IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ashkan Nikiyan
10th March 2025

Outline

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

Executive Summary

- **Summary of methodologies**

  - Collected data through SpaceX API and web scraping techniques.
  - Performed data wrangling and preprocessing to clean and structure the data.
  - Conducted exploratory analysis using SQL queries and Python visualization tools.
  - Developed interactive maps and dashboards for enhanced data analytics.
  - Applied machine learning classification algorithms for predictive analysis.

- **Summary of all results**

  - Key insights derived from Exploratory Data Analysis (EDA).
  - Findings from interactive dashboards and geographic visualizations.
  - Evaluation metrics and interpretation of predictive models.
  - Identification of the most effective classification model for launch success prediction.

Introduction

Project Background and Context:

- SpaceX is renowned for its advancements in reusable rocket technology, particularly with the Falcon 9 series.
- Falcon 9 rockets have demonstrated high reliability in launching payloads into space, although occasional failures still occur.
- This Applied Data Science Capstone project focuses on predicting the success of Falcon 9 first-stage landings using classification models.
- The main objective is to determine whether the first stage of a rocket can be reused based on various launch factors.

Problems to Address:

- How does the payload size influence the launch success rate?
- Does the launch site location impact the outcome of the rocket launch?
- What role does the orbit type play in determining launch success?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was sourced from the SpaceX API and web scraping techniques. API calls were used to extract structured launch data, while web scraping was employed to gather additional insights.
- Perform data wrangling
  - Data was cleaned and formatted to ensure consistency and remove missing values. Transformation techniques were applied to enhance data usability for analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Conducted data exploration using SQL queries and Python visualization libraries. Generated statistical summaries, distribution plots, and trend analyses.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Developed Folium-based interactive maps to visualize launch site locations and success rates. Created Plotly Dash dashboards for an interactive exploration of key metrics.
- Perform predictive analysis using classification models
  - Implemented machine learning classification models to predict first-stage landing success. Performed hyperparameter tuning and model evaluation to identify the best-performing model.

Data Collection

Data Collection Process:

- Data was sourced using two primary methods:
    1. SpaceX API Calls – Extracted structured data related to rocket launches.
    2. Web Scraping – Retrieved additional launch details from Wikipedia.

Web Scraping Process:

- Tools Used: requests, BeautifulSoup, and pandas libraries.
- Steps:
    - Requested data from Wikipedia using requests.get().
    - Created a BeautifulSoup object to parse the HTML response.
    - Extracted all tables using soup.find_all().
    - Retrieved the relevant table and iterated through it to extract key headers and values.
    - Constructed a dataframe containing:
        - Flight No, Launch Site, Payload, Payload Mass, Orbit, Customer, Launch Outcome, Booster Version, Booster Landing, Date, and Time.

Data Collection – SpaceX API

Process of Data Collection via SpaceX API:

- Utilized SpaceX REST API to retrieve real-time rocket launch data.
- Sent API requests using requests library in Python.
- Parsed JSON responses to extract key launch parameters such as:
    - Flight Number, Launch Site, Payload Mass, Orbit, Booster Version, Landing Outcome, and Date.
- Converted extracted data into a structured Pandas DataFrame for further analysis.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%201%20-%20Data%20Collection%20with%20API/jupyter-labs-spacex-data-collection-api.ipynb

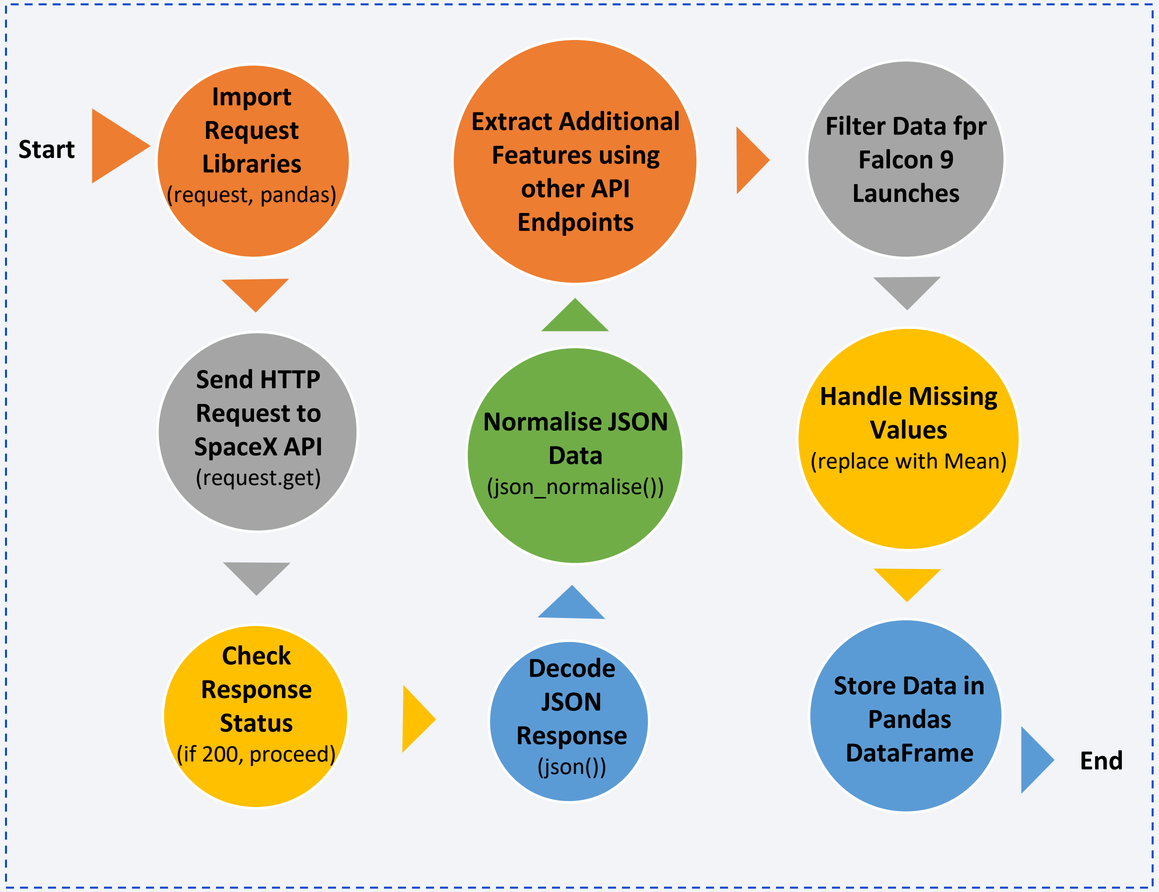

Figure 1: Flowchart of the Data Collection Process Using the SpaceX API

Data Collection - Scraping

- The flowchart in Figure 2 illustrates the web scraping process using Python's requests and BeautifulSoup libraries.
- Extracted structured data from Wikipedia by parsing HTML tables.
- The cleaned data was stored in a Pandas DataFrame and saved as a CSV file using:

```
data_falcon9.to_csv("spacex_falcon9_data.csv",
index=False)
```

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%201%20-%20Data%20Collection%20with%20Web%20Scraping/jupyter-labs-webscraping.ipynb

