IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rahul Gopal
2022-11-15

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Falcon 9 launch details were extracted from Wikipedia.
- Perform data wrangling
  - Filtered the data using the Boot Version column to only keep the Falcon 9 launches
- 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 – SpaceX API

- Data can be collected from SpaceX API.
- SpaceX-API is a open-source REST API for rocket, core, capsule, pad and launch data.
- The flowchart for the data collection is shown in figure 1.

- GitHub - Link

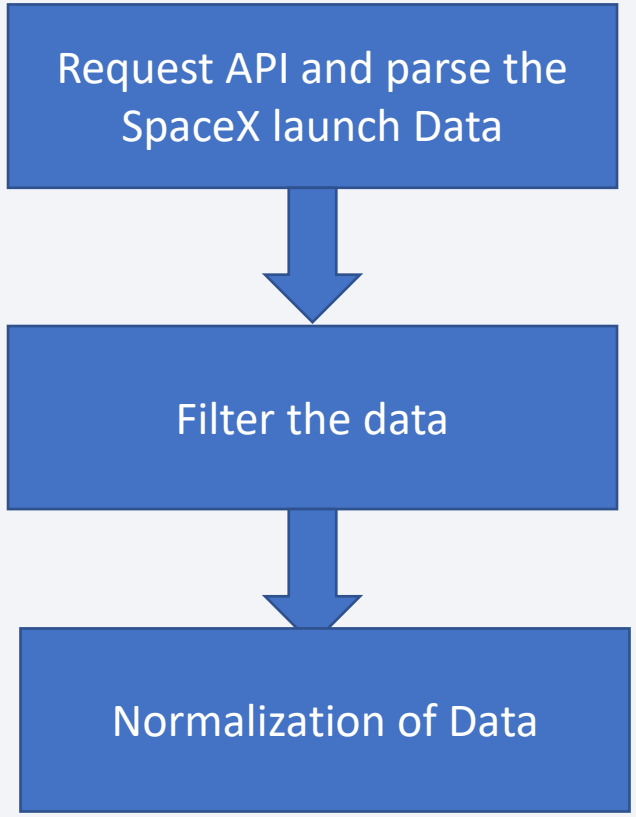


Figure 1:Flow chart of data collection

Data Collection - Scraping

- Web scraping is the process of using bots to extract the content and data from website.
- Web scraping extracts the underlying HTML code and stores in a database.
- The flowchart for the data collection is shown in figure 1.
- GitHub link

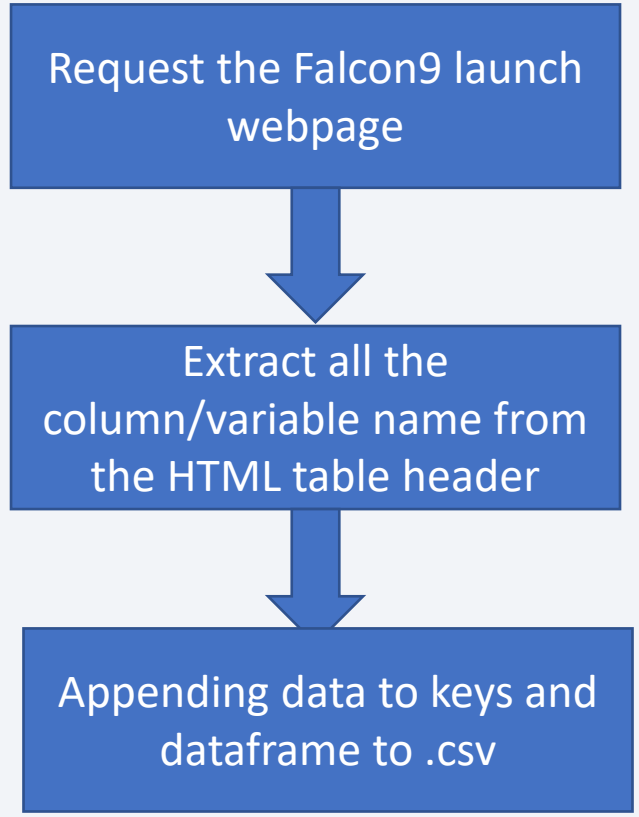
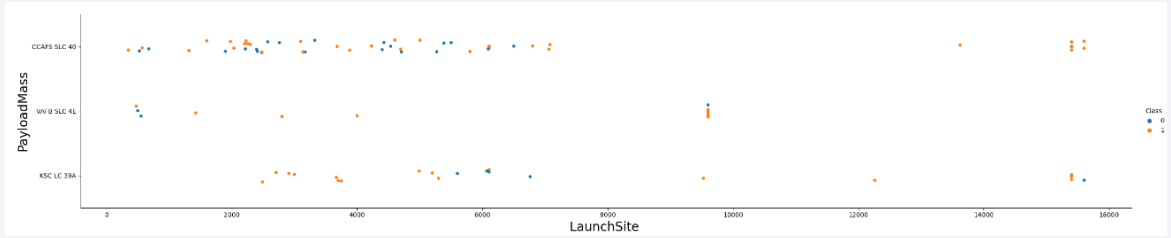


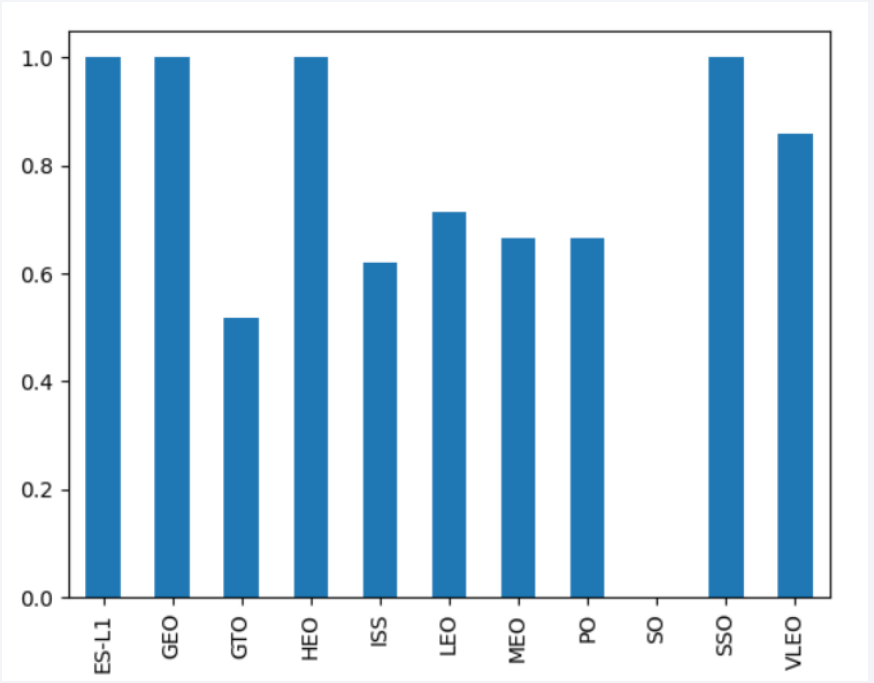
Figure 1:Flow chart of data collection

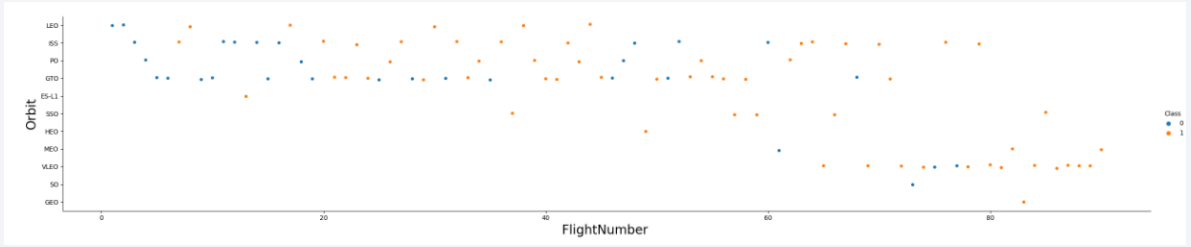
Data Wrangling

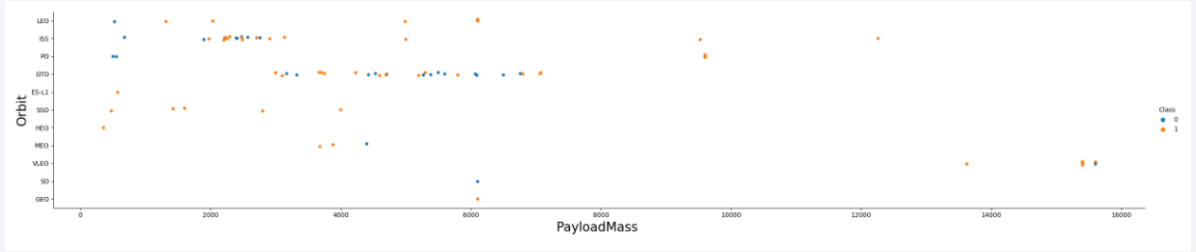
- Check Null values
- Calculate the number of launch on each site
- Calculate number of orbit
- Calculate the no. of occurrence and mission outcome per orbit type
- Create a landing outcome column
- Handle null values

