
# THE WORLD OF REUSABLE SPACESHIPS

OGBUEHI AGATHA UJUNWA

29th September 2022

IBM Developer

SKILLS NETWORK

# OUTLINE

1. Executive Summary
2. Introduction
   - Background of the study
   - Statement of problem
   - Question(s) to answer
3. Methodology
   - Data collection using API request to webscrapping and BeautifulSoup.
   - Data wrangling.
   - Exploration of data using SQL.
   - Visualization of data using matplotlib and pandas.
   - Interactive map built with folium and interactive dashboard built with plotly dash.
   - Machine learning prediction using SVM, DecisionTree, KNN and Logistic Regression.
- Results
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

- SpaceX falcon9 has built a spaceship with two stages of which stage one (the main component) has a higher probability of being reusable. This reusable tendency when successful has made the launching of spaceship cheaper than it used to be.
- To bid against SpaceX in the space world, it is important to determine the successful landing of stage one, use the specification to build a Space Y spaceship which will be used to bid against SpaceX falcon9 for a better prize.
- To do these, many analytic process will be done, and they include;
  - Data collection using API and webscrapping, wrangling of data to obtain a better comprehensive data, exploration of data using SQL, visualization of data using folium map and plotly dash.
  - Prediction of successful landing using machine learning.
- This whole process will help in predicting successful landing keeping in consideration the features of the spaceship and how it affect landing success. This information will help in the building of Space Y spaceship which will be used to bid against Space X.

# INTRODUCTION

## BACKGROUND

The exploration of space has been ongoing since 1961 with the first cosmonaut in the person of Yuri Gagarin landing in space. Since then, exploration of space has continued with spaceships being destroyed after launching. These destructions has made landing in space very expensive with other spaceships announcing the use of $165 million for manufacturing and launching of spaceship, while Elon Musk is announcing $65million for reusable manufacturing of spaceships since the major component is already available. The reusable spaceship has made landing in space cheap. Bidding against SpaceX will require the manufacturing of reusable spaceships too.

## THE PROBLEM STATEMENT

- What are the factors that determine the landing success of the reusable component of Space X?
- Can a prediction of successful landing be made using SpaceX falcon 9 data?
- Is there a relationship between the features of spaceship and the landing success?

## PURPOSE OF RESEARCH

- To predict the successful landing of SpaceX falcon 9 spaceship to Build Space Y which will bid against Space X.

# METHODOLOGY

## DATA COLLECTION

- Data was collected in different ways.

1. The REST API was used to obtain target launch data Like launch specification, payl oad delivery, rocket used, landing specification and landing outcome. The URL obtained was used to obtain past launch data. A get request from request library was used to obtain the launch data and the result was passed to .json(), normalized and into a dataframe.

Link to complete process: https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/SPACEX%20data%20collection.ipynb

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

```
response = requests.get(spacex_url)
```

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/AP
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
from pandas.io.json import json_normalize
import json
```

```
j=response.json()
```

```
data=pd.json_normalize(j)
```

1. Webscrapping from Wikipedia was done using BeautifulSoup to obtain launch dictionary. The content of the table was extracted using launch_dict= dict.fromkeys(column_names). all the column names were gradually obtained and converted to pandas DataFrame and saved as a csv file.

full process link:https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/spacex%20webscraping.ipynb

```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# Assign the result to a list called `html_tables`
html_tables=soup.find_all('table')
```

Starting from the third table is our target table contains the actual launch records.

```
# Let's print the third table and check its content
first_launch_table = html_tables[2]
```

```
column_names = []
# Apply find_all() function with `th` element on first_la
# Iterate each th element and apply the provided extract_
# Append the Non-empty column name (`if name is not None
spe=soup.find_all('th')
for i in range (len(spe)):
    try:
        ace= extract_column_from_header(spe[i])
        if (ace is not None and len(ace)>0):
            column_names.append(ace)
    except:
        pass
```

Check the extracted column names

```
print(column_names)
```

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

# DATA WRANGLING

The data obtained from Beautifulsoup was wrangled using python. Finding the data type, sampling the data, getting the value count, viewing the distinct launch site, including an outcome column to show if the landing was successful or not and dealing with null value to obtain a more comprehensive data which will be visualized, tested and be used for prediction.

Link to complete process: https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/DATA%20WRANGLING%20.ipynb

```
df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/datas
df.head(10)
```

```
df.isnull().sum()/df.count()*100
```

```
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
```

Identify which columns are numerical and categorical:

```
df.dtypes
```

# DATA EXPLORATION USING SQL AND MATPLOTLIB

- Exploration of data was done in two ways.

1. It was first queried with SQL to obtain important information like average payload, the booster that has carried the highest load, the distinct spaceship, successful and failed missions, etc.

For full process, refer to the link below

https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/SPACE%20sql.ipynb

```
[7]: %sql select DISTINCT LAUNCH_SITE from SPACEXTBL
 * sqlite:///my_data1.db
Done.
```

[7]:

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

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER LIKE 'NASA (CRS)'
 * sqlite:///my_data1.db
Done.
```

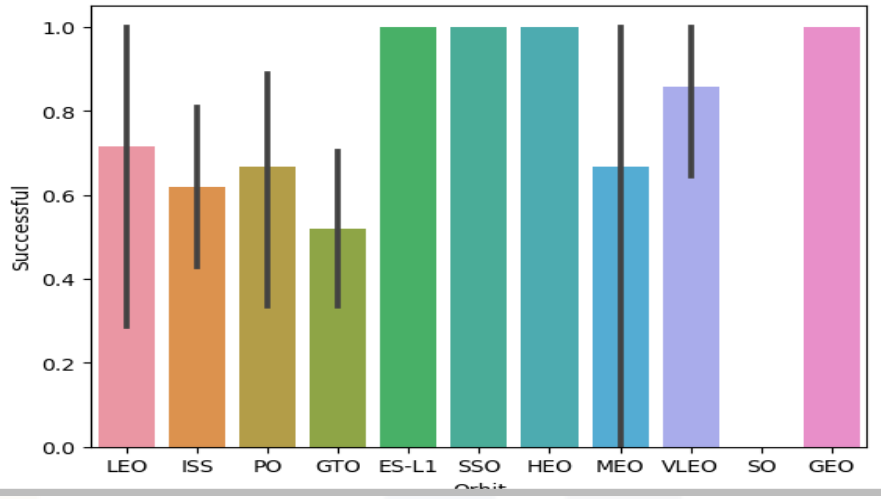
| sum(PAYLOAD_MASS__KG_) |
|---|
| 45596 |

