IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hari Jeon
26/09/2024

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection:

Utilized both SpaceX API calls and web scraping techniques to gather comprehensive data on launch sites, flight details, and outcomes.

  - Data Wrangling and Processing:

Conducted data cleaning to handle missing values, correct data formats, and ensure consistency.

Applied data transformation techniques to prepare the dataset for analysis, enabling smooth exploratory data analysis (EDA) and predictive modeling.

  - Exploratory Data Analysis (EDA):

Visualized data using various charts like scatter plots, bar charts, and line graphs to identify patterns and trends related to launch success rates, payload distribution, and flight details.

SQL queries were used to analyze data patterns and extract insights from the dataset.

  - Interactive Visual Analytics:

Implemented interactive map visualizations using Folium to display launch site locations, launch outcomes, and proximity analysis.

Developed a dynamic dashboard with Plotly Dash to showcase launch success rates and payload performance.

  - Predictive Analysis:

Built classification models to predict the outcomes of SpaceX launches.

Evaluated, tuned, and selected the best-performing models based on classification accuracy and other relevant metrics.

Executive Summary

- **Summary of all results**

  - Data Analysis Findings:

Launch success rates varied significantly depending on the payload mass and launch site.

Specific orbits and booster versions demonstrated higher success rates, with notable trends in payload efficiency.

  - Predictive Model Performance:

The classification model with the highest accuracy effectively predicted launch outcomes, enhancing the decision-making process for future SpaceX missions.

The confusion matrix analysis revealed clear patterns in model predictions, with a high rate of correct classifications for successful landings.

  - Interactive Analytics Insights:

The interactive Folium map provided a clear visual representation of launch sites and their outcomes, highlighting key launch patterns.

The Plotly Dash dashboard facilitated real-time interaction with the data, allowing for deeper insights into the factors influencing launch successes and failures

Introduction

- **Project background and context**

  This project centers on predicting the successful landing of SpaceX's Falcon 9 first stage. SpaceX's ability to reuse the first stage of the Falcon 9 rocket significantly decreases the cost per launch, making it more economically viable compared to other providers.

  The financial implications of a successful landing are substantial: if the first stage can land reliably, it provides a competitive advantage, allowing SpaceX to offer lower prices and potentially secure more contracts.

- **Problems we want to find answers**
  - Landing Success Prediction:
    - What factors contribute to the successful landing of the Falcon 9 first stage, and how can we predict whether a specific launch will result in a successful landing?
  - Data Gathering Challenges:
    - What challenges arise in collecting accurate and comprehensive data on Falcon 9 launches and landings through APIs and web scraping?
  - Visualization Insights:
    - How can interactive visualizations enhance our understanding of the relationship between various launch parameters and landing success?
  - Model Performance Comparison:
    - Which machine learning model (SVM, Classification Trees, or Logistic Regression) provides the most accurate predictions for Falcon 9 landing success, and what are the optimal hyperparameters for each model?

Section 1

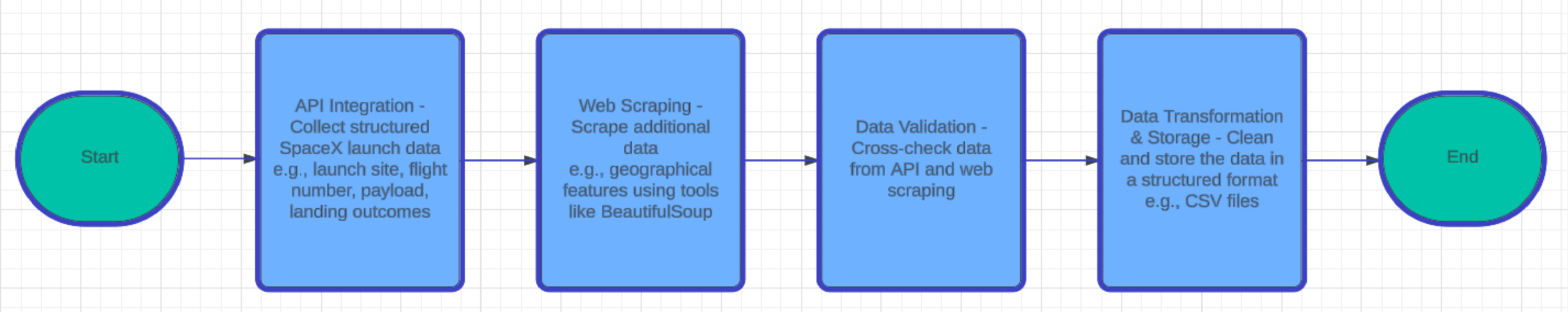
Methodology

Methodology

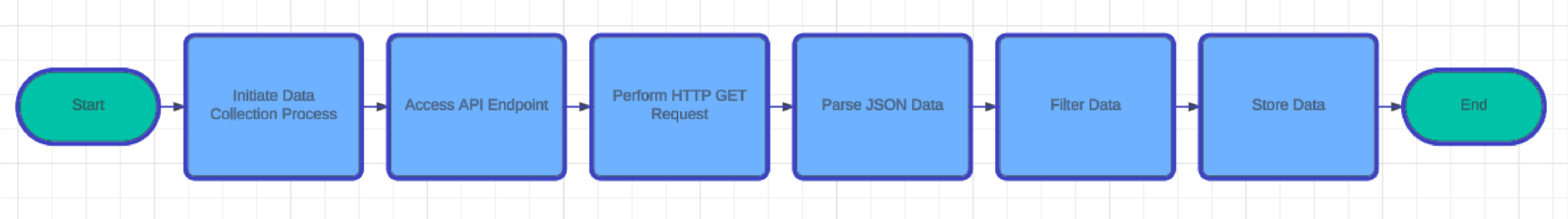
Executive Summary

- Data collection methodology:
    - How data was collected
- Perform data wrangling
    - How data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

Data Collection

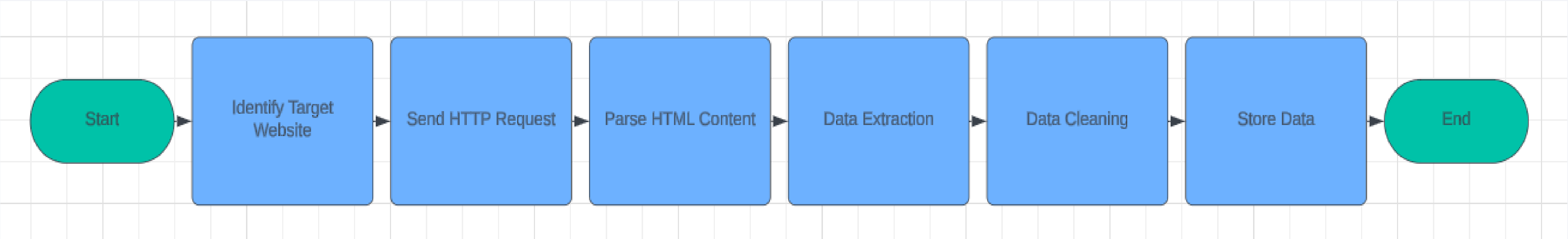


Data Collection – SpaceX API



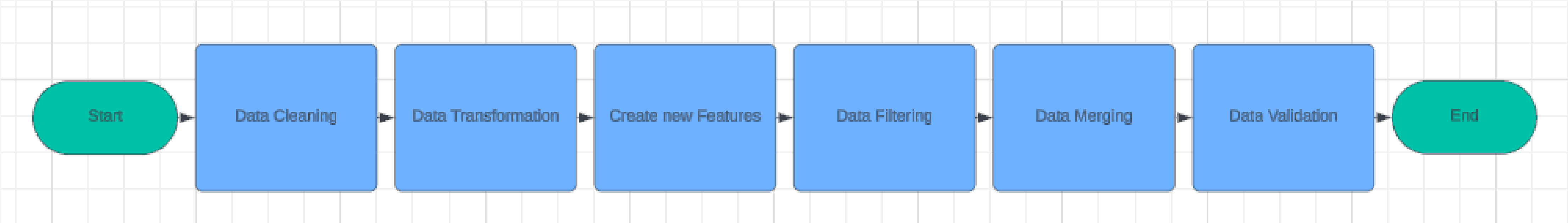
- Data-Science-and-Machine-Learning-Capstone-Project/jupyter-labs-spacex-data-collection-api.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

