IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Heidi Leonard
November 15, 2024

Outline

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

Executive Summary

Summary of Methodologies

The project utilized data from the SpaceX API and web scraping, followed by cleaning, filtering, and enrichment to analyze launch performance. Key metrics such as payload mass, launch site, orbit type, and landing outcomes were visualized using SQL, Folium maps, and Plotly Dash dashboards. Predictive models, including Logistic Regression, SVM, KNN, and Decision Tree, were developed and fine-tuned to forecast first-stage landing success. Decision Tree emerged as the best-performing model, providing actionable insights for mission planning.

Summary of Results

Launch Performance:

- **KSC LC-39A**: Highest success rate (76.9%) for payloads between 2,000–7,000 kg, making it the most reliable site.
- **VAFB SLC-4E**: Moderate success (40%) for polar orbits with low payloads (<4,000 kg).
- **CCAFS LC-40**: Lower success rate (26.9%) with mixed outcomes.
- **Orbit Challenges**: GTO orbit had the lowest success rate, indicating areas for improvement.

Predictive Model Results:

- **Decision Tree**: Best accuracy (87.5%), with no false negatives and minimal errors.
- Other models performed competitively (Logistic Regression: 84.64%, SVM/KNN: 84.82%).

Key Insights:

- Prioritize **KSC LC-39A** for mid-range payloads.
- Address challenges in GTO orbit missions.
- Use Decision Tree predictions to guide mission planning and resource allocation.

Introduction

Project Goal

- Enhance Space Y's competitiveness by improving first-stage rocket landing success through data analysis and predictive modeling.

Data Collection & Preparation

- Sources: SpaceX API and web scraping.
- Focus: Launch data including payload mass, launch sites, orbit types, and outcomes.

Exploratory Data Analysis (EDA)

- Tools: SQL, Folium maps, Plotly Dash.
- Key Insights:

1. **KSC LC-39A**: Most reliable site (76.9% success for 2,000–7,000 kg payloads).
2. **GTO Orbit**: Lowest success rate, highlighting challenges.

Predictive Modeling

- Models: Logistic Regression, SVM, KNN, Decision Tree.
- Best Model: Decision Tree (87.5% accuracy, no false negatives).

Actionable Insights

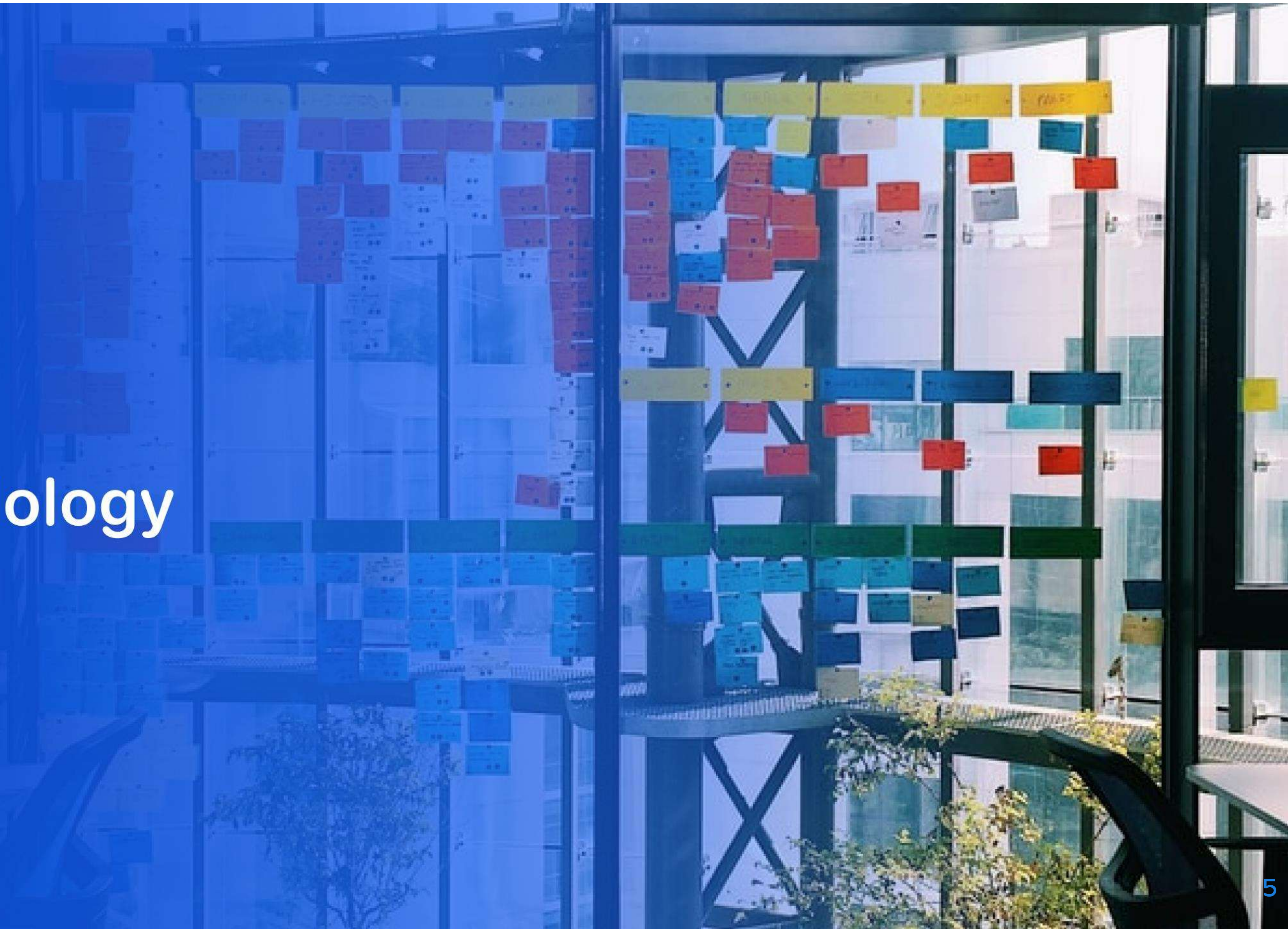
- Dashboards for site, payload, and orbit analysis.
- Data-driven recommendations to optimize launches and resource allocation.

Conclusion

- Provides a robust framework for Space Y to enhance its competitive position.

Section 1

Methodology



Methodology & Executive Summary

Methodology Overview

1. Data Collection

- Retrieved SpaceX launch data via API and Wikipedia.
- Enriched data with booster, payload, site, and outcome details.

2. Data Wrangling

- Filtered Falcon 9 launches, handled missing values, and reset flight numbers.

3. Exploratory Data Analysis (EDA)

- Used SQL and visualizations to analyze launch sites, payloads, and outcomes.

4. Interactive Visual Analytics

- **Folium**: Mapped launch sites and outcomes.
- **Plotly Dash**: Interactive analysis of site success, payloads, and boosters.

5. Predictive Analysis

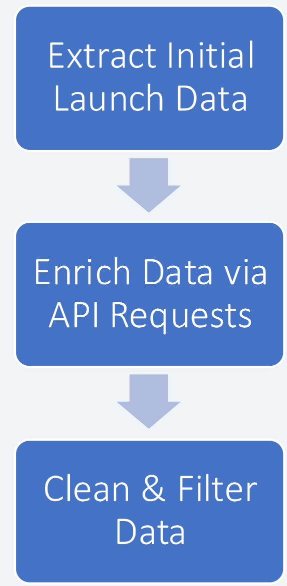
- Trained Logistic Regression, SVM, Decision Tree, and KNN models.
- Decision Tree identified as the best model after tuning and evaluation.

Summary

- Comprehensive pipeline from data collection to modeling.
- Key insights: site performance, payload ranges, booster effectiveness.
- Interactive tools for data-driven mission planning.

