Winning Space Race with Data Science

Maria Nataqi

10/18/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- The following methodologies were used to analyze data:

✓ Data Collection using web scraping and SpaceX API;
✓ Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
✓ Machine Learning Prediction.

- Summary of all results

✓ It was possible to collected valuable data from public sources;
✓ EDA allowed to identify which features are the best to predict success of launchings;
✓ Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- The objective is to evaluate the viability of the new company Space Y to compete with Space X.
- Desirable answers:
- The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets;
- Where is the best place to make launches.

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
- Data from Space X was obtained from 2 sources:
- Space X API (https://api.spacexdata.com/v4/rockets/)
- WebScraping(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform datawrangling
- Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

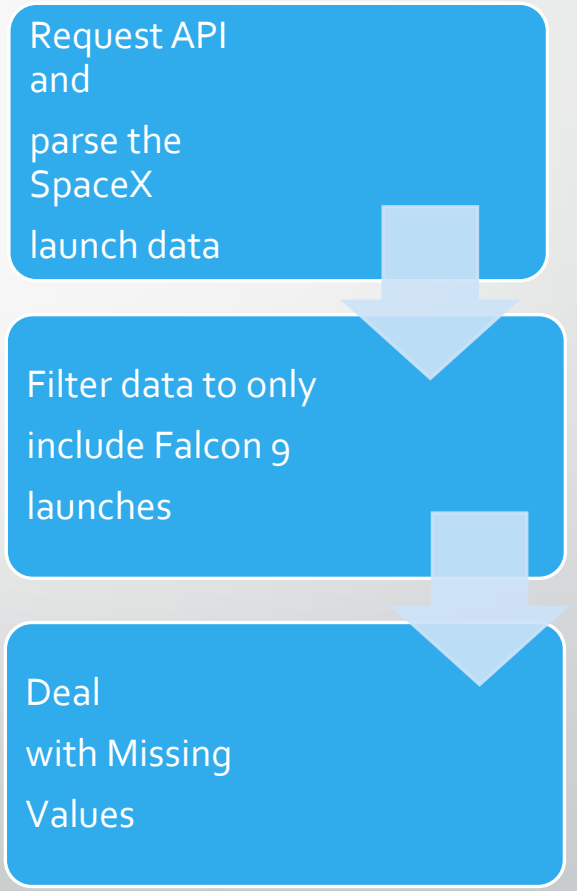
- Perform interactive visual analytics using Folium and PlotlyDash
- Perform predictive analysis using classification models
- Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection& Methods

- ✓ get request to the SpaceX API.
- ✓ json() function : decoded the response content as a Json
- ✓ json_normalize(): call and turn it into a pandas data frame using
- ✓ cleaned the data, checked handle missing values
- ✓ web scraping with BeautifulSoup. Source(from Wikipedia for Falcon 9 launch records)
- ✓ It was intended to extract the launch records as an HTML table, parse the information, and then transform the table into a pandas data frame for upcoming analysis

Data Collection – SpaceX API

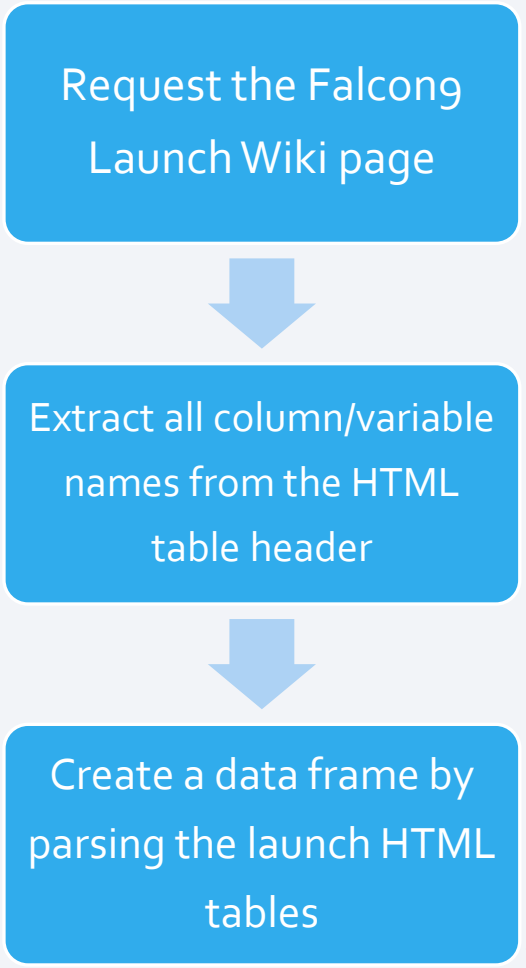
- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside and then data is persisted.



- Source code : https://github.com/marianataqi/IBM-Final-Capstone/blob/main/spacex-data-collection-api%20.ipynb

Data Collection -Scraping

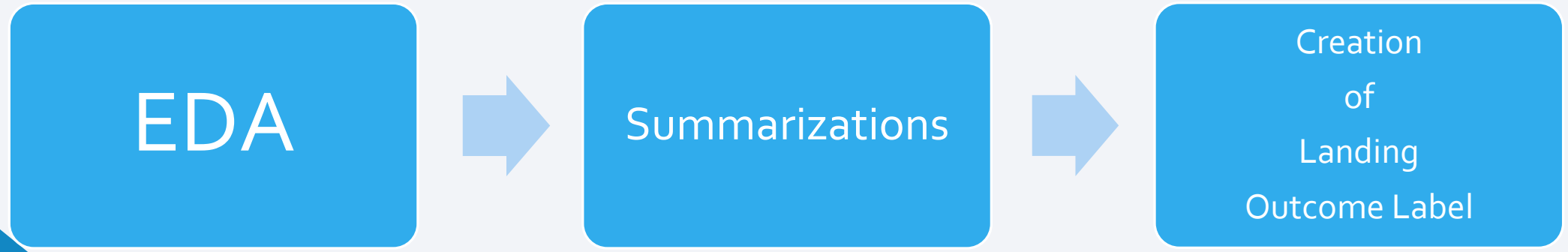
- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.