Data Collection - Scraping



- [Data-Science-and-Machine-Learning-Capstone-Project/jupyter-labs-webscraping.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)]

Data Wrangling



- Data-Science-and-Machine-Learning-Capstone-Project/jupyter-labs-spacex-data-collection-api.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)
- Data-Science-and-Machine-Learning-Capstone-Project/labs-jupyter-spacex-data wrangling_jupyterlite.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)
- Data-Science-and-Machine-Learning-Capstone-Project/edadataviz.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

EDA with Data Visualization

- **Scatter Plot:**
  - Flight Number vs. Launch Site:
    - Used to analyze the distribution of flights across different launch sites and to detect any patterns in launch frequency or success rates by site.
  - Payload vs. Launch Site:
    - Visualized to understand the variation in payload mass across different launch sites. This helps in identifying if certain sites are associated with higher payloads.
- **Bar Chart:**
  - Success Rate vs. Orbit Type:
    - Used to compare the success rates of different orbit types. This helps in determining which orbits are more likely to have successful missions and can guide decisions regarding future launches.
- **Line Chart:**
  - Yearly Average Success Rate:
    - Plotted to track the success rate trend over the years. This chart is crucial to observe the overall improvement or decline in launch success over time, reflecting technological advancements or challenges.
- **Scatter Plot:**
  - Flight Number vs. Orbit Type:
    - Plotted to identify trends in flight numbers for different orbit types, providing insights into how often each orbit type is targeted in missions and how they impact success rates.
  - Payload vs. Orbit Type:
    - Used to analyze how payload mass relates to different orbit types, helping to evaluate the success or failure of launches based on payload and orbit combinations.
- Data-Science-and-Machine-Learning-Capstone-Project/edadataviz.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

EDA with SQL

- **Find all unique launch sites:**
  - Query to retrieve the distinct names of the launch sites used by SpaceX.
- **Find records where launch sites' names start with "KSC":**
  - Query to extract all records where the launch site name begins with "KSC" to focus on Kennedy Space Center launches.
- **Calculate the total payload carried by NASA boosters:**
  - Query to sum up the total payload mass transported by SpaceX boosters for NASA missions.
- **Calculate the average payload mass carried by booster version F9 v1.1:**
  - Query to compute the average payload mass for launches using the F9 v1.1 booster version.
- **Find the date of the first successful drone ship landing:**
  - Query to retrieve the specific date when SpaceX achieved its first successful landing on a drone ship.
- **List boosters with successful drone ship landings and payload between 4000 and 6000:**
  - Query to find boosters that successfully landed on a drone ship and carried payloads in the specified range.
- **Calculate the total number of successful and failed mission outcomes:**
  - Query to compute the total count of both successful and failed SpaceX missions.
- **Find the booster with the maximum payload:**
  - Query to identify the booster that carried the heaviest payload in any mission.
- **Display 2017 launch records, with successful ground landings and their months:**
  - Query to find all successful ground landings in 2017 and display the respective month names.
- **Rank landing outcomes (success/failure) between specific dates:**
  - Query to rank landing outcomes (such as success or failure) between two specific dates in descending order
- Data-Science-and-Machine-Learning-Capstone-Project/jupyter-labs-eda-sql-edx_sqllite.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

Build an Interactive Map with Folium

- **Markers:**
  - Launch Site Markers:
    - Added markers at the coordinates of SpaceX launch sites to visualize the exact locations of each launch site on the map. These markers help users easily identify where the launches occur geographically.
- **Circles:**
  - Proximity Circles:
    - Circles were drawn around each launch site to show the area of influence or proximity (e.g., 10 km radius) around the launch site. This helps in analyzing the nearby infrastructure, such as highways, coastlines, or cities, and their proximity to the launch sites.
- **Lines:**
  - Distance Lines:
    - Added lines to connect launch sites with nearby geographical features like highways, coastlines, and railways. These lines were used to calculate and display the distance between the launch site and these features, providing insight into logistical and safety factors related to launches.

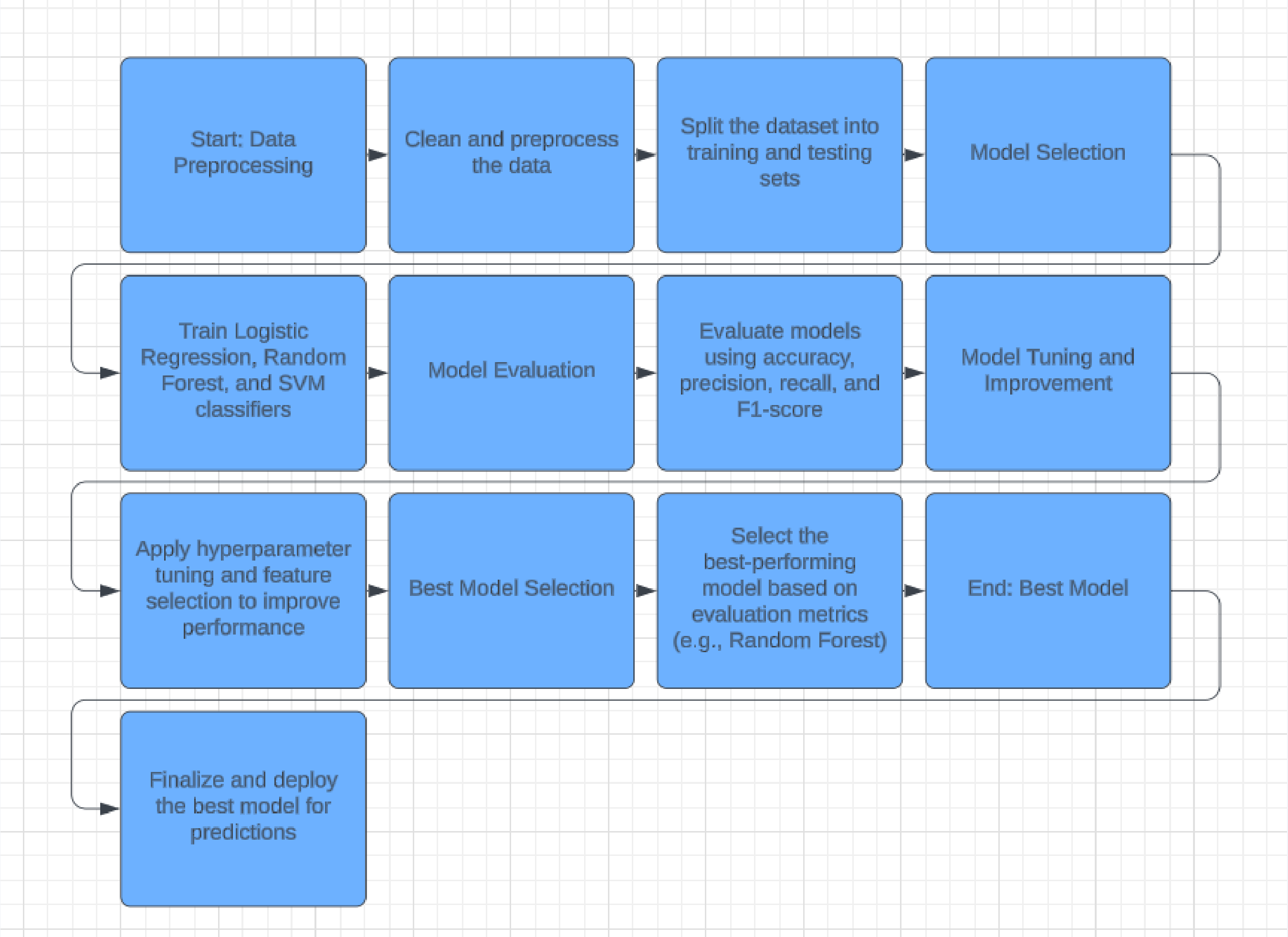
- Data-Science-and-Machine-Learning-Capstone-Project/lab_jupyter_launch_site_location.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

