IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

M.reza Asgari
30 March 2025

Outline

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

Executive Summary

- Summary of methodologies

  We gathered data from Wikipedia and the SpaceX REST API. After data collection, we performed data cleaning and wrangling to prepare the dataset. We then conducted exploratory data analysis (EDA) using various visualization techniques. Additionally, we performed interactive visual analytics and concluded the analysis by predicting future outcomes using machine learning techniques.

- Summary of all results

  We identified key features and algorithms that accurately predict the likelihood of successfully landing the first stage of rockets. This capability can help save costs and enhance competitiveness with other aerospace companies.

Introduction

- Project background and context

In this capstone project, we analyzed Falcon 9 launch records obtained from the SpaceX API and Wikipedia. Our goal was to determine if the first stage of the rocket landed successfully, as this can significantly affect the cost of each launch. SpaceX's Falcon 9 rockets are advertised at $62 million per launch, whereas competitors' costs can exceed $165 million. Much of SpaceX's cost savings are due to the reusability of the first stage. Therefore, predicting the likelihood of a successful landing is crucial for estimating launch costs and enabling competitors to bid effectively against SpaceX.

Introduction

- Problems you want to find answers
  - **Predicting Landing Success:**

    Can we accurately predict if the Falcon 9 first stage will land successfully?
  - **Cost Estimation:**

    How does predicting landing success impact launch cost estimation?
  - **Feature Importance:**

    Which features (payload mass, orbit type, etc.) most influence landing success?
  - **Model Performance:**

    Which machine learning algorithms best predict landing success?
  - **Launch Site Impact:**

    How do different launch sites affect landing success rates?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  The data was obtained through SpaceX's REST API and supplemented with information from Wikipedia tables on Falcon 9 launches. The dataset comprehensively covers all launches from the inception of the Falcon 9 program up to the year 2020.

- Perform data wrangling

  We cleaned the data by removing duplicates from the JSON files received from the API and created the necessary columns from the raw data.

- Perform exploratory data analysis (EDA) using visualization and SQL

  Utilized visualization and SQL techniques to uncover initial insights and correlations in the data.

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash

  Employed Folium and Plotly Dash for dynamic and interactive visual analysis.

- Perform predictive analysis using classification models

  Developed and evaluated models using:

  1. Logistic Regression
  2. Support Vector Machine (SVM)
  3. Decision Tree
  4. K-Nearest Neighbors (KNN)

  Outcome: Logistic Regression outperformed the other models in predicting landing success.

Data Collection

- The data for this project was collected from two main sources:

1. **SpaceX API**: Provides comprehensive launch data.

   URL: SpaceX API - Past Launches

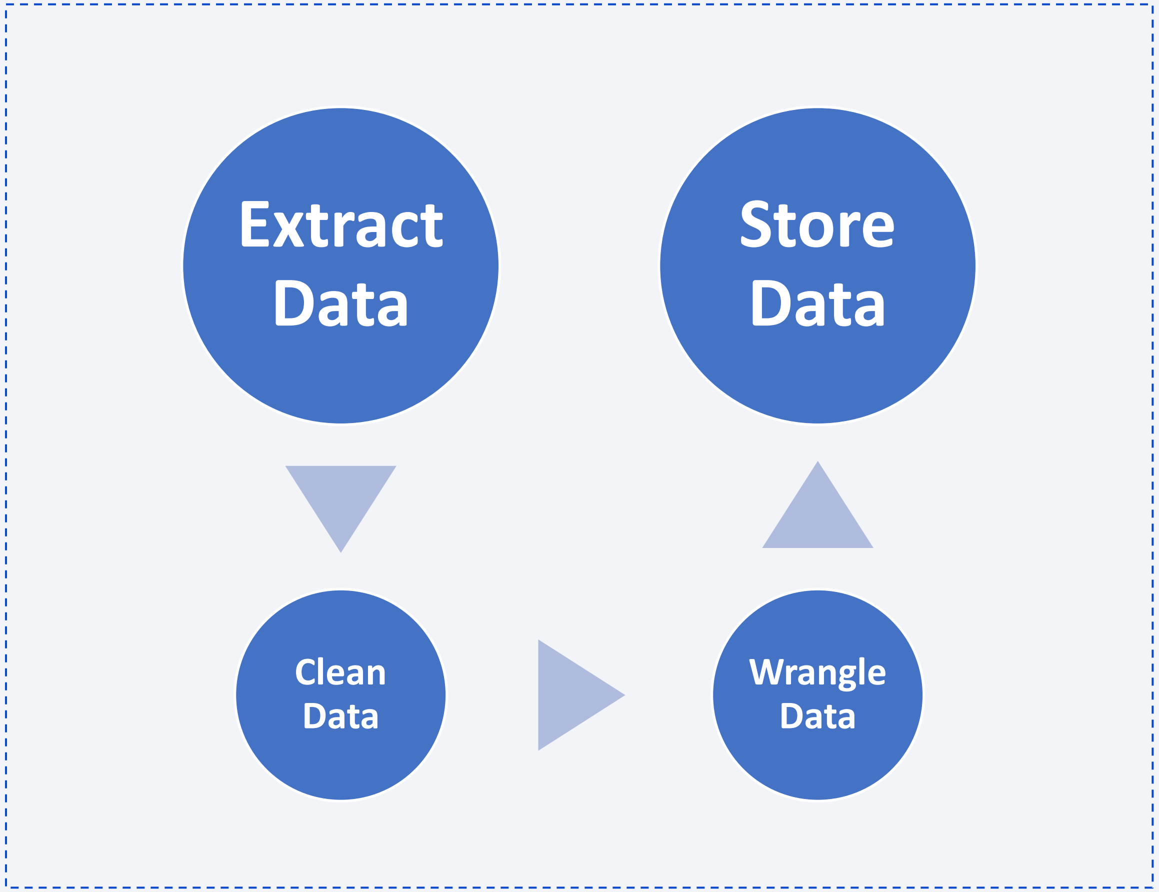
2. **Wikipedia**: Offers detailed information on Falcon 9 and Falcon Heavy launches.

   URL: Wikipedia - Falcon 9 Launches

Data Collection – SpaceX API

Process

1. **Data Extraction**:
    - **API Endpoint**: SpaceX API - Past Launches
    - **Method**: Used HTTP GET requests to retrieve JSON data.
2. **Data Cleaning**:
    - Parsed JSON data to extract relevant fields.
    - Removed any redundant or irrelevant information.
3. **Data Wrangling**:
    - Standardized column names and formats.
    - Created new columns for detailed analysis, such as landing outcome.
4. **Data Storage**:
    - Stored the cleaned and structured data in CSV format for further analysis.

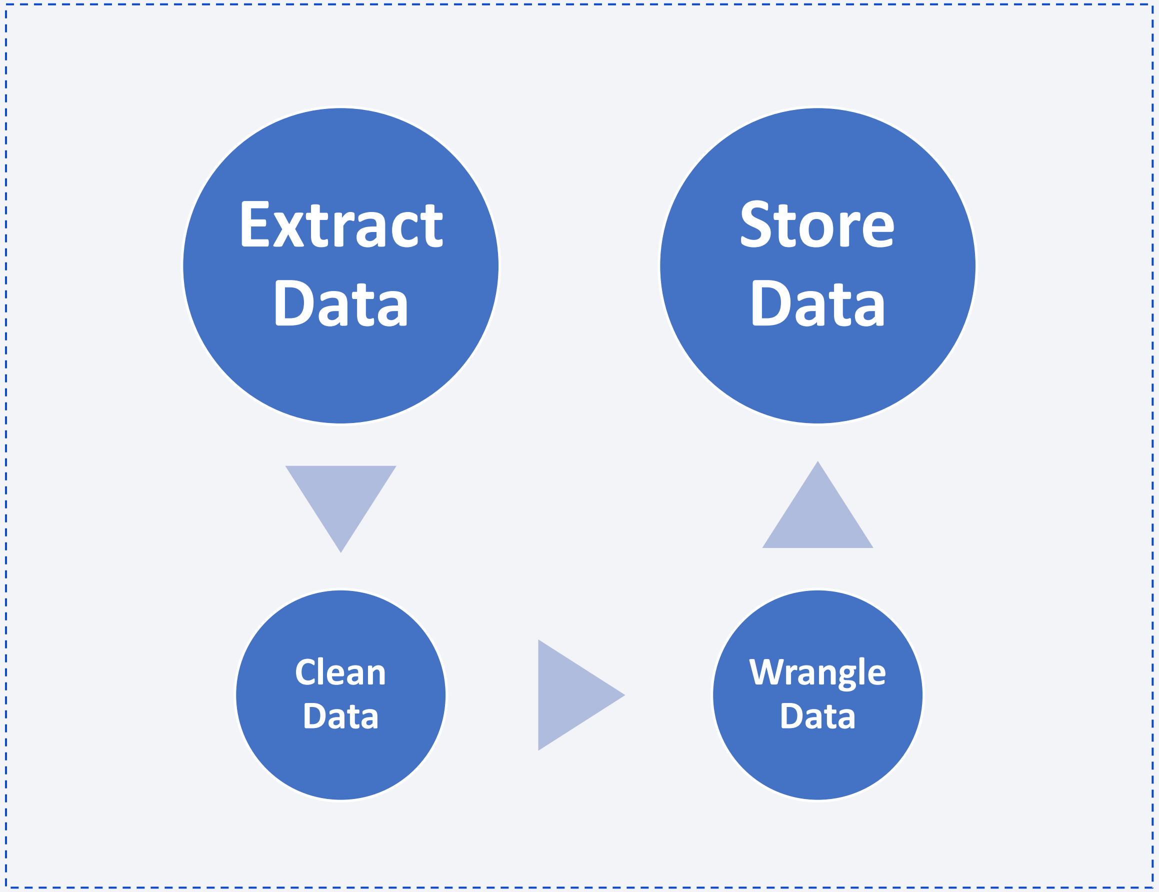


GitHub URL : **Jupyter notebook**

Data Collection - Scraping

Process

1. **Data Extraction**:
   - **Source URL**: Wikipedia - Falcon 9 Launches
   - **Method**: Used web scraping techniques to extract tabular data from the webpage.
2. **Data Cleaning**:
   - Converted the HTML tables to Pandas DataFrames.
   - Removed any redundant or irrelevant information.
3. **Data Wrangling**:
   - Standardized column names and formats.
   - Created new columns for detailed analysis, such as landing outcome.
4. **Data Storage**:
   - Stored the cleaned and structured data in CSV format for further analysis.



