Data Science Capstone-SpaceX

REDHWAN MOHAMMED

OUTLINE

- INTRODUCTION
- SUMMARY
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX

INTRODUCTION

Project Background:
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.

Problem:
Predicting if the Falcon 9 first stage will land successfully.

EXECUTIVE SUMMARY

- **Methodologies Summary**

  The research attempts to identify the factors for a successful rocket landing. To make this determination, the following methodologies where used:

  - Collect data using SpaceX REST API and web scraping techniques
  - Wrangle data to create success/fail outcome variable
  - Explore data with data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend
  - Analyze the data with SQL, calculating the following statistics: total payload, payload range for successful launches, and total # of successful and failed outcomes
  - Explore launch site success rates and proximity to geographical markers
  - Visualize the launch sites with the most success and successful payload ranges
  - Build Models to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

EXECUTIVE SUMMARY

- Results Summary

Exploratory Data Analysis:

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate

Visualization/ Analytics:

- Most launch sites are near the equator, and all are close to the coast

Predictive Analytics

- All models performed similarly on the test set. The decision tree model

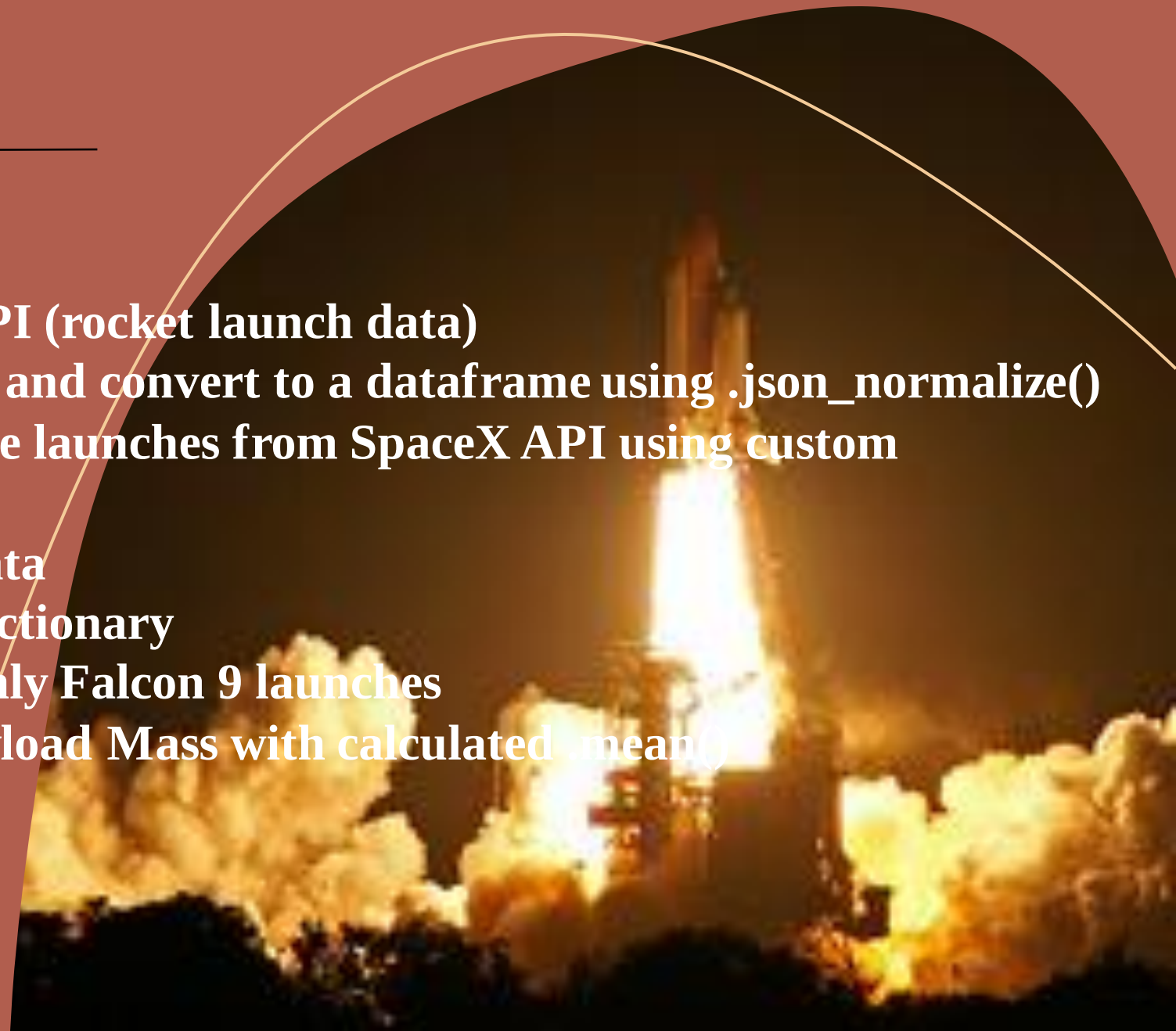
METHODOLOGY

- Data Collection
- Data Wrangling
- EDA Data visualization.
- EDA Sql data
- Machine Learning Prediction
- Interactive Visual Analytics with Folium
- Build a Dashboard Using Plotly Dash

METHODOLOGY

- Data Collection

- Request data from SpaceX API (rocket launch data)
- Decode response using .json() and convert to a dataframe using .json_normalize()
- Request information about the launches from SpaceX API using custom functions
- Create dictionary from the data
- Create dataframe from the dictionary
- Filter dataframe to contain only Falcon 9 launches
- Replace missing values of Payload Mass with calculated mean
- Export data to csv file



METHODOLOGY

- **Data Collection**

  SpaceX API:

  Data collection with SpaceX REST API

- **Link:** https://github.com/REDHWAN0/IBM-Data-Science-Capstone-SpaceX/Data-collection-api.ipynb

METHODOLOGY

- **Data Collectionl**

  Scraping:

  Wikipedia has a page that has tables of data about SpaceX launches

- **Link:** https://github.com/REDHWAN0/IBM-Data-Science-Capstone-SpaceX/Web Scraping.ipynb

