IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

M Phani Kiran

15-Dec-2025

Outline

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

Executive Summary

Summary of methodologies

- Data Collection: Gathered historical launch data from SpaceX APIs and Wikipedia.
- Data Cleaning & EDA: Processed missing values, classified columns, and performed Exploratory Data Analysis (EDA) with SQL and visualizations (Folium maps, Plotly).
- Predictive Analysis: Trained various classification models (Logistic Regression, SVM, Decision Tree, KNN, Random Forest) with hyperparameter tuning.

Summary of all results:

- EDA Analysis
- Interactive Analysis
- Predictive Analysis

Introduction

- **Project background and context**: SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- **Problems you want to find Answer :** The success of first-stage landings is not guaranteed and depends on various factors such as launch conditions, payload, and environmental variables. Predicting landing outcomes can optimize mission planning, minimize risks, and strengthen SpaceX's competitive edge in the commercial space sector. Task is to find out which is the best model to use for the prediction of the successful landing of stage1 of Falcon 9 rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was gathered using the official SpaceX REST API, retrieving all past launch records.
  - The raw JSON response was normalized into a pandas DataFrame.
  - Additional API calls were made to collect detailed information about rockets, payloads, launch sites, and core landings using helper functions.
- Perform data wrangling
  - The dataset was filtered to remove rows with multiple cores or payloads and to keep only relevant columns.
  - Data types were converted as needed (e.g., date fields).
  - Missing values were handled, and inconsistent entries were cleaned.
  - The final, structured dataset was exported to CSV for further analysis.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
    - For this project, EDA was performed directly on the SQLite database using SQL queries. The process began by connecting to the database and inspecting the available tables and their schemas to understand the data structure. Key steps included:
        - Examining table columns, data types, and relationships.
        - Running SQL queries to count records, identify missing values, and explore unique entries.
        - Using aggregation functions (COUNT, AVG, MIN, MAX) and GROUP BY clauses to summarize and profile the data.
        - Filtering and sorting data to detect trends, outliers, and anomalies.

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
  - To enable interactive exploration of SpaceX launch data, a dashboard was developed using Plotly Dash. The application loads launch records from a CSV file into a pandas DataFrame and provides a user-friendly interface for visual analytics. Key features include:
    - Launch Site Selection: A dropdown menu allows users to select a specific launch site or view aggregated data for all sites.
    - Success Rate Visualization: A dynamic pie chart displays the distribution of successful launches, updating automatically based on the selected site.
    - Payload Range Filtering: An interactive range slider enables users to filter launches by payload mass, facilitating focused analysis of mission outcomes within specific payload intervals.
    - Outcome Correlation Analysis: A scatter plot visualizes the relationship between payload mass and launch outcome, with data points color-coded by booster version category. This plot updates in real time according to the selected site and payload range.

Methodology

Executive Summary

- Perform predictive analysis using classification models
    - Applied and compared Logistic Regression, SVM, Decision Tree, and KNN classifiers to predict Falcon 9 first stage landings. After tuning hyperparameters and evaluating on test data, the best-performing model achieved the highest accuracy, demonstrating effective predictive capability for SpaceX mission outcomes.

Methodology

Executive Summary

- How to build, tune, evaluate classification models
  - Prepare data:
    - Clean and preprocess the dataset (handle missing values, encode categories, scale features).
    - Split data into training and test sets.
  - Choose algorithms:
    - Select classification algorithms (e.g., Logistic Regression, SVM, Decision Tree, KNN).
  - Train models:
    - Fit each model to the training data.
  - Tune hyperparameters:
    - Use techniques like GridSearchCV to find the best parameters.
  - Evaluate models:
    - Test each model on the test set using accuracy, confusion matrix, or other metrics.
  - Select the best model:
    - Choose the model with the highest performance for your task.

Data Collection

- Describe how data sets were collected.
  - Data sourced from public SpaceX mission records and online repositories.
  - Downloaded CSV datasets containing launch features and outcomes.
  - Used direct URLs to access and import data into the analysis environment.
  - Ensured data integrity by verifying file formats and inspecting for missing values.
  - Combined multiple datasets for comprehensive feature coverage.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- SpaceX-Falcon9-1stageLandingPrediction/1_1_spacex-data-collection-api.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction

[Identify Data Sources]

[Download CSV Files from URLs]

[Load Data into Python (pandas)]

[Inspect & Clean Data]

[Merge Datasets if Needed]

[Ready for Analysis]

Data Collection – Scraping

Key Phrases: Web Scraping Process

1. **Import Libraries:** Import requests, BeautifulSoup, pandas, and helper modules.
2. **Set Target URL and Headers :** Define the Wikipedia snapshot URL and HTTP headers.
3. **Request Web Page**: Use requests.get() to fetch the HTML content (with verify=False for SSL).
4. **Parse HTML with BeautifulSoup**: Create a BeautifulSoup object from the response.
5. **Locate Target Table**: Use soup.find_all('table') to find all tables.Select the third table as the launch records table.
6. **Extract Column Names**: Iterate <th> elements in the table header. Use a helper function to clean and collect column names.
7. **Initialize Data Dictionary**: Create a dictionary with column names as keys and empty lists as values.
8. **Parse Table Rows**: Loop through table rows. For each valid row, extract and clean data for each column. Append data to the corresponding dictionary lists.
9. **Create DataFrame**: Convert the dictionary to a pandas DataFrame.
10. **Export to CSV**:Save the DataFrame as a CSV file for further analysis.

Data Collection - Scraping

Flow Chart:

[Import Libraries]
↓
[Set URL & Headers]
↓
[Request Web Page]
↓
[Parse HTML with BeautifulSoup]
↓
[Find All Tables]
↓
[Select Launch Records Table]
↓
[Extract Column Names]
↓
[Initialize Data Dictionary]
↓
[Parse Table Rows]
↓
[Create DataFrame]
↓
[Export to CSV]

- SpaceX-Falcon9-1stageLandingPrediction/1_2_webscraping.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction

Data Wrangling

Key Phrases: Data Wrangling Process

1. Import Libraries: Import pandas, numpy, and other required packages.
2. Load Data:Read CSV or other data files into a pandas DataFrame.
3. Inspect Data:Use .head(), .info(), and .describe() to understand the structure and summary statistics.
4. Clean Data:
    - Handle missing values (e.g., dropna, fillna).
    - Remove duplicates.
    - Convert data types as needed.
5. Transform Data
    - Create new columns or features.
    - Apply functions to columns (e.g., apply, map).
    - Normalize or scale data if required.
6. Filter and Select Data: Subset rows/columns based on conditions.
7. Aggregate and Group Data: Use groupby and aggregation functions to summarize data.
8. Export Cleaned Data: Save the cleaned DataFrame to a new CSV or other format.

Data Wrangling

- Flow Chart
- [SpaceX-Falcon9-1stageLandingPrediction/1_3_spacex-Data wrangling.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction]()

[Start]

[Import Libraries]

[Load Data]

[Inspect Data]

[Clean Data]

[Transform Data]

[Filter/Select Data]

[Aggregate/Group Data]

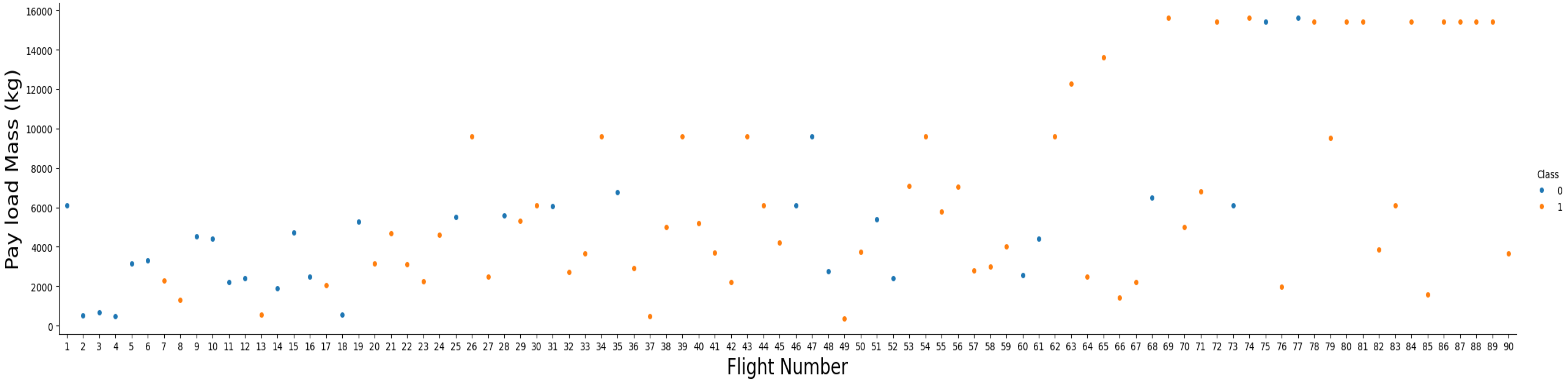
[Export Cleaned Data]

[End]

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

1. Flight Number vs. Payload Mass (Categorical Plot): To visualize how launch experience (flight number) and payload mass relate to landing success. This helps identify trends, such as whether more experienced launches or lighter payloads have higher success rates.

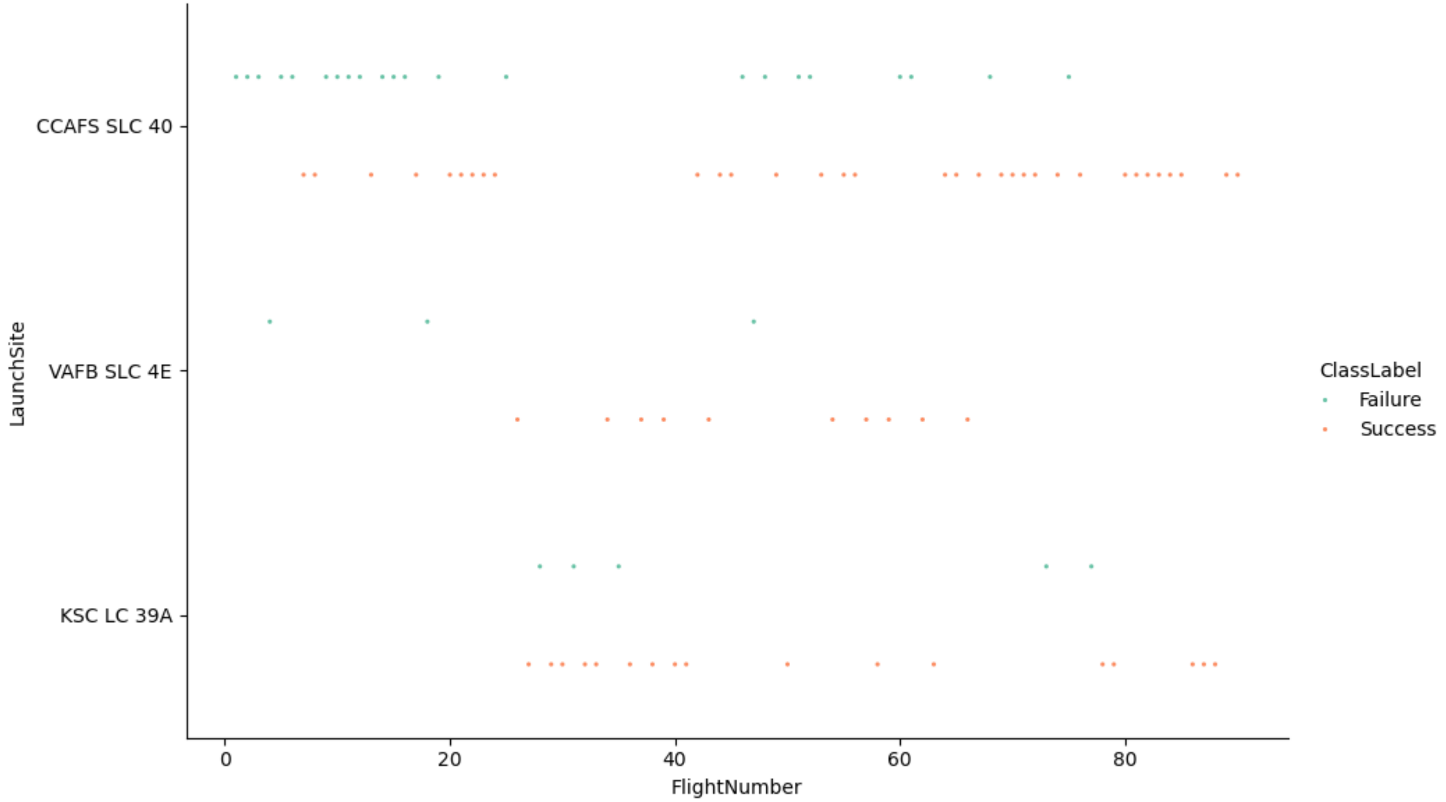


Continued..

