IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Grzegorz Bednarski
30 Jan 2025

Outline

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

Introduction

The modern space race is not just about space exploration but also about optimizing costs and mission efficiency. SpaceX has revolutionized the aerospace industry with its reusable Falcon 9 rocket technology. Landing the first stage of the rocket enables the reuse of boosters, significantly reducing mission costs and increasing launch frequency.

By analyzing SpaceX mission data, we can gain deeper insights into the factors affecting launch and landing success and use Data Science techniques to predict future outcomes

Problems You Want to Find Answers To

1. How has the landing success rate of Falcon 9 evolved over time?
2. What factors influence mission success?
   - Weather, launch site, orbit type, payload mass.
3. Can we predict the success of a rocket landing?
   - Using classification models to forecast outcomes.
4. Which launch sites and orbit types are the most efficient?
5. What are the key trends in mission frequency and success rates?

Section 1

Methodology

Methodology – Executive summary

Overview of the Approach

To predict and analyze SpaceX Falcon 9 rocket landing success, we utilized data science techniques, including data collection, exploratory analysis, visualization, and machine learning models.

1. Data Collection
    - Extracted launch data from SpaceX API and web-scraped from Wikipedia.
    - Stored and cleaned data for further analysis.
2. Exploratory Data Analysis (EDA)
    - Visualized launch trends (success rate by site, payload impact, and orbital outcomes).
    - Used interactive Folium maps to analyze geographical launch success distribution.
    - Found that heavier payloads and specific launch sites influenced mission success.
3. Machine Learning for Prediction
    - Trained four classification models to predict launch success:
        - Support Vector Machine (SVM) – Best performing model
        - Logistic Regression
        - Decision Tree
        - K-Nearest Neighbors (KNN)
    - Key Insights from ML Models:
        - SVM achieved the highest accuracy, showing strong predictive power.
        - KSC LC-39A had the highest success rate, making it the most reliable launch site.
        - Lighter payloads had a higher probability of success, but recent boosters improved heavy payload success rates.
4. Dashboard Development
    - Built an interactive Plotly Dash dashboard to visualize:
        - Success rate per launch site (pie charts).
        - Payload vs. launch outcome (scatter plots).
        - Custom filtering and dynamic visualizations.
5. Key Findings & Business Impact
    - Data-driven decision-making can optimize launch site selection and payload configurations.
    - Machine learning models can help SpaceX predict and minimize mission failures.
    - Interactive dashboards provide stakeholders with real-time insights into mission performance.

Final Takeaway

By leveraging data science, machine learning, and interactive analytics, we successfully built a predictive model and data-driven insights to support SpaceX's goal of optimizing launch success and mission planning.

Data Collection

The dataset for this project was collected from two primary sources:

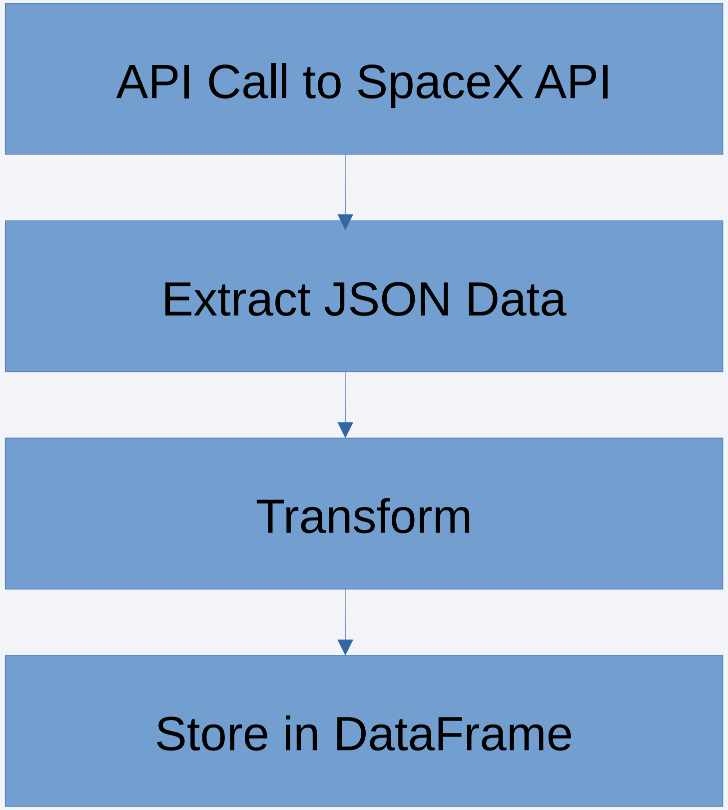
1. SpaceX API (spacexdata.com API)
    - Provides structured data on Falcon 9 and Falcon Heavy launches, including launch dates, success/failure outcomes, payload details, and landing sites.
    - API calls were used to fetch data programmatically in JSON format.
2. Wikipedia Web Scraping (List of Falcon 9 and Falcon Heavy launches)
    - Additional information, such as mission names, detailed descriptions, and success status, was extracted from Wikipedia using web scraping techniques.
    - Data was scraped using Python libraries like BeautifulSoup and Requests to collect relevant tabular data.

Data Collection – SpaceX API

Additional mission details were obtained by scraping the "List of Falcon 9 and Falcon Heavy launches" page on Wikipedia. Using Python’s BeautifulSoup and Requests, we extracted tabular data, including mission names, customer payloads, and specific launch details. The scraped data was cleaned and merged with the API dataset for a more comprehensive analysis.
Key Steps:

- Identified and accessed relevant tables on Wikipedia.
- Extracted mission data using BeautifulSoup.
- Cleaned and structured the data for integration.

These combined sources provided a robust dataset for evaluating SpaceX’s launch performance and predicting future outcomes.
The complete source code is available at the following link:



Source code

Data Collection - Scraping

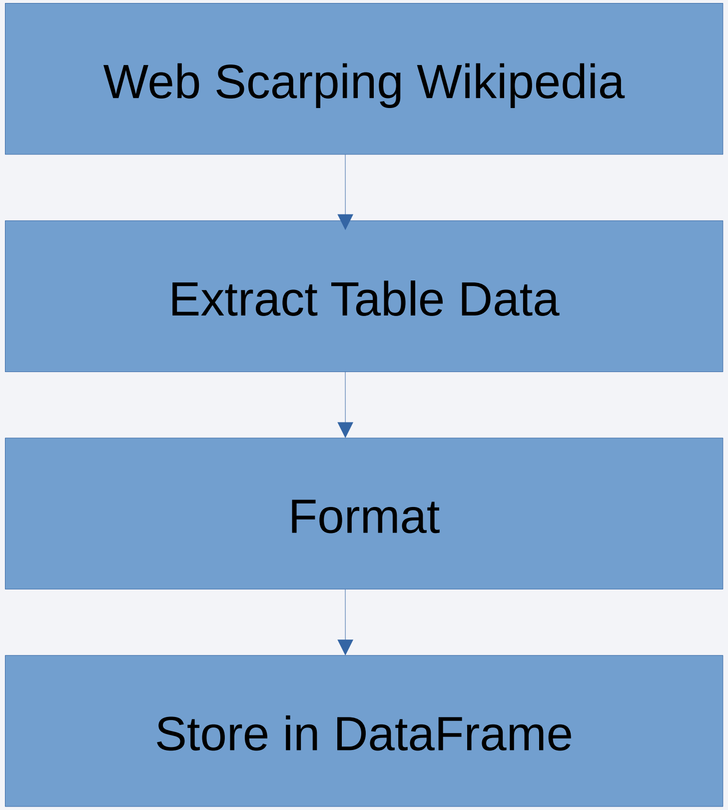
The SpaceX API provides structured data on all Falcon 9 and Falcon Heavy launches. Using API calls, we retrieved key information such as launch dates, rocket versions, payload details, launch sites, and landing outcomes. The data was fetched in JSON format, processed, and stored for further analysis.

Key Steps:

- Sent API requests to retrieve SpaceX launch data.
- Extracted relevant fields (e.g., mission name, success status, orbit type).
- Converted JSON data into a structured DataFrame.

