Predicting if Falcon9 first stage lands successfully

Krishna Teja

20-06-2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- We are predicting if Falcon9 first stage lands successfully because by knowing this we can determine the cost and client can make informed bid against Spacex for a rocket launch.
- We collect the data regarding Spacex's previous launches and their results and split it into test and train data.
- Train different models and choose the one that best fits the train data and have high accuracy with test data.
- Then train the selected model using the entire data.
- Now we can predict weather the Falcon9 first stage lands successfully by using this model.

IBM Developer
SKILLS NETWORK

INTRODUCTION

- We are predicting if the first stage of the SpaceX Falcon 9 rocket 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.
- Point2
- Point3
- Point4
  - Sub Point1
  - Sub Point2

IBM Developer
SKILLS NETWORK

METHODOLOGY

- Data Collection
  - Collecting data from API
- Data Wrangling
  - Preparing data for analysis
- Exploratory Data Analysis
  - Exploratory Analysis using sql
  - Exploratory Analysis using Pandas and Matplotlib
- Interactive Visual Analytics and Dashboard
  - Interactive Visual Analytics with Folium Lab
  - Interactive Visual Analytics with Plotly and Dash.
- Predictive Analysis
  - Classification

Data Collection

- Data Collection via Spacex Rest API:
  - Get response from API and convert it into .json file,
  - Clean the data and make data frame and send it to .csv file.
- Data Collection via Web Scraping
  - Get response from HTML and create Beautiful Soup object.
  - .Parse the HTML using beautiful soup and make a dictionary with column names and values.
  - Convert the dictionary into data frame and send it to .csv file.

Data Wrangling

- Data Wrangling is the process of cleaning the data we obtained from the data collection process.
- Load the data we obtained previously(.csv file from data collection process) into a data frame.
- Choose all the columns that are unnecessary and drop them from the data frame.
- Convert all the categorical columns into numerical columns using dummies(one – hot encoding) and send the data frame to .csv filie.

Exploratory Data Analysis(EDA)

- EDA using with Data Visualization
  - We draw several scatter plots to observe the relation between various columns.
    - Payload vs Flight Number
    - Flight Number vs Launch Site
    - Payload vs Launch Site
    - Flight Number vs Orbit Type
    - Payload vs Orbit Type
- EDA using sql
  - We use sql to for finding the following
    - Total Payload Mass
    - Average Payload Mass carried by booster version F9v1.1
    - The date where successful landing outcome in drone ship was achieved etc..

Interactive Visual Analytics

- Interactive Map with Folium
  - We use Folium to make an interactive leaflet map.
  - We use the coordinates of all the launch sites and add a circle marker with launch site name as label.
  - Then we differentiate success and failure with green and red colors and a number on the circle representing number of failures or success.

- Interactive Map with Plotly Dash
  - We build a dash app with call back function that out puts:
    - Pie-chart(Percentage of success in relation to launch site)
    - Scatter P lot(Success Rate vs Booster Version Category)

Predictive Analysis

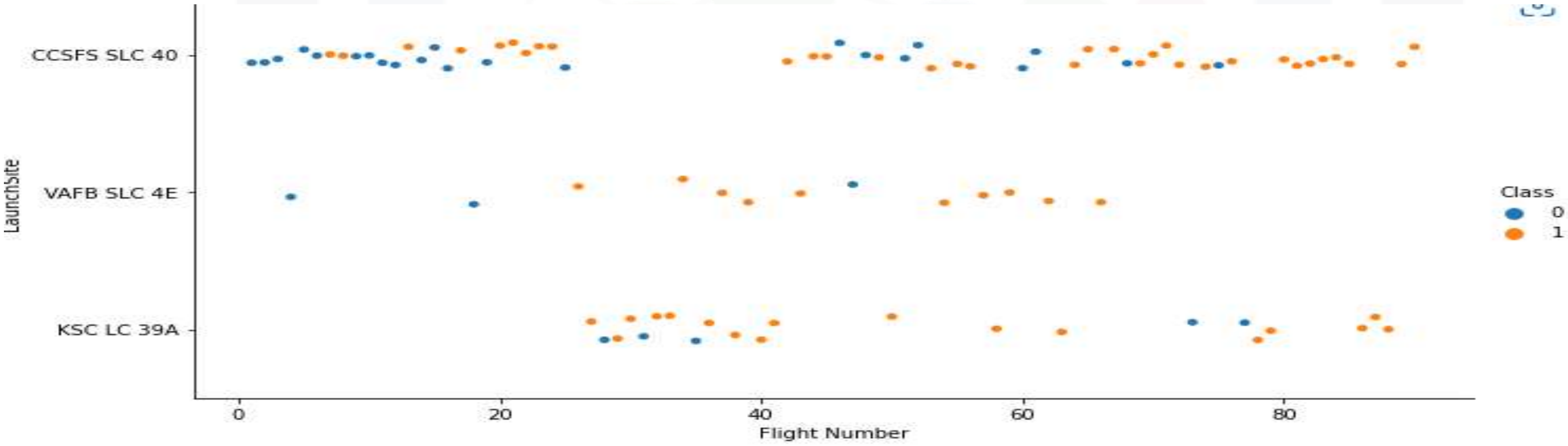
- Building Model
  - Load and transform the data.
  - Standardize the data and split it into test and train data.
  - Train the models using GridSearchCV and train data.
- Evaluating Model
  - Calculate the accuracy of the models and plot confusion matrix.
- Comparing the Model's Accuracy
  - Compare the accuracy and confusion matrices of different matrices and choose the best model.

