IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anansh Sinha

27th May 2025

Outline

- Introduction
- Methodology
- Results
- Conclusion

Introduction

Project background and context

- The commercial space race has ushered in a new era where private companies are making space travel more accessible.
- Key players like Virgin Galactic, Rocket Lab, and Blue Origin have made significant contributions, but SpaceX stands out for its innovative approach, particularly in reusing the first stage of its Falcon 9 rockets.
- This reusability significantly reduces launch costs — $62 million compared to $165 million for competitors — making space missions more economically viable. SpaceX's success is attributed in part to its technological advancement in recovering the first stage of the rocket, which is the most expensive part.
- However, not all missions result in successful landings. Factors such as payload, orbit, and mission parameters influence whether the first stage is recovered or sacrificed.

Introduction

Problem Statement

In this project, we assume the role of a data scientist at a fictional competitor company, SpaceY, founded by Allon Musk. The aim is to understand and replicate SpaceX’s cost advantages by answering the following key questions:

1. ***Can we predict the price of a SpaceX launch using publicly available data?***
2. ***Can we accurately predict whether the first stage of a Falcon 9 rocket will land successfully?***

To answer these, we'll leverage exploratory data analysis, interactive dashboards, and machine learning models trained on historical SpaceX launch data.

Section 1

Methodology

Methodology

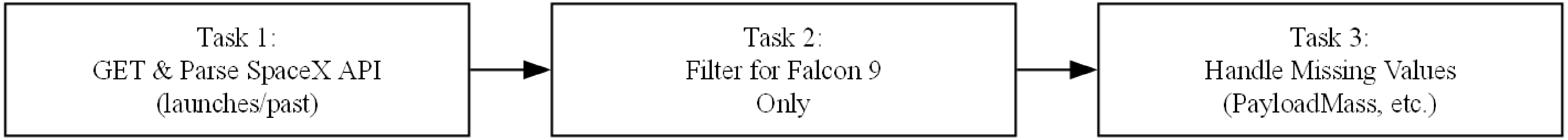
Executive Summary

- Data collection methodology:
  - Collected launch data via SpaceX REST API and web scraped launch tables from Wikipedia.
- Perform data wrangling
  - Cleaned, filtered, and engineered features such as landing outcome and mission success labels.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using *Folium* and *Plotly, Dash*
- Perform predictive analysis using classification models
  - Trained, tuned, and tested multiple models (Logistic Regression, SVM, Decision Tree, KNN) on labeled data.
  - Applied *GridSearchCV* for hyperparameter tuning and selected the best model based on test accuracy.

Data Collection

We gathered SpaceX Falcon 9 launch data using a combination of:

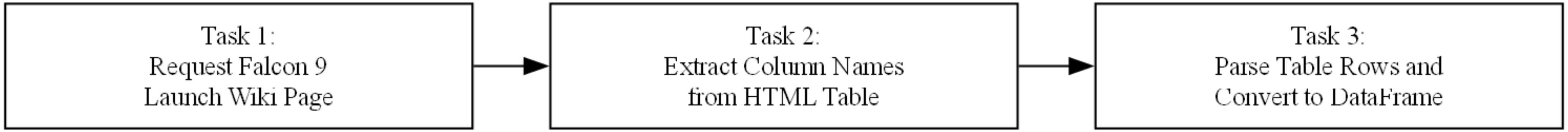
- API extraction from the SpaceX REST API, which provided structured JSON data for past launches.
- Web scraping HTML tables on Wikipedia using BeautifulSoup, extracting additional mission details.
- Data wrangling steps included:
- Normalizing nested JSON data into flat tables.
- Joining auxiliary details from other endpoints like /rockets, /payloads, /cores, and /launchpads.
- Filtering out Falcon 1 launches to focus only on Falcon 9.
- Handling NULL values, especially in ‘PayloadMass’, where missing values were filled using the column’s mean.



Data Collection – SpaceX API

- **API Endpoint Used:** https://api.spacexdata.com/v4/launches/past
- **Tool Used:** requests.get() to perform a GET requestResponse
- **Format:** JSON → normalized using pandas.json_normalize()
- **Filtering:** Only Falcon 9 launches retained
- **Data Cleaning:** Dropped unnecessary or null-heavy columns, Filled missing values (e.g., PayloadMass) with mean imputation, Additional Data: Joined with rockets, cores, payloads, launchpads via ID fields

GitHub link for the notebook



Task 1:
Request Falcon 9
Launch Wiki Page

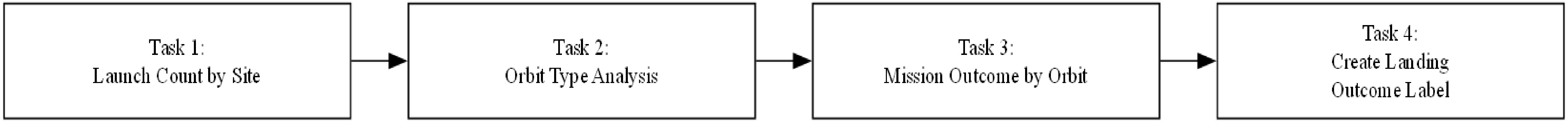
Task 2:
Extract Column Names
from HTML Table

Task 3:
Parse Table Rows and
Convert to DataFrame

Data Collection - Scraping

- **Objective:** Scrape Falcon 9 launch data from Wikipedia
- **Tool Used:** BeautifulSoup (Python web scraping library)
- **Target Page:** Wikipedia Falcon 9 Launches
- **Data Extracted:** Launch dates, payloads, outcomes, etc.
- **Process:** Parse HTML <table> tags, Extract header rows and records, Convert parsed data into a structured Pandas DataFrame

GitHub link for the notebook



Task 1:
Launch Count by Site

Task 2:
Orbit Type Analysis

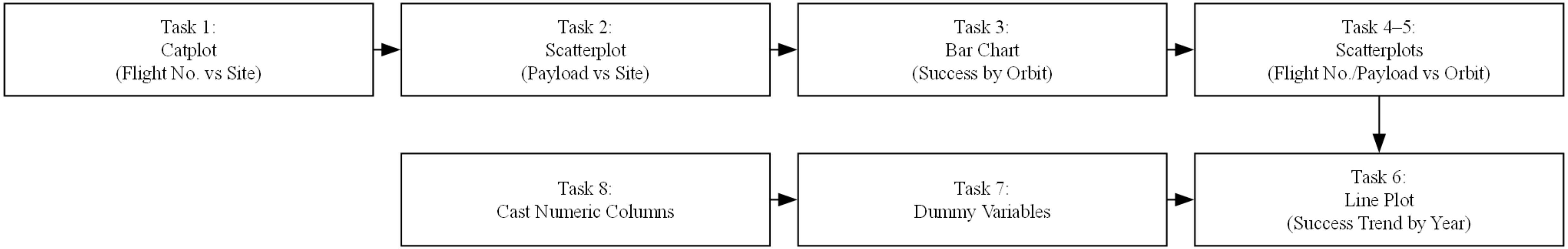
Task 3:
Mission Outcome by Orbit

Task 4:
Create Landing
Outcome Label

Data Wrangling

- **Objective:** Perform EDA and engineer training labels
- **Data Processed Using:** pandas, matplotlib, seaborn
- **Exploratory Data Analysis (EDA):** Analyzed launch site activity, Reviewed orbit types and frequencies, Summarized mission outcomes
- **Label Engineering:** Derived binary classification label: 1 = successful landing, 0 = failure/None, Based on Outcome column

- GitHub link for the notebook



Task 1:
Catplot
(Flight No. vs Site)

Task 2:
Scatterplot
(Payload vs Site)

Task 3:
Bar Chart
(Success by Orbit)

Task 4–5:
Scatterplots
(Flight No./Payload vs Orbit)

Task 6:
Line Plot
(Success Trend by Year)

Task 7:
Dummy Variables

Task 8:
Cast Numeric Columns

EDA with Data Visualization

- **Tools Used:** Pandas, Matplotlib, Seaborn
- **Objective:** Gain insights & prepare data for modelling
- **Why Visualization?:** Understand launch behavior across sites and payloads, Detect patterns, trends, and outliers, Identify feature correlations with success
- **Charts Used:** Catplot: *Flight Number vs Launch Site*, Scatter Plots: *Payload Mass vs Launch Site, Flight Number vs Orbit, Payload Mass vs Orbit*, Bar Chart: *Success rate by orbit type*, Line Plot: *Yearly trend of successful launches.*
- **Feature Engineering:** Created dummy variables for categorical features, Cast numeric features to float64 for consistency

GitHub link for the notebook

EDA with Data Visualization

Chart Type	What It Shows	Why It Was Used
Catplot	Flight Number vs. Launch Site	To observe launch frequency patterns by site over time
Scatterplot	Payload Mass vs. Launch Site	To examine how payload mass varies by site
Bar Chart	Success Rate by Orbit Type	To evaluate which orbits have higher launch success rates
Scatterplot	Flight Number vs. Orbit Type	To identify orbit usage patterns across mission history
Scatterplot	Payload Mass vs. Orbit Type	To detect if certain orbits are linked to heavier/lighter payloads
Lineplot	Yearly Launch Success Trend	To analyze overall improvement or decline in SpaceX launch outcomes over time

EDA with SQL

Summary of SQL Queries Executed

- Task 1: Retrieved unique launch site names
- Task 2: Fetched 5 launch records from sites starting with 'CCA’
- Task 3: Calculated total payload mass for boosters launched by NASA (CRS)
- Task 4: Computed average payload mass for booster version F9 v1.1
- Task 5: Identified the first successful ground pad landing date using MIN()
- Task 6: Listed boosters with successful drone ship landings and payload mass between 4000–6000 kg
- Task 7: Counted total successful and failed mission outcomes
- Task 8: Found booster versions that carried the maximum payload mass using a subquery
- Task 9: Extracted 2015 failure records in drone ship landings, showing month name, booster version, and site
- Task 10: Ranked landing outcome counts between 2010-06-04 and 2017-03-20 in descending order

GitHub link for the notebook

Build an Interactive Map with Folium

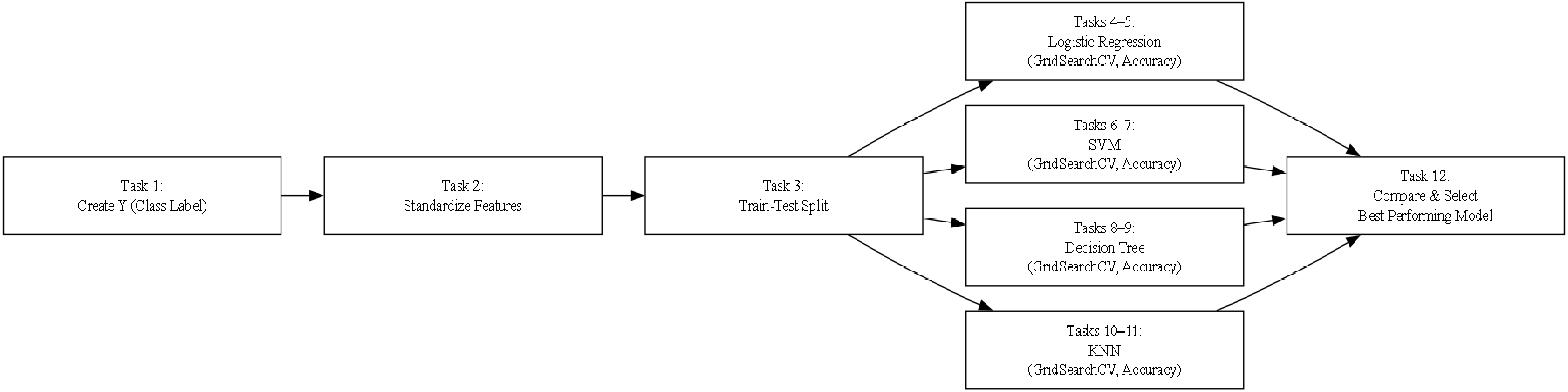
Map Objects Created in Folium

- **folium.Marker:** Placed at each launch site and individual launch results, to pinpoint exact locations and visualize successful vs. failed launches using color-coded icons.
- **folium.Circle:** Drawn around each launch site, to visually highlight the launch site area and improve clarity on zoomed-out views.
- **folium.MarkerCluster:** Clusters individual launch result markers by site, to reduce clutter on the map and enhances user experience when exploring dense data.
- **folium.PolyLine:** Lines from launch sites to coastlines, highways, cities, and railways, to calculate and visualize proximities and possible logistical access points for launch infrastructure.
- **MousePosition Tool:** Used to capture coordinates for nearby features, assisted in manually plotting relevant nearby geographic references like coastlines or roads

Build a Dashboard with Plotly Dash

- **Launch Site Dropdown:** Interactive filter to select a specific site or view all launches. Enables focused exploration of outcomes by site.
- **Pie Chart:** Displays success distribution: total successful launches by site (All Sites) or success vs failure (Single Site). Offers quick visual comparison.
- **Payload Mass Range Slider:** Filters launches based on payload mass range (0–10,000 kg). Helps analyze how payload influences success rates.
- **Scatter Plot:** Shows correlation between payload mass and launch success. Colored by booster version to highlight patterns across booster types.
- **Interactive Callbacks:** Dropdown and slider update charts in real-time. Enhances user-driven data exploration and analysis.

