IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Samritha A.R
05-Dec-2023

Outline

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

Executive Summary

Summary of Methodologies

- Data Collection
- Data Scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with with Data Visualization
- Interactive Visual Analytics with Folium
- Interactive Visual Analytics with Ploty Dash
- Machine Learning Prediction Analysis (Classification)

Summary of All Results

- EDA Results
- Interactive Analytics
- Predictive Analytics

Introduction

Project Background and Context

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 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 is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully or not?

Section 1

Methodology

Methodology

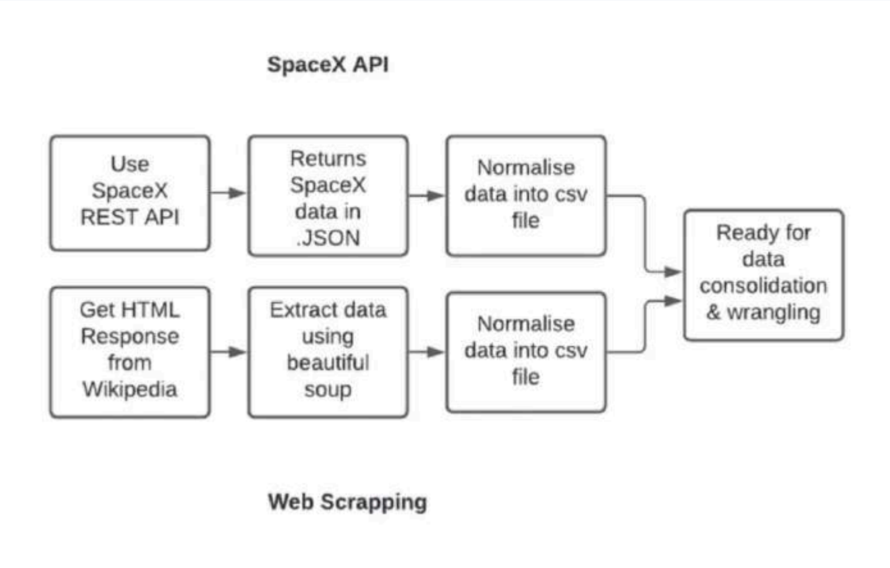
Executive Summary

- 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 columns
- 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:

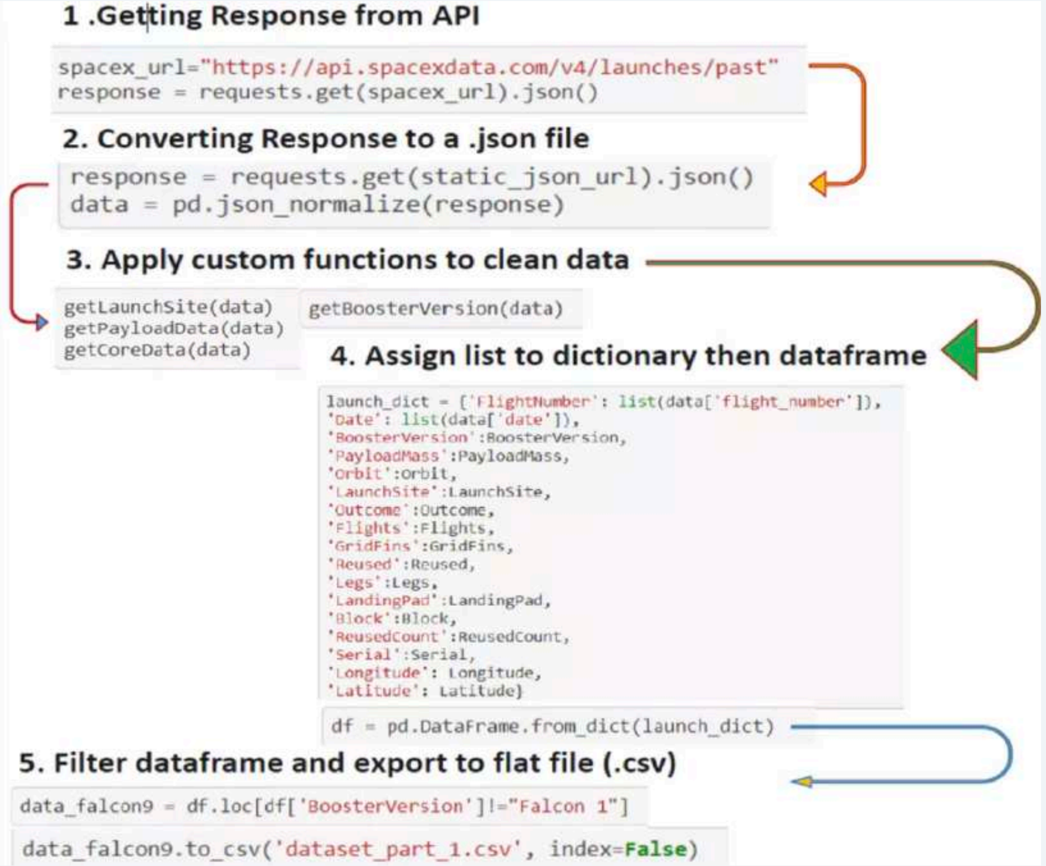
- SpaceX launch data that is gathered from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
- Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup



Data Collection – SpaceX API

- Data collection with SpaceX Rest calls
- GitHub URL:

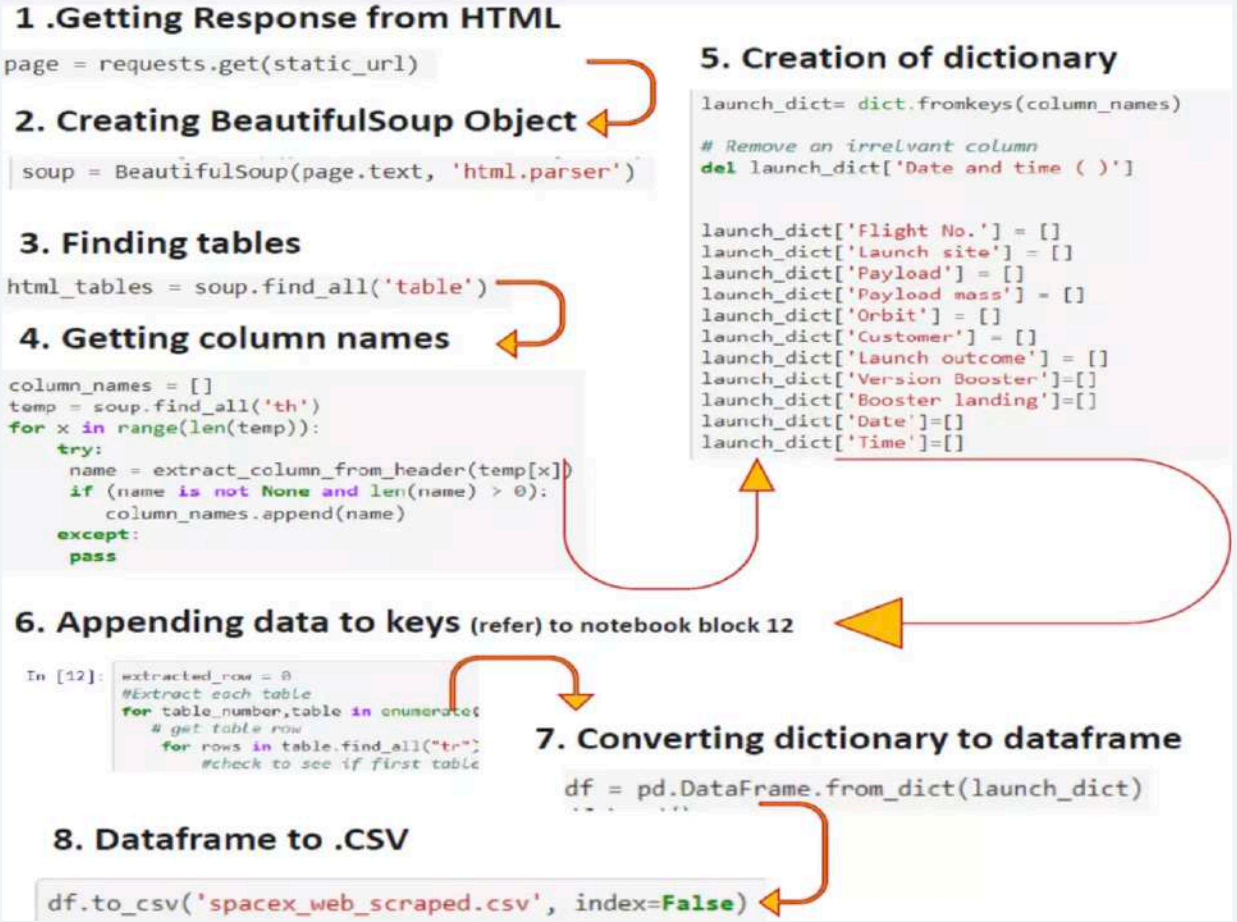
  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%201-Data_Collection.ipynb



Data Collection - Scraping

- Web scraping from Wikipedia
- GitHub URL:

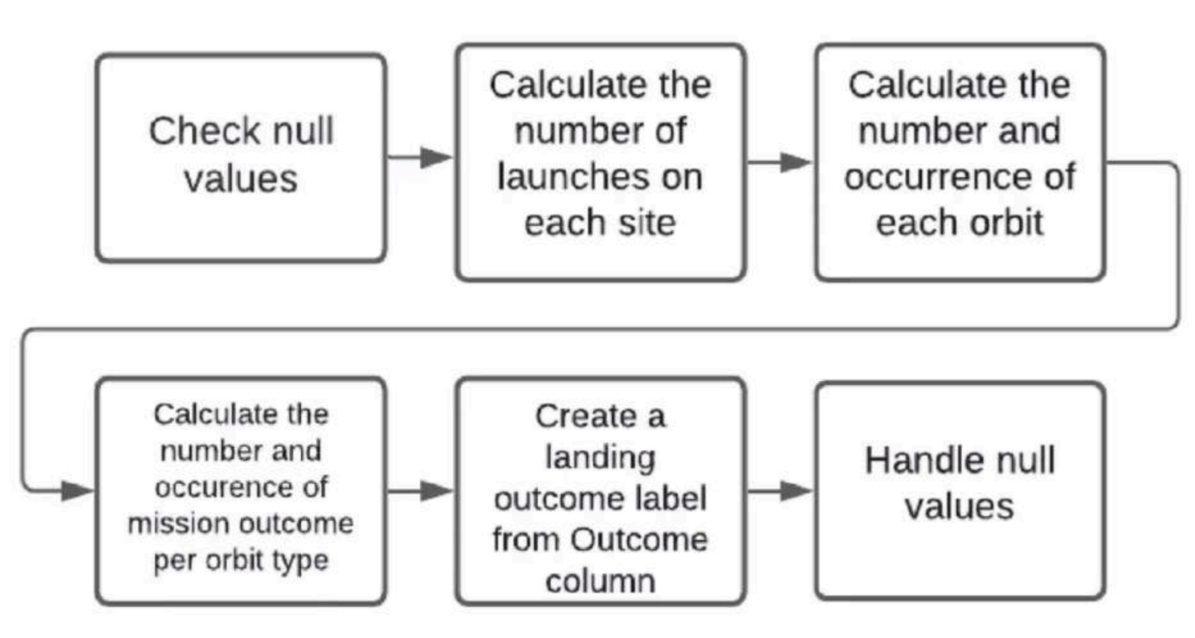
  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%202-Web_Scraping.ipynb



Data Wrangling

- Transforming and mapping raw data into a more suitable format for analysis
- GitHub URL:

  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%204-ExploratoryDataAnalysis_SQL.ipynb

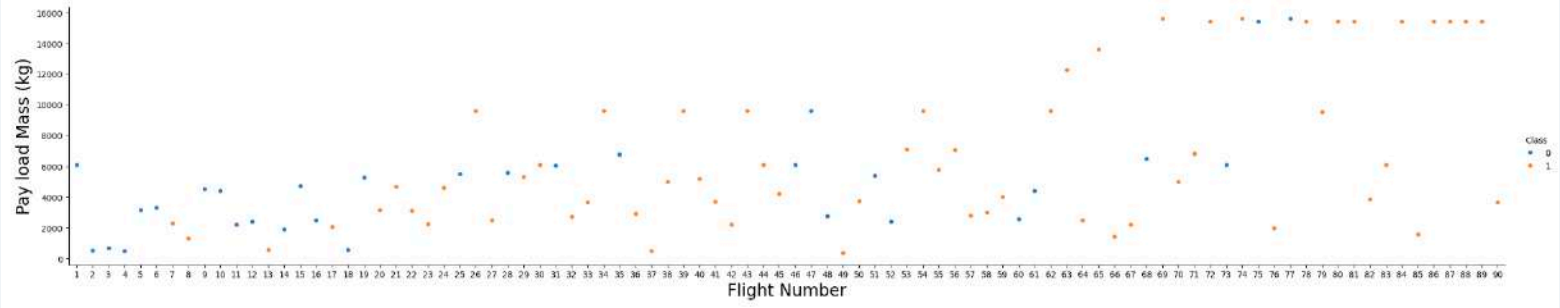


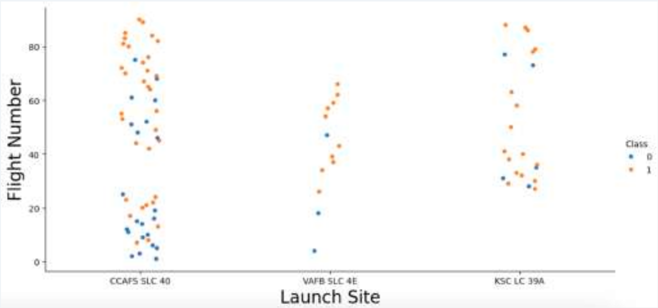
EDA with SQL

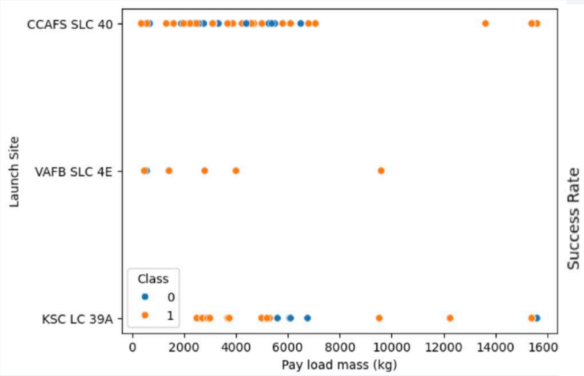
- SQL queries performed include:
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string 'KSC'
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 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 booster_versions 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 of successful landing_outcomes between the date 2010 06 04 and 2017 03 20 in descending order
- GitHub URL:

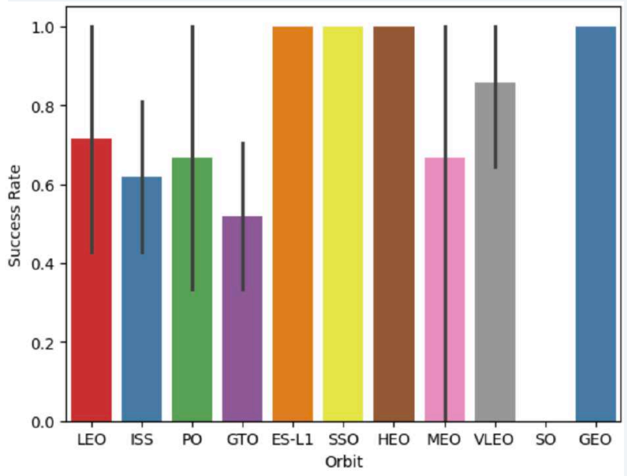
  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%204-ExploratoryDataAnalysis_SQL.ipynb

