IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Harshavardhan
26/04/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection using web scrapping and SpaceX APIs
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics.
  - Prediction using Machine Learning models.
- Summary of all results
  - It was possible to collect valuable data from public sources.
  - EDA allows us to identify features that are best to predict success of landing.
  - Using machine learning models, we can predict the success rate with more than 80% accuracy.

Introduction

- Project background and context
  - The objective is to predict the success rate of launches for new company spaceY to compete with spaceX
- Problems you want to find answers
  - The best way to estimate the total cost for launches, by predicting the success rate of first stage rockets landing
  - Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is obtained from 2 sources:
    - Space X API(https://api.spacexdata.com/v4/rockets/)
    - Web scraping the Wikipedia page.
- Perform data wrangling
  - Data wrangling and cleaning was done using pandas and numpy libraries
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

- Perform predictive analysis using classification models
    - Data was normalized, divided into training and testing sets and evaluated by 4 different classification models.

Data Collection

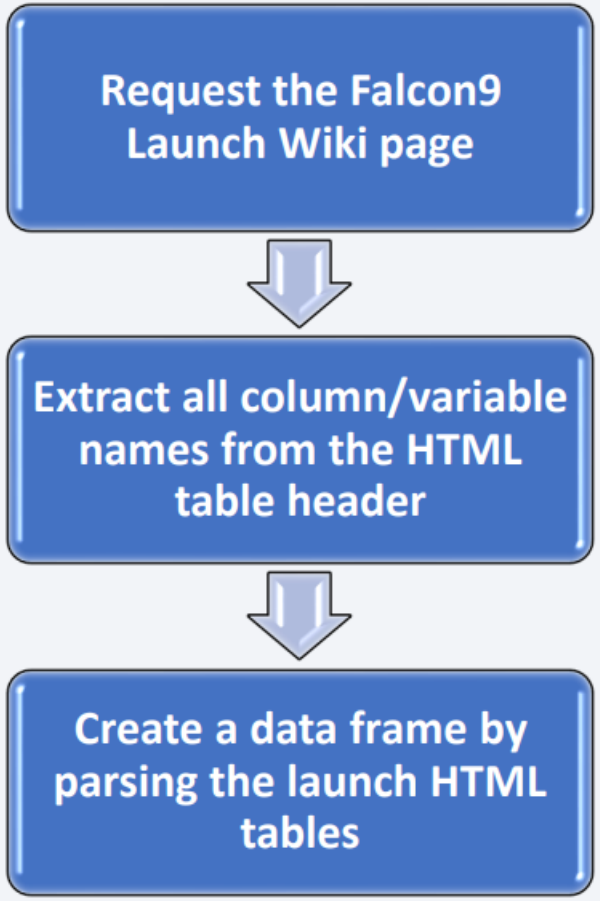
- Describe how data sets were collected.
  - Data is obtained from 2 sources:
    - Space X API(https://api.spacexdata.com/v4/rockets/)
    - Web scraping the Wikipedia page.

Data Collection – SpaceX API

- Public API can be used to get the data.
- The data can be loaded according to the flow chart.
- Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

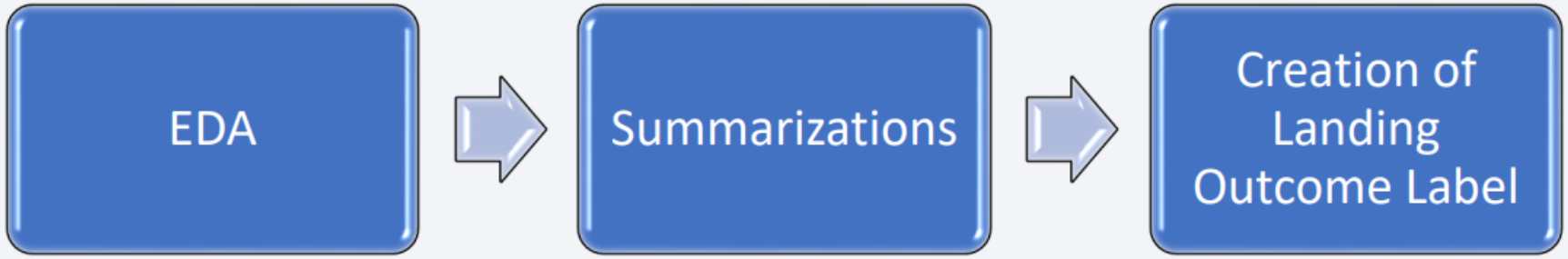
Data Collection - Scraping

- Wikipedia contains the launch data, web scraping can be used to get the data.
- Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

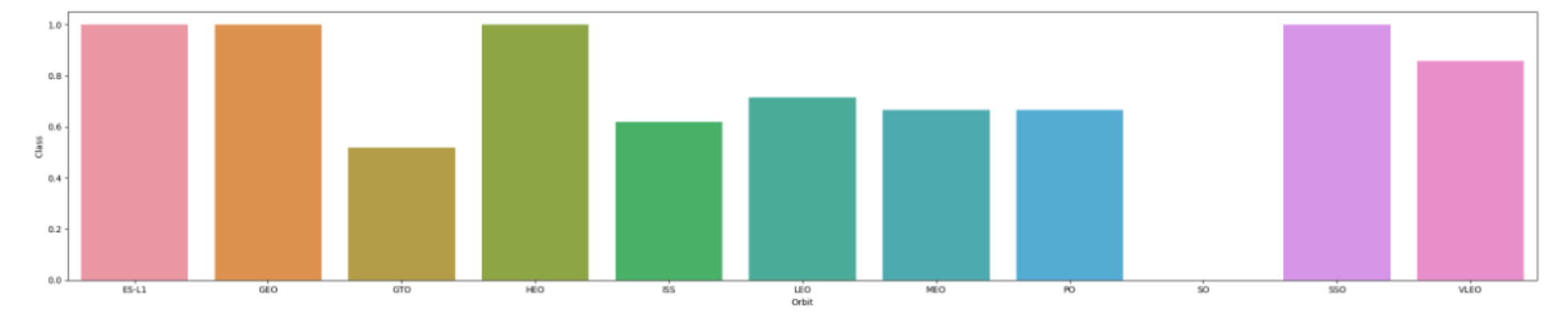
- Some Exploratory data analysis is performed on the data.
- Summaries of launch site, success rate of mission outcome per orbit are calculated.
- Finally landing outcome column was created.



Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/jupyter-labs-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Scatter and bar plots were used to visualize the data to establish relationship between features
- Below image is success rate w.r.t orbit.



Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- The following tasks were performed using SQL:
  - Display launch site names
  - Top 5 launch sites whose names begins with 'CCA'
  - Display the total pay load mass carried by boosters launched by NASA
  - Display average pay load mass launched by booster version F9 v1.1
  - List the date when the first successful landing outcome in the ground pad was performed
  - List the names of boosters which have success in drone ship and have pay load mass between 4000 and 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions that carried maximum payload mass
  - List the records which display the month names, failure landing outcomes in drone ship, booster version, launch site for the months in 2015
  - Count of successful landing outcomes between 04-06-2010 and 20-03-2017 in descending order

Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.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 highlight the area of interest
  - Marker cluster group if events like launches in launch sites
  - Lines are used to indicate distances between two points.

Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/Jupyter-labs-analysis_with_folium.ipynb

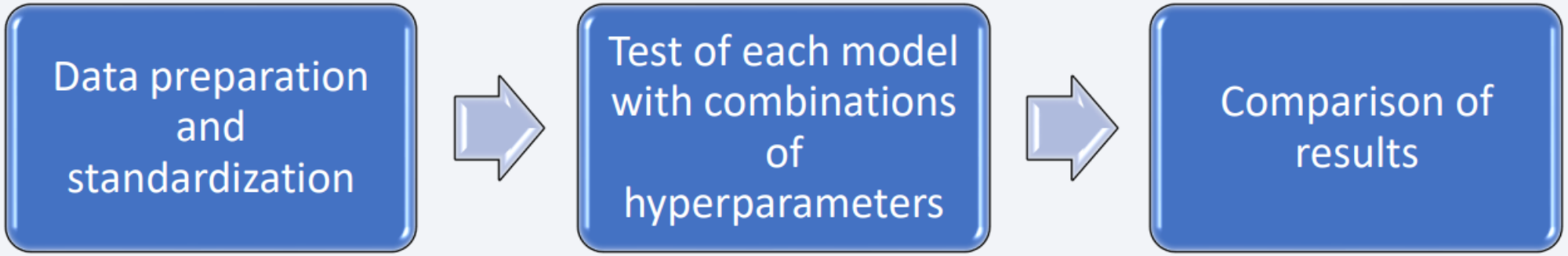
Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allows us to quickly analyze the relationship between payloads and launch sites, helping to identify best places for the launch according to payload.

Code: https://github.com/harshalingala/datascience_capstone_project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were evaluated: logistical regression, support vector machine, decision tree and k nearest neighbors.



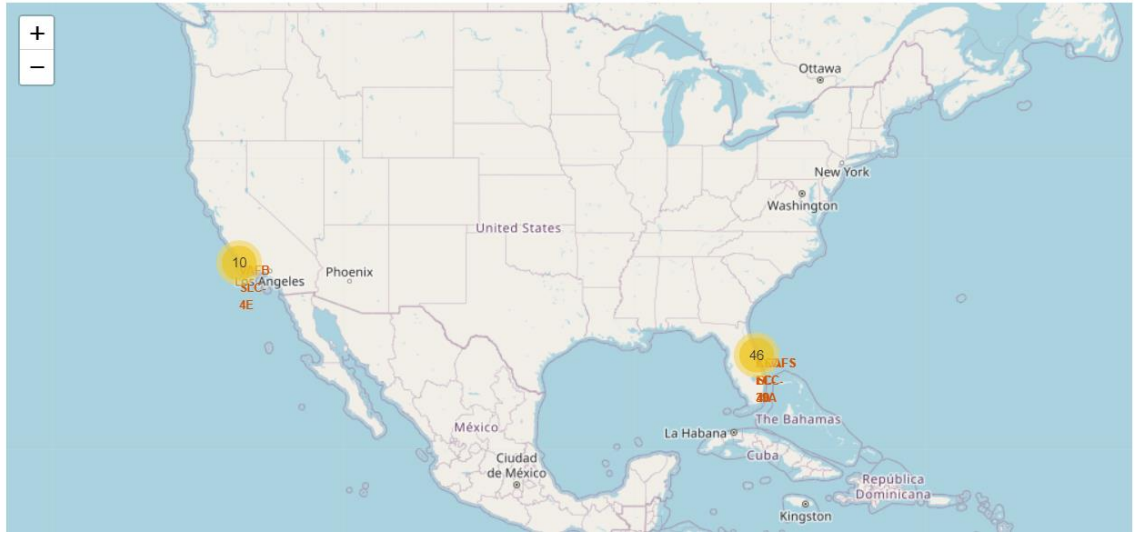
Code:https://github.com/harshalingala/datascience_capstone_project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

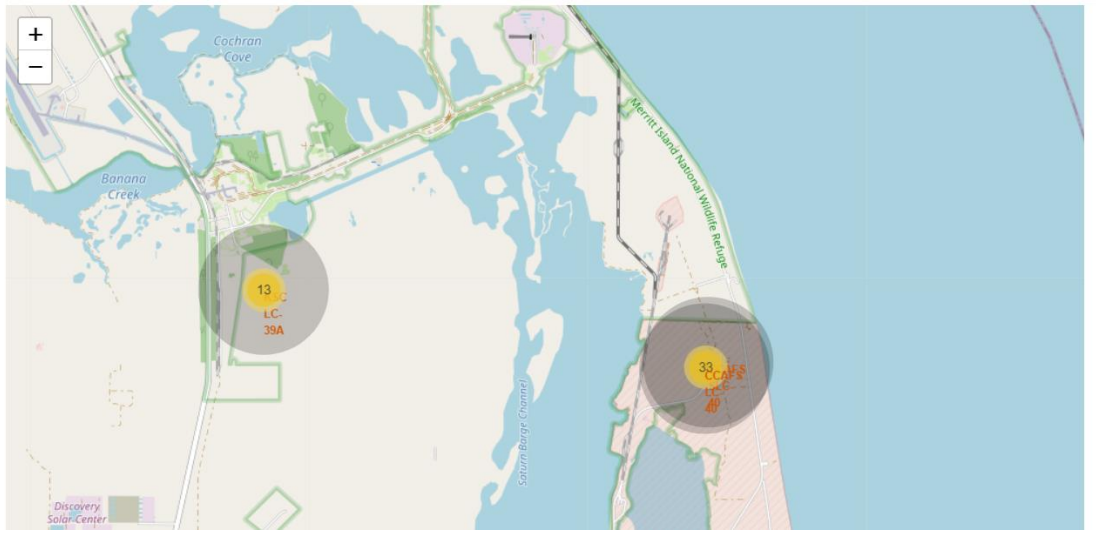
Results

- Exploratory data analysis results
  - spaceX uses 4 different launch sites
  - Average load for F9 v1.1 booster is 2928KG
  - Many falcon 9 landing were successful at landing in drone ships having pay load above 4000KG
  - There are 2 failures in 2015
  - The success rate improved over years

Results

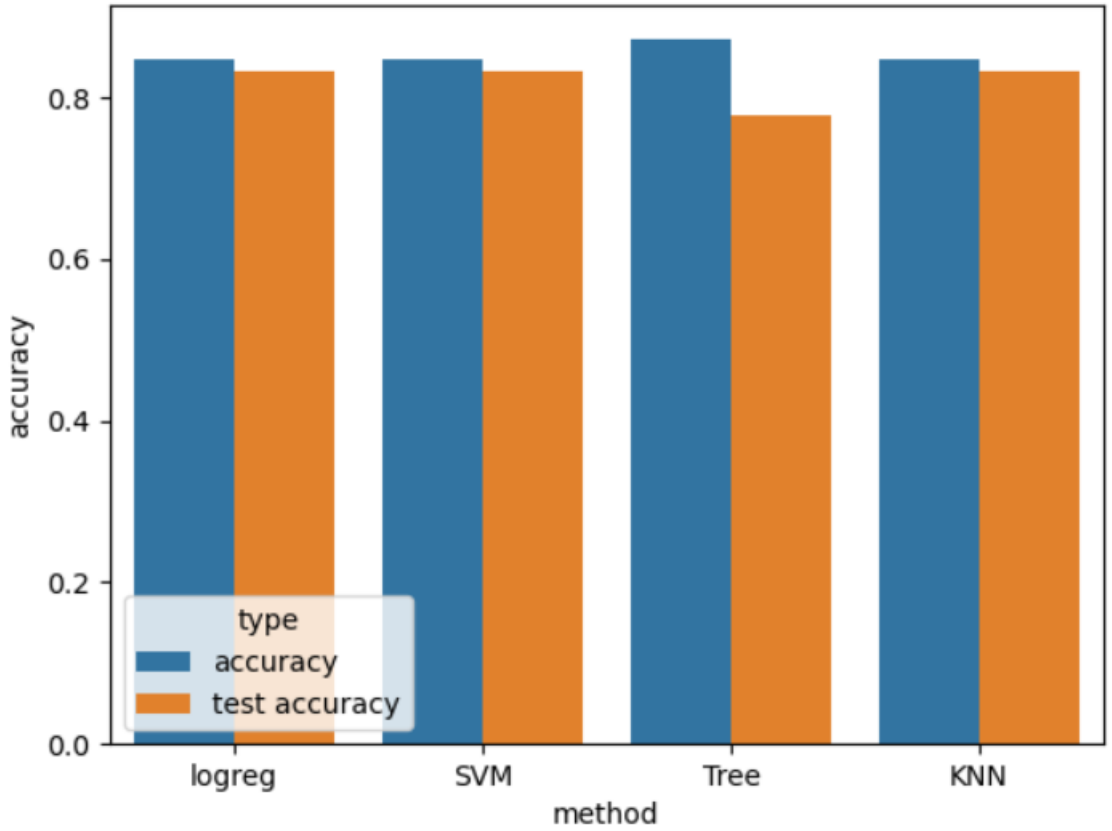
- Using interactive analytics, it was possible to identify that the launch sites are in safe places, near sea and have good infrastructure around
- Most launches happen at east coast

