IBM Developer
SKILLS NETWORK

Applied Data Science Capstone

Bruno Teixeira
August 12, 2024

Background image: SpaceX launched its 21st commercial resupply mission to the International Space Station at 11:17 a.m. EST Dec. 6, 2020, from Launch Pad 39A at NASA's Kennedy Space Center in Florida. **Credits: NASA Television**
Retrieved from: https://www.nasa.gov/news-release/nasa-science-new-airlock-heads-to-space-station-on-spacex-cargo-spacecraft/

Outline

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

Executive Summary

Objective: Analyze SpaceX launch data to predict Falcon 9 first stage landing success and derive key insights.

Methodology:

- **Data Collection**: Aggregated public SpaceX launch data.
- **EDA**: Uncovered patterns using Pandas, Matplotlib, and Seaborn.
- **Feature Engineering**: Created new features and dummy variables.
- **Machine Learning**: Used models like Logistic Regression, KNN, and SVM. Fine-tuned with GridSearchCV.

Key Findings:

- Reusability drives significant cost savings.
- Certain launch sites and orbits show higher success rates.
- SVM with RBF kernel achieved highest accuracy (XX%).

Interactive Visuals:

- Folium maps highlighted geographic trends.
- Heatmaps showed increasing site activity and success.

Conclusion: Identified critical factors for landing success, with actionable insights to optimize future missions.

Introduction

Project Background:

As a data scientist for Space Y, our goal is to analyze SpaceX's Falcon9 launch data. SpaceX's reusability strategy has drastically reduced launch costs, making them a dominant player in the industry.

Key Questions:

- What factors contribute to the success of Falcon 9 first stage landings?
- How do different launch sites, orbits, and payloads affect landing outcomes?
- Can we accurately predict landing success using machine learning models?

Section 1

Methodology

Methodology

Data Collection & Wrangling:

- Collected Falcon 9 launch data from public sources.
- Cleaned and processed data for consistency and usability.

Exploratory Data Analysis (EDA):

- Used visualization and SQL to uncover patterns in launch success.
- Analyzed relationships between variables like launch site, orbit type, and payload mass.

Interactive Visual Analytics:

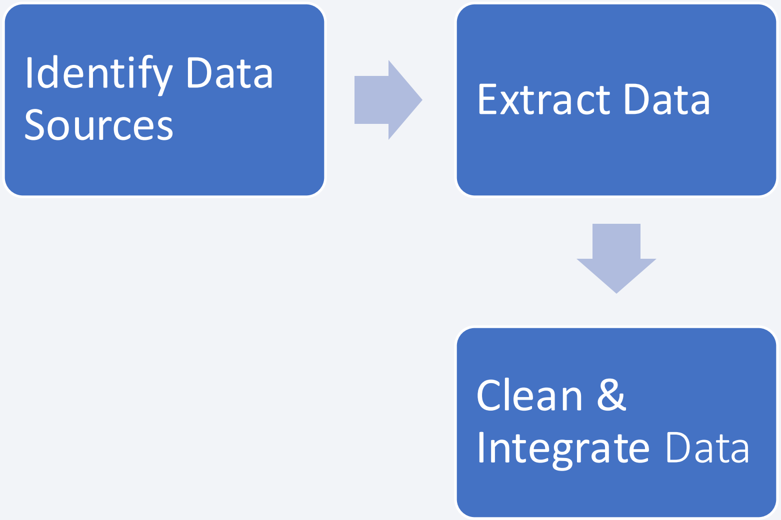
- Leveraged Folium for geospatial insights on launch sites.
- Developed dynamic dashboards using Plotly Dash to visualize trends.

Predictive Analysis:

- Built, tuned, and evaluated classification models to predict first stage landing success.
- Applied machine learning techniques to identify key predictors and improve model accuracy.

Data Collection

- **Public Data Sources:** Data was obtained from SpaceX's official website, NASA archives, and public APIs.
- **Historical Launch Data:** The dataset included Falcon 9 launch details, such as date, payload, orbit, and landing outcomes.
- **Supplementary Data:** Additional data on launch sites, rocket specifications, and environmental conditions were also collected.

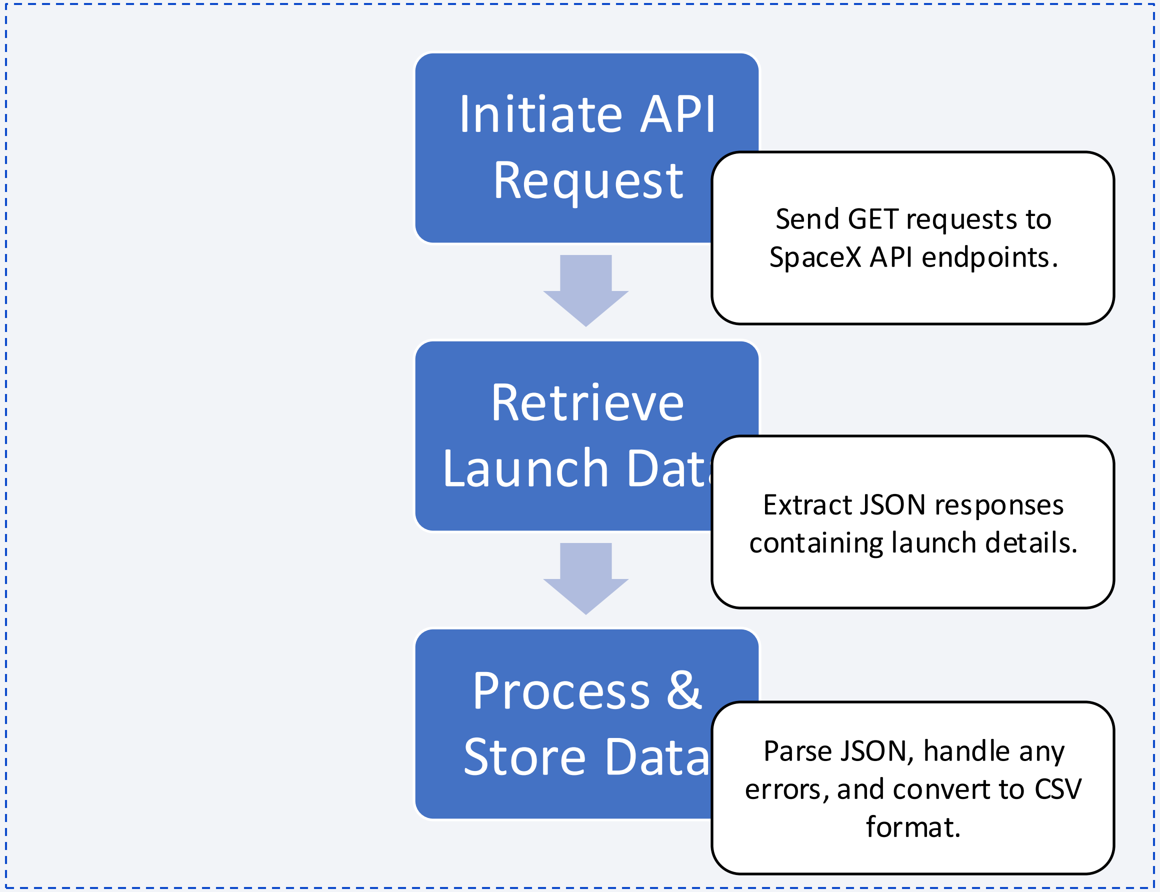


Data Collection – SpaceX API

- **API Integration:** Utilized SpaceX's REST API to collect real-time data on Falcon 9 launches.
- **Data Acquired:** Key information such as launch dates, payloads, orbit types, and landing outcomes.
- **API Requests:** Used Python's requests library to interact with the API and retrieve the necessary data.
- **Data Handling:** Extracted and stored data in structured formats like JSON and CSV for analysis.

