- BY ABIR CHAKRABORTY

IBM DATA SCIENCE CAPSTONE PROJECT

Space X Falcon 9 Landing Analysis

Executive Summary

- Data Collection

- Data Wrangling

- Exploratory Data Analysis

- Predictive Analysis (Classification Model)

Introduction

SpaceX launches Falcon 9 rockets costs $62m. This is considerably cheaper than other providers (which usually cost upwards of $165m), and much of the savings are because SpaceX can land, and then re-use the first stage of the rocket.

- If we can make predictions on whether the first stage will land, we can determine the cost of a launch, and use this information to assess whether or not an alternate company should bid and SpaceX for a rocket launch.

This project will ultimately predict if the Space X Falcon 9 first stage will land successfully through a classification model

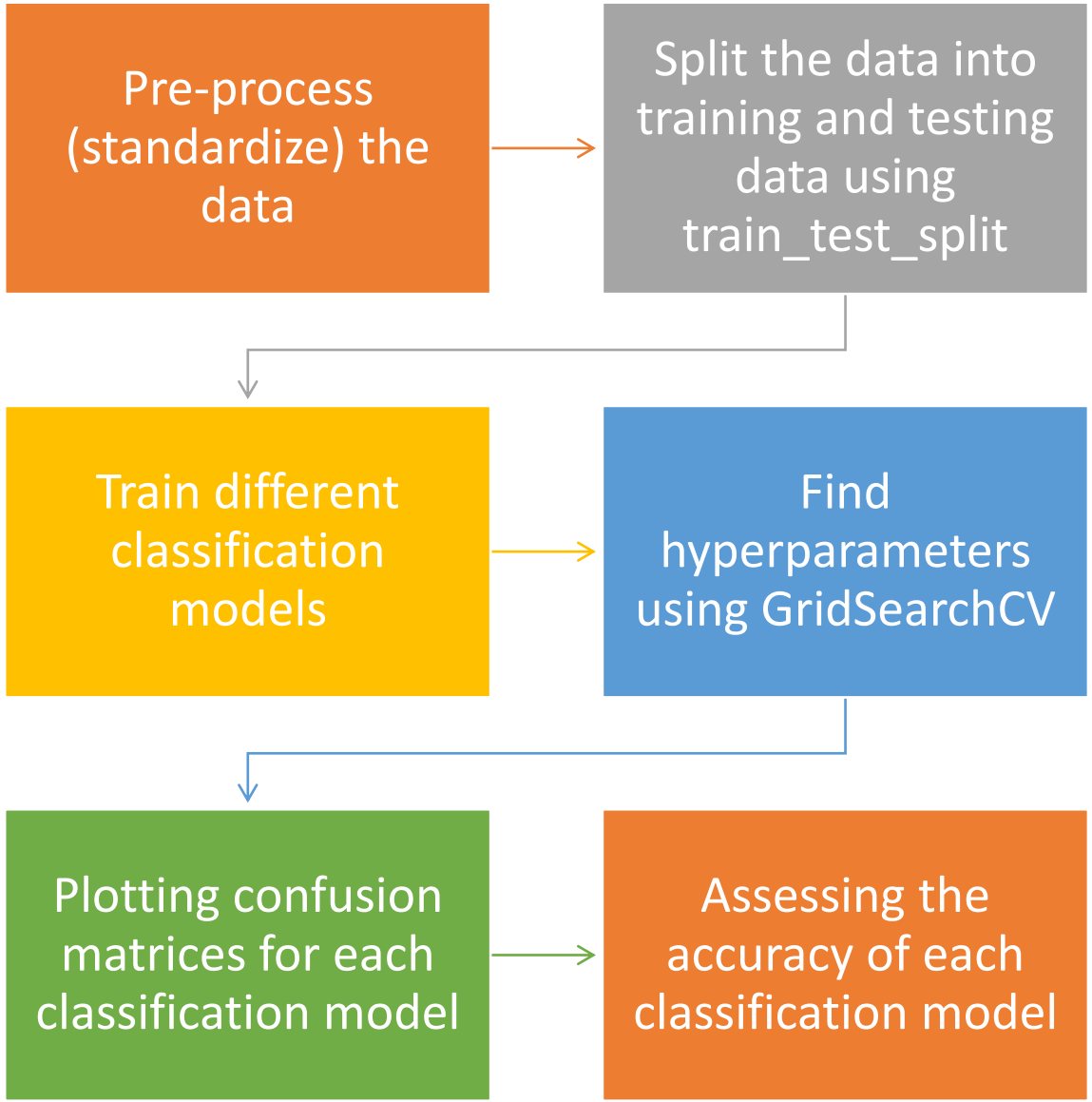
Data collection and data wrangling methodology

