SKILLS NETWORK

Falcon 9 First Stage Success Prediction

Mai Aljuaid

22/06/2022

IBM Developer

OUTLINE

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

Executive Summary

- Summary of Methodology
  - Data Collection from API and Web Scrapping using Beautiful Soup
  - EDA with Python using pandas and matplotlib
  - EDA with SQL
  - Interactive Map with Folium
  - Interactive Dashboard
  - Predictive Analysis
- Summary of Results
  - EDA results
  - Interactive Map and dashboard
  - Predictive results

Introduction

- Predicting whether the first stage of the falcon 9 will land successfully is the goal of this research.

- There is a significant difference between the cost of SpaceX launch in comparison with other companies. The reason is that SpaceX reuses the first stage “if successfully launched”

- This project will help other companies who are willing to compete with SpaceX in determining the cost of launch by predicting whether Falcon 9 land or not.

Methodology

- Data Collection
- Data Wrangling
- Exploratory Data Analysis
    - SQL
    - Pandas & Matplotlib
- Interactive Dashboards
- Predictive Analysis

Methodology – Data Collection

- Data collection phase consisted of two ways to collect data (“SpaceX REST API + Web Scrapping from Wikipedia
- SpaceX REST API > The data is returned in a JSON file which then was turned to data frame using pandas.
- Web Scrapping > Data was extracted using Beautiful Soup library in Python

Methodology – Data Wrangling

- Identified and calculating the percentage of null values in the dataset.
- Determined categorical and numerical variables.
- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculated the number and occurrence of mission outcome per orbit type
- Created a landing outcome label from Outcome column

Methodology – EDA - SQL

- Connected to database in order to perform several queries which helped to better understand the data
- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites begin with the string 'CCA’
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first succesful landing outcome in ground pad was achieved
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- Listed the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Ranked the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

Methodology – EDA – Pandas/Matplotlib

- Feature Engineering:
- Create dummy variables to categorical columns
- Cast all numeric columns to float
- Data was visualized using matplotlib to create bar graphs, scatter plots and line graphs:
- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload and Orbit type
- Visualize the launch success yearly trend

Methodology – Interactive Dashboards - Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

- This was used to find geographical patterns about launch sites and identify successful and failure launch sites.

Methodology – Interactive Map – Plotly Dashboard

- Dashboard created has many components to it. Dropdown menu, pie charts, range slider, and bar graphs.
- Dropdown menu to choose launch site or select ALL.
- Pie chart illustrates success and fail total depending on chosen site.
- Scatter plot shows the relationship between the variables (Success and Payload Mass)
- Range slider to choose the payload mass.

Methodology – Predictive Analysis

- **Step 1**: Prepare the data for modeling by loading the dataset, normalizing the data, then splitting it up for training and testing.
- **Step 2**: Modeling > to select the right machine learning algorithm (we went with classification since it is a classification problem) > Used grid search to determine the best parameters for each algorithm > Trained the model.
- **Step 3**: Model evaluation was conducted based on the result of the testing set for each model after determining the best hyperparameters for each model. Confusion Matrix was also used for each model.
- **Step 4**: Comparison of Models: this was done by comparing the accuracy of test dataset for each model and the best performing model was chosen

