IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

NIVEDHIKA N
27 September 2022

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

EXECUTIVE SUMMARY

Summary of methodologies

- Data Collection via API, web scraping and SQL
- Data Wrangling and Analysis
- Interactive Maps with Folium
- Predictive Analysis for each Classification model

Summary of all results

- Data analysis with interactive visualisations
- Best model for predictive analysis

INTRODUCTION

Project background and context

- SpaceX can reuse the first stage of Falcon 9 rocket, and therefore, bids for a cost of 62 million dollars.
- It is more than 50% of reduced cost from other bidders (165 million dollars).

Problems you want to find answers

- *Factors that contribute in landing a rocket successfully at the first stage.*

Section 1

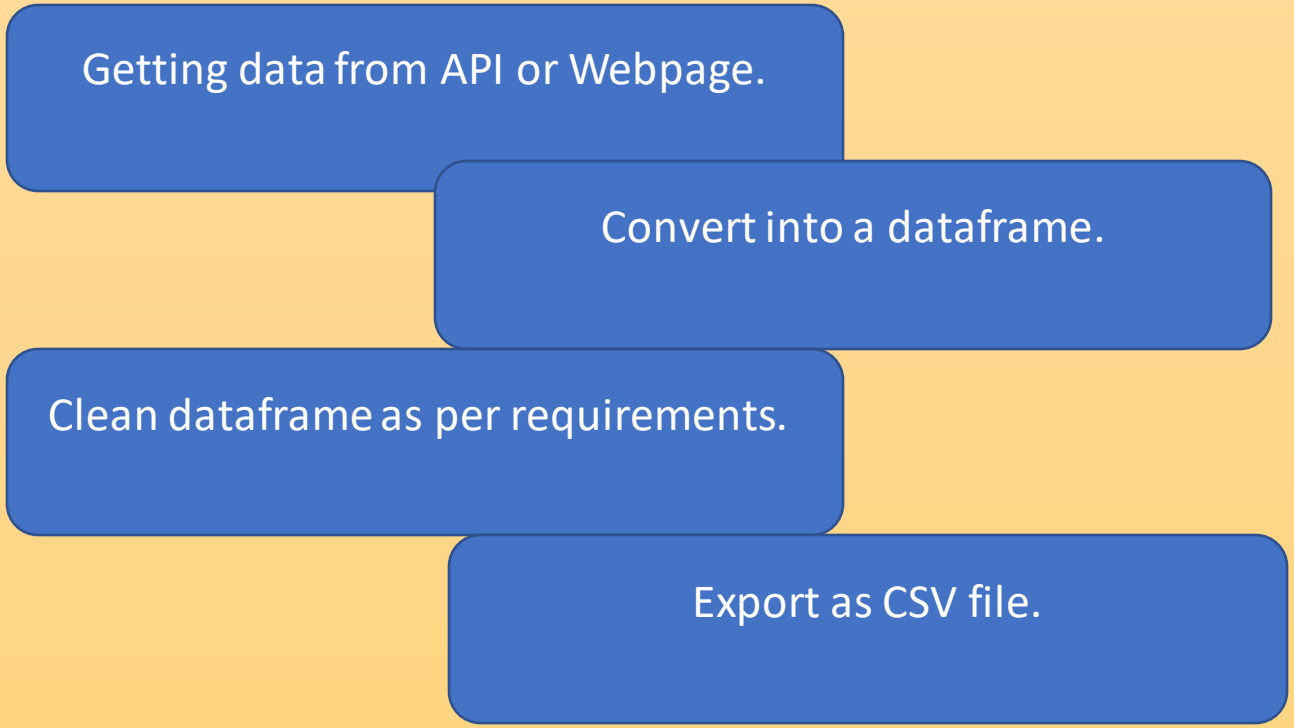
Methodology

Methodology

Data collection	Via SpaceX REST API Web scrapping from Wikipedia
Data wrangling	Performed one hot encoding to deploy machine learning algorithms.
Perform exploratory data analysis (EDA) using visualization and SQL	Plotted Scatter and bar graphs.
Perform interactive visual analytics using Folium and Plotly Dash	Using Folium and plotly dash applications
Perform predictive analysis classification	Using Machine Learning models.

