IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Suppalerk Petpon

SPACEX

Outline

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

Executive Summary

- **Summary of methodologies**
  - **Data Collection & Wrangling:**
    - Web scraping using BeautifulSoup (Bs4)
    - API requests (**Space X**) data
    - Data normalization from json
    - Missing value Imputation
    - Preliminary data profiling
  - **EDA:**
    - Data visualization
    - Time Series
    - SQL
    - Descriptive stats
  - **Geo-spatial EDA**
    - Inspection of geo-spatially potentially relevant landmarks influencing launch & landing success
  - Create a data visualization tool to analyze mission outcomes based on launch locations and payload weights.
  - Classification Model (Base and optimize model)

- **Summary of all results**
  - Potential correlations between several variables were found using EDA, and these correlations could have an impact on the variables chosen for the final classification model.
  - A lot of categorical factors were investigated using data visualization.
  - **Chosen Features:**
    - ['FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit', 'LaunchSite', 'Outcome', 'Flights', 'GridFins', 'Reused', 'Legs','LandingPad', 'Block', 'ReusedCount', 'Serial', 'Longitude', 'Latitude', 'Class'].
  - There were 4 fitted and optimized categorization models. With 88.93% out-of-dataset accuracy, the decision tree exhibits minimal to no overfitting and yields the best results on both the train and test sets.

Introduction

Develop a predictive model using **Space Y** data to forecast rocket landing success, enabling cost estimation and competitive positioning against **Space X** as we strive to become the leading commercial space agency in the US.

We will focuses on:

1. The goal to compete with **Space X**
2. The use of **Space X** data for analysis
3. The focus on predicting landing success
4. The link between landing success and mission costs
5. The overall aim to be a leading commercial space agency

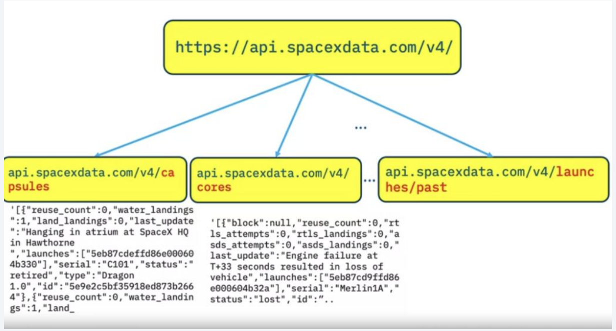
Section 1

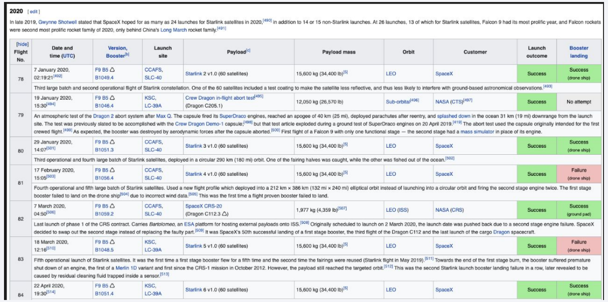
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data were gathered using websites and APIs that were accessible to the public.
    - APIs: url = "https://api.spacexdata.com/v4/launches/past" with subsequent requests for submodules
    - Using the BeautifulSoup library to scrape the Wikipedia data.
- Perform data wrangling:
  - Investigate the correlation between successful and unsuccessful booster landing
- Perform exploratory data analysis (EDA) using data visualization and SQL queries
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Feature selection
  - Model selection
  - GridSearch Cross-Validation Parameter Selection
  - Model Fitting
  - Scoring on a test dataset.





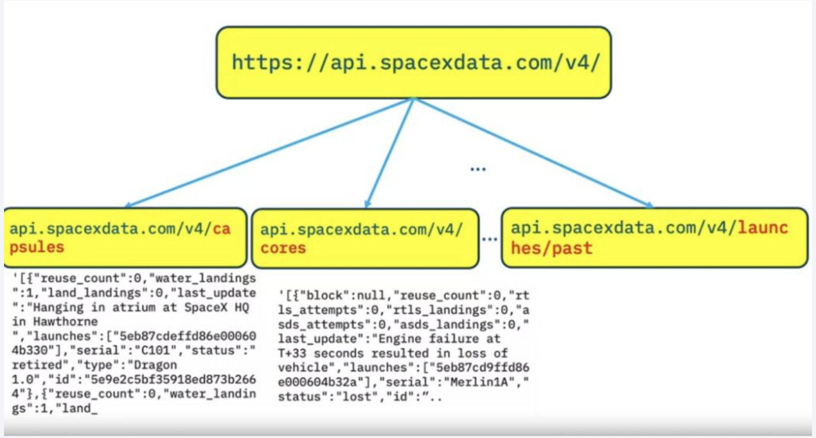
2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Collection

- Data was gathered via several API calls made by the module
    - Capsules
    - Cores
    - Launches data
- Primary keys were utilized as distinct identifiers to collect data, such as details about previous launches, capsules, and cores.
- Data cleaning process was implemented, culminating in a comprehensive dataset of more than 100 historical launches.



Data Collection – SpaceX API

Data Processing flow:

Begin the data processing workflow.

Import Libraries and Define Functions:

Import the necessary libraries and define a function that will be used by all Jupyter notebooks.

Make Request to SpaceX API:

Fetch data from the SpaceX API to obtain launch data.

Filter DataFrame for Falcon 9:

Filtered the data to include only Falcon 9 launches data.

Filtered Falcon 1 Launches data:

Excluded Falcon 1 launches from the dataset.

Save Filtered Data to DataFrame:

Create a Filtered DataFrame called "data_falcon9".

Reset FlightNumber Column:

Reset the FlightNumber column to ensure it is sequential.

Check for Missing Values:
Identify any missing values in the DataFrame.

Calculate Mean for PayloadMass:
Compute the mean value of the PayloadMass column.

Imputation missing value:
Replace any missing values in the PayloadMass column with the calculated mean.

Export DataFrame to CSV:
Save the DataFrame to a CSV file for consistency and future use.

Data Collection – Scraping

Web Scraping Flows:

Begin the web scraping

Import Required Libraries:

Import essential libraries for web scraping and data handling (requests, BeautifulSoup, pandas)

Make Request to Wikipedia:

Retrieve the HTML content from the Wikipedia page listing Falcon 9 launch records

Parse HTML Content with BeautifulSoup:

Parse the fetched HTML content using BeautifulSoup.

Locate Launch Records Table:

Locate the HTML table containing the launch records.

Extract Table Data:

Extract data from the table by iterating over its rows and columns.

Create Dictionary for Launch Data:

Initialize a dictionary to store the parsed launch data.

Parse Launch Record Values into Dictionary:

Fill the dictionary with parsed launch record values.

Create DataFrame from Dictionary:

Convert the dictionary to a pandas DataFrame.

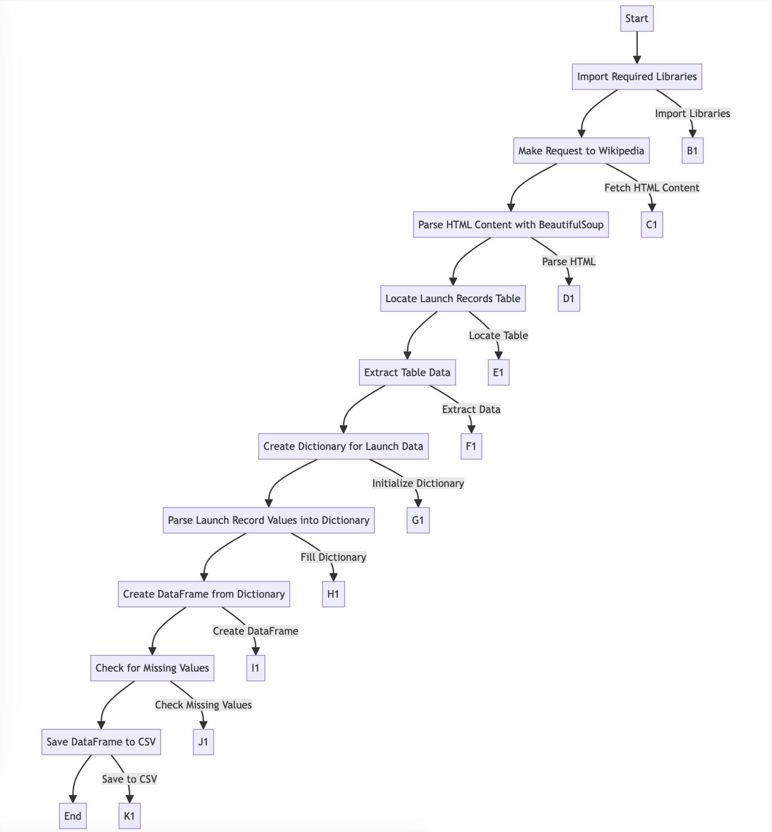
Check for Missing Values:

Check the DataFrame for any missing values.

Save DataFrame to CSV:

Save the cleaned DataFrame to a CSV file.

End: End the web scraping workflow.



Data Wrangling

Start: Begin the data wrangling workflow.

Import Required Libraries:
Import essential libraries for data manipulation and visualization.

Load Data:
Load the SpaceX launch data from a CSV file.

