IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Oseni Abdul-Afeez
July 2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context

  SpaceX advertises Falcon 9 launches on its website with a cost of 62 million dollars each, other providers cost upward of 165 dollars each, much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers

  To predict if the first stage of Space X Falcon 9 will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection
  - Data were collected from SpaceX Rest API and webscrapping was done from the Wikipedia

Perform data wrangling

  - One hot encoding data fields for machine learning and data cleaning of null values and irrelevant columns

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM, DT model were built and evaluated for the best classifier

Data Collection

- SpaceX launch data is gathered from the SpaceX rest API, the API will give us data about launches, information on the rocket used, payload delivered, launch specifications, landing specifications and landing outcome. We can also obtain Falcon 9 launch data with webscrapping using beautifulsoup

Data Collection – SpaceX API

- Data collection with SpaceX REST calls

https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-data-collection-api.ipynb

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Webscrapping from Wikipedia

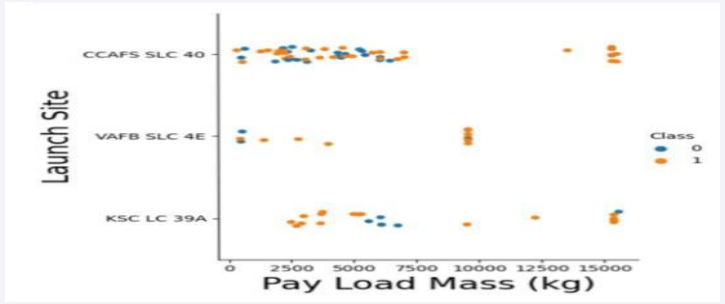
https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-webscraping.ipynb

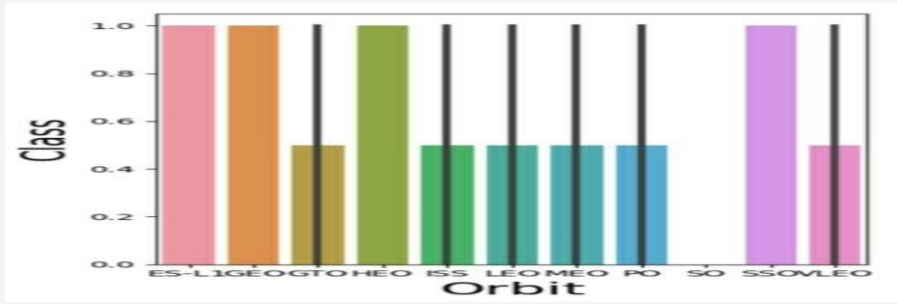
Data Wrangling

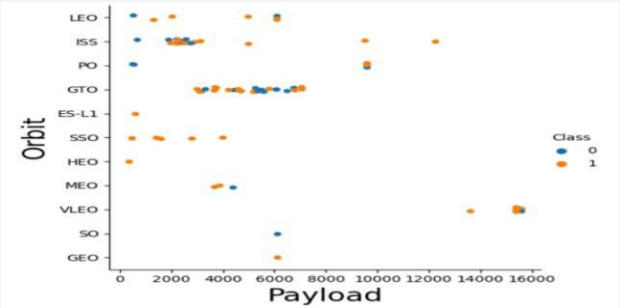
- Null values was calculated then the number of launches on each site was calculated then the number and occurrence of each orbit was calculated then the number and occurrence of mission outcome per orbit type was calculated then a landing outcome label from the outcome column was created then the null values were handled
- You need to present your data wrangling process using key phrases and flowcharts

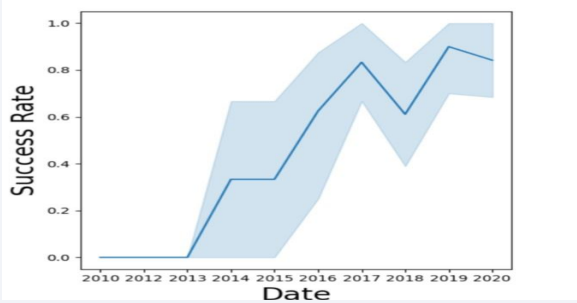
https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

