IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anurak Jittang
18/10/2565

Outline

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

Executive Summary

- Summary of methodologies
  - SpaceX API and Web Scraping method were used to gather data.
  - Preprocess the data by the data wrangling method and visualized them
  - Fit them into machine learning model
- Summary of all results
  - Data were successfully gathered with methods mentioned earlier
  - Preprocessing data made the data easier to work with
  - Machine learning can predict the outcome with informative features of the data

Introduction

- The objective of this project is to find new knowledge to remain in business
- The answer we wanted are the best way to predict the cost and where is the best place for the rocket to launch.

Section 1

Methodology

Methodology

Executive Summary

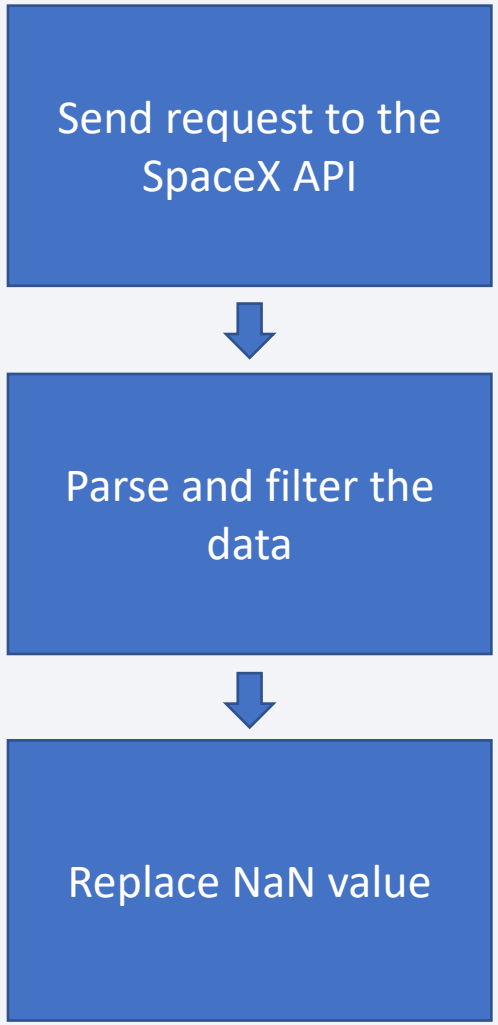
- Data collection methodology:
  - Space X API
  - Web Scraping
- Perform data wrangling
  - Informative features such as Classes were added to the data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data were gathered from 2 channels which are SpaceX API and Wikipedia (Web Scraping method)

Data Collection – SpaceX API

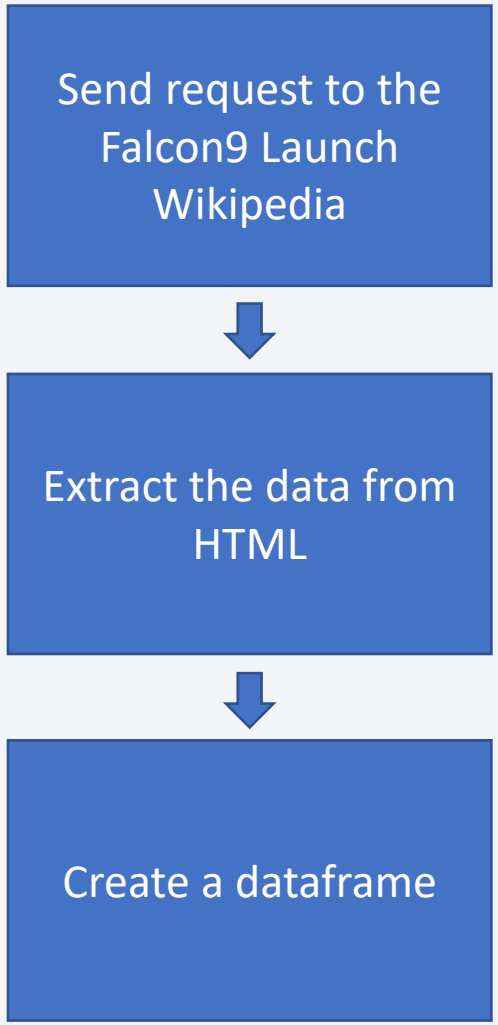
- SpaceX REST API is publicly available on the internet.
- The data were obtained according to the flowchart



- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/data_collection.ipynb

Data Collection - Scraping

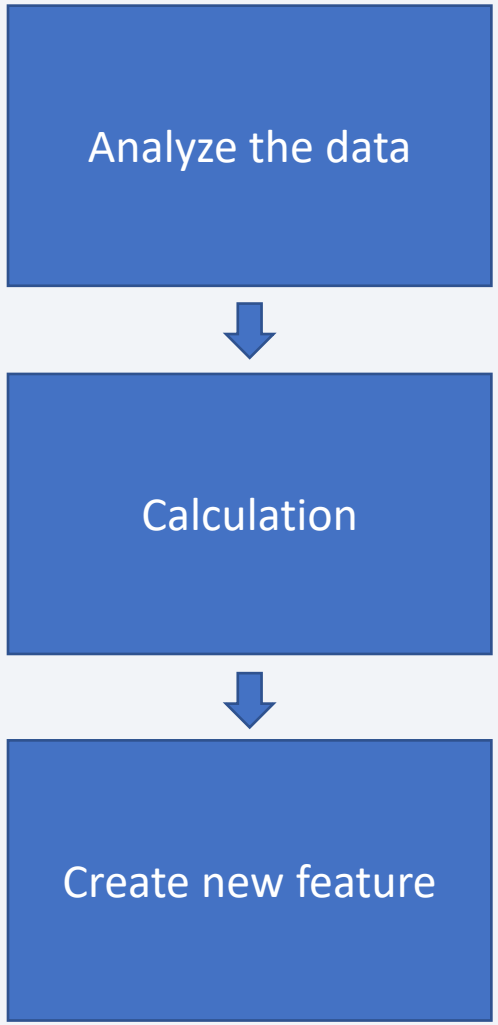
- SpaceX Data is also available on the Wikipedia.
- The data were obtained according to the flowchart
- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/data_collection.ipynb



Data Wrangling

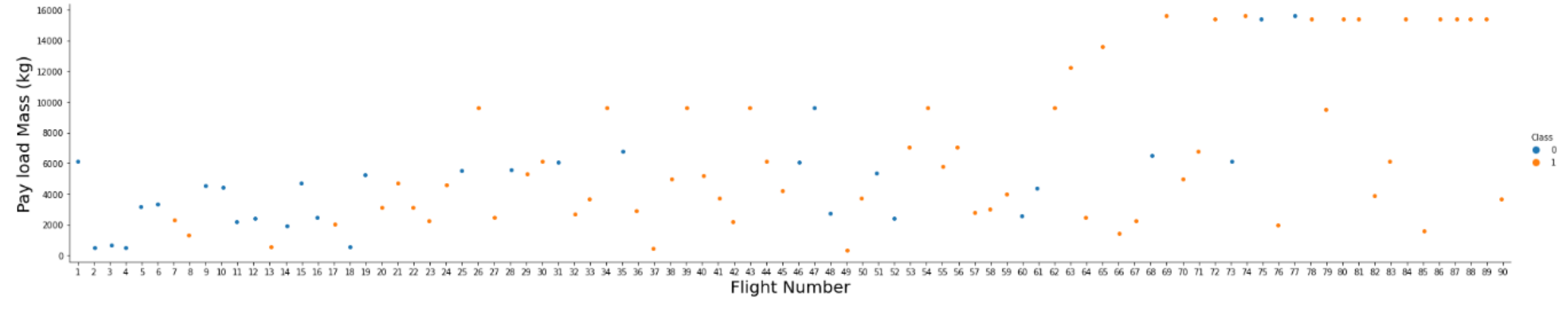
- First, analyze the data
- Second, Calculate the number and the occurrence of mission outcome per orbit type
- Create new feature

- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/Data-wrangling.ipynb



EDA with Data Visualization

- Scatter plots were used to display a relationship between features and Flight Number
  - The Scatter plot below represent a relationship between Payload Mass and Flight Number



- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/eda1 1dataviz.ipynb

EDA with SQL

- List of SQL Queries used
  - 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
  - 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: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/eda_sql.ipynb

Build an Interactive Map with Folium

- Folium Maps and features including markers, circles, lines and marker clusters were used
  - Marker: Determine launch site
  - Circle: Highlighting some area
  - Marker Cluster: Determine group of launch site
  - Line: Determine distance
- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/folium.ipynb

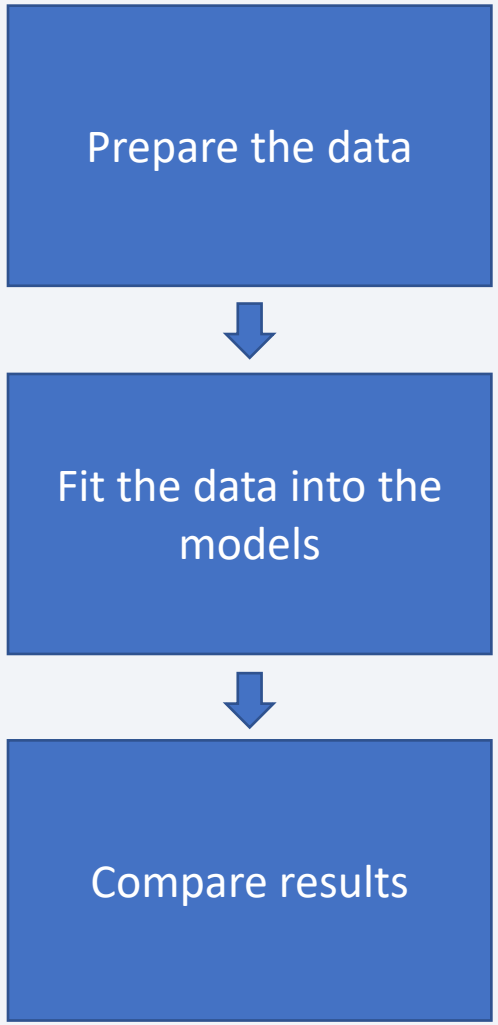
Build a Dashboard with Plotly Dash

- Percentage of launches by site and Payload range were used to visualize the graph
  - These are two were very effective to help identify the best launch place.
- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/spacex_dash_app.py

Predictive Analysis (Classification)

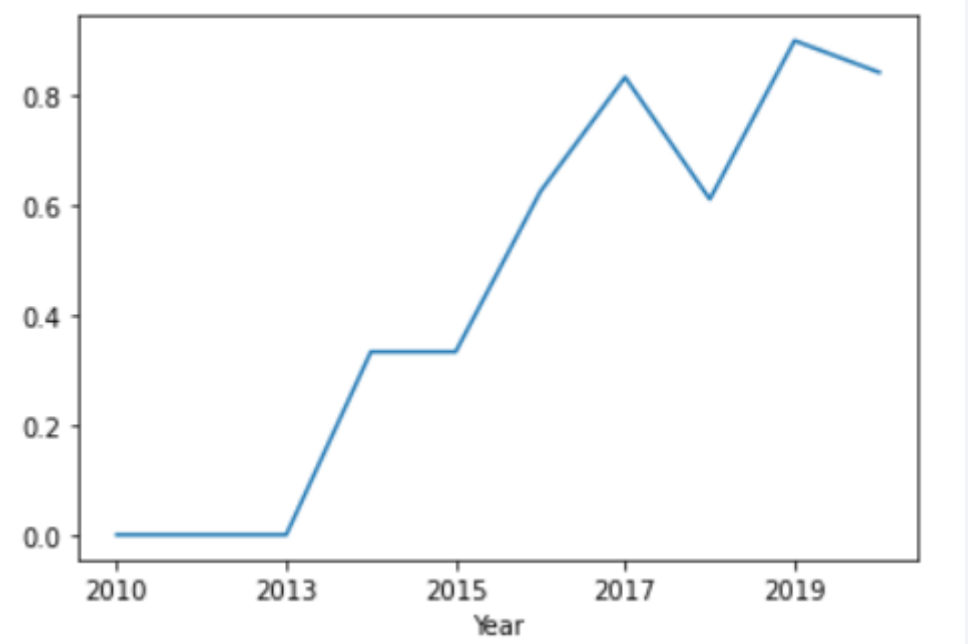
- Logistic regression, support vector machine, decision tree and k nearest neighbors were used and compared to each other.

- Source: https://github.com/anurakamity/assignment_IBM/blob/main/Final%20Assignment/machine_learning.ipynb



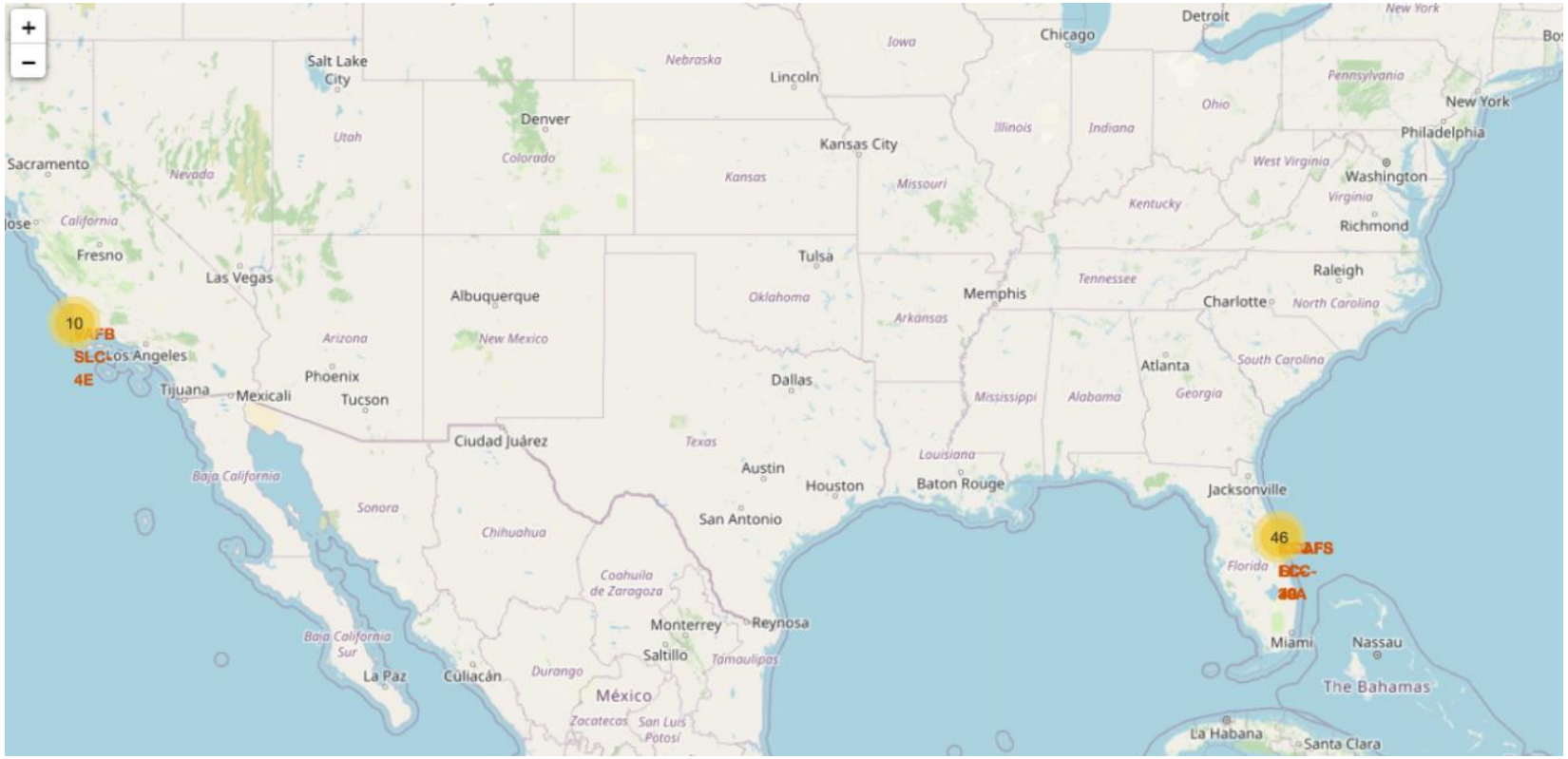
Results

- Exploratory data analysis results
  - The line chart below represents the relation ship between Successful Launch rate and Year.
  - We can see that the trend is increasing