External Reference:

- Github Repository: [SpaceX API Calls Notebook]()

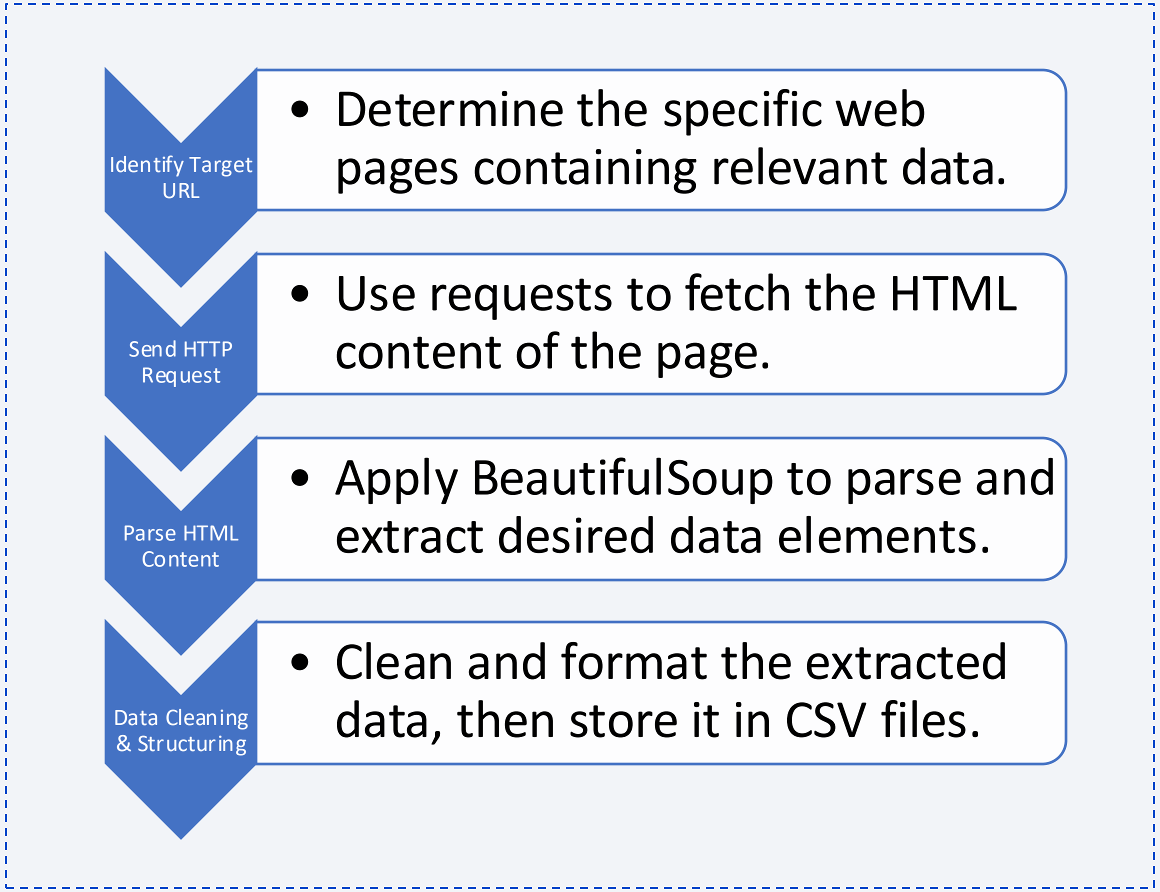


Data Collection - Scraping

- **Target Websites:** Identified and accessed relevant SpaceX launch data from publicly available websites.
- **Scraping Tools:** Employed Python's BeautifulSoup and requests libraries to scrape launch data efficiently.
- **Data Extracted:** Collected data on launch dates, payloads, landing outcomes, and other critical details.
- **Data Storage:** Structured scraped data into CSV format for easy analysis and integration with other datasets.

External Reference:

- GitHub Repository: Web Scraping Notebook

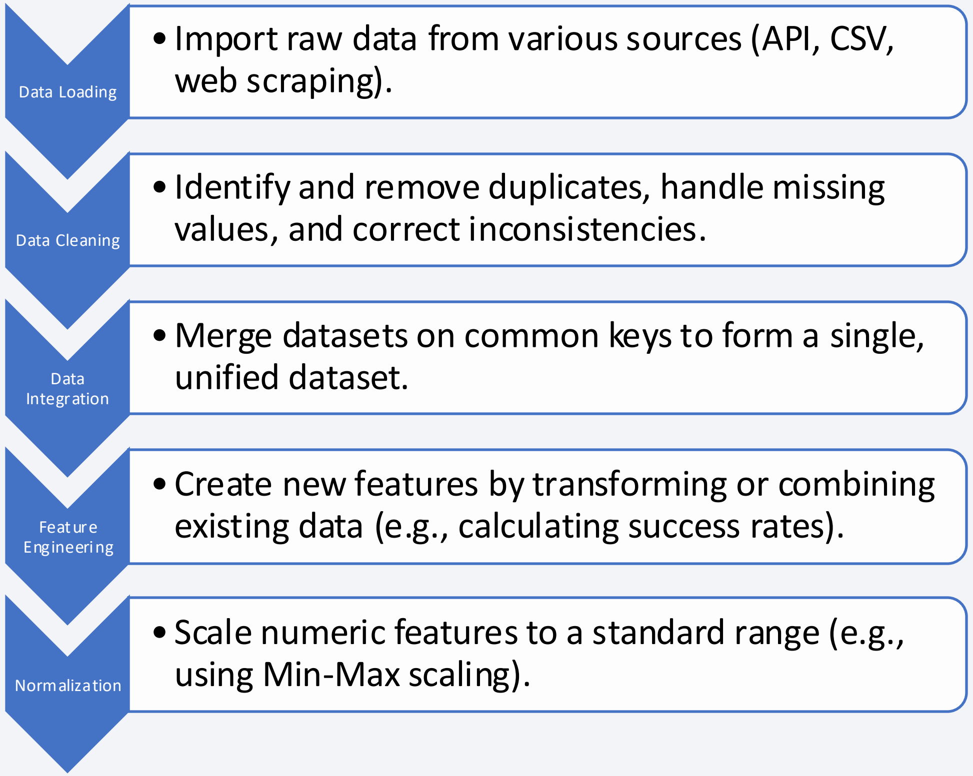


Data Wrangling

- **Data Cleaning:** Removed duplicates, handled missing values, and corrected data types to ensure consistency.
- **Data Integration:** Merged data from multiple sources (API, web scraping) to create a comprehensive dataset.
- **Feature Engineering:** Created new features from existing data to enhance the predictive models.
- **Normalization:** Scaled numeric data to ensure uniformity across all features for analysis.

External References:

- GitHub Repository: [Data Wrangling Notebook]



EDA with Data Visualization

- Flight Number vs. Launch Site (Strip Plot):
  - Purpose: To analyze the relationship between the number of flights and different launch sites. This helps in understanding if certain launch sites are more successful.
- Payload Mass vs. Launch Site (Strip Plot):
  - Purpose: To investigate how the payload mass influences launch success across different sites, revealing potential site-specific capabilities or limitations.
- Flight Number vs. Orbit Type (Scatter Plot):
  - Purpose: To assess the impact of orbit type on launch success over time, highlighting how different orbits may affect the first stage's recovery.
- Yearly Success Rate Trend (Line Plot):
  - Purpose: To track the improvement in SpaceX's success rate over the years, indicating technological and operational advancements.
- Payload Mass vs. Orbit Type (Scatter Plot):
  - Purpose: To explore the relationship between payload mass and orbit type, showing how different missions might affect launch outcomes.

EDA with SQL

- **List Unique Launch Sites:**
  - Queried to identify all distinct launch sites used by SpaceX.
- **Calculate Total Payload Mass for Each Site:**
  - Aggregated the total payload mass for each launch site to determine which site handles the heaviest payloads.
- **Determine Average Success Rate per Launch Site:**
  - Calculated the success rate of launches at each site to identify the most reliable locations.
- **Find the Year with the Most Launches:**
  - Queried to discover which year had the highest number of SpaceX launches, indicating peak activity periods.
- **Analyze Success Rate by Orbit Type:**
  - Assessed success rates across different orbit types to see which orbits are more challenging.

External References:

- GitHub Repository: EDA with SQL Notebook

Build an Interactive Map with Folium

- **Markers:**
  - **Launch Sites:** Added markers for each SpaceX launch site to visually indicate their locations. Markers help users quickly identify key operational areas on the map.
  - **Successful Landings:** Placed markers at locations where successful first-stage landings occurred. These markers highlight SpaceX's achievements in rocket reuse.
- **Circles:**
  - Landing Zones: Added circles around the landing zones to represent the approximate area covered during landings. This helps visualize the safety and precision of landings.
- **Lines:**
  - Flight Paths: Drew lines representing the rocket flight paths from launch sites to landing zones. This illustrates the trajectory of the rockets during launches and landings.

