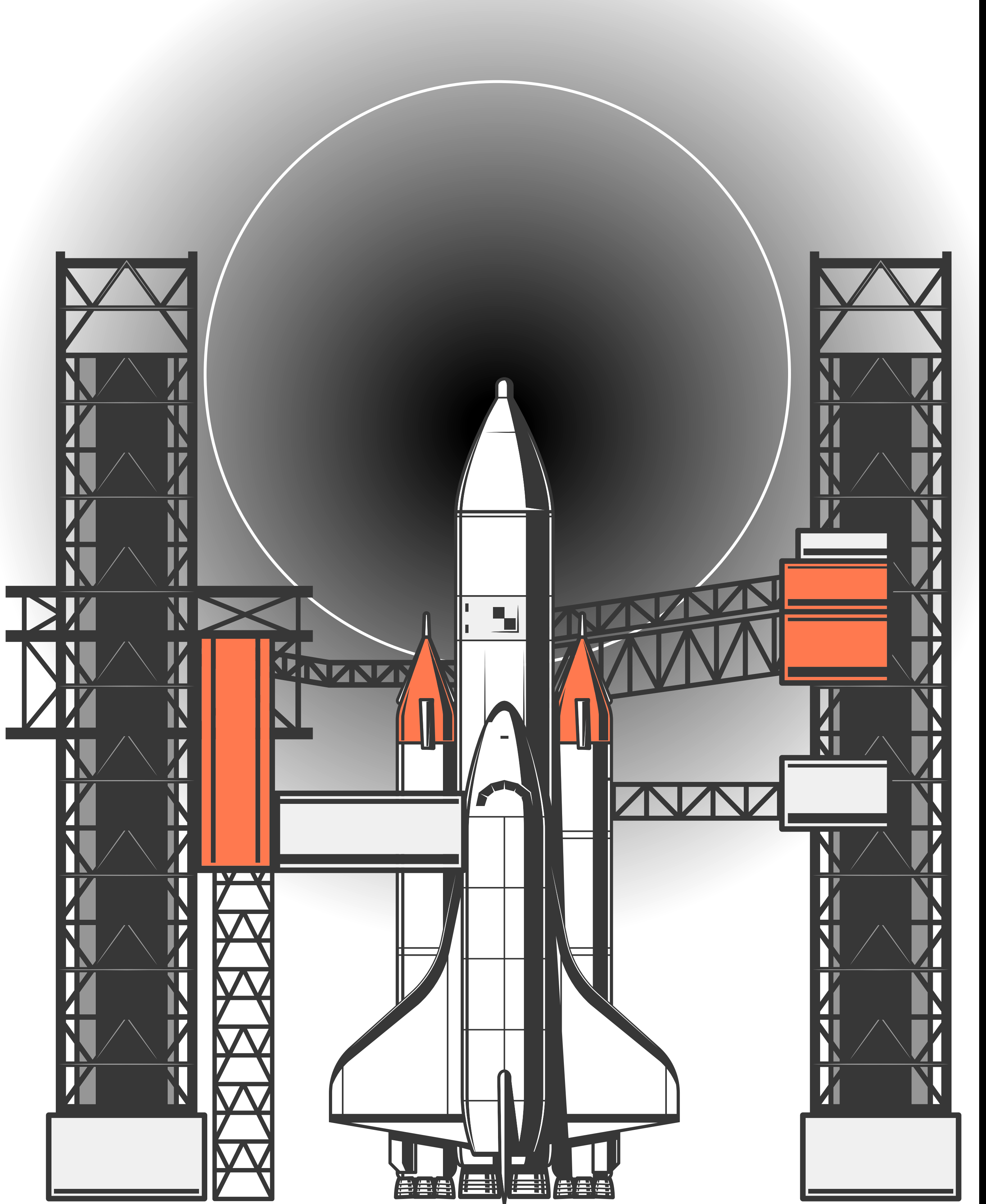
REPORT

“spaceX rocket launch”

GILANG BANYU BIRU

26/06/2023

“content”



Executive Summary

01 SUMMARY OF METHODOLOGIES

- DATA COLLECTION VIA API, WEB SCRAPING
- EXPLORATORY DATA ANALYSIS (EDA) WITH DATA VISUALIZATION
- EDA WITH SQL
- INTERACTIVE MAP WITH FOLIUM
- DASHBOARDS WITH PLOTLY DASH
- PREDICTIVE ANALYSIS

02 SUMMARY OF RESULTS

- EXPLORATORY DATA ANALYSIS RESULTS
- INTERACTIVE MAPS AND DASHBOARD
- PREDICTIVE RESULTS

“introduction”

SpaceX 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 SpaceX 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 for SpaceY company to bid against SpaceX for a rocket launch.

The aim of this project is to predict if the Falcon 9 first stage will successfully land.

PROBLEMS TO BE ANSWERED

- What variables affect the success of the first stage landing?
- What are the conditions which will allow SpaceX to achieve the best landing success rate ?
- What is the best algorithm that can be used for predicting the successfull of a rocket launch?

Methodology

Summary

Data collection methodology :
- SpaceX REST API
- Web Scrapping from Wikipedia

Data wrangling :
- Dropping unnecessary variables
- One Hot Encoding for classification models

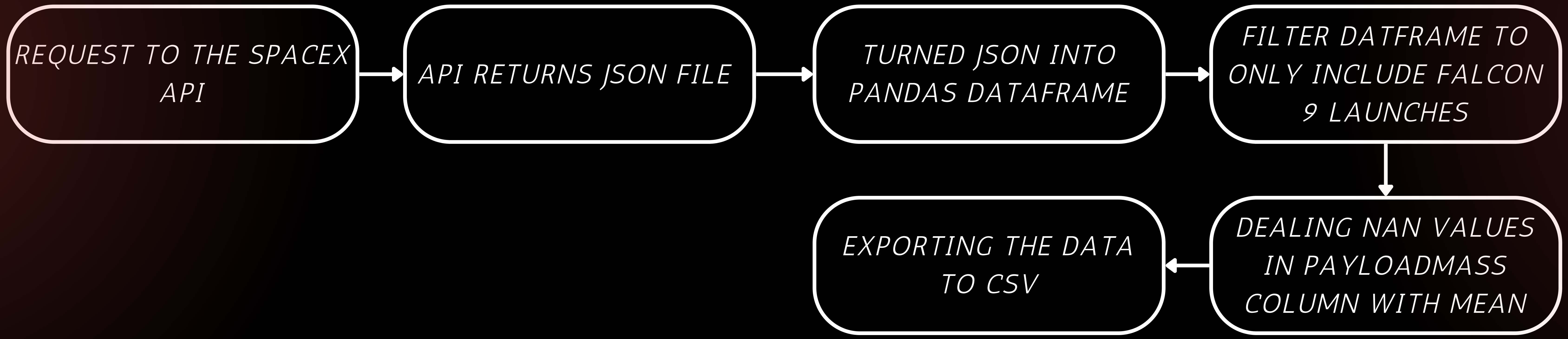
Perform exploratory data analysis (EDA) using visualization and SQL

Build an Interactive Map with Folium

Build a Dashboard with Plotly Dash

“data collection SpaceX API”

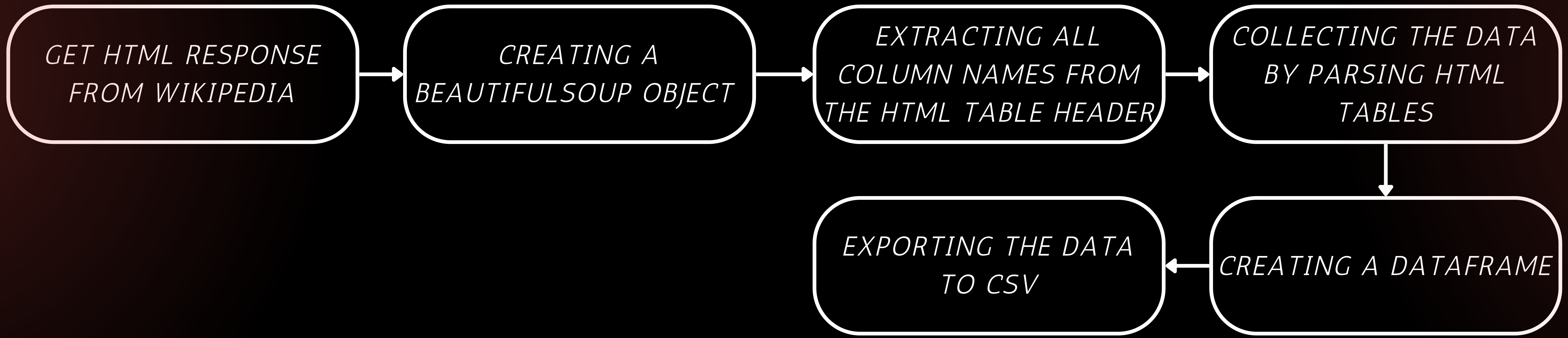
The information obtained by the API are rocket, launches, payload information.



The Space X REST API URL is api.spacexdata.com/v4/

data collection
Web Scraping

The information obtained by the webscrapping of Wikipedia are launches, landing, payload information.



