CapstoneProject

Final Data Report

Matthew Robert Gluski Balanda

09/01/2022

Table of Contents

EXECUTIVE SUMMARY

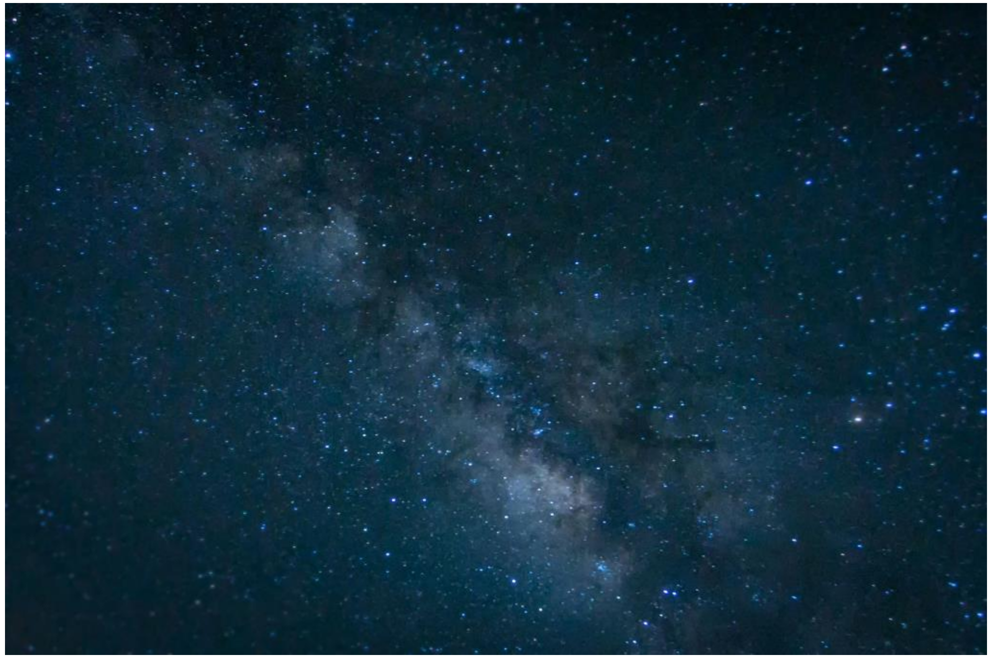
This is the final project of the IBM Certificate course. Is a capstone where we use everything we learned from Data Collection to Machine Learning and prediction.

In this presentation we will resume the obejctives of the project and the results when tackling it. By using the methodologies told in the course, it was created by combining and organizing each task with the main goal to tell a story .

INTRODUCTION

Imagine yourself as inspiring entrepenor with the goal to reach the stars. Space always fascinated people so naturally pursued it. Space Y has that same goal. But we are newcomers to the business of spaceships. Our competition has a headfirst and are way adept in that area.

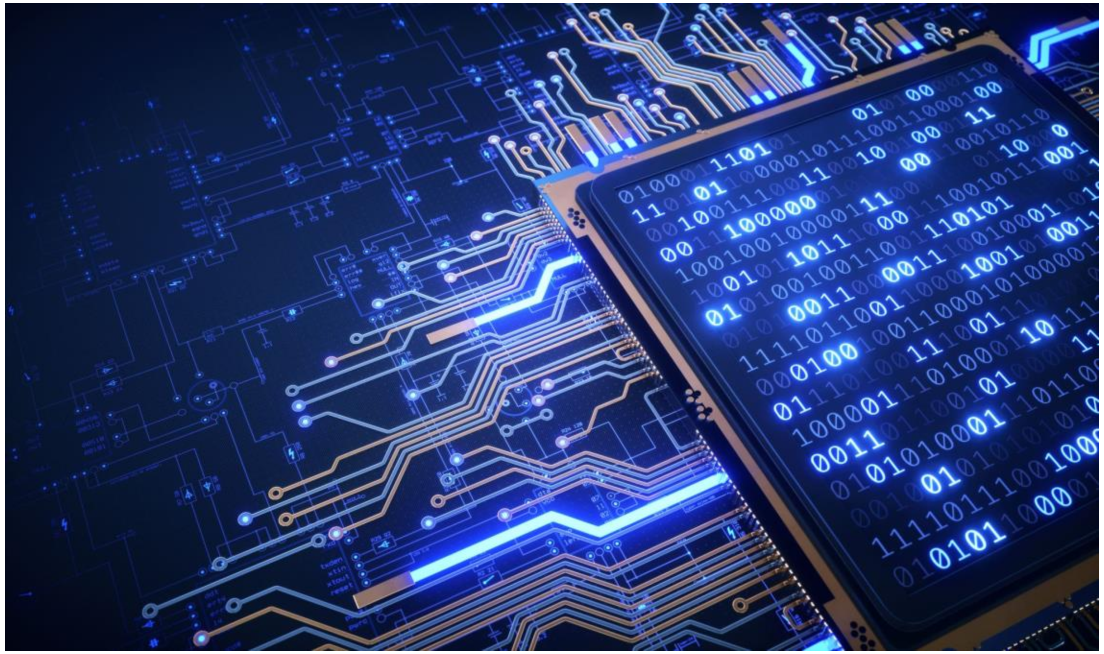
So what should we do first to caught up with the competition?



INTRODUCTION

We use the competition themselves.

More specifically we use their own data to benefit us.



INTRODUCTION

Our competition in question is Space X and they have years of experience and that inclides trial and error. By exploring their past data we can make massive leap to catching up, or even surpassing them.

First we need to collect their public information for setup so that we can estabilish moderm rockets from the start and by using their past data we can avoid any problems already solved.

Our main goal is caught up to their standards and even predict future data, which could potentially surpass them. But for now we going to colelct data and form our dataframe.

METHODOLOGY

Data Set: Collection and Wrangling

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.

Witht that we will collect data from their past launches

"https://api.spacexdata.com/v4/launches/past"

For the Data Collection and Wrangling we used the Pythn language and its libraries like pandas that it had its own methods for easier use of the data. Pandas library allowed us to make easy tables so we can store our data into.

We sued the BeautifullSoup method for the web scrapping process for the past data and stored in a data frame. For that we ahd to establish a connection via BeautifulSoup so we got a response and access to this data

By getting the data response we acquire ton of information that will need to organize in a data frame.

0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0
1	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage	[]	[]	[]	[5eb0
2	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Residual stage 1 thrust led to collision between stage 1 and stage 2	[]	[]	[]	[5eb0 5eb0
3	2008-09-20T00:00:00.000Z	1.221869e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	True	Ratsat was carried to orbit on the first successful orbital launch of any privately funded and developed, liquid-propelled carrier rocket, the SpaceX Falcon 1	[]	[]	[]	[5eb0
4	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	True	None	[]	[]	[]	[5eb0

We have a data frame but still remains messy. With that we cleane it up byt making a new table extraction the necessary information while providing more corcive names instead of using serial numbers.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False

GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Now we have a legible dataframe to work with. Now we need to filter it to only include Falcon 9 Launches as the more moderm rockets, our goal to replicate.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	Lanc
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None
...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb

Final touches was to deal with missing values, specifically on pay Load Mass column, so we simply filled it with the mean value. Landing Pad we didnt touched because been missing implies the landing pad wasn't used, meaning relevant information.

METHODOLOGY

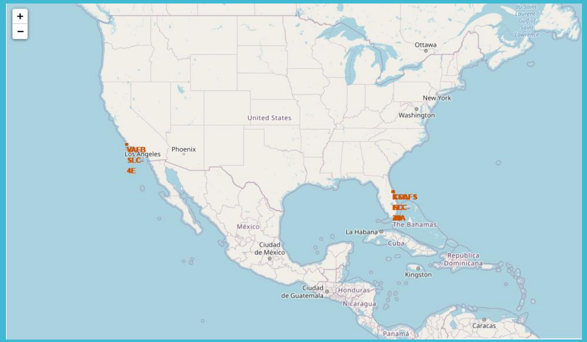
EDA and interactive visual analytics

```
#df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv")
```

With the data wrangled we saved in CSV file and import into DB2 database for SQl tasks. We used Python and pandas to connect this database to the notebook for coding

We also imported into the notebook and used the plotly library to create multiple graphs for easier analysis.

We also imported into the notebook and used the plotly librar y to create multiple graphs for easier analysis. Including Scatter plot and Bar charts.



Another library for analysis was Folium, where we marked on the real life map the launch sites and relevant landscapes

METHODOLOGY

Predictive Analysis

For this method we used these python libraries:

Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis.

Numpy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

Matplotlib
Matplotlib is a plotting library for python and pyplot gives us a MatLab like plotting framework. We will use this in our plotter function to plot data.
import **.pyplot as plt**

Seaborn is a Python
data visualization library based on matplotlib. It provides a high-level interface
for drawing attractive and informative statistical graphics

Sklearn

This library and its other functions will allow us to:

- Preprocess to standarsize our data
- Split our data into training and testing data
- Test parameters of classification algorithms and find the best one. Using GridSearchCV

We using these Classifier models:

- Logistic Regression
- DecisionTree
- K Nearest Neighbors
-
- Support Vector Machine

RESULT

EDA with SQL

SpaceX has gained worldwide attention for a series of historic milestones.