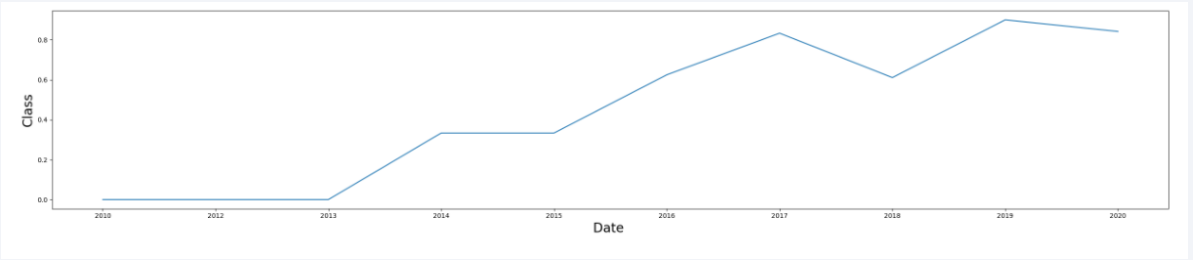
EDA with Data Visualization











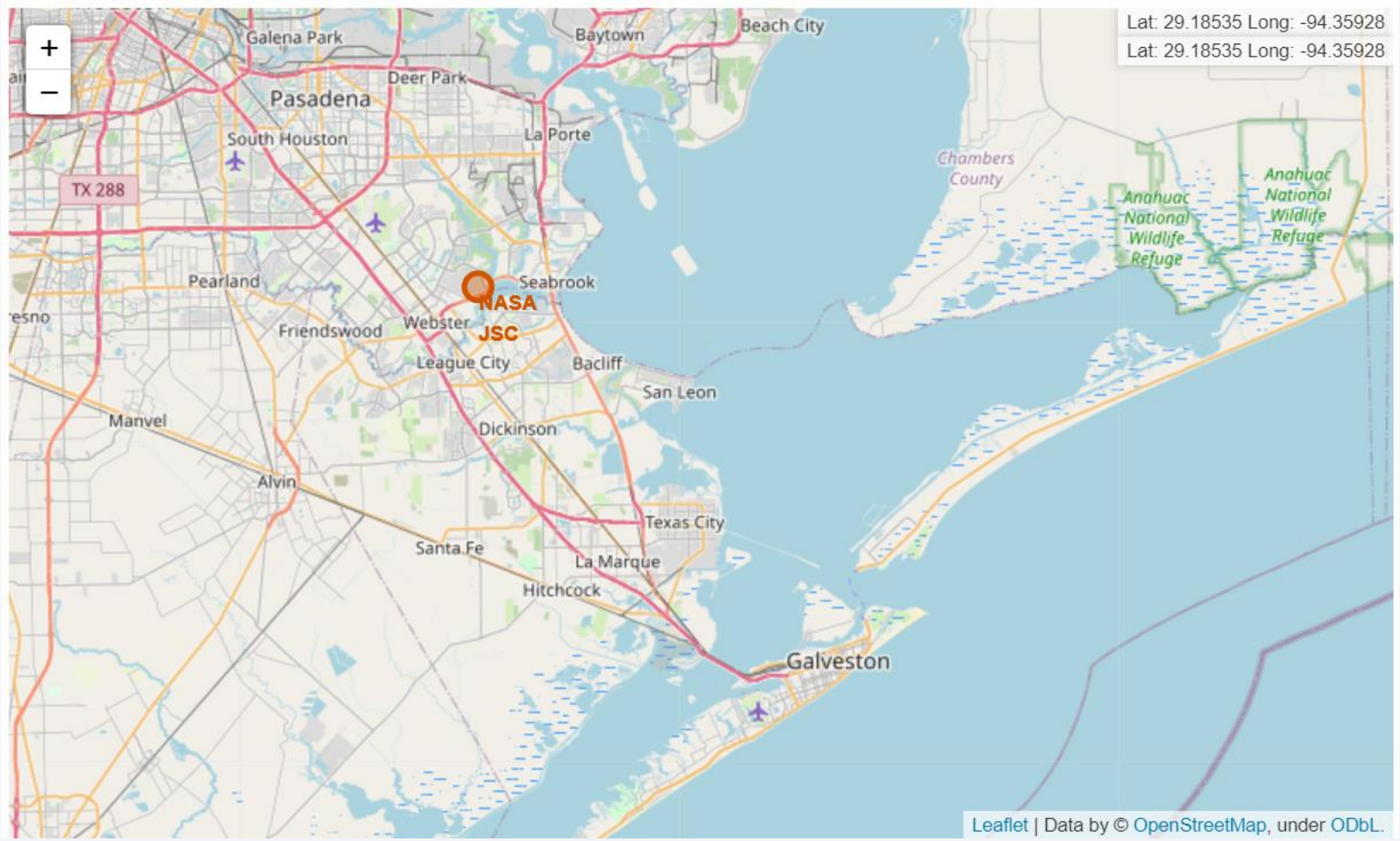
https://github.com/Rahul-Gopal/IBM-Data-Science-Final-Project/blob/f13aaa2a168eb31458e402636760876d7a85d7b0/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

