IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chandima Dabare
08 Apr 2023

Outline

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

Executive Summary

- Summary of methodologies
- In this project the Data Collection was done through API with Web Scrapping.
- The data issues were handled and made a usable data set through Data Wrangling Techniques.
- Exploratory Data Analysis was performed with SQL and necessary Data Visualizations along with Interactive Visual Analytics with Folium.
- Several Machine Learning techniques were used for the prediction.

- Summary of all results
- During the analysis process, we found the best launch site, the best weight range of the launches.
- Due to the improvements the mission are successful based on the successful landing outcomes.
- The predictions based on the Machine Learning indicates the best model with successful landing and optimal resources and profits.

Introduction

- SpaceX is a company with evolutionary ideas offering low-cost rocket launches. Mainly reusing the launches for re-landing on the next mission is the main concept of cost savings.
- This project is crucial in identifying the right price to bid against SpaceX for a rocket launch.
- The problems included:
  - Identifying the significant factors influencing the successful landing.
  - The co-relation among the features and significant features.
  - Estimate the total cost for launches, by predicting successful landings of the first stage of rockets;

Section 1

Methodology

Methodology

Executive Summary

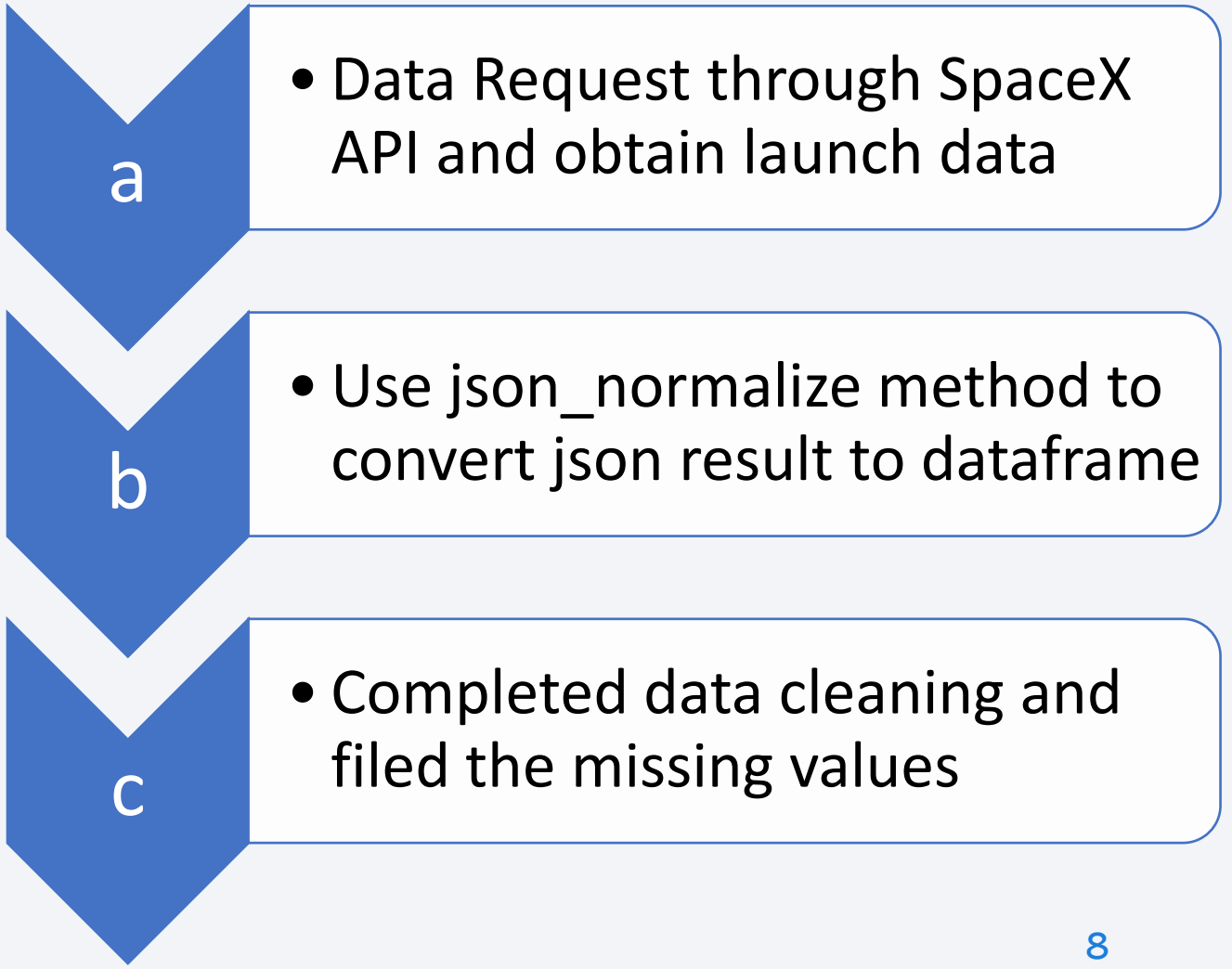
- The data was collected SpaceX REST API and web scrapping from Wikipedia.
  Space X API (https://api.spacexdata.com/v4/rockets/) & WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- The data was enriched with landing outcome and performed data wrangling based on labels on outcome data after summarizing and analyzing respective features.
- The interactive visual analytics was included using Folium and Plotly & Dash
- The models were created based on classification methodologies.
- Data that was segregated in to training and test data sets and trained the classification models with training data and tested with test data and found the model with best fit using serveral evaluation methods (Confusion Matrix)

Data Collection

- The dataset was collected through API, and normalization was done. After this process the data was cleansed and checked for the missing values and filled the missing values using the necessary methodologies.
- Data Collection Space X API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics

Data Collection – SpaceX API

- SpaceX offers a public API and the data can be extracted for usage.
- The methodology used is represented in the flowchart,



- Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/1)%20Data_Collection_SpaceX.ipynb

Data Collection - Scraping

- Data pertaining to SpaceX launches can be obtained from Wikipedia.

a
- Request the Falcon9 Launch Wiki page

b
- Extract all column/variable names from the HTML table header

c
- Create a data frame by parsing the launch HTML table

Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/2)%20Data%20Collection%20with%20Web%20Scraping%20SpaceX.ipynb

Data Wrangling

a
- Data Cleansing & Exploratory Data Analysis (EDA)

b
- Data was summarized

c
- Creation of a new data element (Landing out come)

- In Data Wrangling the data was cleansed performed Exploratory Data Analysis (EDA)

Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/3)%20Data%20Wrangling%20SpaceX.ipynb

