SpaceX Falcon 9 Analysis

Harshvardhan Sharma

18th Feb 2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

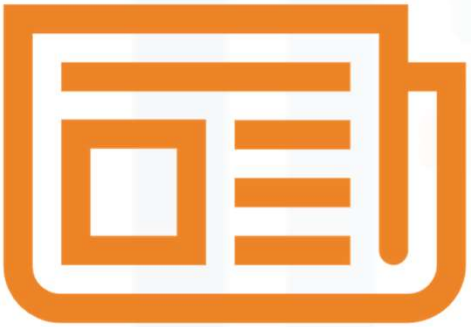
EXECUTIVE SUMMARY

- Methodology Summary:
  - Data collected via API and Webscraping Wikipedia page
  - Data Wrangling to clean data
  - Exploratory Data Analysis using SQL and Visualisation techniques
  - Interactive Plotly web app to visualise payload and success launch data at each launch site
  - Exploring launch analysis using Folium maps
  - Predictive Analysis for classification of Falcon 9 landing success
- Summary of Results

INTRODUCTION

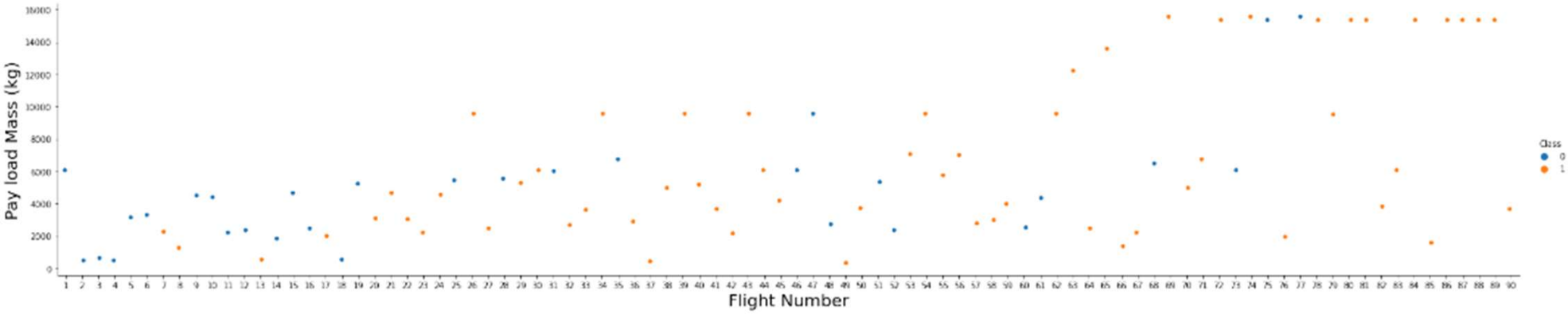
- SpaceX advertises Falcon 9 rocket launches with a cost of 62 million dollars
- Other providers cost upward of 165 million dollars each
- Savings are because SpaceX can reuse the first stage
- What we aim to find
  - Understand Falcon 9 history of success and failures
  - Predict if Falcon 9 rocket will land successfully

