IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jinesh Gohil
December 01, 2023

Outline

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

Executive Summary

Purpose & Scope:

- **Launch Price Determination:** Understand and analyze the factors that contribute to the cost of a rocket launch, especially focusing on SpaceX's Falcon 9. Gather data on historical launch costs and mission parameters to establish a pricing model for Space Y's rocket launches.

- **First Stage Reusability Prediction:** Develop a machine learning model to predict whether SpaceX will successfully reuse the first stage of the Falcon 9 rocket for each launch. Use public information and historical data to train the model, considering mission parameters, success/failure of first stage landings, and any other relevant features.

- **Competitive Analysis:** Enable Space Y to compete with SpaceX by understanding and incorporating pricing strategies and reusability into their business model. Provide insights that help Space Y make strategic decisions in the commercial space industry.

Executive Summary

Methodologies:

- **Objective Definition and Requirement Analysis:** We initiated the project by clearly defining its objectives and thoroughly analyzing all associated requirements.

- **Data Requirement Identification:** Identified the types and quantity of data needed, exploring reliable sources for collection.

- **Data Collection:** Utilized the SpaceX API for comprehensive Falcon 9 launch data. Extracted additional information from Wikipedia, covering launch sites, orbits, customers, booster landings, and mission outcomes.

- **Data Collection:** Utilized the SpaceX API for comprehensive Falcon 9 launch data. Extracted additional information from Wikipedia, covering launch sites, orbits, customers, booster landings, and mission outcomes.

- **Data Wrangling:** Performed data wrangling, addressing missing values, handling diverse data types, and standardizing/normalizing data to prepare it for in-depth analysis.

Executive Summary

- **Model Training:** Trained models based on insights from EDA, testing various types of models and comparing their accuracy to select the optimal model.

- **Interactive Dashboard Preparation:** Developed a simple and effective interactive dashboard for executives to gain real-time insights efficiently.

Introduction

Project Background and Context:

- In the rapidly evolving landscape of commercial space exploration, Space Y is poised to enter the market as a competitor to industry leader SpaceX. Founded by Billionaire industrialist Allon Musk, Space Y aspires to make space travel more accessible and cost-effective.

Problems to Address:

- ***Launch Cost Determination:*** The accurate determination of launch costs is paramount for Space Y's competitiveness. Understanding the factors influencing pricing and assessing how SpaceX achieves cost efficiency, especially through first stage reusability, is crucial.

- ***First Stage Reusability Prediction:*** Predicting the success of first stage landings is a key challenge. Establishing a reliable method, such as leveraging machine learning, to predict whether SpaceX will reuse the first stage for a given launch will greatly aid in pricing strategy and resource allocation.

Introduction

- ***Competitive Analysis:*** Understanding the strengths and weaknesses of SpaceX's pricing model and reusability practices is essential for Space Y's strategic positioning. Identifying areas where Space Y can differentiate itself in terms of pricing and launch reliability is a critical aspect of this analysis and making data driven decisions.

- ***Data-Driven Decision-Making:*** Developing a robust data science framework to gather, analyze, and interpret relevant information from various sources is vital. The goal is to empower Space Y with actionable insights, enabling informed decision-making in a dynamic and competitive space industry.

Section 1

Methodology

Methodology

Data Collection Methodology:

- As we delved into understanding our data requirements, we simultaneously initiated the data gathering process from reliable sources, including the SpaceX API and Wikipedia pages. Leveraging API calls and web scraping techniques, we utilized Python libraries such as the "requests" and "beautifulsoup" libraries for seamless data collection.

Data Wrangling:

- Following data collection, we undertook comprehensive data wrangling. This phase involved addressing missing values, outliers, and random data mistakes. Employing the "pandas" library, a high-level data structure tool, we created dummy variables for categorical data within the dataset. To enhance data integrity, we replaced missing values with feature means and converted categorical features into numerical representations.

Explanatory Data Analysis:

- The exploratory data analysis phase was both enjoyable and insightful. Utilizing SQL, pandas syntax, and Python libraries such as Matplotlib, Plotly, and Seaborn, we visualized various data features. We explored correlations, utilized the "group by" function for improved data clarity, and employed SQL for mission outcomes and average payload calculations.

Methodology

Interactive Visual Analytics:

- To ensure transparency and engagement, we developed an interactive dashboard. Stakeholders can seamlessly interact with the data, assess performance, and make informed decisions. Additionally, we utilized "Folium" to plot launch sites on a map, facilitating an examination of nearby elements like railways, highways, cities, and coastal lines.

Prediction and Modeling:

- Having garnered insights from various stages of the process, we proceeded to model development. We trained and tested the models for accuracy, employing preprocessing techniques such as the "StandardScaler" from the popular machine learning library "scikit-learn" in Python. Furthermore, we utilized grid search to optimize hyperparameters, ensuring the selection of the most accurate model.

Data Collection – SpaceX API

Initiating SpaceX REST Calls:

- Commenced the data collection process by initiating SpaceX REST calls to retrieve essential information about rocket launches.

Utilizing SpaceX API Endpoints:

- Leveraged the SpaceX API's comprehensive endpoints, such as /rockets, to gather detailed data on Falcon 9 launches.

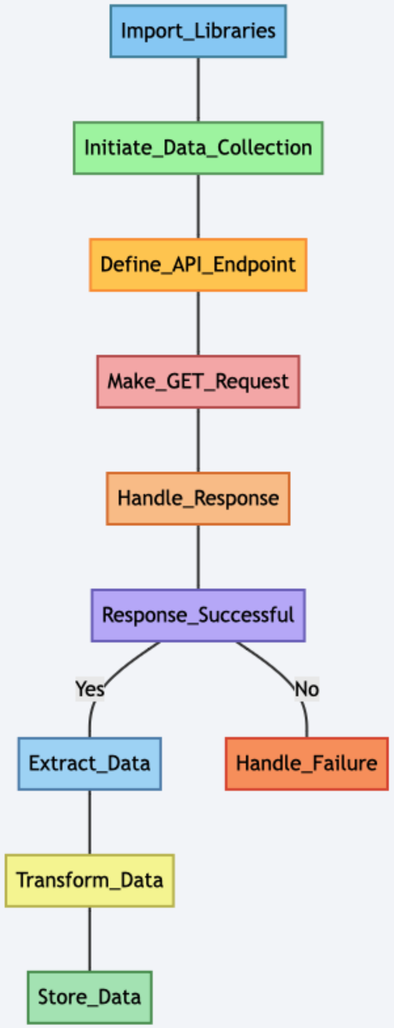
Extracting Launch Data:

- Extracted launch-specific details, including dates, mission parameters, payloads, and outcomes, to build a rich dataset for analysis.

Ensuring Real-time Updates:

- Utilized SpaceX REST calls to ensure real-time updates, enabling the dataset to reflect the latest information on space missions.

GitHub Data Collection – SpaceX API



Data Collection – Scraping

Web Scraping Wikipedia for Additional Information:

- To complement our SpaceX REST data, we implemented web scraping techniques on Wikipedia pages. This allowed us to gather additional details such as launch sites, specific orbits, customers, and mission outcomes.

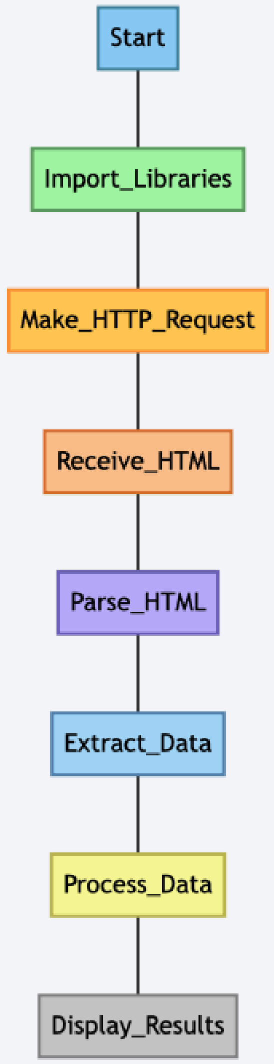
Python Libraries for Streamlined Collection:

- Employing Python libraries, specifically "requests" for HTTP requests and "beautifulsoup" for parsing HTML, we streamlined the web scraping process. This ensured efficient and accurate extraction of supplementary data.
- In the web scraping process, we focused on extracting additional information from Wikipedia pages, specifically targeting tables for structured data.

Scraping Tables with BeautifulSoup:

- Leveraging Python's beautifulsoup library, we employed the find_all method to identify and extract pertinent tables containing valuable details such as launch sites, specific orbits, customers, and mission outcomes.

[GitHub Data Collection - Scraping]



Data Wrangling

Handling Missing Values:

- Identified and addressed missing values in the dataset, employing strategies such as imputation with the mean of the respective feature.

Outlier Management:

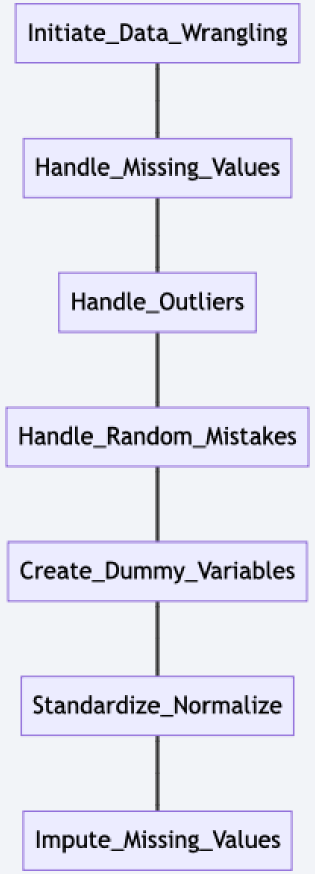
- Dealt with outliers to ensure that anomalous data points did not skew the analysis, employing appropriate methods for detection and correction.

Handling Random Data Mistakes:

- Addressed any random data mistakes that may have been present, ensuring consistency and accuracy in the dataset.

Creating Dummy Variables:

- Employed the "pandas" library to create dummy variables for categorical data, facilitating the inclusion of categorical features in the analysis.



