IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohammad Yousefi ghaleh joogh
9/22/2023

Outline

Executive Summary

- Summary of methodologies
    1. Collecting Data by using API and Web Scraping
    2. Performing Data Wrangling and EDA
    3. Create Interactive charts and maps
    4. Performing ML models for predicting outcomes
- Summary of all results
    1. Performing EDA analysis
    2. Performing interactive analysis
    3. Performing Predictive Analytics

Introduction

- SpaceX's Falcon 9 is a two-stage partially reusable rocket which is capable to carry cargo and crew into Earth orbit and beyond

- The reusability of first stage of the rocket helped SpaceX to outperformed his rivals and it is important to predict successful landing of the rocket, Since it has direct impact on the final cost of the shipment and its competitiveness price which is 62 Million Dolor**.**

What is our aim in this project?

We will explore most factors that affect Falcon9 landing and find most important ones and relation between those factors, then by using different Machine Learning models we try to predict probability of successful/failed landing of the rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Gathering data by using SpaceX official API
  - Gathering data from web by web scraping (Wikipedia)
- Perform data wrangling
  - Converting raw data into a usable form for analysis including cleaning, organizing, and transforming
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Converting raw data into a usable form for analysis including cleaning, organizing, and transforming

Data Collection

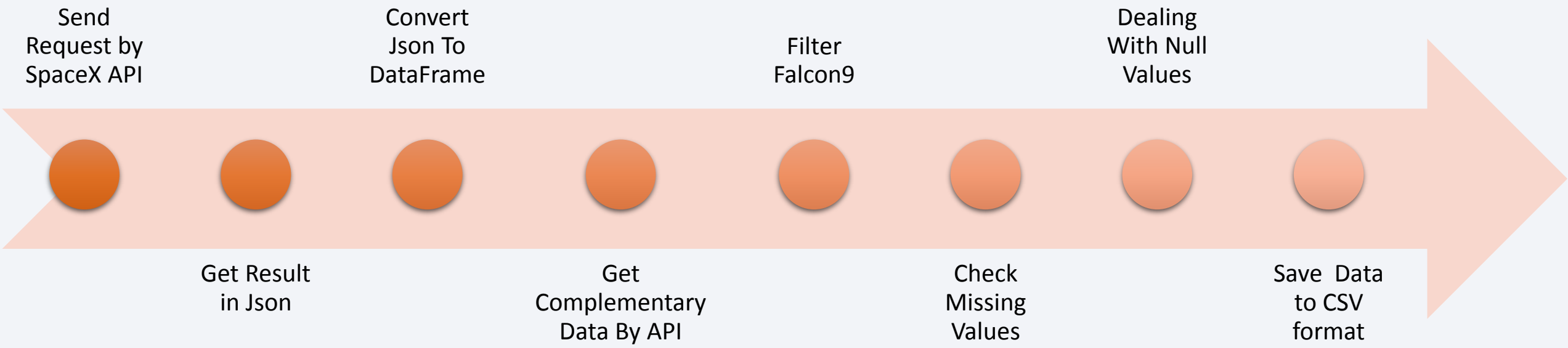
The First Step Is Gathering Data Needed For Analysis. This Is Down By Two Methods:

1. The Spacex Company Officially Publish The Data And We Can Get Datas In Form Of Json Files Which Can Later Converted To Dataframe To Be Use In Python (https://api.spacexdata.com/v4/)

2. In Second Method Data Can Be Access By Web Scraping From Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

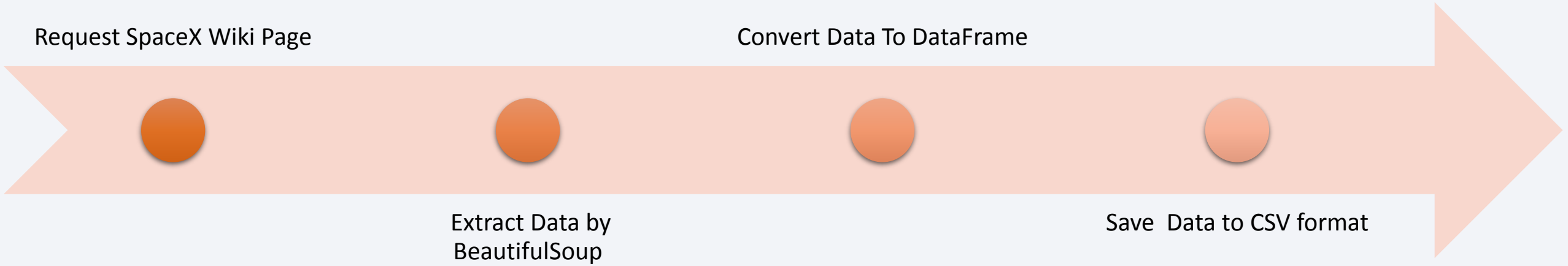
With Data In Hand:

Data Collection – SpaceX API



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/p1_data_collection.ipynb

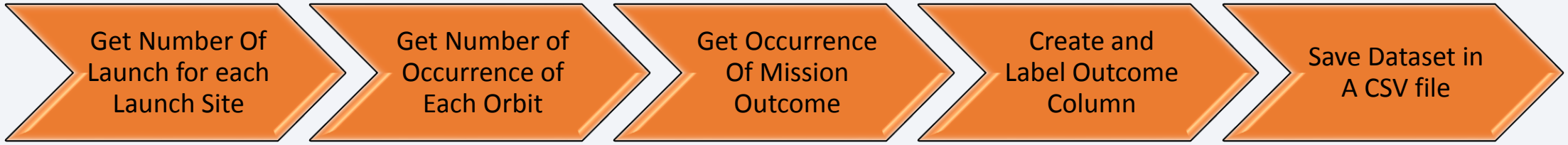
Data Collection - Scraping



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/p2_jupyter-labs-webscraping.ipynb

Data Wrangling

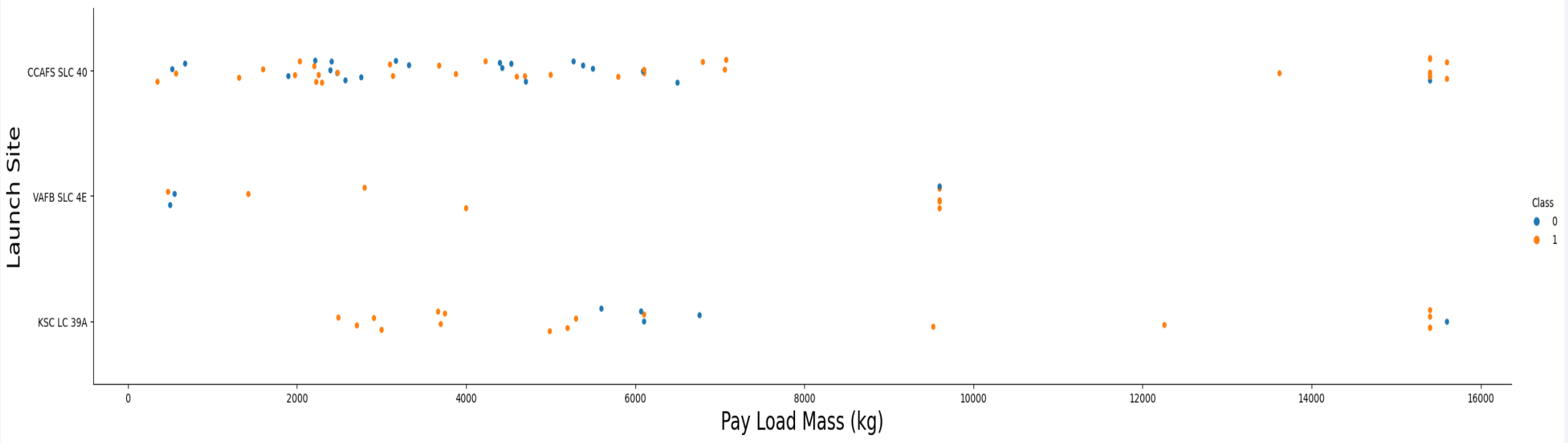
In this section we performed Exploratory Data Analysis and Determine Training Labels