- Source code : https://github.com/marianataqi/IBM-Final-Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

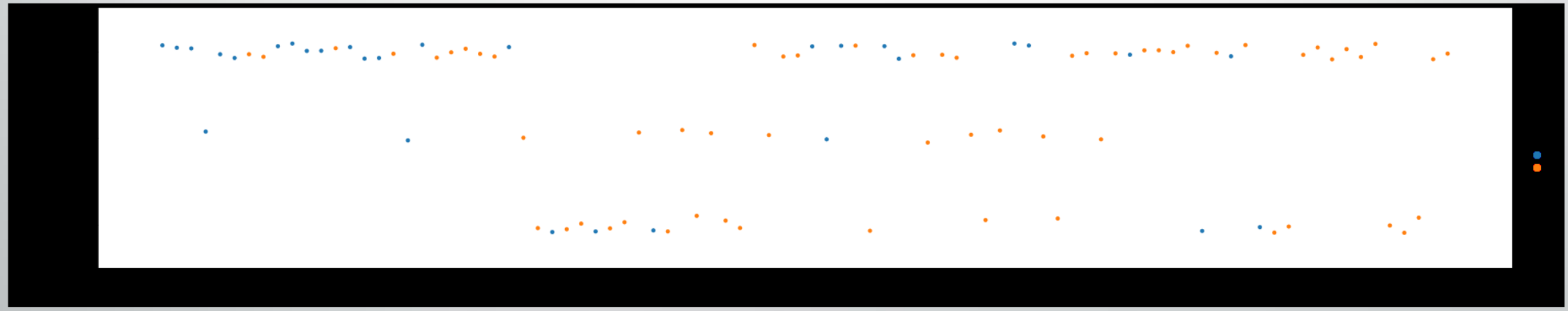
- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.



Source code: https://github.com/marianataqi/IBM-Final-Capstone/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplotswere used to visualize the relationship between pair of feature :
- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



- Source code : https://github.com/marianataqi/IBM-Final-Capstone/blob/main/EDA%20with%20Visualization.ipynb

EDA with SQL

The following SQL queries were performed:

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.
- Source code: https://github.com/marianataqi/IBM-Final-Capstone/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps;
- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site;
- Lines are used to indicate distances between two coordinates.

Source code : https://github.com/marianataqi/IBM-Final-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

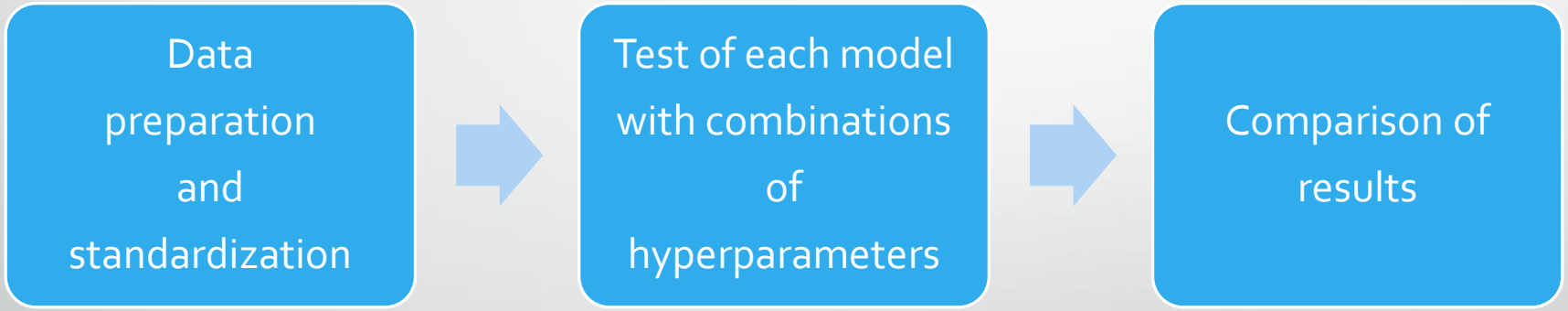
Build a Dashboard with PlotlyDash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads

- Source code: https://github.com/marianataqi/IBM-Final-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

• Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



• Source code: https://github.com/marianataqi/IBM-Final-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

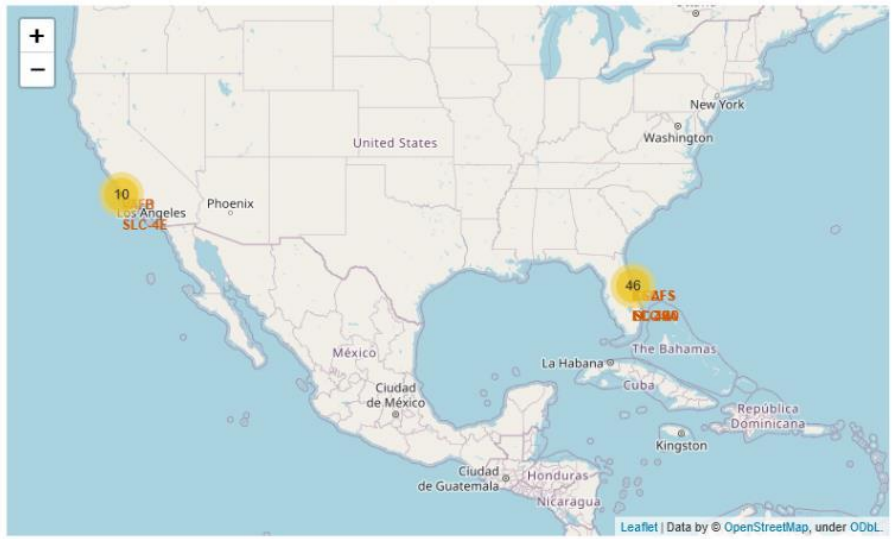
Result

Exploratory data analysis results:

- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed.

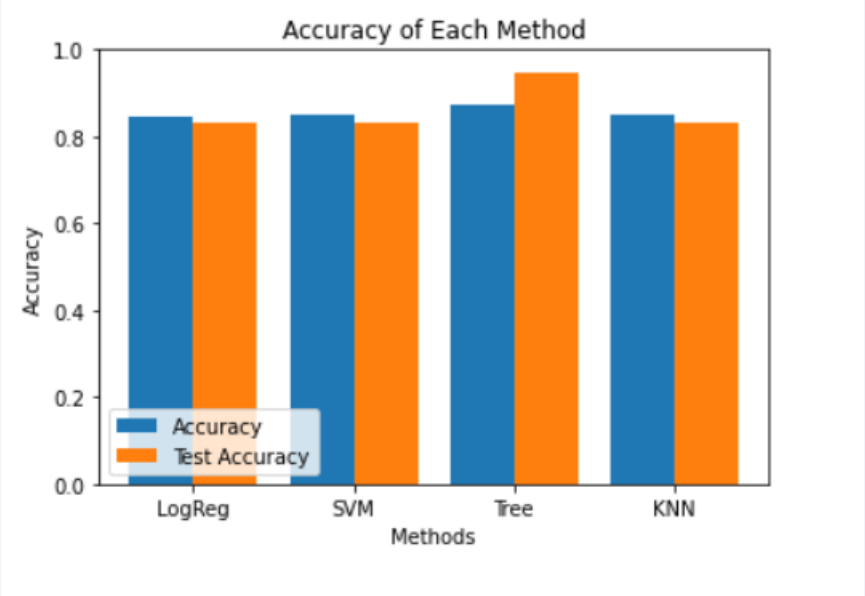
Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.



