DATA SCIENCE CAPSTONE PROJECT

Rodrigo Ribeiro Nunes

10/25/2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

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

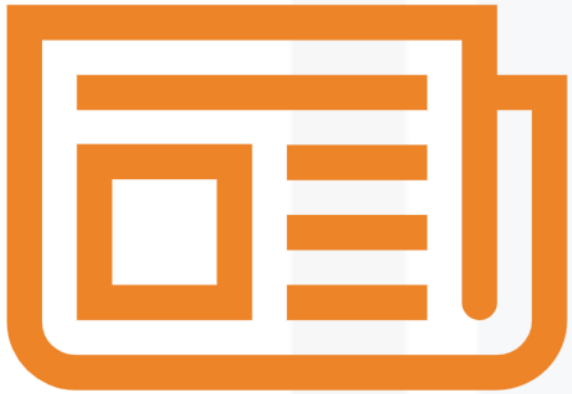
- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

INTRODUCTION

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 needs to be in place to ensure a successful landing program.

METHODOLOGY



Data collection methodology: Data was collected using SpaceX API and web scraping from Wikipedia.

Perform data wrangling: One-hot encoding was applied to categorical features

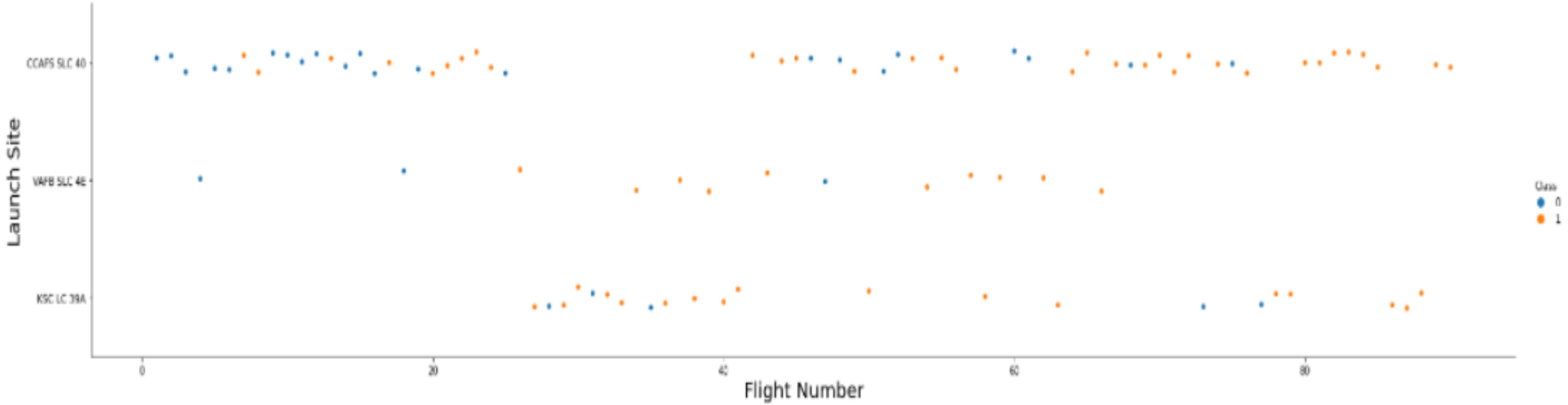
Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification model

