IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Parit Pitakhirunpong
2 Jan 2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

This project applies the methodologies I've learned in the IBM Data Science course to predict the success of Falcon 9 first-stage landings. The objective is to build a predictive model using historical launch data to assess whether a given Falcon 9 rocket will successfully land. The project leverages data collected from the SpaceX API, which includes various features such as payload, launch site, orbit type, and flight number. Through exploratory data analysis (EDA), insights were derived from the relationships between these features and launch success. Various machine learning models, including Logistic Regression, Support Vector Machines (SVM), and Decision Trees, were tested and optimized to identify the most accurate model for predicting landing outcomes. The Decision Tree model performed the best, and its performance was evaluated through accuracy metrics and a confusion matrix.

Section 1

Methodology

Methodology

Data Collection:

- A series of helper functions were developed to request rocket launch data from the SpaceX API.
- The response content was decoded into JSON format and then transformed into a structured DataFrame using the `json_normalize` method.

Data Wrangling:

- Missing values were handled by replacing them with the column mean.
- Key features such as the number of launches for each site and orbit were calculated, and a landing outcome label was created from the "Outcome" column for machine learning purposes.

Exploratory Data Analysis (EDA):

- EDA was conducted using SQL queries and visualizations with Pandas, Matplotlib, and Seaborn.
- Key insights from the EDA include:
    - Flight number and launch site: Higher flight numbers show improved success rates for sites like CCAFS SLC 40 and KSC LC 39A.
    - Payload mass vs. launch site: VAFB SLC 4E did not have launches for heavy payloads (>10,000 kg).
    - Success rate vs. orbit type: Orbits like GEO and VLEO exhibit high success rates, while SO (Sun-synchronous Orbit) has 0% success.
    - Payload vs. orbit type: Heavy payloads tend to have higher success rates for orbits like Polar and LEO.
    - Yearly success trends show a steady increase in success rates from 2013 to 2020.

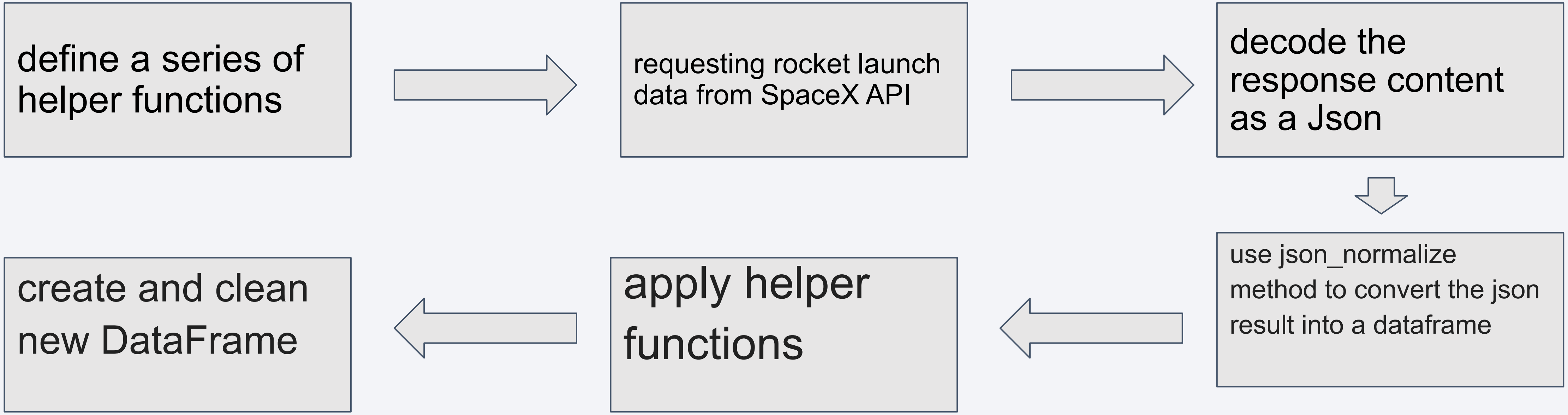
Interactive Data Visualization:

- Interactive visualizations were built using Folium and Plotly Dash to explore spatial and temporal patterns in the data.

Predictive Analysis:

- The data was standardized and split into training and testing sets using the `train_test_split` function.
- Three machine learning models—Logistic Regression, SVM, and Decision Trees—were trained, with the Decision Tree model achieving the highest accuracy.
- Model performance was evaluated through accuracy scores and a confusion matrix, which showed high accuracy in predicting successful landings and some confusion in predicting unsuccessful landings.

Data Collection – SpaceX API



Data Wrangling

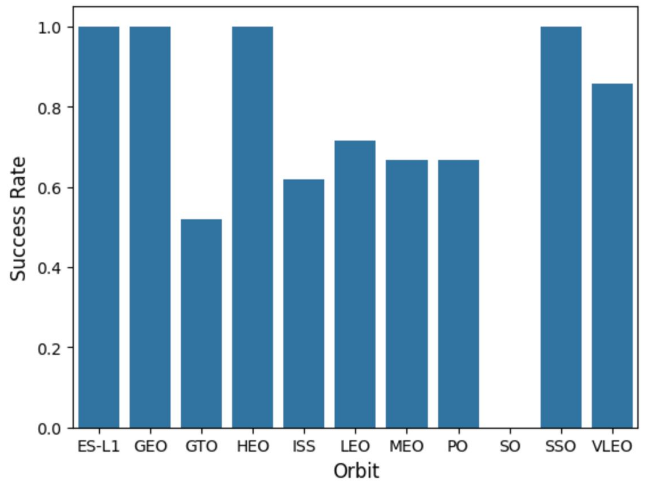
```
LaunchSite
CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
Name: count, dtype: int64
```

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

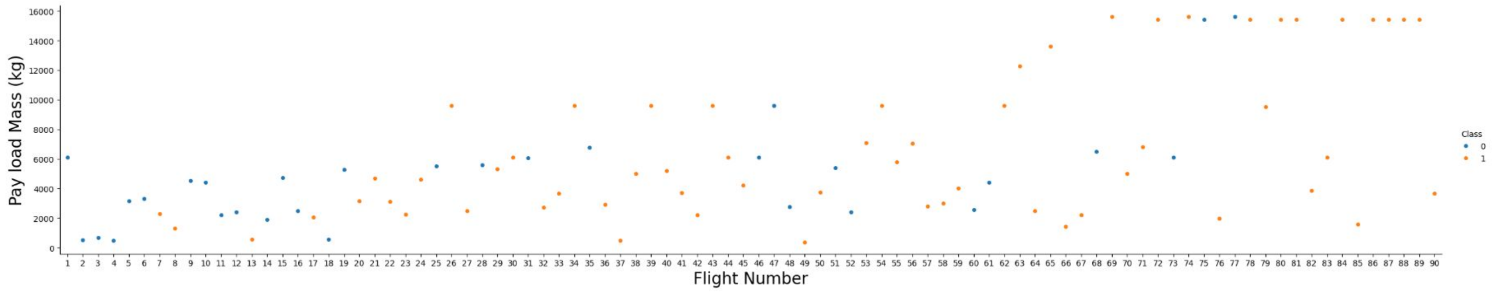
Calculate the number of launches on each site and orbit

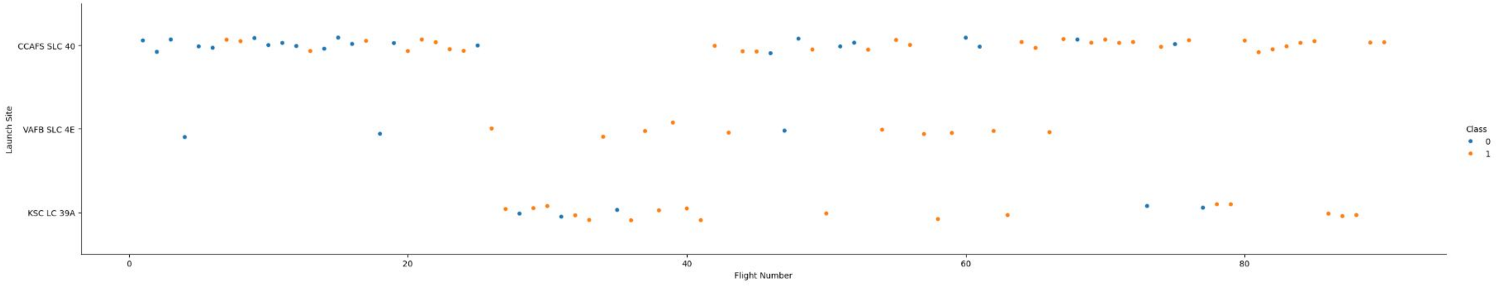
Create a landing outcome label from Outcome column

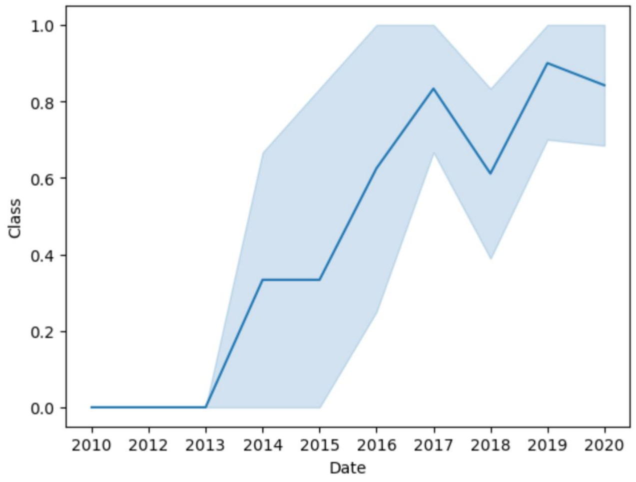
EDA with Data Visualization



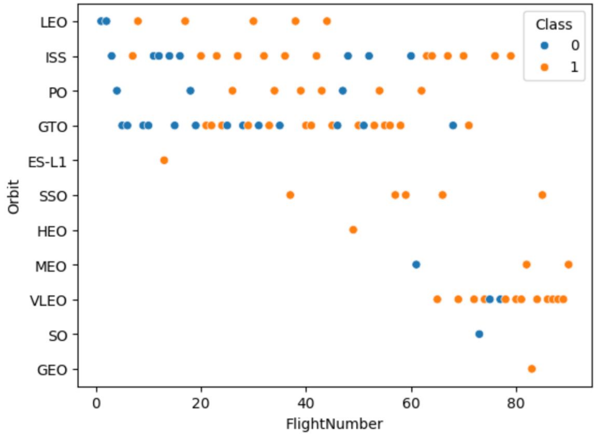
Analyze the plotted bar chart to identify which orbits have the highest success rates.

