IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of Methodologies:

1. Data Collection: Data was collected from various sources, including SpaceX launch records, and processed for analysis.
2. Data Wrangling: The collected data underwent cleaning, transformation, and integration to prepare it for analysis.
3. Exploratory Data Analysis (EDA): Both SQL queries and visualization techniques were employed to explore the dataset, identify patterns, correlations, and key insights.
4. Interactive Visual Analytics: Interactive visual analytics tools like Folium and Plotly Dash were utilized to create dynamic and intuitive visualizations for deeper exploration of geospatial and dashboard insights.
5. Predictive Analysis: Classification models were developed, tuned, and evaluated using historical data to predict launch success, providing predictive capabilities for informed decision-making.

- Summary of Results:

1. Insights into Launch Success Factors: EDA revealed several factors influencing launch success, including payload mass, launch site, and historical trends.
2. Geospatial Analysis: Visualizations using Folium provided geographical insights into launch locations and success rates across different sites.
3. Interactive Dashboards: Plotly Dash enabled the creation of interactive dashboards, facilitating dynamic exploration of launch data and insights.
4. Predictive Modeling: Classification models demonstrated the potential to predict launch success with reasonable accuracy, offering valuable decision support for SpaceX and stakeholders.
5. Continuous Improvement: Iterative model building and evaluation highlighted opportunities for continuous improvement in predictive accuracy and decision support capabilities.
6. Strategic Decision Support: The combined methodologies provided valuable insights and tools for strategic decision-making in space exploration and rocket launch operations, enhancing efficiency and success rates.

Introduction

- Introduction:
- The commercial space industry is booming with companies like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX making space travel more accessible. SpaceX stands out for its achievements in sending spacecraft to the International Space Station, deploying Starlink satellite internet, and conducting manned missions. One key factor in SpaceX's success is its ability to reuse rocket stages, significantly reducing launch costs.

- Problems to Address:
- 1. Determining the cost of each SpaceX launch, which heavily depends on the successful recovery and reuse of the first stage.
- 2. Predicting whether SpaceX will be able to reuse the first stage for a particular launch based on mission parameters like payload, orbit, and customer requirements. This prediction will be made using machine learning techniques and publicly available data.

Section 1

Methodology

Methodology

- Executive Summary
- Data Collection:
  Data was collected from diverse sources including SpaceX launch records and NASA databases using web scraping, API calls, and manual entry.
- Data Wrangling:
  Raw data underwent cleaning and preprocessing, addressing issues like missing values, duplicates, and format standardization.
- Exploratory Data Analysis (EDA):
  EDA was conducted visually using histograms, scatter plots, and bar charts, supplemented by SQL queries for deeper analysis.
- Interactive Visual Analytics:
  Folium and Plotly Dash facilitated dynamic and interactive visualization for geospatial analysis and dashboard creation.
- Predictive Analysis:
  Classification models like logistic regression and decision trees were employed for predictive modeling, with rigorous evaluation using metrics like accuracy and F1-score.

Data Collection

- Data Collection:

1. Utilize SpaceX REST API to gather launch data, specifically from the endpoint api.spacexdata.com/v4/launches/past.
2. Make a GET request to the API using the requests library.
3. Convert the JSON response to a dataframe using json_normalize function.
4. Extract relevant information from the dataframe, such as booster version, payload mass, orbit, launch site, landing outcome, etc.
5. Use additional API calls targeting different endpoints for detailed information about each launch, such as rocket, payloads, launchpad, and cores.
6. Apply helper functions to extract data from the rocket, payloads, launchpad, and cores columns.
7. Store the extracted data in lists and combine them to create a new dataframe.
8. Filter the dataframe to include only Falcon 9 launches.
9. Deal with missing values, such as replacing NaN values in the PayloadMass column with the mean.
10. Export the cleaned dataset to a CSV file for further analysis.

Data Collection - FlowChart

- Start
- --> Make GET request
- --> Convert JSON to dataframe
- --> Extract relevant information
- --> Apply helper functions
- --> Combine extracted data
- --> Filter data
- --> Deal with missing values
- --> Export cleaned dataset
- --> End

Data Collection – SpaceX API

- Data Collection – SpaceX API:
- For the data collection process using the SpaceX REST API, the following key phrases summarize the steps involved:

1. Make a GET request to the SpaceX API.
2. Convert the JSON response to a dataframe.
3. Extract relevant information such as booster version, payload mass, orbit, launch site, landing outcome, etc.
4. Utilize additional API calls targeting different endpoints for detailed information about each launch, including rocket, payloads, launchpad, and cores.
5. Apply helper functions to extract data from specific columns.
6. Combine the extracted data into a comprehensive dataset.
7. Filter the dataset to include only Falcon 9 launches.
8. Handle missing values, such as replacing NaN values in the PayloadMass column with the mean.
9. Export the cleaned dataset to a CSV file for further analysis.

```
Start
  |
  |
  V
Make GET request --> Convert JSON to dataframe --> Extract relevant information
  |
  |
  V
Apply helper functions --> Combine extracted data --> Filter data --> Deal with missing values
  |
  |
  V
Export cleaned dataset --> End
```

Data Collection - Scraping

- Data Collection - Scraping:
- For the web scraping process to collect Falcon 9 launch records from the Wikipedia page, the following key phrases summarize the steps involved:

1. Make an HTTP GET request to the Wikipedia page containing Falcon 9 launch records.
2. Create a BeautifulSoup object to parse the HTML content.
3. Extract relevant column/variable names from the HTML table header.
4. Iterate through each table row and extract launch record details.
5. Fill up a dictionary with extracted launch records.
6. Create a Pandas dataframe from the dictionary.

```
Start
    |
    |
    V
Make HTTP GET request --> Create BeautifulSoup object -->
Extract column/variable names
    |
    |
    V
Iterate through table rows --> Extract launch record details -->
Fill up dictionary
    |
    |
    V
Create Pandas dataframe --> End
```

Data Wrangling

- Data Wrangling:
- For the data wrangling process, the following key phrases summarize the steps involved:

1. Load the Space X dataset.
2. Identify and calculate the percentage of missing values in each attribute.
3. Identify numerical and categorical columns.
4. Calculate the number of launches on each site.
5. Calculate the number and occurrence of each orbit.
6. Calculate the number and occurrence of mission outcomes.
7. Create a landing outcome label based on the Outcome column.

