SpaceX Success Through Data Science

Eilene Kruger

2024/02/08

EileneNPCartel/Capstone-Project (github.com)

OUTLINE

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

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- Commercial space travel market is expanding
  - Various companies launched products to commercialise space travel
  - SpaceX is regarded as the most successful
- Success rate of SpaceX and price of each rocket launches is determined through data analysis using:
  - Data Collection through API/Webscraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Data Visualization
  - Machine Learning
- Results are summarized through:
  - Interactive dashboards and visualisations

INTRODUCTION

- The commercial space travel market is expanding
  - Various companies launched products to commercialise space travel
  - SpaceX is regarded as the most successful

- SpaceY is a new rocket launch company attempting to match the rocket launch costs and success of SpaceX
  - Aim is to determine if SpaceX will reuse the first stage, to predict the cost of launches using success rate, and the best rocket launch location
  - Gathered information is analysed for trends
  - Dashboards are built to visualise the data

METHODOLOGY

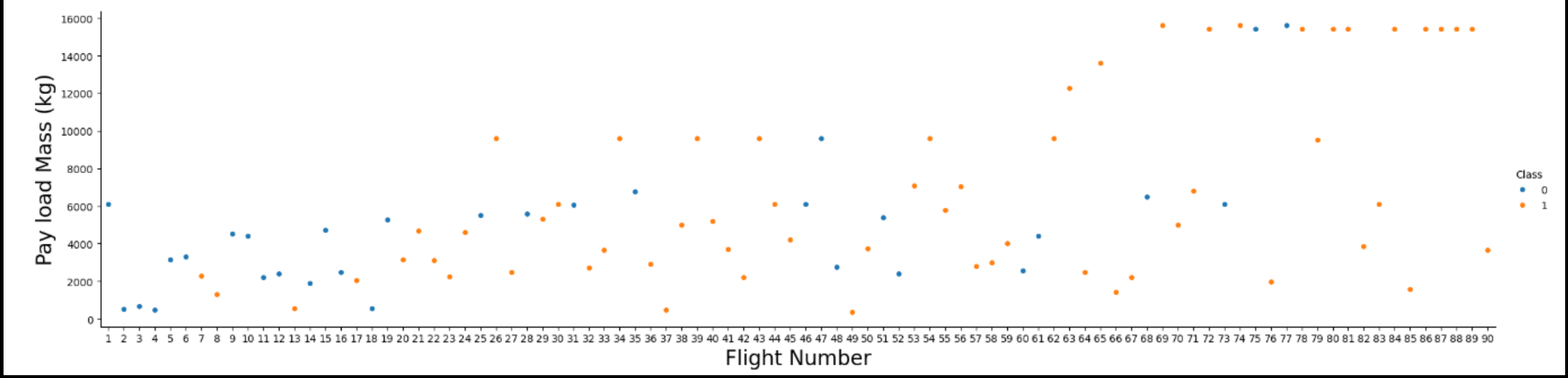
1. Data collected using SpaceX API from Wikipedia
   - Information gathered on launches using columns rocket (booster name), payloads (mass of payload and orbit it is going to), launchpad (location of launch site) and cores (outcome of the landing)
   - Data filtered to only include Falcon 9 launches
2. Web scraping completed
   - Falcon 9 launch records extracted from HTML table and converted into a Pandas dataframe
3. Data wrangling applied to simplify the dataset
   - Success or failure of launches assigned a 1 or 0 pending on successful landing or failure thereof
4. Data explored using SQL
   - Basic summary of data extracted:
     - Launch site names
     - Total and average payload mass
     - Year of first successful landing
     - Boosters names having success in drone ship with payload mass > 4000 but < 6000
     - Total successful and failure outcomes
     - Booster versions carrying maximum payload mass
     - Ranking landing outcomes

METHODOLOGY

4. Data explored using SQL (continued)
   - Relationships determined between the following attributes and displayed via scatter point charts and bar charts:
     - flight number/launch sites,
     - payload/launch sites,
     - success rate/orbit types,
     - flight number/orbit types, and
     - payload/orbit types
   - Launch success across several years visualized on a line graph
   - Dummy variables created to categorical columns
5. Data visualized using Folium
   - Launch sites and success rates marked on a map using coordinates
6. Dashboard built using Plotly Dash
   - Launch records success rates shown
7. Prediction analyses completed using Machine Learning
   - Finding best method to predict the data