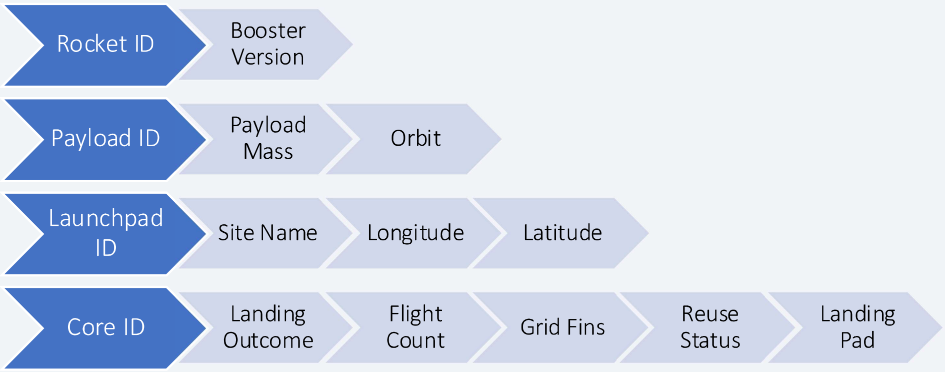
Data Collection

- **Initial Data Retrieval**
  - Access SpaceX API
  - Retrieve launch data in JSON format
  - Convert JSON data to flat DataFrame using pd.json_normalize()
- **Data Enrichment via API Calls**
  - Extract detailed attributes using unique IDs:
    - **Rocket ID** → Booster Version
    - **Payload ID** → Payload Mass, Orbit
    - **Launchpad ID** → Launch Site Name, Longitude, Latitude
    - **Core ID** → Landing Outcome, Flight Count, Grid Fins, Reuse Indicators, Landing Pad
- **Data Cleaning & Filtering**
  - Filter dataset to include only Falcon 9 launches
  - Impute missing `PayloadMass` with mean value
  - Retain None values in `LandingPad` to indicate non-use
  - Reset `FlightNumber` to sequential order for Falcon 9



Data Collection – SpaceX API

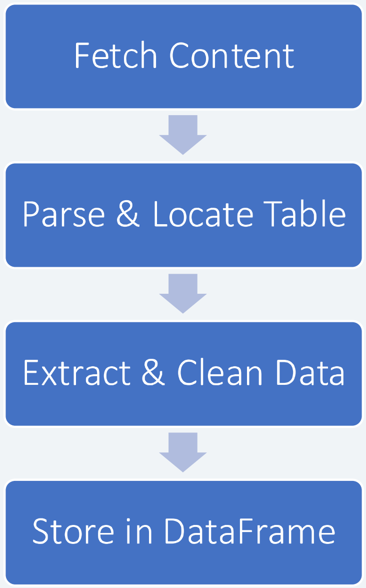
- Key details about launches are accessed via specific API endpoints. A series of helper functions loop through unique IDs in columns like rocket, payloads, launchpad, and cores to extract additional information.
- Each helper function is designed to make API calls using the relevant IDs:
  - **Booster Version**: Retrieved using the rocket ID, providing specific details about the booster used for each launch.
  - **Payload Details**: Accessed with the payload ID to retrieve PayloadMass (in kilograms) and Orbit.
  - **Launch Site Information**: Using the launchpad ID, the launch site's name, longitude, and latitude are retrieved.
  - **Core Details**: Extracted using the core ID, providing features such as landing success, grid fins, reused core indicators, and landing pad use.



GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

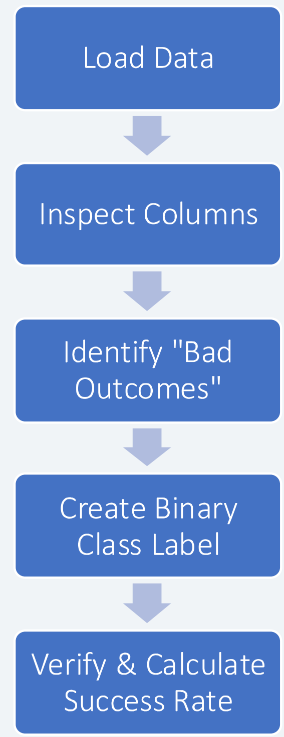
- **Fetch Content**
  - Use requests.get() to retrieve HTML from Wikipedia.
- **Parse HTML & Locate Table**
  - Use BeautifulSoup to parse HTML and identify the launch data table.
- **Extract & Clean Data**
  - Loop through table rows, applying helper functions to extract and clean values (e.g., date, payload, orbit).
- **Store in DataFrame**
  - Populate launch_dict and convert to a pandas DataFrame, cleaning any final inconsistencies.



GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- **Load & Inspect Data**
  - Load dataset and check structure for completeness.
  - Review key columns: launch details, landing outcomes.
- **Define "Bad Outcomes"**
  - Identify failure cases in Outcome column.
  - Group failure types (e.g., "None None", "False ASDS") as "bad outcomes."
- **Create Class Label**
  - Generate binary labels: 1 for success, 0 for failure.
  - Apply Class label based on "bad outcomes" set.
- **Verify & Analyze**
  - Confirm accuracy of Class labels.
  - Calculate success rate as a metric for model baseline.



GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with SQL

- **Unique Launch Sites**: Retrieved distinct SpaceX launch sites to analyze frequently used locations.
- **Sample of 'CCA' Launch Sites**: Displayed 5 records of launches from sites starting with "CCA" for targeted analysis.
- **Total Payload for NASA (CRS) Missions**: Calculated total payload mass for NASA missions under the CRS program to understand payload contributions.
- **Average Payload for F9 v1.1**: Computed average payload carried by the F9 v1.1 booster to assess its payload performance.
- **Date of First Ground Pad Success**: Identified the date of SpaceX's first successful ground pad landing, a key operational milestone.
- **Moderate Payload Drone Ship Landings**: Listed boosters that successfully landed on drone ships with payloads between 4000-6000 kg.
- **Count of Mission Outcomes**: Tallied successful and failed mission outcomes for an overview of mission reliability.
- **Boosters with Maximum Payload Mass**: Identified boosters carrying the highest payloads, highlighting SpaceX's most powerful launches.
- **Failed Drone Ship Landings in 2015**: Listed failed drone ship landings in 2015, detailing month, booster version, and launch site.
- **Landing Outcome Rankings**: Ranked landing outcomes by frequency between 2010 and 2017 to understand common successes and failures.

GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with Data Visualization

- **Flight Number vs. Payload Mass (Scatter Plot)**: Created to observe how flight number and payload mass relate to the success of the landing outcome (Class). This analysis indicates that success likelihood increases with more flight experience and that payload mass affects outcomes across sites.
- **Flight Number vs. Launch Site (Scatter Plot)**: Used to examine the relationship between flight number and launch site, with color-coded success indicators. This chart provided insight into how experience at different sites might impact success rates.
- **Payload Mass vs. Launch Site (Scatter Plot)**: Plotted to explore how different launch sites handle various payload masses, helping identify any patterns between launch sites and payload capacity.
- **Success Rate by Orbit Type (Bar Chart)**: Used to compare success rates across different orbit types. This visualization highlighted which orbits were more consistently successful, informing which mission types tend to have higher landing success.
- **Flight Number vs. Orbit Type (Scatter Plot)**: Visualized to check for patterns between flight experience and orbit type, providing insight into how mission type impacts success as flight experience increases.
- **Payload Mass vs. Orbit Type (Scatter Plot)**: Plotted to observe how payload mass influences success across orbit types, showing which orbits are more likely to succeed with heavier payloads.
- **Yearly Success Rate Trend (Line Chart)**: Tracked the yearly trend in success rate to reveal improvements in landing success over time, likely due to advancements and operational experience.

GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/edadataviz.ipynb

Build an Interactive Map with Folium

Map Objects Added

- **Launch Site Markers**: Labelled markers for each launch site location.
- **Outcome Markers**: Clustered, color-coded markers (green for success, red for failure) showing individual launch outcomes.
- **Highlight Circles**: Circles around each launch site to visually emphasize locations; yellow circle for NASA JSC as a reference point.
- **Distance Lines**: Lines between launch sites and key infrastructure (e.g., coastlines, highways) to show proximities.

Purpose

- **Markers**: Identify site locations and visualize success/failure rates.
- **Circles**: Highlight launch areas and establish reference points.
- **Lines**: Illustrate proximity to infrastructure, informing site selection factors like safety and accessibility.

GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Launch Site Dropdown & Success Pie Chart

- **Function**: Allows users to select a launch site or view all sites collectively.
- **Plot**: Pie chart displays total successful launches across all sites or a success (class=1) vs. Failure (class=0) breakdown for a chosen site.
- **Insight**: Identifies sites with the highest number of successful launches or the highest success rate, enabling comparisons and supporting site-specific performance insights.