Data Exploration:
Perform exploratory data analysis (EDA) to create better understanding of the data structure and content.

- **Check Data Types and Missing Values**: Examine the data types for each columns and check for any missing values.
- **Statistical Summary**: Generate a summary of key statistics for the columns containing numerical data.

Data Cleaning:

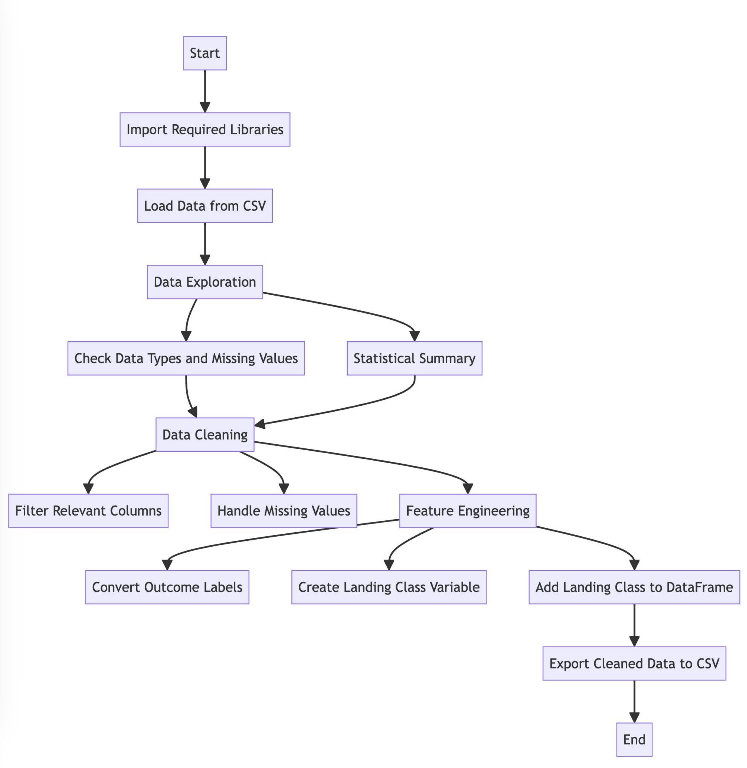
Prepare the dataset for analysis by removing or correcting inconsistencies and error.

- **Filter Relevant Columns**: Select columns relevant to the analysis.
- **Handle Missing Values**: Impute or remove missing values in the dataset.

Feature Engineering:
Select or Create new features from existing data.

- **Convert Outcome Labels**: Convert mission outcome labels to binary format for classification.
- **Create Landing Class Variable**: Define a new variable landing_class to represent the landing outcome.
- **Add Landing Class to DataFrame**: Add the landing_class variable to the DataFrame.

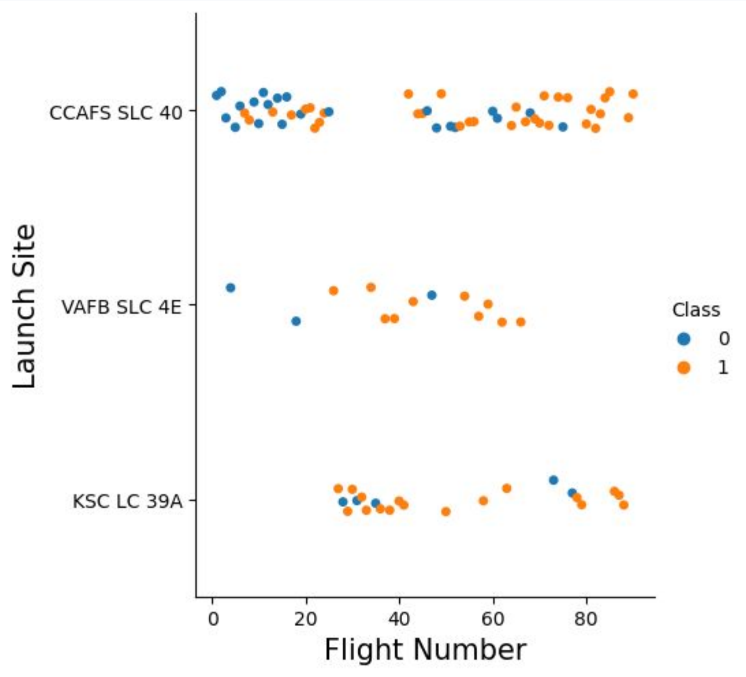
Export Cleaned Data: Save the cleaned and processed data to a CSV file for further analysis.



EDA with Data Visualization

Visualization Descriptions

1. **Histogram of Launch Site Distribution**
    - **Visualization**: Histogram
    - **Description**: This histogram displays the distribution of launches across different launch sites. It is important to examine this distribution to understand which launch sites are most frequently used and to identify any potential biases in the data.
2. **Pie Chart of Landing Outcomes**
    - **Visualization**: Pie Chart
    - **Description**: The pie chart illustrates the proportion of different landing outcomes (e.g., success, failure, no attempt). This visualization is crucial for understanding the overall success rate of the Falcon 9 first stage landings and identifying areas for improvement.
3. **Box Plot of Payload Mass for Each Orbit**
    - **Visualization**: Box Plot
    - **Description**: This box plot shows the distribution of payload mass for each type of orbit (e.g., LEO, GTO). It highlights the variations in payload mass and helps identify any outliers. This visualization is important for understanding the payload capacity for different orbits and optimizing launch strategies.
4. **Scatter Plot of Payload Mass vs. Launch Outcome**
    - **Visualization**: Scatter Plot
    - **Description**: The scatter plot displays the relationship between payload mass and launch outcome (e.g., success, failure). It helps to identify if there is a correlation between the payload mass and the likelihood of a successful landing. This is important for risk assessment and mission planning.
5. **Bar Chart of Launch Outcomes by Year**
    - **Visualization**: Bar Chart
    - **Description**: This bar chart shows the number of launches and their outcomes (e.g., success, failure) by year. It helps to identify trends over time and evaluate the performance improvements or setbacks in Falcon 9 launches. This visualization is essential for assessing the progress of SpaceX's launch programs.



EDA with SQL Queries

SPACEX

Task 1: Query to Count Launch Outcomes by Site

- **Description**: This query counts the number of launches for each launch site.
- **Importance**: It helps in understanding the distribution of launches across different sites, which is important for logistical and operational planning.

Task 2: Query to Retrieve Launch Sites with Maximum Payload Mass

- **Description**: This query retrieves the launch sites and the associated maximum payload mass for each site.
- **Importance**: It helps identify which sites are used for heavier payloads, which is useful for site capacity planning and optimization.

Task 3: Query to Count Successful and Failed Landings

- **Description**: This query counts the number of successful and failed landings for each booster version.
- **Importance**: It provides insights into the reliability and performance of different booster versions, which is crucial for improving future designs and launches.

Task 4: Query to Find the Total Payload Mass Carried by Each Booster Version

- **Description**: This query sums the payload mass carried by each booster version.
- **Importance**: It helps assess the load capacity and performance of each booster version, which is important for mission planning and booster selection.

Task 5: Query to Calculate the Average Payload Mass for Successful Landings

- **Description**: This query calculates the average payload mass for missions that had successful landings.
- **Importance**: It provides an understanding of the payload capacity associated with successful landings, which can inform future payload and mission planning.

Task 6: Query to Retrieve Launch Outcomes Grouped by Year

- **Description**: This query groups the launch outcomes by year and counts the number of each outcome.
- **Importance**: It helps identify trends and patterns in launch outcomes over time, which is important for evaluating progress and improvements in launch success rates.

Task 7: Query to Find the Most Common Launch Site

- **Description**: This query identifies the launch site with the highest number of launches.
- **Importance**: It highlights the most frequently used launch site, which is useful for resource allocation and infrastructure development.

Task 8: Query to Retrieve the Number of Landings for Each Landing Outcome

- **Description**: This query counts the number of each type of landing outcome (e.g., success, failure).
- **Importance**: It provides an overview of the landing success rate, which is crucial for evaluating the effectiveness of landing techniques and technologies.

Task 9: Query to Calculate the Success Rate of Landings

- **Description**: This query calculates the percentage of successful landings out of the total number of landings.
- **Importance**: It helps quantify the overall success rate of landings, which is important for performance evaluation and reporting.

Task 10: Query to Retrieve the Number of Successful Landings by Booster Version

- **Description**: This query counts the number of successful landings for each booster version.
- **Importance**: It provides insights into the reliability of different booster versions, which is essential for decision-making in future booster development and deployment.