DATA COLLECTION

https://github.com/Nivedhika/Final-project-for-Data-Science



DATA COLLECTION – SpaceX API

```
Response=requests.get(spacex_url)
```

Getting response from API

```
data_falcon9 = pd.DataFrame(launch_dict)
```

Create a Pandas dataframe from the dictionary

Apply custom functions

```
getLaunchSite(data)
getPayloadData(data)
```

- https://github.com/Nivedhika/Final-project-for-Data-Science

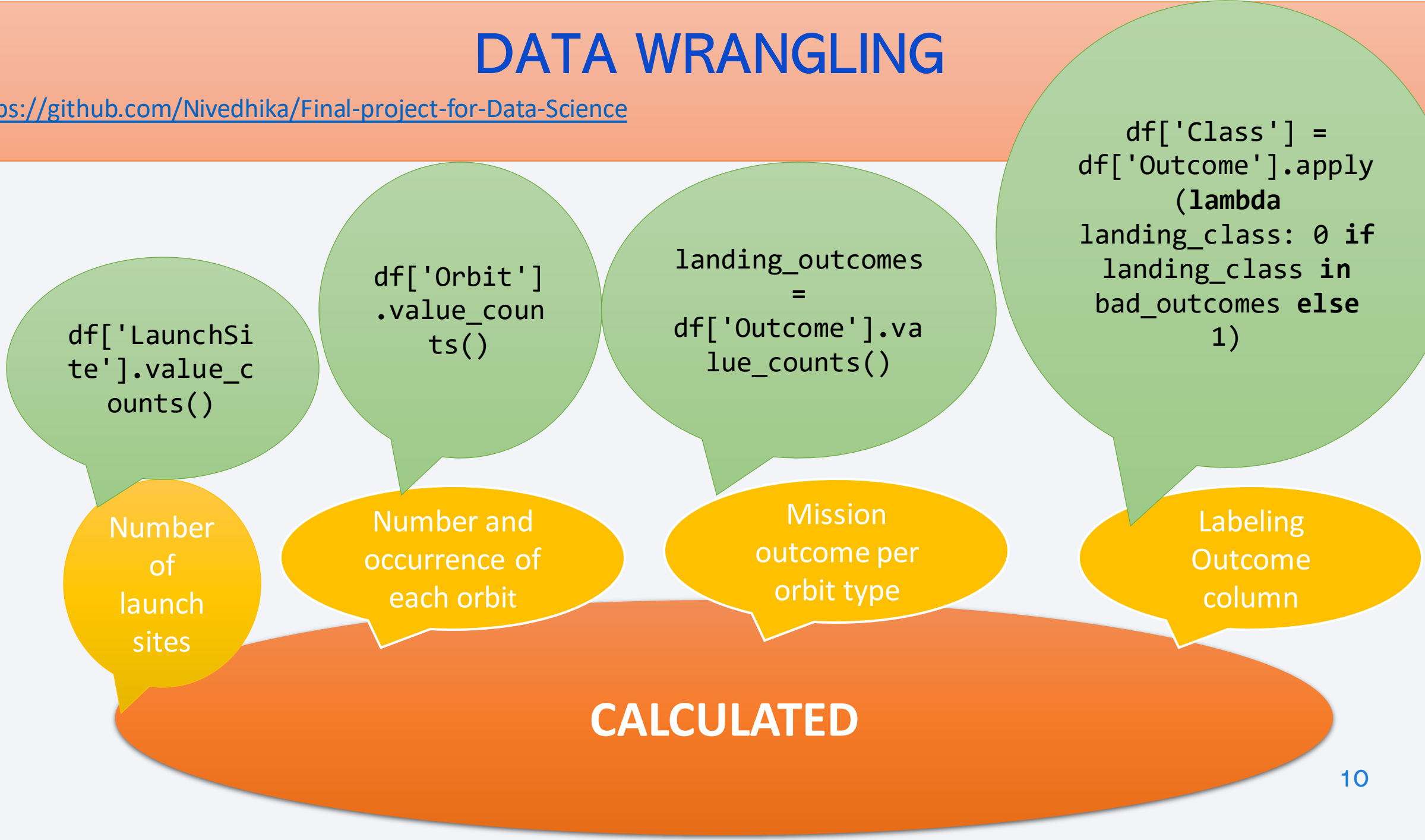
DATA COLLECTION – SCRAPING

https://github.com/Nivedhika/Final-project-for-Data-Science

Creating BeautifulSoup object	`soup = BeautifulSoup(data, 'html5lib')`
Finding Tables	`html_tables=soup.find_all("table")`
Creating dictionary and appending data to keys	`launch_dict= dict.fromkeys(column_names)`
Converting dictionary to dataframe	`df=pd.DataFrame(launch_dict)`
Export dataframe to .CSV	`df.to_csv('spacex_web_scraped.csv', index=False)`

DATA WRANGLING

https://github.com/Nivedhika/Final-project-for-Data-Science



EDA WITH DATA VISUALISATION

https://github.com/Nivedhika/Final-project-for-Data-Science

Scatter plots :

shows relationship between variables

Plotted between these variables:

Payload and Flight Number
Payload and orbit
Flight Number and Launch site
Payload and orbit type
Flight Number and orbit

Bar Graphs:

Shows relationship between variables.

Plotted between Success rate and orbit type.

Line Graphs:

Shows trends clearly.

Plotted to determine the yearly trends of Spacex success rate.

EDA WITH SQL

- Culled out names of the unique launch sites in the space mission.
- Calculated total payload mass carried by boosters launched by NASA (CRS).

- Calculated average payload mass carried by booster version F9 v1.1
- Calculated count of successful landing outcomes.

- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- Listed the total number of successful and failure mission outcomes.
- Listed launch sites begin with the string 'CCA'

https://github.com/Nivedhika/Final-project-for-Data-Science

Build an Interactive Map with Folium

Map Marker (folium.marker)

Makes a mark on map.

PolyLine (folium.PolyLine)

Create a line between points.

Icon Marker(folium.Icon)

Create an icon on map.

MarkerCluster()

Clusters markers of same coordinates.

Circle Marker (folium.Circle)

Create a circle where marker is being placed.

AntPath (Folium.plugins.Antpath)

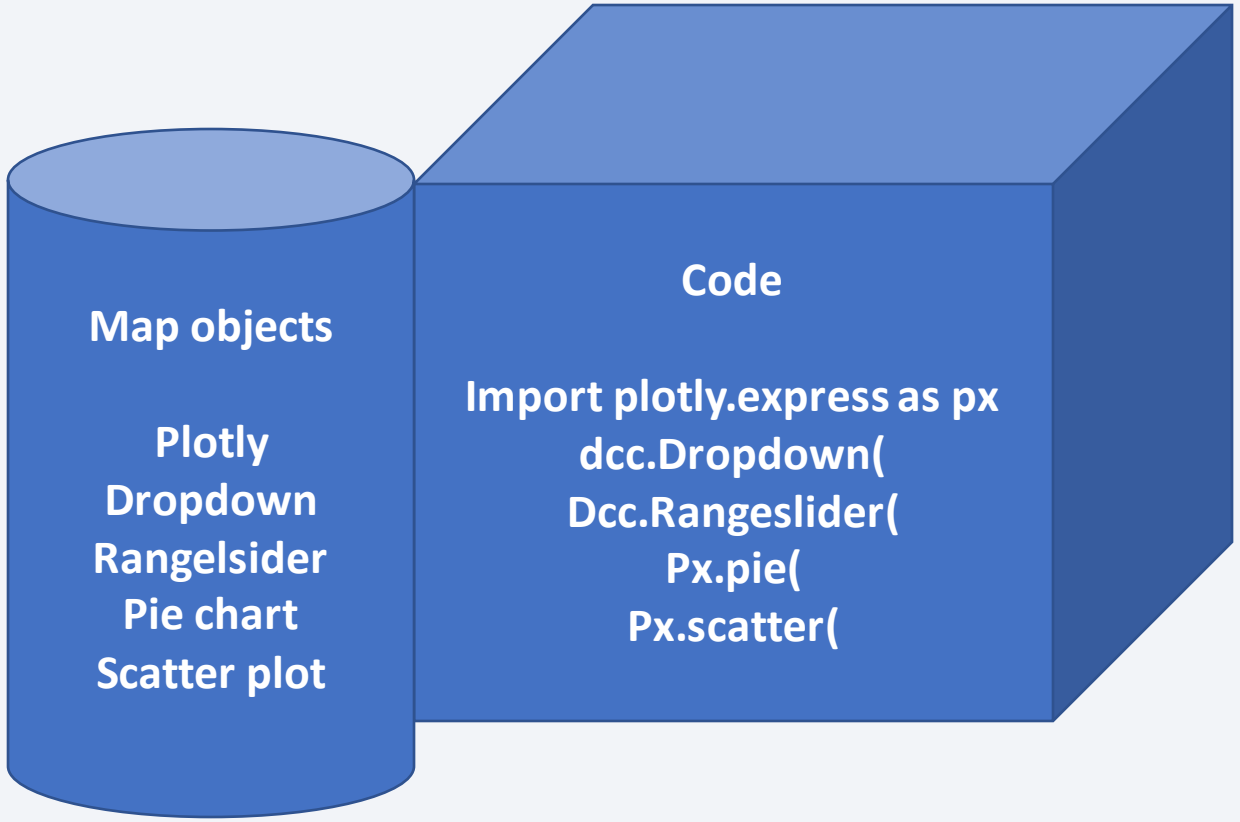
Creates animated line between points.

https://github.com/Nivedhika/Final-project-for-Data-Science

Build a Dashboard with Plotly Dash

Scatter plot
Correlation between payload and success of all sites or by certain sites.

Pie Chart
Showing the total success of all sites or by certain sites.