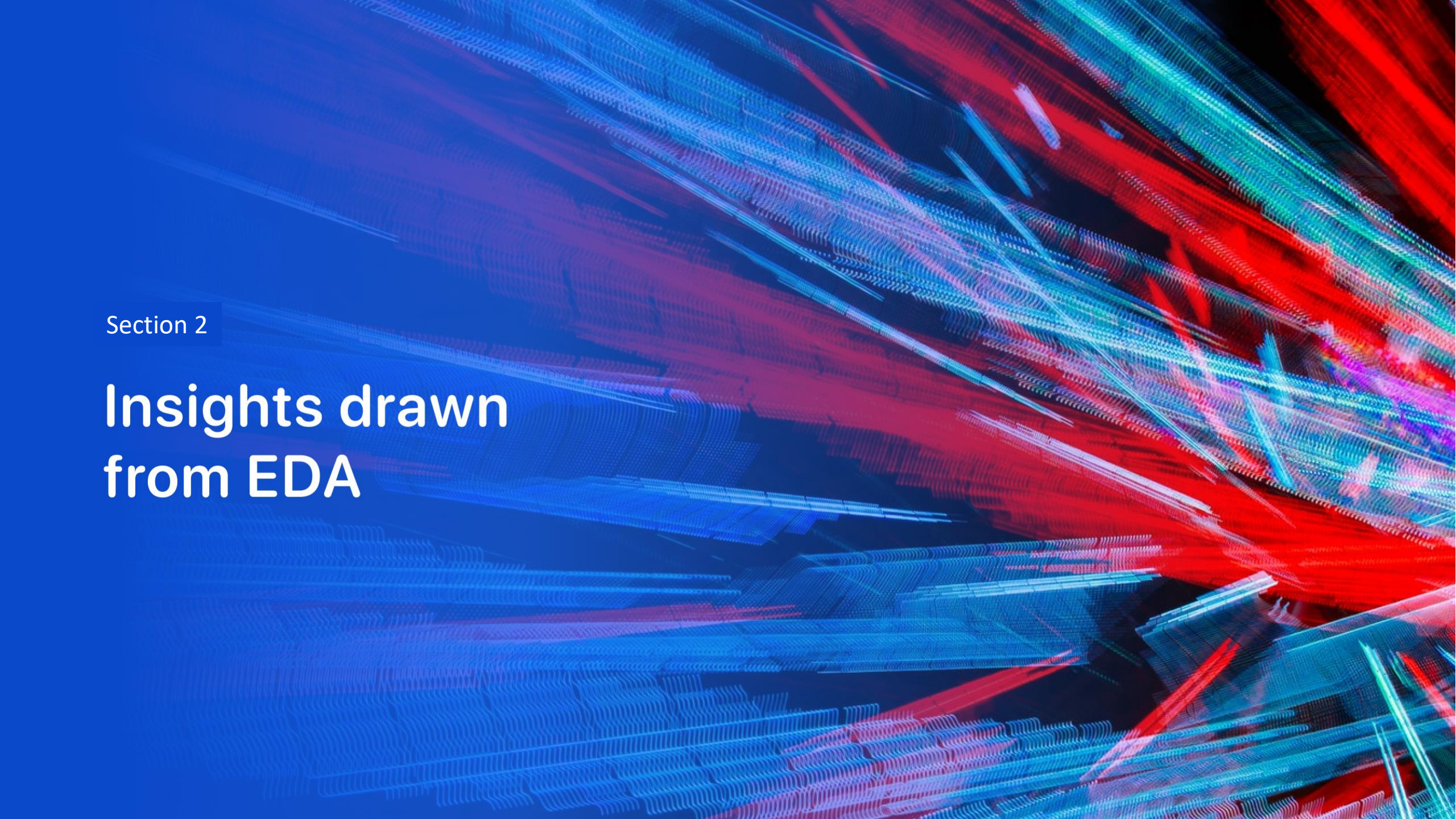




Results

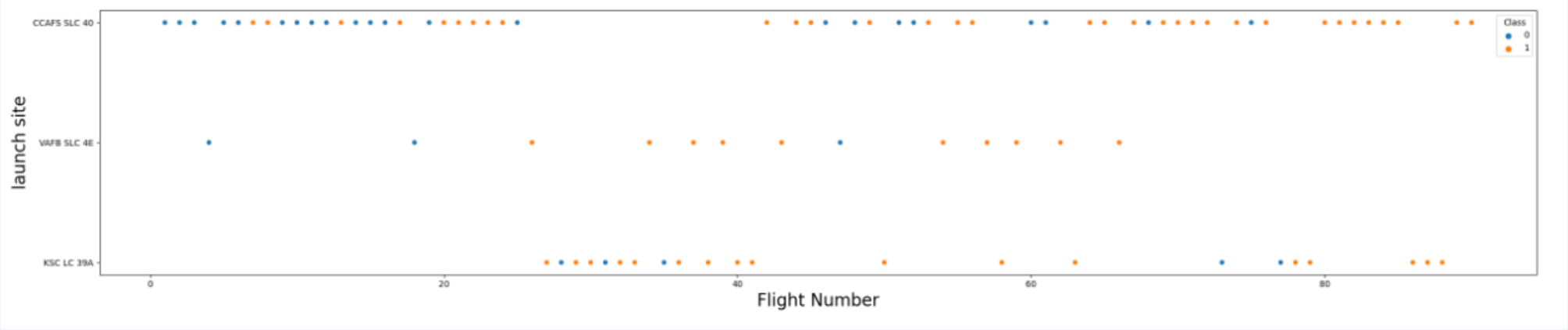
- Predictive analysis show that all models are closer in their performance.
- Tree classification have higher training accuracy and lower test accuracy suggesting a overfitting.