RESULTS

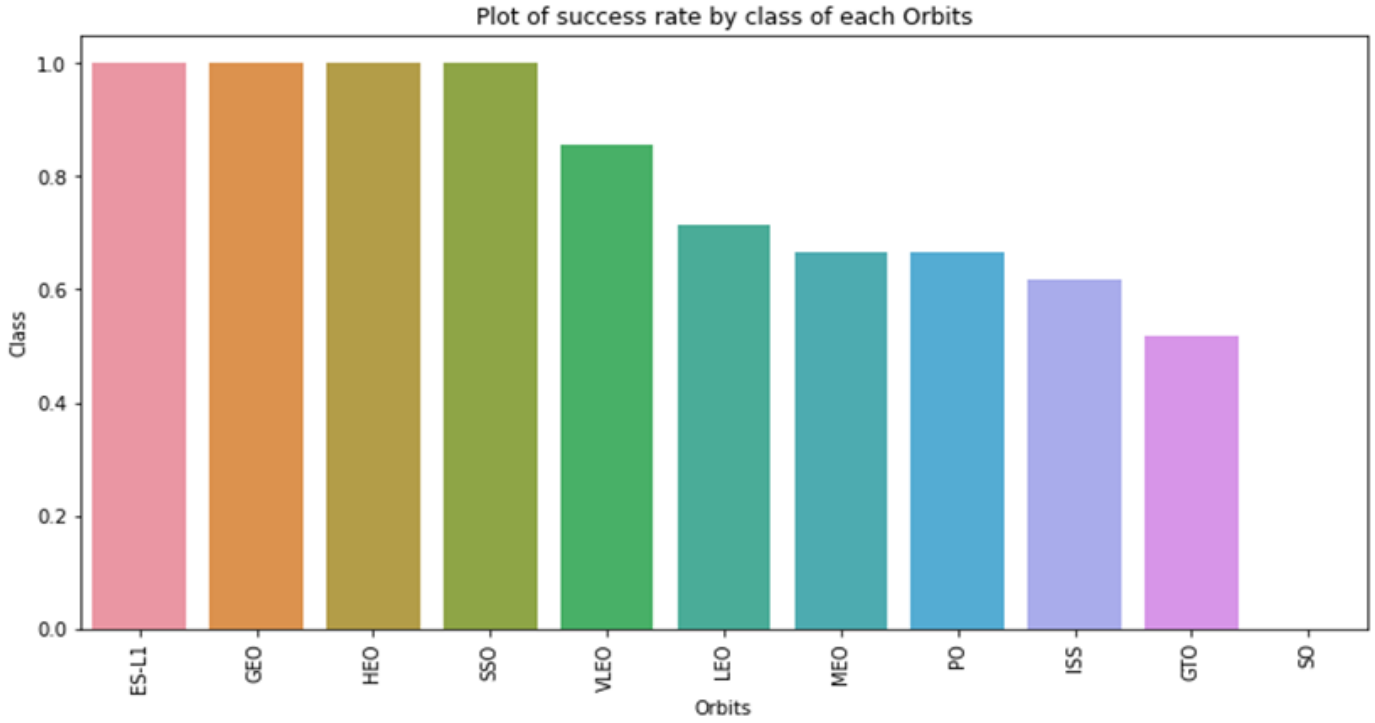
Flight Number X Launch Site



- We can see that the larger the Flight Number, the greater the success in the Launch Site