GitHub link for the app



Predictive Analysis (Classification)

Key Steps in Model Development

- **Create Labels:** Extracted binary class column for landing outcome.
- **Standardize Data:** Scaled input features for optimal model performance.
- **Train-Test Split:** Split dataset (80% train, 20% test) with fixed seed.
- **Model Training + Hyperparameter Tuning:** Applied *GridSearchCV* (cv=10) on *Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN)*
- **Evaluation:** Measured test accuracy using *.score()* method.
- **Selection:** Compared test accuracies to determine best-performing model.

GitHub link for the notebook

Results

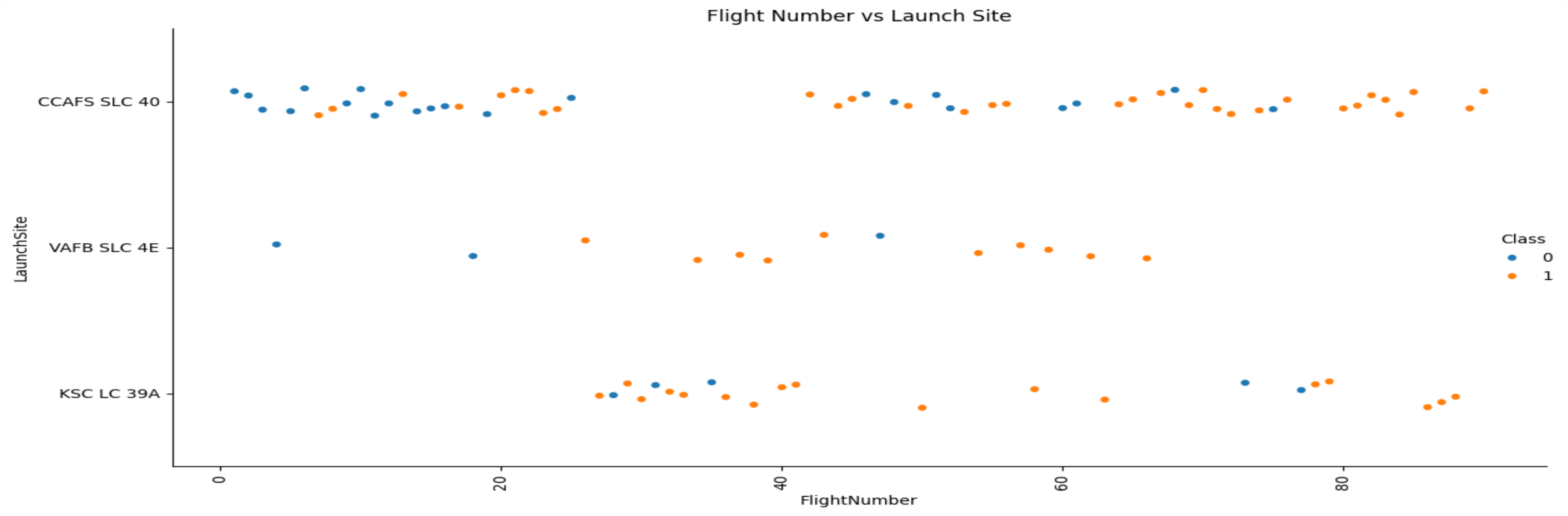
In the subsequent slides, following will be discussed:

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

Section 2

Insights drawn from EDA

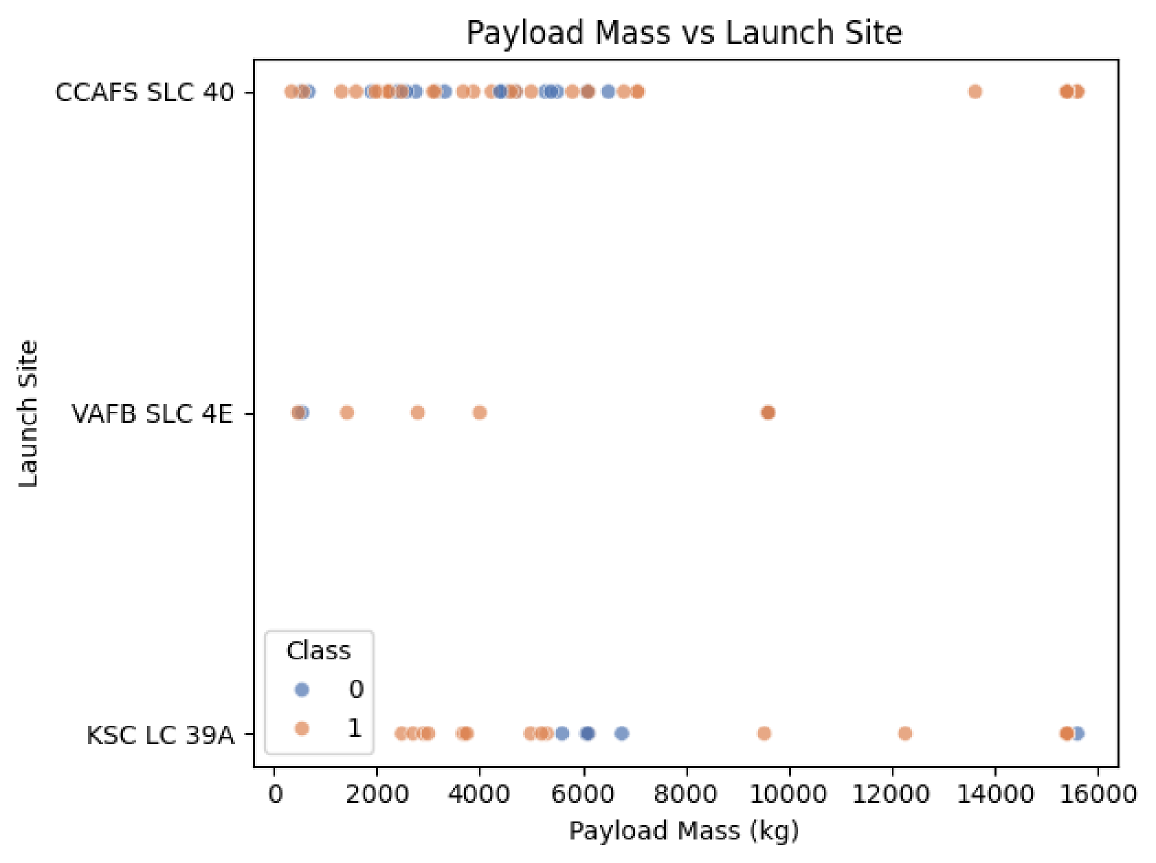
Flight Number vs. Launch Site



- Early Falcon 9 launches mostly failed, while recent ones have consistently succeeded.
- The CCAFS SLC-40 site accounts for nearly half of all launches.
- VAFB SLC-4E and KSC LC-39A show comparatively higher success rates.
- Launch success rates have steadily improved over time with each new mission.

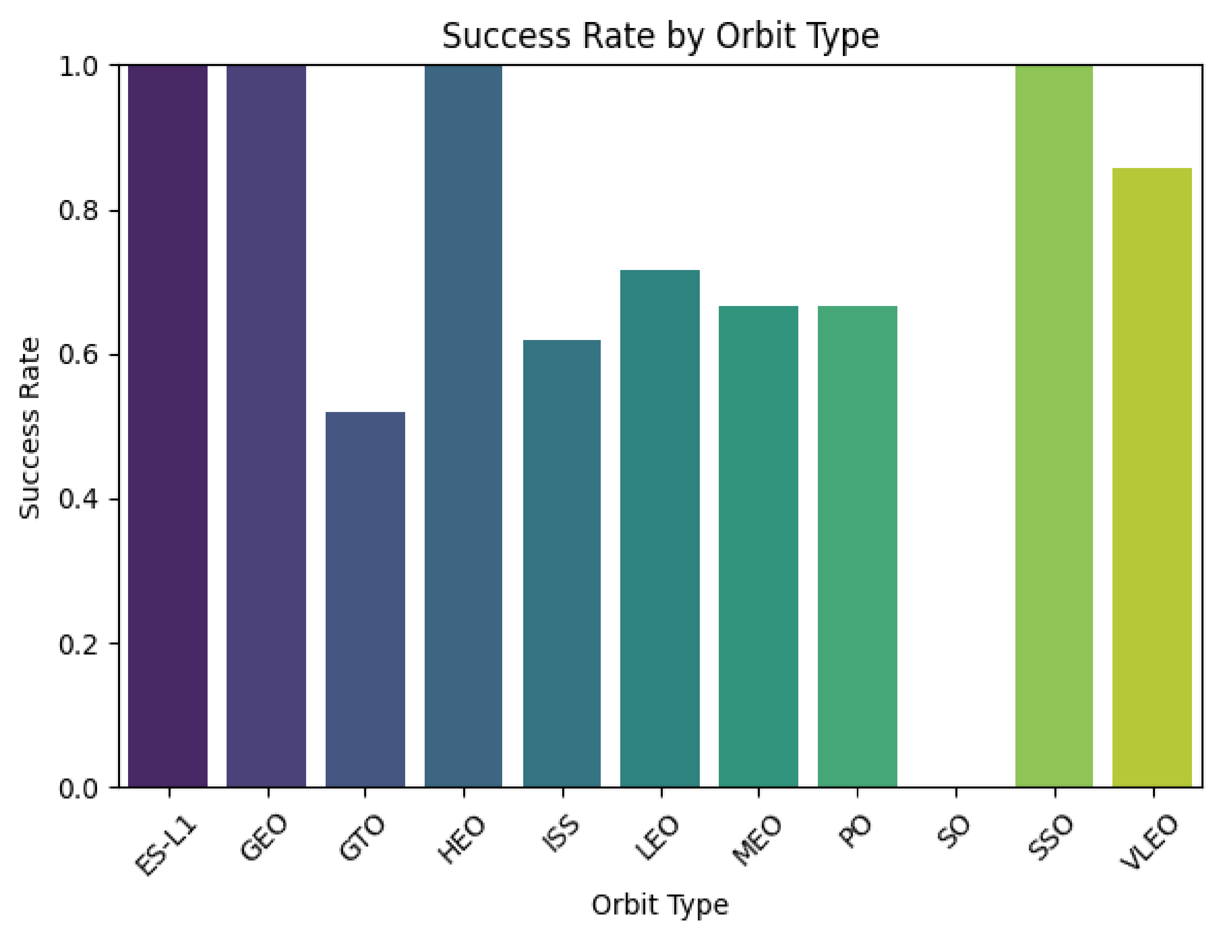
Payload vs. Launch Site

- Across all launch sites, higher payload mass is generally associated with higher success rates.
- Launches carrying payloads over 7000 kg were predominantly successful.
- KSC LC-39A achieved a 100% success rate for payloads below 5500 kg.



