IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nizar Boussabat
22/09/2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Space X is leading the companies that provide space travels and the key for their success is due to the falcon 9 rocket which can reuse the first stage so it lower the travel cost
- As a new company Space Y ,we want to lead the market and the first step is to analyze space X rockets launches specially falson 9 rocket ones also we will try to predict whether the rockets landing are successful or not

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data sets were collected using two methods:

1-Using SpaceX Rest API

2-Web scraping related wiki pages

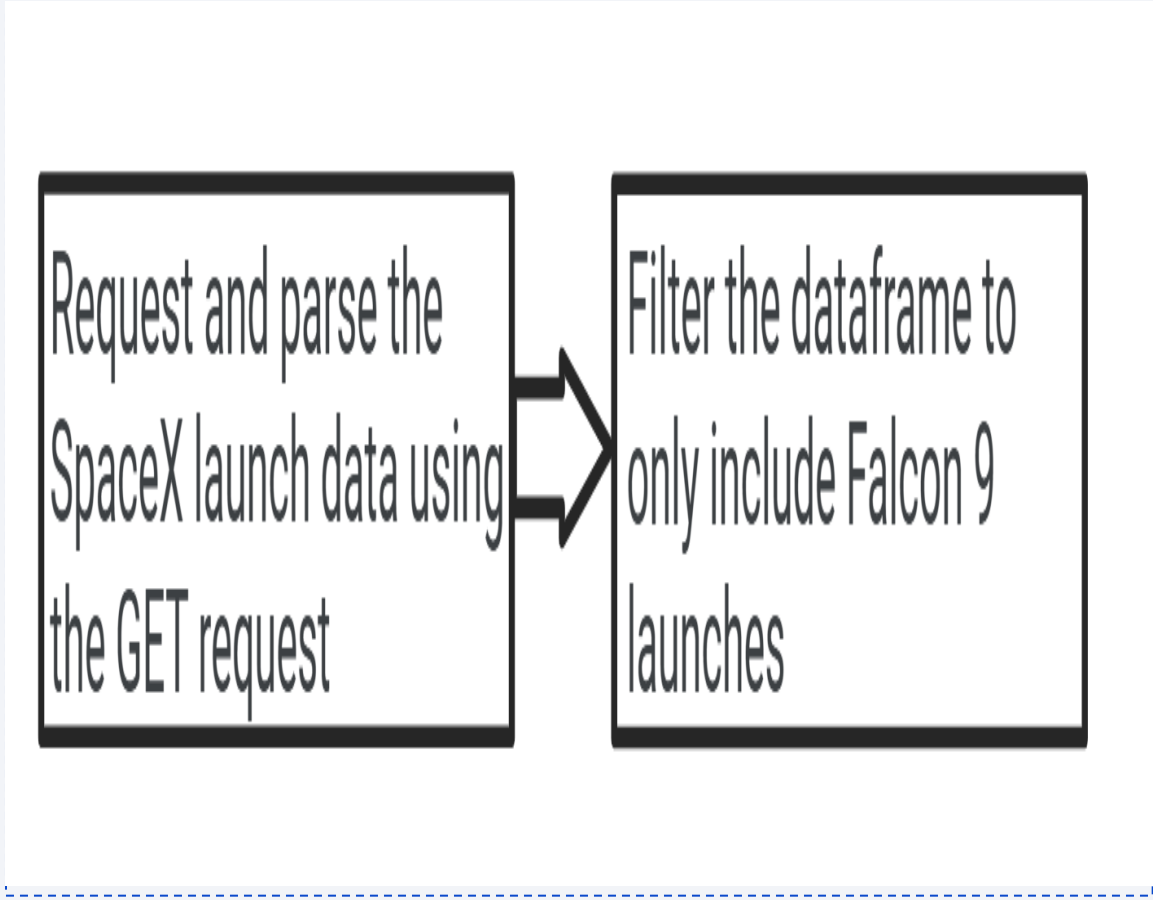
Data Collection – SpaceX API

- The data collection is composed of two steps:

1-Request and parse the SpaceX launch data using the GET request

2-Filter the dataframe to only include Falcon 9 launches

- https://github.com/nizar7702/project_capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

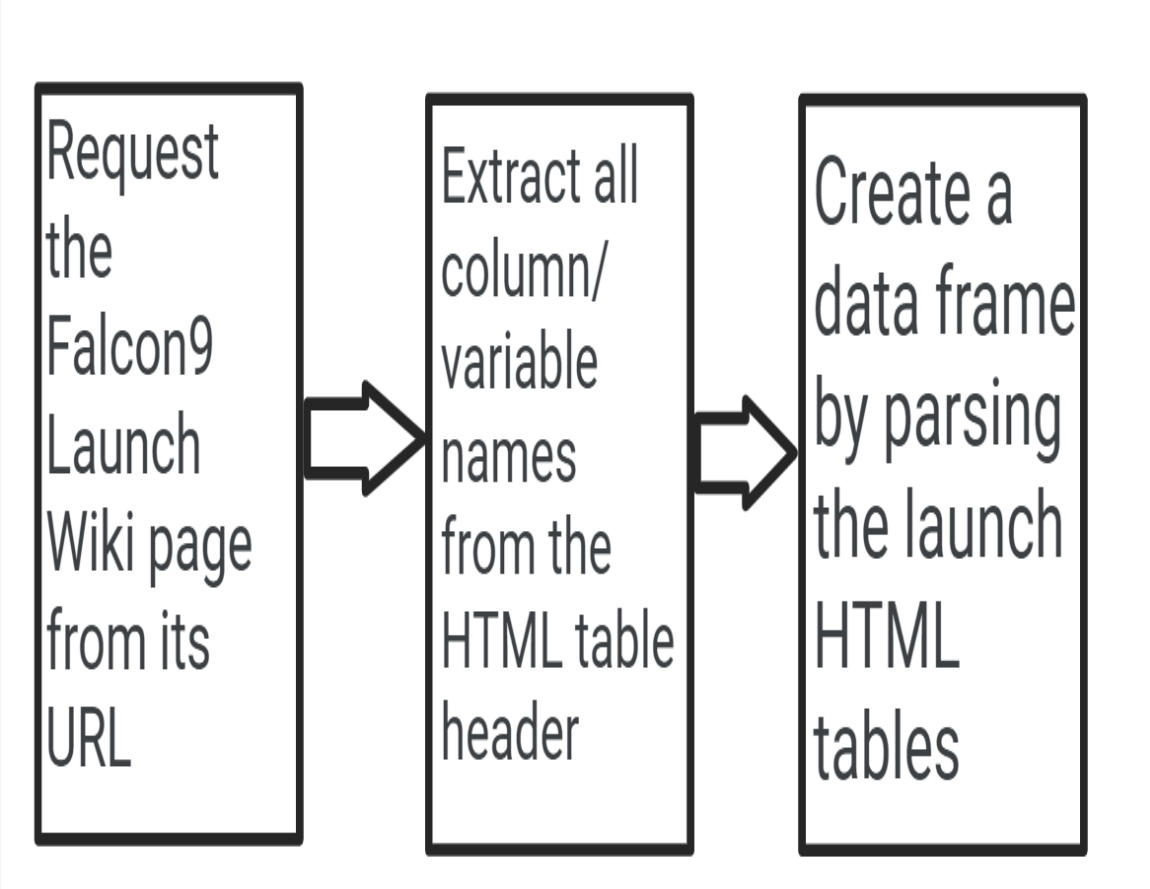
- Web scaping is composed of 3 steps:

1-Request the falcon9 Launch Wiki page from its URL

2-Extract all colun/variable/names from the html table header

3-Create a data frame by parsing the launch HTML tables

- https://github.com/nizar7702/project_capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Web scaping is composed of 5 steps:

1-Dealing with missing values

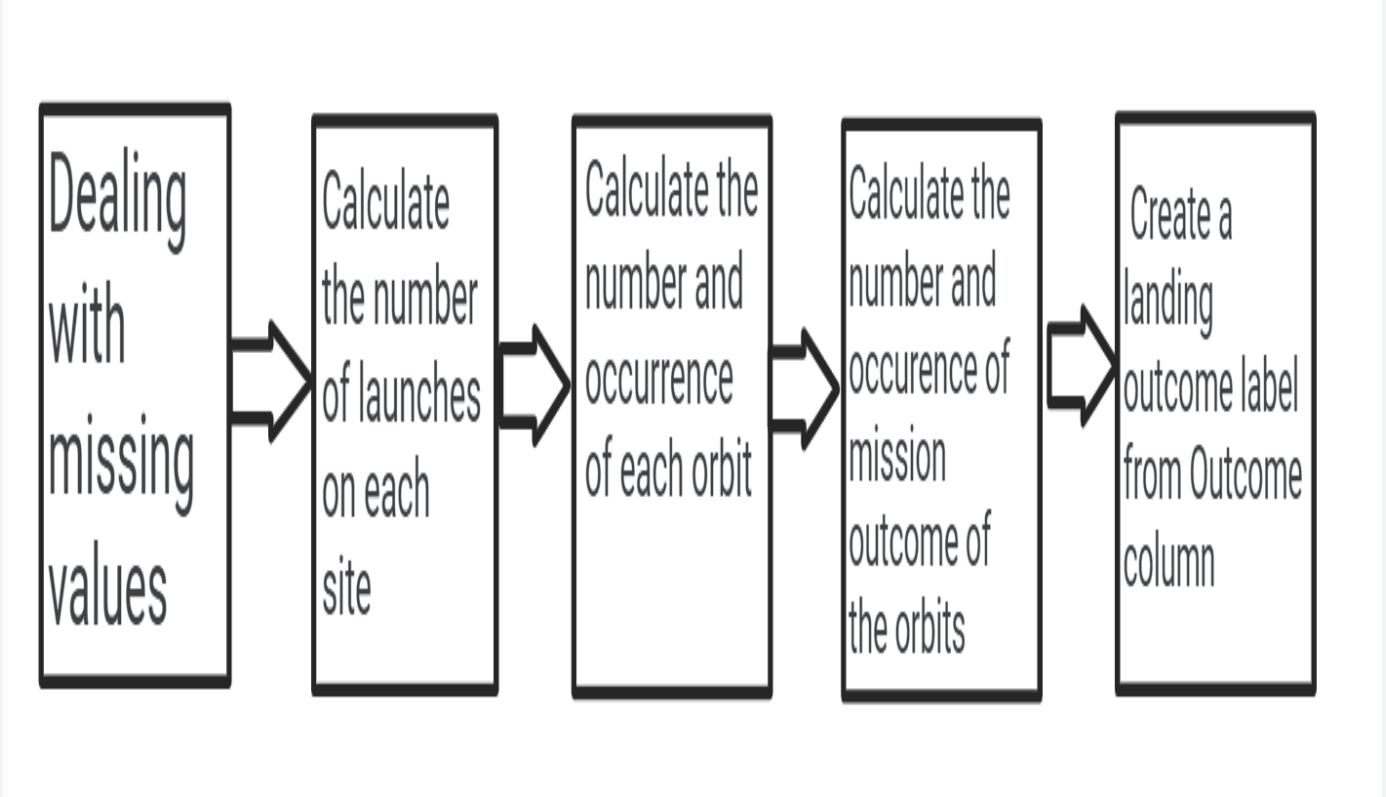
2-Calculate the number of launches on each site

3-Calculate the number and occurrence of each orbit

4-Calculate the number and occurrence of mission outcome of the orbits

5-Create landing outcome label from Outcome column

- https://github.com/nizar7702/project_capstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb



EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

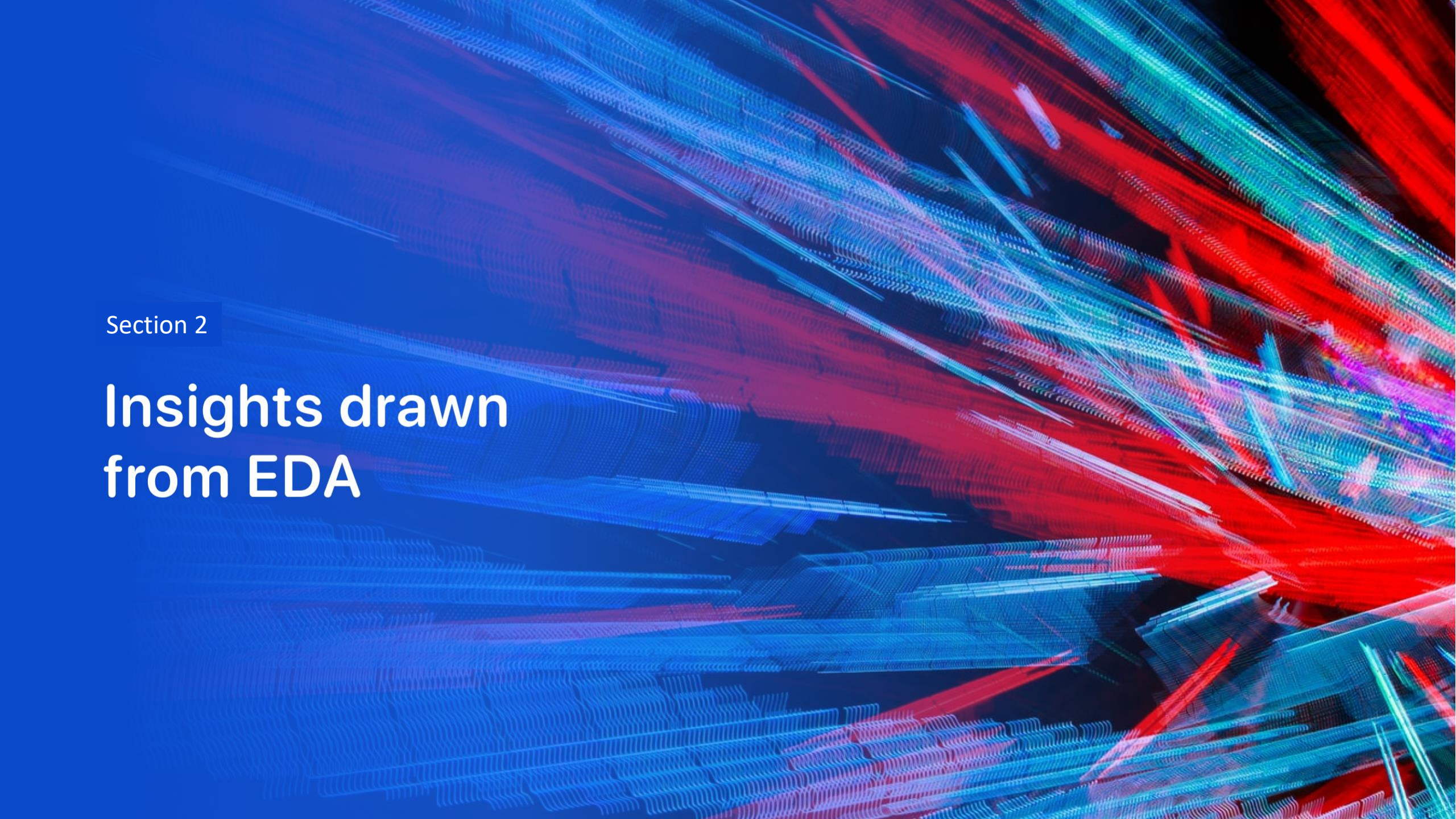
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