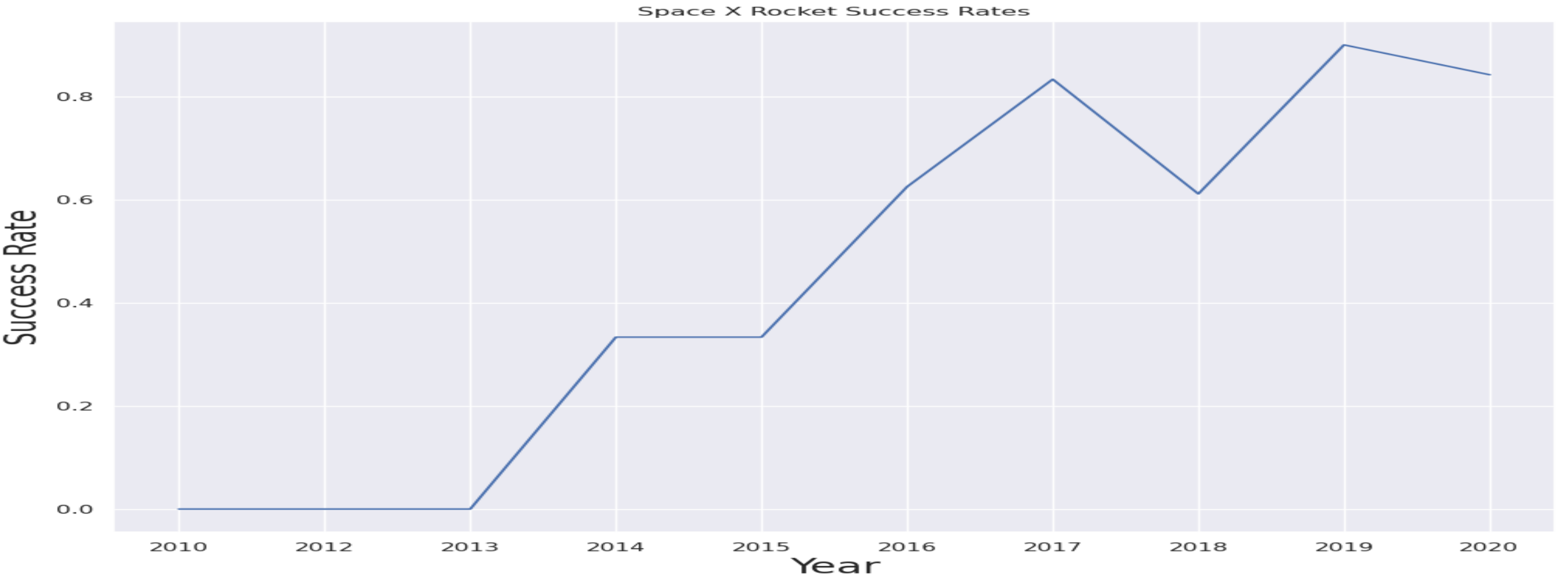


https://github.com/Nivedhika/Final-project-for-Data-Science

Launch Success Yearly Trend

https://github.com/Nivedhika/Final-project-for-Data-Science

The upward trend is visible during the years: 2013-14, 2015-17 and 2018-2019.



Predictive Analysis (Classification)

Load featured engineered data into dataframe.

Transform into NumPy arrays.

Standardize and transform data.

Split data into training and test datasets.

Check how many test samples has been created.

List down machine learning algorithms that to be used in the project.

Set parameters and algorithms to GridSearchCV.

Fit datasets into GridSearchCV model and train the dataset.

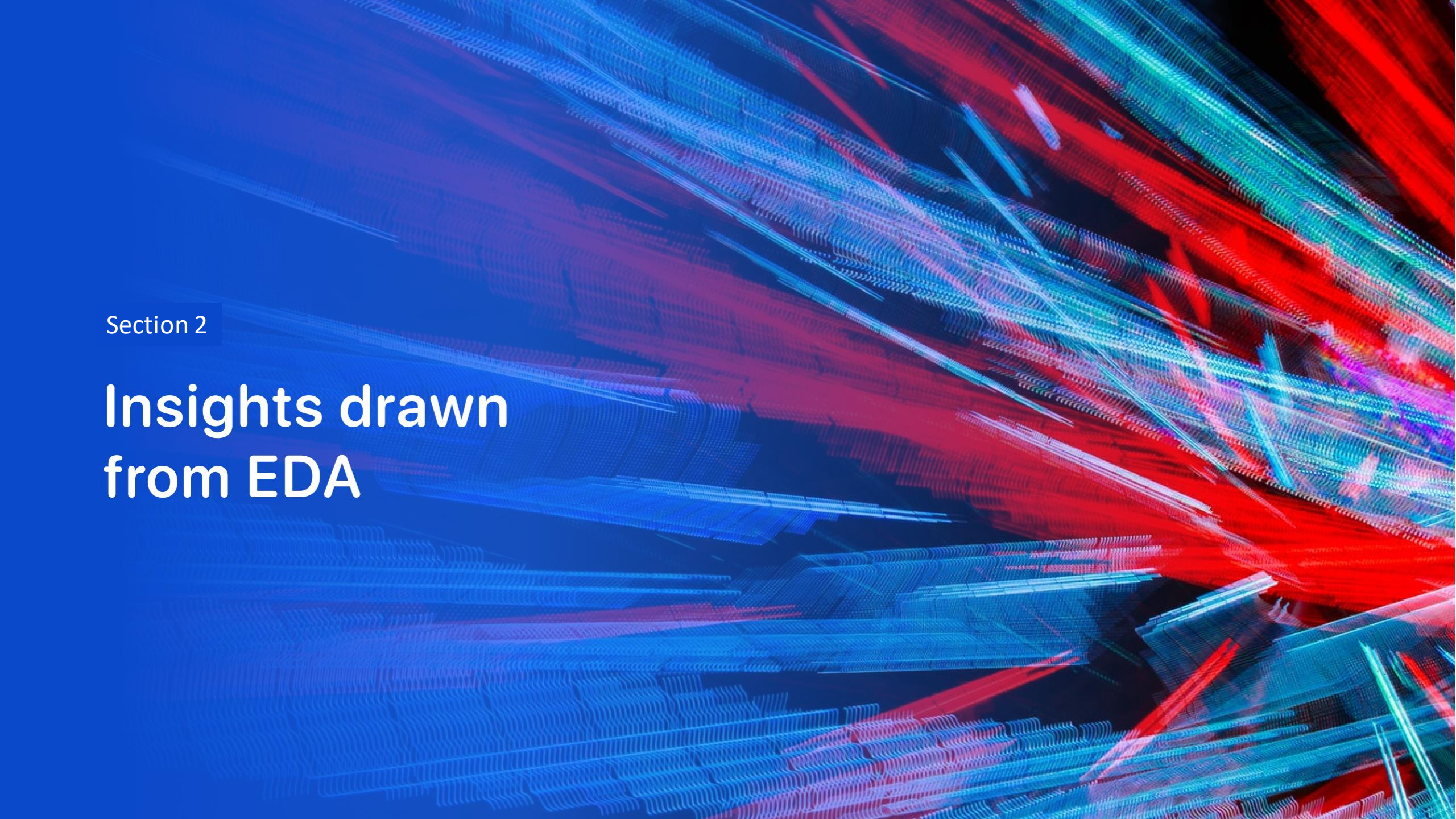
https://github.com/Nivedhika/Final-project-for-Data-Science

