IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thato Manamela
04-01-2025

Outline

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

Executive Summary

Summary of Methodologies

- **Data Collection**:
  - Data requests from the SpaceX API
  - Web Scraping other relevant information on the internet
- **Data Cleaning and Wrangling**:
  - Missing values were handled, and categorical variables were encoded where necessary.
  - Data preprocessing included standardizing numeric features using `StandardScaler`.
- **Data Exploration**:
  - Explored relationships between features and the target variable (`class`) using descriptive statistics and visualizations.
  - Assessed the influence of key variables like payload mass, booster version, and launch site on the landing success.
- **Machine Learning Pipeline**:
  - Implemented a pipeline with preprocessing steps and model training/testing.
  - Split data into training and testing subsets to evaluate performance.
- **Model Selection and Hyperparameter Optimization**:
  - Four machine learning algorithms were used: Logistic Regression, Support Vector Machines (SVM), Decision Tree Classifier, and K-Nearest Neighbors (KNN).
  - Hyperparameters were optimized using **GridSearchCV** with 10-fold cross-validation.
  - Models were evaluated based on accuracy and confusion matrix analysis.

Executive Summary

Summary of Results

- **Logistic Regression**:
  - Achieved a test accuracy of **83.33%**.
  - Best hyperparameters: `C=1`, `penalty='l2'`, `solver='lbfgs'`.
- **Support Vector Machines (SVM)**:
  - Achieved a test accuracy of **80.00%**.
  - Best kernel: **'sigmoid'**.
  - Optimized hyperparameters: `C=1.0`, `gamma=0.32`.
- **Decision Tree Classifier**:
  - Achieved a test accuracy of **73.33%**.
  - Best hyperparameters included `criterion='entropy'`, `max_depth=16`, and `min_samples_split=2`.
- **K-Nearest Neighbors (KNN)**:
  - Achieved a test accuracy of **66.67%**.
  - Best hyperparameters: `n_neighbors=10`, `algorithm='auto'`, `p=1`.
- **Best Performing Algorithm**:
  - **Logistic Regression** outperformed other models with the highest accuracy on the test data.

Executive Summary

Conclusions & Implications

- **Best Performing Algorithm**:
  - **Logistic Regression** outperformed other models with the highest accuracy on the test data.
- **Insights on Falcon 9 Landings**:
  - Launch sites, payload mass, and booster versions significantly influenced the success of first-stage landings.
  - Predictive models demonstrated that Falcon 9 landing success can be reliably forecasted using historical data.
- **Practical Applications**:
  - Insights from this analysis can guide SpaceX in optimizing launch conditions and selecting mission parameters.
  - Improved predictive accuracy can reduce costs and enhance the reliability of reusable rockets.

Introduction

Project background and context

- SpaceX has revolutionized space exploration with its innovative reusable rocket technology, significantly reducing the cost of space missions.
- A critical factor in achieving this milestone is the ability to land the first stage of Falcon 9 rockets successfully, enabling their reuse for future missions.
- Despite advancements, landing success remains influenced by several variables such as payload mass, launch site, orbit type, and booster version, among others.
- The use of machine learning provides a powerful tool to analyze historical data, uncover patterns, and predict the likelihood of successful landings.
- This project applies machine learning algorithms to SpaceX's Falcon 9 launch data to predict the success of first-stage landings, providing valuable insights for mission planning and operational optimization.

Introduction

Problems We Want To Find Answers To

- **What factors influence the success of Falcon 9 first-stage landings?**
  - By exploring variables such as payload mass, booster type, and launch site, we aim to identify the most impactful factors.
- **Can we predict whether the first stage will land successfully?**
  - Using machine learning models, we aim to develop a reliable predictive tool to forecast landing success.
- **Which machine learning algorithm is most effective for this problem?**
  - By testing models like Logistic Regression, Support Vector Machines, Decision Trees, and K-Nearest Neighbors, we will identify the model that delivers the best performance based on accuracy and validation metrics.
- **How can this analysis guide SpaceX's future mission planning?**
  - By understanding the conditions that maximize landing success, SpaceX can optimize resource allocation and improve overall efficiency.

Section 1

Methodology

Methodology

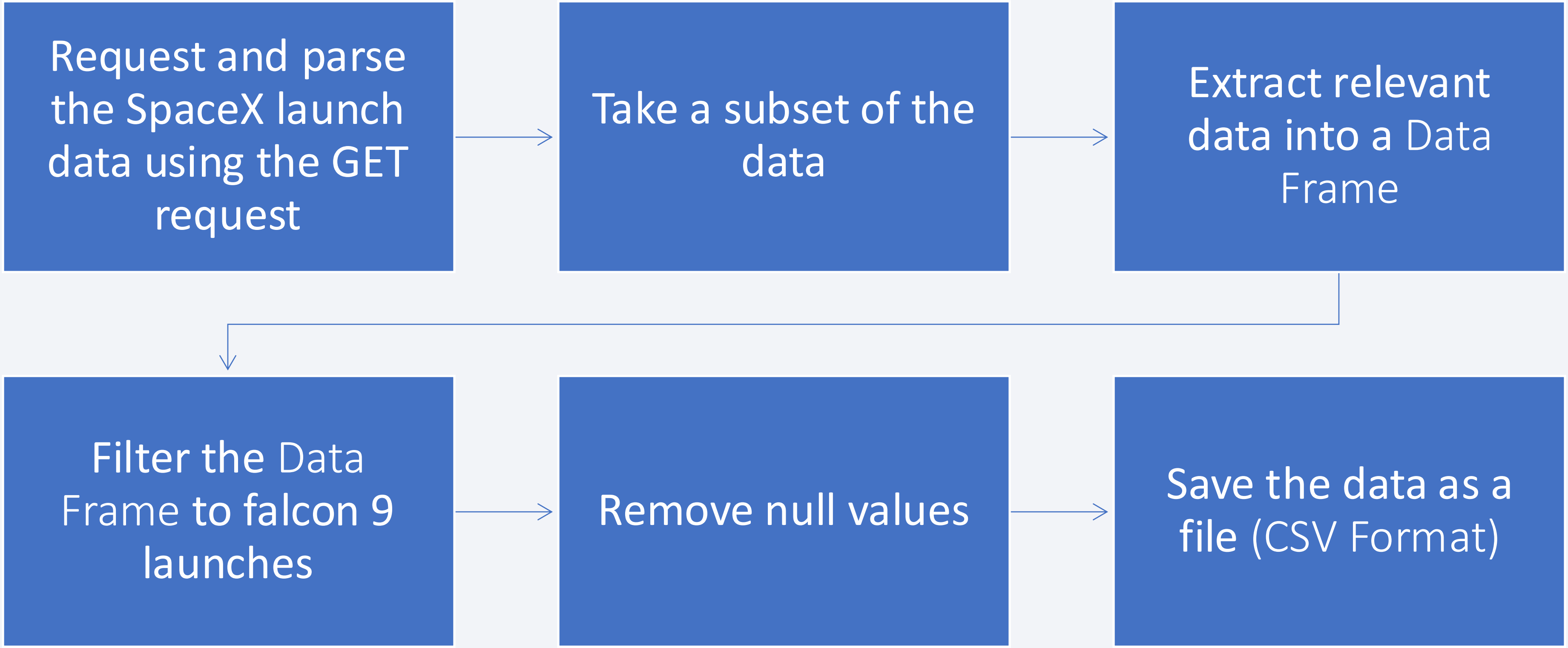
Executive Summary

- Data collection methodology:
  - Data Collection From API's
  - Web Scraping
- Perform data wrangling
  - Removing null values and imputing with mean values.
- 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

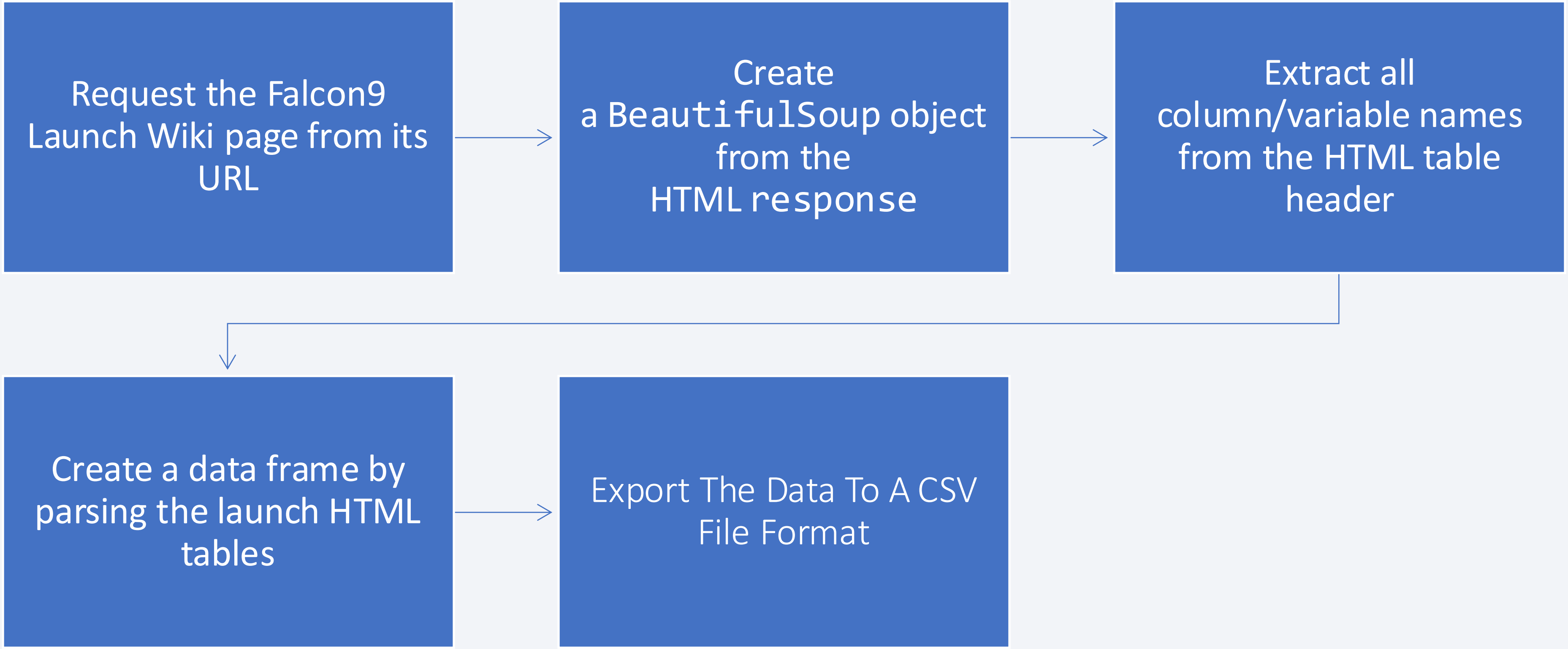
- Data was collected through API calls and Web Scraping