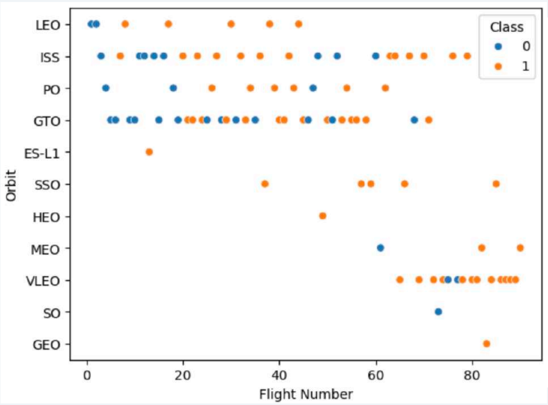
EDA with Data Visualization

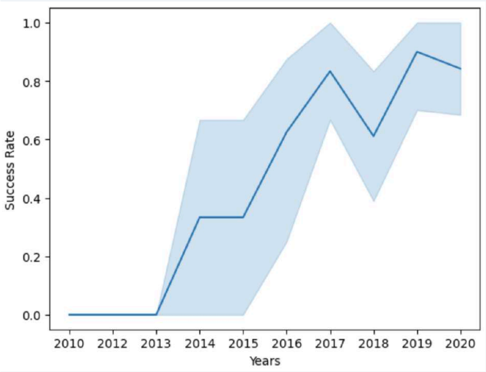








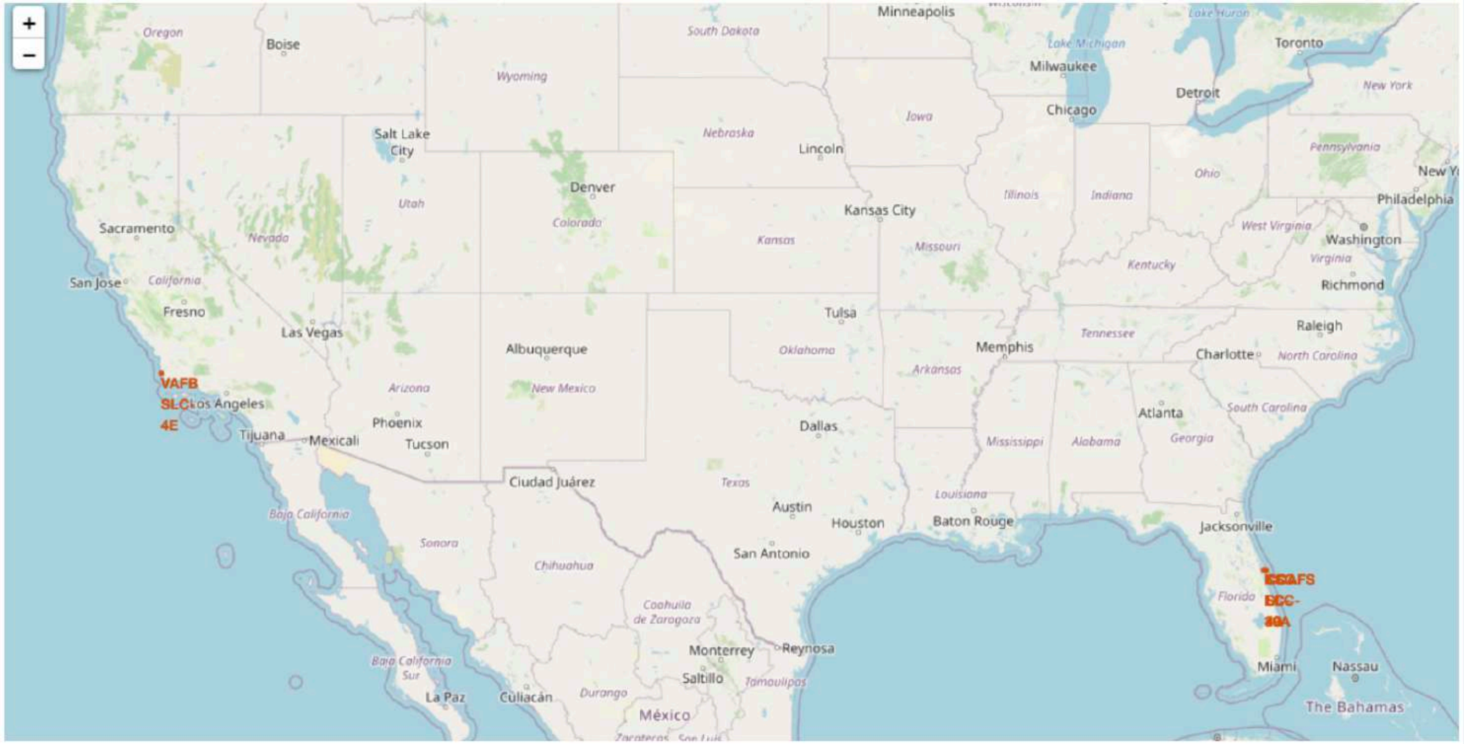




- GitHub URL:

  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%205-ExploratoryDataAnalysis_Visualization.ipynb

Build an Interactive Map with Folium



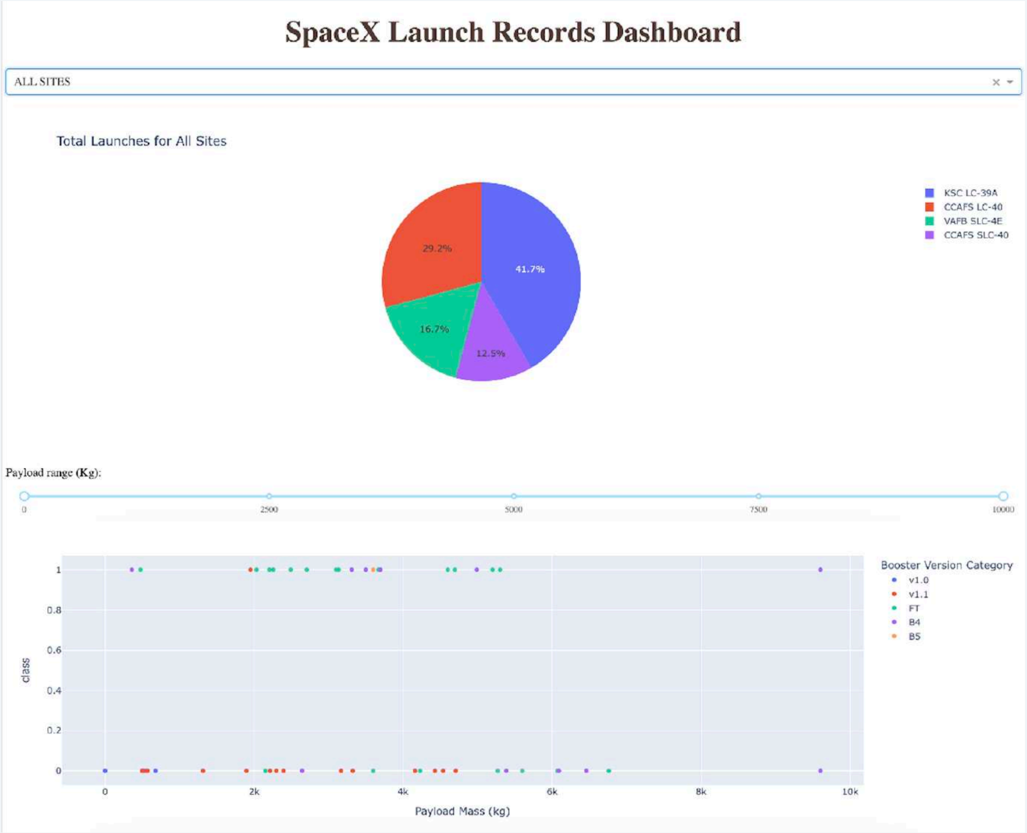
- Map markers have been added to the map with aim to find an optimal location for building a launch site
- GitHub URL:

  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%206-InteractiveVisualAnalytics_Folium.ipynb

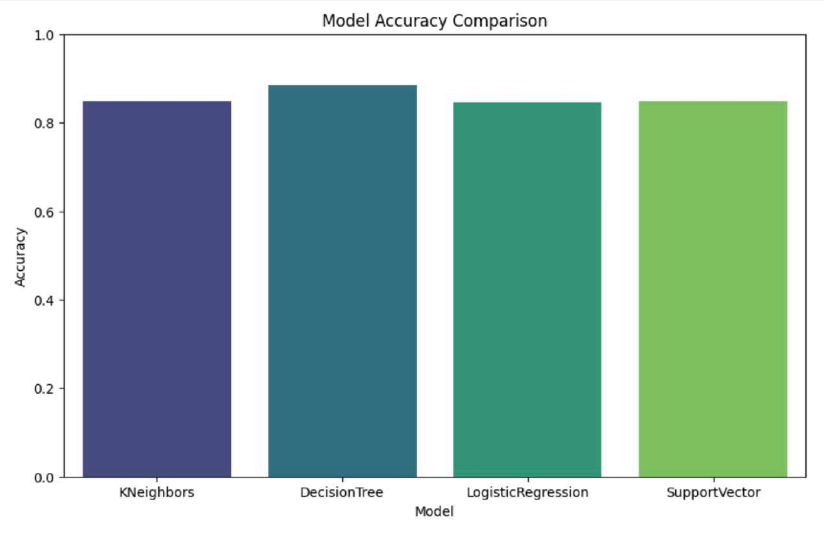
Build a Dashboard with Plotly Dash

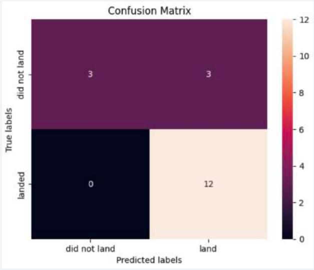
- Plots, such as scatter plots, and piechart, provide a visual representation of the data which helps in understanding the distribution of data, identifying trends, and spotting outliers
- GitHub URL:

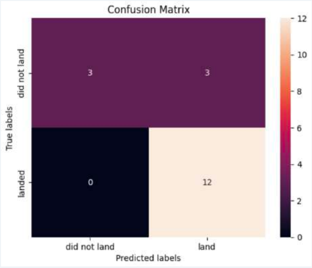
  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%207-InteractiveVisualAnalytics_PlotyDash.py

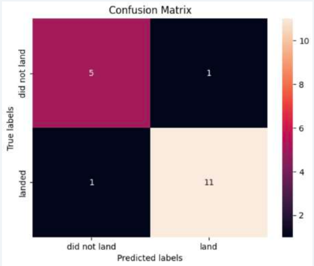


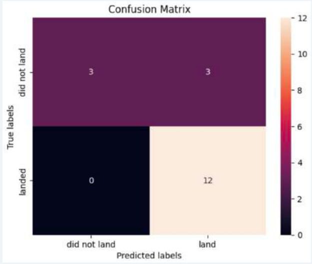
Predictive Analysis (Classification)












```
Best model is DecisionTree with a score of 0.8857142857142858
Best params is : {'criterion': 'entropy', 'max_depth': 14, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samp
les_split': 10, 'splitter': 'best'}
```

- The LR, SVM, KNN model achieved the accuracy at 83.3%, while the Decision Tree model performs the best at 0.88 score
- GitHub URL:

  https://github.com/Samritha03/IBM-Data-Science-Course/blob/main/SpaceX%208-MachineLearning_Prediction.ipynb