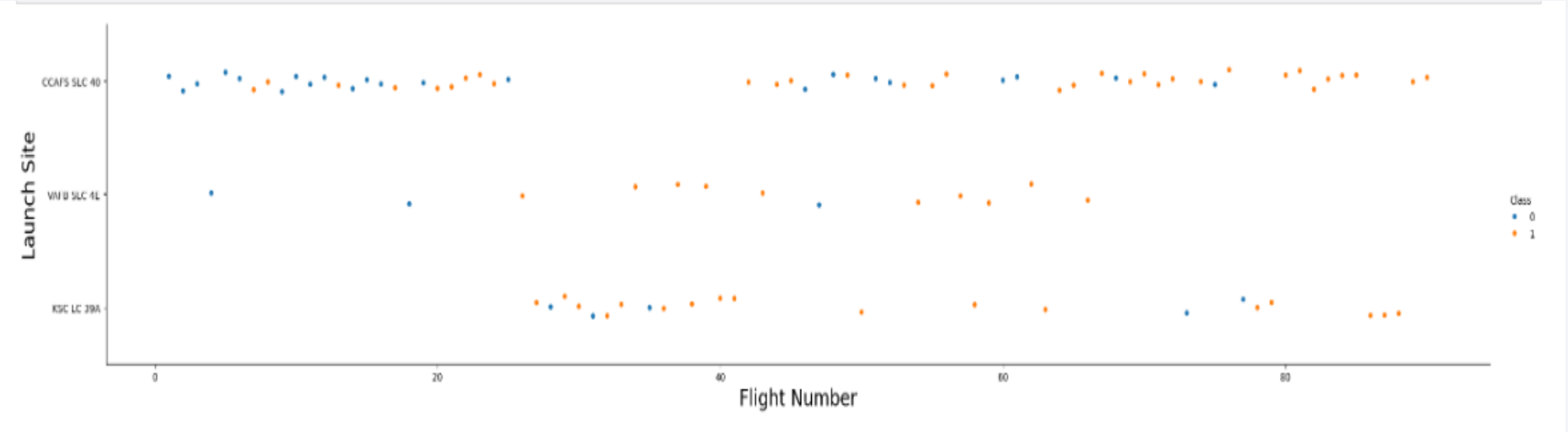
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

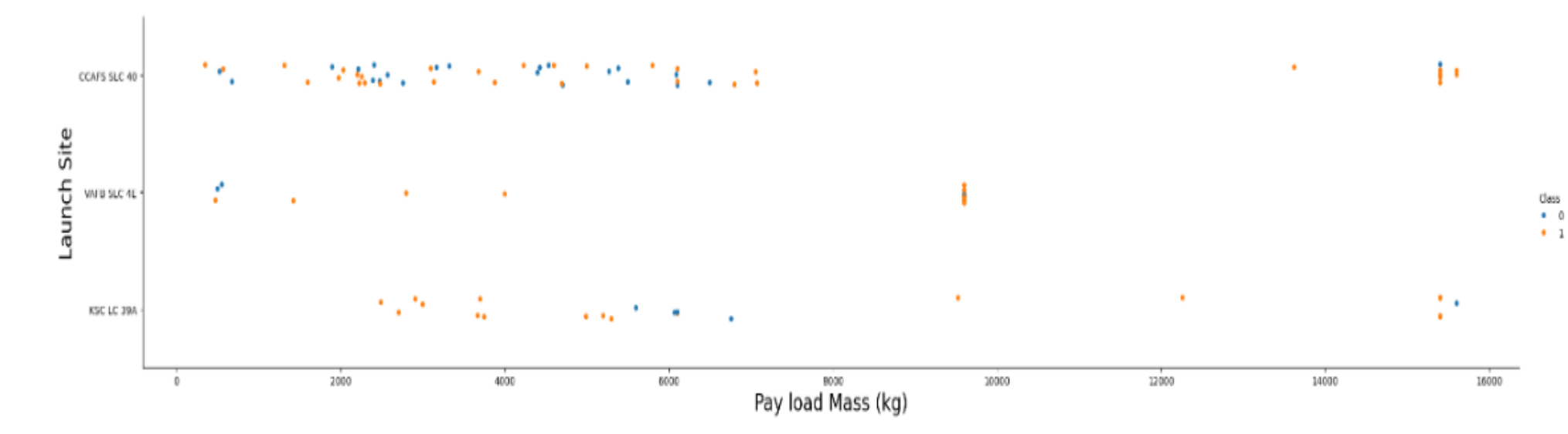
- we will see now how flightNumber and launch site variables would affect the launch outcome



- We see that the main factor that influences the outcome is the flight number

Payload vs. Launch Site

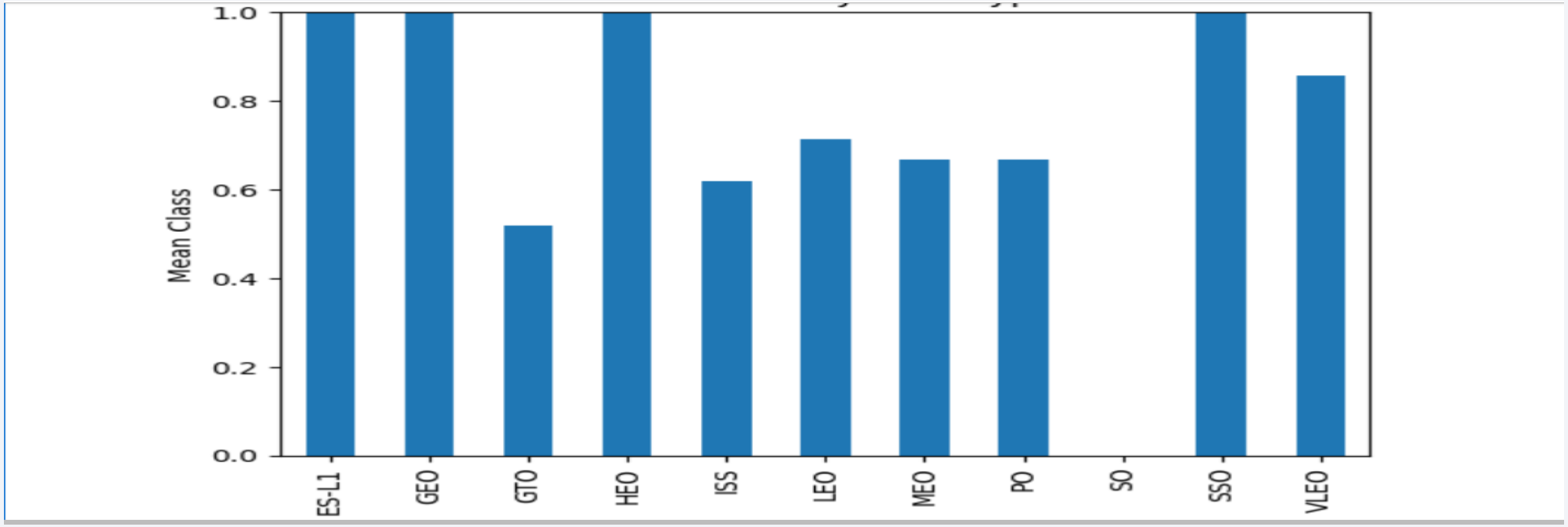
- we will see now how payloadmass and launch site variables would affect the launch outcome



