SPACE ROCKET LAUNCH PROJECT

OKOH ASARE AMOABENG

SEPTEMBER 20, 2023

OUTLINE

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

EXECUTIVE SUMMARY

- **Summary of Methodology**
  - Data Collection with API & Web Scraping
  - Exploratory Data Analysis (EDA) with Data Visualization
  - Exploratory Data Analysis (EDA) with SQL
  - Interactive Map with Folium
  - Predictive Analysis
- **Summary of Results**
  - Exploratory Data Analysis Results
  - Interactive Maps
  - Predictive Results

INTRODUCTION

- Background and Context

  The objective of this project is to predict if the Falcon 9 first stage will successfully land or not. Different companies have had different costs related to their launch. The cost incurred by SpaceX to launch Falcon 9 was 62 million dollars as stated on their website. Other companies had their costs as high as 165 million dollars. The difference in the costs is mainly due to the fact that SpaceX can reuse the first stage. To determine if the stage will land, we can determine the cost of a launch. This information will be of interest for another company if it wants to compete with SpaceX for a rocket launch.

- Problems you want to find answers
  - What are the main characteristics of a successful or failed landing ?
  - What are the effects of each relationship of the rocket variables on the success or failure of a landing ?
  - What are the conditions which will allow SpaceX to achieve the best landing success rate ?

METHODOLOGY

Summary

Two methods were used in data collection. The data was cleaned and exploratory data analysis (EDA) was performed on it to provide much insight.

- Data Collection Methods:
  - SpaceX REST API
  - Web Scrapping from Wikipedia
- Data Cleaning/Wrangling:
  - Dropping unnecessary columns
  - One Hot Encoding for classification models
- Performance of EDA using Visualization and SQL
- Performance of Interactive Visual Analytics using Folium
- Performance of Predictive Analysis using Classification

DATA COLLECTION

The datasets were collected from Rest SpaceX API and through Web Scrapping from Wikipedia.

The information obtained by API include payload information, rocket and success/failure outcome.

Information obtained from Web Scrapping of Wikipedia include landing, launches and payload information.

DATA COLLECTION WITH API

The stages in obtaining the data are explained below;

- Requesting rocket launch data from SpaceX API with the following URL: https://api.spacexdata.com/v4/launches/past
- Convert data into a dataframe
- Filter data to include only Falcon9
- Remove missing values
- And finally convert the output to csv

DATA COLLECTION WITH WEB SCRAPING

The data was obtained from Wikipedia using Web Scraping. Activities performed to obtain our results are as follows;

- Performed an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.
- Created a BeautifulSoup object
- Extract all columns from the HTML table header
- Create a dataframe by parsing
- And finally convert the output to csv file format.

DATA WRANGLING

During data wrangling, the data was loaded and converted to a dataframe, the number of launches were computed, the number of occurrences for each Orbit as well as mission outcome per Orbit were calculated.

The landing outcome for each Orbit was obtained.

After computation and retrieval of all relevant information, the output generated was converted to a csv file format.

EDA WITH DATA VISUALIZATION

During this stage, data was obtained an converted to a dataframe. Relationships were found between certain parameters to gain insight. The relationships were performed using Scatter Graph, Bar Chart and Line Graph. These include

- Scatter Graph
  - Payload Mass vs Flight Number
  - Launch Site vs Flight Number
  - Launch Site vs Payload Mass
  - Orbit vs Flight Number
  - Orbit vs Payload Mass
- Bar Chart
  - Orbit vs Class
- Line Graph
  - Average Success Rate over time

EDA WITH SQL

SQL queries were performed to gather and understand insights in the dataset. Tasks performed here includes;

- 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
- Listing the date when the first successful landing outcome in ground pad was acheived.
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- Listing the failed landing outcomes in drone ship with other features.
- Ranking the count of landing between the date 2010-06-04 and 2017-03-20, in descending order.

INTERACTIVE VISUAL ANALYSIS WITH FOLIUM

The Folium map was created with an initial center location to be NASA Johnson Space Center at Houston, Texas. The following activities were performed to generate the Maps;

- We marked all launch sites on the map
- The success and failed launches for each site was marked on the map
- Distances between launch sites to their proximities were computed

Plotting the distance lines of the launch sites to their proximities helped in answering the following questions;

- Are launch sites in close proximity to railways?
- Are launch sites in close proximity to highways?
- Are launch sites in close proximity to coastline?
- Do launch sites keep certain distance away from cities?

PREDICTIVE ANALYSIS USING CLASSIFICATION MODEL

The dataset obtained was converted into a dataframe and normalized. The dataset was split into training and testing sets. Further, the accuracy of the dataset was ascertained using four methods before their respective confusion matrices were developed.

