IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Thomas Wood
6/13/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

The analysis aims to assess what factors leads to a successful rocket landing.

Summary of methodologies

- The following data science techniques were used:
  - Collect: web scrapping of Space X data using their API
  - Wrangling the data to create a fail/success variable.
  - Explore and analyze the data using data visualization, SQL, dash boards and maps.
  - Apply various ML techniques to aim to predict successful launch factors.
- Summary of all results
  - Launch success has improved over time. Orbits GEO, HEO, SSO and ES-L1 have 100% success rate.

Introduction

Project background and context

- The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX.
- SpaceX’s accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.
- The payload is enclosed in the fairings. Stage two, or the second stage, helps bring the payload to orbit, but most of the work is done by the first stage. This stage does most of the work and is much larger than the second stage.
- This stage is quite large and expensive. Unlike other rocket providers, SpaceX's Falcon 9 Can recover the first stage.
- Sometimes the first stage does not land. Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer.

Introduction

Problems you want to find answers

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

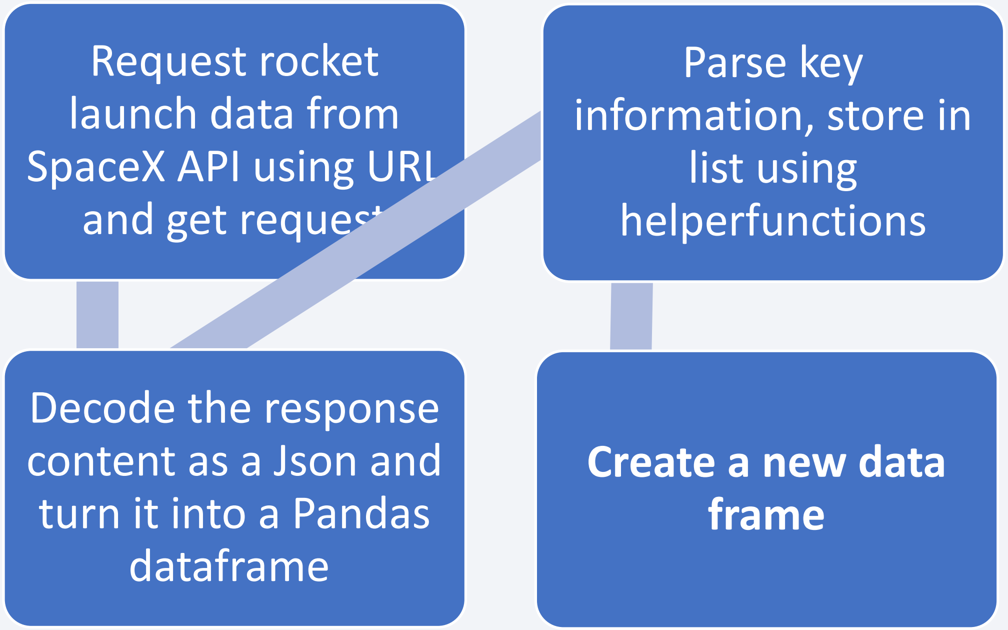
- Data collection methodology:
  - Data was collected using Space X REST API and web scrapping Wikipedia
- Perform data wrangling
  - One-hot encoding applied the categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/. We have the different end points, for example: /capsules and /cores We will be working with the endpoint api.spacexdata.com/v4/launches/past.
- We will perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API. This result can be viewed by calling the .json() method. Our response will be in the form of a JSON, specifically a list of JSON objects. The we convert the JSON objects into a dataframe.
- We will also be using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records in Wikipedia.

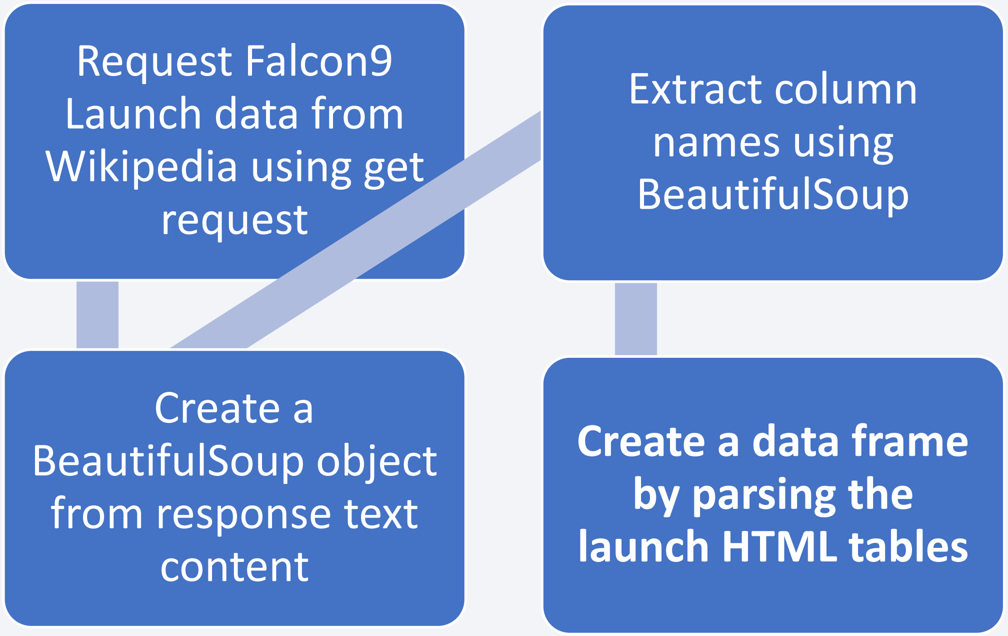
Data Collection – SpaceX API

- Source SpaceX data using json, get , parse and save as dataframe.
- GitHub URL: [Data collection API]



