SpaceX Falcon 9 First Stage Landing Prediction

Cindy Zhang

9/6/2021

OUTLINE

- Executive Summary
- Introduction
- Methodology
  - Data Collection & Wrangling Dashboard
  - EDA & Interactive Visual Analytics
  - Predictive Analysis
- Results
  - EDA with Visualization Dashboard
  - EDA with SQL
  - Interactive Map with Folium
  - Predictive Analysis
- Conclusion

EXECUTIVE SUMMARY

The Problem

We were going to predict if the Falcon 9 first stage will land successfully

Models Used

Logistic regression, support vector machine, classification trees.

The Solution

We achieved an accuracy of 87.3% on the test set.

INTRODUCTION

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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

We used data from SpaceX website and Wikipedia. First, removed some irrelevant information from the data sets and set the data to the format we desired. Next, we used Matplotlib and Plotly to visualized the data so we were able to keep a few features that was related to the success of the launch, such as payload mass, launch sites. Lastly, we set up various classification models working on a training set to find out the best hyper parameters. With the best model among these, we achieved an accuracy 83.3% in determine if the first state of Falcon 9 will land, based on the test set.

METHODOLOGY- Data Collection & Wrangling

Data Collection and Wrangling Methodology

- We sent requests to the SpaceX API and built a table for Falcon 9 launches. Independent variables include launch date, payload mass, launch site, orbit and so on.
- We also scrapped data from Wikipedia page “List of Falcon 9 and Falcon Heavy launches” and exported to csv files.

METHODOLOGY- EDA & Interactive Visual Analytics

Used Matplotlib and Seaborn to visualize the pattern among different independent variables and between independent variables and outcome

Used SQL to filter some results, such as the count of landing outcomes

Used Folium to mark launch sites with success or failure on the map

Used Dash to make an interactive page

METHODOLOGY- Predictive Analysis

Predictive Analysis Methodology

Our data set contains more than 80 features of Falcon 9 launch. We split the data 80% for training and 20% for testing.

The classifiers we used are KNN classifier, decision tree classifier, logistic regression and support vector machine.

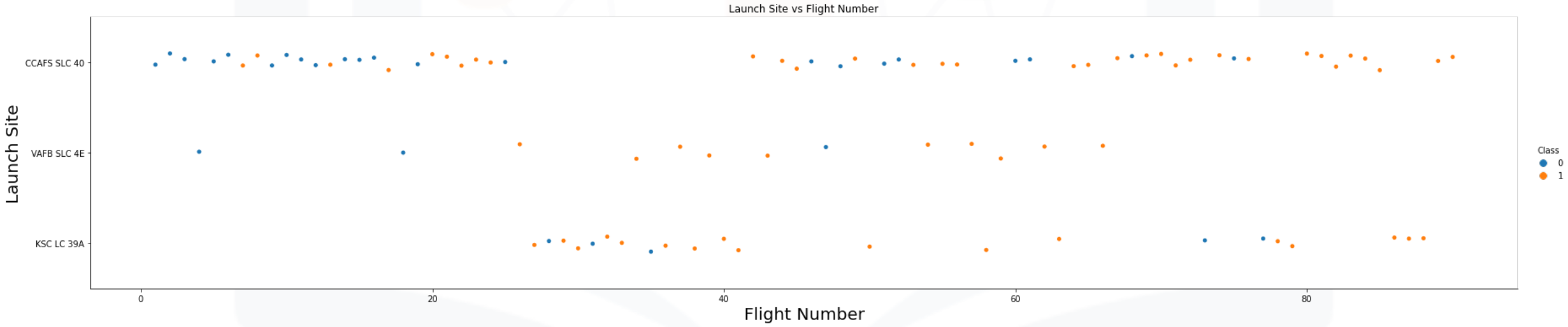
We defined a hyperparameter set for each classifier and used Grid Search with 10-fold cross-validation to find the best hyperparameter set within each classifier.

After setting the hyperparameters with as best ones found in the previous step, we fit the models with the training data and computed the accuracy scores and confusion matrices for these models with testing data. Then we visualized the confusion matrices for easier understanding.

RESULTS-EDA with Visualization

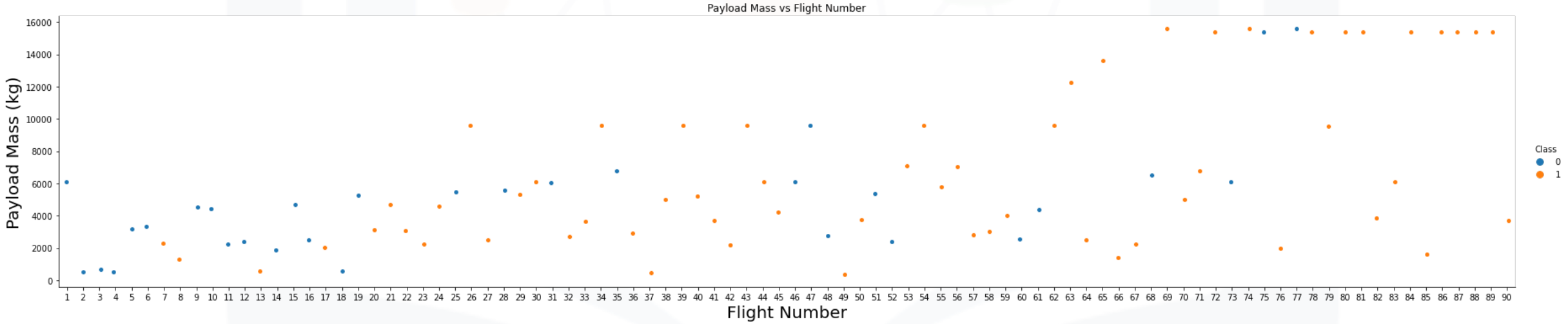
For independent variables, we tried to find if there was any relation between Flight Number, Launch Site, Payload Mass and Orbits. The yellow dots (class 1) indicate successful landings.