The four methods include;

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K-Nearest Neighbors

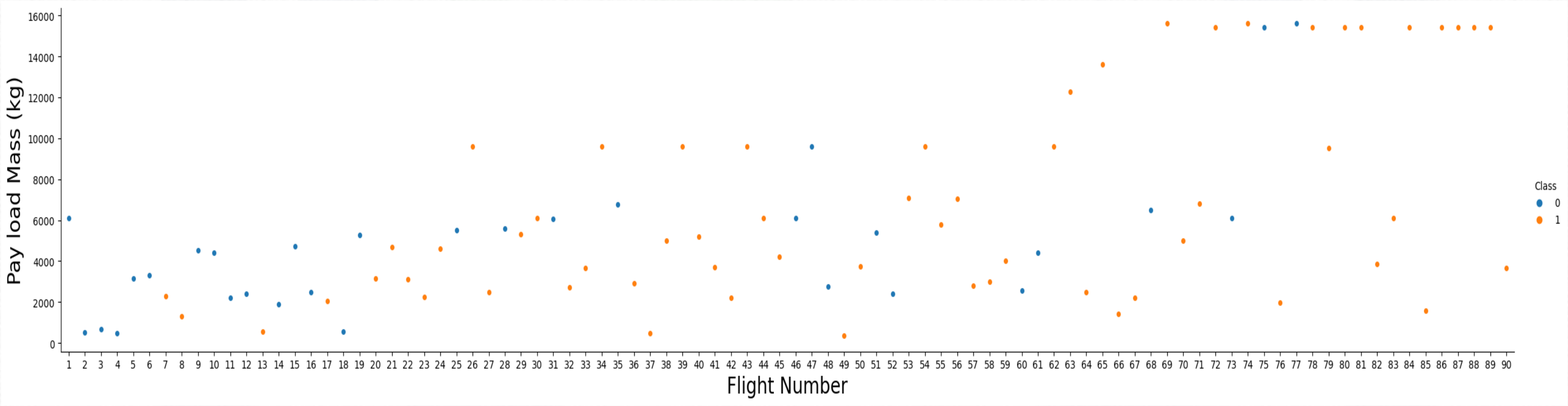
RESULTS

- Results on the Exploratory Data Analysis
- Results on Interactive Visual Analytics using Folium
- Results on Predictive Analysis using Classification

NB: The insights drawn from the analysis above with visuals have been captured in the subsequent slides.

INSIGHTS ON EDA WITH DATA VISUALIZATION

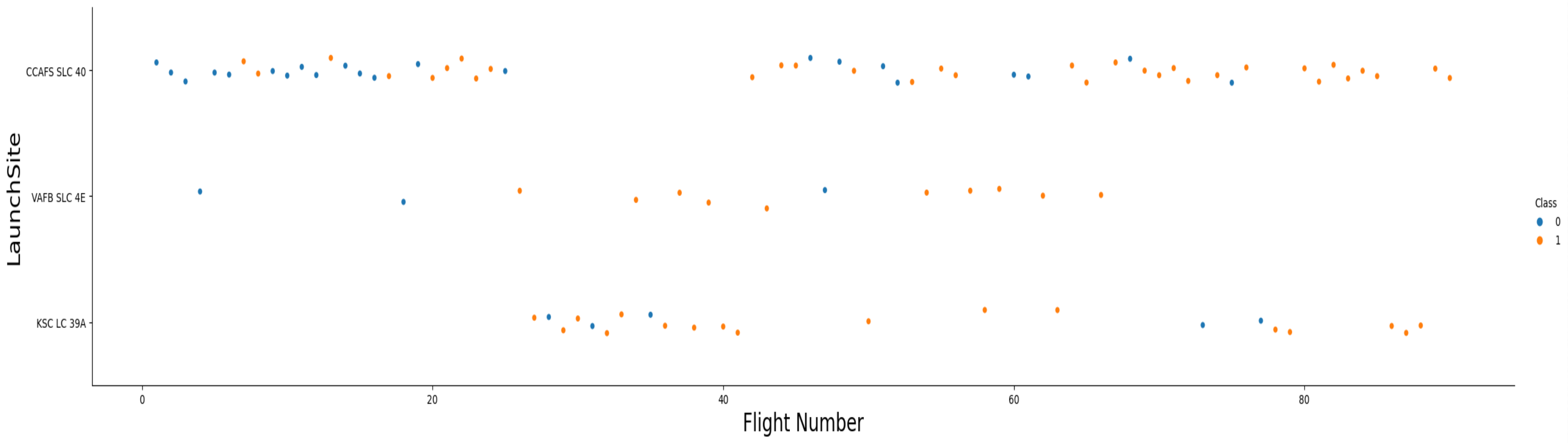
(Flight Number vs Payload Mass)



You can tell from the graph that, the launch success increased as the payload mass increase.

