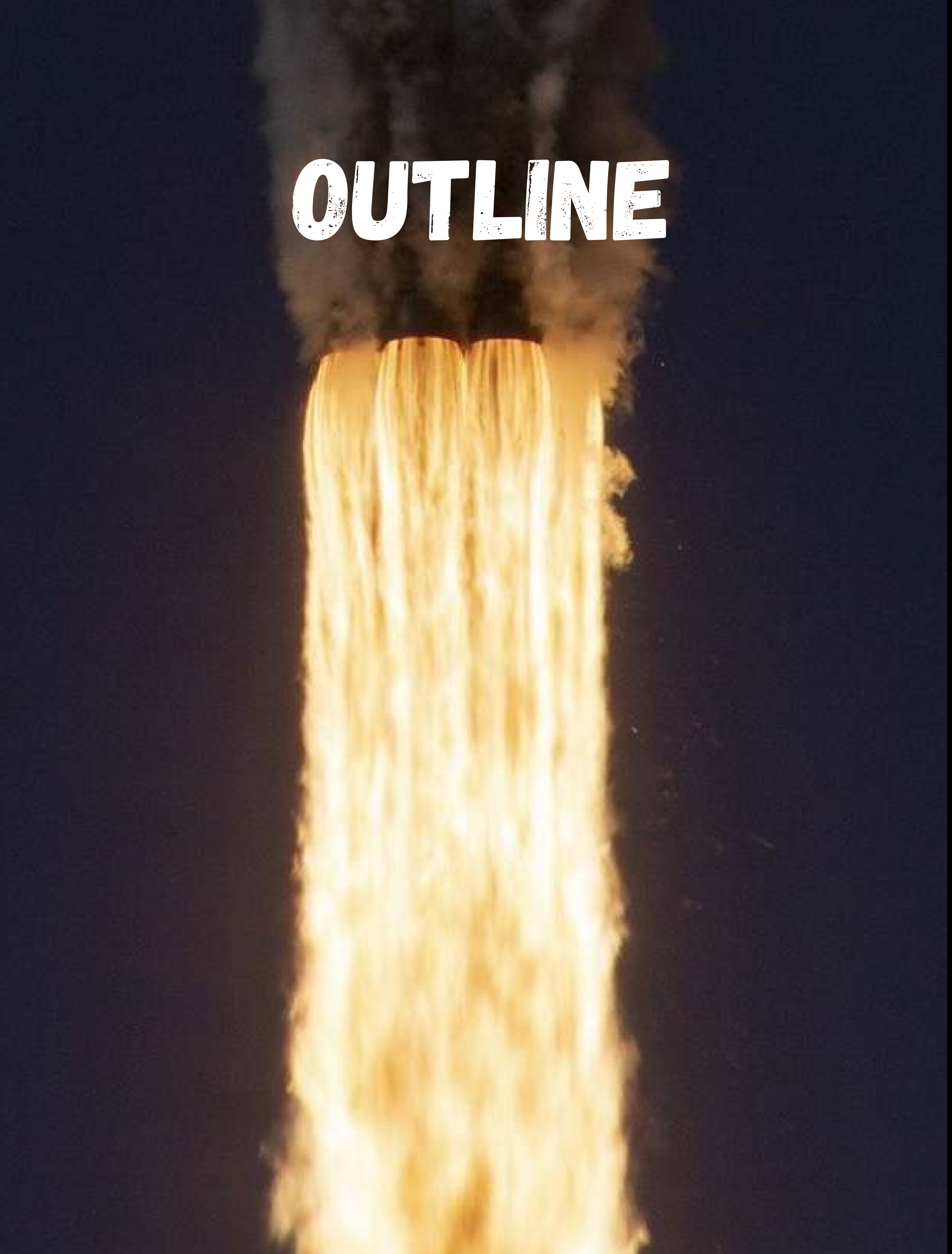
SPACEX FALCON-9 FIRST STAGE LANDING PREDICTION

FINAL PROJECT CAPSTON
IBM DATA SCIECE PROFESSIONAL CERTIFICATE

KENBOUCH AYMANE

OUTLINE



- WHAT TO DO ?
- UNDERSTANDING THE PROBLEME SOURCE
- METHODOLOGY
- RESULTS (CONCLUSION)

01 - WHAT TO DO ?

DATA COLLECTION
API OR WEBPAGE

DATA WRANGLING

EDA VISUALISATION WITH DATA OR SQL

INTERACTIVE VISUALISATION WITH FOLIUM

PREDECTIVE ANALYSIS BY
MACHINE LEARNING

02 - UNDERSTANDING THE PROBLEME :

SpaceX’s Goal ?

- Sending spacecraft to the international space station
- Providing satellite internet to the whole world with Starlink technology
- Taking people and cargo into space and contributing to space exploration.

Falcon 9

FIRST ORBITAL CLASS ROCKET CAPABLE OF REFLIGHTg

Falcon 9 is a reusable, two-stage rocket designed and manufactured by SpaceX for the reliable and safe transport of people and payloads into Earth orbit and beyond. Falcon 9 is the world’s first orbital class reusable rocket. Reusability allows SpaceX to refly the most expensive parts of the rocket, which in turn drives down the cost of space access.

