IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Filipe Pereira
20-09-2024

Outline

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

Executive Summary

Summary: In this capstone, we (a SpaceX rival company) will analyze SpaceX launch data and predict if the Falcon 9 first stage will land successfully. A summary for the methodologies and results described in this report is outlined below.

- Summary of methodologies:
  - Data Collection: In this final project I used API and Web Scrapping to collect data.
  - Data Wrangling: Exploratory Data Analysis (EDA) was used to find some patterns in the data and determine what would be the label for training supervised models.
  - EDA with SQL and Visualization: The goal was to uncover more insights about the data.
  - Visual analytics and Dashboards: Creation of interactive maps and dashboards to get a better sense of the variables and their effect.
  - Machine Learning (ML) methods used to determine if the first stage of Falcon 9 will land successfully:
    - Support Vector Machine (SVM)
    - KNeighborsClassifier
    - Classification Trees
    - Logistic Regression

Executive Summary

- **Summary of all results**
  - **Key Metrics and Data**

    The most importante variables to determine na sucessfull launch are:

    1. **Launch Site**: Mentioned in the exploratory analysis.
    2. **Payload Type and Mass**: Discussed in terms of payload capacity.
    3. **Booster Version and Landing Success**: Mentioned as a key technological factor.
  - **Major Findings**
    1. **Launch Site Insights**: Most launches occurred at CCAFS-SLC-40 and KSC-LC-39A, with earlier launches concentrated at CCAFS-SLC-40.
    2. **Payload Impact**: Heavier payloads, especially **to VLEO, GEO, and ISS, were more likely launched from KSC-LC-39A**.
    3. **Success Rates**: Higher success rates were observed for launches to orbits like GEO, HEO, and ES-L1.
    4. **Booster Landings**: The probability of successful booster landings improved over time, with the first successful landing on 06/05/2016.

Introduction

Project background and context

This project is part of the final Capstone for the IBM Data Science certification. It integrates knowledge acquired from nine course modules, covering a full data science skill set, including methodologies, mathematics, and practical applications. As part of the project, we assume the role of a Data Scientist working at a SpaceX rival company. The objective is to analyze SpaceX's launch data to extract as many insights as possible, particularly regarding launch site performance, payloads, and booster landing success rates, to gain a competitive edge

Problems you want to find answers

1. What factors influence the success rate of SpaceX launches?
   - Analyzing launch sites, payloads, and mission types.
2. How does booster landing technology affect cost savings and launch success?
   - Investigating the reuse of boosters and its impact on reducing costs.
3. Which launch sites and payload types have the highest success rates?
   - Comparing performance across different launch sites and orbits.
4. Can we predict booster landing success using machine learning models?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data about the launches was collect through SpaceX API.
  - Web scrapping was used to get Falcon 9 and Falcon Heavy Launches Records from a Wikipedia page.
- Perform data wrangling
  - Filtering the data.
  - Dealing with missing values.
  - Using One Hot Encoding to prepare the data to a binary classification.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning and evaluation of classification models to ensure the best results.

Data Collection

Data collection process involved a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX's Wikipedia entry.

We had to use both of these data collection methods in order to get complete information about the launches for a more detailed analysis.

Data Columns are obtained by using SpaceX REST API:

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite,

Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount,

Serial, Longitude, Latitude

Data Columns are obtained by using Wikipedia Web Scraping:

Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch

outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API

GITHUB: Data Collection API

1. Requesting data from SpaceX API
2. Decoding the response content using .json() and turning it into a dataframe using .json_normalize()
3. Requesting information from SpaceX API by applying custom functions
4. Create a Dictonary with the Data
5. Create a dataframe from the dictionary
6. Filtering and replacing missing values
7. Exporting the data to CSV

Data Collection - Scraping

GITHUB URL: Web-scrapping

Data Wrangling

GITHUB URL: Data Wrangling

This SpaceX Falcon 9 data wrangling procedure focuses on preparing the data for machine learning by converting landing outcomes into binary labels for success (1) or failure (0).

Key steps include:

Exploratory Data Analysis (EDA): Initial data inspection is done to understand the dataset, check for patterns, and determine which labels to use for training. Specifically, the goal is to classify whether the booster successfully landed.

Handling Landing Outcomes: The dataset contains several different landing outcomes (e.g., True Ocean, False ASDS). These are converted into binary labels where 1 means a successful landing and 0 represents a failure.

Loading and Analyzing Data: The SpaceX dataset is loaded, and essential insights such as missing values, numerical vs categorical data, and launch site distribution are explored. This step also identifies key metrics like the number of launches per site and the frequency of different orbit types.

Labeling Outcomes: Specific landing outcomes are mapped to binary values (1 for successful, 0 for unsuccessful). These labels will later serve as the target variable for supervised learning.

Exporting Cleaned Data: The cleaned and labeled dataset is saved for further modeling and analysis, ready to be used in machine learning tasks.

Data Wrangling

GITHUB URL: Data Wrangling

1. Exploratory Data analyses and determine training labels.
2. Calculate number of launches in each site.
3. Calculate the number of occurances in each orbit.
4. Calculate the number and occurrence of mission outcome per orbit type
5. Create a landing outcome label from Outcome column

EDA with Data Visualization

GITHUB URL: Data Visualization

- **Flight Number vs. Payload Mass (with success overlay):**
  - Used to examine the relationship between flight number, payload mass, and landing success. It shows that as flight number increases, success is more likely, even with heavier payloads.
- **Flight Number vs. Launch Site:**
  - Plotted to explore the correlation between flight number and launch sites, with successful outcomes highlighted. It reveals which sites experienced more launches and their success rates.
- **Payload Mass vs. Launch Site:**
  - This scatter plot highlights the payload masses associated with different launch sites and their outcomes. It helps identify which sites handle heavier payloads.
- **Success Rate by Orbit Type:**
  - A bar chart displaying the success rate of each orbit type. This helps to compare how different orbit types influence landing success.

EDA with Data Visualization

GITHUB URL: Data Visualization

- **Flight Number vs. Orbit:**
  - A scatter plot to investigate the connection between flight number and orbit type, indicating which orbits have a higher success rate as flights progress.
- **Payload Mass vs. Orbit:**
  - Used to analyze how payload mass influences success across various orbit types, revealing that heavier payloads are more successful in specific orbits like LEO.
- **Yearly Success Trend:**
  - A line chart that tracks the average success rate over time, showing how SpaceX's success rate improved over the years.

EDA with SQL

GITHUB URL: EDA with SQL

Bullet-point summary of the SQL queries performed:

Task 1: Displayed unique launch sites in the space mission.

```
SELECT DISTINCT Launch_Site FROM SPACEXTBL
```

Task 2: Retrieved 5 records where launch sites begin with 'CCA'.

```
SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE "CCA%" LIMIT 5
```

Task 3: Displayed total payload mass carried by boosters launched by NASA (CRS).

```
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTBL WHERE Customer LIKE '%CRS%’
```

Task 4: Displayed the average payload mass carried by booster version F9 v1.1.

```
SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1’
```

Task 5: Listed the date of the first successful landing outcome on a ground pad.

```
SELECT Date FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)' ORDER BY Date ASC LIMIT 1
```

EDA with SQL

GITHUB URL: EDA with SQL

Task 6: Listed the booster versions that succeeded on drone ships and carried payloads between 4000 and 6000 kg.

```
SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000
AND PAYLOAD_MASS__KG_ < 6000
```

Task 7: Displayed the total number of successful and failed mission outcomes.

```
SELECT Mission_Outcome, COUNT(*) AS Total_Count FROM SPACEXTBL GROUP BY Mission_Outcome
```

Task 8: Listed the booster versions that carried the maximum payload mass.

```
SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

Task 9: Listed records showing month names, failure outcomes on drone ships, booster versions, and launch sites for 2015.

```
SELECT Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTBL WHERE substr(Date,0,5)='2015' AND Landing_Outcome
= 'Failure (drone ship)’
```

Task 10: Ranked landing outcomes between 2010-06-04 and 2017-03-20, in descending order.

