IBM SpaceX Capstone

Dalia Bajriy

29/08/2022

IBM Developer

SKILLS NETWORK

Outline

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

Executive Summary

- Summary of methodologies
- Data Collection
  - Data Wangling
  - EDA with data visualization
  - EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)
- Summary Of all results
- EDA Result
- Interactive analytics
- Predictive analysis

Introduction

- Project background and context
- SpaceX advertises Falcon 9 rocket launches on its website, with a cost 62 million dollars; other providers cost upward of 165 ,million dollars each, ,much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers
- The project task to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Methodology

Section 1

IBM Developer

SKILLS NETWORK

Methodology

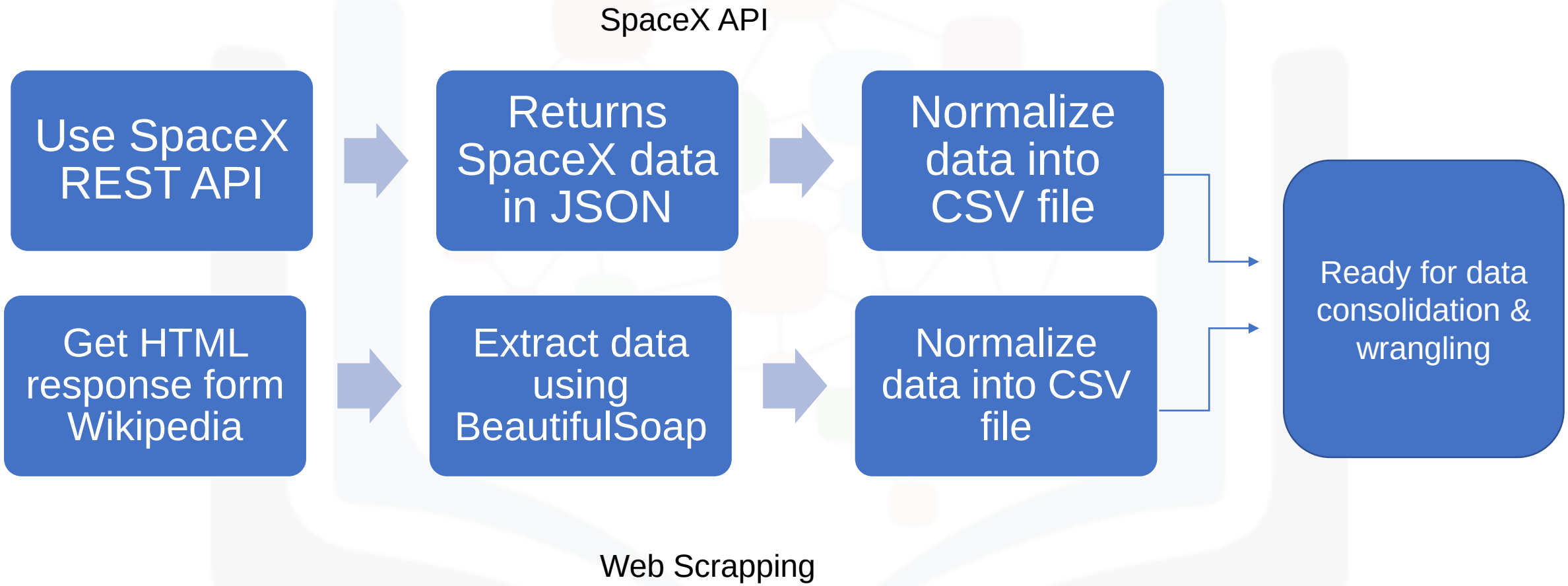
Executive Summary

- **Data collection methodology:**
- SpaceX Rest API
- Web Scrapping from Wikipedia
- **Perform data wrangling**
  - One Hot encoding data fields for Machine Learning and data cleaning of null values and irrelevant.
- **Perform exploratory data analysis (EDA) using visualization and SQL.**
- **Perform interactive visual analytics using Folium and Plotly Dash.**
- **Perform predictive analysis using classification models.**
- LR, KNN, SVM, DT models have been built and evaluated for the best classifier.

Data Collection

- **The following datasets was collected:**
-
- SpaceX launch data that is gathered from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.

Data Collection



Data Collection – SpaceX API

1. Getting response from API

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

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

2. Converting response to .JSON file

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

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

3. Apply custom function to clean data

```
# Call getBoosterVersion
getBoosterVersion(data)
```

```
BoosterVersion[0:5]
```

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(data)
```

4. Assign list to dictionary the dataframe

```
# Call getBoosterVersion
getBoosterVersion(data)
```

the list has now been update

```
BoosterVersion[0:5]
```

```
['Falcon 1', 'Falcon 1', 'Falcon 1', 'Falcon 1', 'Falcon 9']
```

we can apply the rest of the functions here:

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(data)
```

```
# Create a data from launch_dict
data = pd.DataFrame(launch_dict)
```

5. Filter dataframe and export to flat file (CSV)

```
data_falcon9 = data[data.BoosterVersion == 'Falcon 9']
data_falcon
```

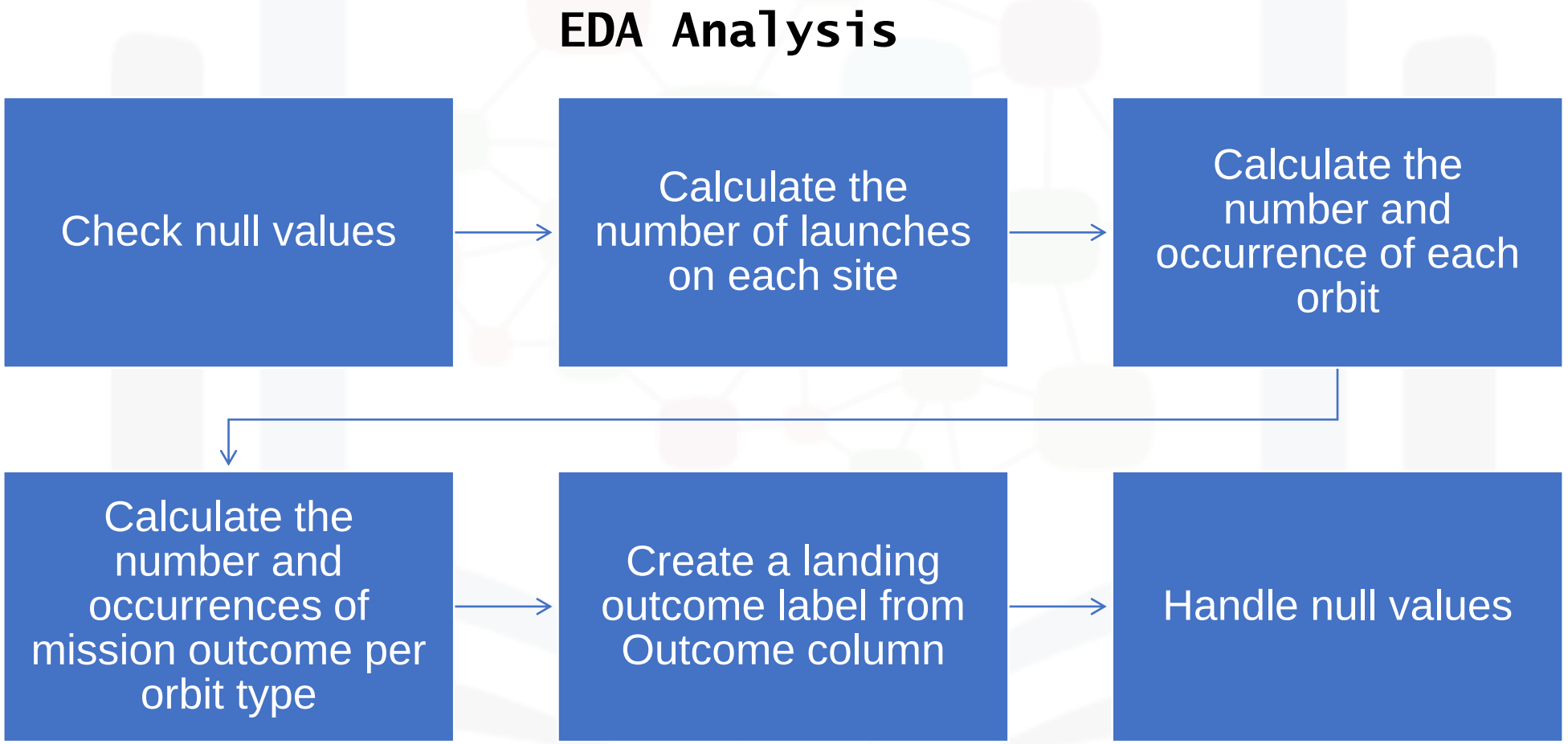
```
data_falcon9.loc[:,'FlightNumber'] = list(range(1, data_falcon9.shape[0]+1))
data_falcon9
```

