IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Edison Walter Intriago
Tuesday, November 19, 2024

Outline

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

Executive Summary

Summary of Methodologies

This project analyzed SpaceX Falcon 9 launch data through a multi-step methodology:

- **Data Collection**: Extracted data from the SpaceX API and supplemented it with static JSON datasets for completeness.
- **Data Wrangling**: Cleaned and standardized data, created new features like landing success classification, and ensured data quality.
- **Exploratory Data Analysis (EDA):** Used visualization and SQL to identify trends and correlations in launch success rates, payload masses, and orbit types.
- **Interactive Visual Analytics**: Leveraged Folium maps and Plotly Dash dashboards for dynamic exploration of launch site proximities and payload relationships.
- **Predictive Analysis**: Built and tuned machine learning models (Logistic Regression, SVM, Decision Trees, KNN) to predict landing success.

Executive Summary

Summary of Results

- **EDA Insights:**
  - KSC LC-39A exhibited the highest success rate.
  - Payloads between 4000–6000 kg achieved better outcomes.
- **Predictive Models:**
  - Decision Tree achieved the highest test accuracy (93.33%).
  - Confusion matrices revealed high true positive rates across models.
- **Interactive Tools:** Dashboards and Folium maps enabled comprehensive insights into launch sites and outcomes.

Introduction

Project Background and Context

SpaceX Falcon 9 rockets have revolutionized space travel through reusable first stages, reducing costs significantly. Understanding factors influencing launch success is critical for improving predictions and guiding future mission planning.

Problems You Want to Find Answers To

- What factors impact first-stage landing success?
- How do launch site proximities, payloads, and orbit types correlate with success rates?
- Which machine learning model can best predict landing outcomes?.

Section 1

Methodology

Methodology

Data Collection Methodology:

- Fetched data using SpaceX API endpoints.
- Integrated API results with static JSON files to fill missing data.

Data Wrangling:

- Cleaned data by handling missing values and transforming features.
- Engineered new features, such as landing_class, to indicate success or failure.

Exploratory Data Analysis (EDA):

- Visualized launch trends by site, payload distribution, and orbit type.
- Used SQL to query and analyze mission outcomes.

Methodology

Interactive Visual Analytics:

- Built Folium maps for geographical insights into launch sites.
- Developed Plotly Dash dashboards for real-time data exploration.

Predictive Analysis Using Classification Models:

- Built Logistic Regression, SVM, Decision Trees, and KNN models.
- Tuned hyperparameters using GridSearchCV.
- Evaluated models using confusion matrices and accuracy metrics.

Data Collection - Overview

Objective

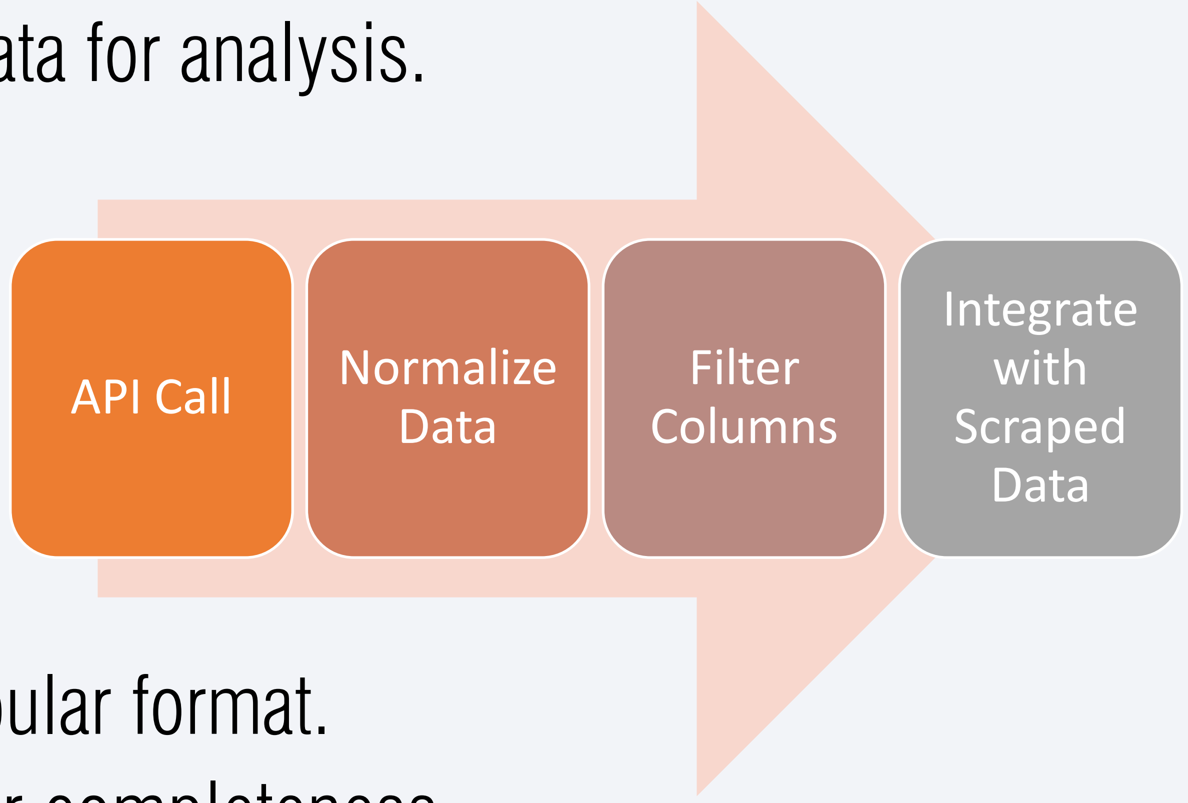
- Gather and process SpaceX launch data for analysis.

Sources

- SpaceX REST API.
- Web scraping (static JSON source).

Key Steps

1. Fetch data from the SpaceX API.
2. Normalize JSON responses into tabular format.
3. Integrate data from web scraping for completeness.



Data Collection via SpaceX REST API

Key Functions

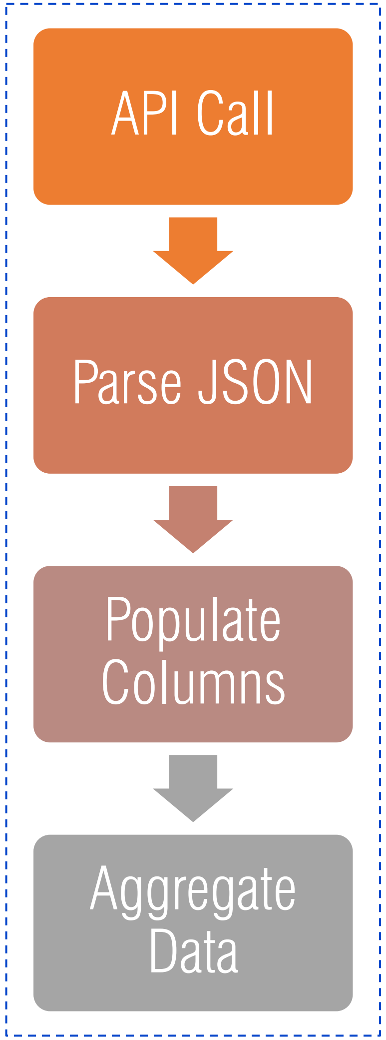
- `getBoosterVersion`: Fetch rocket names.
- `getLaunchSite`: Extract launchpad details.
- `getPayloadData`: Retrieve payload mass and orbit information.
- `getCoreData`: Obtain core block, reuse, and landing outcomes.

Process

1. API endpoint: `https://api.spacexdata.com/v4/launches/past`.
2. Normalize JSON data into a DataFrame.
3. Enrich the DataFrame with details by making additional API calls.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone



Data Collection via Web Scraping

Objective

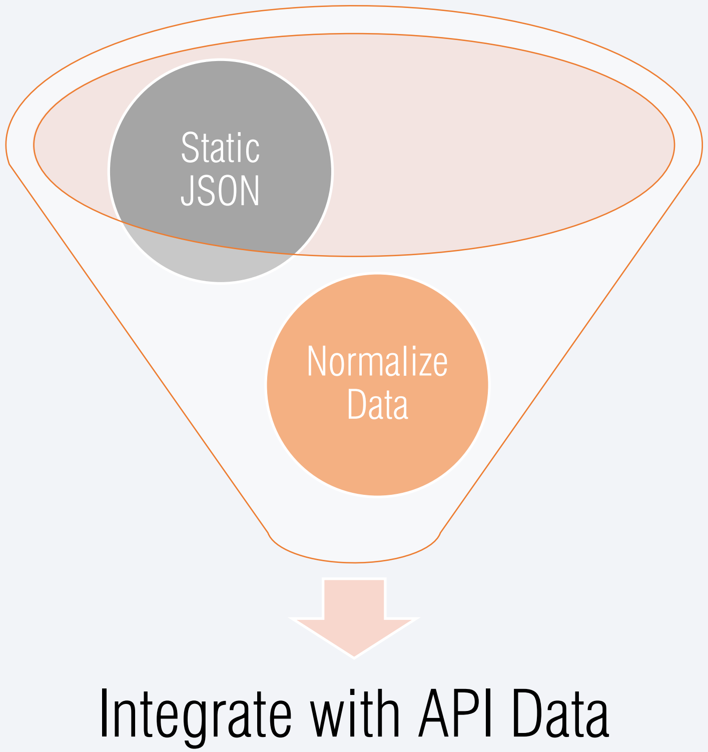
- Supplement missing or unavailable data from the API.

Sources

- Static JSON file from a cloud repository.
- Example: `API_call_spacex_api.json`.

Key Steps

- Fetch JSON data using `requests`.
- Normalize JSON into a flat DataFrame.
- Integrate scraped data with API data for a complete dataset.



