IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammet Emin Yaylı
16/10/2023

Outline

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

Executive Summary

METHODOLOGIES

Data Collection: Data is collected from the SpaceX REST API and terough web scrapping the SpaceX site.

Data Wrangling: The raw data from the SpaceX REST API and web scraping was cleaned and structured for analysis.

Visualization: The dataset was analysed for relationships and dependencies between features.

Predictive Analytics: *The data is split into training and testing sets. Multiple machine learning models are considered to work on the data.*

RESULTS

*Launching success rate increasing from 2013 till 2020.

*Although several classification models achieved an accuracy of 83.3%, the Decision Tree stood out by delivering a slightly superior F1 score.

Introduction

- According to SpaceX's website, the cost of a Falcon 9 rocket launch is 62 million dollars, which is much cheaper than other providers that cost upward of 165 million dollars each. This is mainly due to SpaceX's ability to reuse the first stage. If we can accurately predict whether the first stage will land or not, we can determine the cost of a launch. This information could be really useful for an alternate company that wants to bid against SpaceX for a rocket launch.
- Verifying the successful touchdown of Falcon 9's first stage.
- Is it possible to precisely ascertain the launch cost?"

Section 1

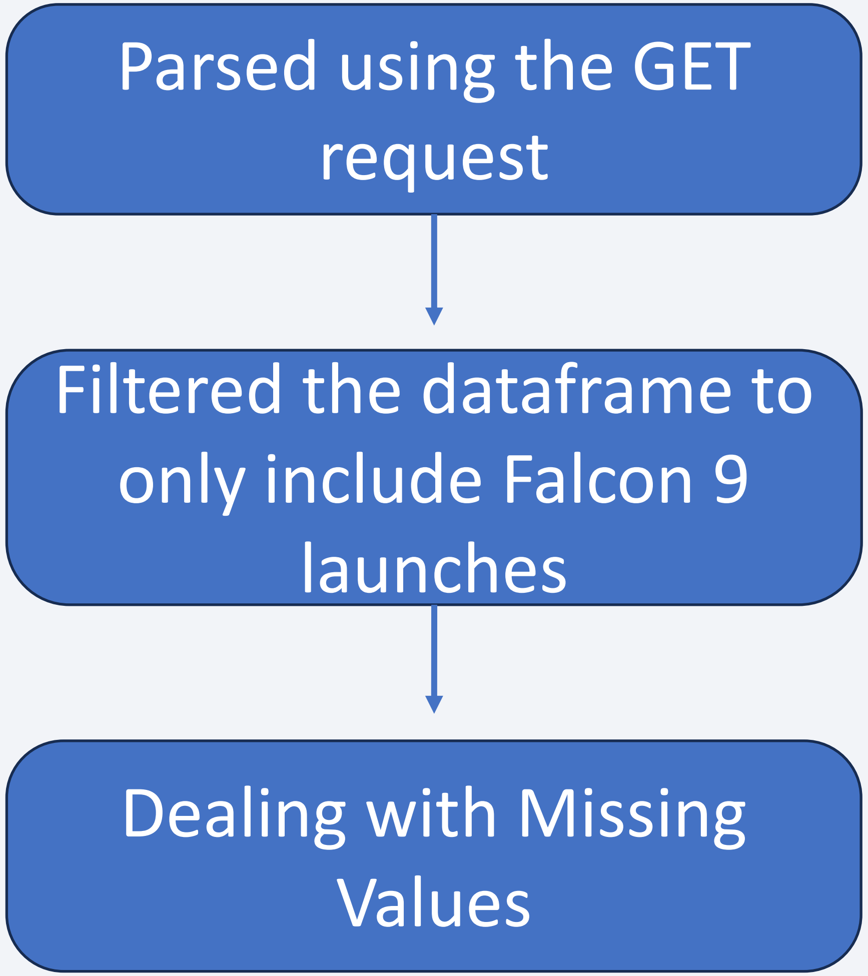
Methodology

Methodology

Executive Summary

- Data collection methodology:
- Data is collected from the SpaceX REST API and SpaceX site.
- Perform data wrangling
- Data cleaning, data structuring, datatransforming ,data integrating.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Use machine learning models to fit the training data.

Data Collection



Data Collection – SpaceX API

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

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

```
# Use json_normalize meethod to convert the json result into a dataframe
a=response.json()
data=pd.json_normalize(a)
```

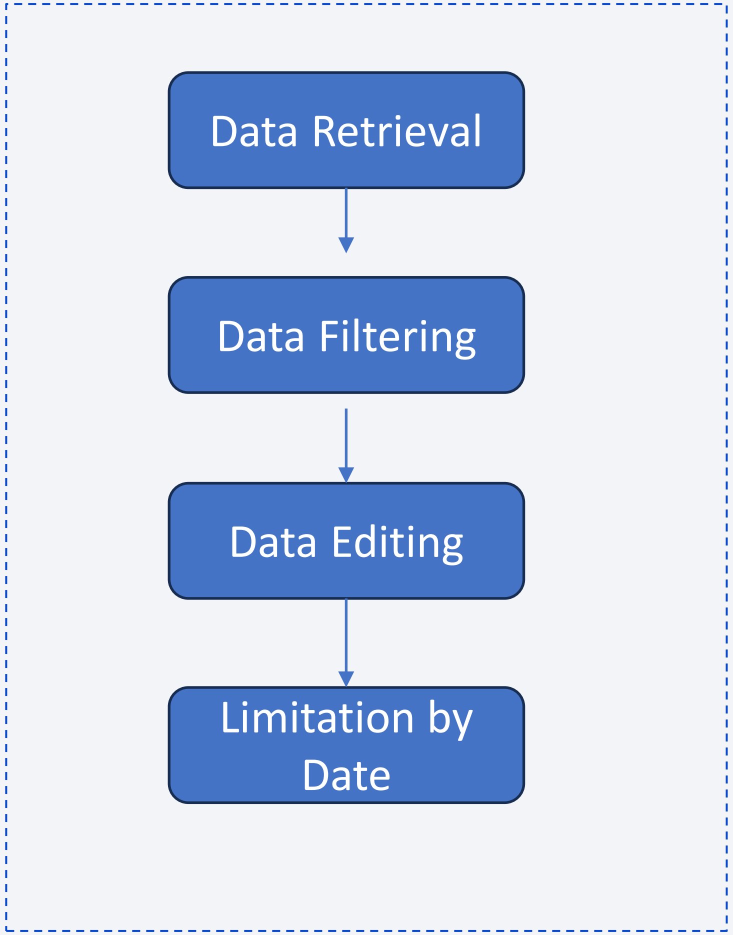
```
# Lets take a subset of our dataframe keeping only the features we want and the flight nu
data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']]

# We will remove rows with multiple cores because those are falcon rockets with 2 extra r
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]

# Since payloads and cores are lists of size 1 we will also extract the single value in t
data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])

# We also want to convert the date_utc to a datetime datatype and then extracting the dat
data['date'] = pd.to_datetime(data['date_utc']).dt.date

# Using the date we will restrict the dates of the launches
data = data[data['date'] <= datetime.date(2020, 11, 13)]
```