IBM Developer

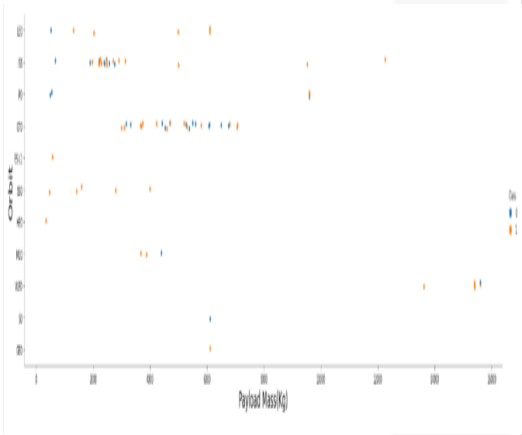
Data collection with SpaceX REST API

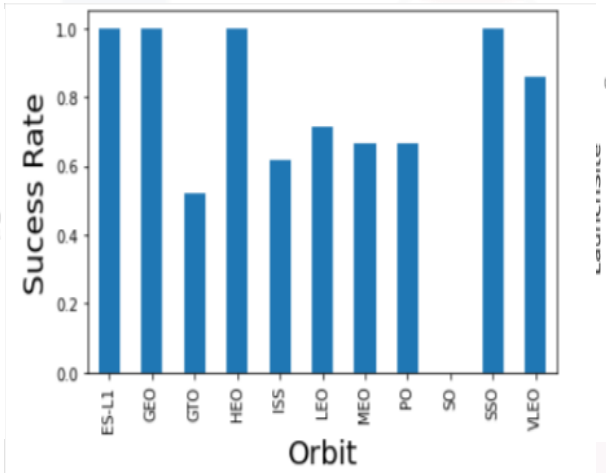
SKILLS NETWORK

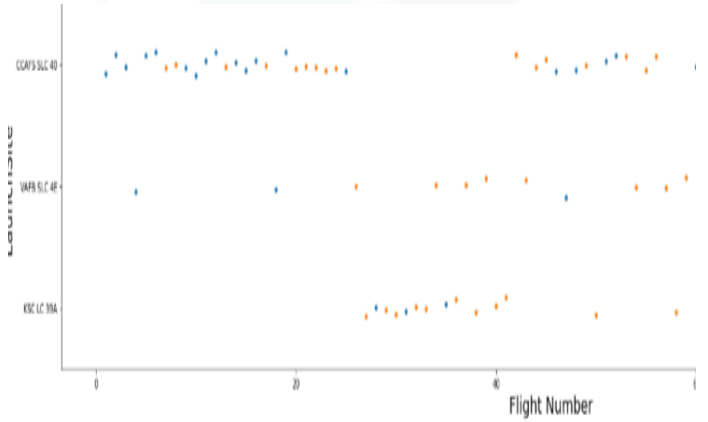
Data Wrangling

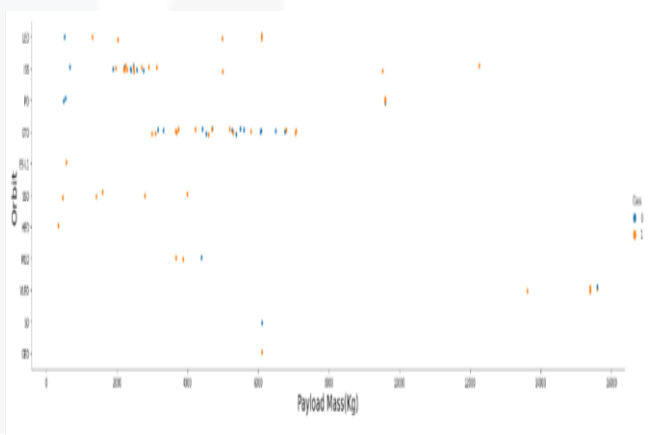


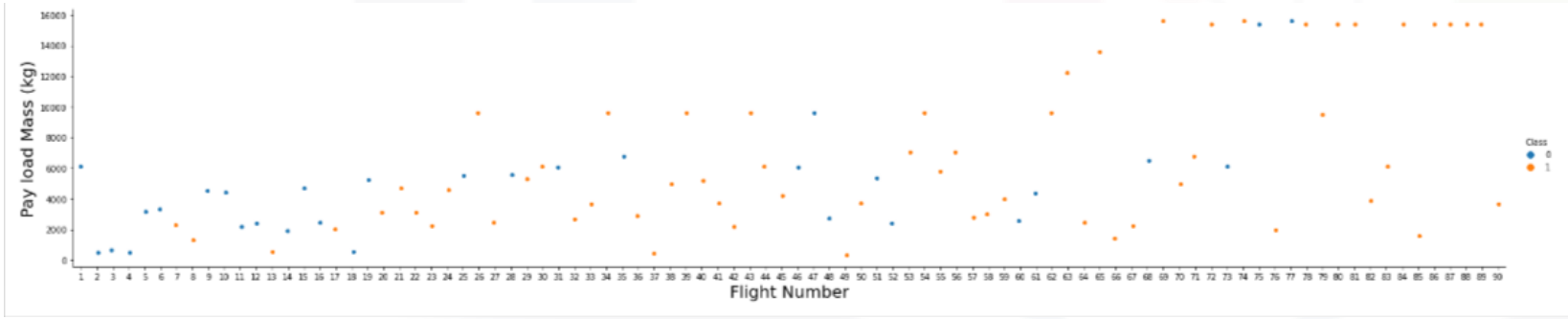
EDA With Data Visualaization

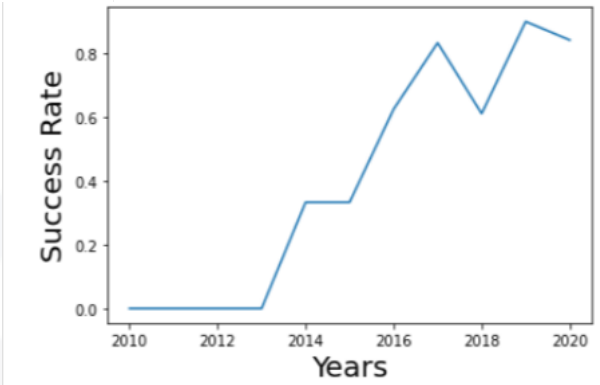










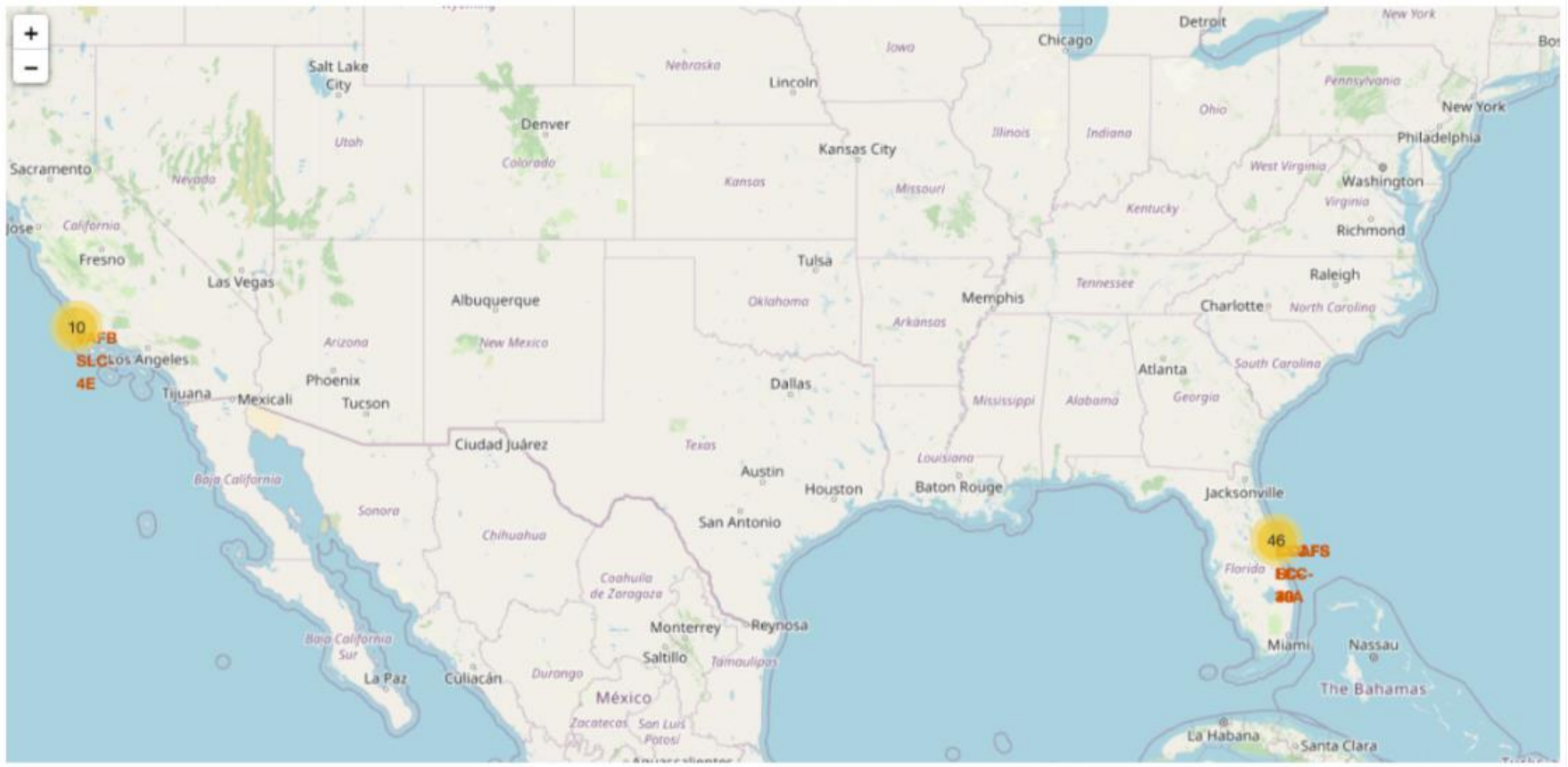


EDA with SQL

https://github.com/daliabajri/IBMCapstone/blob/main/EDA_SQL.ipynb

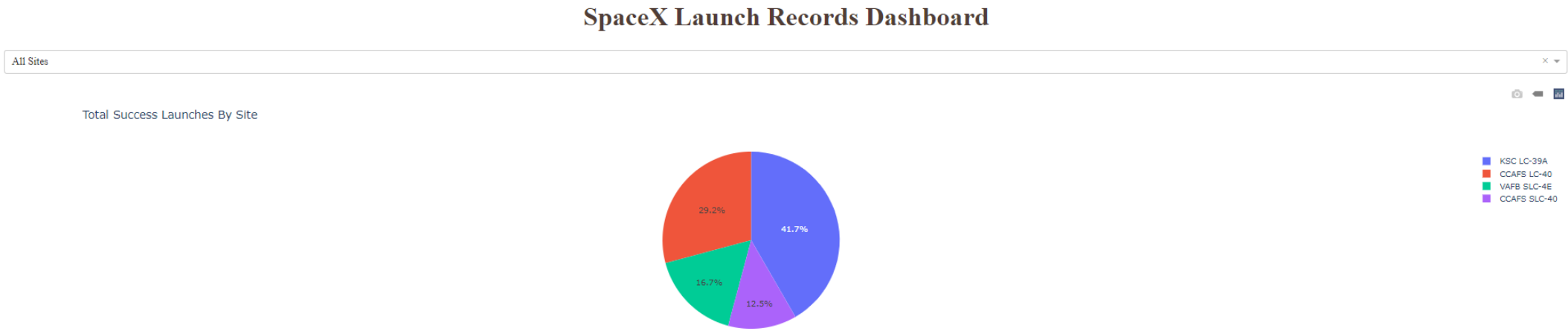
- **SQL queries performed include:**
- Displaying the names of the unique launch sites in the space mission.
- Displaying 5 records where launch sites begin with the string “KSC”.
- Displaying the total payload mass carried by boosters launch by NASA (CRS).
- Displaying average payload mass carried by booster version v1.1
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes.
- Listing the names of the boosters version which have carried the maximum payload mass.
- Listing the records which will display the month names, successful landing, outcomes in ground pad booster.
- Versions launch site for the months in year 2017.
- Ranking the count successful landing outcomes between the date 2010 06 04 and 2017 03 20 in descending order.

Build An Interactive Map With Folium



Map markers have been added to the nap with aim to finding an optimal location for building launch site
https://github.com/daliabajri/IBMCapstone/blob/main/lab_jupyter_launch_site_location.ipynb.ipynb