Data Wrangling

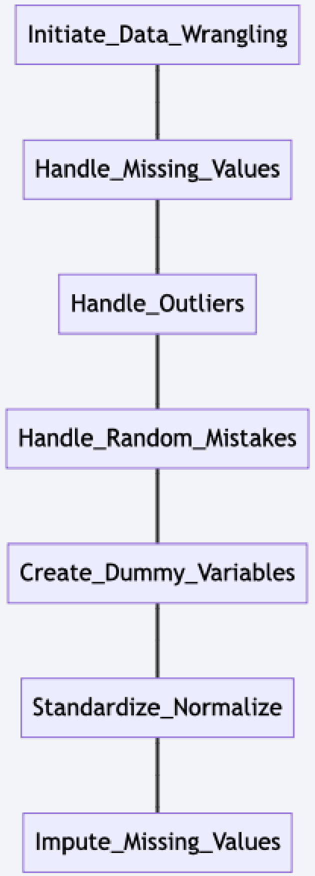
Standardizing/Normalizing Data:

- Ensured uniformity in the dataset by applying standardization or normalization techniques, depending on the nature of the features.

Imputing Missing Values:

- Imputed missing values using the mean of the respective feature, maintaining data completeness and integrity.

[GitHub Data Wrangling]



EDA with Data Visualization

Flight Number vs. Launch Site:

- Purpose: Uncover patterns in the sequencing of Flight Numbers concerning different Launch Sites.
- Insight: Identify if certain Launch Sites exhibit specific trends or anomalies in the ordering of flight missions.

Payload vs. Launch Site:

- Purpose: Investigate the distribution of Payload values across various Launch Sites.
- Insight: Examine if there are variations in payload sizes based on the Launch Site, revealing potential site-specific payload preferences or constraints.

Success Rate of Each Orbit Type:

- Purpose: Demonstrate success rates across different types of orbits.
- Insight: Understand how the success of space missions varies for distinct orbit types, providing insights into the reliability of launches in specific orbital trajectories.

EDA with Data Visualization

Flight Number vs. Orbit Type:

- Purpose: Identify correlations between sequential Flight Numbers and types of orbits.
- Insight: Determine if certain orbit types are consistently associated with particular ranges or sequences of Flight Numbers, indicating potential patterns in mission planning.

Payload vs. Orbit Type:

- **Purpose:** Compare payload sizes for different types of orbits.
- **Insight:** Evaluate how payload sizes vary based on the orbital trajectory, revealing potential payload considerations specific to certain types of space missions.

Launch Success Yearly Trend:

- **Purpose:** Visualize the yearly trend of launch success to discern patterns over time.
- **Insight:** Understand the overall performance trajectory of space launches on an annual basis, identifying trends and variations that may impact decision-making and planning.

GitHub EDA with data visualization

EDA with SQL

- **Distinct Launch Sites:**
  Retrieve distinct Launch Sites from the SPACEXTABLE.

- **Sample Data for Launch Site 'CCAFS LC-40':**
  Retrieve the first 5 records from SPACEXTABLE where the Launch Site is 'CCAFS LC-40'.

- **Total Payload for NASA (CRS) Missions:**
  Calculate the total payload mass for missions with the customer 'NASA (CRS)'.

- **Total Payload for Booster Version 'F9 v1.1':**
  Calculate the total payload mass for missions with the booster version 'F9 v1.1'.

- **Earliest Successful Landing Date:**
  Find the earliest date with a successful landing outcome, specifically 'Success (ground pad)'.

- **Booster Versions for Payload Between 4000 and 6000 KG:**
  Retrieve the unique booster versions for missions where the payload mass is between 4000 and 6000 kilograms.

EDA with SQL

- **Mission Outcome Counts (Success/Failure):**
  Count the number of missions categorized as 'Success' and 'Failure' in the SPACEXTABLE.

- **Booster Version for Maximum Payload:**
  Identify the booster version associated with the maximum payload mass.

- **Failed Landings in 2015 with Drone Ship:**
  Retrieve the date, mission outcome, booster version, launch site, and landing outcome for missions in 2015 with a landing failure specifically on the drone ship.
- **Landing Outcome Counts:**
  Count the number of missions for each landing outcome and order the results by rank in descending order.

GitHub EDA with SQL

Build an Interactive Map with Folium

NASA Johnson Space Center:

- Circle: Represents a 1000-meter radius area in orange.
- Marker: Indicates NASA JSC's location with a text label.
- Purpose: Highlights NASA JSC's position on the map.

Launch Sites:

- Circles: 100-meter radius circles for each site in orange.
- Markers: Text labels at each launch site's coordinates.
- Purpose: Visualizes launch site distribution and provides location info.

Mission Outcomes:

- Marker Cluster: Efficiently groups mission outcome markers.
- Markers: Color-coded (green for success, red for failure).
- Purpose: Manages numerous markers, displaying mission outcomes clearly.

Distance Information:

- Markers: Displayed for coastline, railway, highway, and city points.
- PolyLines: Connect each point to the launch site.
- Purpose: Provides distance info and illustrates spatial relationships.

Build an Interactive Map with Folium

Explanation:

NASA Johnson Space Center Circle and Marker:

- Highlights the specific location of NASA Johnson Space Center, a significant space-related facility.
- The circle serves to delineate the area around the center, and the marker provides a label for easy identification.

Launch Site Circles and Markers:

- Offers a visual representation of the distribution of SpaceX launch sites.
- Each circle denotes the general area of a launch site, and markers provide labels for identification.

Marker Cluster for SpaceX Mission Outcomes:

- Handles the visualization of numerous SpaceX missions efficiently.
- Clusters markers based on proximity, improving map readability.
- Markers with different colors clearly distinguish between successful and unsuccessful missions.

Build an Interactive Map with Folium

Markers and PolyLines for Distance Information:

- Provides additional context by indicating distances from the launch site to various features like the coastline, railway, highway, and city.
- PolyLines visually connect the launch site to each point, enhancing spatial understanding.
- *Overall, these map objects aim to convey spatial relationships, highlight specific locations, and efficiently present information about SpaceX launch sites and mission outcomes.*

GitHub Interactive Maps

Build a Dashboard with Plotly Dash

Plots/Graphs Added:

1. Pie Chart for Success Counts:

- *Purpose:* Displays the distribution of success and failure counts for all launch sites or a specific site.

2. Scatter Chart for Payload vs. Launch Success:

- *Purpose:* Illustrates the correlation between payload mass and launch success for all sites or a selected launch site.

Interactions Added:

1. Launch Site Dropdown:

- *Purpose:* Enables the selection of a specific launch site or displays data for all sites when set to 'ALL'.

2. Payload Range Slider:

- *Purpose:* Allows users to filter data based on a selected payload range, refining the scatter chart's focus.

Build a Dashboard with Plotly Dash

Explanation:

Pie Chart for Success Counts:

- The pie chart provides an overview of success and failure distribution for all launch sites or a specific site. This helps users quickly assess the overall success rate and identify variations among different launch sites.

Scatter Chart for Payload vs. Launch Success:

- This scatter chart explores the relationship between payload mass and launch success. It allows users to analyze whether there is a correlation between payload size and the likelihood of a successful launch. The chart includes additional information, such as the booster version category, for a more comprehensive analysis.

Build a Dashboard with Plotly Dash

Interaction Rationale:

1. Launch Site Dropdown:

- *Purpose:* Users can focus on a specific launch site to understand its individual success and failure distribution. This dropdown provides flexibility in exploring site-specific data.

2. Payload Range Slider:

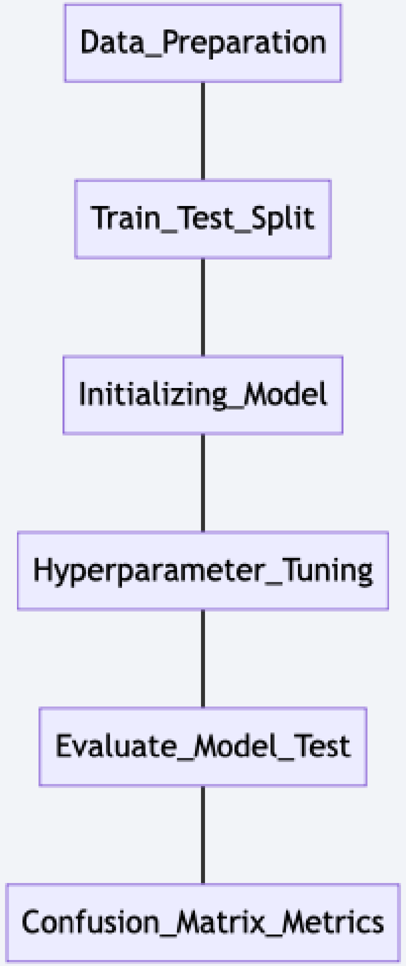
- *Purpose:* The payload range slider empowers users to filter data based on their payload mass of interest. This interaction allows for a more detailed examination of how payload mass correlates with launch success, offering insights into potential patterns or constraints.

Overall Purpose: The dashboard aims to provide a dynamic exploration of SpaceX launch data, allowing users to investigate the success distribution across launch sites and understand the relationship between payload mass and launch success. The chosen plots and interactions facilitate a user-friendly and insightful exploration of key metrics in SpaceX's launch history.

[GitHub DashBoard]

Predictive Analysis (Classification)

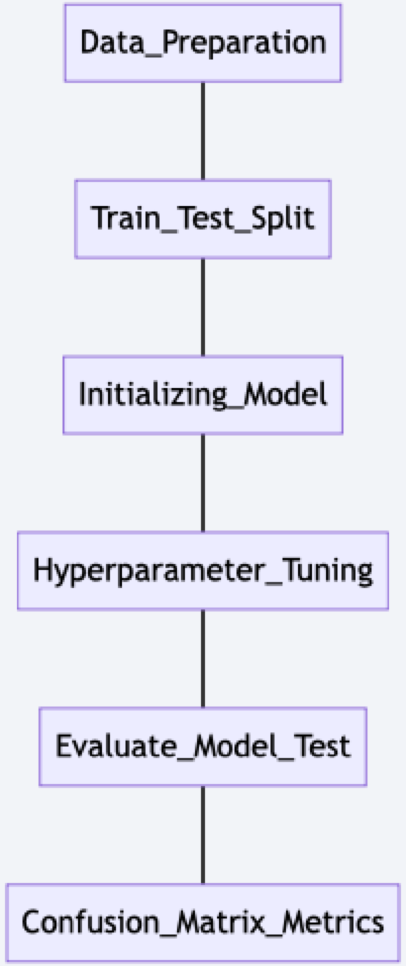
- **Data Preparation:**
  - Extract the target variable and features from the dataset.
  - Format the data appropriately, ensuring compatibility with the chosen models.
- **Train-Test Split:**
  - Split the dataset into training and testing sets for model evaluation.
- **Model Selection:**
  - Choose a classification algorithm (SVM, Decision Tree, k-Nearest Neighbors, Logistic Regression).
- **Hyperparameter Tuning:**
  - Define a set of hyperparameters for the chosen model.
  - Utilize grid search or another hyperparameter tuning technique (e.g., random search) with cross-validation to find the best hyperparameters.



