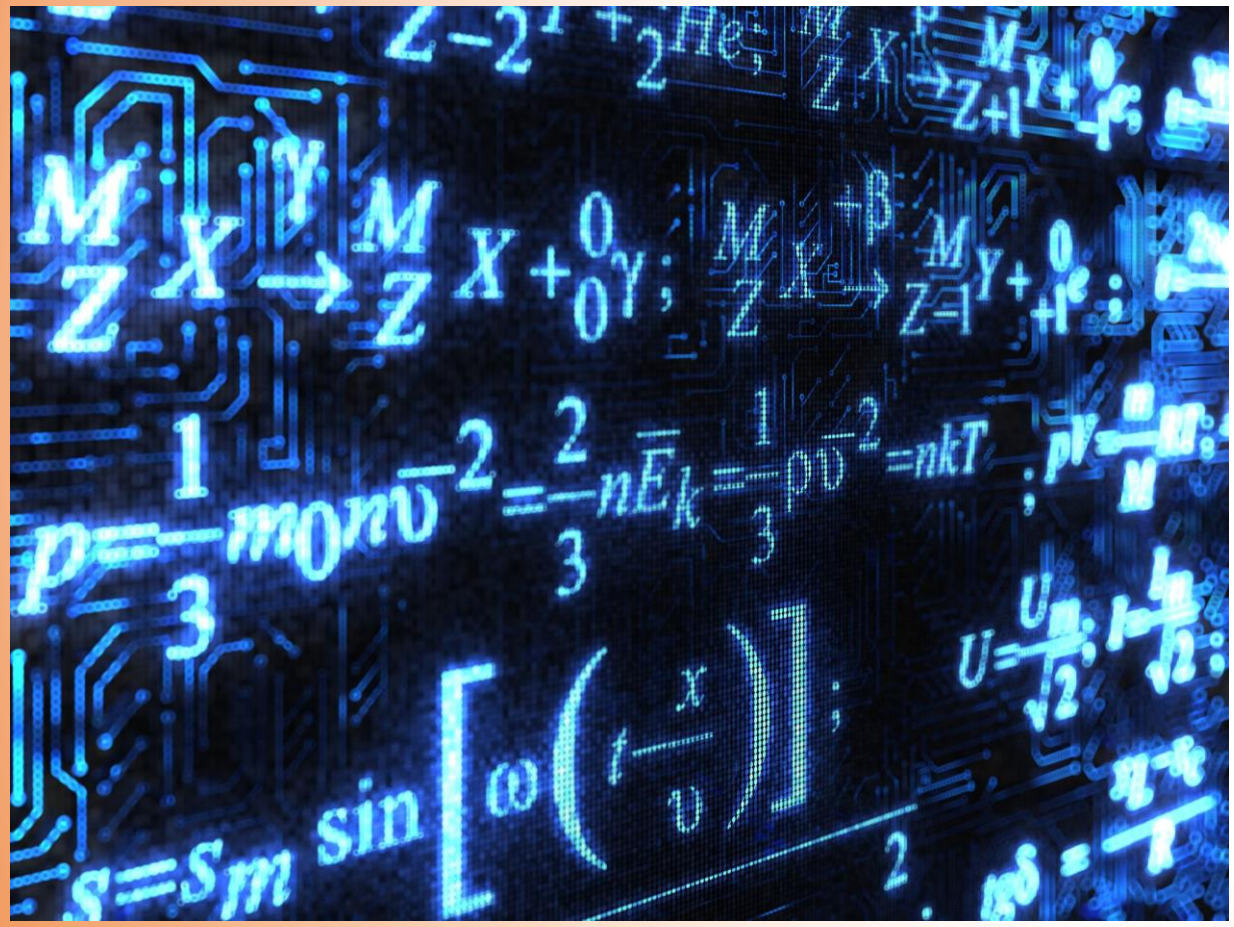
Data Science Capstone Project

William Bottazzi
November 18, 2023

OUTLINE



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

EXECUTIVE SUMMARY

Summary of Methodologies

- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an Interactive Map with Folium
- Building a Dashboard with Plotly Dash
- Predictive Analysis (Classification)

Summary of All Results

- Exploratory Data Analysis Results
- Interactive Analytics
- Predictive Analytics

INTRODUCTION

Project Background and Context

SpaceX advertises the Falcon 9 rocket launches at a cost of 62 million dollars; while other providers have a cost upwards of 165 million dollars, much of the savings comes from the ability to reuse the first stage.

Question to be answered

For a given set of features will the first stage of the rocket land successfully?

METHODOLOGY

- Data Collection
  - SpaceX API
  - Web Scraping
- Data Wrangling
  - Filter Data
  - One Hot Encoding
- Perform Exploratory Data Analysis (EDA) Using Visualization and SQL
- Perform Predictive Analysis Using Classification Models

DATA COLLECTION – SPACEX API

- Request the SpaceX launch data using the SpaceX API
- Normalize the returned JSON file and create a Pandas Data frame
- Parse the Data frame to only include the Falcon9 launches
- Replace missing values with the mean value

DATA COLLECTION – WEB SCRAPING

- Request the Falcon9 Launch HTML
- Using Beautiful Soup extract all the tables
- Create a data frame by parsing the launch HTML tables

DATA WRANGLING

1. **Discovery**

   In the discovery step you think about the questions and the types of data you will need to answer them. Also, you will figure out how to clean, structure and organize the data in the following stages.

2. **Transformation**

   During the transformation step you will structure, normalize and clean the data that you discovered.

3. **Validation**

   During the validation step you will check the work done during the transformation step to insure consistent and quality data.