RESULTS

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

Section 2

Insights drawn from EDA

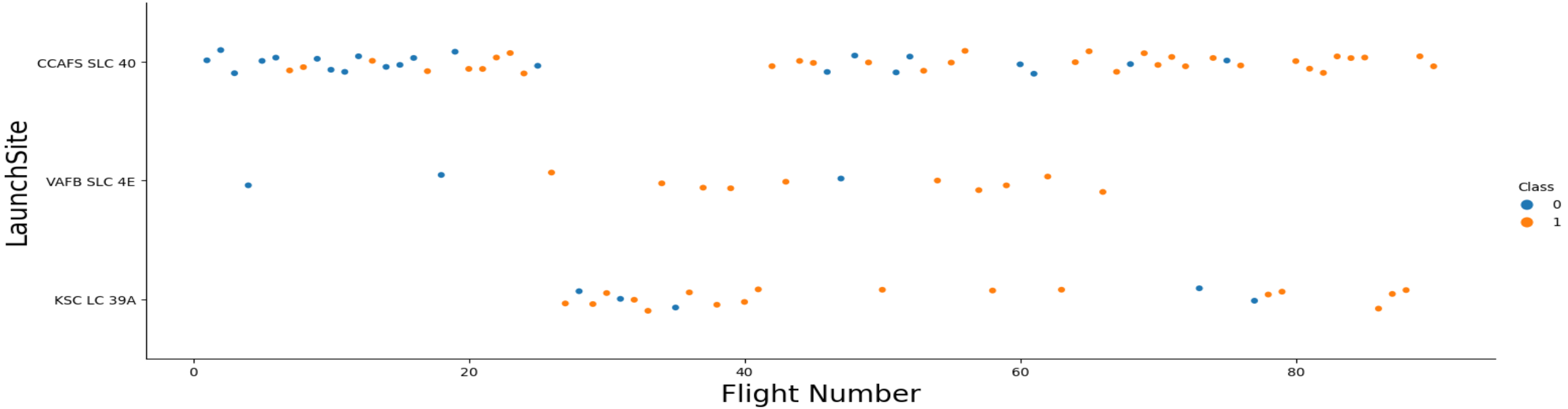


Flight Number vs. Launch Site

Findings:

The success rate of a Launch Site increases with the increase in Flight Number.

It is evident that the success rate has increased after 20 (Flight Number).



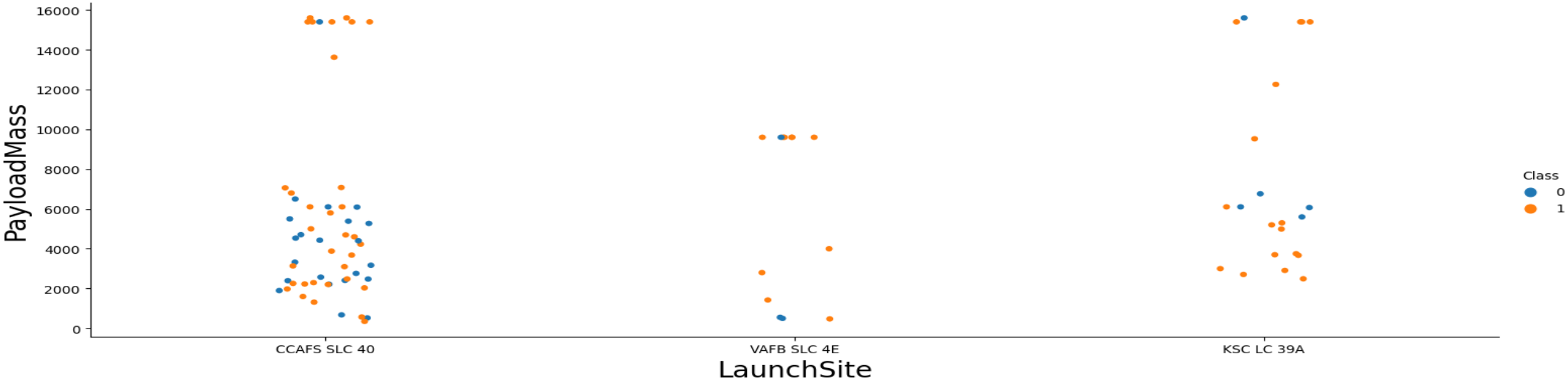
https://github.com/Nivedhika/Final-project-for-Data-Science

Payload vs. Launch Site

The success rate is higher with the increasing Payload mass at the launch site (CCAFS SLC 40).

The success rate is higher with the Payload mass less than 10000 kg at VAFB SLC 4E and KSC LC 39A.

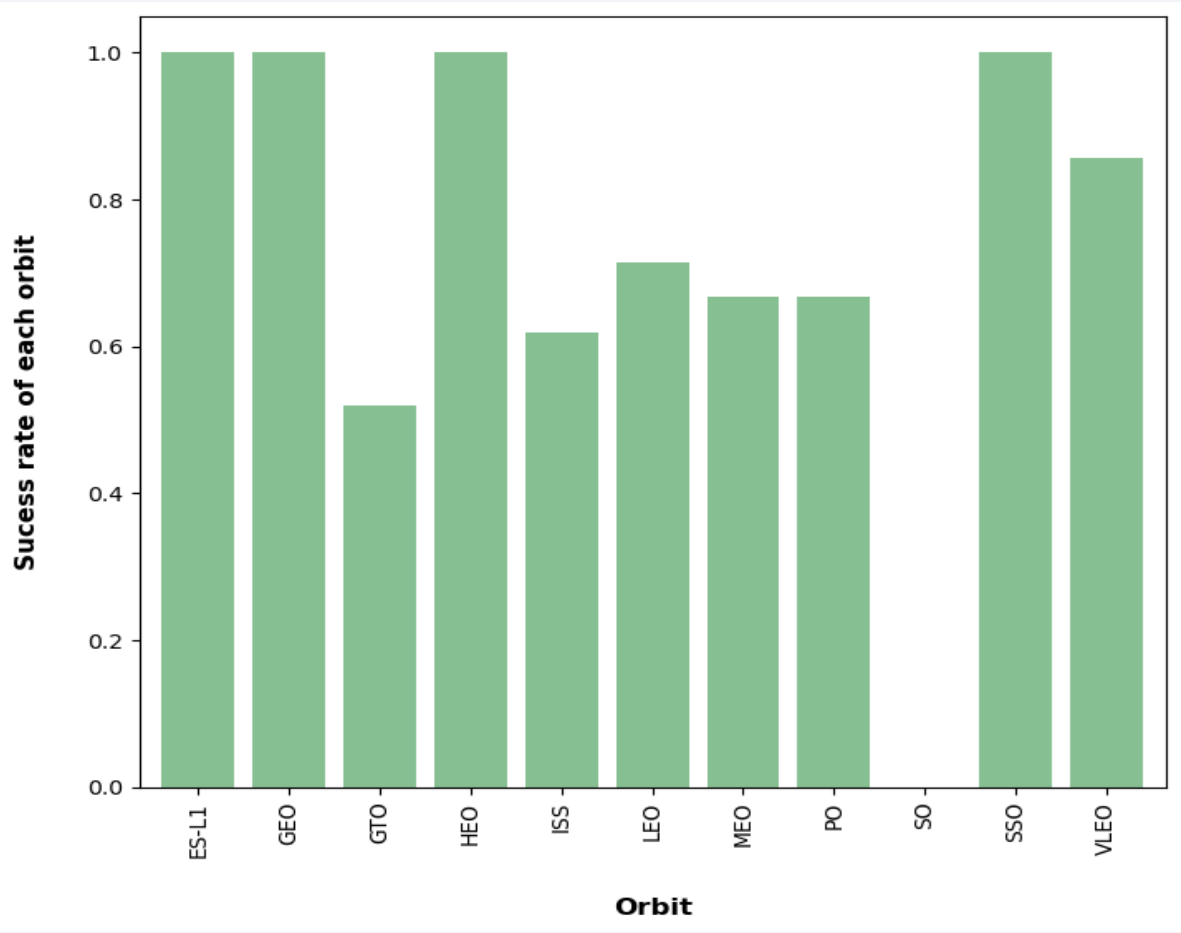
Also, no clear pattern to indicate if Payload mass can be considered as an important variable.



