IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

VISHAL GAURAV
AUGUST 7, 2024

Outline

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

Executive Summary

Methodologies:

- Data Collection: Collected data from SpaceX databases and online sources.
- Descriptive Analysis: Conducted a thorough descriptive analysis to gain insights into the dataset.
- Predictive Modeling: Developed and assessed multiple machine learning models to perform predictive analysis.

Key Results:

- The CCAFS LC-40 launch site exhibited the highest success rate, particularly for payloads between 2000-5000 kg and SSO orbit types.
- The predictive model achieved 83.33% accuracy in forecasting launch success based on test data.

Introduction

- Overview:

This project aims to predict the successful landing of the SpaceX Falcon 9 first stage.

- Objective:

The goal is to collect, process, and analyze data, and then build and compare various machine learning models to identify key factors that influence landing success rates.

Section 1

Methodology

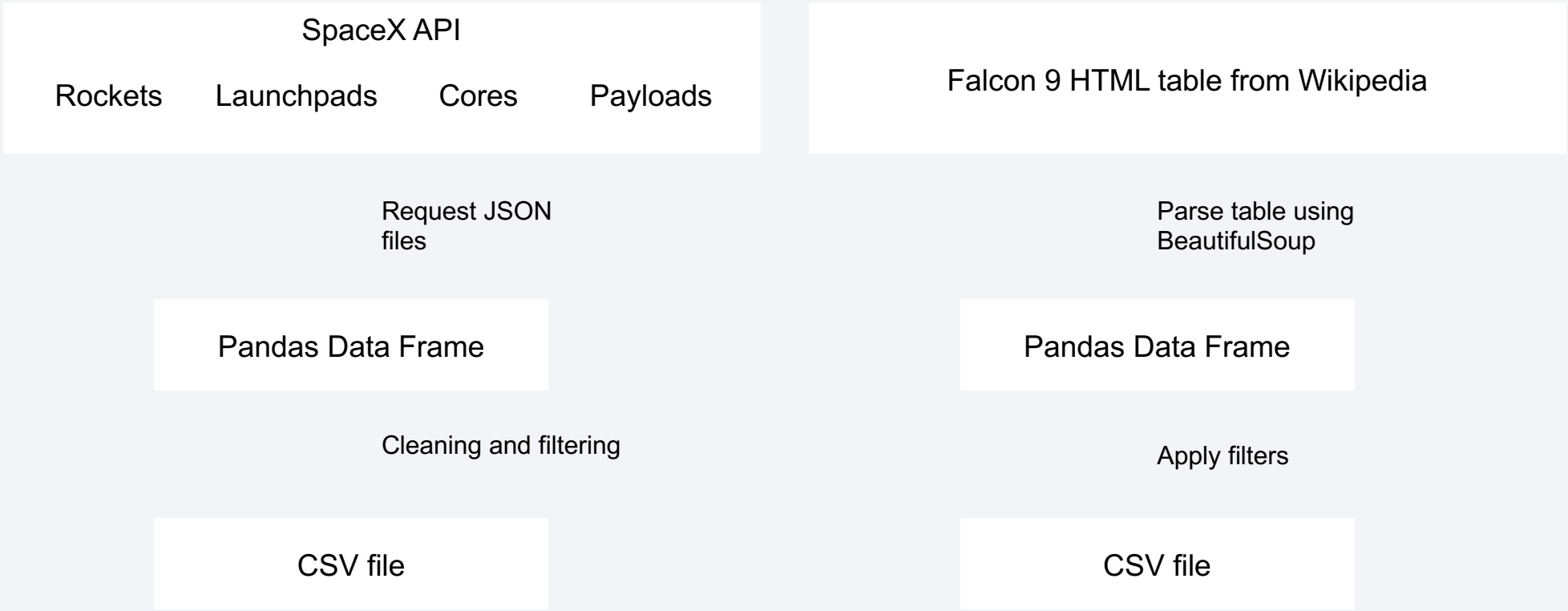
Methodology

Executive Summary

- Data collection methodology:
  - From SpaceX API and web scraping
- Perform data wrangling
  - Handling Missing Values
  - Class Labeling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium & Plotly Dash
- Perform predictive analysis using classification models
  - SVM
  - Classification Trees
  - Logistic Regression

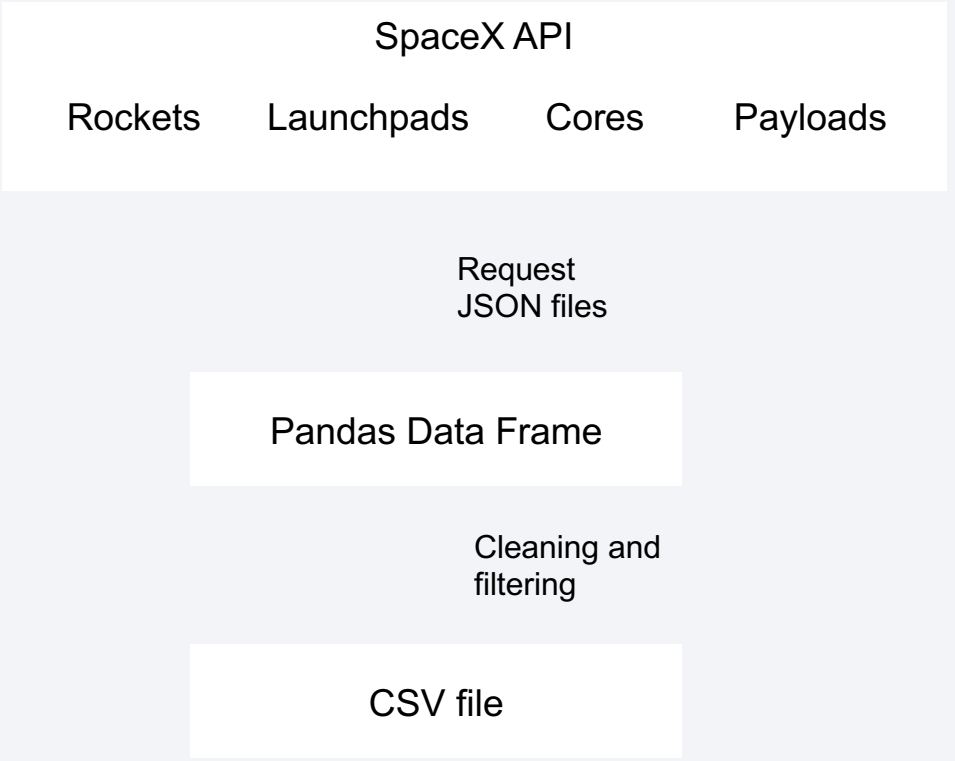
Data Collection

SpaceX API and web scraping:



Data Collection – SpaceX API

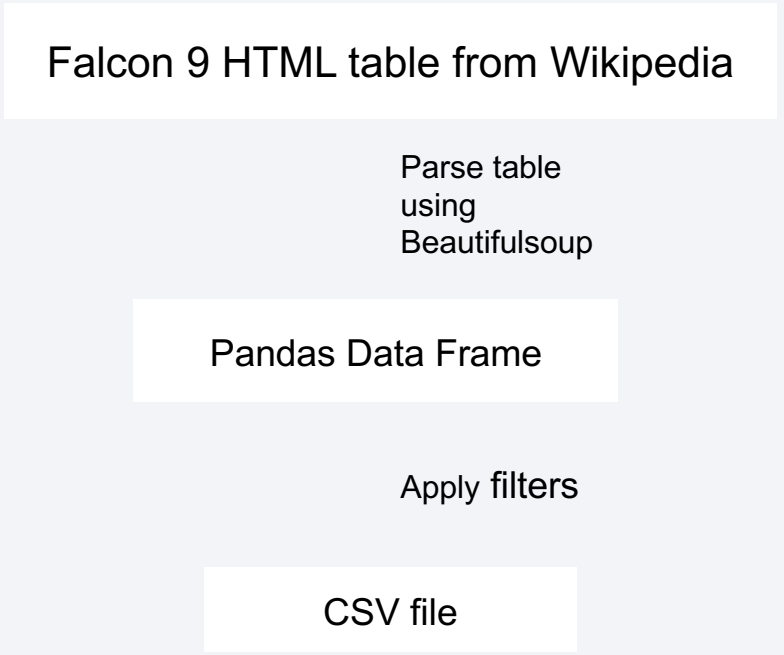
- Use the requests library to extract and concatenate past launch data.
- Replace any missing payload mass values with the column average.
- Filter the data by the relevant features for analysis.



Link-
https://github.com/VishalGaurav109/VishalGaurav109/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Store the HTML table response in a text file.
- Parse the headers, rows, and columns.
- Create a Pandas DataFrame from the parsed data.
- Export the DataFrame to a CSV file.



Link-
https://github.com/VishalGaurav109/VishalGaurav109/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Import the CSV files created during data collection.
- Calculate the number of launches for each site.
- Find the number of occurrences for each orbit.
- Create a landing outcome label from the Outcome column.

Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Create a scatter plot of FlightNumber vs. PayloadMass, colored by Launch Site, to understand how these variables affect the launch outcome.
- Use bar plots to understand the relationship between the success rate of each orbit type.
- Visualize the yearly trend of launch success using a line chart.

Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/edadataviz.ipynb

EDA with SQL

- Display the average payload mass carried by the booster version F9 v1.1.
- List the names of the boosters that have succeeded in landing on a drone ship and have a payload mass greater than 4000 but less than 6000.
- List the names of the booster versions that have carried the maximum payload mass.
- List the failed landing outcomes on drone ships, along with their booster versions and launch site names, for the year 2015.

Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark the successful and failed launches for each site on the map.
- Identify the closest point on the coastline to each launch site and calculate the distance between this coastline point and the launch site.
- Draw lines from each launch site to its closest city, railway, and highway.

Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Create a pie chart to show the total count of successful launches for all sites.
- Create a scatter chart to show the correlation between payload mass and launch success.
- Create a pie chart to display the counts of successful versus failed launches for each site.

Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/Interactive%20Dashboard%20with%20Ploty%20Dash

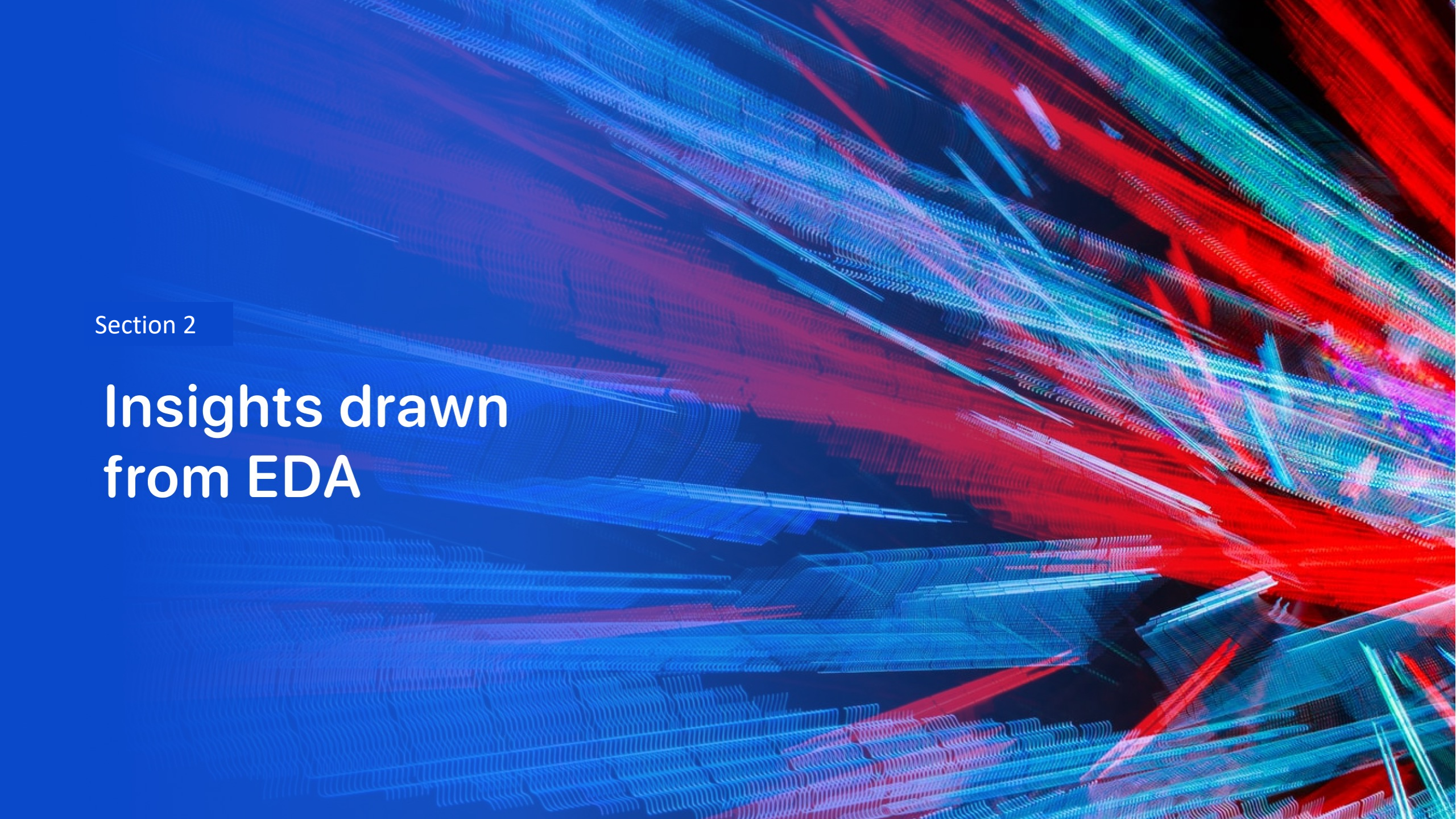
Predictive Analysis (Classification)

- Standardize the data.
- Split the data into training and test sets.
- Use cross-validation to find the best hyperparameters for SVM, Classification Trees, and Logistic Regression.
- Determine which method performs best using the test data.

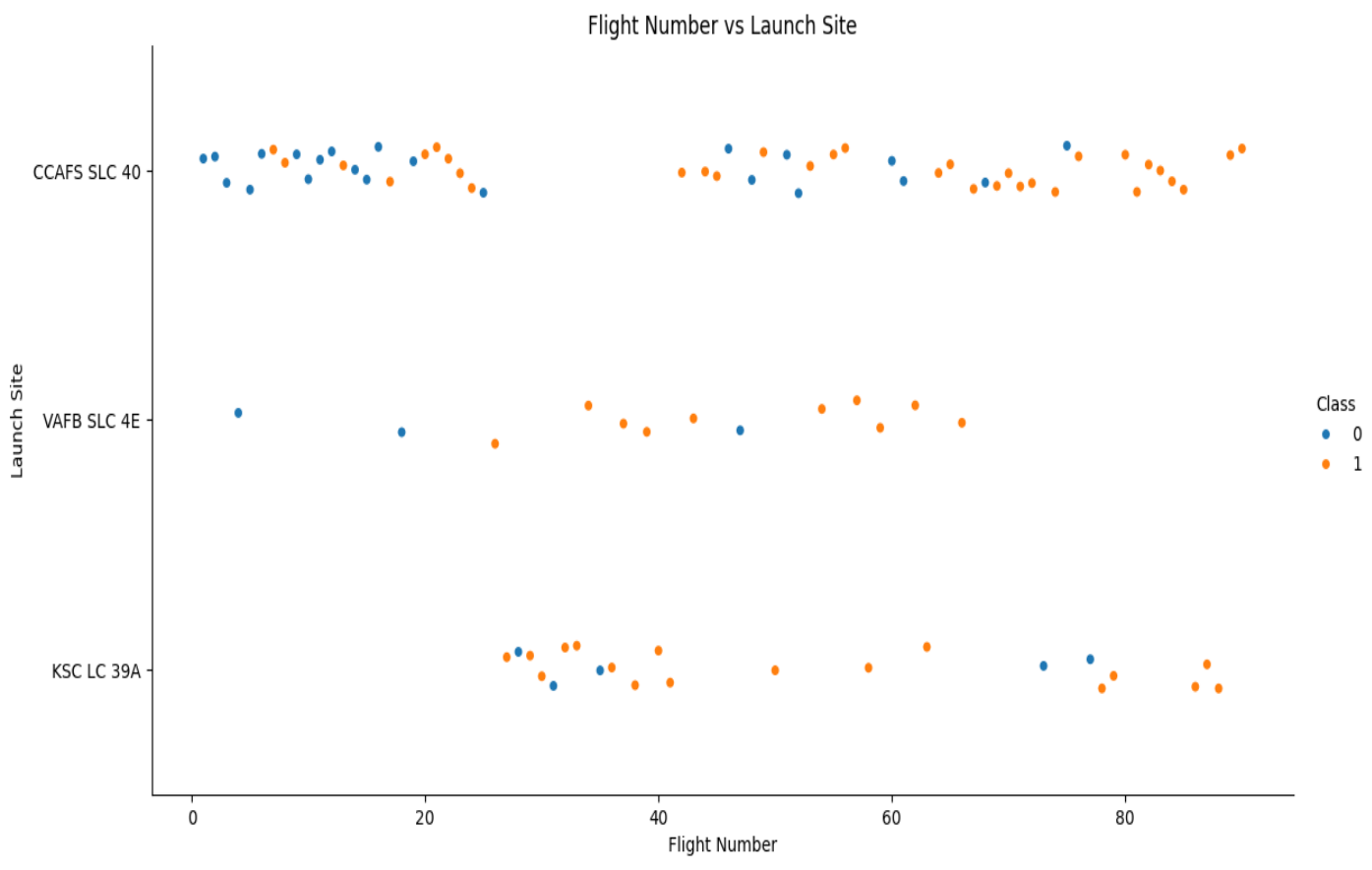
Link- https://github.com/VishalGaurav109/VishalGaurav109/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- The success rate improves with the flight number across all sites.
- Orbits such as ES-L1, GEO, HEO, and SSO have high success rates.
- The success rate has been increasing steadily from 2013 to 2020.
- Payload masses in the range of 2000-5000 kg have the highest launch success rates.
- The best performing method is: Logistic Regression with an accuracy of 0.8333