RESULTS

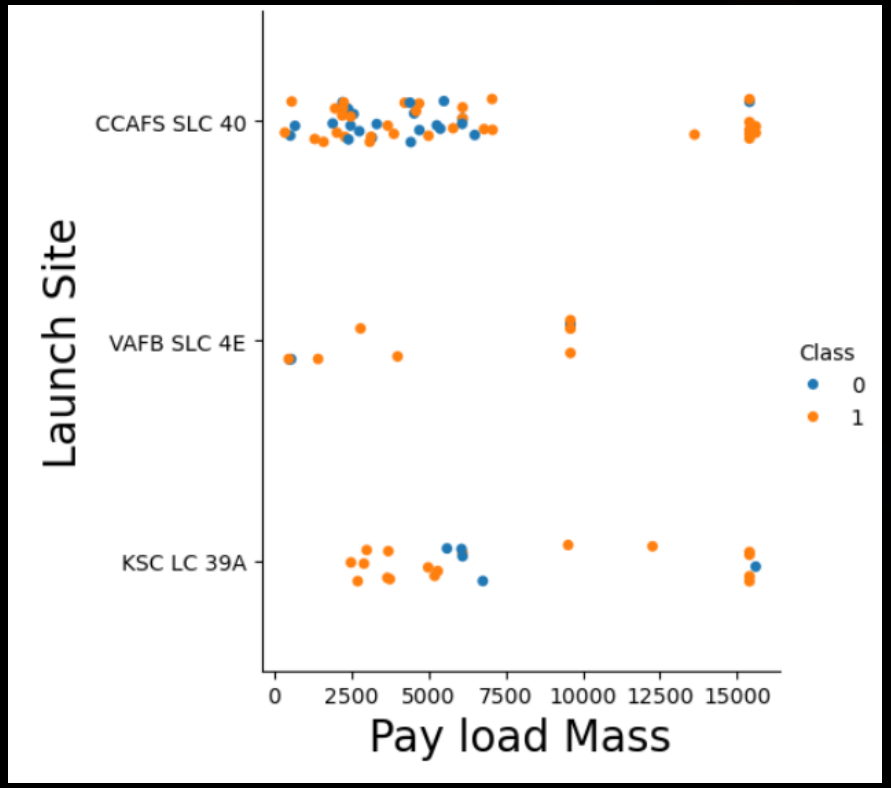
- Exploratory analysis:
  - SpaceX uses 4 different launch sites
  - Average payload mass = 2928,4kg
  - First successful landing in ground pad took place in 2015

