IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Phillemon
2022/05/17

Outline

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

Executive Summary

This project analyze the Space X Falcon 9 landing for the success and failure throughout the sources that affect the success landing rate. Machine models are being used for classification and prediction of the failure.

- Summary of methodologies
    - There are various Machine Learning methods being used.
        - The data collection using the API and the Web Scrapping
        - Data Cleaning by and wrangling (Shaping our data)
        - EDA performed using SQL and Python from IBM cloud packs and DB2
        - Building Models for predicting the failure rate and success rate on possible factors.
- Summary of all results
        - EDA Plotly charts visuals
        - Modelling and evaluation of the predictive modeling.

Introduction

- Project background and context
  - The failure to land of the Falcon 9 has a high impact of loss. The Falcon launch cost almost $62m. The project use machine learning techniques to predict the failure and success rate of the Falcon 9 landing. The cost of successful landing can be determined only if the Falcon 9 landed successfully.
- Problems you want to find answers
  - What are the factors affecting the success and failure to land of the falcon 9.
  - Feature made impact in a successful landing.
  - The ML models that has the higher prediction of failure and success.
  - Most cost-efficient factors.

Section 1

Methodology

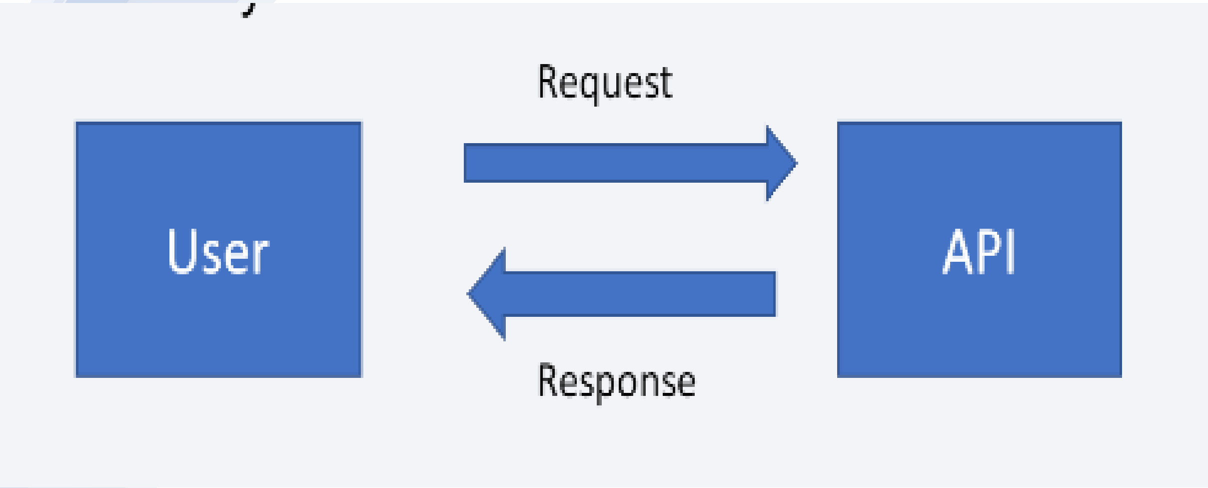
Methodology

Executive Summary

- Data collection methodology:
  - The data has been collected using the Space X API.
  - The web scrapped data using Beautiful soup has been used to collect data on the website.
- Perform data wrangling
  - The data contained not a number and it has been replaced using
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

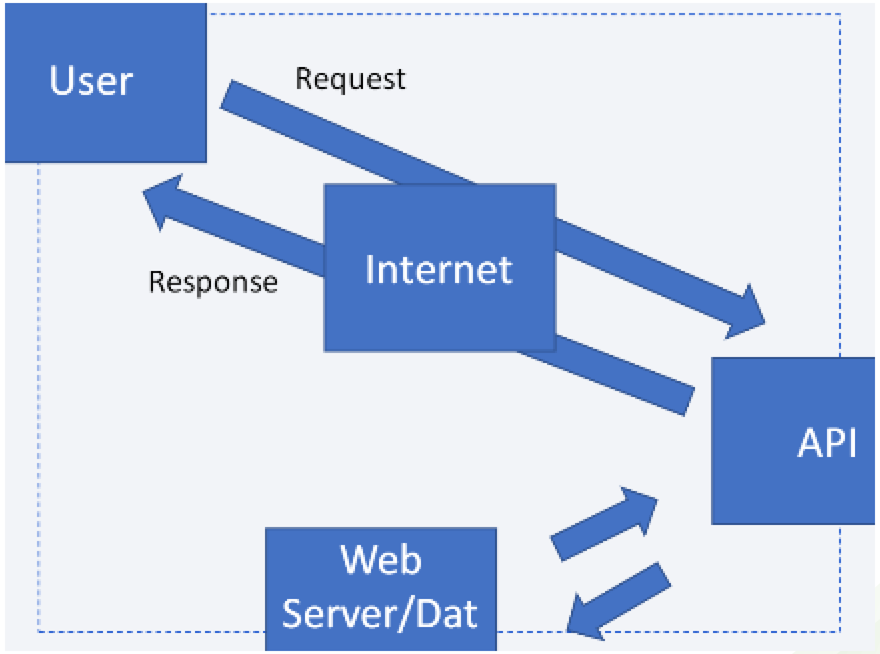
- Describe how data sets were collected.
    - The library “Request” was used for data collection as get request for the Space X API.
    - The Dataset extracted using API, was with features launch site.
- You need to present your data collection process use key phrases and flowcharts
    - All the data extraction process (data collection) has been presented below:
        - Data Collection – SpaceX API
        - Data Collection – Web Scaping



