Winning Space Race with Data Science

<Seliana Kaguamba>

Oct 2022

Executive Summary

- The dataset was constructed from a SpaceX API and Wikipedia page on Falcon 9. Key variables to get were BoosterVersion, Launchsite, Payload (missing values were replaced the the mean of the data set), other core data and all this was filtered for the Falcon 9 Booster version.
- Site KSC LC-39A has the highest successful launchs at 41.7%. Followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7% and lastly CCAFS SLC-40 at 12.5%.
- Payload range from 2,000-5,000kg has the highest successful launch rate.
- Payload range from 6,000-10,000kg has the lowest successful launch rate.
- F9 Booster version FT has the highest launch success rate.

Introduction

- Project background and context

SpaceX flies rockets into space much cheaper than any other private or public company in the world. They do this by reusing the first stage of their rockets. For the company to continue being successful or for competitors to to understand what they are up against, it is useful to predict the landing success rates of the first stage of their rockets.

- Problems you want to find answers

What factor or factors gives the best possible landing success rate?

Section 1

Methodology

Data Collection from SpaceX API

Import relevant libraries and define auxiliary functions

Define booster name, launchpad coordinates, payload, and other data from cores

Request API and decode to Json to then make it a Pandas dataframe

Use get calls to call data to construct dataset

BoosterVersion	Launchsite	Payload	coredata	Filter for Falcon 9

Data Wrangling

Replace missing payloadmass valued with mean payloadmass value

[API Data Collection - Github URL]

Data Collection - Scraping

1

Start

- Import required packages (beautifulsoup4 and requests) and define functions to help
- Request for the Falcon 9 wiki page from provided URL

2

Extract Column names

- Find the tables first to see that the 3rd table has relevant launch data
- Iterate through the <th> elements use provided helper function to extract column names.

