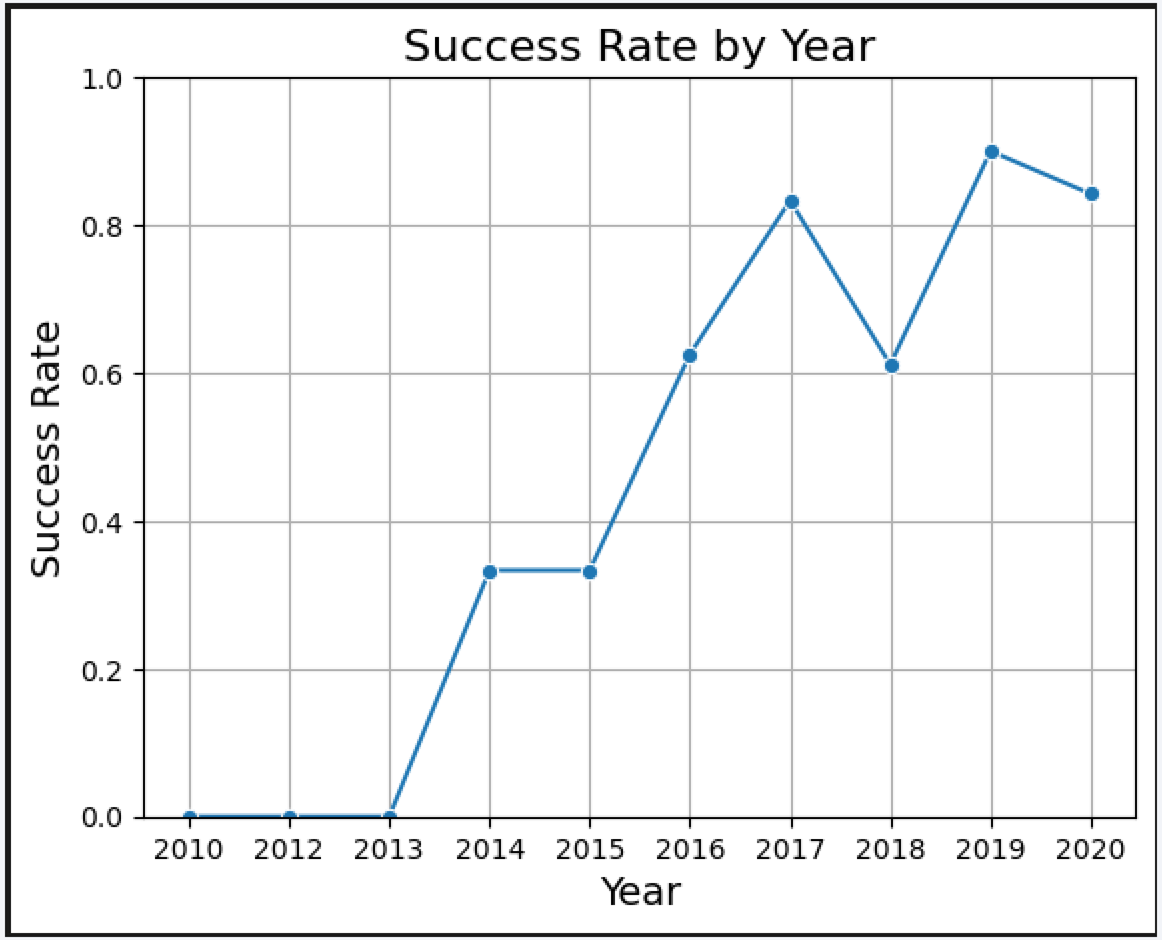
Payload vs. Orbit Type

- **Purpose: Visualize how payload mass correlates with orbit type and launch success.**
    - X-axis: Payload Mass (kg)
    - Y-axis:Orbit Type (e.g., GTO, LEO, ISS, SSO)
    - Data Points:
        - Orange Dots (1): Successful launches
        - Blue Dots (0): Failed launches
- Insights:
    - GTO and VLEO launches handle higher payload masses (~15,000+ kg).
    - LEO and ISS missions mostly carry lighter payloads (<7,000 kg).
    - Higher success rates are observed across all orbit types, even with heavier payloads, especially in later missions.
    - Failures are relatively more common at lower payloads, possibly indicating earlier missions.
- Conclusion:
    - SpaceX successfully launches heavy payloads to complex orbits, with improved outcomes over time, showcasing operational reliability.

