Falcon 9 Data Analysis and Predictions

Felix Jwakdak Joshua

January 2024

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- Data collection and data wrangling methodology
- Predictive analysis methodology
- Predictive analysis (classification) results
- EDA and interactive visual analytics
- EDA with visualization results
- EDA with SQL results
- Interactive map with Folium results
- Plotly Dashboard results

INTRODUCTION

In this presentation, we will be showing results of all the points mentioned in the summary page with regards to the SpaceX Falcon 9 rocket launches. From the Data collection and Data wrangling process to the interactive visuals' analytics and Dashboard results display. This will give an insight on the whole Space X Falcon 9 rocket launches.

METHODOLOGY

Data collection and data wrangling methodology (Collecting the data):

In this part We made a get request to the SpaceX API. We also did some basic data wrangling and formatting.

- Request to the SpaceX API
- Clean the requested data

METHODOLOGY

Data collection and data wrangling methodology (Collecting the data):

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A
2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A
4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C
5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C
6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003

METHODOLOGY

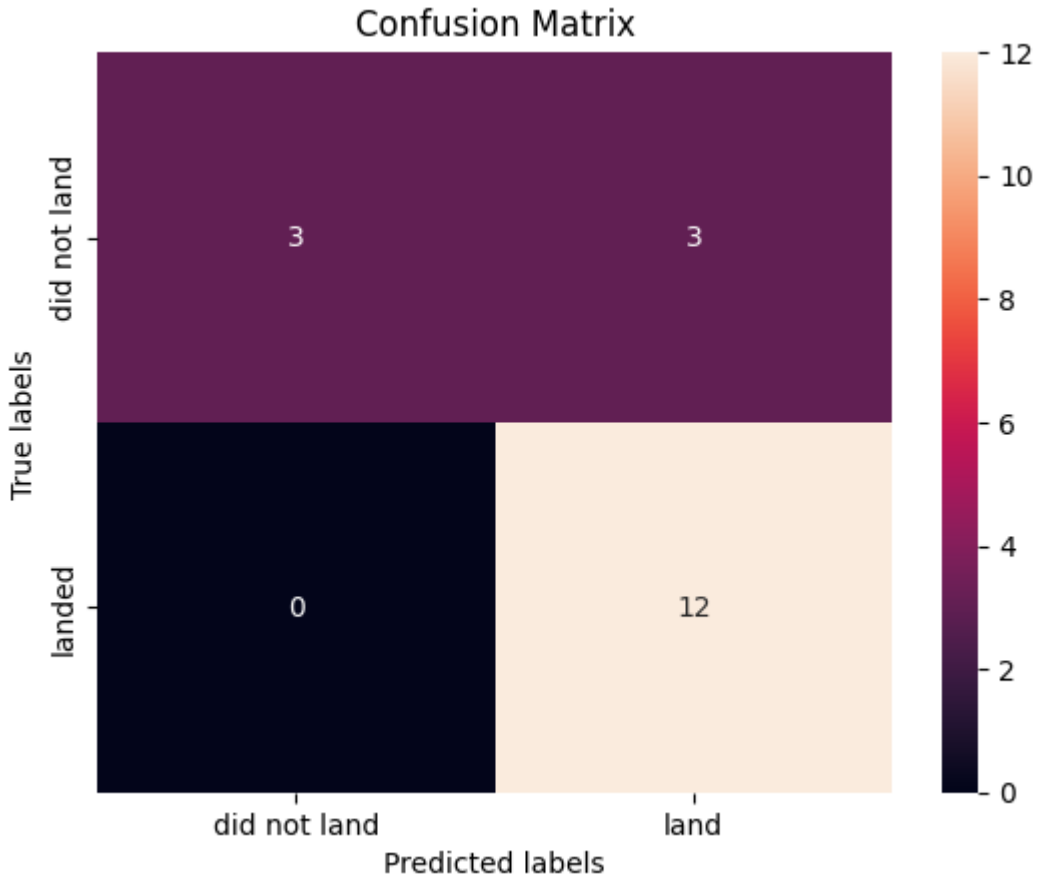
Predictive analysis methodology

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. We created a machine learning pipeline to predict if the first stage will land given the data we got.

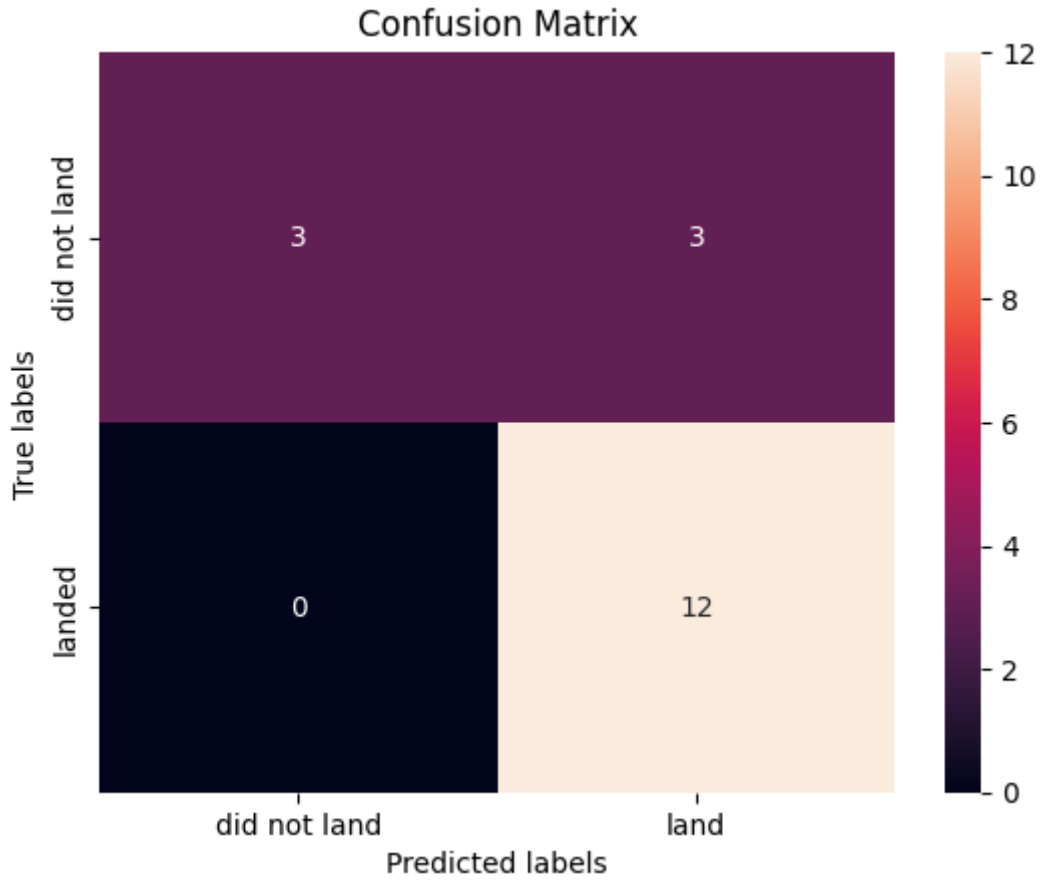
Objectives

- Perform exploratory Data Analysis and determine Training Labels
- create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