Payload Range Slider & Success-Outcome Scatter Plot

- **Function**: Filters payload mass range and launch site to examine mission outcomes.
- **Plot**: Scatter plot visualizes payload mass (x-axis) vs. mission success (y-axis), with each point color-coded by booster version.
- **Insight**: Reveals correlations between payload range and success rate, highlighting optimal payload sizes and successful booster versions. Users can analyze if certain booster types perform better within specific payload ranges.

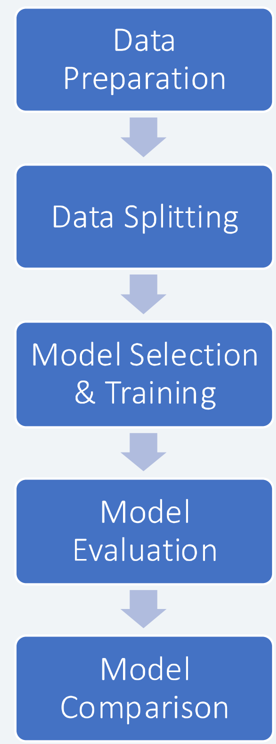
Overall Purpose

- Provides a flexible, interactive platform for analyzing SpaceX launch data.
- Enables data-driven insights into factors influencing launch success, such as site performance, payload ranges, and booster effectiveness.
- Supports dynamic exploration to answer key questions, including identifying optimal launch sites, payload configurations, and booster versions for successful missions.

GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- **Data Preparation**
  - **Exploratory Data Analysis (EDA)**: Inspected the dataset, identified missing values, and visualized feature distributions.
  - **Target Label Creation**: Created a binary label (Class) indicating successful or unsuccessful landing.
  - **Feature Standardization**: Standardized features to ensure consistent scales for model training.
- **Data Splitting**
  - **Training and Test Split**: Divided the dataset into training (80%) and test (20%) sets for model validation.
- **Model Selection and Training**
  - **Logistic Regression, SVM, Decision Tree, KNN**: Selected four classification models for comparison.
  - **Hyperparameter Tuning**: Used GridSearchCV with 10-fold cross-validation to find the optimal hyperparameters for each model.
- **Model Evaluation**
  - **Accuracy Calculation**: Evaluated each model's accuracy on the test set.
  - **Confusion Matrix Analysis**: Examined confusion matrices to identify common types of errors (e.g., false positives).
- **Model Comparison**
  - **Performance Summary**: Compiled cross-validated accuracy scores for each model.
  - **Best Model Selection**: Identified the Decision Tree model as the best-performing model.



GitHub URL: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results: Exploratory Data Analysis (EDA)

- **Launch Site Performance**
- **CCAFS SLC 40**: Most launches, high success for >14,000 kg payloads.
- **KSC LC 39A**: High success for 2,000–7,000 kg payloads.
- **VAFB SLC 4E**: Consistent success for payloads <4,000 kg.
- **Orbit & Payload Insights**
- **High Success**: Orbits ES-L1, GEO, SSO, VLEO.
- **Challenges**: GTO orbit has lowest success rate.
- **Payloads**:
  - <6,000 kg: High success across most orbits.
  - 12,000 kg: Limited data, but high success for ISS and VLEO.

Yearly Trends

- Success rates stabilized at ~90% since 2019.

Project Impact

- Insights inform predictive models for landing success.
- Supports data-driven decisions on payloads, sites, and orbits.

Results: Interactive Analytics

Key Findings

- **Launch Sites**:
  - **KSC LC-39A**: Highest success rate (76.9%) for payloads 2,000–7,000 kg.
  - **VAFB SLC-4E**: Moderate success (40%) for polar orbits and low payloads (<4,000 kg).
  - **CCAFS LC-40**: Lowest success rate (26.9%) with mixed payload performance.

- **Payload and Orbit**:
  - **Low Payloads (<6,000 kg)**: High success across most orbits.
  - **High Payloads (>12,000 kg)**: Limited data, but high success for ISS and VLEO.
  - **GTO Orbit**: Faces the most challenges, with the lowest success rate.

Actionable Insights

- **Focus on Equatorial Launch Sites**: Prioritize KSC LC-39A for reliable performance.
- **Optimize Payload Configurations**: Match payload mass to site and orbit for better outcomes.
- **Address GTO Orbit Challenges**: Investigate improvements for this low-success orbit.

Results: Predictive Analytics

Model Performance

- **Decision Tree**: Best-performing model with **87.5% accuracy**.
- Logistic Regression: 84.64%, SVM & KNN: 84.82%.

Decision Tree Strengths

- Handles non-linear relationships and complex interactions effectively.
- Outperforms other models for predicting Falcon 9 first-stage landings.

Confusion Matrix Results

- **True Positives (TP)**: 12 | **True Negatives (TN)**: 5
- **False Positives (FP)**: 1 | **False Negatives (FN)**: 0

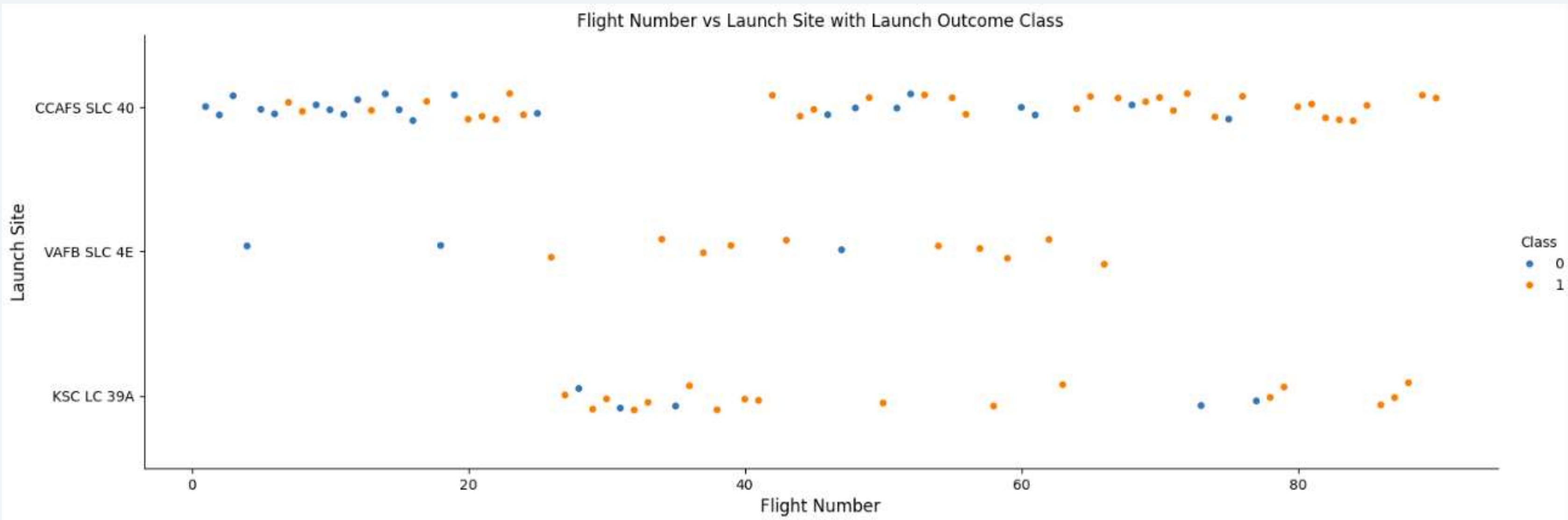
Key Insights

- No false negatives: All successful landings correctly identified.
- Minimal misclassification: Only one false positive.
- Conclusion: Decision Tree offers high accuracy and reliability for landing predictions.