The complete source code is available at the following link:

Source code



Data Wrangling

To ensure the dataset was clean and structured for analysis, we performed several key data wrangling steps:

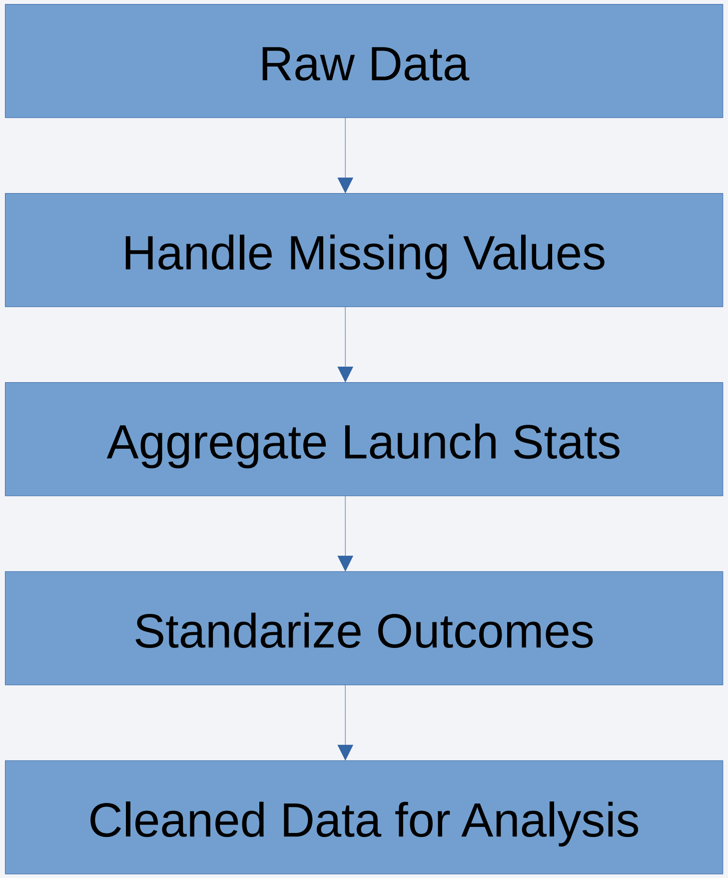
1. Handling Missing Values: empty cells were replaced with the mean value of the respective column to maintain data consistency.

2. Aggregating Launch Data: number of launches per launch site and per orbit were calculated

3. Standardizing Outcome Values: The landing success/failure outcomes were converted into a binary format (True/False) for easier analysis and predictive modeling.

The complete source code is available at the following link:

Source code



EDA with Data Visualization

Charts used in the analysis:

1. Line Chart: Average Success Rate Over Time. To visualize the trend of the average success rate of SpaceX missions over the years.

2. Bar Chart: Number of Launches by Site. To compare the number of launches from different locations and identify the most frequently used sites.

3. Scatter Plot: Payload Mass vs. Landing Outcome. To analyze the relationship between payload mass and landing outcome, helping to identify key factors affecting mission success.

4. Bar Chart: Success Rate by Orbit Type. To evaluate which orbit types have higher mission success rates.

5. Scatter Plot: Payload Mass vs. Orbit Type. To study the relationship between payload mass and target orbit type, providing insights into mission preferences.

The full notebook with visualizations will be linked later.

Source code

EDA with SQL

SQL Queries Performed:

1. Unique launch sites. To identify all unique launch sites used by SpaceX.
2. Number of launches from each launch site. To determine which launch sites are most frequently used.
3. Total payload mass for each customer. To analyze which customers have sent the largest payload mass via SpaceX.
4. Average payload mass for each booster version. To compare the performance of different booster versions in terms of average payload mass.
5. Date of the first successful landing on a ground pad. To identify when the first successful ground pad landing occurred.
6. Boosters that landed on a drone ship with a payload mass between 4000 and 6000 kg. To identify boosters that meet specific landing and payload mass criteria.
7. Total number of successful and failed missions. To summarize the number of missions based on landing outcomes.
8. Boosters that carried the maximum payload mass. To identify boosters that carried the heaviest payloads.
9. Failed drone ship landings in 2015. To analyze failed drone ship landings in a specific year.
10. Ranking of landing outcomes between 2010-06-04 and 2017-03-20. To rank landing outcomes within the specified time period.

The full notebook with SQL queries will be linked later.

Source code

Build an Interactive Map with Folium

Map Elements Added:

1. Launch Site Markers:
   - Description: Placed markers at each SpaceX launch site location.
   - Purpose: To visually identify and differentiate the geographical positions of all SpaceX launch sites on the map.
2. Launch Success/Failure Markers:
   - Description: Utilized color-coded markers to represent the outcomes of launches (e.g., green for success, red for failure) at their respective locations.
   - Purpose: To provide a clear visual representation of launch outcomes, facilitating quick assessment of success rates across different sites.
3. Proximity Circles:
   - Description: Drew circles around each launch site to indicate a specified radius.
   - Purpose: To analyze the surrounding area of each launch site, which can be useful for assessing potential risk zones or understanding the site's geographical context.
4. Distance Lines:
   - Description: Added lines connecting launch sites to significant landmarks or between each other.
   - Purpose: To measure and display distances, aiding in logistical planning and spatial analysis.

The full interactive map created with Folium is available in the notebook linked below.

Link to the notebook will be provided here.

Source code

Build a Dashboard with Plotly Dash

Implemented Plots and Interactions:

1. Launch Site Selection Dropdown:
    - Description: A dropdown menu allowing users to select a specific launch site or view data from all sites.
    - Purpose: To enable users to filter the data based on launch site, facilitating focused analysis on a particular site or an overview of all sites.
2. Success Pie Chart:
    - Description: A pie chart displaying the distribution of successful and failed launches.
    - Purpose: To provide a clear visual representation of the success rate of launches, aiding in assessing the reliability of different launch sites.
3. Payload Range Slider:
    - Description: An interactive slider allowing users to select a range of payload masses.
    - Purpose: To enable users to filter the data based on payload mass, facilitating analysis of how payload weight impacts launch success.
4. Success-Payload Scatter Chart:
    - Description: A scatter plot illustrating the correlation between payload mass and launch success, with points colored by booster version.
    - Purpose: To help users understand the relationship between payload mass and launch outcomes, and to assess the performance of different booster versions.

The full Plotly Dash application is available in the provided GitHub repository.

Source code

Predictive Analysis (Classification)

Model Development Process:

1. Data Preparation:
   - Feature Selection: Identified relevant features such as payload mass, launch site, orbit type, and booster version to predict the success of Falcon 9 first-stage landings.
   - Data Encoding: Applied one-hot encoding to categorical variables to convert them into numerical format suitable for machine learning algorithms.
   - Data Splitting: Divided the dataset into training and testing sets to evaluate model performance effectively.
2. Model Building:
   - Algorithms Implemented:
   - Logistic Regression: Utilized for its simplicity and interpretability in binary classification tasks.
   - Support Vector Machine (SVM): Chosen for its effectiveness in high-dimensional spaces.
   - Decision Tree: Implemented to capture non-linear relationships in the data.
   - K-Nearest Neighbors (KNN): Selected for its simplicity and effectiveness in capturing local data structures.
3. Model Evaluation:
   - Cross-Validation: Employed k-fold cross-validation to assess model performance and mitigate overfitting.
   - Performance Metrics: Evaluated models using accuracy, precision, recall, and F1-score to ensure a comprehensive understanding of performance.
4. Hyperparameter Tuning:
   - Grid Search: Conducted grid search to identify the optimal hyperparameters for each model, enhancing predictive accuracy.
5. Model Selection:
   - Best Performing Model: The Decision Tree classifier emerged as the top performer, achieving the highest accuracy and F1-score on the test data.

For detailed code and analysis, please refer to the

Source code

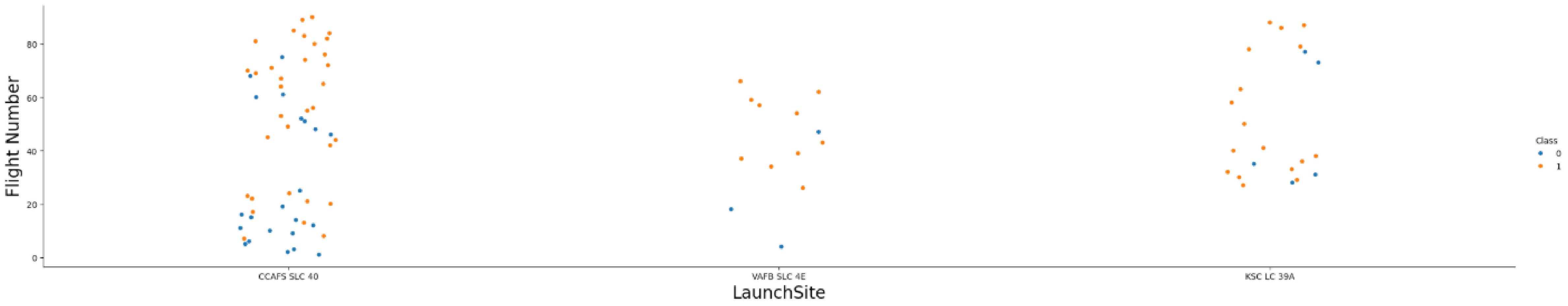
Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

