SpaceX Costing

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

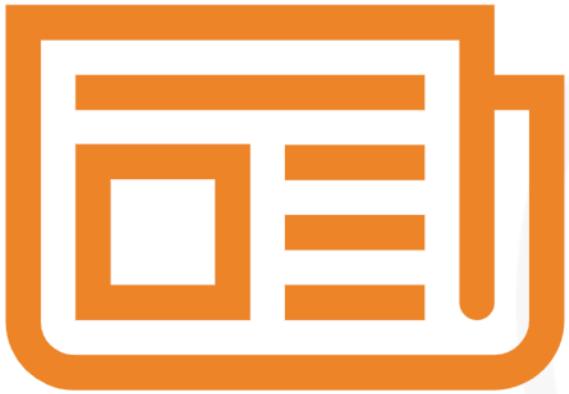
EXECUTIVE SUMMARY

- Data collection
- Data Wrangling
- Exploratory Data Analysis
- Interactive Visual Analytics and Dashboards
- Predictive Analysis
- Explaining Data Insights

INTRODUCTION

- SpaceX is the most successful company in carrying out space exploration programs.
- What makes it more unique is it's low cost operating programs.
- This is achieved by reusing the first stage which makes the total cost around 62 million dollars which is 100 million less than the competitors.
- We focus on the first stage process of this program and derive data insights about the different features to predict success or failure in the space run.

METHODOLOGY



- Data collection methodology:
  - Using SpaceX Rest API
  - Using Web Scrapping from Wikipedia
- Performed data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Building, tuning and evaluation of classification models to ensure the best results

Data Collection

- API:
  - Used SpaceX API
  - Connected with different URL endpoints
  - Extracted json objects
  - Converted into DataFrame
  - Handled missing values

- Data Wrangling:
  - Sourced from Wikipedia
  - Extracted Tables using BeautifulSoup
  - Parsed data
  - Created DataFrame

URL: Data_Collection_API
URL: Data_Collection_WebScraping

EDA

- With SQL:
  - Displayed unique launch sites, total payload mass, average mass, successful landing, failure missions.
  - Counted landing outcomes as Success or Failure for the launch details.

- With Data Visualization:
  - Plotted charts
  - Scatter Plots
  - Showed comparison

URL: EDA_with_SQL

URL: EDA_with_Data_visualization

Visual Analytics and Dashboard

- Folium:
  - Applied markers on launch sites.
  - Colored markers of launch outcomes.
  - Distance between launch site.
- Plotly Dash:
  - Dropdown list of Launch Site
  - Pie Chart of Successful Launches
  - Payload Mass Slider
  - Scatter plot

URL: Folium

URL: Dashboard

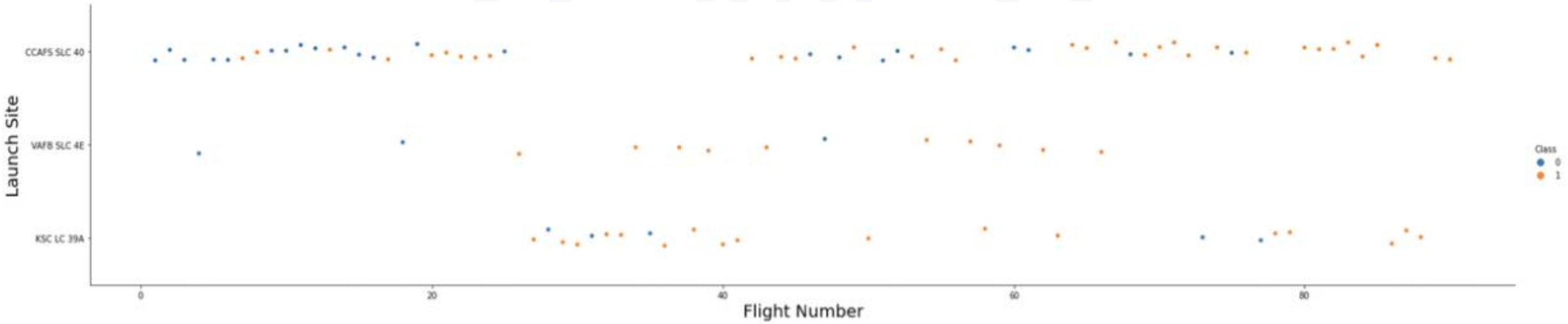
Predictive Analysis

- Classification model:
  - Target variable: Class
  - Standardize data
  - Splitting dataset
  - GridSearchCV
  - Logreg, SVM, DecisionTree, KNN
  - Find accuracy
  - Jaccard distance
  - F1-score

