DATA SCIENCE PROJECT - SPACEX

DHARANI ARUMUGAM

DATE: 13-06-2023

OUTLINE

- Executive summary
- Table of contents
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Gathered spacex data from public spacex API and Wikipedia page. Created class column to tell about successful landings since our analysis is to find successful landing of first stage. Explored data using SQL ,Visualisation and Folium maps. Done with feature engineering to determine which among the given features will get us to our expected result with high accuracy. Changed all categorical variable to numberic with one hot coding. Gridsearch CV used to find the hyperparameters to fine tune our model.

- Machine learning algorithms used: Logistic regression, SVM, Decision tree,KNN. All produced similar results with 83.33% accuracy.

TABLE OF CONTENTS

INTRODUCTION

- SpaceX given the past data about rocket launch to determine whether first stage will land successfully or not. Space Y wants to compete with spaceX. Space Y given us a task to train a model to predict successful stage 1 recovery

METHODOLOGY

Methodology includes

- Data cleansing and pre-processing

  Combined data from spaceX public API and spaceX Wikipedia page

- EDA(Exploratory Data Analysis)

  Determine the successful landings

- Data Visualisation

  Folium and Plotly Dash

- Machine Learning Algorithms

DATA COLLECTION

- Data collection process involved a combination of API request from spaceX public API and Wikipedia website.
- Request and parse the spaceX launch data using GET request
- Filter data frame to only Falcon 9 launches

All the required data gathered from the above two sites.

DATA WRANGLING

EDA WITH VISUALISATION

- Exploratory Data Analysis performed on variabes like flight no,payload mass,launch site,orbit,year and class.
- Charts used:

Scatter plot, line,bar were used to compare relaionships between variables to decide if a relationship exists so that they could be used in training machine learning model.

EDA WITH SQL

Steps to be followed

- Load the dataset in IBM DB2
- Query the dataset with python integration
- Query the data to get better understanding of the data
- Queried info about columns in the dataset

INTERACTIVE MAP WITH FOLIUM

- Folium is used to build interactive maps
- Folium map is used to mark launch sites ,successful and unsuccessful landings and proximity to key locations

BUILD A DASHBOARD WITH PLOTLY DASH

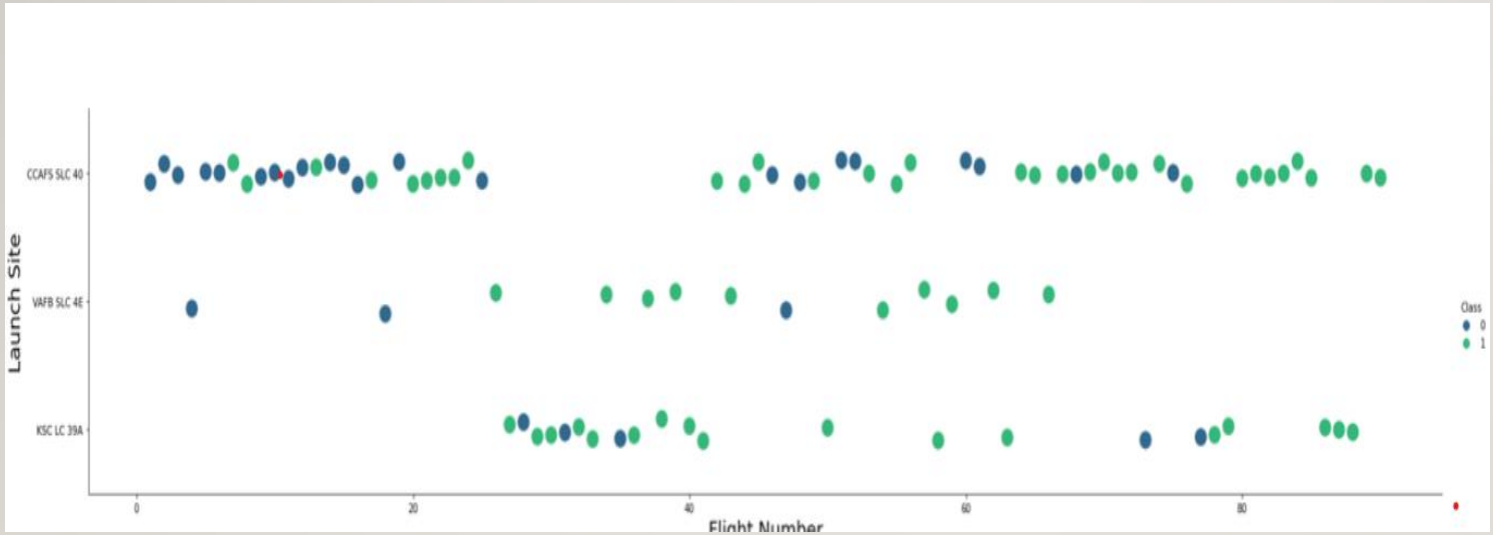
- Dashboard includes a pie chart and a scatter plot
- Pie chart can be selected to show distribution of successful landings across all launch sites.
- Pie chart shows us launch site success rate