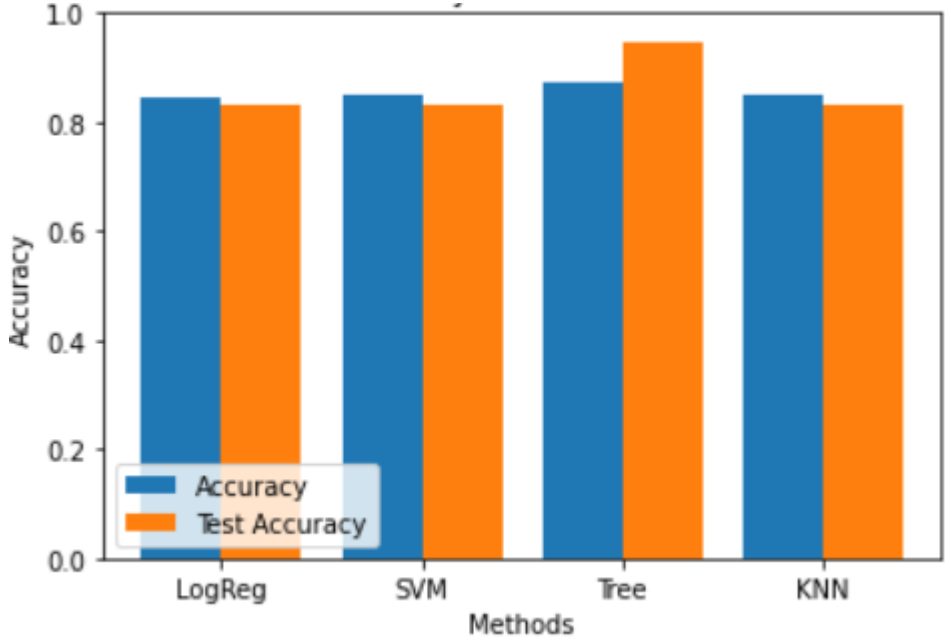
Results

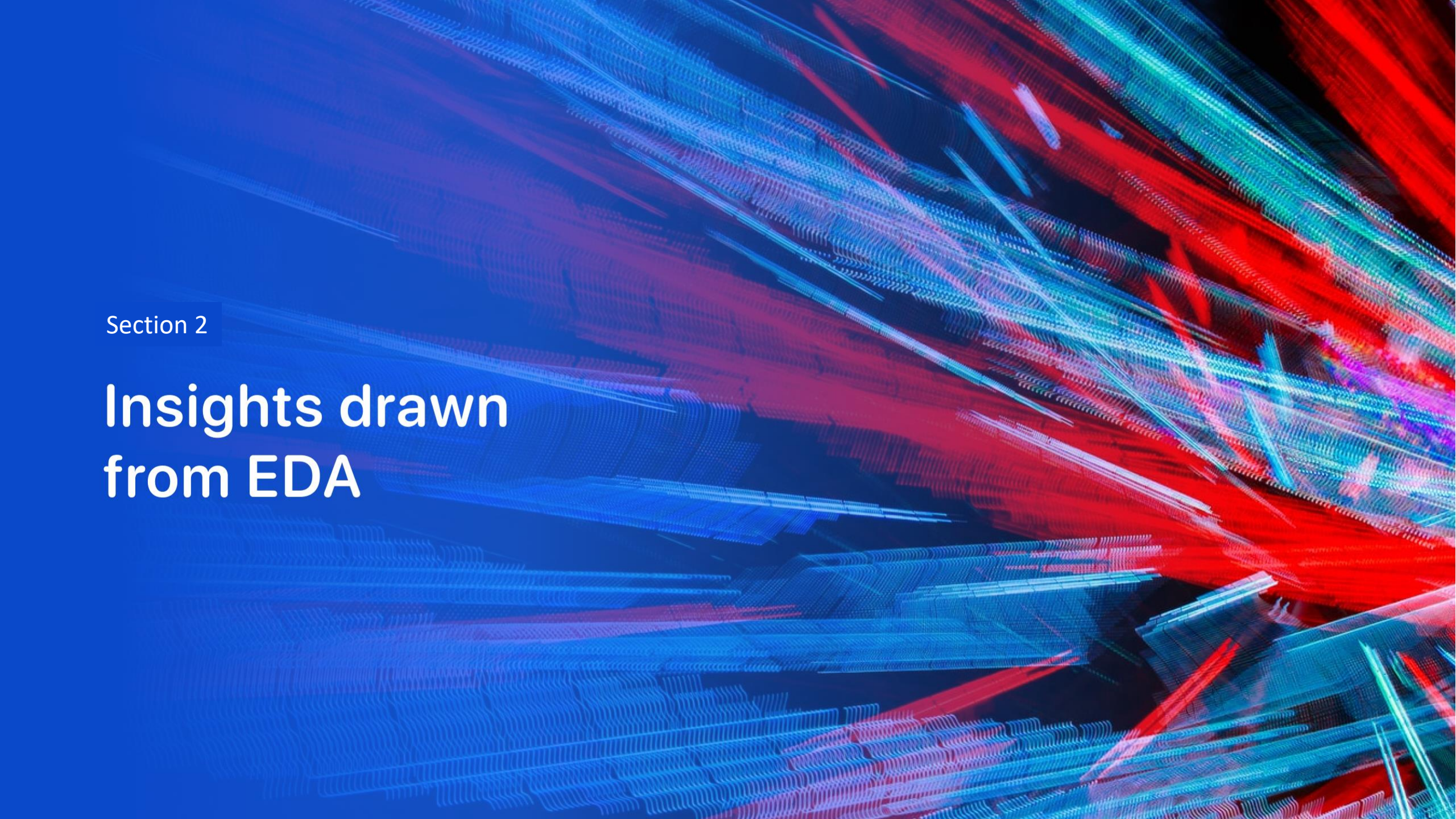
- Interactive analytics demo in screenshots
  - Most of the launches were performed on the East coast
  - All of them were performed near sea area for safety purposes



Results

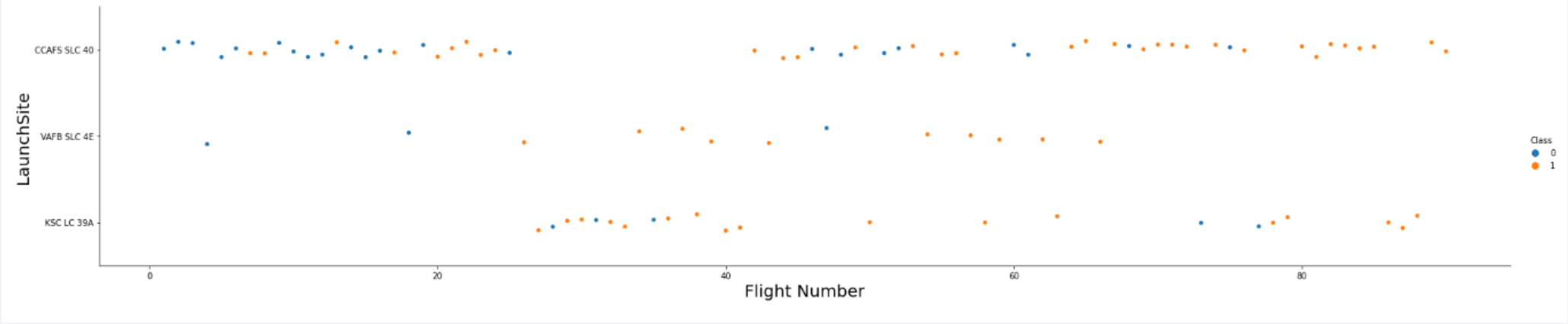
- Predictive analysis results
  - The bar chart below represents accuracy for each method
  - The best method is Decision Tree