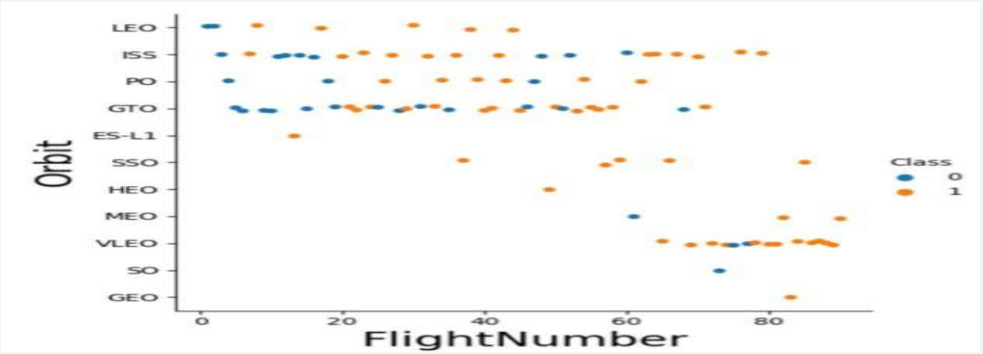
EDA with Data Visualization









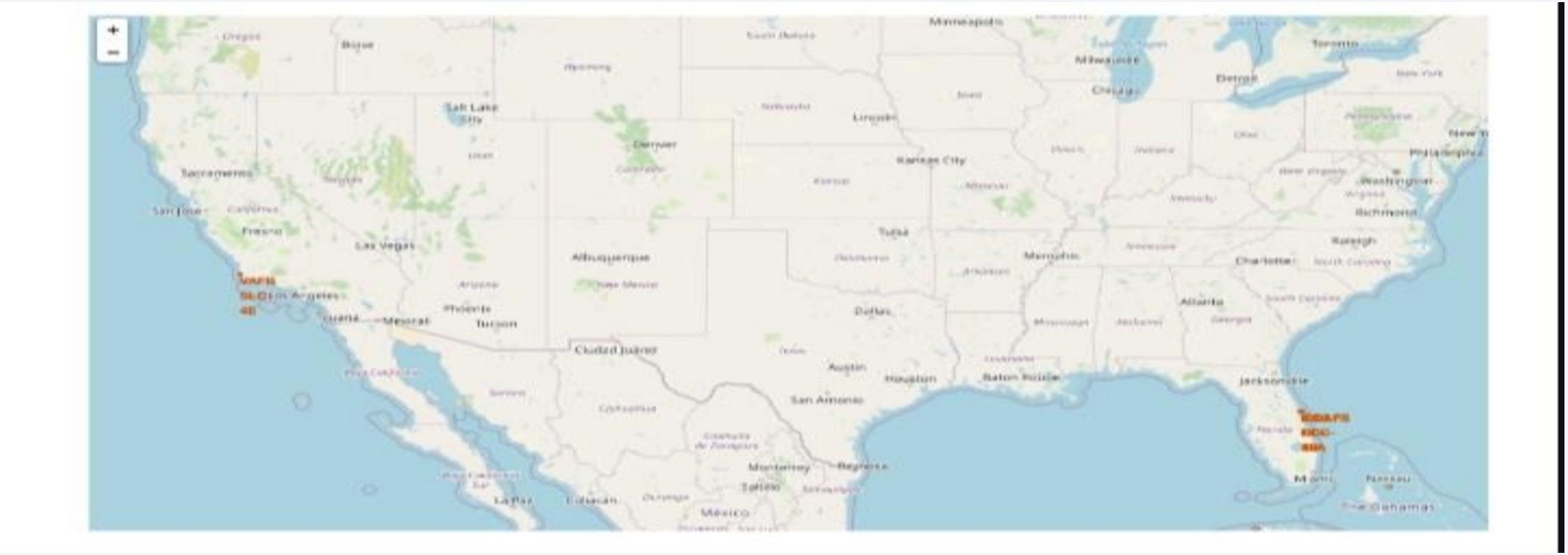


https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

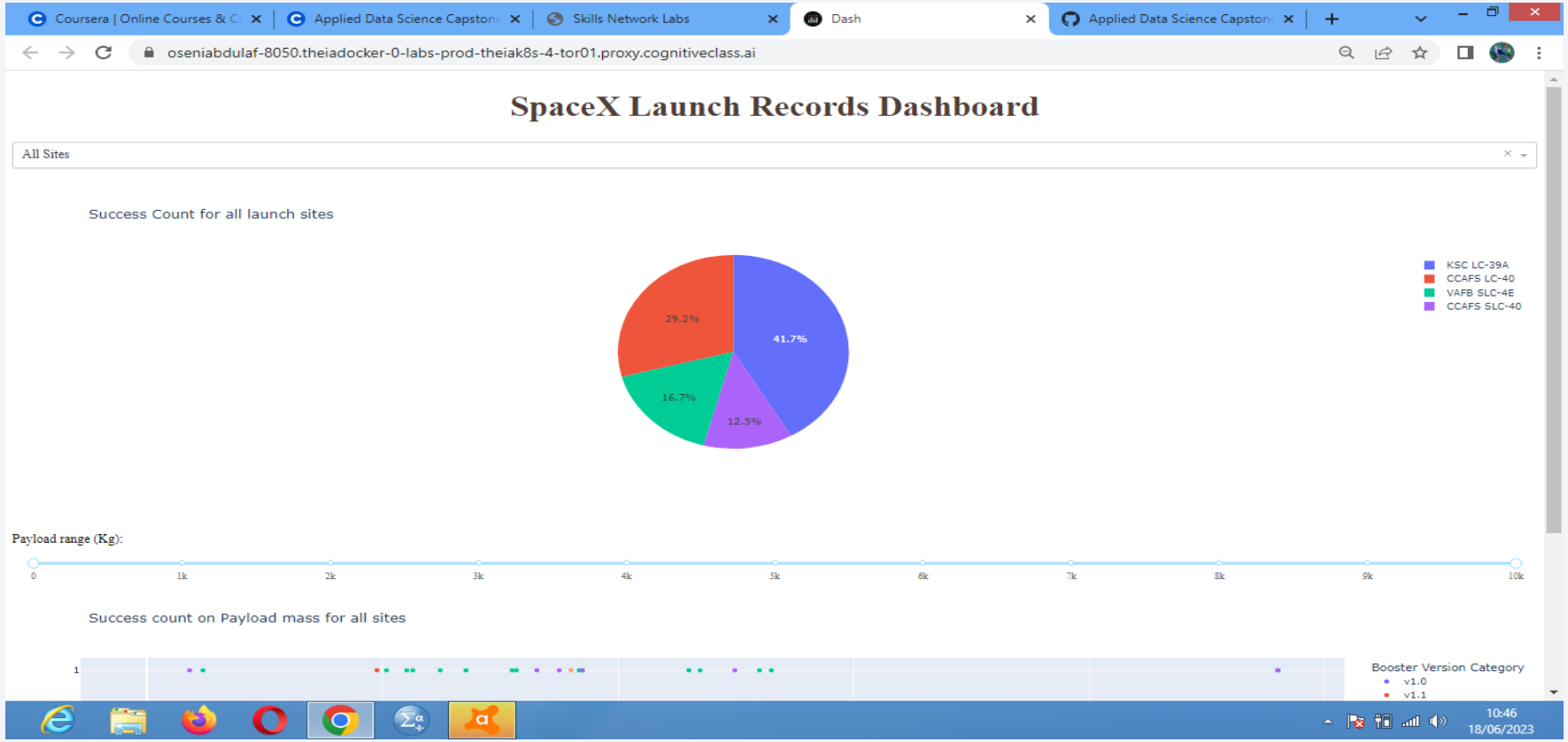
- Displaying the number of the unique launch sites in the space mission
- Displaying 5 records where launch sites where launch sites begins with the string KSC
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by boosters version F9 v1.1
- Listing the date where the successful landing outcome in drone ship was achieved
- Listing the names of the booster which have success in ground pad and have payload mass greater than 4000 but less than 6000
- Listing the number of successful and failed mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the records which will display the month names, successful landing_outcomes in ground pad, booster versions, launch_sites for the months in year 2017
- Ranking the counts of successful landing_outcome between the date 2010 06 04 to 2017 03 20 in descending order
- https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium



https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex-foliumlab-launch_site_location.jupyterlite.ipynb