3

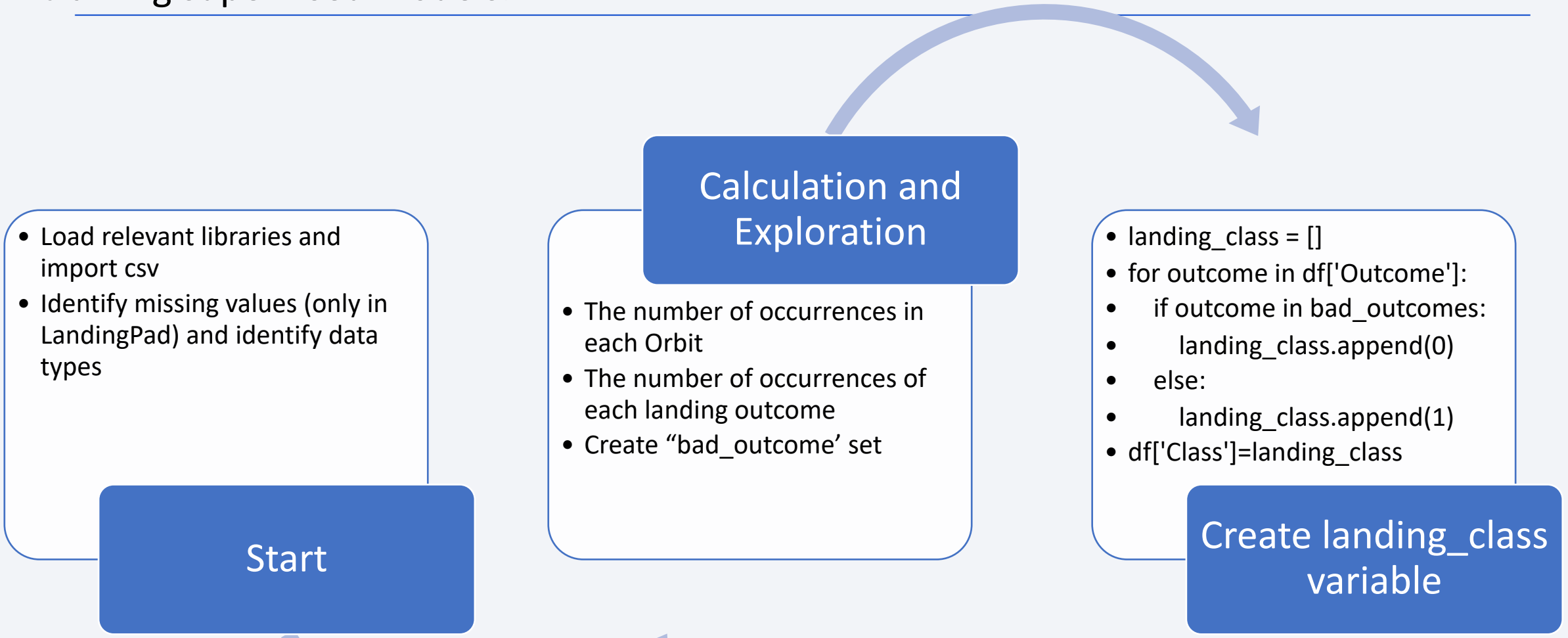
Create dataframe

- Create empty dictionary (launch_dict) with columns names previously extracted
- Parse the HTML tables to fill the dictionary with data
- Lastly create dataframe from dictionary with filled data

Github - Web Scraping URL

Data Wrangling

- The aim was to find some patterns in the data to determine what would be the label for training supervised models.



https://github.com/eatingfingers/coursera/blob/Coursera-Data-Science-Capstone/labs-jupyter-spacex-Data wrangling.ipynb

EDA with Data Visualization

- The plotted charts were
  - Scatterplot to visualize the relationship between flight number and launch site
  - Scatterplot to visualize the relationship between payload and launch site
  - Bar chart to visualize the success rate of launches to each Orbit
  - Scatterplot to visualize the relationship between Orbit and flight number
  - Scatterplot to visualize the relationship between payload and orbit
  - Line chart to to visualize launch success trend.
- GitHub URL
  - https://github.com/eatingfingers/coursera/blob/master/jupyter-labs-eda-dataviz.ipynb

Build an Interactive Map with Folium

- Launch sites were first marked. They are all along a coastline.
- Successful and failed launches were market in green or red markers respectively.
- Lastly distances to the nearest coastline, road, railway line and nearest city were calculated for the launchsite KSC LC-39A.
- GitHub URL https://github.com/eatingfingers/coursera/blob/Coursera-Data-Science-Capstone/lab_jupyter_launch_site_location%20(2).ipynb