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

The Main Objective In This Section Is To Perform EDA And Data Feature Engineering.
EDA Is Visual Representation Of Our Data Which Help Us To Explore Relationships Between Different Features And Make It Easier For Us To Understanding Them. For Example help us to understand relationship between Launch Site and Orbit. Or relationship between Payload and Launch Site



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-dataviz_final.ipynb

EDA with SQL

In This Part We Use SQL To Retrieve Data Then We Accomplish Following Tasks:

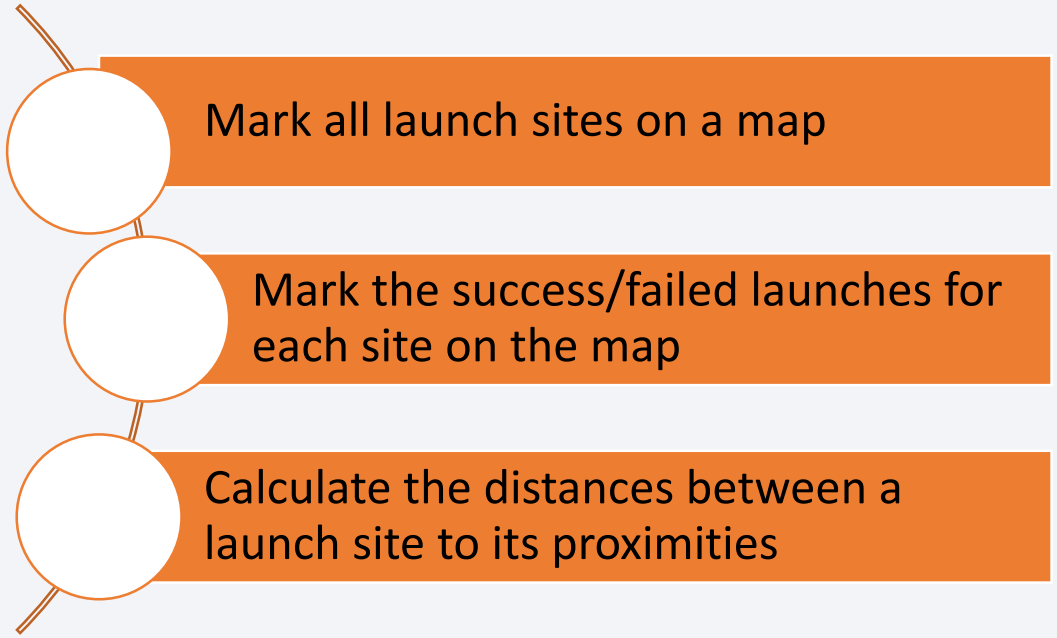
1. Display the names of the unique launch sites in the space mission
2. Display the total payload mass carried by boosters launched by NASA (CRS)
3. Display average payload mass carried by booster version F9 v1.1
4. List the date when the first successful landing outcome in ground pad was achieved.
5. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
6. List the total number of successful and failure mission outcomes
7. names of the booster_versions which have carried the maximum payload mass

https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Folium is a powerful Python library that helps you create several types of Leaflet maps. In this part By Folium API, in generated maps:

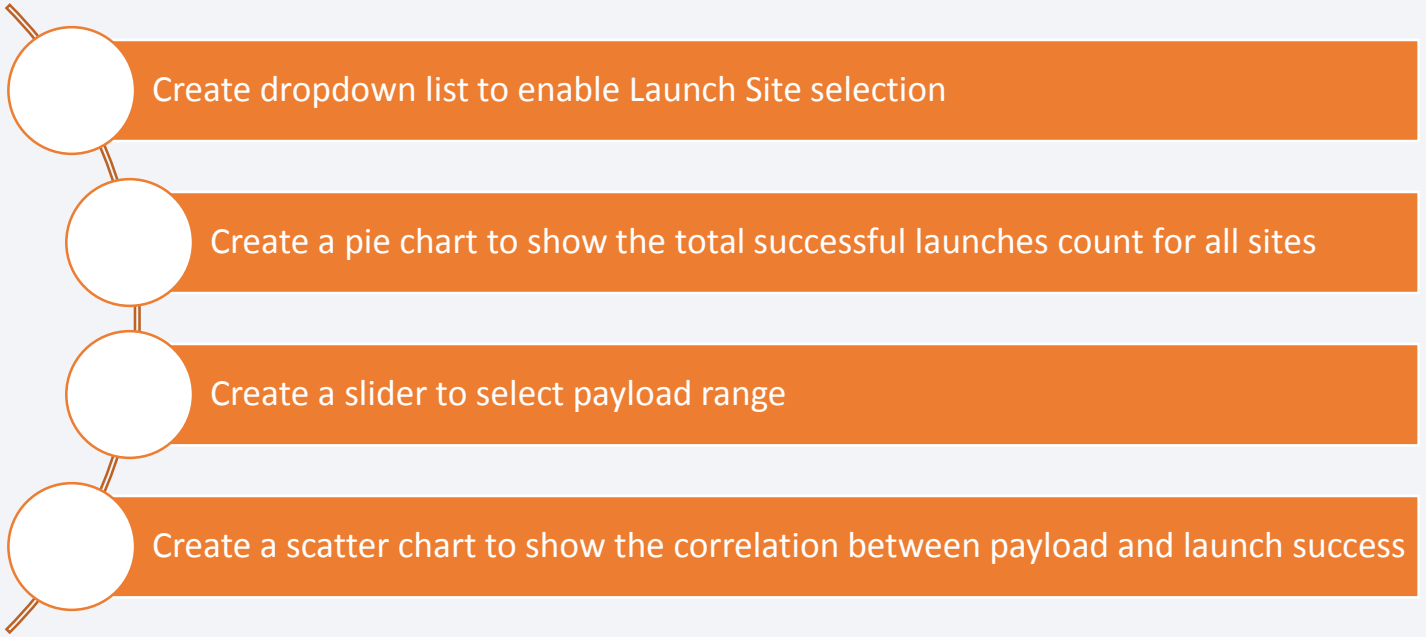
- Circles: represents Launch Sites
- Colored Markers: Success/Failures in each launch Site
- Lines: to show distances to nearest City, railway, coastline and highway