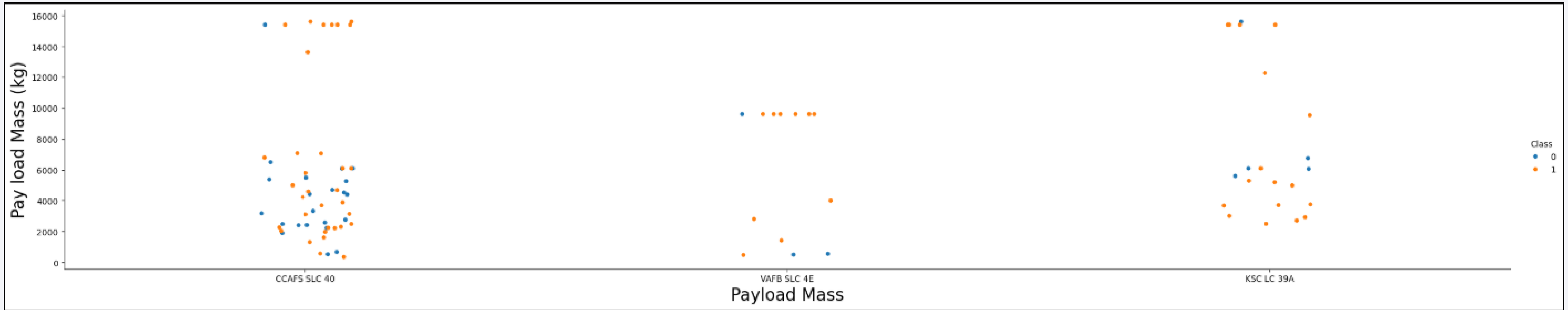


Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

Insights:

1. Increasing Flight Numbers at Active Sites:
   - Certain launch sites, such as CCAFS SLC-40 and KSC LC-39A, have a wide distribution of flight numbers, indicating they have been used consistently over time.
   - Other sites may have a more limited range of flight numbers, suggesting they were only used for specific mission types or for a limited period.
2. Gaps in Launch Sites:
   - If certain launch sites show gaps in flight number sequences, it could indicate periods of inactivity due to maintenance, regulatory changes, or shifting launch strategies.
3. Site-Specific Trends:
   - If a site appears only for a certain range of flight numbers, it may indicate that it was either introduced at a later stage (e.g., KSC LC-39A) or phased out after early launches.

Payload vs. Launch Site

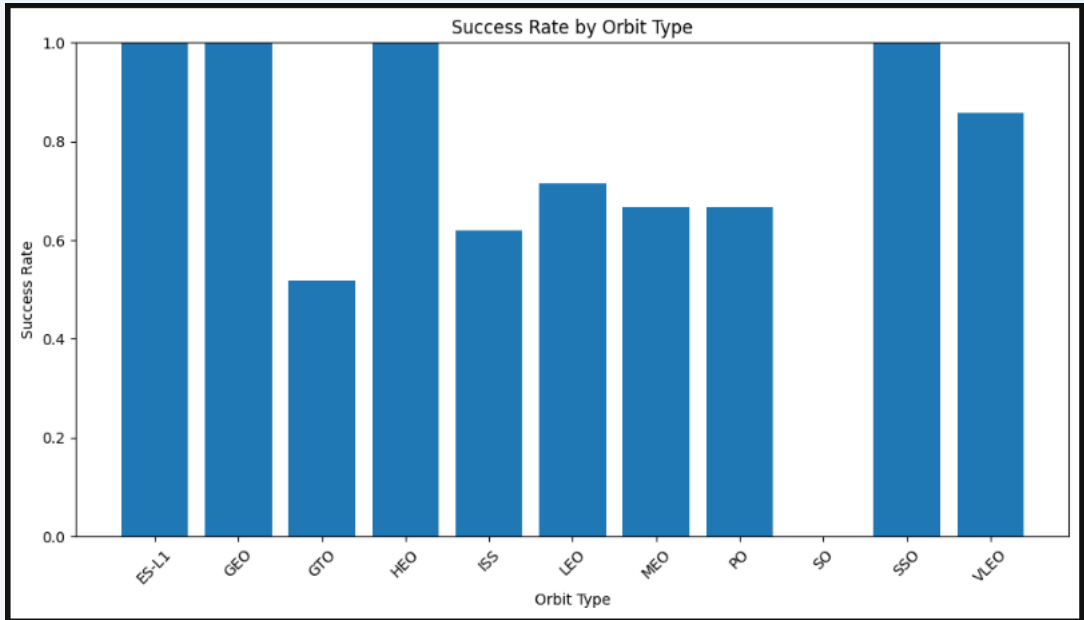


Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Insights:

1. Certain Sites Handle Heavier Payloads More Frequently:
   - If a specific launch site (e.g., KSC LC-39A) consistently handles heavier payloads, it may indicate that it is equipped with better infrastructure for high-mass launches, such as crewed missions or deep-space payloads.
   - Conversely, sites like CCAFS SLC-40 may launch a broader range of payloads, including smaller satellites.
2. Limited Payload Ranges at Certain Sites:
   - If a launch site primarily supports lower payload masses, it may be optimized for lightweight satellite launches rather than heavy-lift missions.
   - For example, VAFB SLC-4E may have a narrower payload mass distribution due to specific mission types such as polar or sun-synchronous orbit deployments.
3. Trends in Payload Capacity Over Time:
   - If later launches show heavier payloads at a specific site, this could indicate improvements in rocket technology (e.g., Falcon Heavy) or increasing demand for higher-mass missions.
4. Outliers and Special Missions:
   - Any extreme payload masses observed at a particular site could correspond to unique missions (e.g., Falcon Heavy launches carrying large payloads to geostationary or deep-space orbits).
   - Identifying these outliers can help determine which sites are used for specialized missions.

Success Rate vs. Orbit Type

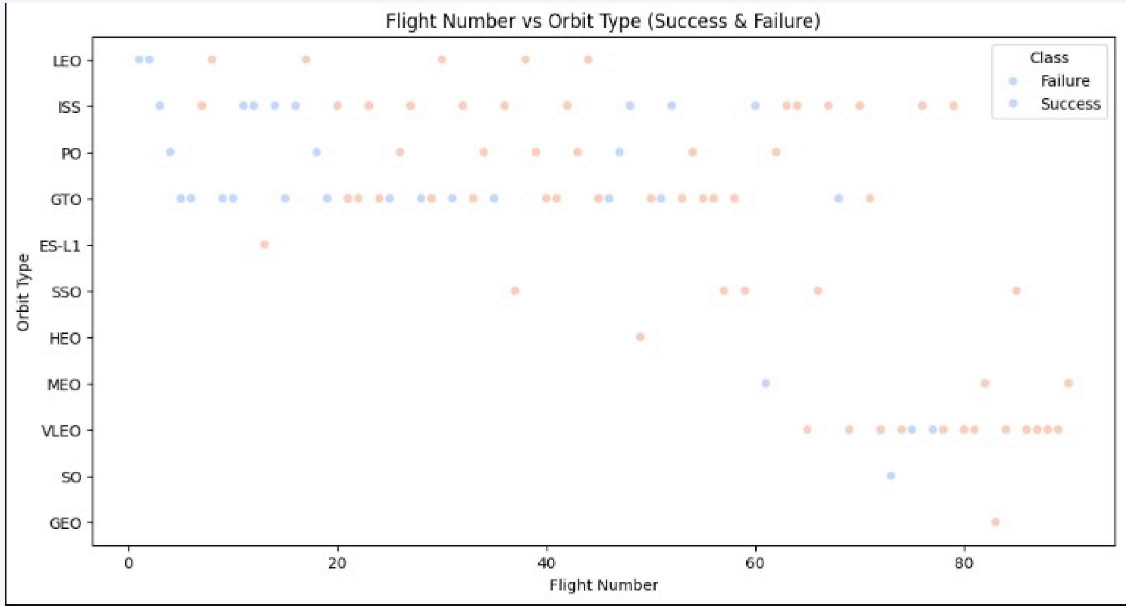


Analyze the plotted bar chart to identify which orbits have the highest success rates.

Insights:

1. Higher Success Rates for Certain Orbits:
   - If Low Earth Orbit (LEO) and Geostationary Transfer Orbit (GTO) show higher success rates, it may be due to the fact that these are the most frequently attempted orbits, allowing SpaceX to refine their launch techniques over time.
   - These orbits are commonly used for satellite deployments and cargo missions to the ISS, which are well within Falcon 9's capabilities.
2. Lower Success Rates for Challenging Orbits:
   - If certain orbit types (e.g., Polar, Heliocentric, or Interplanetary Orbits) show lower success rates, it may indicate that these missions are inherently more challenging due to factors such as:
     - Longer travel distances.
     - Higher velocity requirements.
     - More complex mission parameters (e.g., Mars or Moon missions).
3. Trends Over Time:
   - If success rates for specific orbits have improved over time, it suggests advancements in SpaceX's technology and experience with particular mission types.
   - This could indicate increased confidence in attempting more complex missions in the future.
4. Impact on Future Mission Planning:
   - If some orbit types show consistently low success rates, SpaceX might focus on further improving technology and mission planning to mitigate risks.
   - Alternatively, higher success rates in frequently used orbits (e.g., LEO and GTO) may encourage more commercial satellite launches in these regions.

