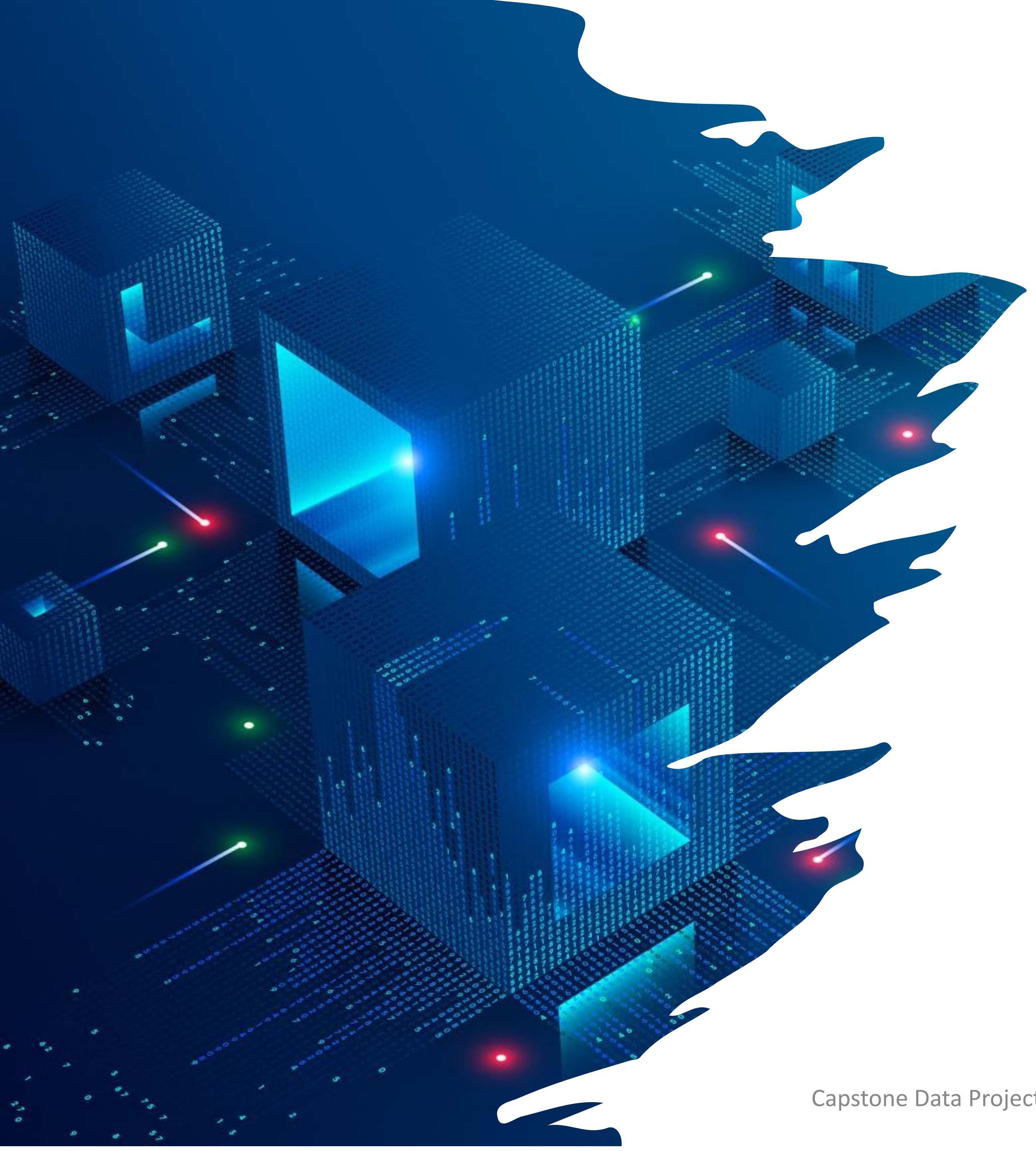
SpaceX Launch Success Prediction

IBM Data Science Capstone Project

Presented on: 30th June 2023

Presented by: Saif UR Rehman



Project Overview

- Data collected from Wikipedia and Data repository of SpaceX
- Explored data with SQL and Python
- Folium was used for visualization and presentation of data in shape of graphs and maps
- Various Machine Learning Models were used to prepare a data model for decision making purposes

Introduction to SpaceX Launch System

- As per data from the SpaceX a routine launch of Falcon9 rocket costs 62 million dollars, a competitor on the other hand costs the same endeavor more than 165 million dollars.
- SpaceX can reap the cost advantage based on its state-of-the-art rockets and launch and land system for its various stage rockets.
- Using data science and machine learning we can predict that what are the probabilities of a launch to be successful by successfully landing its various stages.
- This project will try to predict the same using data already available and feeding it to the MLM already developed

Project Objectives

- To find out the successful landing factors
- To determine how various launch parameters impact the rate of landing success
- To get the set of conditions necessary to ensure a successful landing

Methodology

Following Steps will be completed in this section using various tools already learnt in the process

- Data Collection
- Data Wrangling
- Exploratory Data Analysis
- Visualization
- Predictive Analysis

Data Collection

Data was collected using various methods and techniques to extract launch records, parse the table and convert into dataframe for future analysis these steps are as described in bullets below

- GET Request from SpaceX API
- .Jason() function was used to decode responses
- .Jason_normalize() to convert data into pandas dataframe
- Removed missing values and outliers using data cleaning
- Web Scraping with BeautifulSoup method was used to collect data from Wikipedia

Method1: Data Collection using API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
print(response.content)
```

Requesting data from SpaceX web link and printing the file contents to verify contents are loaded

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

Getting data from URL

```
response.status_code
```

check if the above GET request was successful using status_code, remember 200 shows a successful response

```
data = pd.json_normalize(requests.get(static_json_url).json())
data.head()
```

using .json() to decode the response and convert into a data frame for other mathematical operations possible

data = data[['rocket', 'payloads', 'launchpad', 'cores', 'flight_number', 'date_utc']]

From many data fields only keeping those necessary for our project

```
In [ ]: getBoosterVersion(data)
        BoosterVersion[0:5]
```

```
In [ ]: getLaunchSite(data)
        LaunchSite[0:5]
```

```
In [ ]: getPayloadData(data)
```

```
In [ ]: getCoreData(data)
```

Getting the data ready for usage

data=pd.DataFrame.from_dict(launch_dict,orient='index').transpose()

Create a pandas data from the data dictionary

Method2: Data Collection using Web Scraping

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

Requesting HTML Page and Assigning to an Object