It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars wheras 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 SpaceX for a rocket launch.

This dataset includes a record for each payload carried during a SpaceX mission into outer space.

We imported the dataframe into an SQL database and performed some SQL tasks to see if all was on order.

Task 1
Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1
Done.
```

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

Task 2
Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/BLUDB
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Task 3
Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as Total_payload FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.app
Done.
```

total_payload
45596

Task 4
Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomai
Done.
```

average_payload
2928

Task 5
List the date when the first successful landing outcome in ground pad was achieved.

```
%sql SELECT MIN(DATE) AS Date_Success_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain
Done.
```

date_success_gp
2015-12-22

Task 6
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 PAYL
OAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/BLUDB
Done.
```

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

Task 7
List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS Total FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.dat
Done.
```

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Task 8
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249
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

Task 9
List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND DATE LIKE '2015%';
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/BLUDB
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Task 10
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 as Type, COUNT(LANDING__OUTCOME) as Total FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' A
ND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY Total DESC;
```

```
 * ibm_db_sa://cyy93976:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/BLUDB
Done.
```

TYPE	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

RESULT

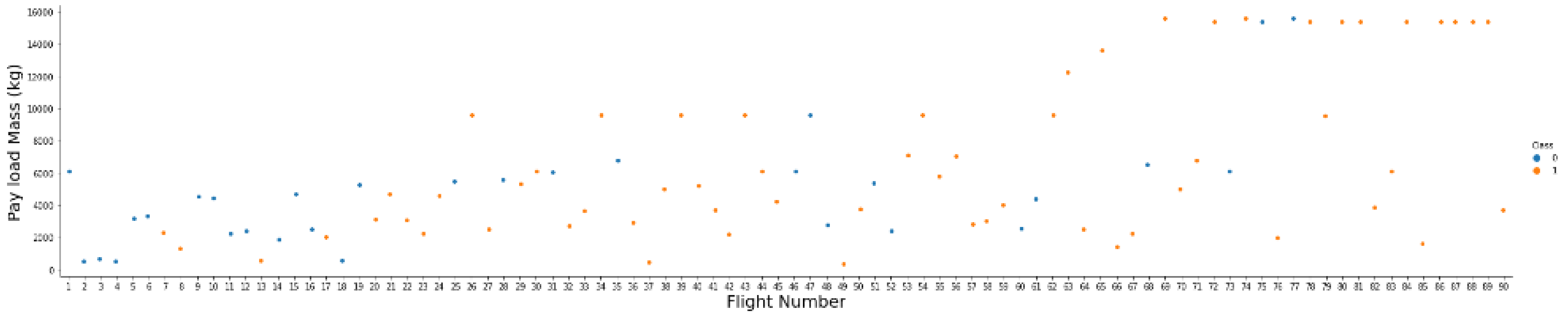
EDA with Data Visualization

Now for better understand the data, we will use graphics to better visualization. Heres the current sumary of our Dataframes in table form

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False

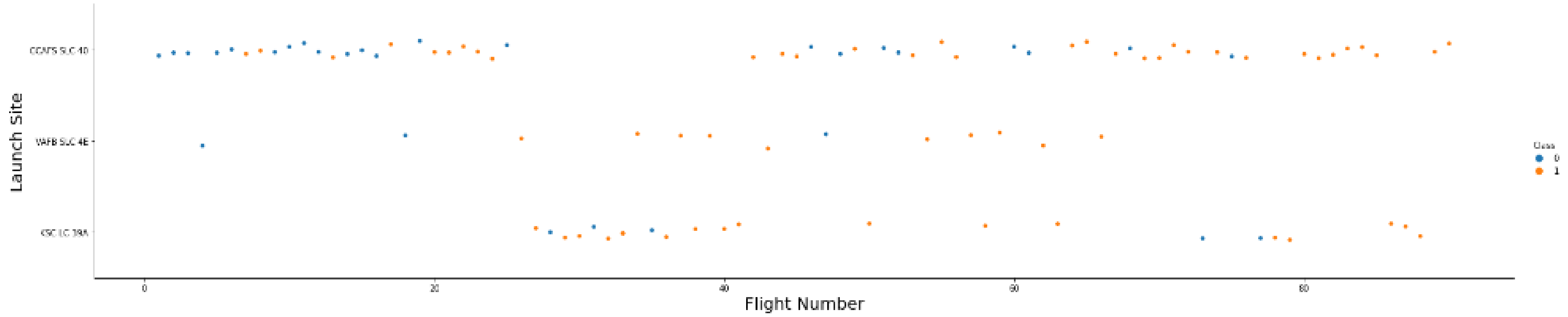
GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Now we will visualize the relationship between certain columns to analyze they are related to launch success rate.



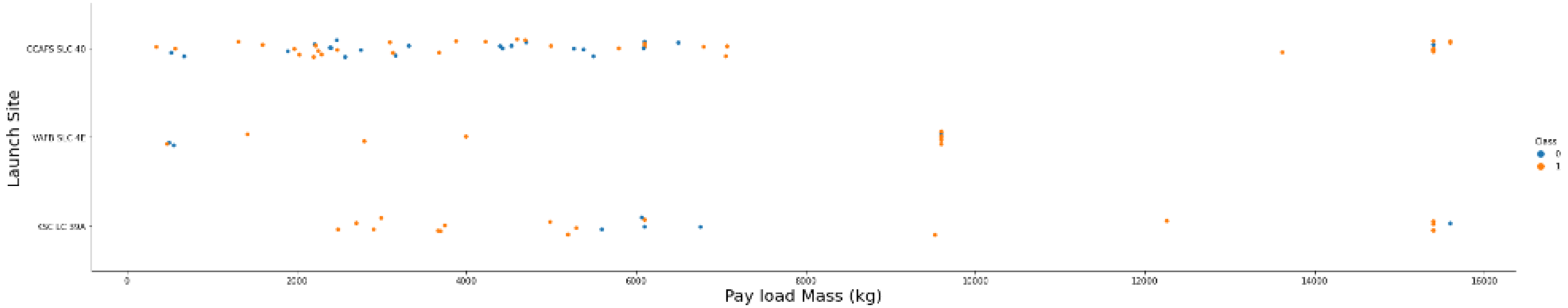
FlightNumber vs PayloadMass

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.



Flight Number vs Launch Site

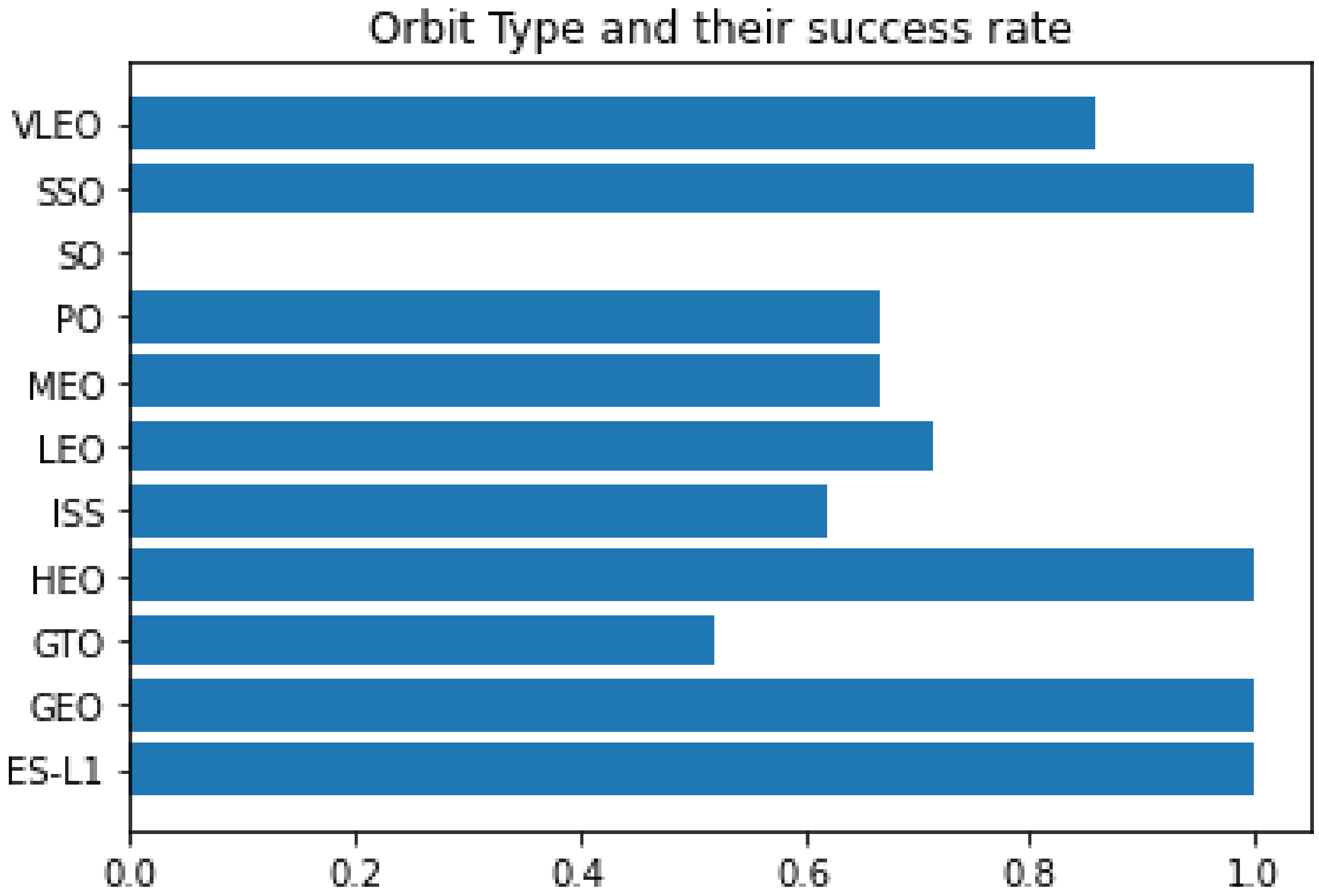
CCAFS LC-40 success rate increased with the humber of flights, while having the most attempts among them. VAFB SLC 4E site laos had a bad start despite having less flights, but it shows increase in success. KSC LC-39A site launched later so is harder to see any notable different as it remain ratehr consitant in comparison