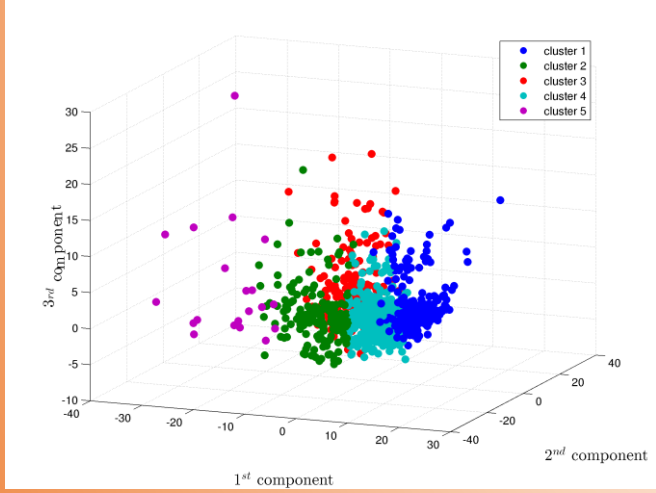
4. **Publishing**

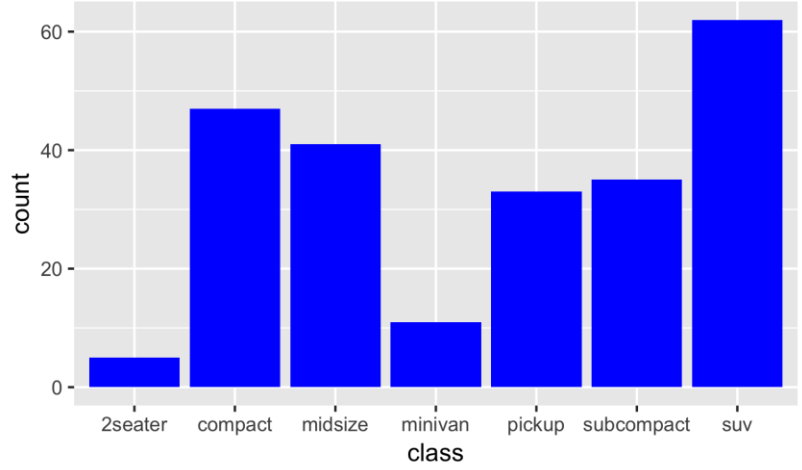
   After finishing the validation step you will publish your data using whatever file format you prefer to share.

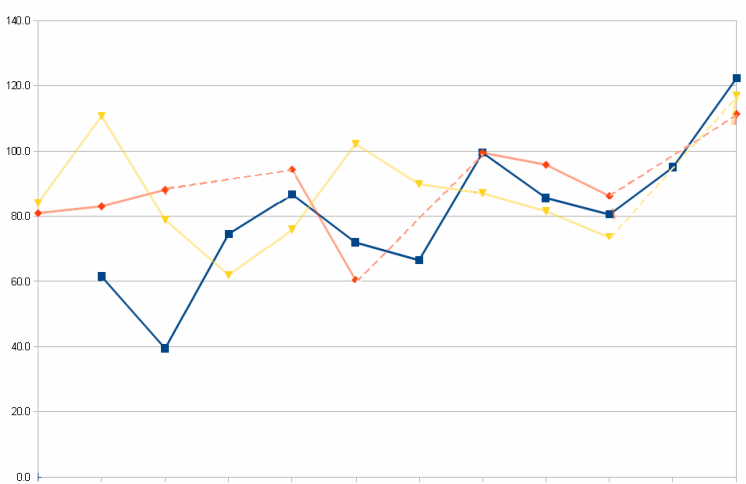
EDA WITH DATA VISUALIZATION

Exploratory Data Analysis was performed using variables including; Flight Number, Payload Mass, Launch Site, Orbit Class and Year.

Scatter plots, bar charts and line charts were used to compare the relationships with the above variables.







EDA WITH SQL

Performed queries to a database to get a better understanding of the dataset.

Queries to the database included information on launch site names, mission outcomes, payload sizes and booster versions.

INTERACTIVE MAP USING FOLIUM

An interactive map was created using Folium.

Markers were added for all of the launch sites using their latitude and longitude coordinates.

Colored markers were added for the launch outcomes. Green for success and red for fails.

Colored lines were added to show proximities from a launch site to the coastline, closest city and a railway

PREDICTIVE ANALYSIS

Build a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.

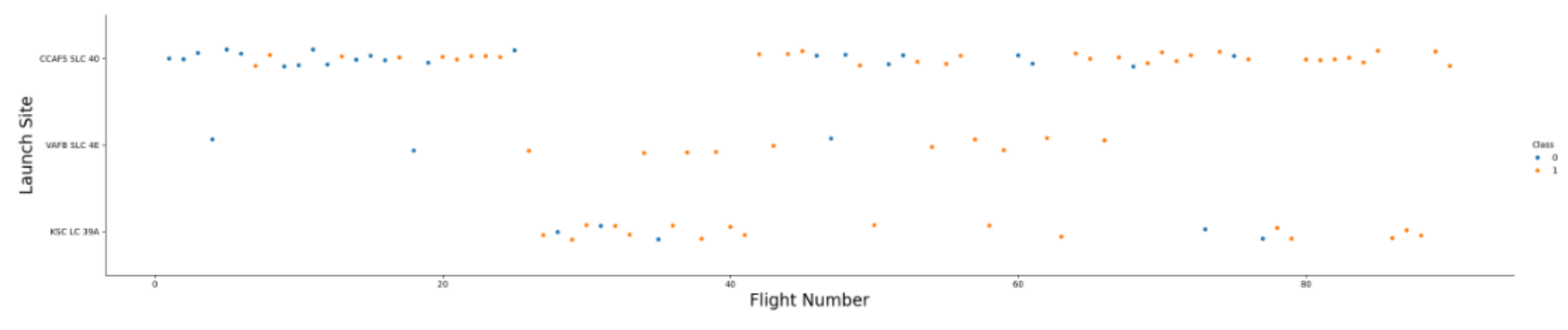
- Preprocessing will allow the data to be standardized
- Splitting the data into training and testing data
- Train the model and perform a Grid Search
  - This will find the hyperparameters
- Using the best hyperparameter values we will find the model with the best accuracy
- Test Logic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors
- Output the confusion matrix