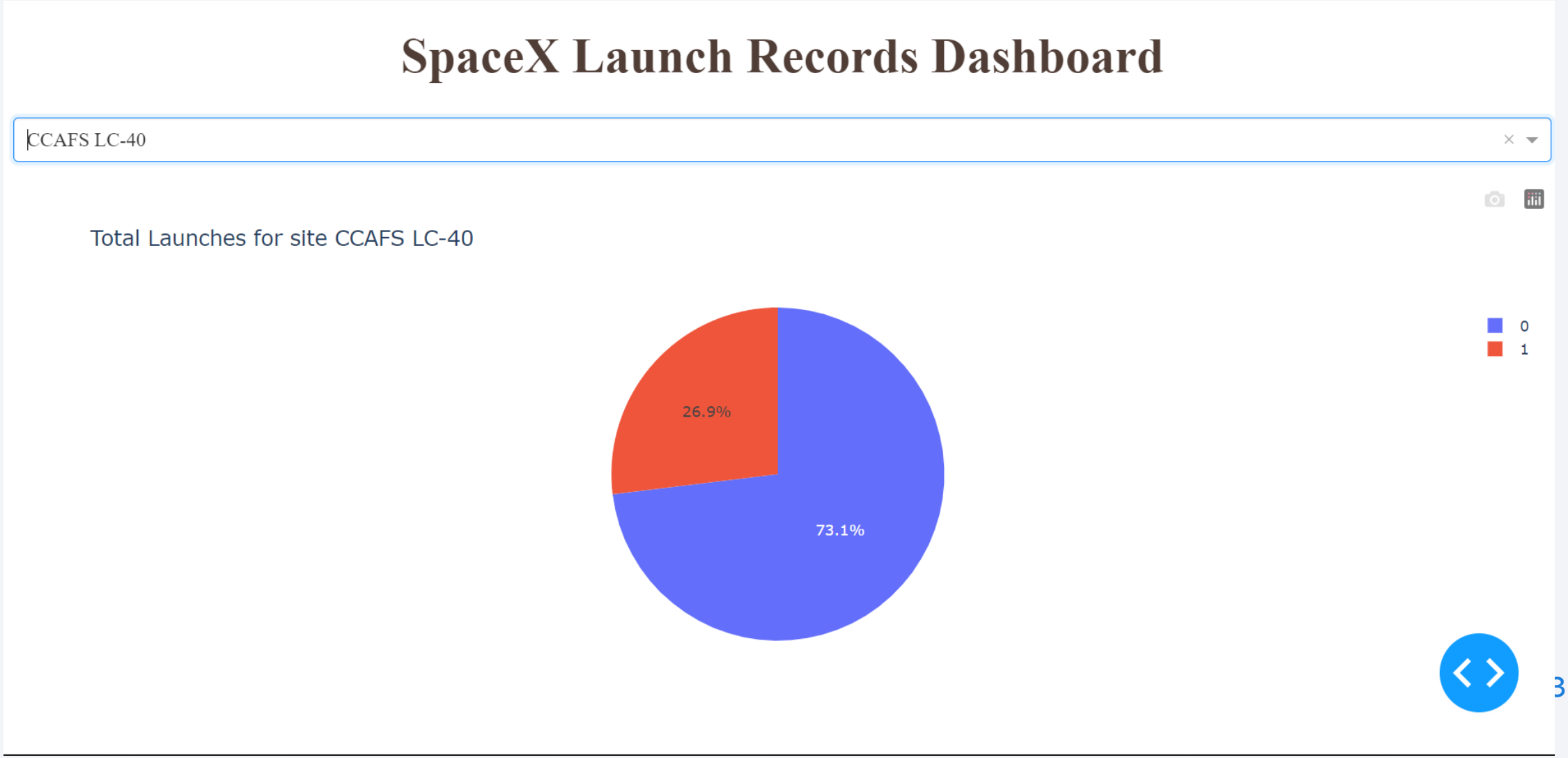
- Following query were performed
  - Display the names of the unique launch sites in the space mission
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- https://github.com/Rahul-Gopal/IBM-Data-Science-Final-Project/blob/f13aaa2a168eb31458e402636760876d7a85d7b0/EDA-SQL.ipynb

Build an Interactive Map with Folium

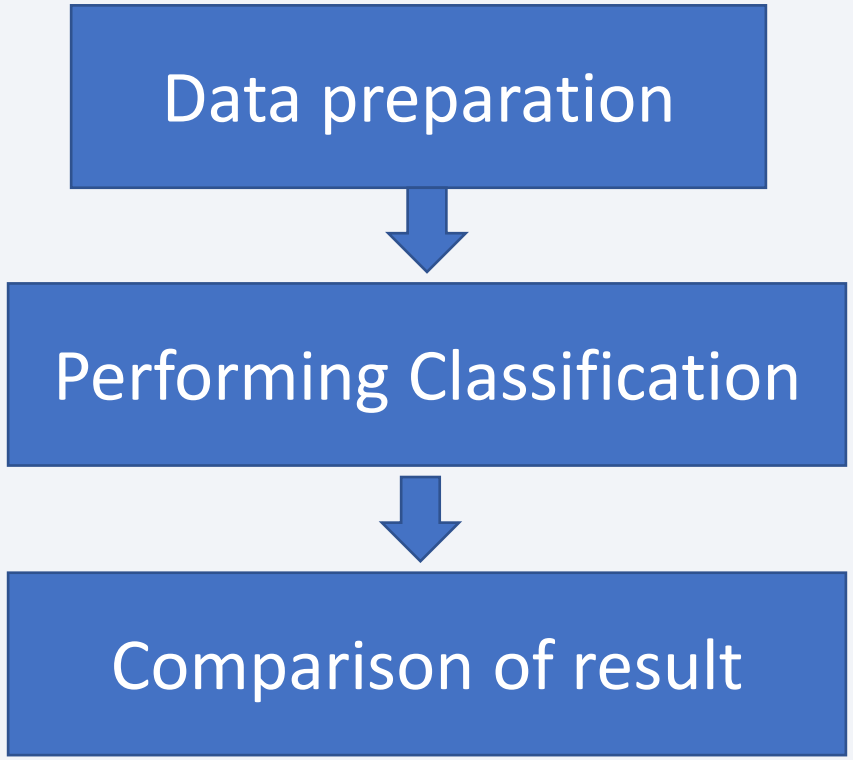


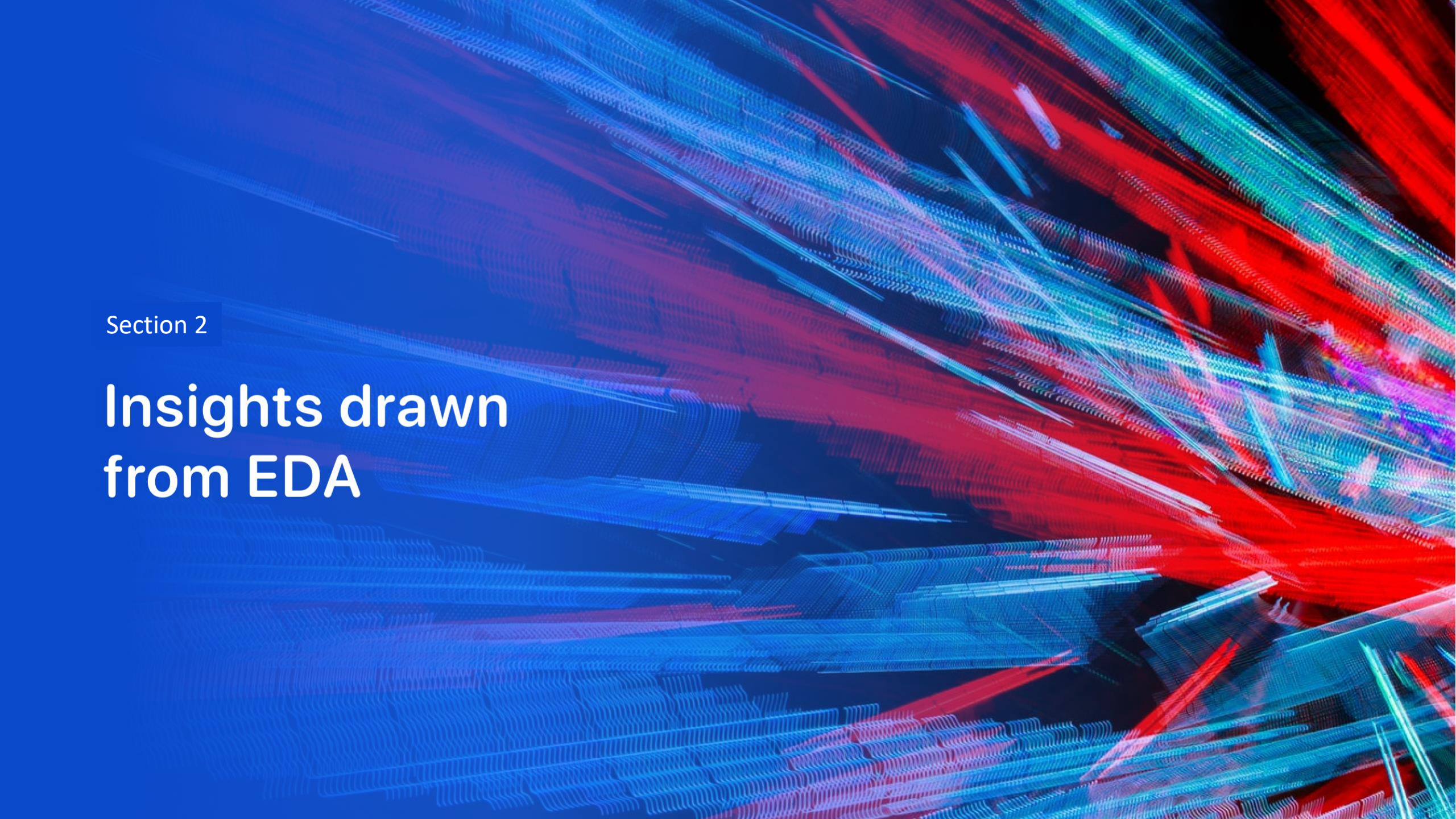
Build a Dashboard with Plotly Dash



Predictive Analysis (Classification)

