IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Lennart Kuhse
Date: 20 November 2024

Outline

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

Executive Summary

Summary of Methodologies

- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Executive Summary

Types of results:

- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Key Findings:

- No clear correlation was found between higher payload masses and lower success rates
- The SVM-Predictive-Models reaches the highest accuracy on the given data
- The "CCAFS LC-40" launch site has the highest success rate

Introduction

Project background and context

This capstone project focuses on predicting the successful landing of SpaceX’s Falcon 9 first stage. SpaceX significantly reduces launch costs compared to competitors by reusing the rocket’s first stage. Accurately forecasting landing success can help estimate launch costs and offer strategic insights to important factors in this process.

Questions to be answered

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- Which machine learning model delivers the most accurate predictions of landing success?

Section 1

Methodology

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 data collection methods to get complete information about the launches for a more detailed analysis.

Data Columns are obtained by using SpaceX REST API:

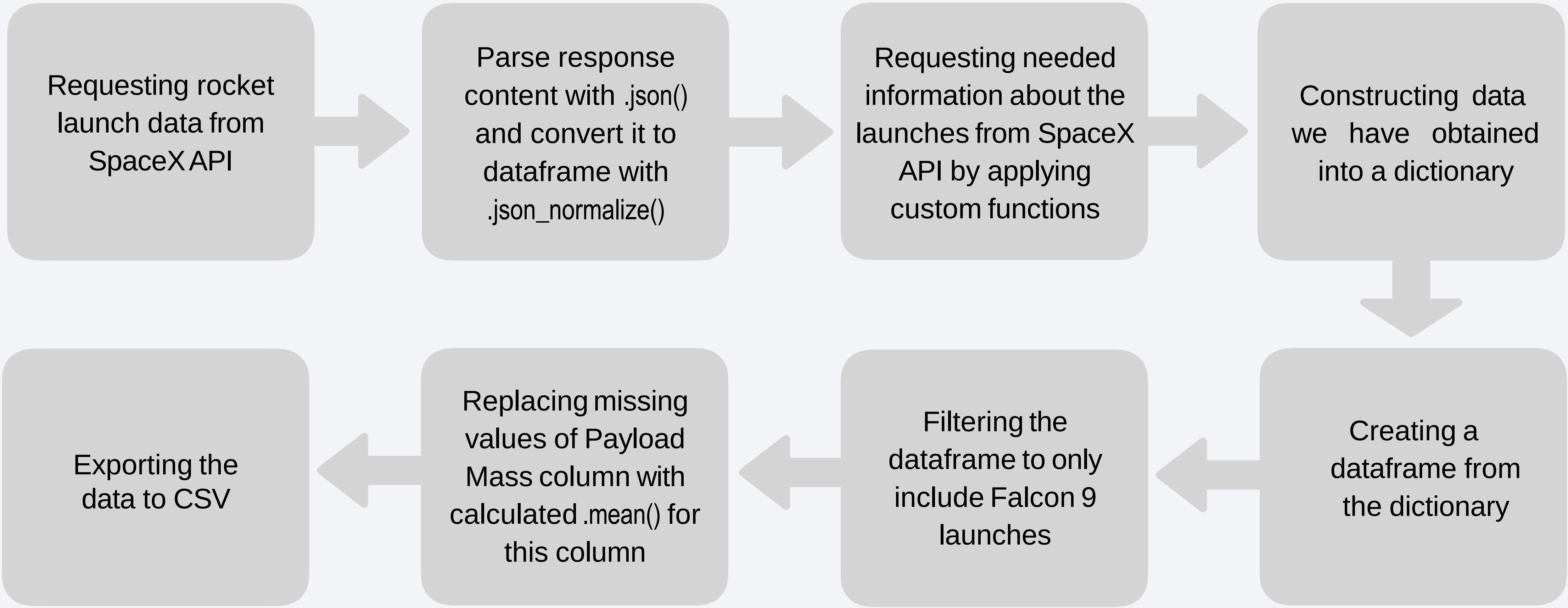
FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

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

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P1_jupyter-labs-spacex-data-collection-api_LK.ipynb

Data Collection – SpaceX API



Requesting rocket launch data from SpaceX API

Parse response content with .json() and convert it to dataframe with .json_normalize()

Requesting needed information about the launches from SpaceX API by applying custom functions

Constructing data we have obtained into a dictionary

Creating a dataframe from the dictionary

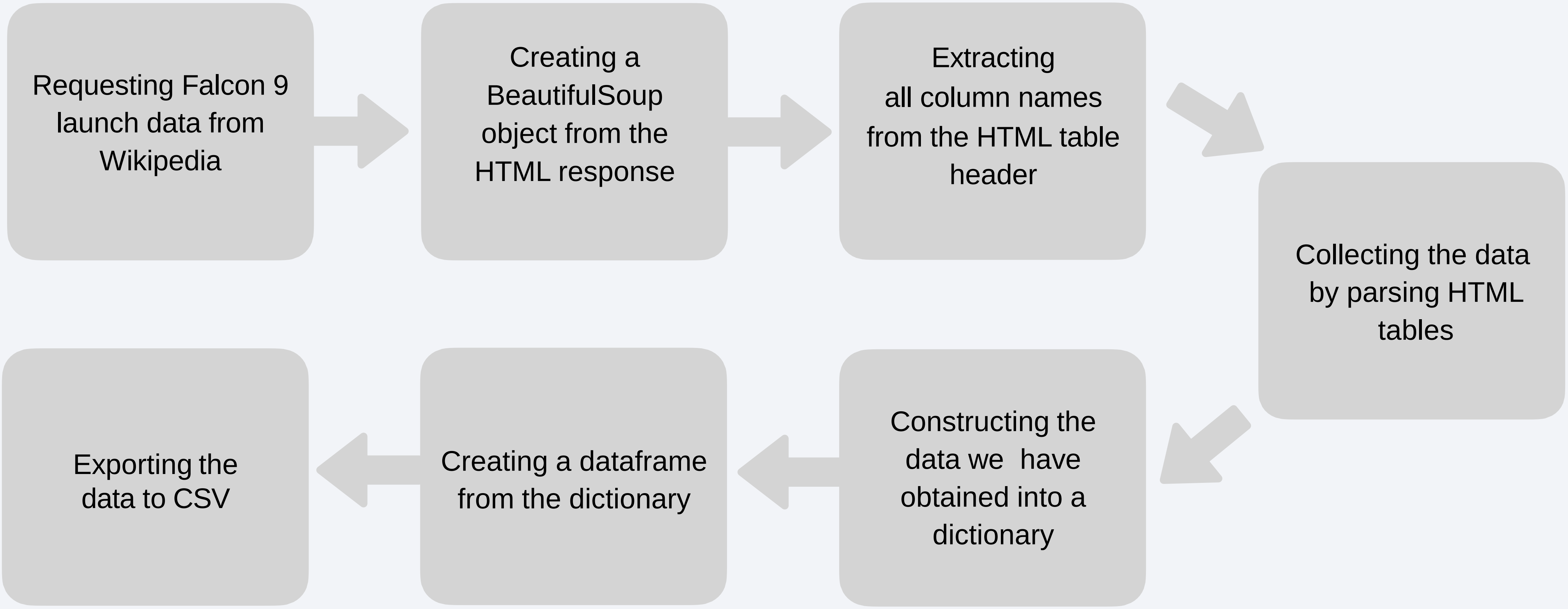
Filtering the dataframe to only include Falcon 9 launches

Replacing missing values of Payload Mass column with calculated .mean() for this column

Exporting the data to CSV

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P1_jupyter-labs-spacex-data-collection-api_LK.ipynb

Data Collection – Web Scraping



Requesting Falcon 9 launch data from Wikipedia

Creating a BeautifulSoup object from the HTML response

Extracting all column names from the HTML table header

Collecting the data by parsing HTML tables

Constructing the data we have obtained into a dictionary

Creating a dataframe from the dictionary

Exporting the data to CSV

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P2_jupyter-labs-webscraping_LK.ipynb

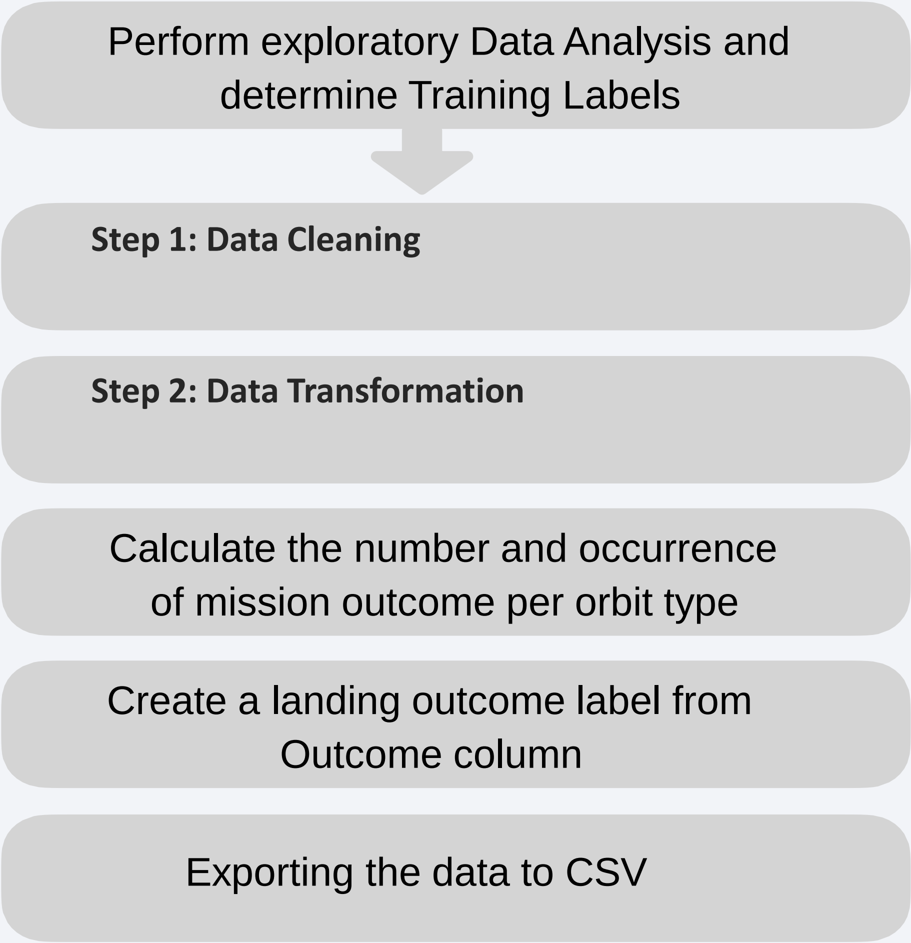
Data Wrangling

Data wrangling transforms raw, unorganized data into a structured and analyzable format through a systematic process. The dataset contains several cases where the booster did not land successfully. Some missions attempted a landing but failed due to accidents. For example:

- True Ocean: Successful landing in a specific ocean region.
- False Ocean: Unsuccessful landing in a specific ocean region.
- True RTLS: Successful landing on a ground pad.
- False RTLS: Unsuccessful landing on a ground pad.
- True ASDS: Successful landing on a drone ship.
- False ASDS: Unsuccessful landing on a drone ship.

For training labels, we convert these outcomes into binary values:

- 1: Successful booster landing.
- 0: Unsuccessful booster landing.



Perform exploratory Data Analysis and determine Training Labels

Step 1: Data Cleaning

Step 2: Data Transformation

Calculate the number and occurrence of mission outcome per orbit type

Create a landing outcome label from Outcome column

Exporting the data to CSV

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P3_labs-jupyter-spacex-Data%20wrangling_LK.ipynb

Data Wrangling

Overview

Data wrangling transforms raw, unorganized data into a structured and analyzable format through a systematic process.