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location_final.ipynb

Build a Dashboard with Plotly Dash

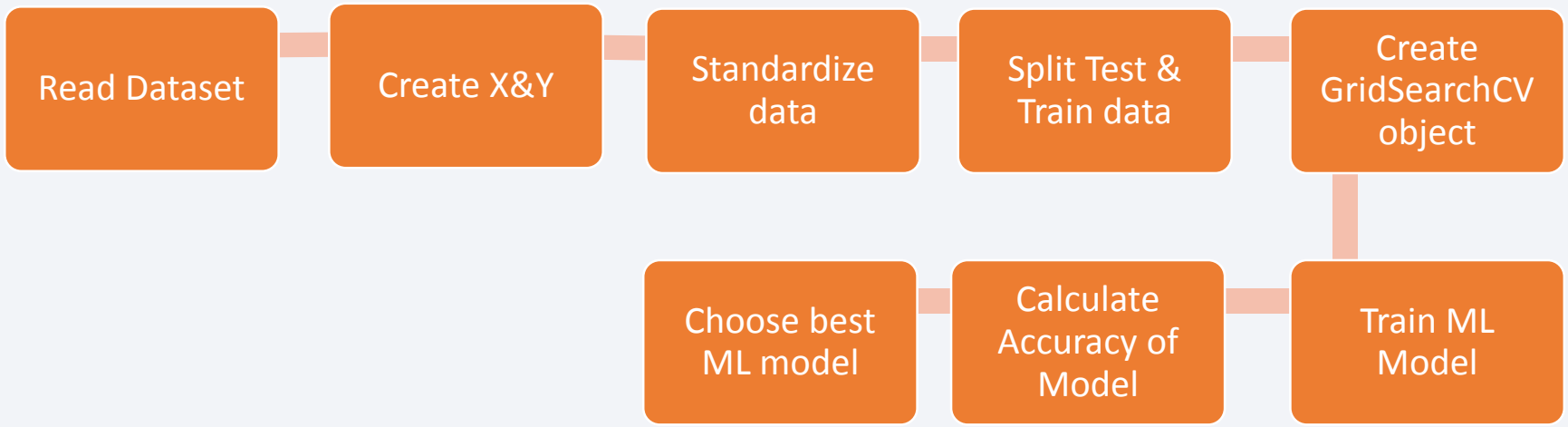
In this section we utilize the power of Plotly and Dash library to create a web app to interactively visualizing the data. This enable us to study relationship between Launch sites, Payload mass and booster types easily.



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/spacex_dash_app_final.py

Predictive Analysis (Classification)

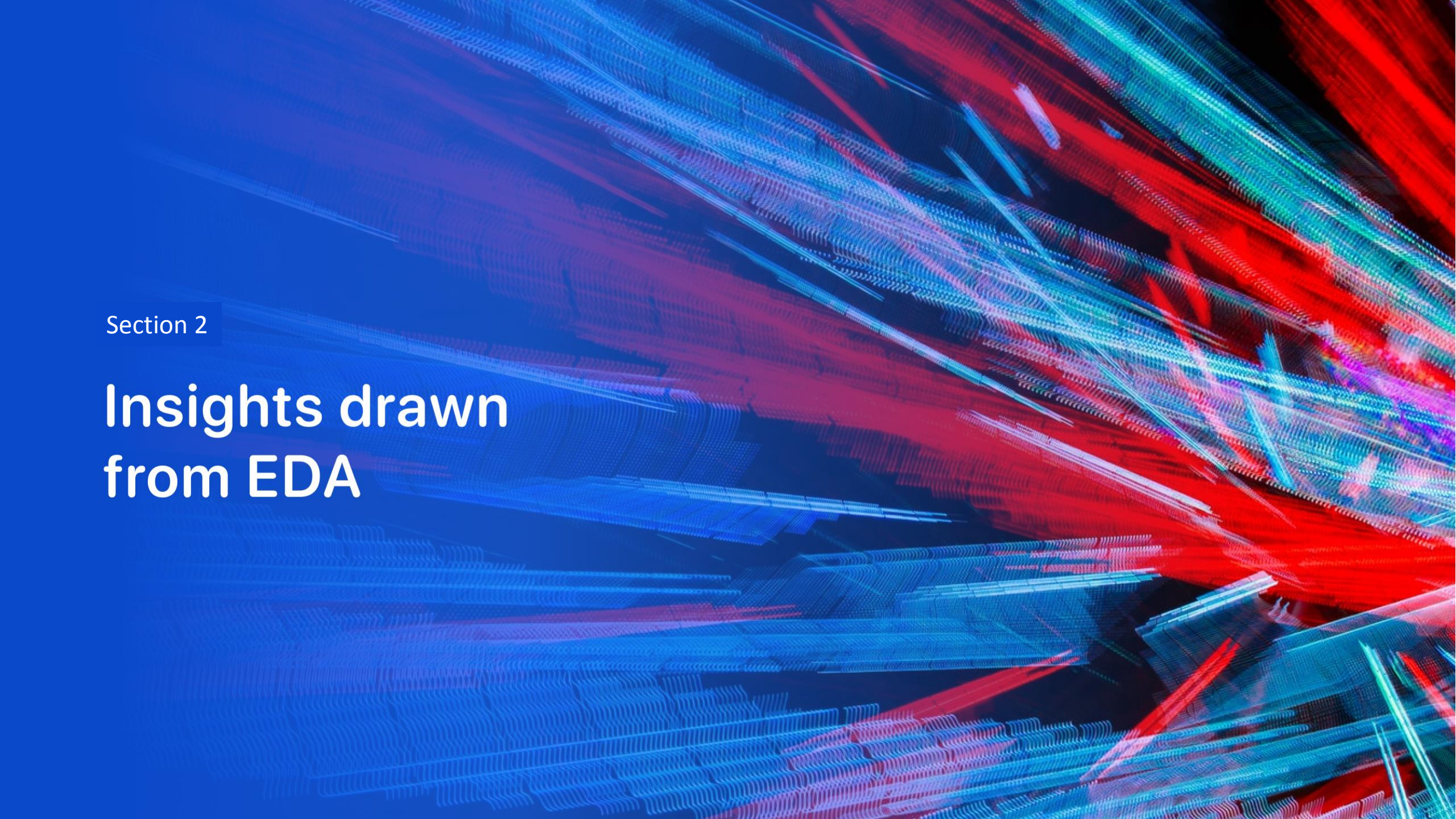
In This Part We Use Scikit-learn Python Library For Performing Machine Learning For Prediction Falcon 9 Rocket Success/Failure Base On Some Selected Features. Then We Try To Find Best Hyper-paramters For One Of The Models And Determine Which One Of The Machine Learning Models Predict With Higher Accuracy.



https://github.com/mygir77/IBM-Applied_Data_Science_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

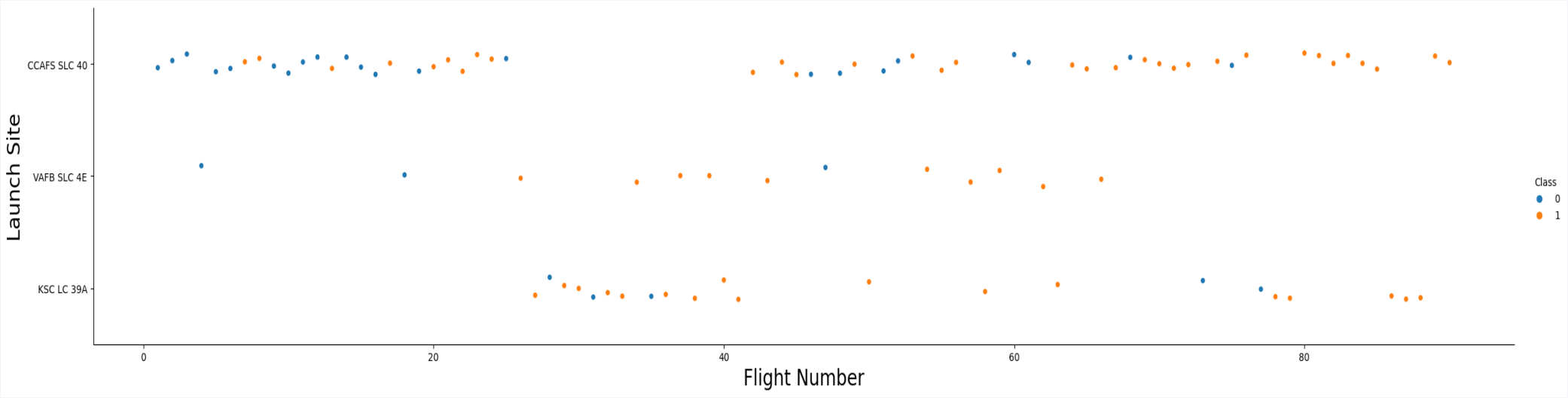
Results

- Exploratory data analysis demonstrate by increasing Flight Number from 2013 to 2020 success rate increased some launch site have Higher success ratio
- Interactive analytics web app, shows 2000-6000 kg with FT booster have performed better
- In Predictive analysis section, we come up with Decision tree has better performance with 83% accuracy.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

