IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexander Schmidt
20.01.2025

Outline

Executive Summary

This report aims to predict the success of **Falcon 9's first-stage landing** using historical launch data. SpaceX's ability to reuse the first stage significantly reduces launch costs, and predicting landing success helps estimate these costs. The analysis involves **collecting, cleaning, and preprocessing** publicly **available SpaceX launch data**, along with exploring launch sites using Folium to understand their proximity to coastlines and cities, shedding light on their strategic selection. The final step is to apply insights from **exploratory data analysis** to **build machine learning models**, including Decision Trees and Logistic Regression, to predict landing success. The Decision Tree model achieved the highest accuracy, correctly predicting 94.44% of outcomes. This model offers valuable insights into launch cost estimation, providing a competitive edge for companies like Space Y in assessing SpaceX’s pricing strategy. GridSearchCV was used to finetune hyperparameters for ML-models, but further model tuning could improve prediction accuracy even more.

Introduction

SpaceX has revolutionized the space industry by reducing the cost of rocket launches through reusability, particularly with the **Falcon 9's first-stage landing**. This cost-saving feature, which allows SpaceX to reuse the first stage of its rockets, gives the company a competitive edge, offering launches at a significantly lower price compared to other providers. This project focuses on **predicting whether the Falcon 9 first stage will land successfully**. Understanding the factors influencing this success can provide valuable insights into estimating launch costs and reusability. By analyzing historical data, exploring launch site locations, and applying machine learning models, we aim to develop an accurate prediction model. This information could potentially help companies like Space Y strategize against SpaceX by evaluating their pricing structure and making informed decisions.

Section 1

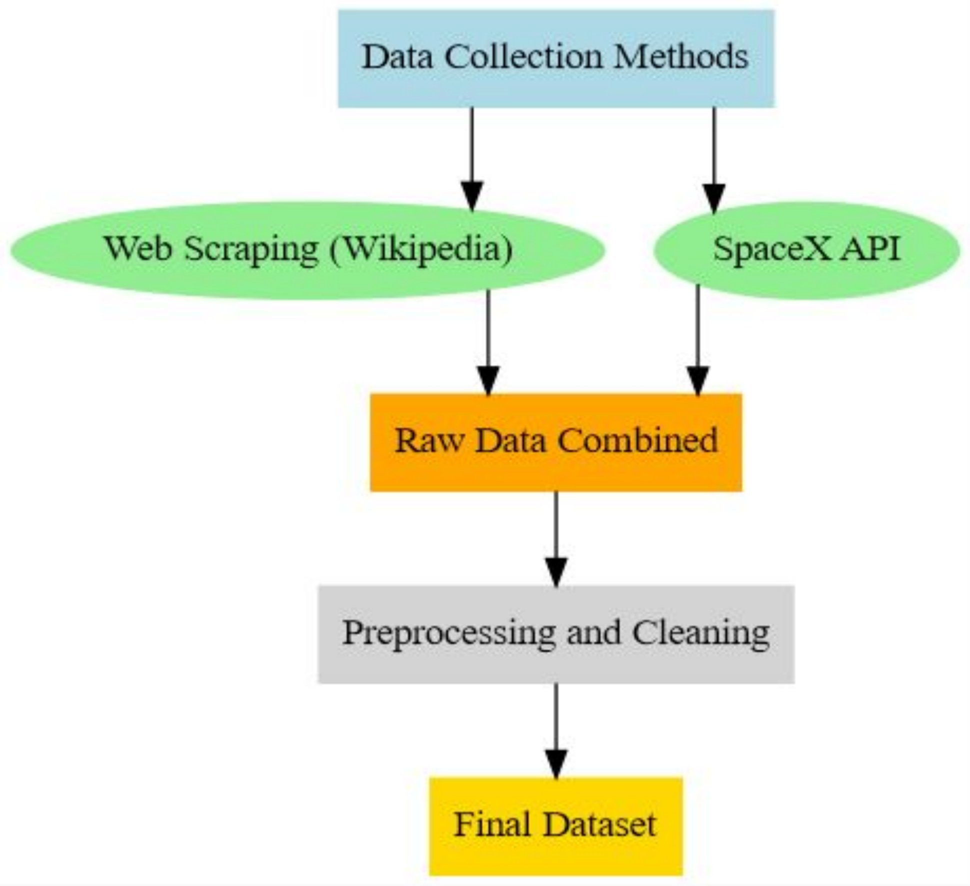
Methodology

Methodology

- Data collection **methodology**:
  - SpaceX launch data was obtained via Wikipedia web scraping (BeautifulSoup) and the SpaceX API.
- Perform **data wrangling**
  - Raw data from the API and web scraping were cleaned, structured into a Pandas DataFrame, and enriched with derived columns like date, time, and payload mass.
- Perform **exploratory data analysis (EDA)** using **visualization** and **SQL**
- Perform interactive visual analytics using **Folium** and **Plotly Dash**
- Perform predictive analysis using **classification models**
  - ML Models were built with Scikit-learn to predict Falcon 9 first-stage landing success.

Data Collection

- **Data Sources:**
- Wikipedia: Scraped Falcon 9 and Falcon Heavy launch data using BeautifulSoup from the "List of Falcon 9 and Falcon Heavy launches" page.
- SpaceX API: Retrieved additional launch details (payload, outcomes, orbital parameters) via SpaceX's public API.
- **Data Collection:**
- Web Scraping: Sent an HTTP GET request to Wikipedia and parsed the table into a Pandas DataFrame.
- API Data: Retrieved JSON launch data from SpaceX API and transformed it into a structured table.
- **Final Output:**
- Enriched Dataset: Merged the data from both sources into a clean, structured Pandas DataFrame ready for analysis.

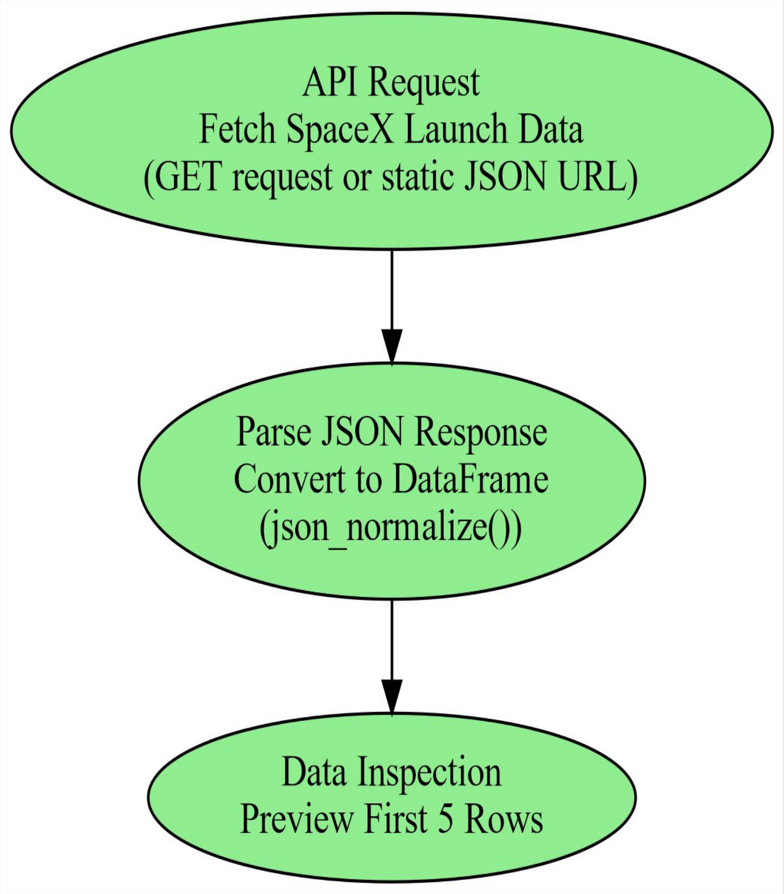


Data Collection – SpaceX API

SpaceX API Calls for Launch Data

- Fetch SpaceX Launch Data
  Use a GET request to retrieve SpaceX launch data.
- Parse and Structure Data
  Convert the JSON response into a Pandas DataFrame with the help of json_normalize().
- Inspect Data
  Preview the first five rows to ensure the data is correctly formatted and ready for analysis. (with data.head())

GitHub Reference:
[SpaceX API Calls Notebook]

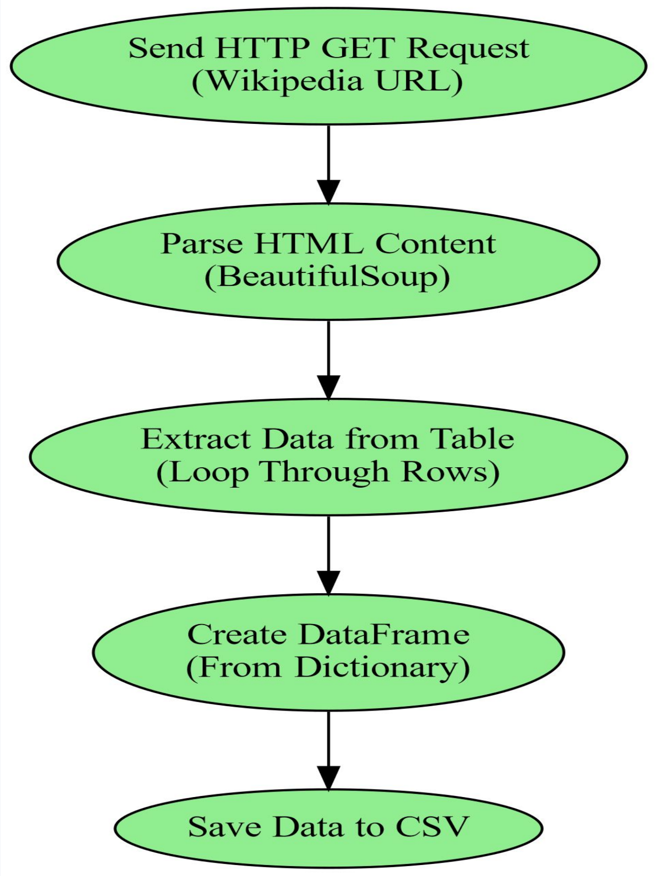


Data Collection - Scraping

Web Scraping Falcon 9 Launch Data

- Fetch Data
  Retrieve HTML content from the Wikipedia page using a GET request.
- Parse Content
  Use BeautifulSoup to extract the table with launch details.
- Extract Data
  Gather key information like flight number, date, and payload.
- Create DataFrame
  Structure the extracted data and convert it into a Pandas DataFrame.