We notice that some sites launch more rockets than the others.



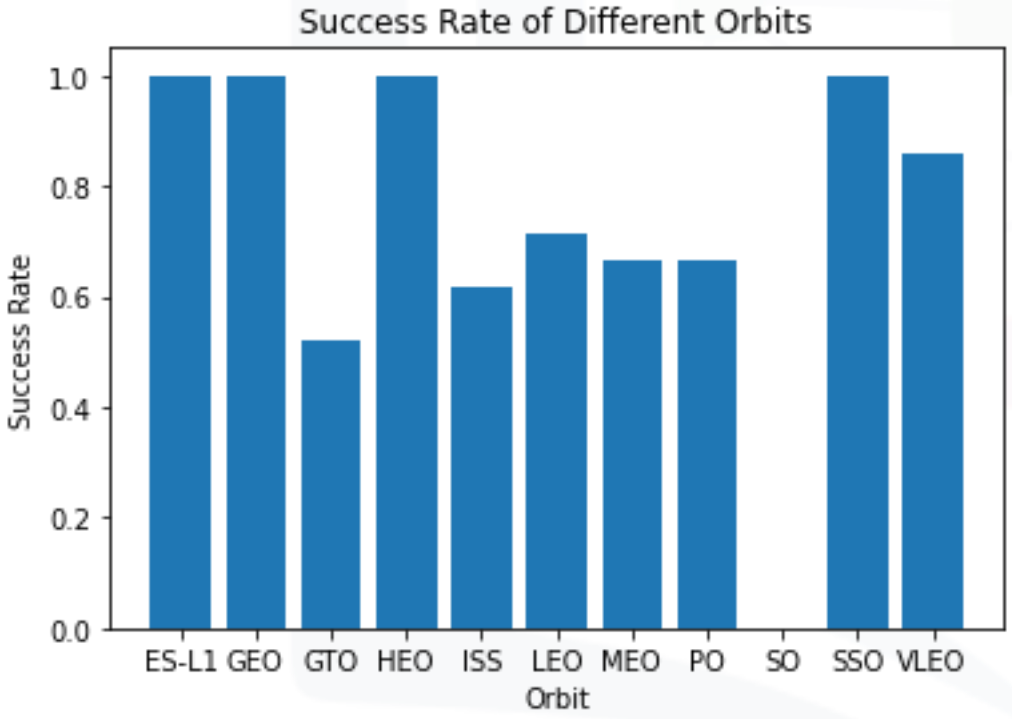
RESULTS-EDA with Visualization

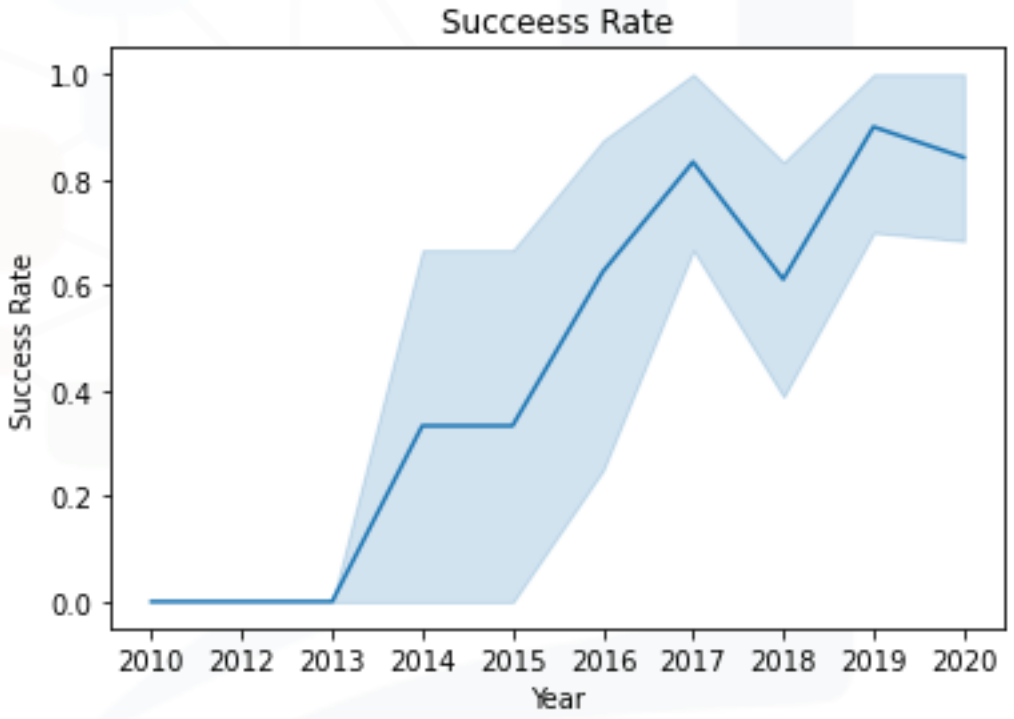
We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.