METHODOLOGY

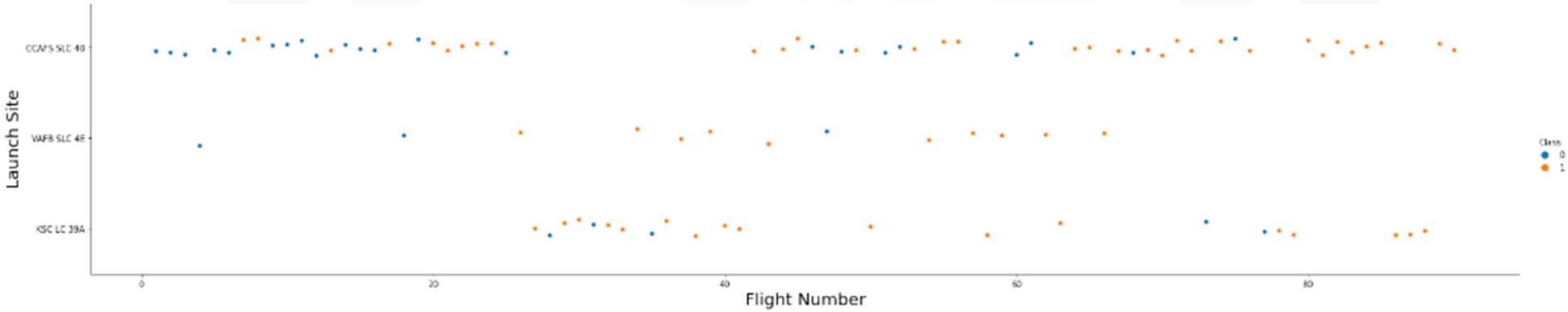


- Data collection methodology
  - SpaceX REST API
  - Web scraping from Wikipedia
- Data wrangling
  - One hot encoding data fields for ML and data cleaning of null values/irrelevant columns
- Exploratory data analysis (EDA)
- Interactive visual analytics
  - Folium maps
  - Plotly dashboard
- Predictive analysis
  - LR, KNN, SVM, DT models were built and evaluated

Data Visualization

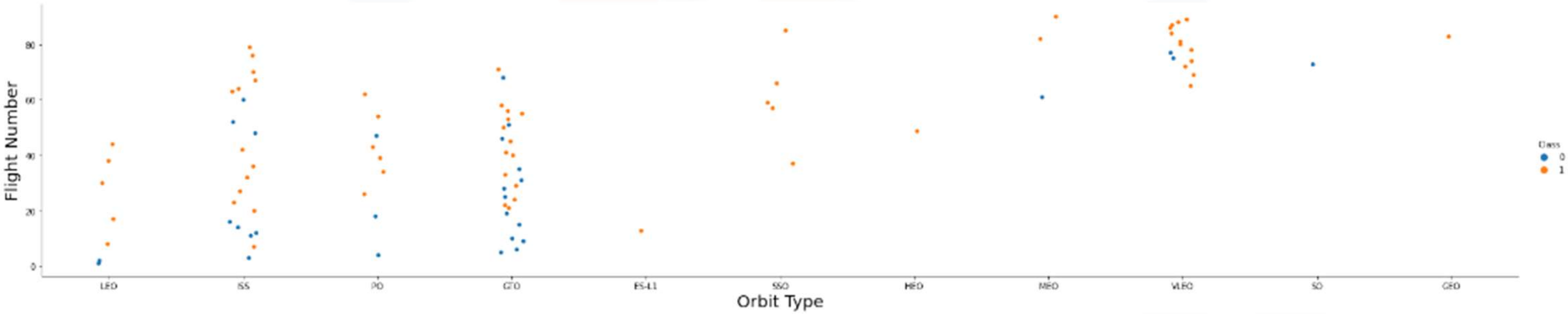


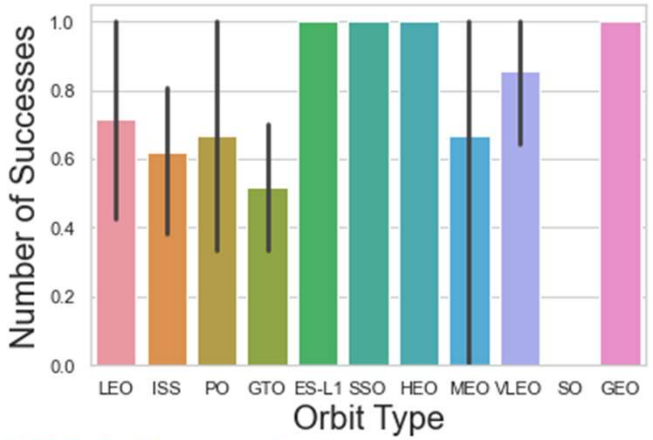
It appears that Flight Number and Pay Load Mass has a positive correlation



SLC4E and LC39A have a high success rates, however after flight 40, even SLC30 begins to reach high success levels