METHODOLOGY

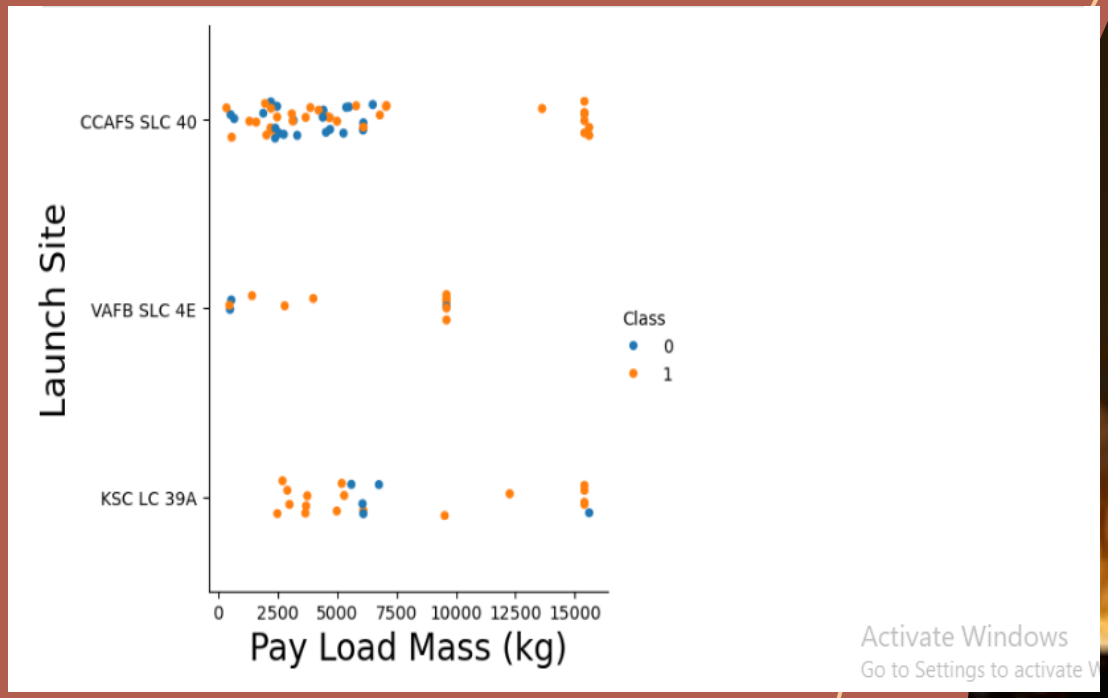
- Data Wrangling
  - Convert outcomes into 1 for a successful landing and 0 for an unsuccessful landing
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification

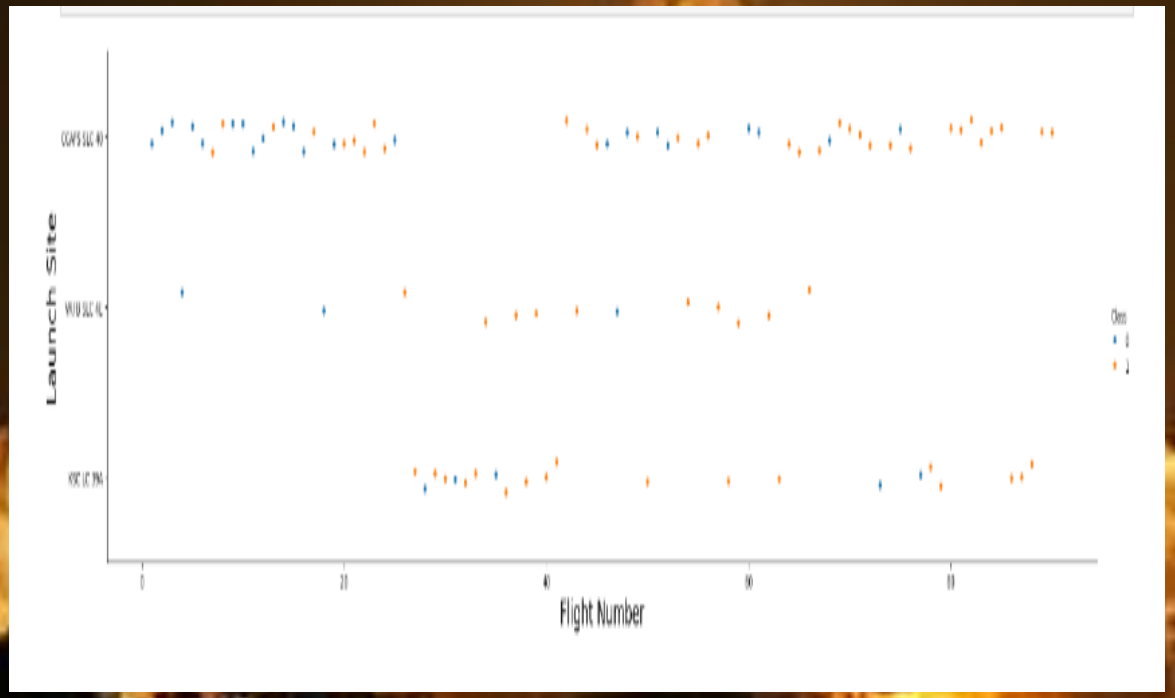
- Link: https://github.com/REDHWAN0/IBM-Data-Science-Capstone-SpaceX/Data wrangling.ipynb