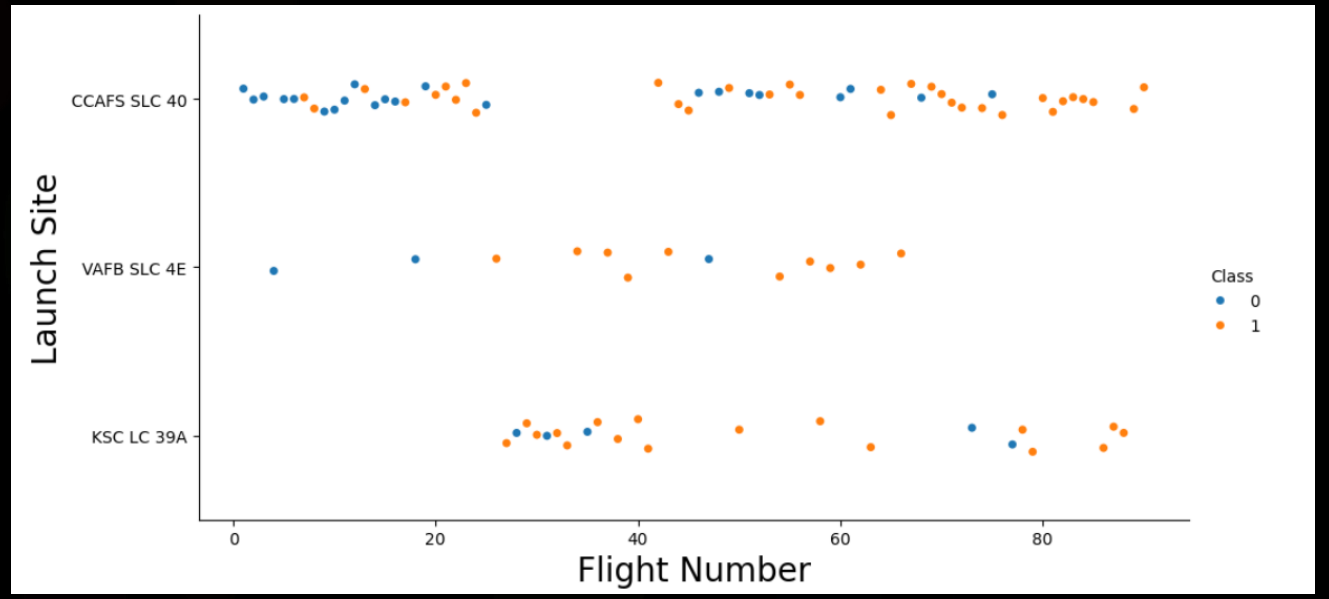


As flight number increases, chances of successful landing increases. Pay load mass increase leads to smaller chance of second stage to return

RESULTS

- Exploratory analysis (cont.):
  - Success rate of each launch site is determined by specific flight numbers



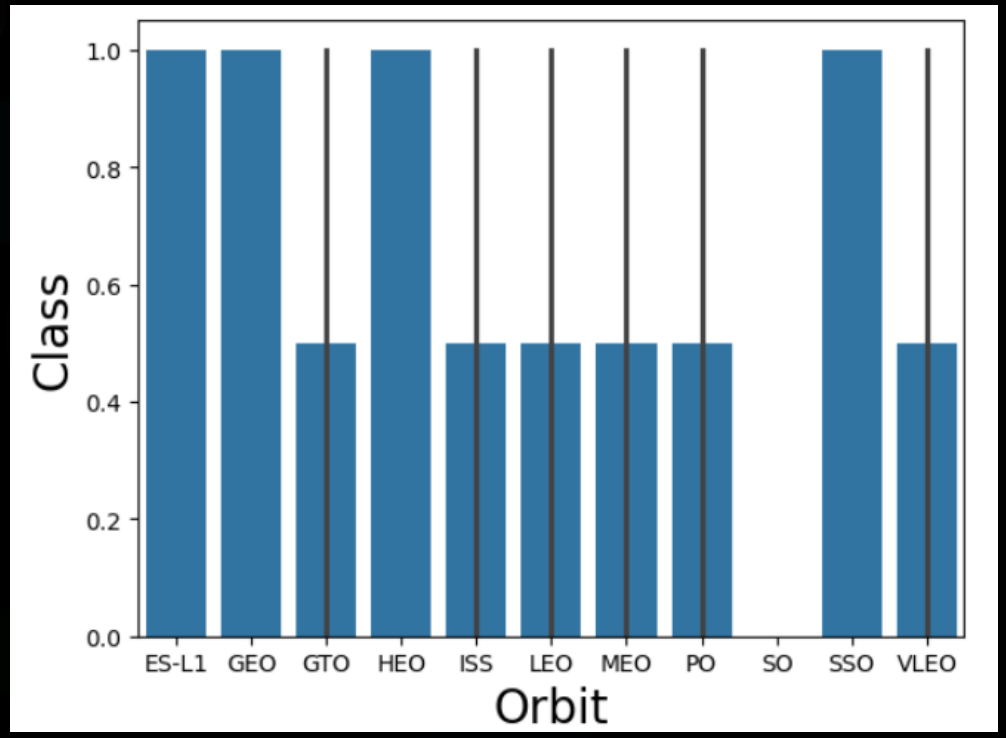


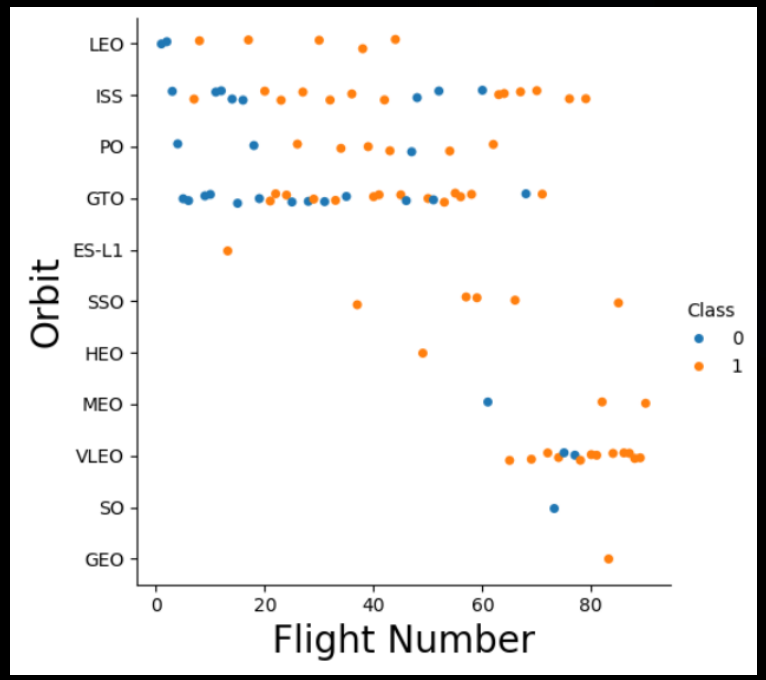
- For CCAFS SLC 40 success rate increases with larger payload masses
- VAFB SLC 4E no rockets with a mass greater than 10,000kg are launched
- KSC LC 39A shows higher success rate for rockets launched with a mass up to 5000kg