RESULTS

- Exploratory Data Analysis
  - EDA with visualization
  - EDA with sql
- Interactive Visual Data Analysis
  - Interactive Map with Folium
  - Plotly Dashboard
- Predictive Analysis

EDA with visualization

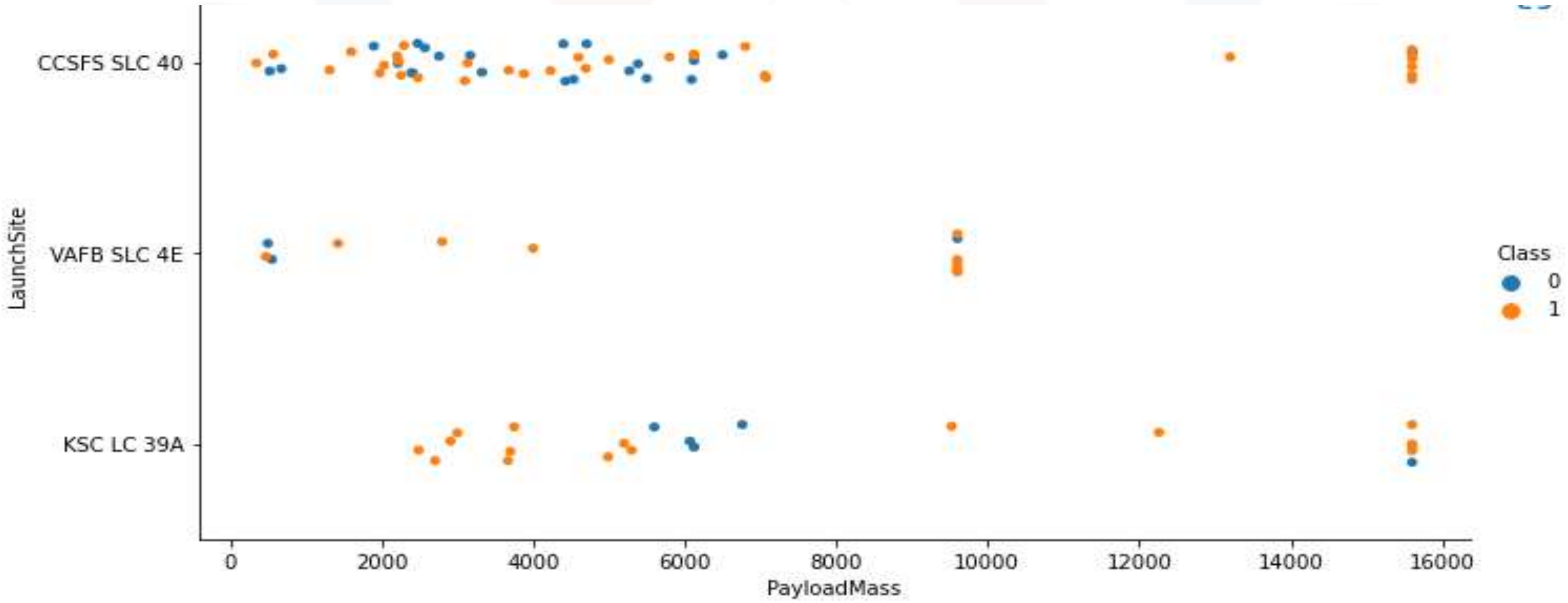
Flight Numbers vs Launch Site



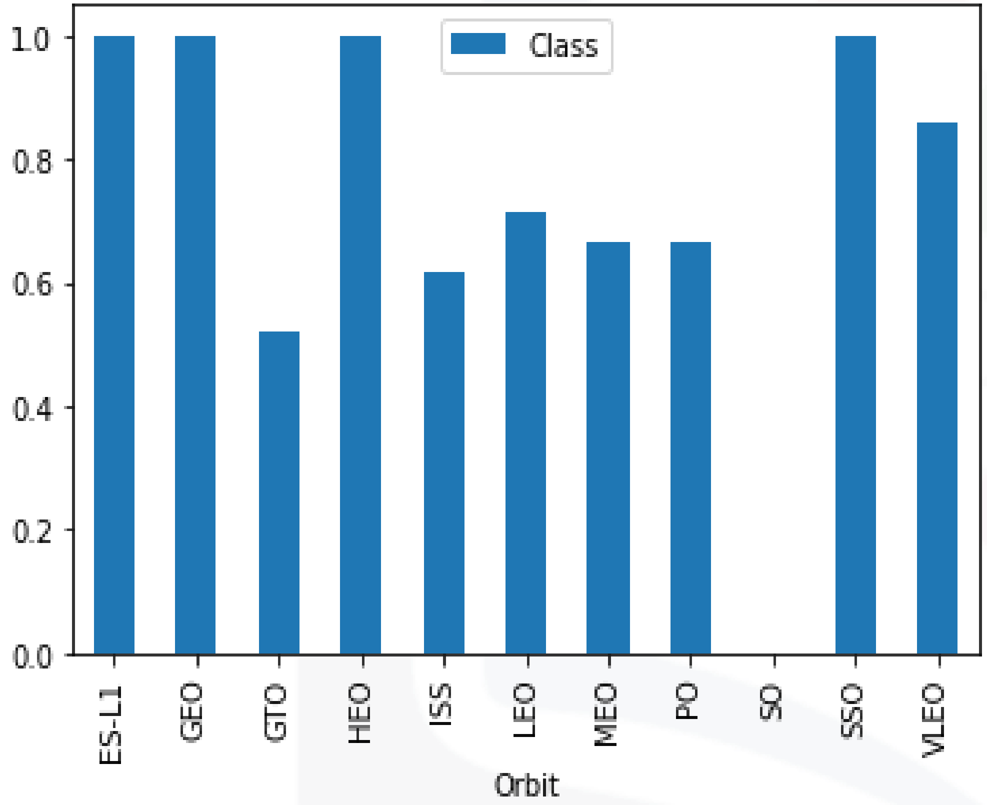
IBM Developer

SKILLS NETWORK

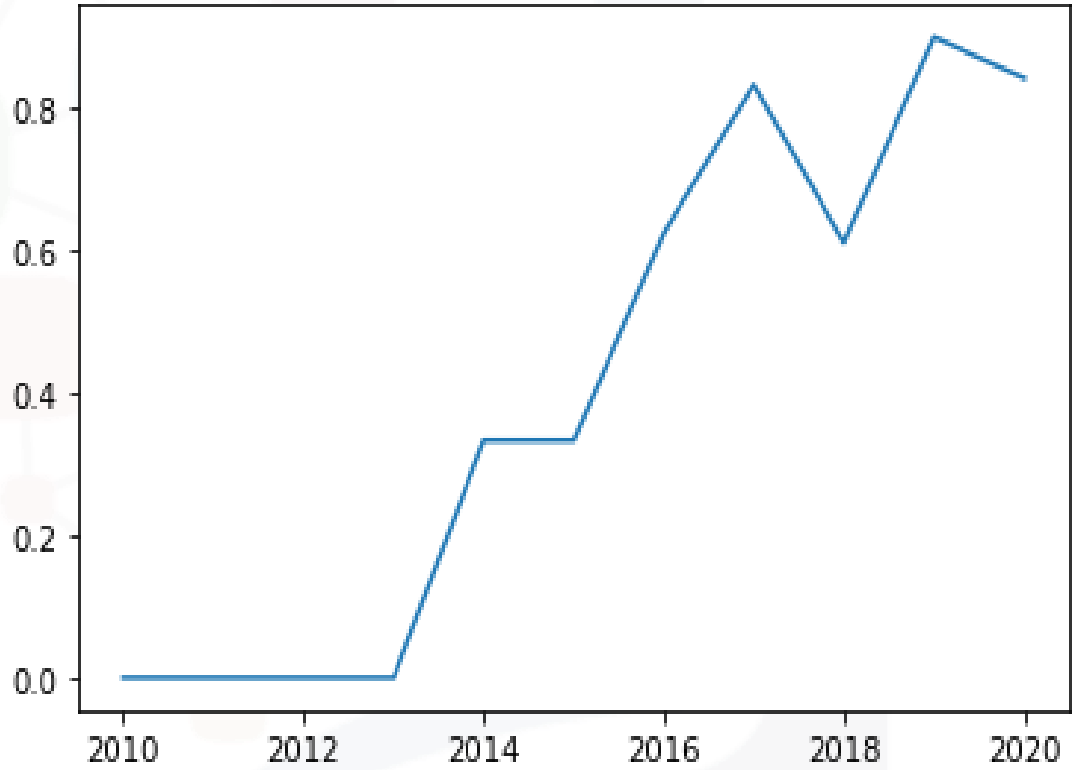
Payload vs Launch Site



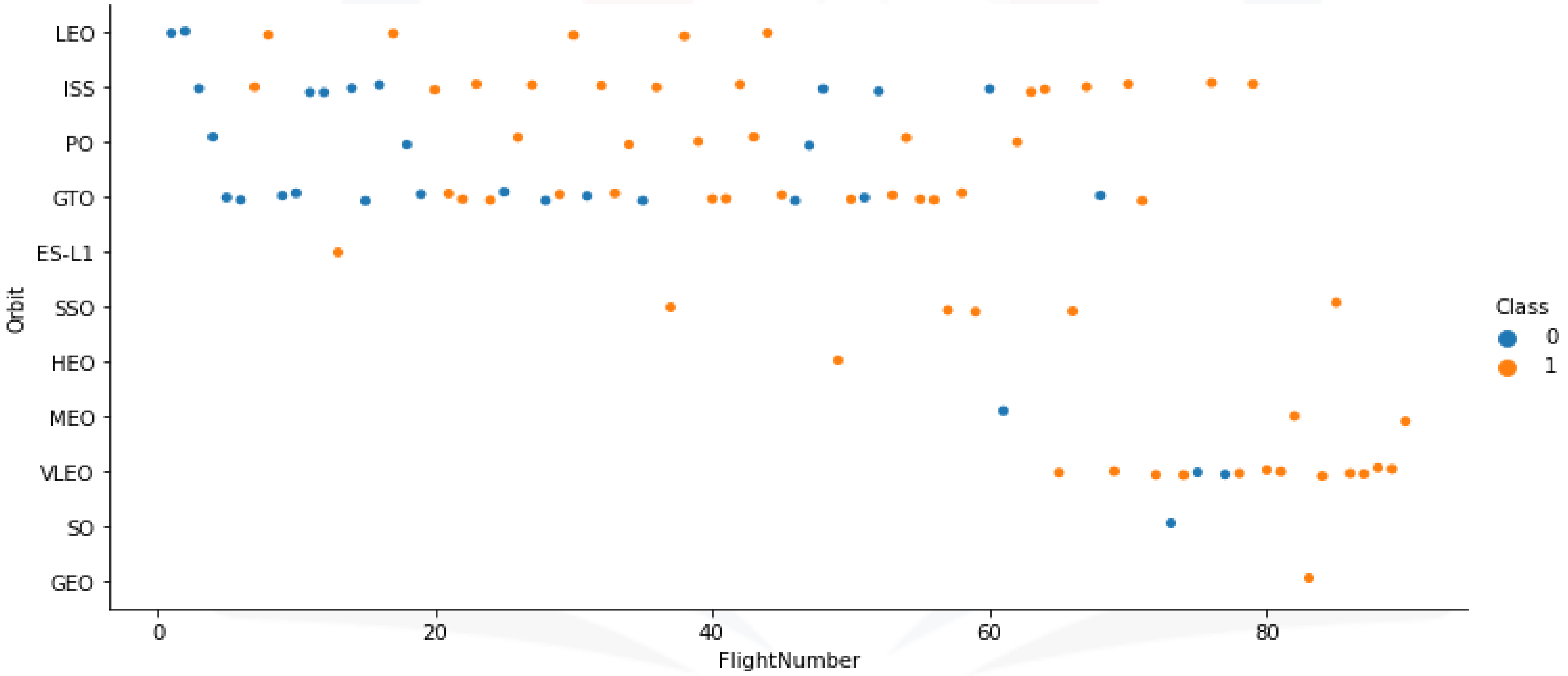
Success Rate vs Orbit Type



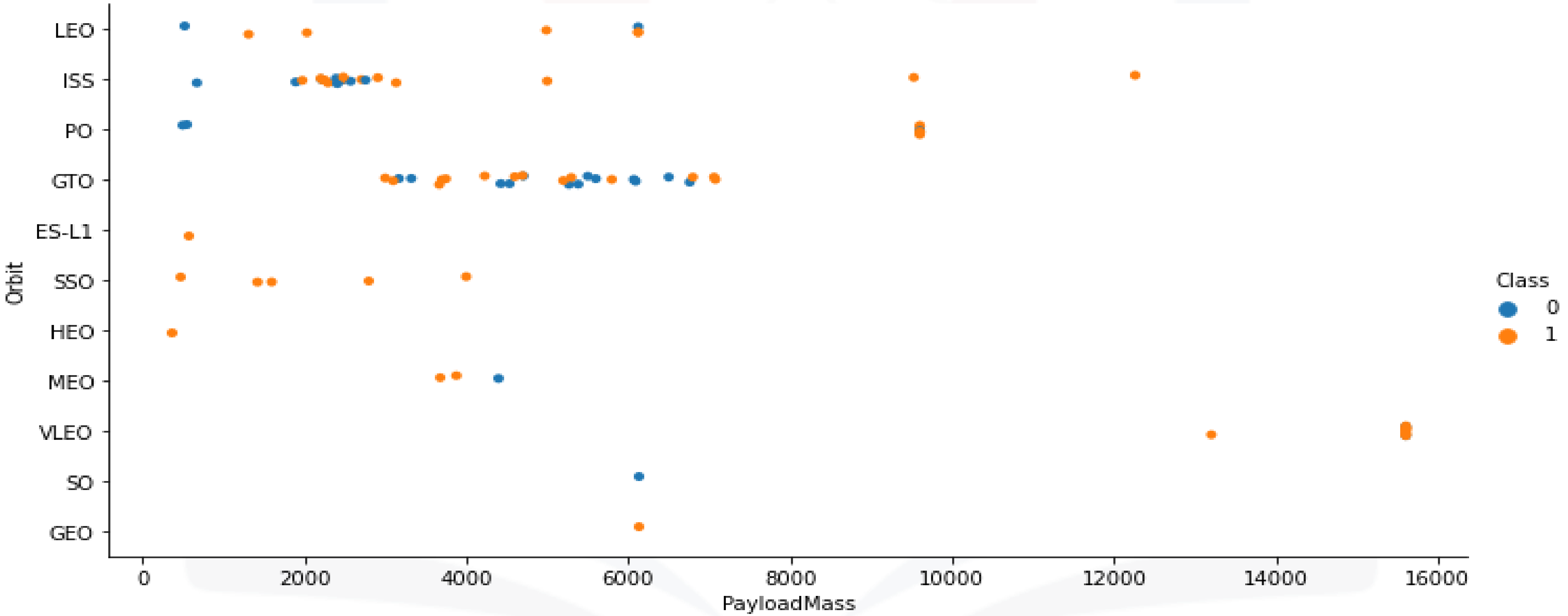
Launch Success Yearly Trend



Flight Number vs Orbit Type



Payload Mass vs Orbit Type



EDA with sql

- The names of the unique launch sites in the space mission
  - **'CCAFS LC-40','CCAFS SLC-40','KSC LC-39A','VAFB SLC-4E'.**
- 5 records where launch sites begin with the string 'CCA'

	Date	Time	Booster_Version	Launch_Site	Payload	Payload_mass(kg)	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Total payload mass carried by boosters launched by NASA (CRS)
  - **999,880 KG.**
- Average payload mass carried by booster version F9 v1.1
  - **2,534.67 KG.**
- The date when the first successful landing outcome in ground pad was achieved.
  - **22-12-2015**
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - **`('F9 FT B1022'), ('F9 FT B1026'), ('F9 FT  B1021.2'), ('F9 FT B1031.2')`**

- The total number of successful and failure mission outcomes
  - `('Failure', 1)`
  - `('Success', 100)`

- Names of the booster_versions which have carried the maximum payload mass.
  - `('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')`

- The records that contain month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

	Month	Booster_Version	Landing_Outcome	Launch_Site
0	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- The count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
  - **61**

Interactive Visual Analysis with Folium

- **Launch Sites on Map**

The success/failed launches for each site on the map

Lest Map shows the number of rocket launches at each location on the map.

Right map show the landing outcome with green color marker as success and red color marker as failure.