Predictive Analysis (Classification)

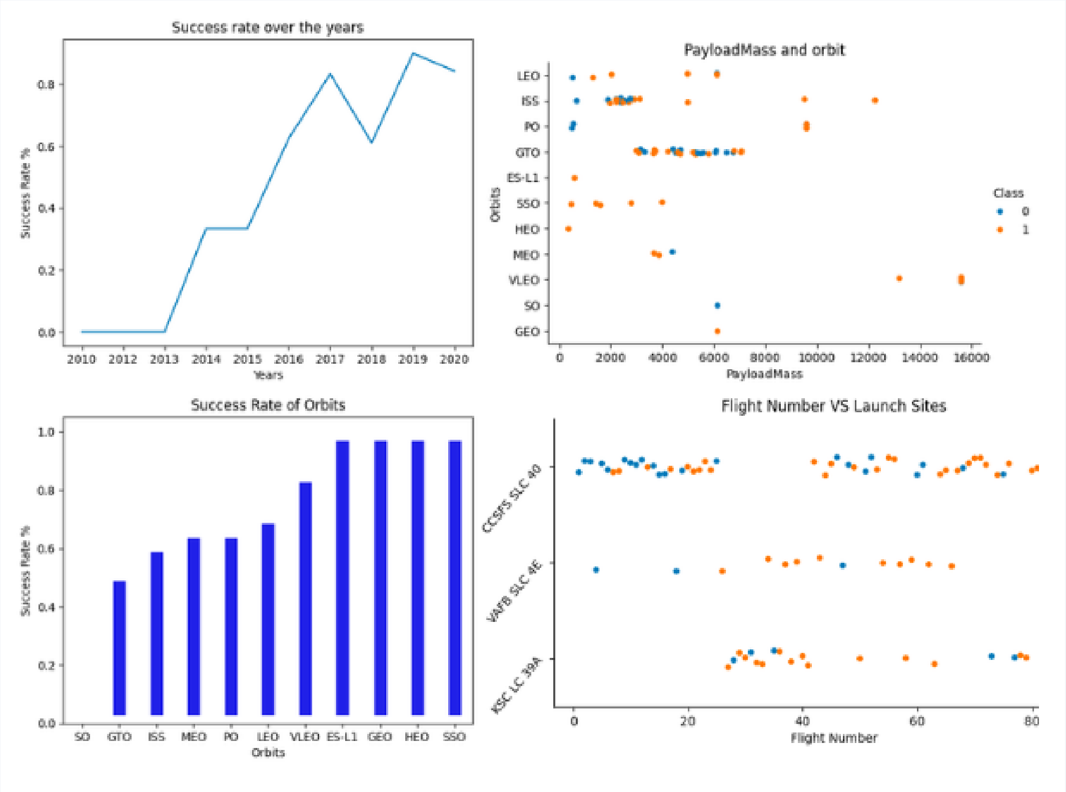
- **Evaluate Model on Test Data:**
  - Apply the tuned model to the test set to assess its generalization performance.
  - Print the accuracy of the model on the test data.
- **Confusion Matrix and Additional Metrics:**
  - Generate a confusion matrix to visualize the model's performance in terms of true positives, true negatives, false positives, and false negatives.
  - Consider additional classification metrics such as precision, recall, and F1-score for a more detailed assessment.

GitHub Model Development



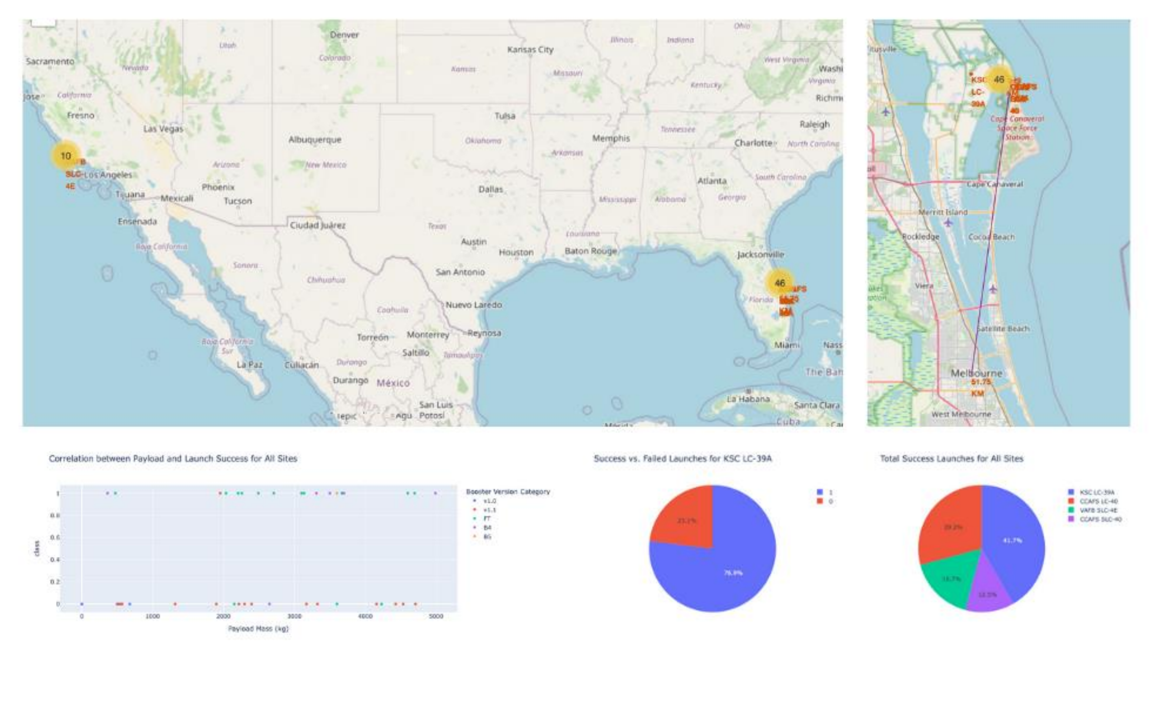
Results

- The EDA combined data visualization and SQL-based analysis, revealing insights into space missions. Visualization highlighted site-specific Flight Number patterns and payload distributions across Launch Sites, uncovering potential preferences. Success rates for orbit types clarified reliability, while correlating Flight Numbers with orbit types suggested mission planning patterns. Payload sizes in relation to orbit types revealed mission-specific considerations. Yearly trends in launch success were visualized. The SQL analysis identified unique Launch Sites, calculated total payload for NASA (CRS) missions, and provided insights into mission performance, booster versions, and landing outcomes. Together, these findings offer a comprehensive foundation for informed decision-making in space mission planning.



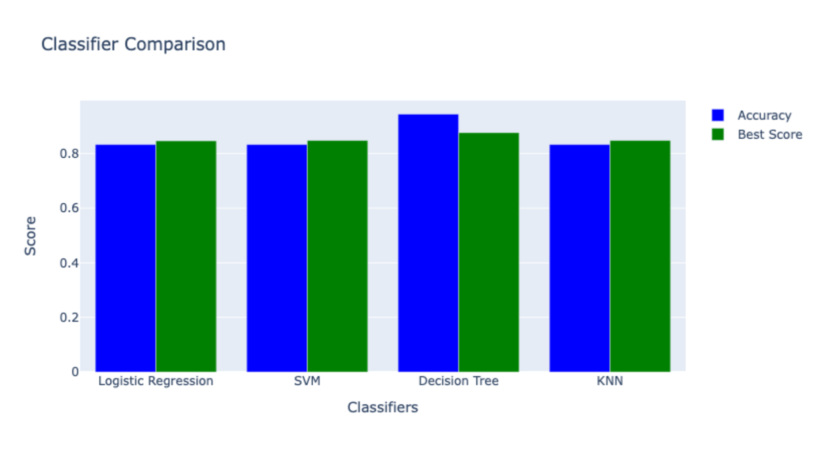
Results

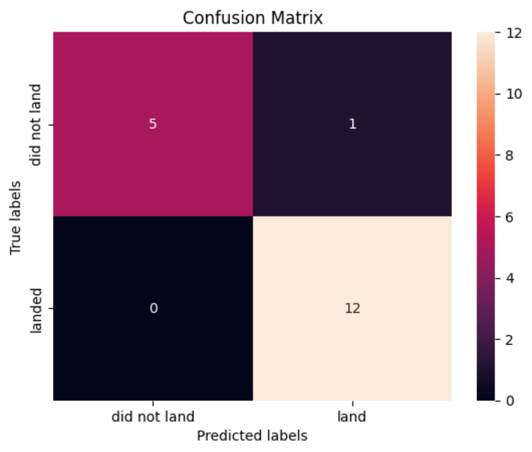
- The Folium map highlights SpaceX launch sites, including the NASA Johnson Space Center, with mission outcomes color-coded and efficiently clustered.
- The Plotly Dash dashboard showcases pie charts for success/failure counts and a scatter chart analyzing payload mass and launch success, with booster version details. User-friendly interactivity is provided through a launch site dropdown and a payload range slider, enabling focused exploration.
- The dashboard aims to offer dynamic insights into SpaceX launch data, providing a quick overview of success rates across sites and analyzing the relationship between payload mass and launch outcomes.



Results

- In our predictive analysis focused on classification, we meticulously prepared the data by extracting target variables and features, ensuring compatibility with the chosen models. After a train-test split, we considered four classification algorithms—Logistic Regression, SVM, Decision Tree, and KNN. Through hyperparameter tuning, the Decision Tree model emerged as the most effective, displaying the highest accuracy among the alternatives.
- Visualized in a bar chart, the Decision Tree's superior accuracy solidified its selection as the optimal model. The Confusion Matrix provided additional insights into the model's performance metrics, highlighting True Negatives, True Positives, False Positives, and False Negatives. This comprehensive assessment, including precision, recall, and F1-score, reinforced the Decision Tree's efficacy in classifying data compared to other models.