# DATA EXPLORATION MATPLOTLIB

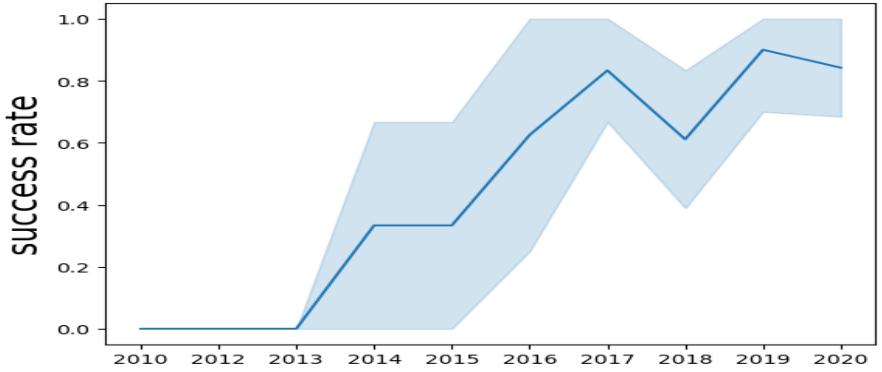
- The relationship between different parameters of the spaceship was plotted to determine if there is a relationship. Payload, launch site, year of launch were all plotted to determine the relation between parameters and the outcome.

Full process link: https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/SPACEX%20DATA%20VIZUALIZATION.ipynb

LAUNCH SITE AGAINST SUCCESS RATE




YEAR AGAINST SUCCESS RATE



# INTERACTIVE MAP WITH FOLIUM

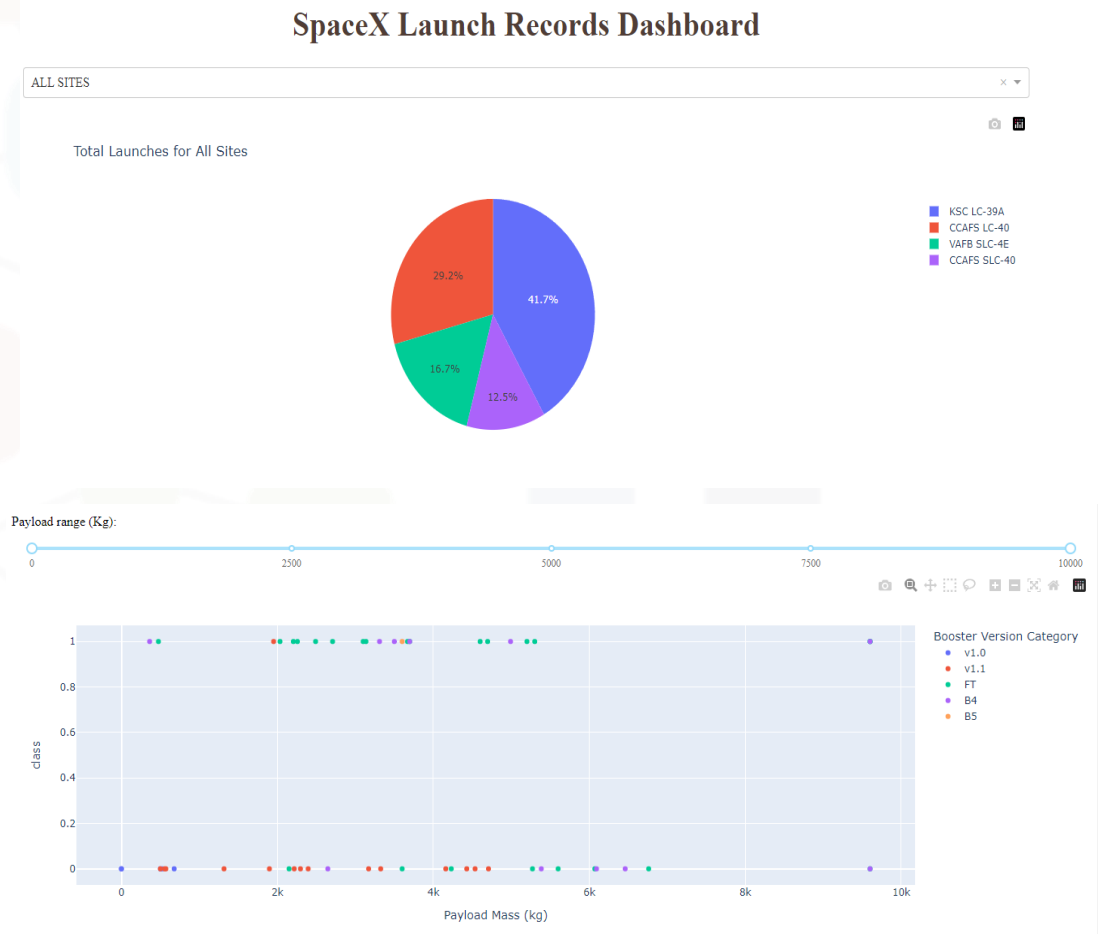
Interactive map of all launch site was done in the following ways;

- ✓ First the sites of the major launch site were extracted and located on the map;
- ✓ Successful and failed outcome were marked on the map;
- ✓ The distance between the map and its proximity was marked;
- ✓ A line was drawn to show the launch site with their proximities.

Full process link: https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/MAP%20INTERACTIVE%20OF%20SPACEX%20.ipynb


# DASHBOARD WITH PLOTLY DASH

- A drop down for selection of all launch sites was created.
- A pie chart indicating all launch site was created to show the percentage of successful and failed outcome.
- A relationship was established between the payload and outcome.