GitHub Reference:
[SpaceX Web Scraping Notebook]()

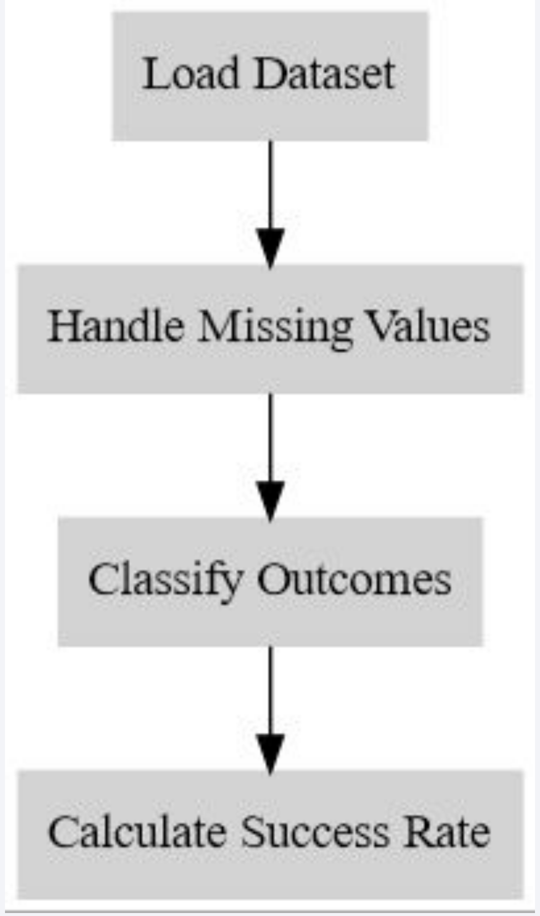


Data Wrangling

Data Wrangling Process

- Load Dataset:
  Import data using pd.read_csv().
- Handle Missing Values:
  Calculate missing values percentage.
- Classify Outcomes:
  Create a column for successful (1) or unsuccessful (0) landing based on Outcome.
- Calculate Success Rate:
  Calculate the success rate of landings.

GitHub Reference:
[SpaceX Data Wrangling Notebook]



EDA with Data Visualization

Performed **Exploratory Data Analysis (EDA)** on the SpaceX Falcon 9 dataset.

- **Key visualizations included:**
- **Success rate per Orbit**:
  Insight: Polar, LEO, and ISS orbits had higher success rates compared to GTO.
- **Flight Number vs. Payload Mass:**
  Insight: Higher flight numbers correlate with increased success rates
- **Flight Number vs. Launch Site:**
  Insight: The CCAFS site had a higher failure rate compared to other sites.
- **Launch Success Trend Over Time:**
  Insight: Success rates steadily improved after 2015.

GitHub Reference:
SpaceX Data Wrangling Notebook

EDA with SQL

Key Insights

- **Launch Sites**: Identified the unique launch sites used in SpaceX missions.
- **Filter Launch Sites**: Extracted launch sites starting with 'CCA' and displayed a sample.
- **Payload Mass**: Calculated the total and average payload mass for specific customers and booster versions.
- **Landing Outcomes**: Analyzed successful and failed landing outcomes, including the first successful landing on a ground pad.
- **Boosters & Payload**: Investigated boosters with successful drone ship landings and payloads within a specified range.
- **Mission Outcome Analysis**: Counted the number of successful and failed missions.
- **Maximum Payload**: Found booster versions that carried the maximum payload mass.
- **Monthly Analysis**: Analyzed landing failures on drone ships in 2015.
- **Landing Outcome Ranking**: Ranked landing outcomes by count within a specified date range.

GitHub Reference:

SQL EDA Notebook - SpaceX

Build an Interactive Map with Folium

Key Map Objects and Their Purposes:

- **Markers**:
  Placed on launch sites to show their geographical location with popup labels for site names.
- **Circles**:
  Highlighted launch site areas with a radius to emphasize the proximity of each site.
- **Marker Clusters**:
  Grouped launch outcomes (successful or failed) to make the map less cluttered and more informative.
- **PolyLines**:
  Drew lines to show distances from launch sites to nearby geographical features (e.g. coastlines, cities).

- **Map Purpose:**
  The map visually represents launch sites, marks launch outcomes, shows proximity to geographical features, and helps analyze distances between sites and nearby features.

GitHub Reference:
[Interactive Map with Folium - SpaceX Launch Sites Analysis]

Build a Dashboard with Plotly Dash

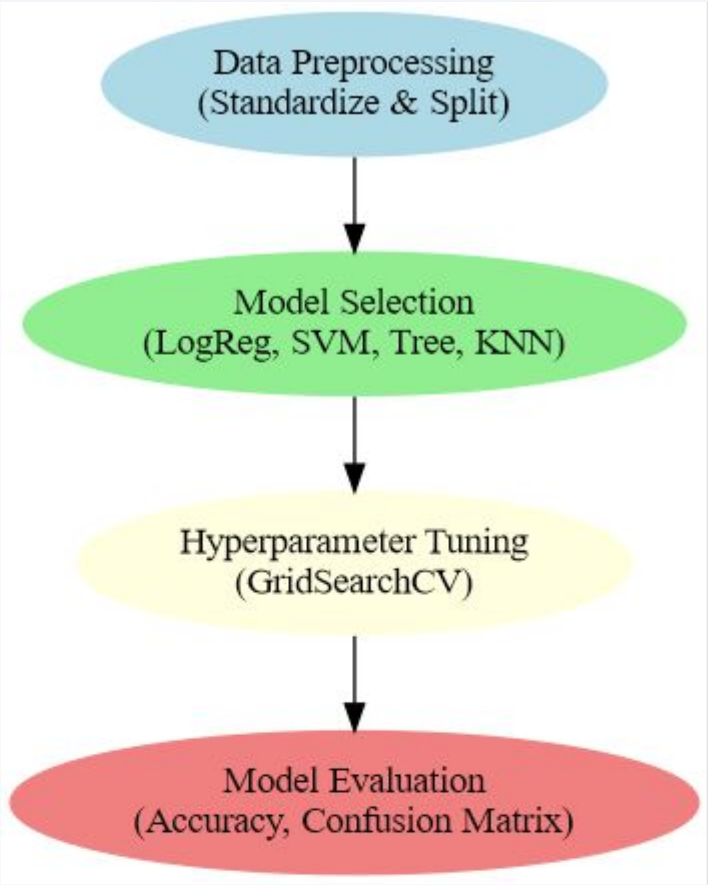
- **Summary of Plots/Graphs and Interactions:**
- **Pie Chart for Launch Success:**
  - Purpose: Displays the success rate of SpaceX launches for either all sites or a specific site.
  - Interaction: Users can select a launch site from a dropdown, updating the pie chart to show success/failure for that site or all sites.
- **Scatter Plot for Payload vs Launch Success:**
  - Purpose: Shows the relationship between payload mass and launch success, along with booster version categories.
  - Interaction: Users can filter data by launch site and payload range using a range slider, updating the plot accordingly.
- **Why These Plots and Interactions Were Added:**
- **Pie Chart**: Provides a clear view of launch success/failure, with interactivity for site-specific exploration.
- **Scatter Plot**: Reveals the correlation between payload mass and success, with filtering options to explore different payload ranges and sites.

GitHub Reference: SpaceX Dashboard

Predictive Analysis (Classification)

Predictive Analysis Summary

- **Data Preprocessing**:
  - Loaded data and standardized features using StandardScaler.
  - Split data into training and test sets (80%/20%).
- **Model Selection & Tuning**:
  - Evaluated four models: Logistic Regression, SVM, Decision Tree, and KNN.
  - Tuned hyperparameters using GridSearchCV.
- **Evaluation**:
  - Models evaluated on accuracy and confusion matrix to identify performance.
  - Best model selected based on the highest accuracy score.



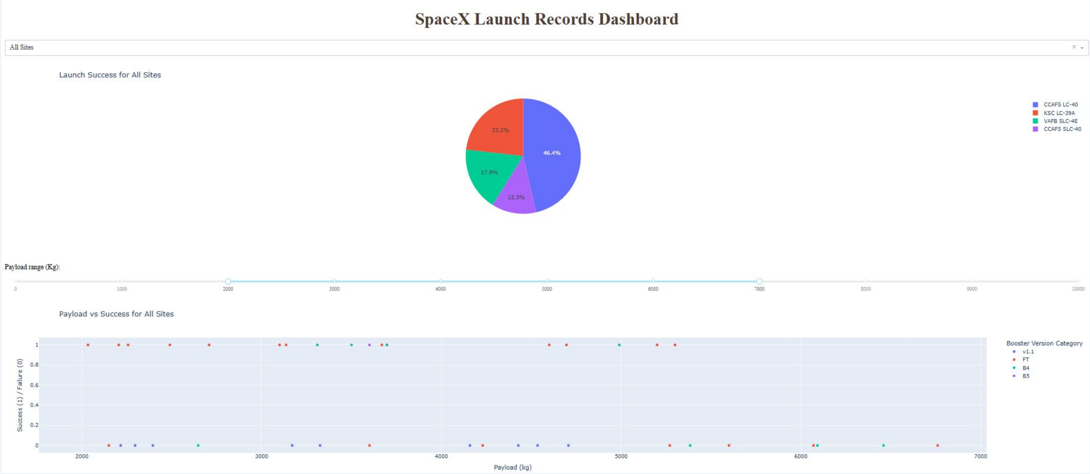
GitHub Reference:
GitHub - SpaceX Machine Learning Prediction