Section 2

Insights drawn from EDA

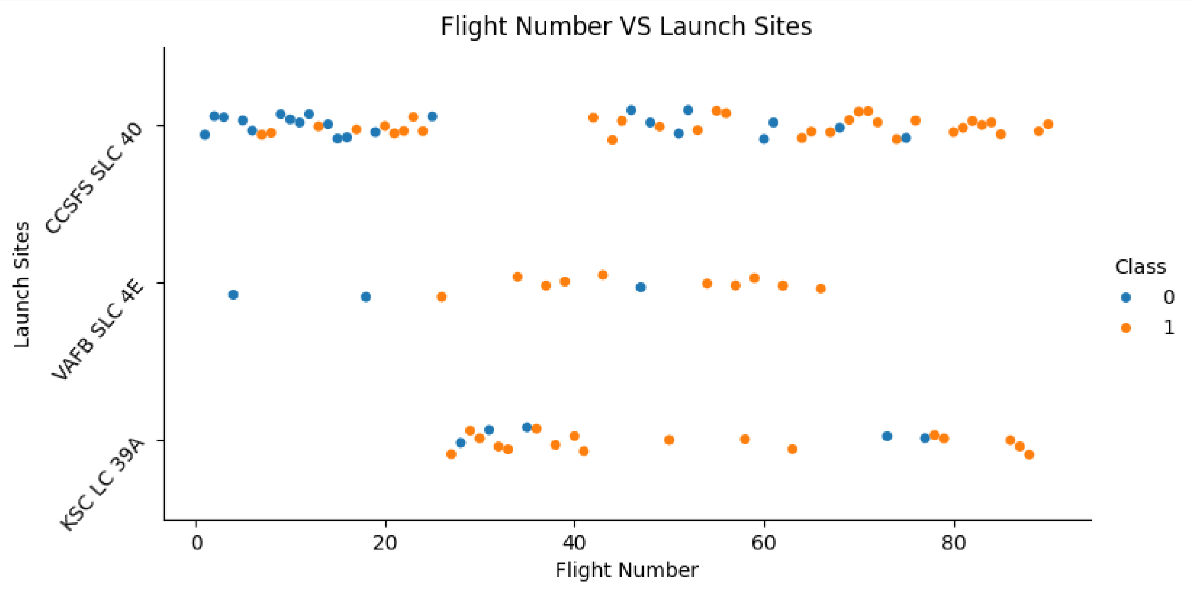
Flight Number vs. Launch Site

Charts Plotted:

- Scatter plot with Flight Number on the x-axis and Launch Site on the y-axis.

Reasoning:

- The plot visually depicts the relationship between flight numbers and launch sites, with each point representing a specific flight number at a launch site. Different classes are color-coded, aiding in understanding class distribution. It helps identify patterns, anomalies, and launch site performance trends across flight numbers and classes. The concise goal is to use visualization for easy comprehension of complex data relationships and trends.



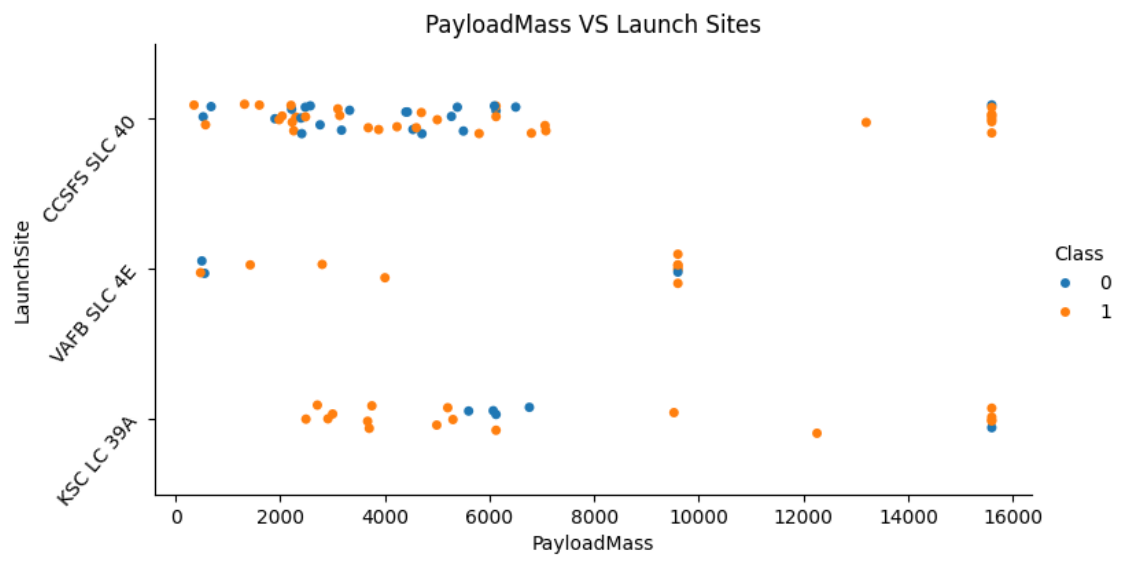
Payload vs. Launch Site

Charts Plotted:

- Scatter plot with Payload on the x-axis and Launch Site on the y-axis.

Reasoning:

- The plot illustrates the relationship between payload mass and launch sites, with each point denoting a specific payload mass at a launch site. Different classes are distinguished by colors. It offers insights into payload mass distribution across launch sites and helps identify patterns or trends. The goal is to efficiently communicate complex information about payload masses, launch sites, and classes. The y-axis label rotation enhances readability for potentially long launch site names.



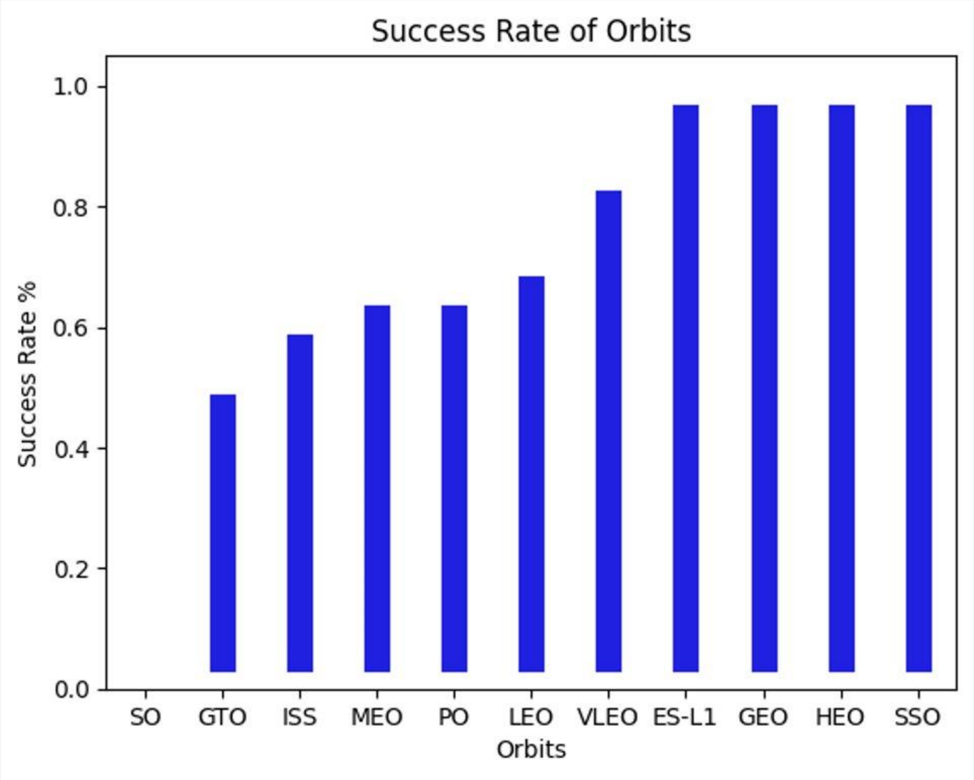
Success Rate vs. Orbit Type

Charts Plotted:

- Bar chart or pie chart representing the success rate of each orbit type.

Reasoning:

- The purpose of this chart is to visually represent the success rate percentage of different orbits. The blue bars indicate the mean success rate for each orbit, providing a clear comparison of success rates across different orbital paths. The chart aims to communicate the varying success rates in a concise manner.



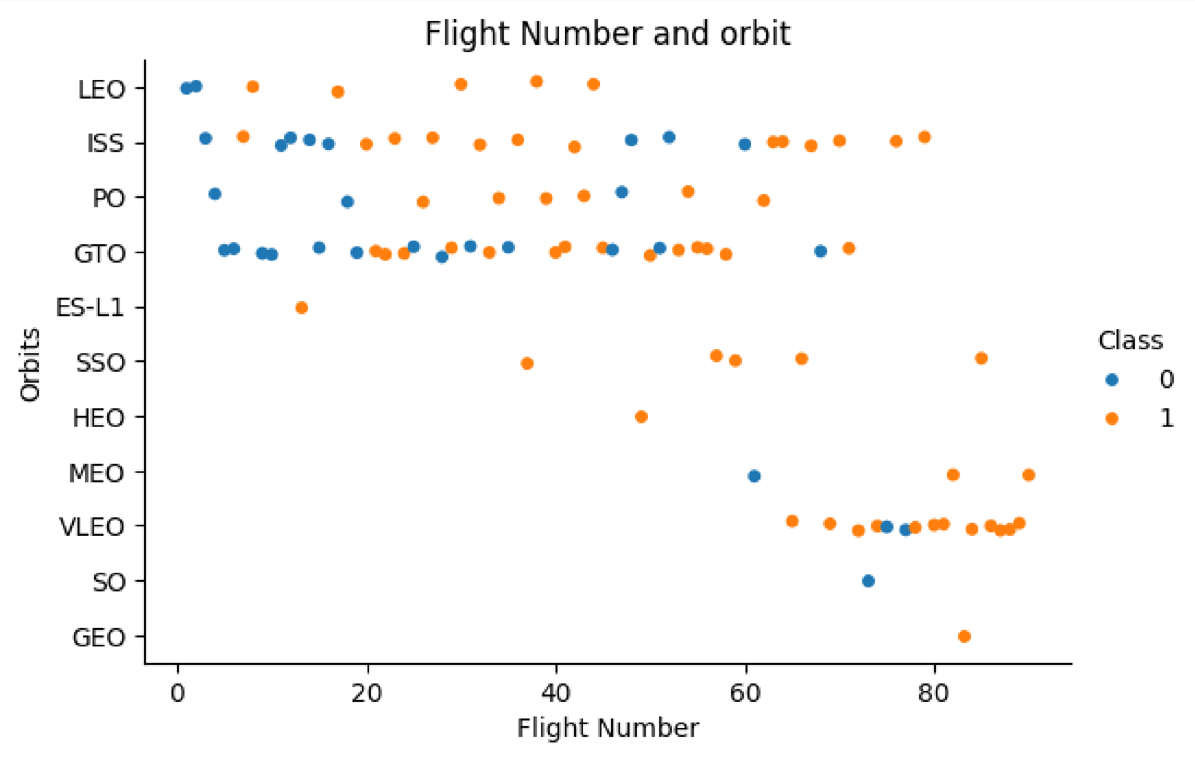
Flight Number vs. Orbit Type

Charts Plotted:

- Scatter plot with Flight Number on the x-axis and Orbit Type on the y-axis.