URL: Predictive Analysis

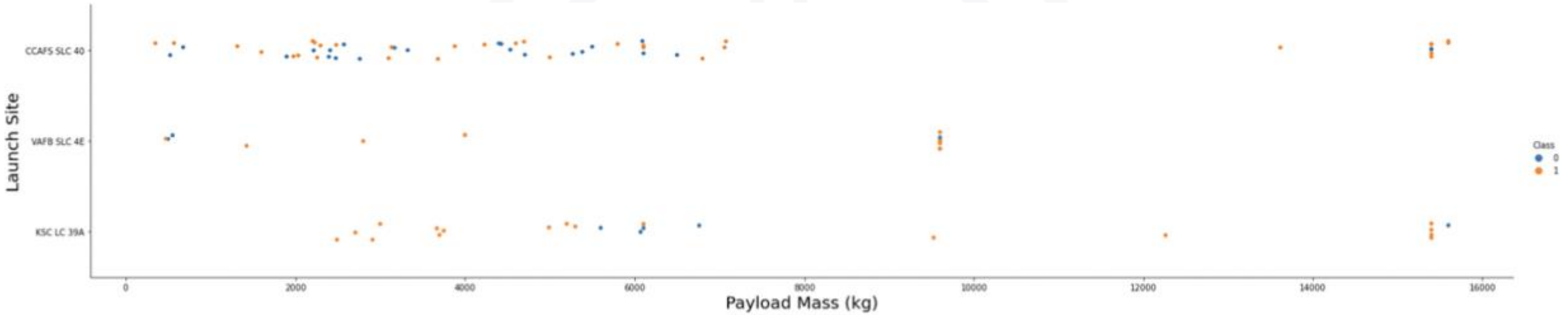
RESULTS

Flight number vs Launch Site



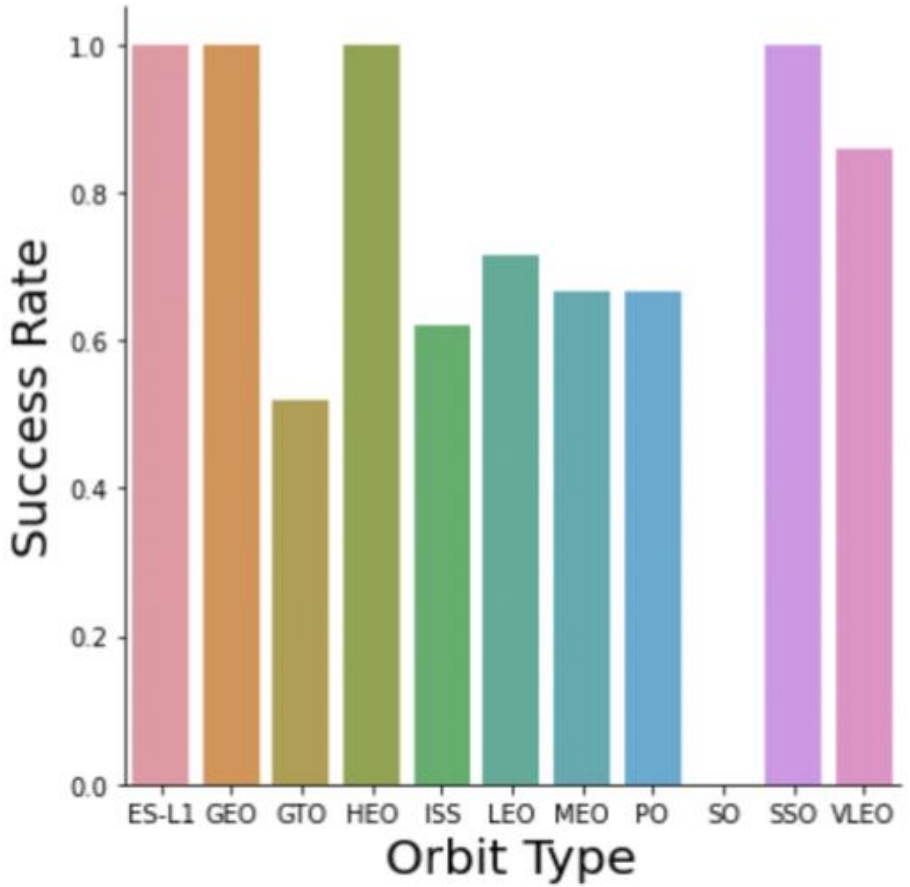
RESULTS

Payload vs Launch Site



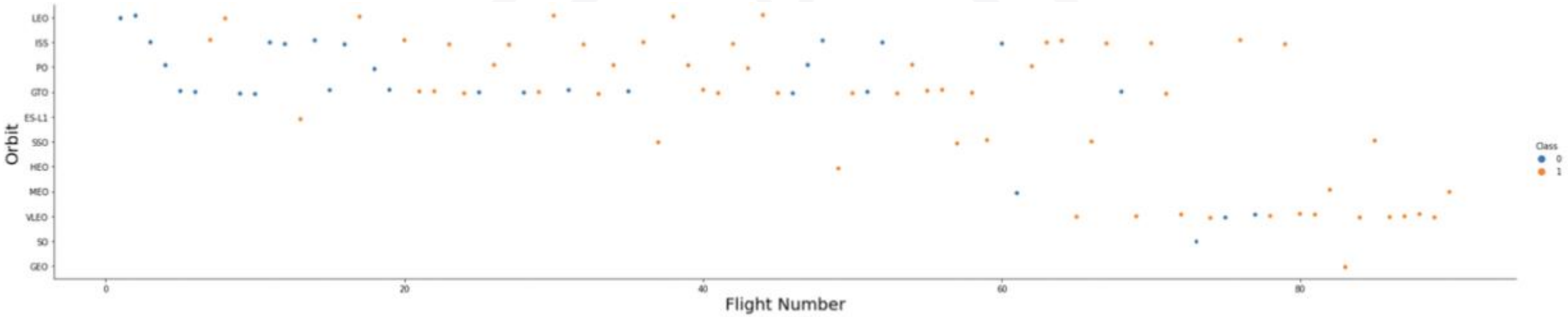
RESULTS