RESULTS-EDA with Visualization

We see that success rates vary from site to site, but success rate has an overall ascending trend.





RESULTS-EDA with SQL

Here are the result of some queries in SQL. The data we used has the following features:

Date, time, booster version, launch site, payload, payload mass (kg), orbit, customer, mission outcome, landing outcome.

Display the names of the unique launch sites in the space mission

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

RESULTS-EDA with SQL

Display 5 records where launch sites begin with the string 'CCA'

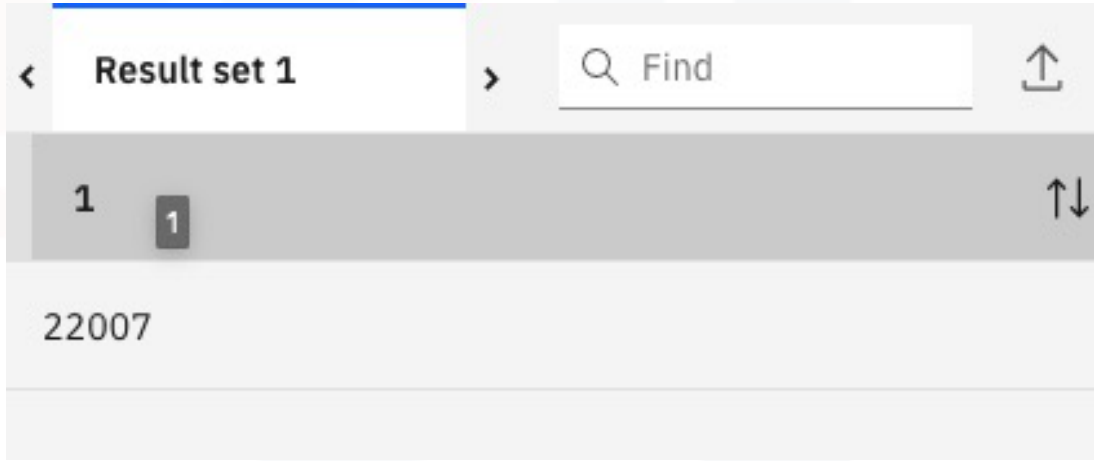
LAUNCH_SITE
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

RESULTS-EDA with SQL

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

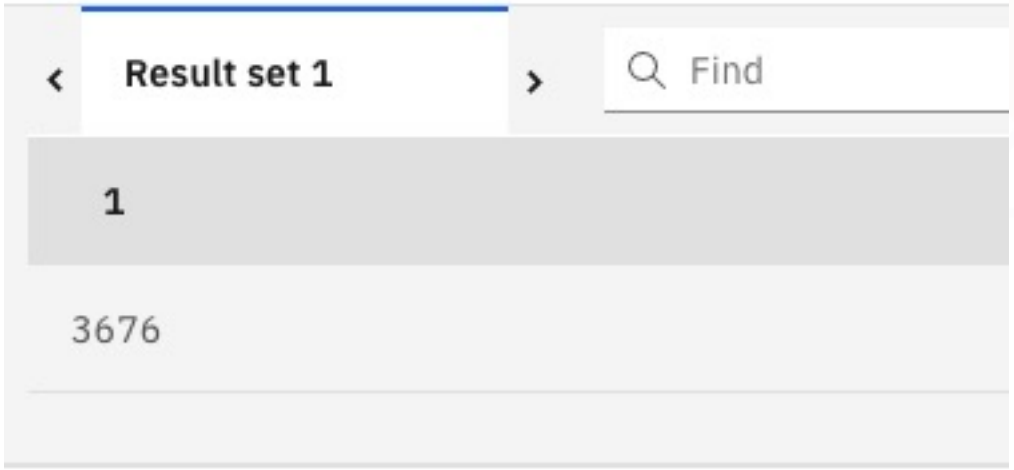
LANDING__OUTCOME	BOOSTER_VERSION
Failure (drone ship)	F9 v1.1 B1012

Display the total payload mass carried by boosters launched by NASA (CRS)

