Applied Data Science Capstone

Andrea da Costa

7 June 2022

OUTLINE

- **Executive Summary**
- **Introduction**
- **Methodology**
- **Results**
- **Conclusion**

EXECUTIVE SUMMARY

Methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Results

- EDA results
- Interactive analytics
- Predictive analysis

IBM Developer

SKILLS NETWORK

INTRODUCTION

Context:

A commercial space exploration company, SpaceY, advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, whereas competitors cite their costs as upward of 165 million dollars each. Much of the savings is due to the face that SpaceY can reuse the first stage of the rocket launch.

This project explores and analyses the different aspects relevant to rocket launches with the aim of finding the best method to predict the likelihood that the first stage of the SpaceY Falcon 9 rocket landing successfully.

METHODOLOGY

- Data collection via
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Data Wrangling
  - One Hot Encoding data fields for Machine Learning and cleaning of null values and irrelevant columns
- Exploratory data analysis (EDA) with SQL and data visualization
- Interactive visual analytics with Folium and Plotly Dash
- Predictive analysis using classification models
  - LR, KNN, SVM, DT models were built and evaluated for the best classifier

SUMMARY OF RESULTS

- The success rates for SpaceX launches is directly proportional time in years spent perfecting the launches.

- Low weighted payloads perform better than the heavier payloads.

- KSC LC 39A had the most successful launches from all the sites.

- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

- In terms of prediction accuracy of successful launches, the SVM, KNN, and Logistic Regression models work best for this dataset.
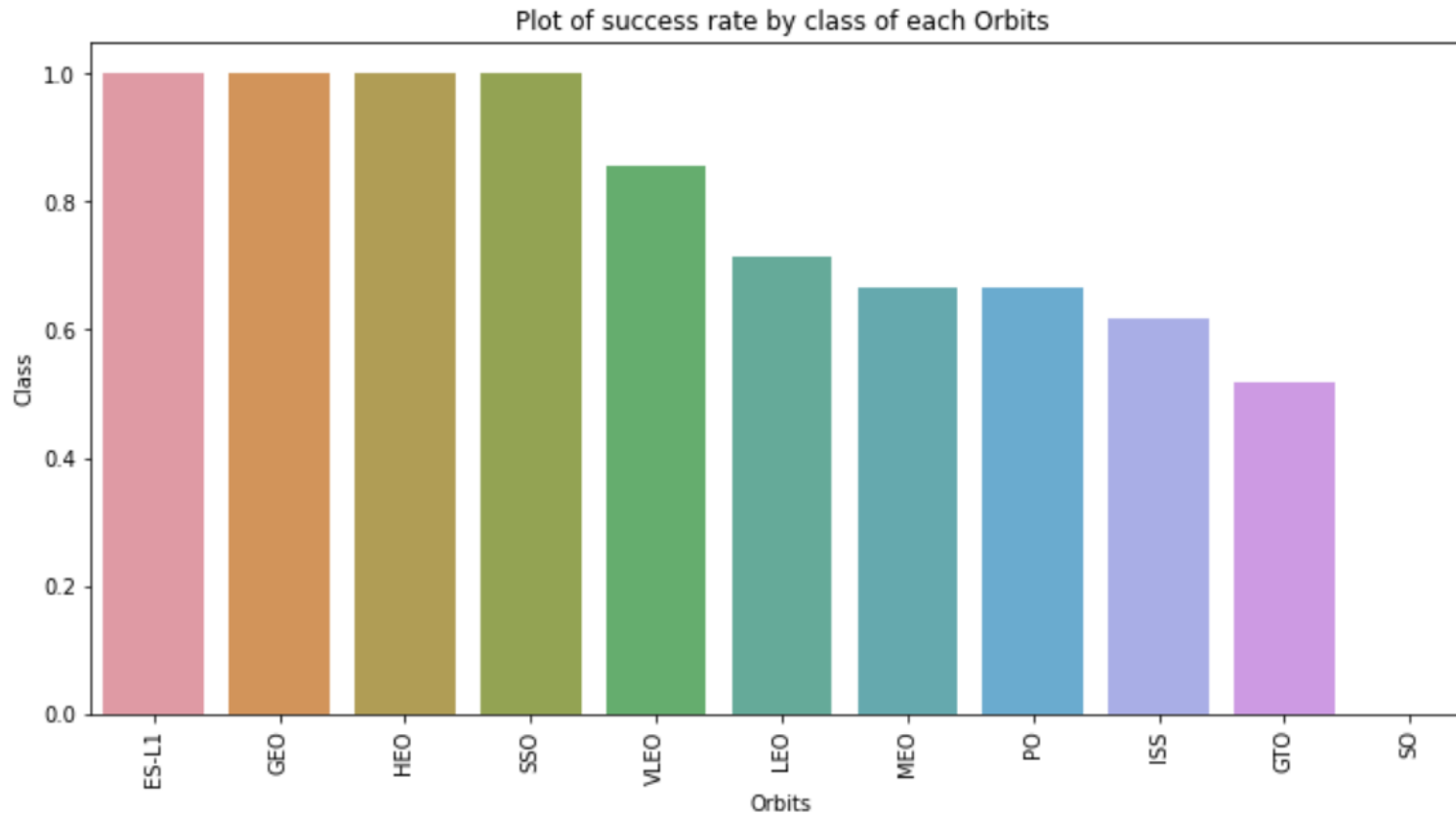
Flight Number vs. Launch Site
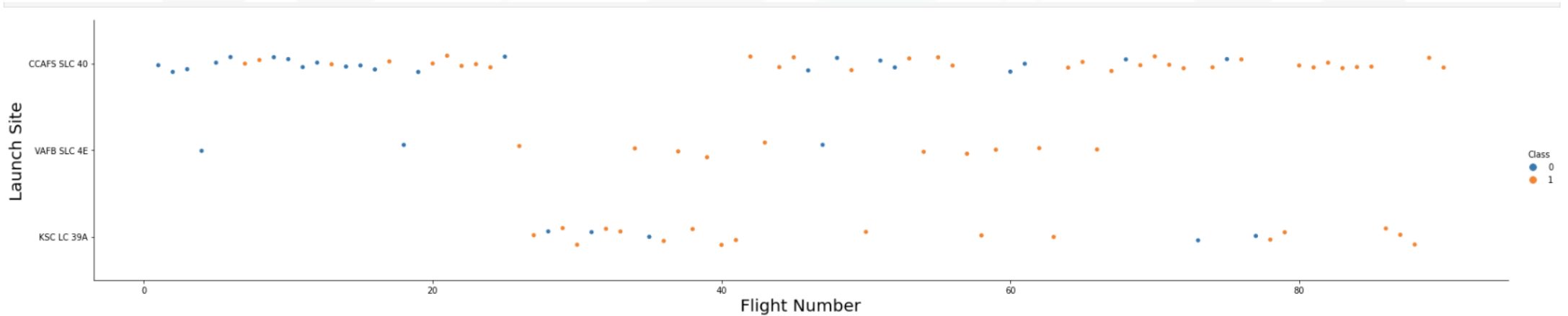