Data Collection – SpaceX API



Link To The Completed Data Collection Notebook: Click *Here*

Data Collection - Scraping



- Link To The Completed Web Scraping Notebook: Click *Here*

Data Wrangling

1.
- Check For Missing Values
- Check for Outliers

2.
- Impute Missing Values with Mean

3.
- Feature Engineering
- Landing class column that shows if a landing was succesful

Link to complete data wrangling notebook: Click Here

EDA with SQL & Data Visualization

- Used scatter plots to visualize correlations between variables
- Also scatter plots helped identify possible useful variables for building a predictive model
- Used Folium interactive charts to visualize launch sites and their proximities
- Used SQL Analysis to answer some basic questions we had about the dataset while performing EDA
- Used Pie Charts to visualize information about launch sites and how factors such as payload affect the success of a landing

EDA with SQL

Summary of SQL Queries Performed

- **Retrieve Distinct Launch Sites**
  - Query: `SELECT DISTINCT Launch_Site FROM SPACEXTABLE`
  - Purpose: To identify all unique launch sites used by SpaceX.
- **Retrieve Records Matching a Launch Site Pattern**
  - Query: `SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;`
  - Purpose: To fetch the first 5 records where the launch site starts with "CCA".
- **Calculate Total Payload Mass for NASA (CRS)**
  - Query: `SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass FROM SPACEXTABLE WHERE Customer LIKE 'NASA (CRS)'`
  - Purpose: To compute the total payload mass for launches carried out for NASA (CRS).

EDA with SQL

Summary of SQL Queries Performed

- **Calculate Average Payload Mass for a Specific Booster Version**
  - Query: `SELECT AVG(PAYLOAD_MASS__KG_) AS AveragePayloadMass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'`
  - Purpose: To find the average payload mass for launches using the "F9 v1.1" booster version.
- **Determine the First Successful Ground Pad Landing Date**
  - Query: `SELECT MIN(Date) AS FirstSuccessfulLandingDate FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'`
  - Purpose: To identify the earliest date a Falcon 9 first stage successfully landed on a ground pad.
- **Retrieve Boosters Used for Drone Ship Success in a Specific Payload Range**
  - Query: `SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000`
  - Purpose: To list boosters that successfully landed on a drone ship with payloads between 4000 and 6000 kg.

EDA with SQL

Summary of SQL Queries Performed

- **Count Landing Outcomes**
  - Query: `SELECT Landing_Outcome, COUNT(*) as Total FROM SPACEXTABLE GROUP BY Landing_Outcome`
  - Purpose: To group all landings by their outcome and count occurrences.
- **Retrieve Booster with Maximum Payload Mass**
  - Query: `SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);`
  - Purpose: To find the booster version associated with the heaviest payload.

EDA with SQL

Summary of SQL Queries Performed

- **Retrieve Failures by Month in 2015**
  - Query: Query to extract the month, booster version, and launch site for failures on drone ships in 2015.
  - Purpose: To analyze failure patterns and identify trends during 2015.

- **Count Landing Outcomes for a Specific Time Range**
  - Query: `SELECT Landing_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;`
  - Purpose: To analyze the distribution of landing outcomes for a defined period.

EDA with SQL

Insights Gained

- **Launch Sites**
  - SpaceX operates multiple distinct launch sites, highlighting its geographic and operational diversity.
- **NASA Payload Contributions**
  - NASA (CRS) missions have contributed a significant amount of payload mass, showcasing their reliance on SpaceX for logistical support.
- **Booster Version Performance**
  - Specific booster versions, such as "F9 v1.1", have a unique payload profile, indicating potential differences in design or mission scope.
- **Successful Ground Pad Landings**
  - The first successful ground pad landing date marks a milestone in SpaceX's reusability goals.

EDA with SQL

Insights Gained

- **Payload-Driven Success on Drone Ships**
  - Payloads between 4000 and 6000 kg have seen successful landings on drone ships, suggesting a sweet spot for this operational mode.
- **Landing Outcome Trends**
  - Certain landing outcomes are more frequent, which helps in identifying areas for improvement or consistency.
- **Payload Extremes**
  - The booster associated with the maximum payload mass demonstrates SpaceX's capability to handle heavy-lift missions.
- **Failures in 2015**
  - Failures concentrated in 2015 provide an opportunity to investigate and address historical challenges.
- **Performance Over Time**
  - Landing outcomes over the 2010-2017 timeframe reveal the evolution of SpaceX's operational efficiency.
- Link To Completed SQL Analysis Notebook: Click *Here*

Build an Interactive Map with Folium

• 1. Summary of Map Objects Added

- **Markers:**
  - **NASA Johnson Space Center Marker:**
    - A blue marker with a text label showing "NASA JSC."
    - Added to represent NASA's Johnson Space Center as the starting reference location.
  - **Launch Site Markers:**
    - Markers with text labels showing the names of different SpaceX launch sites.
    - These were used to pinpoint and label each launch site.
  - **Closest Coastline Marker:**
    - A green marker showing the closest coastline to a launch site with a popup displaying the distance.
  - **Closest Point Marker:**
    - A marker for the nearest city, highway, or railway from the launch site, with a popup displaying the calculated distance.

Build an Interactive Map with Folium

- **1. Summary of Map Objects Added**

- **Circles:**
  - **NASA Johnson Space Center Circle:**
    - A blue circle with a 1000-meter radius around the NASA JSC location.
  - **Launch Site Circles:**
    - Blue circles with a 500-meter radius around each SpaceX launch site.
  - **Success/Failure Circles:**
    - Green circles for successful launches and red circles for failed launches around the respective launch locations.
- **Lines (Polylines):**
  - **Launch Site to Coastline Line:**
    - A blue line connecting a launch site to the nearest coastline.
  - **Launch Site to Closest Point Line:**
    - A red line connecting a launch site to the nearest city, highway, or railway.
- **Mouse Position Plugin:**
  - Displays the real-time latitude and longitude of the mouse pointer.

Build an Interactive Map with Folium

- **2. Reasons for Adding These Objects**

- **Markers:**
  - To visually identify and label key locations like launch sites, coastlines, and nearby infrastructure points.
  - Helps provide clear context for spatial analysis.
- **Circles:**
  - To highlight specific areas of interest (e.g., launch sites) and to indicate launch outcomes (success or failure).
  - The radius provides a sense of scale and proximity for geographical reference.
- **Lines (Polylines):**
  - To illustrate distances and connections between key points, such as launch sites to the coastline or nearest infrastructure.
  - Useful for understanding logistical and safety considerations.
- **Mouse Position Plugin:**
  - Enables users to interactively determine coordinates on the map, aiding further exploration and analysis.

Build a Dashboard with Plotly Dash

- **1. Plots/Graphs Added:**

- **Pie Chart for Success Count of All Sites:**
  - Displays the total number of successful launches per launch site.
  - Aggregates data across all SpaceX launch sites to provide an overview of their success distribution.
- **Pie Chart for Success Count of a Specific Site:**
  - Focuses on a single launch site selected from the dropdown menu.
  - Compares the number of successful launches to failed launches for that specific site.
- **Scatter Plot: Payload Mass vs. Success:**
  - Plots payload mass on the x-axis and success/failure on the y-axis.
  - Uses color-coding for different "Booster Version Categories."
  - Dynamically updates based on user-selected payload mass range and launch site.

Build a Dashboard with Plotly Dash

- **2. Interactions Added:**

- **Dropdown for Launch Site Selection:**
  - Allows users to filter data for all sites or a specific launch site.
  - Dynamically updates both the pie chart and the scatter plot based on the selected site.
- **Payload Range Slider:**
  - Allows users to adjust the range of payload masses displayed in the scatter plot.
  - Enables analysis of success rates for specific payload ranges.
- **Dynamic Data Filtering:**
  - Ensures that all visualizations update in real time based on the dropdown and slider inputs.

Build a Dashboard with Plotly Dash

- Reasons for Adding These Plots and Interactions:

- **Pie Charts:**
  - Provide clear, intuitive insights into success rates at different launch sites.
  - Help identify sites with the highest activity and/or reliability.
  - Allow comparisons between sites or an in-depth look at a single site.
- **Scatter Plot:**
  - Visualizes the relationship between payload mass and launch success.
  - Highlights trends, such as optimal payload ranges for success or the performance of specific booster versions.
  - Adds depth to the analysis by incorporating payload mass, a key factor in mission success.
- **Dropdown and Slider Interactions:**
  - Enhance the dashboard's usability by allowing users to explore data interactively.
  - Facilitate site-specific and payload-specific analyses without overwhelming users with static visuals.
  - Provide a tailored view of the data to address specific research or operational questions.