GitHub URL : **Jupyter notebook**

Data Wrangling

- **Overview**

  Data wrangling involves cleaning and transforming raw data into a structured format suitable for analysis. This process includes handling missing values, removing duplicates, standardizing formats, and creating new features.

  1. **Initial EDA** : Performed initial EDA to understand the dataset better.
  2. **Summarize Launches** : Calculated the number of launches per site, occurrences of each orbit type, and mission outcomes per orbit type.
  3. **Create Landing Outcome Label** : Derived the 'Landing Outcome' label from the 'Outcome' column for further analysis.

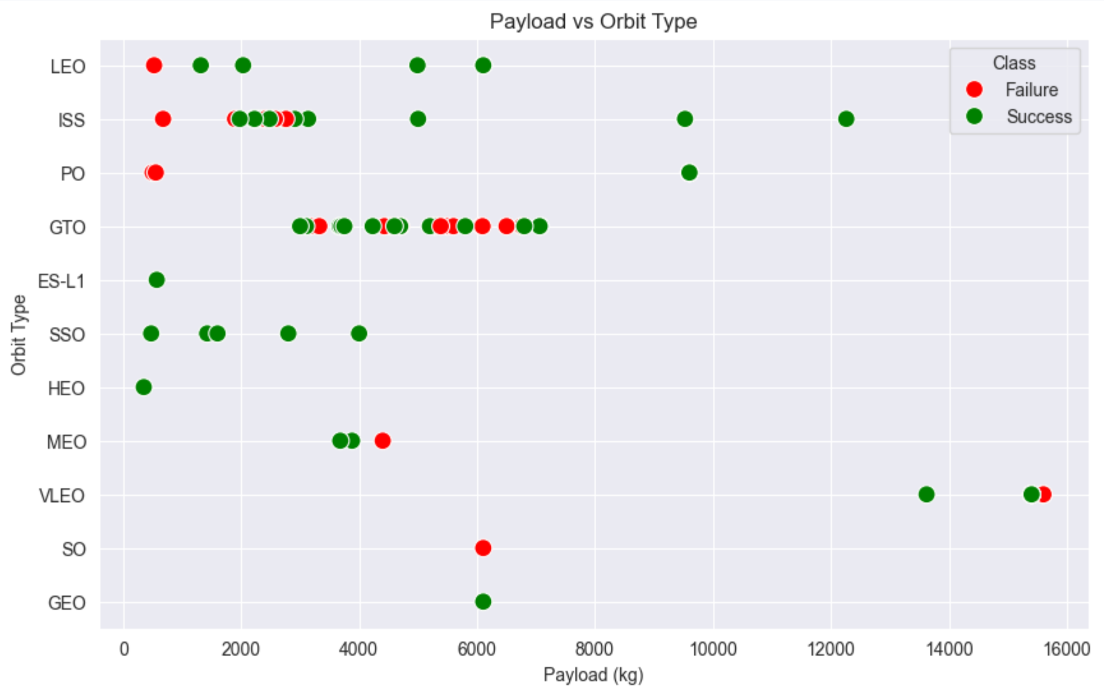
GitHub URL : [Jupyter notebook]

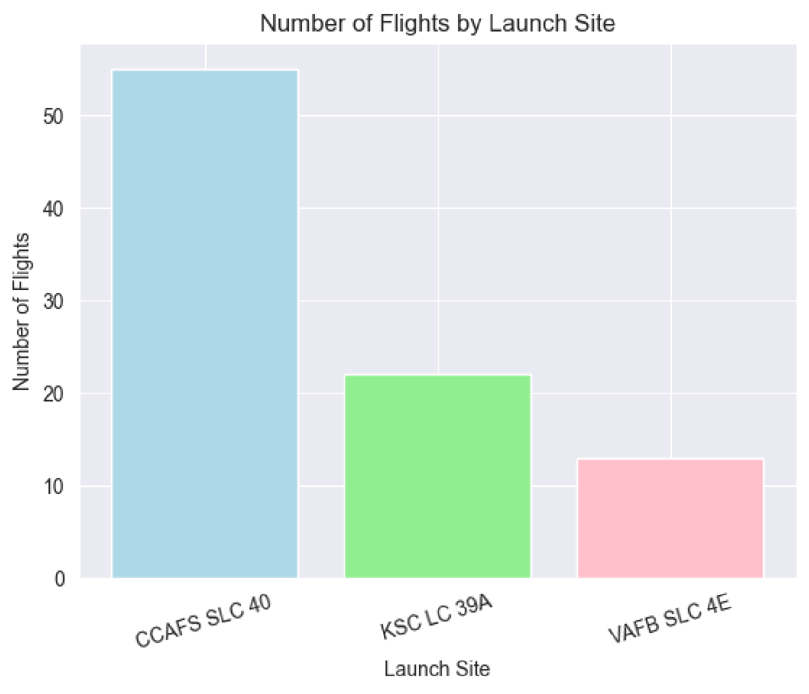
EDA with Data Visualization

The use of scatterplots and barplots provided valuable insights into the SpaceX launch data. These visualizations helped in identifying patterns, trends, and anomalies, which are crucial for predictive analysis and decision-making.

- **Scatterplots** revealed trends and patterns in payload mass, flight numbers, and orbits.
- **Barplots** highlighted the most frequently used launch sites and the reliability of different orbits.

These visualizations provided crucial insights for predictive analysis and decision-making.





GitHub URL : [Jupyter notebook]

EDA with SQL

SQL Query Summary

- Retrieved the names of all unique launch sites involved in the space missions.
- Extracted the first five records where the launch sites begin with 'CCA'.
- Calculated the total payload mass carried by boosters launched by NASA (CRS).
- Computed the average payload mass carried by the booster version F9 v1.1.
- Identified the date of the first successful landing outcome on a ground pad.
- Listed the names of boosters that successfully landed on a drone ship with a payload mass between 4000 and 6000 kg.
- Summarized the total number of successful and failed mission outcomes.
- Listed the names of booster versions that carried the maximum payload mass using a subquery.
- Displayed records of failure landing outcomes on drone ships, including booster versions and launch sites, for each month in 2015.
- Ranked the counts of different landing outcomes (e.g., Failure on drone ship, Success on ground pad) between June 2010 and March 2017 in descending order.

GitHub URL : **Jupyter notebook**

Build an Interactive Map with Folium

These objects were added to enhance the visualization of geographical data, making it easier to understand the spatial distribution and relationships between various launch-related events and locations.

1. **Markers :**

   Purpose: Indicate specific points of interest such as launch sites.

   Usage: Marked locations like NASA Johnson Space Center and various SpaceX launch sites.

2. **Circles :**

   Purpose: Highlight areas around specific coordinates.

   Usage: Drew attention to areas such as NASA Johnson Space Center with a radius circle.

3. **Marker Clusters :**

   Purpose: Group events occurring at the same or nearby coordinates.

   Usage: Aggregated launch events at each launch site for better visualization of clustered data.

4. **Lines :**

   Purpose: Show distances and connections between two coordinates.

   Usage: Indicated the distances between launch sites and their proximities to nearby locations.

GitHub URL : [Jupyter notebook]

Build a Dashboard with Plotly Dash

Plots/Graphs Added:

- **Percentage of Launches by Site :**
  - **Purpose:** To visualize the distribution of launches across different launch sites.
  - **Usage:** A pie chart was used to show the percentage of successful launches at each site, helping to identify the most active and successful launch sites.
- **Payload Range :**
  - **Purpose:** To analyze the relationship between payload mass and launch success rates.
  - **Usage:** A scatter plot was used to show the correlation between payload mass and launch outcomes, both overall and for individual launch sites, highlighting trends and patterns in payload capacity and success rates.

Interactions:

- **Dropdown for Launch Site Selection :**
  - **Purpose:** To filter the data and visualize the success rates for specific launch sites.
  - **Usage:** Allowed users to select a launch site from a dropdown menu, dynamically updating the plots to show data relevant to the selected site.
- **Range Slider for Payload Selection :**
  - **Purpose:** To filter the data based on payload mass range.
  - **Usage:** Enabled users to adjust the payload mass range using a slider, updating the scatter plot to reflect the selected range and providing insights into how different payload masses affect launch success rates.