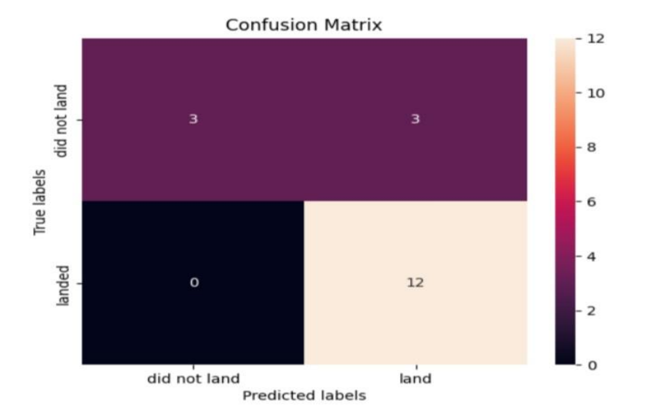
Build a Dashboard with Plotly Dash

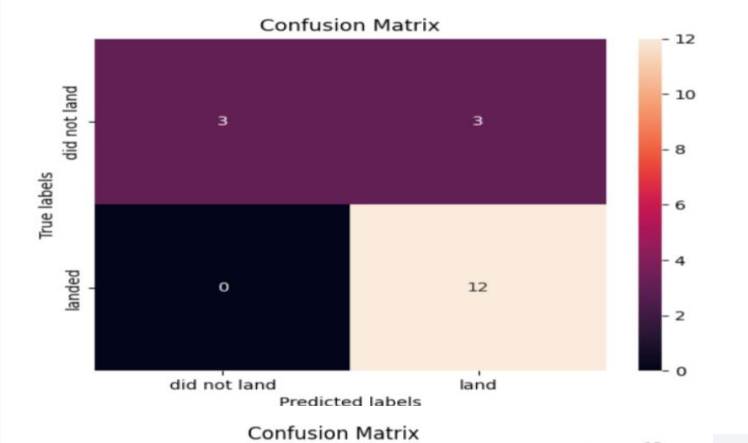


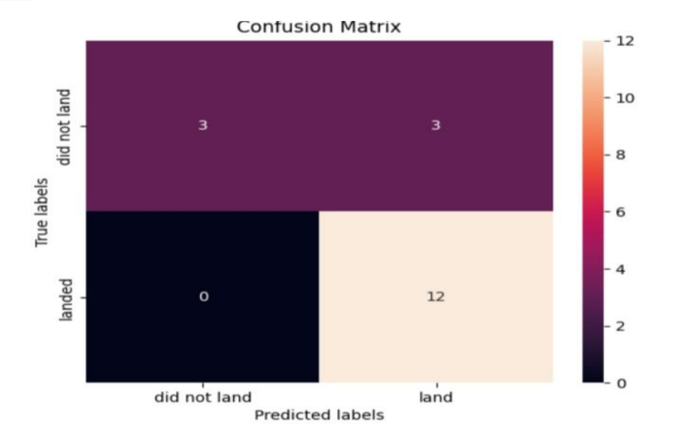
https://github.com/OseniAbdulafeez/DataScienceCapstoneProject/blob/main/spacex.py

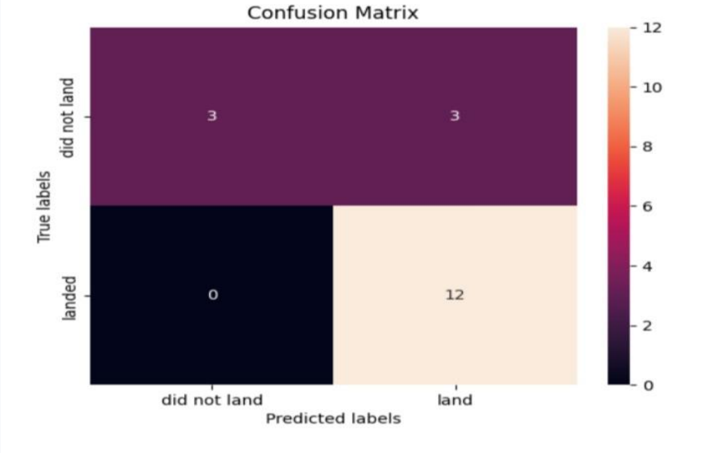
Predictive Analysis (Classification)

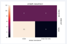
- The SVM, KNN and Logistic regression models are the best in terms of prediction accuracy of the datasets