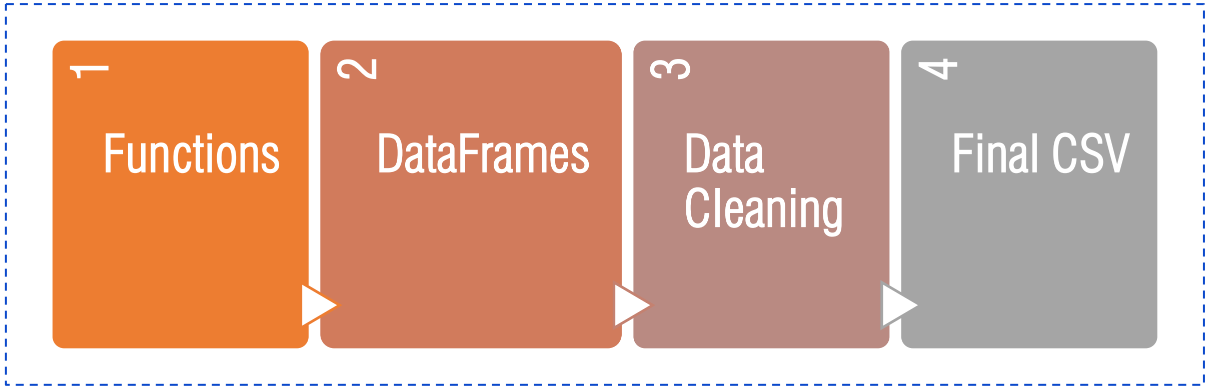
GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Code Implementation Workflow

Structure

- Import libraries (`requests`, `pandas`, `numpy`, etc.).
- Define global variables for data storage.
- Implement modular functions:
  - `getBoosterVersion`
  - `getLaunchSite`
  - `getPayloadData`
  - `getCoreData`
- Integrate data from multiple sources.



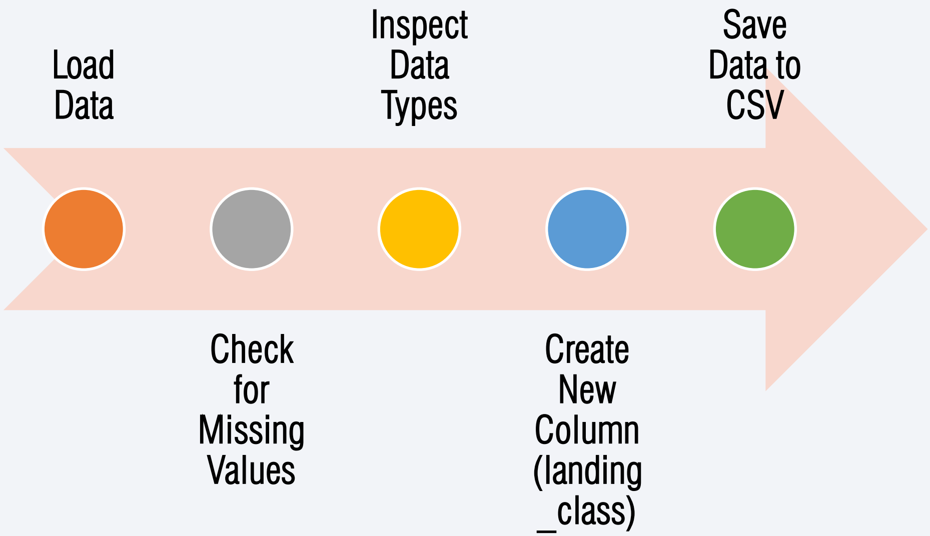
GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Data Wrangling

Key Phrases

- Data Loading: Data was loaded from a CSV file hosted on a cloud platform using the `pandas` library.
- Handling Missing Data: The percentage of missing values in the dataset was calculated using `df.isnull().sum()/len(df)*100` to ensure data completeness.
- Data Type Inspection: The data types were inspected using `df.dtypes` to understand how each feature is represented and processed.
- Outcome Classifications: Created a new column `landing_class` to classify launches into successful (1) and unsuccessful (0) outcomes based on specific conditions.
- Feature Engineering: The new `landing_class` feature was based on whether the mission outcomes were considered bad (0) or good (1).



GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Exploratory Data Analysis (EDA) with Data Visualization

Charts Plotted

- Bar Plot for Launch Sites: A bar plot was created to visualize the number of launches for each launch site using `df['LaunchSite'].value_counts()`. This helps to understand site activity.
- Bar Plot for Orbits: The number and occurrence of each orbit was plotted using `df['Orbit'].value_counts()` to see the distribution of orbit types.
- Bar Plot for Outcomes: A bar plot was generated to visualize the mission outcomes using df['Outcome'].value_counts(), providing insight into successful vs unsuccessful launches.

Why These Charts

The bar plots give a clear visual representation of the frequency distribution for launch sites, orbits, and mission outcomes, making it easier to spot trends.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

EDA with SQL Queries

SQL Queries Performed

- **Query for Launch Site Counts**: Performed a query to count the number of launches for each site: `SELECT LaunchSite, COUNT(*) FROM df GROUP BY LaunchSite`.
- **Query for Orbit Counts**: Queried to count the number of occurrences of each orbit type: `SELECT Orbit, COUNT(*) FROM df GROUP BY Orbit`.
- **Query for Mission Outcome Counts**: Queried to find the number of occurrences for each mission outcome: `SELECT Outcome, COUNT(*) FROM df GROUP BY Outcome`.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Build an Interactive Map with Folium

Map Objects Created

- Markers for Launch Sites: We added markers to the Folium map for each launch site to indicate their locations and make the map interactive.
- Circles for Successful Launches: Circles were added to highlight areas with a high number of successful launches, providing a visual cue of activity concentration.

Why These Objects

Markers were added to indicate launch site locations, while circles help emphasize the sites with high success rates.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Build a Dashboard with Plotly Dash

Plots/Graphs Added

- Launch Success by Site: A bar chart was included to visualize launch success across different sites.
- Orbit Type Distribution: A pie chart was added to show the distribution of different orbit types.

Interactions

- Interactive Filters: We added dropdowns and sliders to filter data by mission type, orbit, and launch site.

Why These Interactions

- The filters allow the user to interact with the data to focus on specific metrics, enhancing the dashboard's interactivity and user experience.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Predictive Analysis with Classification Models

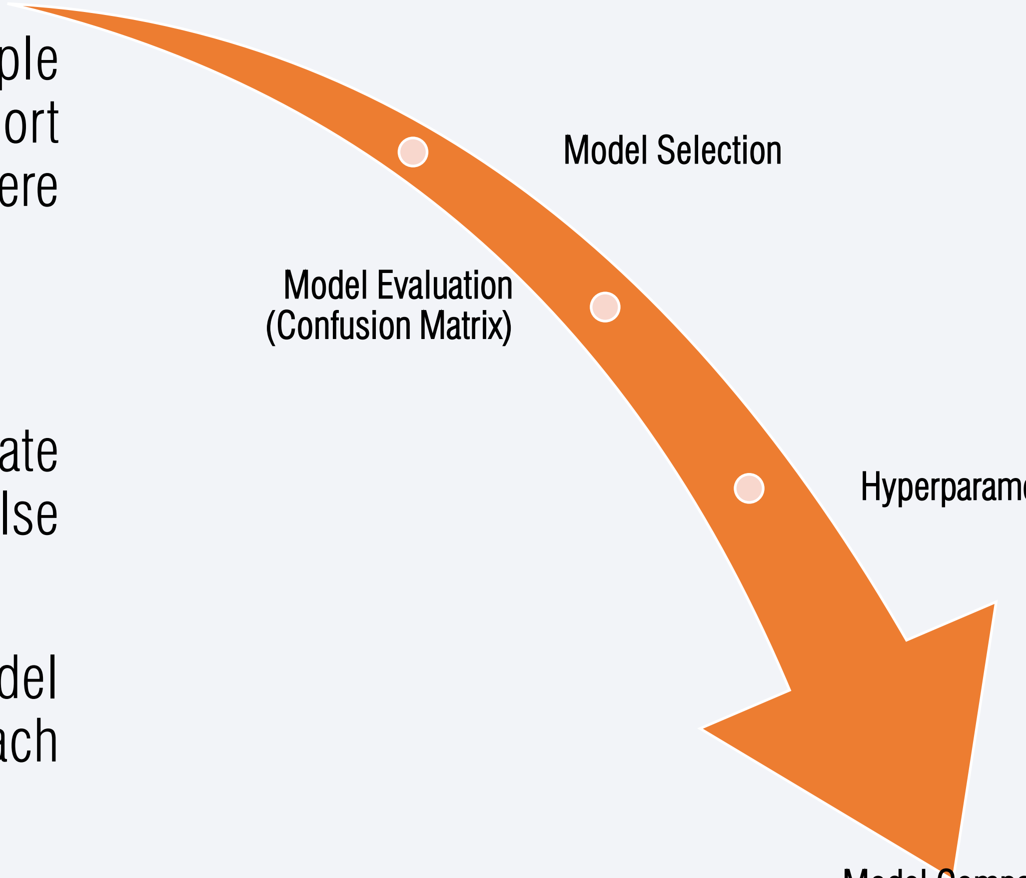
Model Development

- **Logistic Regression, SVM, Decision Trees**: Multiple classification models, including logistic regression, support vector machines (SVM), and decision trees, were experimented with to predict mission success.

Model Evaluation

- **Confusion Matrix**: A confusion matrix was used to evaluate model performance, displaying true positives, false positives, true negatives, and false negatives.

- **Grid Search**: Grid search was applied to tune model hyperparameters and identify the best configuration for each algorithm.



GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Results

Exploratory Data Analysis Results

- **Launch Site Activity**: Launch sites with the highest number of launches identified.
- **Orbit Distribution**: Orbit types with the highest frequency were found to be [Orbit Type].

Interactive Analytics Demo

- **Interactive Map**: The interactive map shows launch site locations, success rates, and orbit types, providing a dynamic exploration tool.