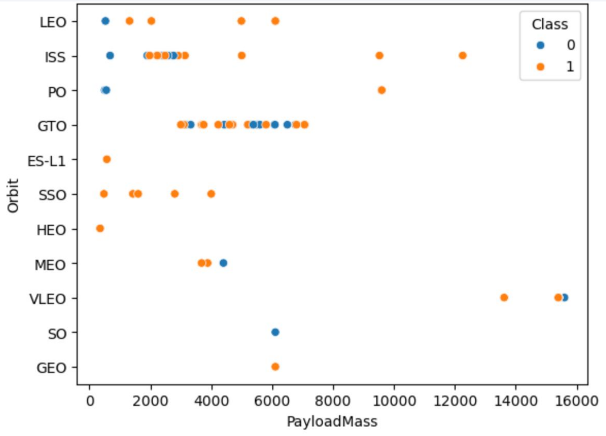






you can observe that the sucess rate since 2013 kept increasing till 2020





EDA with SQL

Here are some key insights from SQL queries.

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 DISTINCT(Landing_Outcome), COUNT(Landing_Outcome) FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND
```

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

Landing_Outcome	COUNT(Landing_Outcome)
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

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

Hint:Use min function

```
%sql SELECT DATE FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)' ORDER BY DATE LIMIT 1
```

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

Date
2015-12-22

List the total number of successful and failure mission outcomes

```
%sql SELECT (SELECT COUNT(Landing_Outcome) FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Failure%') as Faliure, (SE
```

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

Faliure	Success
10	61

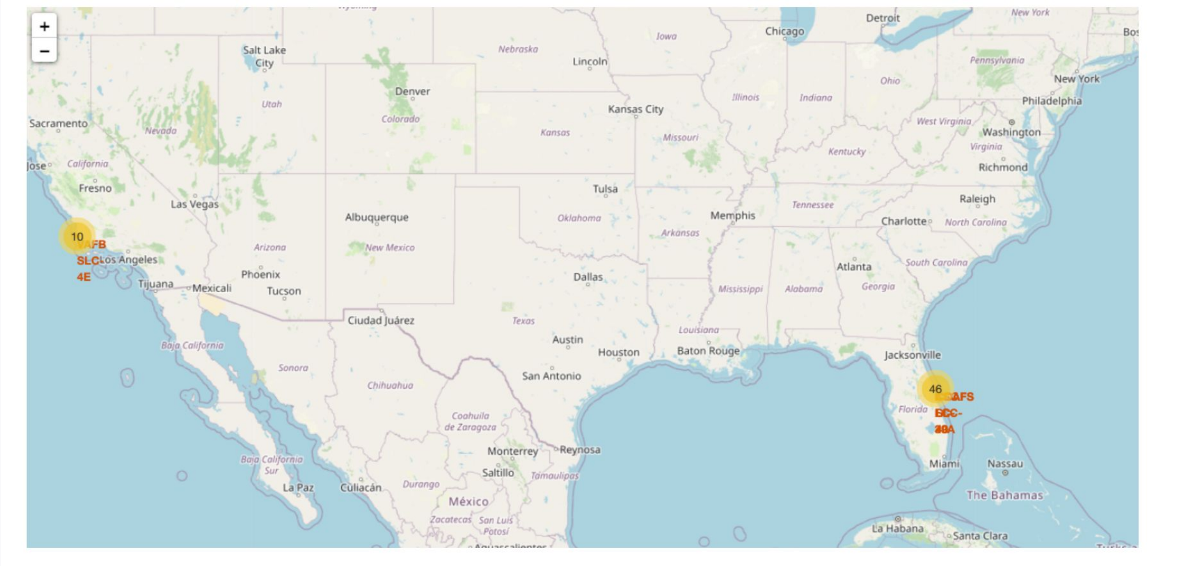
Display the total payload mass carried by boosters launched by NASA (CRS)

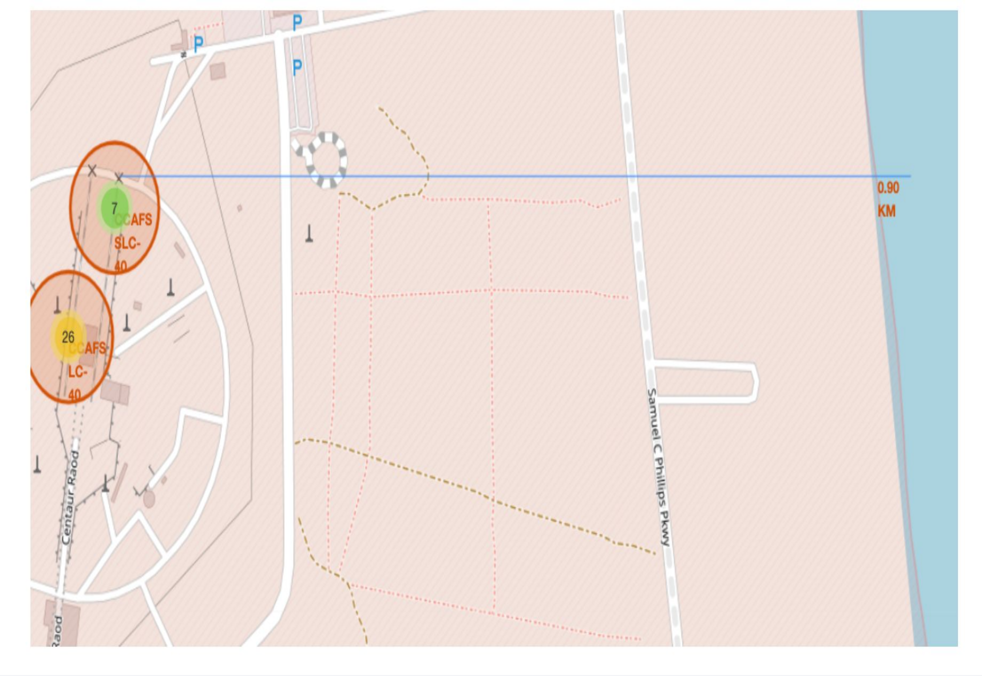
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total payload mass carried by boosters launched by NASA (CRS)'  FROM SPACE
```

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

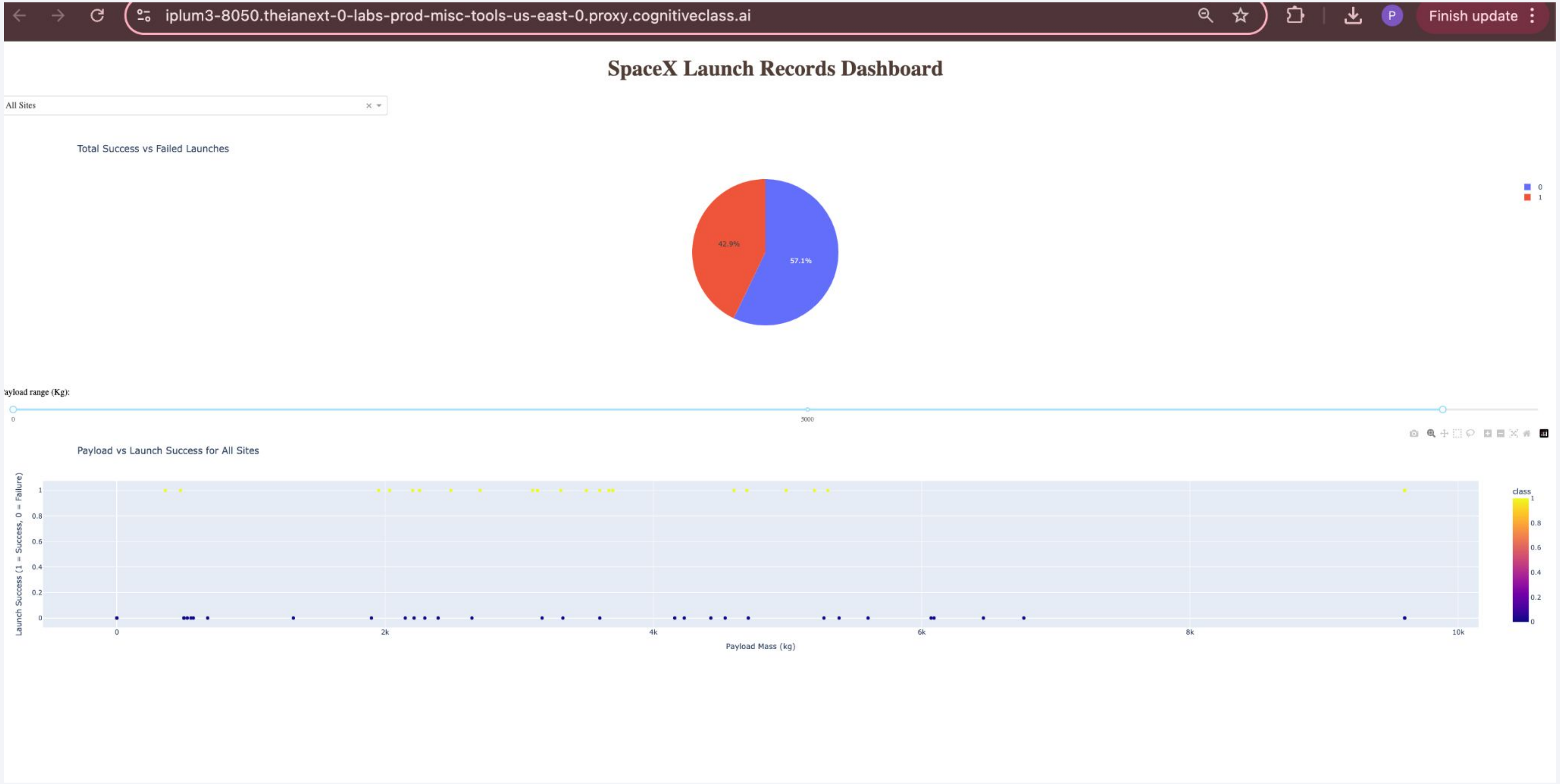
Total payload mass carried by boosters launched by NASA (CRS)
48213