- Link To Completed Notebook: Click Here

Predictive Analysis (Classification)

Model Development Report: Classification Model for SpaceX Launch Success

- **Model Building Process:**
  - Utilized **logistic regression, support vector machines (SVM), decision trees, and k-nearest neighbors (KNN)** as baseline classification models.
  - Extracted key features such as **payload mass, launch site, booster version, and success metrics** for training the models.
- **Evaluation Metrics:**
  - Assessed model performance using **cross-validation scores** and accuracy metrics on the test data.
  - Computed individual accuracies for each model:
    - Logistic Regression: 83%
    - SVM: 82%
    - Decision Tree: 80%
    - KNN: 80%.
- **Model Improvement Techniques:**
  - Performed **hyperparameter tuning** using GridSearchCV to optimize model parameters (e.g., regularization strength for Logistic Regression and kernel types for SVM).
  - Standardized the dataset to improve performance for models sensitive to scale, such as SVM and KNN.
- **Best Performing Model:**
  - Identified **Logistic Regression** as the best-performing model with the highest accuracy (83%) and consistent results across test data.
- **Key Insights:**
  - Logistic Regression excelled due to its ability to handle binary classification effectively, while other models were sensitive to feature distribution and dataset size.
  - Hyperparameter tuning and feature scaling were critical steps in achieving optimal performance.
- **Outcome:**
  - The final model is robust and efficient, providing actionable predictions for SpaceX launch success based on historical data.

Results: Part 1

Short Report: Results and Insights

- **Dataset Characteristics:**
  - The dataset contains key features such as `Launch Site`, `Payload Mass (kg)`, `class` (success/failure), and `Booster Version`.
  - Payload mass ranges widely, with notable clustering at specific intervals.
- **Success Distribution:**
  - Success rates vary significantly across launch sites, highlighting site-specific performance differences.
  - Booster versions and payload mass appear to influence launch success.
- **Correlation Insights:**
  - Higher payload mass may reduce success probability, indicating a need for payload optimization.
  - 2. Results/Insights from Folium Maps and Plotly Dashboards
- **Folium Maps:**
  - Launch sites are geographically distributed, with notable clusters in regions supporting specific space missions.
  - Proximity to water bodies (safety measure for launches) is a consistent pattern across sites.
- **Pie Charts:**
  - The majority of launches occur at specific high-performing sites, such as `CCAFS LC-40`, which exhibits a strong success ratio.

Results: Part 2

Short Report: Results and Insights

- **Scatterplots:**
  - Clear correlation between payload mass and success rate:
    - Lighter payloads generally show higher success probabilities.
    - Some booster versions demonstrate consistent high performance irrespective of payload.
- **Interactive Features:**
  - Filters like launch site and payload range allow granular insights, showing nuanced patterns by site or payload type.
  - 3. Results from Predictive Analysis Modeling with Machine Learning
- **Model Comparison:**
  - Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors were evaluated.
  - Accuracy scores ranged between 0.8 and 0.9, demonstrating reliable predictions based on launch data.
- **Best Model:**
  - Logistic Regression achieved the highest accuracy, suggesting strong predictive power for binary outcomes like success/failure.
- **Key Predictive Insights:**
  - Payload mass and booster version are influential features for predicting launch success.
  - Site-specific trends indicate that incorporating location data improves prediction reliability.
- **Application:**
  - The models provide actionable insights to optimize payload configurations and select suitable launch sites for future missions.
  - This multi-faceted analysis bridges historical data, geographic insights, and predictive modeling to enhance decision-making in space launch operations.

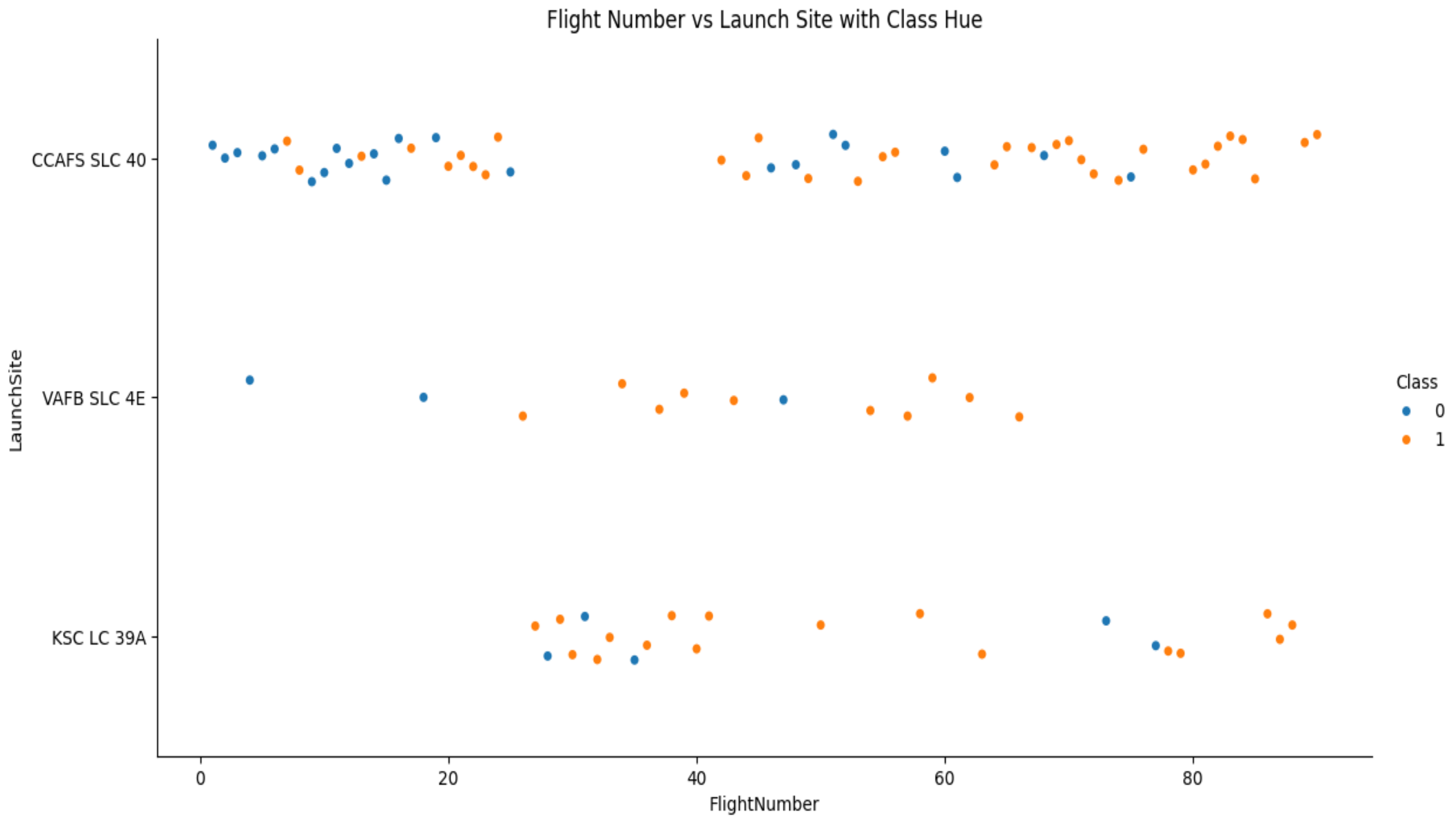
Section 2

Insights drawn from EDA

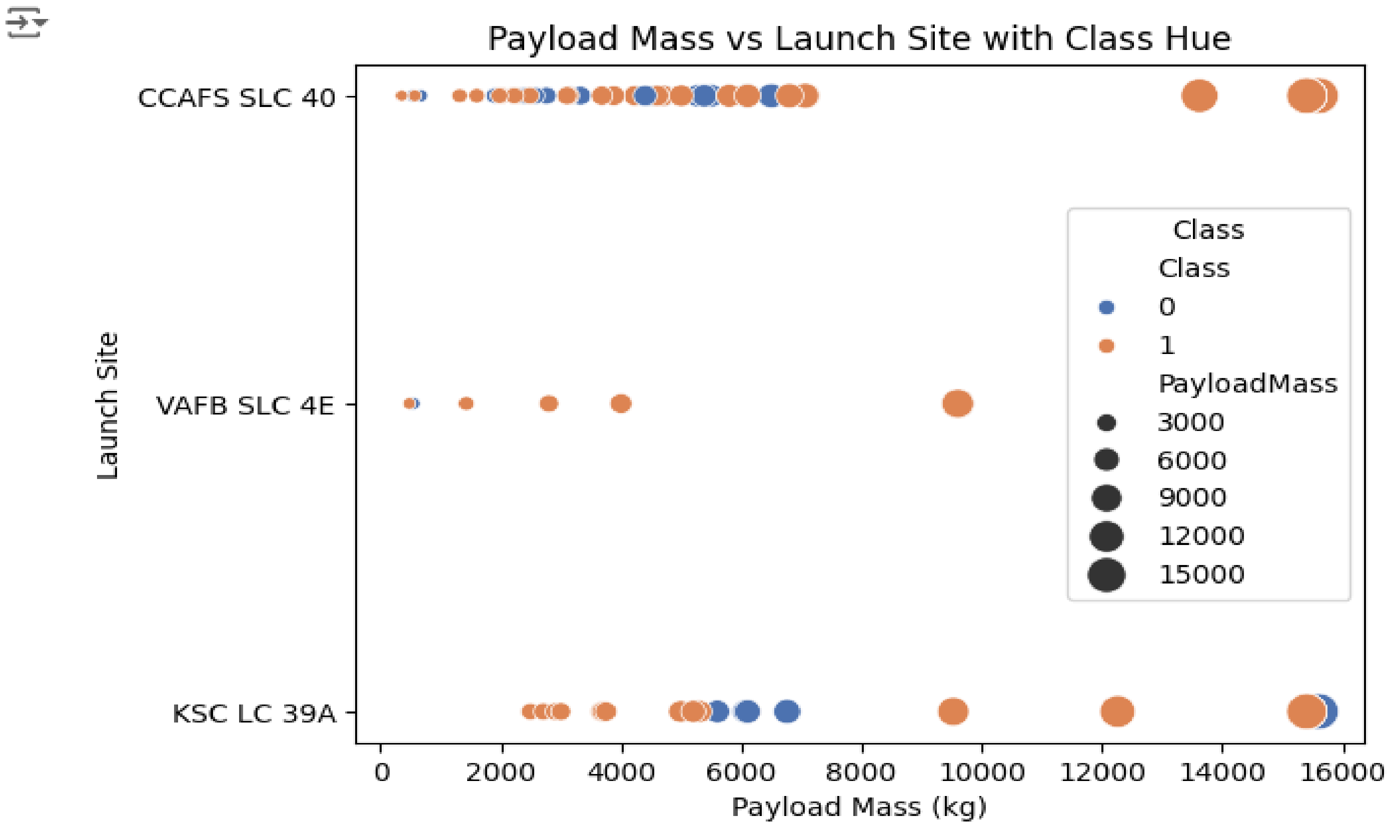
Flight Number vs. Launch Site

```
    aspect=2
)

plt.title('Flight Number vs Launch Site with Class Hue')
plt.show()
```