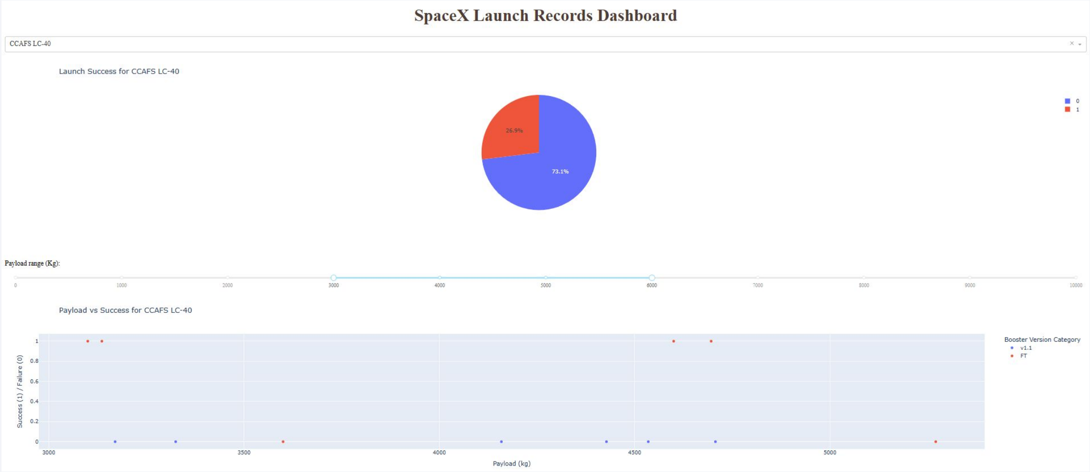
Results

Exploratory Data Analysis (EDA): Key features like Class, BoosterVersion, and Mission were analyzed. Significant correlations were found between certain features and the target variable (Class).
Visualization: Graphs like histograms and bar charts show trends in the data, such as a higher chance of successful landings over time.

Predictive Analysis Results:
Models Evaluated: Logistic Regression, SVM, Decision Trees, and KNN.
Best Model: The Decision Tree classifier achieved the highest accuracy on the first run with a score of 0.94 on test data.
Hyperparameter Tuning: GridSearchCV was used to optimize models.

Interactive Dashboard Examples

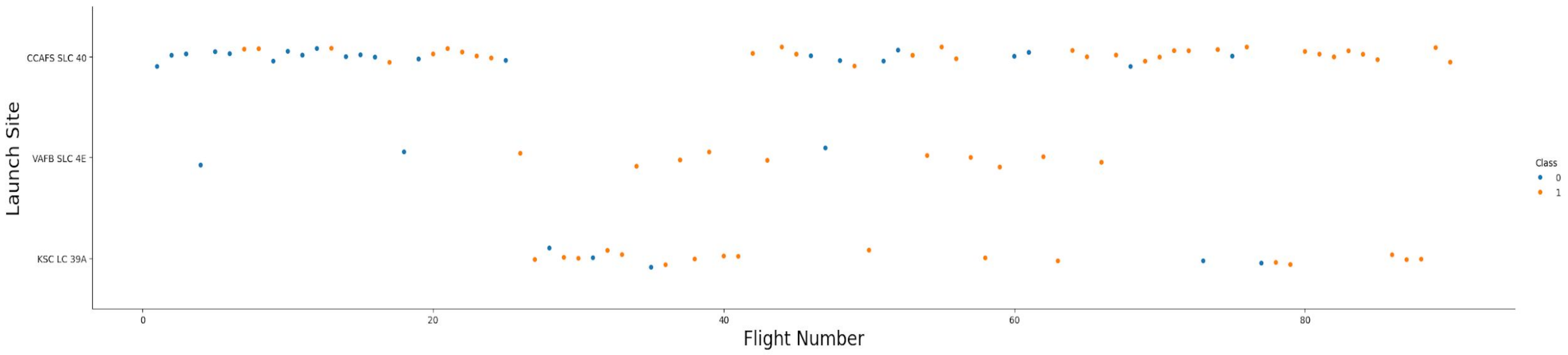




Section 2

Insights drawn from EDA

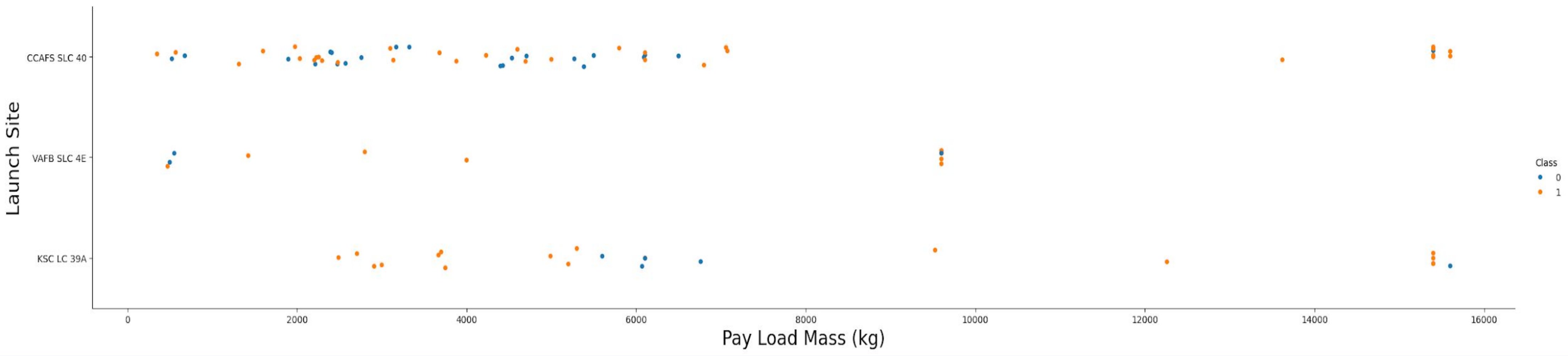
Flight Number vs. Launch Site



Key Insights / Explanations:

- For the CCAFS SLC 40 and VAFB SLC 4E launch sites:
  A clear correlation exists between the number of flights and an increasing rate of successful
  landings (represented by a value of 1).
  In contrast, the KSC LC 39A launch site does not show this trend as clearly.
- Among all the launch sites, CCAFS SLC 40 has the highest number of flights, underscoring its significant role in SpaceX's operations.

Payload vs. Launch Site



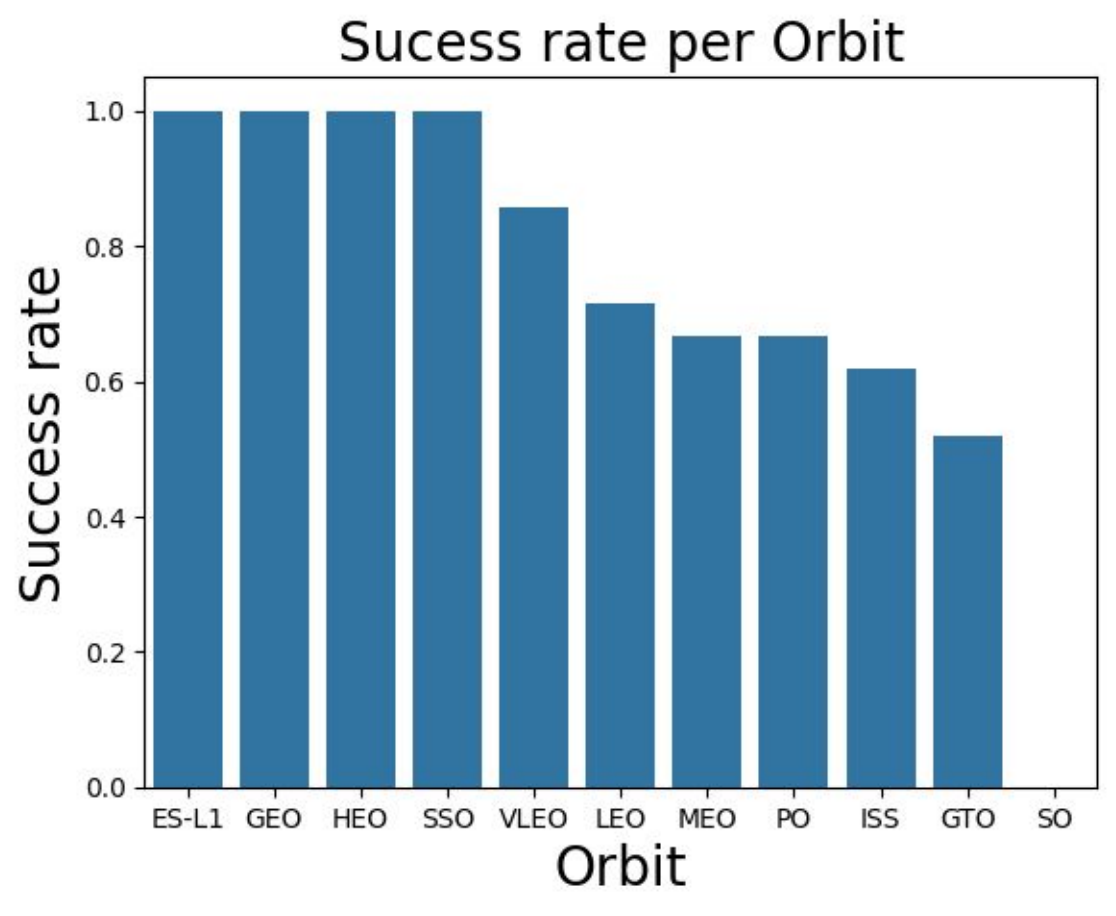
Key Insights / Explanations:

- Payloads exceeding 9000 kg tend to have a higher success rate compared to those below 8000 kg.
- All three launch sites show a greater number of flights associated with lower payload masses than with higher payload masses (in kg).