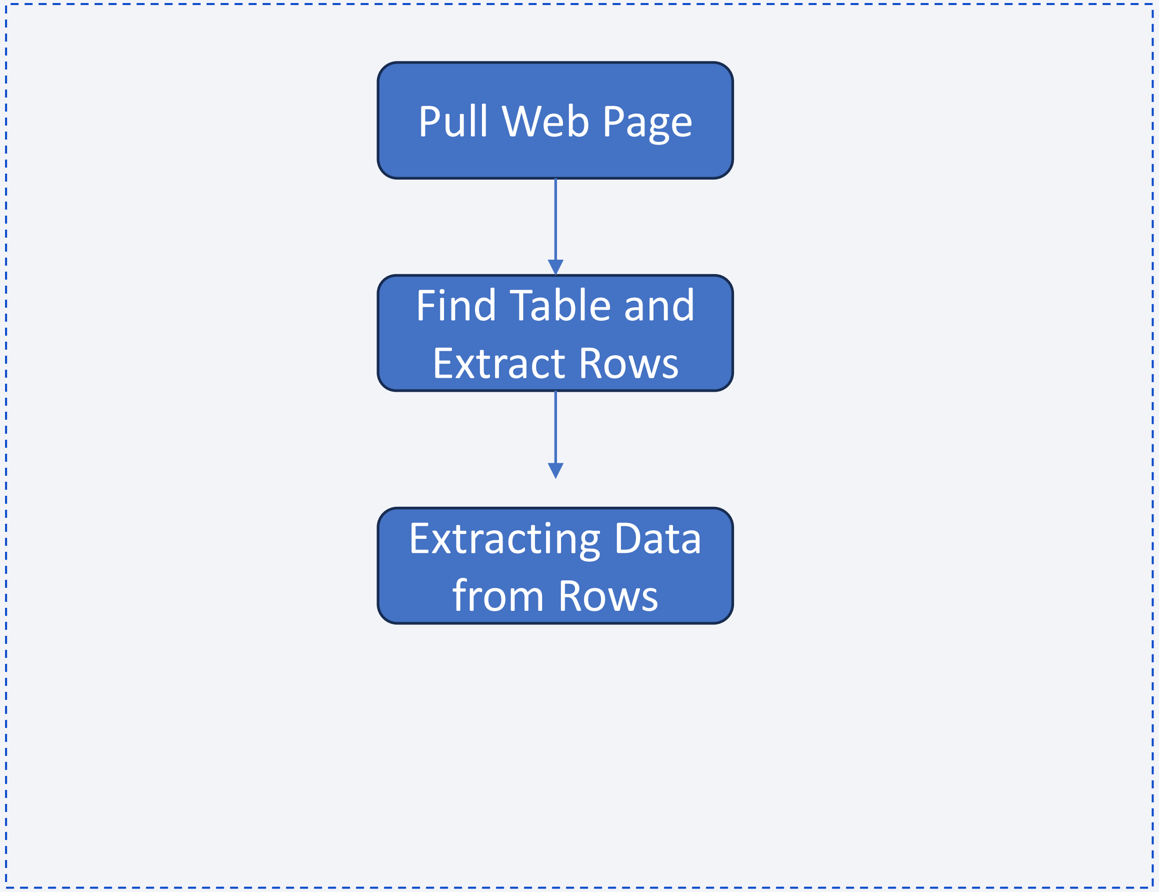


Data Collection - Scraping

```python
q=requests.get(static_url)
q=q.text

soup=BeautifulSoup(q,'html.parser')
```

```python
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launch a number
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()
        else:
            flag=False
```



Git-Hub Links

- https://github.com/muhammeteminyayli/ibm_data_science/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
- https://github.com/muhammeteminyayli/ibm_data_science/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

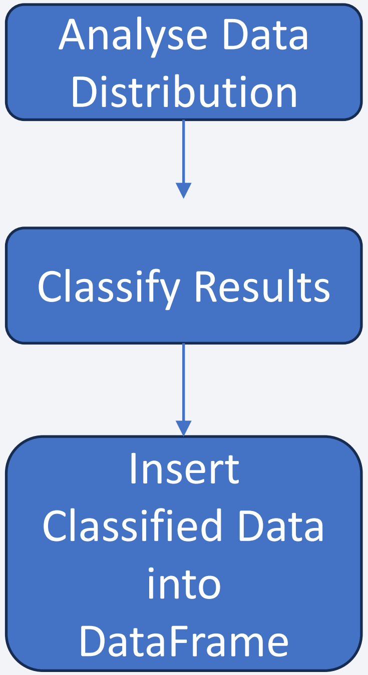
```
a=pd.value_counts(df["LaunchSite"])

b=pd.value_counts(df['Orbit'])

landing_outcomes = pd.value_counts(df["Outcome"])

landing_class = [0 if outcome in bad_outcomes else 1 for outcome in df['Outcome']]

df['Class']=landing_class
```



EDA with Data Visualization

- **Flight Number and Launch Site**:bar plot
- **Payload and Launch Site:** Scatter plot
- **Success Rate vs. Orbit Type:** bar plot
- **FlightNumber and Orbit type:** Scatter plot
- **Payload and Orbit type:**Scatter
- **Visualize the launch success yearly trend:** line plot
- Scatter plot is a type of graph used to show the relationship between two variables.
- A bar plot is a type of graph used to visualise the values of categorical data.

