IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anand kumar
16th December 2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- In this We are going to use the skills learned throughout the specialization on the single project
- Encountered various difficulty with data cleaning, data imputation
- Made interactive plots with Folium and Web app with plotly dash
- Used Machine Learning to predict the landing outcome of a payload

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- 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 – SpaceX API

- Data Source Wikipedia wikipedia data source

GitHub: https://github.com/Anandkumar267/Data-Science/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection - Scraping

1. Used BeautifulSoup to traverse the web page and construct the dataset
2. Made custom function to extract each column
3. Made a Dataframe in pandas for web scrapped data
4. Exported to CSV File

Data Wrangling

1. Collected data from wikipedia
2. Applied custom function to clean the values and remove unwanted hyperlinks and references
3. Made a final DataFrame in Pandas
4. Converted date column to Datatime dtype
5. Numerical column to float

EDA with Data Visualization

GitHUB:
https://github.com/Anandkumar267/Data-Science/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

SQL IS VERY HANDY IN EDA INVOLVING SMALL TO LARGE DATABASES

Columns of data are

Name

Date

Time (UTC)

Booster_Version

Launch_Site

Payload

PAYLOAD_MASS__KG_

Orbit, Customer, Mission_Outcome, Landing_Outcome

Build an Interactive Map with Folium

Graphs are best way to visualize the location based data

Easy to understand and can identify factors like distance between location

Folium is python package enables to visualize map

Build a Dashboard with Plotly Dash

Sucess rate is more than 70% for all of them individually

Background color and Font color is changed for graphs for better visibility

GitHub: https://github.com/Anandkumar267/Data-Science/blob/main/plotly_spacex.py

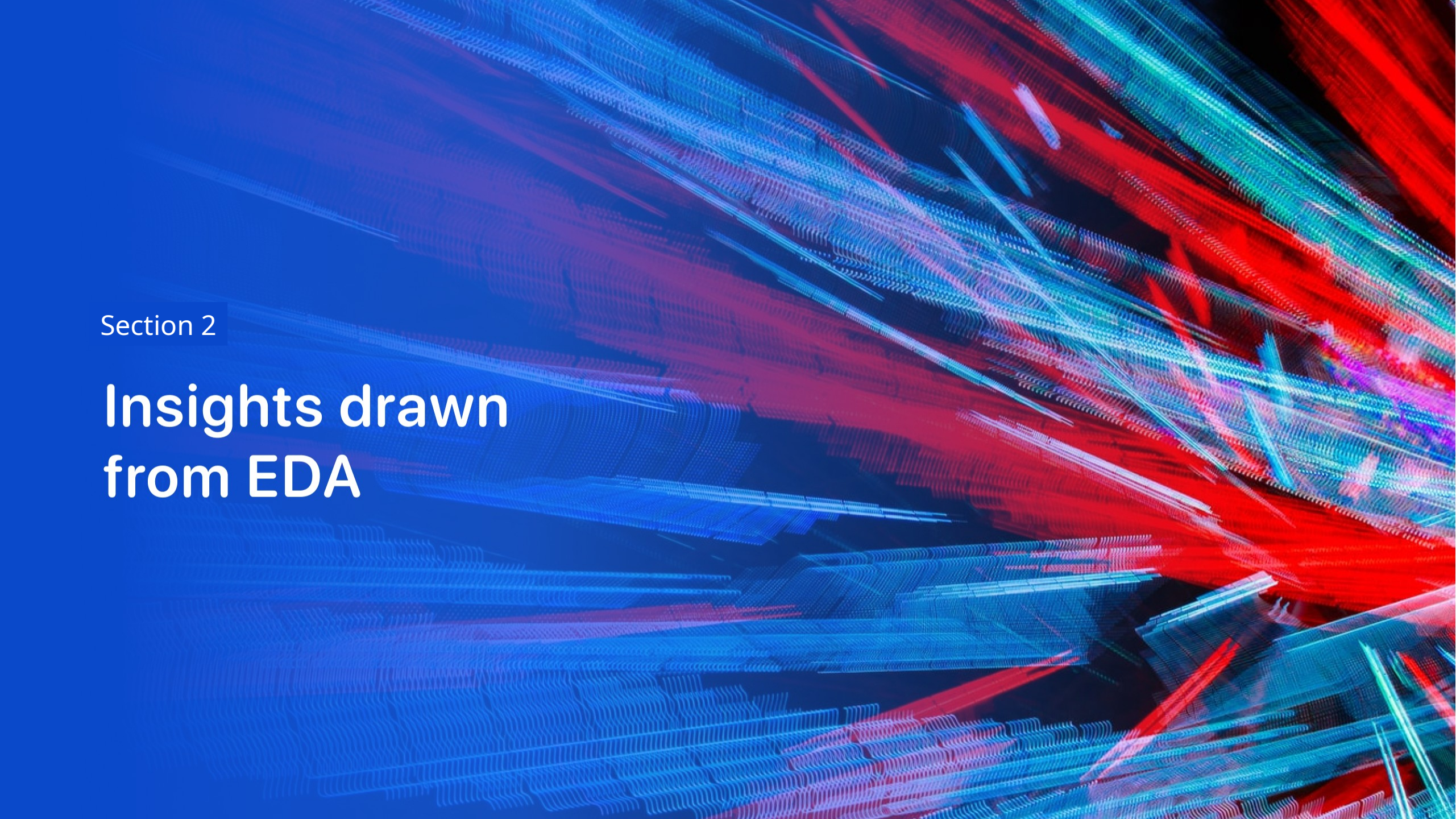
Predictive Analysis (Classification)