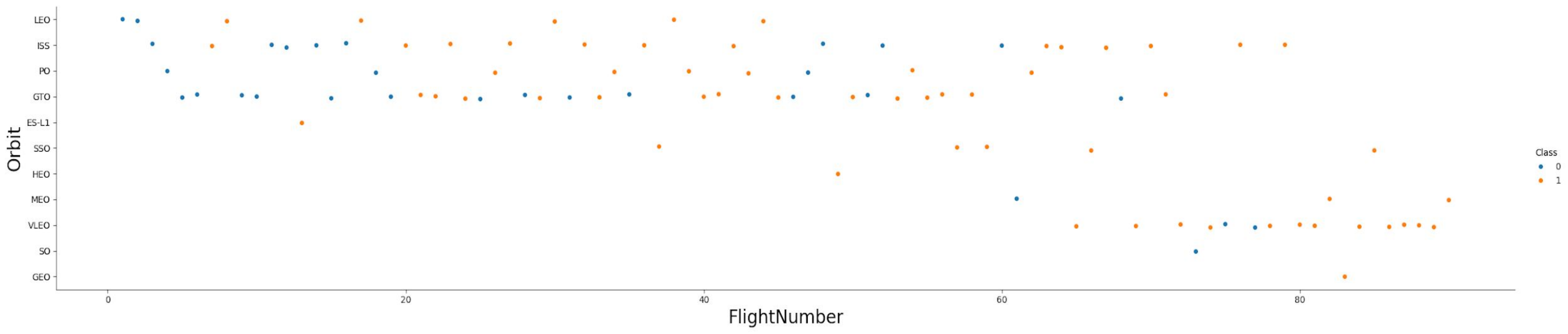
Success Rate vs. Orbit Type

Key Insights / Explanations:

- ES-L1, GEO, HEO, and SSO orbits all show a success rate very close to 1, indicating high reliability for launches to these orbit types.
- SO has the lowest success rate (only one flight to that orbit in the data), with GTO (Geostationary Transfer Orbit) having the second lowest.
- Orbits such as VLEO, LEO, MEO, PO, and ISS fall in between in terms of success rate



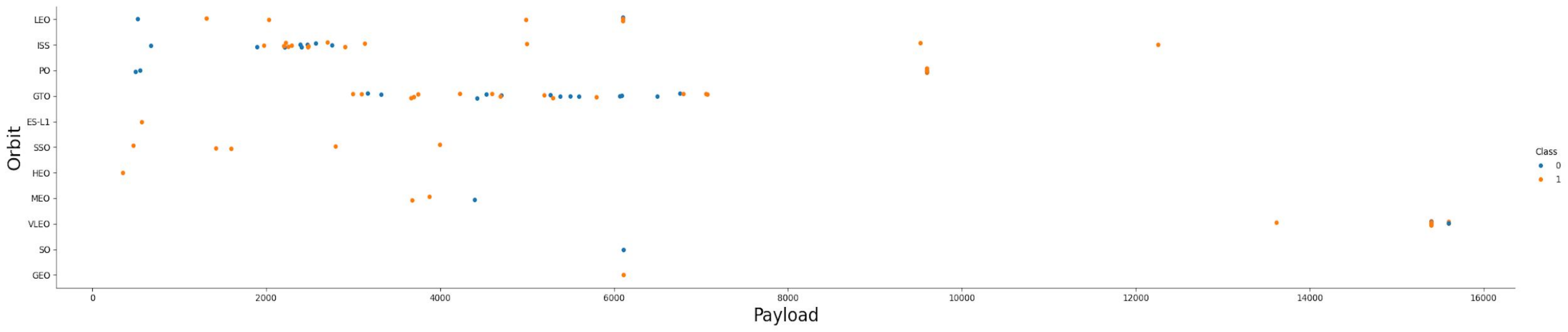
Flight Number vs. Orbit Type



Key Insights / Explanations:

- The early flights were primarily targeted at orbits such as LEO, ISS, PO, and GTO.
- Over time, later flights expanded to include a wider variety of orbits.
- As the flight number increases, there is a noticeable improvement in the success rate, suggesting better experience and optimization in later missions.

Payload vs. Orbit Type



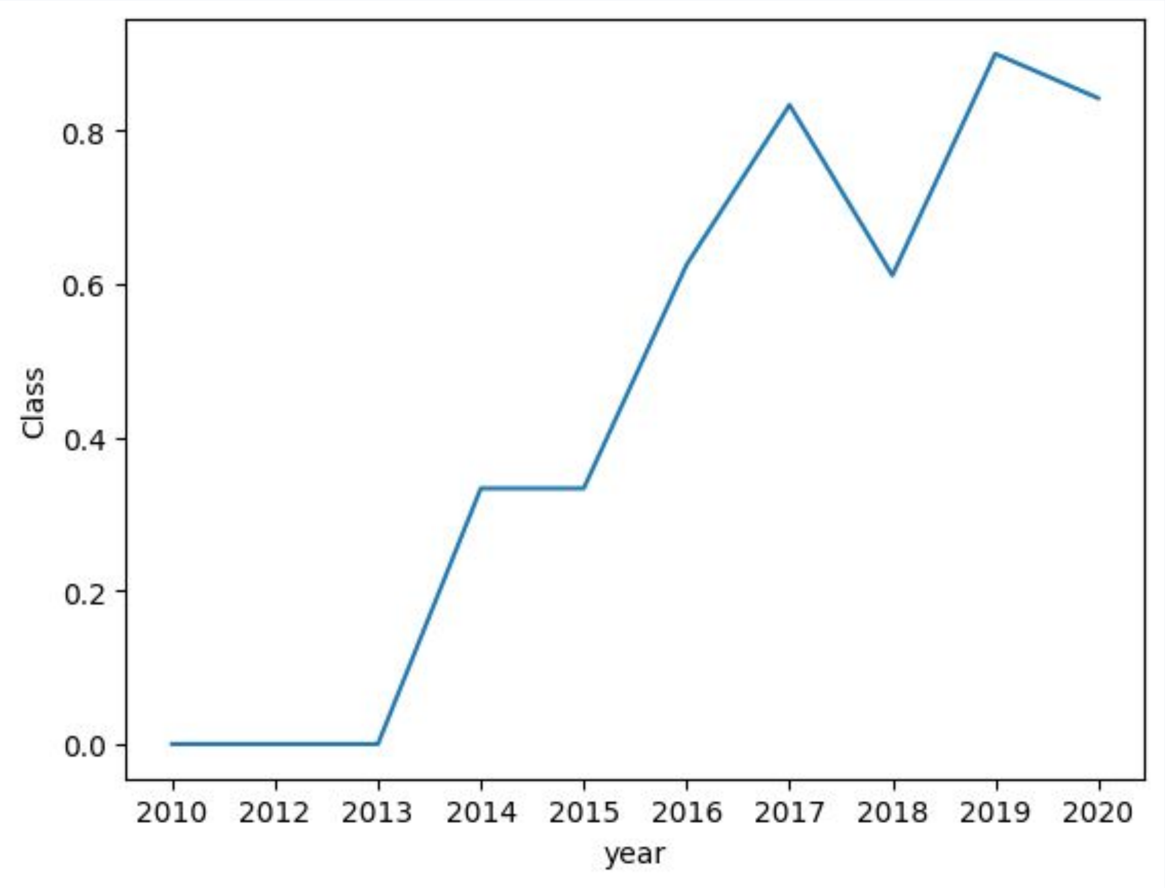
Key Insights / Explanations:

- The are far more Flights with lower Payloads
- ISS, PO and GTO have many unsuccessful landings with Payloads <8000kg

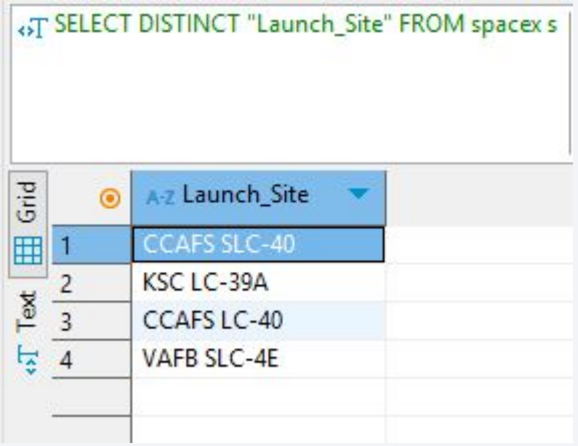
Launch Success Yearly Trend

Key Insights / Explanations:

- The success rate shows a clear trend and grows over time
- There is a small decline between around 2017 and 2018
- There is another even smaller decline starting in 2019



All Launch Site Names



```sql
SELECT DISTINCT "Launch_Site" FROM spacex s
```

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

Explanation:

- This query result shows the names of the four different Launch Sites in the Dataset.

Launch Site Names Begin with 'CCA'

spacex 1

```
SELECT * FROM spacex s WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5
```

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
1	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2	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)
3	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
4	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
5	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Explanation:

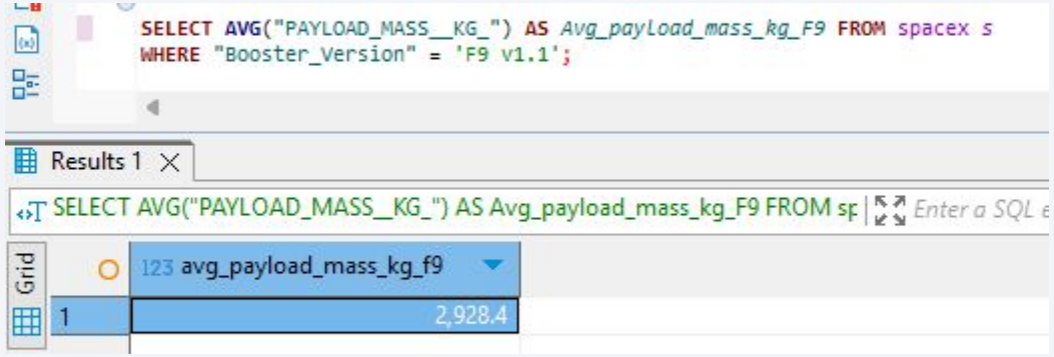
- This query result shows the first five records where the Launch Site begins with `CCA`

Total Payload Mass

Explanation:

- The above query result shows the total payload in kg carried by boosters from NASA.
- It is 45.596 kg, so almost 46 tons