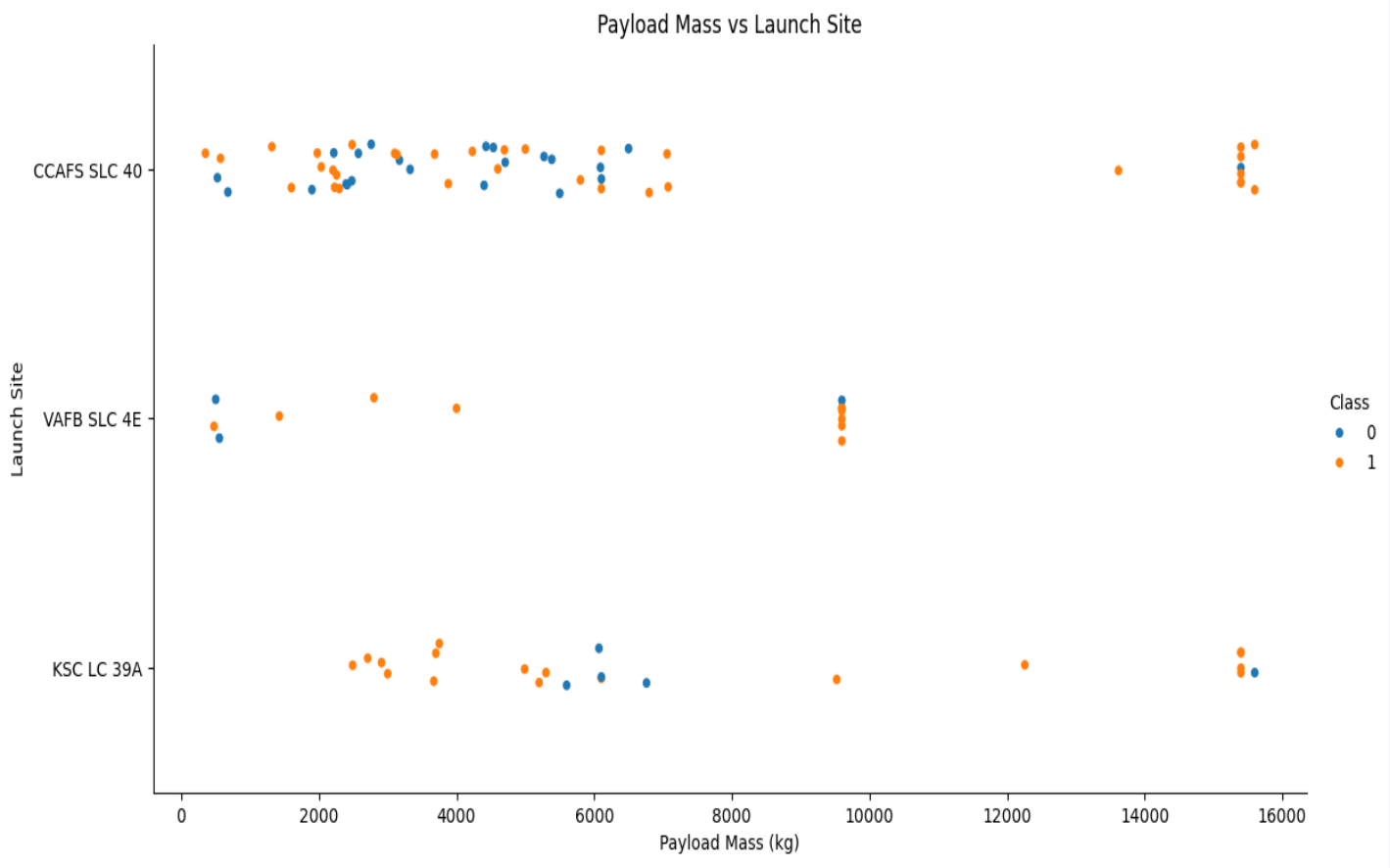

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



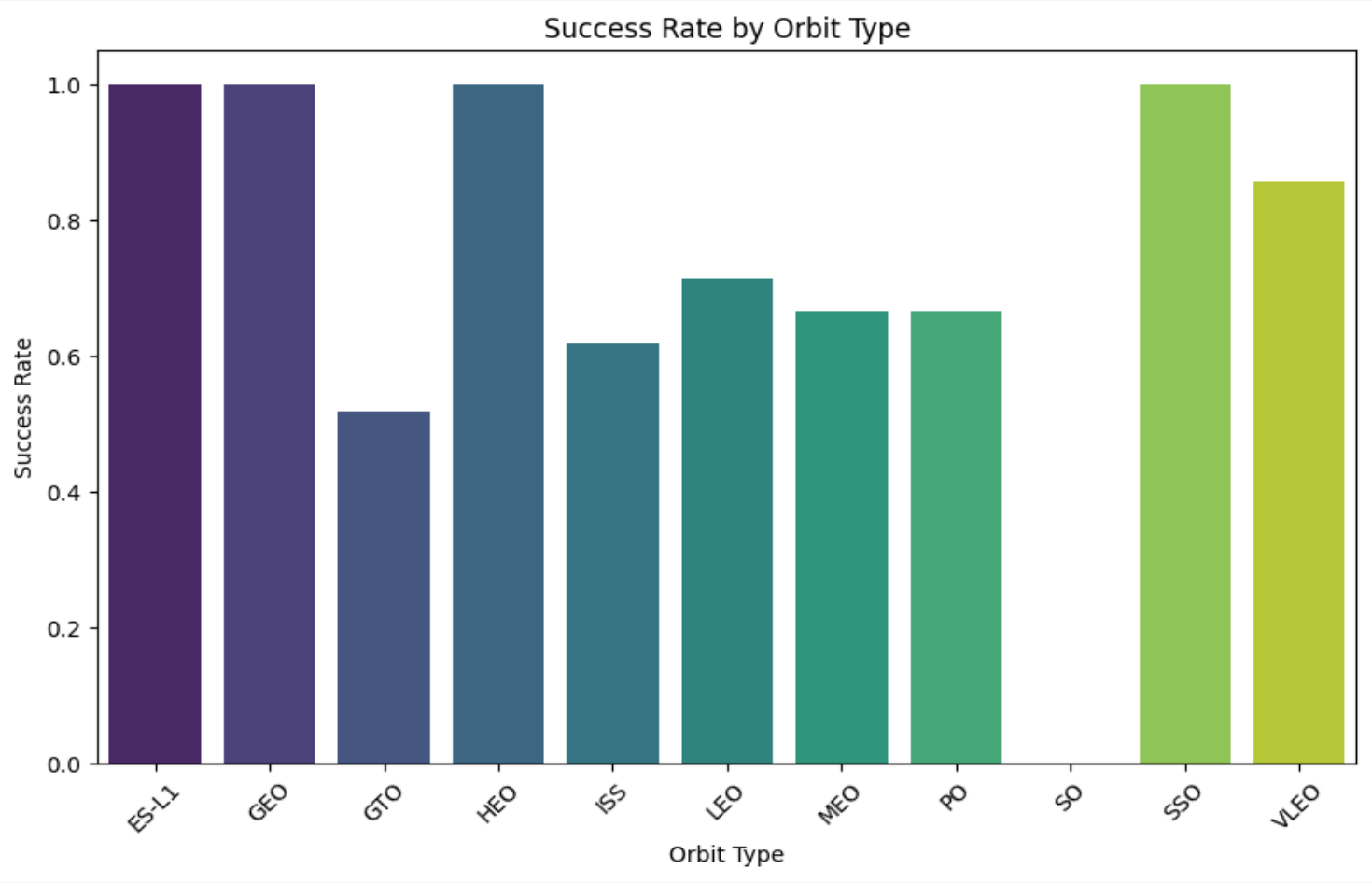
- The success rate improves with the flight number across all sites.
- CCAFS SLC 40 had a significantly higher number of launches compared to other sites.