METHODOLOGY

- EDA with Data Visualization

The following charts were created to look at Launch Site trends

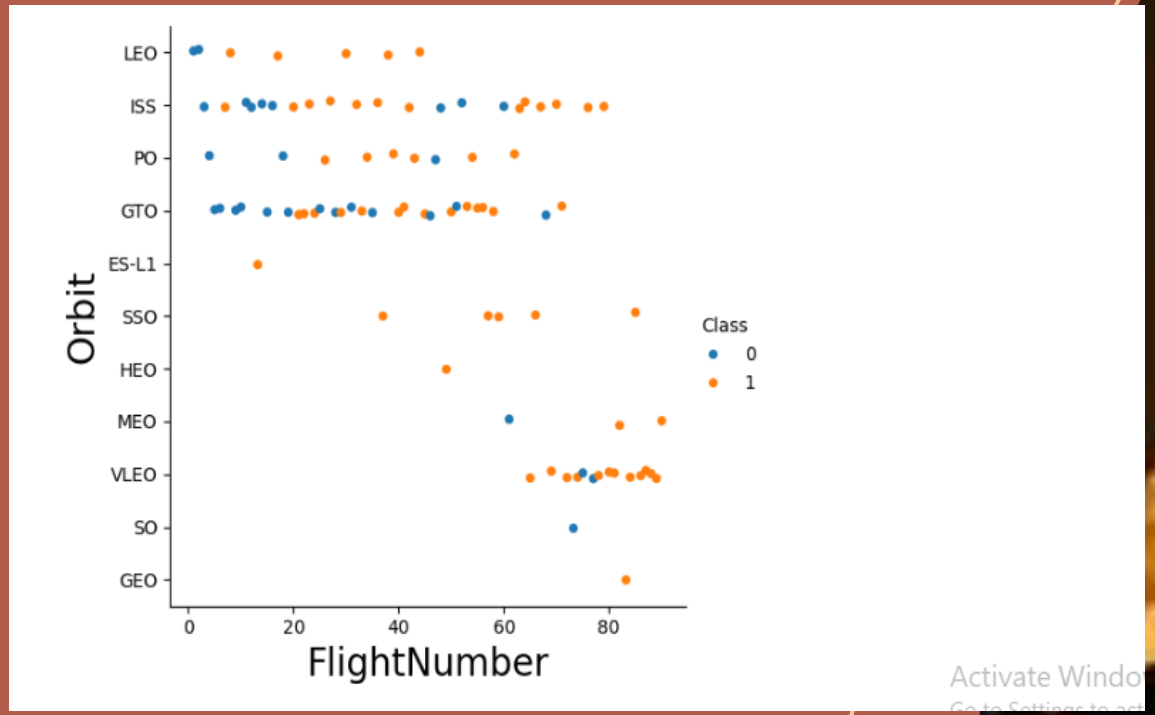


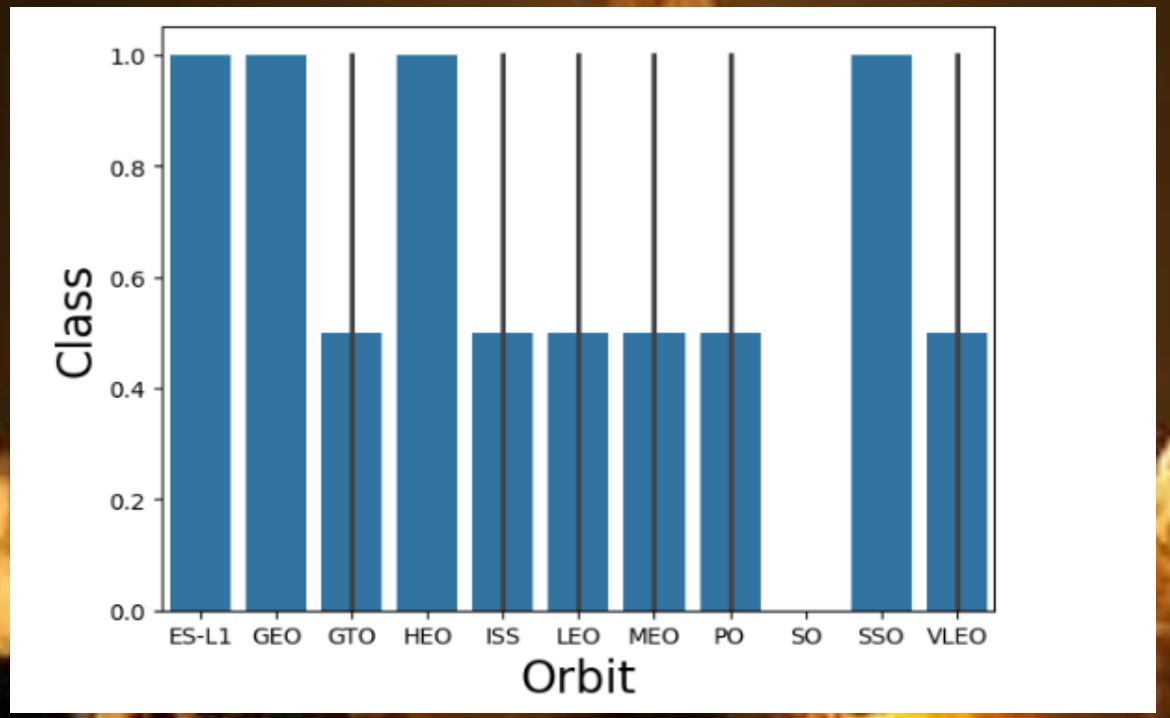


METHODOLOGY

- EDA with Data Visualization

  The following charts were created to look at Orbit Type trends

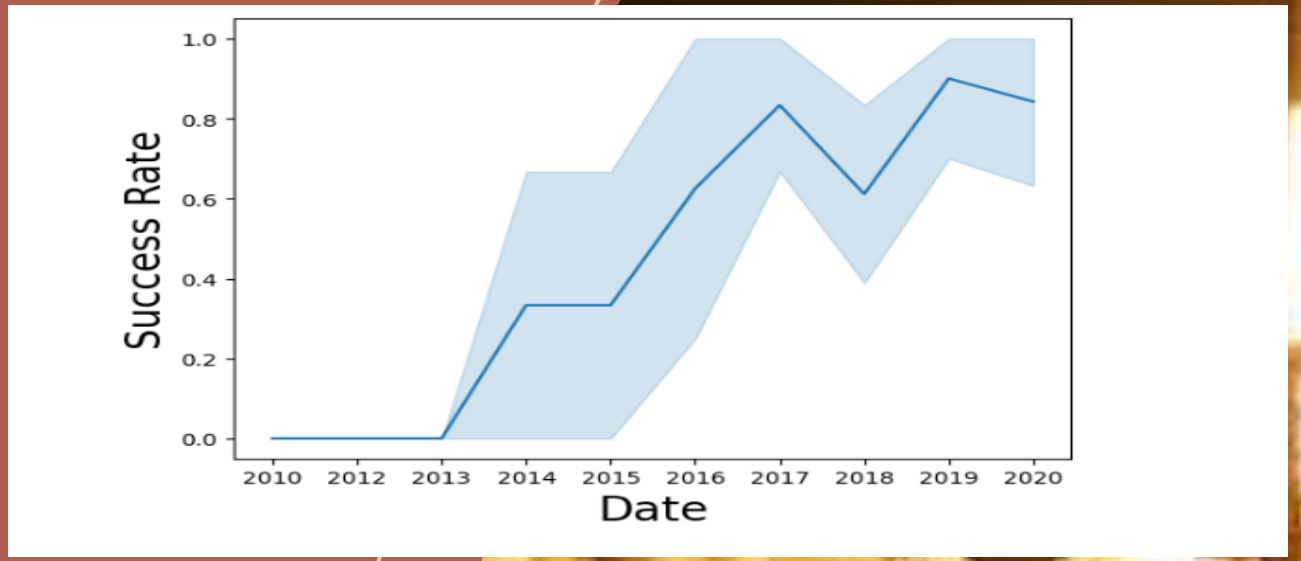




METHODOLOGY

- **EDA with Data Visualization**

  The following chart was created to look at trends based on time



- **GitHub URL (EDA with Data Visualization): :** https://github.com/REDHWAN0/IBM-Data-Science-Capstone-SpaceX/EDA-Data visualization.ipynb