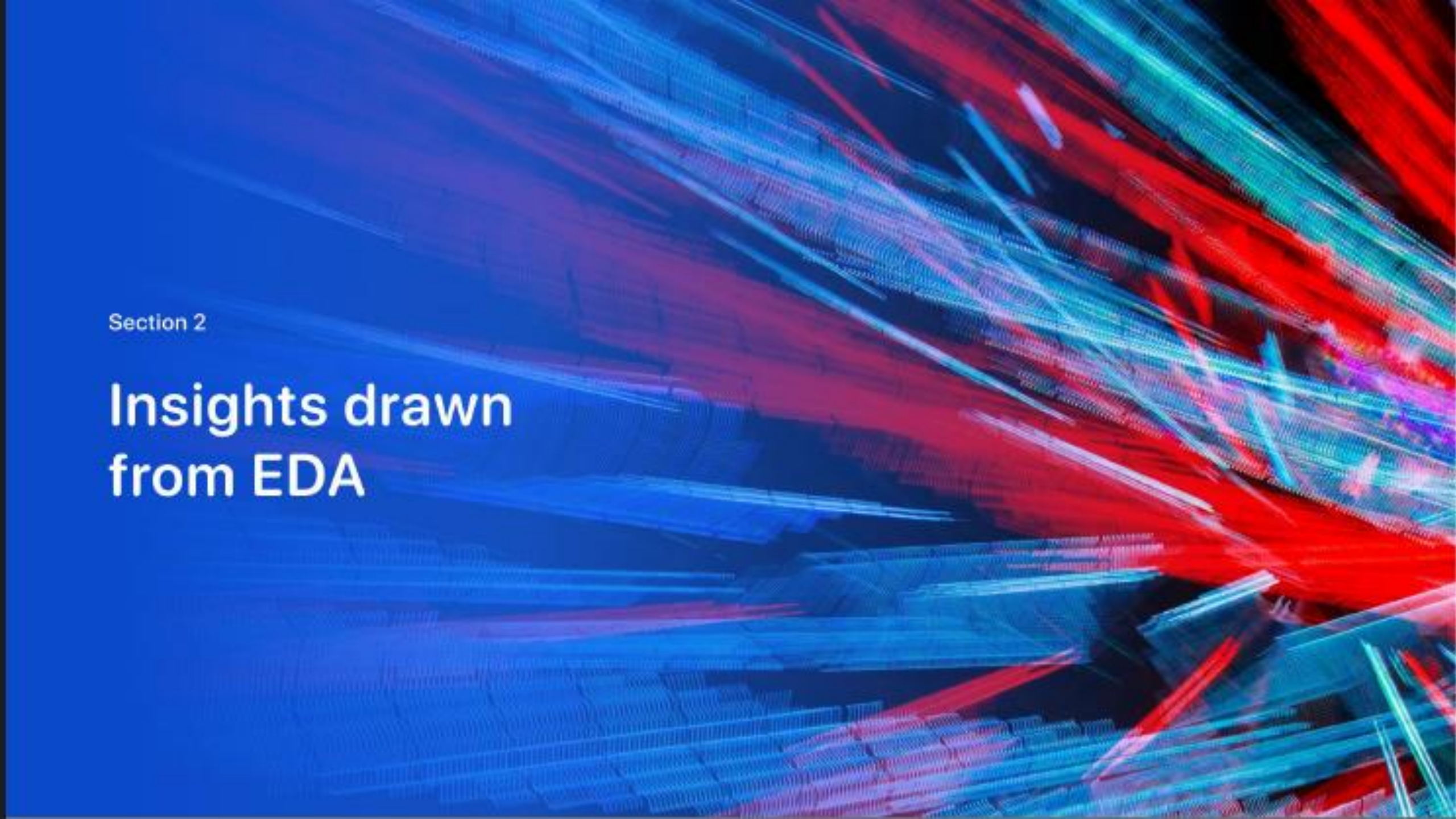
Results

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



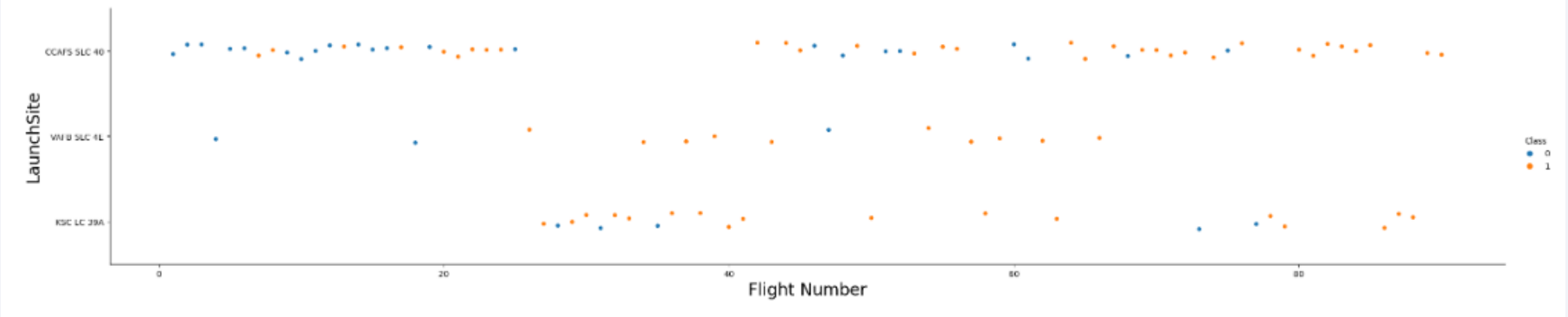
Section 2

Insights drawn from EDA



Flight Number vs. LaunchSite

- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.