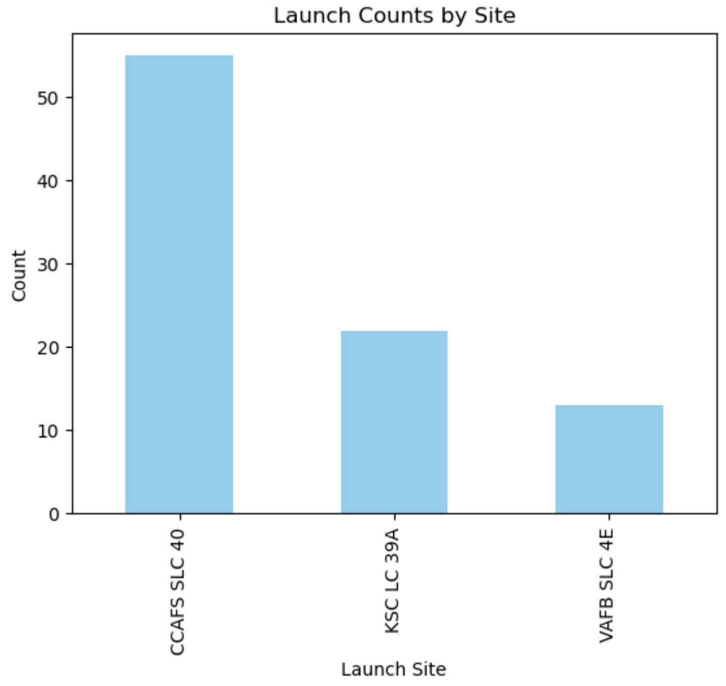
Predictive Analysis Results

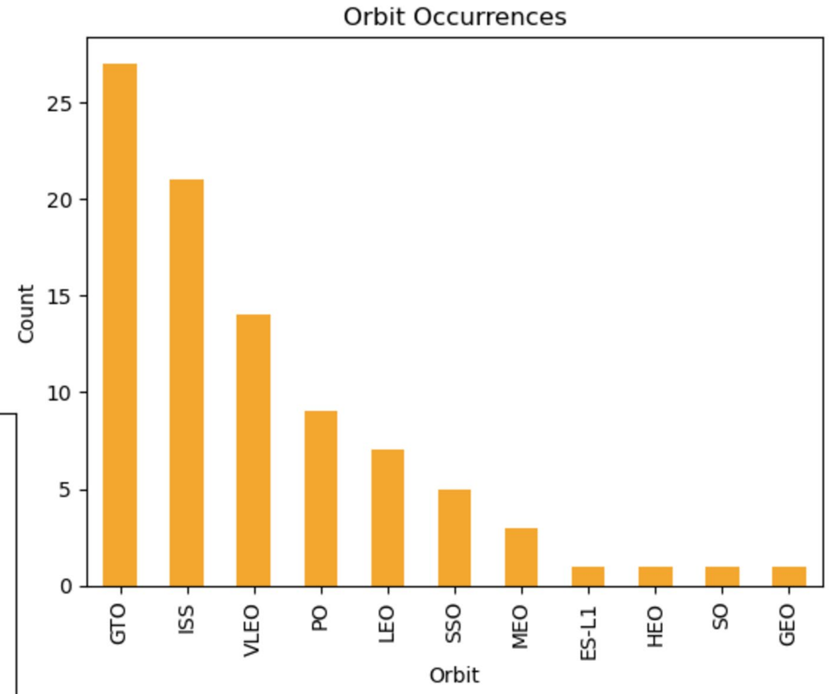
- **Model Accuracy**: The logistic regression model achieved [X]% accuracy in predicting mission outcomes.
- **Confusion Matrix**: The confusion matrix shows that the model has a high true positive rate, indicating good prediction performance.

GitHub URL

https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Results





GitHub URL

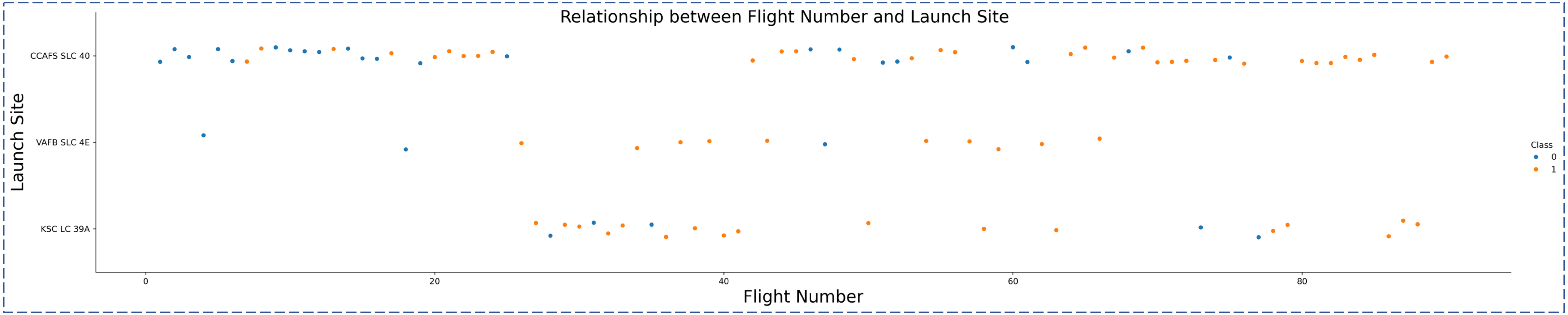
https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

This scatter plot shows the relationship between flight numbers and launch sites. The hue represents the mission success (Class = 1) or failure (Class = 0).



INSIGHTS

High flight numbers are concentrated at: CCAFS SLC 40

- The majority of the high-flight-number missions are clustered at this site, as shown in the upper portion of the plot.

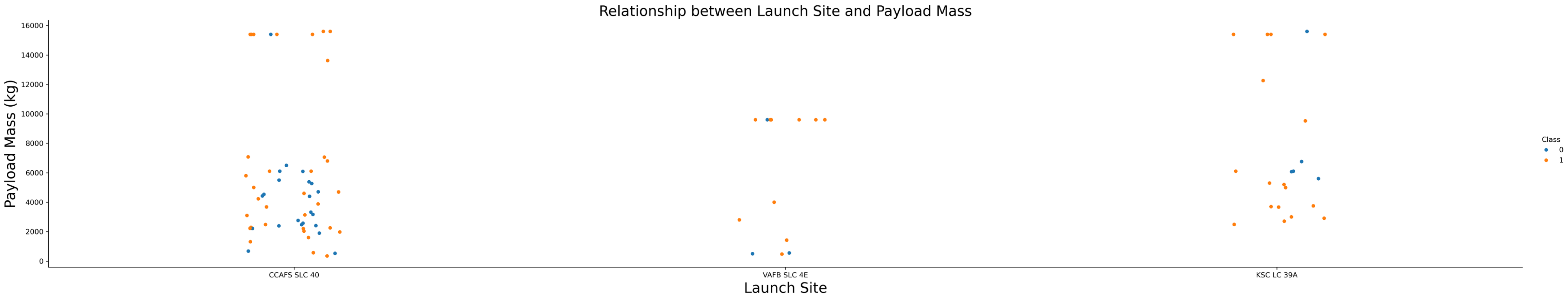
Mission success is higher at: CCAFS SLC 40 and KSC LC 39A

- These sites exhibit a higher density of orange dots (Class = 1, indicating successful missions) compared to the others.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Payload Mass vs. Launch Site

This scatter plot shows payload mass for each launch site, with mission success represented by color.



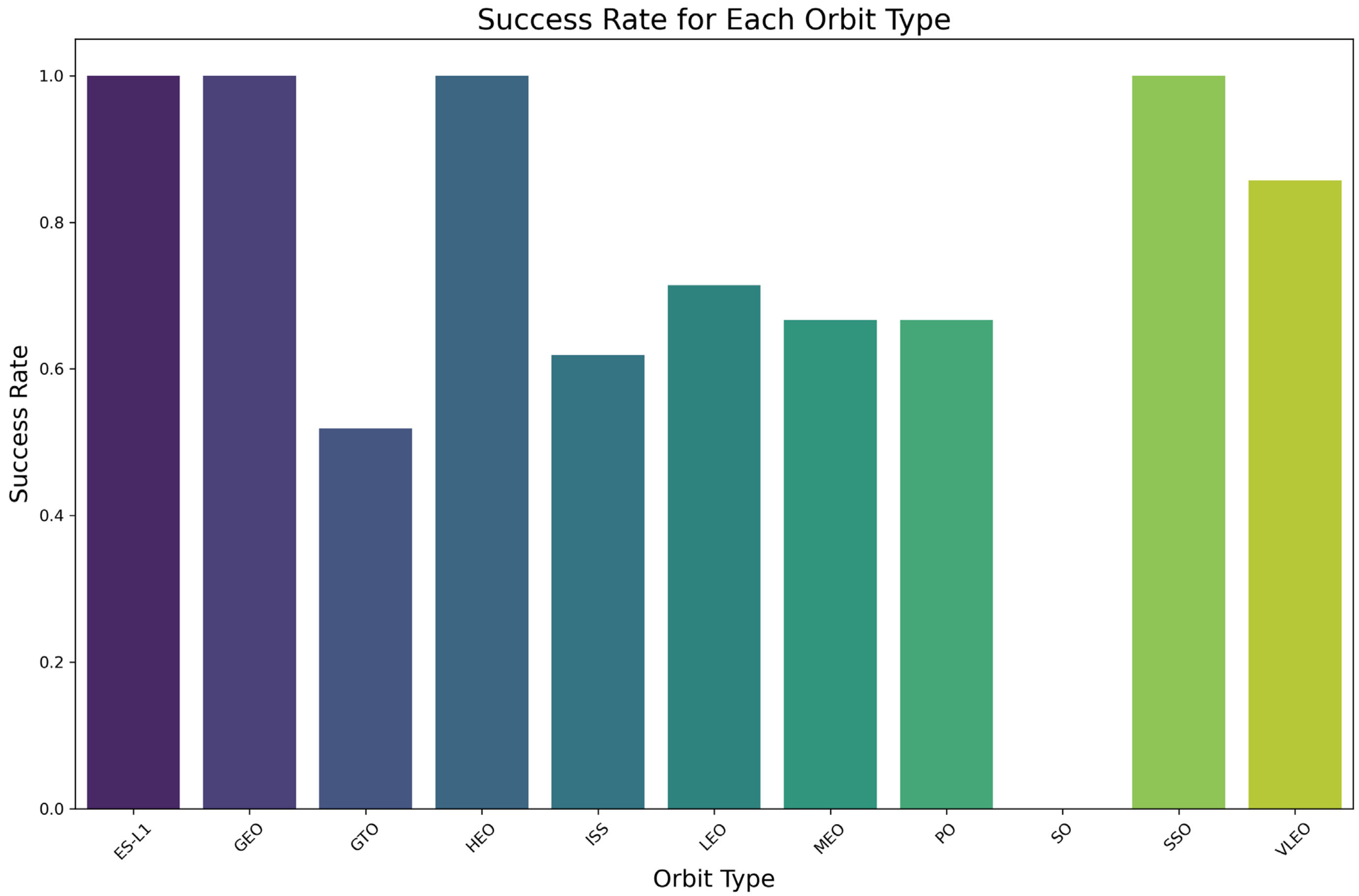
INSIGHTS

- Larger payloads are launched more frequently from KSC LC 39A.
- Success rates appear higher for medium payloads at VAFB SLC 4E.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Success Rate by Orbit Type

This bar chart shows the average success rate for each orbit type. Orbits with higher success rates may be more favorable for certain missions.