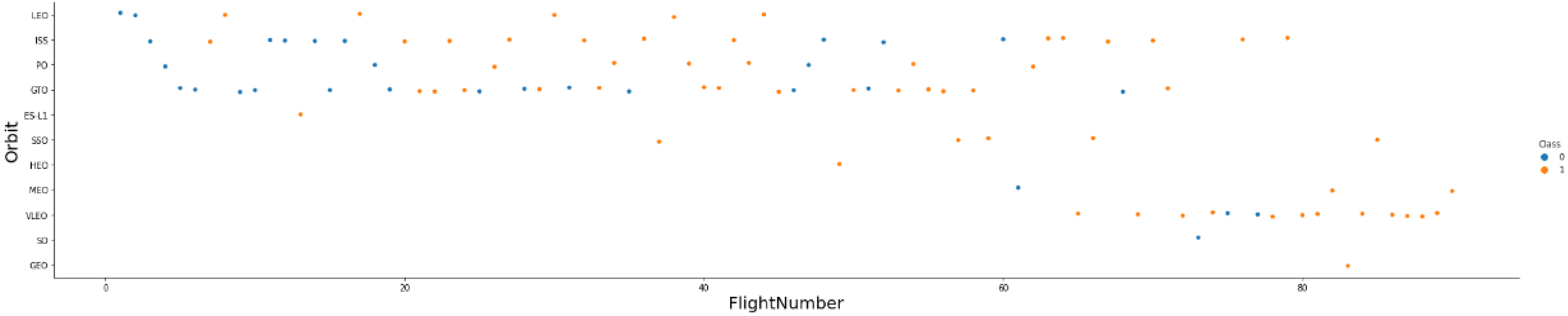


Payload vs Launch Site

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).

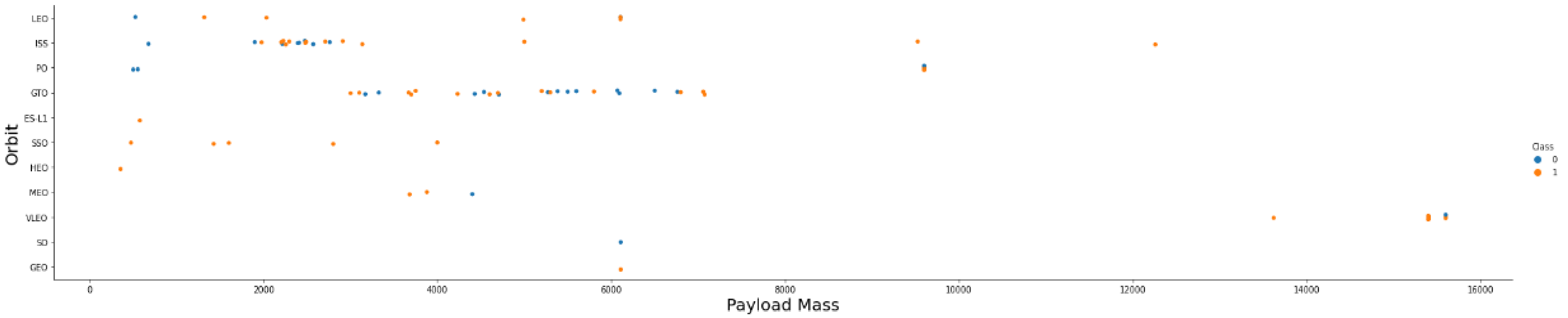
Now we are going visualize relationship between orbit and success rate. Will be a bar chart for a more straightforward interpretation.





FlightNumber vs Orbit type

You should 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.

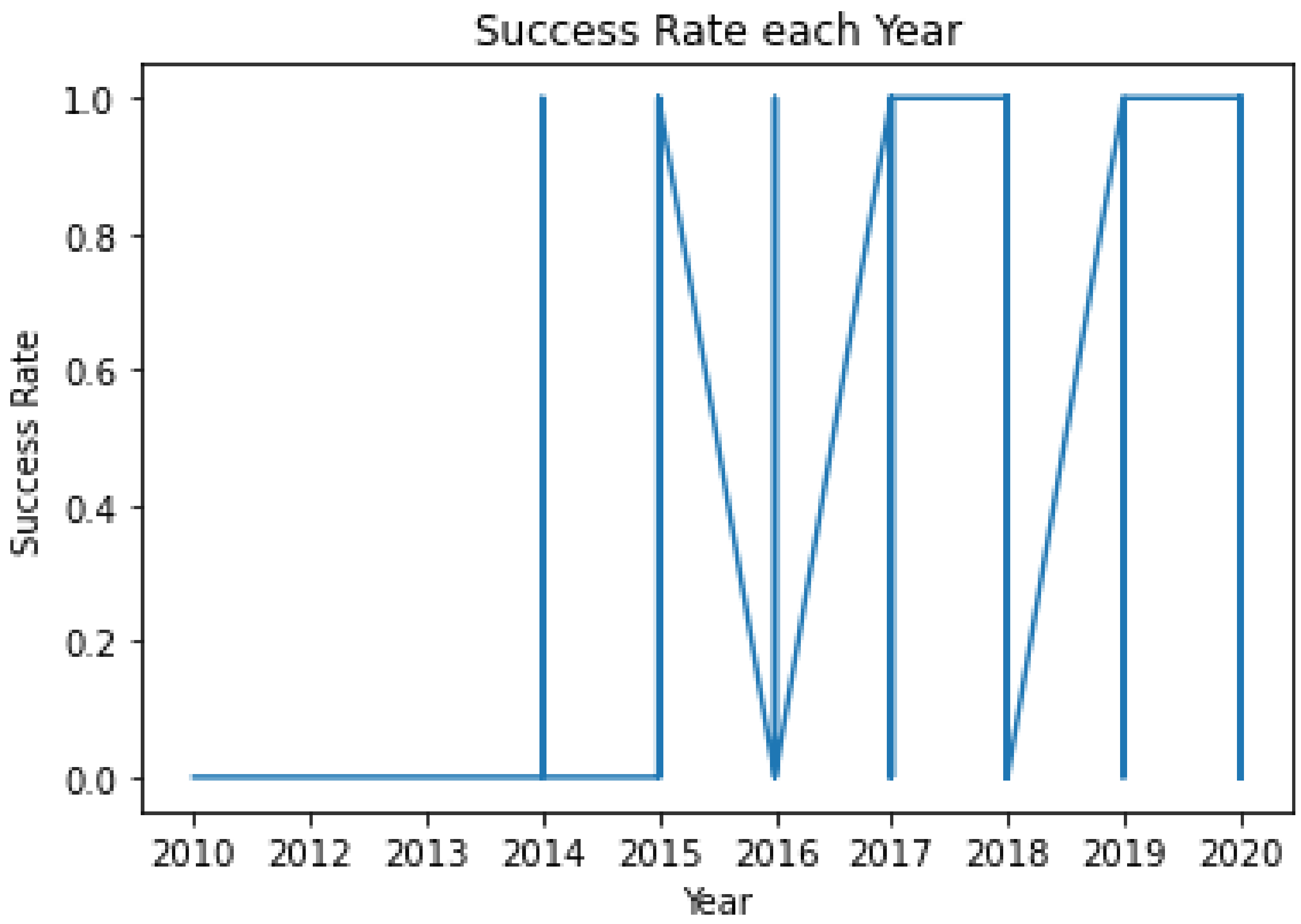


Payload vs Orbit type

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.