Launch Success Yearly Trend

Trend Highlights:

- **2010–2013**: 0% success rate — early development and trial phase.
- **2014**: Success rate jumps to ~33% — beginning of operational launches.
- **2016–2017**: Rapid improvement to over **80%**, reflecting technology maturation.
- **2018**: Slight dip (~60%), possibly due to increased launch frequency or anomalies.
- **2019–2020**: Maintained **high success rates** (85–90%), indicating strong reliability.



All Launch Site Names

SELECT DISTINCT Launch_Site FROM SPACEX

Purpose: To retrieve all distinct launch site entries for further EDA filtering and mapping.

Unique Launch Site Names:

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

Launch Site Names Begin with 'CCA'

```
SELECT *

FROM spacex_df

WHERE "Launch Site" LIKE
'CCA%'

LIMIT 5;
```

	Launch Site	Flight Number		
0	CCAFS LC-40	1		
1	CCAFS SLC-41	2		
3	CCAFS LC-40	4		
5	CCAFS LC-40	6		
7	CCAFS SLC-41	8		

This query filters the spacex_df table to include only rows where the Launch Site starts with "CCA" (e.g., "CCAFS LC-40"). The LIMIT 5 clause ensures only the first five matching records are returned for quick analysis.

Total Payload Mass

```
SELECT SUM(Payload_Mass__kg_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Customer='NASA (CRS)';
```

This query filters all missions with a customer containing “NASA” and sums their payload mass.

The result shows that NASA missions accounted for a total of approximately **45596 kg** of payload mass launched by SpaceX boosters.

Average Payload Mass by F9 v1.1

```
SELECT AVG(Payload_Mass__kg_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version='F9 v1.1'
```

This query calculates the average payload mass for all SpaceX launches that used the F9 v1.1 booster.

The result shows that the F9 v1.1 booster averaged a total of approximately **2928.4 kg** of payload mass.

First Successful Ground Landing Date

- SELECT Date
- FROM SPACEXTABLE
- WHERE Landing_Outcome = 'Success (ground pad)'
- ORDER BY Date
- LIMIT 1;

Explanation: This query filters all records for landings with the outcome "Success (ground pad)". It then sorts those results by the launch date in ascending order.

The earliest date is returned using LIMIT 1.

This identifies the first-ever successful Falcon 9 booster landing on a ground landing pad — on **12-22-2015**.

Successful Drone Ship Landing with Payload between 4000 and 6000

The query filters for boosters with Successful landings on drone ships, Payload mass between 4000 and 6000 kg.

The result lists specific booster versions that met these criteria.

These boosters demonstrate the balance between medium-heavy payloads and SpaceX's reusability via drone ship landings.

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (drone ship)'
    AND Payload_Mass__kg_ > 4000
    AND Payload_Mass__kg_ < 6000;
```

Booster Version:
F9 FT B1022
F9 FT B1025
F9 FT B1031
F9 Block 5 B1046
F9 Block 5 B1051

Total Number of Successful and Failure Mission Outcomes

```
SELECT Mission_Outcome, COUNT(*) AS Total
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- The query groups all missions by their outcome and counts how many times each occurred.
- This reveals the total number of successful and failed missions.
- Most missions were successful, reflecting SpaceX's strong launch performance.

Boosters Carried Maximum Payload

```
SELECT Booster_Version

FROM SPACEXTABLE

WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)

FROM SPACEXTABLE);
```

These booster versions are identified from the query that filters for the maximum value in the Payload_Mass__kg_ column.

The repeated appearance of Block 5 boosters (F9 B5) shows they are consistently used for heavy payload missions, reflecting their advanced capability in the Falcon 9 series.