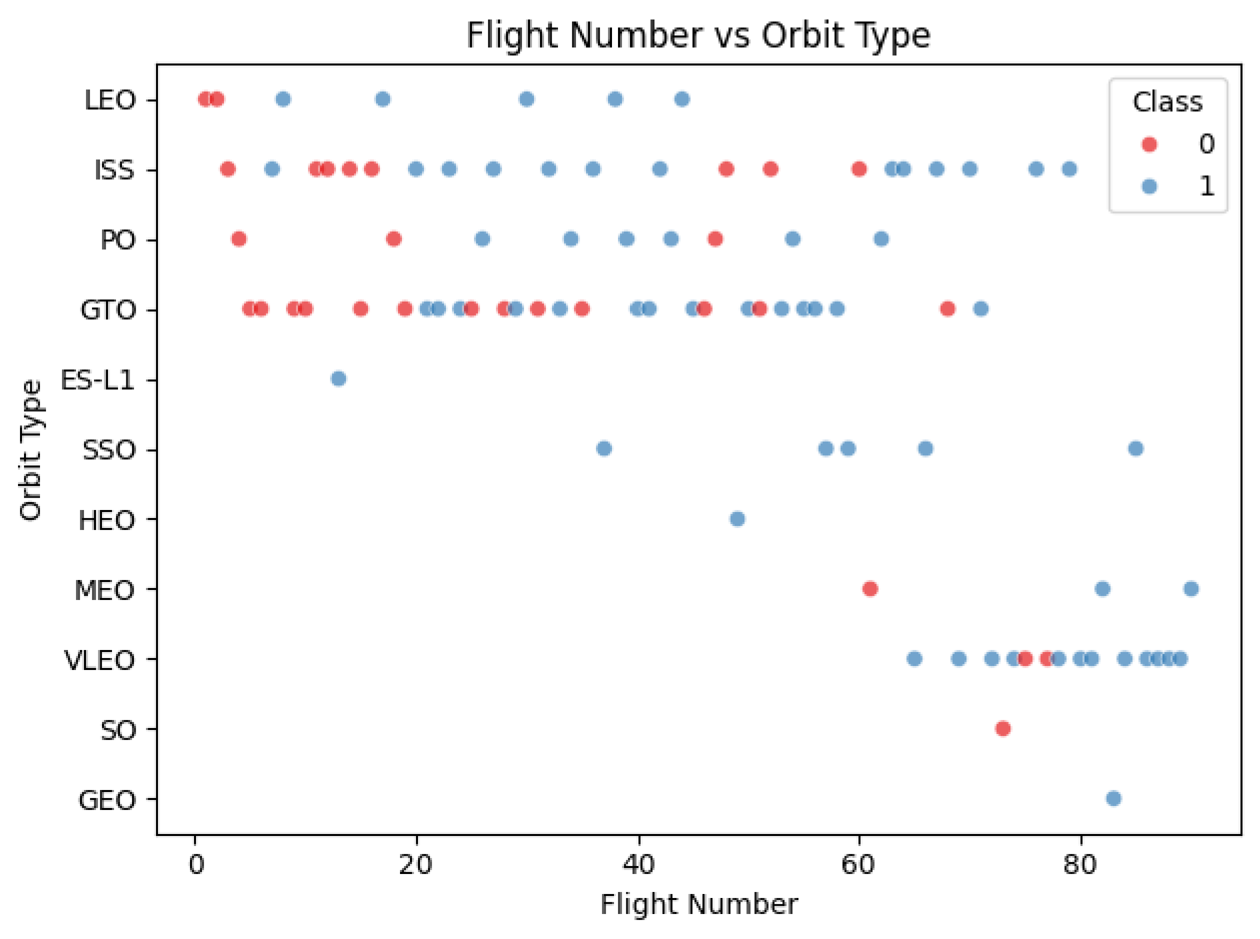
Success Rate vs. Orbit Type

- Orbits with 100% success rate:
  - ES-L1, GEO, HEO, SSO
- Orbits with 0% success rate:
  - SO
- Orbits with success rate between 50% and 85%:
  - GTO, ISS, LEO, MEO, PO



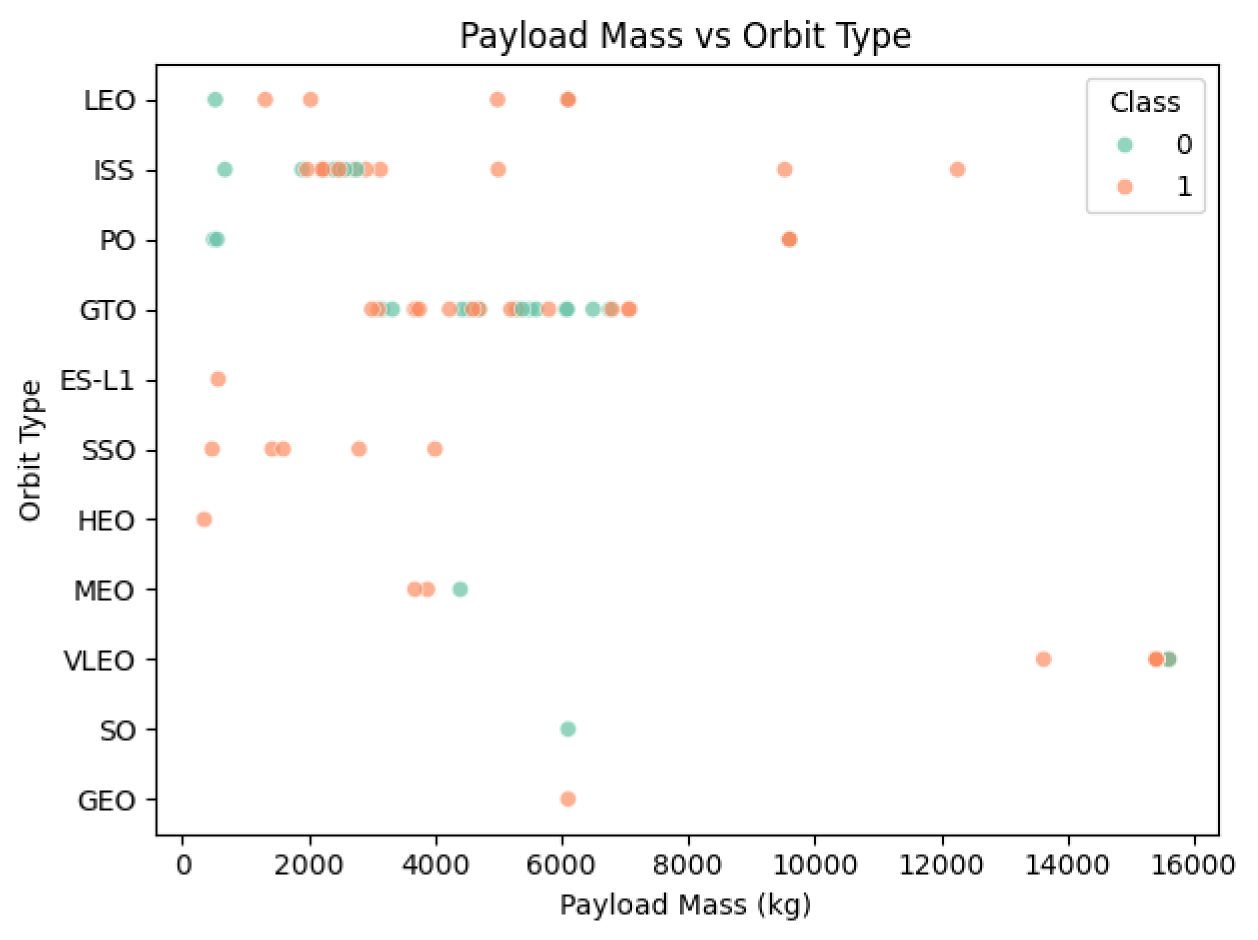
Flight Number vs. Orbit Type

- In LEO, success rates appear to improve with the number of flights, suggesting experience plays a role. However, in the GTO, no clear pattern is observed between flight count and success.
- VLEO has a very good success rate with a large number of flight count.



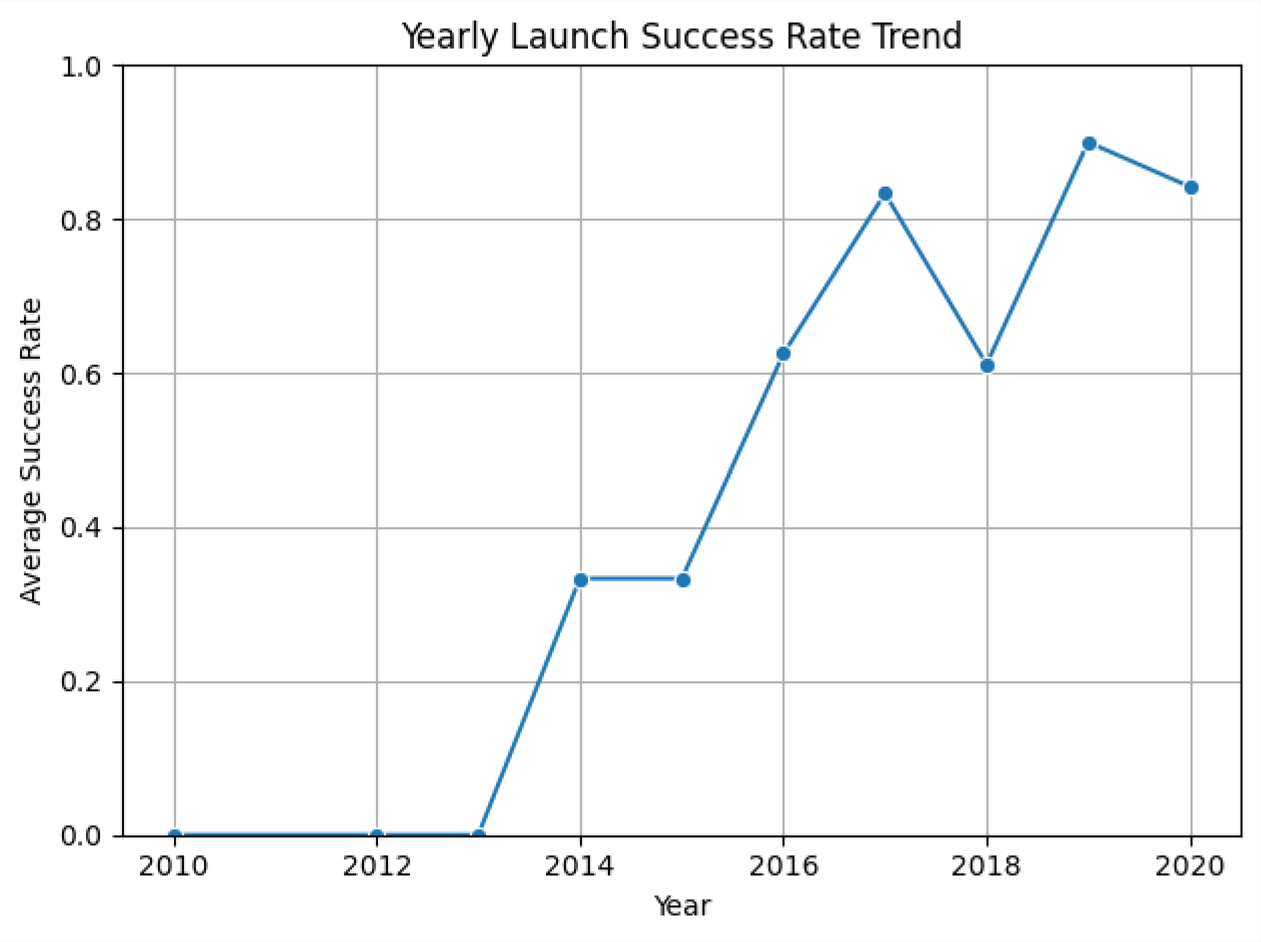
Payload vs. Orbit Type

- Heavy payloads tend to reduce success rates for GTO orbits but positively influence outcomes for Polar LEO and ISS orbits.
- SSO has a 100% success rate with relatively lighter load, whereas VLEO has seen some success with heavier loads.



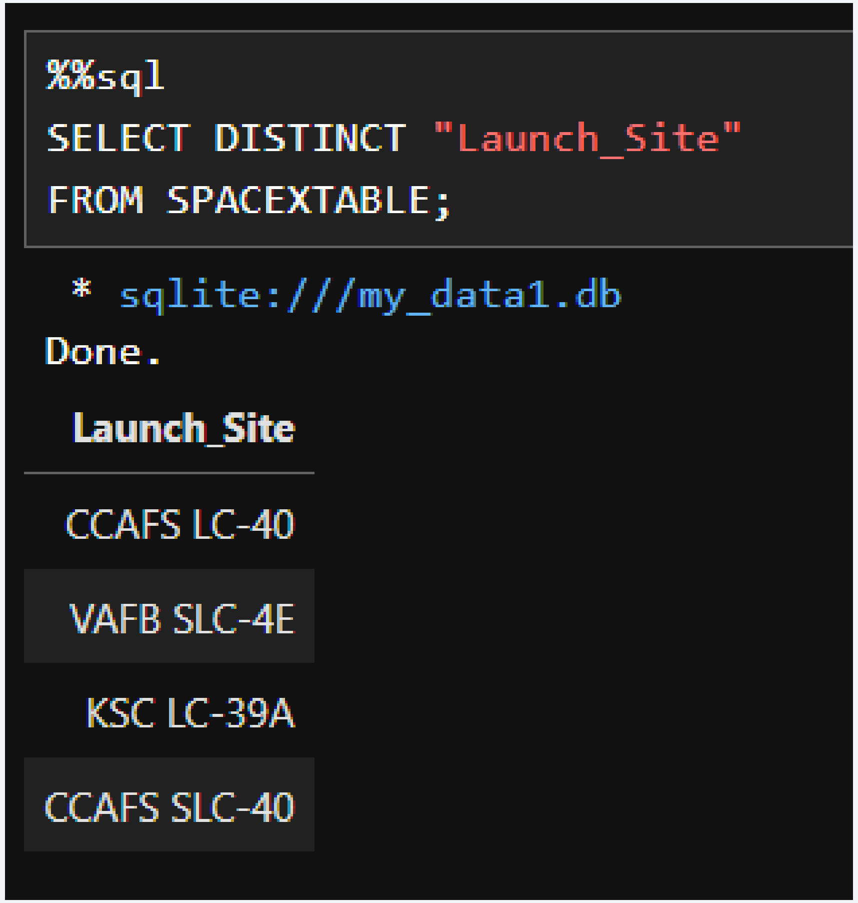
Launch Success Yearly Trend

- The launch success rates keep on increasing year on year since 2013, with a slight dip in 2018 but a subsequent upward trend after that.