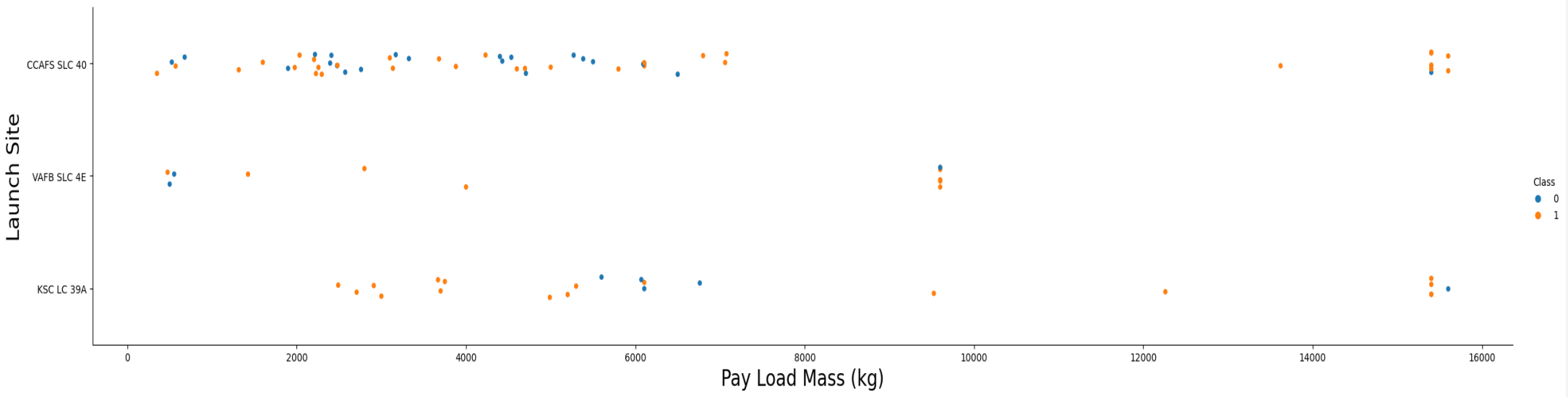
Scatter plot of Flight Number vs. Launch Site



- As the project go further, Obviously FlightNumber 80, Successful landing increased.
- Most used site is CCSFS SLC 40 (**Cape Canaveral Space Launch Complex**)

Payload vs. Launch Site

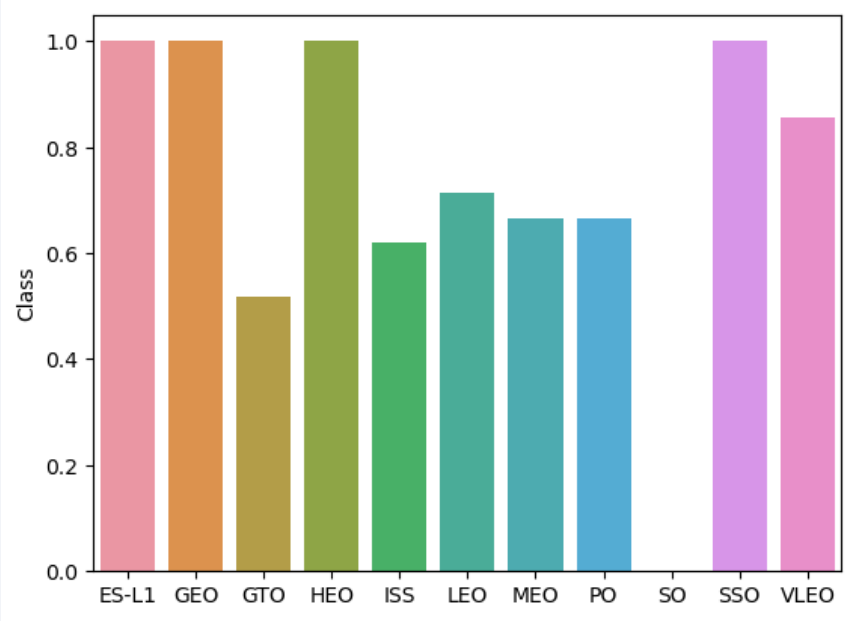
Scatter plot of Payload vs. Launch Site



- Base on this Plot VAFB SLC 4E did not used for PayLoad more then 10000Kg

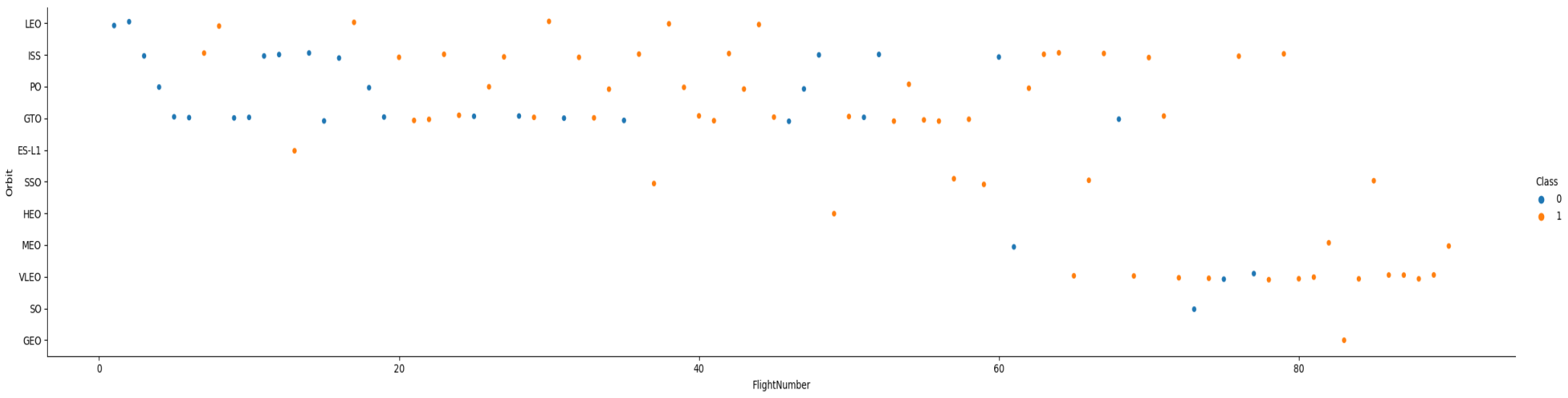
Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type