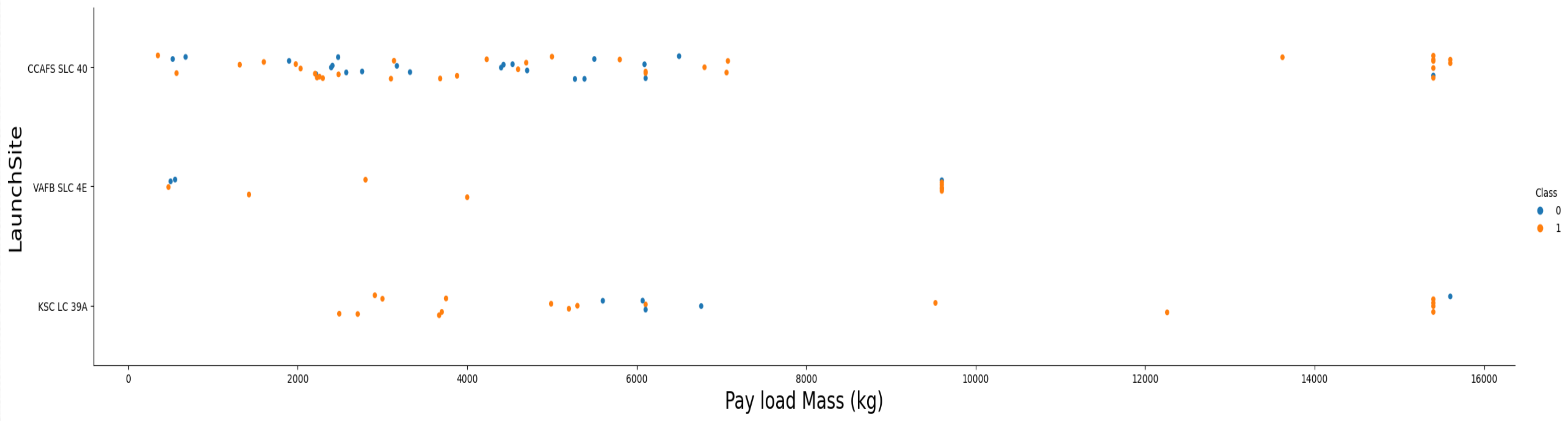
INSIGHTS ON EDA WITH DATA VISUALIZATION

(Flight Number vs Launch Site)



The launch sites had more success as the flight number increased

INSIGHTS ON EDA WITH DATA VISUALIZATION
(Launch Site vs Payload Mass)

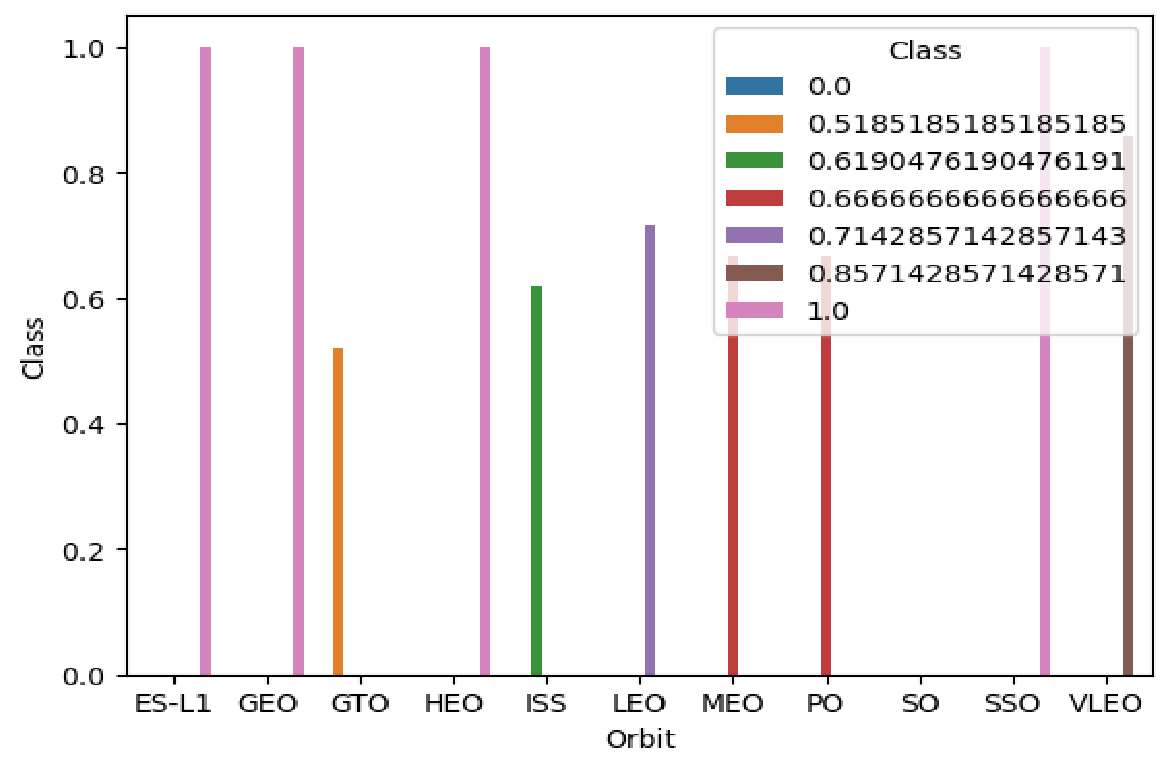


The launch site, CCAFS, initially had a mixed outcome when the payload mass was low but had success at very high payload mass. The other two sites also had inconsistent outcome.