Flight Number vs. Orbit Type

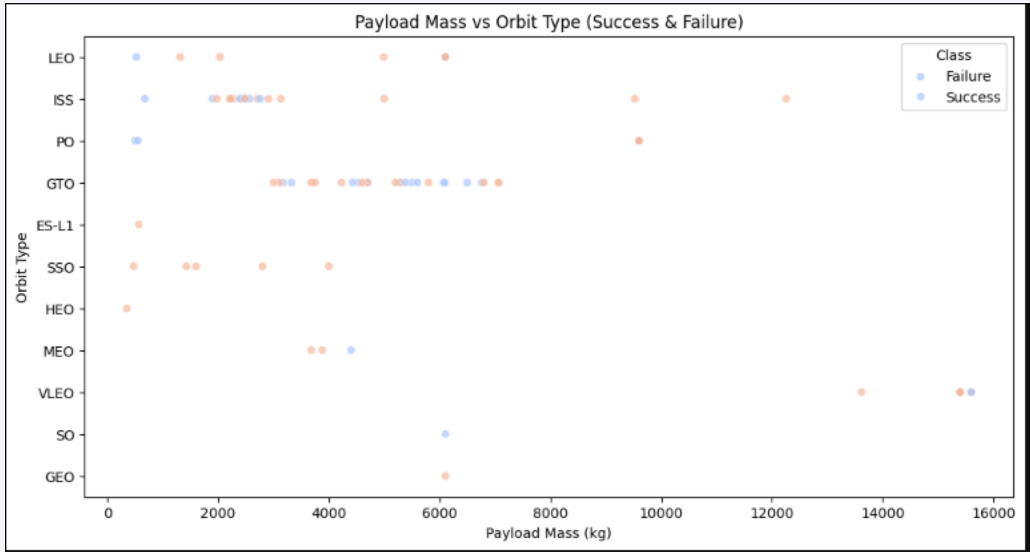


You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Insights:

1. Early Missions Focused on Specific Orbit Types:
   - If lower flight numbers are concentrated in orbits like Low Earth Orbit (LEO) or Geostationary Transfer Orbit (GTO), it suggests that SpaceX initially prioritized these relatively simpler and well-established mission types.
   - These orbits are commonly used for satellite launches and cargo missions to the International Space Station (ISS).
2. Increasing Variety of Orbits Over Time:
   - If later flight numbers show a wider spread across multiple orbit types, it indicates that SpaceX expanded its capabilities to support a greater diversity of missions.
   - The introduction of Polar, Sun-Synchronous, and Interplanetary orbits in later flights would suggest increased confidence in handling complex mission requirements.
3. Flight Experience vs. Risk-Taking:
   - If missions targeting higher-risk orbits (e.g., interplanetary missions) appear only at later flight numbers, it suggests that SpaceX strategically waited until they had sufficient technical expertise and proven reliability before attempting more difficult launches.
   - A lack of failed missions in certain orbit types at higher flight numbers would further confirm SpaceX's improved mission planning.
4. Potential Clusters for Specific Missions:
   - If there are clusters of flight numbers for particular orbit types, it may indicate:
     - Dedicated launch programs for specific customers (e.g., Starlink for LEO).
     - Batch missions of similar type grouped together.
     - Changes in SpaceX's business strategy over time (e.g., shifting focus to deep-space missions).

Payload vs. Orbit Type



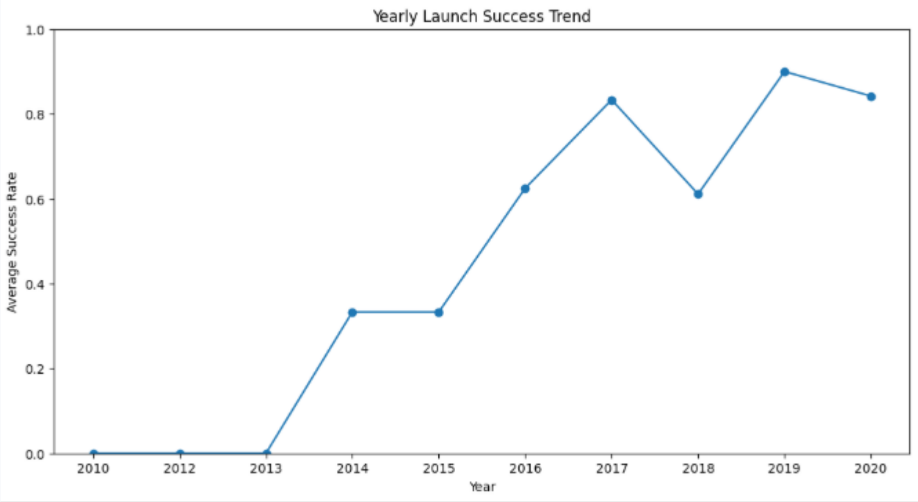
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Insights:

1. Lower Payload Mass for LEO and Sun-Synchronous Orbits (SSO):
   - If Low Earth Orbit (LEO) and Sun-Synchronous Orbit (SSO) show a wide range of payload masses, but mostly on the lower end, it indicates that these orbits are commonly used for lighter payloads, such as small satellites and Starlink missions.
   - LEO missions are more frequent because they require less fuel and lower launch energy.
2. Higher Payload Mass for GTO and Beyond:
   - If Geostationary Transfer Orbit (GTO), Polar, and Interplanetary orbits show heavier payloads, it suggests that missions to these orbits typically require more powerful rockets (e.g., Falcon Heavy for deep-space missions).
   - GTO is often used for communication satellites, which tend to be heavy due to the need for on-board propulsion systems.
3. Payload Constraints for Certain Orbits:
   - If some orbit types show a narrow range of payload masses, it may indicate technical or cost constraints that prevent heavier payloads from being sent to those orbits.
   - For example, Polar and Sun-Synchronous Orbits (SSO) might only accommodate specific satellite sizes due to their altitude and energy requirements.
4. Outliers Indicating Special Missions:
   - If a few points appear at extremely high payload masses, these could correspond to unique missions such as Falcon Heavy launches carrying interplanetary probes or large commercial payloads.
   - Notable outliers may include missions like Tesla Roadster to Heliocentric Orbit or deep-space exploration missions.
5. Correlation Between Payload Mass and Orbit Difficulty:
   - Generally, heavier payloads are associated with higher-energy orbits, meaning more fuel and advanced rocket technology are required.
   - If there is a trend where lighter payloads are consistently associated with specific orbits, it can help SpaceX optimize its launch strategies for efficiency.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

Insights:

1. Steady Growth in Successful Launches Over Time:
   - If the success rate increases year after year, it indicates continuous improvement in SpaceX's launch technology, mission planning, and operational efficiency.
   - A rising trend suggests that SpaceX has optimized its rocket reusability and reliability over time.
2. Early Years Had a Lower Success Rate:
   - If the earlier years show fewer successful launches, it is likely due to experimental and developmental stages of Falcon 9 technology.
   - Initial failures were expected as SpaceX refined its landing techniques and propulsion systems.
3. Significant Jumps Indicating Breakthroughs:
   - If there are sharp increases in success rate in certain years, these could be linked to:
     - The introduction of Falcon 9 Block 5, a more reliable and reusable rocket version.
     - The first successful booster landings and subsequent reusability of rockets.
     - Increased demand for commercial satellite launches (e.g., Starlink program).
4. Identifying Periods of Setbacks or Challenges:
   - If there are dips or stagnation in success rates, these may be caused by:
     - Technical failures (e.g., Falcon 9 explosion incidents).
     - Mission complexity increases, such as interplanetary or crewed missions.
     - Regulatory or economic factors affecting launch schedules.
5. Acceleration in Recent Years:
   - If there is a steep upward trend in the last few years, it suggests that:
     - SpaceX has achieved higher launch frequency and reliability.
     - The Starlink program has significantly contributed to the increasing number of launches.
     - Falcon 9's reusability has allowed for cost-effective and frequent launches.

All Launch Site Names

Explanation:

- SELECT DISTINCT is used to retrieve unique values from the specified column.
- Launch_Site is the column containing the names of the launch sites.
- FROM SPACEXTBL specifies the table from which to retrieve the data.

Expected Outcome:

Executing this query will yield a list of unique launch site names, providing a clear view of the different locations utilized by SpaceX for their launches.