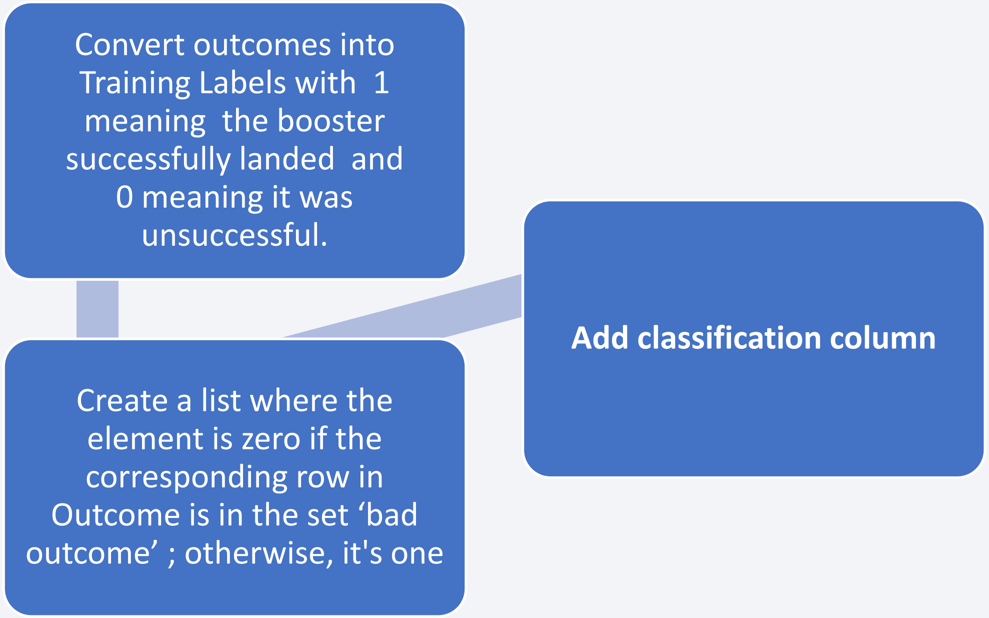
Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- GitHub URL: [Data Collection with Web Scraping]



Data Wrangling

- Created a classification column with good or bad outcomes from the different type of landing outcomes
- Add the GitHub URL: [Data Wrangling]



EDA with Data Visualization

- Following charts are presented:
- **FlightNumber vs. PayloadMass + Outcome Overlay:** shows increased 1st stage landing success as number of flights increase. But shows lesser chance of stage 1 return, when payloads increase.
- **Relationship between Flight Number and Launch Site:** increased success with increasing flight numbers on all launch sites.
- **Relationship between Payload and Launch Site:** VAFB-SLC launch site there are no rockets launched for heavy pay load mass (greater than 10000).
- **Relationship between success rate of each orbit type:** ESL1, GEO, HEO and SSO have 100% success rate.
- **Relationship between Flight Number and Orbit type:** LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- **Relationship between Payload and Orbit type:** heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- **Launch success yearly trend:** shows increase success over the years
- GitHub URL: EDA with Data Visualization

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

After downloading the datasets:

- Identified the unique launch sites
- 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. Use a subquery
- List the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- GitHub URL: [Complete the EDA with SQL.ipynb]

Build an Interactive Map with Folium

- Adding markers at the different launch sites and the launch outcomes for each site, and see which sites have high success rates.

- GitHub URL: [Locations Analysis with Folium]

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard: Input Launch site and Payload Range (Slider)
- Output:
  - Pie chart showing:
    - Success of launches at all launch sites combined
    - Success and failure for each selected launch site
  - Scatter Plot showing success rate for each Booster Version and Payload Mass
- Explain why you added those plots and interactions: The graphs allow to assess the successful relationship of Booster versions and Payload Mass in relationship to the different Launch Sites.
- GitHub URL: Space X Dash

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Task 1: Created a separate array with all outcomes (‘Class’/ Y)
- Task 2: Standardized the data frame (X)
- Task 3: Split the data into train and test data.
- Applied Logistic regression, Support Vector Machine, Decision Tree Classifer and K-nearest Neighbors to the training data. Used the best parameters to predict and then measure the accuracy.
- Best models are Decision Tree Classifier and K-Nearest Neighbors with similar outcomes (SVM did not work due to lack of computing power.
- You need present your model development process using key phrases and flowchart
- GitHub URL: Machine Learning Prediction