```
soup = BeautifulSoup(page.text, 'html.parser')
```

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

Create BeautifulSoup and fill the table

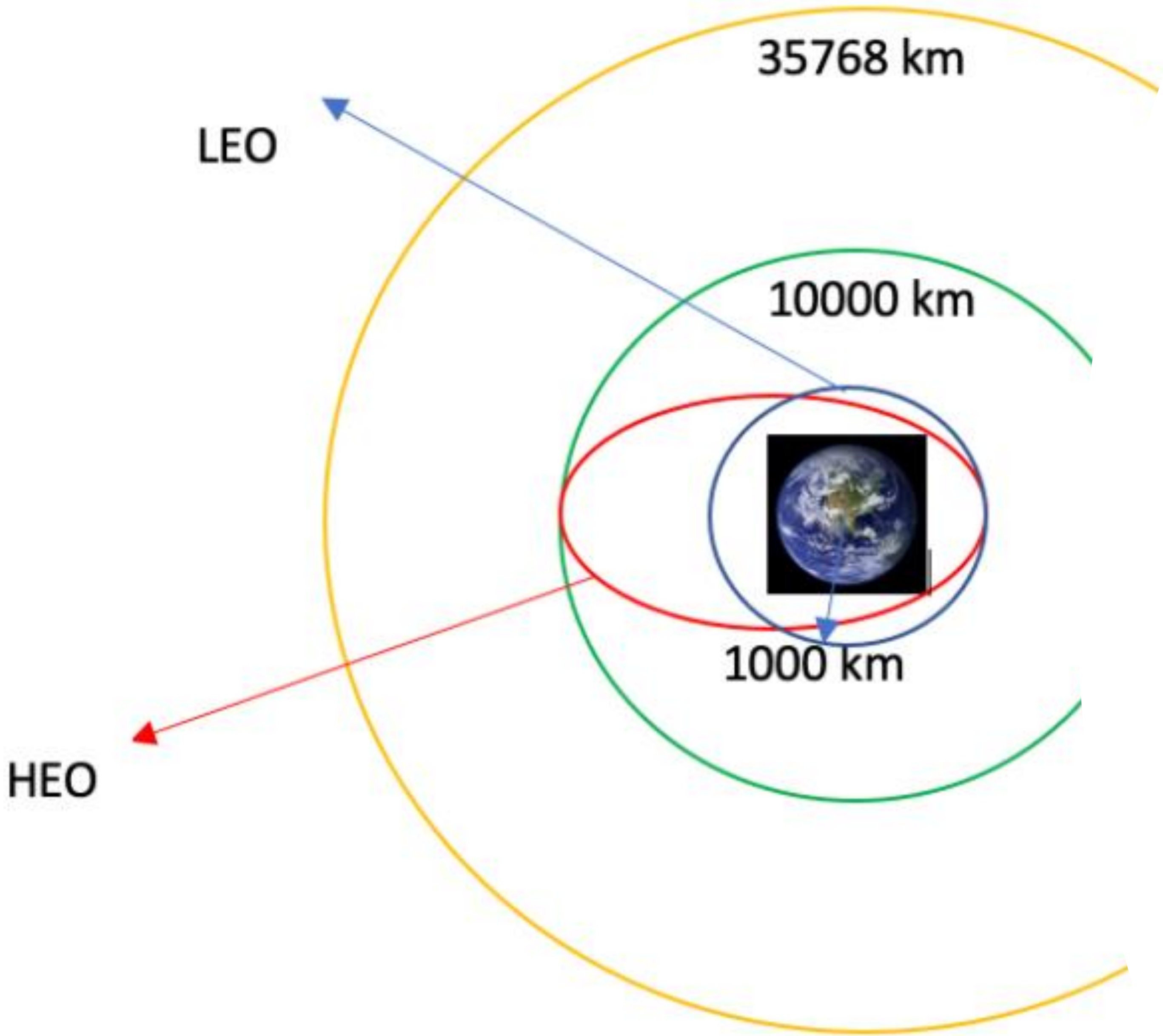
```
column_names = []

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

Extract Column Names

```
launch_dict= dict.fromkeys(column_names)
del launch_dict['Date and time ( )']
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'] = []
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

