Data Analysis of SpaceX

Ritu Vyaghrambare

04/06/2022

OUTLINE

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

EXECUTIVE SUMMARY

- Data collection from the SpaceX API
- Exploratory data analysis to determine the attributes responsible for landing of Falcon9 second stage
    - Landing prediction of first stage to estimate the cost of launch
    - Estimation of launch success rate with respect to payload mass, orbit type, and launch site location.
- Interaction dashboard to compare different aspects
- Predictive analysis using a machine learning model and output is presented in the form of confusion matrix

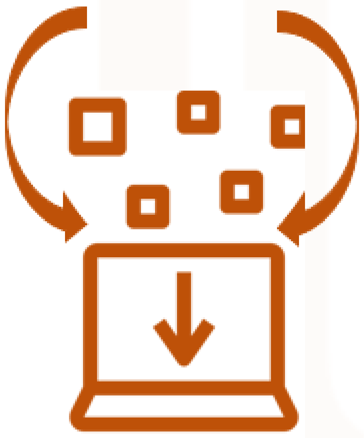
INTRODUCTION

- In the commercial space age, launching a rocket with a minimum cost is crucial. SpaceX does it with an ease. So, by analyzing the data from SpaceX API, the details can be studied.
- The exploratory data analysis is done to find the relation between different features and their affect on cost.
- Different graphs shows the relation of payload mass, orbit type, and the location of launching can help to build a machine learning model.
- An interactive dashboard is given to study the analysis.
- The findings are presented in terms of confusion matrix.

METHODOLOGY

- The data for analysis is taken from the SpaceX API using the web scrapping. Data wrangling is done to clean the data.
- Exploratory data analysis is done with the help of SQL and Python to study the relationship between the different features.
- The interactive dashboard is created to compare the features and predict if the rocket will land successfully.
- The train set and test set is taken from the data to run the different methods. The output of it is used to create the machine learning model for easy prediction.

DATA COLLECTION

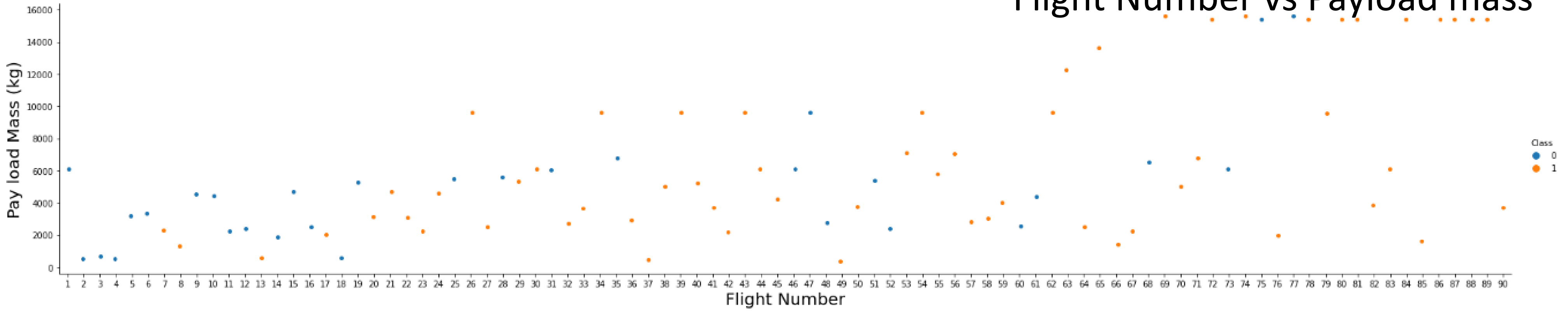


- The rocket launch data was requested from the SpaceX API. The Json content was decoded and converted to pandas data frame.
- Null values in the 'PayloadMass' column were identified and by data wrangling. The missing values were replaced by the mean of the data.

