IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yashada Rane
08-06-2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Visualization
  - Building an interactive map with Folium
  - Building a dashboard with Plotly
  - Predictive Analysis

Summary of all results

1. Exploratory Data Analysis Results
2. Geospatial Analytics
3. Interactive Dashboard
4. Predictive Analysis of classification models

Introduction

- Project background and context

SpaceX revolutionized space travel by reducing launch costs through rocket stage reuse. Predicting whether the first stage will land successfully can influence launch pricing. Using open-source data and machine learning, we aim to determine what affects landing success.

- Problems you want to find answers

1. Which factors are behind the failure of landing?
2. Will the rockets land successfully?
3. What is the accuracy of successful landing?
4. Do payload weight, launch site, flight history, or orbit type affect landings?
5. Is there a trend of increasing landing success?
6. Which classification model best predicts success?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Used the SpaceX REST API and scraped Wikipedia.
- Perform data wrangling
- 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 classification models (LogReg, SVM, Decision Tree, KNN) to assess outcome predictors.

Data Collection

Requested launch data from SpaceX’s API.

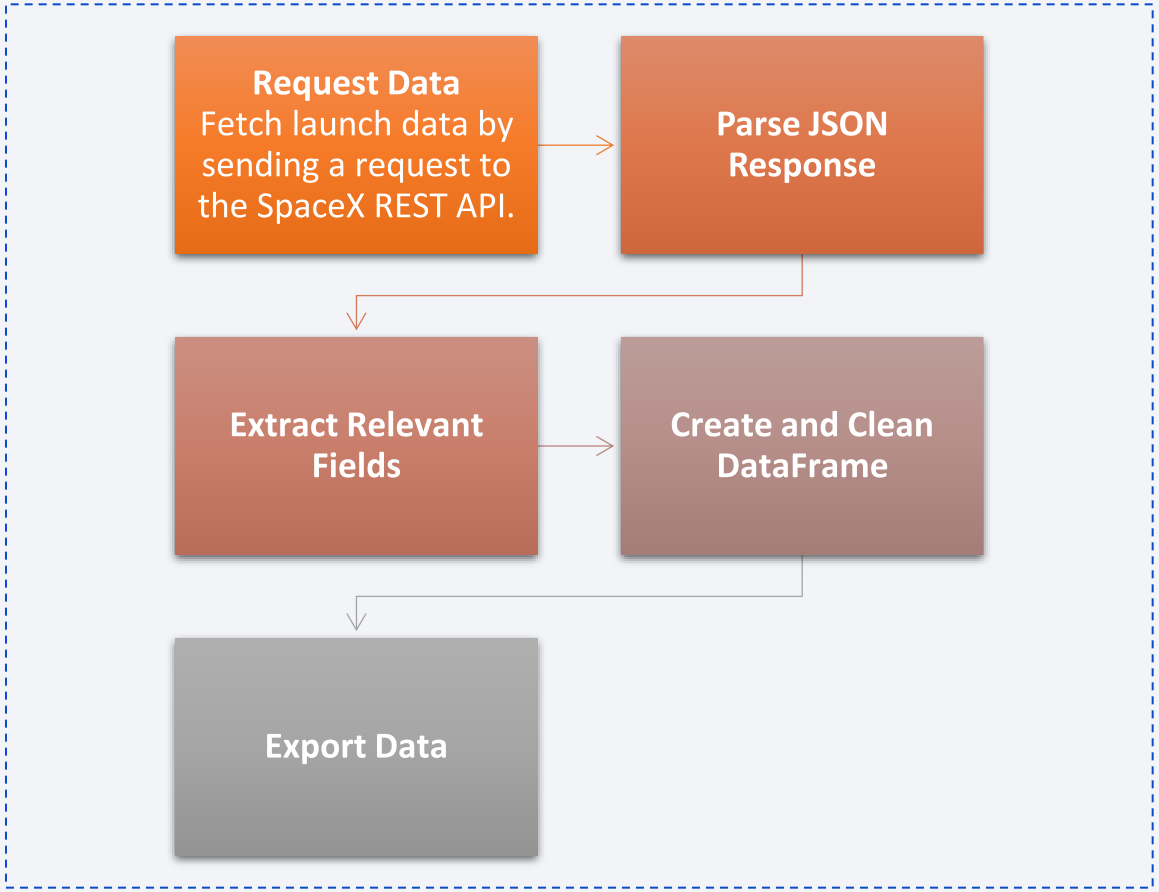
Parsed JSON responses into structured DataFrames.

Filtered for Falcon 9 missions only.

Handled missing values and exported clean data.

Data Collection – SpaceX API

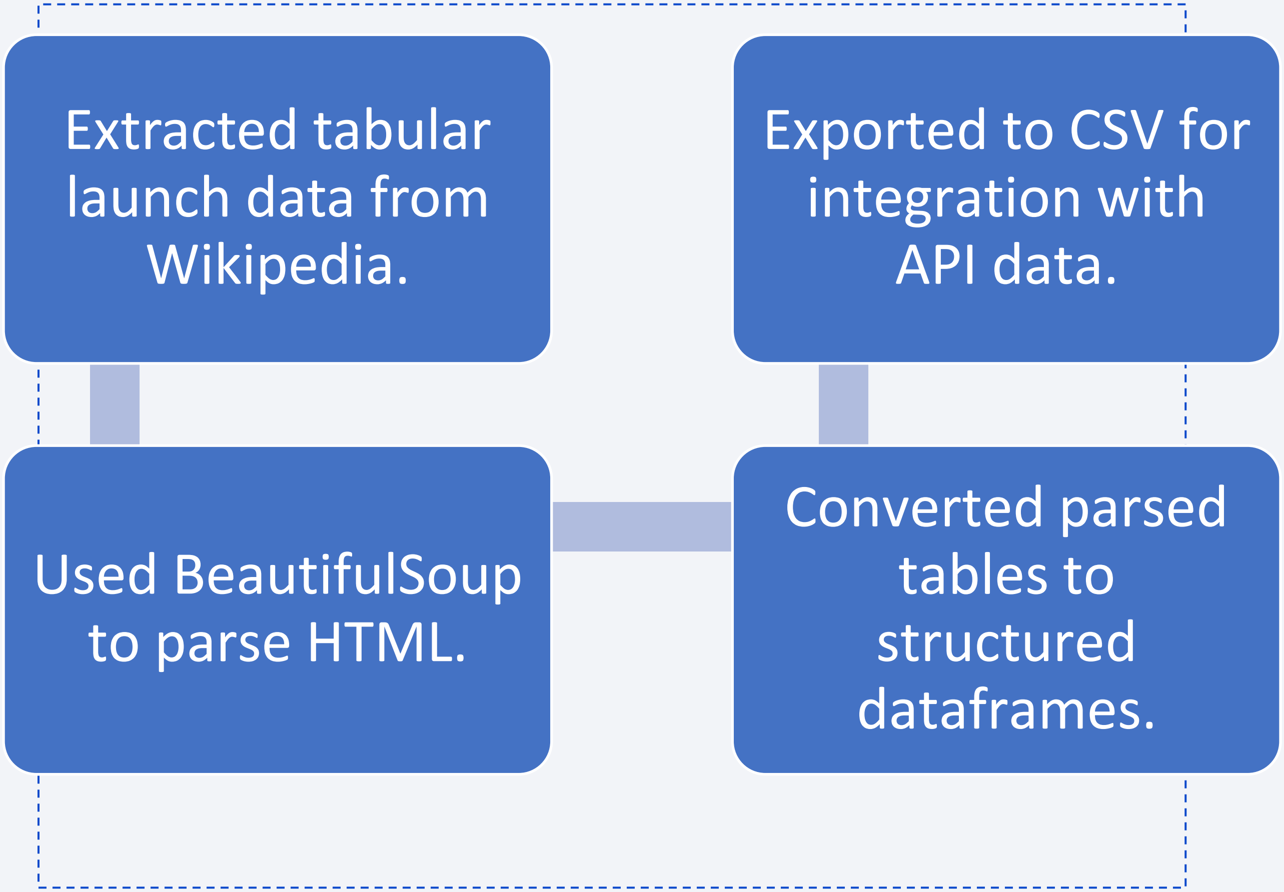
- Retrieve Falcon 9 launch information from Wikipedia
- Generate a BeautifulSoup object using the HTML response
- Identify column headers from the HTML table structure
- Extract launch data by parsing the HTML tables
- Organize the extracted data into a dictionary format
- Convert the dictionary into a structured DataFrame
- Save the finalized data into a CSV file