Section 2
Insights drawn from EDA

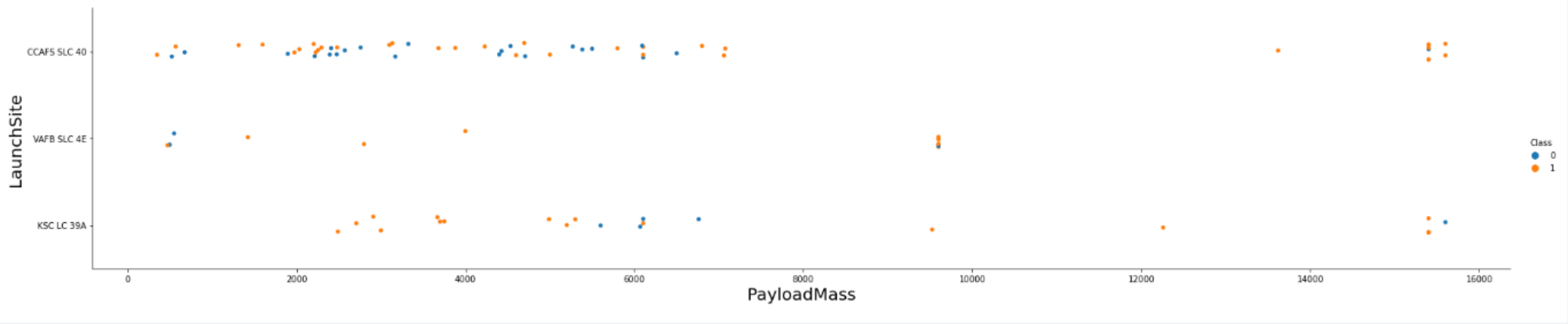
Flight Number vs. Launch Site

- KSC LC-39A is the best launch site because of its high success launch rate



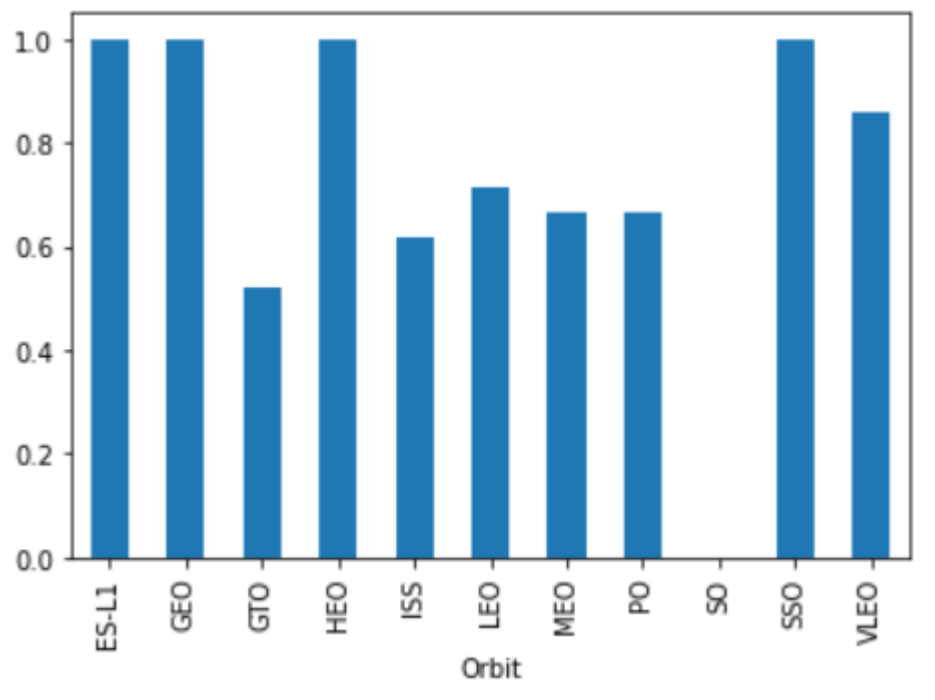
Payload vs. Launch Site

- Payloads that are heavier than 8000 kg seems to have higher success rate



Success Rate vs. Orbit Type

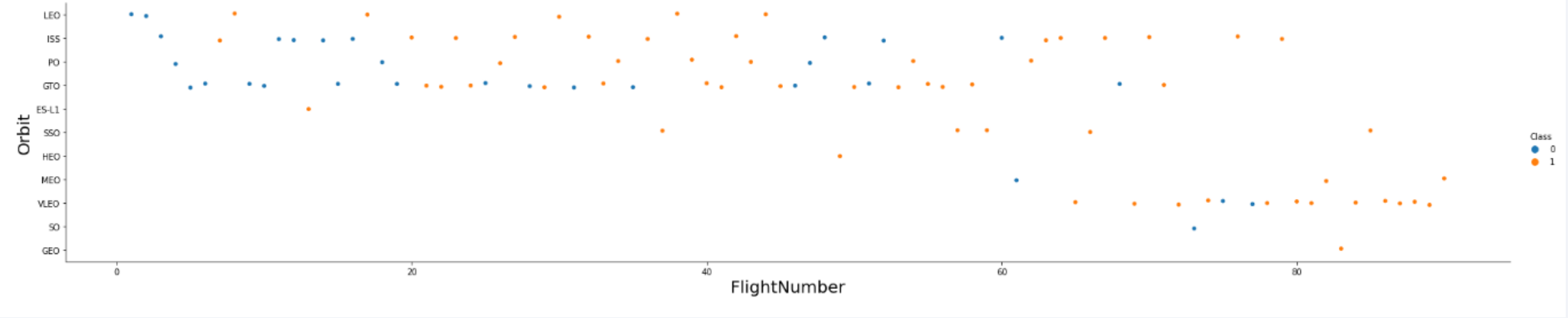
- This is a bar chart of Success Rate vs Orbit Type
- The orbit types with the greatest success rate are ES-L1, GEO, HEO, and SSO
- The orbit type with the least success rate is GTO.



Flight Number vs. Orbit Type

- This is a scatter point of Flight number vs. Orbit type

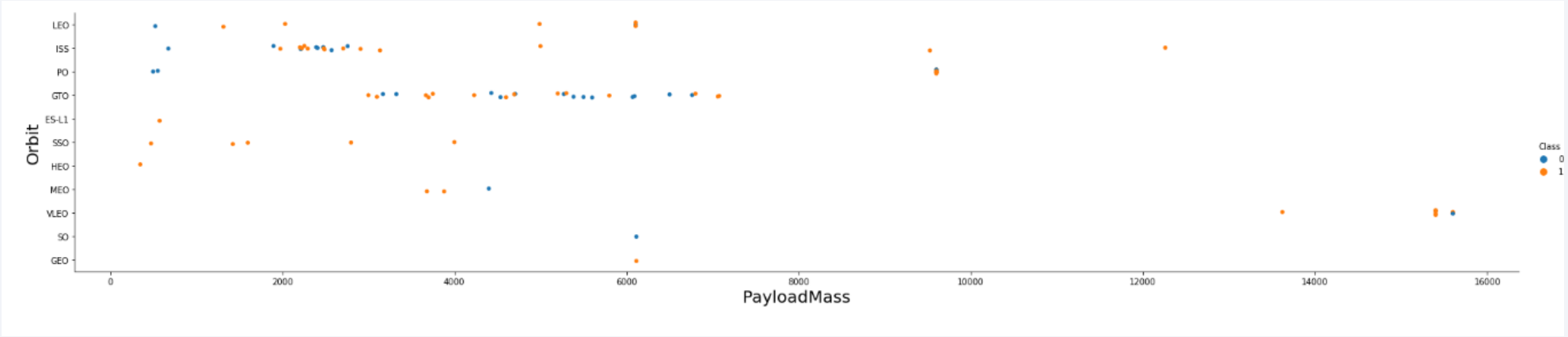
- VLEO type gain popularity in the last 20 launches, 14 of them being VLEO