EDA with SQL

The following are various aspects of Exploratory Data Analysis using SQLDisplay of the names of unique launch sites in the space mission

- Displaying 5 records where launch sites begin with the string 'CCA’
- Demonstration of the total payload mass carried by NASA-launched boosters (CRS)
- Display of the average payload mass carried by the booster version F9 v1.1Listing of the date of the first successful landing on the ground runway.
- Listing of the names of boosters with a payload mass of more than 4000 but less than 6000 that have been successful on board the drone ship
- Listing the total number of successful and unsuccessful mission results
- List the names of the booster versions carrying the maximum payload mass.
- List records for months in 2015 showing month names, failed landing_results on drone ship, booster versions, launch_location.
- List the number of landing results (such as Failed (drone ship) or Successful (ground pad)) between 2010-06-04 and 2017-03-20 in descending order.

Build an Interactive Map with Folium

- The factors analysed using Folium are:
- Marking of all launch sites on a map
- Marking of successful/failed launches for each site on the map
- Calculation of distances between a launch site and its vicinity
- Drawing a line between a launch site and a nearby coastline point
- Drawing a line between a launch site and the nearest city, railway, motorway

Build a Dashboard with Plotly Dash

- Plotly is known for its ability to create interactive graphics. These generated graphs can be easily viewed, shared and even edited through web browsers.We drew a pie chart showing total lounges by specific sitesWe have drawn a scatter plot showing the relationship between the resultant and payload mass for different booster versions.

Predictive Analysis (Classification)

- We added data
- We created a prediction column
- Split the data into training and test sets
- We trained the logistics with Regression Model, SVM, KNN, Decission Tree model
- We compared the models.
- We found the best model

Results

- CCAFS SLC 40, KSC LC 39A have high success rates for high payload, ES-L1, GEO, HEO, SSO orbits have 100% success rate and success rates increase as more flights are launched from LEO, PO, SSO, VLEO, ISS orbits.
- The Decision Tree Classifier is best used to model SpaceX booster launch data as it has the highest accuracy.

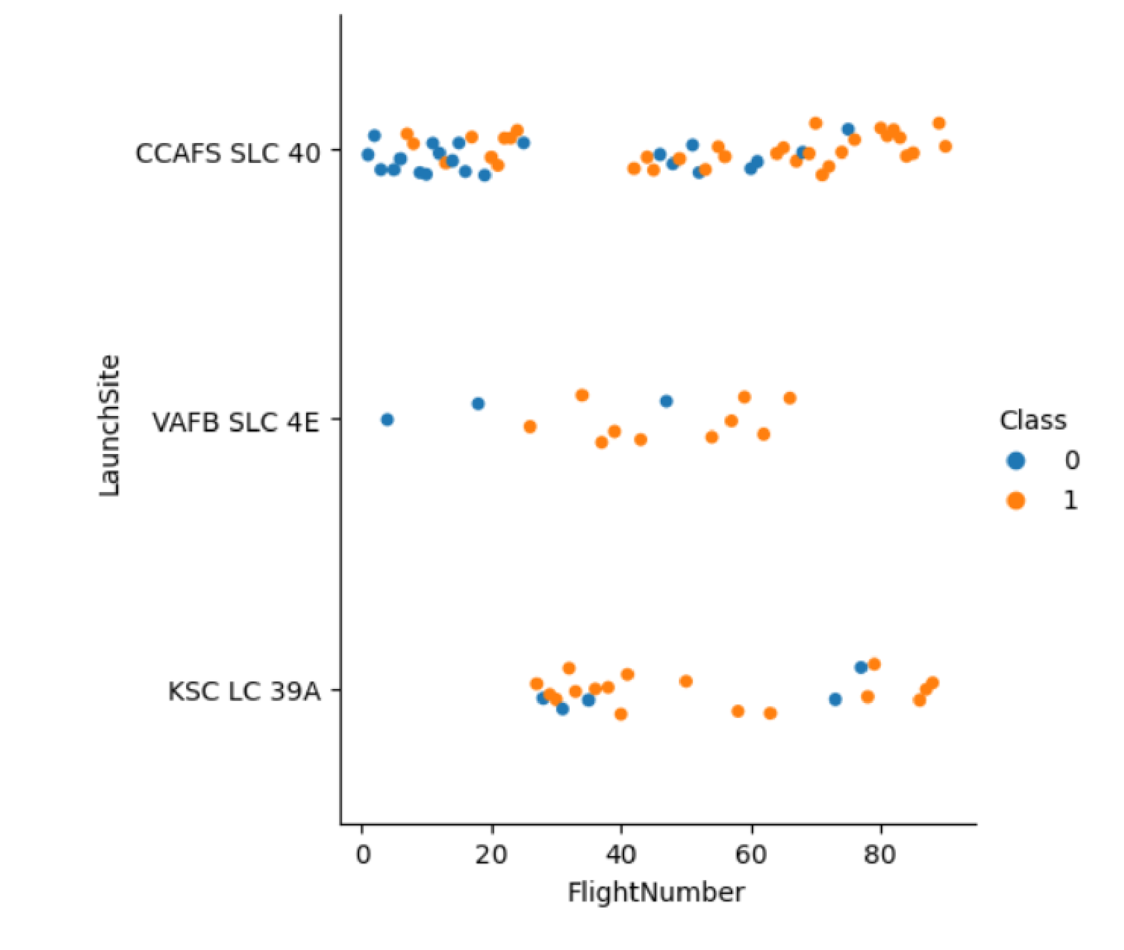
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