RESULTS

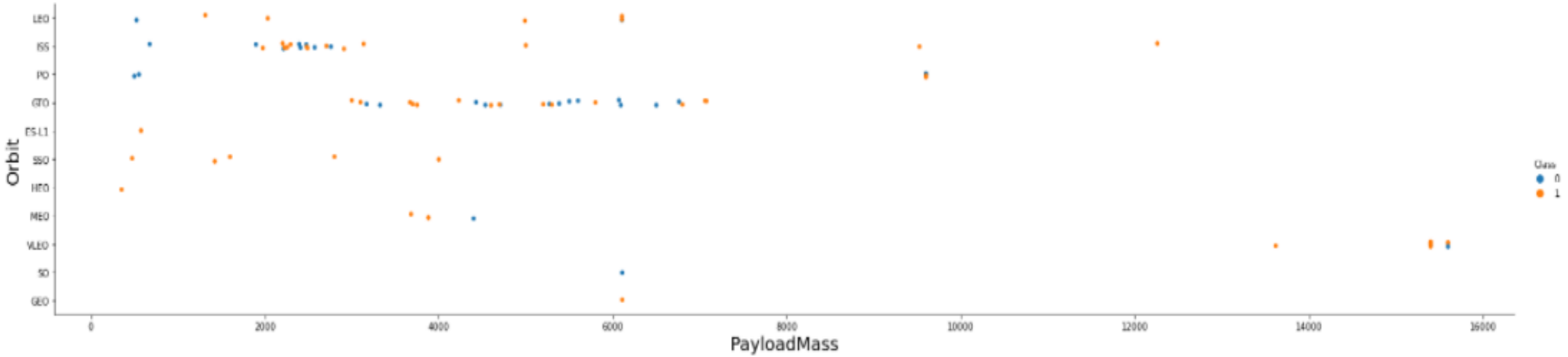
Success rate X Orbit



From the plot we can see that ES-L1 has 100% success rate

RESULTS

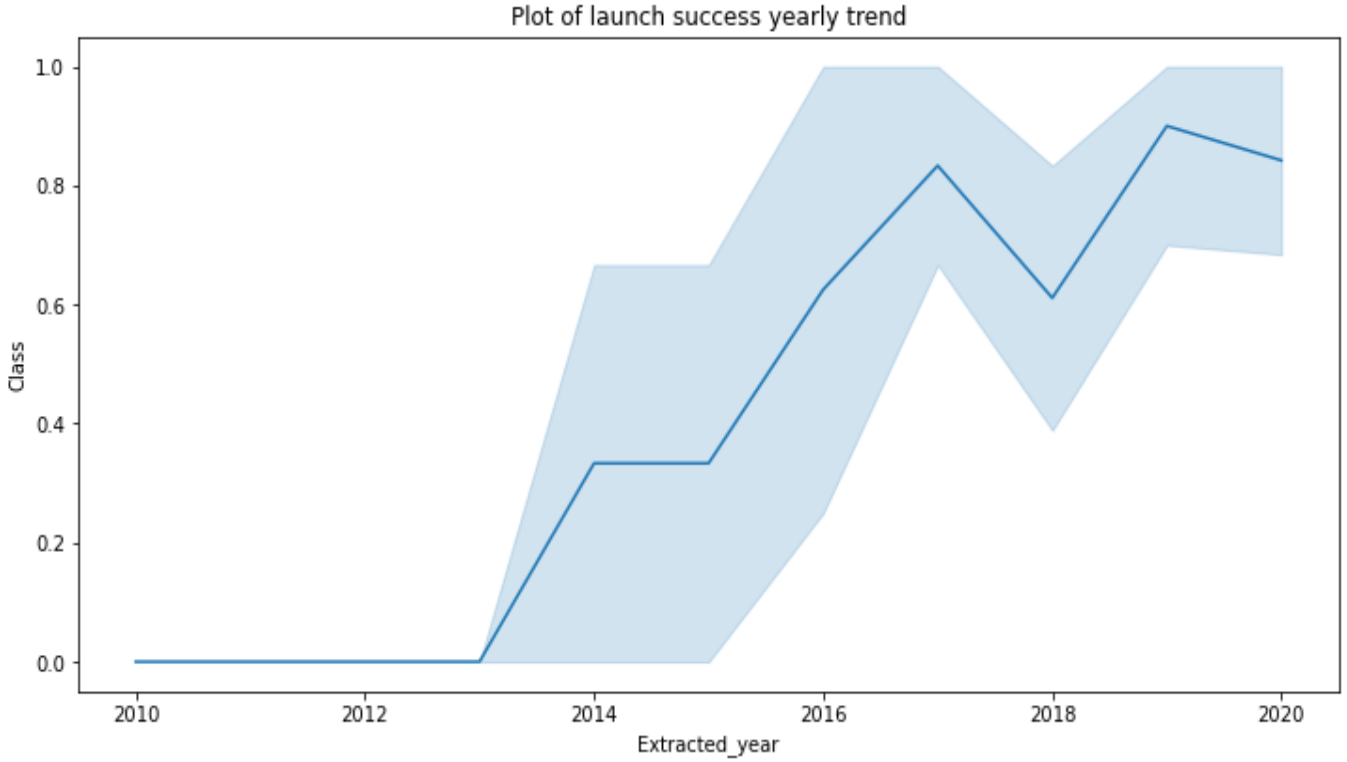
Payload X Orbit



- As we can see, for some orbits higher payloads are more successful

RESULTS

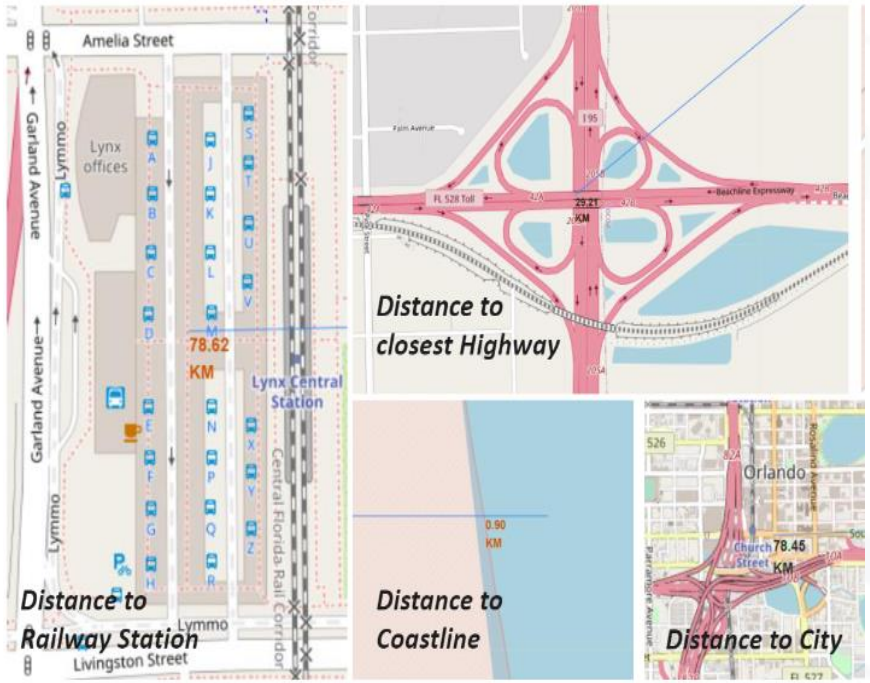