Data Collection – SpaceX API

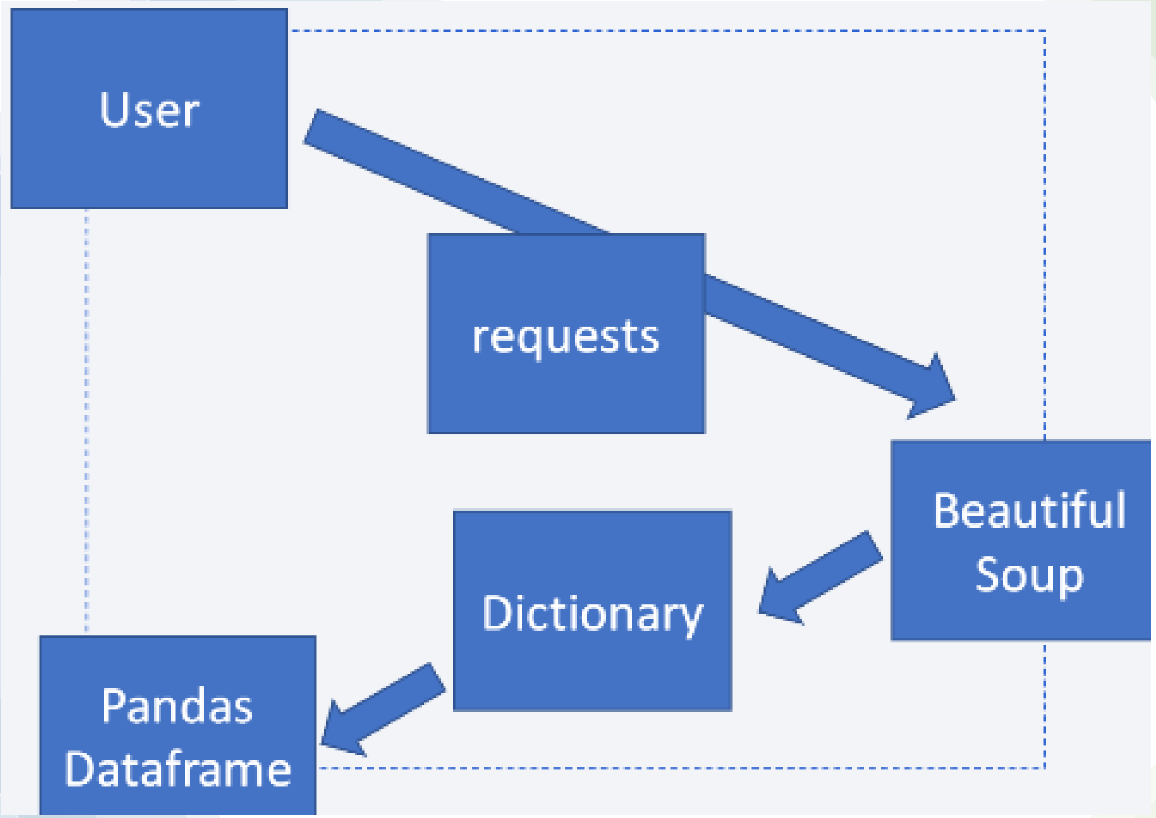
- The Get request from the ‘request’ module by the help of Space X API.

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- HTML link have been used to extract the data.
- The Jason type Dictionary file have been used after extracting data using Beautiful soups.
- The data gets converted to a pandas data frame from the dictionary.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

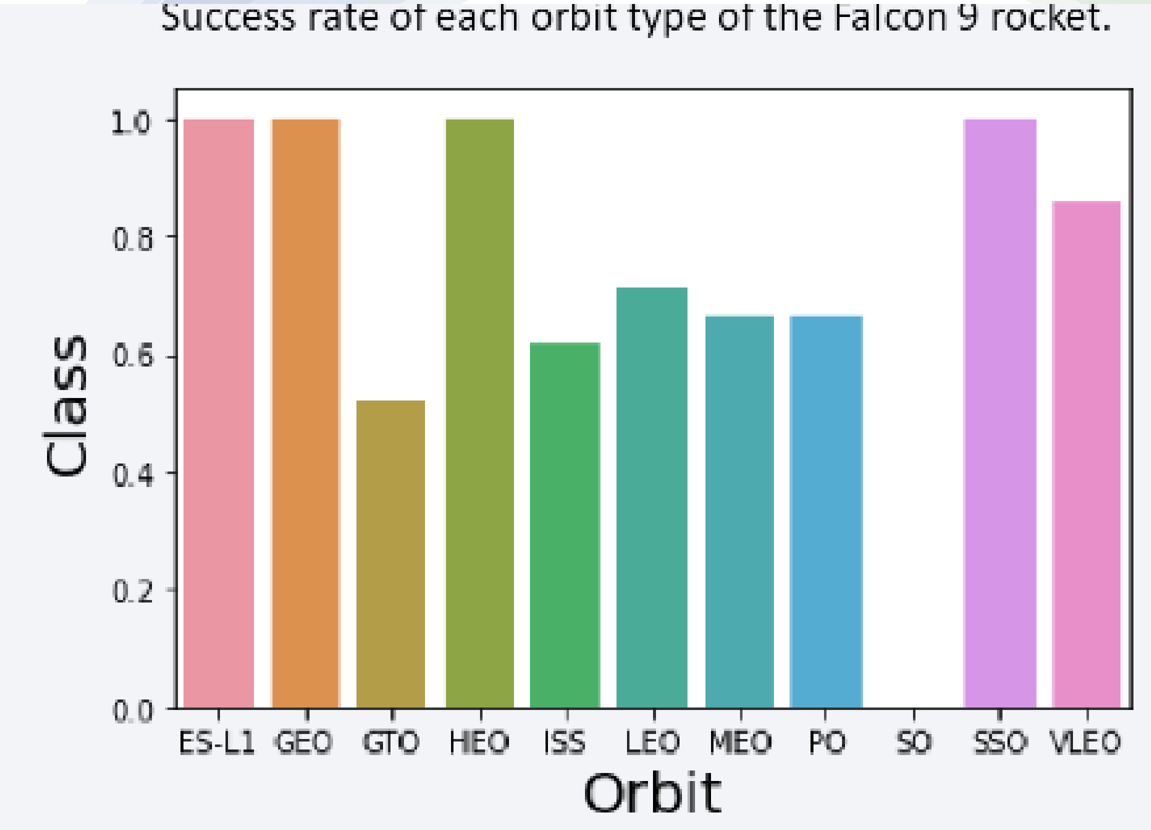
- The removal of null values.
- One hot encoding was used to transform the categorical data so that the models will be able to perform predictions

```
df.isnull().sum()
```

```
FlightNumber       0
Date               0
BoosterVersion     0
PayloadMass        0
Orbit              0
LaunchSite         0
Outcome            0
Flights            0
GridFins           0
Reused             0
Legs               0
LandingPad        26
Block              0
ReusedCount        0
Serial             0
Longitude          0
Latitude           0
dtype: int64
```

EDA with Data Visualization

- The seaborn (sn) library was used to perform the data visualizations.,
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose



EDA with SQL

- The SQL IBM DB2 was used to select the multiples views.