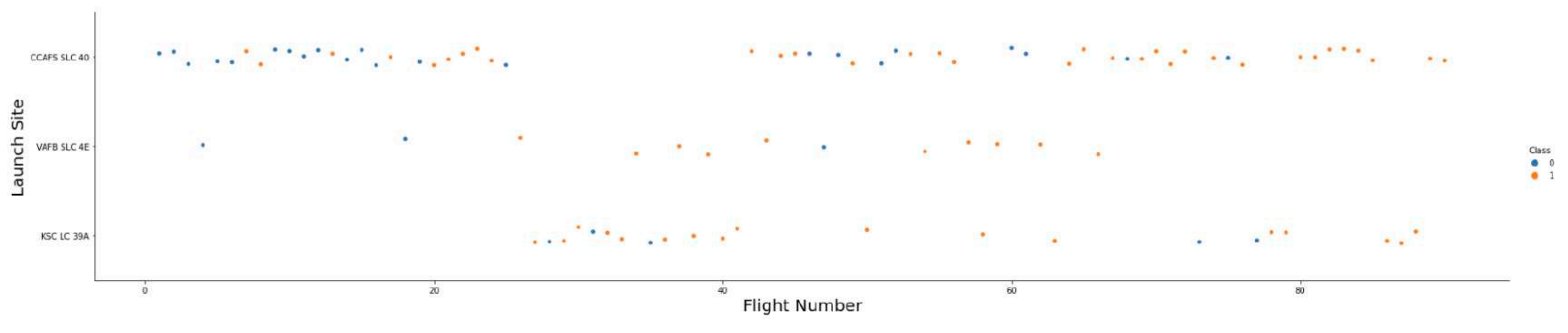
Results

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform 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