Section 2
Insights drawn from EDA

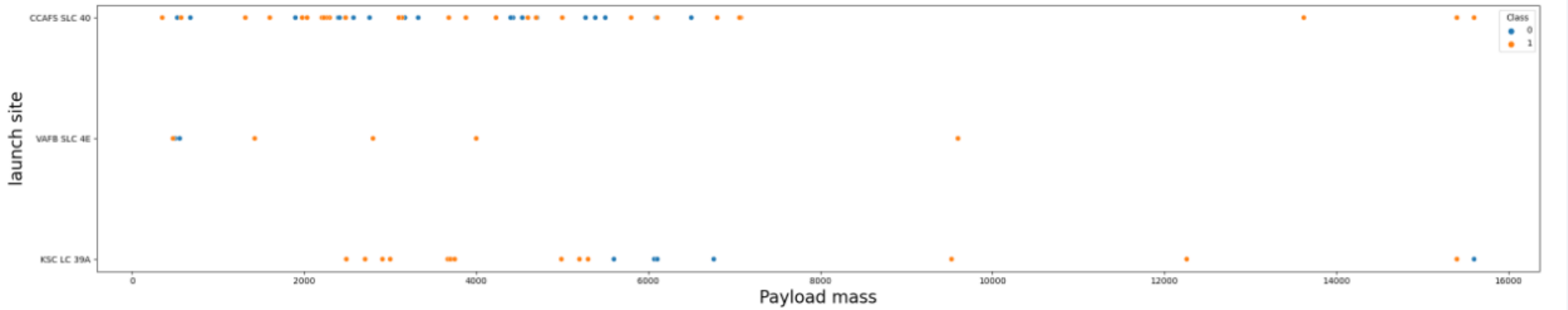
Flight Number vs. Launch Site

- It is possible to notice CCAF5 SLC 40 is best launch site with most of recent launches being successful.
- It is possible to see increase in success rate with time.



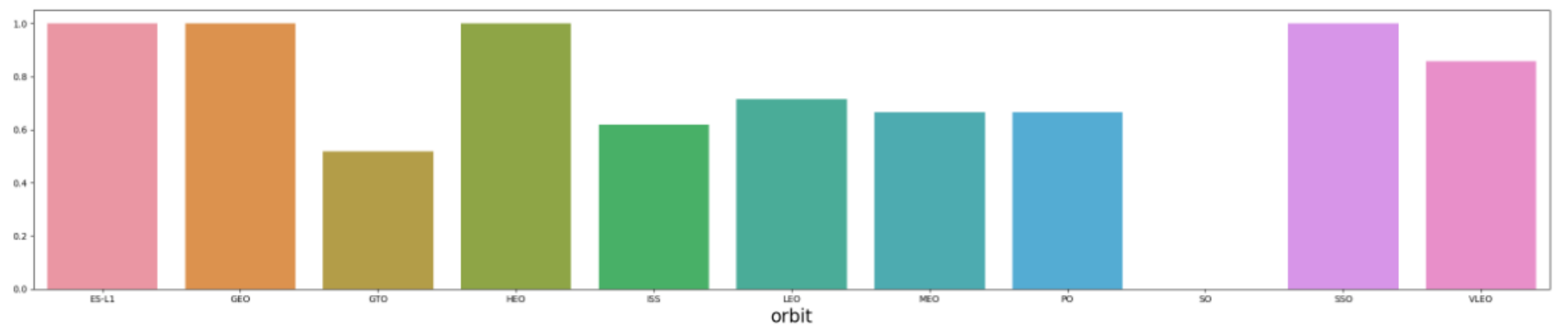
Payload vs. Launch Site

- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.



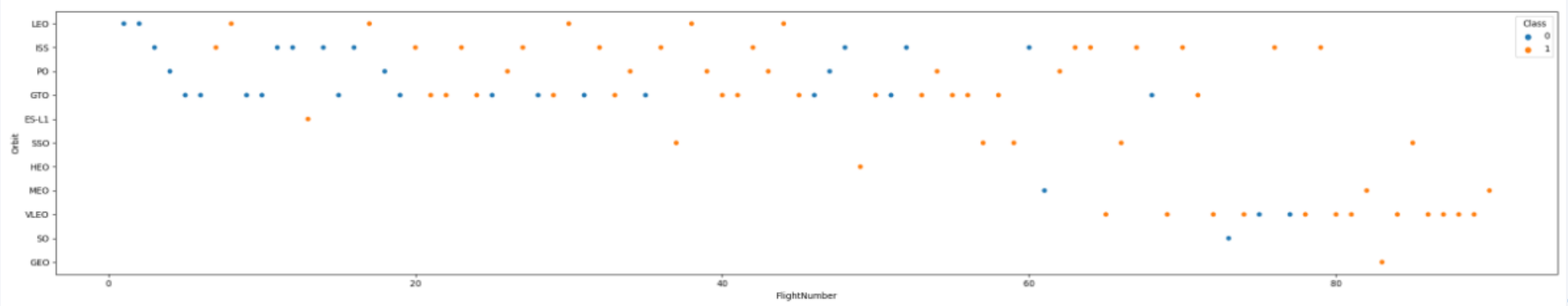
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO orbits have biggest success rate.



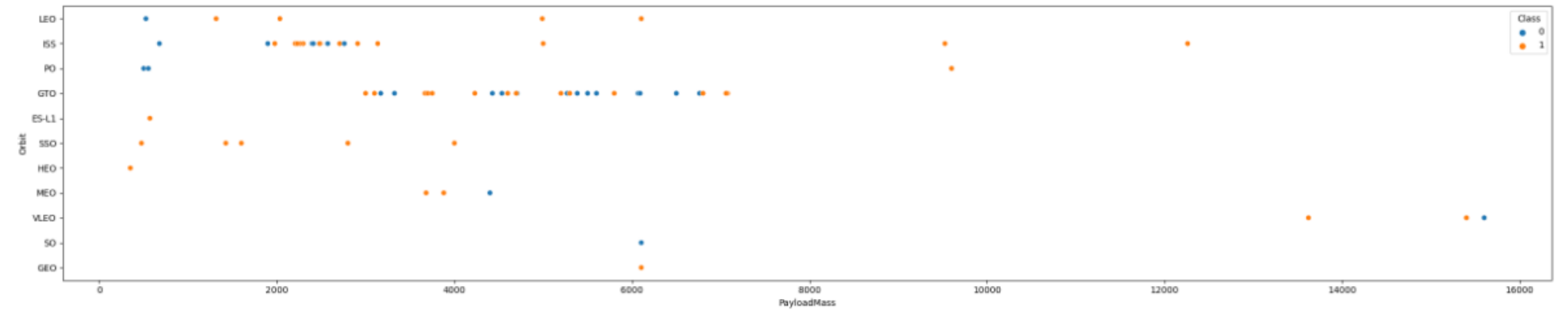
Flight Number vs. Orbit Type

- Success rate improves for all orbits over time.