Build an Interactive Map with Folium

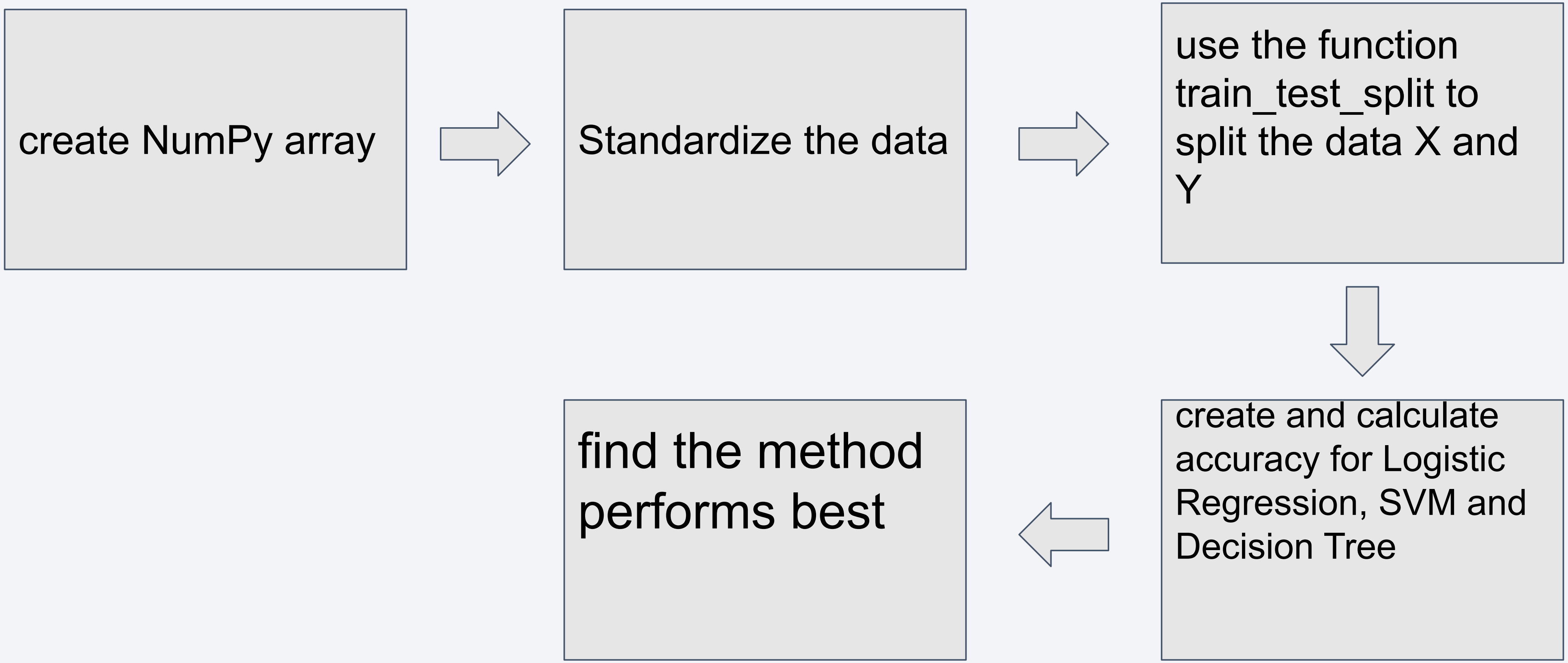




Build a Dashboard with Plotly Dash



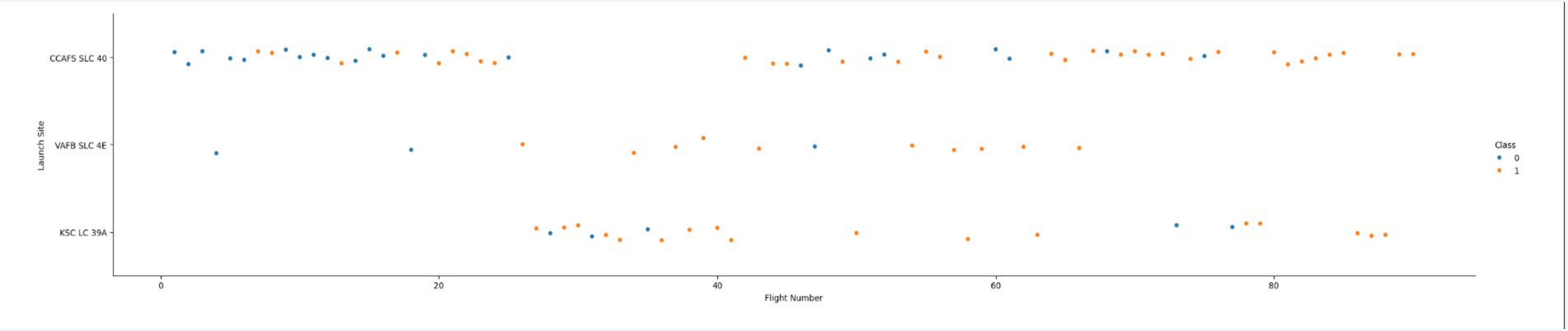
Predictive Analysis (Classification)



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



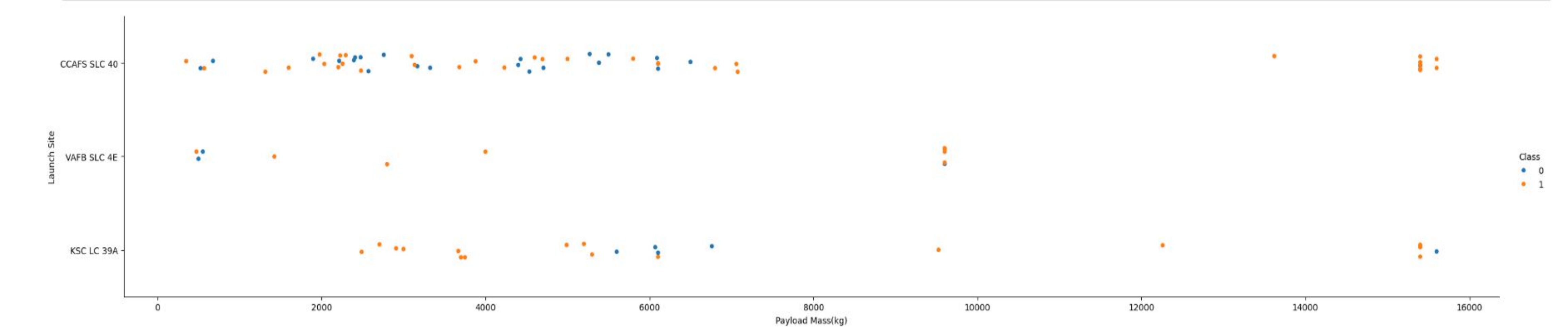
For **CCAFS SLC 40**, there isn't a strong relationship initially, but as the flight number exceeds 60, there's a noticeable increase in the success rate.

Similarly, for **KSC LC 39A**, higher flight numbers also show a significant improvement in success rates.

However, **VAFB SLC 4E** has limited data, making it challenging to draw meaningful conclusions.