KNN is best performing model in my test with highest test accuracy over LogisticRegression, DecisionTreeClassifier, SVC,

GitHub URL :
https://github.com/Anandkumar267/Data-Science/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- There is some decrease in Success Rate in 2018 of SpaceX
- Success is more in number than failure
- With upcoming the success rate is increasing

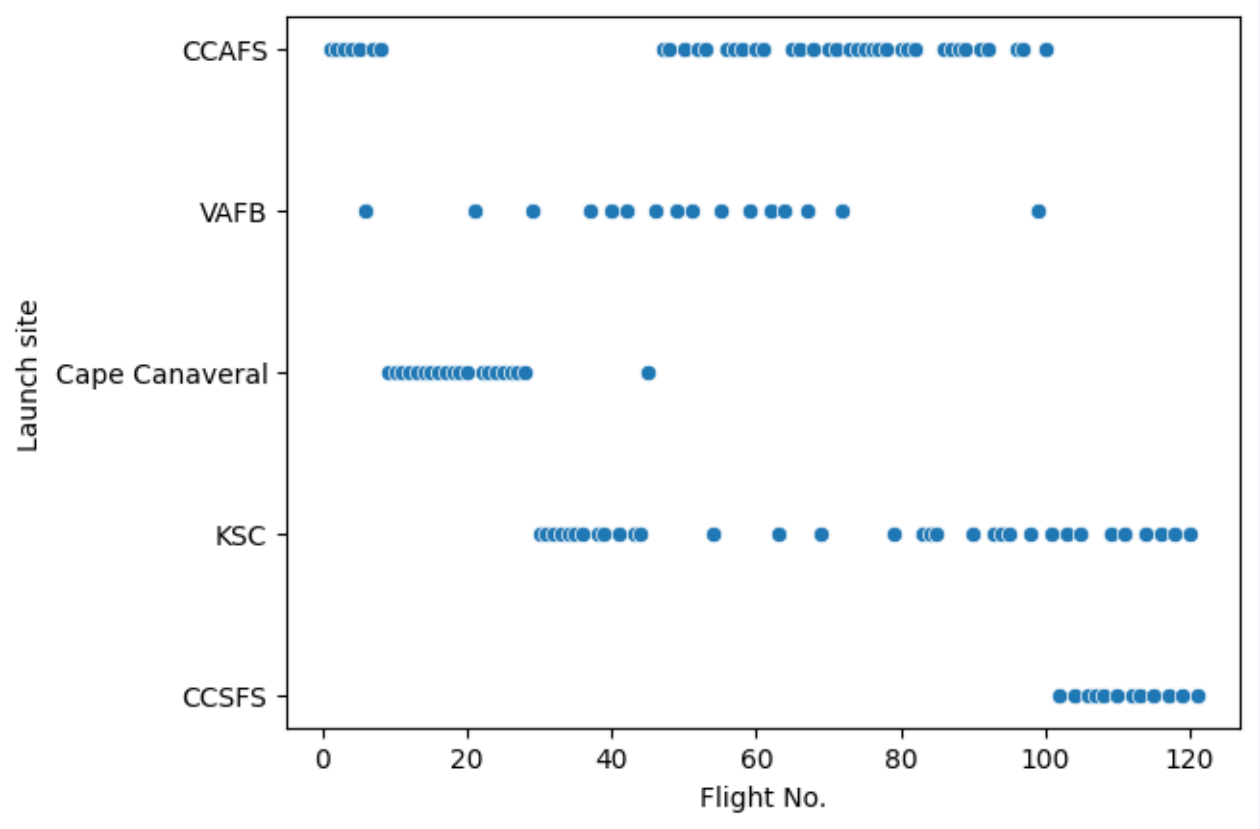


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

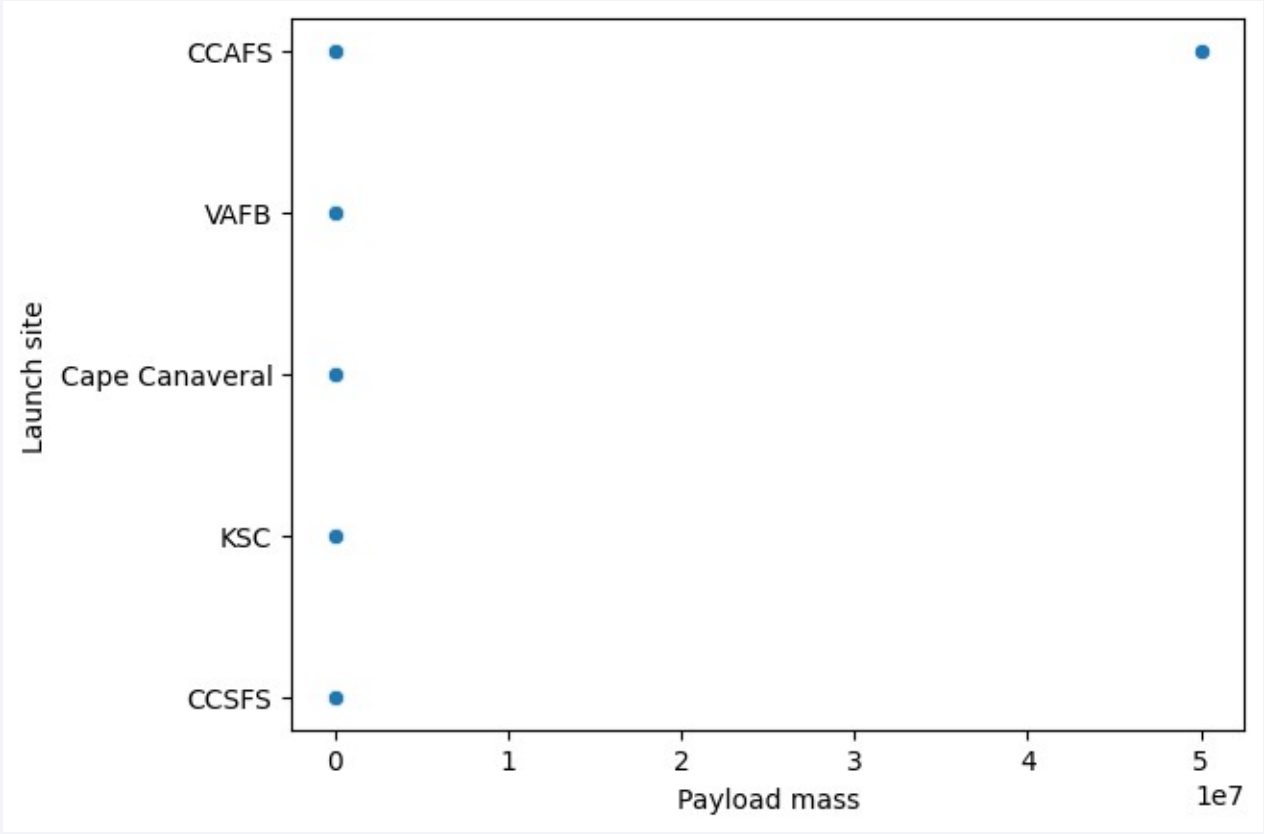
- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



This shows CCSFS only launch with Flight No. 100 to 120 and Cape Canaveral Flight No. 0 to 50 Rest are evenly distributed.