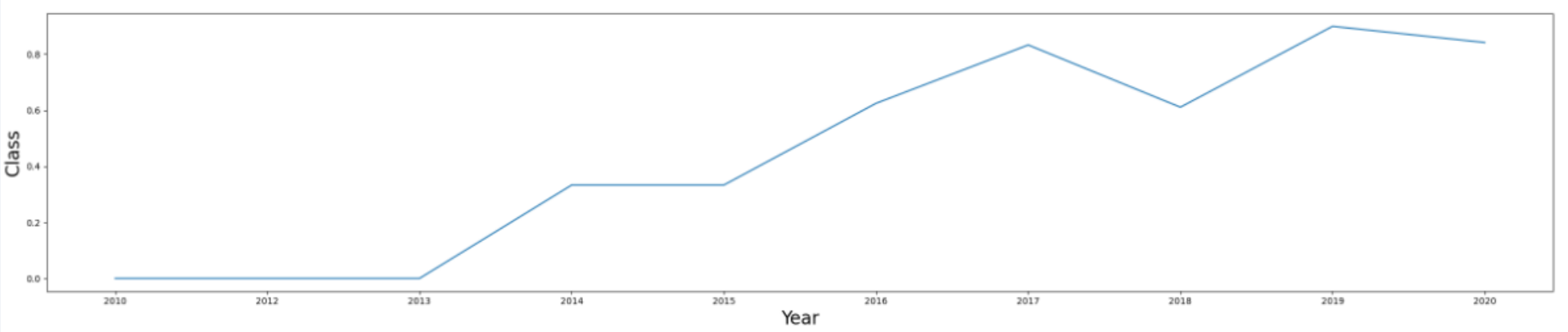
Payload vs. Orbit Type

- The launches for SO and GEO are low.
- For GTO, no good correlation between pay load mass and success rate.



Launch Success Yearly Trend

- Success rate shows improvement over the years.



All Launch Site Names

- Following are the 4 launch sites

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

Launch Site Names Begin with 'CCA'

- The query was able to complete using LIKE operator.
- Below are the 5 records for launch sites beginning with ‘CCA’

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total pay load mass for NASA is more than 45000 KGs
-

SUM(PAYLOAD_MASS__KG_)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- Average pay load mass for F9 v1.1 is ~2900KGs

AVG(PAYLOAD_MASS__KG_)	Booster_Version
2928.4	F9 v1.1

First Successful Ground Landing Date

- First successful landing outcome on ground pad was in 2015

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Here are the boosters the successfully landed on drone ship with pay load between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

- Here is the mission success and fail report

COUNT(*)	Mission_Outcome
1	Failure (in flight)
98	Success
1	Success
1	Success (payload status unclear

Boosters Carried Maximum Payload

- Here are the list of booster version that carried maximum load

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

- List of failures in year 2015

month	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	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

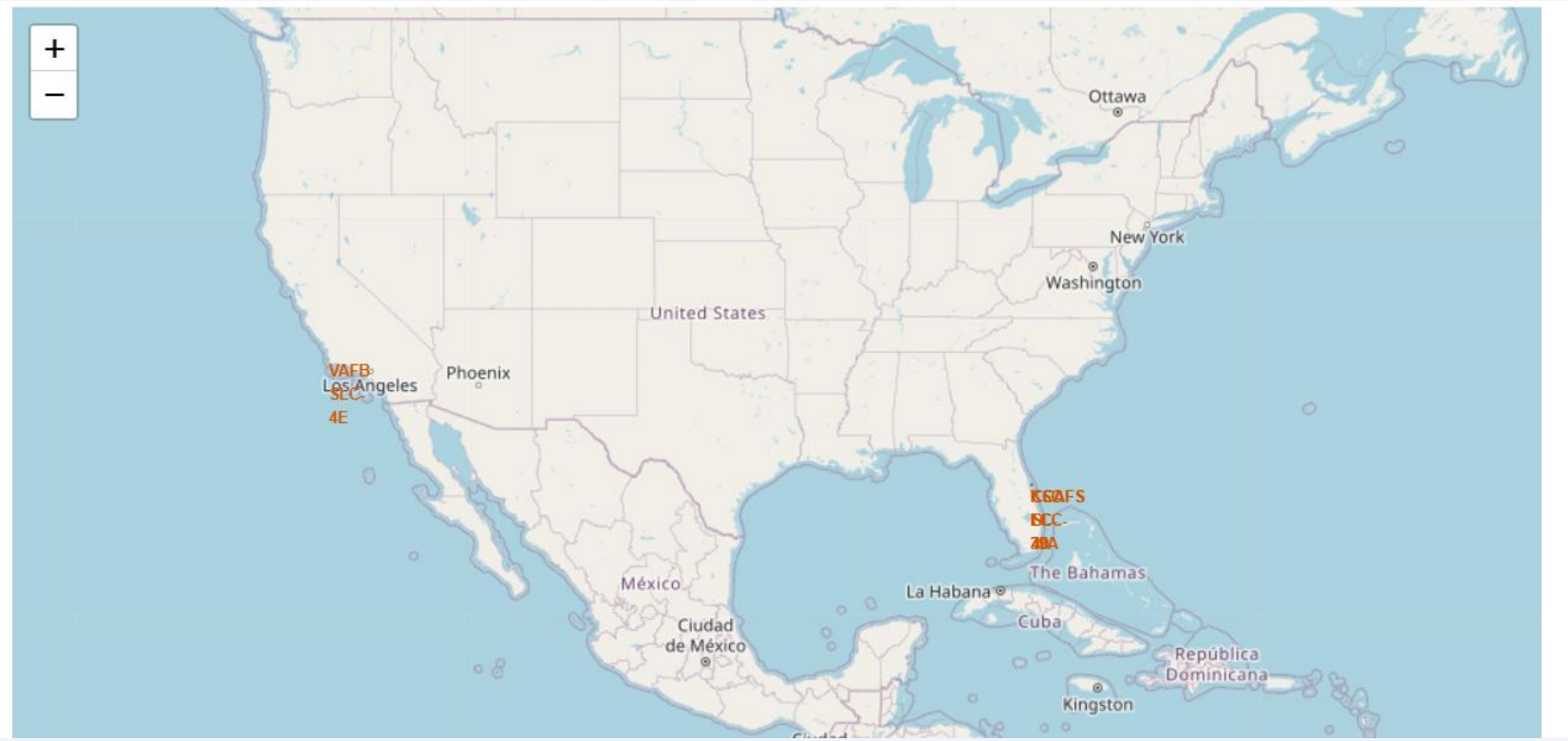
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