https://github.com/Nivedhika/Final-project-for-Data-Science

Success Rate vs. Orbit Type

The orbits ES-L1, GEO, HEO and SSO have recorded a higher success rate when compared with the rest.

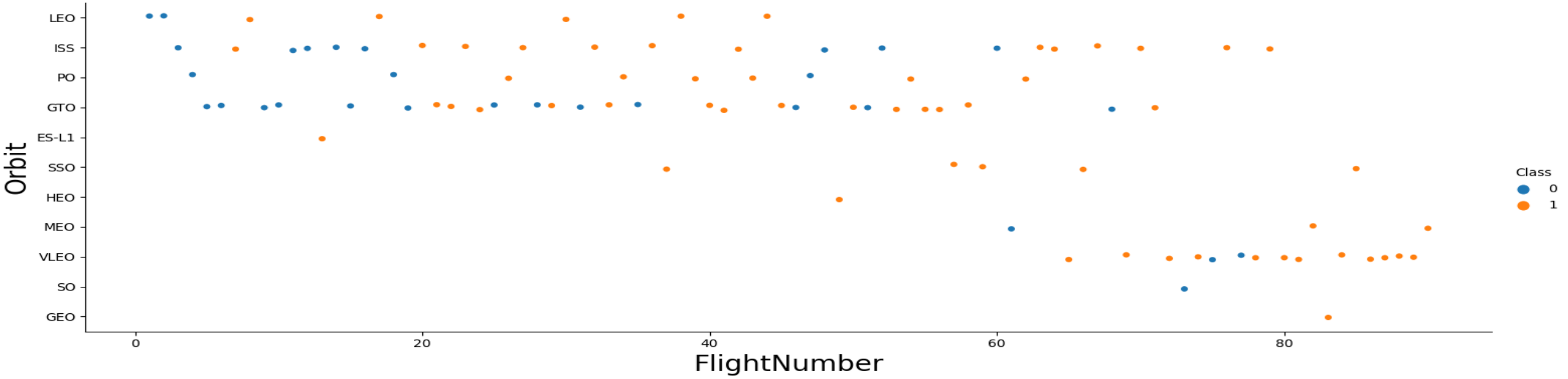


https://github.com/Nivedhika/Final-project-for-Data-Science

Flight Number vs. Orbit Type

Findings:

- The success rate in LEO with increasing Flight Number is evident.
- The ISS and GTO orbit shows no such patterns.



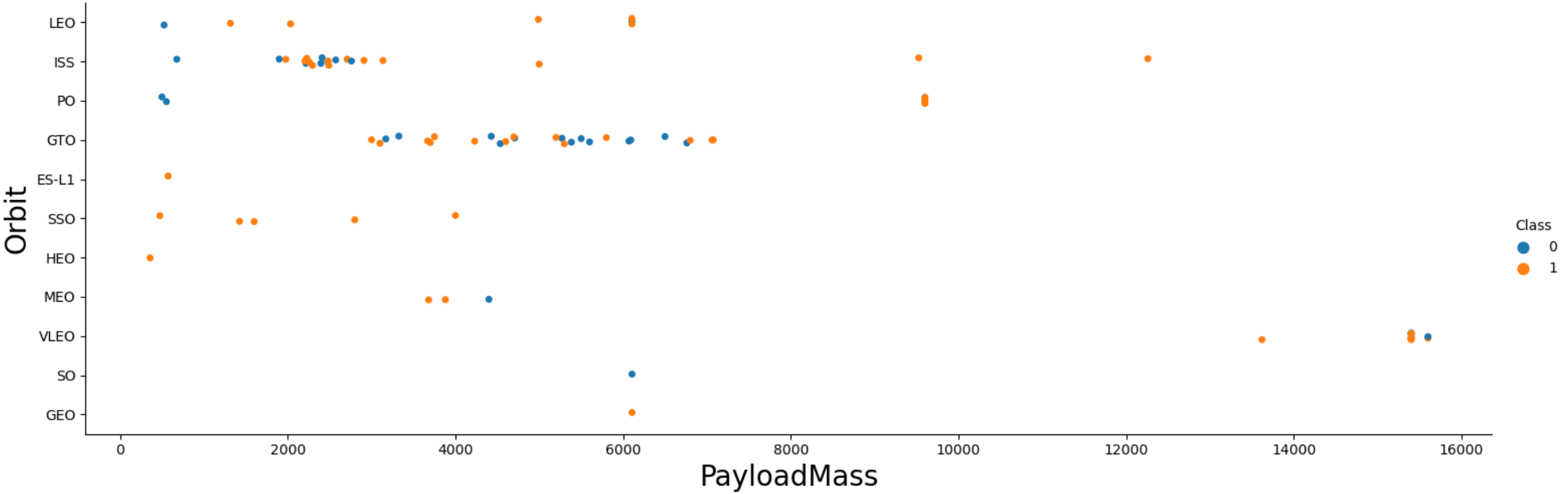
https://github.com/Nivedhika/Final-project-for-Data-Science

Payload vs. Orbit Type

VLEO has performed well with high payload mass between 14,000-16,000 kg.

SSO and LEO have shown a good record with the payload mass of less than 6000 kg.

GTO exhibits no clear pattern. https://github.com/Nivedhika/Final-project-for-Data-Science



All Launch Site Names

https://github.com/Nivedhika/Final-project-for-Data-Science

Listed names of all launch sites with the keyword 'DISTINCT' from the dataframe.

The launch is happening from four sites.

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

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

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

Launch Site Names Begin with 'CCA'

Using the predicate 'WHERE', listed first five rows containing the data related to launch site 'CCAFS-LC 40'.

- https://github.com/Nivedhika/Final-project-for-Data-Science

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

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

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

Applied sum function on Payload mass column of NASA(CRS).