All Launch Site Names

- Extracted unique launch site names from the dataset.
- Sample Data Includes:
  - Cape Canaveral Space Launch Complex 40
  - VAFB SLC 4E
  - Vandenberg Air Force Base Space Launch Complex 4E (SLC-4E)
  - Kennedy Space Center Launch Complex 39A (KSC LC 39A)

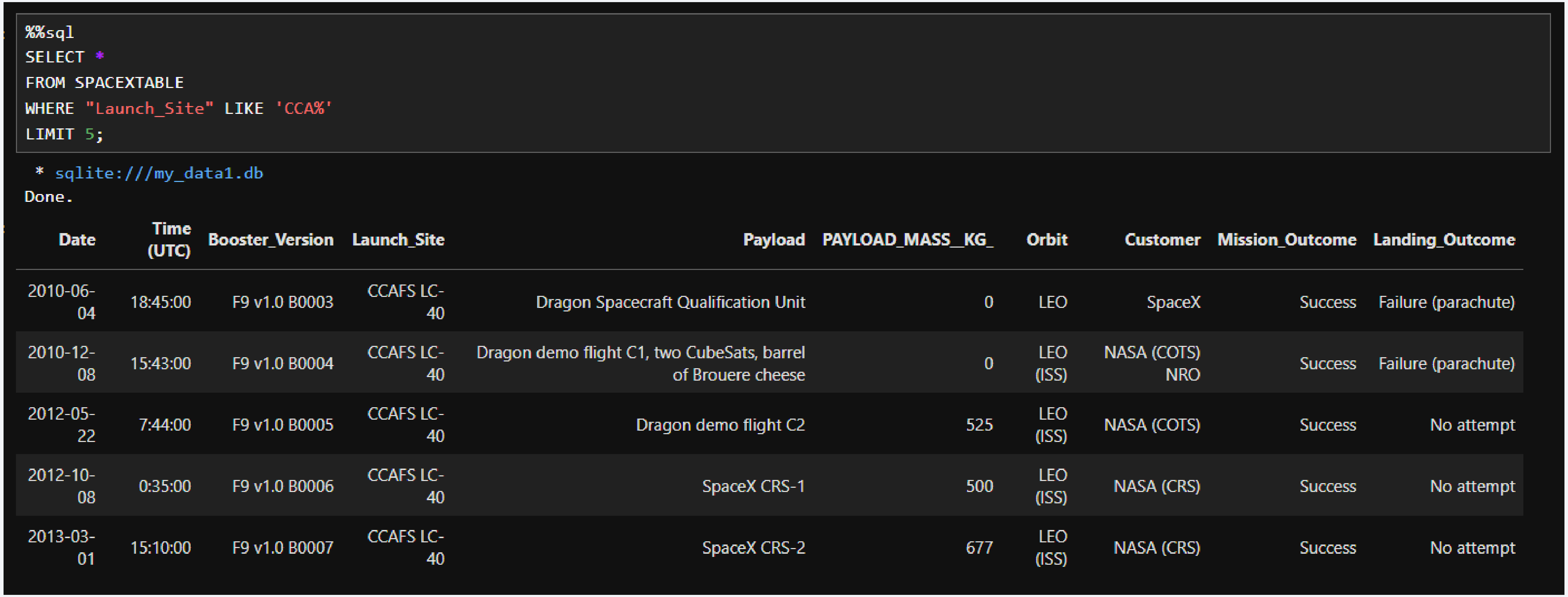


```
%%sql
SELECT DISTINCT "Launch_Site"
FROM SPACEXTABLE;
```

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

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

Launch Site Names Begin with 'CCA'



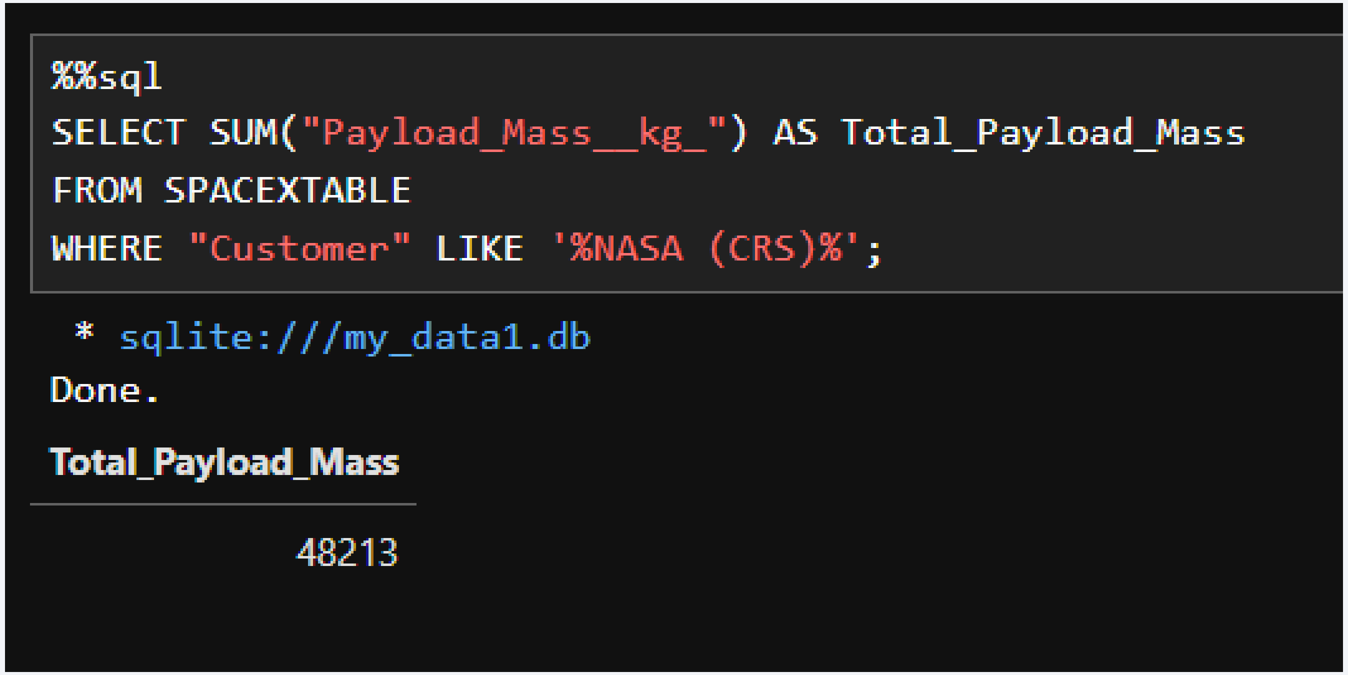
```
%%sql
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
```

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

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

Displaying 5 records where launch sites begin with `CCA`

Total Payload Mass



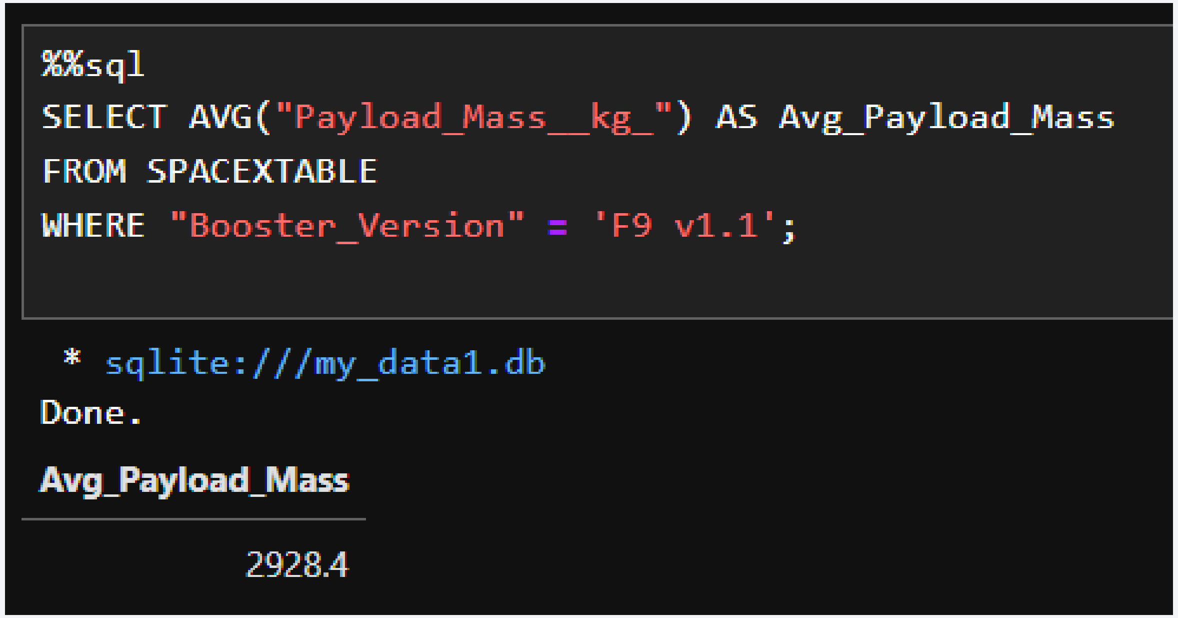
```
%%sql
SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass
FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA (CRS)%';
```

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

Total_Payload_Mass
48213

The total payload mass carried by Falcon 9 rockets is 48213 kgs.