SKILLS NETWORK

RESULTS

- Exploratory analysis (cont.):
  - Orbits with highest success rates are ES-L1, GEO, HEO, SSO



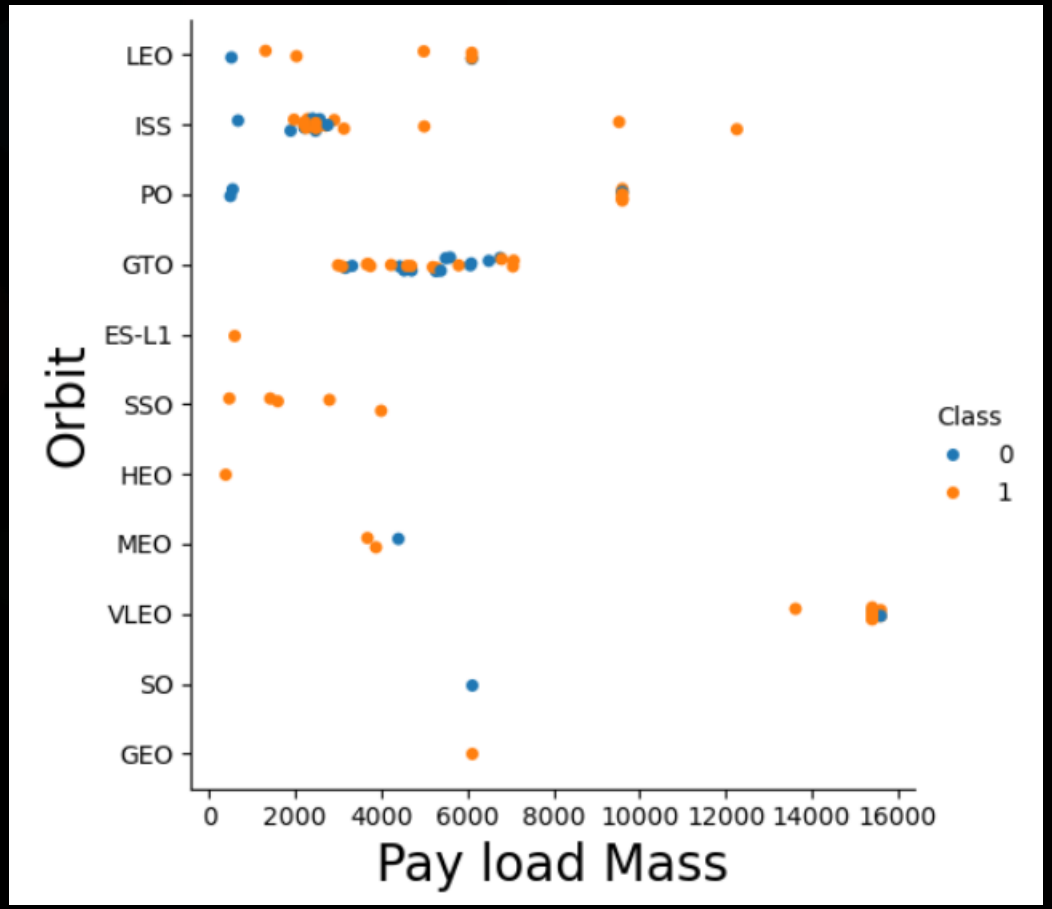


- Most orbits show an increased success rate with an increase in flight number, except for GTO

