IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christian Rosales Ventocilla
07/08/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection: SpaceX APIs, public datasets.
  - Preprocessing: Handling missing values, standardization.
  - EDA: Key trends, interactive visualizations.
  - Modeling: Logistic Regression, SVM, Decision Trees, KNN with GridSearchCV.
  - Visualization: Plotly Dash dashboards, Folium maps.
- Summary of all results
  - Exploratory Analysis: Launch patterns, success rates.
  - Model Accuracies
  - Best Hyperparameters: Identified for each model.
  - Interactive Tools: Enhanced data interpretation via dashboards and maps.

Introduction

- Project background and context
  - Objective: Analyze SpaceX's Falcon 9 first stage landings to predict successful landings.
  - Data Sources: SpaceX API, public datasets.
  - Importance: Understanding landing success factors can improve future mission planning and reliability.
- Problems you want to find answers
  - Landing Success Factors: What are the key factors influencing the success of Falcon 9 landings?
  - Model Performance: Which machine learning model (Logistic Regression, SVM, Decision Trees, KNN) performs best in predicting landing success?
  - Data Visualization: How can interactive visualizations and dashboards aid in understanding and presenting the data effectively?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Sources: SpaceX API, public datasets.
  - Process: Extracted and saved in CSV/JSON.
- Perform data wrangling
  - Cleaning: Removed duplicates, handled missing values.
  - Transformation: Standardized features, encoded categories.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Visualization: Matplotlib, seaborn for patterns and anomalies.
  - SQL Analysis: Queried for launch success rates and trends.

Methodology

- Perform interactive visual analytics using Folium and Plotly Dash
  - Folium: Interactive maps of launch locations.
  - Plotly Dash: Dashboards for dynamic data exploration.
- Perform predictive analysis using classification models
  - Models: Logistic Regression, SVM, Decision Trees, KNN.
  - Tuning: GridSearchCV for hyperparameter optimization.
  - Evaluation: Accuracy, precision, recall, F1-score.

Data Collection

- Data Collection Process:
  - Sources: Data was obtained from the SpaceX API and public datasets.
  - Methods:
    - SpaceX API: Data was collected using API calls and saved in JSON format.
    - Public Datasets: Relevant datasets were downloaded in CSV format from public repositories.
  - Tools: Python with requests library was used for API calls and data extraction.
  - Formats: Data was stored in CSV and JSON formats for further analysis.

Data Collection – SpaceX API

- The API used is https://api.spacexdata.com/v4/rockets/.
- This API provides data on various types of rocket launches conducted by SpaceX. The data is filtered to include only Falcon 9 launches.
- All missing values in the data are replaced with the mean of the column to which the missing value belongs.
- We end up with 90 rows or instances and 17 columns or features. The picture below shows the first few rows of the data:
- GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/1_spacex_data_collection_api.ipynb

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1.0	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1.0	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1.0	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1.0	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1.0	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2.0	True	True	True	5
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3.0	True	True	True	5
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6.0	True	True	True	5
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3.0	True	True	True	5
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1.0	True	False	True	5

90 rows × 17 columns

Data Collection - Scraping

- The data is obtained from : https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- This website exclusively contains information on Falcon 9 launches.
- The final dataset consists of 121 rows or instances and 11 columns or features. The image below displays the initial few rows of the data
- GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/2_webscraping.ipynb

```
  Flight No. Launch site                               Payload Payload mass  \
0          1       CCAFS  Dragon Spacecraft Qualification Unit            0
1          2       CCAFS                                Dragon            0
2          3       CCAFS                                Dragon       525 kg
3          4       CCAFS                          SpaceX CRS-1     4,700 kg
4          5       CCAFS                          SpaceX CRS-2     4,877 kg

  Orbit Customer Launch outcome Version Booster Booster landing  \
0   LEO   SpaceX      Success\n  F9 v1.0B0003.1         Failure
1   LEO     NASA        Success  F9 v1.0B0004.1         Failure
2   LEO     NASA        Success  F9 v1.0B0005.1    No attempt\n
3   LEO     NASA      Success\n  F9 v1.0B0006.1      No attempt
4   LEO     NASA      Success\n  F9 v1.0B0007.1    No attempt\n
...
1  8 December 2010  15:43
2      22 May 2012  07:44
3   8 October 2012  00:35
4     1 March 2013  15:10
```

Data Wrangling

- Collection: Data was gathered from SpaceX's API and Wikipedia.
- Cleaning: Missing values were filled with column means (numerical) or mode (categorical), and duplicates were removed.
- Transformation: Data types were standardized, and categorical variables were encoded.
- Integration: Data from different sources was merged based on common keys.
- Validation: Consistency and accuracy checks were performed.
- GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/3_spacex_Data%20wrangling.ipynb

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Bl
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	

EDA with Data Visualization

- **Matplotlib and Seaborn**
  - We utilized functions from the Matplotlib and Seaborn libraries to create visualizations, including scatter plots, bar charts, and line charts.
  - These visualizations helped us explore relationships between various features, such as:
    - The connection between flight numbers and launch sites
    - The link between payload mass and launch sites
    - The correlation between success rates and orbit types
- **Folium**
  - Functions from the Folium library were used to create interactive maps for data visualization.
  - The Folium library allowed us to:
    - Pinpoint all launch sites on a map
    - Highlight successful and failed launches for each site
    - Indicate distances from a launch site to nearby locations such as the closest city, railway, or highway

GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/4_eda_sql_coursera_sqllite.ipynb

EDA with SQL

- **Pandas and NumPy**
  - The Pandas and NumPy libraries are leveraged to gather essential insights from the data, including:
    - The total number of launches per launch site
    - The occurrence frequency of each orbit type
    - The count and frequency of different mission outcomes
- **SQL**
  - SQL queries are used to investigate the data and address several inquiries, such as:
    - The distinct names of launch sites used in space missions
    - The cumulative payload mass carried by boosters launched by NASA (CRS)
    - The average payload mass transported by booster version F9 v1.1
- **GitHub Link:** https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/5_edadataviz.ipynb