Step 1: Data Cleaning

- Handle Missing Values
- Ensure Data Integrity

Step 2: Data Transformation

- Convert Data Types
- Standardize Text
- Feature Engineering
- Normalize and Scale

Step 3: Data Integration

- Merge Datasets
- Ensure Consistency

Step 4: Data Validation

- Eliminate Duplicates
- Verify Accuracy

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P3_labs-jupyter-spacex-Data%20wrangling_LK.ipynb

EDA with Data Visualization

Overview: Exploratory Data Analysis (EDA)
EDA involves the use of visualizations and summary statistics to explore a dataset's core attributes. The aim is to understand variable distributions, detect patterns, and identify relationships within the data.

Scatter plots show the relationship between variables. If a relationship exists, they could be used in machine learning model.

Bar charts show comparisons among discrete categories. The goal is to show the relationship between the specific categories being compared and a measured value.

Heatmaps depict correlation matrices among numerical variables. They highlight strong positive or negative correlations, aiding in feature selection and understanding multicollinearity.

Line charts show trends in data over time (time series).

Charts plotted

Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

EDA with SQL

Performed SQL queries:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Build an Interactive Map with Folium

Interactive Map Features with Folium

- **Markers**: Indicate SpaceX launch locations for precise geographic reference.
- **Circles**: Highlight proximity zones around launch sites, illustrating potential impact areas or operational boundaries.
- **Lines**: Connect launch sites to relevant locations, providing spatial context and visualizing relationships.

These features enhance user understanding of launch site geography, safety zones, and spatial connections.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P6_SpaceX_Interactive_Visual_Analytics_Folium_LK.ipynb

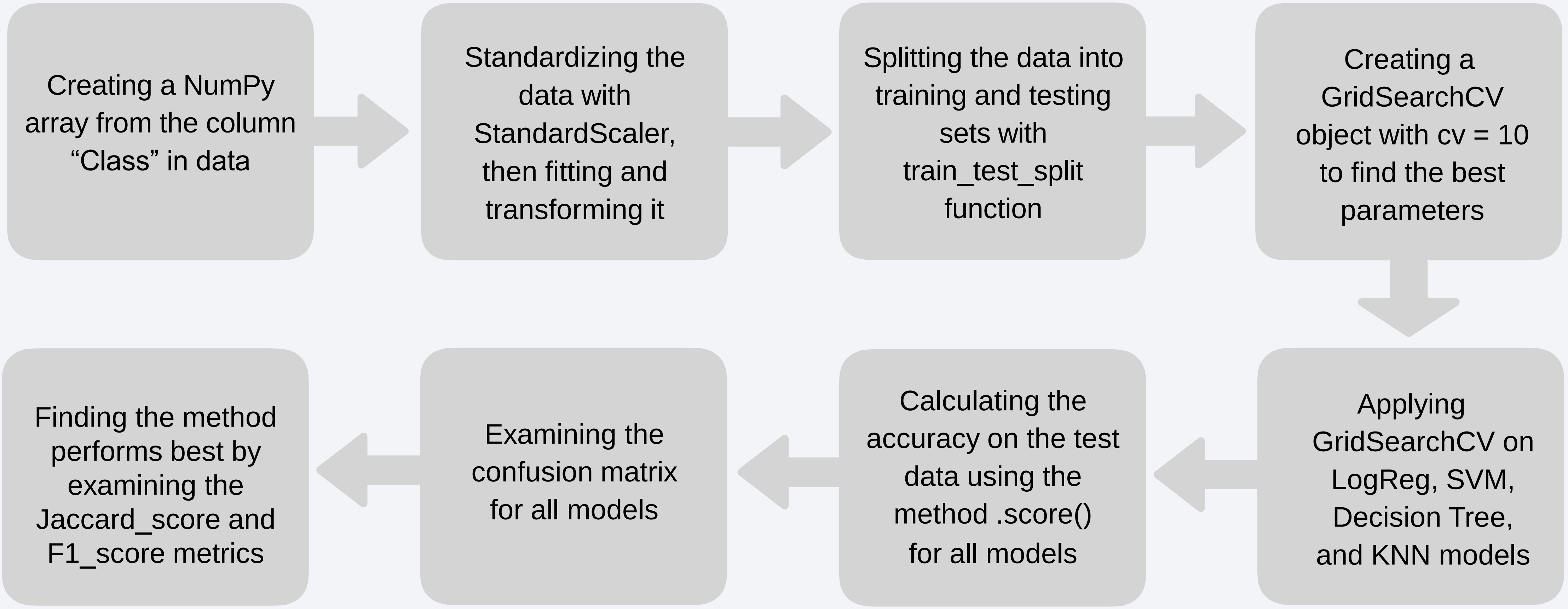
Build a Dashboard with Plotly Dash

Visualizations and Interactive Dashboard Features

- **Launch Site Dropdown**: Filters data by specific launch sites for targeted analysis.
- **Success Pie-Chart**: Visualizes the proportion of successful and failed launches to highlight overall performance trends.
- **Success Payload Scatter-Plot**: Illustrates the relationship between payload mass and launch success, showing its impact on outcomes.
- **Payload Range Slider**: Adjusts payload mass ranges dynamically to explore success rates across different payloads.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/spacex-dash-code.py

Predictive Analysis (Classification)



Creating a NumPy array from the column “Class” in data

Standardizing the data with StandardScaler, then fitting and transforming it

Splitting the data into training and testing sets with train_test_split function

Creating a GridSearchCV object with cv = 10 to find the best parameters

Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models

Calculating the accuracy on the test data using the method .score() for all models

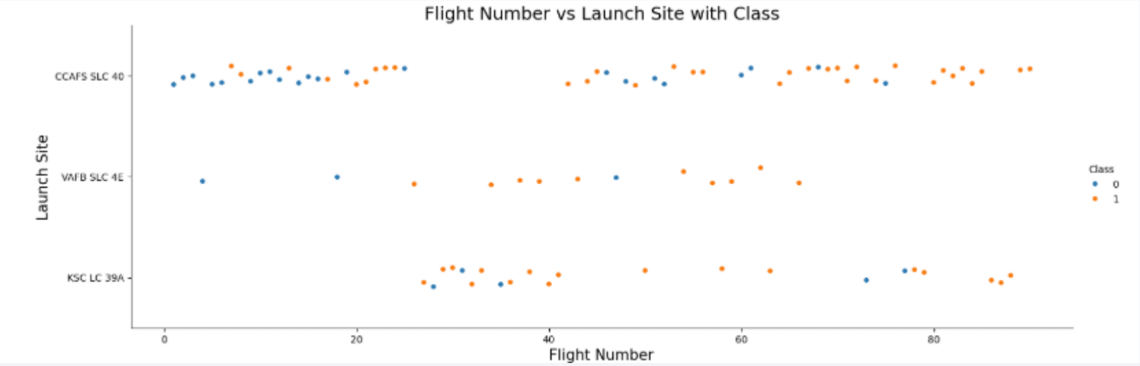
Examining the confusion matrix for all models

Finding the method performs best by examining the Jaccard_score and F1_score metrics

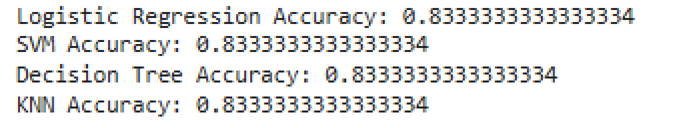
GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P7_SpaceX_Machine%20Learning%20Prediction_Part_5_LK.ipynb

Results

Exploratory data analysis results

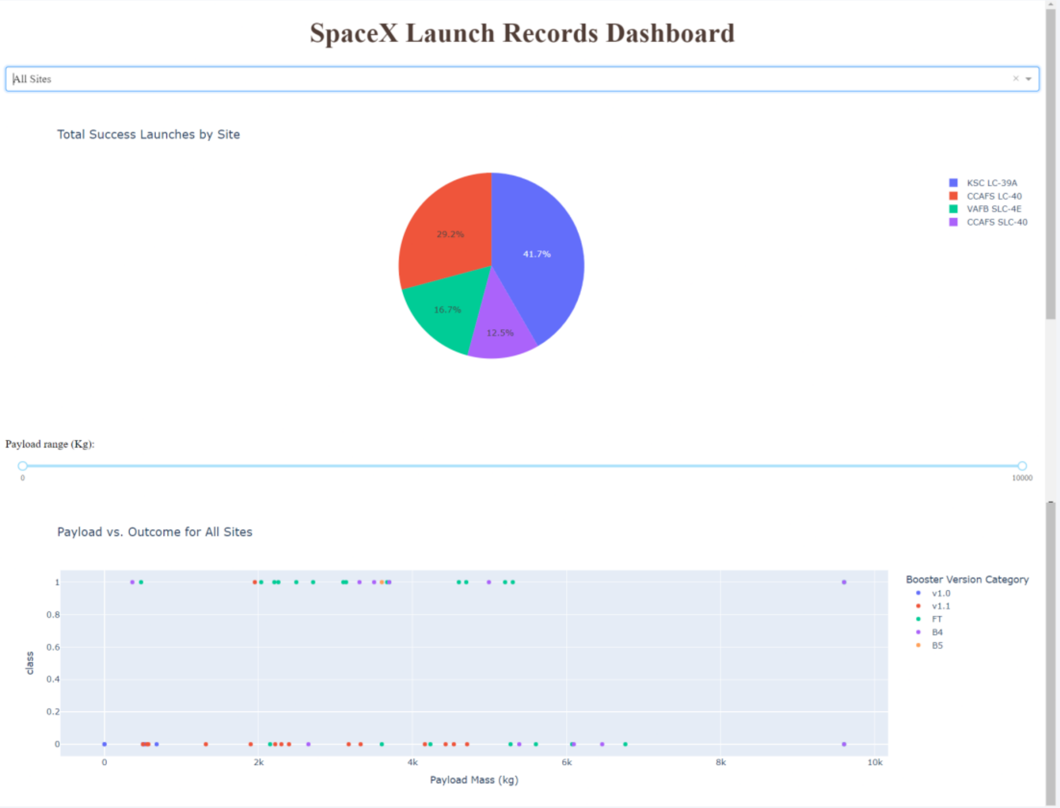


Predictive analysis results



```
Logistic Regression Accuracy: 0.8333333333333334
SVM Accuracy: 0.8333333333333334
Decision Tree Accuracy: 0.8333333333333334
KNN Accuracy: 0.8333333333333334
```