EDA with SQL

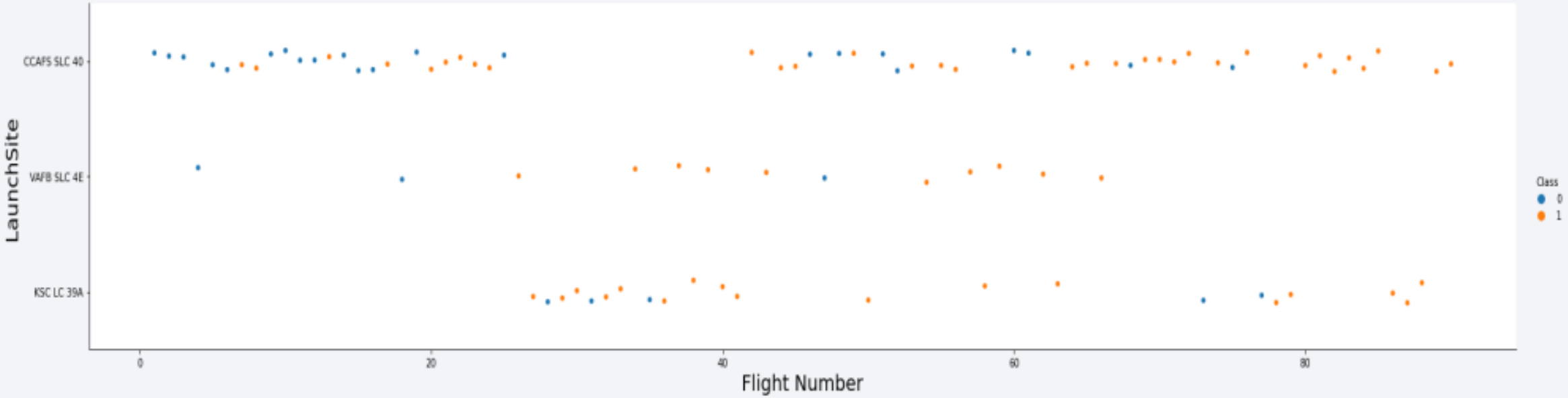
Following SQL queries were prepared…..

- 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: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/5)%20EDA%20SpaceX.ipynb

EDA with Data Visualization

- The below chart demonstrates relationship between the Flight Number vs Launch Site.



Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/4)%20EDA%20with%20Data%20Visualization%20SpeaceX.ipynb

Build an Interactive Map with Folium

- The latitude and longitude coordinates at each launch site and added a circle marker around each launch site with a label of the name of the launch site in order to visualize the launch data into an interactive map.
- Assigned the dataframe launch_outcomes (failure,success) to classes 0 and 1 with
- Red and Green markers on the map in MarkerCluster().
- The Haversine's formula is used to calculated the distance of the launch sites to Various landmark to find the solutions to the questions of:
    - How close the launch sites with railways, highways and coastlines?
    - How close the launch sites with nearby cities?

Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/6)%20Interactive%20Visual%20Analytics%20with%20Folium%20SpaceX.ipynb

Build a Dashboard with Plotly Dash

- An interactive dashboard was developed with Plotly dash.
- Dashboard includes;
  - Pie charts showing the total launches by a certain sites.
  - The scatter plots showing the relationship with Outcome and Payload

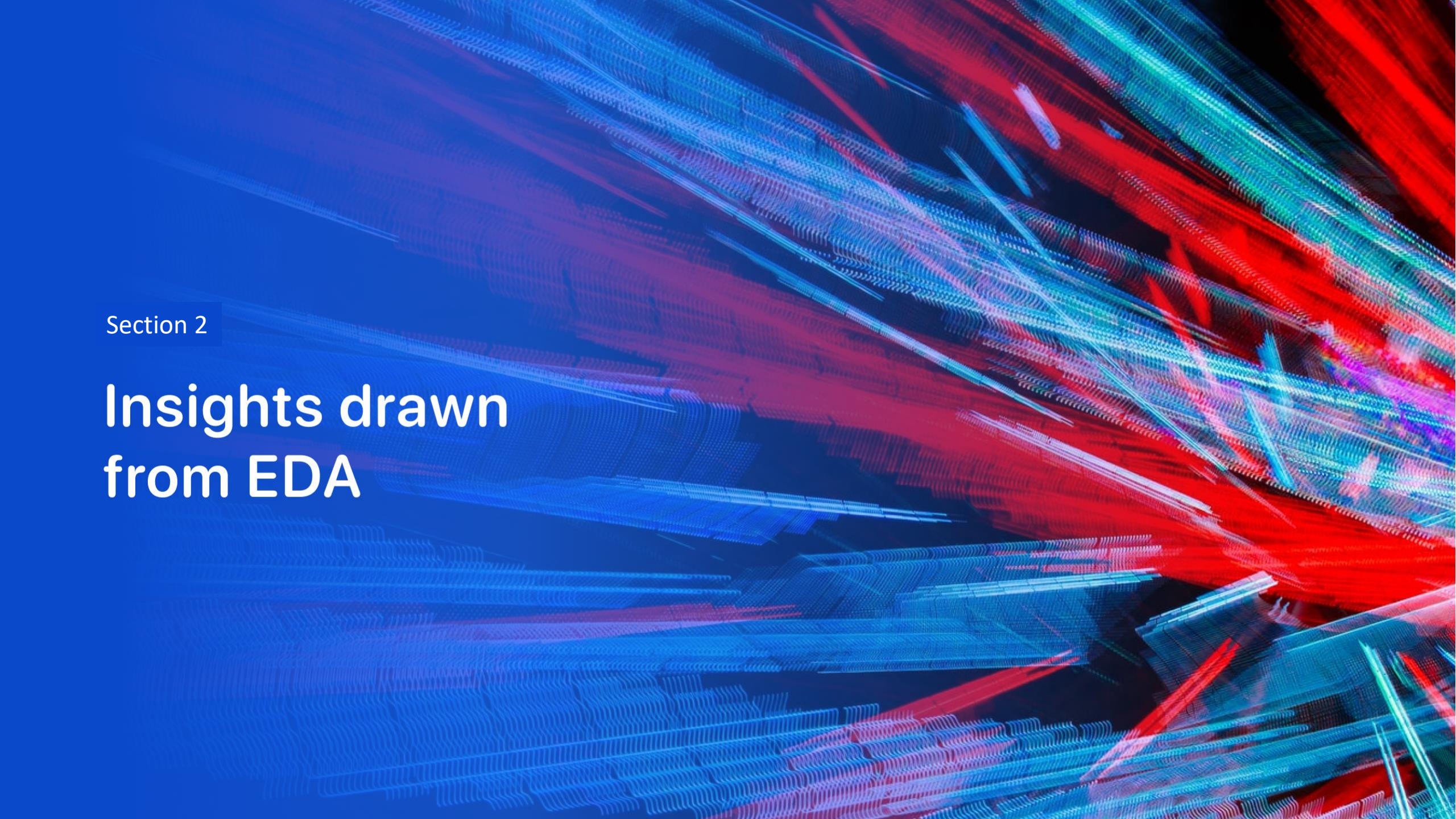
https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/7)%20dash_app_spacex.ipynb

Predictive Analytics (Classification)

Developed Four classification models and the model results were compared to find the best model;