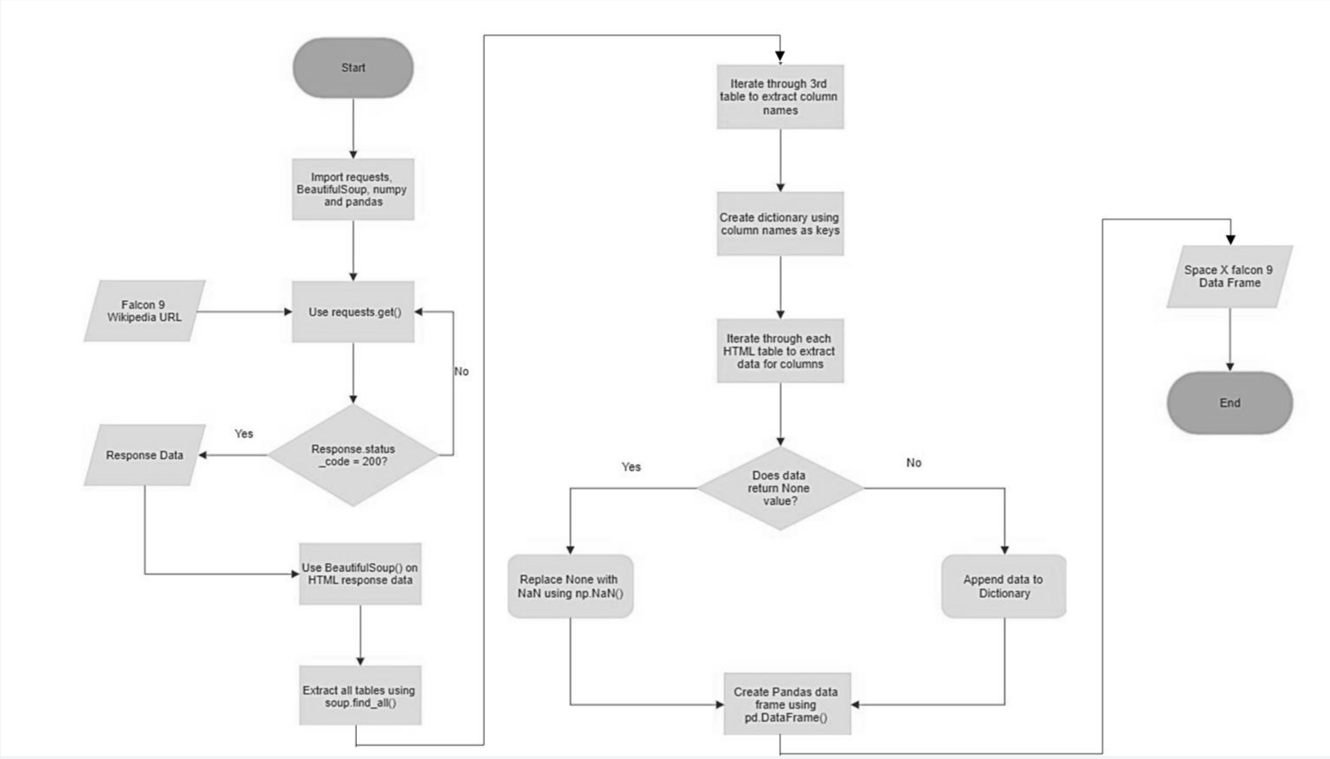

Figure 2: Flowchart of the Web-Based Data Collection Process

Data Wrangling

- Performed Exploratory Data Analysis (EDA) to understand the dataset structure and identify missing values.
- Converted the Outcome column into binary classification labels:
    - 1 → Successfully landed
    - 0 → Otherwise
- Figure 3 illustrates the data wrangling process using a structured flowchart.
- The cleaned dataset was prepared for machine learning modeling and visualization.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%201%20-%20Data%20Wrangling/labs-jupyter-spacex-Data%20wrangling.ipynb

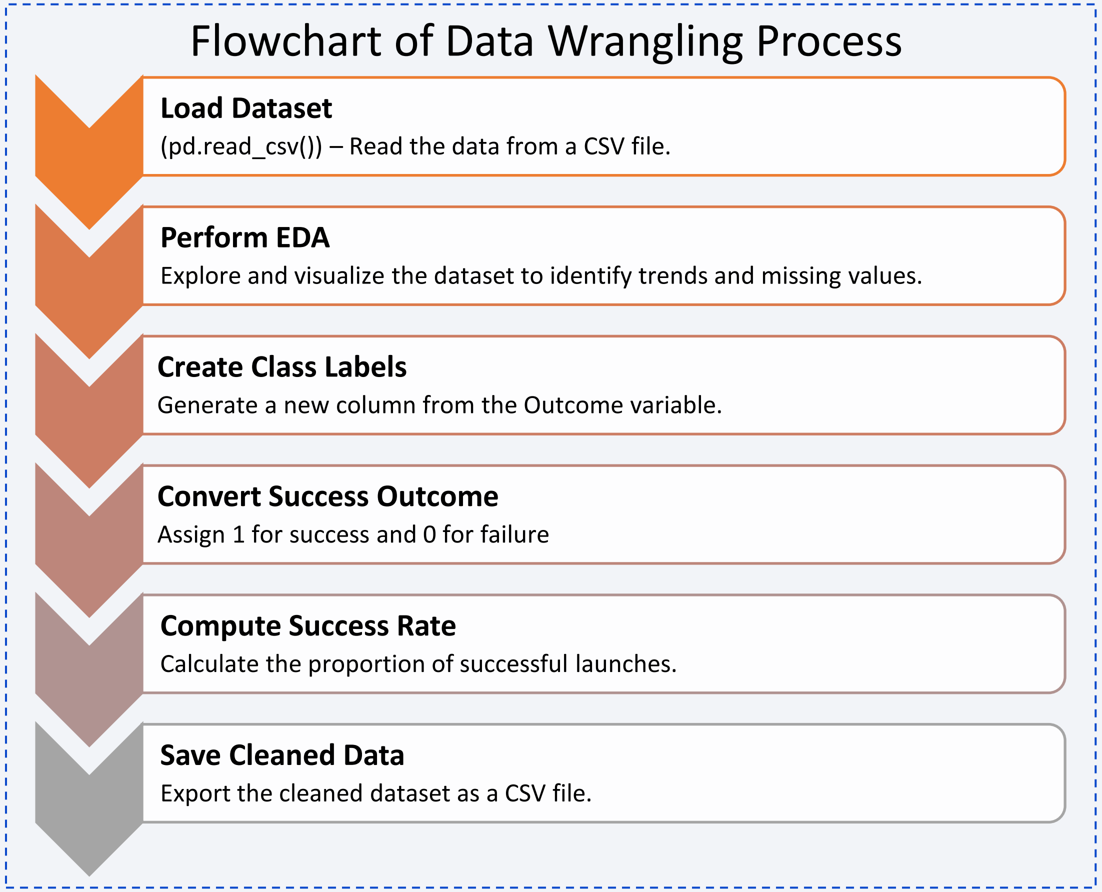

Figure 3: Flowchart of the Data Wrangling Process

EDA with Data Visualization

- Scatter plots were used to analyze how different variables affect launch outcomes:
  - Payload mass vs. Flight number
  - Flight number vs. Launch site
  - Payload mass vs. Launch site
- Bar chart was used to visualize the success rate of each orbit type.
- Scatter plots explored relationships between:
  - Flight number and Orbit type
  - Payload mass and Orbit type
- Line plot was used to track the yearly trend of launch successes.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%202%20-%20EDA%20with%20Visualisation/edadataviz.ipynb

EDA with SQL

Exploratory data analysis was conducted using SQL queries to extract insights from the dataset. The analysis focused on understanding launch site distributions, payload trends, mission success rates, and booster performance. Various SQL queries were executed to filter, aggregate, and rank the data based on key parameters.

Some of the key analyses performed include:

- Displaying the **unique launch sites** involved in space missions.
- Filtering launch sites that **start with ‘CCA’** to examine specific launch locations.
- Calculating the **total payload mass** carried by boosters launched by **NASA (CRS)**.
- Computing the **average payload mass** carried by booster version **F9 v1.1**.
- Identifying the **first successful landing outcome** on a ground pad.
- Displaying successful boosters that landed on a **drone ship** and carried a **payload between 4000kg and 6000kg**.
- Summarizing the **total number of successful and failed mission outcomes**.

EDA with SQL

Additional SQL-Based Analysis:

To gain deeper insights into launch patterns and booster performance, the following analyses were performed:

- **Identifying booster versions** that have carried the **maximum payload mass**.
- **Examining failed landing outcomes** in drone ships during the year **2015**, including **month names, booster versions, and launch sites**.
- **Ranking successful landing outcomes** between **June 4, 2010, and March 20, 2017**, in descending order.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%202%20-%20EDA%20with%20SQL/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

To visualize launch sites and their geographical surroundings, Folium was used to create an interactive map. Several Folium objects were implemented to enhance interactivity and provide useful insights.