Payload vs. Orbit Type

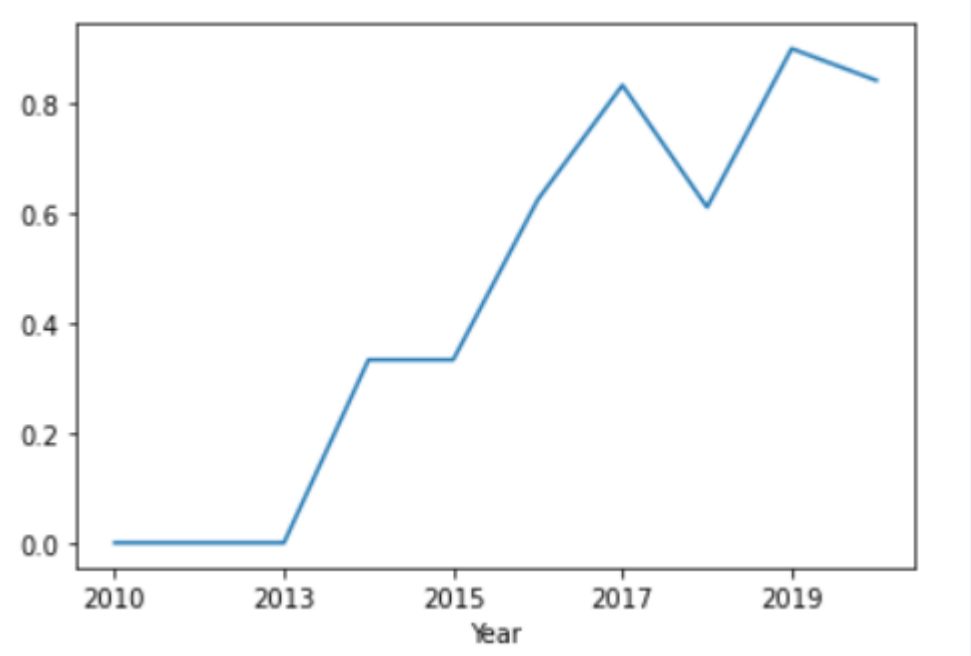
- Below is a scatter point of payload mass vs. orbit type

- Most of the launches had a payload mass that is less than 8000 kg
- Only few of them is higher than 8000 kg



Launch Success Yearly Trend

- This is a line chart of yearly average success rate
- The trend is increasing meaning that if the trend continue the success rate is the future will be higher.



All Launch Site Names

- The names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload carried by boosters from NASA

Total Payload Mass
111268

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

Average Payload Mass by F9 v1.1
2928

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

First Successful Ground Landing Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

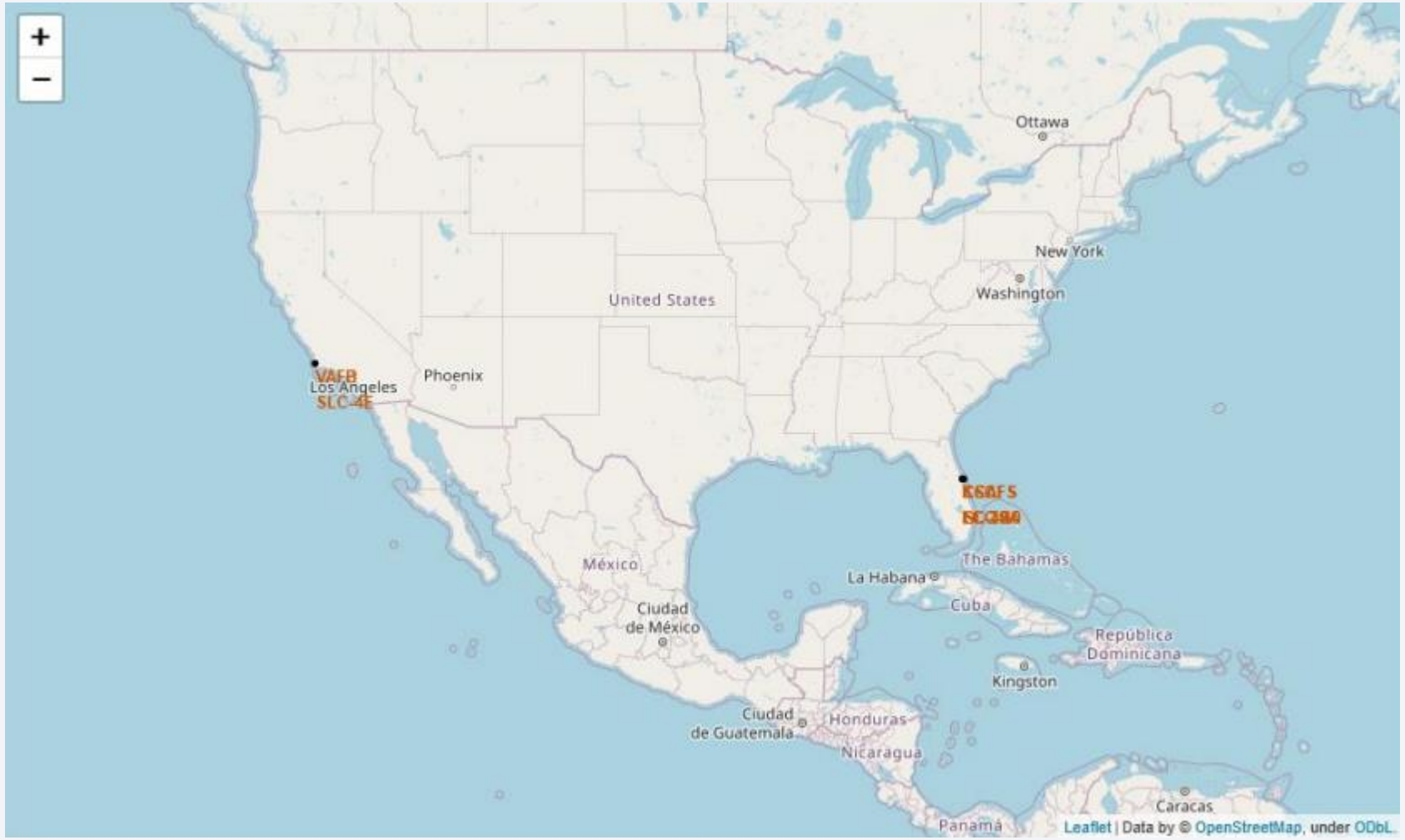
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

