SPACE RACE WITH DATA SCIENCE

KENNY HO

OUTLINES

- Summary
- Method
- Data Collection
- SpaceX API
- Web Scraping
- Data Wrangling
- EDA with Visualization
- EDA with SQL
- Maps with Folium
- Plotly Dash and Dashboard
- Predictive Analysis
- Final Results
- EDA Visualization
- Payload vs Orbit
- Flight vs Orbit
- Launch Sites
- Interactive maps
- Dash
- Conclusion

SUMMARY

- Methodologies that were used to analyze the data
  - Data Collection using Web Scraping
  - EDA including data wrangling, and visualization and interactive analytics
  - ML predictions
- Summary
  - Collected valuable data from public sources
  - EDA identified the features to predict success of launchings
  - ML show model to predict characteristic

METHOD

- Data collection method
  - Space X API using Https://api.spacexdata.com/v4/rockets/
  - Webscraping:
    - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
- Data wrangling
  - Collected data create a landing outcome based on the outcome data after summarizing and analyzing different key features

METHOD

- Visualization analytics done through Folium and Plotly Dash
- Predictive analysis using Classification models

DATA COLLECTION

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

SPACE X API

- SpaceX provide public API from where the data can be gathered and utilized

- Source: https://github.com/hokenny9/Capstone-Project/blob/main/Data%20Collection%20API.ipynb

WEB SCRAPING

- Source: https://github.com/hokenny9/Capstone-Project/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

DATA WRANGLING

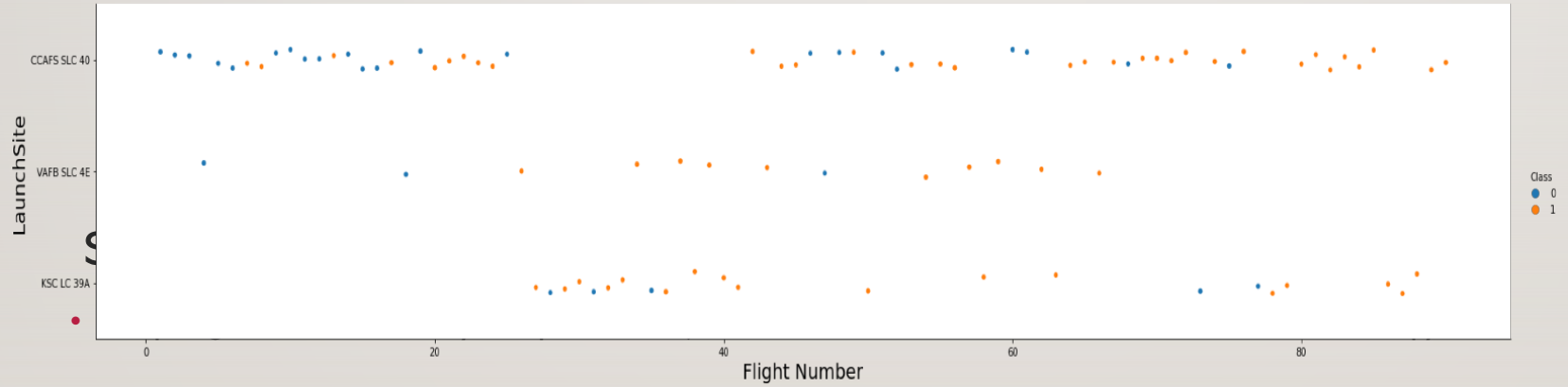
- EDA was performed on the dataset
- EDA provides a summary of launches per sites and around each orbit
- The landing outcome label created and generated from the Outcome Colum

- Source: https://github.com/hokenny9/Capstone-Project/blob/main/Data%20Wrangling.ipynb

EDA WITH VISUALIZATION

- scatterplots and barplots were used to visualize the relationship between pair of features:

• Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



EDA WITH SQL

- Following SQL Queries:
  - Names of unique launch site
  - 5 launch sites begin with CCA
  - Average payload for booster F9V1.1
  - Names of boosters for mass of 4000 and 6000 kg
  - Number of successful and failure missions
  - Failed landing outcomes

Source: https://github.com/hokenny9/Capstone-Project/blob/main/EDA.ipynb

MAP WITH FOLIUM

- Circle / markers / Lines / marker cluster were utilized with the Folium Maps
- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and
- Lines are used to indicate distances between two coordinates.