Payload vs. Launch Site

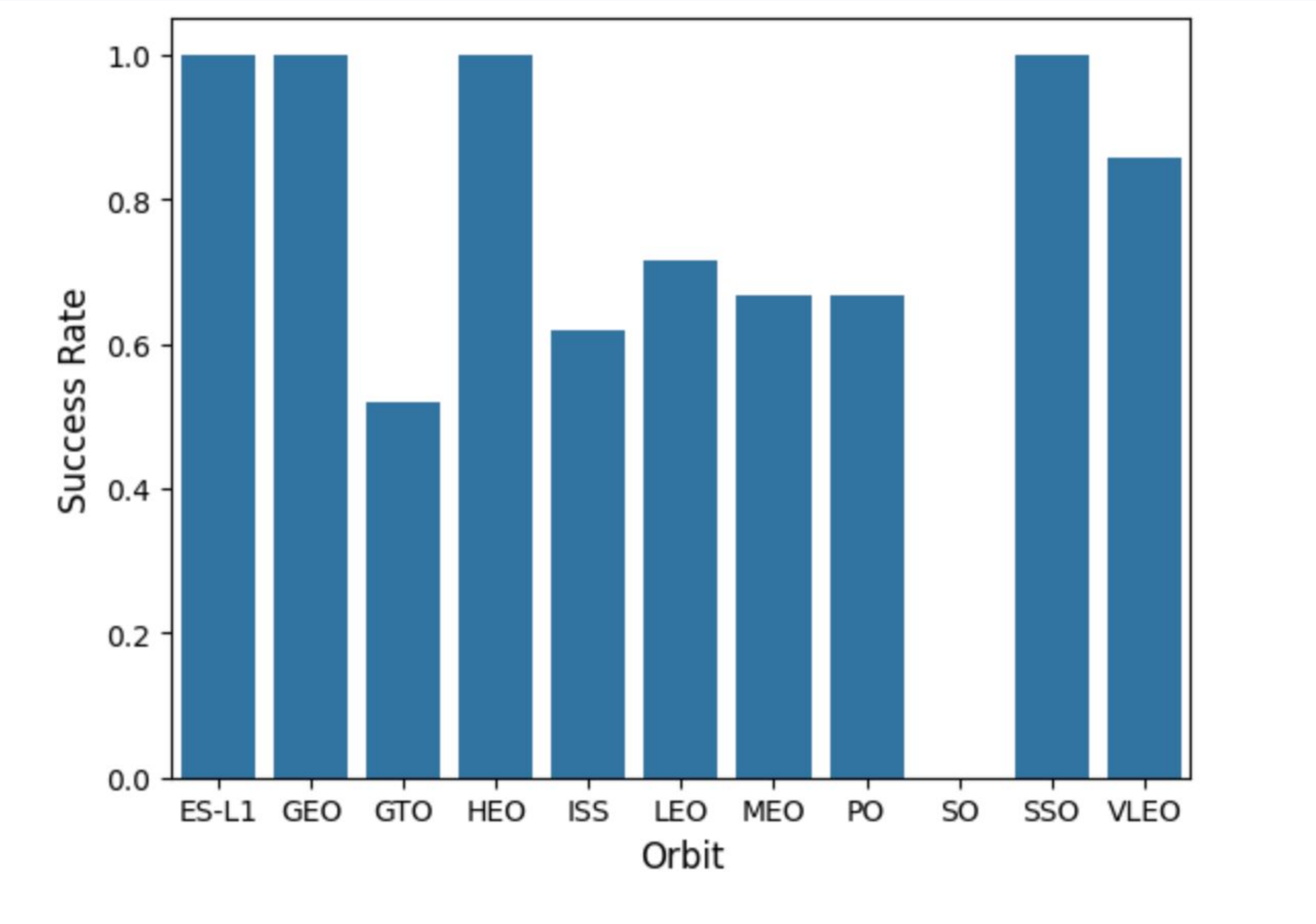
Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).



Success Rate vs. Orbit Type

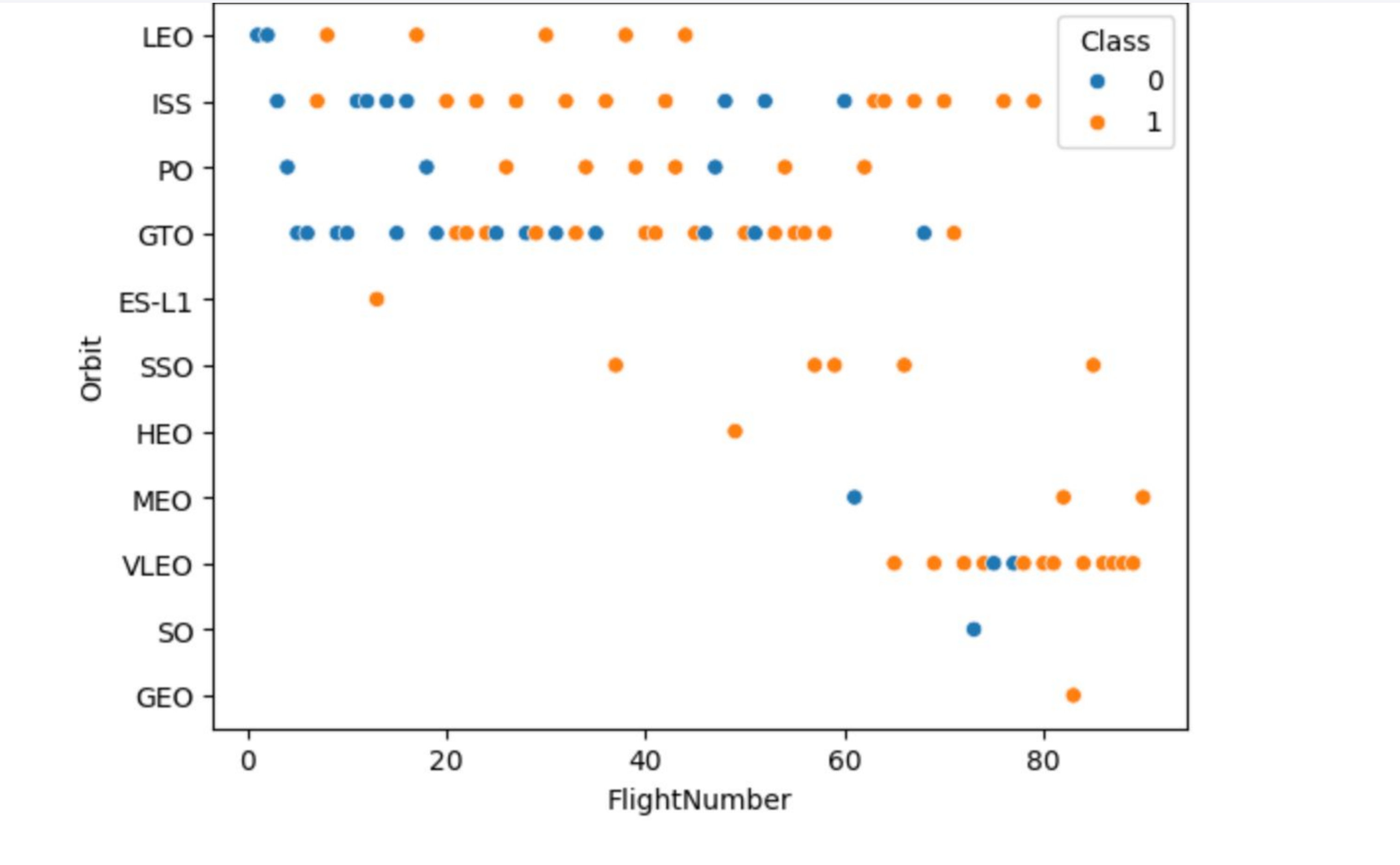
The orbit types can be grouped by success rate into three categories:

1. **High success rate (100% or nearly so):** Orbits like GEO and VLEO.
2. **Moderate success rate (around 50%):** Orbits like MEO and PO.
3. **Low success rate (0%):** SO, the single orbit with no successful launches.



Flight Number vs. Orbit Type

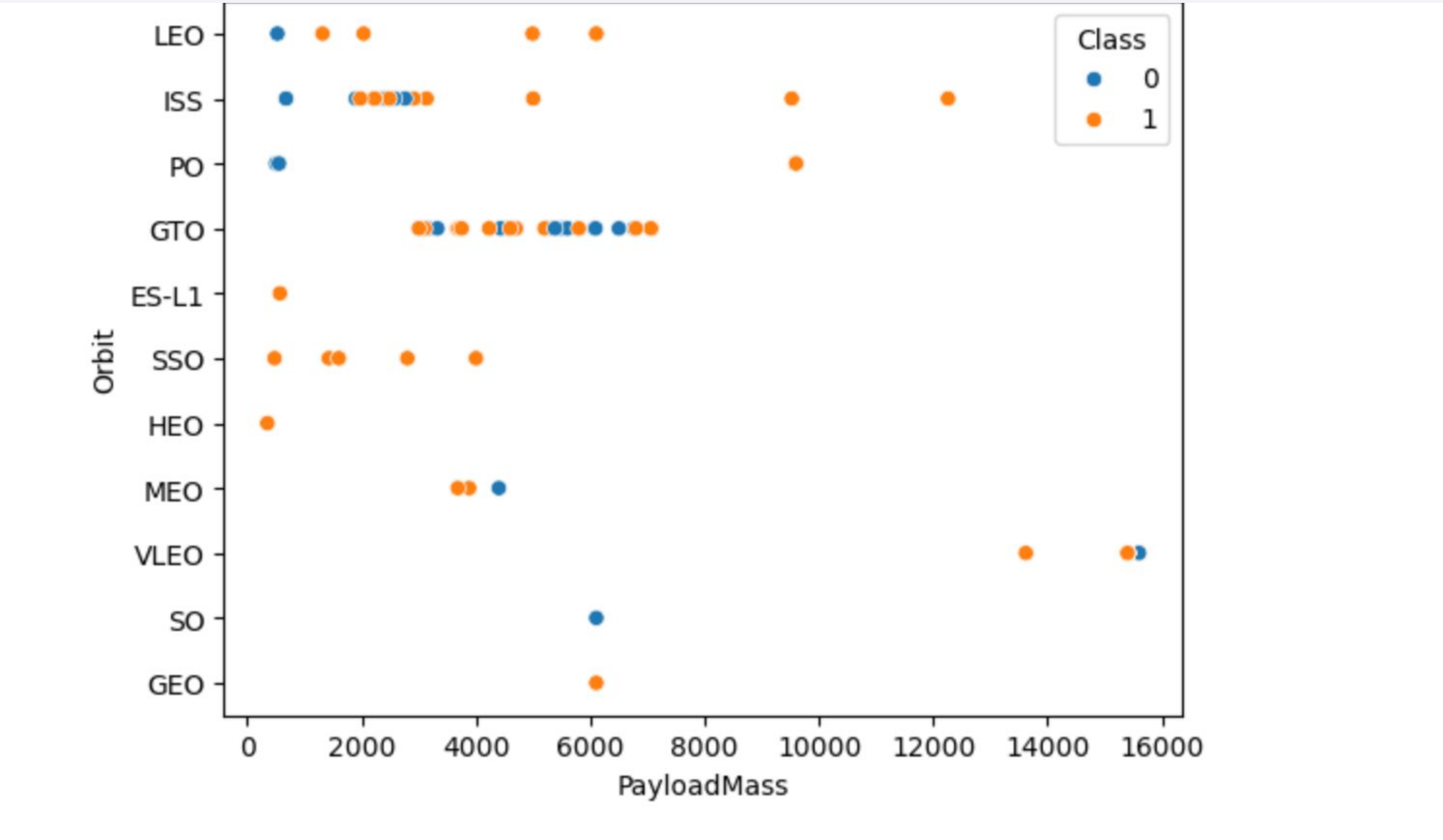
You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.