SKILLS NETWORK

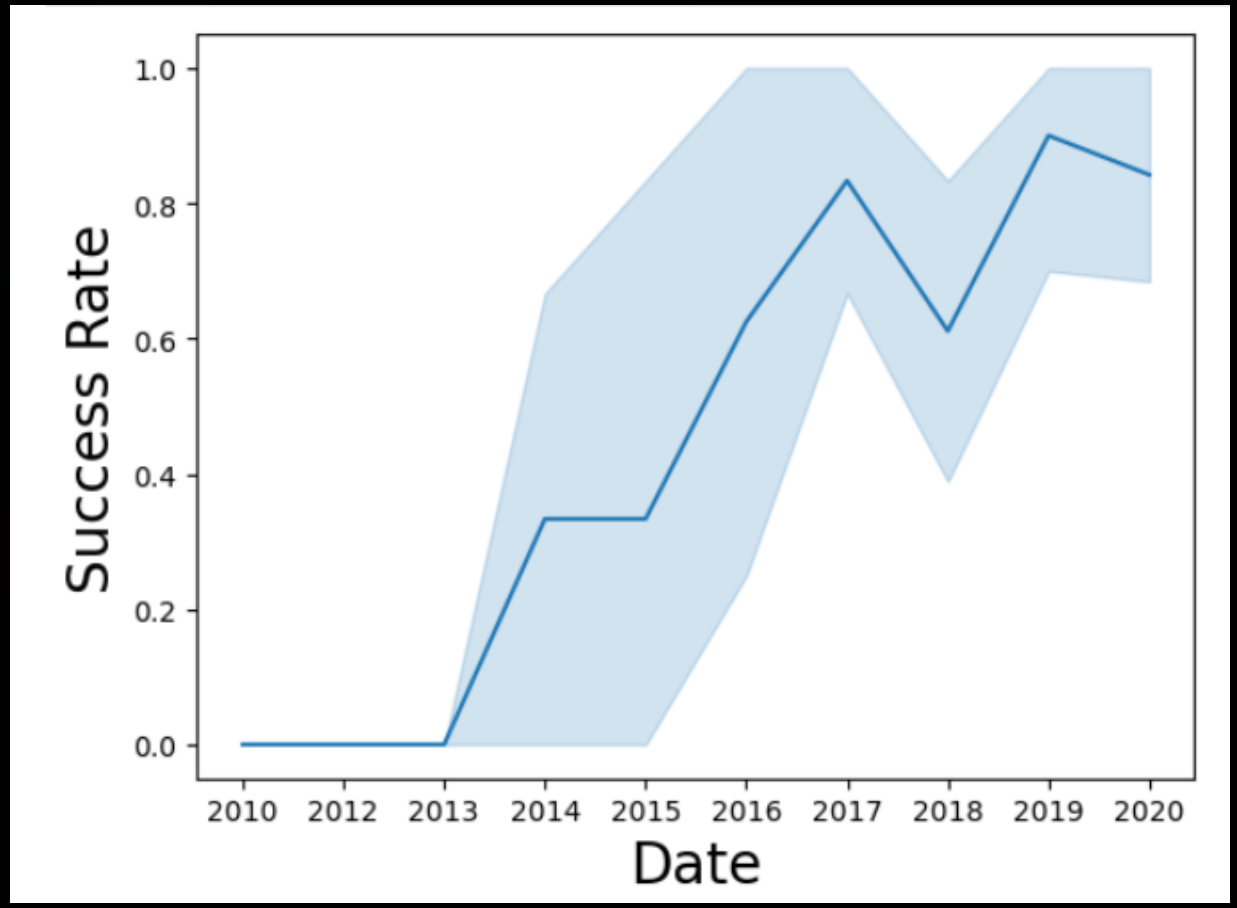
RESULTS

- Exploratory analysis (cont.):
  - Success rate increase with payload mass for:
    - VLEO
    - PO
    - ISS
    - LEO
  - Success rate increase with smaller payload mass for:
    - ES-L1
    - SSO
    - HEO
    - MEO
    - GEO
  - GTO shows no relationship between success rate and payload mass