- https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/DASHBOARD%20SPACEX%20PLOTLY%20DASH.py

# PREDICTION (CLASSIFICTAION)

Prediction was done using machine learning classification. The steps followed include;

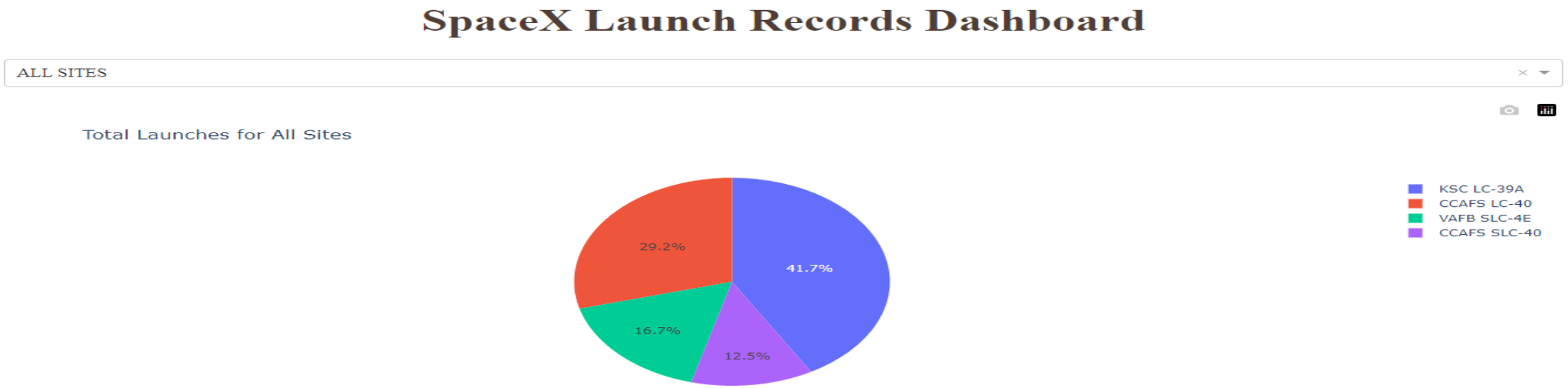
- All needed libraries were imported;
- All data was loaded into pandas, transformed and standardized;
- The hyperparameters were selected using GridSearchCV
- The data was split into test and train and was tested with four classification methods which include KNN, DecisionTree, LogisticRegression and SVM.
- The accuracy score of each method was calculated using .score
- The best classification method (DecisionTree) with the highest accuracy of 83.34% was obtained.

# RESULTS

- We had more unsuccessful mission outcome than successful mission outcome.
- Some launch sites like HEO, SSO, GEO AND ES-L1 has a high success rate.
- Only LEO has a relationship between the success rate and flight number
- Heavy payload leads to increase in successful outcome rate in LEO, ISS and POLAR.
- From 2014, we had constant increase with successful launch.

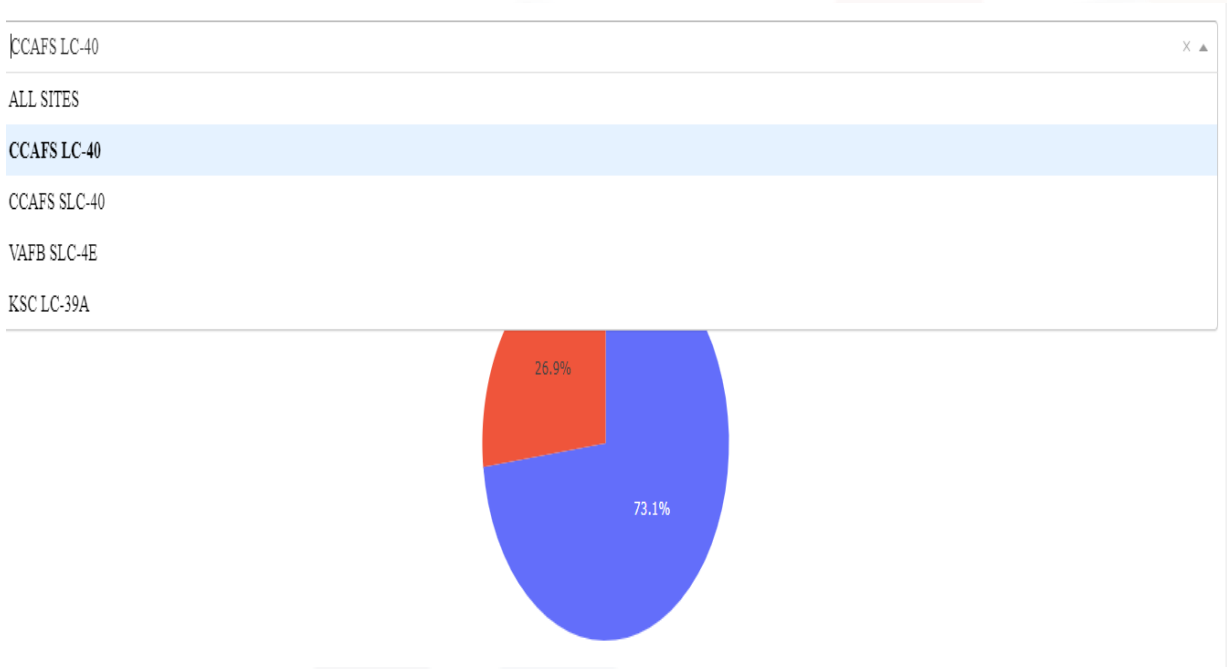
# DISCUSSION

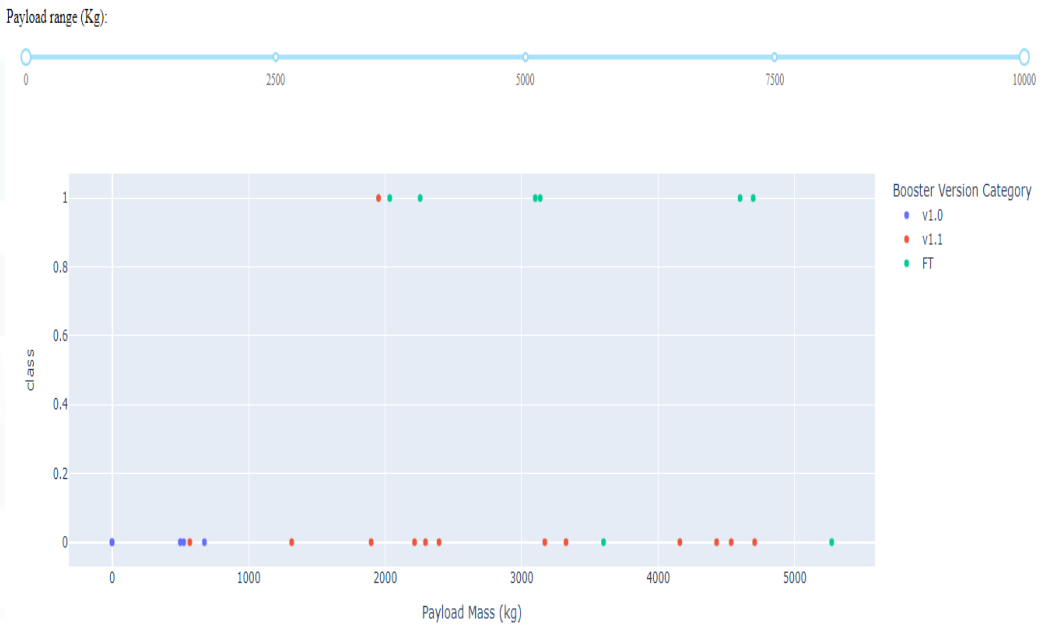
## DASHBOARD TRENDS (PAYLOAD, SUCCESS AND FAILED OUTCOME FOR ALL FLIGHTS)