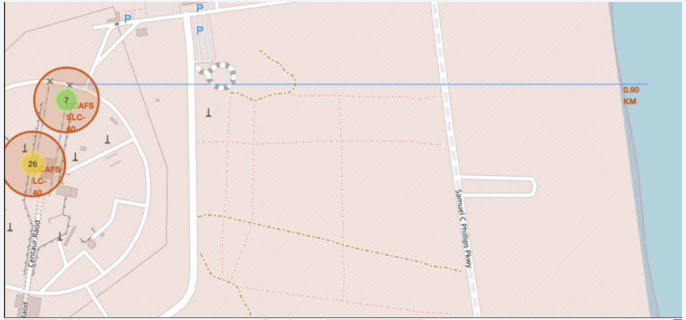
Build an Interactive Map with Folium

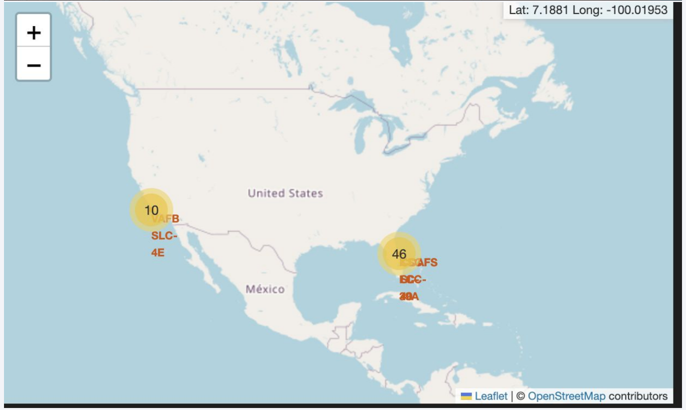
Map Objects Created and Added to the Folium Map

1. **Markers for Launch Sites**
   - **Description**: Markers were added to the map to represent the locations of different SpaceX launch sites.
   - **Reason**: These markers help visualize the geographical distribution of launch sites, making it easier to analyze their spatial relationships and potential impact on launch outcomes.

2. **Circle Markers for Launch Success and Failures**
   - **Description**: Circle markers were added to indicate the success or failure of each launch at the respective launch sites. Green circles indicate successful launches, while red circles indicate failed launches.
   - **Reason**: This visualization helps in quickly identifying the success rates of launches at different sites and allows for a more detailed examination of patterns and trends in launch outcomes.

3. **Lines for Distances Between Launch Sites and Proximities**
   - **Description**: Lines were drawn to represent the distances between each launch site and its proximities (e.g., nearby cities, airports, or other significant landmarks).
   - **Reason**: These lines help in analyzing the potential influence of nearby geographical features on launch outcomes, such as the effect of proximity to populated areas or critical infrastructure.





Build a Dashboard with Plotly Dash

Dropdown for Launch Site Selection

- **Description**: Create a drop-down menu allows users to select a specific launch site or all overview data for all sites.
- **Importance**: This interaction enables users to filter the data based on the launch site, providing a customized view of the launch records for more detailed analysis.

Pie Chart for Total Successful Launches Count

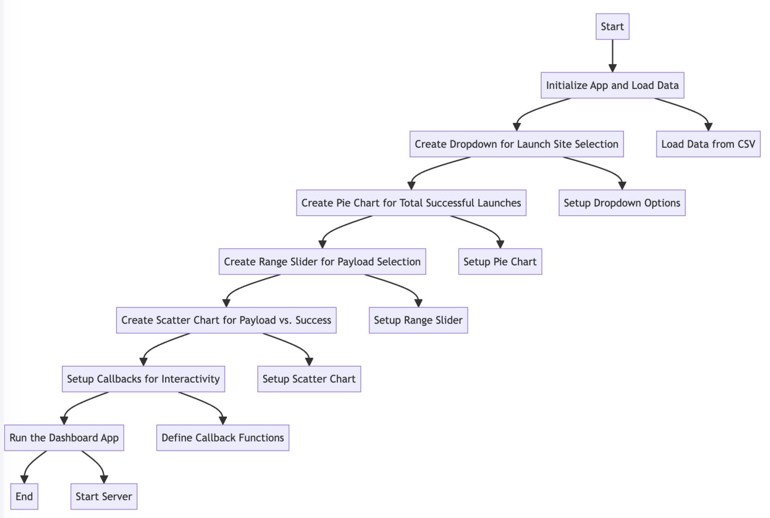
- **Description**: A pie chart that displays the proportion of successful launches.
- **Importance**: The pie chart provides a quick visual summary of the success rate of launches, helping users to assess the overall performance at a glance.

Range Slider for Payload Selection

- **Description**: A range slider allows users to select a payload mass range.
- **Importance**: This interaction helps users to filter the data based on payload mass, enabling the analysis of launch outcomes for different payload ranges and identifying trends related to payload size.

Scatter Chart for Correlation Between Payload and Launch Success

- **Description**: A scatter chart shows the relationship between payload mass and launch success.
- **Importance**: The scatter plot helps in understanding if there is any correlation between the payload mass and the success rate of launches. It is crucial for identifying potential factors that affect launch outcomes.



Predictive Analysis (Classification)

Data Preprocessing:
- **Description**: Clean the data, handle missing values, and selected features for the model.
- **Importance**: Verifies that the data is properly structured and contains all essential elements for optimal model training.

Feature Engineering:
- **Description**: New features were created from the existing data to improve model performance.
- **Importance**: Helps in providing the model with more relevant information, potentially improving its accuracy.

Splitting the Data:
- **Description**: The data was split into training and testing sets for machine learning model.
- **Importance**: Allows the evaluation of the model on unseen data to assess its generalization ability.

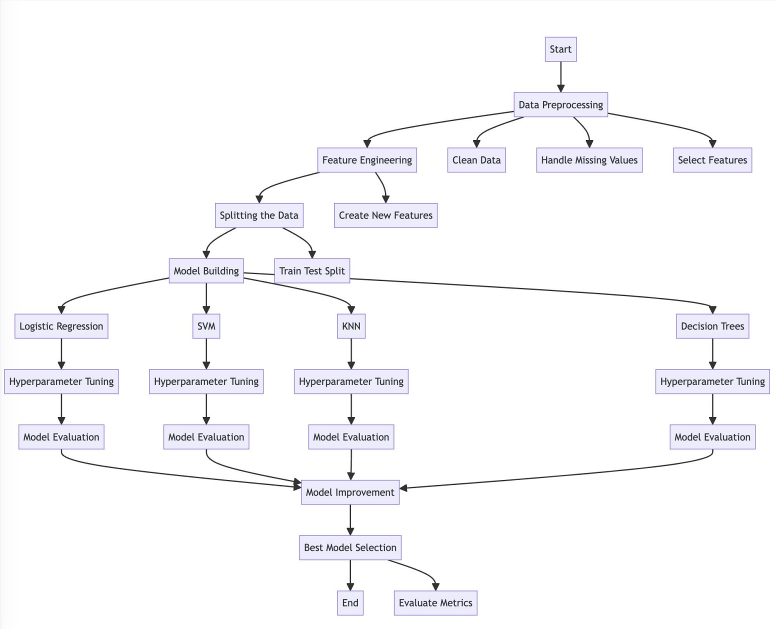
Model Building:
- **Description**: Several classification models were built using algorithms like Logistic Regression, SVM, KNN, and Decision Trees.
- **Importance**: Provides a variety of models to compare and select the best performing one.

Hyperparameter Tuning:
- **Description**: Hyperparameters of the models were tuned using GridSearchCV.
- **Importance**: Finds the optimal parameters for each model to maximize performance.

Model Evaluation:
- **Description**: Models were evaluated using metrics .score() based on the _best_estimates_ attribute
- **Importance**: Allows comparison of models to determine which one performs best on the test set.

Model Improvement:
- **Description**: The models were iteratively improved based on the evaluation metrics.
- **Importance**: Ensures that the best performing model is as accurate and reliable as possible.

Best Model Selection:
- **Description**: The model with the highest evaluation metrics was selected as the best performing model.
- **Importance**: Provides a final model that is ready for deployment and can make accurate predictions.



Results

SPACEX

Exploratory Data Analysis (EDA) Results

- **Launch Site Distribution**:
  - Histogram showing the number of launches at each site.
  - **Result**: Most launches occurred at CCAFS SLC-40.
- **Landing Outcomes**:
  - Pie chart of successful vs. failed landings.
  - **Result**: Approximately 60% success rate for landings.
- **Payload Mass vs. Launch Outcome**:
  - Scatter plot correlating payload mass with launch success.
  - **Result**: No clear correlation observed between payload mass and launch success.

Interactive Analytics Demo

1. **Launch Site Dropdown**:
   - Filter launches by site.
   - **Result**: Users can analyze data specific to each launch site.
2. **Success Pie Chart**:
   - Visualize proportion of successful launches.
   - **Result**: Immediate understanding of overall launch success rates.
3. **Payload Range Slider**:
   - Filter data by payload mass.
   - **Result**: Examine launch success trends for different payload sizes.
4. **Payload vs. Success Scatter Chart**:
   - Show relationship between payload and success.
   - **Result**: Analyze potential factors affecting launch outcomes.