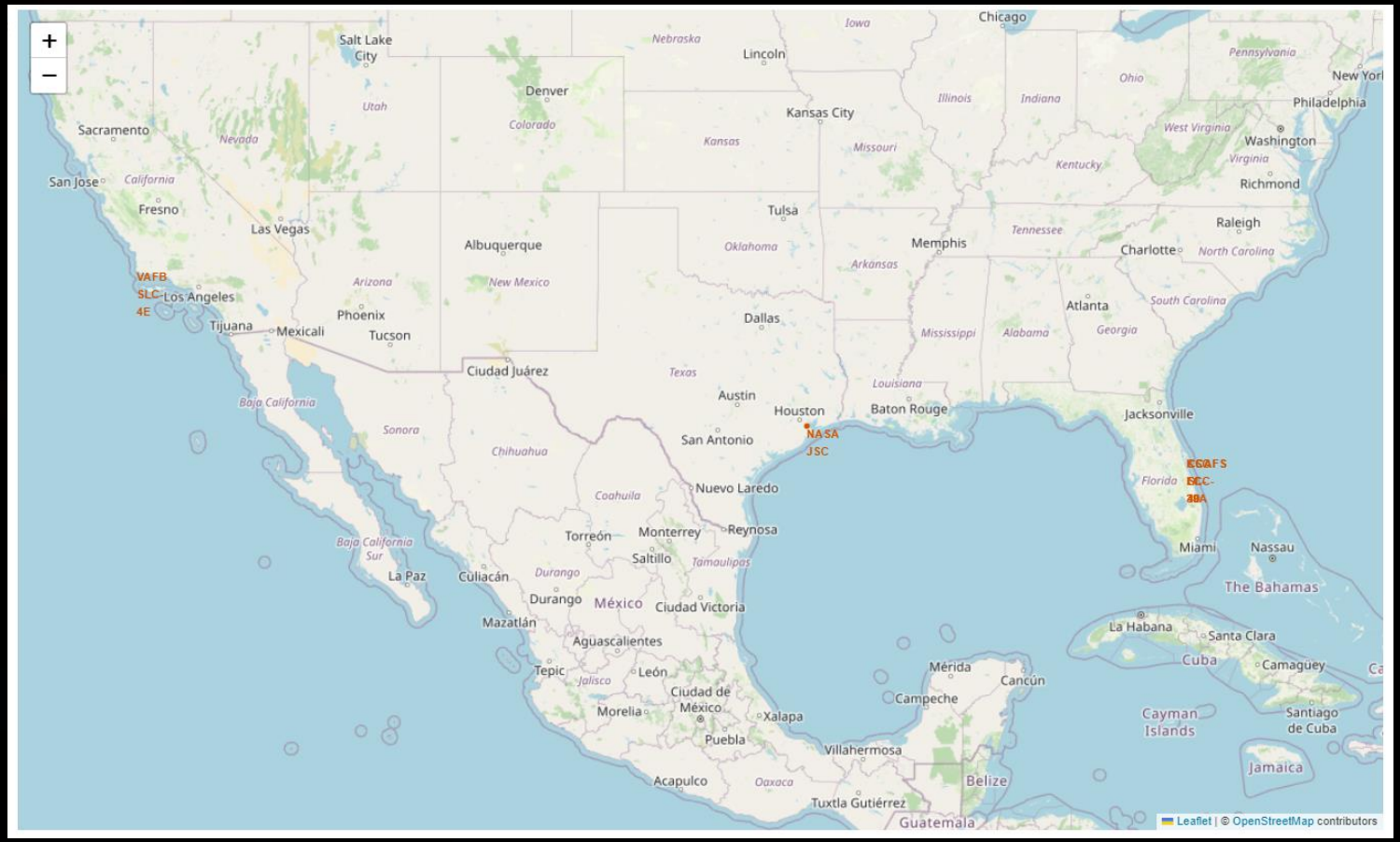
RESULTS

- Exploratory analysis (cont.):
  - An increase in landing success rate is seen from the year 2013