Build A Dashboard With Plotly Dash

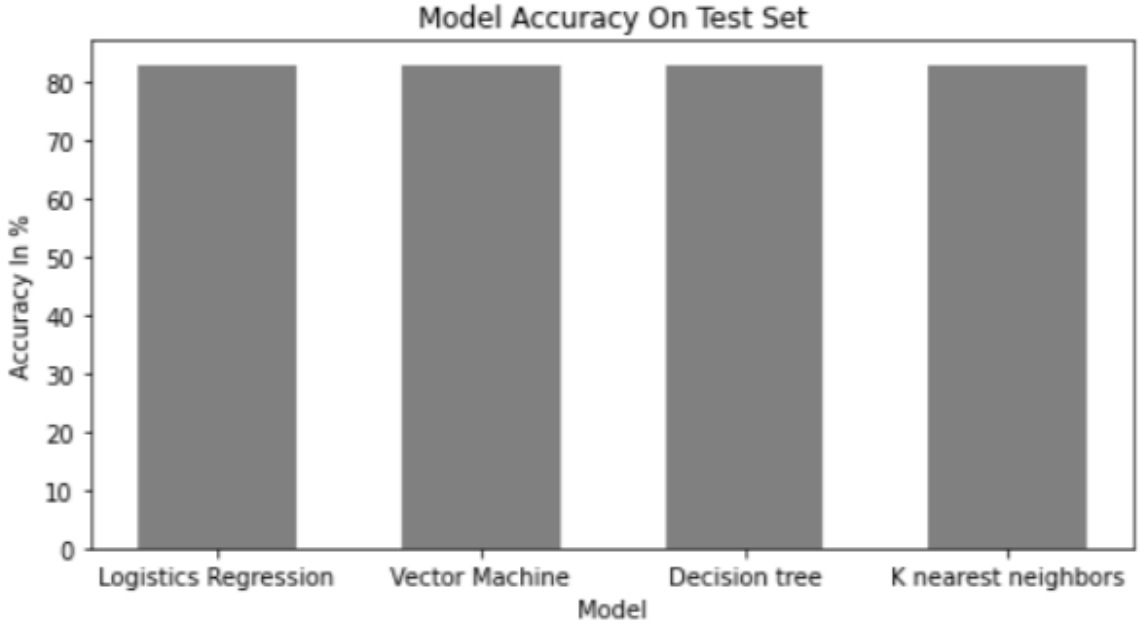


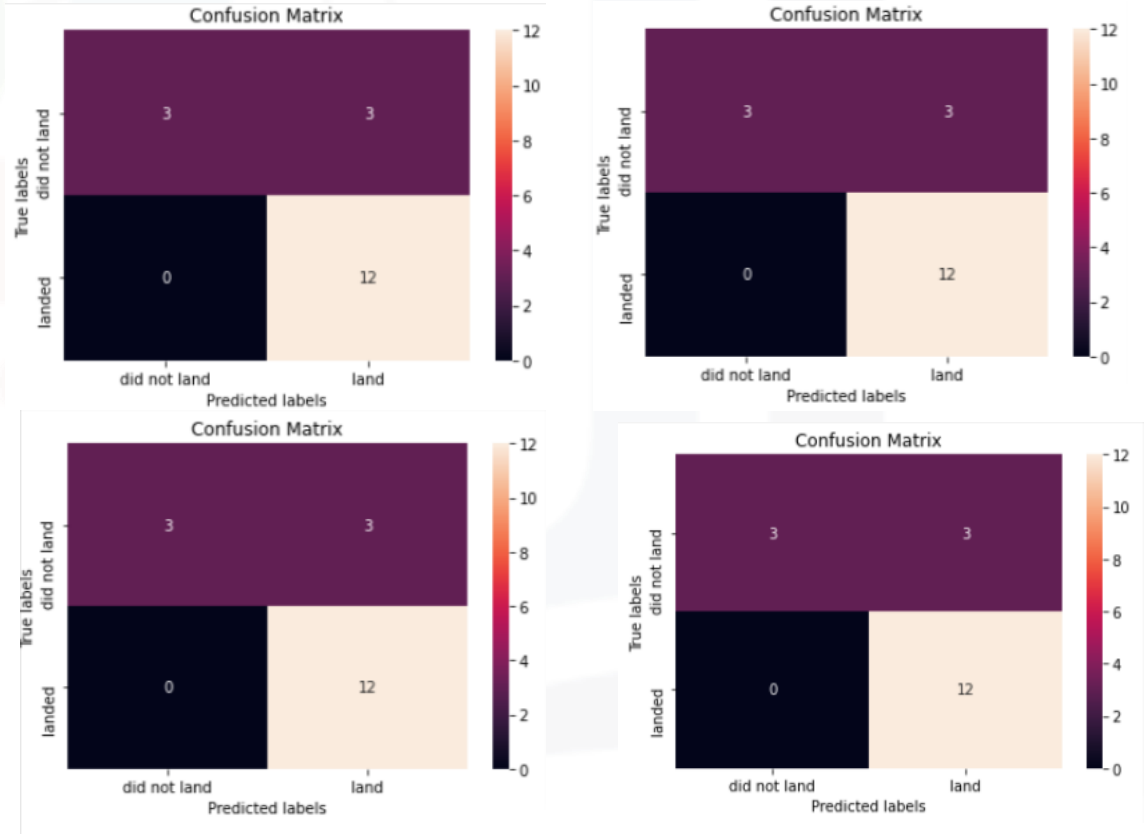
We can see that KSC LC-39A had the most successful launches from all sites

We can see that KSC LC-39A had the most successful launches from all sites

Predictive Analysis (Classification)

- The SVN, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.





Practically all these algorithms give the same result".

IBM Developer

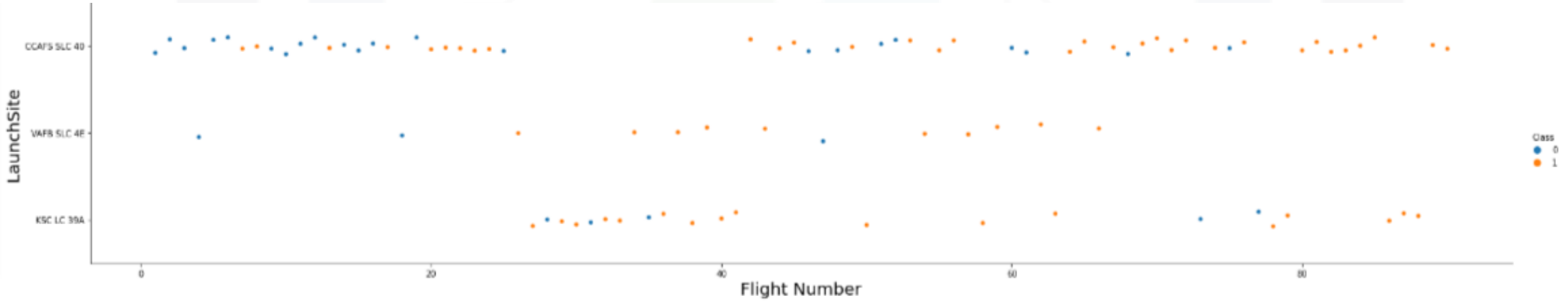
SKILLS NETWORK

Results

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for the dataset.
- Low weighted payloads performs better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO, HEO, SSO, ES L1 has the best Success Rate.