- Making GET requests to the SpaceX REST API
- Web Scraping
- Using the .fillna() method to remove NaN values
- Using the .value_counts() method to determine the following:
- Number of launches on each site
- Number and occurrence of each orbit
- Number and occurrence of mission outcome per orbit type

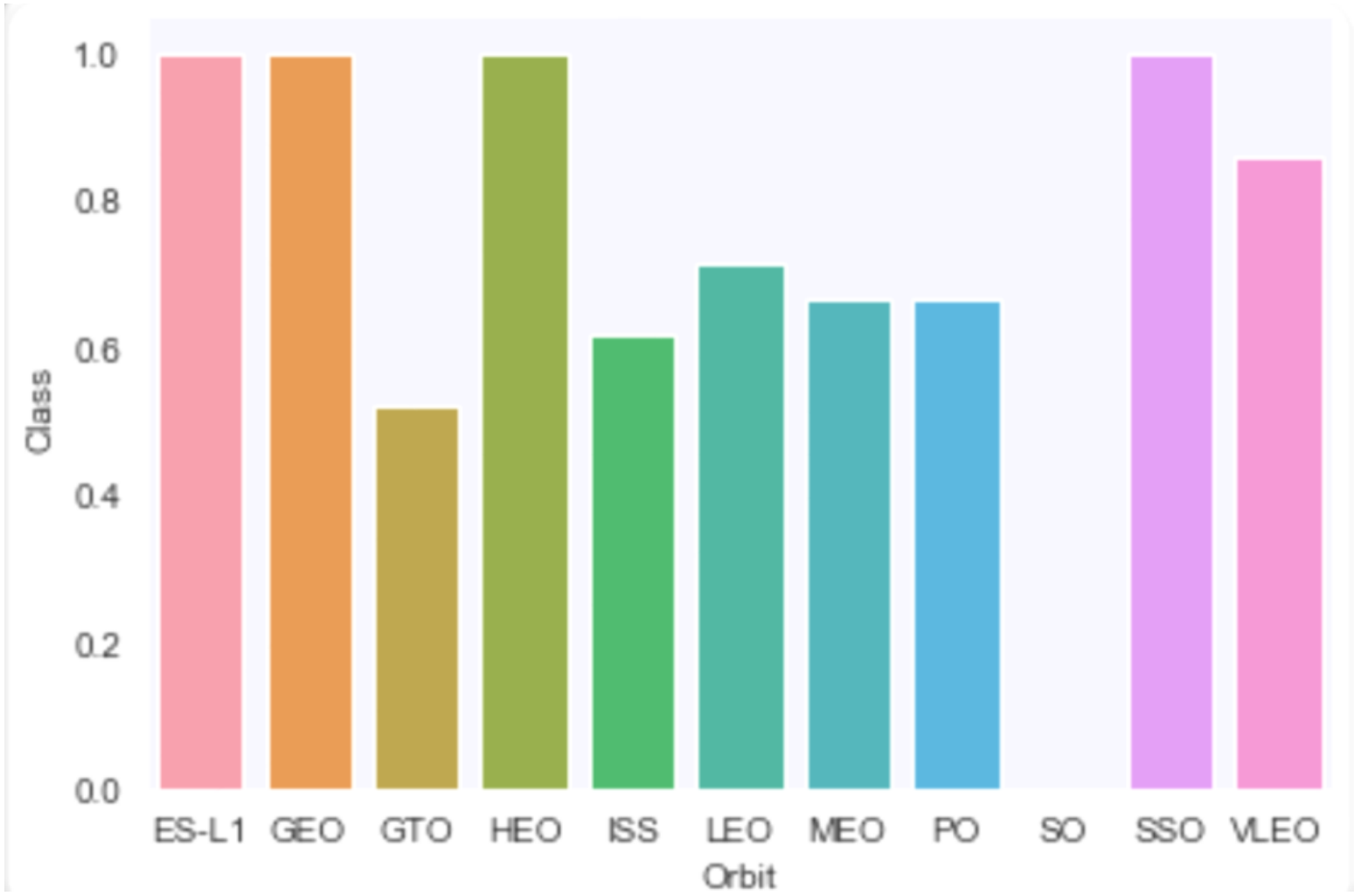
EDA and interactive visual analytics methodology