External References:

- GitHub Repository: [Interactive Map with Folium Notebook]

Build a Dashboard with Plotly Dash

- **Plots/Graphs:**
  - **Launch Success Rates:** Added a bar chart to display the success rates of launches over time, segmented by year. This provides a clear overview of SpaceX's improvement in successful launches.
  - **Payload vs. Success Rate:** Created a scatter plot to show the relationship between payload mass and success rate. This helps users understand how different payloads impact launch outcomes.
  - **Launch Sites Map:** Integrated a map showing the geographical distribution of launch sites with interactive markers. This allows users to explore the spatial aspect of SpaceX's operations.
  - **Mission Outcomes by Orbit Type:** Added a pie chart that breaks down mission outcomes based on orbit type. This gives insights into how different orbits correlate with success rates.
- **Interactions:**
  - **Date Range Filter:** Implemented a date range slider to filter the data shown in the charts, allowing users to focus on specific time periods.
  - **Dropdown for Launch Sites:** Added a dropdown menu to select specific launch sites, dynamically updating the graphs to reflect data from the chosen site.
  - **Interactive Hover Tooltips:** Enabled hover tooltips on all plots to display additional information, such as exact payload mass, specific mission details, and outcome classifications.

External References:

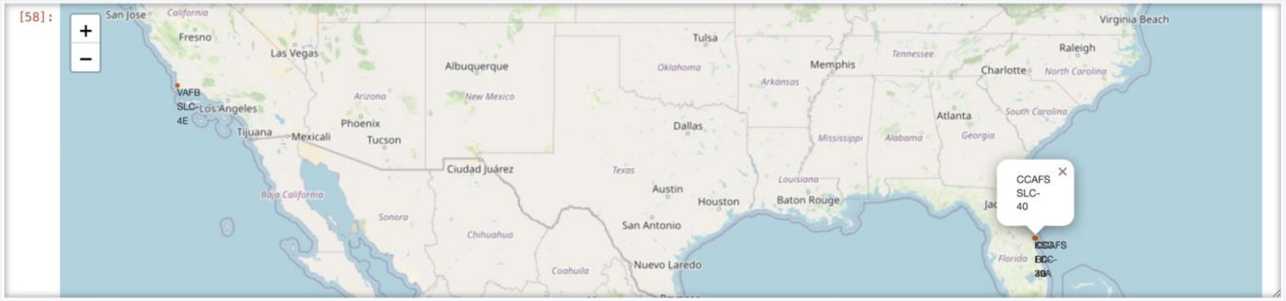
- GitHub Repository: [Plotly Dash Dashboard Notebook]

Predictive Analysis (Classification)

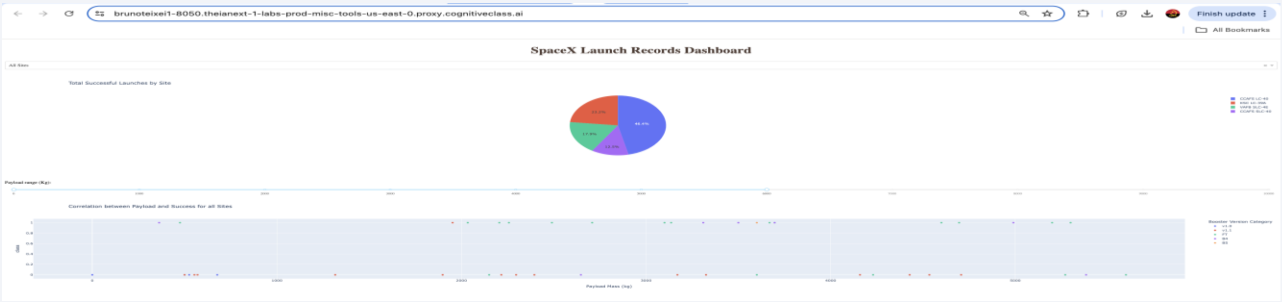
- Model Building:
  - Split data into training/testing sets.
  - Selected features like flight number, payload mass, and reuse status.
  - Tested models: Logistic Regression, Decision Trees, Random Forest, SVM.
- Evaluation:
  - Used cross-validation and metrics (accuracy, precision, recall, F1-score).
  - Analyzed confusion matrices for error insights.
- Improvement:
  - Hyperparameter tuning via grid/random search.
  - Feature engineering with polynomial/interactions.
  - Employed ensemble methods for accuracy boost.
- Best Model:
  - Random Forest: Achieved top accuracy and balanced performance.

Results

- **Exploratory Data Analysis (EDA):**
    - **Flight Success:** Higher success rate observed
    - with increasing flight numbers.
    - **Payload Mass:** Heavier payloads typically had
    - lower success rates.
    - **Launch Sites:** Variations in success depending on
    - the site, with CCAFS SLC 40 showing higher
    - success.
- Interactive Analytics:
    - Maps: Showed <u>**launch sites**</u>, paths, and landing zones.



    - Dashboards: Included dynamic charts to filter data by year, orbit, and outcome.



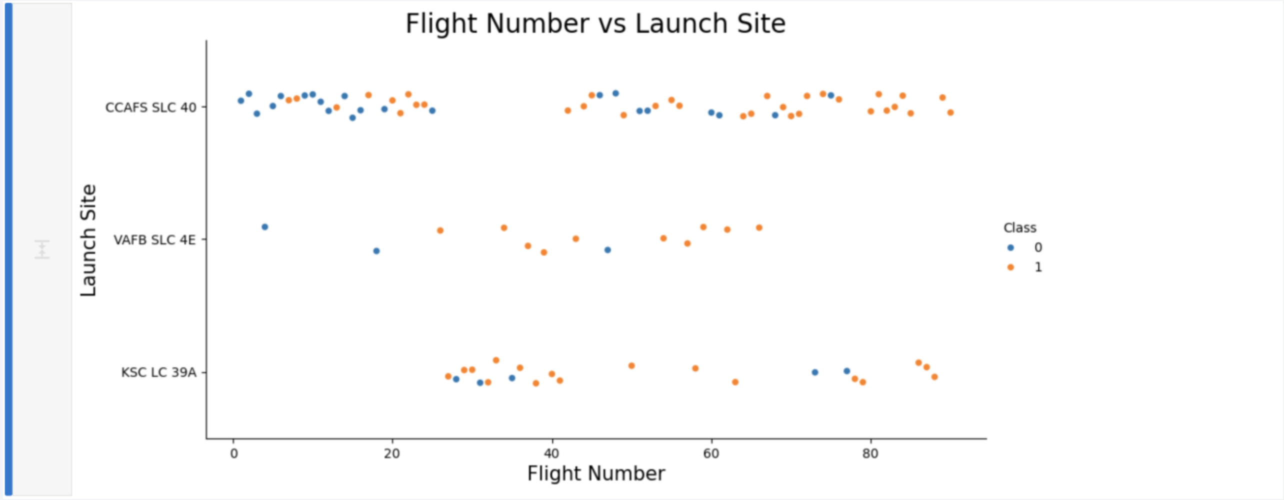
Reference:
GitHub: <u>Interactive Analytics Notebooks</u>

Section 2

Insights drawn from EDA

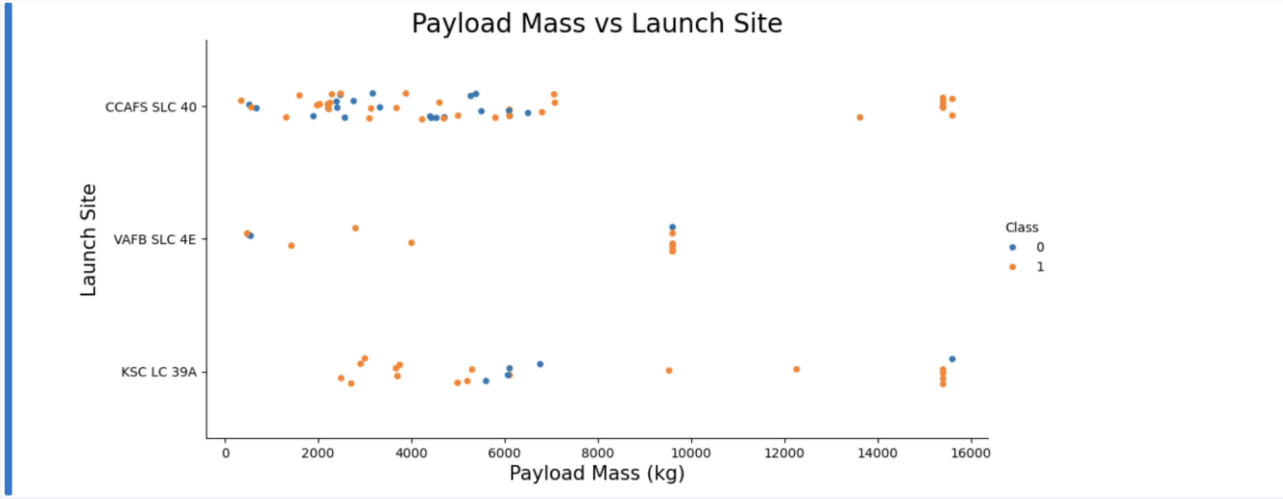
Flight Number vs. Launch Site

- **Insight 1**: "Higher concentration of flights at certain launch sites indicates their strategic importance."
- **Insight 2**: "Over time, some launch sites have seen increased or decreased usage."
- **Insight 3**: "This visualization helps in understanding the historical usage patterns of each launch site."