IBM Developer

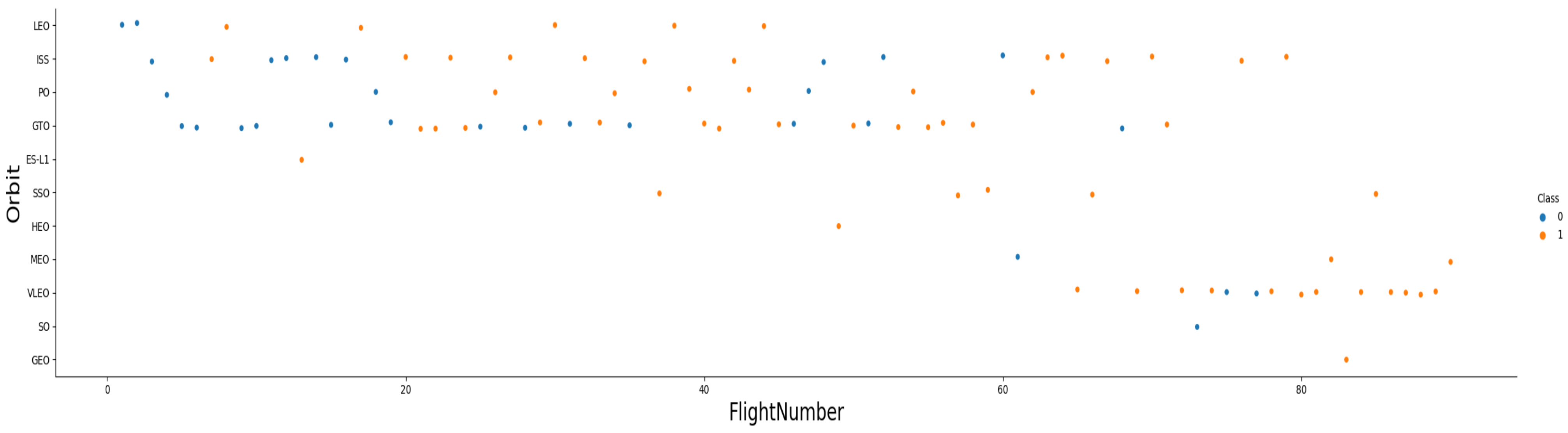
SKILLS NETWORK

INSIGHTS ON EDA WITH DATA VISUALIZATION
(Class vs Orbit)



Four out of ten orbits had the best success rate.

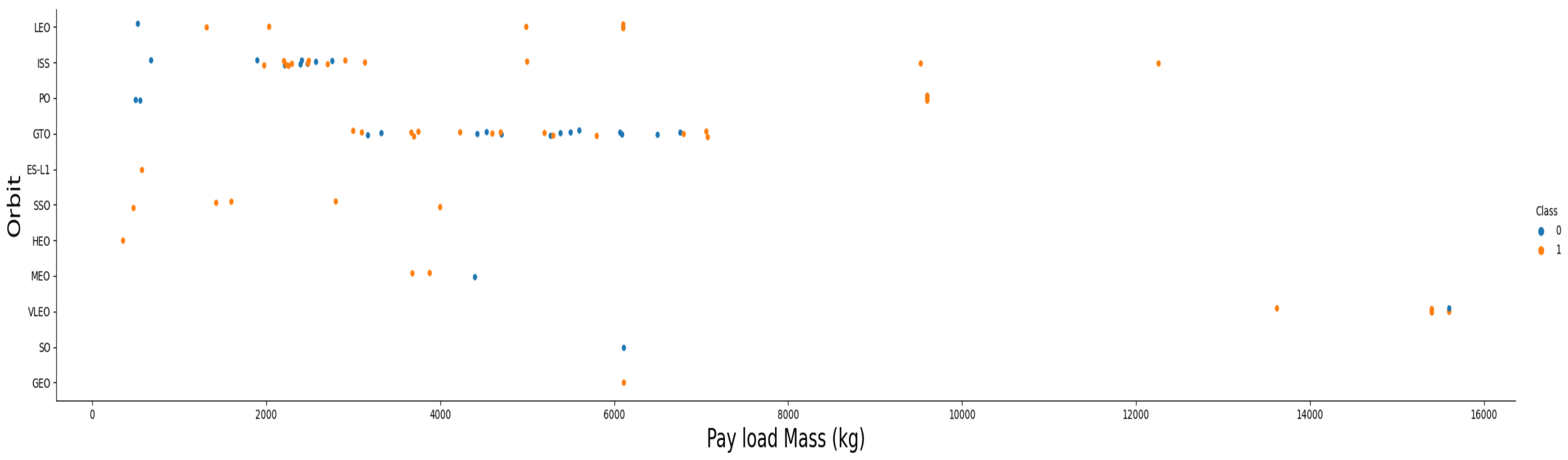
INSIGHTS ON EDA WITH DATA VISUALIZATION
(Flight Number vs Orbit)



