SpaceX Project

Predicting The success of Rocket Launches

Author : Anas Zaggar
as: Azero
Email me ,

Date of Publish:
5/ 6 / 2023

Table of Contents:

Executive Summary

Companies are making space travel affordable for everyone,

SpaceX sends rockets to I.S.S ,starlink for sending manned missions to space. What is interesting about new SpaceX's Rocket

(Falcon9) that is can reuse its first stage

(a part of the rocket itself) and that make these rockets cheap compared to others because it reuses their first stage.

Our goal is to predict if the first stage

will land successfully or not depending on many features about the rocket like Payload mass , Launch Site , Orbit and many more.

.

Starlin

Introductory Section

SpaceX have used many rockets and it needs to know what features to consider the most when making a rocket to get the successful land of the first stage, Speaking about the first stage , it makes the cost of the rocket start from 68 million dollars instead of 168 million dollars for other rockets like Falcon1 and many other by reusing the first stage and saving its components.

The problem is there are many failed launches and so much money is spent on this project.

Our job is minimize the lost of all this sources and make some observations to make the perfect rocket with nearly 100% success rate of landing successfully.

We will view our datasets and make some operations on them and then see visualizations to state some problems, Filtering some unique records , make use of maps and see the relation between launchsites.

Relevant Resources if you want :

1 – Data Base showing Flight Number with some features about the rockets

2 – Data Base with other Features

3 – SpaceX Website

Questions:

1 – How can We deal with data to get observations and ideas from old launches ?

2 – Is there a Launch Site better than any other one ?

3 – How much is the Impact of each feature ?

4 – Can We predict some old launches successfully so we can use the model for future launches ?

Data Collection :

Using Get Request to get the json file dealing with sets

Remove all records with Nan Values and more than one booster Version (cores , payloads)

Selecting Important Features from each column like :
Payload : [orbit , mass]
Launchpad :
[Lat,long,LaunchSite]
Cores = [NumberFlight , Gridfine , block of the core]

Filtering the Set to only Include Falcon 9 "our target"

Relisting The Flight Number

Data Wrangling

1. We search for missing Values and the get the percentage of missing values

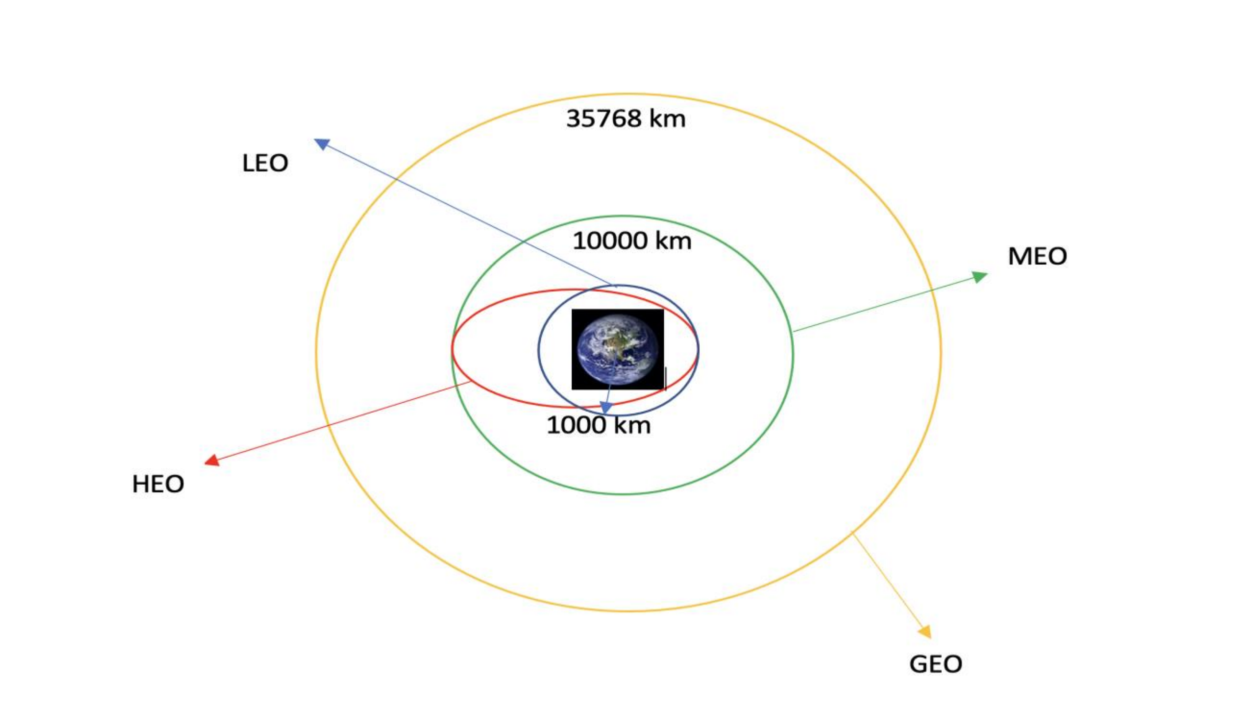
```
...  FlightNumber       0.000
     Date               0.000
     BoosterVersion     0.000
     PayloadMass        0.000
     Orbit              0.000
     LaunchSite         0.000
     Outcome            0.000
     Flights            0.000
     GridFins           0.000
     Reused             0.000
     Legs               0.000
     LandingPad        40.625
     Block              0.000
     ReusedCount        0.000
     Serial             0.000
     Longitude          0.000
     Latitude           0.000
     dtype: float64
```

2. Identify which columns are numeracle or categorical

```
FlightNumber        int64
Date               object
BoosterVersion     object
PayloadMass       float64
Orbit              object
LaunchSite         object
Outcome            object
Flights             int64
GridFins             bool
Reused               bool
Legs                 bool
LandingPad         object
Block             float64
ReusedCount         int64
Serial             object
Longitude         float64
Latitude          float64
```

3. Each Launch aims to a specific orbit

So we need to see the occurrence of each orbit



Orbit	count
GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
ES-L1	1
HEO	1
SO	1
GEO	1

4 . There are many ways of Failing or launching successfully but we only need to know if it was successful or not so we will turn every failure to one cluster so do successful launching.

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

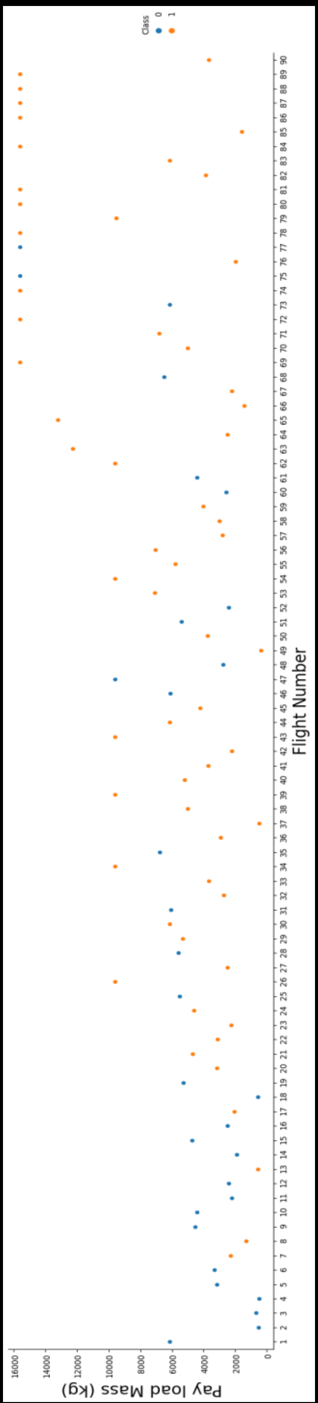
EDA (Exploratory data analysis)