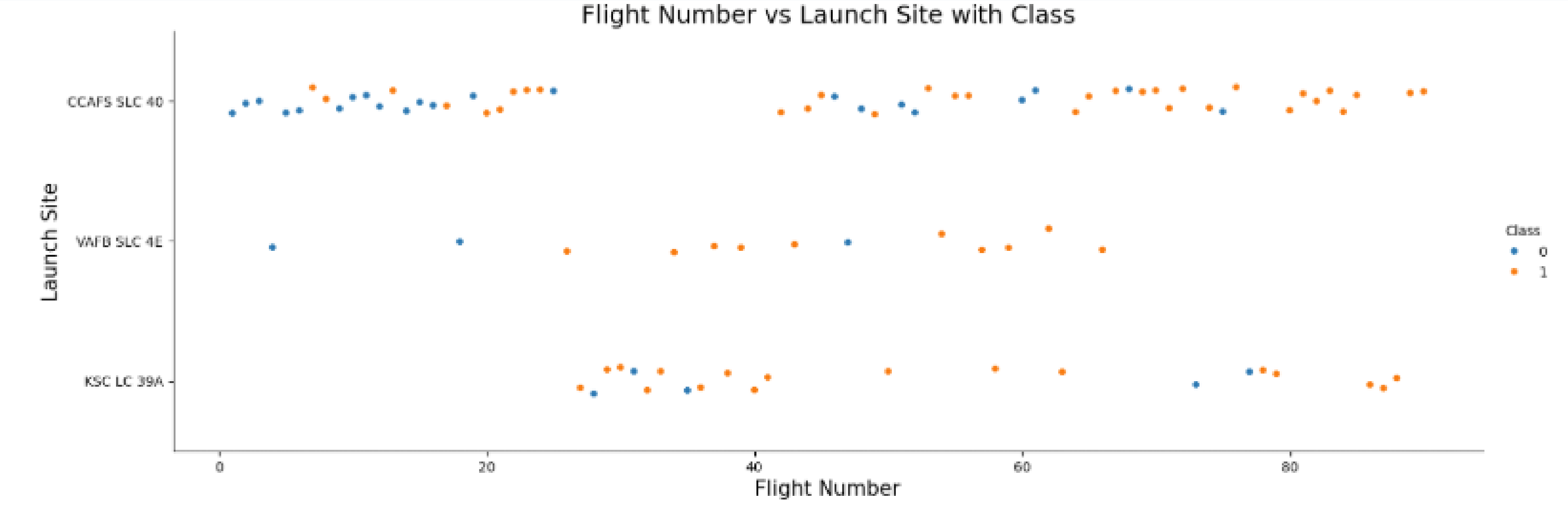
Interactive analytics demo



Section 2

Insights drawn from EDA

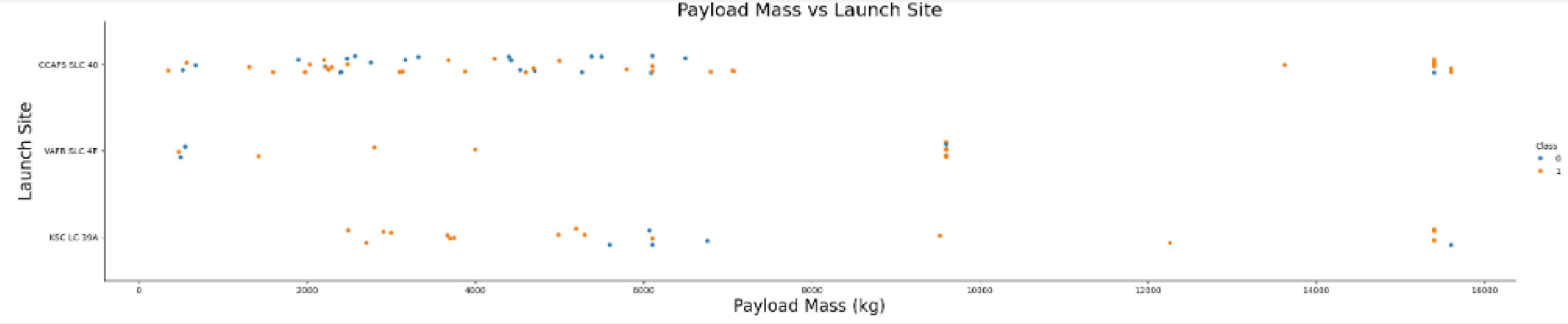
Flight Number vs. Launch Site



- **Mixed Results at Key Launch Sites:** Both CCAFS SLC 40 and KSC LC 39A show a blend of successful (orange) and unsuccessful (blue) landings, suggesting that variables beyond the launch site may affect landing outcomes.
- **Consistent Launch Activity:** Launches are distributed across various flight numbers at all sites, indicating steady operations with a visible trend in improved landing success over time.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

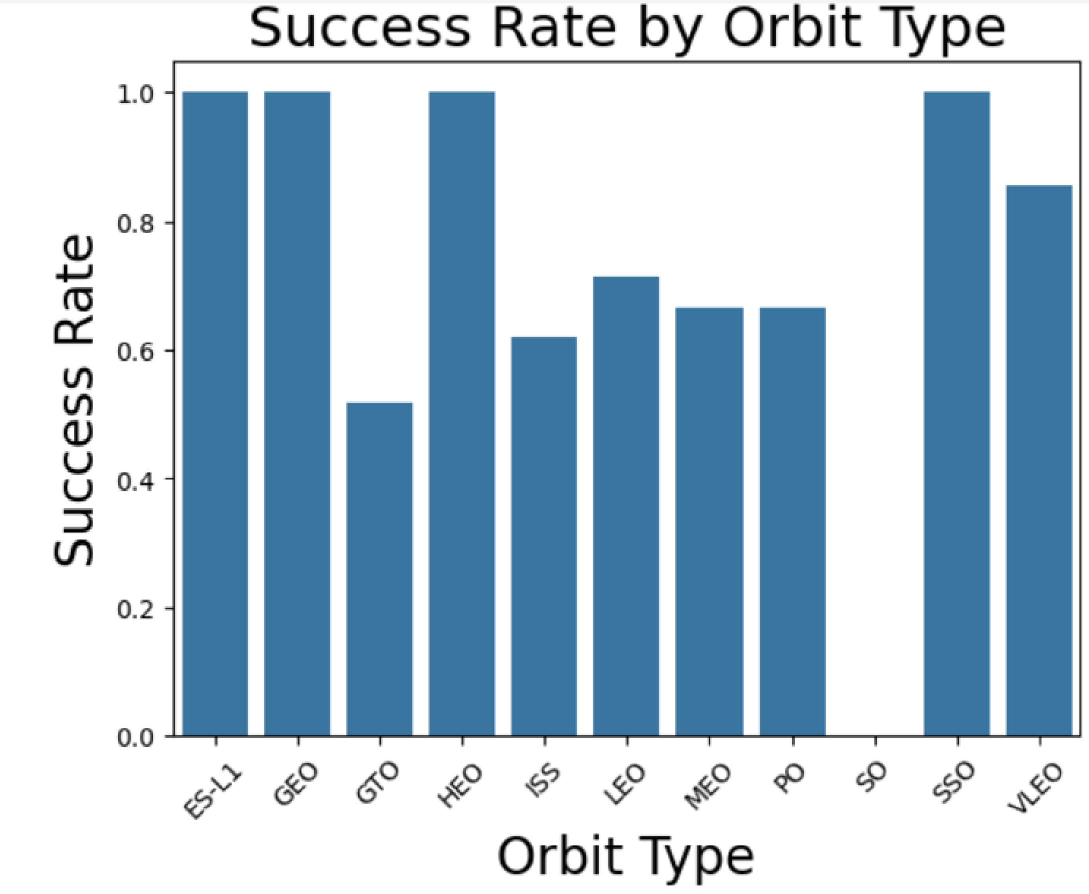
Payload vs. Launch Site



- **Payload Distribution:** Most launches from CCAFS SLC 40 carry payloads under 10,000 kg, while VAFB SLC 4E and KSC LC 39A accommodate a broader range of payload masses, reflecting diverse mission types.
- **Heavy Payload Launches:** KSC LC 39A is commonly used for high-capacity missions, frequently launching payloads exceeding 15,000 kg, indicating its capability for heavy payload operations.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

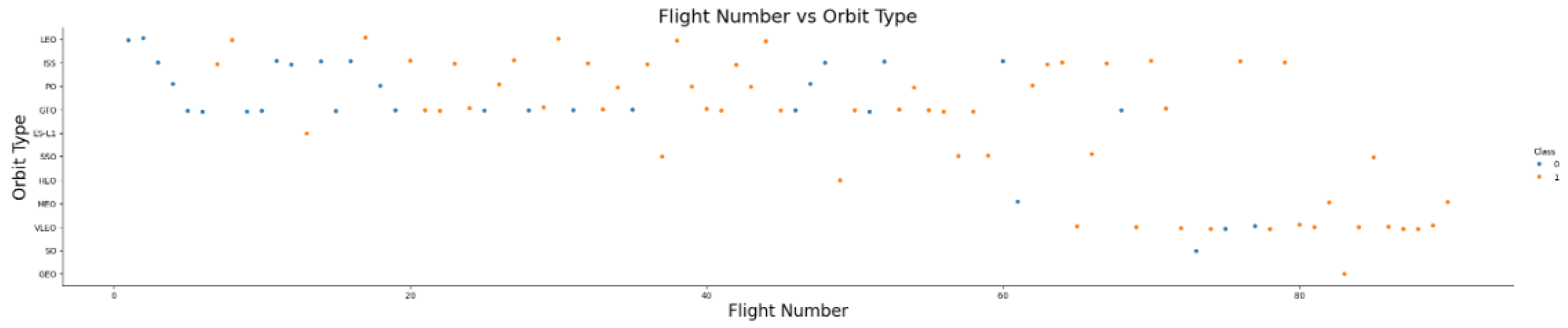
Success Rate vs. Orbit Type



- **High Success Rates**: Missions to VLEO, ES-L1, GEO, HEO, and SSO orbits have achieved a 100% success rate, demonstrating these orbits' reliability for successful first-stage landings.
- **Lower Success Rate for GTO**: GTO missions exhibit a notably lower success rate, indicating potential challenges or complexities associated with landing in this orbit.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

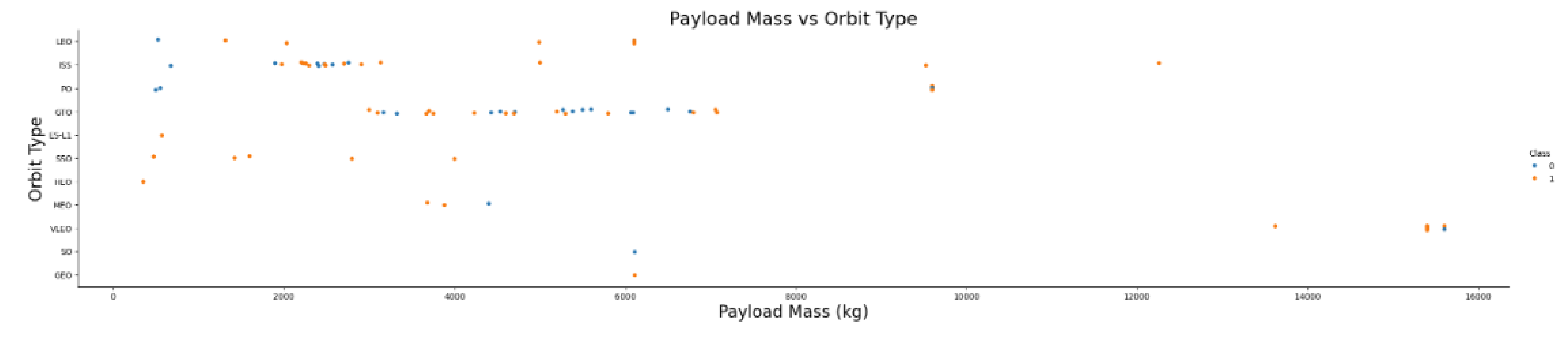
Flight Number vs. Orbit Type



- **Improved Success Over Time:** Falcon 9’s success rate increases with higher flight numbers, highlighting the impact of experience and continuous improvements.
- **Orbit-Specific Performance**: Early GTO and ISS missions showed mixed results, but recent missions to these orbits exhibit higher success rates, reflecting enhanced mission planning and execution.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