- We see that for VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

Success Rate vs. Orbit Type

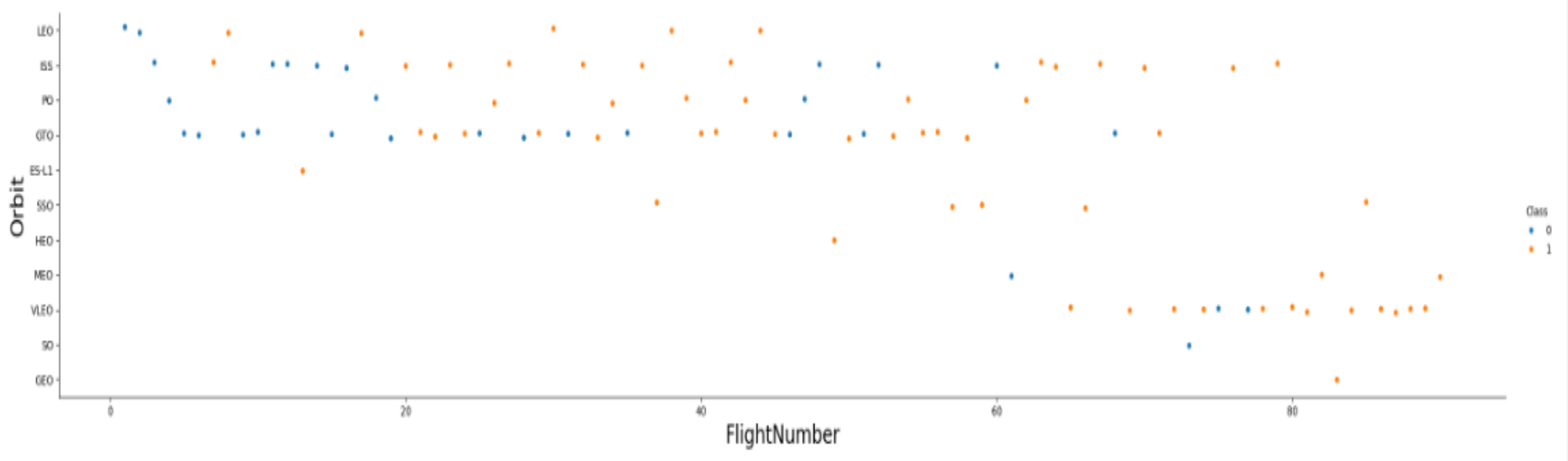
- we will see now how the orbit type would affect the success rate



- We see that some orbits type has more success rates than others

Flight Number vs. Orbit Type

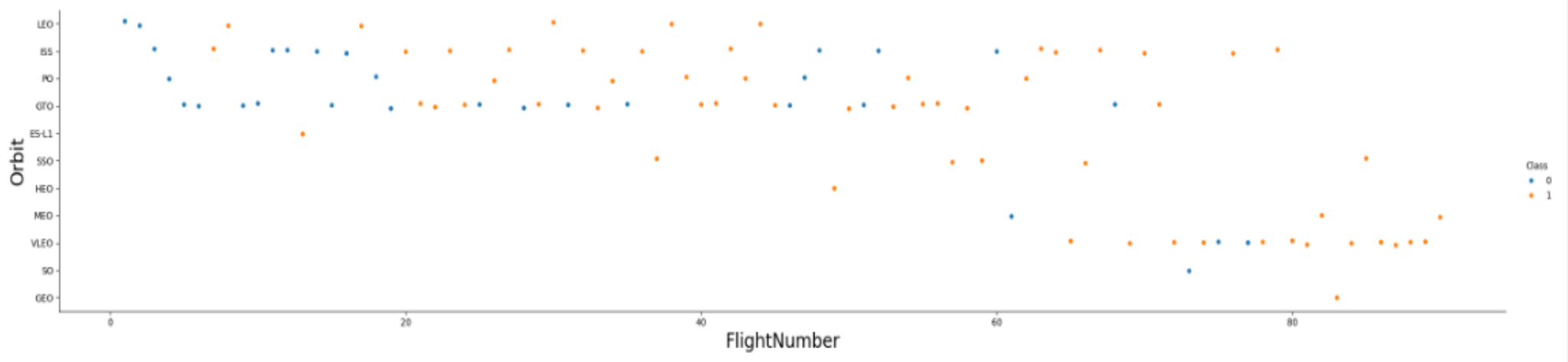
- we will see now how the flightnumber and orbit type variables would affect the launch outcome



- We see that LEO ORBIT seems to be related to the number of flights,conversely,in the GTO orbit,there appears to be no relationship between flight number and success

Payload vs. Orbit Type

- we will see now how the Payload Mass and orbit type variables would affect the launch outcome



- With heavy payloads the successful landing or positive landing rate are more for Polar,Leo and ISS
- However,for GTO,it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present

Launch Success Yearly Trend

- we will see now the launch success yearly trend

- We see that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- %sql select DISTINCT "Launch_Site" from SPACEXTABLE
- We see that there are 4 launch sites names

```
[11]: %sql select DISTINCT "Launch_Site" from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
```

[11]:

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;
- We see 5 records where laucn sites name start with ‘CCA’

[13]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

- %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE Customer LIKE 'NASA (CRS)';
- We see the total payload mass caried by boosters launched by Nasa (CRS)

```
[19]: %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE Customer LIKE 'NASA (CRS)';
 * sqlite:///my_data1.db
Done.
[19]: SUM("PAYLOAD_MASS__KG_")
                        45596
```

Average Payload Mass by F9 v1.1

- %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1';
- We see the average payload mass caried by booster version F9 v1.1

```
[20]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1';
 * sqlite:///my_data1.db
Done.
[20]: AVG("PAYLOAD_MASS__KG_")
      2928.4
```

First Successful Ground Landing Date

- %sql select MIN(DATE) from SPACEXTABLE
- We see the date when the first successful landing outcome in ground pad was achieved

```
[21]: %sql select MIN(DATE) from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
[21]: MIN(DATE)
      2010-06-04
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select Booster_Version from SPACEXTABLE where (PAYLOAD_MASS__KG_>4000) AND (PAYLOAD_MASS__KG_<6000)
- We see the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[22]: %sql select Booster_Version from SPACEXTABLE  where (PAYLOAD_MASS__KG_>4000) AND (PAYLOAD_MASS__KG_<6000)
 * sqlite:///my_data1.db
Done.
```

[22]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1

Total Number of Successful and Failure Mission Outcomes

- %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome="Success“
- %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome!="Success"
- We see the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
[24]: %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome="Success"
 * sqlite:///my_data1.db
Done.
```

[24]:

count(Mission_Outcome)
98

```
[25]: %sql select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome!="Success"
 * sqlite:///my_data1.db
Done.
```

[25]:

count(Mission_Outcome)
3

Boosters Carried Maximum Payload

- %sql select Distinct Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
- We see the names of the booster_versions which have caried the maximum payload mass

```
[28]: %sql select Distinct Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG
 * sqlite:///my_data1.db
Done.
[28]:
```

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

2015 Launch Records

- %sql select substr(Date, 6,2),Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015')
- We see the list of records which will display the month names,failure landing_outcome in drone ship,booster versions,launch_site for the months in year 2015

```
[30]: %sql select substr(Date, 6,2),Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE  where substr(D
```