Reusability : One of the notable features of Falcon 9 is its reusability. The first stage of the rocket is designed to return to Earth after launch, landing vertically either on land (at SpaceX's landing zones) or on an autonomous drone ship in the ocean. This reusability significantly reduces the cost of space launches

03 - METHODOLOGY :

Stepe 1 : Data collection

With webpage : url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

1. Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
2. Request the Falcon9 Launch Wiki page from its URL
3. Extract all column/variable names from the HTML table header
4. Create a data frame by parsing the launch HTML tables

With API : spacex_url="https://api.spacexdata.com/v4/launches/past"

1. Request and parse the SpaceX launch data using the GET request
2. Filter the dataframe to only include Falcon 9 launches

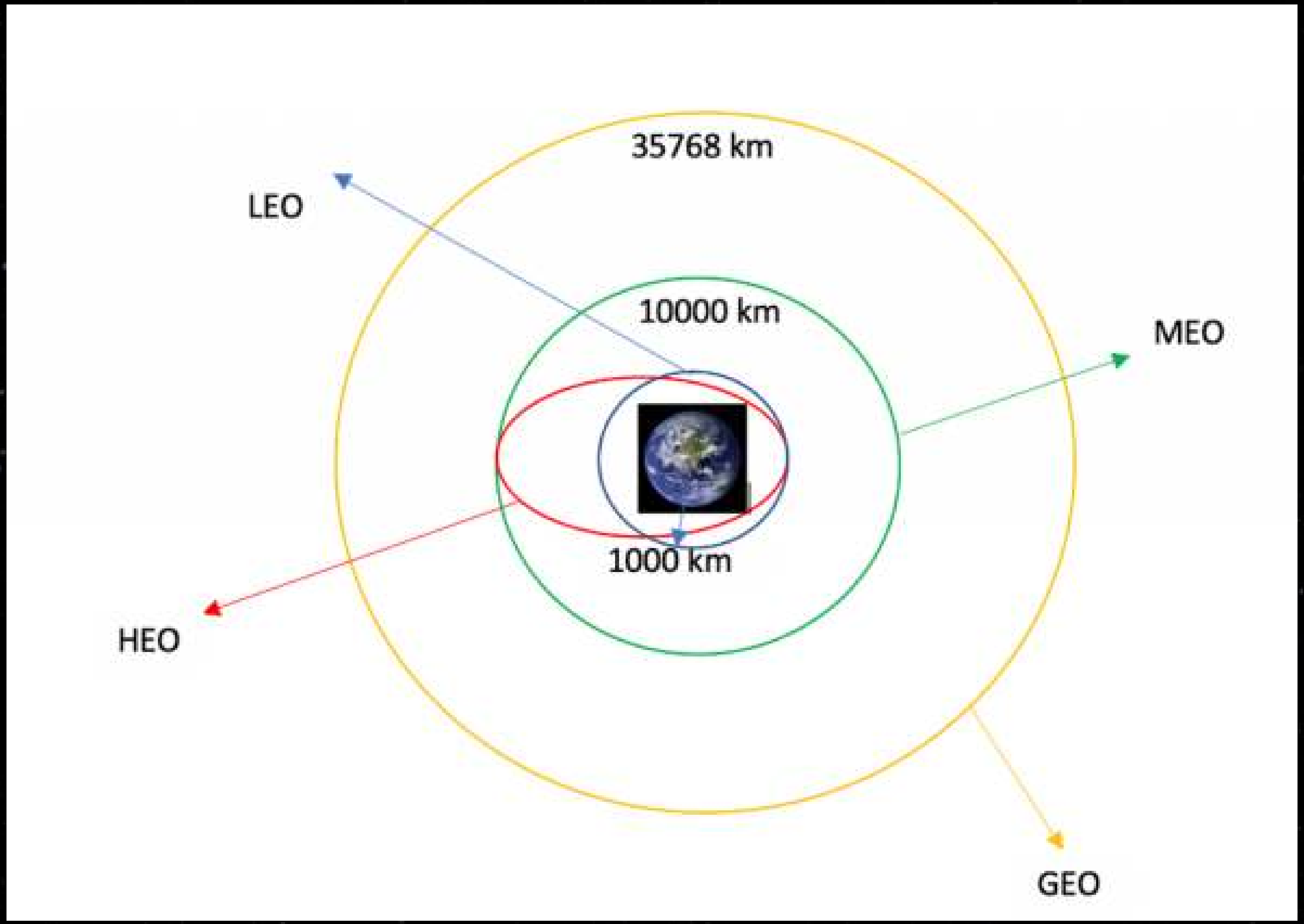
Stepe 1 : Data collection

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	Reus
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	

90 rows × 17 columns

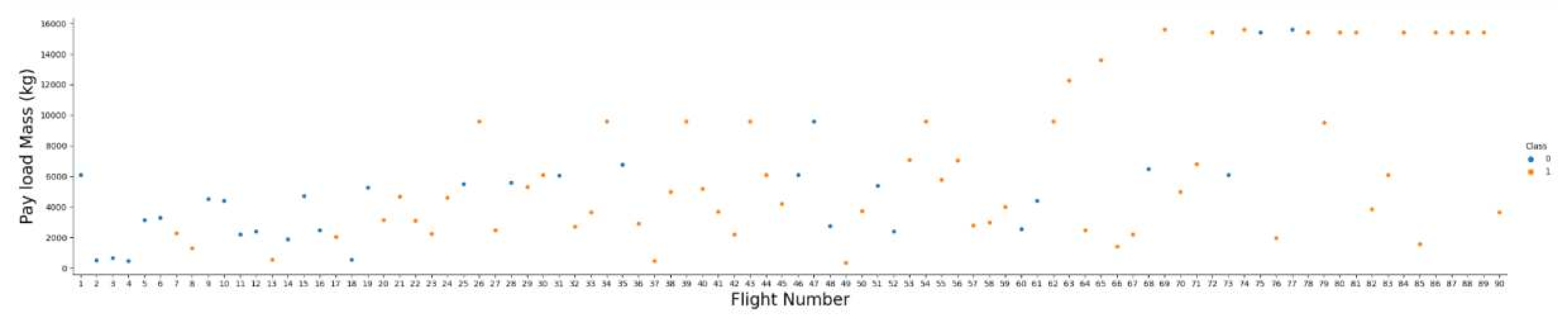
Stepe 2 : Data wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column