RESULTS Predictive analysis (classification) results

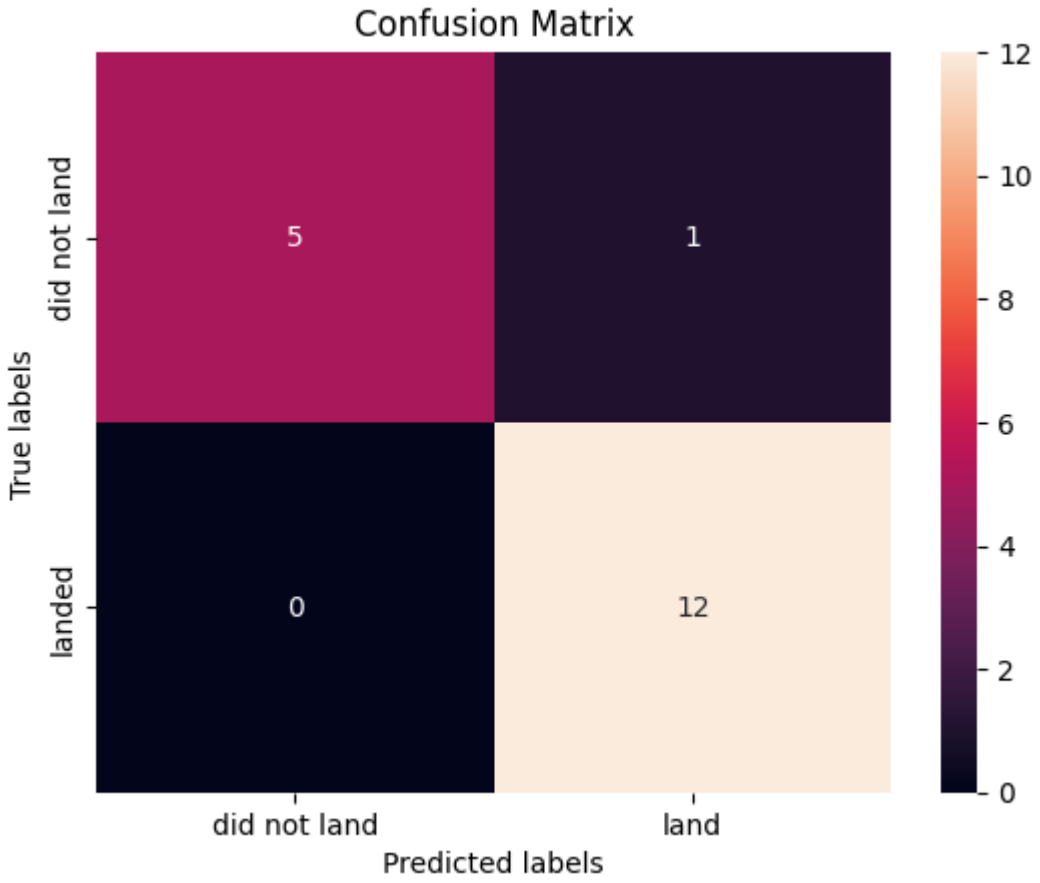


Logistic Regression

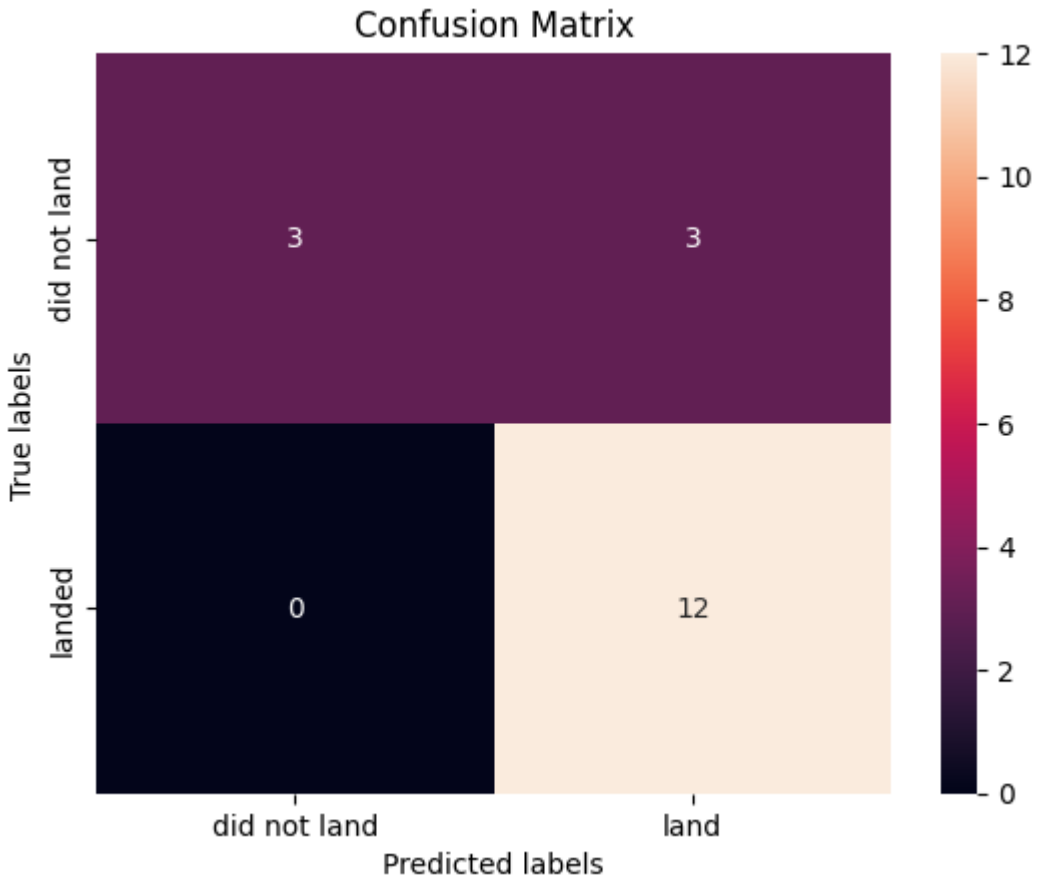


SVM

RESULTS Predictive analysis (classification) results



Decision Tree

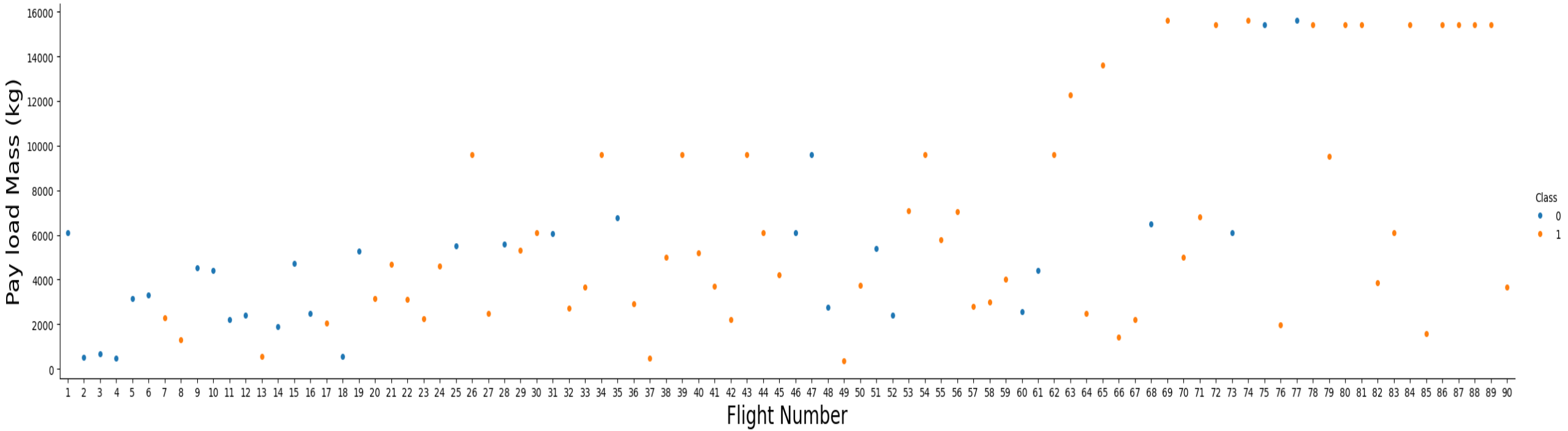


KNN

RESULTS Predictive analysis (classification) results

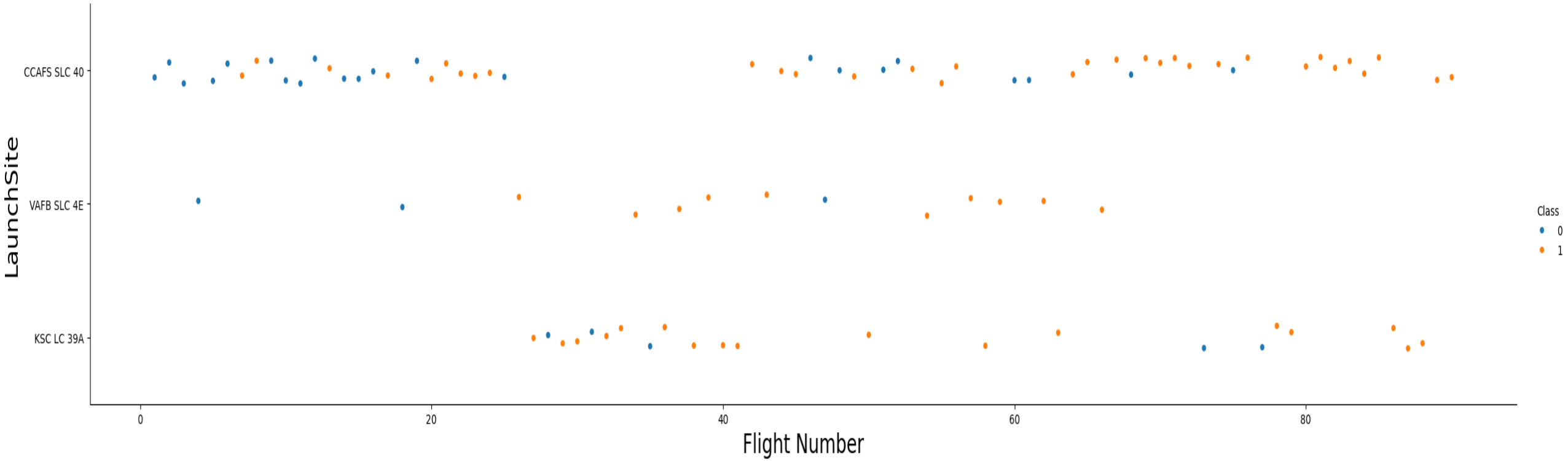
```
Accuracy on test data - Logistic Regression: 0.8333333333333334
Accuracy on test data - Support Vector Machine: 0.8333333333333334
Accuracy on test data - Decision Tree: 0.9444444444444444
Accuracy on test data - K-Nearest Neighbors: 0.8333333333333334
```

RESULTS EDA and interactive visual analytics



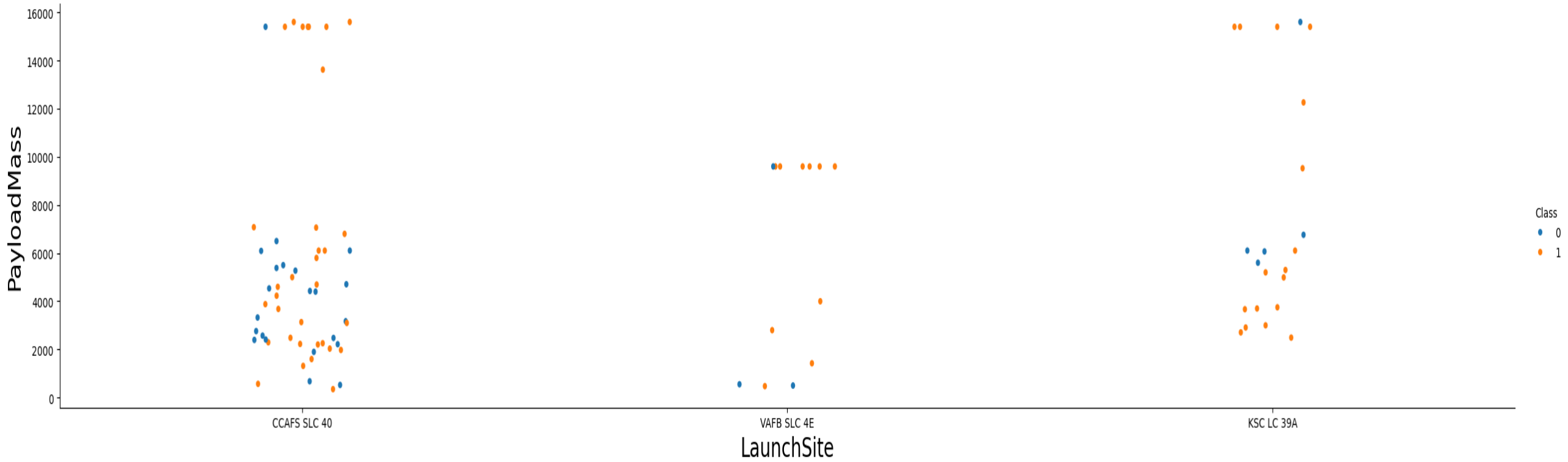
First, let's try to see how the Flight Number (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome.

RESULTS EDA and interactive visual analytics



Next, let's drill down to each site visualize its detailed launch records.

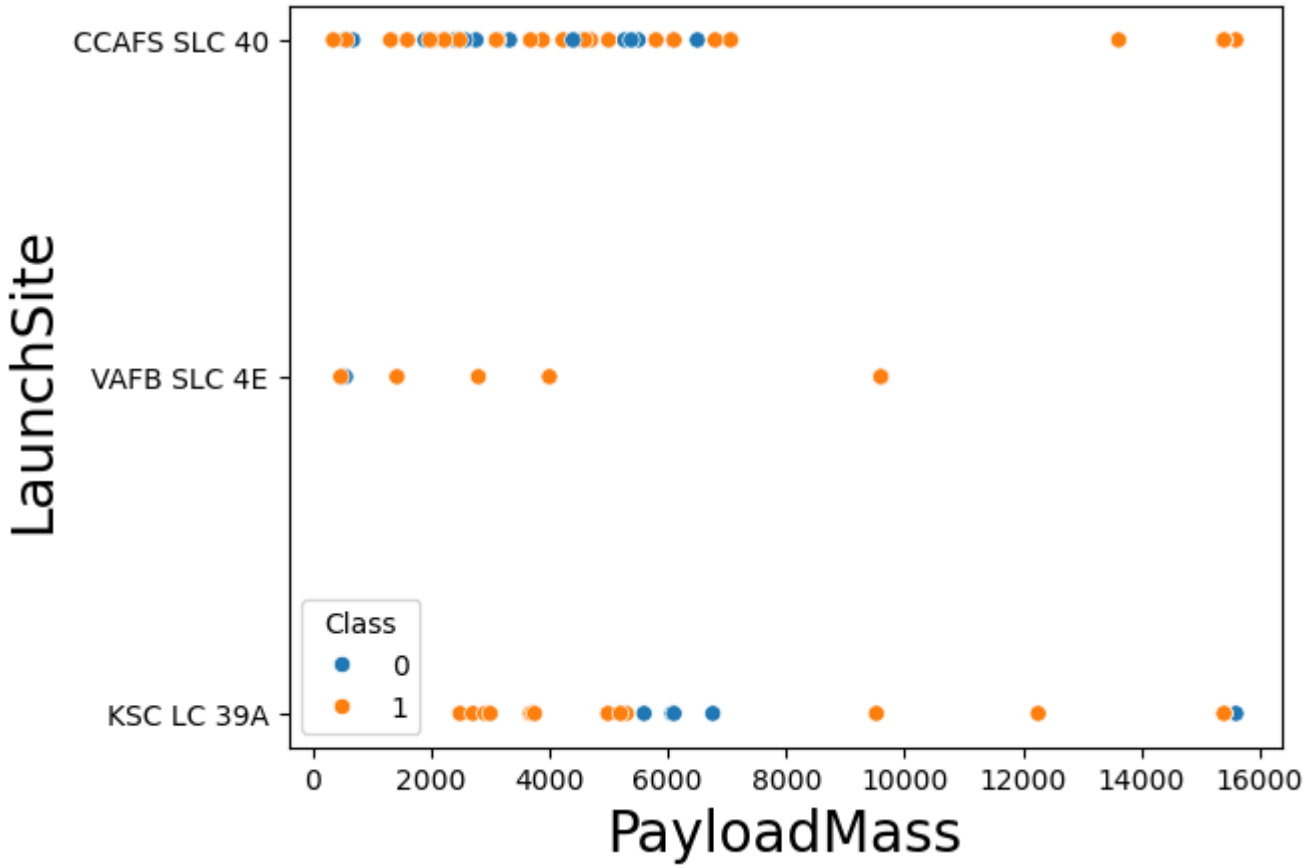
RESULTS EDA and interactive visual analytics



We also want to observe if there is any relationship between launch sites and their payload mass.