- Source: https://github.com/hokenny9/Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

PLOTLY DASH DASHBOARD

- Percentage launches by site and Payload range were used to visualize the data

- Source: https://github.com/hokenny9/Capstone-Project/blob/main/spacex_dash_app.py

PREDICTIVE ANALYSIS

- Four classification were used to compare
  - Logistic Regression
  - Decision Tree
  - K nearest Neighbors
  - Support Vecotrs

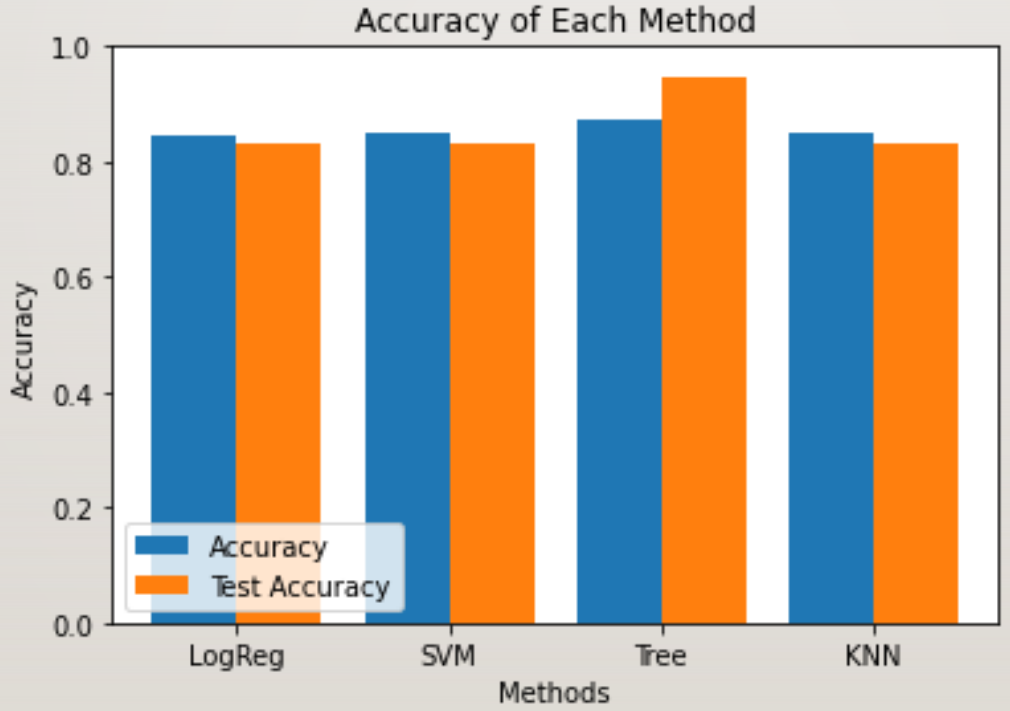
- Source: https://github.com/hokenny9/Capstone-Project/blob/main/Machine%20Learning%20Prediction.ipynb

FINAL RESULT

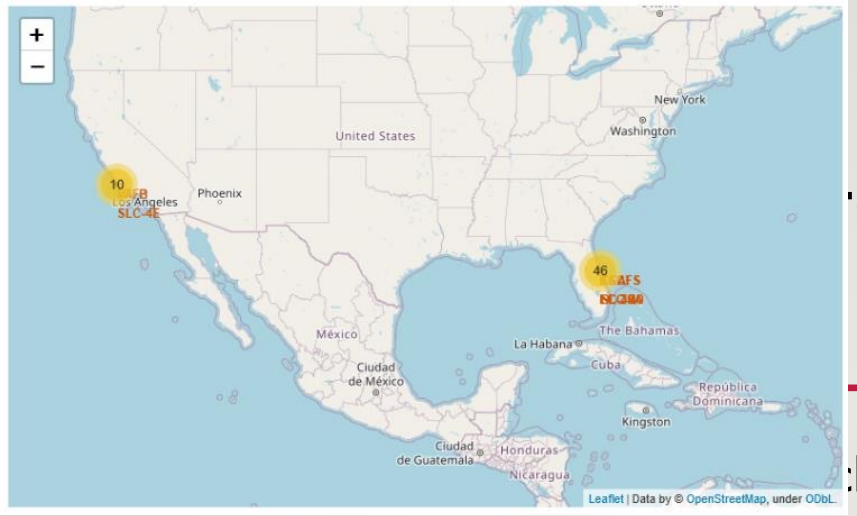
- Space X use 4 different sites
- Launches Number 1 by SpaceX and Nasa
- F9V1.1 Booster is 2928 KG
- 100% mission success
- Number of landing success increased as years go by
- Falcon 9 booster versions were successful