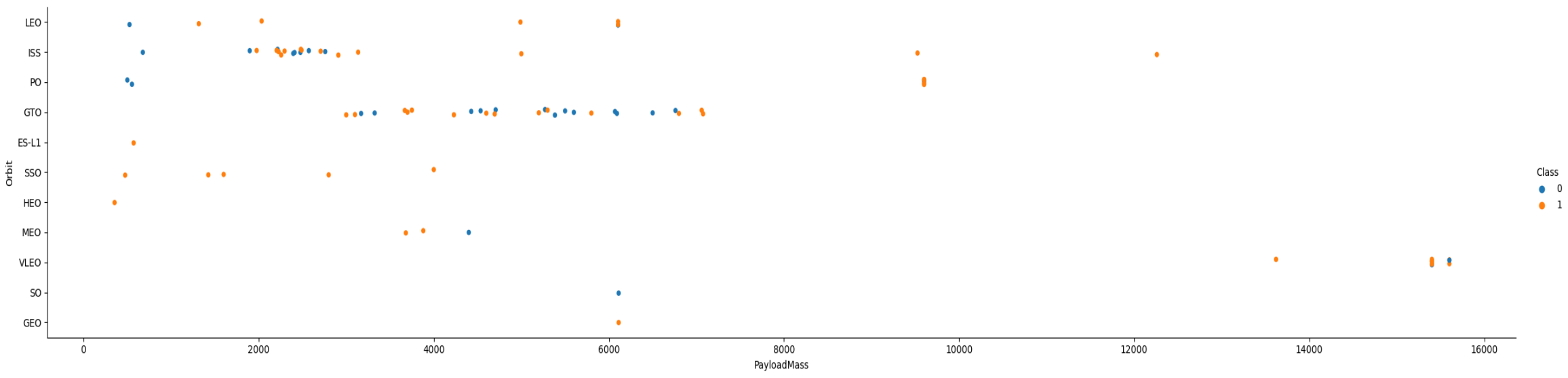
- This Bar chart propose ES-L1, GEO, HEO, SSO orbits have more success rate
- Going to GTO orbit has the lowest success rate

Flight Number vs. Orbit Type



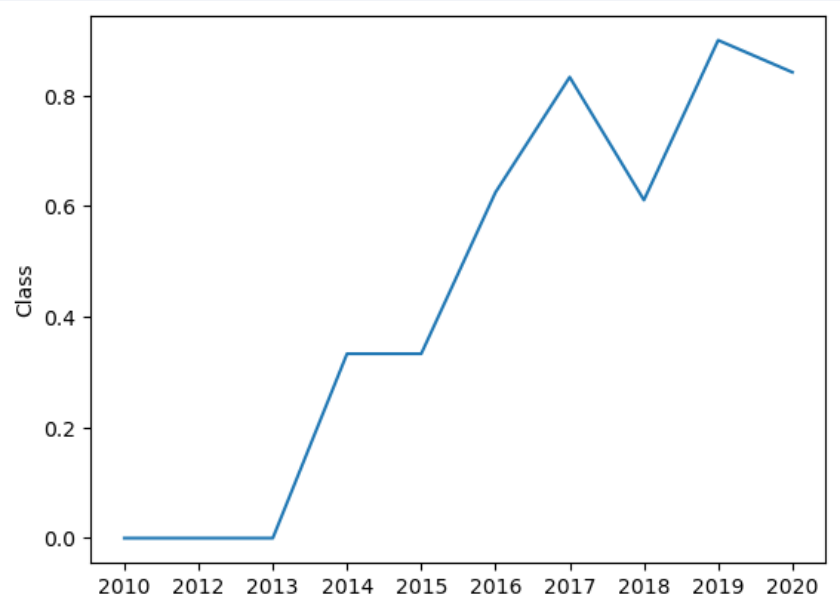
- Previous Bar cart shows High success rate for GEO, SSO and HEO, but this plot indicates these 3 orbits have the lowest Flight attempts and their success cannot be that much accountable.
- In recent Flight numbers VLEO orbit has most success landing

Payload vs. Orbit Type



- PO, LEO, LEO have the most success rate on Heavy Loads
- VLEO not used for light Payloads

Launch Success Yearly Trend



- After 2013 success rate keep increasing

All Launch Site Names

%sql SELECT DISTINCT Launch_Site from SPACEXTABLE;

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

Launch Site Names Begin with 'CCA'

%sql SELECT * from SPACEXTABLE where LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where (Customer) LIKE 'NASA (CRS)';
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

```
%sql SELECT avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1';
```

Avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

```
%sql SELECT min(Date) from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)' ;
```

Min(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

%sql SELECT Booster_Version from SPACEXTABLE where (Landing_Outcome = 'Success (drone ship)') and (PAYLOAD_MASS__KG_ BETWEEN 4000 and 5999)

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) as MissionResult from SPACEXTABLE
GROUP BY Mission_Outcome;
```

Mission_Outcome	MissionResult
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

```
%sql SELECT BOOSTER_VERSION from SPACEXTABLE WHERE
PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE);
```

2015 Launch Records

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