METHODOLOGY

- **EDA Sql data**

Queries were written to extract information about:

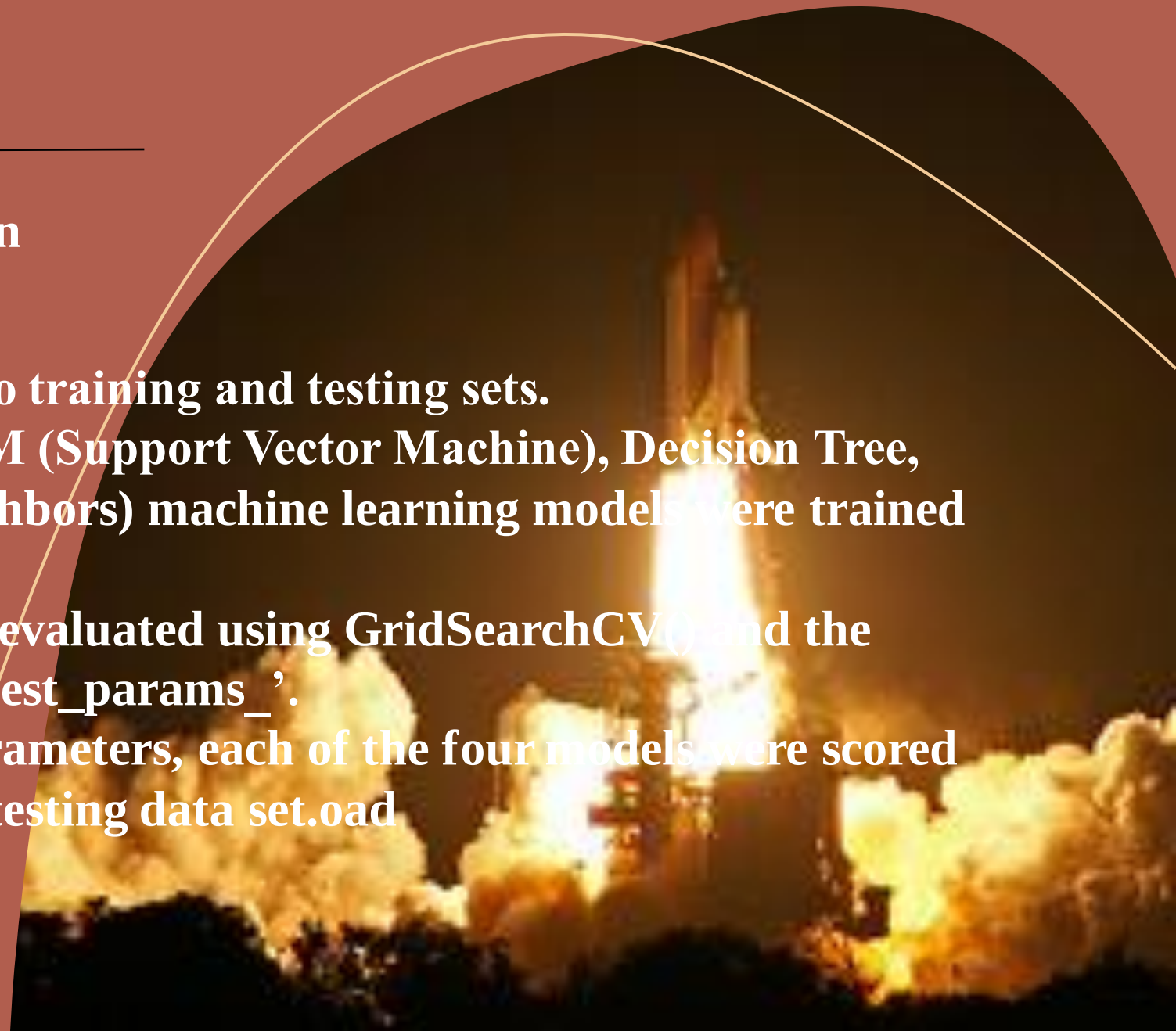
- **Launch sites**
- **Payload masses**
- **Dates**
- **Booster types**
- **Mission outcomes**

- **GitHub URL (EDA Sql data) :** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX%20/EDA-Sql%20data.ipynb

METHODOLOGY

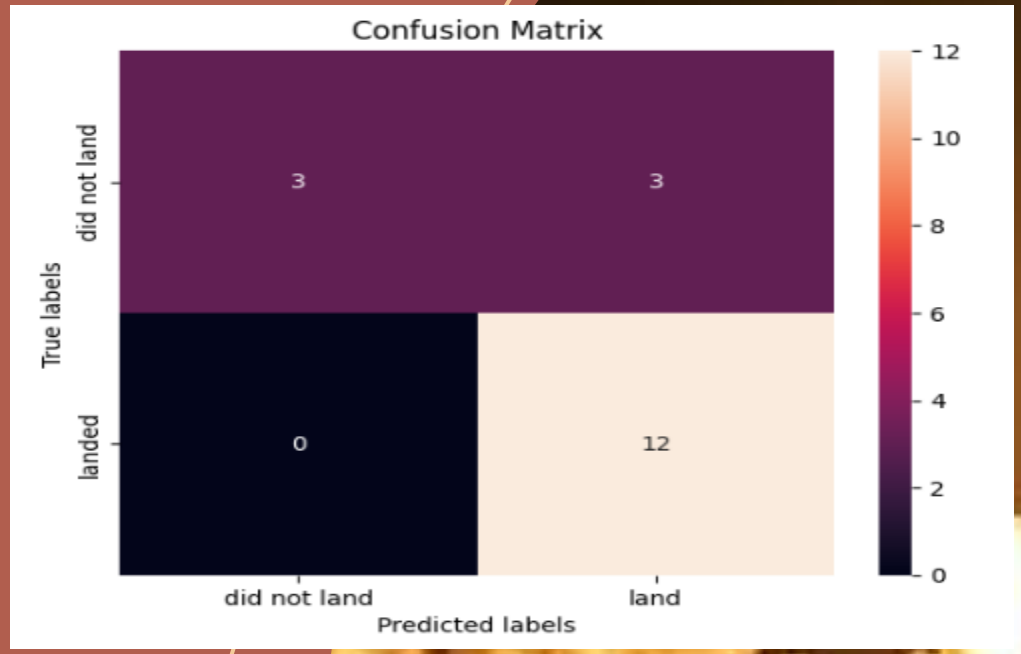
- **Machine Learning Prediction**

  - **The dataset was split into training and testing sets.**
  - **Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) machine learning models were trained on the training data set.**
  - **Hyper-parameters were evaluated using GridSearchCV() and the best was selected using '.best_params_'.**
  - **Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set.oad**