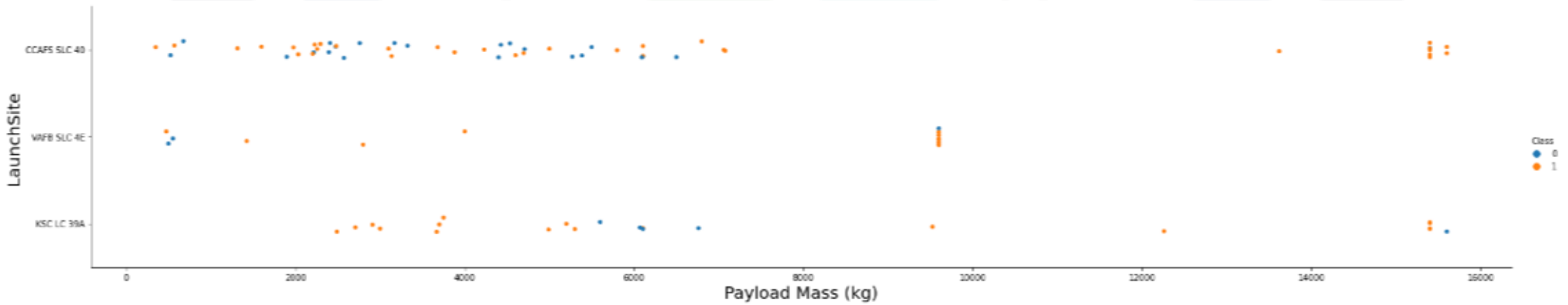

Section 2
Insights Drawn From EDA
IBM Developer
SKILLS NETWORK

Flight Numbers Vs. Launch Site



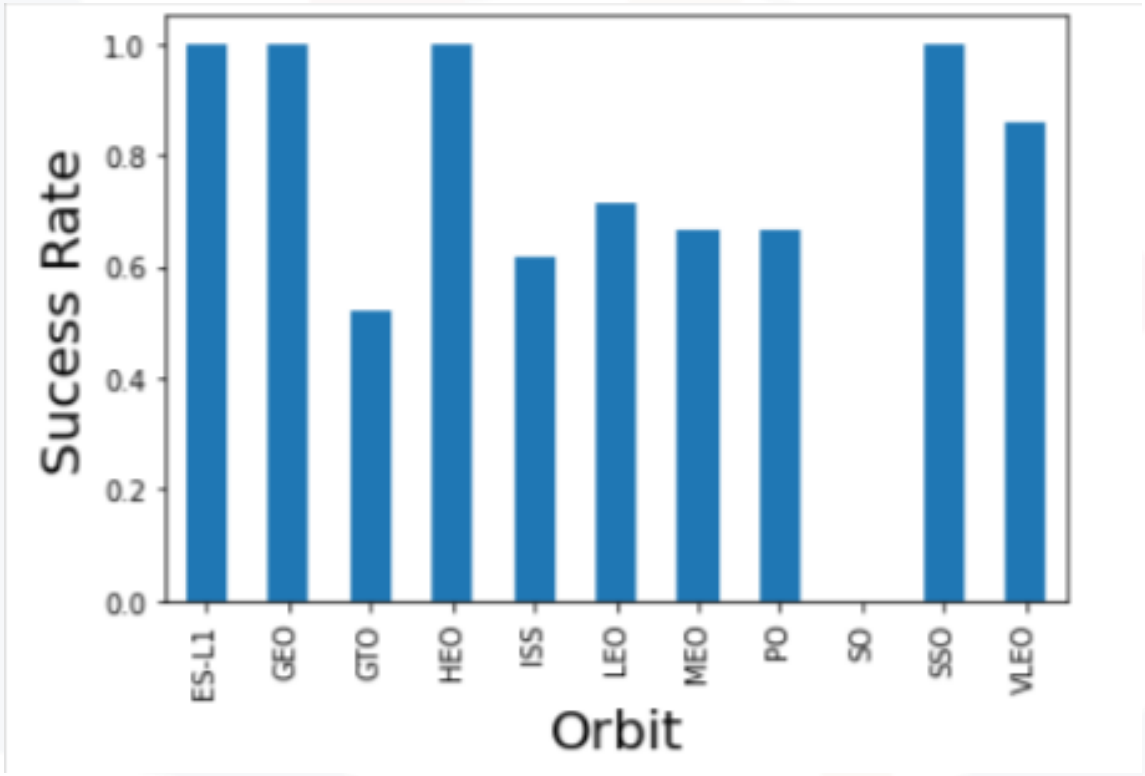
- Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites.

Payload Vs. Launch Site



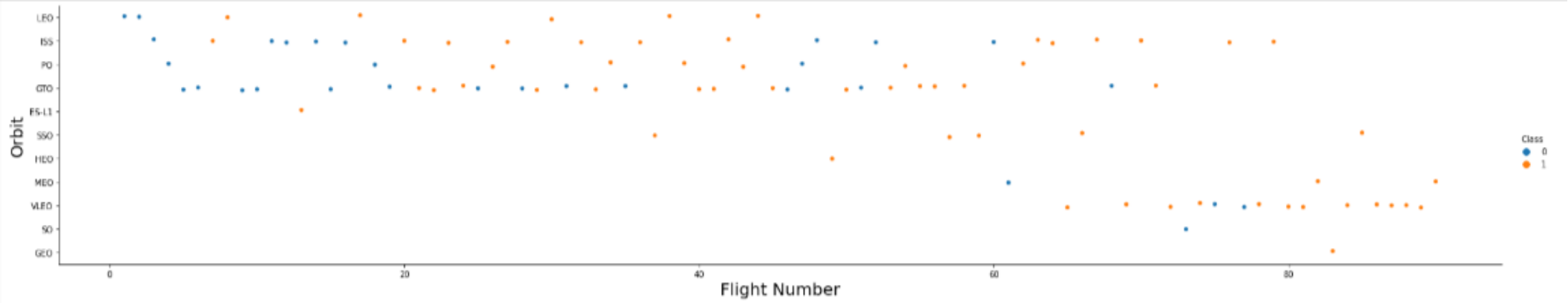
- The majority of Payloads with lower Mass have been launched from CCAFS SLC 40.

Success Rate Vs. Orbit Type



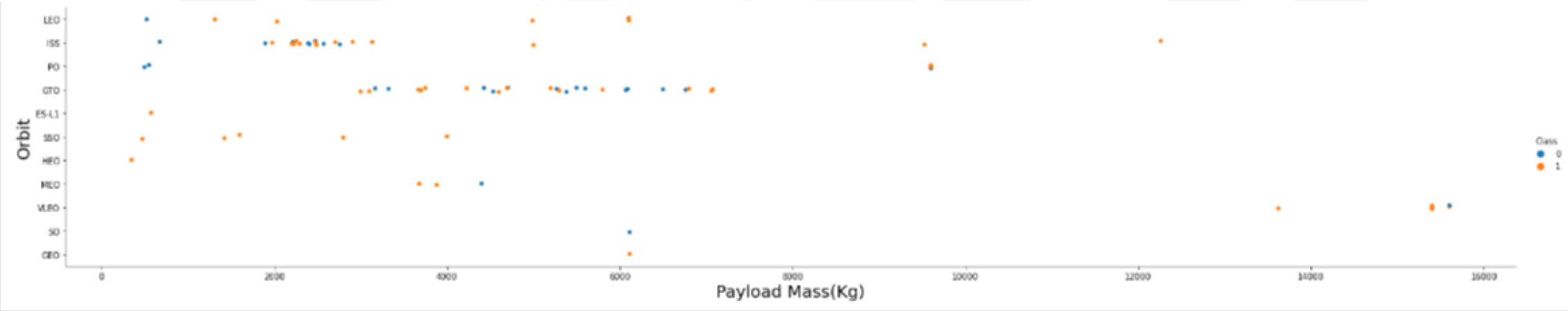
- The Orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number Vs. Orbit Type



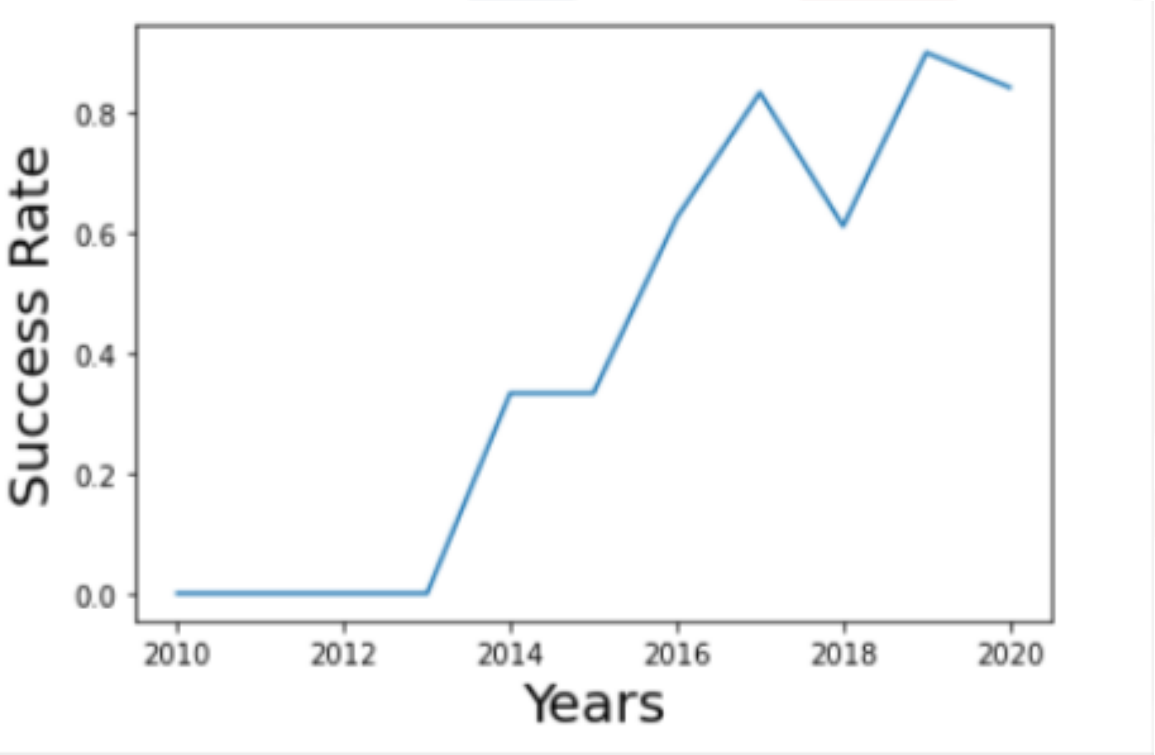
- A trend can be observed of shifting to VLEO launches in recent years.