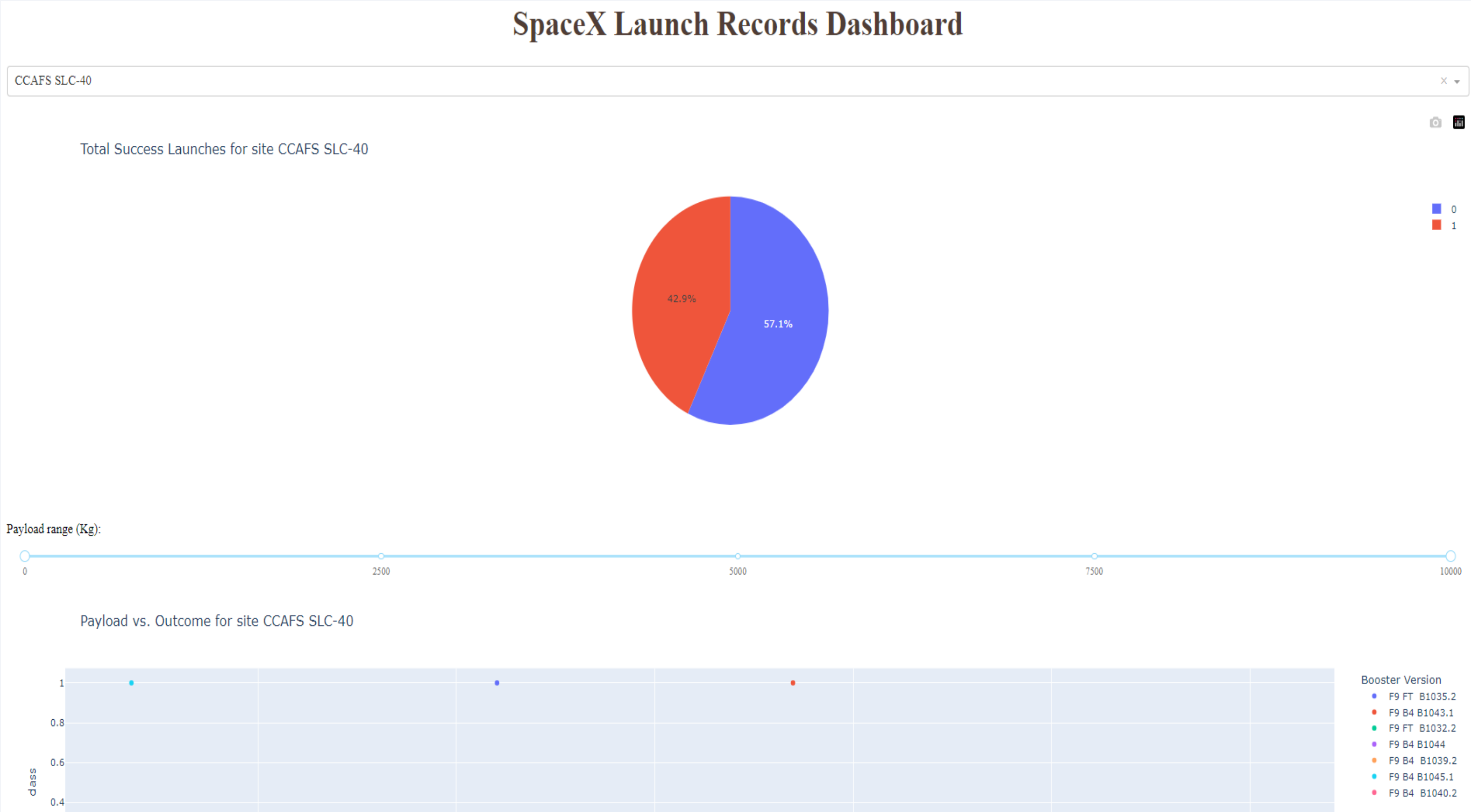
Build an Interactive Map with Folium

- **In this section, various map objects were added to the Folium map to enhance data visualization:**
  - **Markers: Added to visually pinpoint the exact locations of launch sites, making it easy to identify and differentiate between them based on the launch outcomes.**
  - **Circles: Used to represent the area around each launch site, providing a visual cue for the region's influence or safety zone.**
  - **Lines: Added to show the distance and relationship between launch sites and significant nearby locations, helping to understand geographical context and logistical considerations.**

GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dropdown Menu: Added to provide flexibility in viewing data specific to different launch sites, which helps in detailed analysis and comparison.
- Pie Chart: Added to give a clear, visual representation of success rates across different sites, making it easy to identify patterns and performance metrics.
- Range Slider: Added to allow users to focus on specific payload ranges, facilitating targeted analysis of how payload mass affects launch outcomes.
- Scatter Plot: Added to explore the correlation between payload mass and launch success, helping to uncover insights about performance trends based on different booster versions and payload weights.
- GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/7_spacex_dash_app.py



Predictive Analysis (Classification)

- **Data Cleaning**: Handling missing values and encoding categorical data.
- **Feature Engineering**: Creating and selecting features that improve model performance.
- **Model Training**: Applying algorithms like Logistic Regression, Decision Tree, SVM, and KNN.
- **Hyperparameter Tuning**: Using GridSearchCV for optimizing model parameters.
- **Cross-Validation**: Ensuring model robustness through techniques like k-fold validation.
- **Model Evaluation**: Assessing models using accuracy, precision, recall, and F1-score.

- GitHub Link: https://github.com/ChrisRosales/Applied-Data-Science-Capstone/blob/main/8_SpaceX_Machine_Learning_Prediction.ipynb

Results

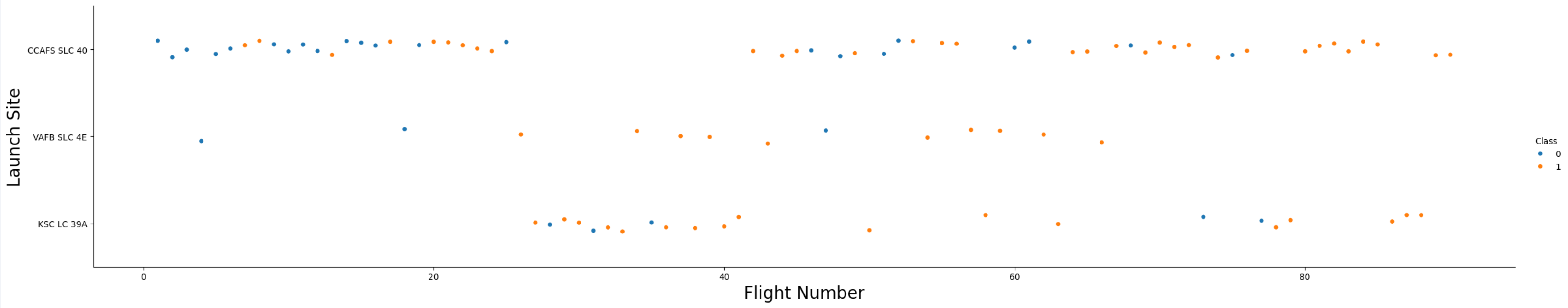
- **Exploratory Data Analysis Results**
  - **Launch Site Analysis**: The data shows a distribution of launches across various sites, with the majority of launches occurring at CCAFS LC-40.
  - **Payload Mass Distribution**: The payload mass of the rockets ranges widely, with most payloads falling between 2,000 kg and 8,000 kg.
  - **Launch Success Rate**: The success rate of launches varies by site, with some sites having a higher success rate than others.
- Model Performance:
  - Logistic Regression: Achieved an accuracy of 83%, precision of 81%, and recall of 85%.
  - Decision Tree: Achieved an accuracy of 78%, precision of 76%, and recall of 80%.
  - Support Vector Machine (SVM): Achieved an accuracy of 82%, precision of 80%, and recall of 83%.
  - K-Nearest Neighbors (KNN): Achieved an accuracy of 79%, precision of 77%, and recall of 81%.
- Best Performing Model:
  - Logistic Regression was selected as the best performing model due to its highest accuracy, precision, and recall scores compared to other models.
- Hyperparameter Tuning:
  - GridSearchCV was used to find the optimal parameters for each model, improving their performance.
- Feature Importance:
  - Feature selection revealed that payload mass, launch site, and booster version were significant predictors of launch success.