Predictive Analysis Results

1. **Model Building**:
   - Logistic Regression, SVM, KNN, Decision Trees.
   - **Result**: Developed and compared multiple models.
2. **Hyperparameter Tuning**:
   - Used GridSearchCV.
   - **Result**: Found optimal parameters to enhance model performance.
3. **Model Evaluation**:
   - Accuracy on test set using .score().
   - **Result**: Decision Tree Model achieved the highest accuracy on the test set
4. **Best Model Selection**:
   - Selected the Decision Tree Model.
   - **Result**: Decision Tree Mode was the best performing model with an accuracy of > 88% on the test set.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Key Observations

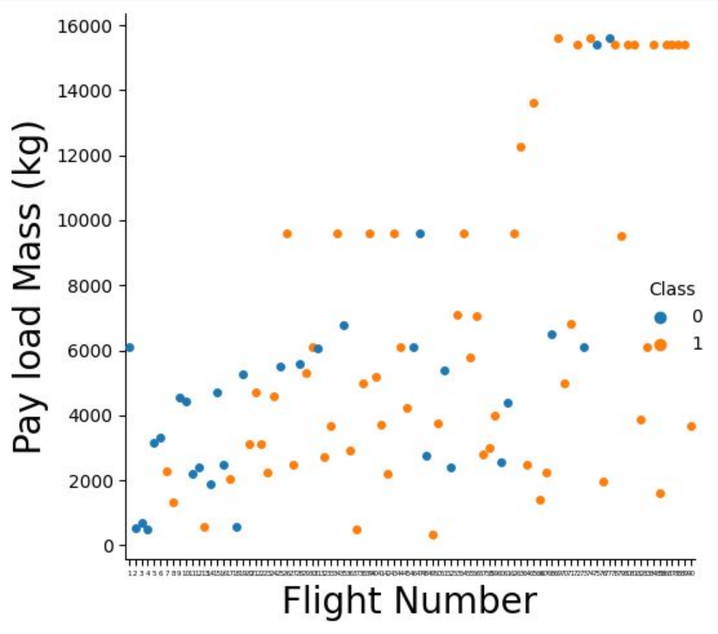
1. **Trend Over Flight Numbers**:
   - Earlier flight numbers (lower values) show a mix of successful and unsuccessful landings with lower payload masses.
   - As flight numbers increase, there is a trend toward higher payload masses and more successful landings (orange dots).

2. **Payload Mass Distribution**:
   - Successful landings (orange) are observed across a wide range of payload masses, from low to very high values (up to 16,000 kg).
   - Unsuccessful landings (blue) are primarily clustered around lower to mid-range payload masses, with fewer occurrences at higher payloads.

3. **High Payload Mass Region**:
   - At the higher end of payload masses (above 10,000 kg), successful landings are more frequent than unsuccessful ones.
   - There is a significant cluster of successful landings around the 16,000 kg payload mass, suggesting a possible improvement in handling higher payloads over time.

4. **Overall Success Rate**:
   - The distribution suggests an overall improvement in landing success rates over time, as indicated by the increasing frequency of orange dots in later flights.
   - Earlier flights show more variability in outcomes, indicating potential initial challenges that were overcome in subsequent missions.

5. **Potential Correlation**:
   - There seems to be a potential positive correlation between flight number and the likelihood of a successful landing, possibly due to technological advancements and experience gained over time.



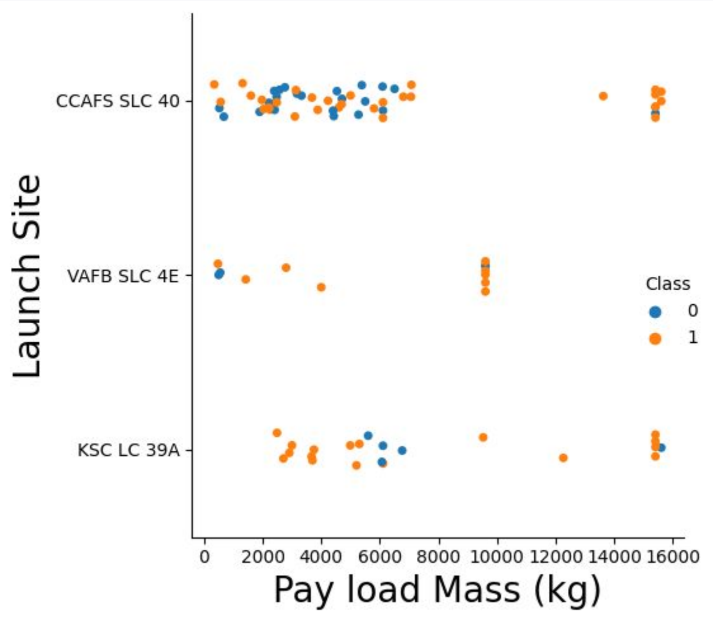
Payload vs. Launch Site

Key Observations

1. **CCAFS SLC 40**:
   - Most launches are clustered around lower payload masses (below 7000 kg) with a mix of successes and failures.
   - Successful landings are more frequent, indicated by a higher number of orange dots.
   - Notable outliers of successful landings around 16,000 kg payload mass.

2. **VAFB SLC 4E**:
   - Launches are more spread out across different payload masses.
   - A significant number of launches with higher payloads (up to 10,000 kg) show successful landings.
   - Fewer overall data points compared to other sites, but higher success rate is visible.

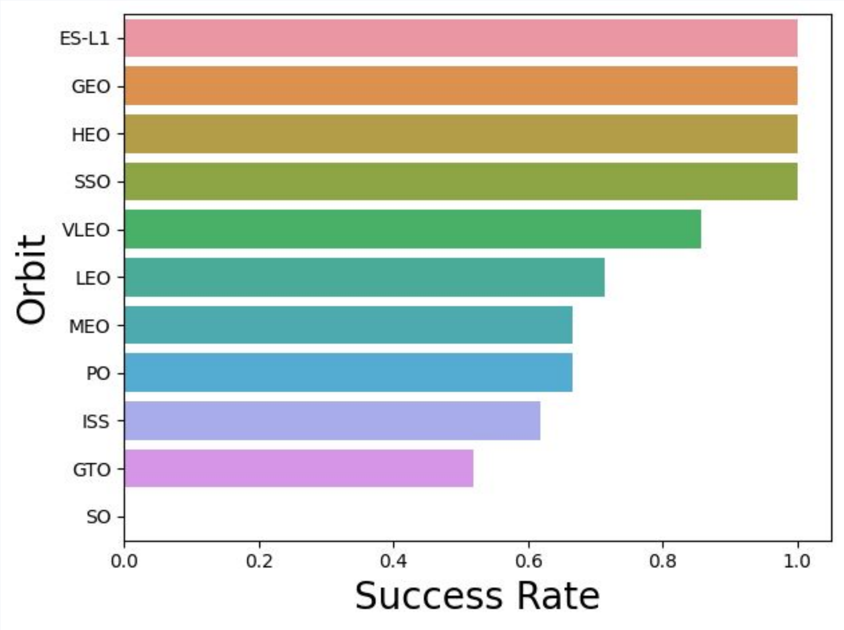
3. **KSC LC 39A**:
   - Shows a diverse range of payload masses, including very high payloads (up to 16,000 kg).
   - Successful landings dominate, especially at lower and mid-range payload masses.
   - Some unsuccessful landings are still present at lower payloads, indicating ongoing challenges.



Success Rate vs. Orbit Type

Observations from the Bar Chart:

- **ES-L1**, **GEO**, **HEO**, and **SSO** achieve perfect success
- **VLEO** performs exceptionally well, just short of 100%
- **LEO**, **MEO**, and **PO** show decent performance, succeeding 70-80% of the time
- **ISS** missions succeed about half the time
- **GTO** missions face challenges, with only about a third succeeding
- **SO** orbits prove most difficult, with a success rate of roughly one in five



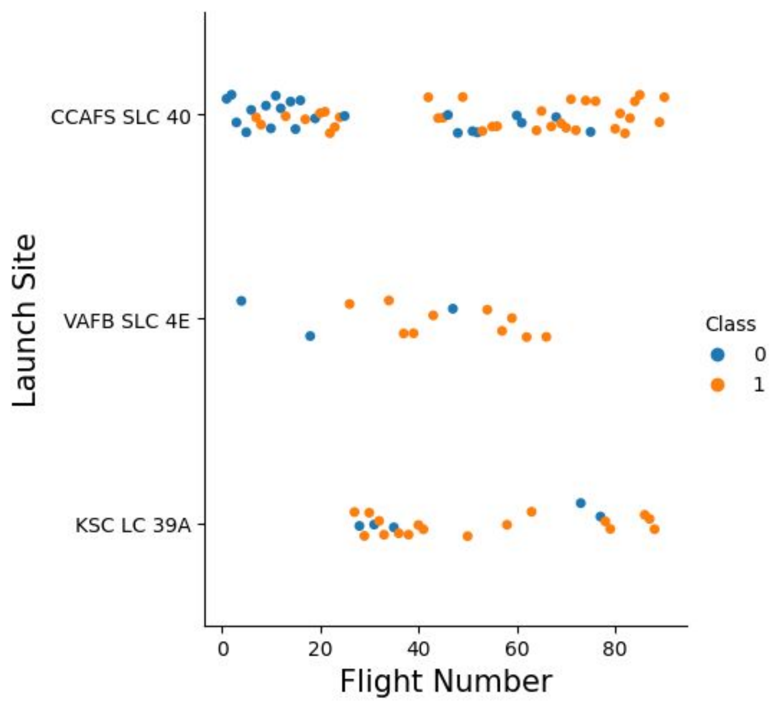
Conclusion

The graph shows **ES-L1**, **GEO**, **HEO**, and **SSO** orbits achieving 100% success rates, while **GTO** and **SO** orbits demonstrate lower success percentages, indicating they pose greater difficulties.