Stepe 3 : EDA Visualization with Data

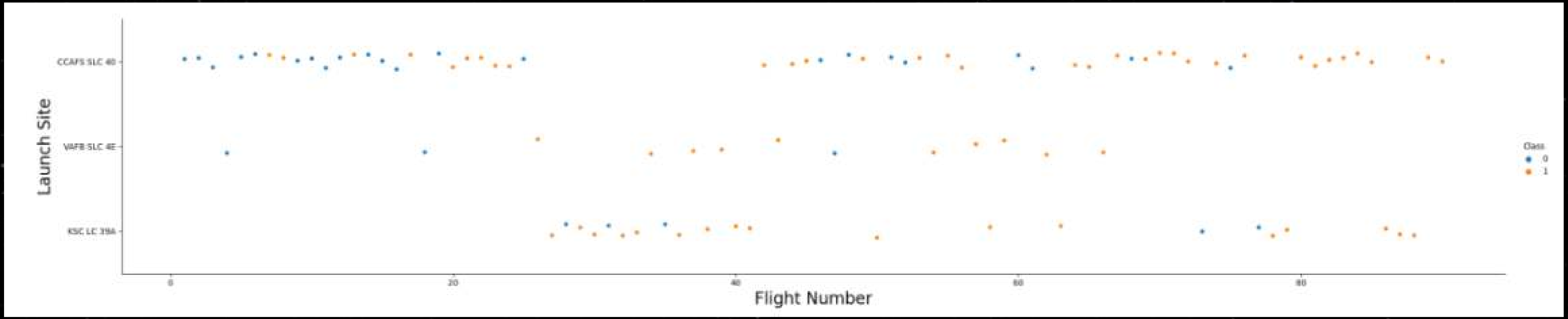
We plot out the FlightNumber vs PayloadMassand overlay the outcome of the launch



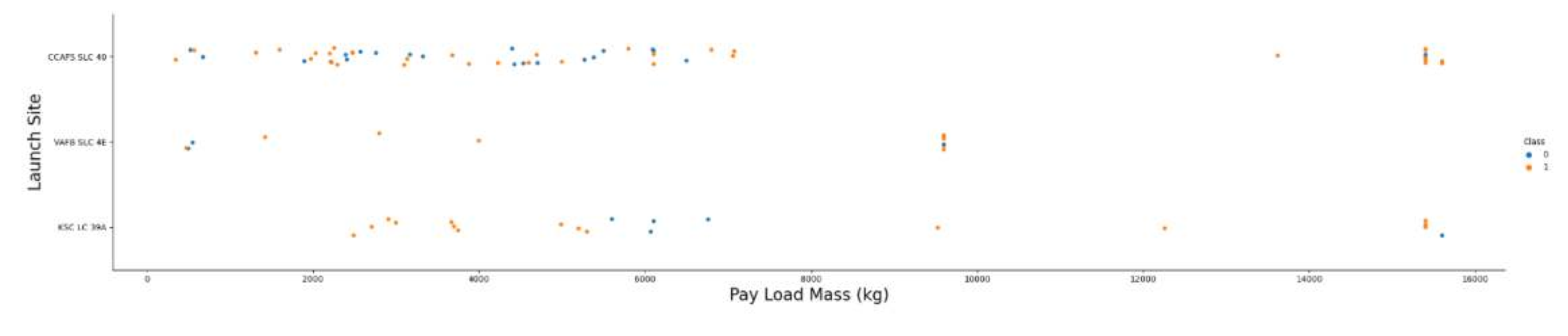
We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Stepe 3 : EDA Visualization with Data

We plot out the FlightNumber vs LaunchSite



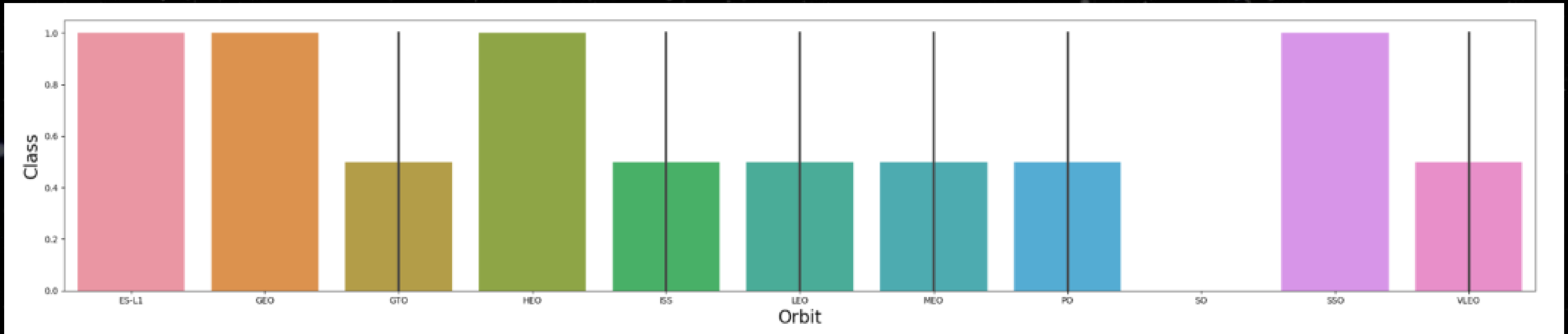
We plot out the PayLoadMass vs LaunchSite