URL is https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

“Data Wrangling”

perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

In the data set, there are several different cases of the booster landing, those cases are:

- **True Ocean** means the mission outcome was successfully landed to a specific region of the ocean
- **False Ocean** means the mission outcome was unsuccessfully landed to a specific region of the ocean.
- **True RTLS** means the mission outcome was successfully landed to a ground pad
- **False RTLS** means the mission outcome was unsuccessfully landed to a ground pad.
- **True ASDS** means the mission outcome was successfully landed on a drone ship
- **False ASDS** means the mission outcome was unsuccessfully landed on a drone ship.

convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

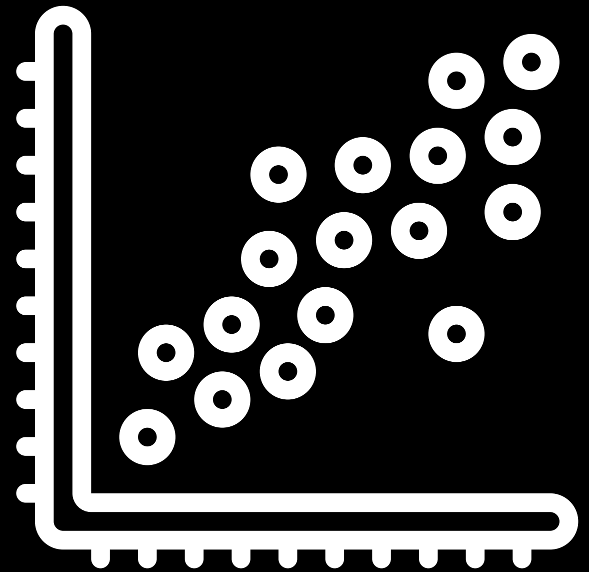
Data Wrangling

1. PERFORM EDA
2. CALCULATE THE NUMBER OF LAUNCHES ON EACH SITE
3. CALCULATE THE NUMBER AND OCCURRENCE OF EACH ORBIT
4. CALCULATE THE NUMBER AND OCCURRENCE OF MISSION OUTCOME PER ORBIT TYPE
5. CREATE A LANDING OUTCOME LABEL FROM OUTCOME COLUMN
6. EXPORTING THE DATA TO CSV

“EDA with data visualization”

Scatter Charts

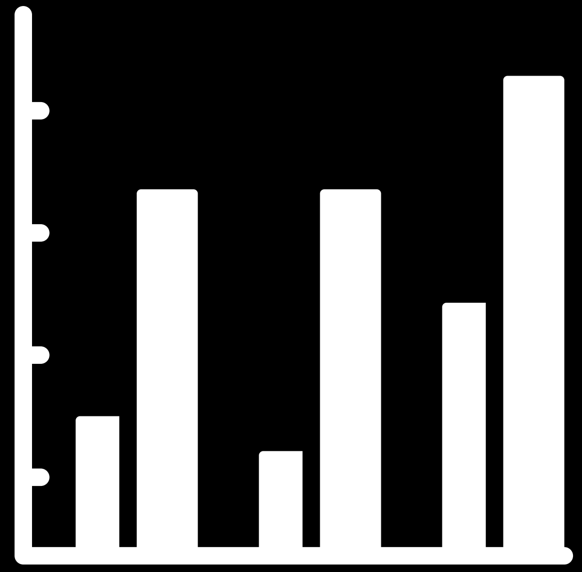
- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Orbit vs. Flight Number
- Payload vs. Orbit Type
- Orbit vs. Payload Mass



Scatter plots is best used for show relationship between variables. This relationship determined the correlation between variables

Bar Charts

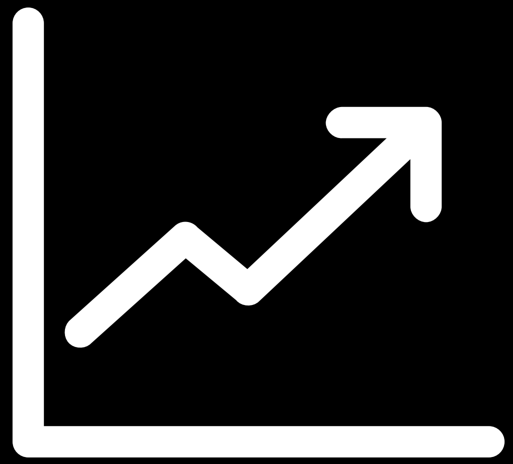
- Success rate vs. Orbit



Bar graphs is best used for show the relationship between numeric and categoric variables.

Line Charts

- Success rate vs. Year



Line charts show trends in data over time (time series).

EDA with SQL

Performed SQL queries to gather and understand data from dataset:

- Displaying the names of the unique lauunch sites in the space mission.
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1.
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass.
- List the records which will display the month names, faiilure landing_ouutcomes in drone ship, booster versions, launch_site for the months in year 2015.
- Rank the count of successful landiing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Build an Interactive Map with Folium

Folium map object is a map centered on NASA Johnson Space Center at Houson, Texas

- Red circle at NASA Johnson Space Center's coordinate with label showing its name *(folium.Circle, folium.map.Marker).*
- Red circles at each launch site coordinates with label showing launch site name *(folium.Circle, folium.map.Marker, folium.features.DivIcon).*
- The grouping of points in a cluster to display multiple and different information for the same coordinates *(folium.plugins.MarkerCluster).*
- Markers to show successful and unsuccessful landings. Green for successful landing and Red for unsuccessful landing. *(folium.map.Marker, folium.Icon).*
- Markers to show distance between launch site to key locations (railway, highway, coastway, city) and plot a line between them. *(folium.map.Marker, folium.PolyLine, folium.features.DivIcon)*

Build a Dashboard with Plotly Dash

Dashboard has dropdown, pie chart, rangeslider and scatter plot components

- Dropdown allows a user to choose the launch site or all launch sites *(dash_core_components.Dropdown)*.
- Pie chart shows the total success and the total failure for the launch site chosen with the dropdown component *(plotly.express.pie).*
- Rangeslider allows a user to select a payload mass in a fixed range *(dash_core_components.RangeSlider).*
- Scatter chart shows the relationship between two variables, in particular Success vs Payload Mass *(plotly.express.scatter)*

“ Predictive Analysis ”

Data preparation

- Load dataset
- Normalize data
- Split data into training and test sets.

Model preparation

- Selection of machine learning algorithms
- Set parameters for each algorithm to GridSearchCV
- Training GridSearchModel models with training dataset

Model evaluation

- Get best hyperparameters for each type of model
- Compute accuracy for each model with test dataset
- Plot Confusion Matrix

Model comparison

- Comparison of models according to their accuracy
- The model with the best accuracy will be chosen (see Notebook for result)

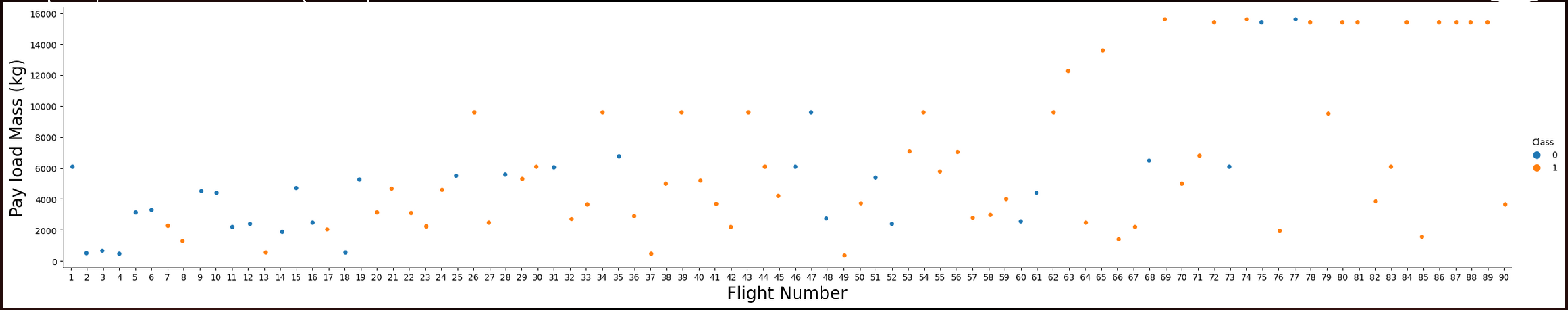
"Results"

Summary

- EDA with visualization results
- EDA with SQL results
- interactive map with Folium results
- Plotly Dash dashboard results
- predictive analysis (classification) results

❝ EDA with visualization results ❞

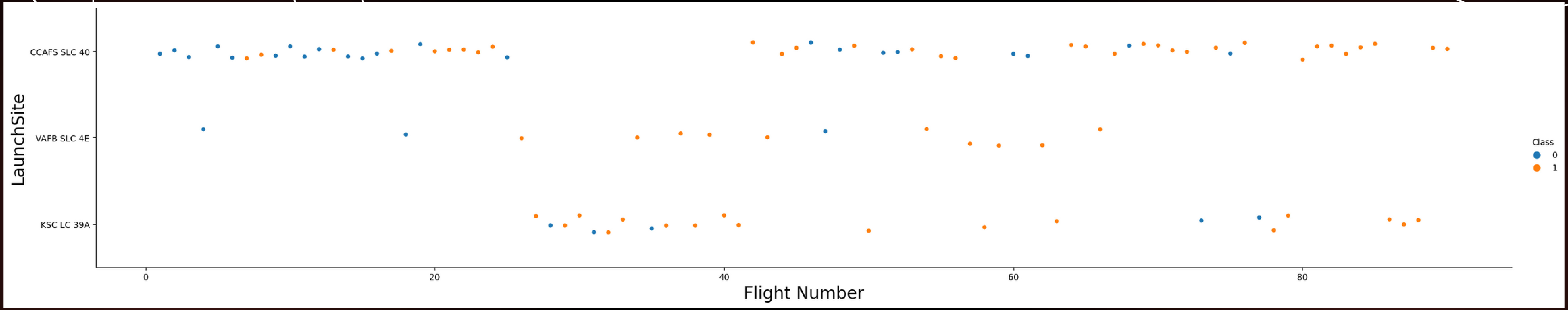
Flight Number vs Payload Mass



- as the flight number increases, the first stage is more likely to land successfully.
- it seems the more massive the payload, the less likely the first stage will return.