Section 2

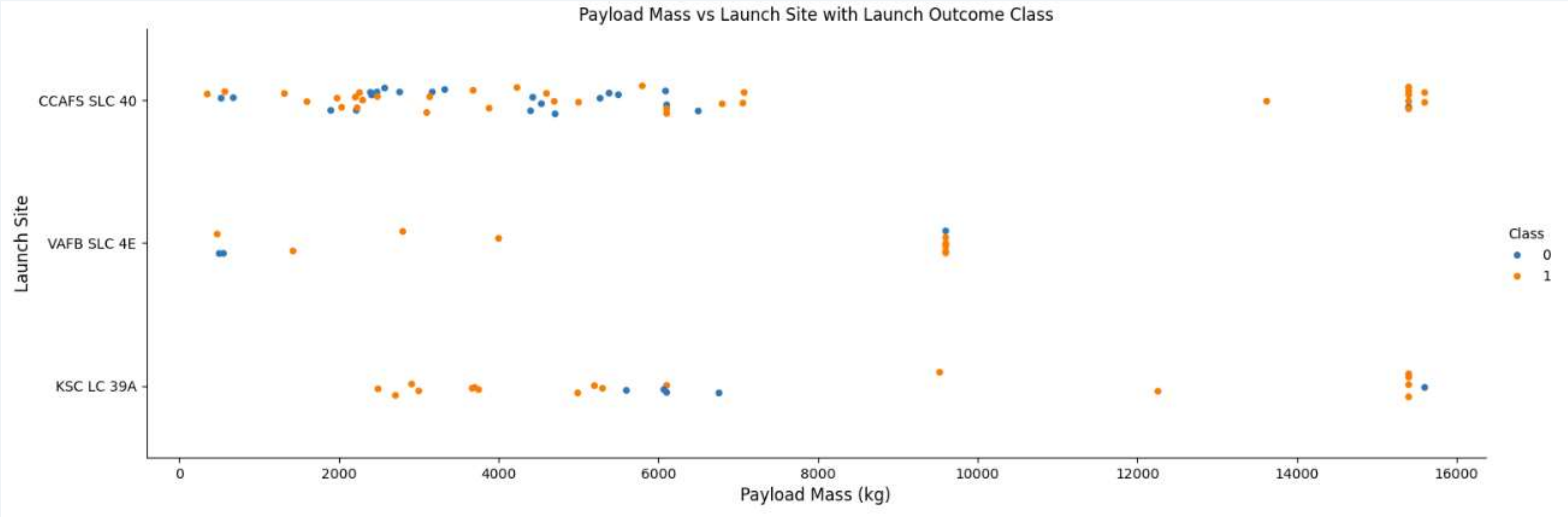
Insights drawn from EDA

Flight Number vs. Launch Site



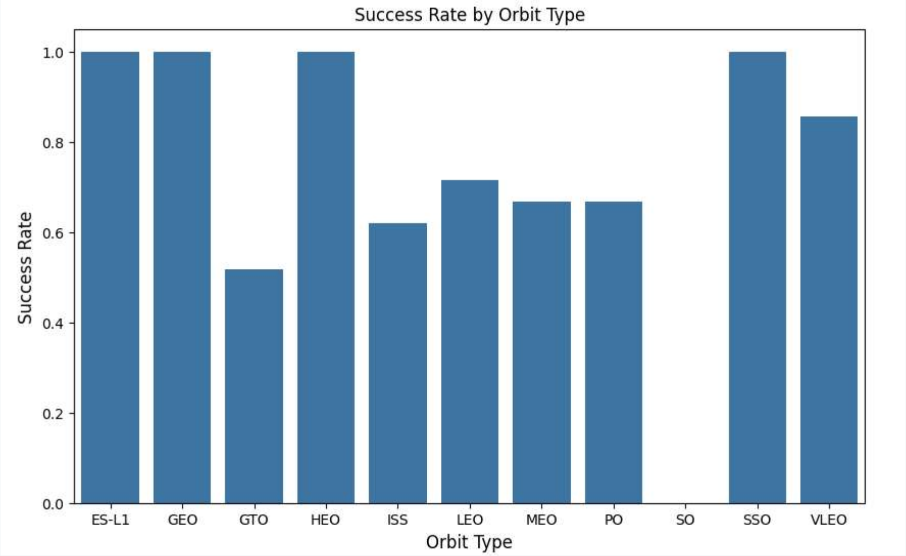
- **CCAFS SLC 40**: Shows the highest number of launches with a gradual increase in success rate as flight numbers increase, indicating steady improvements over time.
- **VAFB SLC 4E**: With fewer launches overall, this site displays mostly successful outcomes concentrated around mid-range flight numbers, suggesting favorable conditions or specific mission types at this range.
- **KSC LC 39A**: Demonstrates a high success rate between mid-range and higher flight numbers, indicating strong performance at this site as flight numbers increase.

Payload vs. Launch Site



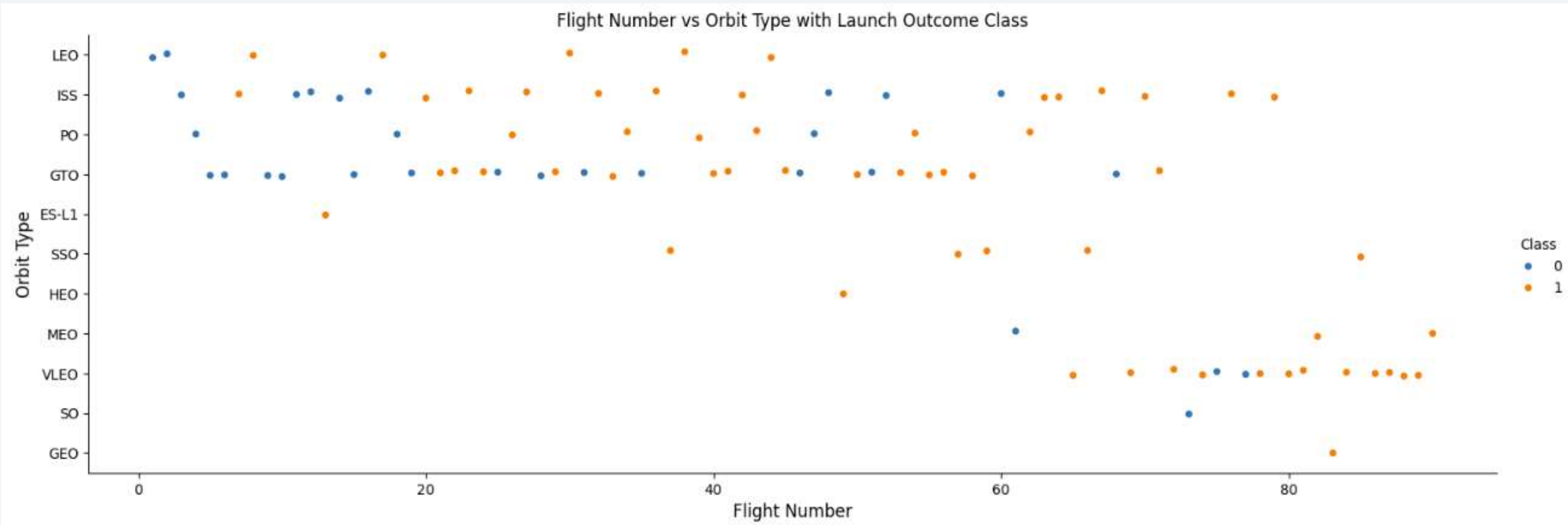
- **CCAFS SLC 40**: Handles the widest range of payloads. Shows mixed outcomes for payloads between 2000–4000 kg, no payloads between 7000–14,000 kg, and a high success rate for payloads above 14,000 kg.
- **VAFB SLC 4E**: Has the fewest launches overall, mostly successful. Payloads are concentrated in the low-range (up to 4000kg), with no launches between 4000–10,000 kg or above 10,000 kg.
- **KSC LC 39A**: Primarily supports payloads between 2000–7000 kg with high success rates and a few successful heavy payloads between 9000–16,000 kg.