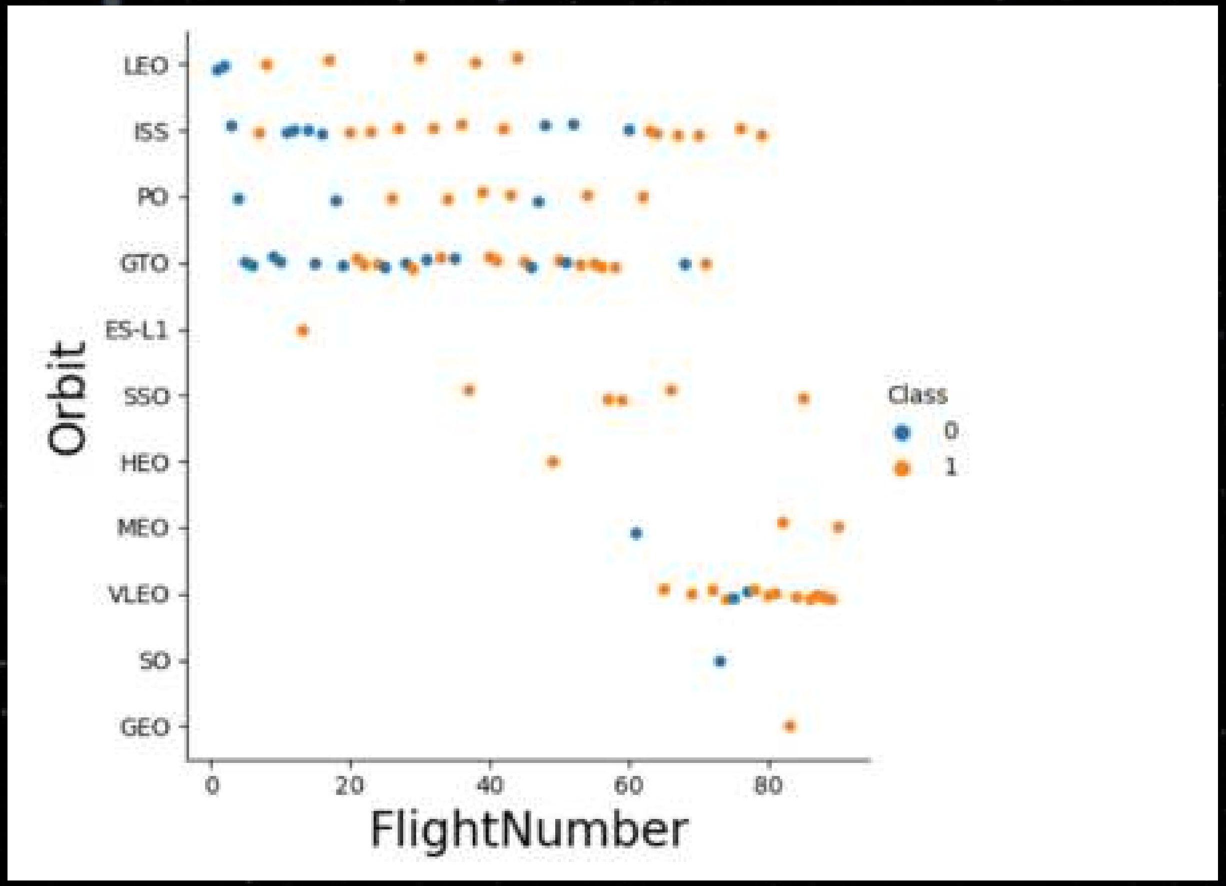


For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass > 10000

Stepe 3 : EDA Visualization with Data

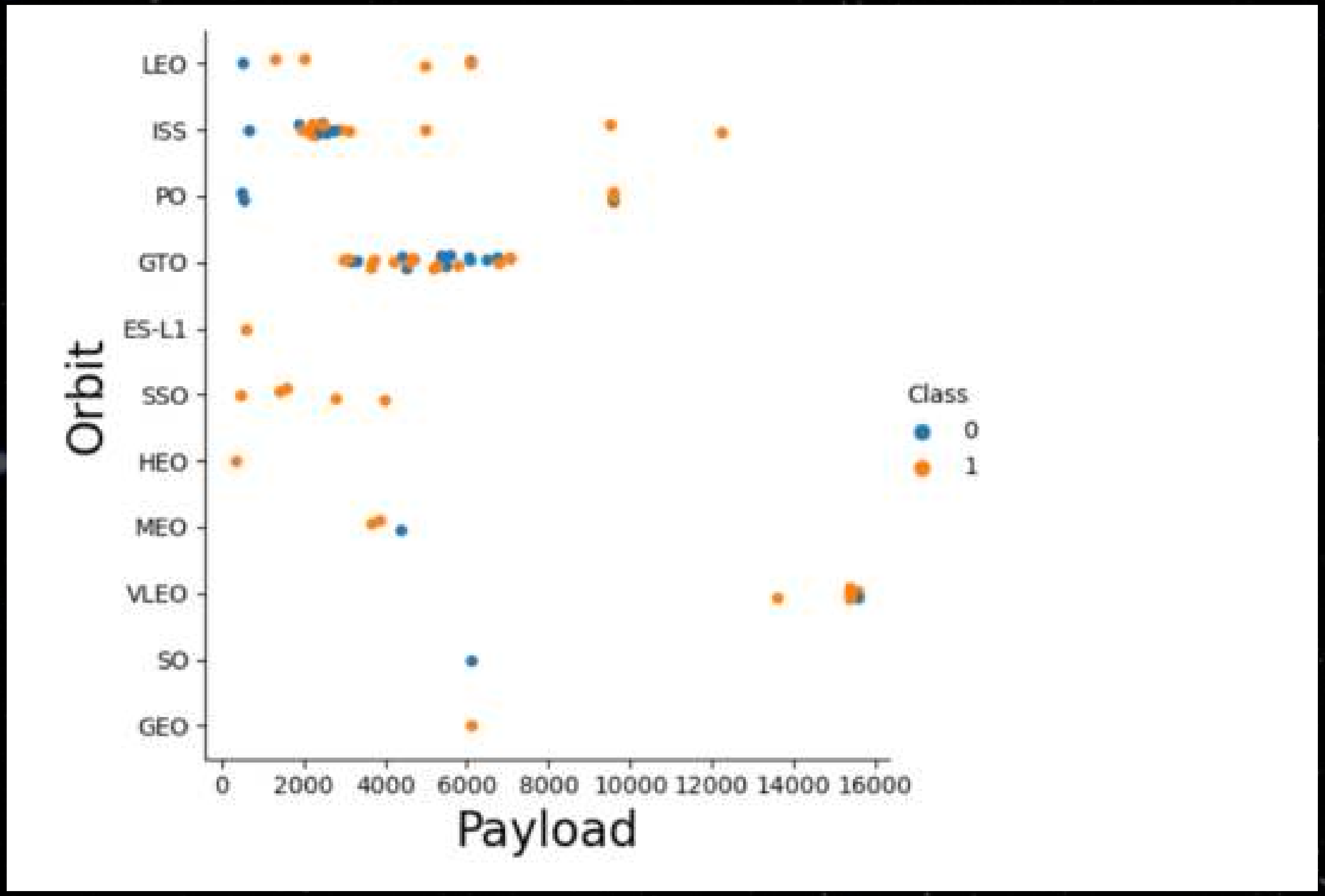
After analysing this plot <u>below</u> we figure out that the most orbits which have high success rate are :