Payload Vs. Orbit type



- There are strong correlation between ISS and Payload at the range around 2000. as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stablished since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

Launch Site Names Begin With ‘CCA’

- **%sql** SELECT LAUNCH_SITE from SPACEXTBL where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

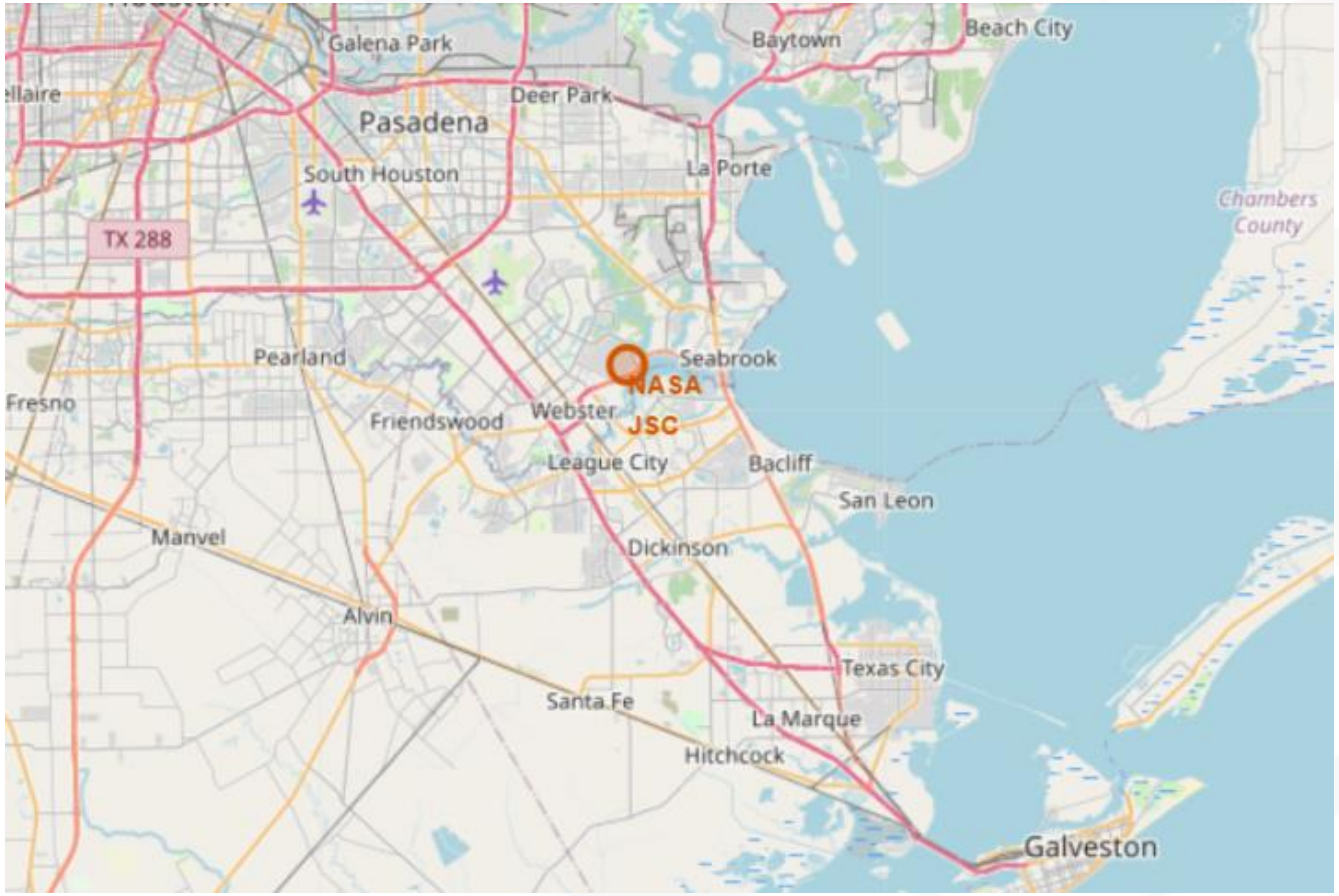
Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

- **%sql** select count(MISSION_OUTCOME) as missionoutcomes from SPACEXTBL GROUP BY MISSION_OUTCOME;

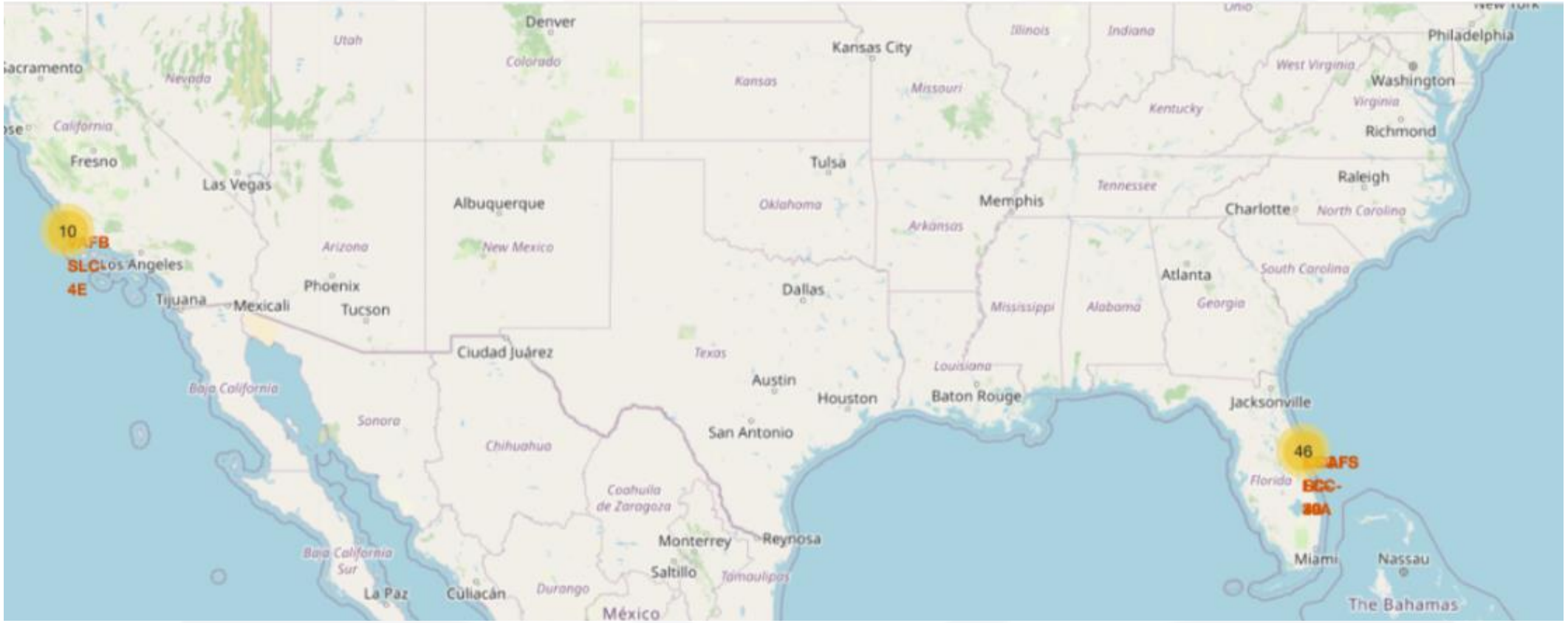
landing__outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)
No attempt
Controlled (ocean)
Failure (drone ship)