“ EDA with visualization results ”

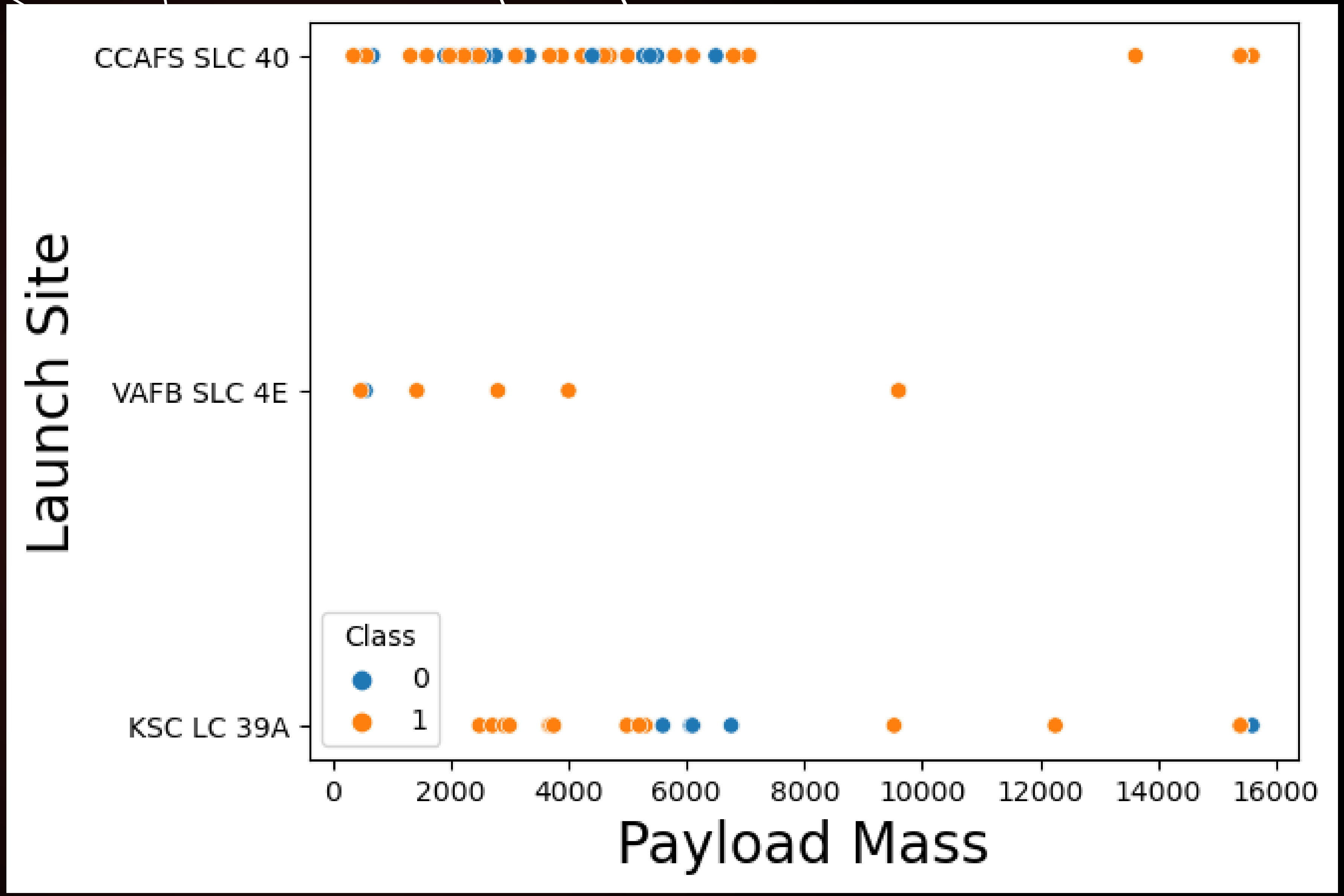
Flight Number vs Launch Site



- The earliest flights all failed while the latest flights more likely to succeed
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- We can assumed that the general success rate improved over time.

“ EDA with visualization results ”

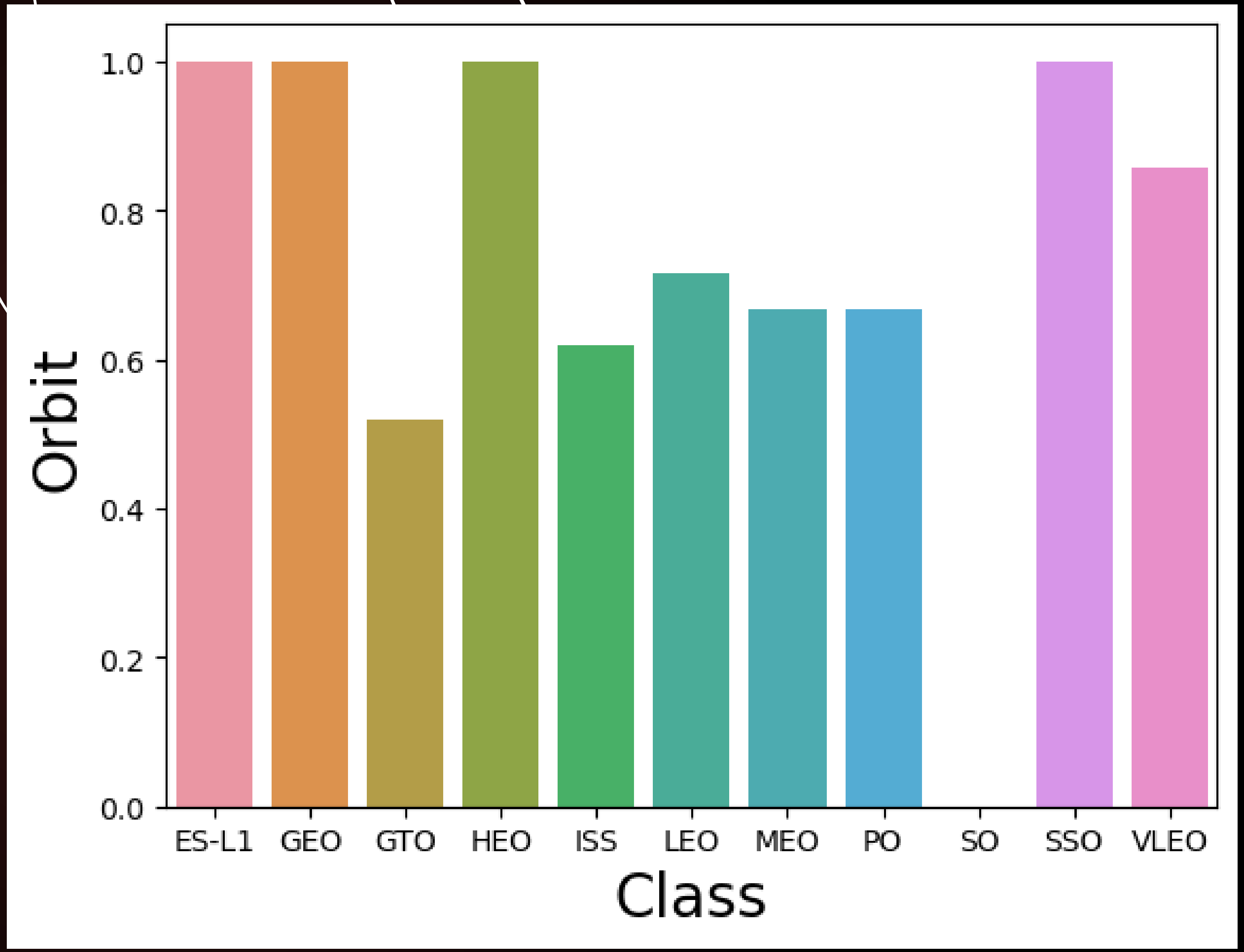
Payload Mass vs Launch Site



- CCAFS SLC 40 and KSC LC 9A has higher rate of success when payload mass is higher than 10000kg
- VAFB SLC 4E there are no rockets launched for heavy payload mass(greater than 10000).
- KSC LC 39A has 100% success rate when payload mass is less than 5000kg