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

2015 Launch Records

```
SELECT Landing_Outcome, Booster_Version,
Launch_Site
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE '%Failure%'
      AND Landing_Pad LIKE '%drone%'
      AND Date LIKE '2015%';
```

MONTH	DATE	booster_version	launch_site	landing_outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

In 2015, SpaceX attempted multiple drone ship landings using the F9 v1.1 booster version from CCAFS SLC 40. However, these missions failed to land successfully on the drone ship, reflecting the early stage of Falcon 9's landing technology.

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

```
SELECT Landing_Outcome, COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
  AND Landing_Outcome IN ('Failure (drone ship)', 'Success (ground pad)')
GROUP BY Landing_Outcome
ORDER BY Outcome_Count DESC;
```

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

This query filters launches between June 4, 2010 and March 20, 2017, then counts how many times each selected landing outcome occurred.

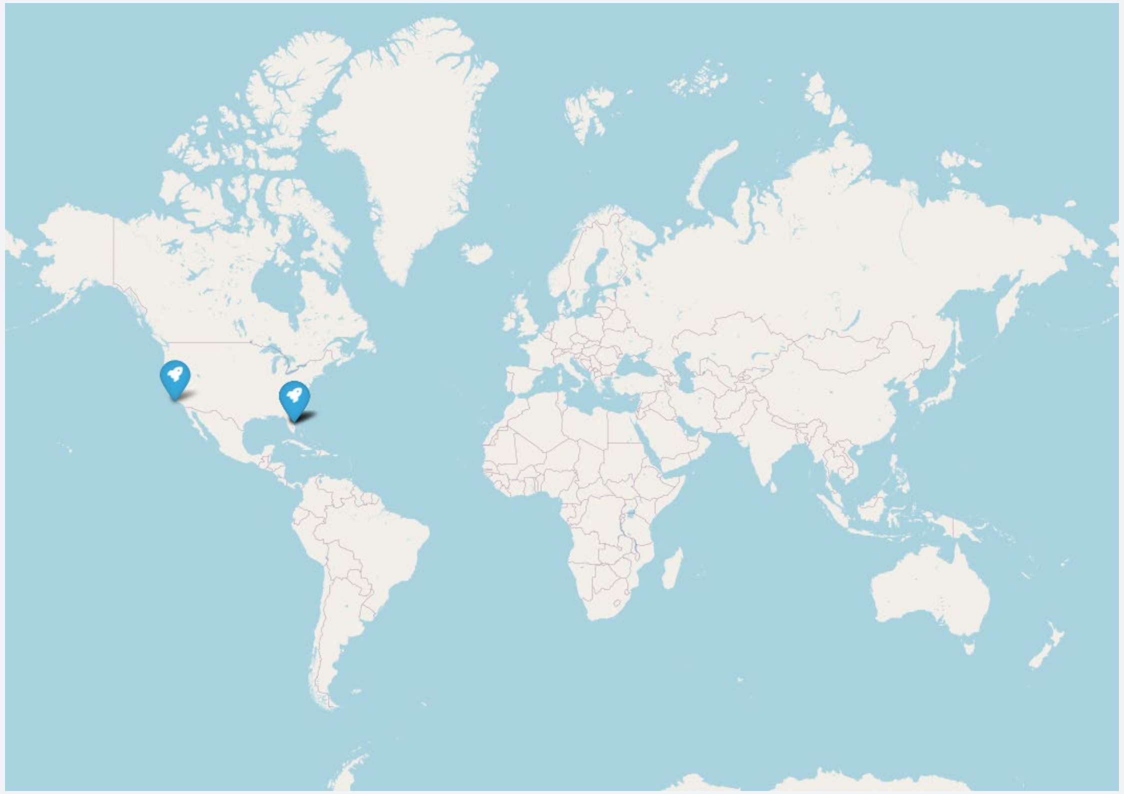
The results show that “Failure (drone ship)” happened more frequently than “Success (ground pad)” in this period.

Section 3

Launch Sites Proximities Analysis

Launch Sites on a Global Map

This folium map displays the geographic locations of all SpaceX launch sites worldwide using interactive markers. Each marker pinpoints a specific launch facility and provides a pop-up label with the site’s name.



- **Key Findings:**
  - All SpaceX launch sites are located in the United States.
  - Major launch locations include Cape Canaveral, Kennedy Space Center, and Vandenberg Air Force Base.
  - The map highlights SpaceX's focus on coastal launch facilities to optimize trajectory and safety for orbital insertions.