Data Wrangling - Flowchart

- Start
- --> Load Space X dataset
- --> Identify missing values
- --> Identify numerical and categorical columns
- --> Calculate number of launches on each site
- --> Calculate number and occurrence of each orbit
- --> Calculate number and occurrence of mission outcomes
- --> Create landing outcome label based on Outcome column
- --> End

EDA with Data Visualization

- Here are the charts plotted during the Exploratory Data Analysis with Data Visualization:

1. FlightNumber vs. PayloadMass vs. Launch Outcome: A scatter plot showing the relationship between FlightNumber and PayloadMass, with the launch outcome (success or failure) indicated by color. This chart helps visualize the impact of FlightNumber and PayloadMass on the success of the launch.
2. FlightNumber vs. LaunchSite vs. Launch Outcome: A scatter plot showing the relationship between FlightNumber and LaunchSite, with the launch outcome (success or failure) indicated by color. This chart helps visualize if there's any correlation between the launch site and the success of the launch.
3. PayloadMass vs. LaunchSite vs. Launch Outcome: A scatter plot showing the relationship between PayloadMass and LaunchSite, with the launch outcome (success or failure) indicated by color. This chart helps visualize if there's any correlation between the payload mass, launch site, and the success of the launch.
4. Success Rate vs. Orbit Type: A bar chart showing the success rate of each orbit type. This chart helps identify which orbit types have the highest success rates.
5. FlightNumber vs. Orbit Type vs. Launch Outcome: A scatter plot showing the relationship between FlightNumber and Orbit Type, with the launch outcome (success or failure) indicated by color. This chart helps visualize if there's any correlation between the flight number, orbit type, and the success of the launch.
6. PayloadMass vs. Orbit Type vs. Launch Outcome: A scatter plot showing the relationship between PayloadMass and Orbit Type, with the launch outcome (success or failure) indicated by color. This chart helps visualize if there's any correlation between the payload mass, orbit type, and the success of the launch.
7. Yearly Launch Success Trend: A line chart showing the trend of launch success rate over the years. This chart helps visualize how the success rate of launches has changed over time.

EDA with SQL

- Here are the SQL queries performed during the Exploratory Data Analysis:
- Query 1: Retrieve data on Falcon 9 launches from the database.
- Query 2: Calculate the success rate of Falcon 9 launches since 2013.
- Query 3: Determine the success rates of different launch sites (CCAFS LC-40, KSC LC-39A, VAFB SLC 4E).
- Query 4: Overlay the landing outcomes with launch site information to analyze success rates based on launch site and payload mass.
- Query 5: Identify attributes correlated with successful landings.

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- In the Plotly Dash dashboard, the following plots/graphs and interactions have been added:

1. Dropdown List for Launch Site Selection: Users can select a launch site from the dropdown list to filter the data accordingly.
2. Pie Chart Showing Total Successful Launches by Site: This pie chart displays the total count of successful launches for all sites or for a specific site selected from the dropdown list.
3. Slider for Payload Range Selection: Users can select a payload range using the slider to filter the data based on payload mass.
4. Scatter Chart Showing Correlation Between Payload and Launch Success: This scatter chart displays the correlation between payload mass and launch success. Users can see how different payload masses affect the success of launches, with options to filter data by launch site and payload range.

- These plots and interactions were added to provide users with an interactive dashboard that allows them to explore the SpaceX launch records visually and gain insights into the relationship between launch success, payload mass, and launch site.

Predictive Analysis (Classification)

- Here's how the classification model development process was conducted:

1. Data Preparation: The dataset was loaded and preprocessed to ensure it was in a suitable format for modeling.
2. Train-Test Split: The data was split into training and testing sets using the train_test_split function.
3. Model Training: Several classification algorithms were trained on the training data, including Logistic Regression, Support Vector Machines (SVM), Decision Tree Classifier, and K-nearest Neighbors (KNN).
4. Hyperparameter Tuning: Grid search was performed to find the best hyperparameters for each model, maximizing their performance.
5. Model Evaluation: The trained models were evaluated using the testing data to assess their performance metrics, such as accuracy.
6. Improvement: Based on the evaluation results, adjustments were made to the models or hyperparameters to improve their performance.
7. Finding the Best Model: The model with the highest accuracy or best performance metric was identified as the best performing classification model.
8. Output Confusion Matrix: Finally, a confusion matrix was generated to visualize the performance of the best performing model.

Predictive Analysis (Classification) - Flowchart

- Start
- --> Data Preparation
- --> Train-Test Split
- --> Model Training
- --> Hyperparameter Tuning
- --> Model Evaluation
- --> Improvement
- --> Finding the Best Model
- --> Output Confusion Matrix
- --> End

Results

- Exploratory Data Analysis Results:

1. Launch Success Factors: Identified key factors influencing launch success, such as payload mass, launch site, and historical trends.
2. Launch Site Analysis: Analyzed success rates across different launch sites, revealing variations in performance.
3. Payload Analysis: Explored the relationship between payload mass and launch success, identifying potential correlations.

- Interactive Analytics Demo in Screenshots:

1. Folium Geospatial Visualization: Screenshots showcasing interactive maps displaying launch sites and success rates.
2. Plotly Dash Dashboard: Screenshots illustrating interactive dashboards with dropdowns, sliders, and dynamic charts for exploration.