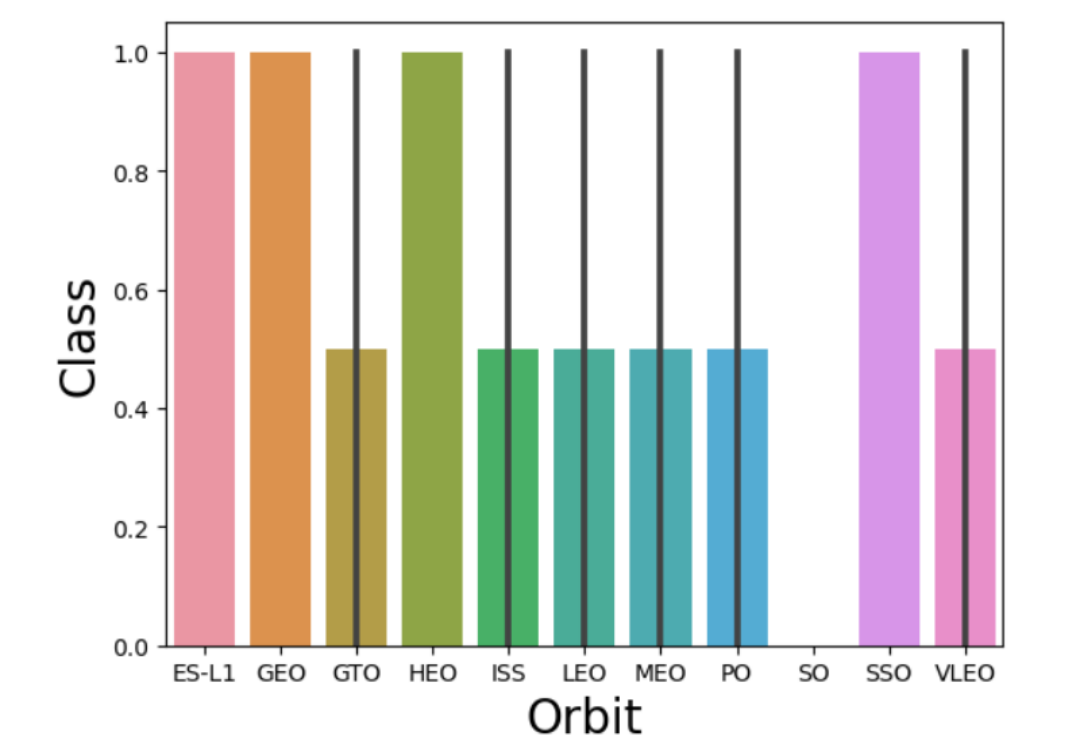
Success Rate vs. Orbit Type

The biggest success rates happens to orbits:

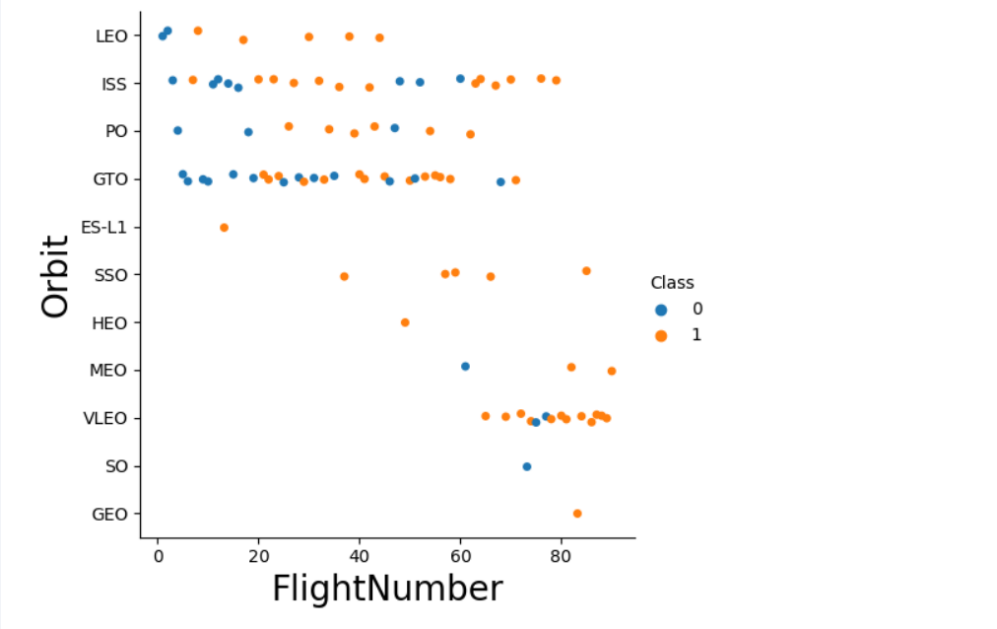
- ES-L1;
- GEO;
- HEO;
- SSO.

Followed by:

- VLEO (above 80);
- LFO (above 70%).