ES-L1 ,GEO , HEO & SSO





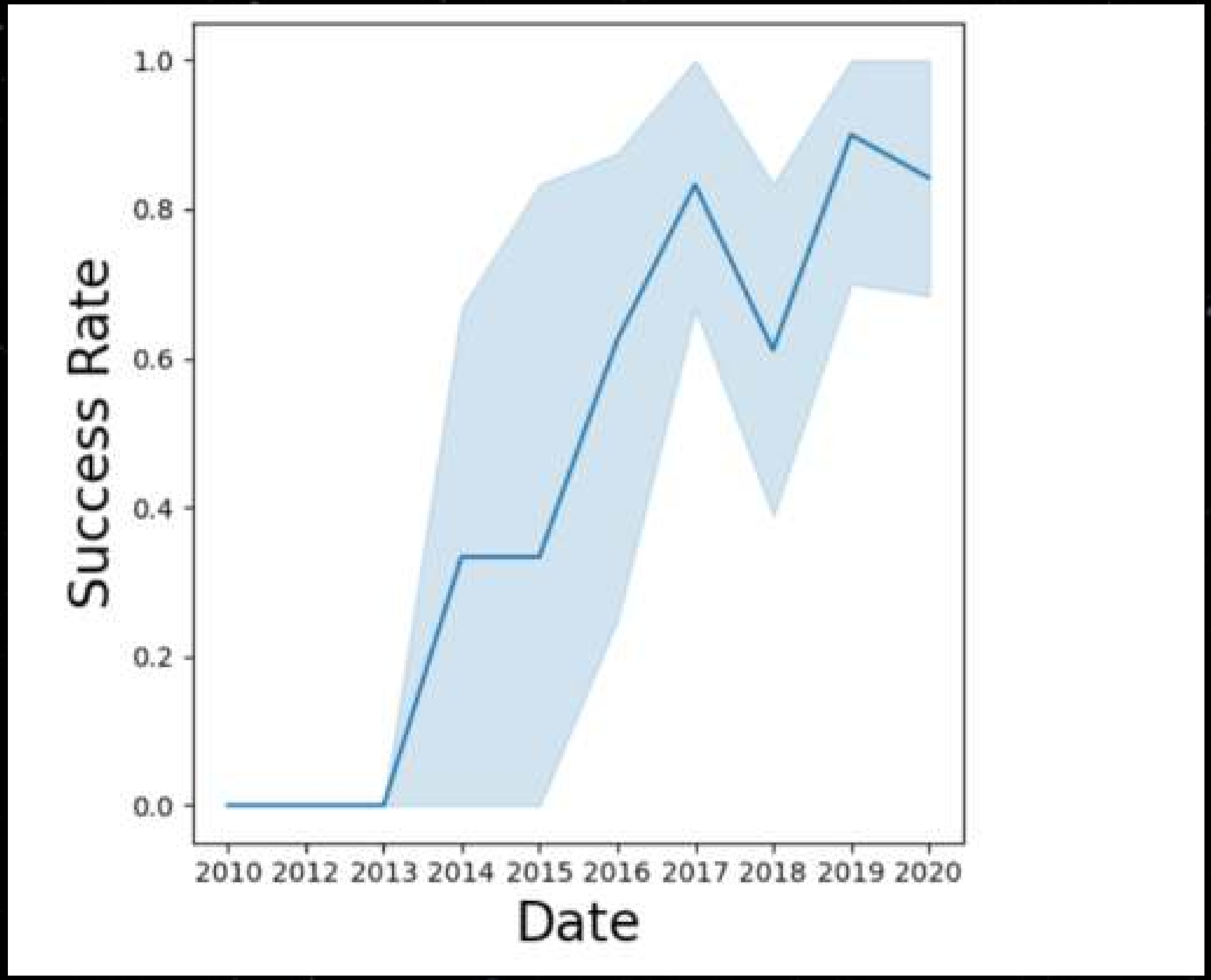
In the LEO orbit the Success appears related to the number of flights
on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Stepe 3 : EDA Visualization with Data