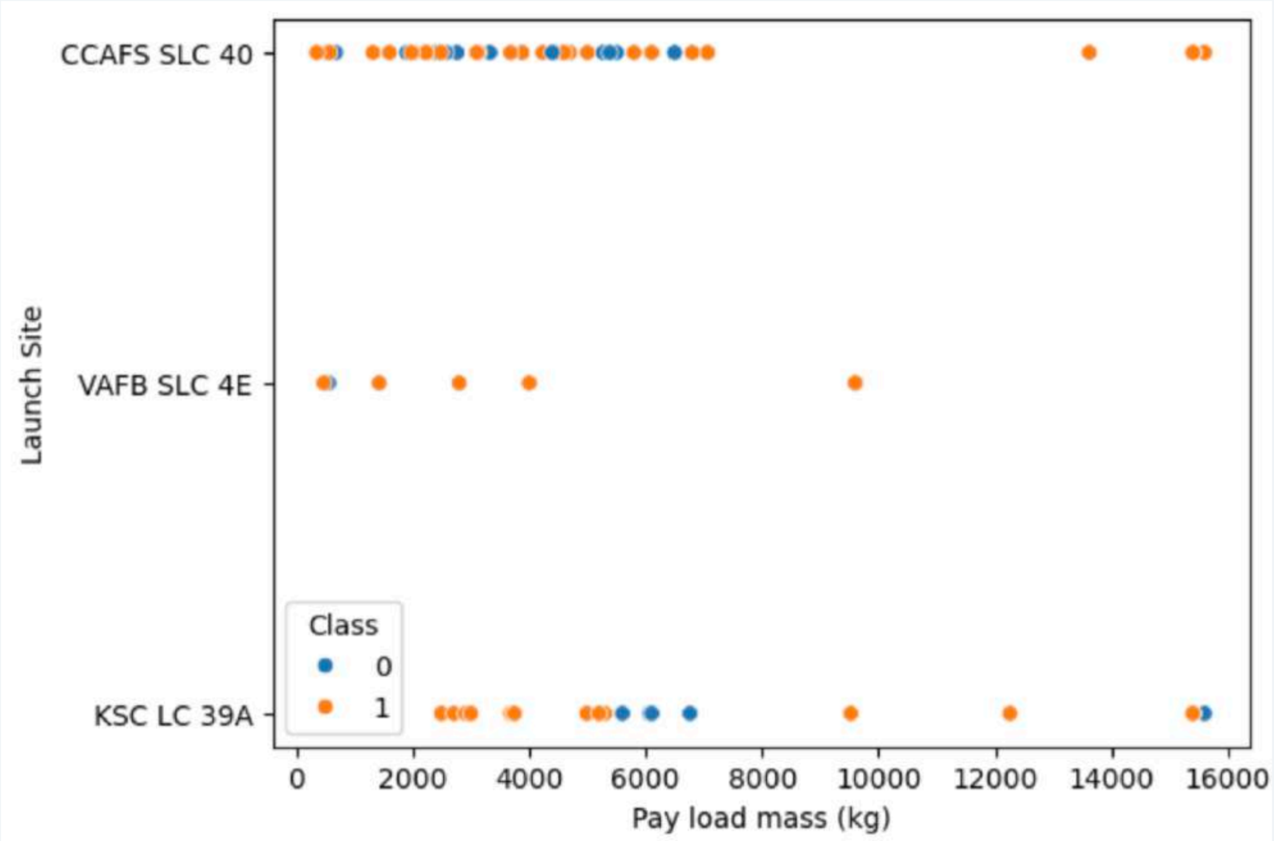
Flight Number vs. Launch Site



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

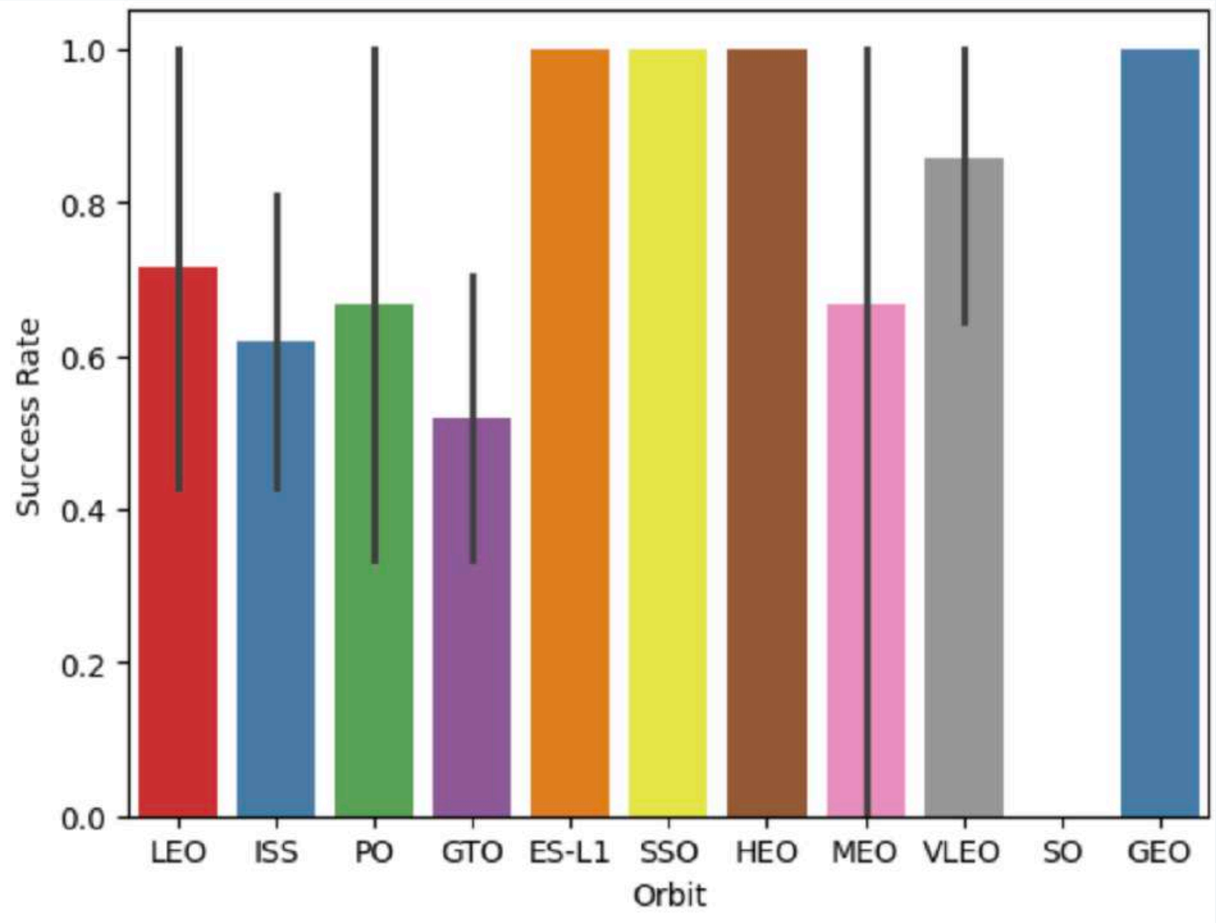
Payload vs. Launch Site

- The majority of Pay Loads with lower Mass have been launched from CCAFS SLC 40



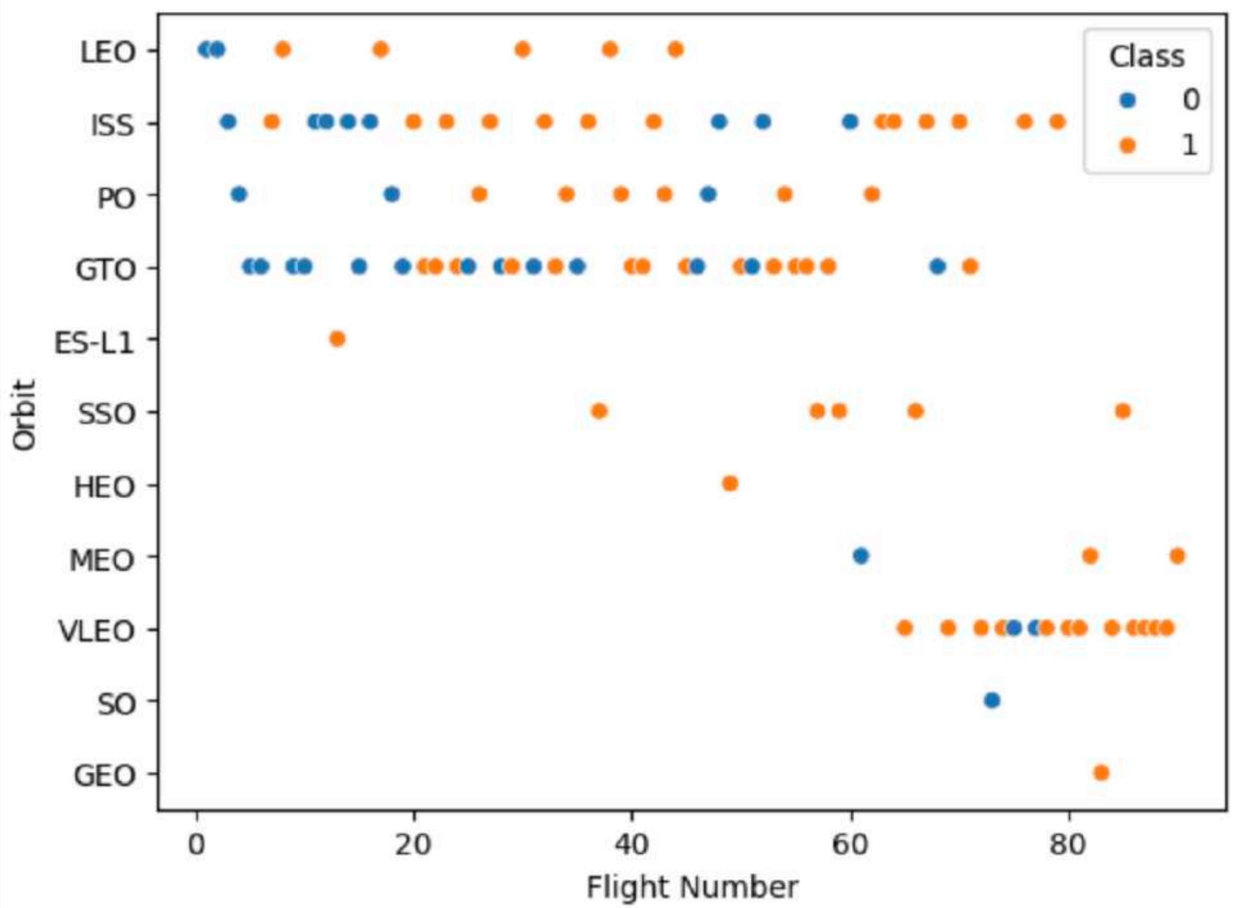
Success Rate vs. Orbit Type

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

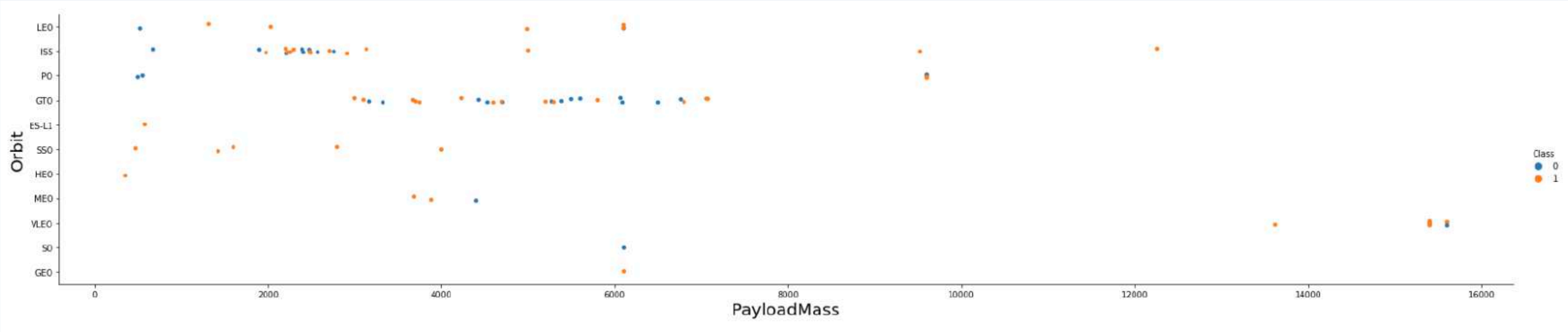


Flight Number vs. Orbit Type

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



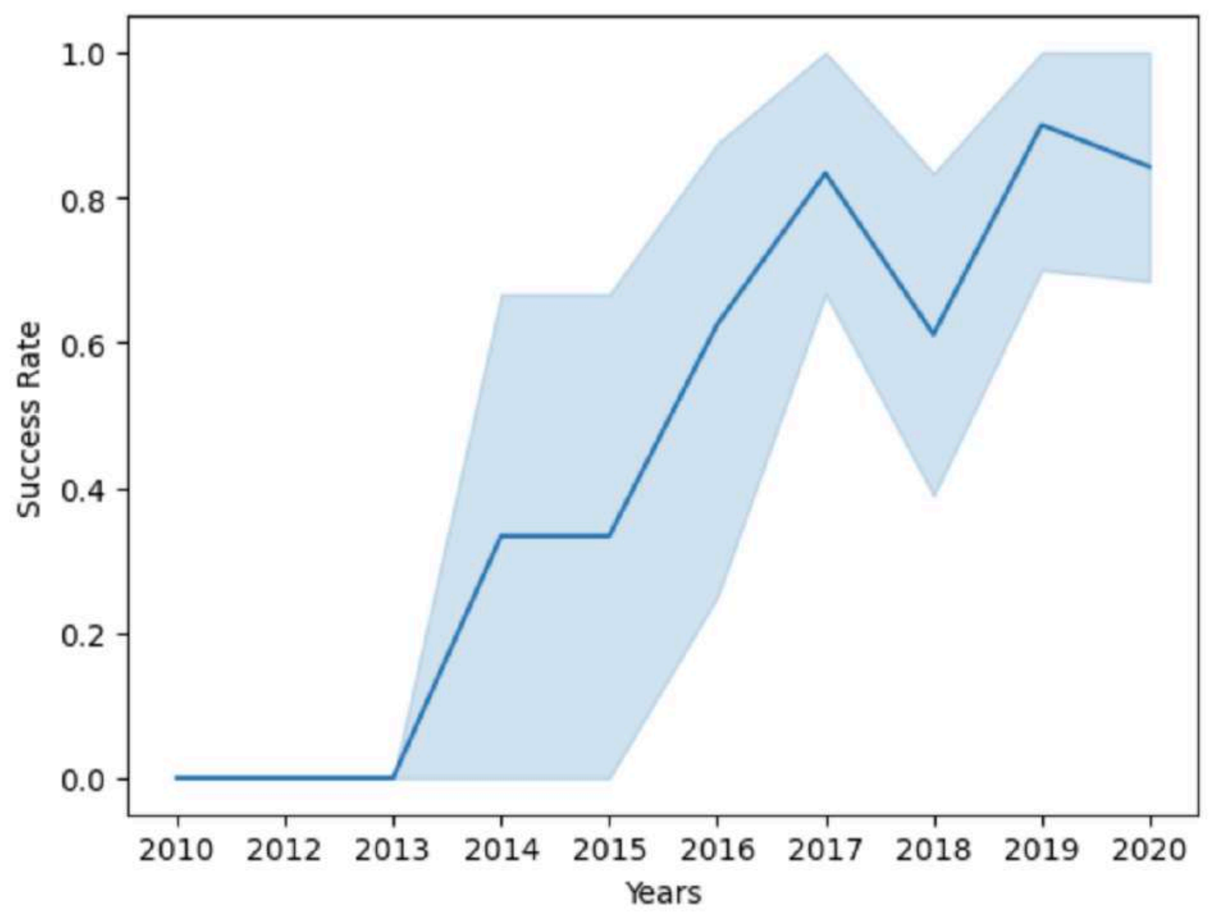
Payload vs. Orbit Type



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

Launch Success Yearly Trend

- Launch success rate has increased significantly since 2013 and has stabilised 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
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