Payload vs. Launch Site



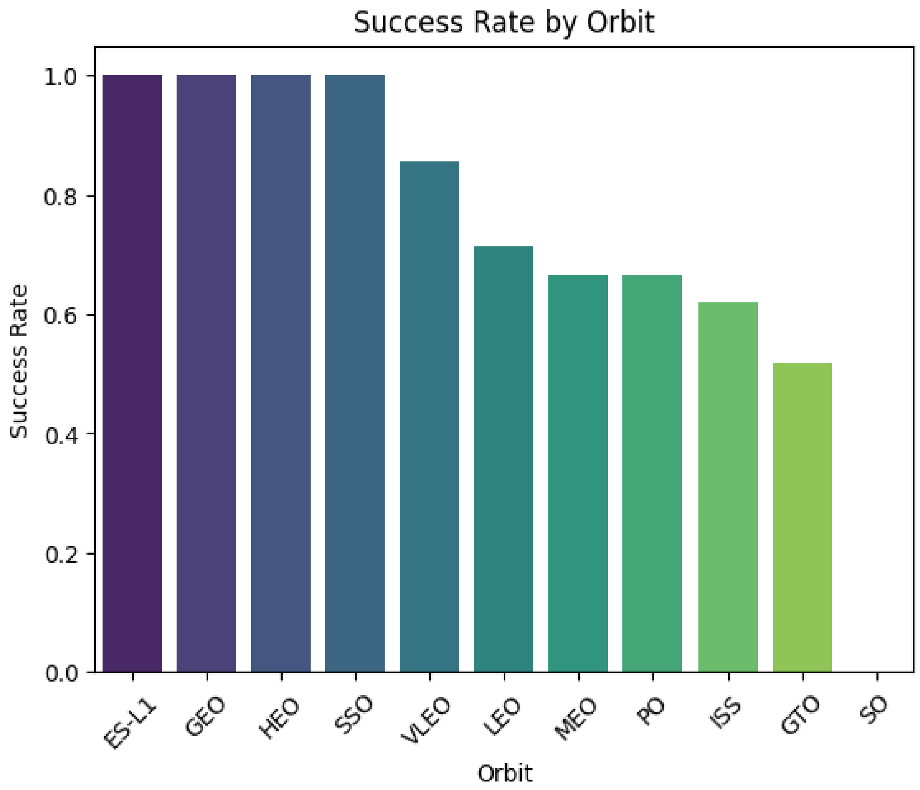
Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

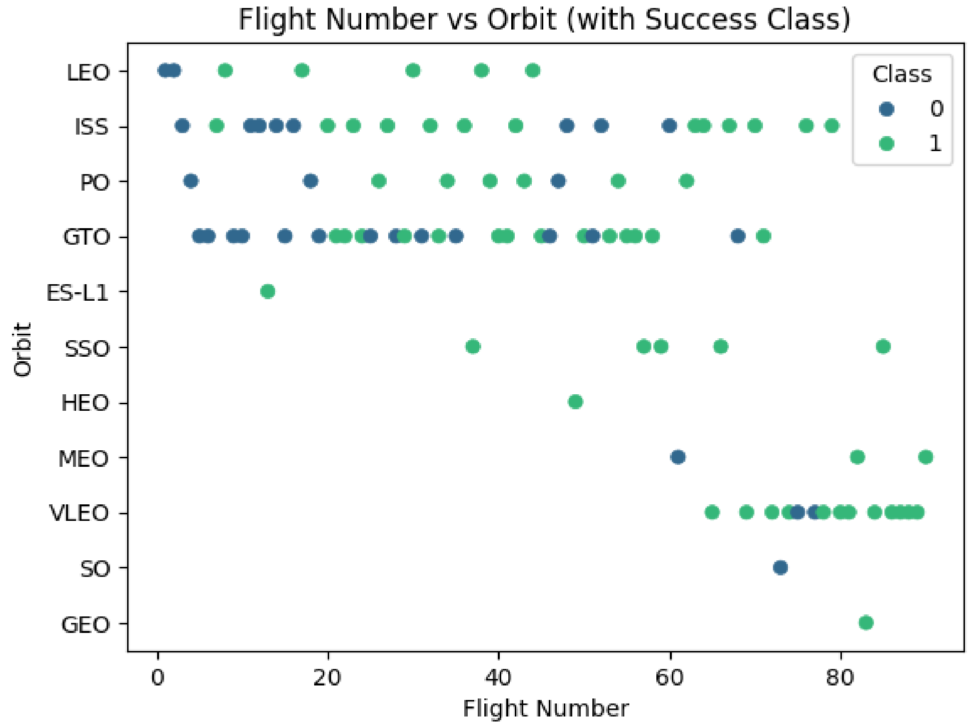
```
<ipython-input-6-2e584c81d3cb>:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x='Orbit', y='Class', data=success_rate_by_orbit, palette='viridis')
```

