# Applied Data Science Capstone project

# Outline

- Executive summary
- Introduction
- Methodology
- Results
- Conclusion

# Summary of methodologies

- Data Collection through API
- Data Collection with Web Scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Data Visualization
- Interactive Visual Analytics with Folium
- Machine Learning Prediction

# Summary of all results

- Exploratory Data Analysis result
- Interactive analytics in screenshots
- Predictive Analytics result

# Introduction

- **Project background and context**

  Space X 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 Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch. The goal of the project is to create a machine-learning pipeline to predict if the first stage will land successfully.

- **Problems you want to find answers**

  - What factors determine if the rocket will land successfully?
  - The interaction amongst various features that determine the success rate of a successful landing.
  - What operating conditions need to be in place to ensure a successful landing program?

# Methodology

- Data collection
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

# Data collection

- Data collection was done using get request to the SpaceX API.
- Next, the response was decoded using.json() function.
- Then return into a pandas data frame using .json_normalize().
- Cleaned data and checked missing values by filling in the missing values.
- Web scraping was also performed using Wikipedia for the Falcon 9 launch.

# Data collection-SpaceX API

Link to the notebook

https://github.com/freteka/SpaceX_project/blob/main/Space%20X_Data%20Collection.ipynb

# Data collection-Scraping

- Falcon9 was launched using BeautifulSoup
- The link

https://github.com/freteka/SpaceX_project/blob/main/Space%20X_WebScraping.ipynb

# Data Wrangling

- The number of launches were calculated at each site, and the number and occurrence of each orbits.
- landing outcomes were created from the outcome column.
- Exported the results to CSV.
- The link

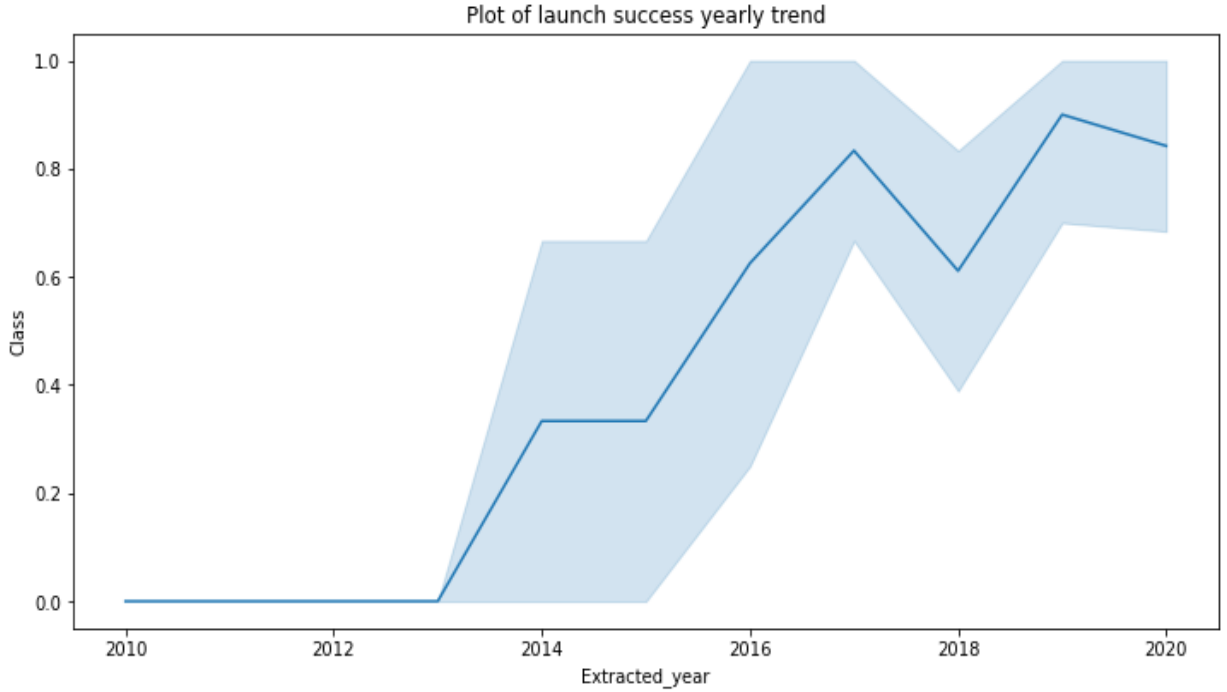
https://github.com/freteka/SpaceX_project/blob/main/SpaceX_Data%20Wrangling.ipynb