```
df=pd.DataFrame(launch_dict)
```

Convert to pandas data sets and dictionaries

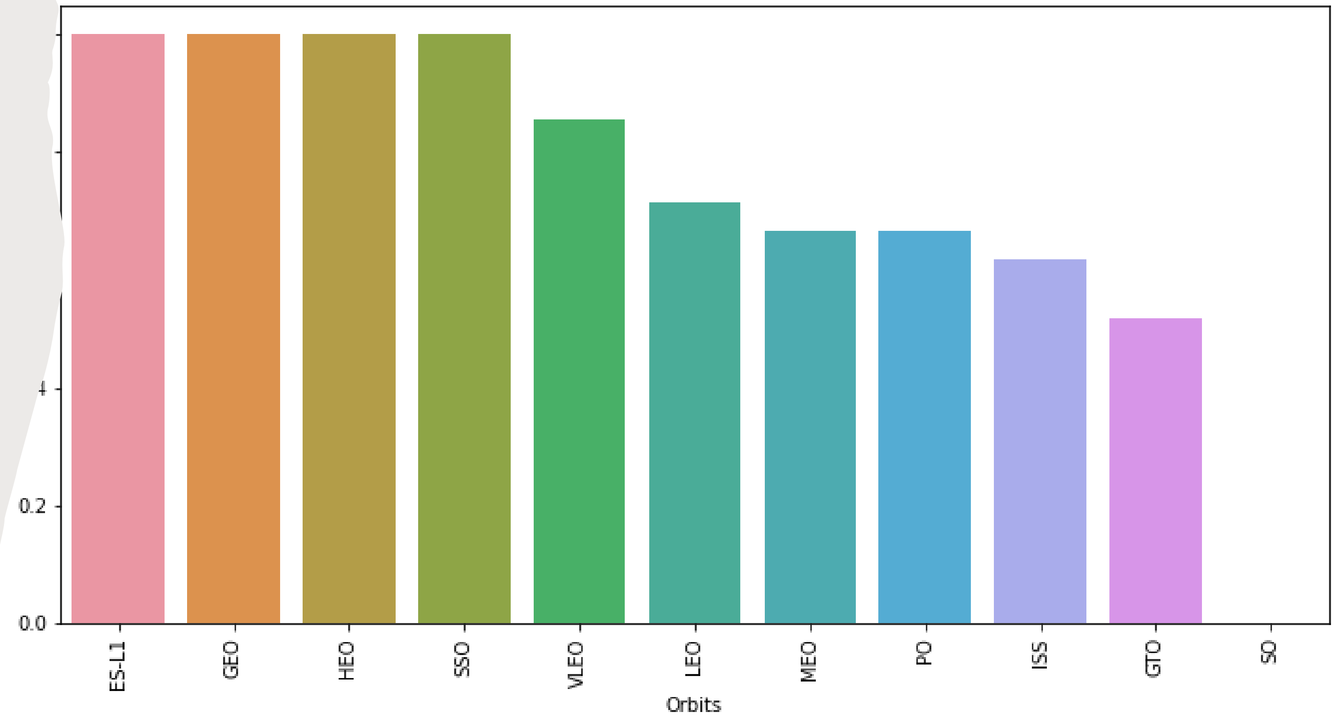


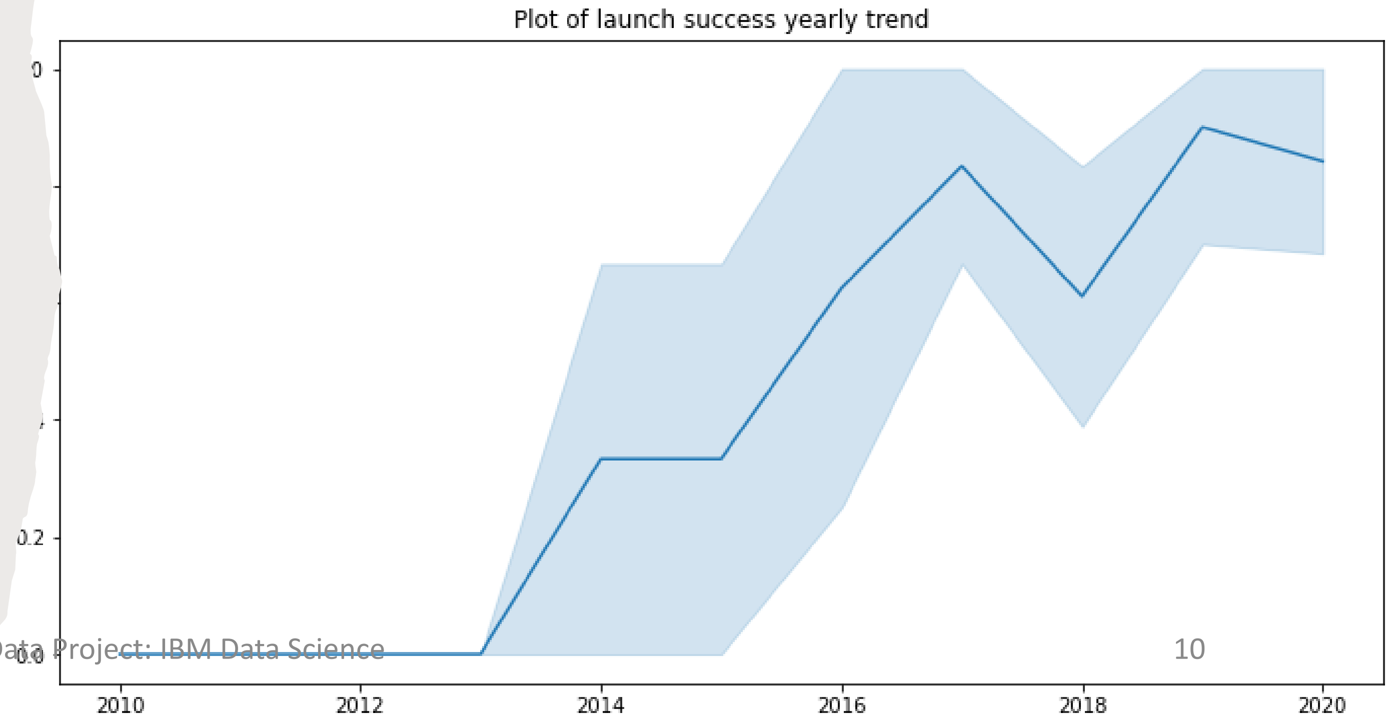
Method3: Data Wrangling

- Exploratory Data Analysis (EDA) helped in determining the training labels.
- Launch Statistics and Orbit Occurrences were calculated for each site
- Outcome of landing of each launch was exported to a csv file for further processing

Method4: Data Visualization

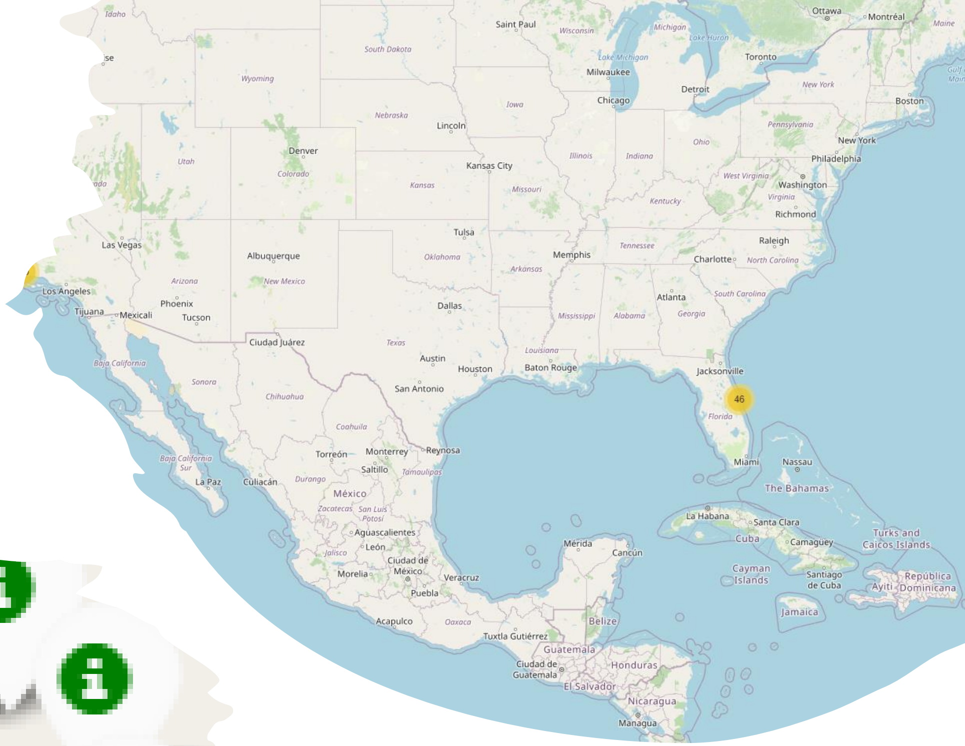
- As shown in the left side a line-graph and a bar chart was prepared using various tools
- The graphs show the relationship between payload, launch site, type of orbit and flight number of each Falcon9 rocket.
- A trend can be seen which represents a sharp rise in success rate from year 2013 to 2020.

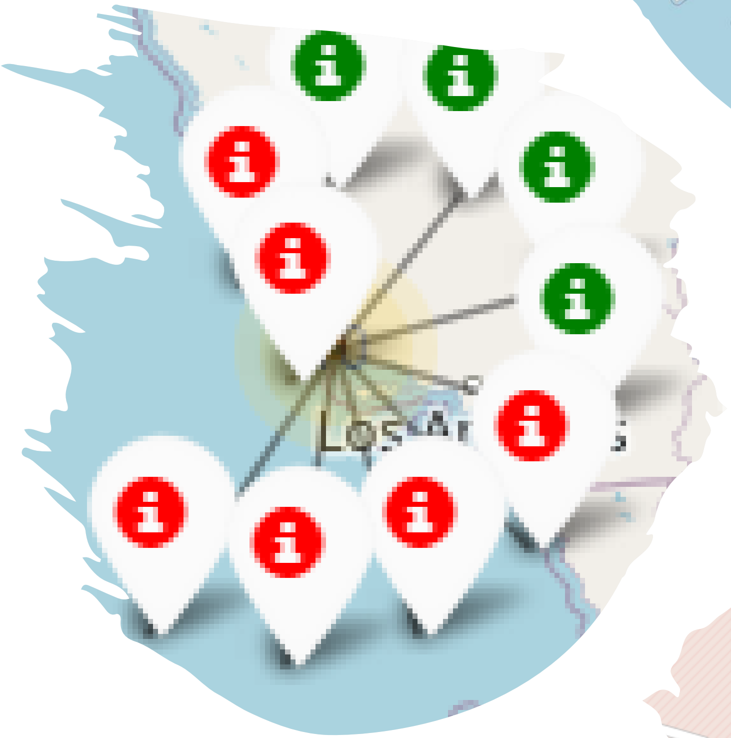


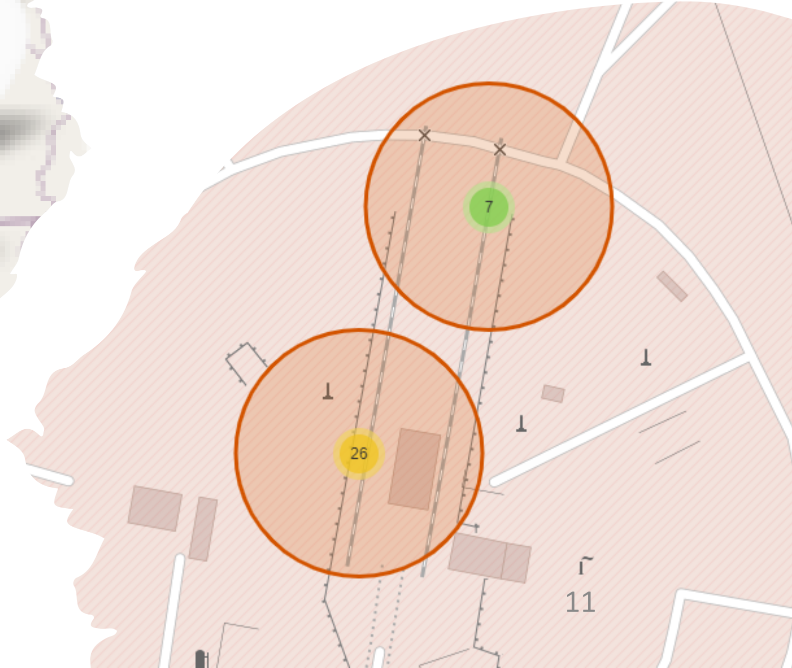


Method5: Using Folium for interactive mapping