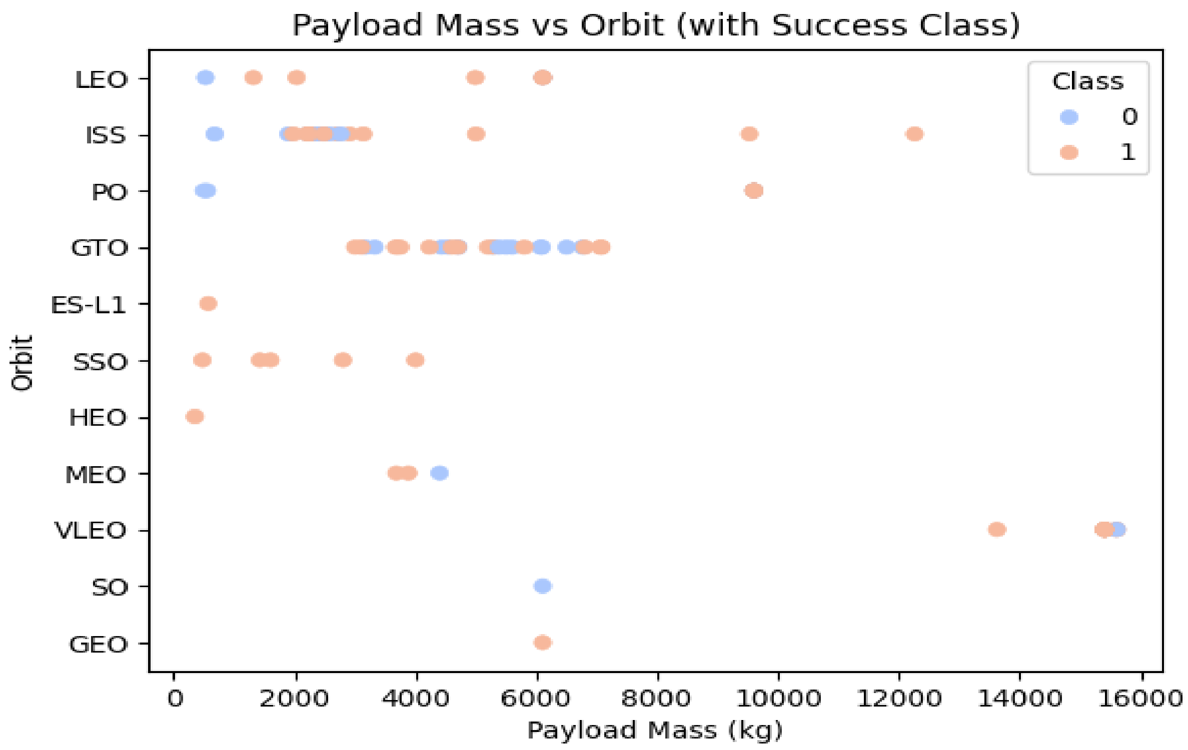


Flight Number vs. Orbit Type



You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

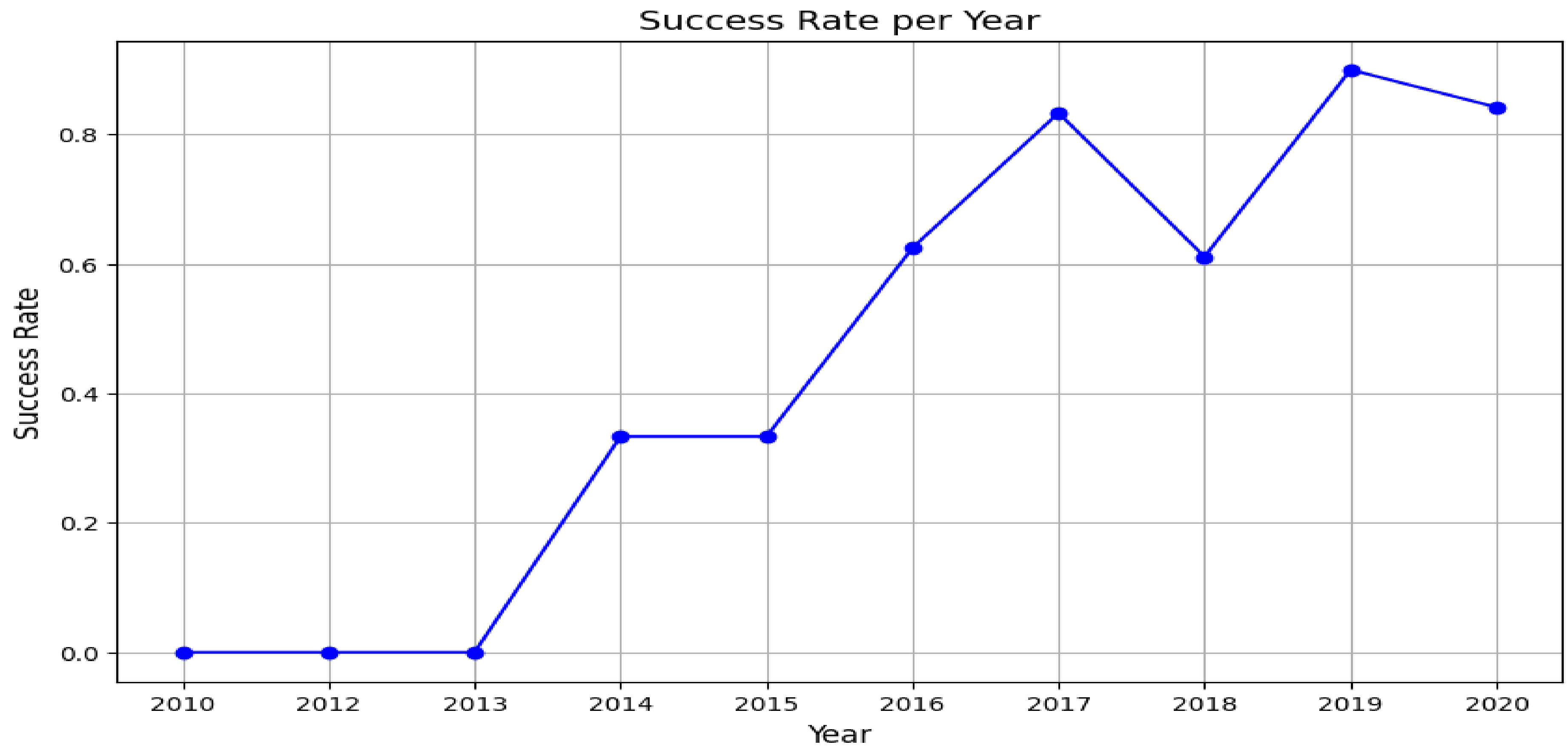
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- **Retrieve Distinct Launch Sites**
  - Query: `SELECT DISTINCT Launch_Site FROM SPACEXTABLE`
  - Purpose: To identify all unique launch sites used by SpaceX.

- **QUERY RESULT**

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- **Retrieve Records Matching a Launch Site Pattern**
  - Query: `SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;`
  - Purpose: To fetch the first 5 records where the launch site starts with "CCA".