Launch Success Yearly Trend



- We can see from the plot that as years go by, the success rate is increasing, showing a trend.

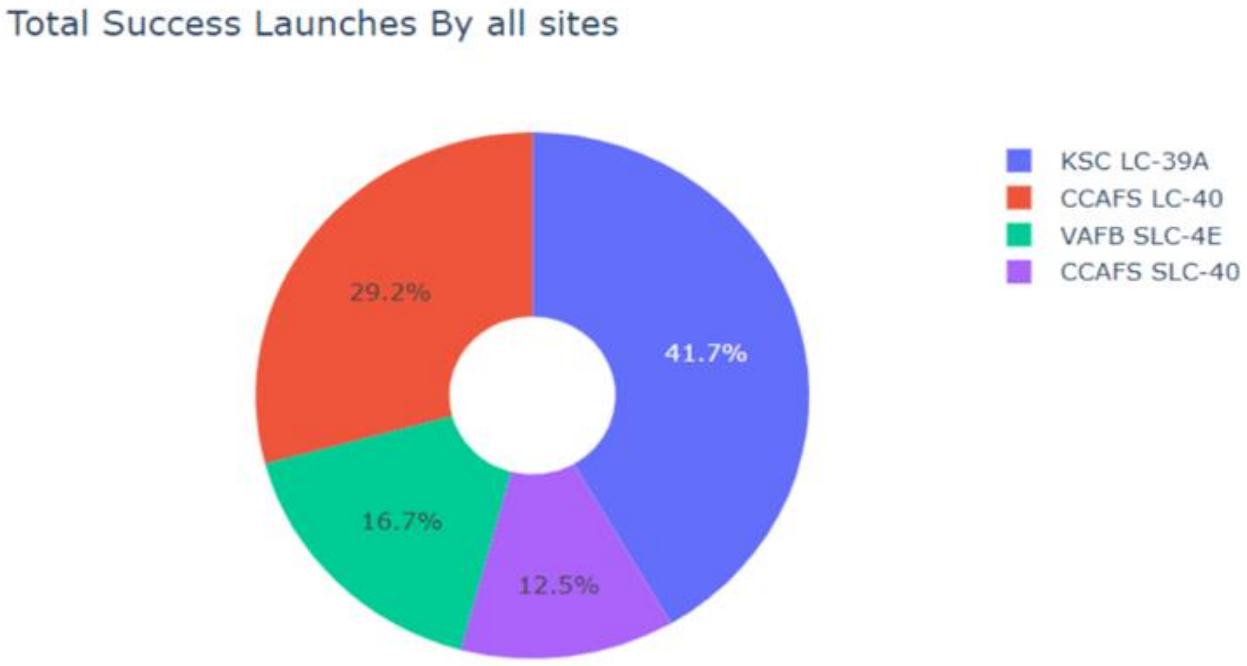
RESULTS

Launch Site distance to landmarks



- We checked that launching sites are far away from cities, but near railroads and coastline