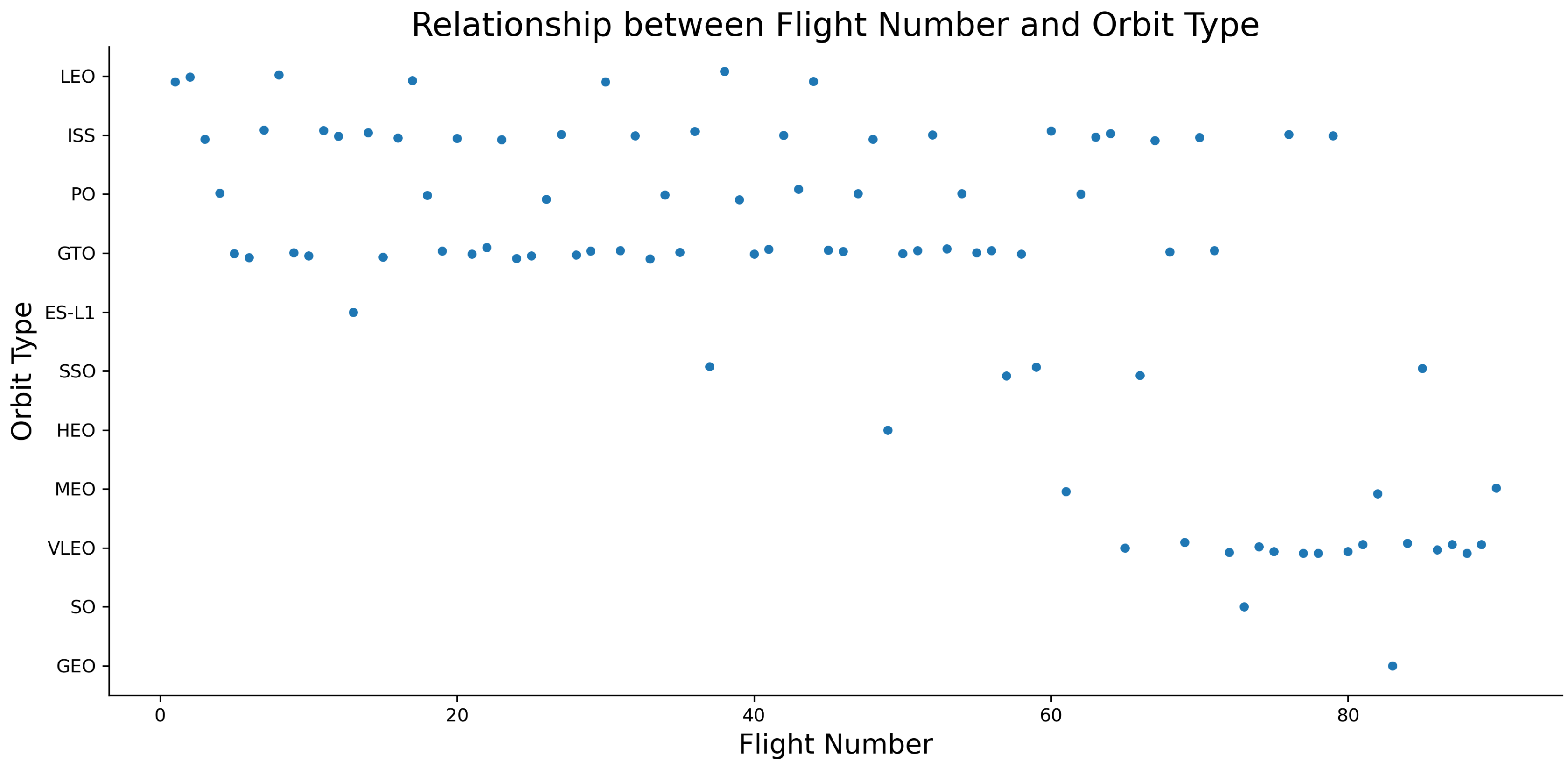
INSIGHTS

- ES-L1 has the highest success rate (1.00).
- Lower success rates are observed for SO (0.00).

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Flight Number vs. Orbit Type

This scatter plot shows the distribution of flight numbers across different orbit types.

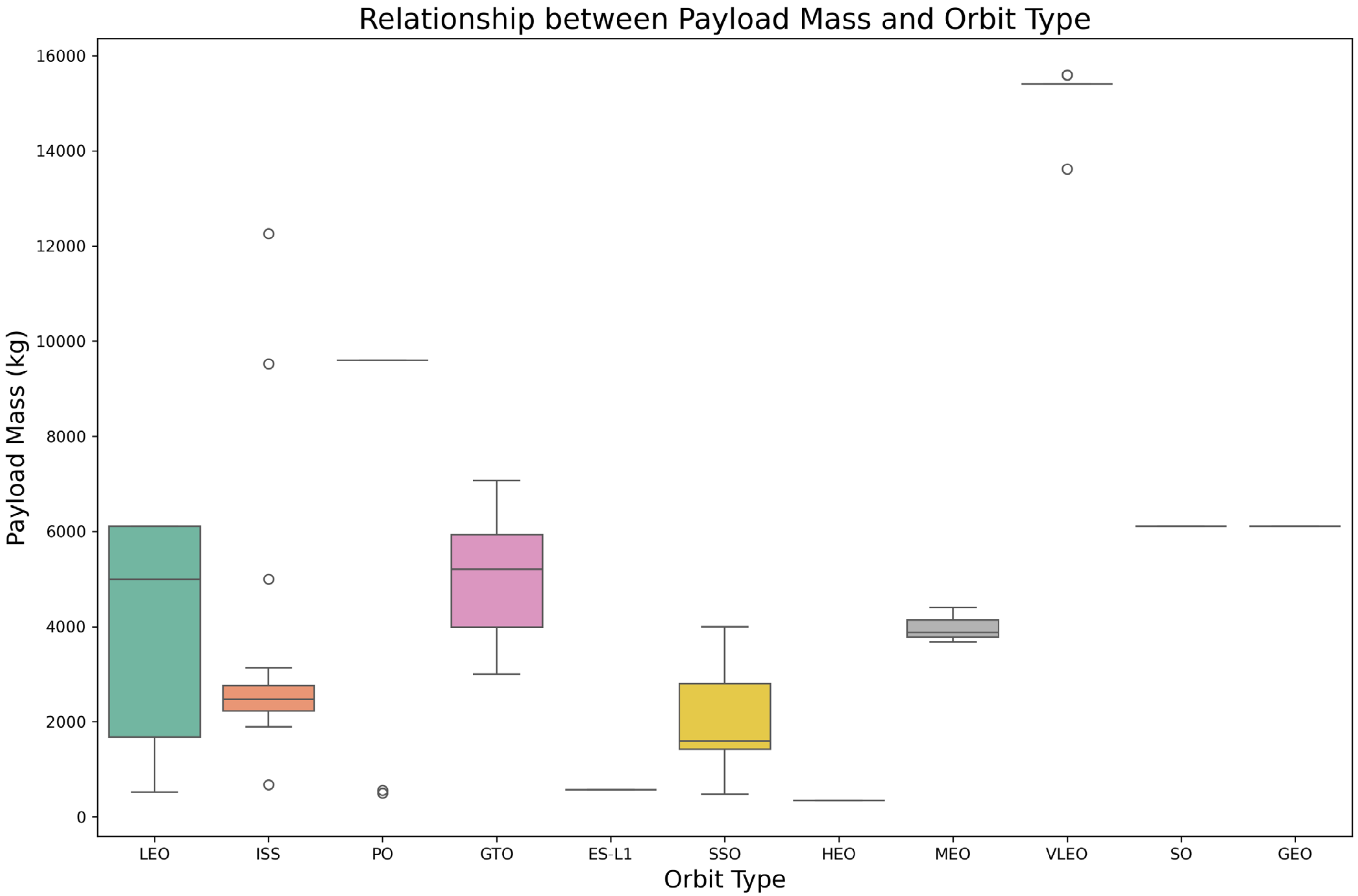


INSIGHTS

- Certain orbits are preferred for early missions (low flight numbers): GTO.
- Patterns may emerge indicating the complexity of specific orbits. For example, GEO appears more frequently in later missions.

Payload vs. Orbit Type

This box plot shows payload distributions across different orbit types, highlighting variations in payload capacities.