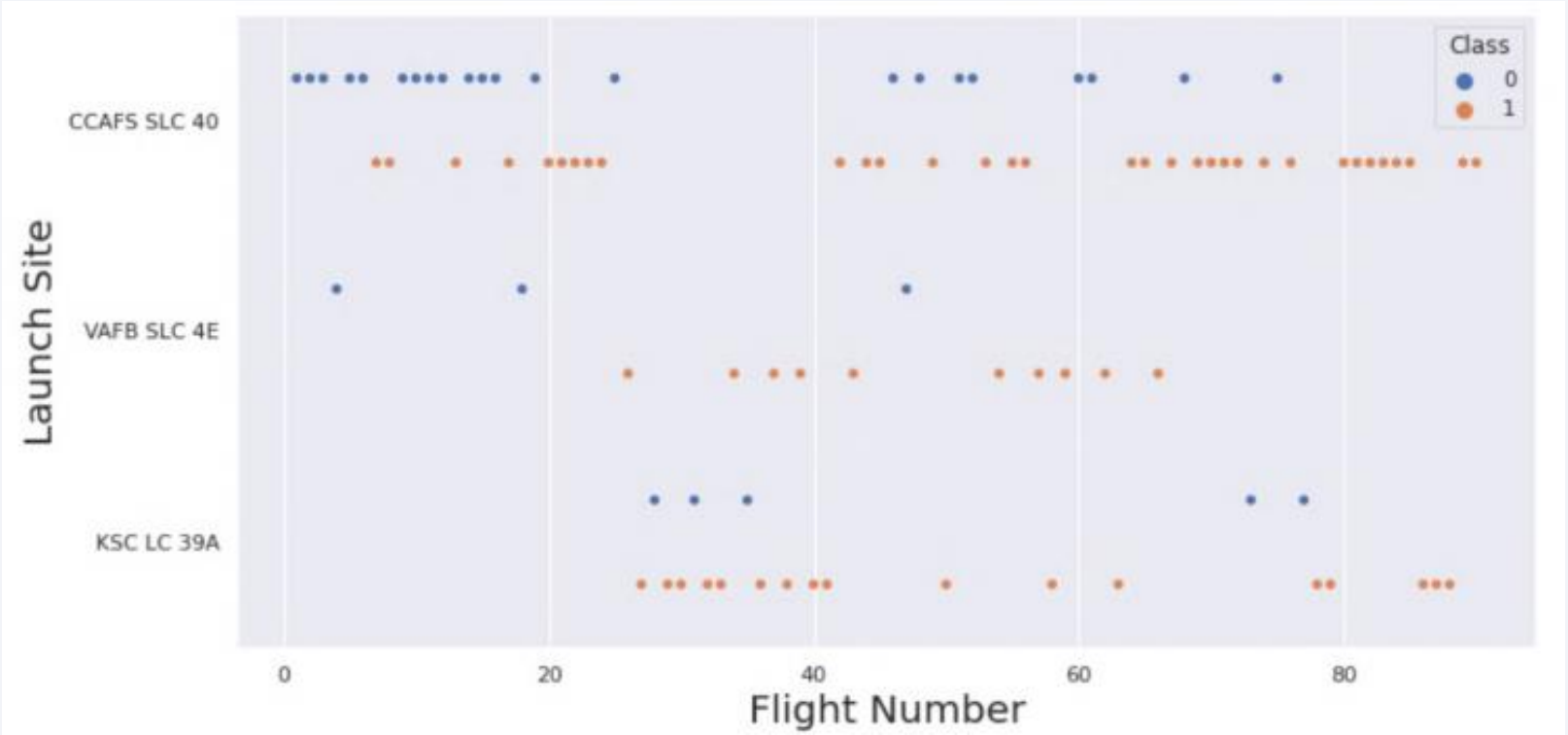
- Logistic regression
- Support vector machine (SVM)
- Decision tree
- Knn (K Nearest Neighbors)

Source: https://github.com/chandimadabare/DC_Capstone_SpaceX_CD/blob/main/8)%20Machine%20Learning%20Prediction%20Spacex.ipynb

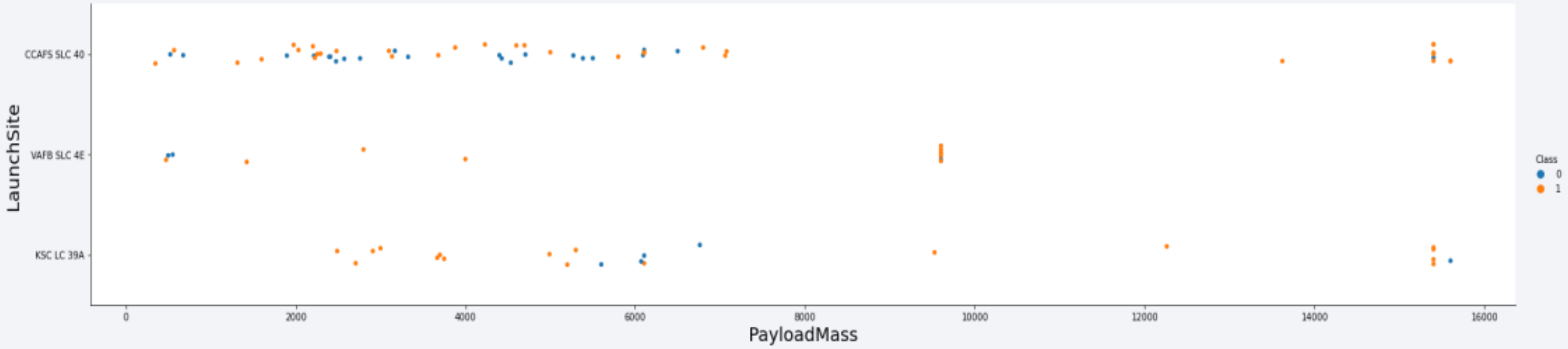

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- CCAF5 SLC 40 was the best & successful launch site.
- In second best site was VAFB SLC 4E and third best site was KSC LC 39A.