Average Payload Mass by F9 v1.1



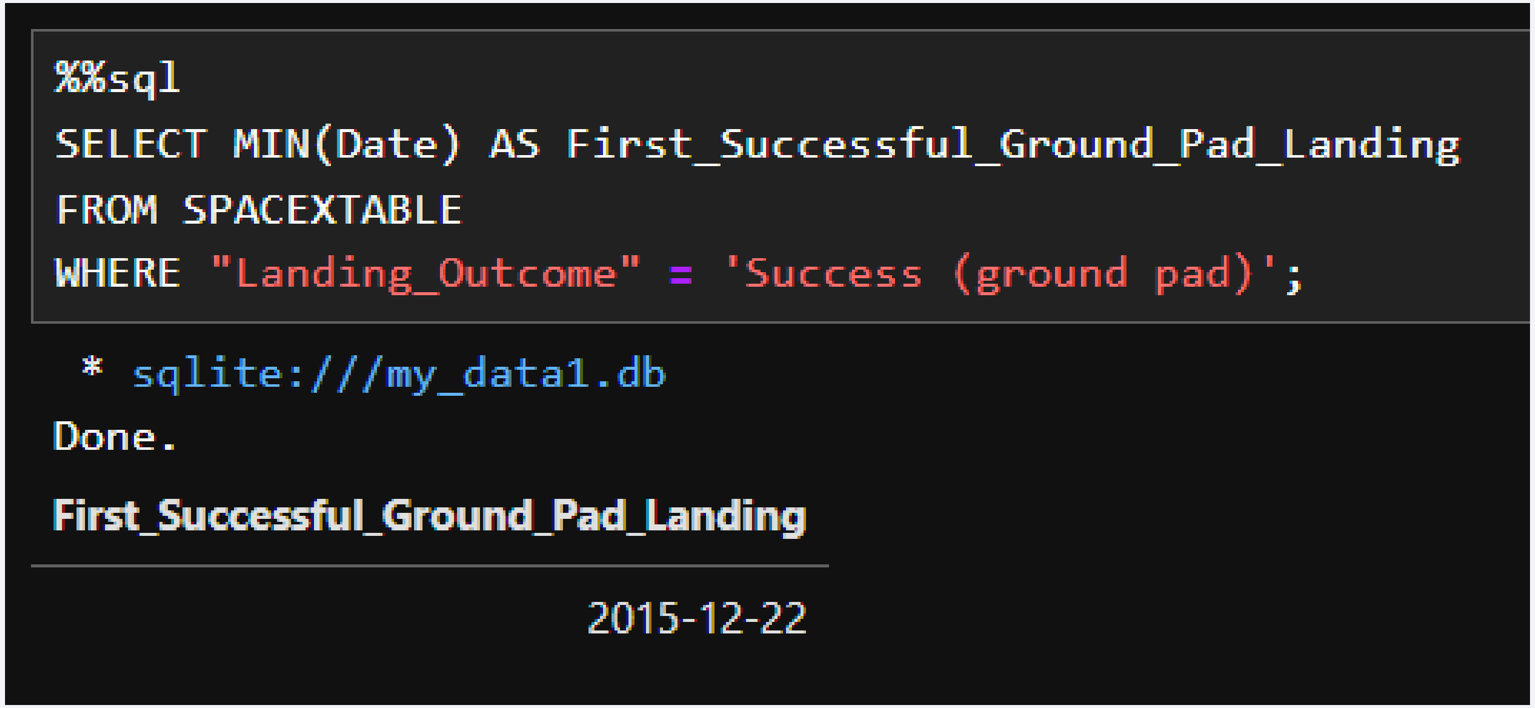
```
%%sql
SELECT AVG("Payload_Mass__kg_") AS Avg_Payload_Mass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

Avg_Payload_Mass
2928.4

The average payload mass carried by booster version F9 v1.1 is 2,928.4 kgs.

First Successful Ground Landing Date

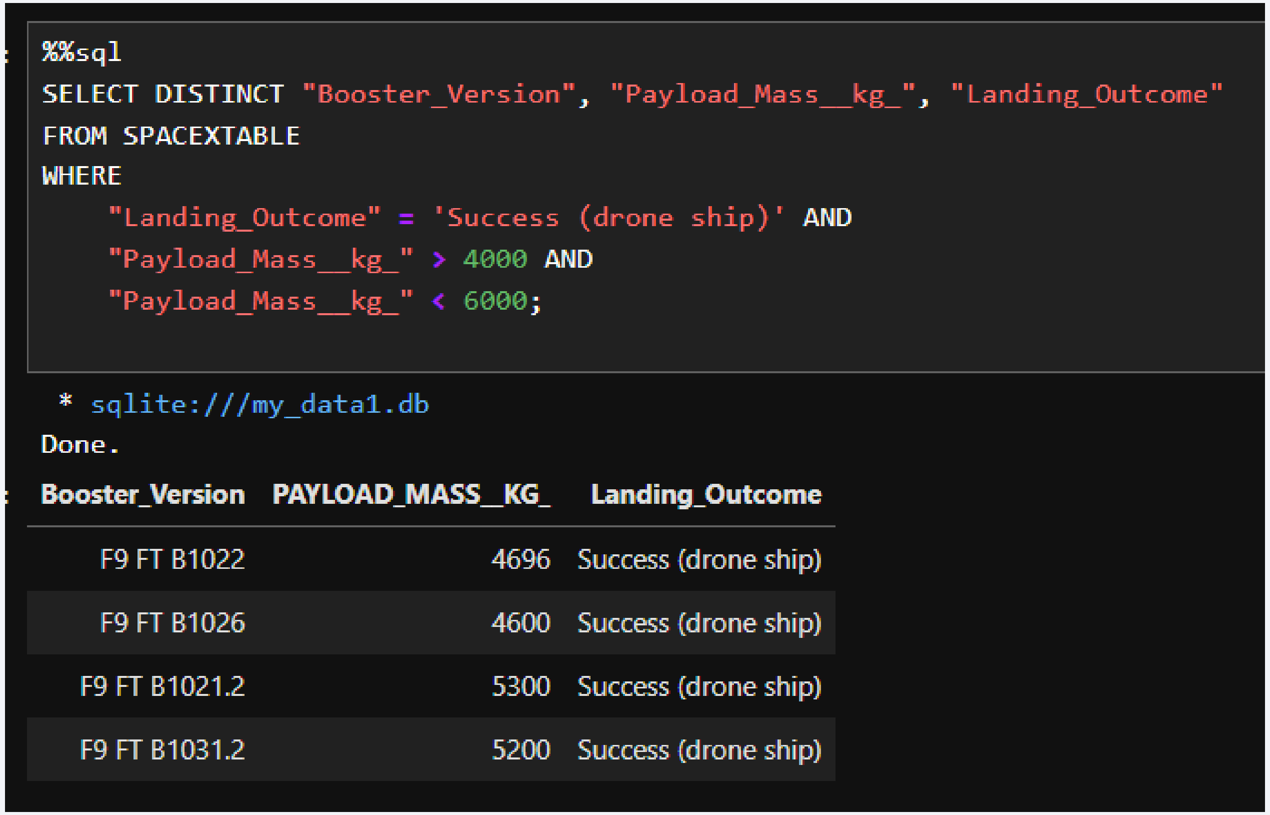


```
%%sql
SELECT MIN(Date) AS First_Successful_Ground_Pad_Landing
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
First_Successful_Ground_Pad_Landing
                         2015-12-22
```

The date of the first successful landing outcome on ground pad was 22nd December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000



```
%%sql
SELECT DISTINCT "Booster_Version", "Payload_Mass__kg_", "Landing_Outcome"
FROM SPACEXTABLE
WHERE
    "Landing_Outcome" = 'Success (drone ship)' AND
    "Payload_Mass__kg_" > 4000 AND
    "Payload_Mass__kg_" < 6000;
```

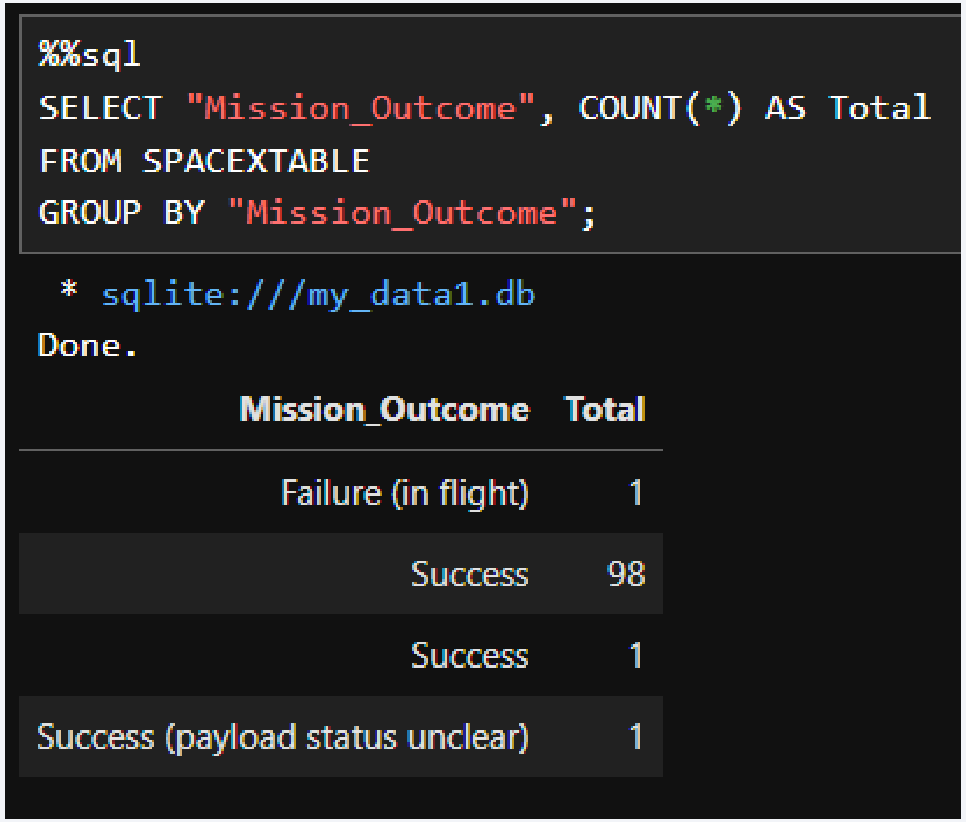
```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Here are the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

Total Number of Successful and Failure Mission Outcomes

The total number of successful mission outcomes is 100, whereas the number of failure mission outcomes is 1.



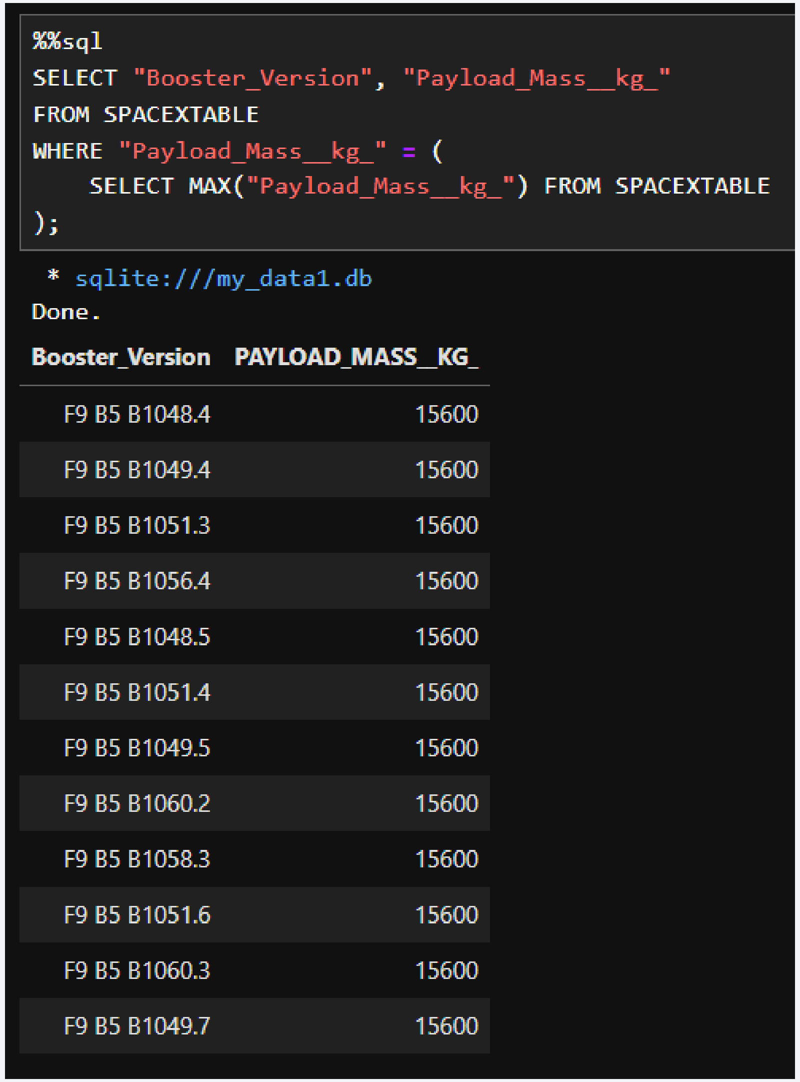
```
%%sql
SELECT "Mission_Outcome", COUNT(*) AS Total
FROM SPACEXTABLE
GROUP BY "Mission_Outcome";
```

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

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

Boosters Carried Maximum Payload

Here are the names of the booster which have carried the maximum payload mass.