1. Data visualization

- *Now we will move on to some visualization to make some observations and details out of plots.*

1. Flight Number VS Payload Mass :

We see that as the flight number Increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

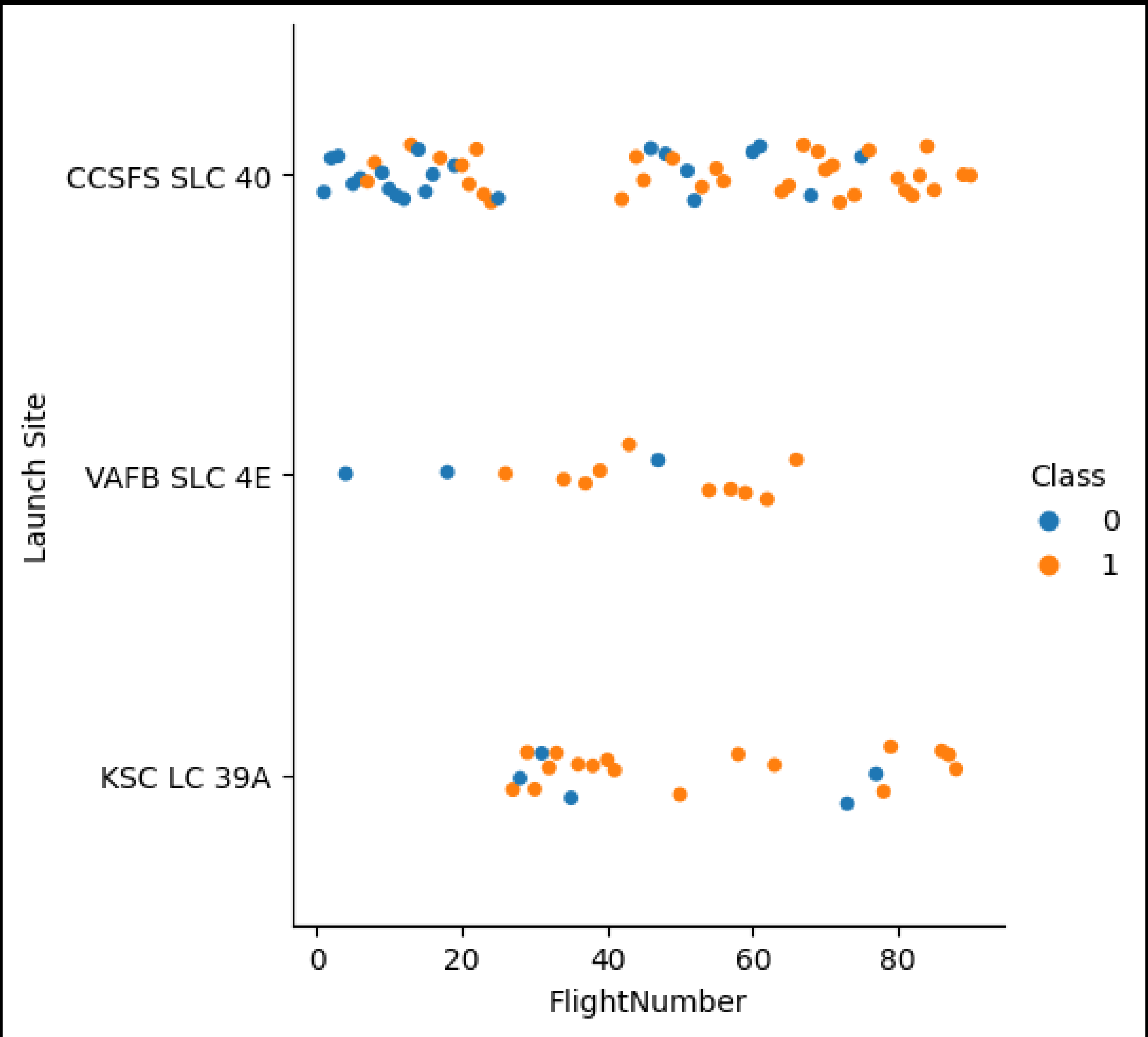


Keep in mind that the color point on this plot
Is represnting the class of each landing.
(if it fails or not)

2. Flight Number VS Launch site

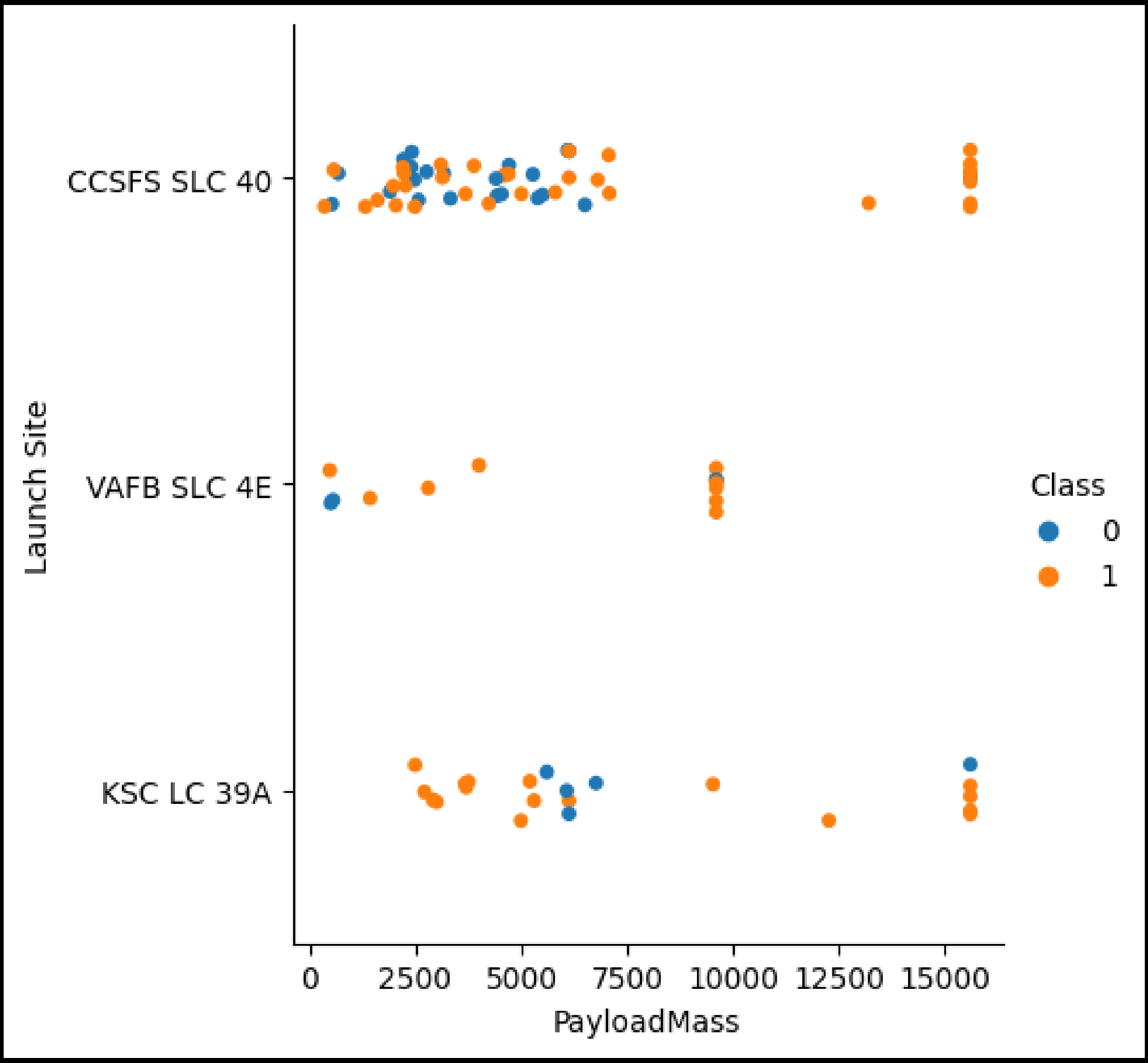
- *We see that different launch Sites have different success rates,*
- *CCAFC LC-40 has a success rate of 60%,while KSC LC-39A,VAFB SLC 4E*

  Has a success rate of 77%.