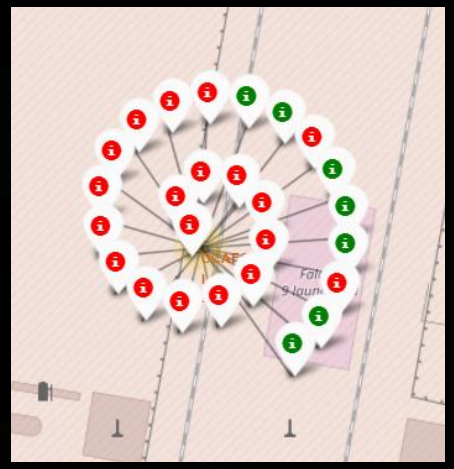
RESULTS

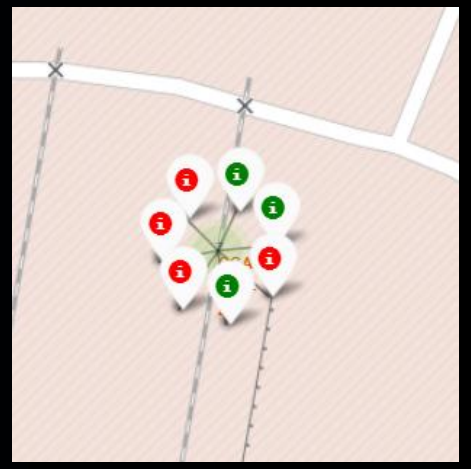
- Folium analysis:
  - Launch sites are close the coastlines

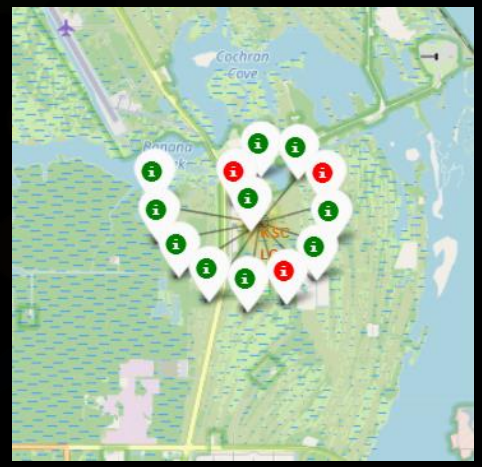


RESULTS

- Folium analysis:
  - Some launch sites showed >50% failure rate
    - CCAFS SLC-40
    - CCAFS LC-40
    - VAFB SLC-4E
  - KSC LC-39A showed a >70% success rate
  - Launch sites are located close to railways and highways