GitHub-Link

Data Collection - Scraping

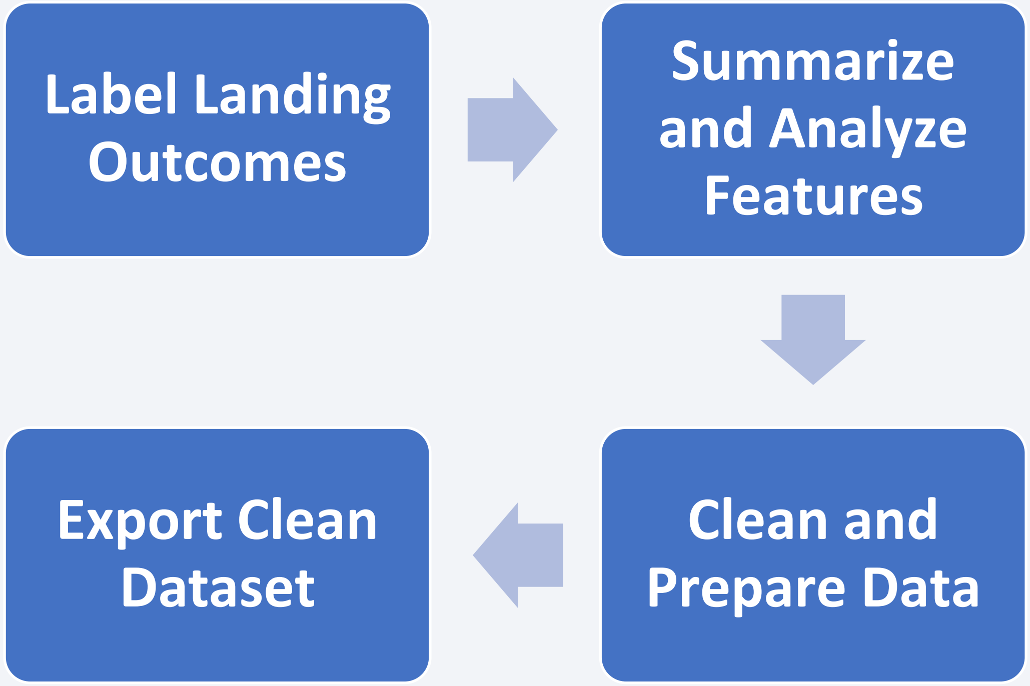
- **Send Request**
Fetch Falcon 9 launch data from Wikipedia.
- **Parse HTML**
Create a BeautifulSoup object from the HTML response.
- **Extract Table Headers**
Identify column names from the HTML table.
- **Collect Data**
Parse table rows and extract relevant launch data.
- **Build Structured Data**
Organize the data into a dictionary, then convert it into a DataFrame.
- **Export to CSV**
Save the structured data to a CSV file for further analysis.



GitHub-Link

Data Wrangling

- Created a binary “success” label for landing outcomes.
- Categorized landings: True/False RTLS, ASDS, Ocean.
- Labeled all outcomes as 1 (success) or 0 (failure).
- Summarized launches per site and orbit.
- Exported processed data for further analysis.



GitHub-Link

EDA with Data Visualization

- Used scatter and bar plots to examine:
- Flight number vs. Payload mass and Launch Site
- Payload vs. Orbit
- Orbit type vs. success
- Success trends over time
- Charts helped reveal potential predictors for model training.

- GitHub Link

EDA with SQL

- Queried the database for:
- Unique launch sites
- NASA payload totals
- Dates of key milestones (e.g., first ground landing)
- Payload stats by booster version
- Distribution of successes/failures
- High/low performing orbits and boosters

- GitHub Link

Build an Interactive Map with Folium

- Mapped launch sites using their coordinates.
- Color-coded markers indicated successful vs. failed launches.
- Visualized proximity to coastlines, highways, cities, and railways.
- Noted that most sites were near the equator and ocean for performance and safety.

- Site Markers with Labels:
  Placed at each launch site to show names and locations.
- Color-Coded Outcomes:
  Green = success, Red = failure; helps visualize site performance.
- Distance Lines:
  From launch site to nearest city, highway, railway, and coast; shows safety and accessibility.

GitHub-Link

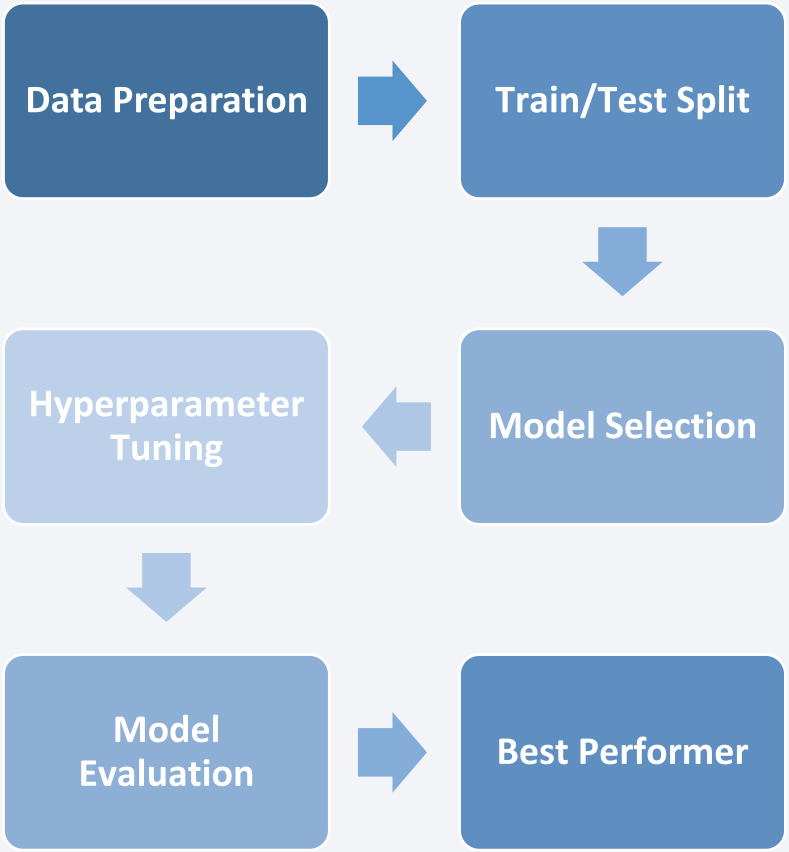
Build a Dashboard with Plotly Dash

- **Launch Site Dropdown:**
  Allows users to filter the data by selecting specific launch sites, enabling targeted analysis.
- **Pie Charts & Scatter Plots:**
  Pie charts show the ratio of successful vs. failed launches.
  Scatter plots display the relationship between payload mass and launch success, grouped by booster version.
- **Payload Range Slider:**
  Interactive slider lets users select a payload mass range to analyze how different weights impact success rates across sites and boosters.