Now we are going visualize the success rate over the years

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



Now we have enough results and evaluation, we selected the relevant features and made a new table out of it to be used in success prediction

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007
3	4	500.000000	PO	VAFB SLC 4E	1	False	False	False	NaN	1.0	0	B1003
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004

Finishing touches by adding dummy variables and converting them to float for consistancy

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	ES-L1	GEO	...	B1048	B1049	B1050	B1051	B1054	B1056	B
0	1	6104.959412	1	False	False	False	1.0	0	0	0	...	0	0	0	0	0	0	0
1	2	525.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	0	0	0
2	3	677.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	0	0	0
3	4	500.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	0	0	0
4	5	3170.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	0	0	0

5 rows × 80 columns

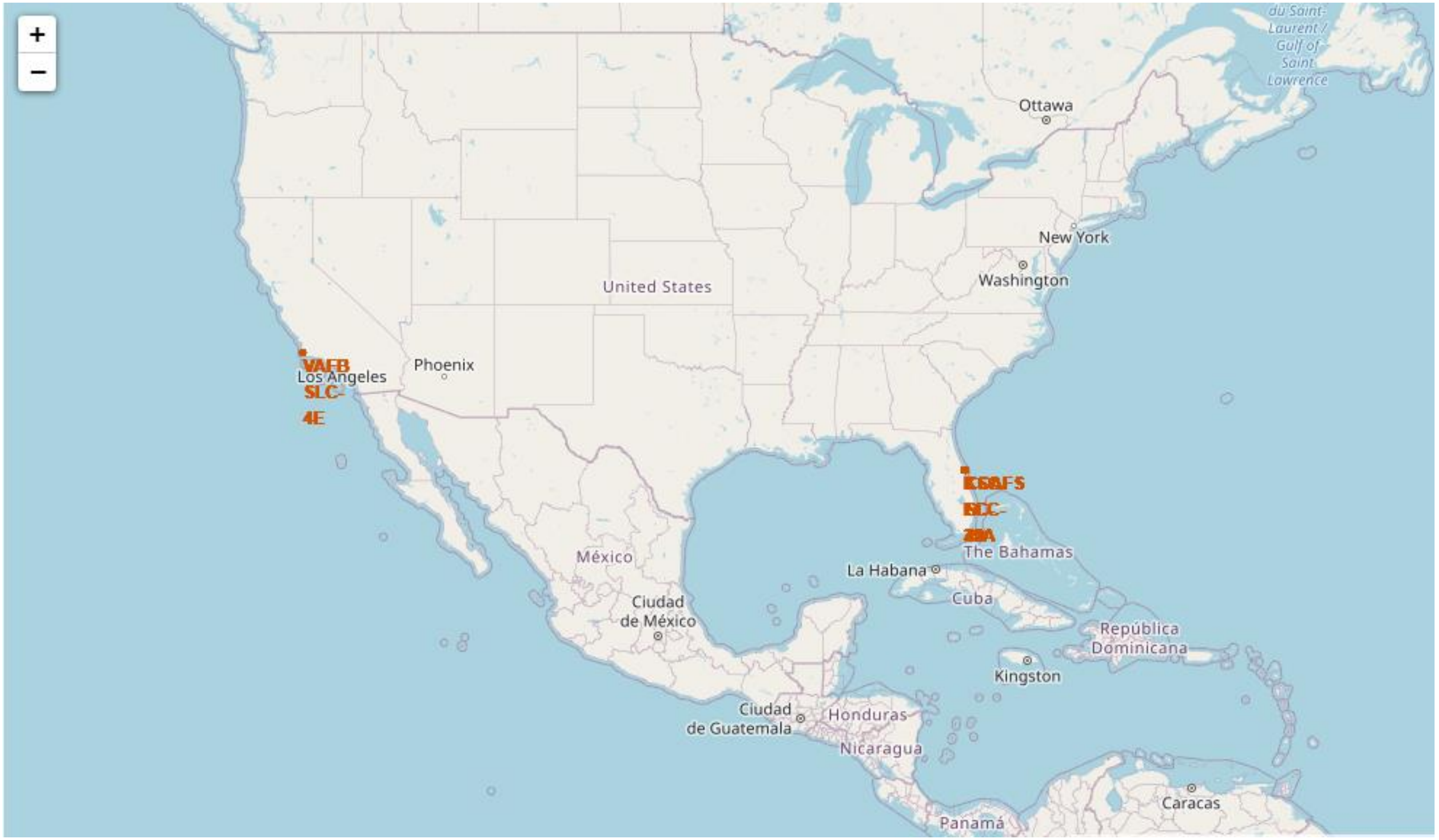
```
ES-L1     float64
GEO       float64
GTO       float64
HEO       float64
ISS       float64
            ...
B1056     float64
B1058     float64
B1059     float64
B1060     float64
B1062     float64
Length: 72, dtype: object
```

RESULT

Launch Sites Locations Analysis with Folium

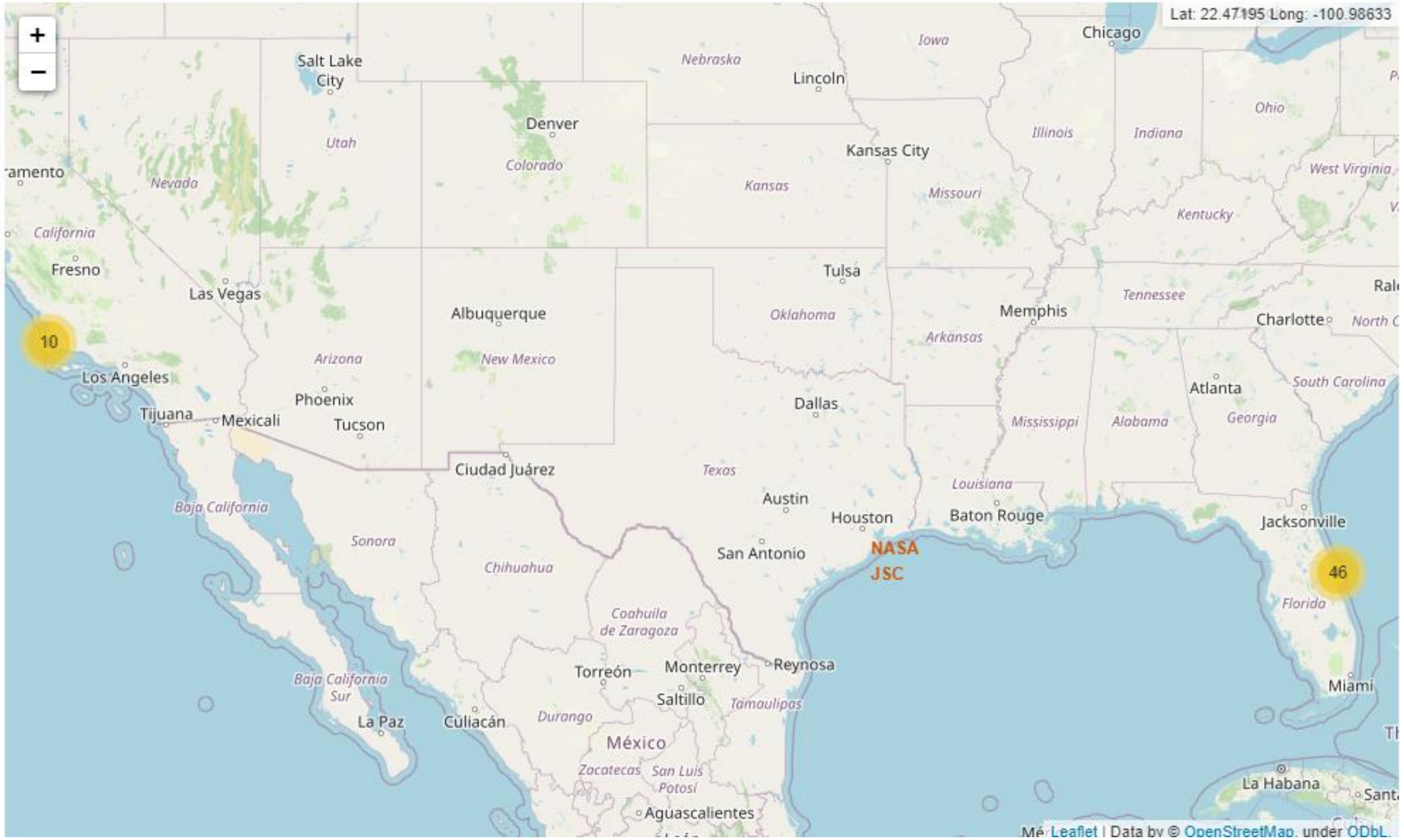
The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

We used the folium package to display the map and we marked the site locations for analysis.

