IBM Developer
SKILLS NETWORK

Joseph C
08/06/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

1. Collecting Data and API from a webpage
2. Transforming and mapping data
3. EDA with SQL
4. Visualizing EDA information
5. Predictive analysis with ML

Introduction

- The project investigates SpaceX's Falcon 9 rocket launches to determine the cost of each launch
- To determine if SpaceX will reuse the first stage
- Train a machine learning model and use public information to predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

- Data collection methodology:
  - Get data from SpaceX API
  - Web-scrape data from a Wiki page
- Perform data wrangling
  - Perform EDA
  - Determine label for training supervised model

Perform exploratory data analysis (EDA) using visualization and SQL

- Interactive visual analytics using Folium
- Predictive analysis using classification models

Data Collection

API

spacexAPI_url=https://api.spacexdata.com/v4/ launches/past

- Using this API get the required launch data from SpaceX

Web-Scraping

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Using this link to Web-Scrape data from WIKI and using it to make a Dataframe using the pandas module.

Data Collection - SpaceX API

- Request data from SpaceX API
- Convert the json result into a dataframe
- Filter dataframe to only Falcon 9` launches and data wrangling
- Export to csv

Request data

Convert into dataframe

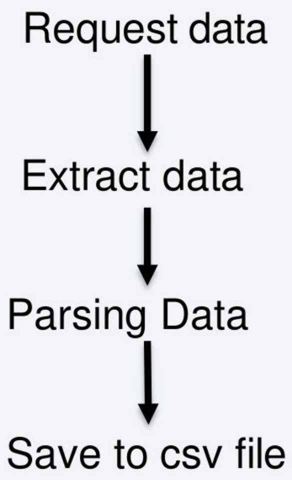
Data wrangling

Save to csv file

Data Collection - Scraping

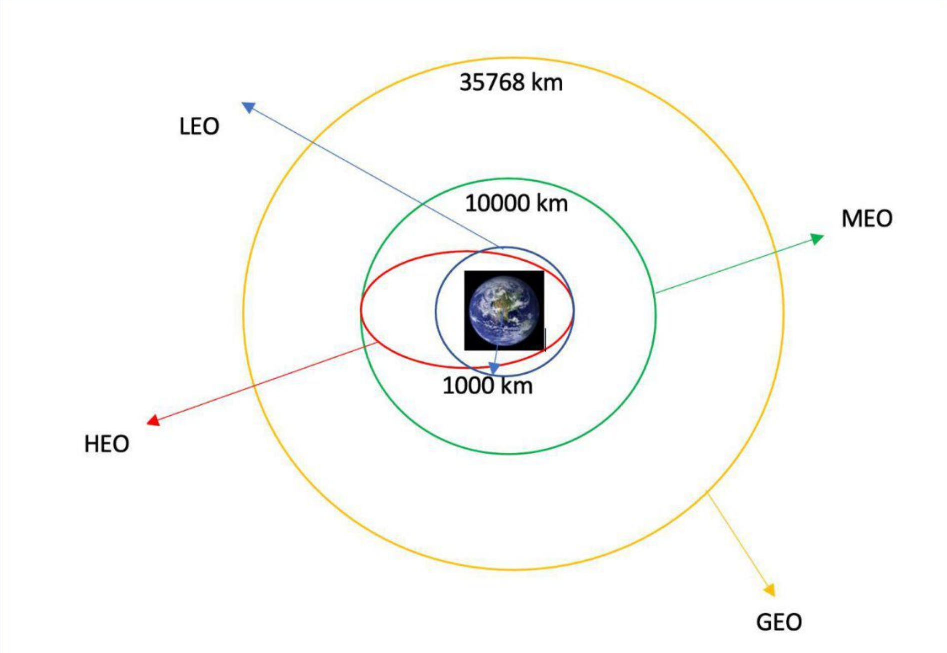
SPACEX

- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the launch HTML tables
- Export to csv



Data Wrangling

- Add a missing value
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of each orbit
- Create a landing outcome label from Outcome column



Link: https://github.com/JC-project-works/public

EDA with Data Visualization

- The relationship between Flight Number and Launch Site -> scatter plot
- The relationship between Payload and Launch Site -> scatter plot
- The relationship between success rate of each orbit type -> bar plot
- The relationship between Flight Number and Orbit type -> scatter plot
- The relationship between Payload and Orbit type -> scatter plot
- The launch success yearly trend -> line chart
- The scatter plot is the best to describe the relation between two categorical data
- The bar plot is the best to compare several categorical data
- The line plot is the best to show the time series data
- https://github.com/JC-project-works/public

EDA with SQL

- Display the names of the unique launch 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
- Display average payload mass carried by the falcon 9
- 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.
- https://github.com/JC-project-works/public

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- To find similarities between launch sites based on the geographical topology
- https://github.com/JC-project-works/public

Predictive Analysis (Classification)