Payload vs. Launch Site



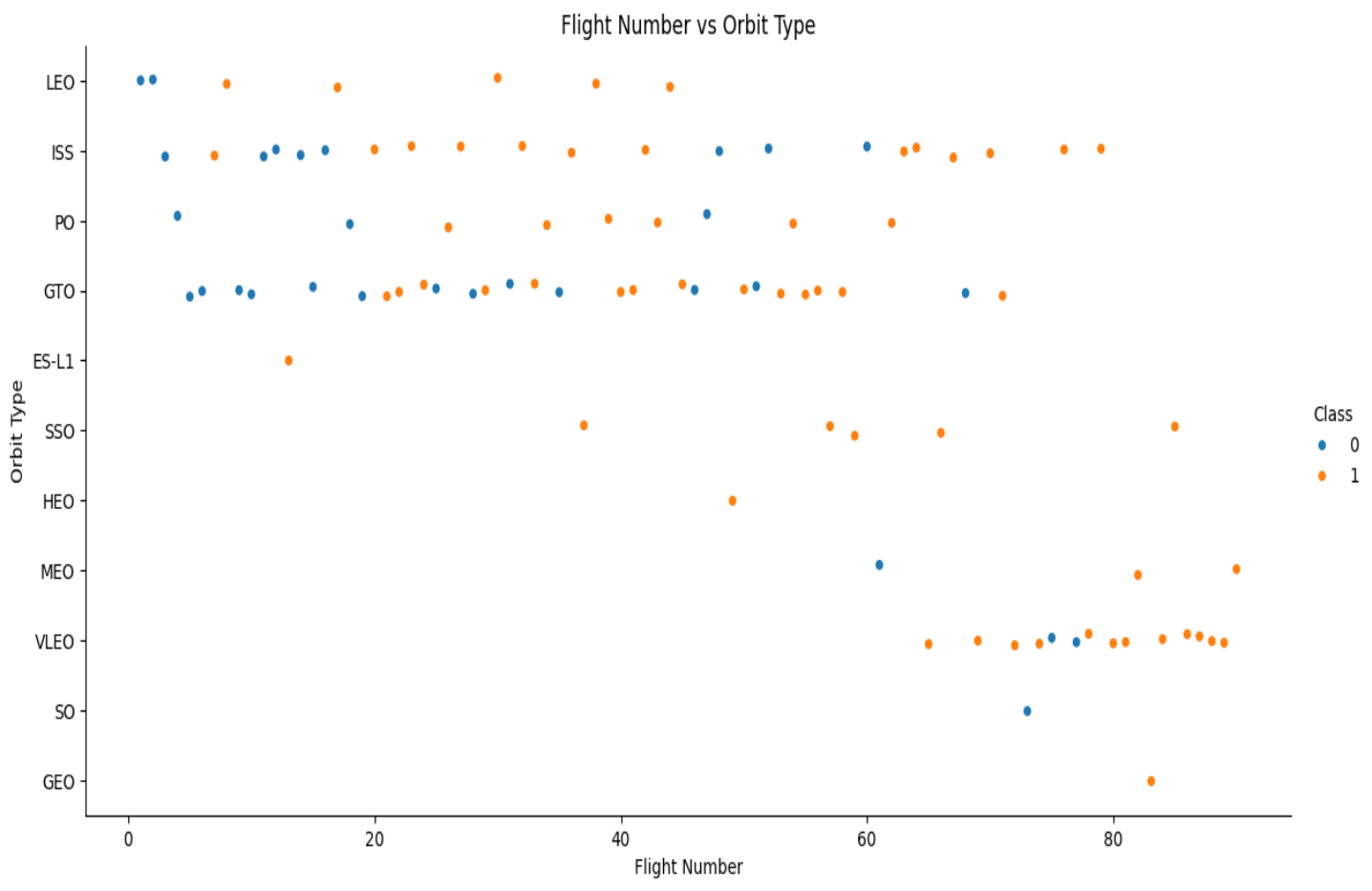
- There are more launches for lighter stages.
- At the VAFB-SLC launch site, no rockets were launched with a payload mass greater than 10,000 kg.

Success Rate vs. Orbit Type



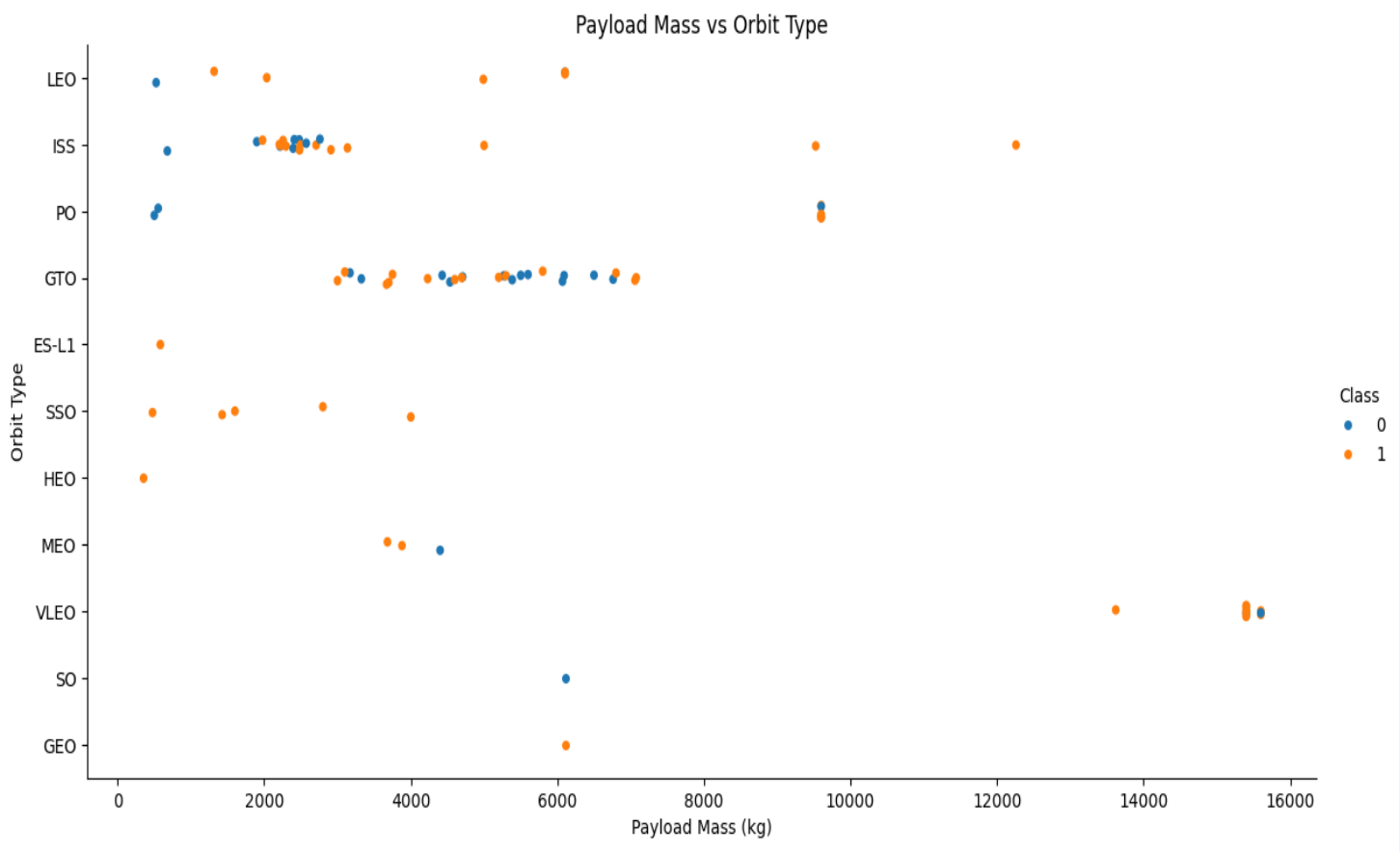
- Orbits ES-L1, GEO, HEO, and SSO have high success rates.
- The success rate for the SO orbit is zero.