```
SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC
```

Build an Interactive Map with Folium

[GITHUB URL: Map with Folium]

Markers for Launch Sites:

Description: For each launch site in the dataset, a marker was added to pinpoint its location on the map.

Why Added: Markers clearly identify the specific location of each launch site (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E), allowing for better geographic context.

Circles around Launch Sites:

Description: A circle was drawn around each launch site with a radius of 1,000 meters.

Why Added: These circles visually highlight the area surrounding each launch site, making it easier to recognize their spatial proximity to features such as coastlines and other infrastructures.

Marker Clusters for Launch Outcomes:

Description: A MarkerCluster object was used to group multiple launch records at the same location. Each launch is represented with a marker, colored green for success and red for failure.

Why Added: Marker clusters condense information where many launches occur at the same location, enabling a clearer overview of launch outcomes at each site.

Mouse Position Plugin:

Description: The MousePosition plugin was enabled to display latitude and longitude coordinates dynamically as the mouse pointer moves across the map.

Why Added: This feature helps in identifying the geographic coordinates of points of interest (like coastlines, highways, and cities) relative to the launch sites for further proximity analysis.

Build an Interactive Map with Folium

GITHUB URL: Map with Folium

Distance Markers and Polylines:

Description: Distance markers and Polylines were added to draw lines between launch sites and nearby geographic features such as coastlines, highways, railways, and cities. Distance markers display the calculated distances in kilometers.

Why Added: Polylines visualize the spatial relationship between launch sites and nearby infrastructure. Distance markers make the map more informative by showing the proximity of launch sites to significant features like coastlines, railroads, and highways.

Equator Line:

Description: A PolyLine was drawn along the equator to visually represent its location.

Why Added: The equator serves as a reference line for determining whether launch sites are located near this important latitude, which can be relevant for orbital mechanics and launch efficiency.

Coastline, Highway, and City Markers:

Description: Markers with distance labels were added to represent the closest coastline, highway, and city to each launch site.

Why Added: These markers and labels provide insight into how close launch sites are to critical infrastructures like coastlines (for trajectory and safety reasons), highways (for transportation), and cities (for safety and regulatory considerations).

Build a Dashboard with Plotly Dash

GITHUB URL: Ploty Dash

Summary of Plots/Graphs and Interactions in the Dashboard

Dropdown for Launch Site Selection:

Plot: This dropdown allows users to filter the data based on specific SpaceX launch sites, with options for individual sites and all sites combined.