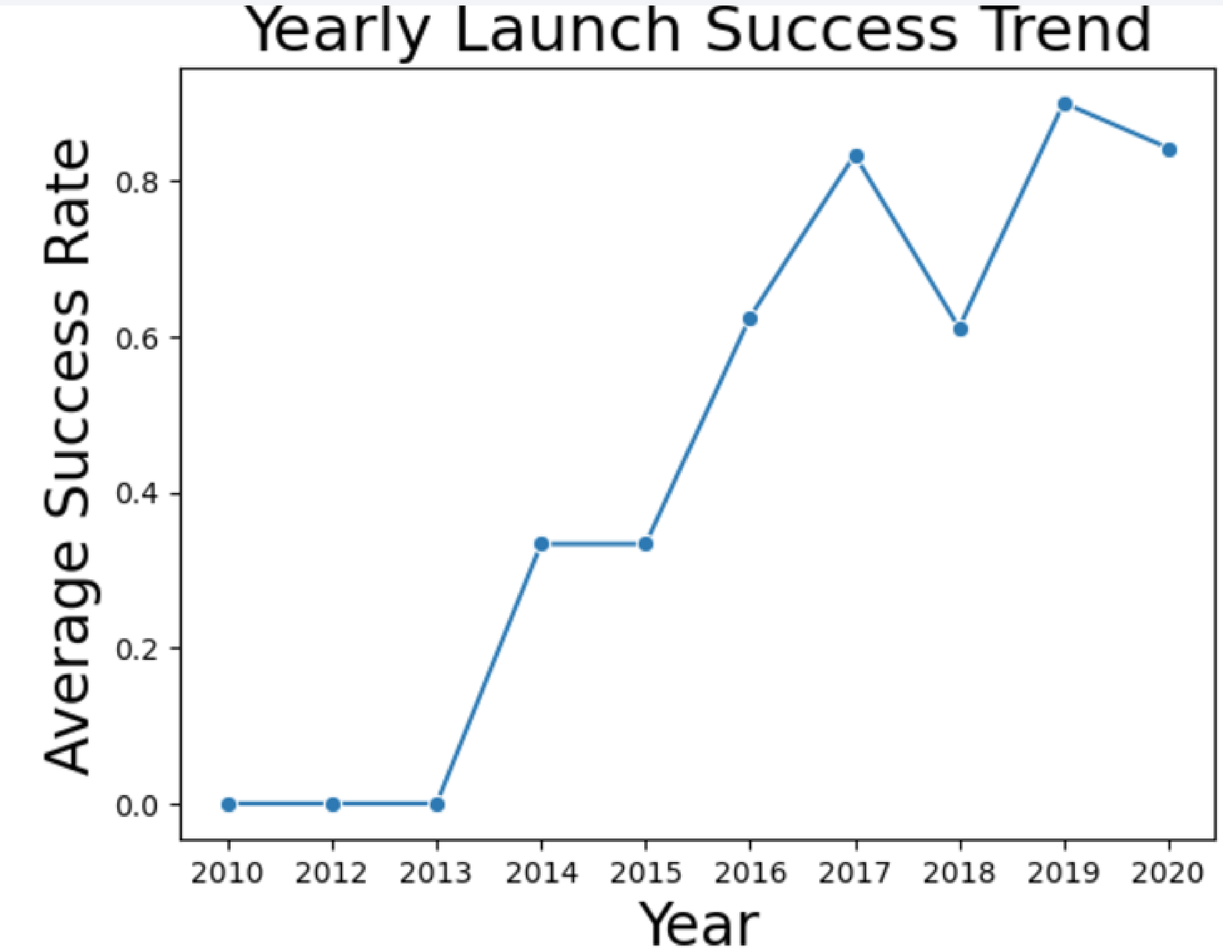
Payload vs. Orbit Type



- **Higher Success for Lighter Payloads**: Successful landings are more common across all orbit types, particularly for payloads under 6000 kg.
- **Challenges with Heavier Payloads**: Payloads exceeding 10,000 kg show a mix of successful and failed landings, suggesting greater difficulty with larger payloads.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

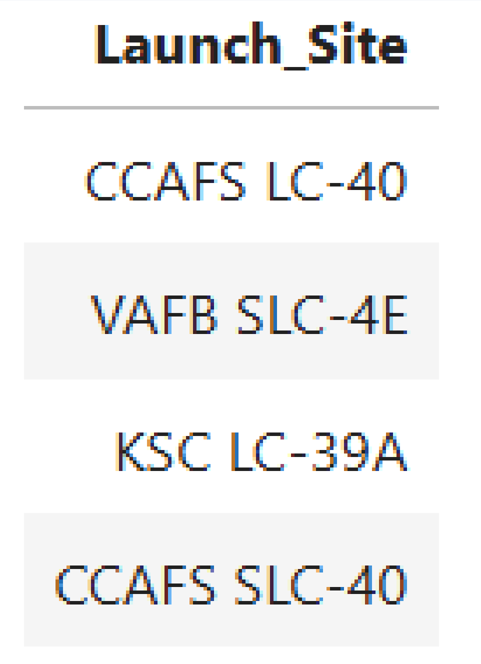
Launch Success Yearly Trend



- **Improved Success Over Time**: The annual success rate of Falcon 9 launches has notably increased since 2013, surpassing 80% by 2020.
- **Temporary Dip in 2018**: Although there was a decline in 2018, the overall trend reflects growing reliability and success in subsequent years.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P5_edadataviz_LK.ipynb

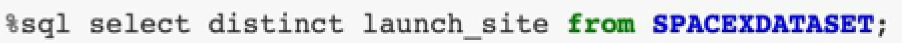
All Launch Site Names



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

List of all unique launch sites of SpaceX rockets.

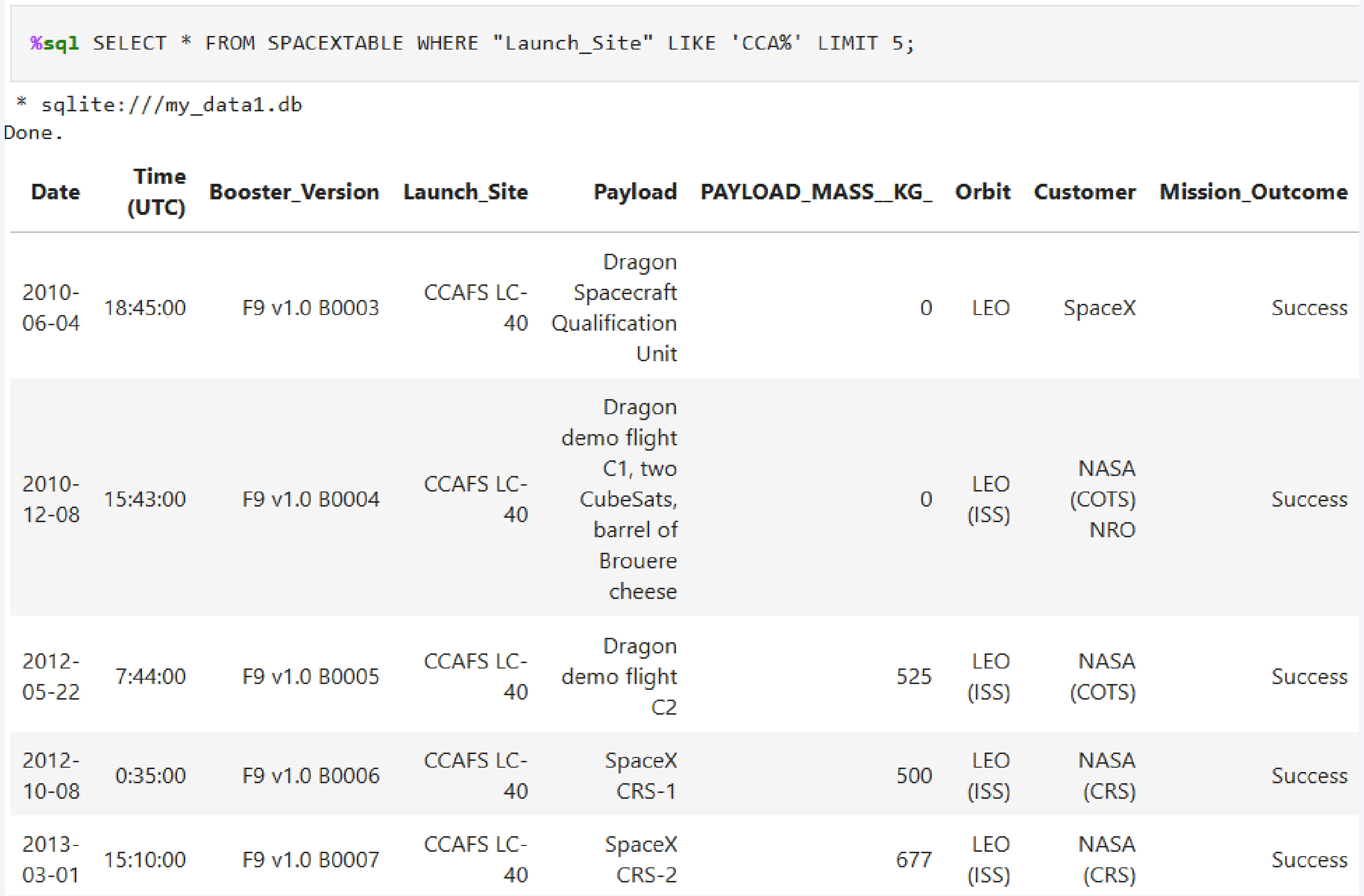
Listed through the following SQL query:



```
%sql select distinct launch_site from SPACEXDATASET;
```

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Launch Site Names Begin with 'CCA'



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

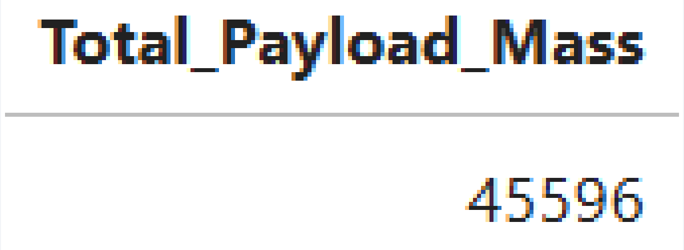
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Using a SQL query to list all launch sites beginning with the letters “CCA”.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

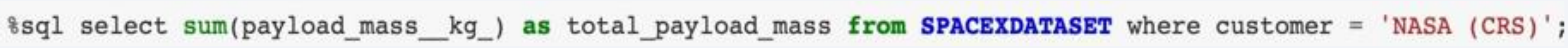
Total Payload Mass



Total_Payload_Mass
45596

As show in the image, the total payload mass for the customer “NASA (CRS)” is 45596kg

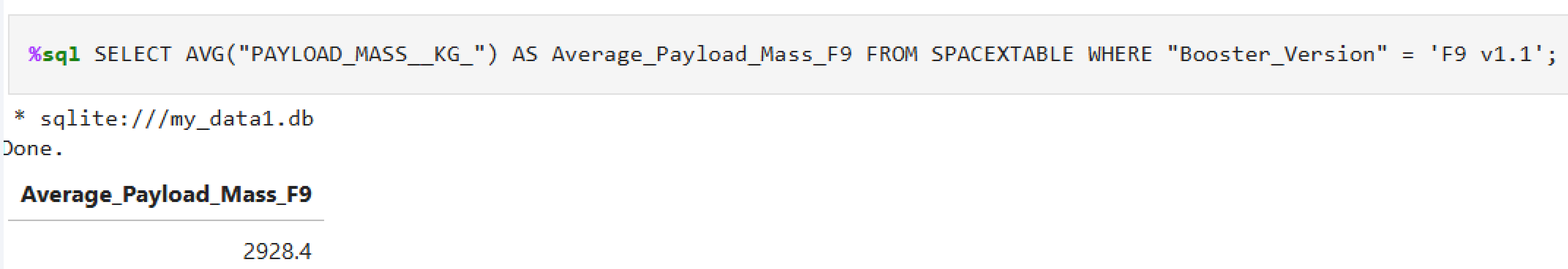
The image is created by the following SQL query:



```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA (CRS)';
```

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Average Payload Mass by F9 v1.1



```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass_F9 FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

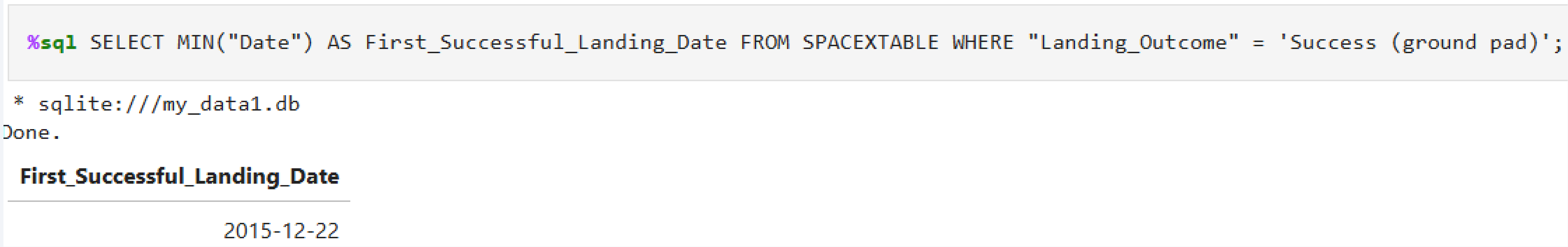
```
 * sqlite:///my_data1.db
Done.
```

Average_Payload_Mass_F9
2928.4

As shown in the image, the average payload of the F9 v1.1 booster is 2928.4kg

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

First Successful Ground Landing Date