Payload vs. Orbit Type

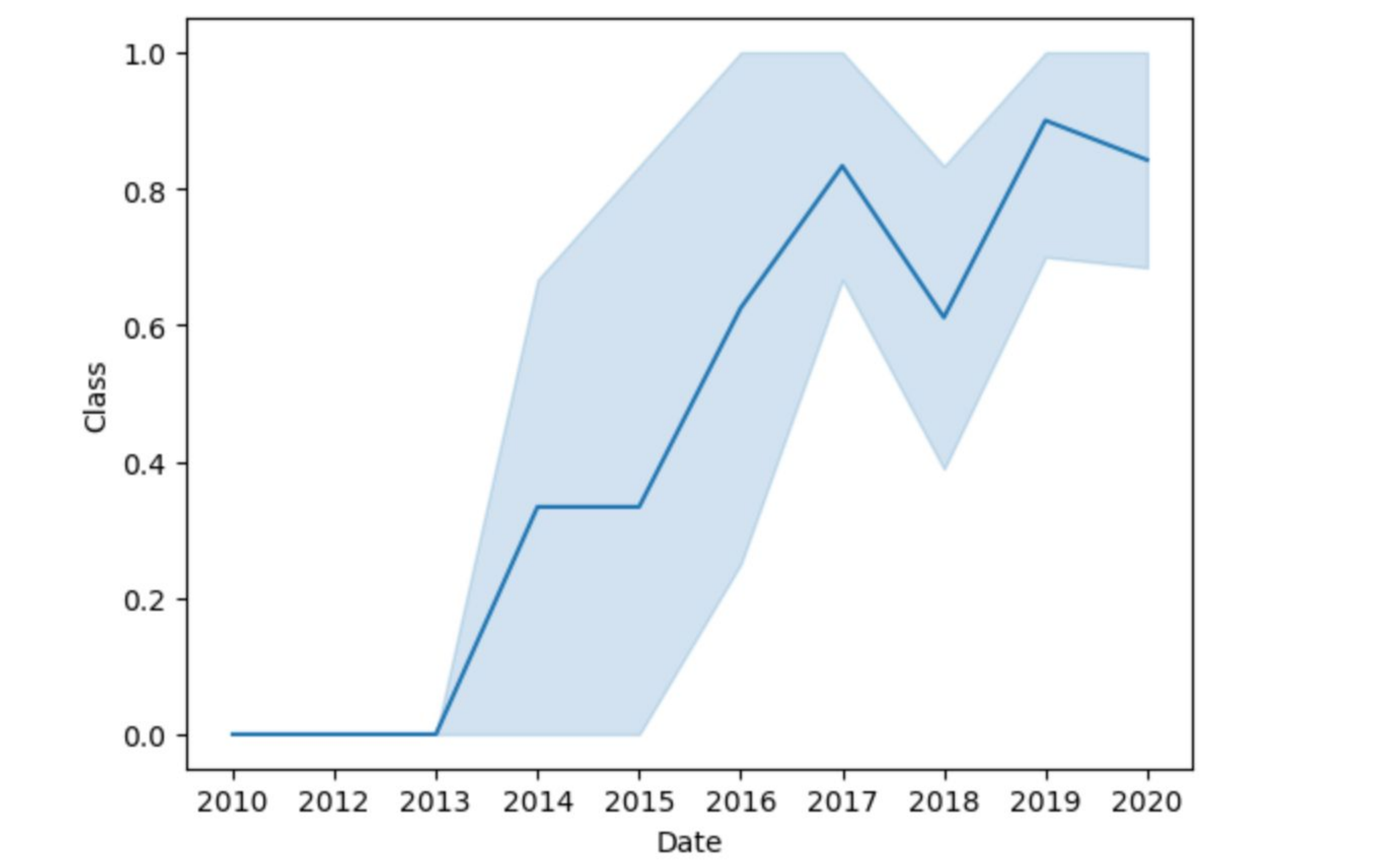
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

you can observe that the success rate since 2013 kept increasing till 2020



All Launch Site Names

```
select DISTINCT(Launch_Site) from SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5

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

```
SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total payload mass carried by boosters launched by NASA (CRS)'  FROM SPACEXTBL WHERE Customer LIKE '%CRS%'
```

Total payload mass carried by boosters launched by NASA (CRS)
48213

Average Payload Mass by F9 v1.1

```
SELECT AVG(PAYLOAD_MASS__KG_) AS 'Average payload mass carried by booster version F9 v1.1' FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1'
```

Average payload mass carried by booster version F9 v1.1
2928.4

First Successful Ground Landing Date