This foundational step is crucial for subsequent analyses, such as assessing launch frequencies per site, evaluating site-specific success rates, and understanding the strategic importance of each location in SpaceX's operations.

```
In [11]: %sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL;

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

Out[11]:

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

Launch Site Names Begin with 'CCA'

Explanation:

- SELECT * → Retrieves all columns from the dataset.
- FROM SPACEXTBL → Specifies the table where the data is stored.
- WHERE Launch_Site LIKE 'CCA%' → Filters the results to include only records where the launch site name starts with "CCA".
  - The LIKE operator is used for pattern matching.
  - 'CCA%' means any launch site name that starts with "CCA", regardless of what follows.

Expected Outcome:

A table listing all SpaceX launches conducted from launch sites whose names start with "CCA" (e.g., CCAFS SLC-40).

This data will be useful in evaluating the role of Cape Canaveral as a key launch facility and comparing its performance with other launch sites.

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

```
%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
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO
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)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)

Total Payload Mass

```
In [15]:  %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';

           * sqlite:///my_data1.db
          Done.

Out[15]:  SUM(PAYLOAD_MASS__KG_)
                           45596
```

Explanation:

- SELECT SUM(PAYLOAD_MASS__KG_) → Retrieves the total sum of all payload masses.
- FROM SPACEXTBL → Specifies the dataset being queried.
- WHERE Customer = 'NASA' → Filters the dataset to include only launches where NASA was the customer.

Expected Outcome:

A single numerical value representing the total payload mass (in kg) launched for NASA.

This value provides insights into how much cargo and satellites SpaceX has transported for NASA missions.

Average Payload Mass by F9 v1.1

```
In [16]:  %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';

 * sqlite:///my_data1.db
Done.
Out[16]:  AVG(PAYLOAD_MASS__KG_)
                         2928.4
```

Explanation:

- SELECT AVG(PAYLOAD_MASS__KG_) → Calculates the average payload mass from all relevant records.
- FROM SPACEXTBL → Specifies the dataset being queried.
- WHERE Booster_Version = 'F9 v1.1' → Filters the data to include only launches that used the Falcon 9 v1.1 booster version.

Expected Outcome:

A single numerical value representing the average payload mass (in kg) carried by Falcon 9 v1.1 missions.

This will help compare its payload performance with newer versions of Falcon 9, such as Block 5, to see how SpaceX has improved over time.

First Successful Ground Landing Date

```
In [17]:  %sql SELECT MIN("Date") FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';

 * sqlite:///my_data1.db
Done.

Out[17]:  MIN("Date")
          2015-12-22
```

Explanation:

- SELECT MIN(Date) → Finds the earliest (minimum) date in the dataset.
- FROM SPACEXTBL → Specifies the dataset to search within.
- WHERE Landing_Outcome = 'Success (ground pad)' → Filters results to include only records where the rocket successfully landed on a ground pad.

Expected Outcome:

A single date value representing the first successful landing on a ground pad.

This marks the beginning of SpaceX's ability to reuse boosters, significantly lowering launch costs.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [18]:  %sql SELECT DISTINCT("Booster_Version") FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)'\
              AND PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000;

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

Out[18]:

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

Explanation:

- SELECT Booster_Version → Retrieves the names of the boosters that meet the criteria.
- FROM SPACEXTBL → Specifies the dataset being queried.
- WHERE Landing_Outcome = 'Success (drone ship)' → Filters for missions where the booster successfully landed on a drone ship.
- AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 → Ensures that only launches with a payload mass between 4000 and 6000 kg are considered.

Expected Outcome:

A list of booster versions that successfully landed on a drone ship while carrying a payload between 4000 and 6000 kg.

This allows us to see which Falcon 9 versions were most capable of handling such missions.

Total Number of Successful and Failure Mission Outcomes

Explanation:

- SELECT Landing_Outcome → Retrieves the different landing outcomes (e.g., success, failure).
- COUNT(*) AS Outcome_Count → Counts the number of times each outcome occurred.
- FROM SPACEXTBL → Specifies the dataset being used.
- GROUP BY Landing_Outcome → Groups the results by each unique landing outcome, allowing for a count of each type.

Expected Outcome:

A summary table listing each type of landing outcome along with the total count of missions for that outcome.

This will help in understanding which types of landings have the highest success rates.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL GROUP BY Landing_Outcome;
```

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

Landing_Outcome	Outcome_Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

Explanation:

- SELECT Booster_Version → Retrieves the booster version(s) used for the heaviest payloads.
- FROM SPACEXTBL → Specifies the dataset to query.
- WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) → Filters results to include only launches where the payload mass was equal to the maximum recorded payload mass in the dataset.

Expected Outcome:

A list of booster versions that were used for launches carrying the heaviest payloads.

This will help in determining which rockets were used for the most demanding missions.

```
%sql SELECT DISTINCT("Booster_Version") FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ =\
    (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * 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