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

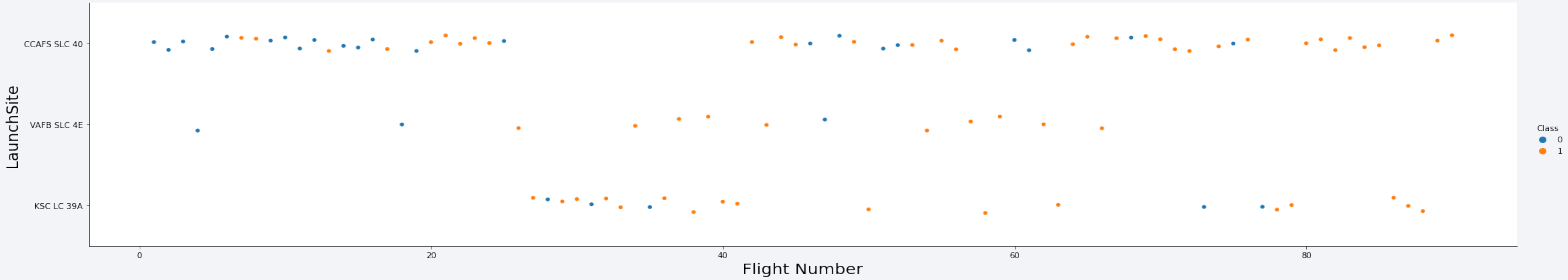
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Regardless of launch site, success rates increased over time.

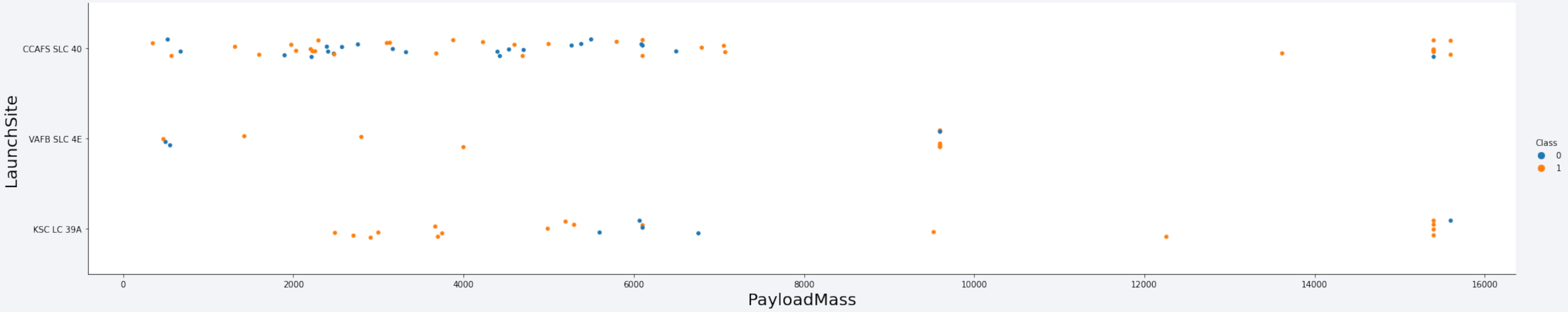
Classification is 1 = Good outcome, 0 = Bad outcome



Payload vs. Launch Site

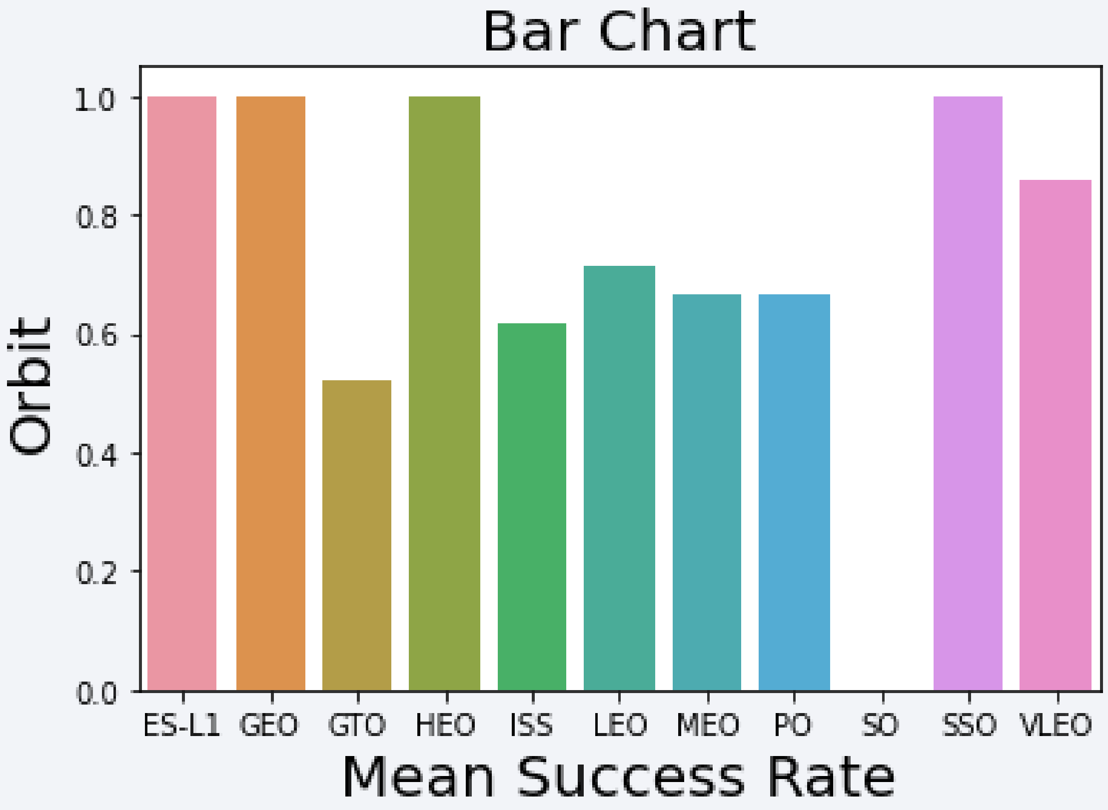
VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