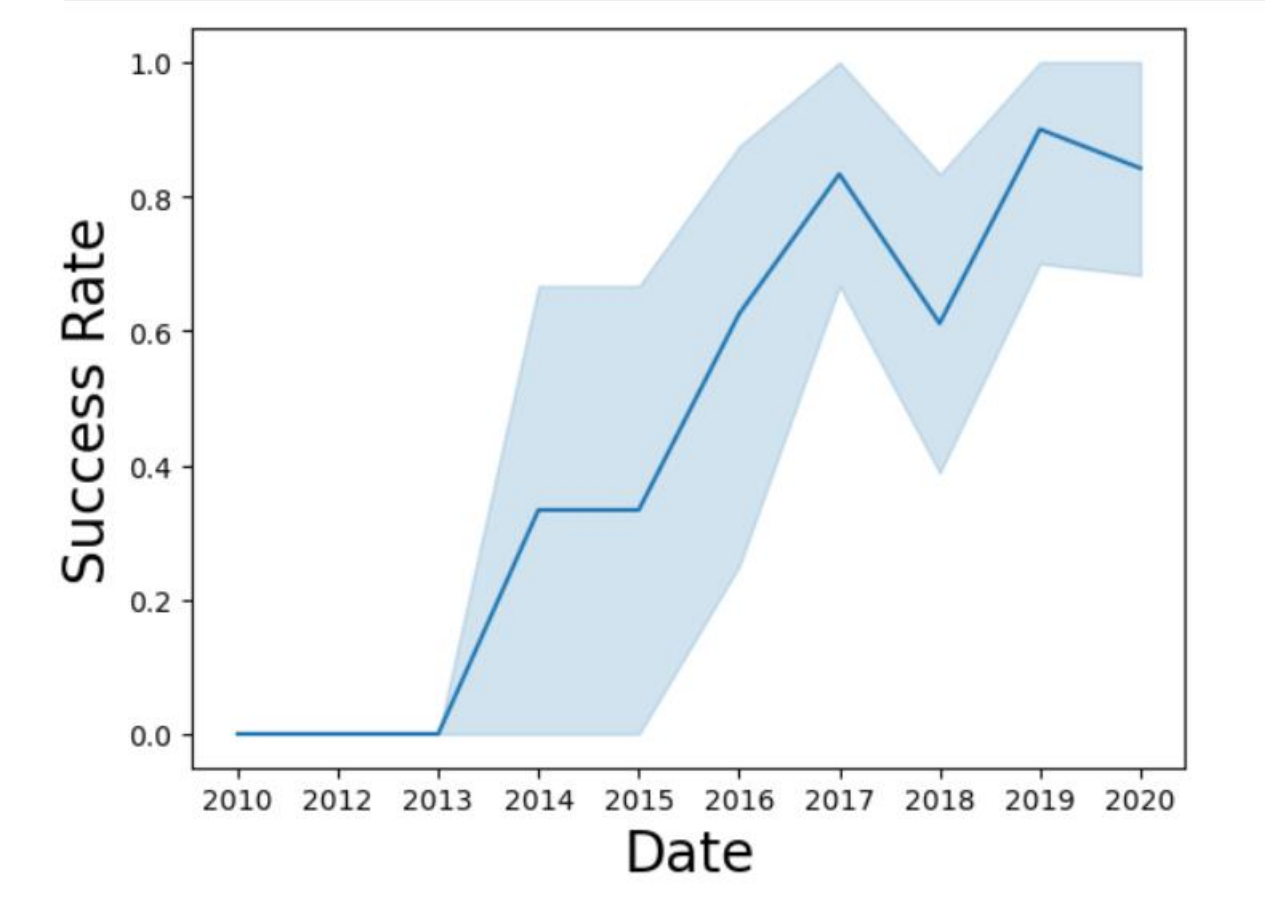
Flight Number vs. Orbit Type



- Apparently, success rate improved over time to all orbits;
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.