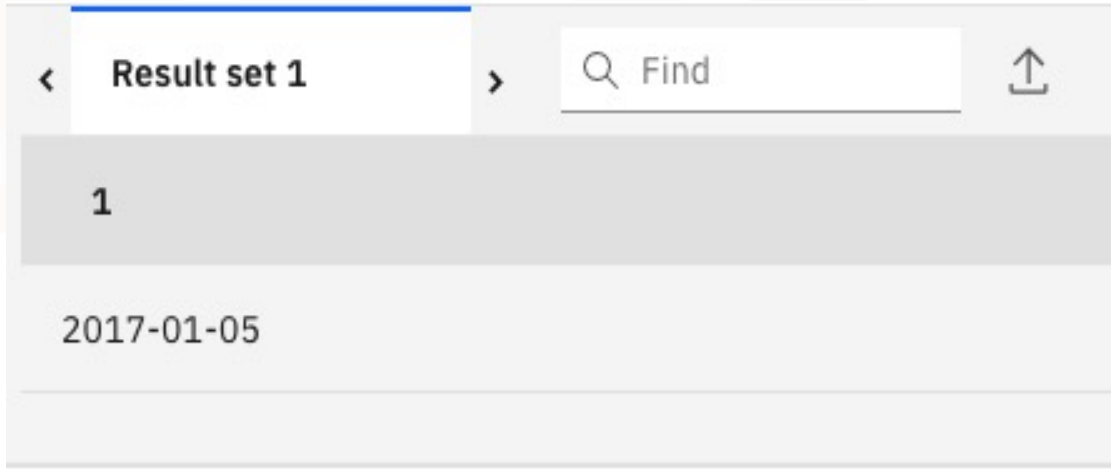


RESULTS-EDA with SQL

Display average payload mass carried by booster version F9 v1.1

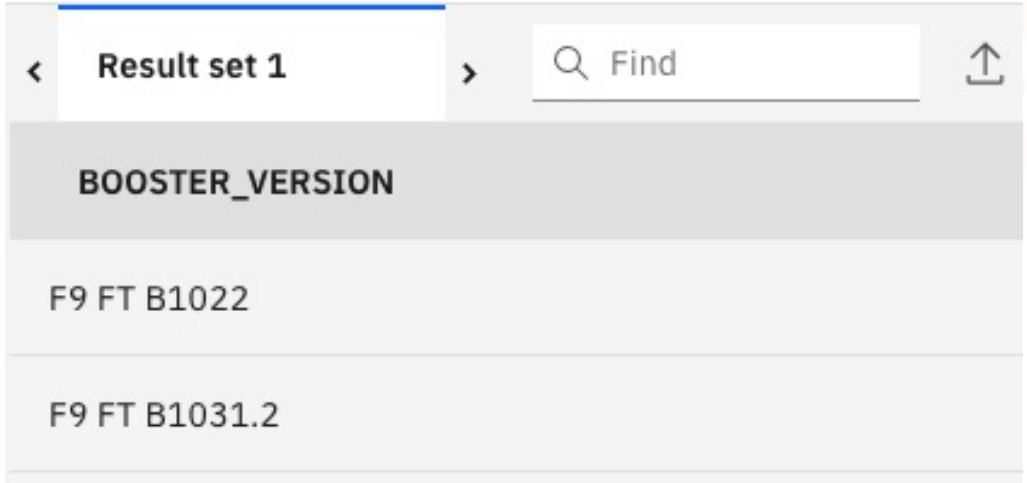


List the date when the first successful landing outcome in ground pad was achieved.