Data wrangling

Data standarization

Split into traning and test datasets

Predictive model evalutation

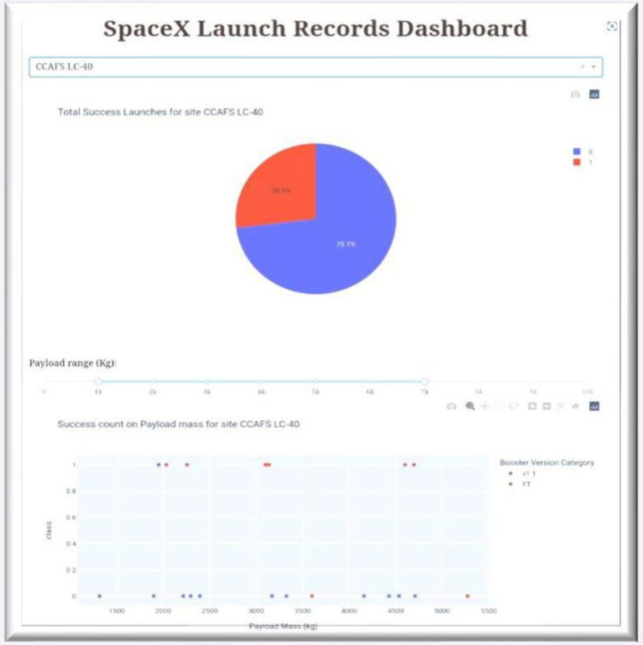
Predictive model selection

https://github.com/JC-project-works/public

Results

EDA Information

- KSC LC-39A and VAFB SLC
  4E has a success rate of 77%
- VAFB SLC 4E has no payload above 10000 kg
- In the LEO orbit the Success appears related to the number of flights

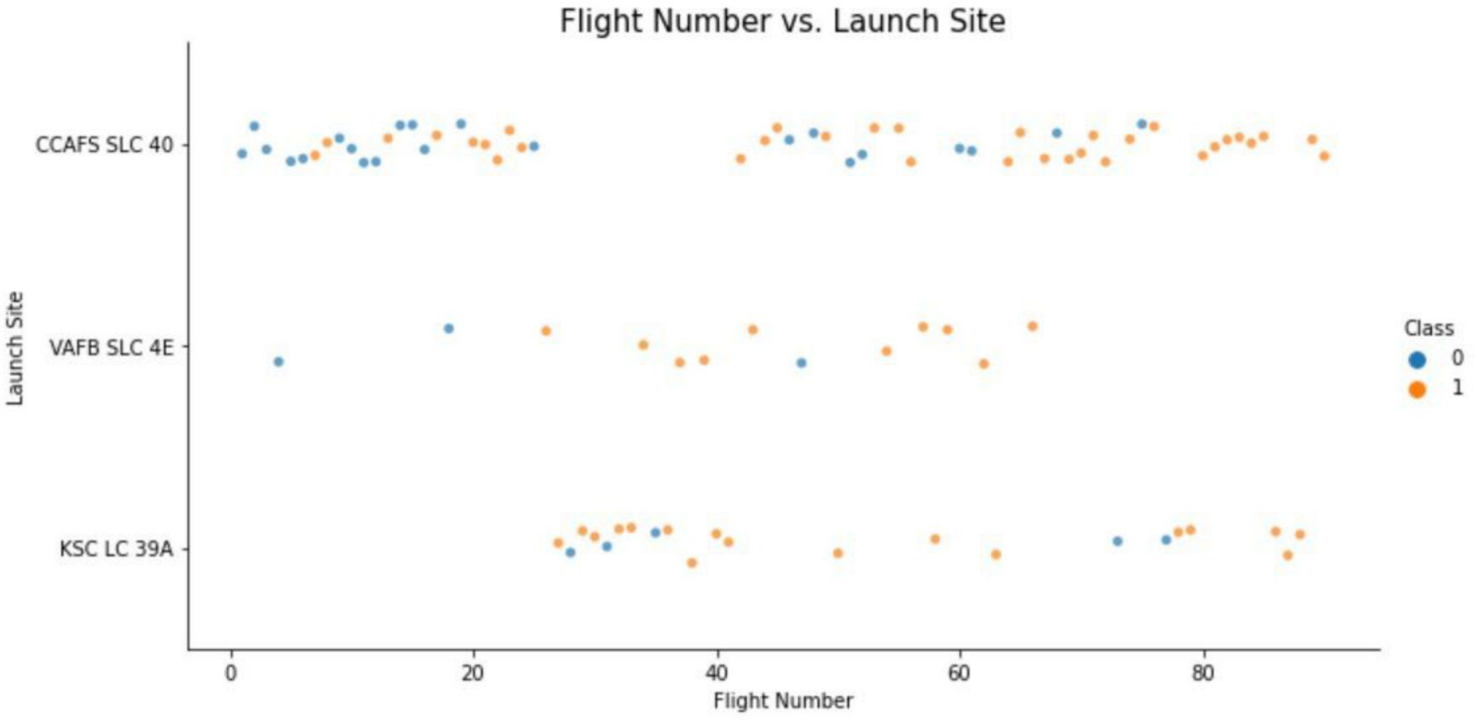


According to the classifier model, it tells us that there is an accuracy of around 83% with the best parameters.

