Commercial Space Flights

Barbara van Zyl

10 September 2023

OUTLINE

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

EXECUTIVE SUMMARY

- **The Purpose**:
  - Predict if SpaceX Falcon 9 stage 1 will land successfully.
- **The Data**:
  - Data was collected using the SpaceX API and web scraping from public sources.
  - The data was cleaned using the Pandas library.
  - EDA was performed to determine training labels.
- **The Analysis:**
  - Data wrangling, data visualisation, geo analytics
  - Machine Learning predictions
- **The Conclusion**:

INTRODUCTION

- The main objective is to assess the ability of SpaceY to compete with SpaceX.
- Having Falcon 9 stage one land successfully significantly reduces launch costs.
- **Desired outcome of analysis:**
  - Predict the successful landing of stage one rocket launches based on launch sites.

METHODOLOGY

Data Collection

- **SpaceX API**
  - Source data from the SpaceX API (https://api.spacexdata.com/v4/launches/past)
  - Filter data to retain information about inly Falcon 9 launches
- **Web scraping**
  - Source Falcon 9 data from Wikipeida using BeuatifulSoup (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Wrangling

- Calculate launches per site and outcomes per orbit.
- Create a numeric outcome column per status.

Exploratory Data Analysis (EDA)

- **SQL:**
  - Identify distinct launch sites.
  - Calculate total payload mass carries by NASA-launched boosters.
  - Calculate average payload mass carried by Faclon 9.
  - Identify the first successful landing in ground pad.
  - Calculate total successful and failed launches.
  - Identify booster versions that carried the largest payload mass.

Exploratory Data Analysis (EDA)

- **Data Visualisation:**
  - Explore relationships between variables using scatterplots, bar charts, and line charts.

Interactive Data Visualisation

- **Geo analytics with Folium:**
  - Identify launch sites on a map.
  - Identify successes and failures at each location.
- Dashboarding with Plotly
  - Pie chart shows the percentage of successful launches per site.
  - Scatterplot shows the payload mass of successful luanches.
  - The combination of these two visuals indicates which site has been most successful in the past and what the associated payload masses were.

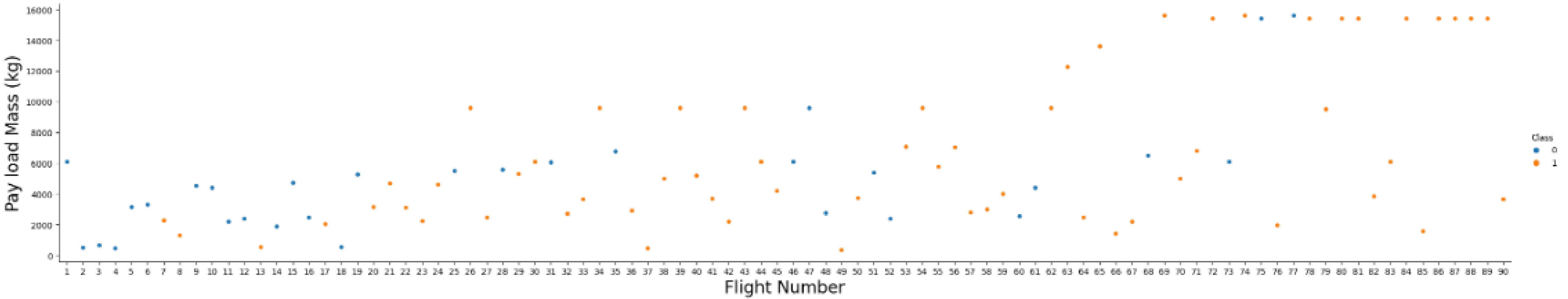
Predictive Analytics

- Data was split into training and testing sets.
- Data was evaluated using these models:
    - Logistic regression
    - Support vector machine
    - Decision tree
    - K-Nearest Neighbour (KNN)