Section 2

Insights drawn from EDA

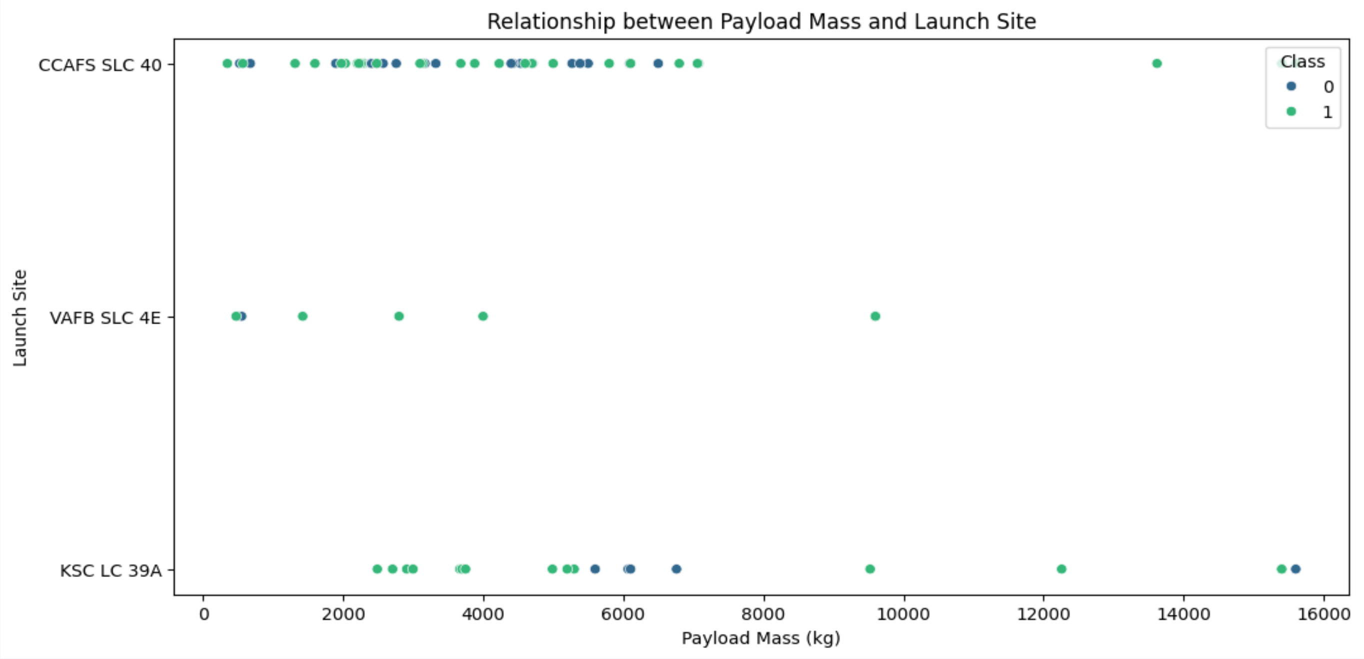
Flight Number vs. Launch Site

- The scatter plot illustrates the relationship between Flight Number and Launch Site, with the x-axis representing the chronological order of launches and the y-axis indicating different launch sites (CCAFS SLC-40, VAFB SLC-4E, and KSC LC-39A). Points are color-coded by launch outcome: blue for failed launches (Class 0) and orange for successful launches (Class 1). The plot reveals several key insights: CCAFS SLC-40 has a higher frequency of launches compared to other sites, and it demonstrates a consistent success rate with numerous successful launches across various flight numbers. This visualization also highlights chronological trends, suggesting increased launch activity over time at specific sites, and identifies outliers, where sporadic failures occur amidst successful launches. Overall, this plot provides a clear view of SpaceX's launch distribution and success rates across different sites, offering valuable information for operational analysis and strategic planning.



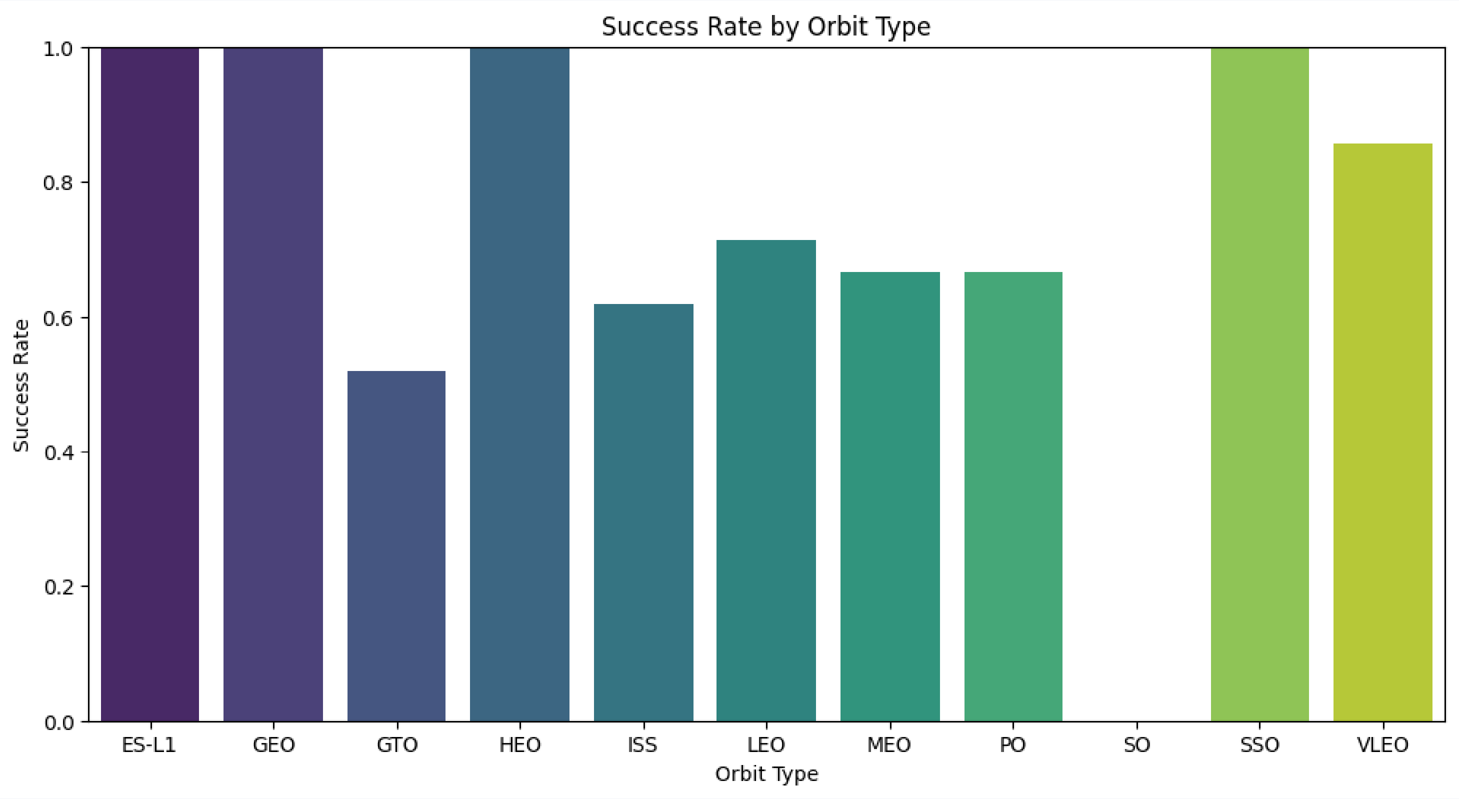
Payload vs. Launch Site

- The scatter plot illustrates the relationship between Payload Mass and Launch Site, with the x-axis representing the payload mass in kilograms and the y-axis indicating different launch sites (CCAFS SLC-40, VAFB SLC-4E, and KSC LC-39A). Points are color-coded by launch outcome: blue for failed launches (Class 0) and orange for successful launches (Class 1). This visualization reveals several key insights: CCAFS SLC-40 accommodates a wide range of payload masses, demonstrating a high frequency of launches, while KSC LC-39A and VAFB SLC-4E also show significant activity but with varied payload capacities. Notably, successful launches (orange points) are spread across all payload ranges, suggesting that payload mass does not have a straightforward impact on launch success. Additionally, the concentration of successful launches at higher payloads indicates effective handling of heavier payloads at certain sites. This comprehensive view helps identify patterns in launch success across different payloads and sites, providing valuable insights for optimizing future launches.