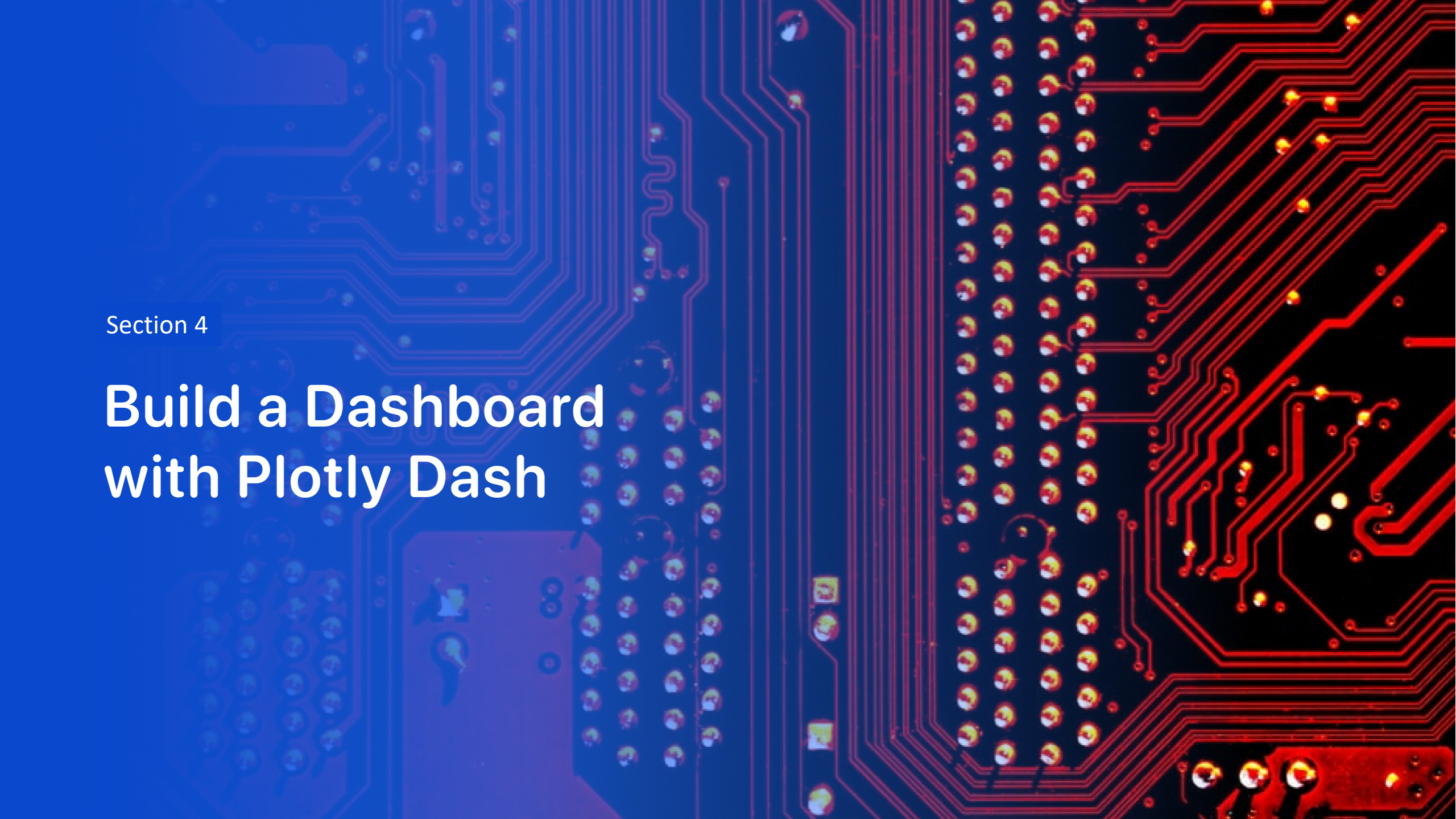
Build a Dashboard with Plotly Dash

- The dashboard has a Launch site drop-down component and a success pie chart to easily visualize which site has the largest successful launches and white site has the highest launch success rate.
- The dashboard also has ranger slider representing the payload range. Moving this generates a success payload scatter chart with coloured dots to easily identify each booster version category.
- GitHub URL - https://github.com/eatingfingers/coursera/blob/master/spacex_dash_app.py

https://selirocker-8050.theiadocker-1-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai/

Predictive Analysis (Classification)