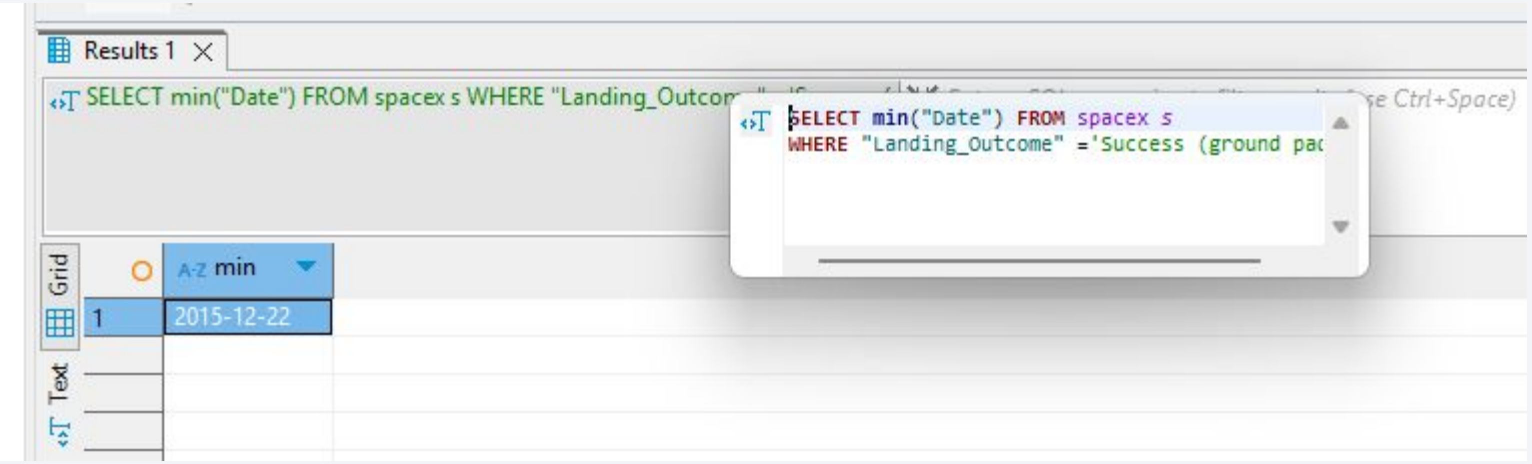
Average Payload Mass by F9 v1.1



Explanation:

- This query result shows that F9 boosters version 1.1 carry almost 3000kg payload mass on average on their flights.
- The query filters for the booster version and calculates the average payload mass.

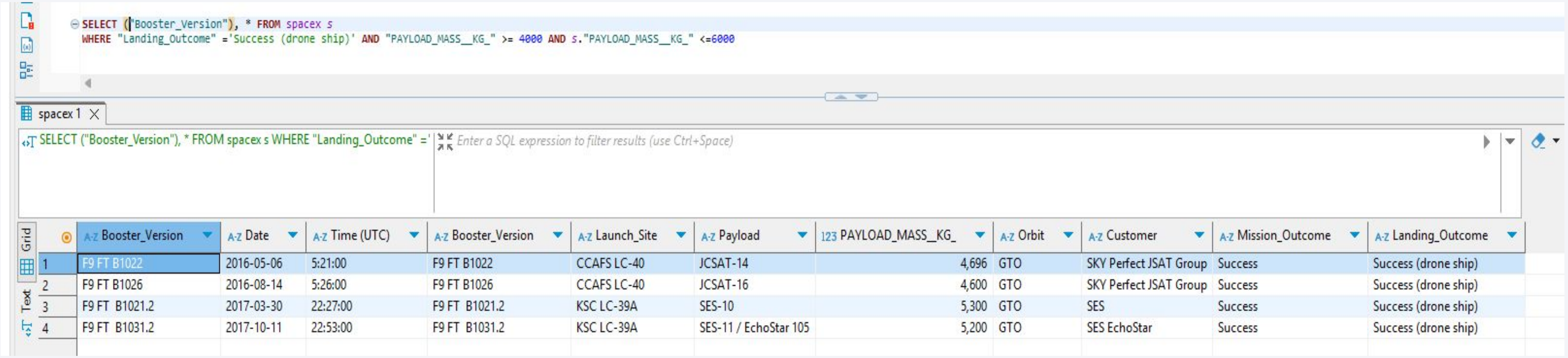
First Successful Ground Landing Date



Explanation:

- The above query result shows that the first successful landing on a ground pad happened two days before Christmas in 2015.

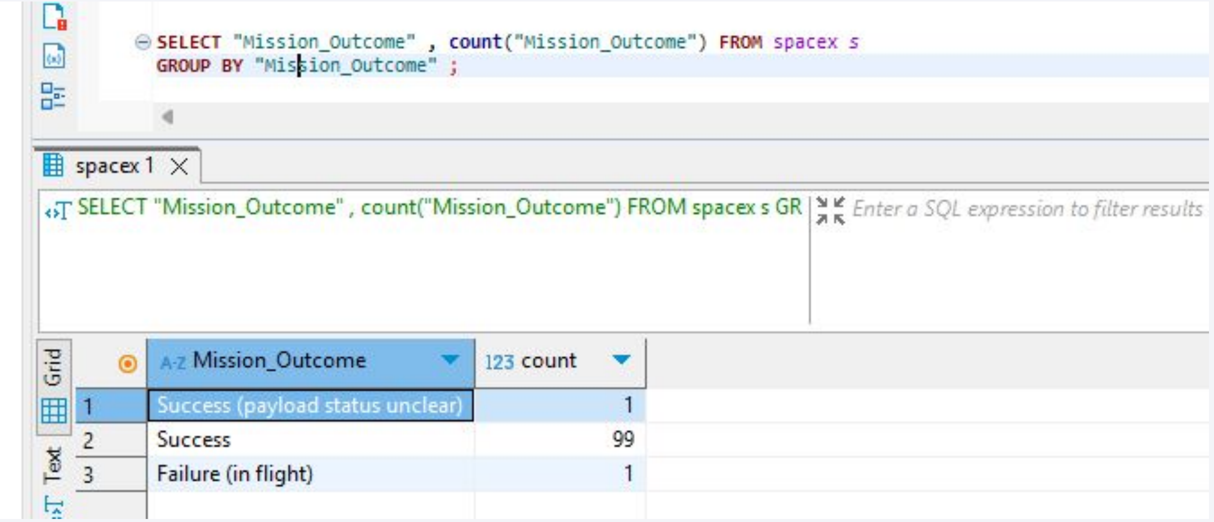
Successful Drone Ship Landing with Payload between 4000 and 6000



Explanation:

- This query result shows the names of boosters which have successfully landed on a drone ship and had payload mass greater than 4000 kg but less than 6000 kg
- There are four flights in the dataset that meet these conditions

Total Number of Successful and Failure Mission Outcomes



Explanation:

- This query result shows that out of the 101 flights:
  99 fulfilled their mission,
  1 fulfilled it, but the payload status is unclear
  1 mission ended in a failure
- note that mission success is different from landing success and reuse of the first stage

Boosters Carried Maximum Payload

```sql
SELECT "Booster_Version", * FROM spacex s
WHERE s."PAYLOAD_MASS__KG_" IN (SELECT MAX("PAYLOAD_MASS__KG_") FROM spacex s2);
```

spacex 1

SELECT "Booster_Version", * FROM spacex s WHERE s."PAYLOAD_MASS__KG_

Enter a SQL expression to filter results (use Ctrl+Space)

	Booster_Version	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
1	F9 B5 B1048.4	2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15,600	LEO	SpaceX	Success	Success
2	F9 B5 B1049.4	2020-01-07	2:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15,600	LEO	SpaceX	Success	Success
3	F9 B5 B1051.3	2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15,600	LEO	SpaceX	Success	Success
4	F9 B5 B1056.4	2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15,600	LEO	SpaceX	Success	Failure
5	F9 B5 B1048.5	2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15,600	LEO	SpaceX	Success	Failure
6	F9 B5 B1051.4	2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15,600	LEO	SpaceX	Success	Success
7	F9 B5 B1049.5	2020-06-04	1:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15,600	LEO	SpaceX, Planet Labs	Success	Success
8	F9 B5 B1060.2	2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15,600	LEO	SpaceX	Success	Success
9	F9 B5 B1058.3	2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15,600	LEO	SpaceX	Success	Success
10	F9 B5 B1051.6	2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15,600	LEO	SpaceX	Success	Success
11	F9 B5 B1060.3	2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15,600	LEO	SpaceX	Success	Success
12	F9 B5 B1049.7	2020-11-25	2:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15,600	LEO	SpaceX	Success	Success

Explanation:

The query result shows the booster version of the 12 flights with the maximum payload mass of 15600 kg

2015 Launch Records

```sql
SELECT TO_CHAR("Date"::date, 'FMMonth'), * FROM spacex
WHERE "Landing_Outcome" = 'Failure (drone ship)' AND EXTRACT (YEAR FROM "Date" ::date) = 2015;
```

spacex 1

SELECT TO_CHAR("Date"::date, 'FMMonth'), * FROM spacex WHERE "Landin

Enter a SQL expression to filter results (use Ctrl+Space)

	to_char	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
1	January	2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2,395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2	April	2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1,898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

Explanation:

- This query result shows the two failed landing_outcomes in drone ship with their booster versions and launch site names in the year 2015
- There are two NASA flights that fulfill these specific conditions

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

Explanation:

- This query result shows the rank according to the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 ordered by outcome_count in descending order
- “No attempt” is the most common Landing_Outcome

```sql
SELECT
    "Landing_Outcome",
    COUNT("Landing_Outcome") AS outcome_count,
    RANK() OVER (ORDER BY COUNT("Landing_Outcome") DESC) AS rank
FROM spacex s
WHERE "Date"::date >= '2010-06-04' AND "Date"::date <= '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY rank;
```

spacex 1

SELECT "Landing_Outcome", COUNT("Landing_Outcome") AS outcome_cou

	Landing_Outcome	outcome_count	rank
1	No attempt	10	1
2	Success (drone ship)	5	2
3	Failure (drone ship)	5	2
4	Success (ground pad)	3	4
5	Controlled (ocean)	3	4
6	Uncontrolled (ocean)	2	6
7	Failure (parachute)	2	6
8	Precluded (drone ship)	1	8

Note: %sqllite queries can be found on **GitHub:** [SQL EDA Notebook - SpaceX]