Successful & Failed Launch Sites

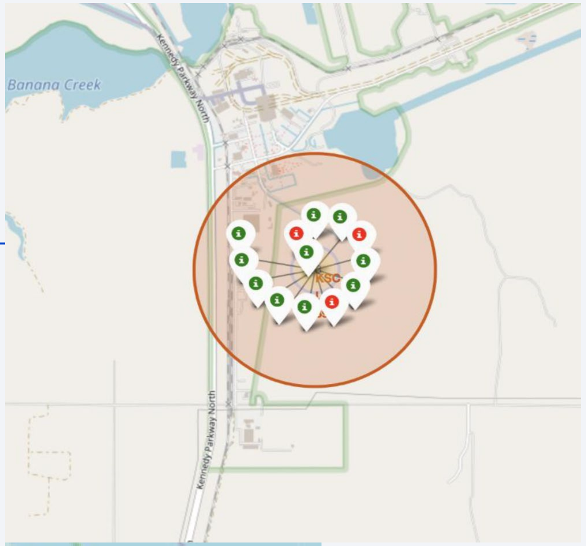
- **Map Description:**

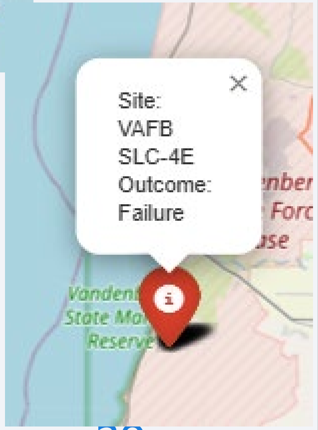
This folium map visually represents SpaceX launch outcomes by geographic location. Each launch is marked with a color-coded icon:

  - **Green**: Successful launch
  - **Red**: Failed launch

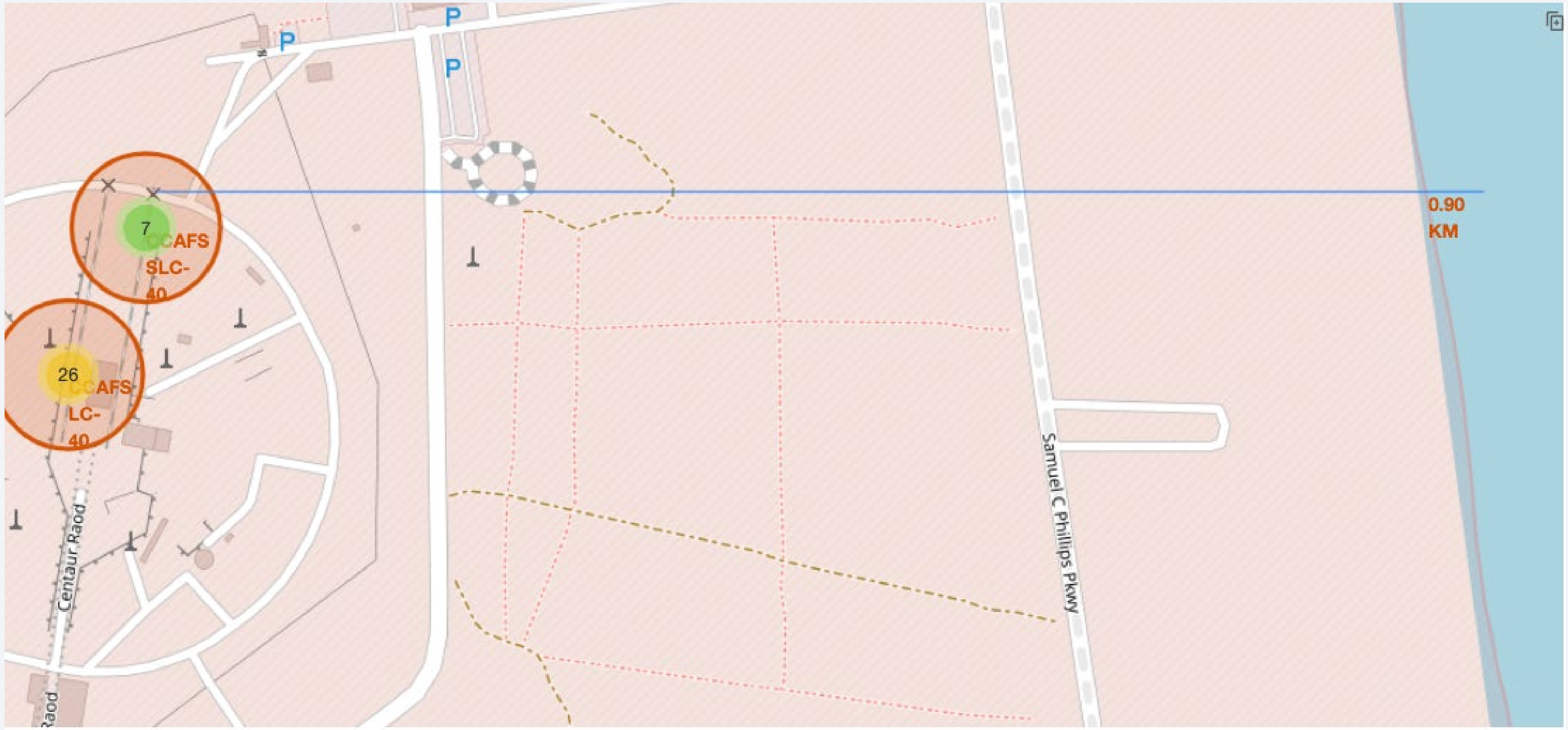
- **Key Elements:**
  - **Markers**: Each represents an individual launch event.
  - **Popups**: Display the launch site name and the outcome.
  - **Color Coding**: Provides immediate visual feedback on performance per site.