- KSC-LC-39A has the highest with successful percentage of 41.7 followed by CAFS LC-40 with 29.2%, VAFB SLC-4E with 16.7% and the rest is for CCAFS SLC-40.

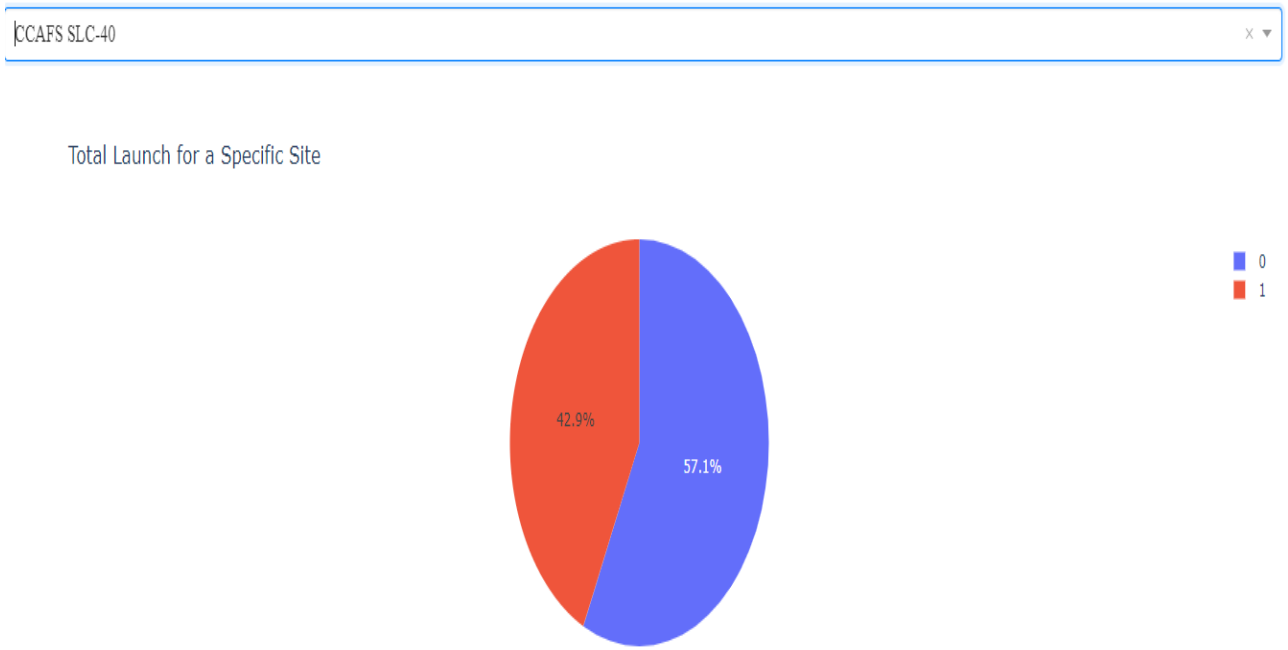
# DASHBOARD TRENDS (PAYLOAD, SUCCESS AND FAILED OUTCOME FOR ALL FLIGHTS)

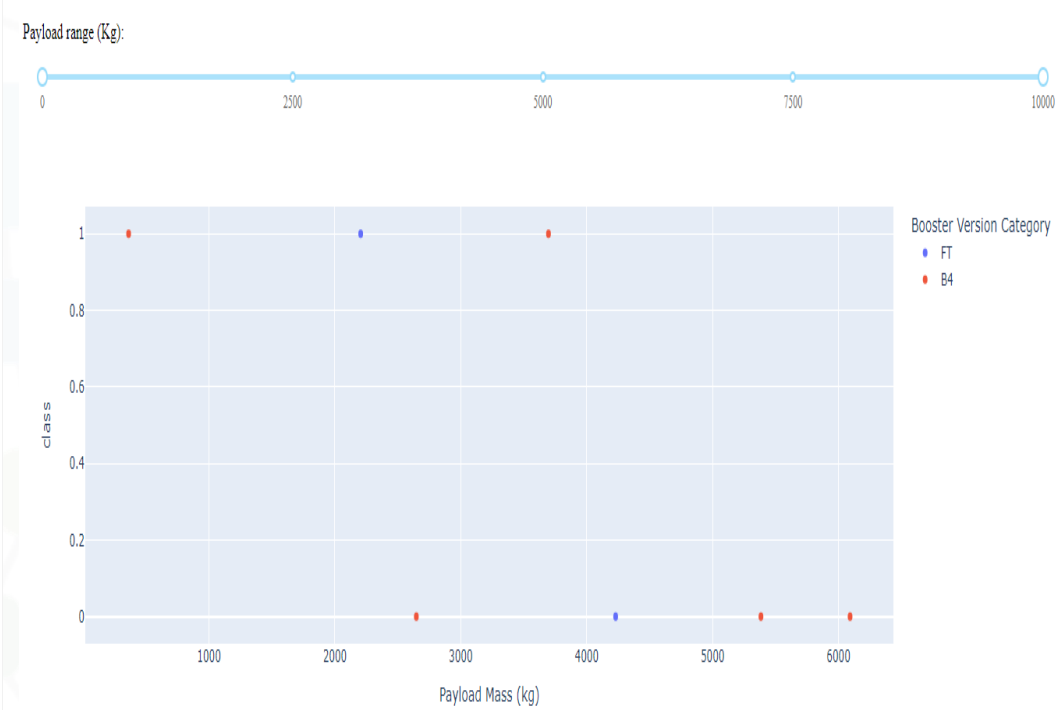



The percentage of successful outcome to failed outcome for CCAFS LC-40 is 73.1% for unsuccessful outcome and 26.9% of successful outcome