Reasoning:

- The purpose of this plot is to visually depict the relationship between flight numbers and the orbits of spaceflights. Each point on the scatter plot represents a specific flight number and its corresponding orbit, with different classes color-coded for clarity. The chart aims to provide insights into the distribution of flight numbers across different orbits and how classes are associated with specific flights. Overall, it serves to visually explore patterns, trends, or anomalies in the data related to flight numbers and orbits.



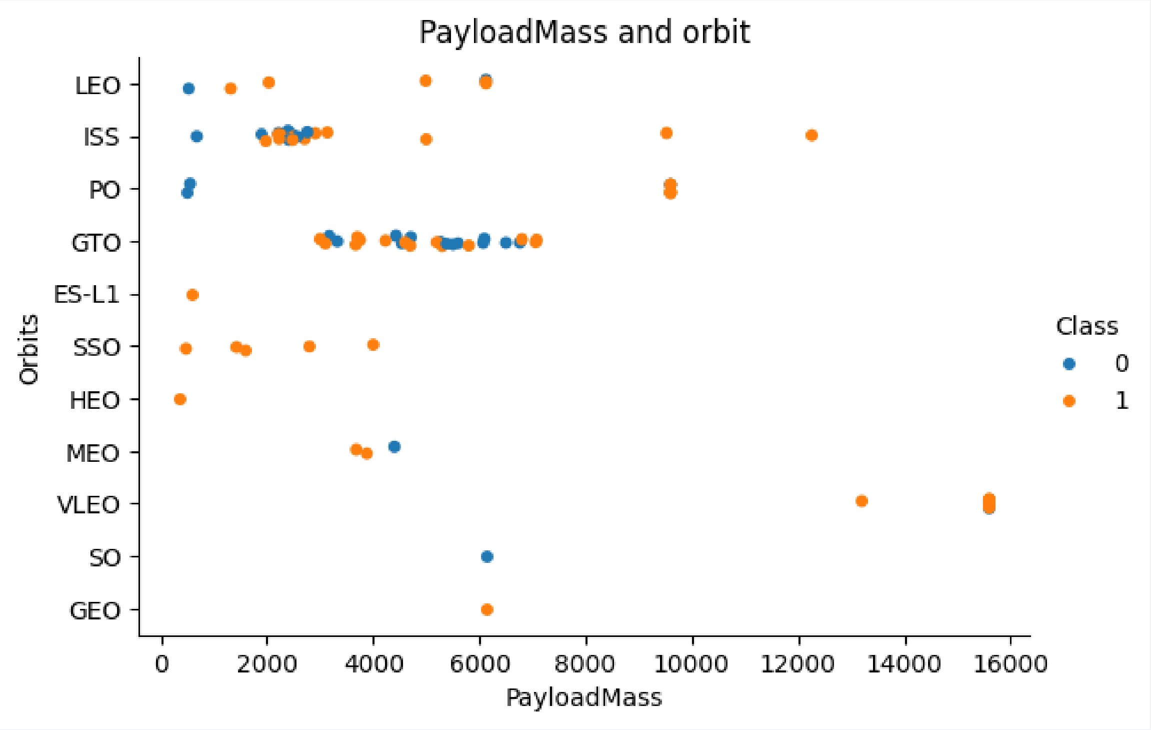
Payload vs. Orbit Type

Charts Plotted:

- Scatter plot with Payload on the x-axis and Launch Site on the y-axis.

Reasoning:

- The purpose of this chart is to visually represent the relationship between payload mass and the orbits of spaceflights. Each point on the scatter plot corresponds to a specific payload mass and its associated orbit, with different classes distinguished by color. The chart aims to provide insights into how payload masses are distributed across different orbits and how different classes are associated with specific payload masses and orbits. Overall, it serves as a visual exploration of patterns, trends, or distinctions in the data related to payload masses and orbits.



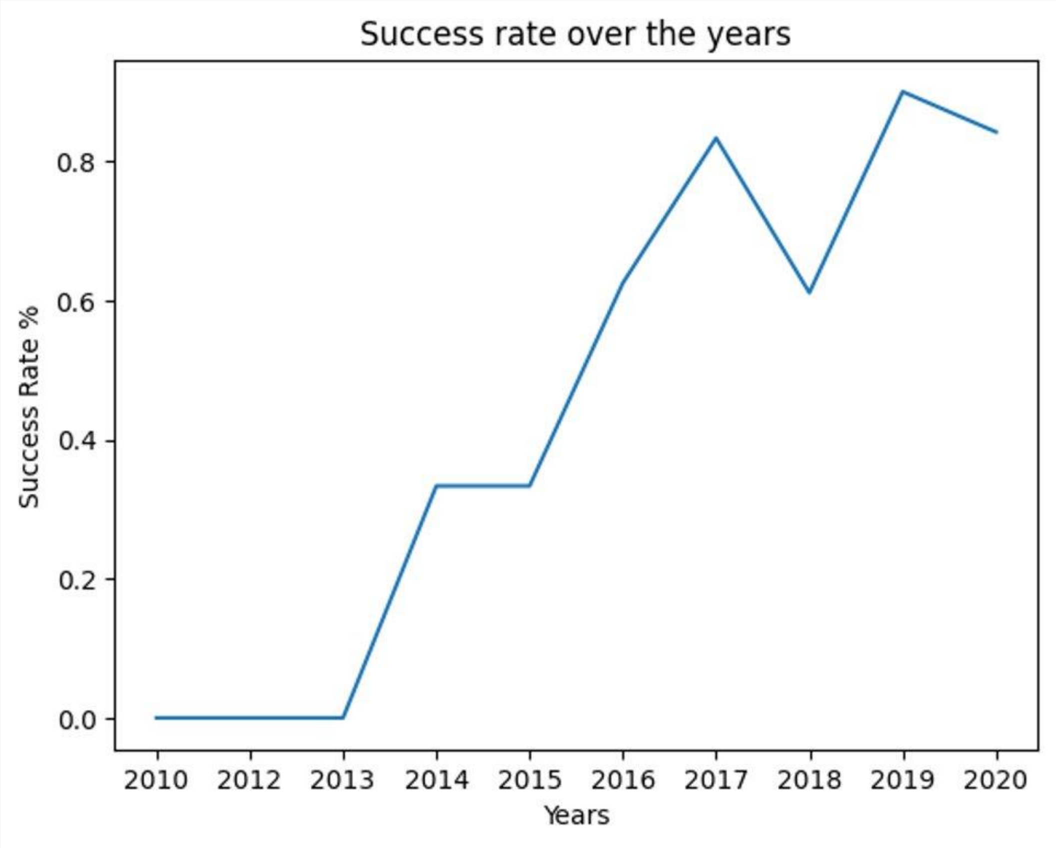
Launch Success Yearly Trend

Chart Plotted:

- Line plot with the year on the x-axis and the count or percentage of successful launches on the y-axis.

Reasoning:

- The purpose of this line chart is to visually depict the success rate percentage of spaceflights over the years. The x-axis represents the extracted years from the data, and the y-axis represents the corresponding success rates. The line chart helps to illustrate trends and variations in the success rates of spaceflights across different years, providing a clear overview of how success rates have evolved over time. The chart serves as a visual tool for understanding the historical performance of spaceflights in terms of their success rates.



All Launch Site Names

SQL QUERY:

- SELECT DISTINCT(Launch_Site) FROM SPACEXTABLE;

Distinct Launch Sites:

- This query identifies and returns a list of distinct launch sites from the SPACEXTABLE. It helps in understanding the variety of locations from which SpaceX launches its missions.

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

Launch Site Names Begin with 'CCA'

SQL QUERY:

- SELECT DATE ,LAUNCH_SITE,BOOSTER_VERSION,ORBIT,MISSION_OUTCOME FROM SPACEXTABLE WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5:

Sample Data for Launch Site Beginning with 'CCA':

- By specifying the launch site begins as 'CCA', this query retrieves a sample of 5 rows from the SPACEXTABLE to provide a quick look at the data associated with launches from this specific sites.

Date	Launch_Site	Booster_Version	Orbit	Mission_Outcome
2010-06-04	CCAFS LC-40	F9 v1.0 B0003	LEO	Success
2010-12-08	CCAFS LC-40	F9 v1.0 B0004	LEO (ISS)	Success
2012-05-22	CCAFS LC-40	F9 v1.0 B0005	LEO (ISS)	Success
2012-10-08	CCAFS LC-40	F9 v1.0 B0006	LEO (ISS)	Success
2013-03-01	CCAFS LC-40	F9 v1.0 B0007	LEO (ISS)	Success

Total Payload Mass

SQL QUERY:

- SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total Payload' FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';

Total Payload for NASA (CRS) Missions:

- The query calculates the total payload mass for missions where the customer is 'NASA (CRS)'. This can be useful in assessing the cumulative payload contribution of SpaceX to NASA's Commercial Resupply Services (CRS) missions.

Total Payload
45596

Average Payload Mass by F9 v1.1

SQL QUERY:

- SELECT AVG(PAYLOAD_MASS__KG_) AS 'AVG Payload' FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

Average Payload for Booster Version 'F9 v1.1':

- Similar to the previous query, this one calculates the Average payload mass, but specifically for missions that used the booster version 'F9 v1.1'. It helps in understanding the payload capacity of this particular booster version.

Avg Payload
2928.4

First Successful Ground Landing Date

SQL QUERY:

- SELECT MIN(Date) AS DATE FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';

Earliest Successful Landing Date:

- This query finds and returns the earliest date on which SpaceX successfully landed a booster on a ground pad. It provides insights into the historical timeline of successful landings.

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL QUERY:

- SELECT DISTINCT(Booster_Version) FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

Booster Versions for Payload Between 4000 and 6000 KG:

- By filtering payload masses between 4000 and 6000 kilograms, this query lists distinct booster versions associated with missions falling within this payload range. It aids in understanding the variety of boosters used for medium-range payloads.

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

SQL QUERY:

- SELECT 'Success' AS Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Success%' UNION ALL SELECT 'Failure' AS Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Failure%';

Mission Outcome Counts (Success/Failure):

- This query presents a summary of mission outcomes, counting the occurrences of success and failure. It combines data from both ground pad and drone ship landings, offering an overall view of mission success rates.

Outcome	Count
Success	100
Failure	1

Boosters Carried Maximum Payload

SQL QUERY:

- SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

- **Booster Version for Maximum Payload:**
  - Identifying the booster version used for the mission with the maximum payload mass, this query provides information on the most powerful booster in terms of payload capacity.

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 QUERY:

- SELECT Date, Mission_Outcome, Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE Landing_Outcome = 'Failure (drone ship)' AND Date LIKE '2015%';

Failed Landings in 2015 with Drone Ship:

- Focusing on the year 2015, this query retrieves details of missions where the landing outcome was a failure specifically on a drone ship. It helps in understanding challenges faced during early drone ship landings.