RESULTS

Ploty dashboard
It shows the result of our model with 83.33% accuracy.

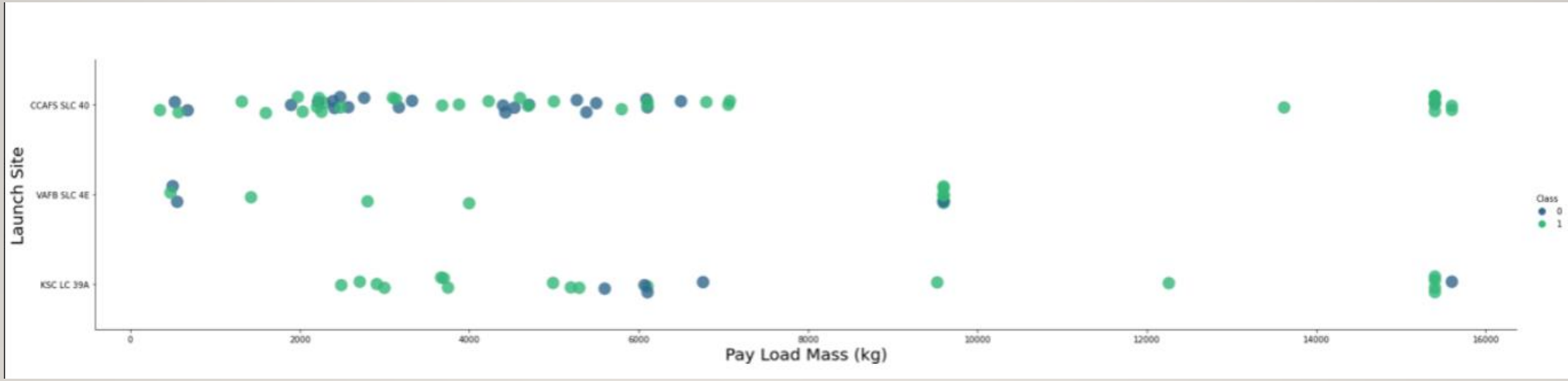
It shows EDA with visualisation ,EDA with SQL, Interactive map.