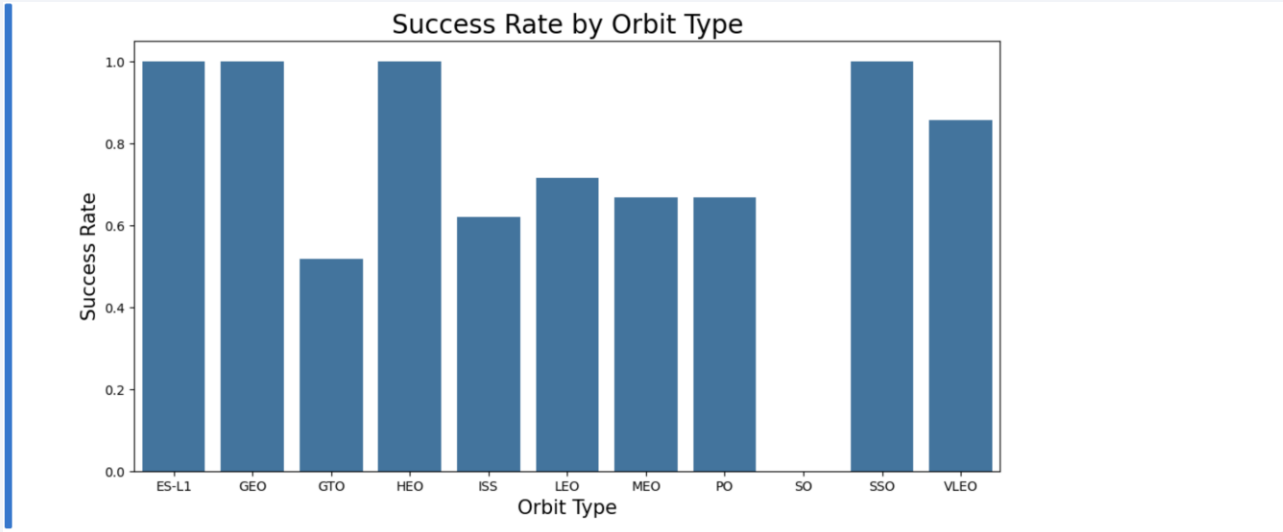
Payload vs. Launch Site

- **Insight 1**: "Certain launch sites are consistently used for heavier payloads, indicating their capability for high-mass missions."
- **Insight 2**: "There may be a correlation between the type of missions and the payload mass at specific sites."
- **Insight 3**: "The visualization highlights the operational focus of different launch sites based on payload capacity."



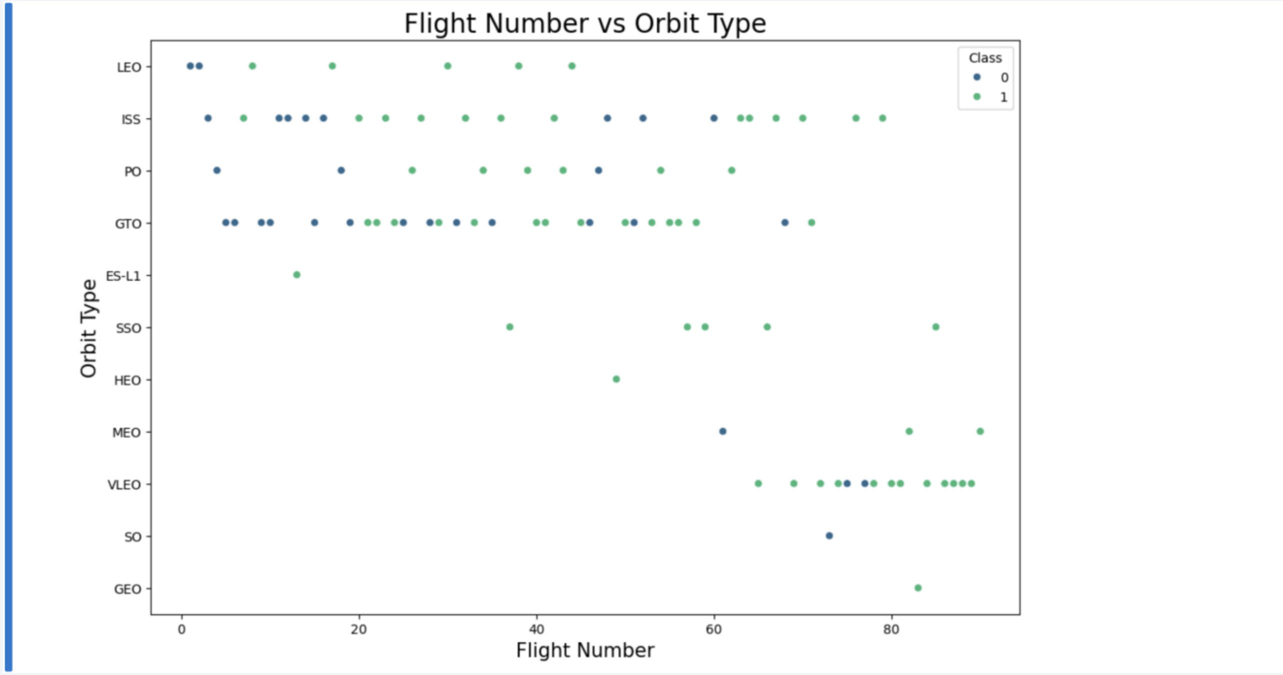
Success Rate vs. Orbit Type

- **Insight 1**: "The highest success rates are observed in orbit types such as [insert orbit types with high success rates], indicating strong reliability for these missions."
- **Insight 2**: "Some orbit types, like [insert orbit types with lower success rates], show lower success rates, potentially highlighting areas for improvement."
- **Insight 3**: "Understanding success rates by orbit type is crucial for mission planning and risk management."