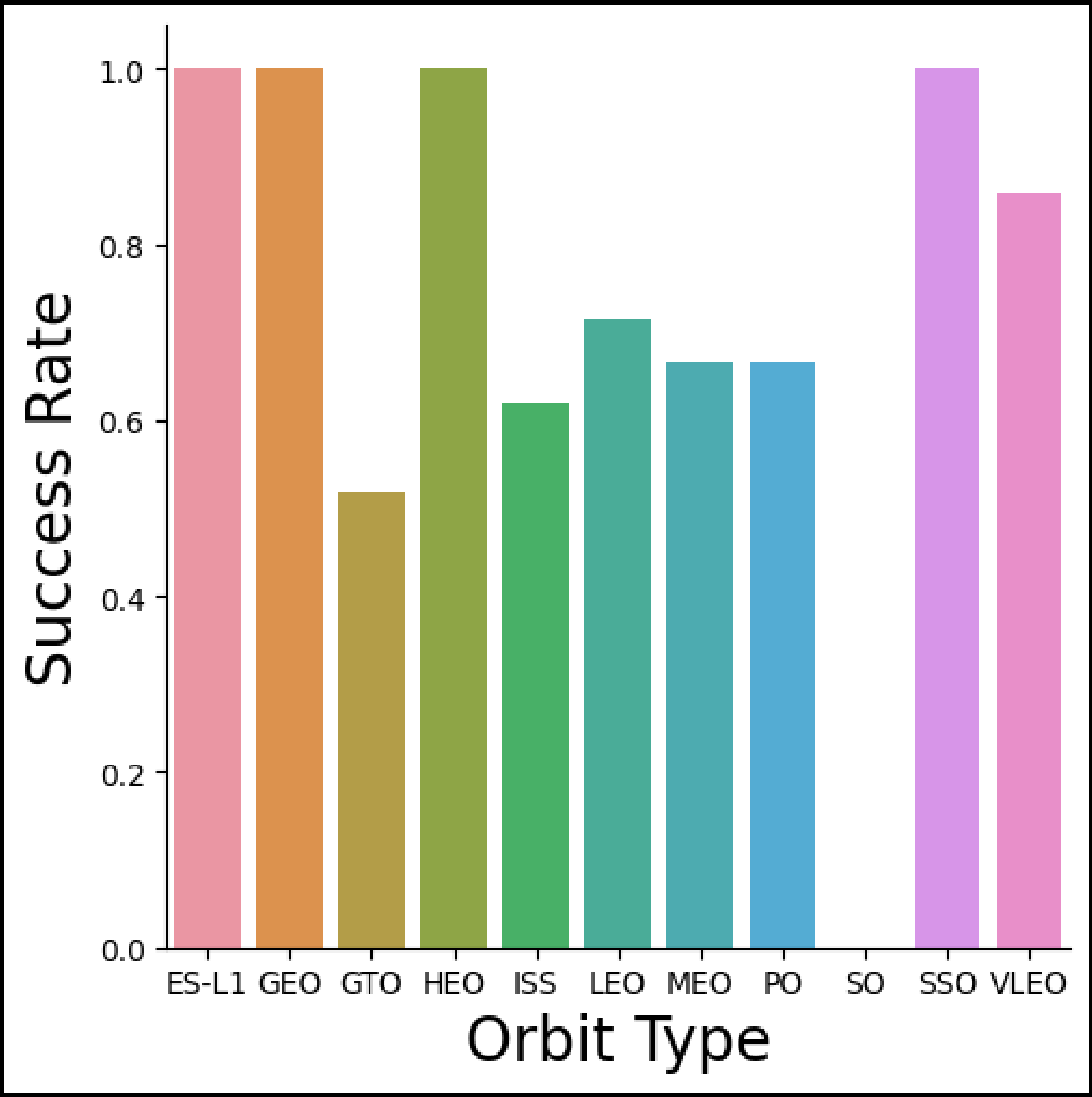
3. Payload Mass VS Launch Site

- We can easily see that higher payload masses are more likely to succeed and for VAFB-SLC there are no rockets launched higher than 10000kg.



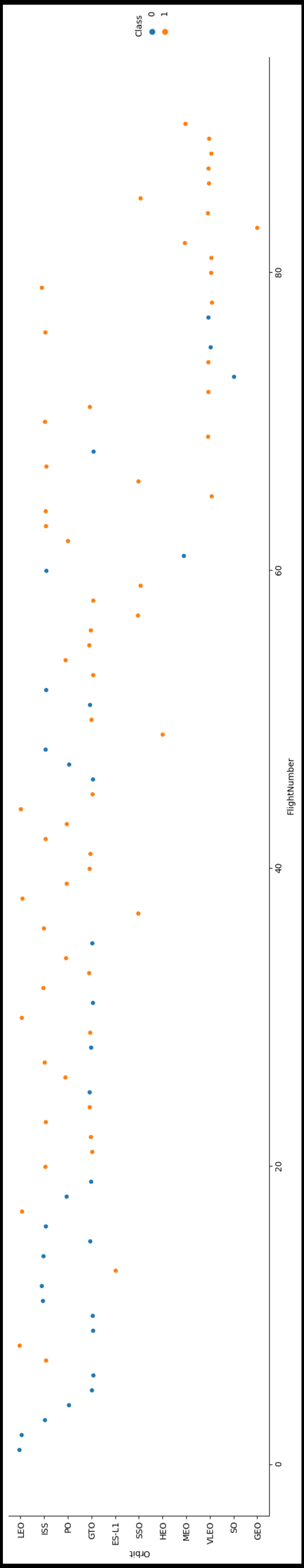
4. Success rate of each orbit

As we can clearly see that ES-L1,GEO,HEO,SSO and VLEO have high success rates where GTO,ISS Are the lowest and SO have no success rate at all because there are no flight from it



5. Flight Number VS Orbit Type

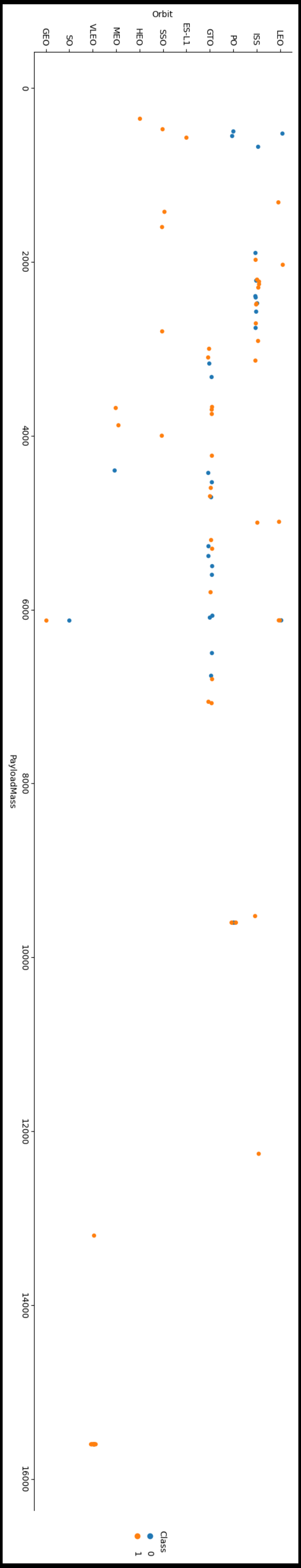
- In LEO orbit the success Rate can appear clearly
- there seem to be no Relationship between GTO orbit and the Success rate



