DATA SCIENCE CAPSTONE PROJECT

SPACE X

Anshu Singh

Skills Network

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - ➢Visualization – Charts
  - ➢Dashboard
- Discussion
  - ➢Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an Interactive map with Folium
- Building a dashboard with Plotly
- Predictive analysis using classification
- Results
- Exploratory data analysis results
- Interactive data visualization screen shots
- Results of the predictive analysis with classification module

INTRODUCTION

Project Context

The objective of the project is to predict whether the first stage of Falcon 9 will land successfully or not. The motive behind the prediction is that the Spacex advertises that cost of launching Falcon 9 rocket is 62 million dollars whereas the costing incurred by other service providers is close to 165 million dollars.

The reason that Spacex is able to launch at a considerably lower cost is due to reuse of its first stage. So, if we are able to predict if the first stage of Falcon 9 will be successful , then other providers can also benefit from this and bid for rocket launch at a competitive pricing.

Challenges :

- Identifying the factors that affect the successful launching of Falcon 9.
- Determining the impact of certain variables on success of the launch.

METHODOLOGY

- Data collection methodology
- Spacex Rest API
- Web scrapping from Wikipedia using BeautifulSoup
- Data Wrangling using an API.
- Performing exploratory data analysis (EDA)using visualization techniques
- Performing exploratory data analysis using SQL
- Plotting Scatter Graphs and Bar Graph to show the relatioships between variables impacting the success or failure of launch result
- Performing interactive visual analytics using Folium and Plotly Dash.
- Identifying the best suited machine learning model to carry out predictive analysis to predict the outcome of launch result.

DATA COLLECTION –SPACEX API

- **Request and parse the Spacex launch data using the GET request**
- static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
- **Decoding the response as a json file and converting it into pandas dataframe.**
- data=pd.json_normalize(response.json())
- **Extracting the required data and cleaning.**
- **Constructing and combining the dataset columns into a dictionary.**
- launch_dict = {'FlightNumber': list(data['flight_number']),
- 'Date': list(data['date']),
- 'BoosterVersion':BoosterVersion,
- 'PayloadMass':PayloadMass,
- 'Orbit':Orbit,
- 'LaunchSite':LaunchSite,
- 'Outcome':Outcome,
- 'Flights':Flights,
- 'GridFins':GridFins,
- 'Reused':Reused,
- 'Legs':Legs,
- 'LandingPad':LandingPad,
- 'Block':Block,
- 'ReusedCount':ReusedCount,
- 'Serial':Serial,
- 'Longitude': Longitude,
- 'Latitude': Latitude}

Github link:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb:

- **Data wrangling to deal with the missing values**
- data_falcon9.isnull().sum()
- mean_value=data_falcon9['PayloadMass'].mean()
- data_falcon9['PayloadMass'].fillna(value=mean_value, inplace=True)
- data_falcon9.isnull().sum()

DATA COLLECTION –WEB SCRAPPING

Data Collection through Web scrapping Falcon 9 records from Wikipedia using BeautifulSoup

- **Request the Falcon9 Launch Wiki page from its URL**

```
data=requests.get(static_url).text
```

Create a BeautifulSoup object from HTML response

```
soup=BeautifulSoup(data)
```

Extract all column/variable names from the HTML table header

```
html_tables=soup.find_all('table')
```

Extract column names one by one

```
for row in first_launch_table.find_all('th'):
        name=extract_column_from_header(row)
if (name!=None and len(name)>0):
        column_names.append(name)
```

Create a data frame by parsing the launch HTML tables → **Create a dataframe after filling the parched launch records into launch_dict.**

Skills Network

Github url:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/jupyter-labs-webscraping.ipynb

DATA WRANGLING

We performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

We convert the launch outcomes into Training Labels with 1 meaning the booster landed successfully whereas 0 means the landing was unsuccessful.

The main tasks performed were as follows:

- Calculating the number of launches on each site.

```
df['LaunchSite'].value_counts()
```

- Calculating the number and occurrence of each orbit.

```
df['Orbit'].value_counts()
```

- Calculating the number and occurrence of mission outcomes of the orbit.