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations

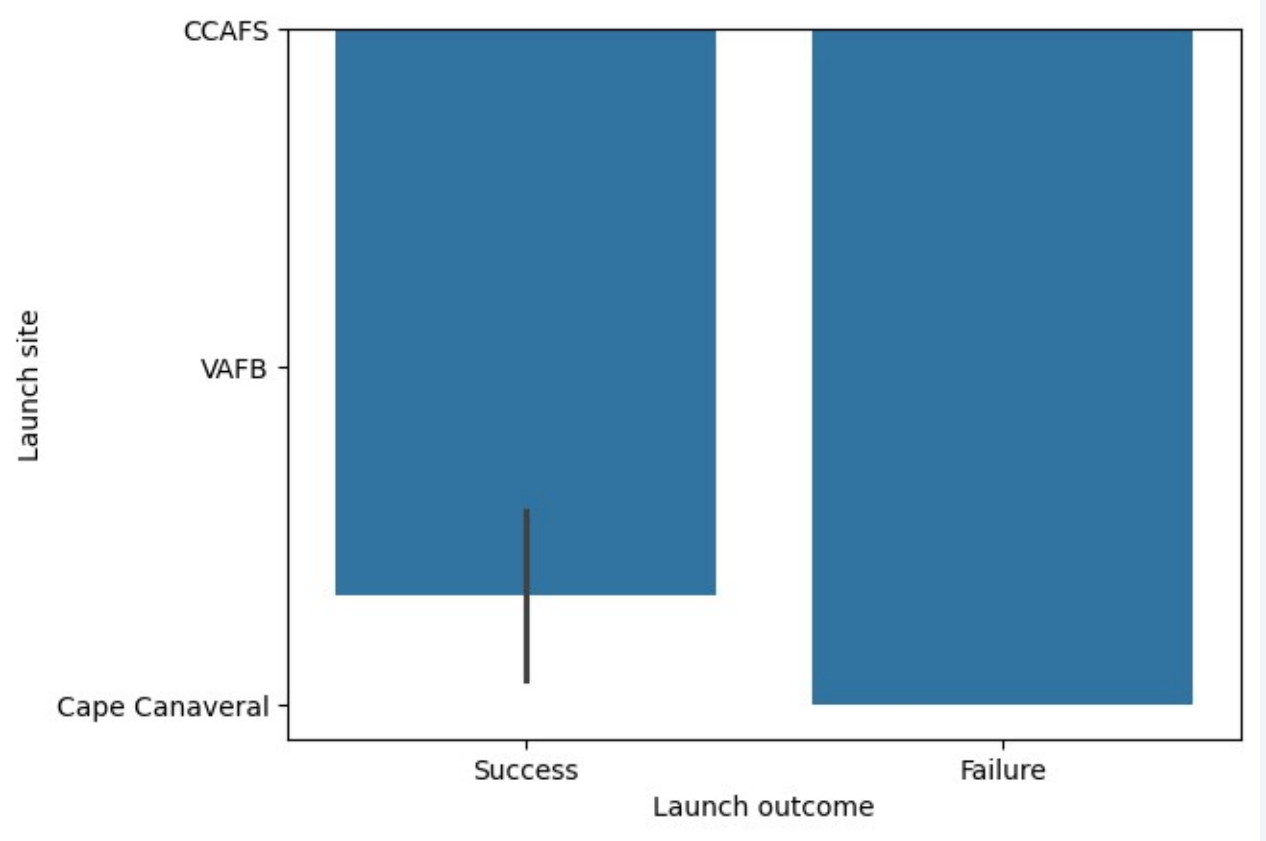


Payload mass 0 to 14000 kg maximum but for CCAFS there is anomly of 50000000kg

Success Rate vs. Orbit Type