RESULTS

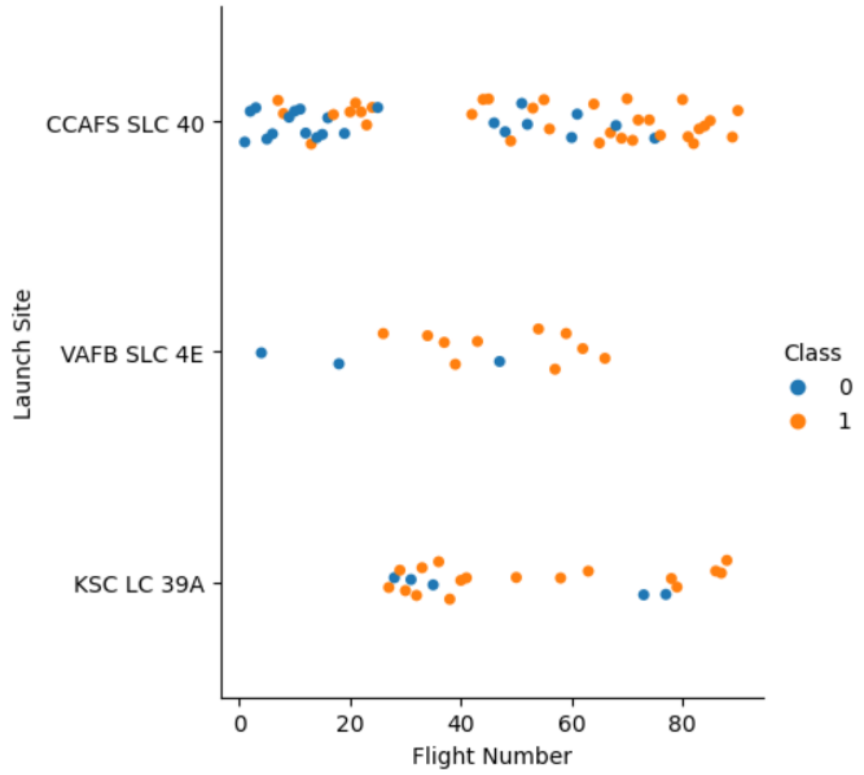
- The four launch sites are:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- The average payload mass for booster version F9 v1.1 is 2928.4kg.
- The first successful outcome on ground pad occurred on 22 December 2015.
- The boosters that have success in drone ship with payload mass between 4000kg and 6000kg:
  - F9 FT B1021.2
  - F9 FT B1031.2
  - F9 FT B1022
  - F9 FT B1026

RESULTS



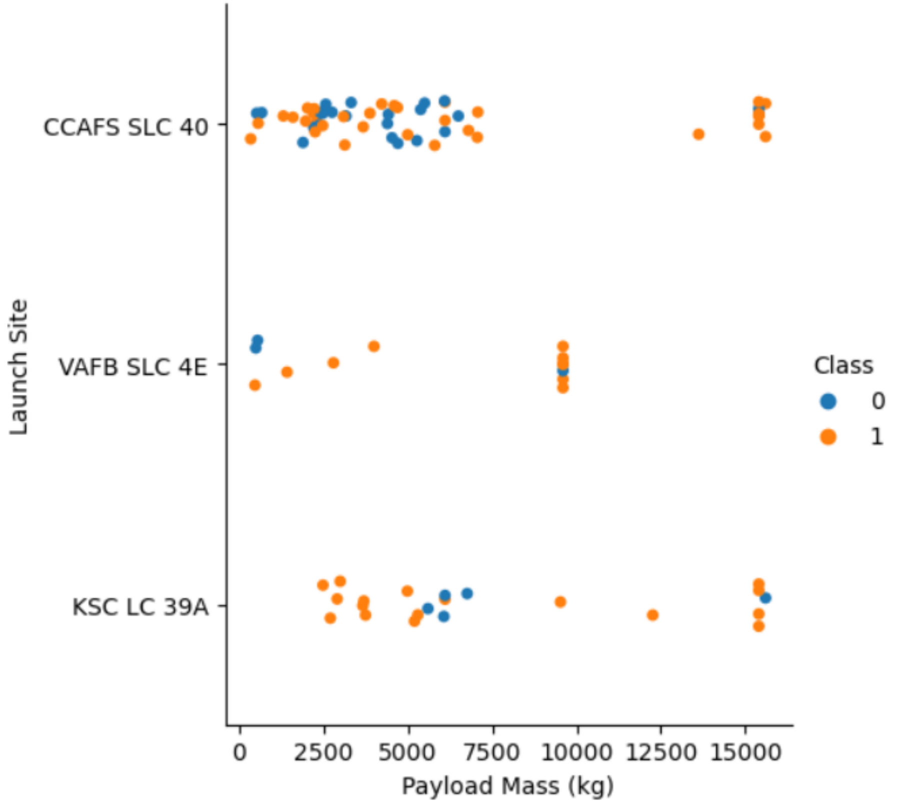
- Increased flight numbers results in more successful first stage landings.
- Higher payload masses result in fewer success successful first stage landings.
- The combination of increased flight numbers and high payload mass seems to result in successful landings.