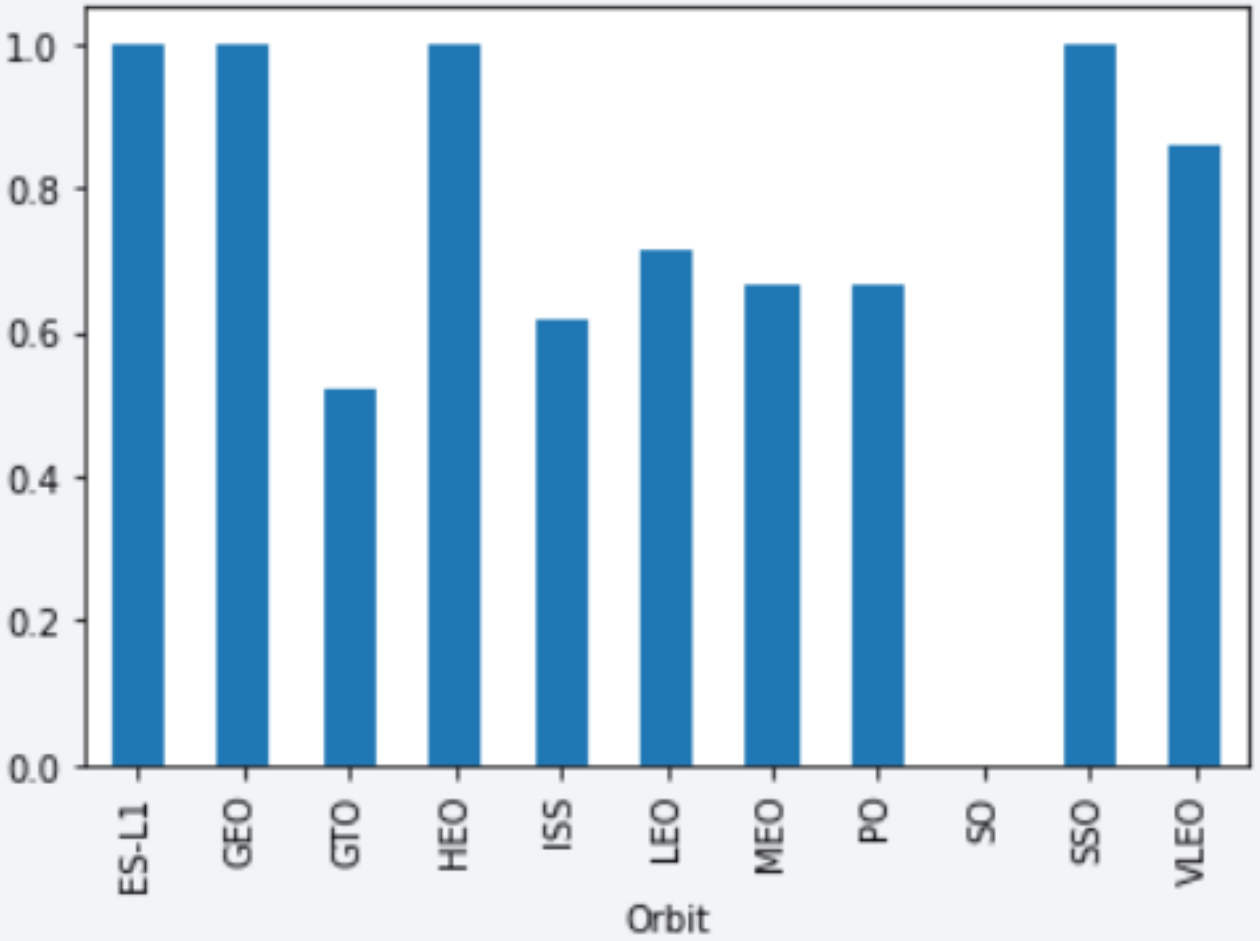


Payload vs. Launch Site



- Payloads over 9,000kg indicates excellent success rate;
- Payloads over 12,000kg indicates success on CCAFS SLC 40 and KSC LC 39A launch sites

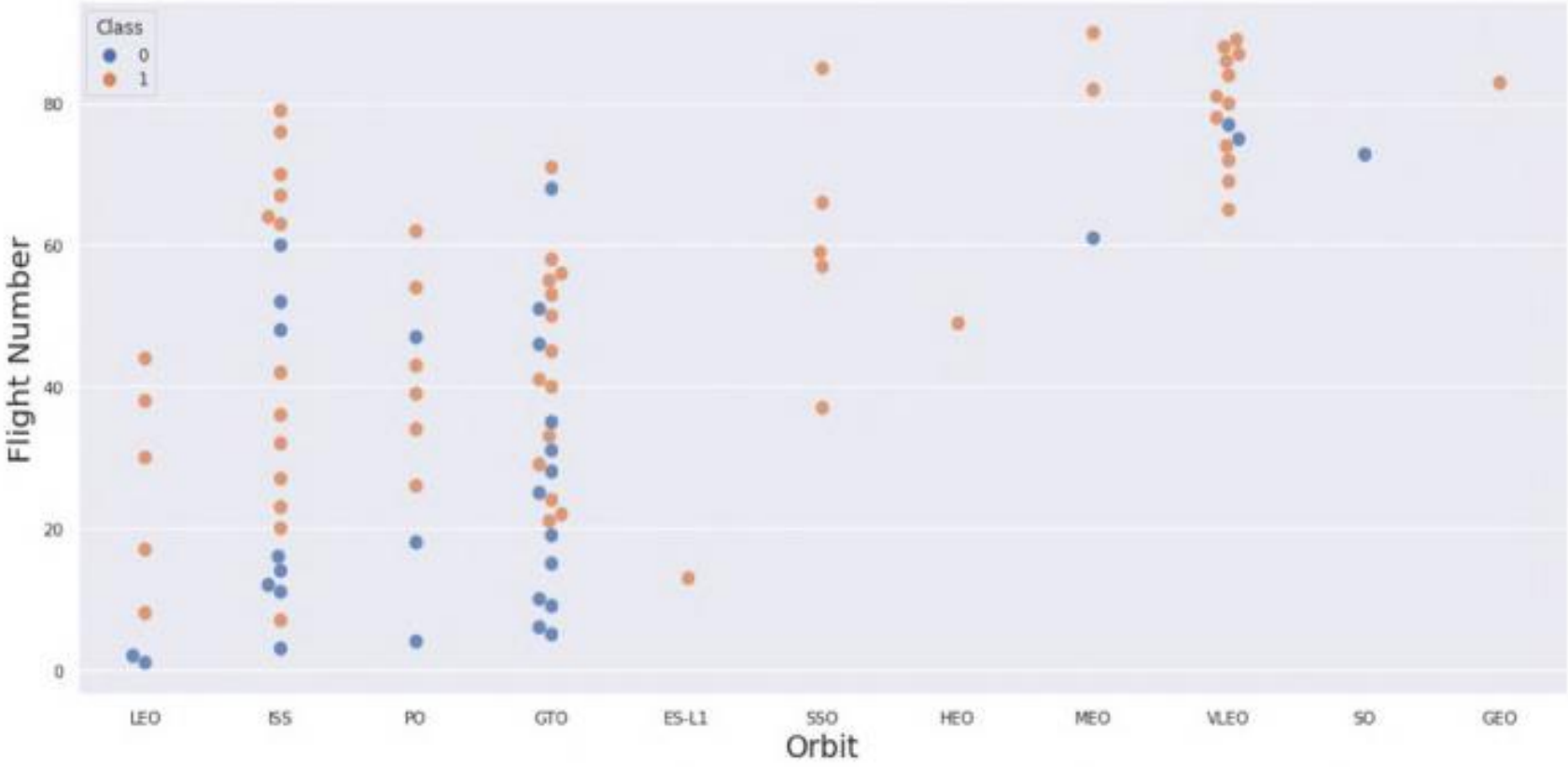
Success Rate vs. Orbit Type



This bar graph indicate ;

- Influences the landing outcomes as some orbits has 100% success rate such as SSO, HEO, GEO AND ES-L1 while SO orbit produced 0% rate of success.
- Moreover, it indicates shows that some of orbits has 1 occurrence such as GEO, SO, HEO and ES-L1 .