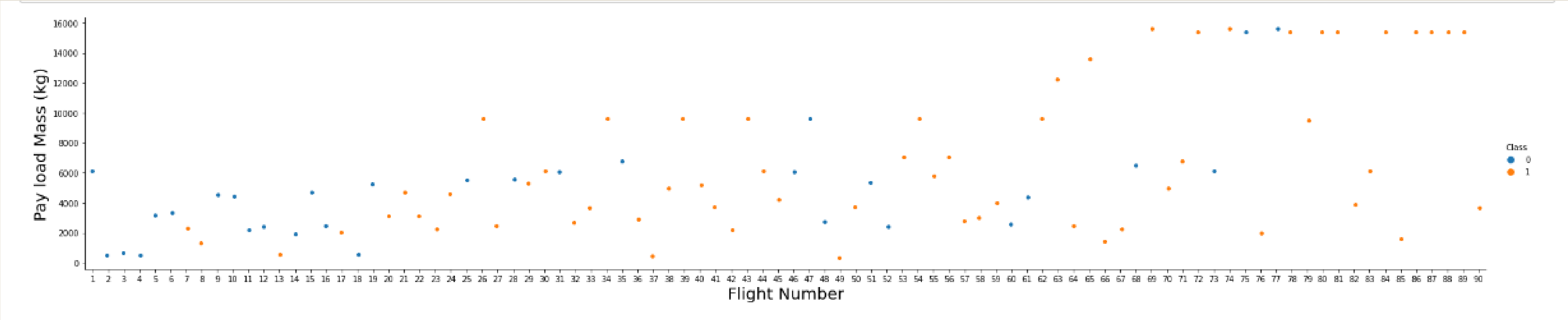
Results

- Exploratory Data Analysis
- Interactive Map & Dashboard
- Predictive Analysis

Exploratory Data Analysis

SKILLS NETWORK

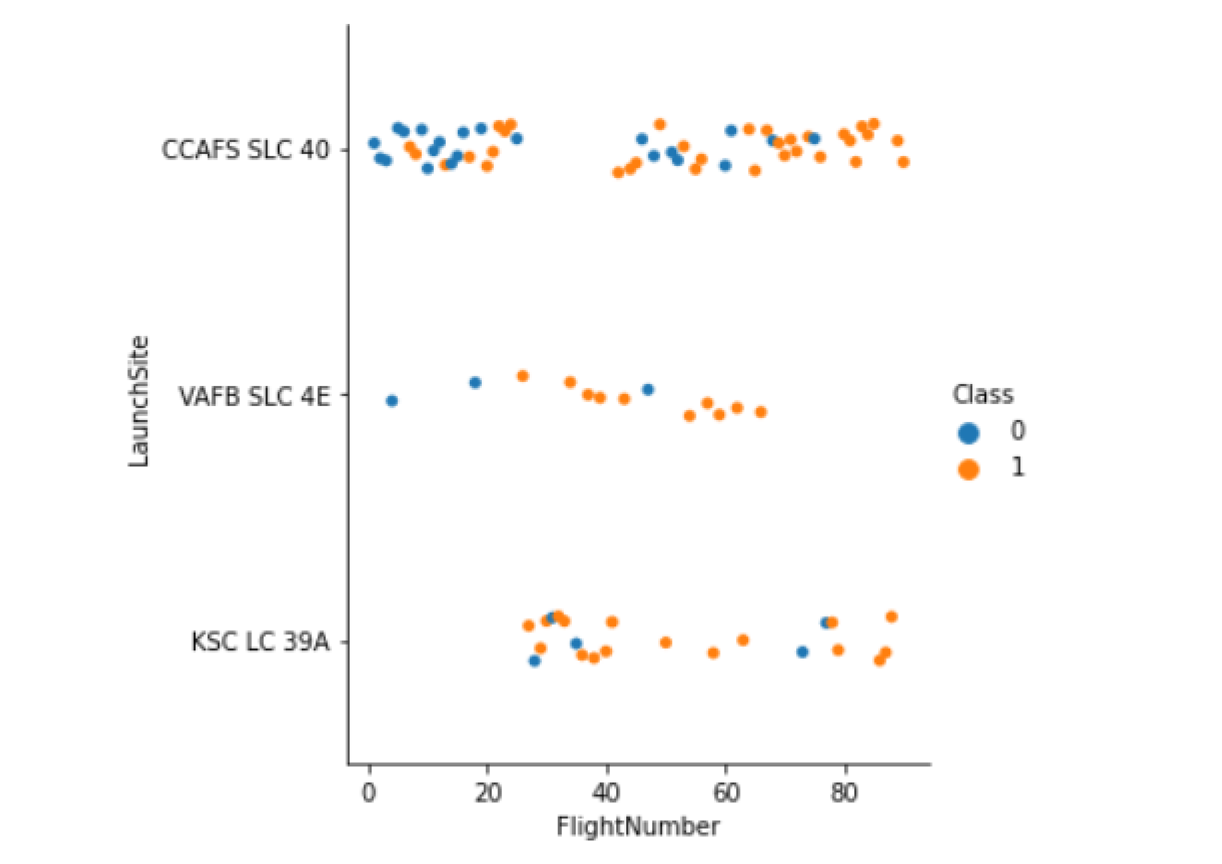
Payload Mass vs. Flights number



This graph illustrates the payload by flight number. We can observe that the heavier the payload the less is the success