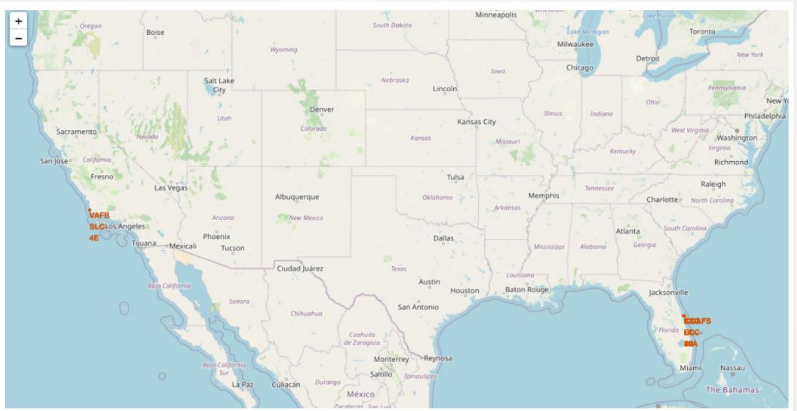


Section 4

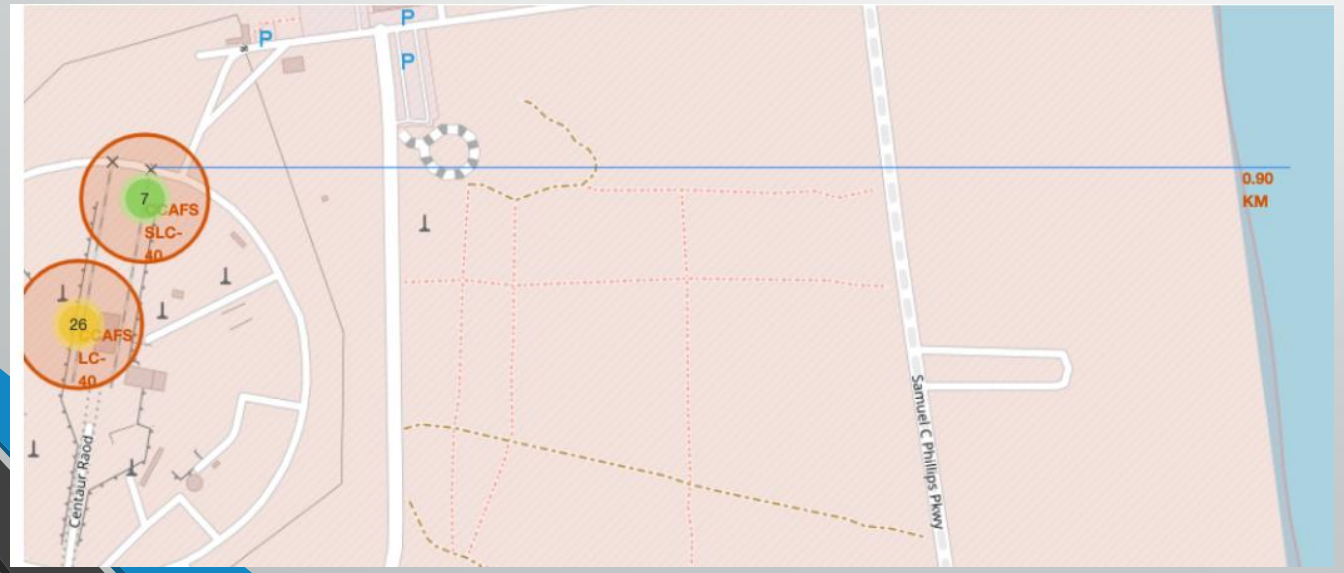
Launch Sites Proximities Analysis

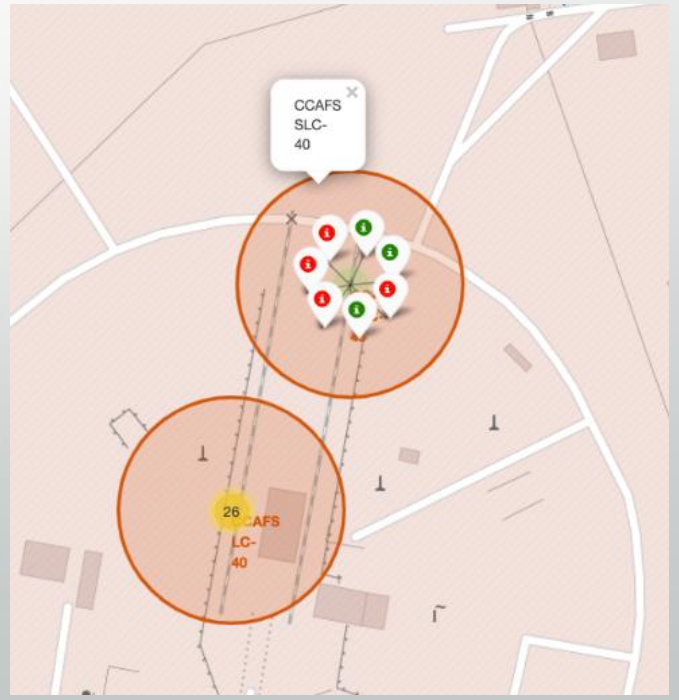
All launch sites

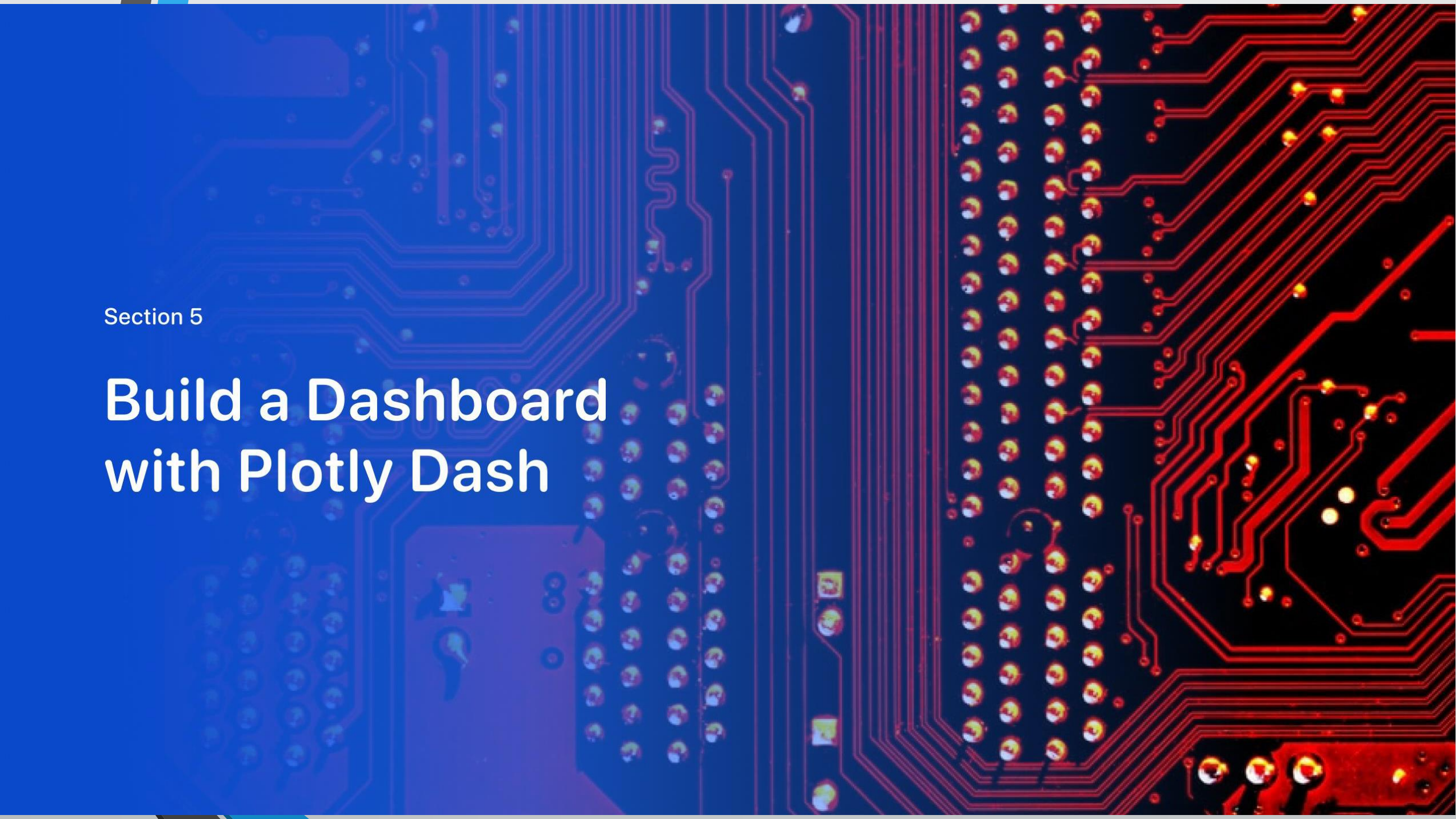
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.