```
%sql SELECT MIN("Date") AS First_Successful_Landing_Date FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

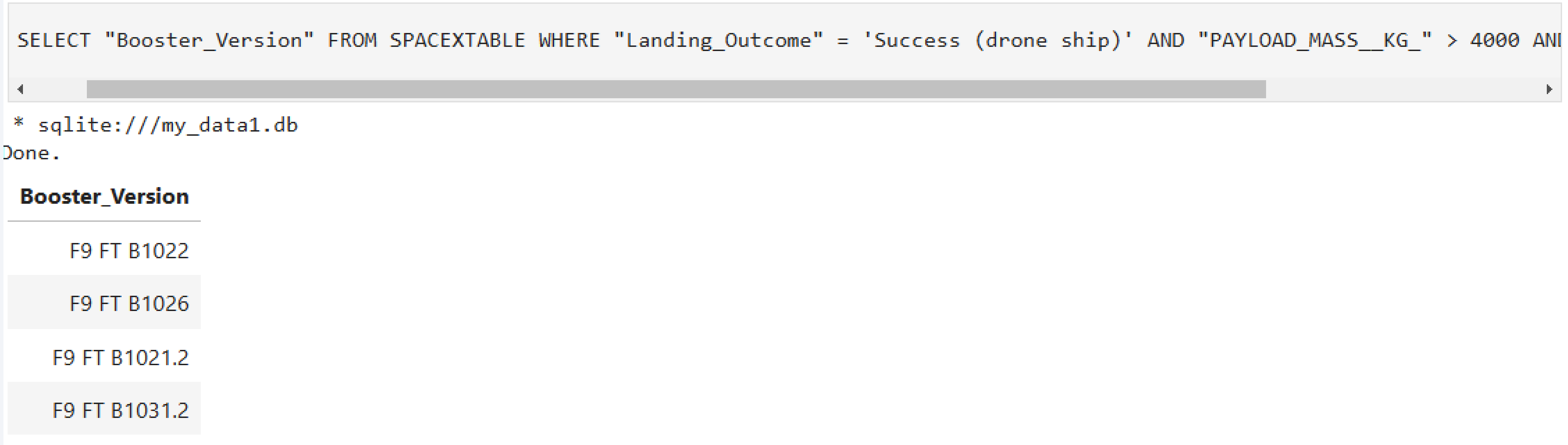
```
 * sqlite:///my_data1.db
Done.
```

First_Successful_Landing_Date
2015-12-22

As shown in the image, the first successful ground landing was achieved 2015-12-22.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Successful Drone Ship Landing with Payload between 4000 and 6000



```
SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 ANI
```

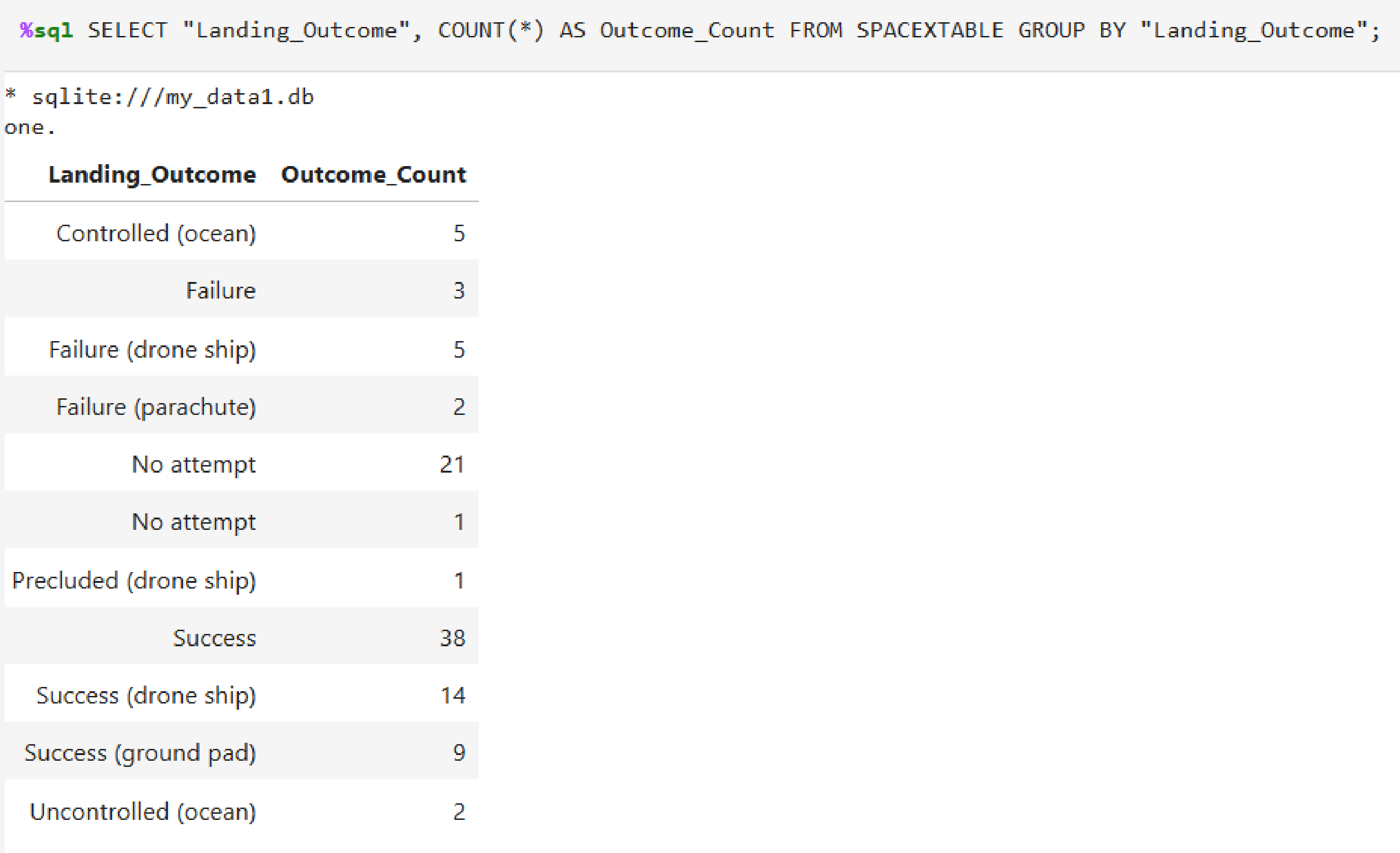
```
 * sqlite:///my_data1.db
Done.
```

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

The image shows the names of the boosters which were successful in drone ship landing and had a payload mass greater than 4000kg but less than 6000kg.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Total Number of Successful and Failure Mission Outcomes



```
%sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE GROUP BY "Landing_Outcome";
```

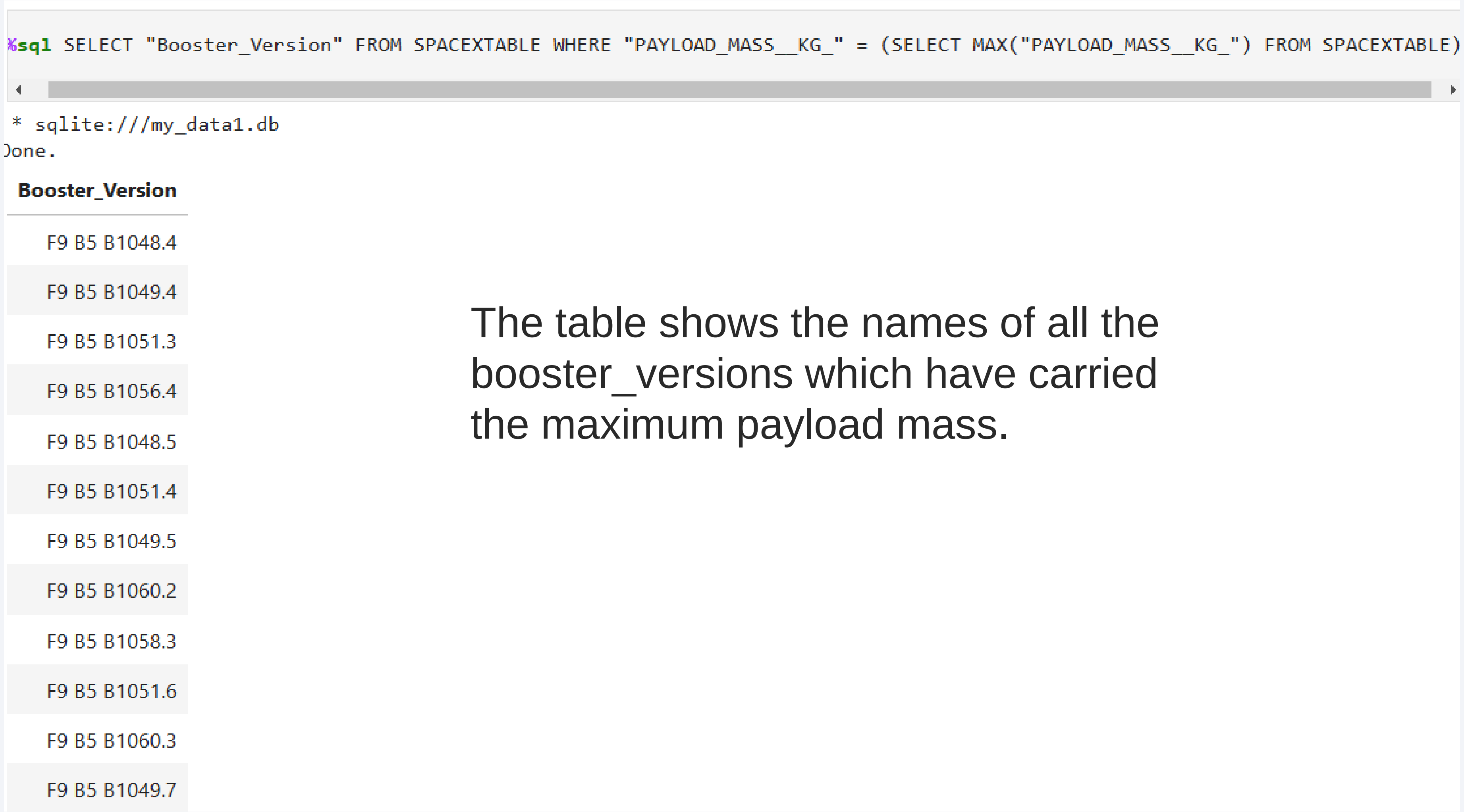
```
* sqlite:///my_data1.db
one.
```

Landing_Outcome	Outcome_Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

The table created by SQL query shows the total number of successful and failure mission outcomes.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

Boosters Carried Maximum Payload



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

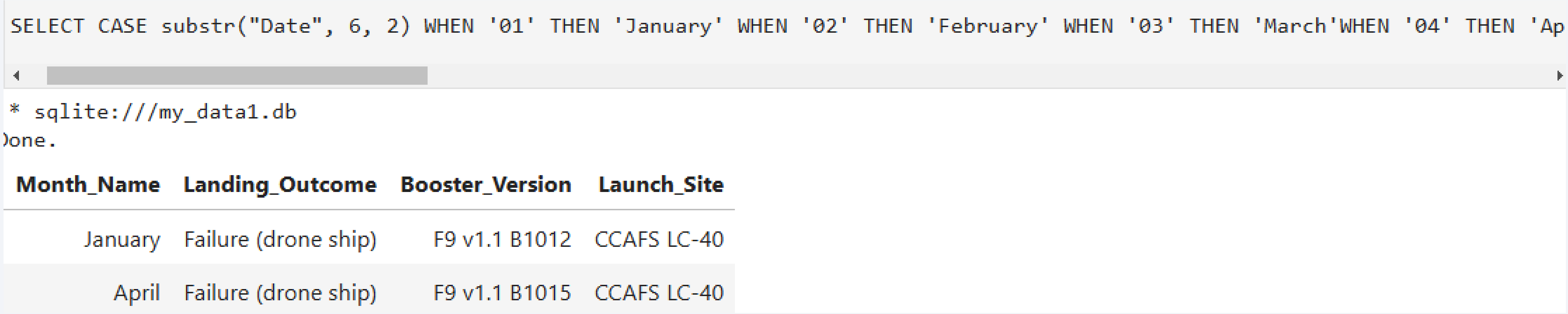
```
 * sqlite:///my_data1.db
Done.
```

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

The table shows the names of all the booster_versions which have carried the maximum payload mass.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