Classification is 1 = Good outcome, 0 = Bad outcome



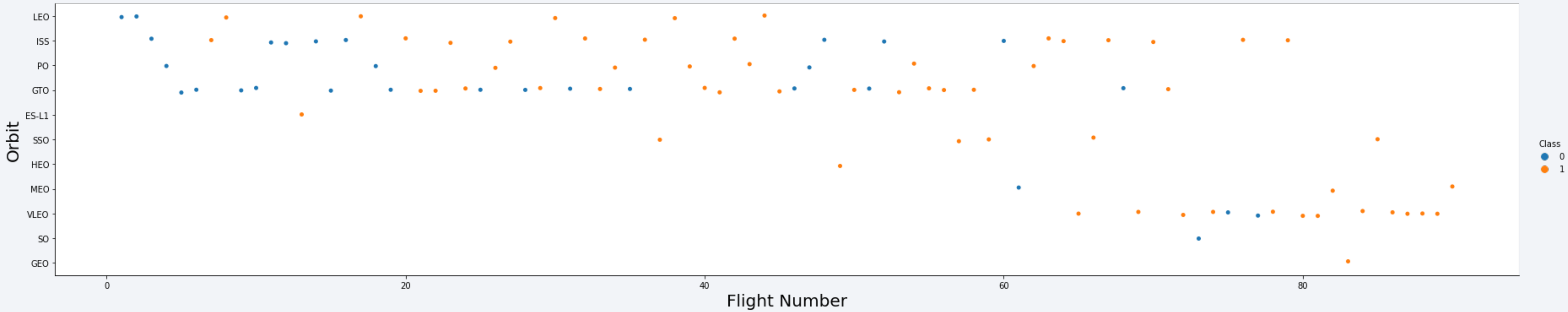
Success Rate vs. Orbit Type

- ES-LI, GEO, HEO & SSO have a 100% success rate



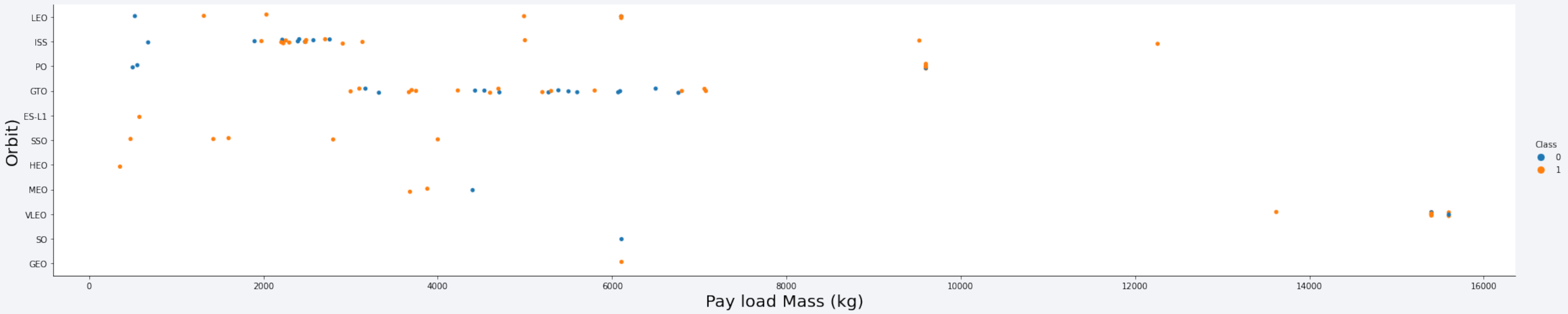
Flight Number vs. Orbit Type

- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



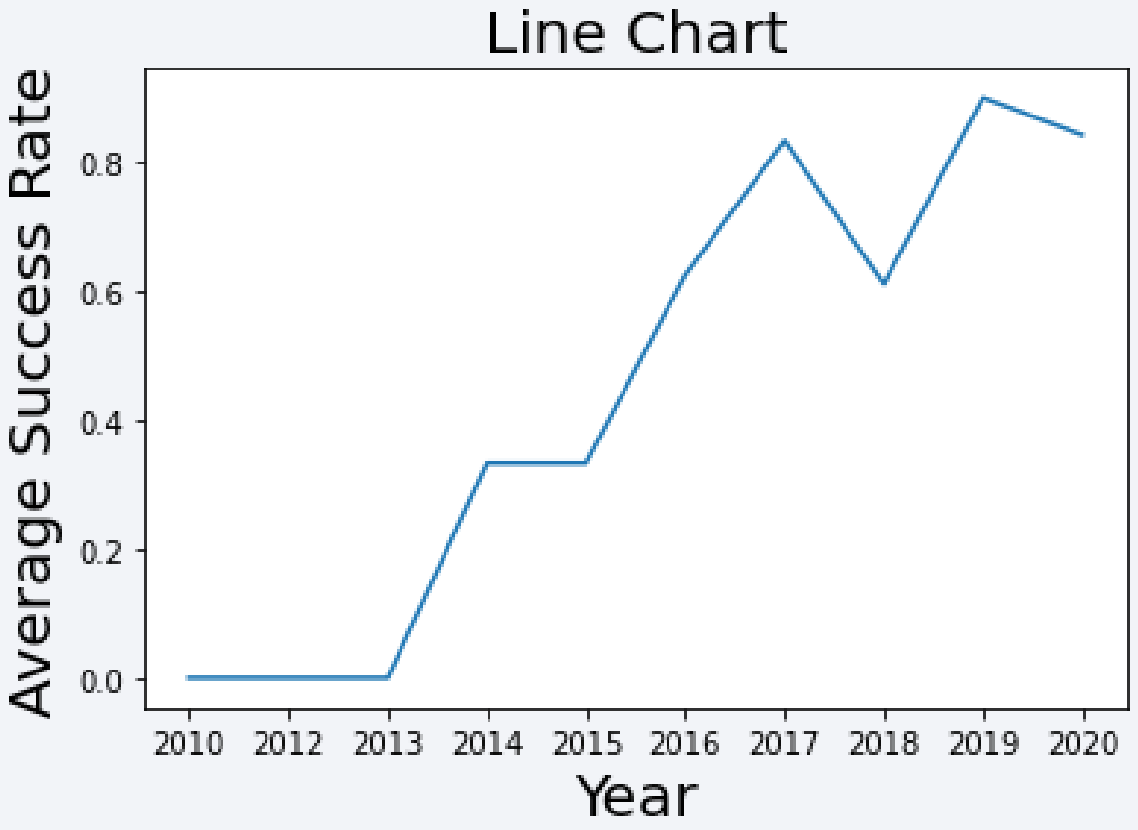
Payload vs. Orbit Type

- With eavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



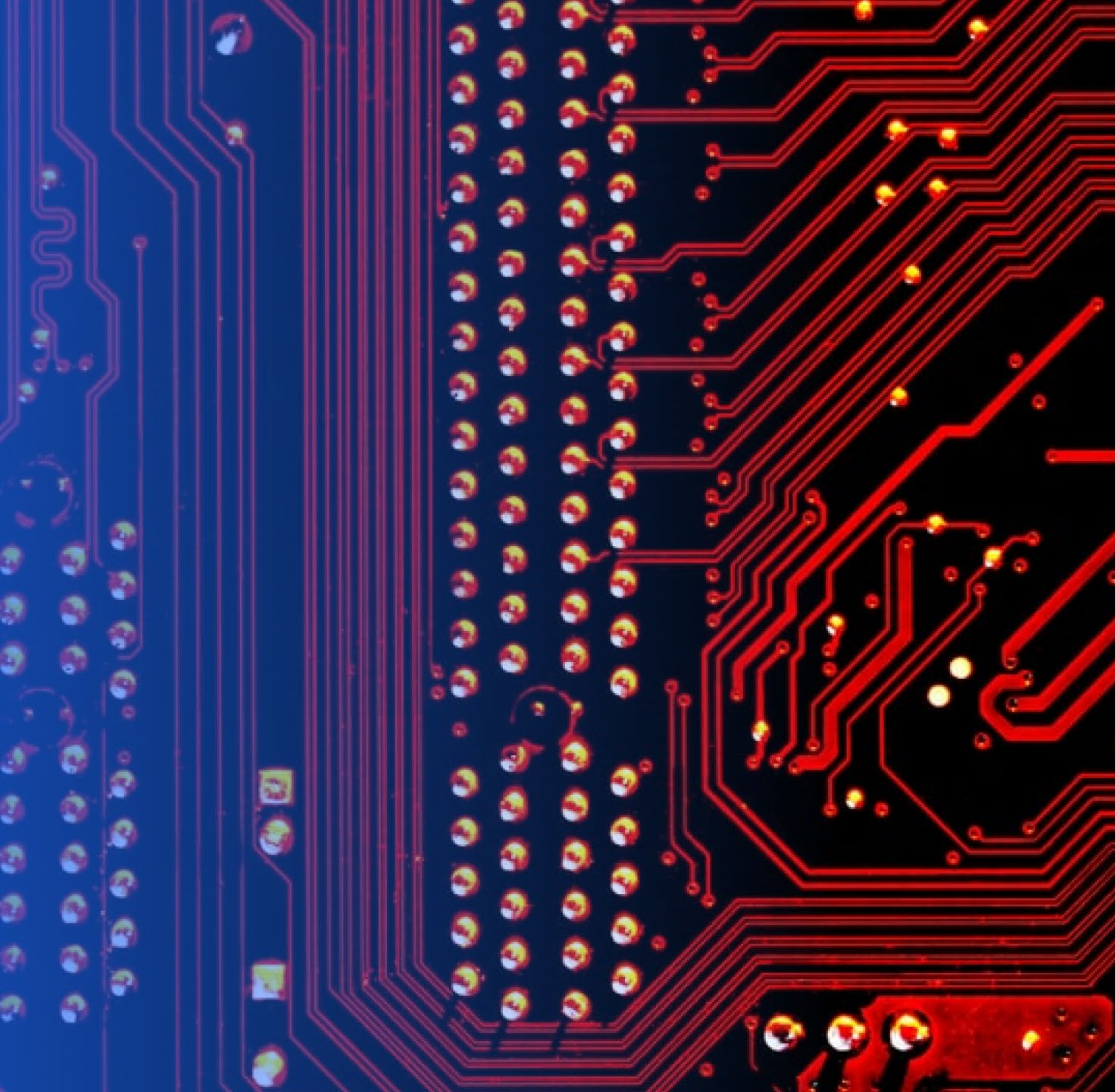
Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020