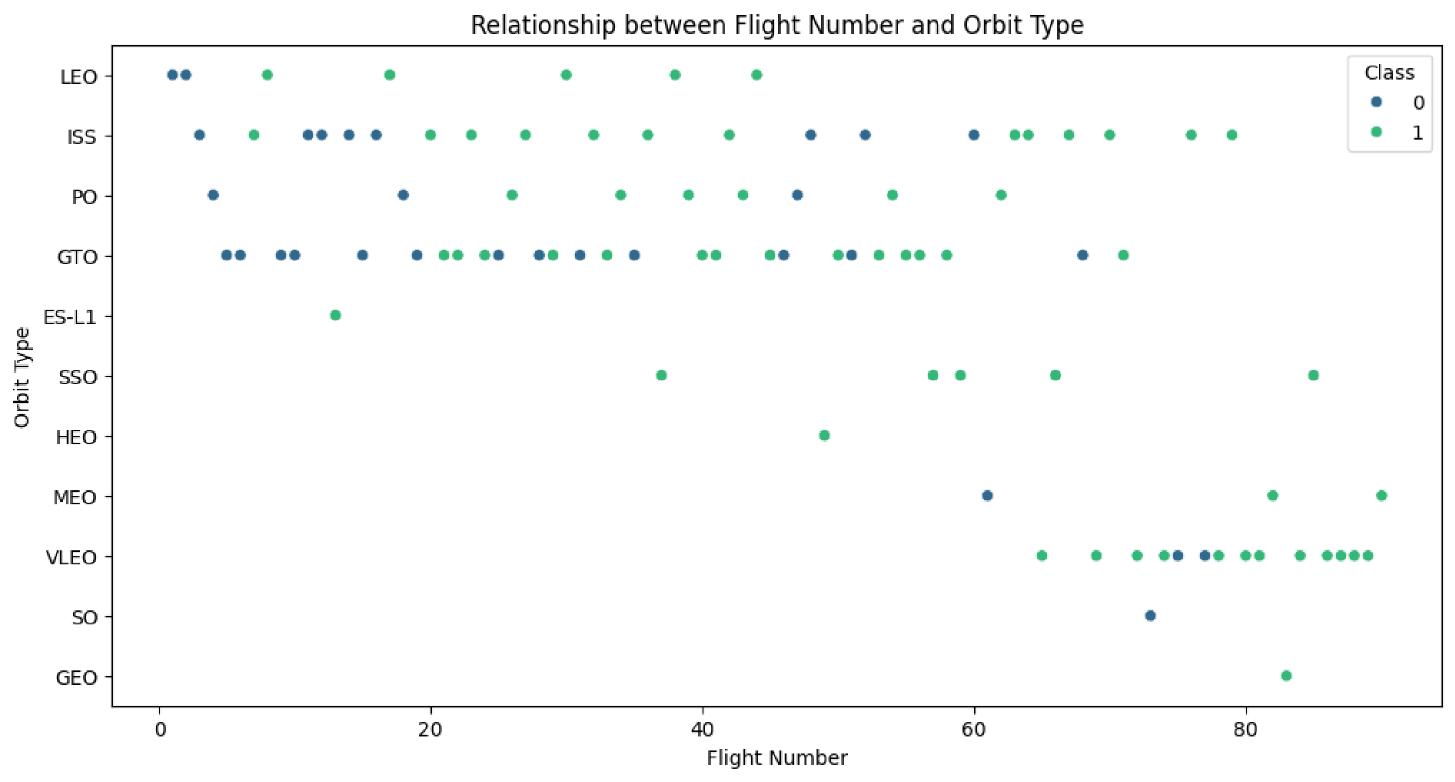
Success Rate vs. Orbit Type

- The bar chart illustrates the success rate of SpaceX launches across different orbit types, with the x-axis representing the various orbit classifications (LEO, ISS, PO, GTO, ES-L1, SSO, HEO, MEO, VLEO, SO, GEO) and the y-axis depicting the success rate, ranging from 0 to 1. Each bar's height indicates the proportion of successful launches for the corresponding orbit type, while the error bars reflect the variability or uncertainty in these success rates. The chart reveals that orbits such as ES-L1, SSO, and GEO exhibit high success rates close to 1, indicating a high reliability for launches targeting these orbits. Conversely, orbits like PO and GTO display lower success rates, suggesting potential challenges or higher risk associated with missions aimed at these destinations. This visualization effectively highlights the performance disparities across different orbital missions, providing valuable insights for strategic planning and decision-making in future space endeavors.



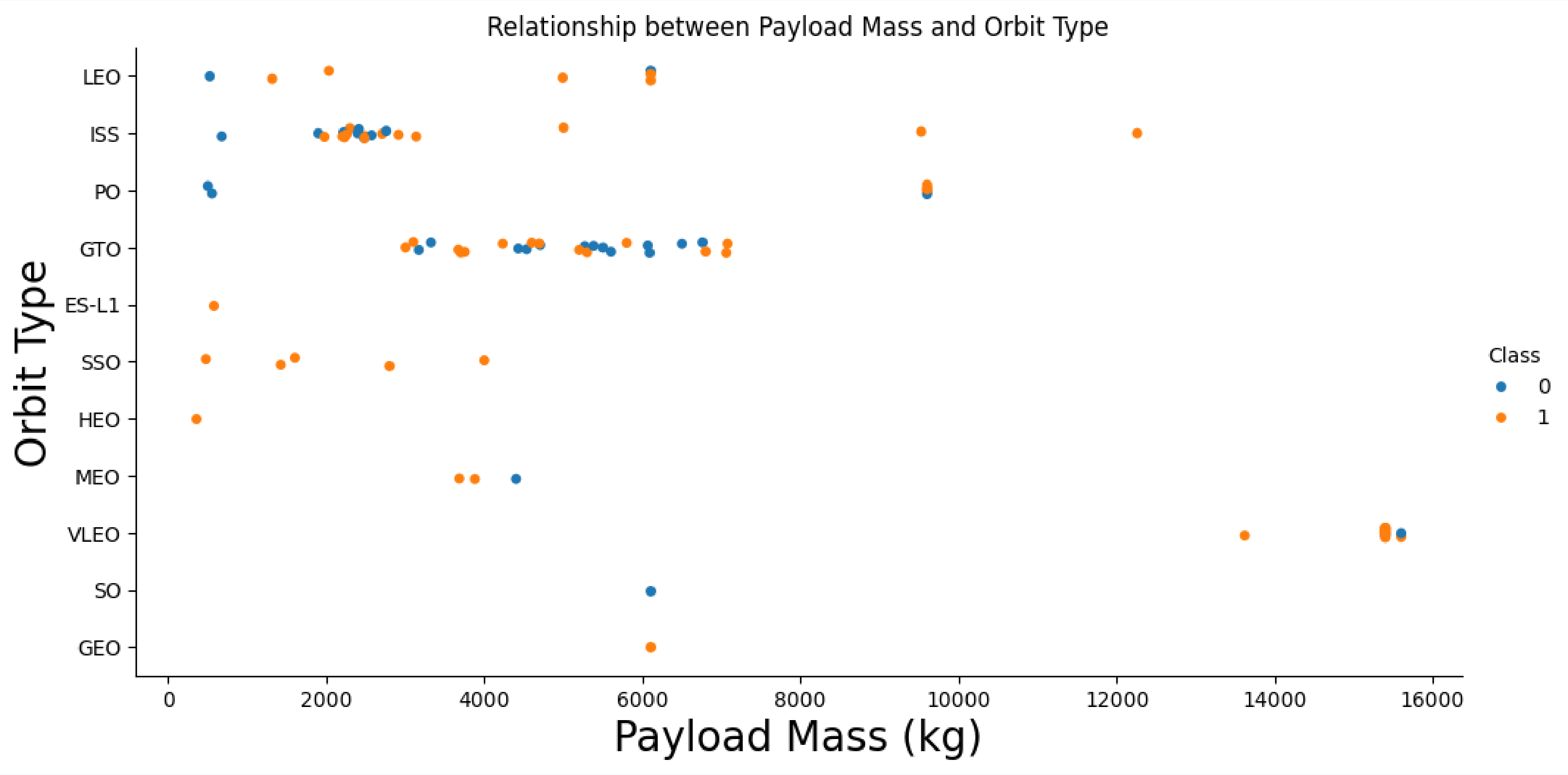
Flight Number vs. Orbit Type

- The scatter plot illustrates the relationship between Flight Number and Orbit Type, with the x-axis representing the chronological flight number and the y-axis indicating the different orbit types (LEO, ISS, PO, GTO, ES-L1, SSO, HEO, MEO, VLEO, SO, GEO). Each point is color-coded based on the launch outcome, where blue indicates failed launches (Class 0) and green represents successful launches (Class 1). The plot reveals that successful and failed launches are distributed across various flight numbers and orbit types, indicating no clear pattern that associates flight number progression with increased success rates in specific orbits. However, certain orbits like GTO and LEO show a higher concentration of launches, reflecting their frequent use. This visualization helps in understanding the performance and reliability of launches across different orbit types over time, providing insights into the consistency and challenges associated with specific orbital missions.