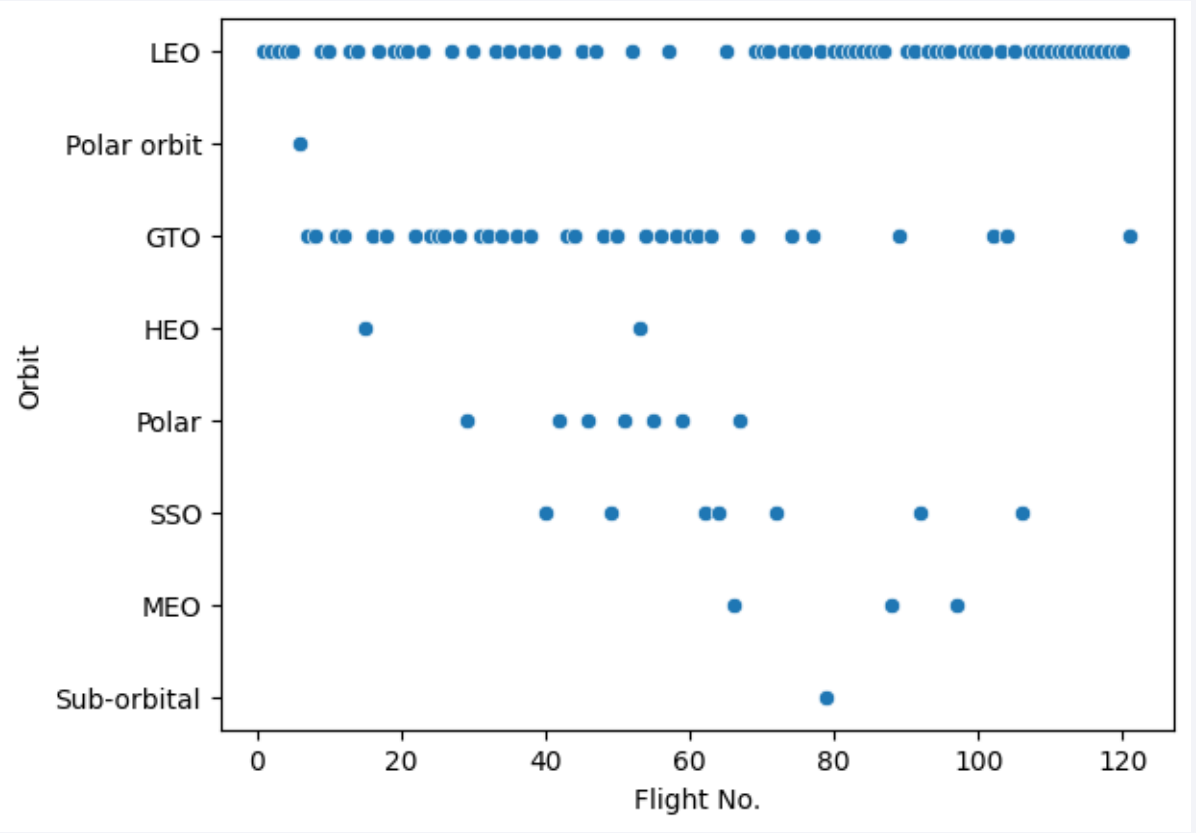
- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations



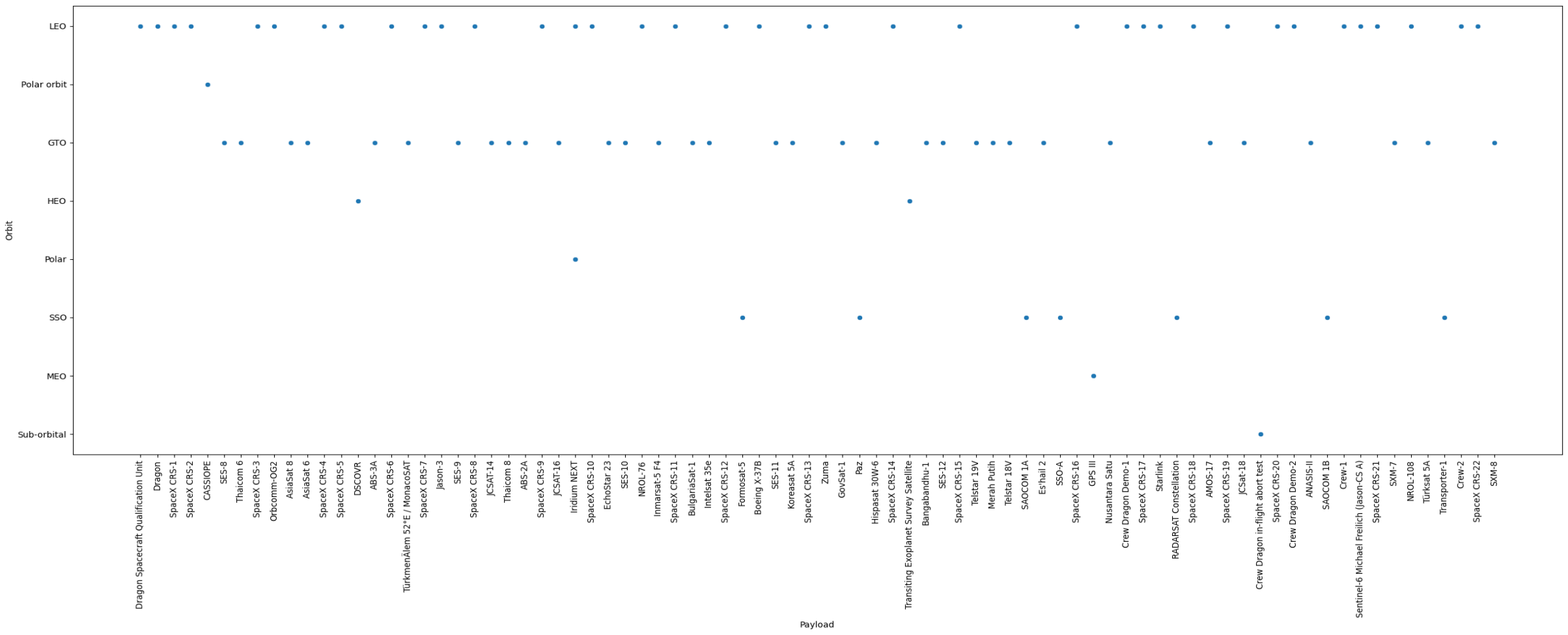
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