FINAL RESULTS

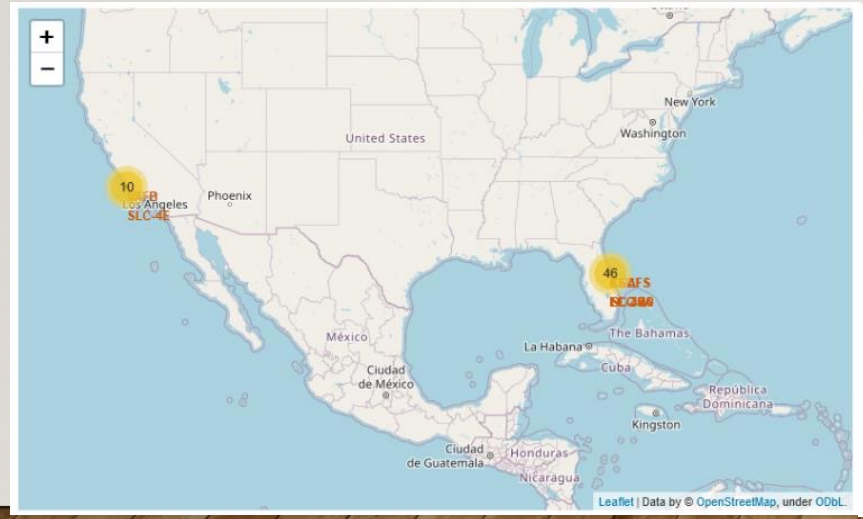
- Decision Tree Classifier is the best model for predicting successful landings, over 87% accuracy and test data at 94#