Folium Map Features Used:

folium.Circle – Added a highlighted circle marker at NASA JSC as the initial center location.folium.map.Marker – Placed markers at specific launch locations to display their exact positions.

MarkerCluster() – Created cluster markers for successful and failed launches at specific sites.MousePosition() – Displayed latitude and longitude coordinates dynamically based on cursor movement. Used to calculate the distance from launch sites to the coast.

folium.PolyLine() – Drew connected line segments on the map to mark distances from launch sites to coasts, railways, highways, and major cities.

These interactive elements enhanced spatial understanding of launch site locations and their proximity to essential infrastructures.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%203%20-%20Visual%20Analytics%20with%20Folium/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

A **Plotly Dash interactive dashboard** was developed to analyze **launch success rates** and **payload correlations** dynamically. Users can interact with different filters to gain insights.

Dashboard Features:

- **Drop-down Input** – Select a **launch site** to filter data.
- **Pie Chart** – Displays **success rates** based on the selected launch site.
- **Range Slider** – Adjust the **payload mass range** dynamically.
- **Scatter Plot** – Shows **payload mass vs. launch success** correlation.

The dashboard provides **real-time visual insights**, allowing users to explore launch trends efficiently.

GitHub Reference:

https://github.com/anikiyan/SpaceX_Falcon-9_First_stage_landing_prediction/blob/main/Week%203%20-%20Dashboard%20with%20Ploty%20Dash/Dahsboard__Ploty_Dash.ipynb

Predictive Analysis (Classification)

The process of building and evaluating classification models involved several key steps. First, the dataset was loaded and divided into features and target variables. To ensure consistency, feature columns were normalized, and the target variable was converted into a NumPy array.

Next, the data was split into training and test sets, allowing for proper evaluation of model performance. Hyperparameter tuning was performed using GridSearchCV, which helped determine the best parameters and scores to optimize classification accuracy.

Once the model was trained, its accuracy was assessed on the test set using the .score() method. Finally, a confusion matrix was plotted to visualize classification performance and analyze misclassifications.

📌 Key Steps in Model Development:

- Data Loading & Splitting – Features and target variables were extracted.
- Normalization – Feature columns were standardized, and the target variable was converted to a NumPy array.
- Train-Test Split – Data was divided into training and test sets.
- Hyperparameter Tuning – Used GridSearchCV to identify the best parameters (.best_params_, .best_score_).
- Model Evaluation – Accuracy was calculated using the .score() method.
- Confusion Matrix – Visualized model predictions to assess performance.

📌 Figure 4: Flowchart illustrating the model building and evaluation process.

Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

From the analysis, it is evident that **CCAFS SLC 40** has recorded the **highest number of rocket launches** compared to other sites. This suggests that it is a **primary launch hub** for space missions.
Additionally, trends indicate that **later flights from VAFB SLC 4E and KSC LC 39A** have demonstrated a **higher success rate** than earlier flights. This improvement could be attributed to **technological advancements and operational optimizations** over time.

📌 **Key Observations:**

- **CCAFS SLC 40** has the **highest number of launches** among all sites.
- **Success rates improved in later flights** from **VAFB SLC 4E and KSC LC 39A**, likely due to **better engineering and mission enhancements**.

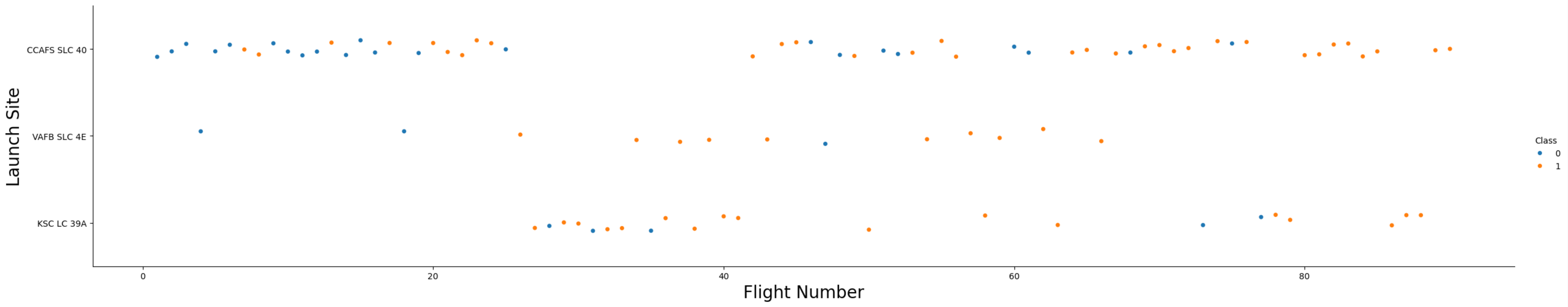

Figure 4. Scatter plot of Flight Number Vs. Launch Site

Payload vs. Launch Site

The analysis reveals that **VAFB-SLC 4E has no rocket launches** for payloads exceeding **10,000 kg**, indicating that this site primarily handles **lighter payload missions**.
Additionally, across all launch sites, the majority of rockets carry **payloads below 9,000 kg**, suggesting that **lighter payload launches are more common** in the dataset.
Furthermore, when comparing **CCAFS SLC 40** to **VAFB-SLC 4E and KSC LC 39A**, it is evident that **CCAFS SLC 40 has a higher success rate** for **heavier payloads (14,000 kg - 16,000 kg)**, making it more reliable for **high-mass missions**.

📌 **Key Observations:**

- **VAFB-SLC 4E has no launches** for payloads greater than **10,000 kg**.
- **Most rockets launched** at all sites carry a **payload of less than 9,000 kg**.
- **CCAFS SLC 40 has the highest success rate** for **heavier payloads (14,000 kg - 16,000 kg)** compared to other sites.

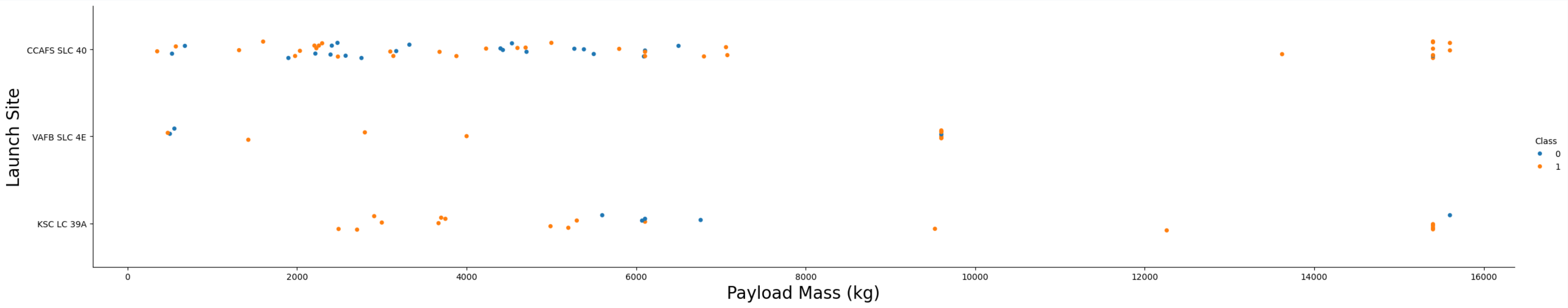

Figure 5. Scatter plot of Payload Vs. Launch Site

Success Rate vs. Orbit Type

The analysis indicates that **orbits VLEO, ES-L1, GEO, HEO, and SSO** have the **highest success rates**, making them more reliable for space missions. These orbits are likely preferred for their **stability and favorable conditions for successful deployments**. Conversely, **orbit SO** has the **lowest success rate**, suggesting that launches into this orbit face **higher challenges or lower mission reliability**.

📌 **Key Observations:**

- **VLEO, ES-L1, GEO, HEO, and SSO** have the **highest success rates** among all orbit types.
- **Orbit SO has the least success rate**, indicating possible **higher risks or mission challenges** in this orbit.