METHODOLOGY

- **Machine Learning Prediction**



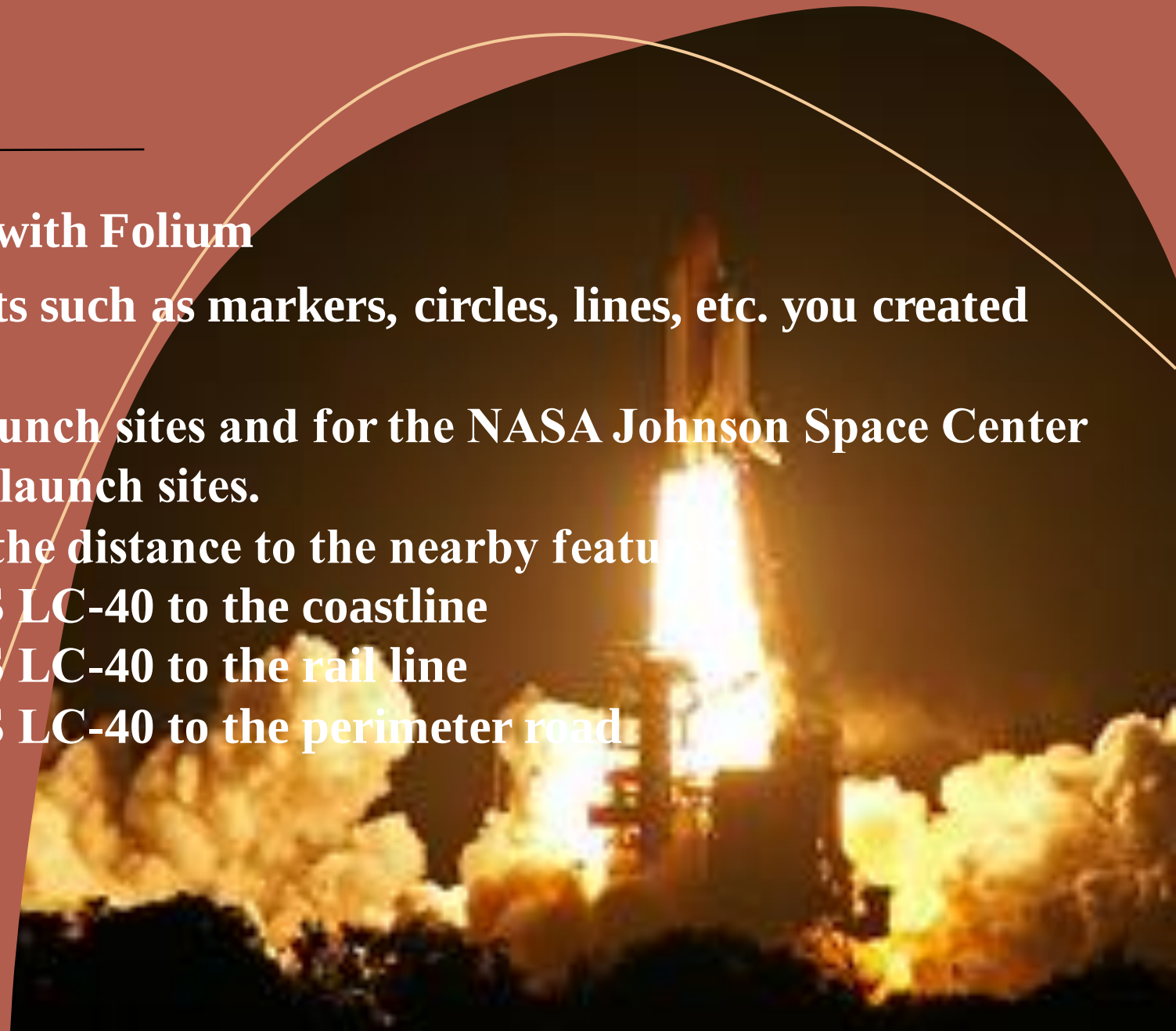
- **GitHub URL (Machine Learning Prediction) :** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX%20/Machine_Learning_Prediction.ipynb

METHODOLOGY

- **Interactive Visual Analytics with Folium**

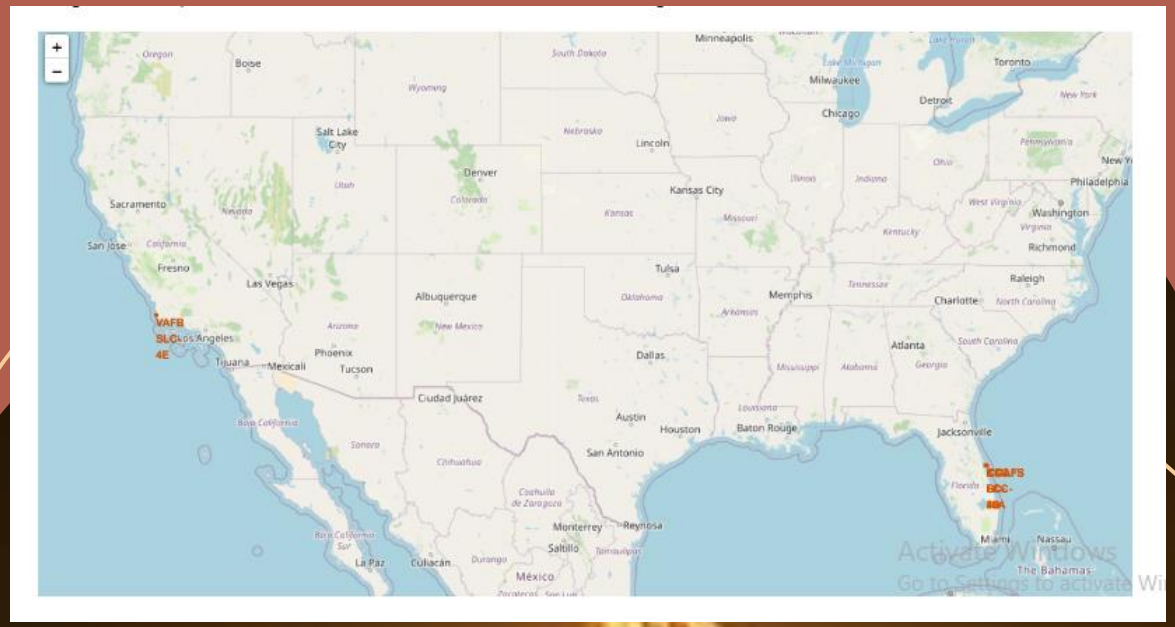
  Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