RESULTS



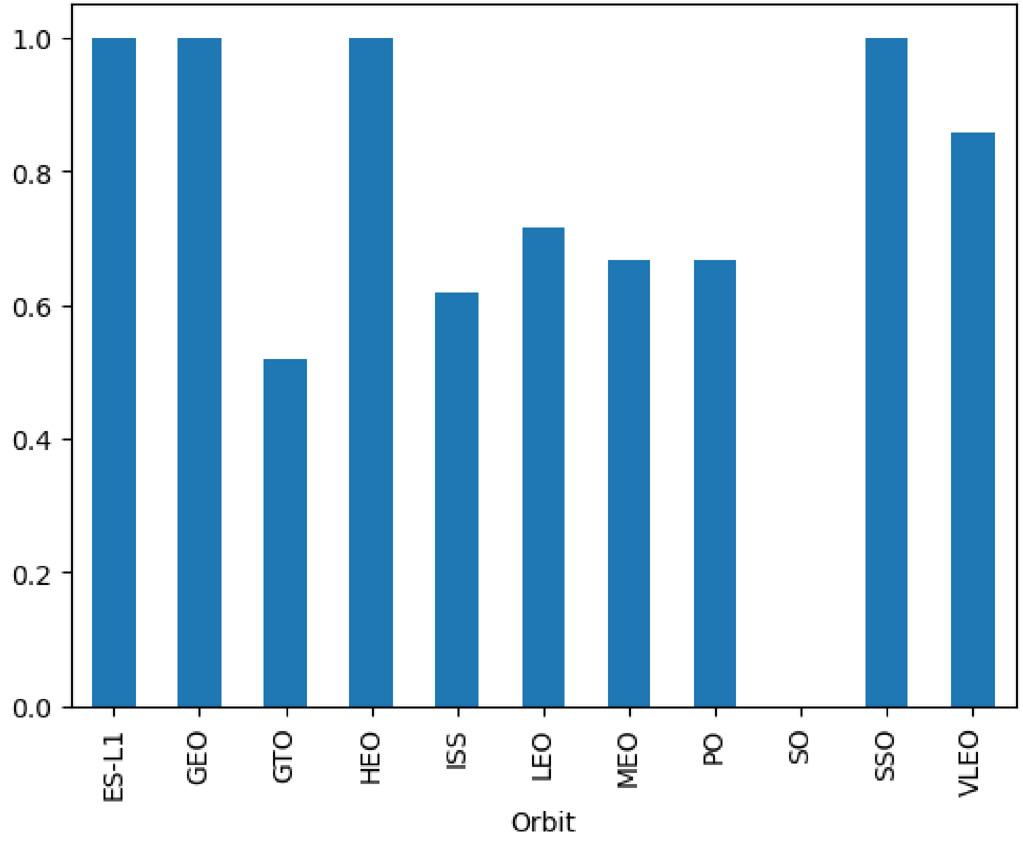
- Launch site CCFAS SLC 40 has the greatest number of flights.
- VAFB SLC 4E has the lowest number of flights, but also the lowest number of failures.
- KSC LC 39A has more flights than VAFB SLC 4E, but the number of failures are comparable.
- KSC LC 39A looks like a good launch location from this data.