FLIGHT NUMBER VS LAUNCH SITE



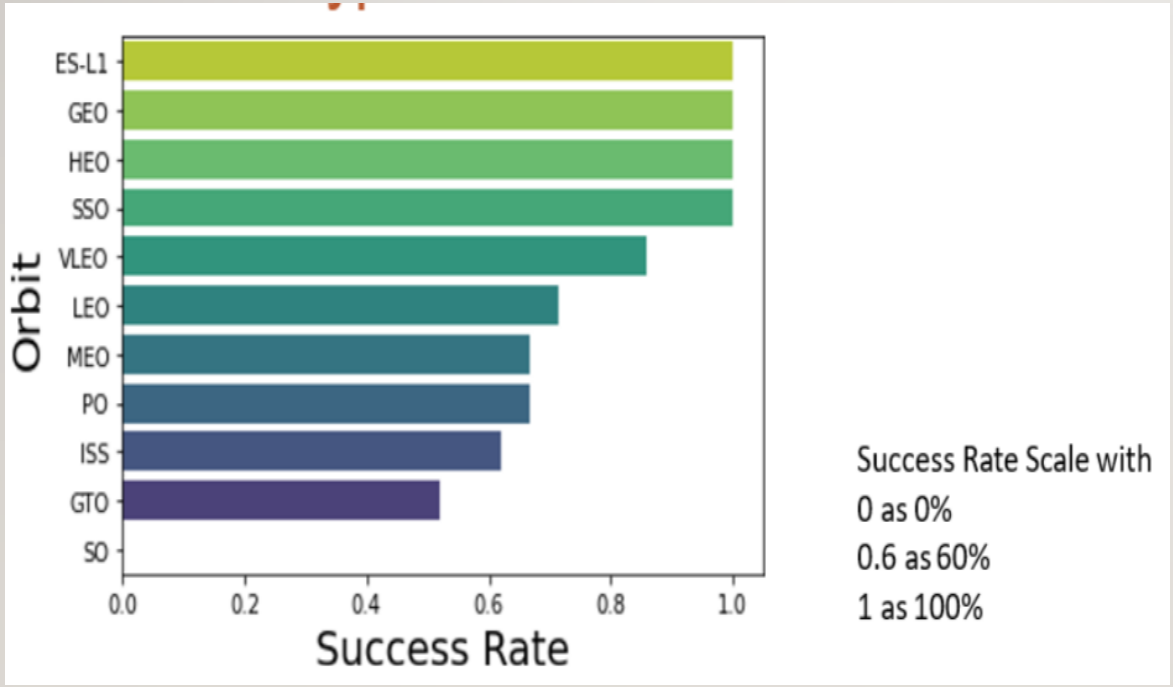
Green indicates successful launch and purple indicates unsuccessful launch