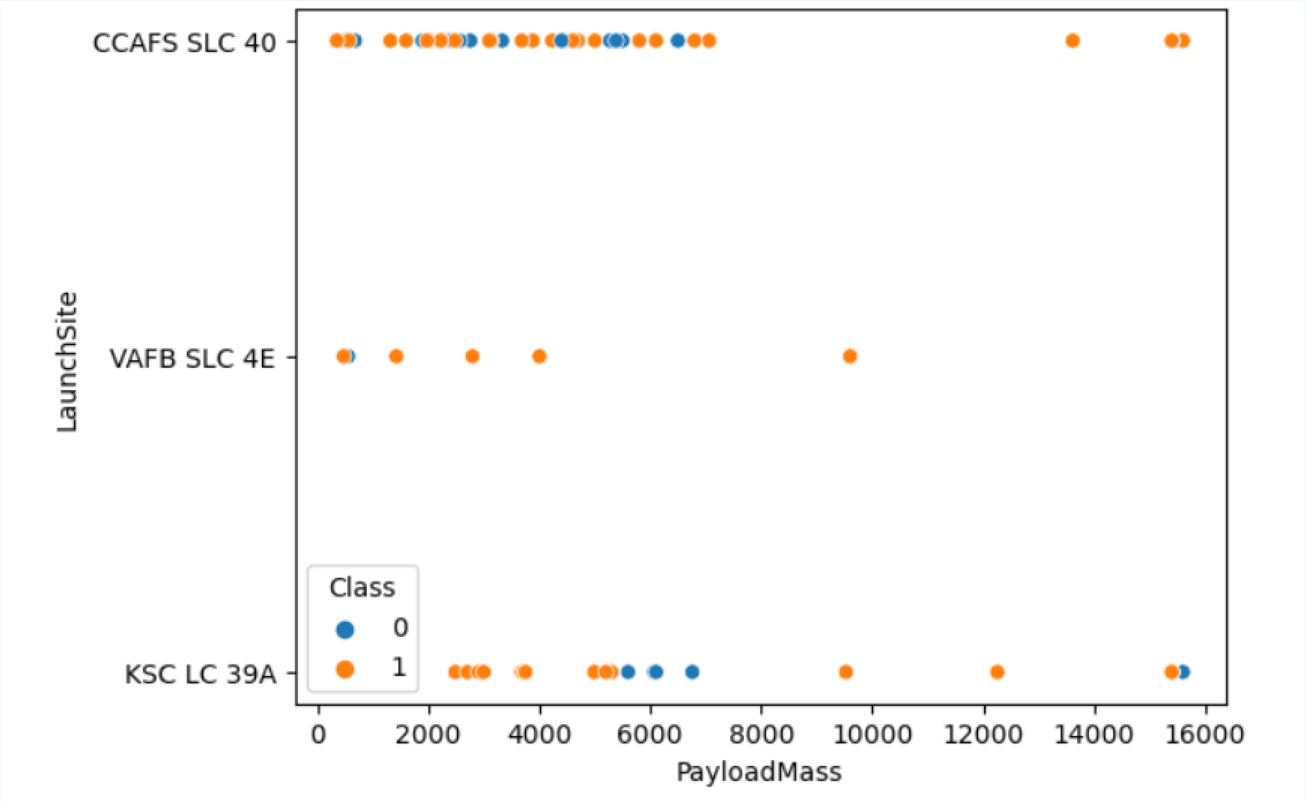
Here the number of flights launched from different launch locations is shown.