- Using SQL queries to manipulate and evaluate the SpaceX dataset
- Using Pandas and Matplotlib to visualize relationships between variables, and determine patterns
- Geospatial analytics using Folium
- Creating an interactive dashboard using Plotly Dash

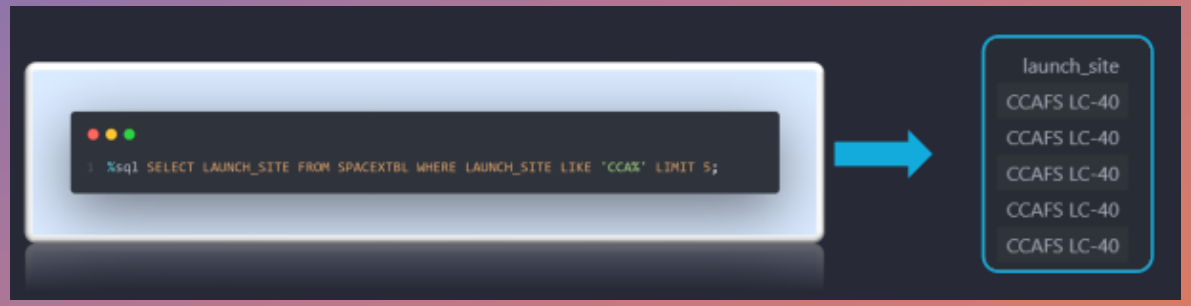
Predictive analysis methodology

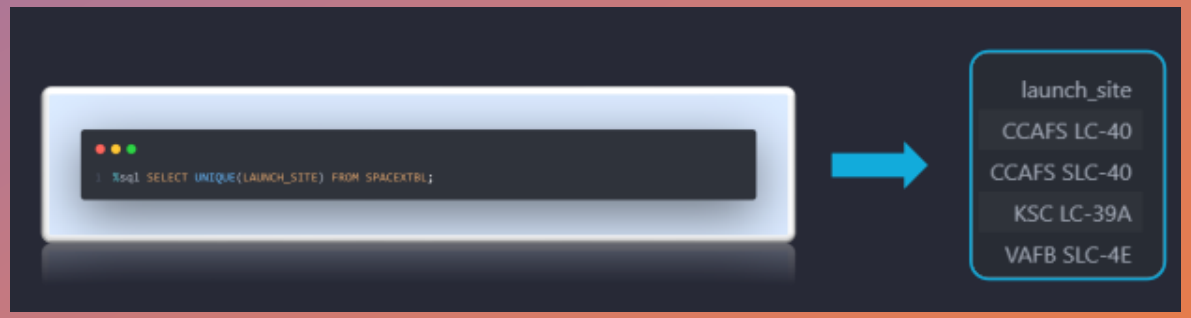


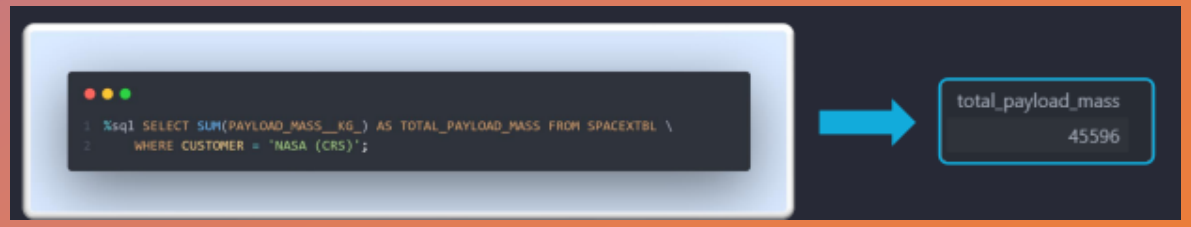
SUCCESS RATE VS. ORBIT TYPE



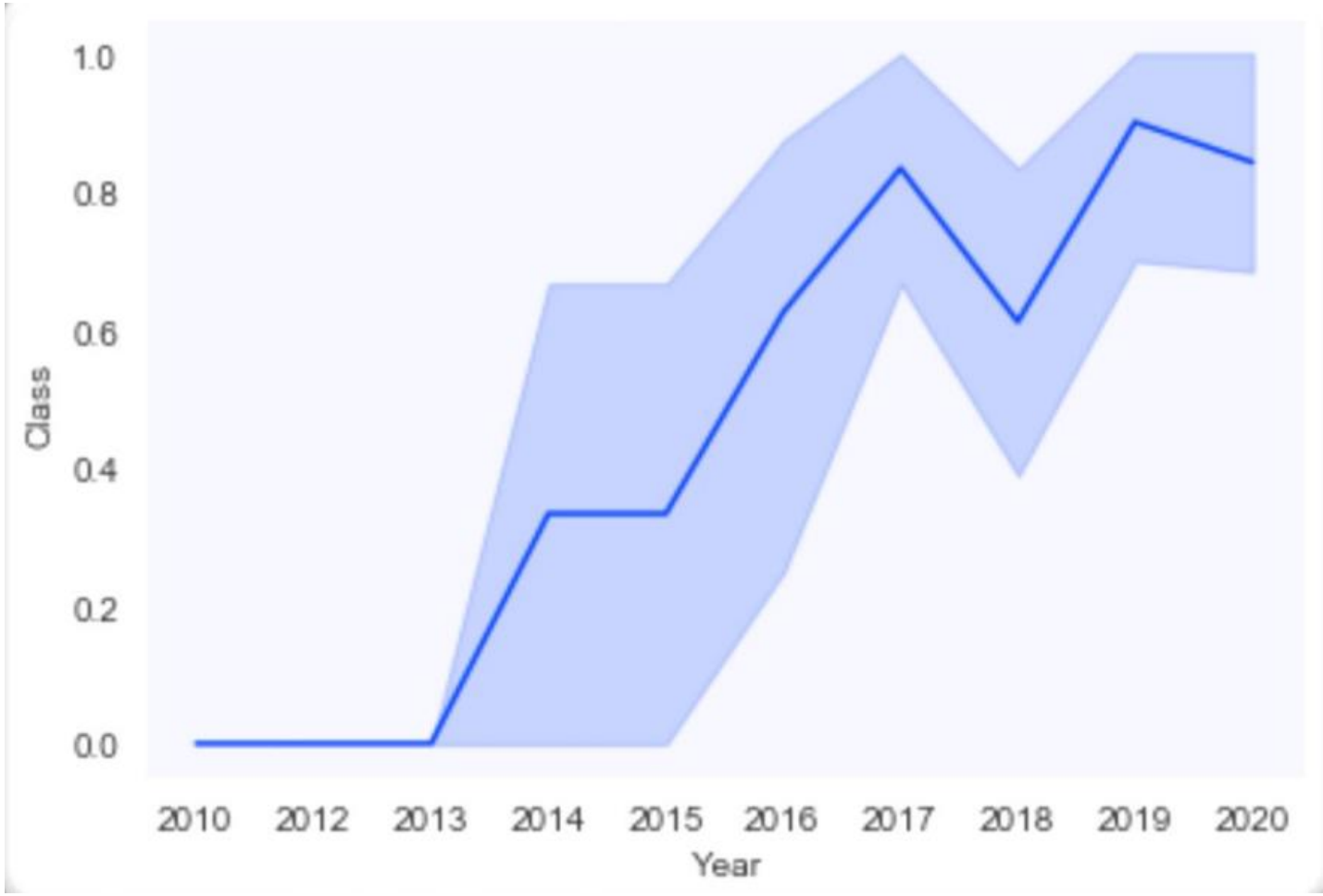
EDA with SQL


%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