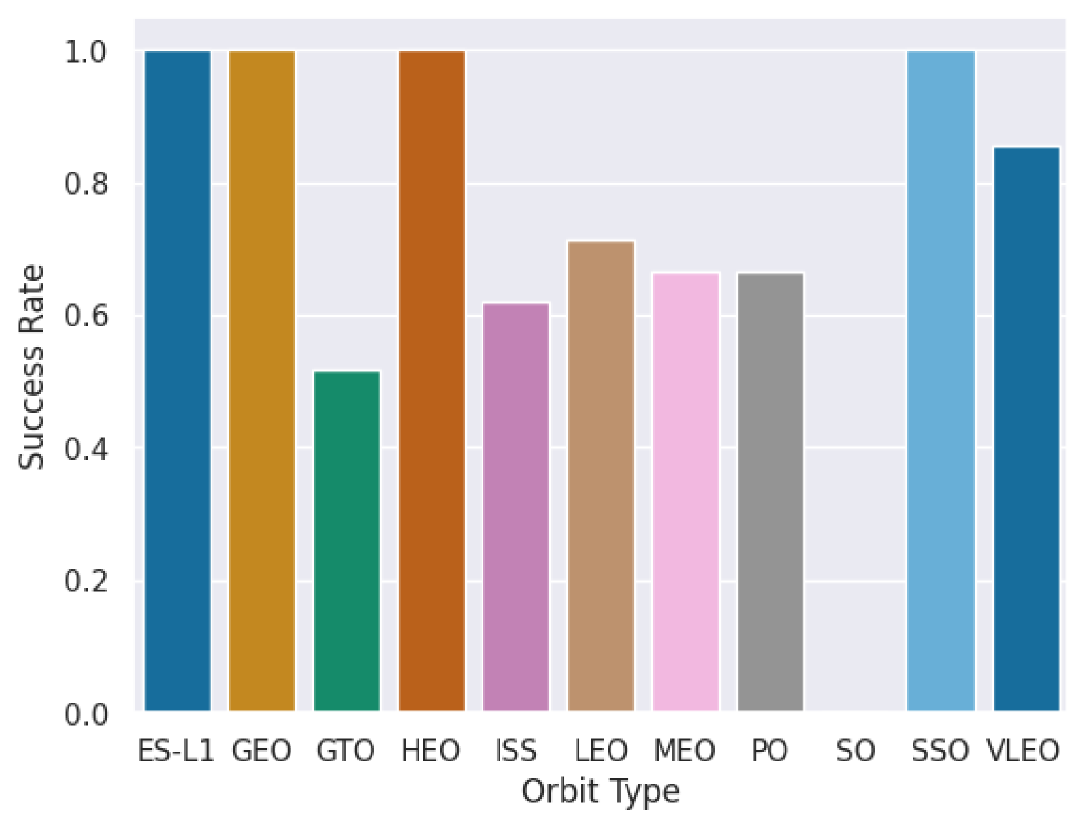

Figure 6. Bar plot of Success rate Vs. Orbit Type

Flight Number vs. Orbit Type

The analysis reveals that **LES, ISS, PO, GTO, and VLEO** have recorded the **highest number of rocket launches** compared to other orbit types. This suggests that these orbits are **preferred destinations for various space missions**.

Additionally, in **LEO (Low Earth Orbit), success rates appear to be correlated with the number of flights**, indicating that **higher launch frequency may contribute to improved mission success**. However, for **other orbits, there is no clear relationship** between **flight number and success rate**, suggesting that **other factors influence mission outcomes** beyond just the number of flights.

📌 **Key Observations:**

- **LES, ISS, PO, GTO, and VLEO** have the **highest number of rocket launches**.
- **LEO orbit success rates** appear to be **positively influenced by the number of flights**.
- **No clear relationship** exists between **flight number and success rate** for other orbit types.

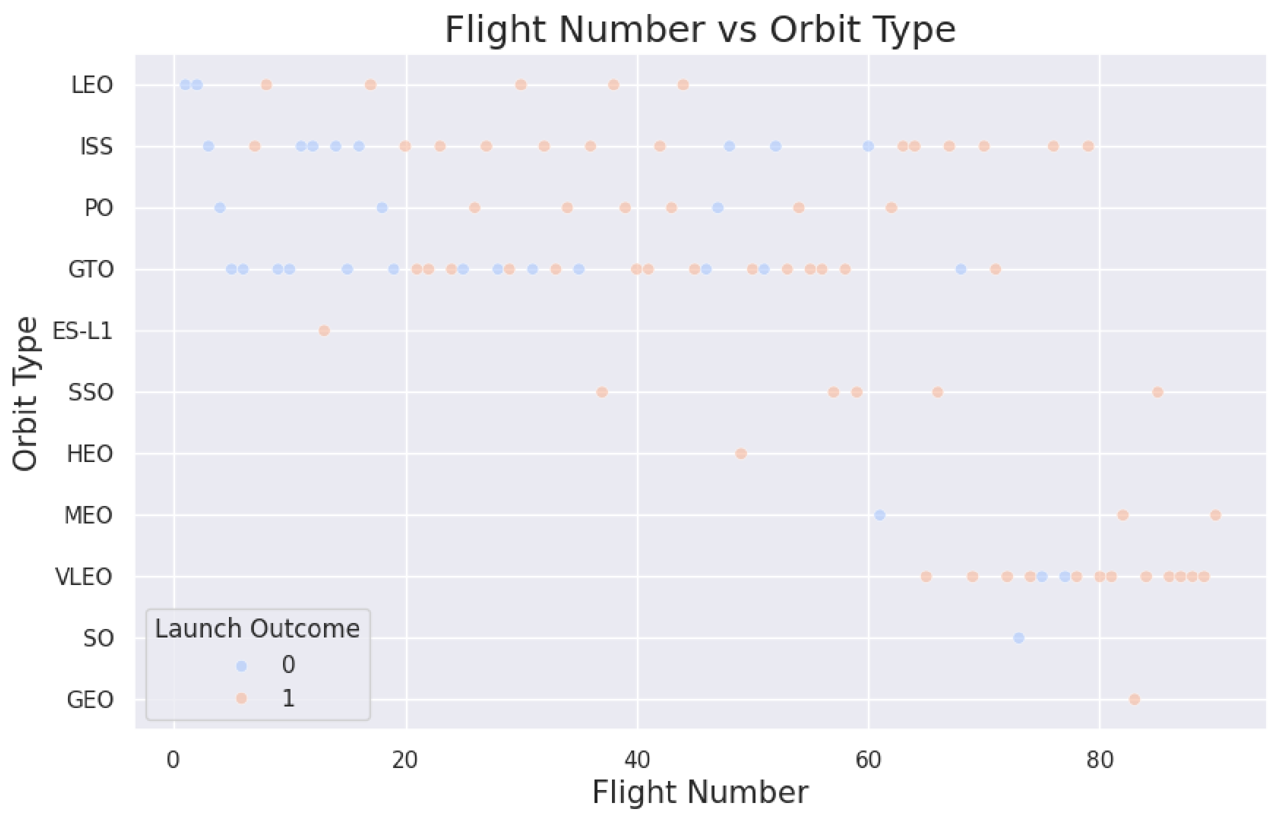

Figure 7. Scatter plot of Flight Number Vs. Orbit type

Payload vs. Orbit Type

The analysis indicates that **rockets with heavier payloads** tend to have a **higher success rate** when launched into **PO, LEO, and ISS** orbits. This suggests that these orbits are well-suited for handling **larger payload missions**.

Conversely, **SSO and MEO orbits** show **higher success rates with lighter payloads**, indicating that these orbits may be **optimized for smaller satellite deployments**.

Additionally, **rockets launched in GTO (Geostationary Transfer Orbit) exhibit both positive and negative landing outcomes**, regardless of payload size. This suggests that **other factors beyond payload mass** influence landing success in GTO missions.