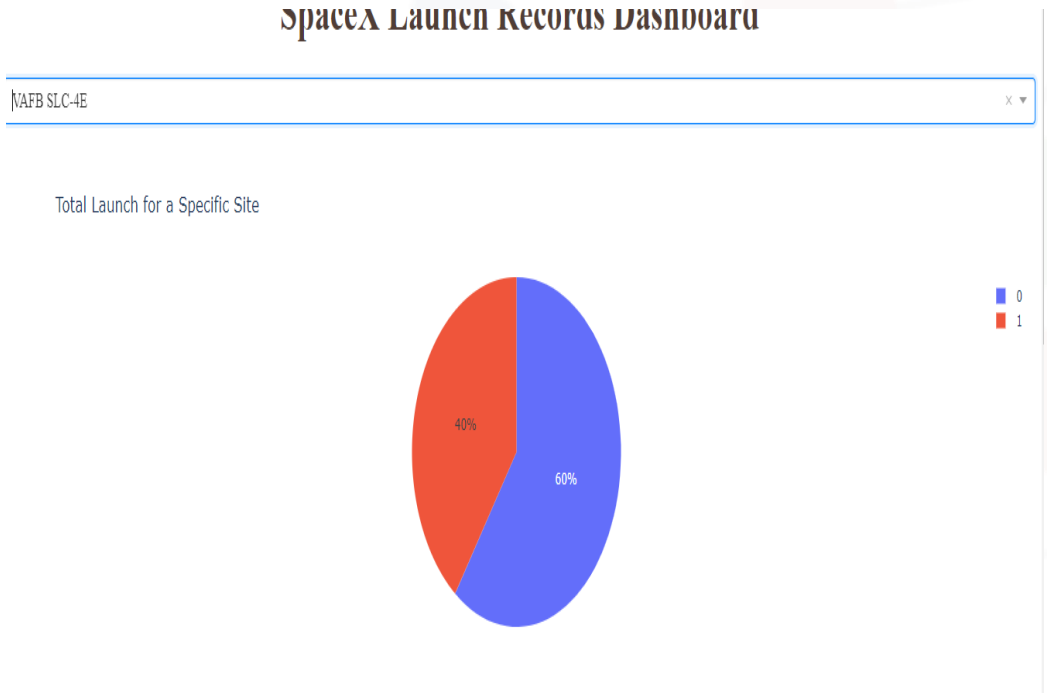
# DASHBOARD TRENDS (PAYLOAD, SUCCESS AND FAILED OUTCOME FOR ALL FLIGHTS)

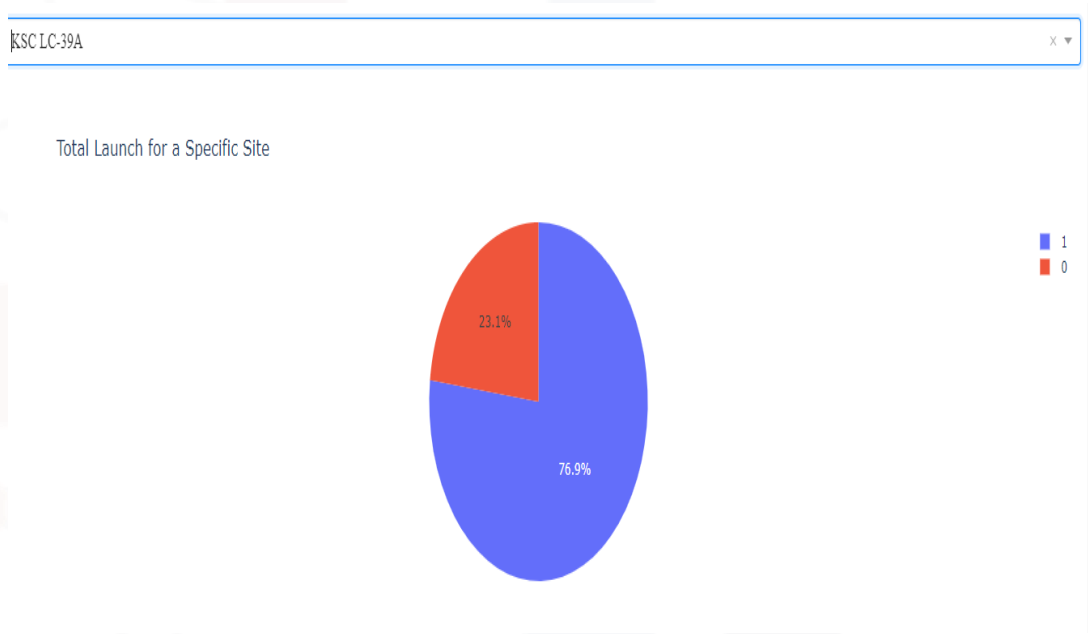



The percentage of successful outcome to failed outcome for CCAFS SLC-40 is 57.1% for unsuccessful outcome and 42.9% of successful outcome.