EXPLORATORY DATA ANALYSIS

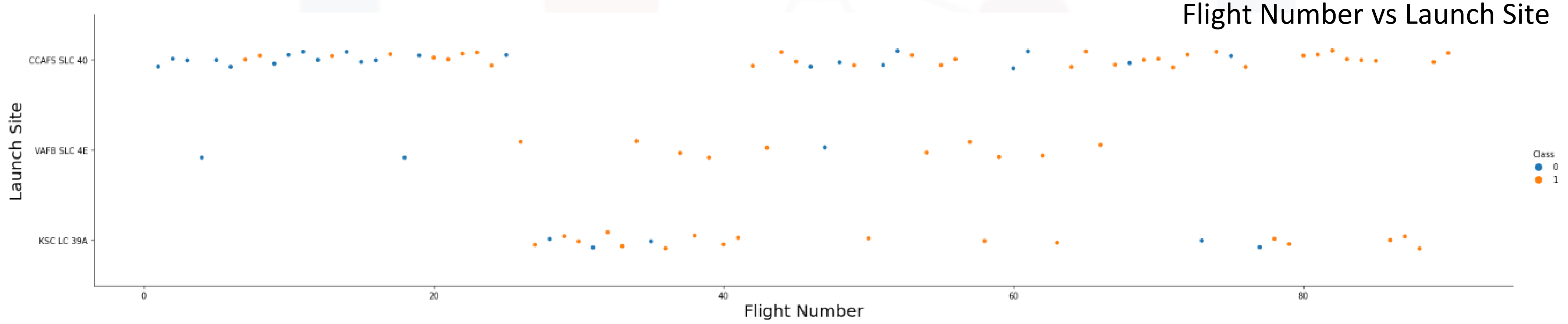
- The connection to the data base was made to query the data for important information.
- The different features were compared as shown:

Flight Number vs Payload mass

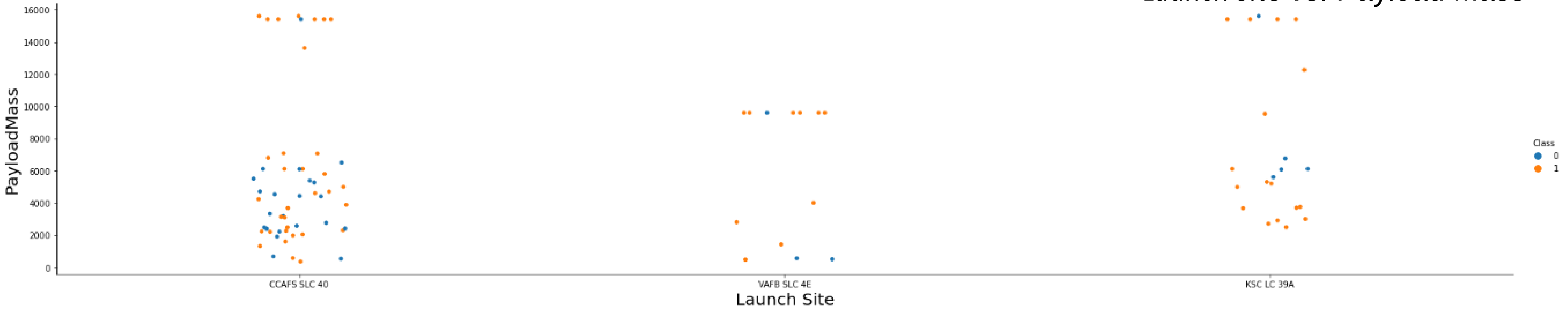


EXPLORATORY DATA ANALYSIS

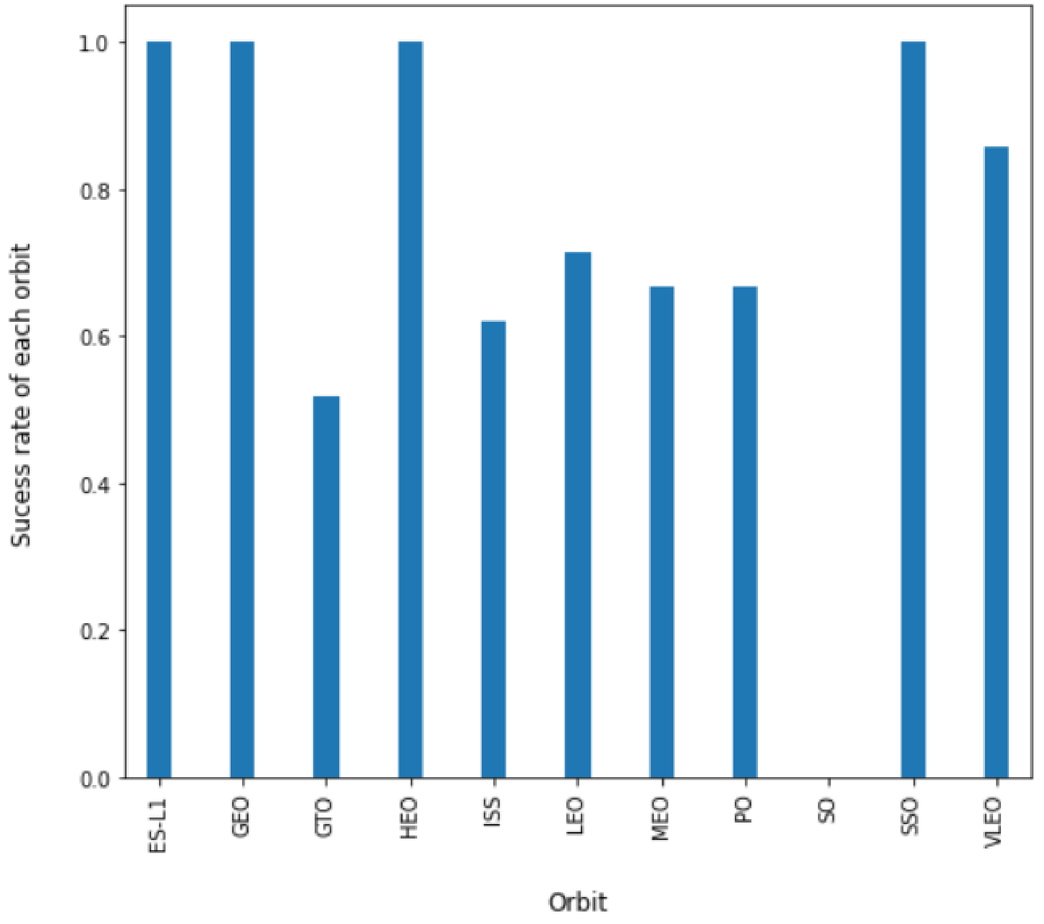