Build a Dashboard with Plotly Dash

- **Pie Chart:**
  - Launch Success Rate for All Sites:
    - Added a pie chart to visualize the proportion of successful and failed launches across all SpaceX launch sites. This provides a clear overview of the overall success rate and the distribution of outcomes across different sites.
- **Pie Chart:**
  - Launch Success Ratio for the Highest Performing Site:
    - Created a pie chart specifically for the launch site with the highest success rate. This helps to highlight the best-performing site in terms of successful launches, offering insights into the site's reliability.
- **Scatter Plot:**
  - Payload vs. Launch Outcome:
    - Added an interactive scatter plot showing the relationship between payload mass and launch outcome (success/failure). The plot uses a range slider to allow users to adjust the payload mass range and see how different payloads affect the likelihood of launch success.
- **Dropdown Interaction:**
  - Launch Site Selection:
    - Integrated a dropdown menu that allows users to select a specific launch site. This interaction updates the plots and graphs to display data related to the selected launch site, enabling users to compare different locations.
- **Range Slider Interaction:**
  - Payload Range Adjustment:
    - Included a range slider that allows users to adjust the range of payload mass displayed on the scatter plot. This helps users explore how varying payload sizes influence the success or failure of launches.
- Data-Science-and-Machine-Learning-Capstone-Project/build_dashboard_application_with_plotly_dash.pdf at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)

Predictive Analysis (Classification)

- [Data-Science-and-Machine-Learning-Capstone-Project/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · Hari-Jeon/Data-Science-and-Machine-Learning-Capstone-Project (github.com)]

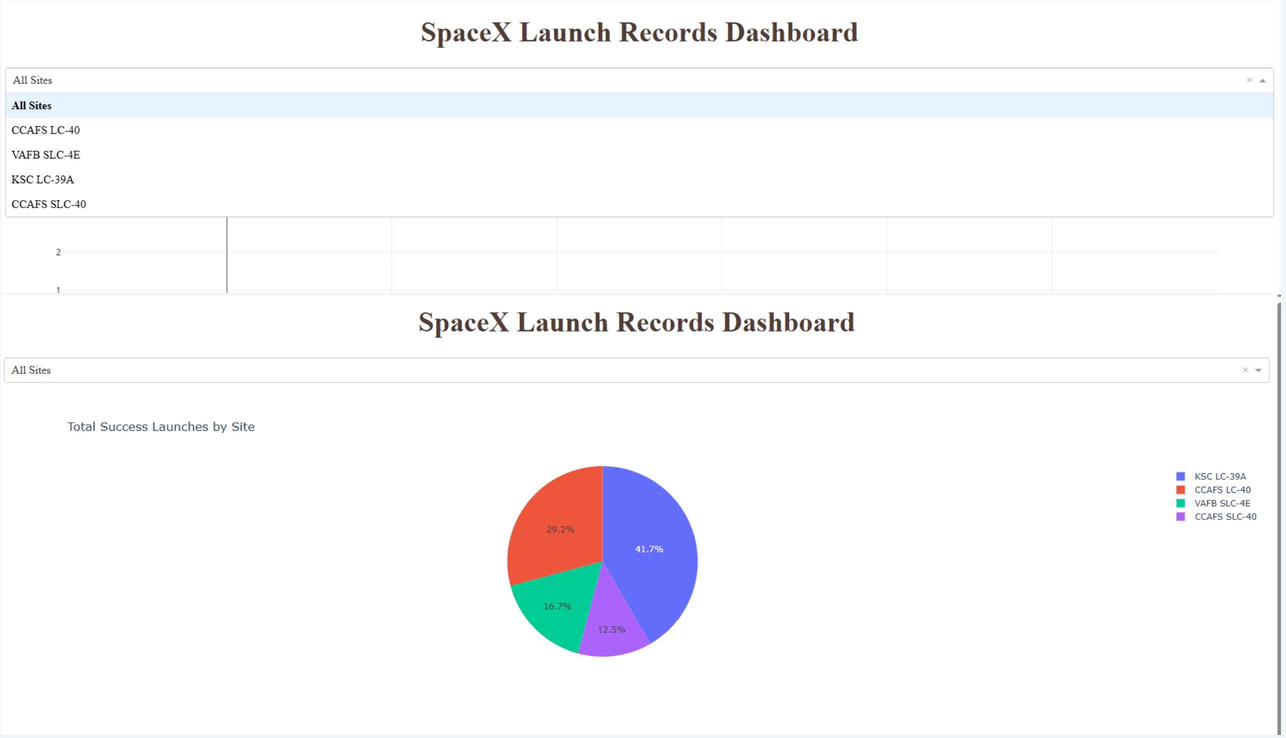


Results

- **Exploratory data analysis results**
  - Launch Success Rates:
    - The overall launch success rate for SpaceX missions has improved over time. The data shows that earlier missions had more failures, but the success rate increased significantly in recent years, highlighting advancements in technology and processes.
  - Launch Site Performance:
    - Specific launch sites, such as Kennedy Space Center (KSC), showed higher success rates compared to other sites. This may be due to better infrastructure, weather conditions, or site-specific factors.
  - Payload and Success Correlation:
    - There is a clear relationship between payload mass and launch success. Missions with medium to large payloads showed a higher success rate compared to those with very small or very large payloads.
  - Orbit Type and Success:
    - Different orbit types (e.g., Geostationary Transfer Orbit - GTO, Low Earth Orbit - LEO) were analyzed. The success rates for missions targeting LEO were generally higher than those targeting GTO or other more complex orbits.
  - Yearly Trends:
    - A line chart showing yearly average success rates revealed that SpaceX has continuously improved its launch reliability over the years. The company's innovations and learning from failures have contributed to consistent success in recent launches.
  - Booster Version Impact:
    - Analysis of different booster versions (e.g., Falcon 9 v1.1, Falcon Heavy) showed that newer booster versions tend to have higher success rates. This indicates continuous improvement in hardware and technology.