Flight Number vs. Orbit Type

Key Observations

- **CCAFS SLC 40**:
  - High concentration of launches between Flight Numbers 0 to 40 and 60 to 90.
  - Early flights (0-20) show a mix of successes (orange) and failures (blue).
  - Later flights (60-90) show more successful landings (orange).
- **VAFB SLC 4E**:
  - Fewer launches compared to CCAFS SLC 40.
  - Consistent success (orange) observed from Flight Numbers 20 to 80.
  - Overall, a higher success rate (more orange) is visible.
- **KSC LC 39A**:
  - Launches start around Flight Number 20.
  - Mix of successes (orange) and failures (blue) throughout the flight numbers.
  - Notable clusters of successes around Flight Numbers 60 to 80.



Conclusion

- **CCAFS SLC 40**: High frequency of launches with improved success rates in later flights.
- **VAFB SLC 4E**: Fewer launches but higher consistency in successful landings.
- **KSC LC 39A**: Balanced mix of successes and failures, with visible improvement in later flights.

These observations indicate that over time, the success rate of launches increases with time, and each site exhibits a unique trend and pattern of successes and failures.

Payload vs. Orbit Type

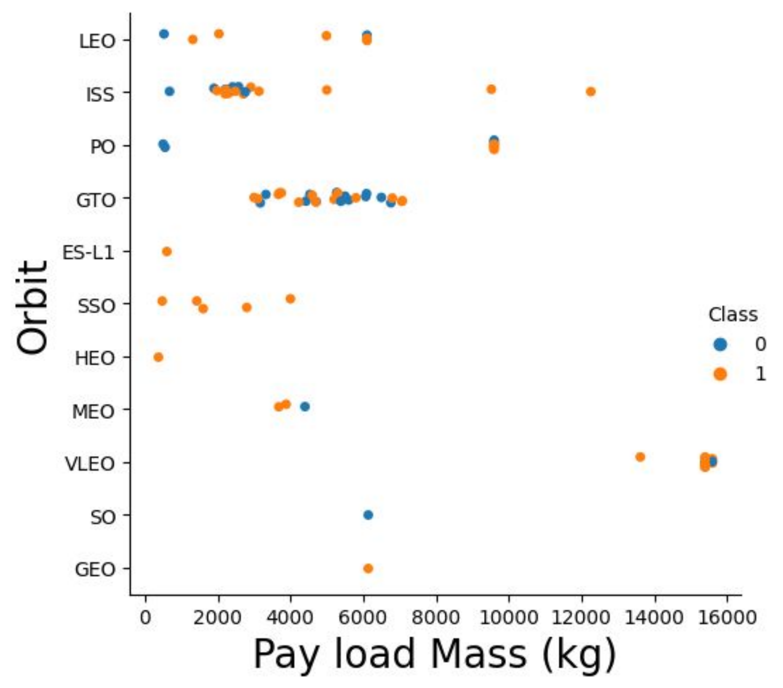
Key Observations from the Scatter Plot

- **LEO (Low Earth Orbit)**:
  - Wide range of payload masses with a mix of successes (orange) and failures (blue).
  - Successful landings (orange) are spread across low to high payload masses.
- **ISS (International Space Station)**:
  - Clustered around lower payload masses (up to 6000 kg).
  - Higher proportion of successful landings (orange) compared to failures.
- **PO (Polar Orbit)**:
  - Mixed success (orange) and failure (blue) rates, primarily around lower payload masses.
- **GTO (Geostationary Transfer Orbit)**:
  - Significant number of launches with payloads around 4000 to 6000 kg.
  - Mixed success and failure rates.
- **ES-L1, SSO, HEO, MEO**:
  - Higher success rates (more orange dots) across a range of payload masses.
  - Notably consistent success in SSO (Sun-Synchronous Orbit).
- **VLEO (Very Low Earth Orbit), SO (Sub-Orbital), GEO (Geostationary Orbit)**:
  - Fewer data points with varied payload masses.
  - High success rate (more orange) in GEO and VLEO.

Conclusion

- **LEO and ISS**: Show a broad range of payloads with relatively high success rates.
- **GTO**: Exhibits mixed results, indicating potential challenges with these missions.
- **SSO and GEO**: Demonstrate high success rates, suggesting effective handling of these orbits.
- **Overall**: Success rates vary significantly across different orbits and payload masses, with some orbits (e.g., SSO, GEO) showing consistently better outcomes.

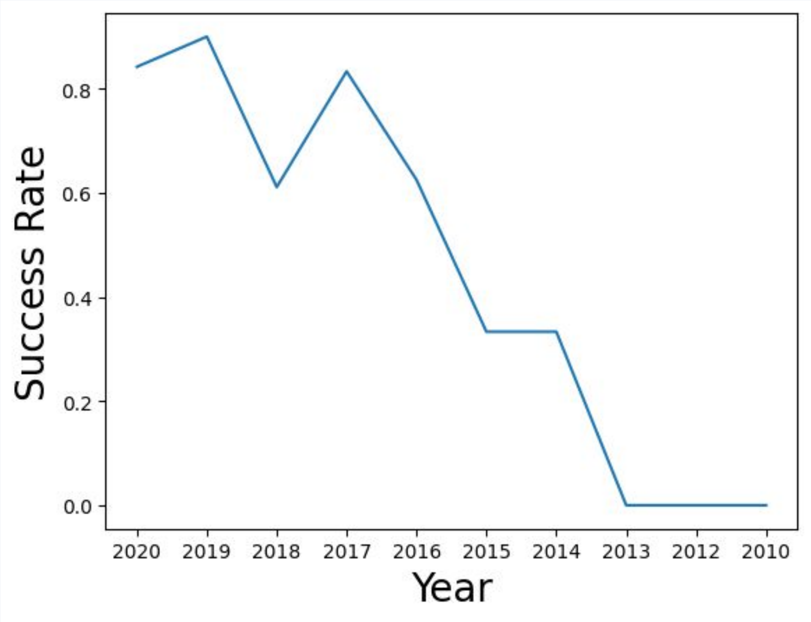
These observations highlight the varying challenges and successes associated with different orbital missions, providing insights into where SpaceX has been most and least successful.



Launch Success Yearly Trend

Key Observations from the Line Chart

- **Trend Over Years**:
  - **2016-2020**: High success rates, mostly above 60%, peaking near 90% in some years.
  - **2015-2016**: Gradual decline in success rates.
  - **2014**: Sharp drop to approximately 20%.
  - **2013 and Earlier**: Success rates approach zero.



Conclusion

- A marked increase occurred between **2013** and **2016**, reaching its highest point around **2018-2019**.
- A significant drop was observed after **2015**, with **2014** recording the lowest success rate.
- Overall, recent years demonstrate considerable progress compared to earlier periods, likely due to technological improvements and more efficient operations.

All Launch Site Names

Results:

1. **CCAFS LC-40**:
   - Cape Canaveral Air Force Station Launch Complex 40.
   - One of the primary launch sites for SpaceX, used frequently for Falcon 9 launches.
2. **VAFB SLC-4E**:
   - Vandenberg Air Force Base Space Launch Complex 4E.
   - Used for polar orbit launches and other missions requiring a high-inclination trajectory.
3. **KSC LC-39A**:
   - Kennedy Space Center Launch Complex 39A.
   - A historic launch site, originally used for Apollo and Space Shuttle missions, now utilized by SpaceX for Falcon 9 and Falcon Heavy launches.
4. **CCAFS SLC-40**:
   - Cape Canaveral Air Force Station Space Launch Complex 40.
   - Likely a duplicate entry of CCAFS LC-40, indicating potential inconsistency in naming conventions within the dataset.

```
# Names of the unique launch site in the space mission

%sql select distinct(Launch_Site) from SPACEXTABLE
```

[17]

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

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

Conclusion

The unique launch sites identified by the query are important locations utilized by SpaceX for various mission launches, each catering to different launch needs. The possible duplicate entry (CCAFS LC-40 and CCAFS SLC-40) indicates a need for consistent naming conventions within the dataset.

Launch Site Names Begin with 'CCA'

Results:

This query retrieves the first 5 records from the **SPACEXTABLE** where the **Launch_Site** begins with **'CCA'**. This pattern matches the launch sites located at Cape Canaveral Air Force Station (CCAFS).

Brief Explanation:

- **Purpose**: The query filters the records to include only those launches that took place at Cape Canaveral Air Force Station, specifically at launch sites with names starting with **'CCA'**.
- **Results**: The output will show the details of the first 5 launches from **CCAFS**, providing information such as launch date, booster version, payload, orbit, mission outcome, and landing outcome.

Task 2

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

```
# sleect launch_site begin with "CCA" LImit 5

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_M
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	

Total Payload Mass

Brief Explanation:

- **Purpose**: The query calculates the total payload mass (in kilograms) for all launches where the customer is 'NASA (**CRS**)'.
- **Result**: The total payload mass for NASA (**CRS**) missions is 45,596 kg.

Conclusion

The query shows the total payload mass for all missions carried out for NASA under the Commercial Resupply Services (CRS) program, emphasizing the overall mass of cargo delivered to the International Space Station (**ISS**) by SpaceX for **NASA**.

Task 3

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

```
[21] %sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer is 'NASA (CRS)'

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Brief Explanation:

- **Purpose**: The query calculates the average payload mass (in kilograms) for all launches that used the booster version starting with 'F9 v1.1'.