RESULTS



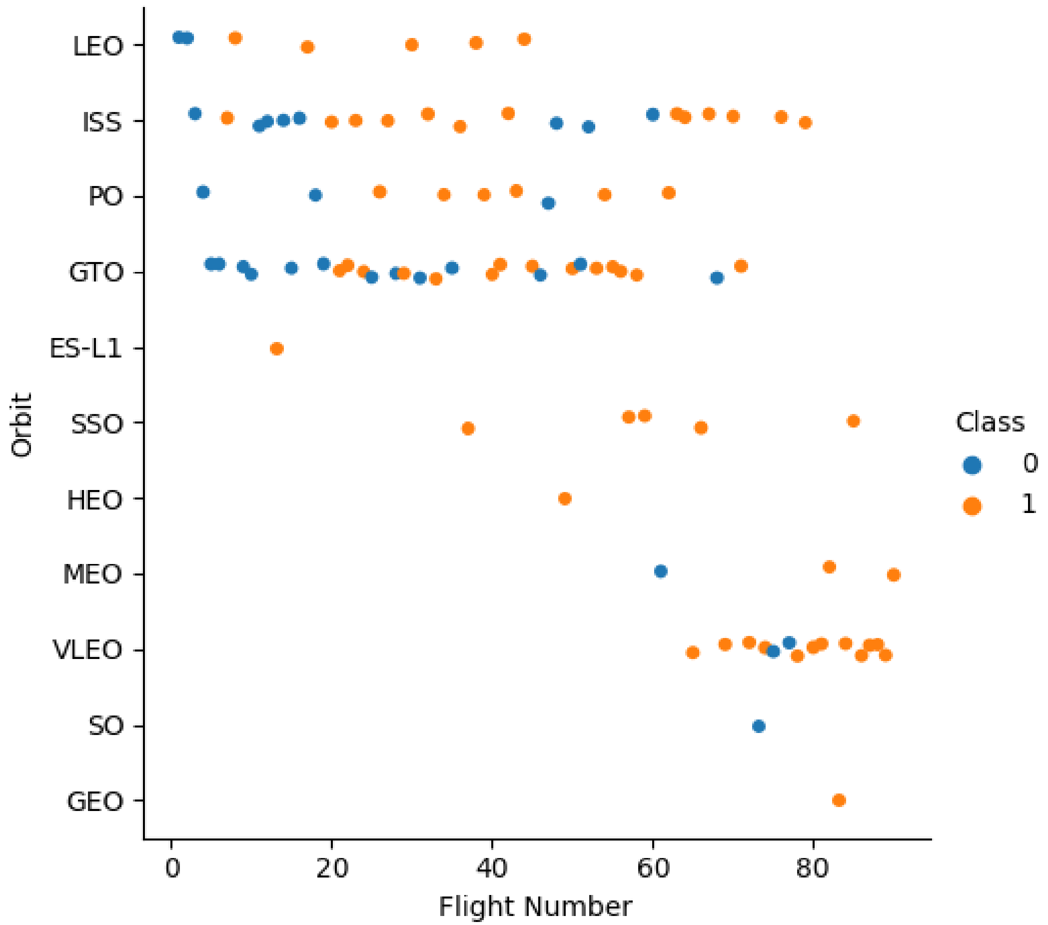
- CCFAS SLC 40 has greater success with payload masses of 15000kg.
- VAFB SLC 4E has greater success with payload masses of 10000kg.
- KSC LC 39A has greater success with payload masses of 15000kg.
- CCFAS SLC 40 and KSC LC 39A are comparable for payload masses of 15000kg.

RESULTS



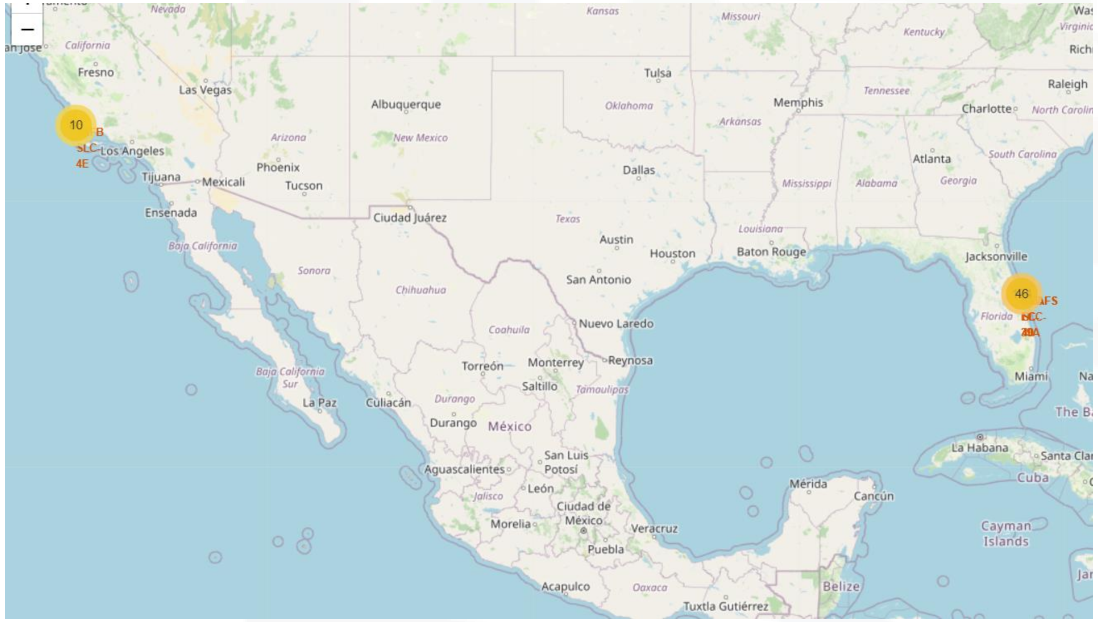
- Orbits ES-L1, GEO, HEO, and SSO results in the greatest successful landings.