- Distic launch sites
- Average payload

```
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']=[]
```

Next, we just need to fill up the `launch_dict` with launch records extracted from table rows

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- The machine learning algorithms,
  - Logistic regression
  - Decision tree
  - SVM
  - KNN

  - The data has been split in to 20% test and 80% training
  - The s

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

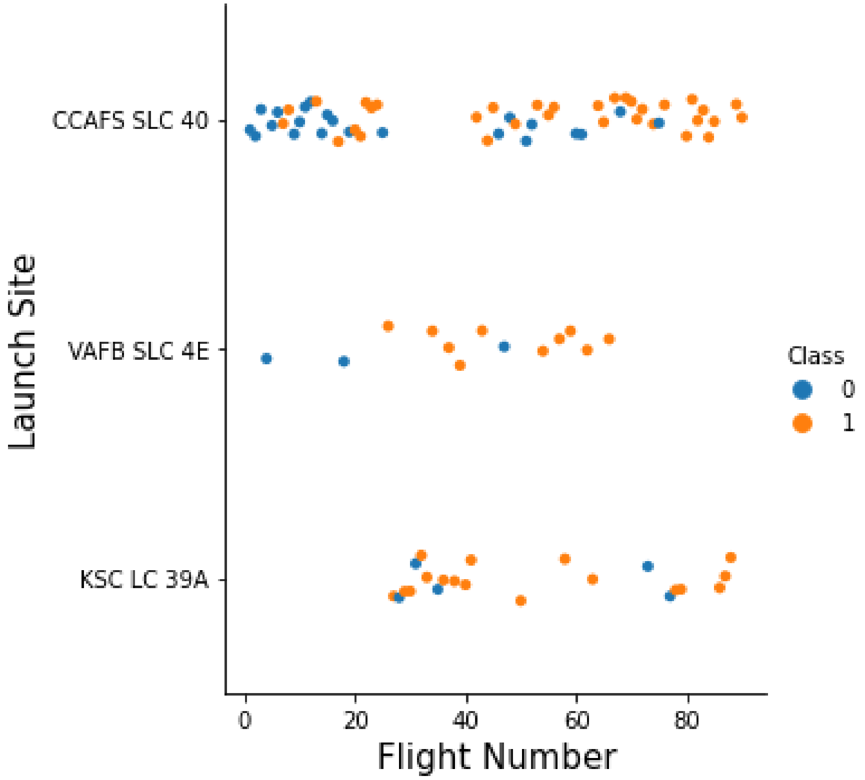
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

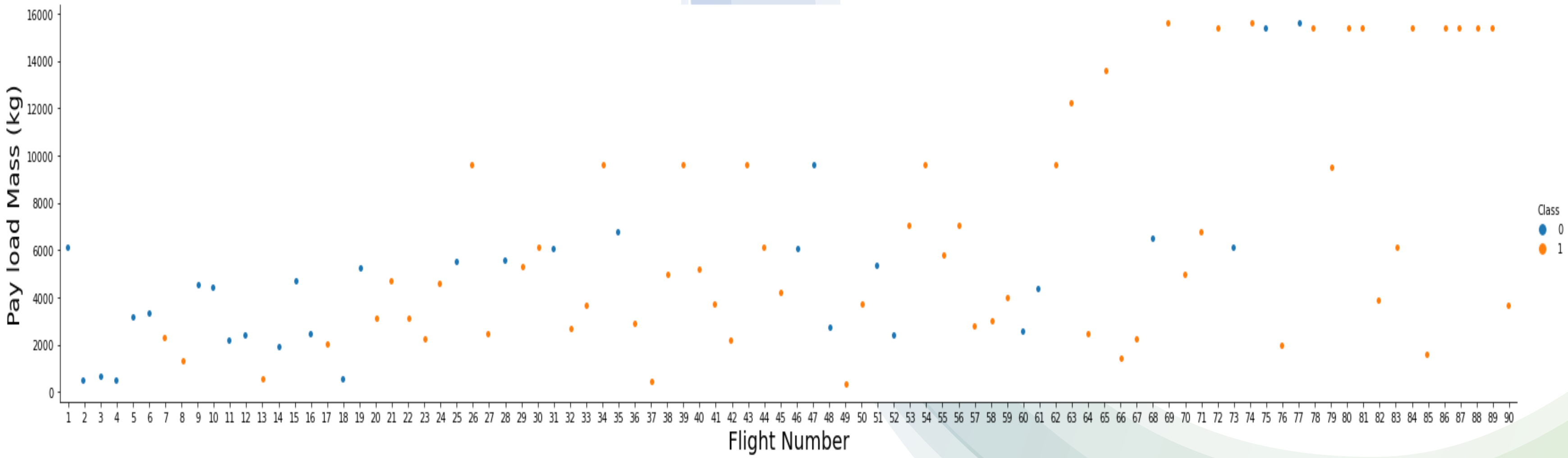
- Show the screenshot of the scatter plot with explanations