6. Payload Mass VS Orbit type

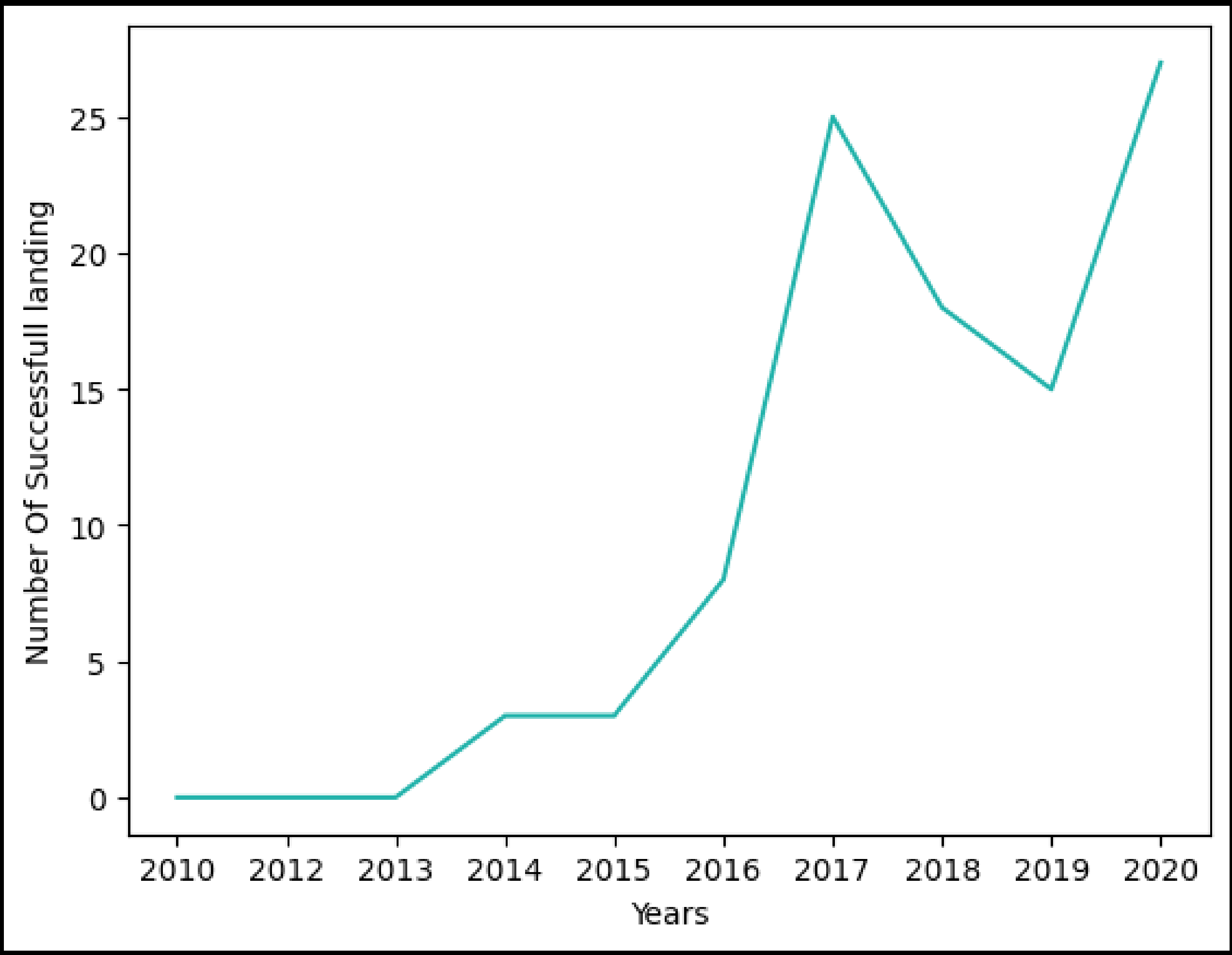
- With Heavy payloads the successful landing or positive landing are more from polar,LEO, And ISS

-but GTO have the Two types of landings (positive and negative)



7.Successful landing each year

- We can clearly see that the number of successful landing is increasing each year in general and year 2020 and 2017 have the highest number of successful landing whereas from 2010 to 2013 there are no successful landing !.



2.EDA With SQL

- There will be some unique distributions for the data with specific attributes.

1. Launch Sites

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

2. Some records from

CCAFS LC-40 , SLC-40

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

3. Total Payload Mass carried by NASA

45596

4. Average Payload Mass carried by F9 v1.1

2534

5. First Successful Landing

22-12-2015

6.Boosters that have success and Mass between 4000 and 6000

Done.

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

7 Failing and Success Landings

Fail : 10

Success : 61

8. Boosters Carrying Maximum Payload Mass

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

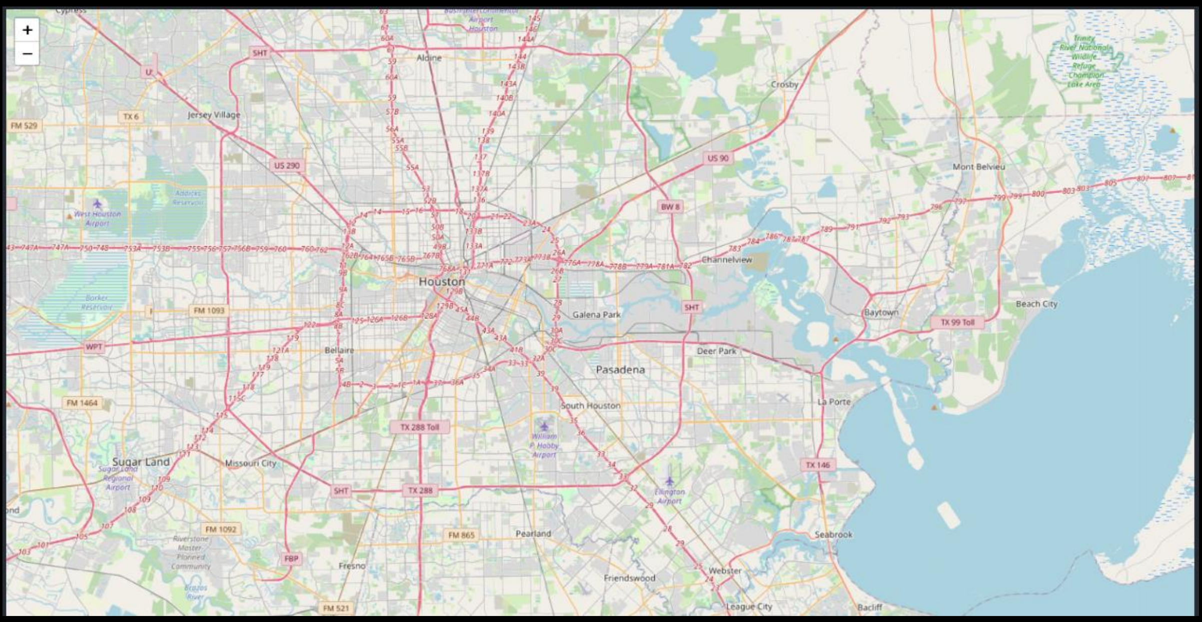
9. Landings in by month

Month	Date	Landing_Outcome	Booster_Version	Launch_Site
10	01/10/2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
11	02/11/2015	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
02	03/02/2015	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	14/04/2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	27/04/2015	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	28/06/2015	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	22/12/2015	Success (ground pad)	F9 FT B1019	CCAFS LC-40