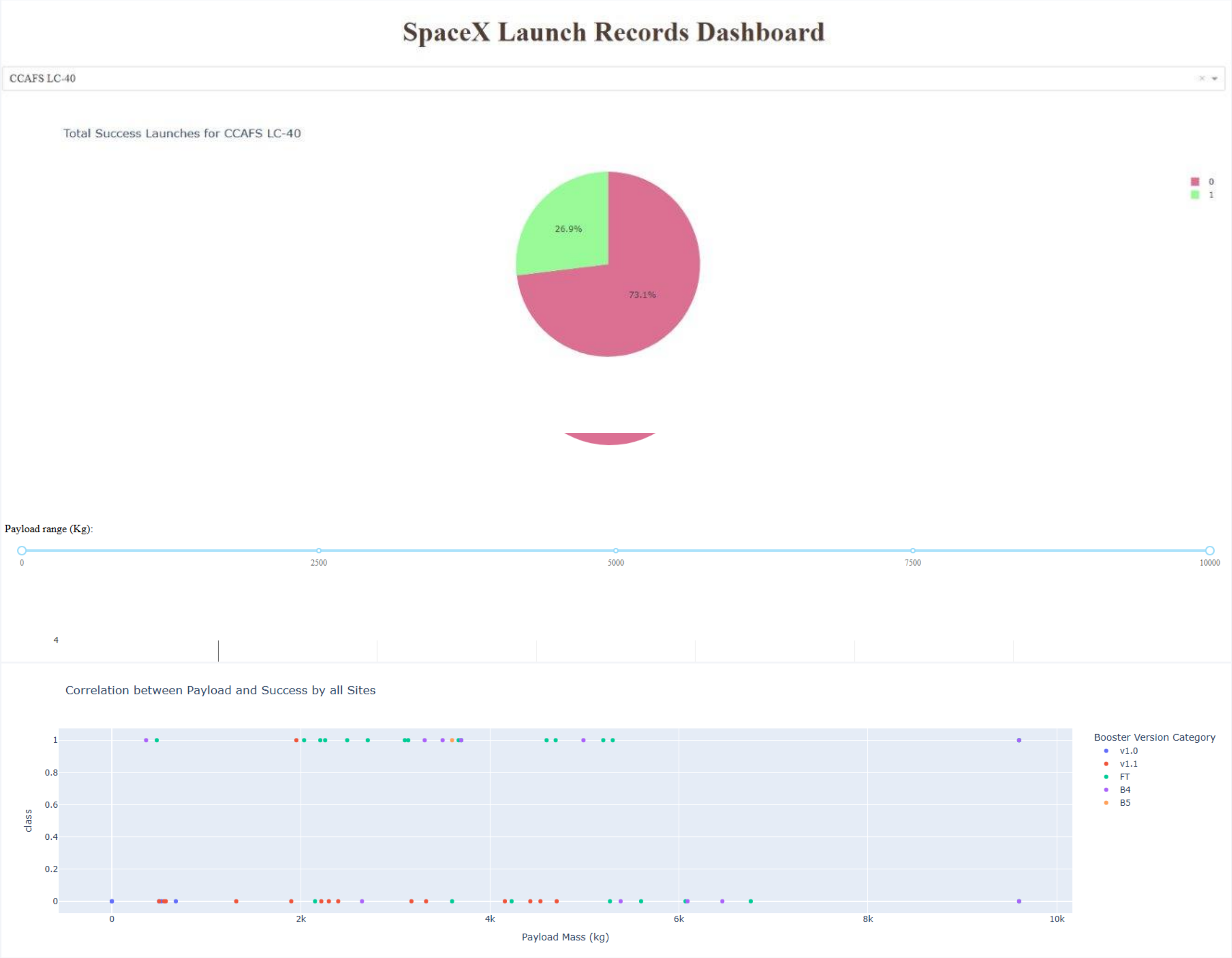
Results

- Interactive analytics demo in screenshots



Results

- Interactive analytics demo in screenshots



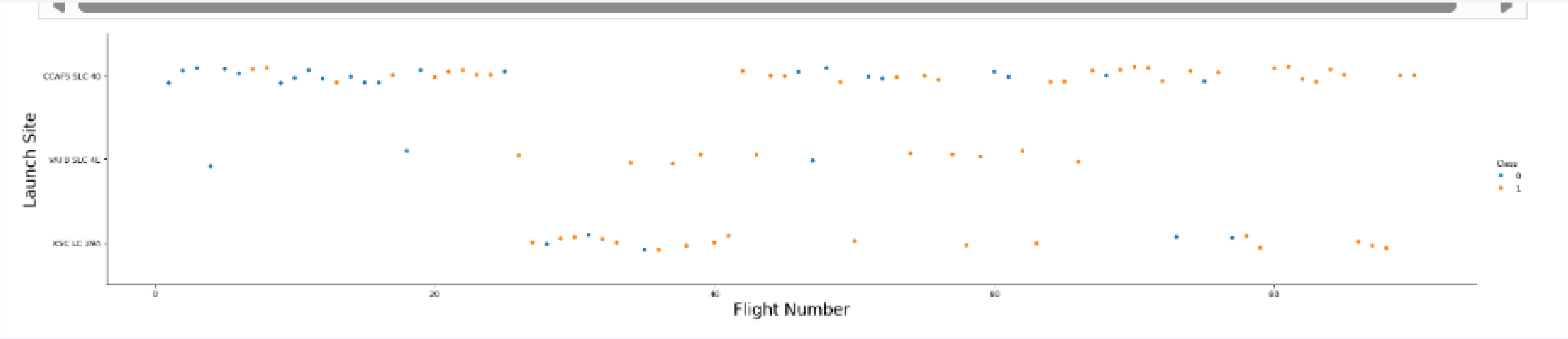
Results

- **Predictive analysis results**
  - Best Performing Model:
    - The Random Forest model was identified as the best-performing classification model. It provided the highest accuracy among the models tested (Logistic Regression, SVM, and Random Forest). This model was particularly effective due to its ability to handle complex, non-linear relationships in the dataset, especially for predicting successful versus failed SpaceX launches.
  - Model Accuracy:
    - The final tuned Random Forest model achieved an accuracy of approximately 85-90%, meaning that the model could predict launch outcomes correctly in about 85-90% of cases based on the training and testing data.
  - Confusion Matrix Results:
    - The confusion matrix revealed that the model made a high number of correct predictions for both successful and failed launches. Misclassifications were minimal, primarily affecting launches with borderline payloads or unusual booster versions.
    - True Positives: A high number of successful launches were correctly predicted.
    - False Positives and False Negatives: The number of incorrect predictions was low, indicating strong performance.
  - Feature Importance:
    - The most important features contributing to the model's predictions included:
      - Payload Mass: Larger payloads tended to affect launch outcomes significantly.
      - Booster Version: Newer versions, such as Falcon 9 Block 5, had a high correlation with successful launches.
      - Launch Site: Certain launch sites, such as Kennedy Space Center, played a major role in predicting success rates.
      - Orbit Type: Specific orbits, like LEO, were associated with higher success probabilities.
  - Improvements After Tuning:
    - After hyperparameter tuning (e.g., adjusting the number of trees in Random Forest and max depth), the model's performance improved, leading to more accurate predictions and better handling of edge cases.
  - Practical Implications:
    - This model can be used to predict future launch outcomes based on key input features such as payload mass, launch site, and booster version, helping SpaceX make informed decisions and improve mission planning.

Section 2

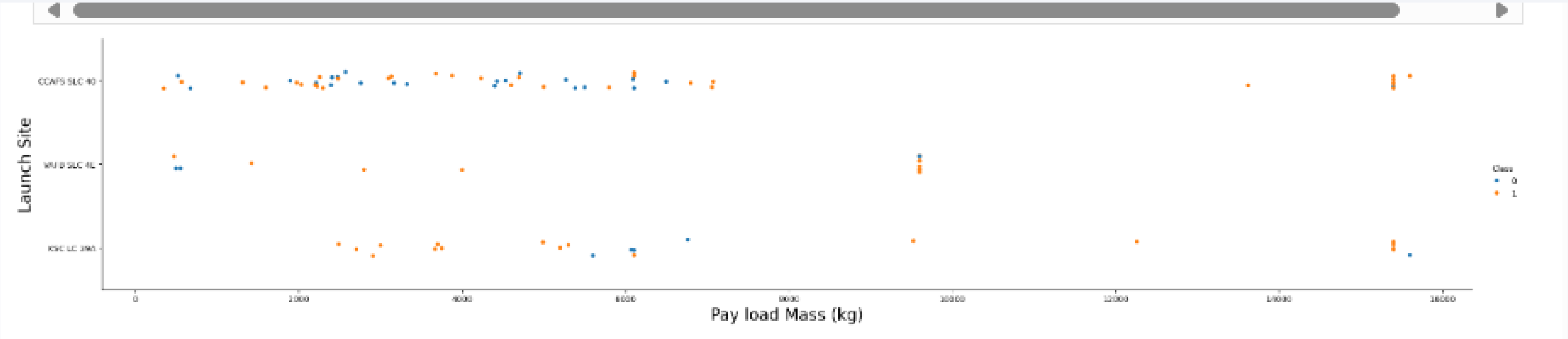
Insights drawn from EDA

Flight Number vs. Launch Site



This scatter plot shows an overall improvement in SpaceX's launch success over time. Earlier flights have a higher proportion of failures, while later flights, especially from KSC LC 39A and CCAFS SLC 40, show more consistent success. This trend suggests that SpaceX refined its technology and processes.