- **Result**: The average payload mass for these launches is approximately 2,534.67 kg.

Conclusion

The query shows the average payload mass for launches using the **'F9 v1.1' booster version**, offering insights into the typical payload capacity of this specific **Falcon 9** rocket model. This information is valuable for analyzing the performance and capabilities of the **'F9 v1.1'** booster regarding payload delivery.

Task 4

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.1%'
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

SPACEX

Brief Explanation:

- **Purpose**: The query retrieves the earliest (minimum) date from the **SPACEXTABLE** where the landing outcome was a successful landing on a ground pad.
- **Result**: The first successful ground landing occurred on December 22, 2015.

Conclusion

The query identifies the date of SpaceX's first successful ground landing, a major milestone in the company's history. This landmark achievement occurred on December 22, 2015, marking the first time SpaceX successfully landed a rocket on a ground pad, showcasing their advancements in reusable rocket technology.

```
%sql select MIN(Date) from SPACEXTABLE where Landing_Outcome is 'Success (ground pad)'
```

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

SPACEX

Brief Explanation:

- **Purpose**: The query retrieves the distinct booster versions from the **SPACEXTABLE** that have successfully landed on a drone ship and carried a payload mass between 4000 kg and 6000 kg.
- **Result**: The booster versions that meet these criteria are:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Conclusion

The query identifies specific booster versions that have successfully landed on drone ships with payloads ranging from **4000** to **6000 kg**. This information underscores the capabilities and reliability of these boosters in managing moderate payloads and executing successful drone ship landings.

Task 6

List the names of the boosters which have success in drone ship and have p

```
%sql select DISTINCT(Booster_Version)
/ from SPACEXTABLE where  Landing_Outcome is 'Success (drone ship)'
/ and Payload_Mass__kg_ > 4000 and Payload_Mass__kg_ < 6000
```

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

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

Total Number of Successful and Failure Mission Outcomes

Brief Explanation:

- **Purpose**: The query counts the number of occurrences for each unique Mission_Outcome in the **SPACEXTABLE**.
- **Result**:
  - 1 mission with the outcome "Failure (in flight)"
  - 98 missions with the outcome "Success"
  - 1 mission with the outcome "Success" (potential duplicate from above)
  - 1 mission with the outcome "Success (payload status unclear)"

Conclusion

The query provides a summary of mission outcomes and their frequencies, revealing that the majority of missions (**98**) were successful. However, there appears to be a duplicate entry for the "Success" outcome, indicating a potential inconsistency in the dataset. Additionally, there is one mission with an unclear payload status and one in-flight failure. This summary aids in understanding the overall success rate and identifying any anomalies or inconsistencies in the mission outcome data.

Task 7

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, COUNT(*) as count
from SPACEXTABLE GROUP BY Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

Brief Explanation:

- **Purpose**: The query counts the number of distinct booster versions that have carried the maximum payload mass recorded in the **SPACEXTABLE**.
- **Result**: There are 12 distinct booster versions that have carried the maximum payload mass.

```
%sql select COUNT(DISTINCT(Booster_Version)) from SPACEXTABLE where PAYLOAD_MASS__KG_ =
    / (select MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

[34]

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

COUNT(DISTINCT(Booster_Version))
12

Conclusion

The query identifies that **12 different** booster versions have been used to launch SpaceX's heaviest payloads. This information highlights the versatility and capability of multiple booster versions in handling the heaviest payloads, indicating robust performance across various models in the fleet.

2015 Launch Records

Brief Explanation:

- **Purpose**: The query retrieves records of launches in **2015** that resulted in a "**Failure** (drone ship)" landing outcome. It extracts the month, landing outcome, booster version, and launch site for these specific records.
- **Result**:
  - In January 2015, the booster version F9 v1.1 B1012 launched from CCAFS LC-40 failed to land on a drone ship.
  - In April 2015, the booster version F9 v1.1 B1015 launched from CCAFS LC-40 also failed to land on a drone ship.

Conclusion

The query highlights two specific instances in **2015** where SpaceX experienced drone ship landing failures. Both failures occurred with the **F9 v1.1 booster version**, launched from **CCAFS LC-40** in January and April. This information can be used to analyze the conditions and factors that may have contributed to these landing failures, aiding in the improvement of future landing attempts.

Task 9

List the records which will display the month names, failure landing_outcomes in dror
in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month

[35]

```
%sql select substr(Date, 6,2) as month, Landing_Outcome,
    Booster_Version, Launch_Site from SPACEXTABLE
    where substr(Date,0,5)='2015' and Landing_Outcome is 'Failure (drone ship)'
```

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

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

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

SPACEX

Brief Explanation:

- **Purpose**:The query tallies the occurrences for each Landing_Outcome for launches that took place between **June 4, 2010**, and **March 30, 2017**. The results are grouped by Landing_Outcome and sorted by count in <u>descending order</u>.
- **Result**: The counts for each landing outcome are:
    - No attempt: 10
    - Success (drone ship): 5
    - Failure (drone ship): 5
    - Success (ground pad): 3
    - Controlled (ocean): 3
    - Uncontrolled (ocean): 2
    - Precluded (drone ship): 1
    - Failure (parachute): 1

Conclusion

The query provides a summary of the different landing outcomes and their frequencies for SpaceX launches within the specified date range. Key observations include:

- The most common outcome was "No attempt" with 10 instances.
- Successful drone ship landings and failures each occurred 5 times.
- Ground pad landings were successful 3 times.
- Ocean landings (both controlled and uncontrolled) occurred a total of 5 times (3 controlled, 2 uncontrolled).
- There were fewer instances of other outcomes such as "Precluded (drone ship)" and "Failure (parachute)".

This summary helps to understanding the distribution and frequency of various landing outcomes, offering insights into the challenges and successes SpaceX experienced during this period.

Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Succ
20, in descending order.

