SpaceX Falcon 9 successful landings

Jonas Karosas

jonaskarosas919@gmail.com

2022-08-12

EXECUTIVE SUMMARY

SpaceX is a very successful company which can save around 100 millions dollars on a rocket launch because it reuses the first stage. But sometimes the first stage does not land and can crash. By collecting the existing data and using data science tools, it is possible to make predictions about landing outcomes. It this works it is presented that the more massive the payload, the less likely the first stage will return except for Polar, LEO and ISS orbits. Also, no rockets were launched for heavy payload mass in VAFB-SLC launch site.

In this work it is showed that in the LEO orbit the success is related to the number of flights, but in GTO orbit there are no relationship between flight number. Success rate since 2013 kept increasing till 2020 and overall, the KSC LC-39A launch site has the highest successful outcome rate.

INTRODUCTION

- Perhaps the most successful companies among commercial space age is SpaceX. It advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage [1]. Sometimes the first stage does not land or even crash. The aim of this project is to use the public information and with data science tools predict if SpaceX will reuse the first stage or not.

[1] Lance a. Davis, First Stage Recovery, Engineering 2 (2016), 152-153

METHODOLOGY: DATA COLLECTION AND WRANGLING

Data collection

1) Data collection:

- Request and parse the SpaceX launch data using the GET request;
- Decode the response content as a Json using .json() and turn it into a Pandas dataframe using .json_normalize();
- Filter the dataframe to only include Falcon 9 launches.

2) Data wrangling:

- Clean the data from missing values;
- Create the outcome label from the ‘Outcome’ column, where 1 stands for successful landing and 0 - for failure.

METHODOLOGY: SQL

3) Structured Query Language:

Use SELECT FROM statement together with the following clauses:

- WHERE; LIKE; BETWEEN; GROUP BY; ORDER BY; COUNT; SUM; AND.

Make different findings:

- Unique launch site names;
- Total payload mass carried by boosters launched by NASA (CRS);
- An average payload mass carried by booster version F9 v.1.1;
- The data of the first successful outcome in the ground pad;
- The total number of successful and failure mission outcomes;
- Booster versions names which have carried the maximum payload mass.

METHODOLOGY: DATA VISUALIZATION

4) Data visualization:

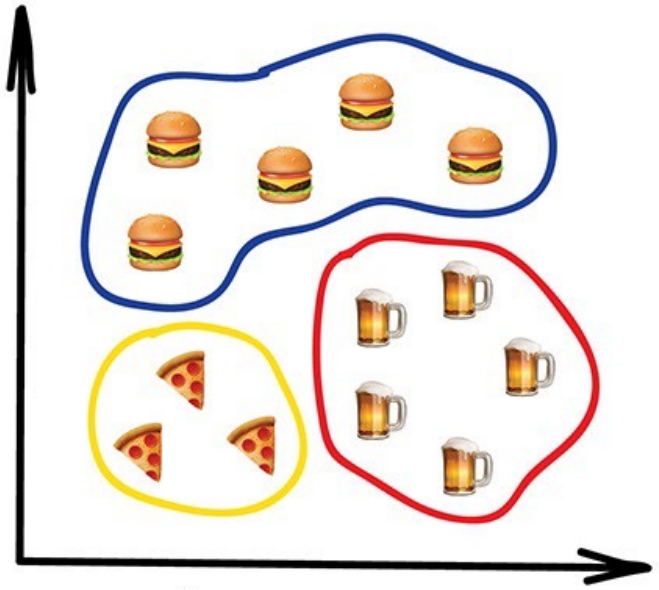
Plot different relations using seaborn library:

- Flight Number vs Payload Mass or Launch Site or Orbit type;
- Payload Mass vs Launch Site or Orbit type;
- Success rate of each orbit type;
- Yearly trend of the success launch.

Show launch site location on the map using folium library:

- All launch sites;
- Success/failed launches for each site;
- Distances between a launch site to its proximities

METHODOLOGY: MACHINE LEARNING PREDICTION



5) Machine learning prediction:

- Create a column for the class;
- Standardize the data;
- Split the data into training data and test data;
- Find the best Hyperparameters for:
  a) Support Vector Machine (SVM);
  b) Decision Tree;
  c) Logistic Regression;
  d) K Nearest Neighbour (KNN);
- Find which method performs best using test data.

RESULTS: SQL

Unique launch sites in the space mission:

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

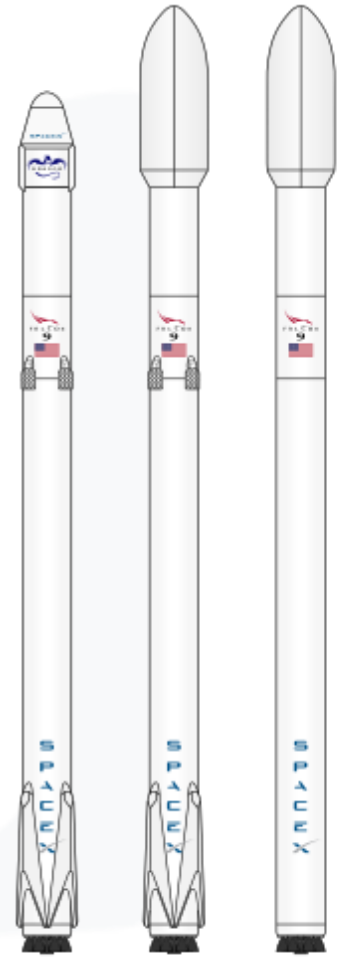
Total number of successful and failure mission outcomes:

- Failure (in flight) - 1
- Success – 99
- Success (payload status unclear) – 1

Total payload mass carried by boosters launched by NASA (CRS): 48213 KG

Average payload mass carried by booster version F9 v1.1: 2534.66 KG

The first successful landing outcome in ground pad was achieved: 22-07-2018



Falcon 9 v1.1

RESULTS: Flight Number vs Payload Mass

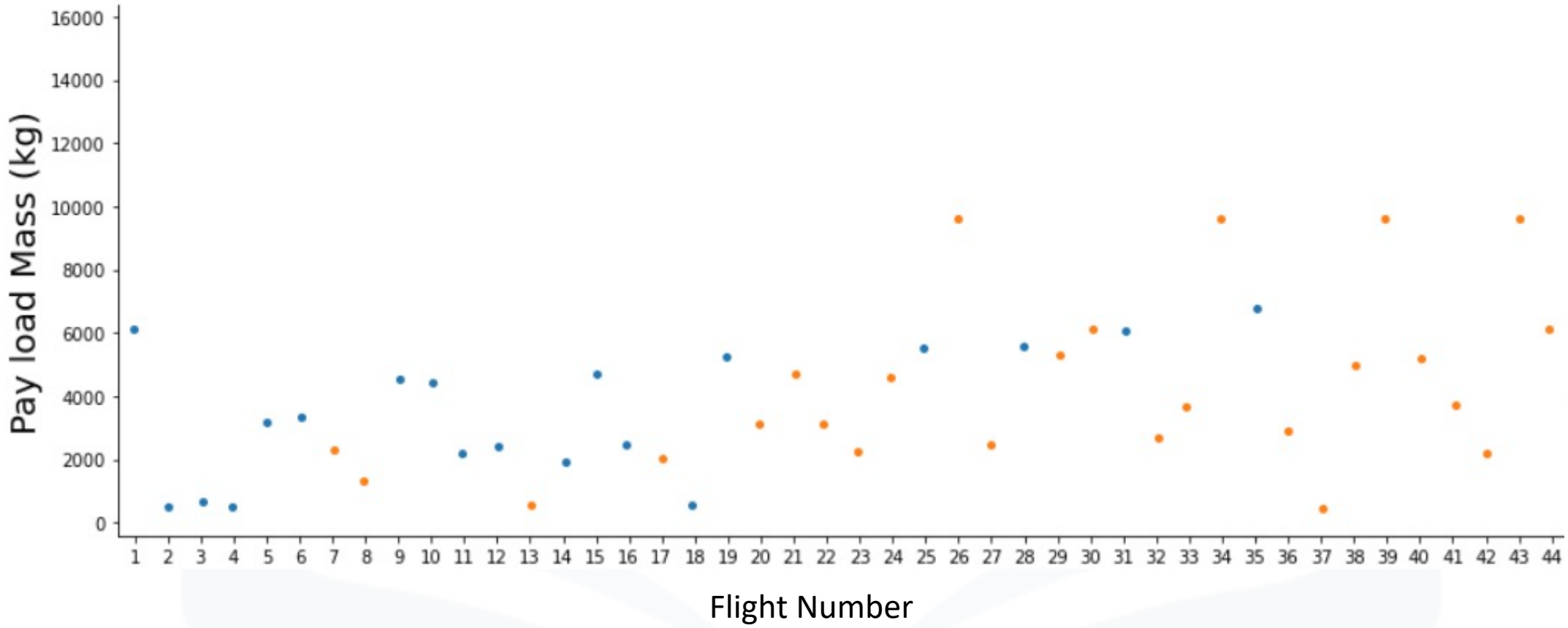


Fig 1. Flight Number (1-44) vs Payload Mass.