📌 **Key Observations:**

- **PO, LEO, and ISS** support **heavier payloads** with **higher success rates**.
- **SSO and MEO** show **better success rates with lighter payloads**.
- **GTO launches have mixed landing results,** independent of payload size

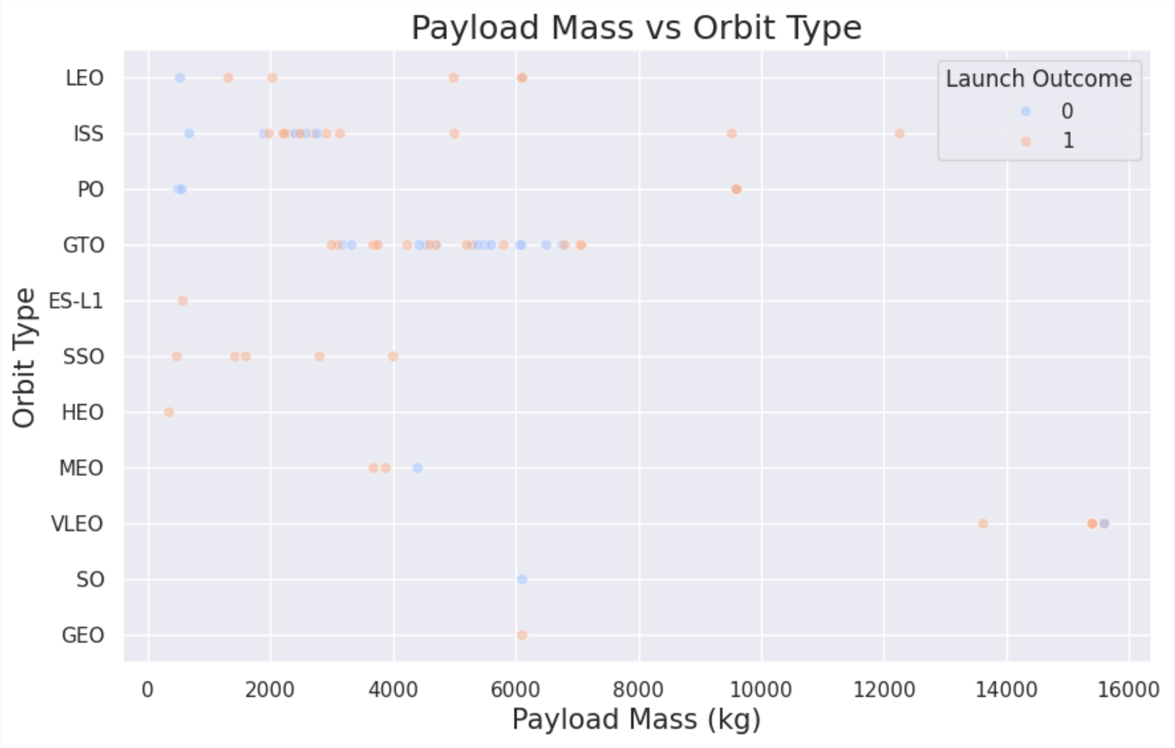

Figure 8: Scatter Plot of Payload vs. Orbit Type

Launch Success Yearly Trend

The analysis in **Figure 10** reveals a **consistent increase in launch success rates** from **2013 to 2020**. This upward trend suggests **advancements in rocket technology, improved engineering processes, and enhanced mission execution strategies**, contributing to a **higher rate of successful launches over time**.

📌 **Key Observations:**

- **Steady improvement** in launch success rates, indicating **increased reliability**.
- **Technological advancements** and **better operational strategies** have contributed to higher success rates.
- **Continuous optimizations in rocket design** and **mission planning** have minimized failures over the years.

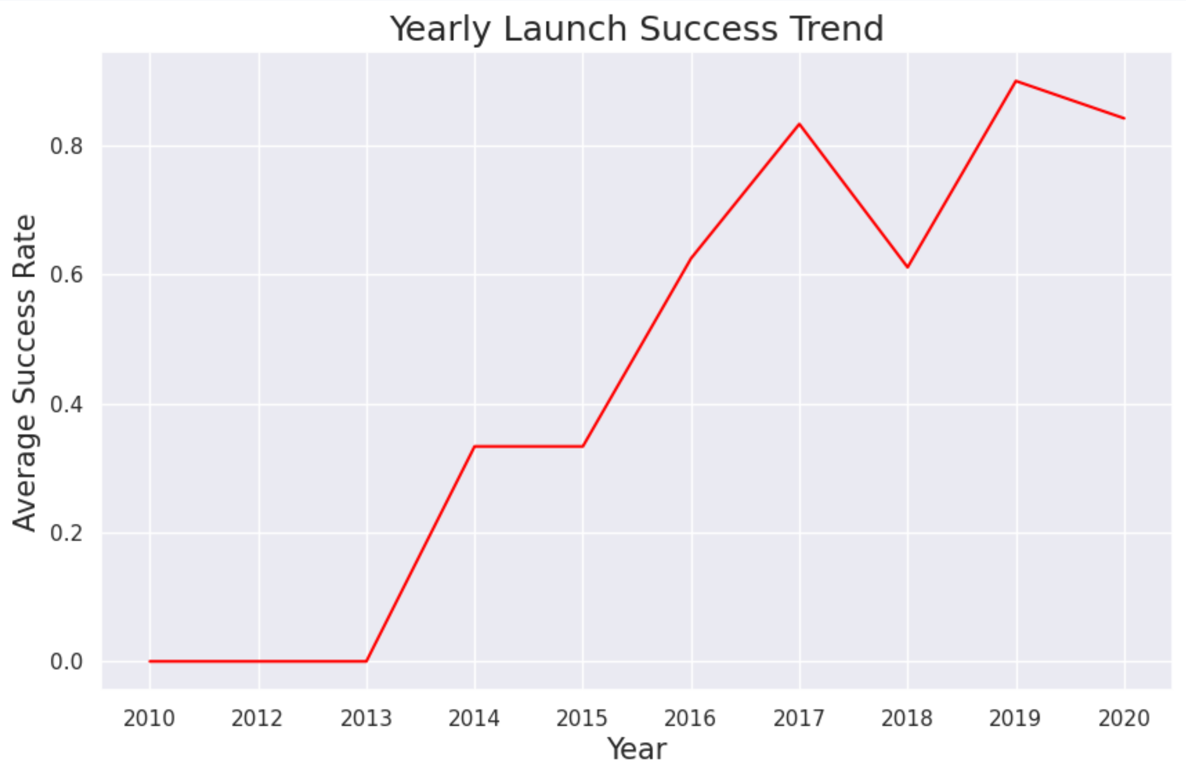

Figure 9: Line Plot of Yearly Launch Success Trend

All Launch Site Names

An SQL table named SPACEXTBL was created using the existing DataFrame to store and analyze launch data.

📌 Key Steps:

- The DISTINCT keyword was applied to the launch site column to extract unique launch site names.
- This approach ensured that each launch site appeared only once in the results, helping identify all locations used for rocket launches.

In [15]:

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

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

Out[15]:

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

Figure 10. SQL query for unique launch site names

Launch Site Names Begin with 'CCA'

To retrieve launch site names that start with "CCA", an SQL query was executed using specific filtering conditions.

📌 Key Steps:

- The LIKE 'CCA%' keyword was used to filter launch site names that begin with 'CCA’.
- The LIMIT 5 clause was applied to display only the first five records matching the condition.

```
In [17]: %%sql
             SELECT *
             FROM SPACEXTBL
             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

Figure 11. SQL query for Launch site names beginning with CCA

Total Payload Mass

To calculate the total payload mass for missions conducted by NASA (CRS), the SUM function was applied in the SQL query.

📌 Key Steps:

- The SUM function was used to aggregate the payload mass values.
- A WHERE clause filtered the data to include only records where the customer name is 'NASA (CRS)'.

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

In [19]:
```
%%sql
    SELECT SUM(PAYLOAD_MASS__KG_) AS Total_payload_NASA_CRS
    FROM SPACEXTBL
    WHERE Customer = 'NASA (CRS)'
```

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

Out[19]:

Total_payload_NASA_CRS
45596

Figure 12: SQL Query for Calculating Total Payload Mass for NASA (CRS)

Average Payload Mass by F9 v1.1

The average payload mass for booster version F9 v1.1 was determined using the AVG function in SQL.

📌 Key Steps:

- The AVG function was applied to calculate the mean payload mass.
- A WHERE clause was used to filter records for booster version F9 v1.1.

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

In [20]:
```
%%sql
    SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_F9V1_1
    FROM SPACEXTBL
    WHERE Booster_Version LIKE 'F9 v1.1%'
```

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

Out[20]:

Average_Payload_F9V1_1
2534.6666666666665

Figure 13: SQL Query for Calculating Average Payload Mass for F9 v1.1

First Successful Ground Landing Date

An SQL query was executed to identify the first successful landing on a ground pad.

📌 Key Findings:

- The query returned the date December 22, 2015, as the first recorded successful ground landing.

📌 Key Steps:

- A filter condition was applied to extract only successful ground landings.
- The MIN function was used to retrieve the earliest landing date from the dataset.