Used 'WHERE' predicate to classify data from the column NASA (CRS).

https://github.com/Nivedhika/Final-project-for-Data-Science

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass by NASA (CRS)" FROM SPACEX WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Total Payload Mass by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

- The average payload mass of Booster Version remains in the range of 2000 kg.

https://github.com/Nivedhika/Final-project-for-Data-Science

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

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

Average Payload Mass by Booster Version F9 v1.1
2928.4

First Successful Ground Landing Date

- Used MIN method to arrive at the first landing date.

https://github.com/Nivedhika/Final-project-for-Data-Science

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

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Successful Drone Ship Landing with Payload between 4000 and 6000

- F9 FT version has remained more successful in the drone ship category.

https://github.com/Nivedhika/Final-project-for-Data-Science

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

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

https://github.com/Nivedhika/Final-project-for-Data-Science

The total number of successful and failure missions stands at 101.

The data was collected using the predicates 'COUNT' and 'WHERE'.

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Total Number of Successful and Failure Mission" FROM SPACEX \
WHERE MISSION_OUTCOME LIKE 'Success%' OR MISSION_OUTCOME LIKE 'Failure%';
```

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Total Number of Successful and Failure Mission
101

BOOSTERS CARRIED MAXIMUM PAYLOAD

https://github.com/Nivedhika/Final-project-for-Data-Science

Booster versions which carried a maximum payload were selected using the predicate 'WHERE'.

```
%sql SELECT DISTINCT BOOSTER_VERSION AS "Booster Versions which carried the Maximum Payload Mass" FROM SPACEX \
WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX);
```

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

https://github.com/Nivedhika/Final-project-for-Data-Science

Selected Booster version and launch site when drone ship failed during 2015.

```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEX WHERE DATE LIKE '2015-%' AND \
LANDING__OUTCOME = 'Failure (drone ship)';
```

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

- Listed landing outcome and total count of the same between 2010 and 2017.
- Used GROUP BY and ORDER BY methods to segregate it.

https://github.com/Nivedhika/Final-project-for-Data-Science

```
%sql SELECT LANDING__OUTCOME as "Landing Outcome", COUNT(LANDING__OUTCOME) AS "Total Count" FROM SPACEX \
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY  LANDING__OUTCOME \
ORDER BY COUNT(LANDING__OUTCOME) DESC ;
```

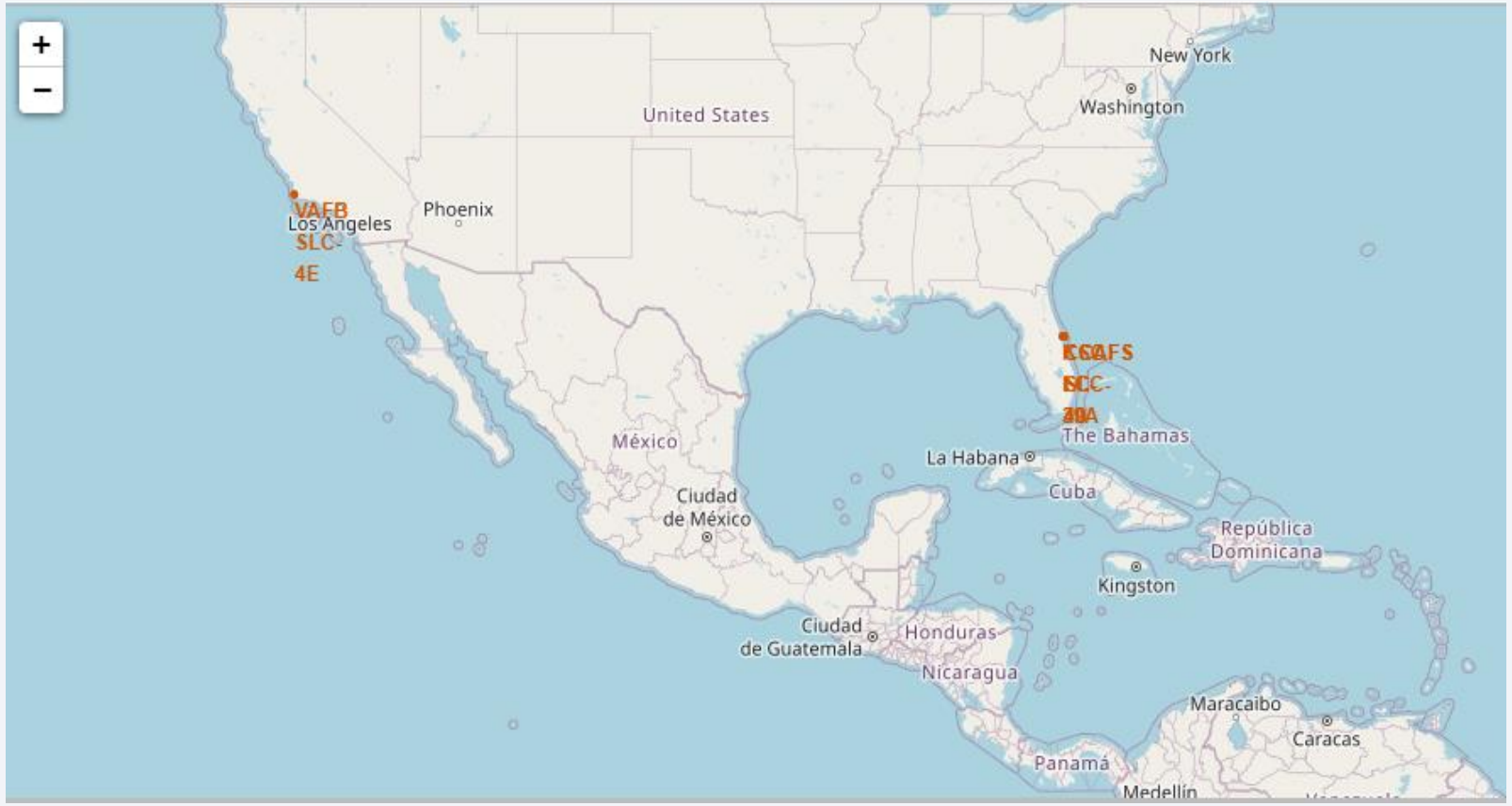
```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

The launch sites - CCAFS LC-40, KSC LC-39A and VAFB SLC 4E – are located near the coasts.