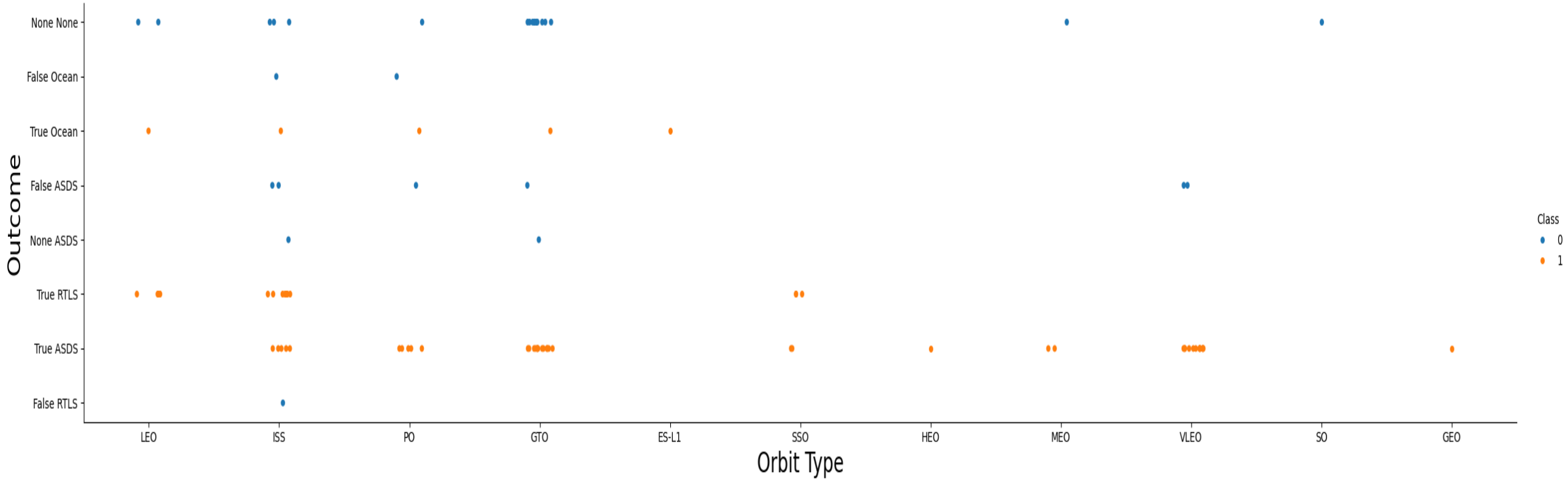
RESULTS

EDA and interactive visual analytics



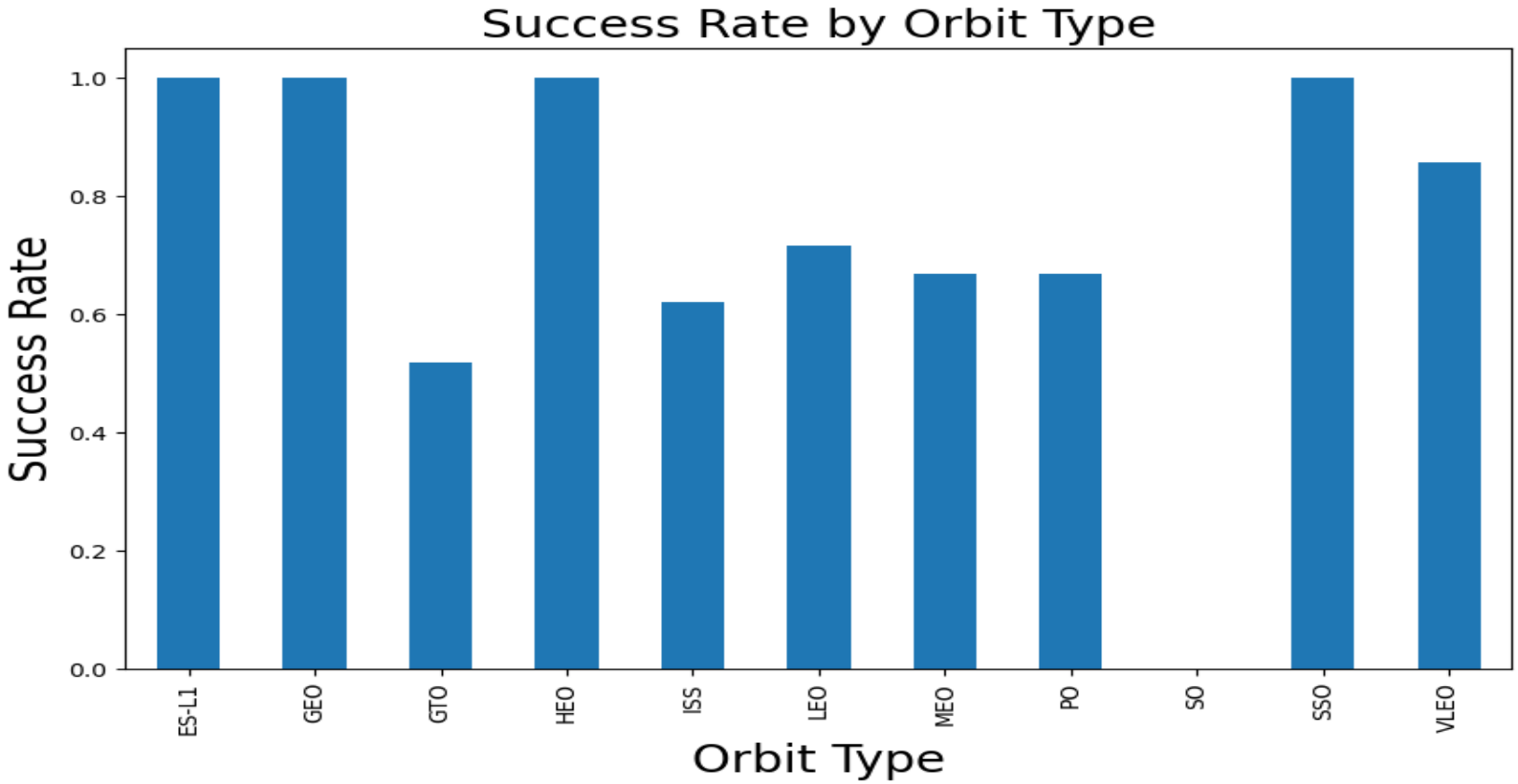
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

RESULTS EDA and interactive visual analytics



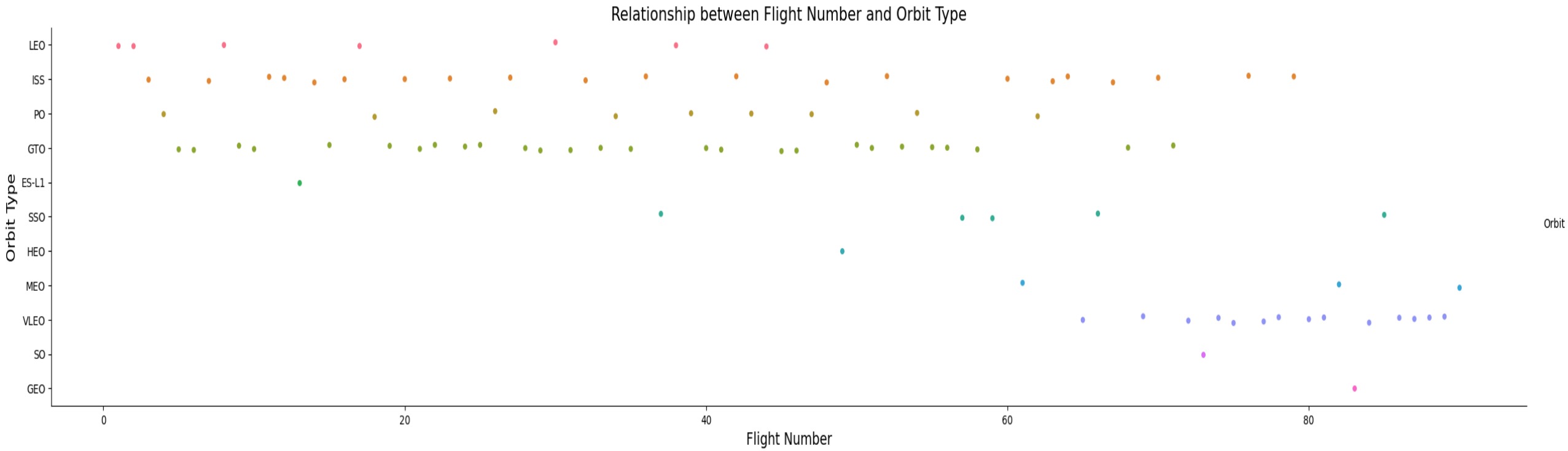
Next, we want to visually check if there are any relationship between success rate and orbit type.

RESULTS EDA and interactive visual analytics



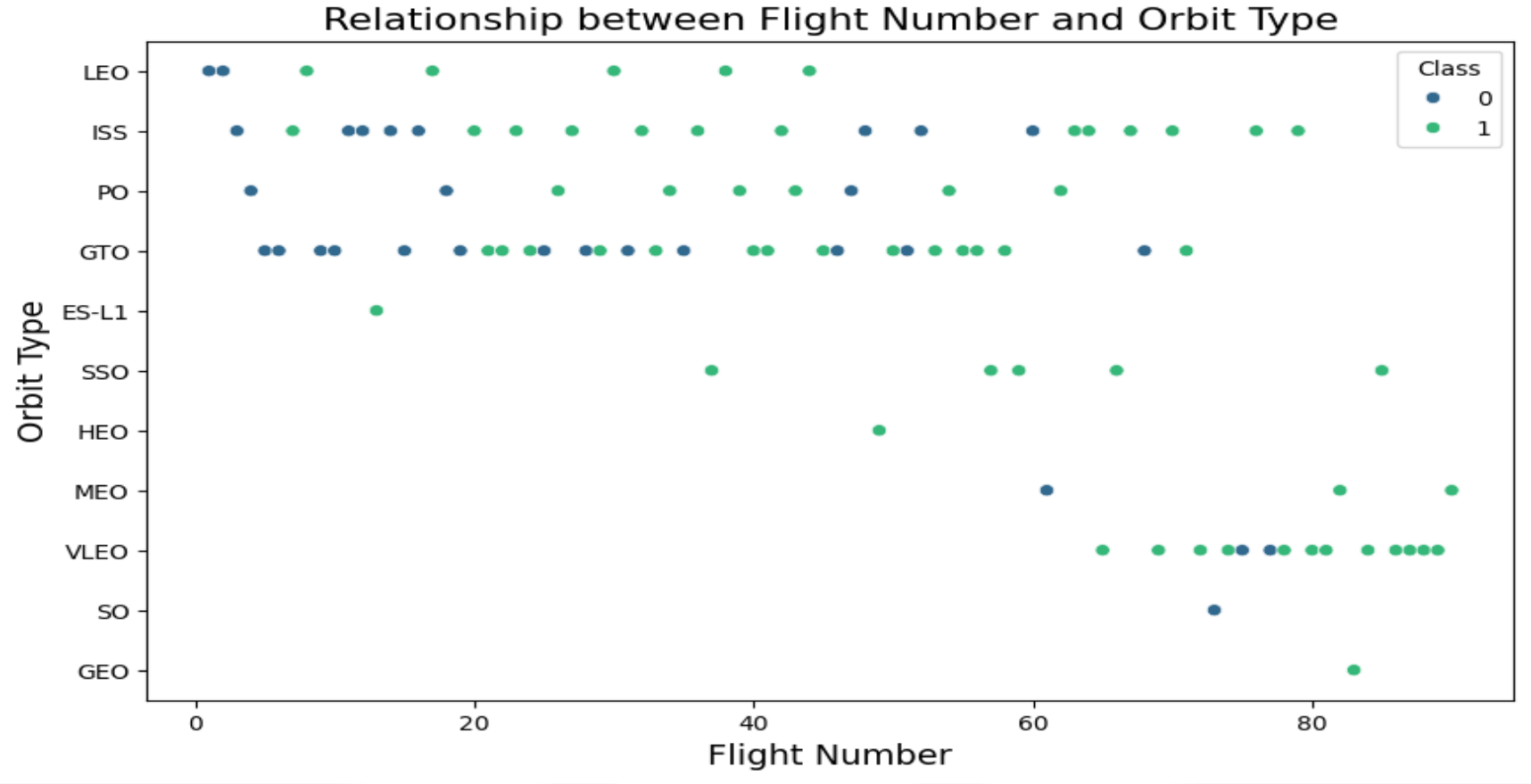
Analyze the plotted bar chart we can see which orbits have high success rate.