- GitHub Link

Predictive Analysis (Classification)

- **Data Preparation**
- **Train/Test Split:**
  Divided the data into training and testing sets using train_test_split.
- **Model Selection:**
  Logistic Regression, SVM, Decision Tree, and K-Nearest Neighbors (KNN).
- **Hyperparameter Tuning**
- **Model Evaluation**
- **Best Performer:**
  Decision Tree model showed the highest overall performance and accuracy across the full dataset.



Results

1. Exploratory Data Analysis (EDA):

- Success over Time:
  Launch success improved significantly from 2013 onward.
- Site Performance:
  KSC LC-39A had the highest success rate among all sites.
- Orbit Insights:
  Orbits like ES-L1, GEO, HEO, and SSO achieved 100% success.
- Payload Trends:
  Higher payloads (2,000–7,000 kg) generally led to more successful landings.

2. Interactive Analytics:

- Folium Map:
  Visualized all launch sites with outcome markers (green for success, red for failure).
  Displayed proximity to key infrastructure (cities, highways, coasts).
- Plotly Dashboard:
  Enabled interactive filtering by launch site and payload range.
  Pie and scatter charts revealed success patterns and payload correlations..

3. Predictive Analysis:

Model Comparison:
Logistic Regression, SVM, Decision Tree, and KNN were trained and evaluated.

Best Model:
Decision Tree outperformed others with highest accuracy and F1 score.

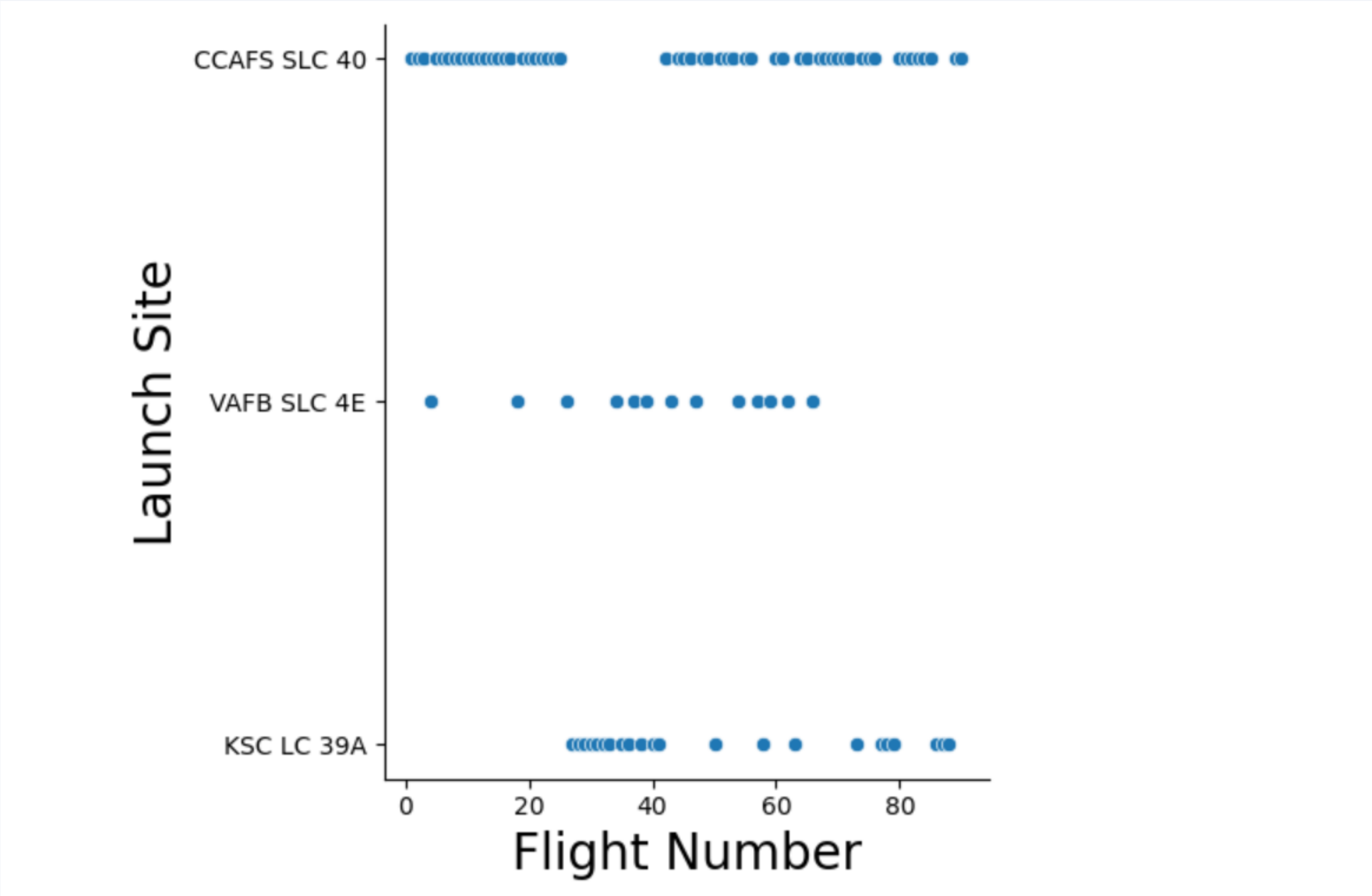
Performance Metrics:
Evaluation based on confusion matrix, Jaccard score, and cross-validation accuracy

Section 2

Insights drawn from EDA

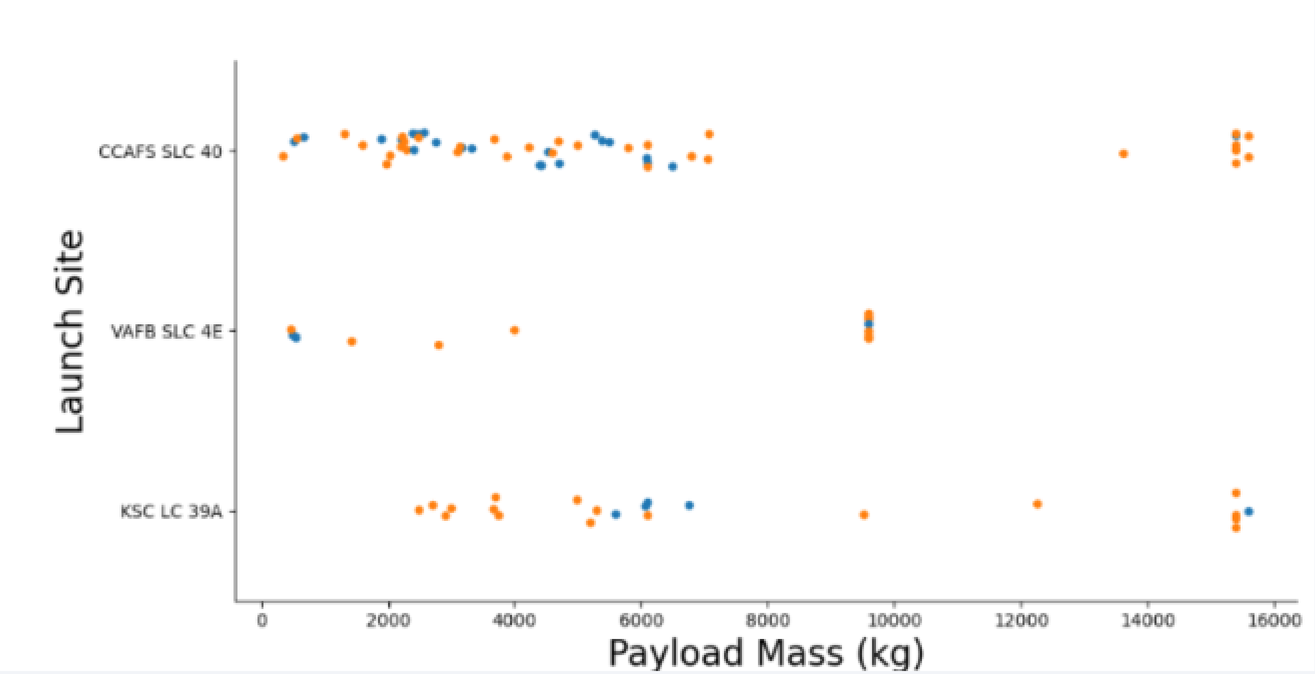
Flight Number vs. Launch Site

- With increase in flight number, the success rate increases as well in launch sites