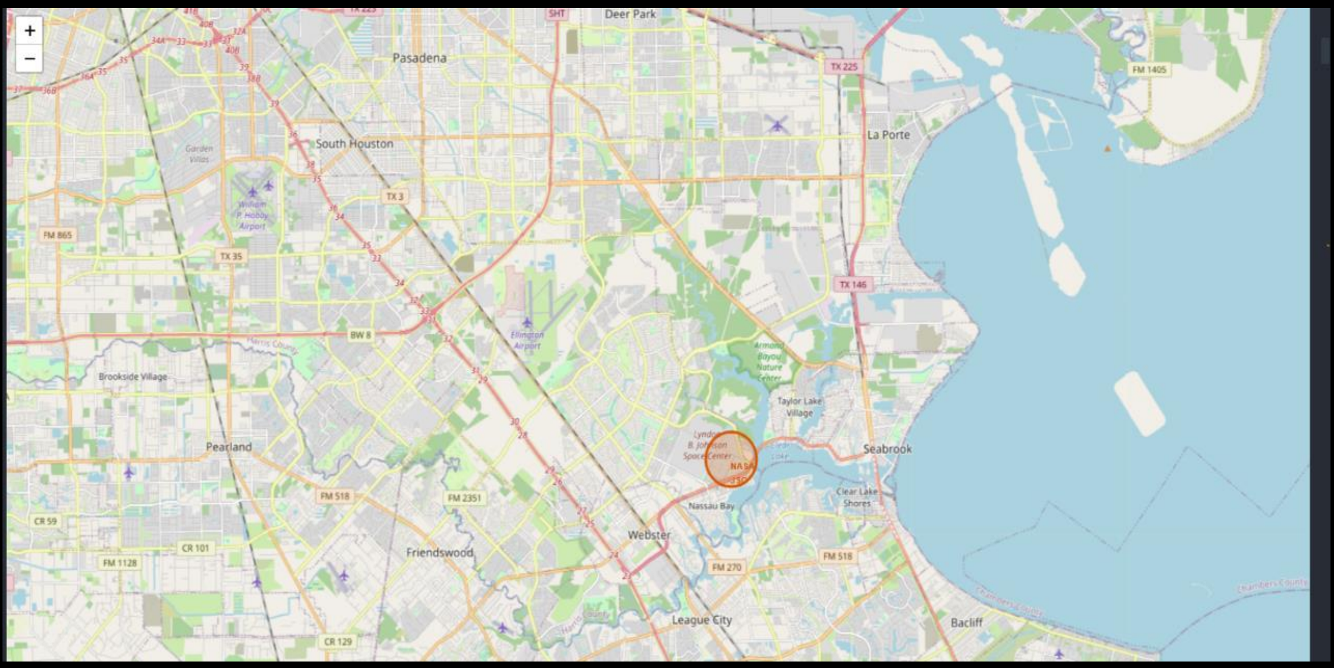
Folium For Maps

- Using Folium is a nice way to visualize some sites and location Sites

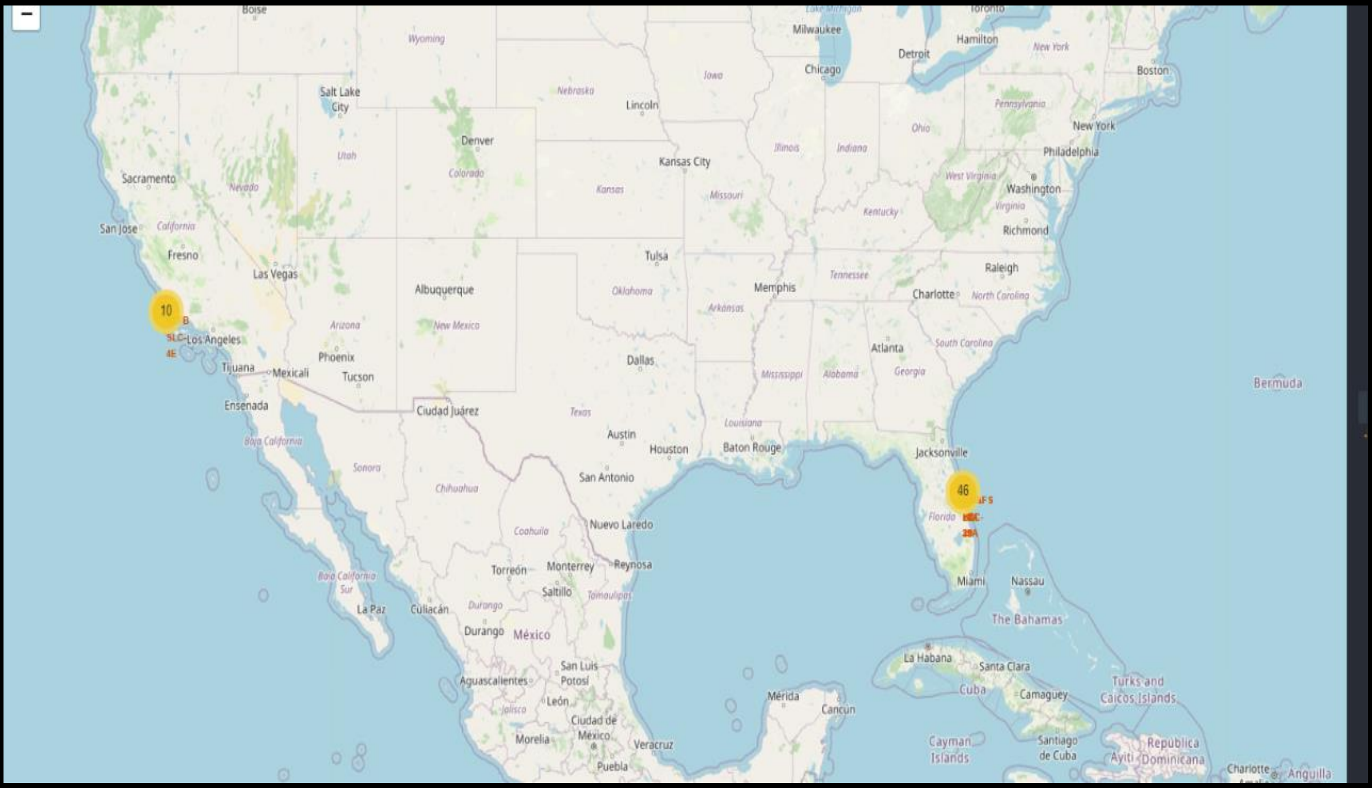
-Nasa Location



-Closer Look

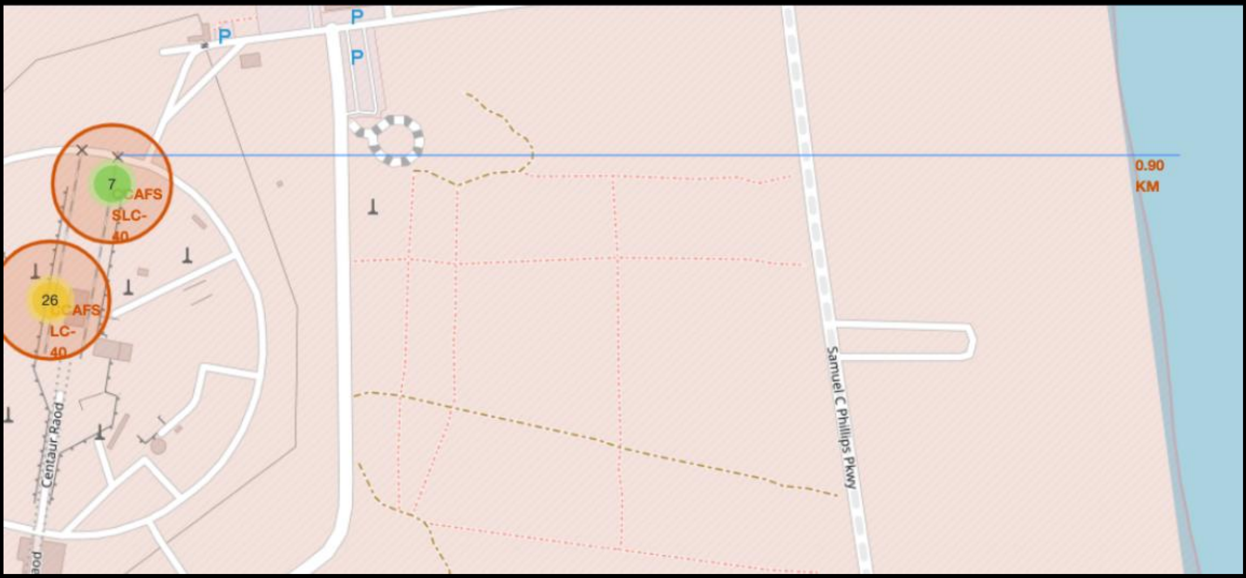


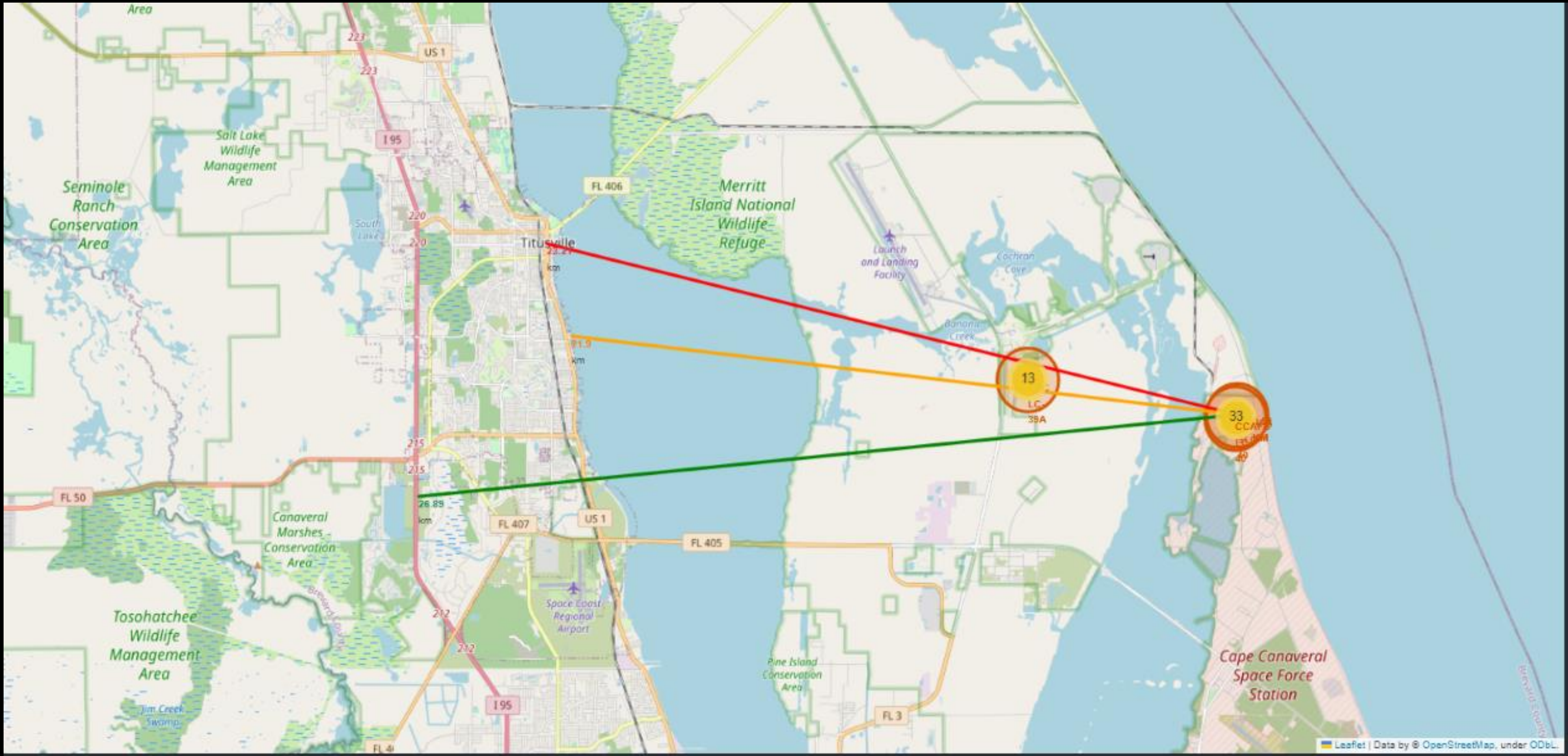
-Launch Sites with Number of Landings



As we Can clearly see:

- Launch Sites are near the cost
- Number of Landing on each Site is represented
- Launch Sites and its approximates




Seminole Ranch Conservation Area
Salt Lake Wildlife Management Area
Merritt Island National Wildlife Refuge
Launch and Landing Facility
Titusville
FL 406
US 1
I 95
FL 50
FL 407
FL 405
FL 3
Canaveral Marshes Conservation Area
Tosohatchee Wildlife Management Area
Space Coast Regional Airport
Pine Island Conservation Area
Cape Canaveral Space Force Station
13
33
39A
26.89 km
Leaflet | Data by © OpenStreetMap, under ODbL.