RESULTS EDA and interactive visual analytics



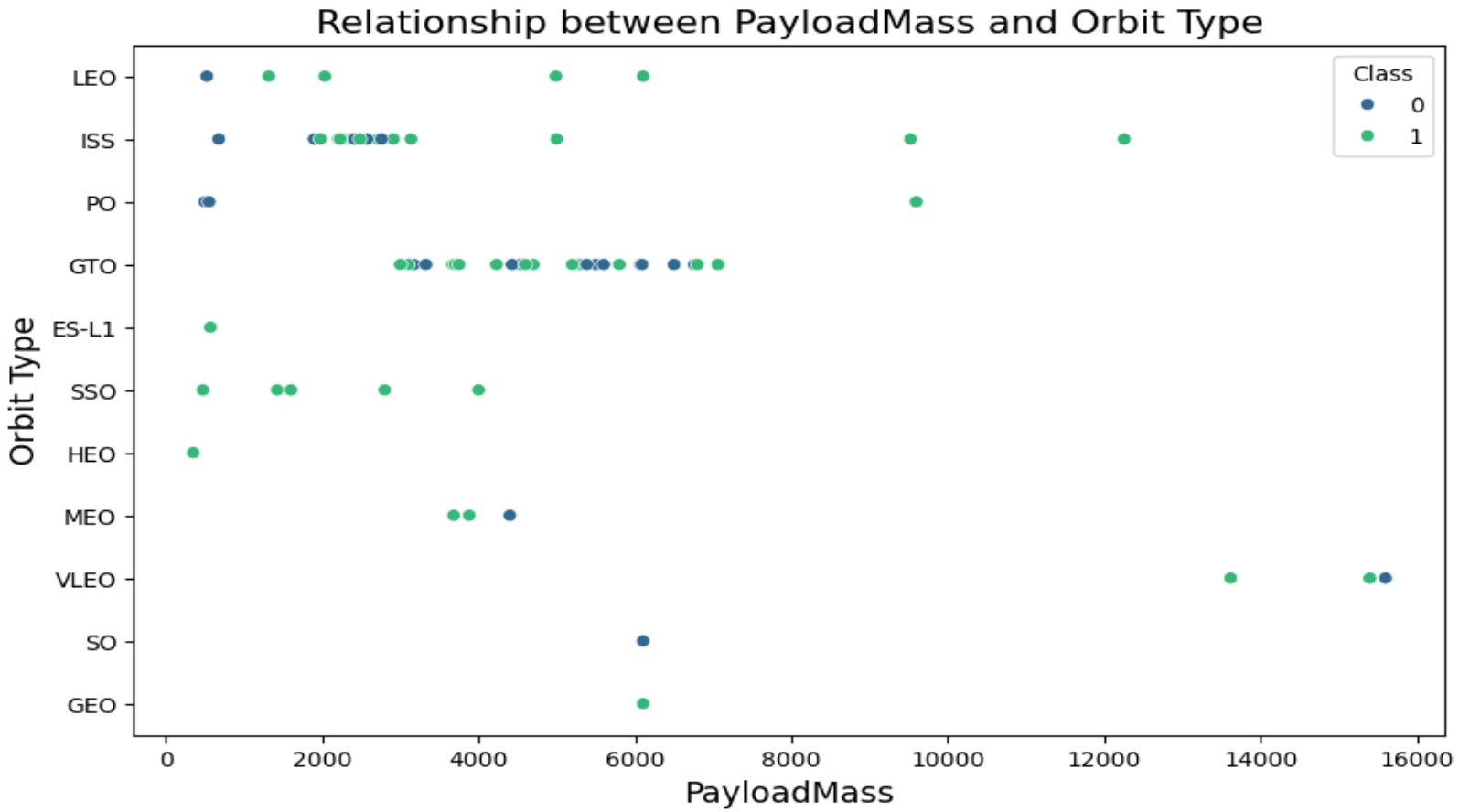
For each orbit, we want to see if there is any relationship between FlightNumber and Orbit type.

RESULTS EDA and interactive visual analytics



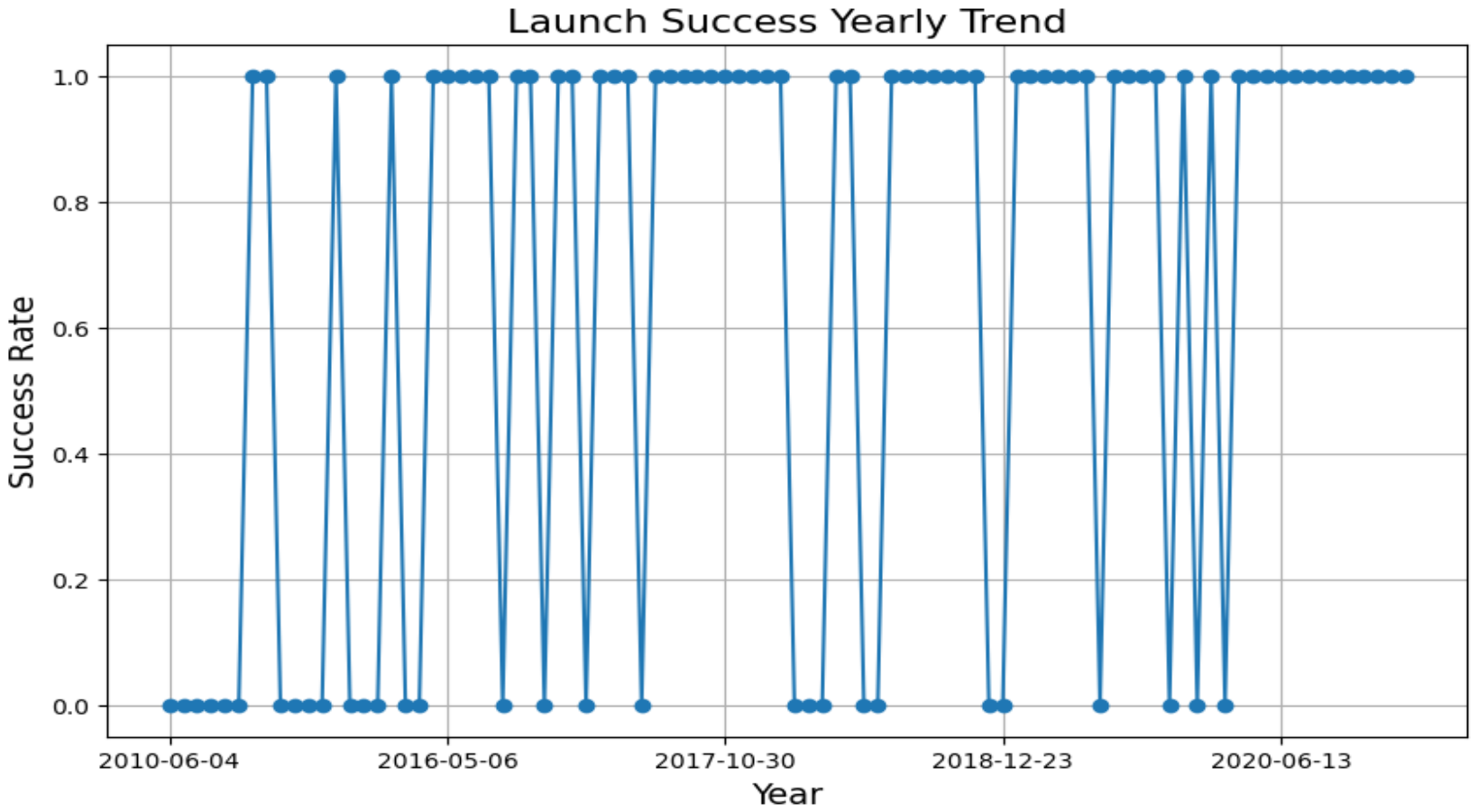
You see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

RESULTS EDA and interactive visual analytics



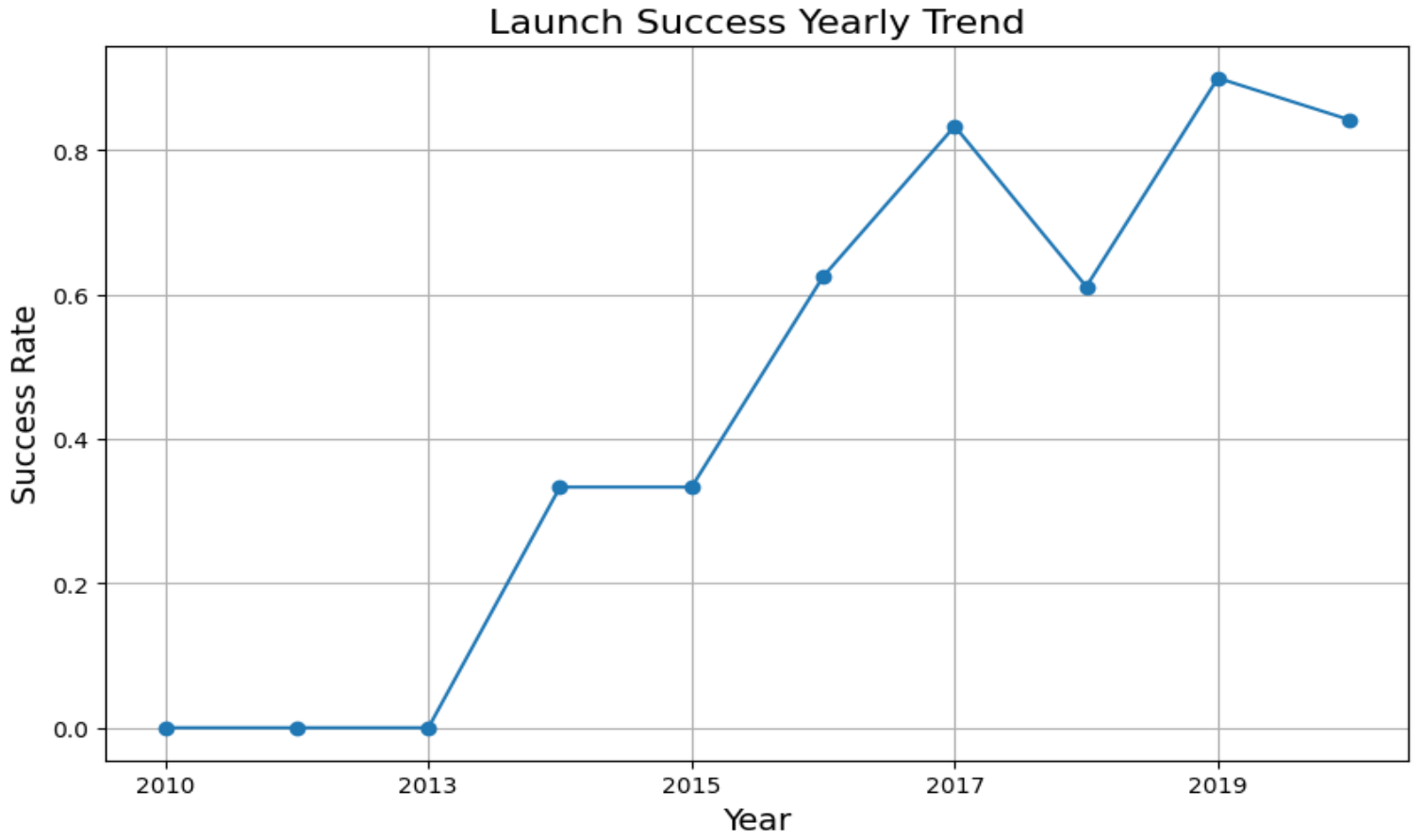
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

RESULTS EDA and interactive visual analytics



Here we Visualize the launch success yearly trend

RESULTS EDA and interactive visual analytics



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

RESULTS EDA with visualization results

By now, we have obtained some preliminary insights about how each important variable would affect the success rate, we then select the features that will be used in success prediction in the future module, Which are ‘FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial’.