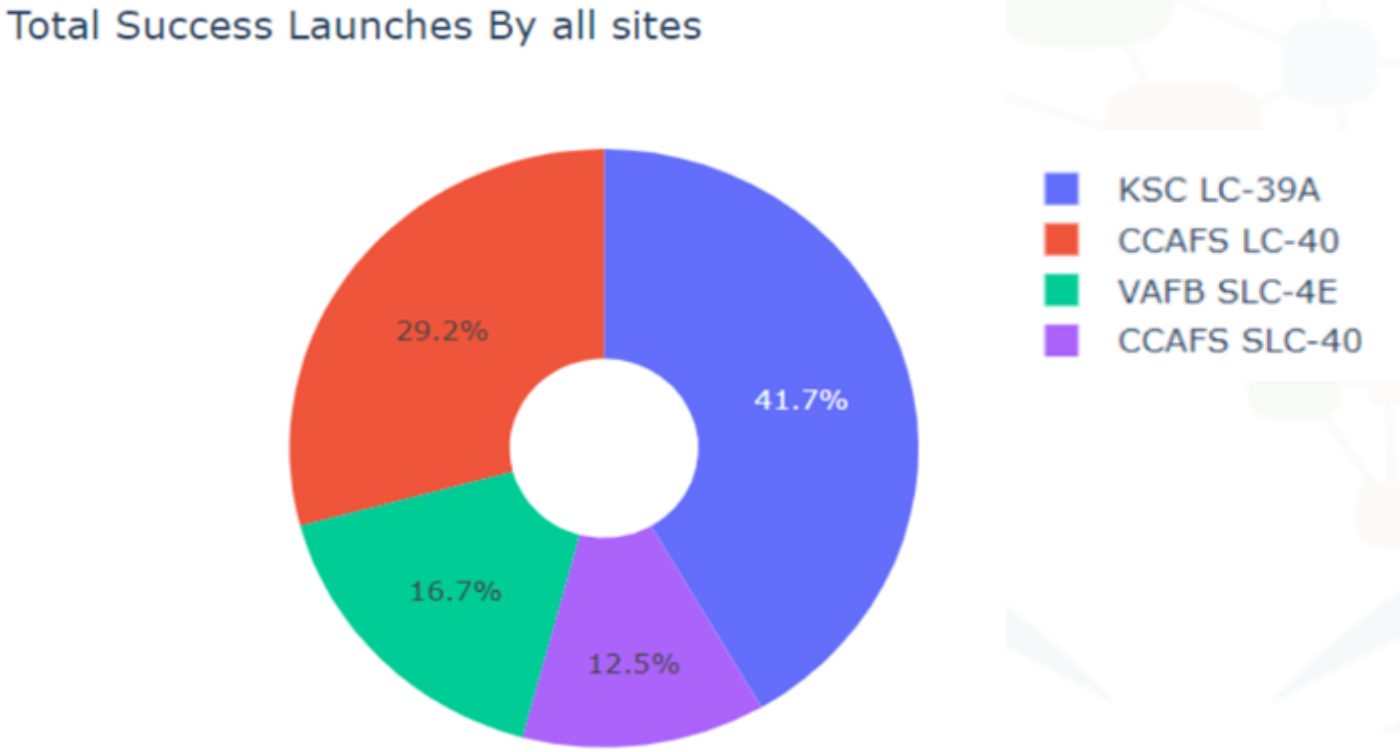
DASHBOARD – TOTAL SUCCESS NUMBER



- We checked that launching sites are far away from cities, but near railroads and coastline