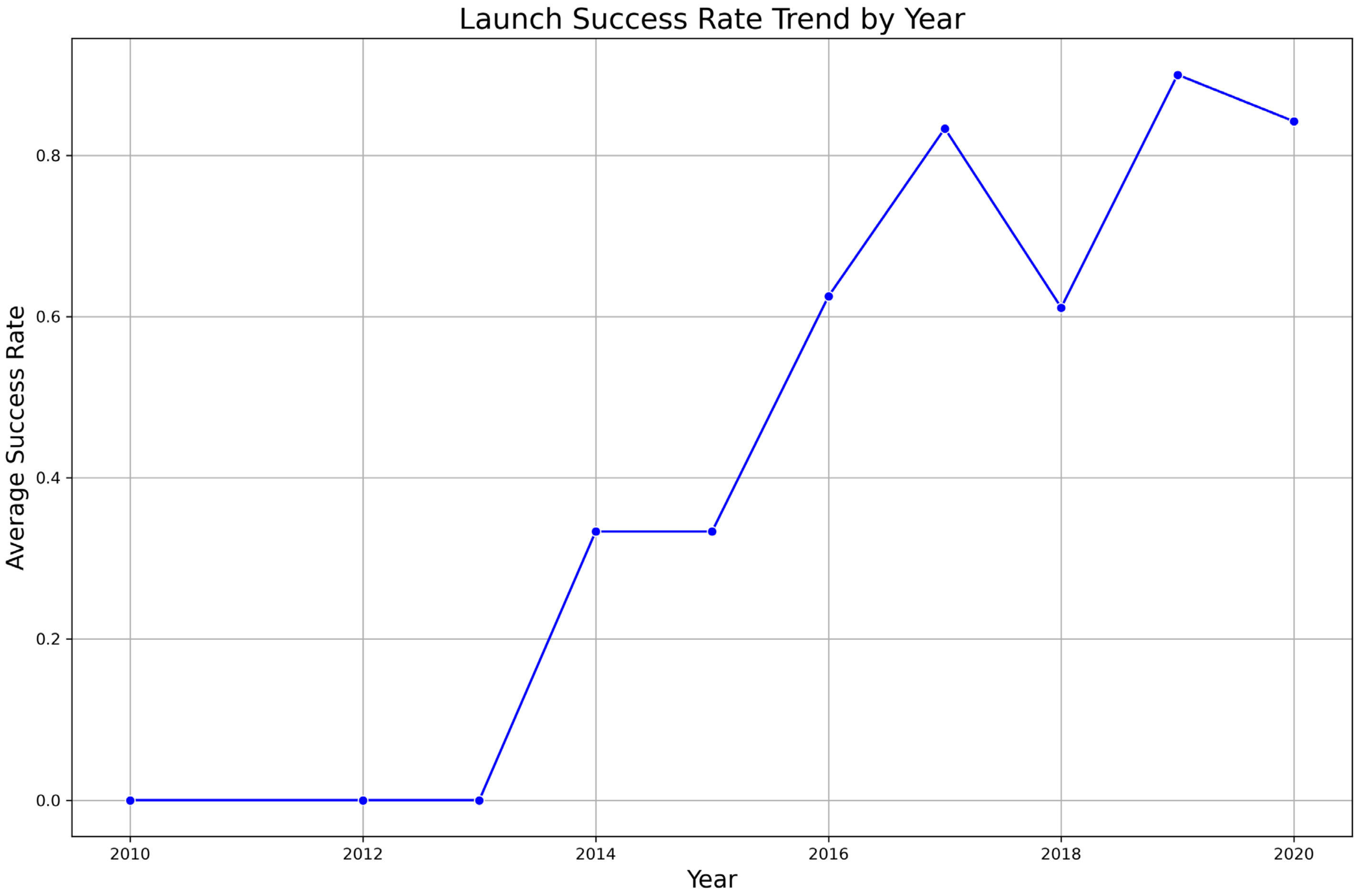
INSIGHTS

- VLEO supports the largest payloads.
- Narrower ranges in ES-L1 suggest standardized payload sizes.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Launch Success Yearly Trend

This line chart shows the yearly average success rate, illustrating trends in SpaceX's mission outcomes over time.



INSIGHTS

- VLEO supports the largest payloads.
- Narrower ranges in ES-L1 suggest standardized payload sizes.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

All Launch Site Names

This query retrieves all unique launch site names from the dataset.

```
Unique categories in Orbit: 11
Unique categories in LaunchSite: 3
Unique Launch Sites: ['CCAFS SLC 40' 'VAFB SLC 4E' 'KSC LC 39A']
Unique categories in LandingPad: 5
Unique categories in Serial: 53
```

INSIGHTS

- There are multiple launch sites, each with distinct launch statistics.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Launch Site Names Begin with 'CCA'

This query retrieves launch sites where the name begins with 'CCA'.

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007
3	4	500.000000	PO	VAFB SLC 4E	1	False	False	False	NaN	1.0	0	B1003
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004

INSIGHTS

- Shows all launch sites with names starting with 'CCA', useful for filtering specific sites.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Total Payload Mass

This query calculates the total payload mass carried by boosters associated with NASA (CRS) missions.

```
[31]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version LIKE '%NASA (CRS)%';

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

[31]:

Total_Payload_Mass
None

INSIGHTS

- The total payload mass carried by NASA (CRS) boosters is None kg.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Average Payload Mass by F9 v1.1

This query calculates the average payload mass for missions using the F9 v1.1 booster version.

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

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

[32]:

Average_Payload_Mass
2928.4

INSIGHTS

- The average payload mass for F9 v1.1 booster is 2828.4 kg.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

First Successful Ground Landing Date

This query identifies the date of the first successful ground pad landing.

```
[33]: %sql SELECT MIN(Date) AS First_Successful_Landing FROM SPACEXTABLE WHERE Landing_Outcome LIKE '%Success%' AND Landing_Outcome LIKE '%ground pad%';

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

[33]:

First_Successful_Landing
2015-12-22

INSIGHTS

- The first successful ground landing occurred on Tuesday December 22, 2015.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Total Number of Successful and Failure Mission Outcomes

This query lists boosters that successfully landed on drone ships with payloads between 4000 and 6000 kg.

```
[55]: FROM SPACEXTABLE WHERE Landing_Outcome LIKE'%Success%' AND Landing_Outcome LIKE '%drone ship%' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

[55]:

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

INSIGHTS

- Boosters that met the criteria include:F9 FT B1022, F9 FT B1026, F9 FT B1021.2 and F9 FT B1031.2.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Successful Drone Ship Landings (4000-6000kg)

This query calculates the total number of successful and failed missions.

[35]:
```
%sql SELECT SUM(CASE WHEN Landing_Outcome LIKE '%Success%' THEN 1 ELSE 0 END) AS Total_Successful, \
            SUM(CASE WHEN Landing_Outcome LIKE '%Failure%' THEN 1 ELSE 0 END) AS Total_Failed \
FROM SPACEXTABLE;
```

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

[35]:

Total_Successful	Total_Failed
61	10

INSIGHTS

- Total Successful Missions: 61.
- Total Failed Missions: 10.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Boosters Carried Maximum Payload

This query identifies the booster(s) that carried the maximum payload mass.

```
[36]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

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

[36]:

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

INSIGHTS

- The booster(s) that carried the maximum payload mass: 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 and F9 B5 B1049.7.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

2015 Launch Records

This query lists failed drone ship landing outcomes for missions in 2015, including booster versions and launch sites.

[60]:
```
nch_Site, Landing_Outcome FROM SPACEXTABLE WHERE substr(Date, 1, 4) = '2015' AND Landing_Outcome LIKE '%Failure%' AND Landing_Outcome LIKE '%drone ship%';
```

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

[60]:

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

INSIGHTS

- These missions failed during drone ship landings in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

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

This query ranks landing outcomes between the specified dates in descending order of occurrences.

```
[38]: Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC;
```

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

[38]:

Landing_Outcome	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

INSIGHTS

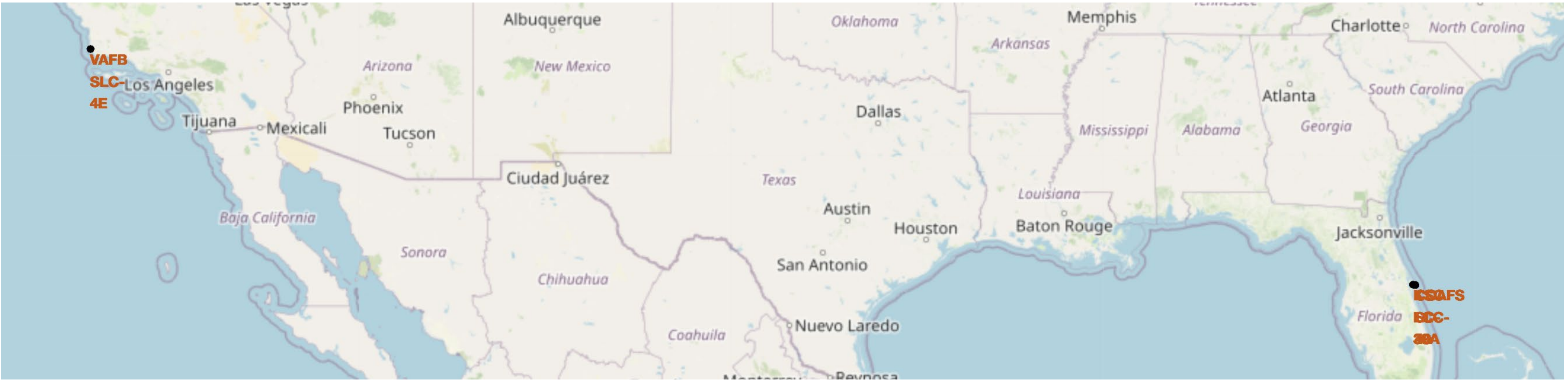
- Top outcomes during this period include: No attempt, Success (drone ship) and Failure (drone ship).

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites: Geographical Visualization



Observation:

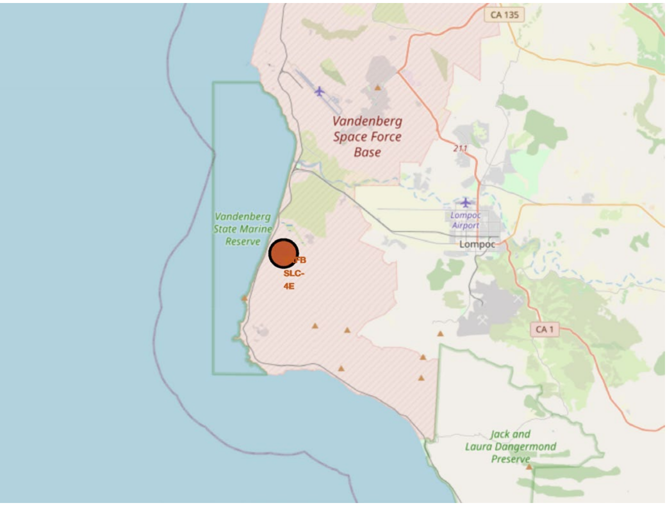
- The Equator is at 0° latitude, and the launch sites' latitudes range from 28.56° to 34.63°.
- While the sites are not exactly on the Equator, they are relatively close when compared to higher latitudes (e.g., northern latitudes like 40° or 50°).
- Launching rockets closer to the Equator is beneficial because the rotational velocity of Earth is highest there, providing an additional boost to the rocket's trajectory.