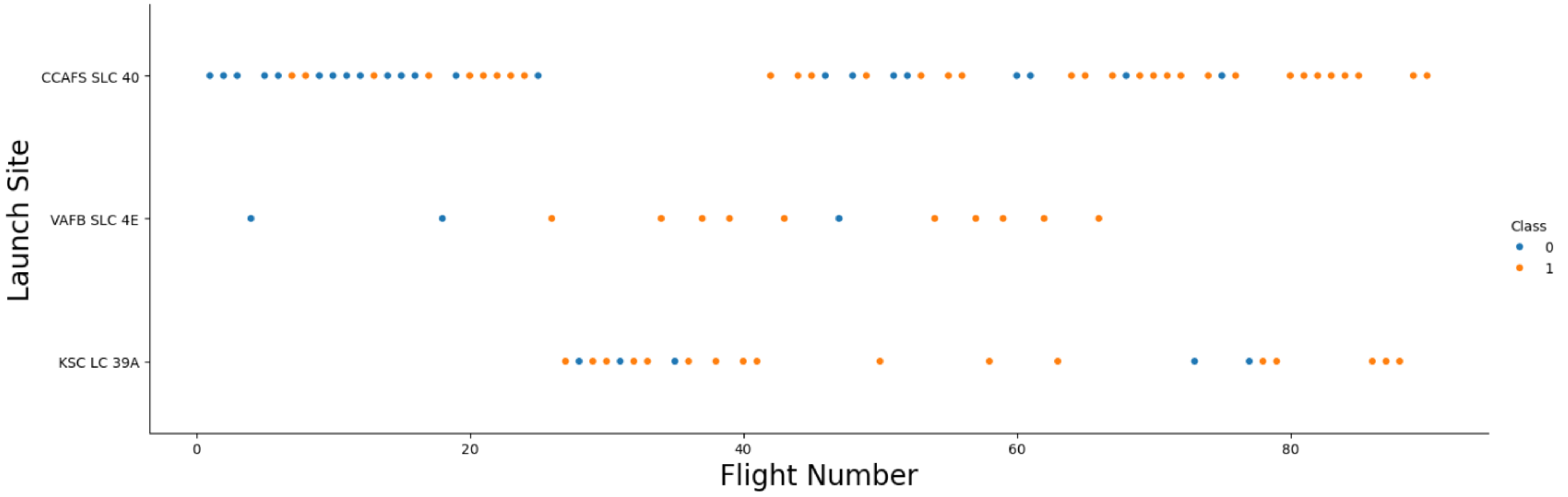
- Predictive Analysis Results:

1. Classification Models: Demonstrated predictive capabilities with classification models achieving reasonable accuracy in predicting launch success.
2. Evaluation Metrics: Results included metrics such as accuracy, precision, recall, and F1-score, providing insights into model performance.
3. Feature Importance: Identified significant features influencing launch success through feature importance analysis.
4. Prediction Visualization: Visualized model predictions and actual outcomes for validation and interpretation.
5. Potential Improvements: Highlighted areas for model refinement and enhancement to further improve predictive accuracy and performance.