RESULTS

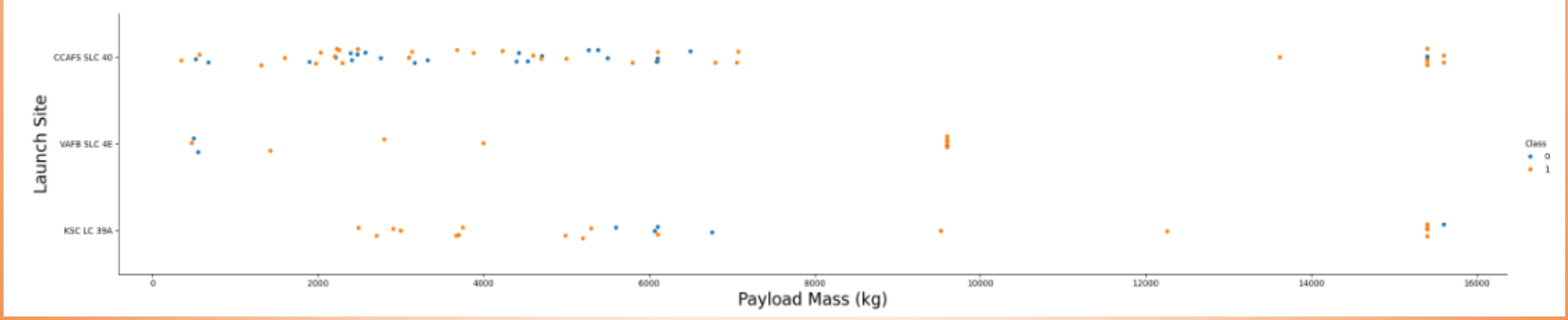
- EDA with Visualization
- EDA with SQL
- Interactive map with Folium
- Plotly Dashboard
- Predictive analysis (Classification)

EDA WITH VISUALIZATION

The relationship between Flight Number and Launch Site

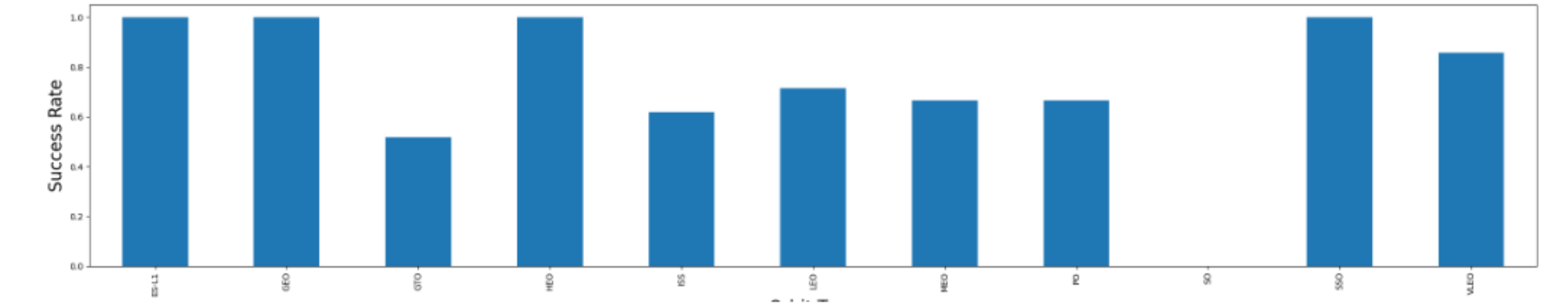


The relationship between Payload and Launch Site

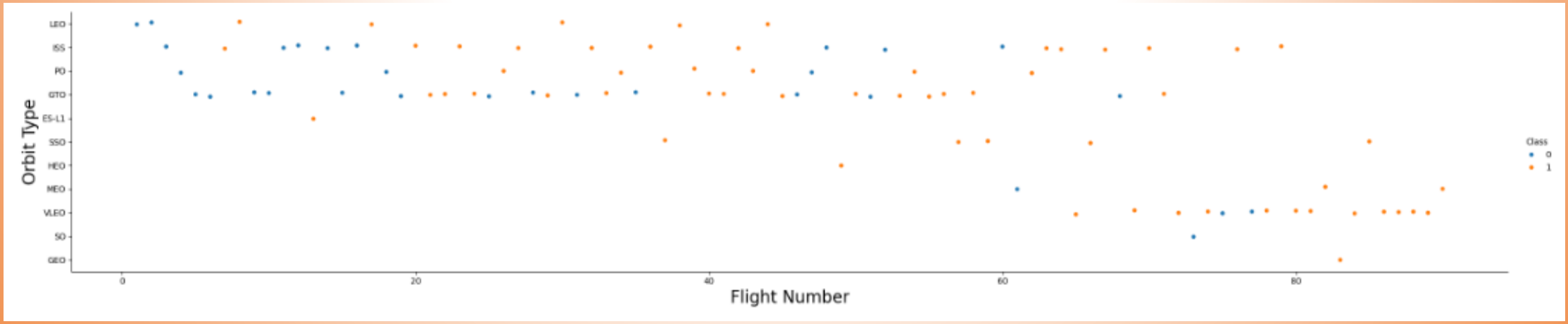


EDA WITH VISUALIZATION CONT.