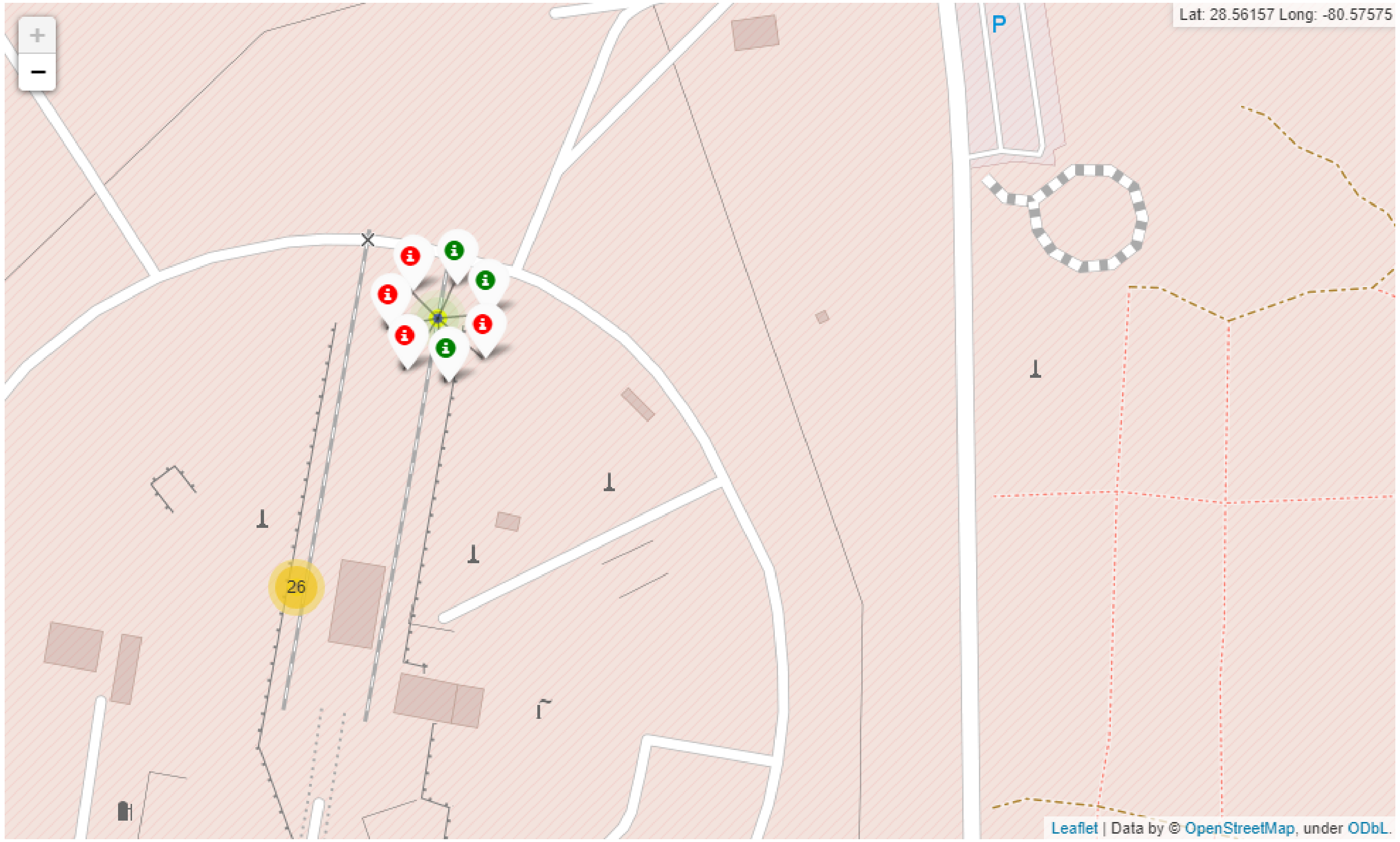


This is the general map showing the main locations of the site in US. As you can see both are enar the coast, so that might be a relevant piece of information.

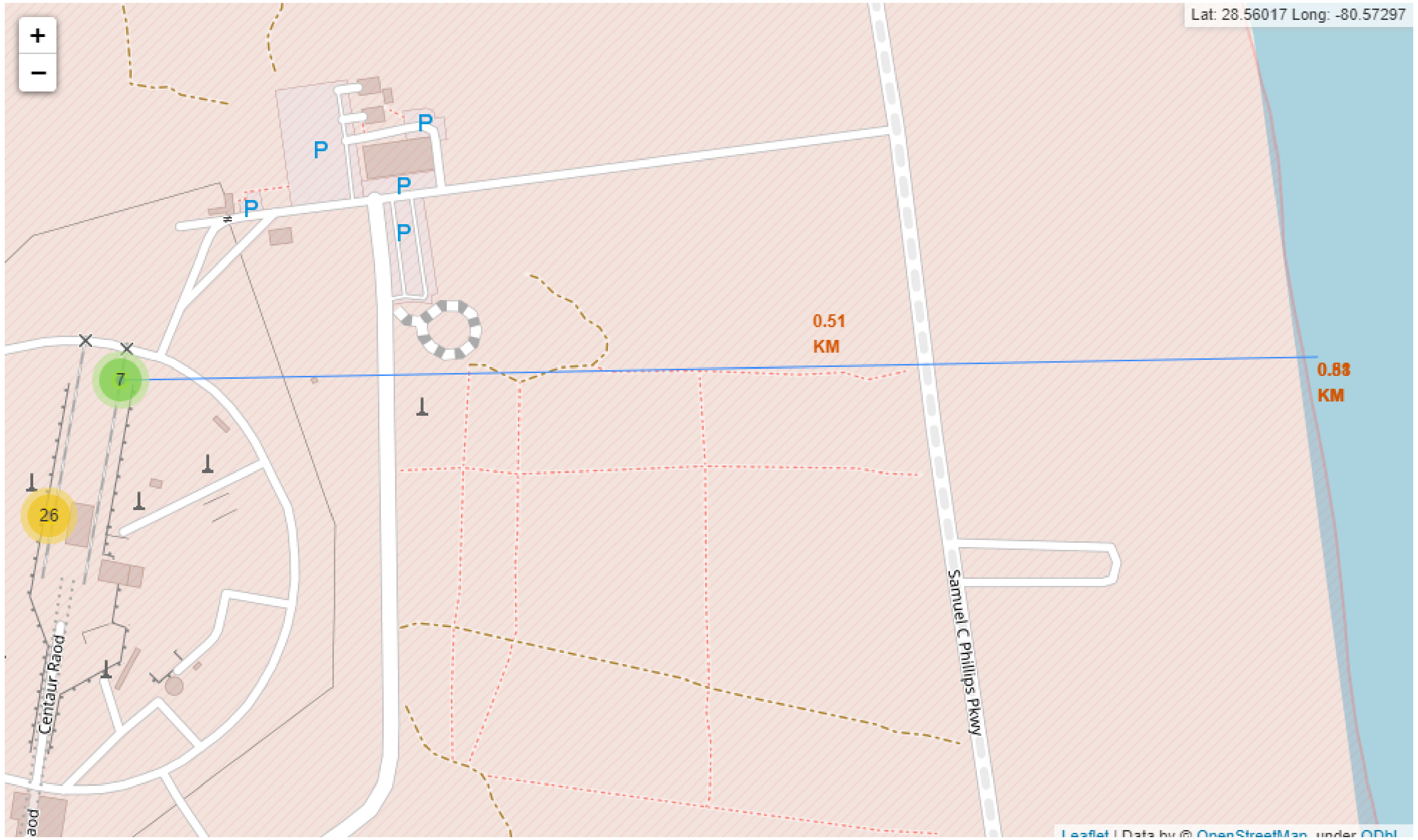
Now will we add adition features, like the coordinates using the mouse pointer and cluster markers to show you the number of sights in the area