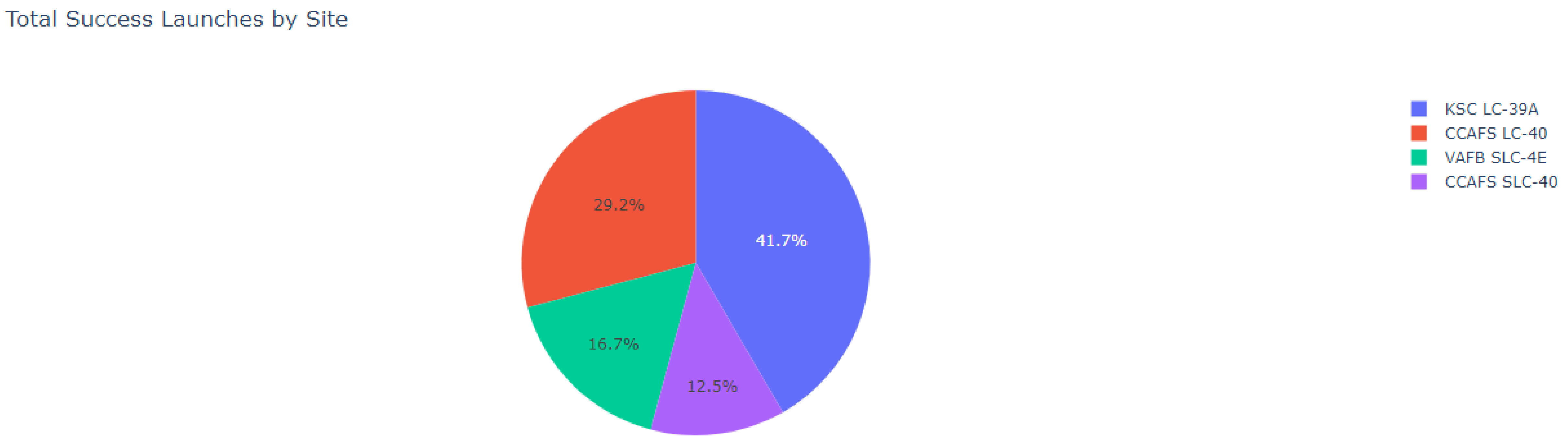
Interactive Visual Analytics with dash and plotly

- Total successful launches by all sites
- Pay Load vs Launch Outcome scatter plots for all sites.
- Launch Site with highest success ratio.
- Launch Site with lowest success ratio.

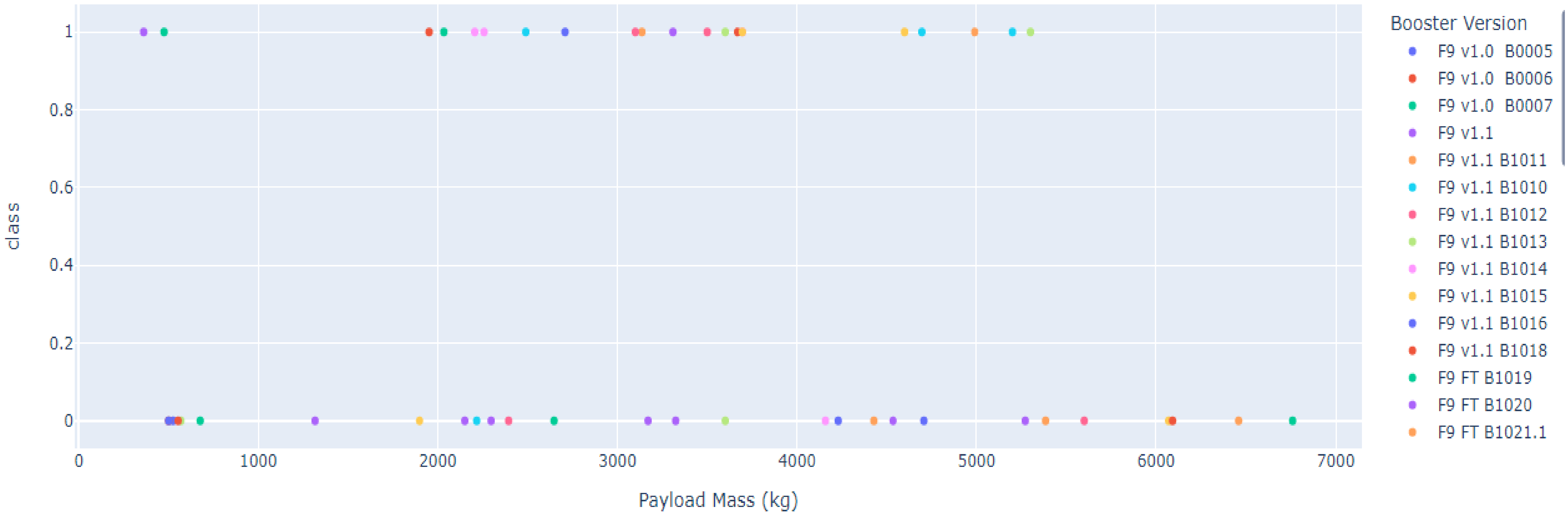
DASHBOARD

Dash is the link for dash board.