The relationship between success rate of each orbit type

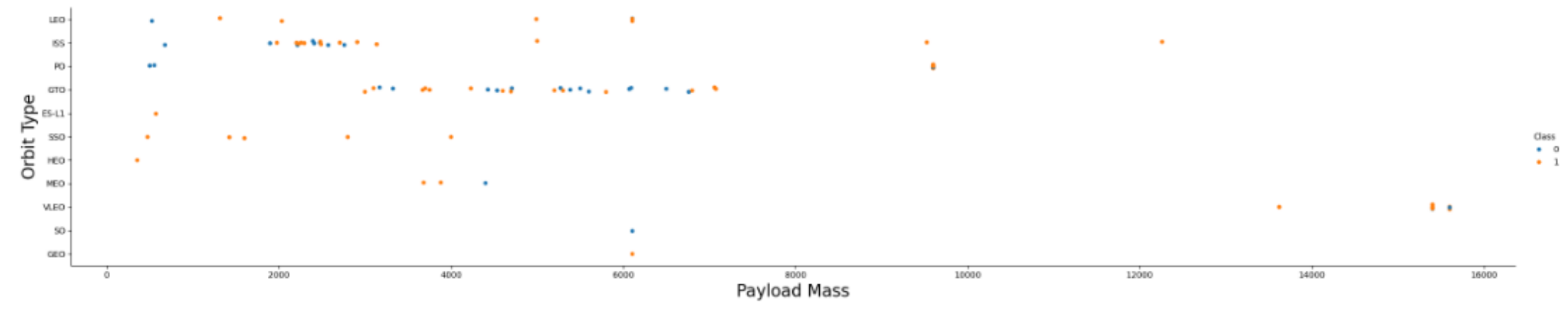


The relationship between Flight Number and Orbit type

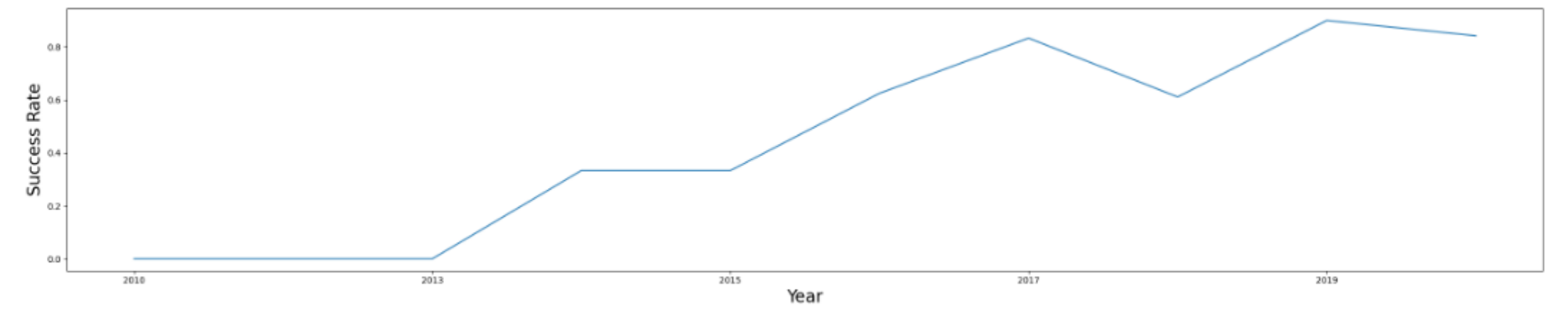


EDA WITH VISUALIZATION CONT.

The relationship between Payload and Orbit type



The launch success yearly trend



EDA WITH SQL

Display the names of the unique launch sites in the space mission

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Display 5 records where launch sites begin with the string 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

EDA WITH SQL CONT.

Display the total payload mass carried by boosters launched by NASA (CRS)

SUM(PAYLOAD_MASS__KG_)
45596

Display 5 records where launch sites begin with the string 'CCA'

AVG(PAYLOAD_MASS__KG_)
2928.4

List the date when the first successful landing outcome in ground pad was achieved.

MIN(DATE)
2015-12-22

EDA WITH SQL CONT.

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

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

List the total number of successful and failure mission outcomes

Successful Missions	Failed Missions
100	1

EDA WITH SQL CONT.

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

List the names of the boosters which have less than 6000

List the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.

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

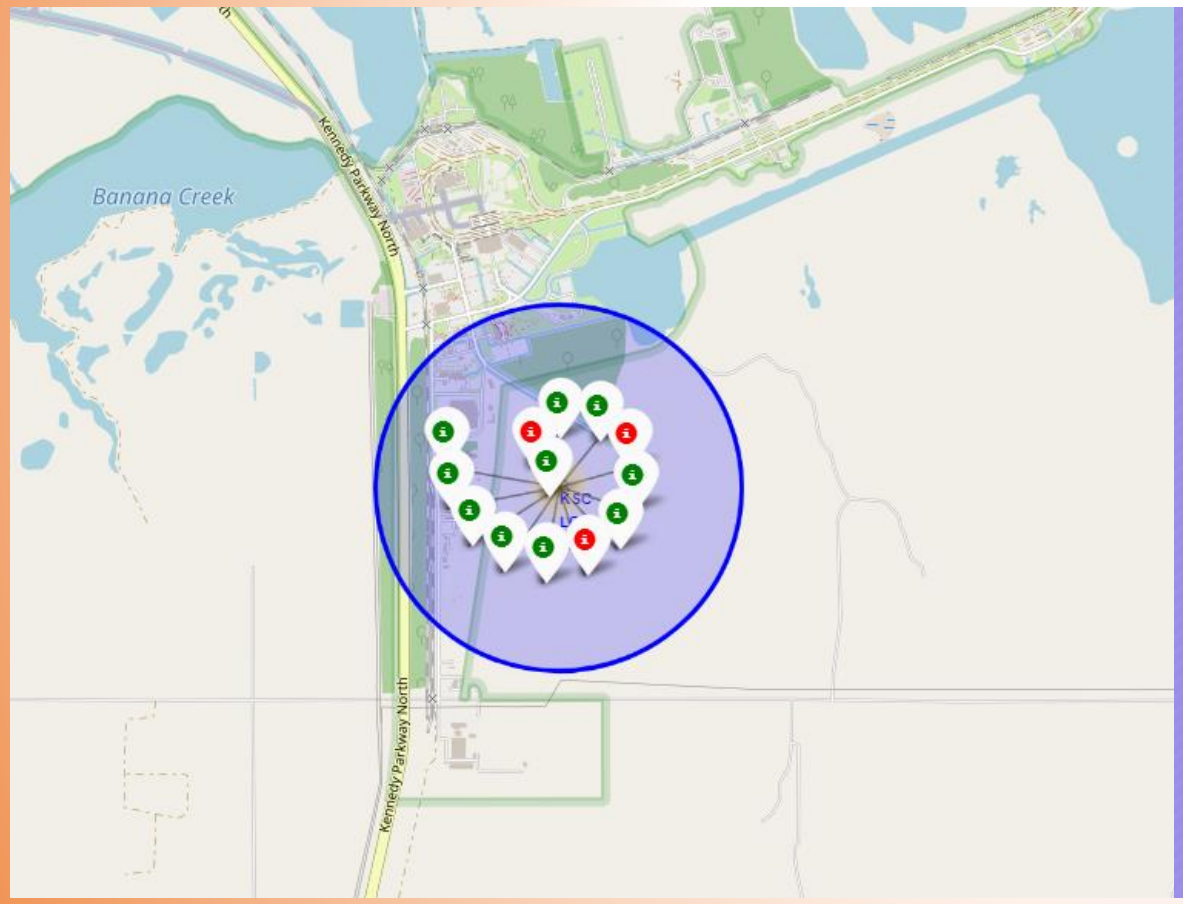
EDA WITH SQL CONT.