Success Rate vs. Orbit Type



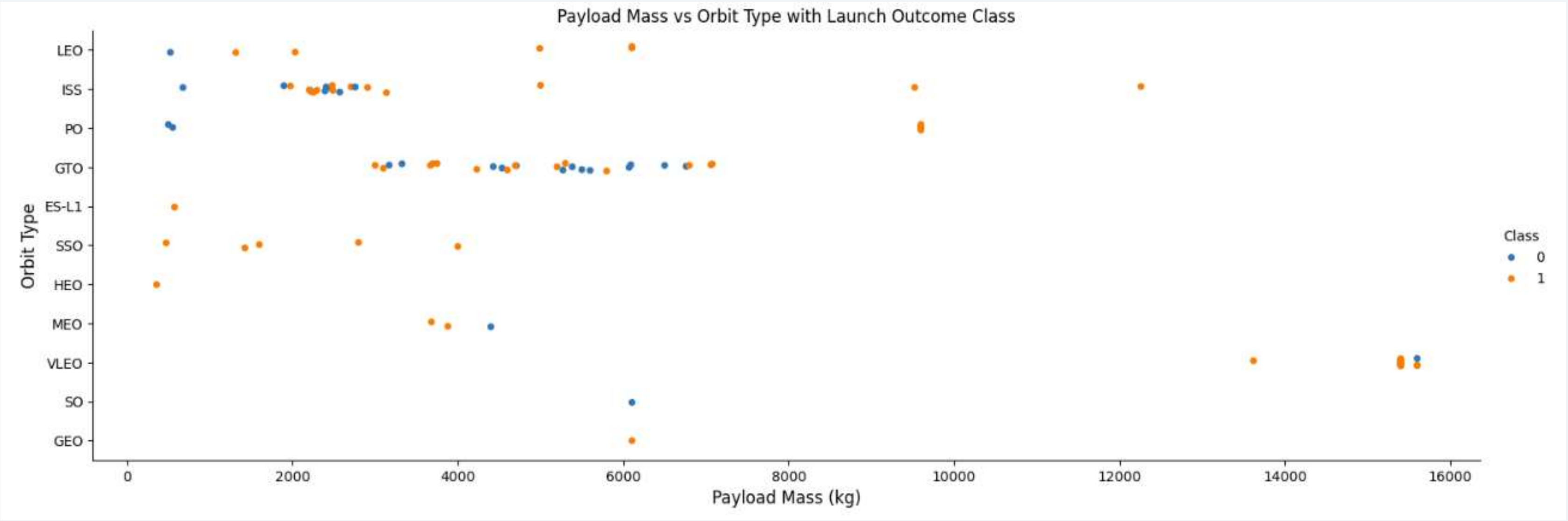
- **High Success Rates**: Orbits **ES-L1**, **GEO**, **HEO**, **SSO**, and **VLEO** achieve success rates close to or at 100%.
- **Moderate Success Rates**: Orbits **LEO**, **MEO**, **ISS**, and **PO** show moderate success rates, indicating some operational challenges.
- **Low Success Rate**: **GTO** has the lowest success rate, suggesting more significant challenges.
- **No Data**: No success data is available for the **SO** orbit.

Flight Number vs. Orbit Type



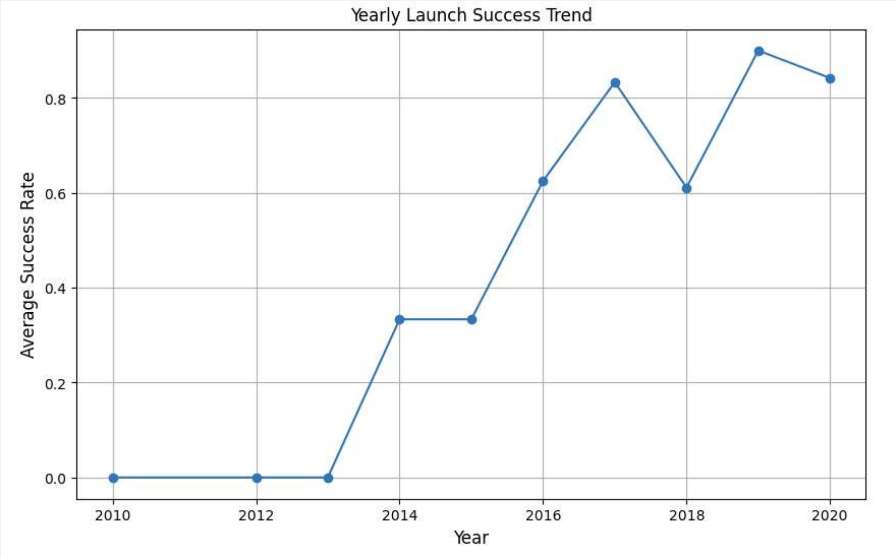
- **Low Flight Numbers (0–20)**: Mixed success across LEO, ISS, PO, and GTO orbits. Limited data available for ES-L1, but success rates are high.
- **Mid Flight Numbers (20–60)**: Success rates improve notably for ISS and PO. GTO continues to show mixed results. Limited data for LEO and SSO, both showing high success in this range.
- **High Flight Numbers (60+)**: High success rates for ISS and VLEO. Limited data for SSO, MEO, and GEO, all with high success.

Payload vs. Orbit Type



- **Low Payload Mass (0–6000 kg)**: **ISS** and **GTO** show mixed results. **LEO**, **ES-L1**, **SSO**, **HEO**, and **MEO** have limited data but generally show high success.
- **Mid Payload Mass (6000–12,000 kg)**: **GTO** continues to show mixed results. **LEO**, **ISS**, **PO**, and **GEO** have limited data but are generally successful.
- **High Payload Mass (12,000+ kg)**: **ISS** and **VLEO** have limited data, with generally high success rates.

Launch Success Yearly Trend



- **2010–2014**: Success rates remained low, with gradual improvement starting around 2014.
- **2015–2018**: Significant upward trend in success rates, reaching above 80% by 2017 but dropping to around 60% in 2018.
- **2019–2020**: Success rates stabilized at high levels, maintaining above 80–90%, indicating consistent reliability.

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%%sql
SELECT DISTINCT "Launch_Site"
FROM SPACEXTABLE;
```

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

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

The output shows a list of unique launch sites where SpaceX has conducted missions. In this case, there are four unique sites:

- **CCAFS LC-40**: Cape Canaveral Air Force Station, Launch Complex 40.
- **VAFB SLC-4E**: Vandenberg Air Force Base, Space Launch Complex 4 East.
- **KSC LC-39A**: Kennedy Space Center, Launch Complex 39A.
- **CCAFS SLC-40**: Cape Canaveral Space Launch Complex 40.

Launch Site Names Begin with 'CCA'

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

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

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

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

- The output displays details of 5 launches from CCAFS LC-40 (Cape Canaveral Air Force Station, Launch Complex 40), including:
  - **Date/Time** of each launch.
  - **Booster Version** and **Payload** details.
  - **Payload Mass (kg)**, **Orbit** destination (mainly LEO).
  - **Customer** (e.g., SpaceX and NASA).
  - **Mission and Landing Outcomes** (e.g., "Success," "Failure (parachute)").

Total Payload Mass

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

```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass
FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA (CRS)%';
```

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

Total_Payload_Mass
48213

- The output shows a total payload mass of 48,213 kg for missions associated with NASA under the Commercial Resupply Services (CRS) program, representing the cumulative mass carried by SpaceX boosters on these missions.
- This query provides insight into the total payload mass delivered for NASA’s CRS missions, highlighting SpaceX's role in supporting NASA’s resupply efforts.

Average Payload Mass by F9 v1.1

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

```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

Average_Payload_Mass
2928.4

- The output shows that the average payload mass carried by the F9 v1.1 booster version is **2928.4 kg**.
- This query provides insight into the typical payload capacity handled by the `F9 v1.1` booster version, offering a metric for its performance and utility.

First Successful Ground Landing Date

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

Hint:Use min function

```
%%sql
SELECT MIN(Date) AS First_Successful_Landing_Date
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Successful_Landing_Date
2015-12-22

- The output shows that the first successful landing on a ground pad occurred on **2015-12-22**.
- This query highlights the significant milestone when SpaceX achieved its first successful ground pad landing, a crucial advancement in reusable rocket technology.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (drone ship)'
AND "PAYLOAD_MASS__KG_" > 4000
AND "PAYLOAD_MASS__KG_" < 6000;
```

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

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

The output provides a list of specific booster versions—**F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2**—that meet two important criteria:

- They achieved a **successful landing on a drone ship**, which is a key goal in SpaceX's reusable rocket program.
- They carried a **payload mass between 4000 and 6000 kg**, which categorizes them as boosters capable of handling moderate payloads.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%%sql
SELECT
    CASE
        WHEN "Mission_Outcome" LIKE 'Success%' THEN 'Success'
        WHEN "Mission_Outcome" LIKE 'Failure%' THEN 'Failure'
        ELSE "Mission_Outcome"
    END AS Outcome,
    COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
GROUP BY Outcome;
```

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

Outcome	Outcome_Count
Failure	1
Success	100

- This query provides a summary of SpaceX's mission reliability by counting successful and failed missions. With 100 successful outcomes and only 1 failure, the results highlight SpaceX's high mission success rate, underscoring the effectiveness of their technology and operational protocols.

Boosters Carried Maximum Payload

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

```
%%sql
SELECT "Booster_Version"
FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (
    SELECT MAX("PAYLOAD_MASS__KG_")
    FROM SPACEXTABLE
);
```

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

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

- The output lists all booster versions that carried the maximum payload mass, including versions such as **F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3**, among others.
- These boosters represent SpaceX's highest payload-carrying missions, highlighting the specific booster versions that achieved this capacity.
- This query provides insights into the boosters with the highest payload capacity, useful for understanding which models in the Falcon 9 series can handle the largest missions. This information can assist in planning future missions that require high payload capacities and further emphasizes the capabilities of certain booster versions.

2015 Launch Records

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.