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

In [13]:
```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Out[13]:

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

Total Payload Mass

- **Calculate Total Payload Mass for NASA (CRS)**
  - Query: `SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass FROM SPACEXTABLE WHERE Customer LIKE 'NASA (CRS)'`
  - Purpose: To compute the total payload mass for launches carried out for NASA (CRS).

QUERY RESULT

	TotalPayloadMass
0	45596

Average Payload Mass by F9 v1.1

- **Calculate Average Payload Mass for a Specific Booster Version**
  - Query: `SELECT AVG(PAYLOAD_MASS__KG_) AS AveragePayloadMass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'`
  - Purpose: To find the average payload mass for launches using the "F9 v1.1" booster version.

QUERY RESULT

	AveragePayloadMass
0	2928.4

First Successful Ground Landing Date

- **Determine the First Successful Ground Pad Landing Date**
  - Query: `SELECT MIN(Date) AS FirstSuccessfulLandingDate FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'`
  - Purpose: To identify the earliest date a Falcon 9 first stage successfully landed on a ground pad.

QUERY RESULT

	FirstSuccessfulLandingDate
0	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Retrieve Boosters Used for Drone Ship Success in a Specific Payload Range**
  - Query: `SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000`
  - Purpose: To list boosters that successfully landed on a drone ship with payloads between 4000 and 6000 kg.

QUERY RESULT

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- **Count Landing Outcomes**
  - Query: `SELECT Landing_Outcome, COUNT(*) as Total FROM SPACEXTABLE GROUP BY Landing_Outcome`
  - Purpose: To group all landings by their outcome and count occurrences.

QUERY RESULT

	Landing_Outcome	Total
0	Controlled (ocean)	5
1	Failure	3
2	Failure (drone ship)	5
3	Failure (parachute)	2
4	No attempt	21
5	No attempt	1
6	Precluded (drone ship)	1
7	Success	38
8	Success (drone ship)	14
9	Success (ground pad)	9
10	Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- **Retrieve Booster with Maximum Payload Mass**
  - Query: `SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);`
  - Purpose: To find the booster version associated with the heaviest payload.

QUERY RESULT

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

2015 Launch Records

- **Retrieve Failures by Month in 2015**
    - Query: Query to extract the month, booster version, and launch site for failures on drone ships in 2015.
    - Purpose: To analyze failure patterns and identify trends during 2015.

QUERY RESULT

	month	Booster_Version	Launch_Site
0	January	F9 v1.1 B1012	CCAFS LC-40
1	April	F9 v1.1 B1015	CCAFS LC-40

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

- **Count Landing Outcomes for a Specific Time Range**
  - Query: `SELECT Landing_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;`
  - Purpose: To analyze the distribution of landing outcomes for a defined period.

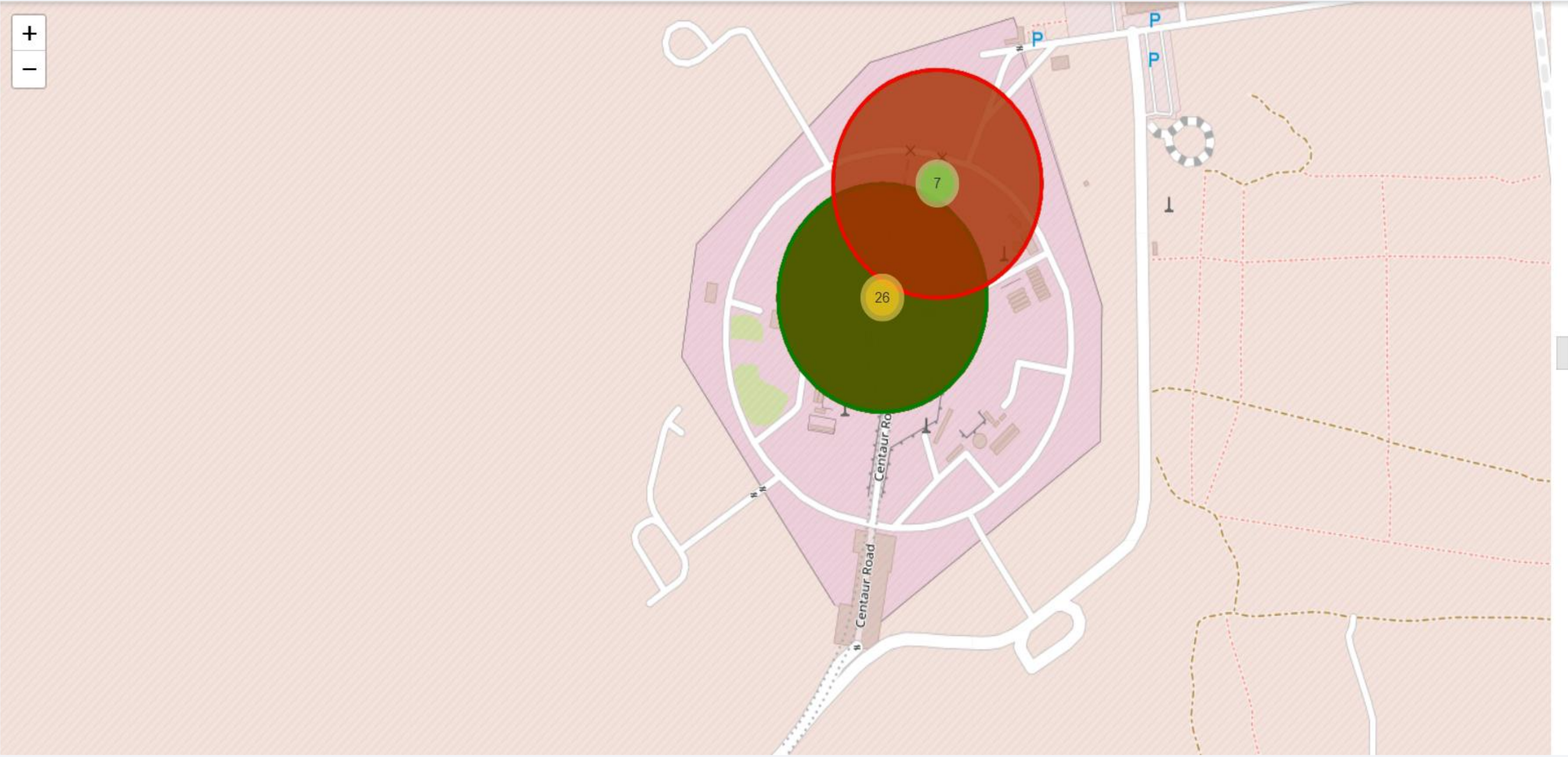
QUERY RESULT

	Landing_Outcome	outcome_count
0	No attempt	10
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2

Section 3

Launch Sites Proximities Analysis

1. Launch Sites and Success/Failure Outcomes Map (Screenshot)



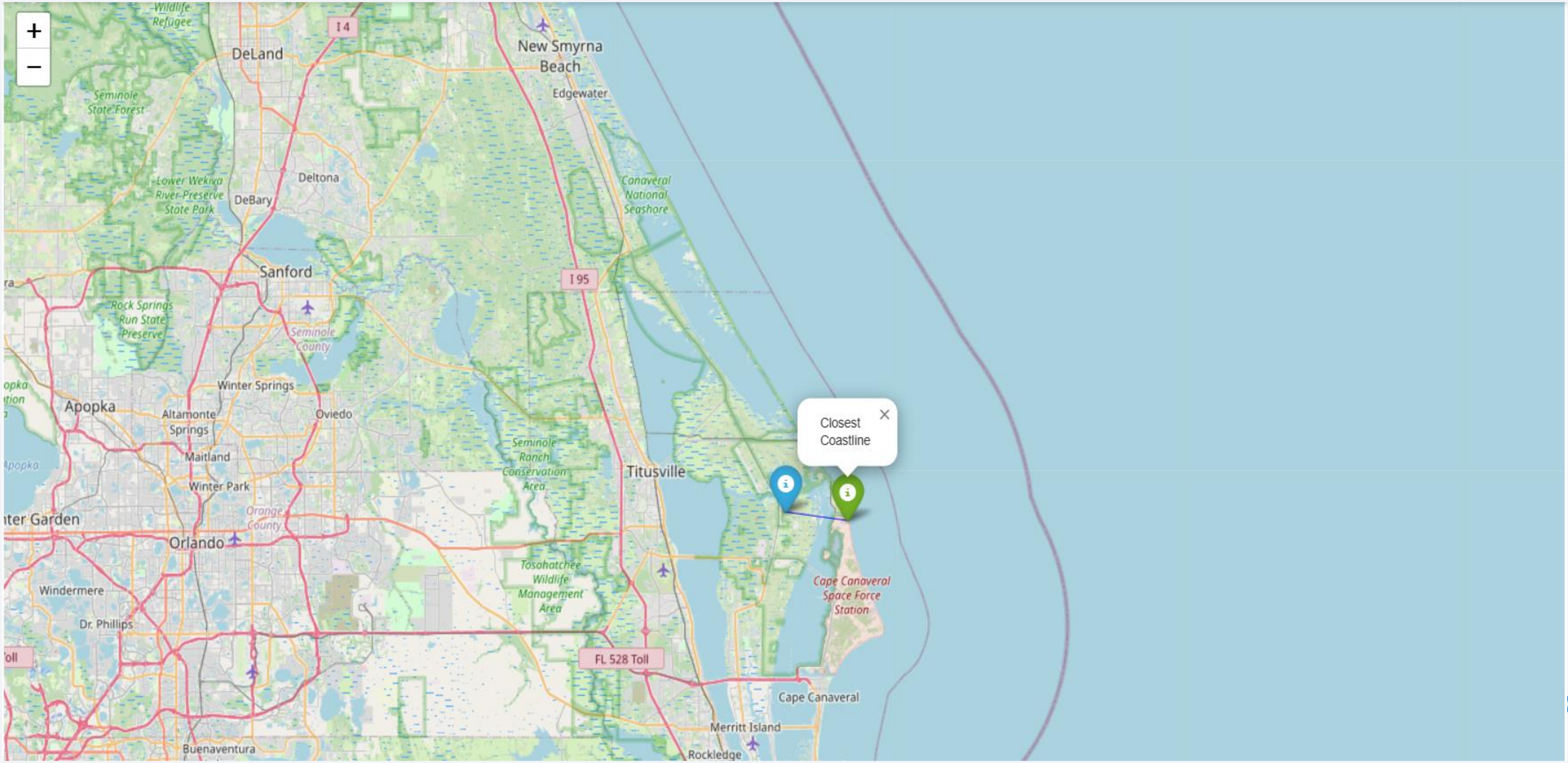
1. Launch Sites and Success/Failure Outcomes Map (Insights + Explanation)

- **Important Elements:**
  - Green circles represent successful launches, and red circles denote failures.
  - Each circle is centered on a specific launch site's coordinates.
  - Popups display the launch site's name and the success or failure status.

- **Findings/Insights:**
  - Visualizing success and failure outcomes at specific locations provides insight into the reliability of launches across different sites.
  - Patterns in success rates at particular launch sites may inform site-specific operational strategies or infrastructure improvements.

- Link To Completed Notebook: Click Here

2. Launch Site and Closest Coastline Distance Map (Screenshot)



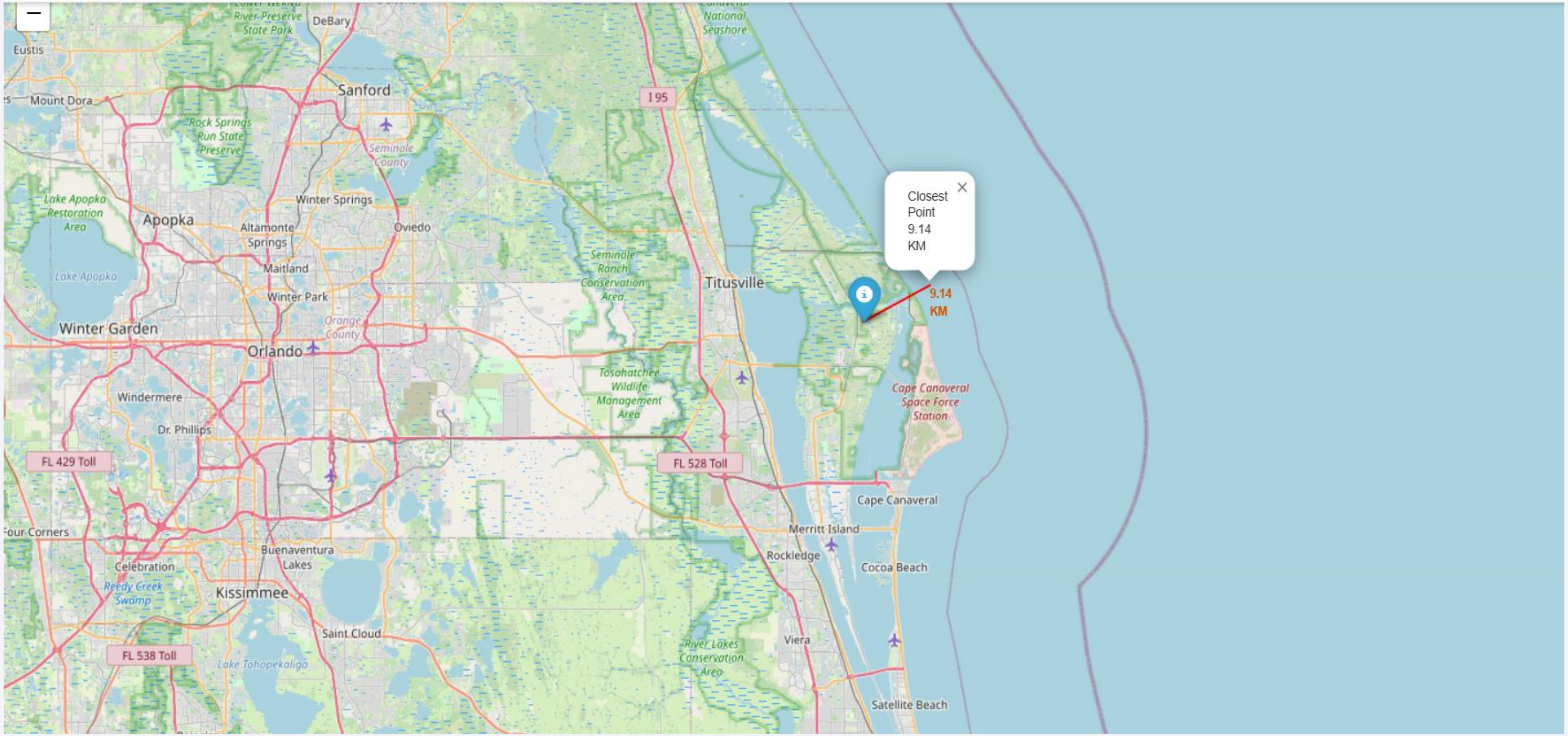
2. Launch Site and Closest Coastline Distance Map (Insights + Explanation)

- **Important Elements:**
  - Markers identify the launch site and the closest point on the coastline.
  - A blue PolyLine connects the launch site to the coastline.
  - A popup on the coastline marker displays the calculated distance in kilometers.

- **Findings/Insights:**