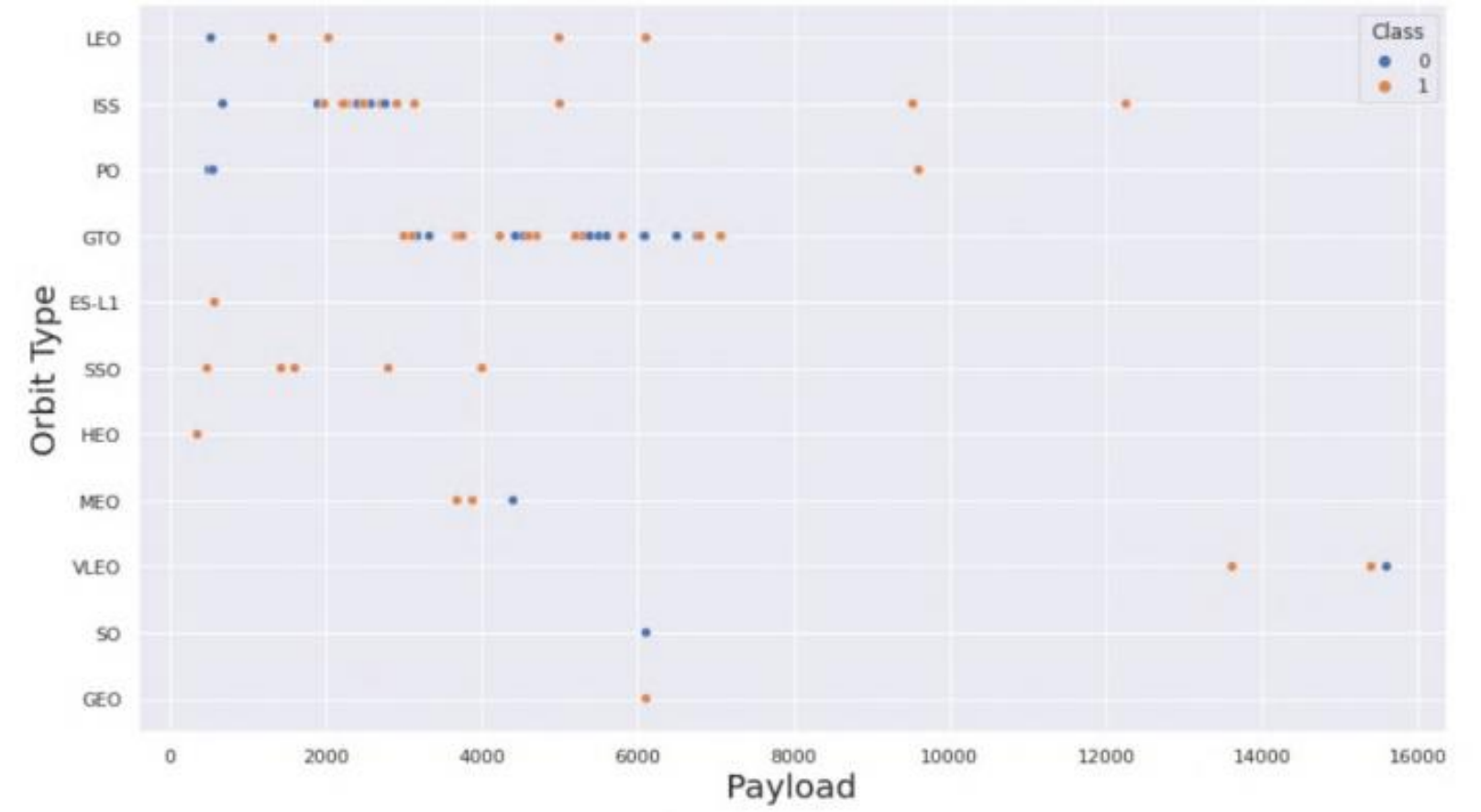
Flight Number vs. Orbit Type



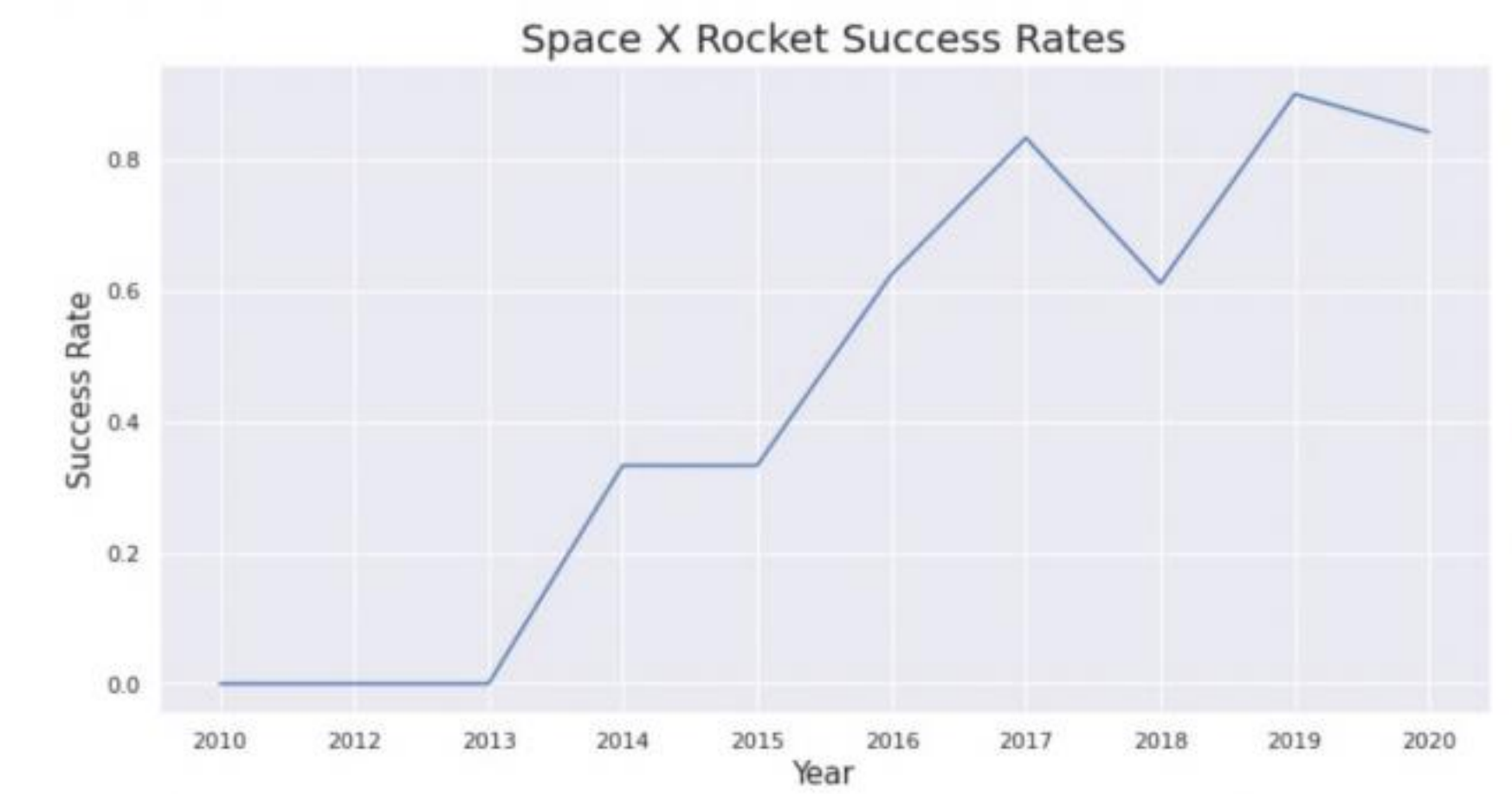
- This scatter plot indicates that larger the flight number on each orbits, it result greater success rate.
- However; GTO & Orbit 1 occurrence can be considered as an outliers.

Payload vs. Orbit Type

- Heavier payload shows positive impact on LEO, ISS and P0 orbit. But it has negative impact on MEO and VLEO orbit.
- GTO orbit indicates no correlation between the attributes.



Launch Success Yearly Trend



- This trend analysis indicates that success rate has a growing trend over the period of time.