# DASHBOARD TRENDS
# (PAYLOAD, SUCCESS AND FAILED OUTCOME FOR ALL FLIGHTS)

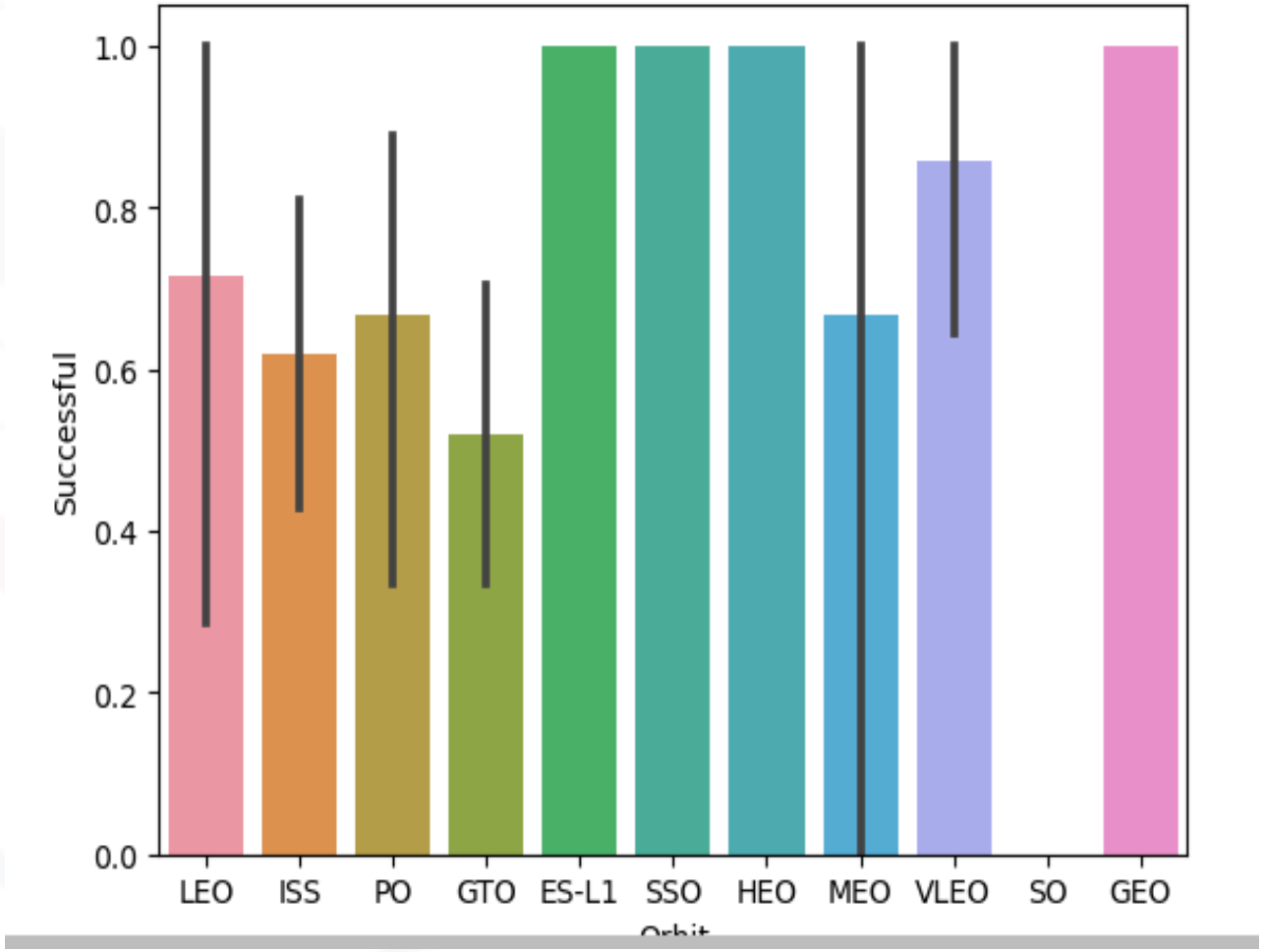



The percentage of successful outcome to failed outcome for KSC-LC-39A is 76.9% successful outcome making it the highest successful launch site and the rest is for of successful outcome. VAFB SLC-4E has a 60% failed outcome with a 40% success rate.

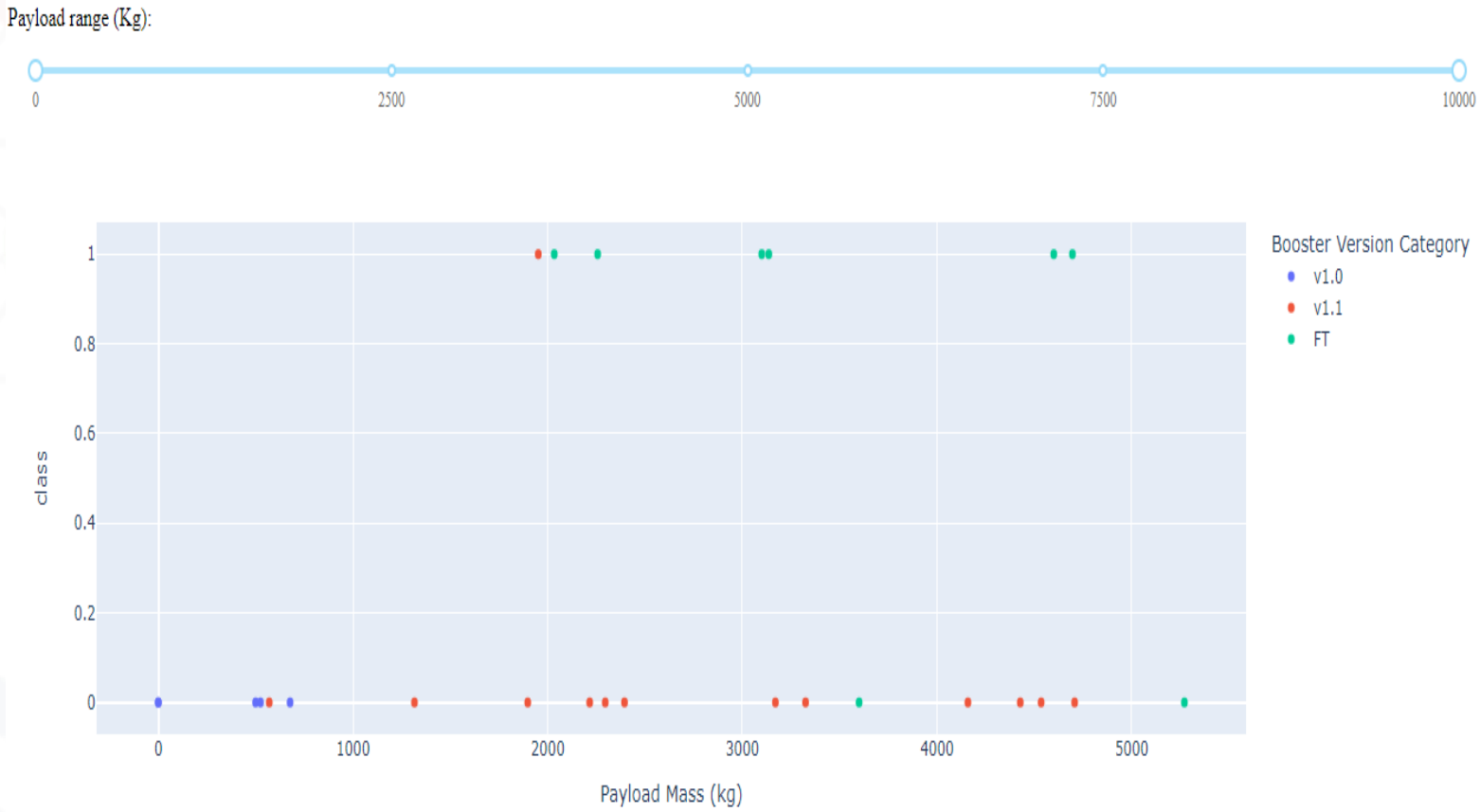
# Orbit relationship with success rate

- GEO, SSO, HEO and ES-L1 have higher successful rate.