```
 * sqlite:///my_data1.db
Done.
```

[30]:

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

- %sql SELECT Landing_Outcome, COUNT(*) AS outcome_count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY outcome_count DESC;
- We see the rank of the count of landing outcomes (such as failures (drone ship) or Success (ground pad) between the date 2010-06-04 and 2017-03-20 in descending order

```
[32]: %sql SELECT  Landing_Outcome, COUNT(*) AS outcome_count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' A
```

```
 * sqlite:///my_data1.db
Done.
```

[32]:

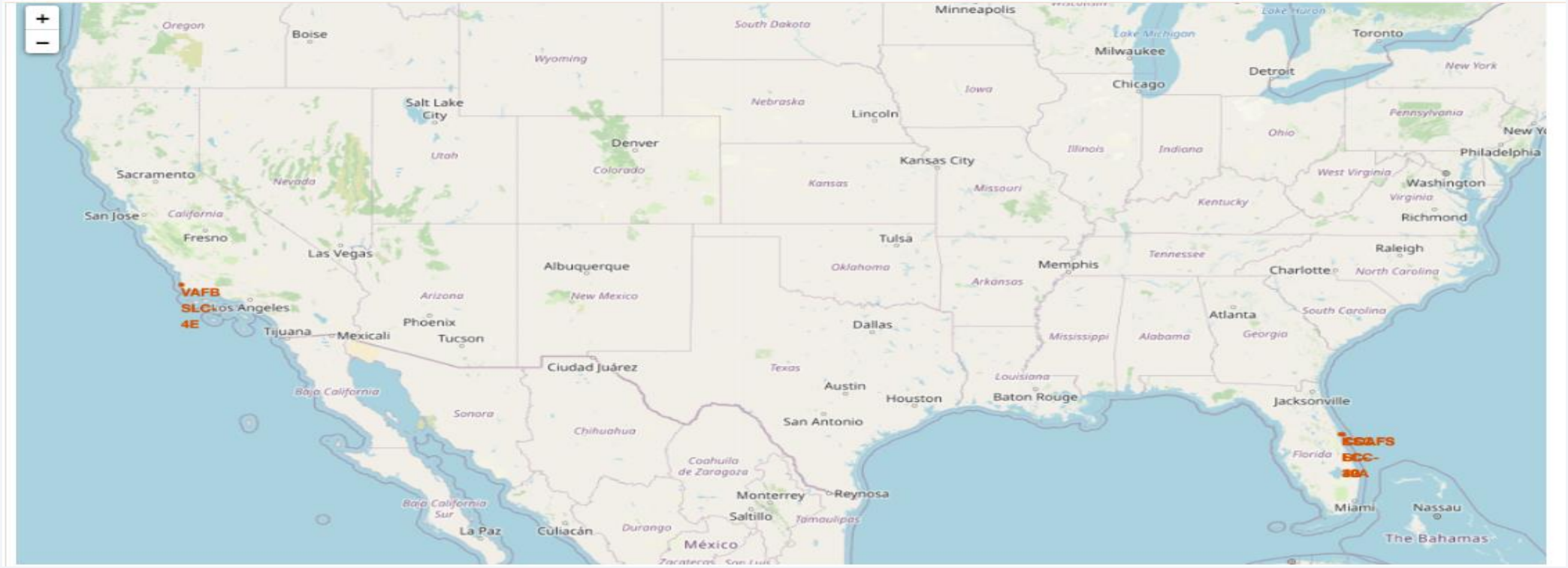
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

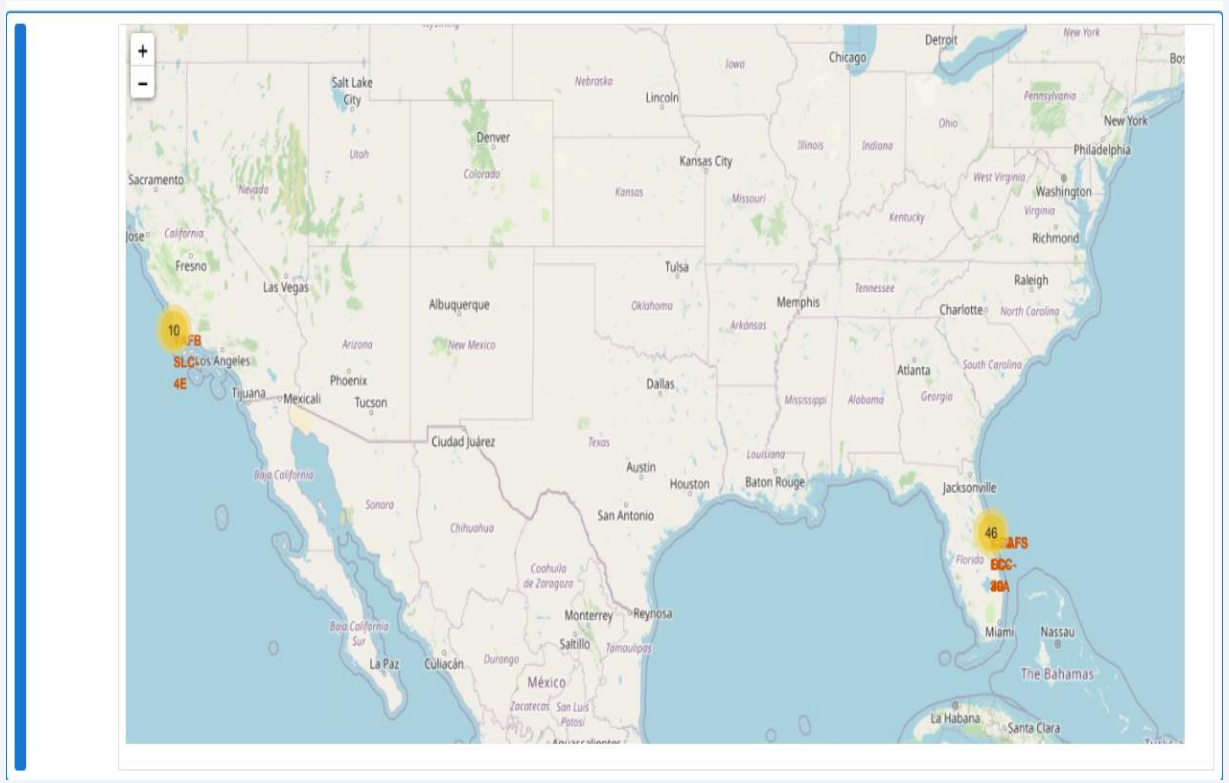
Launch Sites Proximities Analysis

Launch sites location on the map

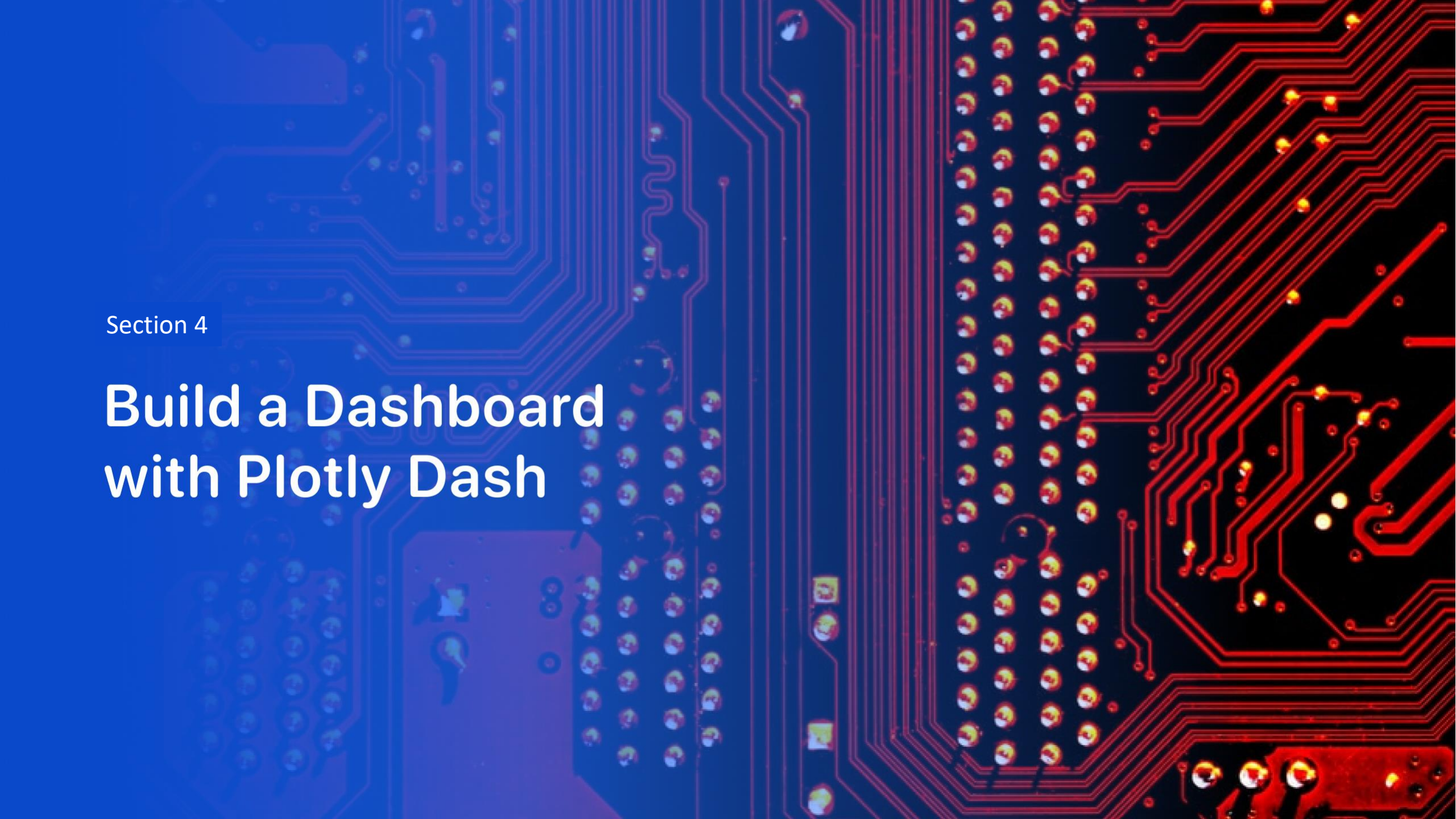
We see that some sites are on the west and others on the east.



Markers creation for launch records

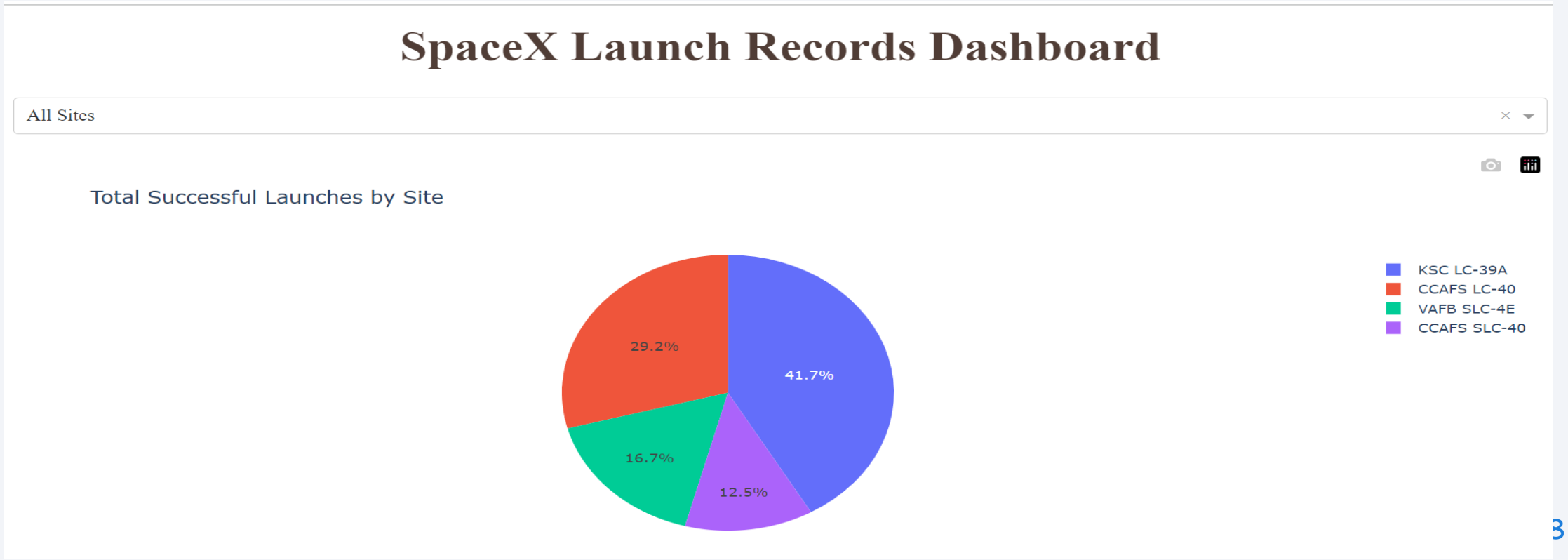


The distance between launche sites and coast lines


Section 4
Build a Dashboard with Plotly Dash

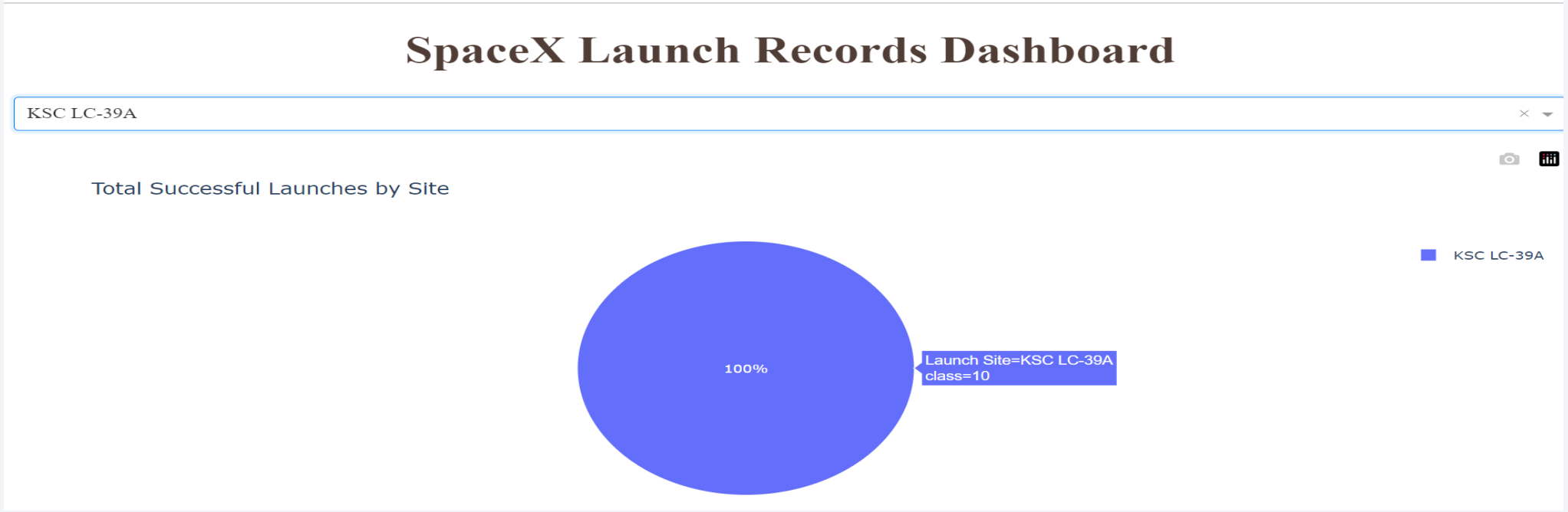
Launch success count for all sites