- Folium was used to mark all launch sites as interactive objects on a map
- Markers like launch sites, distances, success/failure rates were added.
- Predictions of future launches were also plotted using 0 and 1 on the map, further distances from various proximities were marked
- Thus high probability sites for successful launch were highlighted on map

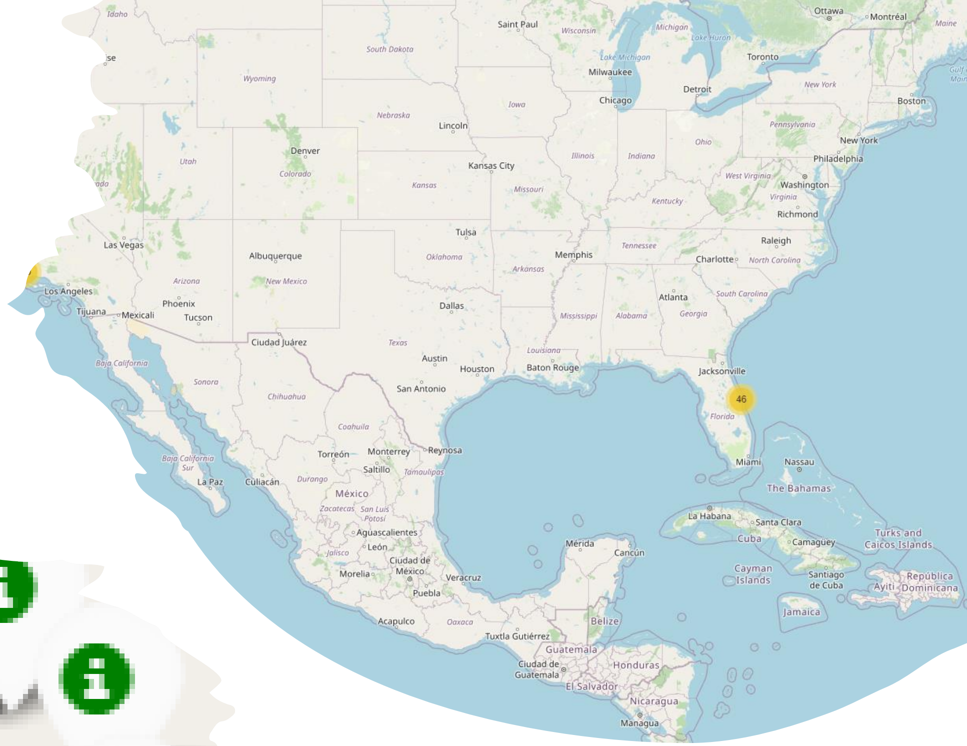


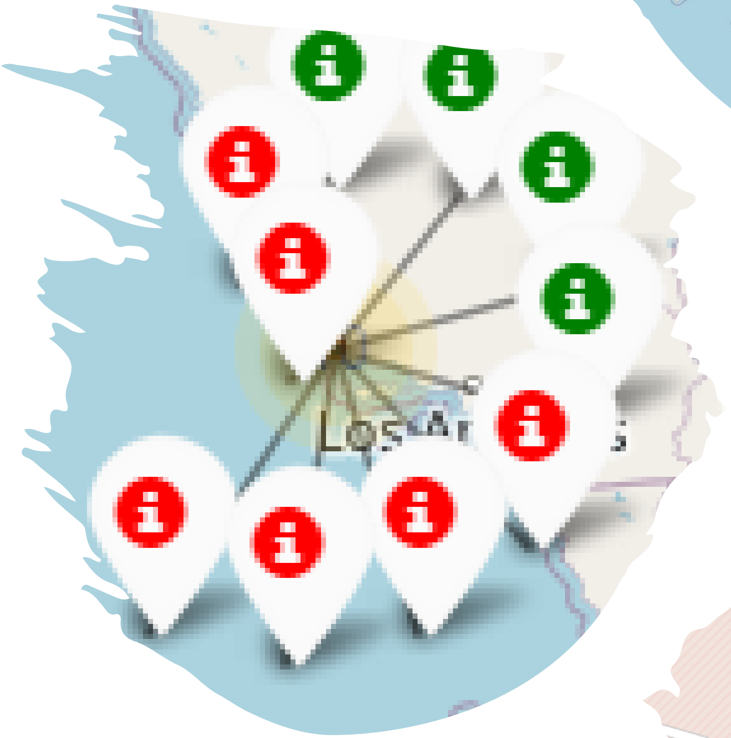
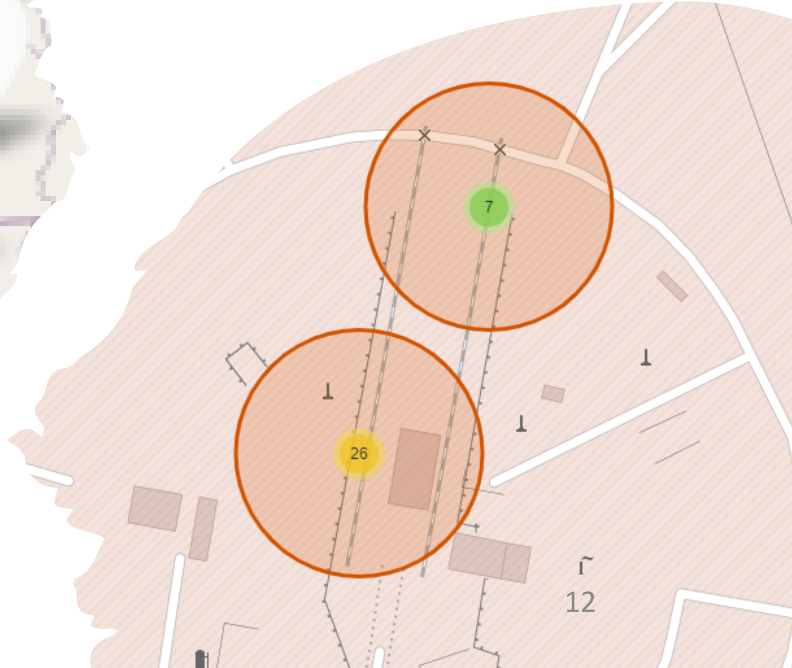




Method6: Building a Dashboard with Plotly Dash

- Plotly dash was used to build an interactive dashboard
- Various Graphs and Charts were drawn to show launch, orbit and landing features.



Method6: Predictive Analysis

- Using numpy and pandas the data we obtained from previous activities were split into training and testing classifications
- Machine learning models were built using GridSearchCV
- Accuracy metrix was used for modeling