RESULTS



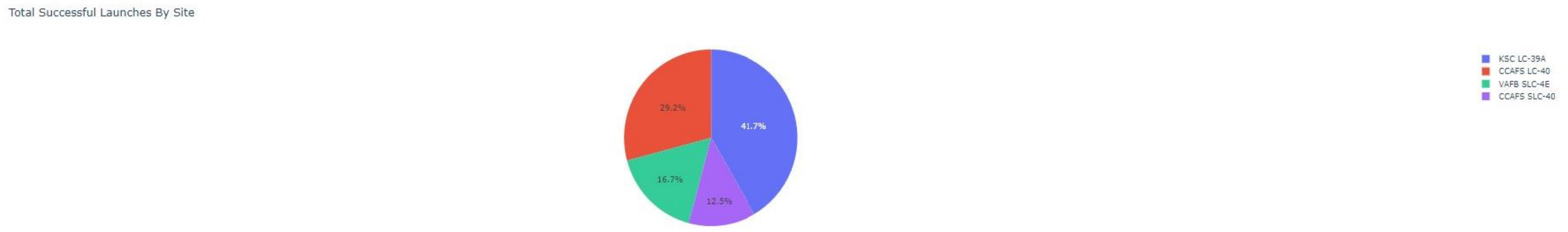
- LEO orbit is the only orbit where success appears realted to number of flights.