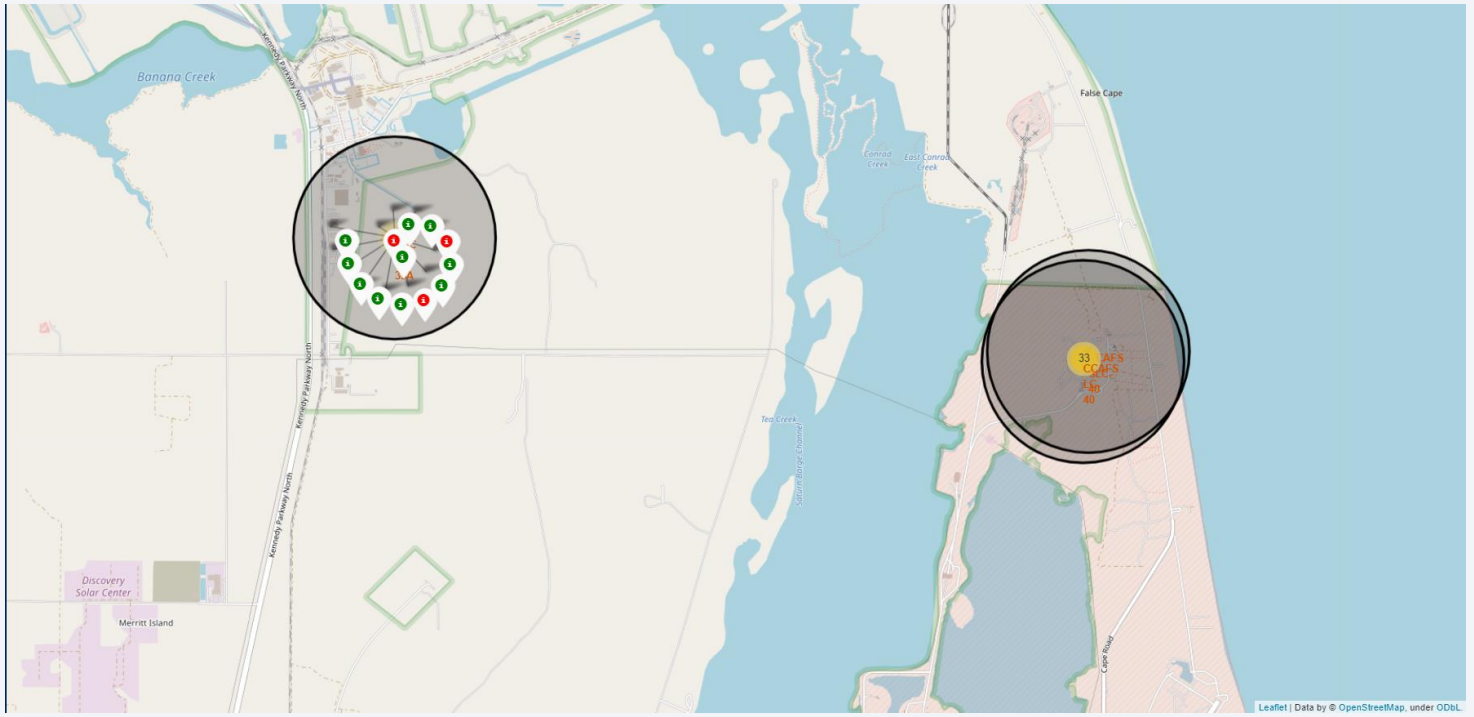
All Launch sites



- All of them are near the sea for safety purposes

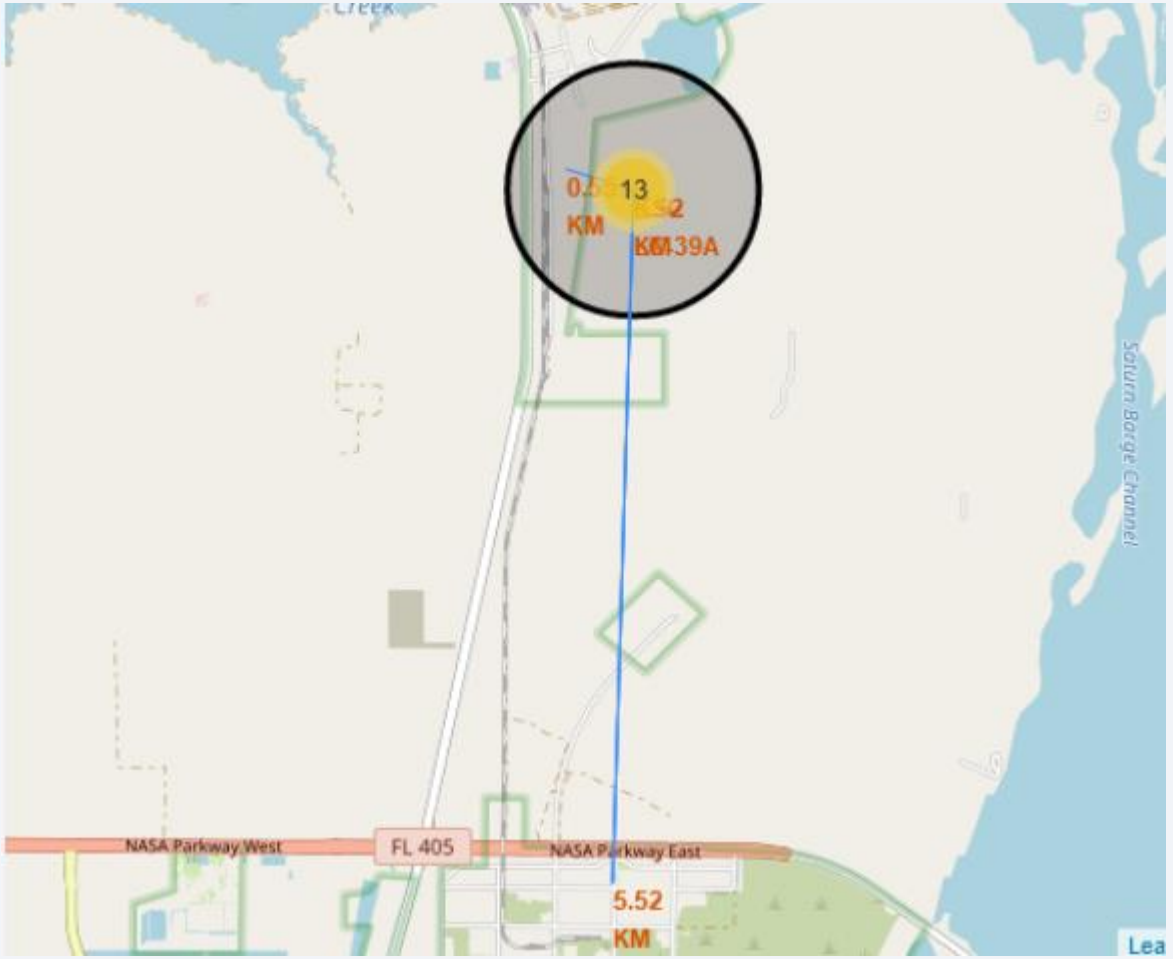
Launch Outcome by site

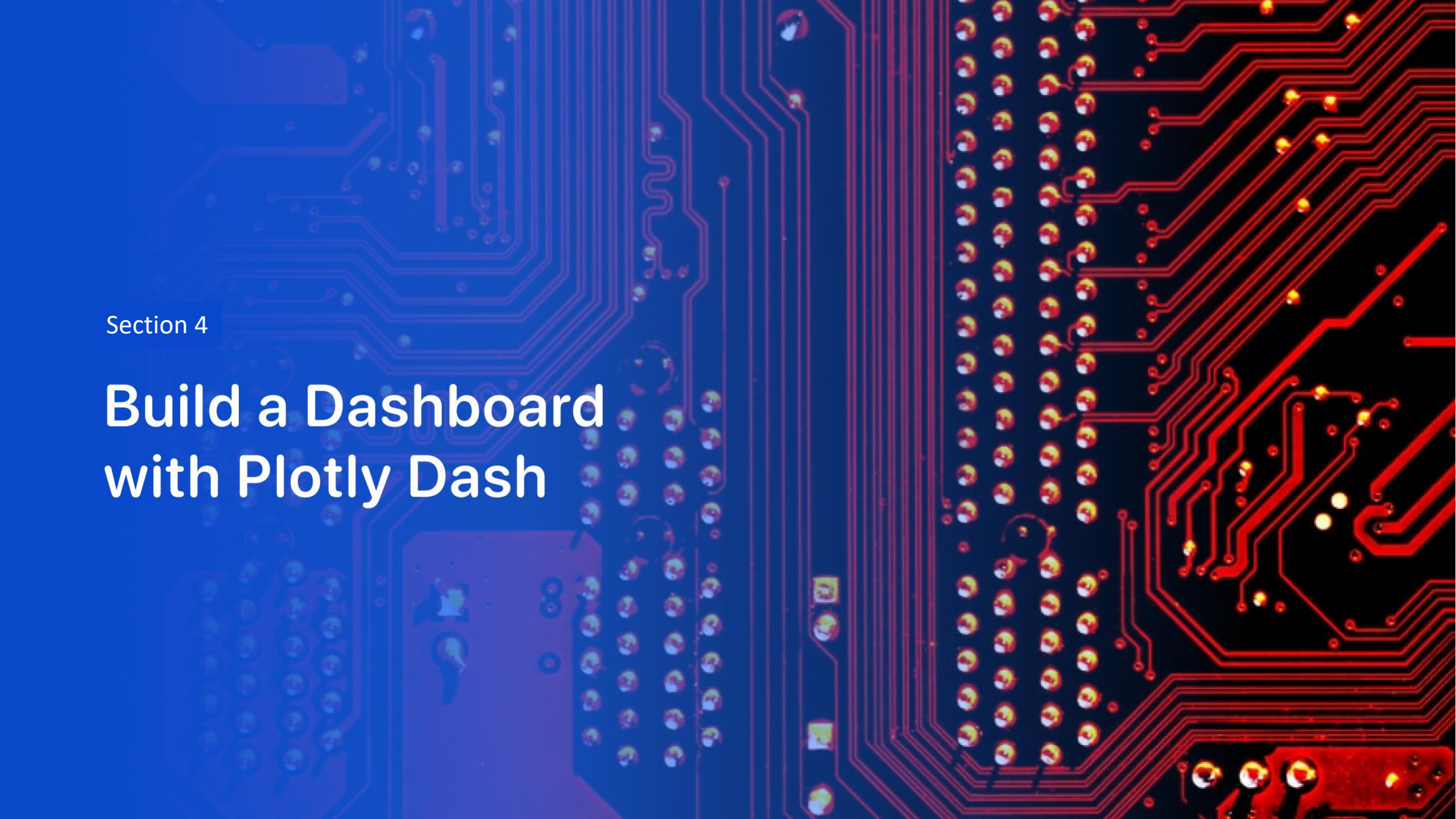
- Red marker: Fail
- Green marker: Success