Create dummy variables to categorical columns 'Orbit', 'LaunchSite', 'LandingPad', 'Serial’,

Then we cast all numeric columns to `float64`

RESULTS

EDA with visualization results

Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	Serial_B1048	Serial_B1049	Serial_B1050	Serial_B1051	Serial_B1054	Serial_[illegible]
1.0	False	False	False	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
1.0	False	False	False	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
1.0	False	False	False	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
1.0	False	False	False	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
1.0	False	False	False	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	

RESULTS EDA with SQL results

In this section we:

1. Understand the SpaceX Dataset
2. Load the dataset into the corresponding table in a Db2 database
3. Execute SQL queries to answer assignment questions

RESULTS EDA with SQL results

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

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

RESULTS EDA with SQL results

```
('6/4/2010', '18:45:00', 'F9 v1.0  B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Succes
s', 'Failure (parachute)')
('12/8/2010', '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)')
('22/05/2012', '7:44:00', 'F9 v1.0  B0005', 'CCAFS LC-40', 'Dragon demo flight C2', 525, 'LEO (ISS)', 'NASA (COTS)', 'Success',
'No attempt')
('10/8/2012', '0:35:00', 'F9 v1.0  B0006', 'CCAFS LC-40', 'SpaceX CRS-1', 500, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attemp
t')
('3/1/2013', '15:10:00', 'F9 v1.0  B0007', 'CCAFS LC-40', 'SpaceX CRS-2', 677, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attemp
t')
```

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

RESULTS EDA with SQL results

```
Total payload mass carried by NASA (CRS) boosters: 45596 kg
```

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

RESULTS EDA with SQL results

```
Average payload mass carried by F9 v1.1 boosters: 2928.4 kg
```

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

RESULTS EDA with SQL results

```
The date of the first successful landing on a ground pad was: 1/8/2018
```

Date when the first successful landing outcome in ground pad was achieved.

RESULTS EDA with SQL results

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

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

RESULTS EDA with SQL results

```
Total successful missions: 98
Total failed missions: 0
```

List the total number of successful and failure mission outcomes

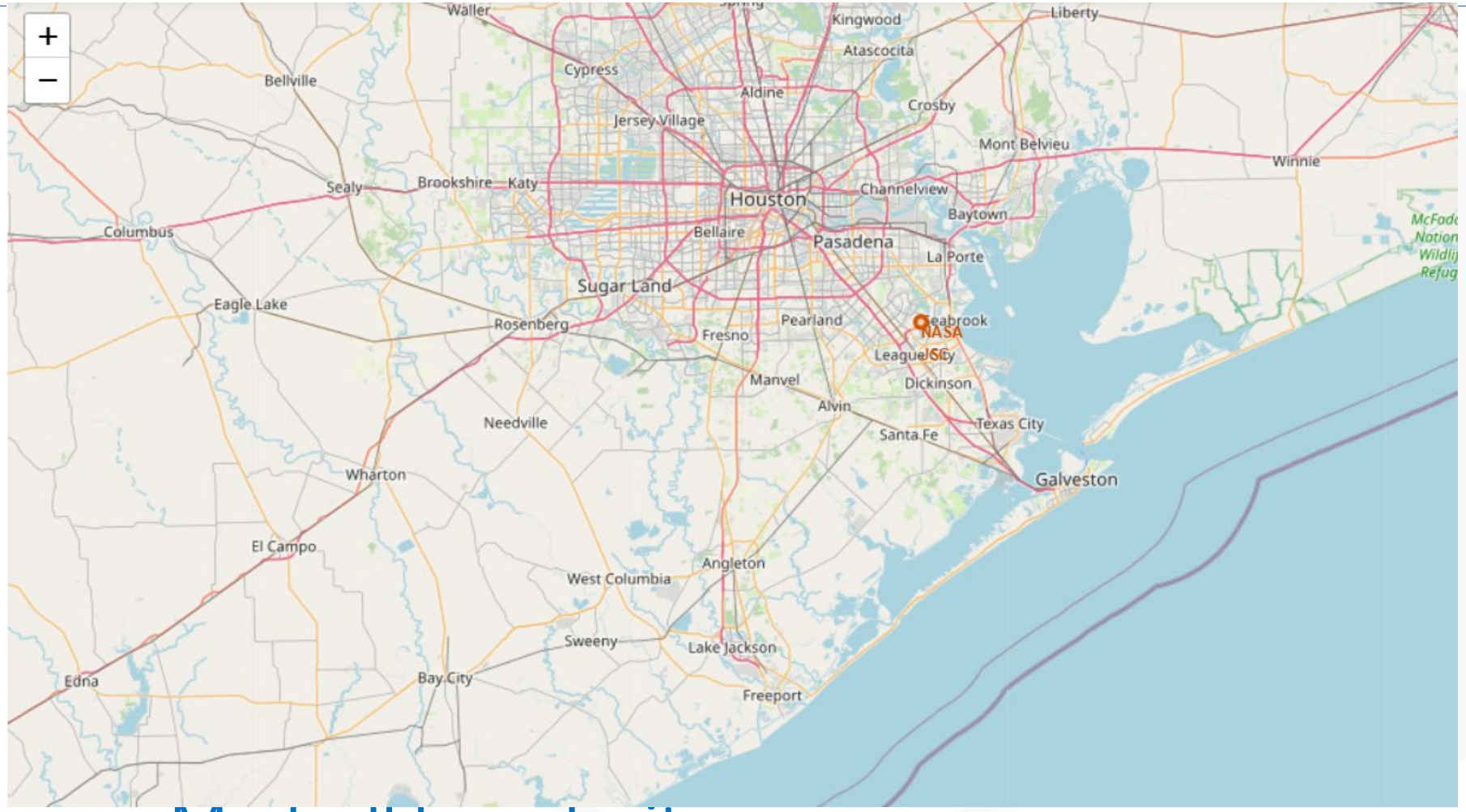
RESULTS EDA with SQL results

```
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
```

Names of the booster versions which have carried the maximum payload mass.