PAYLOADMASS VS LAUNCH SITE



Green indicates successful launch and purple indicates unsuccessful launch

SUCCESS RATE VS ORBIT

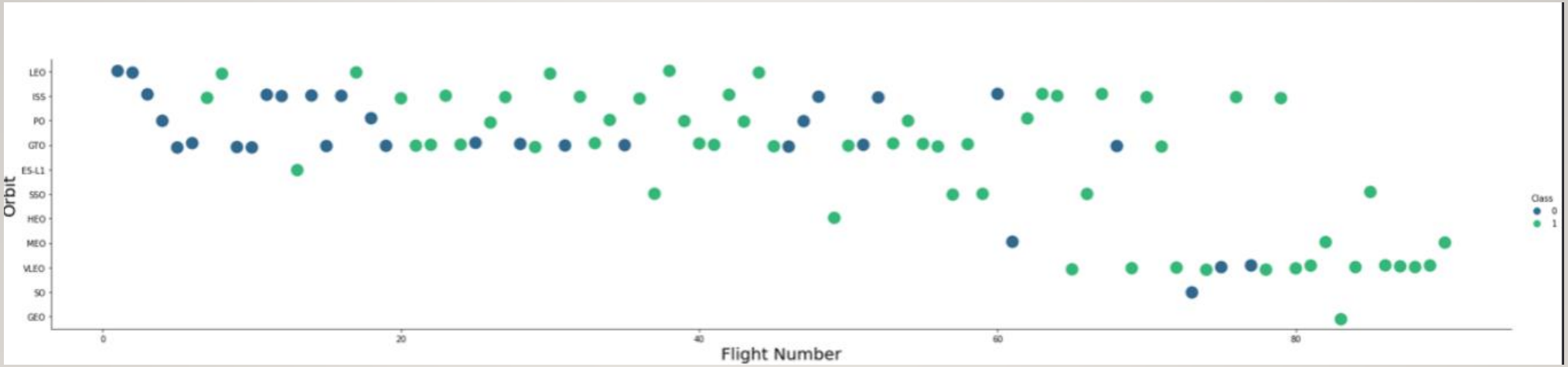


ES-LI,GEO,HEO,SSO has 100% success rate

VLEO has decent success rate and attempts

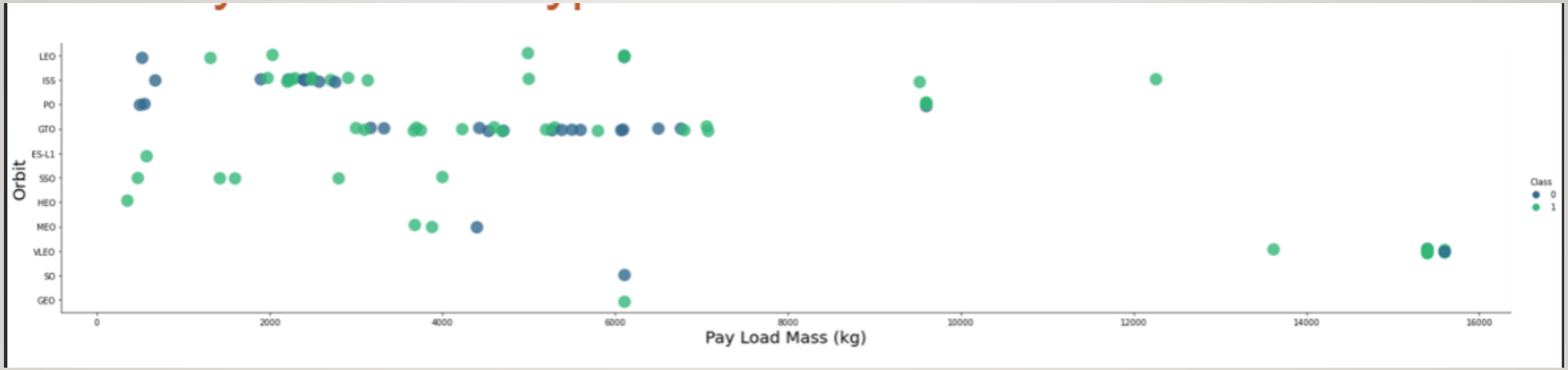