Date	Mission_Outcome	Booster_Version	Launch_Site	Landing_Outcome
2015-01-10	Success	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	Success	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

SQL QUERY:

- SELECT Landing_Outcome, COUNT(*) as Rank FROM spacextable WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Rank DESC;

Landing Outcome Counts:

- This SQL query counts and ranks the occurrences of different landing outcomes in the `spacextable` table for SpaceX missions between June 4, 2010, and March 20, 2017. The results are ordered by the frequency of each outcome in descending order, with the count labeled as 'Rank'.

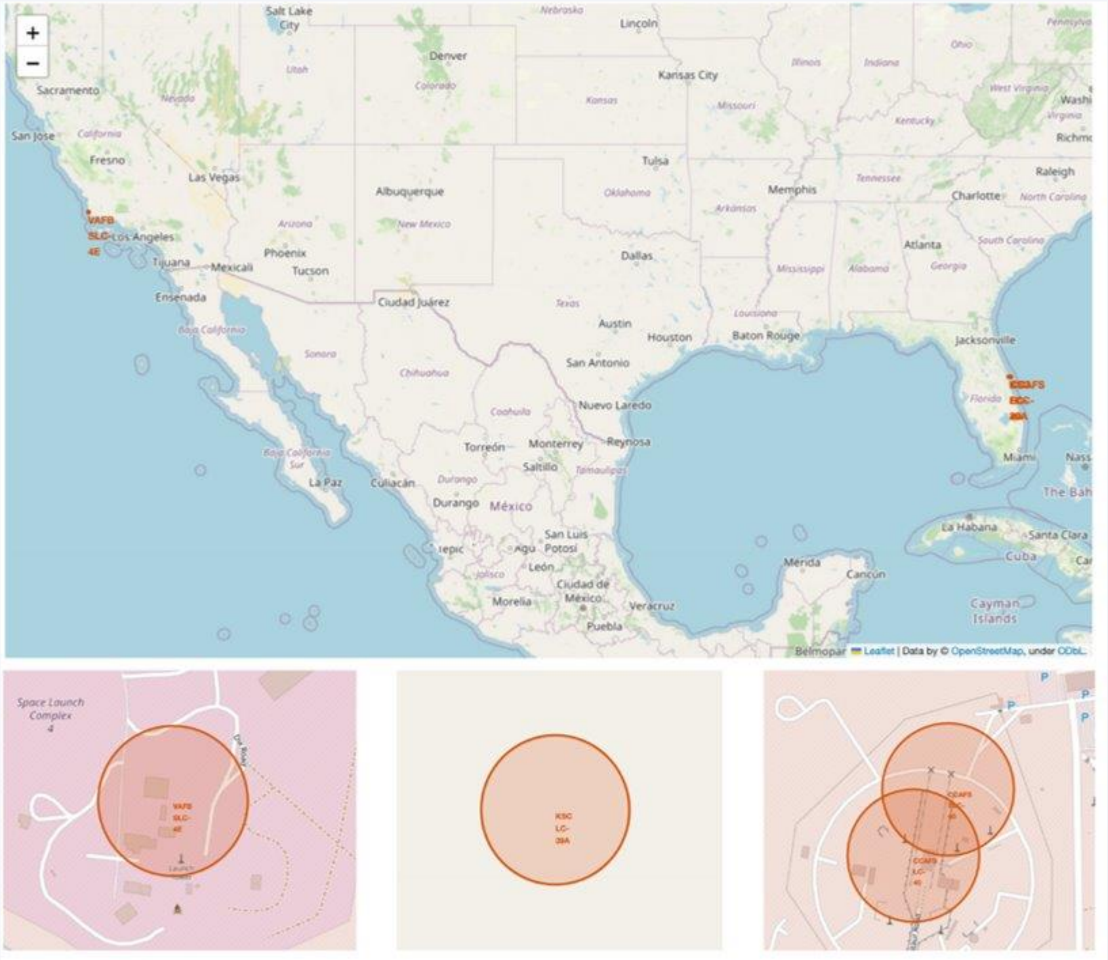
Landing_Outcome	Rank
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

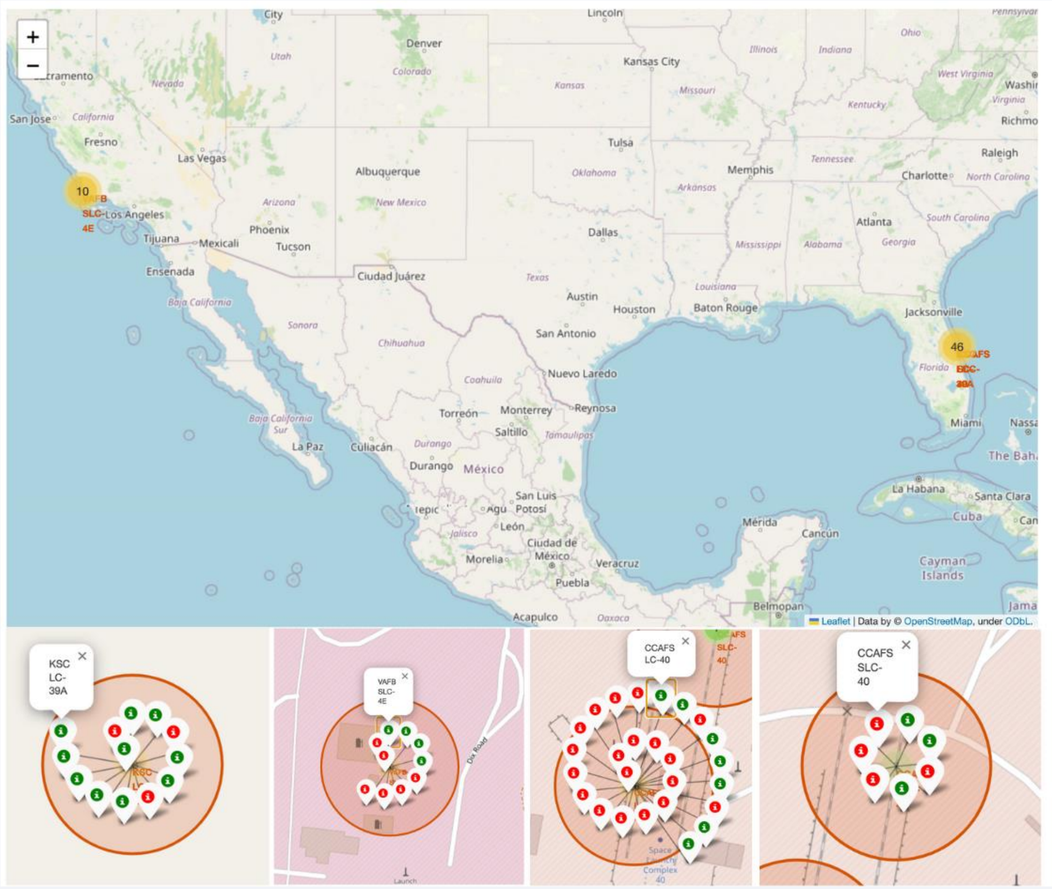
Marking Launch Sites

- The map provides a comprehensive geographical overview of various launch sites, with each site's coverage area delineated by a uniquely styled circle. The map is designed for interactive exploration, allowing users to zoom, pan, and click for a dynamic and detailed examination of launch facility information.
- Markers with unique labels make it easy to identify and distinguish between different launch sites, offering quick access to essential information. Upon clicking a circle, users can access popups that deliver concise launch site names, facilitating a rapid understanding of key details. In conclusion, the map efficiently communicates the spatial distribution of launch sites, highlighting coverage areas and ensuring a user-friendly exploration experience.



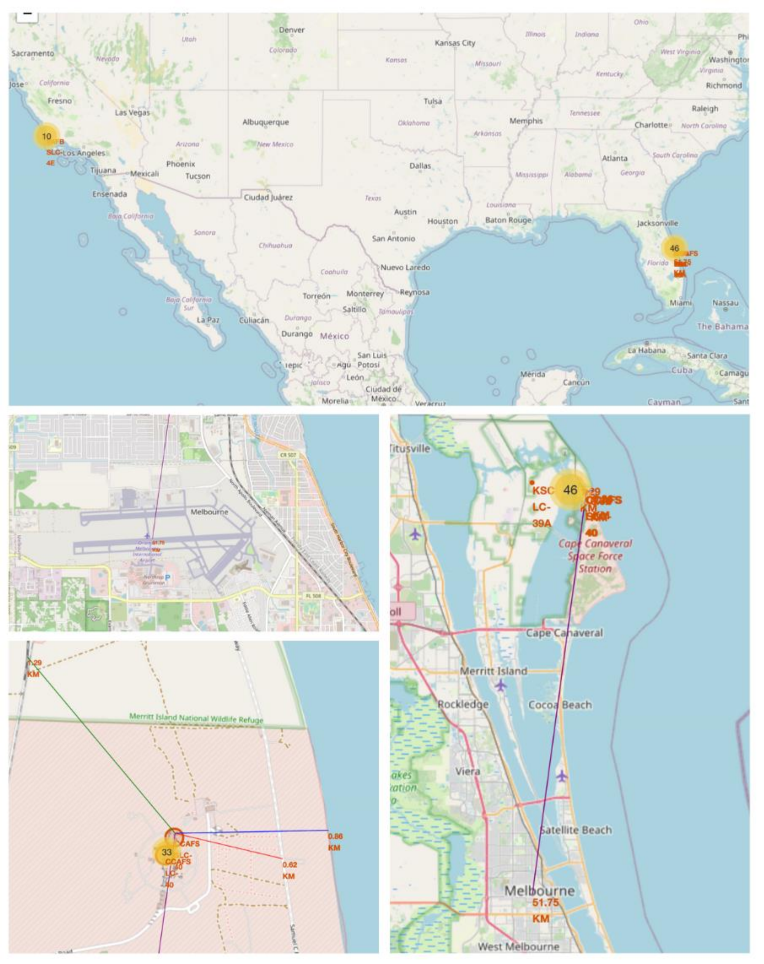
SpaceX Launch Outcome Map

- The map uses circles and markers to illustrate SpaceX launch sites, with marker colors indicating launch outcomes (green for success, red for failure). Each launch site's coverage is represented by a distinctive colored circle. Marker clustering enhances readability, and clicking on a circle or marker reveals detailed information. Customized markers with dynamic colors contribute to a visually informative representation. Overall, the map effectively communicates SpaceX launch sites, outcomes, and coverage areas, providing insights into spatial and operational aspects.