1. In Sub orbital only Flght No. 80 operate
2. Approximatly all Flight No. operate in LEO and GTO

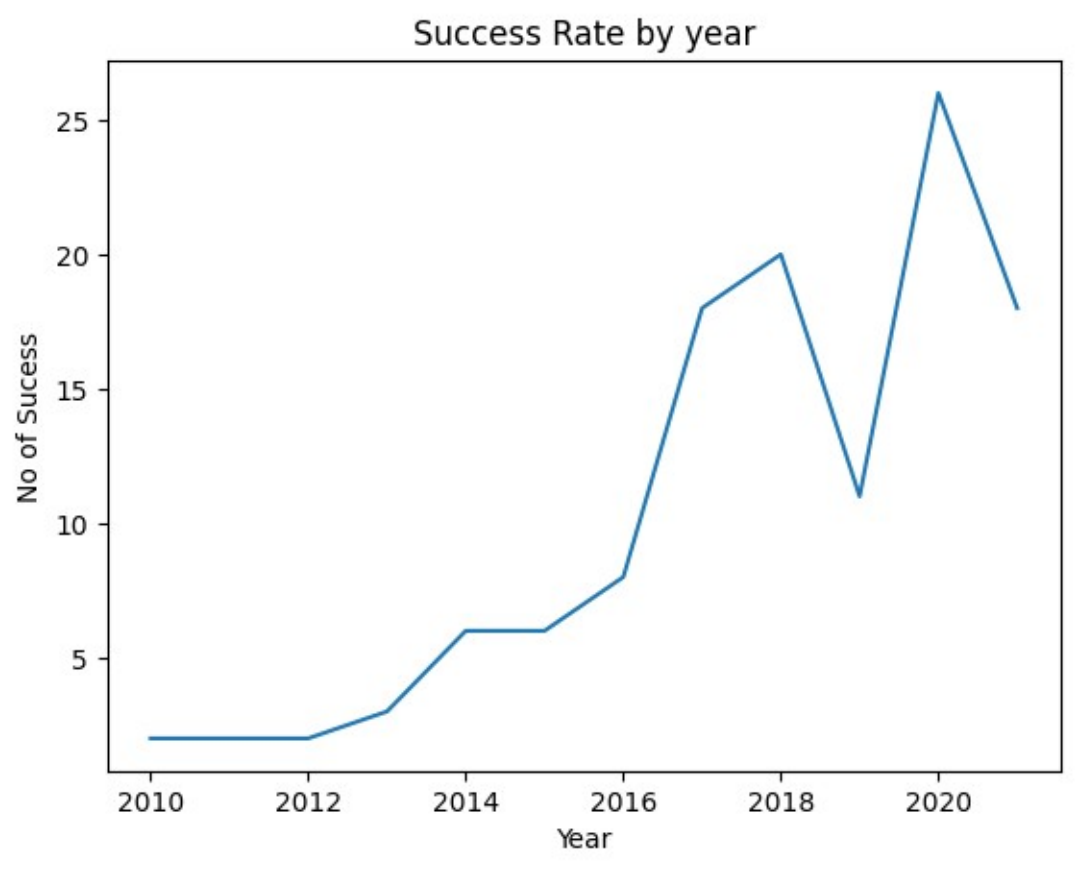
Payload vs. Orbit Type



Only LEO and GTO are orbit which all payload operate

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations



There is decrease in success rate in 2019
Reason for decrease After 2020 is covid-19

All Launch Site Names

Find the names of the unique launch sites

Present your query result with a short explanation here

```
SELECT DISTINCT LAUNCH_SITE FROM SPACEXTABLE
```

THIS PRINT UNIQUE LAUNCH_SITE

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

Present your query result with a short explanation here

```
SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE "CCA%" LIMIT 5
```

“CCA%” search for string begins with CCA ends with anything

Total Payload Mass

Calculate the total payload carried by boosters from NASA

Present your query result with a short explanation here

```
SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE
Customer="NASA (CRS)"
```