Reason: By providing the ability to select a specific launch site, users can analyze the performance (successful or failed launches) of individual sites or get an overview of all sites together.

Pie Chart for Launch Success Counts:

Plot: A pie chart visualizes the total number of successful launches. When a specific launch site is selected, it shows the count of successful vs. failed launches for that site.

Reason: The pie chart offers a clear visual comparison of success rates either across all sites or for individual sites, helping users quickly assess SpaceX's success performance.

Build a Dashboard with Plotly Dash

GITHUB URL: Ploty Dash

Summary of Plots/Graphs and Interactions in the Dashboard

Payload Slider:

Interaction: A slider is added to allow users to filter the payload mass range, affecting the data displayed in the scatter plot.

Reason: Payload mass is a critical factor in launch success, and this interaction gives users flexibility to narrow down launches based on payload weight, aiding in identifying trends in relation to payload mass.

Scatter Plot for Payload vs. Launch Success:

Plot: A scatter plot shows the relationship between payload mass and launch success, color-coded by booster version category. Users can filter by launch site and payload range.

Reason: This scatter plot helps identify any correlations between payload mass and the likelihood of launch success. It also highlights the effect of different booster versions on launch outcomes, which is useful for deeper technical analysis.

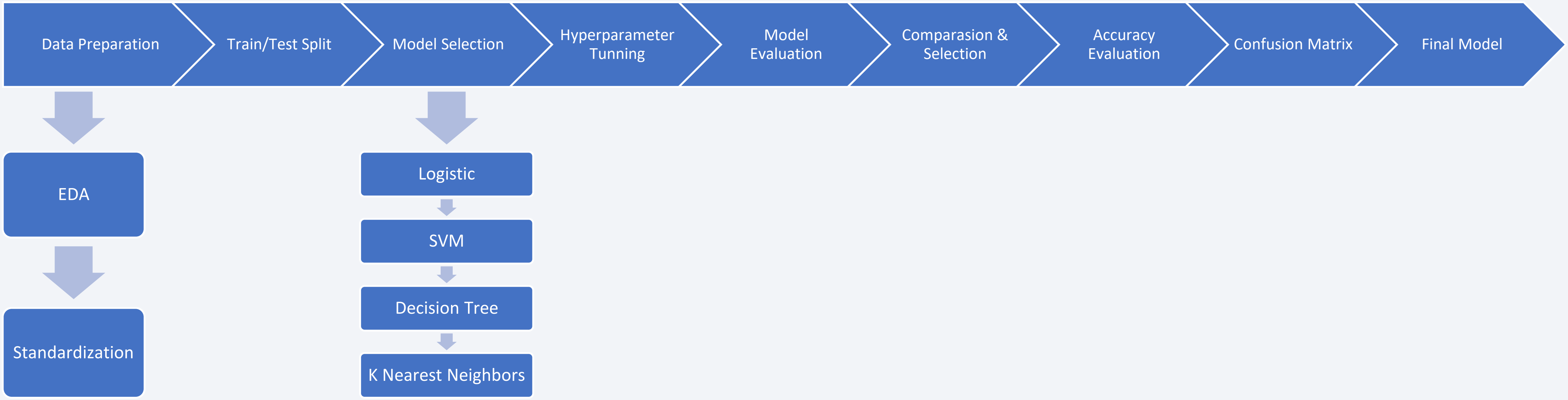
Predictive Analysis (Classification)

GITHUB URL: Predictive Analysis

1. Data Preparation:
    - Exploratory Data Analysis: Understanding the dataset and creating labels (Y = data['Class']).
    - Standardization: Standardize the features (X) using StandardScaler() to ensure all features are on the same scale.
2. Train-Test Split:
    - Use train_test_split() to split the data into training and test sets (X_train, X_test, Y_train, Y_test) with 20% reserved for testing.
3. Model Selection and Hyperparameter Tuning:
    - Logistic Regression: Use GridSearchCV to find the best hyperparameters (e.g., 'C': [0.01, 0.1, 1]).
    - Support Vector Machine (SVM): Tune using grid search across kernel, C, and gamma.
    - Decision Tree Classifier: Grid search for the best parameters, though some hyperparameter combinations (e.g., 'max_features': 'auto') caused errors.
    - K-Nearest Neighbors (KNN): Use GridSearchCV to optimize hyperparameters like n_neighbors, weights, and p for distance metric.
4. Model Evaluation:
    - Evaluate the models using accuracy on the test data (.score()), confusion matrix, and the best hyperparameters (best_params_).
5. Comparison and Selection:
    - Logistic Regression: Achieved 94.44% accuracy on the test set.
    - SVM: Also achieved 94.44% accuracy but tuned with different parameters (best performance).
    - Decision Tree: Struggled due to parameter constraints and fitting issues.
6. Confusion Matrix and Error Analysis:
    - Logistic regression and SVM models showed good performance but some false positives were noted.

Predictive Analysis (Classification)

[GITHUB URL: Predictive Analysis]



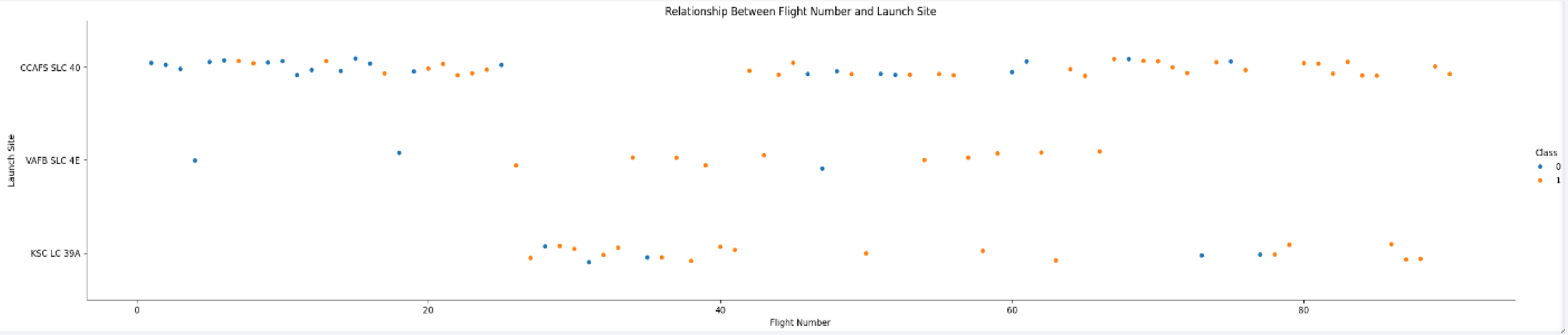
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

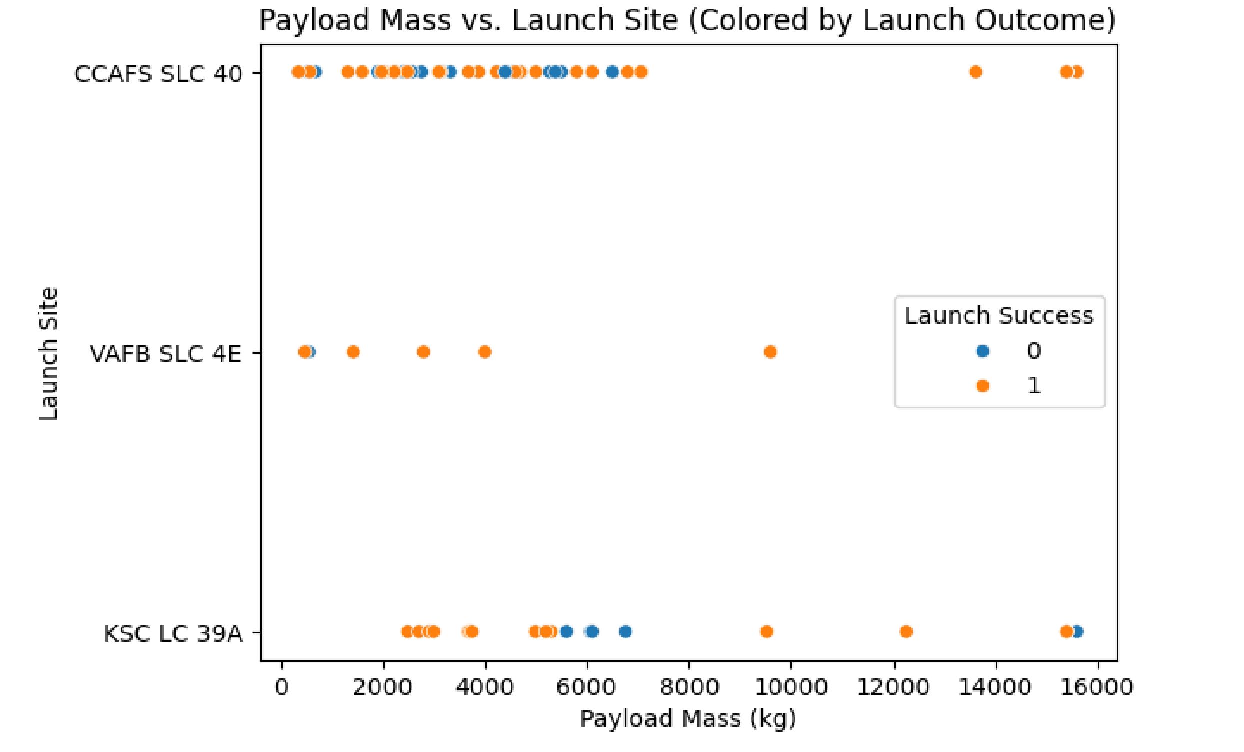
Insights drawn from EDA

Flight Number vs. Launch Site



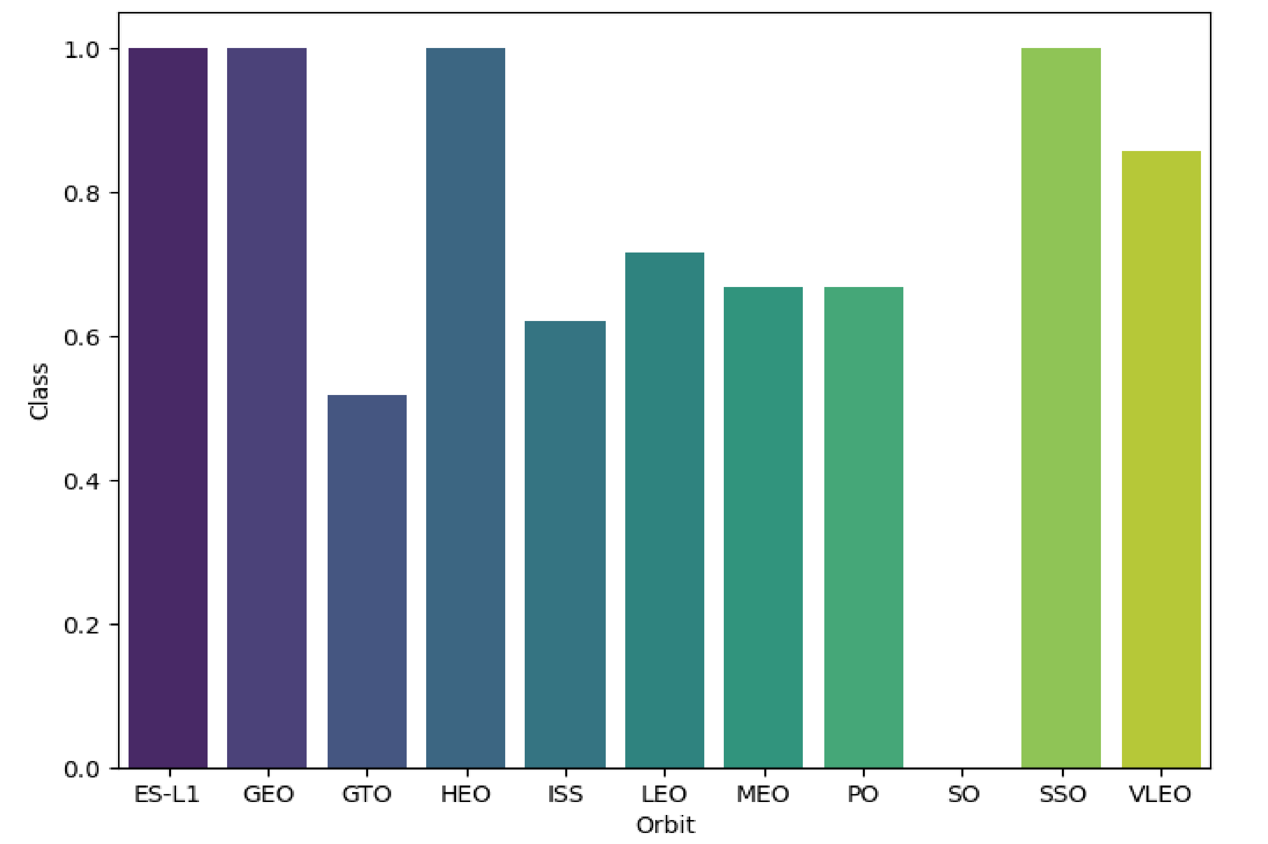
- CCAFS SLC 40 is the most used launch site.
- KSC LC 39A also has a significant number of launches, while VAFB SLC 4E is used less frequently.
- CCAFS SLC 40 and KSC LC 39A show a mix of both successes and failures, though CCAFS SLC 40 seems to have a higher concentration of successful launches.
- VAFB SLC 4E shows fewer total launches, with both successful and unsuccessful launches relatively spread out.
- The success rate improves over time, especially in later flights, with more launches resulting in successful outcomes.

Payload vs. Launch Site



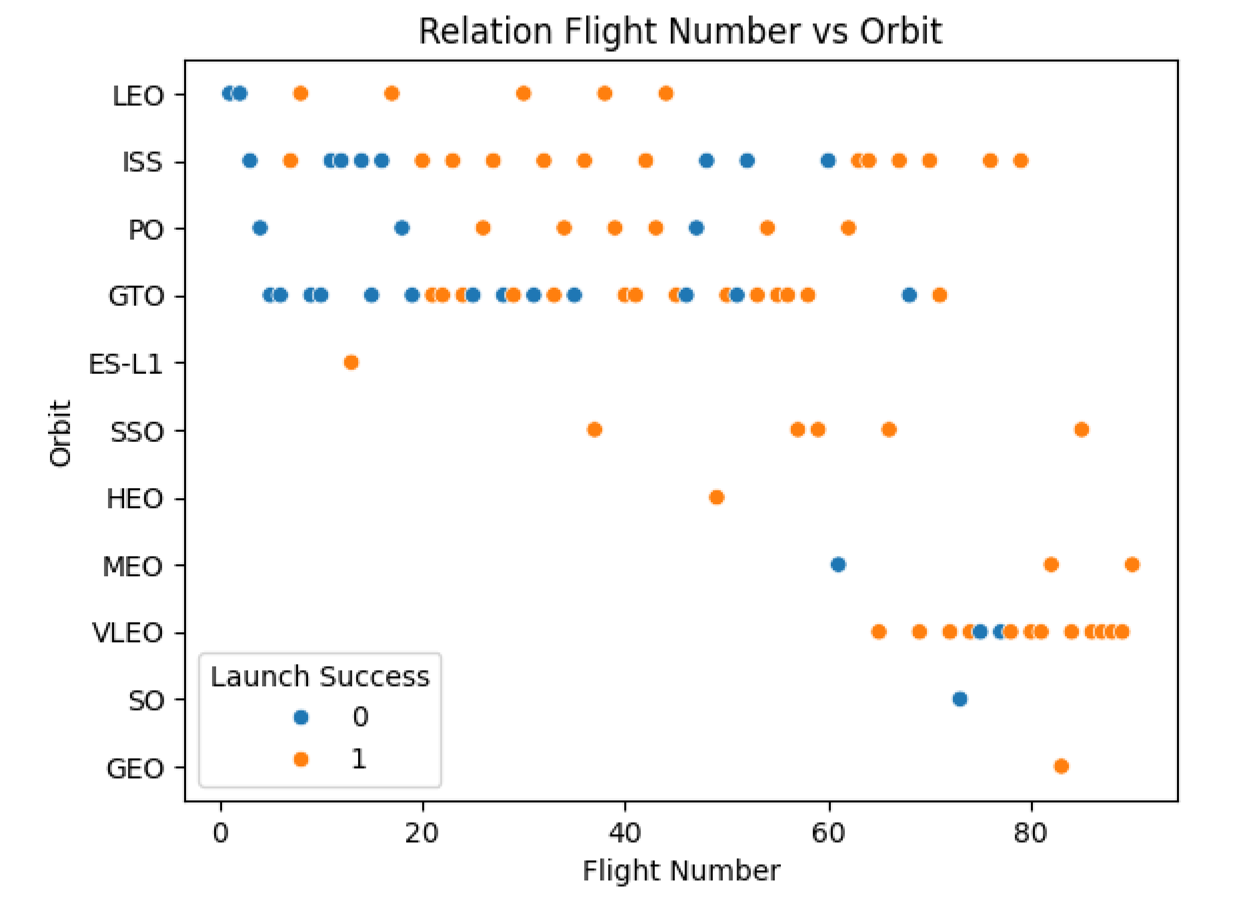
- CCAFS SLC 40 and KSC LC 39A can handle a broader range of payloads, including very heavy ones, while VAFB SLC 4E is more specialized in lighter payloads.
- Heavier payloads (over 10,000 kg) generally have better launch success.
- Smaller payloads show a higher frequency of both successes and failures, possibly due to more diverse or experimental missions.

Success Rate vs. Orbit Type



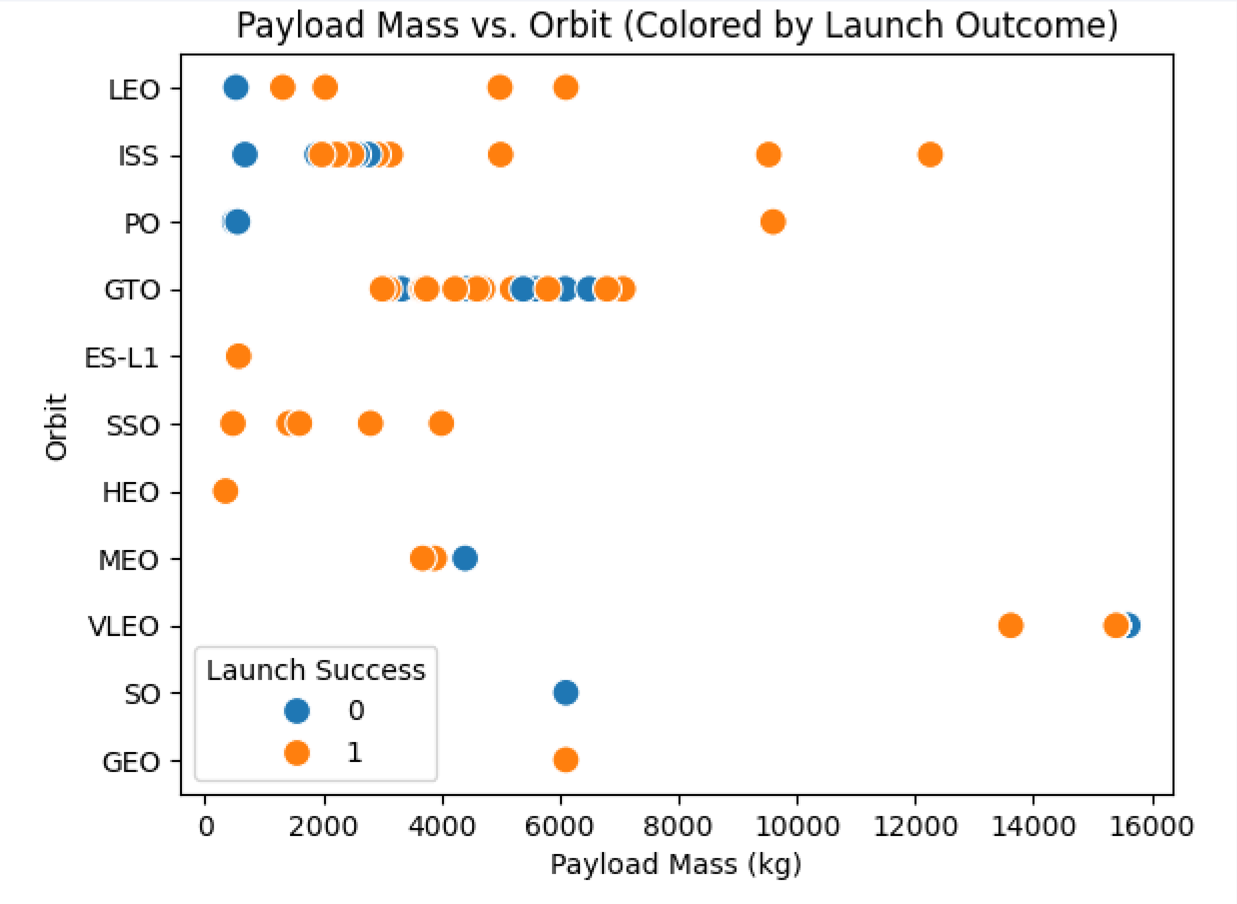
- ES-L1, GEO, HEO, and SSO have a 100% sucess rate.
- SO has a 0% success rate.
- GTO has a 50% success rate.

Flight Number vs. Orbit Type



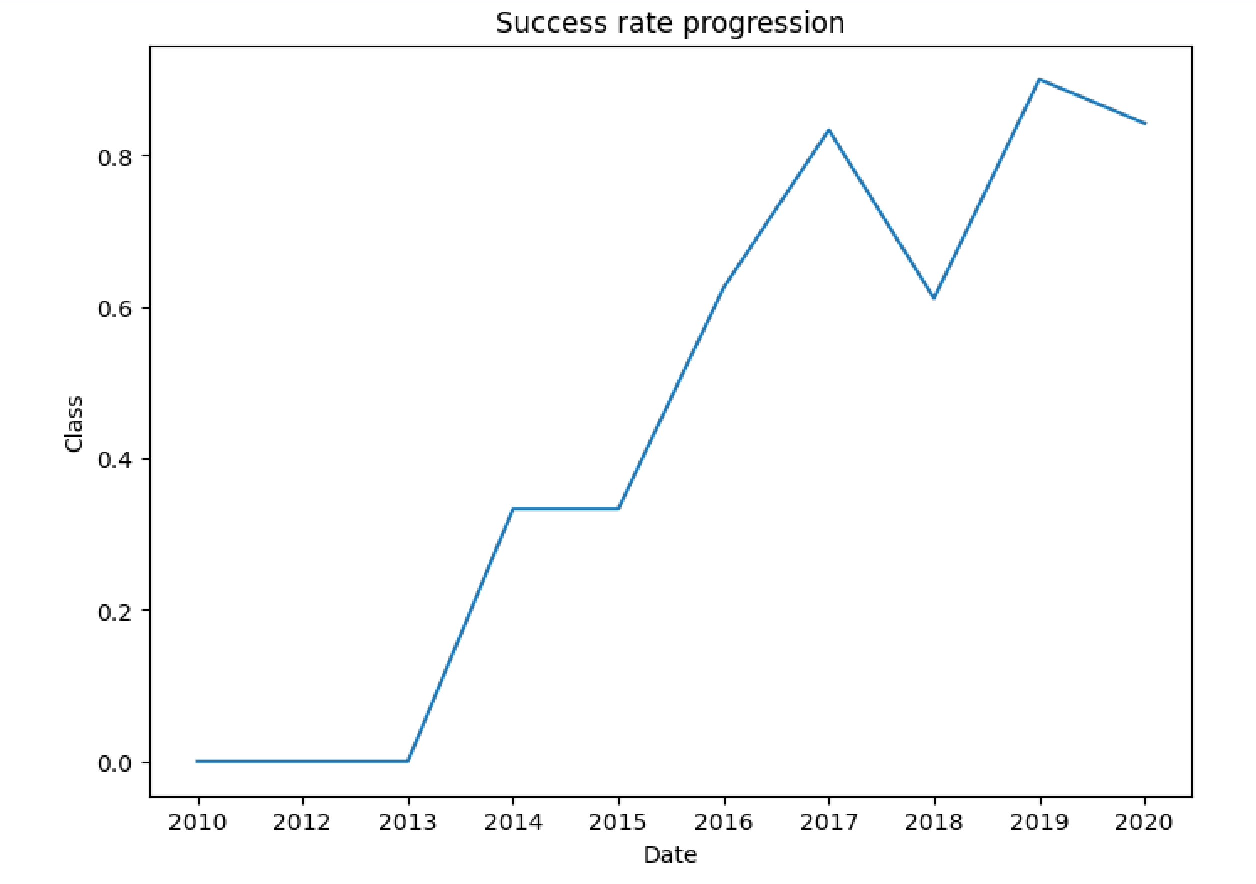
- First flights concentrate at GTO, PO ISS and LEO orbits.
- Last flights concentrate mostly on VLEO orbit
- VLEO orbit has the highest success rate.
- Most failed launches happened on the first flights
- ES-L1 has only 1 launch
- GTO has the most launches

Payload vs. Orbit Type



- Most launches have a payload between 0 and 6000 Kg
- VLEO is the orbit used to the heavier payloads

Launch Success Yearly Trend



- Success rate picked in 2013
- Positive progression with time

All Launch Site Names

Task 1

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

In [13]:
```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL
```

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

Out[13]:

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

This SQL query retrieves a list of unique launch sites from the SPACEXTBL table.

Launch Site Names Begin with 'CCA'

Task 2

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

In [19]:
```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE "CCA%" LIMIT 5
```

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

Out[19]:

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

This SQL query retrieves up to 5 records from the SPACEXTBL table where the Launch_Site field starts with "CCA".

Total Payload Mass

Task 3

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

In [21]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTBL WHERE Customer LIKE '%CRS%'
```