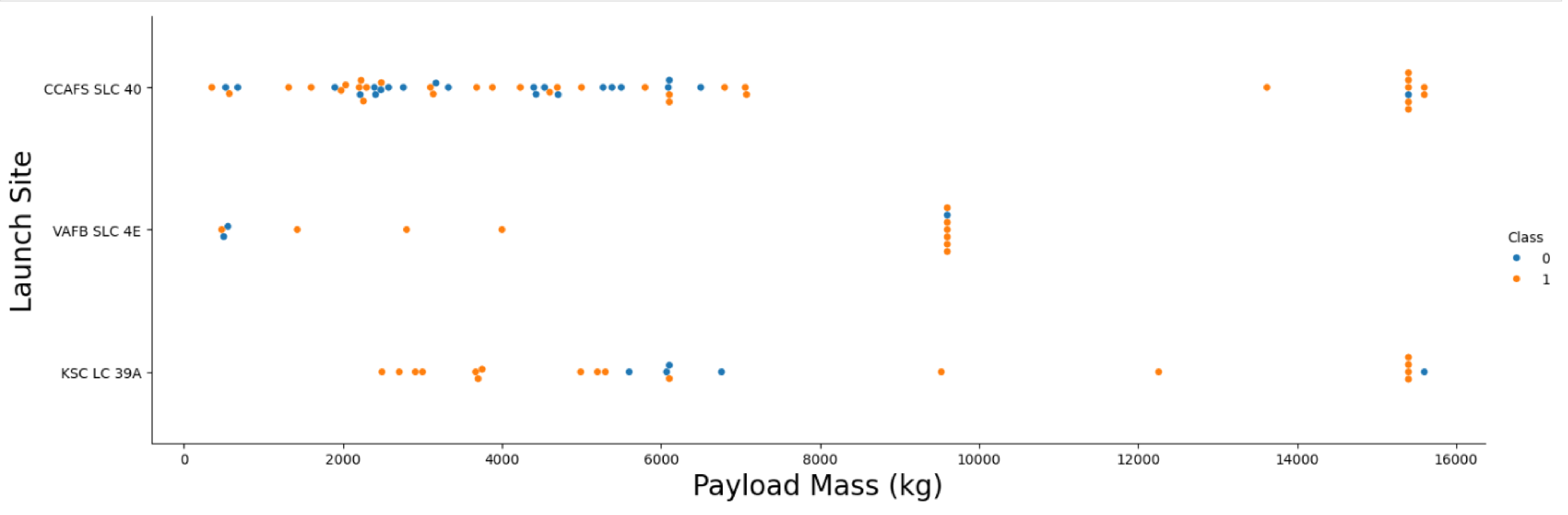
Section 2

Insights drawn from EDA

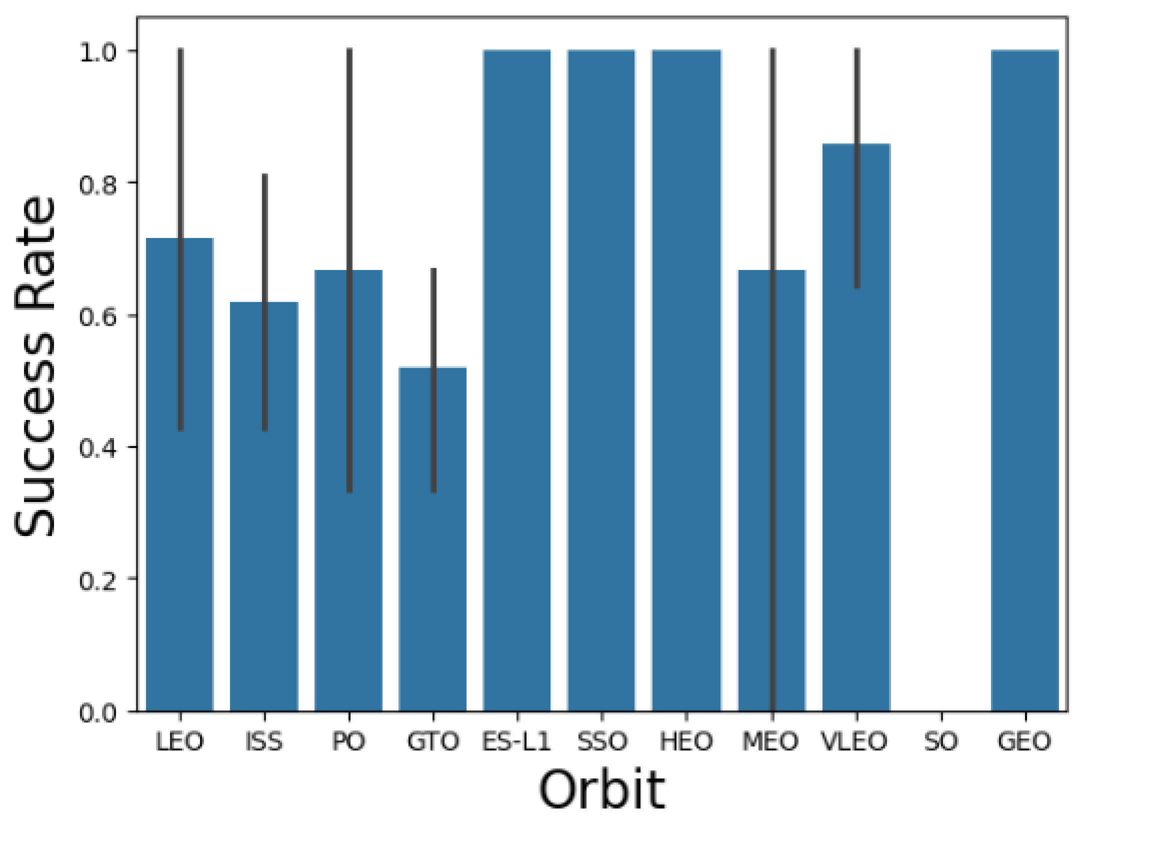
Flight Number vs. Launch Site



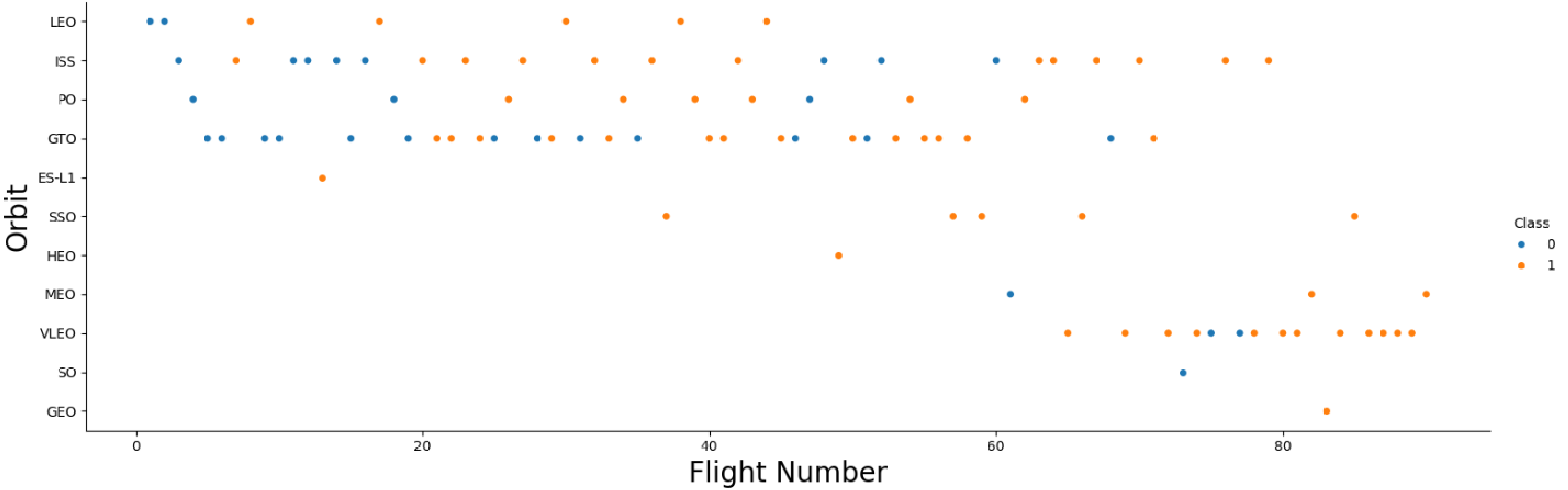