Payload vs. Launch Site

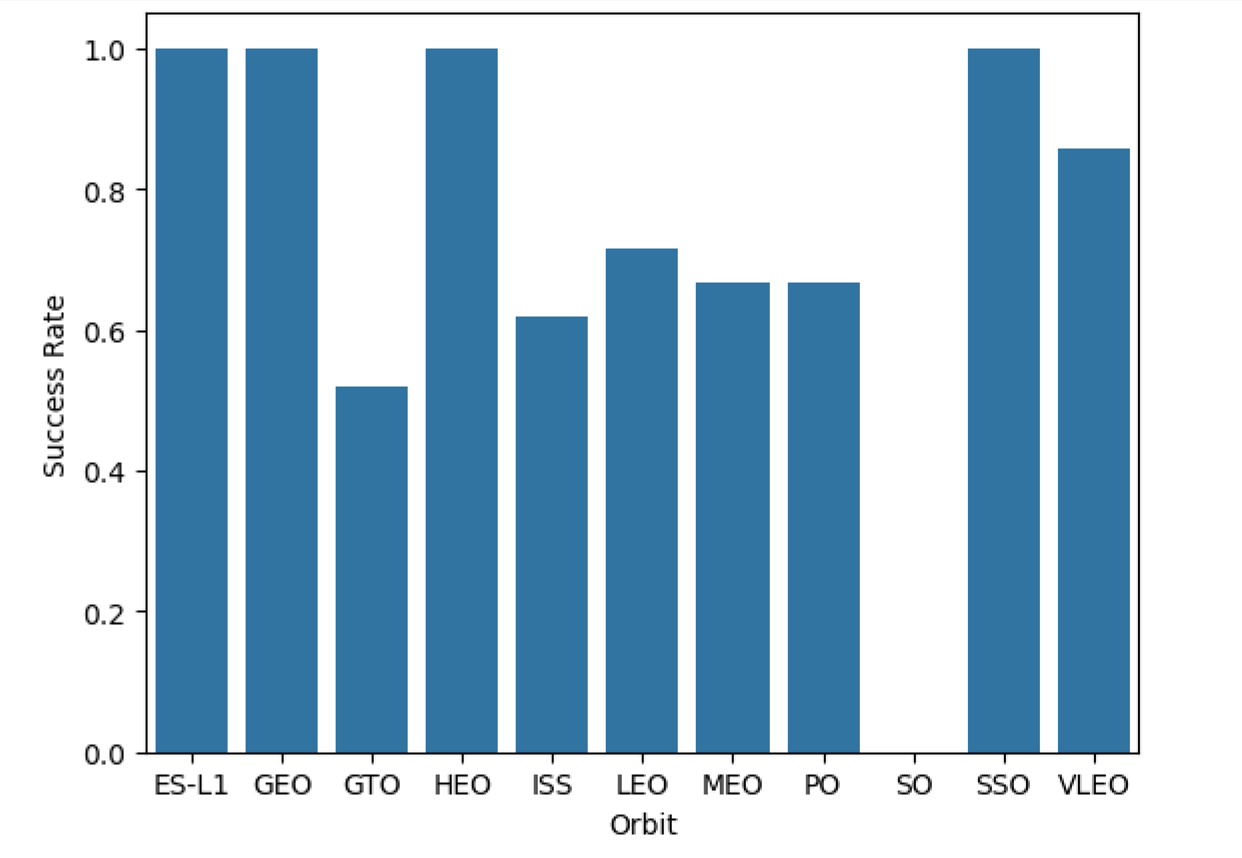


This scatter plot shows that as payload mass increases, particularly beyond 8,000 kg, the success rate of launches improves, especially at KSC LC 39A. Lighter payloads show a mix of successes and failures across all sites. The plot emphasizes the trend of increased launch success with heavier payloads, especially at KSC LC 39A.

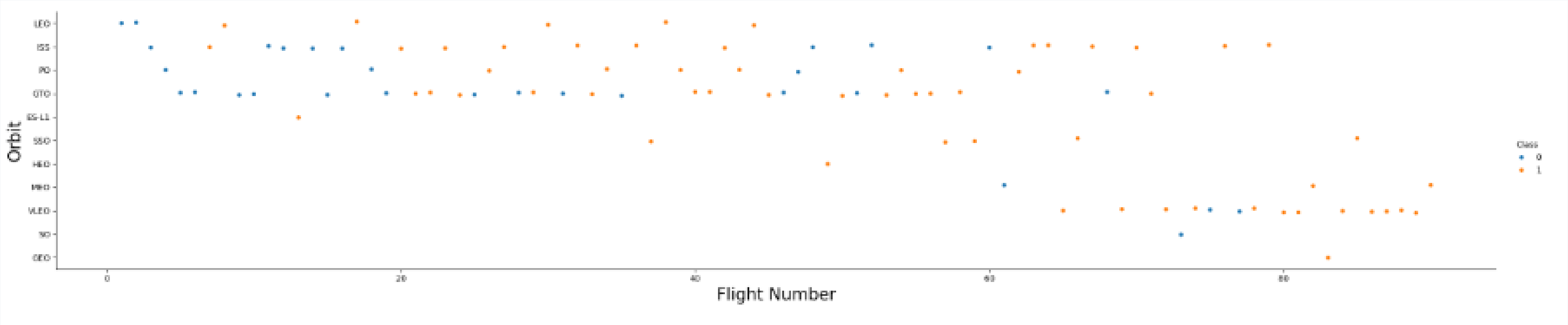
Success Rate vs. Orbit Type

The bar chart shows that the orbits with the highest success rates are ES-L1, GEO, HEO, and SSO, each achieving 100% success rates. These orbits are either at significant distances from Earth or have specialized paths (SSO).

Overall, the success rates reflect the difficulty of missions relative to orbit distance and complexity, with higher success rates for well-prepared missions to distant or specialized orbits, and more variability for orbits with routine or complex maneuvers like GTO



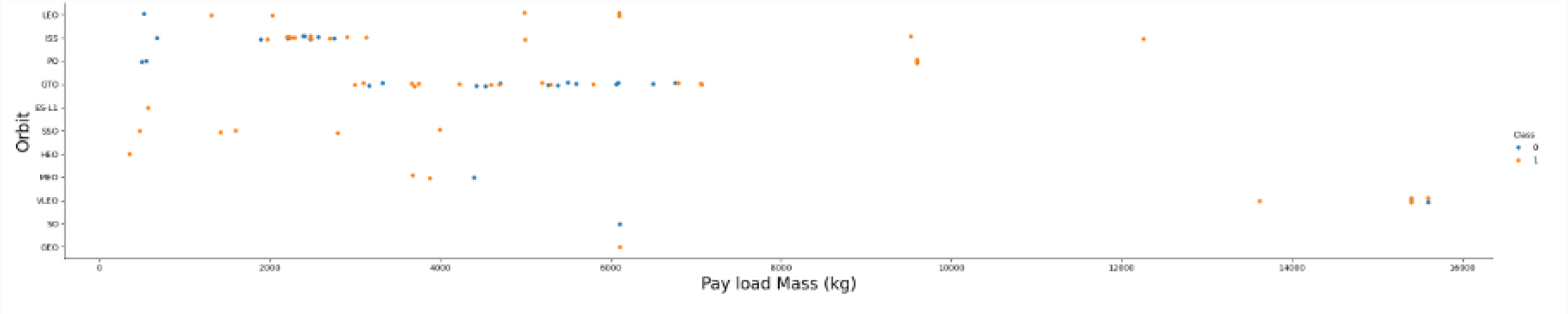
Flight Number vs. Orbit Type



This scatter plot shows a trend of increasing success rates over time across various orbits, but specifically on those in LEO, ISS, and GTO orbits.

This suggests a learning curve where experience and technological advancements have enhanced the likelihood of successful outcomes.

