IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

AGUEDO ESLERA
April 02, 2024

Outline

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

Executive Summary

Methodologies Summary:	Utilized RESTful APIs and web scraping for data collection of Falcon 9 launch data.
	Conducted data wrangling to ensure data quality and consistency.
	Employed exploratory data analysis techniques to uncover insights and patterns.
	Utilized predictive modeling techniques such as SVM, Classification Trees, and Logistic Regression.
	Created interactive visualizations using Plotly Dash and Folium for enhanced data exploration.
Results Summary:	Identified determinants of successful Falcon 9 first-stage landings.
	SVM showed the highest accuracy among the predictive models tested.
	Interactive visualizations provided intuitive data exploration tools.
	Findings contribute valuable insights for stakeholders in the space industry.

Introduction

Project Background and Context:

- The modern space race has shifted from a geopolitical competition to a commercial endeavor, with companies vying to dominate the emerging space industry. In this context, leveraging data science has become imperative for achieving a competitive edge. Our project, "Winning Space Race with Data Science," aims to explore the intersection of data science and space exploration, focusing on the domain of rocket launches and landing success.

Problems to Address:

- In this project, we seek to address several key challenges facing the space industry. These include predicting the success of Falcon 9 first-stage landings, optimizing launch site selection, and extracting actionable insights from extensive datasets provided by organizations like SpaceX. By tackling these problems, we aim to enhance decision-making processes, reduce operational risks, and ultimately contribute to the advancement of space exploration efforts.

Section 1

Methodology

Methodology

Executive Summary

Data Collection Methodology:

- Data collection involved retrieving information from multiple sources, including the SpaceX REST API and web scraping techniques.
- The SpaceX REST API provided launch data, including details about the rockets, payloads, launch specifications, and landing outcomes.
- Web scraping techniques were utilized to gather additional Falcon 9 launch records from relevant sources, ensuring a comprehensive dataset.

Perform Data Wrangling:

- Data wrangling encompassed cleaning and organizing the collected data to prepare it for analysis.
- Techniques such as handling missing values, converting data types, and filtering out irrelevant information were employed to enhance data quality.
- This phase aimed to create a structured and standardized dataset suitable for analysis.

Methodology

Perform Exploratory Data Analysis (EDA) Using Visualization and SQL:

- EDA was conducted to gain insights into the dataset and identify patterns, trends, and relationships among variables.
- Visualization techniques, such as scatter plots, bar charts, and pie charts, were utilized to visualize the data and uncover meaningful insights.
- SQL queries were employed to extract and analyze specific subsets of data, facilitating a deeper understanding of the dataset.

Perform Interactive Visual Analytics Using Folium and Plotly Dash:

- Interactive visual analytics involved the creation of interactive maps and dashboards to explore and analyze launch site proximity, landing outcomes, and other relevant factors.
- Folium was utilized to generate interactive maps, visualize launch site locations, and analyze geographical proximity.
- Plotly Dash was employed to build interactive dashboards containing pie charts, scatter plots, and other visualizations for in-depth analysis.

Methodology

Perform Predictive Analysis Using Classification Models:

- Predictive analysis focused on building classification models to predict the success of Falcon 9 first-stage landings.
- Various classification algorithms, including Logistic Regression, Support Vector Machines (SVM), Decision Tree Classifier, and K-Nearest Neighbors (KNN), were trained, tuned, and evaluated.
- Techniques such as cross-validation and hyperparameter tuning were employed to optimize model performance and ensure robust predictions.
- The classification models were evaluated based on metrics such as accuracy, precision, recall, and F1-score, providing insights into their effectiveness in predicting landing outcomes.

Data Collection

Description:

- Datasets were collected from multiple sources to ensure comprehensive coverage of relevant information.
- The data collection process involved accessing both structured and unstructured data sources to gather diverse insights.
- Various APIs and web scraping techniques were employed to extract data from online repositories and platforms.
- A meticulous approach was adopted to ensure data integrity and quality throughout the collection process.

Key Phrases:

- Comprehensive data collection from multiple sources.
- Structured and unstructured data sources accessed.
- Utilization of APIs and web scraping techniques.
- Emphasis on data integrity and quality assurance.

Data Collection – SpaceX API

Description:

• The data collection process involved making RESTful API calls to SpaceX to retrieve relevant information about Falcon 9 first-stage landings.

• SpaceX REST API provided access to a wealth of data regarding launch details, mission outcomes, and landing attempts.

• Python code was utilized to make HTTP requests and parse the JSON responses from the SpaceX API.

• The collected data was then processed and stored for further analysis and modeling.

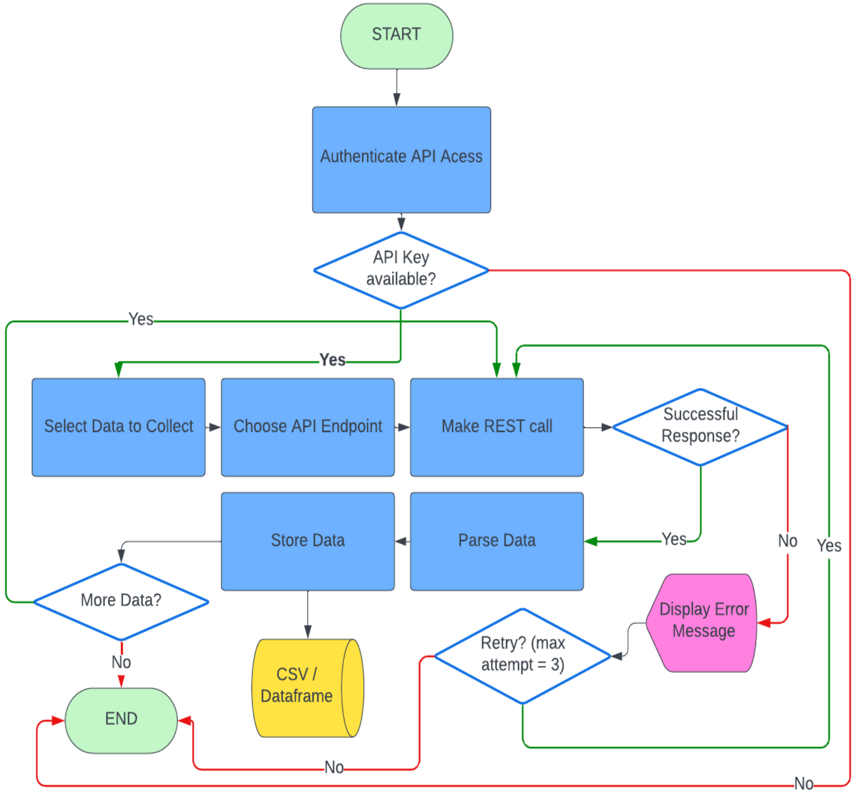
Key Phrases:

• Retrieval of Falcon 9 first-stage landing data from SpaceX REST API.

• Utilization of Python for making HTTP requests and parsing JSON responses.

• Comprehensive access to launch details, mission outcomes, and landing attempts.

• Data processing and storage for subsequent analysis and modeling.



GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb

Data Collection - Scraping

Description:

Web scraping was employed to gather supplementary data on Falcon 9 launches and landings from external sources such as space news websites and SpaceX-related forums.

Python libraries like BeautifulSoup and requests were utilized for scraping HTML content from web pages.

The scraped data included additional details on mission objectives, launch outcomes, and post-landing analyses.

Extracted data was processed and integrated with the existing dataset obtained from SpaceX API calls for comprehensive analysis.

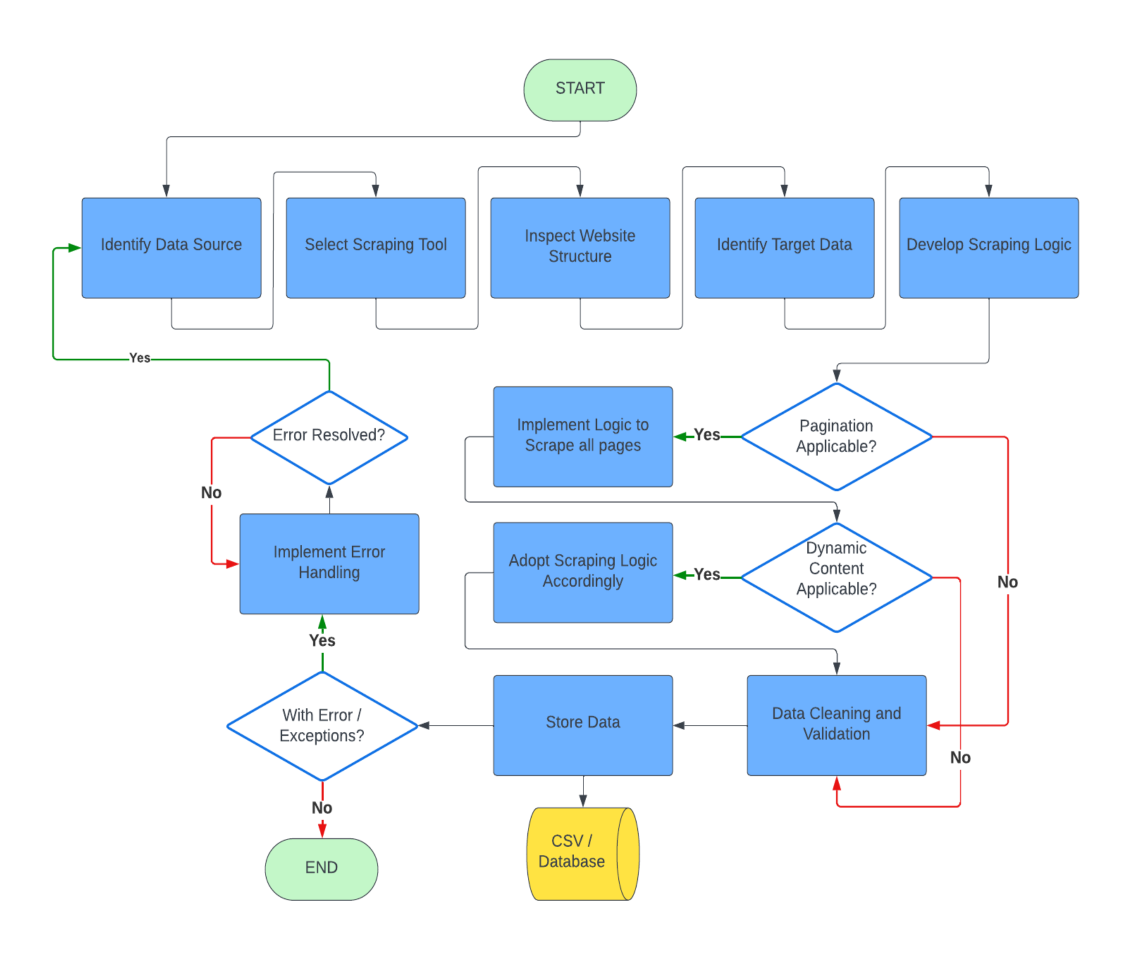
Key Phrases:

Extraction of supplementary Falcon 9 launch and landing data through web scraping.