Flight Number vs. Orbit Type



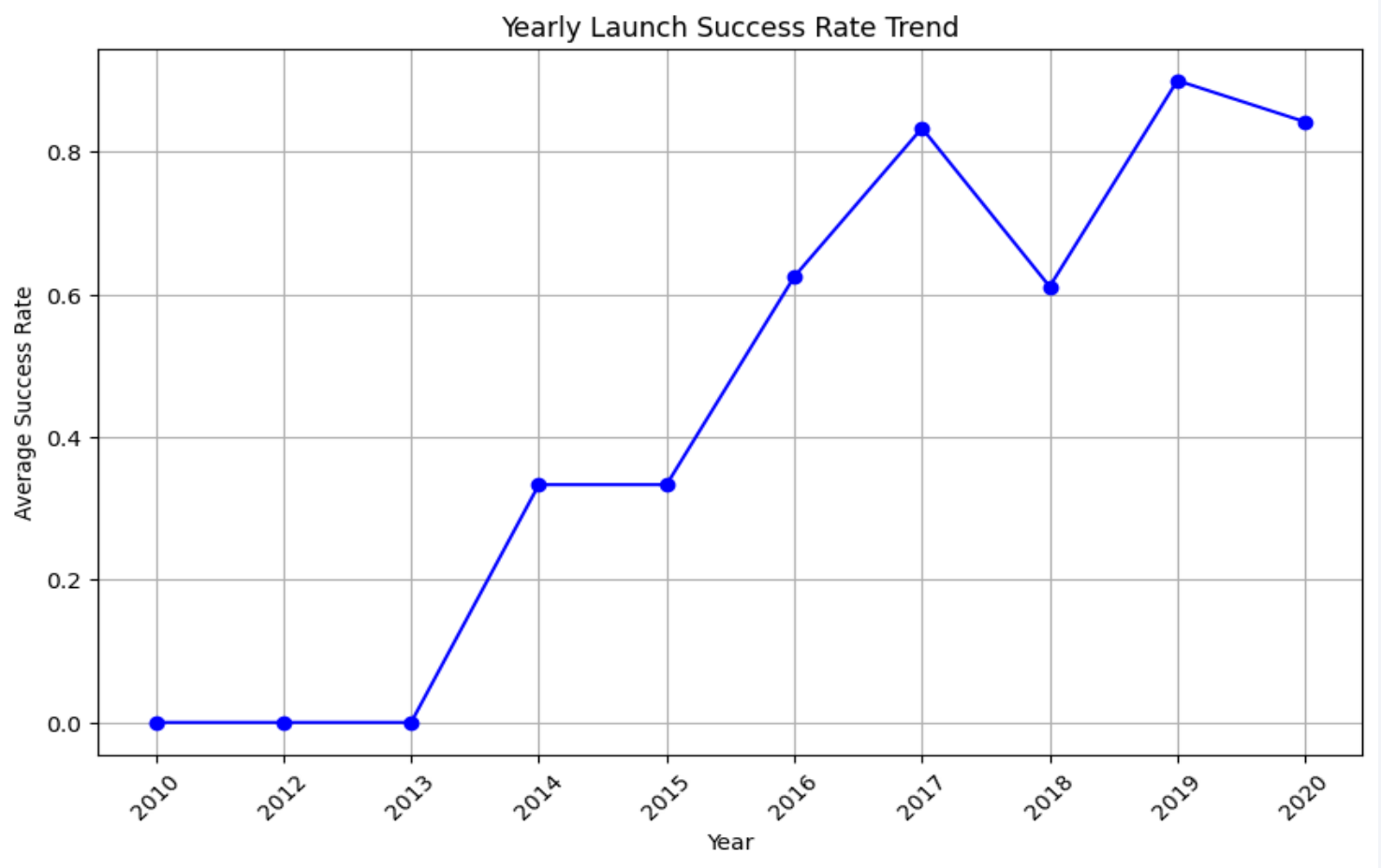
- Despite a success rate of 1, the GEO, ES-L1, and HEO orbits each have only one launch.
- The scatter plot indicates that the success rate appears to increase with flight number for LEO orbits, but there is no observable relationship for GTO orbits.
- SSO orbit appears to be the best choice based on success rates.

Payload vs. Orbit Type



- For heavy payloads, the successful landing rates are higher for Polar, LEO, and ISS orbits.
- For GTO orbits, it is challenging to distinguish the landing rates, as both successful and unsuccessful landing rates are observed.

Launch Success Yearly Trend



- The success rate has consistently increased from 2013 to 2020.

All Launch Site Names

The SQL query to find the distinct launch sites:

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

```
 * 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'

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

```
%sql SELECT * FROM SPACEXTBL 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_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

SQL query to calculate the total payload carried by boosters from NASA

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTBL WHERE
    "Mission_Outcome" LIKE '%NASA%' AND "Mission_Type" = 'CRS';
```

Average Payload Mass by F9 v1.1

SQL query to calculate the average payload mass carried by booster version F9 v1.1

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

```
]:  average_payload_mass
                2928.4
```

First Successful Ground Landing Date

SQL query to find the dates of the first successful landing outcome on ground pad:

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

first_successful_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' \
AND "PAYLOAD_MASS__KG_" > 4000 \
AND "PAYLOAD_MASS__KG_" < 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

SQL query to calculate the total number of successful and failure mission outcomes:

%sql SELECT "Mission_Outcome", COUNT(*) AS total FROM SPACEXTBL \ GROUP BY "Mission_Outcome";

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

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass :

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

```
 * 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

2015 Launch Records

List of the 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 SPACEXTBL \
     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