Section 3
Launch Sites
Proximities Analysis
IBM Developer
SKILLS NETWORK

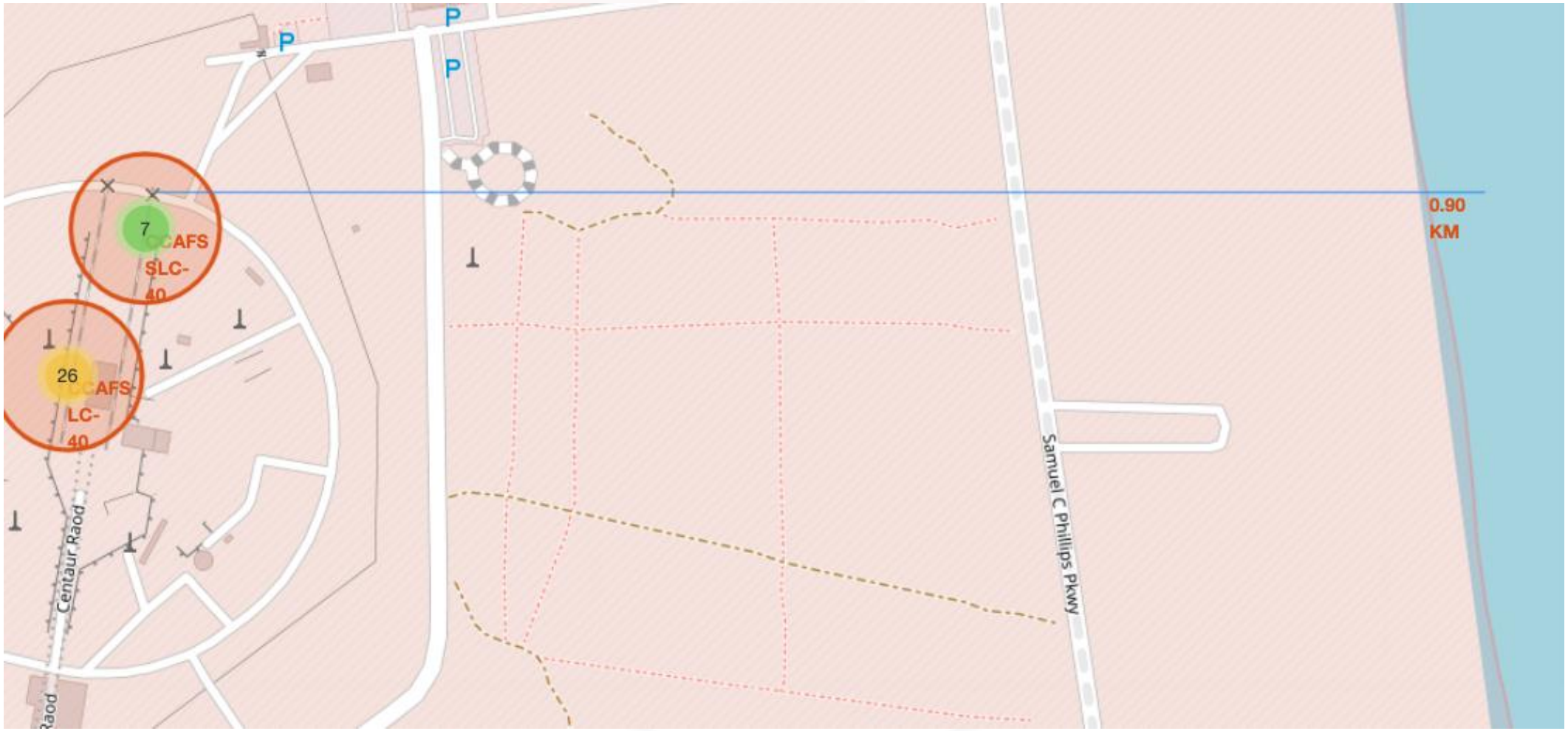
All launch sites on a map



Success/failed launches for each site on the map



The distances between a launch site to its proximities



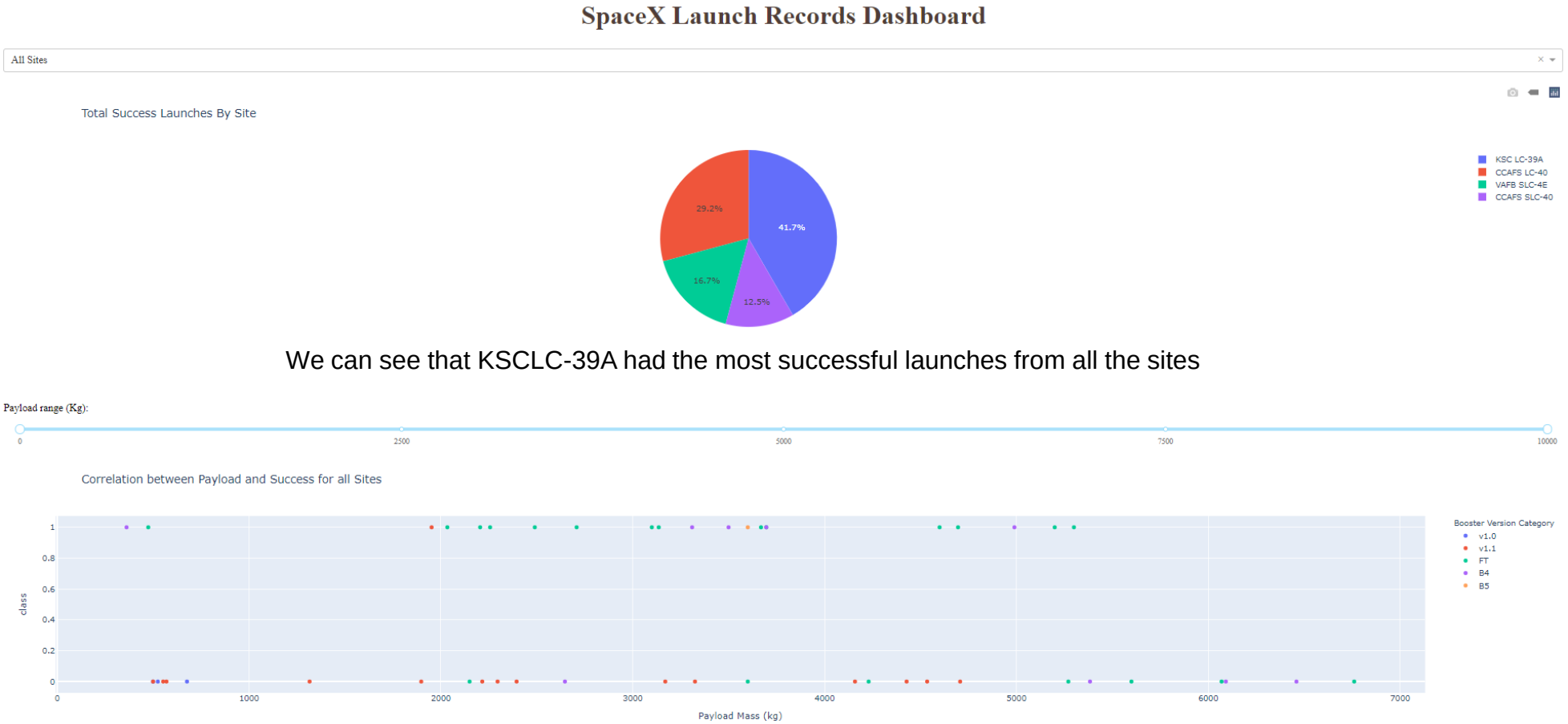
Section 4

Build a Dashboard With Plotly Dash

IBM Developer

SKILLS NETWORK

Payload Vs. Launch Outcome



We can see that KSCLC-39A had the most successful launches from all the sites