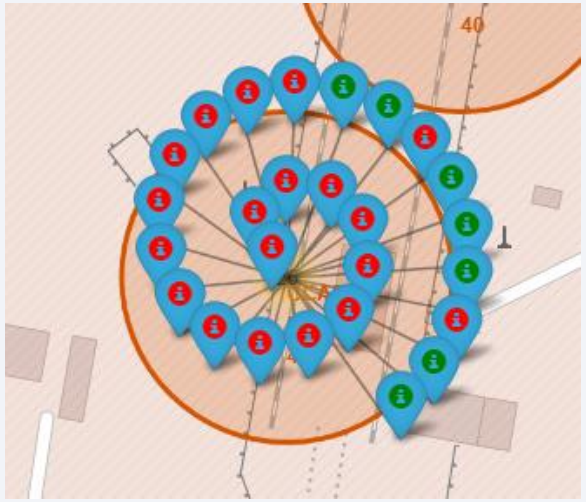
https://github.com/Nivedhika/Final-project-for-Data-Science

FOLIUM-MARKERS

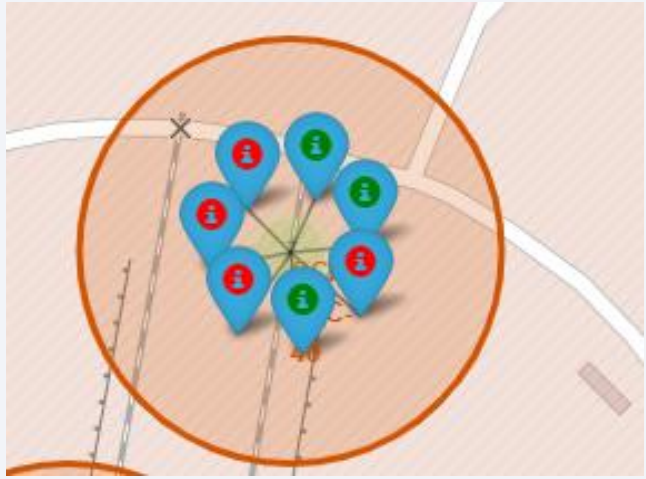
- The red marker indicates failed mission and green reflects successful mission.
- The launch site KSC-LC-39A has recorded 10 successful missions out of 13.