Section 2

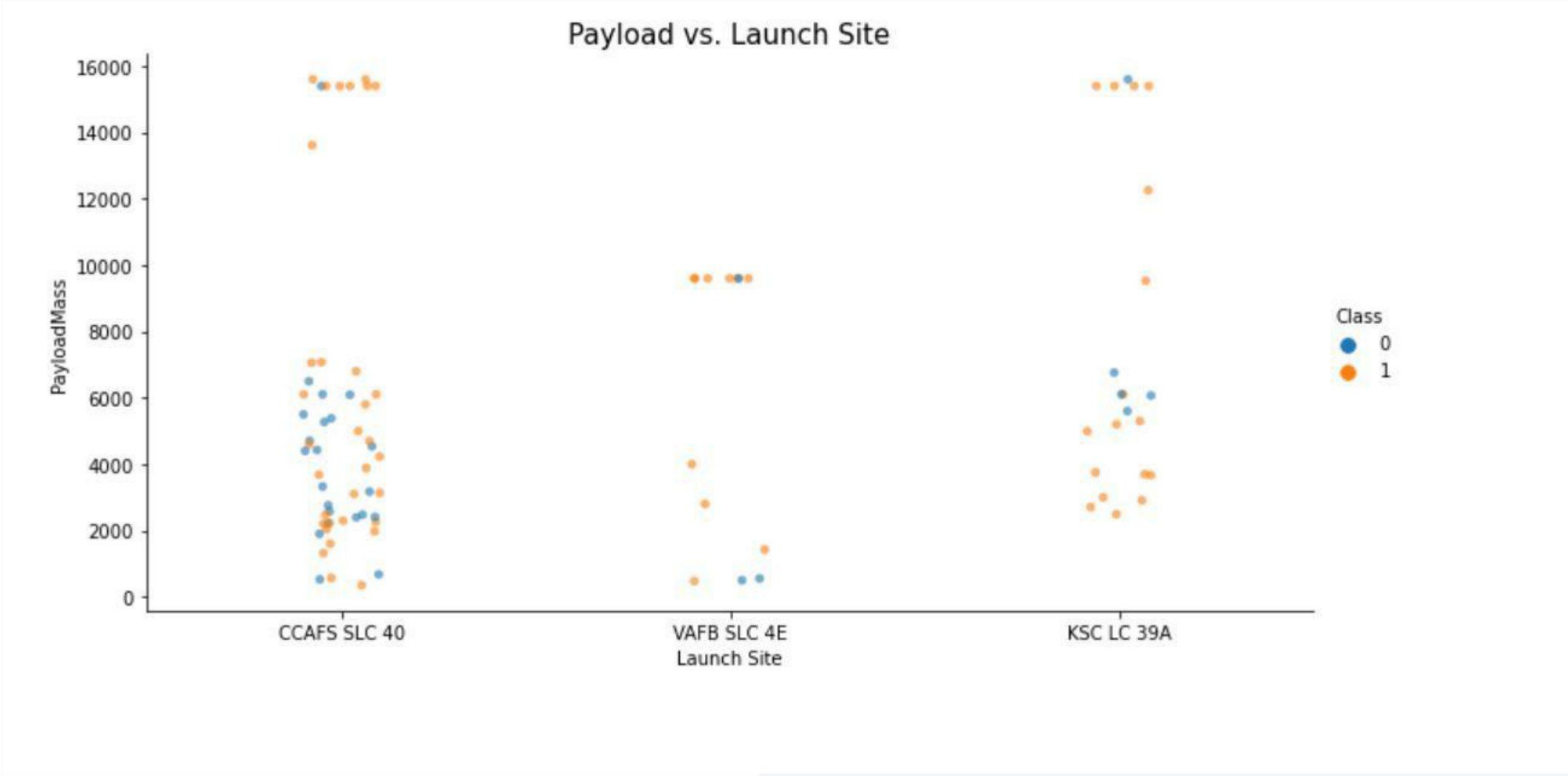
Insights drawn from EDA

Flight Number vs. Launch Site

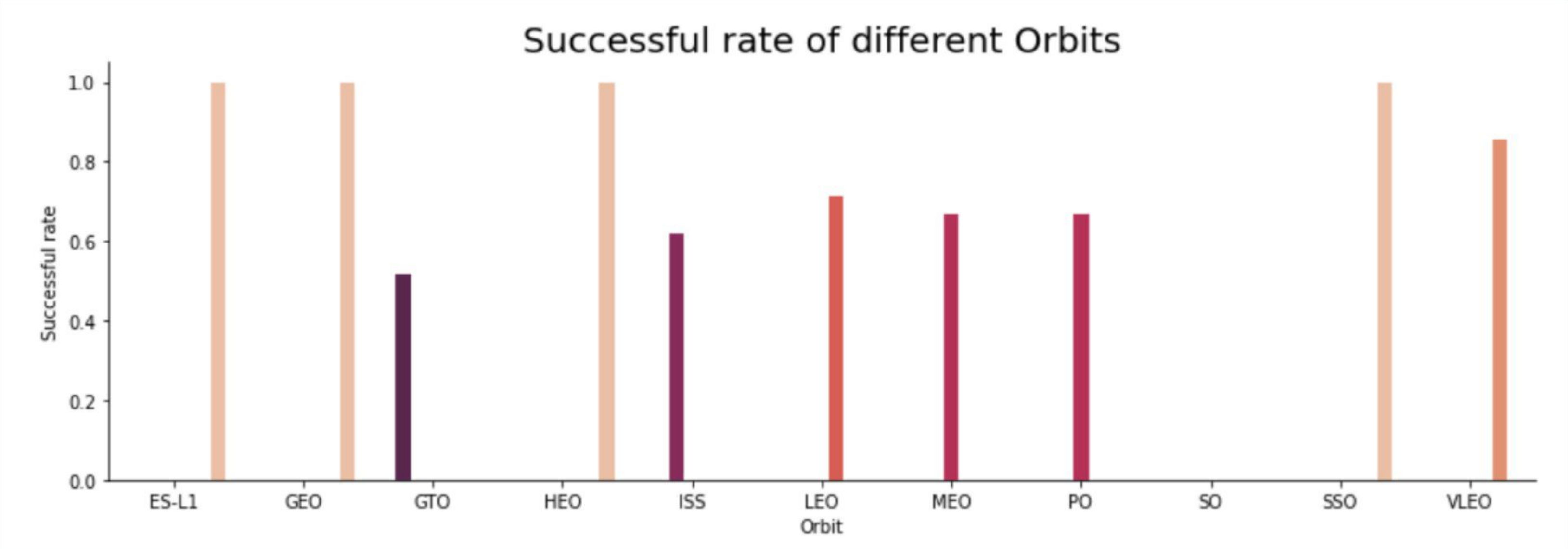


-> KSC LC 39A has the highest sucessful rate, vice verce CCAFS SLC 40 has the lowest