The success rate increased as the flight number increased

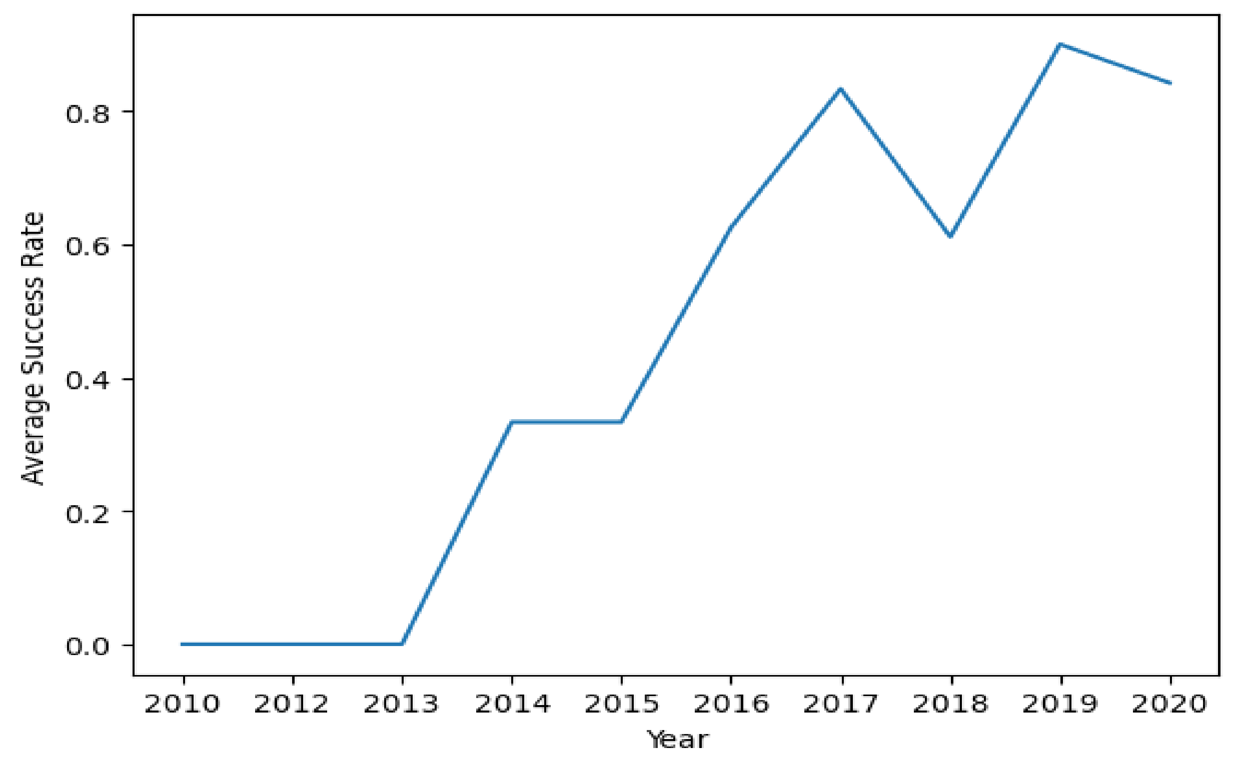
INSIGHTS ON EDA WITH DATA VISUALIZATION

(Payload Mass vs Orbit)



ES-LI, SSO and HEO had better success rate when payload mas was low. The other orbits had mixed results.

INSIGHTS ON EDA WITH DATA VISUALIZATION
(Average Success Rate)



The average success rate started to improve after 2013 and reached its highest in 2019.