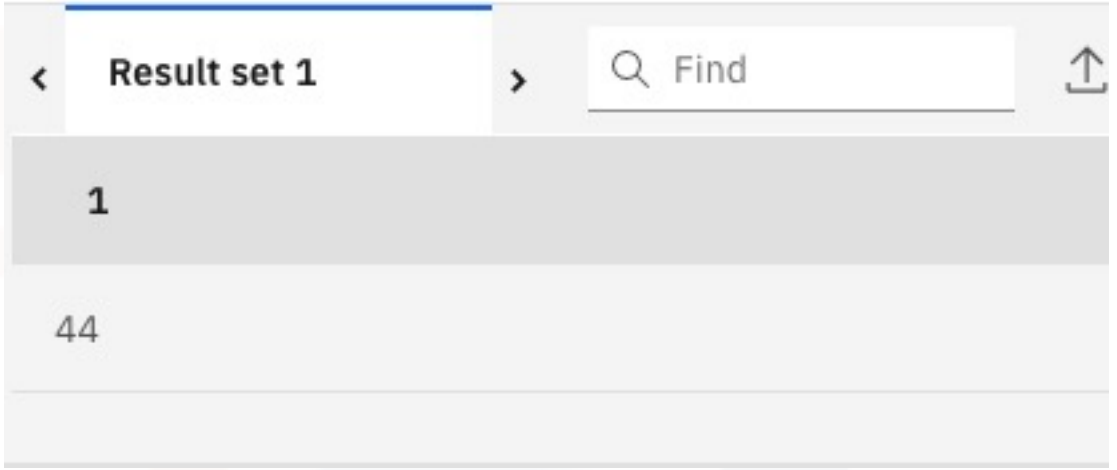


RESULTS-EDA with SQL

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000



List the total number of successful and failure mission outcomes



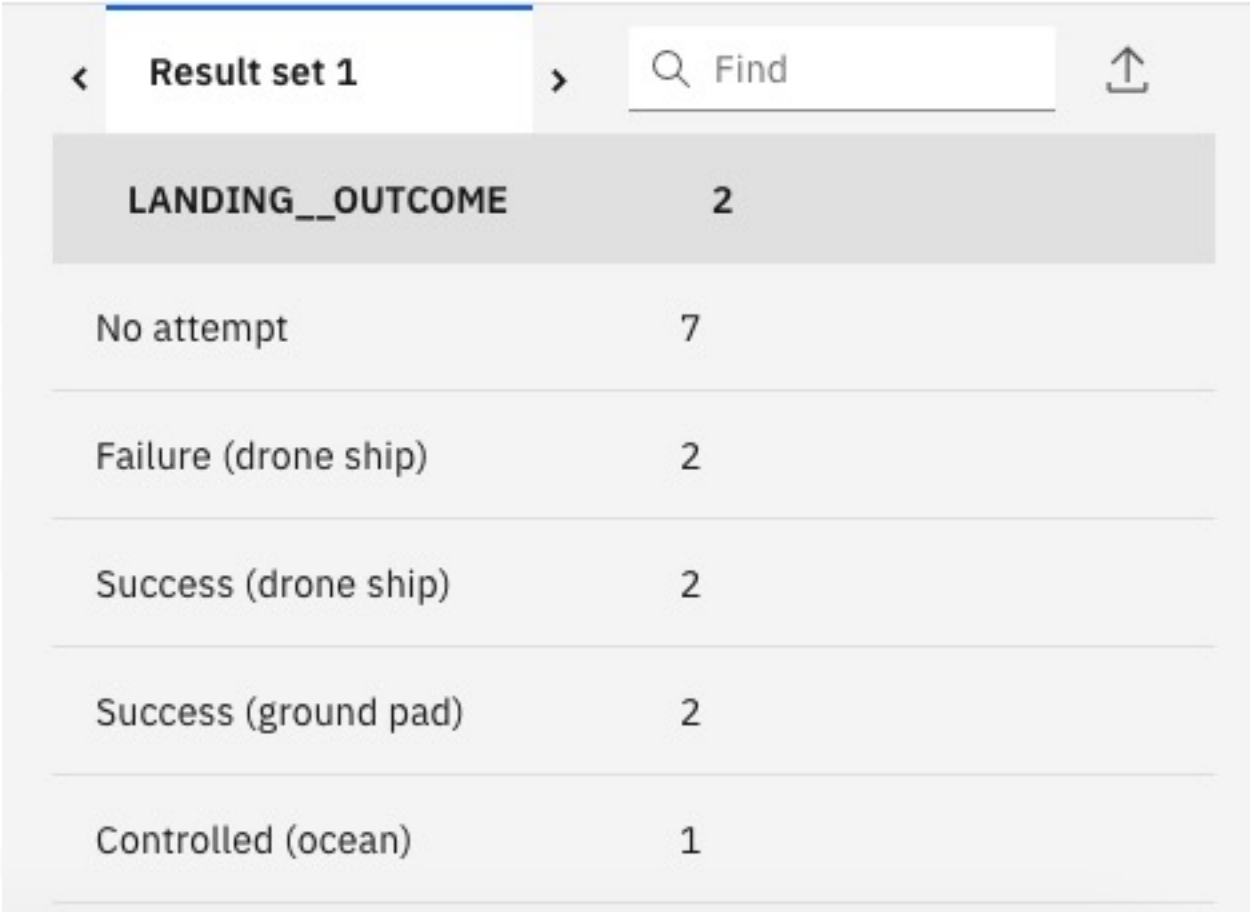
RESULTS-EDA with SQL

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

BOOSTER_VERSION
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

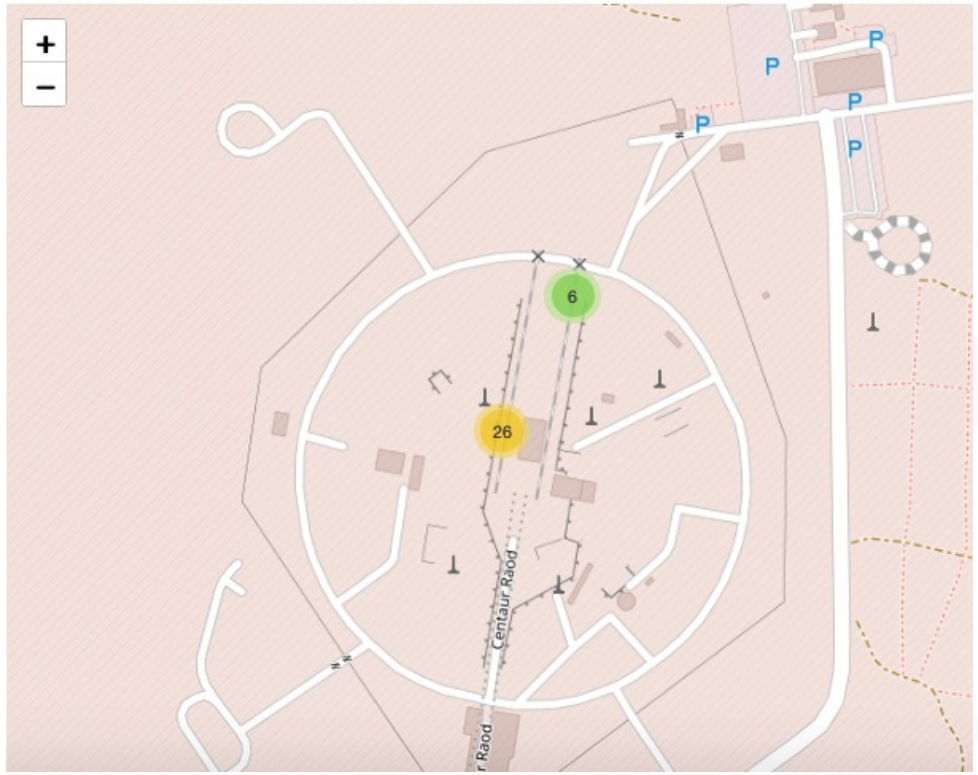