Launches from the site of CCAFS SLC 40 are significantly higher than launches form other sites.
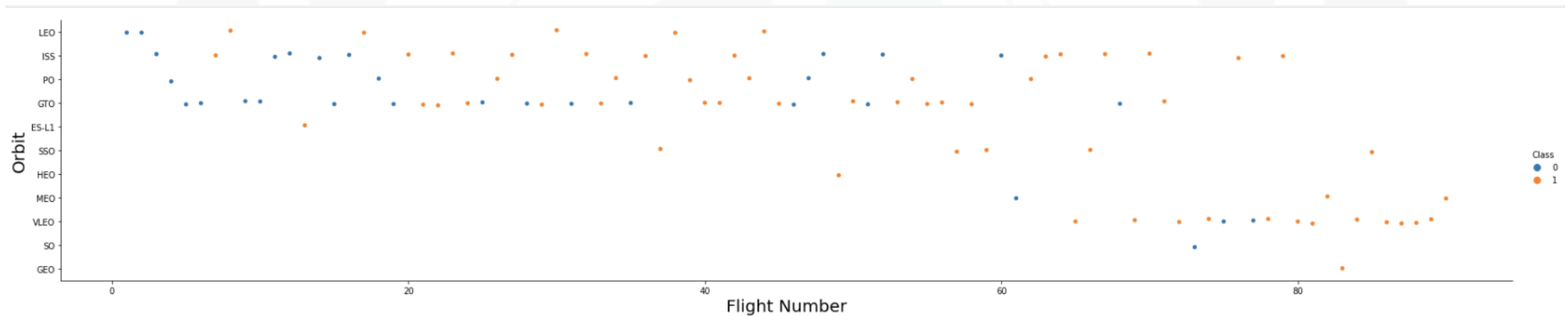
Payload vs. Launch Site
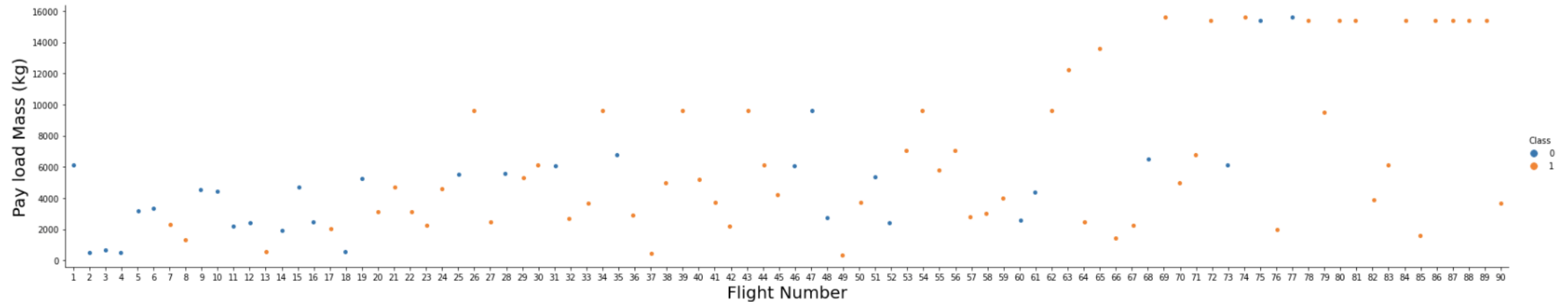
The majority of payloads with lower Mass have been launched from CCAFS SLC 40.
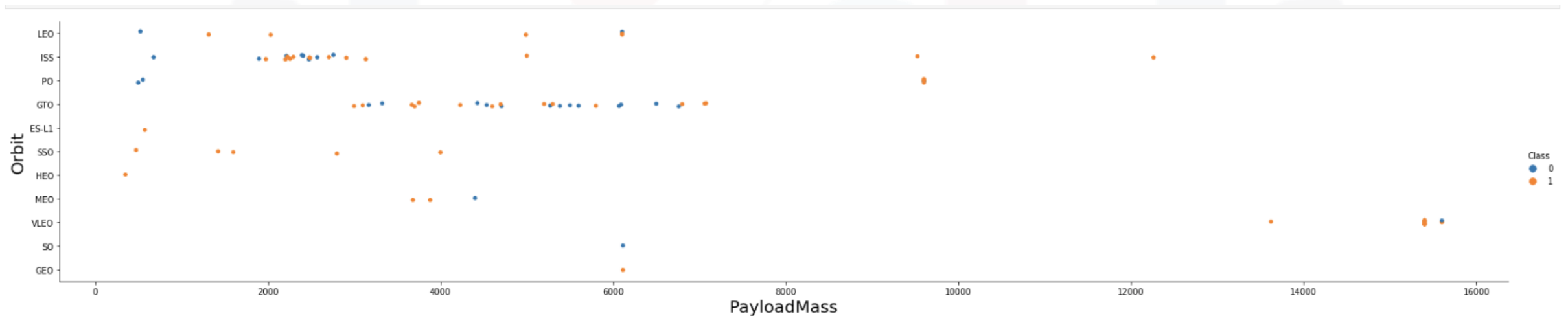
Success Rate vs. Orbit Type