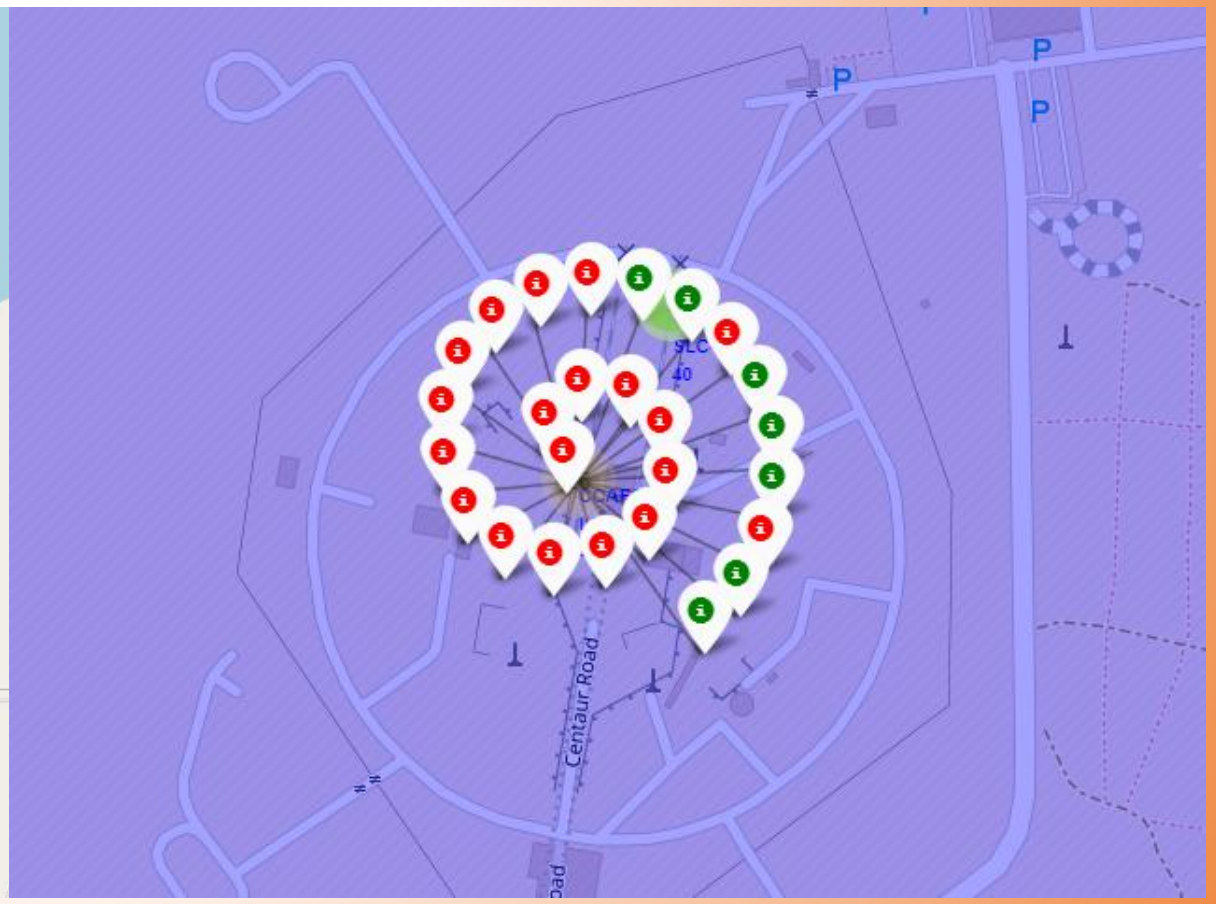
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Landing_Outcome	Total_Outcomes
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