EDA with visualization results

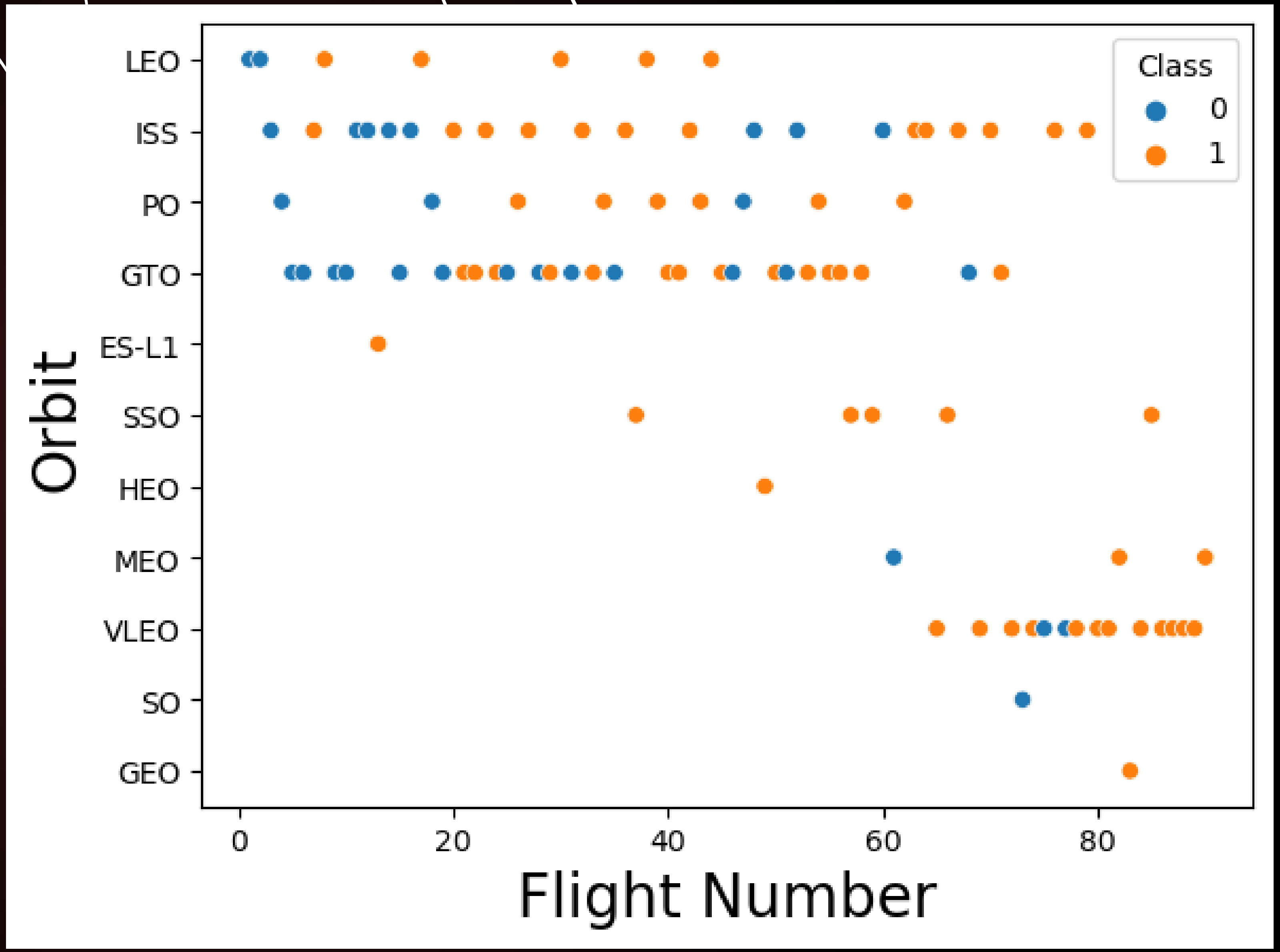
Successful Rate of each Orbit Type



- ES-L1, GEO, HEO, and SSO has 100% success rate
- GTO, ISS, LEO, MEO, PO, and VLEO has success rate between 50% and 85%

EDA with visualization results

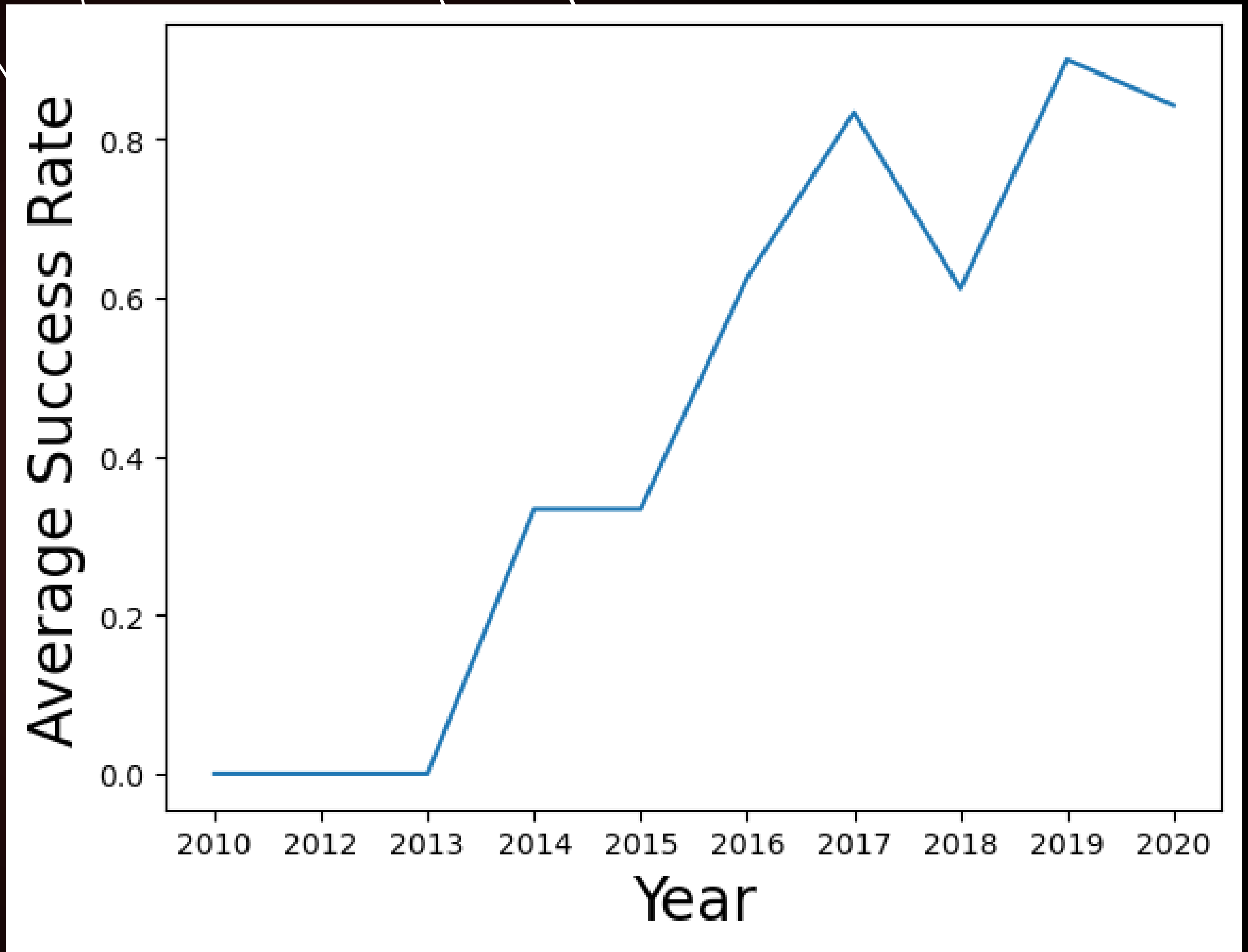
Flight Number vs Orbit Type



Successful rate increased as flight number increase, meaning as more flight happenned the better the results. except for GTO there is no relation with flight number

“ EDA with visualization results ”

Launch Success Yearly Trend



- Success rate started increasing in 2013 and kept until 2020;
- we can assumed that every year the successful rate is increased, with only in 2018 and 2020 the rate was decreased, so further analysis need to be done as to know why

“ EDA with SQL results ”

All Launch Site Names

In [9]:

```
%%sql
SELECT DISTINCT "Launch_Site" from SPACEXTBL
```

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

Out[9]:

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

❝ EDA with SQL results ❞

Launch site names begin with `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_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	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	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

❝ EDA with SQL results ❞

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

```
%%sql
SELECT Customer, SUM(PAYLOAD_MASS__KG_) AS Total_Payload FROM SPACEXTBL
GROUP BY Customer
HAVING Customer = "NASA (CRS)"
```

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

Customer	Total_Payload
NASA (CRS)	45596.0

❝ EDA with SQL results ❞

average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT Booster_Version, AVG(PAYLOAD_MASS__KG_) AS Average_Payload FROM SPACEXTBL
GROUP BY Booster_Version
HAVING Booster_Version = "F9 v1.1"
```

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

Booster_Version	Average_Payload
F9 v1.1	2928.4

“ EDA with SQL results ”

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