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

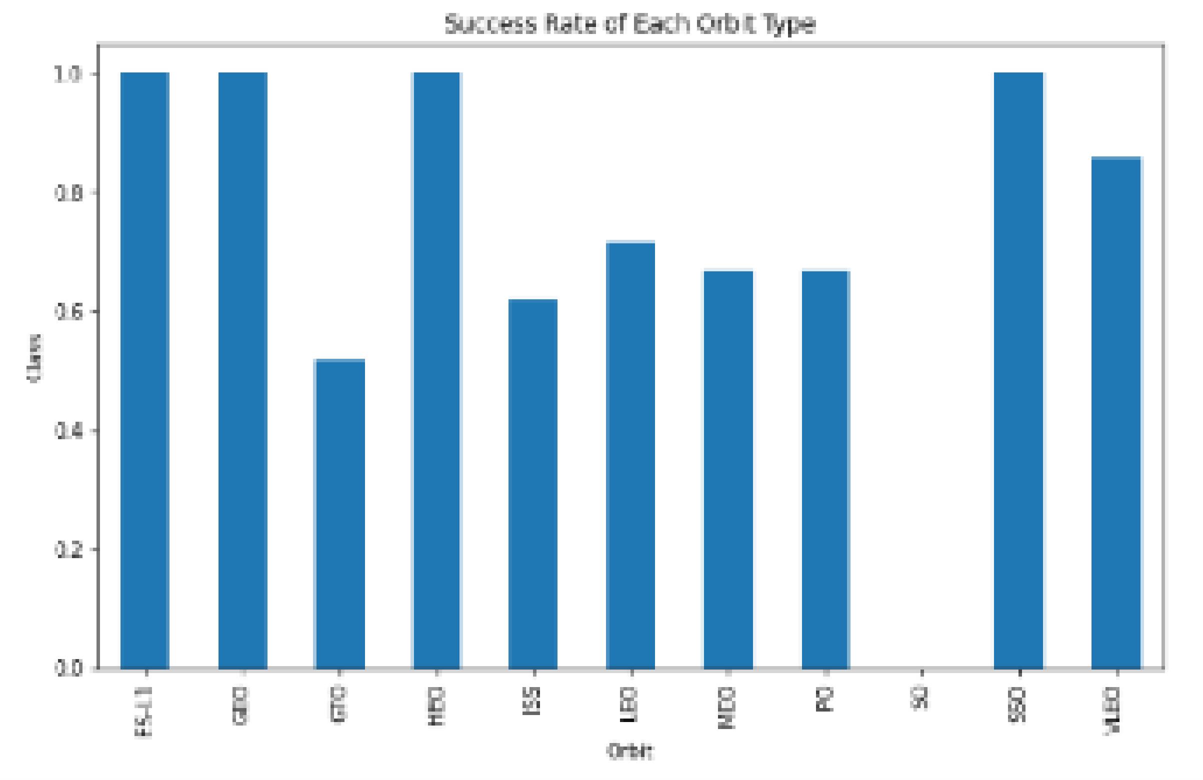
- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

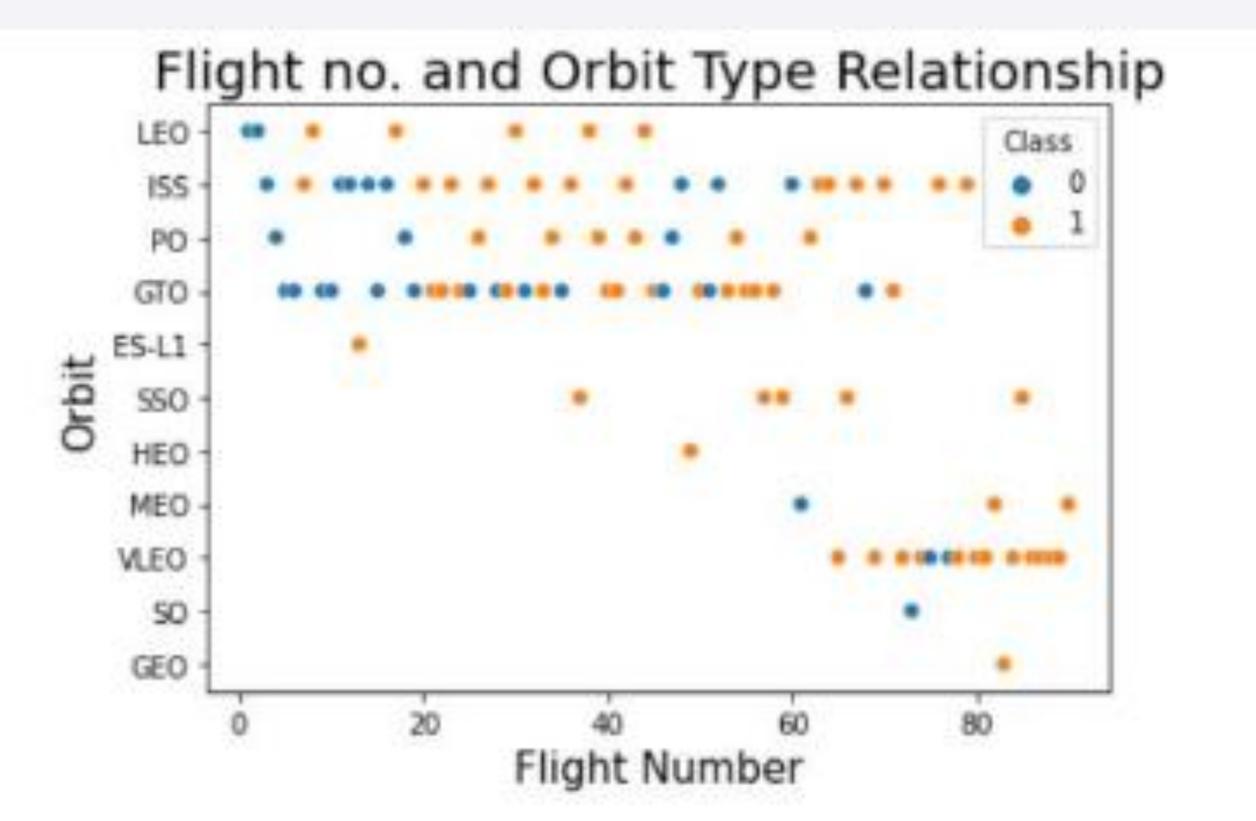
- Show the screenshot of the scatter plot with explanations