INTERACTIVE MAP WITH FOLIUM

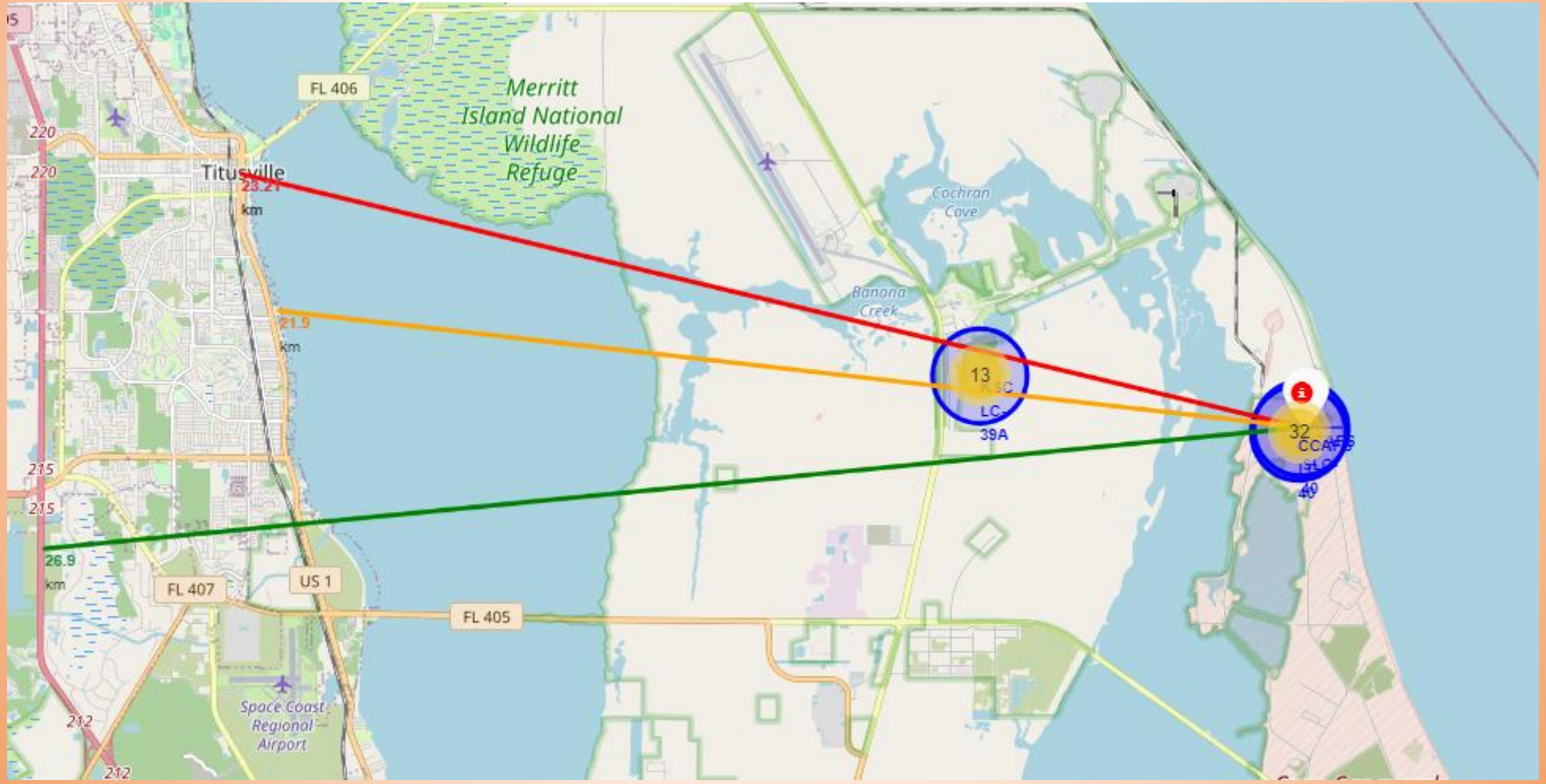
The success/failed launches for each site