Flight Number vs Launch Site



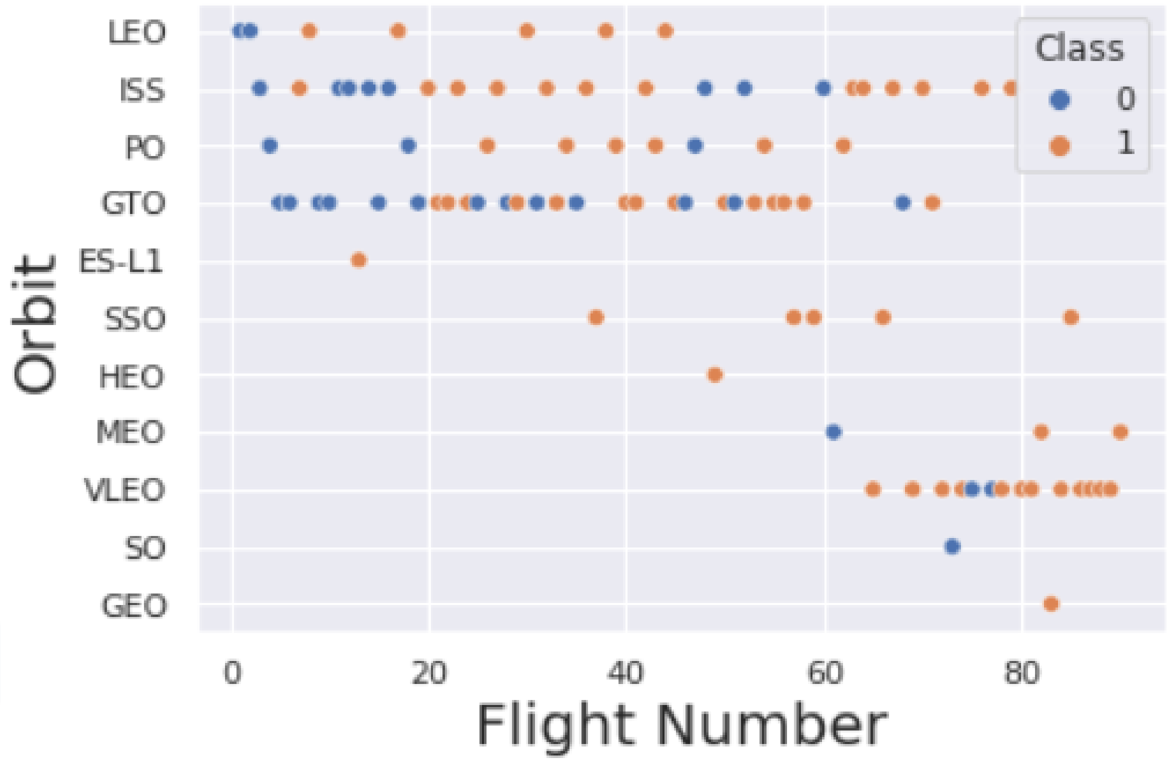
Launch Site vs. Payload Mass



SUCCESS RATE OF EACH ORBIT