- Example of CCAFS SLC-40 launch site launch outcomes

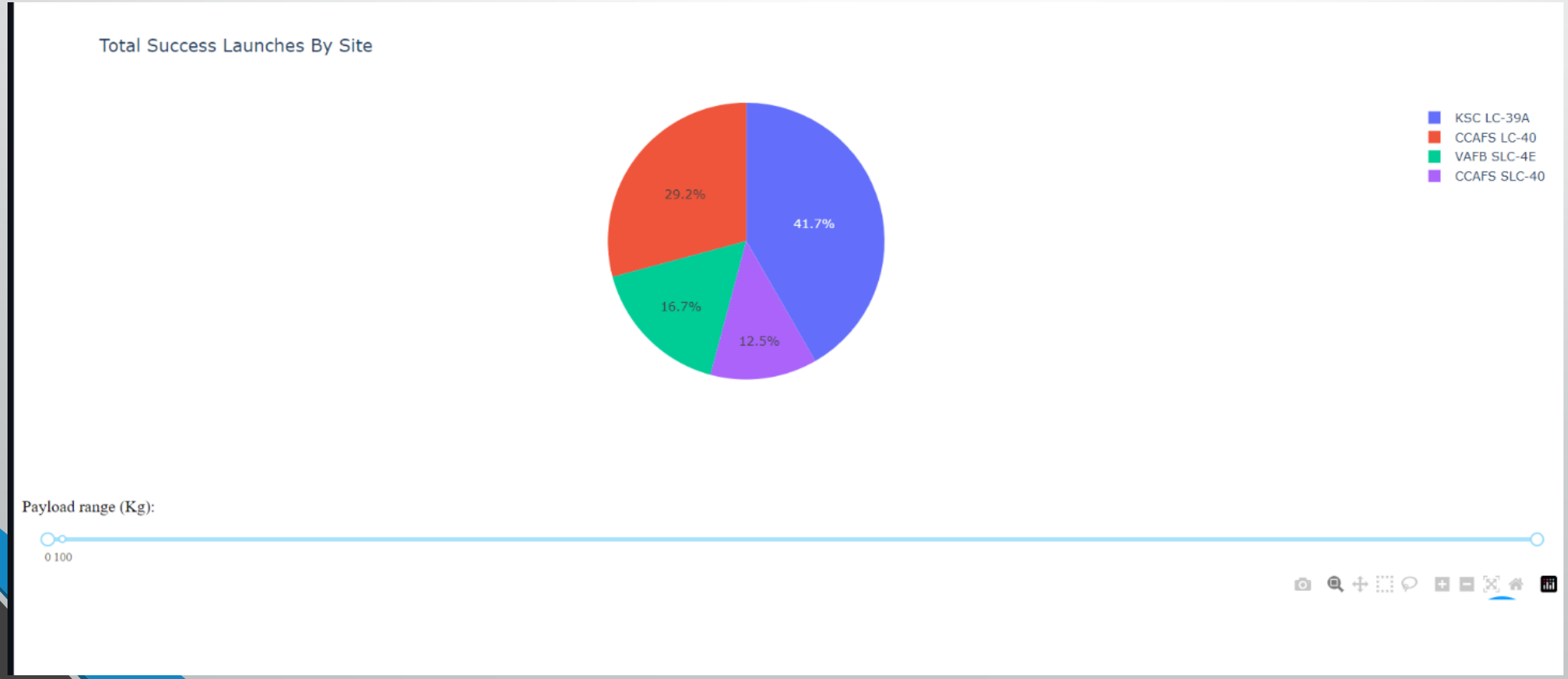




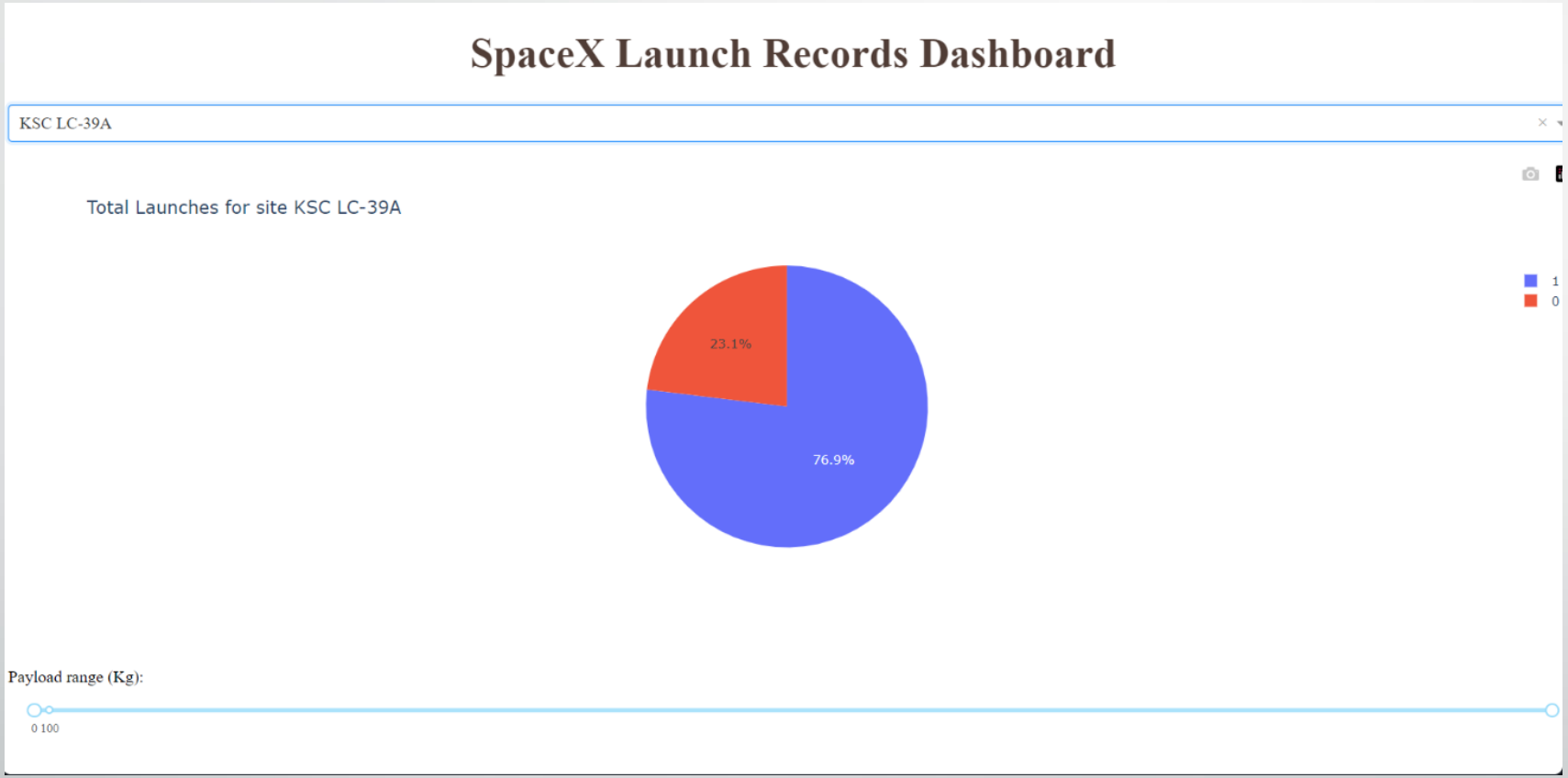

Section 5
Build a Dashboard with Plotly Dash

Successful Launches by Site

•The place from where launches are done seems to be a very important factor of success of missions.