Data Visualization



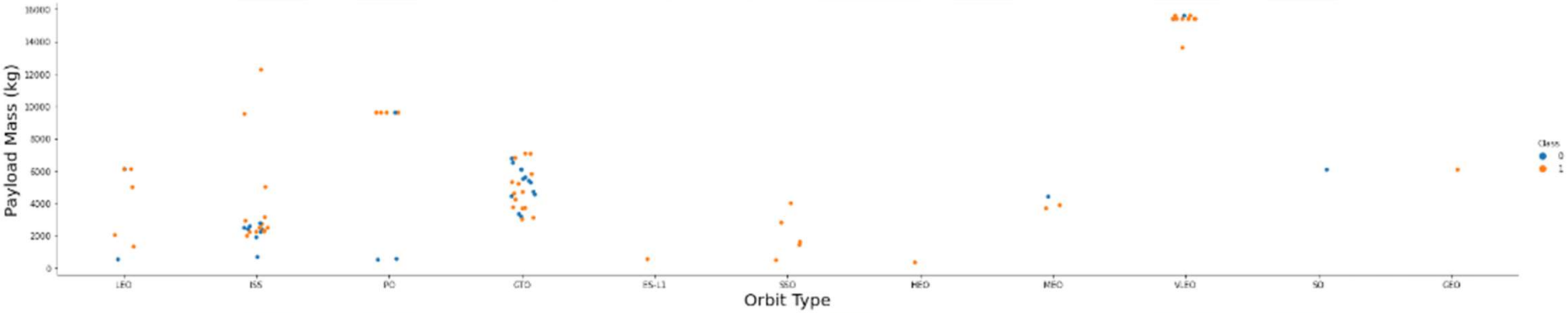


GEO, SO, HEO, ESL-1 each only have one flight and therefore an assessment cannot be made as to the efficacy of its flight success. VLEO has achieved a high success rate in the later flights whereas LEO and ISS also have higher average success throughout all flights.