RESULTS



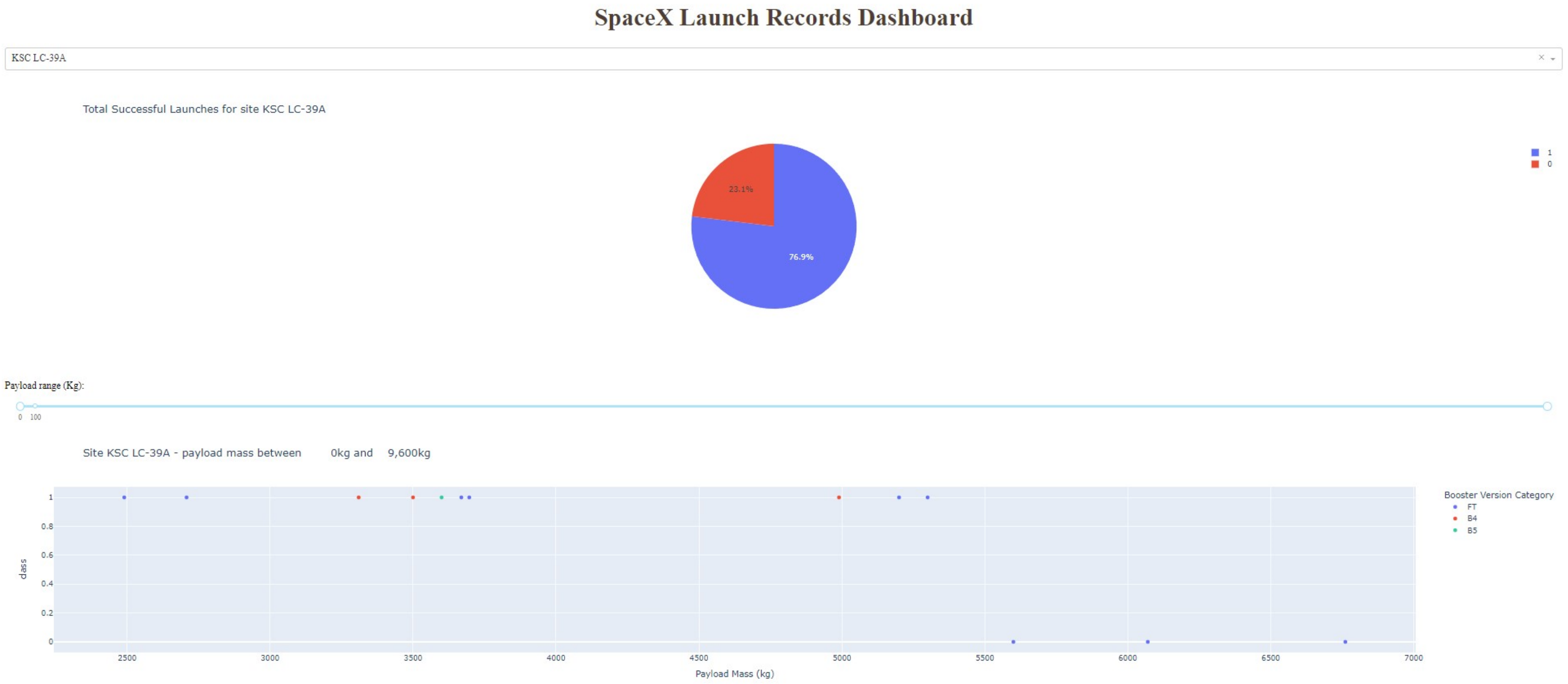
- Launching from sites closer to the coast appears to result in greater success.

RESULTS



- The greatest percentage of successful landings occurred from launch site KSC LC-39A.

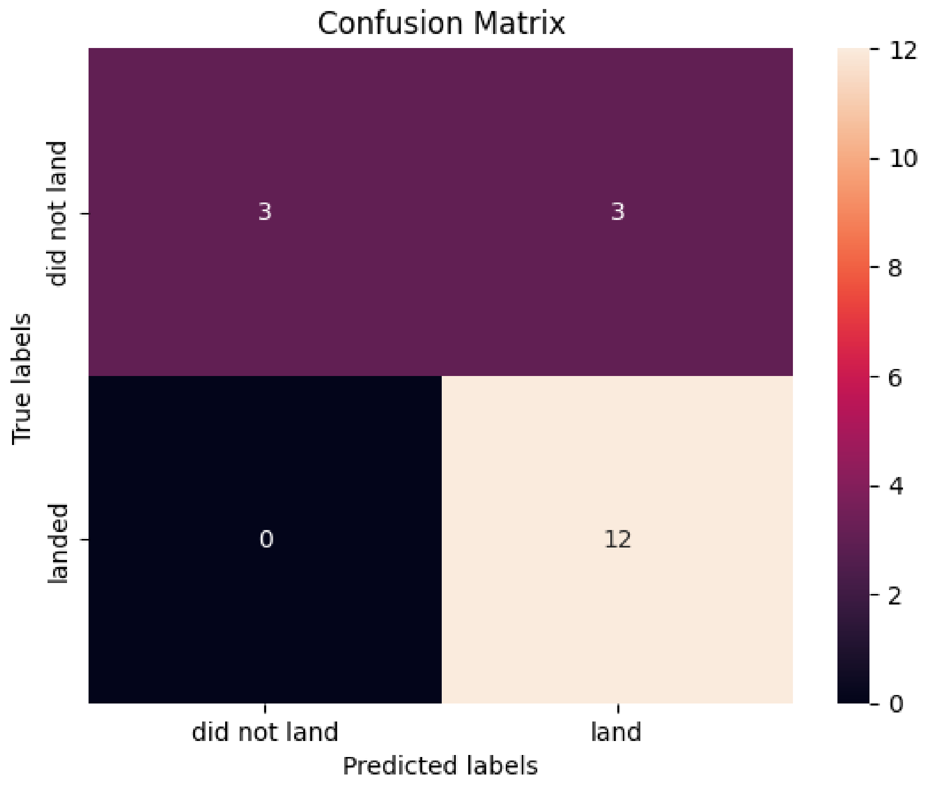
RESULTS



- For site KSC LC-39A, the booster version FT is the only version with failures.