RESULTS-EDA with SQL

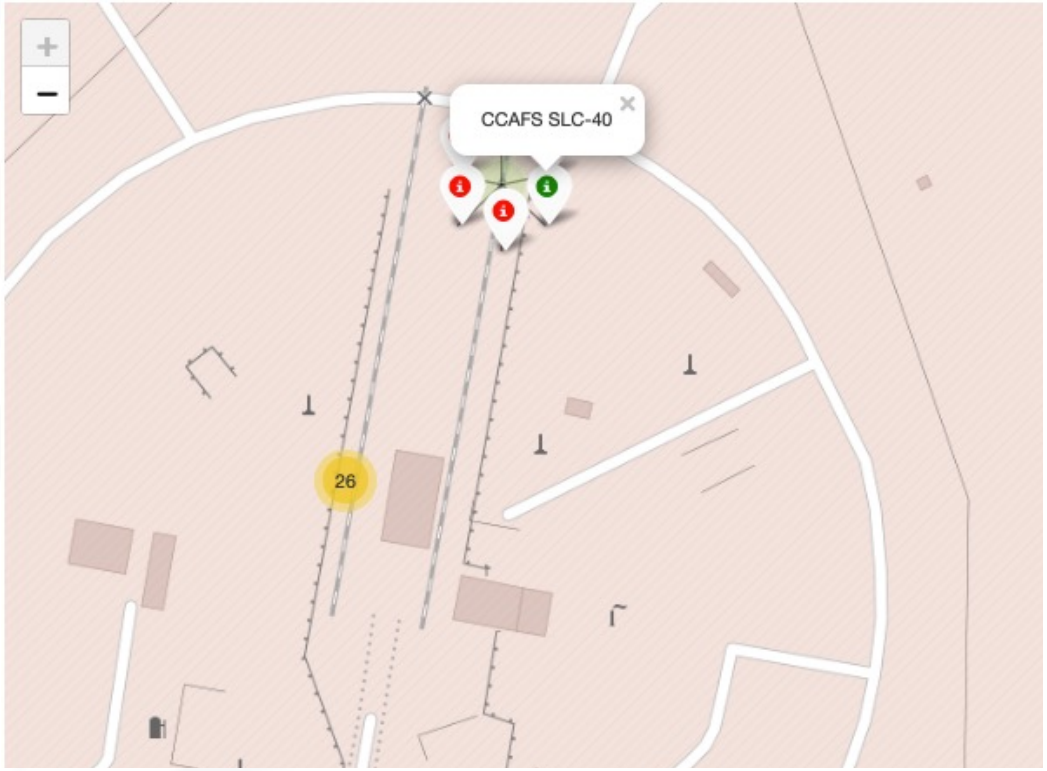
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order



RESULTS-Interactive Map with Folium

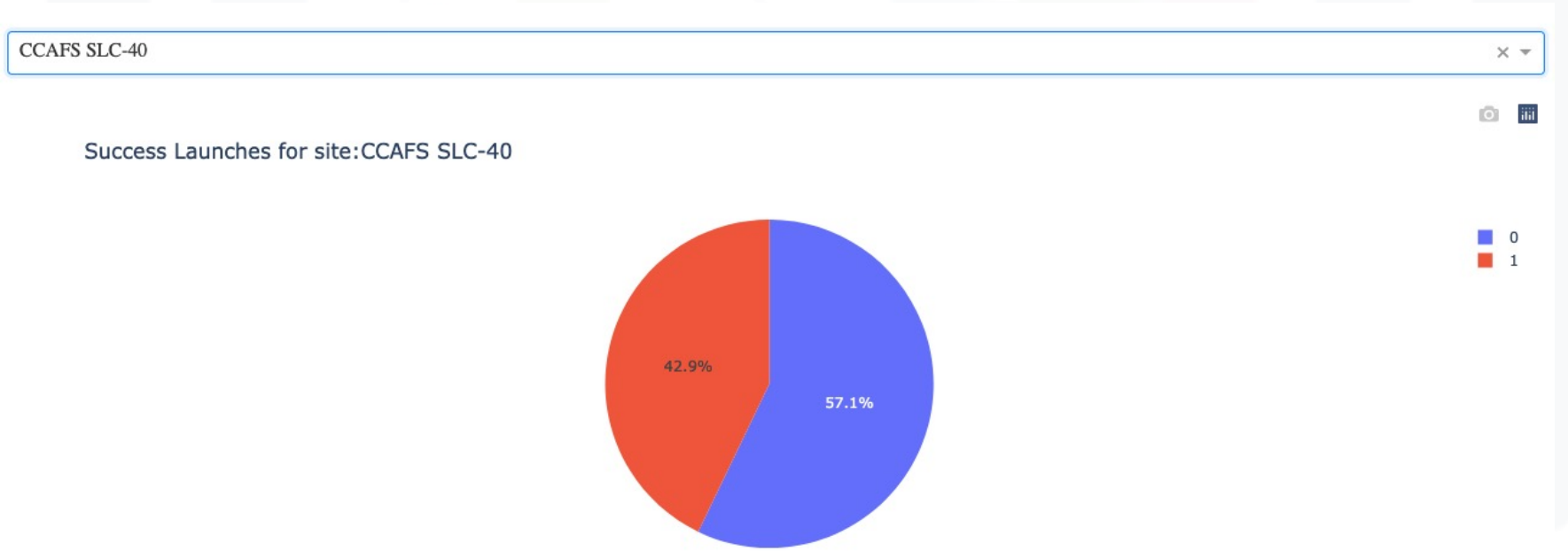
All launch records on marked on the map. The green marker indicates successful launch, the red marker indicates unsuccessful launch. The second picture is what we see after zooming in.