IBM Developer

SKILLS NETWORK

INSIGHTS ON EDA WITH SQL - 1

Display the total payload mass carried by boosters launched by NASA (CRS)

[20]:
```
%sql SELECT SUM(PAYLOAD_MASS_KG_) AS payloadmass FROM SPACEX;
```

```
 * ibm_db_sa://vmj88629:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd
Done.
```

[20]:

payloadmass
619967

Task 4

Display average payload mass carried by booster version F9 v1.1

[22]:
```
%sql select avg(PAYLOAD_MASS_KG_) as payloadmass from SPACEX;
```

```
 * ibm_db_sa://vmj88629:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1og
Done.
```

[22]:

payloadmass
6138

The information above shows the total and average payload mass carries by Boosters launched by NASA

INSIGHTS ON EDA WITH SQL - 2

```
]:  %sql select min(DATE) from SPACEX;

 * ibm_db_sa://vmj88629:***@6667d8e9-9d4d-4c
Done.
```

1
2010-06-04

```
]:  %sql select BOOSTER_VERSION from SPACEX where LANI

 * ibm_db_sa://vmj88629:***@6667d8e9-9d4d-4ccb-ba32-
Done.
```

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

The first image shows the date we recorded the first success whilst the second shows the boosters that were successful with a payload mass between 4000 and 6000.

INSIGHTS ON EDA WITH SQL - 3

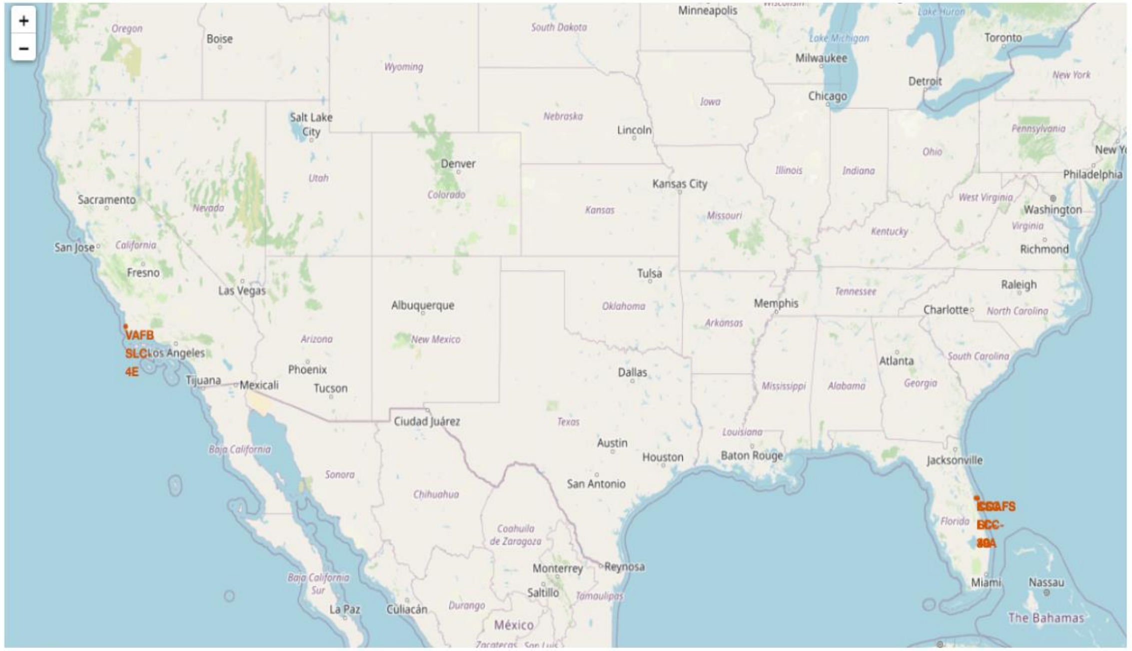
```
[33]: %sql SELECT DATE,MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM S

 * ibm_db_sa://vmj88629:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1og
Done.
```

[33]:

DATE	mission_outcome	booster_version	launch_site
2015-01-10	Success	F9 v1.1 B1012	CCAFS LC-40
2015-02-11	Success	F9 v1.1 B1013	CCAFS LC-40
2015-03-02	Success	F9 v1.1 B1014	CCAFS LC-40
2015-04-14	Success	F9 v1.1 B1015	CCAFS LC-40
2015-04-27	Success	F9 v1.1 B1016	CCAFS LC-40
2015-06-28	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
2015-12-22	Success	F9 FT B1019	CCAFS LC-40

The mission outcome of drone ships launched in 2015

INSIGHTS ON INTERACTIVE ANALYSIS WITH FOLIUM



The launch sites are located on the south eastern and western part of USA.