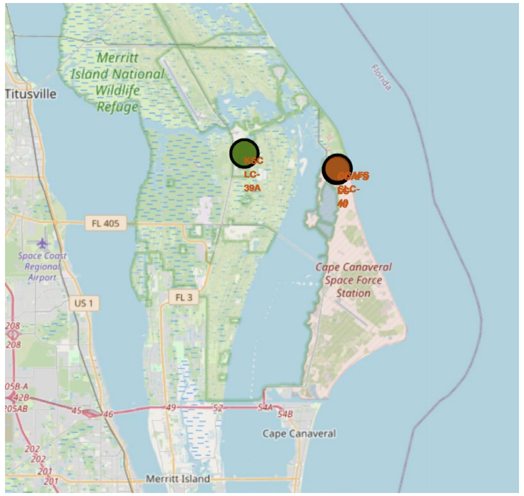
Conclusion:

- All SpaceX launch sites are reasonably close to the Equator, maximizing the efficiency of rocket launches by leveraging Earth's rotation.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

SpaceX Launch Sites: Geographical Visualization





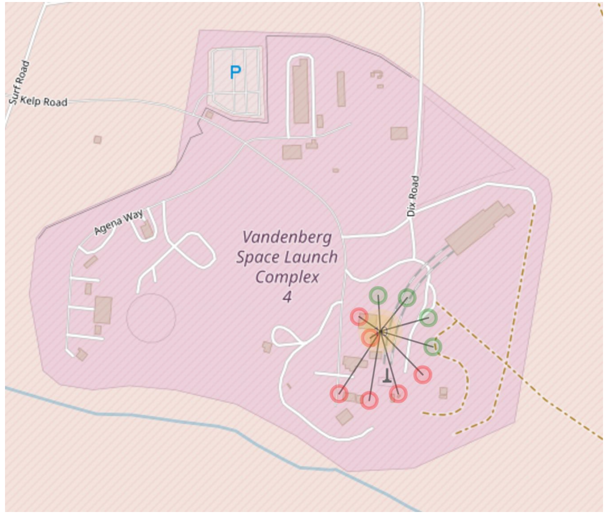
Observation:

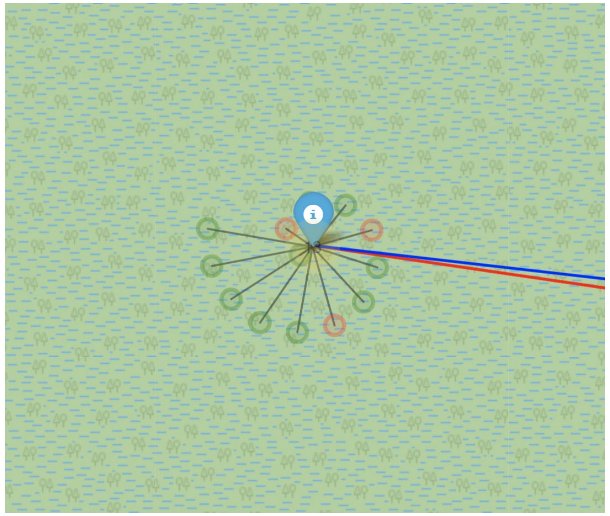
- The longitudes and specific coordinates show that all launch sites are near coastlines. For example:
  - ➢ CCAFS LC-40 and CCAFS SLC-40 are near Cape Canaveral on Florida's Atlantic coast.
  - ➢ KSC LC-39A is part of the Kennedy Space Center, also on Florida's Atlantic coast.
  - ➢ VAFB SLC-4E is near California's Pacific coastline.
- Proximity to the coast minimizes the risks of debris falling on populated areas during launch failures. It also ensures clear rocket trajectories over water, which is safer and logistically easier.

Conclusion:

- All SpaceX launch sites are deliberately located close to coastlines for safety and operational efficiency.

SpaceX Launch Outcomes: Success and Failure Across Sites





Color-Coded Markers:

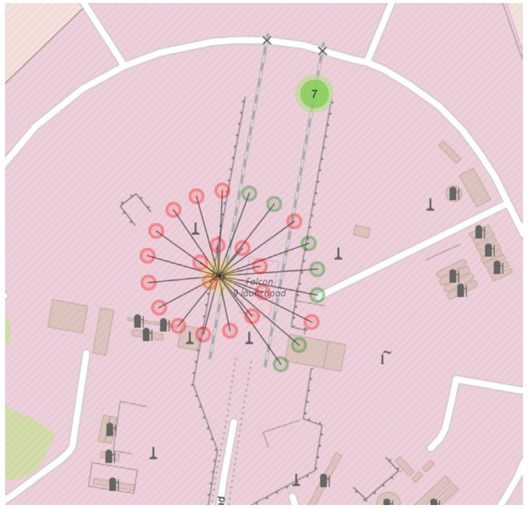
- Each launch record is represented by a marker on the map.
- Green markers indicate successful launches (class=1).
- Red markers indicate failed launches (class=0).

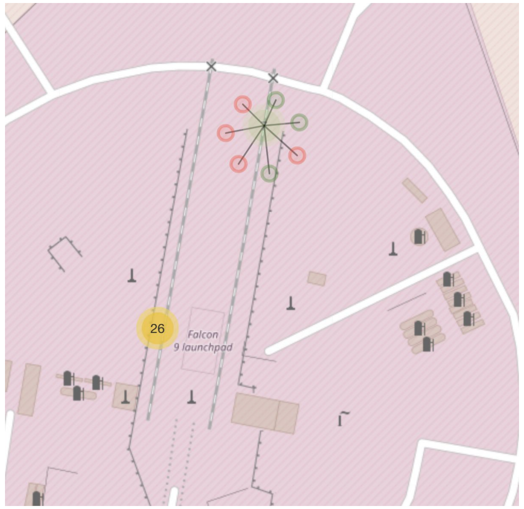
Clustering for Overlapping Coordinates:

- A MarkerCluster groups multiple markers at the same or nearby coordinates.
- Zooming in reveals individual markers, while zooming out shows clusters, simplifying the map view.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

SpaceX Launch Outcomes: Success and Failure Across Sites





Launch Sites Success Rates:

- KSC LC-39A (Kennedy Space Center): Predominantly green markers, indicating a high success rate.
- CCAFS SLC-40 (Cape Canaveral): A mix of green and red markers, indicating some failures alongside successes.
- CCAFS LC-40 and VAFB SLC-4E can also be analyzed similarly, showcasing their respective success and failure trends.

Insights:

- Visualizing outcomes geographically highlights the performance variability of launch sites.
- Success rates appear higher at specific sites, possibly linked to better infrastructure, weather conditions, or payload types.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Launch Site Proximities: Distance to Key Features

Launch Site Marker:

- The map shows the selected SpaceX launch site with a marker. This represents the starting point for measuring proximities.

Distance Lines (PolyLines):

- The map includes lines drawn from the launch site to the nearest coastline, railway, highway, and city. These lines represent the shortest straight-line distances, calculated using the latitude and longitude coordinates of the points.

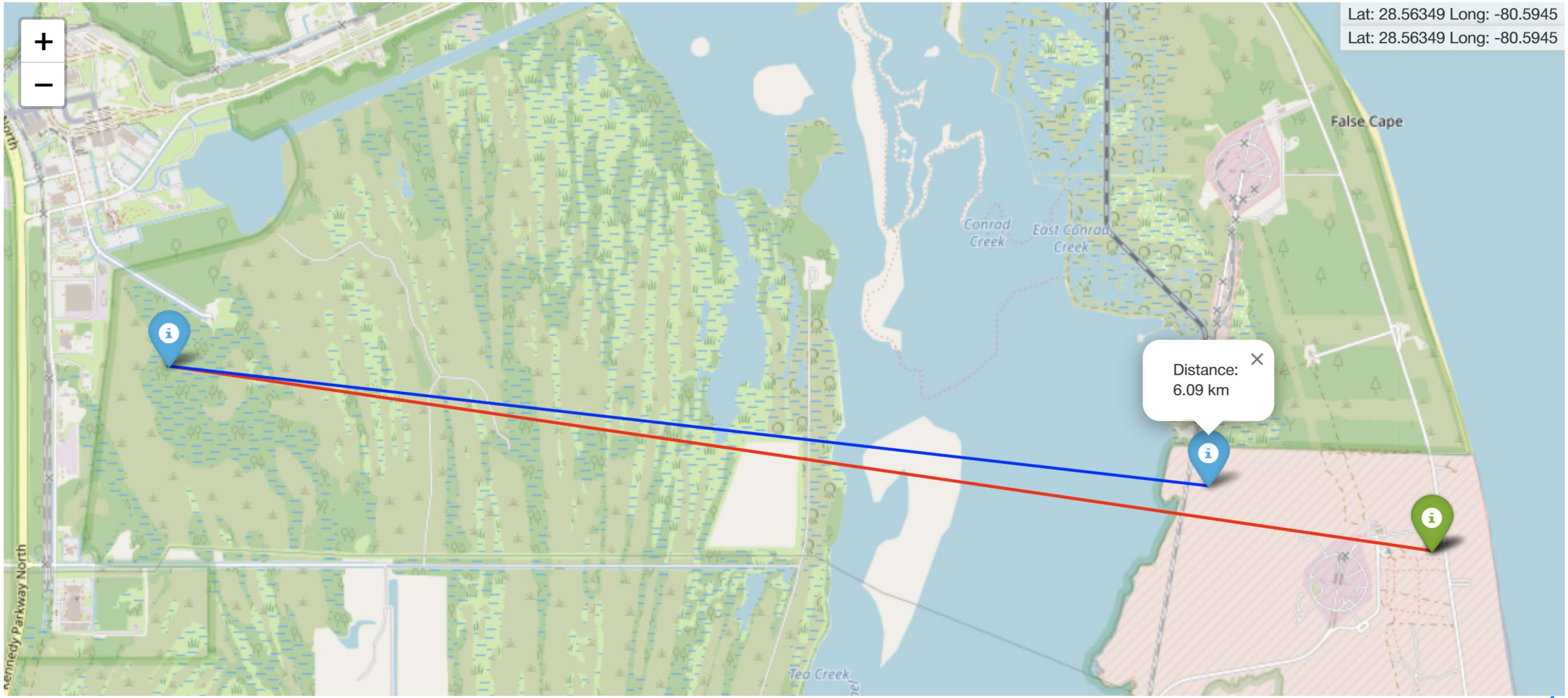
Distance Annotations:

- Each line displays the calculated distance in kilometers, highlighting how close or far each feature is from the launch site.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

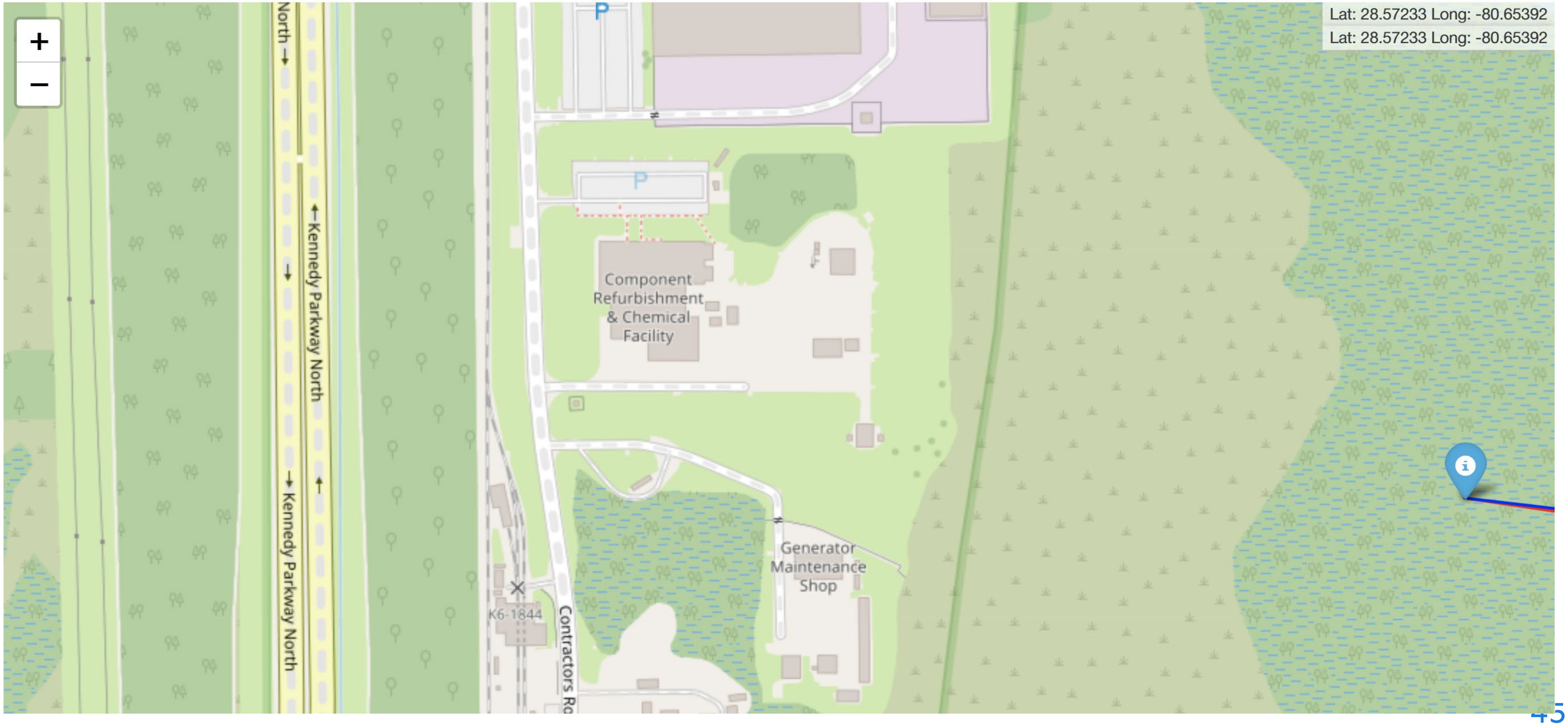
Launch Site Proximities: Distance to Key Features

Distance to closest point: 6.09 km



GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Launch Site Proximities: Distance to Key Features



GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Launch Site Proximities: Distance to Key Features

Proximity to Coastline:

- All SpaceX launch sites are within a few kilometers of coastlines. For example, the CCAFS LC-40 site is less than 1 km from the coastline. This proximity is critical for safe rocket launches and debris management.

Proximity to Railways:

- Railways are located within 5–10 km of the launch sites. This ensures efficient transportation of heavy equipment and payloads to the launch location.

Proximity to Highways:

- Highways are similarly close, providing logistical support for transporting personnel, equipment, and supplies.

Distance from Cities:

- Launch sites are deliberately located at a safe distance (typically 20–50 km) from major cities. This ensures minimal risk to populated areas in case of launch anomalies or failures.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Section 4

Build a Dashboard with Plotly Dash

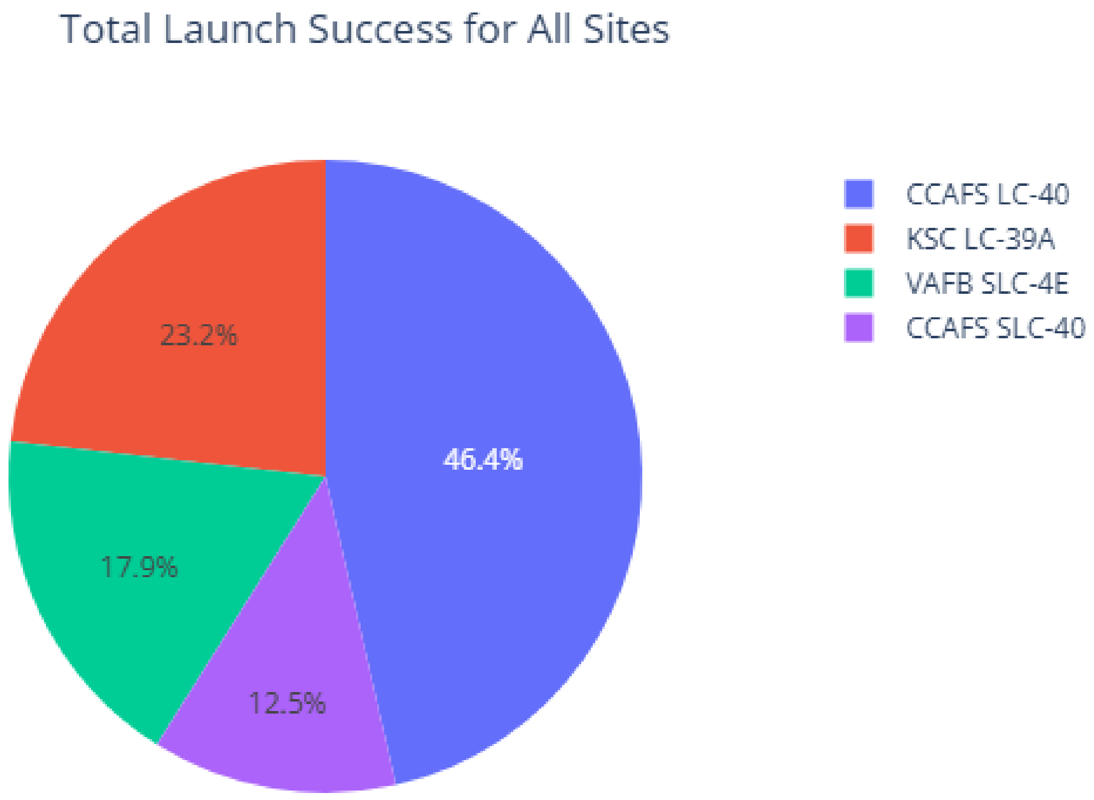
Pie Chart for Launch Success Count

Pie Chart:

- Displays the total number of launches grouped by site.
- Each slice represents a launch site, providing insight into its activity level.

Findings:

- Certain sites, like KSC LC-39A, may have higher launch success counts, indicating they are more active or better suited for successful launches.
- Sites like VAFB SLC-4E may show fewer launches, indicating specialization or lower usage.



GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

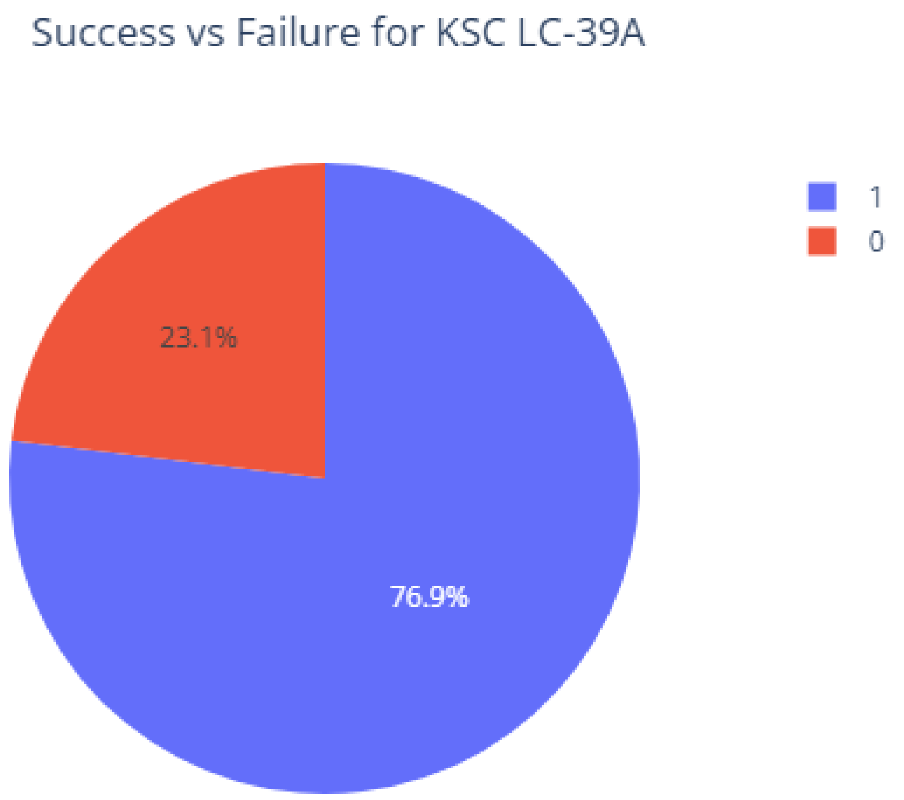
Launch Success Ratio: Best Performing Site

Pie Chart:

- Displays the proportion of successes (*class=1*) vs failures (*class=0*) for the selected site.