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here

Here we see that the sucess rate since 2013 kept increasing till 2020



Stepe 3 : EDA Visualization with SQL

We connect first the the database with Jupyter Notebook

In [2]:
```
%load_ext sql
```

In [3]:
```
import csv, sqlite3

con = sqlite3.connect("my_data1.db")
cur = con.cursor()
```

In [4]:
```
!pip install -q pandas==1.1.5
```

In [5]:
```
%sql sqlite:///my_data1.db
```

Out[5]:
```
'Connected: @my_data1.db'
```

In [6]:
```
import pandas as pd
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/dat
df.to_sql("SPACEXTBL", con, if_exists='replace', index=False,method="multi")
```

```
/home/jupyterlab/conda/envs/python/lib/python3.7/site-packages/pandas/core/generic.py:2882: UserWarning: The spaces in these co
lumn names will not be changed. In pandas versions < 0.14, spaces were converted to underscores.
  both result in 0.1234 being formatted as 0.12.
```

Stepe 3 : EDA Visualization with SQL

- The names of unique Launch Sites in the space mission

```
In [7]: %sql select distinct(LAUNCH_SITE) from SPACEXTBL
 * sqlite:///my_data1.db
Done.
Out[7]:
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

- Displaying the first 5 records that starts with “CCA”

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

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

sum(PAYLOAD_MASS__KG_)
45596.0

- The average payload mass carried by booster version F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

Stepe 3 : EDA Visualization with SQL

- List 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 from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOA
```

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

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

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

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

min(DATE)
01/08/2018

Stepe 3 : EDA Visualization with SQL

- List the names of the booster_versions which have carried the maximum payload mass.

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

Stepe 3 : EDA Visualization with SQL

- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

```
%sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN '04-06-201
```

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

Landing_Outcome	Numbers
Success	20
Success (drone ship)	8
Success (ground pad)	7

- 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.

```
%sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AN
```

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

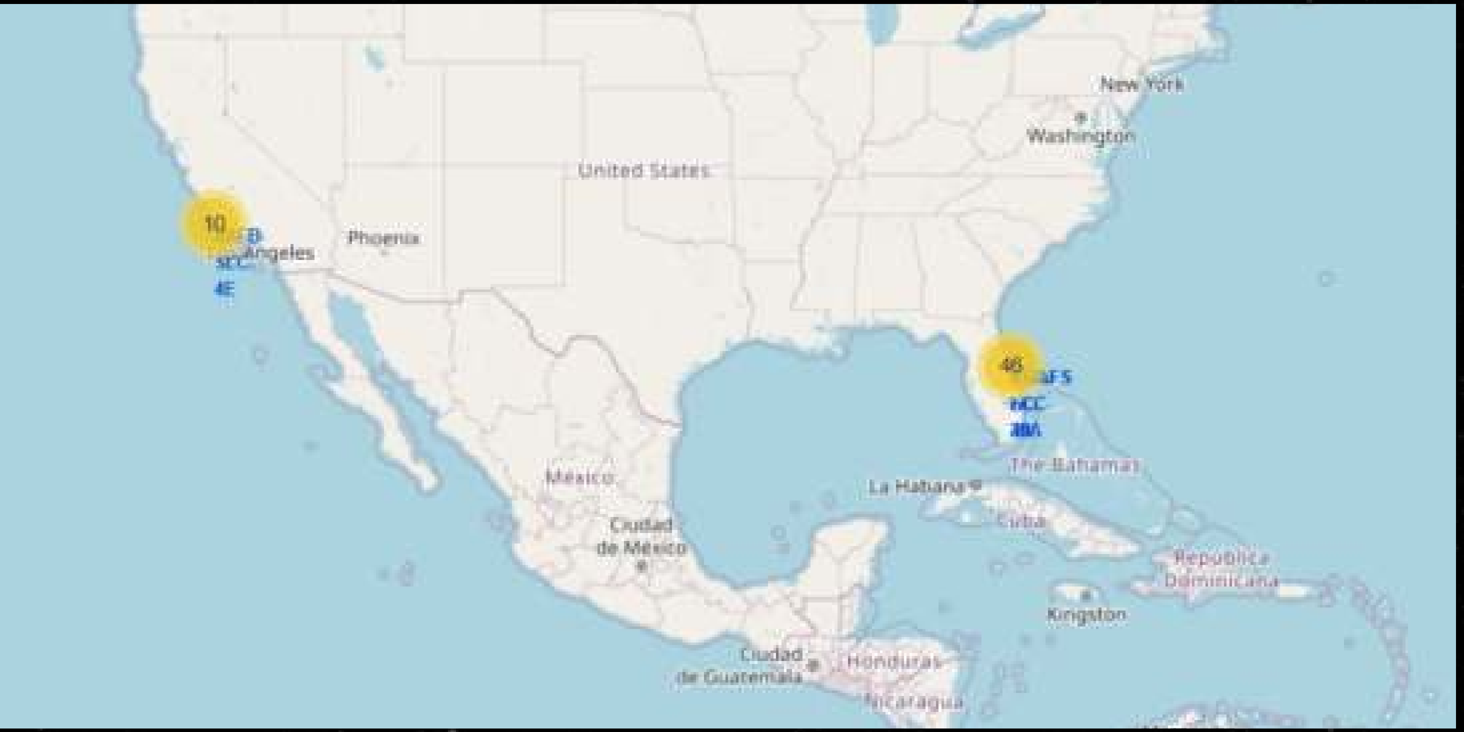
Month	Booster_Version	Launch_Site
10	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