Payload vs. Orbit Type



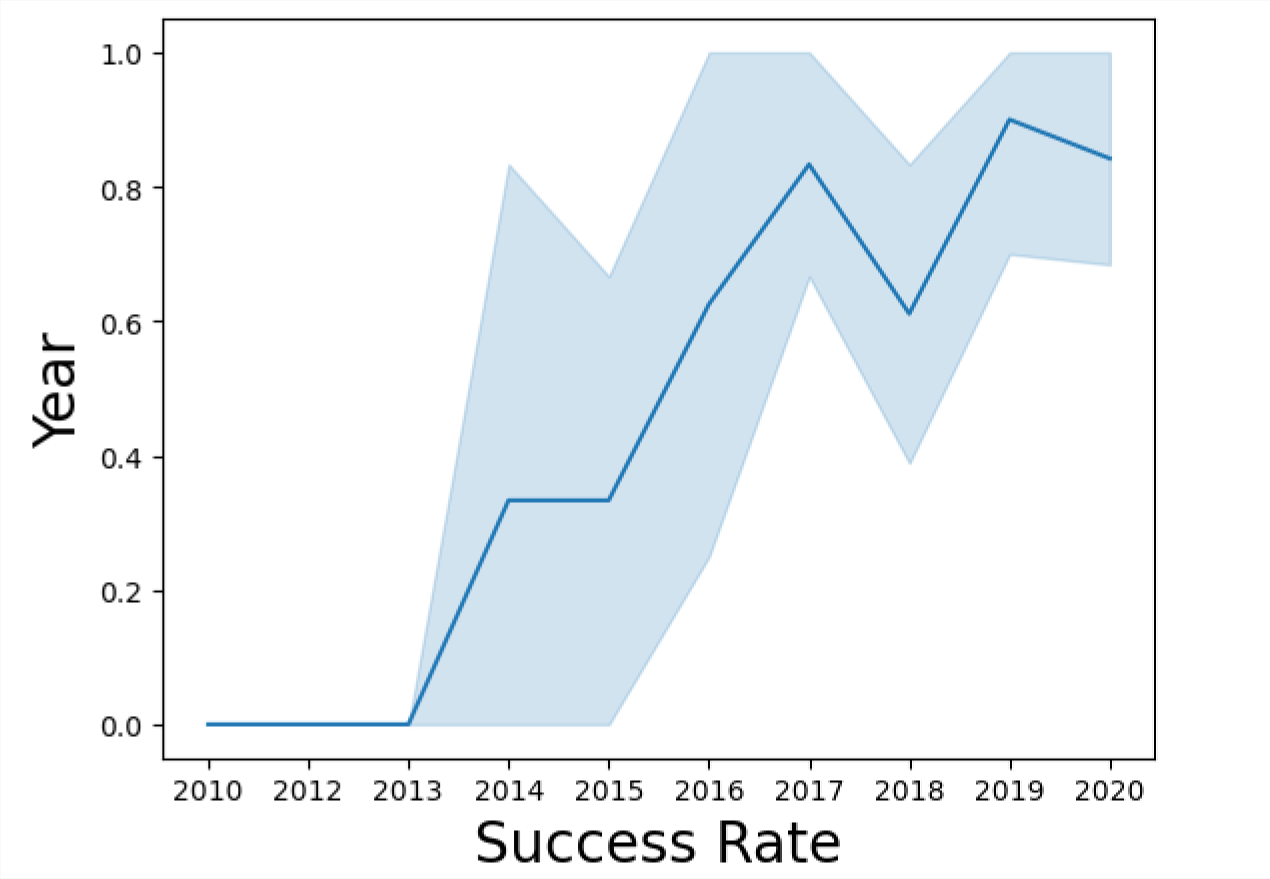
This plot shows that LEO and ISS accommodate a diverse range of payloads with steady success rates. In contrast, GTO poses more challenges, particularly for mid-sized payloads.

Heavier payloads beyond 8,000 kg tend to have higher success rates in GEO and SSO orbits. This suggests that the complexity of the orbit and the payload size significantly impact launch outcomes.

Launch Success Yearly Trend

This line chart reflects SpaceX's trajectory from early struggles to achieving high reliability in launches.

After 2013, the success rate has improved dramatically, reaching high success from 2016 onward.



All Launch Site Names

The query retrieves all unique launch site names from the SPACEXTABLE by using the GROUP BY clause on the Launch_Site column.

Task 1

Display the names of the unique launch sites in the space mission

```
%sql SELECT Launch_Site FROM SPACEXTABLE GROUP BY Launch_Site
```

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

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

Launch Site Names Begin with 'KSC'

The query retrieves the first 5 records from the SPACEXTABLE where the Launch_Site name starts with "KSC".

The LIKE clause with KSC% is used to match any launch site that begins with "KSC".

The LIMIT 5 clause ensures that only 5 records are returned.

Task 2

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

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'KSC%' LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

The query calculates the total payload mass carried by all boosters launched by NASA (CRS), using the SUM function on the PAYLOAD_MASS__KG_ column.

Task 3

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

```
9]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS_KG FROM SPACEXTABLE
```

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

9]:

TOTAL_PAYLOAD_MASS_KG
619967

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS MEAN_PAYLOAD_MASS_KG_F9 FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%'
```

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

MEAN_PAYLOAD_MASS_KG_F9
2534.6666666666665

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

The AVG function is used to find the mean of the PAYLOAD_MASS__KG_ column for all records where the Booster_Version starts with "F9 v1.1".

First Successful Ground Landing Date

Task 5

List the date where the succesful landing outcome in drone ship was acheived.

Hint:Use min function

```
%sql SELECT MIN(Date) AS SUCCESFUL_LANDING_OUTCOME_IN_DRONE_SHIP FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone sh
```

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

SUCCESFUL_LANDING_OUTCOME_IN_DRONE_SHIP
2016-04-08

The query finds the earliest date when a successful landing on a drone ship was achieved by using the MIN function on the Date column.

The WHERE clause filters for records where the Landing_Outcome is "Success (drone ship)".

Successful Drone Ship Landing with Payload between 4000 and 6000

The query retrieves the booster versions that successfully landed on a ground pad and carried a payload mass between 4,000 kg and 6,000 kg.

The WHERE clause filters for records where the Landing_Outcome is "Success (ground pad)" and the PAYLOAD_MASS__KG_ is greater than 4,000 but less than 6,000.

Task 6

List the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000

```
%%sql
    SELECT Booster_Version
    FROM SPACEXTABLE
    WHERE Landing_Outcome = 'Success (ground pad)'
    AND PAYLOAD_MASS__KG_ > 4000
    AND PAYLOAD_MASS__KG_ < 6000
```

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

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

The query calculates the total number of each type of mission outcome by grouping the results based on the Mission_Outcome column and counting the occurrences of each outcome.

Task 7

List the total number of successful and failure mission outcomes

```
%%sql
        SELECT Mission_Outcome, COUNT(*)
        FROM SPACEXTABLE
    GROUP BY Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

This query identifies the booster versions that carried the maximum payload mass.

It uses a subquery to find the maximum payload mass (MAX(PAYLOAD_MASS__KG_)) from the SPACEXTABLE, then selects booster versions that match this payload mass.

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql
        SELECT Booster_Version
        FROM SPACEXTABLE
    WHERE PAYLOAD_MASS__KG_    = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

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

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2017 Launch Records

The query extracts month numbers and lists booster versions and launch sites for missions with a successful landing on a ground pad in the year 2017.

SUBSTR is used to extract the month (from position 6) and the year (from position 0) since SQLite does not directly support month names.

Task 9

List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

Note: SQLLite does not support monthnames. So you need to use substr(Date,6,2) for month, substr(Date,9,2) for date, substr(Date,0,5),='2017' for year.

```
%%sql
        SELECT SUBSTR(Date,6,2) AS MONTH_NUMBER, Booster_Version, Launch_Site
        FROM SPACEXTABLE
    WHERE Landing_Outcome = 'Success (ground pad)'
    AND SUBSTR(Date,0,5)='2017'
```

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

MONTH_NUMBER	Booster_Version	Launch_Site
02	F9 FT B1031.1	KSC LC-39A
05	F9 FT B1032.1	KSC LC-39A
06	F9 FT B1035.1	KSC LC-39A
08	F9 B4 B1039.1	KSC LC-39A
09	F9 B4 B1040.1	KSC LC-39A
12	F9 FT B1035.2	CCAFS SLC-40

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

The query counts the different types of landing outcomes between 2010-06-04 and 2017-03-20.

It groups the results by Landing_Outcome and orders them in descending order of occurrence.

Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

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

Landing_Outcome	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

All launch sites' locations

The map displays the locations of SpaceX launch sites across the U.S.:

•VAFB SLC-4E is on the west coast near Los Angeles.

•KSC LC-39A and CCAFS SLC-40 are on the east coast near Cape Canaveral.