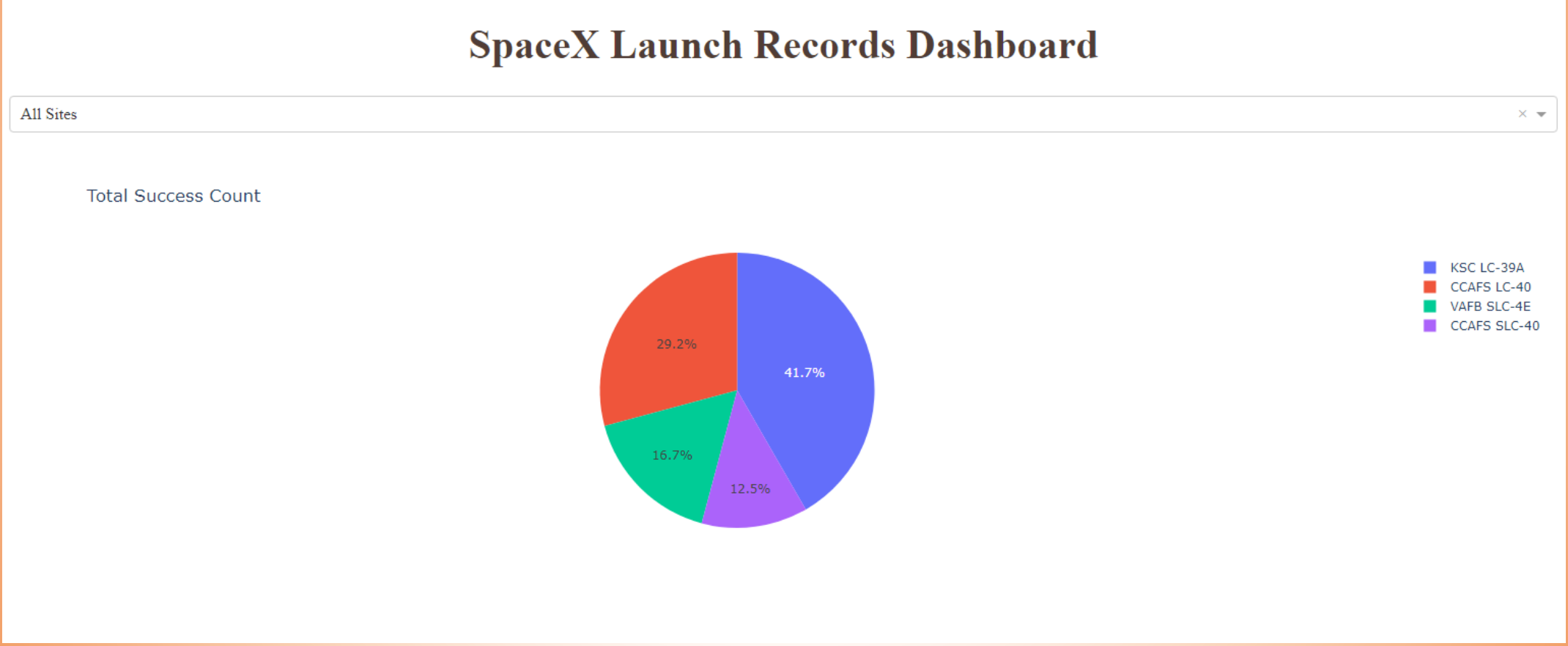
INTERACTIVE MAP WITH FOLIUM

Calculate the distances between a launch site



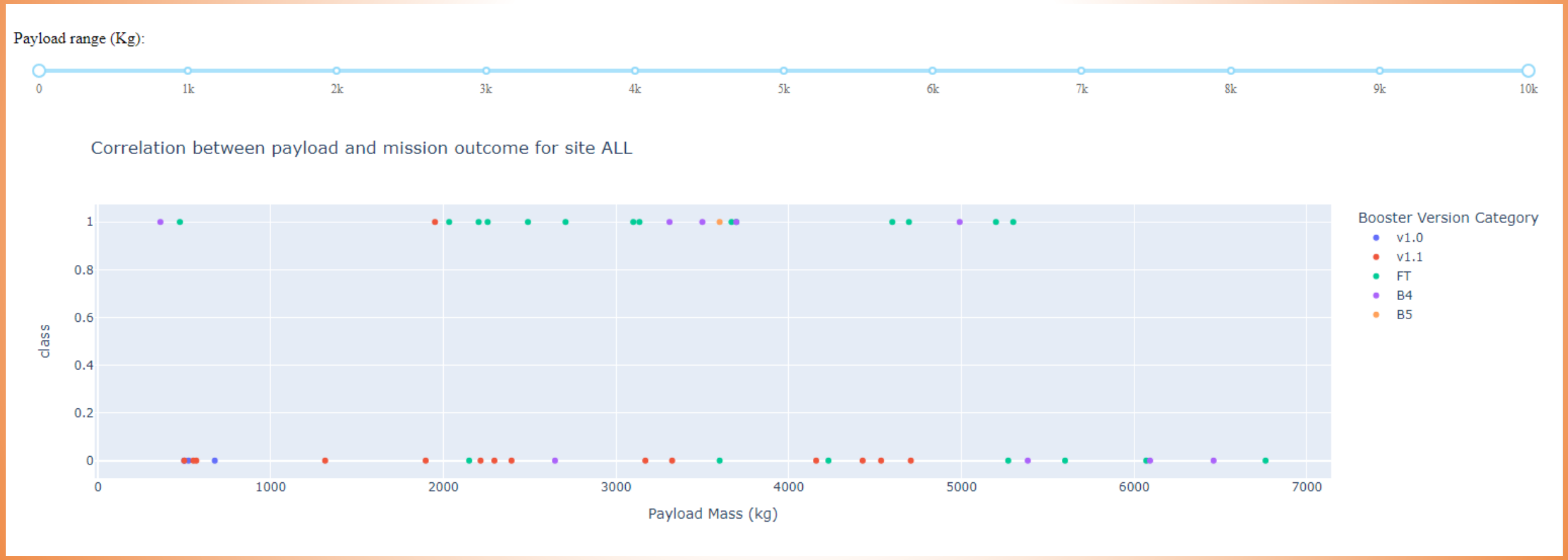
PLOTLY DASH DASHBOARD

Successful launches across all launch sites



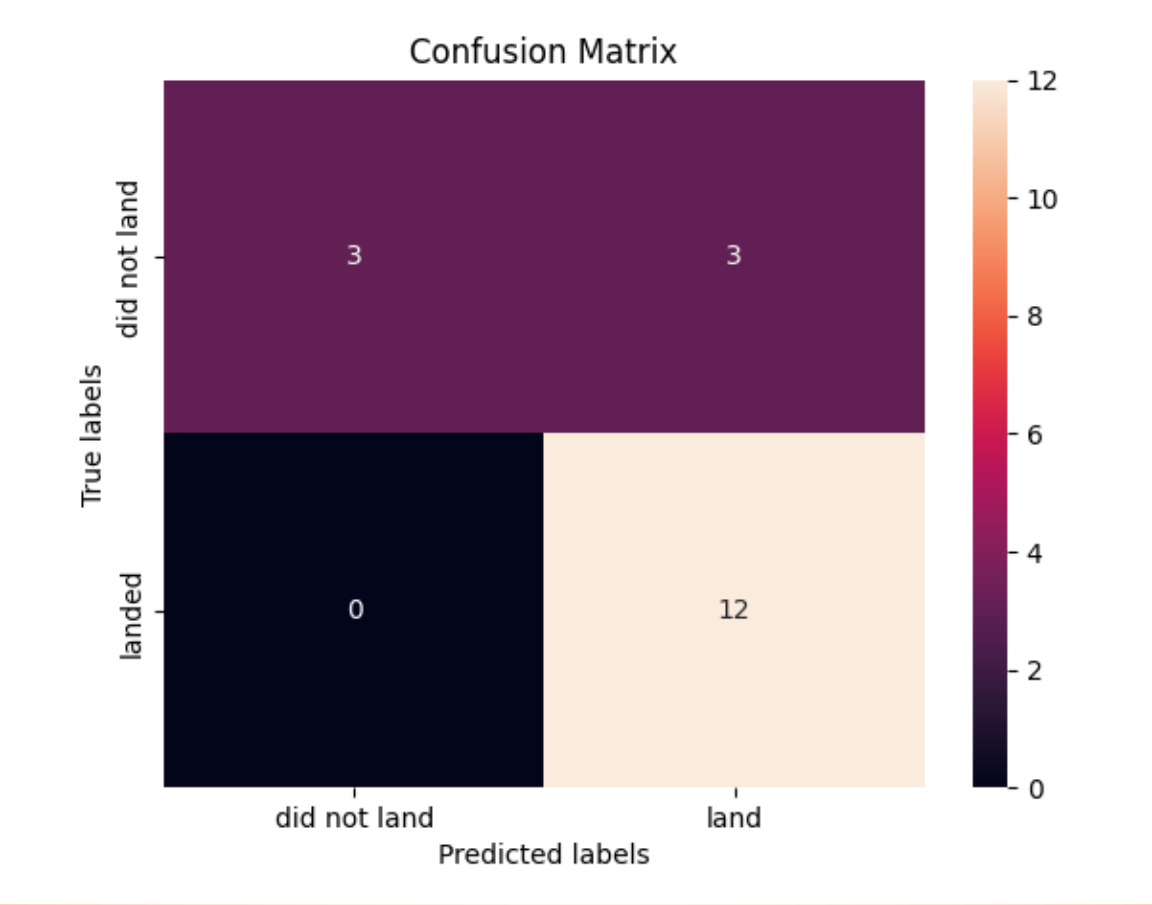
PLOTLY DASH DASHBOARD

Correlation between payload and mission outcome