- **Findings:**
  - Most launch sites show a predominance of successful missions (green markers).
  - A few sites, such as early launches at Vandenberg or Cape Canaveral, show a mix of successes and failures.
  - The clustering of successful launches at certain sites may indicate improved reliability over time or more experienced site management.





Distance to the Coastline

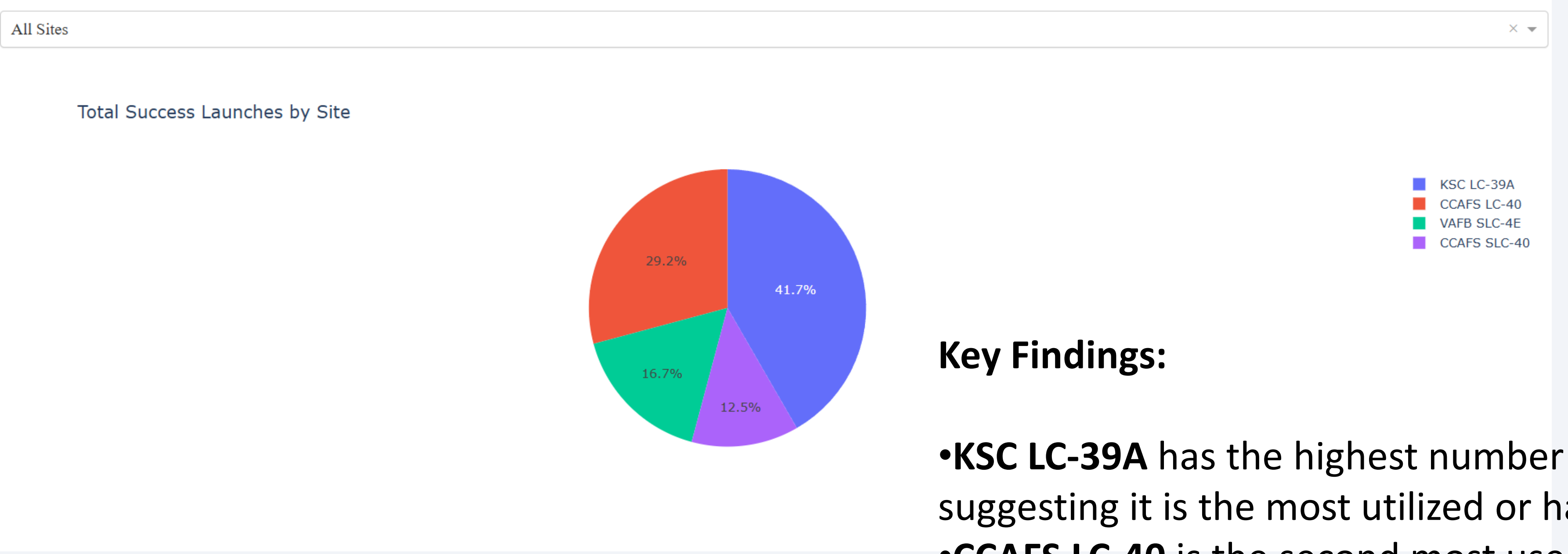


- **Conclusion:**
  Locating launch sites near coastlines and away from infrastructure is essential for public safety, mission flexibility, and operational compliance.

Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site

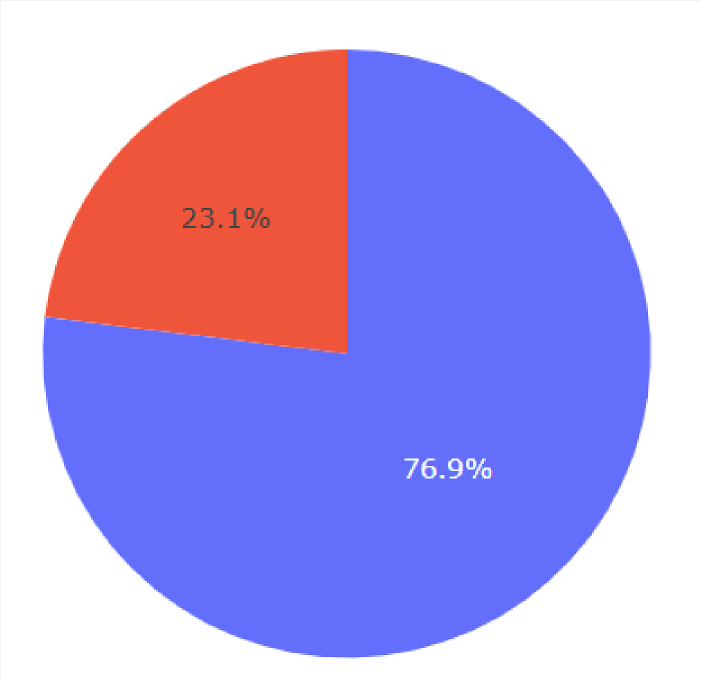


Key Findings:

•**KSC LC-39A** has the highest number of successful launches, suggesting it is the most utilized or has the highest success efficiency.
•**CCAFS LC-40** is the second most used successful launch site, likely serving as a backup or secondary site.
•**Smaller shares** from **VAFB SLC-4E** and **CCAFS SLC-40** indicate either fewer launches or lower success rates from these sites.
•The data helps **visually rank launch sites** based on performance, which is crucial for strategic planning and logistics in future missions.

Success vs Failure at site KSC LC-39A

- **High Success Rate**:
  KSC LC-39A boasts a **76.9% success rate**, demonstrating strong reliability and consistent performance. This aligns with it being the top site for total successful launches.
- **Room for Improvement**:
  While impressive, the **23.1% failure rate** indicates that nearly 1 in 4 launches from this site still face issues — a non-negligible risk in space missions.
- **Operational Preference**:
  The high usage and success make LC-39A a **strategic asset**. It is likely preferred for high-stakes or payload-sensitive missions.
- **Strategic Significance**:
  This success rate reinforces the importance of infrastructure, testing, and design optimization at this flagship NASA/SpaceX facility.



Payload vs. Success Rate – All Sites

Payload range (Kg):

Payload vs. Success Rate – All Sites (Continued)

- Key Findings:
  - Payload Range with High Success:
    - Most successful launches (green dots at class = 1) are concentrated between 2000 and 6000 kg.
    - Very heavy payloads (above 6000 kg) show mixed success, with some failures.
  - Booster Version Performance:
    - FT (green) and B5 (orange) have the highest density of successful launches, especially in the mid-payload range.
    - v1.0 (blue) and v1.1 (red) show more failure instances (dots near class = 0).
  - Booster Advancement Over Time:
    - The newer versions (FT, B5) clearly outperform older ones (v1.0, v1.1), supporting the idea of iterative engineering improvements.