2015 Failed Launch Records



```
SELECT CASE substr("Date", 6, 2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March'WHEN '04' THEN 'Ap
```

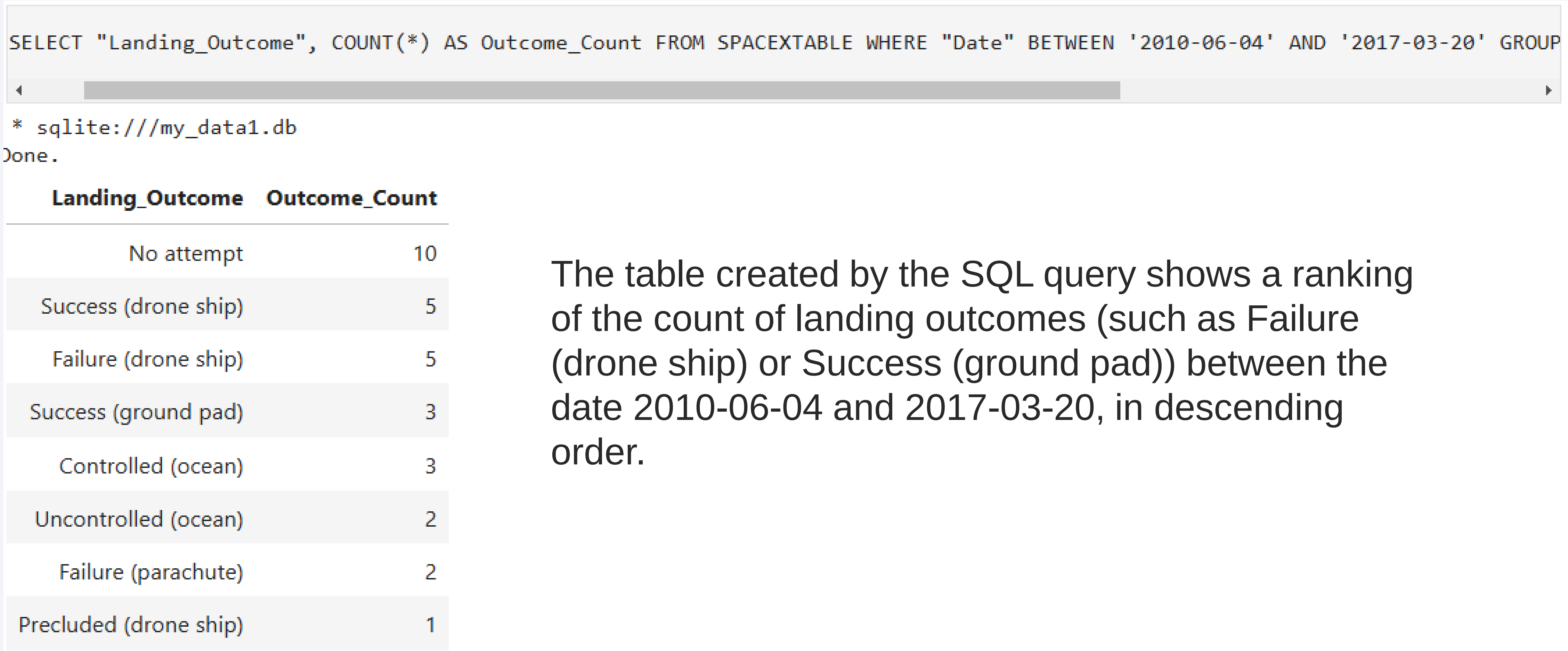
```
* sqlite:///my_data1.db
)one.
```

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

The table created by the SQL query shows the failed landing outcomes in drone ship with the respective booster versions and launch site names in year 2015.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

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



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

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

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

The table created by the SQL query shows a ranking of 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.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P4_jupyter-labs-eda-sql-coursera_sqllite_LK.ipynb

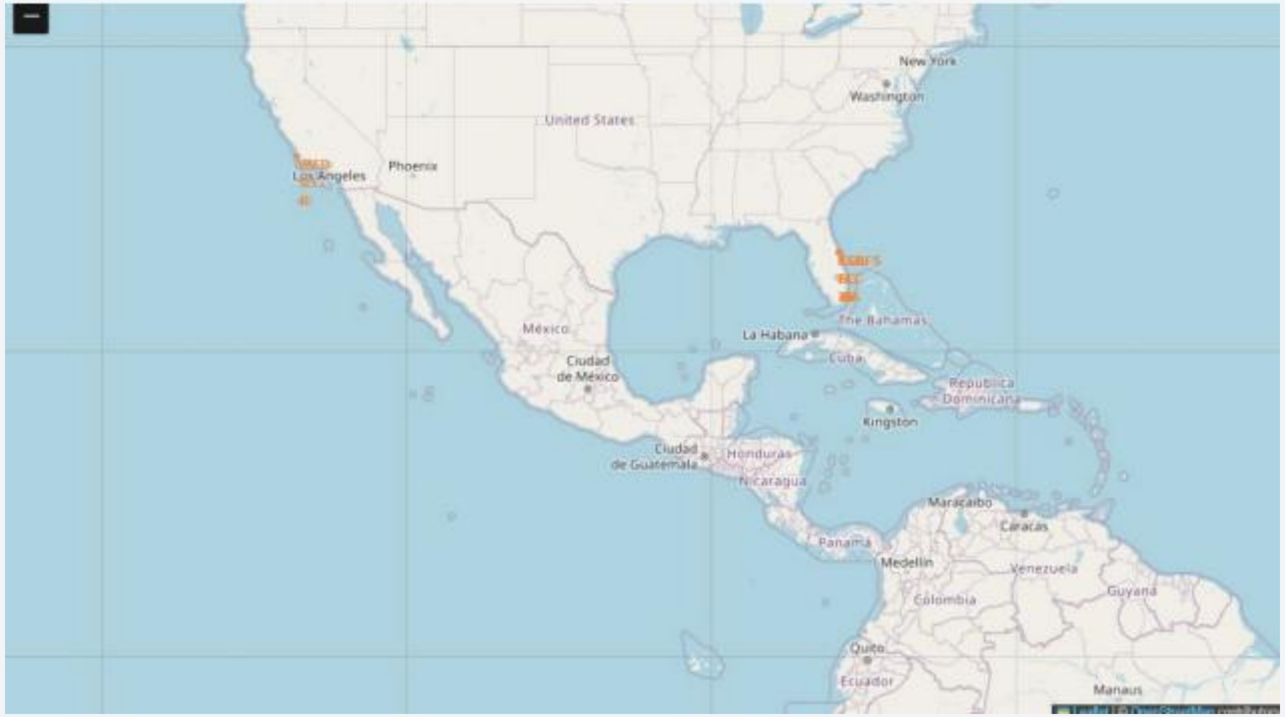
Section 3

Launch Sites Proximities Analysis

Launch Sites Overview

Proximity to the Equator:
Not all launch sites are near the Equator. For instance, Vandenberg Air Force Base (VAFB SLC-4E) is located at 34.63° latitude, further from the Equator than the Florida-based sites.

Proximity to the Coast: All launch sites are located near the coast. Sites in Florida (CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A) and VAFB SLC-4E in California are all coastal.

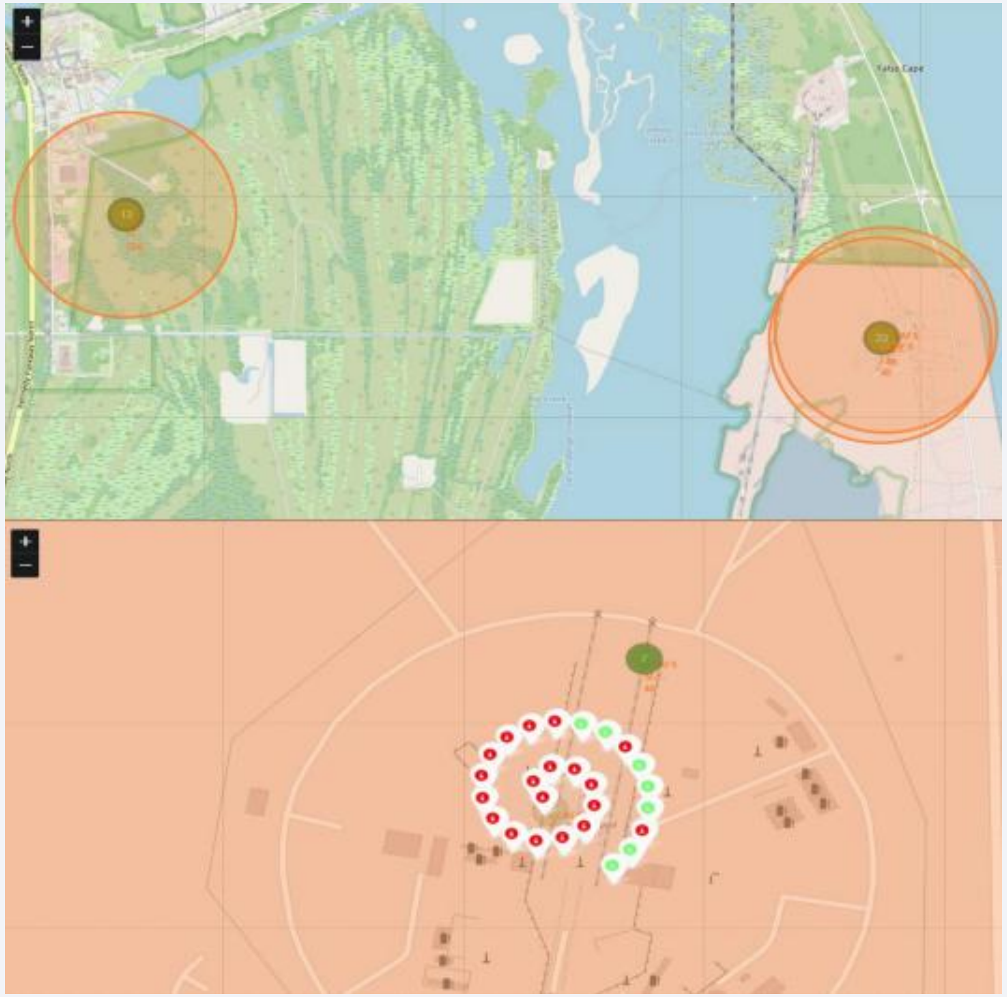


GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P6_SpaceX_Interactive_Visual_Analytics_Folium_LK.ipynb