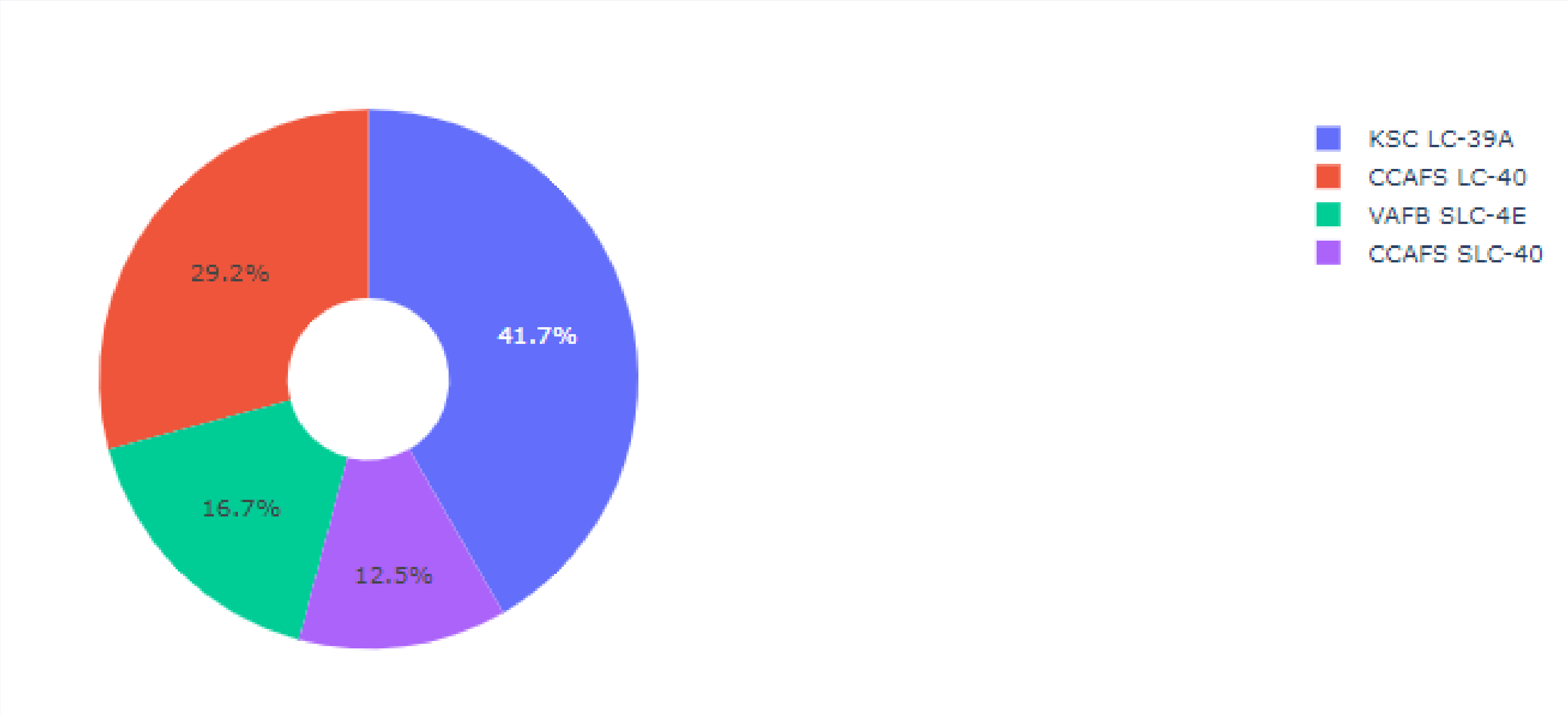
Section 4

Build a Dashboard with Plotly Dash



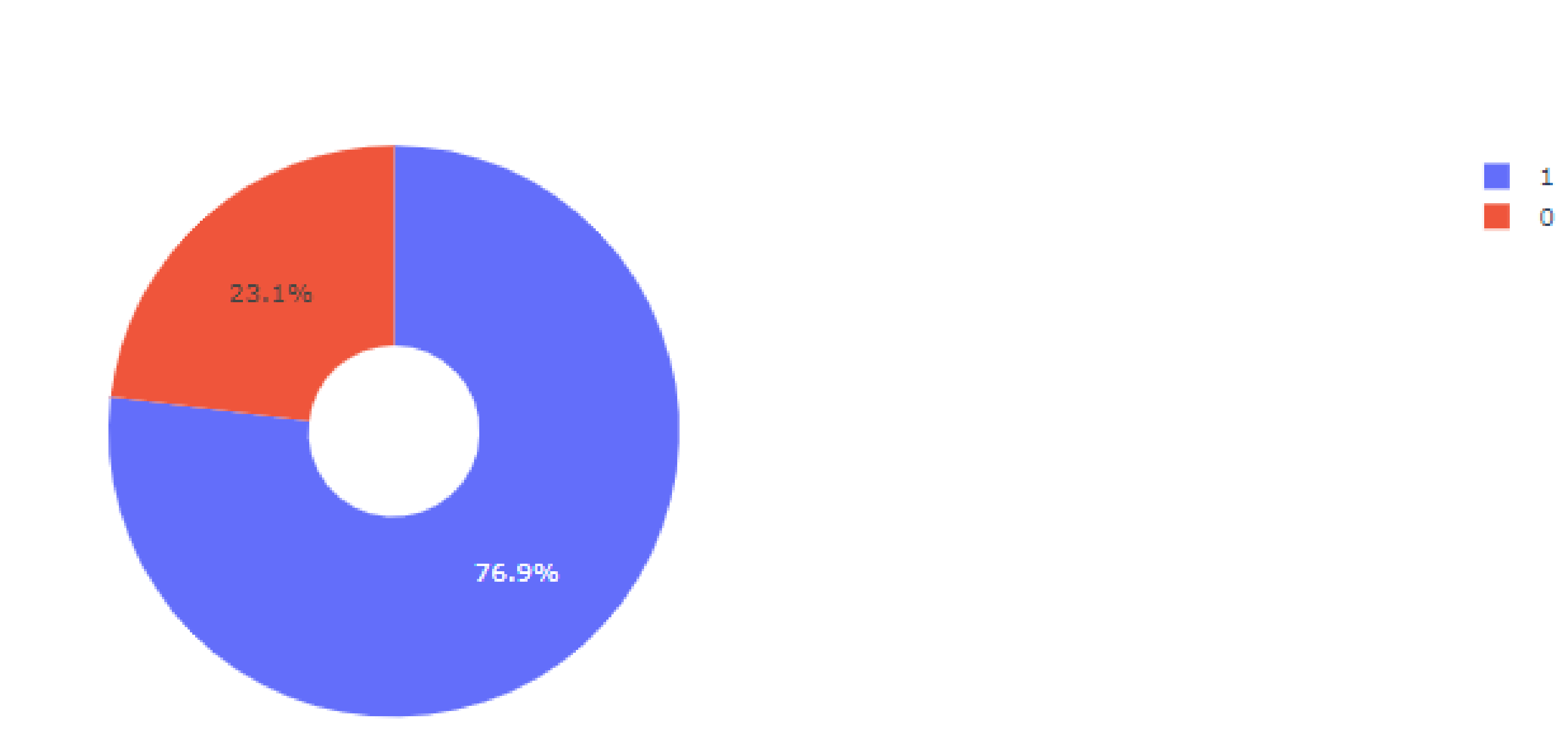
Launch success across launch sites

- Data shows that the KSC LC-39A has the biggest percentage of successful launches overall.