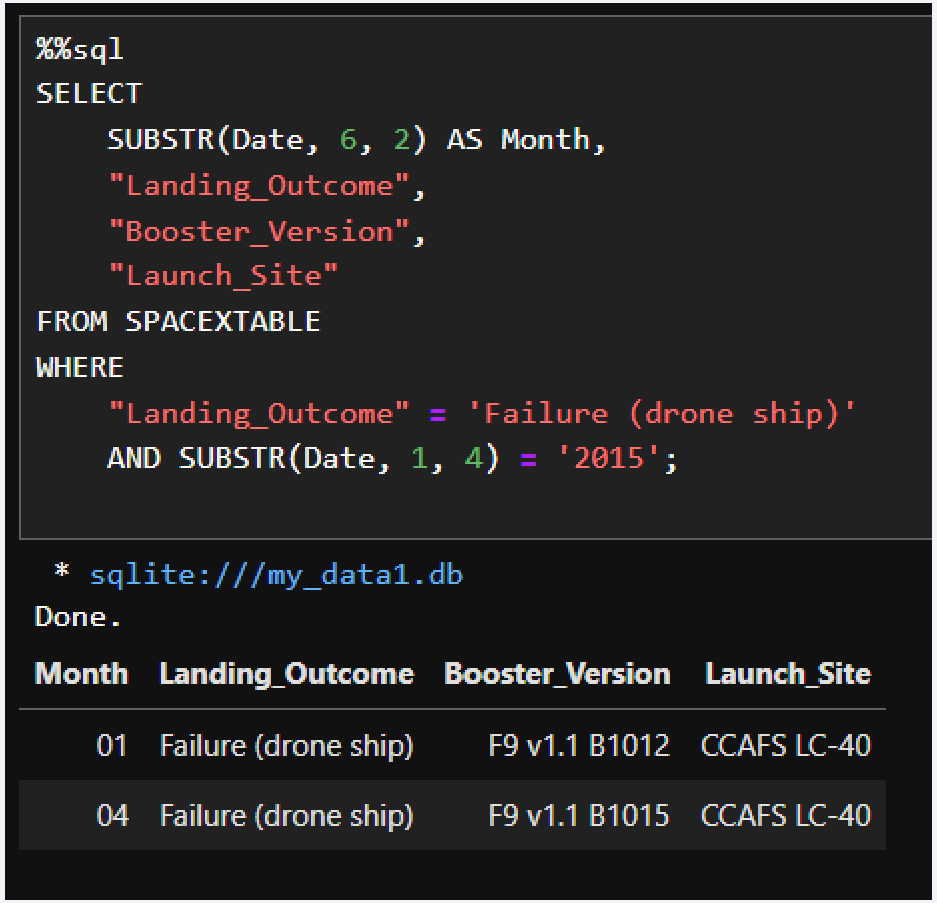
```
%%sql
SELECT "Booster_Version", "Payload_Mass__kg_"
FROM SPACEXTABLE
WHERE "Payload_Mass__kg_" = (
    SELECT MAX("Payload_Mass__kg_") FROM SPACEXTABLE
);
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Here is the list of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015



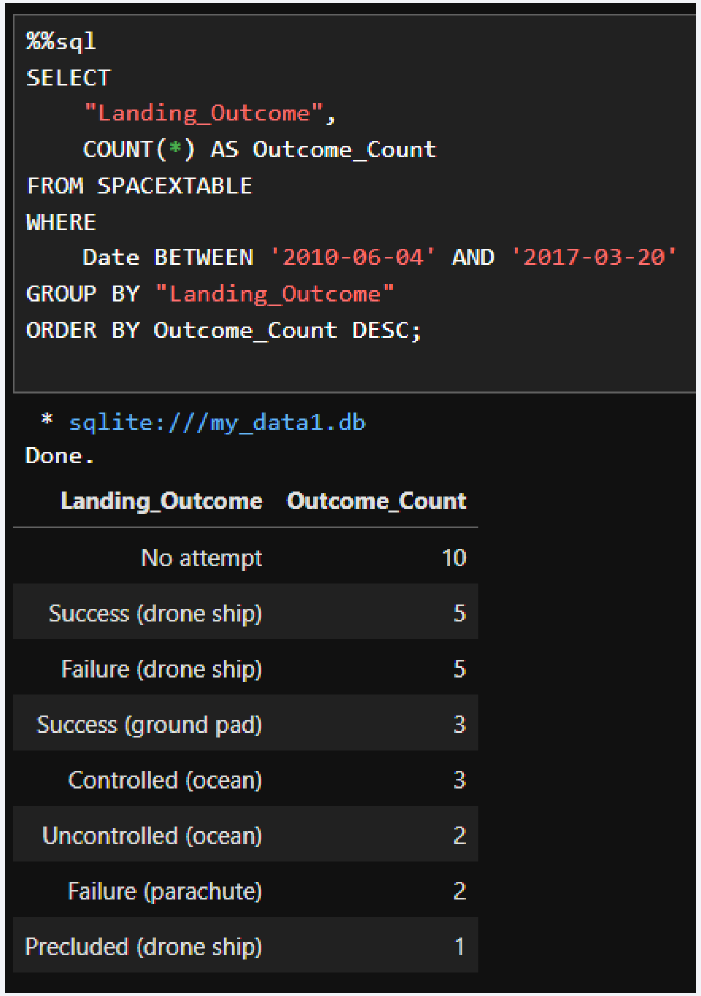
```
%%sql
SELECT
    SUBSTR(Date, 6, 2) AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE
    "Landing_Outcome" = 'Failure (drone ship)'
    AND SUBSTR(Date, 1, 4) = '2015';
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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