https://github.com/Nivedhika/Final-project-for-Data-Science

VAFB-SLC-4E



CCAFS-LC 40



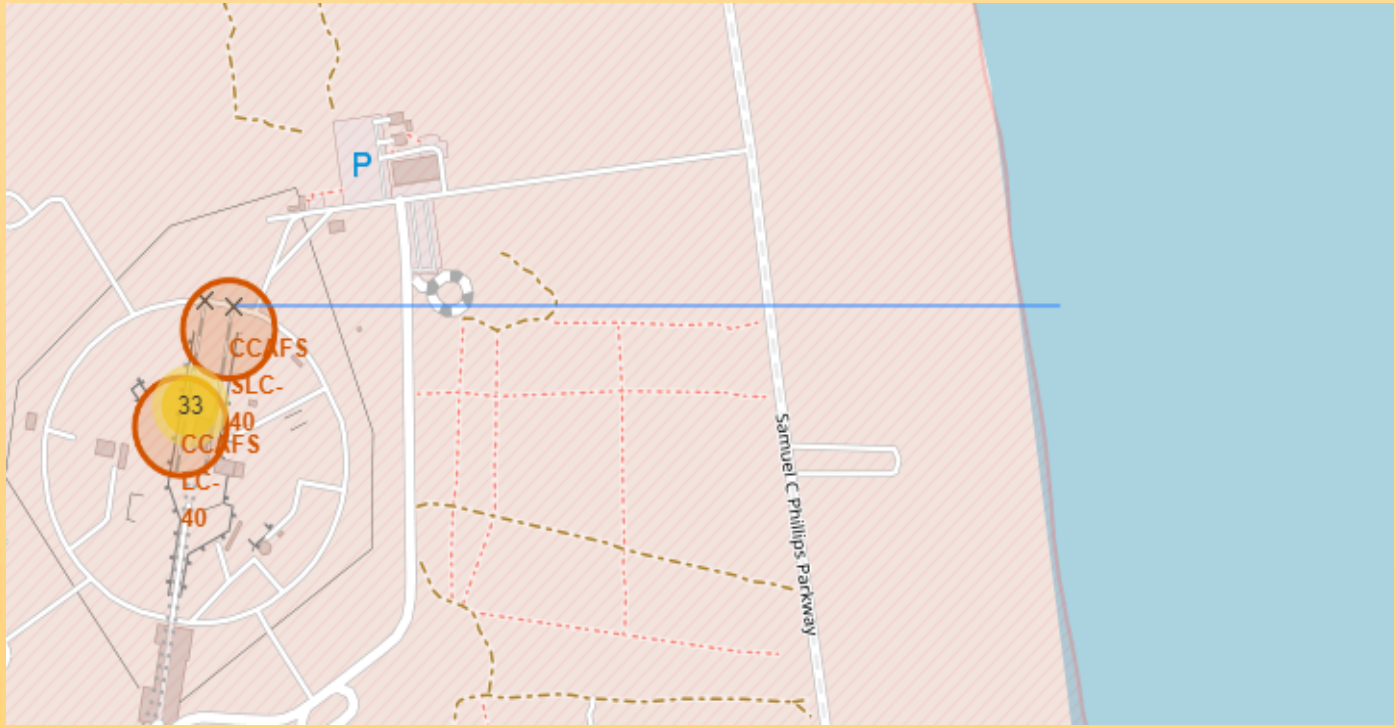
CCAFS-SLC 40

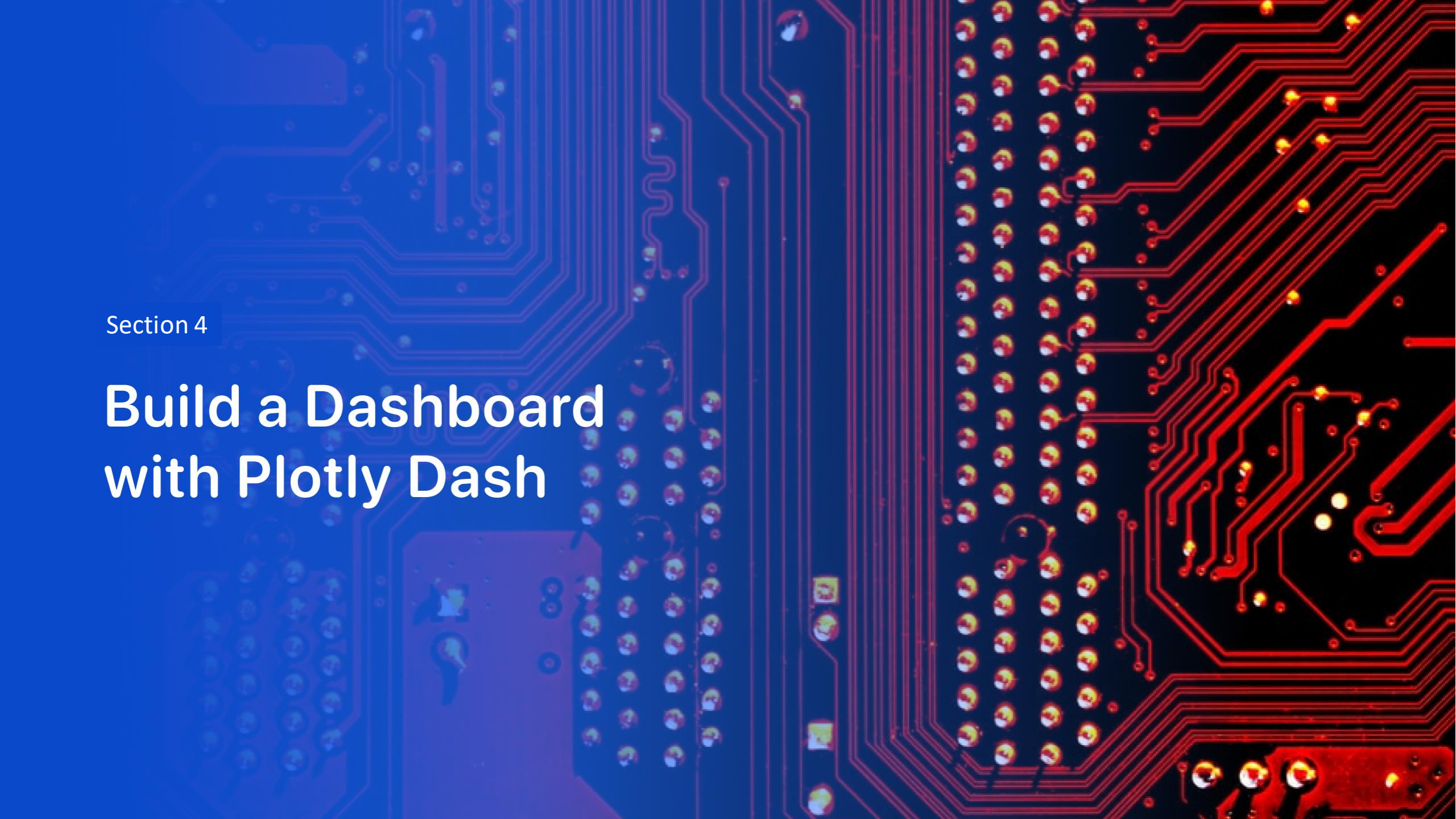
KSC-LC 39A

NEAREST COASTLINE

- The distance between CCAFS SLC-40 and the nearest coastline is 0.90 km.

https://github.com/Nivedhika/Final-project-for-Data-Science




Section 4
Build a Dashboard with Plotly Dash

VISUALISATION - PLOTLY

- KSC LC-39A has the highest success rates.
- The rockets whose Payload mass ranged between 2000 and 10,000 kg have witnessed higher success rates.
- F9 Booster version FT has witnessed highest success rate.

https://github.com/Nivedhika/Final-project-for-Data-Science

Section 5

Predictive Analysis (Classification)

CLASSIFICATION ACCURACY

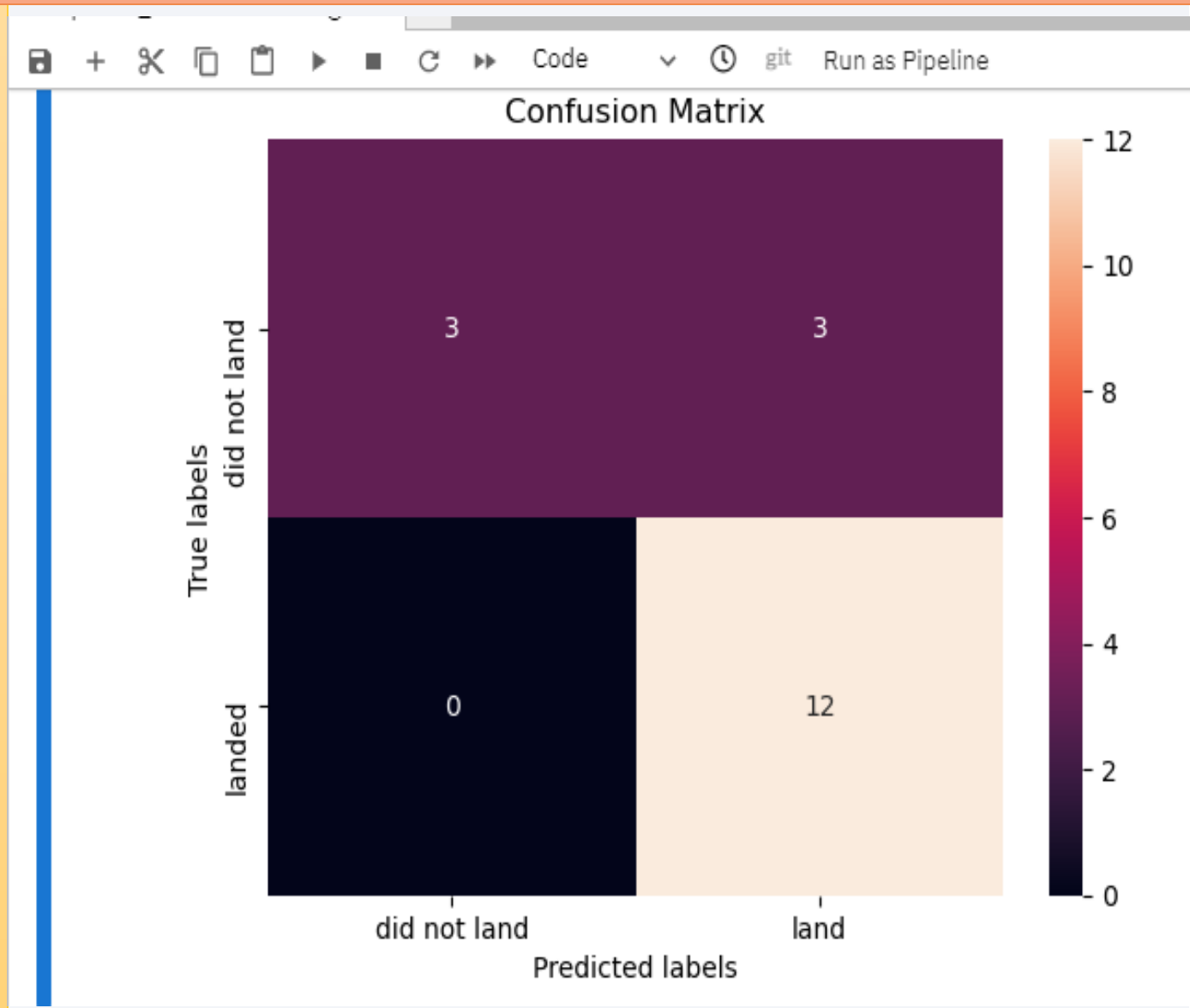
https://github.com/Nivedhika/Final-project-for-Data-Science

MODELS	ACCURACY
LOGISTIC REGRESSION	0.84
SUPPORT VECTOR MACHINE	0.84
DECISION TREE MODEL	0.90
KNN	0.84

CONFUSION MATRIX

https://github.com/Nivedhika/Final-project-for-Data-Science

- Total cases – 18
- Out of 6 actual 'NO', the model has predicted only 3.
- Precision-0.8



CONCLUSION

a. The orbits - ES-L1, GEO, HEO and SSO - have witnessed a higher success rate compared with others.

b. A most of the successful launches were carried out with a Payload mass equal to or less than 6000 kg.

c. Further, the average Payload mass of Booster Version hovered around 2000 kg.

d. Given data suggests that the launch site (KSC LC 39A) has witnessed a higher success rate than others.

https://github.com/Nivedhika/Final-project-for-Data-Science

Thank you!