Descending Order: KSC LC 39, CCAFS SLC 40, VAFB SLC 4E



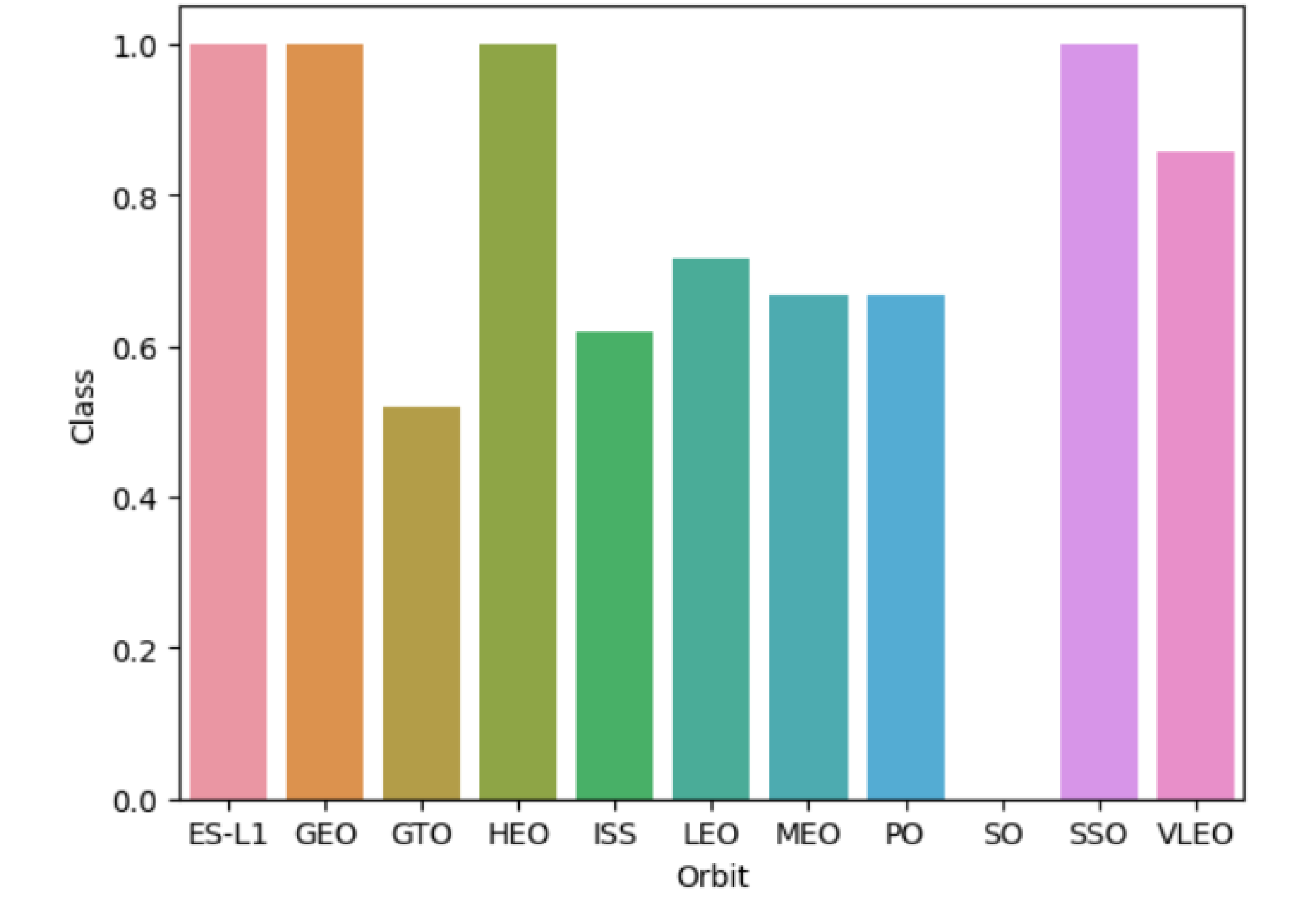
Payload vs. Launch Site

- CCAFS SLC 40 is successful at high payload mass. but mixed at low.
- VAFB SLC 4E is succesful at low payload.
- The KSC LC 39A site has a high success rate at low payloads and high payloads, with a failure rate of around 6000kg.