OUTPUT: 45596

Average Payload Mass by F9 v1.1

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

Present your query result with a short explanation here

```
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
WHERE Booster_Version ="F9 v1.1"
```

OUTPUT = 2928.4kg

First Successful Ground Landing Date

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

Present your query result with a short explanation here

```
SELECT Date FROM SPACEXTABLE WHERE Landing_Outcome="Success
(ground pad)"
```

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

Present your query result with a short explanation here

```
SELECT Booster_Version FROM SPACEXTABLE WHERE
Landing_Outcome="Success (drone ship)" AND PAYLOAD_MASS__KG_ >=4000 AND
PAYLOAD_MASS__KG_ < 6000
```

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

Present your query result with a short explanation here

```
SELECT Landing_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY
Landing_Outcome HAVING Landing_Outcome LIKE "Success%" OR "Failure%"
```

Boosters Carried Maximum Payload

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

Present your query result with a short explanation here

```
SELECT Booster_Version, PAYLOAD_MASS__KG_ 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

Present your query result with a short explanation here

```
SELECT substr(Date, 6,8) AS MONTH, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTABLE WHERE substr(Date, 0,5)='2015' AND Landing_Outcome LIKE "Failure
%"
```

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

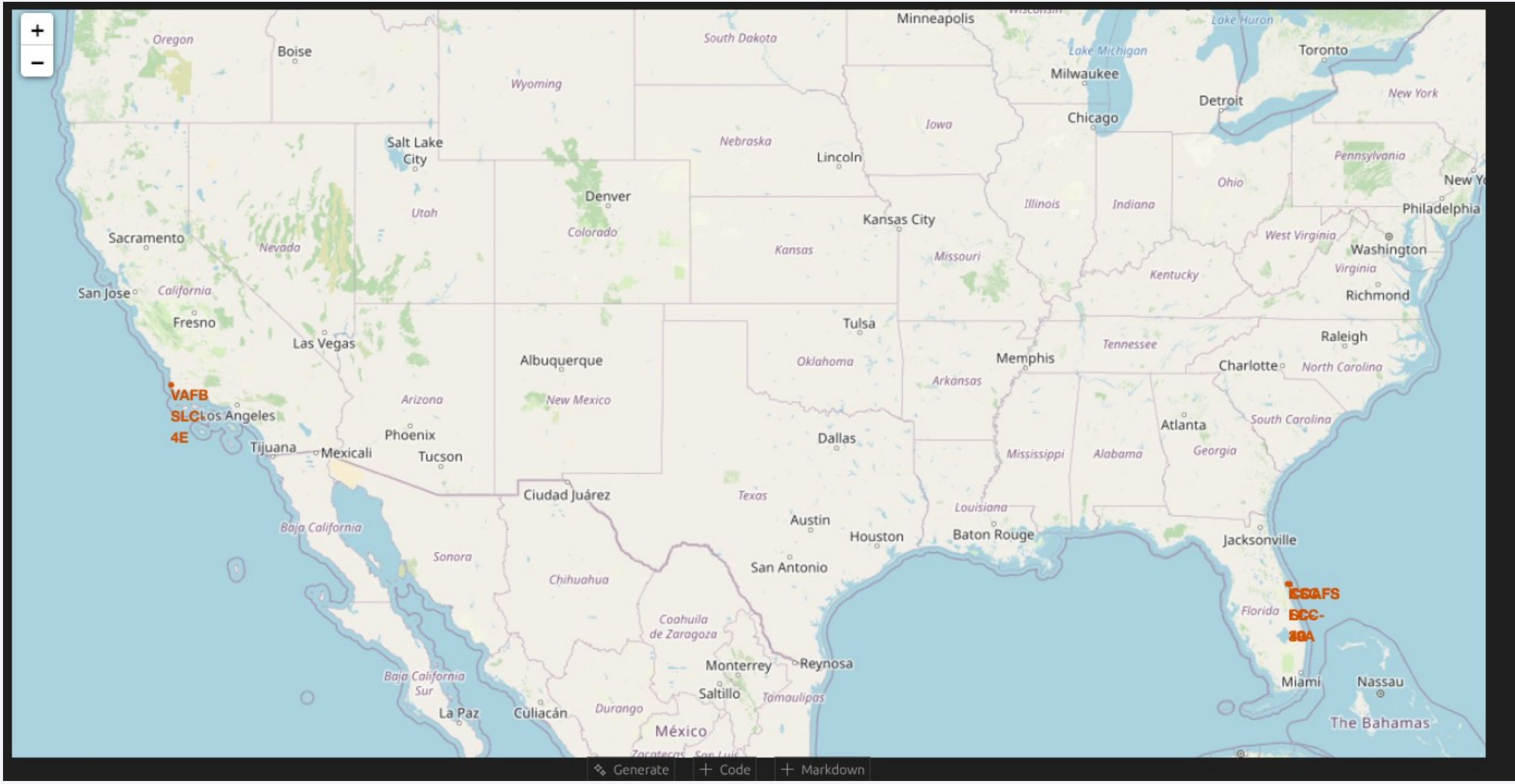
Present your query result with a short explanation here