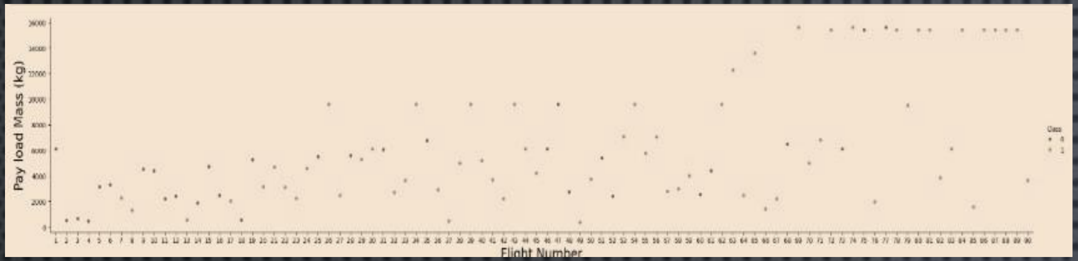
```
landing_outcomes=df['Outcome'].value_counts()
for i,outcome in enumerate(landing_outcomes.keys()):
print(i,outcome)
```

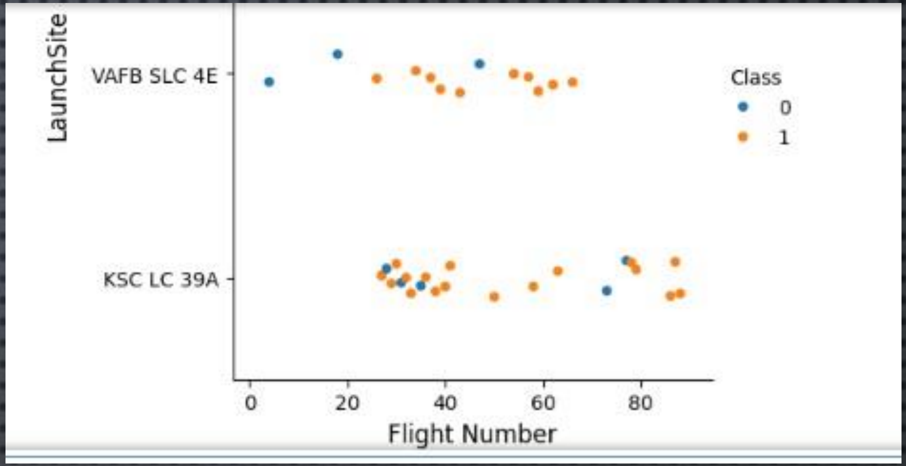
- Creating a landing outcome label from outcome column

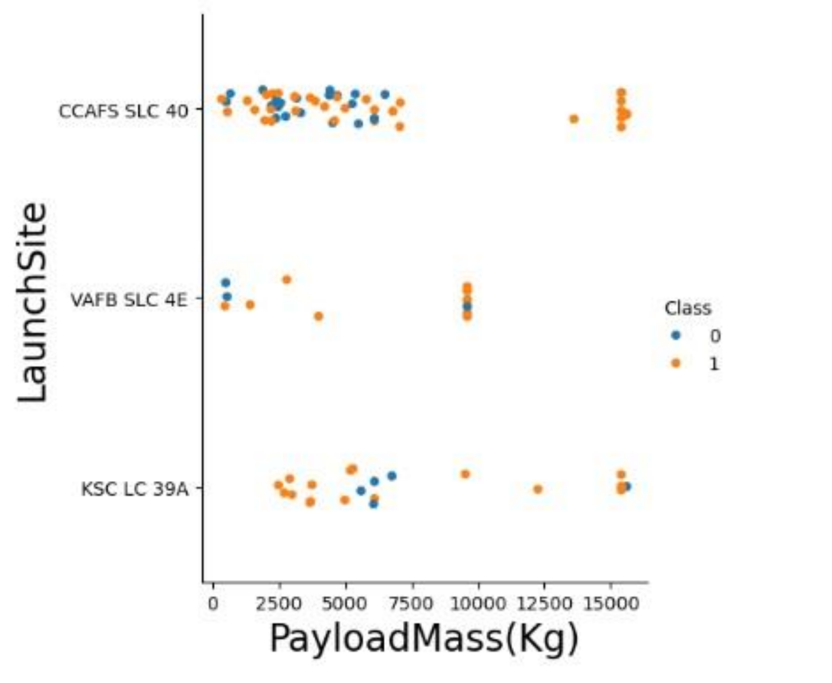
```
landing_class=[]
for outcome in df['Outcome']:
 if outcome in bad_outcomes:
 landing_class.append(0)
 else:
landing_class.append(1)
```