Payload vs. Launch Site



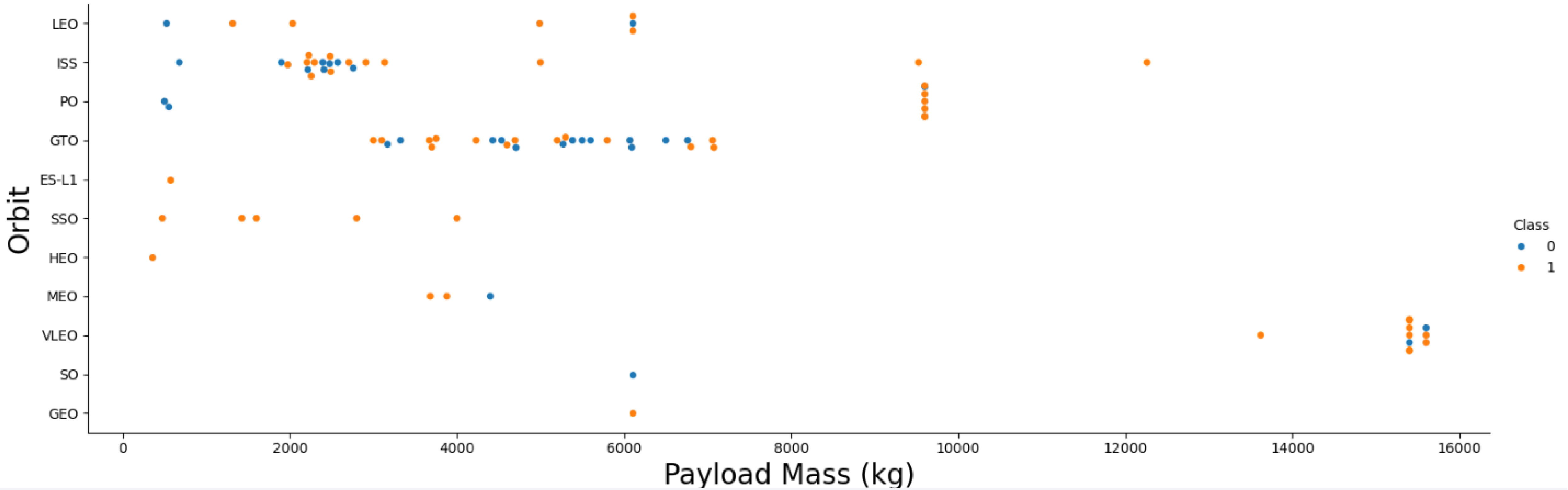
Success Rate vs. Orbit Type



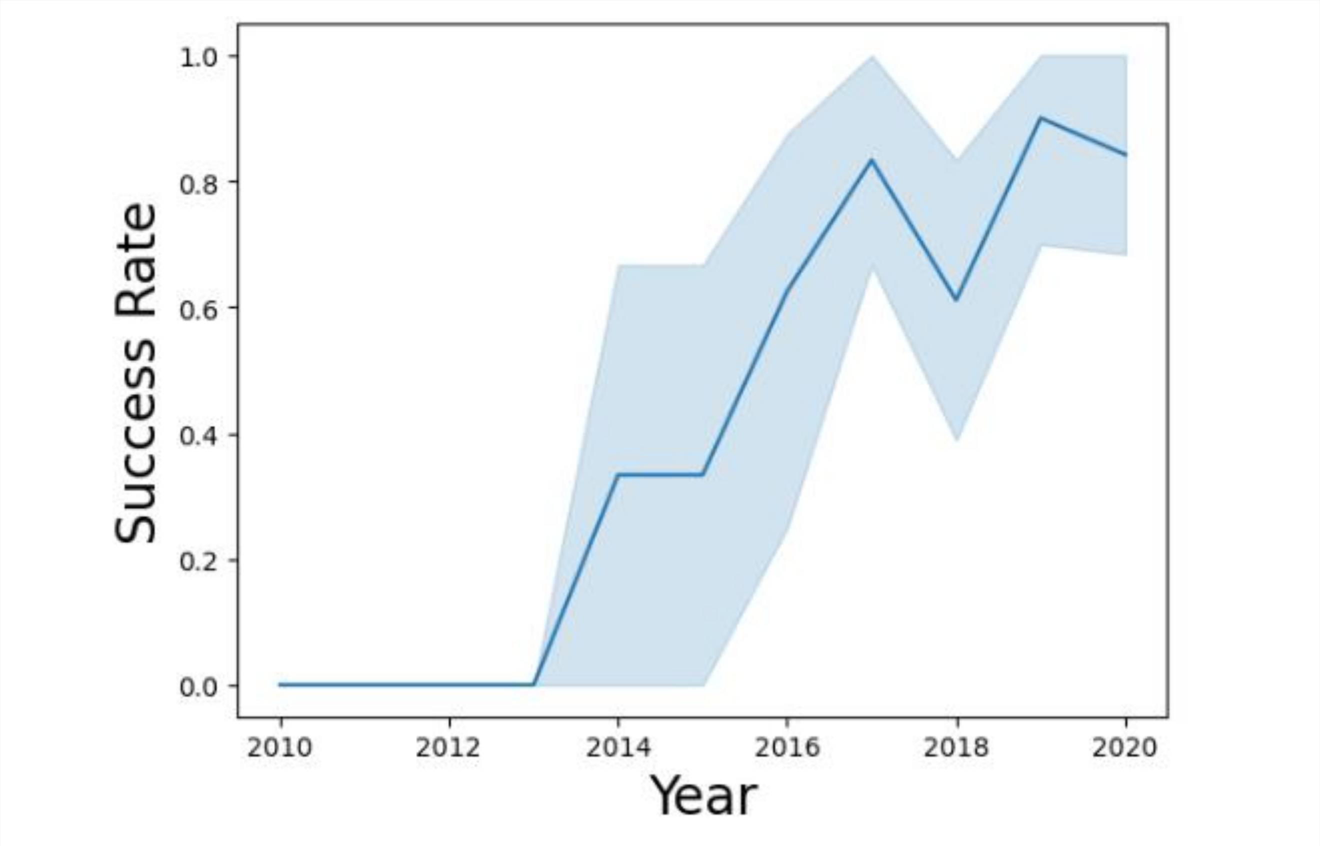
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