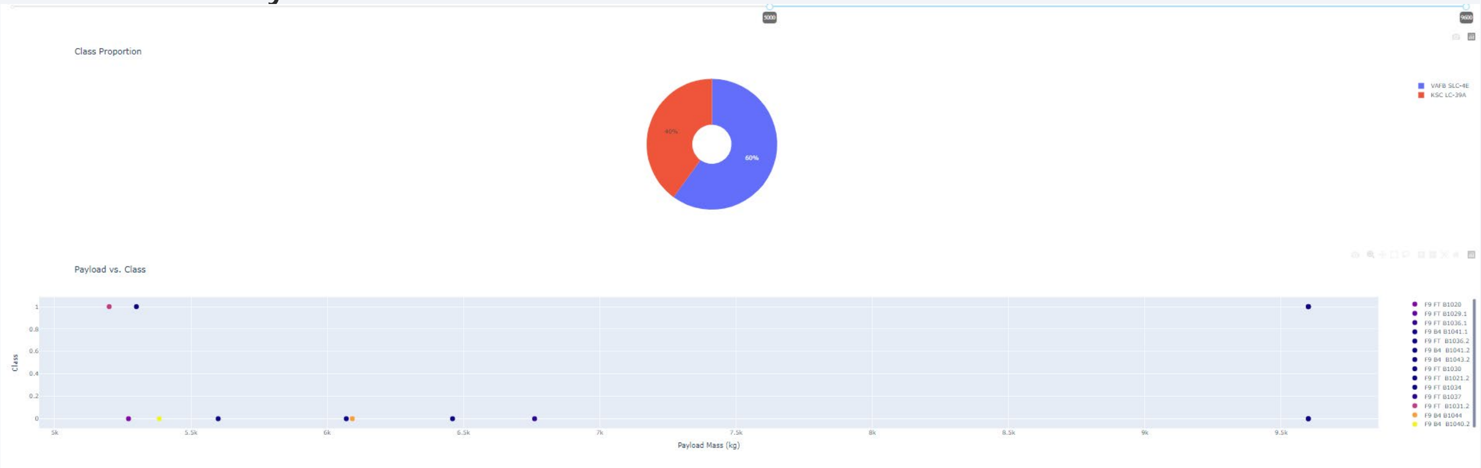
KSC LC has the highest success rate

- KSC LC has the highest success rate at 76.9%

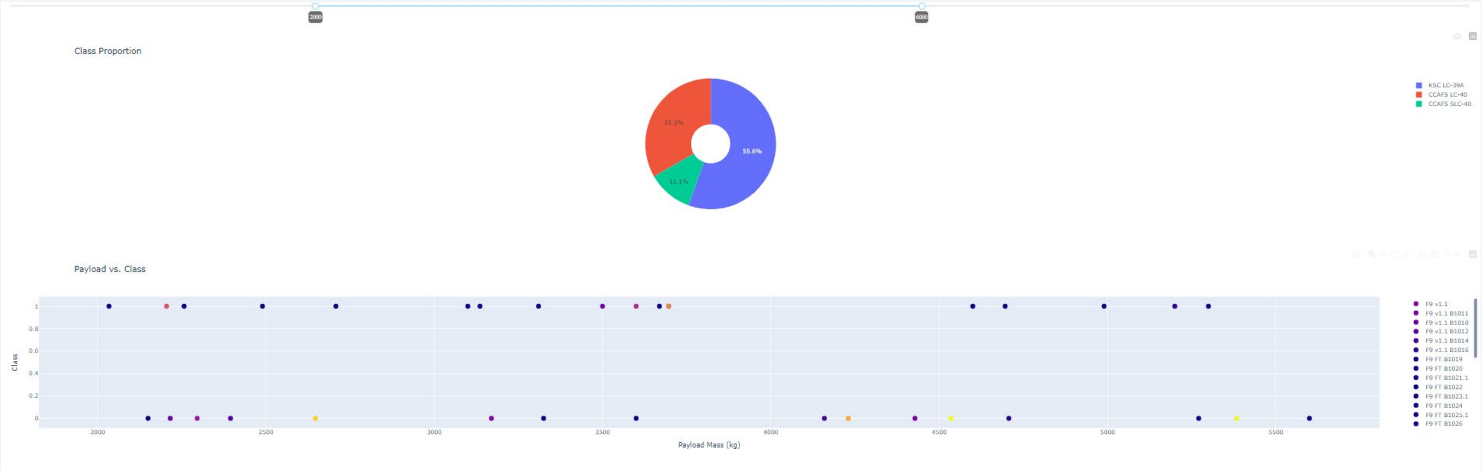


Payload vs. Launch Outcome scatter plot

- Higher payloads on VAFB SLC-4E and KSC LC-39A only



- Lower payloads

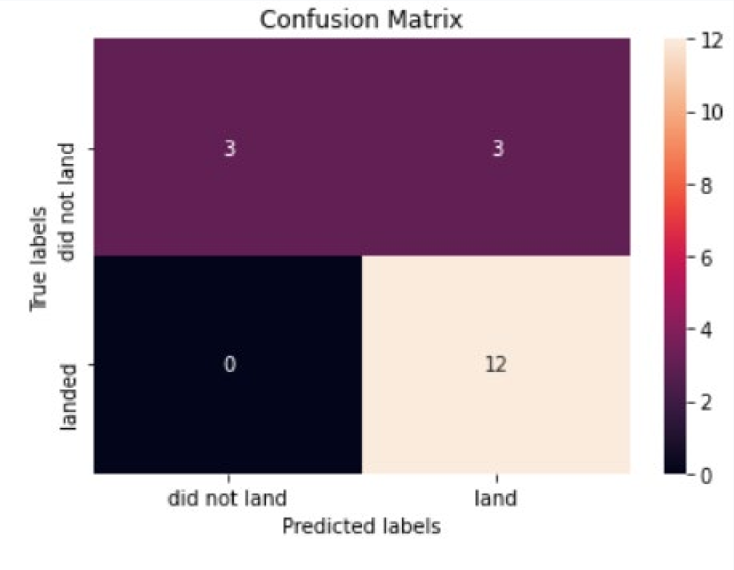


Section 5

Predictive Analysis (Classification)

Confusion Matrix

- Decision tree classifier object has the highest accuracy. The issue is the high number of false predicted successful landings.



Conclusions

- The Decision tree classifier is the best machine learning algorithm for this task.

Appendix

- Git Hub URL: IBM Data Cert

Thank you!