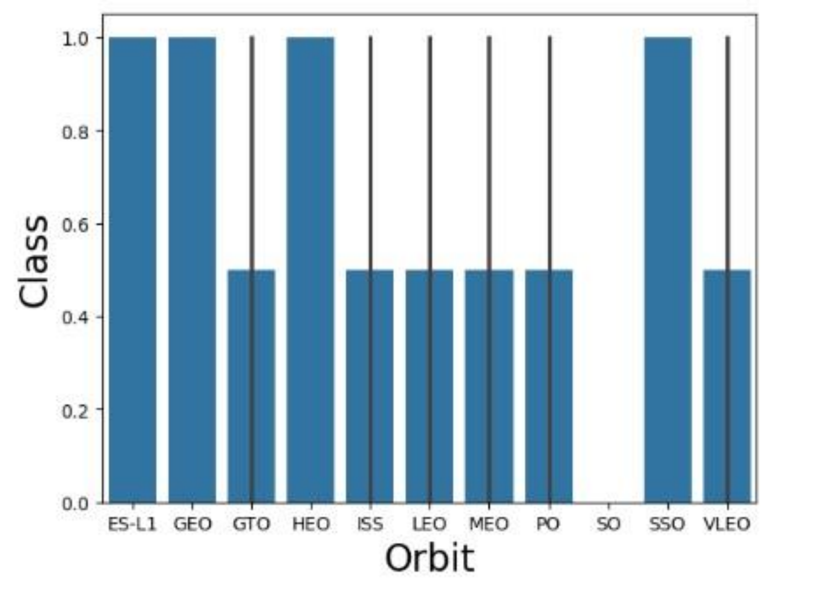
Guthub url:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA WITH DATA VISUALIZATION