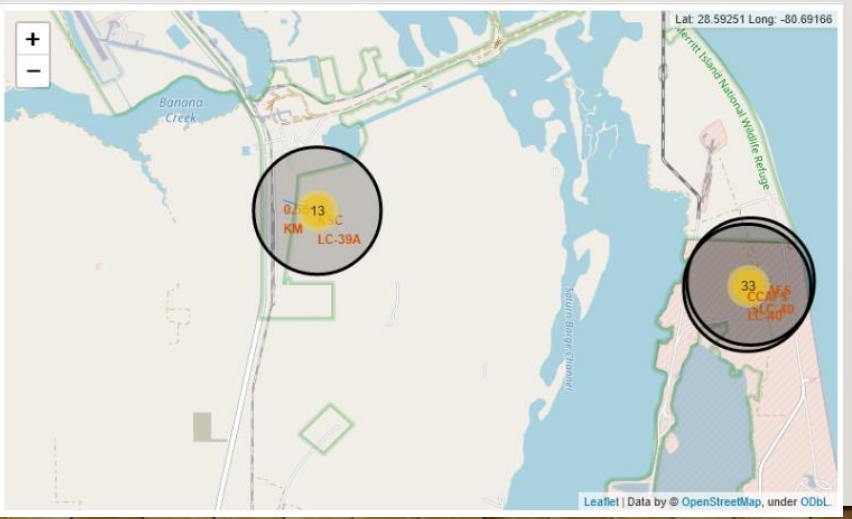


'S



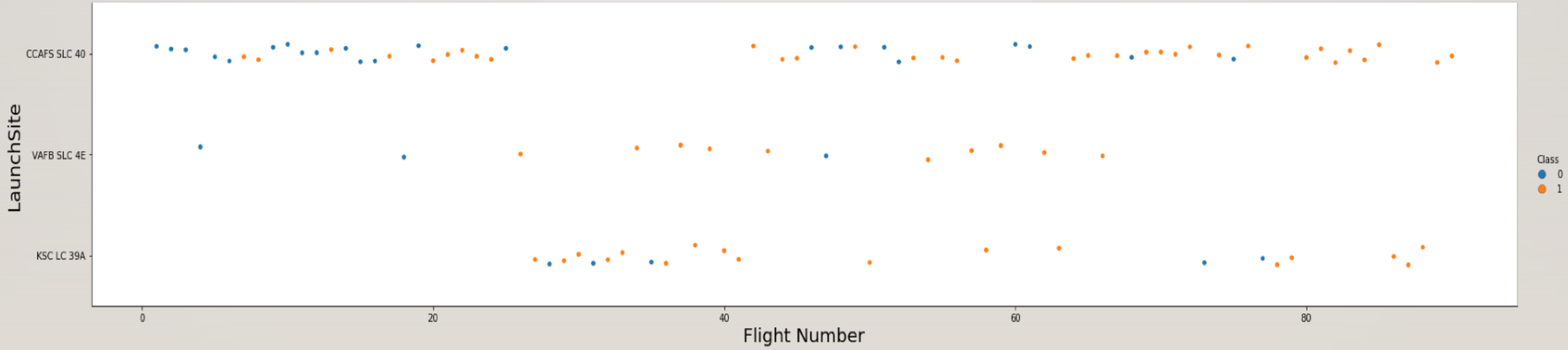
:hes happens at east coast launch site

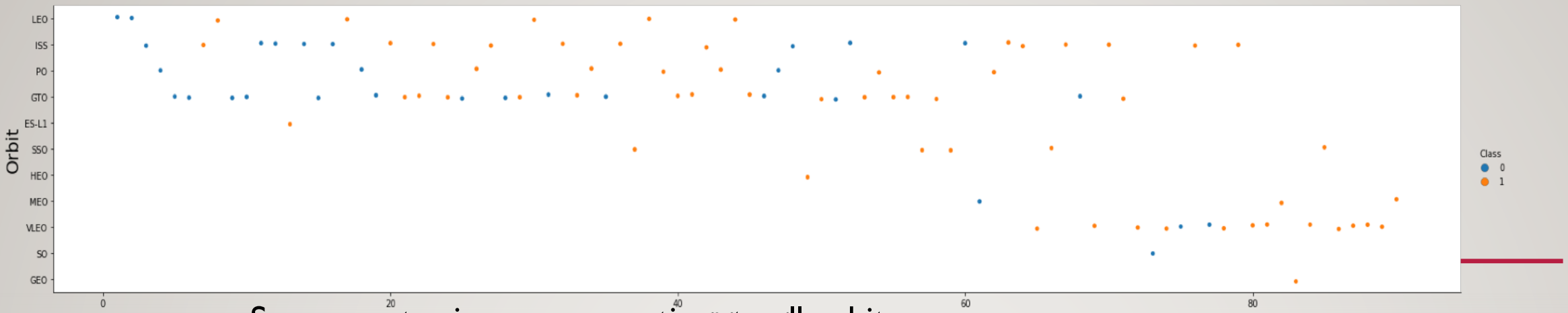




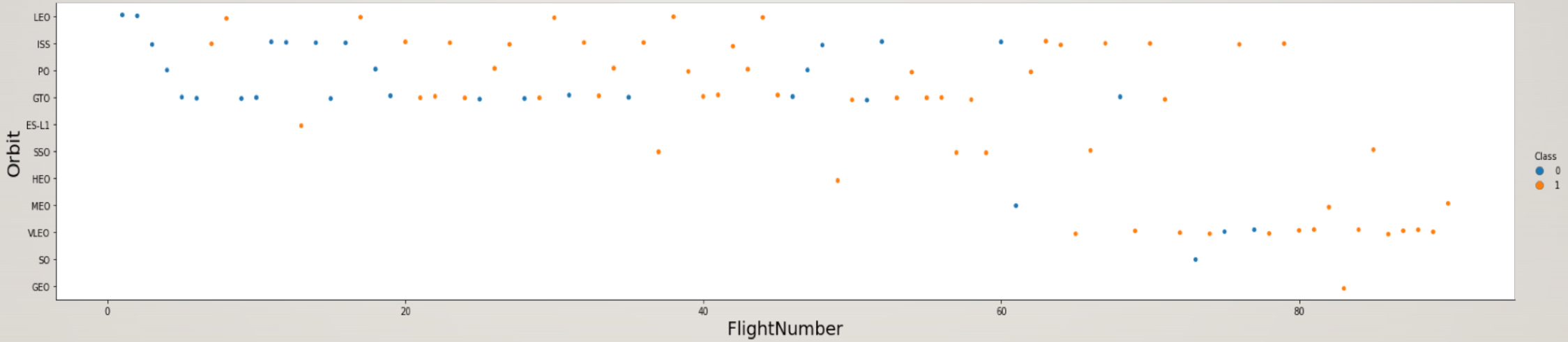
EDA

- Flight Number vs Launch Site
  - It's possible to verify the best site for CCAF5 SLC 40
  - VAFB SLC 4E for 2nd Place