RESULTS

Interactive map with Folium Map



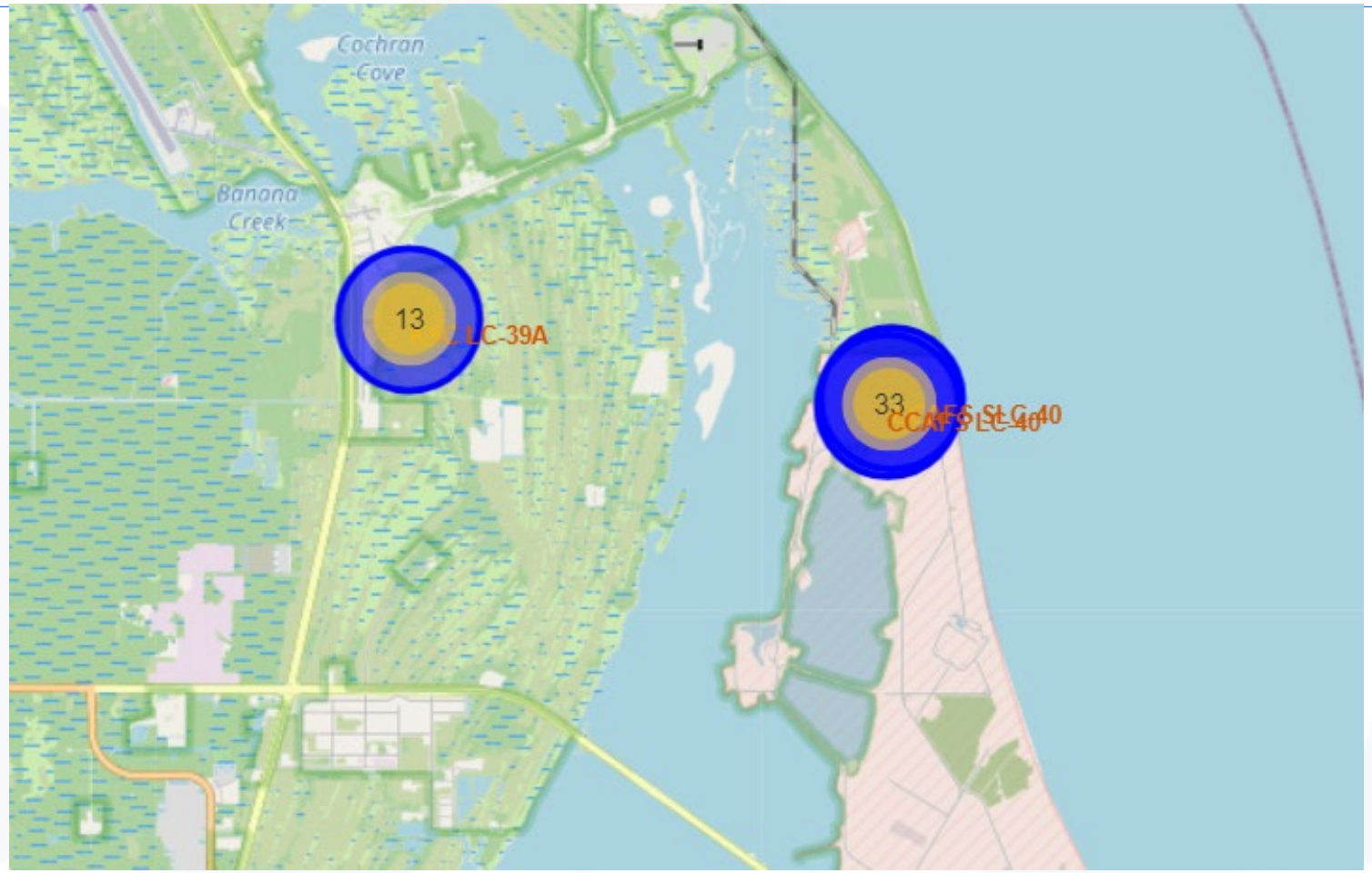
Mark all launch sites on a map

RESULTS Interactive map with Folium Map

Mark all launch sites on a map

RESULTS

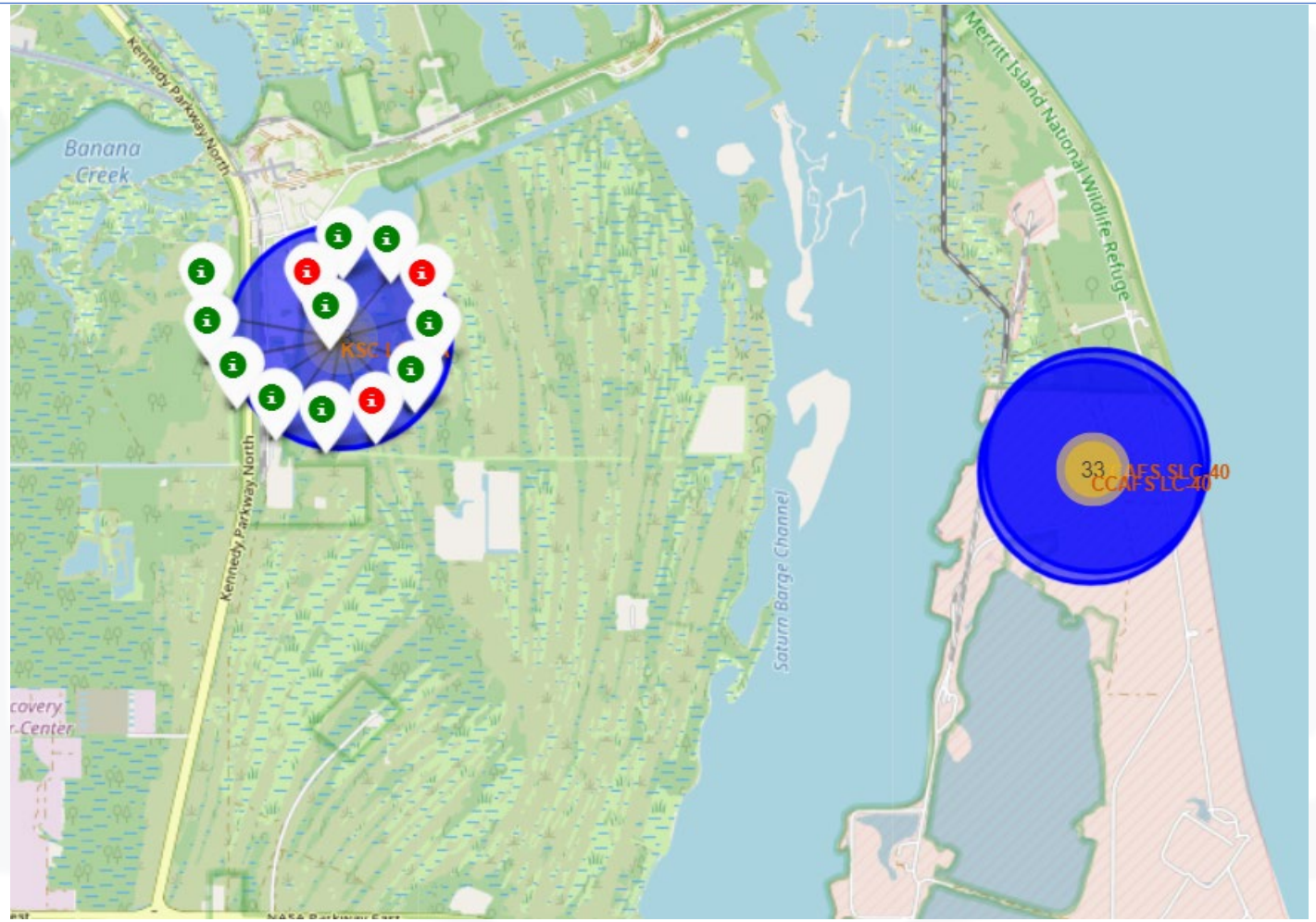
Interactive map with Folium Map



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

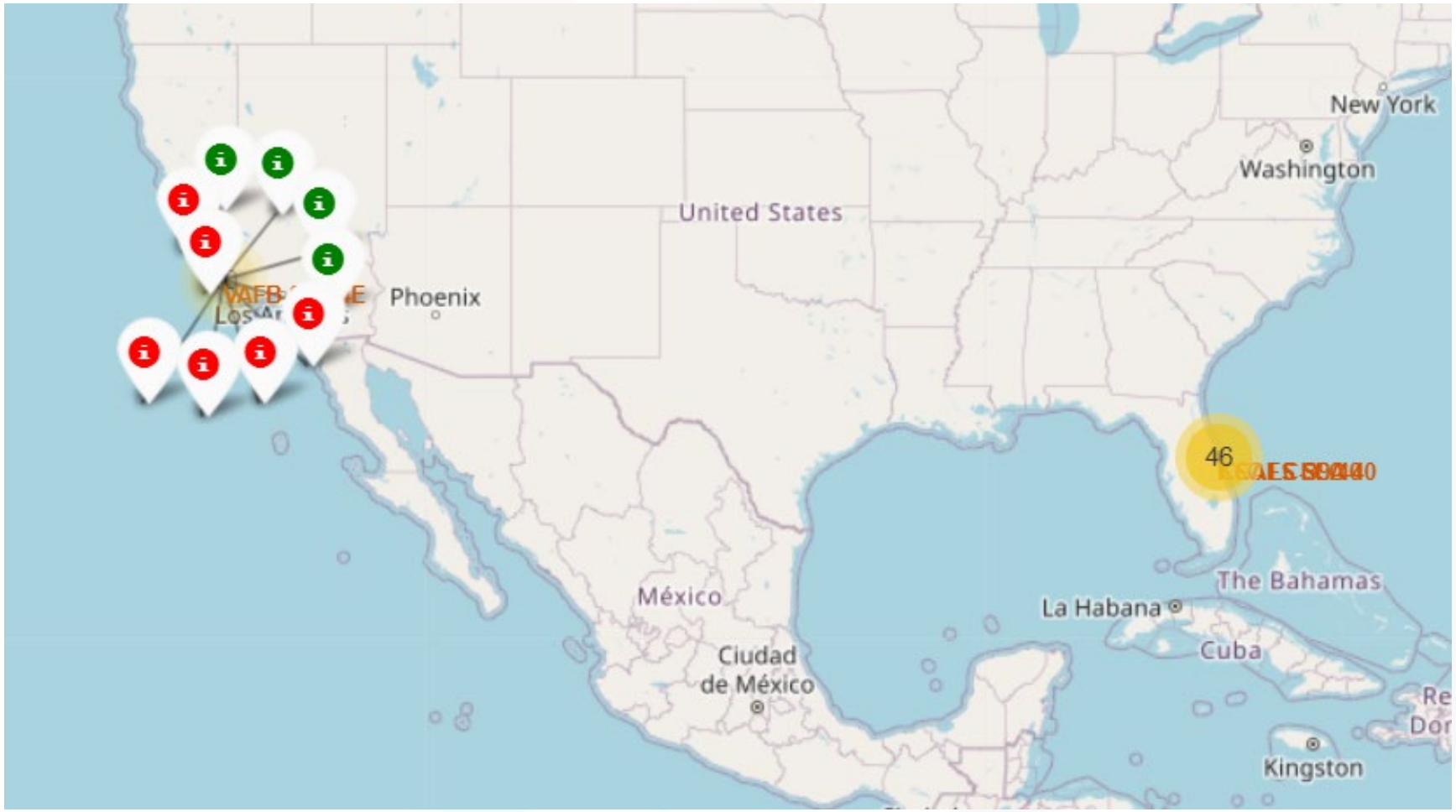
RESULTS

Interactive map with Folium Map



Mark the success(green)/failed(red) launches for each site on the map

RESULTS Interactive map with Folium Map

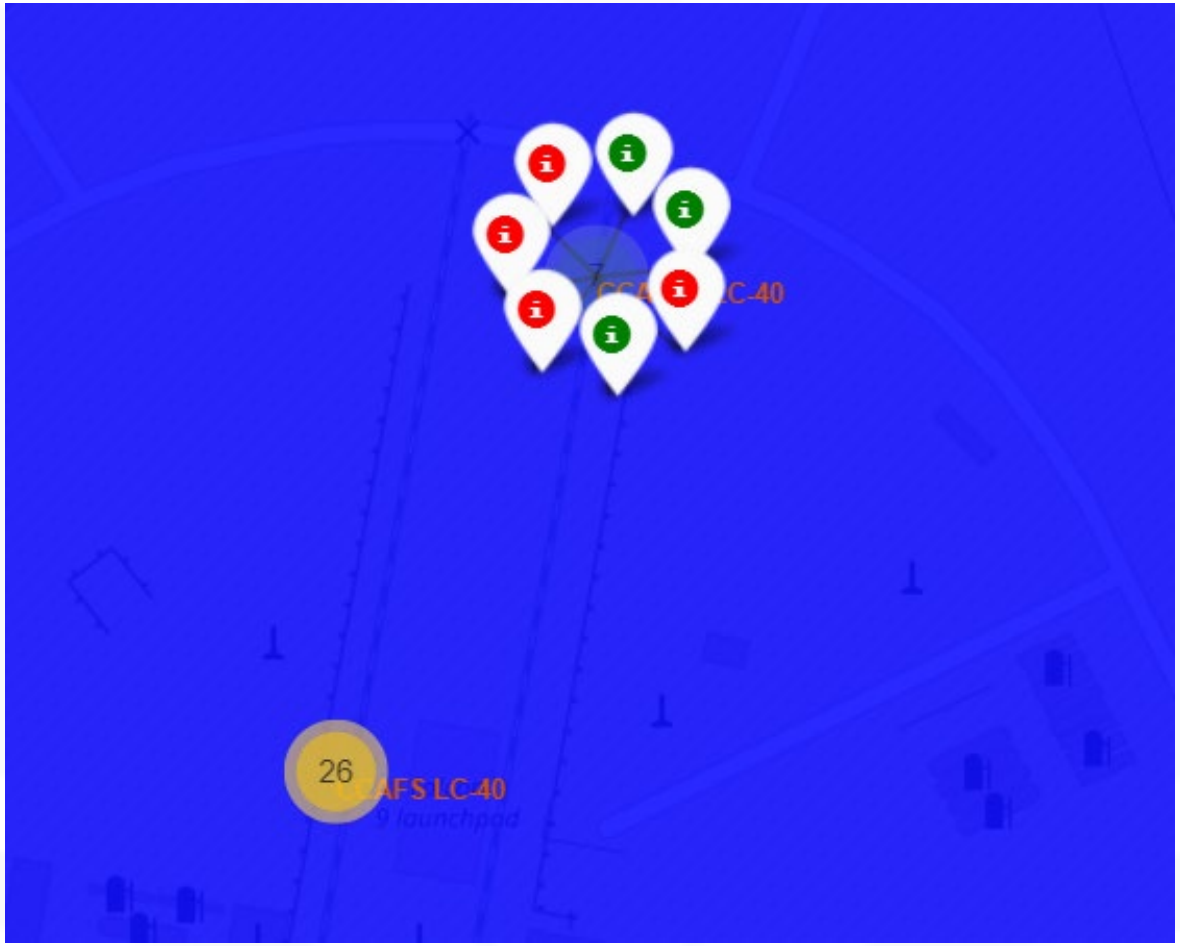


Mark the success(green)/failed(red) launches for each site on the map

RESULTS Interactive map with Folium Map

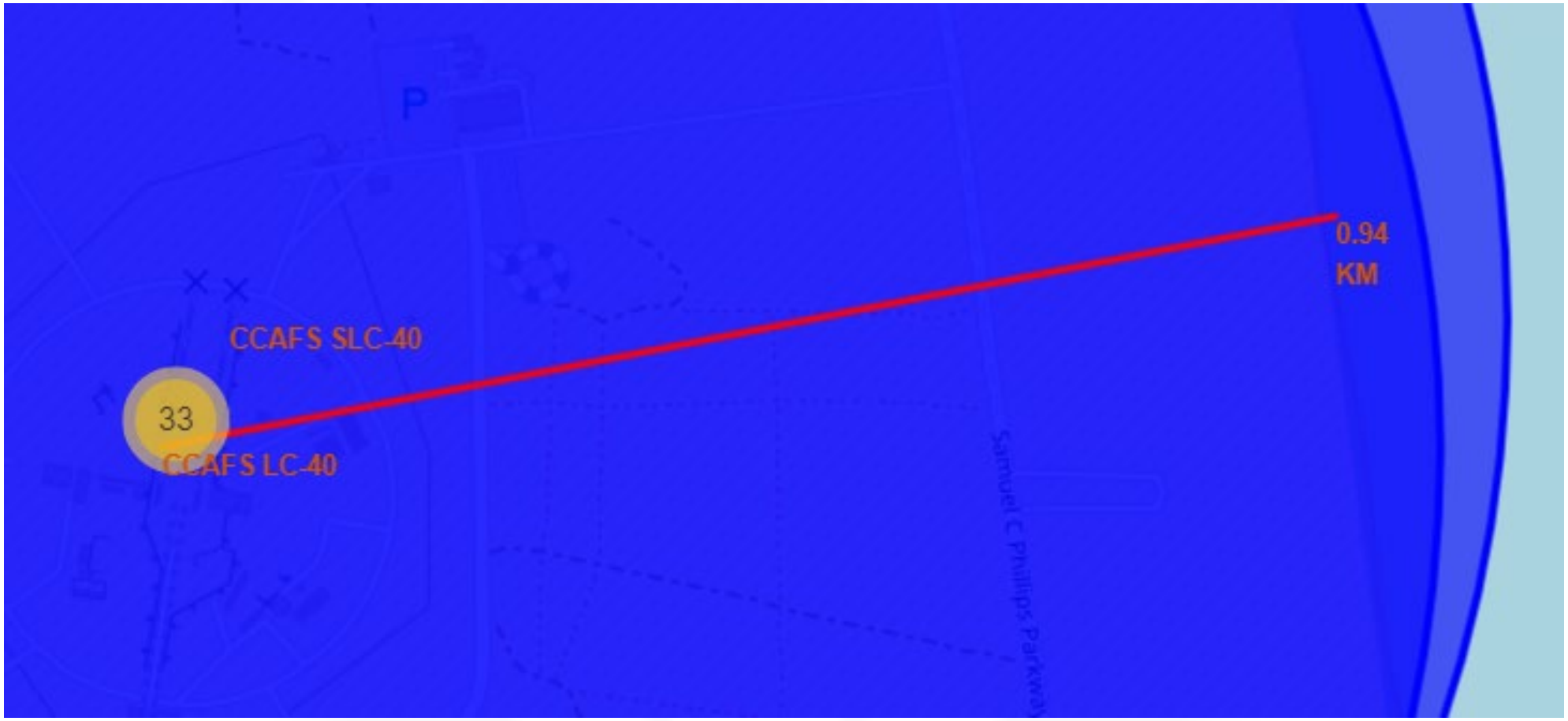
Mark the success(green)/failed(red) launches for each site on the map