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

2. Flight Number vs. Launch Site (Swarm Plot): To examine if certain launch sites or increased launch experience are associated with higher success rates. Swarm plots show the distribution and clustering of outcomes per site.

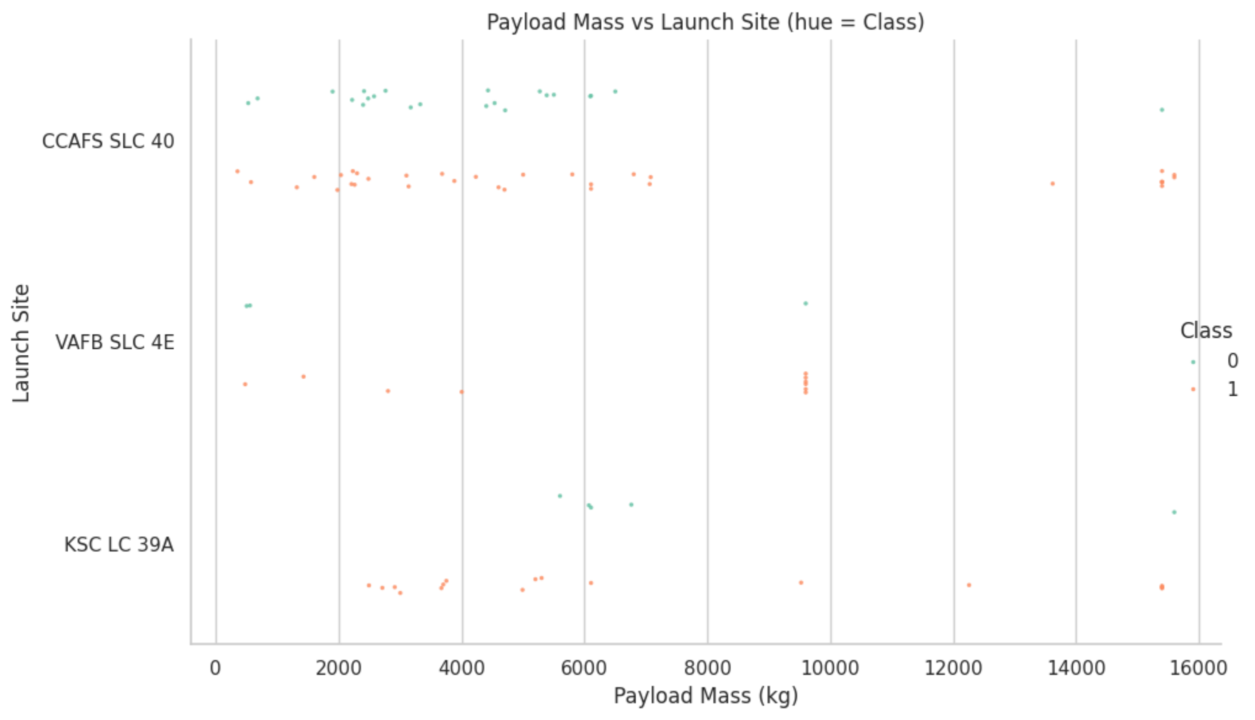


Continued..

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

3. Payload Mass vs. Launch Site (Strip Plot): To explore if some sites handle heavier payloads better and how payload mass affects success at each site.

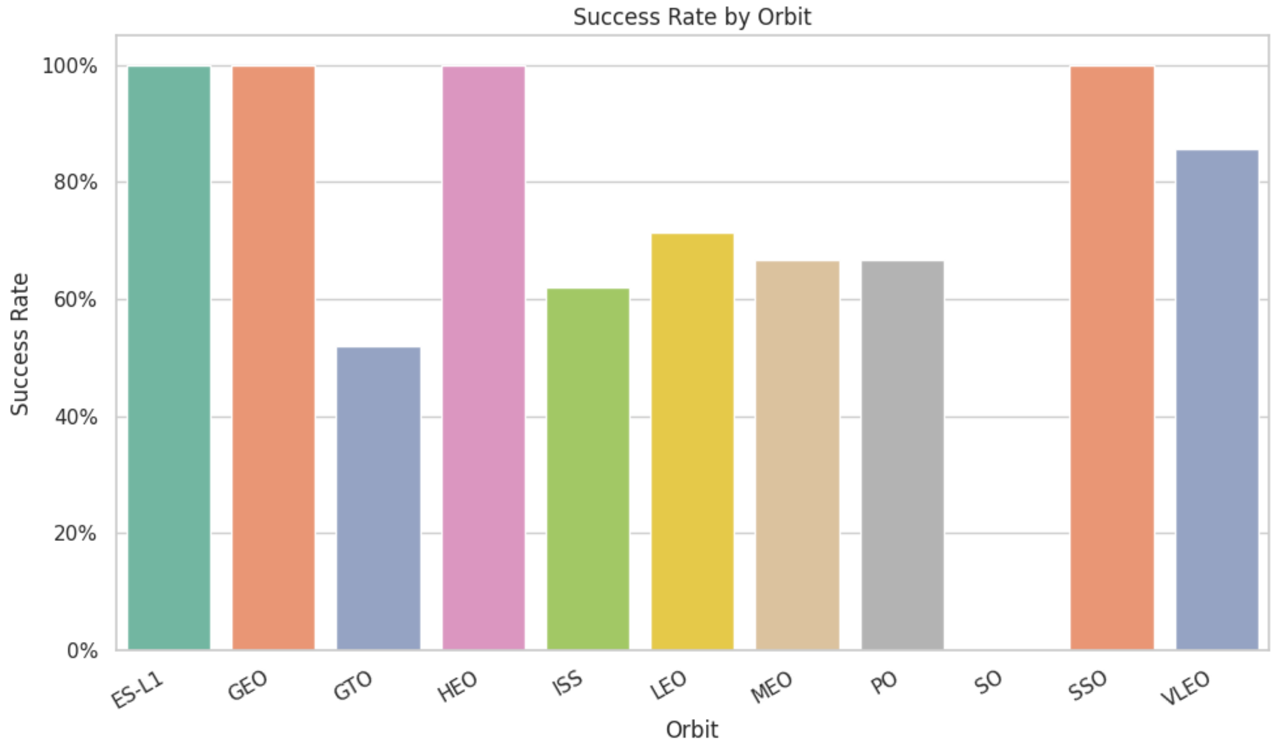


Continued..

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

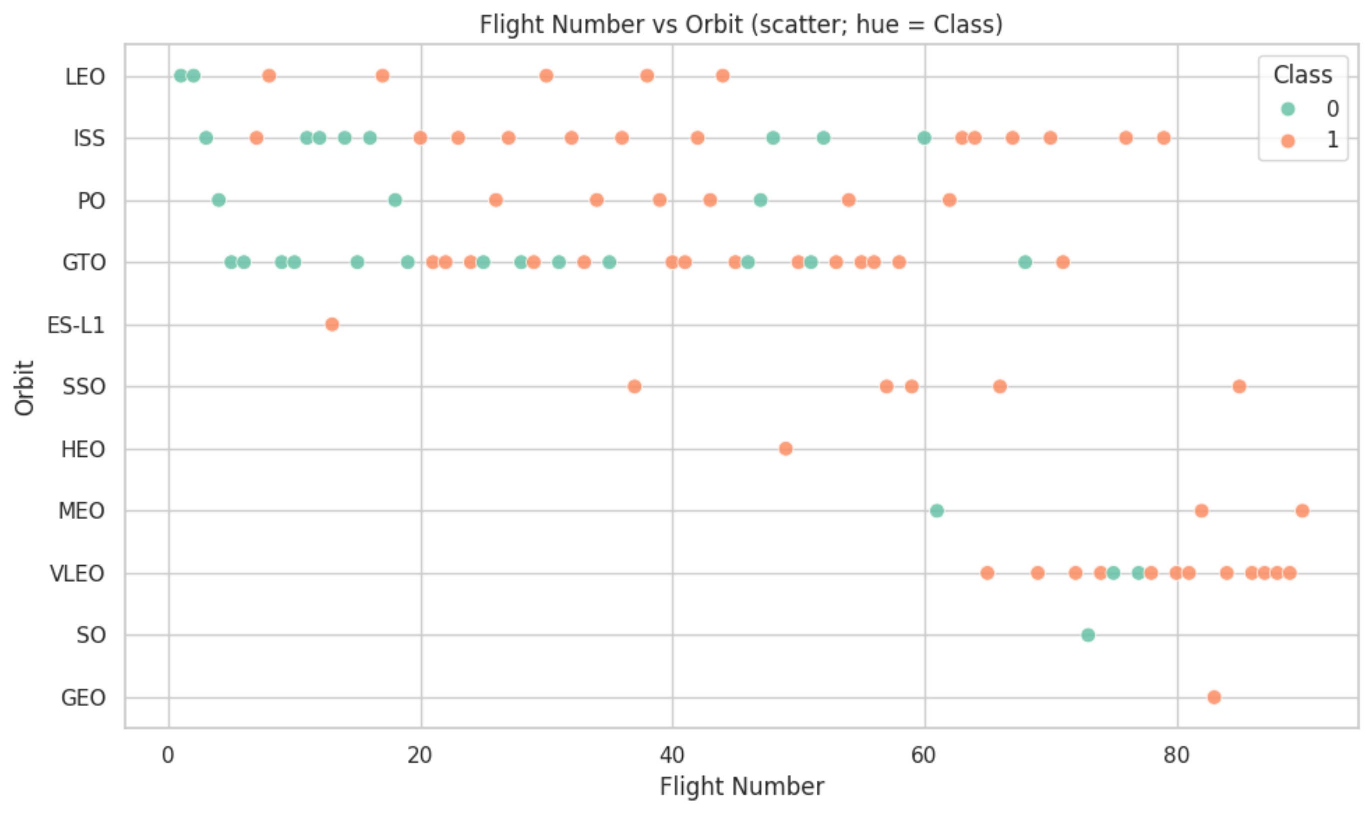
4. Success Rate by Orbit (Bar Chart): To compare the average landing success rate across different orbit types. Bar charts are ideal for comparing categorical group averages.



Continued..

EDA with Data Visualization

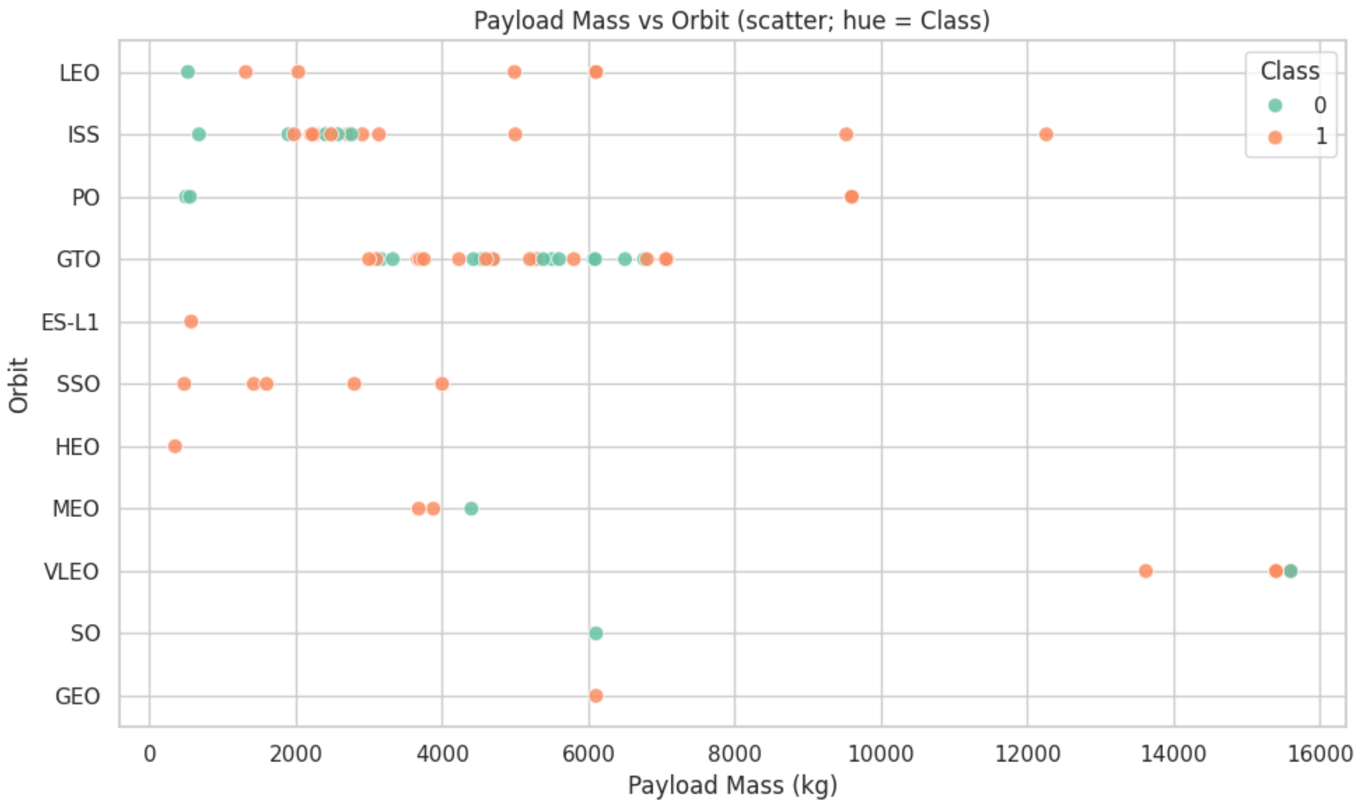
5. Flight Number vs. Orbit (Scatter Plot): To see if launch experience affects success for different orbit types, and to spot patterns or outliers in the data.