```
In [27]: %%sql
             SELECT MIN(Date) AS first_successful_landing_date
             FROM SPACEXTBL
             WHERE Landing_Outcome = 'Success (ground pad)';

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

Out[27]:

first_successful_landing_date
2015-12-22

Figure 14: SQL Query for First Successful Ground Landing Date

Successful Drone Ship Landing with Payload between 4000 and 6000

An SQL query was executed to retrieve the names of boosters that successfully landed on a drone ship and carried a payload mass between 4000kg and 6000kg.

📌 Key Steps:

- The BETWEEN keyword was used to filter payload mass within the range of 4000kg to 6000kg.
- The AND operator ensured that only successful drone ship landings were included.

📌 Key Findings:

- The query result identified four rockets that met the specified conditions.

```
In [29]: %%sql
             SELECT Booster_Version FROM SPACEXTBL
             WHERE Landing_Outcome = 'Success (drone ship)'
             AND
             PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000

          * sqlite:///my_data1.db
         Done.
Out[29]:
```

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

Figure 15: SQL Query for Successful Drone Ship Landings Between 4000kg and 6000kg

Total Number of Successful and Failure Mission Outcomes

An SQL query was executed to determine the total number of successful and failed missions using the COUNT function.

📌 Key Steps:

- The COUNT function was applied to count the occurrences of successful and failed mission outcomes.
- A GROUP BY clause was used to separate results based on mission status.

📌 Key Findings:

- The query results show 100 successful missions and 1 failed mission.

In [31]:
```
%%sql
    SELECT COUNT(Mission_Outcome)
    AS Success_missions
    FROM SPACEXTBL
    WHERE Mission_Outcome LIKE '%Success%'
```

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

Out[31]:

Success_missions
100

In [32]:
```
%%sql
    SELECT COUNT(Mission_Outcome)
    AS Failure_missions
    FROM SPACEXTBL
    WHERE Mission_Outcome LIKE '%Failure%'
```

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

Out[32]:

Failure_missions
1

Figure 16: SQL Query for Total Number of Successful and Failed Mission Outcomes

Boosters Carried Maximum Payload

An SQL query was executed to identify the boosters that carried the highest payload mass using a subquery with the MAX function`.

📌 Key Steps:

- The MAX function was used to determine the highest payload mass in the dataset.
- A subquery was implemented to filter boosters that match this maximum payload value.

📌 Key Findings:

- The results show that 12 boosters carried the maximum payload recorded.

In [34]:

```
%%sql
    SELECT  Booster_Version, PAYLOAD_MASS__KG_
    FROM SPACEXTBL
    WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

Out[34]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

Figure 17: SQL Query for Boosters That Carried the Maximum Payload

2015 Launch Records

An SQL query was executed to retrieve launch records for failed drone ship landings in 2015 using date extraction and filtering techniques.

📌 Key Steps:

- The SUBSTR() function was used to extract the month and year from the Date column.
- The WHERE and AND keywords were applied to filter records for failed landings specifically in 2015.

📌 Key Findings:

- The results indicate that failed drone ship landings occurred in the months of April (04) and October (10) in 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.

In [40]:

```
%%sql
SELECT
    substr(Date, 6, 2) AS Month,
    Booster_Version,
    Landing_Outcome,
    Launch_Site
FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)'
  AND substr(Date, 0, 5) = '2015';
```

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

Out[40]:

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

Figure 18: SQL Query for Failed Drone Ship Landings in 2015

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

An SQL query was executed to rank landing outcomes between June 4, 2010, and March 20, 2017, using aggregation and sorting techniques.

📌 Key Steps:

- COUNT() was used to count occurrences of each landing outcome.
- SUBSTR() was applied to extract relevant date components.
- GROUP BY was used to categorize outcomes.
- ORDER BY DESC ensured results were ranked in descending order.

📌 Key Findings:

- Highest occurrences:
  - No Attempt (10)
  - Successful Landings (Drone Ship: 5, Ground: 5)
- One recorded failure due to parachute deployment malfunction.

```
In [43]:  %%sql
            SELECT "Landing_Outcome", COUNT(*) AS landing_count
            FROM SPACEXTBL
            WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
            GROUP BY "Landing_Outcome"
            ORDER BY landing_count DESC;

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

Out[43]:

Landing_Outcome	landing_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

Figure 19: SQL Query for Ranking Landing Outcomes Between 2010-06-04 and 2017-03-20

Section 3

Launch Sites Proximities Analysis

Launch Site Locations for SpaceX Falcon 9

The **Folium map** visualization highlights the **geographical locations of all SpaceX Falcon 9 launch sites**. The analysis reveals that **all launch sites are strategically positioned in coastal cities across the United States**, allowing for safer launches and efficient payload deployment.

📌 **Key Insights:**

- **Proximity to coastlines** minimizes risks and enhances recovery operations.
- **Strategic locations** enable optimized flight trajectories for various orbital missions.

Figure 20: Folium Map Displaying SpaceX Falcon 9 Launch Site Locations

SpaceX Falcon 9 Launch Outcomes by Site

The **Folium map** visualization illustrates the **launch outcomes across different SpaceX Falcon 9 launch sites**. The map provides a **clear distinction between successful and failed launches** using color-coded markers.

📌 **Launch Sites and Their Positions on the Map:**

- **Top Left:** VAFB SLC-4E
- **Top Right:** KSC LC-39A
- **Bottom Left:** CCAFS SLC-40
- **Bottom Right:** CCAFS LC-40

📌 **Color Indicators:**

- 🟥 **Red Icons:** Represent **failed launches**.
- 🟩 **Green Icons:** Indicate **successful launches**.

Figure 21: Folium Map Showing SpaceX Falcon 9 Launch Outcomes Across Various Launch Sites

Proximity of SpaceX Launch Sites to Key Landmarks

The **Folium map** visualization highlights the **distances of SpaceX Falcon 9 launch sites** from **coastlines, cities, railways, and highways**. The analysis reveals that while **launch sites are positioned near coastlines**, their proximity to other infrastructures varies.

📌 Key Observations:

- **Close to Coastlines:**
  - **CCAFS SLC-40** is **0.95 km** from the coast.
  - **VAFB SLC-4E** is **1.52 km** from the coast.
- **Distance from Major Cities:**
  - **VAFB SLC-4E** is **38.16 km** from Santa Maria.
  - **CCAFS SLC-40** is **56.04 km** from Melbourne.
- **Proximity to Railways & Highways:**
  - Some launch sites are **not as close** to railways and highways, affecting transportation logistics.

Figure 22: Folium Map Showing Launch Site Distances from Coastlines, Cities, Railways, and Highways.

Section 4

Build a Dashboard with Plotly Dash

Launch Success Distribution Across Sites

The **pie chart** in **Figure 24** illustrates the **success rates of launches across all SpaceX Falcon 9 launch sites**. The analysis reveals that **KSC LC-39A has the highest success rate** among all sites.

📌 **Key Observations:**

- **KSC LC-39A** accounts for **41.7% of total successful launches**, making it the **most successful launch site**.
- Other launch sites contribute to the remaining success rates, highlighting **variability in mission outcomes** across different locations.