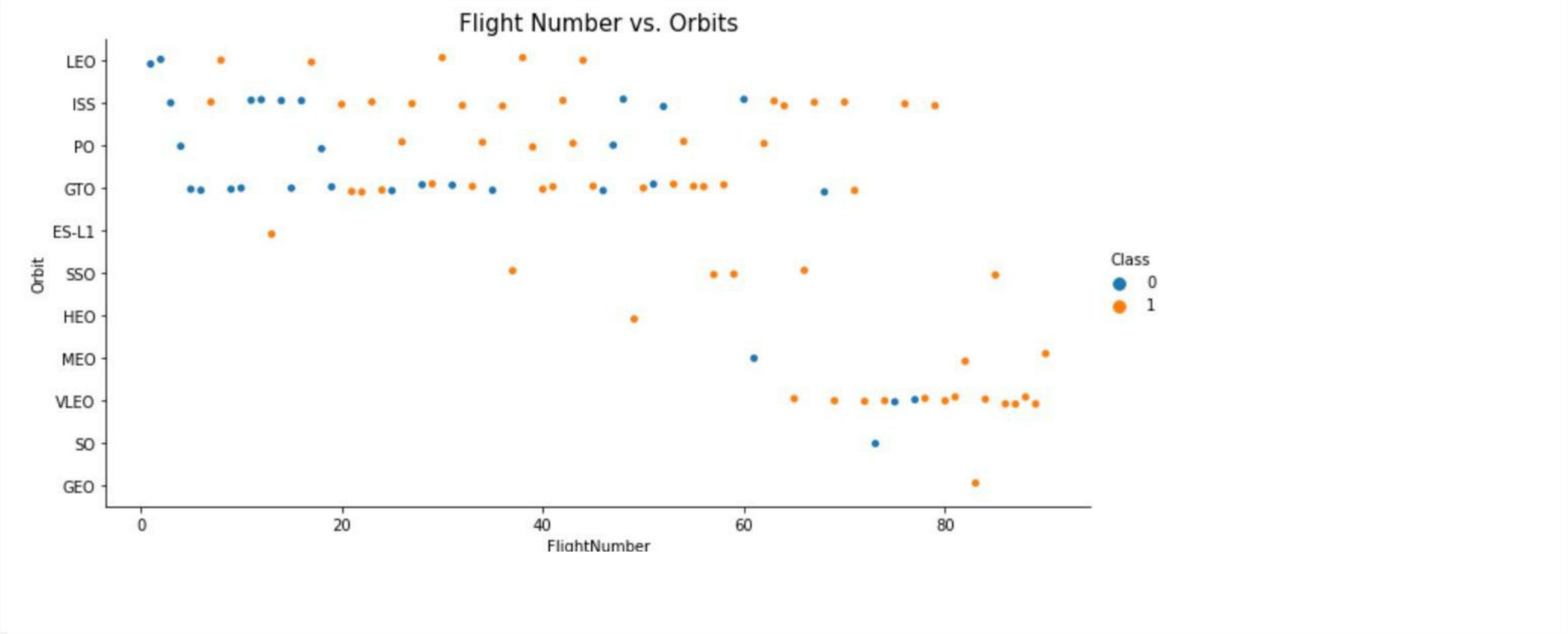
Payload vs. Launch Site



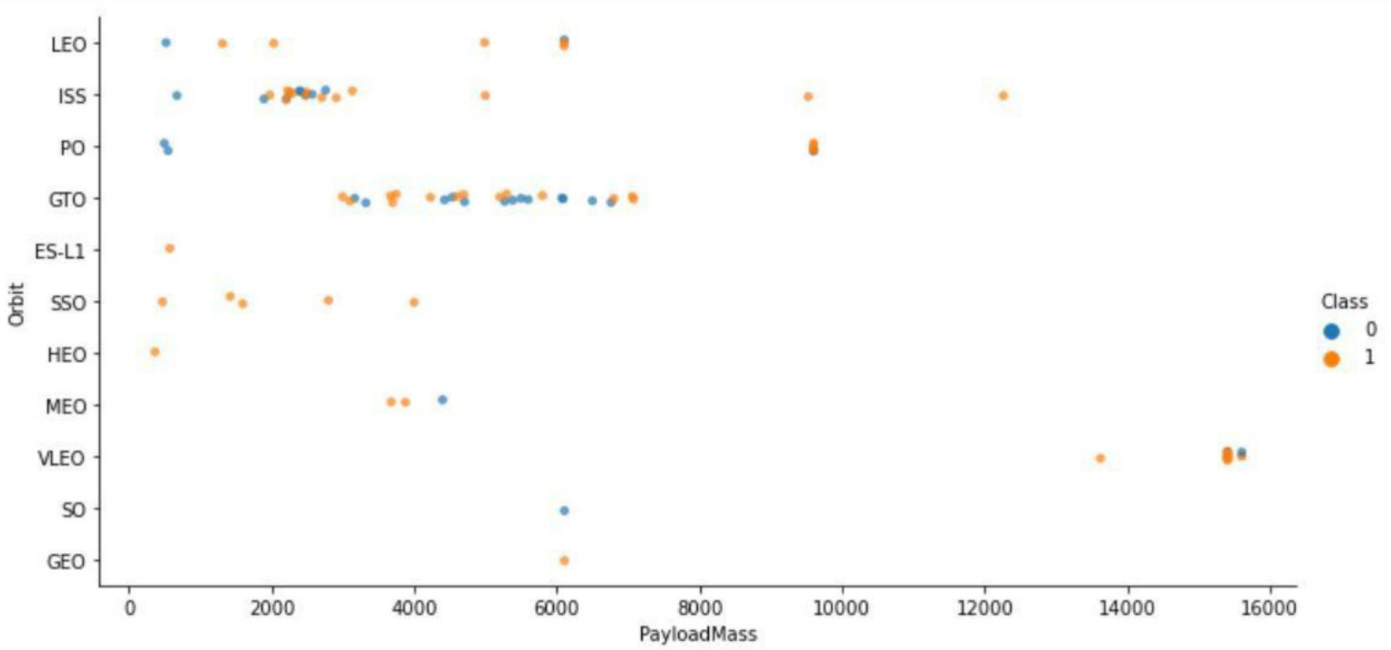
Success Rate vs. Orbit Type



Flight Number vs. Orbit Type

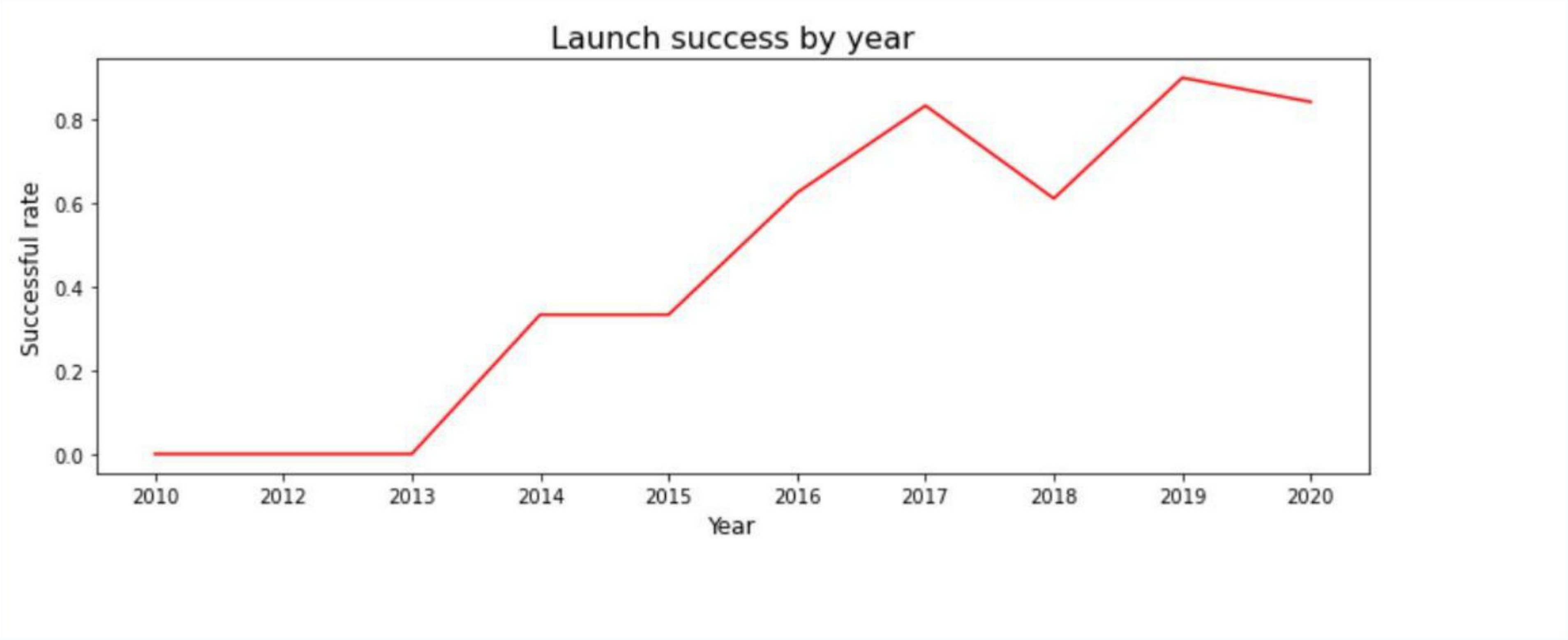


Payload vs. Orbit Type



Payload vs. Launch Outcome

Launch Success Yearly Trend



All Launch Site Names

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

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

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

There are 4 launch sites

Launch Site Names Begin with 'CCA'

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

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

Time JTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
5:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
3:00	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)
4:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
5:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
0:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

It represents only faliures

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload FROM SPACEXTBL WHERE Customer LIKE 'NASA (CRS)';
```

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

total_payload
45596

Total payload mass = 45590

Average Payload Mass by F9 v1.1

```
%sql SELECT avg(PAYLOAD_MASS__KG_) AS Avg_Payload FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1';
```

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

Avg_Payload
2928.4

average payload mass = 2928.4

First Successful Ground Landing Date