IBM Developer

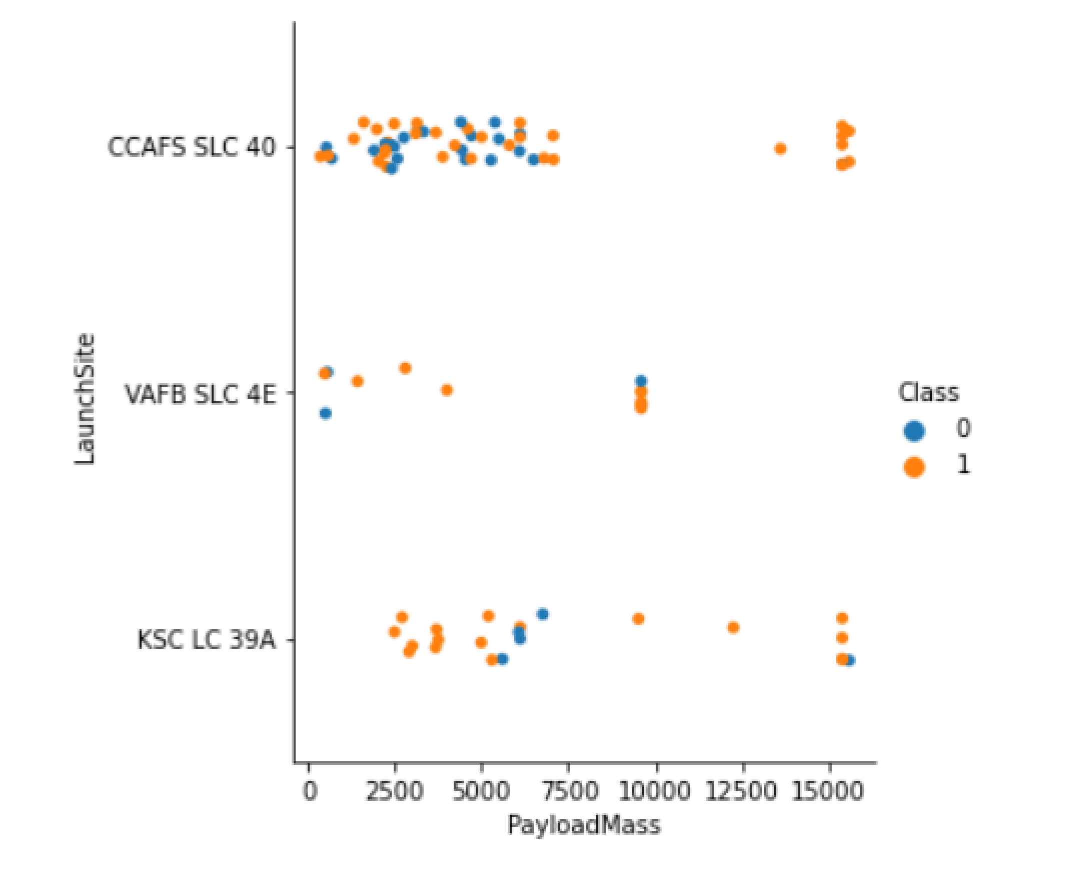
Launch Site vs. Flights number



The graph shows success and failure of launch depending on Launch Site

SKILLS NETWORK

Launch Site vs. Payload Mass

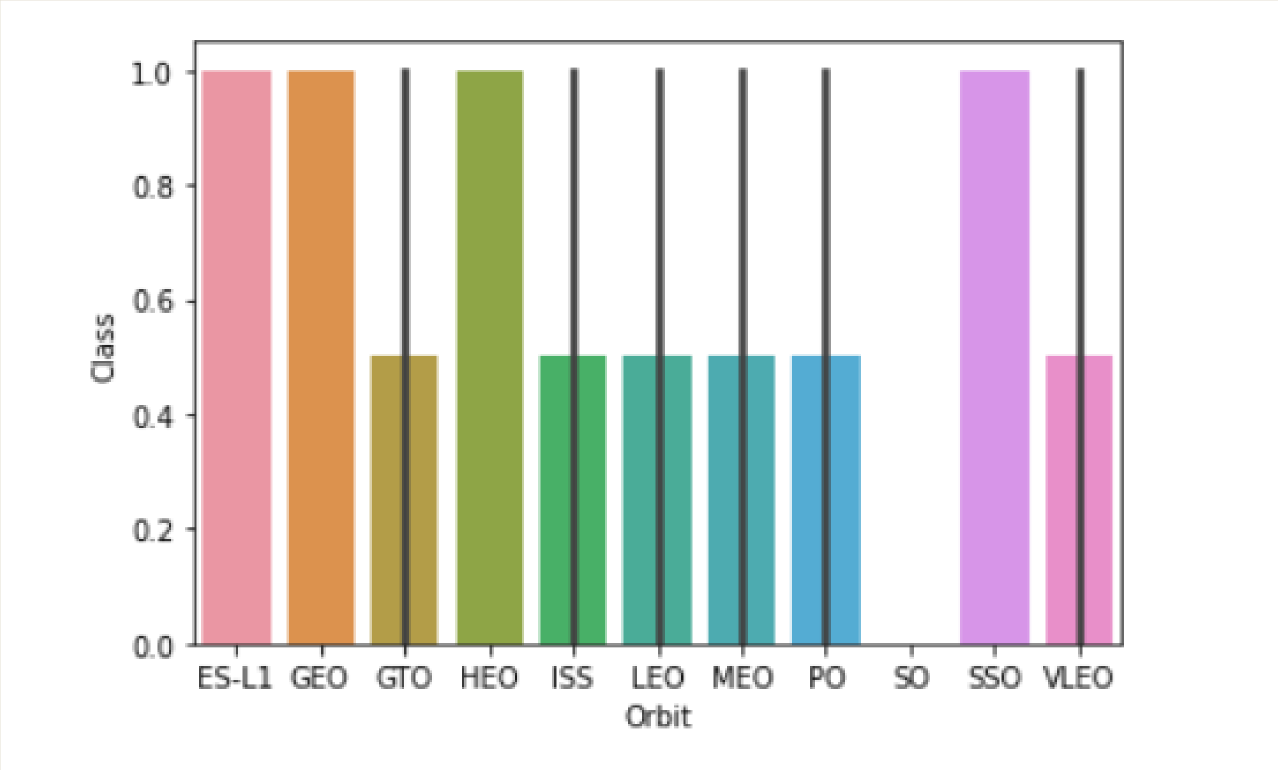


We can observe that depending on the Launch site, we can say the heavier the better success rate.