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

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

Launch Site Names Begin with 'CCA'

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

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

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

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

Total Payload Mass

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 SPACEXTABLE WHERE CUSTOMER='NASA (CRS)';
```

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

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

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

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

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

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

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_Ground_Pad FROM SPACEXTABLE WHERE Landing_Outcome='Success (ground pad)';
```

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

First_Successful_Landing_Ground_Pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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 DISTINCT BOOSTER_VERSION FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 A
```

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

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

Display average payload mass c

```
%sql SELECT AVG(PAYLOAD_MASS
```

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS Total FROM SPACEXTABLE GROUP BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

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

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

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

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

List the total number of success

```
%sql SELECT MISSION_OUTCOME,
```

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

Mission_Outcome	Tot
Failure (in flight)	
Success	
Success	
Success (payload status unclear)	

Snipping Tool
Screenshot copied to
Activate

2015 Launch Records

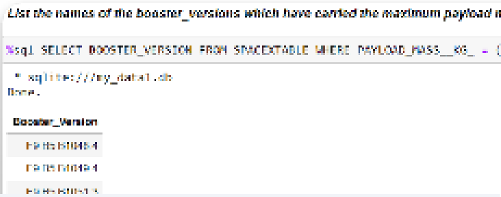
List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT SUBSTR(Date, 6, 2) AS Month, Landing_Outcome, BOOSTER_VERSION, Launch_Site FROM SPACEXTABLE WHERE SUBSTR(Date, 1, 4)
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site



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

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

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

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

Landing_Outcome	Count_Landing_Outcome
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

List the records which will display the mon 2015.

Note: SQLLite does not support monthnam