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

Out[21]:

Total_Payload_Mass
48213

This SQL query calculates the total payload mass (in kilograms) from the SPACEXTBL table for entries where the Customer field contains "CRS".

Average Payload Mass by F9 v1.1

Task 4

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

In [22]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1'
```

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

Out[22]:

Average_Payload_Mass
2928.4

This SQL query calculates the average payload mass (in kilograms) from the SPACEXTBL table for records where the Booster_Version is 'F9 v1.1'.

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

In [23]:
```
%sql SELECT Date FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)'ORDER BY Date ASC LIMIT 1
```

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

Out[23]:

Date
2015-12-22

This SQL query retrieves the earliest date from the SPACEXTBL table where the landing outcome was "Success (ground pad)." It orders the results in ascending order and limits the output to just one record.

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

In [24]:
```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND
```

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

Out[24]:

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

Retrieve the booster versions for successful drone ship landings with payload masses between 4000 kg and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [26]:
```
%sql SELECT Mission_Outcome, COUNT(*) AS Total_Count FROM SPACEXTBL GROUP BY Mission_Outcome
```

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

Out[26]:

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

This query counts the total number of launches grouped by their mission outcomes in the SPACEXTBL table.

Boosters Carried Maximum Payload

Task 8

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

In [18]:
```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

Out[18]:

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

This query retrieves the booster version for the launch that carried the maximum payload mass from the SPACEXTBL table.

2015 Launch Records

Task 9

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.

In [19]:

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

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

Out[19]:

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

This query returns the booster version associated with the launch that had the highest payload mass (PAYLOAD_MASS__KG_) in the SPACEXTBL table. It uses a subquery to find the maximum payload mass and then retrieves the corresponding booster version.

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

Task 10

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.

In [20]:

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

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

Out[20]:

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

This query counts the number of landing outcomes for launches that occurred between June 4, 2010, and March 20, 2017, in the SPACEXTBL table. The results are grouped by landing outcome and ordered by the count in descending order.

Section 3

Launch Sites Proximities Analysis

All Launch Sites Location

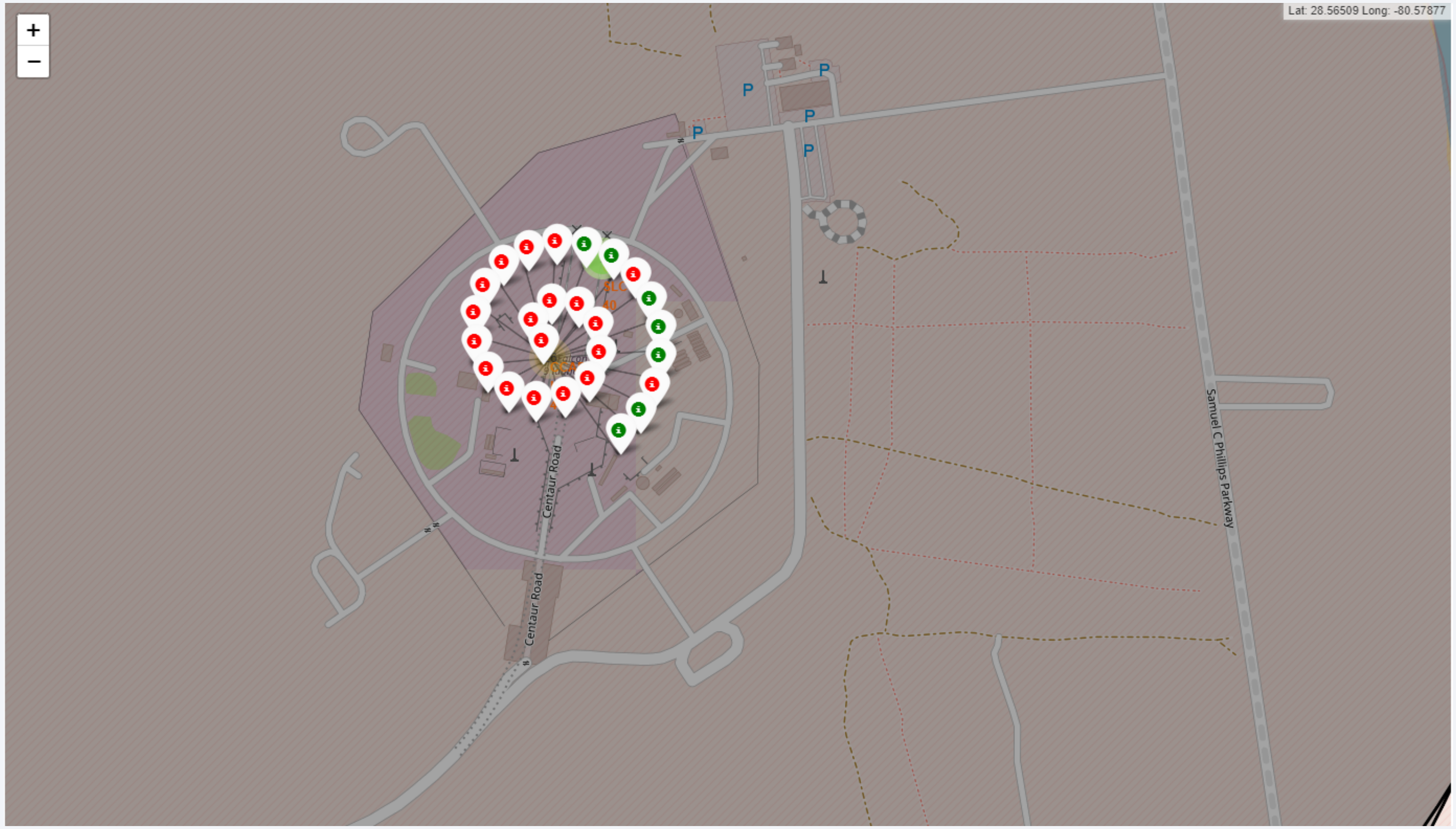
- Most of Launch sites are in proximity to the Equator line. The land is moving faster at the equator than any other place on the surface of the Earth. Anything on the surface of the Earth at the equator is already moving at 1670 km/hour. If a ship is launched from the equator it goes up into space, and it is also moving around the Earth at the same speed it was moving before launching. This is because of inertia. This speed will help the spacecraft keep up a good enough speed to stay in orbit.