RESULTS Interactive map with Folium Map



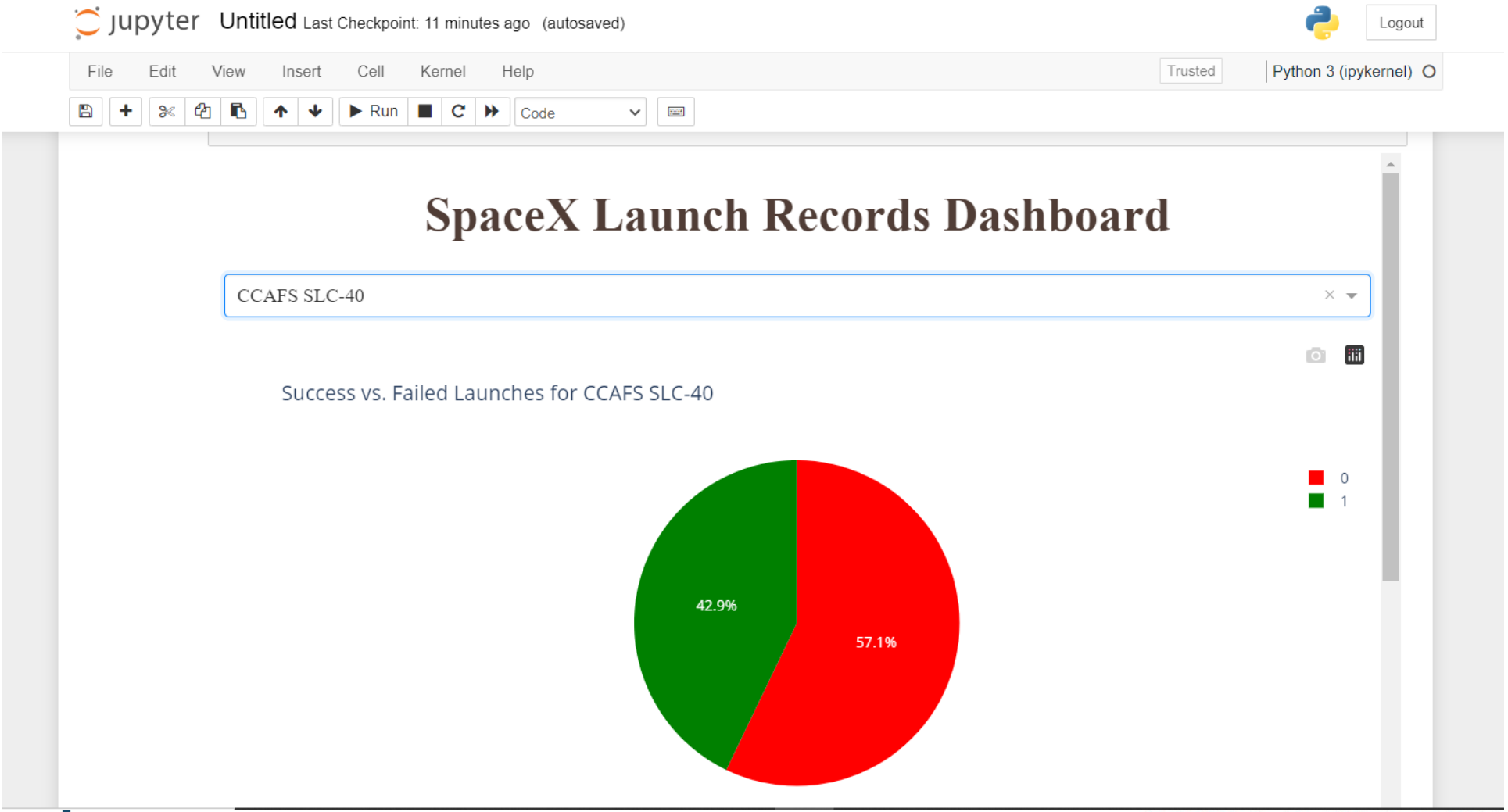
Mark the success(green)/failed(red) launches for each site on the map