  - KSC LC 39A for 3rd Place



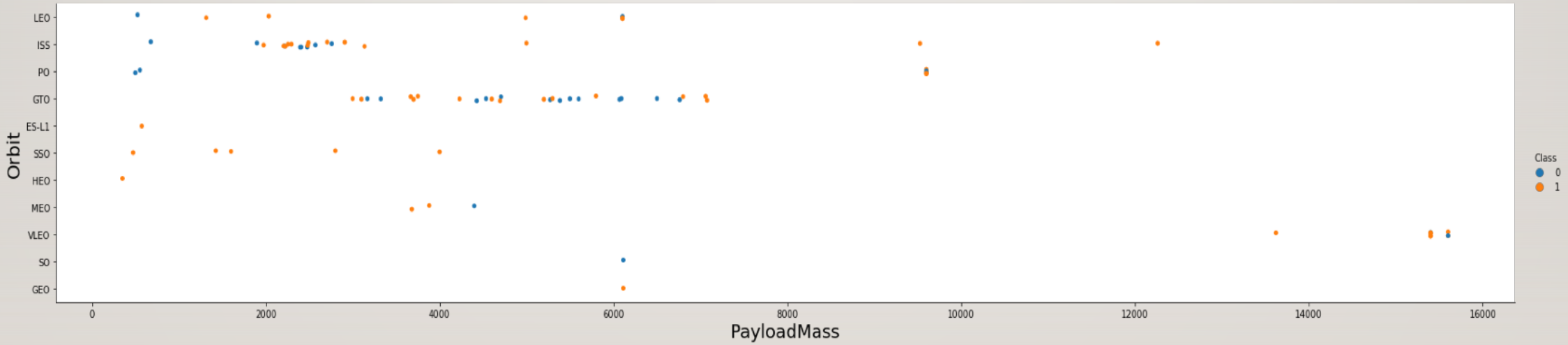


- Success rates improve over time to all orbit



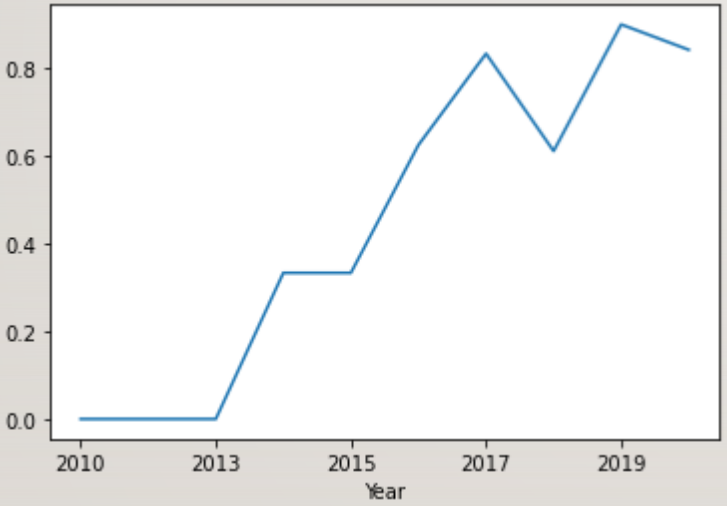
PAYLOAD VS ORBIT

- No relation between the payload and success rate of orbit
- The widest Range of Payload and success rate increase for the ISS