Flight Number vs. Orbit Type

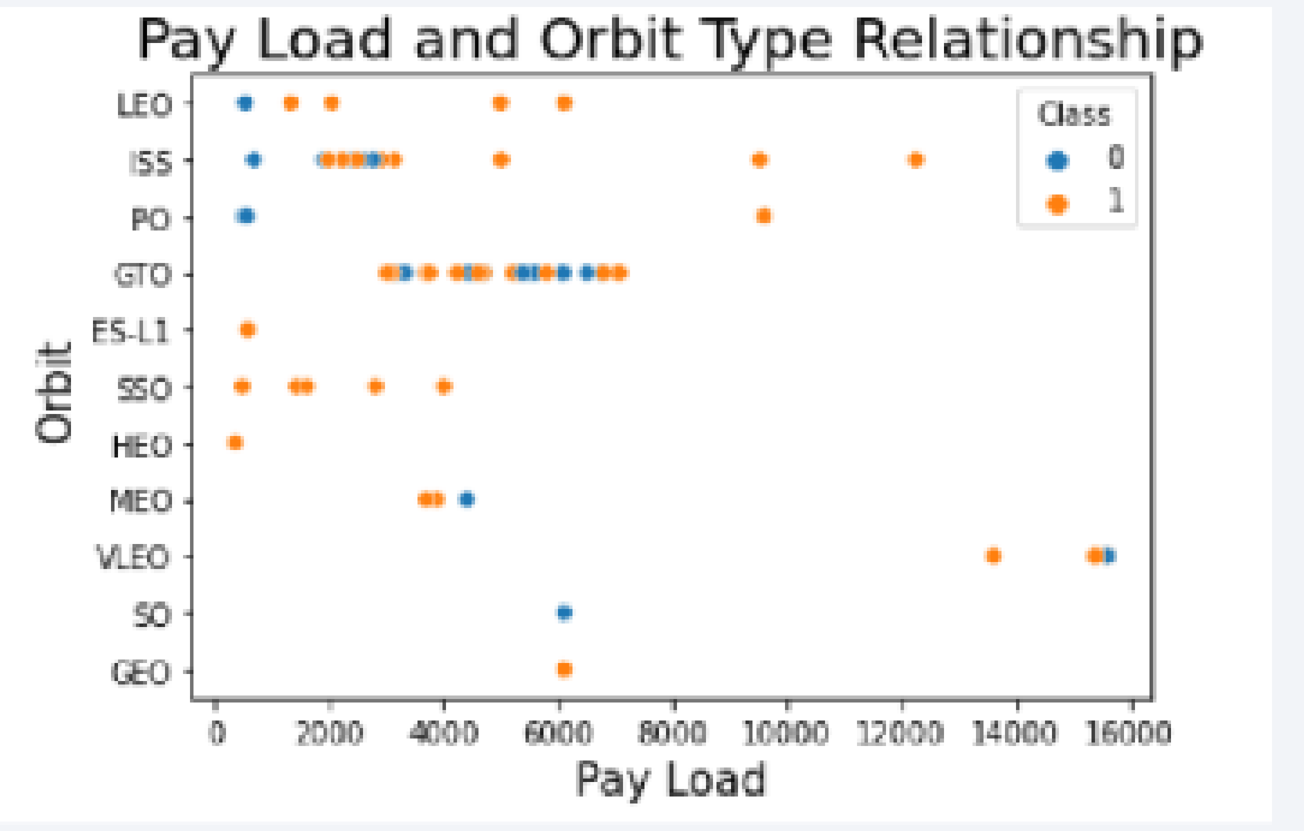
- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations



Payload vs. Orbit Type

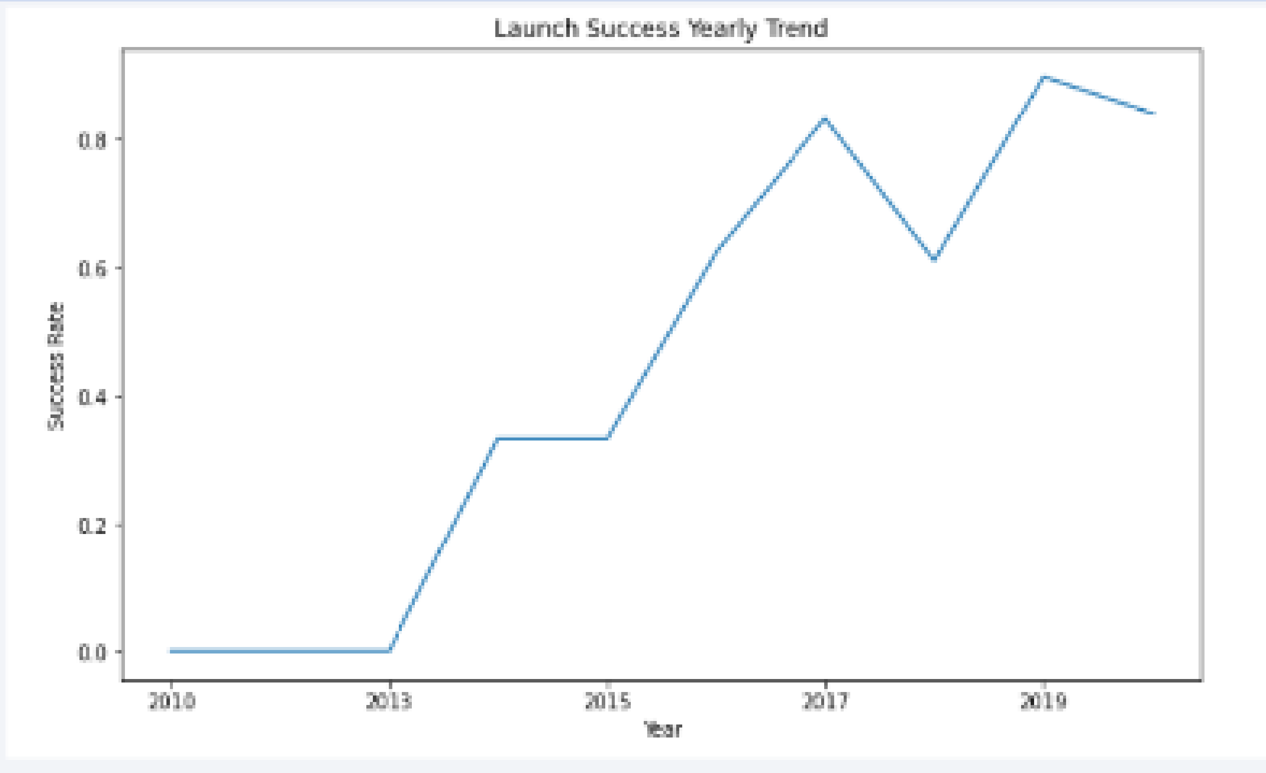
- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- SQL have been used to ectraxt from the loaded Space X database