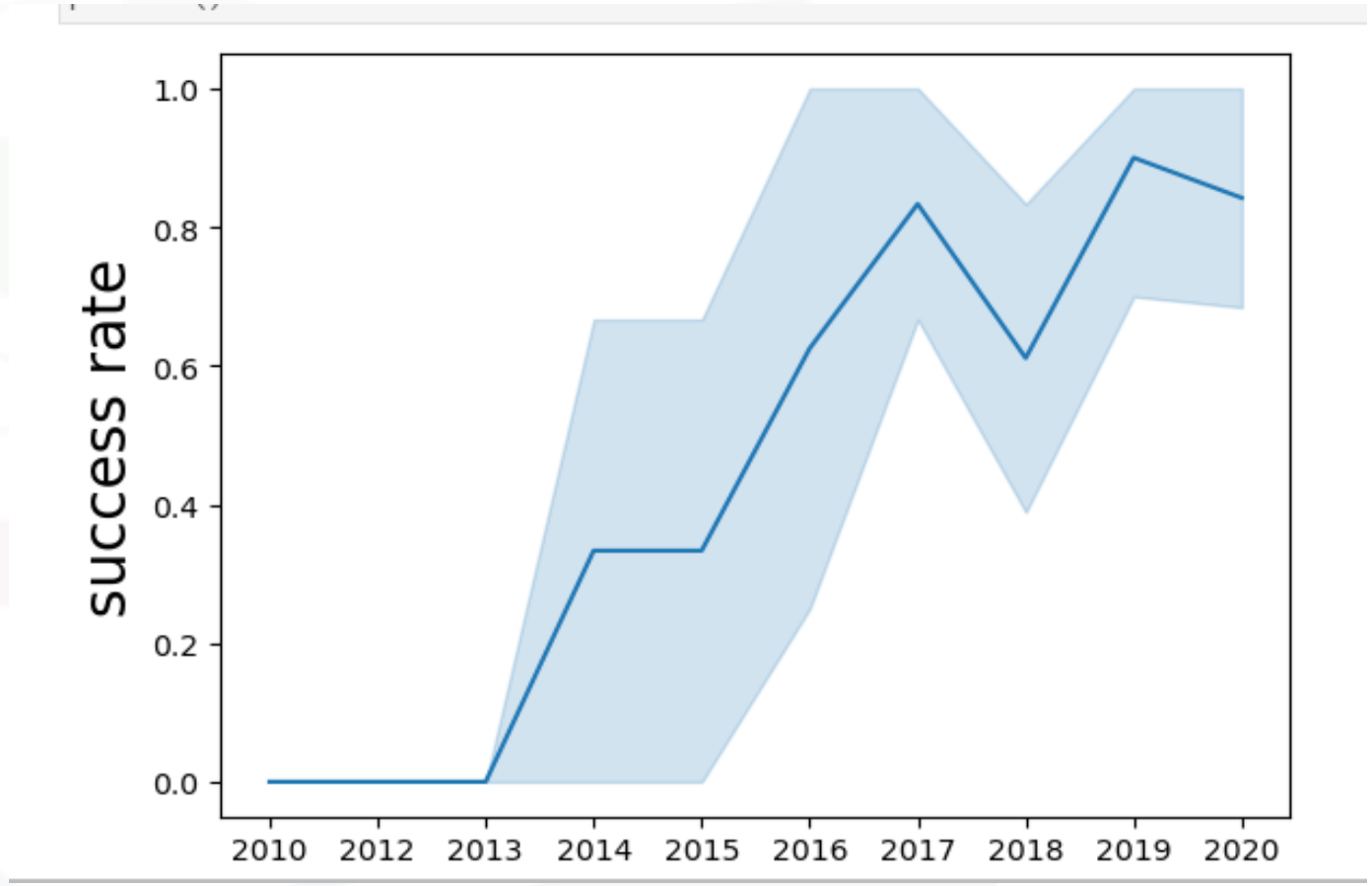
# Payload relationship with success rate

- There is no concrete relationship between some flights and their payload except polar, ISS and LEO.