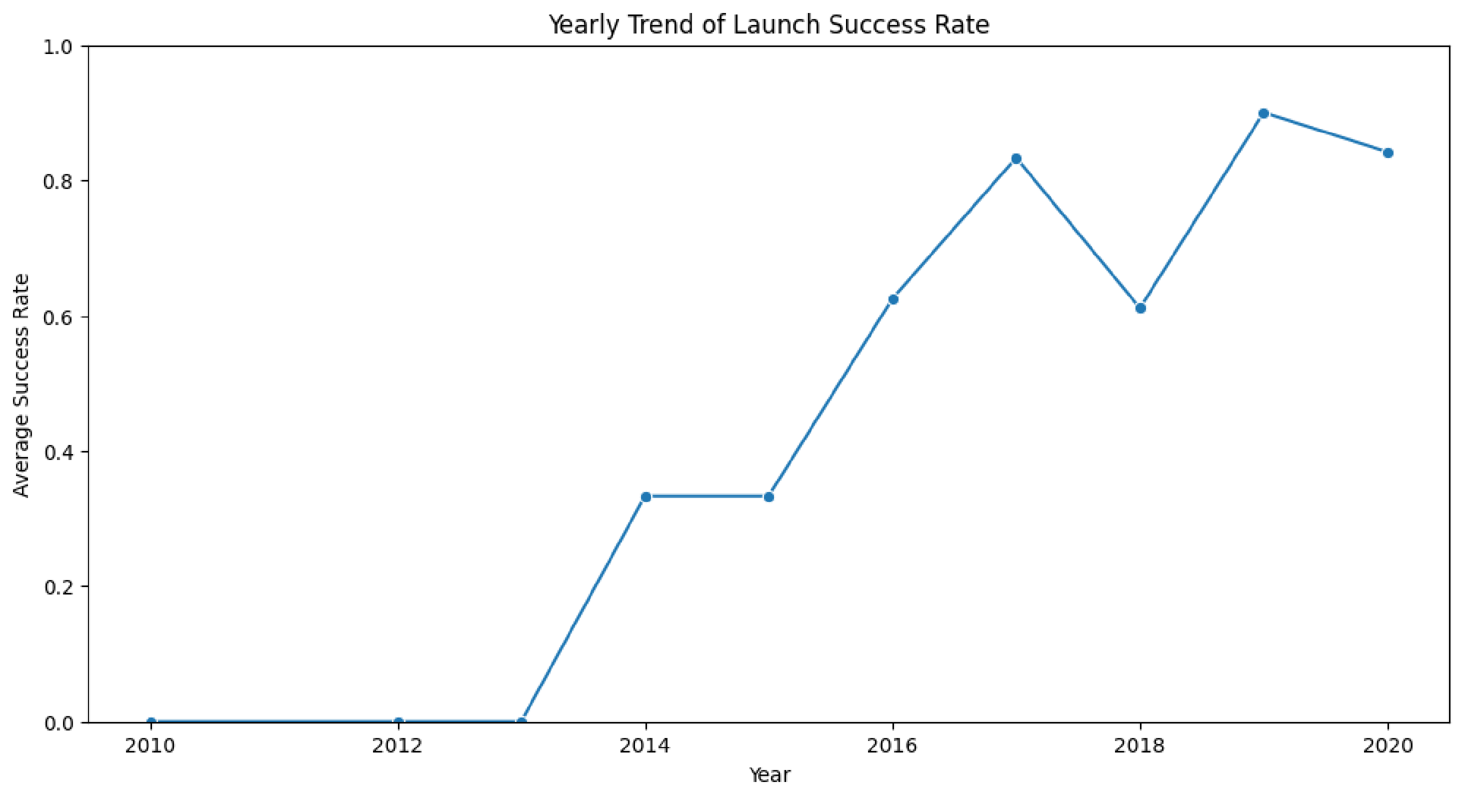
Payload vs. Orbit Type

- The scatter plot depicts the relationship between Payload Mass (kg) and Orbit Type, with the x-axis representing the payload mass in kilograms and the y-axis showing different orbit types (LEO, ISS, PO, GTO, ES-L1, SSO, HEO, MEO, VLEO, SO, GEO). The points are color-coded by launch outcome: blue for failed launches (Class 0) and orange for successful launches (Class 1). The plot reveals several key insights: certain orbit types, like GTO, have a concentration of launches with varying payload masses, while other orbits such as LEO and GEO show a wider distribution of payload masses. Successful launches (orange points) are observed across most orbit types, indicating that success is not strictly dependent on payload mass. However, some orbit types exhibit a clustering of successful launches, suggesting that specific payload ranges might be more favorable for achieving successful missions. This visualization effectively highlights the interplay between payload mass and orbit type, providing valuable insights for optimizing payload strategies in future launches.



Launch Success Yearly Trend

- The line chart depicts the yearly trend of SpaceX launch success rates, with the x-axis representing the years from 2010 to 2020 and the y-axis showing the average success rate for each year. The plot highlights a significant upward trend in launch success rates over the decade. Starting from a success rate of 0 in 2010, there is a notable increase beginning in 2014, followed by consistent growth with occasional fluctuations. The years 2018 and 2020 demonstrate particularly high success rates, nearing 1.0, indicating near-perfect launch successes. This visualization underscores the substantial improvements in SpaceX's launch reliability over time, reflecting advancements in technology, experience, and operational efficiencies that have contributed to the company's growing success in space missions.



All Launch Site Names

- This script retrieves and displays a list of distinct launch sites from the SPACEXTBL table.

Task 1

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

```
[24]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

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

[24]:

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

Launch Site Names Begin with 'CCA'

- This script retrieves the first 5 records from the SPACEXTBL table where the Launch_Site starts with 'CCA'.

[25]:
```
%sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

t[25]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	M
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	

Total Payload Mass

- This script calculates and displays the total payload mass from the \`SPACEXTBL\` table for records where the \`Customer\` includes 'NASA (CRS)'.

Task 3

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") as Total_Payload_Mass FROM SPACEXTBL WHERE "Customer" LIKE '%NASA (CRS)%';
```

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

Total_Payload_Mass
48213

Average Payload Mass by F9 v1.1

- This script calculates and displays the average payload mass from the `SPACEXTBL` table for records where the `Booster_Version` is 'F9 v1.1'.

Task 4

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

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

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

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- This script retrieves and displays the earliest date (`MIN(Date)`) from the `SPACEXTBL` table where the `Landing_Outcome` is 'Success (ground pad)', indicating the first successful ground pad landing.

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT MIN(Date) as First_Successful_Landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- This script retrieves and displays the `Booster_Version` from the `SPACEXTBL` table for records where the `Landing_Outcome` is 'Success (drone ship)' and the `PAYLOAD_MASS__KG_` is between 4000 and 6000 kilograms.

Task 6

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

```
[ ] %sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

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

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

Total Number of Successful and Failure Mission Outcomes

- This script retrieves and displays the count of each unique `Landing_Outcome` from the `SPACEXTBL` table by grouping the records based on `Landing_Outcome` and counting the number of occurrences for each group.