RESULTS: Flight Number vs Payload Mass

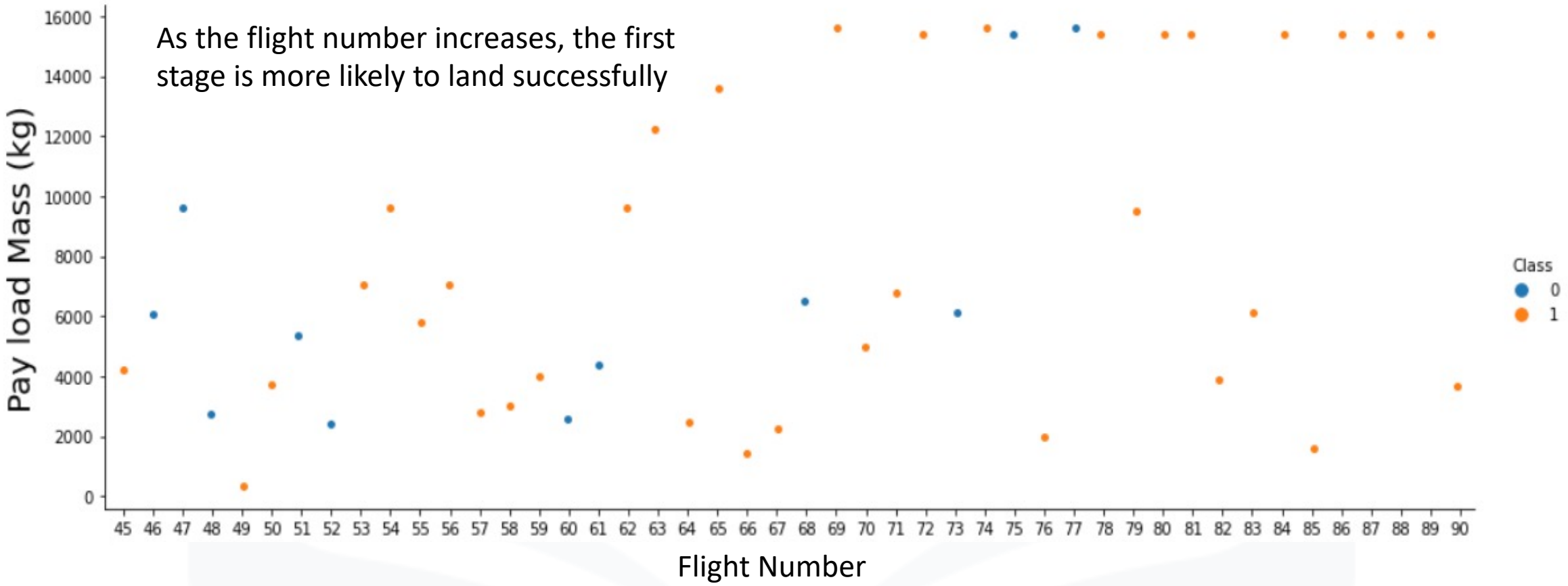


Fig. 2 Flight Number (45-90) vs Payload Mass

RESULTS: Launch Site relations

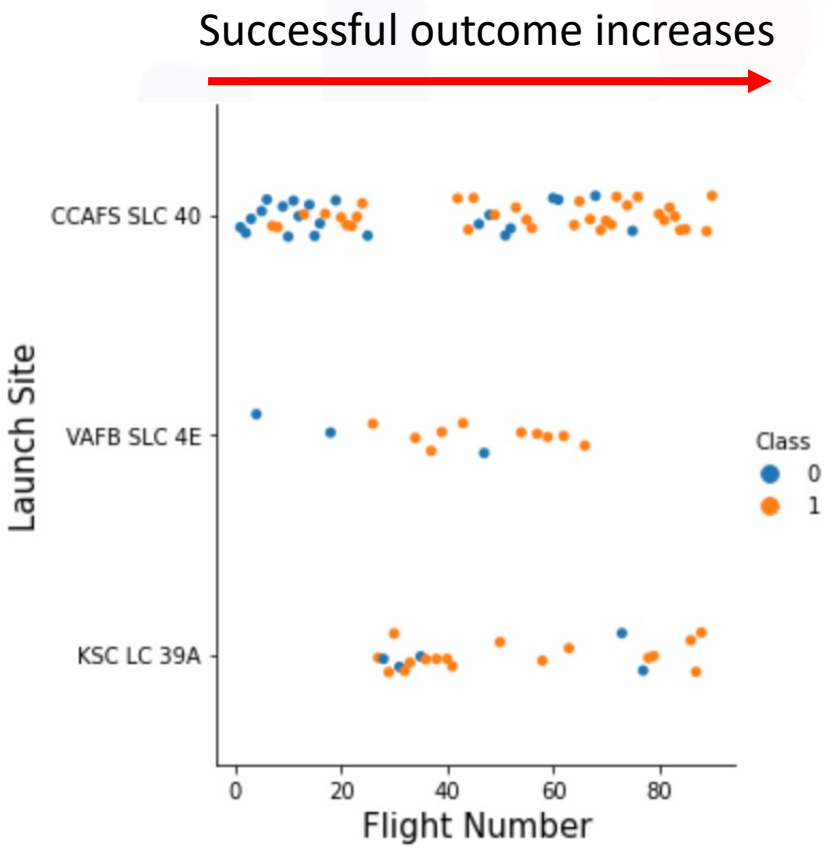


Fig 3. Flight Number vs Launch Site

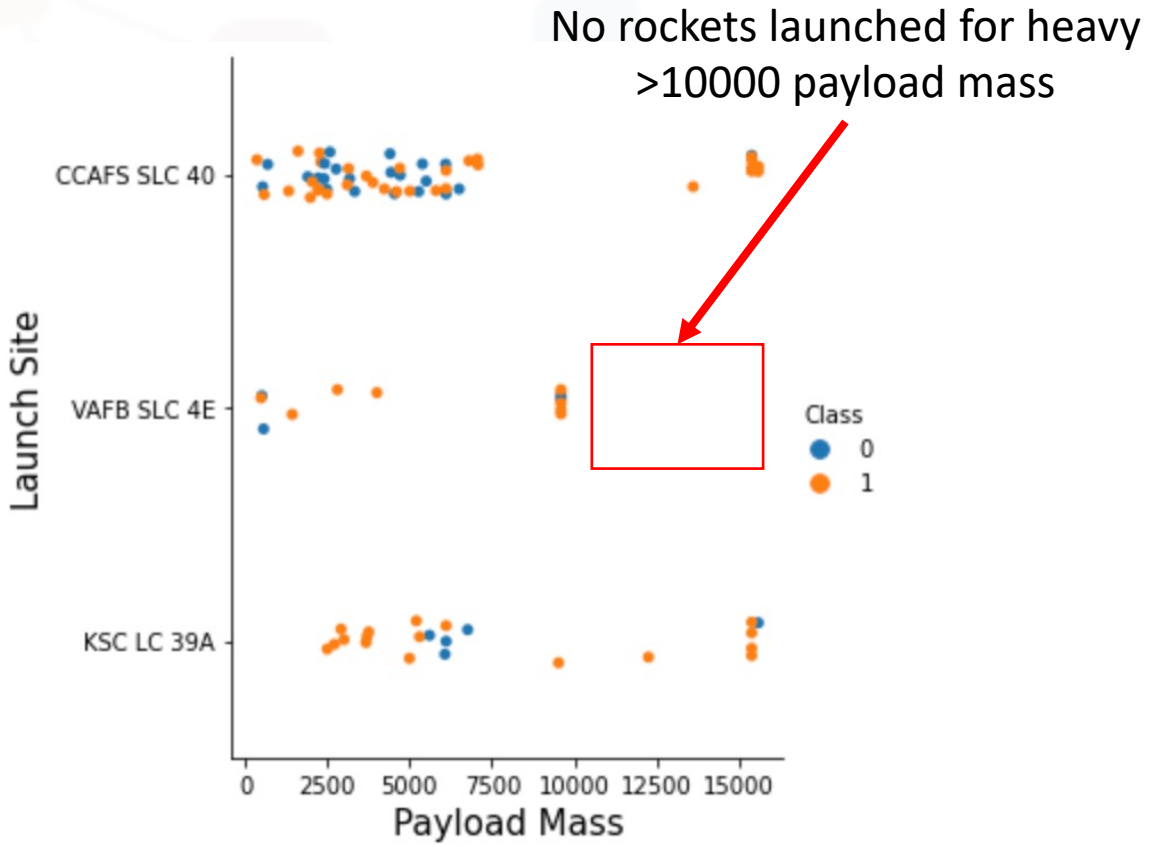


Fig 4. Payload Mass vs Launch Site

RESULTS: Orbit type relations