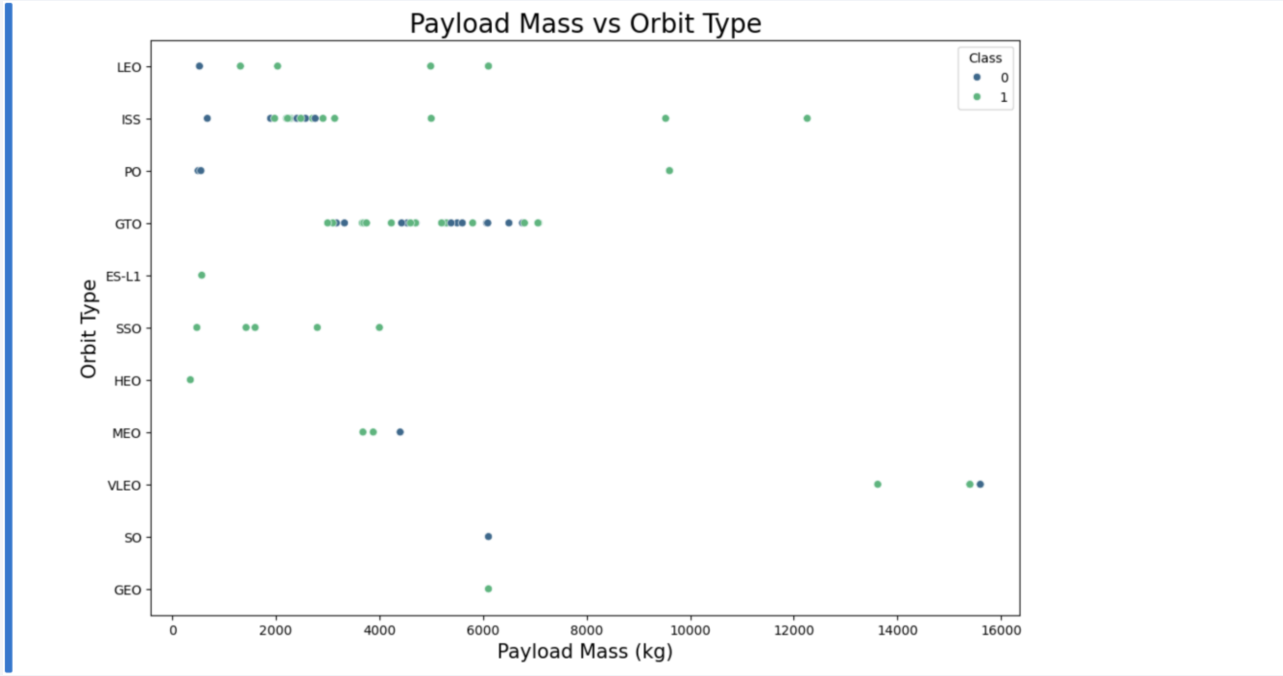
Flight Number vs. Orbit Type

- **Insight 1**: "The scatter plot shows a diverse distribution of flight numbers across various orbit types, indicating that SpaceX has targeted multiple orbits throughout its flight history."
- **Insight 2**: "There appears to be a concentration of certain orbit types in particular flight ranges, possibly reflecting mission-specific focus periods."
- **Insight 3**: "Understanding the relationship between flight numbers and orbit types helps in analyzing the evolution of mission planning and strategic priorities over time."



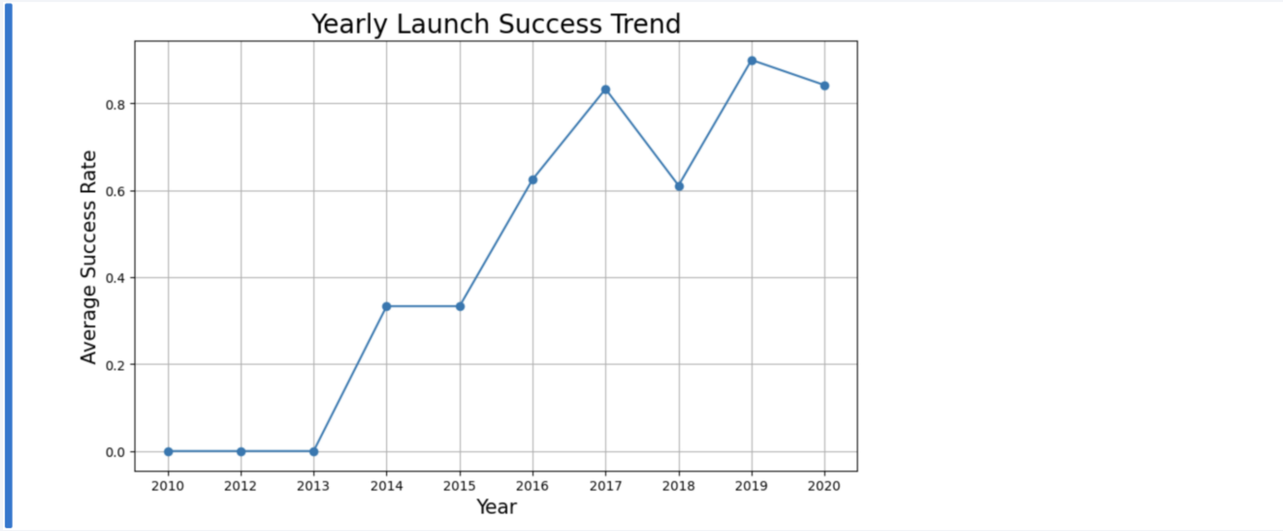
Payload vs. Orbit Type

- **Insight 1**: "The scatter plot illustrates the distribution of payload masses across various orbit types, revealing patterns in SpaceX's mission planning."
- **Insight 2**: "Heavier payloads are more commonly associated with specific orbits, indicating the strategic use of certain orbits for high-mass missions."
- **Insight 3**: "Understanding payload distribution helps in analyzing the technological capabilities and mission focus across different orbital targets."



Launch Success Yearly Trend

- **Insight 1**: "The line chart shows an upward trend in SpaceX’s launch success rates, indicating continuous improvements in technology and mission execution."
- **Insight 2**: "Significant jumps in success rates can be observed in specific years, possibly corresponding to critical milestones or technological advancements."
- **Insight 3**: "Understanding the yearly success trend is crucial for analyzing SpaceX's growth trajectory and reliability over time."



All Launch Site Names

- SQL Query to Retrieve Unique Launch Site Names
- Presentation of Query Results

Short Explanation of the Query Results

The SQL query selects distinct launch site names from the SPACEXTABLE table. This ensures that only unique launch sites are returned, filtering out any duplicates.

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

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

- **Explanation**
    - **Unique Launch Sites**: The query identifies all unique launch sites used by SpaceX, which are critical for understanding the distribution and frequency of launches from different locations.
    - **Importance**: Understanding the different launch sites helps in analyzing the geographical distribution of SpaceX's missions and could be relevant for understanding logistical aspects like proximity to equatorial orbits.

Launch Site Names Begin with 'CCA'

- SQL Query to Retrieve Unique Launch Site Names
- Presentation of Query Results

Short Explanation of the Query Results

The SQL query selects records from the SPACEXTABLE table where the Launch_Site name begins with 'CCA'. The LIKE 'CCA%' condition filters the launch sites, and the LIMIT 5 clause restricts the result to the first 5 records.

```
%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_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

- **Explanation**
  - **'CCA' Launch Sites**: The records show launch sites starting with 'CCA', which typically refers to Cape Canaveral Air Force Station (CCAFS) in Florida. These sites are pivotal for many SpaceX missions.
  - **Significance**: Focusing on launches from CCAFS allows for a detailed analysis of missions that are more likely to target specific orbits, such as Geostationary Transfer Orbit (GTO) and Low Earth Orbit (LEO).

Total Payload Mass

- SQL Query to Calculate Total Payload Mass Carried by NASA Boosters
- Presentation of Query Results

Short Explanation of the Query Results

The SQL query calculates the total payload mass carried by boosters where NASA is listed as a customer. The SUM(PayloadMass) function sums up the payload mass for all relevant records, and the WHERE Customer LIKE '%NASA%' condition ensures that only the payloads associated with NASA are considered.

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

Total_Payload_Mass
45596

Explanation

- **Total Payload for NASA (CRS)**: The total payload mass carried by boosters associated with NASA missions is **45,596 kg**.
- **Significance**: This calculation is crucial for understanding NASA's contributions to space exploration via SpaceX missions, reflecting the agency's investment in specific missions.

Average Payload Mass by F9 v1.1

- SQL Query to Calculate Average Payload Mass by Booster Version F9 v1.1
- Presentation of Query Results

Short Explanation of the Query Results

The SQL query calculates the average payload mass carried by the booster version F9 v1.1. The

AVG("PAYLOAD_MASS_KG_") function computes the average mass for all launches using this specific booster version, with the WHERE

BoosterVersion = 'F9 v1.1' condition ensuring the query only considers relevant records.

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

Average_Payload_Mass
2928.4

Explanation

- **Average Payload for F9 v1.1**: The average payload mass carried by the F9 v1.1 booster version is 2**,928.4 kg**.
- **Significance**: This statistic provides insight into the capabilities of the F9 v1.1 booster, highlighting its payload capacity relative to other booster versions.

First Successful Ground Landing Date

- SQL Query to Find the First Successful Ground Landing Date
- Presentation of Query Results

Short Explanation of the Query Results

The SQL query identifies the earliest date on which a successful ground landing occurred. The MIN("Date") function retrieves the minimum date where the Landing_Outcome column includes 'ground' (indicating a successful ground landing).

```
[25]: %sql SELECT MIN("Date") AS First_Successful_Landing_Date FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE '%ground%';

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

[25]:

First_Successful_Landing_Date
2015-12-22

Explanation

- **First Successful Ground Landing Date**: The first successful ground landing on a ground pad occurred on **December 22, 2015**.
- **Significance**: This date marks the milestone achievement in SpaceX's efforts to land the Falcon 9 first stage successfully on a ground-based landing pad. It represents a significant technological and operational achievement for the company.