Launch Site Infrastructure Map

- The map displays key infrastructure points near the launch site, such as coastline, railway, highway, and city, with distinctive colored markers. Distance information in DivIcons enhances clarity. PolyLines visually connect each point to the launch site, illustrating spatial relationships. Different colors differentiate infrastructure types, contributing to a comprehensive map. This design effectively communicates the launch site's proximity to significant features, offering insights into spatial relationships and accessibility.

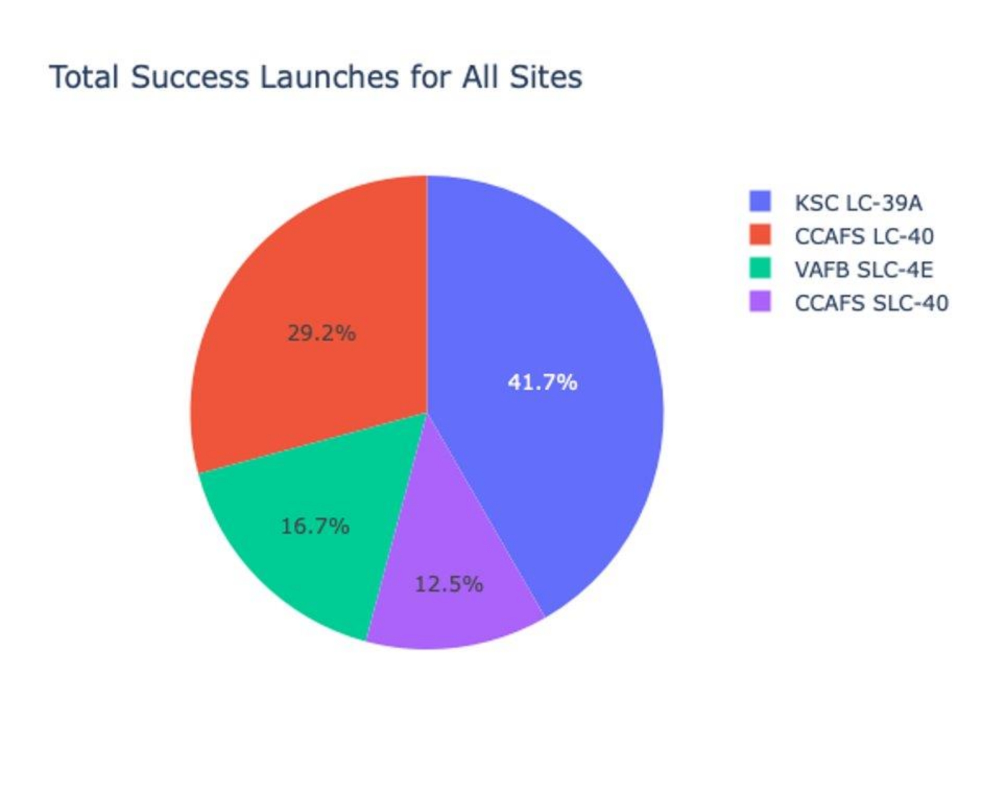


Section 4

Build a Dashboard with Plotly Dash

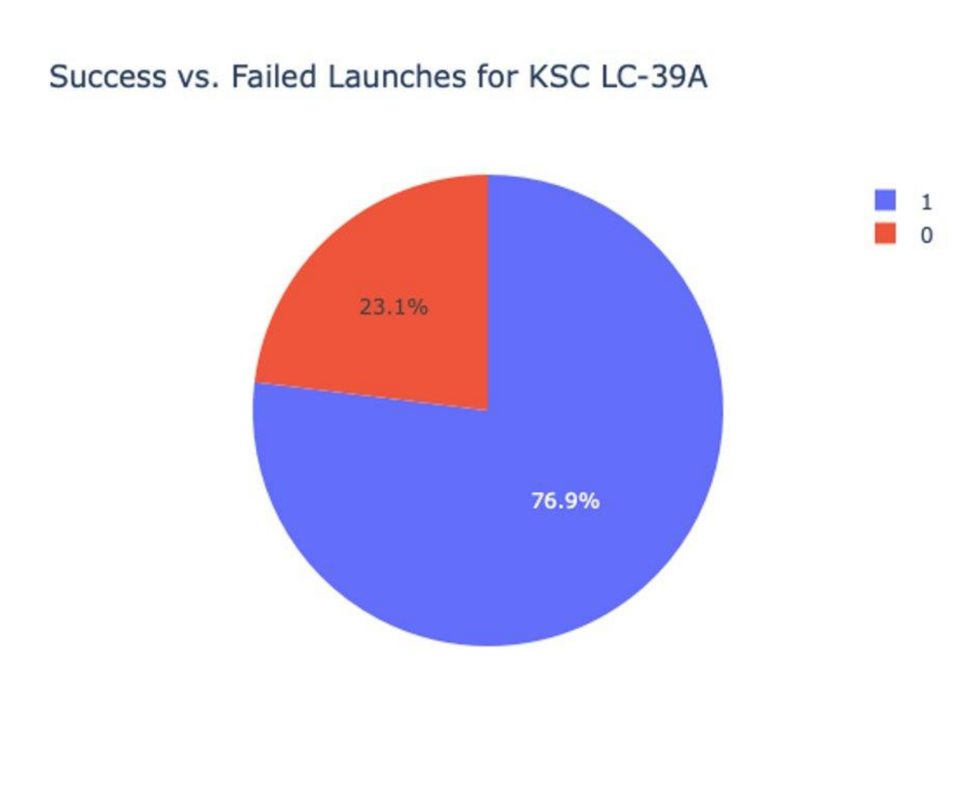
Total Success Launches for All Sites

- The success rates for launches from various sites reveal distinct performance levels. KSC LC-39A stands out with the highest success rate at 41.7%, indicating a reliable track record. CCAFS LC-40 follows closely with a 29.2% success rate, showcasing consistent performance. VAFB SLC-4E has a success rate of 16.7%, while CCAFS SLC-40 trails with the lowest success rate of 12.5%, suggesting potential challenges or issues at this launch site. Stakeholders can leverage this data to inform strategic decisions, emphasizing the need for improvements in sites with lower success rates to enhance overall launch reliability and success.



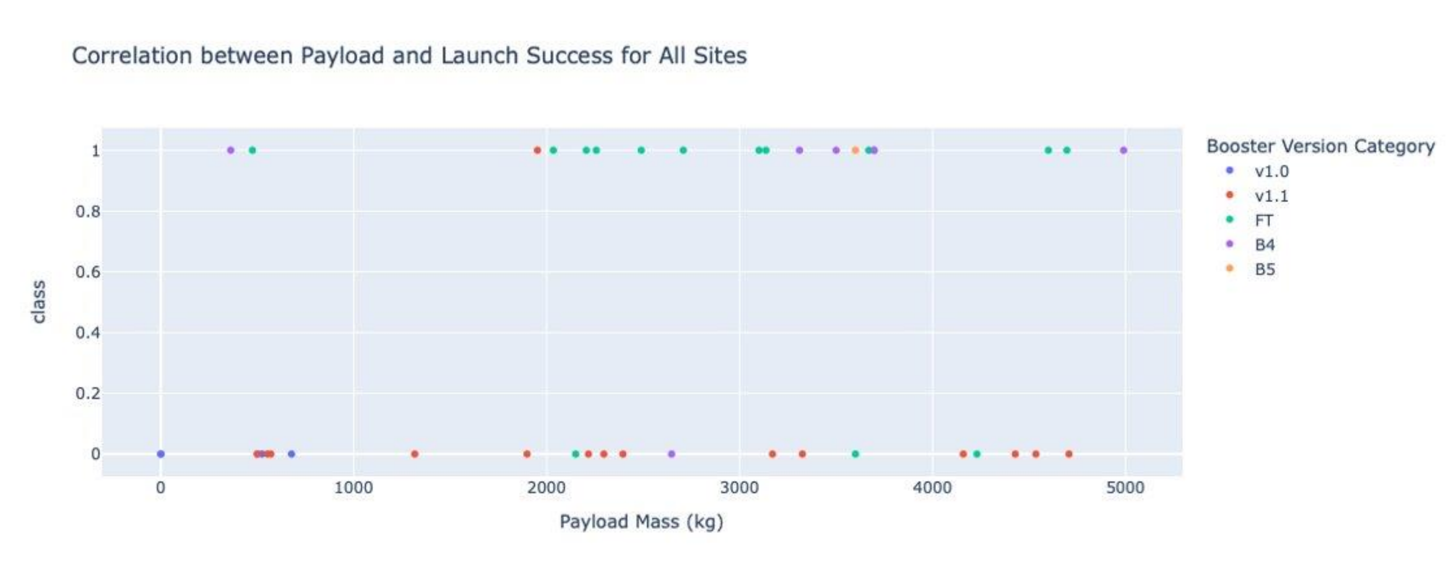
Success v/s Failed Launches For KSC LC-39A

- The success analysis for KSC LC-39A showcases a strong performance with a success rate of 76.9% and an unsuccessful launch rate of 23.1%. This indicates a predominantly successful track record, highlighting the reliability and effectiveness of launches from KSC LC-39A. The majority of launches from this site have been successful, making it a robust choice for future missions. The data suggests that KSC LC-39A is a high-performing launch site with a significantly higher success rate compared to unsuccessful launches, underscoring its importance in the context of successful space missions.