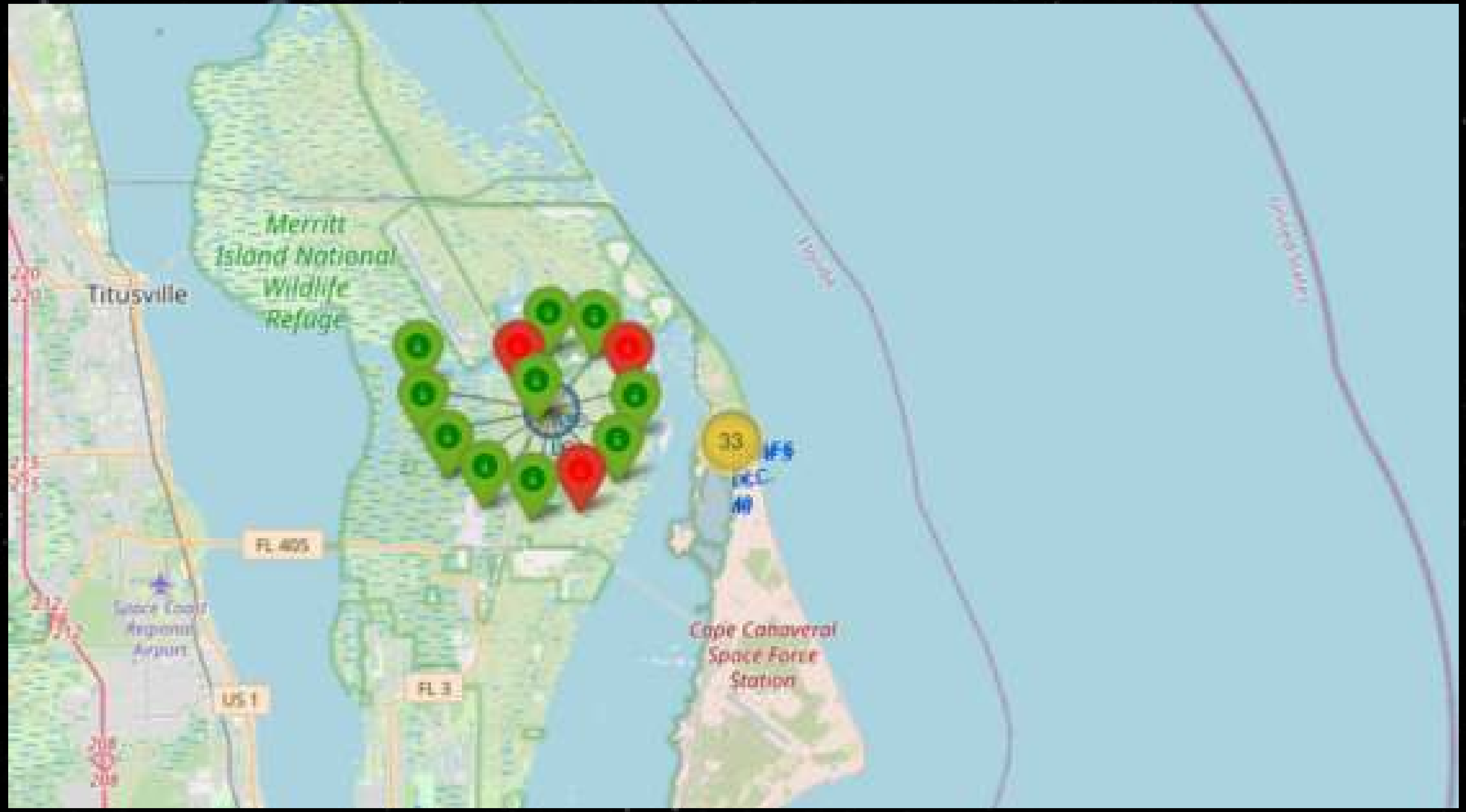
Stepe 4 : Interactive Visualisation with Folium