GTO has very low success rate

FLIGHT NUMBER VS ORBIT



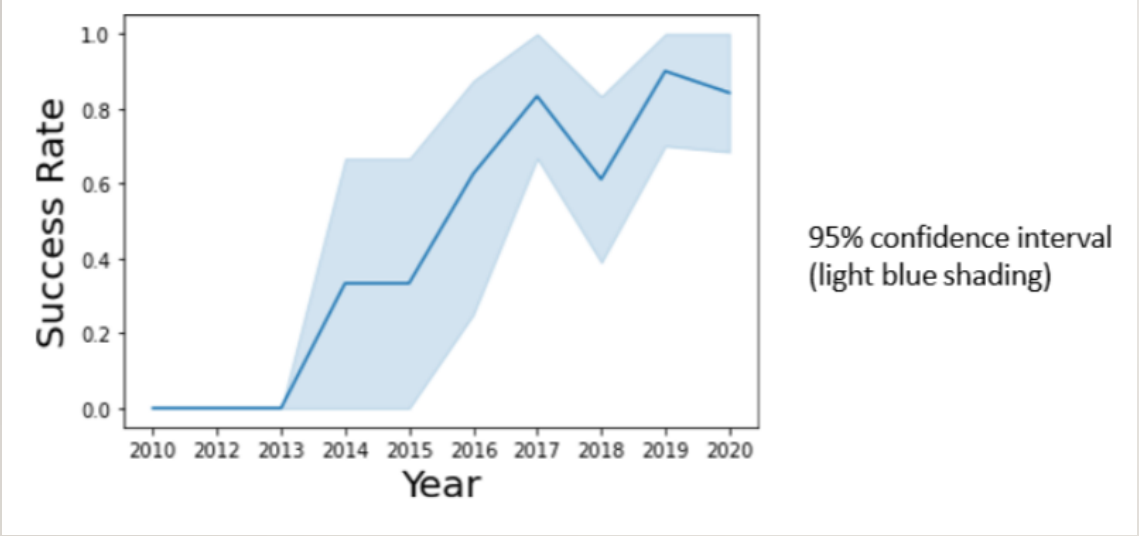
Green – success
Purple - Failure

PAYLOADMASS VS ORBIT



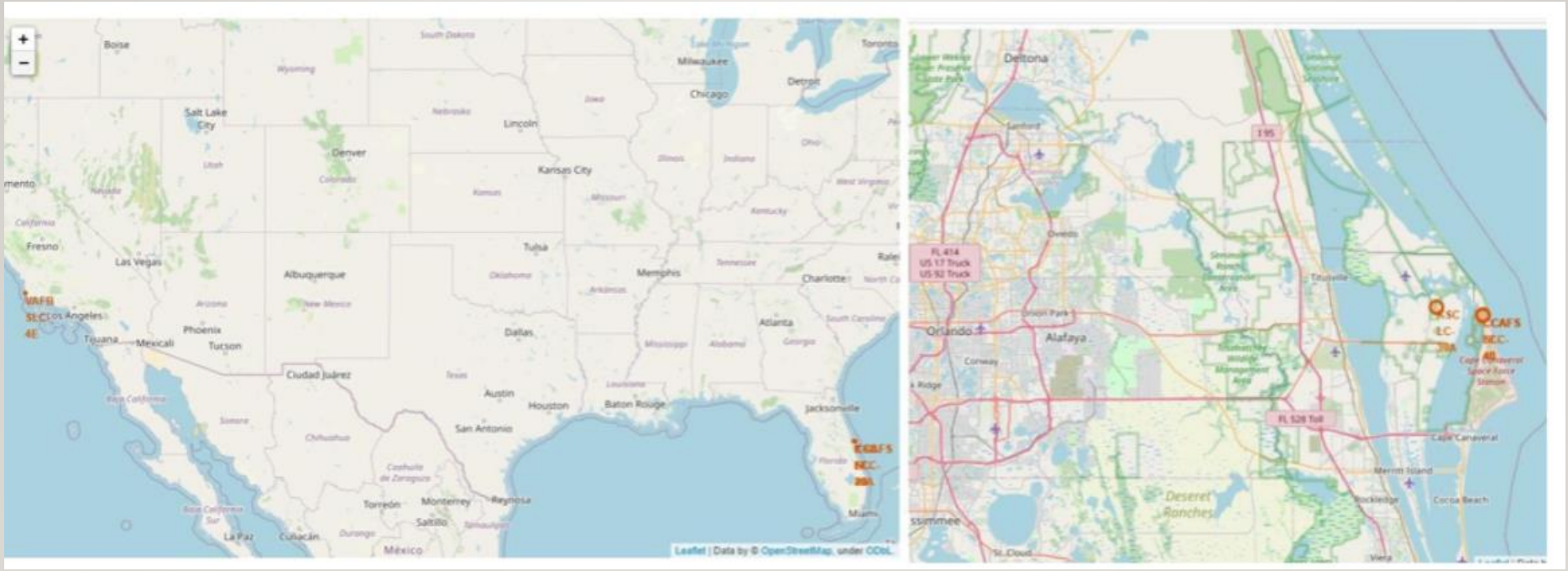
Green – success
Purple - Failure

LAUNCH SUCCESS YEARLY TREND

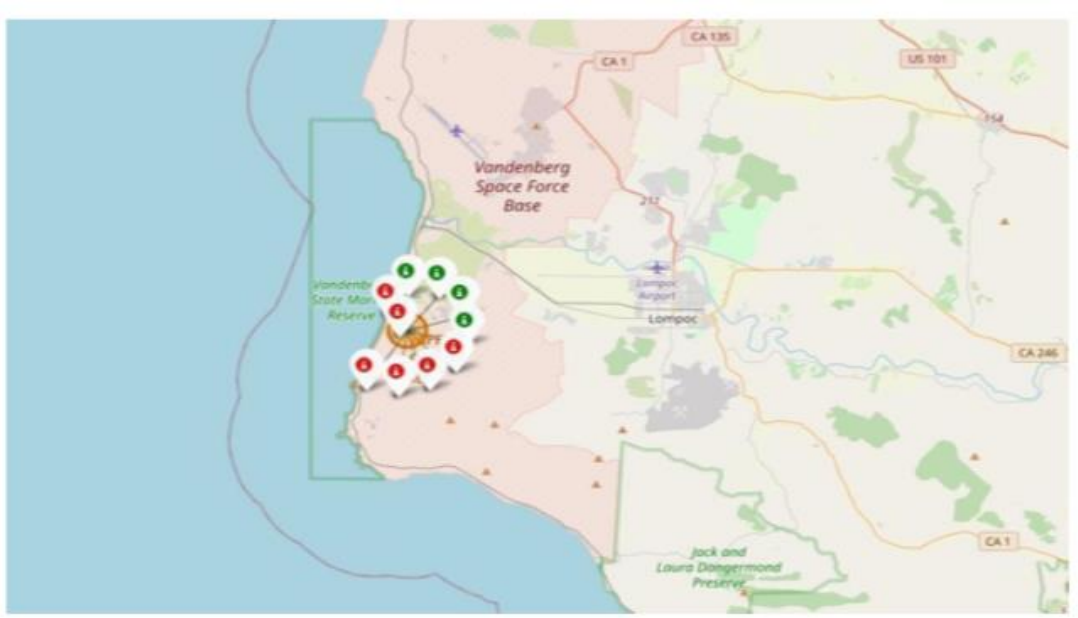