  - The proximity of launch sites to coastlines highlights safety considerations, as launches near the coast minimize risk to populated areas in case of anomalies.
  - Distances also inform logistical planning, including access to recovery ships for booster landings.

- Link To Completed Notebook: Click Here

3. Launch Site and Closest Infrastructure Map (Screenshot)



3. Launch Site and Closest Infrastructure Map (Insights + Explanation)

- **Important Elements:**
  - Markers represent the launch site and the nearest infrastructure (e.g., city, highway, railway).
  - A red PolyLine connects the launch site to the nearest point of infrastructure.
  - A popup at the infrastructure marker displays the calculated distance in kilometers.

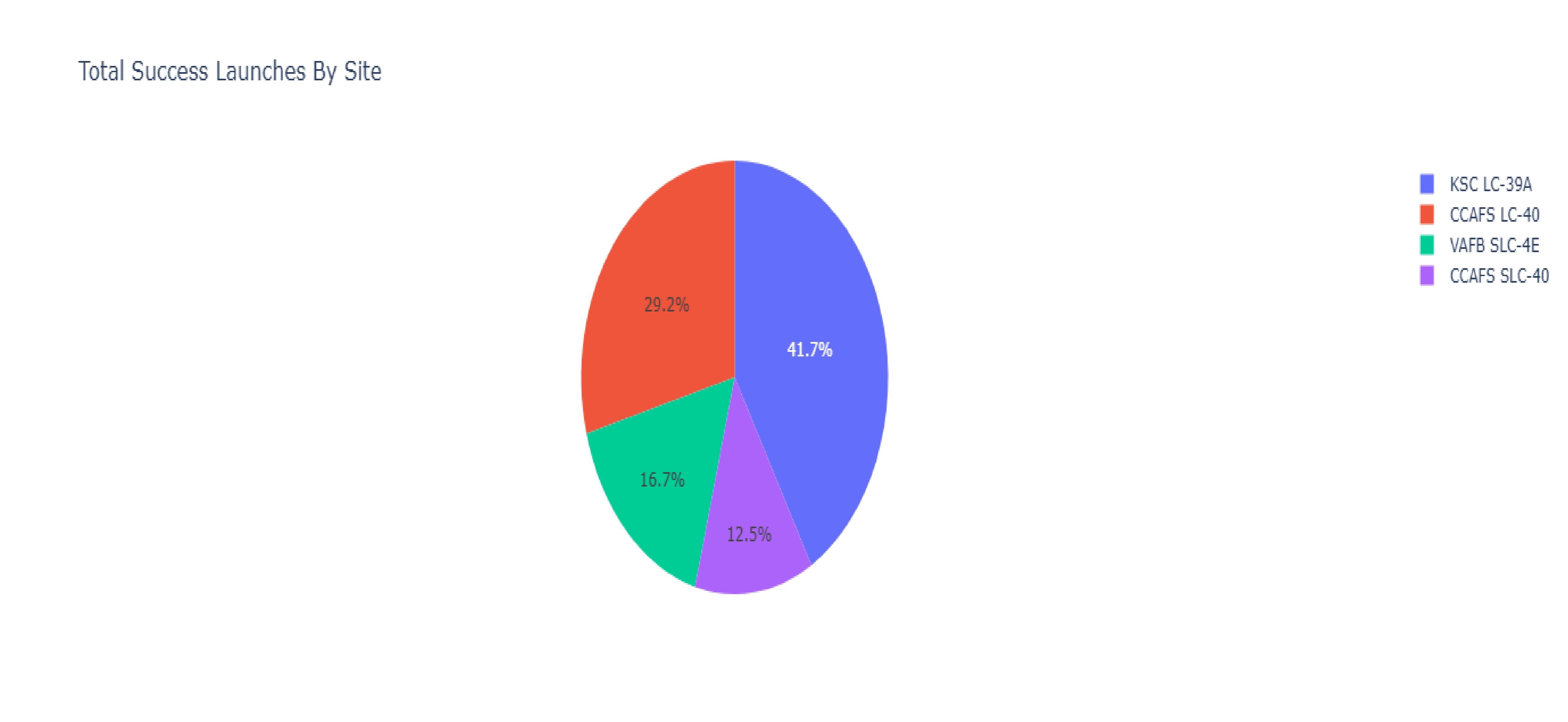
- **Findings/Insights:**
  - This map illustrates the accessibility of critical infrastructure near launch sites.
  - Shorter distances to infrastructure may indicate logistical efficiency for transporting equipment and personnel.
  - Helps assess the feasibility of emergency response or supply chain requirements.

- Link To Completed Notebook: Click Here

Section 4

Build a Dashboard with Plotly Dash

1. Pie Chart Showing the Success Count of All Sites: Screenshot



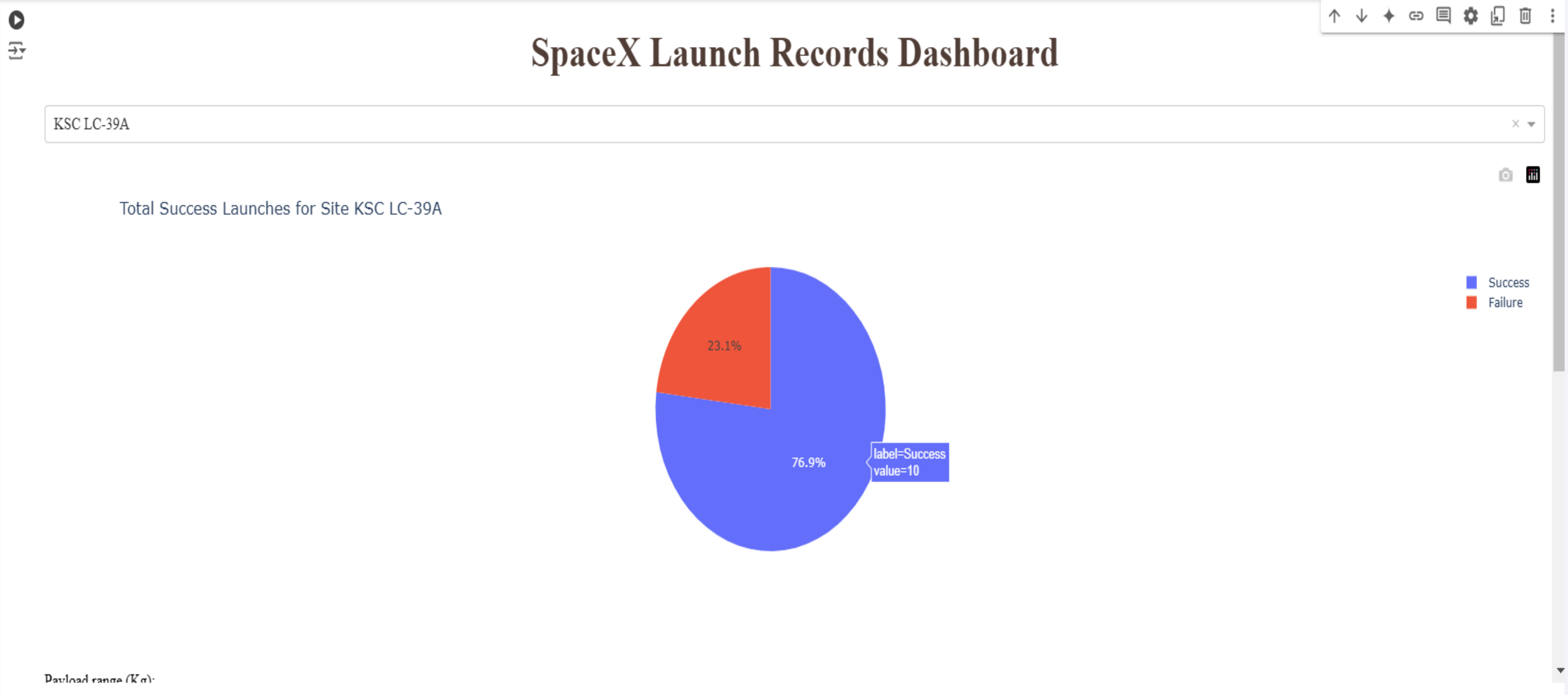
1. Pie Chart Showing the Success Count of All Sites: Insights

- **Important Elements:**
  - Displays the total count of successful launches for each launch site.
  - Aggregates data from all SpaceX sites into a single visualization.
  - Uses a proportional chart format to highlight the relative contributions of each site.

- **Findings/Insights:**
  - The launch site 'KSC LC-39A' has the highest success ratio of all the launch sites
  - Identifies which that this site has the highest number of successful launches, reflecting its reliability or frequency of use.
  - Shows the distribution of success across all SpaceX sites, helping identify the most productive launch locations.
  - Also shows lowest performing launch site showing which launch site can be cut if its not performing well to save money.

- Link To Completed Notebook: Click Here

2. Pie Chart with the Highest Launch Success Ratio: Screenshot



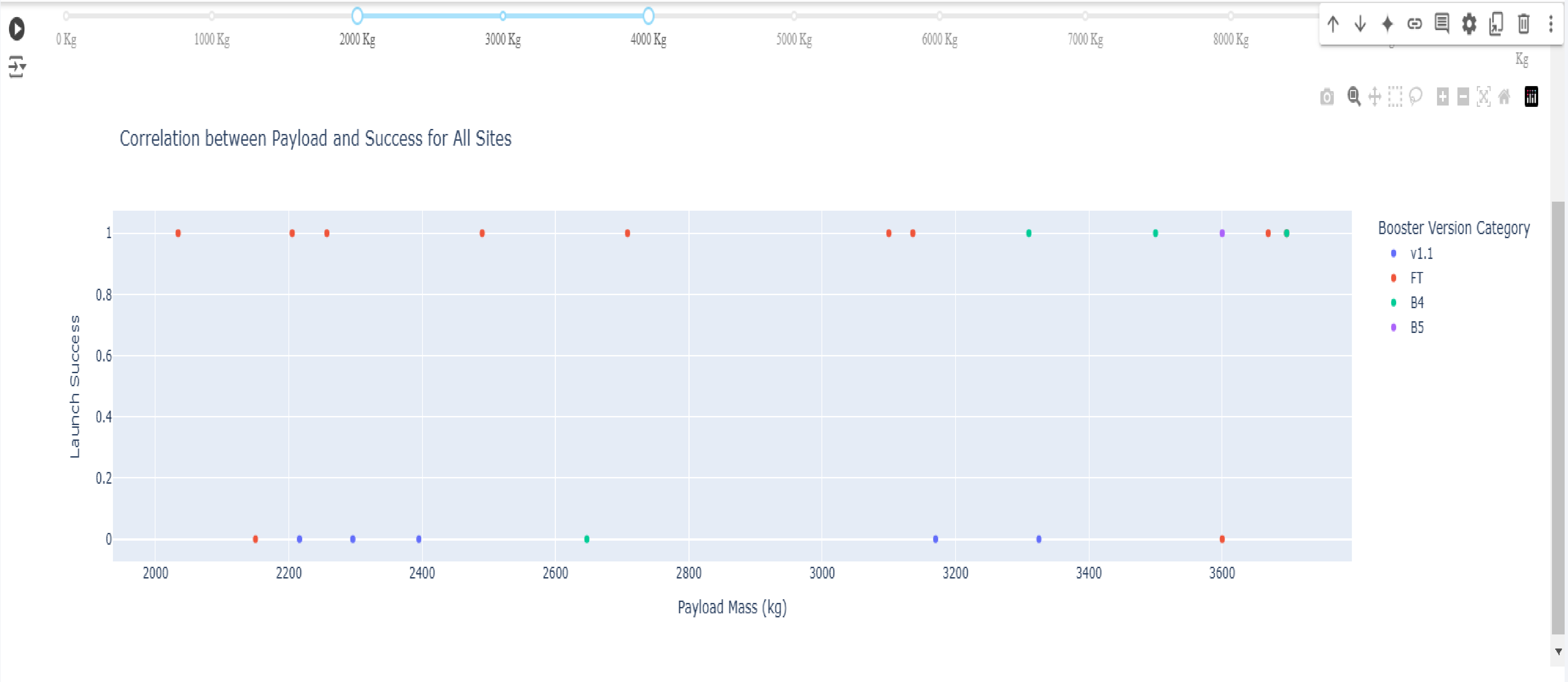
2. Pie Chart with the Highest Launch Success Ratio: Insights + Findings

- **Important Elements:**
  - Focuses on a specific site selected through the dropdown menu.
  - Compares the count of successful and failed launches at the selected site.
  - Provides a clear success-to-failure ratio for site-specific analysis.

- **Findings/Insights:**
  - Highlights the site with the highest success ratio, indicating its efficiency and reliability.
  - A site with a higher success ratio might have better infrastructure, operational strategies, or less challenging launch conditions.

- Link To Completed Notebook: Click Here

3. Scatter Plot: Payload Mass vs. Success: Screenshot



3. Scatter Plot: Payload Mass vs. Success: Insights + Findings

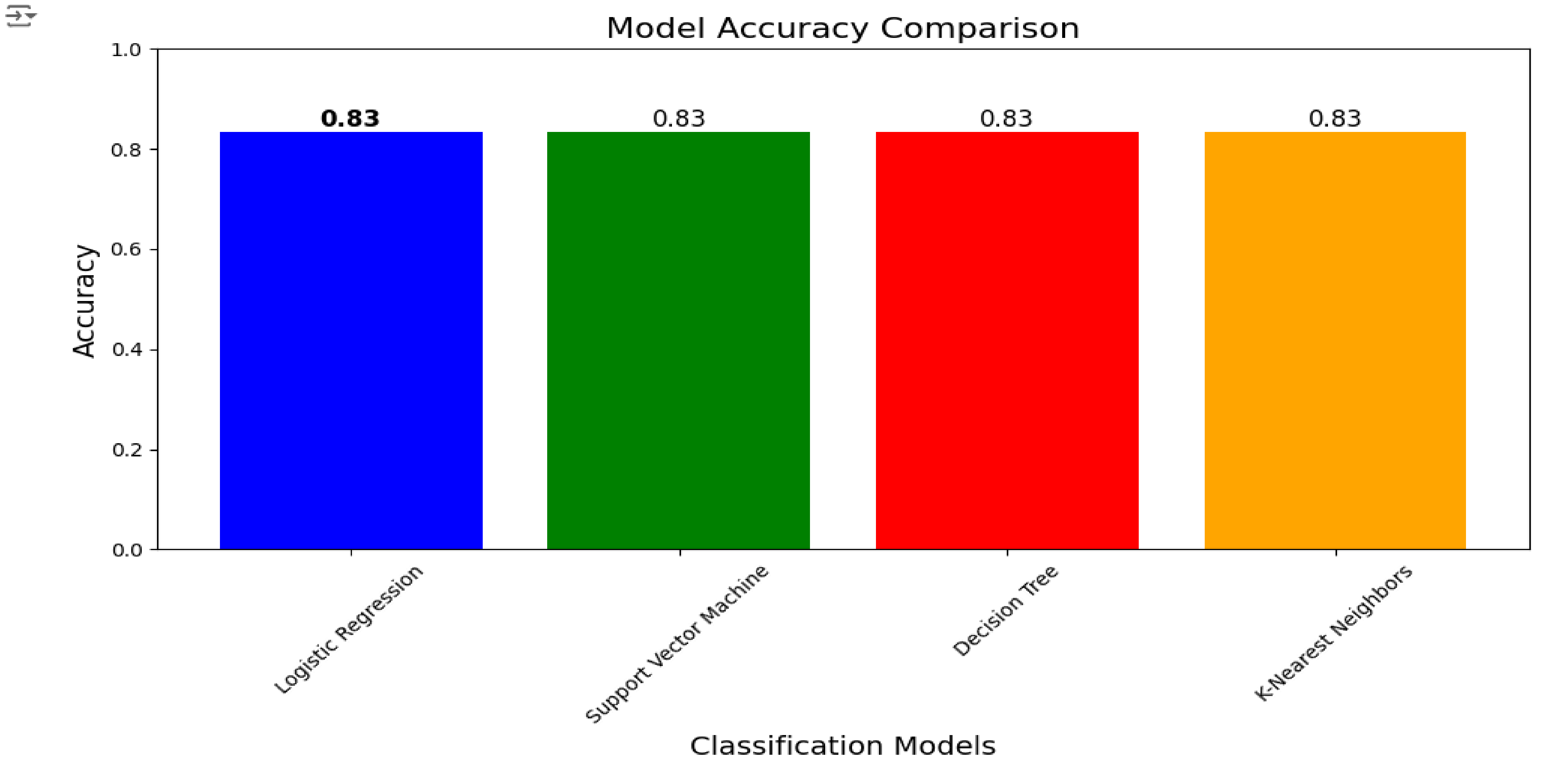
- **Important Elements:**
  - Displays the relationship between payload mass and launch success.
  - Includes a color-coded "Booster Version Category" for additional context.
  - Filters dynamically based on user-selected payload range and launch site.
- **Findings/Insights:**
  - **Payload Range with Highest Success Rate:**
    - Identifies payload ranges that consistently achieve successful launches. From ***2000-5000 KG.***
    - Typically, mid-range payloads might have higher success rates due to engineering and operational constraints.
  - **Booster Version with Largest Success Rate:**
    - Certain booster versions show higher success rates, highlighting advancements in technology and reliability.
    - The **FT** & **B5** Boosters have high success
    - Provides insight into which booster versions are optimal for specific payload ranges.
  - **Correlation Insights:**
    - Larger payloads may have a lower success rate if they push the limits of the rocket's capabilities.
    - This can prove to be costly if finances are a constraint.
    - Identifies trends that could inform future mission planning and design.

- Link To Completed Notebook: Click Here

Section 5

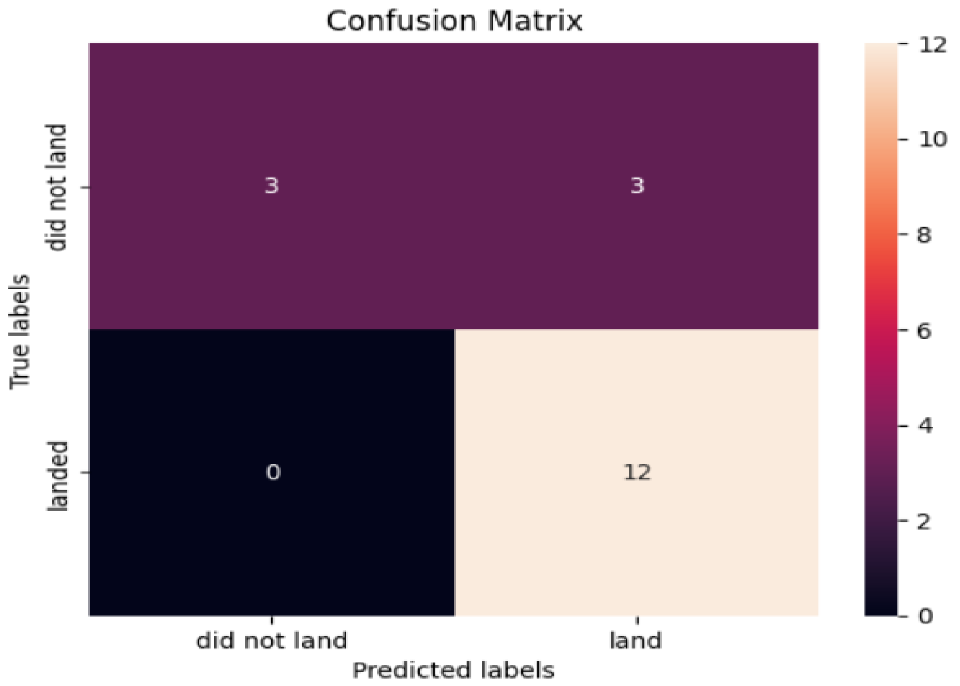
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix (Linear Regression Model)

[17]:
```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the problem is false positives.