Continued..

EDA with Data Visualization

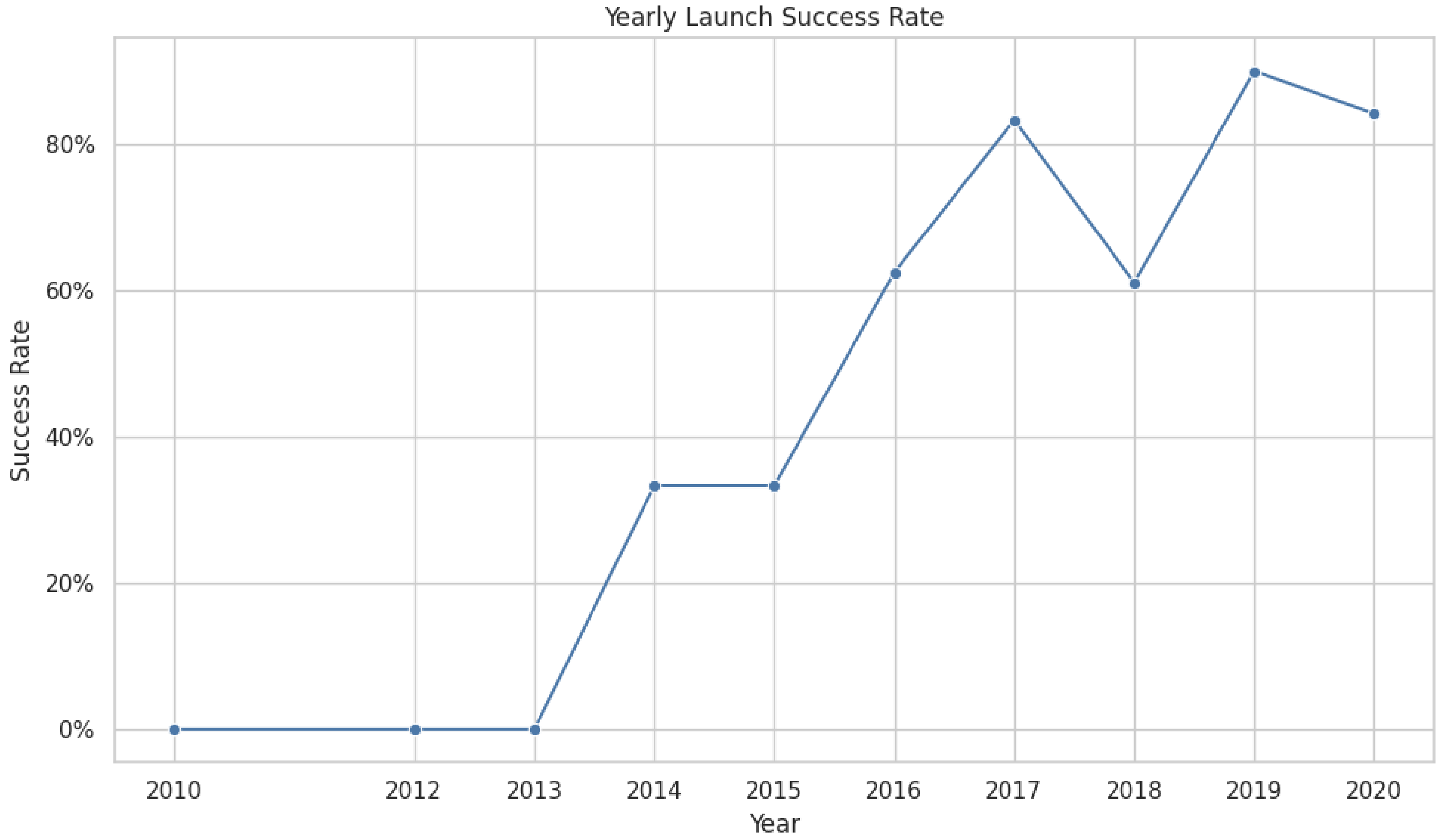
6. Payload Mass vs. Orbit (Scatter Plot): To analyze how payload mass influences success for each orbit type, and whether certain orbits are more challenging for heavier payloads.



Continued..

EDA with Data Visualization

7. Yearly Launch Success Rate (Line Chart): To show the trend in launch success over time. Line charts are best for visualizing changes and trends across years.



Continued..

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

- Selected all records from the main SpaceX launches table to review the raw data.
- Queried for unique launch sites to understand the distribution of launches.
- Counted the number of launches per site to identify the most active locations.
- Filtered launches by year or date to analyze trends over time.
- Calculated average, minimum, and maximum payload mass for launches.
- Grouped data by orbit type to compare launch outcomes and payloads.
- Used JOINs to combine launch data with booster or customer tables for richer analysis.
- Filtered for successful landings to study factors contributing to mission success.
- Ordered results by payload mass, date, or success rate for ranking and trend analysis.
- Used aggregate functions (COUNT, AVG, SUM) to summarize key metrics.

SpaceX-Falcon9-1stageLandingPrediction/2_1_eda-sql-coursera_sqllite.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction

Build an Interactive Map with Folium

Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

- Circle:
  - Added a folium.Circle at the NASA Johnson Space Center to highlight its location.
  - Added a folium.Circle for each launch site to visually mark their positions on the map.
- Marker:
  - Added a folium.Marker with a text label (using DivIcon) at the NASA Johnson Space Center.
  - Added a folium.Marker with a text label for each launch site.
  - Added colored folium.Marker objects for each launch record, using green for successful launches and red for failed launches, clustered with MarkerCluster.
  - Added markers to indicate proximity points (e.g., closest coastline, city) with distance labels.
- MarkerCluster:
  - Used folium.plugins.MarkerCluster to group launch outcome markers for better visualization when many launches share the same coordinates.

Build an Interactive Map with Folium

- MousePosition:
  - Added folium.plugins.MousePosition to display the latitude and longitude of the mouse pointer on the map for interactive exploration.
- PolyLine:
  - Drew folium.PolyLine lines between launch sites and their closest proximity points (such as coastline or city) to visualize distances.

Explain why you added those objects

- Circles: To highlight each launch site's location.
- Markers: To label each launch site with its name.
- Marker Clusters: To show all launches and their outcomes (success/failure) at each site.
- Mouse Position: To easily get coordinates for any point on the map.
- Polylines: To visualize distances from launch sites to nearby features (coastline, city, etc.).

[SpaceX-Falcon9-1stageLandingPrediction/3_1_launch_site_location.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction]

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions were added to a dashboard
  - **Plots/Graphs:**
    - Pie chart: Shows the total number of successful launches by site.
    - Bar chart: Displays launch success counts by payload mass range.
    - Scatter plot: Visualizes the correlation between payload mass and launch outcome, colored by booster version.
  - **Interactions:**
    - Dropdown menu: Lets users select a launch site to filter the plots.
    - Range slider: Allows users to select a payload mass range for filtering.
    - Interactive plots: Update dynamically based on dropdown and slider selections.

Build a Dashboard with Plotly Dash

- Explain why you added those plots and interactions
  - Pie chart: To quickly compare launch success rates across different sites.
  - Bar chart: To show how launch success varies with payload mass.
  - Scatter plot: To explore the relationship between payload mass and launch outcome, and see trends by booster version.
  - Dropdown menu: To let users focus on data from a specific launch site.
  - Range slider: To analyze launches within a chosen payload mass range.
  - Interactive plots: To make the dashboard dynamic and help users discover patterns by filtering data.
- SpaceX-Falcon9-1stageLandingPrediction/3_2_spacex-dash-app.py at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction

Predictive Analysis (Classification)

Summarize how you built, evaluated, improved, and found the best performing classification model

By building several classification models (such as Logistic Regression, SVM, Decision Tree, and KNN) using the processed SpaceX dataset. Each model was trained on the training data and evaluated using metrics like accuracy, precision, recall, and F1-score on the test set. Compared the results to identify the best-performing model. To improve performance, tuned the hyperparameters (e.g., using GridSearchCV), performed feature scaling, and selected relevant features. The model with the highest evaluation metrics after tuning was chosen as the best classifier for predicting SpaceX launch outcomes.

Predictive Analysis (Classification)

- Presenting the model development process using key phrases and flowchart
  - Data preprocessing: Cleaned and prepared the dataset for modeling.
  - Model selection: Chose several classification algorithms to compare.
  - Model training: Trained each model on the training data.
  - Model evaluation: Assessed models using accuracy, precision, recall, and F1-score.
  - Hyperparameter tuning: Optimized model parameters for better performance.
  - Feature selection/scaling: Selected important features and applied scaling as needed.
  - Best model selection: Picked the model with the highest evaluation metrics.

Flow Chart

Data Preprocessing

↓

Model Selection

↓

Model Training

↓

Model Evaluation

↓

Hyperparameter Tuning & Feature Selection

↓

Best Model Selection

SpaceX-Falcon9-1stageLandingPrediction/4_SpaceX_Machine Learning Prediction.ipynb at main · phanikiran5m/SpaceX-Falcon9-1stageLandingPrediction