IBM Developer

SKILLS NETWORK

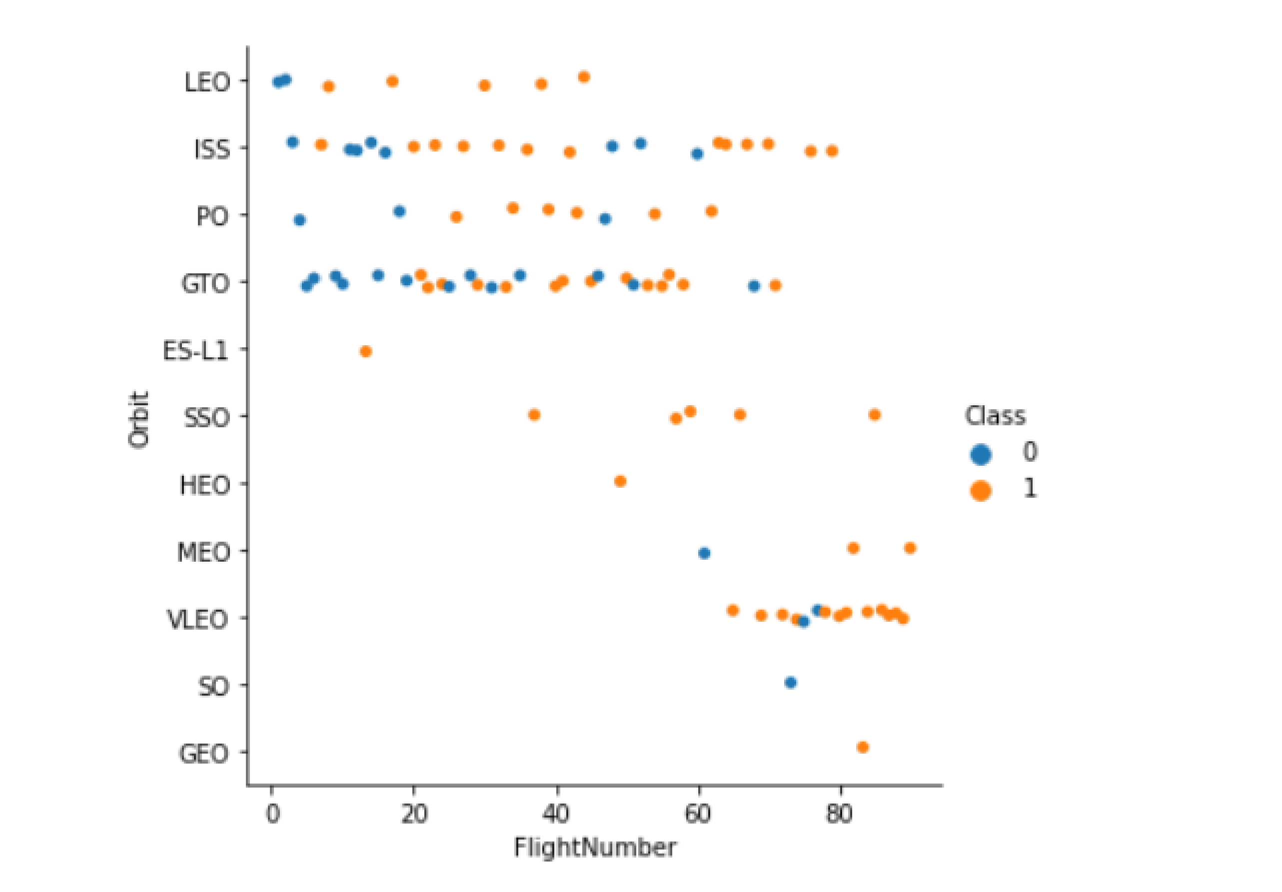
Success Rate per Orbit



This bar graph shows the success rate depending on the Orbit

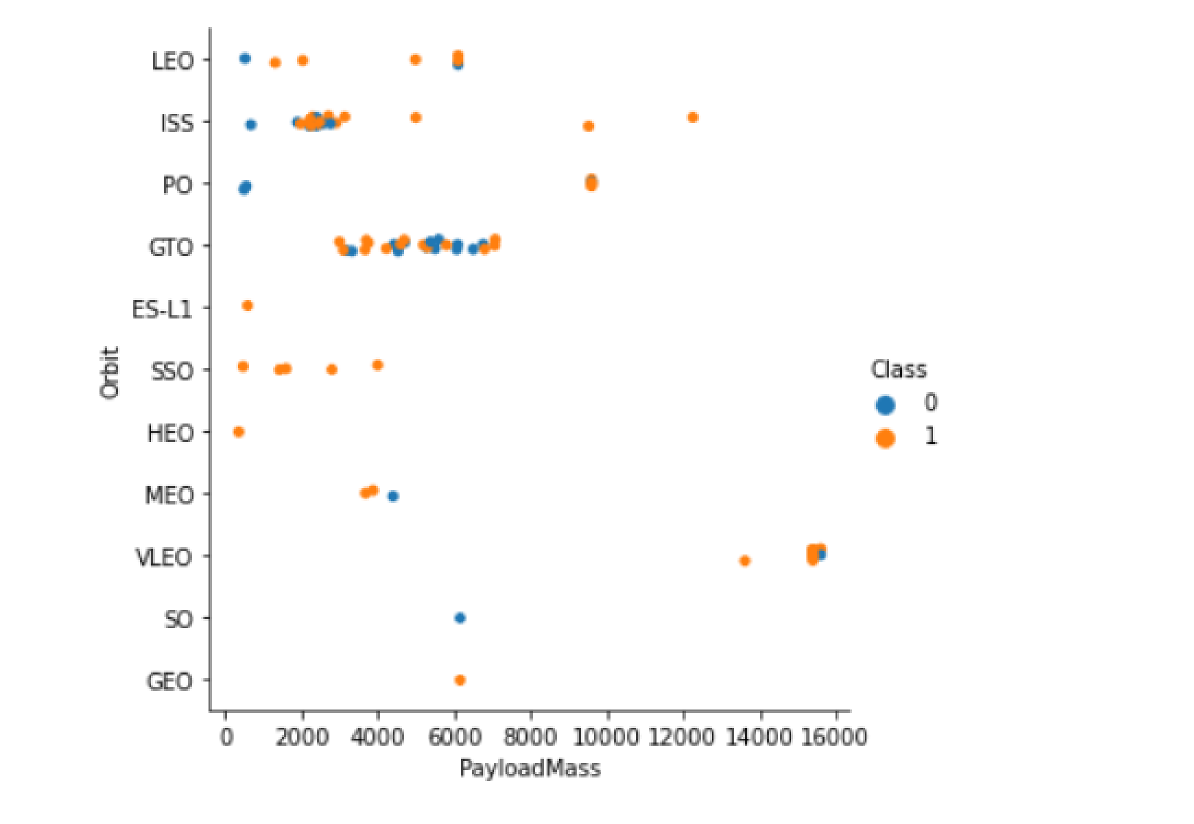
IBM Developer

Orbit vs. Flights number



We can see that there is no direct relationship between the flights number and Orbit.