The launch Sites

Stepe 4 : Interactive Visualisation with Folium

These are the Launch sites , the successfil ones with green color and failed with red color for each site on the previous map





Stepe 5 : Predictive Analysis with Machine Learning

We split the data into training and testing data using the function **train_test_split**. The training data is divided into validation data, a second set used for training data; then the models are trained

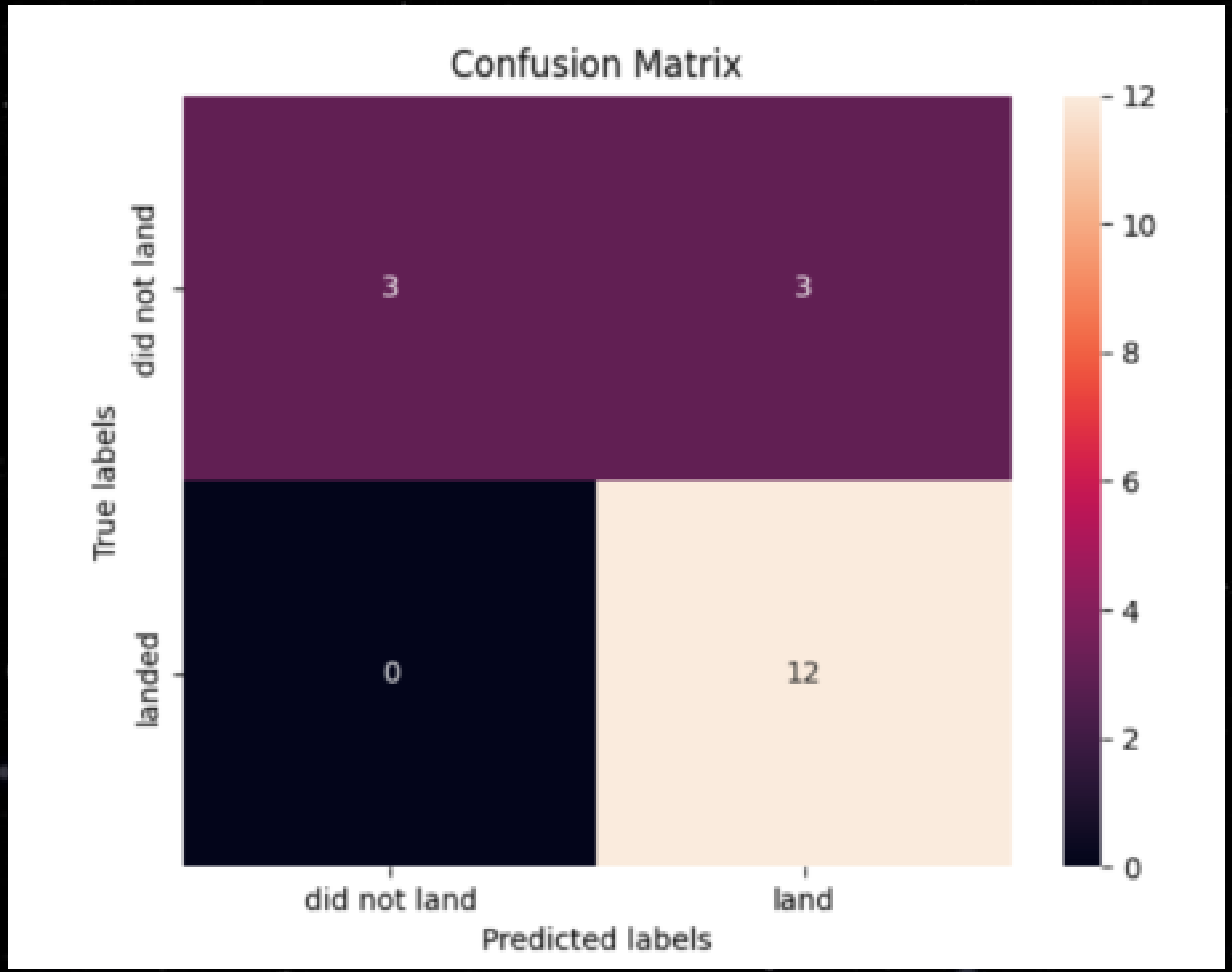
we get the shape of the Train Set and the Test Set :
Train set: (72, 83) (72,)
Test set: (18, 83) (18,)

First Algorithme to use : Logistic Regression

Tha accuracy is : **0.8464285714285713**

Calculating the accuracy using the score methode gives **0.8333333333333334** which is lower than the previous accuracy with the difference of ~0.01 or 1%

Stepe 5 : Predictive Analysis with Machine Learning



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

Stepe 5 : Predictive Analysis with Machine Learning

Second Algorithme : Support vector machine

After Creating the object support vector machine with *SVM()*
The results of accuracy is **0.8482142857142856** and using score methode we have **0.8333333333333334**

The Algorithmes **K Nearest Neighbors KNN &**
Decision Tree Classifier gives the same results and the similar accuracies

04 - CONCLUSION :

- There is a correlation between launch site and success rate Payload mass is also associated with the success rate, the more massive the payload, the less likely the first stage will return .
- For orbit type, SO has the least success rate while ES-L1, GEO, HEO and SSO have the highest success rate According to the yearly trend .
- There has been an increase in the success rate since 201 3 kept increasing till 2020 • With best parameter provided, decision tree classifier used in prediction yielded the highest accuracy of 84% .

Thank You

My Github For this project : https://github.com/AymaneK24/SpaceX