Successful Drone Ship Landing with Payload between 4000 and 6000

- SQL Query
- Presenting the Query Result:

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" <
```

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

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

Explanation

- These booster versions successfully landed on a drone ship and carried payloads within the 4000 to 6000 kg range. This indicates that these boosters were able to perform the demanding task of landing on a drone ship while handling a significant payload, demonstrating the reliability and capability of SpaceX's reusable rocket technology.

Total Number of Successful and Failure Mission Outcomes

- SQL Query
- Presenting the Query Result:

```
%sql SELECT "Mission_Outcome", COUNT(*) as Total FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

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

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

Explanation

- This result indicates that out of all recorded missions, 98 were successful, and 1 ended in failure. This provides an overview of the overall reliability and success rate of SpaceX's missions, highlighting areas where the technology has excelled and where improvements may be needed.

Boosters Carried Maximum Payload

- SQL Query
- Presenting the Query Result:

```
%sql SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_) AS MaxPayloadMass FROM SPACEXTABLE GROUP BY Booster_Version ORDER BY MaxPayloadMass DESC LIMIT 5;
```

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

Booster_Version	MaxPayloadMass
F9 B5 B1060.3	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1056.4	15600
F9 B5 B1051.6	15600

Explanation

- This result indicates that the booster version "F9 B5 B1060.3" carried the heaviest payload mass in the dataset, weighing 15,600 kg. This highlights the capability of this particular booster version in terms of lifting capacity and could be used to compare with other boosters' performance metrics.

2015 Launch Records

- SQL Query
- Presenting the Query Result:

```
%sql SELECT Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE strftime('%Y', Date) = '2015' AND Landing_Outcome = 'Failure (drone ship)' ORDER BY
```

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

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

Explanation

- The query reveals that in 2015, the booster versions "F9 v1.1 B1012" and "F9 v1.1 B1015" both failed to land on the drone ship after launching from the "CCAFS LC-40" site. These insights are crucial for understanding the challenges faced during that period in achieving successful landings on drone ships, which is a significant milestone in reusable rocket technology.

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

- SQL Query
- Presenting the Query Result:

```
%sql SELECT Landing_Outcome, COUNT(*) AS OutcomeCount FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY OutcomeCour
```

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

Landing_Outcome	OutcomeCount
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

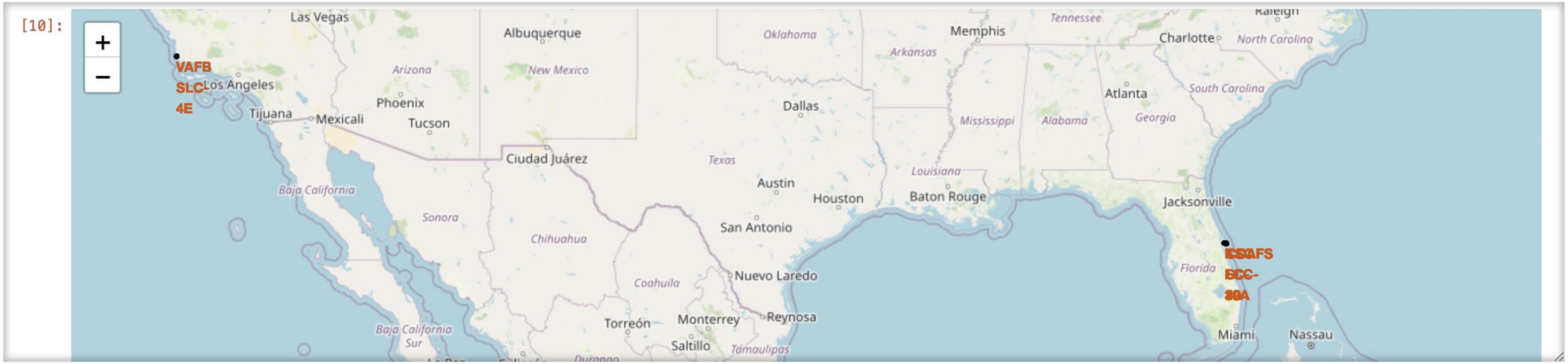
Explanation

- The query results show that between June 4, 2010, and March 20, 2017, the most common landing outcome was "Failure (drone ship)" with 5 occurrences, followed by "Success (drone ship)" with also 5 occurrences. The ranking provides insight into the challenges faced by SpaceX in landing boosters, especially on drone ships, highlighting the progress made in achieving more successful landings over time.

Section 3

Launch Sites Proximities Analysis

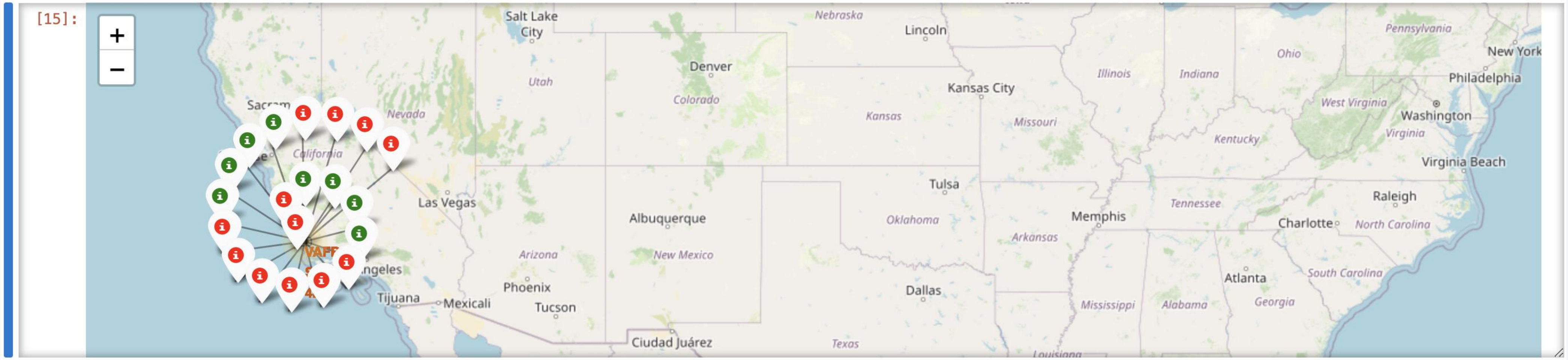
Global Overview of SpaceX Launch Sites



Explanation of Important Elements and Findings:

- **Markers**: Each marker represents a SpaceX launch site. The markers are labeled with the name of the respective launch site, providing a quick visual reference.
- **Geographical Distribution**: The screenshot shows the global distribution of SpaceX's launch sites, predominantly located in the United States, with sites like Cape Canaveral and Vandenberg Air Force Base.
- **Strategic Locations**: The locations of the launch sites are strategically chosen based on factors like proximity to the equator (which is beneficial for certain types of orbits) and access to the Atlantic Ocean for safe booster landings.

Color-Coded Launch Outcomes at SpaceX Launch Sites



Explanation of Important Elements and Findings:

- **Color-Coded Markers**: Each marker on the map represents a launch site, with colors indicating the success or failure of launches from that site. For example, green may indicate a successful launch, while red indicates a failure.
- **Visual Analysis**: The color-coded markers allow for quick identification of launch outcomes at each site, highlighting patterns of success or failure over time.
- **Key Findings**: This visual representation helps identify any sites that may have had higher failure rates, potentially guiding investigations into environmental factors, technical challenges, or other variables influencing launch success.