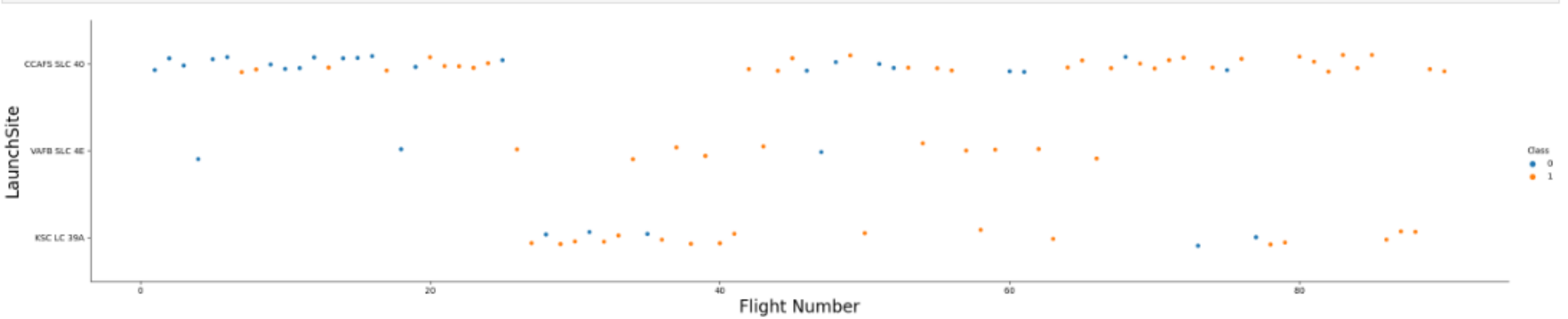
- Logistic regression , support vector machine, decision tree and k nearest neighbors were performed on the dataset and results were compared.




Section 2
Insights drawn from EDA

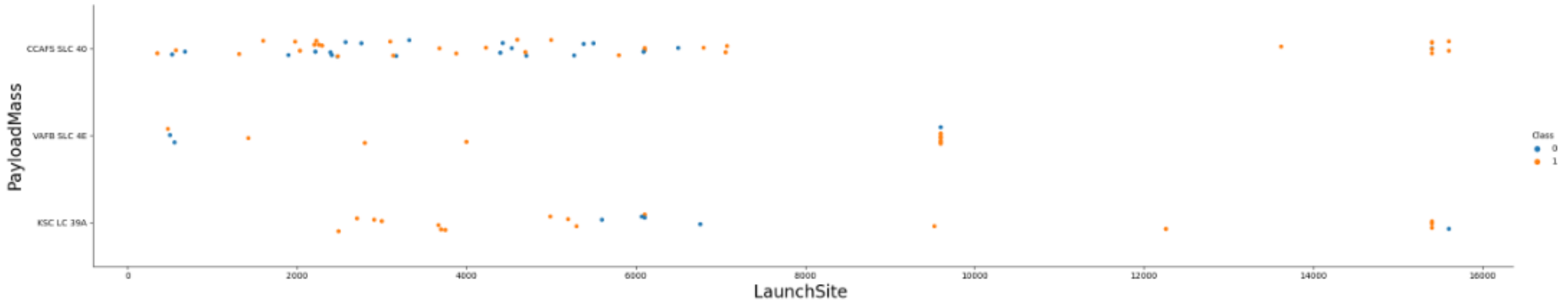
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site



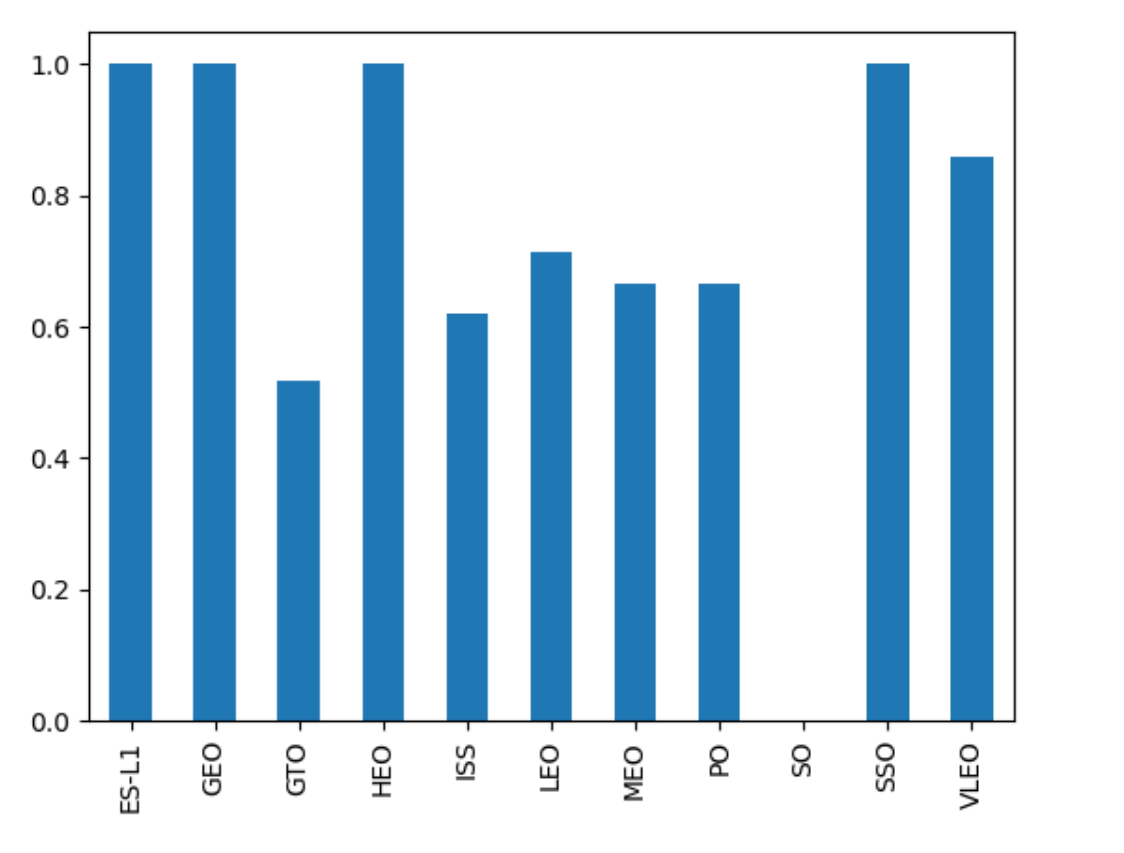
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



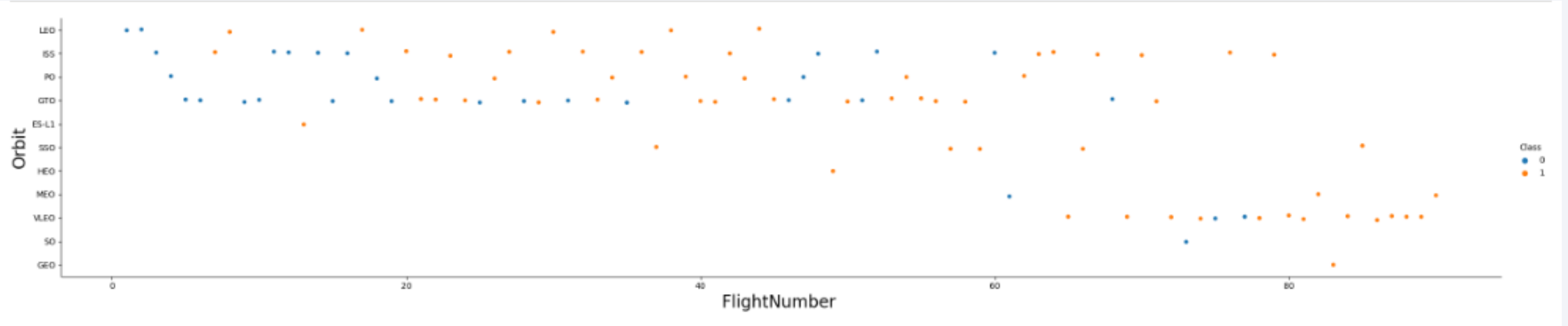
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type