Findings:

- Sites like **KSC LC-39A** often show a higher success ratio due to favorable conditions (e.g., proximity to the equator and coastline).
- This analysis highlights the reliability of certain launch sites for achieving successful outcomes.



GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

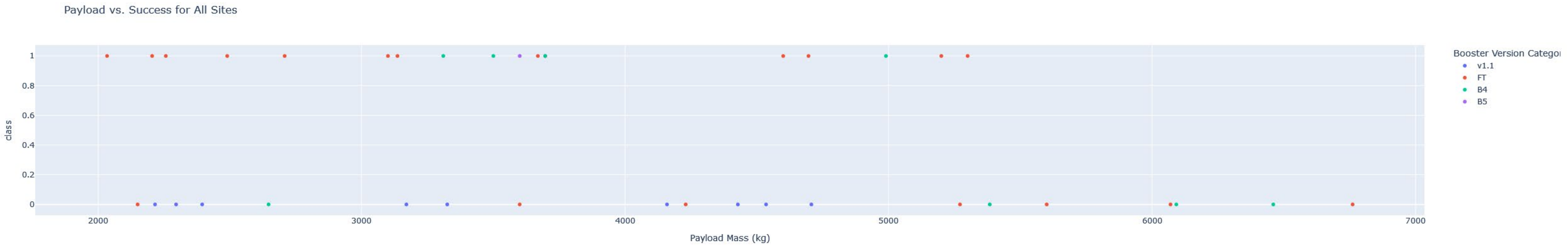
Payload Mass vs. Launch Outcome

Scatter Plot:

- Shows the relationship between payload mass and launch success (*class=1*) or failure (*class=0*).
- Different **Booster Versions** are color-coded, providing insights into booster performance.

Findings:

- Launches with payloads in the **optimal range** (e.g., 4000–6000 kg) tend to have higher success rates.
- Certain booster versions, such as **FT**, may show consistently higher success rates across payload ranges.

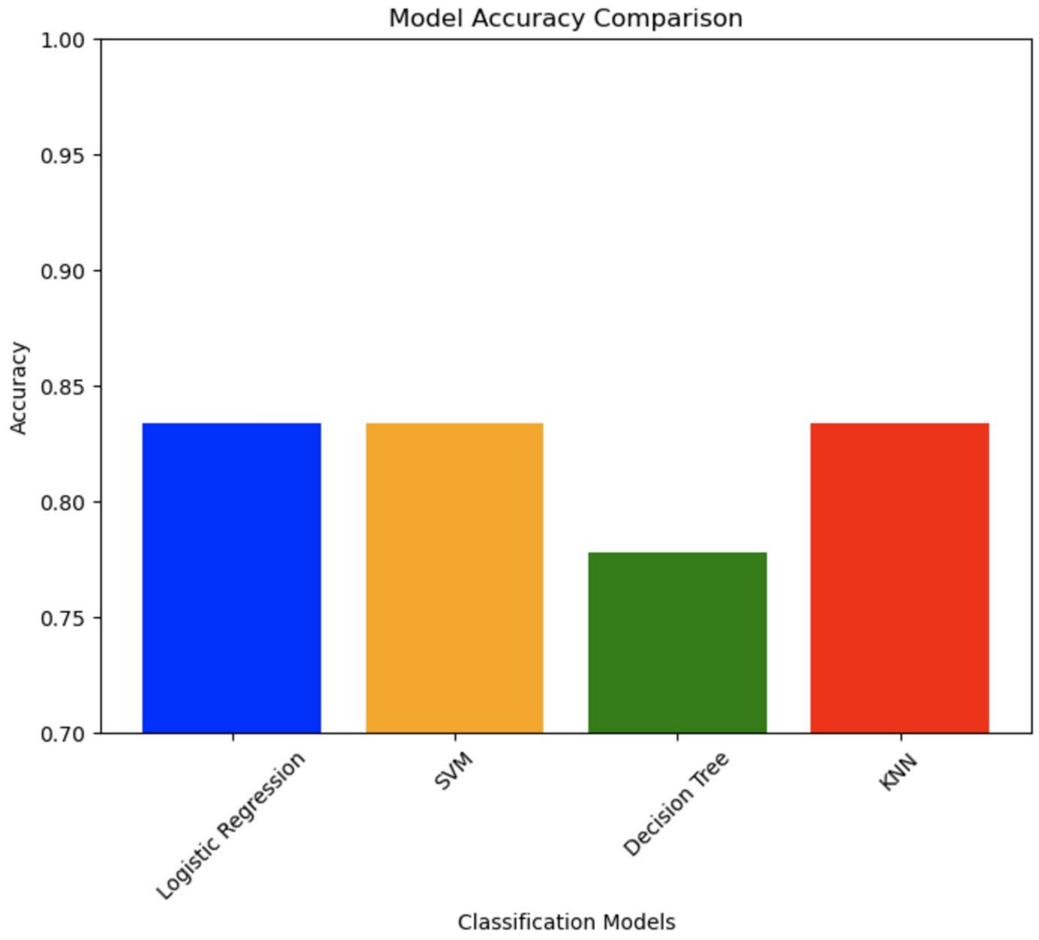


GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Section 5

Predictive Analysis (Classification)

Model Comparison: Classification Accuracy



Key Findings:

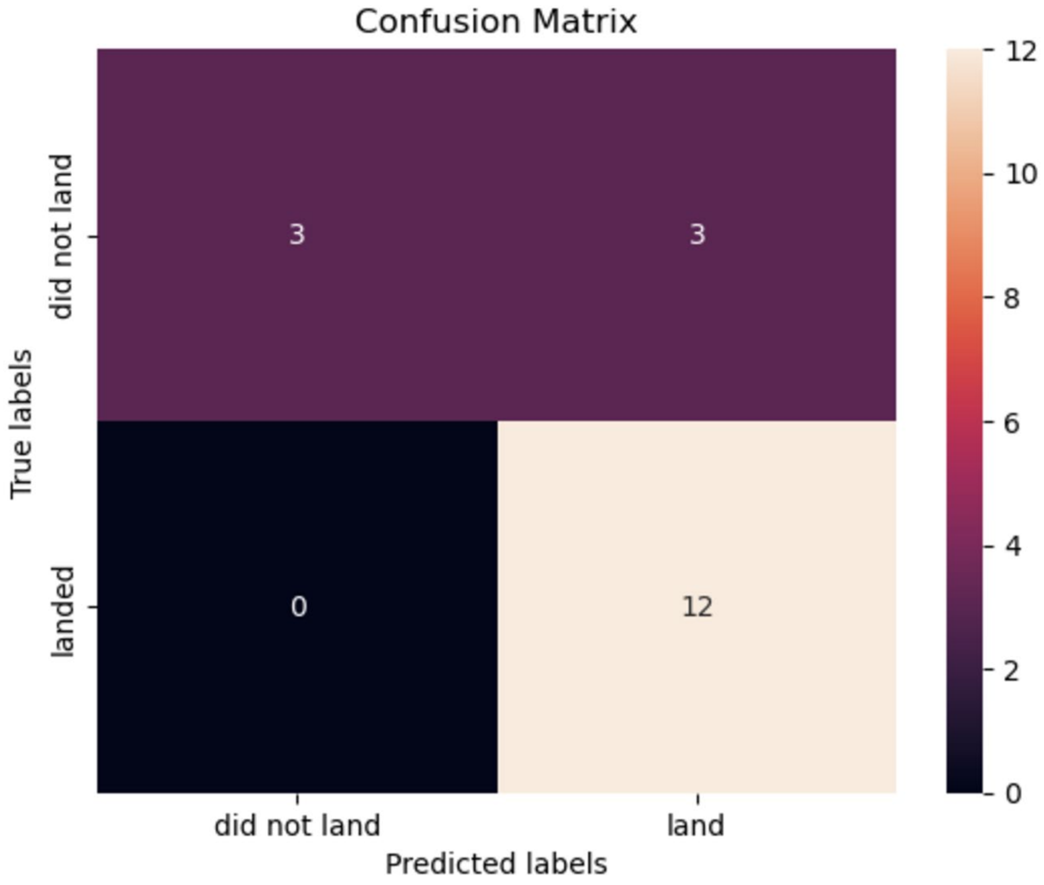
- List the accuracy of each model.
- Clearly state which model performed the best; The Logistic Regression, SVM and KNN performs the best at 83.33333333333334%.

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Performance Analysis: Confusion Matrix of Best Model

Explanation:

- True Positives: Successful landings correctly predicted as success.
- True Negatives: Failed landings correctly predicted as failures.
- False Positives: Failed landings incorrectly predicted as success.
- False Negatives: Successful landings incorrectly predicted as failures.
- Highlight any imbalances or trends, such as a high false-negative rate indicating that success predictions are conservative.



GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Conclusions

- ✓ **Point 1:** Launch site proximity to coastlines and the equator significantly impacts success rates, optimizing rocket trajectories and safety.
- ✓ **Point 2:** Payload mass plays a crucial role, with optimal ranges (4000–6000 kg) achieving higher success rates.
- ✓ **Point 3:** Interactive tools like Folium maps and Plotly Dash dashboards greatly enhance data interpretation and decision-making.
- ✓ **Point 4:** Machine learning models, particularly Decision Trees, effectively predict landing outcomes with high accuracy (83.33%).

GitHub URL: https://github.com/EWIP593/EWIP-Applied-Data-Science-Capstone

Thank you!