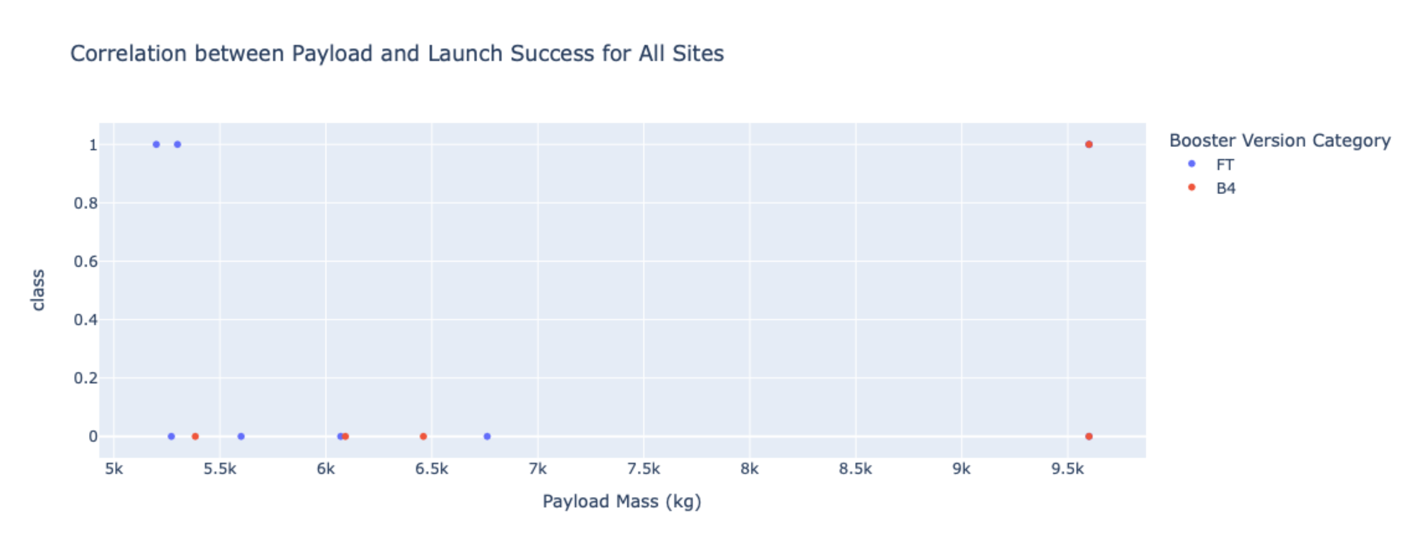
Correlation between Payloads & Launch Success

- The correlation analysis between payload and launch success for all sites reveals insights based on different booster versions and payload categories. The plot, with a range selector between 0 to 5000 kg on the x-axis, illustrates the mission outcomes (successful or unsuccessful) on the y-axis for booster versions v1.0, v1.1, FT, B4, and B5. The data allows stakeholders to observe the relationship between payload weight and launch success for each booster category. This information is crucial for understanding the performance of specific booster versions with varying payload capacities and can guide decision-making in optimizing payload assignments for future launches across different booster categories.



Correlation between Payloads & Launch Success

- The correlation analysis between payload and launch success for all sites focuses on the booster version categories FT and B4. The plot features a range selector between 5000 to 10000 kg on the x-axis, depicting different payloads within this range for both booster categories. The y-axis represents the mission outcomes, indicating whether launches were successful or unsuccessful. This specific analysis allows stakeholders to assess the relationship between payload weight and launch success for FT and B4, providing valuable insights into the performance of these booster versions within the specified payload range. This information can be instrumental in refining payload assignments and decision-making for future launches involving FT and B4 boosters.

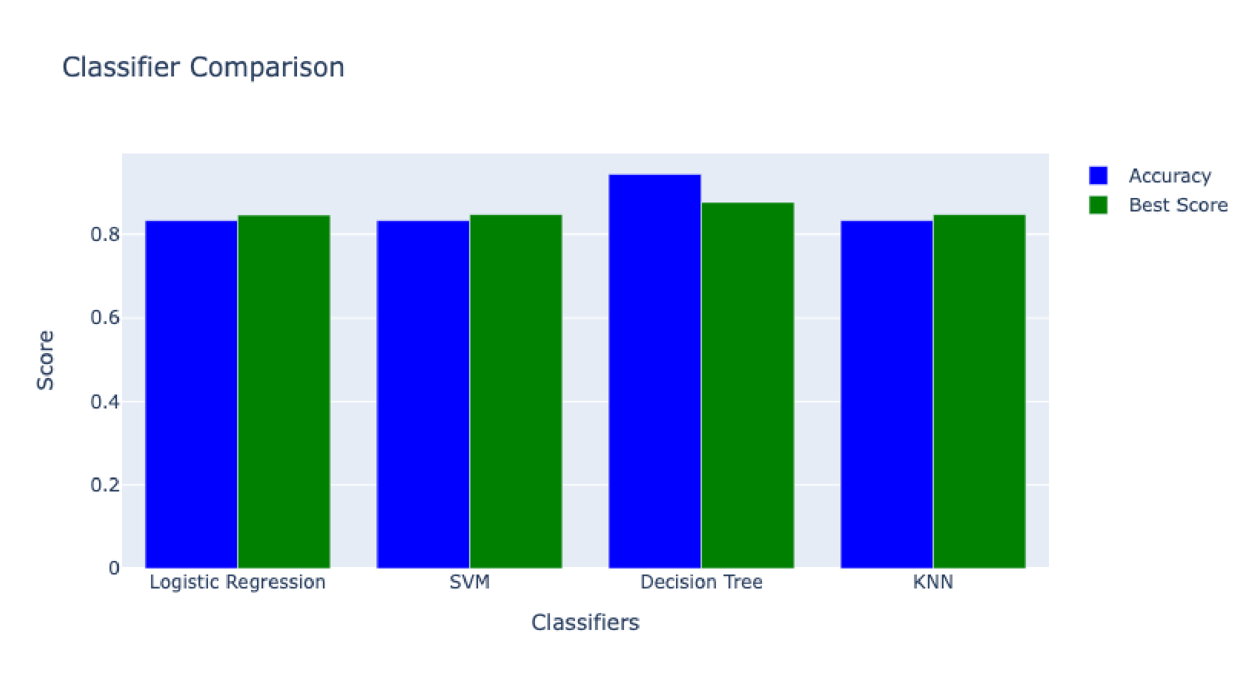


Section 5

Predictive Analysis (Classification)

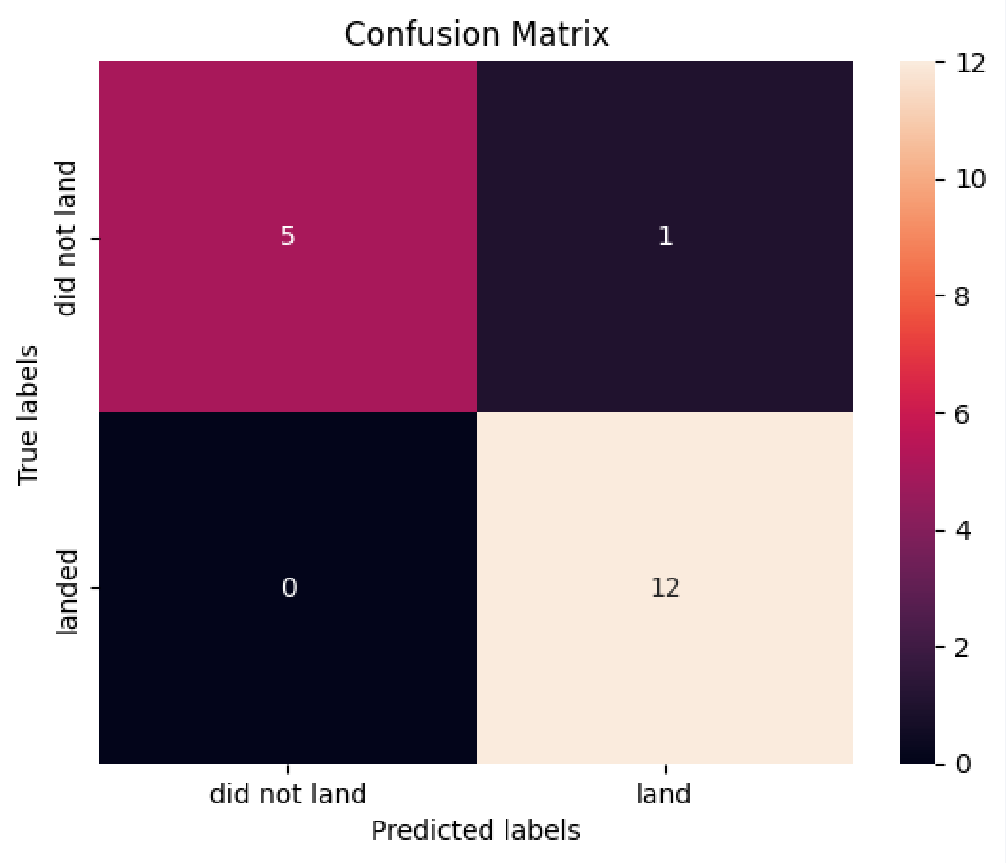
Classification Accuracy

- In our analysis, we employed four classification models: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (KNN). After visualizing the accuracy scores of these models in a bar chart, it became evident that the Decision Tree model outperformed the others, showcasing the highest accuracy. This implies that the Decision Tree algorithm was most effective in correctly classifying the data compared to Logistic Regression, SVM, and KNN. The choice of Decision Tree as the optimal model for this classification task is supported by its superior accuracy score.



Confusion Matrix

- Did not land (True Negatives, TN): There are 5 instances where the model correctly predicted that an event did not happen (e.g., a flight did not land), and it indeed did not happen.
- Landed (True Positives, TP): There are 12 instances where the model correctly predicted that an event happened (e.g., a flight landed), and it indeed happened.
- False Not Landed (False Positives, FP): There is 1 instance where the model incorrectly predicted that an event happened (e.g., a flight landed), but it did not happen (did not land).
- False Landed (False Negatives, FN): There are 0 instances where the model incorrectly predicted that an event did not happen (e.g., a flight did not land), but it did happen (landed).



Conclusion

- In conclusion, this project successfully achieved its objectives in analyzing and visualizing key metrics related to SpaceX's Falcon 9 launches. The comprehensive data collection, including SpaceX API calls and web scraping of Wikipedia, provided a rich dataset for thorough exploration. The methodologies employed, such as data wrangling and exploratory data analysis, ensured the integrity and meaningful interpretation of the data.
- The incorporation of interactive visualizations, including Folium maps and Plotly Dash dashboards, allowed for dynamic exploration and insightful analysis. The addition of pie charts, scatter plots, and interactive features like dropdowns and sliders enhanced user engagement and facilitated a deeper understanding of the data.
- The specific analyses, such as comparing success counts across launch sites, exploring the relationship between payload mass and launch success, and visualizing spatial relationships with Folium, provided valuable insights into SpaceX's launch history. The machine learning model for predicting first stage reusability further added a predictive dimension to the project.
- The GitHub repository serves as a comprehensive reference for the project, enabling external review and validation. The inclusion of thoughtful interactions, such as the launch site dropdown and payload range slider, enhances user flexibility and facilitates a customized exploration experience.
- In essence, this project not only met its technical objectives but also provided a user-friendly and informative platform for stakeholders to gain valuable insights into SpaceX's Falcon 9 launches. The methodologies, visualizations, and interactions collectively contribute to the project's success in achieving its goals.

Appendix

[GitHub Data Collection – SpaceX API]

[GitHub Data Collection - Scraping]

[GitHub Data Wrangling]

[GitHub EDA with data visualization]

[GitHub EDA with SQL]

[GitHub Interactive Maps]

[GitHub DashBoard]

[GitHub Model Development]

Thank you!