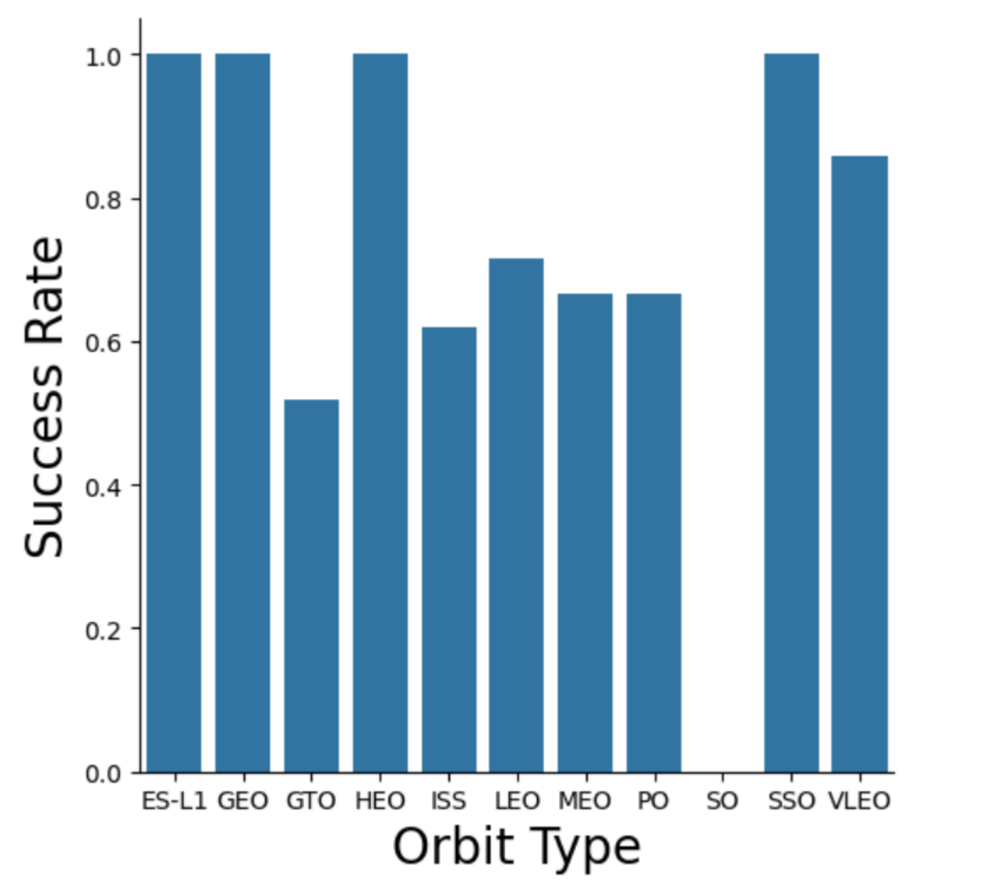
Payload vs. Launch Site

- Usually, higher the payload mass(kg), higher is the success rates.
- Most launches with a payload greater than 7000 kg were successful.

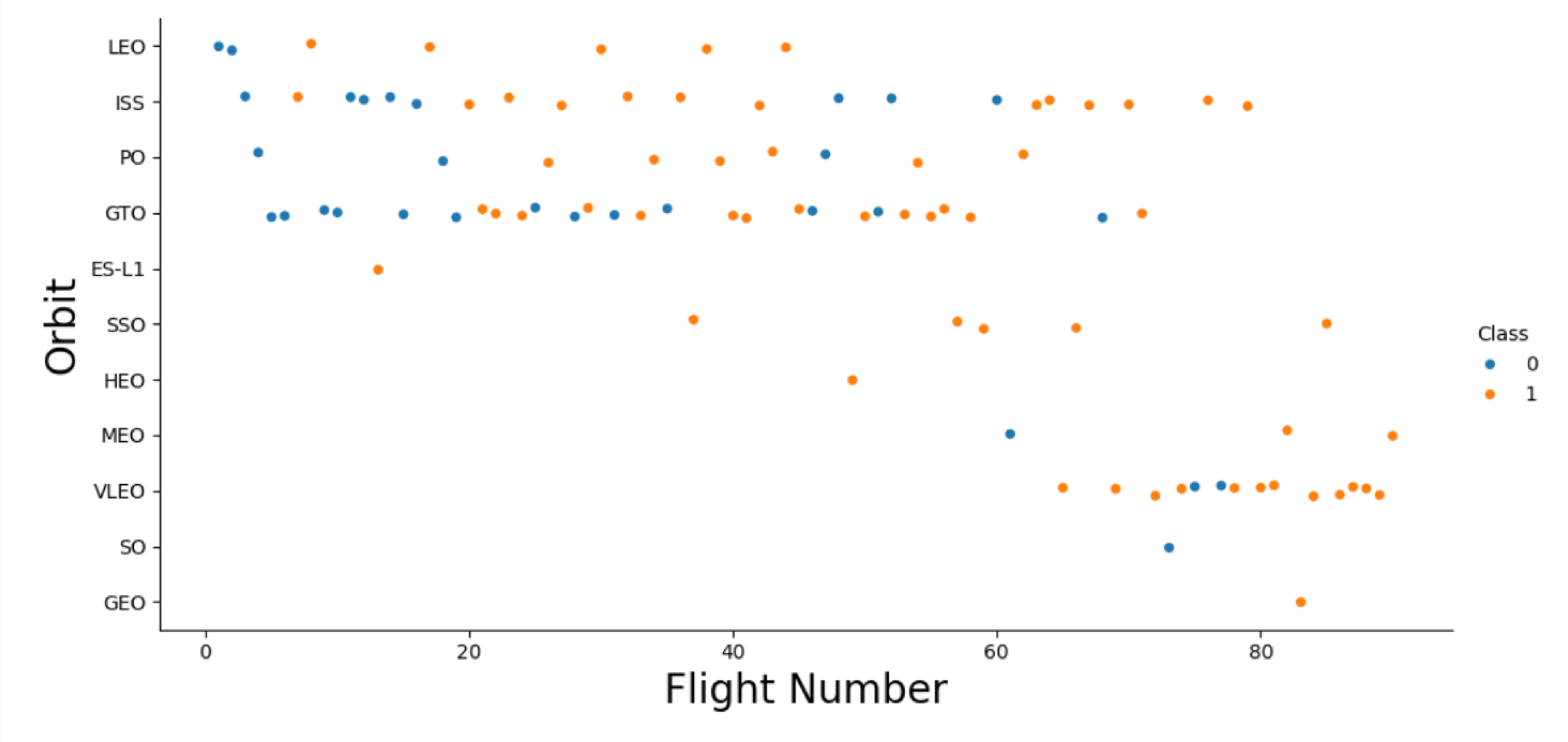


Success Rate vs. Orbit Type

- **100% Success:**
  ES-L1, GEO, HEO, and SSO orbits had perfect landing success rates.
- **Lowest Success:**
  SO orbit had **0%** success rate.
- **Moderate Success (60–75%):**
  ISS, LEO, MEO, PO orbits showed average performance.
- **Below Average:**
  GTO orbit underperformed compared to others (~50%).
- **Strong Performer:**
  VLEO orbit had a relatively high success rate (~88%).