- SELECT DISTINCT(Launch_Site) FROM Spacextbl

- There were four Unique launch sites from the database data

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

Launch Site Names Begin with 'CCA'

- The SQL Query:

Written from the Jupiter notebook

SELECT * FROM spacextbl

WHERE launch_site LIKE “CCA%”

LIMIT 5

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- 45596 Kg.
- The SQL Statement used.
- %sql

SELECT SUM(pay;load_mass_kg_)

FROM spacetbl

WHERE customer = “NASA (CRS)”

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- %sql
- SELECT AVG(payload_mass_kg_) FROM spoacextbl
- WHERE booster_version = “F9v1.1”

- The averga payload mass : 2929 Kg.

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql
SELECT MIN(DATE)
FROM spacextbl
WHERE landing_outcome = “Success (ground pad)”
```

min_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

%sql

Select booster_version

FROM spacextbl

WHERE (PAYLOAD_MASS_KG_) >4000 and (PAYLOAD_MASS_KG_) <6000

Landing_outcome – 'success (drone ship)'

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

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

- 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

- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

Folium Launch locations

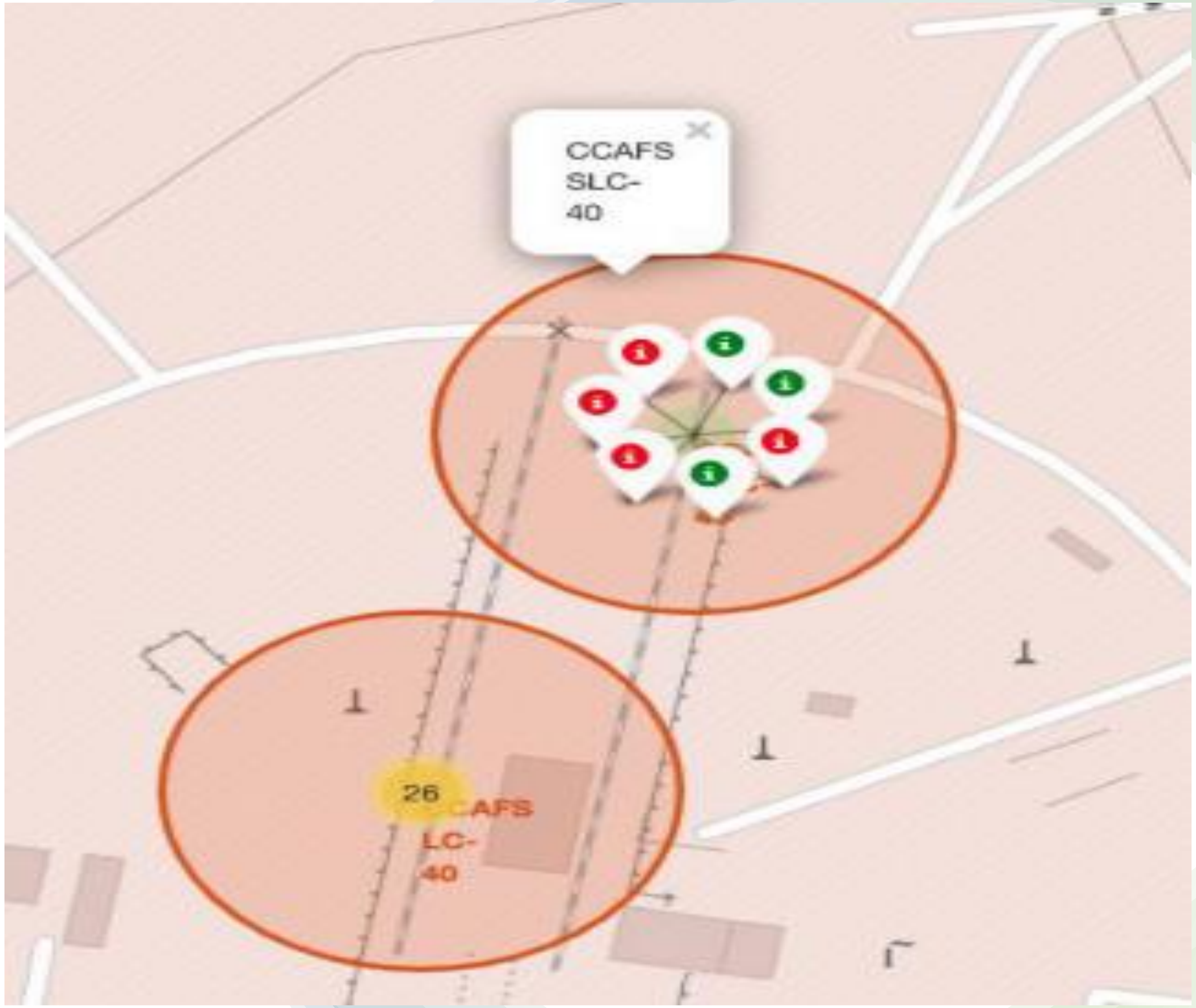
- Presents all the three location in the USA where there is a launch site

Success and Failus loactions

- The green represents the launch been mad and become successful
- The red presents all the success landing locations

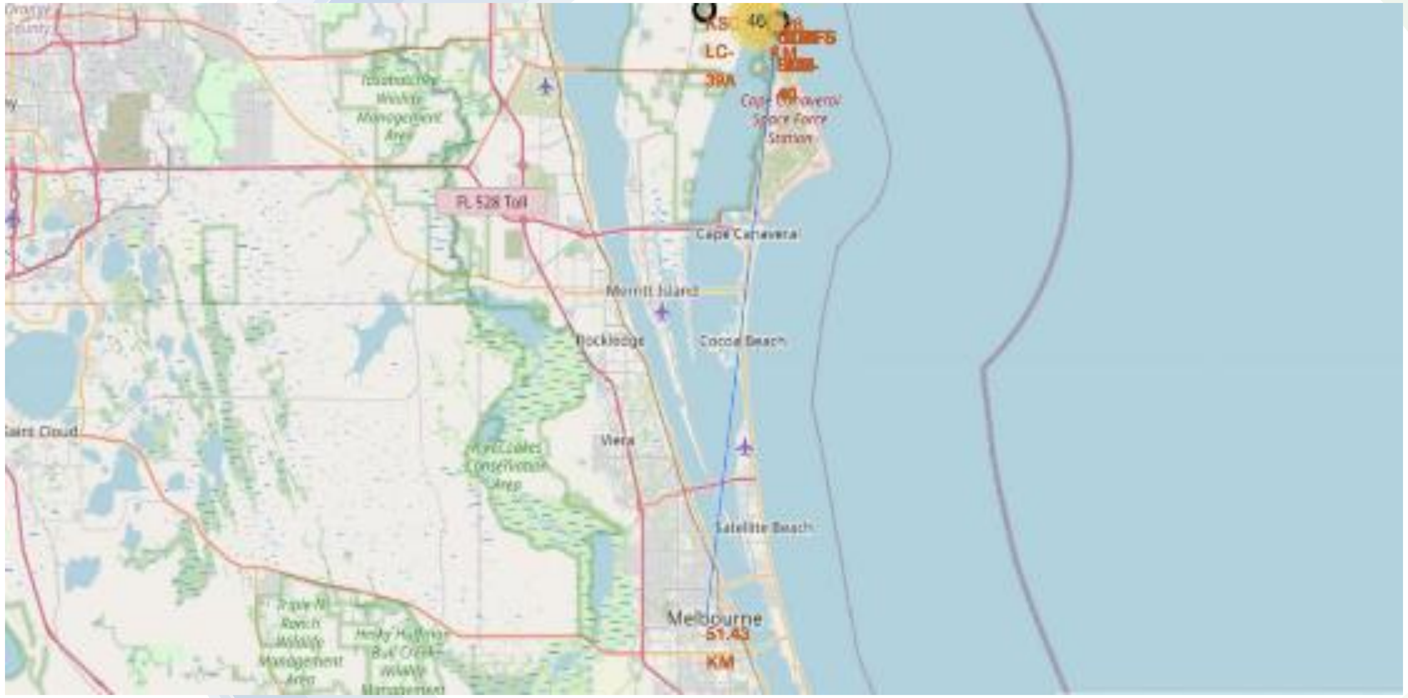
Fig 4 – CCAFS SLC-40 success/failed markers

Fig 5 – CCAFS SLC-40 success/failed markers



Launch Site Distances

- The following map shows how far each launch and landing point is from the city

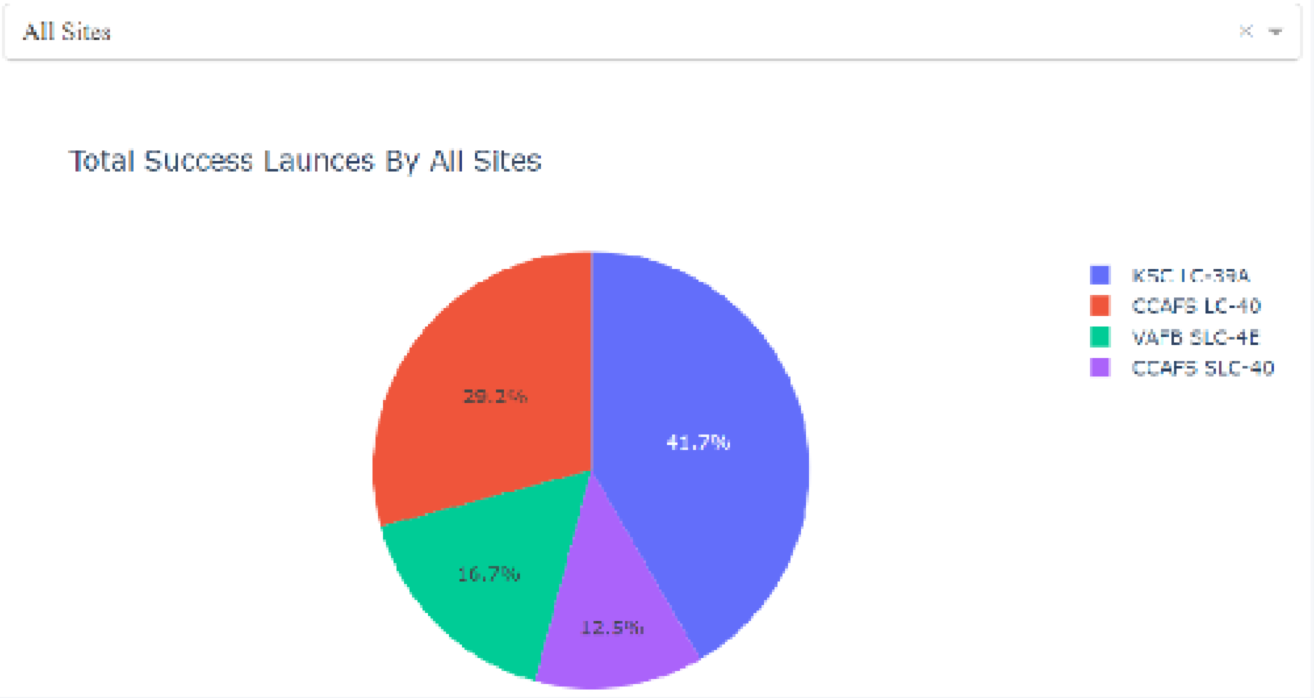


Section 4

Build a Dashboard with Plotly Dash

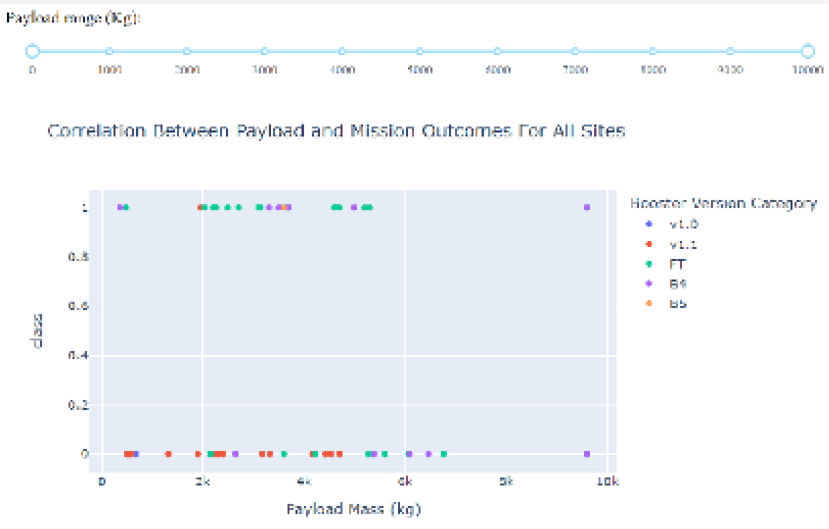
Success lauch sites

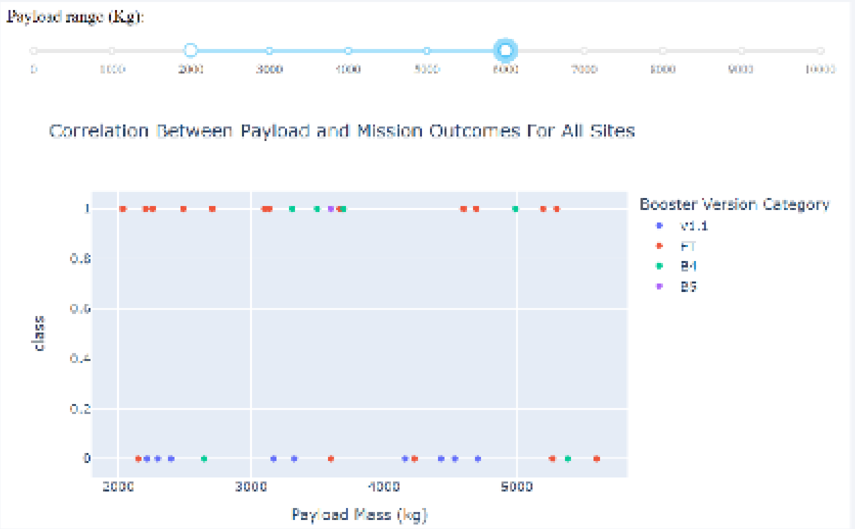