- Selection was made on best performing model

Significant Insights from Exploratory Data Analysis

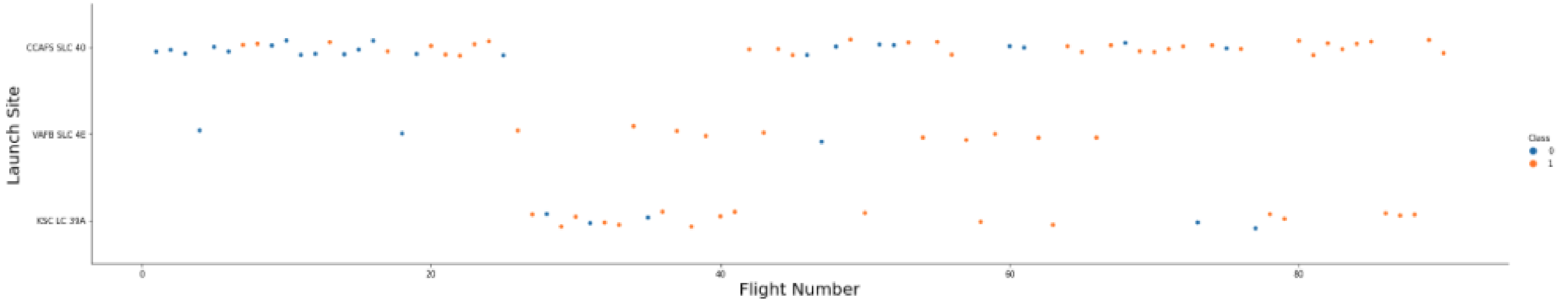
- Following insights were drawn from the data available for the Project
- Success rate depends on the number of flights launched from any site: The site with greatest launches was the most successful
- Success rate depends on the payload mass, the greater the mass of a rocket payload the greater chances it has for success
- Type of Orbit has direct relationship with rate of success, certain orbits hold greater successes
- As discussed in previous slides, the most recent years have seen a greater success rate from 2013 onwards

Some useful information obtained from data using SQL

S/No.	Instance	Readings
1	Total Launch Sites	4
2	Total Payload carried by NASA’s Rockets	45,596 KG
3	Falcon flies with average Mass	29,284 KG
4	First Successful Flight was on	22 December, 2015