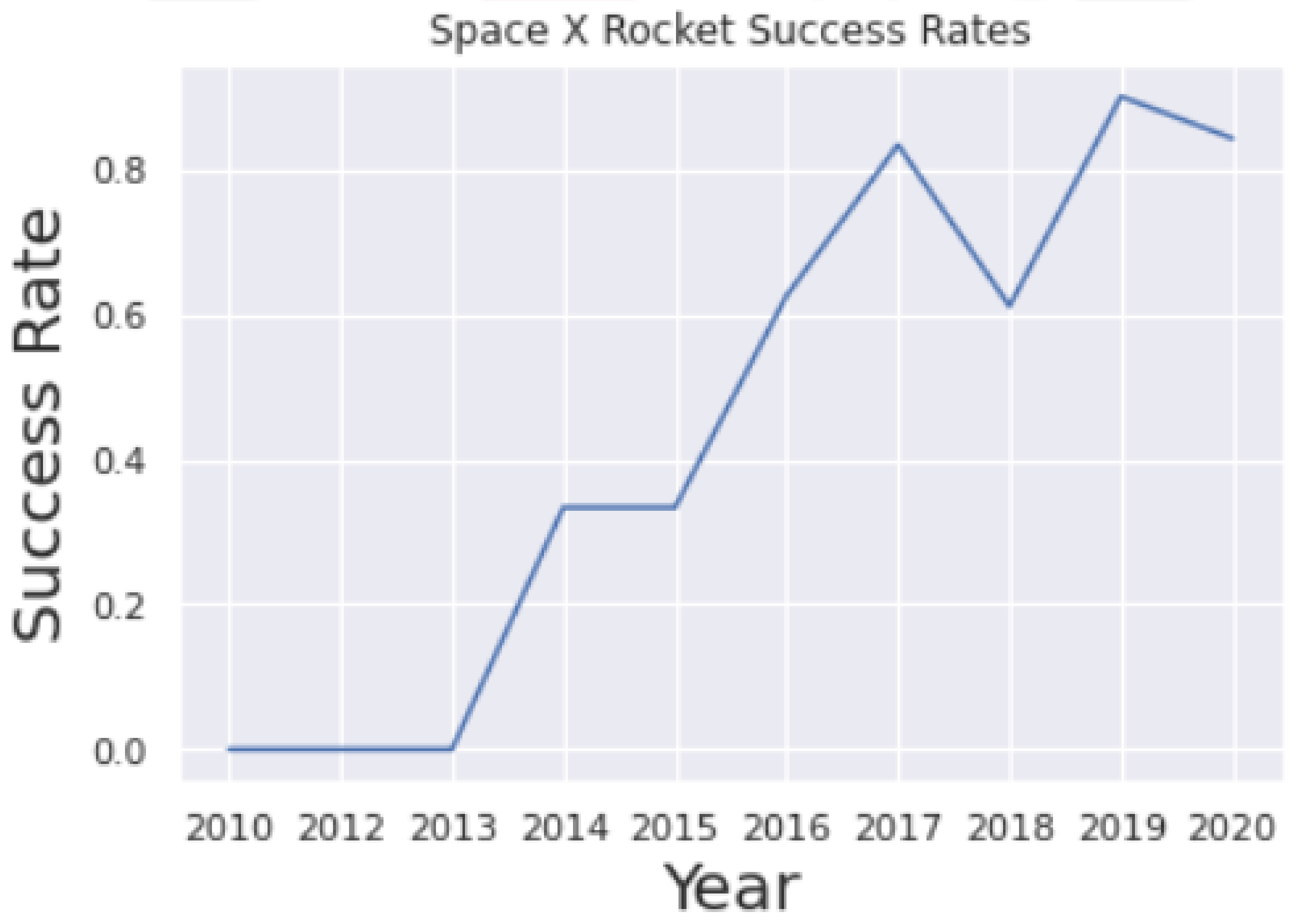


RELATIONSHIP BETWEEN FLIGHTNUMBER AND ORBIT TYPE



RELATIONSHIP BETWEEN PAYLOAD AND ORBIT TYPE