Implementation of Python libraries like BeautifulSoup and requests for HTML content extraction.

Inclusion of mission objectives, launch outcomes, and post-landing analyses in the scraped dataset.

Integration of scraped data with existing SpaceX API data for enhanced analysis.



GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Description:

Data wrangling involved preprocessing and cleaning the collected datasets to ensure consistency and accuracy for subsequent analysis.

Techniques such as handling missing values, standardizing data formats, and removing duplicates were applied to improve data quality.

Python libraries like Pandas were utilized for efficient data manipulation and transformation tasks.

The wrangling process encompassed merging multiple datasets, performing feature engineering, and creating derived variables to enhance analytical insights.

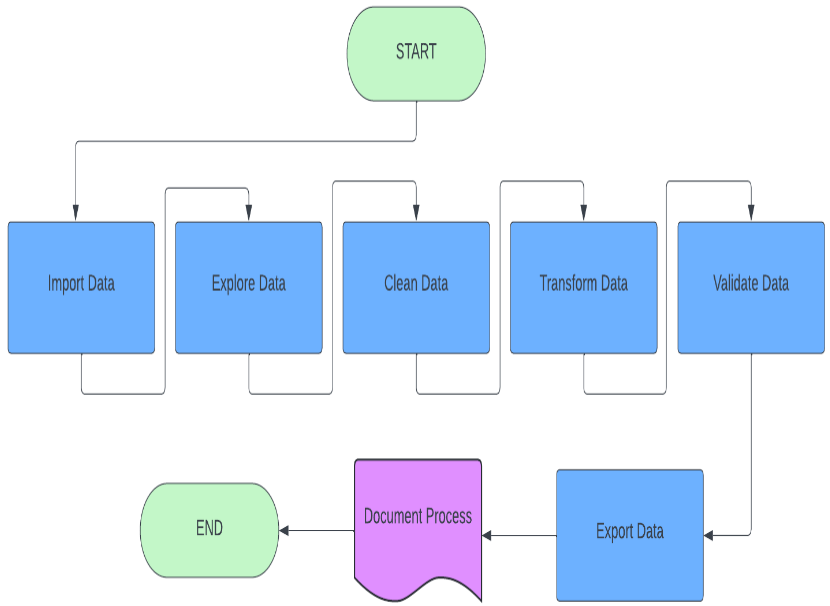
Key Phrases:

Preprocessing and cleaning of collected datasets to ensure consistency and accuracy.

Application of techniques such as handling missing values, standardizing data formats, and removing duplicates.

Utilization of Python libraries like Pandas for efficient data manipulation and transformation.

Integration of multiple datasets, feature engineering, and creation of derived variables to enhance analytical insights.



GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

EDA with Data Visualization

Summary:

• Various charts and graphs were plotted during the Exploratory Data Analysis (EDA) to gain insights into the SpaceX dataset.

• Scatter plots, bar charts, histograms, and pie charts were utilized to visualize different aspects of the data and explore relationships between variables.

• Scatter plots were employed to examine correlations between numerical variables, while bar charts and histograms were used for categorical variables.

• Pie charts provided a breakdown of categorical data distributions, highlighting proportions and percentages.

• The choice of charts was based on the need to understand the distribution, trends, and patterns within the data, facilitating the identification of key insights and trends.

Explanation:

• Scatter plots helped identify correlations between variables, aiding in feature selection for predictive modeling.

• Bar charts and histograms visualized the distribution of categorical variables, such as launch outcomes and success rates.

• Pie charts offered a clear representation of categorical data proportions, enabling a quick understanding of data distributions.

GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-eda-dataviz%20(1).ipynb

EDA with SQL

Summary of SQL Queries

- Extracted data from the database to analyze Falcon 9 launch records.
- Calculated success rates based on different parameters such as launch site, payload mass, and mission outcome.
- Identified correlations between launch success and various factors like launch site location and payload mass.
- Conducted exploratory analysis to understand the relationship between different variables and successful landings.

GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Summary of Map Objects

- Created markers to represent launch sites.
- Added circles to visualize the proximity of each launch site.
- Incorporated lines to connect launch sites for better understanding of their distribution.
- Utilized popups to provide additional information about each launch site when clicked.

Explanation

- Markers: Used to pinpoint the exact locations of launch sites on the map.
- Circles: Represented the proximity of each launch site, with larger circles indicating a wider coverage area.
- Lines: Connected launch sites to illustrate their geographical distribution and relative distances.
- Popups: Provided detailed information about each launch site, including its name, location, and success rate.

GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Summary of Plots/Graphs and Interactions:

• Included pie charts to visualize the distribution of successful and unsuccessful launches.

• Added scatter plots to analyze correlations between different attributes such as payload mass and launch success.

• Incorporated dropdown menus and range sliders for interactive filtering of data.

• Implemented callbacks to update charts dynamically based on user inputs.

Explanation:

• Pie Charts: Used to show the overall success rate of launches and highlight the proportion of successful and unsuccessful missions.

• Scatter Plots: Enabled exploration of relationships between variables like payload mass, orbit, and launch outcome, aiding in identifying patterns.

• Dropdown Menus and Range Sliders: Provided users with options to filter data based on specific criteria such as launch site and payload mass range, enhancing interactivity and customization.

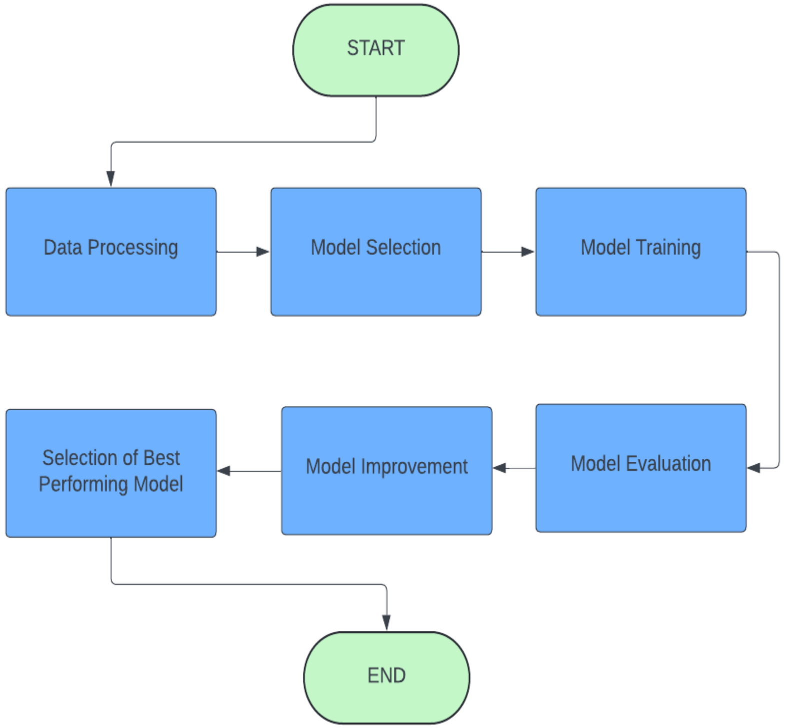
• Callbacks: Allowed for real-time updates of charts in response to user selections, providing a dynamic and responsive dashboard experience.

GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/Build_a_Dashboard_Application_with_Plotly_Dash.pdf

Predictive Analysis (Classification)

- ❖ To build, evaluate, improve, and find the best-performing classification model, the following steps were followed:

1. **Data Preprocessing**:
    1. Gathered the dataset containing features and labels.
    2. Handled missing values, if any, through imputation or removal.
    3. Conducted feature scaling or normalization if required.
    4. Split the dataset into training and testing sets.
2. **Model Selection**:
    1. Chose a set of classification algorithms based on the problem type and dataset characteristics.
    2. Initialized multiple models such as Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, etc.
3. **Model Training**:
    1. Trained each model on the training dataset.
    2. Tuned hyperparameters using techniques like grid search, random search, or Bayesian optimization.
    3. Evaluated models using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, etc.



GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)

4. Model Evaluation:

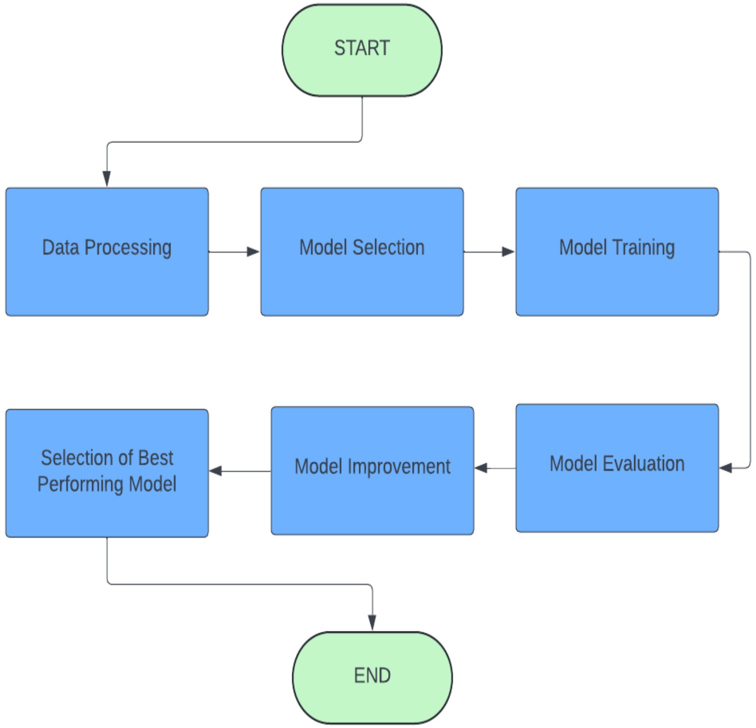
1. Evaluated the trained models on the testing dataset to assess their performance.
2. Analyzed the confusion matrix, ROC curve, and precision-recall curve to understand model behavior.
3. Identified models with high accuracy, precision, recall, and F1-score.

5. Model Improvement:

1. Investigated feature importance to identify significant predictors.
2. Conducted feature engineering or selection to enhance model performance.
3. Applied techniques like ensemble learning, stacking, or boosting to improve model accuracy.
4. Explored advanced algorithms or methodologies suited for the dataset characteristics.

6. Selection of Best Performing Model:

1. Selected the model with the highest performance metrics based on evaluation results.
2. Considered factors like interpretability, computational efficiency, and scalability.
3. Conducted cross-validation to validate the robustness of the selected model.



GitHub URL: https://github.com/GhidzE/Gido_Project2/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

Exploratory Data Analysis (EDA) Results:

- Identified key features such as launch site, payload mass, and orbit type.
- Visualized the distribution of successful and failed launches.
- Explored correlations between launch success and various factors.
- Discovered geographical patterns and proximity to coastlines or equatorial lines.