Launch Outcomes Shown on Map

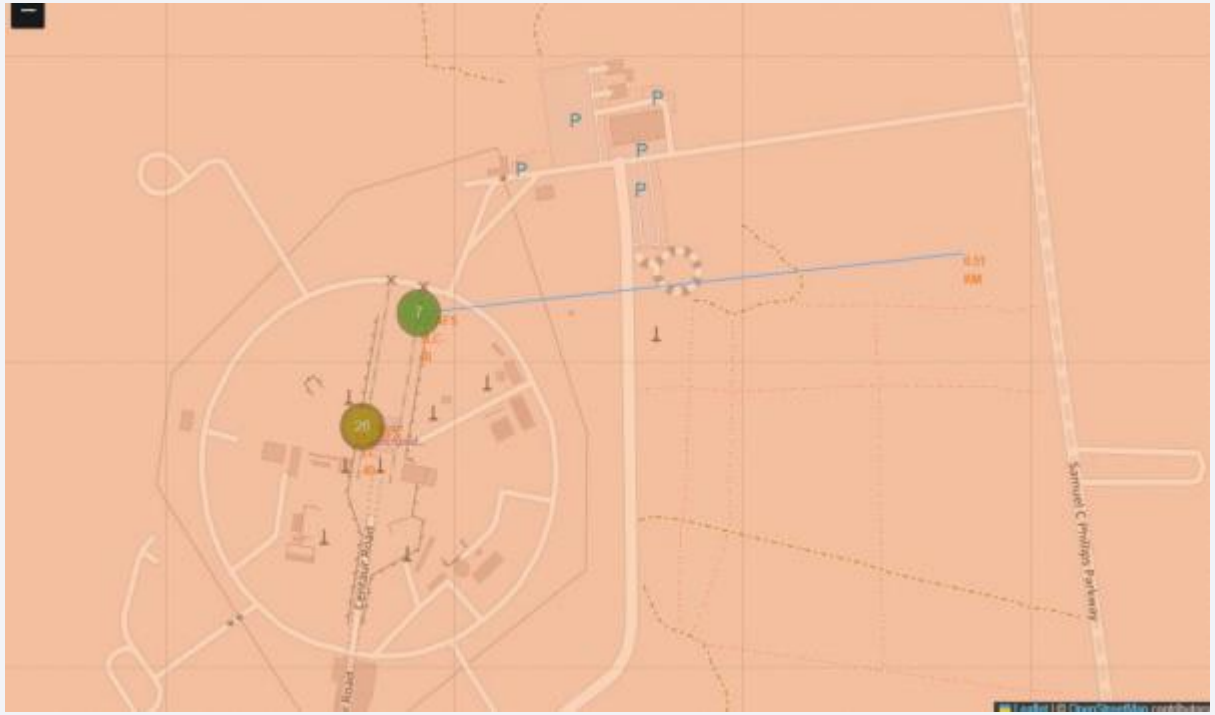
This enhanced visualization with clustered markers simplifies the exploration and analysis of SpaceX launch data. Clustering helps managing large number of markers and highlights patterns that may be overlooked in a less organized plot. By analyzing marker colors and popup details, you can gain valuable insights into launch characteristics and distribution.

For instance, at CCAFS LC-40, the color-coding of 26 launch sites shows 19 red markers (unsuccessful launches) and 7 green markers (successful launches), providing immediate visual feedback on launch performance at this site.



GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P6_SpaceX_Interactive_Visual_Analytics_Folium_LK.ipynb

Distance Between Launch Site & Railway Tracks

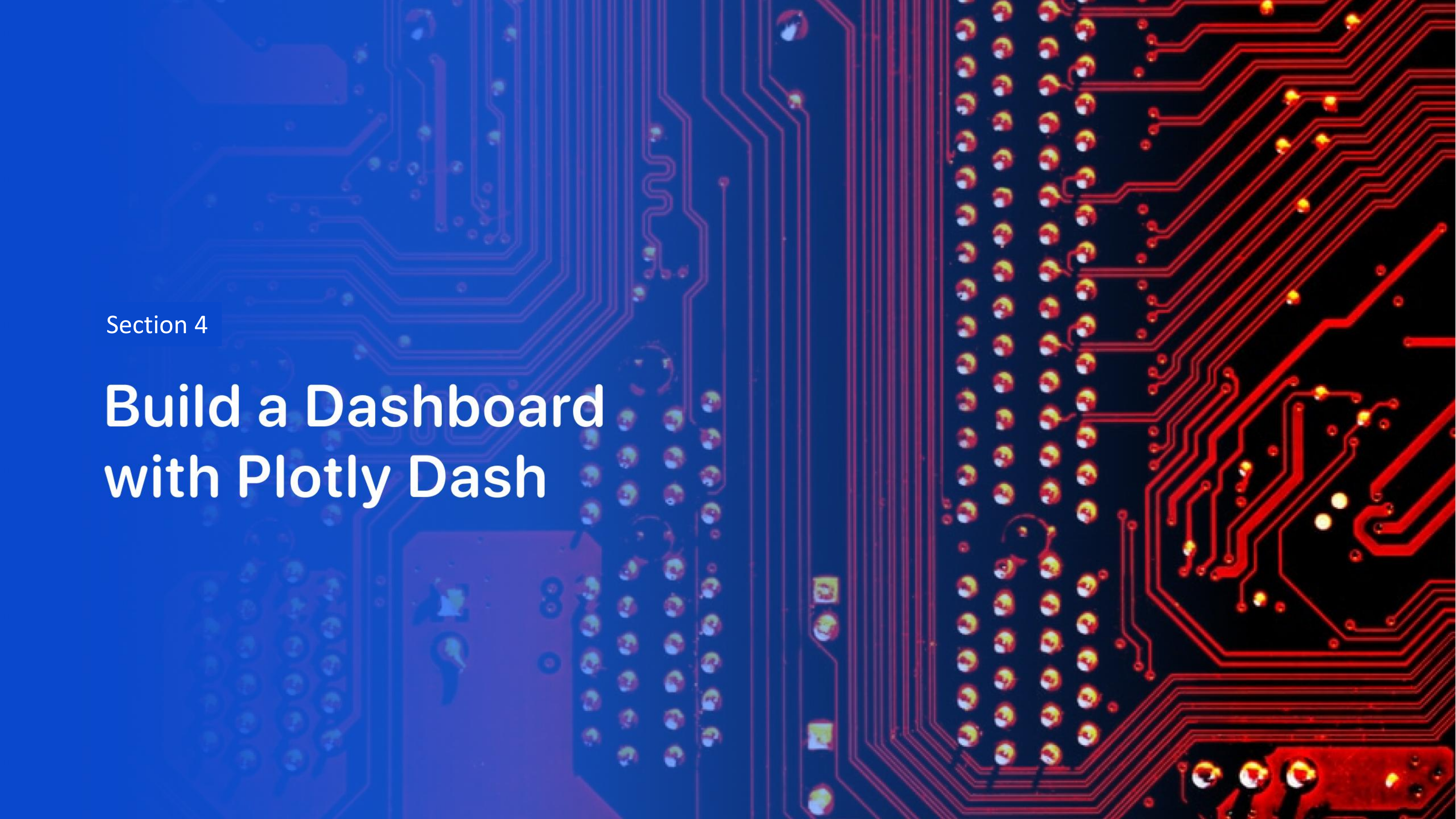


This plot visually represents the distance between the CCAFS SLC-40 launch site and the nearest coastline, approximately 0.51 kilometers. The PolyLine illustrates the direct distance, emphasizing the site's proximity to the railway tracks. The proximity allows for an easier transportation of goods to the launch site.

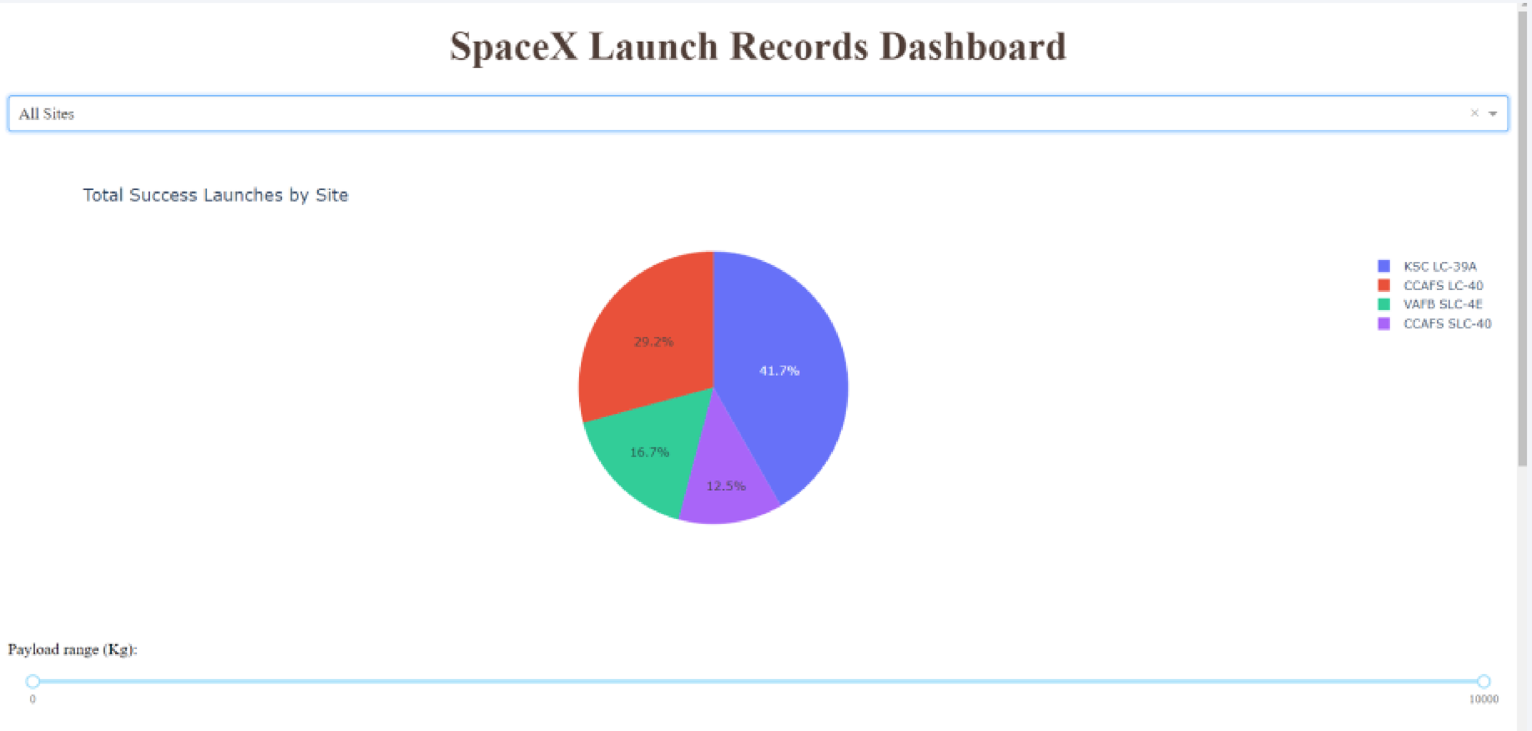
GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P6_SpaceX_Interactive_Visual_Analytics_Folium_LK.ipynb

Section 4

Build a Dashboard with Plotly Dash



Launch Success Count for all sites (in a pie chart)



SpaceX Launch Records Dashboard

All Sites

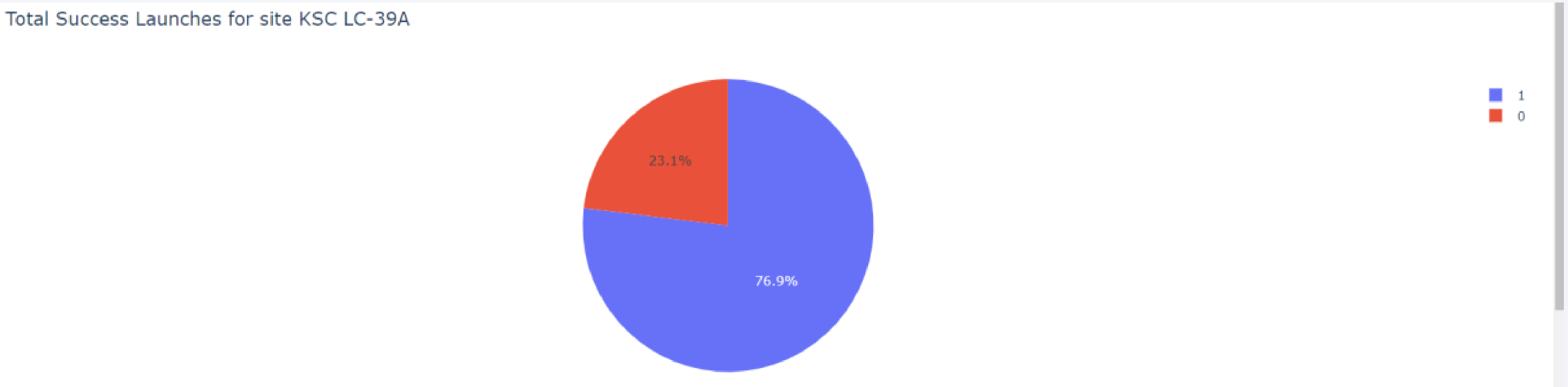
Payload range (Kg):