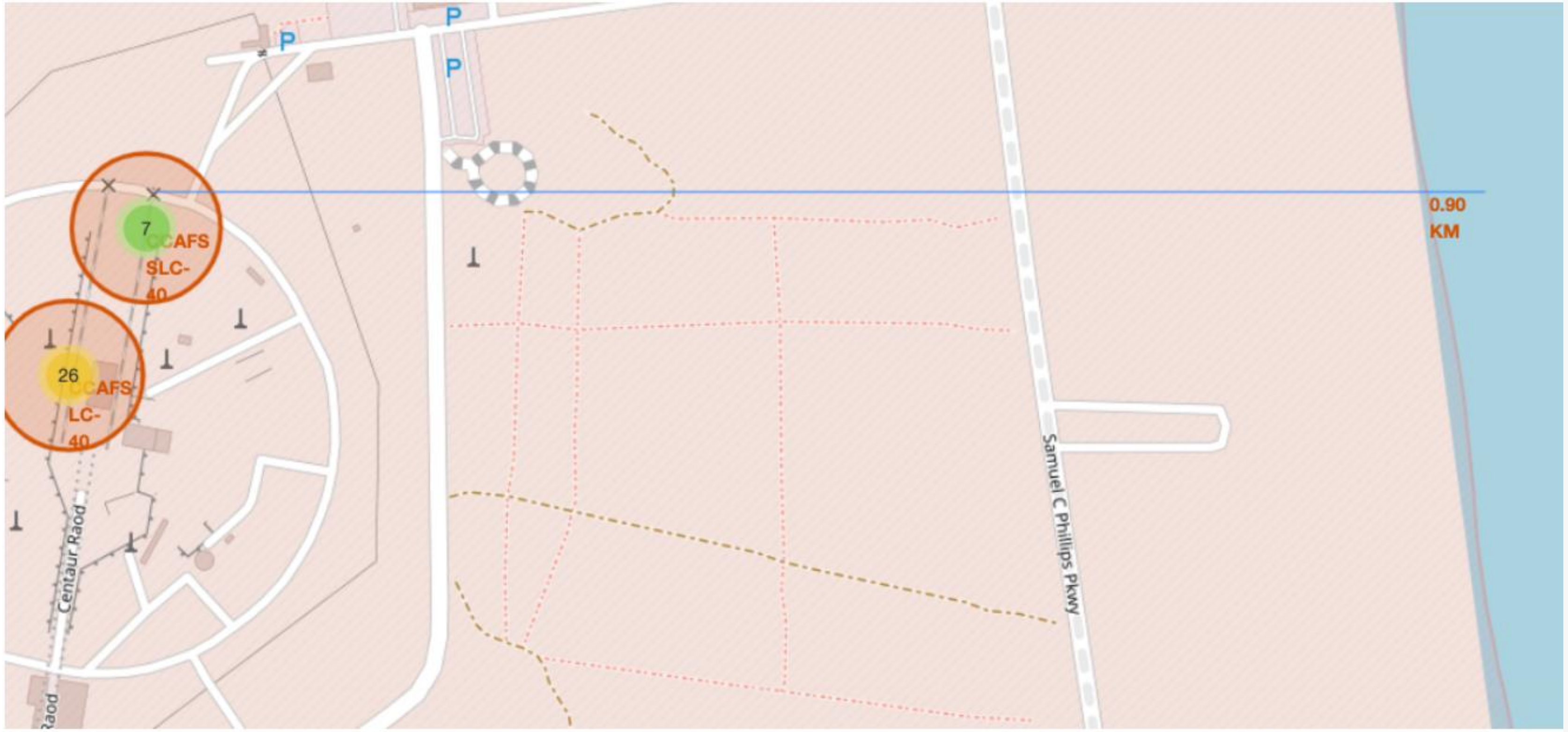
SUCCESS RATE



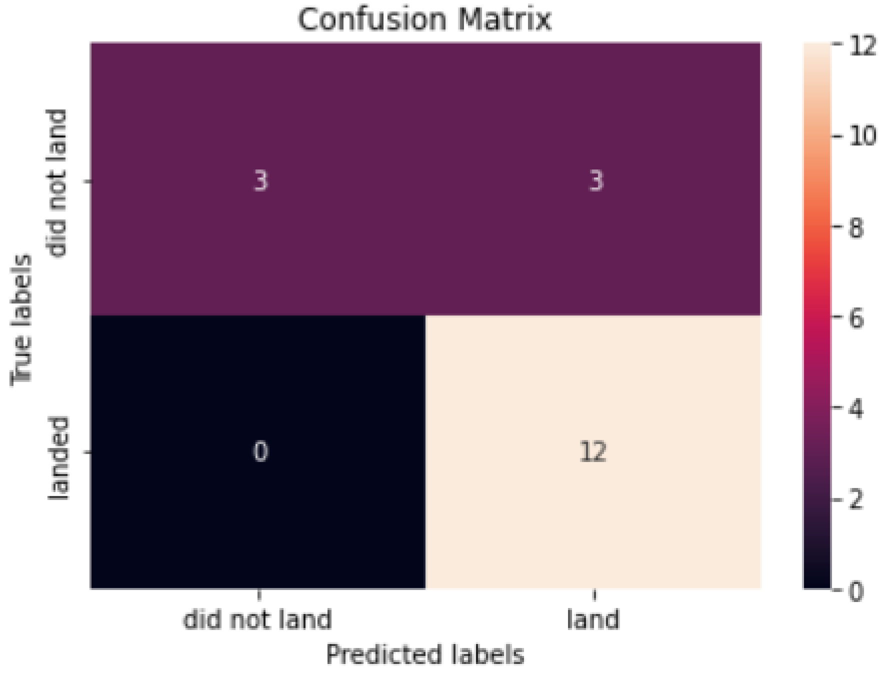
IBM Developer

SKILLS NETWORK