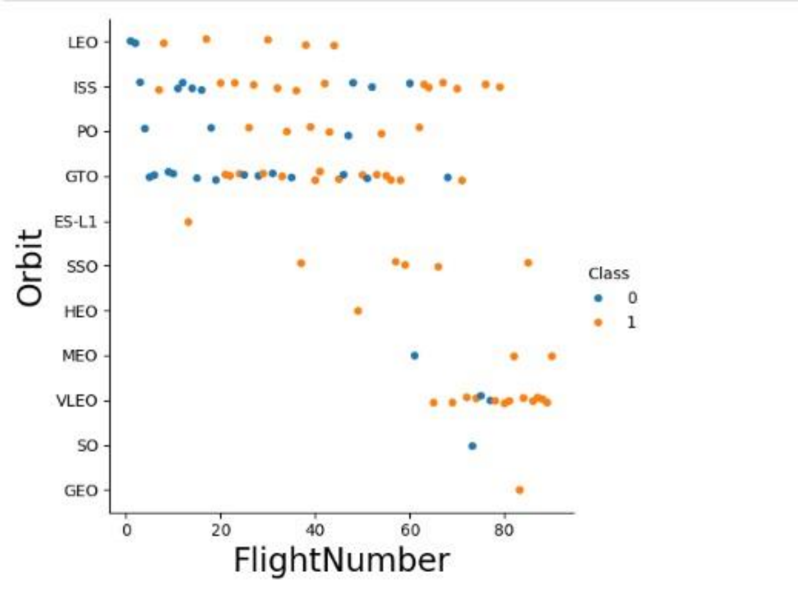


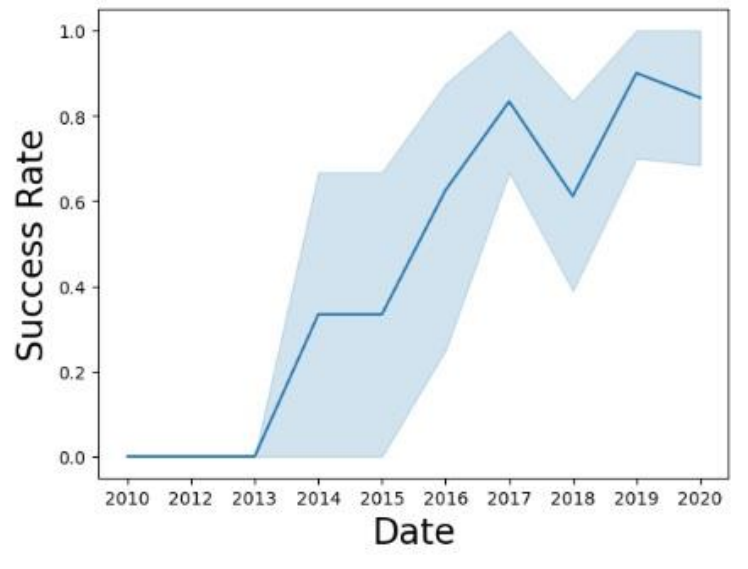


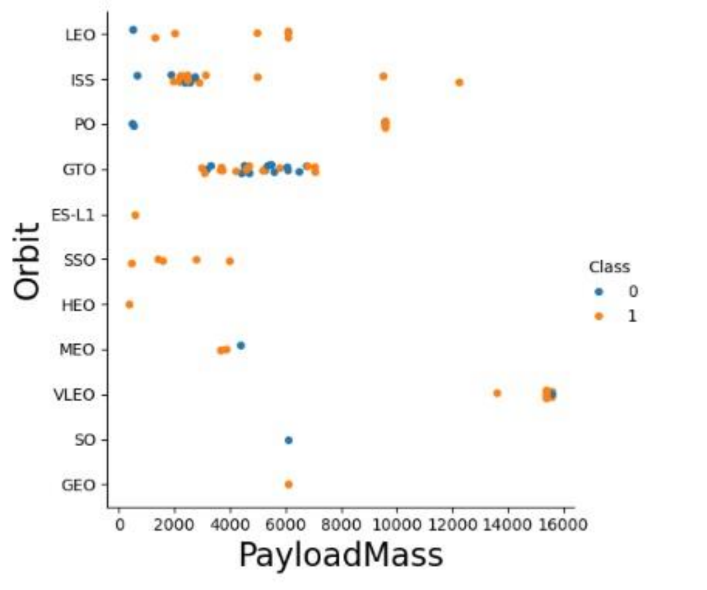




EDA WITH DATA VISUALIZATION-CONTD...







Github link:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/edadataviz.ipynb