5	Success Ratio	89%

Discussion on Results

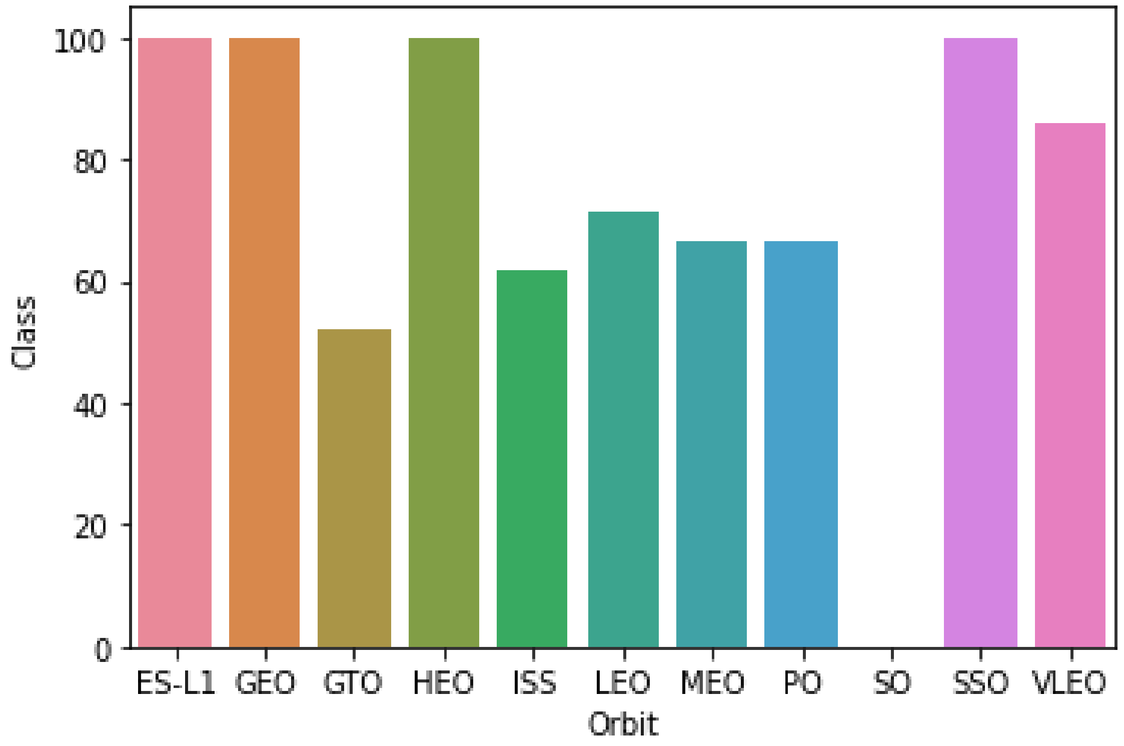
Flight Numbers plotted against Launch Sites



- Over the period of time, launch sites has been shifted
- Most flight successes were obtained from one particular site, i.e KSC LC 39A
- Most unsuccessful flights were made from CCAFS SLC 40

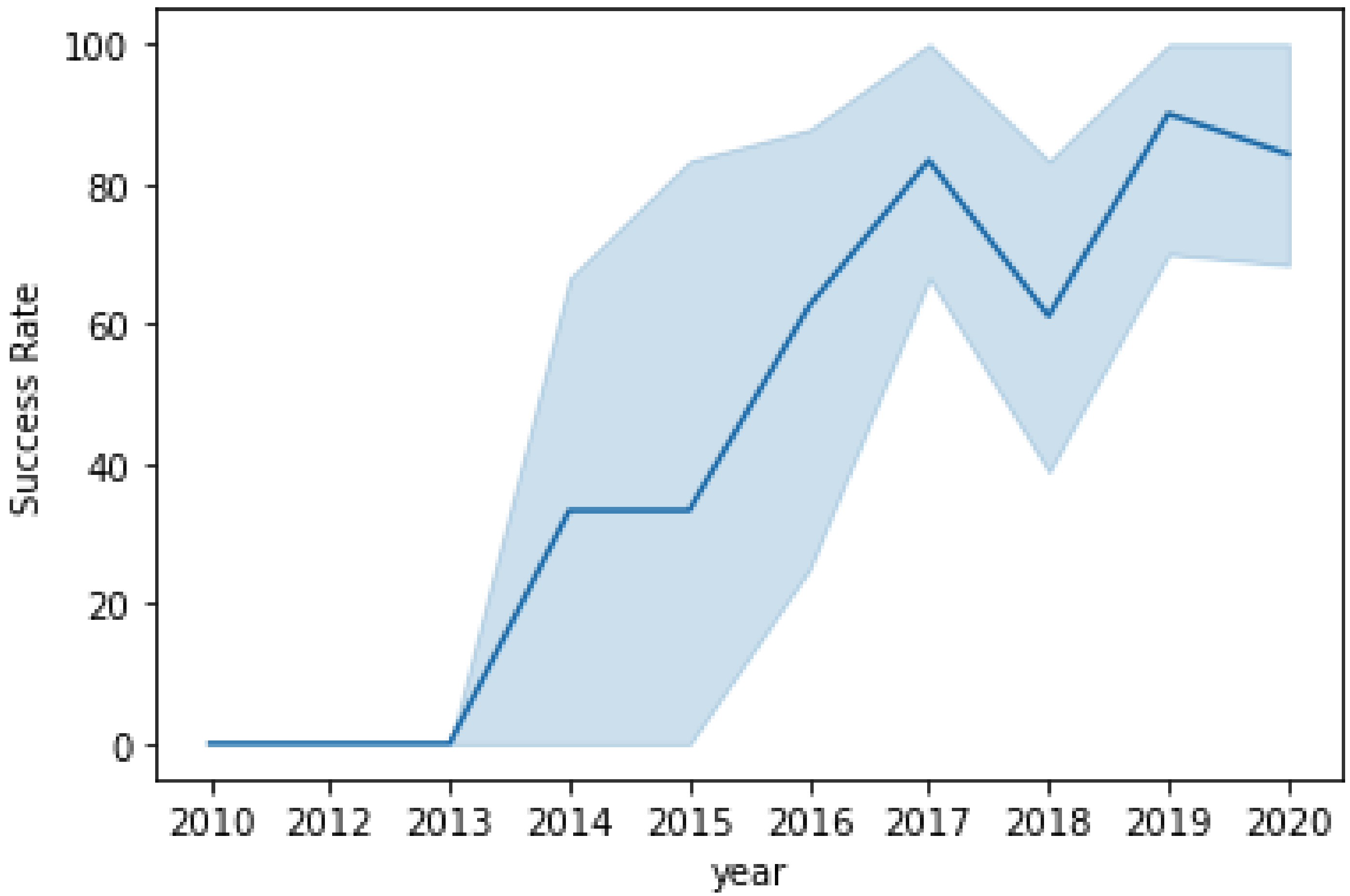
Orbital Success Rate

- Flights flew to different orbits with variable success rates
- No success was observed at SO Orbit.
- Total Success was observed for four orbits, i.e., ES-L1; GEO; HEO; SSO



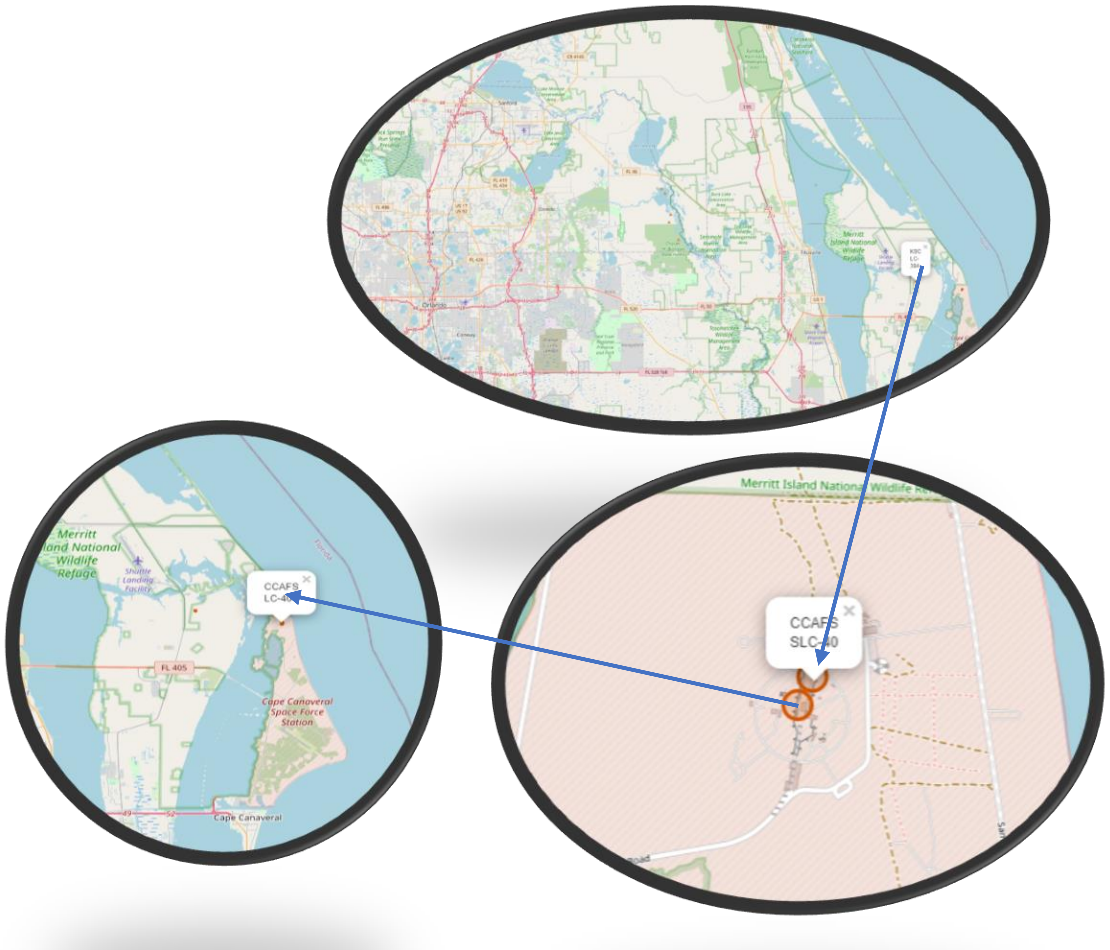
Yearly Recorded Success Trends

- No success observed from 2010 to 2013
- Success started from 2013 and went on to reach maximum in 2019.
- Slight dips in trend were observed in 2015, 2018 and 2020.

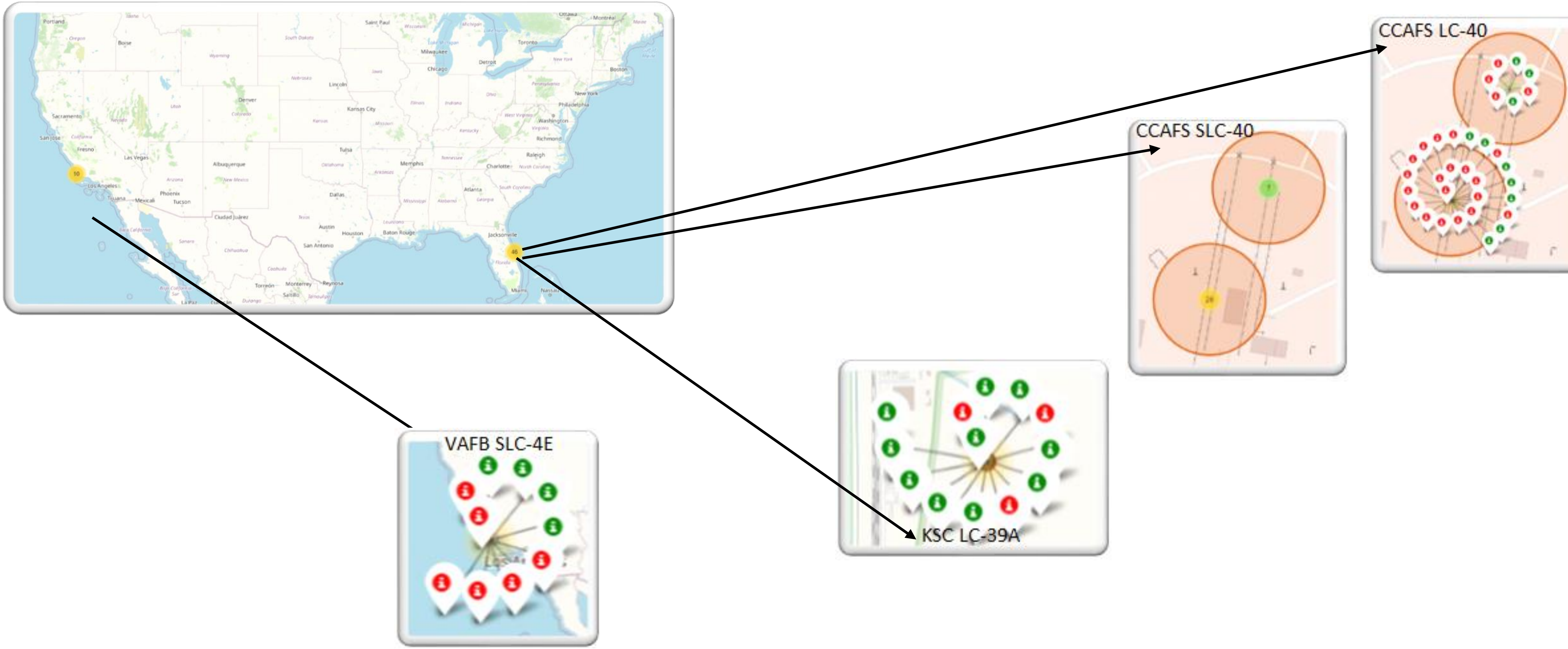


Using Folium for Interactive Mapping

- Launch Sites as shown in terrain maps of USA.