- All launch sites are in very close proximity to the coast, while launching rockets towards the ocean it minimises the risk of having any debris dropping or exploding near people.

Launch Outcome Per Location

From the colour-labeled markers we should be able to easily identify which launch sites have relatively high success rates.

- Green Marker = Successful Launch
- Red Marker = Failed Launch



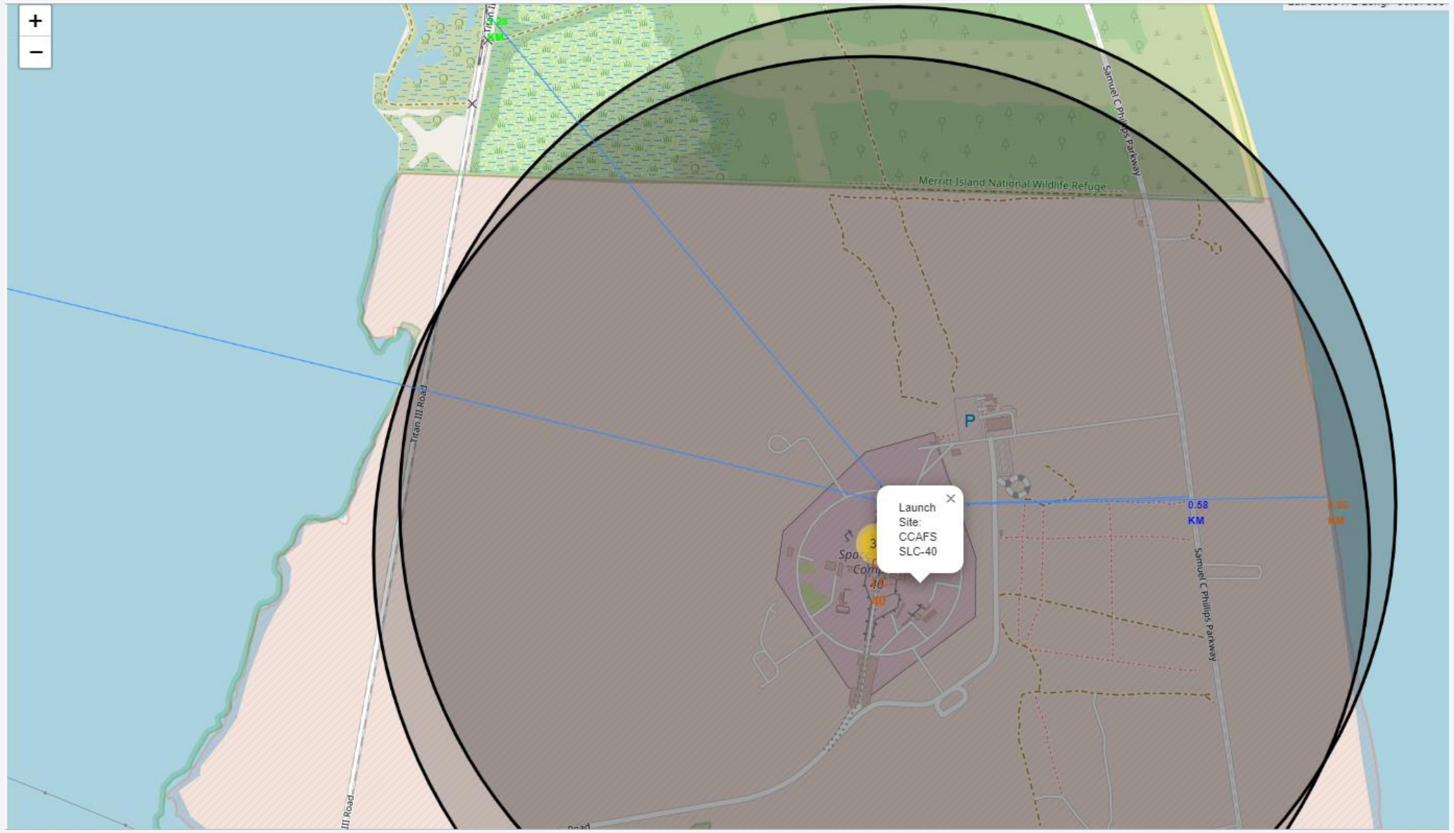
Launch Site Proximities Distance

Launch site: CCAFS SLC -40

With this map iteration, we can display Stratigic points close to the launch site and display the distances.

Points highlighted:

Closest coastline
Closest highway
Closest Railway
Closest city (Titusville 23,5km)