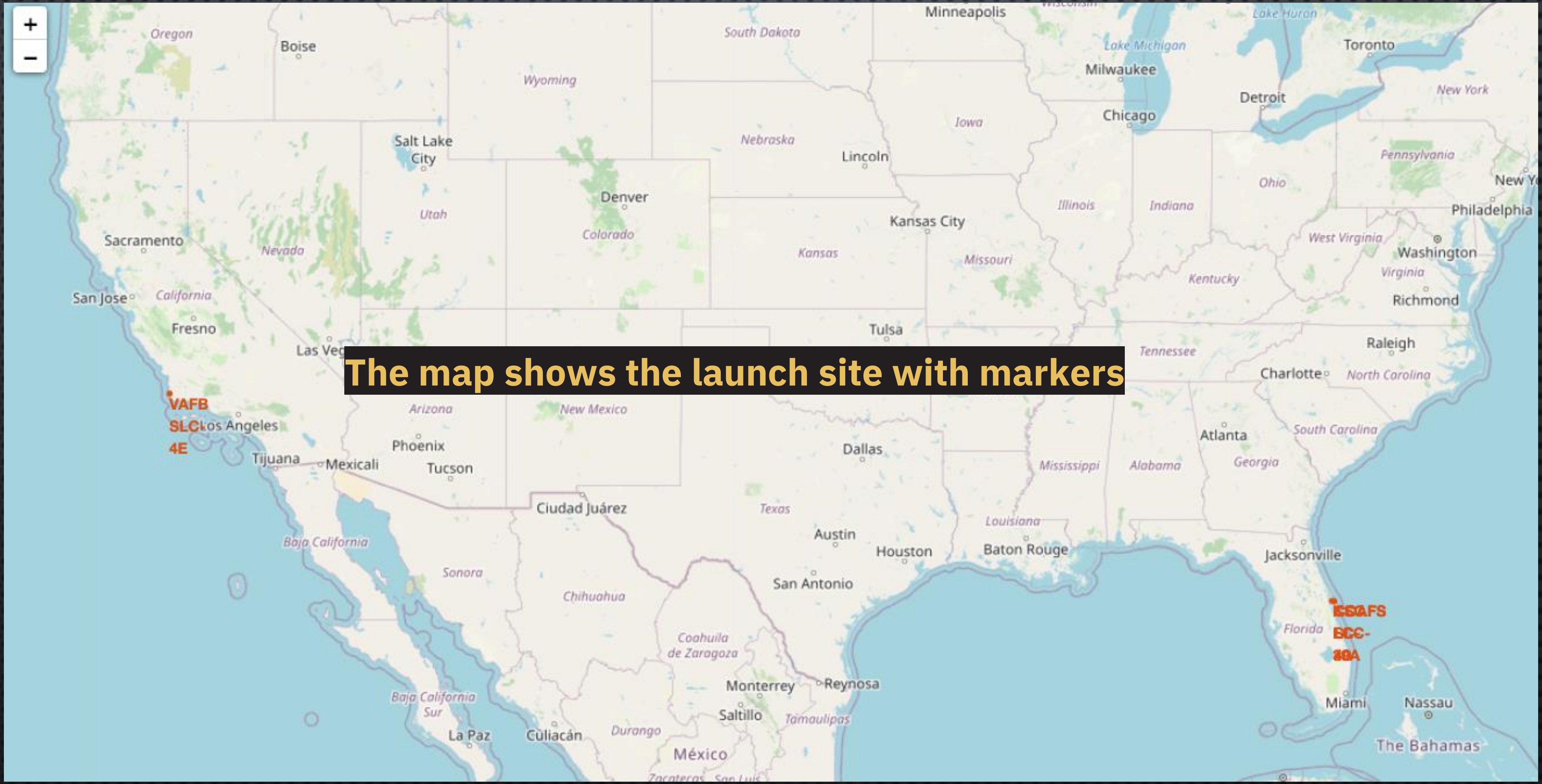
EDA WITH DATA VISUALIZATION- SQL

Performing SQL queries to understand the dataset.

- Display the names of the unique launch sites.
- Display 5 records where launch sites begin with “KSC”
- Display the total payload mass(kg) carried by boosters launched by NASA CRS
- Display the average payload mass carried by booster version F9v1.1
- List the date where successful landing outcome in drone ship was achieved
- List the name of the boosters which have successful landing in ground pad and have payload mass between 4000 and 6000 kg
- List the total number of successful and failure mission outcomes
- List the names of booster versions which have carried the maximum payload mass
- List the records that display the month name, successful landing outcomes in ground pad,booster versions, launch_site for the months in the year 2017
- Ranking the cpunt of successful landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

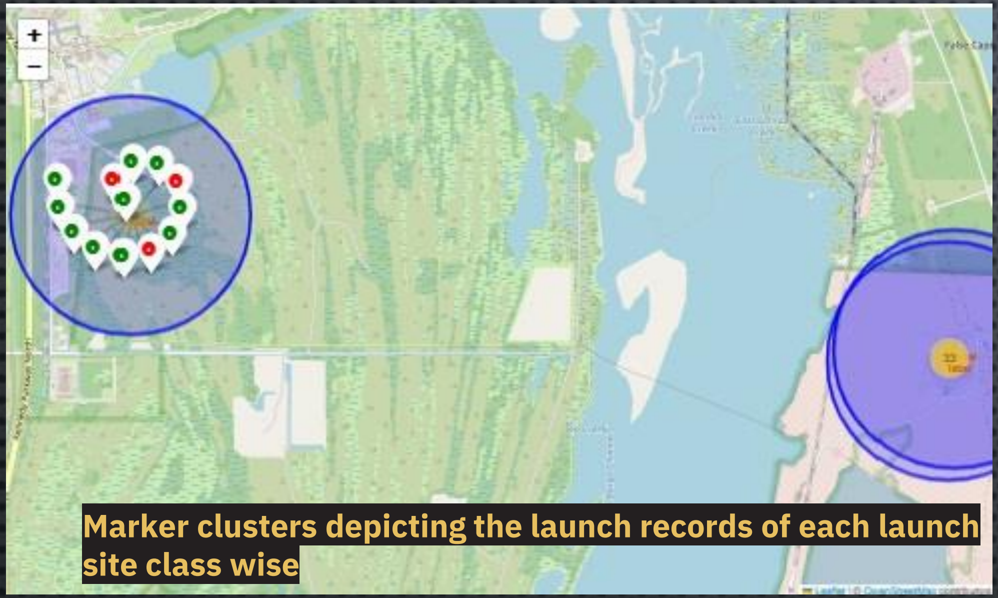
Github url :https://github.com/Anshu0903/Data-science-capstone-project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