Success appears related to the number of flights

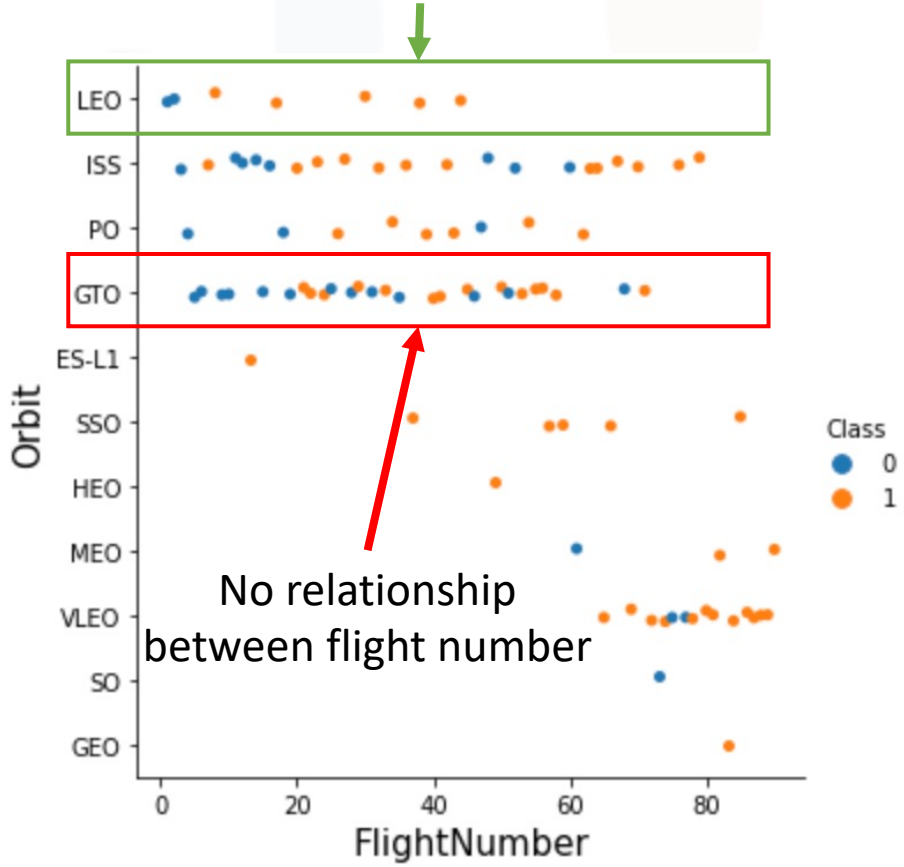


Fig. 5 Flight Number vs Orbit Type

Successful landing with heavy payloads

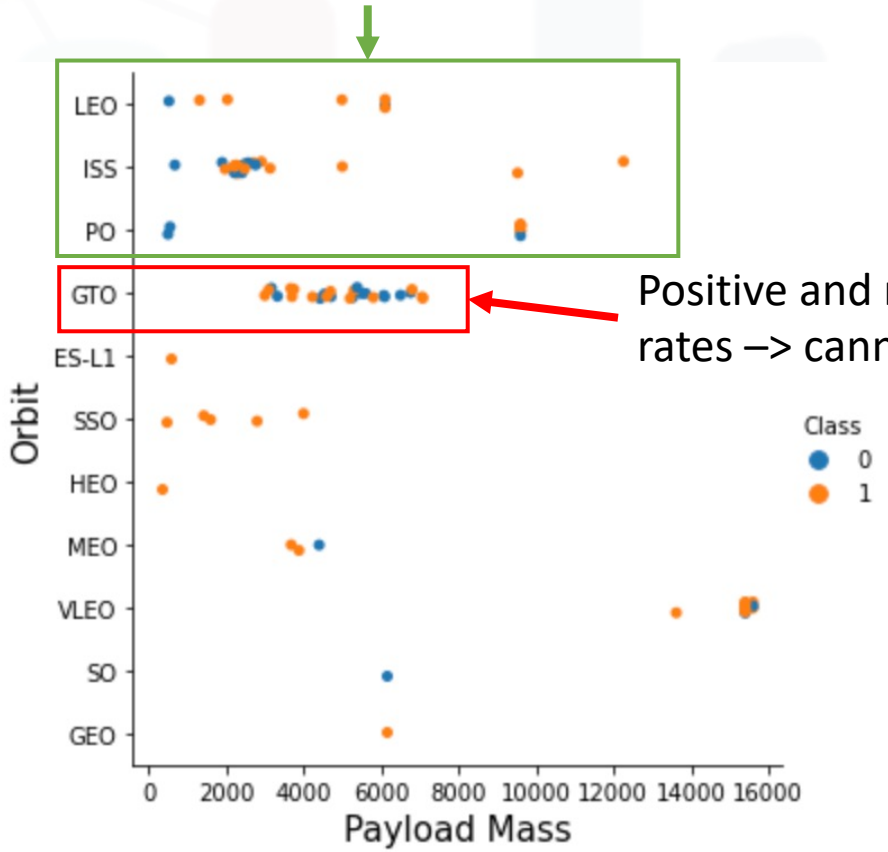


Fig. 6 Payload Mass vs Orbit Type

RESULTS: Success rate and trend

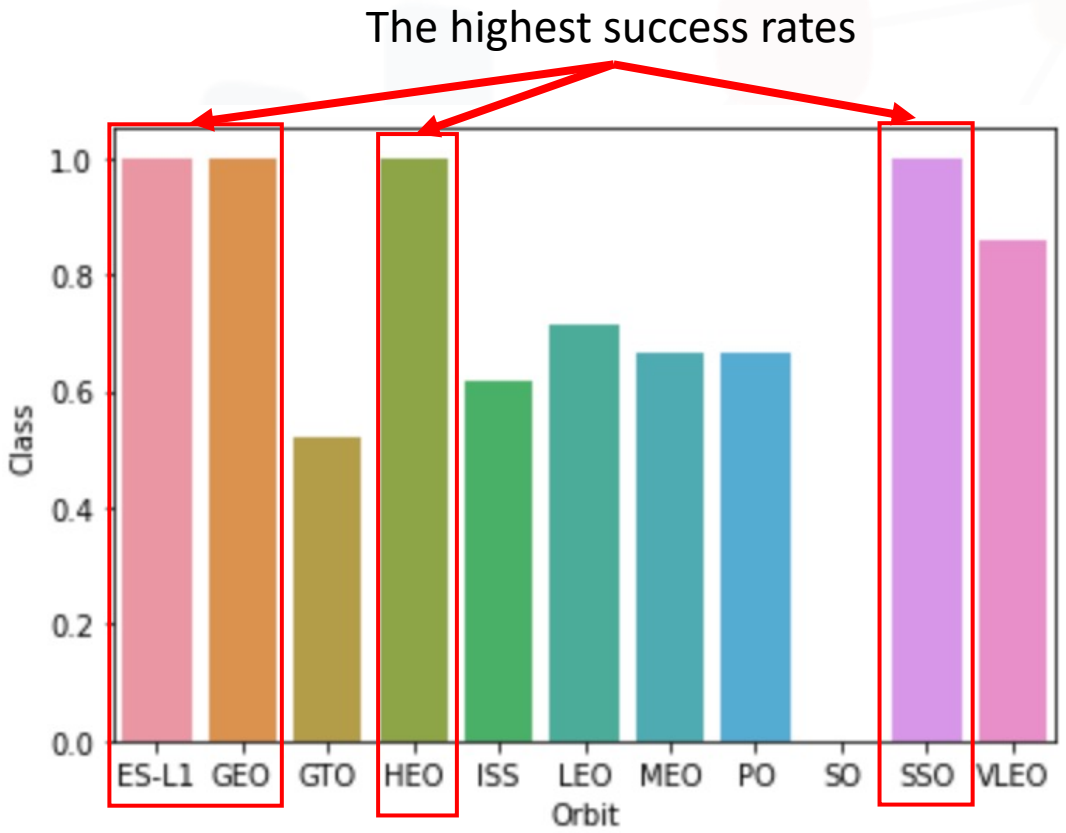


Fig. 7 Success rate of each orbit type.