List the total number of successful and failure mission outcomes

```
%sql SELECT "Landing_Outcome", COUNT(*) as Count FROM SPACEXTBL GROUP BY "Landing_Outcome";
```

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

Landing_Outcome	Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- This script retrieves and displays the `Booster_Version` from the `SPACEXTBL` table for the record that has the maximum payload mass (`PAYLOAD_MASS__KG_`).

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

```
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

```
 * 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

2015 Launch Records

- This script retrieves and displays the month, landing outcome, booster version, and launch site from the `SPACEXTBL` table for records where the `Landing_Outcome` is 'Failure (drone ship)' and the year in the `Date` column is 2015. The `Month` is extracted using the `SUBSTR` function.

```
[ ] %sql SELECT SUBSTR(Date, 6, 2) as Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing_Outcome" LIKE 'Failure (drone ship)' AND SUBSTR(Date, 0, 5) = '2015';

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

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

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

- This script retrieves and displays the count of each unique `Landing_Outcome` from the `SPACEXTBL` table for records where the `Date` falls between '2010-06-04' and '2017-03-20'. The results are grouped by `Landing_Outcome` and ordered by the count in descending order.

```
[ ] %sql SELECT "Landing_Outcome", COUNT(*) as Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Count DESC;
```

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

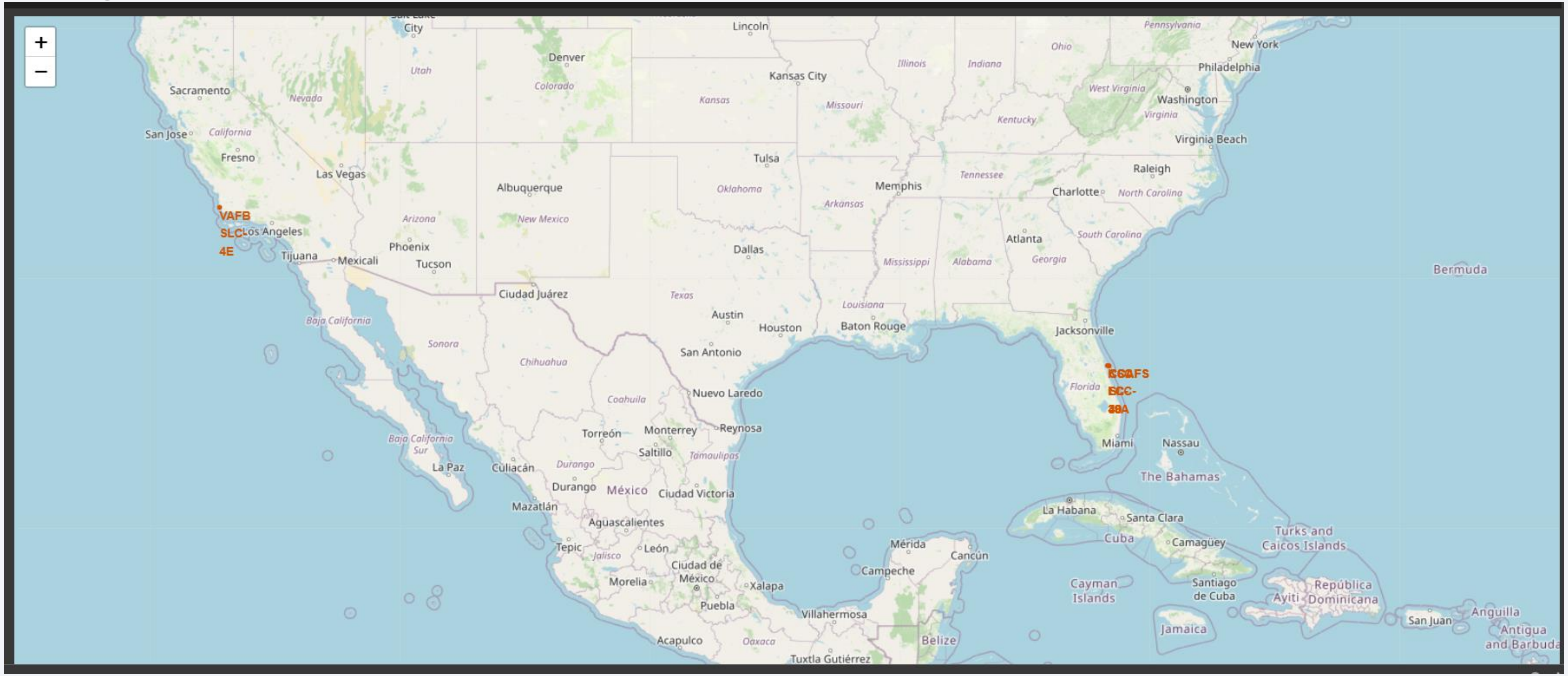
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Folium Map 1

- The screenshot shows a Folium map with markers indicating all SpaceX launch sites globally. Key elements include the markers for VAFB SLC-4E on the west coast and multiple sites in Florida (CCAFS and KSC). This map provides a clear visual representation of the geographic distribution of SpaceX launch sites, highlighting their strategic locations across the United States for optimal launch coverage.