95% confidence interval
(light blue shading)

INTERACTIVE MAP WITH FOLIUM



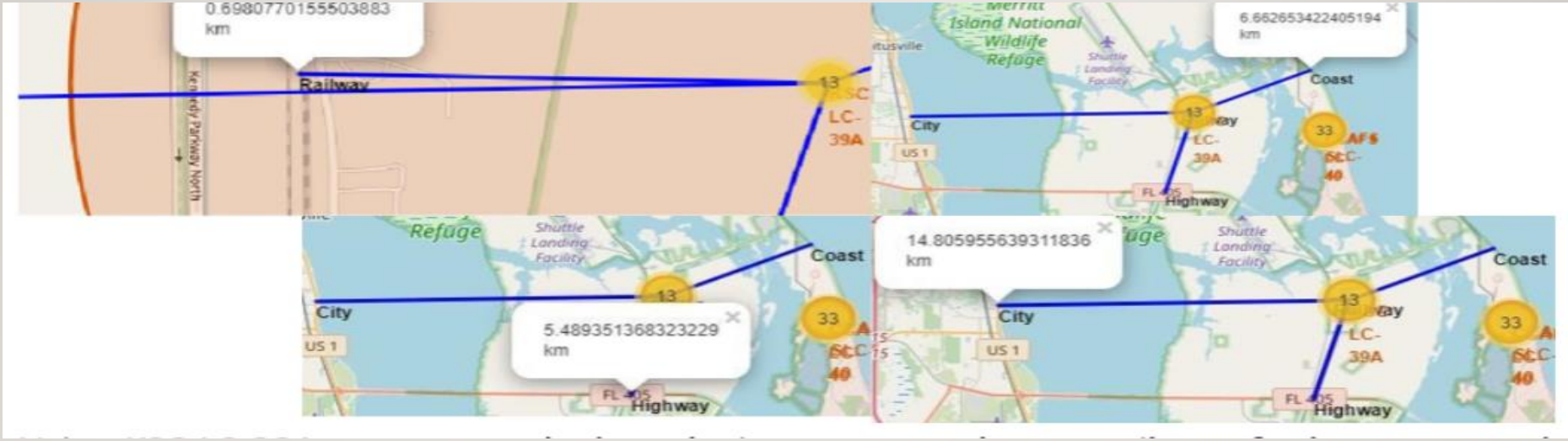
COLOR-CODED LAUNCH MARKERS



Clusters on Folium map can be clicked on to display each successful landing (green icon) and failed landing (red icon).