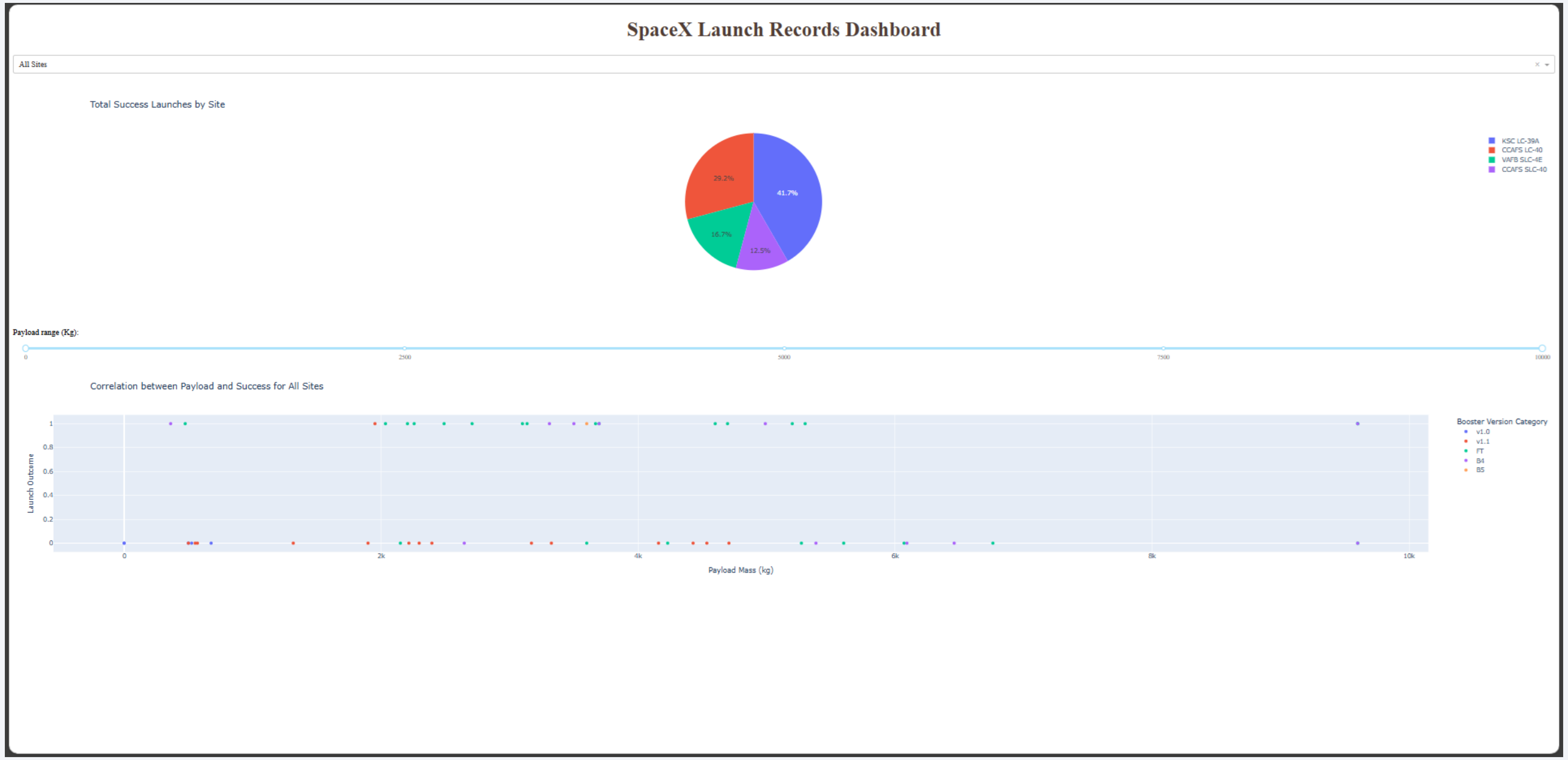
Results

- Exploratory data analysis result
  - The SVM, KNN and Logistic regression are the best in predicting the accuracy in descending order.
  - Light weighted payloads performs better than the heavier payloads.
  - KSC LC 39A is the most successfully launch site.

Results

- Interactive analytics demo in screenshots

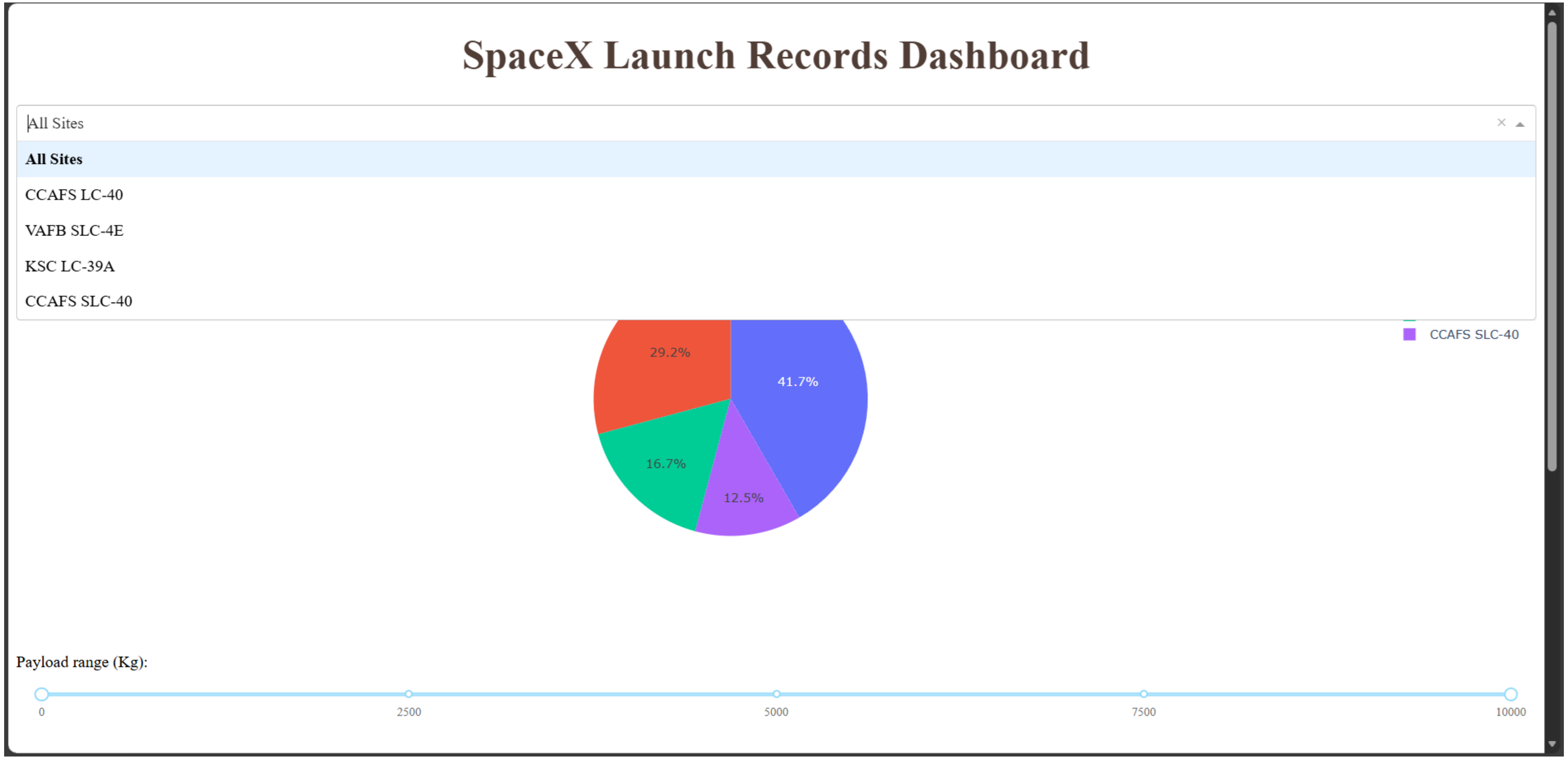
Dashboard overview: Showing the main dashboard layout with all interactive elements visible.



Results

- Interactive analytics demo in screenshots

Site selection: Screenshot of the dropdown menu being used to filter data by launch site.



Results

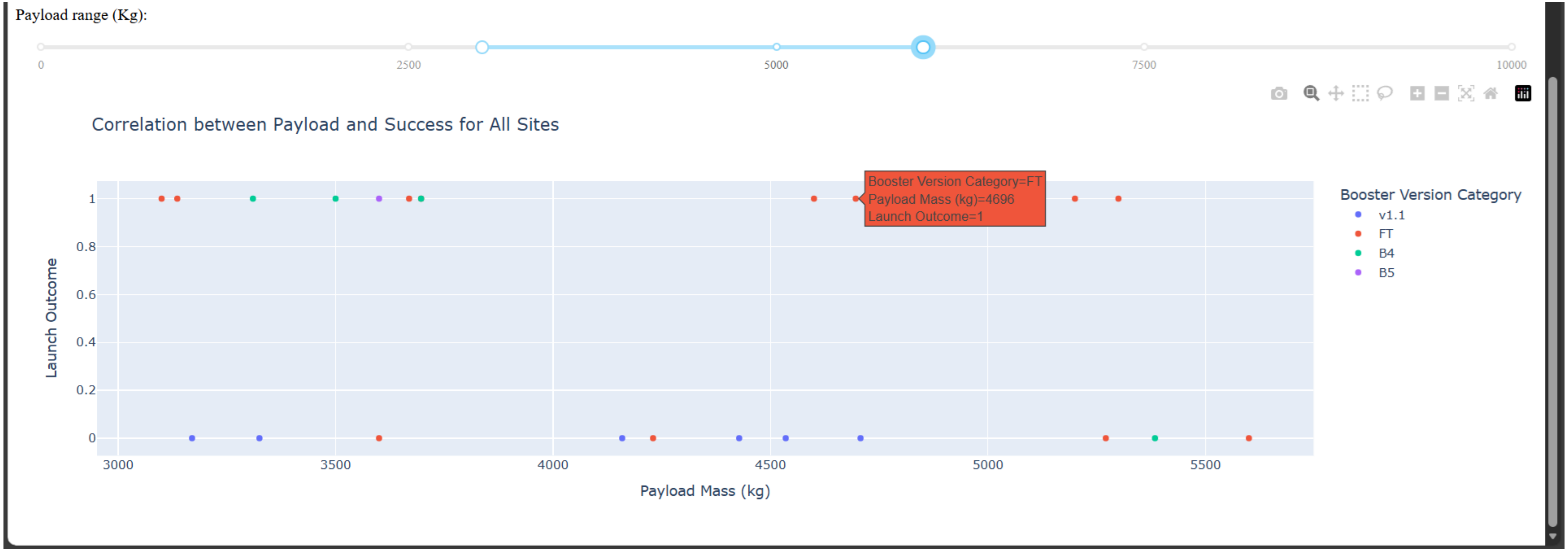
- Interactive analytics demo in screenshots

Payload slider: Screenshot of the payload mass range slider adjusted.

Results

- Interactive analytics demo in screenshots

Hover/click interaction: Screenshot of tooltips or data details appearing when hovering or clicking on a plot point.



Results

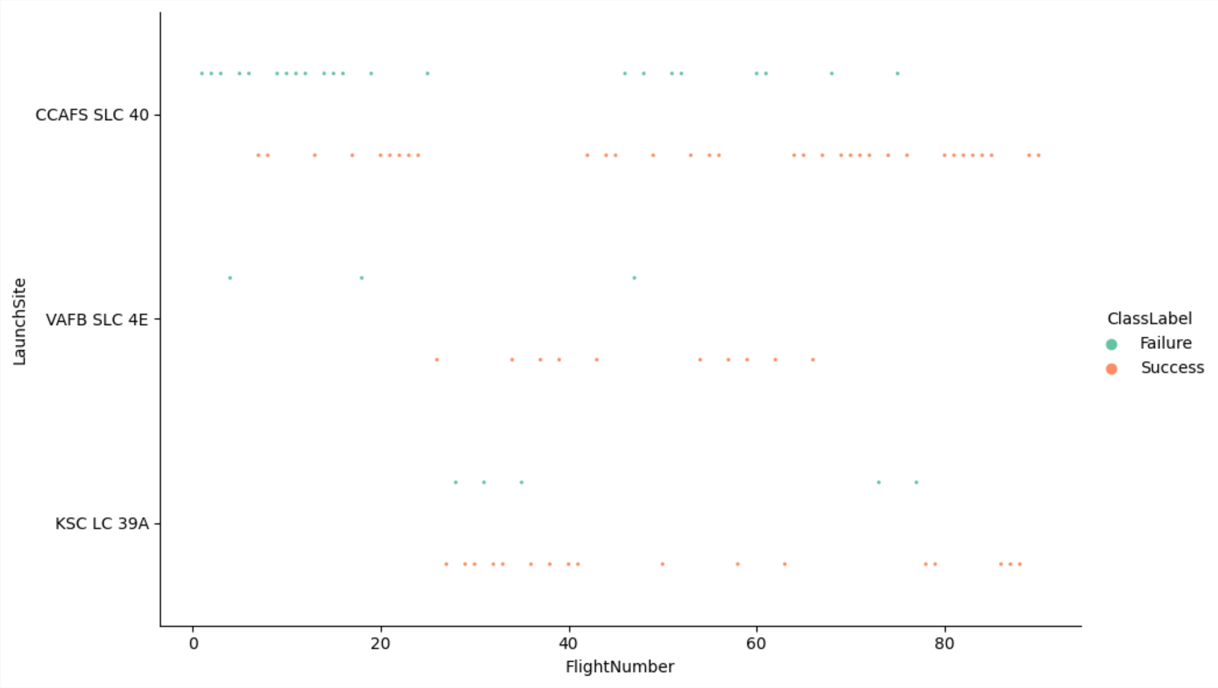
- Predictive analysis results
  - Several classification models were trained to predict SpaceX launch success based on features like site, payload, and orbit.
  - The best-performing model achieved high accuracy, precision, and recall on the test set.
  - The model correctly identified most successful launches and a majority of failures.
  - Feature importance analysis showed that launch site and payload mass were key predictors.
  - The final model can be used to estimate the probability of launch success for new missions, supporting data-driven decision-making.