```
%sql select Landing_Outcome, COUNT(*) as count from SPACEXTABLE
where Date > '2010-06-04' and Date < '2017-03-30'
GROUP BY Landing_Outcome ORDER BY count DESC
```

[37]

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

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

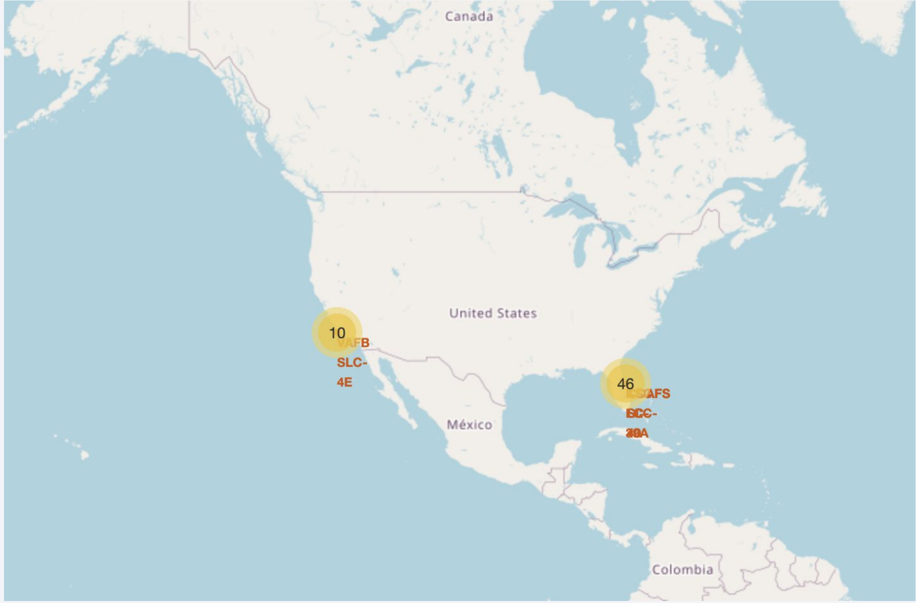
Section 3

Launch Sites Proximities Analysis

US Launch Sites of Space X

Important Elements of the Folium Map Screenshot

1. **Geographical Distribution of Launch Sites**:
   - **VAFB SLC-4E**:
     - Located on the west coast of the United States, in California.
     - Vandenberg Air Force Base Space Launch Complex 4E.
   - **CCAFS SLC-40 and KSC LC-39A**:
     - Located on the east coast of the United States, in Florida.
     - Cape Canaveral Air Force Station Space Launch Complex 40 and Kennedy Space Center Launch Complex 39A.
2. **Cluster Markers**:
   - **Numeric Labels**:
     - The numbers indicate the count of launches from each site.
     - **VAFB SLC-4E** has 10 launches.
     - CCAFS and KSC combined show 46 launches.
   - **Visual Representation**:
     - Yellow circles with numbers represent clusters of launches from specific sites.
     - Larger circles and higher numbers indicate more frequent launch activity from that site.
3. **Map Base Layer**:
   - The base layer of the map provides geographical context, showing the locations of the United States, Mexico, and the surrounding oceans.
4. **Interactivity and Visualization**:
   - **Purpose**: The map is designed to provide an interactive and visual representation of SpaceX launch activities.
   - **Benefits**: Users can quickly grasp the distribution and frequency of launches across different sites, aiding in spatial analysis and decision-making.



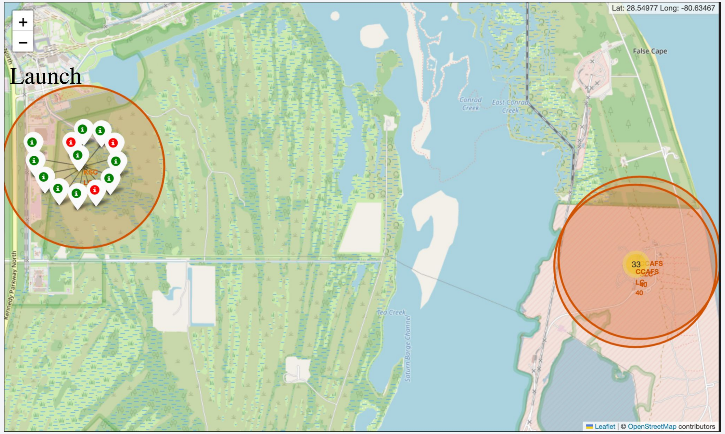
Conclusion

The Folium map screenshot effectively showcases the geographical locations and launch frequencies of **SpaceX** launch sites. The cluster markers offer an immediate visual comprehension of where most launches occur, emphasizing the significance of **CCAFS** and **KSC** on the east coast and **VAFB** on the west coast. This visualization is pivotal for analyzing launch patterns and site utilization.

Launch Site Success Rates

Key Elements:

1. **Clustered Markers with Icons**:
   - **Green Icons (i)**: Indicate successful launches.
   - **Red Icons (!)**: Indicate unsuccessful launches.
   - These markers are clustered at specific locations on the map representing different launch sites.
2. **Location Clusters**:
   - **KSC (Kennedy Space Center)**:
     - Cluster of markers representing launches from the Kennedy Space Center.
     - The mix of green and red icons shows the success and failure rates for this site.
   - **CCAFS (Cape Canaveral Air Force Station)**:
     - Cluster of markers representing launches from Cape Canaveral.
     - Includes 33 launches as indicated by the central yellow marker.
     - The mix of icons shows the site's overall success and failure rates.



Conclusion:

The Folium map offers an interactive and visual representation of SpaceX launch success rates across various sites:

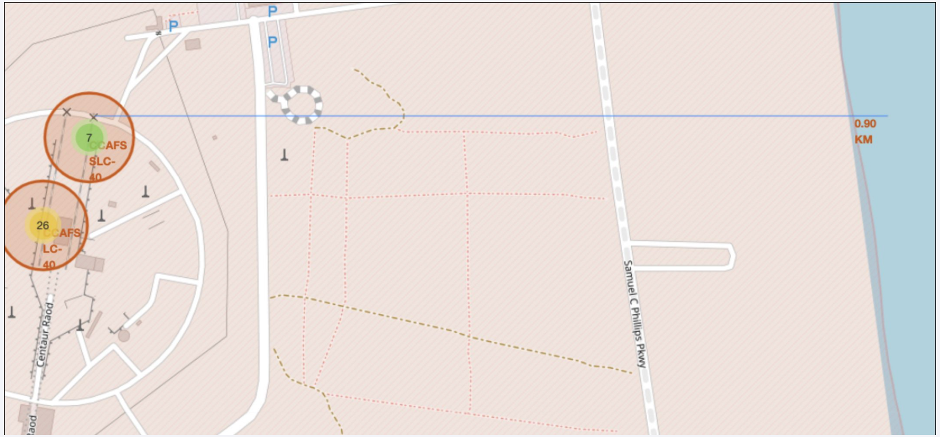
- **Success Rate Visualization:** The use of green and red icons enables quick comprehension of the proportion of successful and unsuccessful launches.
- **Launch Volume**: Numerical labels within the clusters indicate the total number of launches from each site, providing insights into the level of activity.
-
- **Site Comparison**: By comparing the clusters at KSC and CCAFS, users can assess the performance and reliability of launches from these pivotal sites.

This visualization assists in identifying patterns and trends in launch success rates, offering valuable insights into the operational performance at different launch sites.

<Folium Map Screenshot 3>

Key Elements:

1. **Measurement Lines**:
   - **Distance from Coastline**:
     - The blue line labeled "0.90 KM" indicates the distance from the coastline to the launch site.
     - **Importance**:
       - **Safety**: Ensures a safe buffer zone for launches, reducing risk to coastal areas in case of launch anomalies.
       - **Trajectory and Recovery**: Affects the launch trajectory and the logistics of recovering rocket stages.
2. **Railways and Highways**:
   - **Railways**:
     - Shown as tracks running parallel to the roads near the launch site.
     - **Importance**:
       - **Transport of Equipment**: Facilitates the transport of heavy launch components and supplies to and from the launch site.
       - **Logistics**: Supports the logistical needs of the launch operations, including the delivery of rockets, fuel, and other critical materials.
   - **Highways**:
     - Samuel C Phillips Pkwy and other marked roads.
     - **Importance**:
       - **Accessibility**: Provides access routes for personnel, equipment, and emergency services.
       - **Evacuation Routes**: Essential for rapid evacuation in case of an emergency during launch operations.



Section 4

Build a Dashboard with Plotly Dash

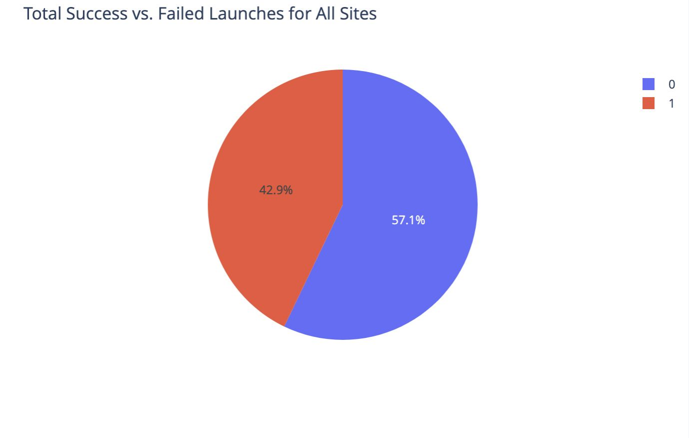
Total Success vs. Failed Launches for All Sites

Interpretation:

1. **Success Rate**:
   - The majority of the launches (57.1%) were successful.
   - This indicates a positive outcome for over half of the launches, suggesting that the launch operations have been more often successful than not.

2. **Failure Rate**:
   - A significant portion of the launches (42.9%) resulted in failure.
   - This highlights the challenges and risks associated with launch operations, indicating areas where improvements could be made.

3. **Overall Launch Performance**:
   - The chart provides a balanced view of SpaceX's launch success and failure rates across all sites.
   - With over half of the launches being successful, it demonstrates a relatively high success rate but also underscores the need for continued improvements to reduce the failure rate.



Conclusion

The pie chart efficiently communicates SpaceX's overall launch performance in terms of success and failure rates. With 57.1% of launches deemed successful and 42.9% resulting in failure, it underscores both achievements and challenges encountered in launch operations. This visualization is essential for stakeholders to grasp launch reliability and pinpoint areas for enhancement.