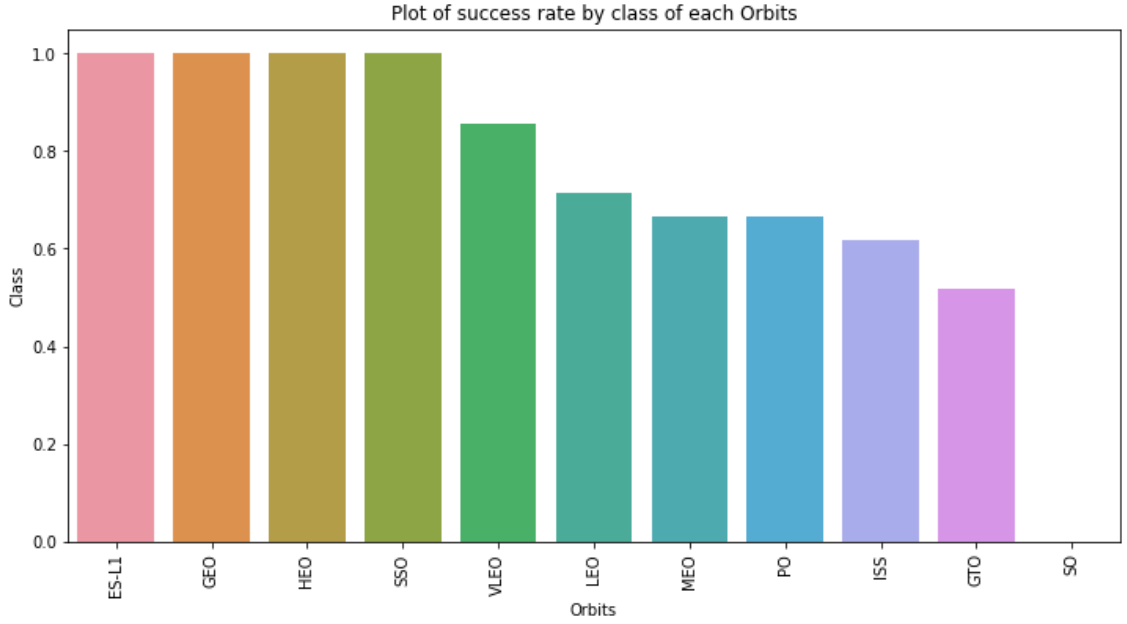
# EDA with data Visualization

- Data was explored by visualizing the relationship between flight number and launch Site, payload and launch site, a success rate of each orbit type, flight number and orbit type, and the launch success yearly trend.

- Link

https://github.com/freteka/SpaceX_project/blob/main/SpaceX_eda%20DataViz.ipynb



# EDA with SQL

- The names of unique launch sites in space missions.
- The total payload mass carried by booster launched by NASA.
- The average payload mass carried by booster version F9.
- The total number of successful and failed mission outcomes.
- The failed landing outcomes in drone ship, their booster version, and launch site names.
- Link https://github.com/freteka/SpaceX_project/blob/main/SpaceX_edaSQL.ipynb

# Interactive map with Folium

- All marks were launched using a folium map.
- The feature launch outcomes (success or failure) were assigned.
- The distances between launch sites to their proximities were calculated.
- Link https://github.com/freteka/SpaceX_project/blob/main/SpaceX_DataVizFolium.ipynb

# Predictive analysis

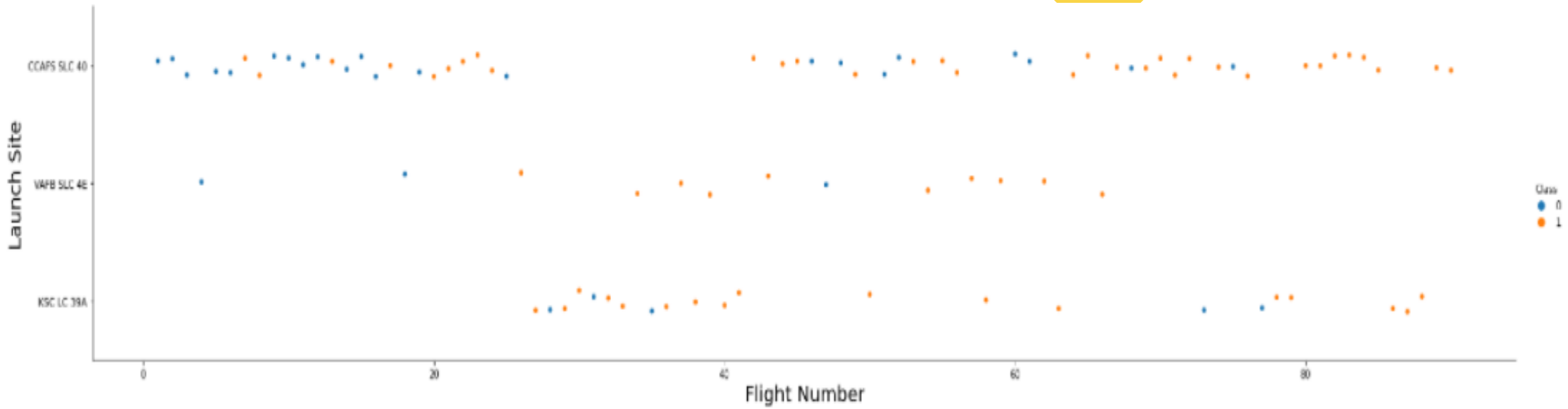
- To improve the model feature engineering and algorithm tunning was used.