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

```
 * 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

Section 3

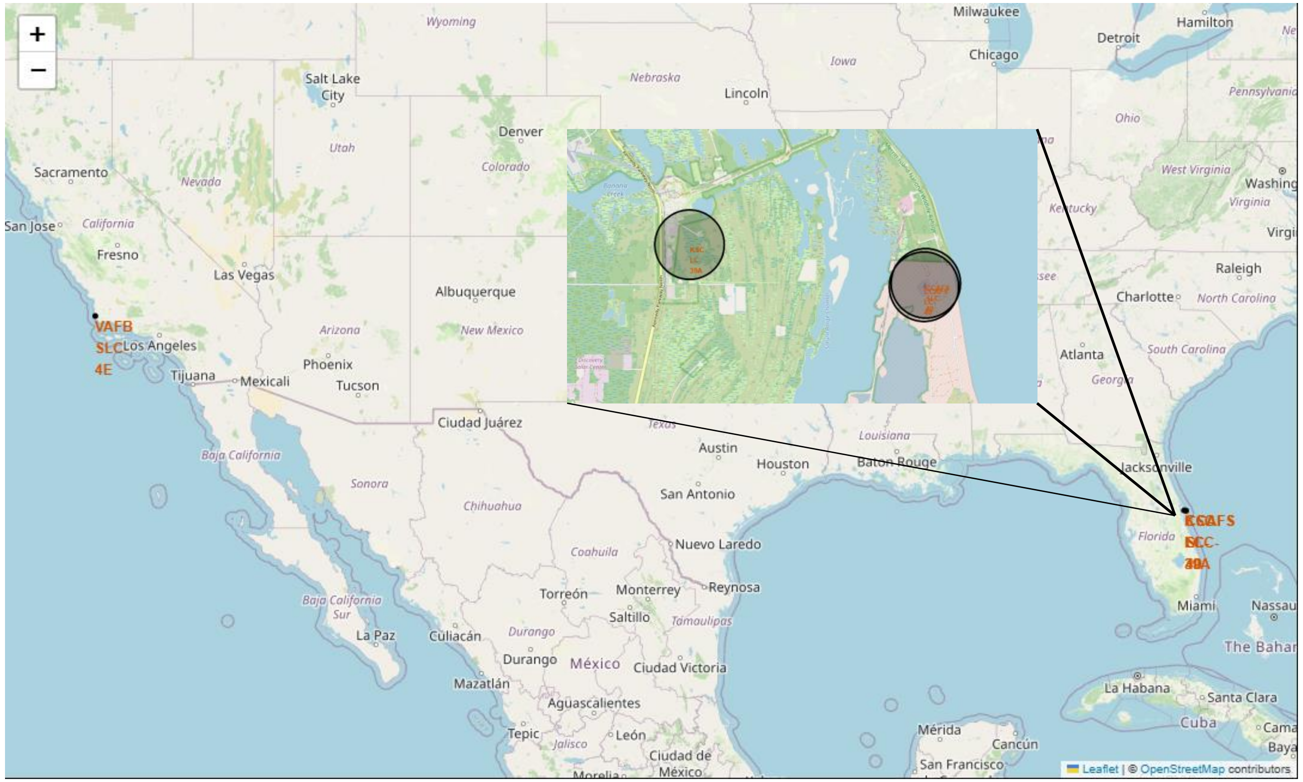
Launch Sites Proximities Analysis

SpaceX Launch Site Locations

Description: This map provides a global overview of SpaceX launch site locations, visualized using the Folium library.

Key Elements in the Map:

- Black Circles: Represent the geographical position of each SpaceX launch site.
- Orange Text Labels: Show the name of each launch site.
- Findings: All launch sites are in the United States, clustered along the East Coast, West Coast, and the Gulf of Mexico. The Kennedy Space Center and Cape Canaveral are situated close together in Florida, highlighting a major launch hub.

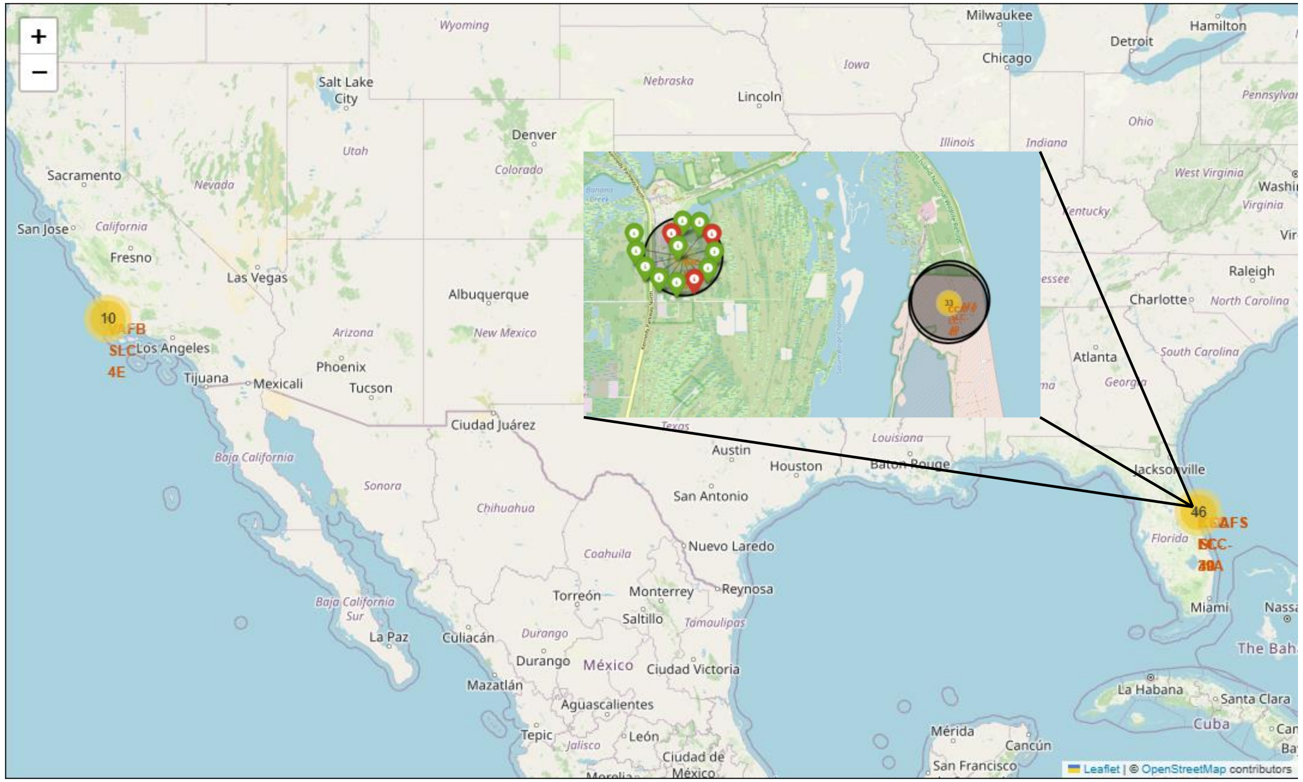


Launch Outcomes by Location

Description: This folium map displays launch sites marked with color-coded indicators representing the outcomes of SpaceX launches. Green markers represent successful launches (class = 1). Red markers represent failed launches (class = 0).

Insights:

- Successful launches (green) dominate most locations, highlighting SpaceX's improved launch performance over time.
- A few scattered red markers indicate early launch failures, particularly in the earlier years of Falcon 9 missions.
- The visual clustering around major launch sites like CCAFS SLC 40 and KSC LC 39A provides intuitive insight into their operational frequency and performance.

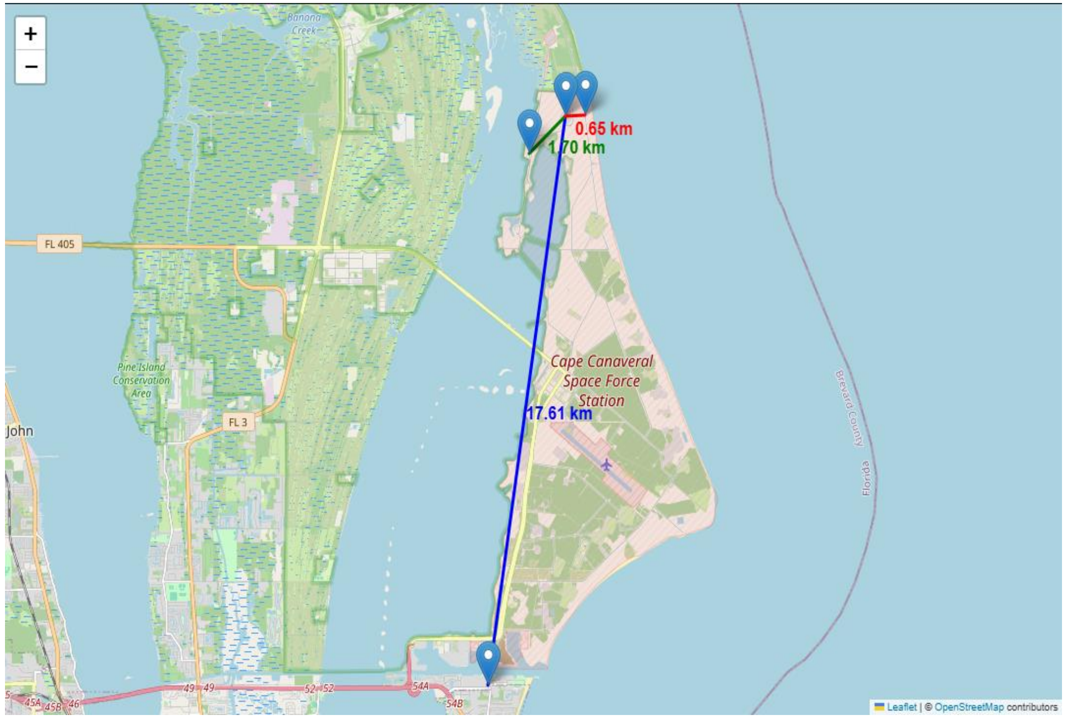


Proximity of Launch Site to Key Infrastructure

Description: This Folium map visualizes the spatial proximity of the CCAFS SLC-40 launch site to nearby geographic and infrastructural features including a city, highway, railway, and the coastline.

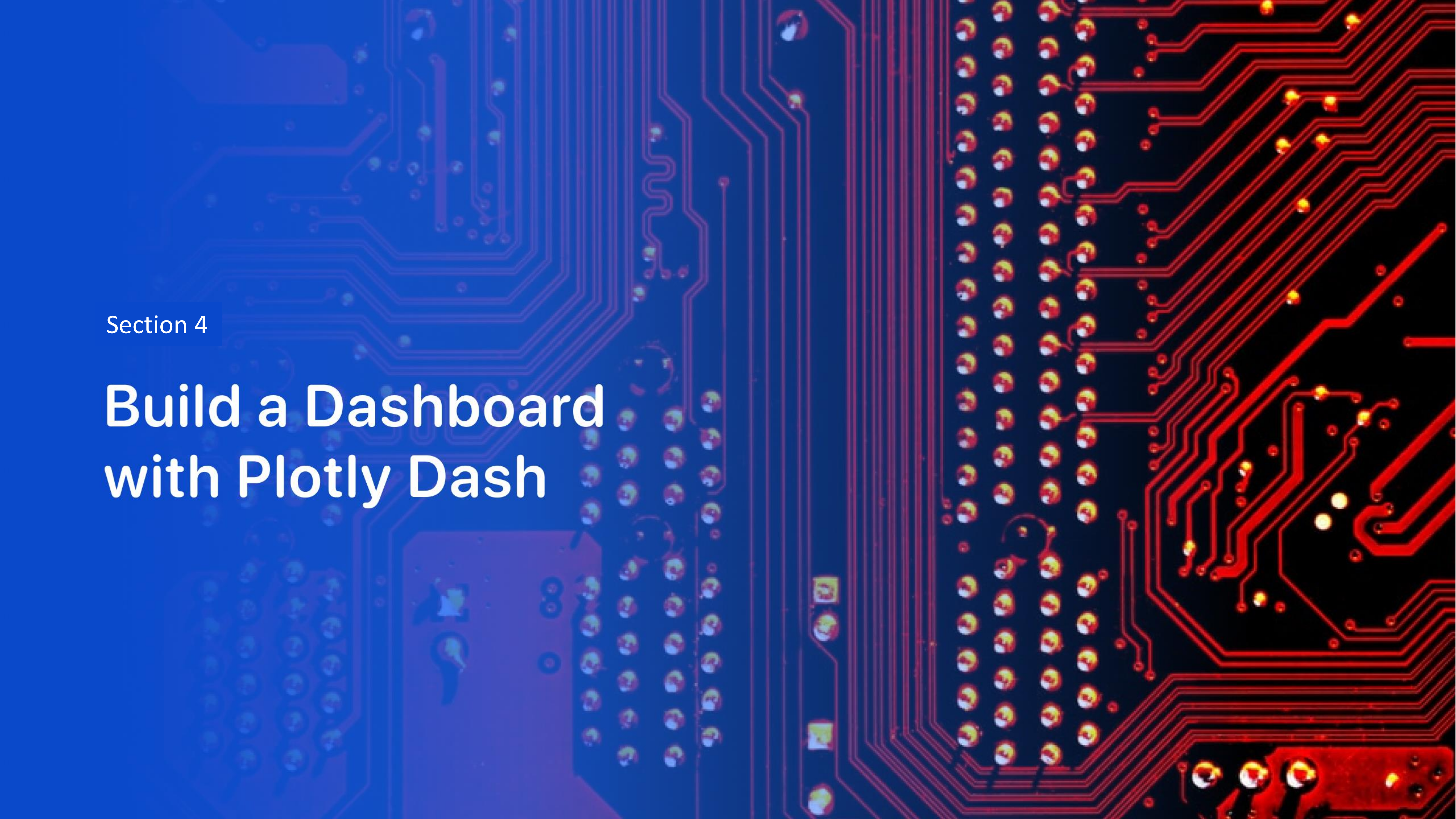
Map Elements & Findings:

- Markers indicate the launch site and each point of interest (city, highway, railway).
- Colored Lines connect the launch site to each feature:
  - Red line to NASA Parkway (highway) – 0.65 km
  - Green line to nearest railway – 1.7 km
  - Blue line to Cape Canaveral – 17.6 km
- **Insights:** The launch site is highly accessible, located just 0.65 km from a major highway and 1.7 km from the nearest railway. While the nearest city is ~17.6 km away, the close infrastructure connectivity makes the site logistically efficient for transportation and operational support.


+
−
Banana Creek
0.65 km
1.70 km
FL 405
Pine Island Conservation Area
FL 3
John
Cape Canaveral Space Force Station
17.61 km
Brevard County
Florida
45A
45B
46
49
49
52
52
54A
54B
Leaflet | © OpenStreetMap contributors

Section 4

Build a Dashboard with Plotly Dash



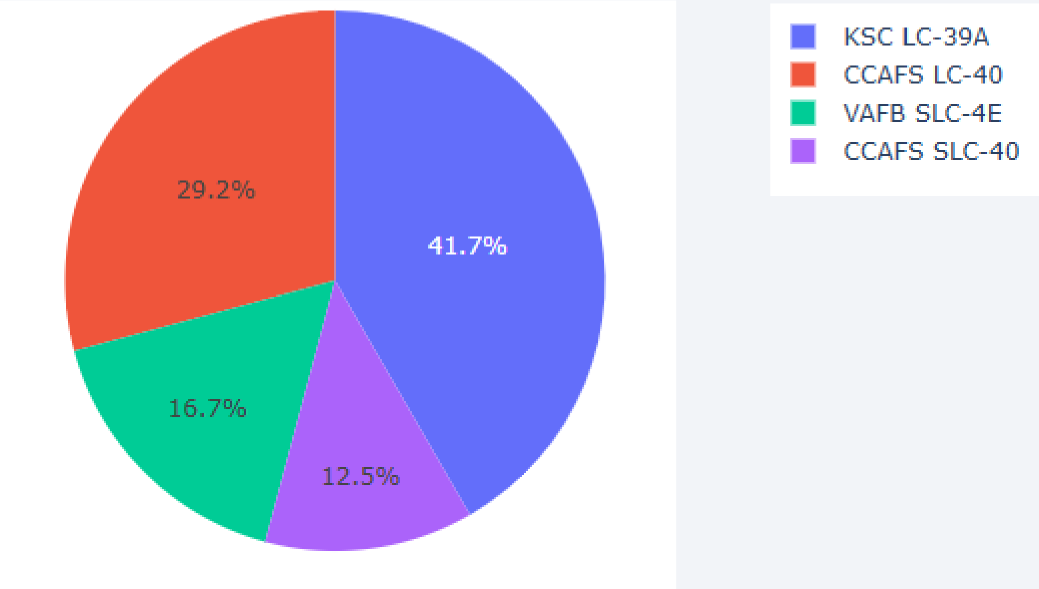
Launch Success Distribution Across All SpaceX Sites

The pie chart visualizes the percentage of successful launches across all launch sites:

- KSC LC-39A: 41.7% of all successful launches – the most active and successful site.
- CCAFS LC-40: 29.2% – the second most used site with strong success record.
- VAFB SLC-4E: 16.7% – consistent but fewer launches.
- CCAFS SLC-40: 12.5% – smallest success share among the four.

Key Insight: Launches from KSC LC-39A dominate the overall success count, indicating higher usage or reliability, while VAFB and CCAFS SLC-40 contribute to a smaller but notable share.

Total Successful Launces by Site

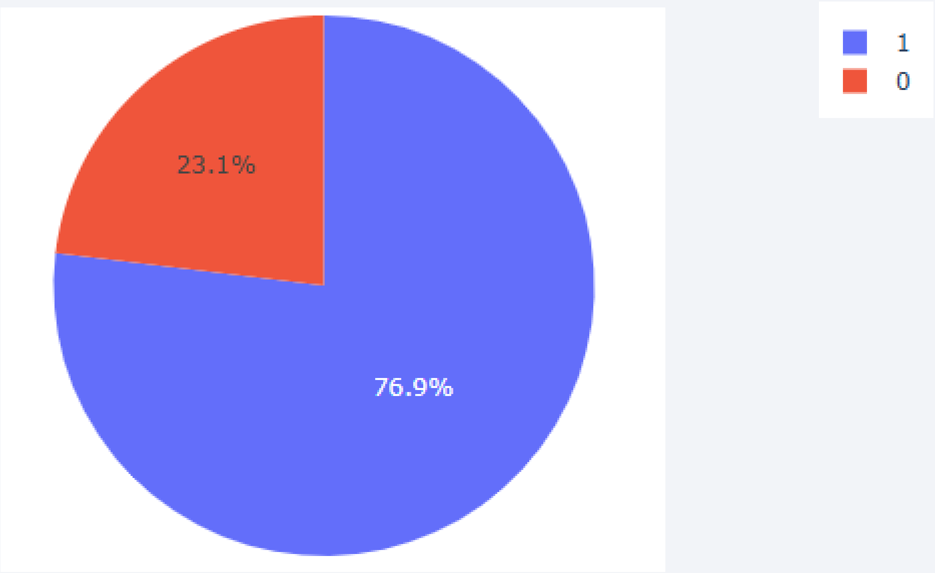


Launch Outcome Breakdown: KSC LC-39A

The pie chart shows the distribution of successful and failed launches from the KSC LC-39A launch site:

- 76.9% of launches were successful (class = 1).
- 23.1% were failures (class = 0).
- **Key Insight:** KSC LC-39A has the ***highest success ratio*** among all launch sites, highlighting its reliability and possibly better operational conditions or upgraded infrastructure.

Launch Outcomes for KSC LC-39A



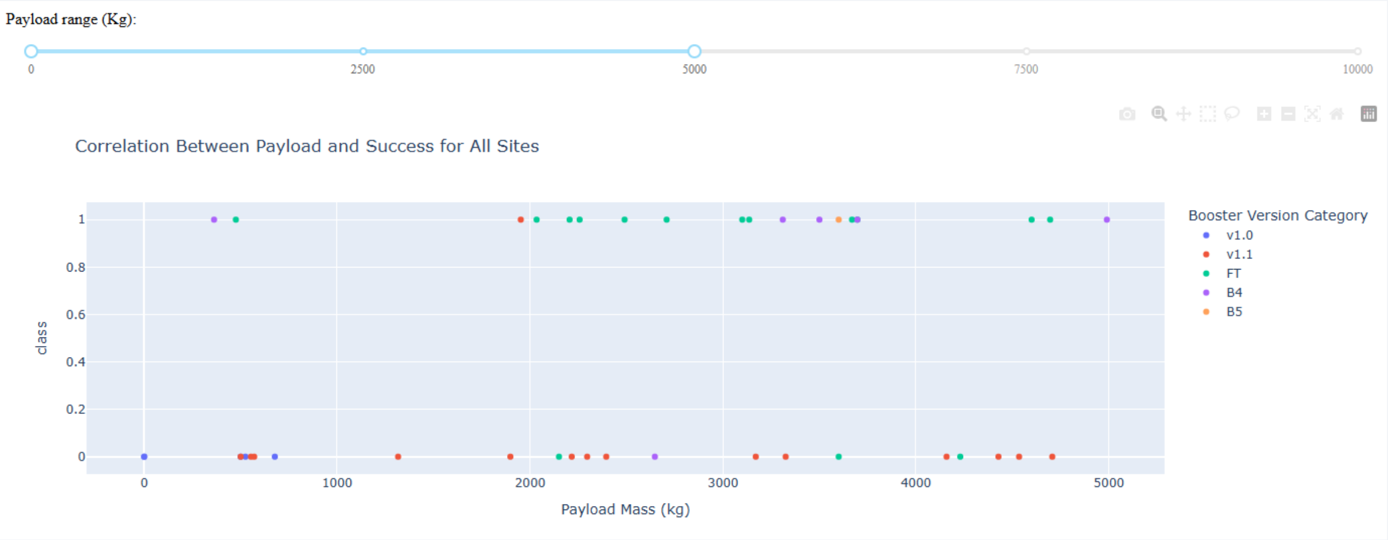
Payload vs. Launch Outcome Scatter Plot for All Sites

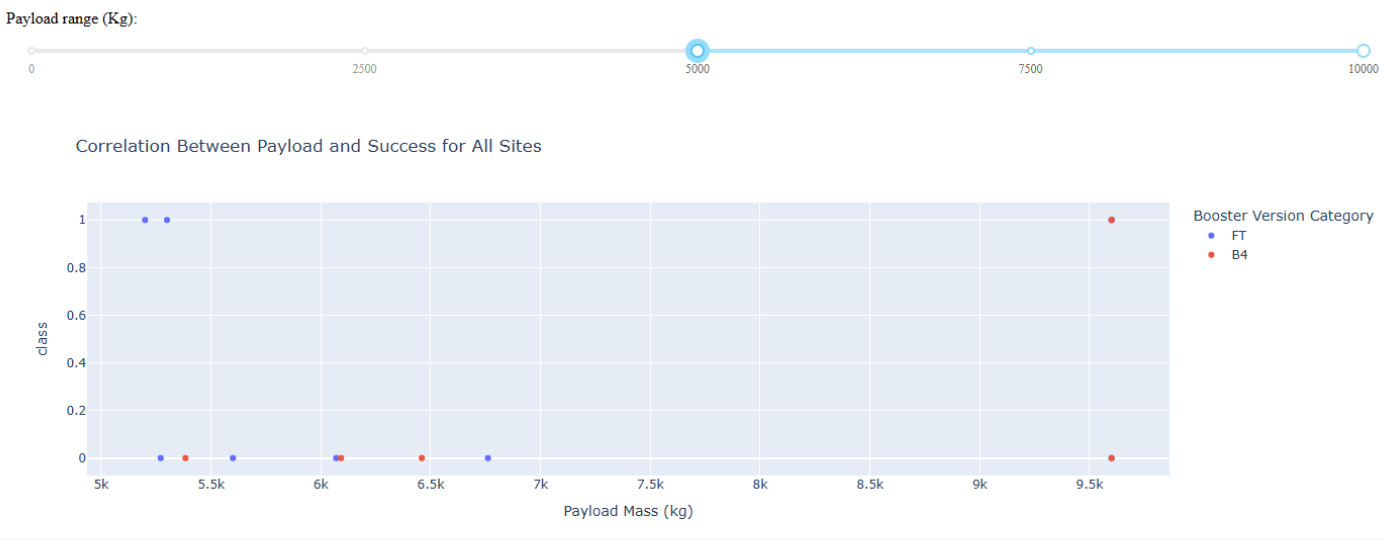
- Overall Pattern: The scatter plots show how launch success (class = 1) or failure (class = 0) varies with payload mass and booster version across all launch sites.

- **0–5000 Kg Payload Range:** Booster versions FT, B4, and B5 show high success rates in this lighter payload range. Launches are more frequent and show a higher concentration of successful outcomes. Failures are spread across versions, but v1.0 and v1.1 seem less reliable.