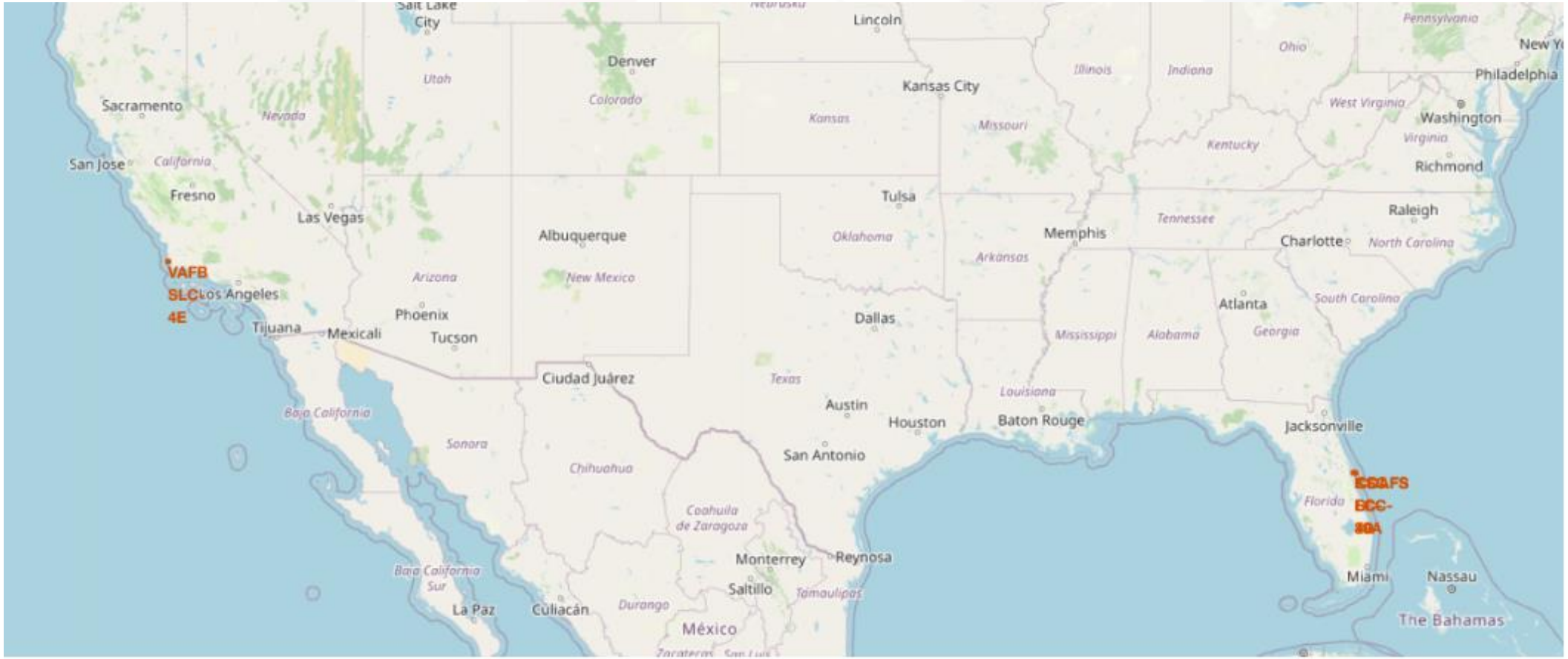


# Year relationship with success rate

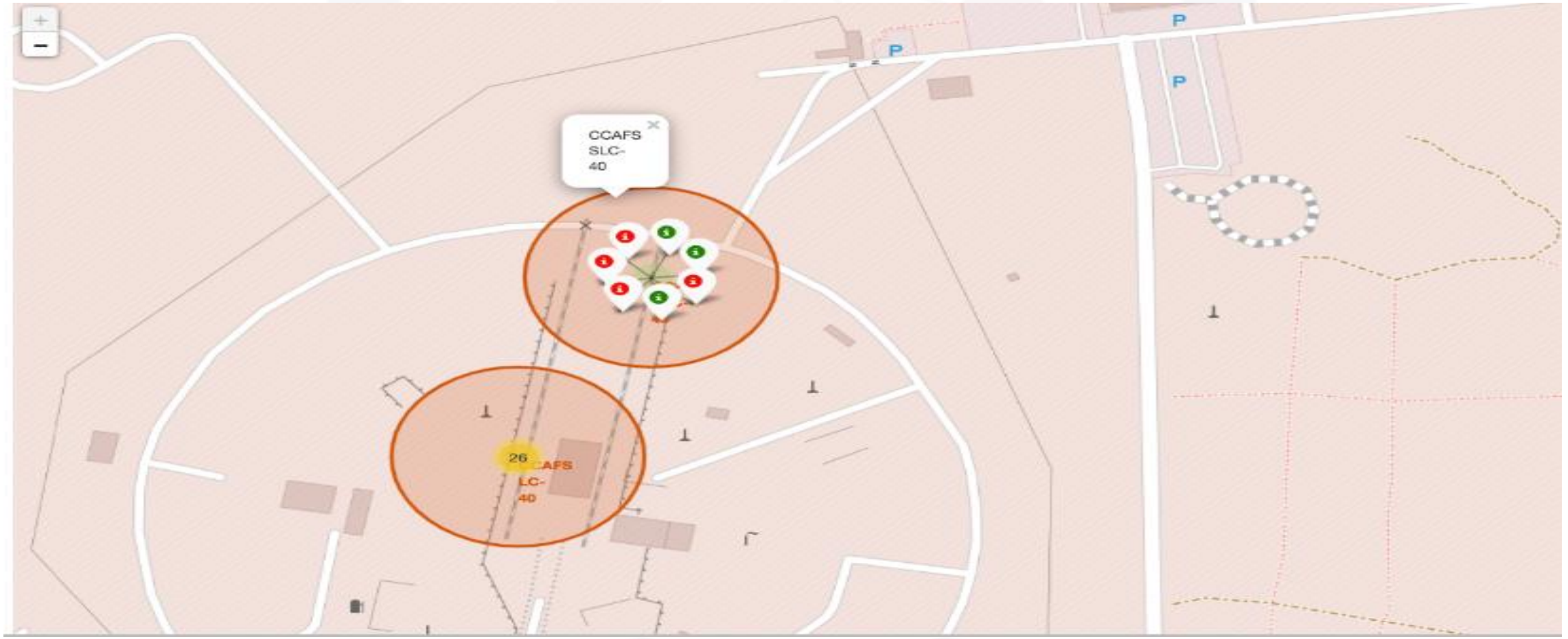
- Success rate increased with the increase in a year even though there was a slight decrease in success rate between 2017 to 2018.



# GLOBAL MAP LOCATION OF LAUNCH SITE



# SUCCESSFUL and failed OUTCOME ON A MAP



# Distance between a laun site and its closest city

# Prediction

- Classification method was used for prediction. The accuracy obtained was 83.33% for the KNN and SVM but for DecisionTree, an accuracy of 83.34% was obtained, making it the most accurate method of classification.
- Confusion matrix was plotted and it showed low level of true negative with a little bit of false negative (false prediction that states that there was a successful landing when in fact there was unsuccessful landing).

# CONCLUSION

- There is no significant relationship between payload and successful launching.
- Higher launch amount increases the tendency of having more successful outcome.
- HEO, GEO, ES-L1 has the highest successful orbit.
- Success rate started to increase from 2013.
- KSC LC-39A is the highest success launch site.
- Decision tree with accuracy of 83.34% is the best method to classify.

# Appendix

## DATA USED

Links to the data used:

- https://github.com/ujunwa-DS/SPACEX-FALCON-9-CAPSTONE-PROJECT/blob/main/Spacex.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
-