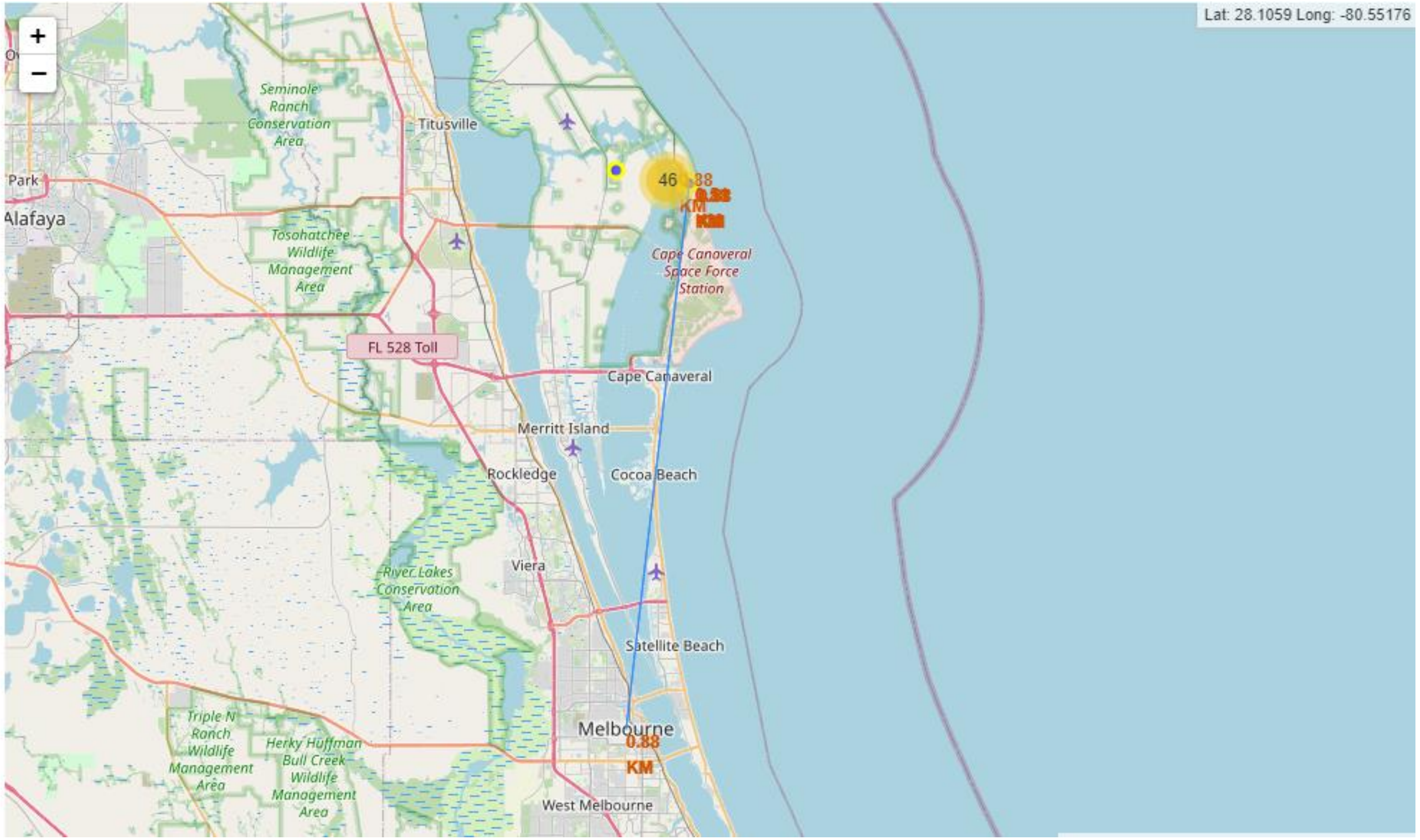
And also giving different colros to each site to signalize their success rate, with red been failure and green been success.



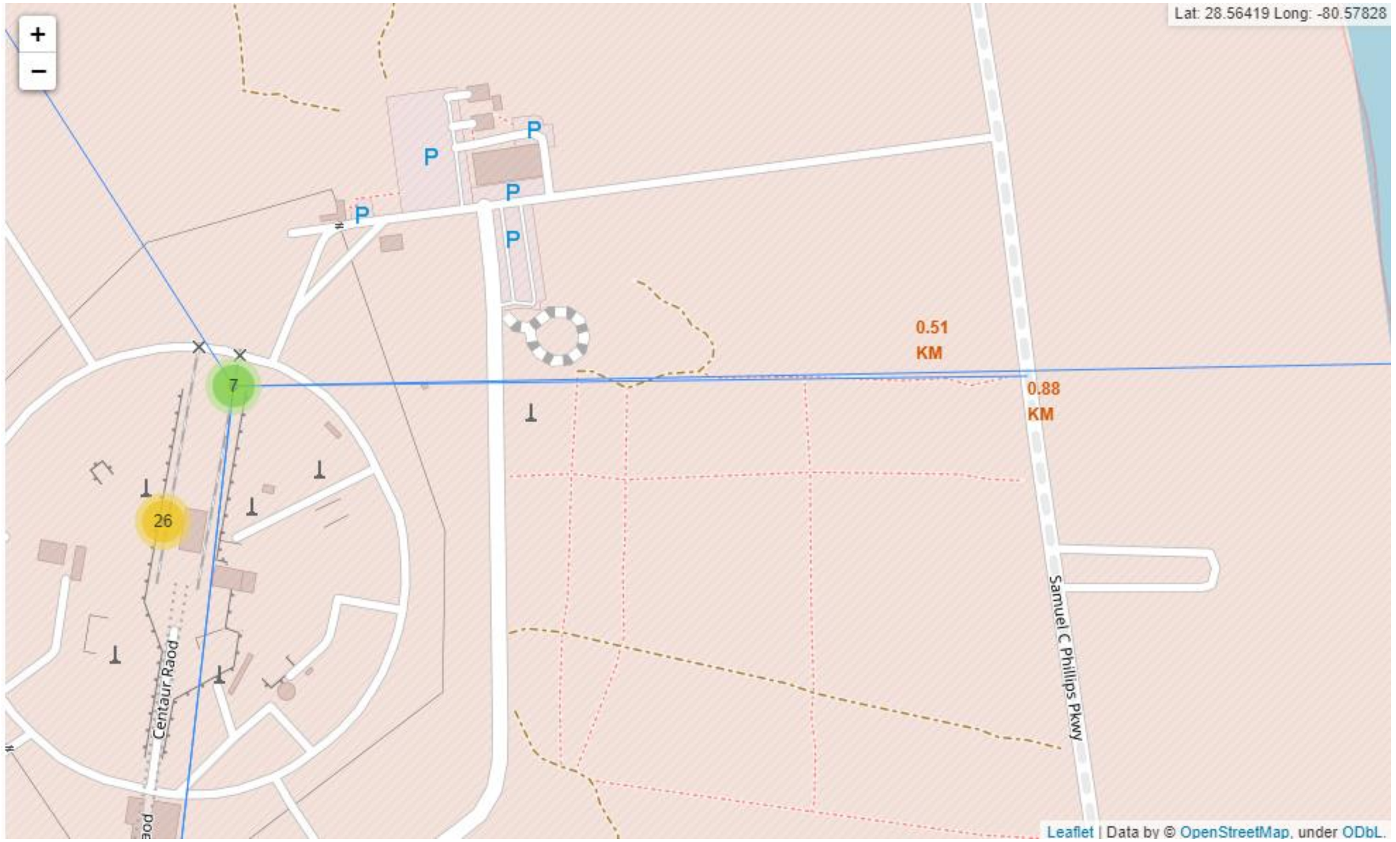
Now we are going pinpoint certain ladnscape features to the eastern cluster launch sites to determine the distance. First a map showing the closest coastline with aprox 0.9KM.



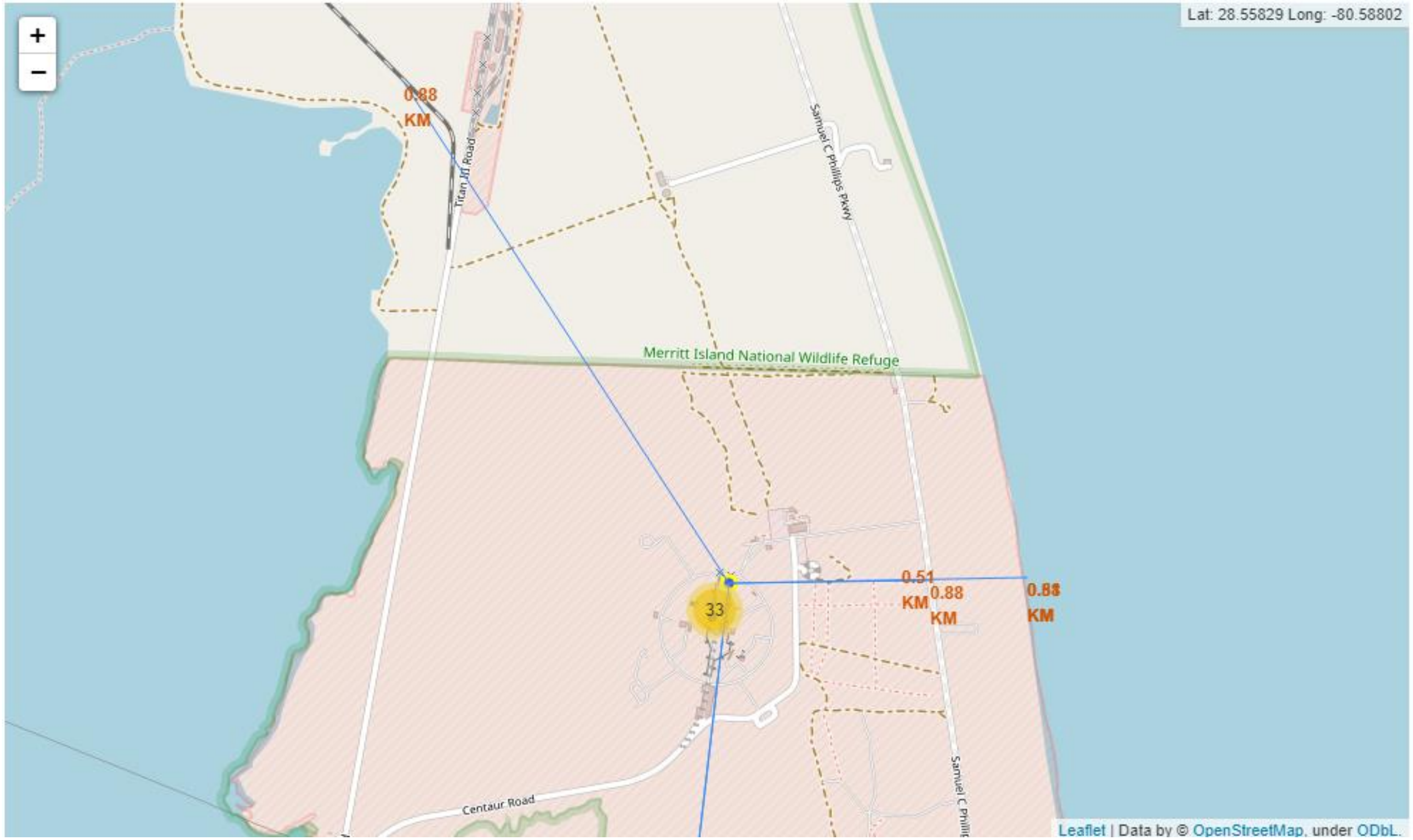
Now map showing the closest city and its distance