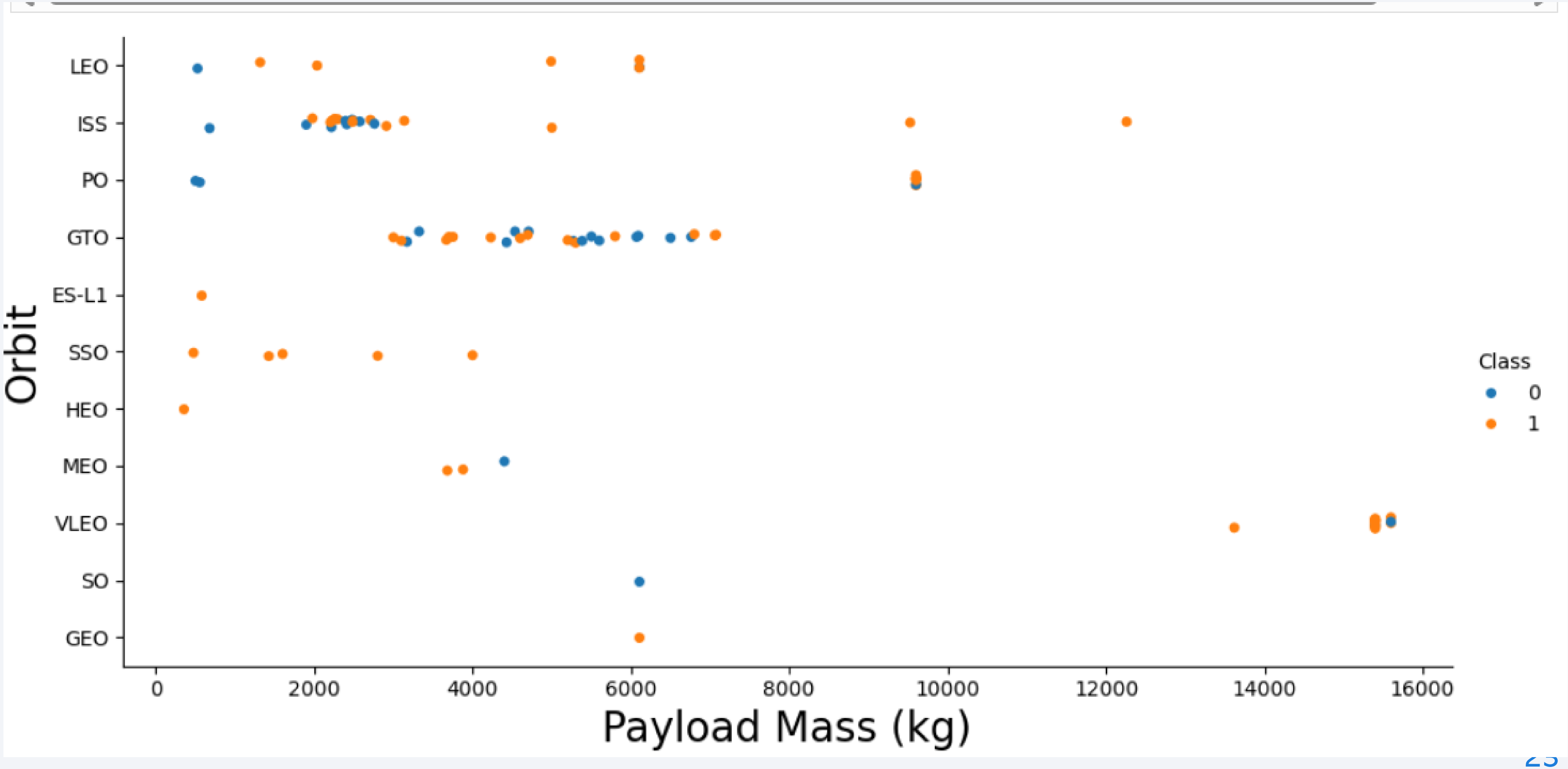
Flight Number vs. Orbit Type



Flight Number vs. Orbit Type

- **LEO & ISS Orbits:**
  Show a clear trend — more recent flights (higher flight numbers) have more successes.
- **Success Improves Over Time:**
  Later flights across most orbits trend towards higher success rates.
- **GTO Orbit:**
  Mixed results across all flight numbers; no strong upward trend in success.
- **MEO, VLEO, SSO:**
  These orbits show mostly successful outcomes in later flights.
- **SO Orbit:**
  All missions in this orbit failed (only blue dots).