```
%%sql
SELECT
    CASE SUBSTR(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,
    "Booster_Version",
    "Launch_Site",
    "Landing_Outcome"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Failure (drone ship)'
AND SUBSTR(Date, 0, 5) = '2015';
```

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

Month	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

This query retrieves records that display the month names, booster versions, launch sites, and landing outcomes for failures on drone ships in the year 2015. The output shows:

- **January**: Booster version F9 v1.1 B1012 launched from CCAFS LC-40 with a landing outcome of Failure (drone ship).
- **April**: Booster version F9 v1.1 B1015 launched from CCAFS LC-40 with a landing outcome of Failure (drone ship).

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.

```
%%sql
SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY Outcome_Count DESC;
```

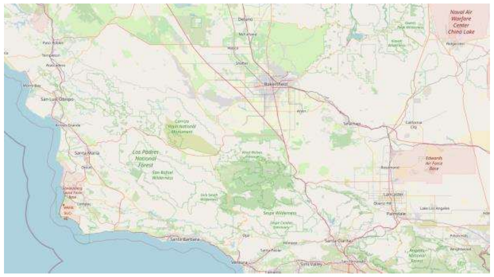
```
 * sqlite:///my_data1.db
Done.
```

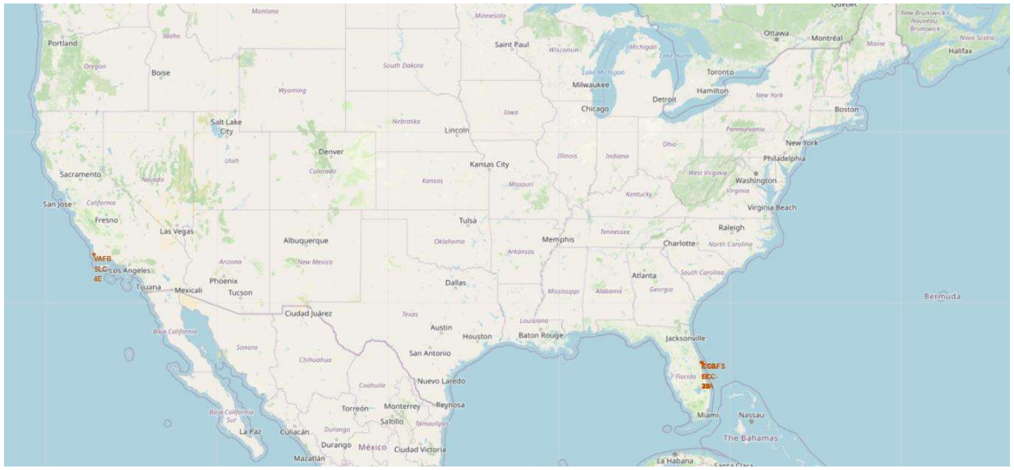
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- This query ranks the frequencies of various landing outcomes for SpaceX missions conducted between June 4, 2010, and March 20, 2017, in descending order.
- The results showcase SpaceX's advancements in achieving successful landings, particularly on drone ships and ground pads, while also identifying areas for improvement, such as minimizing failures on drone ships.

Section 3

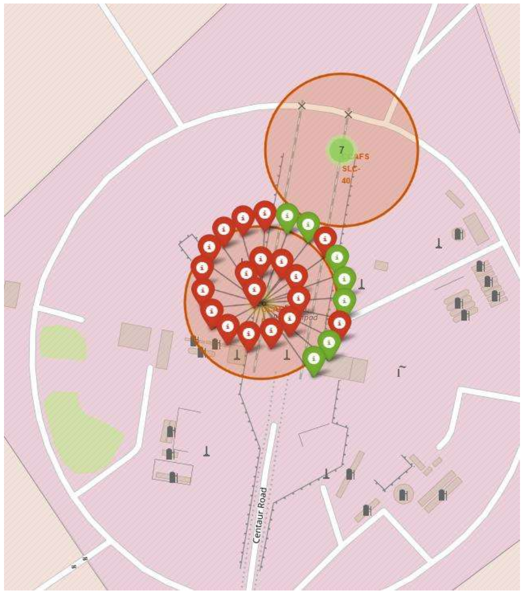
Launch Sites Proximities Analysis

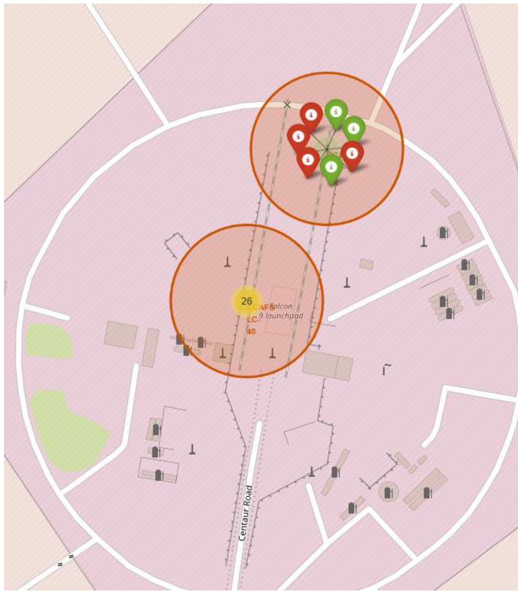


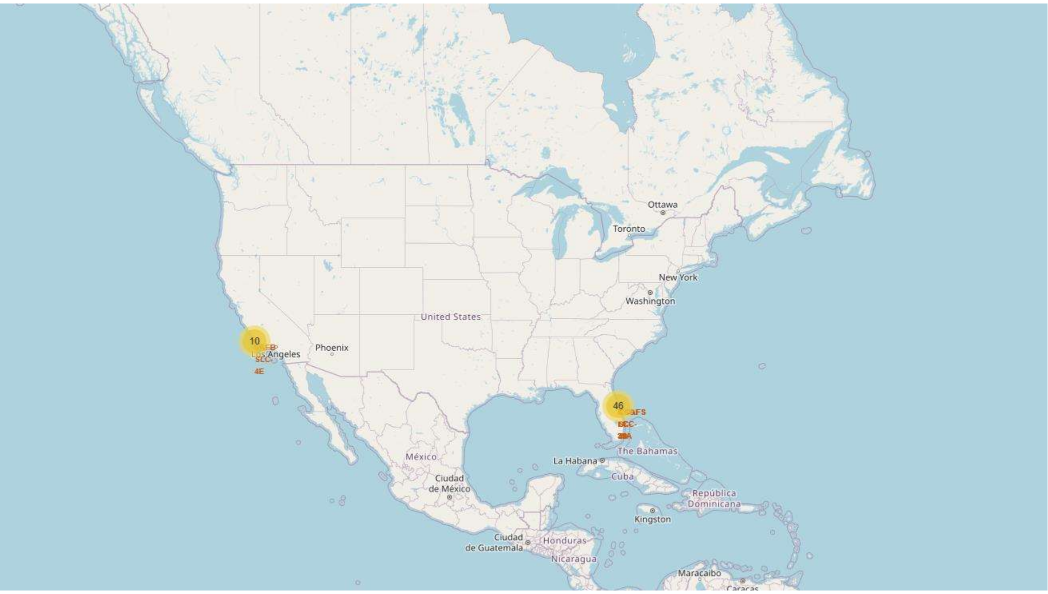


Key Insights from SpaceX Launch Site Locations

- **Geographical Distribution**:
  - All SpaceX launch sites are located in the United States, distributed along the **East Coast (Florida)** and **West Coast (California)**.
  - This distribution ensures flexibility in launch directions and orbits, such as polar orbits from the West Coast and equatorial orbits from the East Coast.
- **Proximity to Coastlines**:
  - All launch sites are situated near coastlines, allowing rockets to launch over open water, reducing risks to populated areas in case of a failure.
- **Clustering of East Coast Sites**:
  - A noticeable clustering of sites is observed in Florida (Top Right Photo: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A), an ideal location for equatorial launches due to its latitude and proximity to the Atlantic Ocean.
- **Strategic Placement of West Coast Site**:
  - **VAFB SLC-4E (Top Left Photo)** in California supports launches into polar orbits, which require trajectories over the Pacific Ocean.







Map Elements and Key Insights

Map Elements

- **Green Markers**: Successful launches.
- **Red Markers**: Failed launches.
- **Labels & Circles**: Identify and highlight launch sites.
- **Clustered Markers**: Group multiple launches for clarity.