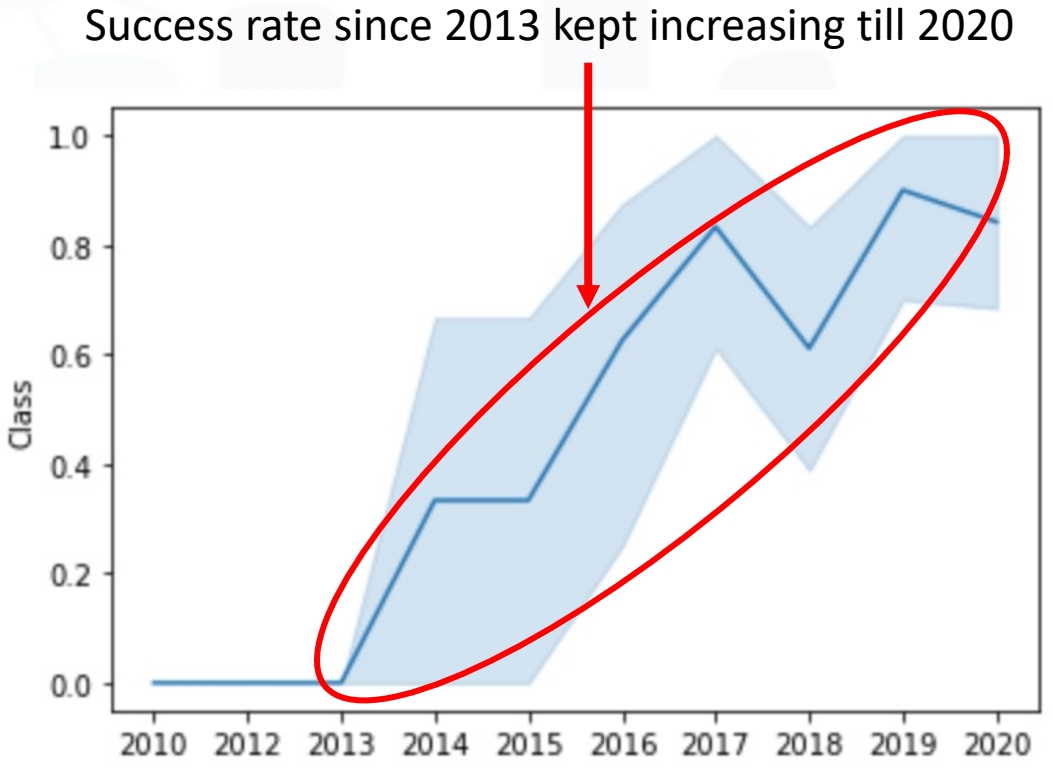
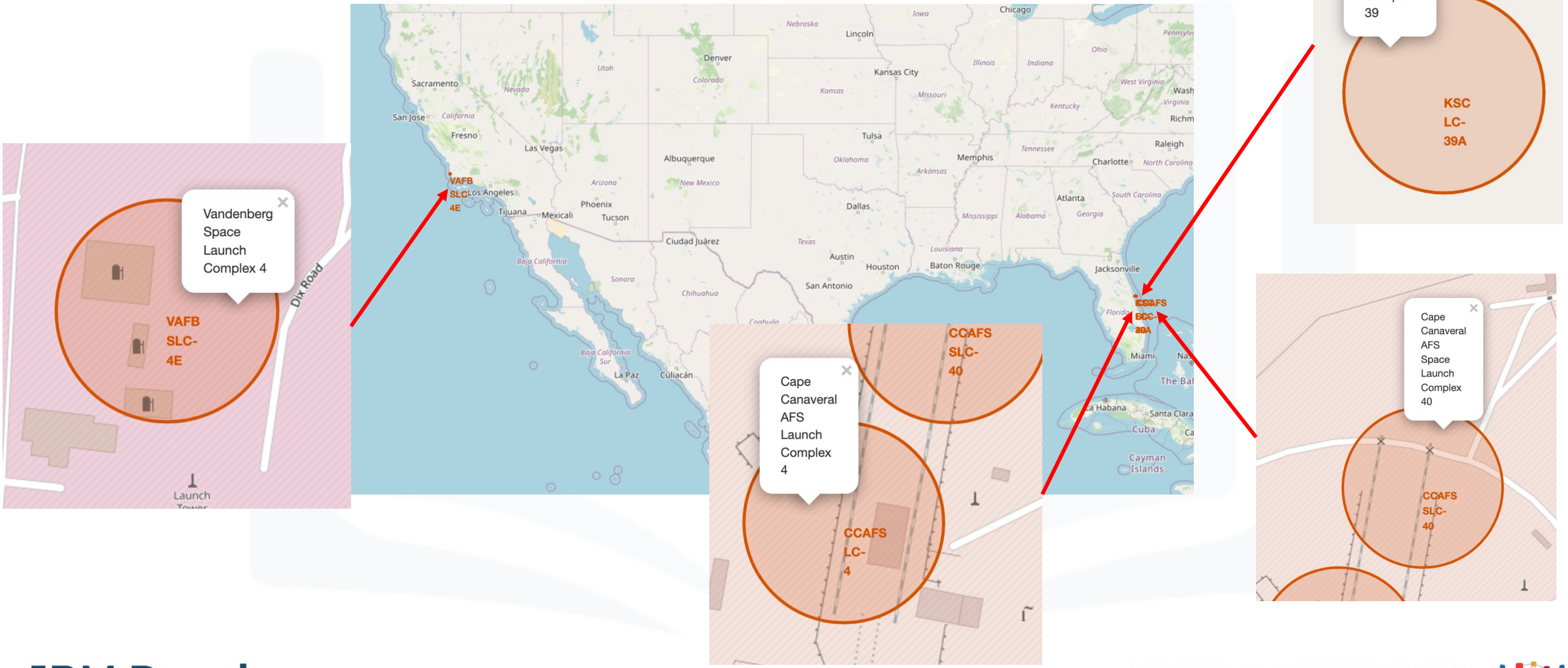


Fig. 8 Yearly trend of the successful launch.