```
%%sql
SELECT date FROM SPACEXTBL
WHERE Landing_Outcome = "Success (ground pad)"
```

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

Date
22/12/2015

“ EDA with SQL results ”

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

```
%%sql
SELECT Booster_Version, PAYLOAD_MASS__KG_, Landing_Outcome FROM SPACEXTBL
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	4696.0	Success (drone ship)
F9 FT B1026	4600.0	Success (drone ship)
F9 FT B1021.2	5300.0	Success (drone ship)
F9 FT B1031.2	5200.0	Success (drone ship)

❝ EDA with SQL results ❞

the total number of successful and failure mission outcomes

```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) from SPACEXTBL
GROUP BY Mission_Outcome
```

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

Mission_Outcome	COUNT(Mission_Outcome)
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

❝ EDA with SQL results ❞

the names of the booster_versions 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

❝ EDA with SQL results ❞

2015 launch records

```
%%sql
SELECT SUBSTR(Date,4,2) as Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL
WHERE Landing_Outcome = "Failure (drone ship)" AND substr(Date,7,4) = "2015"
```

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

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

❝ EDA with SQL results ❞

Rank success count between 2010-06-04 and 2017-03-20

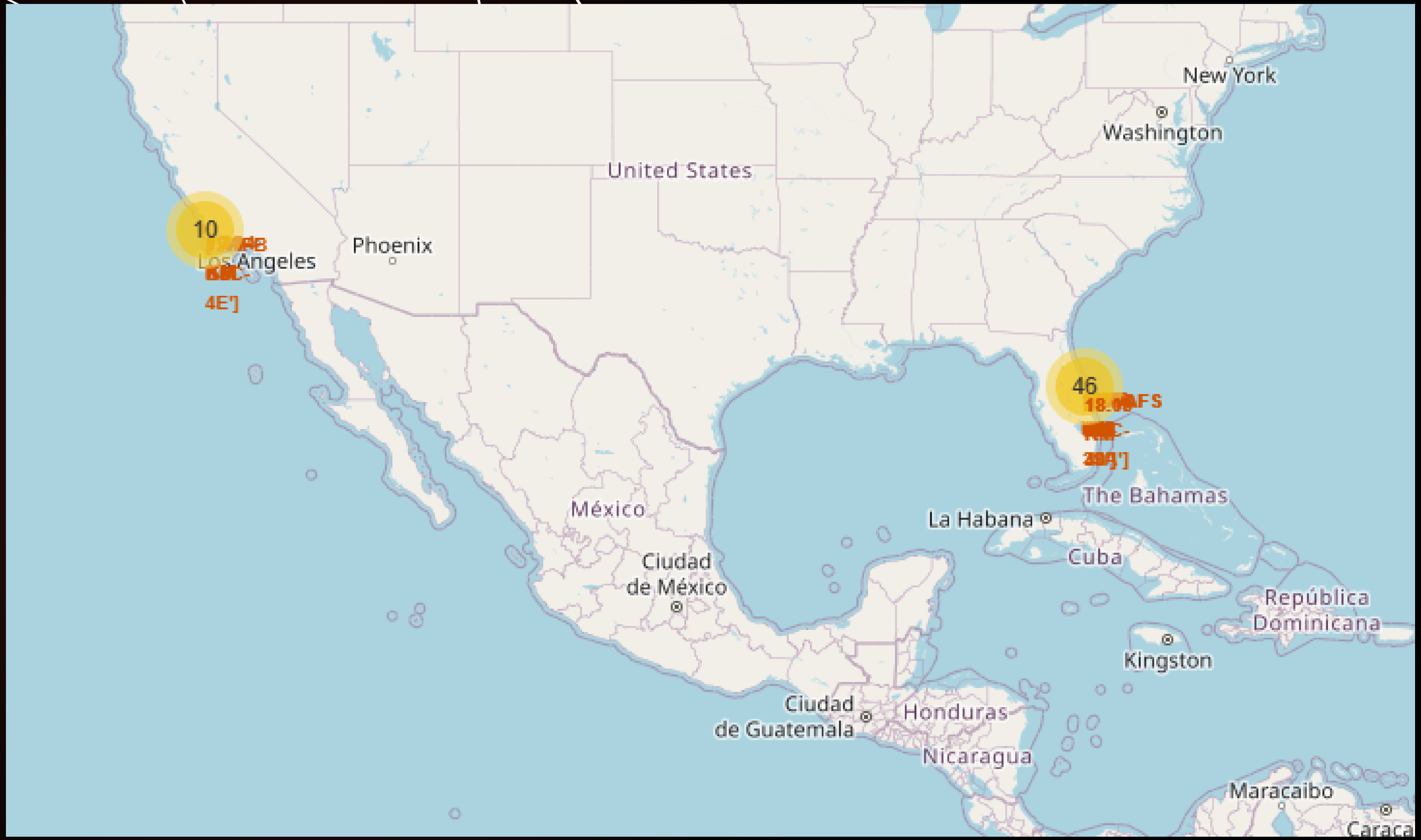
```
%%sql
SELECT Date, Landing_Outcome, COUNT(Landing_Outcome) AS TOTAL_NUMBER
FROM SPACEXTBL
WHERE CAST(substr(Date,7,4) AS INT) > 2010 AND CAST(substr(Date,7,4) AS INT) < 2017
GROUP BY Landing_Outcome
ORDER BY TOTAL_NUMBER DESC
```

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

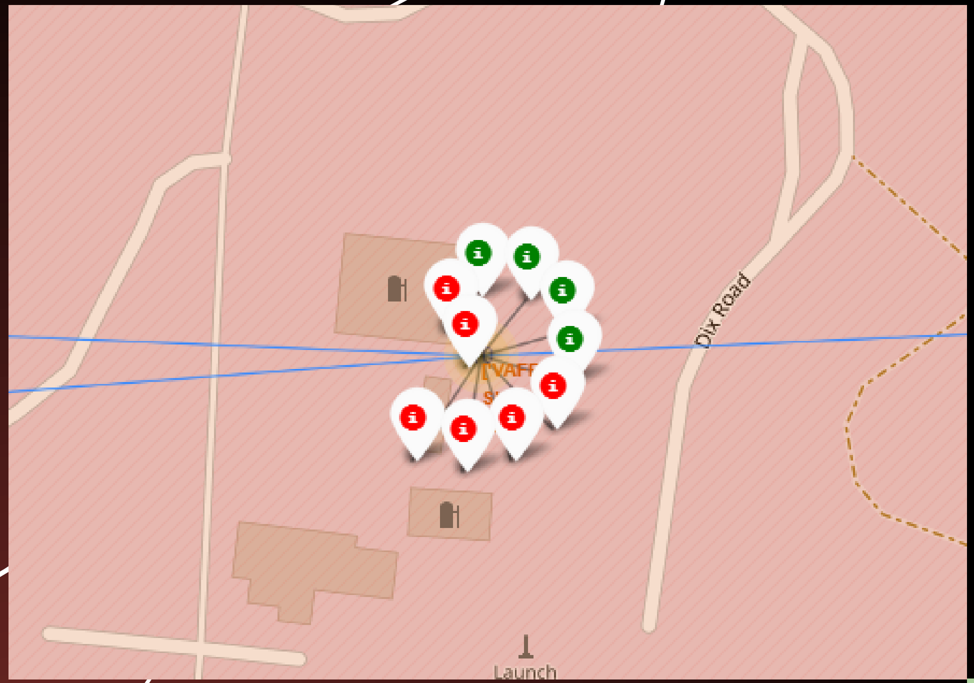
Date	Landing_Outcome	TOTAL_NUMBER
22/05/2012	No attempt	9
01/10/2015	Failure (drone ship)	5
04/08/2016	Success (drone ship)	4
18/04/2014	Controlled (ocean)	3
29/09/2013	Uncontrolled (ocean)	2
22/12/2015	Success (ground pad)	2
28/06/2015	Precluded (drone ship)	1