RESULTS Interactive map with Folium Map

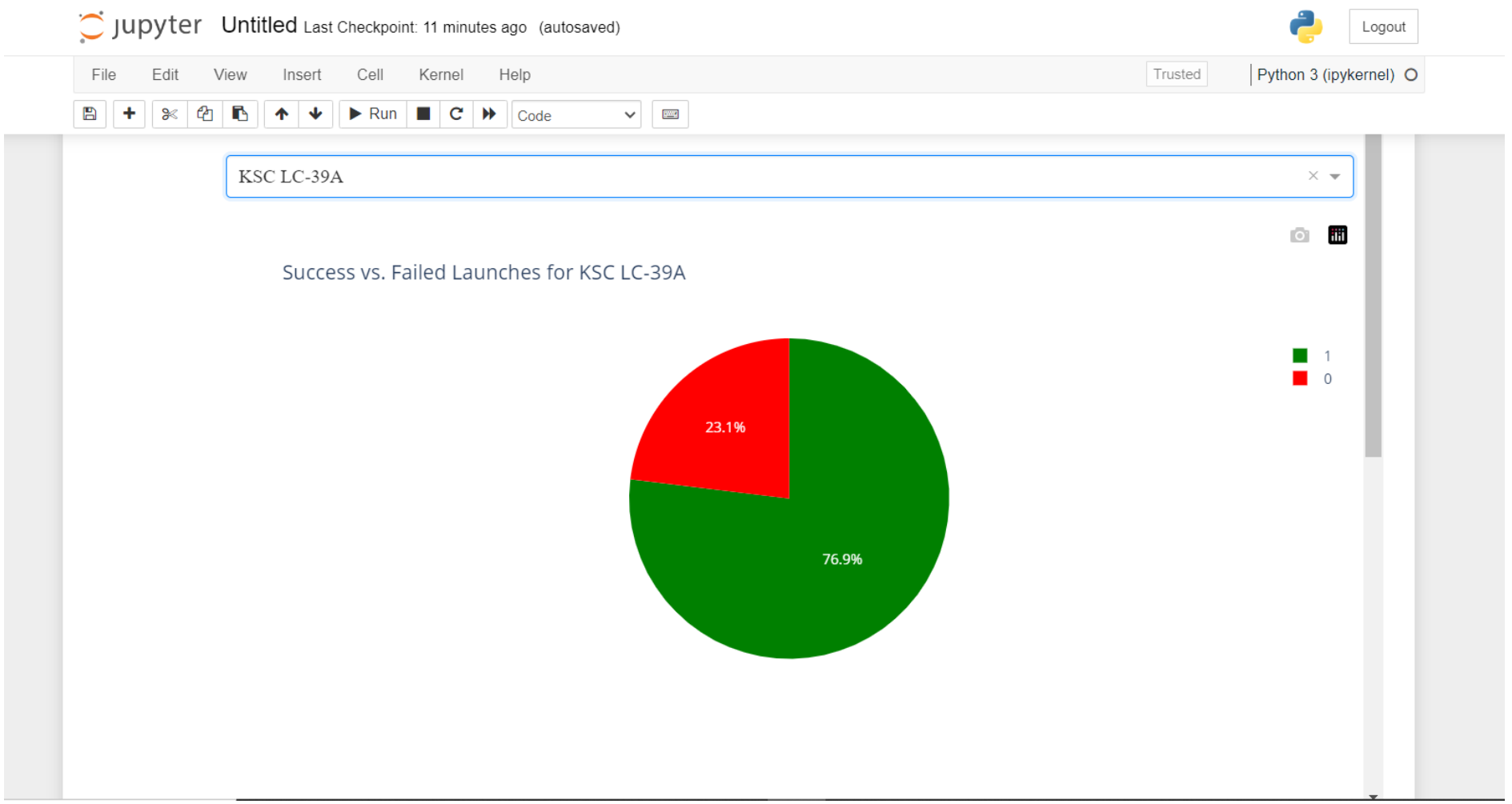


Calculate the distances between a launch site to its proximities

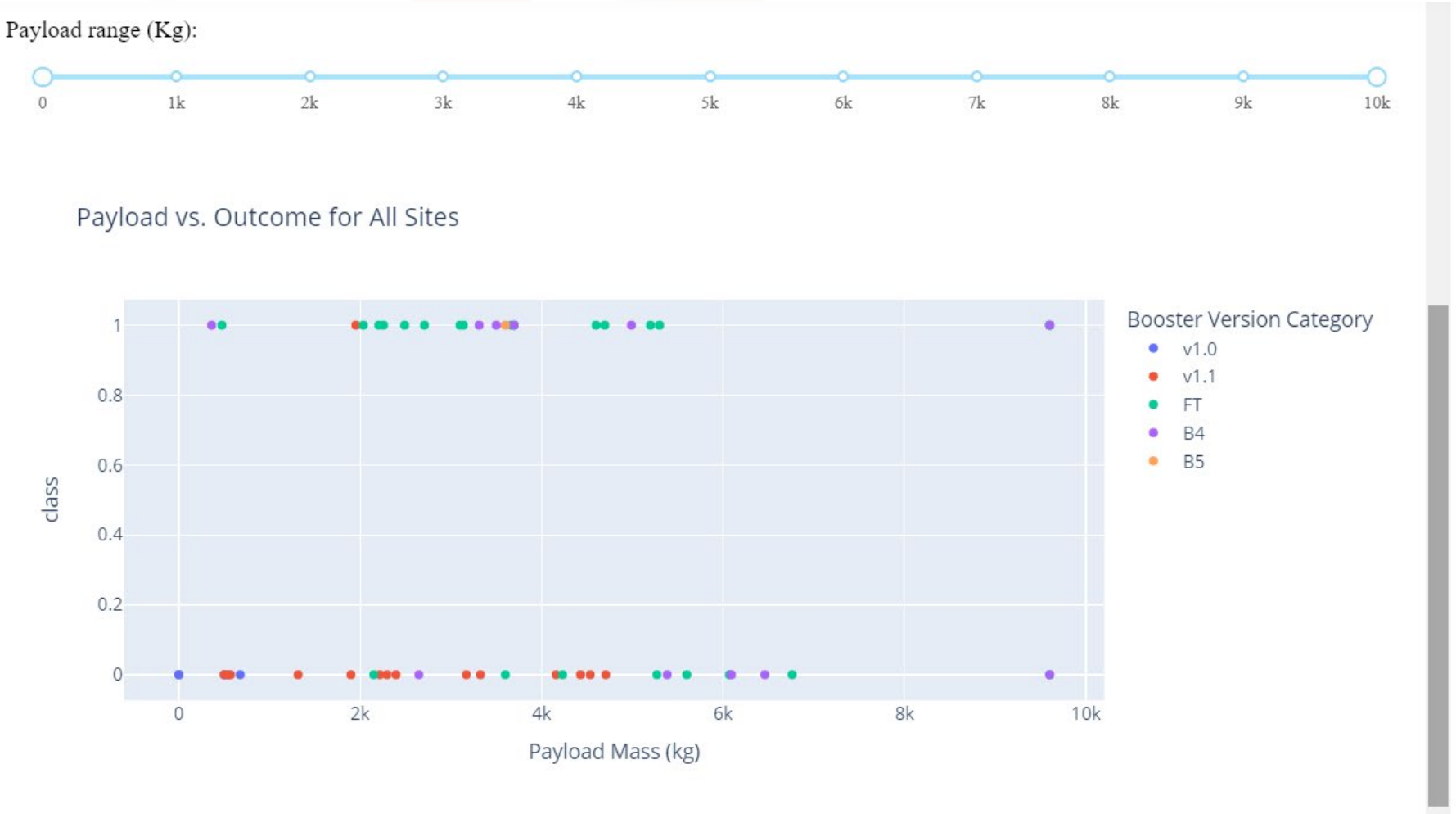
DASHBOARD TAB 1



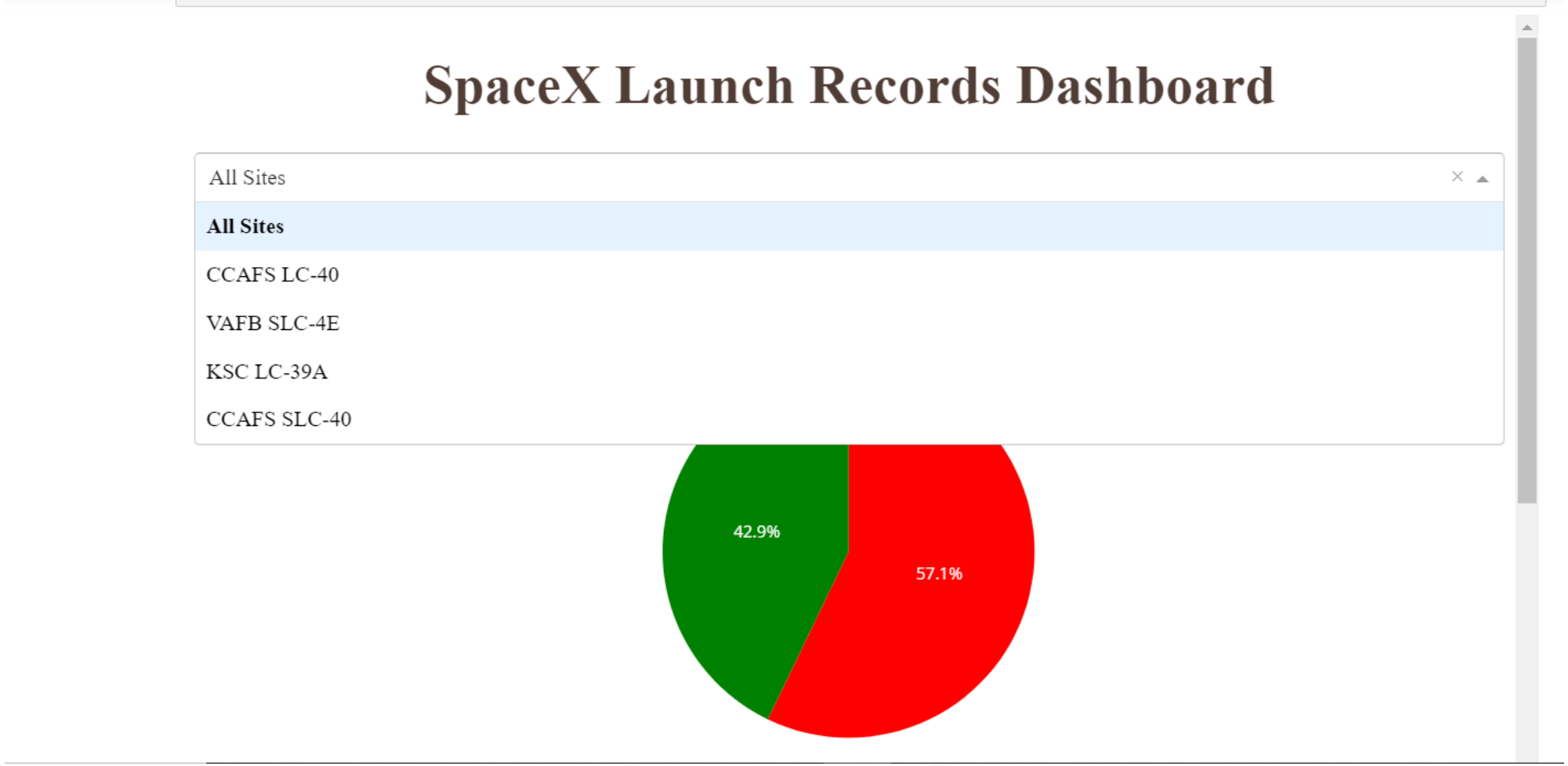
DASHBOARD TAB 2



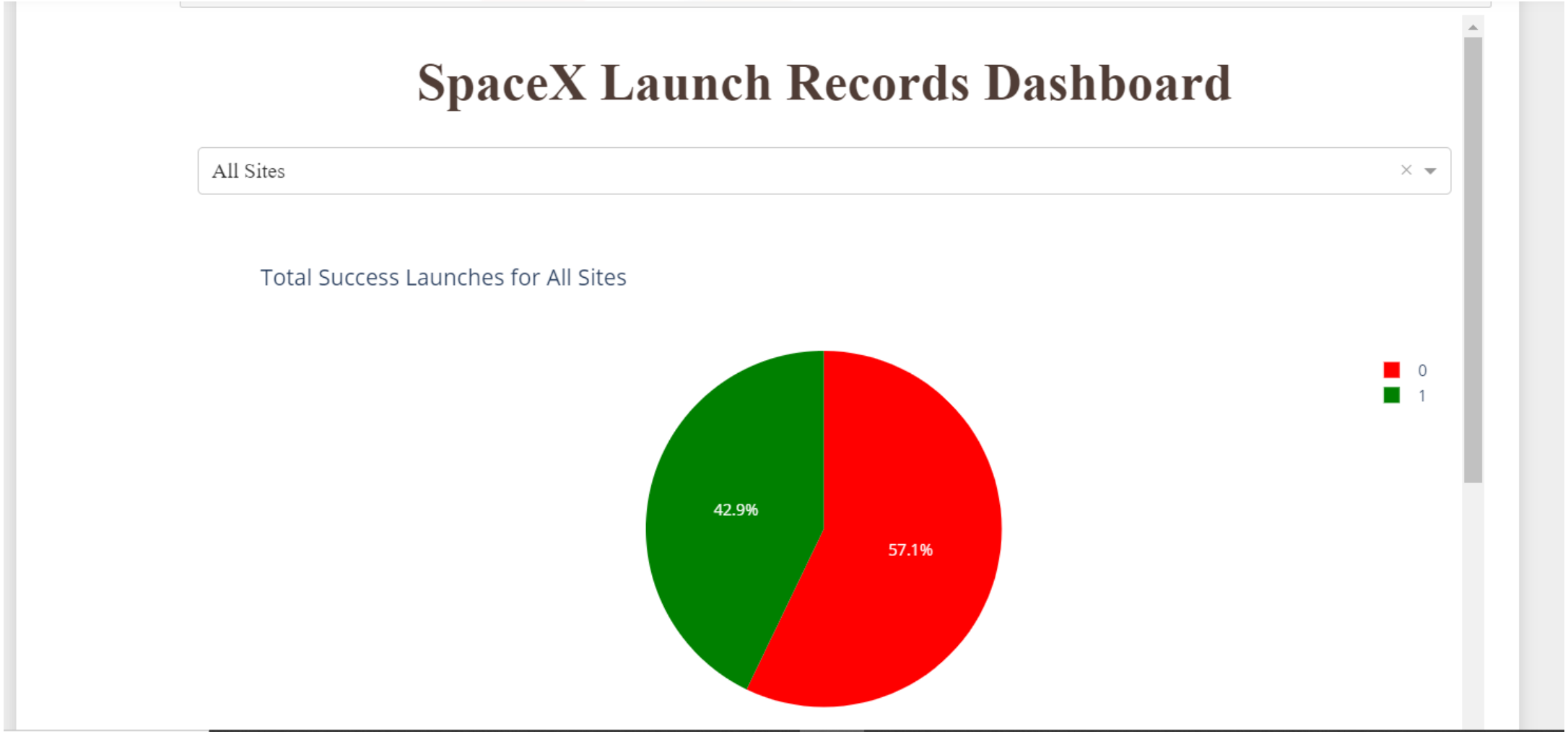
DASHBOARD TAB 3



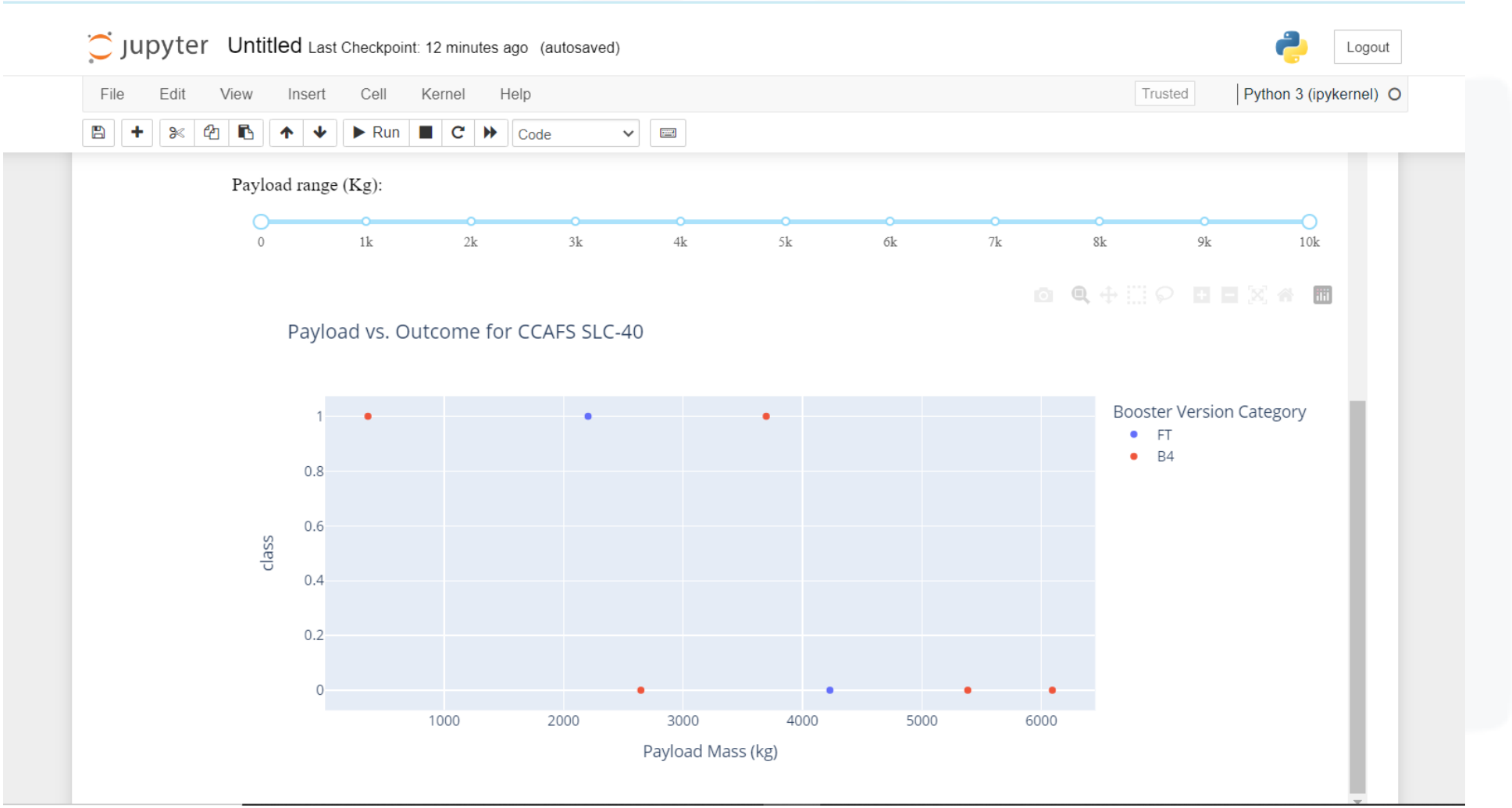
DASHBOARD TAB 4



DASHBOARD TAB 5



DASHBOARD TAB 6



DISCUSSION

- We were able to predict if the Falcon 9 first stage will land successfully.
- We collected data on the Falcon 9 first-stage landings.
- We used a RESTful API and web scraping. And converted the data into a data frame and then perform some data wrangling.
- We build a dashboard to analyze launch records interactively with Plotly Dash.
- And then we build an interactive map to analyze the launch site proximity with Folium.
- We also used machine learning to determine if the first stage of Falcon 9 will land successfully. We split the data into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression.
- Then found the method that performs best using test data.

CONCLUSION

- We were able to predict if the Falcon 9 first stage will land successfully.
- We collected data on the Falcon 9 first-stage landings.
- We used a RESTful API and web scraping. And converted the data into a data frame and then perform some data wrangling.
- We build a dashboard to analyze launch records interactively with Plotly Dash.
- And then we build an interactive map to analyze the launch site proximity with Folium.
- We also used machine learning to determine if the first stage of Falcon 9 will land successfully. We split the data into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression.
- Then found the method that performs best using test data.

CONCLUSION

- We were able to predict if the Falcon 9 first stage will land successfully.
- We collected data on the Falcon 9 first-stage landings.
- We used a RESTful API and web scraping. And converted the data into a data frame and then perform some data wrangling.
- We build a dashboard to analyze launch records interactively with Plotly Dash.
- And then we build an interactive map to analyze the launch site proximity with Folium.
- We also used machine learning to determine if the first stage of Falcon 9 will land successfully. We split the data into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression.
- Then found the method that performs best using test data.