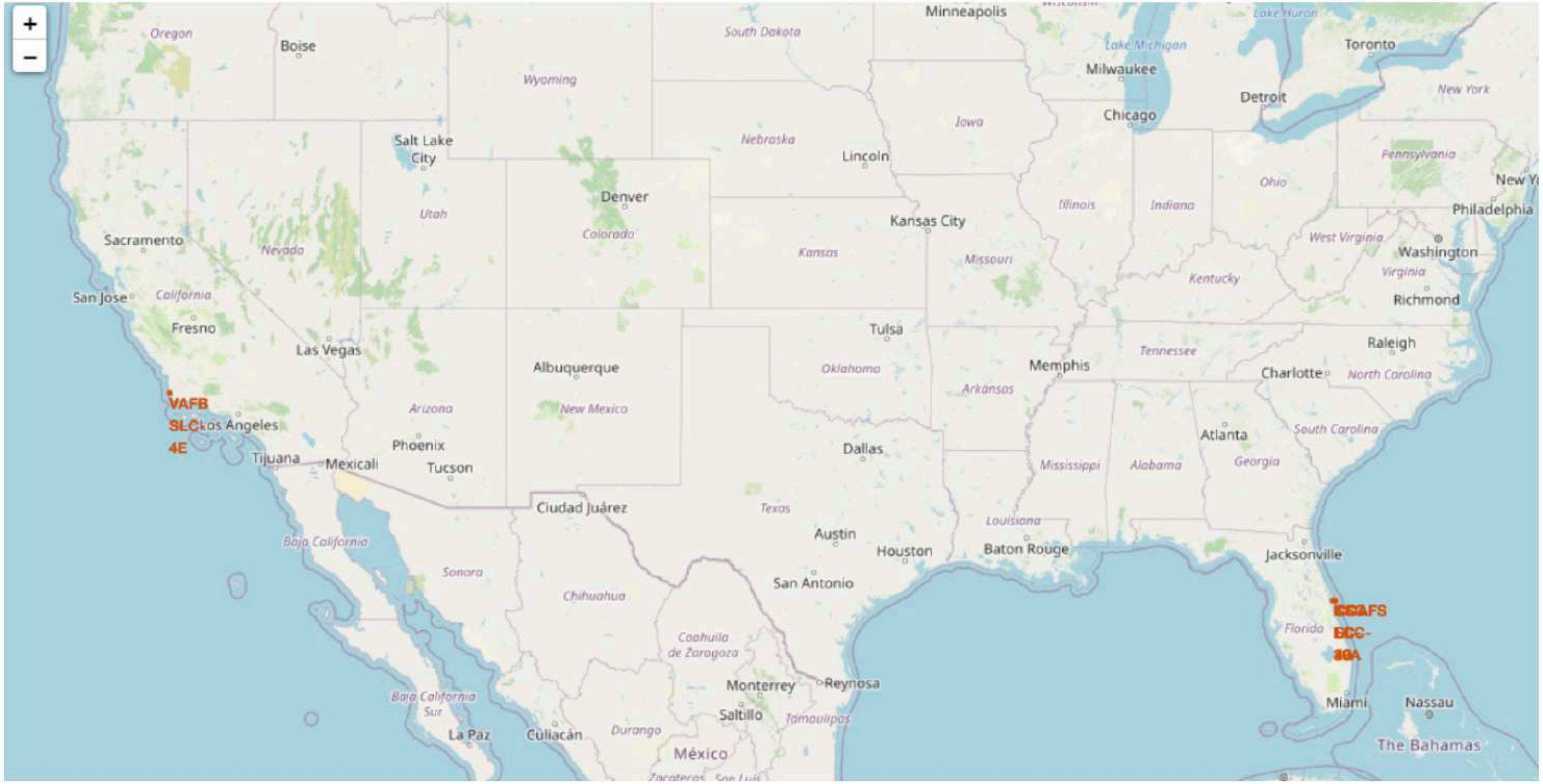
- ```
  %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
  ```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
2010-12-08	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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N
  ```

Total Payload Mass

- **%sql** `SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER='NASA (CRS)'`

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- **%sql** `SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION='F9 v1.1'`

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- `%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

- **%sql** SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ between 4000 and 6000 AND LANDING_OUTCOME='Success (drone ship)’

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

Total Number of Successful and Failure Mission Outcomes

- **%sql** SELECT COUNT(*) FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Success%' OR MISSION_OUTCOME LIKE '%Failure%'

COUNT(*)
101

Boosters Carried Maximum Payload

- **%sql** SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)

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

- `%sql SELECT DATE, LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE \`
  `FROM SPACEXTBL WHERE LANDING_OUTCOME LIKE 'Failure (drone ship)' AND DATE LIKE '%2015%'`

Date	Landing_Outcome	Booster_Version	Launch_Site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- ```
  %sql SELECT "DATE", COUNT(LANDING_OUTCOME) as COUNT FROM SPACEXTBL \
  WHERE "DATE" BETWEEN '2010-06-04' and '2017-03-20' AND
  LANDING_OUTCOME LIKE '%Success%' \
  GROUP BY "DATE" \
  ORDER BY COUNT(LANDING_OUTCOME) DESC
  ```