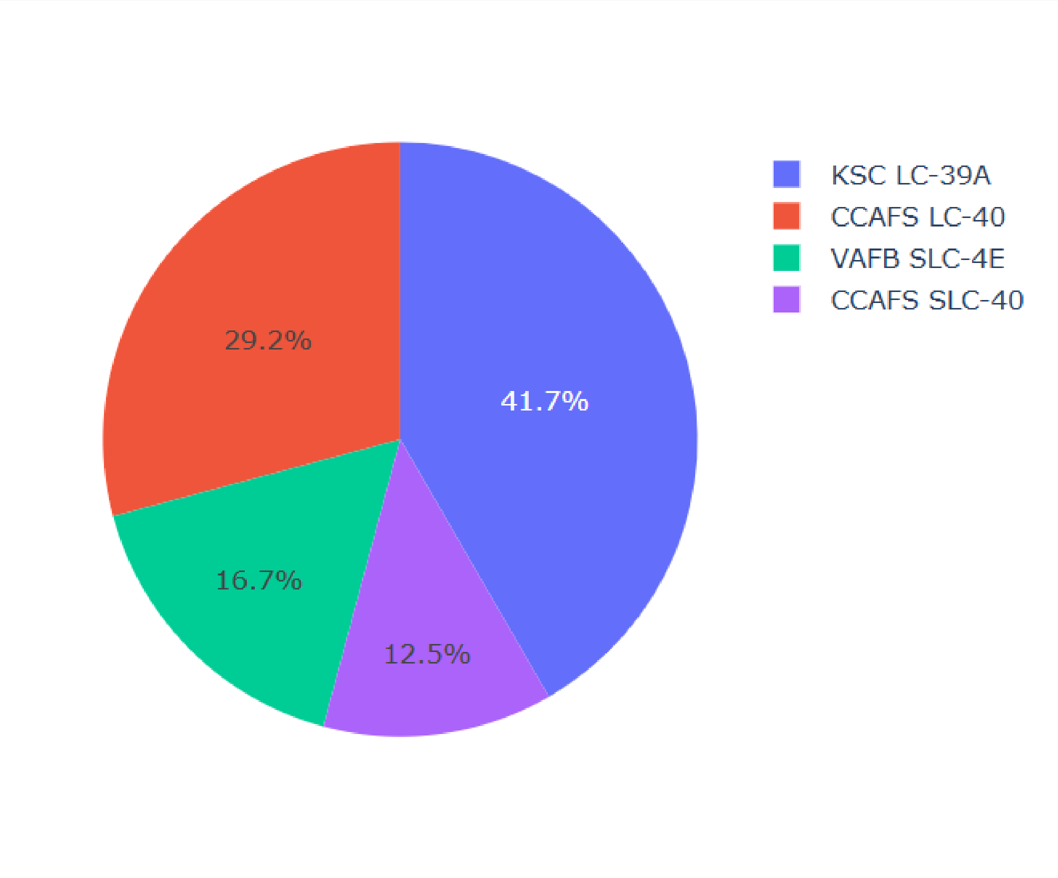

Figure 23: Pie Chart Showing Launch Success Rates Across All Sites.

Launch Site with the Highest Success Ratio

The **pie chart** in **Figure 25** highlights the **success ratios of different SpaceX Falcon 9 launch sites**. The analysis confirms that **KSC LC-39A has the highest success ratio**, outperforming other sites.

📌 **Key Observations:**

- **KSC LC-39A** leads with a **76.9% success rate**, making it the **most reliable launch site**.
- Other success ratios include:
  - **CCAFS LC-40: 73.1%**
  - **VAFB SLC-4E: 60%**
  - **CCAFS SLC-40: 57.1%**

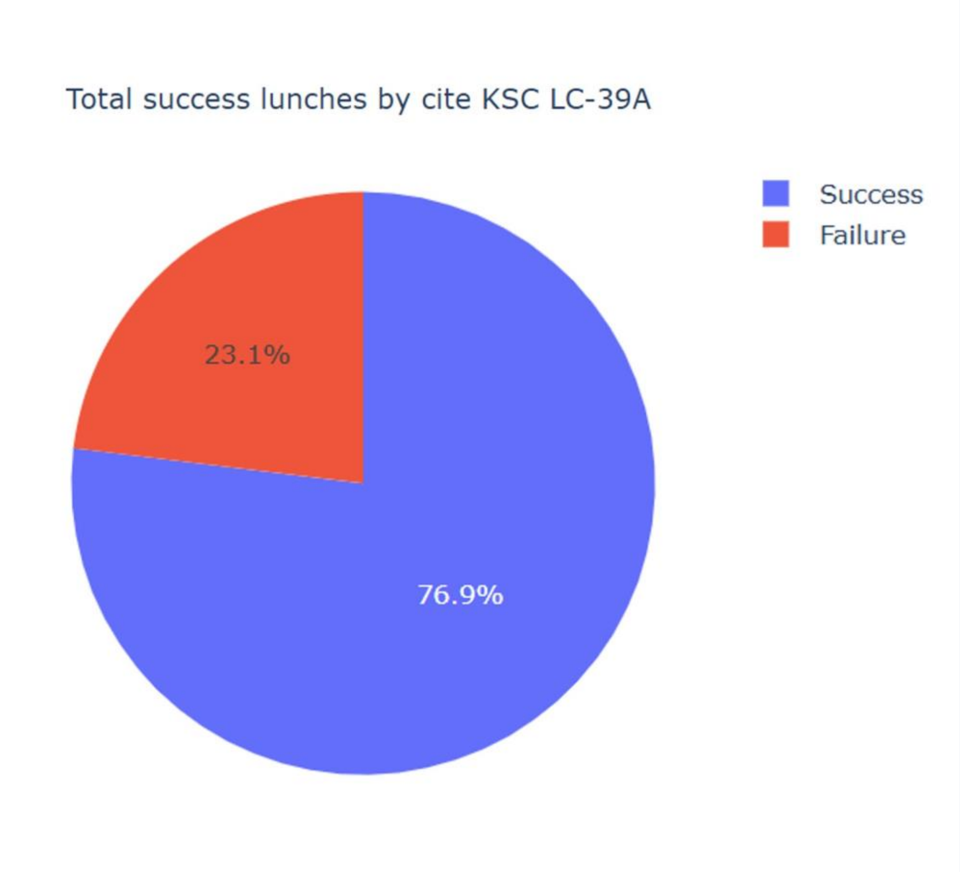

Figure 24: Pie Chart Showing the Launch Site with the Highest Success Ratio.

Payload vs. Launch Outcome Across All Sites

The **scatter plots** in **Figures 25 and 26** illustrate the relationship between **booster versions, payload mass, and launch success rates** across all launch sites.

📌 **Key Observations:**

- **Booster version FT** has the **highest success rate**, particularly for payloads between **700kg and 5,500kg**.
- **Rockets carrying payloads above 5,500kg** exhibit **lower success rates**, indicating that **heavier payloads reduce the chances of a successful launch outcome**.

Figure 25: Scatter Plot Showing Booster Versions with Different Payload Mass Across All Launch Sites.

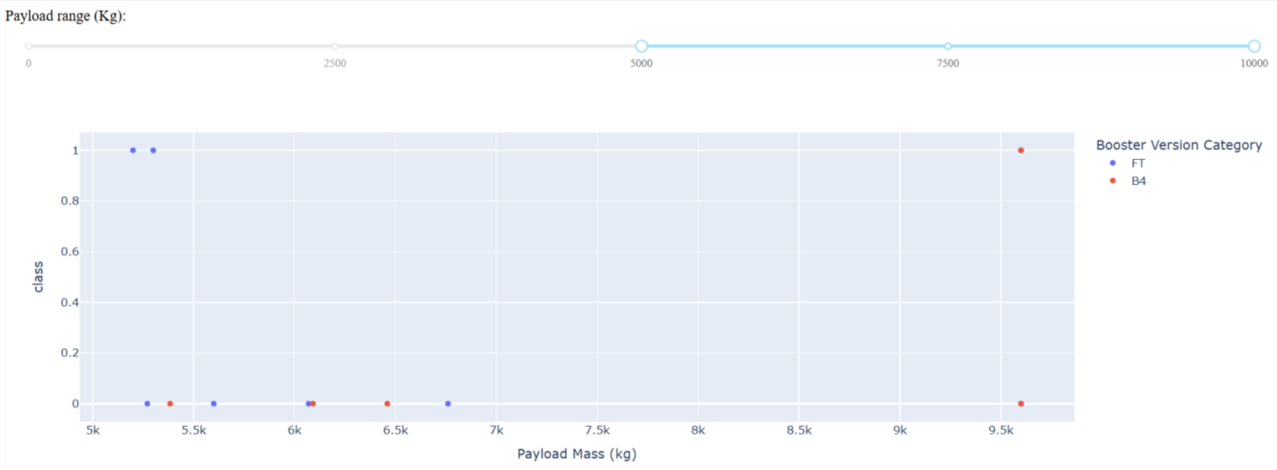

Figure 26: Scatter Plot Showing Booster Versions for Payloads Greater Than 5,500kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

The **bar chart in Figure 27** compares the accuracy scores of different classification models used in the predictive analysis. The results indicate that the **Decision Tree Classifier** outperformed other models in terms of accuracy.