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 SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '20
     GROUP BY "Landing_Outcome" ORDER BY outcome_count DESC;
```

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

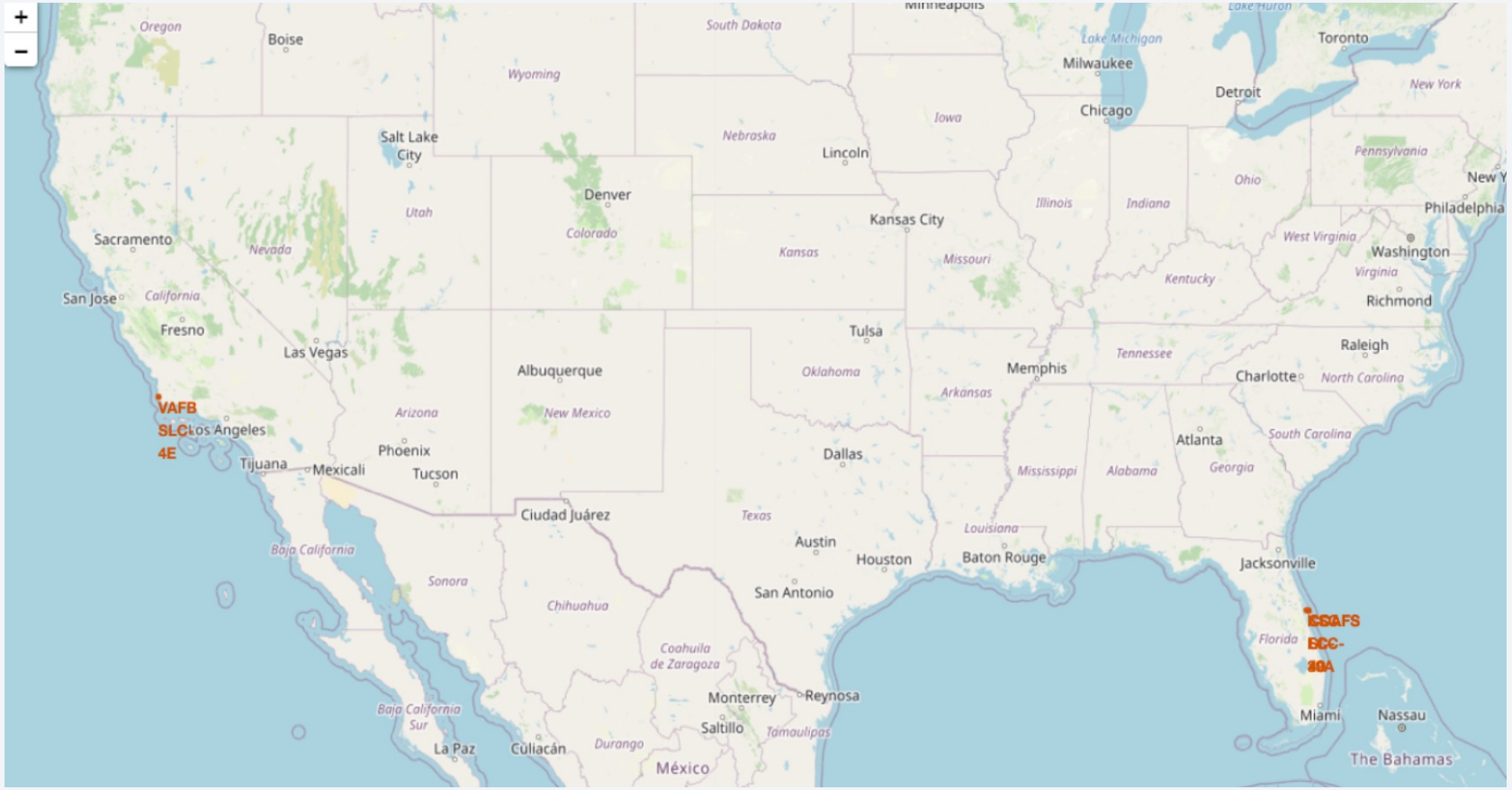
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

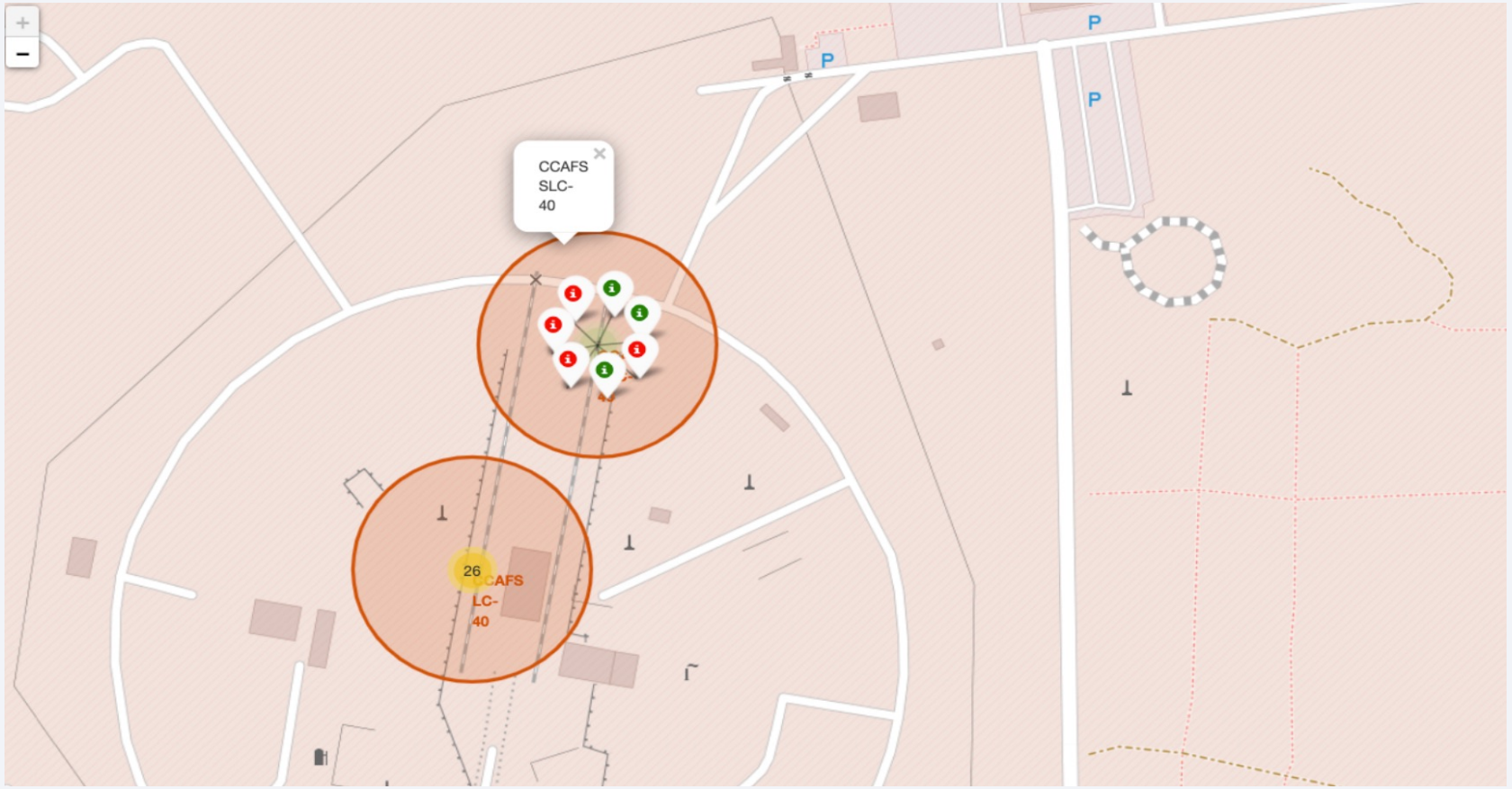
<Folium Map Screenshot 1>

Generated folium map with all launch sites' location markers on a global map



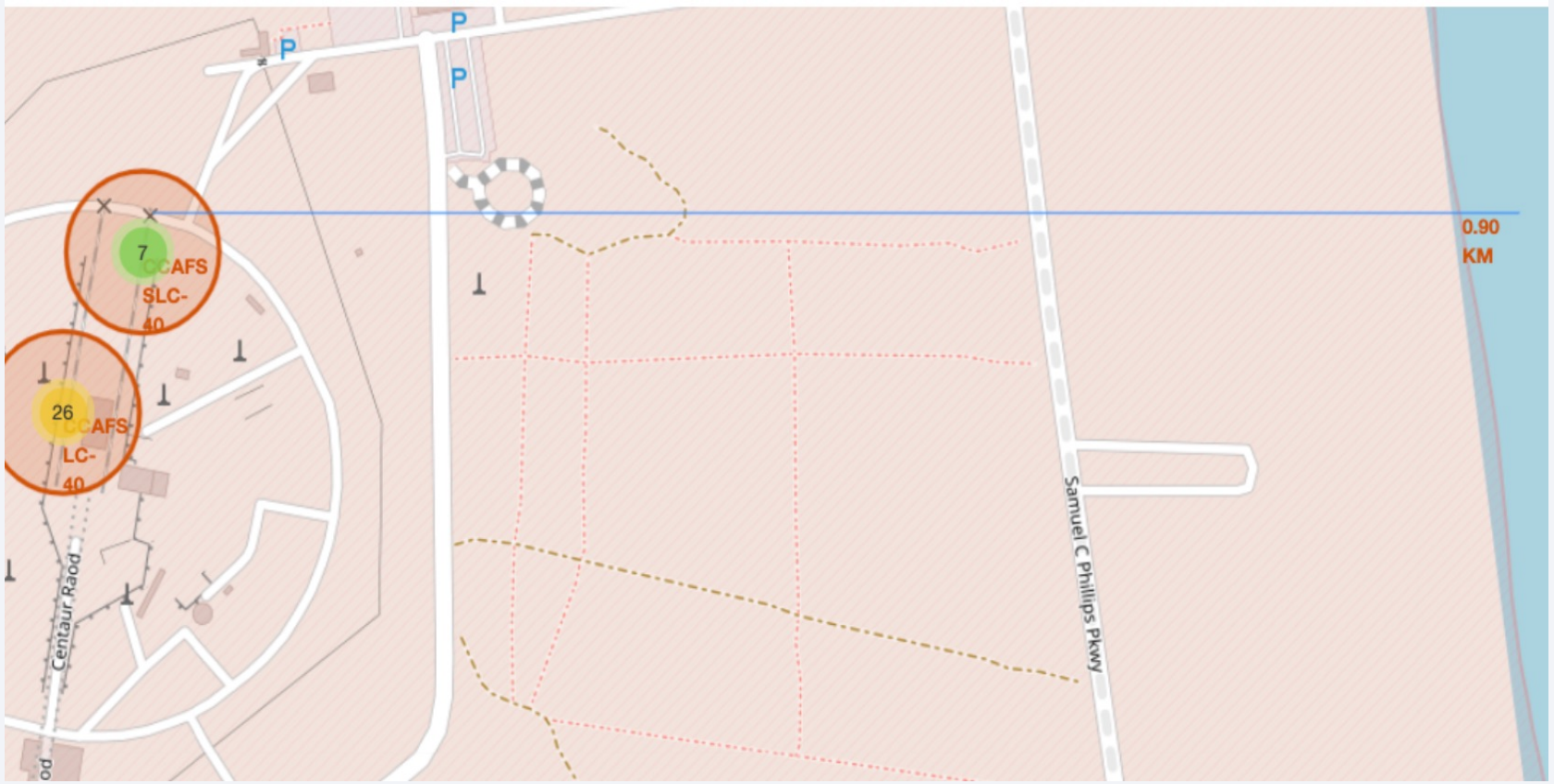
<Folium Map Screenshot 2>

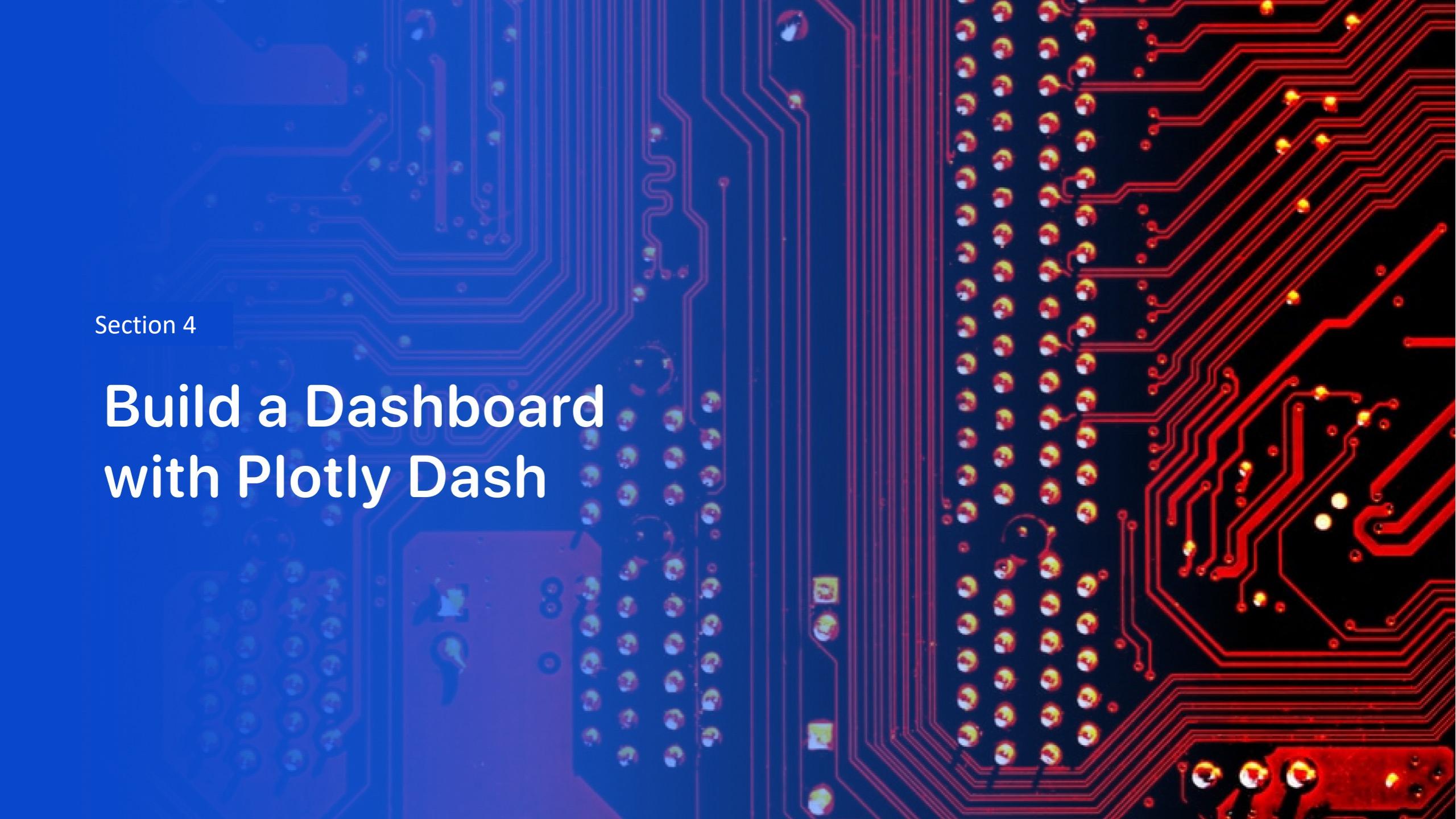
Color-labeled launch outcomes in KSC LS: success (green) failure (red)



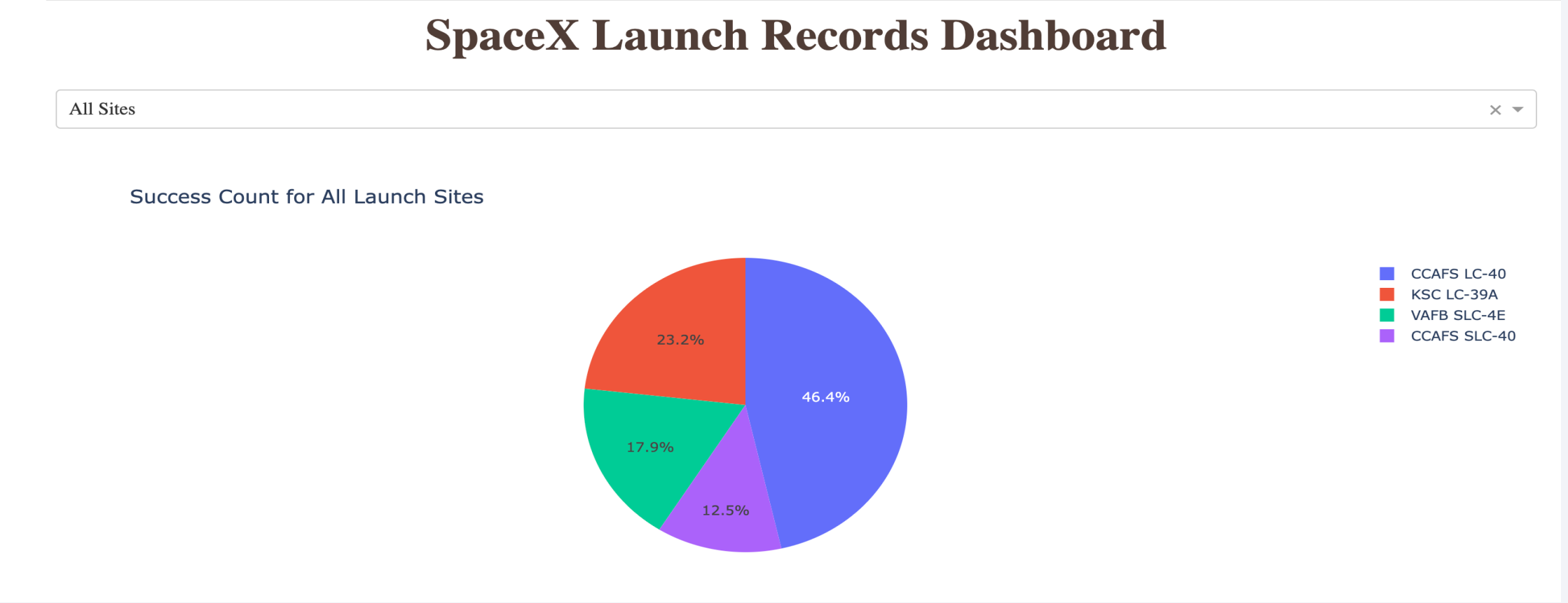