IBM Developer

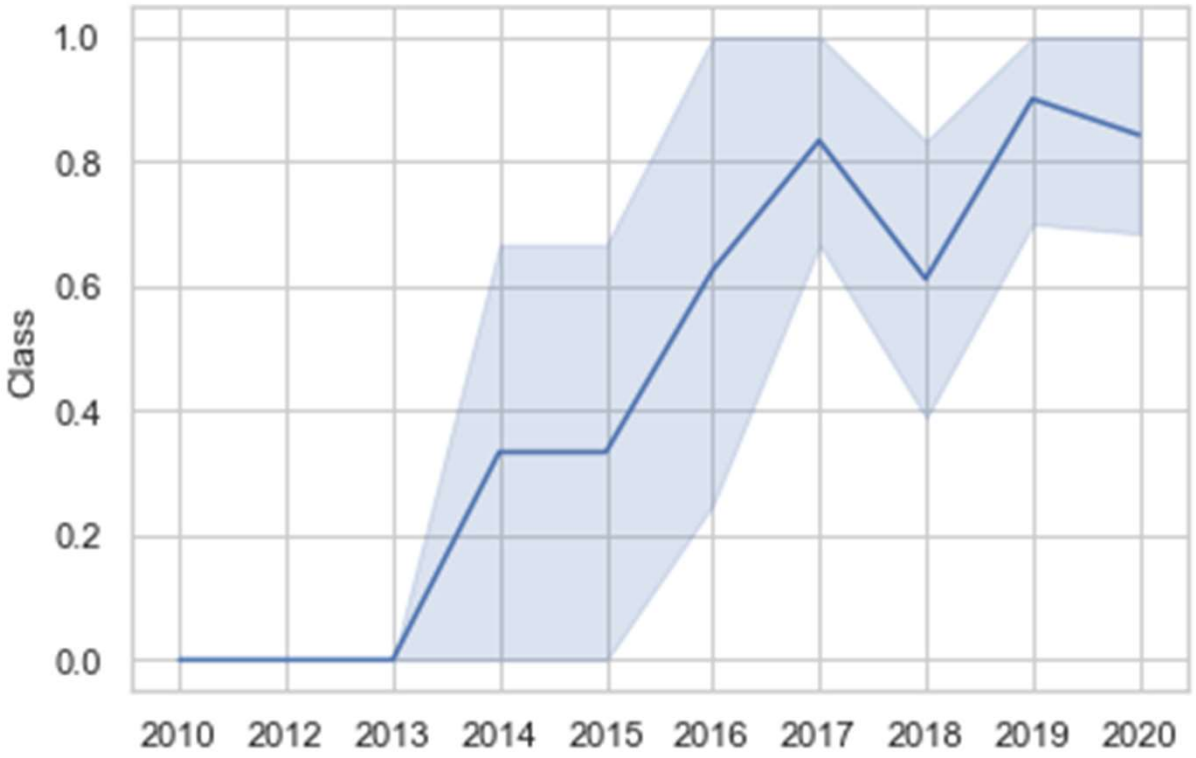
SKILLS NETWORK

Data Visualization



LEO, SSO, and ISS tend to only launch smaller payloads achieving high mission success. GTO on the other hand launches heavier payloads but has seen more mission failure. VLEO almost exclusively launches the heaviest payloads and achieves a very high mission success. The key takeaway: VLEO orbits are very likely to be heavy payloads whereas lighter payloads will be in SSO, LEO, or ISS save a few exceptions.

Data Visualization



There is a positive trend of mission success in the years beyond 2013. There was a negative trend in 2017 and a slight dip in 2019 though the overall trend of mission success has been upward.

Python SQL Query

- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first successful landing outcome in ground pad was achieved**
- **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
- **List the total number of successful and failure mission outcomes**
- **List the names of the booster versions which have carried the maximum payload mass**
- **Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order**

Github Notebook Link

Maps with Folium

- Used Latitude and Longitude coordinates of given launch sites to add markers and labels
- Assigned launch outcome (success/failure) from the data frame to classes 1 and 0 respectively and assigned the class ‘Green’ and ’Red’ markers to marker clusters grouped by launch site
- Measured minimum distances of launch sites to:
  - Cities
  - Highways
  - Coastlines
  - Railways
- It was determined that launch sites are within proximity to **railways**, **highways**, and **coastline**. Launch sites are within 50km of a city.

Github Notebook Link

Build Plotly Dashboard

- Make an interactive web app to visualize launch data
  - Include pie chart to visualize launch landing success by launch site
  - Include scatter plot of payload mass vs. landing success rate color coded bby booster version

IBM Developer

SKILLS NETWORK

Predictive Analysis

Build

Transform data

Split data into Training and Testing Sets

Use ML algorithms to use for classification (KNN, Decision Tree, SVM, Logistic Regression)

Use Grid Search and Cross Validation to find best tuning parameters

Evaluate

Check accuracy of each model

Plot Confusion Matrix

Improve

Feature Engineering

Algorithm Tuning

Select

Model with best accuracy score is chosen on best model

Github Link

ML Method

Find the method performs best:

```
[]: from sklearn.metrics import jaccard_score, f1_score

jaccard_scores = [
                  jaccard_score(Y_test, logreg_yhat, average='binary'),
                  jaccard_score(Y_test, svm_yhat, average='binary'),
                  jaccard_score(Y_test, tree_yhat, average='binary'),
                  jaccard_score(Y_test, knn_yhat, average='binary'),
                 ]

f1_scores = [
              f1_score(Y_test, logreg_yhat, average='binary'),
              f1_score(Y_test, svm_yhat, average='binary'),
              f1_score(Y_test, tree_yhat, average='binary'),
              f1_score(Y_test, knn_yhat, average='binary'),
             ]

accuracy = [logreg_accuracy, svm_accuracy, tree_accuracy, knn_accuracy]

scores = pd.DataFrame(np.array([jaccard_scores, f1_scores, accuracy]), i
scores
```

[]:

	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	0.848214	0.833333	0.833333

After performing analysis on logistic regression, SVM, tree, and KNN methods it was noted that SVM had the highest accuracy and would be used for modeling and prediction

Github Link

CONCLUSION

- Low weighted payloads perform better than heavier payloads on the Falcon 9 rocket system except when in VLEO orbit
- Success rates for SpaceX launches display a positive linear trend indicating that success rates of Falcon 9 rocket systems are increasing
- KSCLC 39A has the most success of all the sites by a large margin
- GEO, SO, HEO, ESL-1 only have one launch and data is inconclusive to their efficacy. VLEO has demonstrated success.

Thank you!