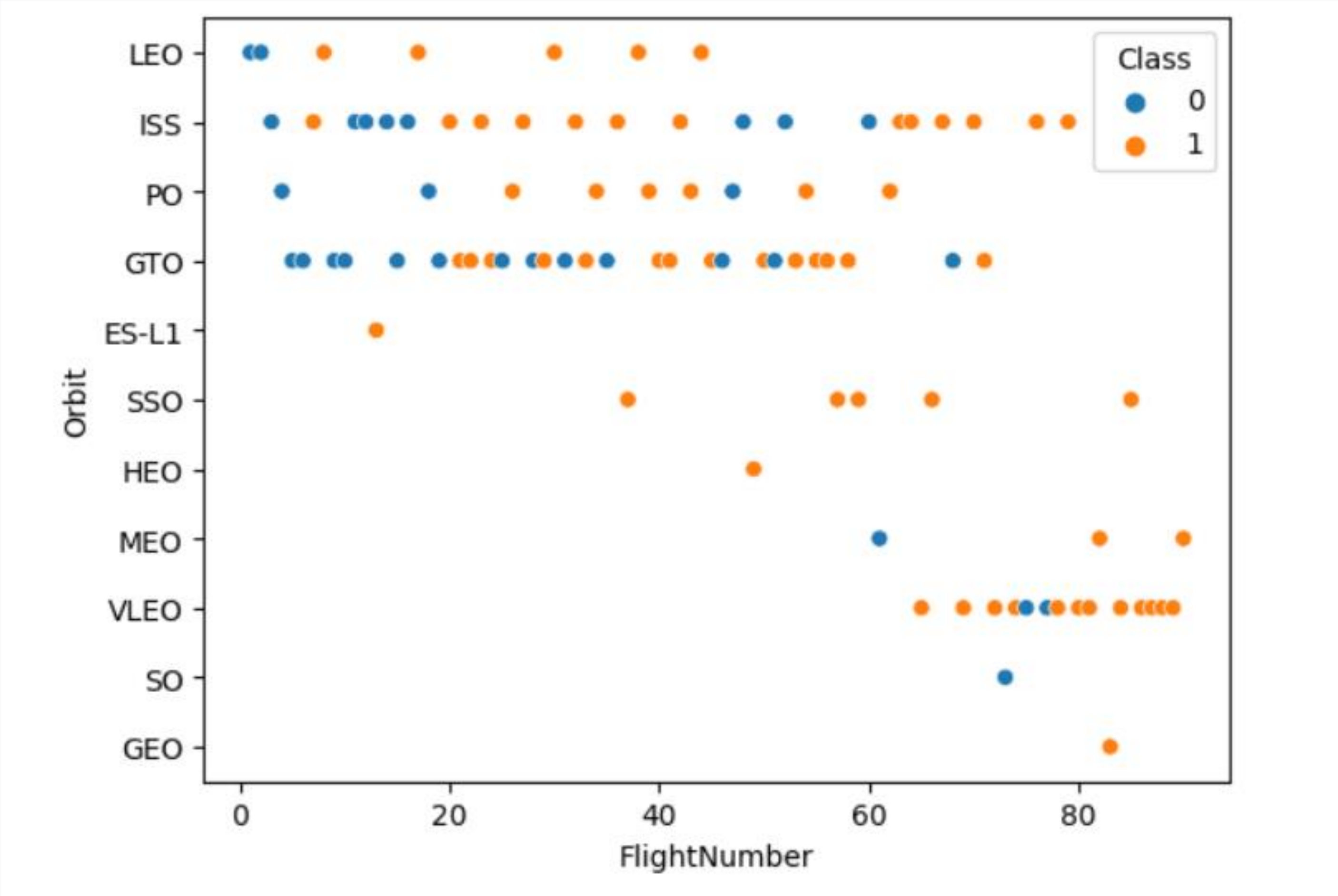
Success Rate vs. Orbit Type

- ES-L1, SSO, HEO, GEO recorded 100% success rate
- SO show 0% success rate



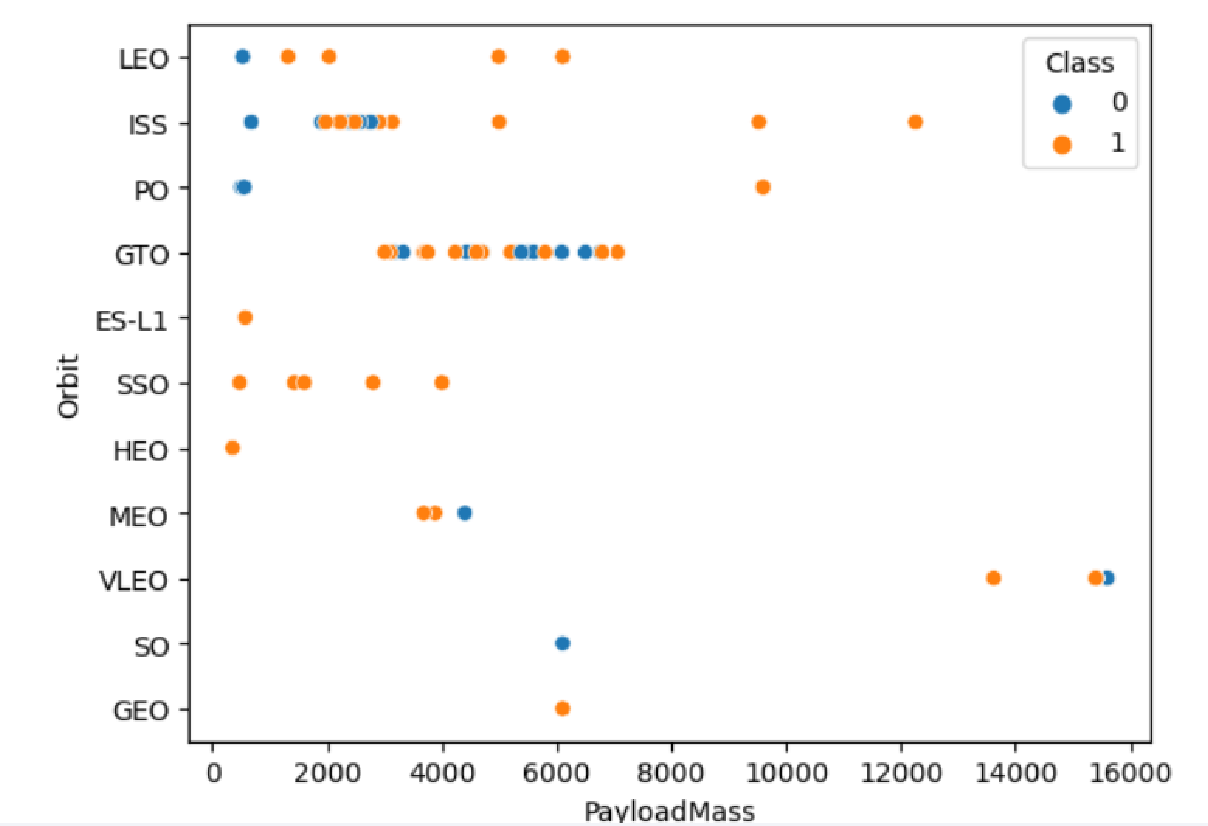
Flight Number vs. Orbit Type

- GTO has successful and unsuccessful launches.
- There were no unsuccessful launches in ES-L1 SSO HEO and GEO.
- LEO and VLEO are more secure.



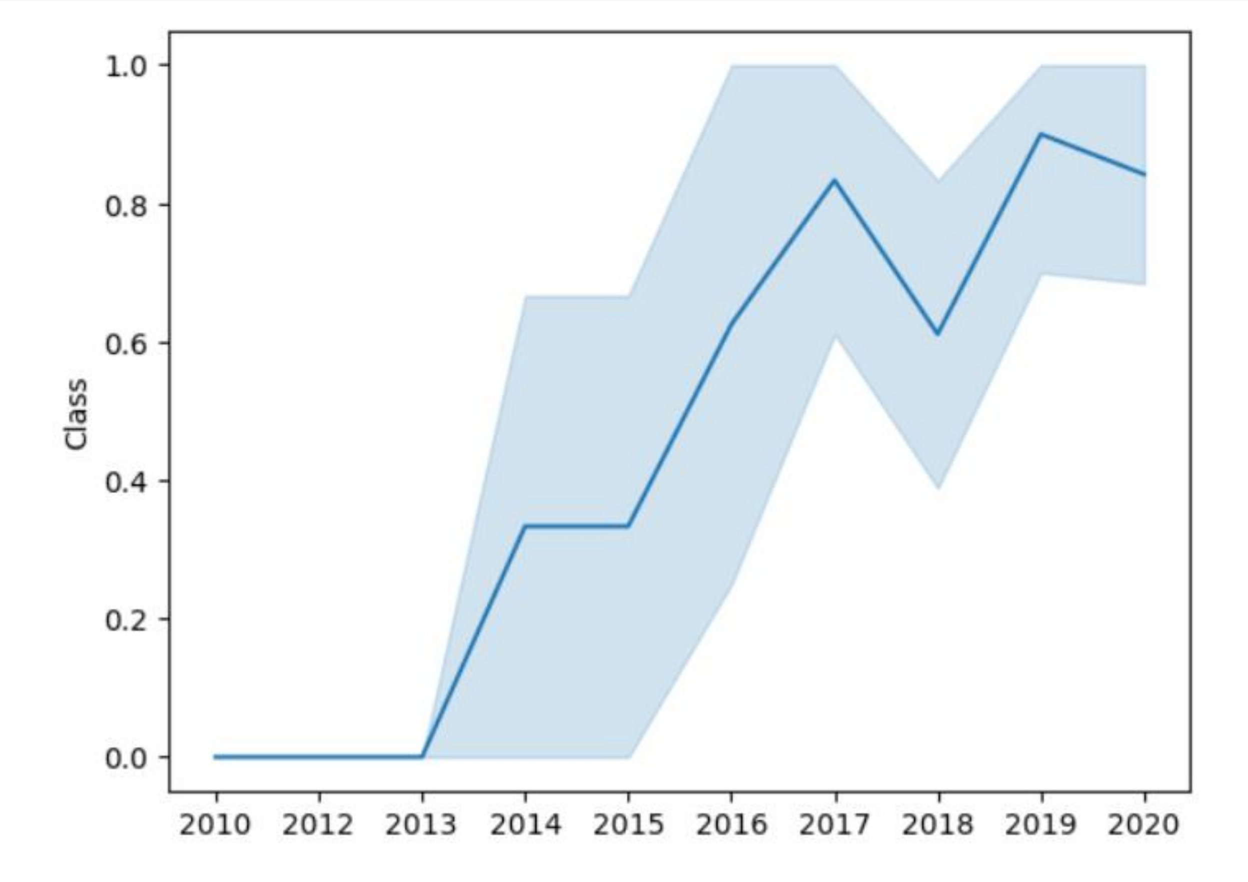
Payload vs. Orbit Type

- Higher loads witnessed almost complete success rate
- ES L1, SSO and HEO have 100% success rates
- GTO has a higher failure rate than others



Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020



All Launch Site Names

THESE ARE THE LAUNCH SITES:

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

Launch Site Names Begin with 'CCA'

WE SEE THE FİRST 5 LİNES STARTİNG WİTH THE LAUNCH SİTE CCA

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	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

- Total payload carried by thrusters from NASA:

  45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

  2928.4

First Successful Ground Landing Date

- The first successful landing on the land runway was achieved on:

  2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of thrusters that have been successful on the drone ship and have a payload mass greater than 4000 but less than 6000:

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes:

  98

Boosters Carried Maximum Payload

Names of the booster versions which have carried the maximum payload mass:

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

2015 Launch Records

- Records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015:

substr(Date, 6,2)	Launch_Site	Booster_Version
10	CCAFS LC-40	F9 v1.1 B1012
04	CCAFS LC-40	F9 v1.1 B1015

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

Number of landing results (such as Failed (drone ship) or Successful (ground pad)) between 2010-06-04 and 2017-03-20:

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

Section 3

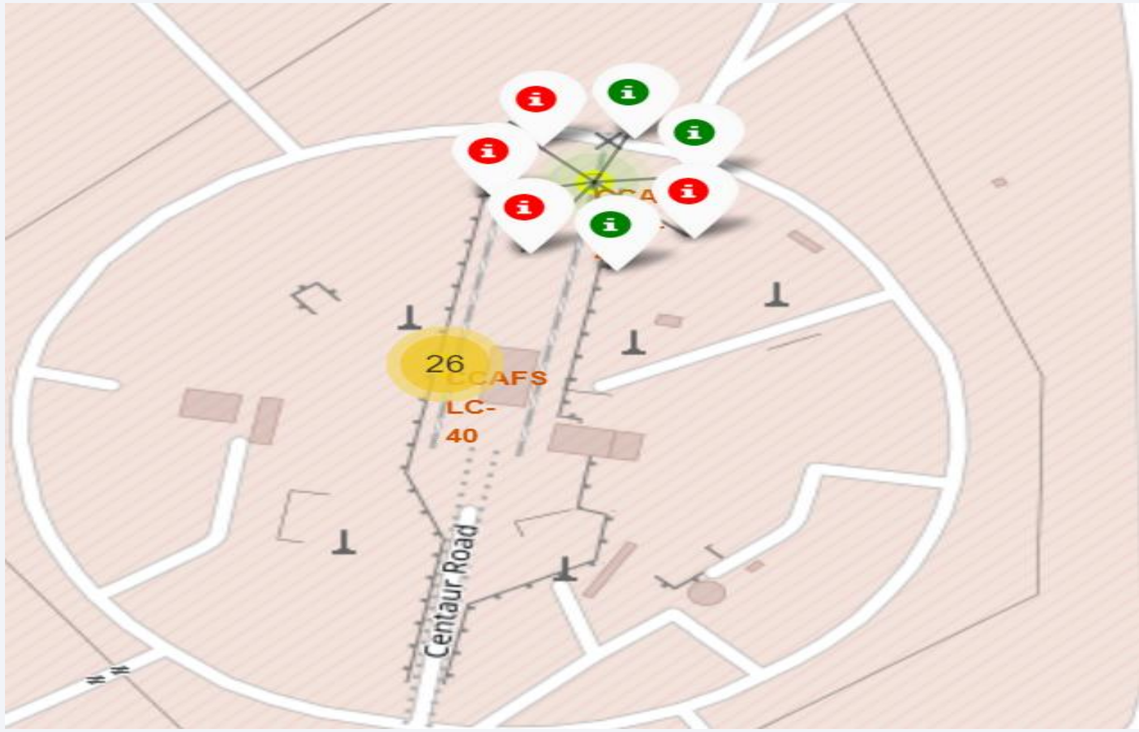
Launch Sites Proximities Analysis

Launch Sites

The launches were made there

- All the launch site are close to the coastline

Launch Records



Green cluster indicates successful take-off. Red cluster indicates unsuccessful take-off.

The Distances Between A Launch Site To Its Proximities

- Distance 0.96 KM
- Close to roads and railways

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

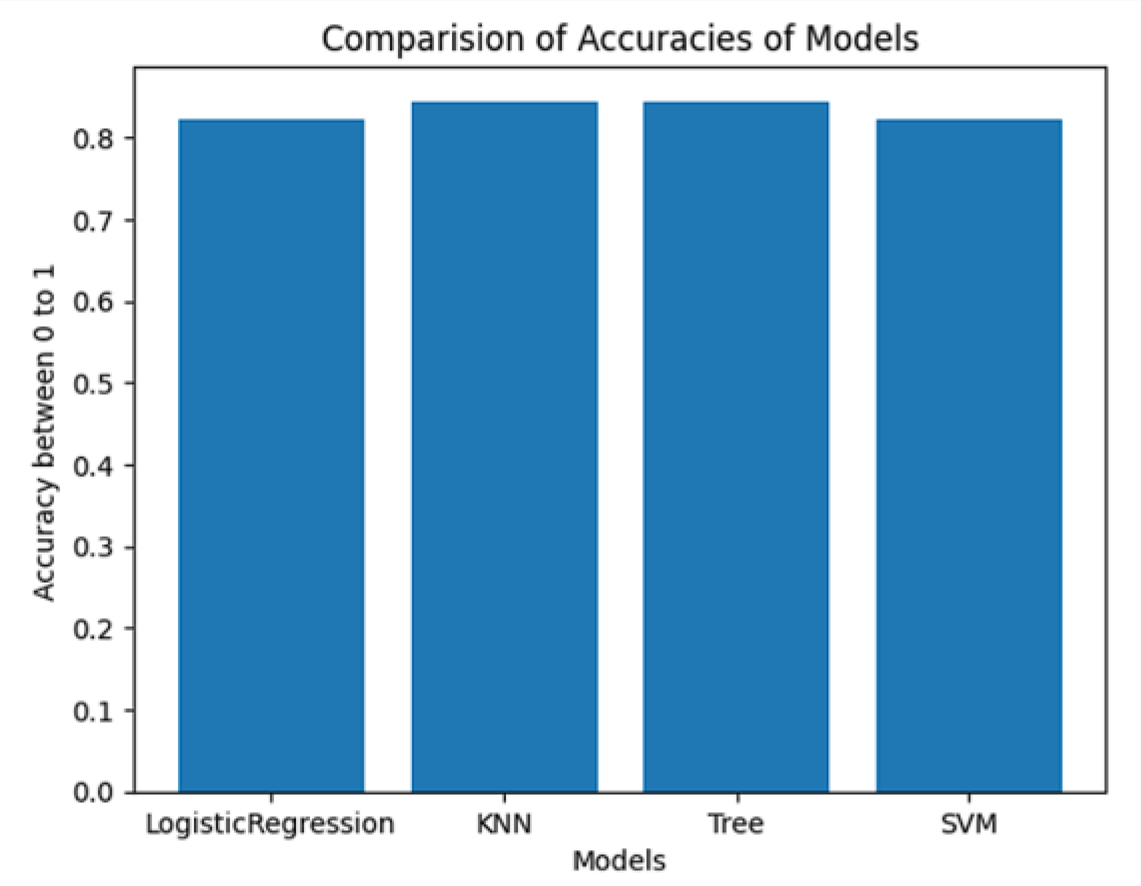
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