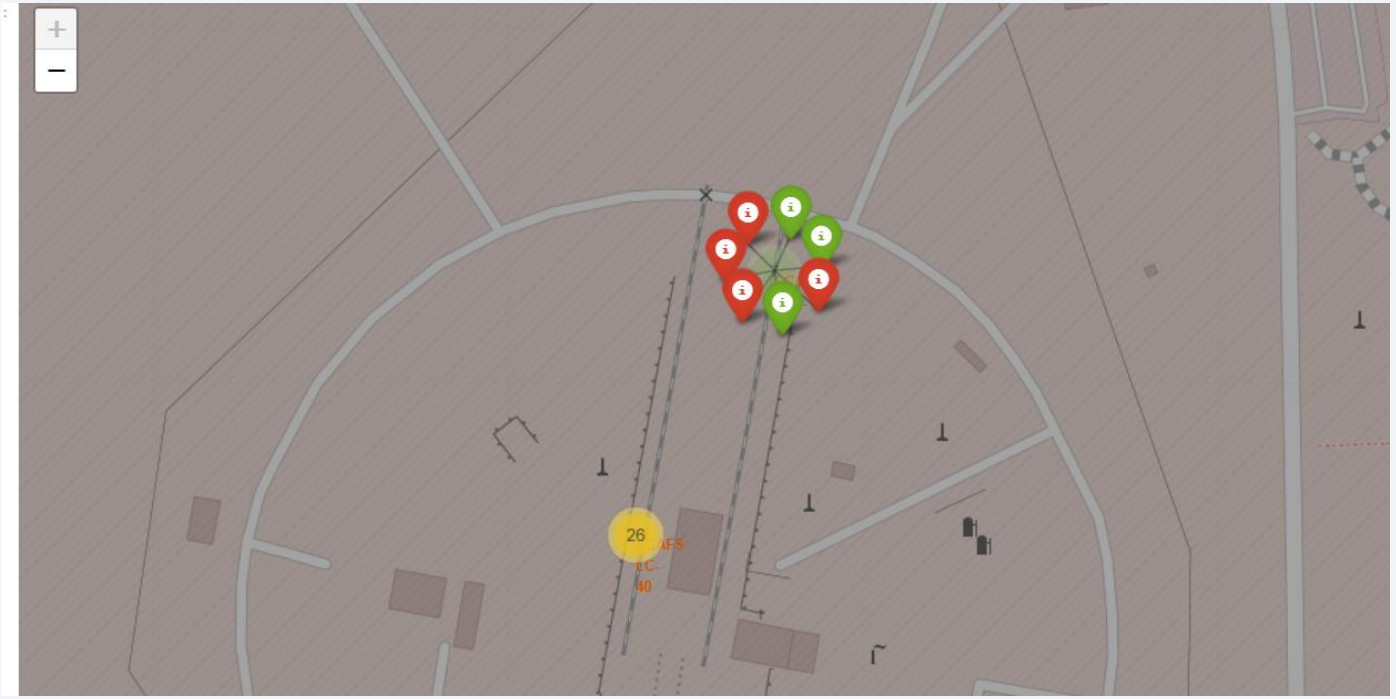
All Launch Sites

- all the launch sites are near coastline.



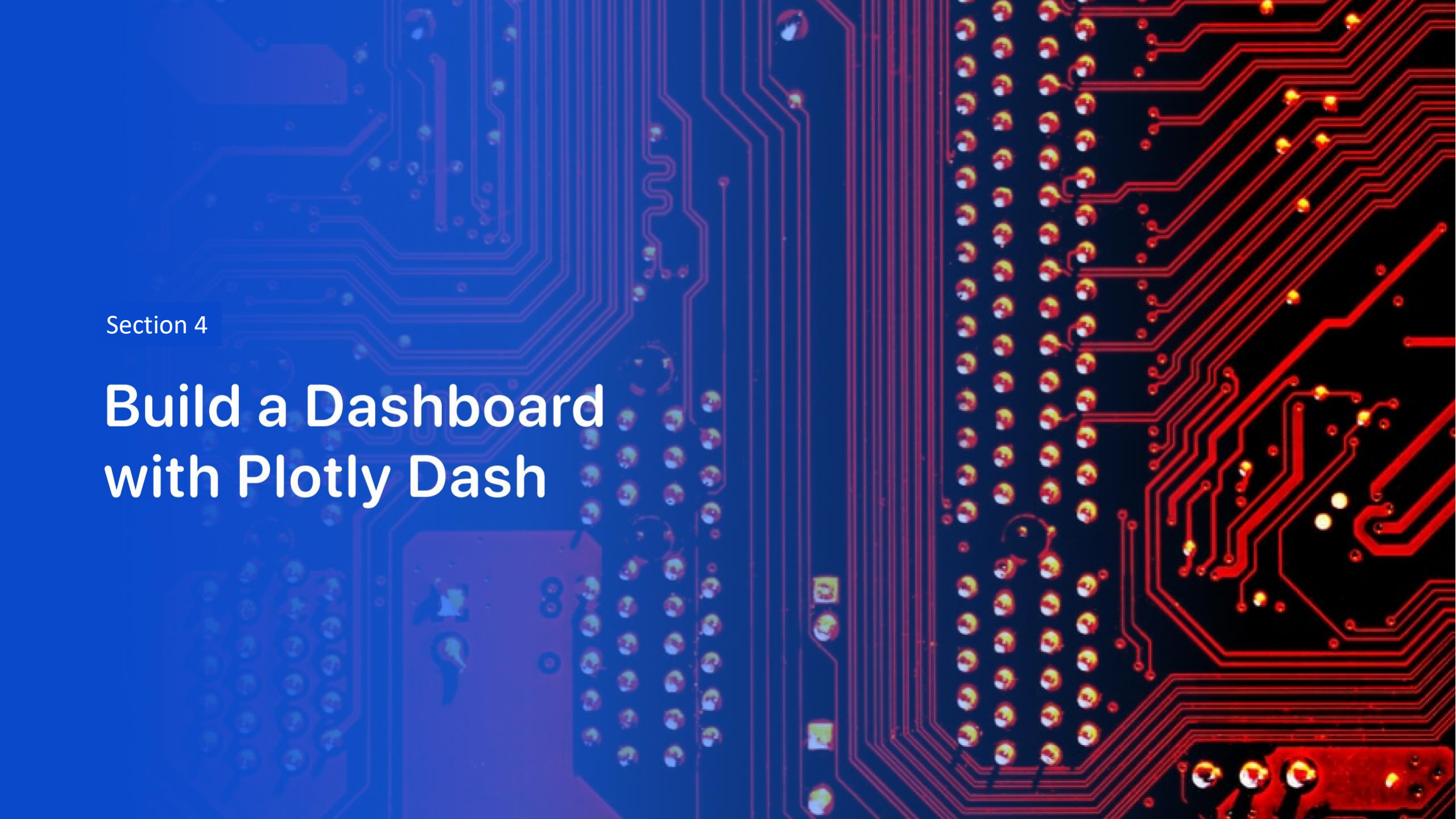
Launch outcomes

- Green indicates success and red indicates fail



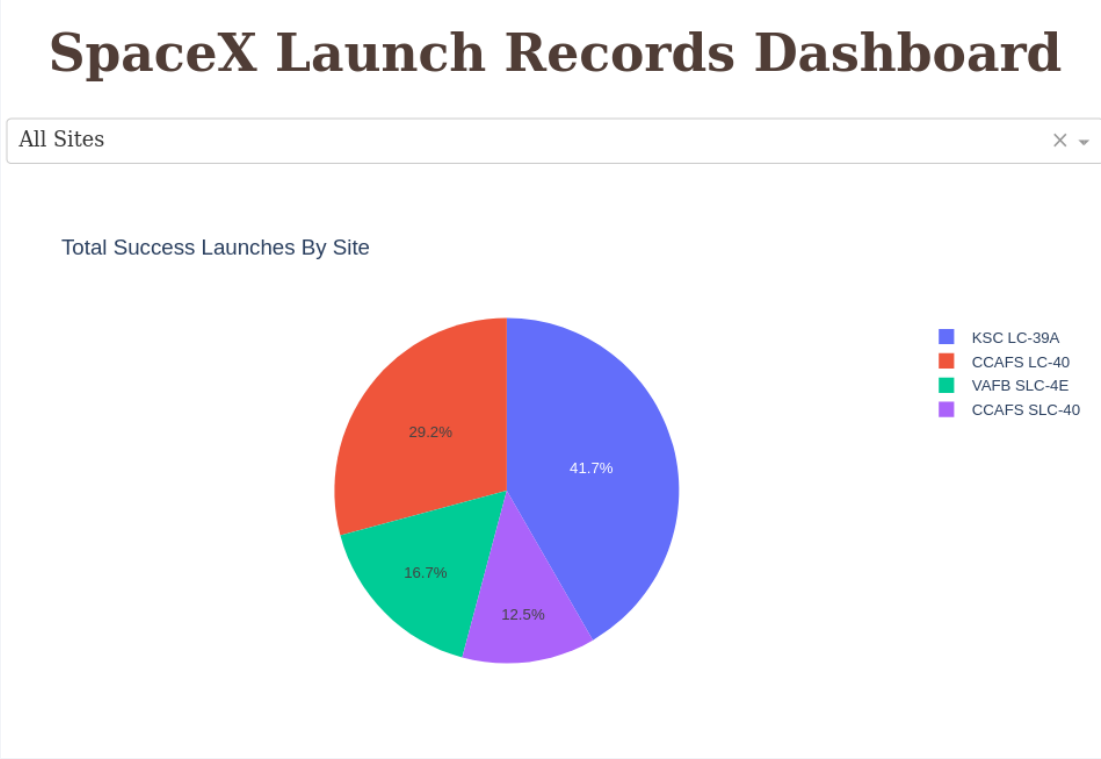
Lines and distances

- Lines can be added to the map and coordinates can be used to get the distance.