Section 2

Insights drawn from EDA

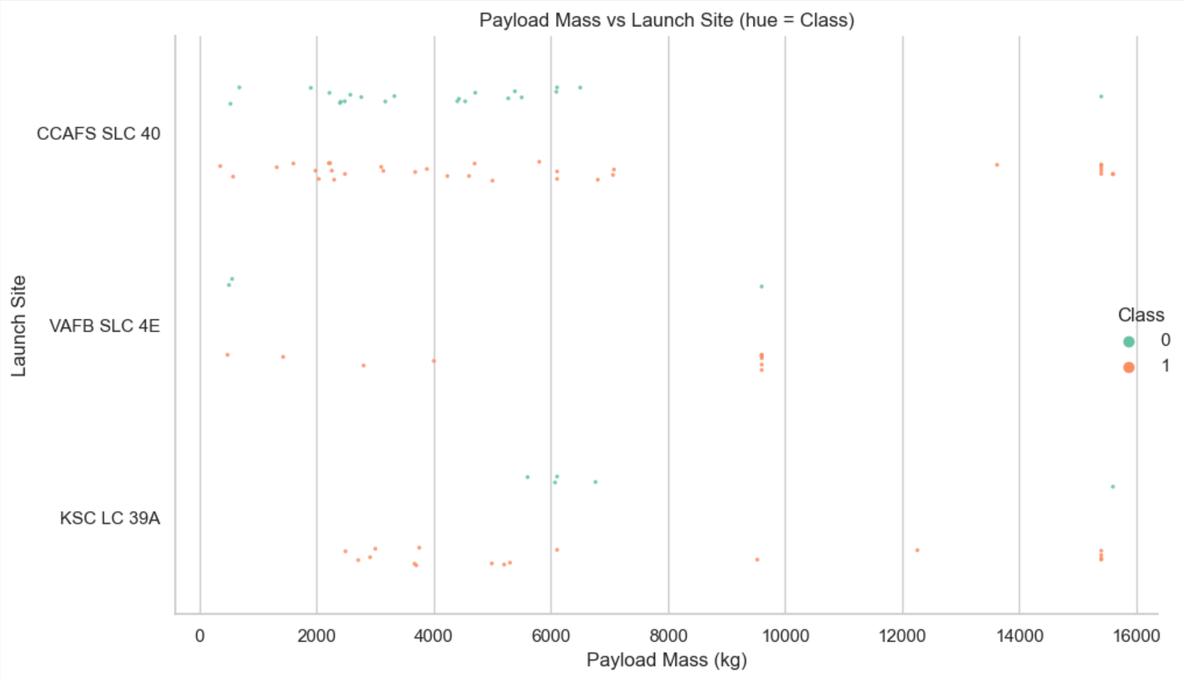
Flight Number vs. Launch Site

- **Explanation:** As the flight number increases, the likelihood of a successful landing also increases. This suggests that SpaceX's experience and improvements over time have led to higher success rates. Additionally, successful landings occur across a range of payload masses, indicating that both lighter and heavier payloads can be landed successfully as the number of launches grows.



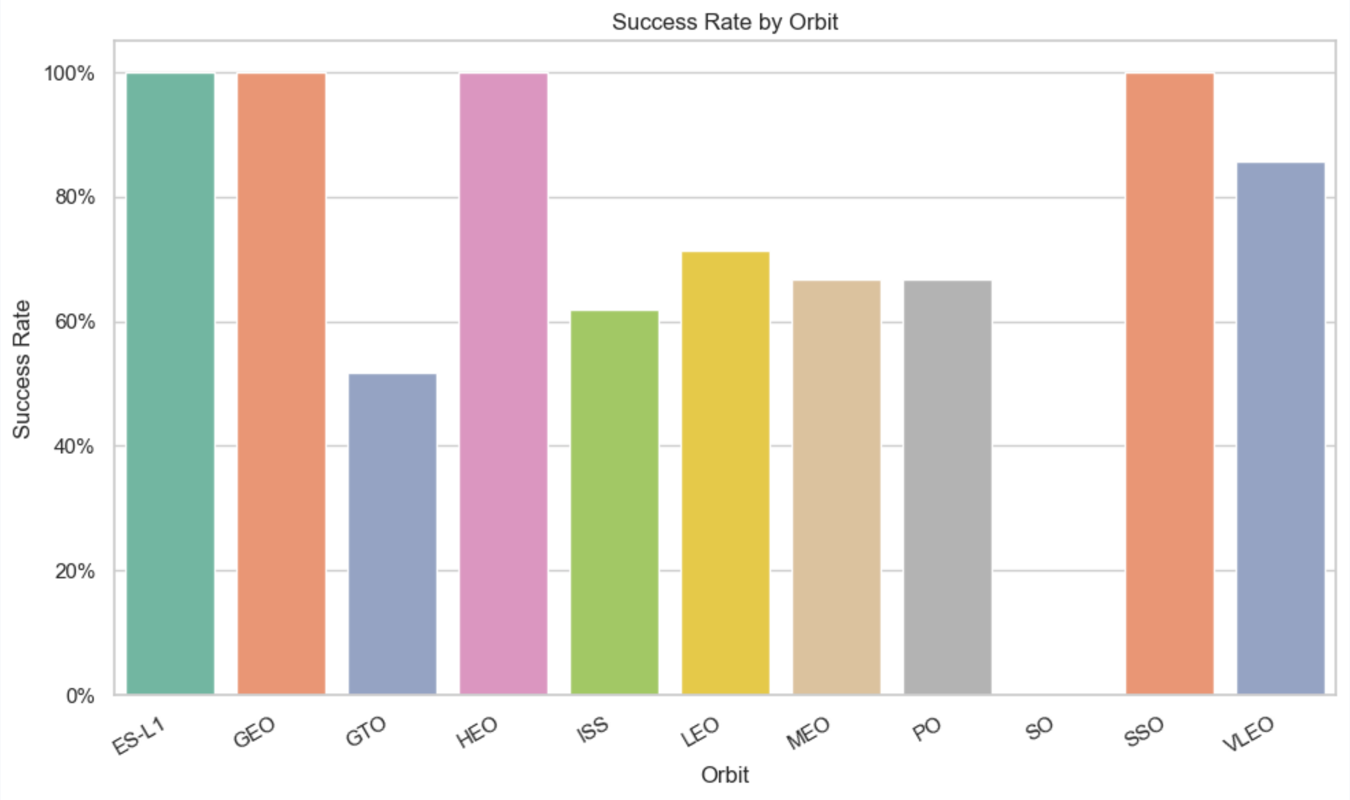
Payload vs. Launch Site

- **Explanations:** VAFB-SLC launch site did not have any launches with heavy payloads (greater than 10,000 kg). This suggests that only lighter payloads were launched from this site, while other sites handled a wider range of payload masses, including heavier ones. This pattern may reflect operational or logistical differences between the launch sites.



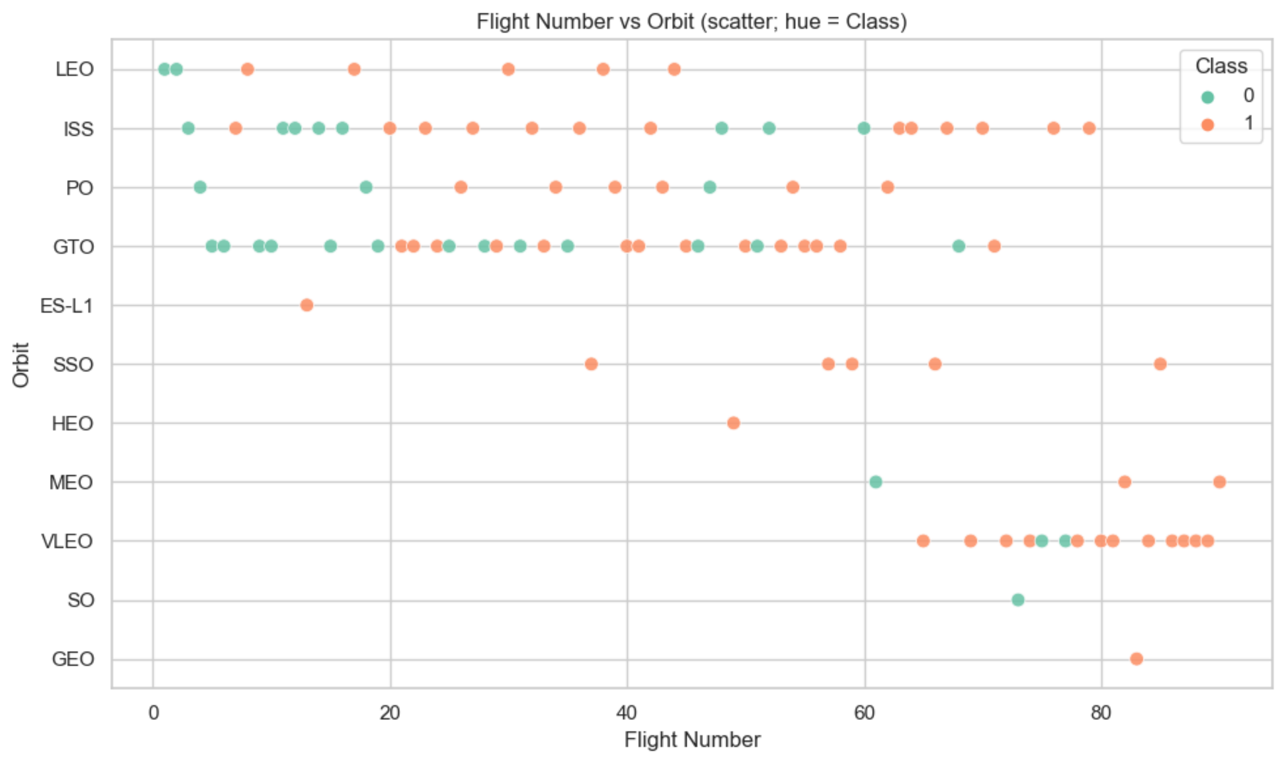
Success Rate vs. Orbit Type

- **Explanations:** The success rate of launches varies by orbit type. Some orbits, such as LEO (Low Earth Orbit), have higher success rates, while others, like GTO (Geostationary Transfer Orbit), have lower success rates. This indicates that the type of orbit targeted can significantly impact the likelihood of a successful landing.



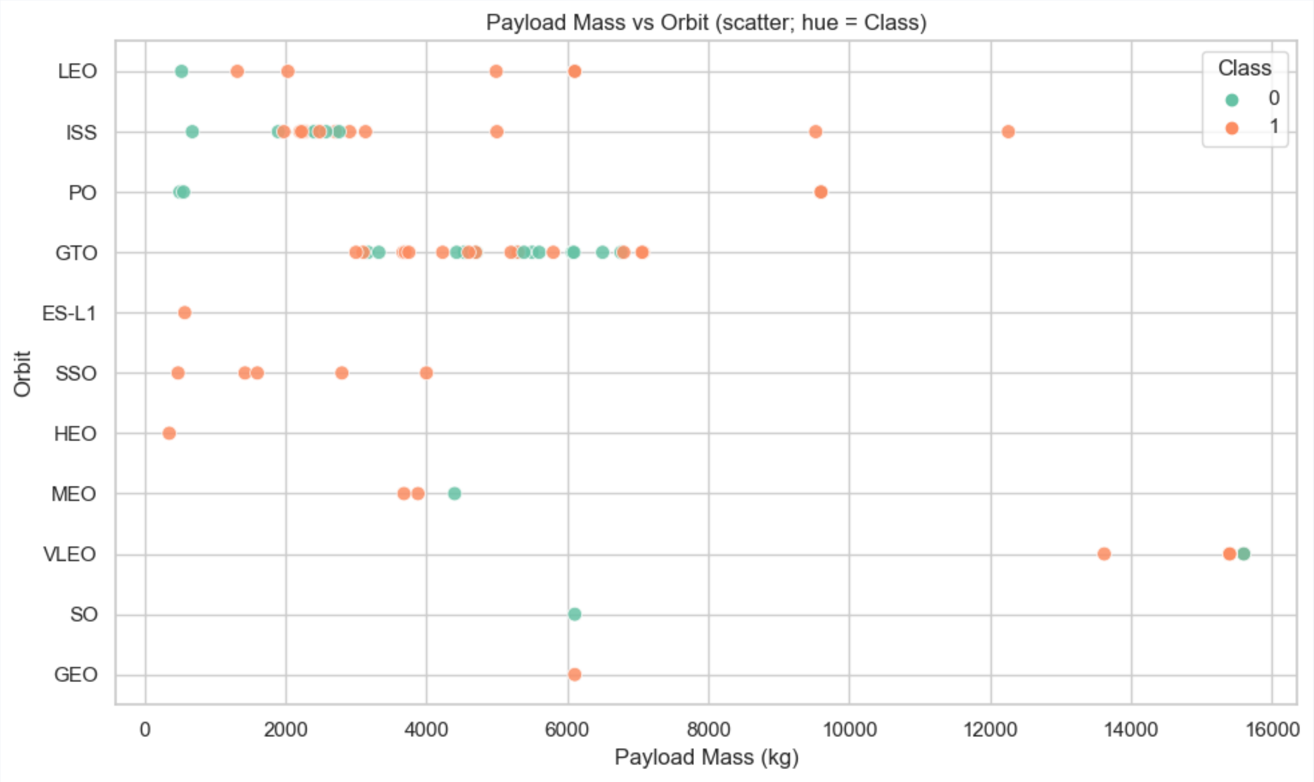
Flight Number vs. Orbit Type

- **Explanation:** For LEO (Low Earth Orbit), the likelihood of a successful landing increases with the number of flights, indicating learning and improvement over time. In contrast, for GTO (Geostationary Transfer Orbit), there is no clear relationship between flight number and success, suggesting that other factors may influence outcomes for this orbit.