Section 3

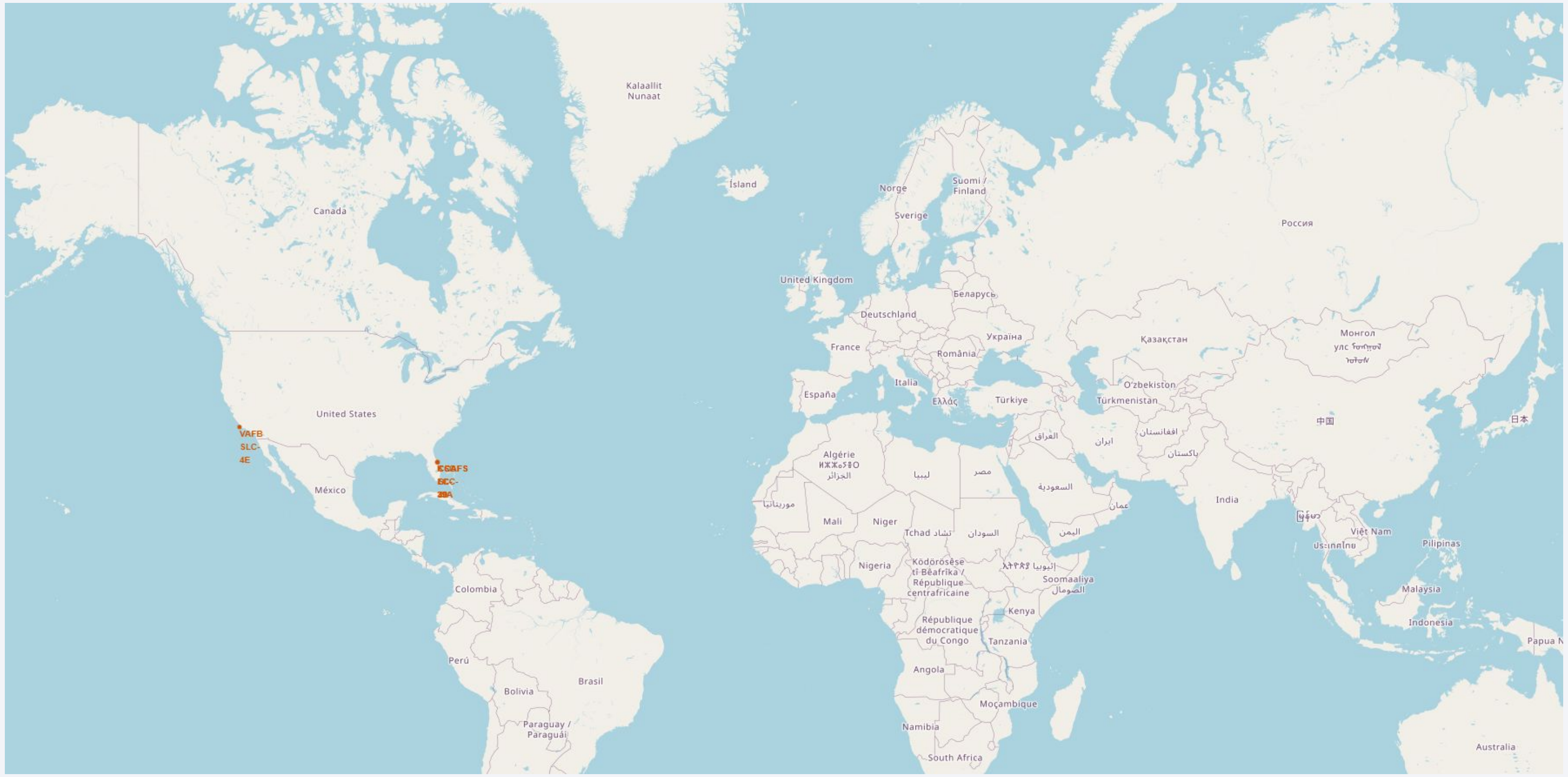
Launch Sites Proximities Analysis

Global Distribution of SpaceX Launch Sites

- **Launch Site Locations:**

  3 sites (**CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A**) are located on the east coast of the United States, in **Florida**.

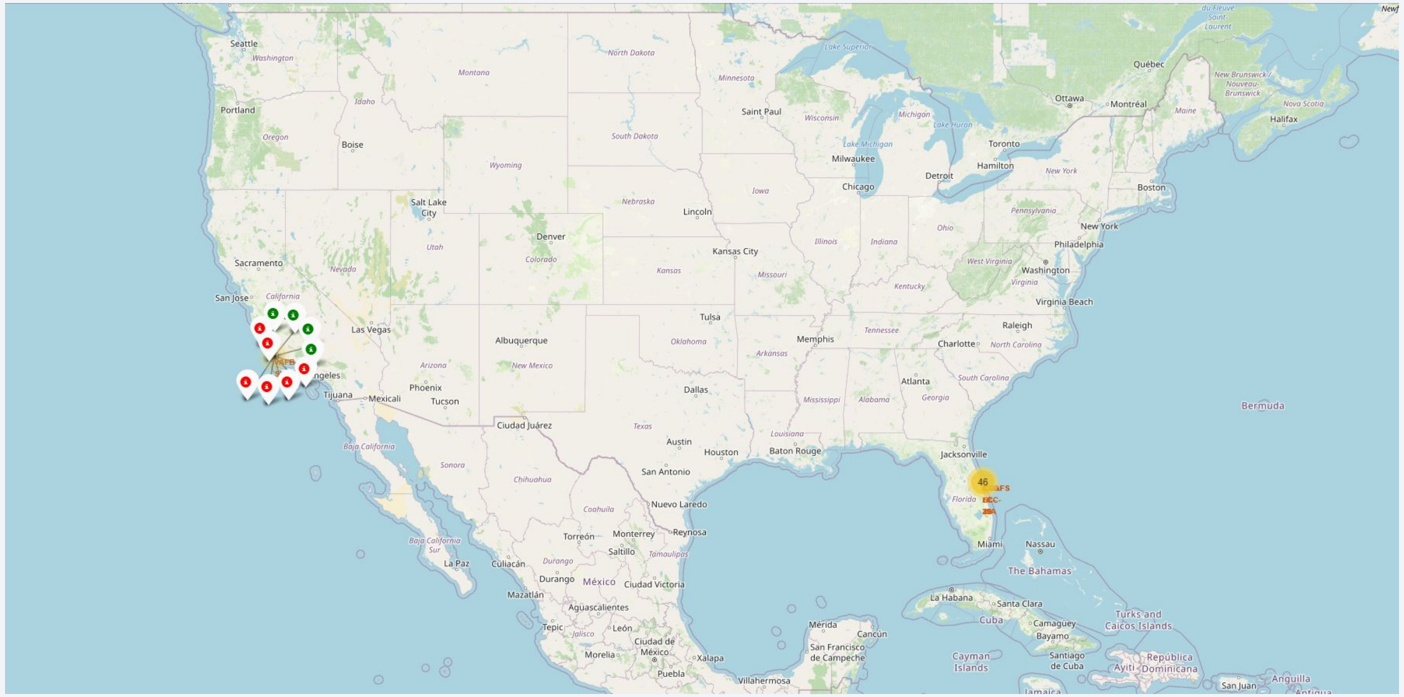
  1 site (**VAFB SLC-4E**) is situated on the west coast of the US, in **California**.



Launch Outcomes by Color

- West Coast (**VAFB SLC-4E**): 10 launches
  6 failed
  4 successful.
- East Coast: 46 launches total

Red Marker color: failed to reuse the first stage (=0 in the class column)
Green Marker color: success to reuse the first stage (=1 in the class column)



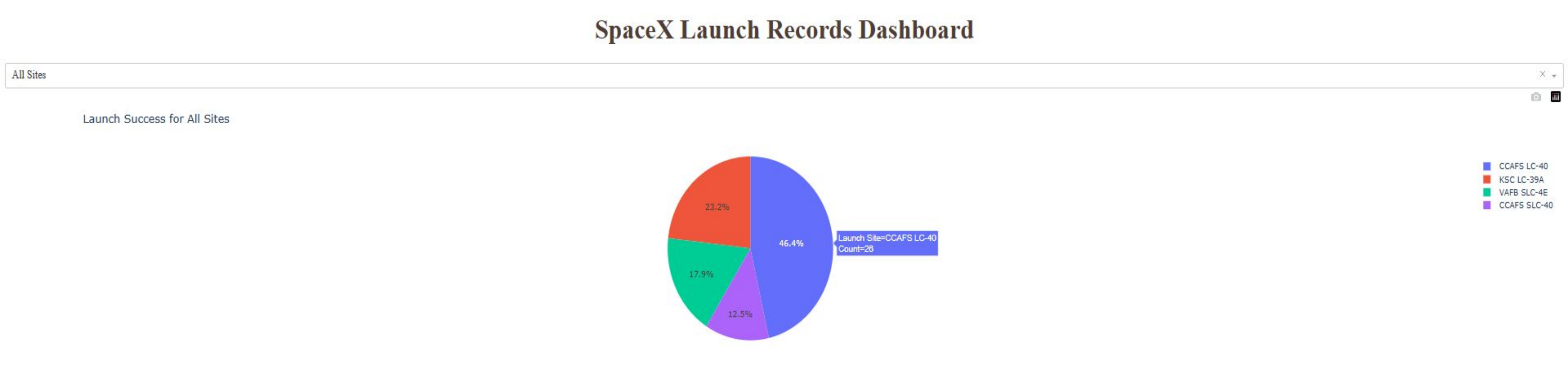
Proximity Analysis of West Coast Launch Site

- **Key Insights**:
- The **VAFB SLC-4E** (west coast launch site) is **1.35 km** from the coastline, highlighting its close proximity to the ocean for safe rocket launches.
- The nearest city, **Santa Maria**, is located **39.04 km** away, ensuring a safe distance from populated areas.