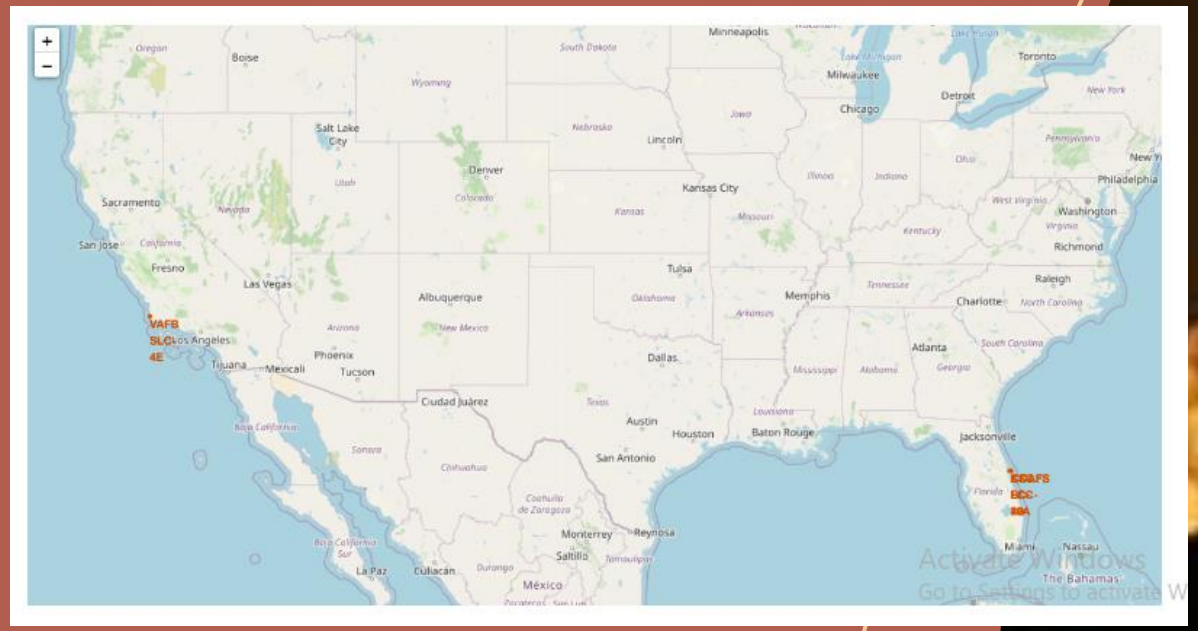
  • Markers were added for launch sites and for the NASA Johnson Space Center
  • Circles were added for the launch sites.
  • Lines were added to show the distance to the nearby featu
    • Distance from CCAFS LC-40 to the coastline
    • Distance from CCAFS LC-40 to the rail line
    • Distance from CCAFS LC-40 to the perimeter road