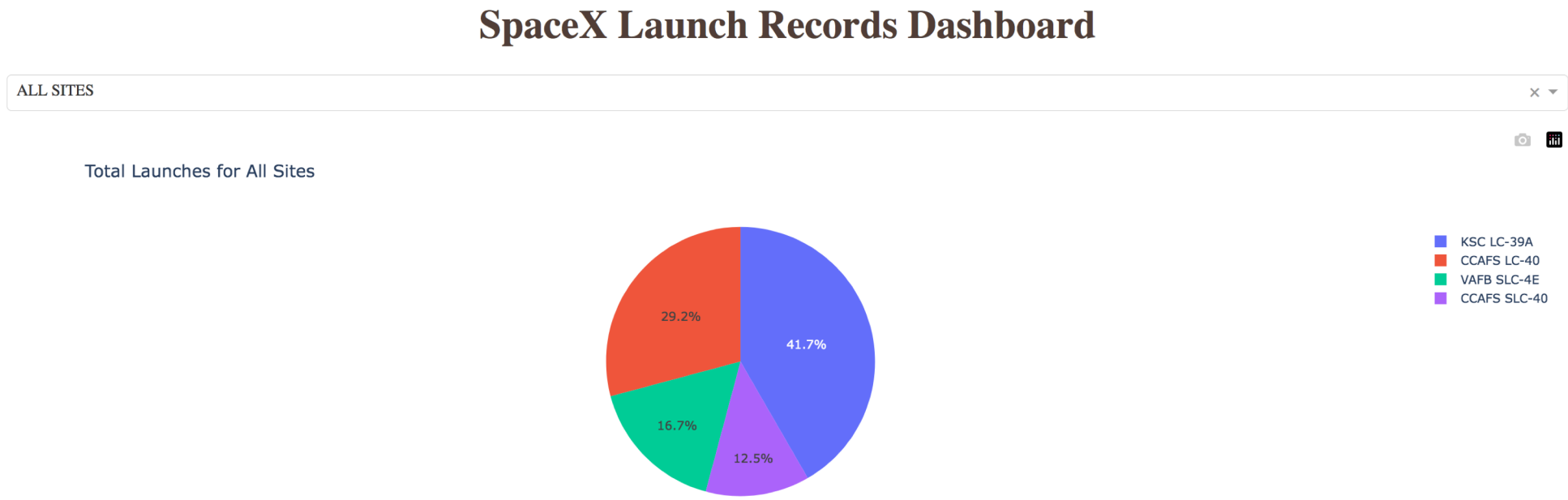
- 4 machine learning models were trained; KNeighboursClassifer (KNN), Decision Tree classifier, Logistics regression, and Support Vector Machine (SVM).
- After standardizing the dataset, the data was split into the training and testing datasets. 18 data points were in the test data.
- The models with the highest accuracy scores were the SVM and KNN. Their confusion matrix is in Section 5 of the report.
- Of the 18 data points of the test data, 4 were mal predicted. The rest were equally correctly predicted by the KNN and the SVM. i.e 10 were predicted to land and they did and 4 were predicted to fail to land and they did in fact fail. The 4 bad predictions were all false positives.
- GitHub URL https://github.com/eatingfingers/coursera/blob/master/SpaceX_Machine%20Learning%20Prediction_Part_5%20(2).ipynb


Section 4
Build a Dashboard with Plotly Dash

Successful Launches

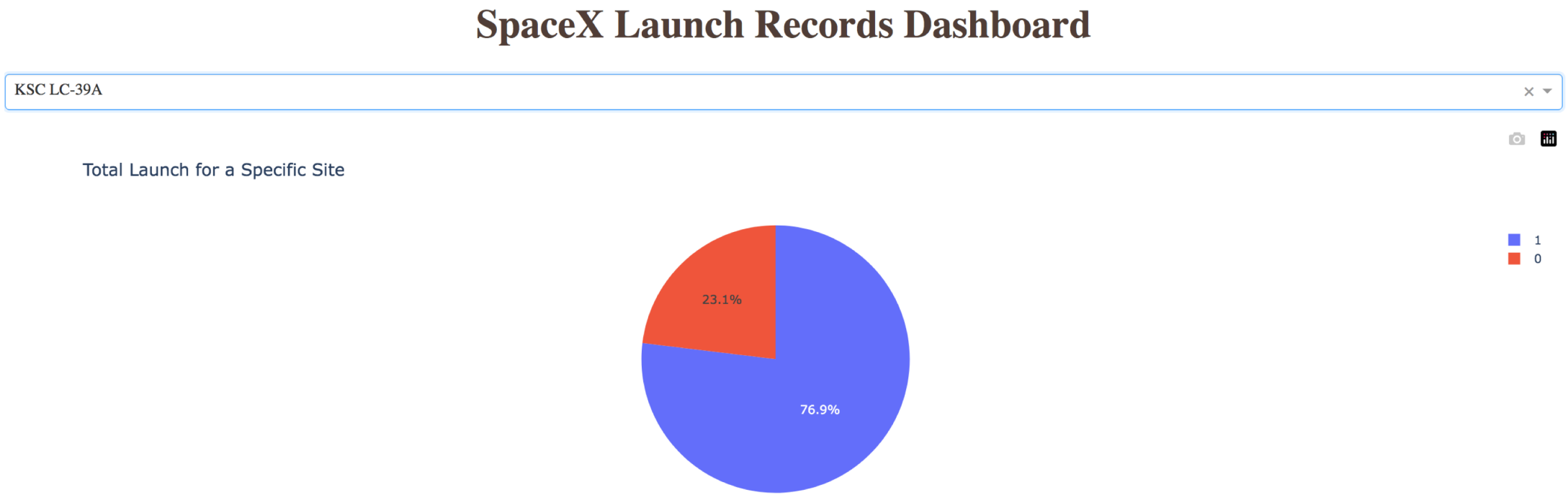
- Site KSC LC-39A has the highest successful launchs at 41.7%. Followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7% and lastly CCAFS SLC-40 at 12.5%.