YEARLY LAUNCH TREND SUCCESS

- Rate began increasing from 2013 and continued to 2020
- First three years were time for improvement for technology



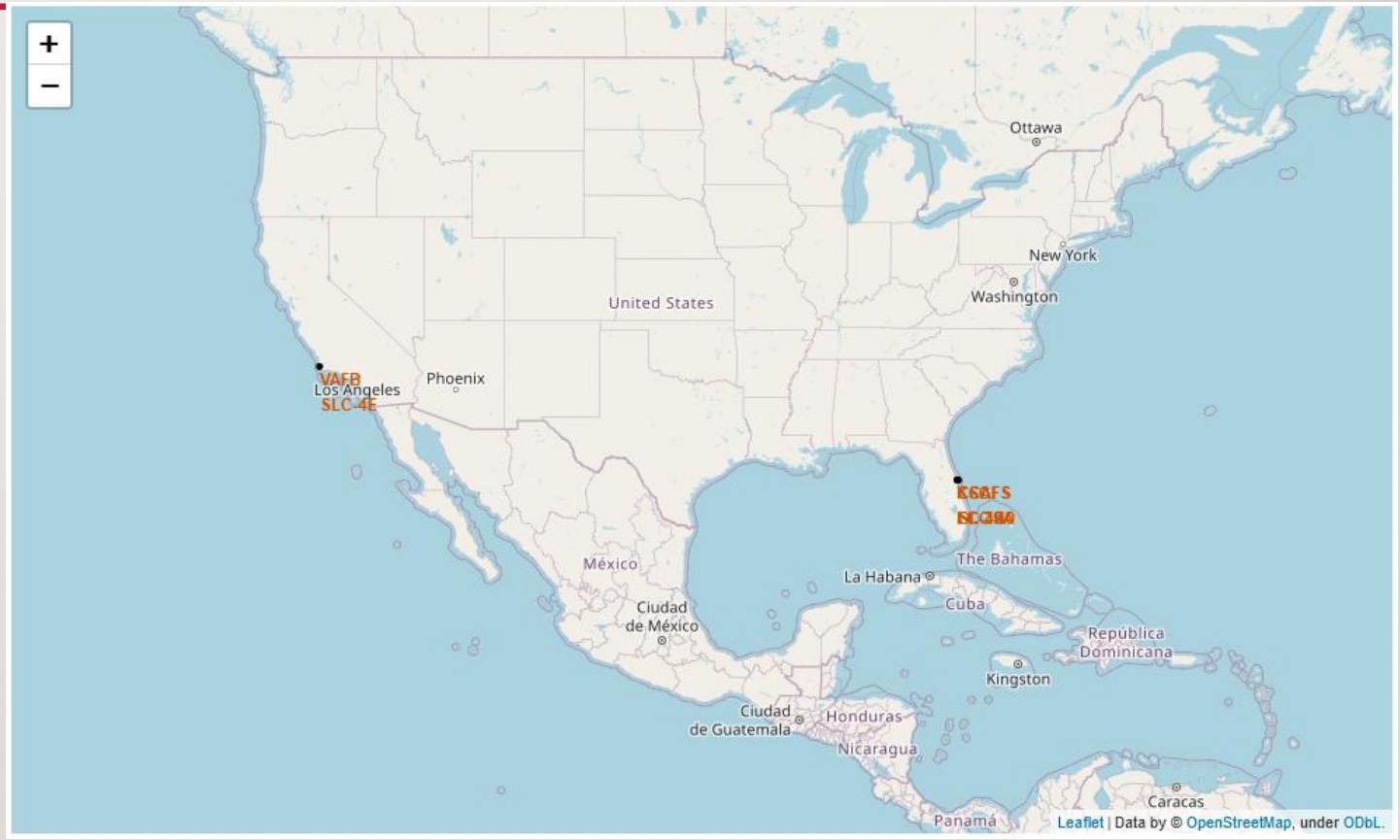
LAUNCH SITE

- Four launch sites
- CCAFS LC-40
- CCAFS SLC-40
- KSC LC-39A
- VAFB SLC-4E

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

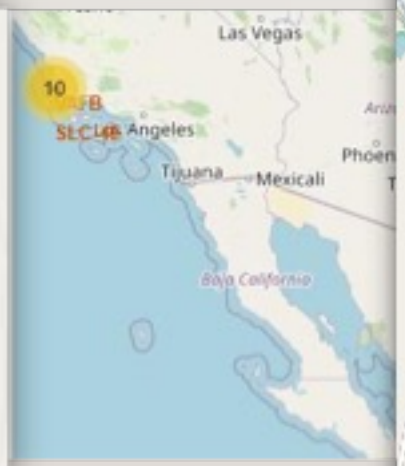
INTERACTIVE MAP AND FOLIUM

- Launch near the ocean



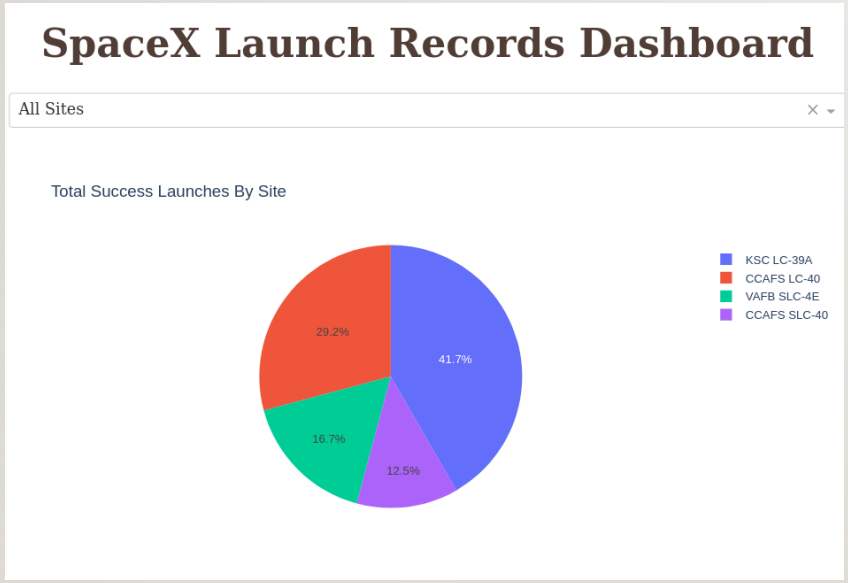