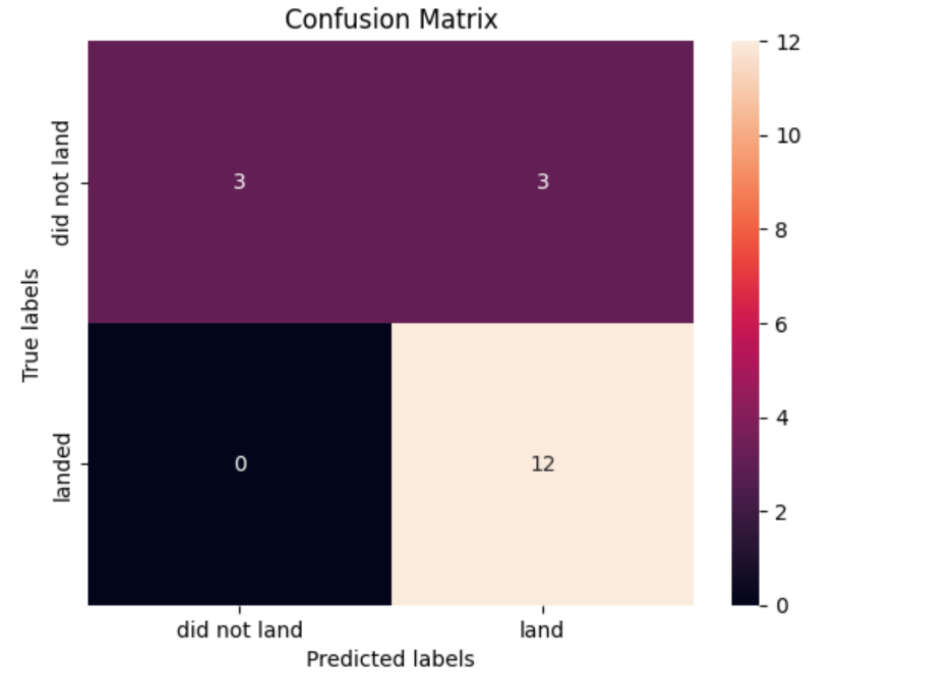
PREDICTIVE ANALYSIS WITH CLASSIFICATION

The test accuracy and confusion matrices were all equal when we used four different methods. From the confusion matrices, the main problem was the false positives. The methods used to check the accuracy are;

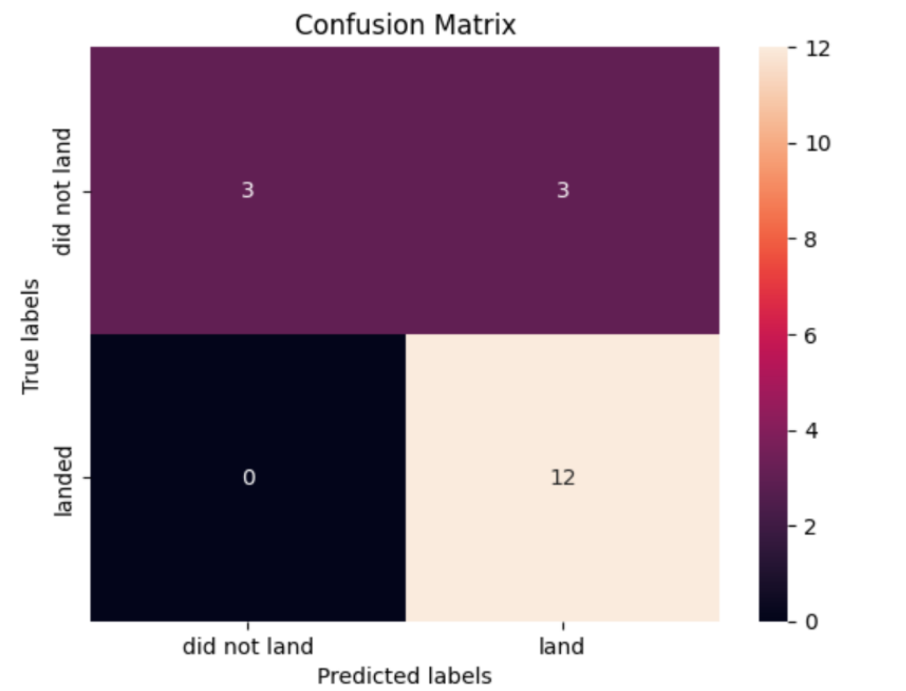
- k-Nearest Neighbors Algorithm (KNN)
- Decision Tree
- Support Vector Machines (SVM)
- Logistic Regression

The next two slides shows the image of the confusion matrices.