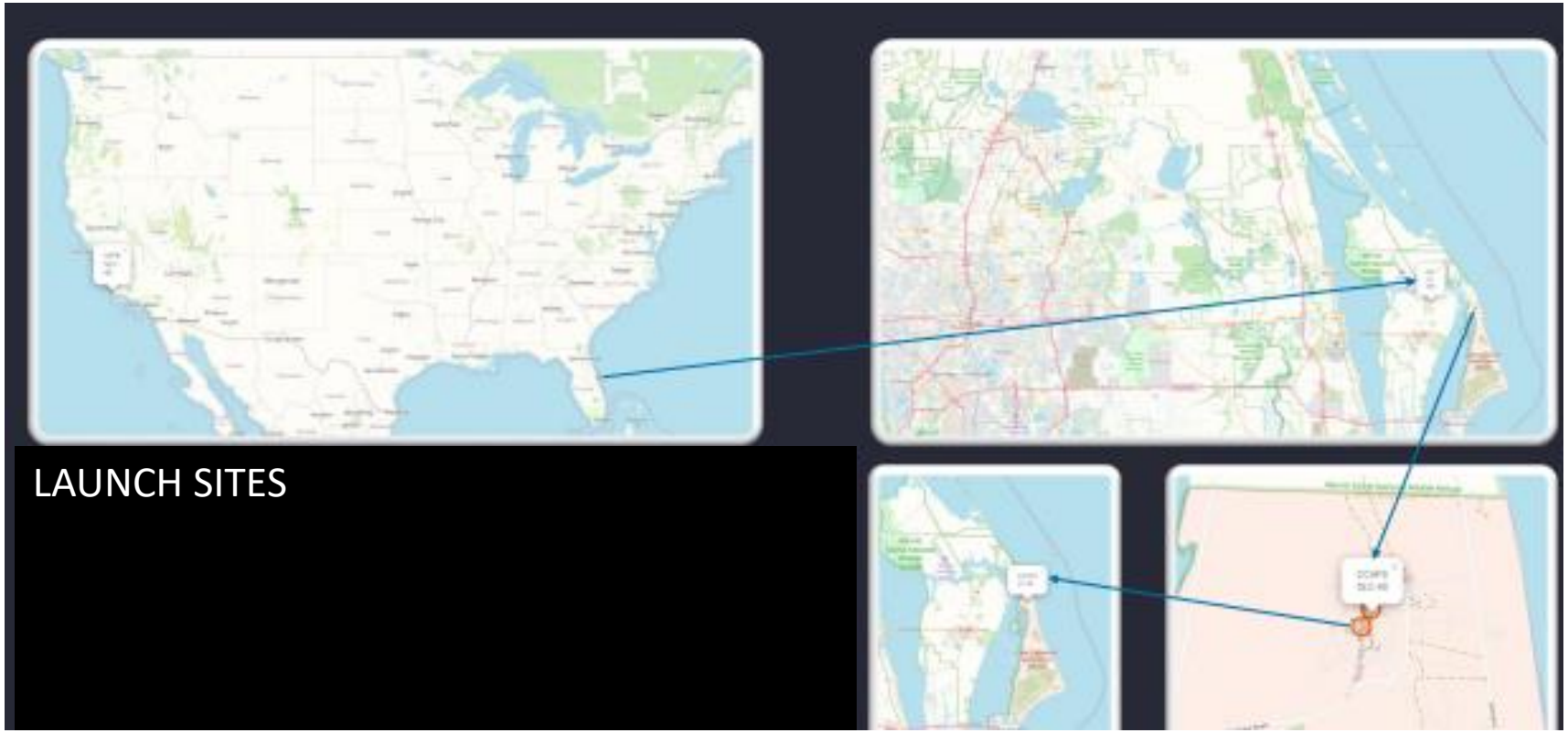

%sql SELECT UNIQUE(LAUNCH_SITE) FROM SPACEXTBL;
launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E


%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL \
WHERE CUSTOMER = 'NASA (CRS)';
total_payload_mass
45596