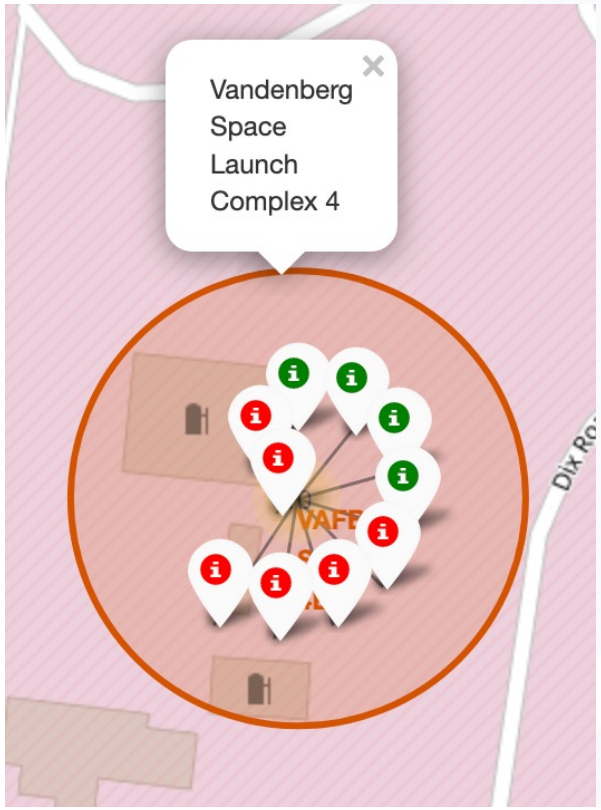
RESULTS: All launch sites on map

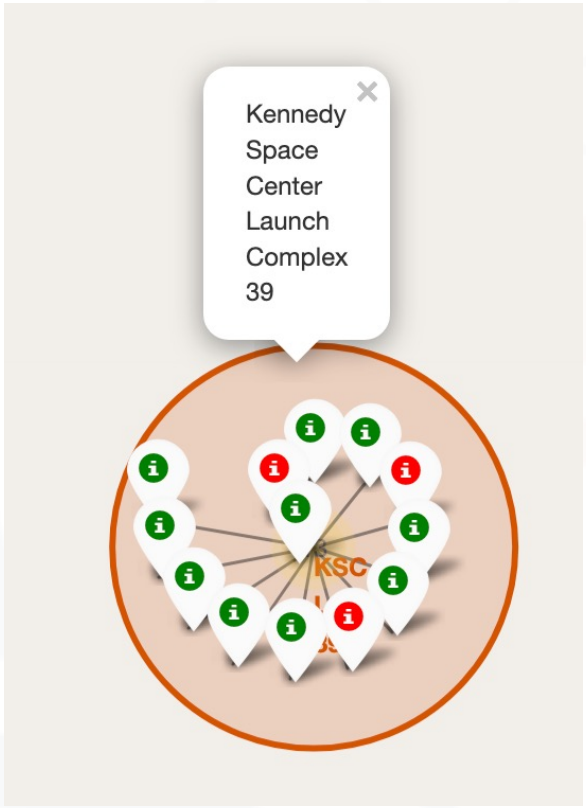


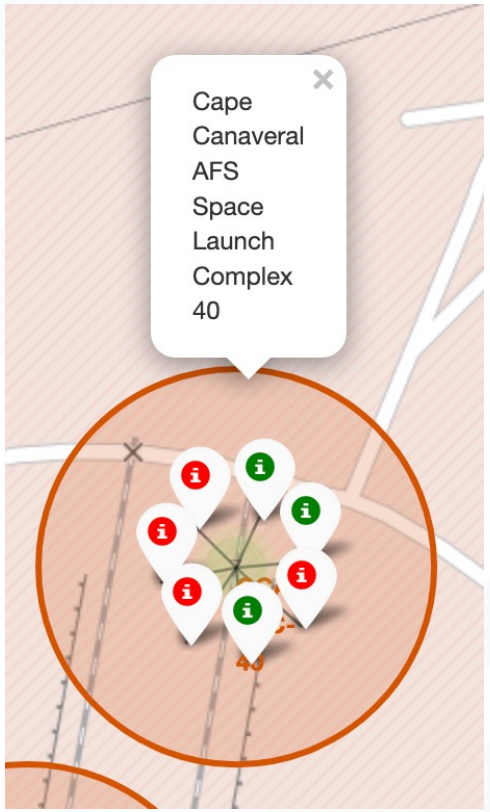
IBM Developer

SKILLS NETWORK

RESULTS: Success/Failed launches

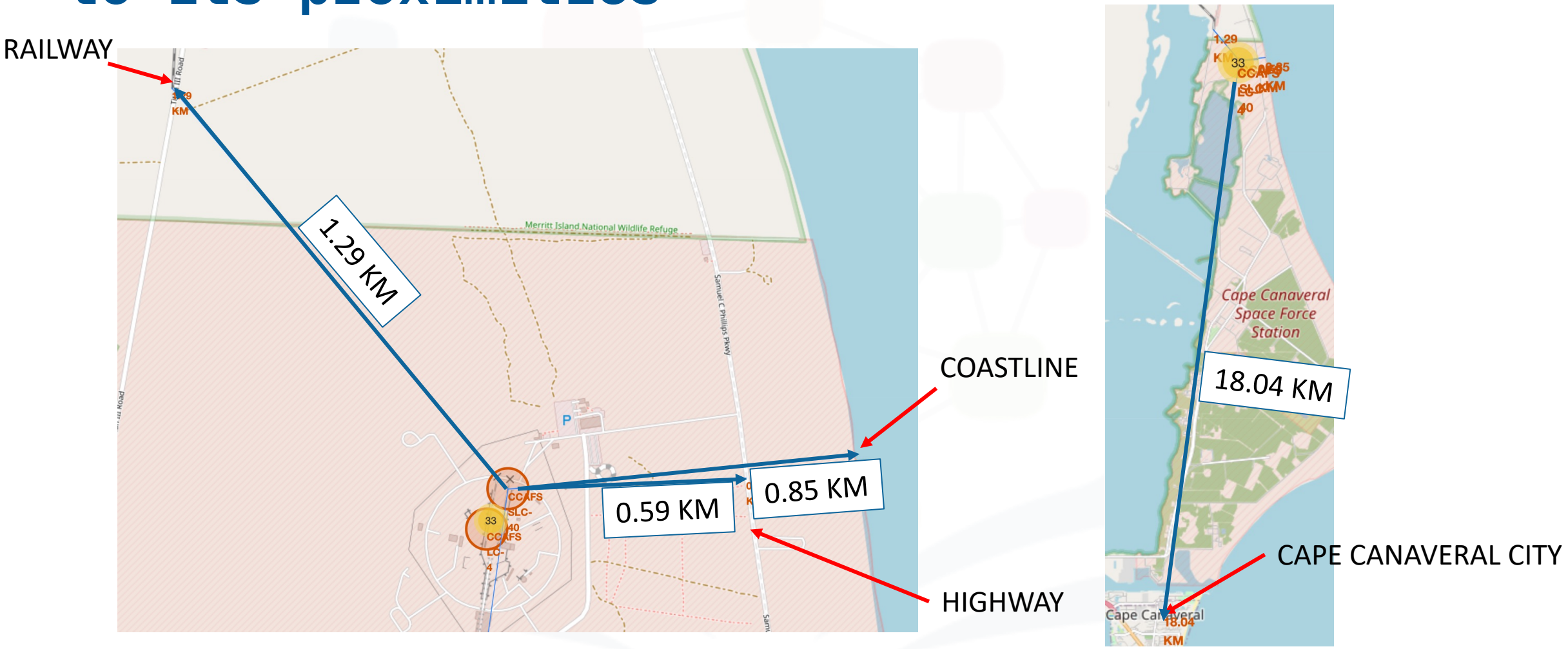




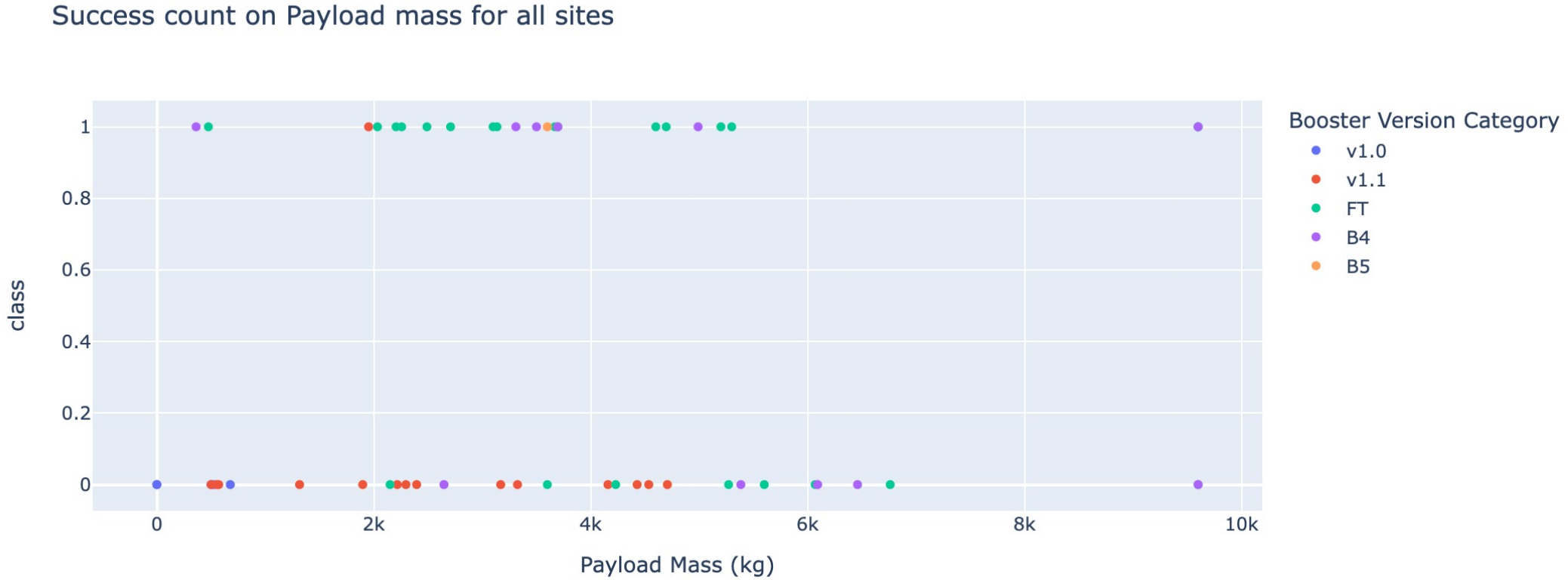


Cape Canaveral AFS Launch Complex 4

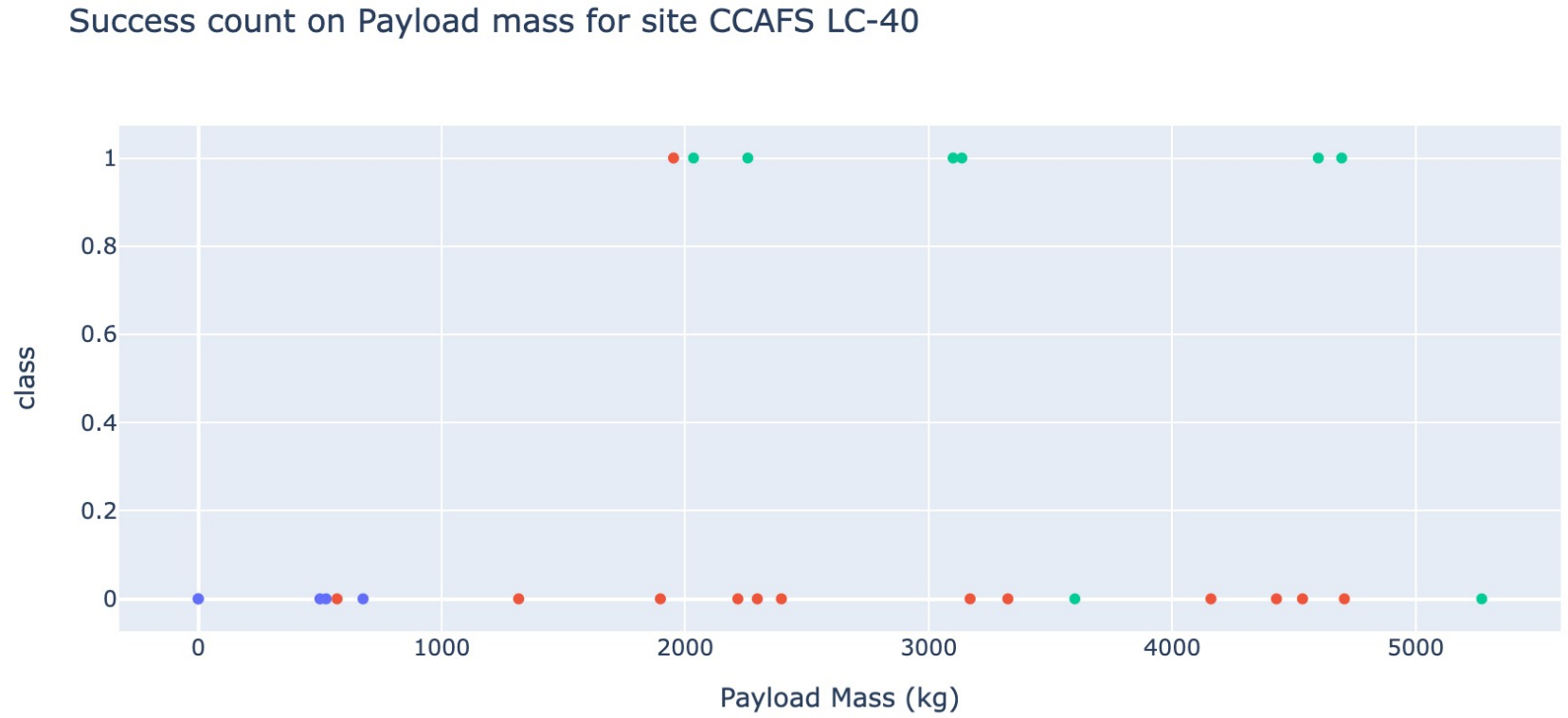
RESULTS: Distances between CCSAFS SLC-40 to its proximities