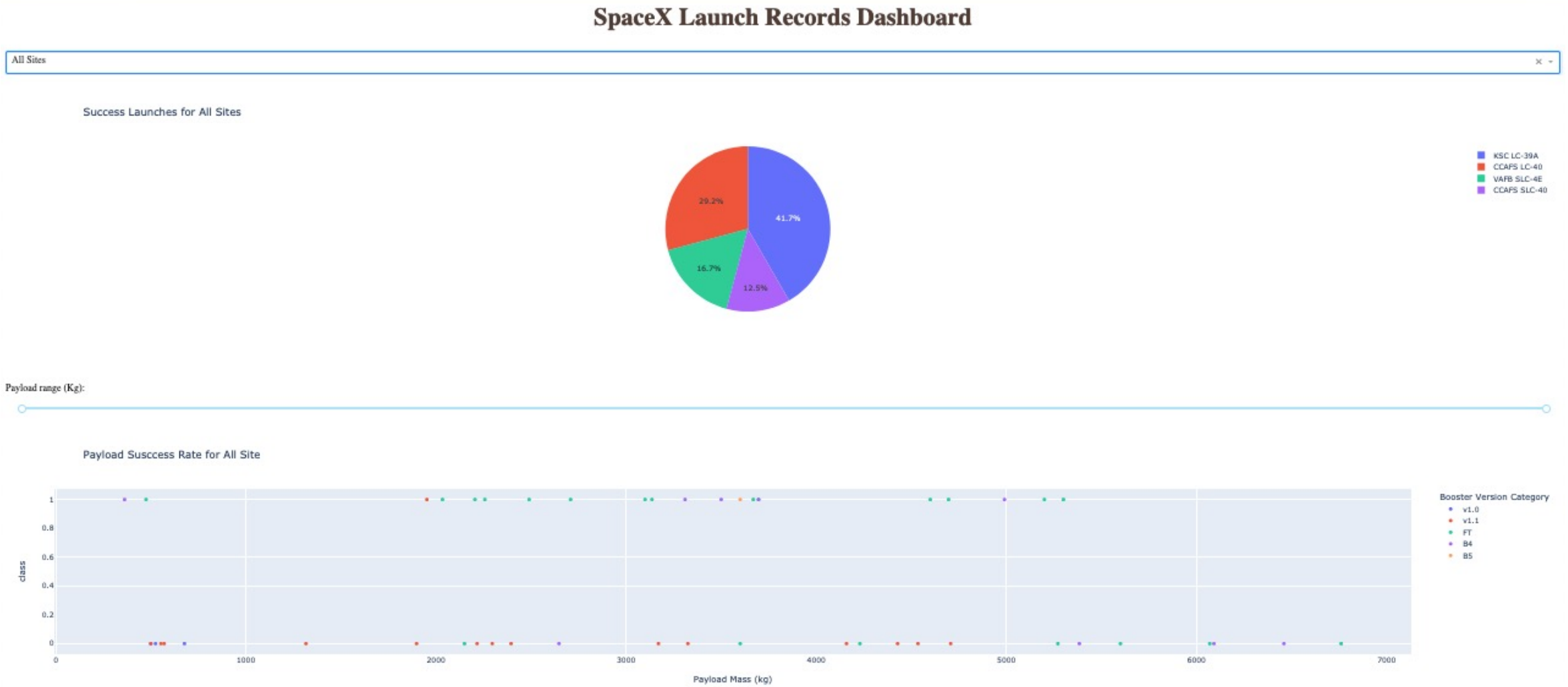


RESULTS-Plotly Dash Dashboard

We built a Dash application using Python Plotly library. The dropdown menu allows us to select all sites or a specific launch site. When “All Sites” is selected, the pie graph show the contribution to the successful launches from each l sites. When a specific launch site is selected, it shows the launch outcome percentage for that site. The scatter plot with a range slide shows the correlation between payload mass and success for all sites or one specific site, labeled by booster versions.