- Link https://github.com/freteka/SpaceX_project/blob/main/SpaceX_machine%20learning.ipynb

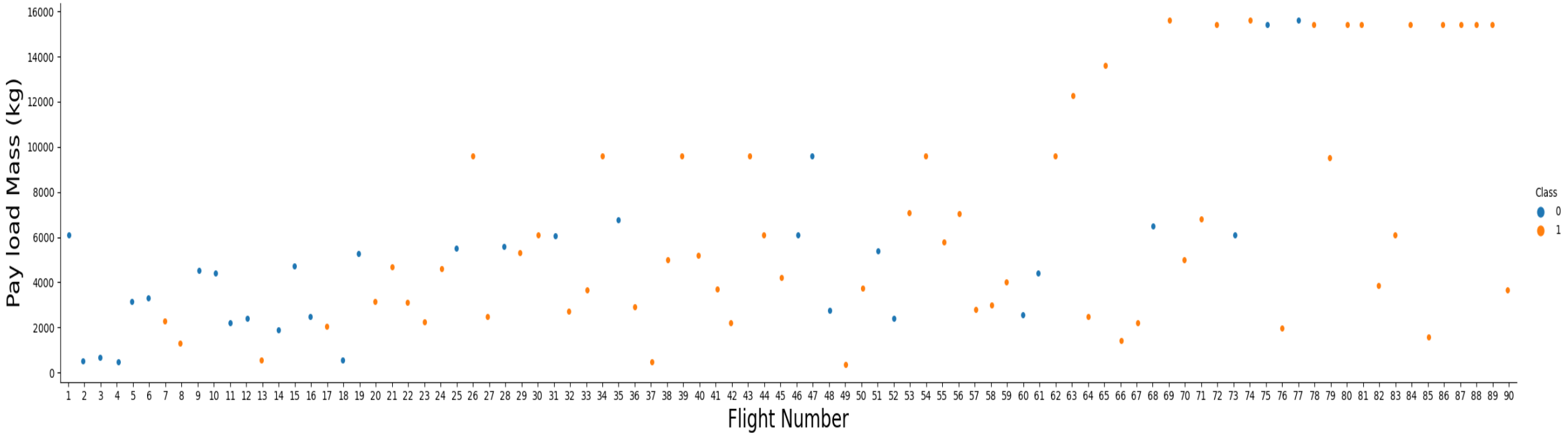
# Results

# Flight number vs Launch site

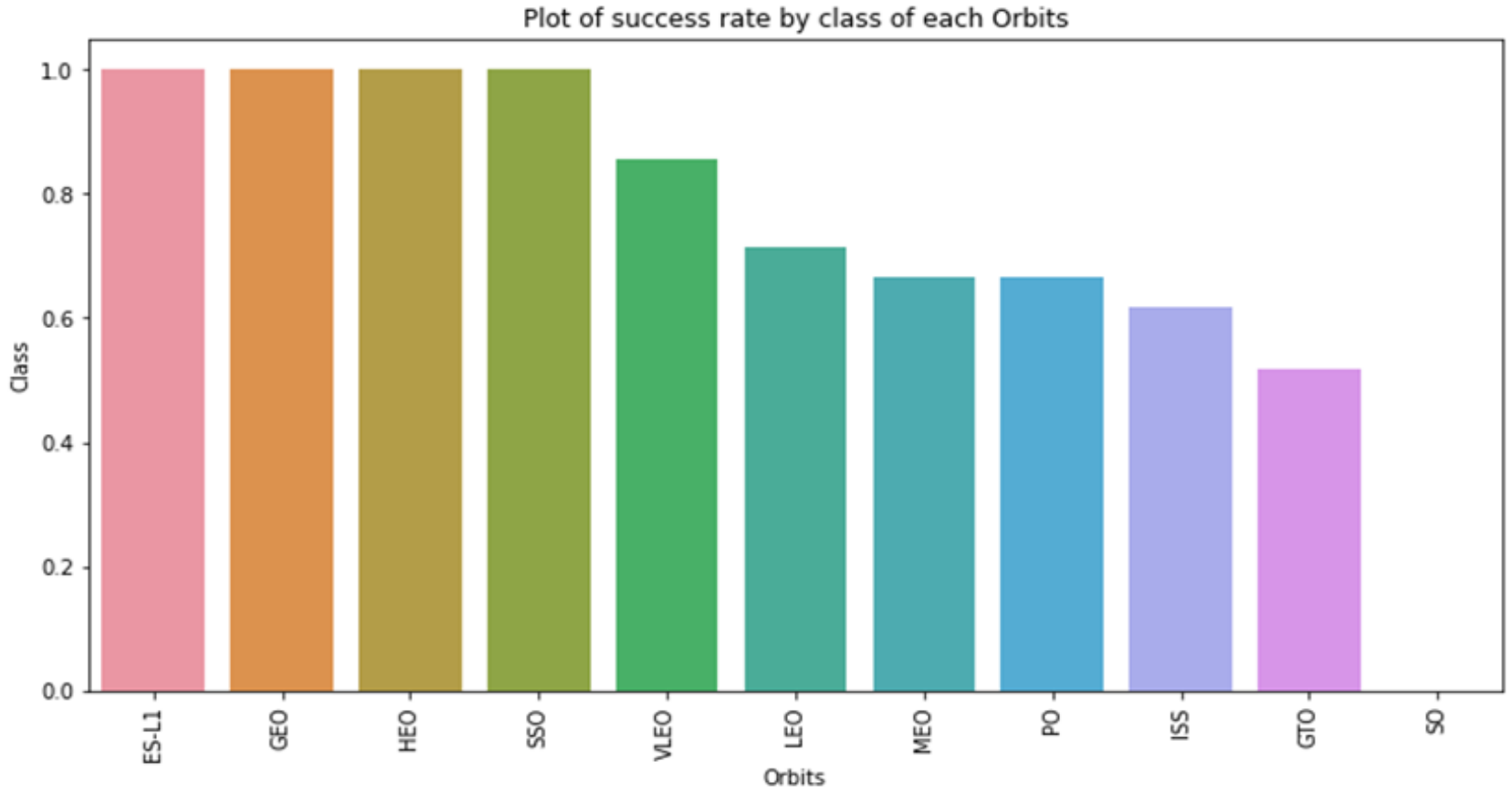
- The greater success rate was found in the launch site



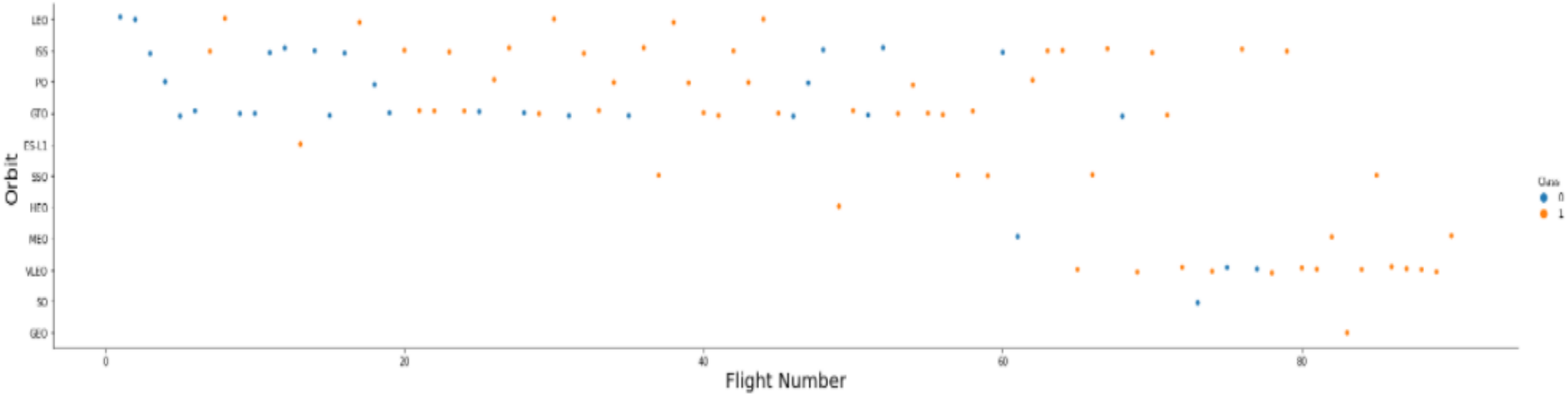
# Payload vs Launch site



# Success rate vs Orbit type



# Flight Number vs Orbit type



# All Launch sites global map markers

# Predictive analysis

## Classification accuracy

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

# Conclusion

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

# Thank you