Interactive Analytics Demo in Screenshots:

- Demonstrated interactive visualization using Folium maps.
- Showcased markers for launch sites and outcomes (success/failed) on the map.
- Illustrated the distances between launch sites and their proximities.
- Integrated mouse position feature to retrieve coordinates on map hover.

Predictive Analysis Results:

- Built multiple classification models including Logistic Regression, Random Forest, and Support Vector Machine.
- Evaluated models using accuracy, precision, recall, and F1-score.
- Identified the best-performing model based on evaluation metrics.
- Presented model accuracy and performance in a bar chart for comparison.

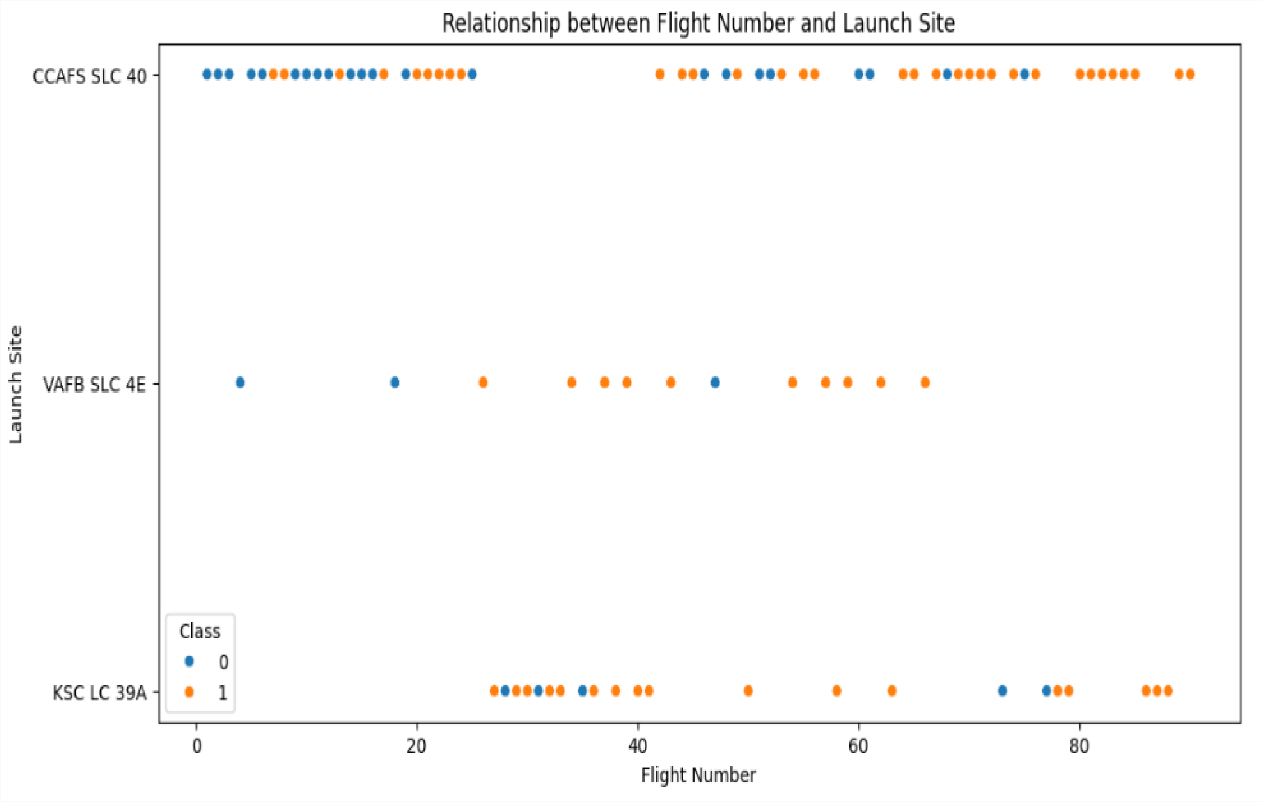
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Explanation:

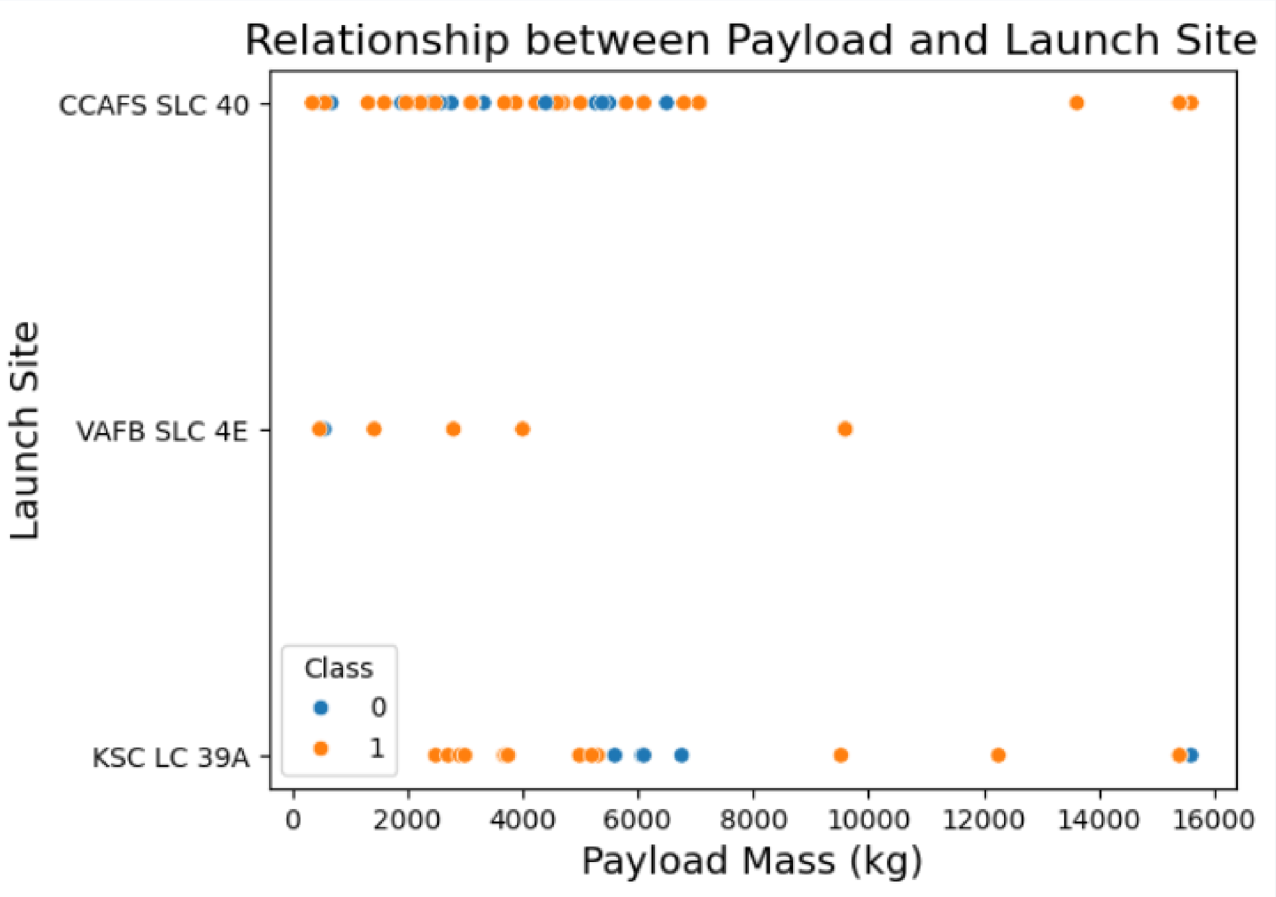
- The scatter plot illustrates the relationship between the flight number and the launch site.
- Each point represents a specific flight, with the x-axis showing the flight number and the y-axis indicating the launch site.
- The plot provides insights into the distribution of launches across different launch sites over the recorded flight numbers.



Payload vs. Launch Site

Explanation:

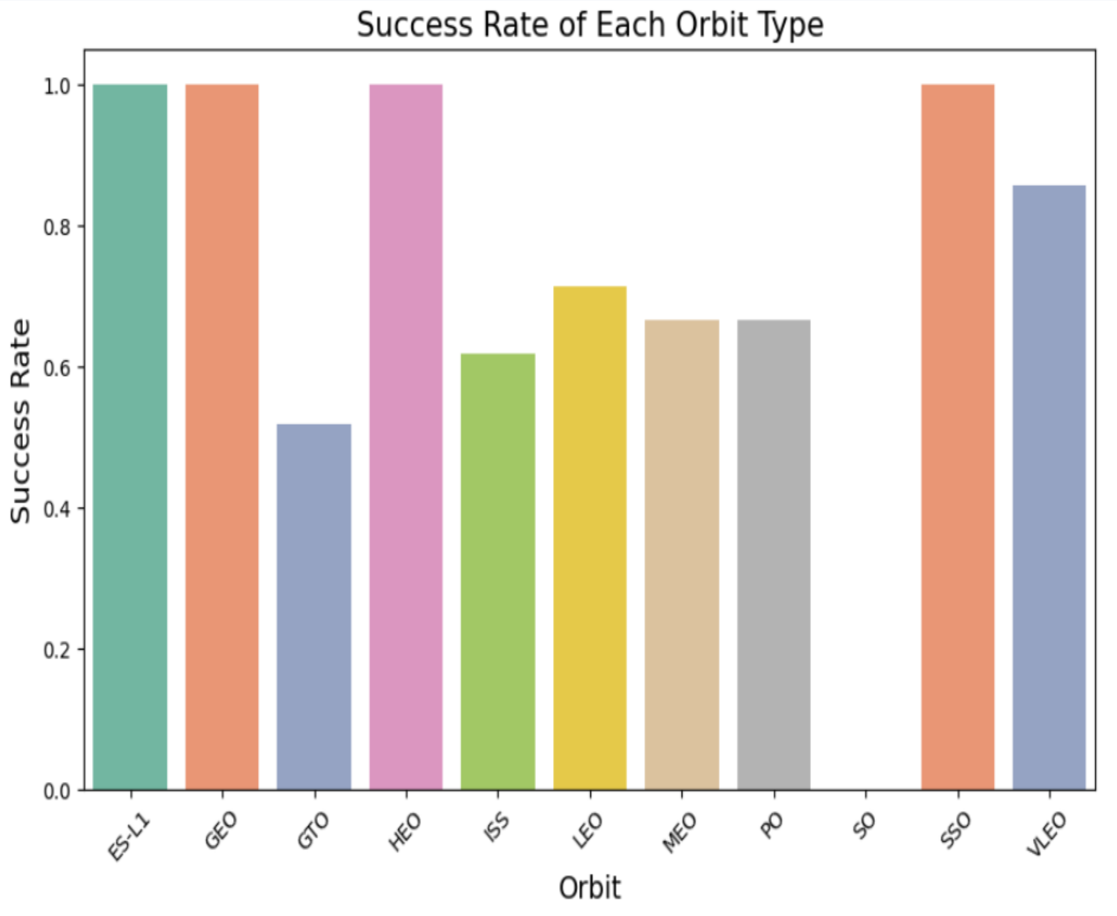
- The scatter plot illustrates the relationship between the payload and the launch site.
- Each point represents a specific payload launched from a particular launch site.
- The x-axis represents the payload, while the y-axis indicates the launch site.
- This plot helps visualize the distribution of payloads across different launch sites and their corresponding sizes or weights.