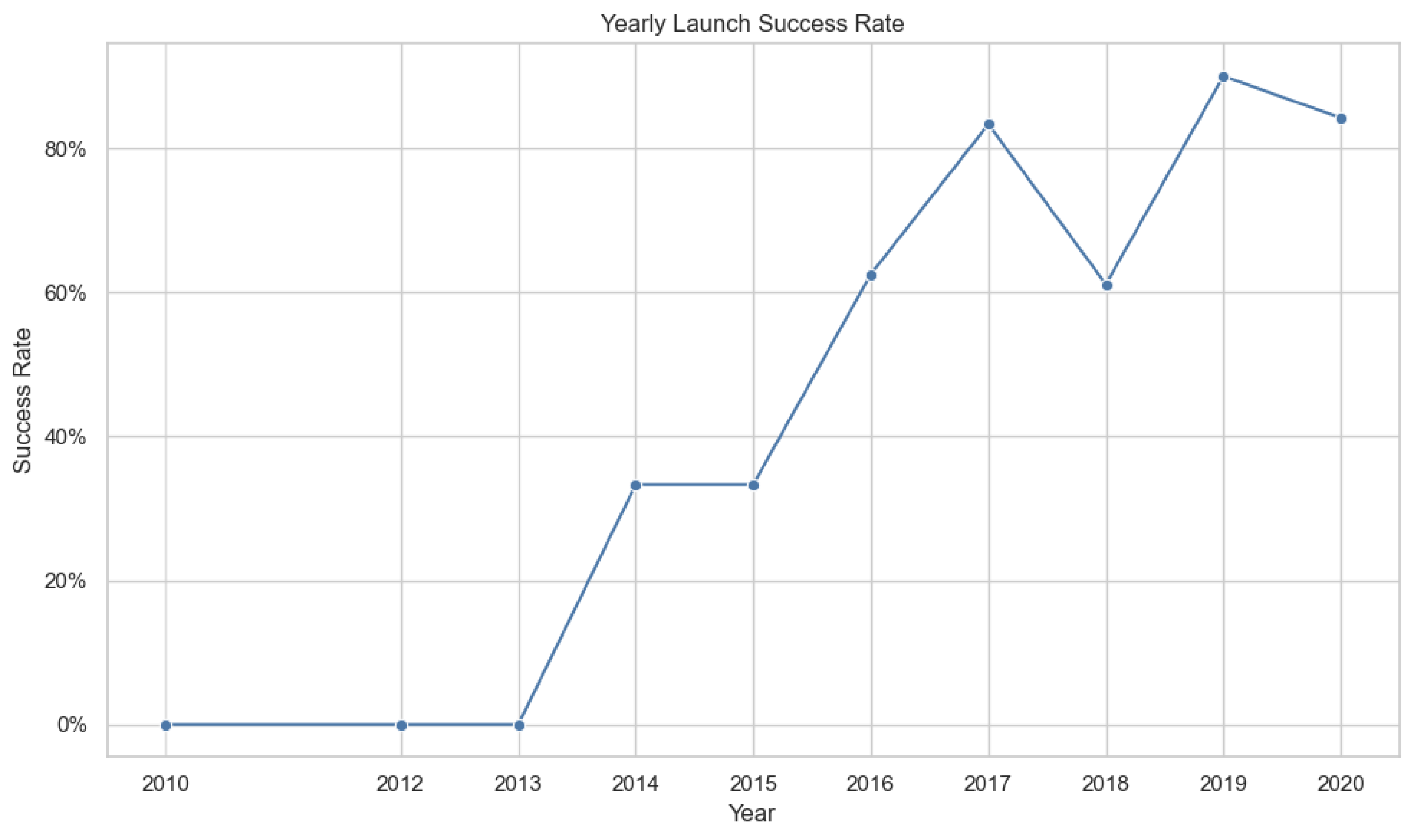
Payload vs. Orbit Type

- **Explanation:** Successful landings with heavy payloads are more common for Polar, LEO, and ISS orbits. For GTO orbits, both successful and unsuccessful landings occur with heavy payloads, making it difficult to distinguish a clear pattern. This suggests that orbit type influences the relationship between payload mass and landing success



Launch Success Yearly Trend

- **Explanation:** The launch success rate has steadily increased from 2013 to 2020. This trend indicates continuous improvement in SpaceX's technology and processes, leading to more frequent successful landings over time.



All Launch Site Names

- Find the names of the unique launch sites
  - 'CCAFS SLC 40’
  - 'VAFB SLC 4E’
  - 'KSC LC 39A’
- Explanation: We can find the unique launch site names in the dataset by using the unique function. This will display the names of all unique launch sites in your SpaceX dataset.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Explanation: We can find 5 records where launch sites begin with CCA using the str.startswith('CCA')] command and then list out only the top 5 records with head(5)

```
   FlightNumber        Date BoosterVersion  PayloadMass Orbit    LaunchSite  \
0             1 2010-01-01       Falcon 9  6104.959412   LEO  CCAFS SLC 40
1             2 2012-01-01       Falcon 9   525.000000   LEO  CCAFS SLC 40
2             3 2013-01-01       Falcon 9   677.000000   ISS  CCAFS SLC 40
4             5 2013-01-01       Falcon 9  3170.000000   GTO  CCAFS SLC 40
5             6 2014-01-01       Falcon 9  3325.000000   GTO  CCAFS SLC 40

     Outcome  Flights  GridFins  Reused   Legs LandingPad  Block  ReusedCount  \
0  None None        1     False   False  False        NaN    1.0            0
1  None None        1     False   False  False        NaN    1.0            0
2  None None        1     False   False  False        NaN    1.0            0
4  None None        1     False   False  False        NaN    1.0            0
5  None None        1     False   False  False        NaN    1.0            0

  Serial  Longitude   Latitude  Class ClassLabel  Year
0  B0003 -80.577366  28.561857      0    Failure  2010
1  B0005 -80.577366  28.561857      0    Failure  2012
2  B0007 -80.577366  28.561857      0    Failure  2013
4  B1004 -80.577366  28.561857      0    Failure  2013
5  B1005 -80.577366  28.561857      0    Failure  2014
```

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Total payload carried by all boosters: 549446.3470588236 kg
- Explanation : To calculate the total payload carried by boosters for NASA, but there is no customer column, we cannot filter for NASA directly with the current data.So the sum of total payload for all launches (since we cannot filter for NASA) we use sum of ‘PayloadMass’
-

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - Average payload mass for Falcon 9 boosters: 6104.959411764707 kg
- **Explanation:** The procedure to find this is by calculating the average payload mass for booster version Falcon 9k

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
First successful landing with a non-blank LandingPad:
         Date                LandingPad BoosterVersion    Outcome
16 2015-01-01  5e9e3032383ecb267a34e7c7       Falcon 9  True RTLS
```

- Explanation: The first successful landing with a non-blank LandingPad. This cell filters for successful landings (Class == 1), ensures LandingPad is not blank, sorts by date, and displays the first such record.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - Booster versions with successful drone ship landings and payload mass between 4000 and 6000 kg: ['Falcon 9’]
- Explanation: This lists the names of booster versions which have successfully landed on a drone ship (Outcome as "True ASDS") and had a payload mass greater than 4000 kg but less than 6000 kg.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - Total successful missions: 60
  - Total failed missions: 30
- Explanation: to calculate and print the total number of successful and failed mission outcomes filter the out come based on class and check the count.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
  - Booster version(s) with the maximum payload mass (15600.0 kg): ['Falcon 9']
- Present your query result with a short explanation here
  - The names of the booster versions that have carried the maximum payload mass can be obtained by checking the max value in the payload.

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
Failed drone ship landings in 2015 (Outcome, BoosterVersion, LaunchSite):
       Outcome BoosterVersion    LaunchSite
11  False ASDS       Falcon 9  CCAFS SLC 40
13  False ASDS       Falcon 9  CCAFS SLC 40
15   None ASDS       Falcon 9  CCAFS SLC 40
```

- **Explanation:** To list the failed landing outcomes on drone ships in 2015, along with their booster versions and launch site names can be obtained by applying the filters to the 'Outcome', 'ADAS' and 'Class'

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

```
Landing outcome counts (2010-06-04 to 2017-03-20), ranked descending:
True ASDS      12
None None      10
True RTLS       8
True Ocean      4
False ASDS      4
False Ocean     2
None ASDS       2
Name: Outcome, dtype: int64
```

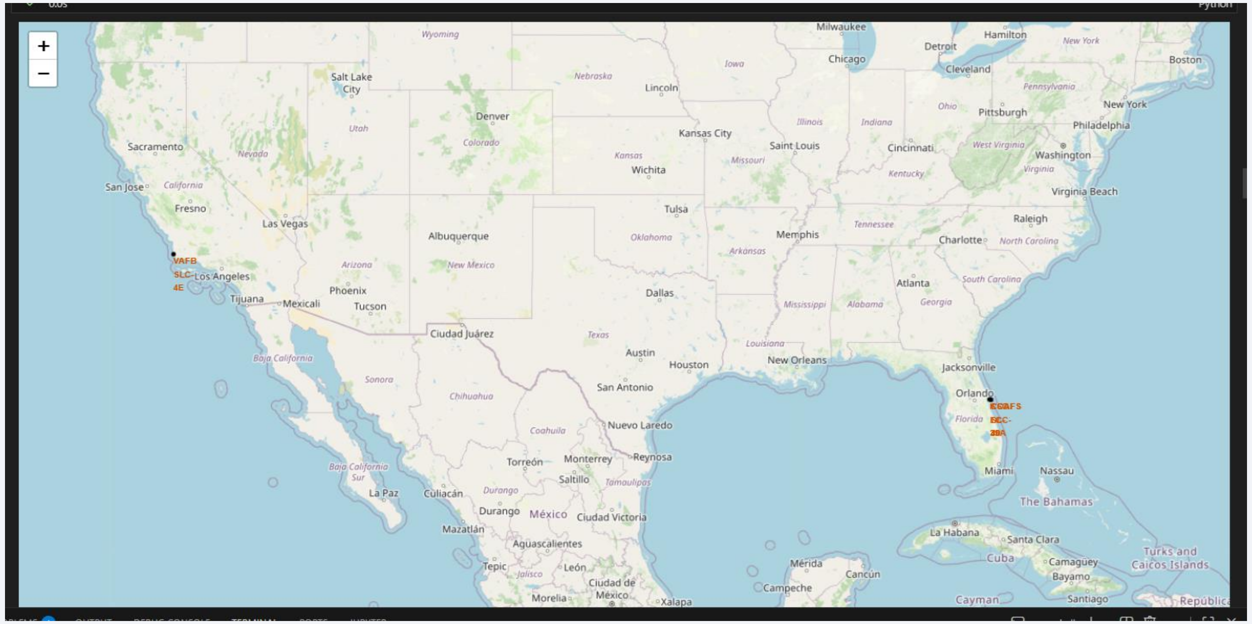
- **Explanation:** To rank the count of landing outcomes (such as "Failure (drone ship)" or "Success (ground pad)") between 2010-06-04 and 2017-03-20 in descending order is obtained by applying a mask of the given dates and extracting the outcome and sort it in descending order.