Payload vs. Orbit Type

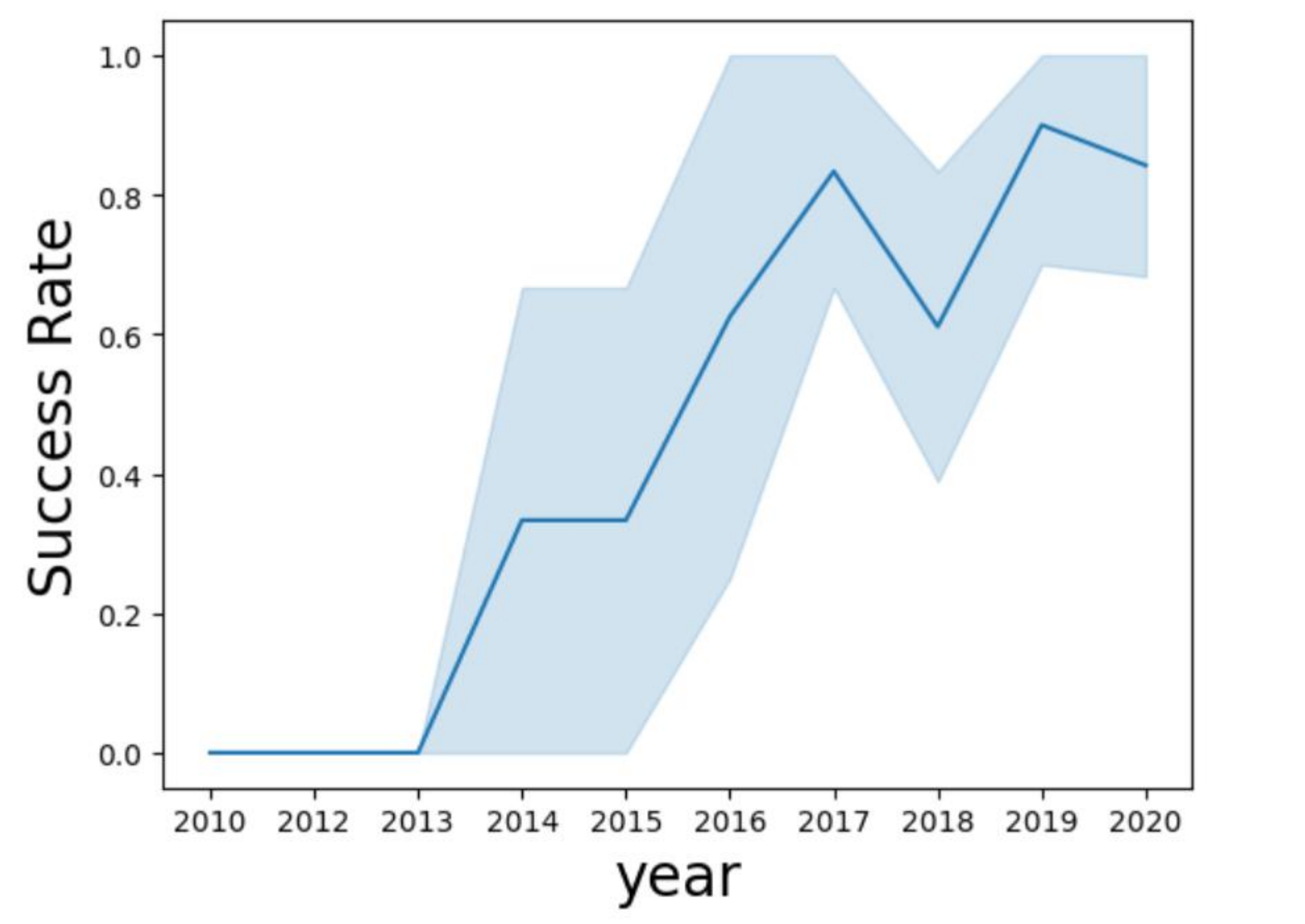


Payload vs. Orbit Type

- **LEO, ISS, PO Orbits:**
  Heavier payloads generally show higher success (more orange points).
- **GTO Orbit:**
  Mixed results even at similar payload ranges; no clear success pattern.
- **SSO & HEO Orbits:**
  Show mostly successful launches across varied payloads.
- **SO Orbit:**
  Failed outcomes dominate, regardless of payload mass.
- **High Payloads (~10,000–15,000 kg):**
  Mostly successful launches across multiple orbits.

Launch Success Yearly Trend

- The success rate improved from 2013 to 2017 and 2018 to 2019.
- The success rate decreased from 2017 to 2018 and from 2019 to 2020.
- Overall, the success rate has improved since 2013.



All Launch Site Names

Task 1

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

```
[10]: %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;

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

[10]:

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

Launch Site Names Begin with 'CCA'

Task 2

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

[11]:
```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
```

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

[11]:

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

Total Payload Mass

Task 3

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

[13]:

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

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

[13]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Task 4

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

[14]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

[14]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[14]: %sql SELECT MIN(DATE) FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';
```

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

[14]:

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

[16]:
```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

[16]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

[17]:
```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;
```

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

[17]:

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

Boosters Carried Maximum Payload

```
[18]: %sql SELECT BOOSTER_VERSION \
      FROM SPACEXTBL \
      WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

[18]:

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

2015 Launch Records

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT substr(Date,4,2) as month, DATE,Booster_Version, Launch_Site \
FROM SPACEXTBL \
where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

MONTH	booster_version	launch_site
1	F9 v1.1 B1012	CCAFS LC-40
4	F9 v1.1 B1015	CCAFS LC-40

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

Task 10

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

[24]:
```
%sql SELECT Landing_Outcome, count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by count_outcomes DESC;
```

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

[24]:

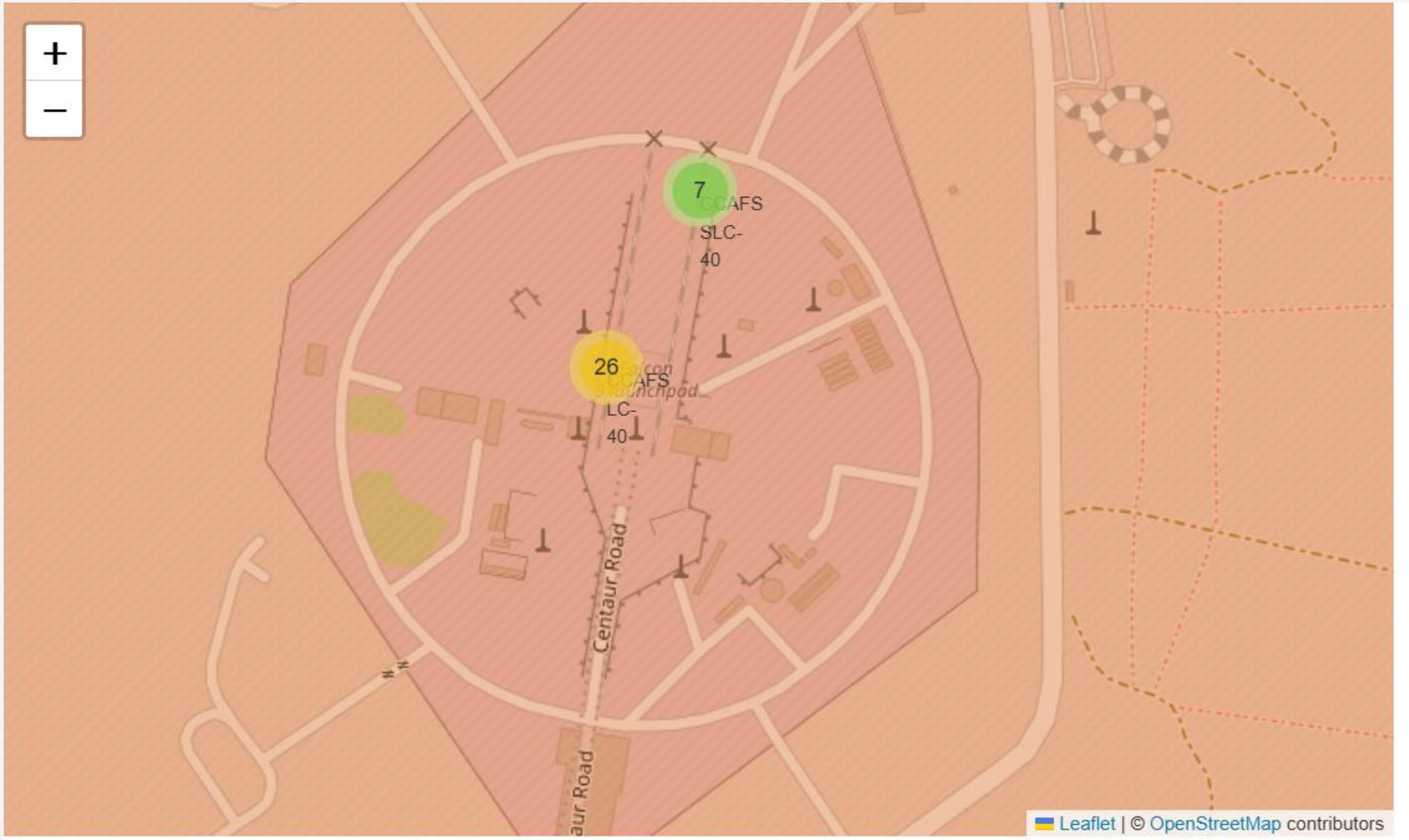
Landing_Outcome	count_outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