```
%sql SELECT SUBSTR(Date,6,2) as Month,Date, Booster_Version,Launch_Site from SPACEXTABLE
WHERE SUBSTR(Date,1,4)='2015' and Landing_Outcome ='Failure (drone ship)';
```

Month	Date	Booster_Version	Launch_Site
10	**2015-10-01**	**F9 v1.1 B1012**	**CCAFS LC-40**
04	**2015-04-14**	**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

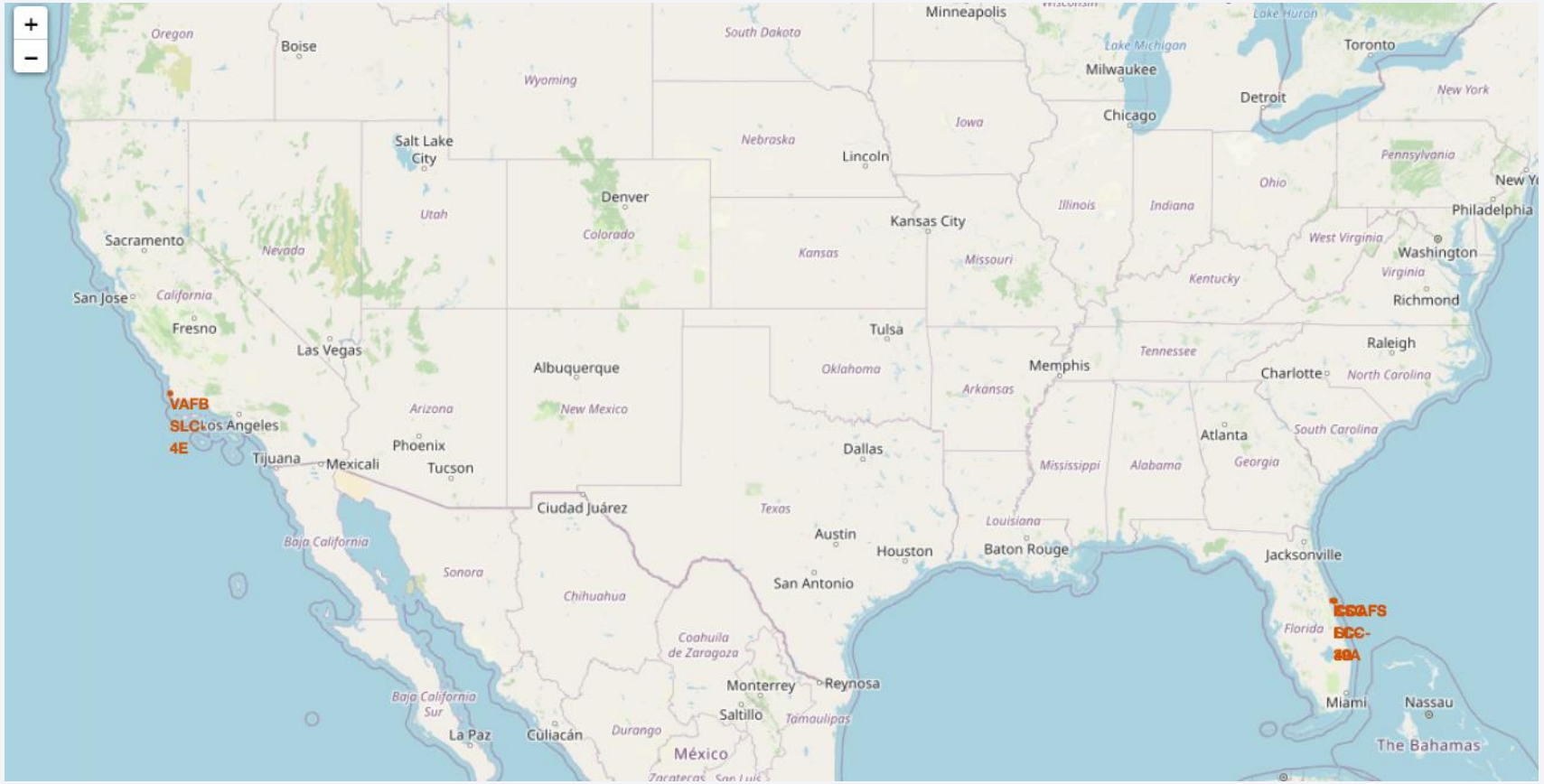
```
%sql SELECT Landing_Outcome, COUNT(*) as count_outcomes from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by [Landing_Outcome] order by count_outcomes DESC;
```

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


Section 3
Launch Sites
Proximities Analysis

launch sites locations on a map

- All Launch Site Sited near the coast lines



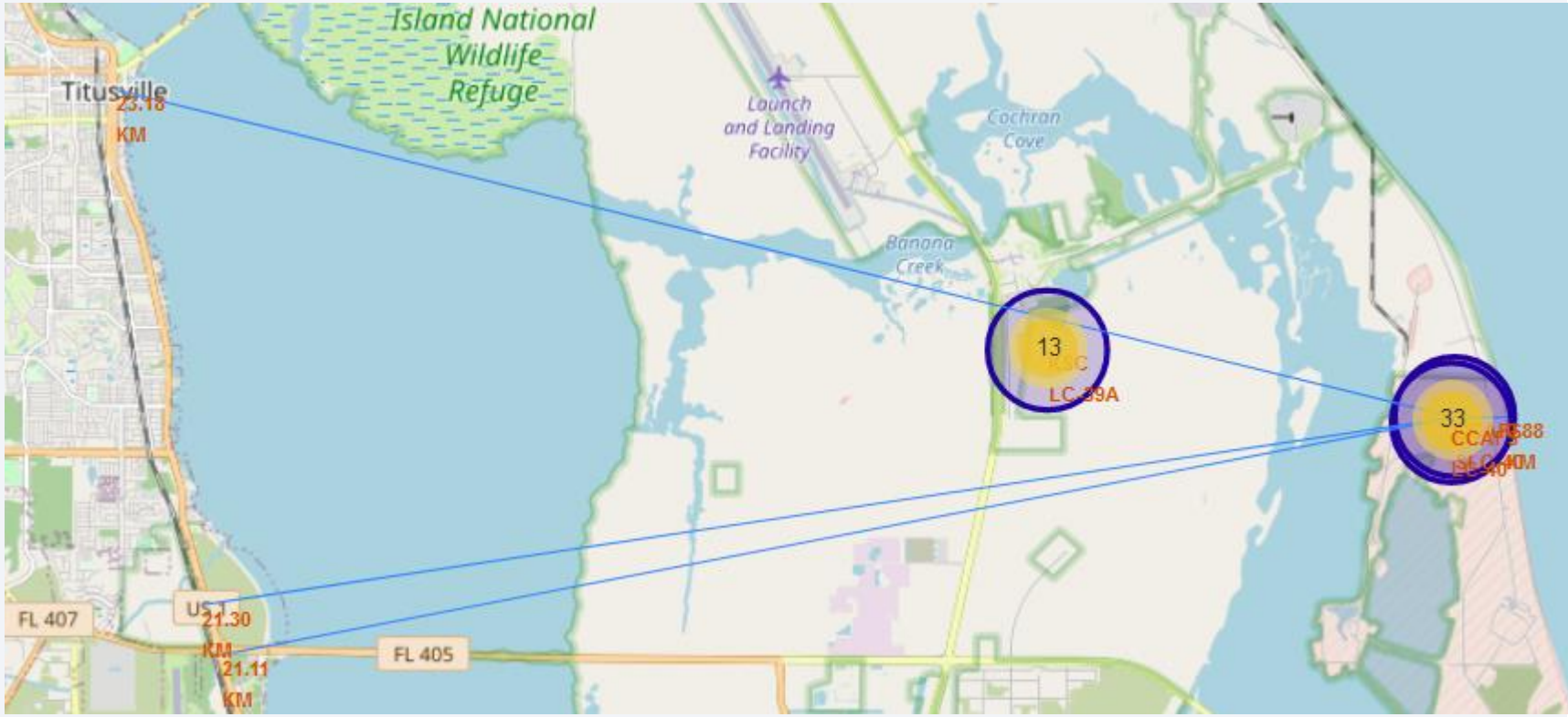
Launch Site Success Rates

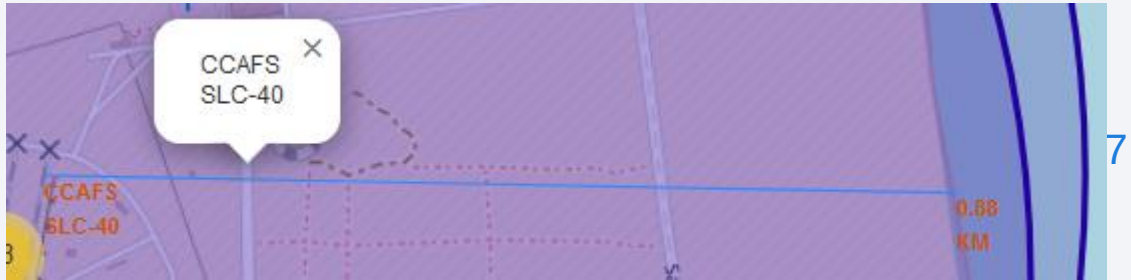