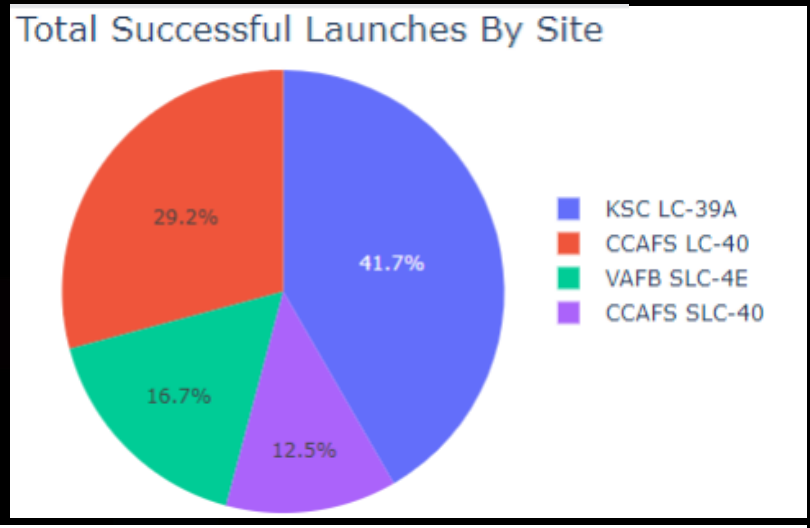




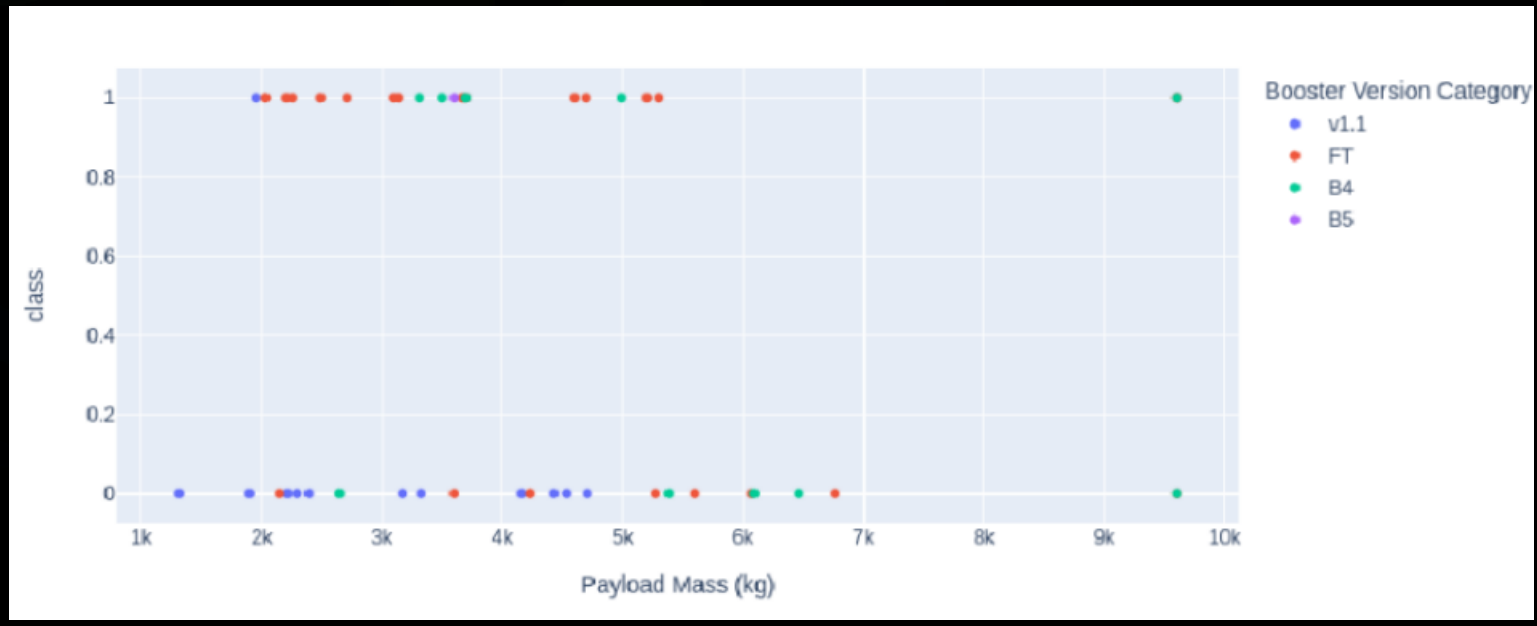


RESULTS

- Dashboard analysis
  - As before, KSC LC-39A showed the highest launch success rate



- The launches with the highest success rates have payload masses between 2000kg and 5500kg



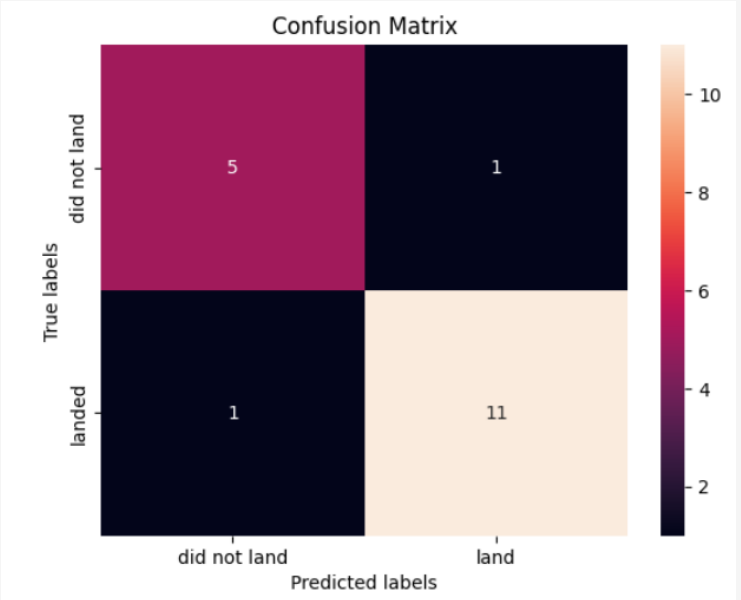
IBM Developer

RESULTS

- Predictive analysis
  - Various classifier models were created to test the predictability of the launch outcomes
  - The Decision Tree model predicted launch outcomes with the highest accuracy (87,5%)

```
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 0.875
Best params is : {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 10, 'splitter': 'random'}
```

CONCLUSION

- Based on obtained results:
  - Successful landing outcomes started to increase from the year 2013, where first successful landing in ground pad took place in 2015
  - Successful landing outcomes seem to correlate with the number of flights
  - Payload mass varied across different launch sites and is a contributing factor to landing success
  - Launch sites are located close to railways and highways (this provides logistic support) and close to coastlines
  - KSC LC-39A had the highest landing success rate
  - A decision tree machine learning model can most accurately predict success of landing outcomes

APPENDIX

- EileneNPCartel/Capstone-Project (github.com)