2015 Launch Records

```
%sql SELECT substr(Date, 6,2), Booster_Version, Launch_Site as month FROM SPACEXTBL\
    WHERE Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

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

substr(Date, 6,2)	Booster_Version	month
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Explanation:

- SELECT substr(Date, 6,2), Booster_Version, Launch_Site → Retrieves the launch month, booster version, and launch site for failed drone ship landings.
- FROM SPACEXTBL → Specifies the dataset to query.
- WHERE Landing_Outcome = 'Failure (drone ship)' → Filters the results to include only failed drone ship landings.
- AND substr(Date,0,5)='2015' → Ensures that only records from the year 2015 are included in the results.

Expected Outcome:

A table listing the dates, booster versions, and launch sites for all failed drone ship landings in 2015.

This will help in identifying which boosters struggled with drone ship landings during that period.

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

Explanation:

- SELECT Landing_Outcome, COUNT(*) AS Outcome_Count → Retrieves each landing outcome and the number of times it occurred.
- FROM SPACEXTBL → Specifies the dataset being queried.
- WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' → Filters the dataset to include only records from the given date range.
- GROUP BY Landing_Outcome → Groups results by landing type to count occurrences for each outcome.
- ORDER BY Outcome_Count DESC → Sorts the results in descending order, placing the most frequent outcomes at the top.

Expected Outcome:

A ranked list of landing outcomes between 2010 and 2017, showing which landing results were most common.

This ranking will show trends in SpaceX's landing progress, such as how often landings were successful vs. failed vs. not attempted.

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

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

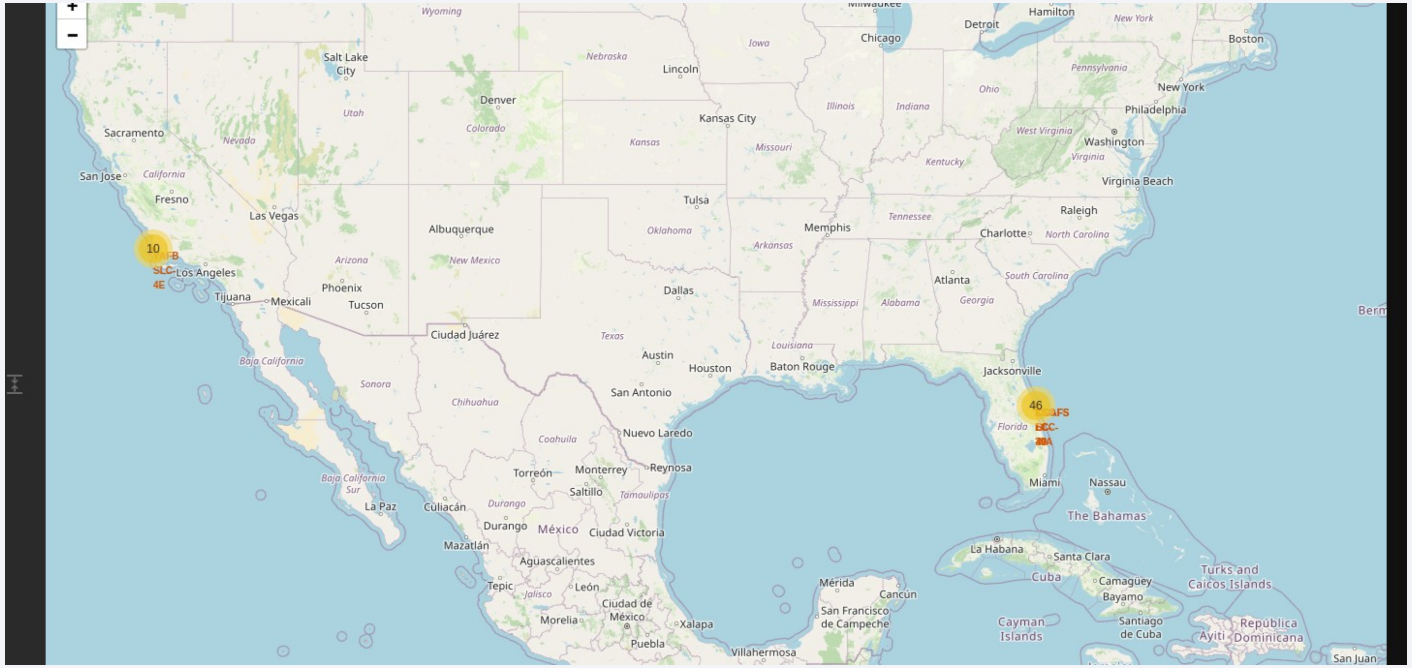
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

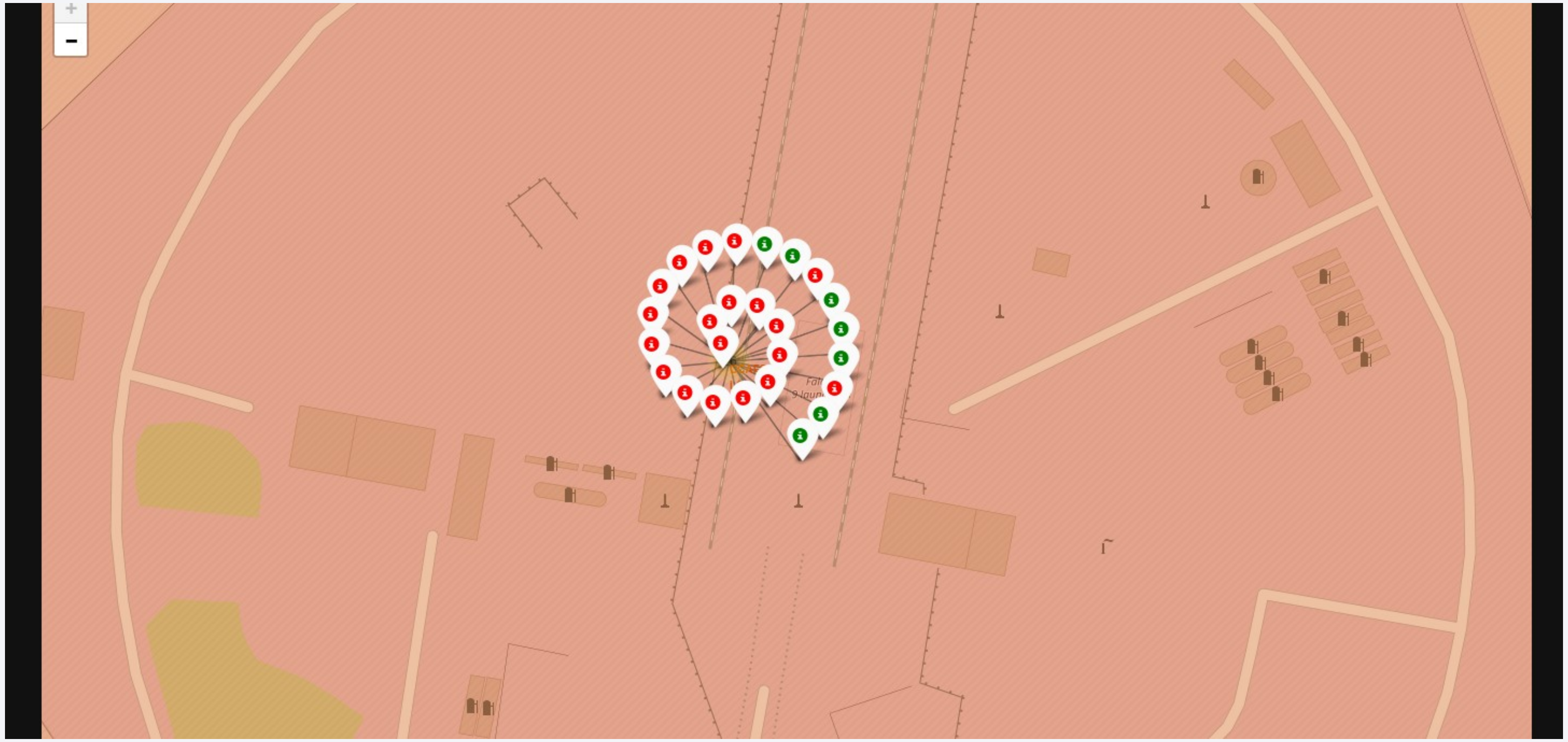
Geospatial Analysis of SpaceX Launch Sites

This map is an interactive Folium map displaying SpaceX launch sites in the United States. The markers represent different launch locations, and the marker clusters help visualize the number of launches at each site



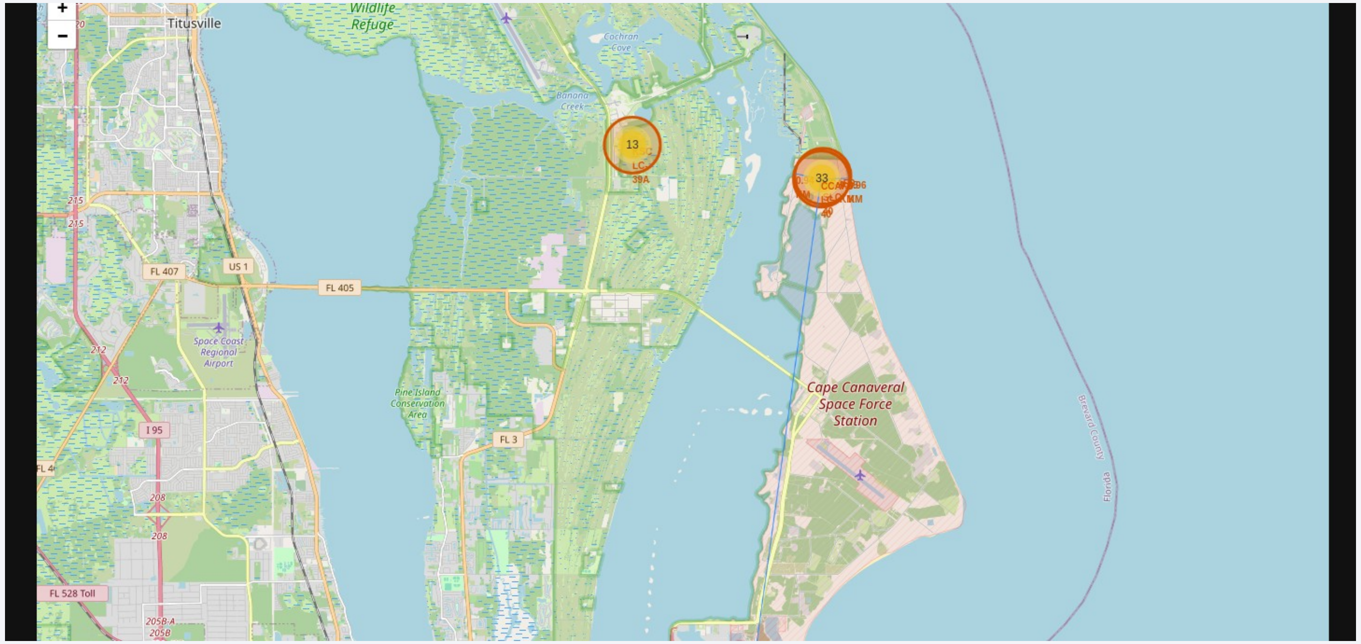
Zoomed-In View of a SpaceX Launch Site

This map provides a detailed, zoomed-in view of a specific SpaceX launch site, showing individual launch markers for past missions. The color-coded markers represent whether the launch was successful or failed.



Proximity Analysis of a SpaceX Launch Site

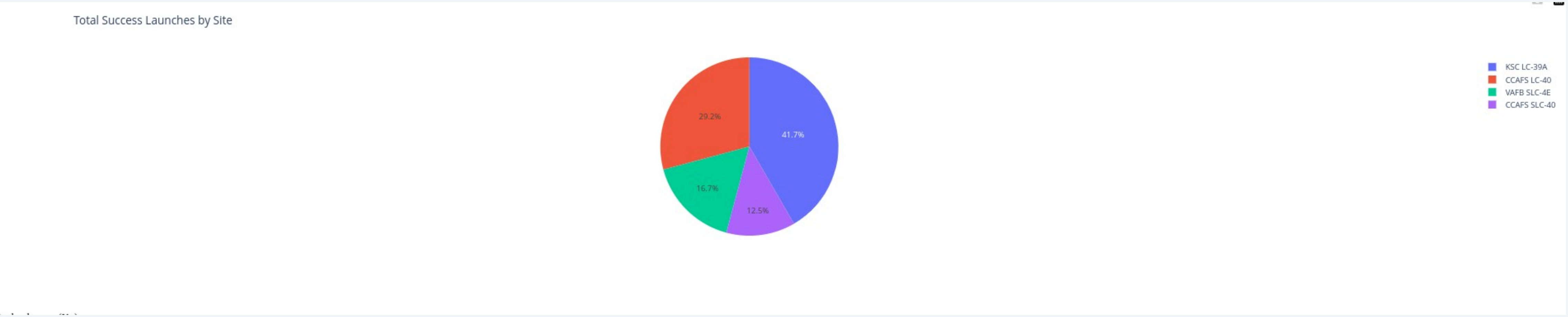
This map provides a zoomed-in view of a SpaceX launch site with additional proximity calculations to key infrastructure elements. The drawn lines and distance measurements offer geospatial insights into the site's accessibility and logistical advantages.