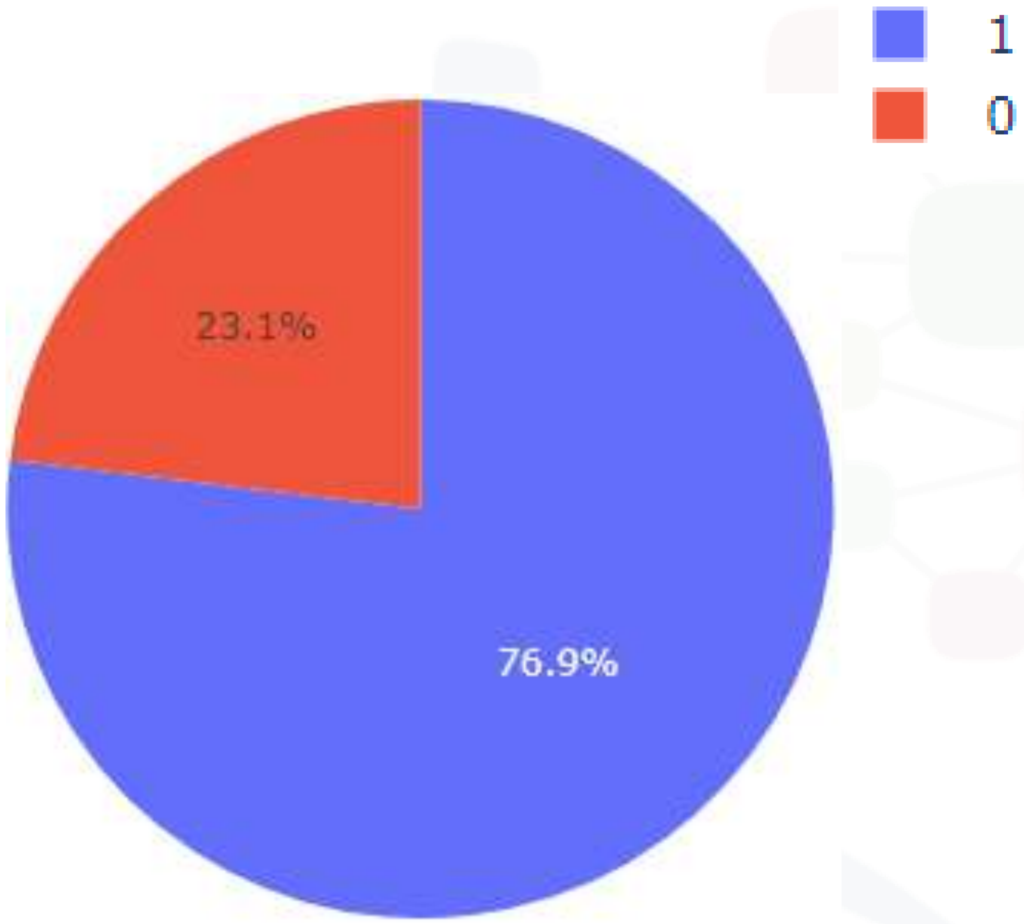
Total successful launches by all sites



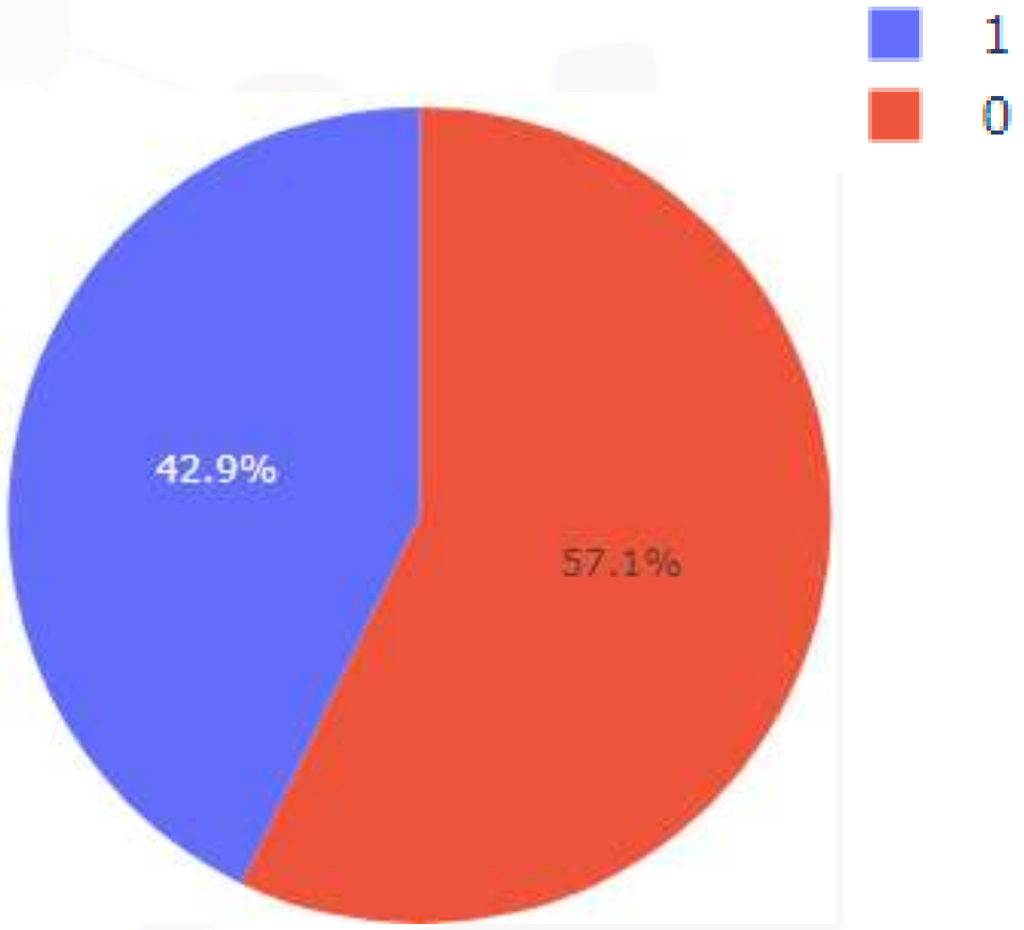
Pay Load vs Launch Outcome



Launch Site with highest success ratio.