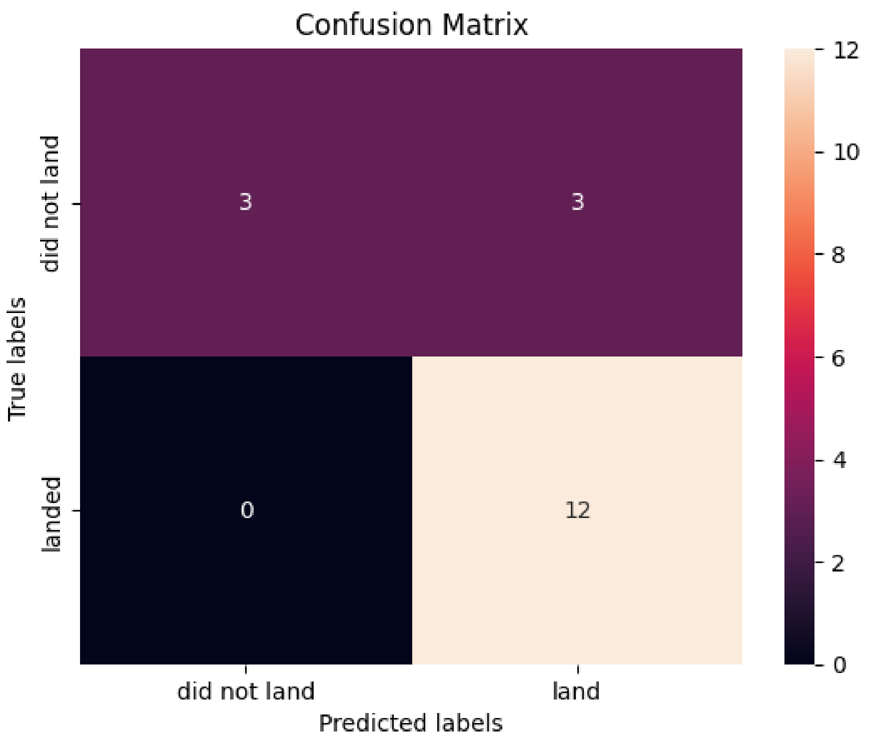
IBM Developer SKILLS NETWORK

RESULTS



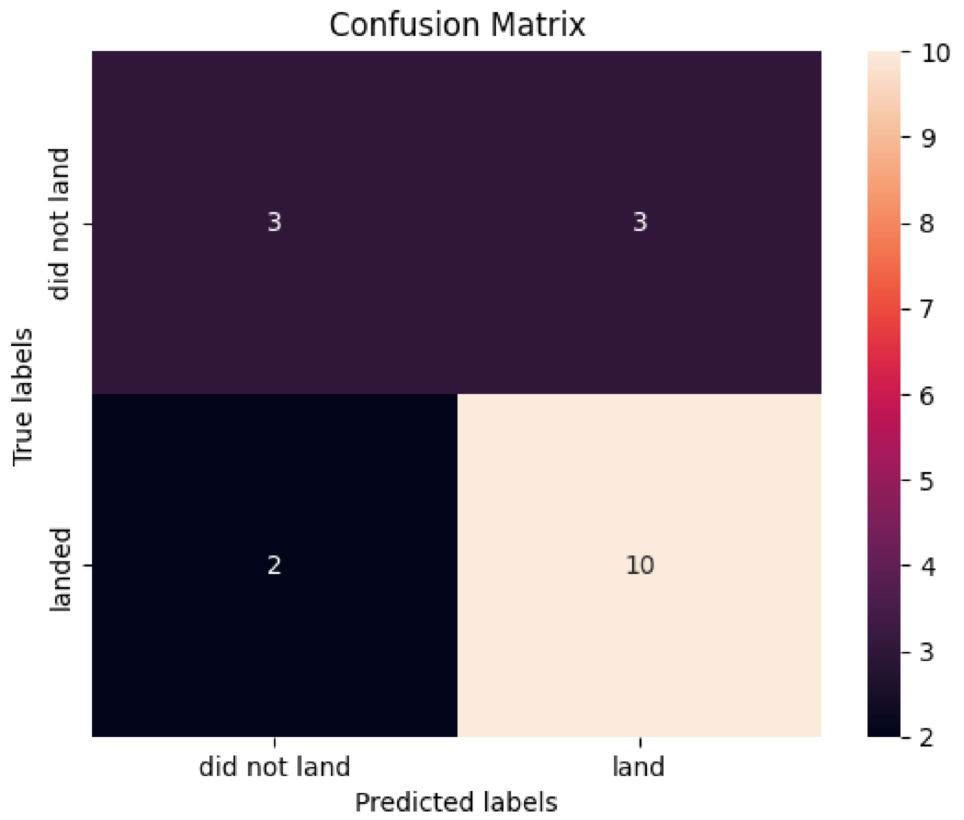
- For regression testing, the true positive is 12 and the true negative is 3.