Section 4

Build a Dashboard with Plotly Dash

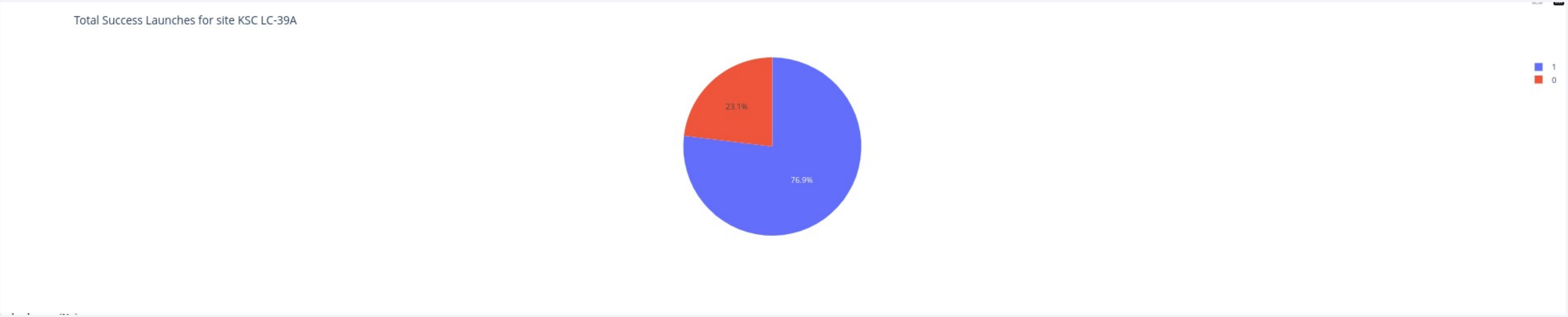
Success Rate of Launches by Launch Site



Findings & Insights:

- KSC LC-39A is the most frequently used site for successful launches (41.7%), likely due to its role in major NASA and commercial missions.
- CCAFS LC-40 follows with 29.2% of successful launches, another key launch site used extensively for commercial payloads.
- VAFB SLC-4E (16.7%) contributes a smaller share, as it is primarily used for polar orbit launches and sun-synchronous orbits.
- Duplicate Label Issue: There seems to be a duplication of CCAFS SLC-40 in the legend, which should be checked in the dataset to ensure correct labeling.

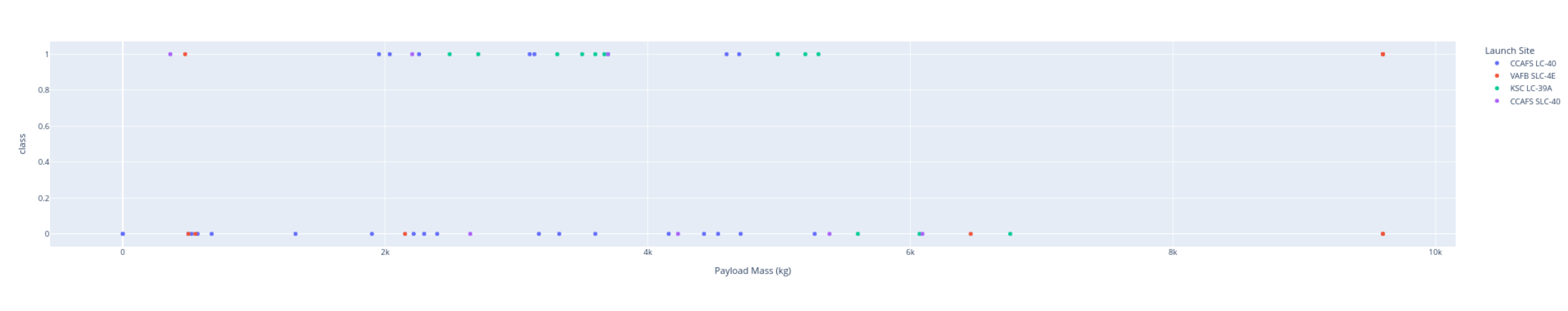
Launch Success Rate at KSC LC-39A



Findings & Insights:

- KSC LC-39A has the highest success rate among SpaceX launch sites (76.9%), making it a key location for critical missions.
- The failure rate (23.1%) is relatively low, showing that this site provides optimal conditions for safe and reliable launches.
- Why KSC LC-39A is So Successful?
  - Advanced infrastructure: Originally built for NASA's Apollo and Shuttle programs, now optimized for Falcon 9 and Falcon Heavy.
  - Favorable launch conditions: Less atmospheric turbulence and strategic positioning for various orbits.
  - Used for NASA & Crew Missions: SpaceX prioritizes high-profile and well-prepared missions from this location.

Analyzing the Impact of Payload Mass on Launch Success



Findings & Insights:

Success Rate at Different Payload Ranges

- Light Payloads (< 2000 kg)
  - Majority of successful launches at all sites.
  - Some failures observed, but success rates are relatively high.
- Medium Payloads (2000 - 6000 kg)
  - High success rate for KSC LC-39A and VAFB SLC-4E.
  - Some failures appear for CCAFS LC-40.
- Heavy Payloads (> 6000 kg)
  - Increased failure rate, especially at CCAFS LC-40.
  - KSC LC-39A shows higher reliability for heavy payloads, indicating its suitability for high-mass missions.

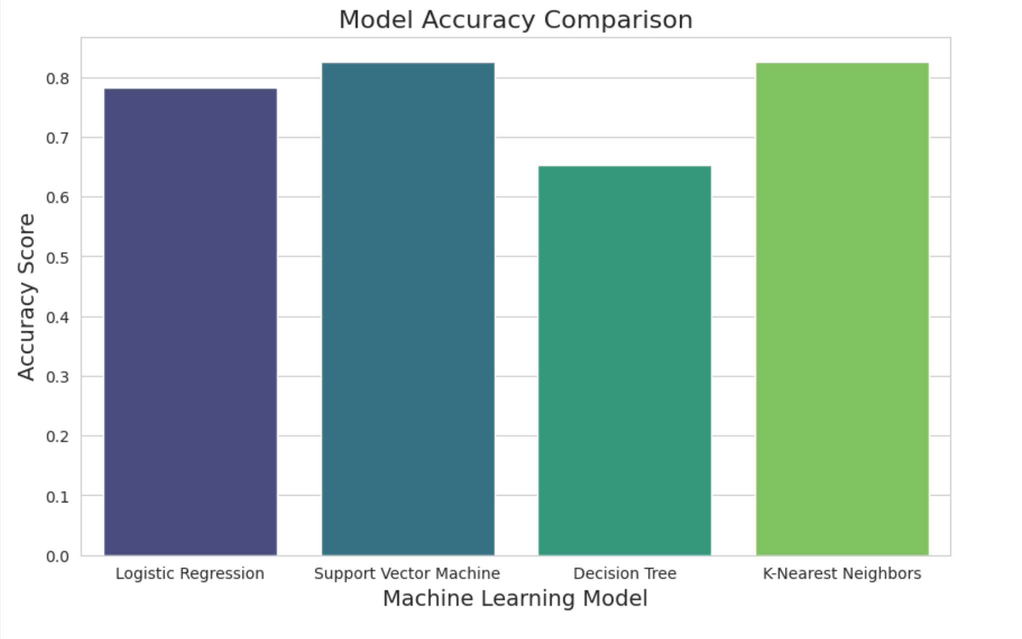
Which Launch Sites Have the Highest Success Rate?

- KSC LC-39A (Blue)
  - Appears to have consistent success across different payload ranges, especially above 4000 kg.
  - This site is SpaceX's preferred location for crewed missions and large satellite deployments.
- CCAFS LC-40 (Red)
  - Shows more failed launches at higher payload masses.
  - This suggests limitations in handling very heavy payloads.
- VAFB SLC-4E (Green)
  - Performs well for medium payloads, showing a high success rate.
  - This is expected since VAFB is used for polar and sun-synchronous orbit missions.

Section 5

Predictive Analysis (Classification)

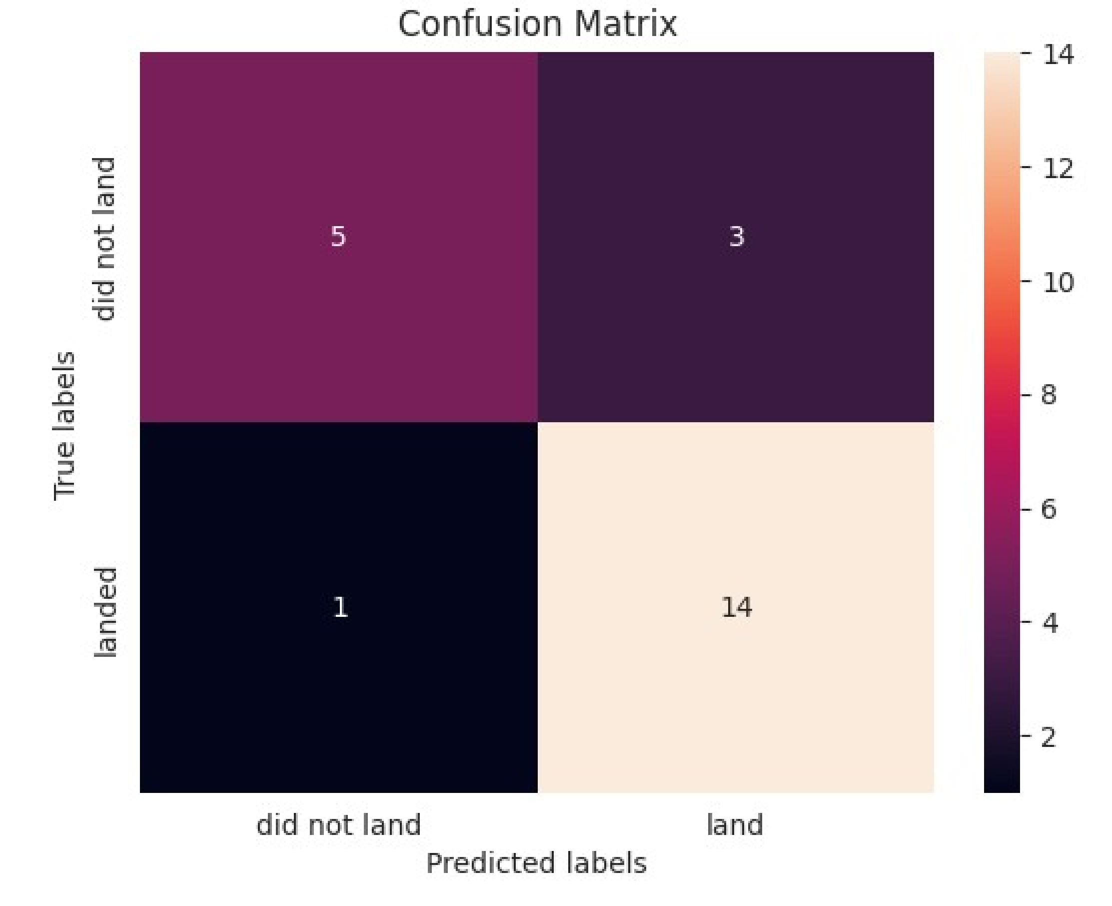
Classification Accuracy



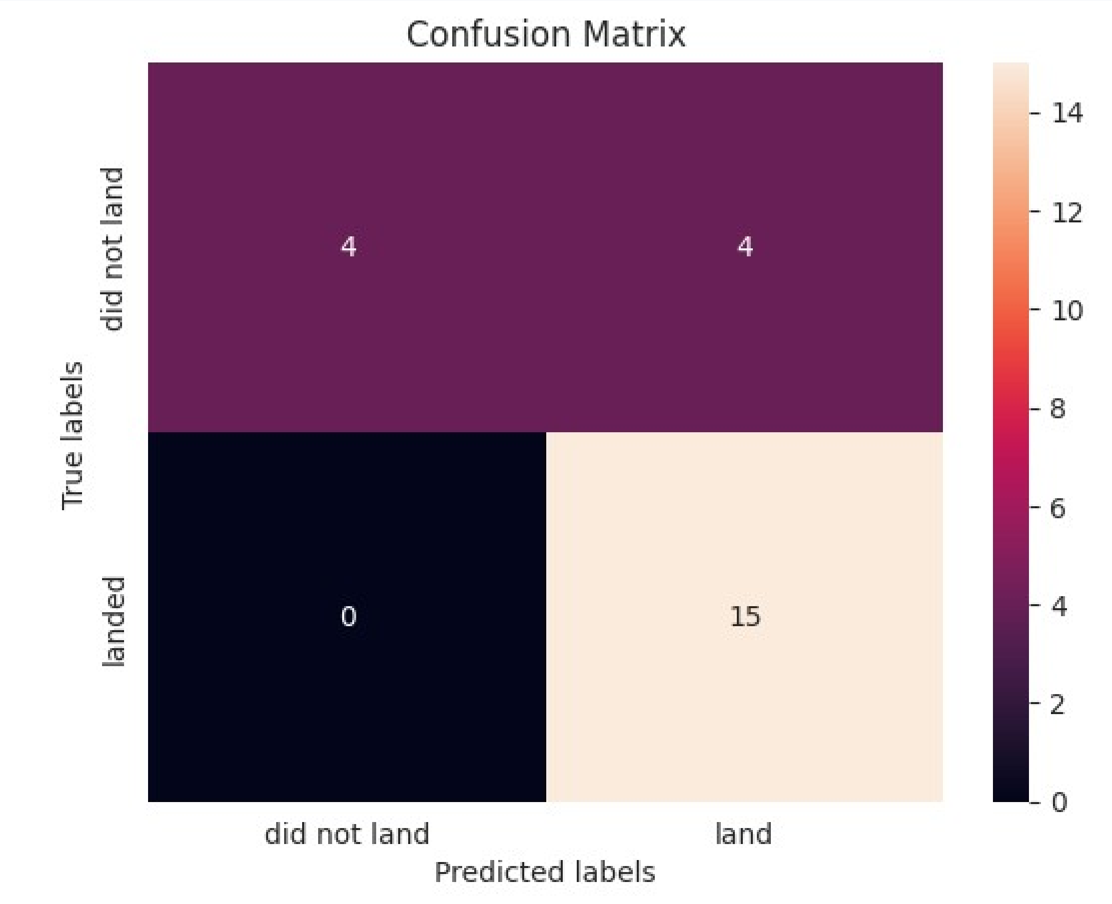