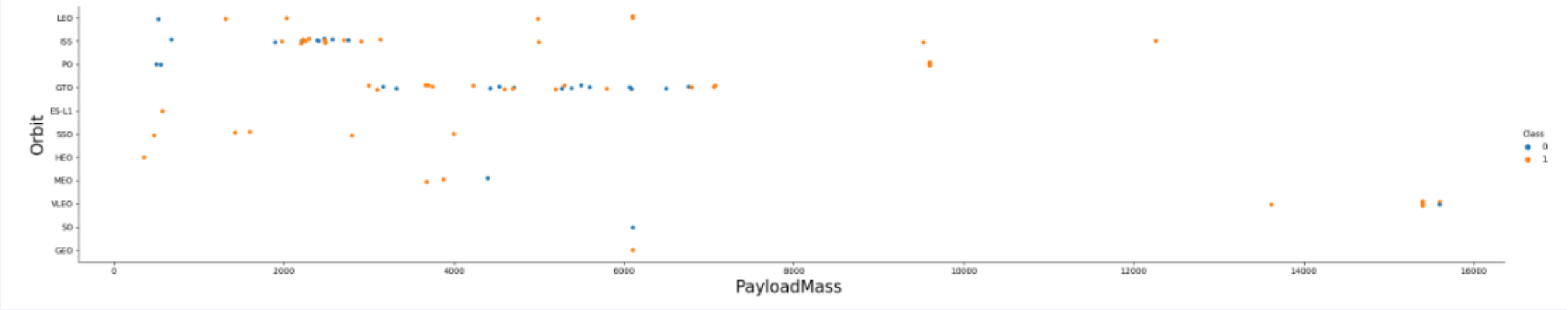
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type



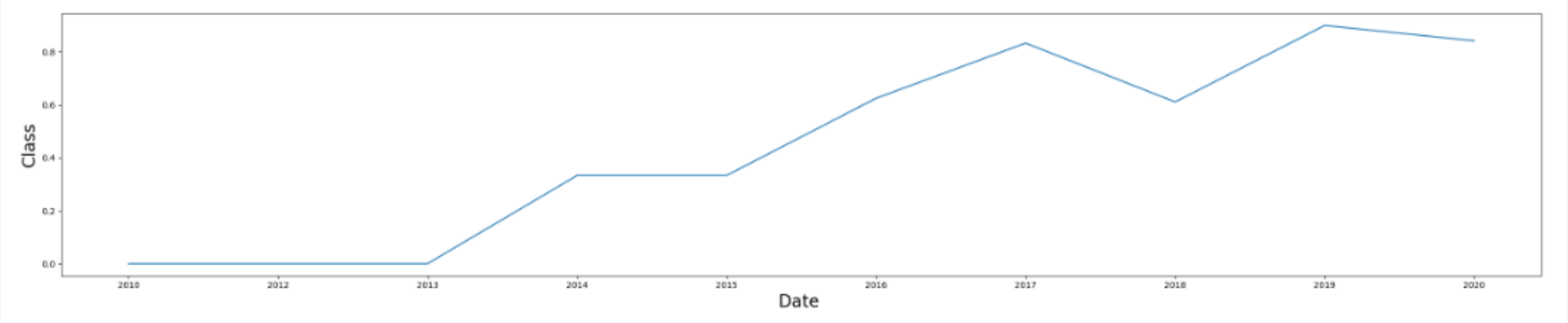
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Show a line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch site

Task 1

Display the names of the unique launch sites in the space mission

In [7]:
```
sql select DISTINCT(Launch_Site) from SPACEXTBL
```

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

Out[7]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

Task 2

Display 5 records where launch sites begin with the string 'CCA'

In [9]:
```
sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
```

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

Out[9]:

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

- Calculate the total payload carried by boosters from NASA

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

In [9]:
```
sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)'
```

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

Out[9]:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

Task 4

Display average payload mass carried by booster version F9 v1.1

In [10]:
```
sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like 'F9 v1.1'
```

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

Out[10]:

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

In [30]:
```
sql select min(Date) from SPACEXTBL where LANDING__OUTCOME = 'Success%'
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: select min(Date) from SPACEXTBL where LANDING__OUTCOME = 'Success%']
(Background on this error at: http://sqlalche.me/e/e3q8)
```

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

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [40]:
```
sql select Booster_Version from SPACEXTBL where payload_mass__kg_ > 4000 and payload_mass__kg_< 6000
```

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

Out[40]:

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
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

In [28]:
```
sql select count(Mission_Outcome),Mission_Outcome from SPACEXTBL group by Mission_Outcome
```

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

Out[28]:

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

Boosters Carried Maximum Payload

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

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [44]:
```
sql select booster_version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

Out[44]:

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

2015 Launch Records

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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

In [47]:
```
sql select substr(Date,4,2),BOOSTER_VERSION,LAUNCH_SITE  from SPACEXTBL where substr(Date,7,4)='2015'
```

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

Out[47]:

substr(Date,4,2)	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
02	F9 v1.1 B1013	CCAFS LC-40
03	F9 v1.1 B1014	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40
04	F9 v1.1 B1016	CCAFS LC-40
06	F9 v1.1 B1018	CCAFS LC-40
12	F9 FT B1019	CCAFS LC-40

Task 10

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

Task 10

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

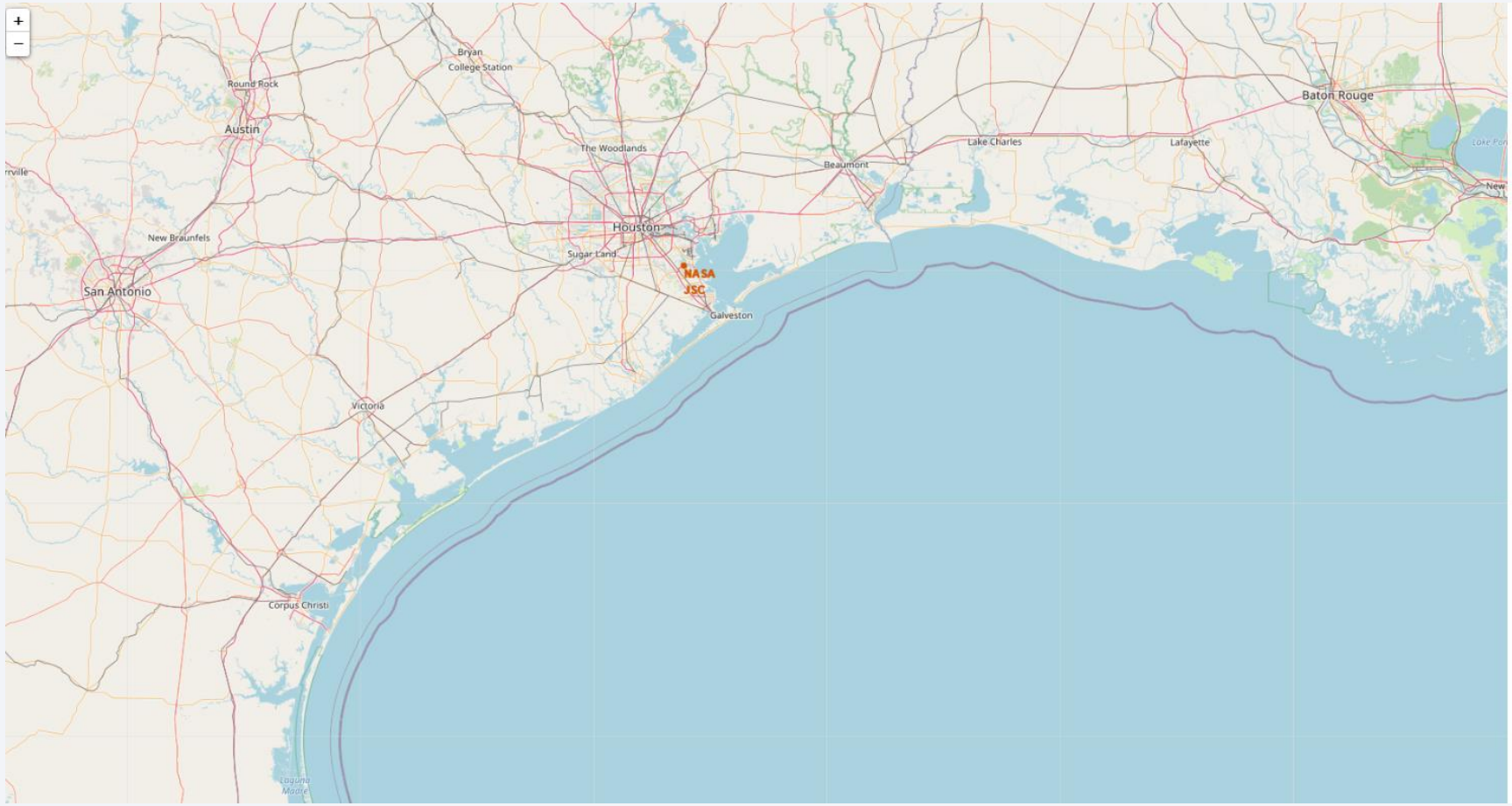
In [48]:
```
sql SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY QTY DE
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY QTY
DESC;]
(Background on this error at: http://sqlalche.me/e/e3q8)
```

Section 3

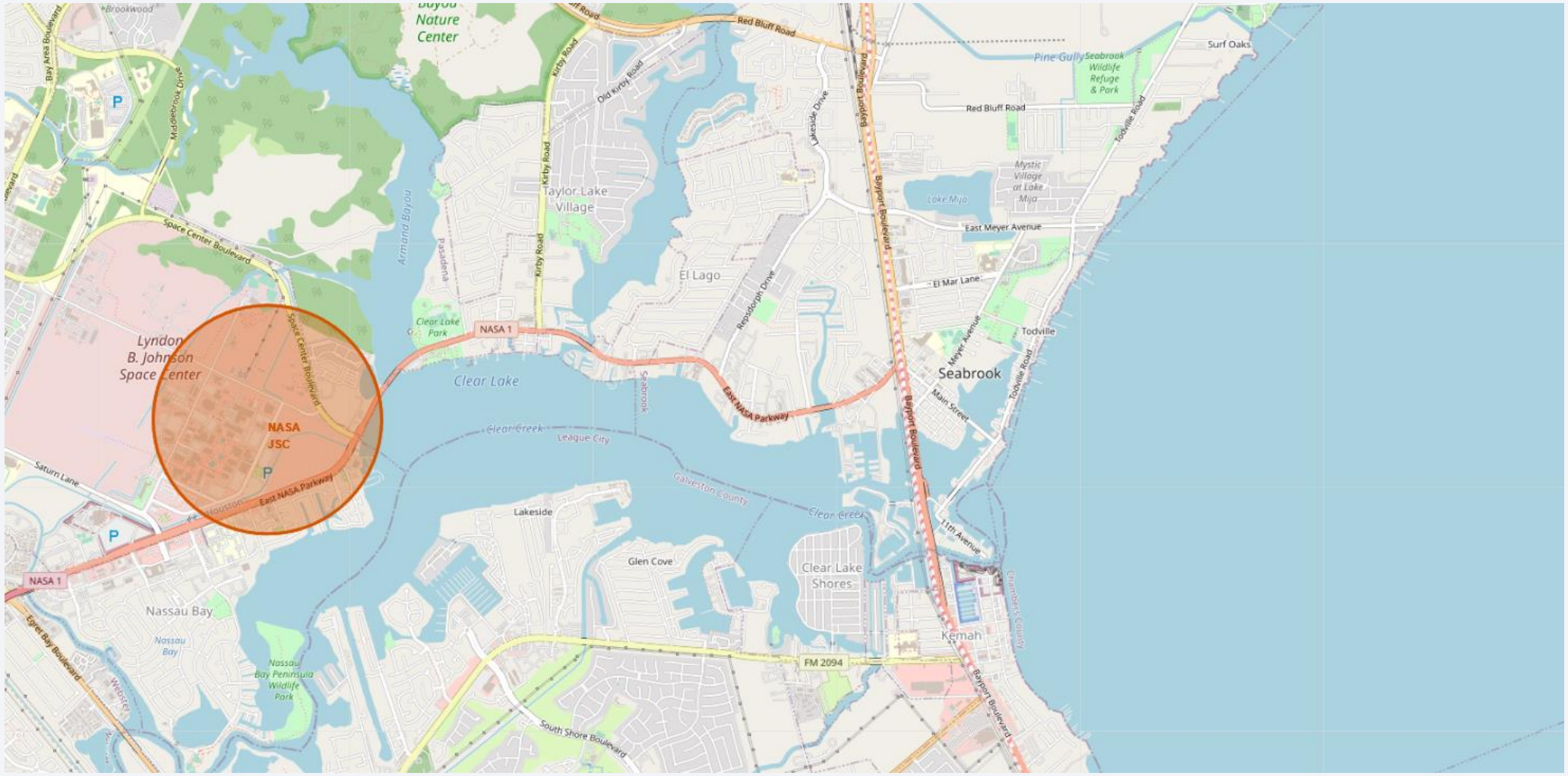
Launch Sites Proximities Analysis

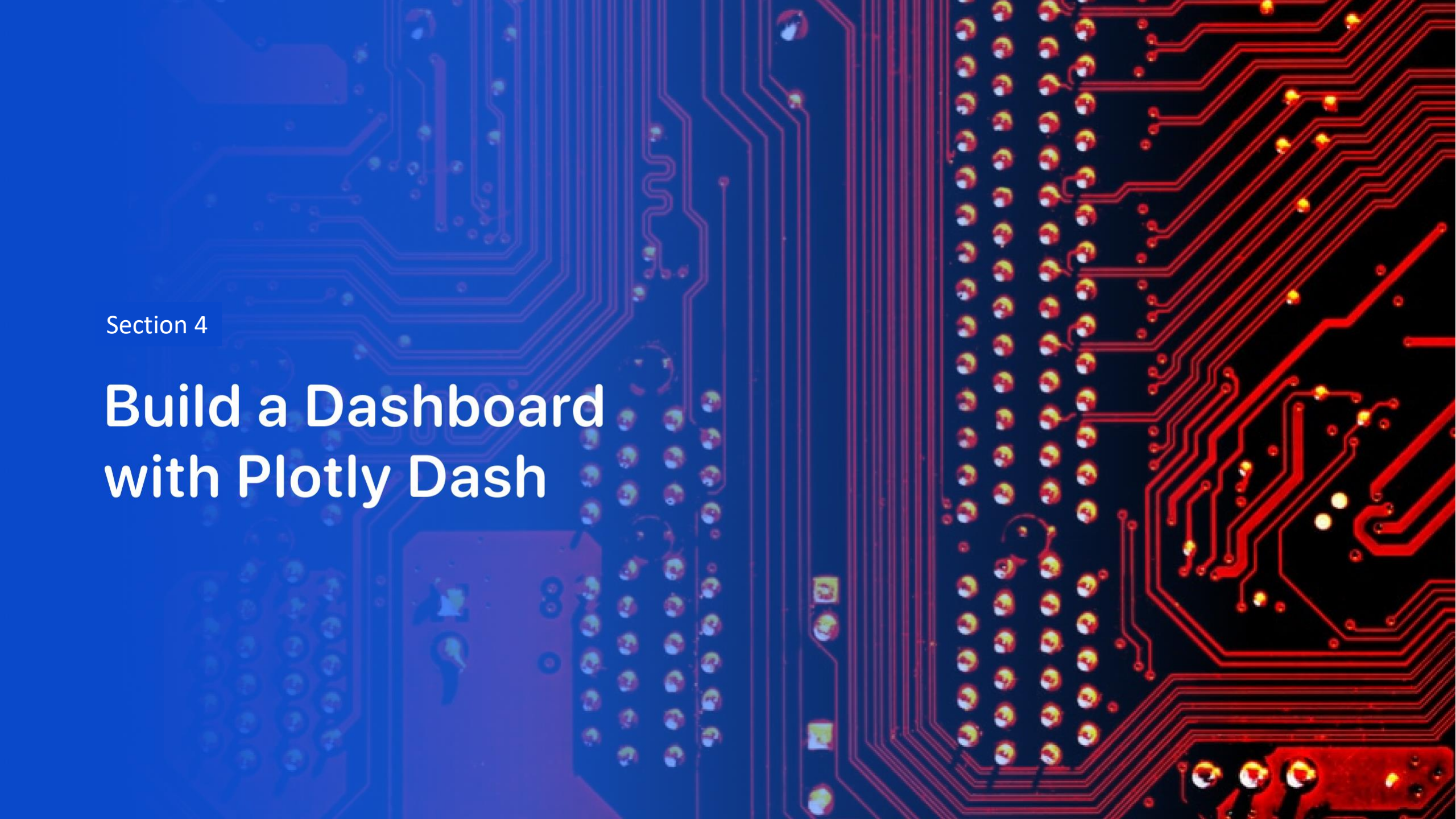
Sites on map


Bryan
College Station
Round Rock
Austin
The Woodlands
Beaumont
Lake Charles
Lafayette
Baton Rouge
Houston
Sugar Land
NASA
JSC
Galveston
New Braunfels
San Antonio