```
%sql SELECT min(date) AS Early_Date from SPACEXTBL where Landing_Outcome LIKE 'Success (ground pad)'
```

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

Early_Date
01-05-2017

The first landing success was on 1/05/17

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Customer, Landing_Outcome,PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE Landing_Outcome ='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Customer	Landing_Outcome	PAYLOAD_MASS__KG_
SKY Perfect JSAT Group	Success (drone ship)	4696
SKY Perfect JSAT Group	Success (drone ship)	4600
SES	Success (drone ship)	5300
SES EchoStar	Success (drone ship)	5200

most successful landing is on the drone ship.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, Count(*) AS Numbers FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

Mission_Outcome	Numbers
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

There is 1 failure in the flights and 99 successes.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version, Max_Payload FROM (SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_)
                                                AS Max_Payload FROM SPACEXTBL GROUP BY Booster_Version)
```

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

Booster_Version	Max_Payload
F9 B4 B1039.2	2647
F9 B4 B1040.2	5384
F9 B4 B1041.2	9600
F9 B4 B1043.2	6460
F9 B4 B1039.1	3310
F9 B4 B1040.1	4990
F9 B4 B1041.1	9600
F9 B4 B1042.1	3500
F9 B4 B1043.1	5000
F9 B4 B1044	6092