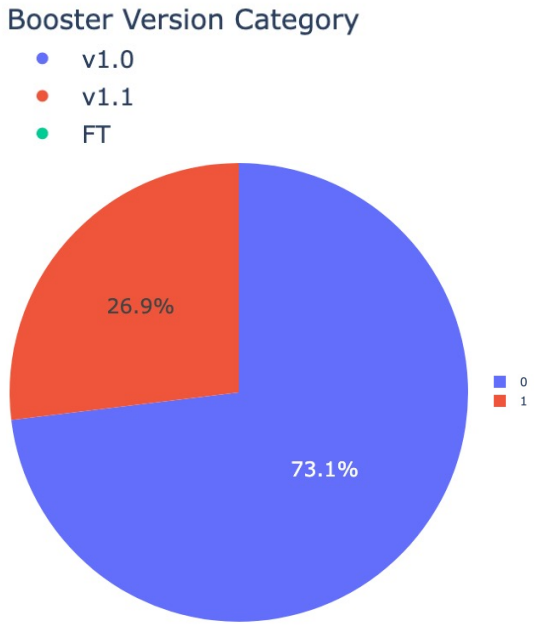


RESULTS: SpaceX Launch Record Dashboard

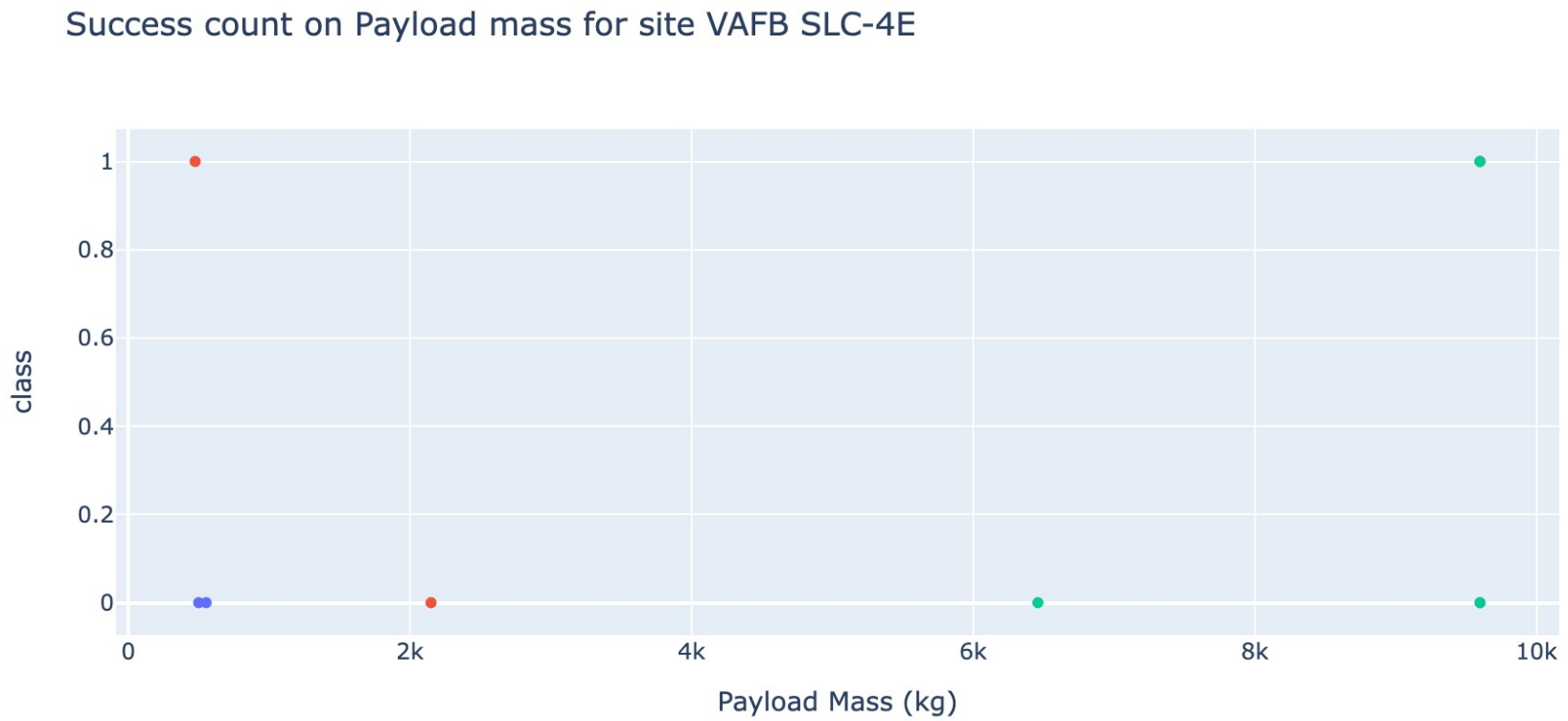


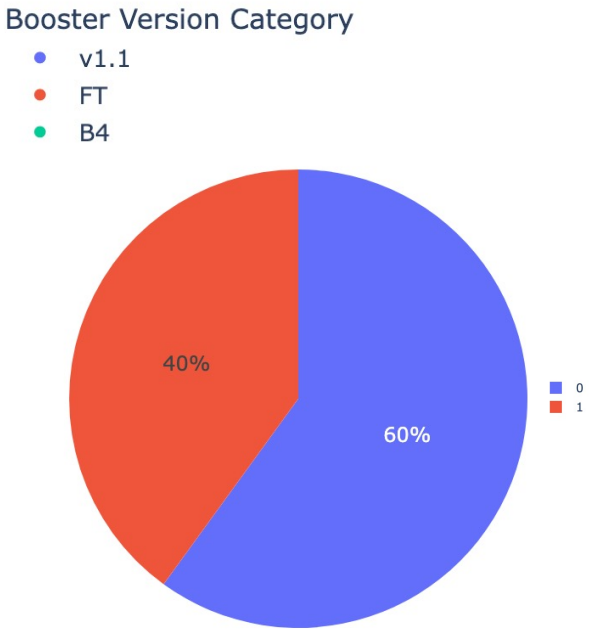
RESULTS: SpaceX Launch Record Dashboard



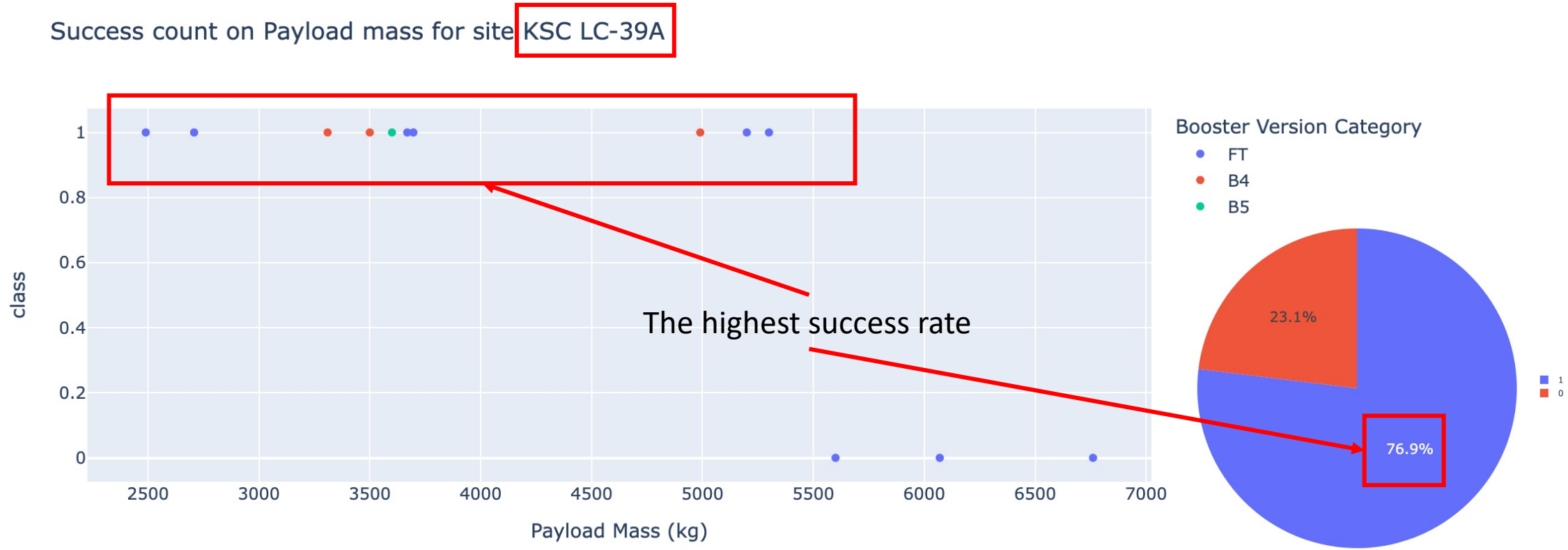


RESULTS: SpaceX Launch Record Dashboard



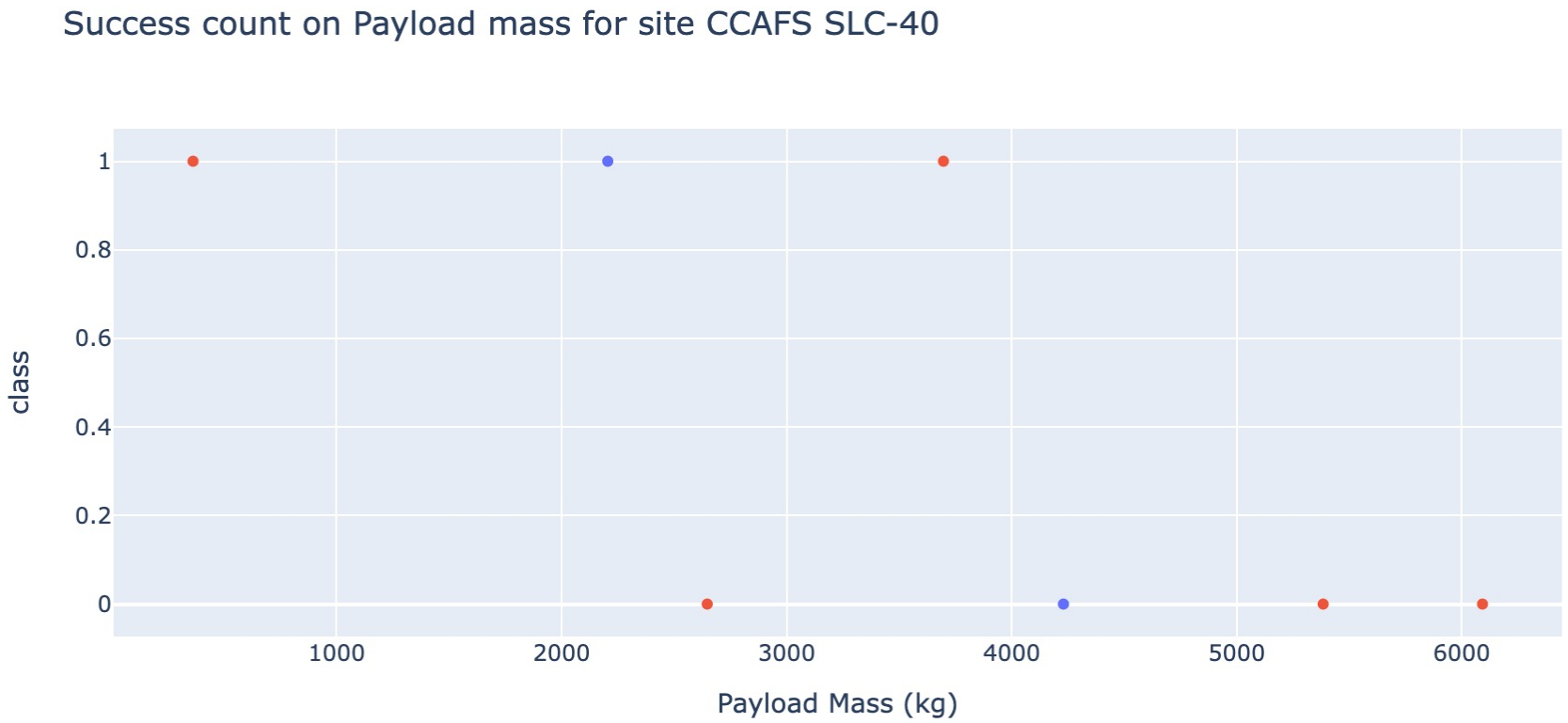


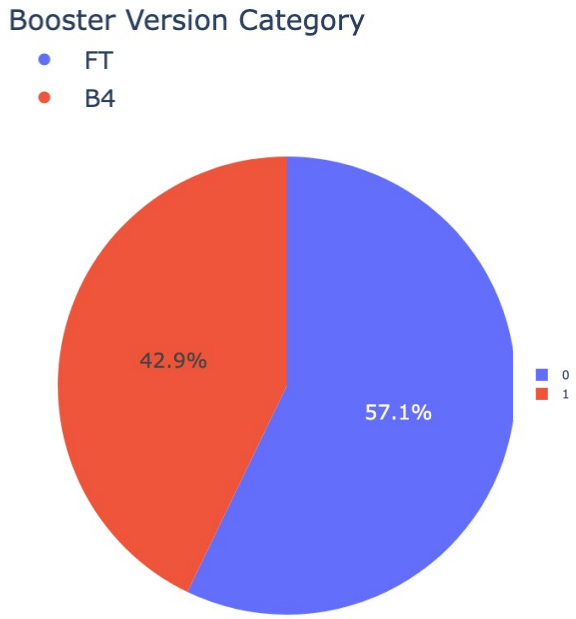