INTERACTIVE MAP WITH FOLIUM



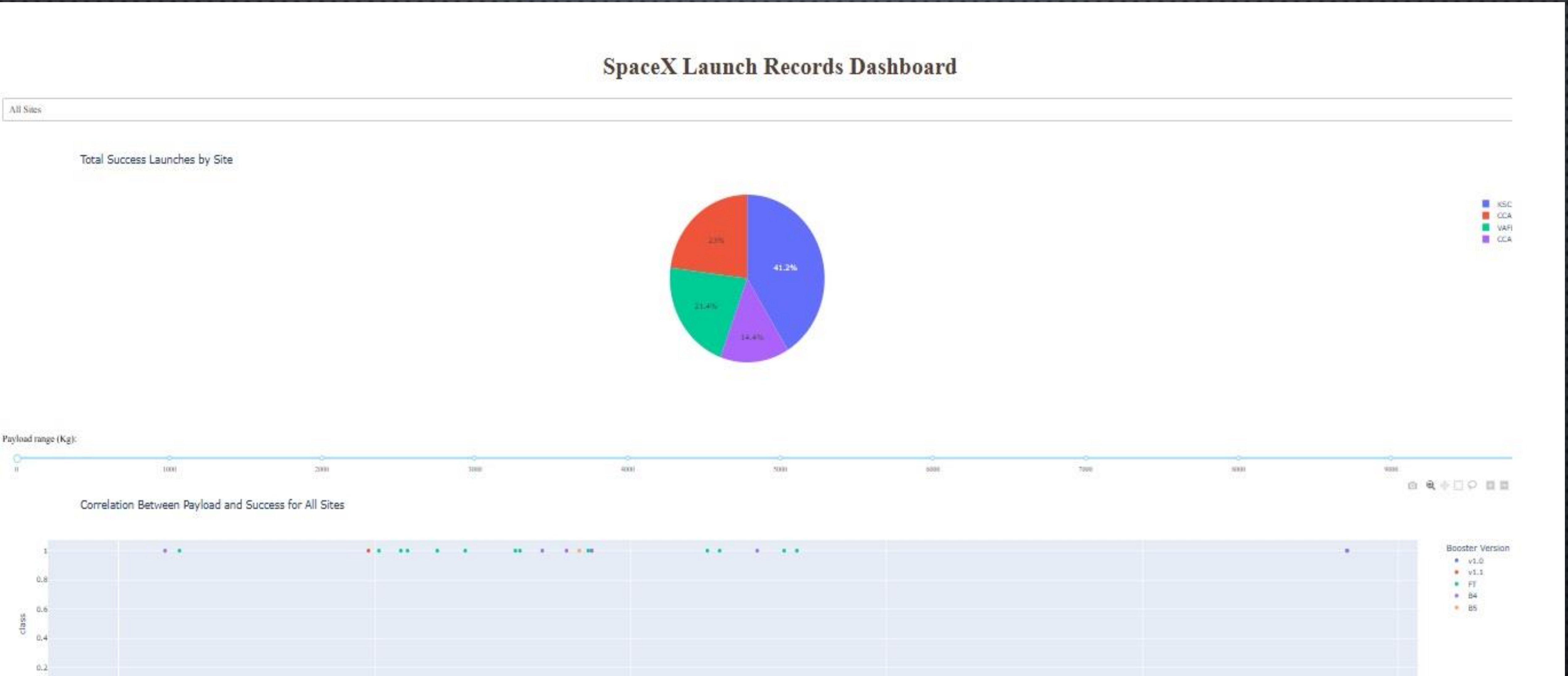
The map shows the launch site with markers

INTERACTIVE MAP WITH FOLIUM – CONTD..