- All this is better visualized in the pie chart



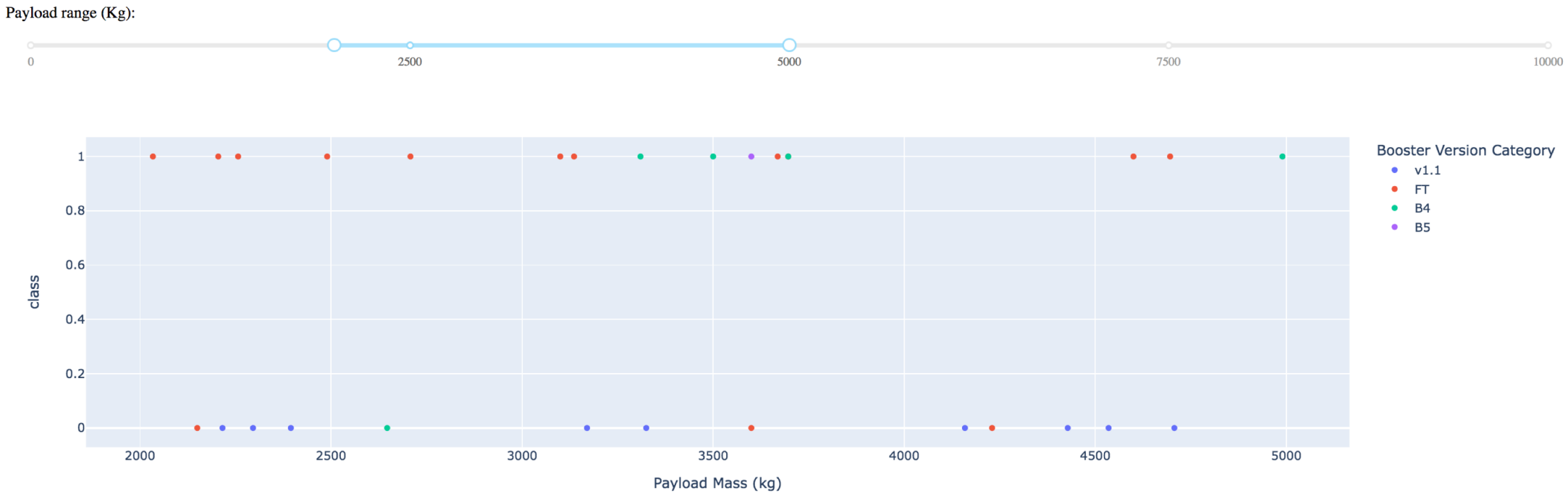
KSC LC-39A success ratio

- This site has the highest launch success ratio at 76.9%



Most successful payload range and booster version

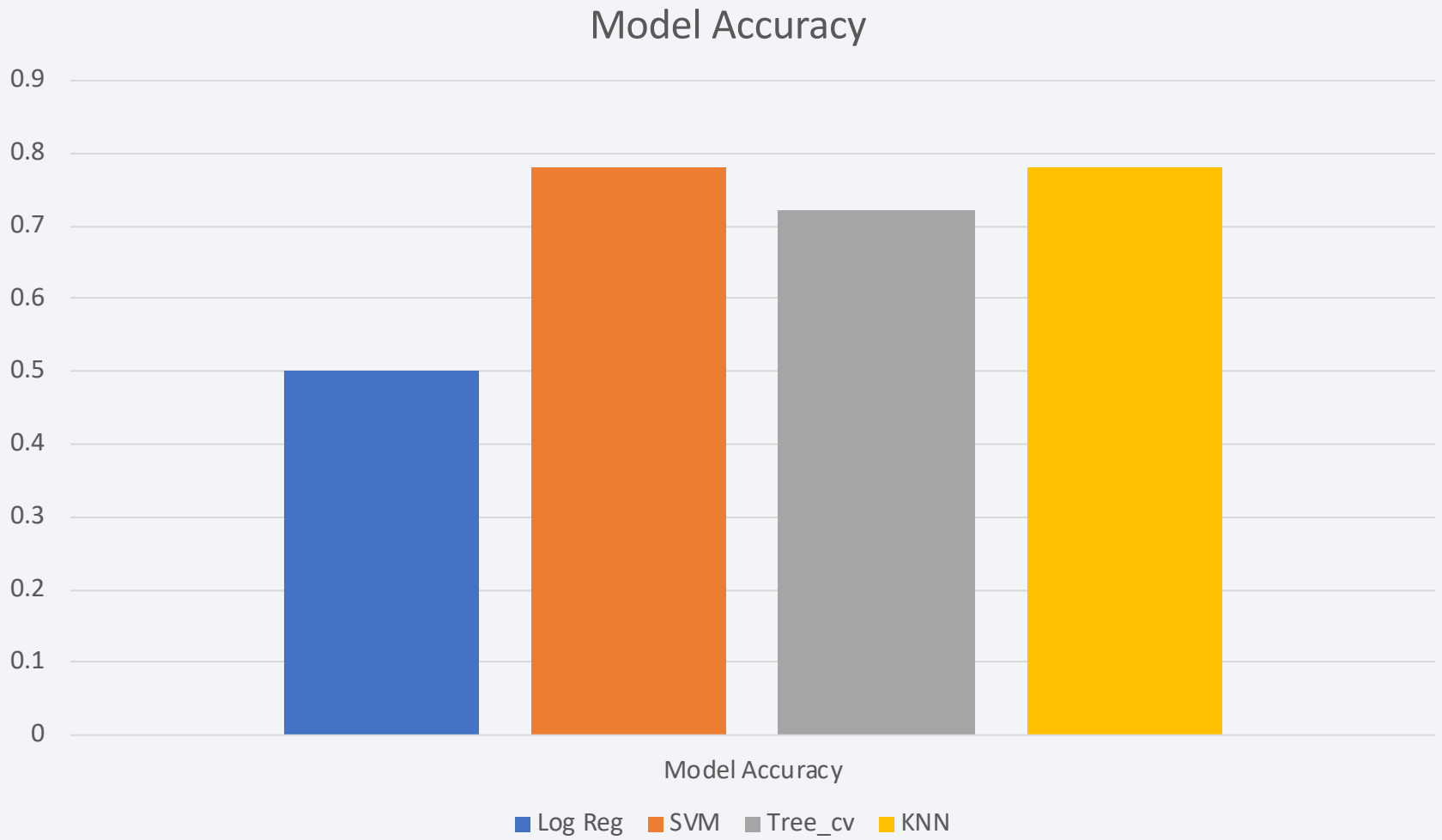
- Payload range from 2,000-5,000kg has the **highest** successful launch rate.
- Payload range from 6,000-10,000kg has the **lowest** successful launch rate.
- F9 Booster version **FT** has the highest launch success rate.



Section 5

Predictive Analysis (Classification)