RESULTS



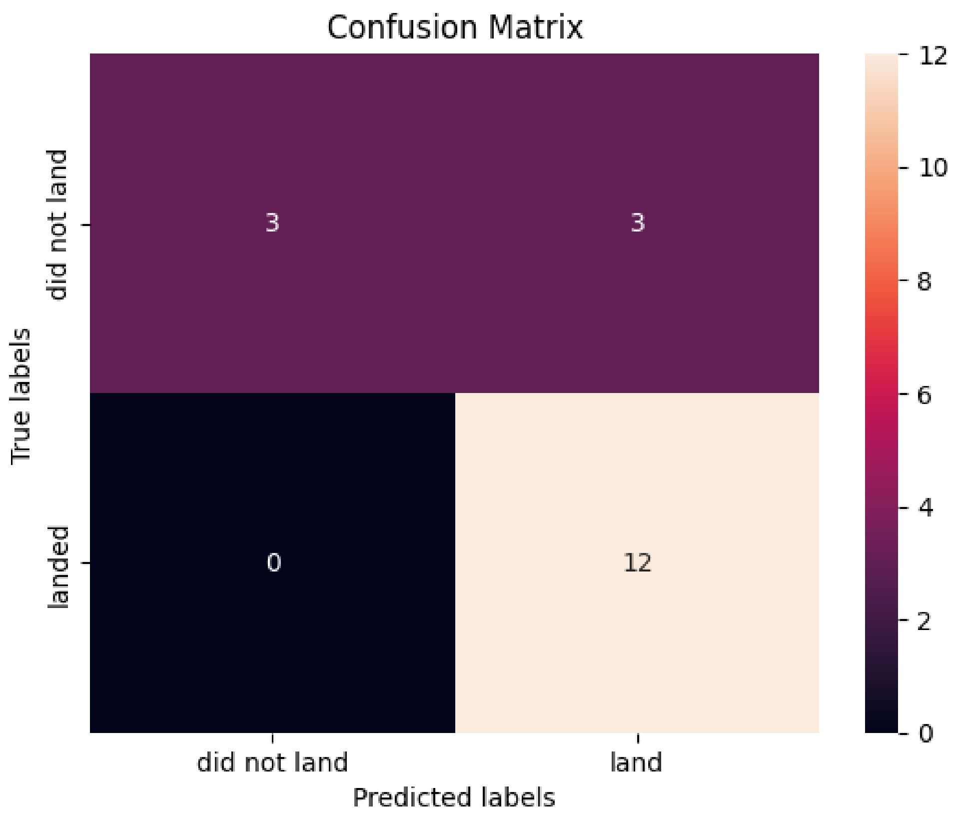
- For the support vector machine model, the true positive is 12 and the true negative is 3.

RESULTS



- For the decision tree model, the true positive is 10 and the true negative is 3.

RESULTS



- For the KNN model, the true positive is 12 and the true negative is 3.

RESULTS

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.90179	0.72222
KNN	0.84821	0.83333

- The least successful predictive model was the decision tree model, based on the confusion matrices.
- The Support Vector Machine and KNN models appear to have the greatest accuracy when predicting successful landings.

CONCLUSION

- The best overall launch site is KSC LC-39A.
- LEO orbit successes are tied to number of flights and therefore the best launch site for this orbit is CCFAS SLC 40.
- Payload masses exceeding 10000kg result in greater success.
- SVM and KNN models should be used to predict landing outcomes.