Distance from Road

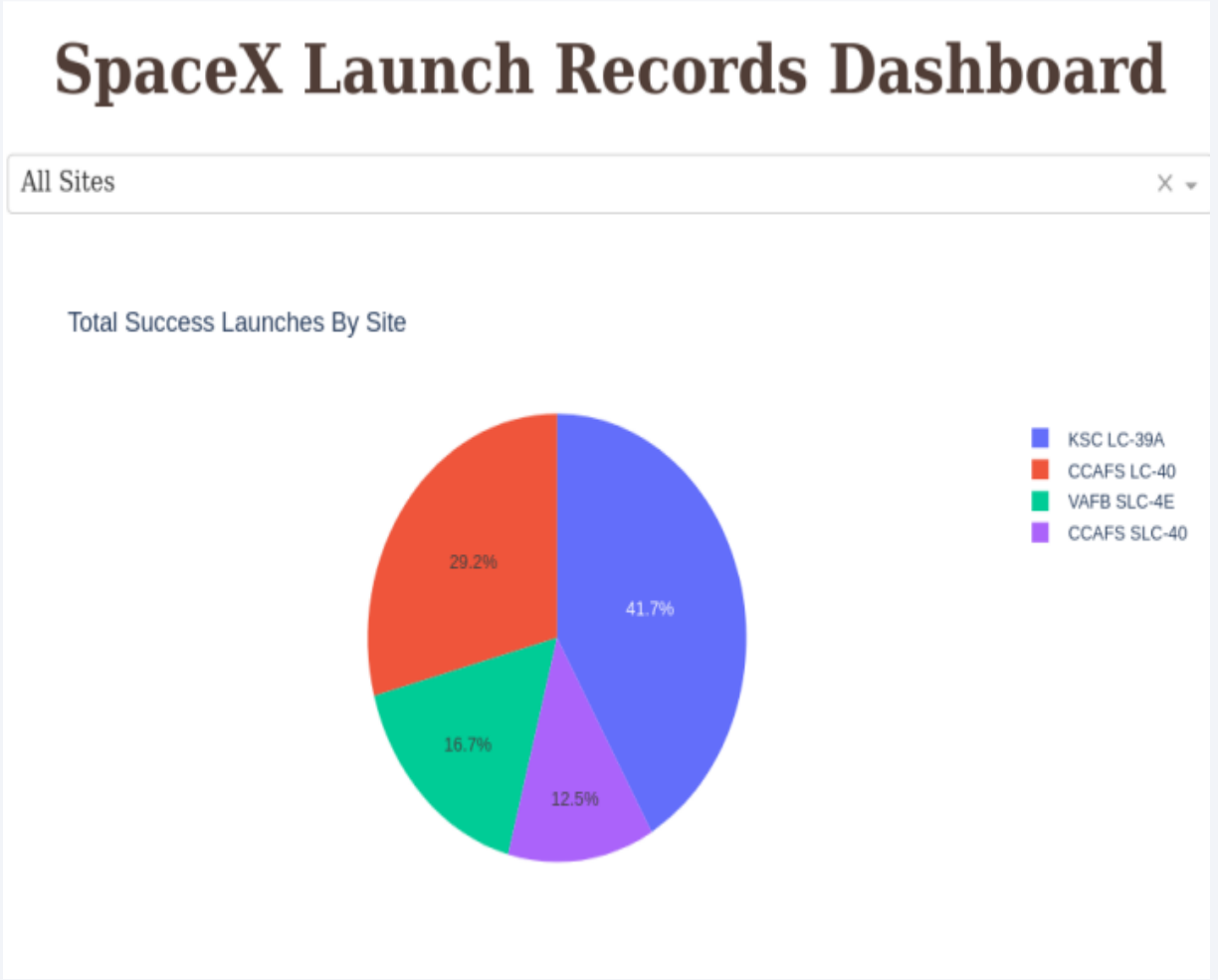
- It is far from community but relatively close to railroad. KSC LC-39A is a good option to consider.




Section 4
Build a Dashboard with Plotly Dash

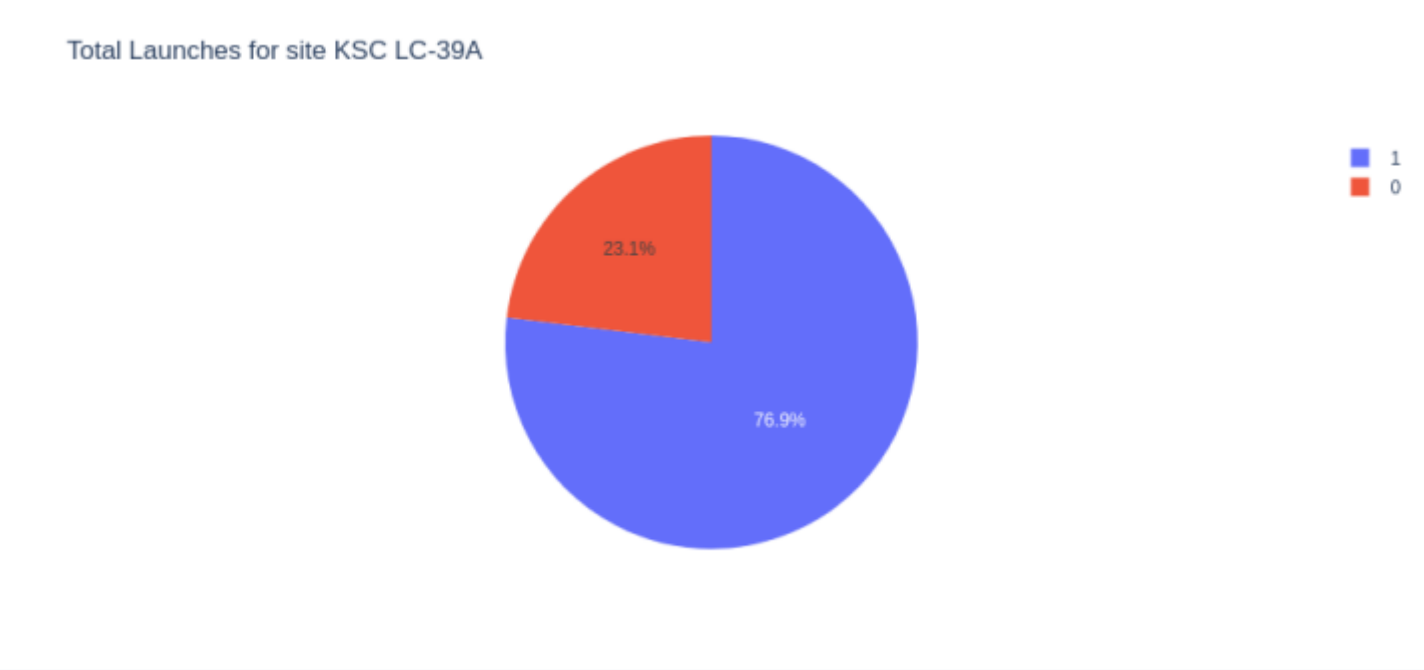
Successful Launches by Site

- KSC LC-39A has the greatest number of successful launches.