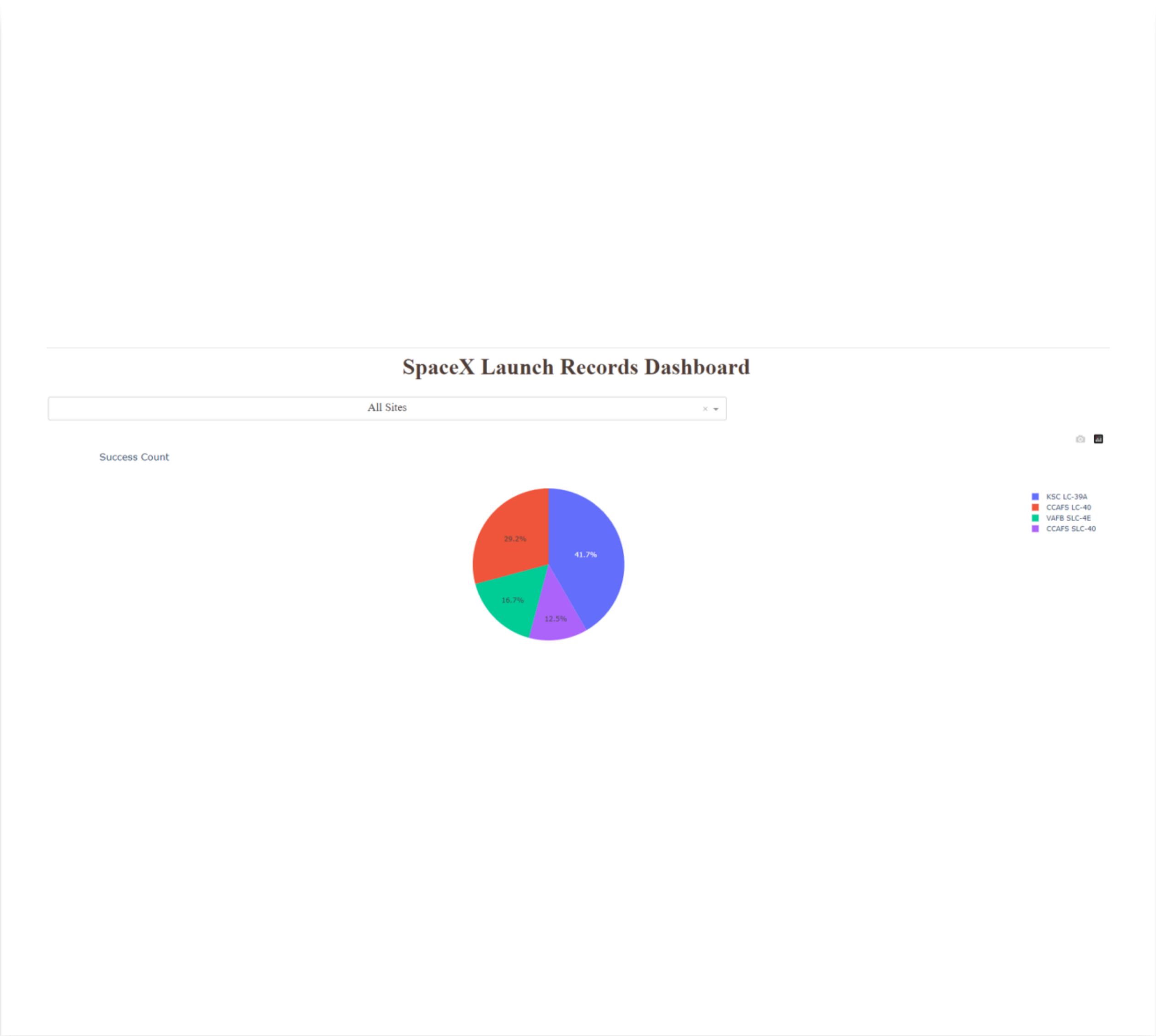


Section 4

Build a Dashboard with Plotly Dash

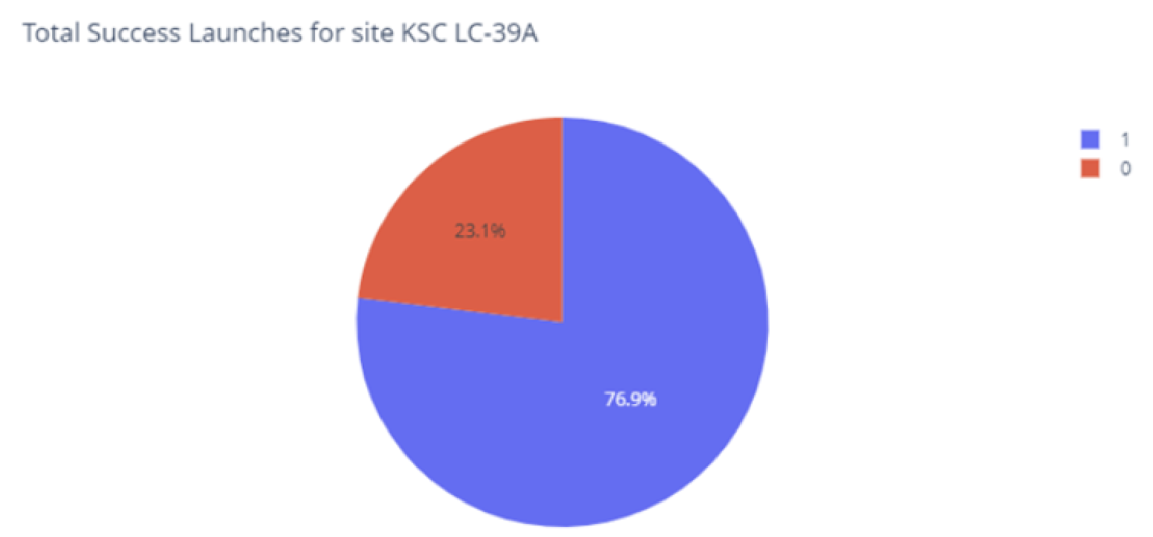
Success Count Dashboard

The chart shows the success count of all sites and contains and filter to choose a specific site.



KSC LC-39 A is the launch site with highest success rate with 76,9%

Lauch Site with highest Success Rate

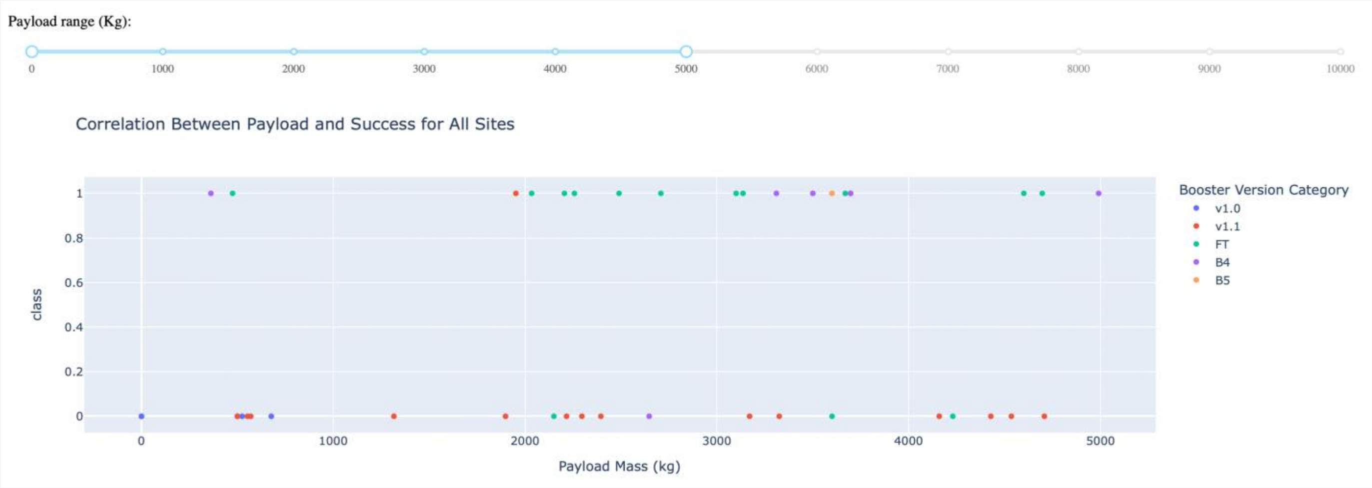


Payload vs. Launch Outcome

This chart shows the variation in success rate with the Payload, allowing us to filter the payload range.

Conclusions:

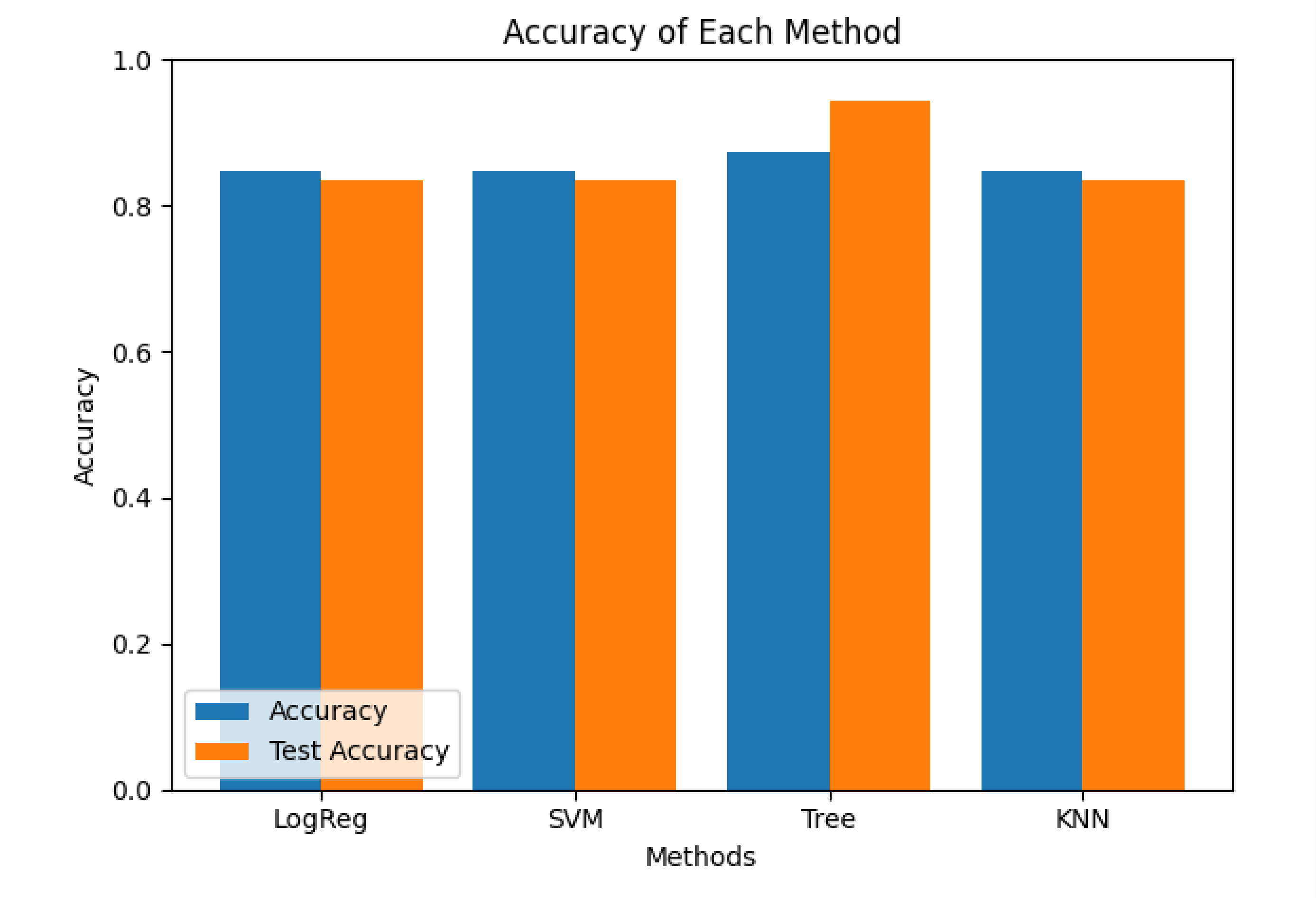
- There are only two types of boosters used for payloads between 5.000Kg and 10 000Kg
- The highest success rates are between 2.000Kg and 5.500Kg.



Section 5

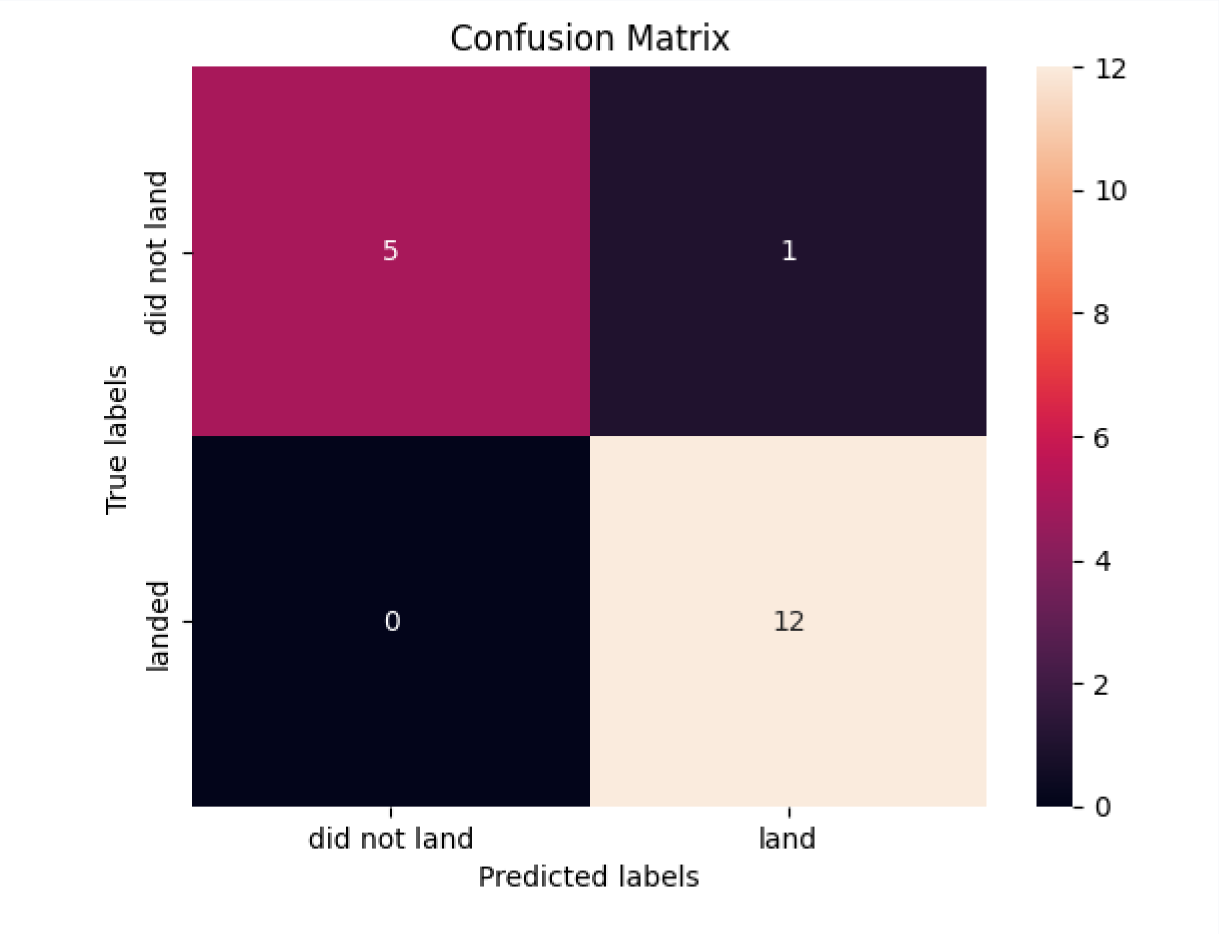
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree presente the highest accuracy

Confusion Matrix



The Confusion Matrix shows that the model only failed in one prediction, being a False Positive.

Conclusions

1.Model Performance:

The Decision Tree model showed the highest accuracy for predicting Falcon 9 landing success, with minimal misclassification (only one false positive, as seen in the confusion matrix).

Logistic Regression and SVM also achieved high accuracy, but the Decision Tree excelled in terms of overall prediction power for this dataset.

2.Key Influencing Factors:

Payload mass, launch site, and orbit type were crucial factors influencing launch success. Launches with payloads between 2,000 kg and 5,500 kg had the highest success rates.

KSC LC-39A had the highest launch success rate among SpaceX sites, followed by CCAFS SLC-40. VLEO orbits were associated with higher payloads and a better success rate.

3.Geographic Insights:

Most of the launch sites are located near the equator and the coastlines, leveraging Earth's rotational speed for orbital mechanics and minimizing the risk of debris falling in populated areas.

Proximity to strategic points like railways, highways, and cities plays a significant role in facilitating logistics and ensuring public safety.

Conclusions

4. Trends Over Time:

    The success rate of launches has improved steadily over time, with a notable spike in 2013. The increased success rate reflects both technological advancements in booster landings and improvements in launch management.

    Orbits like ES-L1, GEO, and HEO demonstrated a 100% success rate, which highlights their stability for specific mission types.

5. Implications for Competitors:

    By analyzing SpaceX's performance, it is clear that successful reusable booster landings significantly reduce costs. Competitors can focus on improving payload management and orbit-specific success strategies to remain competitive.

Appendix

Data Samples:

- CSV exports of the cleaned and labeled dataset, ready for machine learning model training and validation.
- Tables showing key descriptive statistics like payload mass, success rates by orbit, and flight number progression across the years.

Useful Links:

Full Project GITHUB

Falcon9 Wikipedia Page

SpaceX API

Thank you!