- KSC LC-39A has most success for falcon 9 landing
- CCAFS SLC-40 has most failed mission

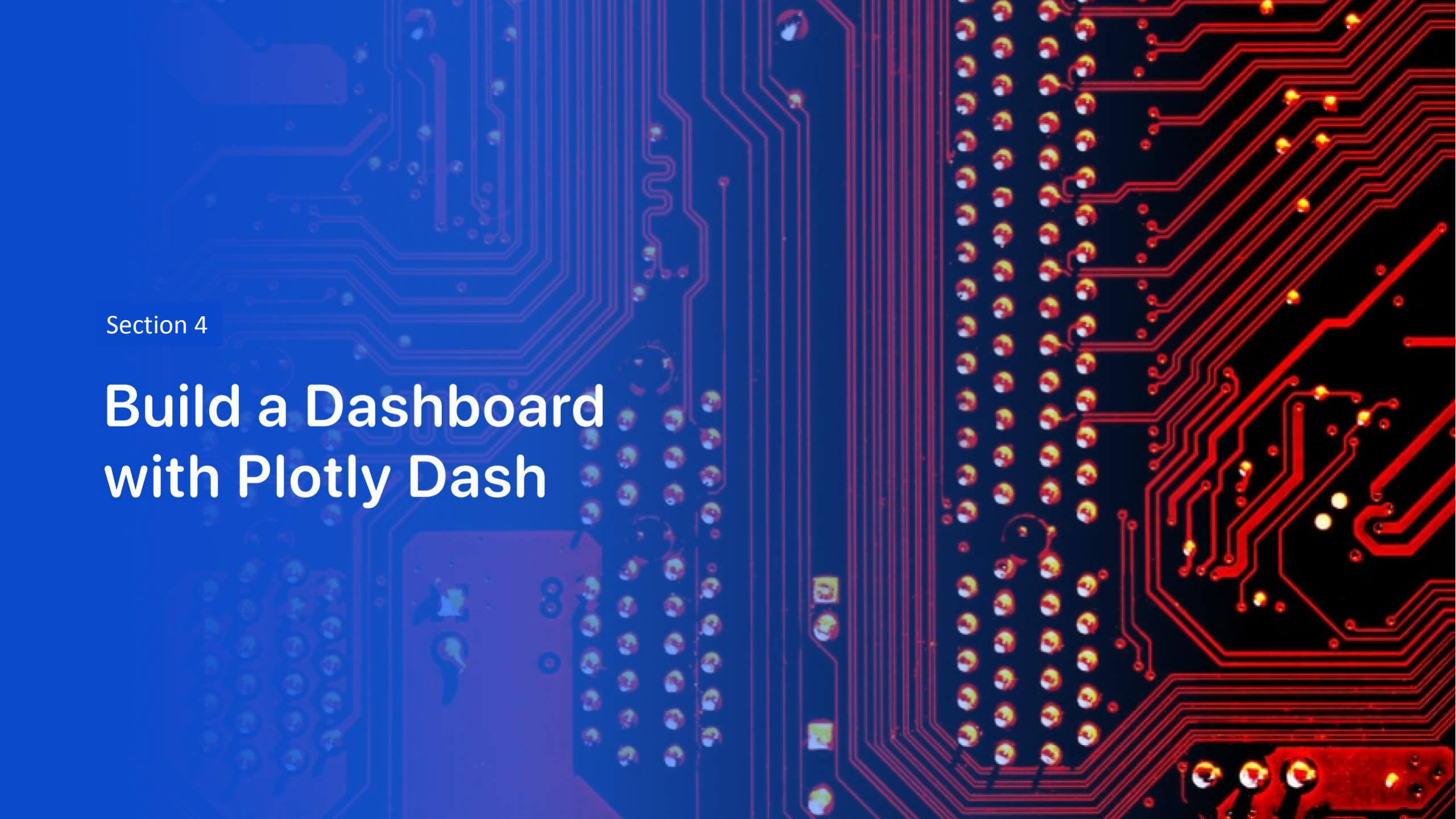
Distances between a launch site to its proximities

AS you can see distance to Nearest :

- City: 23 Km
- Railway: 21 km
- Highway: 21km
- Coastline: 0.9 km



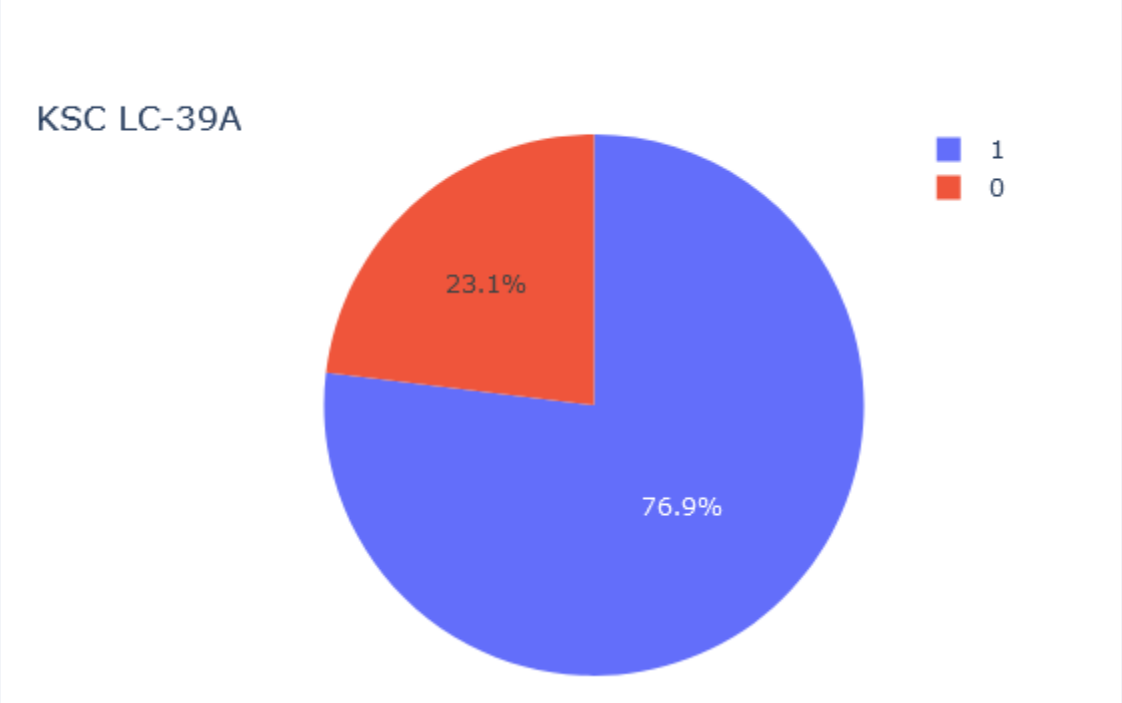



Section 4
Build a Dashboard with Plotly Dash

Total successful launches count for all sites

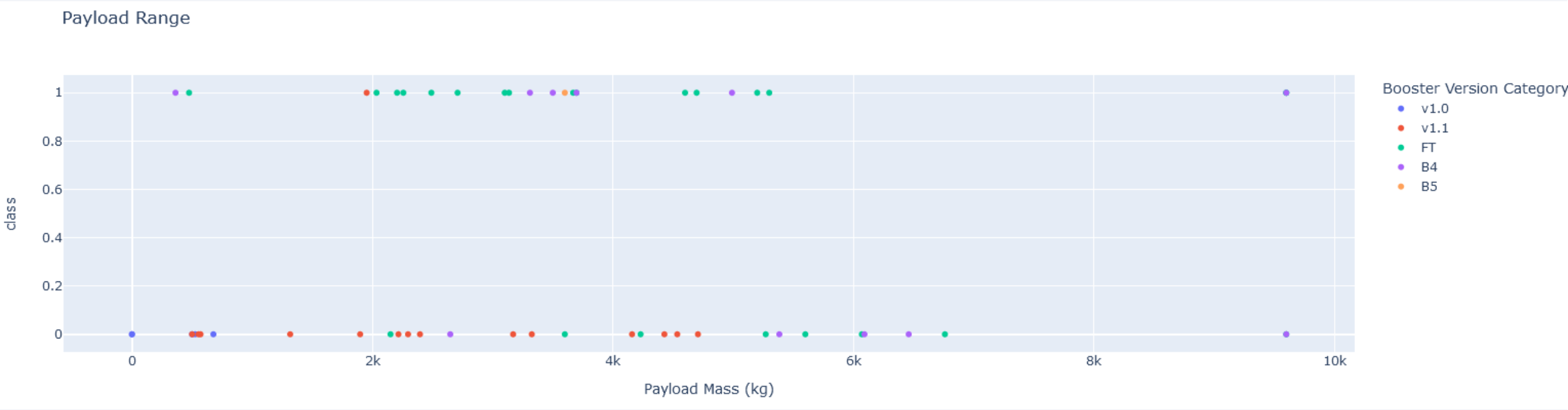
- KSC LC-39A has highest Success among the other launch sites and CCAFS –LC-40 comes next.

Success/Fail of most successful launch site



- KSC LC-39A as most successful launch site has 76.9% success and 23.1% failure.

Scatter for Payload vs Launch Outcome for all sites



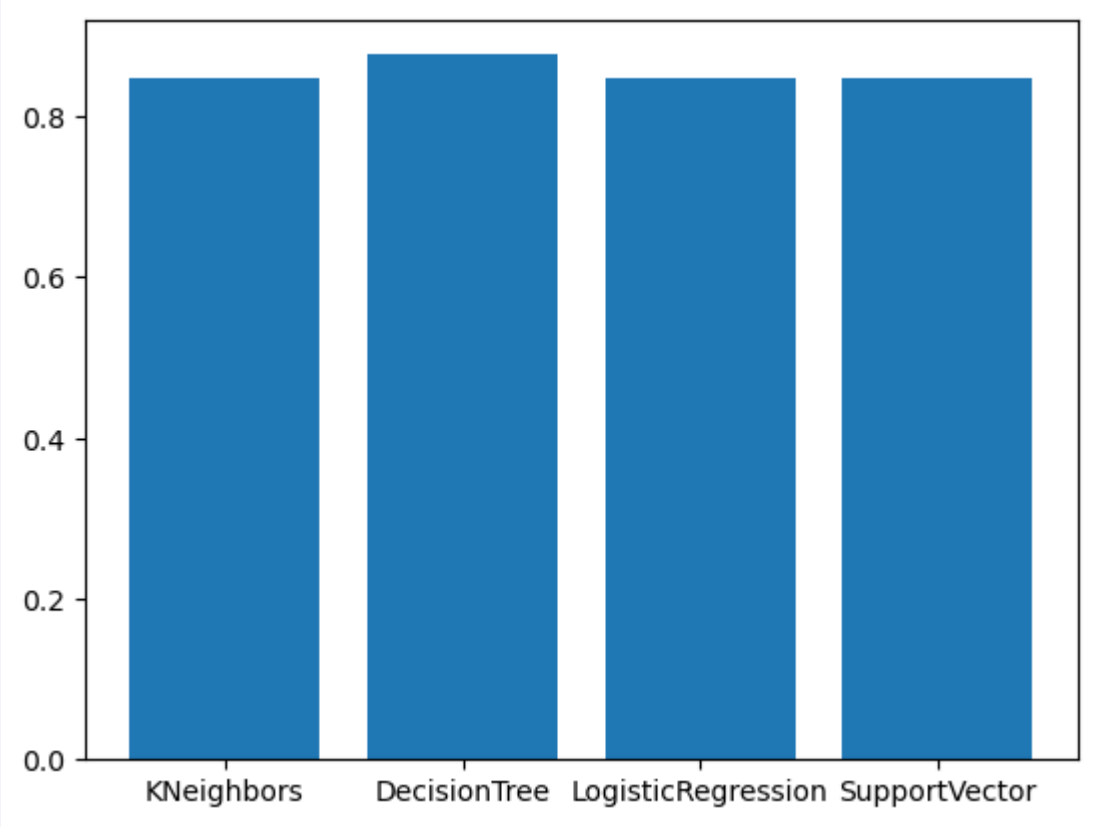
- Payloads between 2000-6000 kg have more chances to land successfully
- FT Booster has higher success rate