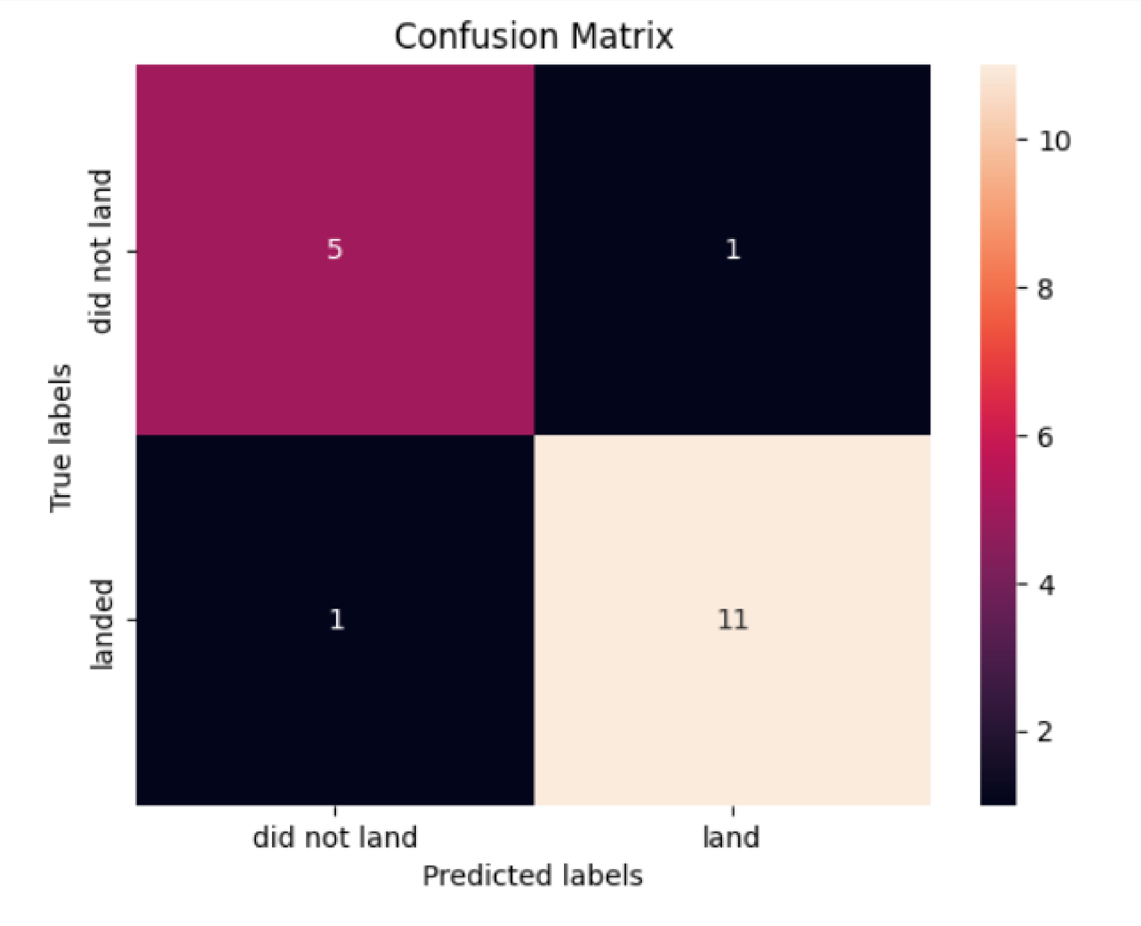
Classification Accuracy

- Decision Tree is best performing Model. accuracy %87.
- KNN is almost as accurate.



Confusion Matrix

- The model was wrong once on the booster that didn't land. The other five times it guessed right.
- The model was wrong once on the landing booster. The other eleven times it guessed right.



Conculusion

- Point 1 : The Tree Decision Classifier and KNN classifier both perform best under the given training and test datasets.The Tree Decision Classifier gives the highest accuracy of 87%
- Launch success rate since 2013 the success rate continued to increase until 2020.
- KSC-LC 39A is the best launch site with 76 per cent.
- Higher loads witnessed almost complete success rate.
- ES-L1, SSO, HEO, GEO are reliable orbits, but SO is not.

Appendix

- SpSpacex.csv
- spacex2_dash_app.py
- SpaceX REST Api
- Jupyter Notebooks:

  jupyter-labs-spacex-data-collection-api.ipynb

  jupyter-labs-webscraping.ipynb

  Labs-jupyter-spacex-Data wrangling.ipynb

  jupyter-labs-eda-dataviz.ipynb

  lab_jupyter_launch_site_location.ipynb

  SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Thank you!