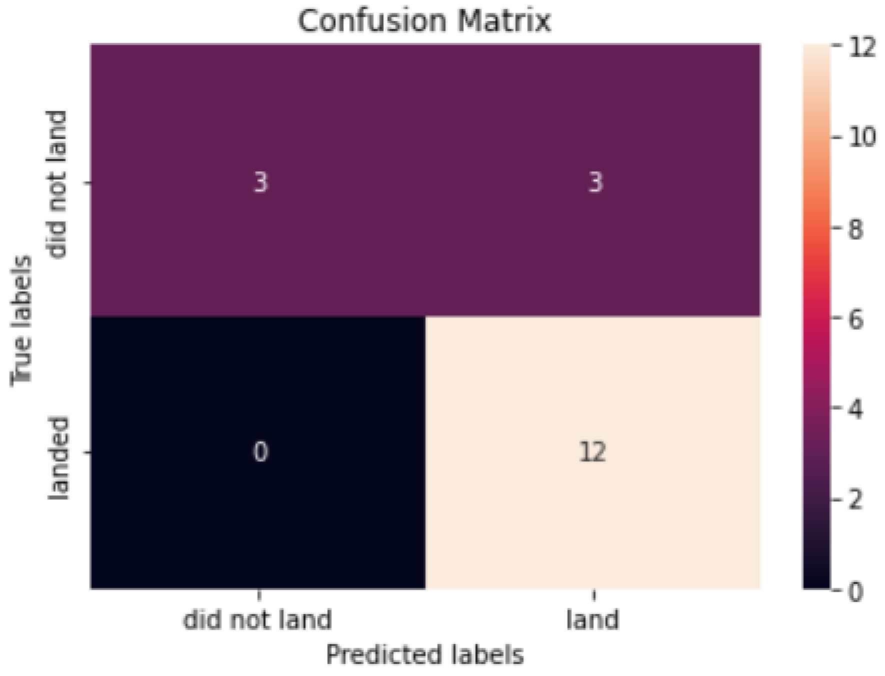
LAUNCHING SITES



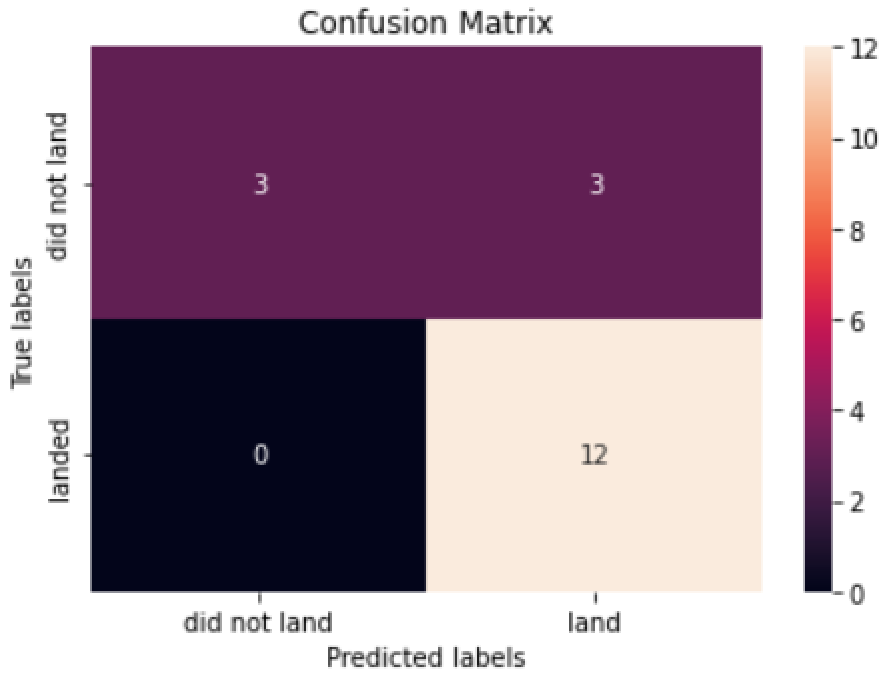
PREDICTIVE ANALYSIS (Logistic regression)