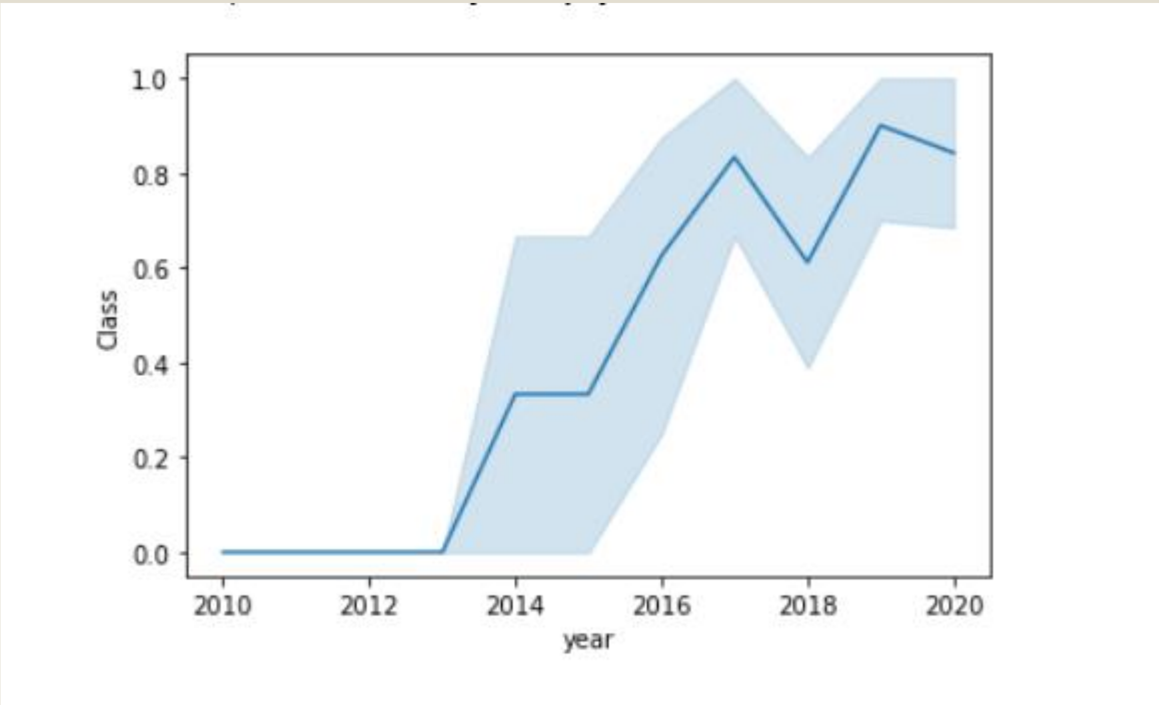
Orbit vs. Payload Mass



Payload mass significantly influences the success of launch depending on the Orbit.

Launch Success by year

- Since 2013, the success rate has increased significantly.



Distinct Launch Site names

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Display launch sites that starts with “CCA” and the first 5 records are displayed