Section 3

Launch Sites Proximities Analysis

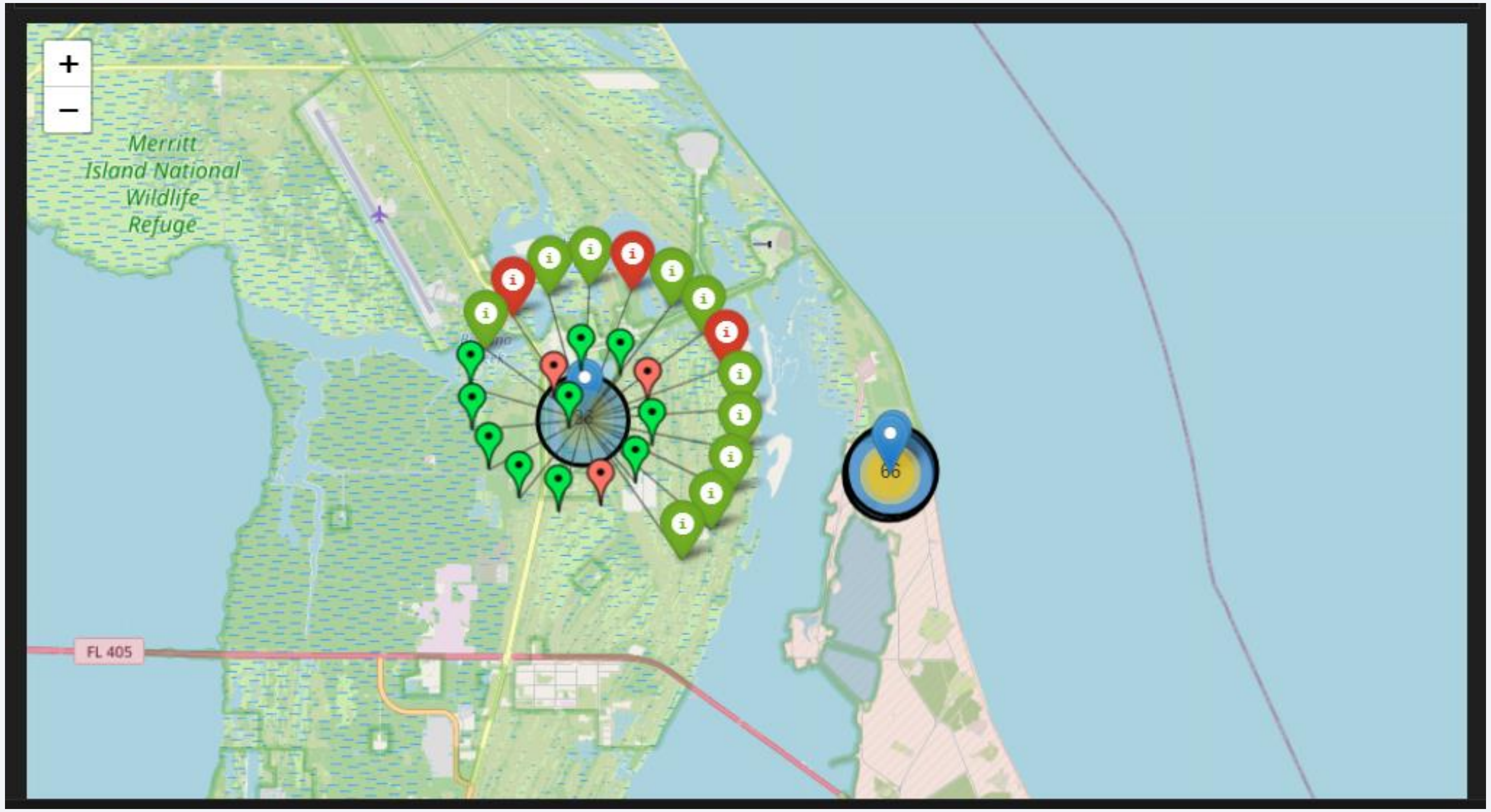
Folium Map – With all the Launch sites

- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map



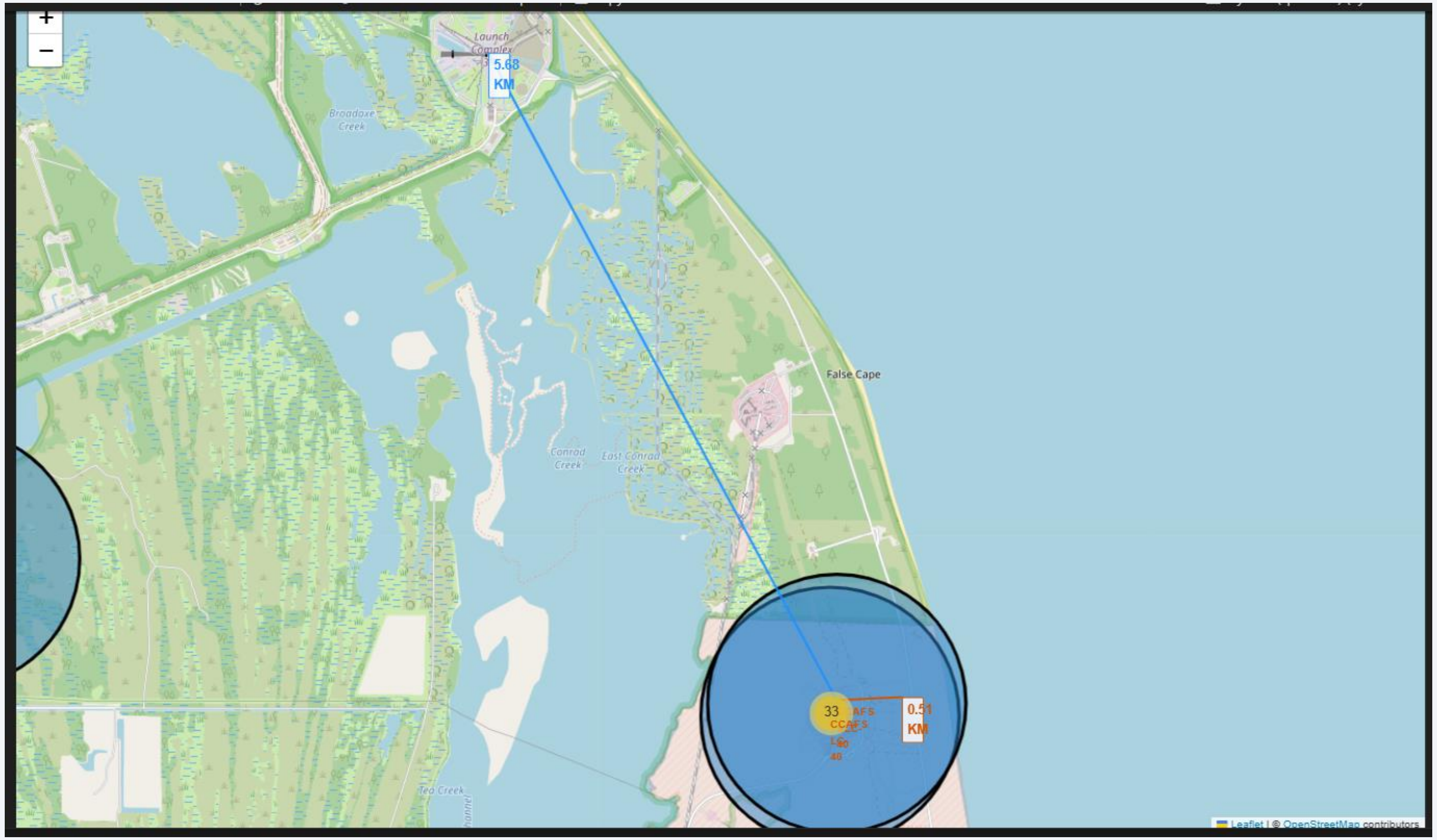
- Important elements and findings: In the map, all the launch sites are marked in red colour

Success & Failure Launches for each site



- Explanation: As there are lot of attempts made from the same sight. It is difficult to explain all the attempts, but a marker cluster feature of Folium can make it easy to visualize. Red indicates failure attempts and green indicates success attempts.

Distance Calculation from Proximities

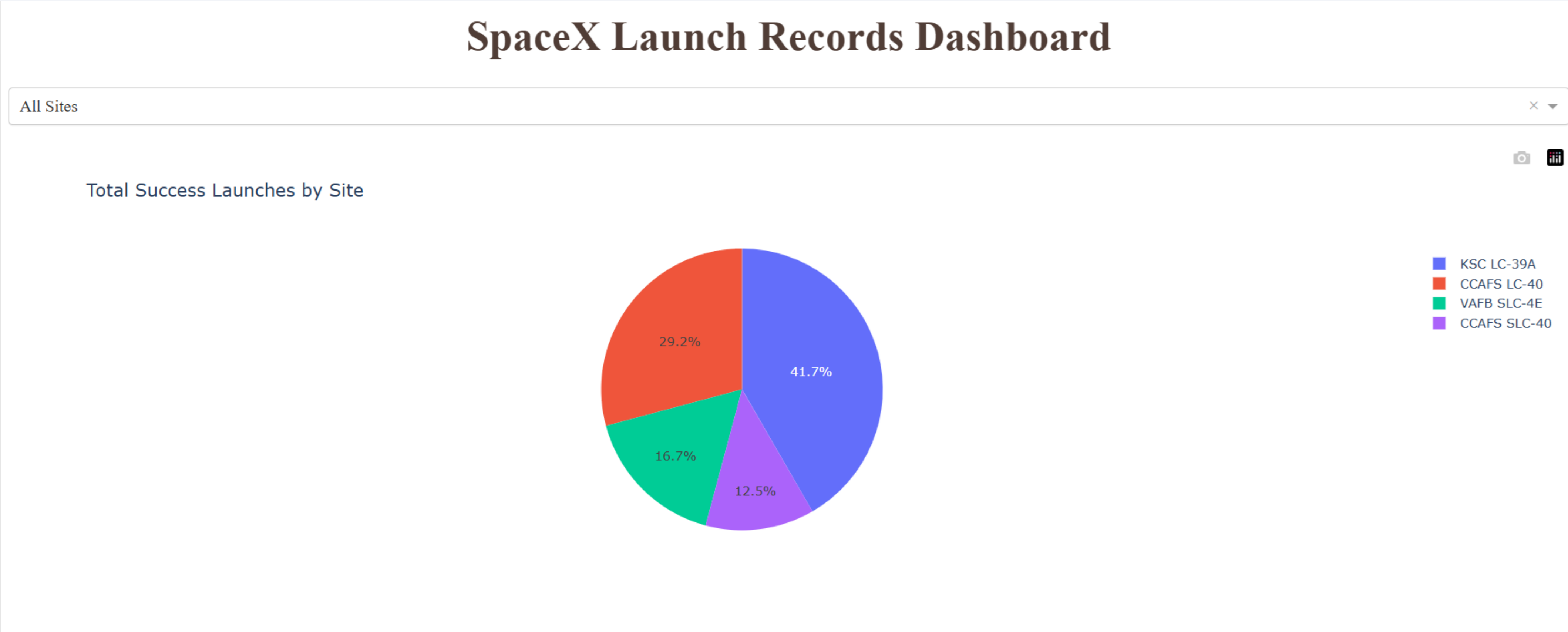


- **Explanation:** The above screen shot shows the distance from the coast as 0.5km and the distance from the Launch complex is 5.68KM. This is possible with the polyline feature.