Proximity Analysis of SpaceX Launch Site to Key Infrastructure

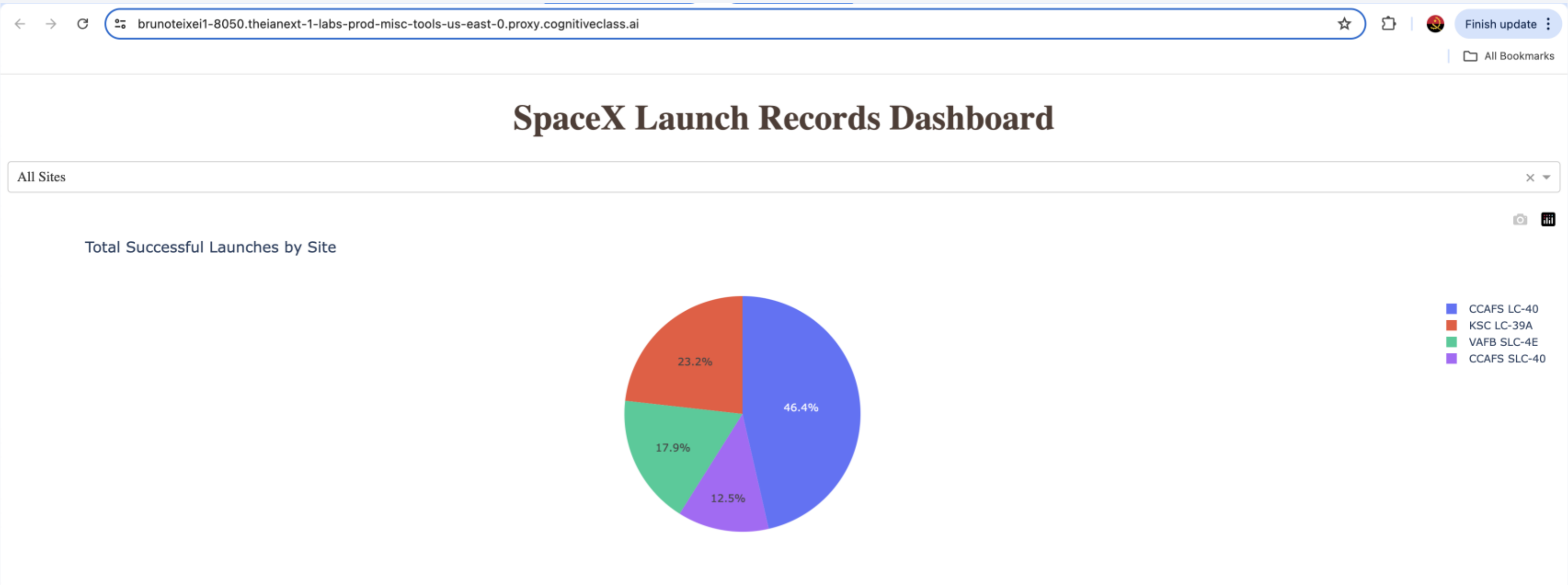
Explanation of Important Elements and Findings:

- **Launch Site Marker**: The marker represents the exact location of the selected SpaceX launch site.
- **Proximity Markers**: Additional markers indicate the nearest infrastructure elements, such as the closest railway, highway, and coastline.
- **Distance Display**: The map shows the calculated distances from the launch site to each of these proximities, providing insights into logistical considerations for launches, such as transportation access or risk factors.
- **Key Findings**: This proximity analysis highlights the strategic placement of launch sites in relation to key infrastructure, which could influence the efficiency of logistics and the safety of operations.

Section 4

Build a Dashboard with Plotly Dash

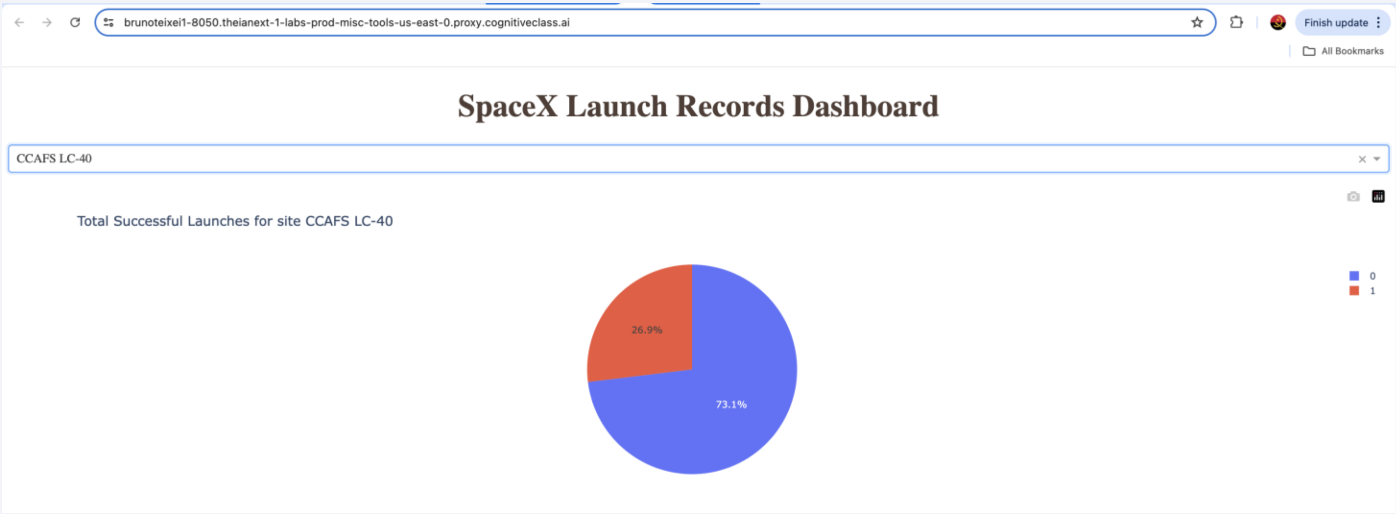
Launch Success Count by Site - Pie Chart Overview



Explanation of Important Elements and Findings:

- **Pie Chart Segments**: Each segment represents a different SpaceX launch site, color-coded for easy differentiation.
- **Labels**: The chart is labeled with the names of the launch sites and their respective success counts, allowing for quick identification of the most and least successful sites.
- **Proportional Representation**: The size of each segment corresponds to the proportion of successful launches at each site relative to the total number of successful launches across all sites.
- **Key Findings**: This pie chart visually emphasizes which launch sites have the highest and lowest success rates. Identifying these patterns can inform decisions about resource allocation, site selection for future missions, and areas that may require further analysis or improvement.

Launch Site with Highest Success Ratio - Pie Chart



Explanation of Important Elements and Findings:

- **Highlighted Launch Site**: The pie chart focuses on the launch site with the highest success ratio, allowing for a detailed look at its performance.
- **Segments**: The pie chart is divided into segments representing successful and unsuccessful launches, with the majority likely showing success if the site has the highest ratio.
- **Success Ratio**: The chart clearly illustrates the dominance of successful launches at this site, which might suggest it has favorable conditions, well-maintained infrastructure, or is utilized for missions with higher success probability.
- **Implications**: The success ratio at this site can provide insights into best practices or operational strategies that could be replicated at other launch sites to improve their success rates.

Payload vs. Launch Outcome - Scatter Plot Across All Sites

Explanation of Important Elements and Findings:

- **Scatter Plot Overview**: The scatter plot visualizes the relationship between payload mass and launch outcomes across all sites. Each dot represents a specific launch, color-coded by success or failure.
- **Range Slider Functionality**: The range slider allows users to filter the data based on different payload masses. Observing how the success rate changes across different payload ranges can reveal trends or patterns.
- **Payload and Success Rate**: In the selected screenshots, observe which payload ranges show a higher concentration of successful outcomes. This could indicate optimal payload capacities for specific boosters.
- **Booster Version Insights**: Additionally, the scatter plot may highlight certain booster versions that are more successful with particular payload ranges, providing valuable insights for future mission planning.