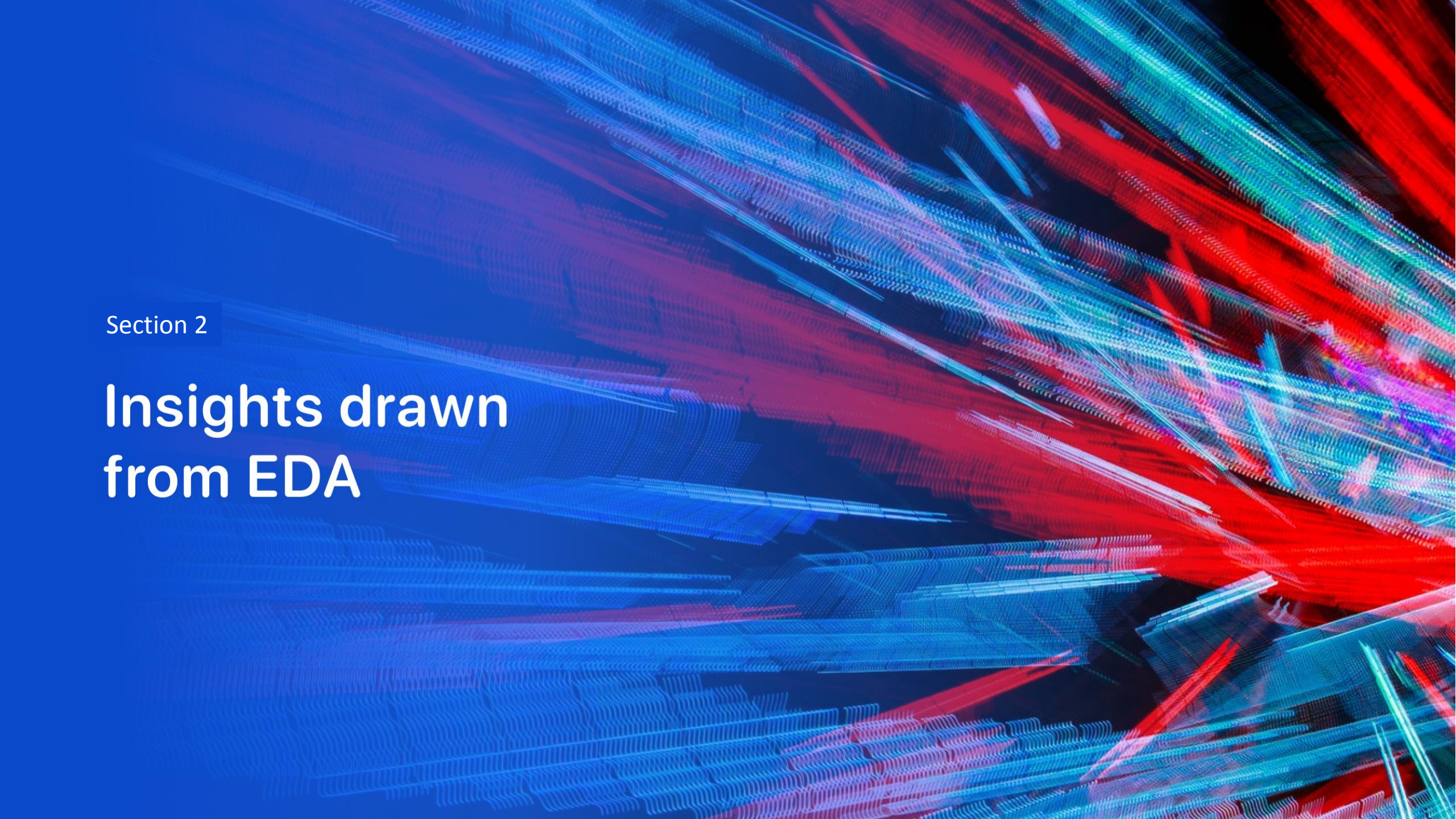




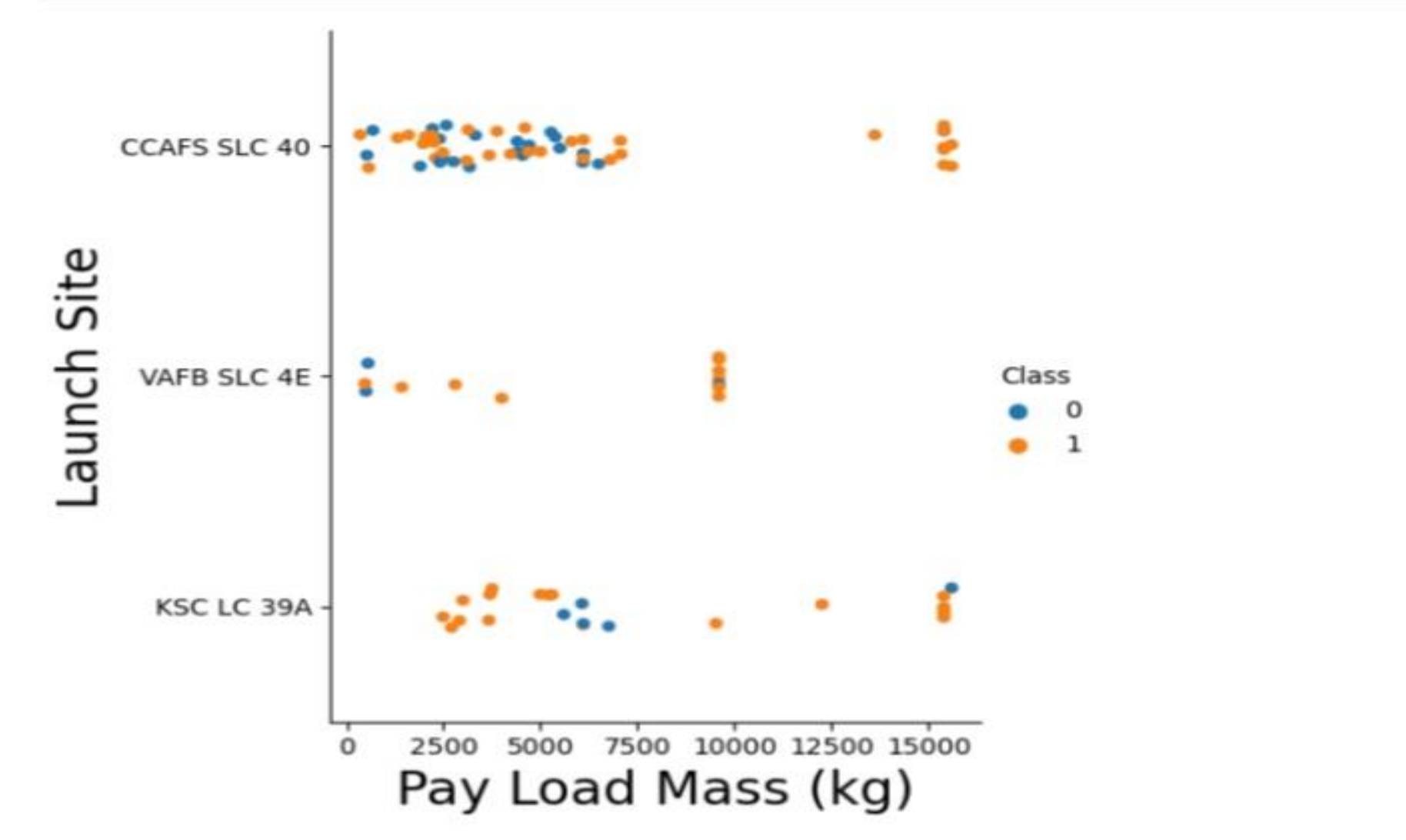
Results

- The SVM, KNN and Logistic regression models are the best in terms of prediction accuracy of the datasets
- Low weighted payloads performs better than the heavier payloads
- KSL LC 39A had the most successful launches from all the sites
- Orbit GEO, SEO, SSO and ES L1 has the best success rate

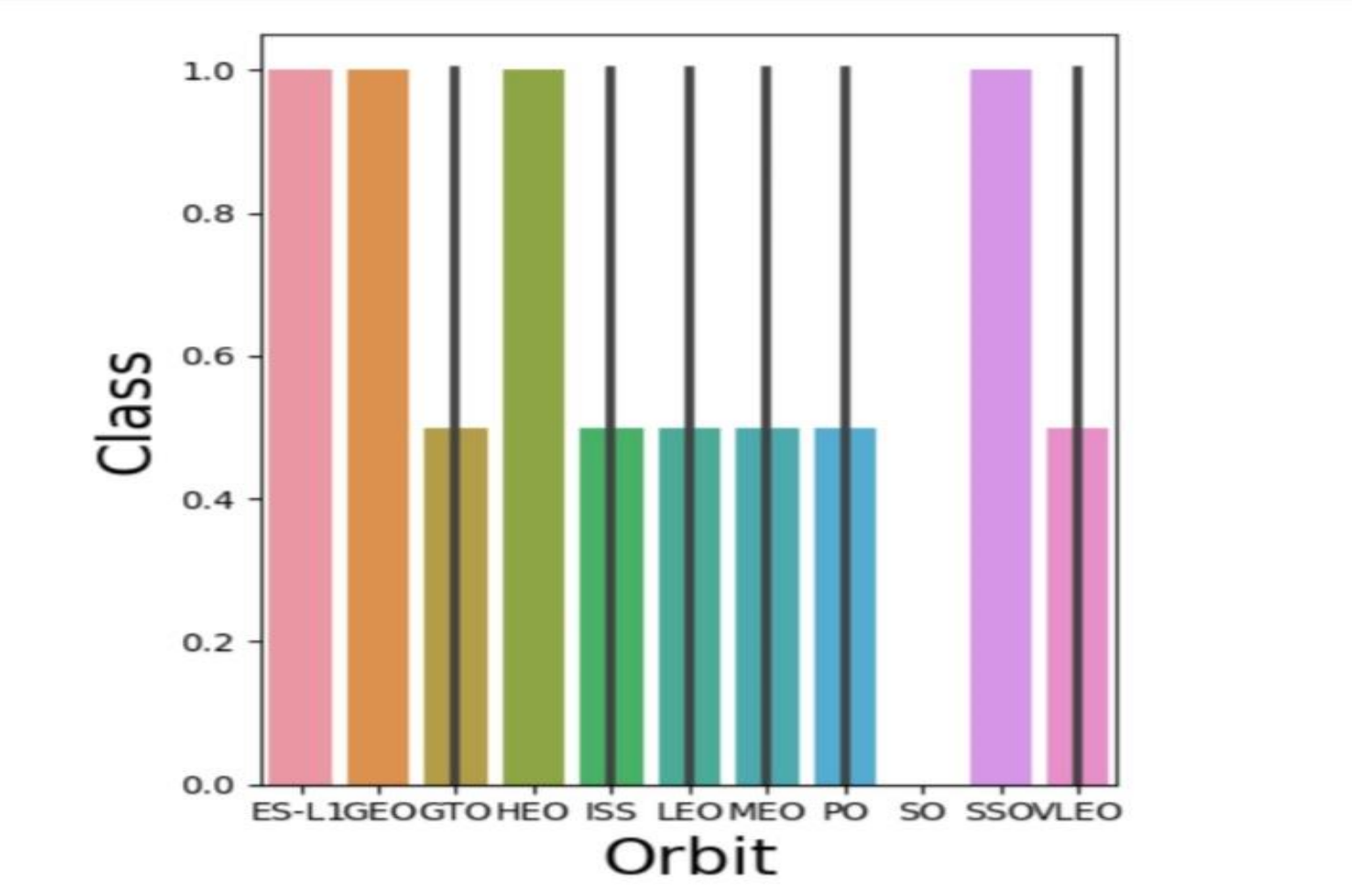
https://github.com/OseniAbdulafeez/DataScienceCapstoneProject.git