Act

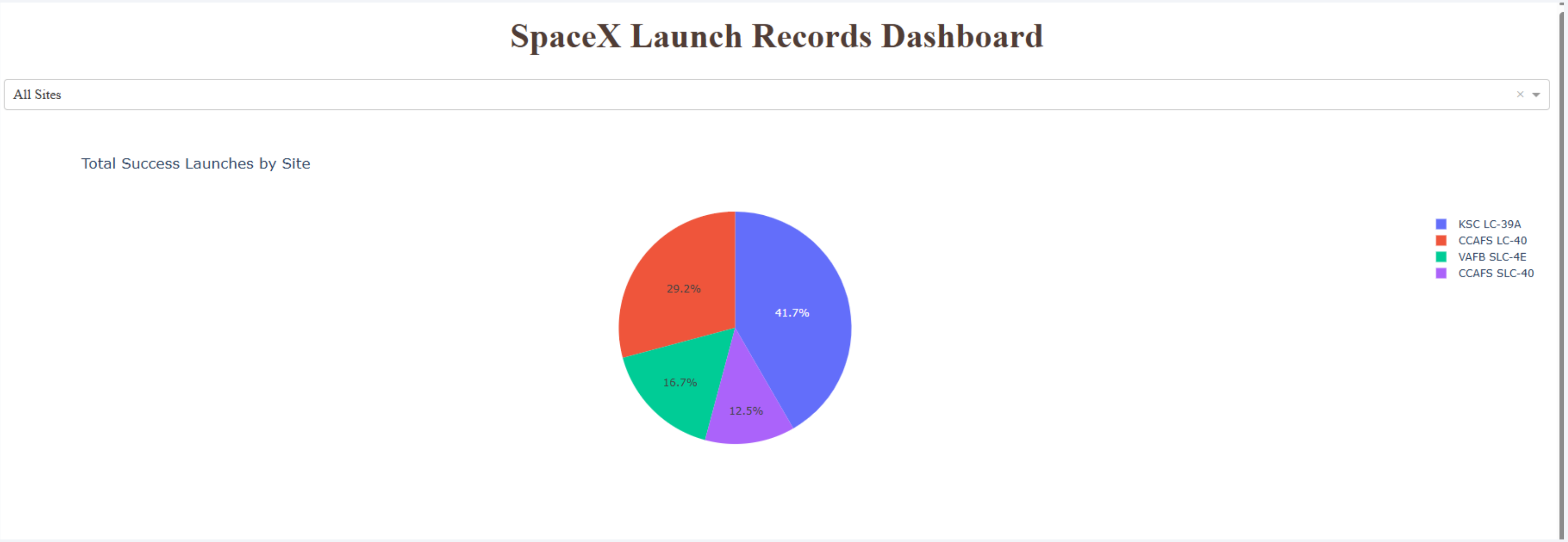
The map shows clustered markers indicating SpaceX's launch activity:

- 46 launches from Florida (KSC LC-39A and CCAFS SLC-40) and 10 launches from VAFB SLC-4E in California.
- The clustering visually represents SpaceX's operational focus and activity levels across launch sites.

Section 4

Build a Dashboard with Plotly Dash

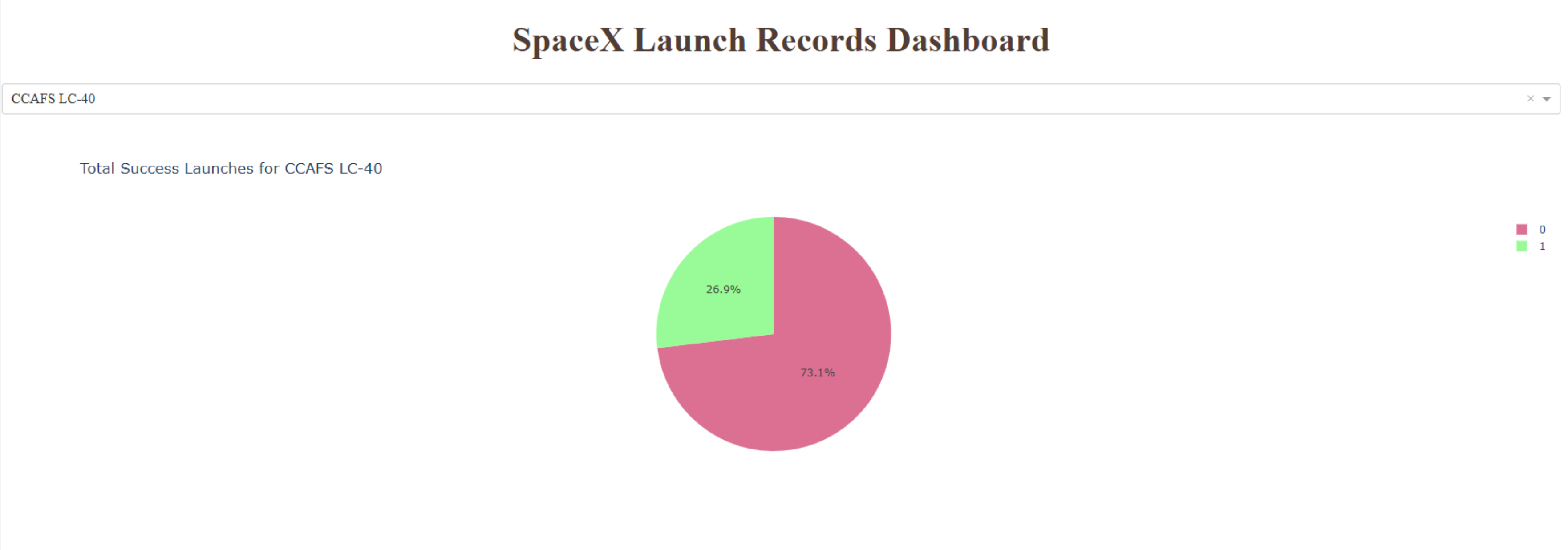
Total success launches by site



The pie chart shows the distribution of successful launches across SpaceX's launch sites:

- KSC LC-39A leads with 41.7% of the total successful launches, highlighting its importance as SpaceX's primary launch site.
- CCAFS LC-40 follows with 29.2%, making it another significant site for SpaceX’s missions.
- VAFB SLC-4E accounts for 16.7%, indicating its role in specialized launches, such as polar orbits.
- CCAFS SLC-40 contributes 12.5% to the success tally, reflecting its complementary role in the overall launch operations.

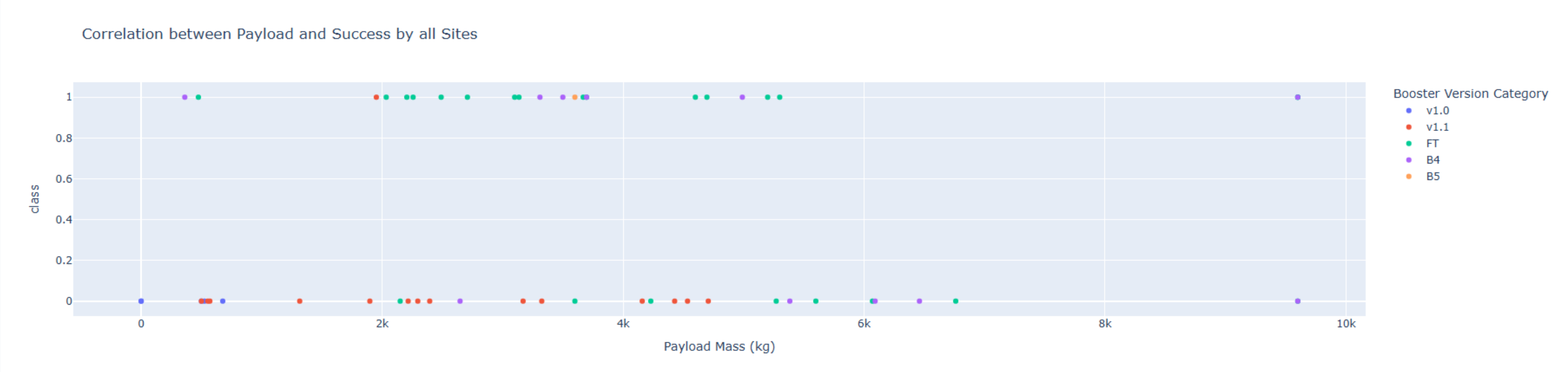
Total success launches for CCAPS LC-40



The pie chart shows the success vs. failure rate of launches specifically at CCAFS LC-40:

- 73.1% failures (indicated in pink) and 26.9% successes (indicated in green).
- This data highlights that CCAFS LC-40 has faced a significant number of unsuccessful launches compared to successful ones.

Correlation between Payload and Success by all Sites



This scatter plot shows that newer booster versions, especially B5, have the highest success rates, even with larger payloads up to 10,000 kg.

In contrast, older versions like v1.0 and v1.1 have a higher frequency of failures, particularly with payloads below 6,000 kg.