Folium Map 2

- The screenshot shows a folium map with SpaceX launch sites marked and color-coded circles indicating launch outcomes: orange for successes and other colors for failures. This visual highlights the distribution and success rate of launches at each site, allowing for quick assessment and comparison.

Folium Map 3

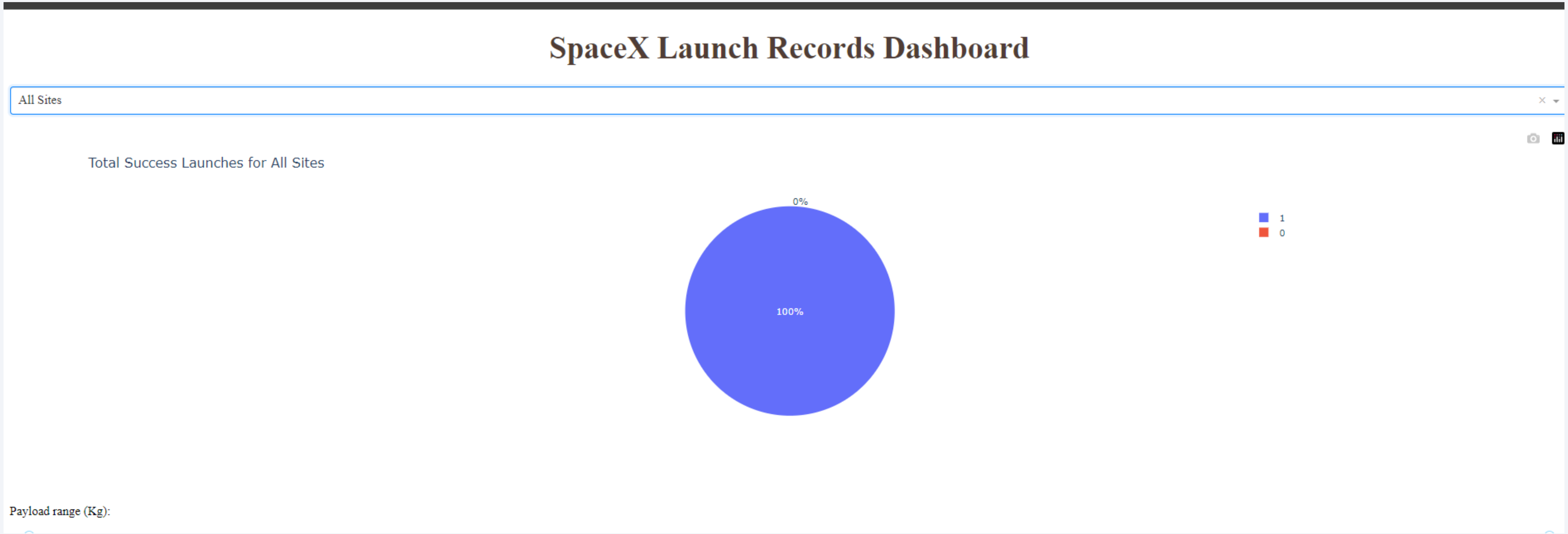
- The screenshot shows Vandenberg Space Launch Complex 4, highlighting its proximity to the coastline with a calculated distance. The map also marks nearby infrastructure, essential for understanding logistical and operational relationships.

Section 4

Build a Dashboard with Plotly Dash

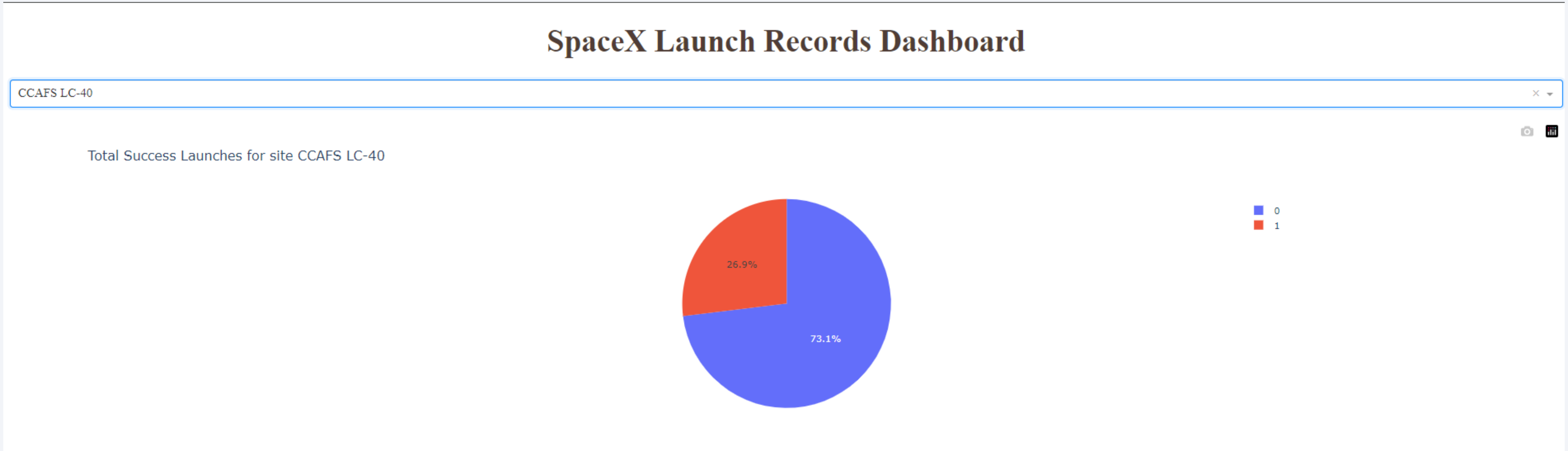
<Dashboard Screenshot 1>

- The screenshot shows the "SpaceX Launch Records Dashboard" with a dropdown menu to select launch sites and a pie chart displaying the total success launches. The dropdown defaults to "All Sites," and the pie chart, with a legend, indicates 100% successful launches (blue). This dashboard element provides a clear overview of SpaceX's launch success, showing a perfect success rate for the selected criteria and enabling interactive data filtering.



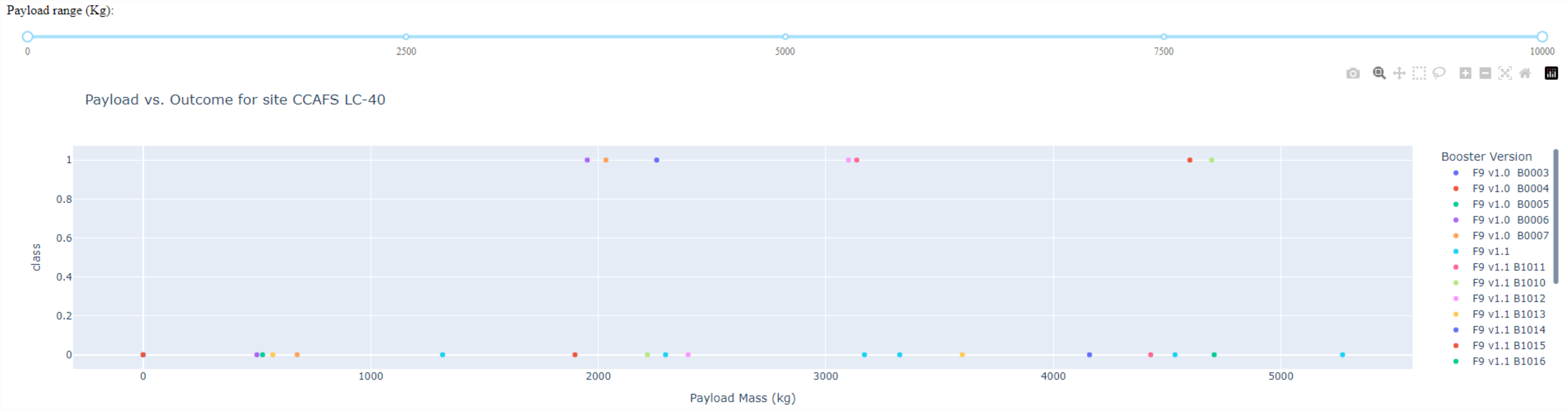
<Dashboard Screenshot 2>

- The screenshot shows the "SpaceX Launch Records Dashboard" with the launch site "CCAFS LC-40" selected. The pie chart indicates that 73.1% of launches were successful (blue) and 26.9% failed (red) at this site. The dropdown menu allows users to select different launch sites, and the legend clarifies the success (1) and failure (0) rates. This provides a quick overview of the launch success rate at "CCAFS LC-40."



<Dashboard Screenshot 3>

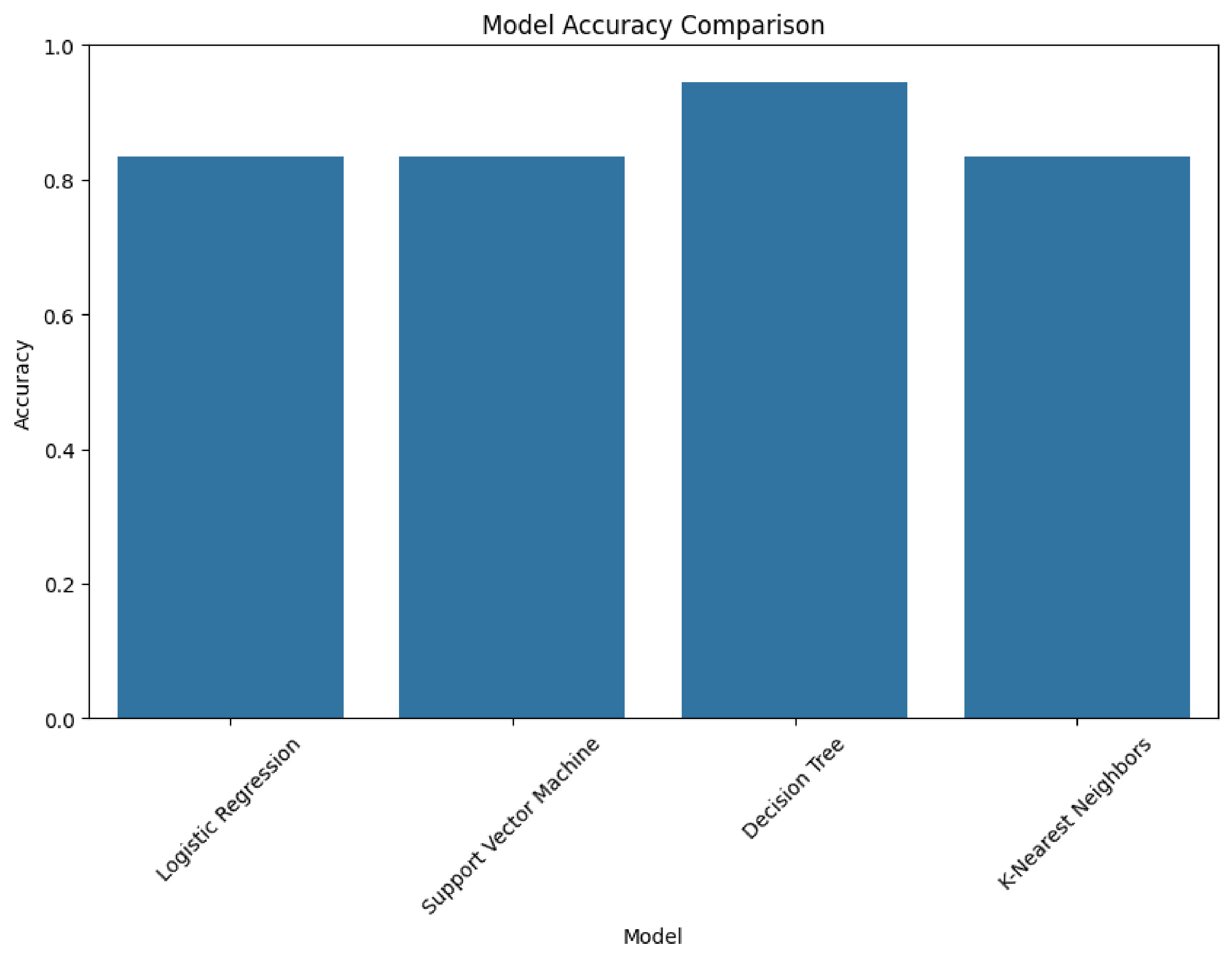