KNN & SVM

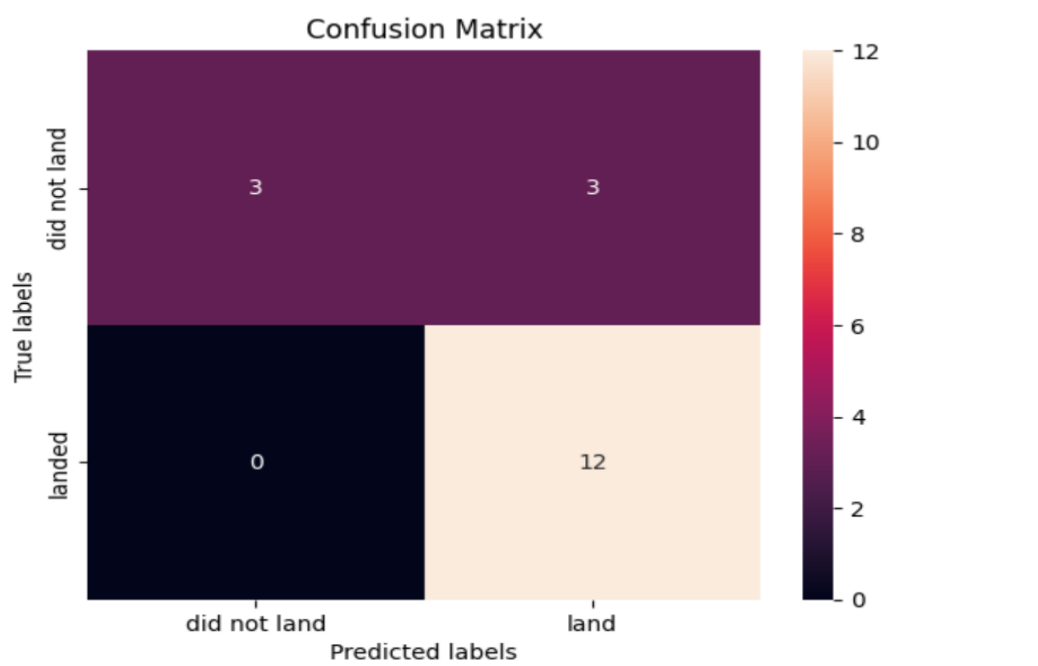


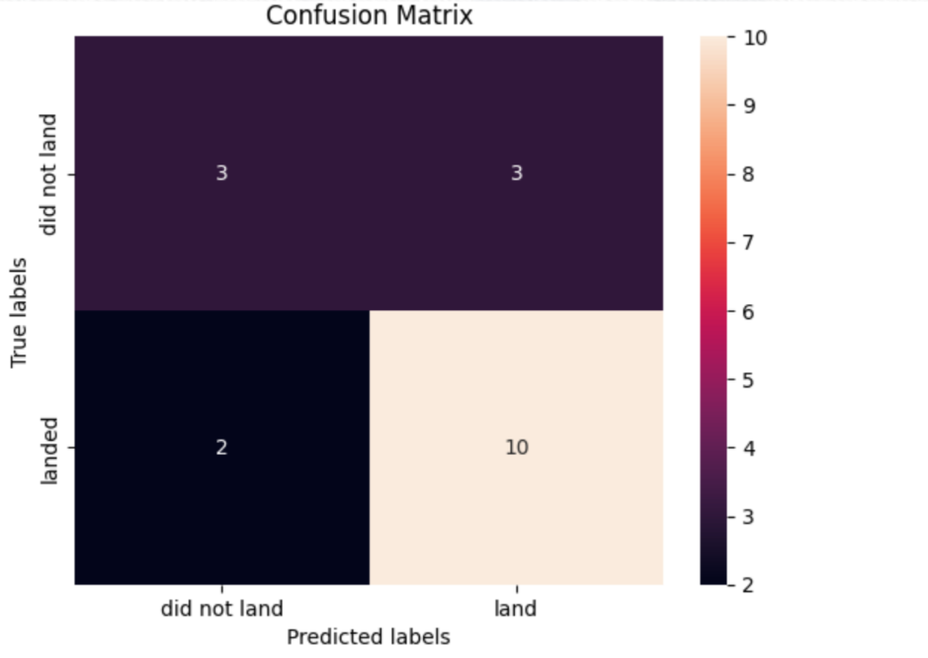
Confusion Matrix
True labels
did not land
landed
Predicted labels
did not land
land
3
3
0
12
0
2
4
6
8
10
12



Confusion Matrix
True labels
did not land
landed
Predicted labels
did not land
land
3
3
0
12
0
2
4
6
8
10
12

LOGISTIC REGRESSION & TREE





IBM Developer

SKILLS NETWORK

CODE USED FOR PLOTLY DASH

<IBM-Data-Science-Captsone-2023/Plotly_Dash_Dashboard_Resultc.py at main · okohasare/IBM-Data-Science-Captsone-2023 (github.com)>

CONCLUSION

From the insight gained from the visuals and analyses, we can say the orbits with the best success rates are GEO, HEO, SSO and ESL1. The missions' success can be explained by combination of several factors such as the launch site, the orbit, payload mass and so on.

From observation, we can say that the payload mass is one of the major criterion to take into account for the success of a mission. Some orbits require a light payload mass to have a successful launch, others require a heavy payload mass. But generally, orbits with low payload masses perform better than those with heavy payload masses.

The Decision Tree Algorithm had the best train accuracy and we can say it is the best model in determining accuracy even though the test accuracy between all the models used is identical.

The current data cannot give an insight as to why some launch sites are better than the others. Additional information on say, the experience and skill of the scientists performing the launch or weather conditions can also help answer this question.

IBM Developer

SKILLS NETWORK

Thank you.