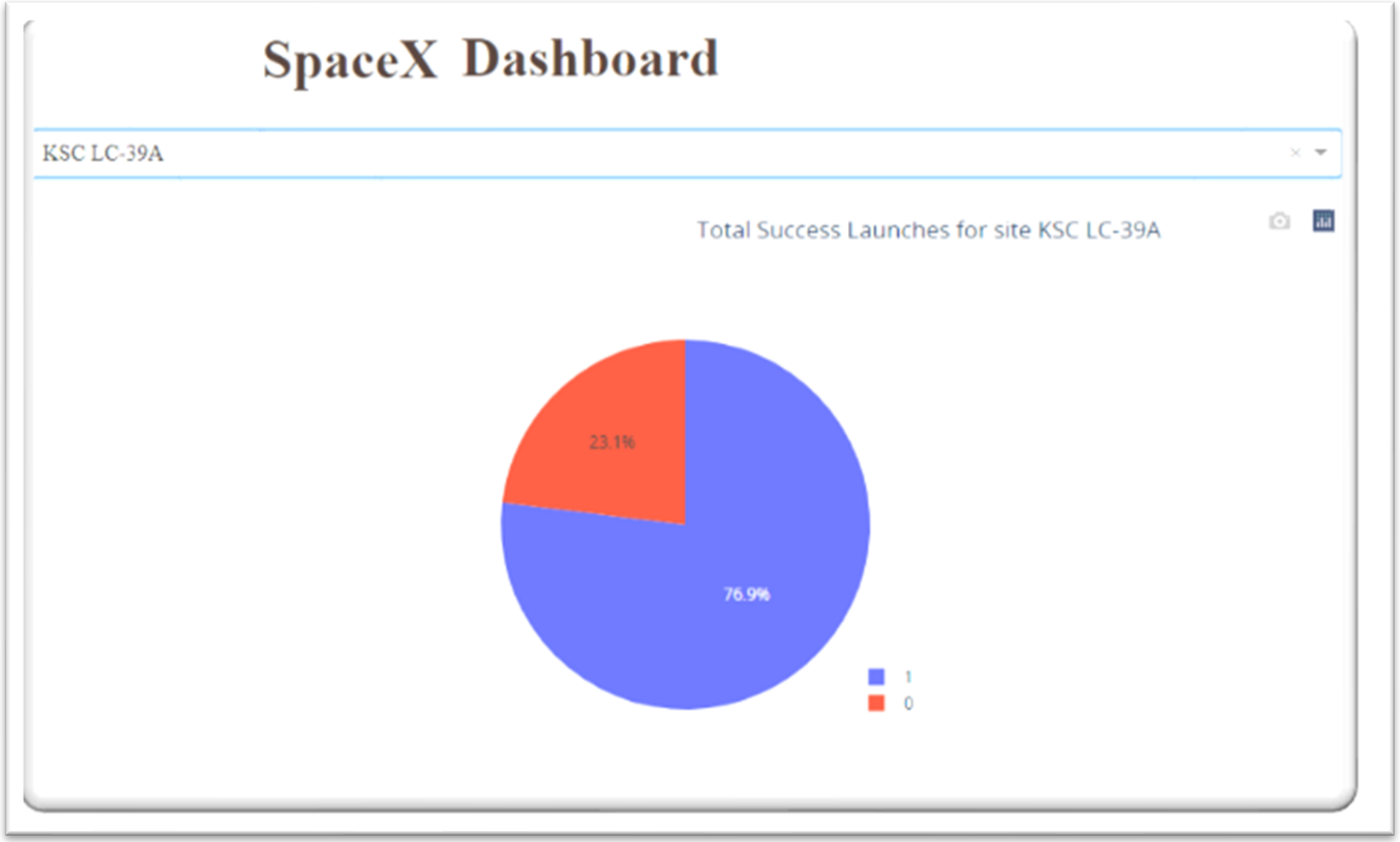


Launch Sites with Success/Failure Statistics



Interactive Dashboard with PLOTLY DASH

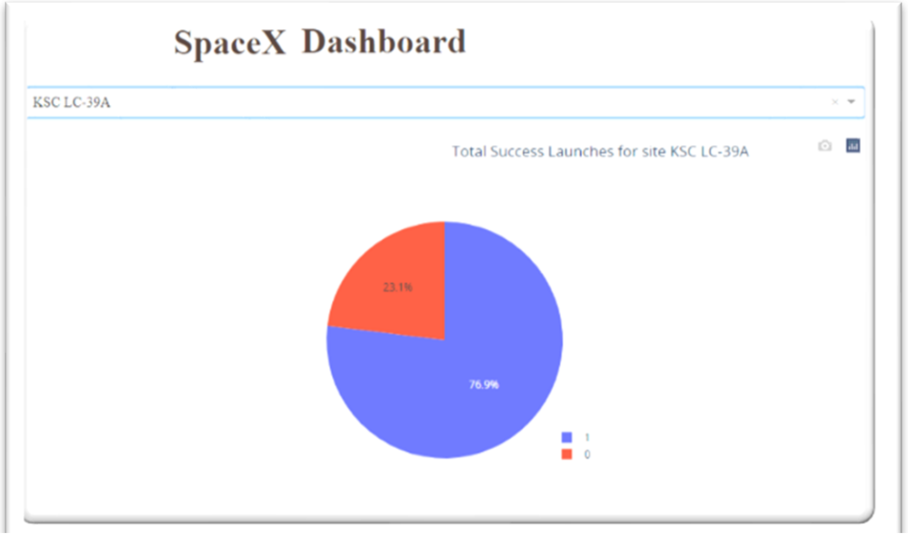
Highest Launching Success Ratio

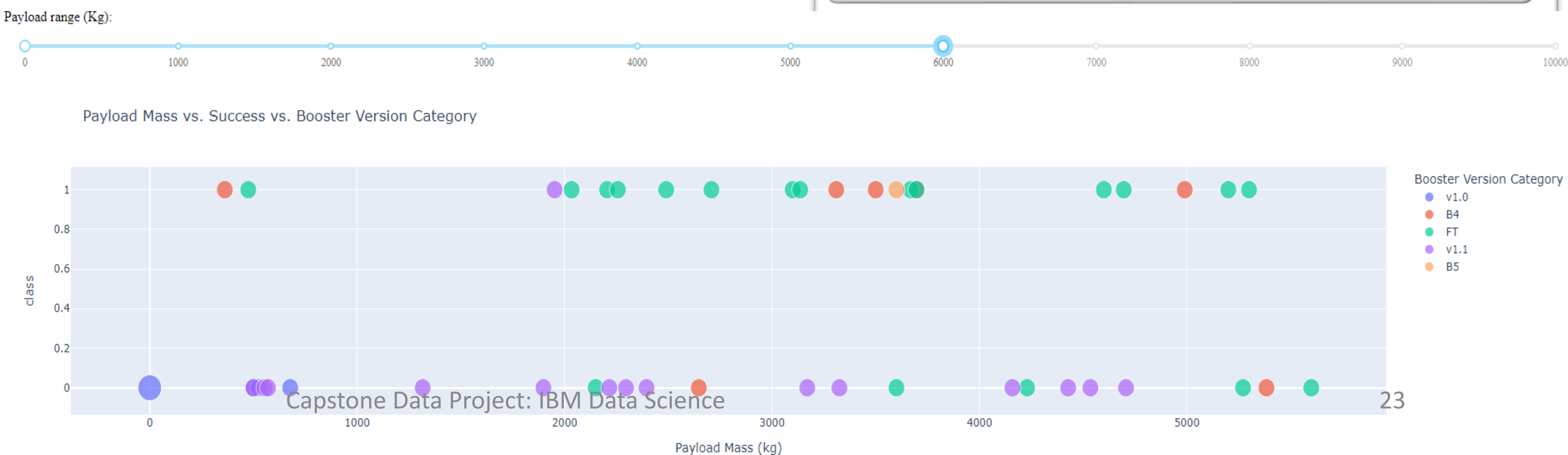


- Launch site KSC LC-39 also has the highest ratio success ratio with a ratio of 76.9%.

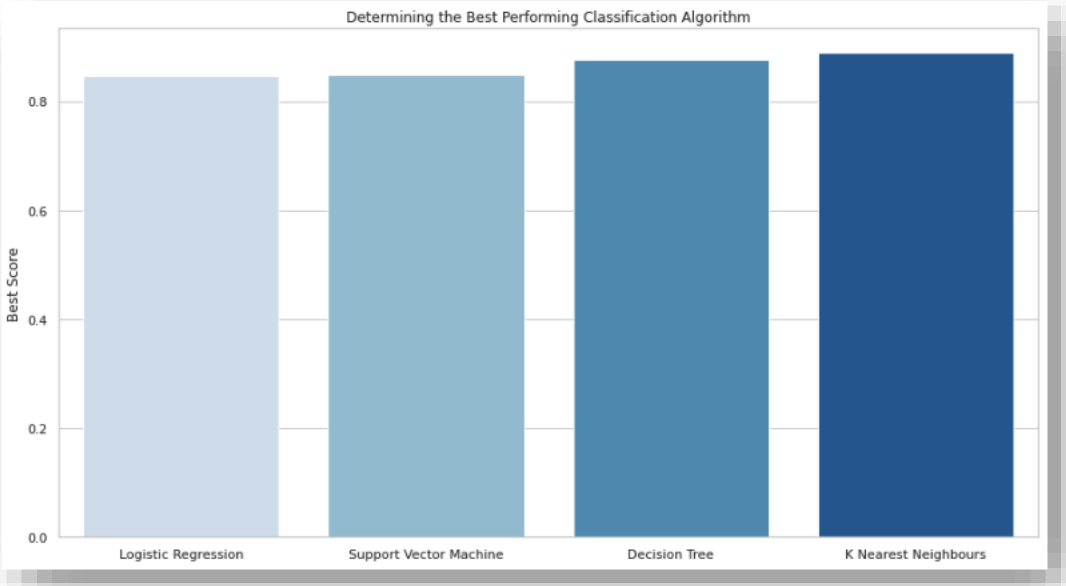
PLOTLY Dash was used for creating Dashboard

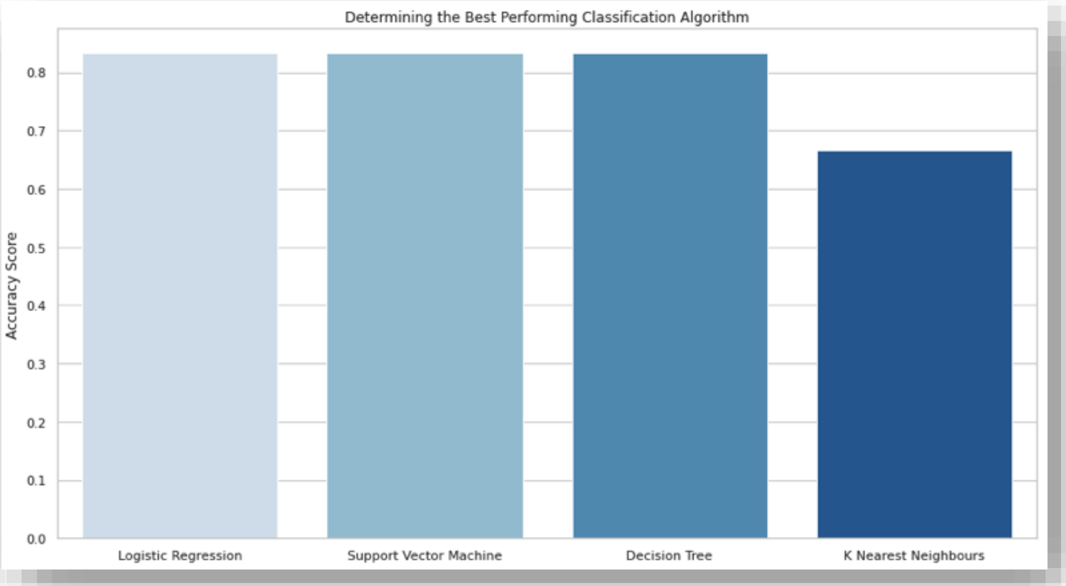
- An interactive dashboard with PI Charts and Bar Charts was created using plotly dash.



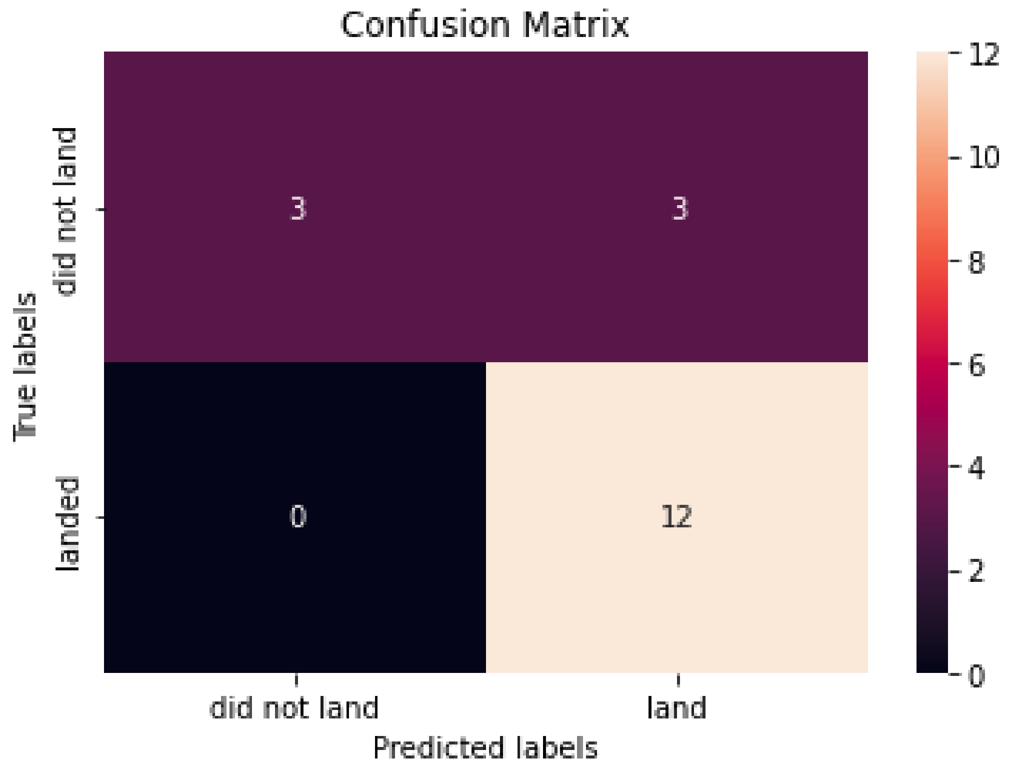


Predictive Analysis Model Selection





	Algorithm	Accuracy Score	Best Score
0	Logistic Regression	0.833333	0.846429
1	Support Vector Machine	0.833333	0.848214
2	Decision Tree	0.833333	0.876786
3	K Nearest Neighbours	0.666667	0.889286



Predictive Analysis

Being the best analysis model available, the Predictive Analysis was completed using decision tree and the confusion matrix thus obtained provided following predictions

Successful Launches predicted = 12

Unsuccessful Launches predicted = 3

False positive launches successful = 3

Conclusion

(Unique Insights from the Project)

As the number of flights increased, the rate of success at a launch site increased.

Most of the early flights were unsuccessful.

Between 2010 and 2013, all landings were unsuccessful

After 2013, the success rate generally increased, despite small dips in 2018 and 2020.

Orbit types ES-L1, GEO, HEO, and SSO, have the highest success rate of 100%.

Launch site KSC LC-39 A had the most successful launches, with 41.7% of the total successful launches, and also the highest rate of successful launches, with a 76.9% success rate.

Thank You

- Following is the link to my git hub resource set.\

https://github.com/rehmansaif/Saif.git