Success rate vs Orbit type



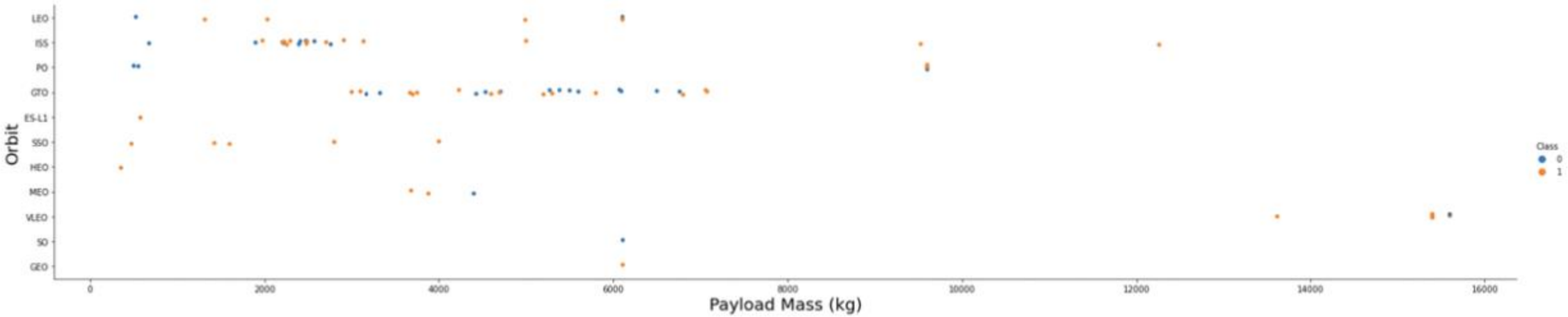
RESULTS

Flight number vs Orbit type



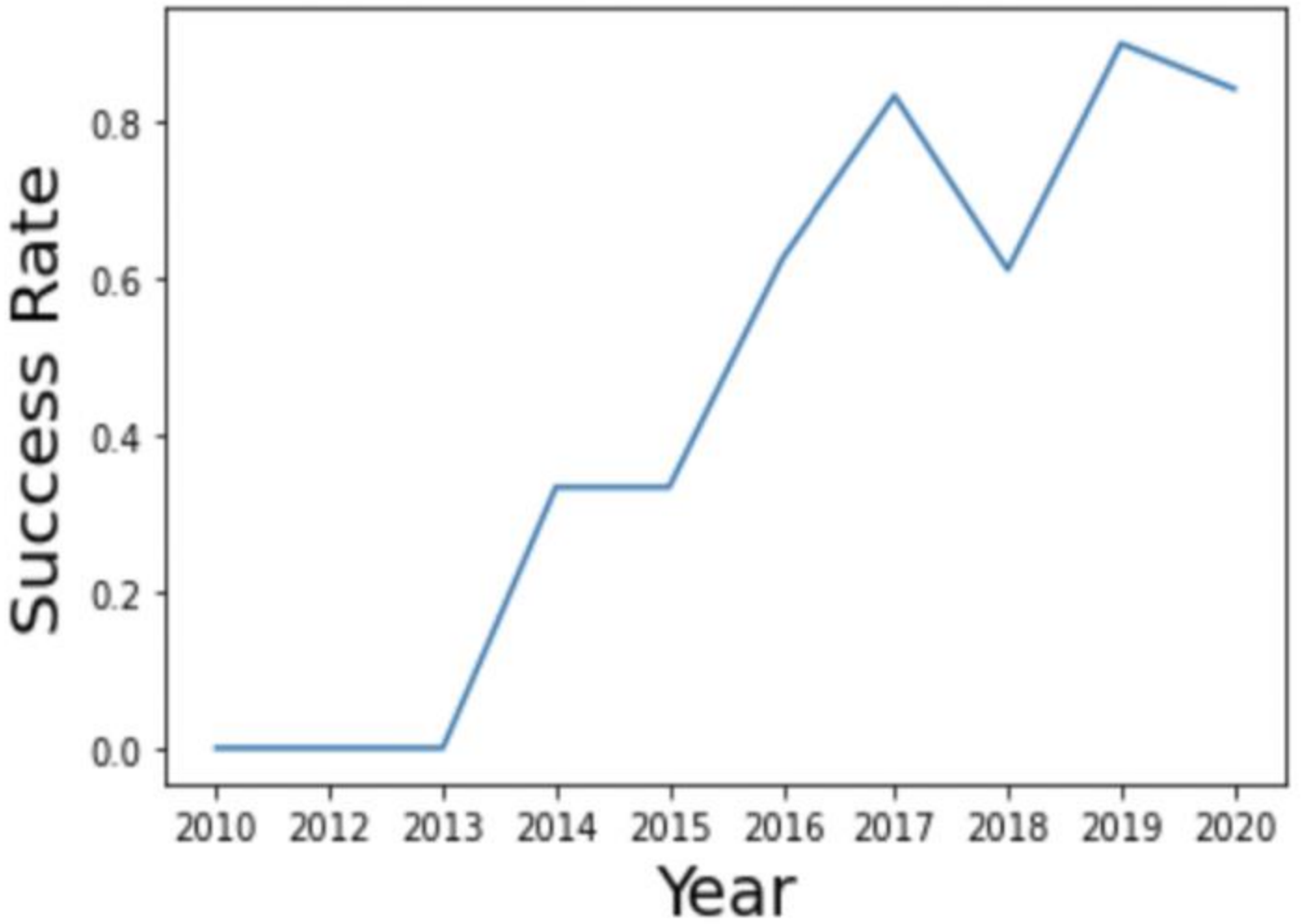
RESULTS

Payload Mass vs Orbit type



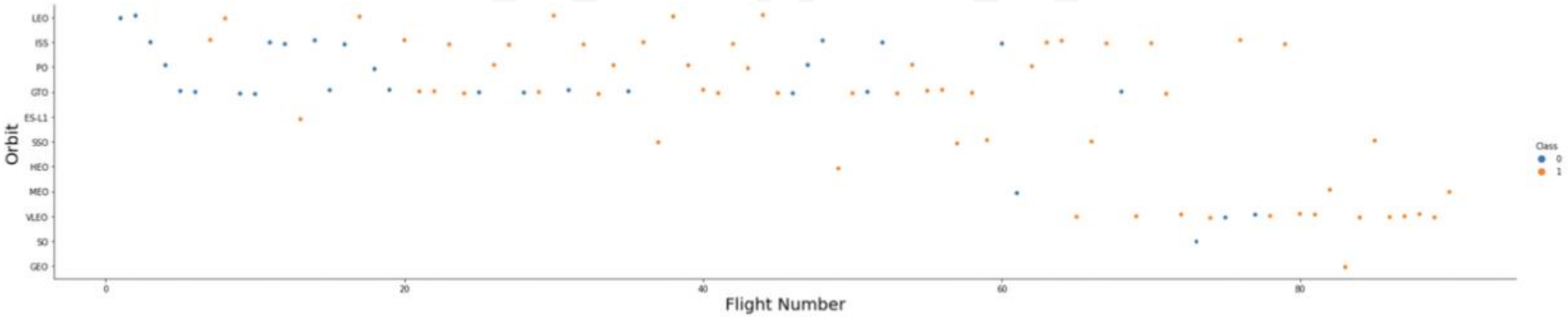