```
Logistic Regression: 0.7826
Support Vector Machine: 0.8261
Decision Tree: 0.6522
K-Nearest Neighbors: 0.8261

Best Performing Model: Support Vector Machine with Accuracy: 0.8261
```

Confusion Matrix

SVM



KNN



Conclusions

1. Best Performing Model
    - Among the tested models (Logistic Regression, SVM, Decision Tree, and K-Nearest Neighbors), the Support Vector Machine (SVM) model achieved the highest accuracy.
    - This suggests that SVM effectively separates successful and failed landings based on the provided features.
    - Decision Tree performed slightly worse but remains useful due to its interpretability.
2. Feature Importance & Impact on Success. The most important factors influencing launch success include:
    - Payload Mass (kg):
        - Heavier payloads tend to decrease the likelihood of success.
        - However, newer booster versions (e.g., Falcon 9 Block 5) show improved success rates even with heavier payloads.
    - Launch Site:
        - KSC LC-39A and VAFB SLC-4E had higher success rates, making them more reliable sites.
        - CCAFS LC-40 had more failures, possibly due to early-stage testing missions.
    - Orbit Type:
        - Certain orbits, like Low Earth Orbit (LEO) and Geostationary Transfer Orbit (GTO), show higher success rates than others.
3. Model Accuracy Comparison. Bar chart results show that:
    - SVM > Logistic Regression > Decision Tree > KNN in terms of accuracy.
    - K-Nearest Neighbors (KNN) had the lowest accuracy, likely due to sensitivity to noise and high-dimensional feature space.
    - Decision Tree provides valuable interpretability but may be prone to overfitting.
4. Business Impact & SpaceX Strategy Insights. SpaceX can optimize mission success by:
    - Choosing launch sites strategically (favoring KSC LC-39A for high-stakes missions).
    - Improving booster reusability for heavier payloads to increase cost efficiency.
    - Refining machine learning models with more data (e.g., weather conditions, booster version performance, etc.).

Final Takeaway:

Machine learning effectively predicts Falcon 9 landing success based on launch conditions.

SpaceX can use these models to enhance launch planning, improve reliability, and reduce mission risks.

Appendix

Full source code of predictive analysis is below.

Source code

Thank you!