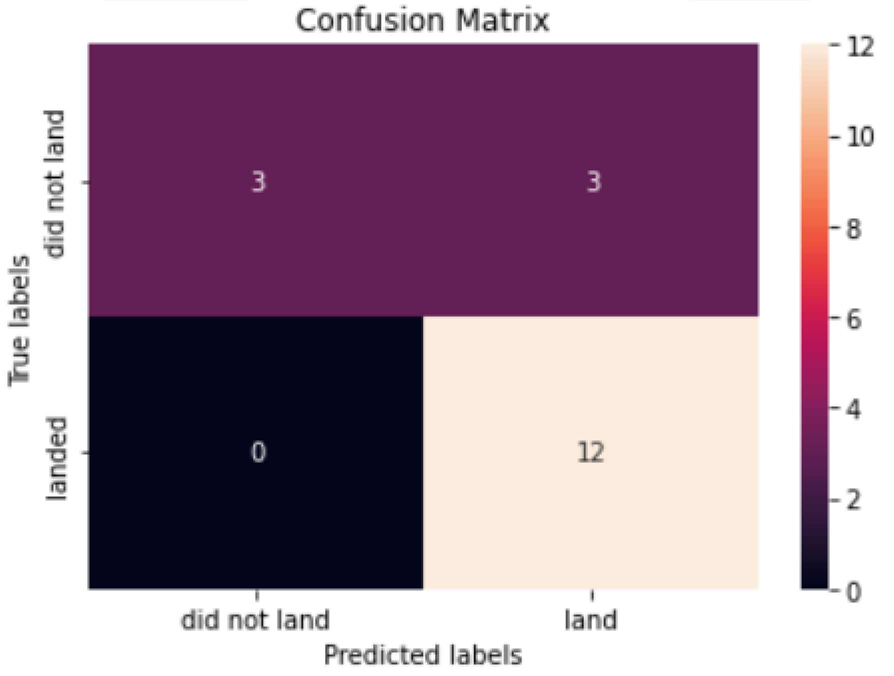
PREDICTIVE ANALYSIS (Support vector machine object)



PREDICTIVE ANALYSIS
(Decision tree classifier object)



PREDICTIVE ANALYSIS
(k nearest neighbors object)



RESULTS

	Model	Accuracy	Prediction score
0	LogisticRegression()	0.8464285714285713	0.8333333333333334
1	SVC()	0.8482142857142856	0.8333333333333334
2	DecisionTreeClassifier()	0.8910714285714286	0.7222222222222222
3	KNeighborsClassifier()	0.8482142857142858	0.8333333333333334

CONCLUSION

- The retrieval of the first stage successfully can make the mission cost efficient.
- For a successful landing, the appropriate launch site, payload mass, and orbit type plays a significant role.
- Success rate is directly proportional to payload mass.
- CCAFS SLC 40 launching site has higher success rate for the payload masses below 6000 kg and above 14000 kg.
- Orbit type VLEO and ISS can be accessed for the launches.
- ISS orbit type is more suitable for the payload masses between 2000 kg and 4000 kg.
- Decision tree classifier is the best fitted machine learning model for the most accurate predictions.