Launch Site with lowest success ratio.



DISCUSSION

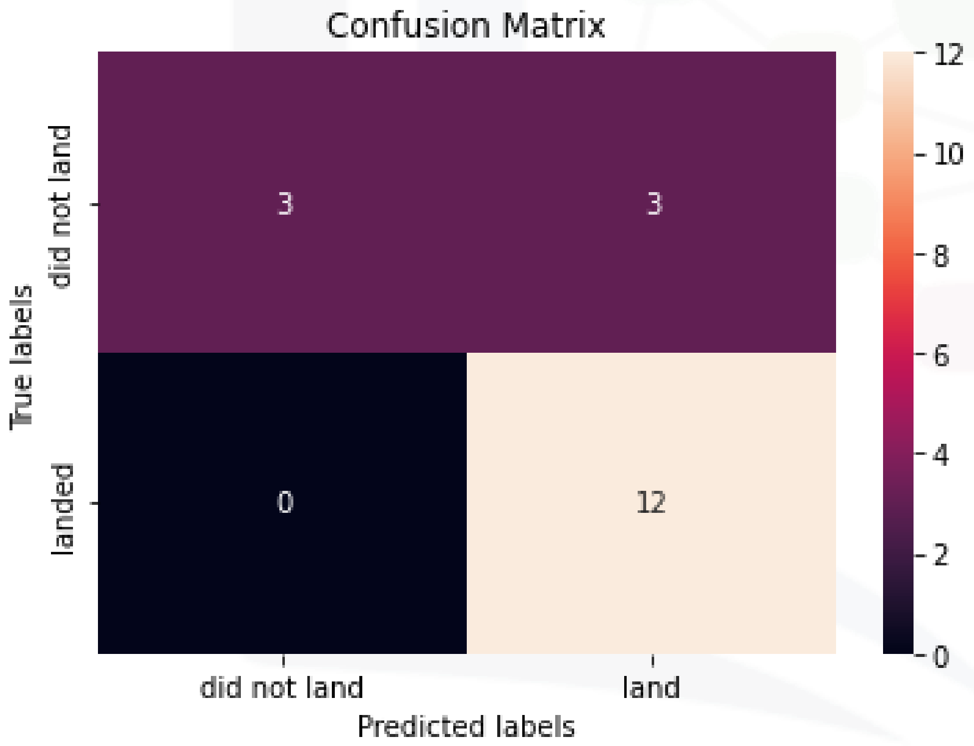
- KSC LC-39A has the highest launch site success rate.
- CCAFS SLC-40 has lowest launch site success rate.
- Payload between 2k-10k has the highest success rate.
- Payload less than 1k has the least success rate
- F9 booster version FT has the highest success rate

Predictive Analysis

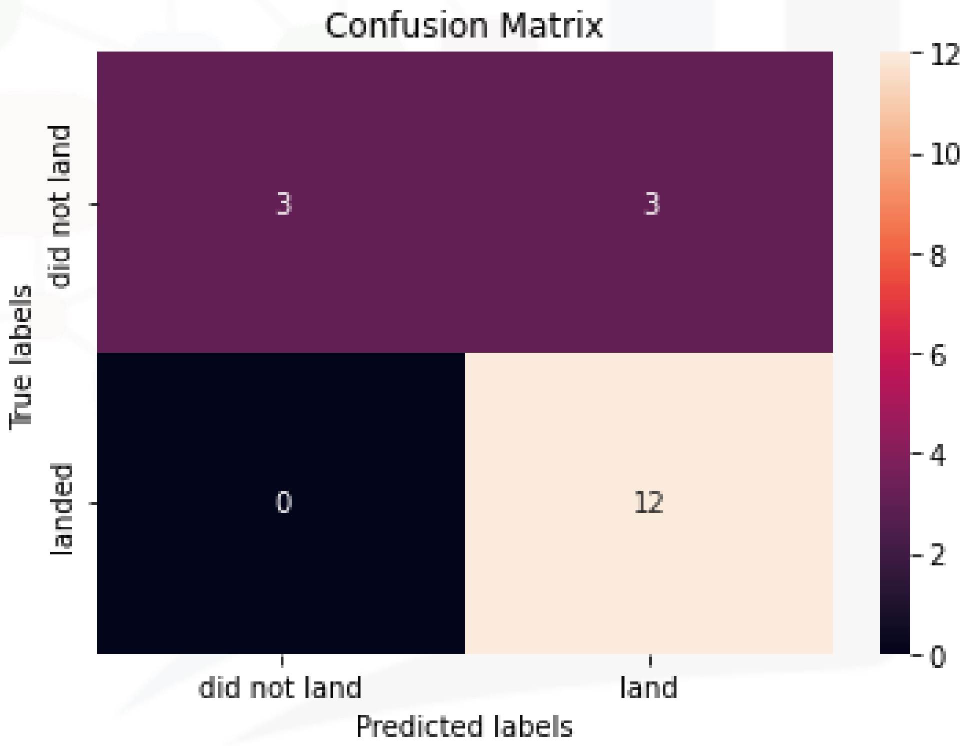
- Logistic Regression
- SVM
- Decision Tree
- KNN