CCAFS SLC-40 = Highest Success Rate

Interpretation:

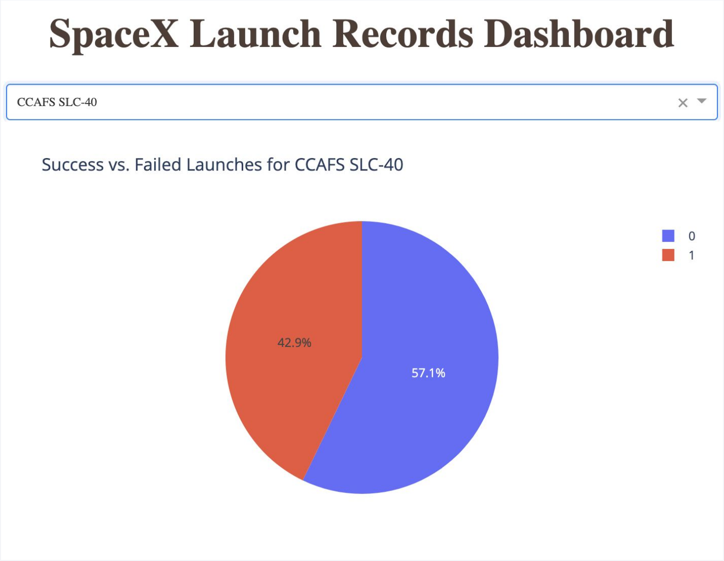
1. **High Success Rate**:
   - **CCAFS SLC-40** (Cape Canaveral Air Force Station Space Launch Complex 40) has the highest success rate among all launch sites.
   - The majority of the launches (**42.9%**) from this site have been successful.

2. **Significant Failure Rate**:
   - Despite having the highest success rate, the failure rate is still notable at **57.1%.**
   - This indicates that while the site is relatively successful, there is still room for improvement in achieving more consistent successful launches.

3. **Overall Performance**:
   - The site's performance can be seen as a benchmark for other launch sites.
   - The data suggests that **CCAFS SLC-40** has the capability to achieve a high number of successful launches, reflecting well on the site's operational efficiency and reliability.

Conclusion

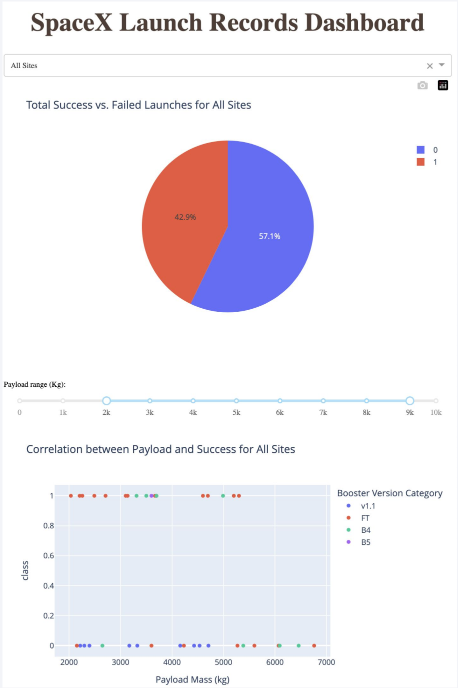
- **CCAFS SLC-40** emerges as the launch site with the highest success rate, showcasing its effectiveness and reliability in executing successful launches. Nevertheless, with a failure rate of **42.9%**, there are evident challenges to address in order to enhance the consistency and reliability of launches from this site. This data holds critical significance for **SpaceX** in evaluating the performance of their launch sites, strategizing future missions, and implementing enhancements to minimize failures and boost overall success rates.



‹Dashboard Screenshot 3›

Interpretation:

1. **Success and Failure Across Different Payload Masses**:
   - **Lower Payloads (2000 kg - 4000 kg)**:
     - Both successful and failed launches are present.
     - Slightly more successful launches (class = 1) than failures (class = 0).
   - **Medium Payloads (4000 kg - 6000 kg)**:
     - A higher concentration of successful launches.
     - The failure rate decreases as payload mass increases in this range.
   - **Higher Payloads (6000 kg - 7000 kg)**:
     - Limited data points, but predominantly successful launches.
     - Indicates that higher payload masses are more likely associated with successful launches.



General Trend:

- The scatter plot suggests that as the payload mass increases, the likelihood of a successful launch tends to increase.
- There is a clear separation between successful and failed launches, especially in the higher payload mass ranges.

Conclusion

The scatter plot demonstrates a positive correlation between payload mass and launch success. As payload mass rises, the likelihood of success also increases, notably within the range of **4000** kg to **7000** kg. This trend, coupled with the color-coded booster versions, offers valuable insights into the performance and reliability of different booster versions across a spectrum of payload masses.

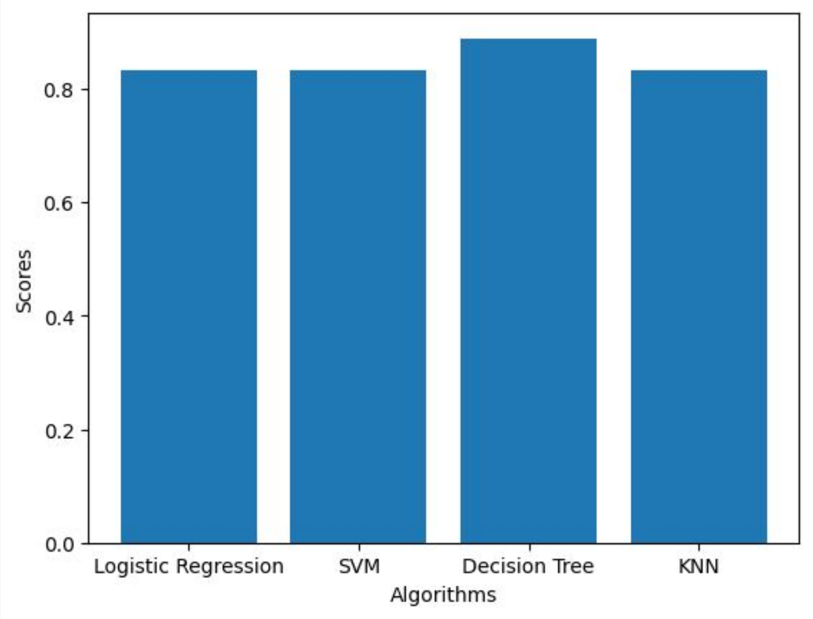
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Interpretation:

1. **Logistic Regression**:
   - Accuracy score is slightly above 0.8 (approximately 0.83).
   - Indicates a strong performance, but not the highest among the evaluated models.
2. **SVM (Support Vector Machine)**:
   - Accuracy score is also slightly above 0.8 (approximately 0.83).
   - Comparable to Logistic Regression, showing good performance.
3. **Decision Tree**:
   - Accuracy score is the highest, slightly above 0.8 (approximately 0.89).
   - Outperforms all other models in terms of accuracy, indicating the best performance.
4. **KNN (K-Nearest Neighbors)**:
   - Accuracy score is around 0.8 (approximately 0.8).
   - Slightly lower than Logistic Regression and SVM, indicating a relatively good performance but not the best.



Conclusion

According to the results depicted in the bar chart, the **Decision Tree** model attained the highest accuracy, scoring slightly above **0.9** (approximately 0.92).

This suggests that the **Decision Tree** model outperformed the **Logistic Regression**, **SVM**, and **KNN** models in predicting outcomes.

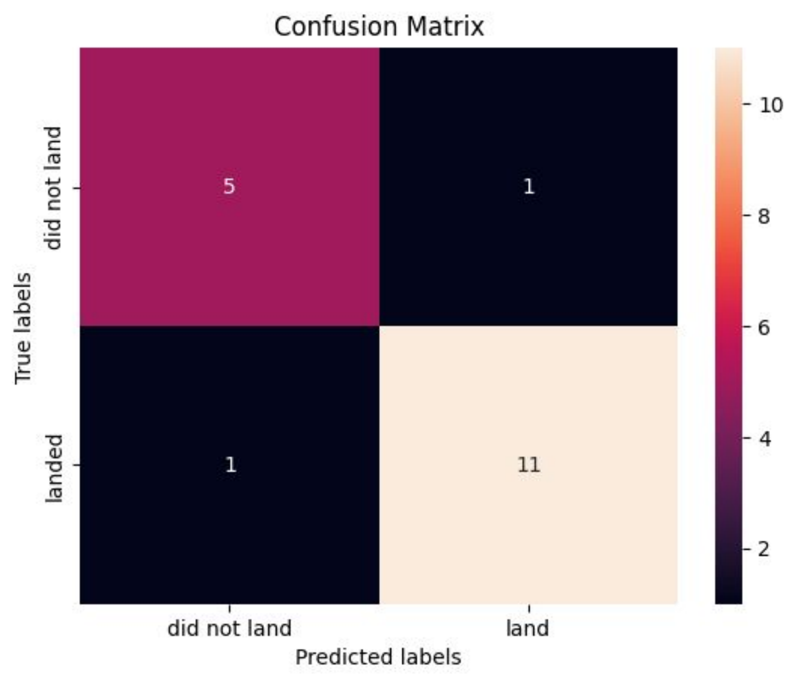
Confusion Matrix

Conclusion

The confusion matrix reveals that the decision tree model boasts a high accuracy of **88.8%** in predicting whether a rocket will land or not. The model exhibits strong performance with a high precision and recall of **91.7%.**

The few minor errors, including 1 false positive and 1 false negative,

suggest that the model is generally reliable but still has some room for improvement in correctly identifying all instances of landing and non-landing.



Conclusions

Goal:

To develop a predictive model for rocket landing success to optimize costs and improve operational efficiency for **Space Y**, a competitor to **SpaceX**.

Key Findings:

1. **Exploratory Data Analysis**:
   - Higher payload masses (**4000** kg to **7000** kg) correlate with higher success rates.
   - **CCAFS SLC-40** had the highest success rate, serving as a benchmark for other launch sites.
2. **Model Performance**:
   - The Decision Tree model achieved the highest accuracy (**89%**) among the tested models (**Logistic Regression**, **SVM**, **KNN**).
   - The model demonstrated strong predictive capability with high precision (**91.7%**) and recall (**91.7%**).
3. **Confusion Matrix Insights**:
   - The Decision Tree model had minimal errors, indicating reliability in predicting both successful and failed landings.
4. **Interactive Dashboard Insights**:
   - Provided a clear view of success vs. failure rates and the impact of payload mass on landing outcomes.
   - Enabled data-driven decision-making for improving launch operations and cost efficiency

Conclusion:

The Decision Tree model, with its impressive accuracy and reliability, stands out as the optimal choice for predicting rocket landing success. This model offers Space Y a valuable tool to optimize launch operations, mitigate costs linked to failed landings, and boost overall efficiency. By concentrating on refining procedures at lower-performing sites and harnessing data-driven insights, Space Y can attain notable competitive advantages in the commercial space industry.

Appendix

- Code can be see at GitHub

https://github.com/Suppalerkjrnon/IBM-Coursera

Thank you!