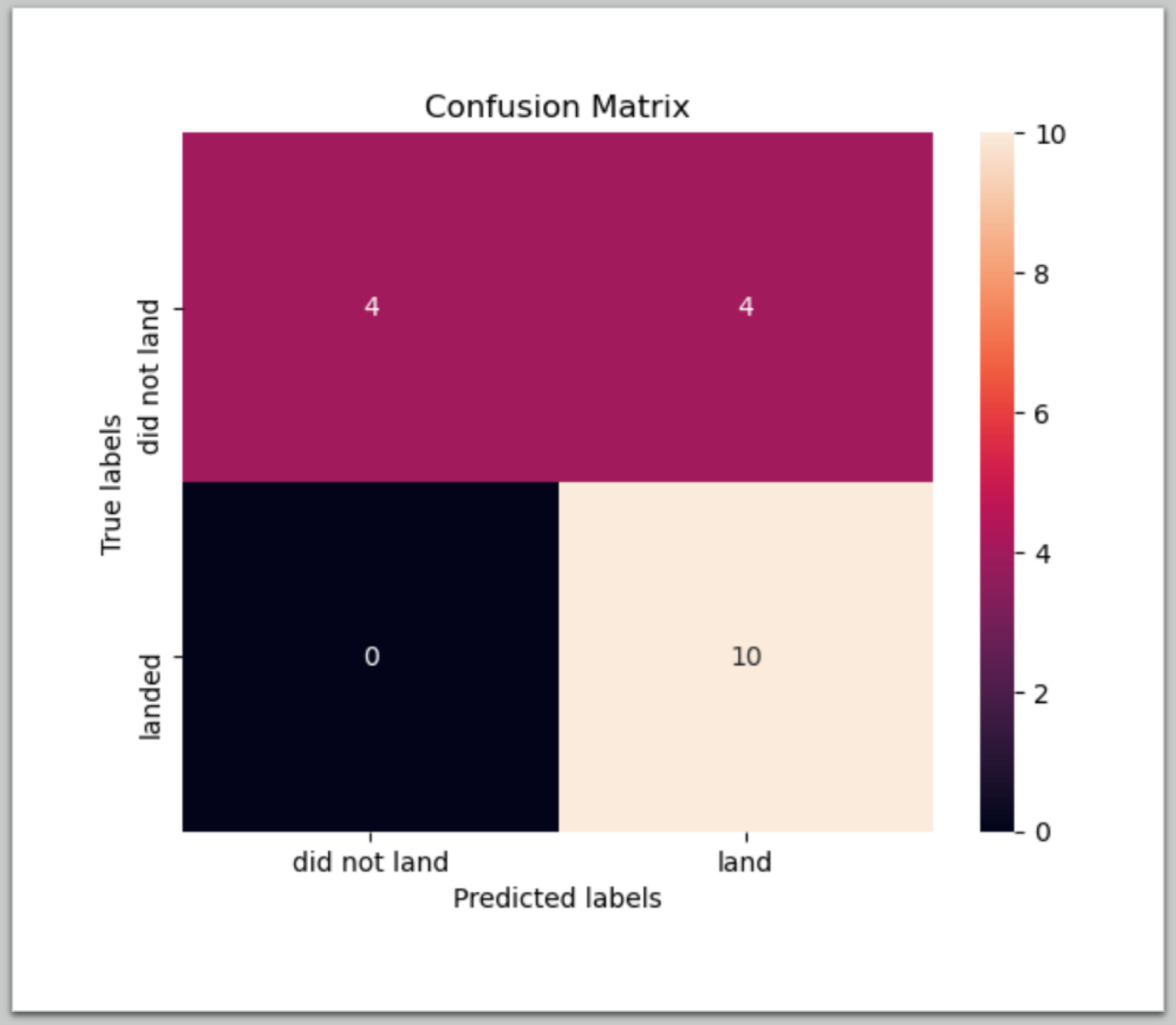
Classification Accuracy



The model Decision Tree Classifier has the highest classification accuracy.

Confusion Matrix of KNN and SVM

- Of the 18 data points of the test data, 4 were mal predicted. The rest were equally correctly predicted by the KNN and the SVM. i.e 10 were predicted to land and they did and 4 were predicted to fail to land and they did in fact fail. The 4 bad predictions were all false positives.



Conclusions

- Site KSC LC-39A has the highest successful launchs at 41.7%. Followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7% and lastly CCAFS SLC-40 at 12.5%.
- Payload range from 2,000-5,000kg has the highest successful launch rate.
- Payload range from 6,000-10,000kg has the lowest successful launch rate.
- F9 Booster version FT has the highest launch success rate.

Thank you!