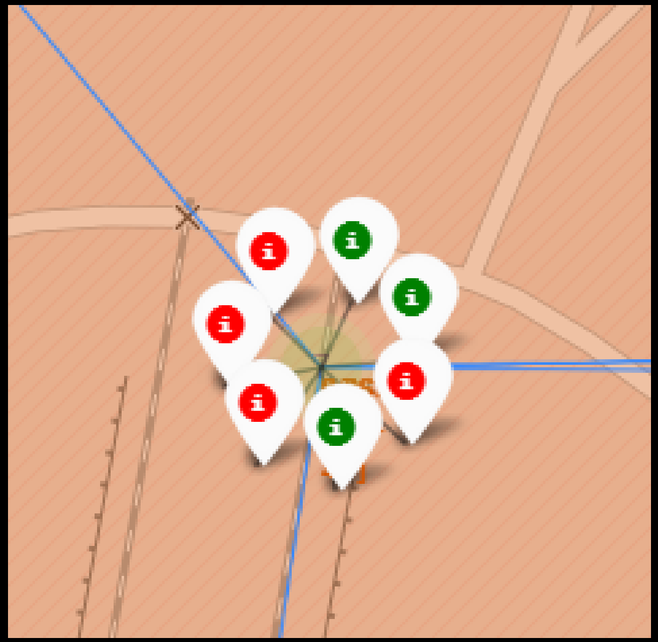
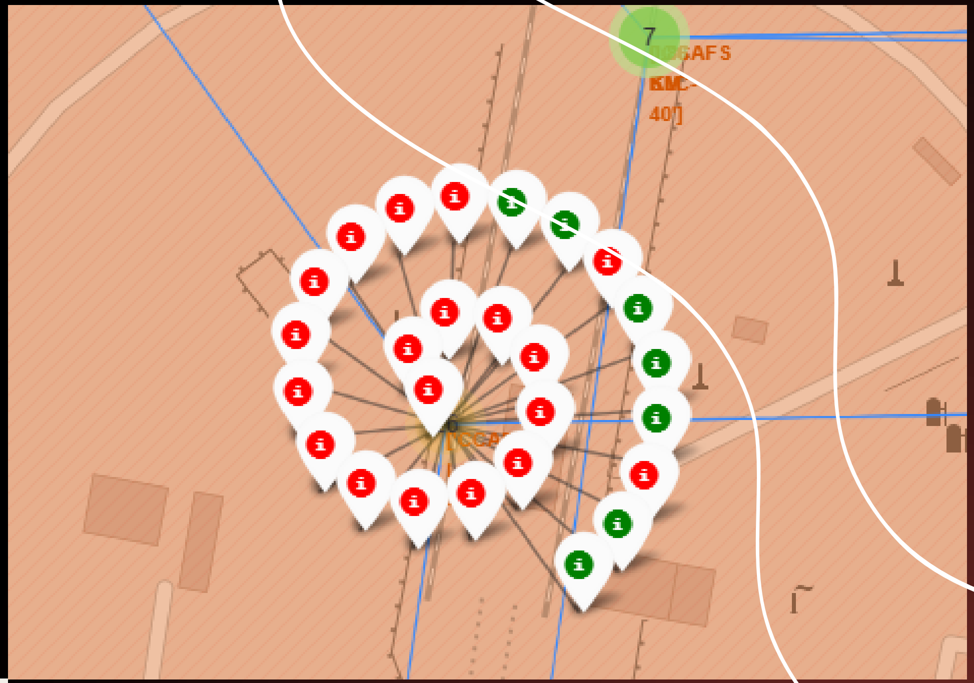
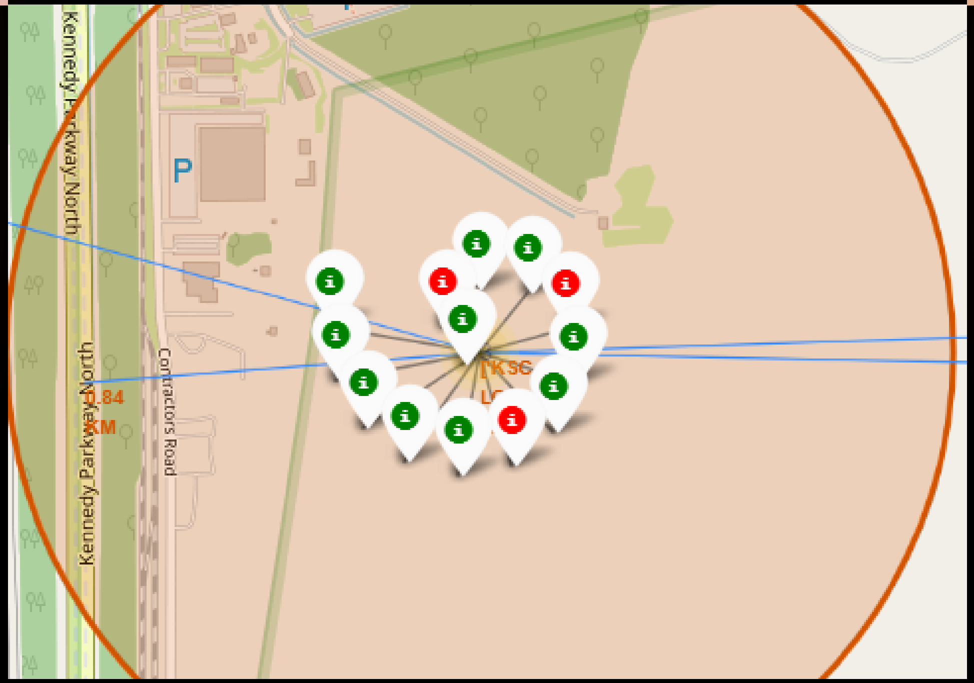
interactive map with Folium results