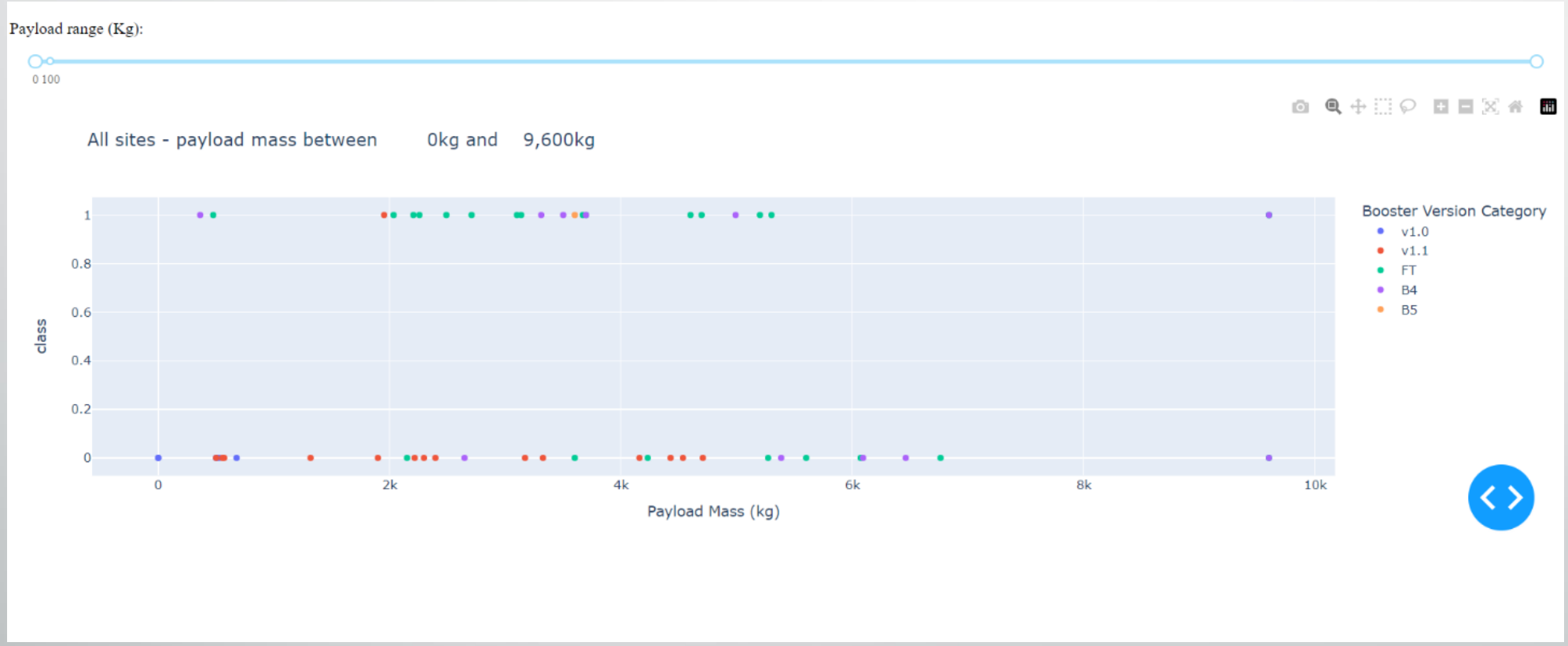


Launch Success Ratio for KSC LC-39A



- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



- Payloads under 6,000kg and FT boosters are the most successful combination.

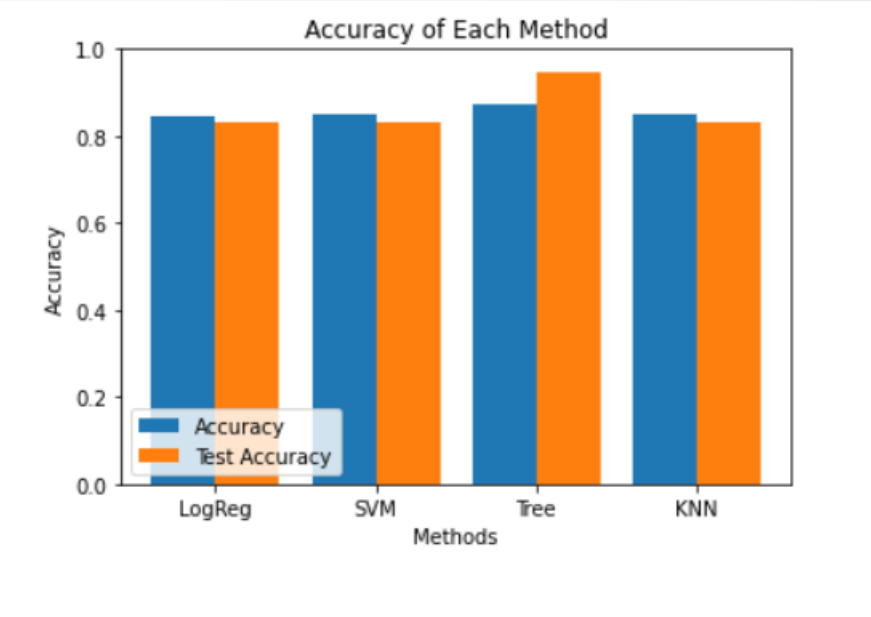
Section 6

Predictive Analysis (Classification)

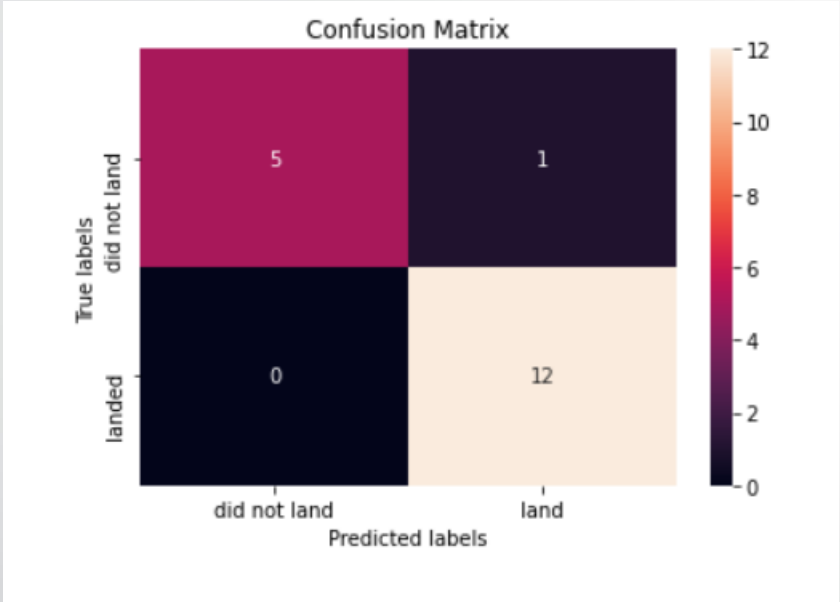
Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside;

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix of Decision Tree Classifier



- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!