Section 4
Build a Dashboard with Plotly Dash

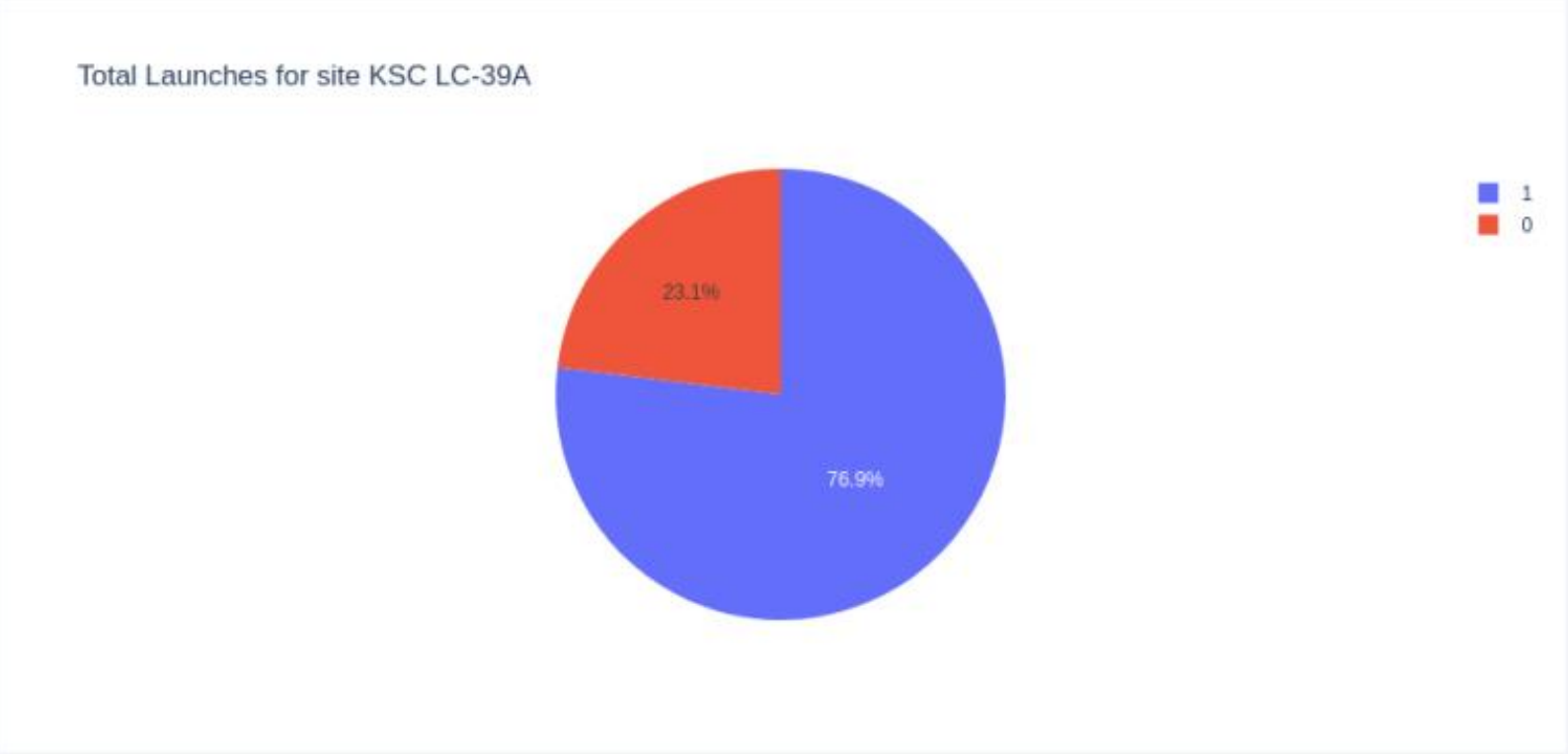
All Sites

- By selecting all sites options, we can show all sites.



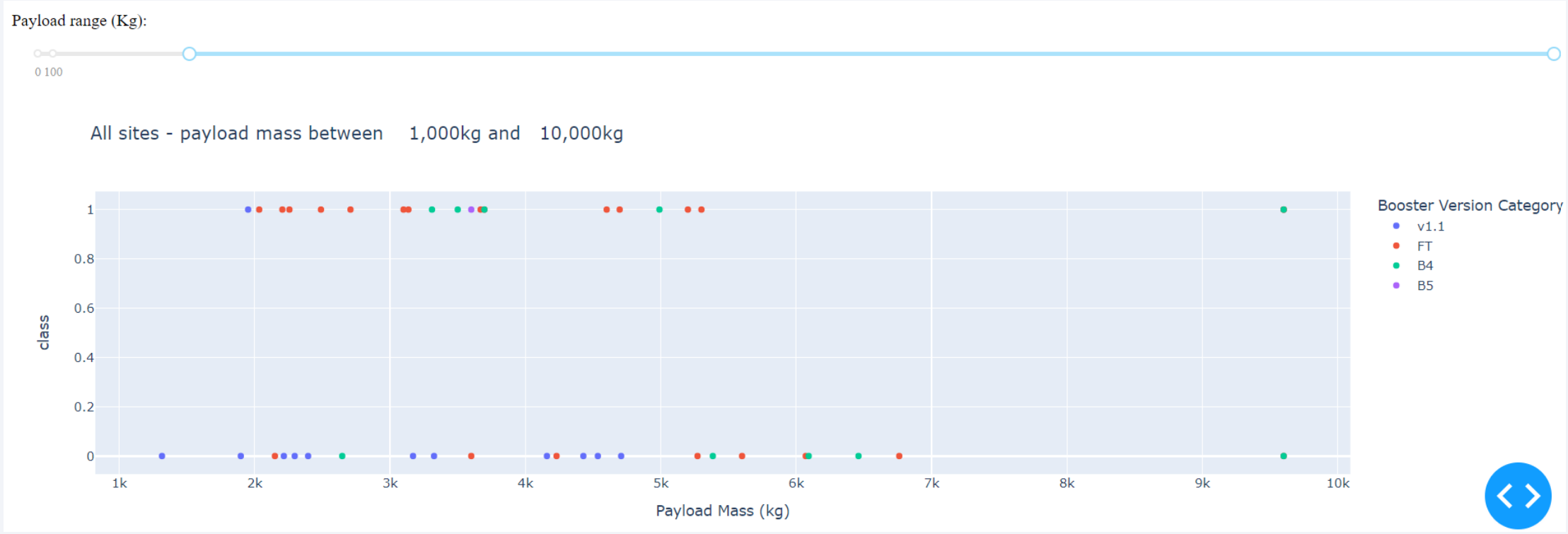
Launch Success Ratio:

- Site KSC LC-39A have highest success rate with 76.9%



Pay load and launch outcome

- Payload above 6000KG are not very successful.
- Payload below 6000KG and FT booster are very successful

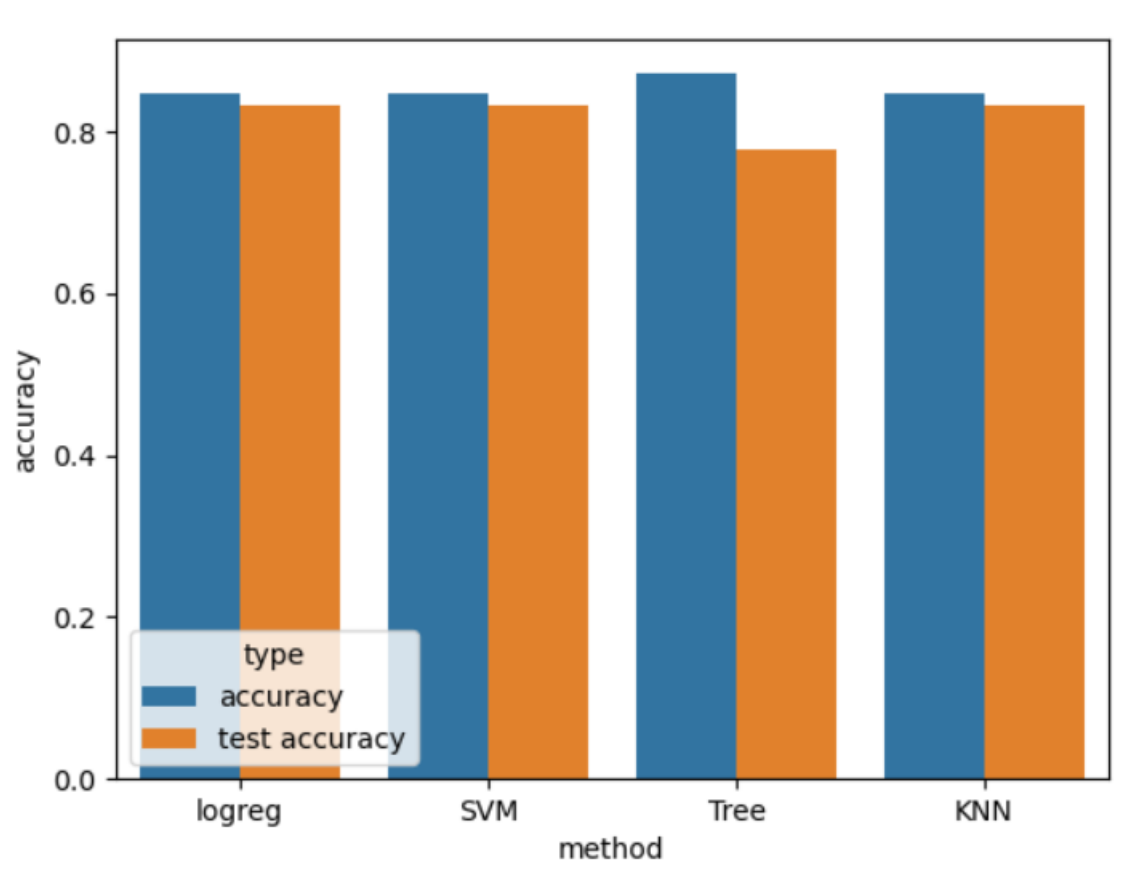


Section 5

Predictive Analysis (Classification)

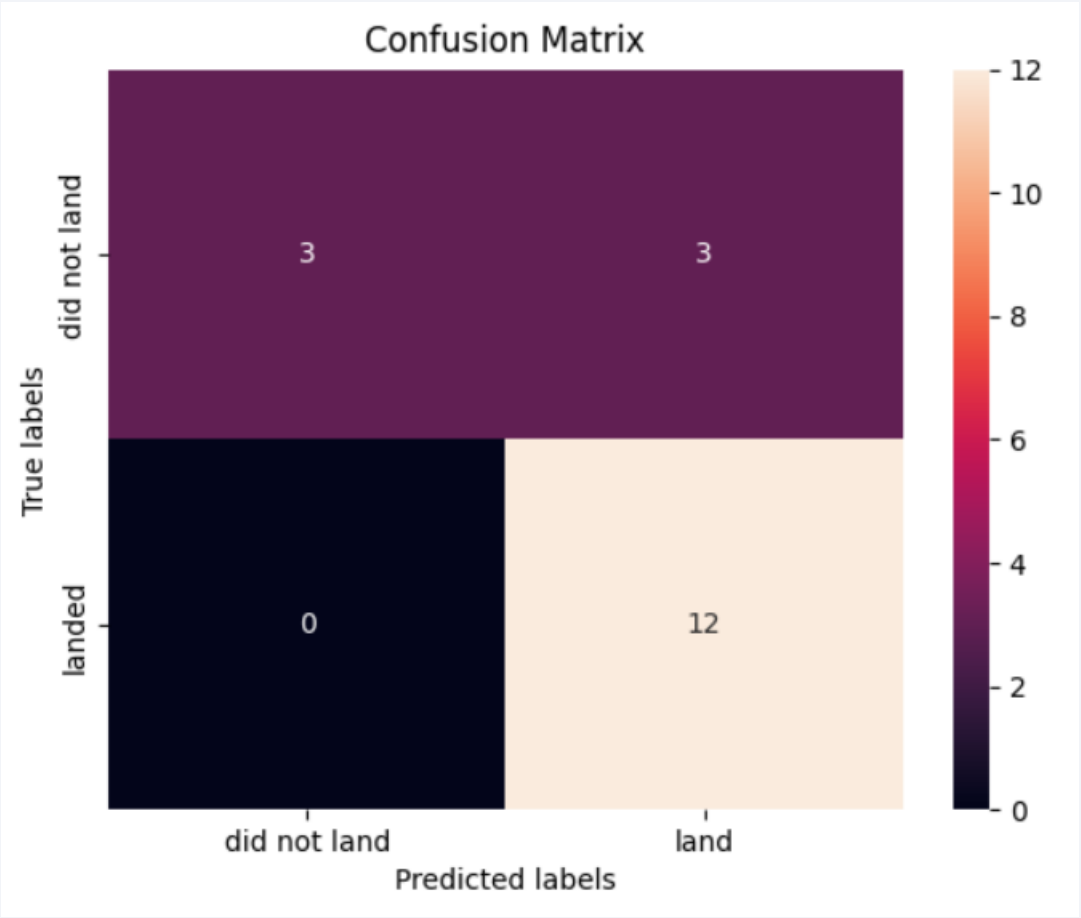
Classification Accuracy

- Here are the accuracies for 4 classification models used.
- Performance of all models are close.
- Decision tree classification may have overfitting as training accuracy is much better that test accuracy.



Confusion Matrix

- Here's the confusion matrix of logistic regression model, it was able to predict all landed cases successfully.
- For 'did not land' cases, it was able to predict only half of it.



Conclusions

- Data from different data sources were analyzed, refining conclusions along the process.
- The best launch site is KSC LC-39A with 76.9% success rate.
- Launches with below 6000KG payload and FT booster are highly successful.
- Most of the mission outcomes are successful.
- Landing outcome improved significantly over years.

Appendix

- All the codes are present in below github repository: https://github.com/harshalingala/datascience_capstone_project

Thank you!