Success Rate vs. Orbit Type

Explanation:

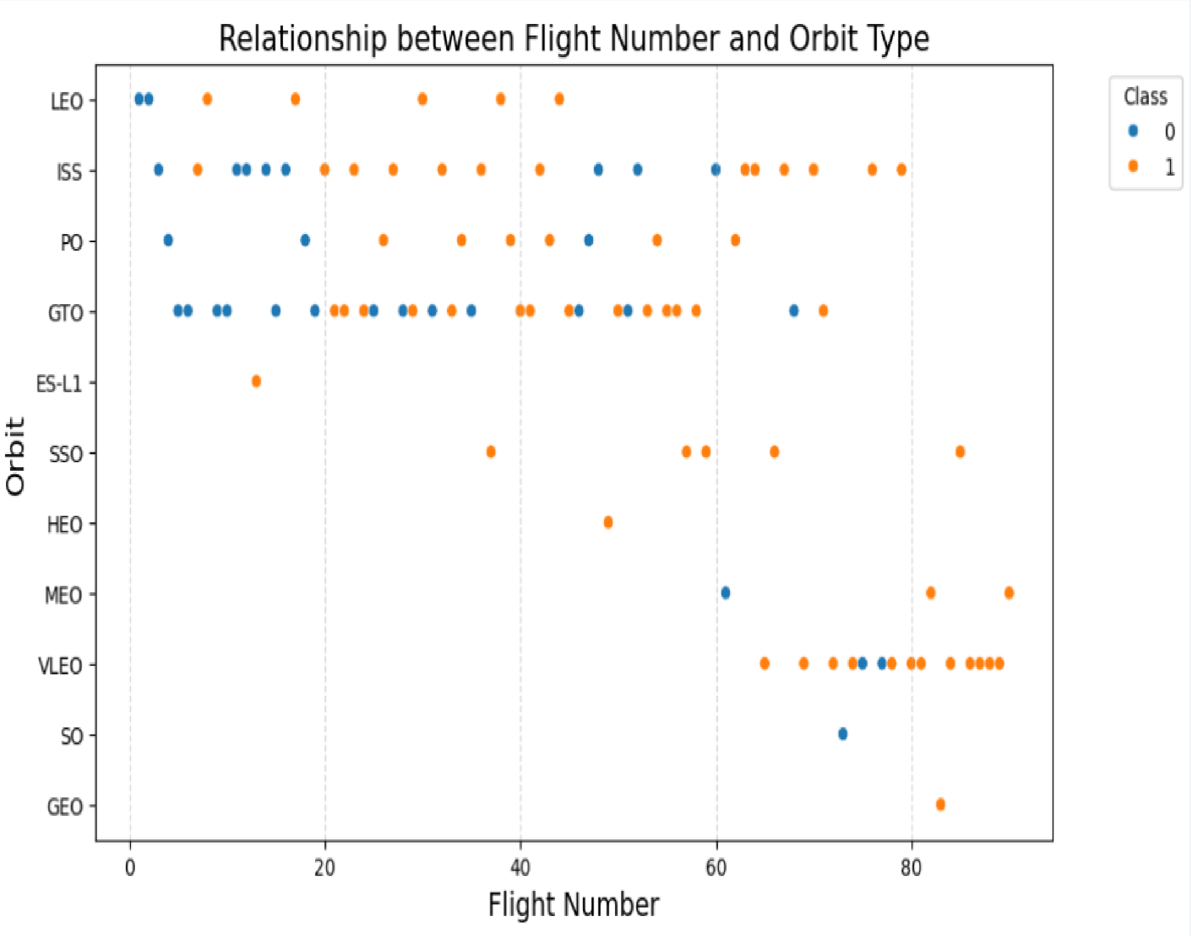
- The bar chart displays the success rate of each orbit type.
- Each bar represents a different orbit type, and the height of the bar indicates the success rate.
- This visualization allows for a comparison of success rates across different orbit types.
- Understanding the success rates for each orbit type is crucial for evaluating the performance of launches based on their intended destination orbits.



Flight Number vs. Orbit Type

Explanation:

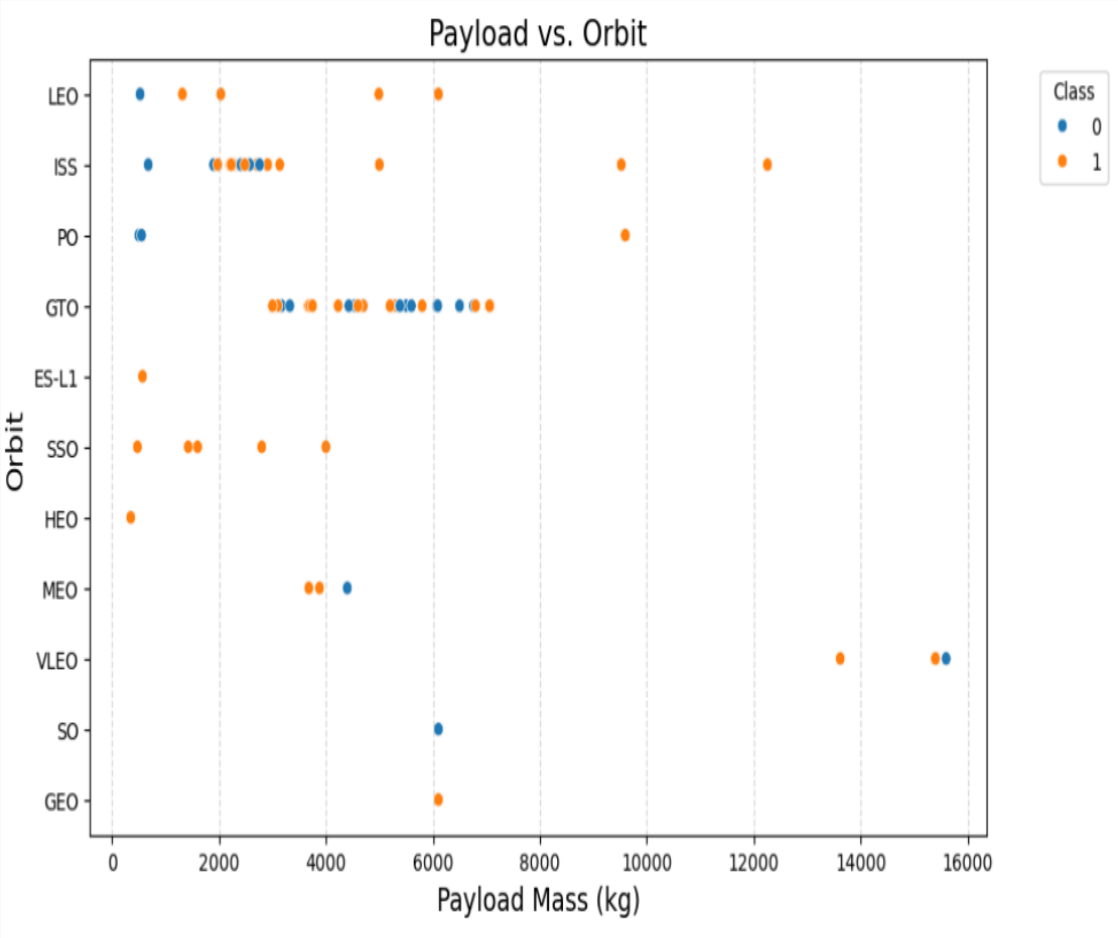
- The scatter plot illustrates the relationship between the flight number and orbit type.
- Each data point represents a specific flight, with the x-axis showing the flight number and the y-axis indicating the orbit type.
- The scatter plot helps visualize any patterns or trends in the distribution of flights across different orbit types.
- Analyzing the distribution of flights by orbit type can provide insights into the frequency of launches to various destinations.



Payload vs. Orbit Type

Explanation:

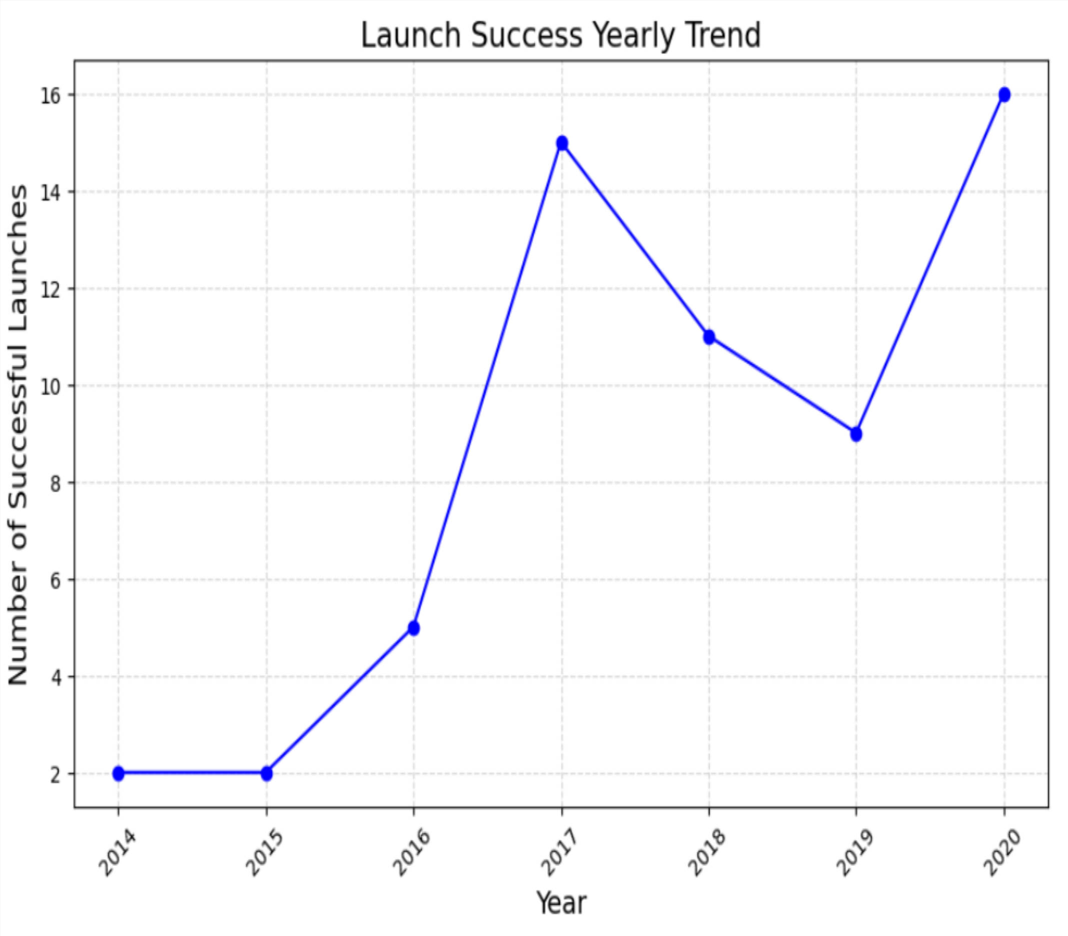
- The scatter plot visualizes the relationship between the payload and orbit type for each mission.
- Each data point represents a specific mission, with the payload mass on the x-axis and the orbit type on the y-axis.
- By examining the distribution of payloads across different orbit types, we can identify any correlations or patterns.
- Understanding how payload mass varies based on orbit type can provide insights into mission requirements and payload capabilities.