Different boosters have different payload capacity

2015 Launch Records

```
%sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,7,4) = '2015'
```

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

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

2015 there are launch failure by booster B1012 and B1015.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
%sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN '04-06-2010' AND '20-03-2017'
GROUP BY Landing_Outcome ORDER BY Numbers DESC;
```

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

Landing_Outcome	Numbers
Success	20
Success (drone ship)	8
Success (ground pad)	6

There are in total

- 20 successful landings
- 8 successful drone ship landings
- 6 successful ground pad landing

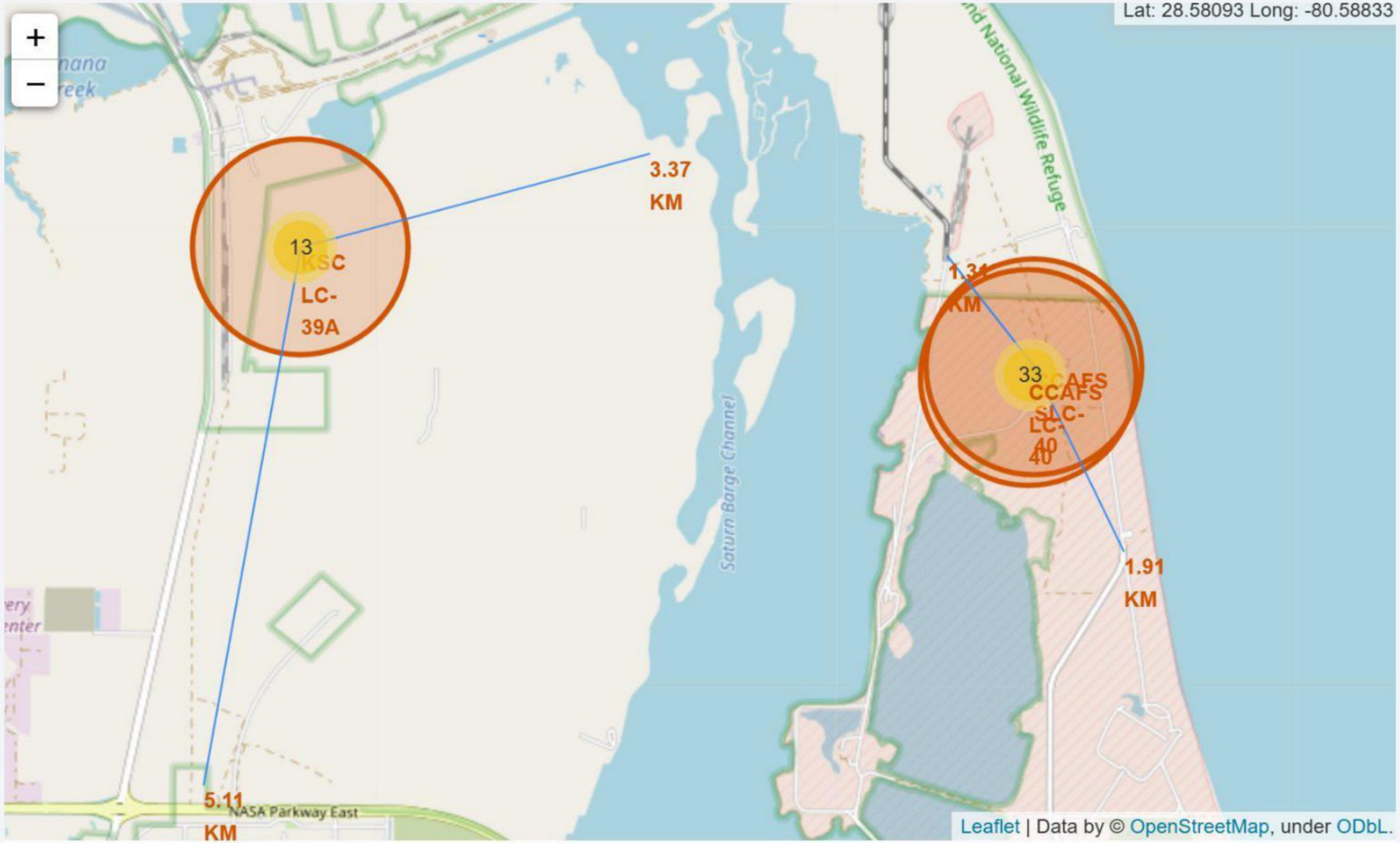
Section 3