Section 4

Build a Dashboard with Plotly Dash

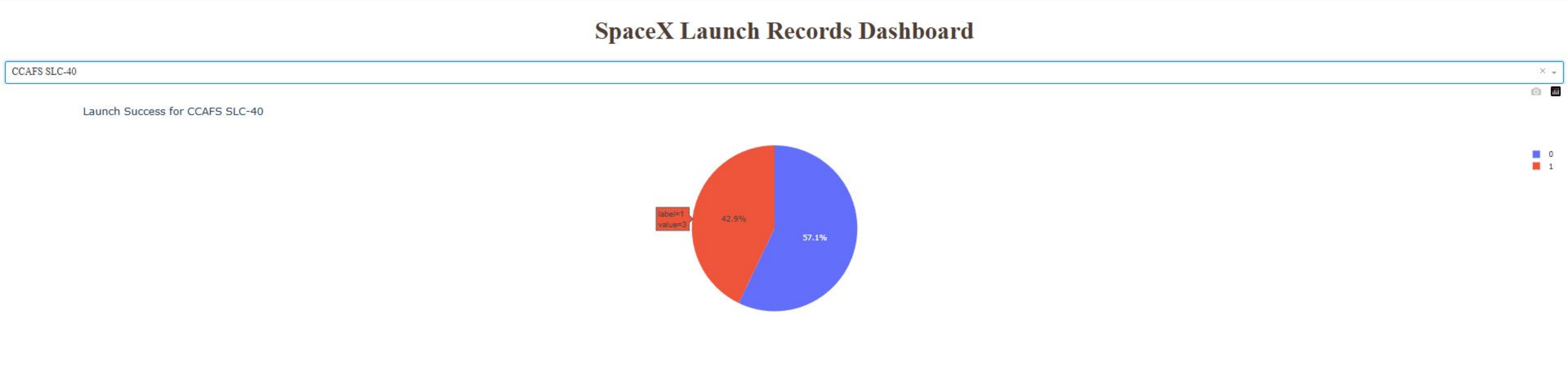
Launch Success Distribution Across Sites



Key Insights:

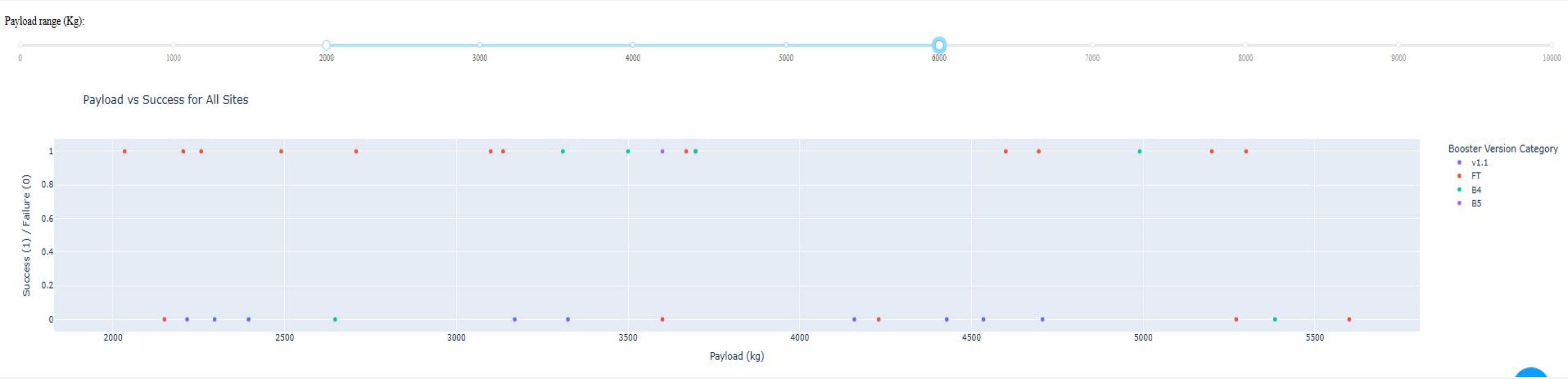
- The pie chart visually represents the **launch success counts** for all four sites.
- **CCAFS LC-40** has the highest share of successful launches, highlighting its prominence in SpaceX's operations.
- **KSC LC-39A** and **VAFB SLC-4E** have smaller but significant portions, indicating diversified launch activities.
- **CCAFS SLC-40** has the smallest slice, reflecting fewer launches or success rates.

CCAFS SLC-40 Launch Site: Success Ratio



- **Key Insights**:
- **CCAFS SLC-40** has the highest **success ratio (42.9%)**, based on its total launch history.
- Out of the 7 launches from this site, **3 launches were successful**

Launch Outcomes Across All Sites for Payload Range 2000–6000 kg



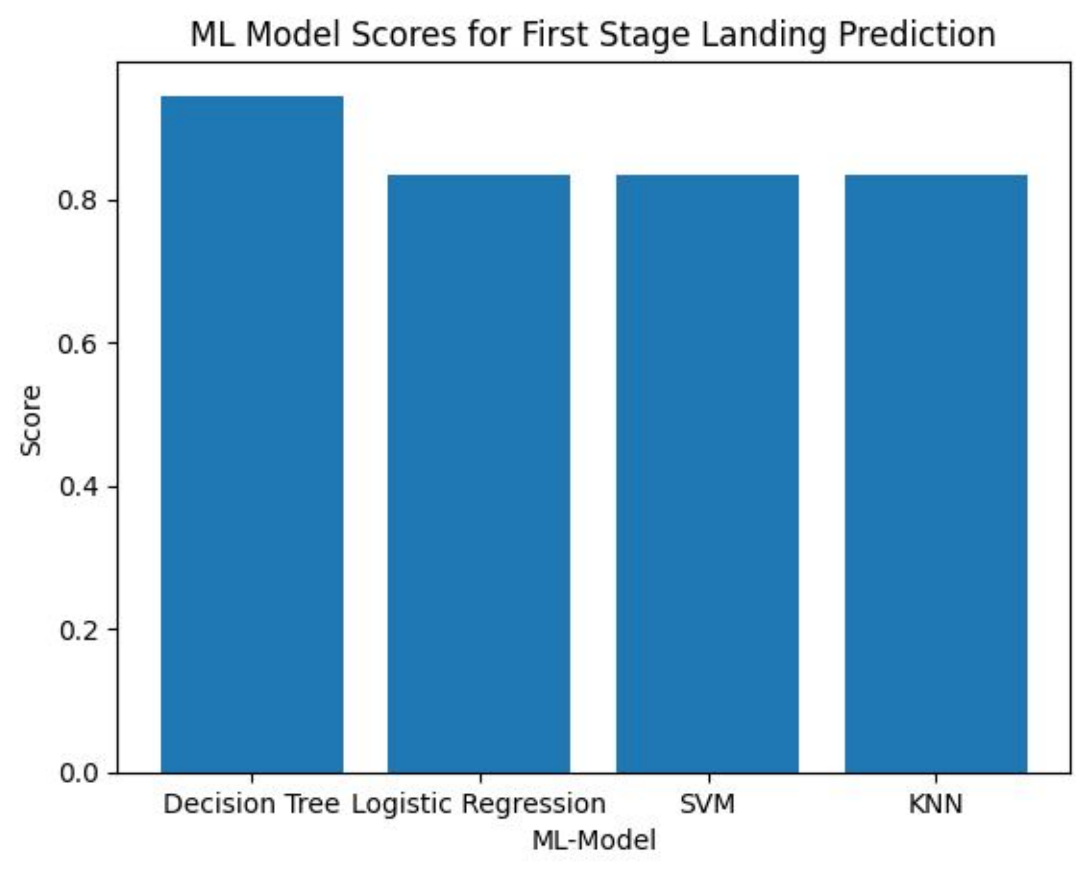
- **Key Insights**:
- The scatter plot focuses on launches with payloads between **2000 and 6000 kg**.
- Many of the successful launches use **Booster Version FT** or **B4**
- **Booster Version v1.1** is not very successful for this payload range

Section 5

Predictive Analysis (Classification)

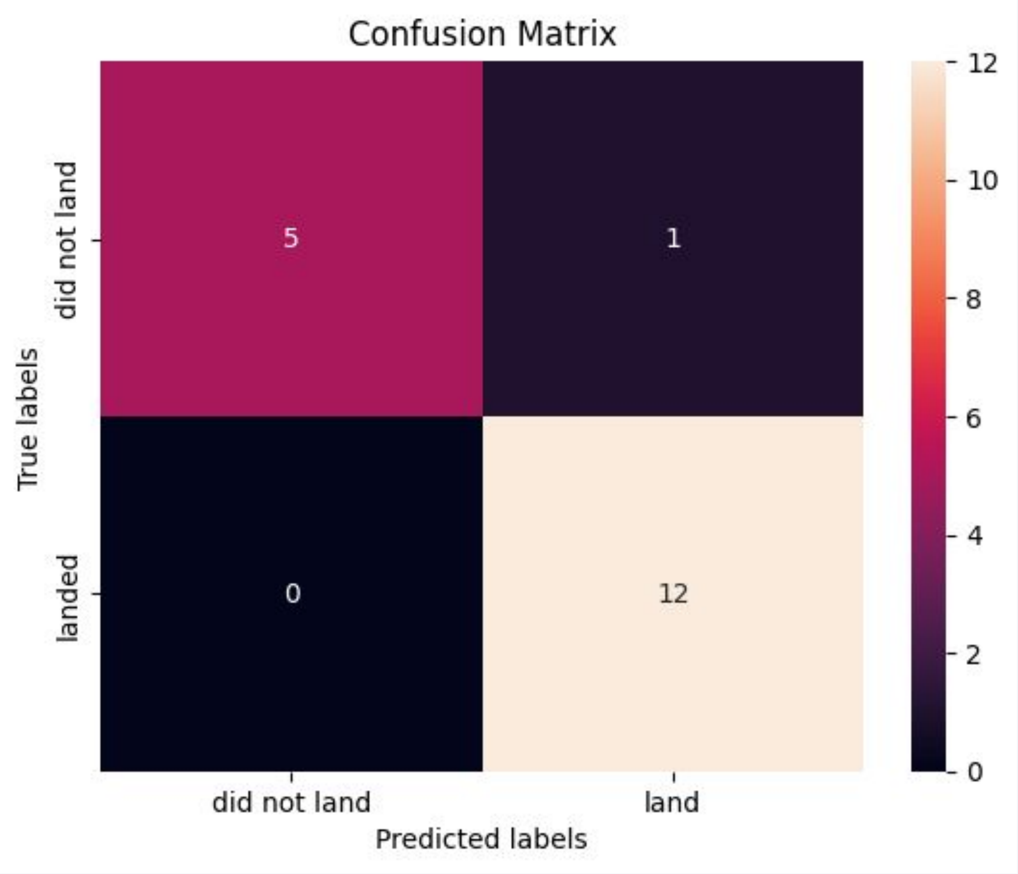
Classification Accuracy

- **Key Insights:**
- Decision Tree achieves the **highest score** on the first notebook run, showcasing strong performance.
- Some **randomness** in the training process can lead to variations in model scores during subsequent runs.