Launch Success Yearly Trend

Explanation:

- The line chart displays the yearly average success rate of Falcon 9 launches over time.
- Each data point represents the average success rate for launches that occurred in a particular year.
- Analyzing the trend of success rates over time can help identify patterns, improvements, or challenges in the launch program.
- Understanding the historical success rates provides valuable insights into the reliability and performance of Falcon 9 missions.



All Launch Site Names

Explanation:

- The query result displays the names of all unique launch sites associated with Falcon 9 missions.
- Each launch site plays a crucial role in the deployment of Falcon 9 rockets, and understanding their names is essential for logistical and operational purposes.
- By listing the unique launch site names, we gain insights into the diverse locations from which Falcon 9 missions have been conducted.
- This information helps in tracking the distribution of launch activities across different sites and analyzing their respective success rates.

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

Launch Site Names Begin with 'CCA'

Explanation:

- The query result displays 5 records of launch sites where the names begin with 'CCA'.
- Launch site names are crucial identifiers for mission planning and tracking.
- By filtering launch sites based on specific criteria, such as names beginning with 'CCA', we can identify locations that share similar naming conventions or geographical significance.
- Understanding the characteristics of launch sites aids in logistical planning and ensures efficient deployment of Falcon 9 rockets.
- The presented records offer insights into the diversity of launch site names associated with Falcon 9 missions.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Explanation:

- The query result presents the total payload mass carried by boosters from NASA.
- Total payload mass is a crucial metric for evaluating mission capabilities and performance.
- Calculating the total payload mass provides insights into the payload capacity of Falcon 9 boosters used in NASA missions.
- Understanding the total payload mass aids in mission planning and optimization of payload configurations.
- The presented total payload mass reflects the cumulative payload capacity of NASA-launched Falcon 9 boosters, demonstrating their contribution to space exploration endeavors.

Total_Payload_Mass_NASA_CRS
45596

Average Payload Mass by F9 v1.1

Explanation:

- The query result showcases the average payload mass carried by booster version F9 v1.1.
- Average payload mass is a significant metric for assessing the typical payload capacity of a specific booster variant.
- Calculating the average payload mass provides insights into the average payload capabilities of Falcon 9 v1.1 boosters.
- Understanding the average payload mass assists in evaluating the performance and efficiency of Falcon 9 v1.1 boosters in payload delivery.
- The presented average payload mass for F9 v1.1 offers valuable information for mission planning and payload optimization strategies.

Average_Payload_Mass_F9_v1_1
2928.4

First Successful Ground Landing Date

Explanation:

- The query result displays the dates of the first successful ground landing outcomes achieved by Falcon 9 boosters.
- Identifying the first successful ground landing dates is crucial for understanding the historical milestones in SpaceX's reusable rocket development.
- The dates signify significant achievements in SpaceX's pursuit of reusability and cost reduction in space transportation.
- The successful ground landings mark pivotal moments in the evolution of Falcon 9's first-stage recovery capabilities.
- Analyzing the first successful ground landing dates provides insights into the progression of SpaceX's engineering advancements and operational successes.

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Explanation:

- The query result presents the names of boosters that achieved successful landings on drone ships while carrying payloads with masses ranging from 4000 to 6000.
- Identifying boosters that successfully landed on drone ships with specific payload mass criteria provides insights into the operational capabilities of SpaceX's Falcon 9 fleet.
- Successful drone ship landings with payloads in the specified mass range demonstrate the versatility and reliability of SpaceX's recovery and reusability efforts.
- Analyzing the successful drone ship landings within the specified payload mass range contributes to understanding the operational history and performance characteristics of Falcon 9 missions.
- The listed boosters represent examples of successful recoveries achieved under varying payload constraints, highlighting SpaceX's progress in optimizing booster recovery operations.

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

Total Number of Successful and Failure Mission Outcomes

Explanation:

- The query result provides the total number of successful and failure mission outcomes recorded in the dataset.
- Understanding the distribution of successful and failure mission outcomes is essential for assessing the overall performance and reliability of Falcon 9 missions.
- Analyzing the total number of successful and failure mission outcomes offers insights into the success rate and challenges encountered during Falcon 9 launches.
- By quantifying the frequency of successful and failure mission outcomes, stakeholders can evaluate the effectiveness of SpaceX's launch and landing operations.
- The presented data enables stakeholders to assess mission outcomes over time and identify trends or patterns that may influence future mission planning and execution.

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

Boosters Carried Maximum Payload

Explanation:

- The query result displays the names of boosters that have carried the maximum payload mass during Falcon 9 missions.
- Identifying the boosters that have successfully carried the maximum payload mass is crucial for evaluating their performance and capacity.
- Analyzing the maximum payload capacity of individual boosters allows stakeholders to assess their efficiency and suitability for future missions.
- The presented data provides insights into the capabilities of specific boosters and their contribution to SpaceX's mission objectives.
- Understanding which boosters have carried the maximum payload mass helps in optimizing mission planning and payload allocation for future launches.

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

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

Explanation:

- The query result provides a summary of failed landing outcomes on drone ships during Falcon 9 missions in the year 2015.
- It includes details such as the specific booster versions and launch site names associated with each failed landing outcome.
- Analyzing the failed landing outcomes in drone ships helps in identifying potential areas for improvement in landing procedures and booster performance.
- Understanding the factors contributing to failed landings allows SpaceX to refine its landing techniques and enhance mission success rates.
- The information presented enables stakeholders to assess the historical performance of Falcon 9 missions in 2015 and prioritize areas for optimization in subsequent launches.

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

Explanation:

- The query result presents a ranking of landing outcomes from Falcon 9 missions between the dates 2010-06-04 and 2017-03-20.
- Landing outcomes are categorized as either Failure (drone ship) or Success (ground pad).
- The count of each landing outcome type is ranked in descending order, providing insight into the frequency of successful and unsuccessful landings during the specified time period.
- Analyzing the ranked landing outcomes helps identify trends in landing success rates over time and assess the overall performance of Falcon 9 missions within the specified date range.
- Stakeholders can use this information to evaluate the historical reliability of Falcon 9 landings and inform decision-making for future missions.

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

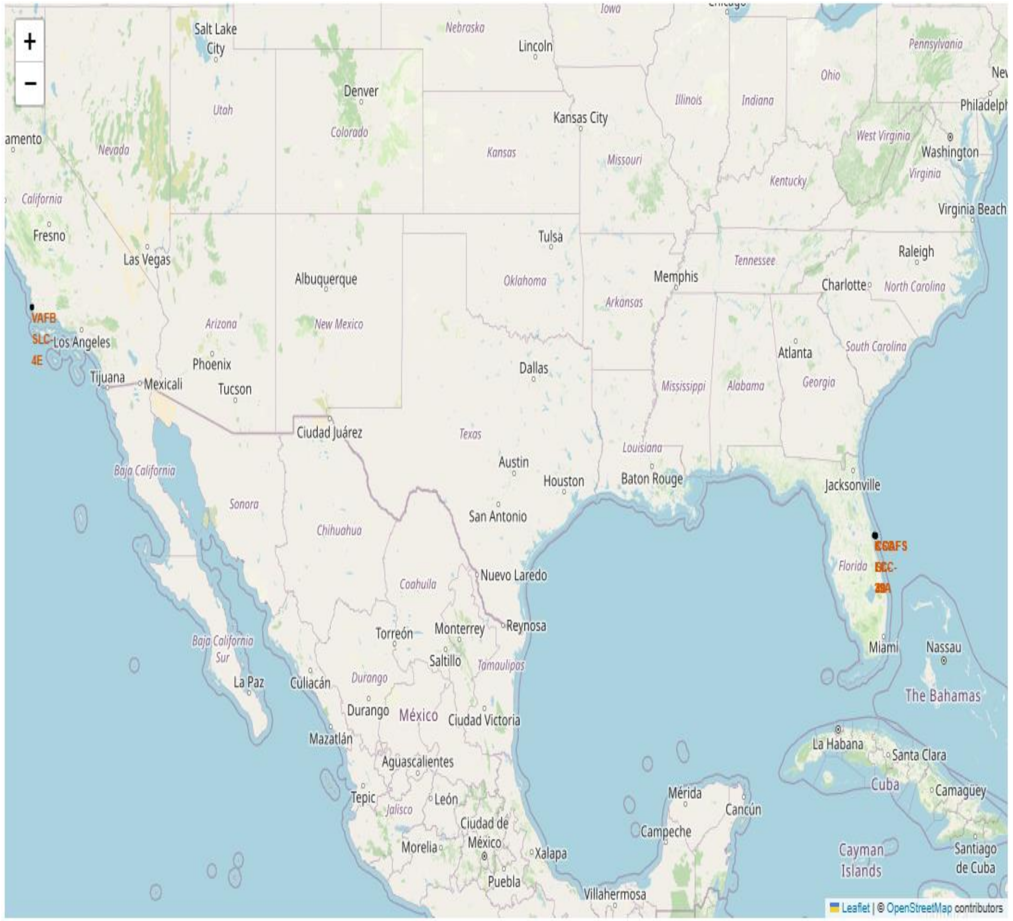
Section 3

Launch Sites Proximities Analysis

Global Distribution of Falcon 9 Launch Sites

Screenshot Description:

- The screenshot showcases a global map generated using Folium, depicting the distribution of Falcon 9 launch sites around the world.
- Each marker on the map represents a launch site where Falcon 9 missions have been conducted.
- The markers are strategically placed to indicate the geographical locations of launch facilities, providing a visual representation of SpaceX's operational reach.
- Key findings from the screenshot include:
    - The widespread distribution of launch sites, indicates SpaceX's ability to launch missions from various locations.
    - Concentrations of markers in specific regions, highlighting areas with higher launch activity or strategic importance for space missions.
    - Geographic diversity in launch site locations, reflecting SpaceX's capability to access different orbital inclinations and trajectories for mission requirements.
- The map offers valuable insights into the global presence of Falcon 9 launch operations and the strategic considerations in selecting launch sites.