These plots and interactions were added to the dashboard to provide a comprehensive and interactive visualization of the SpaceX launch data, facilitating detailed analysis and exploration of the relationships between launch sites, payloads, and success rates.

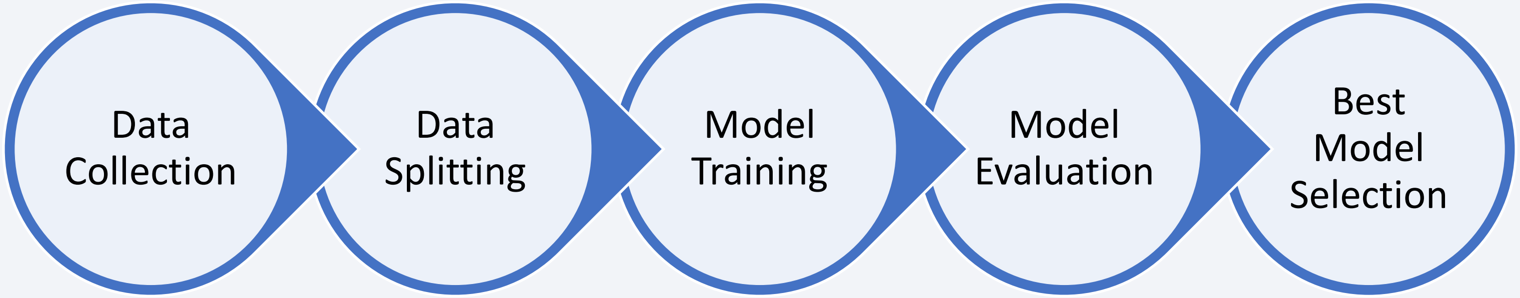
GitHub URL : [Jupyter notebook]

Predictive Analysis (Classification)

Compared Models:

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. K-Nearest Neighbors (KNN)

Best Performing Model : Logistic Regression was determined to be the best performing model for predicting the success of SpaceX launches based on the given dataset.



GitHub URL : [Jupyter notebook]

Results

Exploratory Data Analysis Results :

- **Launch Sites :**

  SpaceX uses 4 different launch sites.

- **Early Launches :**

  Initial launches were for SpaceX itself and NASA.

- **Payload Insights :**

  The average payload mass for the F9 v1.1 booster is 2,928 kg.

- **First Successful Landing :**

  The first successful landing outcome occurred in 2015, five years after the first launch.

- **Booster Versions :**

  Many Falcon 9 booster versions were successful at landing on drone ships with payloads above the average.

- **Mission Outcomes :**

  Almost 100% of mission outcomes were successful.

- **Failures in 2015 :**

  Two booster versions failed at landing on drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.

- **Improvement Over Time :**

  The number of successful landing outcomes improved significantly over the years.

Results

Interactive Analytics Demo in Screenshots :

- **Identification of Launch Sites :**

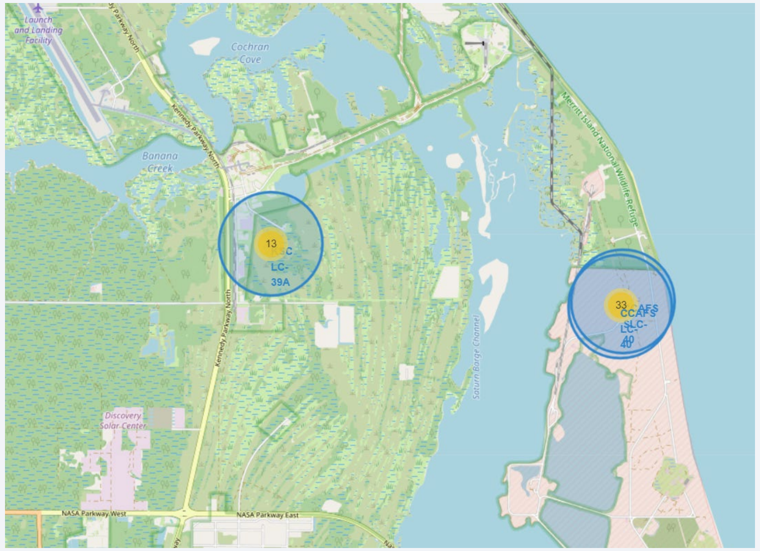
  Launch sites are typically located in safe areas, often near the sea, and are supported by good logistic infrastructure.

- **East Coast Launch Sites :**

  Most launches occur at East Coast launch sites.

- **Insights from Interactive Analytics :**

  Utilizing interactive analytics enabled the identification of patterns and geographical advantages of launch site locations.



Results

Predictive Analysis Results :

- **Logistic Regression Test Data Accuracy :** 0.833
- **Support Vector Machine Test Data Accuracy :** 0.833
- **Decision Tree Test Data Accuracy :** 0.778
- **K Nearest Neighbors Test Data Accuracy :** 0.833

Best Performing Method : Logistic Regression with an accuracy of 0.833

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Explanation :

- **Launch Site Performance :**

  CCAFS SLC 40 : This is the best-performing launch site currently, with most of the recent launches being successful.

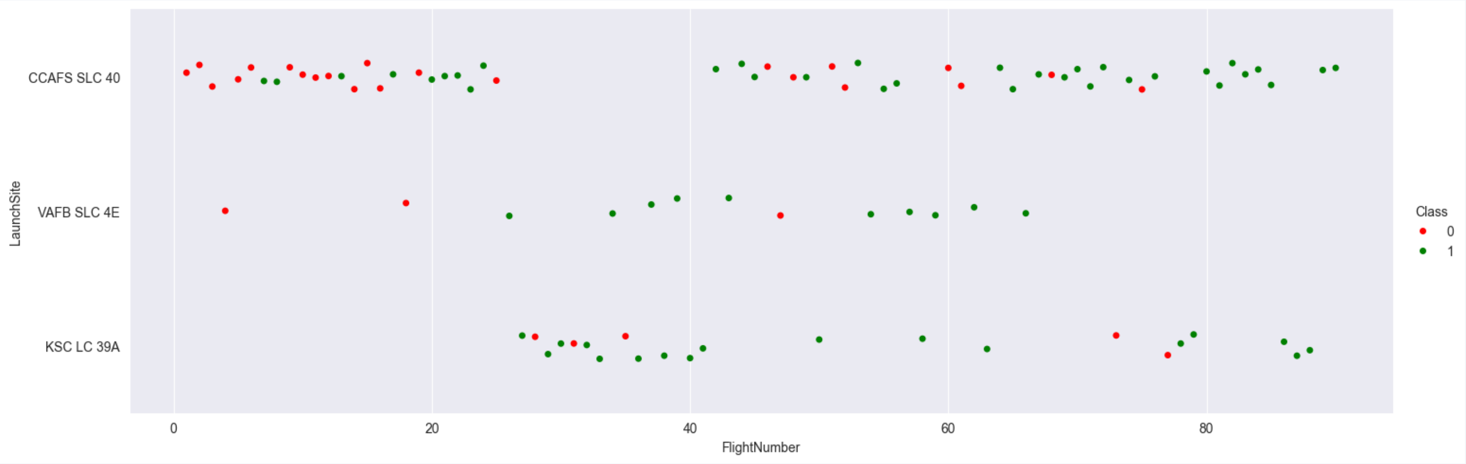
  VAFB SLC 4E : This site is the second-best in terms of launch success rate.

  KSC LC 39A : This site is in the third place for successful launches.

- **General Trend**:

  The overall success rate of launches has improved over time, as indicated by the increasing number of successful launches in the plot.

This scatter plot provides a clear visual representation of the relationship between flight numbers and launch sites, demonstrating the performance and reliability of different launch locations over time.



Payload vs. Launch Site

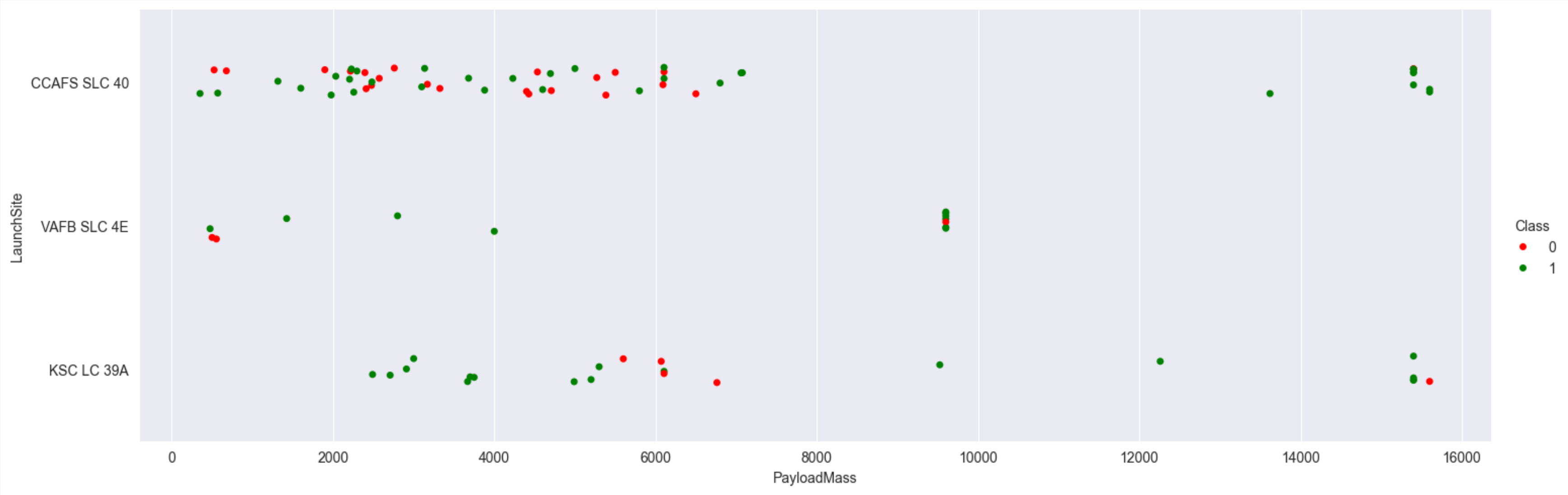
Explanation :

- **Payload Success Rate :**

  Payloads over 9,000 kg : These payloads, which are roughly equivalent to the weight of a school bus, exhibit an excellent success rate.