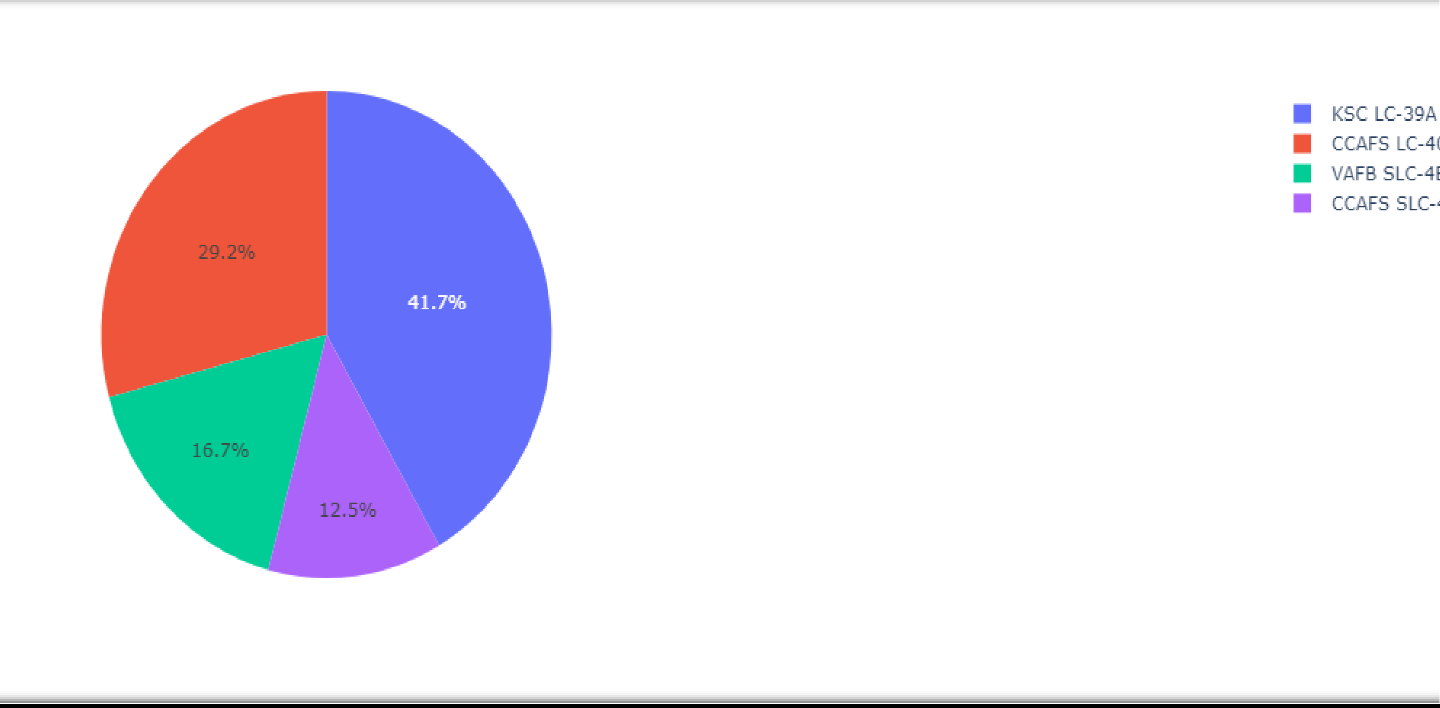
Explanation:

• From the visual analysis of the launch site KSC LC-39A we can clearly see that it is: - relative close to railway (15.23 km) - relative close to highway (20.28 km) - relative close to coastline (14.99 km)

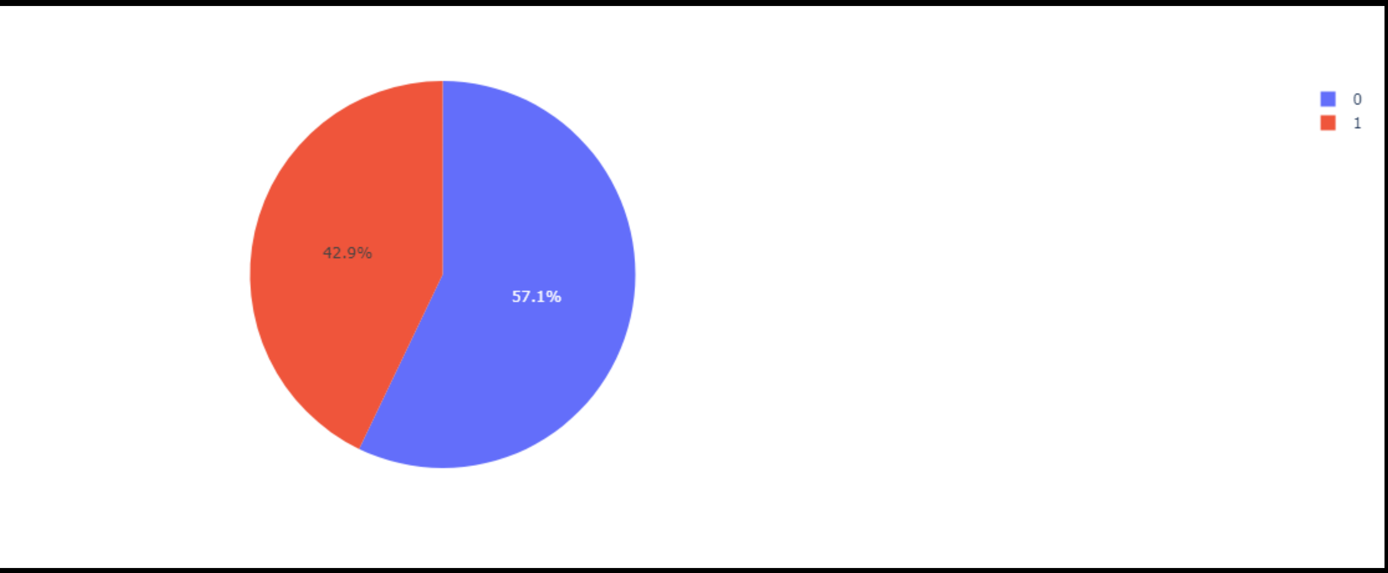
• Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km). • Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially

Interactive Visualizations with Dash

I. Successful landing coming from each launch site

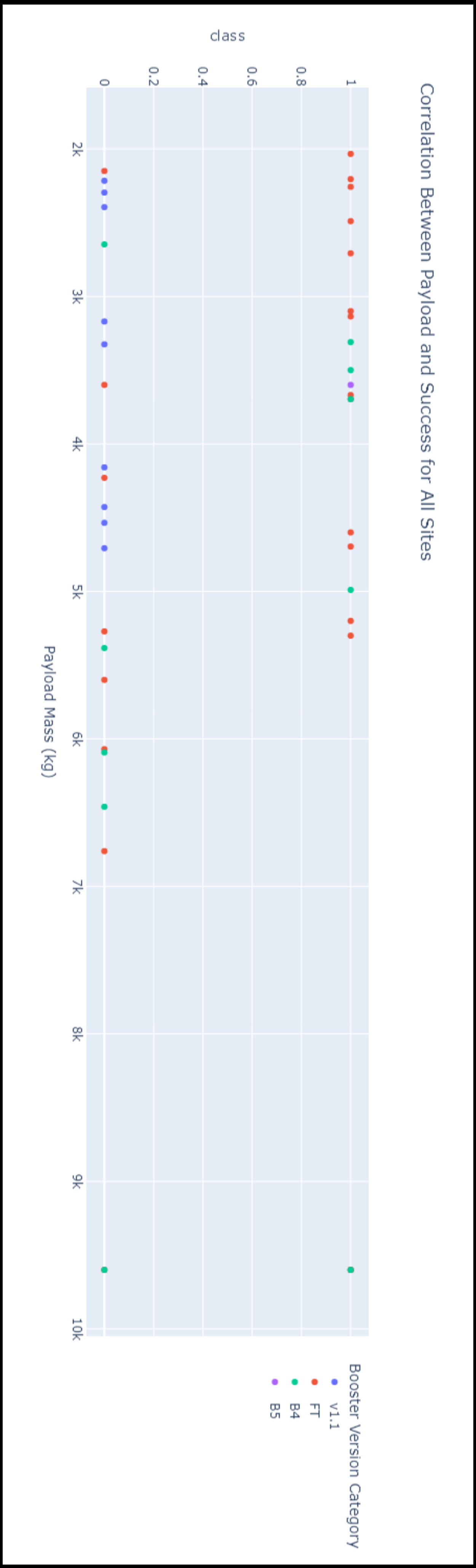


II. Landings from specific orbit (CCAFS SLC 40)

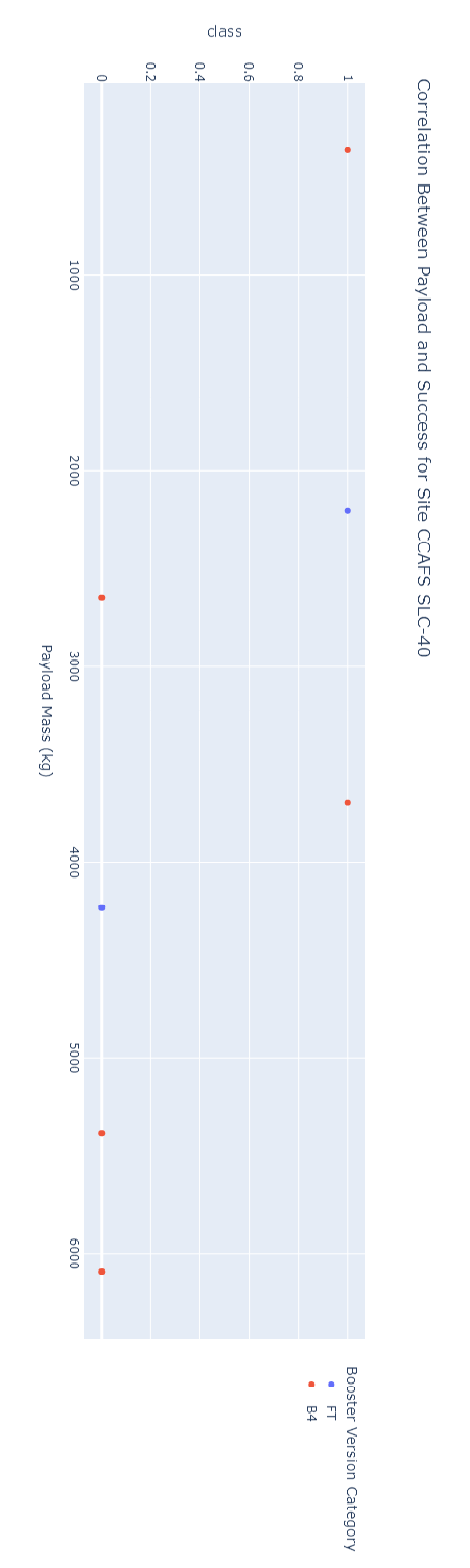


for

III. Correlation between PayloadMass and Success for all sites with ranges for the mass



IV. Payload Mass and CCAFS SLC40 for all masses

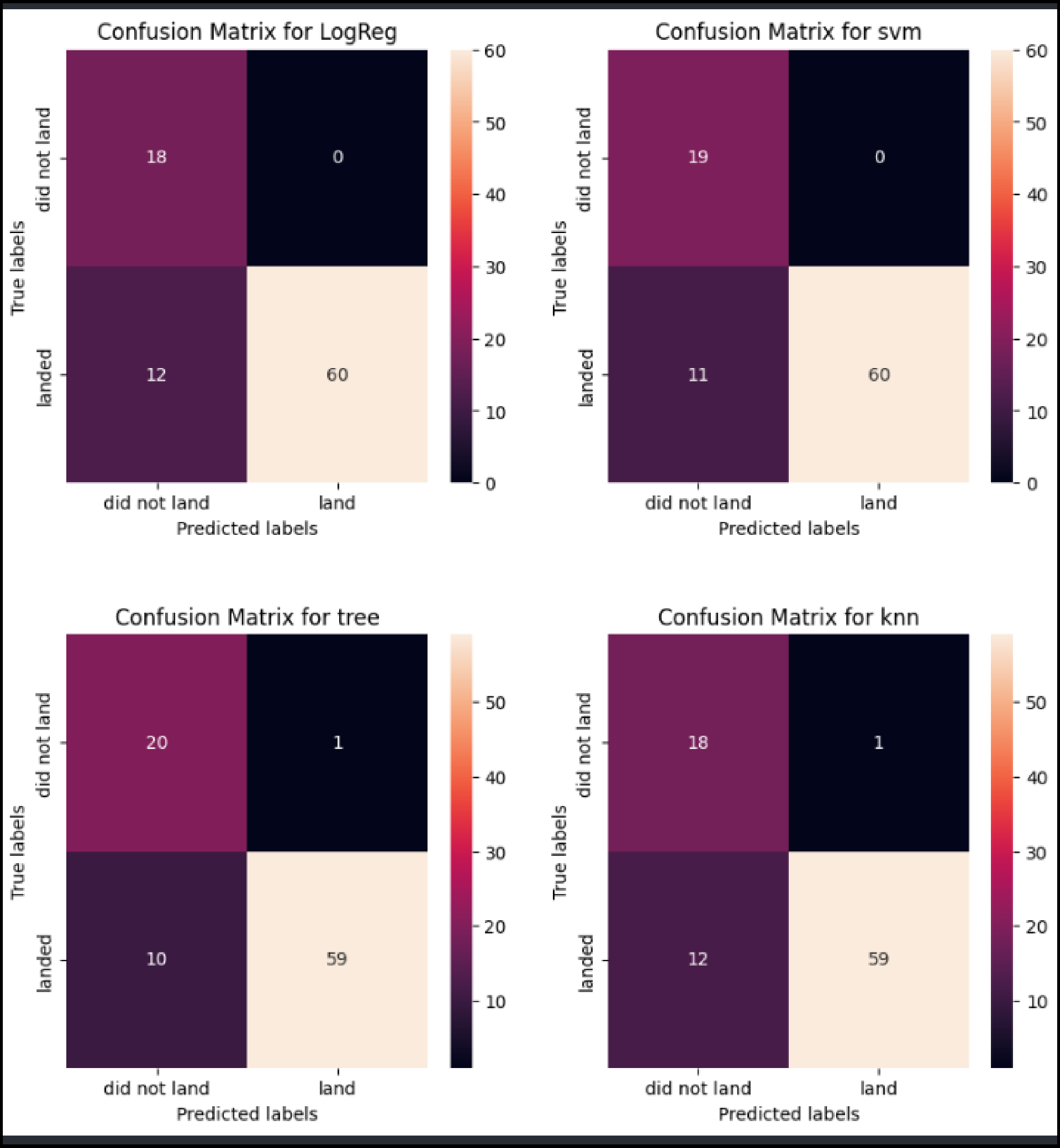


-For more information you can visit the dash page from the source code

Predictive Analysis

- Depending on the sources we put it above , we will use the last landing reports to predict the future launches.
- There are many machine learning algorithms like (Logistic Regression , Decision Tree classifier, Support Vector Machine, KNearestNeighbors)
- We will use train/test split in our data to train the model then test it to see how well it is and then plot the confusion matrix
- After using Grid Search for giving the best fitting parameters for each algorithm and without getting into details of the process.
- We used many ways to know how well the model for each algorithm like (f1 scores , log loss , r2 score , jaccord index)

	f1 scores	log_loss	r2 score	jaccord_index
lr	0.909091	4.805820	0.40	0.833333
svm	0.916031	4.405335	0.45	0.845070
tree	0.914729	4.405335	0.45	0.842857
knn	0.900763	5.206305	0.35	0.819444



As we can see :

- tree classifier have the highest number of true negatives and some scores as well as svm but tree classifier a little bit better

Conclusion

- Decision Tree Model is the best algorithm for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate

Appendix

- Special thanks to IBM environment that gives me these skills to be able to write this project.

[Instructors]

[IBM]

[Coursera]