Section 4

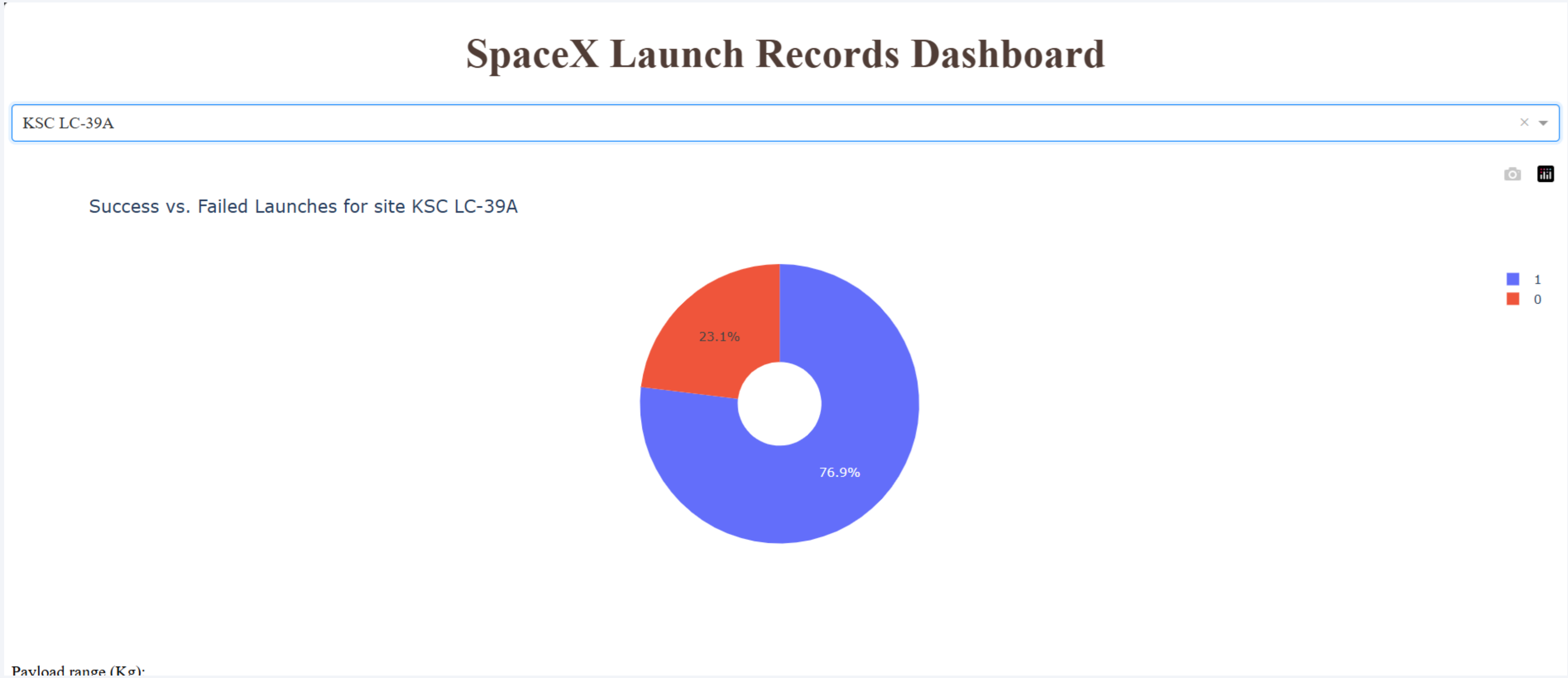
Build a Dashboard with Plotly Dash

Launch Status of All Sites



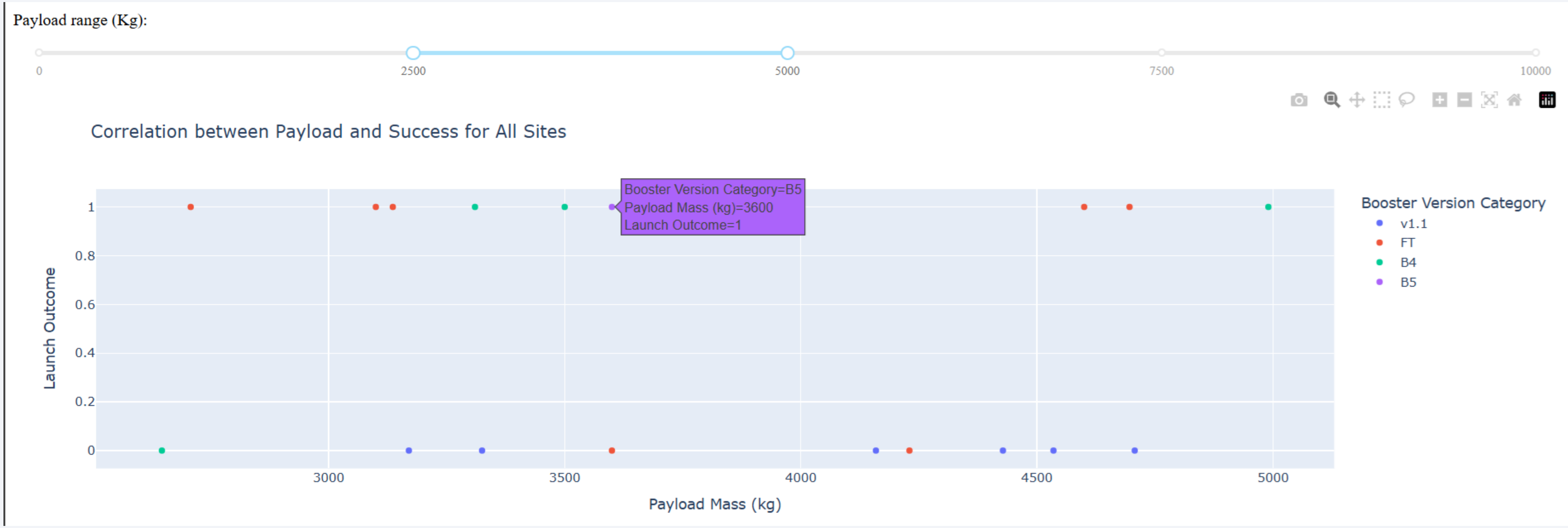
Explanation: The chart shows that among all the types of launch sites, the maximum of 41.7% of the total sites which is KSC LC-39A. And the least among all is the 16.7% which is VAFB SLC-4E

Launch Site with Highest Success Ratio



- **Explanation:** The maximum launches were performed from the KSC LC-39A with 41.7%. Also, in this site success ratio is highest. With a success percentage of 76.9%.

Launch Outcome for all sites for a range of payload

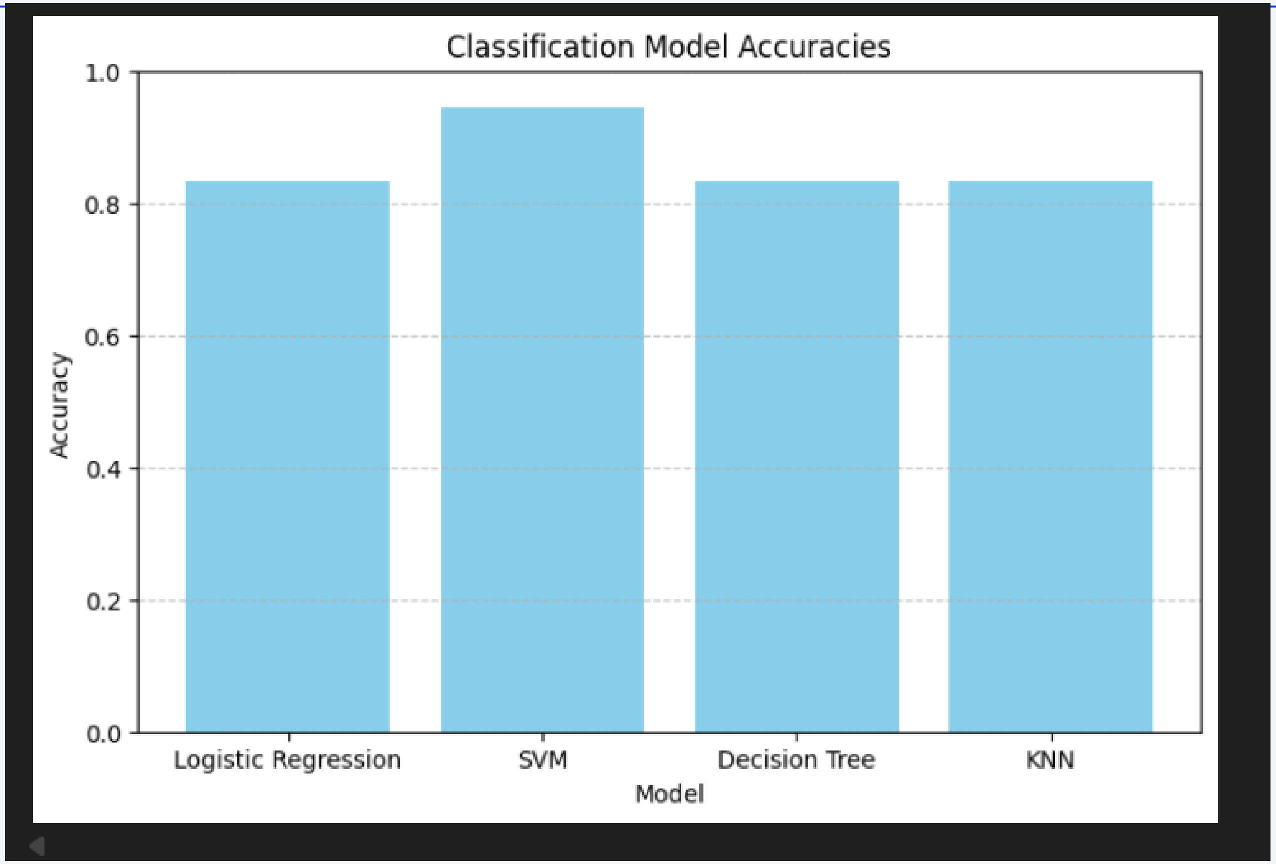


- Explanation: In the given range of payload of 2500kg to 5000kg, maximum success was achieved at the lower range of payloads. The booster version of v.1.1 has the largest success rate.

Section 5

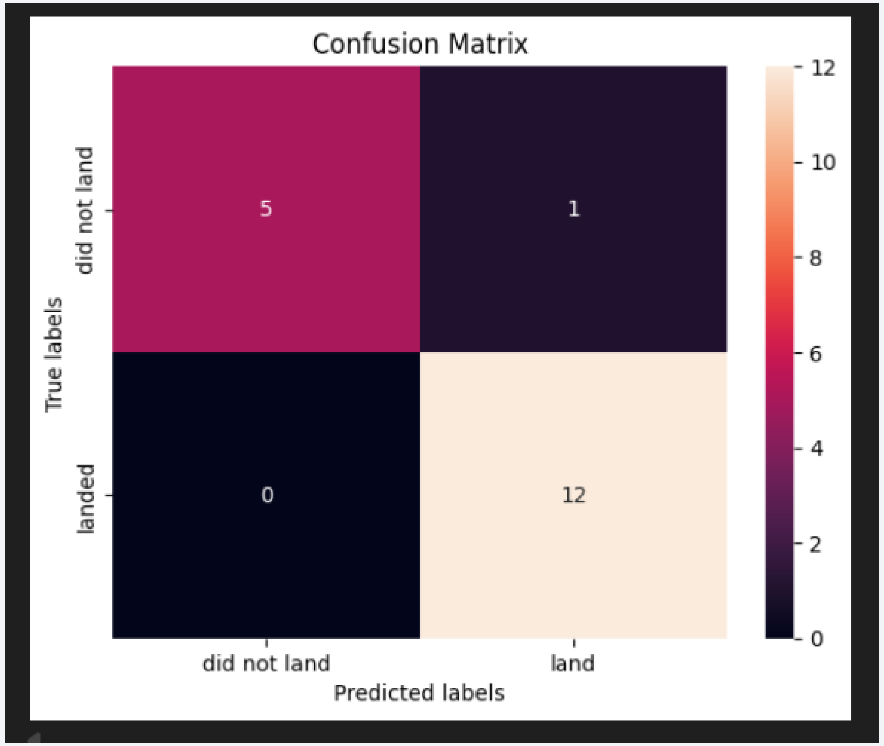
Predictive Analysis (Classification)

Classification Accuracy Comparision



- **Explanation:** From the bar graph, it is evident that the SVM Model has the maximum classification accuracy of 0.94444

Best Confusion Matrix- SVM



- **Explanation:** A good model will have high values on the diagonal (true positives and true negatives) and low values off the diagonal (false positives and false negatives).
- The SVM model is highly accurate, especially for predicting successful landings (no missed landings).
- Only one case was incorrectly predicted as a landing when it was not.
- The model is reliable for both classes, with very few mistakes and no false negatives. This means it is unlikely to miss a successful landing, which is important for mission planning and cost estimation.

Conclusions

- SVM: Test accuracy = 0.944
- Decision Tree: Test accuracy = 0.833
- KNN: Test accuracy = 0.833
- Logistic Regression: Test accuracy = 0.833
- SVM performed the best on this dataset.

Appendix

- Refer to the GitHub links provided in the respective sections for the code snippets and the outputs of all the cells are available.

Thank you!