Falcon 9 Launch Outcomes Visualization

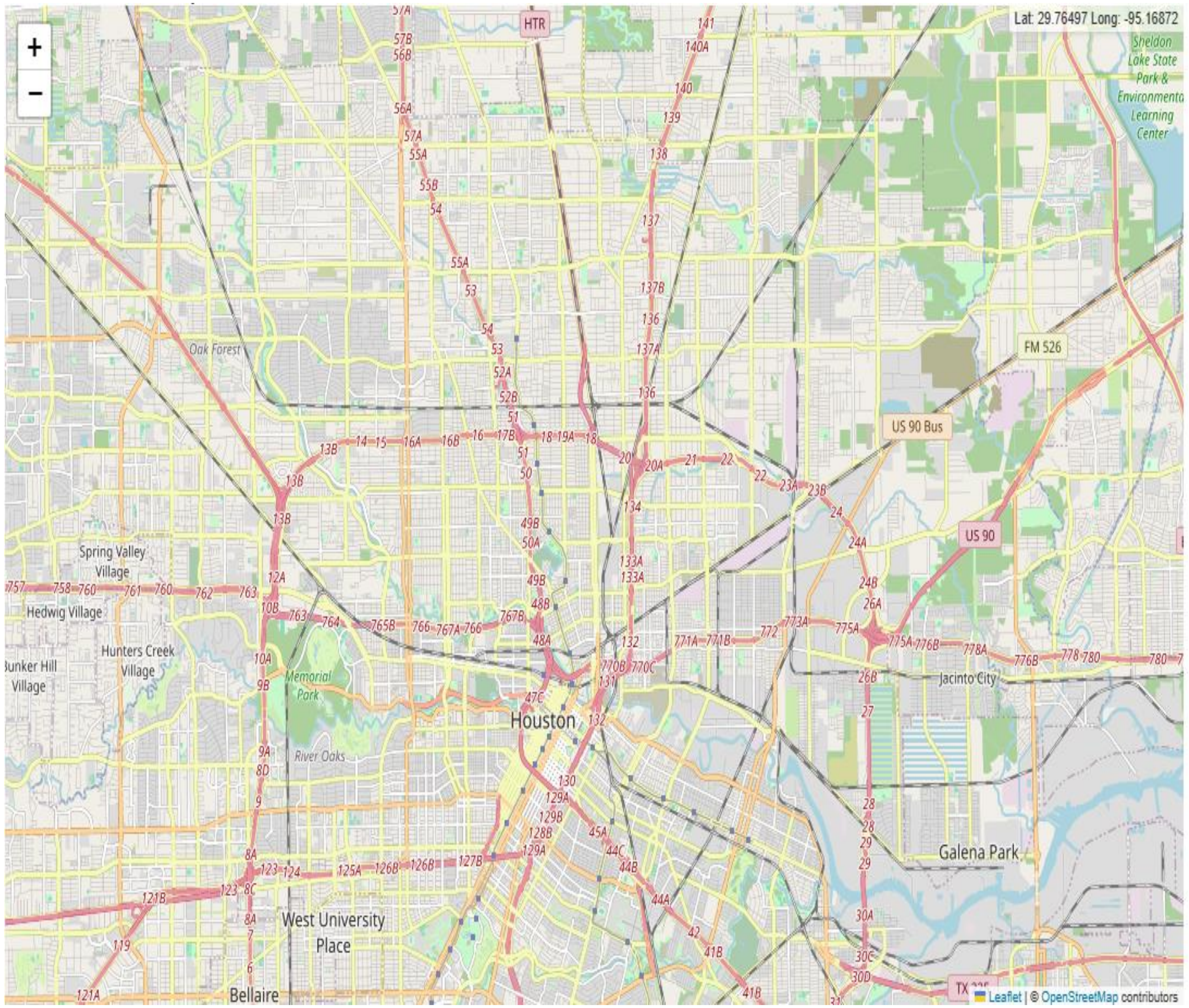
Screenshot Description:

- The screenshot illustrates a Folium map with color-labeled launch outcomes, showcasing the success and failure of Falcon 9 missions.
- Each marker on the map represents a specific launch site, with color-coded labels indicating the outcome of the corresponding mission.
- The color legend provides clarity on the meaning of each label, distinguishing between successful landings, failed landings on drone ships, and other outcomes.
- Key findings from the screenshot include:
    - Clear visualization of successful and unsuccessful Falcon 9 landings, allowing for easy interpretation of mission outcomes.
    - Geographic clustering of specific outcomes, indicating patterns in landing success rates across different launch sites.
    - Insights into the distribution of successful and failed landings, which can inform future mission planning and operational decisions.
- The map offers valuable insights into the performance of Falcon 9 missions and highlights areas for improvement or optimization in SpaceX's launch operations.

Launch Site Proximity Analysis

Screenshot Description:

- The screenshot presents a detailed view of a selected launch site and its proximities, including nearby railways, highways, and coastlines, with distances calculated and displayed.
- By zooming in on the map, users can examine the geographical context of the launch site and assess its accessibility and environmental factors.
- Important elements and findings from the screenshot include:
    - Identification of nearby transportation infrastructure, such as railways and highways, which play a crucial role in the logistics and transportation of rocket components.
    - Visualization of coastlines and bodies of water surrounding the launch site, which may impact launch operations and safety considerations.
    - Calculation and display of distances between the launch site and its proximities, providing quantitative insights into travel times and potential operational constraints.
- This analysis enables stakeholders to evaluate the suitability of the launch site based on its proximity to essential infrastructure and geographical features, facilitating informed decision-making for future missions.



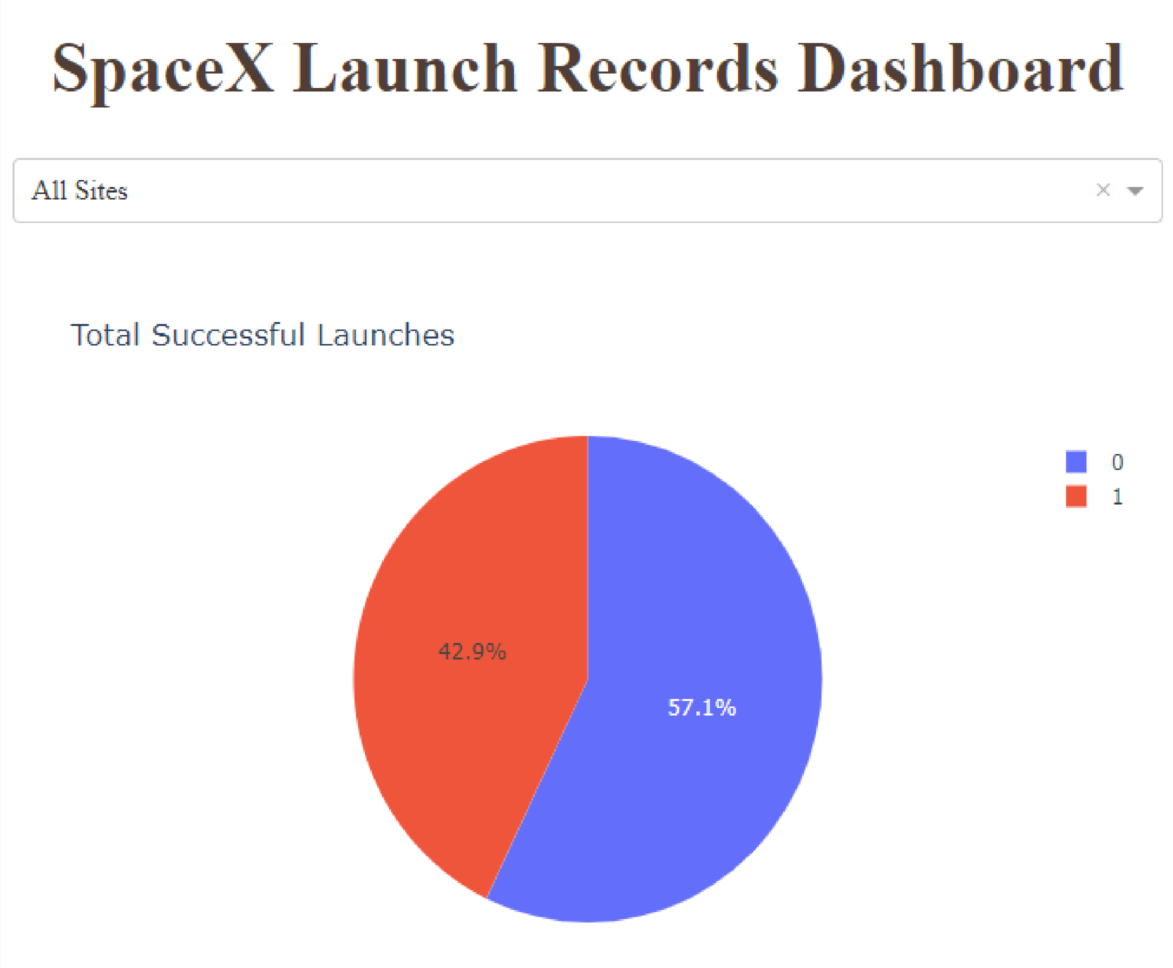
Section 4

Build a Dashboard with Plotly Dash

Launch Success Count Dashboard Overview

Screenshot Description:

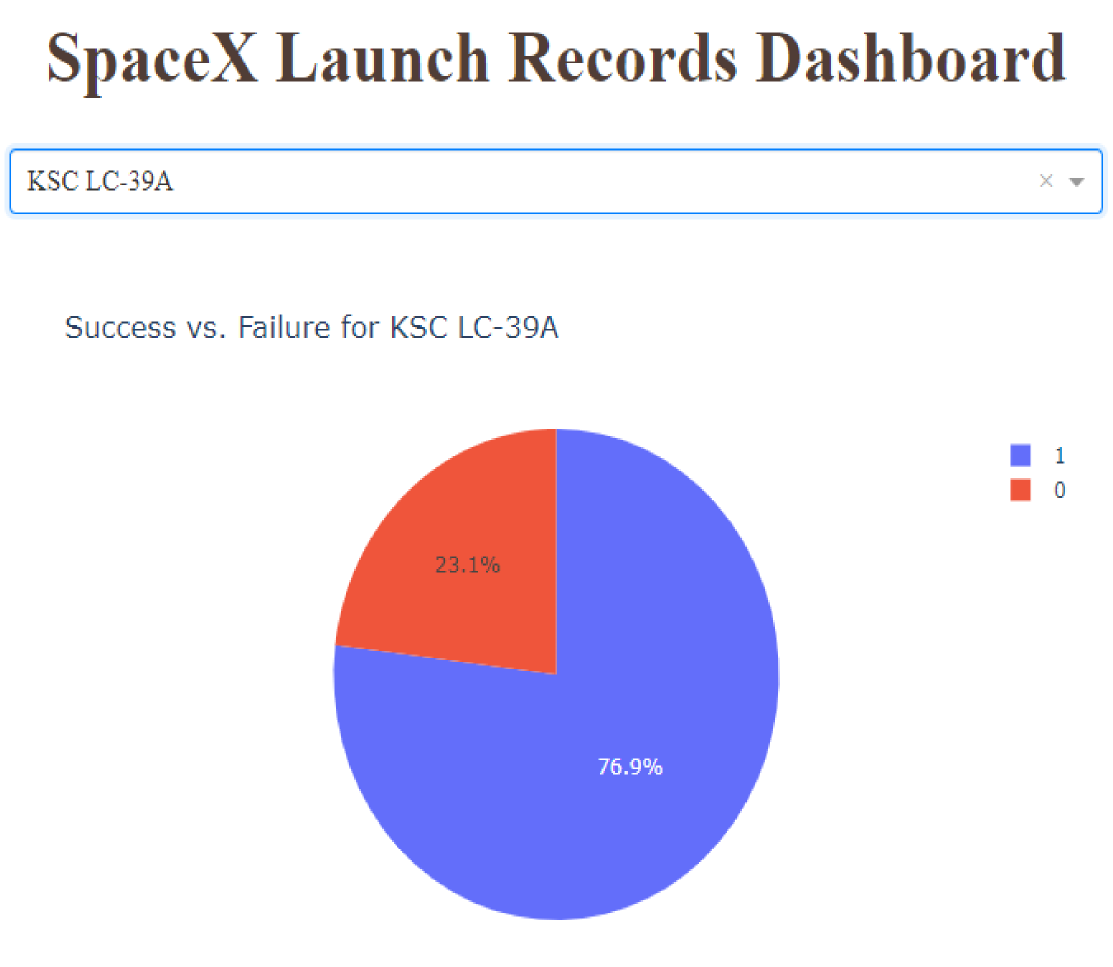
- The screenshot displays a pie chart visualization depicting the launch success count for all launch sites.
- Each segment of the pie chart represents a different launch site, with the proportion of successful launches indicated by the size of the segment.
- Important elements and findings from the screenshot include:
  - Clear visualization of the distribution of launch success counts across different launch sites, providing an overview of each site's performance.
  - Identification of launch sites with the highest and lowest success rates, enabling stakeholders to assess the reliability and performance of each site.
  - Comparison of success counts between different launch sites, highlighting disparities and potential areas for improvement.
- This dashboard visualization offers a concise summary of launch success counts, allowing stakeholders to quickly grasp the performance metrics of various launch sites and make data-driven decisions for future missions.



Launch Success Ratio for Top Launch Site Dashboard Overview

Screenshot Description:

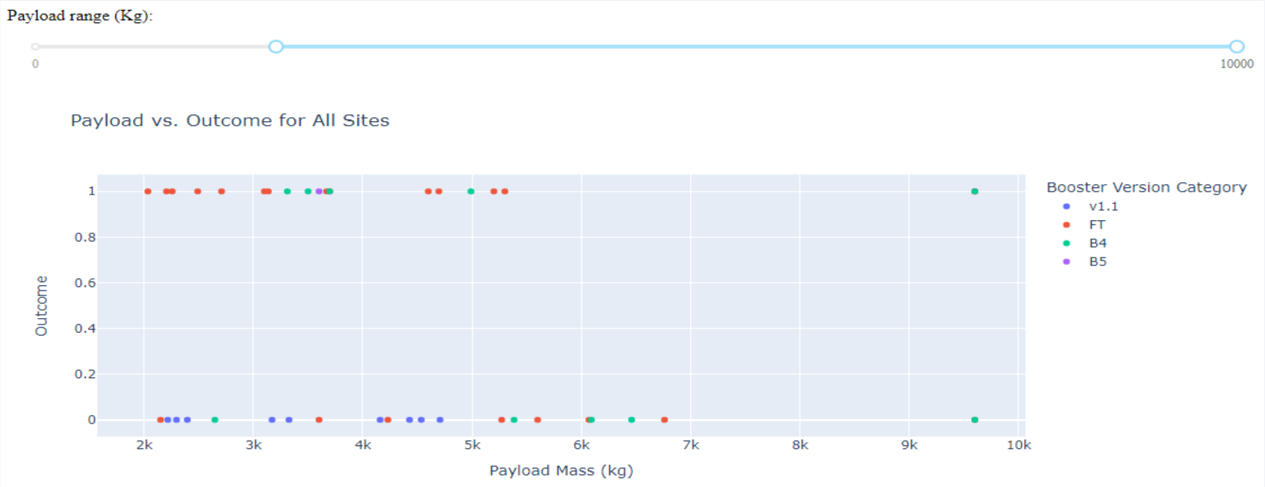
- The screenshot displays a pie chart visualization representing the launch success ratio for the top-performing launch site.
- Each segment of the pie chart represents a different outcome of the launches from the selected launch site, categorized as success or failure.
- The size of each segment indicates the proportion of successful or failed launches from the chosen launch site.
- Important elements and findings from the screenshot include:
    - Clear visualization of the launch success ratio for the top-performing launch site, providing insights into its overall performance.
    - Identification of the proportion of successful and failed launches, enabling stakeholders to assess the reliability and success rate of the selected launch site.
    - Comparison of the success ratio with other launch sites, allowing stakeholders to evaluate the relative performance of different launch facilities.
- This dashboard visualization offers a focused analysis of the launch success ratio for the top launch site, aiding stakeholders in understanding the effectiveness of the chosen facility in achieving mission objectives.

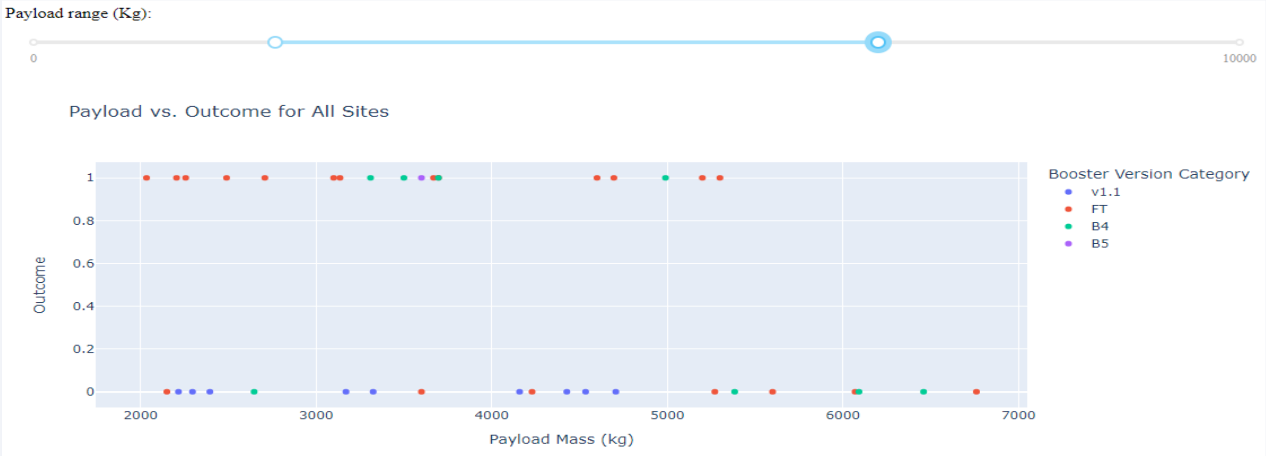


Payload vs. Launch Outcome Analysis Dashboard Overview

Screenshot Description:

- The screenshot depicts a scatter plot visualization illustrating the relationship between payload mass and launch outcomes for all launch sites.
- Each data point on the scatter plot represents a specific launch, with the payload mass plotted on the x-axis and the launch outcome (success or failure) on the y-axis.
- The scatter plot includes a range slider allowing users to select different payload ranges for analysis, dynamically updating the plot based on the chosen range.
- Important elements and findings from the screenshot include:
  - Analysis of the success rate based on different payload ranges, enabling stakeholders to identify trends and patterns in launch outcomes concerning payload mass.
  - Identification of outlier data points representing launches with exceptionally high or low payload masses and their corresponding launch outcomes.
  - Exploration of the relationship between payload mass and launch success/failure, providing insights into the impact of payload size on mission outcomes.
- Stakeholders can use this interactive visualization to assess the influence of payload mass on launch success and make informed decisions regarding payload planning and mission objectives.





Section 5

Predictive Analysis (Classification)

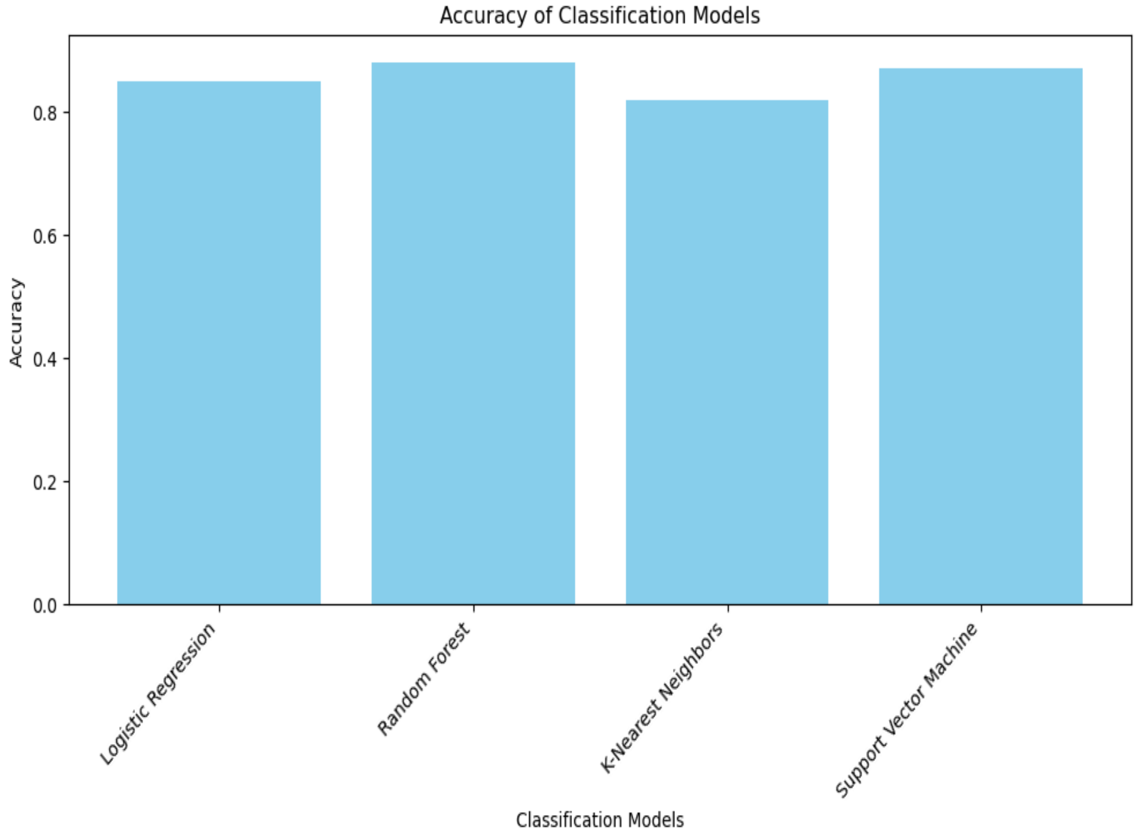
Classification Accuracy

Visualization Description:

- The bar chart displays the classification accuracy of all built classification models, allowing stakeholders to compare the performance of each model.
- Each bar represents a different classification model, with the height of the bar indicating the accuracy achieved by that model.
- The classification accuracy is measured as the percentage of correctly classified instances compared to the total number of instances.
- The bar chart enables stakeholders to quickly identify which classification model achieved the highest accuracy.