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

SKILLS NETWORK

Total Payload Mass

total_payload_mass
45596

Average Payload Mass

avg_payload_mass
2928

First Successful ground landing date

1
2015-12-22

Total # of Failure and Successful mission outcomes

1
101

IBM Developer

Successful drone ship landing with payload between 4000 and 6000kg

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

Boosters that carried maximum payload

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

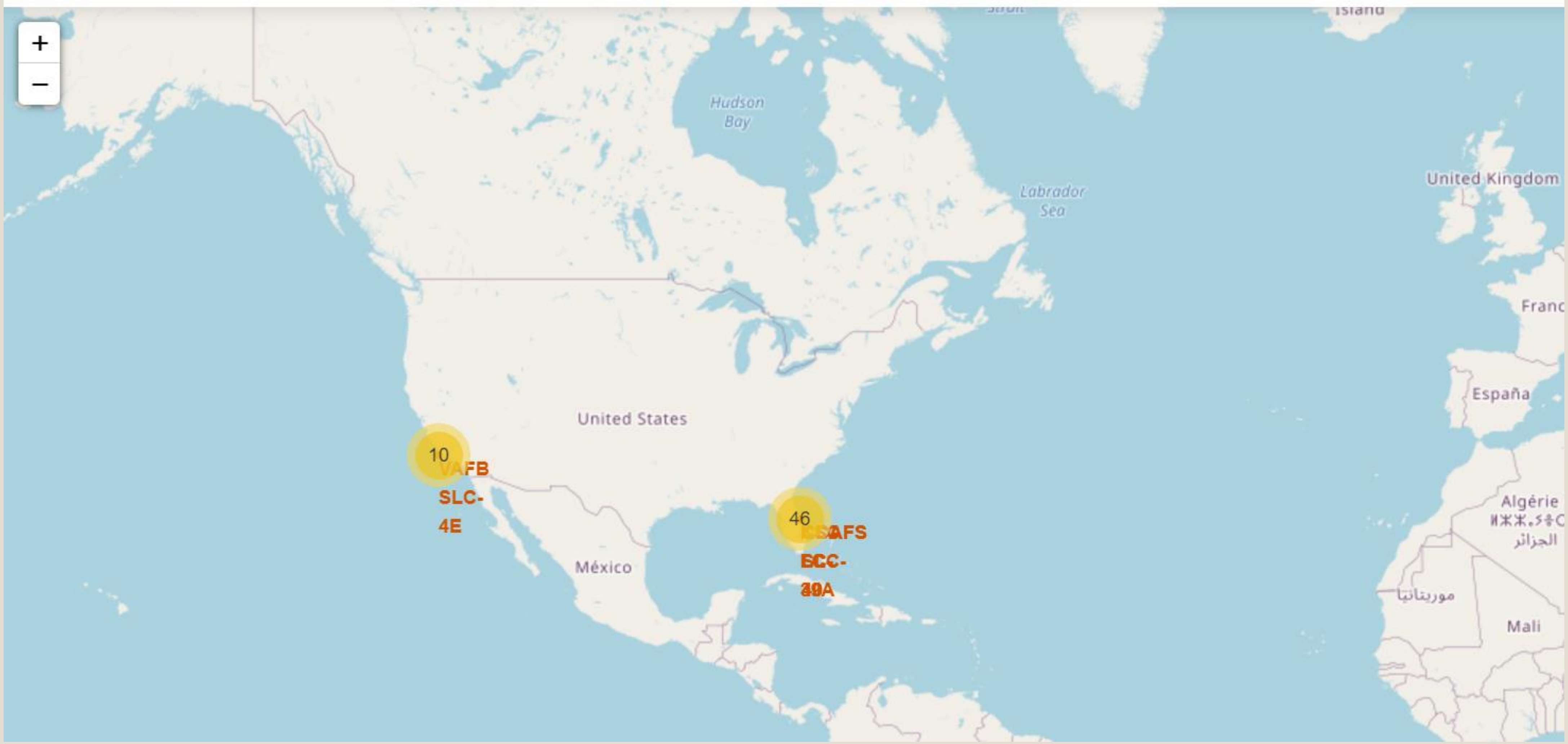
SKILLS NETWORK

2015 Launch Records

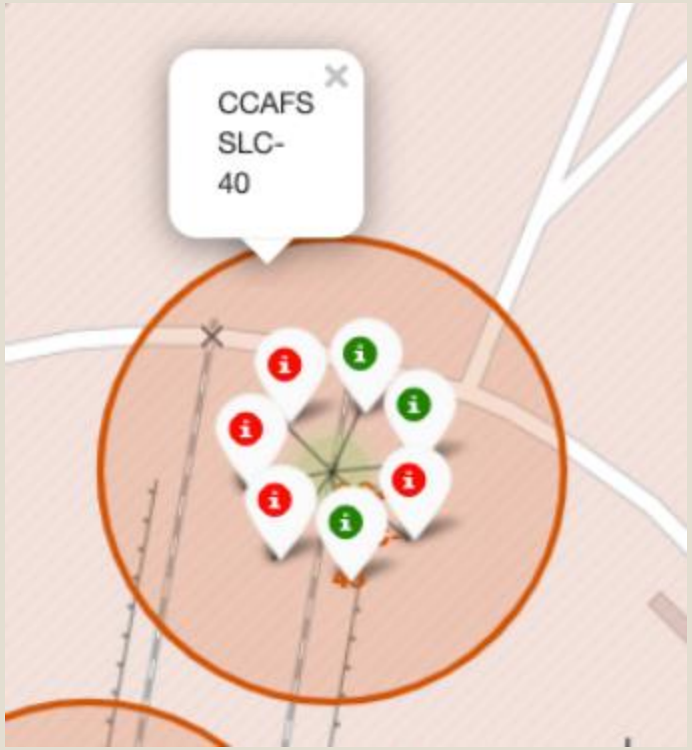
DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