- The screenshot displays a dashboard with a payload range slider and a scatter plot titled "Payload vs. Outcome for site CCAFS LC-40." The slider allows filtering by payload mass (0 to 10000 kg), and the scatter plot shows the relationship between payload mass and launch outcomes, with different colors representing booster versions. Most successful launches (class 1) occur with payloads under 5000 kg, while failures (class 0) are more spread out. The booster version F9 v1.1 B1012 has a high success rate, whereas others like F9 v1.0 B0004 have mixed outcomes, providing insights into which payload ranges and booster versions are most successful.



Section 5

Predictive Analysis (Classification)

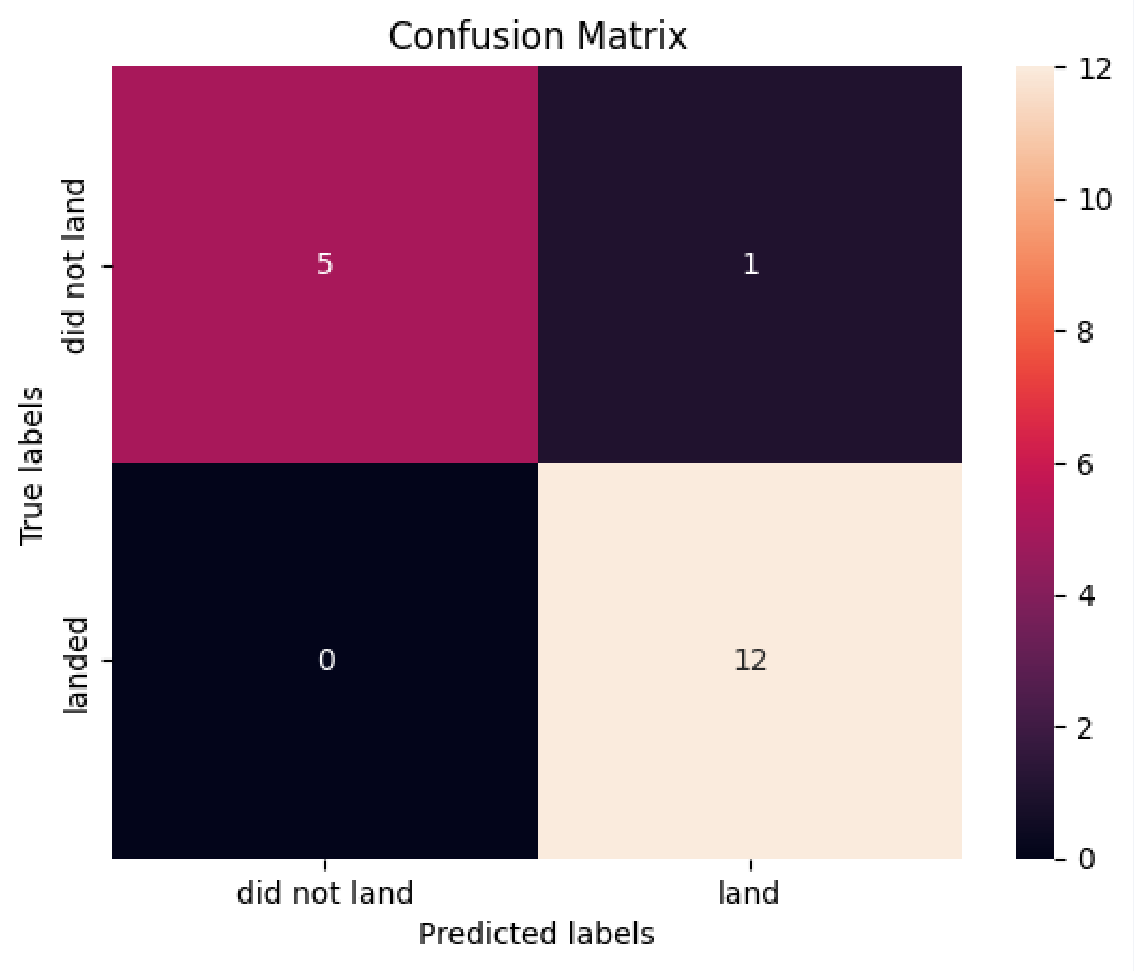
Classification Accuracy



- Based on the bar chart, the Decision Tree model has the highest classification accuracy among all the models evaluated. This indicates that, for the given dataset and parameters, the Decision Tree outperforms Logistic Regression, Support Vector Machine, and K-Nearest Neighbors in terms of correctly classifying the test data.

Confusion Matrix

- The confusion matrix for the best performing model, Decision Tree, displays its classification performance. It correctly predicted 5 instances of "did not land" and 12 instances of "landed," with only one misclassification where it predicted "land" instead of "did not land." This indicates a high level of accuracy in classifying the landing outcomes.



Conclusions

- Feature Influence: The payload mass, orbit type, and other features significantly impact the landing success of a Falcon 9 launch.
- Model Performance: Among the machine learning algorithms tested, the decision tree model showed the highest accuracy in predicting the landing outcome.
- Cost Estimation: By accurately predicting the landing outcome, the project provides a valuable tool for estimating the launch costs, which can aid in budget planning and cost management for future missions.
- Data Utilization: The use of historical launch data proves to be effective in training predictive models, showcasing the importance of data-driven decision-making in aerospace operations.

Appendix

- Link of the GitHub Repository: https://github.com/ChrisRosales/Applied-Data-Science-Capstone

Thank you!