<Folium Map Screenshot 3>

Launch site to its proximity railway, highway, coastline, with distance



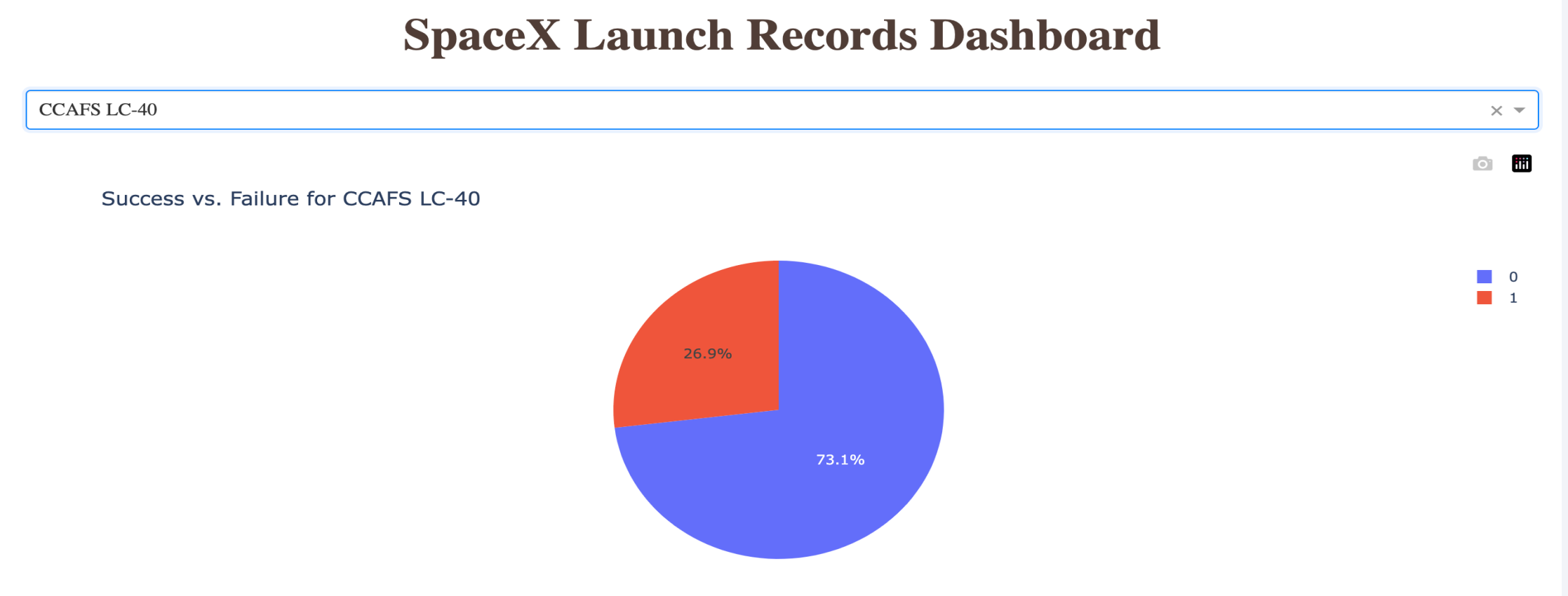

Section 4
Build a Dashboard with Plotly Dash

Interactive Plot: Successful Launches by Site



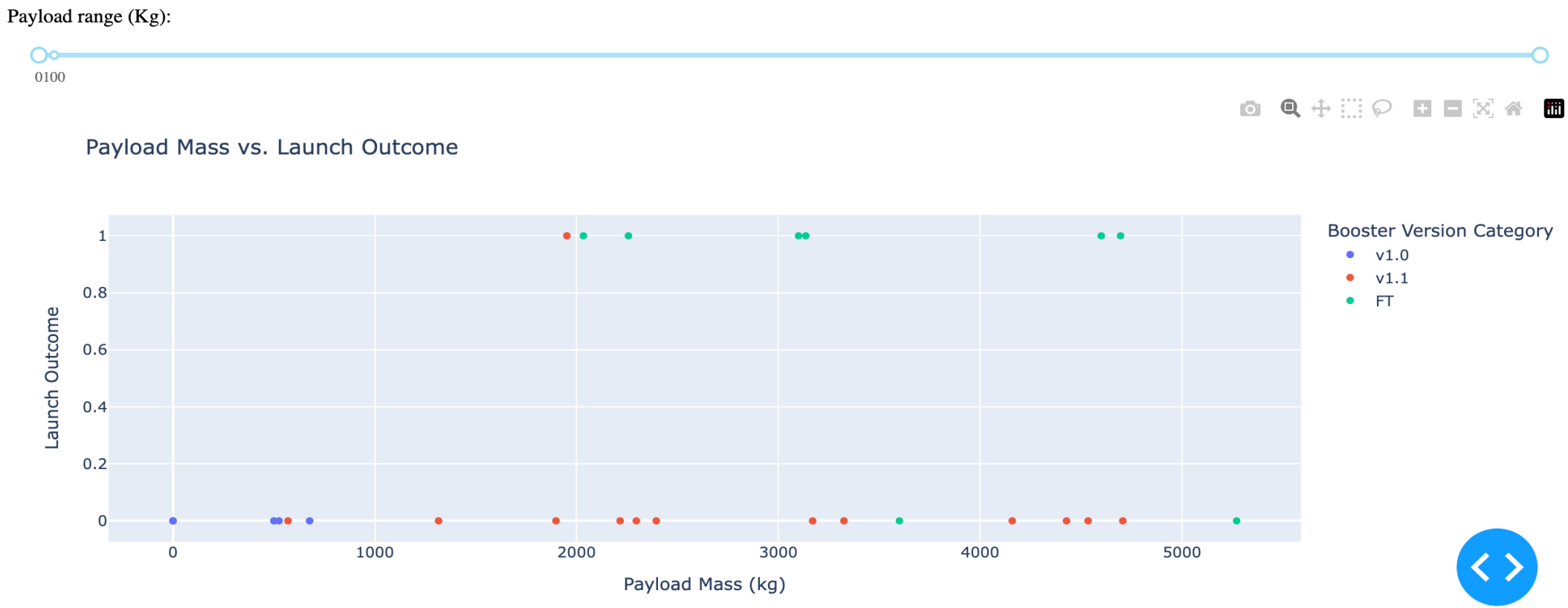
The CCAFS LC-40 site had the highest number of success launches

Interactive Plot: Launch Site with Highest Success Rate



73.1% of launches had success in site CCAFS LC-40

Interactive Plot: Payload Mass vs. Success Class

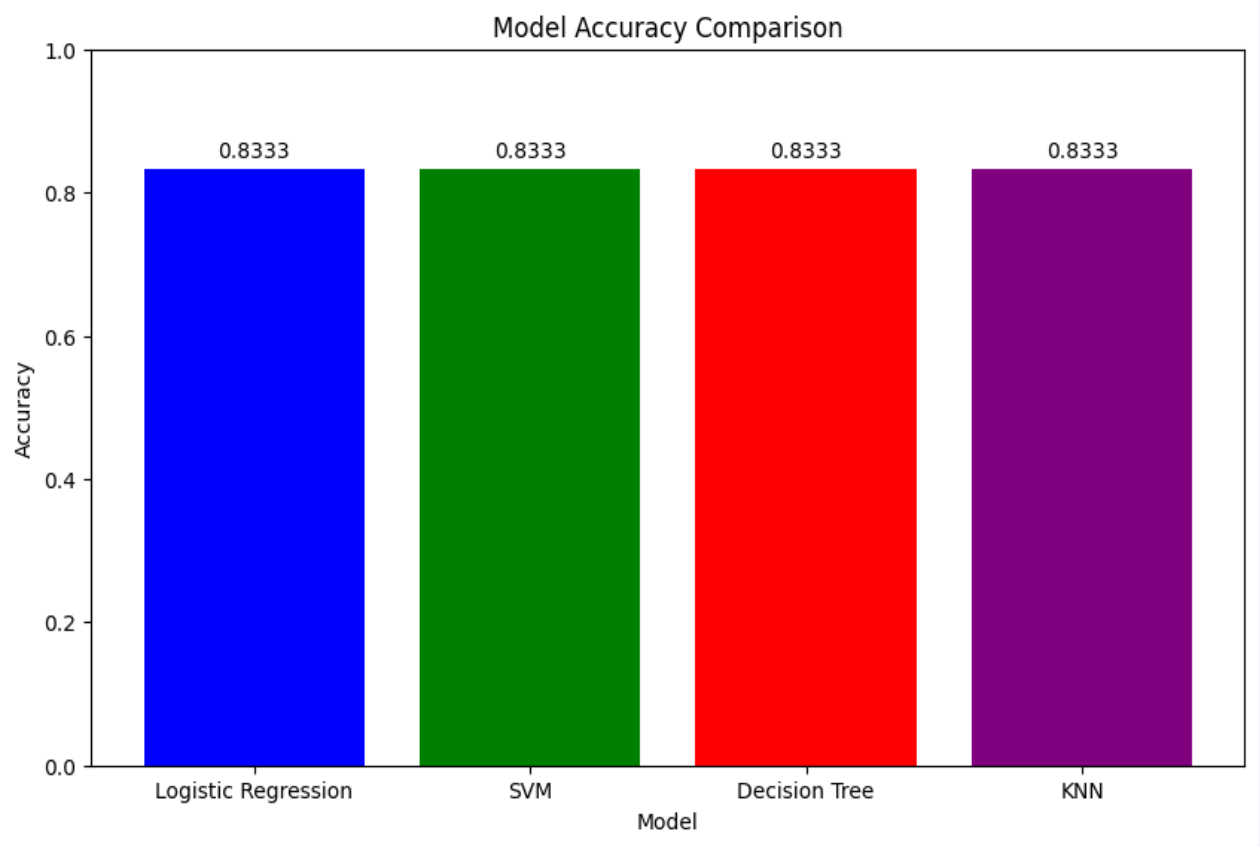


Section 5

Predictive Analysis (Classification)

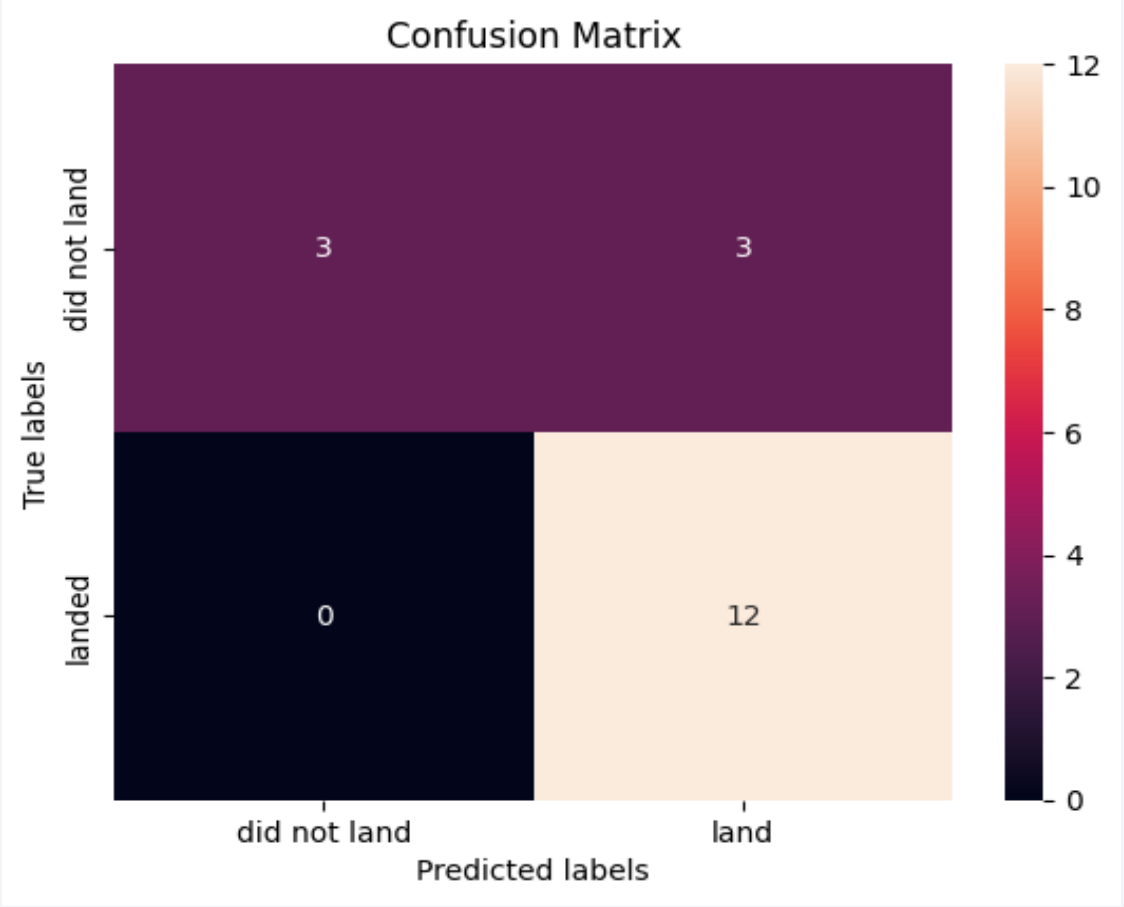
Classification Accuracy

- Logistic Regression accuracy: 0.8333
- SVM accuracy: 0.8333
- Decision Tree accuracy: 0.8333
- KNN accuracy: 0.8333



Confusion Matrix

- The best performing method is: Logistic Regression with an accuracy of 0.8333



Conclusions

- To improve the success rate, the optimal choice is to use the launch site CCAFS LC-40, with a payload mass between 2000-5000 kg and the SSO orbit type.
- Given the payload mass, launch site, and orbit type, we can predict launch success or failure with 87% accuracy using validation data.
- By reusing the first stage of the Falcon 9, SpaceX could potentially save USD 100 million, which could enhance the company's profitability.

Thank you!