We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

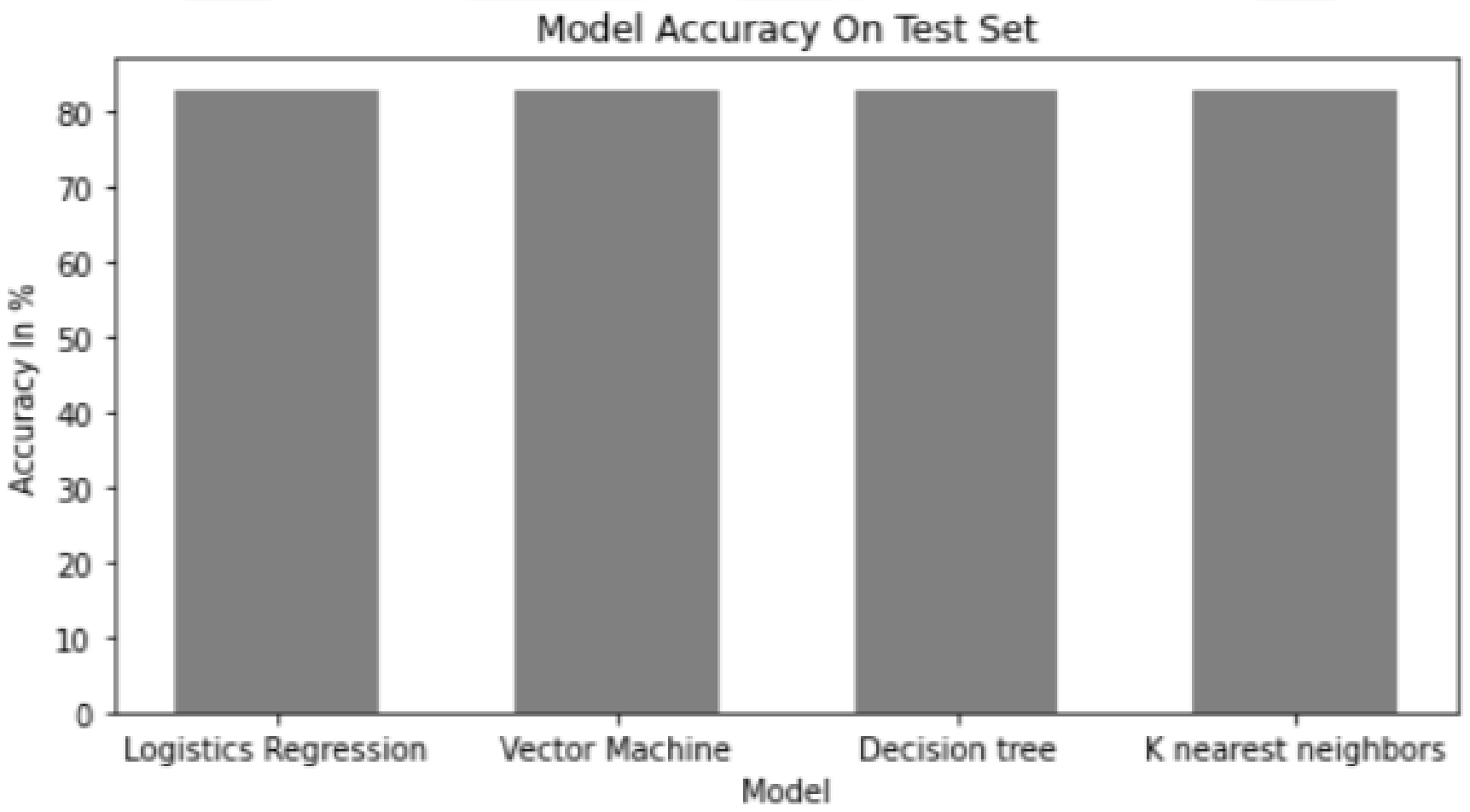
Section 5

Predictive Analysis (Classification)

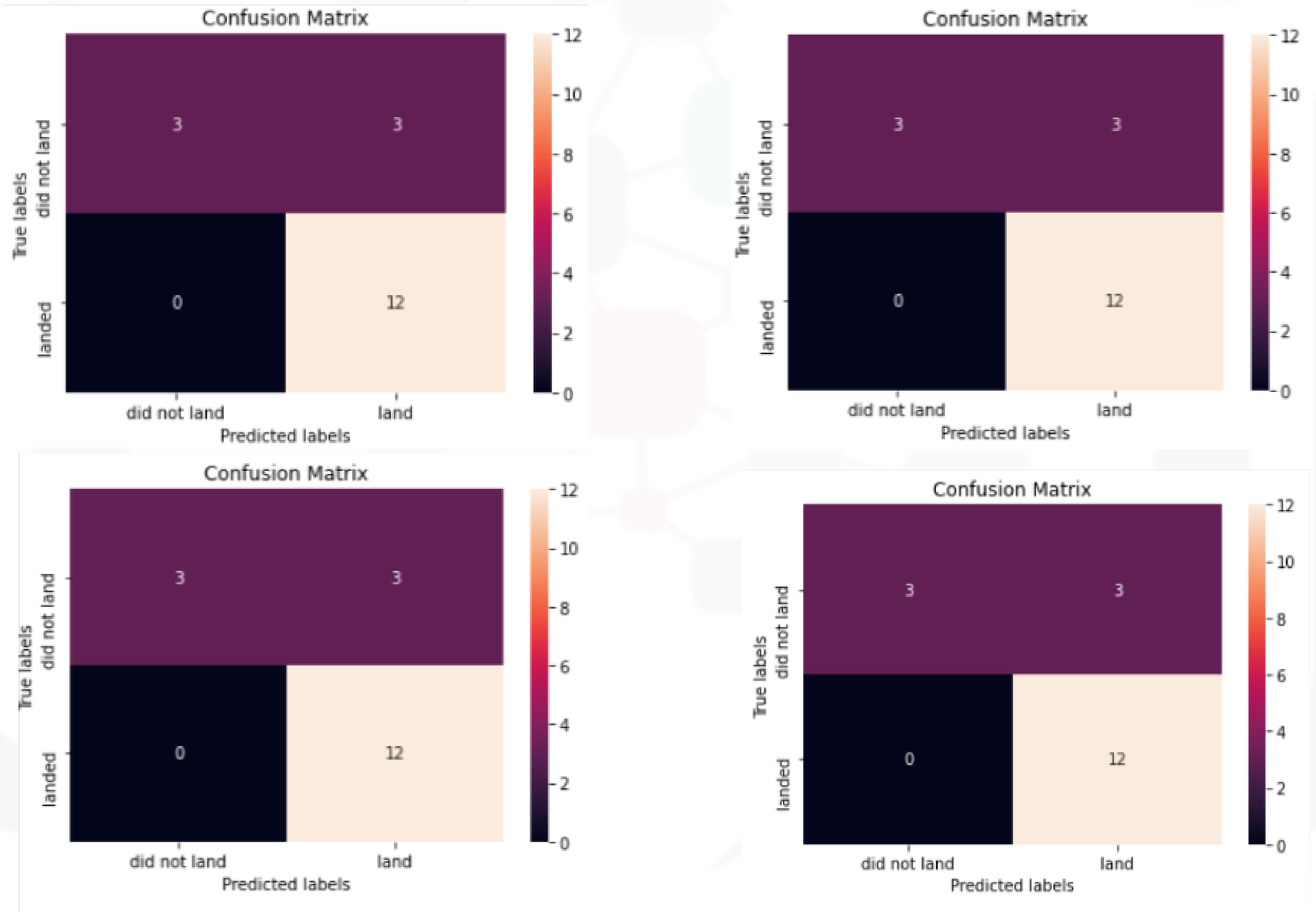
IBM Developer

SKILLS NETWORK

Classification Accuracy



Confusion Matrix



Conclusion

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for the dataset.
- Low weighted payloads performs better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO, HEO, SSO, ES L1 has the best Success Rate.


Thank You
IBM Developer
SKILLS NETWORK