Launch Site with Success Percentages
(Listed as Top Photo from Left to Right)

- **VAFB SLC-4E (California)**: 4/10 Successes (40%).
- **KSC LC-39A (Florida)**: 10/13 Successes (76.9%).
- **CCAFS LC-40 (Florida)**: 7/26 Successes (26.9%).
- **CCAFS SLC-40 (Florida)**: 3/7 Successes (42.9%).

Key Features and Findings for Launch Site CCAFS SLC-40

Map Elements

Popups

- Display the distance and the name of the point of interest.

Polyline Connections

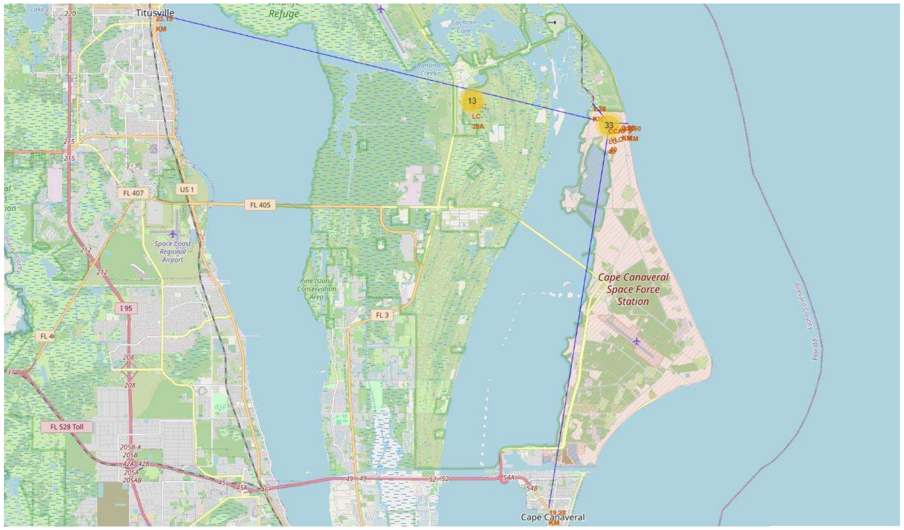
- **Blue Lines**: Visually connect the launch site to each point of interest.
- Highlight the spatial relationship between the launch site and surrounding infrastructure.

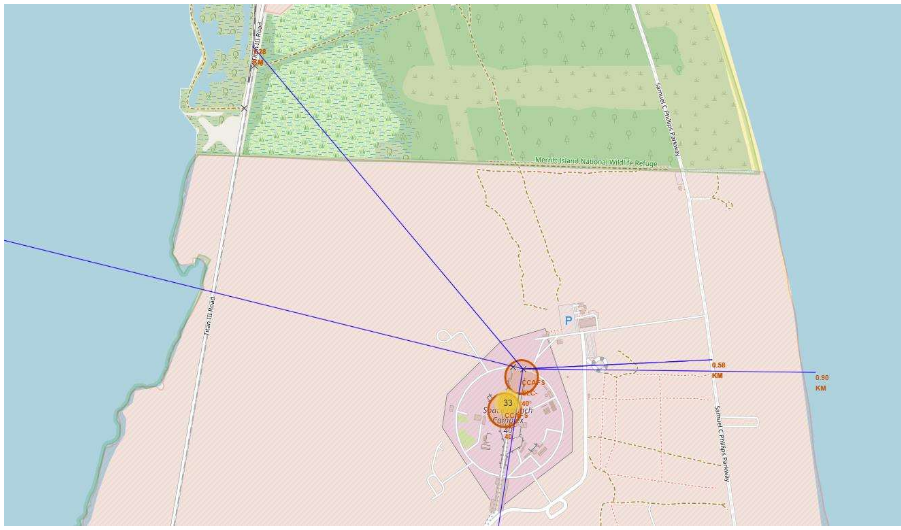
Proximity to Infrastructure

- Launch sites are strategically placed near vital infrastructure such as railways, highways, and coastlines to support operations and ensure safety.

Distances

- **Coastline**: 90 KM
- **Railway**: 1.28 KM
- **Highway**: 0.58 KM
- **City (Cape Canaveral):** 19.28 KM

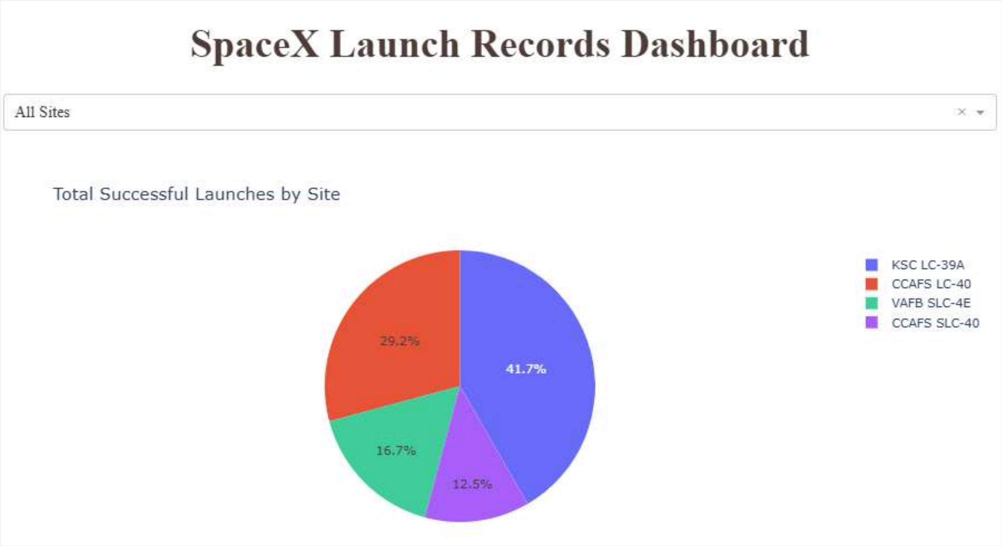




Section 4

Build a Dashboard with Plotly Dash

Total Successful Launches by Site



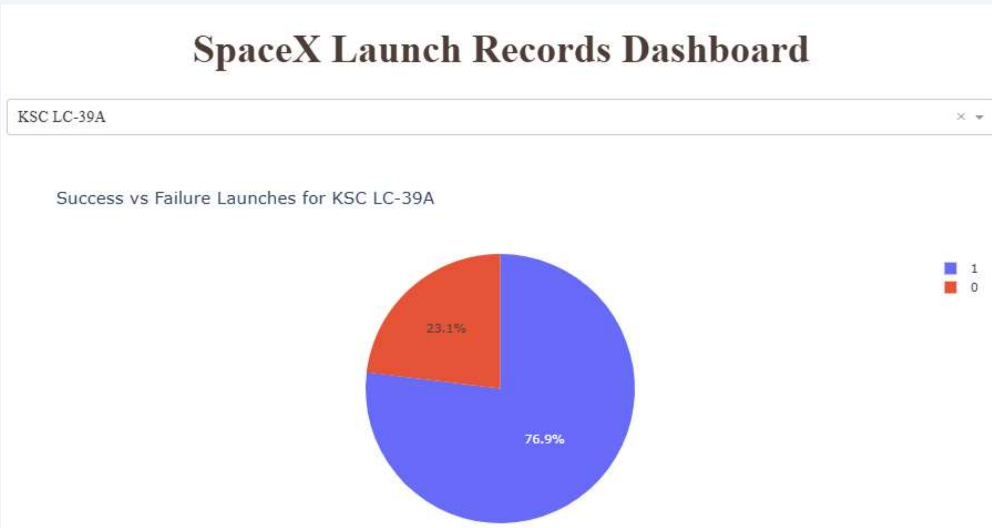
Key Insights:

- **KSC LC-39A (41.7%)**: Most productive site with the highest proportion of successful launches.
- **CCAFS LC-40 (29.2%)**: Second most significant contributor to success.
- **VAFB SLC-4E (16.7%)** and **CCAFS SLC-40 (12.5%)**: Lower contributions to overall success.

Importance:

- **Performance Analysis**: Identifies top-performing and underperforming sites.
- **Resource Optimization**: Guides resource allocation for improved performance.
- **Strategic Insights**: Basis for further exploration into site reliability and efficiency.

Success vs Failure Launches for KSC LC-39A



Success (class=1) vs. Failure (class=0)

Key Insights:

- **Success Rate:** KSC LC-39A achieves a **76.9% success rate**, making it a reliable launch site.
- **Failure Rate:** Only **23.1% of launches** have failed, showing strong performance consistency.

Importance:

- **Performance Analysis:** Highlights the high success rate of KSC LC-39A, reinforcing its reliability.
- **Strategic Focus:** Useful for prioritizing future launches from this site due to its strong performance metrics.

Payload and Booster Performance Insights

Key Insights:

First Plot (Payload: 0 to 2500 kg)

- **Failure Trends:** Higher failures below 1000 kg, linked to v1.0 and v1.1.
- **Consistent Success:** FT boosters perform well in 2000-2500 kg range.

Second Plot (Payload: 0 to 7500 kg)

- **Optimal Payload Range:** 2000-4000 kg with high success rates.
- **Best Boosters:** FT and B4 dominate successful launches.
- **Failures:** More frequent at <2000 kg, especially with v1.0 and v1.1 boosters.

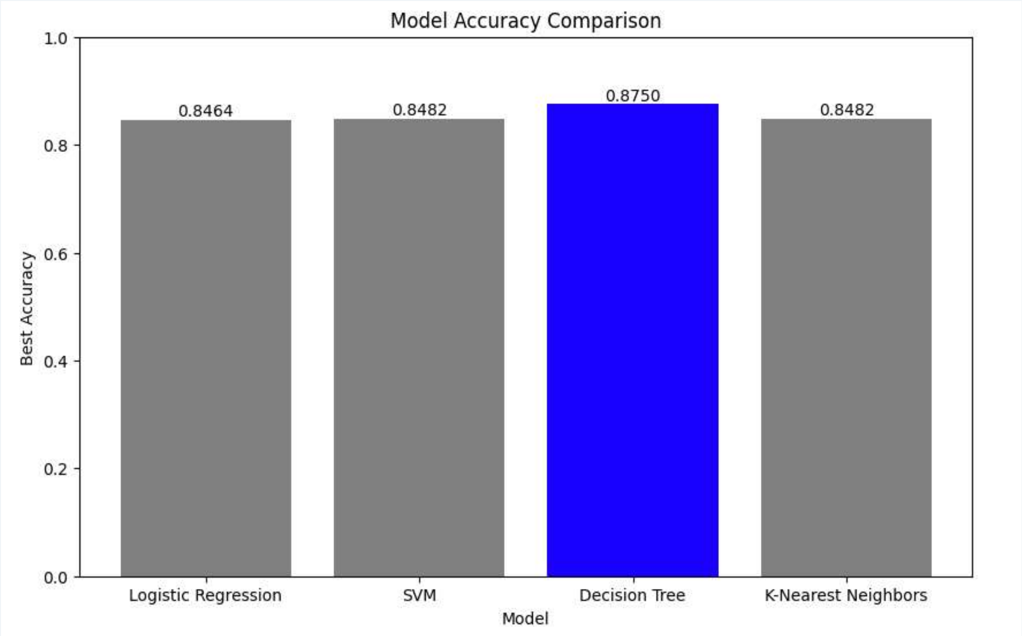
Takeaway:

- Focus on payloads **2000-4000 kg** and boosters **FT/B4** for better success rates.
- Address challenges for lightweight missions (<1000 kg).