RESULTS-Plotly Dash Dashboard



RESULTS-Predictive Analysis

Accuracy for Decision Tree Classifier: 72.2%

Correctly classified 4 successful landings out of 12, 5 unsuccessful landings out of 6 in the test set

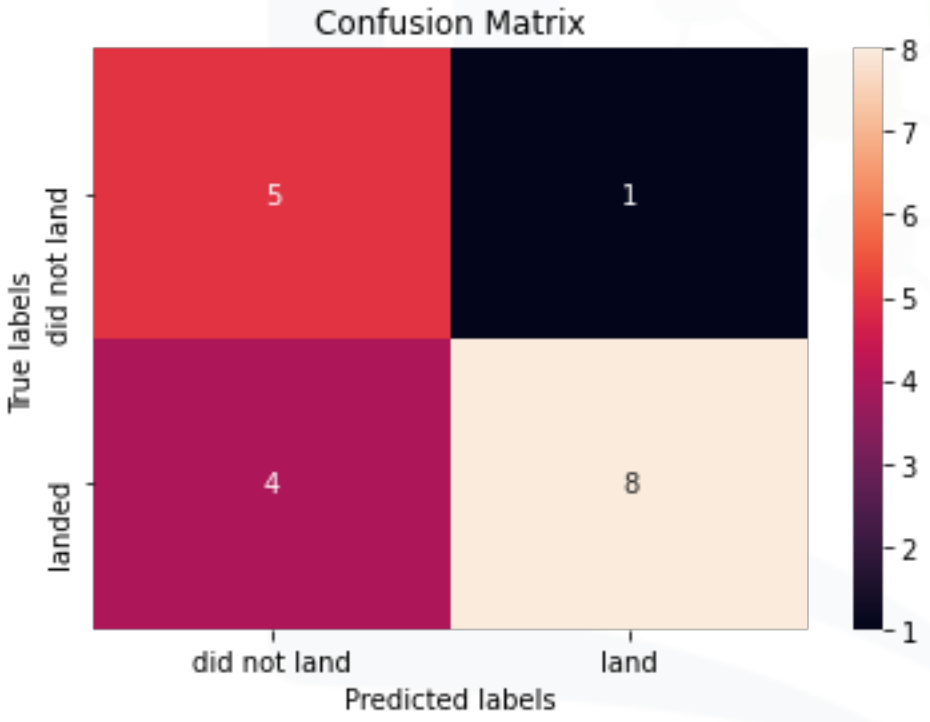
Accuracy for KNN Classifier, SVM, Logistic Regression: 83.3% (Same)

Correctly classified all 12 successful landings in the test set, 3 unsuccessful landings out of 6

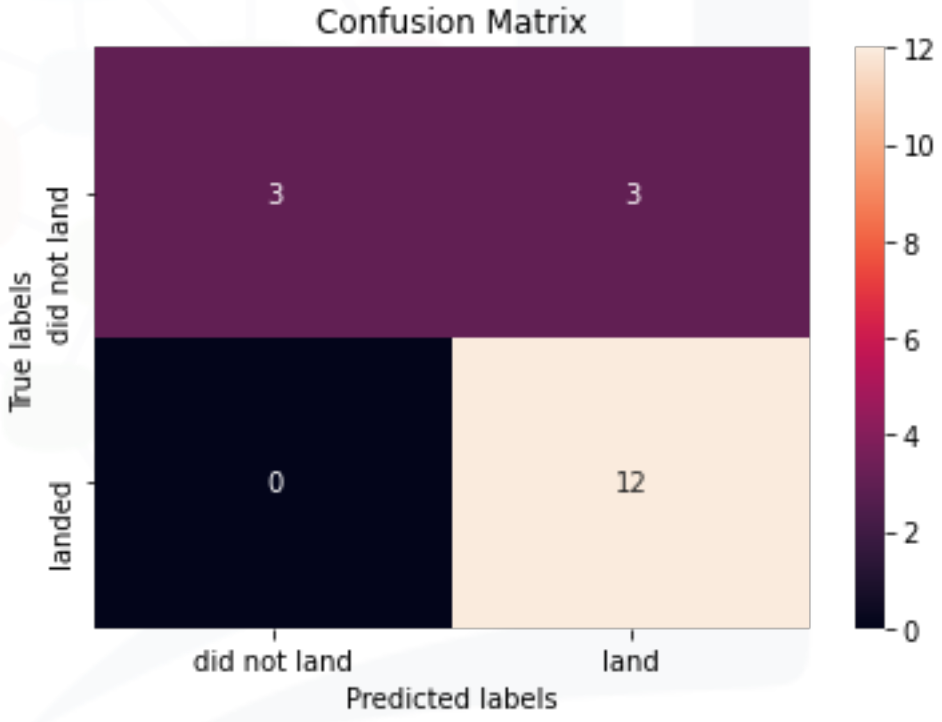
The decision tree classifier has higher accuracy (83.3%) in classifying the unsuccessful landings while the other classifiers have higher accuracy (100%) in classifying successful landings.

RESULTS-Predictive Analysis

Visualized Confusion Matrix for Decision Tree Classifier



Visualized Confusion Matrix for KNN Classifier, SVM, Logistic Regression



CONCLUSION

- Success rate of launching varies from site locations, payload mass, orbits, etc.
- Success rate has an overall increasing trend as researches in spacecrafts develop fast in recent years.
- Multiple features determine if a launch will be successful.
- It’s likely to predict whether if we can get the data beforehand.