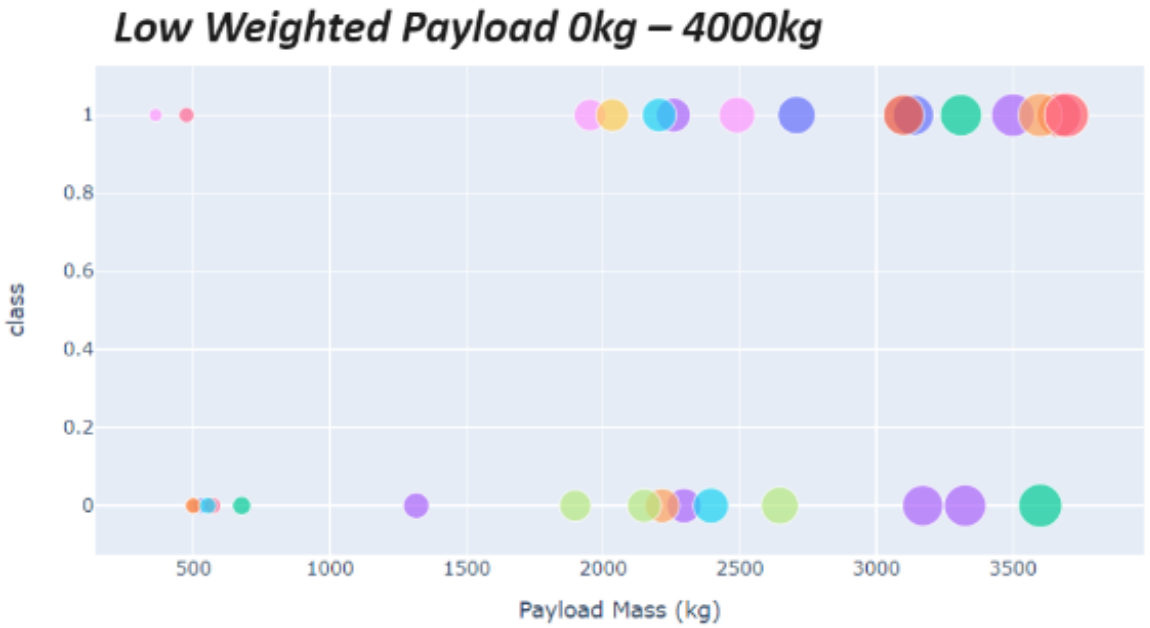
DASHBOARD SUCCESS RATE

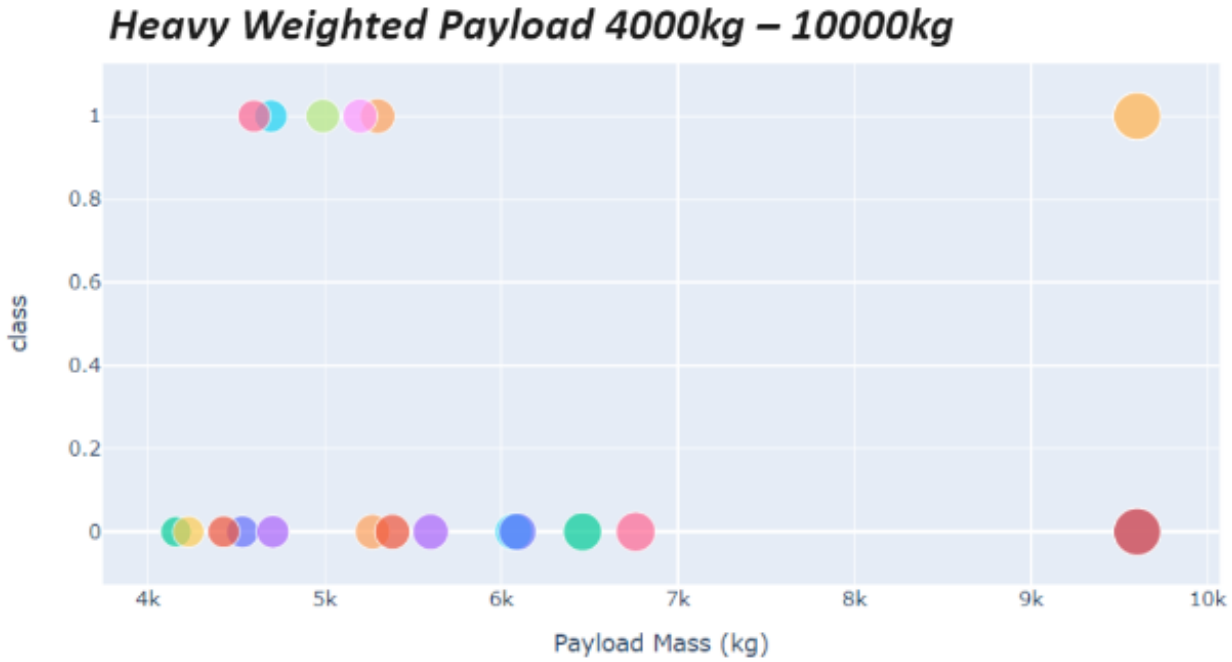
Success rate X Launch site



It can be seen that KSC LC-39 launch site had the best performance and CCAFS SLC-40 the worst.

DASHBOARD – PAYLOAD VS SUCCESS





DISCUSSION

- 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.
- KSC LC-39A had the most successful launches of any sites.

OVERALL FINDINGS & IMPLICATIONS

Machine Learning Algorithims

```
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'}
```

OVERALL FINDINGS & IMPLICATIONS

Machine Learning Algorithims

- Using Scikit Learn package for python we tried many machine learning algorithms to find the best predictive solution.
- We realized that Decision Tree Classifier had the best performance, as it had 87% of accuracy.

CONCLUSION

- Exploratory Data Analysis made possible to us visualize and better understand the data.
- 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!