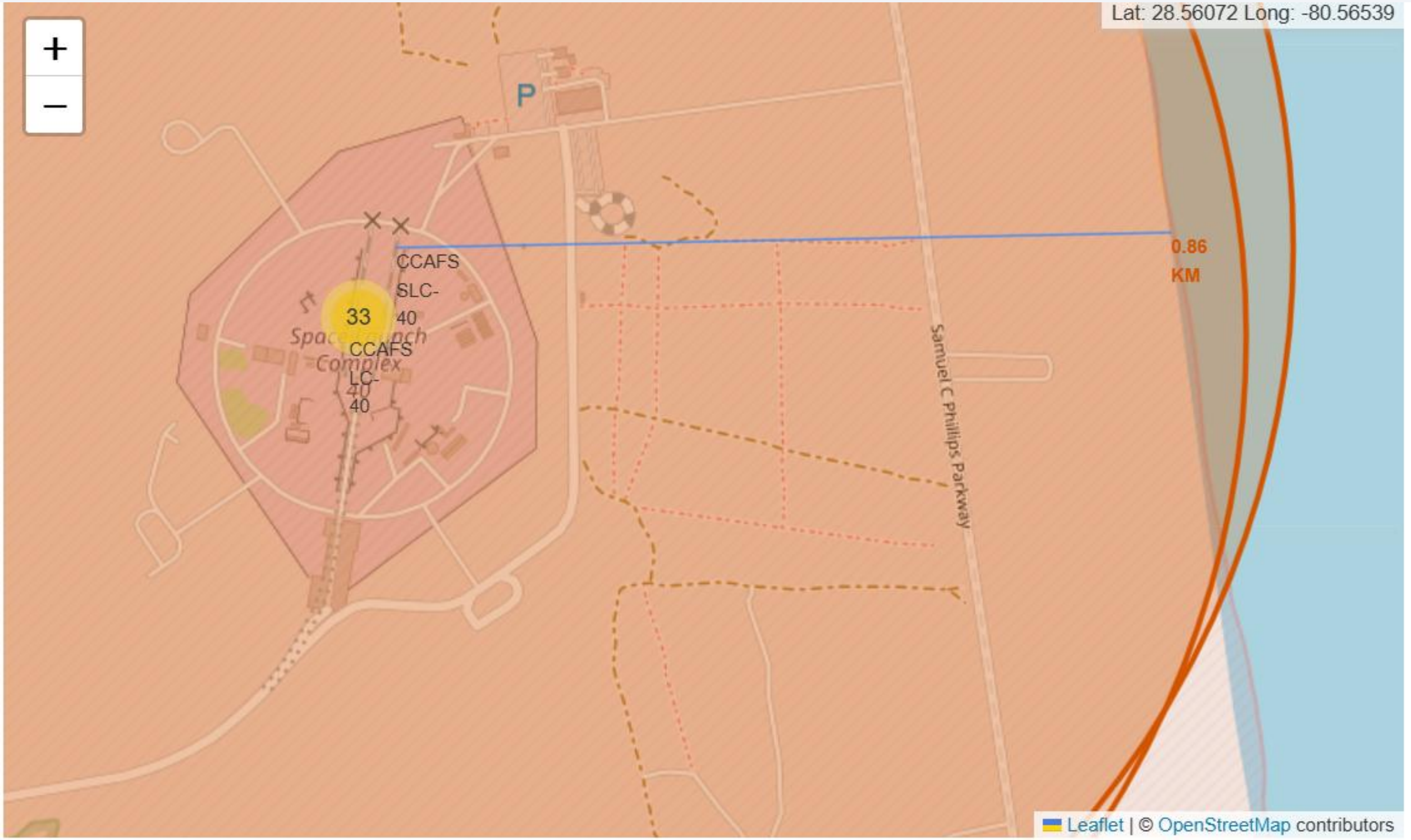
Launch Sites Proximities Analysis

All Launch Sites Markers

Launch Outcomes



Proximity

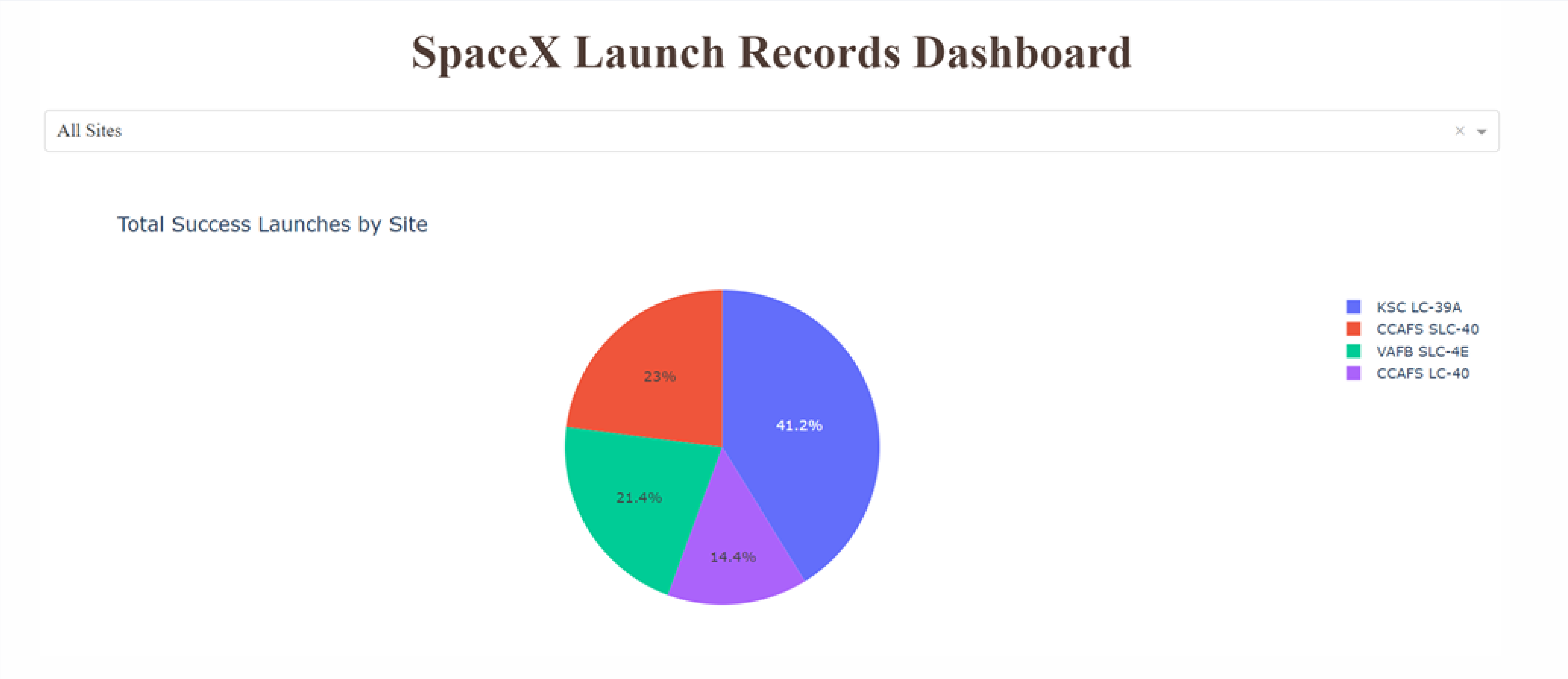


Lat: 28.56072 Long: -80.56539
+
−
P
CCAFS
SLC-
40
33
CCAFS
LC-
40
0.86
KM
Samuel C Phillips Parkway
Leaflet | © OpenStreetMap contributors