Overview:

True Postive - 12 (True label is landed, Predicted label is also landed)

False Postive - 3 (True label is not landed, Predicted label is landed)

Conclusions

Final Conclusions of the SpaceX Data Science Project

- **Launch Success Factors:**
  - Launch success is influenced by a combination of payload mass, launch site, and booster version, with lighter payloads and certain booster versions showing higher success rates.
- **High-Performance Launch Sites:**
  - Specific launch sites, such as KSC LC-39A, demonstrate consistently higher success ratios, making them critical for mission planning and resource allocation.
- **Geographical Insights:**
  - Folium maps revealed that all launch sites are strategically located near water bodies, underscoring the safety considerations inherent in site selection.
- **Interactive Dashboards:**
  - Plotly-based dashboards provided actionable insights, enabling stakeholders to explore success rates by payload range, site, and booster version interactively.
- **Predictive Model Performance:**
  - Logistic Regression emerged as the most accurate classification model, effectively predicting launch outcomes based on key features, with a high accuracy score.
- **Operational Optimization:**
  - The integration of exploratory, geographical, and predictive analyses offers a comprehensive framework for improving mission planning, optimizing payload configurations, and selecting launch sites to maximize success.

Appendix

- 0. Link To All Completed Project Files: Click Here
- 1. Link To The Completed API Data Collection Notebook: Click Here
- 2. Link To The Web Scraping Notebook: Click Here
- 3. Link To Data Wrangling Notebook: Click Here
- 4. Link To EDA With SQL Notebook: Click Here
- 5. Link To Completed Notebook: Click Here
- 6. Link To Launch Site Analysis Notebook: Click Here
- 7. Link To SpaceX Plotly Dashboard Notebook: Click Here
- 8. Link To Machine Learning Notebook: Click Here

Thank you!