Confusion Matrices

Logistic Regression

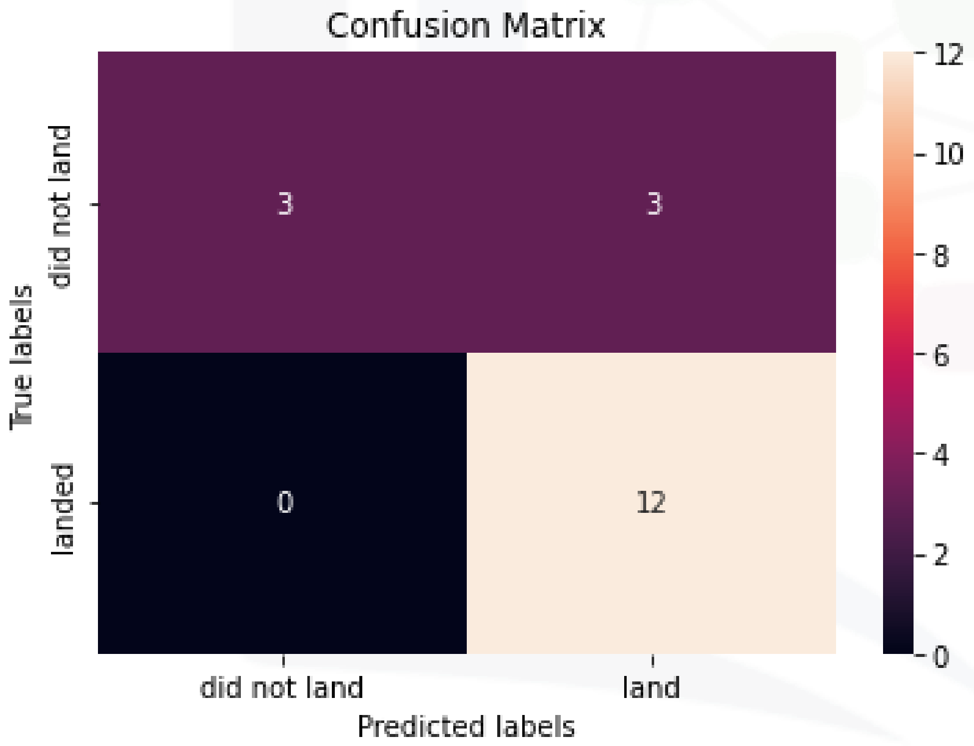


SVM

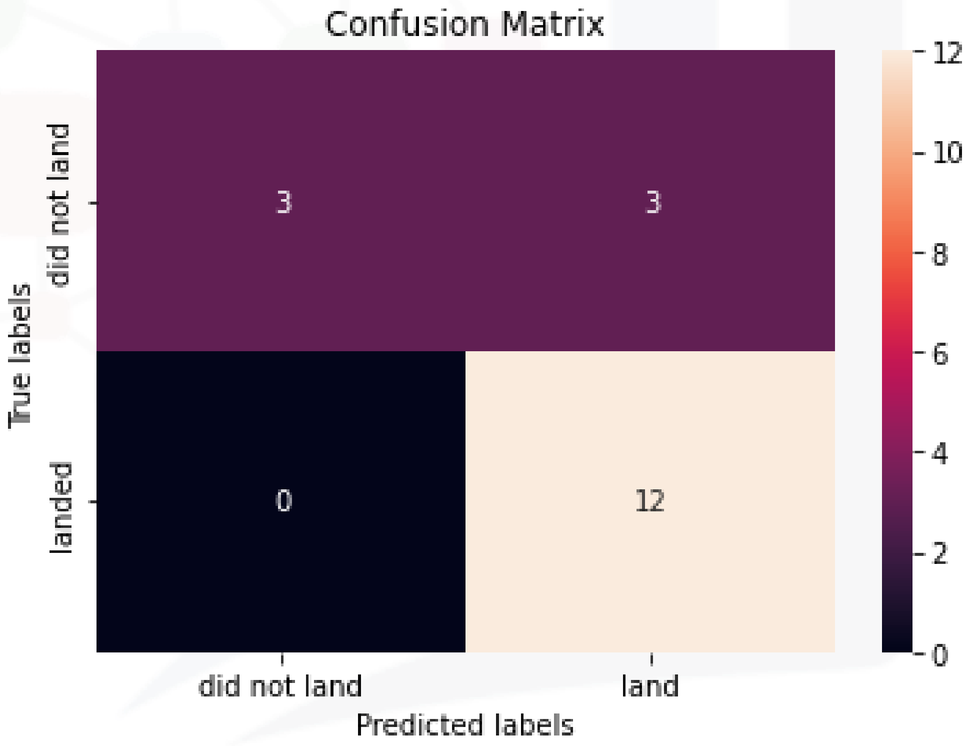


Confusion Matrices

Decision Tree



KNN



Accuracy

Algorithm	Accuracy	Accuracy on Test Data
Logistic Regression	0.846	0.834
SVM	0.835	0.834
Decision Tree	0.875	0.834
KNN	0.860	0.834

From the table we can see that decision tree performs the best with an accuracy of 0.875 which is very high.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Orbits ES-L1,GEO,SSO has highest success rates.
- Success rates are gradually increasing with time.
- Decision Tree algorithm best predicts the outcome the rocket landing.

Implications

- This implies the rockets that go into these orbits are more likely to successfully land.
- Spacex rockets are more likely to be successful in near future.
- Decision Tree algorithm can be used to predict the outcome of the rocket launching.

CONCLUSION

- The success rate of the spacex rockets is increasing with time.
- Success rate of certain orbits is higher than other orbits.
- Rockets with payload 2k-10k are more likely to have a successful outcome.
- Decision Tree algorithm can be used to predict the outcome of the spacex rocket launch.

IBM Developer SKILLS NETWORK