  Payloads over 12,000 kg : Such heavy payloads appear to be launched successfully only from **CCAFS SLC 40** and **KSC LC 39A** launch sites.

This scatter plot effectively demonstrates the relationship between payload mass and launch site, highlighting the capacity and reliability of different sites for handling heavy payloads.

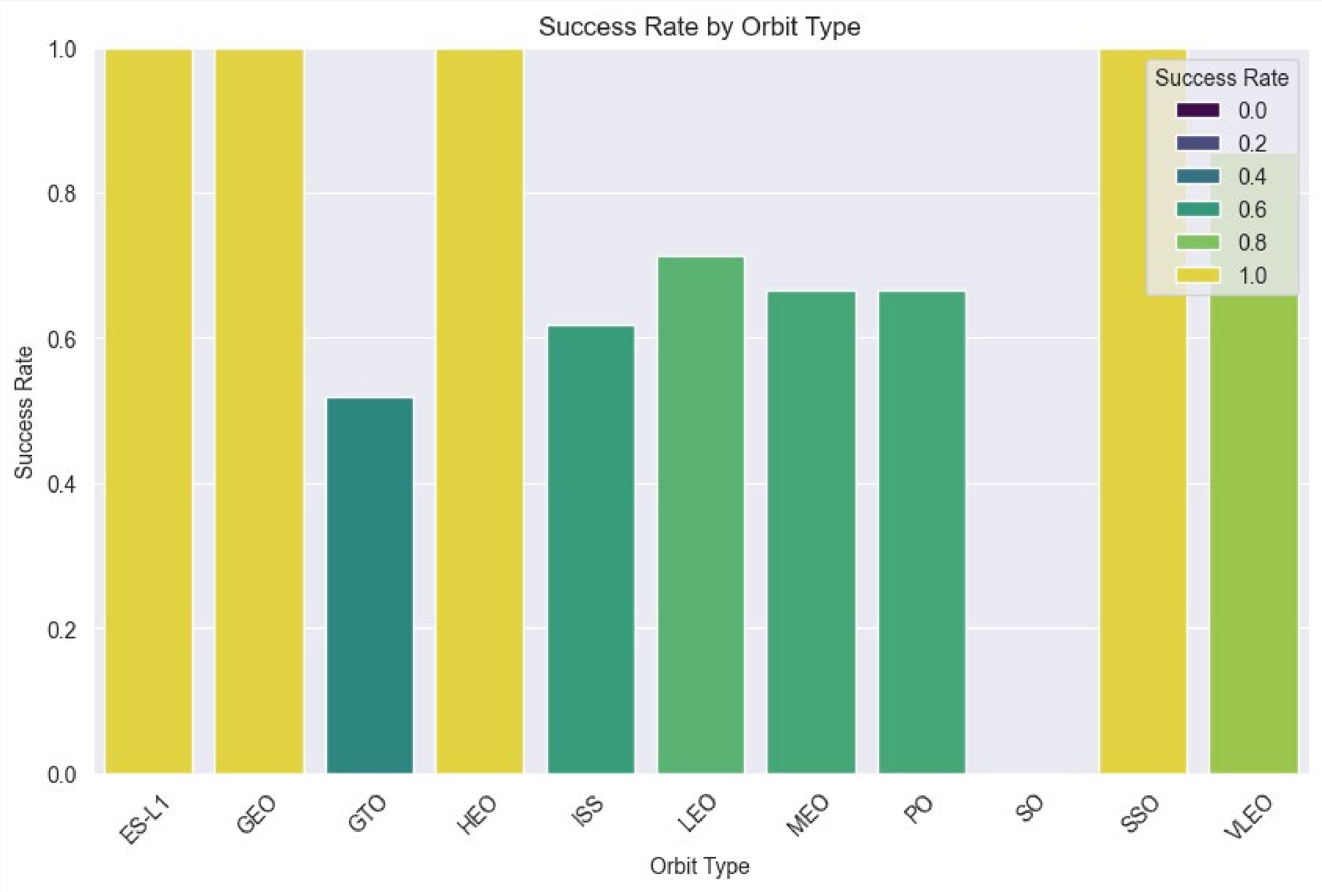


Success Rate vs. Orbit Type

Explanation :

- **Highest Success Rates :**
    - **ES L1**
    - **GEO**
    - **HEO**
    - **SSO**
- **High Success Rates :**
    - **VLEO** : Above 80%
    - **LFO** : Above 70%

This bar chart highlights the success rates of various orbit types, showcasing the reliability of specific orbits for successful launches. It provides a clear comparison of the performance across different orbits, aiding in identifying the most reliable ones for future missions.



Flight Number vs. Orbit Type

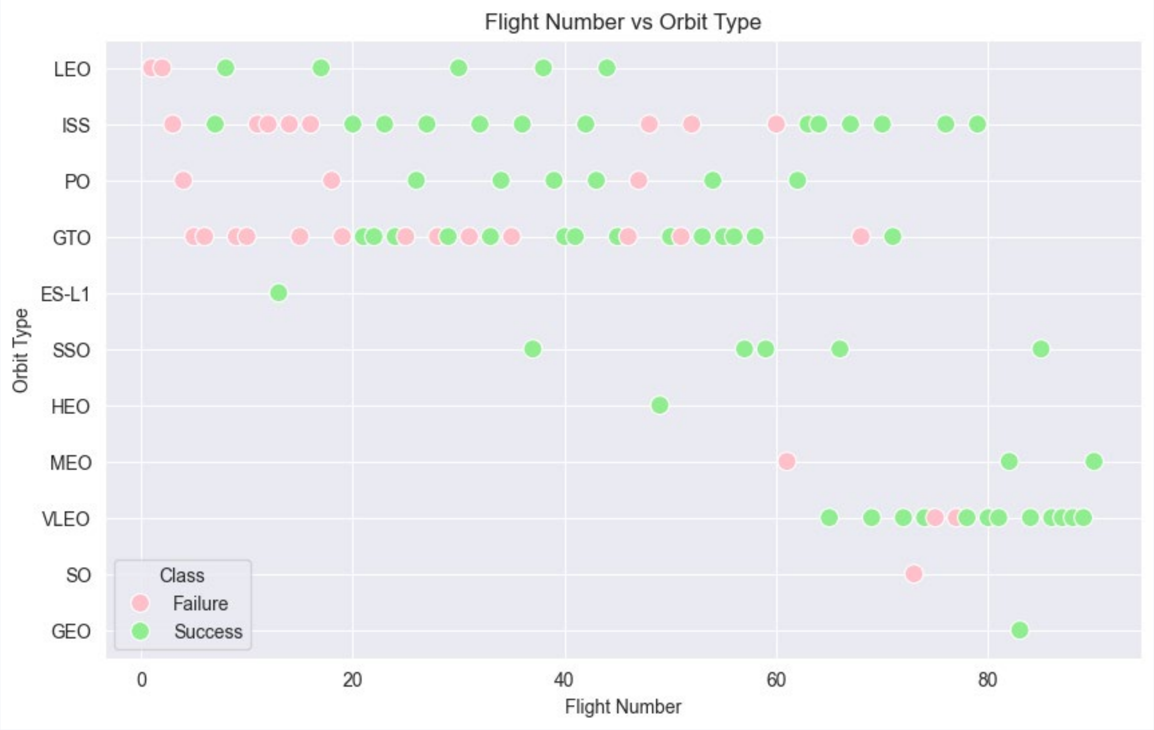
Explanation :

- **Success Rate Improvement :**

  The success rate has improved over time across all orbits, indicating advancements in technology and processes.

- **VLEO Orbit :**

  The VLEO (Very Low Earth Orbit) shows a notable increase in frequency in recent times, suggesting it as a new business opportunity. The trend indicates growing interest and potential for more launches in this orbit.



This scatter plot helps visualize the relationship between flight numbers and orbit types, highlighting the advancements in launch success and emerging opportunities in the space industry.