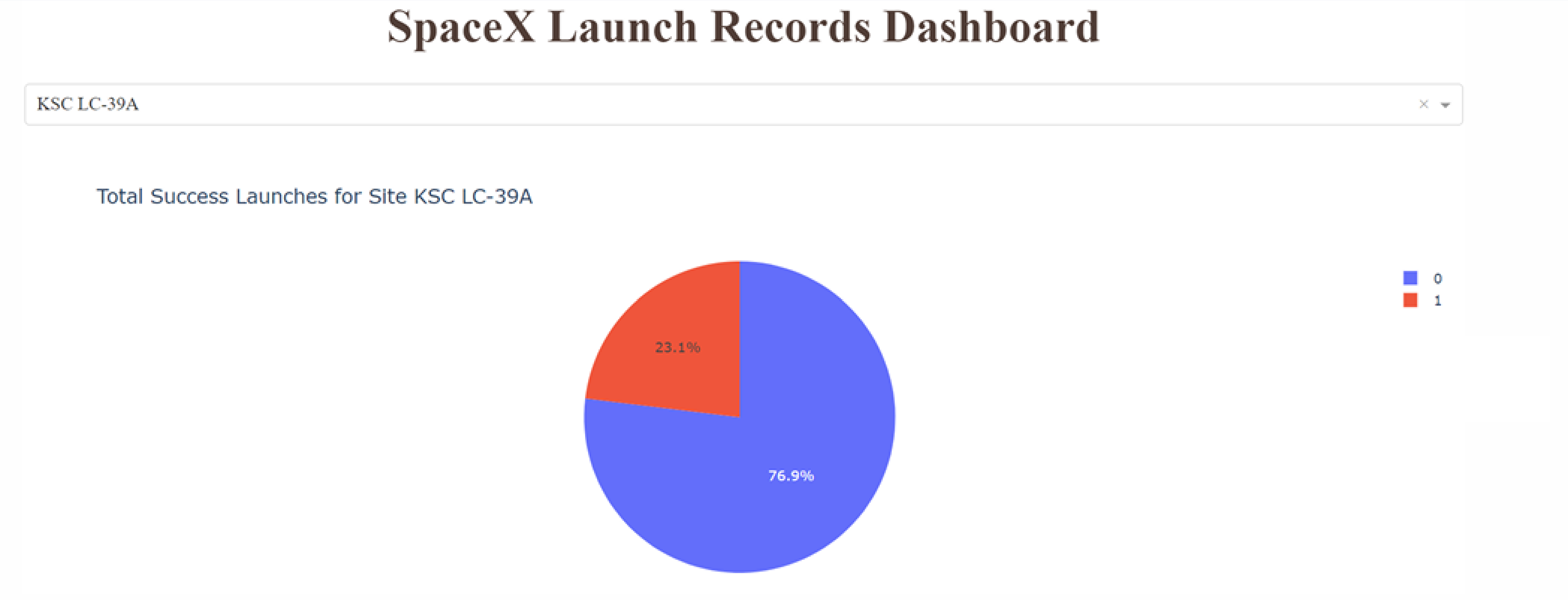
Section 4

Build a Dashboard with Plotly Dash

Launch Success by Sites



Highest Launch Success Ratio



Payload vs Launch Outcome

Section 5

Predictive Analysis (Classification)

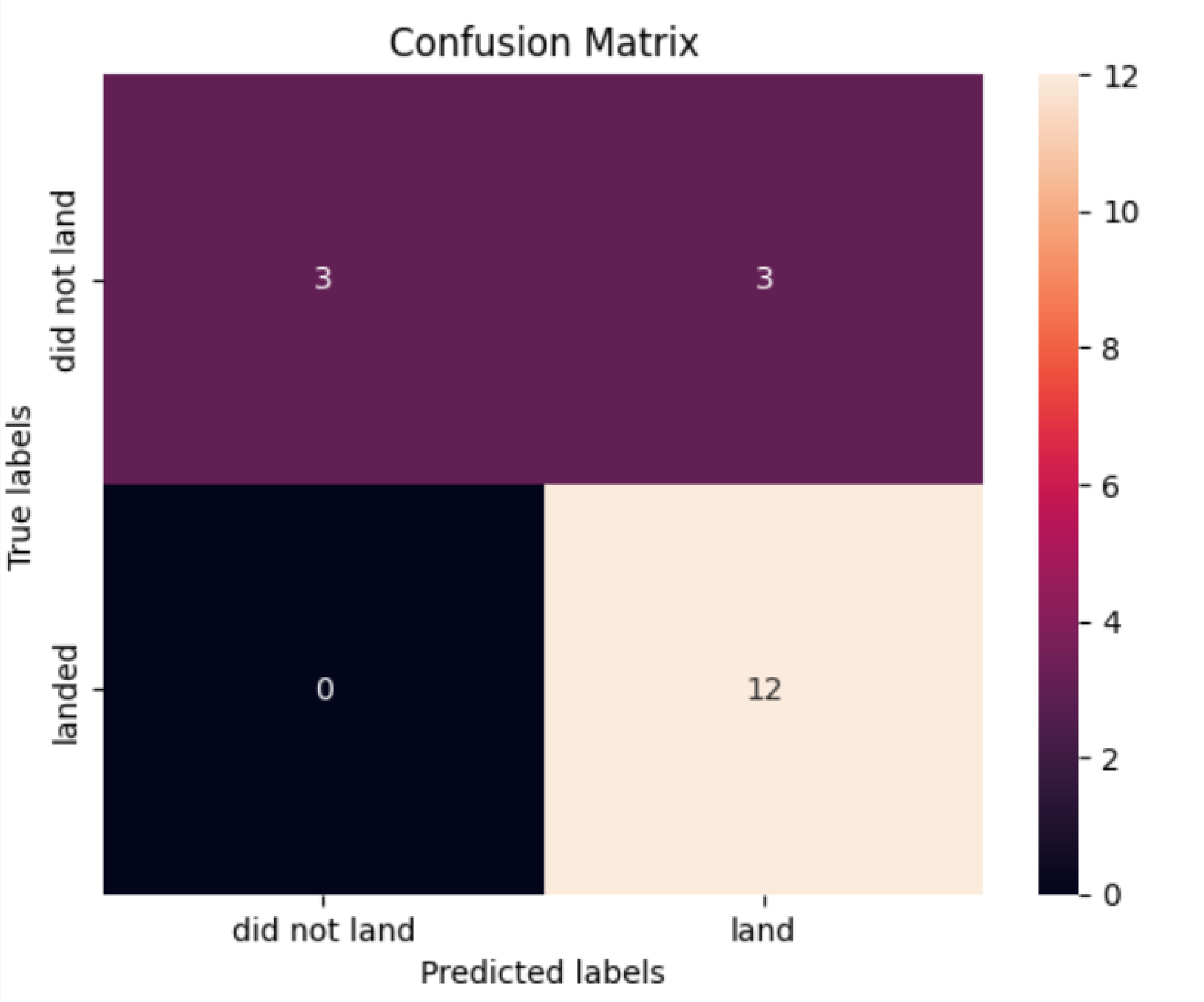
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- **True Positives (12):** Correctly predicted successful landings
- **True Negatives (3):** Correctly predicted failed landings
- **False Positives (3):** Predicted success, but actually failed
- **False Negatives (0):** None — all actual successes were correctly predicted



Conclusions

- **Best Model:**
  The **Decision Tree** algorithm outperformed others with the highest accuracy and F1 score.
- **Success Trends:**
  Launch success rates **improved over time**, especially in recent missions.
- **Orbit Performance:**
  Orbits like **GEO, HEO, ES-L1, and SSO** showed **100% landing success**.
- **Payload Insights:**
  Launches with **2,000–7,000 kg** payloads had **higher success rates** across most orbits.
- **Launch Sites:**
  KSC LC-39A was the most reliable launch site, with the **highest success rate**.
- **Geographical Advantage:**
  Most launch sites were near the **equator** and **coast**, aiding launch performance and safety.

Thank you!