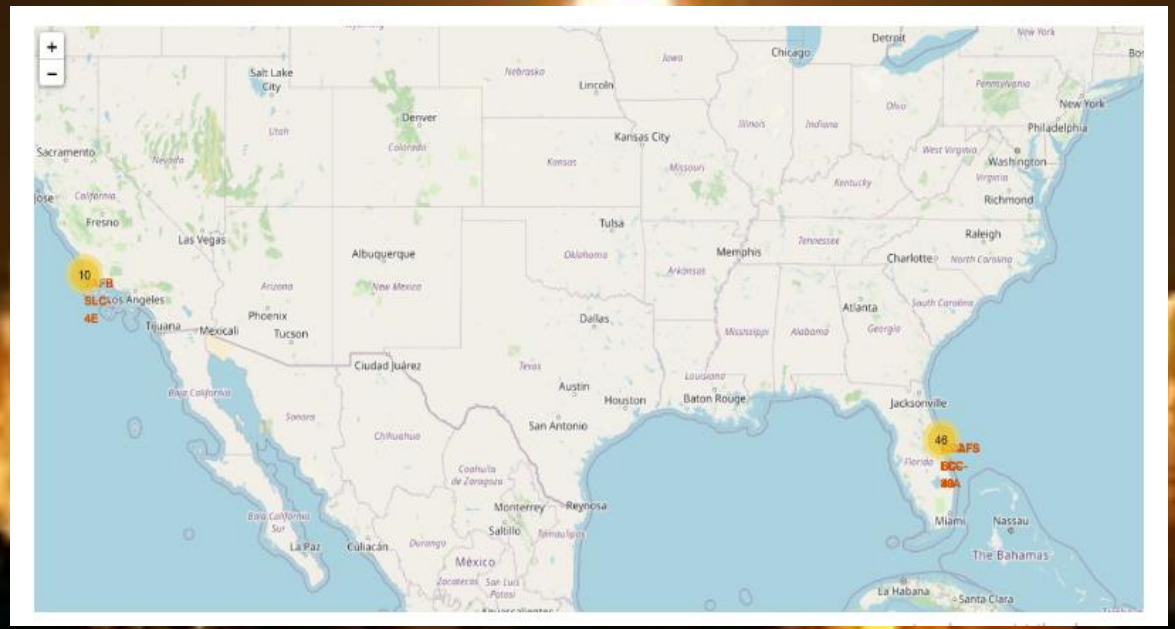


METHODOLOGY

- Interactive Visual Analytics with Folium

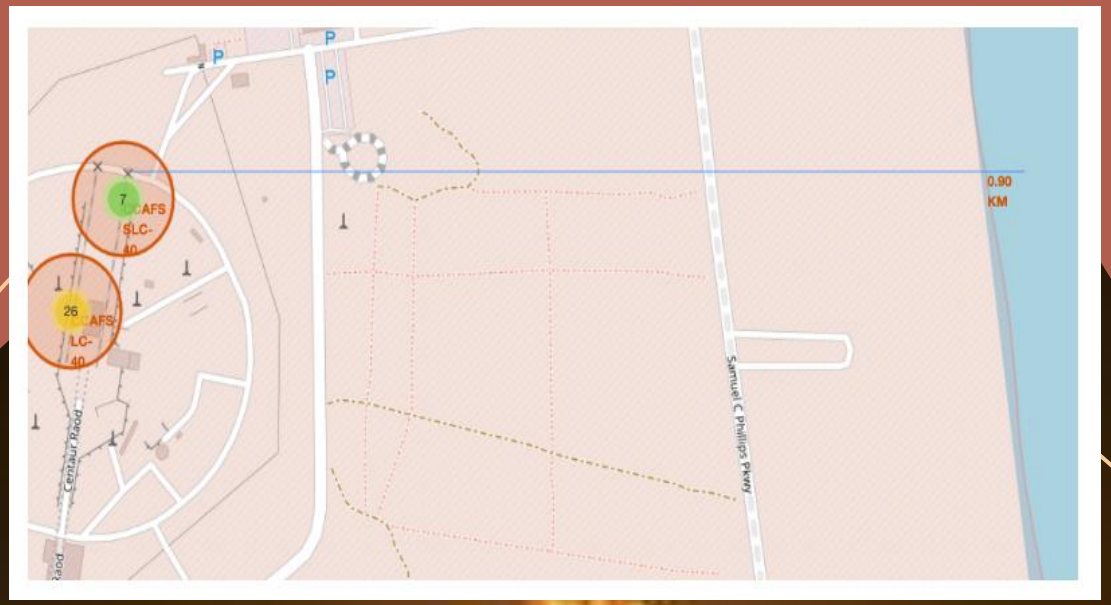


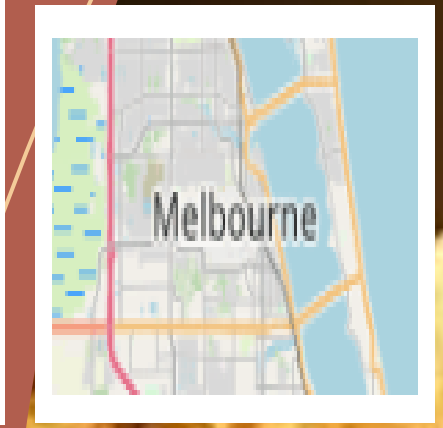


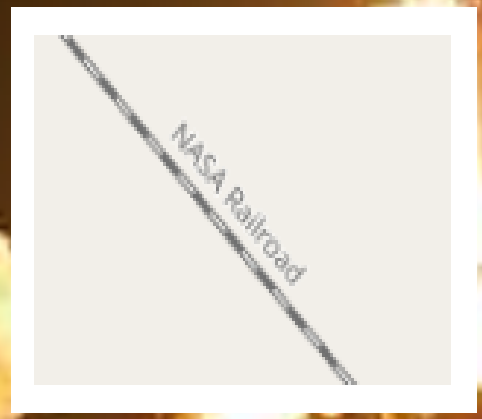


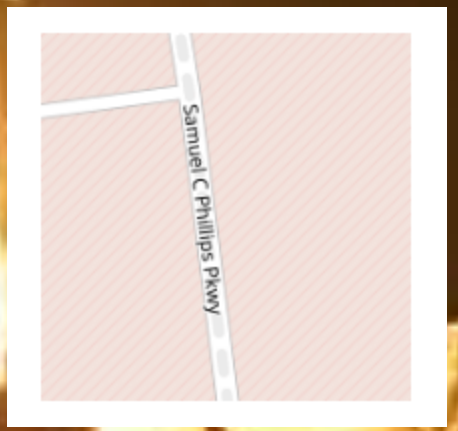
METHODOLOGY

- Interactive Visual Analytics with Folium









- **GitHub URL (Interactive Visual Analytics with Folium) :** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX%20/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

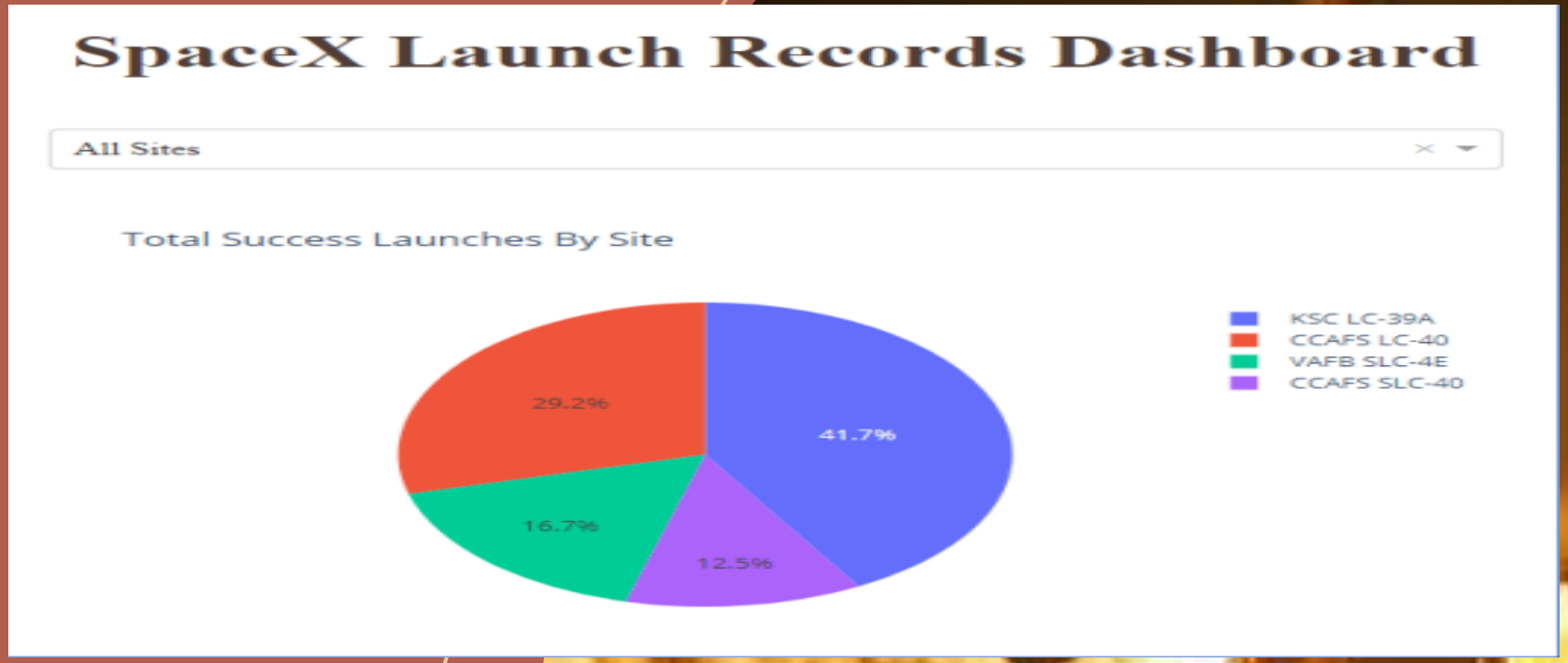
METHODOLOGY

- **Build a Dashboard Using Plotly Dash**

• The input dropdown is used to select one or all launch sites for the pie chart and scatterplot.
• The pie chart displays one of two things:
• For All Sites – the distribution of successful Falcon 9 first stage landings between the sites
• For One Site – the distribution of successful and failed Falcon 9 first stage landings for that site
• The input slider is used to filter the payload masses for the scatterplot. • The scatterplot displays the distribution of Falcon 9 first stage landings split by payload mass, mission outcome and by booster version category.