Payload vs. Orbit Type

Explanation :

- **GTO Orbit :**

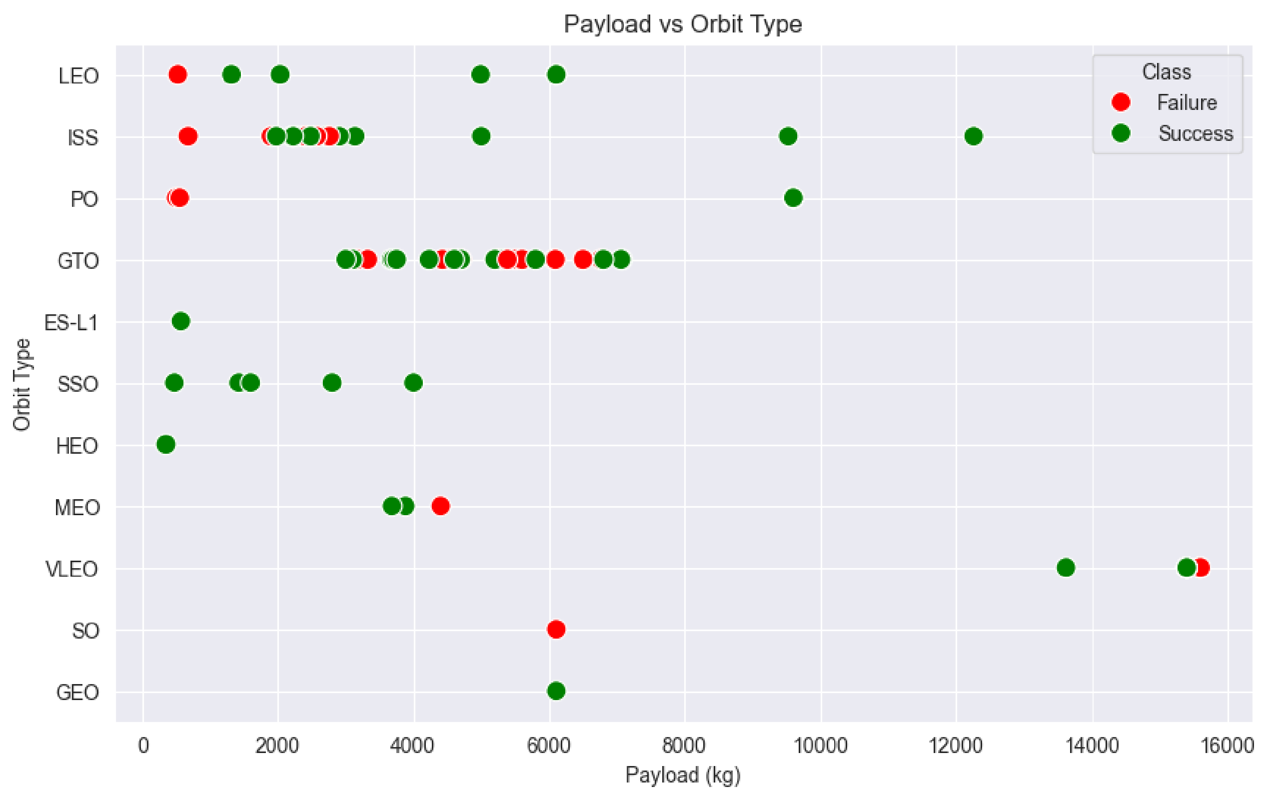
  There appears to be no significant relationship between payload and success rate for the GTO (Geostationary Transfer Orbit), indicating that payload mass does not heavily influence success in this orbit.

- **ISS Orbit :**

  The ISS (International Space Station) orbit supports a wide range of payloads and exhibits a high success rate, showing versatility and reliability in missions targeting this orbit.

- **SO and GEO Orbits :**

  There are relatively few launches to SO (Sun-synchronous Orbit) and GEO (Geostationary Orbit), possibly due to their specialized nature and specific mission requirements.



This scatter plot aids in understanding the distribution and success rates of payloads across different orbit types, providing insights into mission planning and payload management.

Launch Success Yearly Trend

Explanation :

- **Trend Over Time :**

  The success rate for SpaceX launches started to increase significantly in 2014 and continued to improve through 2020. This trend highlights the company's learning curve and technological advancements over time.

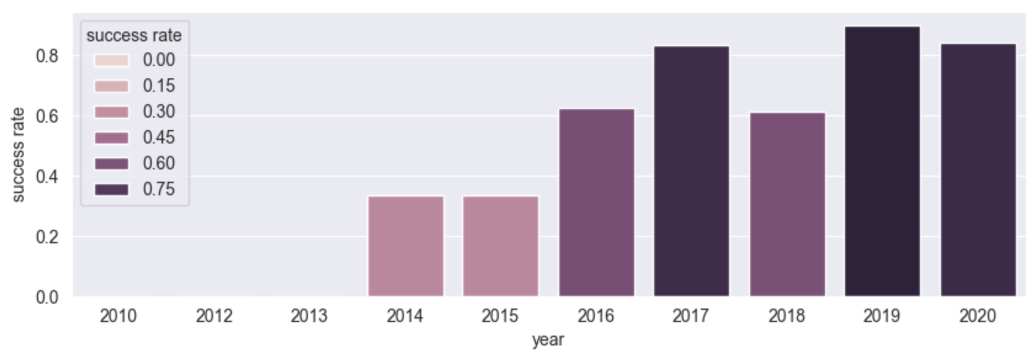
- **Early Years (2010-2013) :**

  The first three years of launches (2010-2013) had lower success rates, indicating a period of adjustments, experimentation, and refinement of launch technologies and processes.

- **Consistent Improvement :**

  From 2014 onward, the success rate shows a consistent upward trend, reflecting improvements in engineering, operational procedures, and experience gained from previous launches.

This line chart provides a clear visual representation of the overall improvement in launch success rates over the years, underscoring SpaceX's progress and increasing reliability in its space missions.



All Launch Site Names

Explanation :

- The query selects distinct values from the launch_site column in the spacex_launch_data table.
- This retrieves the unique names of the launch sites used by SpaceX.

Result :

According to the data, there are four unique launch sites:

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

These launch sites are pivotal for SpaceX's operations, each providing specific logistical and technical advantages for different types of missions.

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE where "Launch_Site" like "CCA%" limit 5
```

Result :

- The result shows five records from the dataset where the launch_site begins with 'CCA'. This includes various launch dates, booster versions, payloads, and mission outcomes.
- The provided screenshot displays the first five records with launch sites starting with 'CCA', including:
  1. **CCAFS LC-40**: Several launches from 2010 to 2013 with various payloads and customers like SpaceX and NASA.
  2. **Mission Outcomes**: All listed missions were successful.
  3. **Landing Outcomes**: Includes instances of 'Failure (parachute)' and 'No attempt'.
- This data highlights the early launch history at the CCAFS LC-40 site, showcasing different mission profiles and outcomes.

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

Total Payload Mass

```
%%sql
select sum("PAYLOAD_MASS__KG_") as "total_payload" from SPACEXTABLE where "Customer" = 'NASA (CRS)'
```

Result :

- The total payload carried by boosters for NASA (CRS) is **45,596 kg**.

This result is derived by summing all the payload masses from records where the customer column contains 'NASA (CRS)'. It confirms the substantial payloads delivered by SpaceX for NASA missions.

Average Payload Mass by F9 v1.1

```
%%sql
select avg("PAYLOAD_MASS__KG_") as "average_payload_mass" from SPACEXTABLE where "Booster_Version" = 'F9 v1.1'
```

Result :

- The average payload mass carried by the booster version **F9 v1.1** is **2,928.4 kg**.

This result is obtained by filtering the data for the booster version 'F9 v1.1' and calculating the average of the payload masses, indicating the typical payload capacity of this specific booster version.

First Successful Ground Landing Date

```
%sql select min(Date) as "first_successful_landing_date" from SPACEXTABLE where Landing_Outcome = "Success (ground pad)"
```

Result :

- The first successful landing outcome on a ground pad occurred on **2015-12-22**.

This result was obtained by filtering the data to only include successful ground pad landings and then finding the earliest date in the filtered set.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct Booster_Version from SPACEXTABLE where "landing_outcome" = "Success (drone ship)" and PAYLOAD_MASS__KG_>4000 and PAYLOAD_MASS__KG_<6000
```

Result :

Boosters which have successfully landed on a drone ship and had payload mass greater than 4000 kg but less than 6000 kg :

1. F9 FT B1021.2
2. F9 FT B1031.2
3. F9 FT B1022
4. F9 FT B1026

These results were obtained by selecting distinct booster versions that meet the specified criteria for landing outcome and payload mass.

Total Number of Successful and Failure Mission Outcomes

```
%%sql
select Mission_Outcome, count(*) as occurrences from SPACEXTABLE group by Mission_Outcome
```

Result :

Mission Outcome Summary :

- **Success :** 99
- **Success (payload status unclear) :** 1
- **Failure (in flight) :** 1

By grouping the mission outcomes and counting the records for each group, we obtained the above summary. This analysis helps to understand the distribution of different mission outcomes in the dataset.

Boosters Carried Maximum Payload

```
%%sql
select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = ( select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

Result :

Boosters which have carried the maximum payload mass :

These are the boosters that have carried the maximum payload mass registered in the dataset.

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

Result :

- **Failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015 :**

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

These are the only two occurrences of failed landing outcomes on the drone ship for the year 2015.