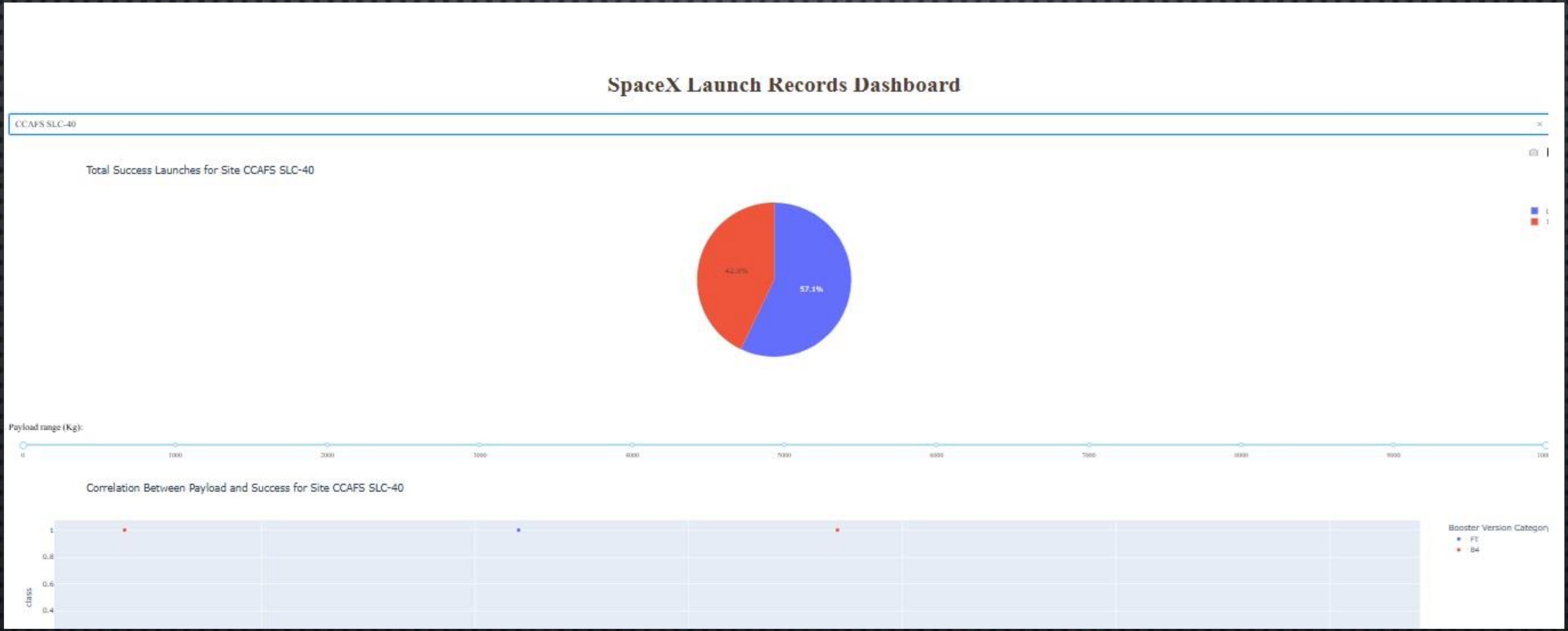


Github url:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/lab_jupyter_launch_site_location.ipynb

DASHBOARD APPLICATION - SPACEX LAUNCH RECORDS



DASHBOARD APPLICATION - SPACEX LAUNCH RECORDS-CONTD..