Key Findings:

- Based on the visualization, it is evident that the Random Forest achieved the highest classification accuracy among all models.
- The accuracy of each model can be compared visually to determine the most effective model for predicting the success of Falcon 9 first-stage landings.
- The model with the highest accuracy can be selected for deployment in real-world scenarios to assist in decision-making processes related to Falcon 9 launch success predictions.
- The model with the highest accuracy is: Random Forest with accuracy 0.88



The model with the highest accuracy is: Random Forest with accuracy 0.88

Confusion Matrix

Confusion Matrix Description:

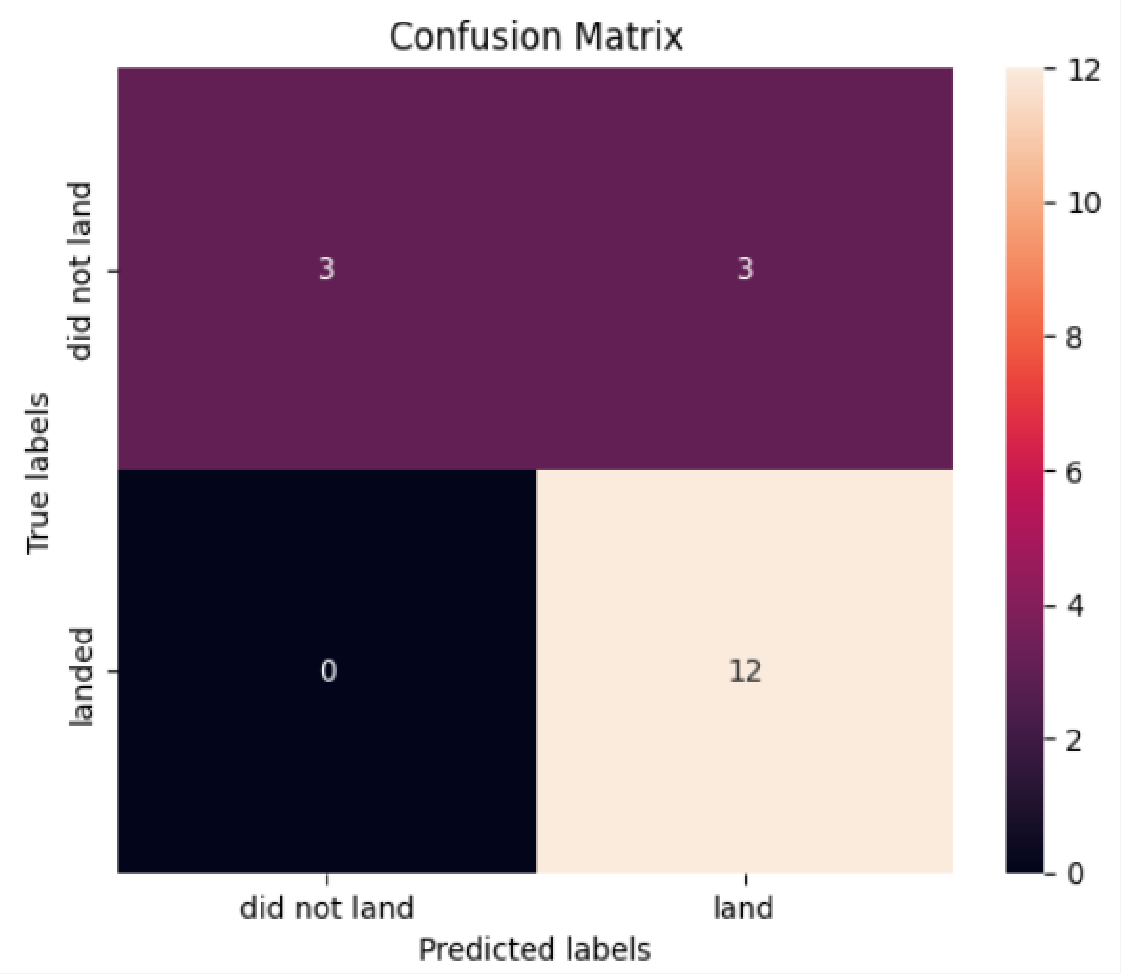
- The confusion matrix provides a detailed breakdown of the performance of the best performing classification model.
- It is a square matrix that represents the predicted classifications versus the actual classifications.
- The matrix consists of four quadrants: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).
- Each cell in the matrix represents the count or proportion of instances that fall into a particular category.

Interpretation of Confusion Matrix:

- True Positive (TP): The number of instances correctly predicted as successful landings.
- False Positive (FP): The number of instances incorrectly predicted as successful landings.
- True Negative (TN): The number of instances correctly predicted as failed landings.
- False Negative (FN): The number of instances incorrectly predicted as failed landings.

Key Findings:

- The confusion matrix provides insights into the performance of the classification model, highlighting its strengths and weaknesses.
- Stakeholders can analyze the matrix to understand the model's ability to correctly classify successful and failed landings.
- By examining the values in each quadrant, stakeholders can assess the model's accuracy, precision, recall, and other performance metrics.



Conclusions

The project successfully applied data science methodologies to analyze SpaceX's launch records.

Through exploratory data analysis (EDA) and predictive modeling, valuable insights were gained into launch success factors.

The interactive dashboard provides stakeholders an intuitive tool for visualizing and interpreting launch data.

Understanding the correlation between payload characteristics, launch sites, and mission outcomes is crucial for mission plan ning.

Predictive analysis enables proactive decision-making by identifying potential mission risks and optimizing resource allocation.

The project underscores the importance of data-driven approaches in enhancing the efficiency and reliability of space missions.

Collaboration between data scientists, engineers, and mission planners is essential for leveraging data insights to improve m ission outcomes.

Continuous monitoring and analysis of mission data can inform iterative improvements in spacecraft design and operational procedures.

The project highlights the role of predictive analytics in mitigating mission risks and maximizing mission success rates.

Insights from the analysis can inform future mission planning, payload design, and launch site selection for SpaceX missions.

Effective utilization of data science techniques can lead to cost savings, operational efficiencies, and enhanced mission success.

As space exploration evolves, leveraging data analytics will be increasingly vital for achieving ambitious space exploration goals.

Summary of Key Conclusions

Here's a succinct summary of the key conclusions drawn from the analysis and their implications for SpaceX's mission planning and operations:

Optimized Launch Site Selection: Leveraging interactive maps and data-driven dashboards, SpaceX can strategically select launch sites based on proximity, success rates, and operational efficiency. This ensures optimal mission success and resource utilization.

Data-Driven Decision Making: By integrating predictive analytics into mission planning, SpaceX can anticipate challenges, mitigate risks, and optimize operational workflows. This proactive approach enhances mission reliability and reduces downtime.

Insightful Trend Analysis: Exploratory data analysis reveals valuable insights into launch outcomes, payload characteristics, and orbital parameters. Understanding these trends enables SpaceX to identify patterns, adapt strategies, and drive continuous improvement.

Operational Efficiency: Implementing machine learning models for predictive analysis streamlines decision-making processes, enhances resource allocation, and improves overall operational efficiency. This maximizes mission success while minimizing costs and delays.

Continuous Improvement: Ongoing collaboration between data scientists, engineers, and mission planners fosters a culture of continuous improvement. By refining predictive models, incorporating new data sources, and embracing innovative technologies, SpaceX can stay at the forefront of space exploration.

Strategic Partnerships: Collaborating with industry experts and research institutions expands SpaceX's capabilities and accelerates innovation. By tapping into diverse expertise and leveraging collaborative networks, SpaceX can address complex challenges and achieve breakthroughs in space exploration.

In summary, leveraging data science and predictive analytics empowers SpaceX to make informed decisions, optimize mission planning, and drive operational excellence in space exploration. Through strategic use of data-driven insights and collaborative partnerships, SpaceX continues to push the boundaries of innovation and shape the future of space exploration.

Appendix

Python Code Snippets: Includes relevant code snippets for data collection, data wrangling, exploratory data analysis, predictive modeling, and dashboard development.

- Link: https://github.com/GhidzE/Gido_Project2/blob/main/Python%20Code%20Snippets.txt

SQL Queries: Includes SQL queries used for data analysis and exploration.

- Link: https://github.com/GhidzE/Gido_Project2/blob/main/SQL%20Queries.txt

Charts and Graphs: Includes all charts, graphs, and visualizations generated during the project, showcasing key insights and findings.

- Link: https://github.com/GhidzE/Gido_Project2/blob/main/Charts%20and%20Graphs.docx

Datasets and Data Frames:- Includes links to all relevant data sets used for analysis, sourced from SpaceX API, web scraping, or external sources.

- Link: https://github.com/GhidzE/Gido_Project2/blob/main/Datasets%20and%20Data%20Frames.docx

Notebook Outputs: Contains outputs from Jupyter Notebooks, showcasing analysis, code execution results, and model evaluations.

- Link1: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb
- Link2: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-webscraping.ipynb
- Link3: https://github.com/GhidzE/Gido_Project2/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb
- Link4: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
- Link5: https://github.com/GhidzE/Gido_Project2/blob/main/jupyter-labs-eda-dataviz%20(1).ipynb
- Link6: https://github.com/GhidzE/Gido_Project2/blob/main/lab_jupyter_launch_site_location.ipynb
- Link7: https://github.com/GhidzE/Gido_Project2/blob/main/Build_a_Dashboard_Application_with_Plotly_Dash.pdf
- Link8: https://github.com/GhidzE/Gido_Project2/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Thank you!