Plot of success rate by class of each Orbits

The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.
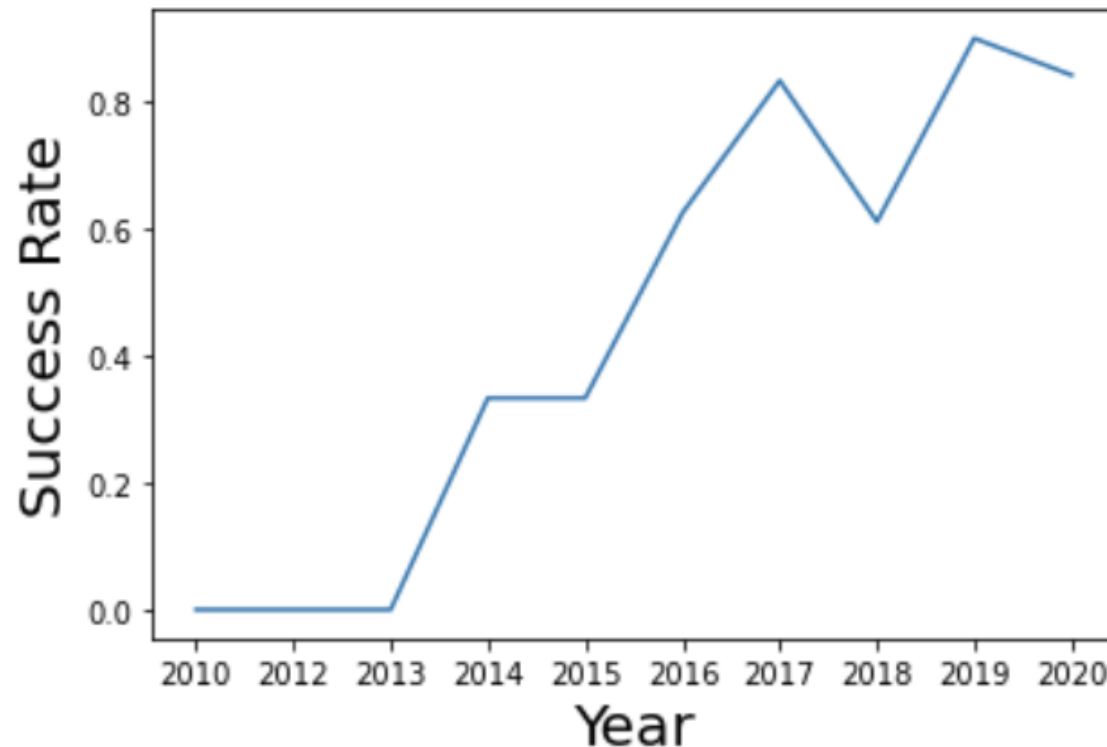
Flight Number vs. Orbit Type

A trend can be observed of shifting to VLEO launches in recent years.

Payload vs. Orbit Type

There is a high launch success rate between the ISS orbit and payload mass at the 2000 range, as well as for the GTO orbit between the 4000-8000 range.

IBM Developer

SKILLS NETWORK

Launch Success Rate YoY

Launch success rate has increased significantly since 2013, potentially due to advances in technology and lessons learned.
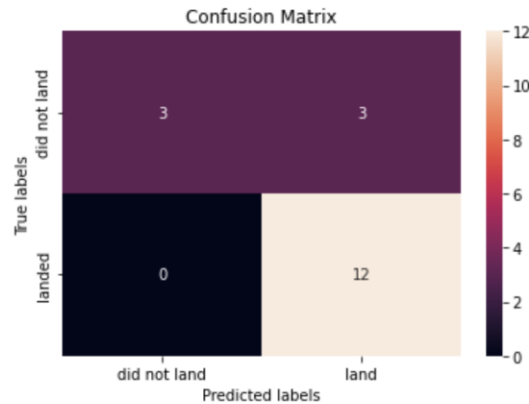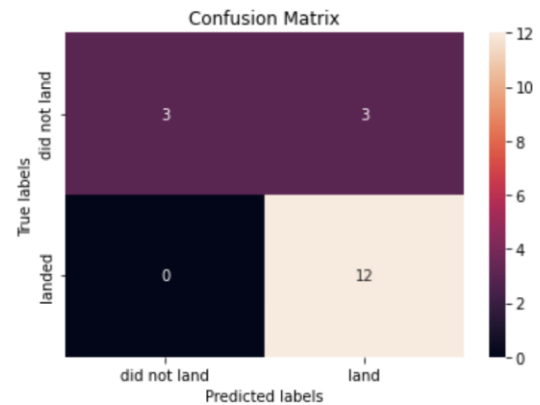
Launch Site Locations
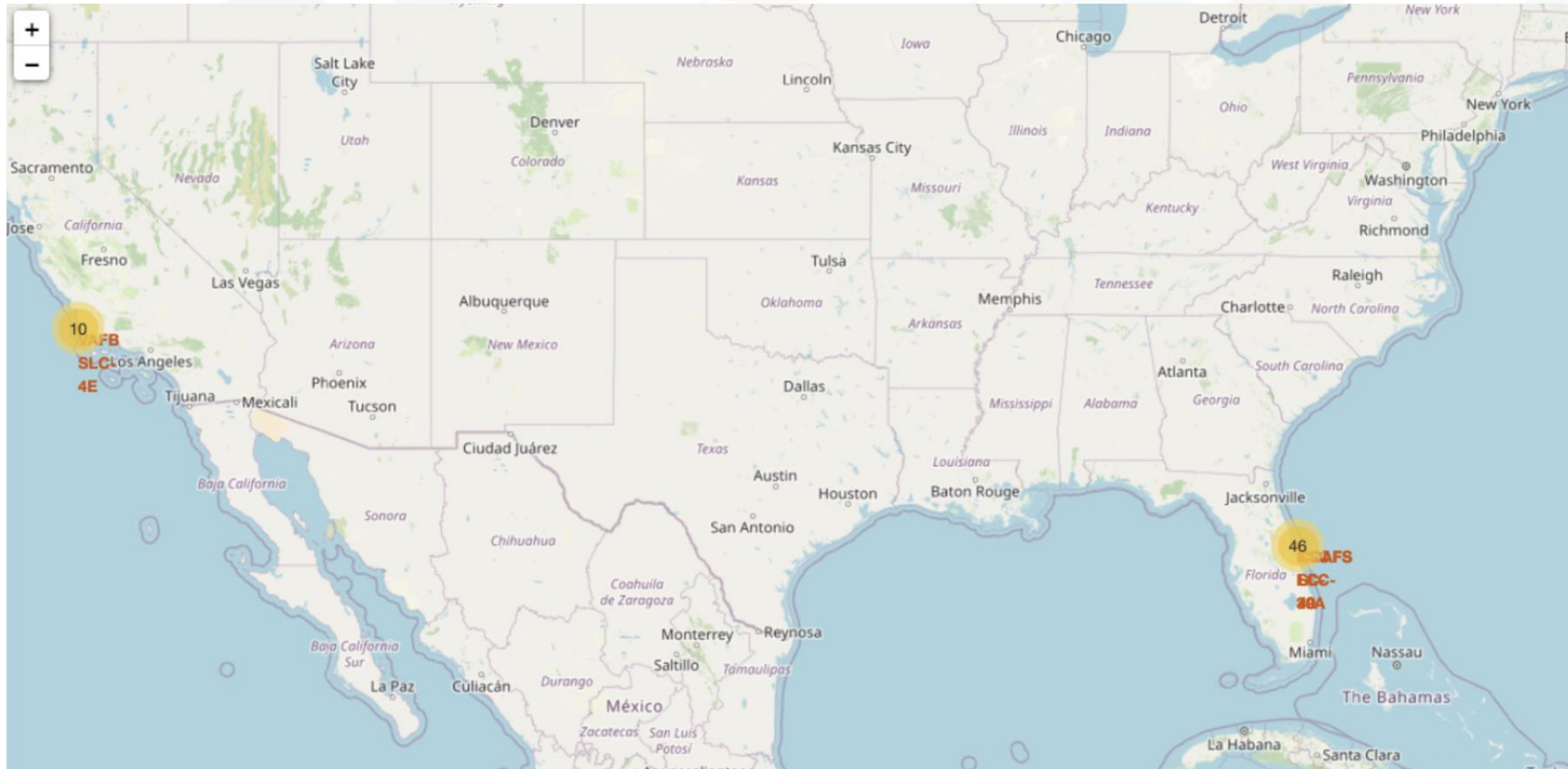
Optimal launch site locations are in close proximity to coastlines