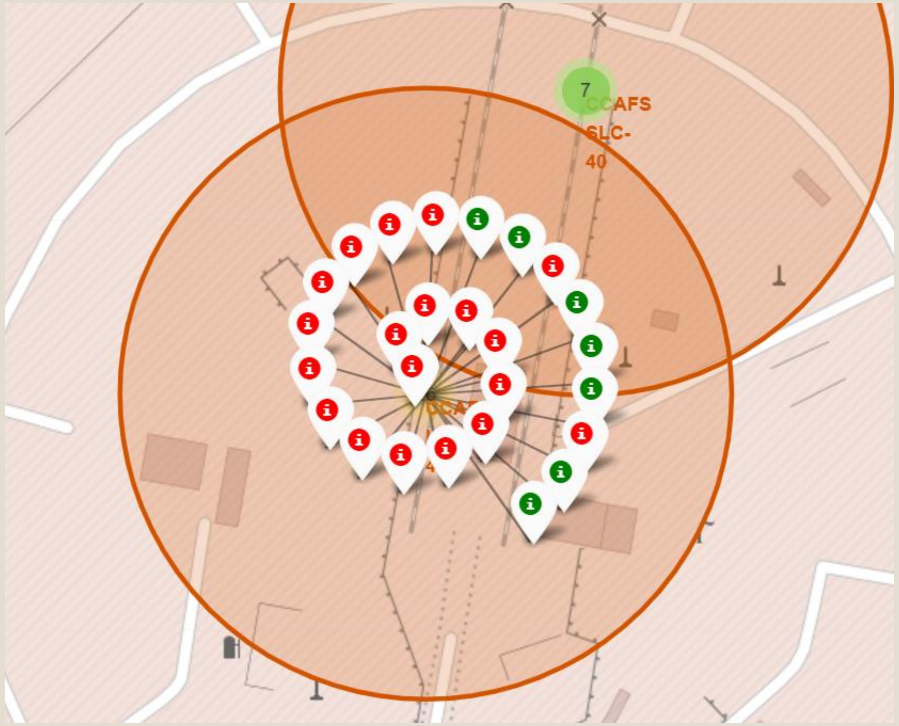
IBM Developer

Interactive Map & Dashboard



Launch sites are on the east and west coasts





The green and red markers here indicate successful and failed launches

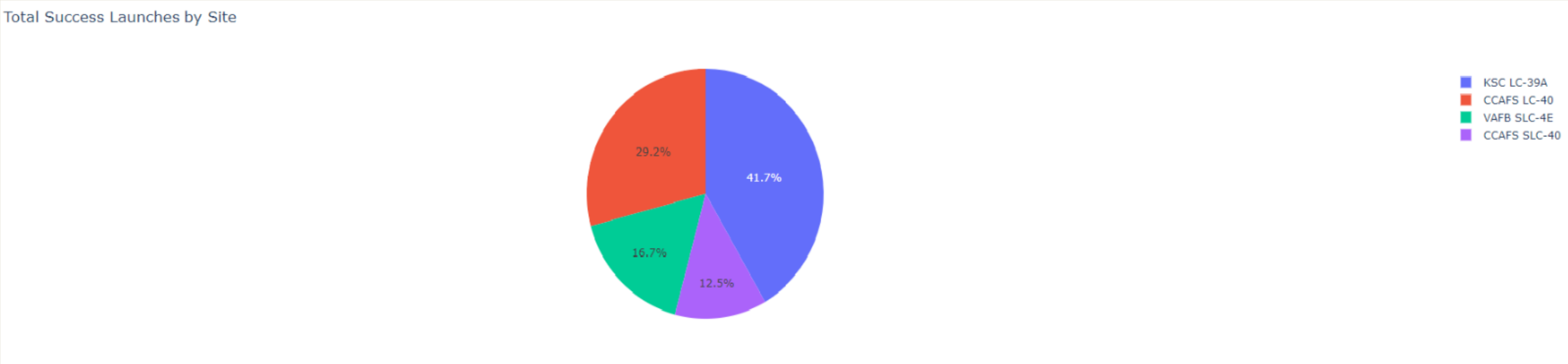
IBM Developer

Distance lines were drawn to illustrate the distance between launch site and other locations

- Are launch sites in close proximity to railways? Yes
- Are launch sites in close proximity to highways? Yes
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? No

IBM Developer

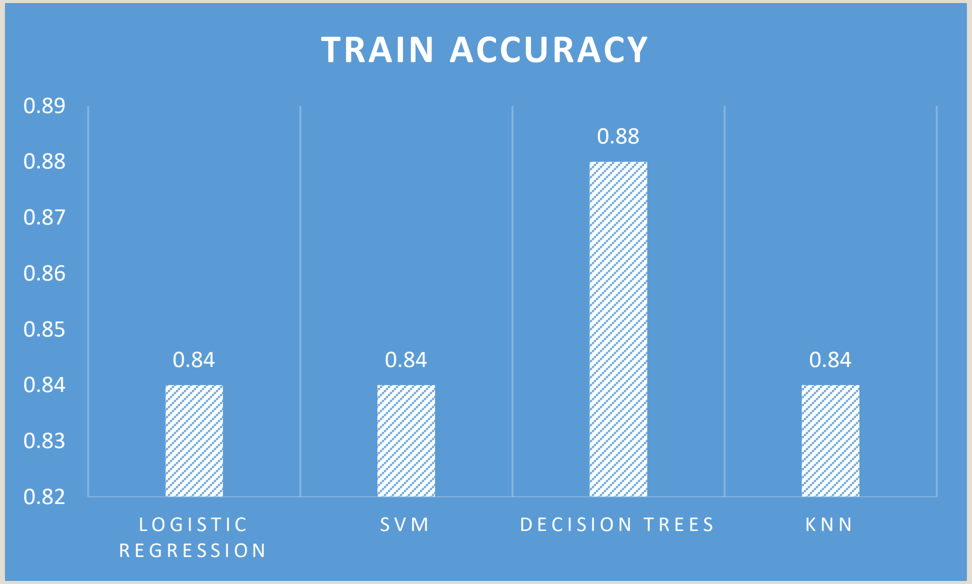
Results – Interactive Dashboards

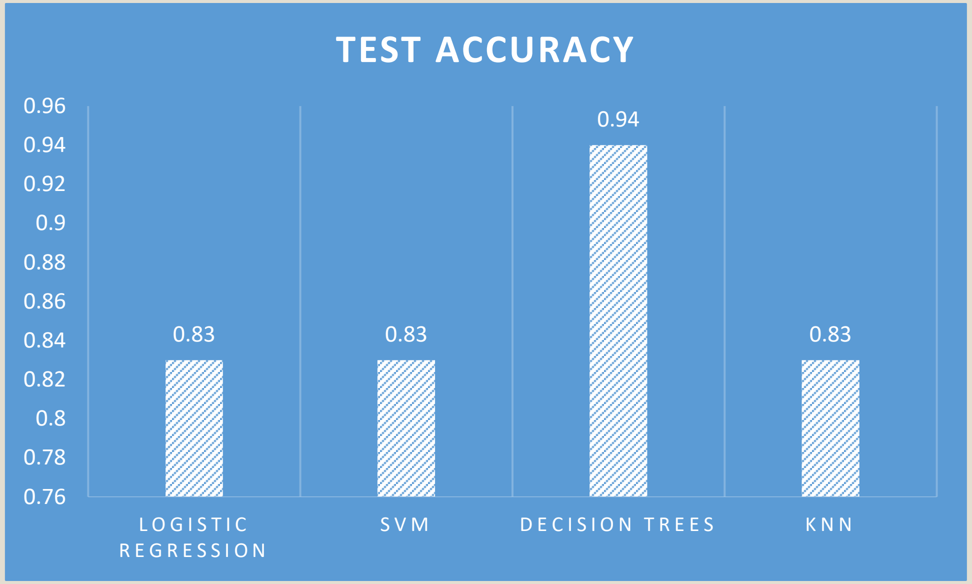


This pie chart is part of the Plotly Dashboard. We can see that KSC LC-39A has the highest success rate.