- We see that KSC LC-39A has the mpst successful lanches among all sites



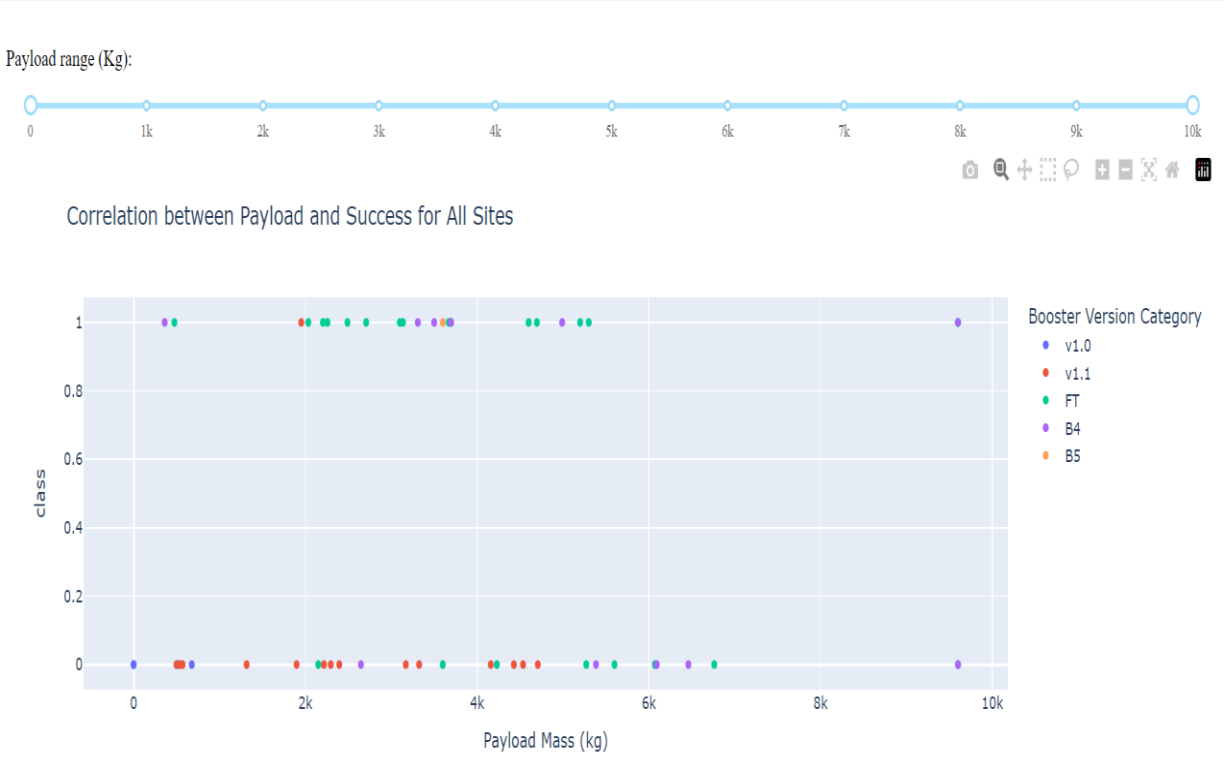
Launch site with highest success ratio

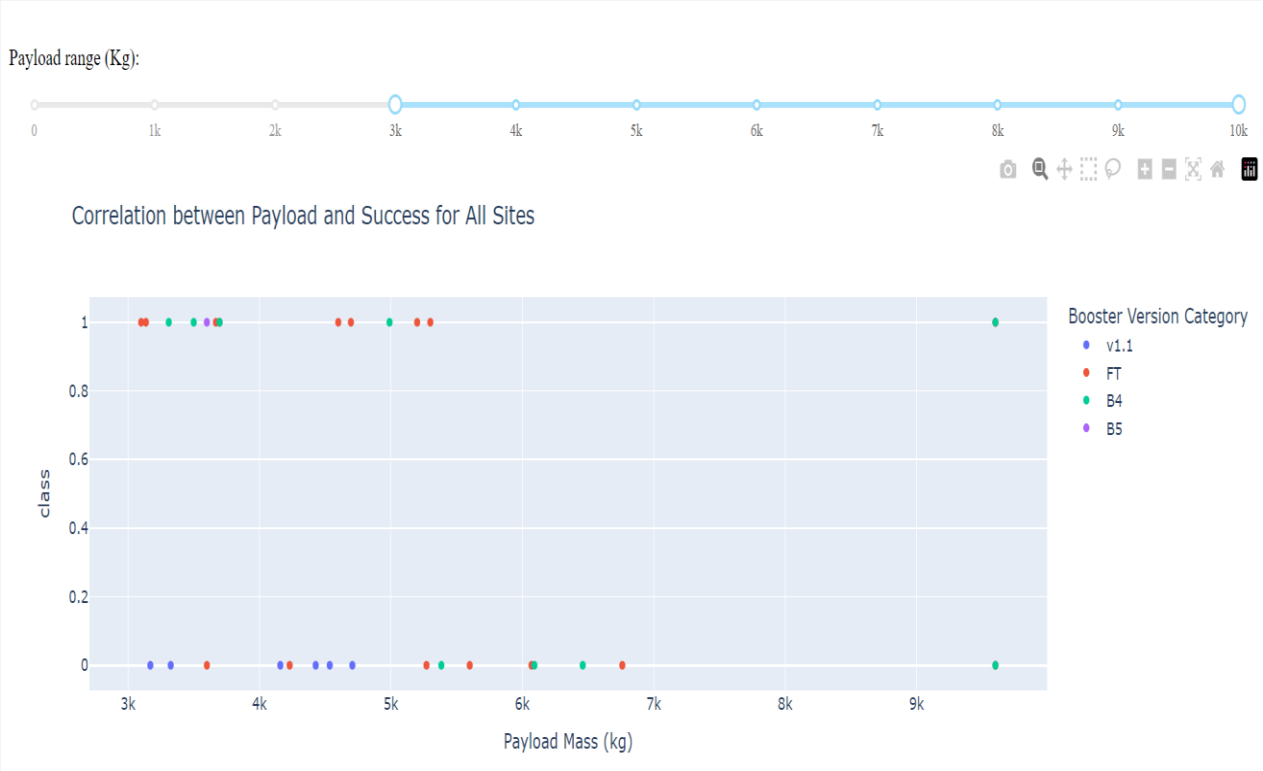
- Here we see the launch site with highest success ratio



Payload vs Launch Outcome with different ranges

We see that payload range from 0k to 10000k has a more success rate than between 3k and 10k and that's completely normal



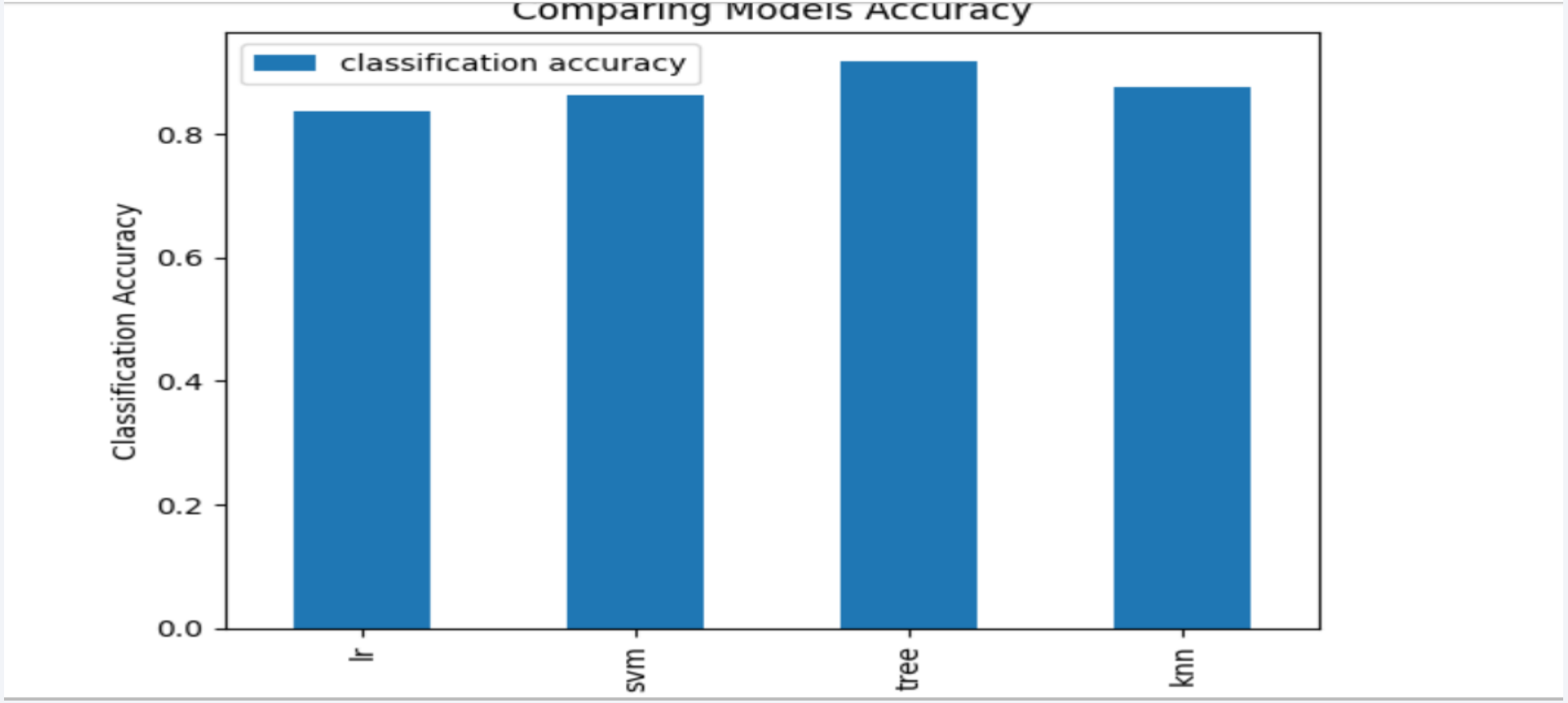


Section 5

Predictive Analysis (Classification)

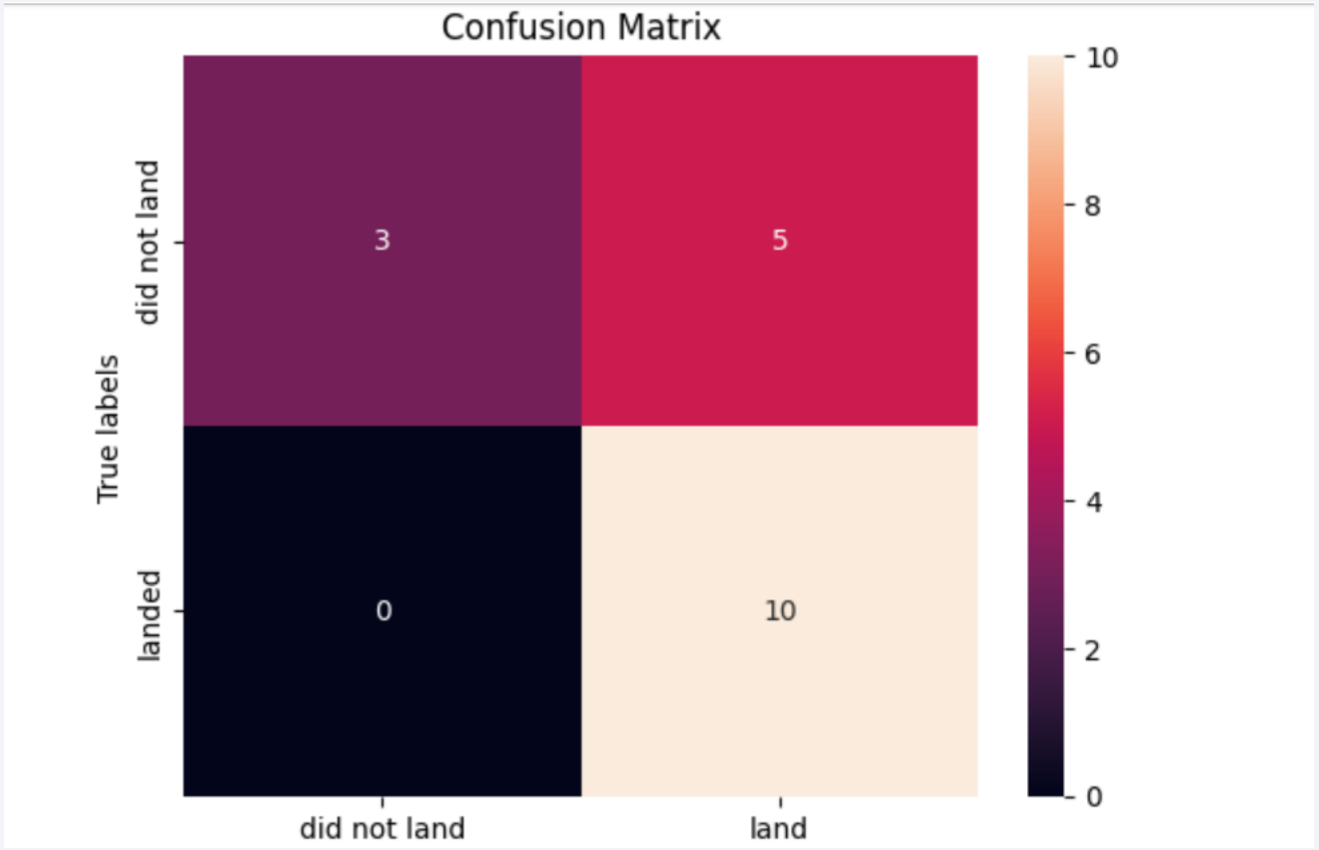
Classification Accuracy

- We see that the decision tree methos has the best classification accuracy



Confusion Matrix

- Here the true positive=10,the true negative=3,the false positive=5,the false negative=



Conclusions

- We see that the model have a sum of 13 of true positive and true negative which is a good sign
- The model has only 5 misclassification which is not that much compared to the total
- This model is quite good in predicting

Appendix

- https://github.com/nizar7702/project_capstone/tree/main
- This is the github link for all the labs shared in this presentation

Thank you!