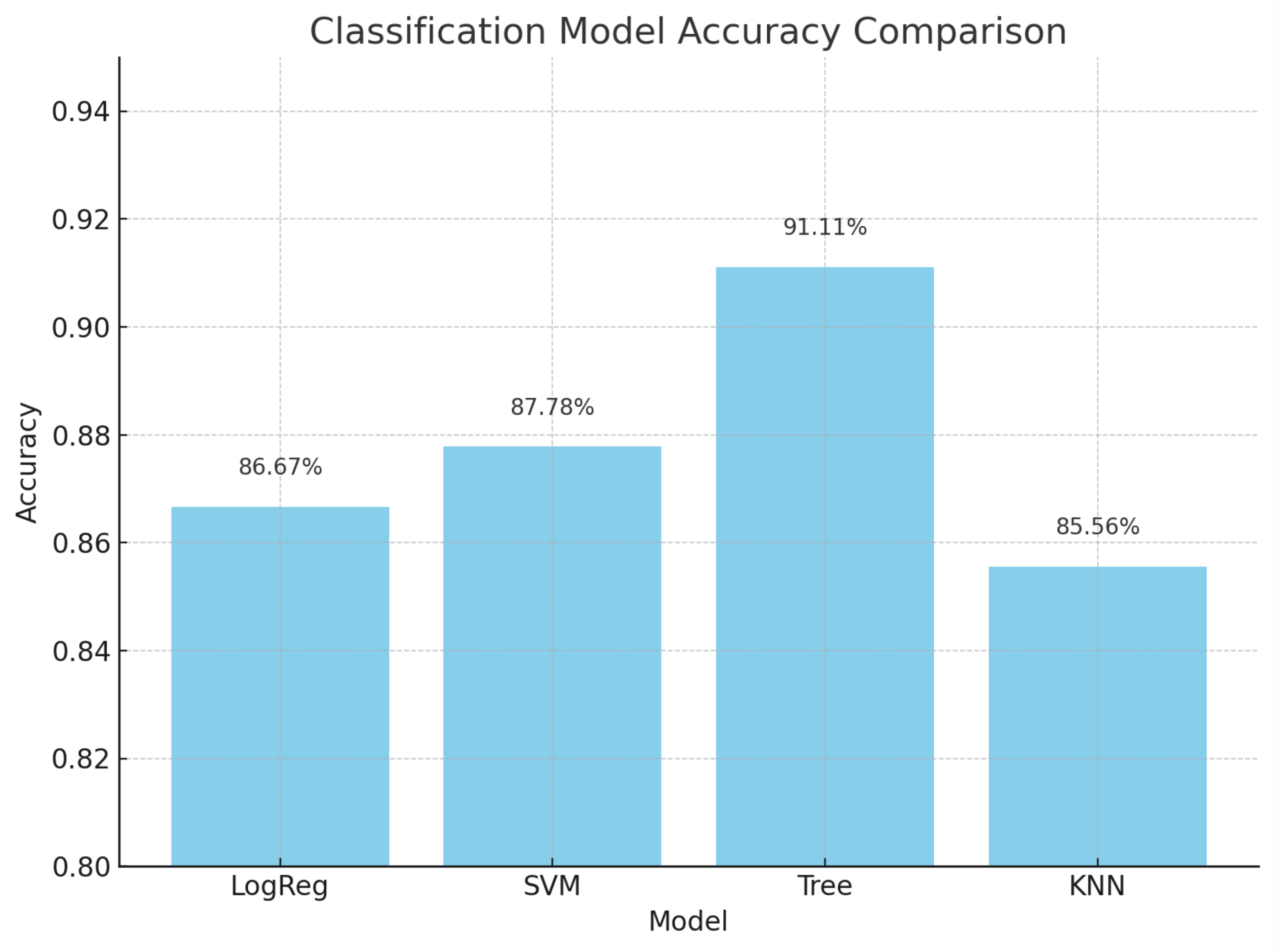
Conclusion:

The FT and B5 booster versions are the most reliable, particularly in the payload mass range of 2000–6000 kg. Launches involving either very light or very heavy payloads tend to have a slightly higher failure rate, emphasizing the importance of optimal payload design and mature booster technology.