Section 5

Predictive Analysis (Classification)

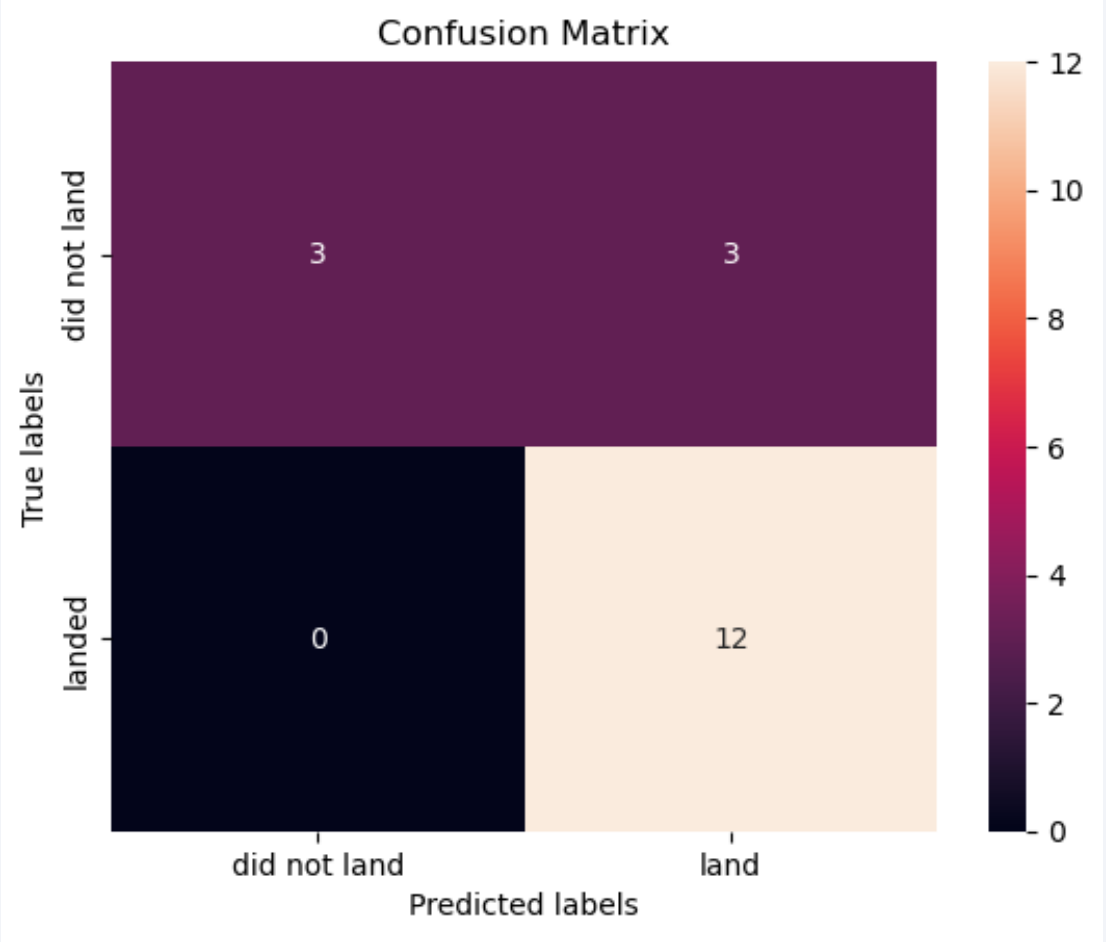
Classification Accuracy

- This chart indicates accuracy of models doesn't different with each other dramatically. But in overall, Decision Tree has higher accuracy.



Confusion Matrix

- Confusion matrix of Decision Tree model indicates this model has some difficulties in predict Failed Landing that predict as Landed mission(FN, Type2 error)



Conclusions

- By increasing Flight Number, Successful landings are increased
- By pass time from 2013 to 2020, successful Landing increased
- KSC LC-39A launch Site is best place for Landing
- In recent Launches, VLEO orbit outperformed the other ones
- PayLoad with 2000-6000 kg have higher success rate with FT booster
- For prediction Purpose, Decision Tree model perform slightly better

Thank you!