RESULTS: SpaceX Launch Record Dashboard



RESULTS: SpaceX Launch Record Dashboard

Payload range (Kg):

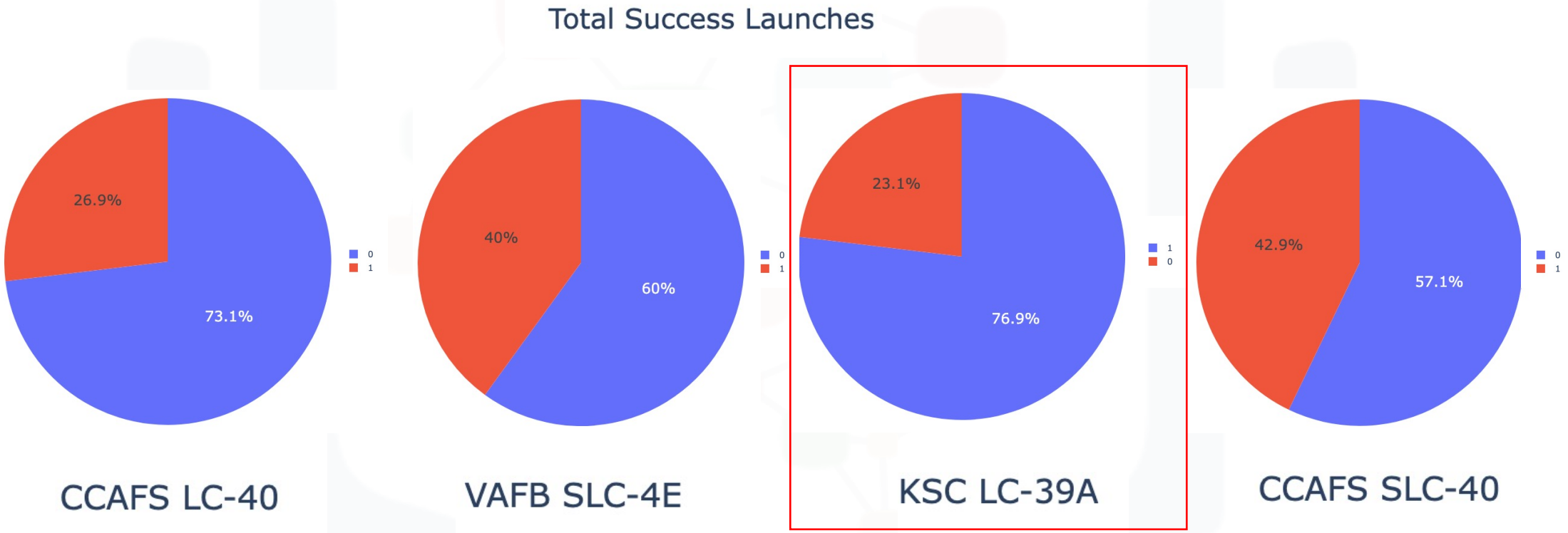




IBM Developer

SKILLS NETWORK

RESULTS: SpaceX Launch Record Dashboard



The highest success rate

RESULTS: Predictive Analysis

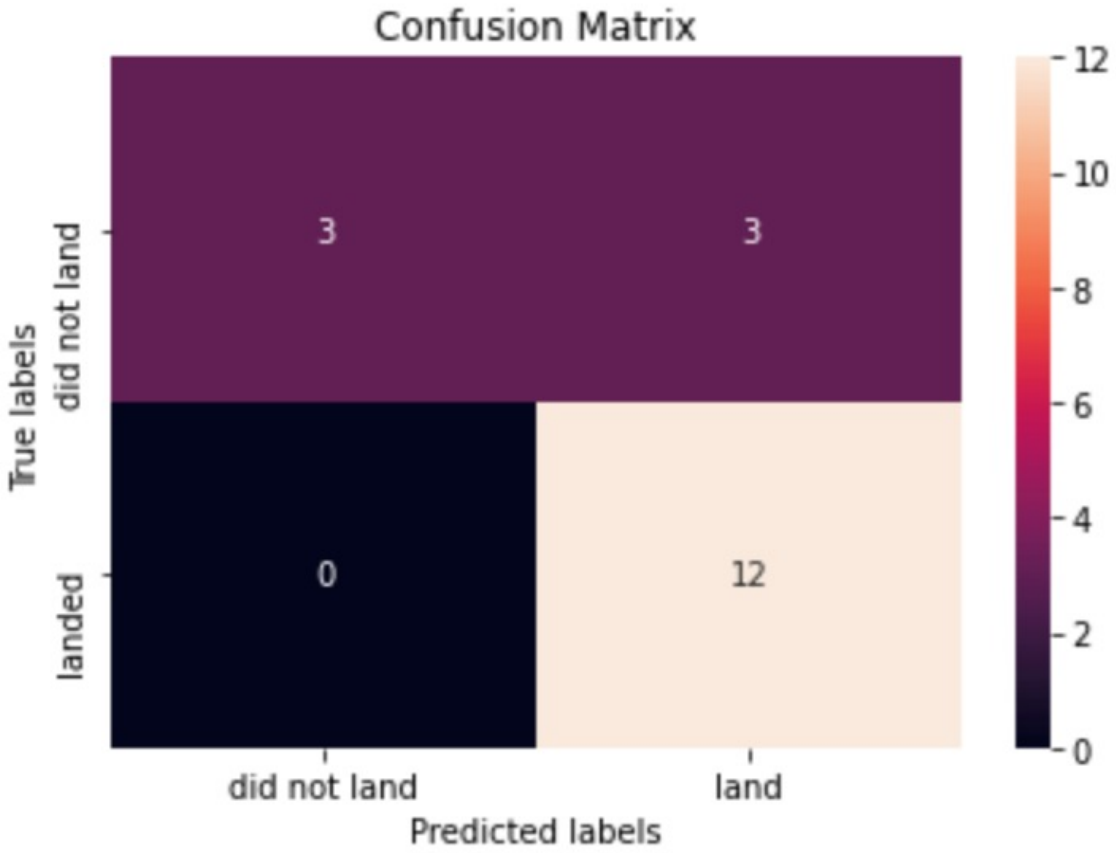
- Logistic Regression method

The best parameters using the data:

{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs’}

The accuracy on the validation data: 0.847

The accuracy of the test data: 0.833



RESULTS: Predictive Analysis

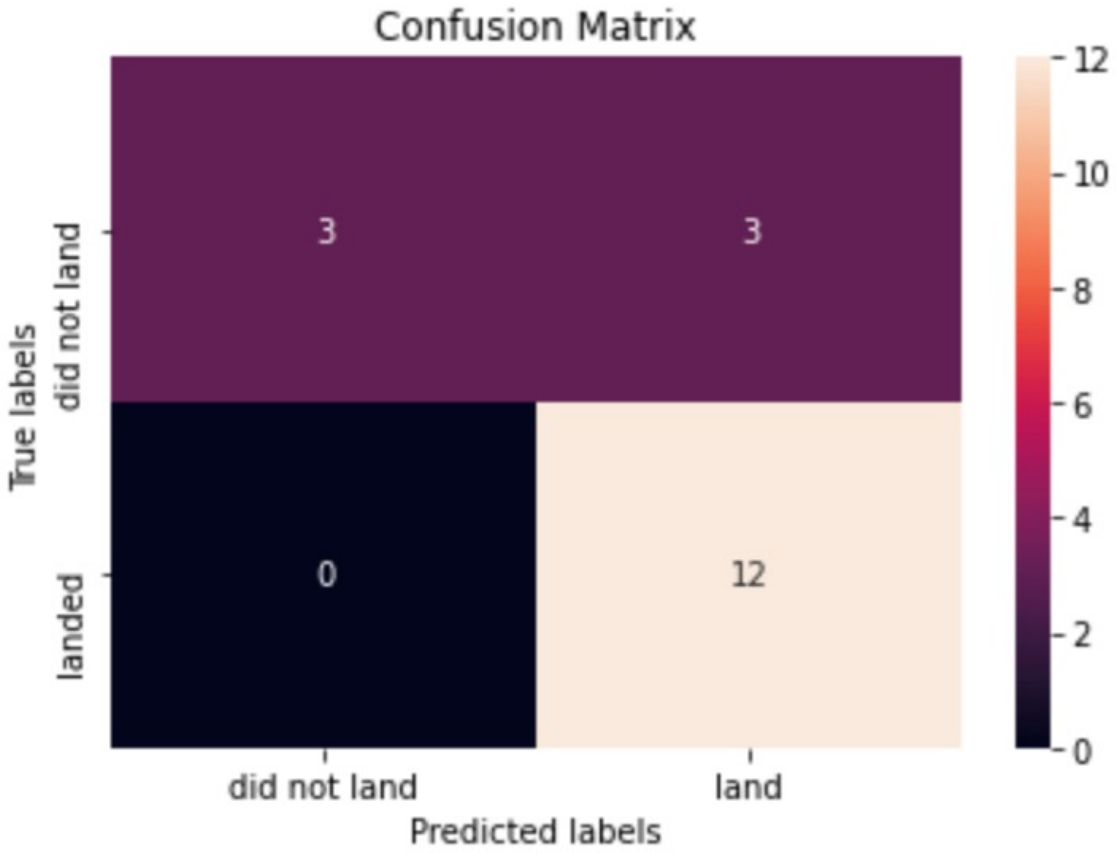
- Support Vector Machine method

The best parameters using the data:

{'C': 1.0, 'gamma': 0.0316227660168379, 'kernel': 'sigmoid'}

The accuracy on the validation data: 0.847

The accuracy of the test data: 0.833



RESULTS: Predictive Analysis

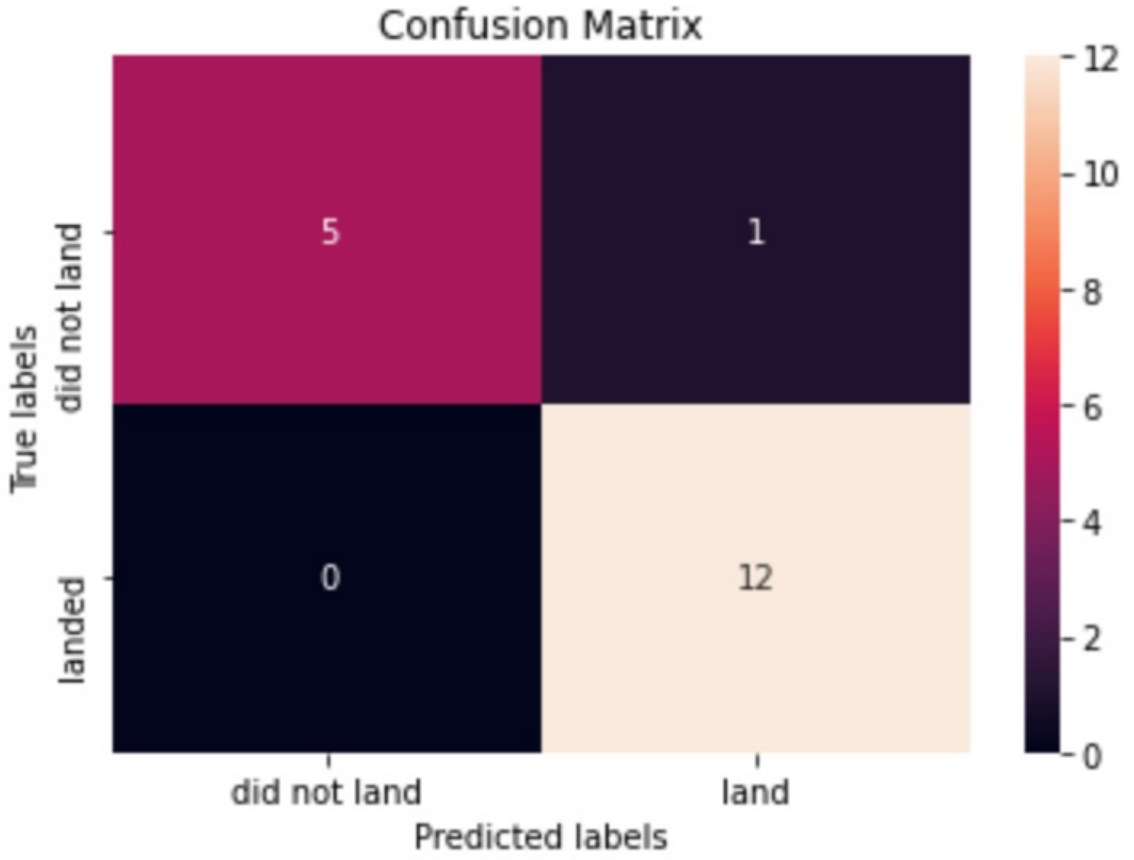
- Decision Tree method (performs the best)

The best parameters using the data:

{'criterion': 'gini', 'max_depth': 4,
'max_features': 'sqrt', 'min_samples_leaf': 4,
'min_samples_split': 10, 'splitter': 'best’}

The accuracy on the validation data: 0.875

The accuracy of the test data: 0.944



RESULTS: Predictive Analysis

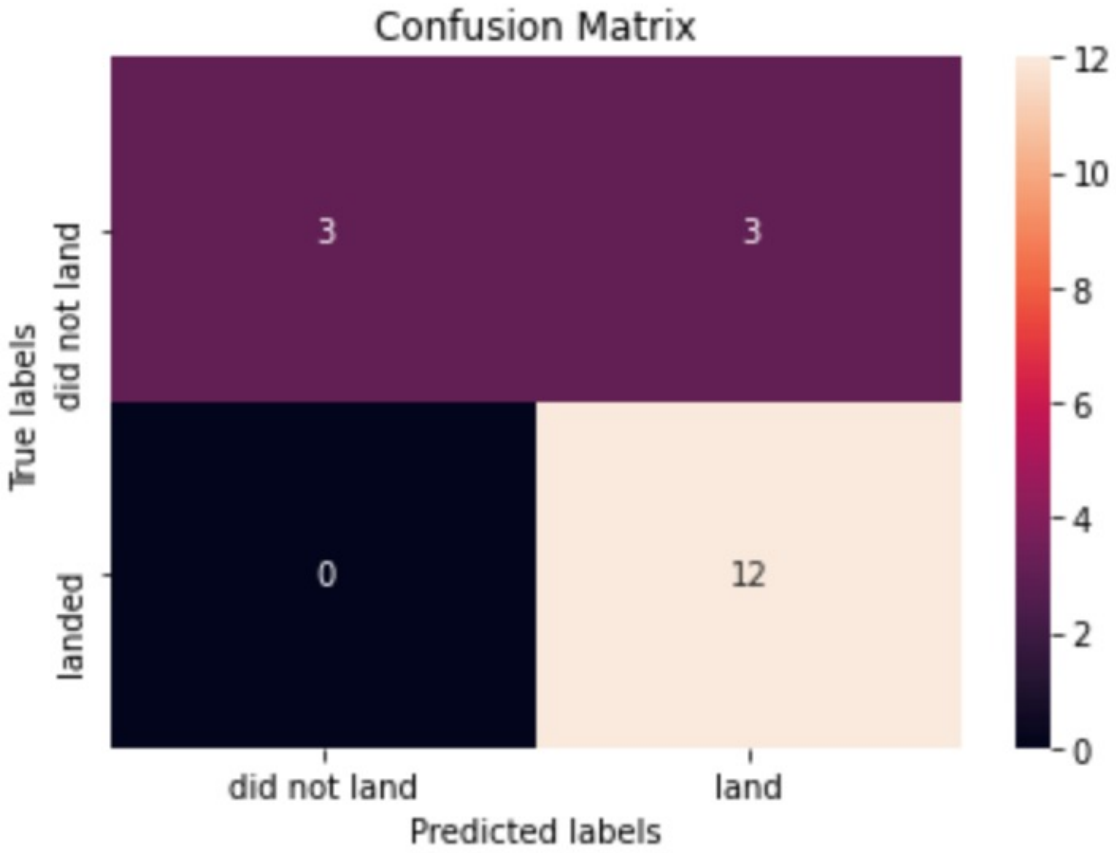
- K Nearest Neighbor method

The best parameters using the data:

{'algorithm': 'auto', 'n_neighbors': 9, 'p': 1}

The accuracy on the validation data: 0.847

The accuracy of the test data: 0.833



CONCLUSION

- Successful outcome increases with the increasing number of flight.
- The more massive the payload, the less likely the first stage will return.
- There are no rockets launched for heavy payload mass in VAFB-SLC launch site.
- In the LEO orbit the success related to the number of flights, but in GTO orbit there are no relationship between flight number.
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS orbits.
- Success rate since 2013 kept increasing till 2020.
- KSC LC-39A launch site has the highest successful outcome rate.
- Decision Tree method works best to predict successful outcomes.