Section 5

Predictive Analysis (Classification)

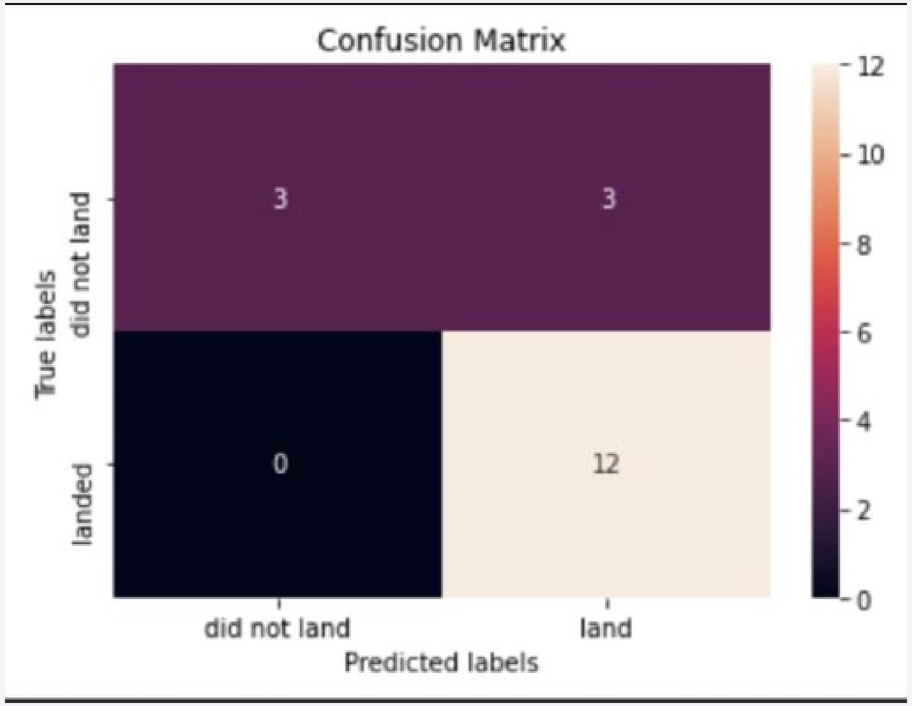
Classification Accuracy



- The **Decision Tree model** has the highest classification accuracy at **91.11%**, making it the best-performing model among the four in this case.

Confusion Matrix Analysis – Rocket Landing Prediction

- Key Insights:
  - 100% Recall for landings – All actual successful landings (12) were correctly predicted.
  - High overall accuracy – 15 out of 18 predictions were correct.
  - Some false positives – 3 failed landings were wrongly predicted as successful.



- Conclusion:
  - The model is highly reliable in predicting successful landings.
  - Improvement opportunity: Reduce false positives to enhance prediction accuracy for failed landings.

Key Point Conclusions

- **Most Launches Are Near Coastlines for Safety**
  Launch sites are strategically placed near coasts and away from population centers to reduce risk from launch failures or falling debris.
- **KSC LC-39A Had the Highest Success Rate**
  Kennedy Space Center Launch Complex 39A demonstrated the highest success rate among all sites, with over 75% successful missions and no false negatives in the confusion matrix.
- **Payload Mass Has Minimal Impact on Success Rate**
  Launches with a wide range of payload masses—from under 1000 kg to nearly 10,000 kg—achieved similar success rates, indicating the models and technology handle varied payloads reliably.
- **FT and B5 Booster Versions Show High Reliability**
  Later booster versions like Falcon 9 Full Thrust (FT) and Block 5 (B5) consistently achieved higher success rates compared to earlier versions like v1.0 and v1.1.
- **Decision Tree Classifier Performed Best**
  Among all machine learning models, the **Decision Tree** achieved the highest classification accuracy (91.1%) and the best F1 Score (0.9375), making it the most effective model for predicting landing outcomes.

Thank you!