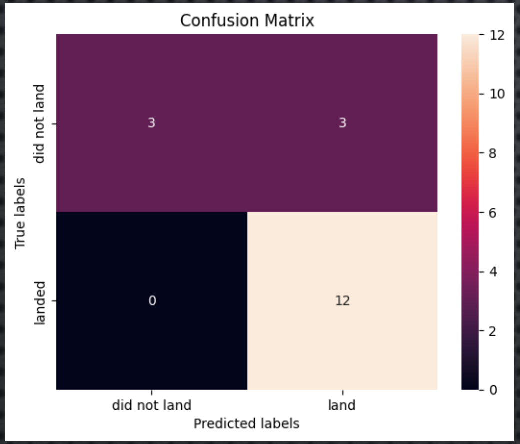
Dashboard depicting launch records of CCAFS SLC-40 as dropdown.

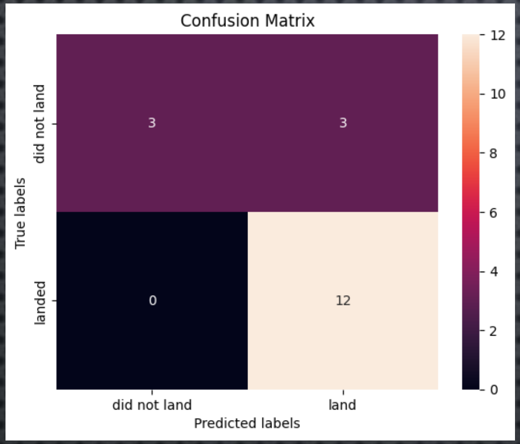
[Github url :https://github.com/Anshu0903/Data-science-capstone-project/blob/main/spacex_dash_app.py](https://github.com/Anshu0903/Data-science-capstone-project/blob/main/spacex_dash_app.py)

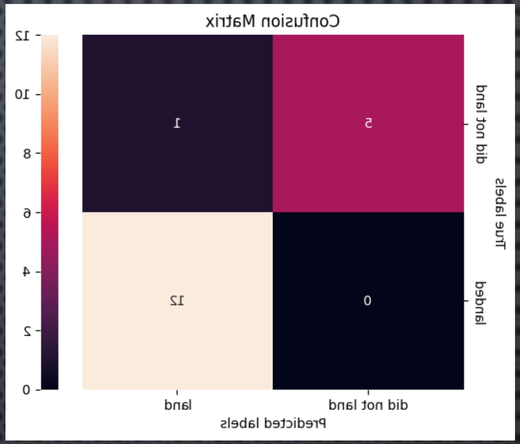
Skills Network

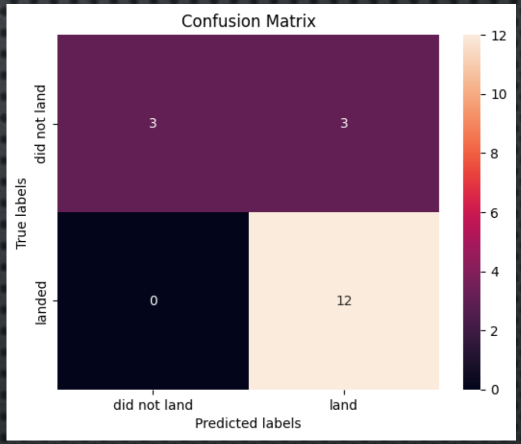
PREDICTIVE ANALYSIS-CLASSIFICATION

The predictive analysis was carried out using classification models Logistic Regression, SVM, Decision Tree, and KNN. The confusion matrix for all the models is shown below:









The summarized report of the accuracy of the models is mentioned below:

Method	Test Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.944444
KNN	0.833333

Github url:https://github.com/Anshu0903/Data-science-capstone-project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

RESULTS

- The Logistic Regression, SVM, and KNN all have an approximate prediction accuracy of 84% whereas Decision Tree has an accuracy score of 93%
- All these models perform predictions with reasonable accuracy
- Low weighted payloads perform better than heavier ones, but after a payload mass of 10000, count of successful launches from all sites relatively increases.
- KSC LC 39A has the most successful launches from all sites
- ES-L1, GEO, HEO, and SSO have higher success rate.
- All the launch sites are in close proximity to coast, railway, highway but at a distance from cities.

THANK YOU.