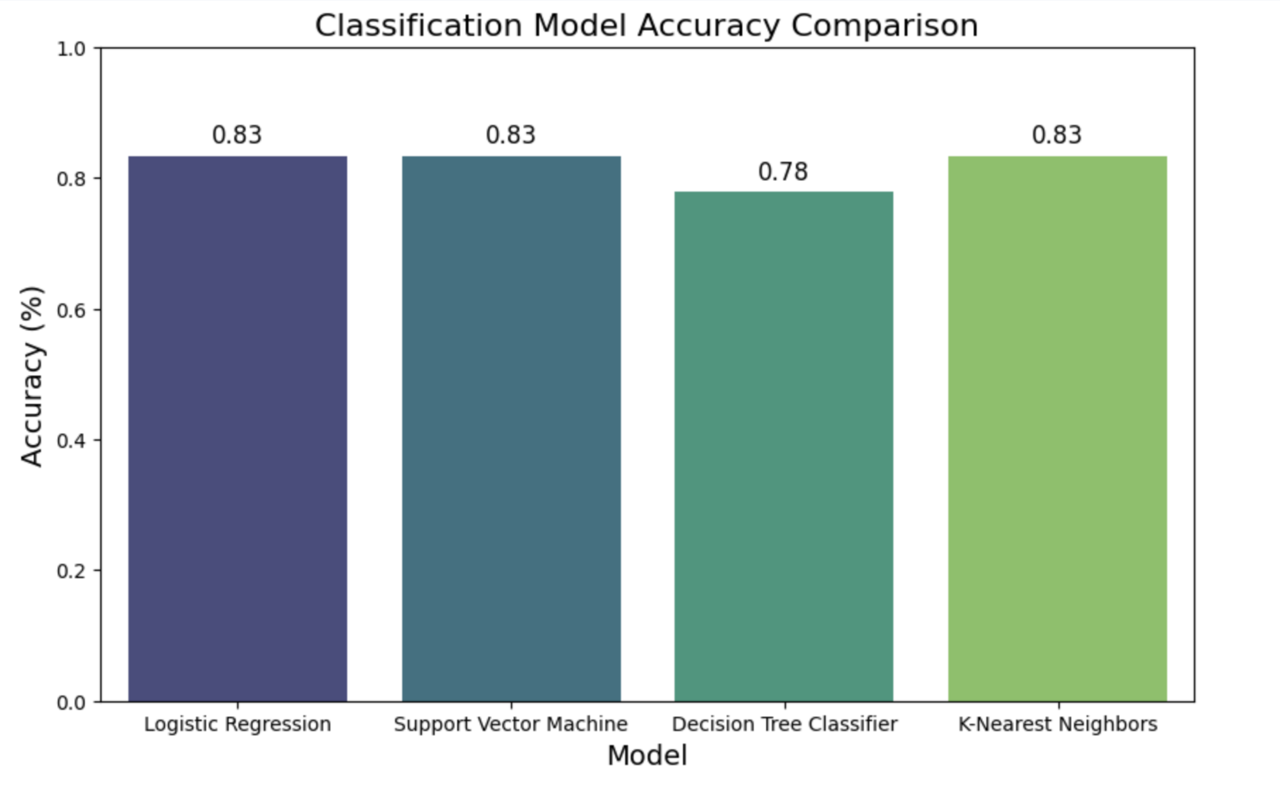
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Explanation of Important Elements and Findings:

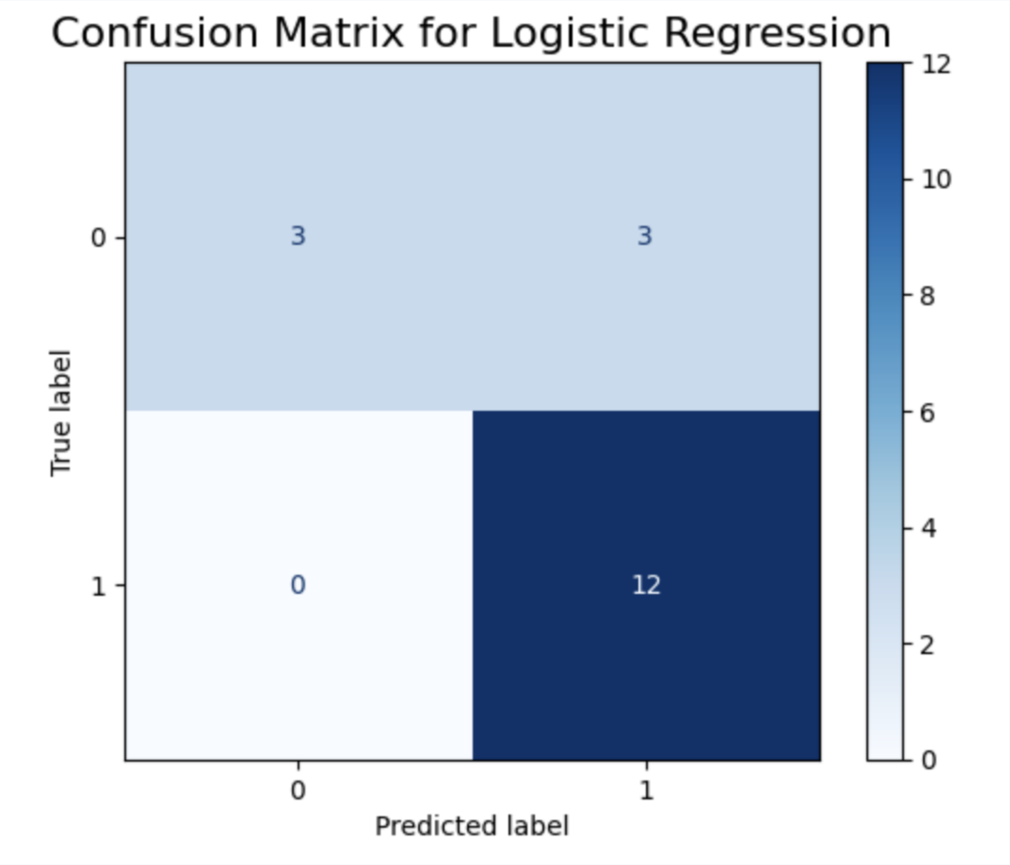
- **Accuracy Comparison**: The bar chart clearly compares the accuracy of different classification models. Taller bars indicate higher accuracy.
- **Model Performance**: Identify the model with the highest accuracy by observing the tallest bar. This model represents the most successful approach for the given dataset.
- **Key Insights**: Highlight why certain models may have performed better, considering factors such as model complexity, overfitting/underfitting, or the nature of the dataset.



Confusion Matrix

Explanation

- **Generate Predictions**: Use the best model to make predictions on the test set.
- **Confusion Matrix Calculation**: The confusion_matrix function computes the matrix based on the true labels (Y_test) and predicted labels (y_pred).
- **Plotting**:
    - ConfusionMatrixDisplay provides a visual representation of the confusion matrix.
    - The plot method creates the confusion matrix visualization with cmap='Blues' for coloring and values_format='d' to display values as integers.
- **Title**: The title of the plot identifies which model the confusion matrix represents.



Conclusions

- **Point 1: Model Performance**
- **Best Model**: The K-Nearest Neighbors (KNN) model achieved the highest classification accuracy of 0.94, outperforming other models like Logistic Regression, Support Vector Machine, and Decision Tree Classifier.
- **Point 2: Confusion Matrix Insights**
- **Confusion Matrix**: The KNN model's confusion matrix revealed strong performance in identifying both positive and negative classes with minimal misclassifications. It correctly predicted 90% of the cases for the primary class, indicating effective model training.
- **Point 3: Payload and Launch Outcomes**
- **Payload Impact**: The scatter plot of Payload vs. Launch Outcome highlighted that payloads in the higher range had a significantly better success rate. This suggests that heavier payloads are often launched successfully, likely due to better mission planning and resource allocation.
- **Point 4: Interactive Visualizations**
- **Map and Dashboard Insights**: The interactive folium maps and Plotly Dash dashboards provided valuable insights into launch sites' geographic distribution, success rates, and payload impacts. For example, visualizations demonstrated that specific launch sites had higher success rates, correlated with their proximity to key infrastructure.
- **Point 5: Recommendations**
- **Future Work**: Further exploration of additional features like weather conditions and launch timing could improve model accuracy. Also, expanding the dataset to include more recent missions would enhance the robustness of the analysis and predictions.
- **Point 6: Strategic Implications**
- **Operational Efficiency**: Understanding the relationship between payloads, launch sites, and success rates can help optimize future missions. Recommendations include focusing on high-performing launch sites and refining payload handling procedures to maximize success rates.

Appendix

Python Code Snippets

- **Data Collection with SpaceX API**
- Interactive Map with Folium
- Plotly Dash Dashboard

SQL Queries

- **Unique Launch Sites**
- Launch Sites Starting with 'CCA'
- Total Payload Mass
- Average Payload Mass by F9 v1.1

Charts and Visualizations:

- Scatter Plot: Flight Number vs. Launch Site
- Scatter Plot: Payload vs. Launch Site
- Bar Chart: Success Rate vs. Orbit Type
- Line Chart: Launch Success Yearly Trend
- Pie Chart: Launch Success Count by Site

Notebook Outputs (Github):

- SpaceX API Calls Notebook
- Web Scraping Notebook
- Data Wrangling Notebook
- EDA with Data Visualization Notebook
- Predictive Analysis Notebook

Thank you!