Confusion Matrix

- **Insights:**
- The Decision Tree model predicted 17/18 results correctly, giving it **94.44% accuracy** on the test data, indicating strong predictive performance.
- The model made only **1 false positive**, where it predicted a landing (land), but the actual result was no landing (did not land). This single mistake doesn't significantly affect the model's overall performance, as the accuracy is very high.



Conclusions

- **High Performance Across Models:**

  All machine learning models achieved scores above 80%, with the Decision Tree model reaching an impressive 94.44% accuracy during testing, demonstrating the potential of machine learning in predicting Falcon 9's first-stage landing success.

- **Key Features Impacting Success:**

  Several key features significantly influence the success of the landing, including the mission's orbit, payload mass, launch site, and the number of previous flights. These factors are crucial in predicting the likelihood of a successful landing.

- **Improvement in Landing Success Over Time:**

  There has been a noticeable improvement in the success rate of first-stage landings over time, highlighting SpaceX's continued advancements in technology, precision, and operational experience.

- **Geographic Commonalities Among Launch Sites:**

  The four primary SpaceX launch sites share common geographic features, such as proximity to railway stations and the ocean, as well as a safe distance from populated areas. These factors are integral to site selection for safety and operational efficiency.

- **Average Payload Mass for Falcon 9 Boosters Version 1.1 is around 3000 kg**

  While the average payload mass for Falcon 9 boosters version 1.1 is approximately 3000 kg, a significant number of these missions do not result in successful first-stage landings. This suggests that payload mass plays a significant role in landing success, alongside other factors like mission parameters and booster version.

Recommended Actions

- **1. Focus on High-Impact Launch Years**
  Action: Allocate more resources to years with high payload mass to maximize operational efficiency.
  Reason: Certain years show significantly higher payloads—understanding the key factors can lead to improved cost efficiency in the future.
- **2. Analyze Landing Success Trends**
  Action: Investigate the factors leading to successful landings (e.g., weather, technology, rocket version).
  Reason: Success in landing correlates with cost savings—further improving landing success rates will reduce overall mission expenses.
- **3. Optimize for Future Growth in Payload Mass**
  Action: Explore potential market opportunities where payload mass is increasing.
  Reason: Years with larger payloads indicate growing customer needs—tailoring rockets for larger payloads could open up new business opportunities.
- **4. Further optimize Machine Learning Models with recent data**
  Action: Source additional data to further optimize models, for example data from 2020 to 2025.
  Reason: The test data was very small. More training and testing data and more recent data can lead to better and more accurate results.

Appendix

Explanation:
This SQL query aggregates the total payload mass for each year, and the corresponding bar plot visually displays the total payload mass (in kg) launched by SpaceX per year.

```
payload_mass_by_year = %sql SELECT \
    substr(Date, 1, 4) AS Year, \
    SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass_kg \
FROM spacex \
GROUP BY Year \
ORDER BY Year;
payload_mass_by_year
```

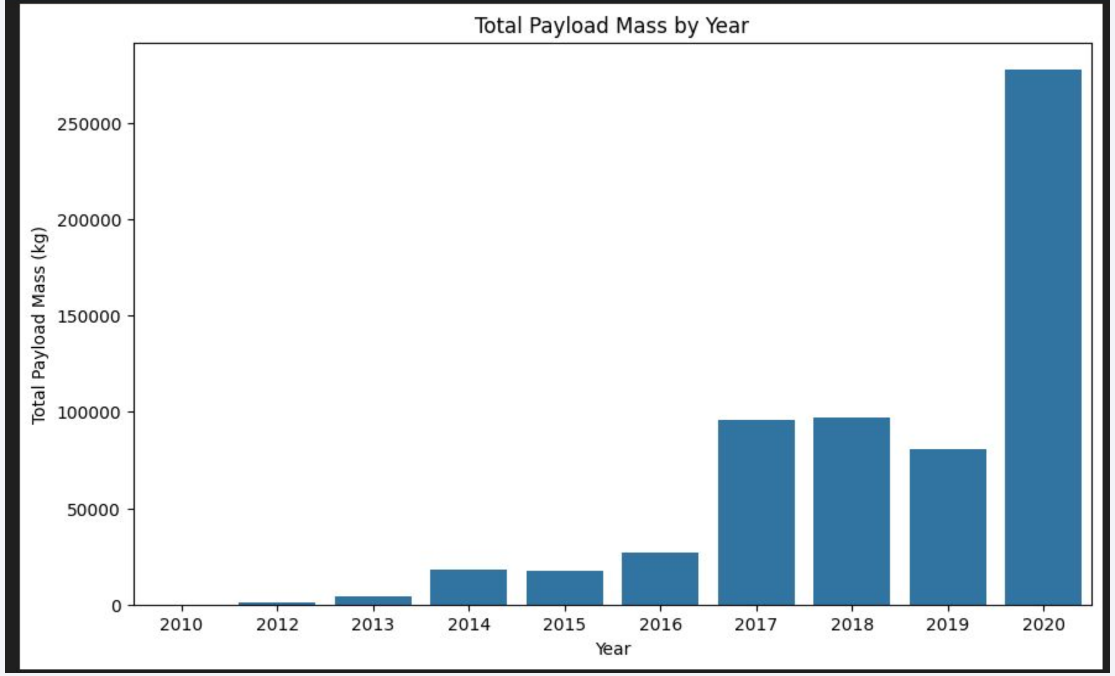
✓ 0.0s

* sqlite:///my_data1.db
Done.

Year	Total_Payload_Mass_kg
2010	0
2012	1025
2013	4347
2014	18116
2015	17715
2016	27213
2017	95978
2018	96957
2019	80761
2020	277855

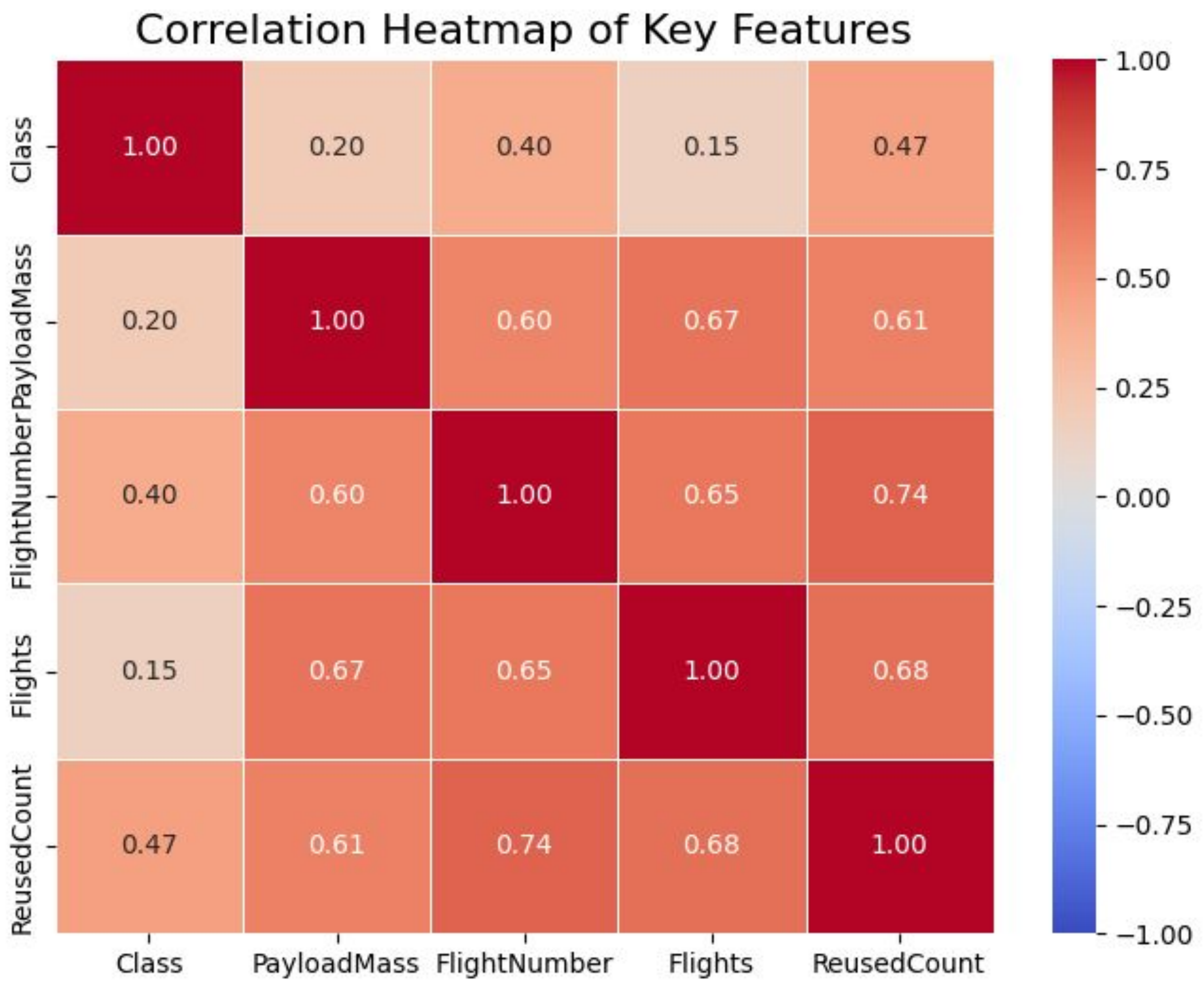
```
plmby = pd.DataFrame(payload_mass_by_year)

plt.figure(figsize=(10,6))
sns.barplot(x='Year', y='Total_Payload_Mass_kg', data=plmby)

plt.title('Total Payload Mass by Year')
plt.xlabel('Year')
plt.ylabel('Total Payload Mass (kg)')
plt.show()
```

✓ 0.2s



Appendix

Explanation:
This Heatmap shows that Class is more closely related to FlightNumber and ReusedCount than to PayloadMass



Thank you!