All Launch Site Names

- There are four launch sites, and it is as follows;

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

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEX;
```

Launch Site Names Begin with 'CCA'

- Launch sites name begin with CCA is appended below;

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	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 Payload Mass is appended below;

Total Payload Mass by NASA (CRS)
45596

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass by NASA (CRS)
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is appended below.

Average Payload Mass by Booster Version F9 v1.1
2928

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS "Average Payload Mass by Booster
WHERE BOOSTER_VERSION = 'F9 v1.1';
```

First Successful Ground Landing Date

- First successful landing outcome on ground pad is appended below;

First Succesful Landing Outcome in Ground Pad
2015-12-22

```
%sql SELECT MIN(DATE) AS "First Succesful Landing Outcome in Ground Pad
WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT BOOSTER_VERSION FROM SPACEX WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

Total Number of Successful and Failure Mission Outcomes

Failure Mission
1

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Failure Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Failure%';
```

Successful Mission
100

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Successful Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Success%';
```

Boosters Carried Maximum Payload

- Boosters Carried Maximum Payload is appended below.

Booster Versions which carried the Maximum Payload Mass
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

- Ranked 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	Total Count
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

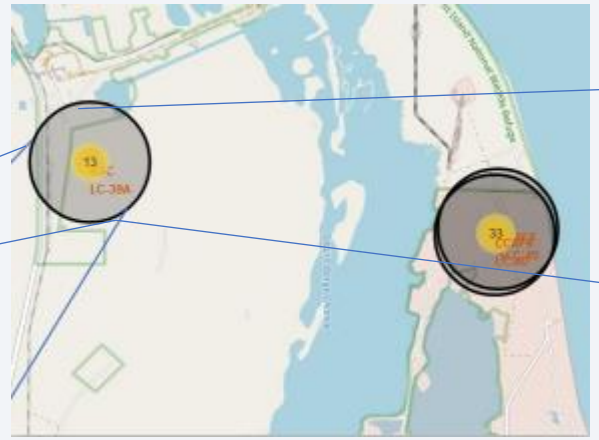
Section 3

Launch Sites Proximities Analysis

Launch Sites

- Launch sites often located proximity to sea.

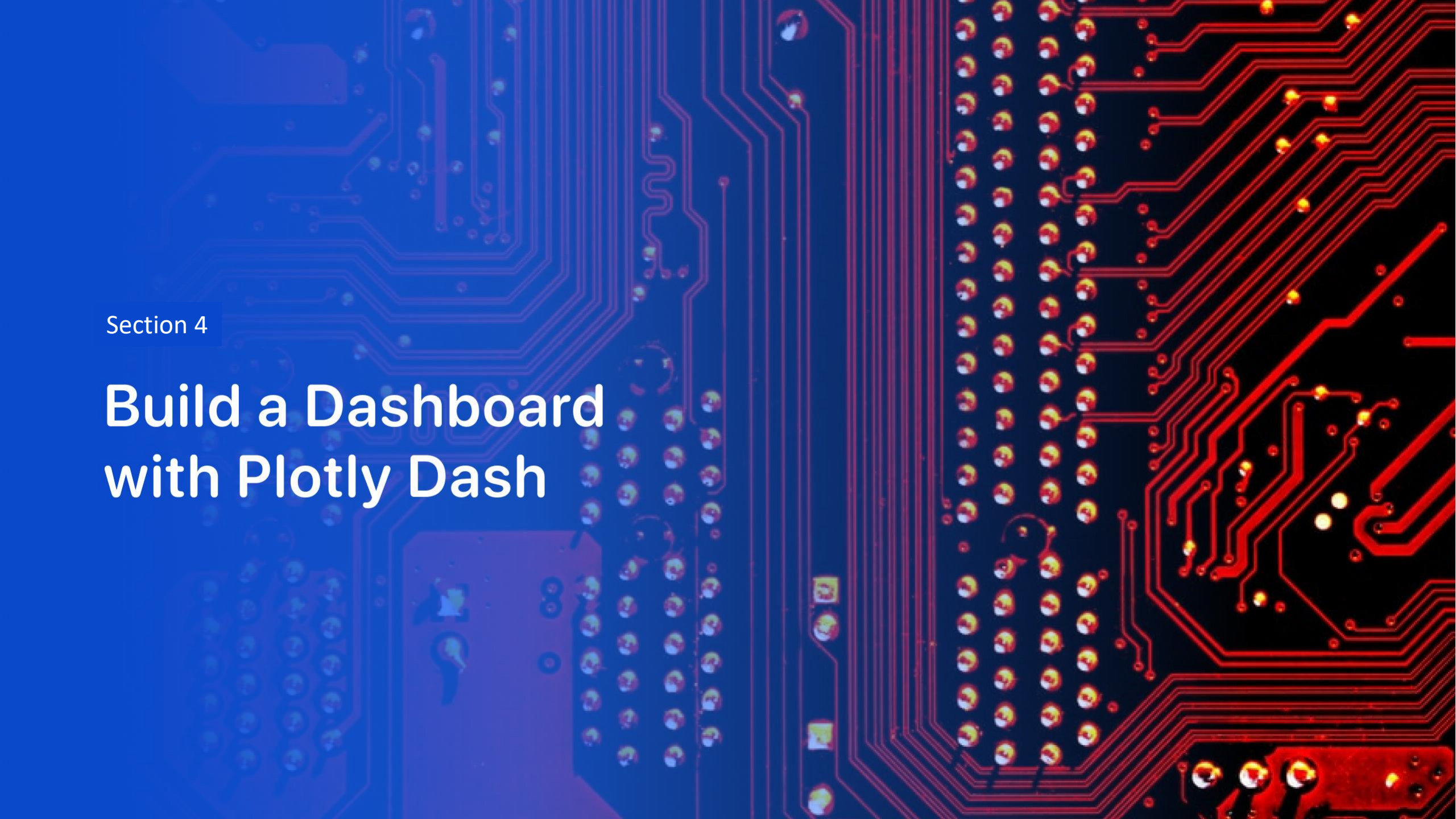
Launch Site Statistics



- Green dots shows successful launches while and red dots shows failures. And mostly the launches are successful.

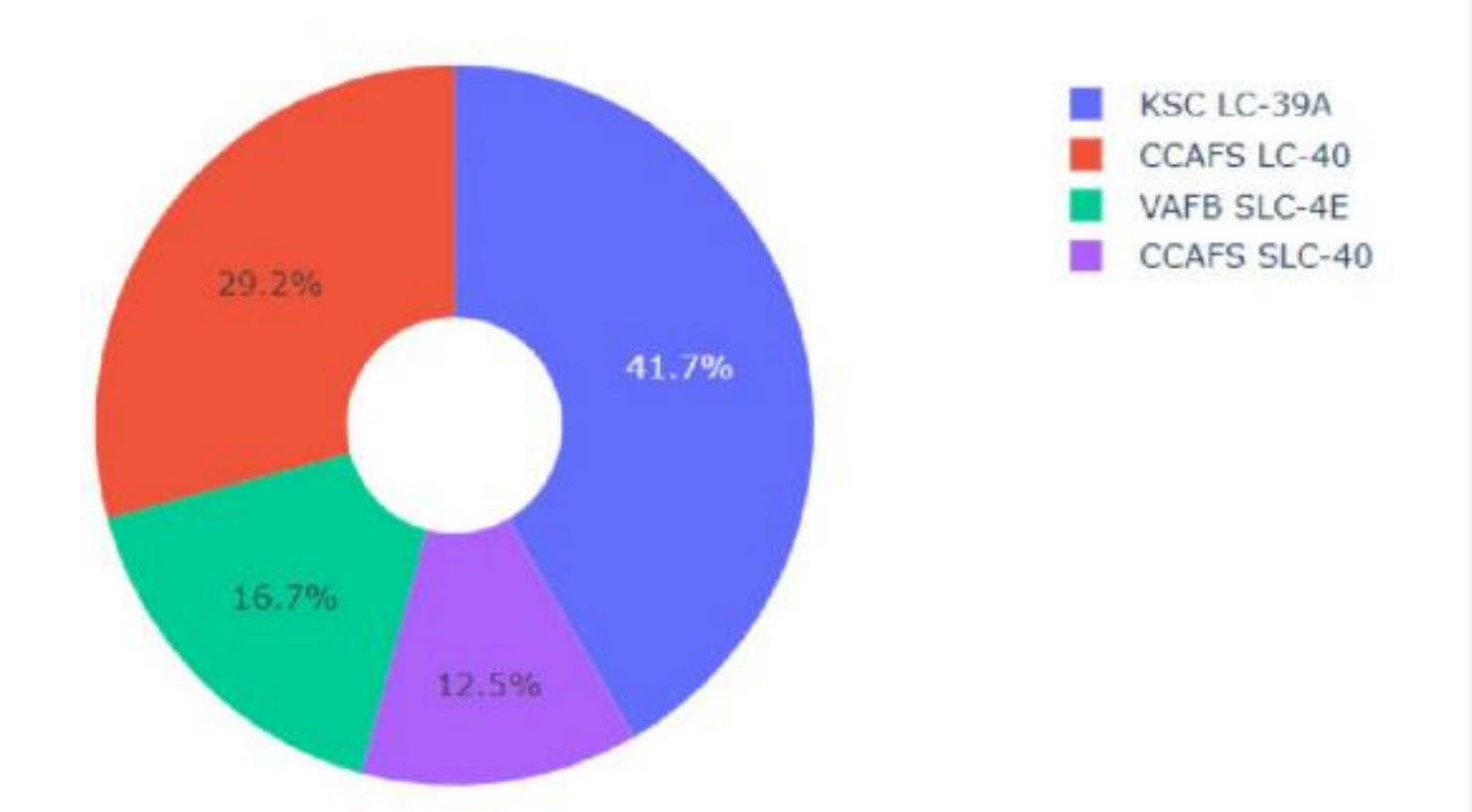
Launch sites features

- Launch sites are located far from the populated areas. Which is a good sign for the safety of the people.

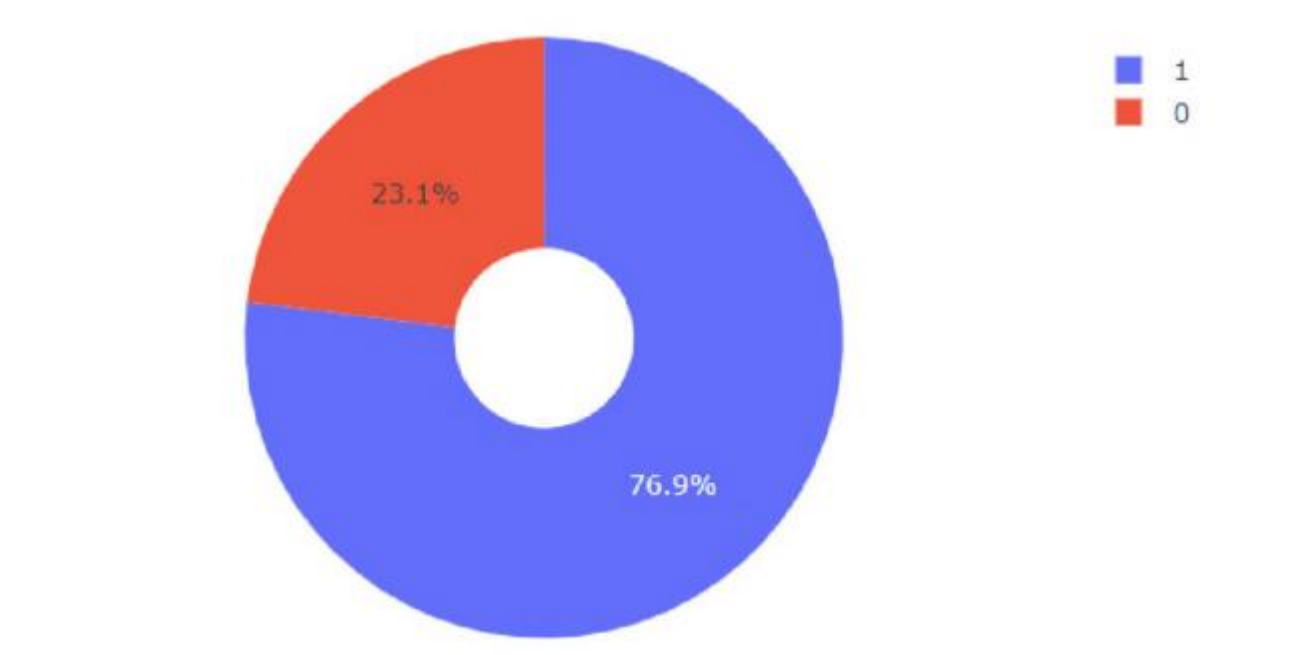

Section 4
Build a Dashboard with Plotly Dash

Success of the operation by launch site