METHODOLOGY

- **Build a Dashboard Using Plotly Dash**



- **GitHub URL (Build a Dashboard Using Plotly Dash) :** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX%20/plotly_dash.py

RESULTS

Exploratory Data Analysis:

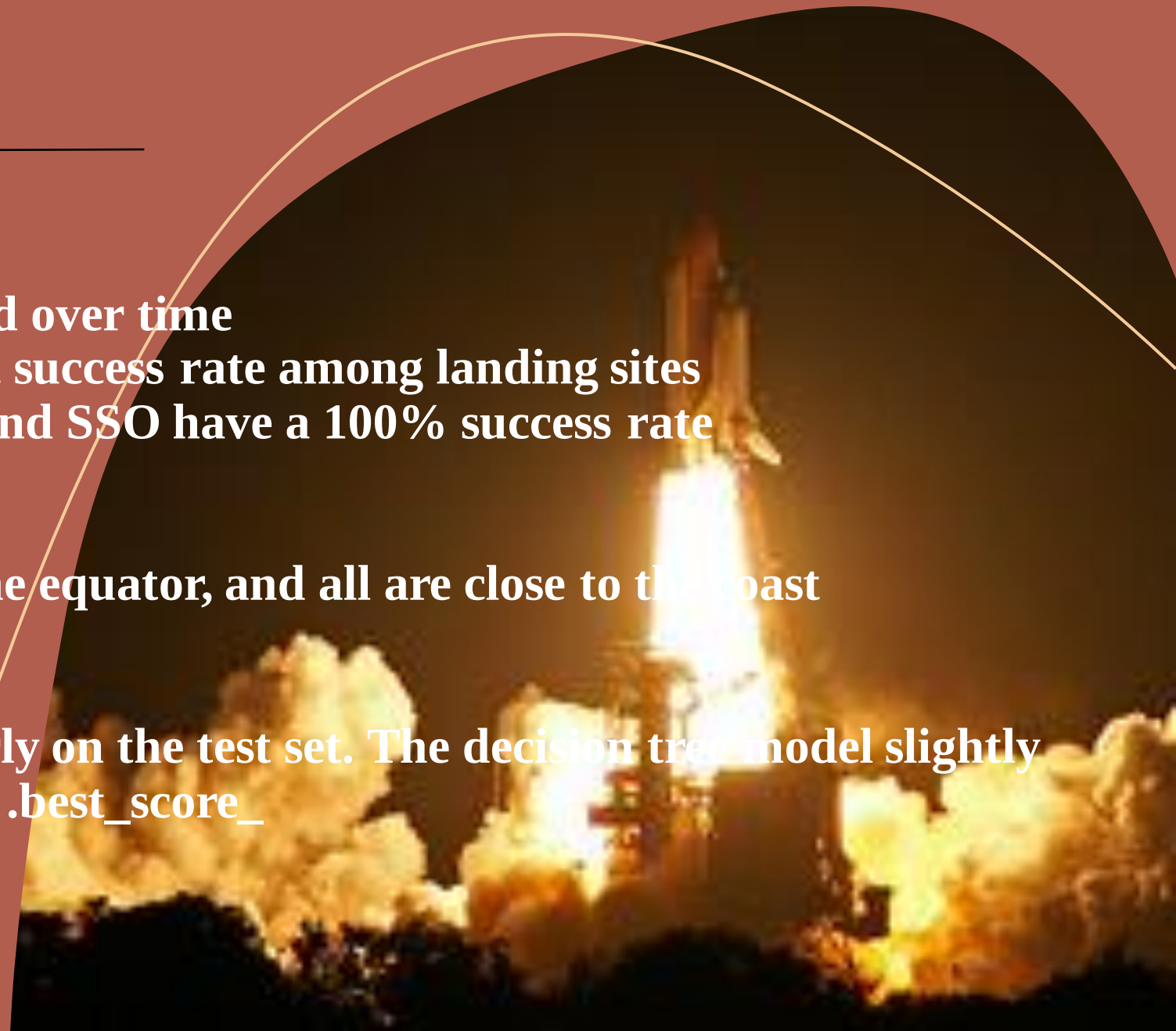
- **Launch success has improved over time**
- **KSC LC-39A has the highest success rate among landing sites**
- **Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate**

Visualization / Analytics:

- **Most launch sites are near the equator, and all are close to the coast**

Predictive Analytics

• All models performed similarly on the test set. The decision tree model slightly outperformed when looking at .best_score_



CONCLUSION

- SpaceX does not have a perfect track record of Falcon 9 first stage landing outcomes.
- SpaceX’s Falcon 9 first stage landing outcomes have been trending towards greater success as more launches are made.
- The machine learning models can be used to predict future SpaceX Falcon 9 first stage landing outcomes

APPENDIX

Initial Data Sets

- Wikipedia (Webpage): https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

APPENDIX

- **Data Sets (.csv files):**

- **GitHub URL (CSV 1):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /dataset_part_1.csv
- **GitHub URL (CSV 2):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /dataset_part_2.csv
- **GitHub URL (CSV 3):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /dataset_part_3.csv
- **GitHub URL (SpaceX):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /Spacex.csv
- **GitHub URL (Launch Geo):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /plotly_dash.py
- **GitHub URL (Plotly Dash):**
  https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /plotly_dash.py/

APPENDIX

Jupyter Notebooks and Dashboard Python File

- **GitHub URL (Data Collection):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /Data-collection-api.ipynb
- **GitHub URL (Web Scraping):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /Web Scraping.ipynb
- **GitHub URL (Data Wrangling):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /Data wrangling.ipynb
- **GitHub URL (EDA Sql data):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /EDA-Sql data.ipynb
- **GitHub URL (EDA with Data Visualization):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /EDA-Data visualization.ipynb
- **GitHub URL (Folium Maps):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /Interactive Visual Analytics with Folium.ipynb
- **GitHub URL (Plotly Dash):** https://github.com/RADHWAN0/IBM-Data-Science-Capstone-SpaceX /plotly_dash.py
- **GitHub URL (Machine Learning):** https://github.com/RADHWAN0/IBM-Data-Science-