```
SELECT DATE FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)' ORDER BY DATE LIMIT 1
```

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
SELECT Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND
PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

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

Total Number of Successful and Failure Mission Outcomes

SELECT (SELECT COUNT(Landing_Outcome) FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Failure%') as Faliure, (SELECT COUNT(Landing_Outcome) FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Success%') as Success

Faliure	Success
10	61

Boosters Carried Maximum Payload

SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)

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

SELECT substr(Date, 6, 2) AS 'Month', (SELECT Landing_Outcome FROM SPACEXTBL WHERE Landing_Outcome = 'Failure (drone ship)') AS 'Landing Outcome', Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr(Date, 1, 4) = '2015'

Month	Landing Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Failure (drone ship)	F9 v1.1 B1013	CCAFS LC-40
03	Failure (drone ship)	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1016	CCAFS LC-40
06	Failure (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Failure (drone ship)	F9 FT B1019	CCAFS LC-40

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

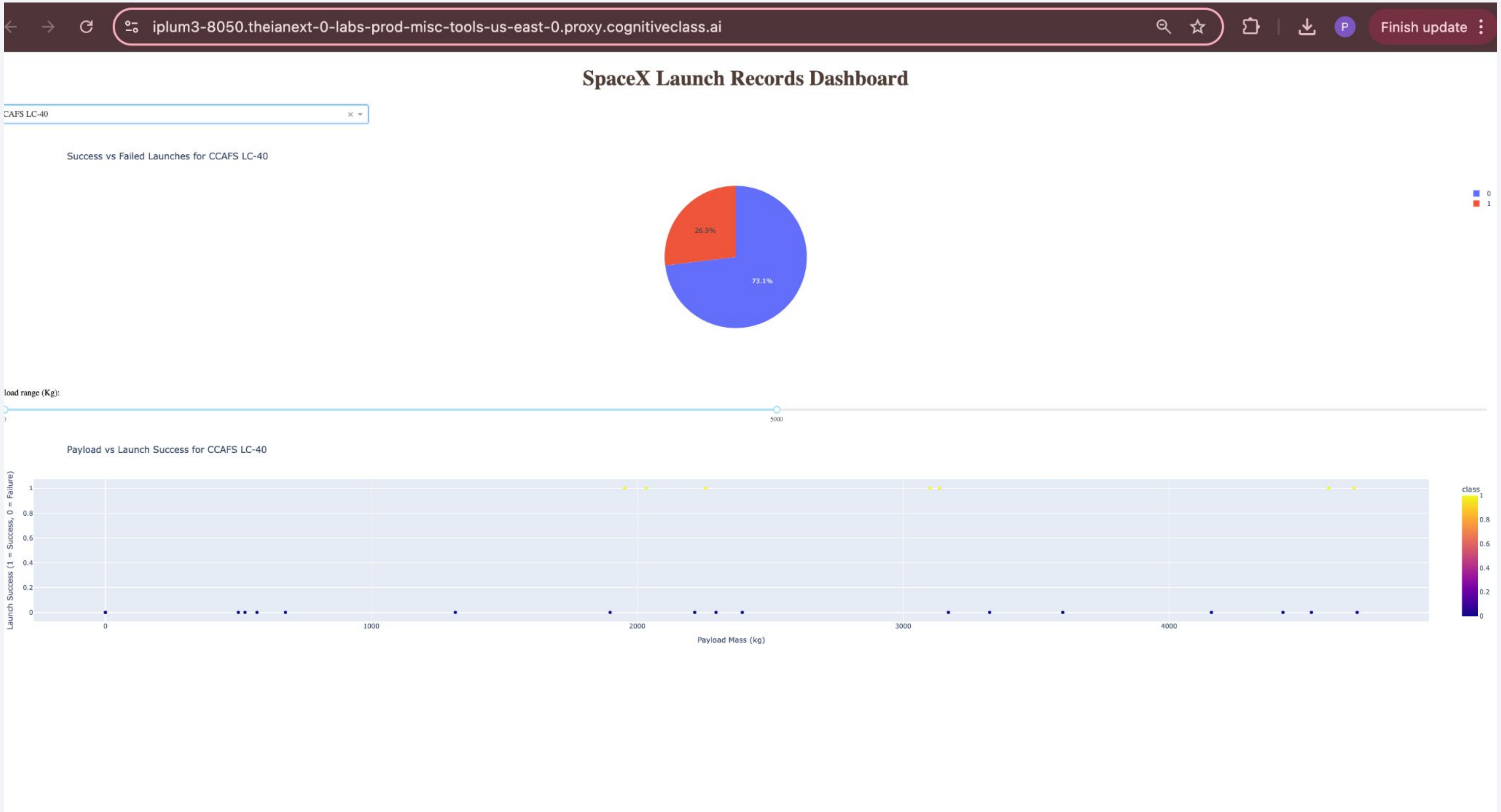
SELECT DISTINCT(Landing_Outcome), COUNT(Landing_Outcome) FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC

Landing_Outcome	COUNT(Landing_Outcome)
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

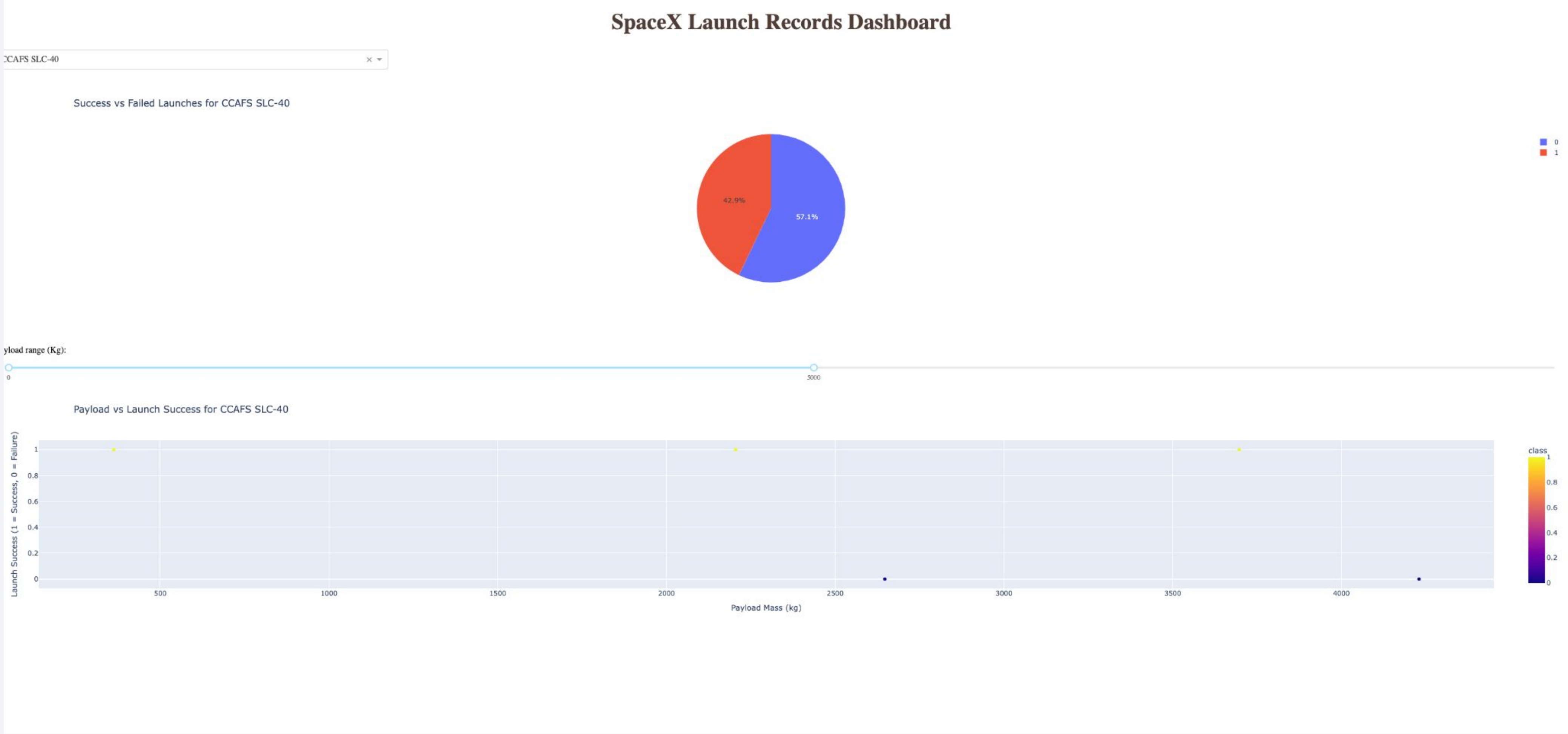
Section 3

Launch Sites Proximities Analysis

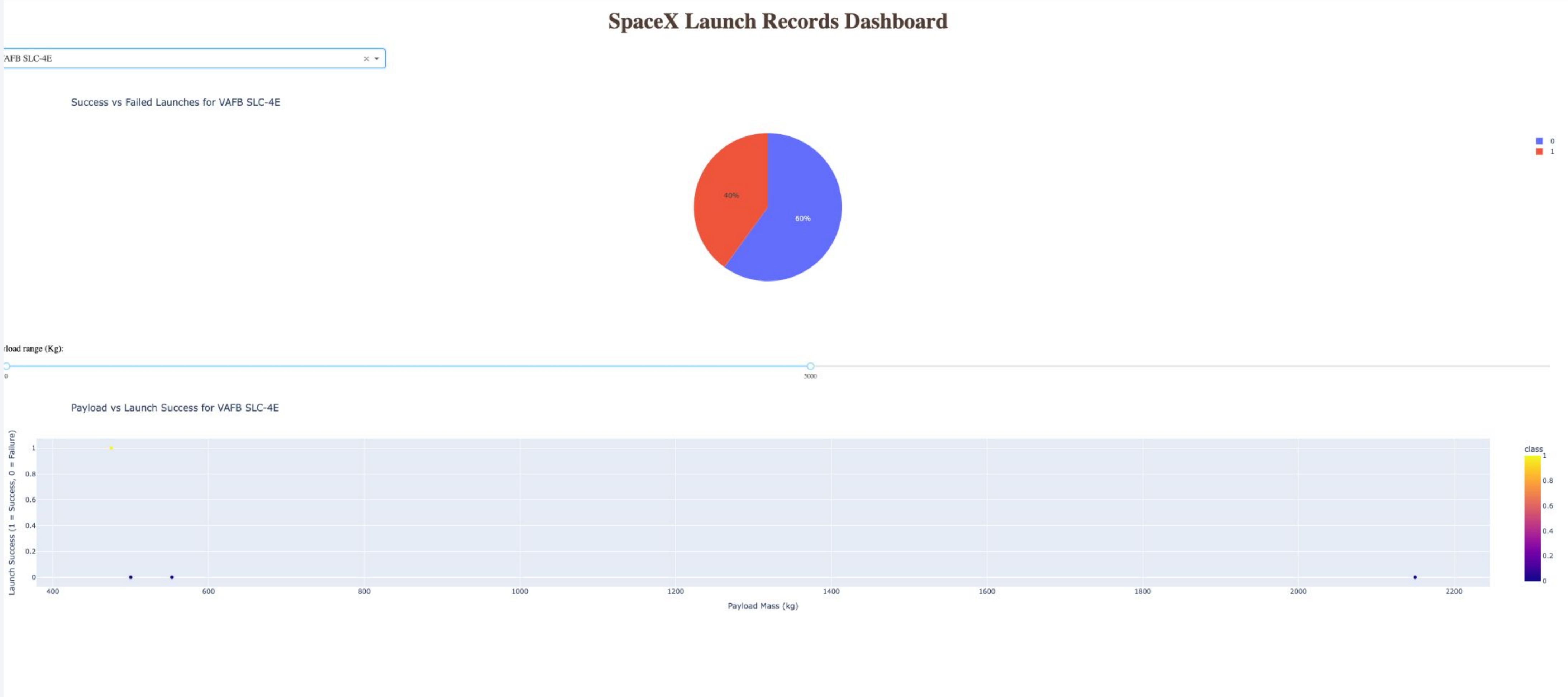
Success vs Failed Launches for CCAFS LC-40



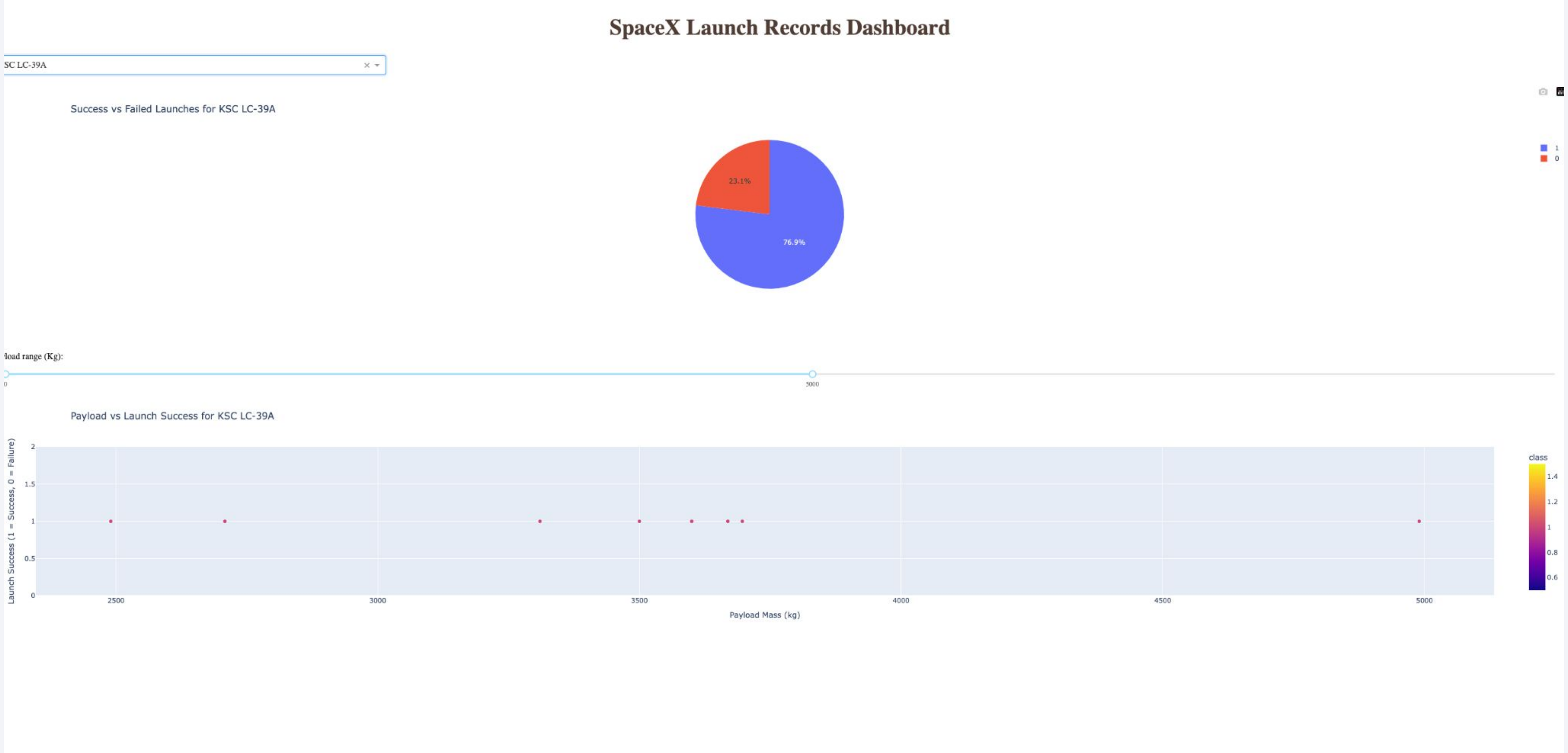
Success vs Failed Launches for CCAFS SLC-40



Success vs Failed Launches for VAFB SLC-4E



Success vs Failed Launches for KSC LC 39-A

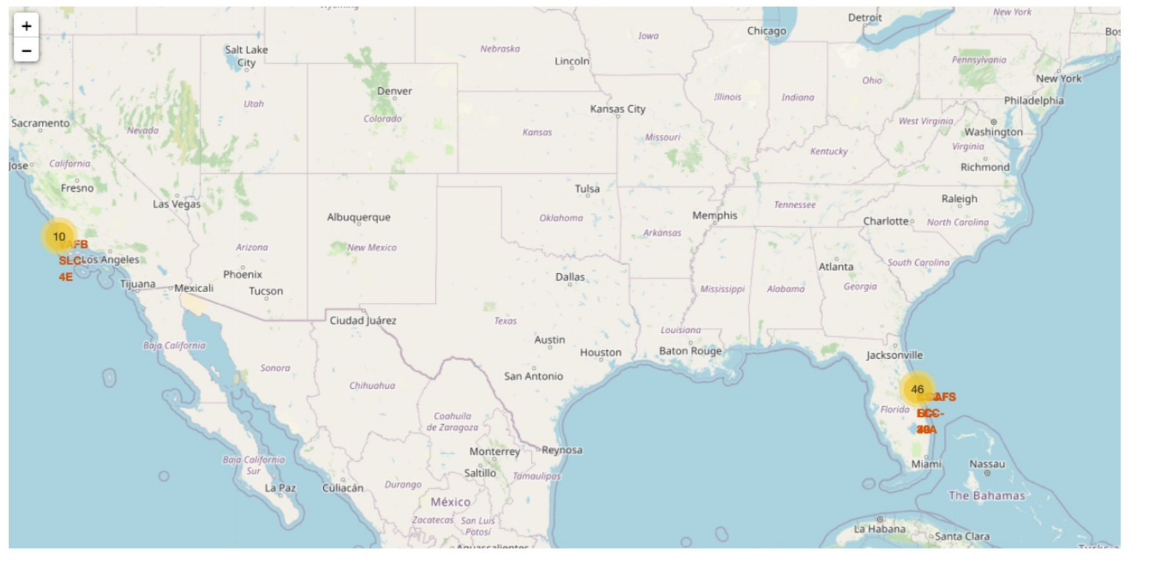


Section 4

Build a Dashboard with Plotly Dash

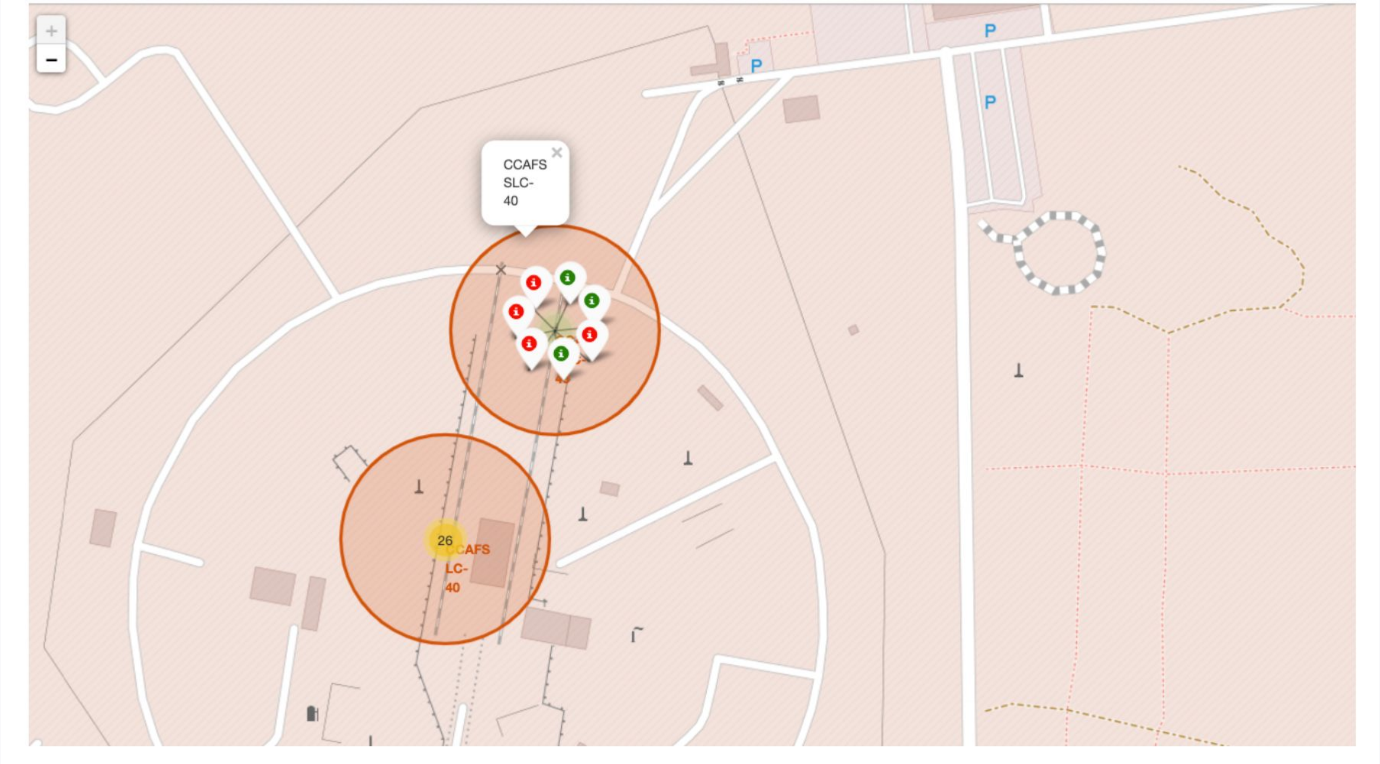
Mark all launch sites on a map

We could use folium.Circle to add a highlighted circle area with a text label on a specific coordinate.



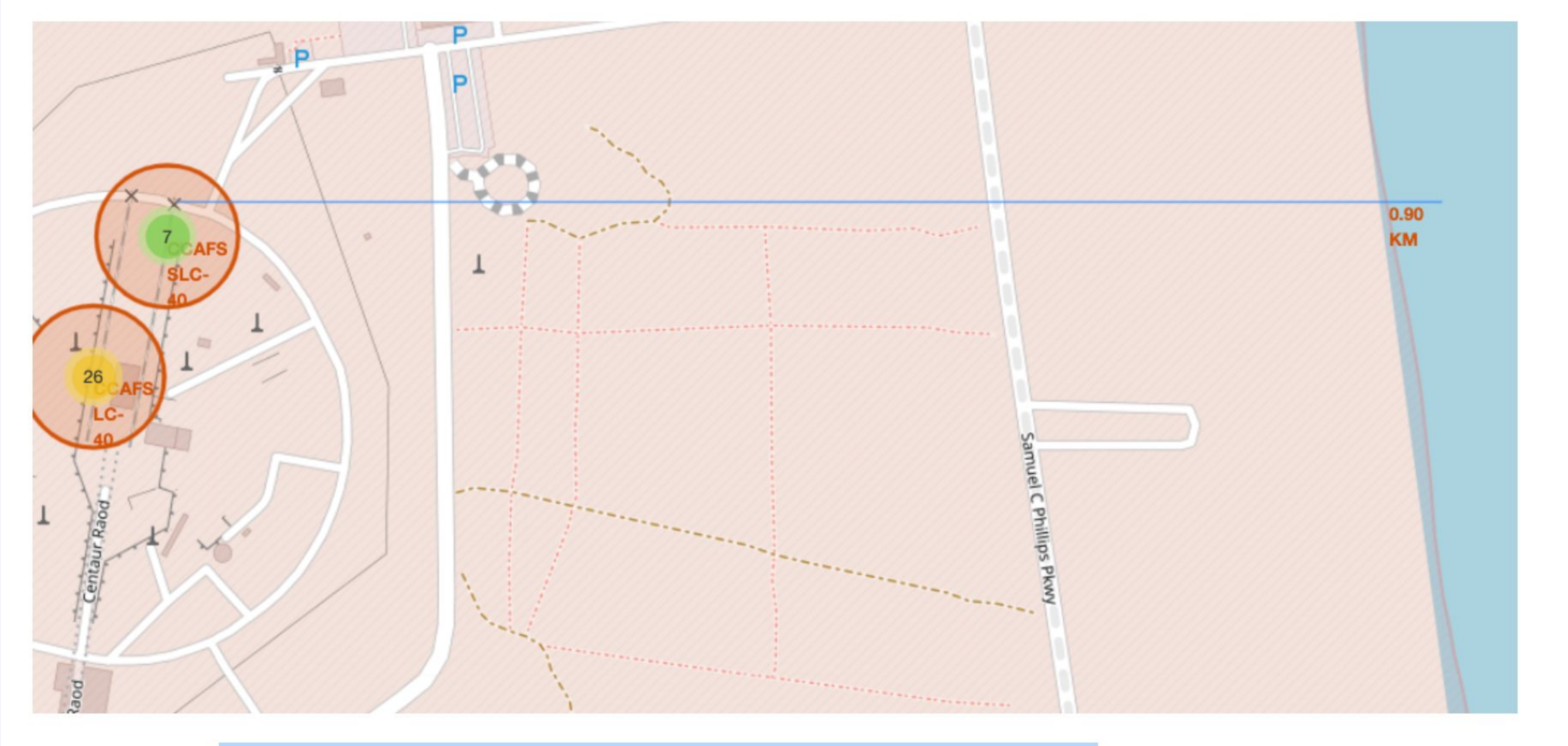
Mark the success/failed launches for each site on the map

From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.



Calculate the distances between a launch site to its proximities

you can draw a line between a launch site to its closest city, railway, highway, etc.

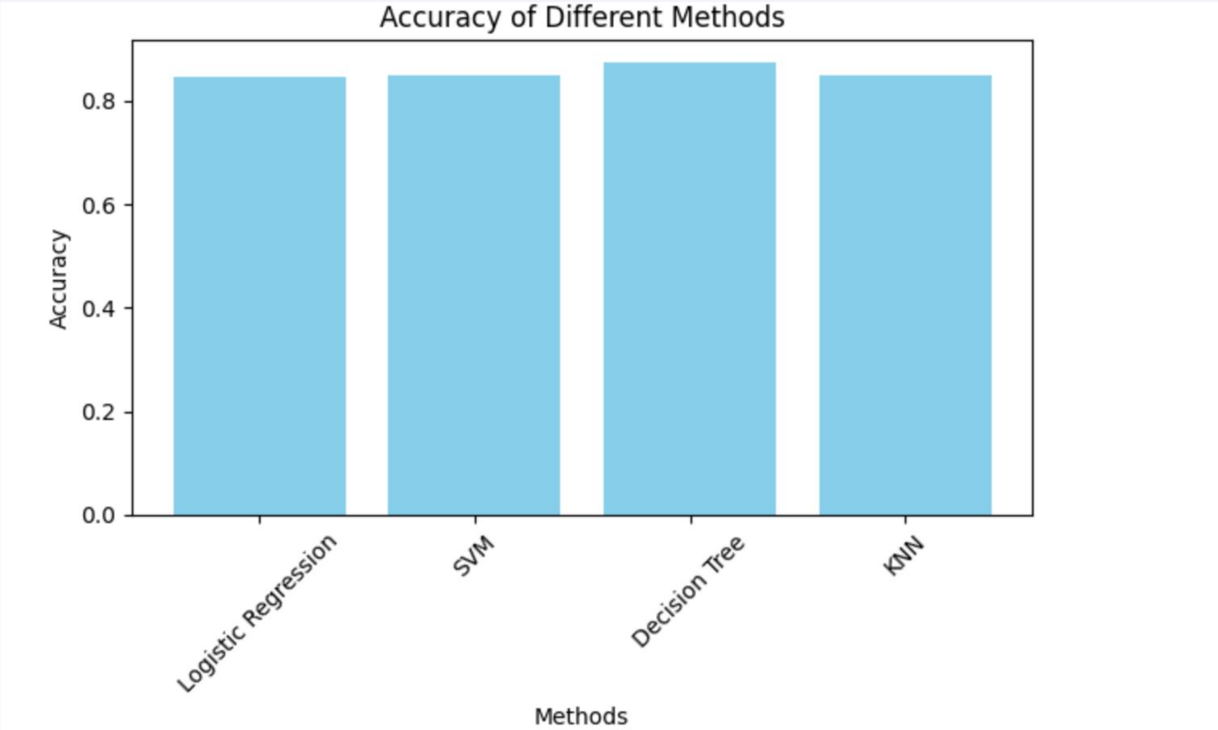


Section 5

Predictive Analysis (Classification)

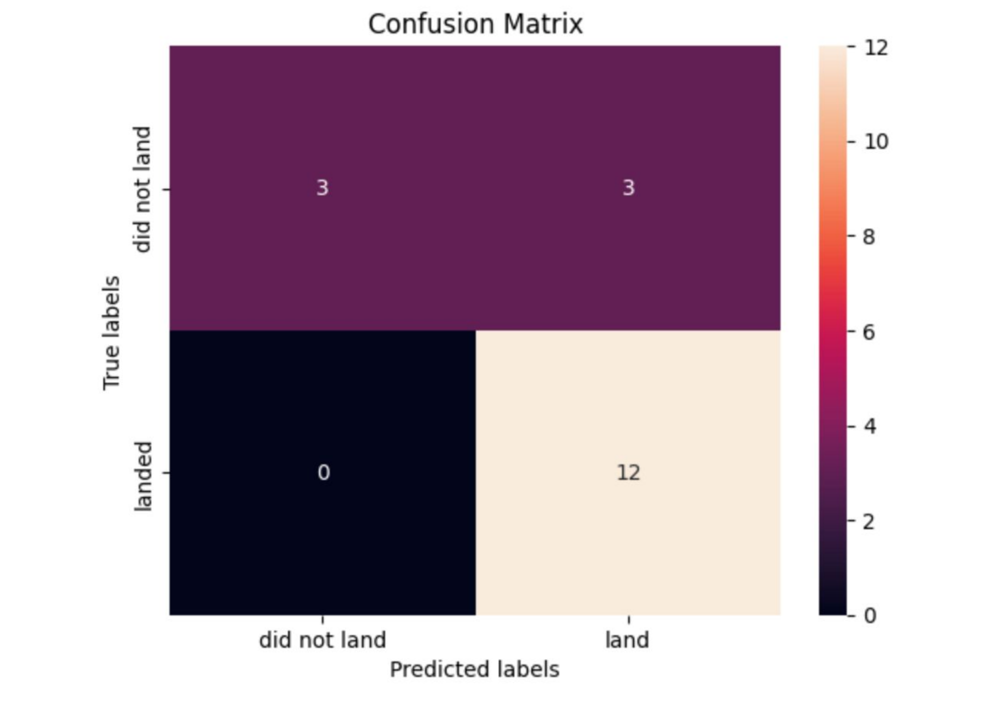
Classification Accuracy

Decision Tree have the highest accuracy.



Confusion Matrix

The model correctly predicted "landed" for 12 instances, correctly predicted "did not land" for 3 instances, incorrectly predicted "landed" for 3 cases where it actually did not land and did not make any errors where it predicted "did not land" but it actually landed.



Observations:

- The model has a **high success rate** for predicting successful landings (12 true positives).
- There is some confusion in predicting unsuccessful landings, leading to 3 false positives.
- A **perfect recall for "landed"** is evident because no false negatives are observed.

Conclusions

This project successfully utilized data science techniques to predict Falcon 9 first-stage landing success. The predictive model demonstrated the importance of flight number, launch site, payload mass, and orbit type in determining landing success. Through the application of machine learning algorithms, the Decision Tree model emerged as the best-performing classifier, with high accuracy and a perfect recall for successful landings. Despite some confusion in predicting unsuccessful landings, the model provides valuable insights into the factors influencing Falcon 9 landing outcomes. These findings contribute to the improvement of predictive modeling in aerospace applications and can inform future SpaceX missions.

Thank you!