Date	COUNT
2017-02-19	1
2017-01-14	1
2016-08-14	1
2016-07-18	1
2016-05-27	1
2016-05-06	1
2016-04-08	1
2015-12-22	1
  ```

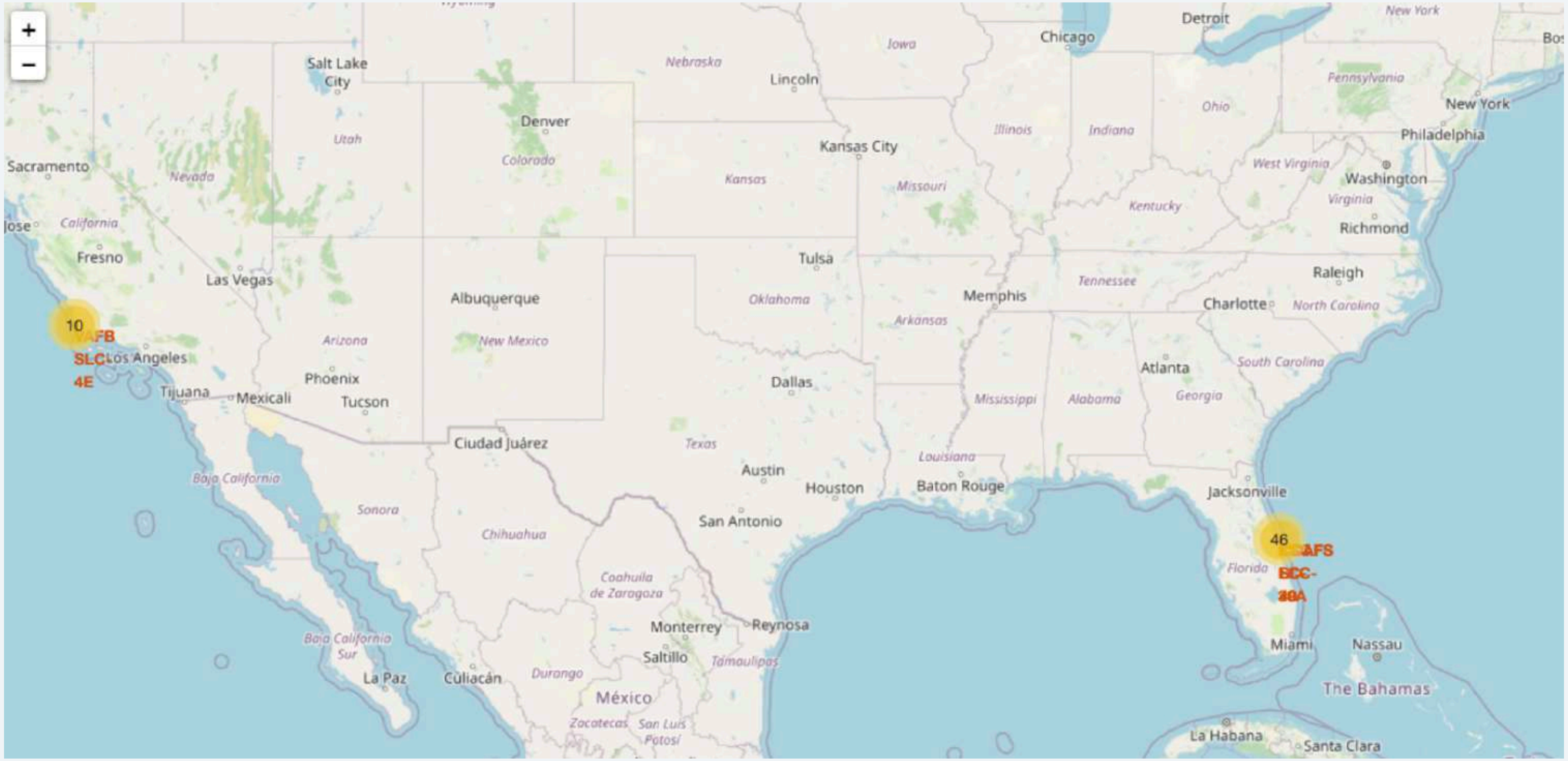
Section 3

Launch Sites Proximities Analysis

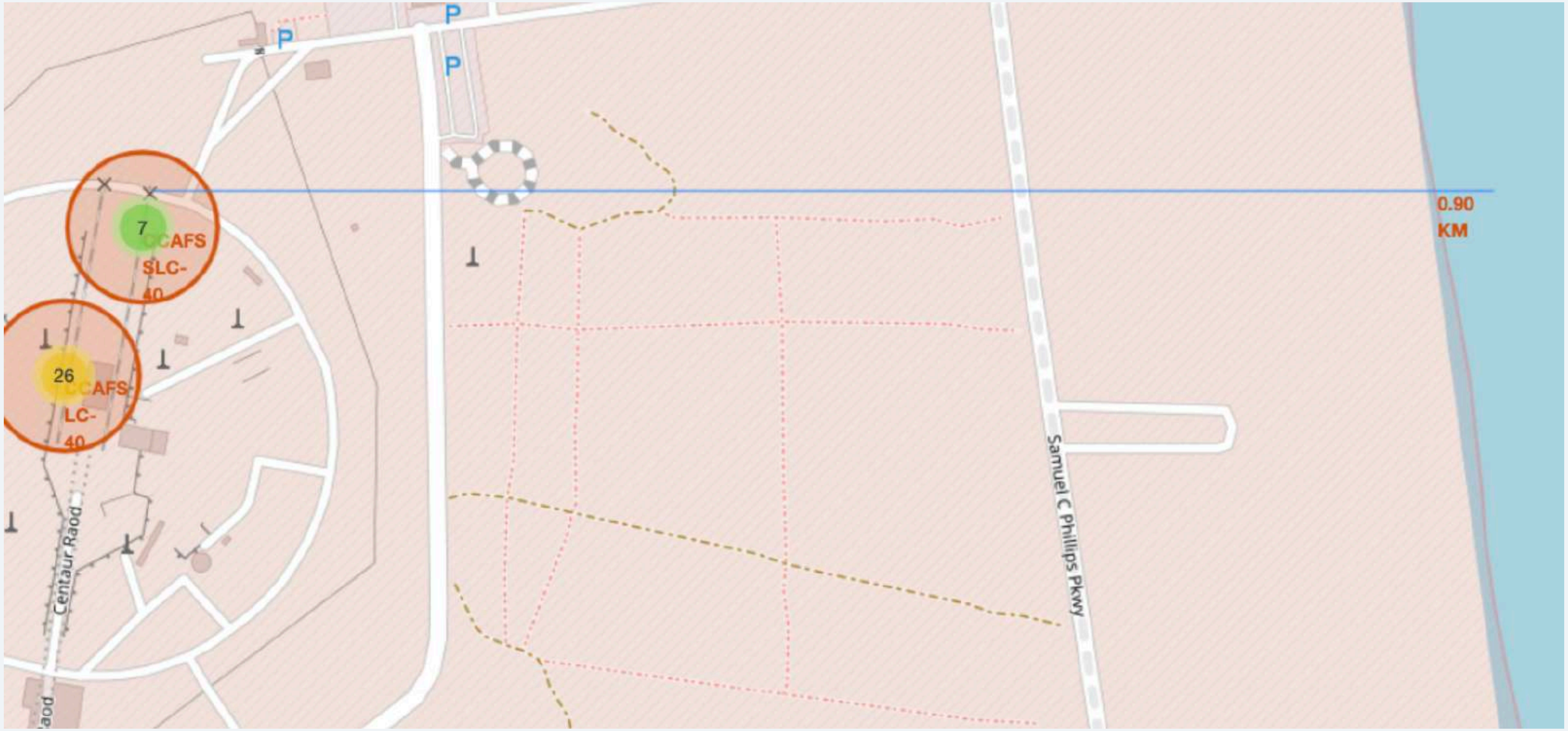
Mark all launch sites on a map



Mark the success/failed launches for each site on the map



Calculate the distances between a launch site to its proximities

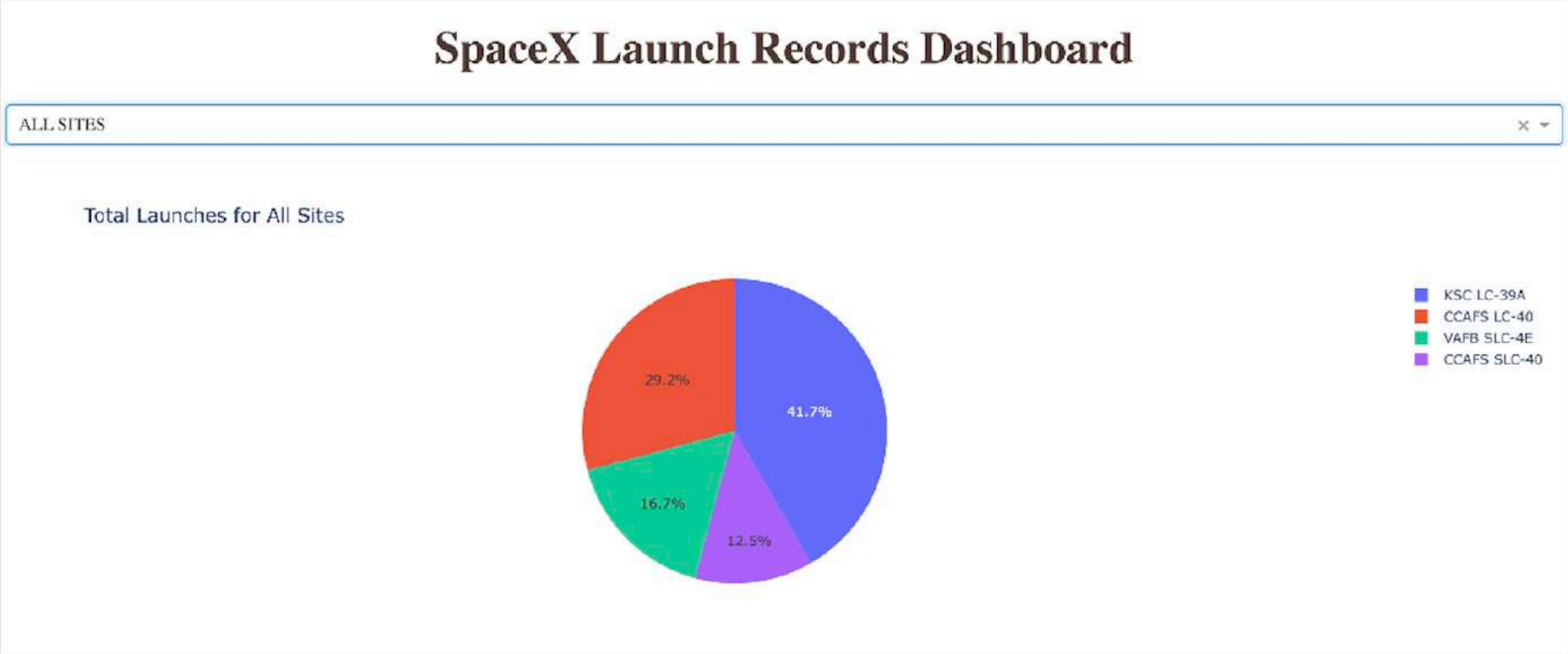


Section 4

Build a Dashboard with Plotly Dash

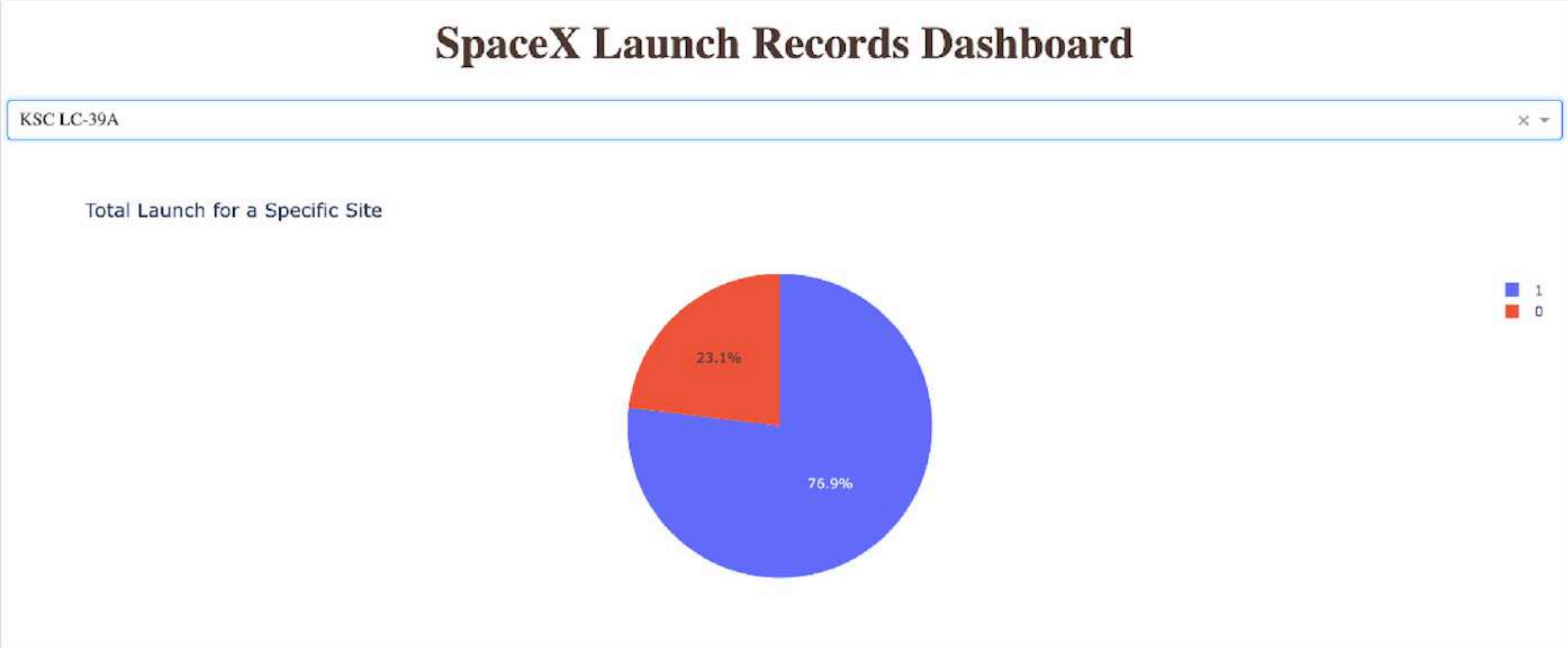
Total success launches by all sites

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



Success rate by site

- KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate



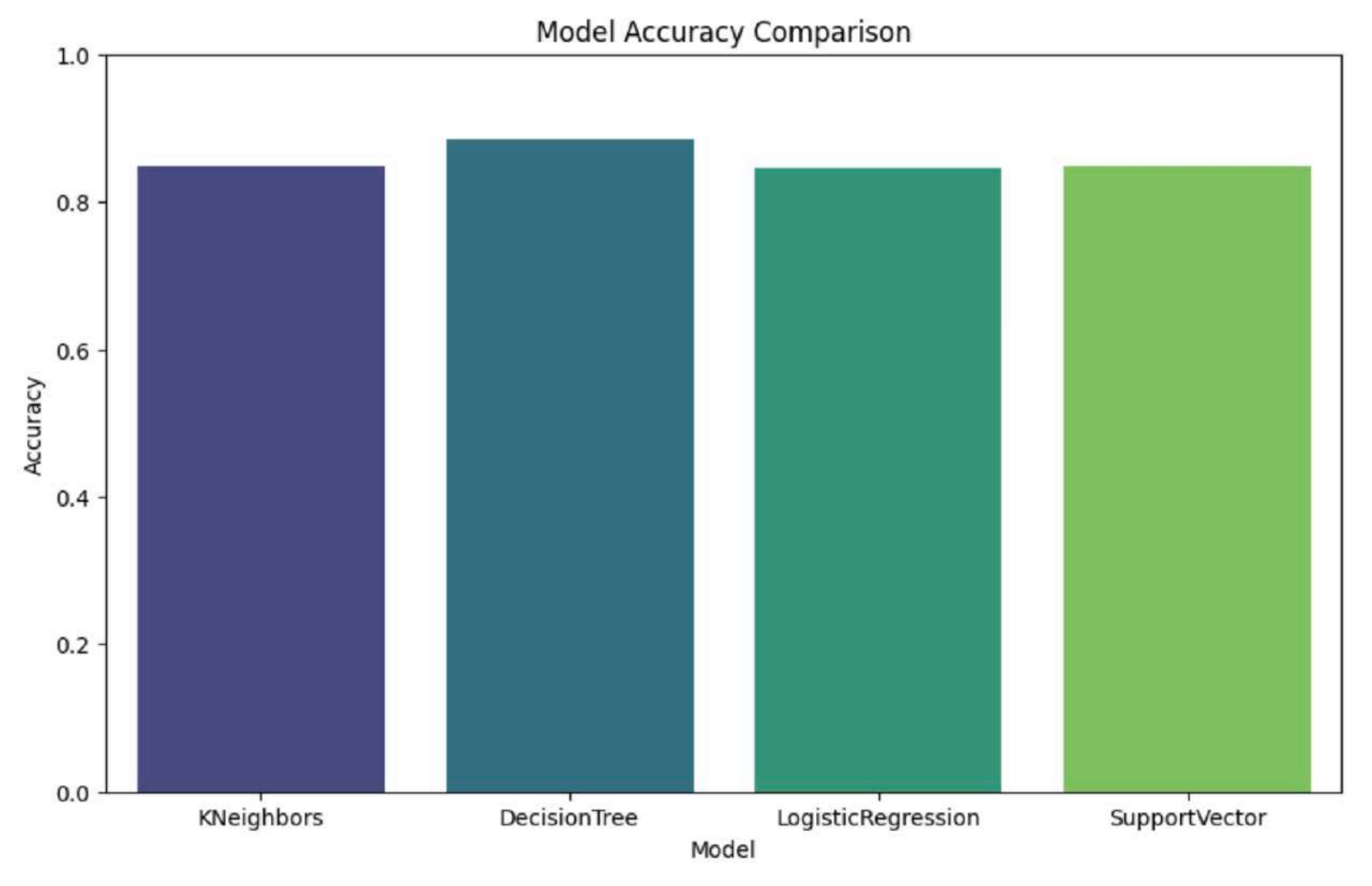
Payload vs Launch outcome

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

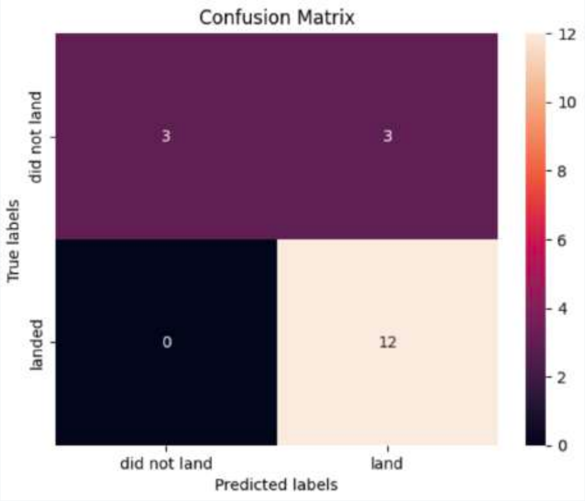
Section 5

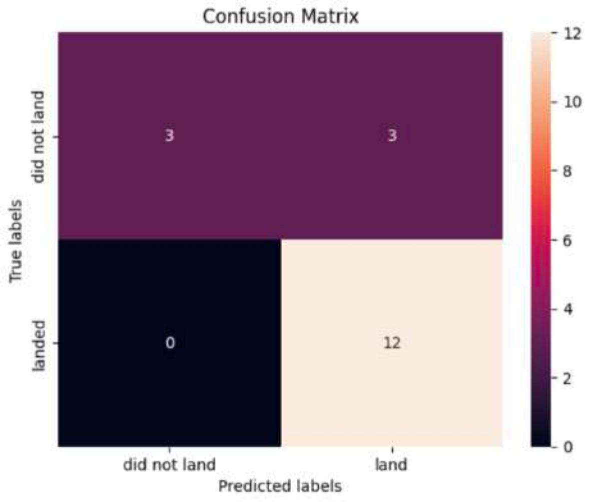
Predictive Analysis (Classification)

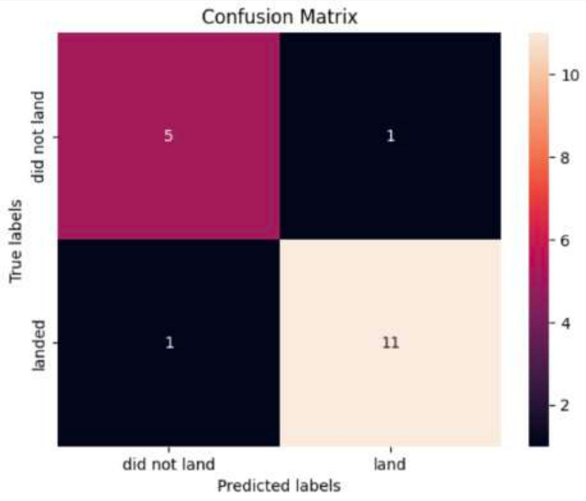
Classification Accuracy

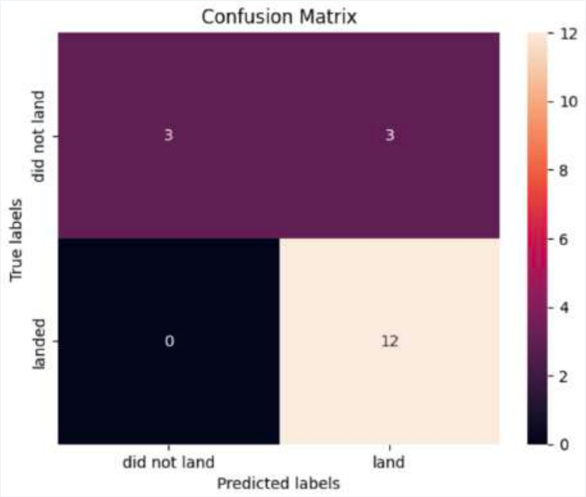


Confusion Matrix









Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform 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

Appendix

- GitHub URL:

  https://github.com/Samritha03/IBM-Data-Science-Course/tree/main

Thank you!