Predictive Analysis

Results – Predictive Analysis

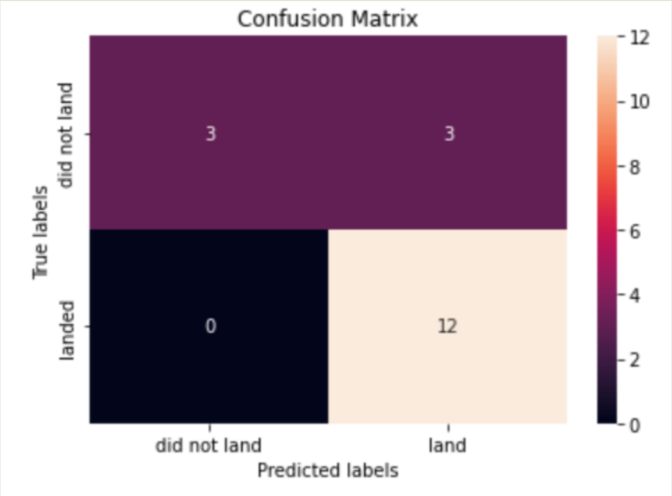




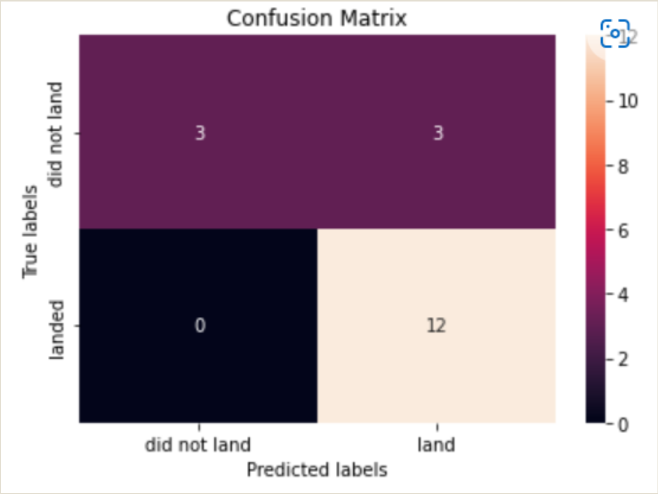
The above bar charts represents the results of the predictive analysis for various classification algorithms. We can observe that the best performing algorithm is “Decision Trees” as it performed best on the training and testing set.

Confusion Matrix

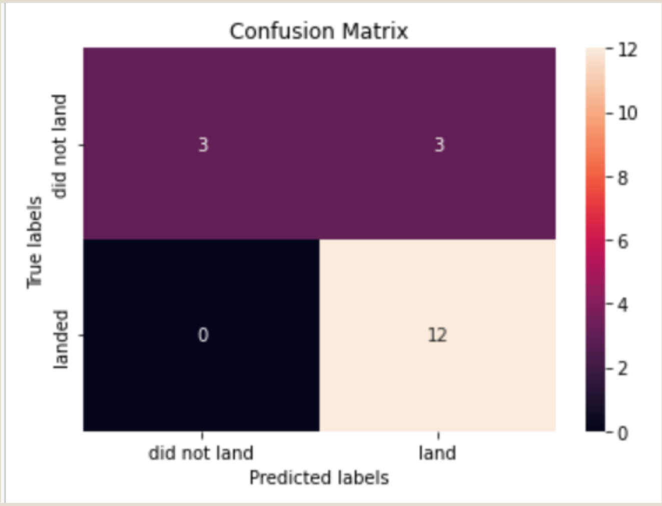
Logistic Regression



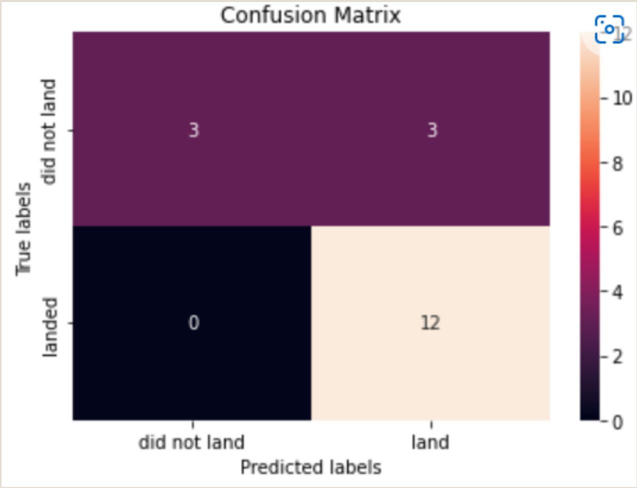
SVM



Decision Trees



KNN



IBM Developer

CONCLUSION

- Success of launch depends on various variables such as payload and launch site.
- ES-L1, SSO and GEO are the orbits with the highest success rate.
- The algorithms for classification are giving almost similar results and that is proven by the confusion matrix.