```
%%sql
select
  case
    when substr("Date", 6, 2) = '01' then 'January'
    when substr("Date", 6, 2) = '02' then 'February'
    when substr("Date", 6, 2) = '03' then 'March'
    when substr("Date", 6, 2) = '04' then 'April'
    when substr("Date", 6, 2) = '05' then 'May'
    when substr("Date", 6, 2) = '06' then 'June'
    when substr("Date", 6, 2) = '07' then 'July'
    when substr("Date", 6, 2) = '08' then 'August'
    when substr("Date", 6, 2) = '09' then 'September'
    when substr("Date", 6, 2) = '10' then 'October'
    when substr("Date", 6, 2) = '11' then 'November'
    when substr("Date", 6, 2) = '12' then 'December'
    else 'Unknown'
  end as month_name,
  "Landing_Outcome",
  "Booster_Version",
  "Launch_Site"
from SPACEXTABLE
where substr("Date", 0, 5) = '2015'
and "Landing_Outcome" = 'Failure (drone ship)'
```

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

```
%%sql
select "Landing_Outcome", COUNT(*) as "Occurrences" from SPACEXTABLE where "Date" between '2010-06-04' and '2017-03-20' and
("Landing_Outcome" like '%failure%' or "Landing_Outcome" like '%Success%') group by "Landing_Outcome" order by "outcome_count" desc
```

Result :

- **Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20 :**

Landing_Outcome	Occurrences
Success (ground pad)	3
Success (drone ship)	5
Failure (parachute)	2
Failure (drone ship)	5

This view of the data highlights the frequency of different landing outcomes during the specified period. Notably, both successful and failed attempts on the drone ship are equal, and the "No attempt" category should be taken into account for a comprehensive analysis.

Section 3

Launch Sites Proximities Analysis

Folium Map Screenshot: Launch Sites

Important Elements and Findings:

- **Markers:** Indicate the precise locations of the launch sites. Each marker is labeled with the site name for easy identification.
- **Geographical Spread:** The launch sites are strategically located along the east and west coasts of the United States, providing optimal locations for different types of orbits and missions.
- **Proximity to the Sea:** Most launch sites are located near the sea, which is a strategic choice to ensure safety and allow for the recovery of reusable rocket components.
- **Logistical Infrastructure:** The locations are chosen for their excellent logistical infrastructure, facilitating transportation, assembly, and support for space missions.

This visual representation helps to understand the distribution and strategic placement of SpaceX's launch facilities, which are crucial for the company's operational efficiency and mission success.

Screenshot Explanation: The map displays the locations of all SpaceX launch sites in the United States. The markers represent the following launch sites:

- **VAFB SLC-4E (Vandenberg Air Force Base Space Launch Complex 4E)** in California.
- **KSC LC-39A (Kennedy Space Center Launch Complex 39A)** in Florida.
- **CCAFS LC-40 (Cape Canaveral Air Force Station Launch Complex 40)** in Florida.

Folium Map Screenshot: Launch Outcomes by Color

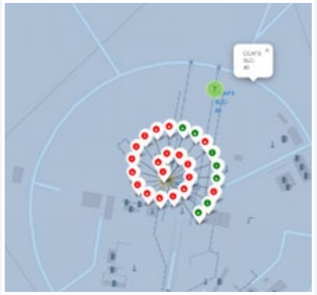
Screenshot Explanation: The map visualizes SpaceX launch sites in Florida and labels the launch outcomes with different colors. It includes:

- **Top Image:** Shows the clustering of launch events around the CCAFS LC-40 and KSC LC-39A sites in Florida.
- **Bottom Left Image:** Highlights individual launch sites with detailed markers showing specific launch locations.
- **Bottom Right Image:** Depicts a color-coded marker cluster showing various launch outcomes.

Important Elements and Findings:

- **Color-Coded Markers:**
  - **Green:** Represents successful landings.
  - **Red:** Represents failed landings.
  - **Blue:** Indicates other specific conditions like "No attempt" or "Success with conditions."
- **Marker Clusters:** Groups multiple events in close proximity to avoid marker overlapping and improve map readability.
- **Proximity to the Sea:** Similar to the first map, the launch sites are near the coastline, emphasizing safety and recovery logistics.