- KSC LC-39A presents the location with the highest launch success rate.
- CCAFS SLC-40 Shows the lowest rate of success.



Payload vs launch Outcome scatter plot for all sites

- Figure below shows the most successful rate range between 2000 and 55000
- The Booster version FT has the most successful lunches,





Section 5

Predictive Analysis (Classification)

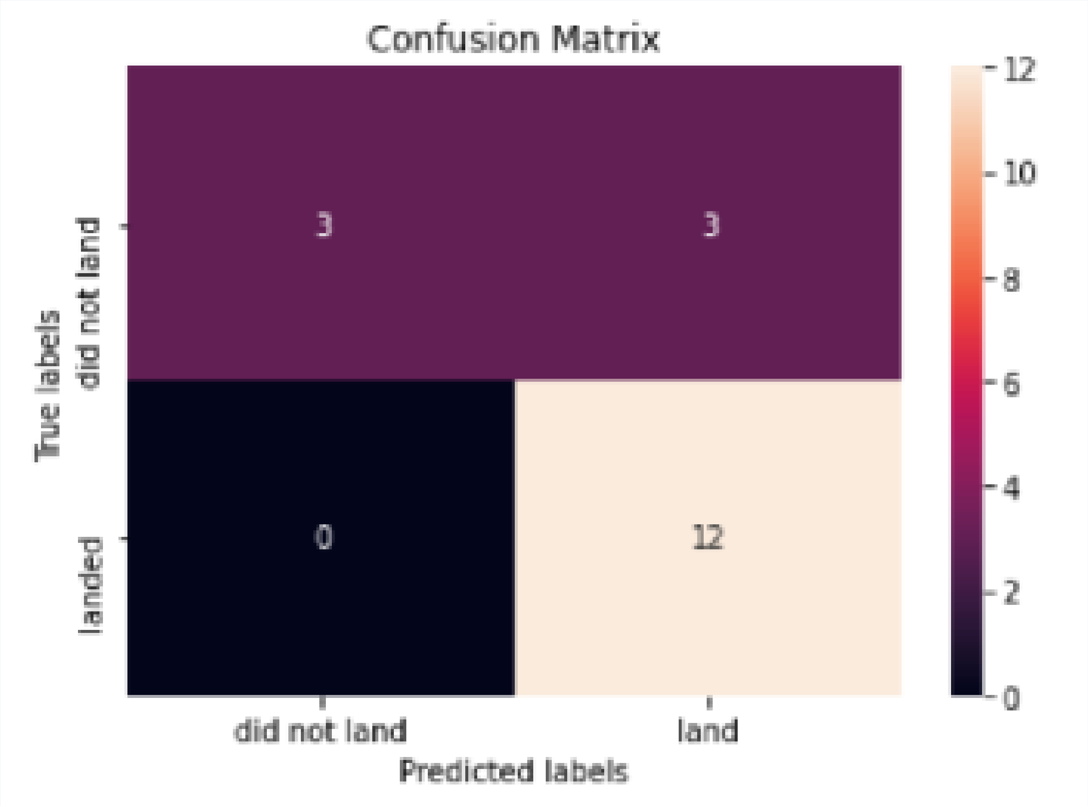
Classification Accuracy and Mattrix

- The logistic regression - 83%
- Support vector machine – 83%
- K-nearest neighbor – 84&

Summary:

The confusion matrix shows the success rate of 83%

The model is accurate in a slightly manner.



Conclusions

- The payloads with lower mass has more success than the one with the more masses.
- The LC-39A site had the most successful landing.
- It is obvious from the maps that the launches site are near the sea and the distances are far apart.
- The success rate of the launched fif increase to at least 80% between 2013 and 14.
- ...

Thank you!