Closest Highway

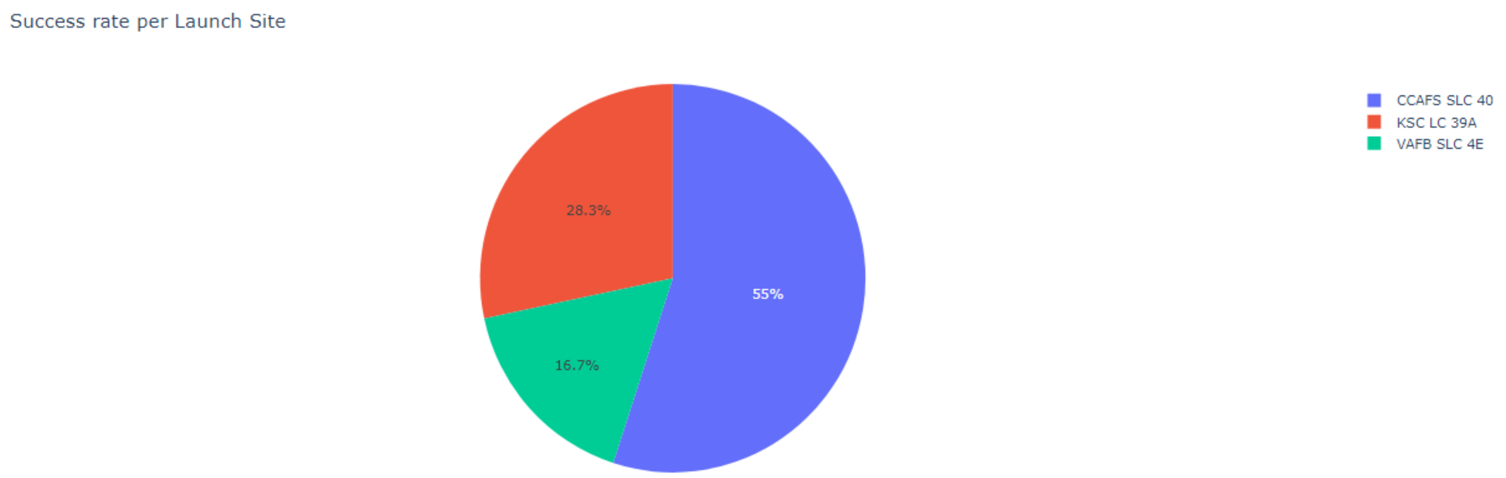


And closest Railway

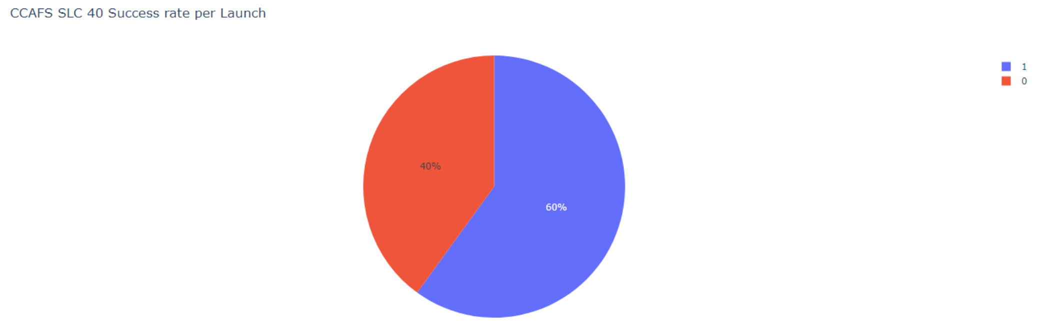


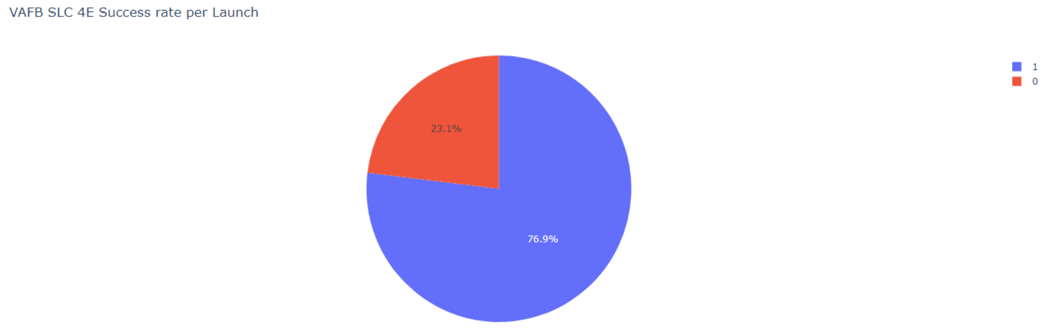
RESULT

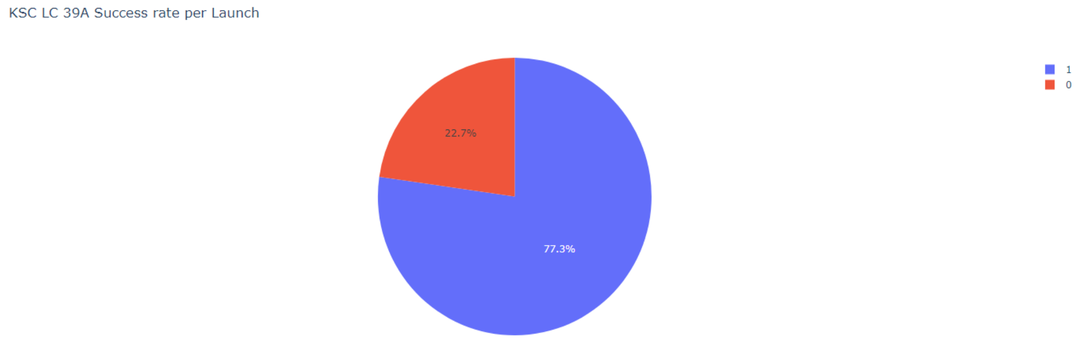
Plotly Dash



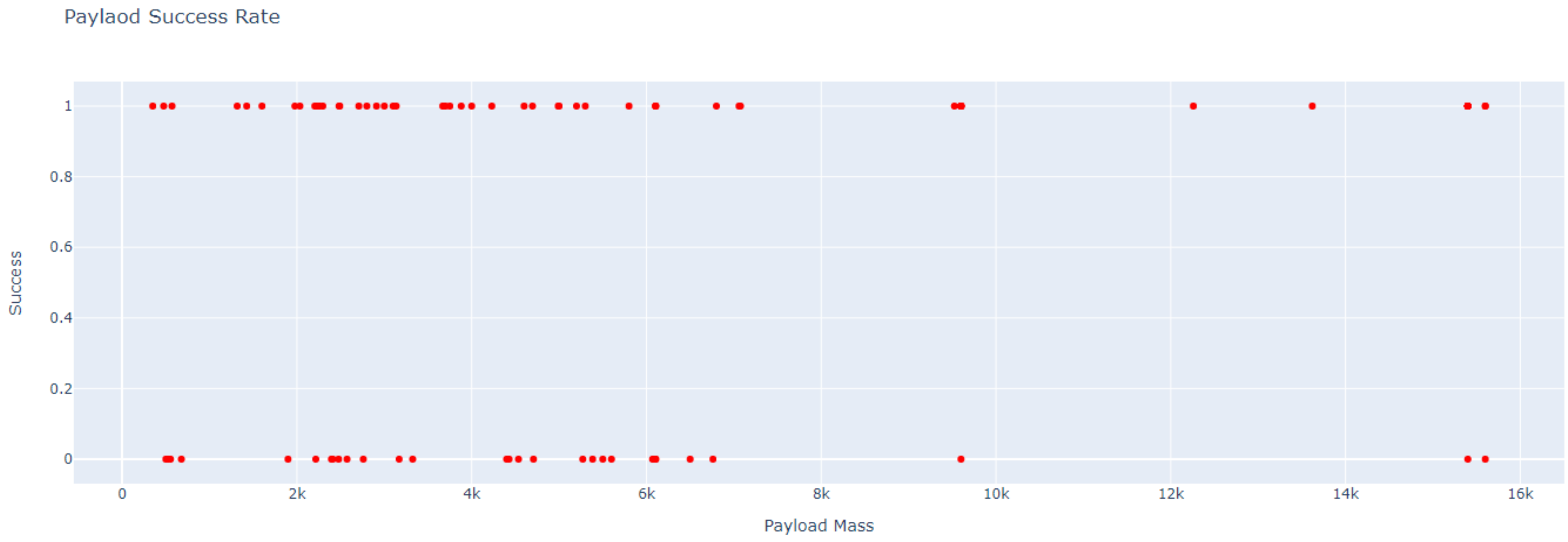
CCAFS SLC 40 has only a 60% success rate, which would deem it less efficient that the other launch sites with VAFB SLC 4E with 76,9% success rate and KSC LC39A with 77,3% success rate respectivelly. I taken at first value, but remembering scatterplot ebfore, CCAFS SLC 40 had the most test flights at the earlly stage more proned to failure.