Section 2
Insights drawn from EDA

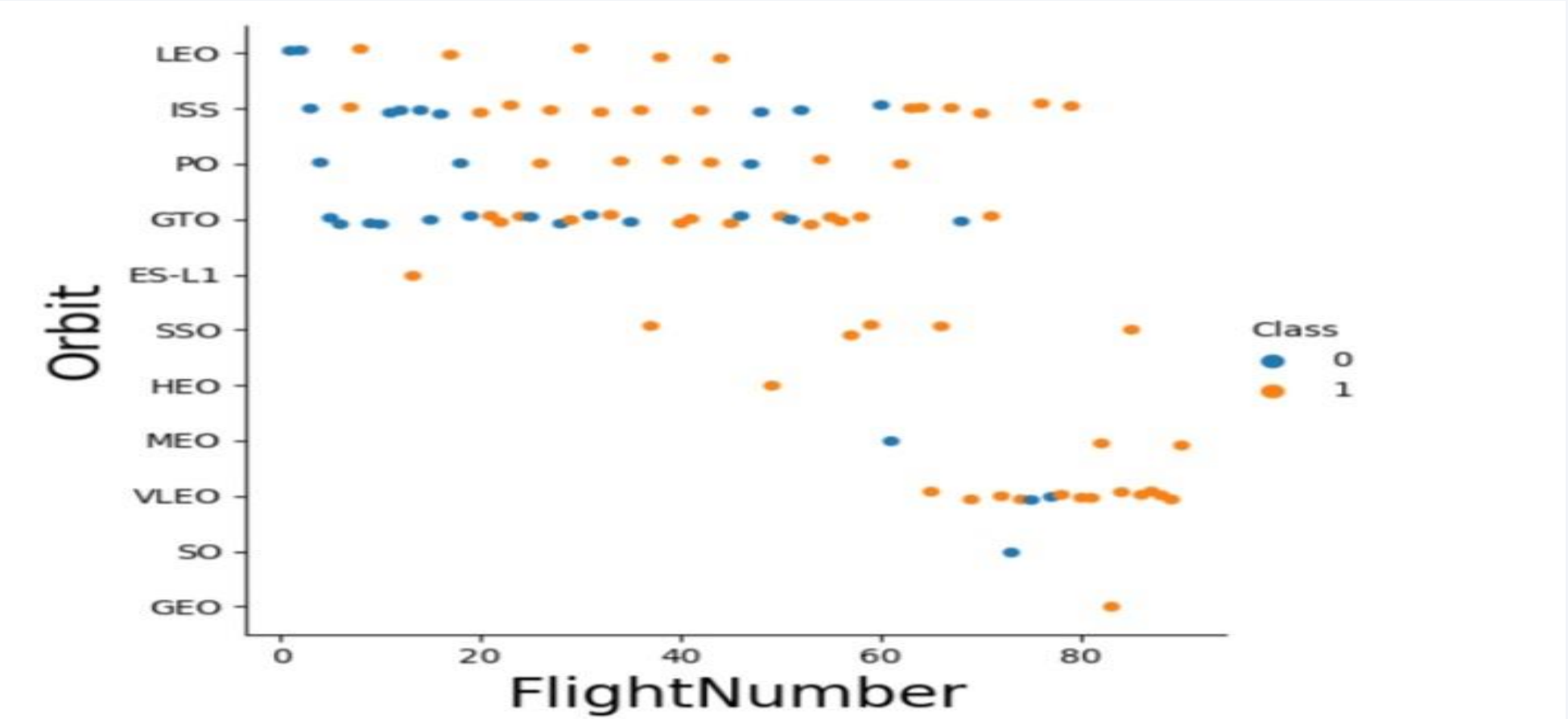
Payload vs. Launch Site



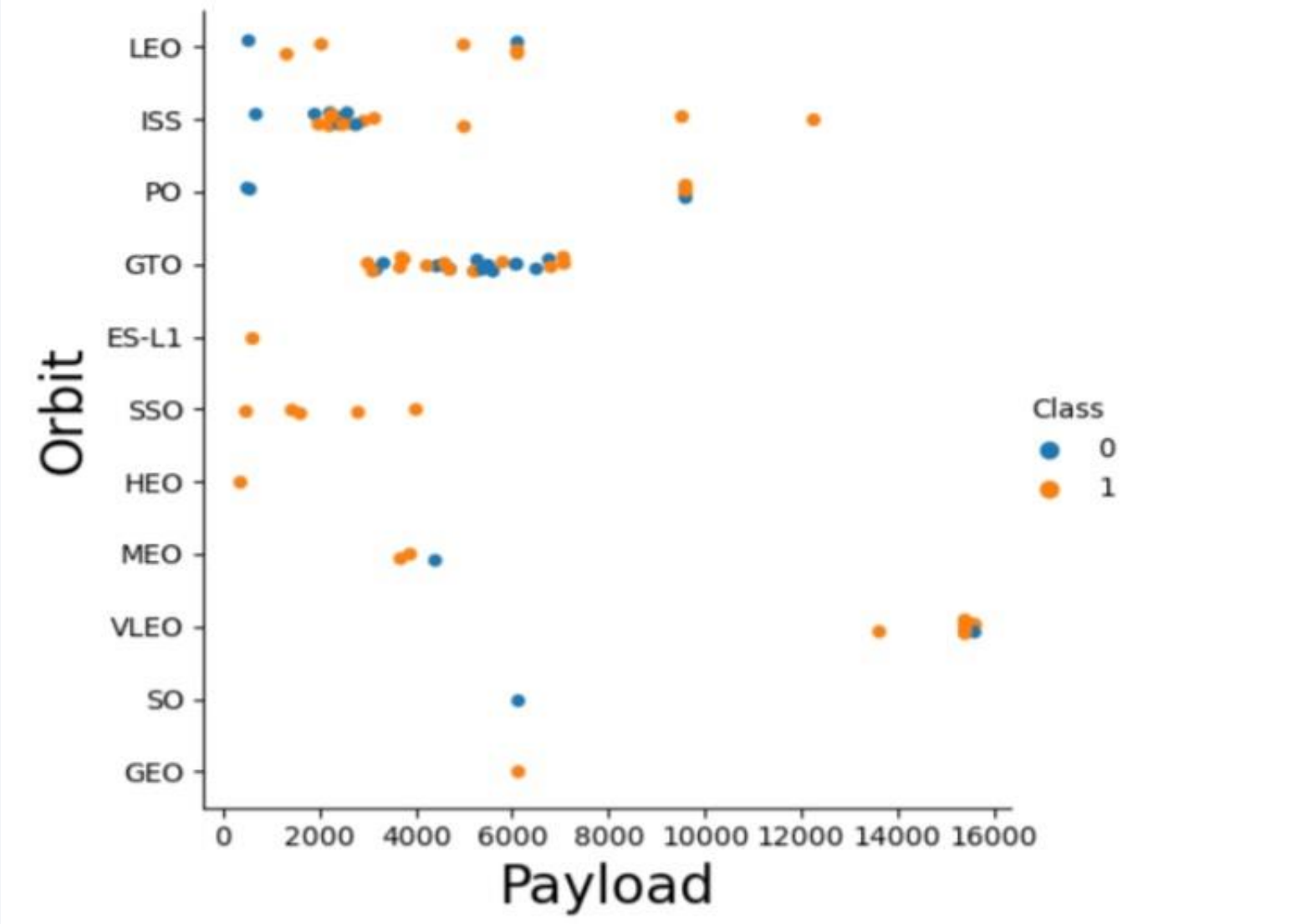
Success Rate vs. Orbit Type



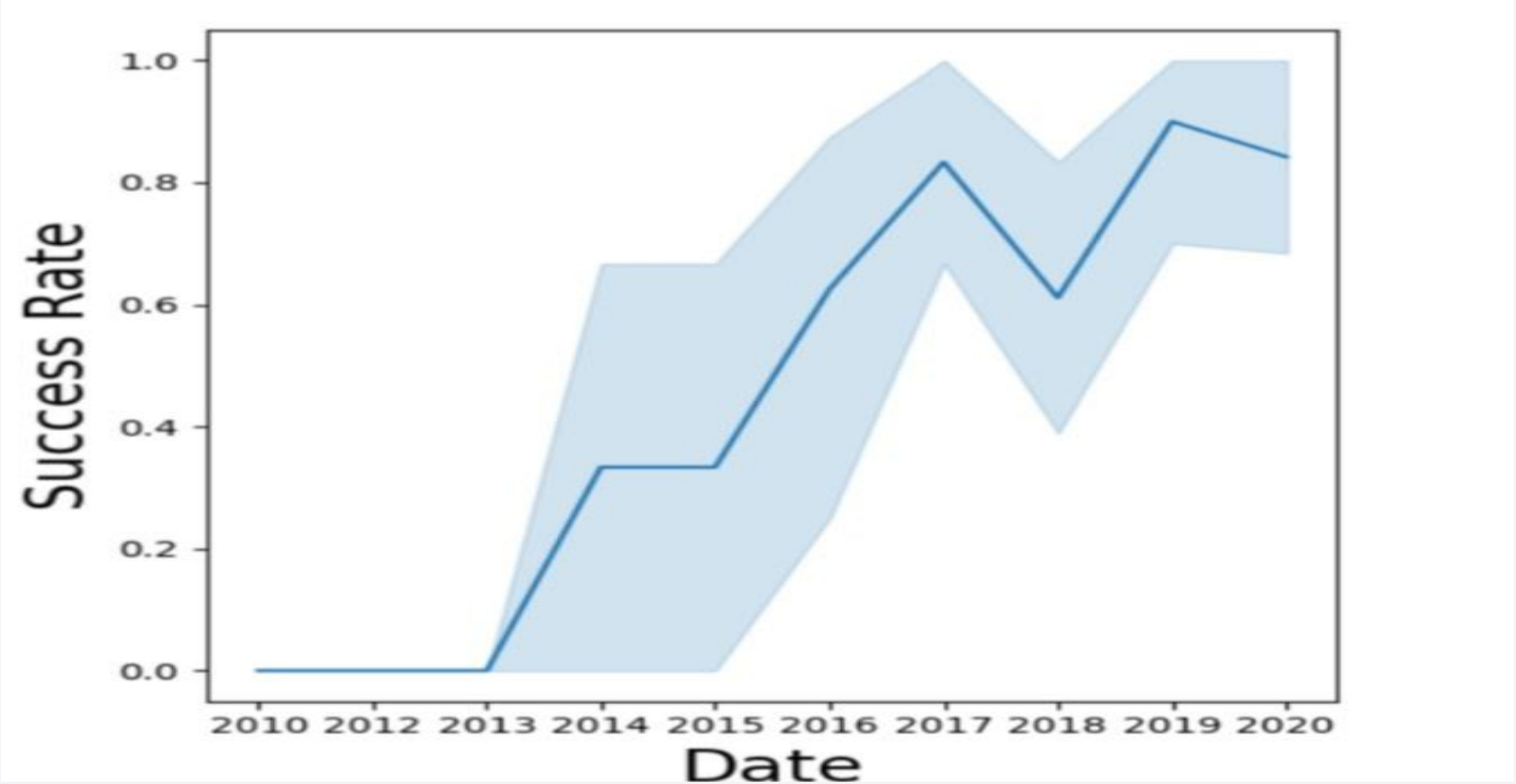
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
In [7]:  %sql SELECT Distinct
          * sqlite:///my_data1.db
        Done.
```

Out[7]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

```
[8]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

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

[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcon
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failu (parachut
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failu (parachut
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attem
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attem
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

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_)
45596.0

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION='F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Hint:Use min function

```
[13]: %sql SELECT min(DATE) FROM SPACEXTBL WHERE LANDING_OUTCOME='Success (ground pad)
```

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

[13]:

min(DATE)
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[14]: %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ between 4000 and 6000 AND LANDING_OUTCOME='Su
 * sqlite:///my_data1.db
Done.
```

[14]:

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

Task 7

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Success%' OR MISSION_OUTCOME LIKE '%Failu
```

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

COUNT(*)
101

Boosters Carried Maximum Payload

```
[6]: %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTB
 * sqlite:///my_data1.db
Done.
```