- KSC LC-39A has 41.7% highest success
- CCAFS LC-40 has 29.2% 2nd highest success
- CCAFS – SLC-40 has 12.5% least success.

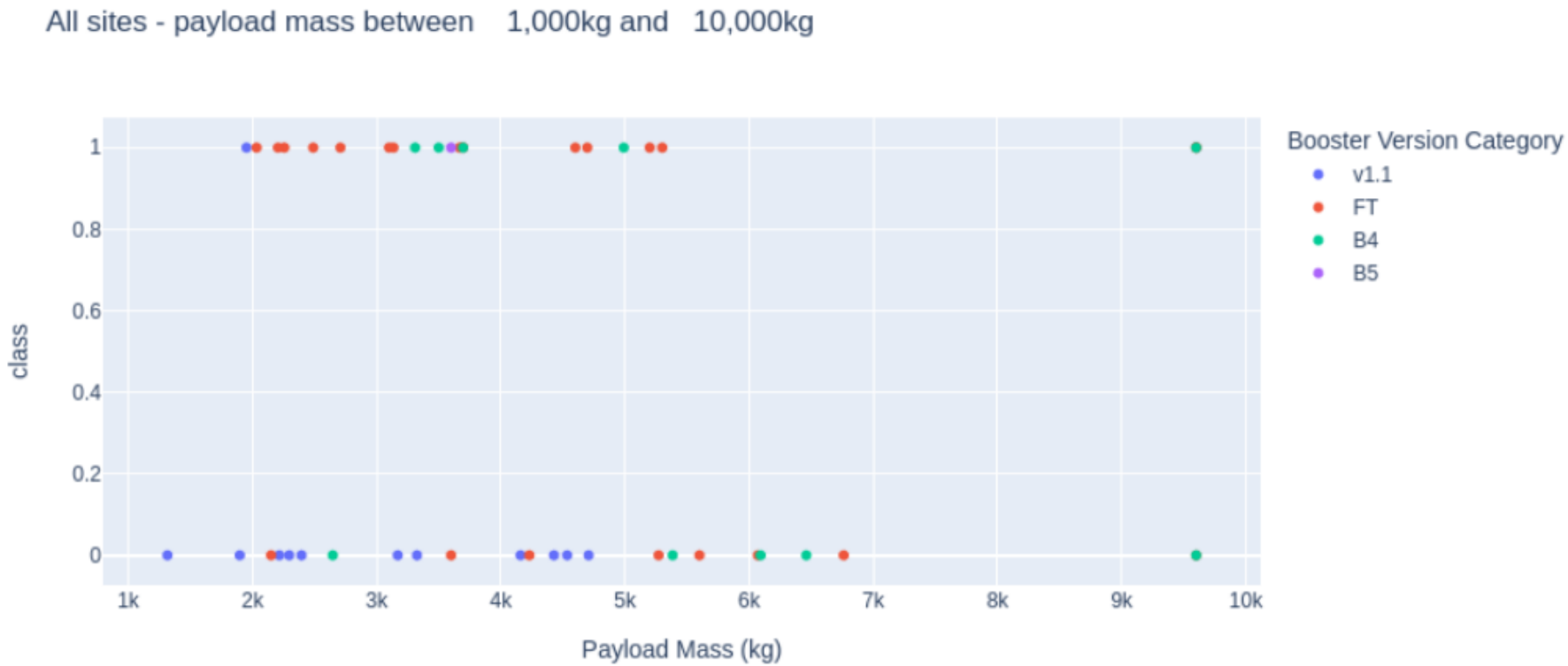


Success rate on KSCLC-39A



- KSCLC-39A has 76.9% success rate with 23.1% failure rate as well.

All Sites Payload mass between 1,000 KG till 10,000 kg



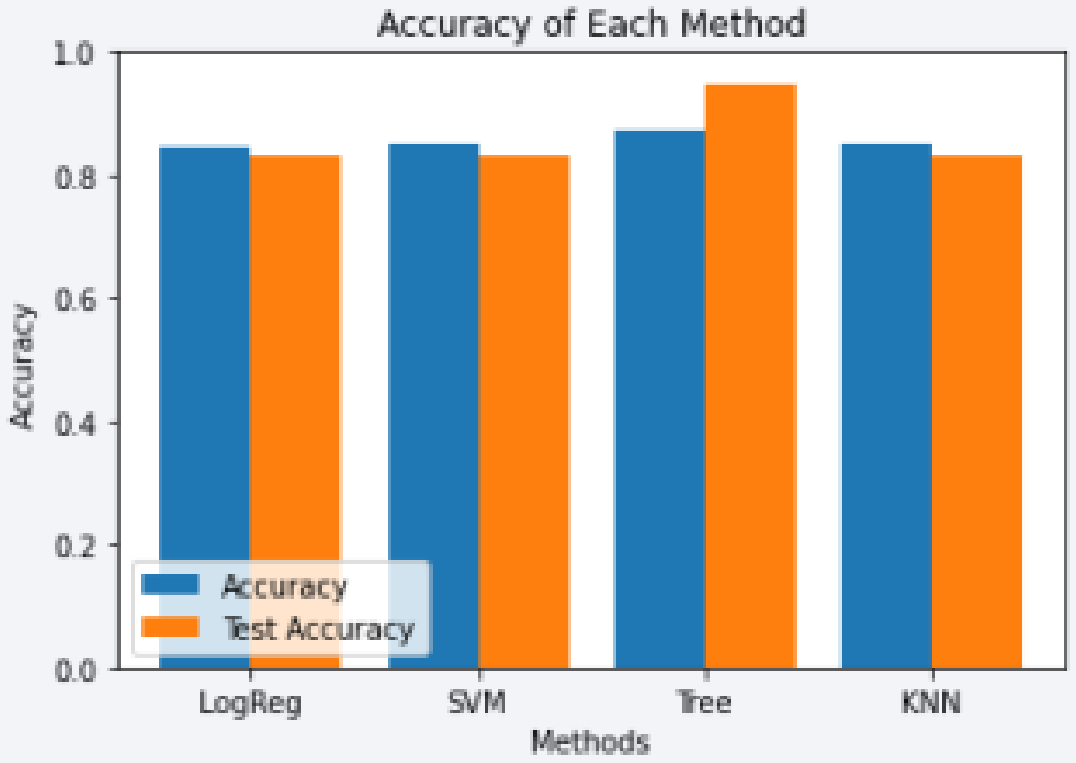
- Payload mass <= 6,000 KG has higher success rate

Section 5

Predictive Analysis (Classification)

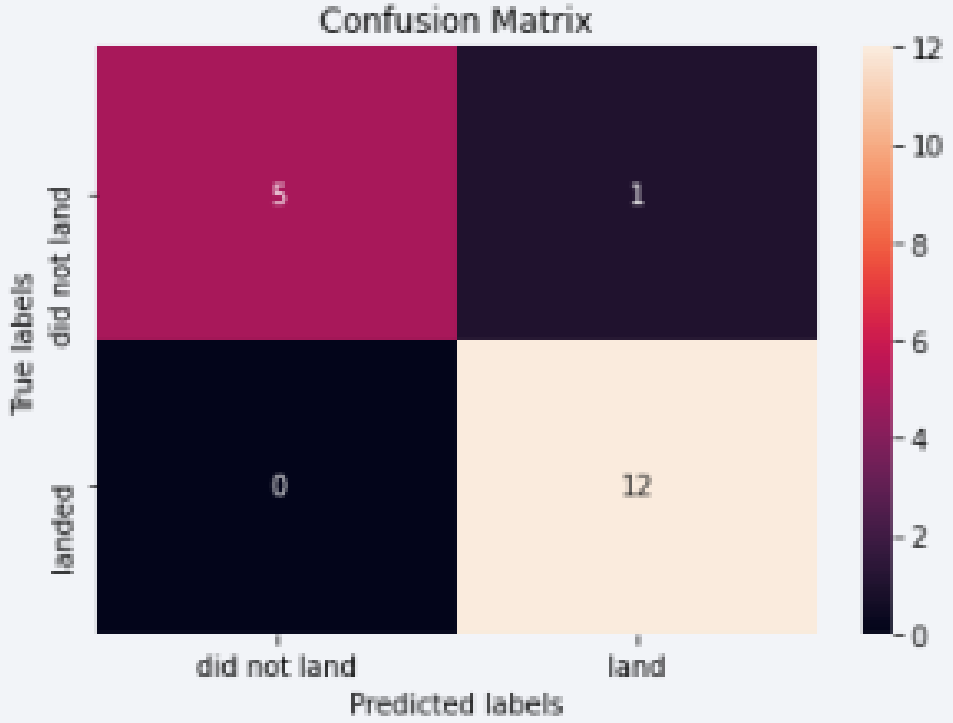
Classification Accuracy

- Four models has been tested, and their accuracies have plotted separately
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over 87%.



Confusion Matrix

- Confusion matrix of Decision Tree model shows higher percentage of true positive.



Conclusions

- The Decision Tree model was the best Machine Learning model for this problem.
- From year 2013, the success rate for SpaceX launches is increased over the period of time.
- Launches above 7,000kg are less risky compared to other launches..
- KSC LC-39A have the most successful launches of any it has 76.9% success rate.
- SSO orbit have the most success rate; of 100% .

Appendix

- https://github.com/chandimadabare/DC_Capstone_SpaceX_CD

Thank you!