IBM Developer

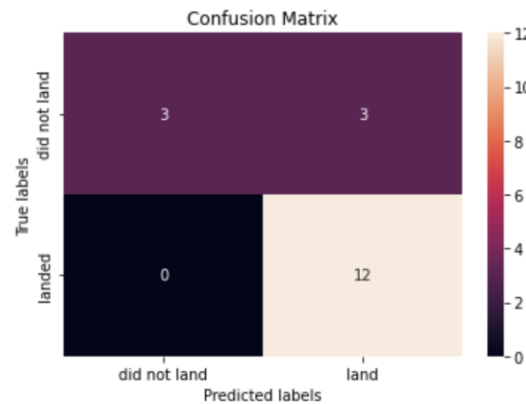SKILLS NETWORK

Success Rate Predictions
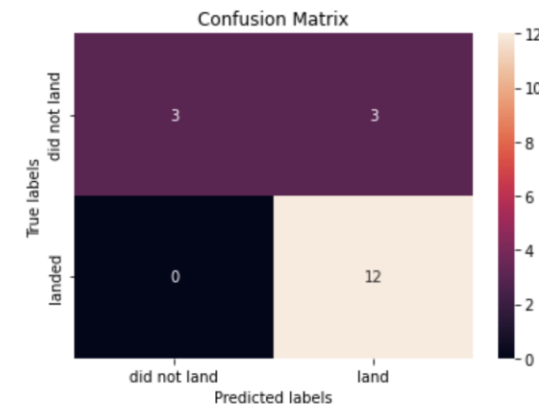
Logisitic Regression

SVM

Decision Tree

KNN

CONCLUSION

- The success rates for SpaceX launches is directly proportional time in years spent perfecting the launches.

- For predicting the success rate of future launches, it is recommended to use any of the SVM, KNN, and Logistic Regression algorithms.