Victoria
Corpus Christi

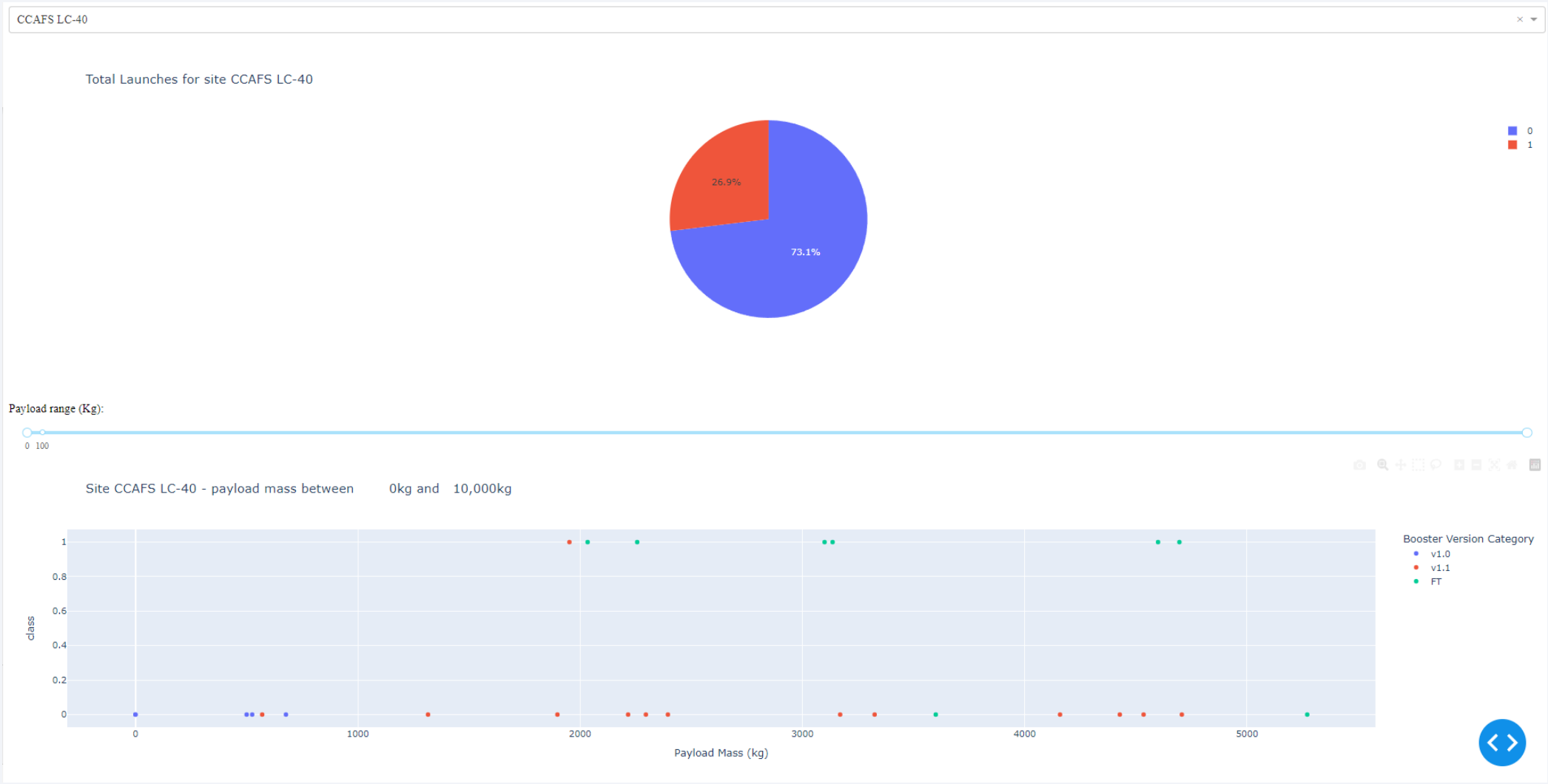
Success and failed launches

Marker with distance to a closest city

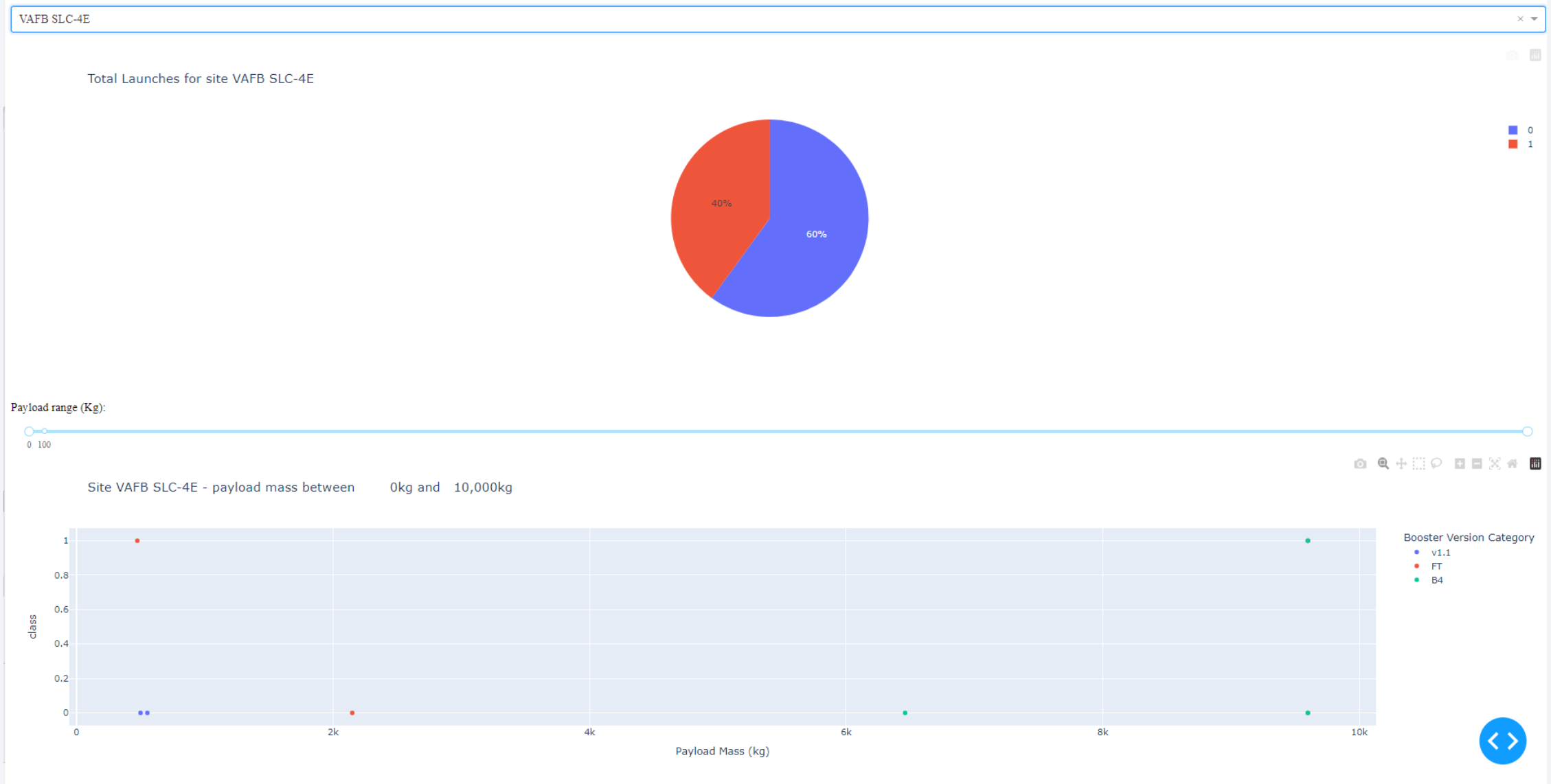



Section 4
Build a Dashboard with Plotly Dash

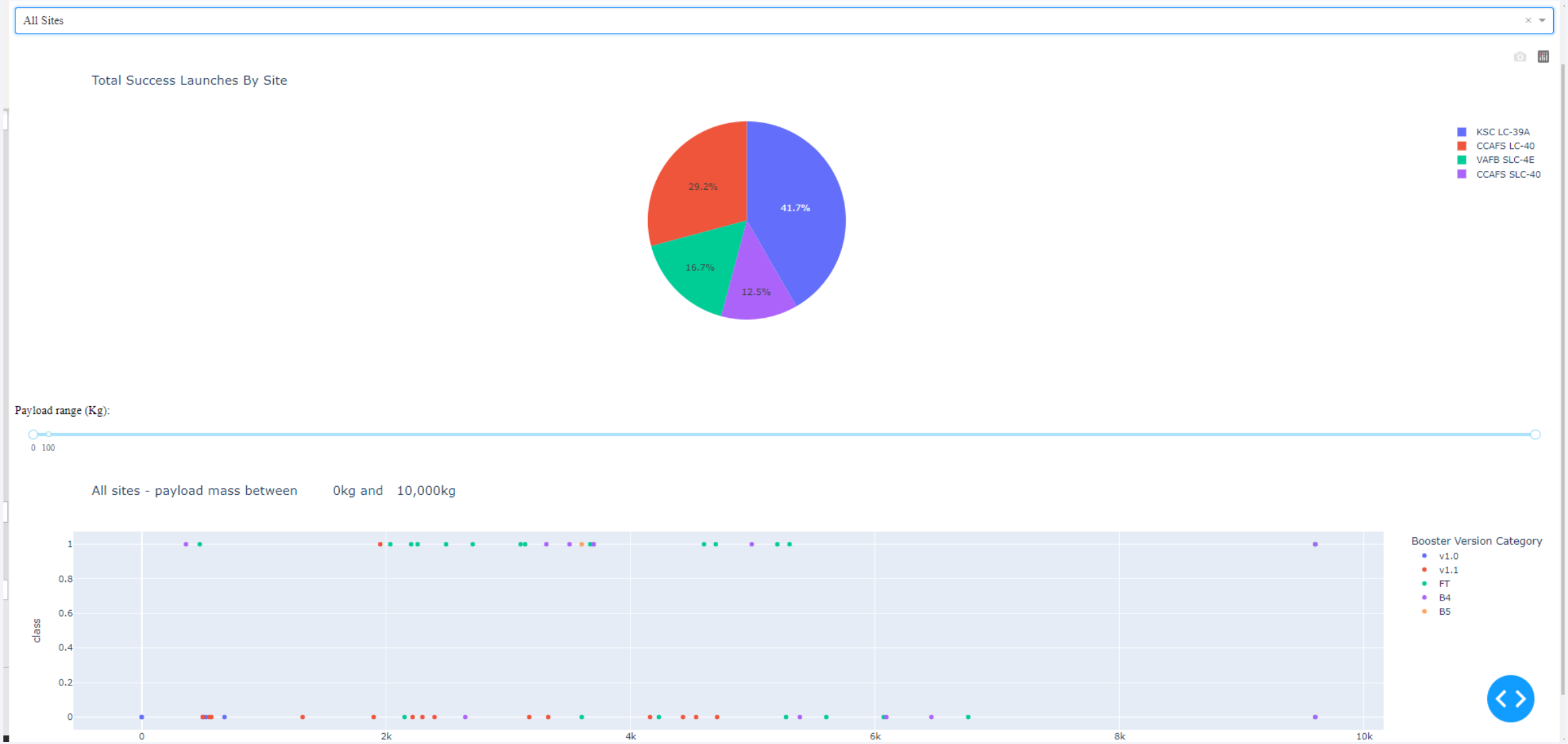
Dashboard for CCADS LC-40



Dashboard for VAFB SLC-4E



Dashboard for all site



Section 5

Predictive Analysis (Classification)

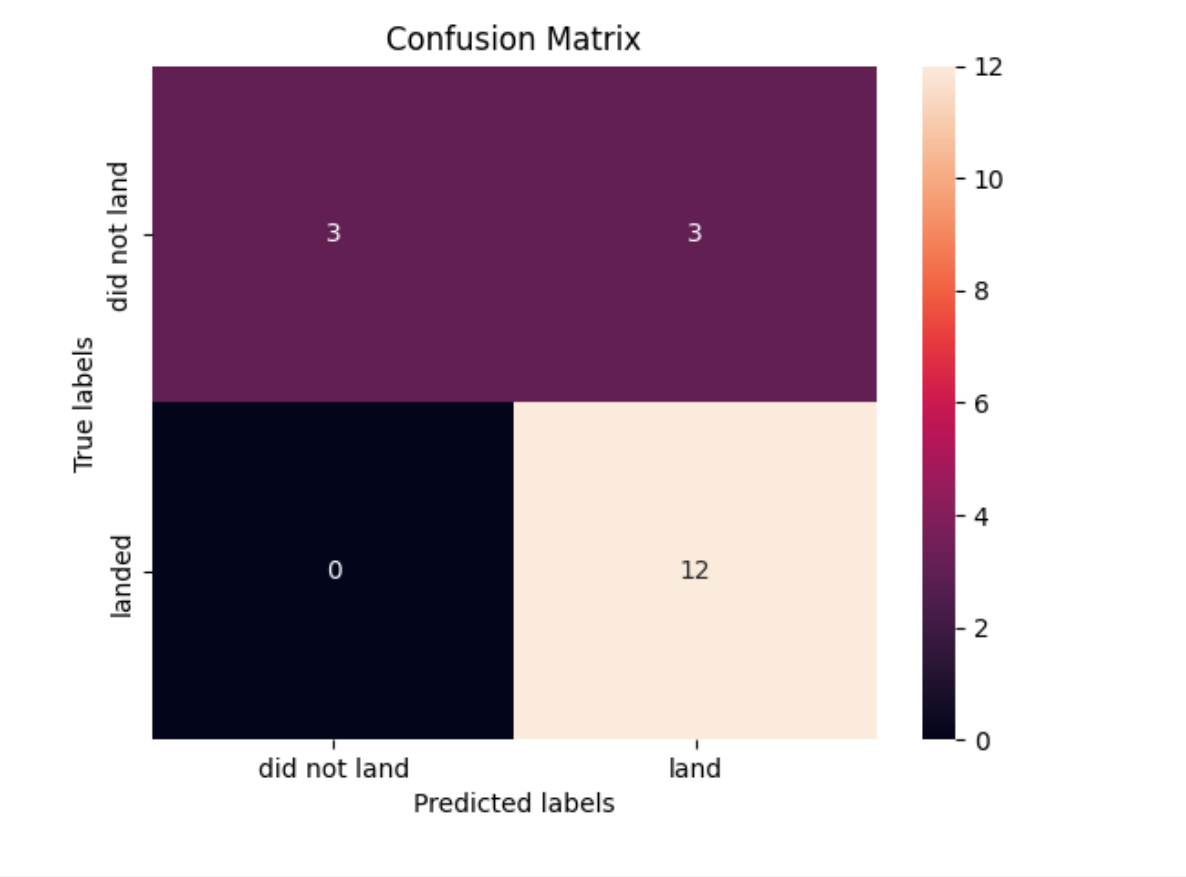
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.87679	0.83333
KNN	0.84821	0.83333

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- SVM,KNN and Logistic Regression models are best in term of prediction accuraccy of dataset
- KSC LC 39A has the most successful launches from all the sites
- Orbit GEO HEO SSO ES has best success rate

Thank you!