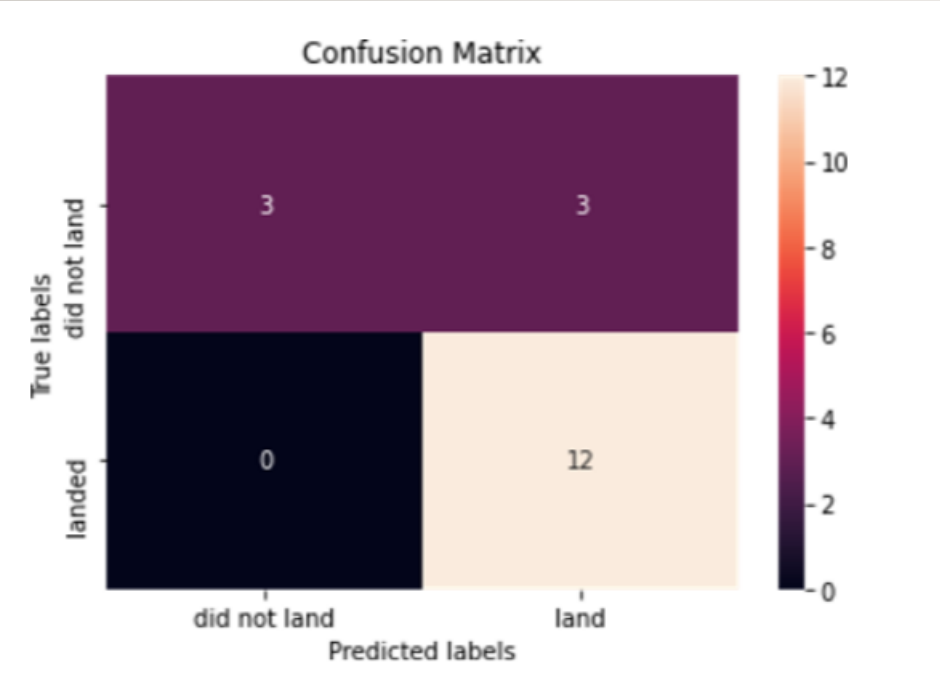
In this example VAFB SLC -4E shows 4 successful landings and 6 failed landings.

KEY LOCATION PROXIMITIES



Launch sites are also close to coasts and relatively far from cities so that launch failures can land in the sea to avoid rockets falling on densely populated areas.

CONFUSION MATRIX



Model predicted 12 successful landings correctly – TT
Model predicted 3 unsuccessful landing correctly – FF

False positive – Model predicted 3 successful landing when true label was unsuccessful

CONCLUSION

- Our task: to develop a machine learning model for Space Y who wants to bid against SpaceX
- The goal of model is to predict when Stage 1 will successfully land to save ~$100 million USD
- Used data from a public SpaceX API and web scraping SpaceX Wikipedia page
- Created data labels and stored data into a DB2 SQL database
- Created a dashboard for visualization
- We created a machine learning model with an accuracy of 83%
- Allon Mask of SpaceY can use this model to predict with relatively high accuracy whether a launch will have a successful Stage 1 landing before launch to determine whether the launch should be made or not
- If possible more data should be collected to better determine the best machine learning model and improve accuracy

APPENDIX

- Github URL
- DharaniAB/Spacex-Project (github.com)

- Thanks to all the instructors in coursera.