All Sites

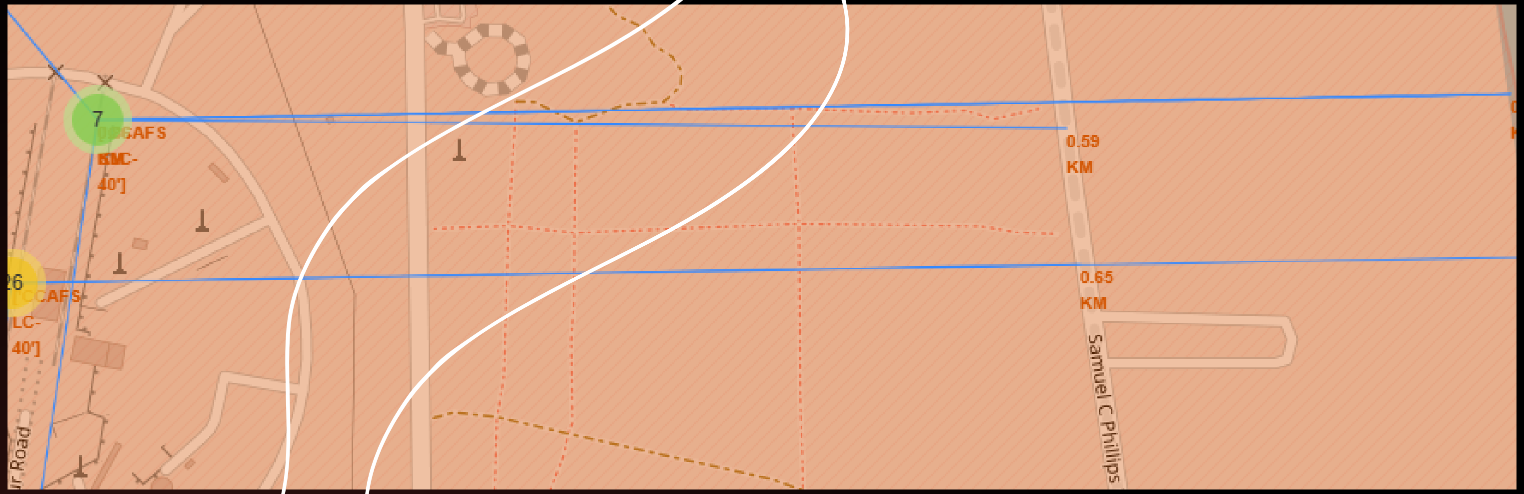


All Sites of launches were near the coast of United States

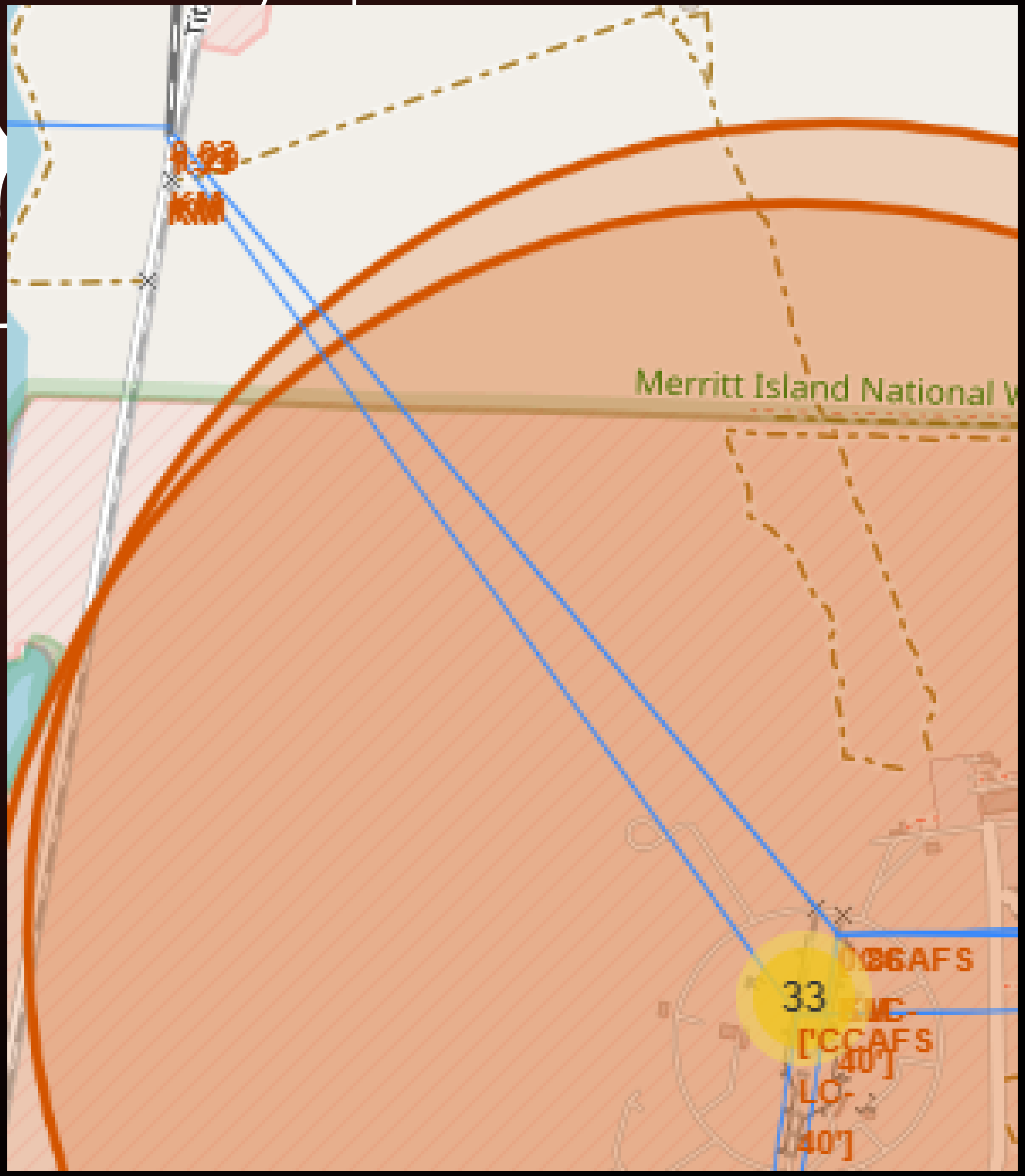


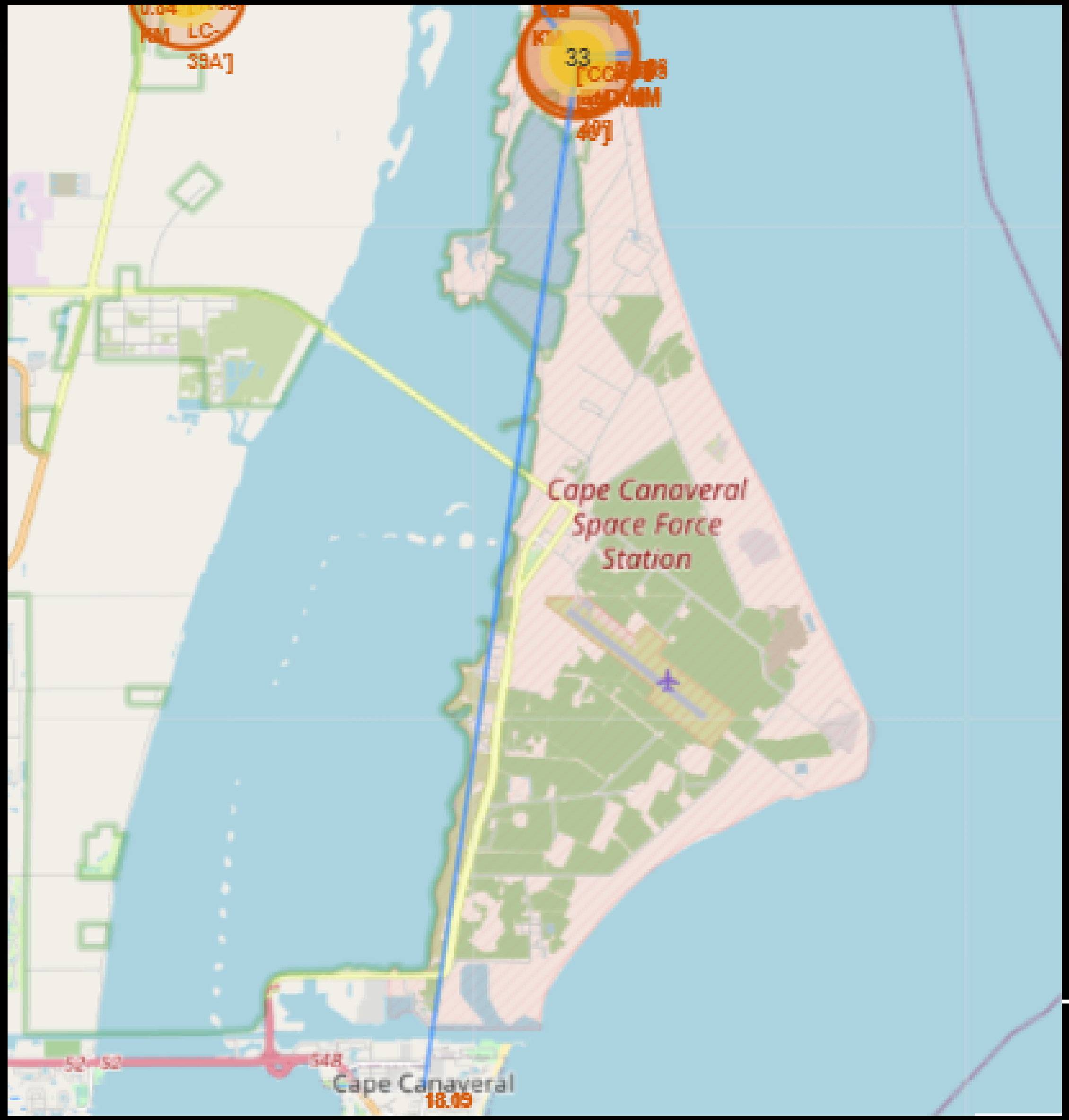




Green marker represents successful launches. Red marker represents unsuccessful launches.



"Distance from the launch site CCAFS to its proximities"

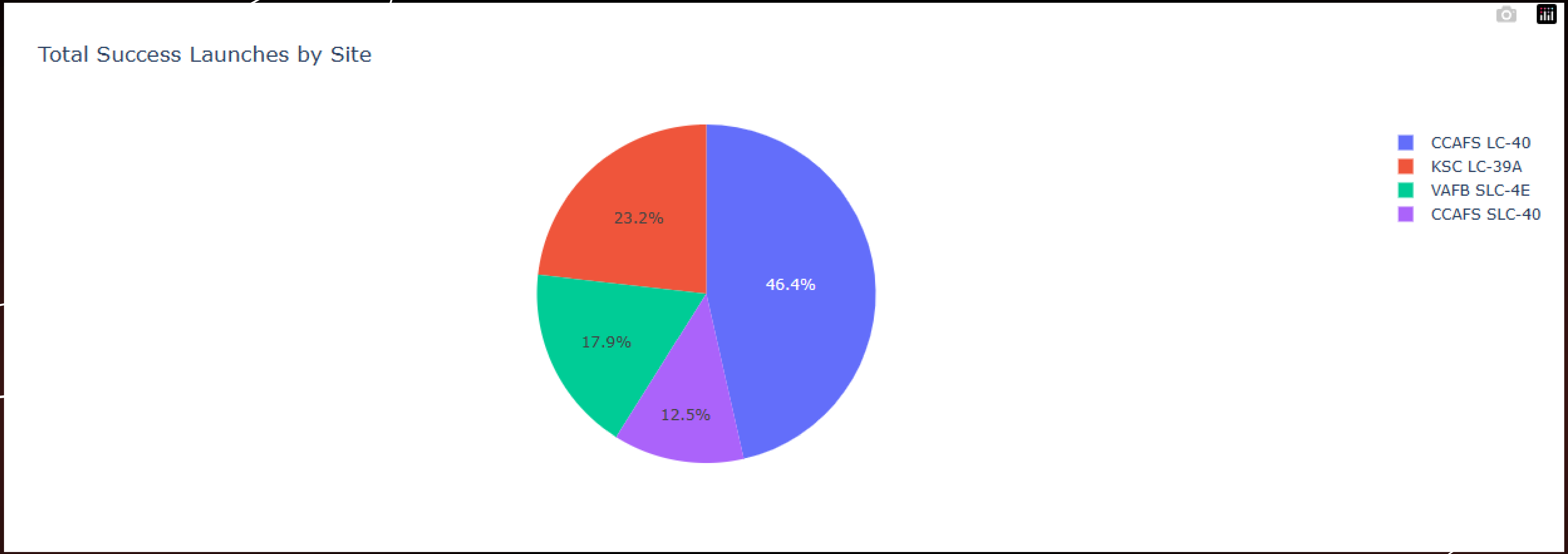




From the visual analysis of the launch site CCAFS we can clearly see that it is:

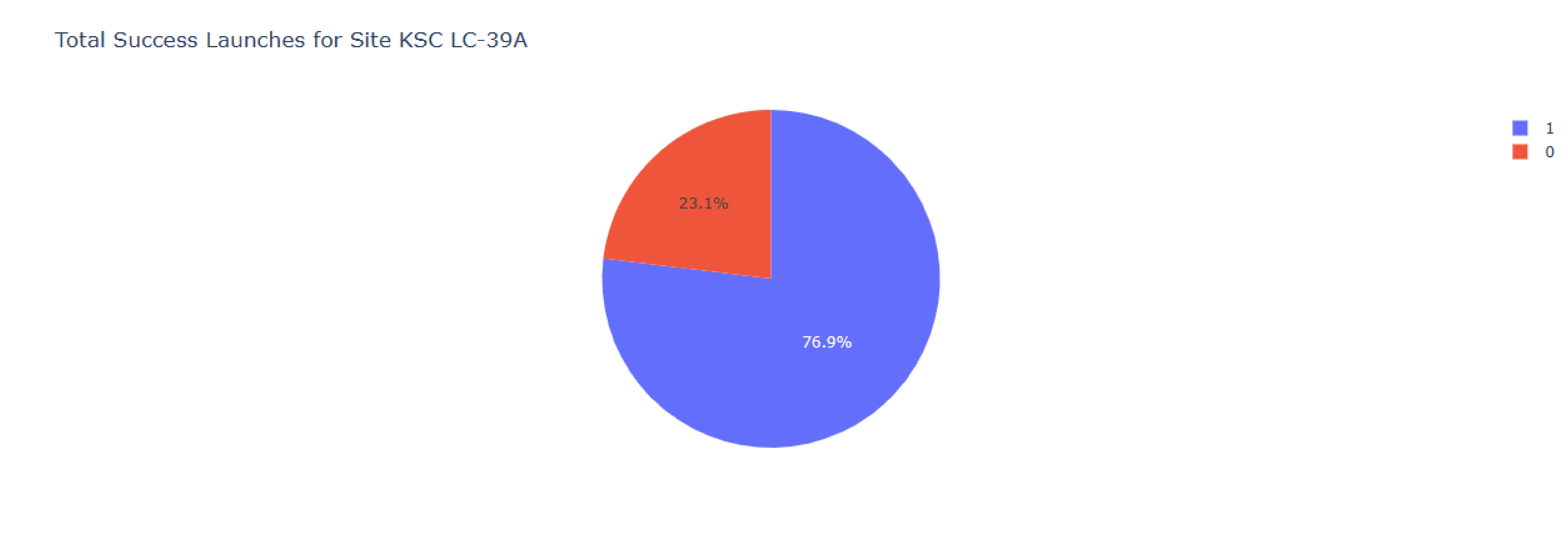
relative close to railway (1,32 km)
relative close to highway (0.60 km)
relative close to coastline (0,90 km)
relatively far away from city (18 km)

Plotly Dash dashboard results



The place from where launches are done seems to be a very important factor of success of missions.

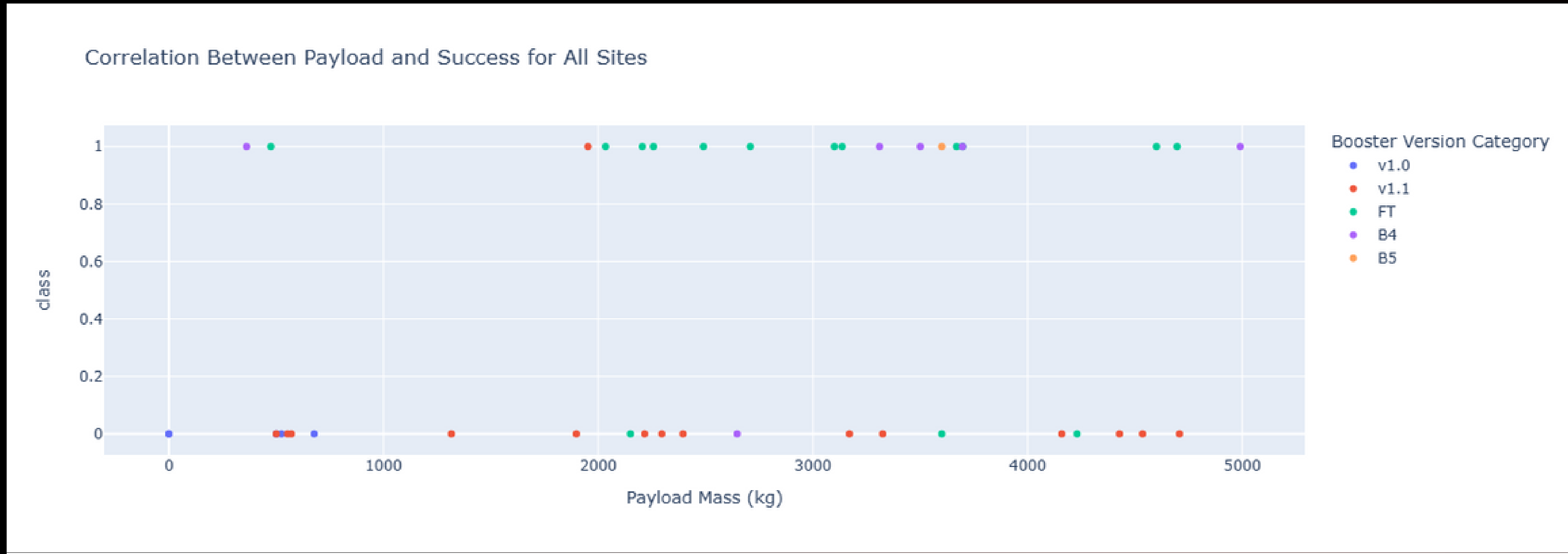
“ Plotly Dash dashboard results ”

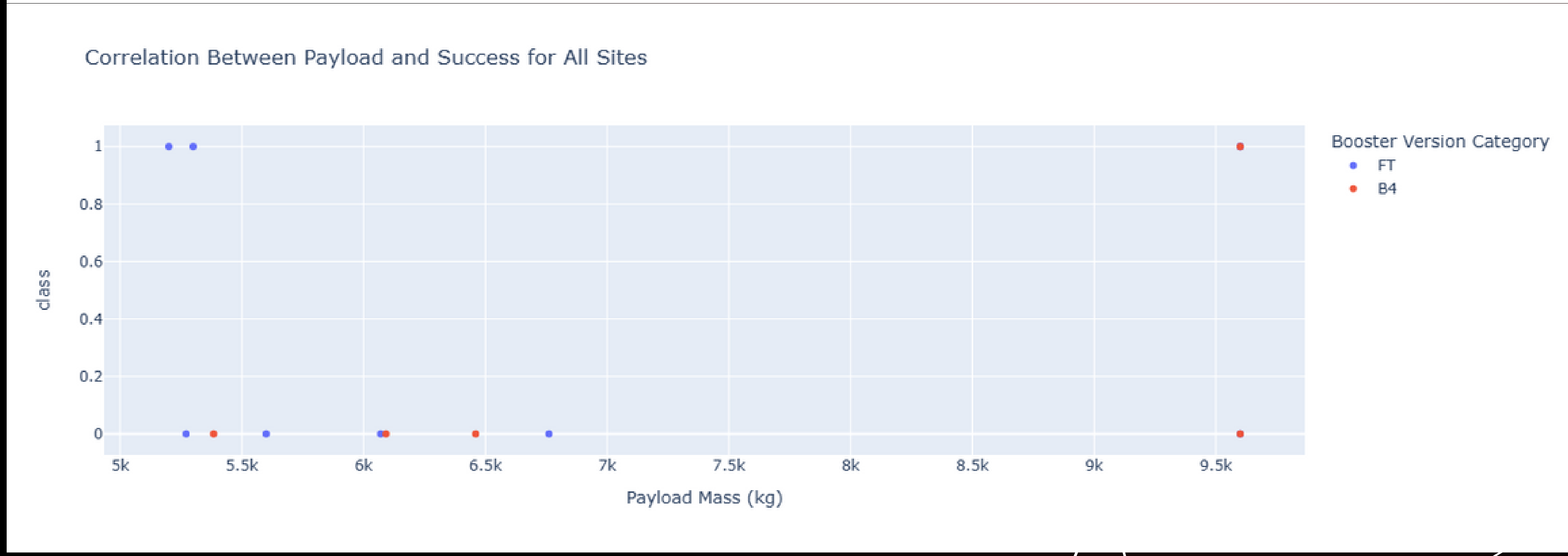


KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings.

“ Plotly Dash dashboard results ”

The charts show that payloads between 2000 and 5500 kg have the highest success rate.





Predictive analysis (Classification) Results

Accuracy

	Accuracy Train	Accuracy Test
Tree	0.876786	0.833333
Knn	0.848214	0.833333
Svm	0.848214	0.833333
Logreg	0.846429	0.833333

- For accuracy test, all methods performed similar. We could get more test data to decide between them.
- But if we really need to choose one right now, we would take the decision tree.

DECISION TREE BEST PARAMETERS

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 5, 'splitter': 'random'}
accuracy : 0.9
```

Predictive analysis (Classification) Results

Confusion Matrix

- As the test accuracy are all equal, the confusion matrices are also identical.
- The main problem of these models are false positives.

❝ Conclusion ❞

- The success of a mission can be explained by several factors such as the launch site, the orbit and especially the number of previous launches. Indeed, we can assume that there has been a gain in knowledge between launches that allowed to go from a launch failure to a success.
- Depending on the orbits, the payload mass can be a criterion to take into account for the success of a mission. Some orbits require a light or heavy payload mass. But generally low weighted payloads perform better than the heavy weighted payloads
- Depending on the launch sites, CCAFS SLC 40 and KSC LC 9A has higher rate of success when payload mass is higher than 10000kg, in contrast where KSC LC 39A has 100% success rate when payload mass is less than 5000kg
- For this dataset, we choose the Decision Tree Algorithm as the best model even if the test accuracy between all the models used is identical. We choose Decision Tree Algorithm because it has a better train accuracy.