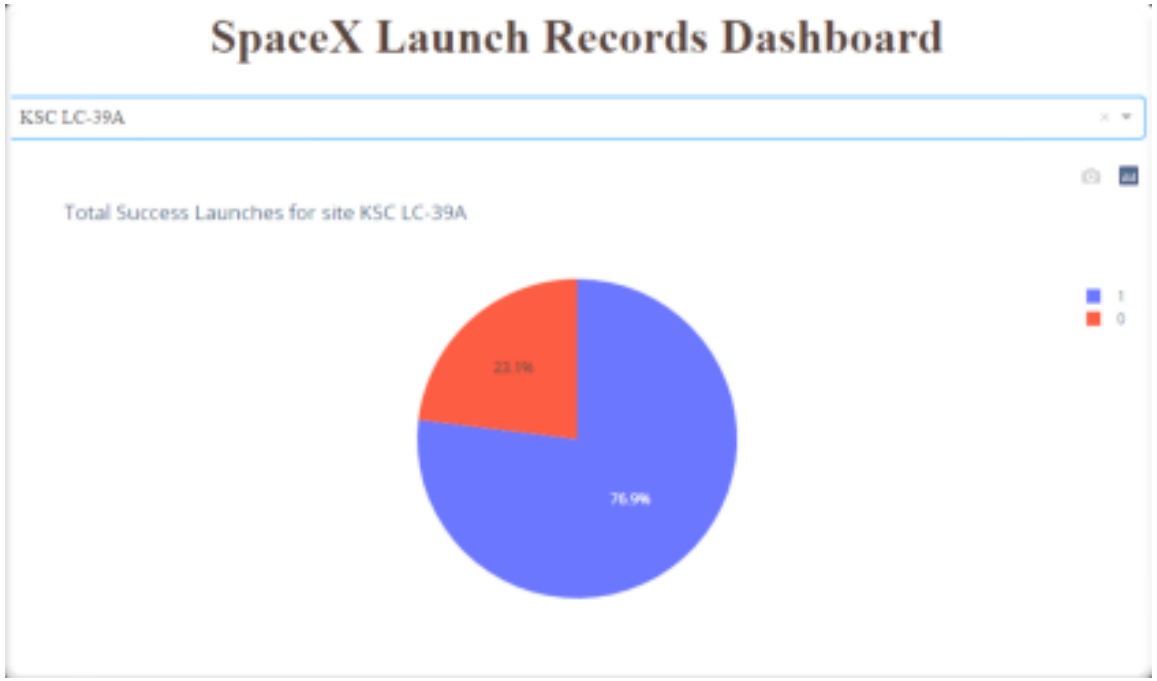
Launch Success Yearly Trend

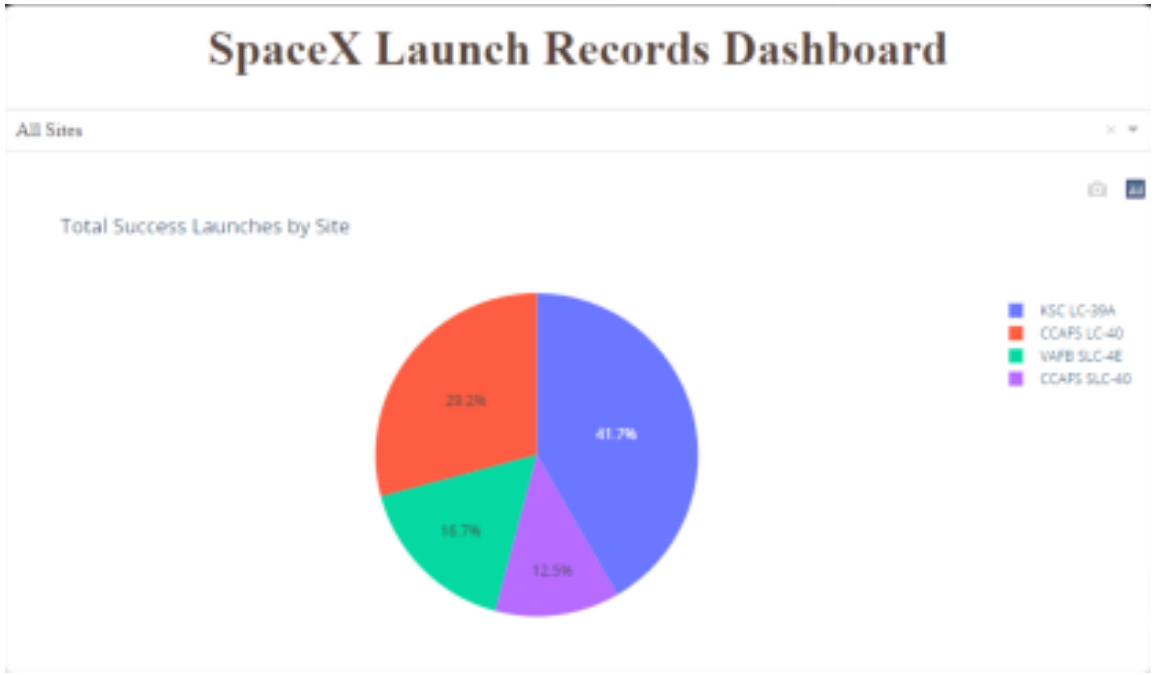


interactive map with Folium results

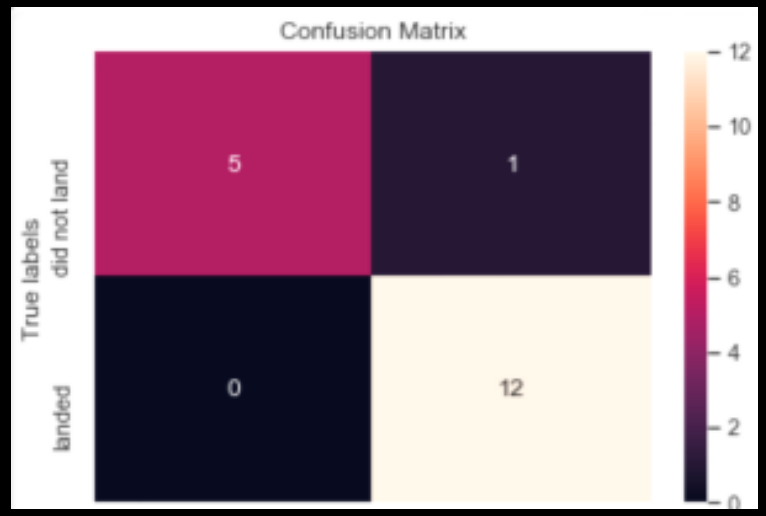


Plotly Dash dashboard results





predictive analysis (classification) results



Algorithm	Accuracy Score	Best Score
Logistic Regression	0.833333	0.846429
Support Vector Machine	0.833333	0.848214
Decision Tree	0.944444	0.903571
K Nearest Neighbours	0.888889	0.876786

Conclusions

As the number of flights increases, the rate of success at a launch site increases, with most early flights being unsuccessful. I.e. with more experience, the success rate increases.

The success for massive payloads (over 4000kg) is lower than that for low payloads.

The best performing classification model is the Decision Tree model, with an accuracy of 94.44%