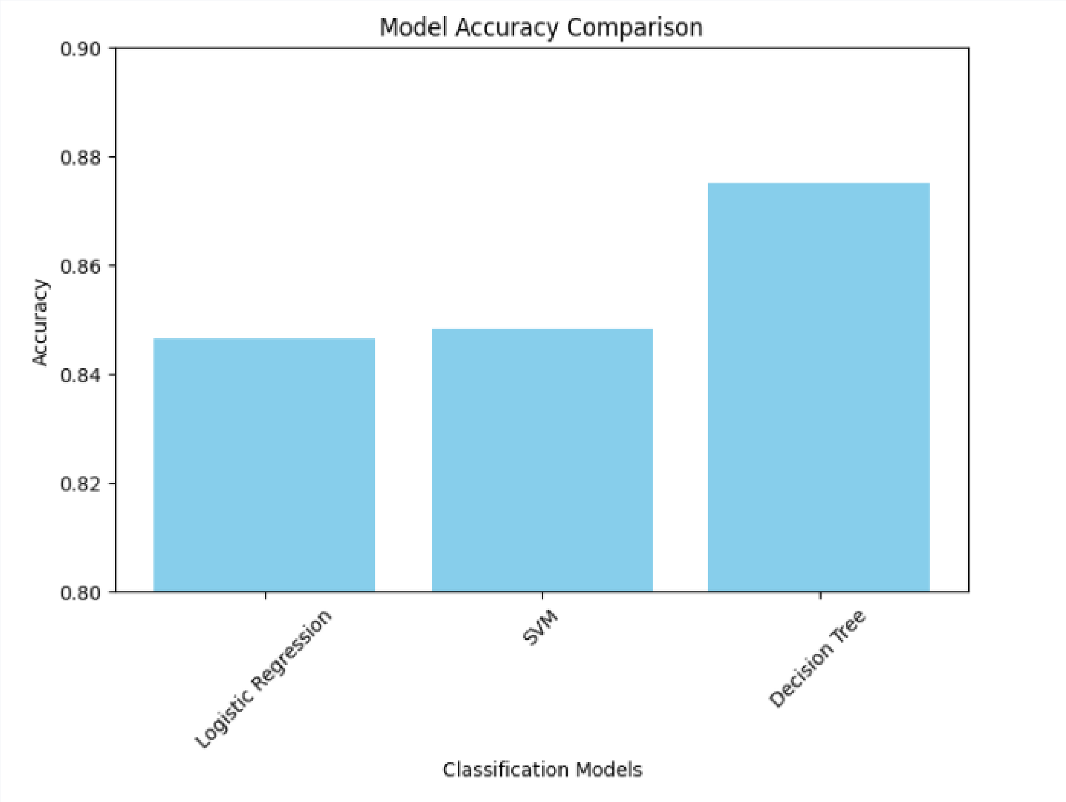
The 2,000 to 6,000 kg payload range demonstrates the most consistent success across different boosters, highlighting it as an optimal range for mission reliability.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

This bar chart represent the accuracy of each model, and it shows that the Decision Tree classifier outperforms the others in this dataset with an accuracy of 0.875.

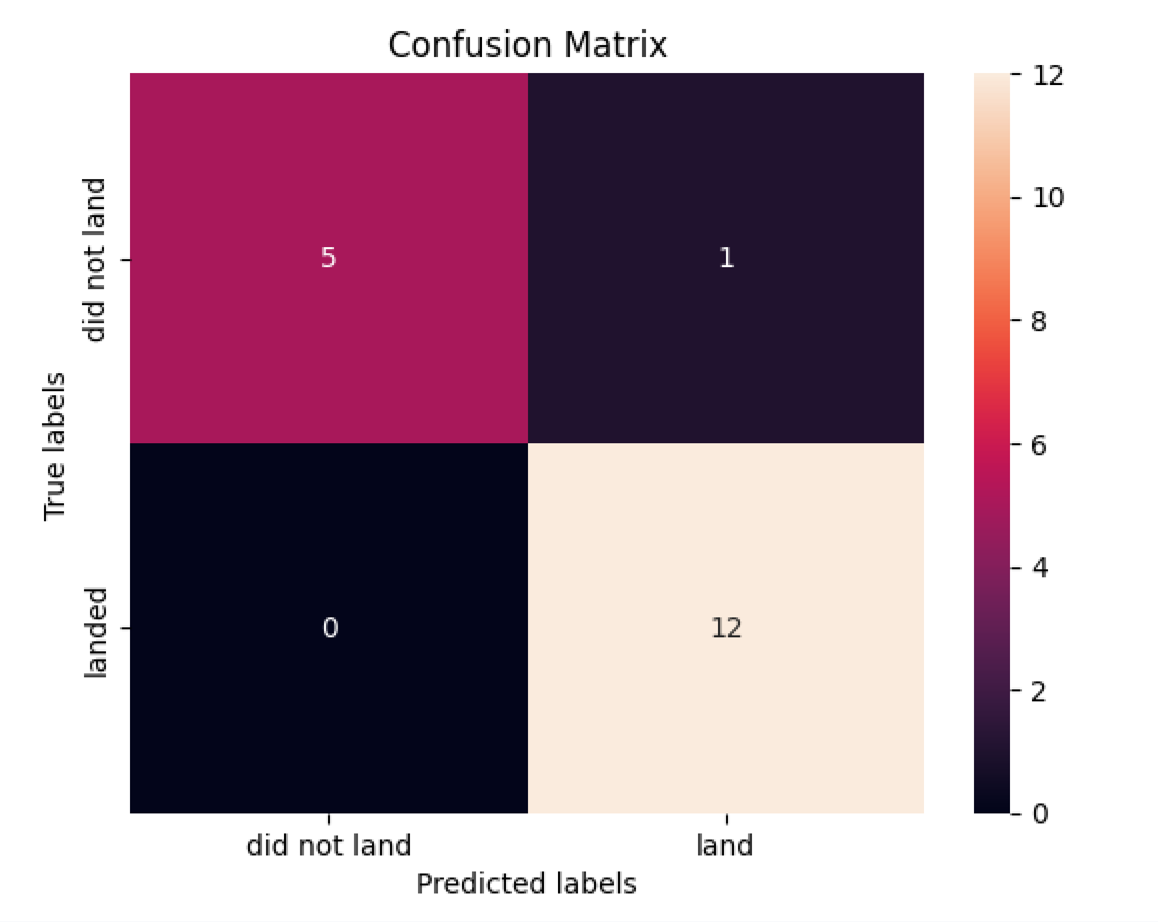


Confusion Matrix

The Decision Tree model, with an accuracy of 0.875, is the best performer. Its confusion matrix shows:

- 12 true positives (correctly predicted successful landings)
- 5 true negatives (correctly predicted unsuccessful landings)
- 1 false positive (incorrectly predicted a landing that did not occur)
- 0 false negatives (no missed successful landings)

This matrix highlights the model's strong ability to predict landings accurately, with only a minor tendency to overestimate success.



Conclusions

- **Data Collection:**
  - Data was collected using SpaceX's API and web scraping methods, focusing on launch details such as payload, launch site, and landing outcomes.
- **Data Wrangling and Cleaning:**
  - Prepared and cleaned data for analysis by handling missing values and standardizing formats to ensure consistency.
- **Exploratory Data Analysis (EDA):**
  - Analyzed trends in launch success across various sites and payload ranges.
  - Visualized data using scatter plots to understand relationships between flight numbers, payload mass, and success rates.
  - Key findings include improved launch success rates over time and higher success with heavier payloads in newer booster versions.
- **Model Building and Evaluation:**
  - Built and tuned Logistic Regression, SVM, and Decision Tree models using GridSearchCV for hyperparameter optimization.
  - Evaluated models using accuracy scores and confusion matrices to compare performance.

Conclusions

- **Model Comparison:**
  - The Decision Tree model achieved the highest accuracy (0.875), outperforming Logistic Regression and SVM.
  - The Decision Tree's confusion matrix showed a high rate of correct predictions with minimal false positives.
- **Interactive Data Visualization:**
  - Built interactive visualizations using Plotly Dash to explore launch success by site and payload range.
  - Created a Folium map to display launch site locations and launch distributions geographically.
- **Insights from Visualizations:**
  - Most successful launches occurred from KSC LC-39A, while CCAFS LC-40 faced more challenges.
  - Success rates increased with the introduction of newer boosters like B5, demonstrating technological advancements.
  - Optimal payload range for consistent launch success was identified between 2,000 to 6,000 kg.
- **Predictive Analysis:**
  - Identified key factors influencing launch success, including booster version and payload mass.
  - The analysis revealed that newer booster versions, particularly B5, had a significant impact on improving launch reliability.

Thank you!