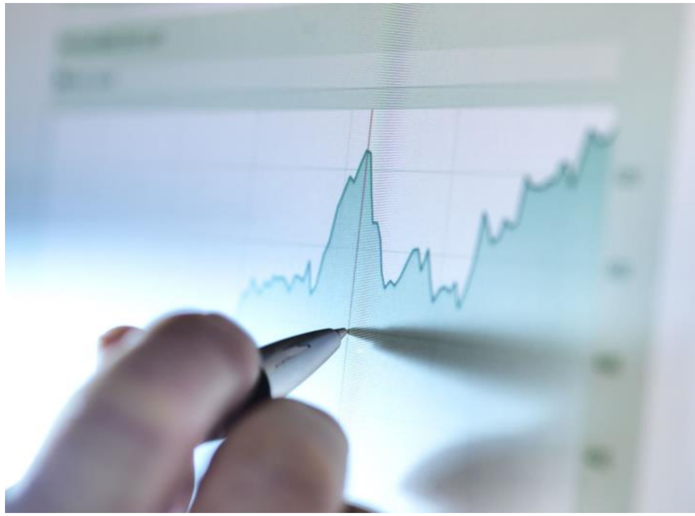
Not alot of disparity between Pay load mass as it has na even spread.



RESULT

Predictive Analysis

Here we will have a pipeline with different models and their prediction accuracy to see which has the best fit for future predictions. We use GridSearchCV objects with cv = 10 to find the best parameters.



Logistic Regression

Best Parameters:

tuned hpyerparameters :(best parameters) {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713

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

```
logreg_cv.score(X_test, Y_test)
```

```
0.8333333333333334
```

Lets look at the confusion matrix:

```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```

Support Vector Machine

Best Parameters

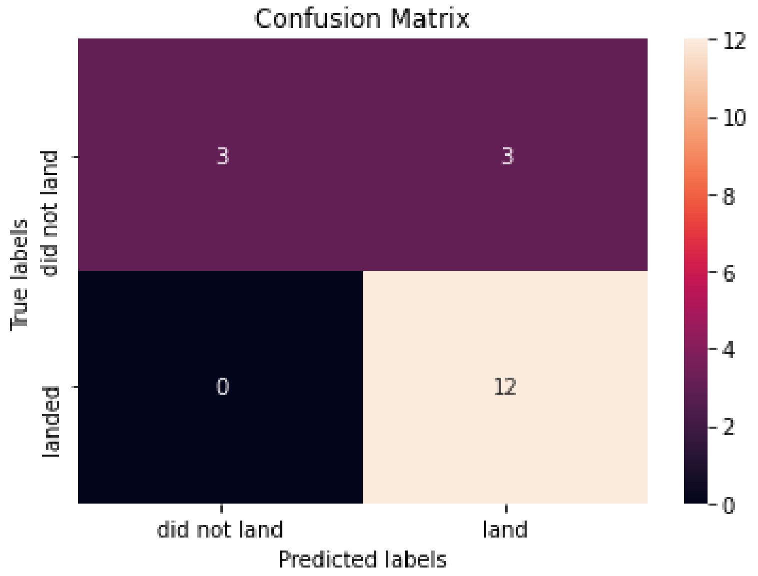
tuned hpyerparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel':'sigmoid'}
accuracy : 0.8482142857142856

```
svm_cv.score(X_test, Y_test)
```

0.8333333333333334

We can plot the confusion matrix

```
yhat=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Best Parameters

tuned hpyerparameters :(best parameters) {'criterion': 'gini', 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'random'}
accuracy : 0.875

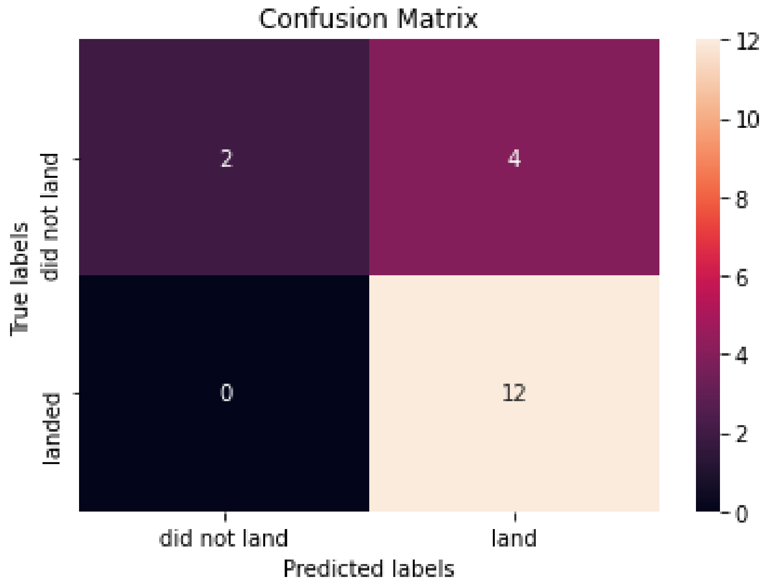
Decision Tree

```
tree_cv.score(X_test, Y_test)
```

0.7777777777777778

We can plot the confusion matrix

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



K Nearest Neighbors

Best Parameters

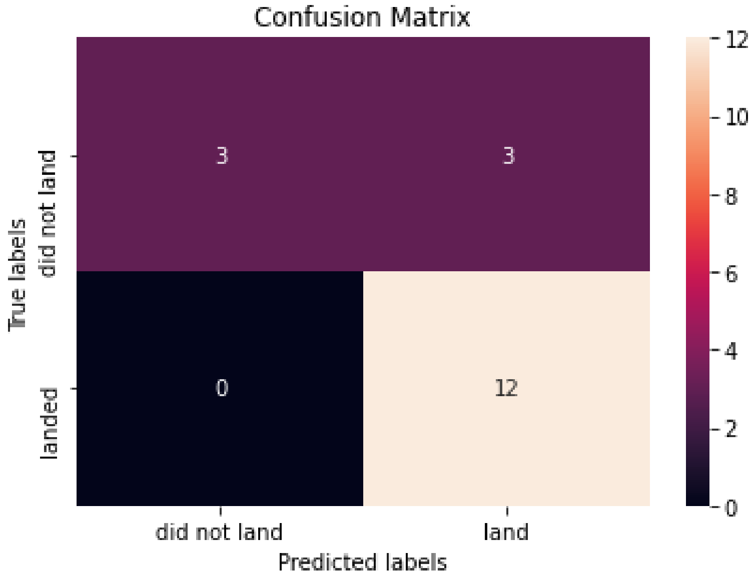
tuned hpyerparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858

```
knn_cv.score(X_test, Y_test)
```

0.8333333333333334

We can plot the confusion matrix

```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



DISCUSSION

Accuracy Score

Logistic Regression = 0.8333333333333334

Support Vector Machine = 0.8333333333333334

Decision Tree = 0.77777777777778

K Nearest Neighbors = 0.8333333333333334

Apart from Decision Tree, the other three models had simillar results. Decision Tree ahd different results while refiting, while the other three remained consistant.
With our current data, using Logistic Regerssion SVM or KNN would be the better choices even if Decision Tree sometimes had better acuracy score some times.

CONCLUSION

Based in our data we have these conclusions:

- Had multiple Booster versions that could carry above the avrage payload mass

- VAFB-SLC isn't feasible for not handling paylaod mass bigger than 10000

- For Orbits SSo, HEO, GEO and ES-L1 had 100% success rate making them the best choices. While SO had 0% success rate.

- Site location is important for having access to the coast for sea absed landing, access to nearby highways and railroads for easy access for our workers and finally far way from populated areas to avoid any pottential accidents during a launch failure.

- CCAFS SLC 40 Launch Site had the most successful flights despite past performances lowering its rate as showned by the graph where successfull flight started to happen after 2013.

CONCLUSION

And finally we got our classifier models with above 80% accuracy for future predictions when moving forward.

Now with the aquired data and information we can start in almsot equal ground by using the most efficient boosters, most effective launch sites and their appropriate locations , the best orbits, the necessary paylaod masses and the predictive analysis to keep going minimizing any pottential problems and error.

By studying the past, even small data, can prove big results for any company, even something big as going to space.

APPENDIX

GithubLink:

https://github.com/RedCheese/capstone