ERACTIVE MAPS AND FOLIUM

- Red indicate failures
- Green indicate success



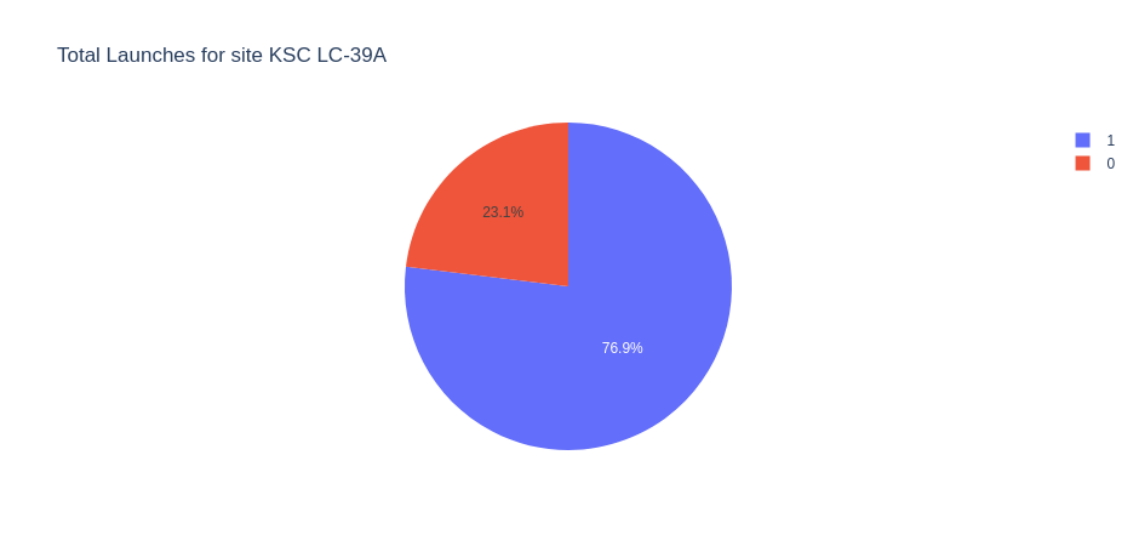
PLOTLY DASH RELATED

- Area where launches were conducted play a crucial role for the success of the mission



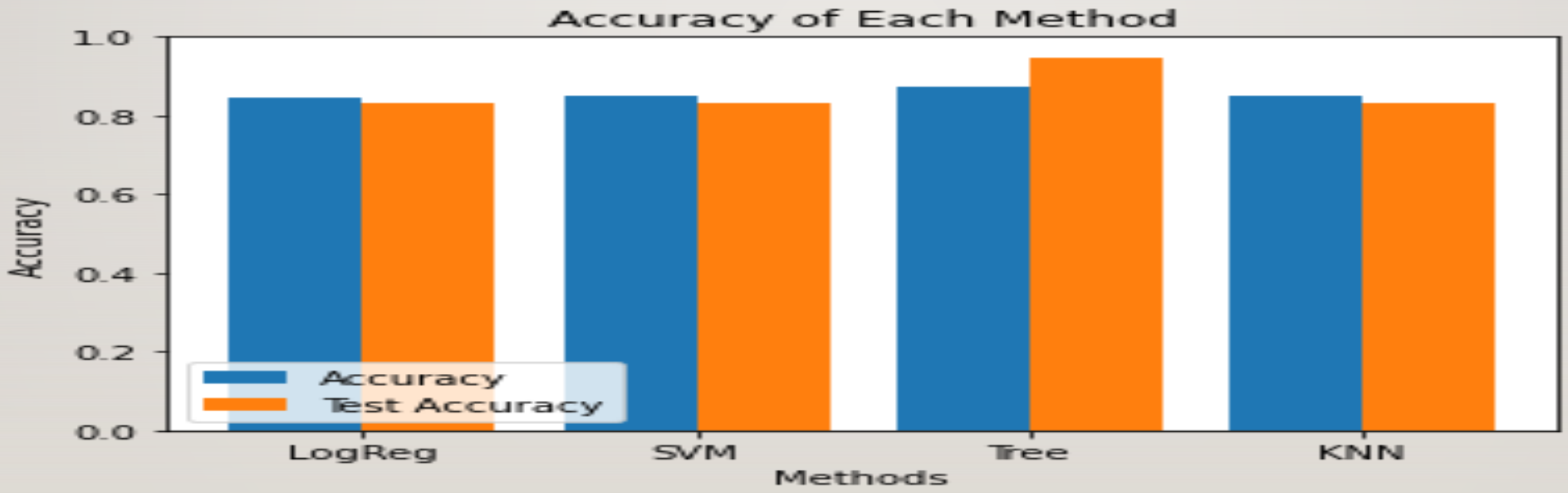
PLOTLY DASH RELATED

- 76.9% launches were success by the KSC LC-39A site



PREDICTIVE ANALYSIS

- Classifications were tested and their accuracies are plotted
- Highest Classification is the Decision Tree Classifier at 87%



PREDICATIVE ANALYSIS

- Confusion Matrix for the Decision Tree Classifier proves the accuracy of true positive and true negative

CONCLUSION

- Best launch site is the KSC LC-39A
- Launches above the 7000 kg weight scale is less risky
- Decision Tree Classifier can be used to predict the success of the landing
- Successful landing outcomes improve over time, according to the improvement of technology and rocket power