```
%sql SELECT SUBSTR(Date, 6, 2) AS Mon
```

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

Month	Landing_Outcome	Booster_Version	L

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

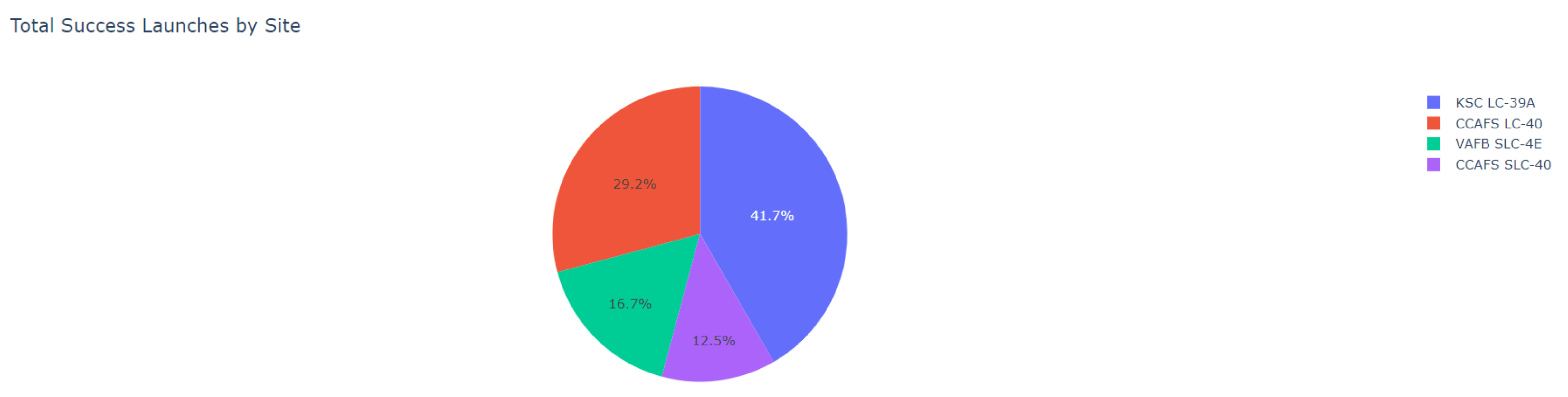
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

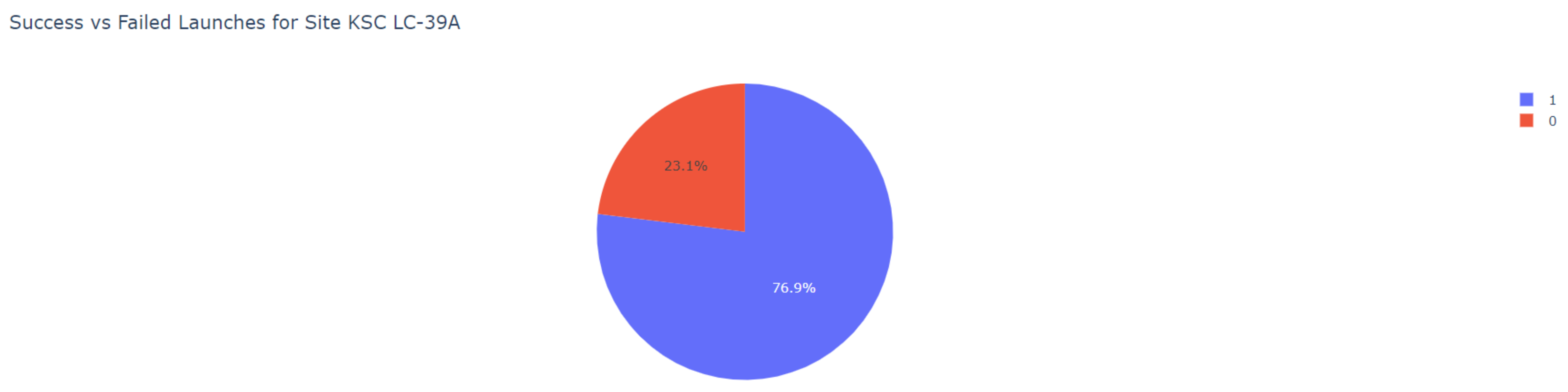
Section 4

Build a Dashboard with Plotly Dash

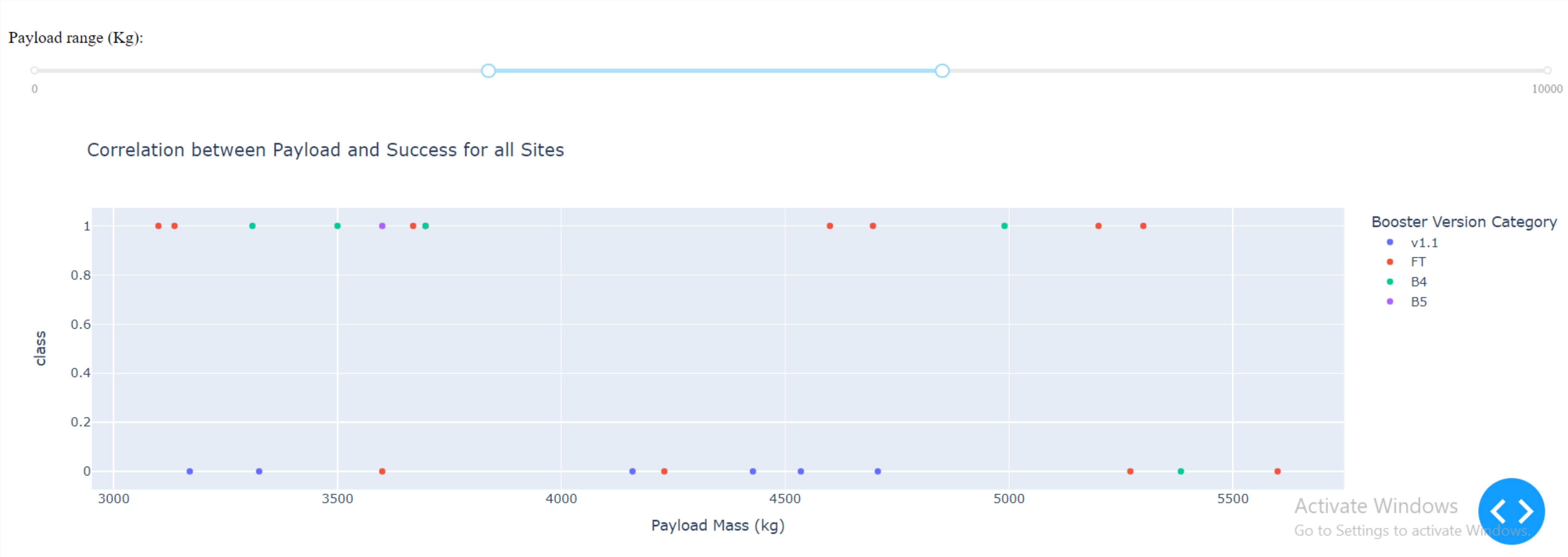
Total Success Launches for all sites



Success vs Fail for KSC LC-39A



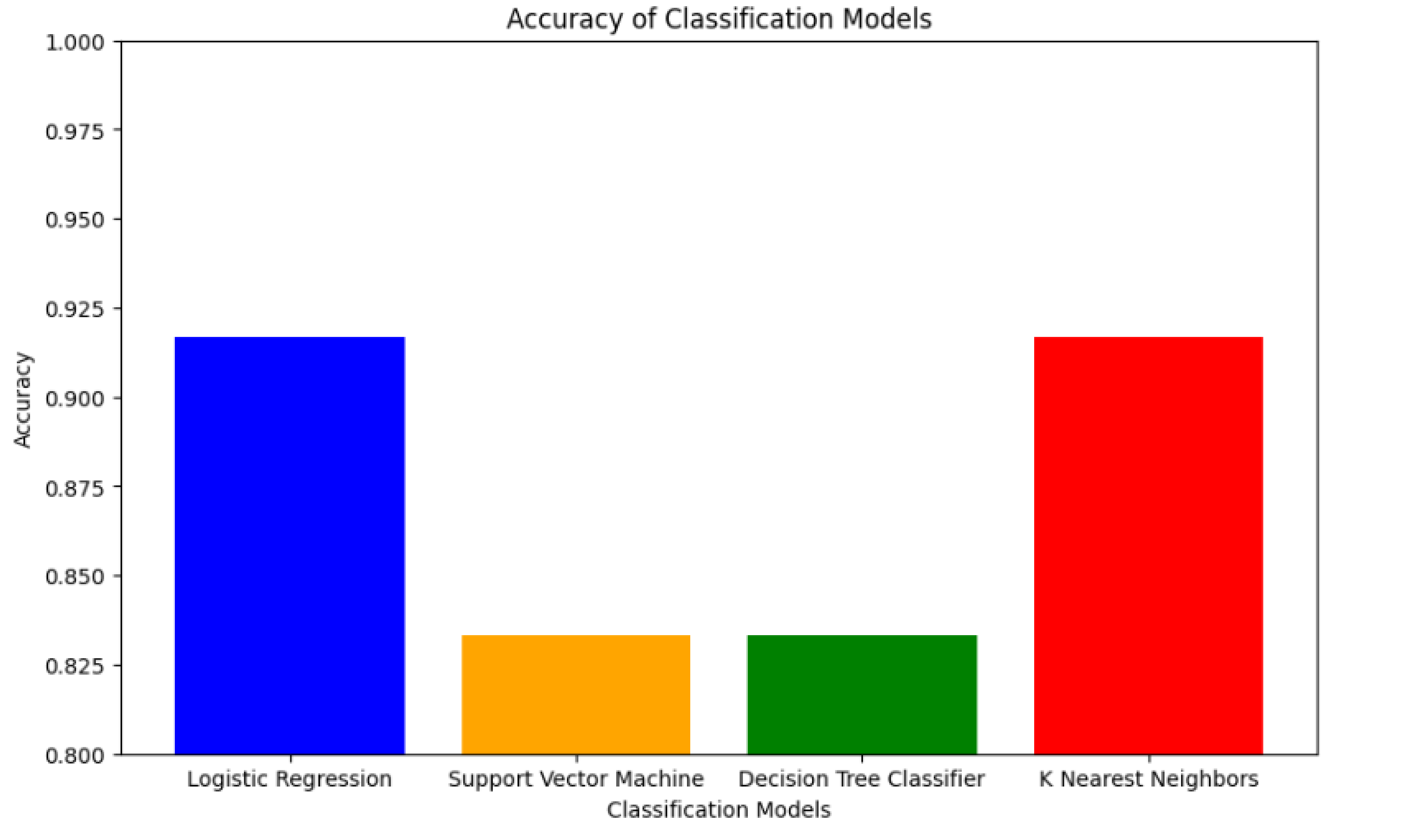
Payload vs Launch Outcome for all sites



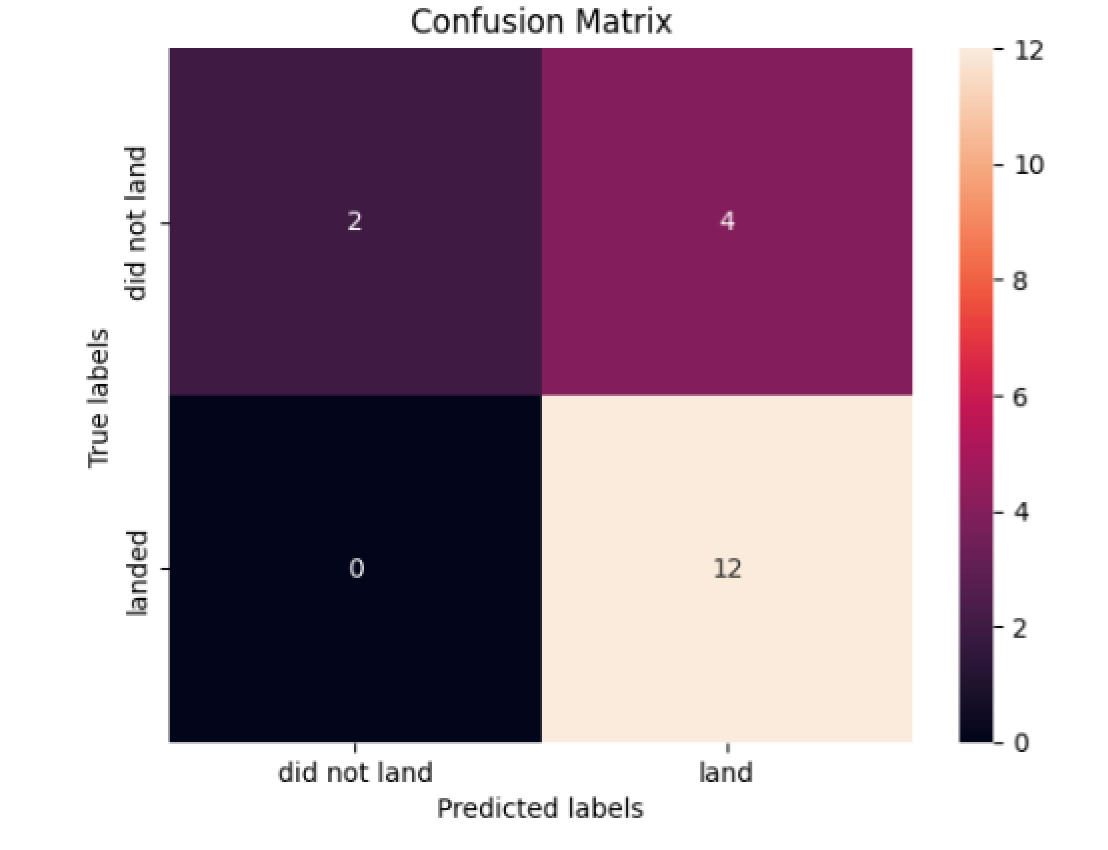
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- In conclusion, this project represents a comprehensive exploration of SpaceX Falcon 9 launch data, encompassing various stages from data collection to predictive analysis. Through meticulous data collection, wrangling, processing, and exploratory analysis, valuable insights were gained into the factors influencing launch success.
- The interactive visual analytics provided by tools like Folium and Plotly Dash enabled intuitive exploration and presentation of geospatial data and dashboard insights, enhancing understanding and communication of complex patterns and trends.
- Moreover, the development and evaluation of classification models for predicting launch success demonstrated the potential for leveraging historical data to inform future decision-making processes. The iterative process of model building, tuning, and evaluation underscored the importance of continuous improvement in refining predictive accuracy and enhancing decision support capabilities.
- Ultimately, this project serves as a testament to the power of data-driven insights in informing strategic decision-making processes, particularly in the context of space exploration and rocket launch operations.

Thank you!