This representation helps in quickly identifying the success rate of launches from different sites and visually assessing the frequency and outcomes of these launches. The clustering and color coding provide an intuitive understanding of the data distribution and highlight areas of interest or concern.



Vandenberg Space Force Base Launch Site and Its Proximity to Key Infrastructure

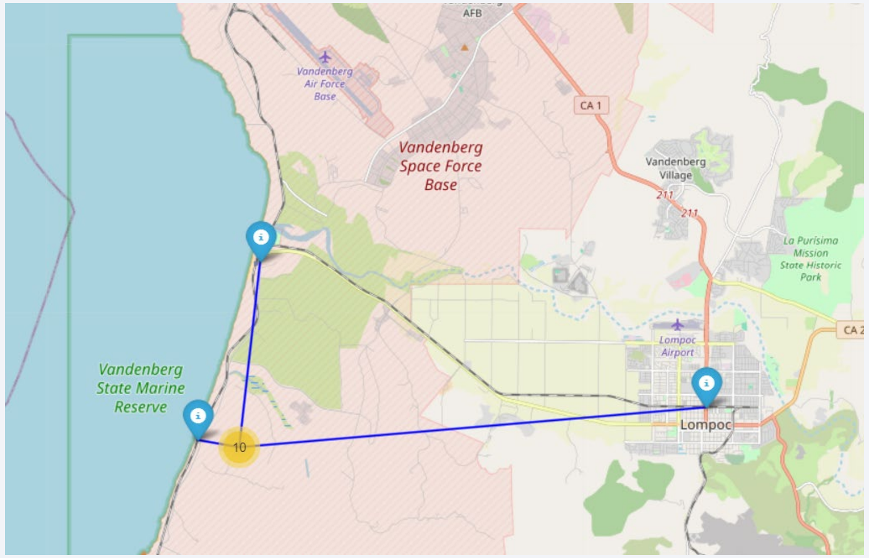
Explanation of the Screenshot:

- The map shows the **Vandenberg Space Force Base** launch site and its proximity to important nearby infrastructure, such as highways, coastlines, and railways.

Important Elements and Findings:

- The proximity to the coastline is crucial for safe launches, as it provides a clear path for rockets to safely exit over the ocean.
- The presence of major highways and the nearby town of Lompoc ensures the launch site is well-supported logistically.
- The calculated distances give a clear idea of how accessible key locations are to the base, helping in planning and emergency response scenarios.

This analysis showcases the strategic importance of **Vandenberg Space Force Base** for conducting safe and efficient launches.



Section 4

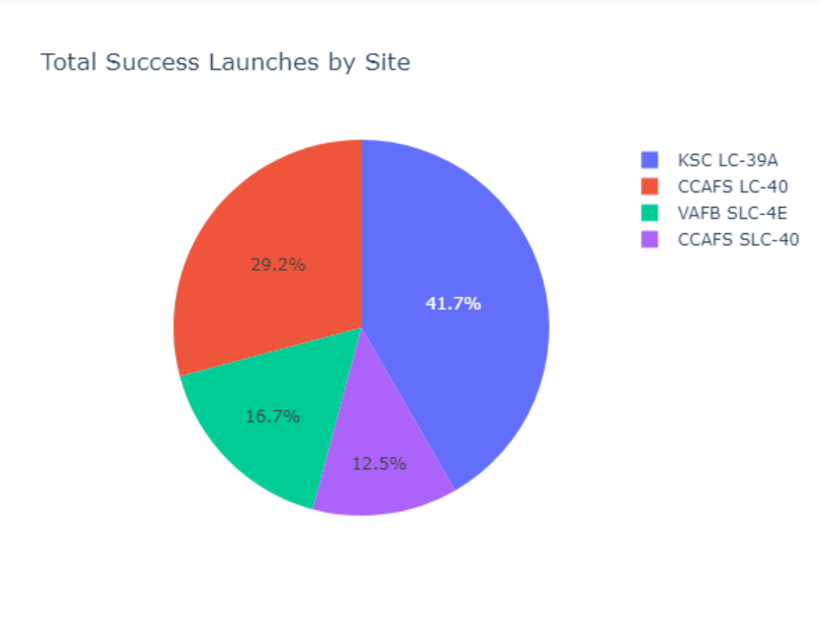
Build a Dashboard with Plotly Dash

Launch Success Count for All Sites

Findings

- **Concentration of Success:** The majority of successful launches are concentrated at specific sites, particularly CCAFS LC-40 and KSC LC-39A.
- **Impact of Launch Site:** The launch site plays a significant role in the success of the missions. This suggests that factors such as infrastructure, location, and support facilities at the launch site are crucial for mission success.
- **Site Performance Comparison:** By visualizing the success rates, it becomes evident which sites have been more reliable and effective for successful missions.

These insights can guide future decisions on site selection for launches to optimize the chances of success.



Explanation

This pie chart illustrates the distribution of successful launches across various SpaceX launch sites. The key elements in the screenshot include:

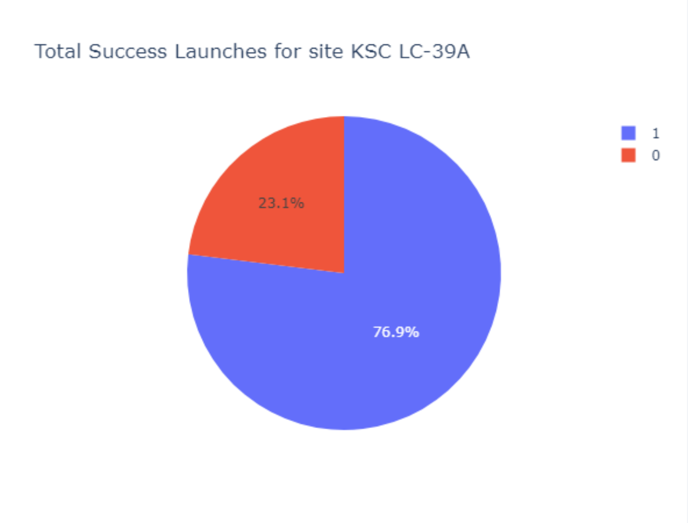
- **Segments:** Each segment represents a different launch site.
- **Labels:** The labels on the segments indicate the name of the launch site.
- **Percentages:** The percentages show the proportion of successful launches from each site.

Launch Site with Highest Launch Success Ratio

Findings

- **High Success Rate:** The launch site depicted in this pie chart has a success rate of 76.9%, indicating that the majority of launches from this site have been successful.
- **Site Efficiency**: This high success ratio underscores the efficiency and reliability of the launch site, making it a preferred location for SpaceX launches.
- **Strategic Importance**: The success rate highlights the strategic importance of this launch site for SpaceX's operations, suggesting that it may have superior infrastructure, better support systems, or other advantageous factors contributing to its high success rate.

Understanding the success rates of different launch sites helps in strategic planning and resource allocation to ensure higher mission success in future launches.



Explanation

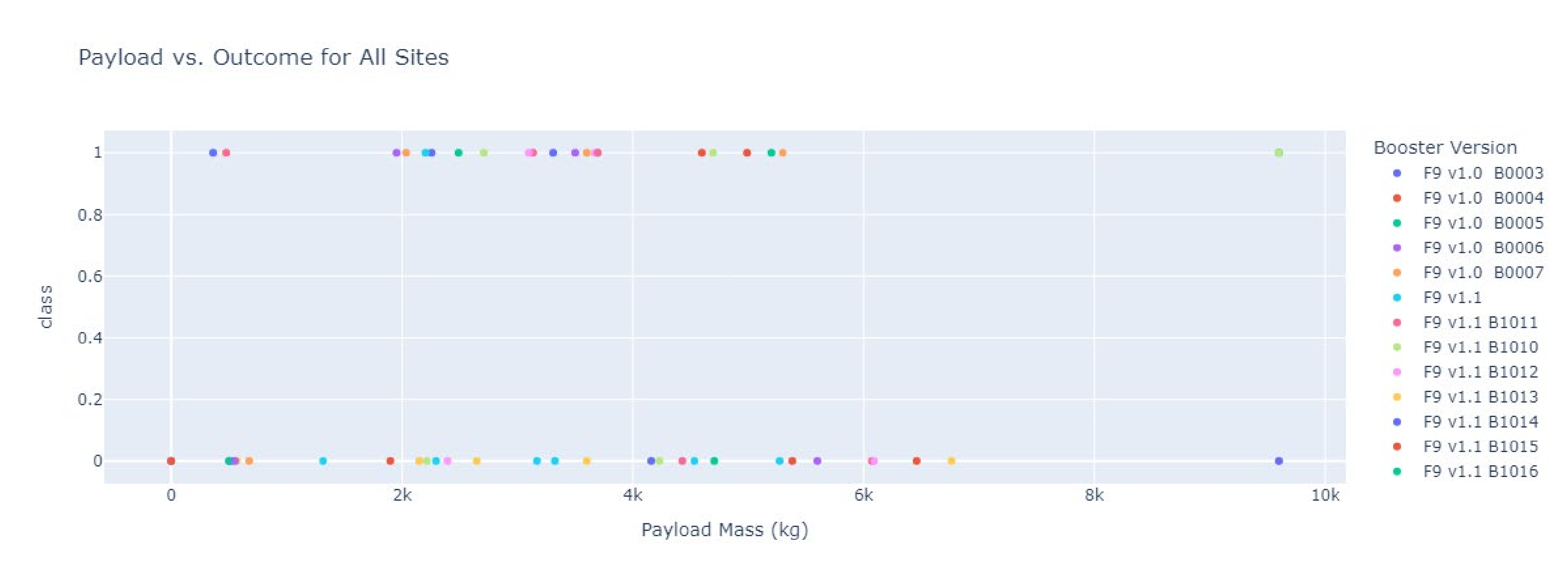
This pie chart highlights the success ratio of launches at the launch site with the highest success rate.

- **Segments**: The chart is divided into two segments: successful launches and unsuccessful launches.
- **Labels**: Each segment is labeled to indicate the outcome of the launches.
- **Percentages**: The success and failure rates are shown as percentages.

Payload vs. Launch Outcome Across All Sites with Different Payload Selections

Payload Range from 0 to 9600 Kg

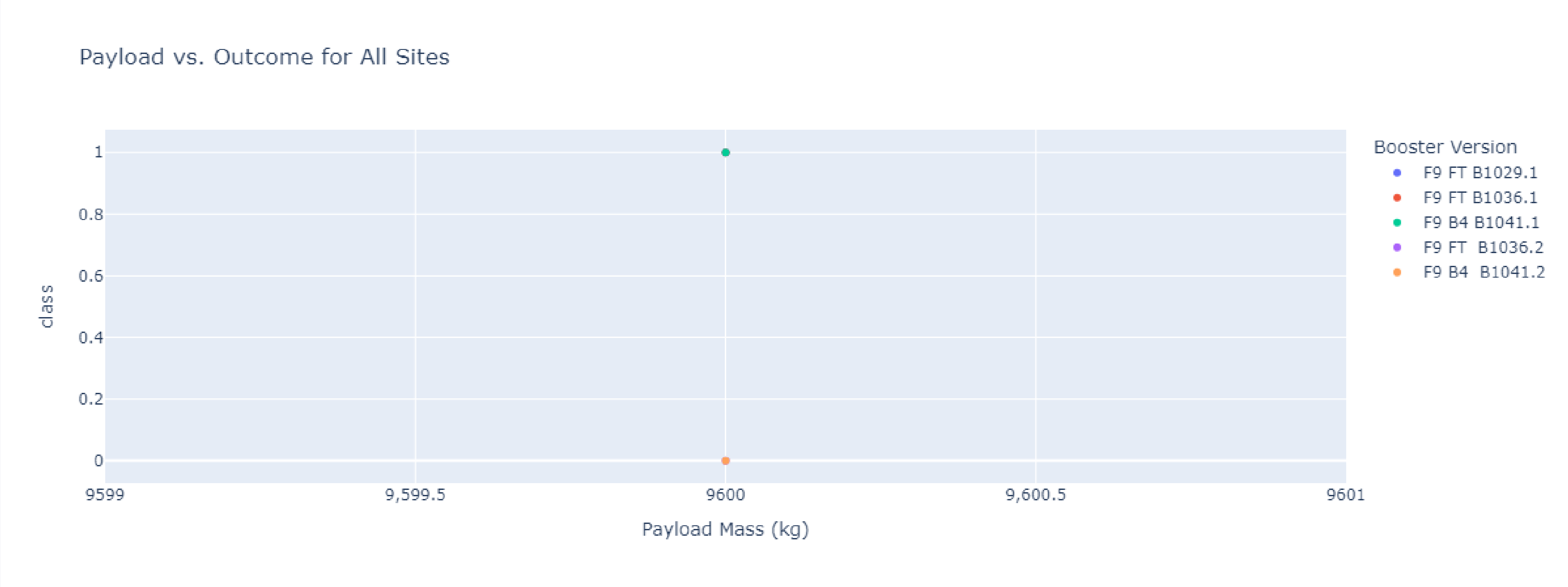
- The scatter plot shows the relationship between payload mass and the launch outcome (success or failure) across all launch sites.