Section 5

Predictive Analysis (Classification)

Classification Accuracy



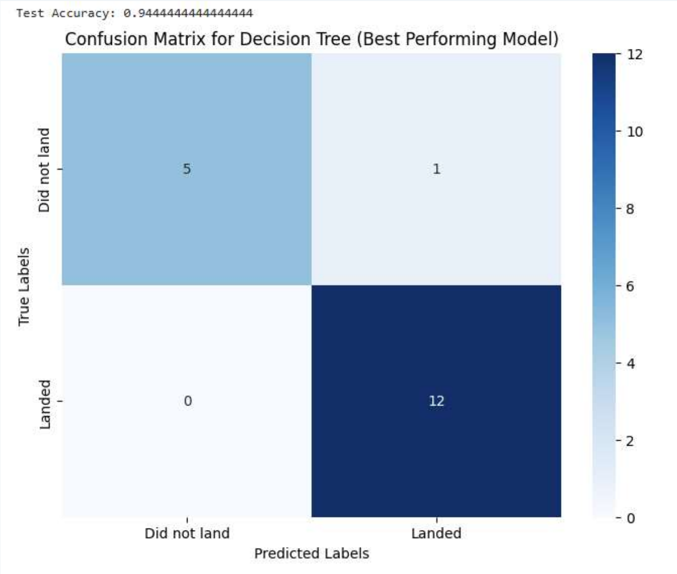
The bar chart shows the accuracy of all the built classification models:

- **Logistic Regression**: Accuracy = 0.8464
- **Support Vector Machine (SVM)**: Accuracy = 0.8482
- **Decision Tree**: Accuracy = 0.8750
- **K-Nearest Neighbors (KNN)**: Accuracy = 0.8482
- Best Performing Model
- The **Decision Tree** model achieved the highest classification accuracy of **0.8750**.

Explanation of Results

- The Decision Tree model outperformed the other models in terms of accuracy, likely due to its ability to handle non-linear relationships and capture complex interactions in the dataset. Logistic Regression, SVM, and KNN also showed competitive performance, but their accuracies were slightly lower, possibly due to their limitations in capturing non-linear patterns.
- This visualization highlights the Decision Tree model as the most effective for this classification task, making it the best choice for predicting Falcon 9 first-stage landings.

Confusion Matrix



Explanation of the Results

- **True Positives (TP)**: 12
  - The model correctly predicted "Landed" when the actual label was "Landed."
- **True Negatives (TN)**: 5
  - The model correctly predicted "Did not land" when the actual label was "Did not land."
- **False Positives (FP)**: 1
  - The model predicted "Landed" when the actual label was "Did not land."
- **False Negatives (FN)**: 0
  - The model did not misclassify any cases where the actual label was "Landed."

Insights

- The **Decision Tree** model demonstrated strong predictive capability with **no false negatives**, indicating that all successful landings were correctly identified.
- The model showed only **one false positive**, where a case was incorrectly predicted as "Landed" when it did not land.
- Overall, the confusion matrix highlights the Decision Tree's high accuracy and its ability to minimize misclassification for this binary classification task.

Conclusions

Key Insights & Recommendations

- **KSC LC-39A (Florida)** is the most reliable launch site with a 76.9% success rate for payloads between 2,000–7,000 kg, ideal for equatorial missions. **VAFB SLC-4E (California)** showed moderate success (40%) for polar orbits, while **CCAFS LC-40 (Florida)** had lower success (26.9%) with inconsistent performance. **GTO orbit** remains the most challenging, with the lowest success rate.
- The **Decision Tree model** achieved the highest accuracy (87.5%), with no false negatives and minimal errors, outperforming Logistic Regression and SVM/KNN. Recommendations include prioritizing **KSC LC-39A** for mid-range payloads, addressing GTO challenges, and using the Decision Tree model for accurate landing predictions and mission planning.

Appendix

- GitHub Repository: https://github.com/haleo9/IBM_Applied_Data_Science_Capstone.git


Thank you!