📌 **Key Observations:**

- The **Decision Tree Classifier** achieved the **highest accuracy** of approximately **0.875 (87.5%)**, making it the most effective model for predicting launch success.
- Other classification models had **lower accuracy scores**, suggesting they may require further tuning or additional feature selection for improvement.

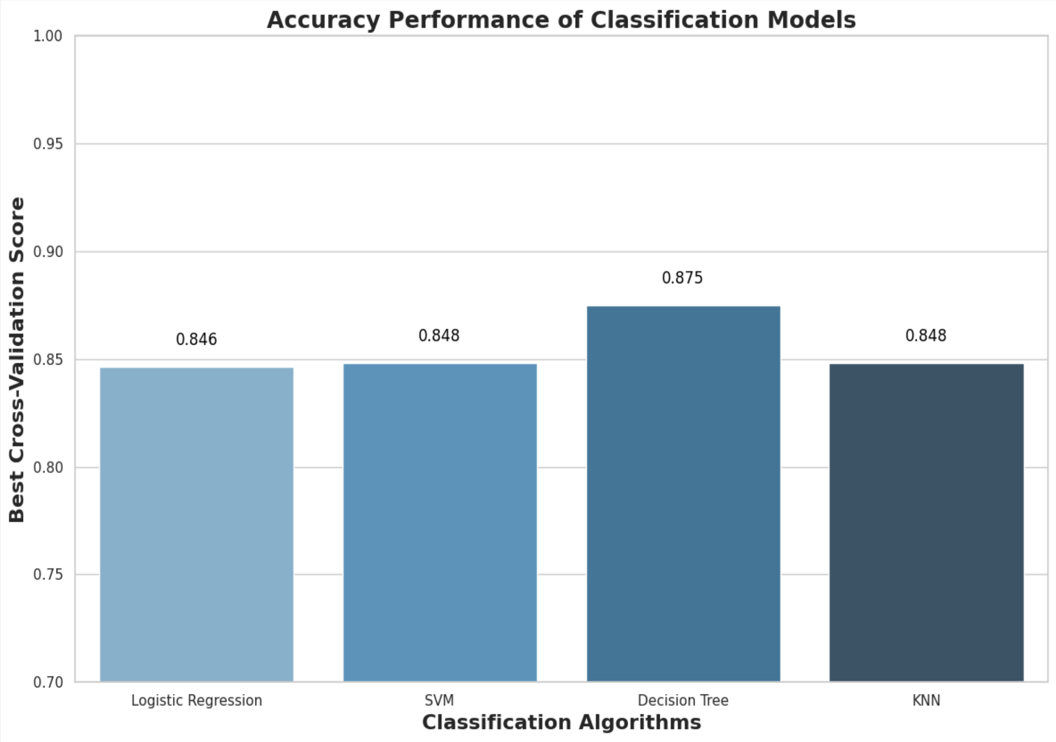

Figure 27: Bar Chart Showing Accuracy Scores of Different Classification Models

Confusion Matrix Analysis

The **confusion matrix in Figure 28** provides insights into the **performance of the Decision Tree Classifier** on the test set. After splitting the dataset into **training and test sets**, the test set contained only **18 samples**, allowing for an evaluation of the model's predictive accuracy.

📌 **Key Observations:**

- The model correctly predicted **12 successful landings** (**True Positives**).
- It also correctly identified **3 unsuccessful landings** (**True Negatives**).
- The classifier had **0 False Negatives**, meaning it did not misclassify any successful landings as failures.
- However, **3 False Positives** were recorded, where the model incorrectly predicted success for failed landings.

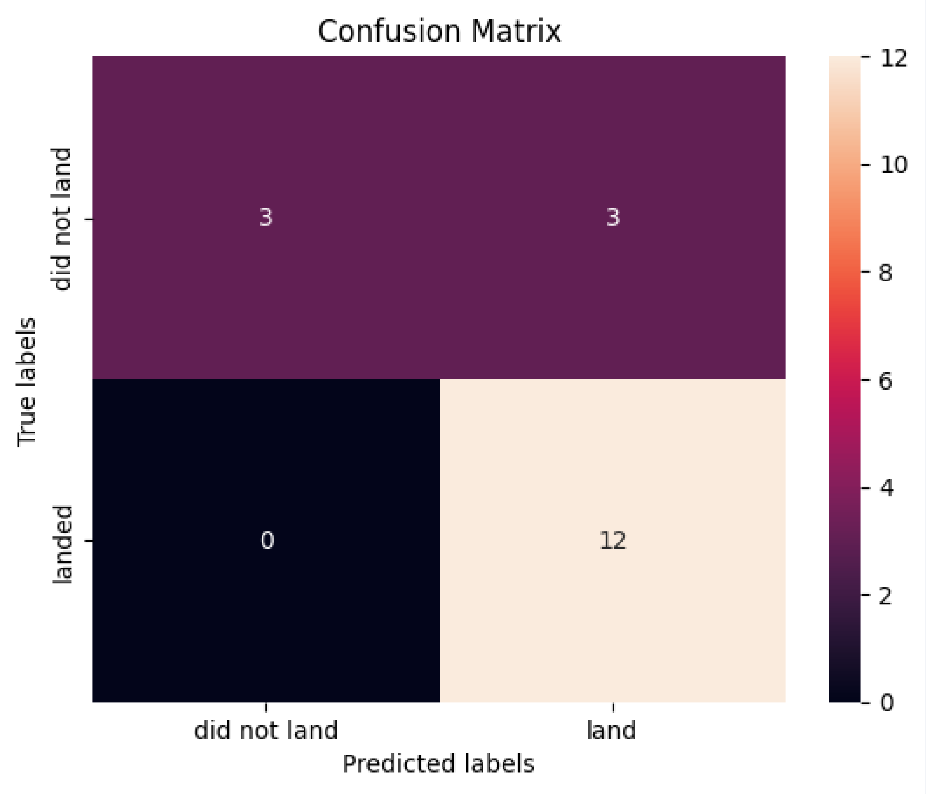

Figure 28: Confusion Matrix of the Decision Tree Classifier

Conclusions

The analysis of SpaceX Falcon 9 launch data provides valuable insights into the factors influencing mission success. Several key aspects should be considered to optimize future launches:

📌 **Key Takeaways:**

- **Payload Mass Matters:** Rockets with **smaller payloads** demonstrated a **higher success rate**, indicating that payload size plays a critical role in launch outcomes.
- **Orbit Selection is Crucial:** Certain orbits, including **VLEO, ES-L1, GEO, HEO, and SSO**, had **higher success rates** compared to others, emphasizing the importance of **orbit selection in mission planning**.
- **Strategic Launch Site Locations:** SpaceX launch sites are **positioned in coastal regions**, facilitating **efficient retrieval and recovery operations** while minimizing risks to populated areas.
- **Improved Success Over Time:** In **recent years, launch success rates have increased**, suggesting that **technological advancements and operational optimizations** have enhanced mission reliability.
- **Model Performance:** Among various classification models, the **Decision Tree Classifier** achieved the **best performance**, with an accuracy of **87%**, making it a **reliable model for predicting landing outcomes**.

These findings highlight the **continuous improvements in rocket technology, launch strategies, and predictive modeling**, contributing to the growing success of **SpaceX Falcon 9 missions**.

Thank you!