- Each dot represents a launch, color-coded by the booster version used.
- From the scatter plot, it can be observed that lower payload masses (0 to 4000 Kg) have a mix of successful and failed launches across various booster versions. However, as the payload mass increases, the launches tend to show a higher rate of success, particularly for payloads between 6000 to 8000 Kg, highlighting a trend of stability in launch outcomes as payload mass increases.



Payload vs. Launch Outcome Across All Sites with Different Payload Selections

Payload Range Close to Maximum (9600 Kg)

- In this view, the payload mass slider is set to focus on the higher end of the payload spectrum (around 9600 Kg).
- It is evident from the scatter plot that the largest payloads tend to be associated with successful launches.
- Specific booster versions such as "F9 FT B1036.1" and "F9 FT B1041.2" have consistently handled these heavier payloads with success, demonstrating their reliability for heavy payload missions.



Payload vs. Launch Outcome Across All Sites with Different Payload Selections

Important Elements and Findings:

- **Success Rate by Payload**: Higher payload masses, particularly in the range of 7000 Kg and above, appear to have a higher success rate, suggesting that missions carrying heavier payloads are generally well-prepared and more likely to succeed.
- **Booster Versions**: Certain booster versions, such as the "F9 FT" series, stand out for their success in handling heavier payloads, indicating their robustness and reliability for high-stakes missions.
- **Implication for Future Launches**: For missions that require the transport of heavy payloads, these findings suggest that leveraging booster versions with a proven track record (such as F9 FT B1036.1) is critical for ensuring mission success.

This analysis underscores the importance of both payload mass and booster version in determining the success of space launches, with heavier payloads showing a higher correlation with successful outcomes in this dataset.

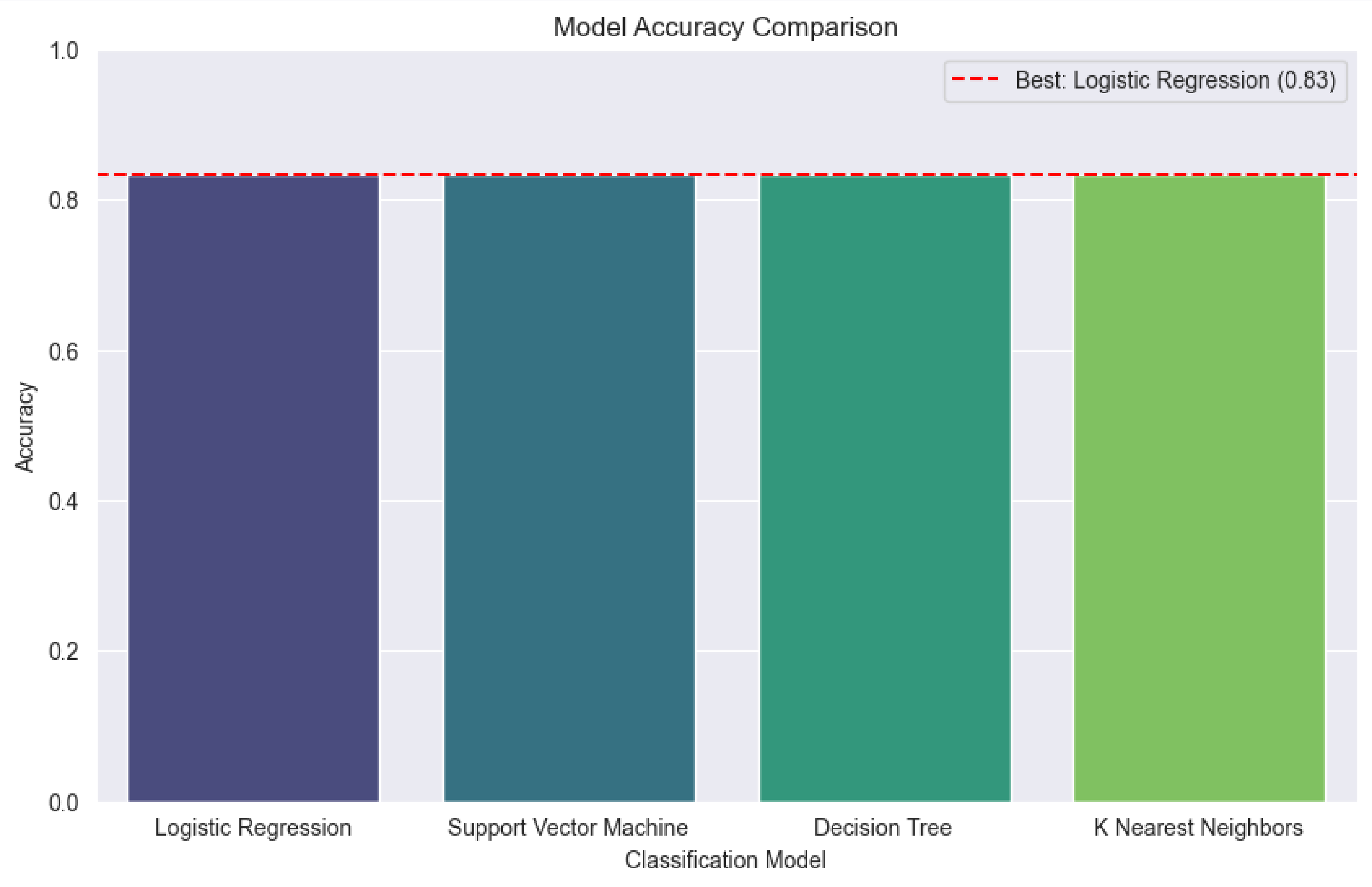
Section 5

Predictive Analysis (Classification)

Classification Accuracy

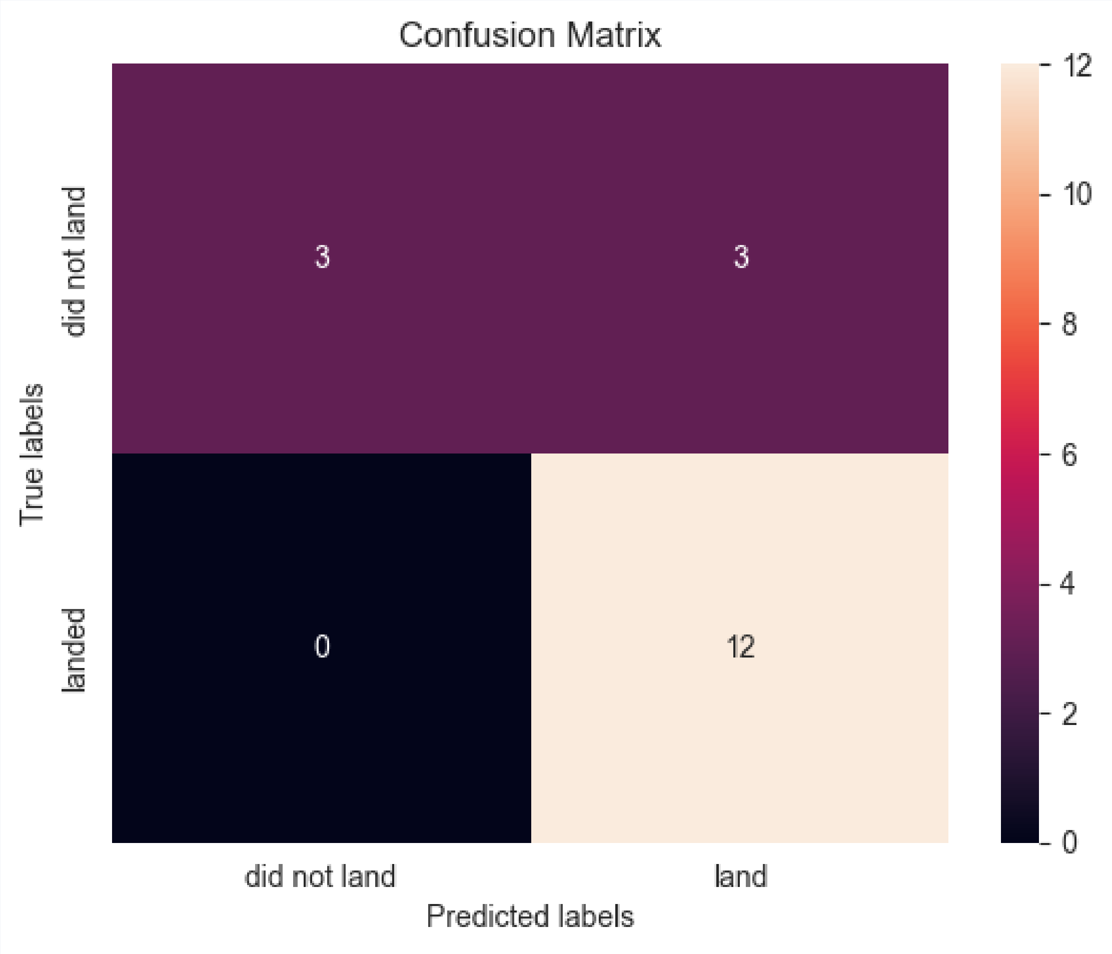
Findings:

- The bar chart visually compares the accuracy of all the built classification models.
- The red dashed line and legend indicate the best performing model and its accuracy.
- By running the above code, you'll be able to visualize the performance of each classification model and easily identify which one has the highest classification accuracy.



Confusion Matrix

The confusion matrix provides a detailed breakdown of the model's performance. A high number of true positives and true negatives compared to false positives and false negatives indicates good model performance, which validates the accuracy of the Logistic Regression model.



Conclusions

1. **Comprehensive Data Analysis:** Different data sources were analyzed, refining conclusions along the process to ensure accuracy and reliability.
2. **Best Launch Site:** The best launch site is KSC LC-39A, identified based on the number of successful launches and favorable conditions for various missions.
3. **Payload Weight and Risk:** Launches above 7,000kg are less risky, indicating that heavier payloads have a higher success rate, potentially due to better planning and more robust launch systems.
4. **Improvement Over Time:** Although most mission outcomes are successful, successful landing outcomes have shown improvement over time, likely due to the continuous evolution of processes and advancements in rocket technology.
5. **Predictive Modeling:** Logistic Regression has proven effective in predicting successful landings, which can be leveraged to increase profits by reducing the risk of failures and optimizing launch operations.

Appendix

The Jupyter Notebook, which contains all the code, visualizations, and analyses performed during this project, has been uploaded to GitHub along with the associated database.

Thank you!