- **5000–10000 Kg Payload Range:** Fewer launches are observed in this heavier payload range. Booster versions FT and B4 dominate. Success rates are mixed; even reliable boosters like FT see failures. Indicates increased risk with heavier payloads.

- **Conclusion:** FT booster version shows consistent performance across payloads. Lighter payloads (0–5000 Kg) have more launches and higher success rates. Heavier payloads (5000–10000 Kg) are less frequent and riskier, despite advanced booster versions.





Section 5

Predictive Analysis (Classification)

Classification Accuracy

Model Evaluation & Comparison:
All models had similar Test Accuracy (~83.3%), making it hard to differentiate performance, thus applied 5-Fold Cross-Validation to assess consistency across data splits.

Cross-Validation Accuracy:

- Logistic Regression: 80.0%; SVM: 80.0%; KNN: 76.7%; Decision Tree: 67.8%

Insights:

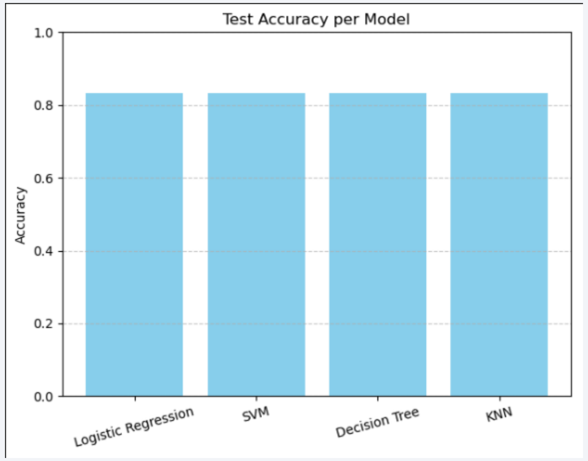
Decision Tree underperformed → least reliable.

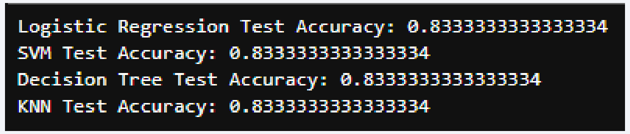
Logistic Regression & SVM tied on accuracy.

Which to Choose?: Logistic Regression: Simpler, faster, more interpretable.

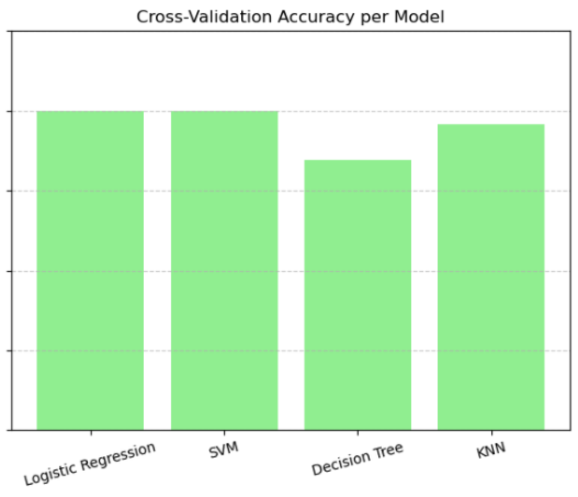
SVM: May perform better with complex, non-linear data (with tuning).

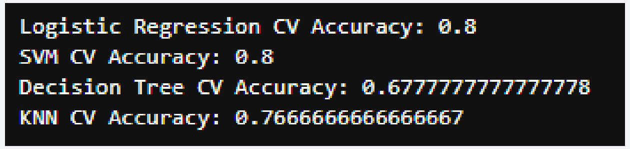
Preferred: Logistic Regression *for its simplicity and stability.*





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





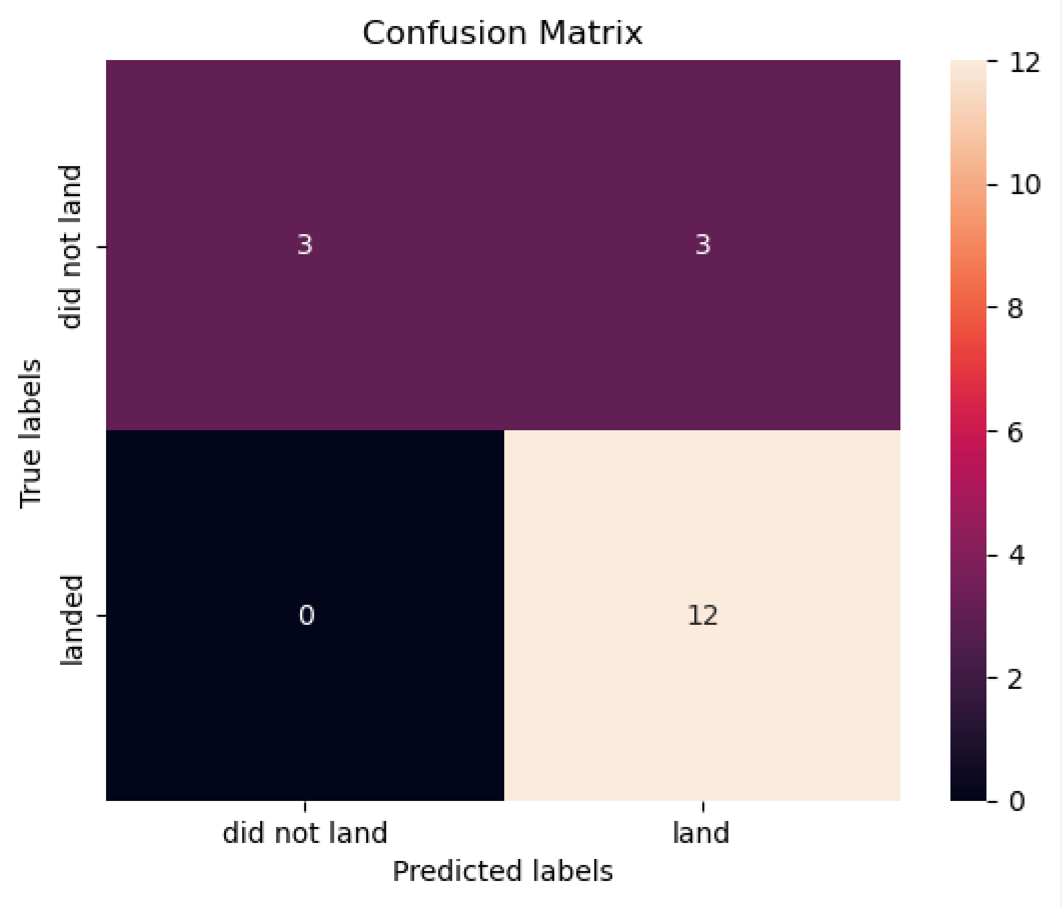
```
Logistic Regression CV Accuracy: 0.8
SVM CV Accuracy: 0.8
Decision Tree CV Accuracy: 0.6777777777777778
KNN CV Accuracy: 0.7666666666666667
```

Confusion Matrix

Interpretation (Logistic Regression)

- True Positives (12): Landed launches correctly predicted.
- True Negatives (3): Non-landings correctly predicted.
- False Positives (3): Non-landings misclassified as landings.
- False Negatives (0): No successful landings were missed.

Insights: This confusion matrix reflects the performance of the Logistic Regression model. No false negatives means it accurately detects all actual landings. A few false positives implies a slight tendency to overpredict landings. Overall, the model shows strong recall and is reliable in identifying success.



Conclusions

- The **Logistical Regression** model performed best overall for this classification task.
- **Lower payload mass** is generally associated with **higher launch success rates**.
- All launch sites are located near **coastlines**, and most are positioned close to the Equator to take advantage of rotational boost.
- The launch success rate has **improved consistently over the years**, indicating advancements in technology and reliability.
- **KSC LC-39A** stands out as the most reliable launch site, with the highest success rate among all sites.
- Missions to **ES-L1, GEO, HEO, and SSO** orbits have shown a **100% success rate**, suggesting these orbits are well-supported by current capabilities.

Answers to the Problem Statement

1. **Can we predict the price of a SpaceX launch using publicly available data?**

   Yes, to a fair extent, though there are some limitations. Public data like payload mass, orbit type, and launch site can help estimate launch costs, since heavier payloads and more complex orbits (like GTO or L1) generally cost more. While SpaceX doesn't share exact prices for individual launches, these features offer good clues, allowing us to make reasonably accurate estimates. That said, for more precise predictions, we'd need details that aren't publicly available, such as contract specifics, reusability savings, and mission configurations.

2. **Can we predict whether the first stage of a Falcon 9 will land successfully?**

   Yes, and with pretty good accuracy. Using data like payload, orbit, launch site, and booster version, we built classification models that can predict landing success with around 83% accuracy on test data. The results held up with cross-validation too, showing a consistent performance of about 80%, which suggests the predictions are reliable. Among the models tested, Logistic Regression stood out, which made very few false predictions for failed landings, making it a dependable choice for this kind of task.

Thank you!