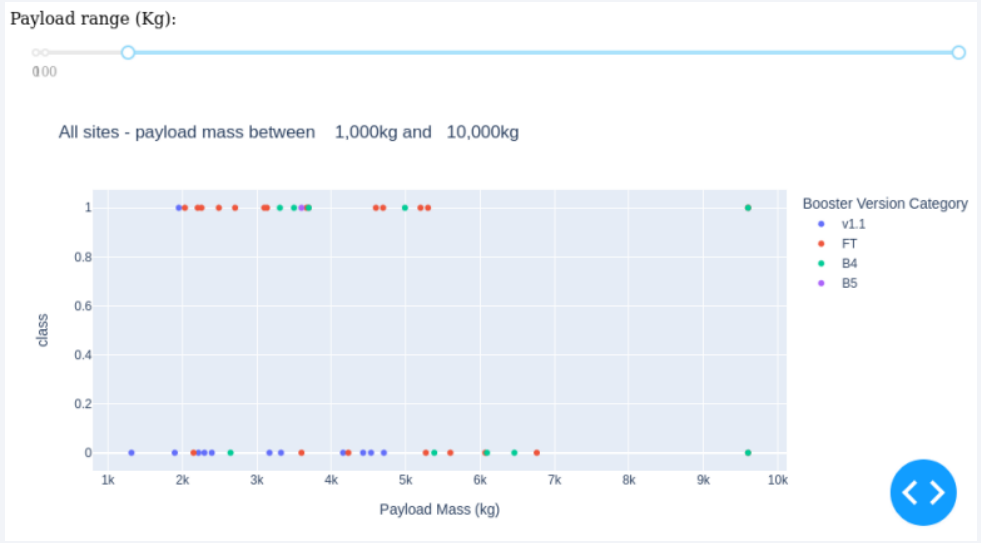


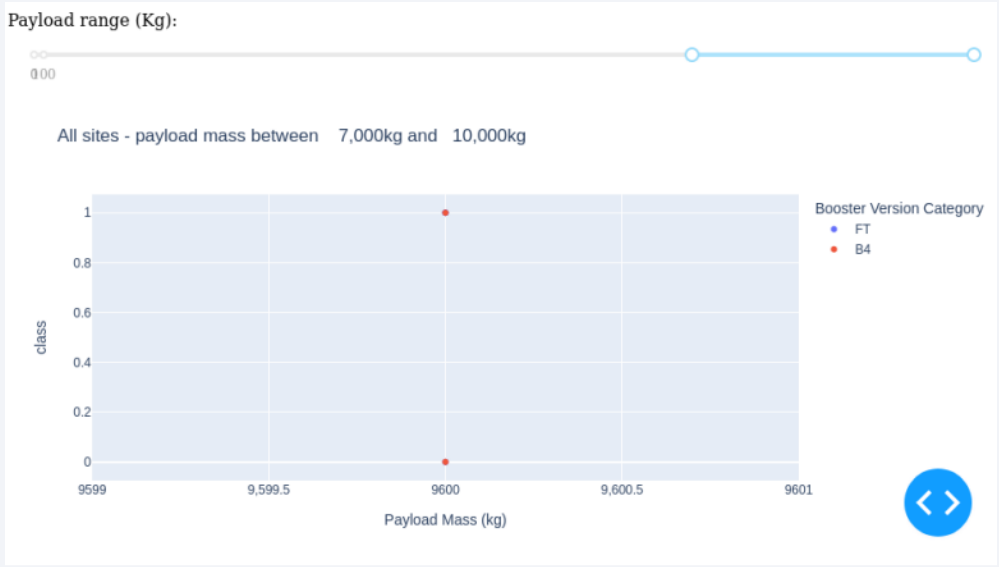
Successful rate for site KSC LC-39A

- 76.9 successful rate for site KSC LC-39A



Payload mass and Outcome





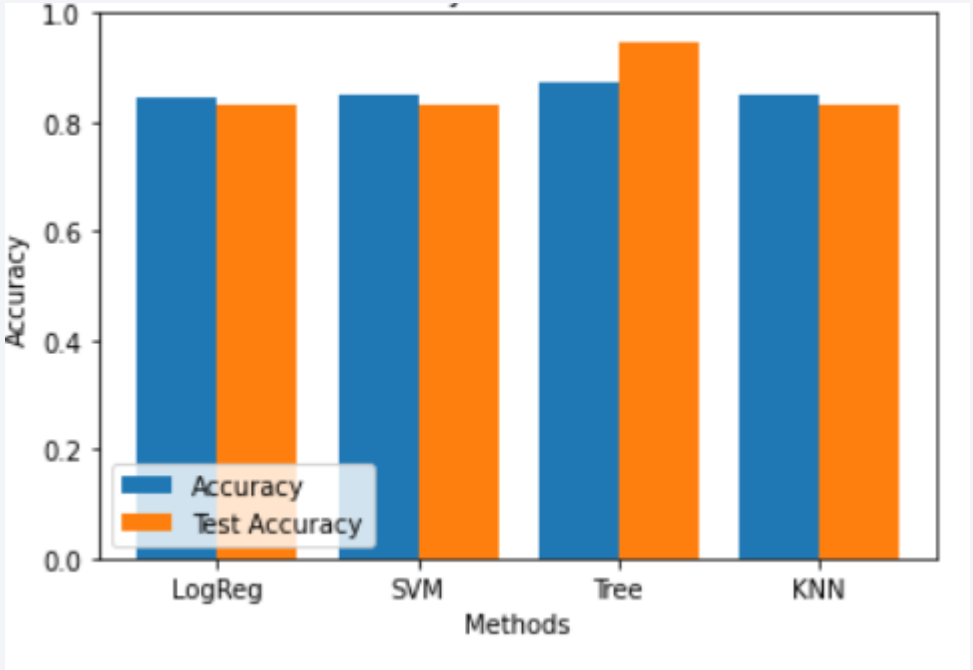
- Most effective combination is FT model with the Payload under 6000 kg

Section 5

Predictive Analysis (Classification)

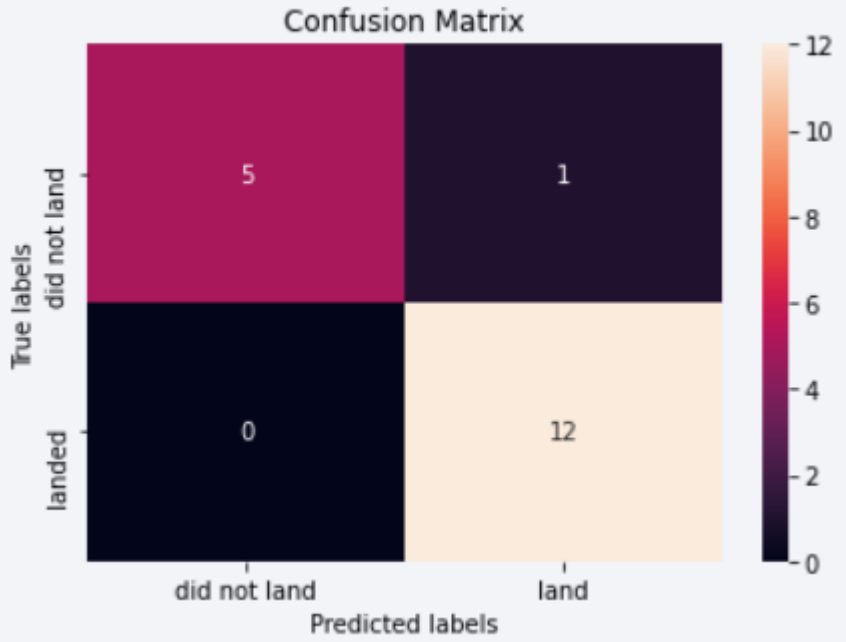
Classification Accuracy

- The bar chart on the right represents accuracy for each method including Logistic Regression, SVM, Decision Tree, KNN
- The best model is Decision Tree



Confusion Matrix

- It has a relatively great number on True positive and True negative. Therefore, it has high accuracy.



Conclusions

- KSC LC-39A is the best launch site
- Payload should be less than 7000kg for minimum risk
- Combination of FT model with the Payload under 6000 kg is effective
- Decision Tree model has the greatest accuracy out of the 4 models.

Appendix

- I spent so much time on this project, please enjoy.

Thank you!