RESULTS

Launches success per year

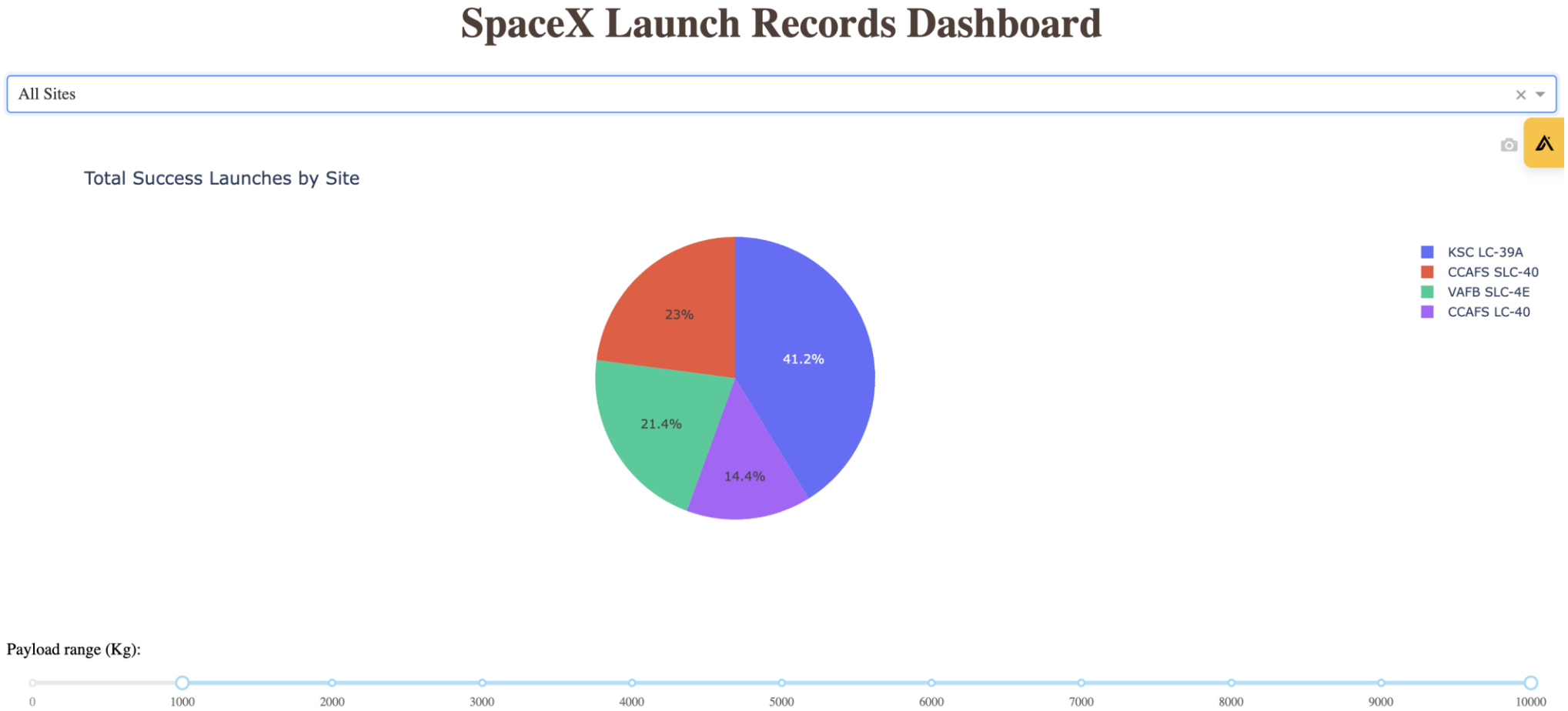


RESULTS

Flight number vs Orbit type



Dashboard



Classification Accuracy

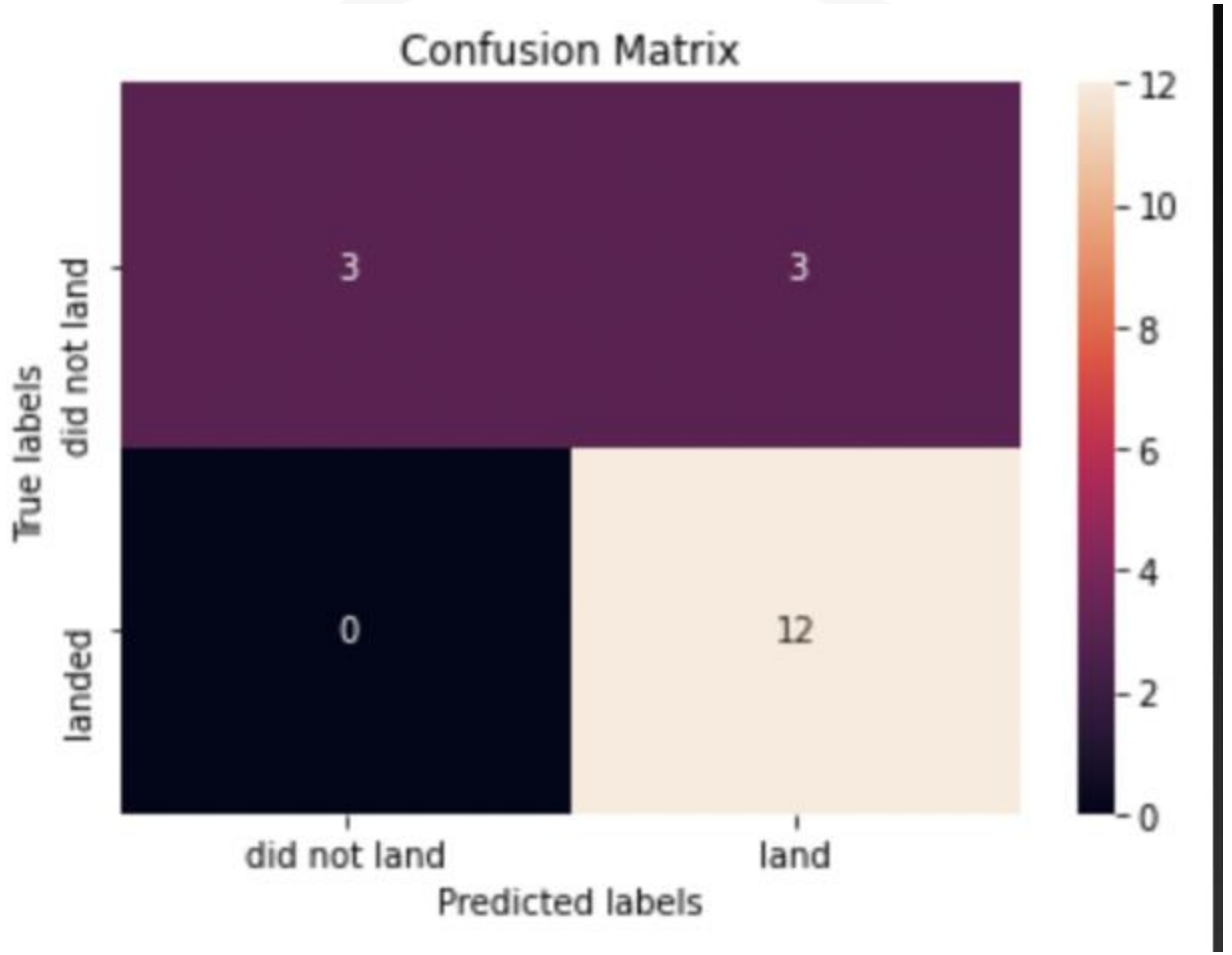
Score and Accuracy of Test set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Score and Accuracy of Entire Data Set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

Confusion Matrix



CONCLUSION

- Decision Tree is the best algorithm fir this dataset.
- Success rate of launches increase over the years.
- Orbits ES-L1, GEO, HEO, SSO have 100% success rate.
- Launch with low payload mass show better results than with higher payload mass.

APPENDIX

Thanks for the instructors, IBM, Coursera.