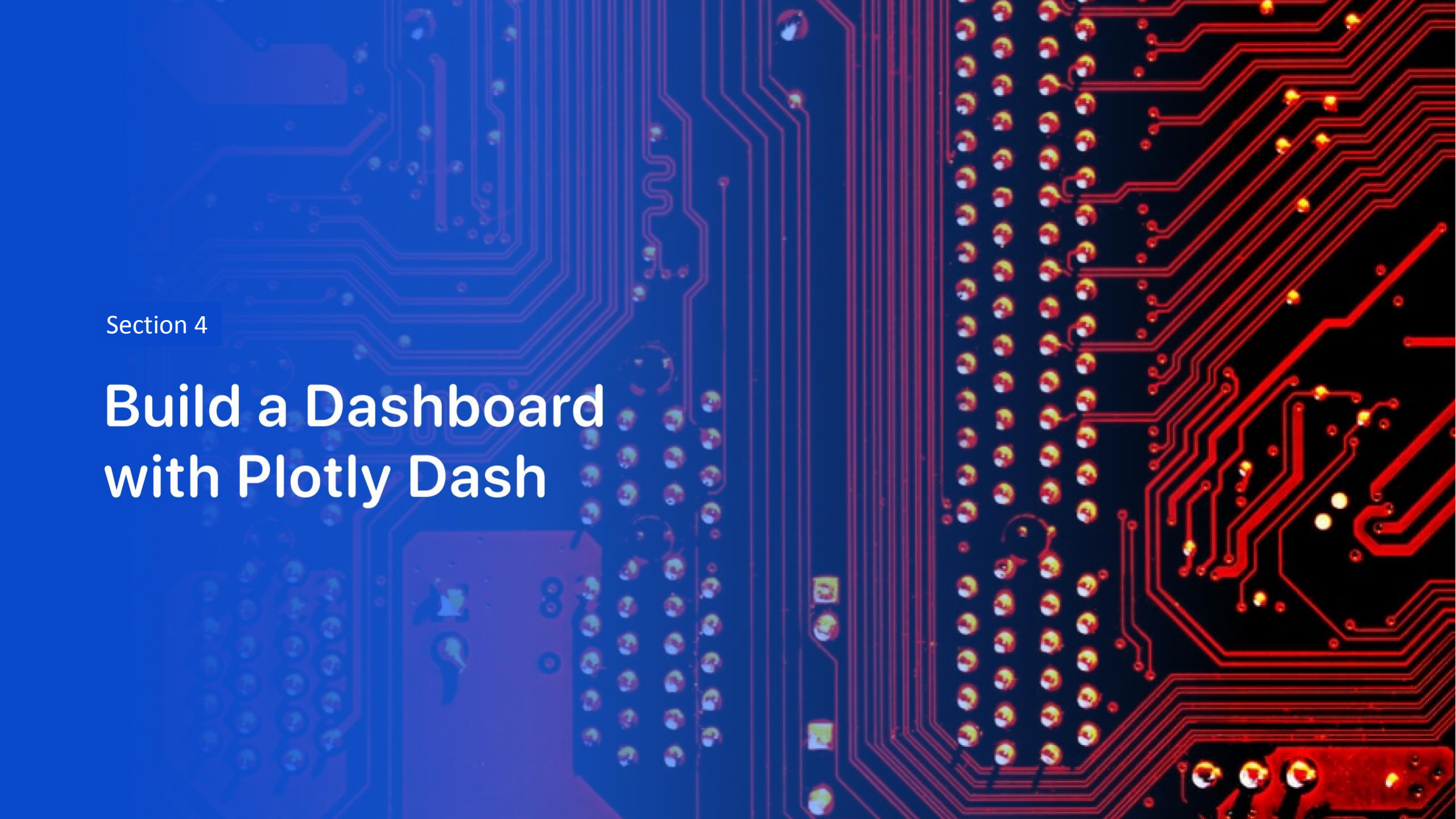
[6]:

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

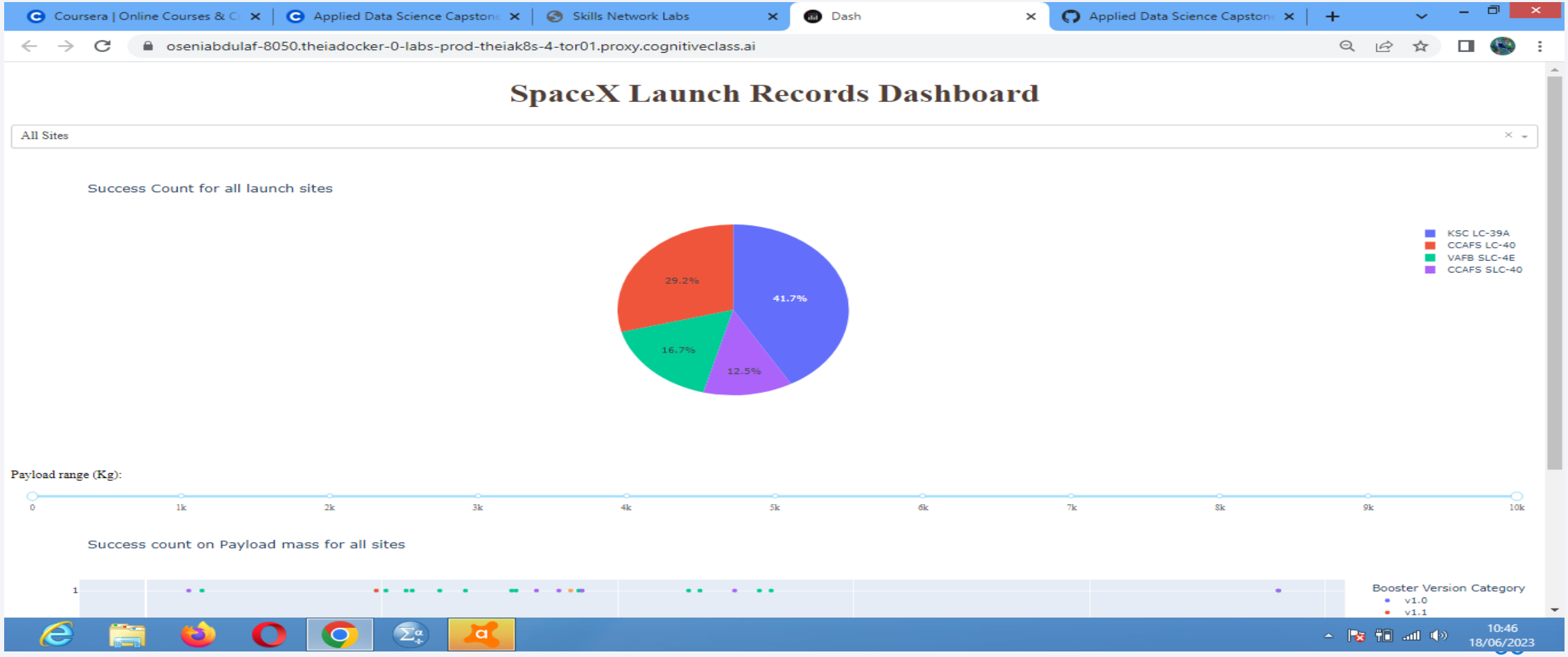

Section 3
Launch Sites
Proximities Analysis

Success/failed launches marked on the map

Distance between a launch site to its proximities


Section 4
Build a Dashboard with Plotly Dash

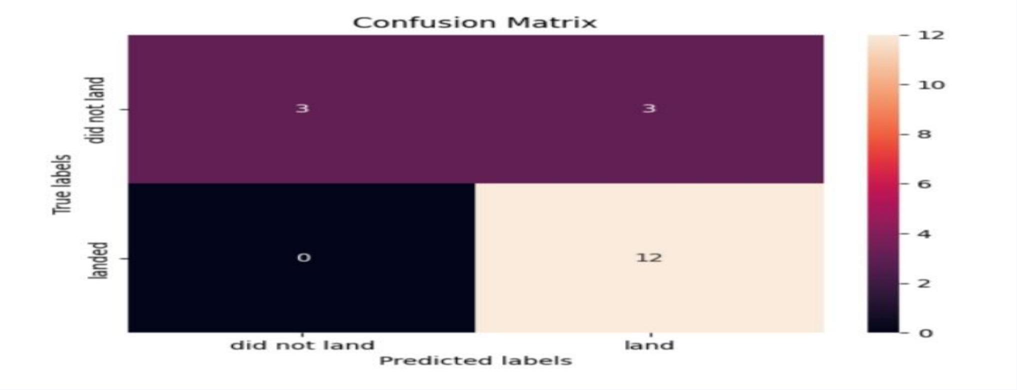
Total success launches by all sites

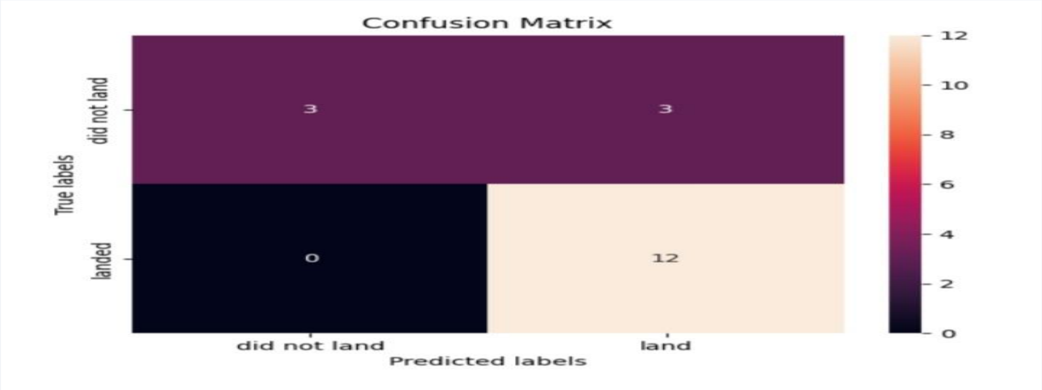


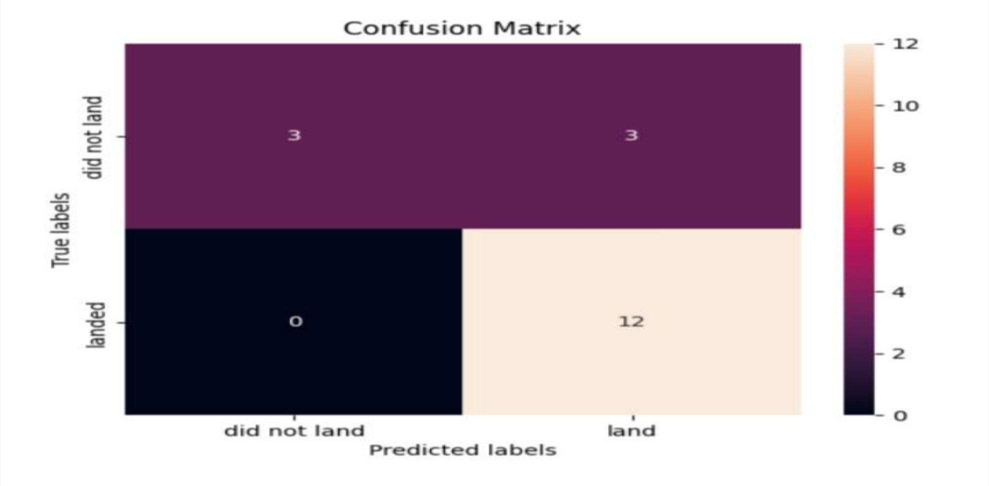
Section 5

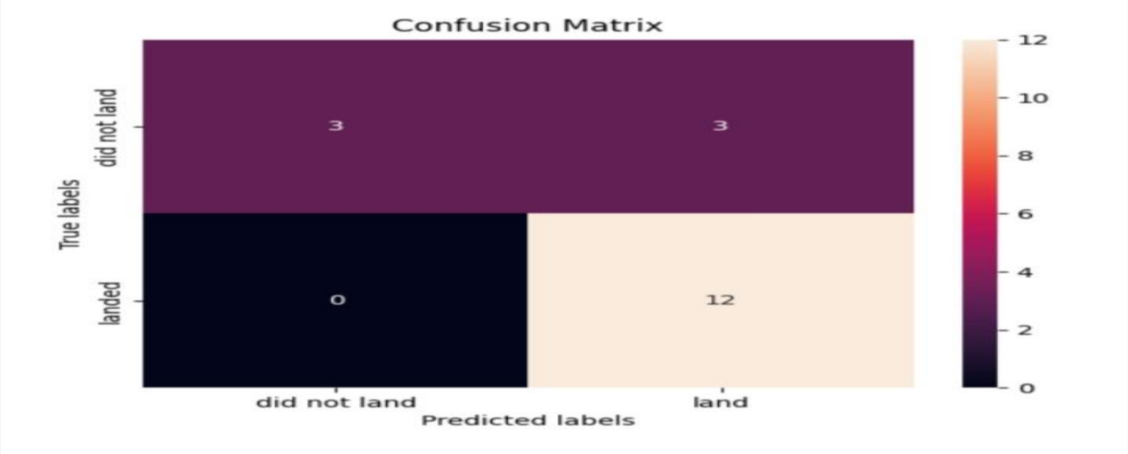
Predictive Analysis (Classification)

Confusion Matrix









Conclusions

- The SVM, KNN and Logistic regression models are the best in terms of prediction accuracy of the datasets
- Low weighted payloads performs better than the heavier payloads
- KSL LC 39A had the most successful launches from all the sites
- Orbit GEO, SEO, SSO and ES L1 has the best success rate

https://github.com/OseniAbdulafeez/DataScienceCapstoneProject.git

Thank you!