Launch Sites Proximities Analysis

launch sites

Launch outcome of different site

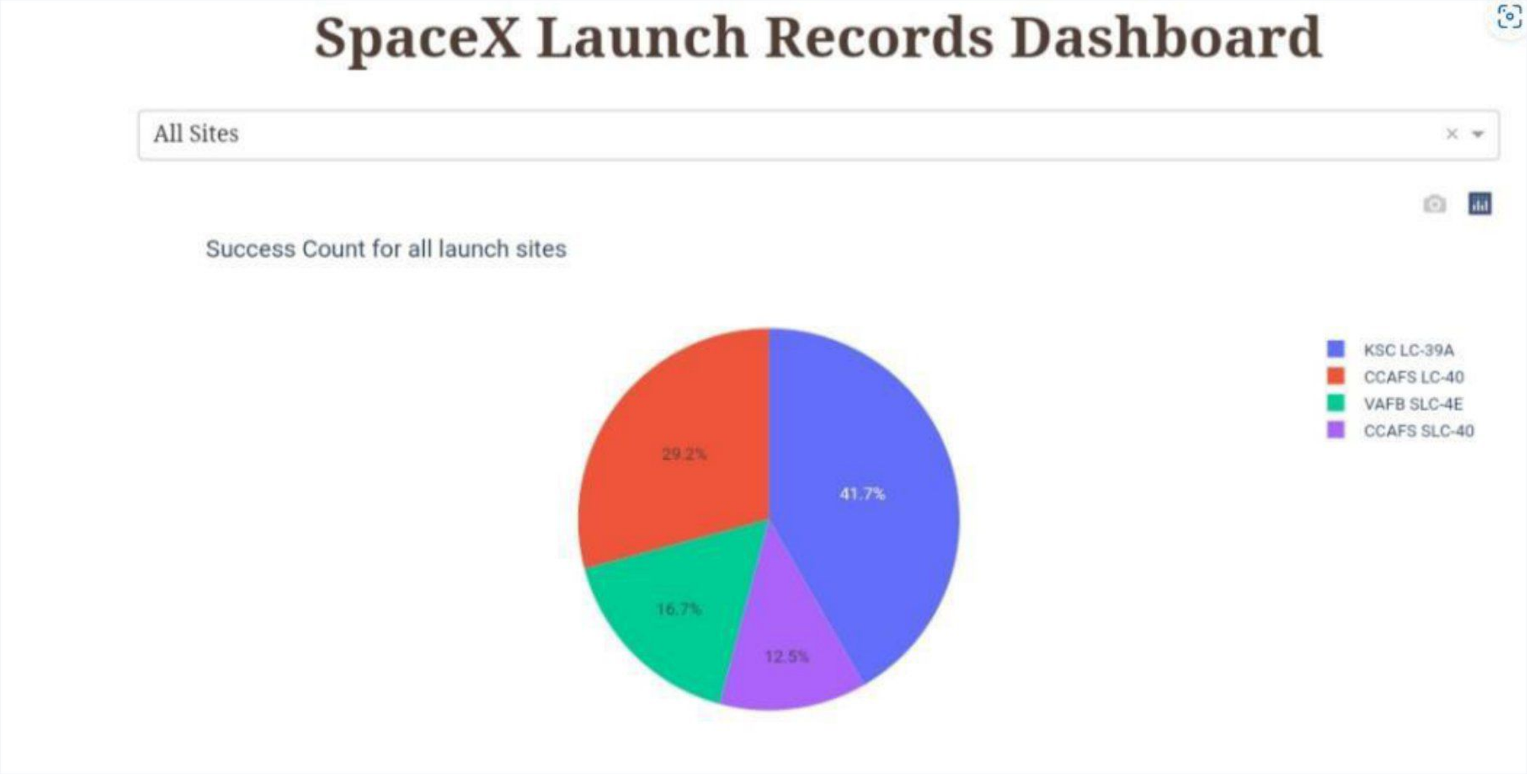
The proximity of the launch sites



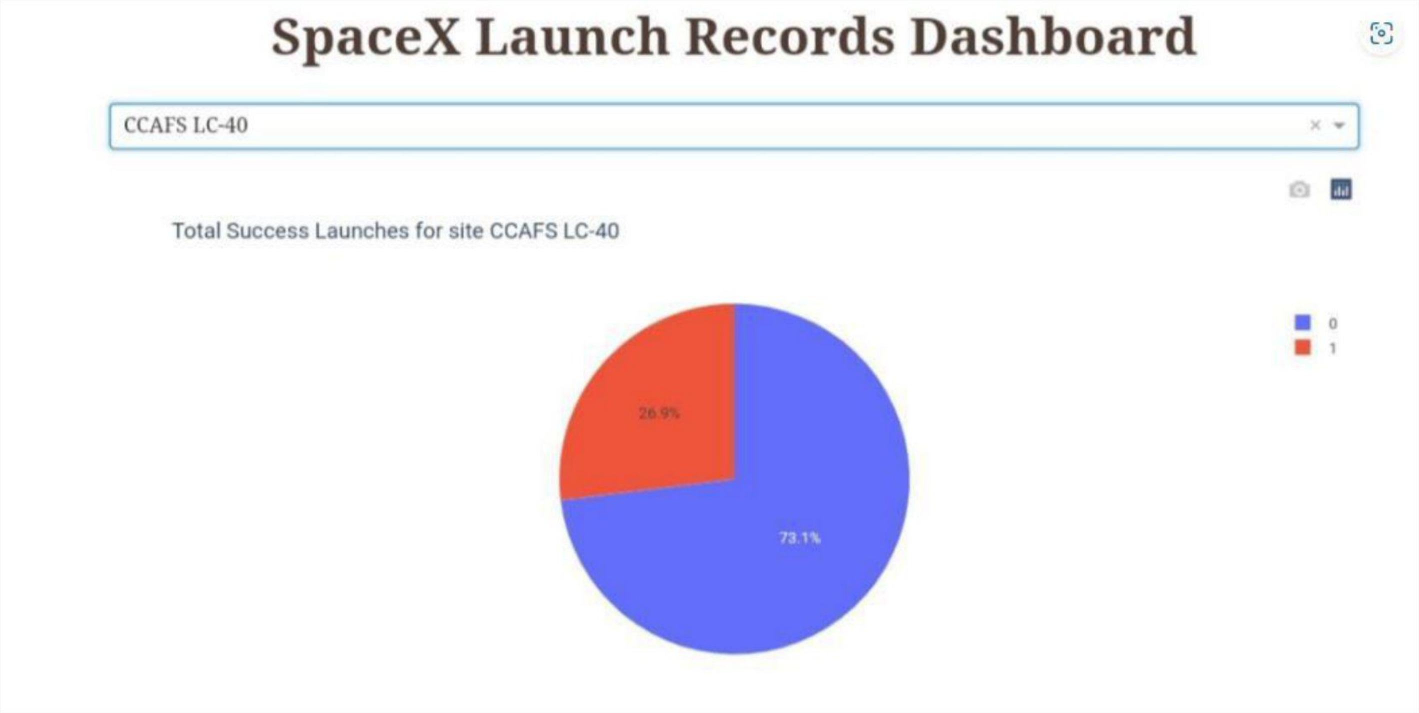
Section 4

Build a Dashboard with Plotly Dash

All site launch



Highest success launch ratio

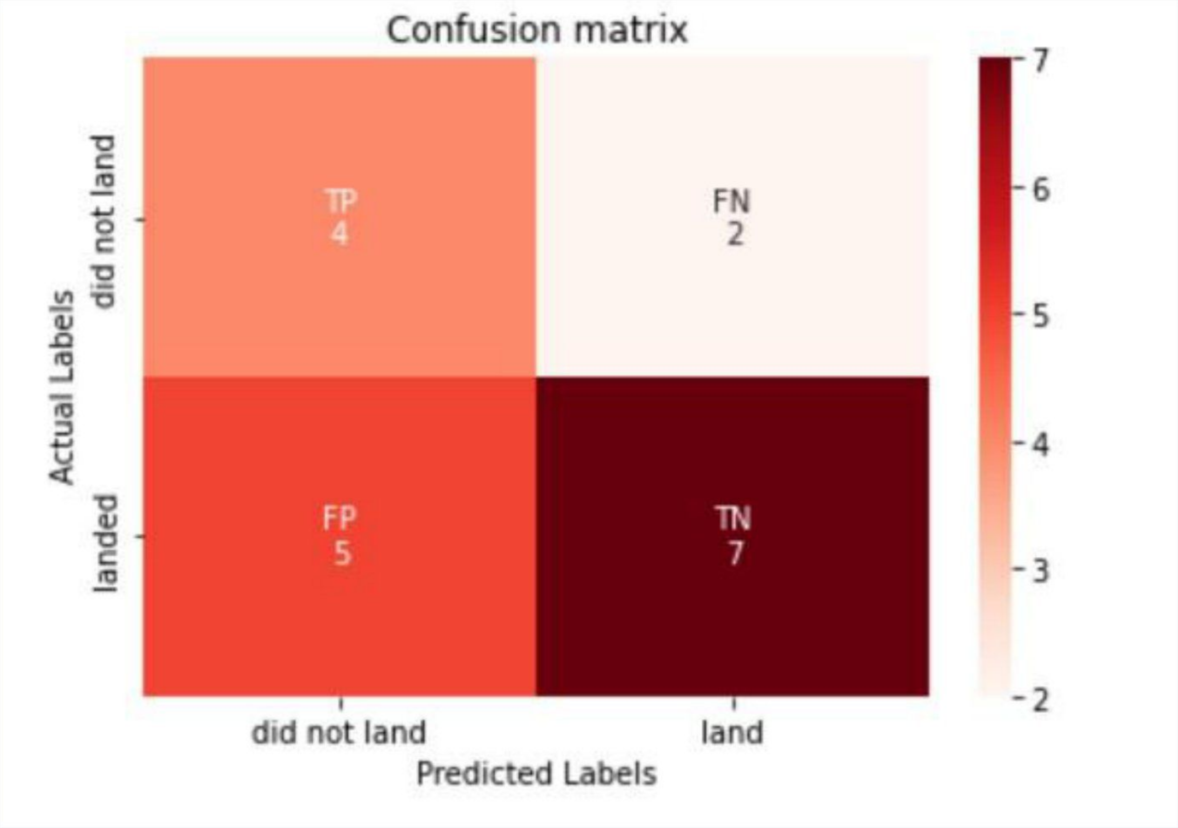


Section 5

Predictive Analysis (Classification)

Classification Accuracy

Confusion Matrix



Conclusions

- The more greater the payload, the less likely the first stage will return.
- Orbits-SO has the least success rate while ES-L1, GEO, HEO and SSO have the highest success rate There has been an increase in the success rate since 2013 kept increasing till 2020
- With best parameter provided, all revealed an accuracy of 83%. .

Appendix

- https://github.com/JC-project-works/public

Thank you!