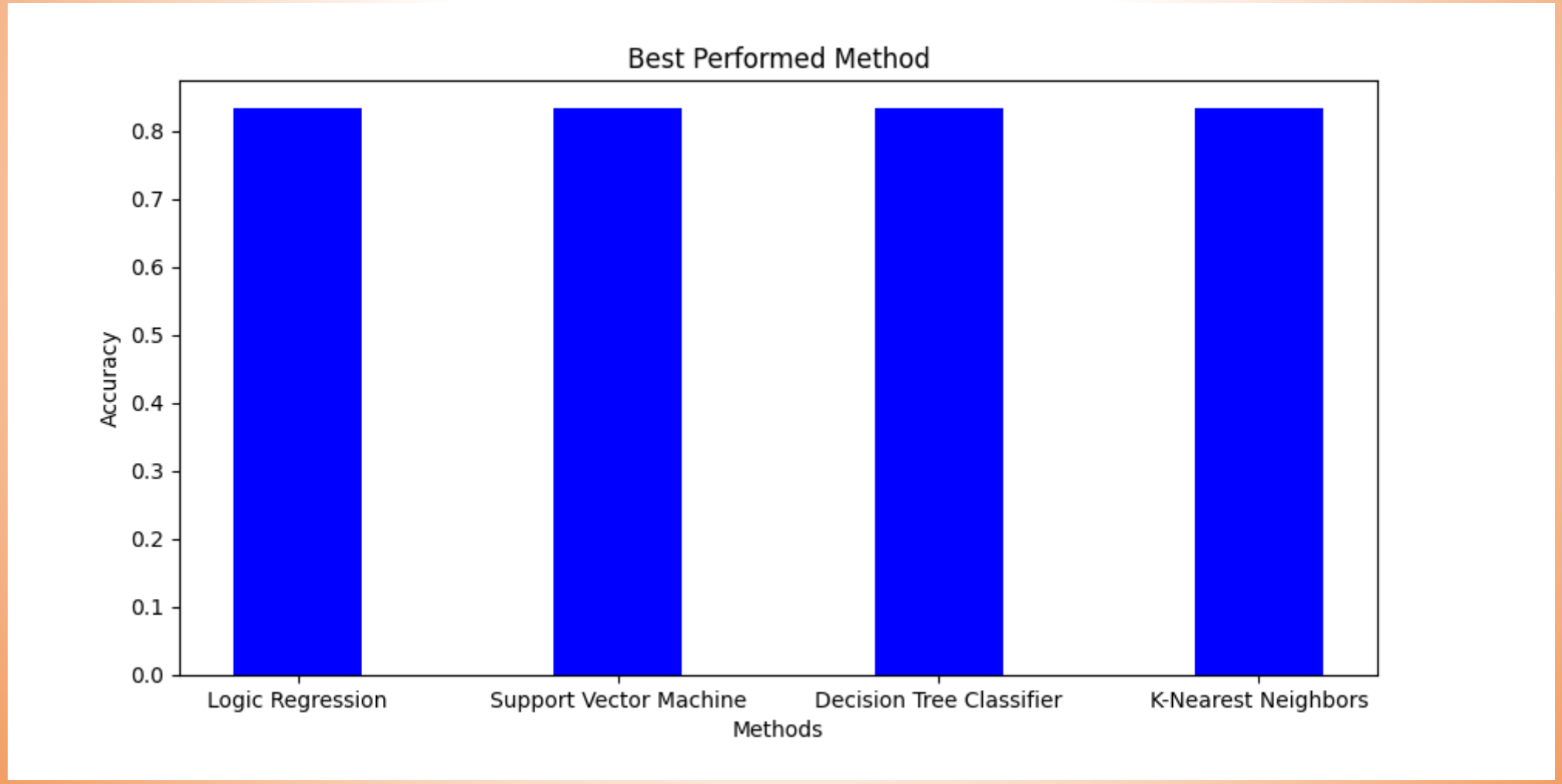
PREDICTIVE ANALYSIS

Confusion Matrix



PREDICTIVE ANALYSIS

All models have similar results



CONCLUSION

All of the models have similar results with an accuracy of around 83%. More data may be needed to find a model with better accuracy.

APPENDIX

GitHub URL

https://github.com/tazs4wheelin/Data-Science-Capstone