- CCAFS LC-40: 29.2%
- CCAFS SLC-40: 12.5%
- VAFB SLC-4E: 16.7%
- KSC LC-39A: 41.7%

The KSC LC-39A launch site has the highest number of successful launches, making up 41.7% of the total successes. This indicates that KSC LC-39A is a highly reliable site for SpaceX launches.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/spacex-dash-code.py

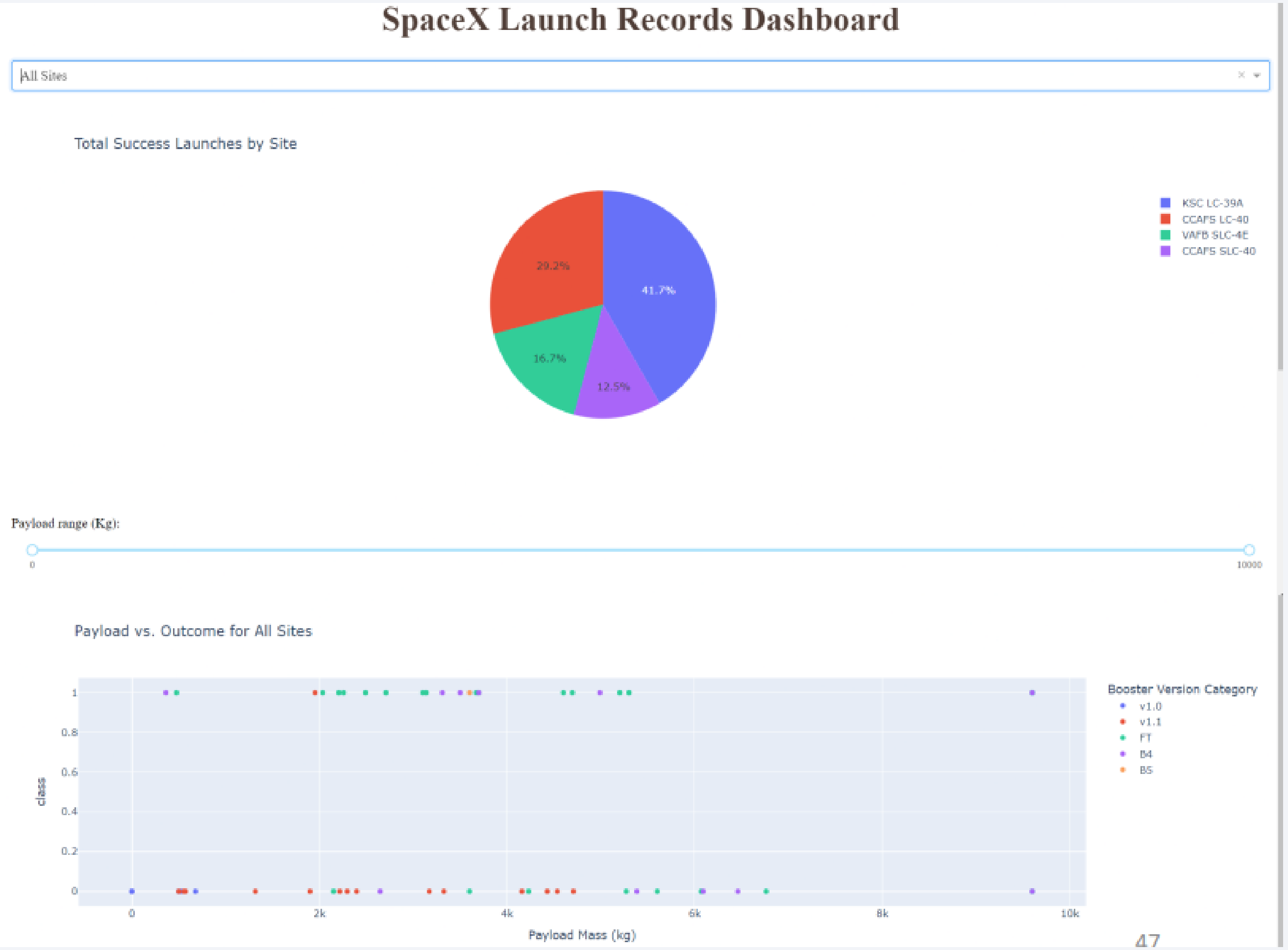
Launch success ratio by launch sites



The high success rate of launches from KSC LC-39A demonstrates its reliability and effectiveness as a launch site. With 76.9% of launches classified as successful (Class 1) and 23.1% as unsuccessful (Class 0), the data underscores the site's strong performance.

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/spacex-dash-code.py

Dashboard Overview



CCAFS LC-40 leads with the highest success rate at 43.7%, making it the most reliable launch site among those analyzed. Other sites, including KSC LC-39A, VAFB SLC-4E, and CCAFS SLC-40, show lower success rates, indicating variability in performance. Regarding booster versions, "FT" is used the most and demonstrates a high success rate across various payloads, while "v1.0" has fewer launches, requiring further analysis. No clear trend is observed between higher payload masses and lower success rates for different booster versions.

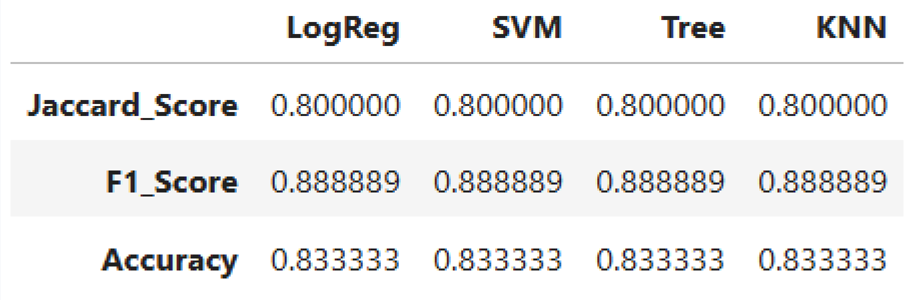
GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/spacex-dash-code.py

Section 5

Predictive Analysis (Classification)

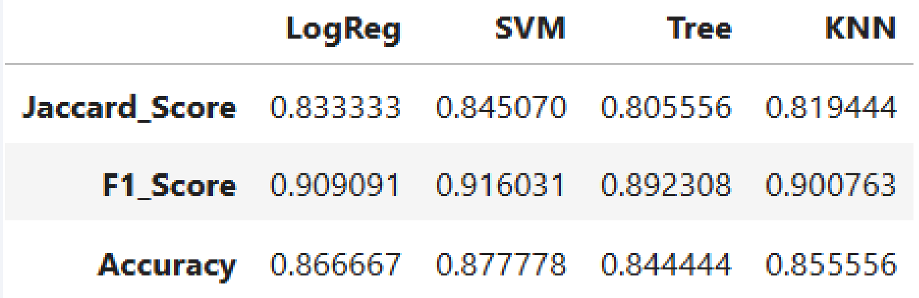
Classification Accuracy

The first table shows that based on the test-set none of the four models outperformed the others. They all scored the same across all three categories.



	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

When looking at the performance of the models on the whole dataset the Support-Vector-Machine (SVM) model outperformed the other models. It has not only higher scores, but also the highest accuracy. This indicates that the SVM-Model is more effective for this dataset than Logistic Regression, Decision-Tree, and K Nearest Neighbors. Although the overall difference between the models is not large.

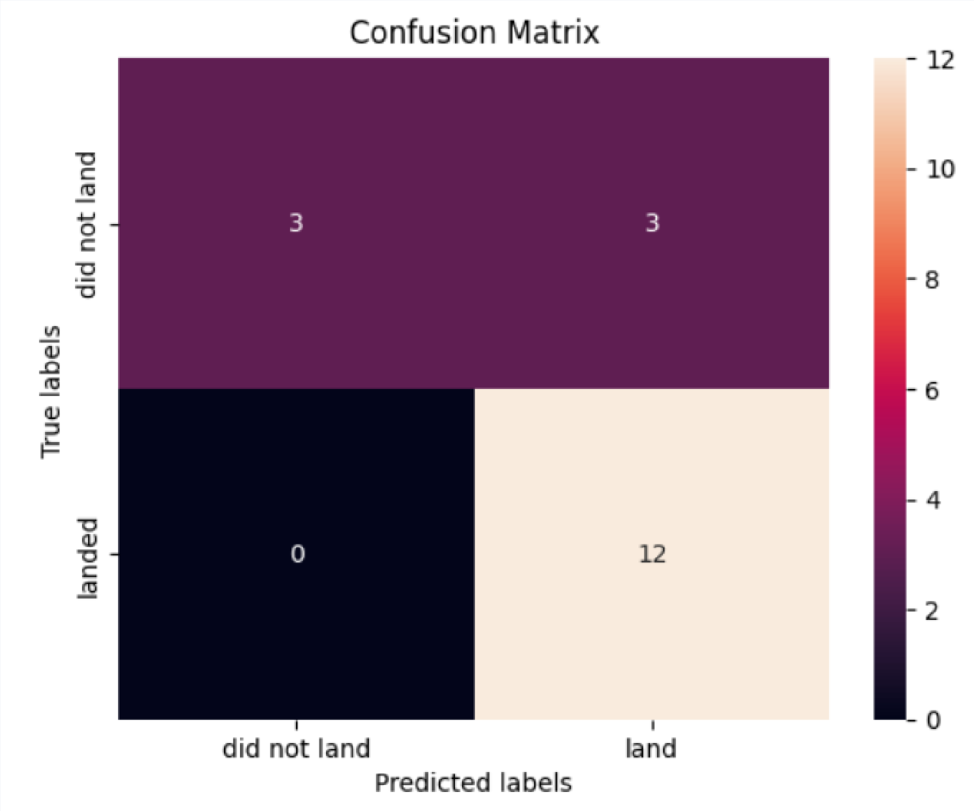


	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.805556	0.819444
F1_Score	0.909091	0.916031	0.892308	0.900763
Accuracy	0.866667	0.877778	0.844444	0.855556

GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P7_SpaceX_Machine%20Learning%20Prediction_Part_5_LK.ipynb

Confusion Matrix

The SVM-Model demonstrated high accuracy at 87.77%, effectively predicting Falcon 9 first stage landings with minimal false positives and no false negatives. The absence of false negatives is crucial for ensuring safety and operational readiness. Although there are false positives, it is less critical than false negatives and manageable in practice. Overall, the model's balanced performance, with a slight bias towards predicting successful landings, aligns well with the aerospace industry's need for reliable predictions in cost estimation and mission planning.



GitHub: https://github.com/raywen117/DSCapstone_LK/blob/main/P7_SpaceX_Machine%20Learning%20Prediction_Part_5_LK.ipynb

Conclusions

- **Finding 1:** The analysis indicates that the "CCAFS LC-40" launch site has the highest success rate, contributing to 43.7% of successful launches. This suggests that the site may offer optimal conditions that support better launch outcomes.
- **Finding 2:** The "FT" booster version demonstrated a high success rate across various payloads, highlighting its reliability and suggesting that future missions could benefit from using this version for improved success rates.
- **Finding 3:** No clear correlation was found between higher payload masses and lower success rates, suggesting that other factors, such as site conditions and booster versions, are more influential in determining launch success.
- **Finding 4:** Interactive visualizations created with Folium and Plotly Dash provided valuable insights into geographical and operational patterns, enabling better-informed decision-making through comprehensive data exploration.

Conclusion: This analysis, supported by predictive models and interactive visual tools, offers key insights into factors affecting SpaceX launch success. These findings can guide future strategies and support the ongoing advancement of reusable rocket technology.

Appendix

The sources and the code used for this report can be found by following the link to my GitHub repository:

https://github.com/raywen117/DSCapstone_LK/tree/main

Thank you!