```
SELECT Landing_Outcome, COUNT(*) AS count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY count DESC
```

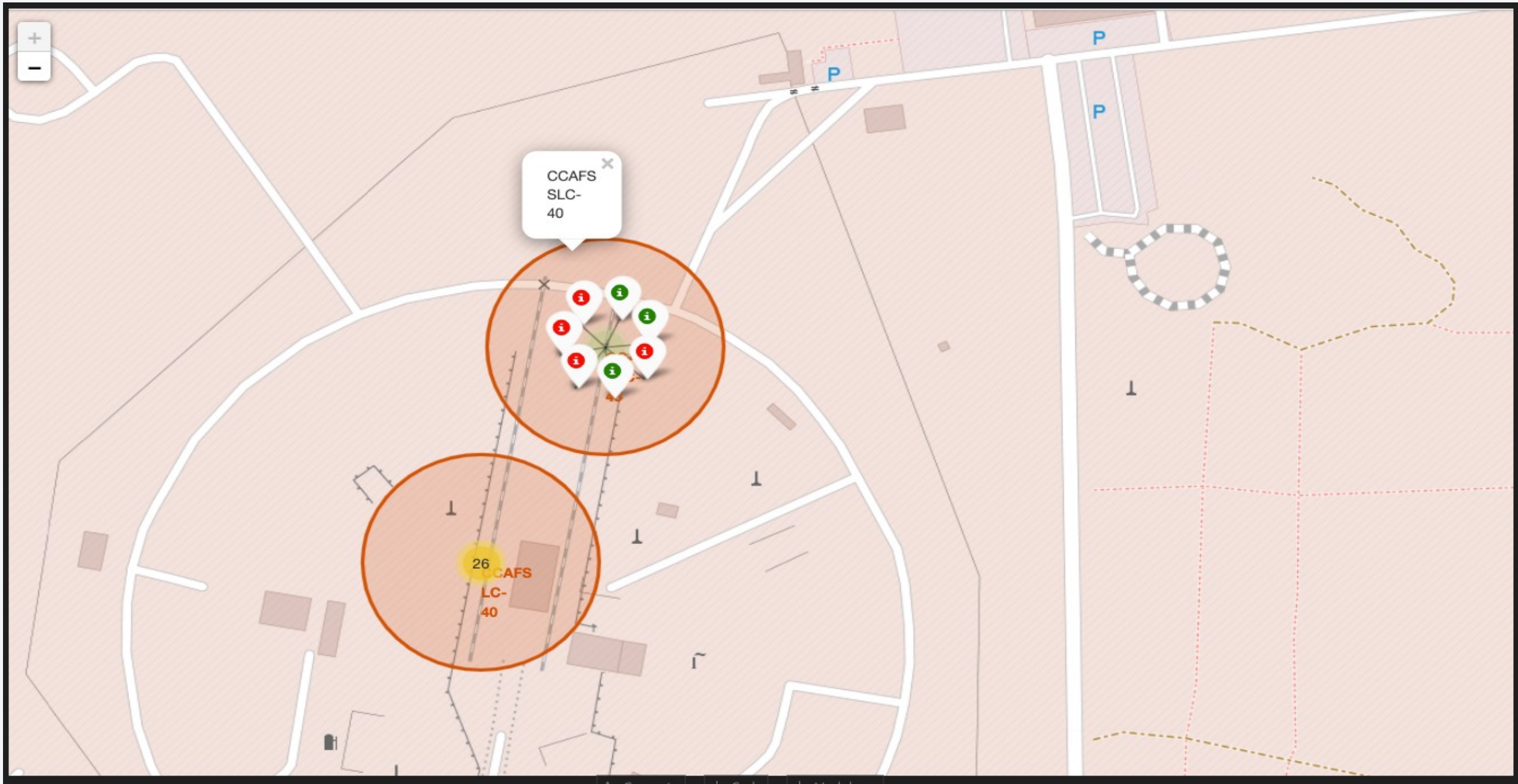
Section 3

Launch Sites Proximities Analysis

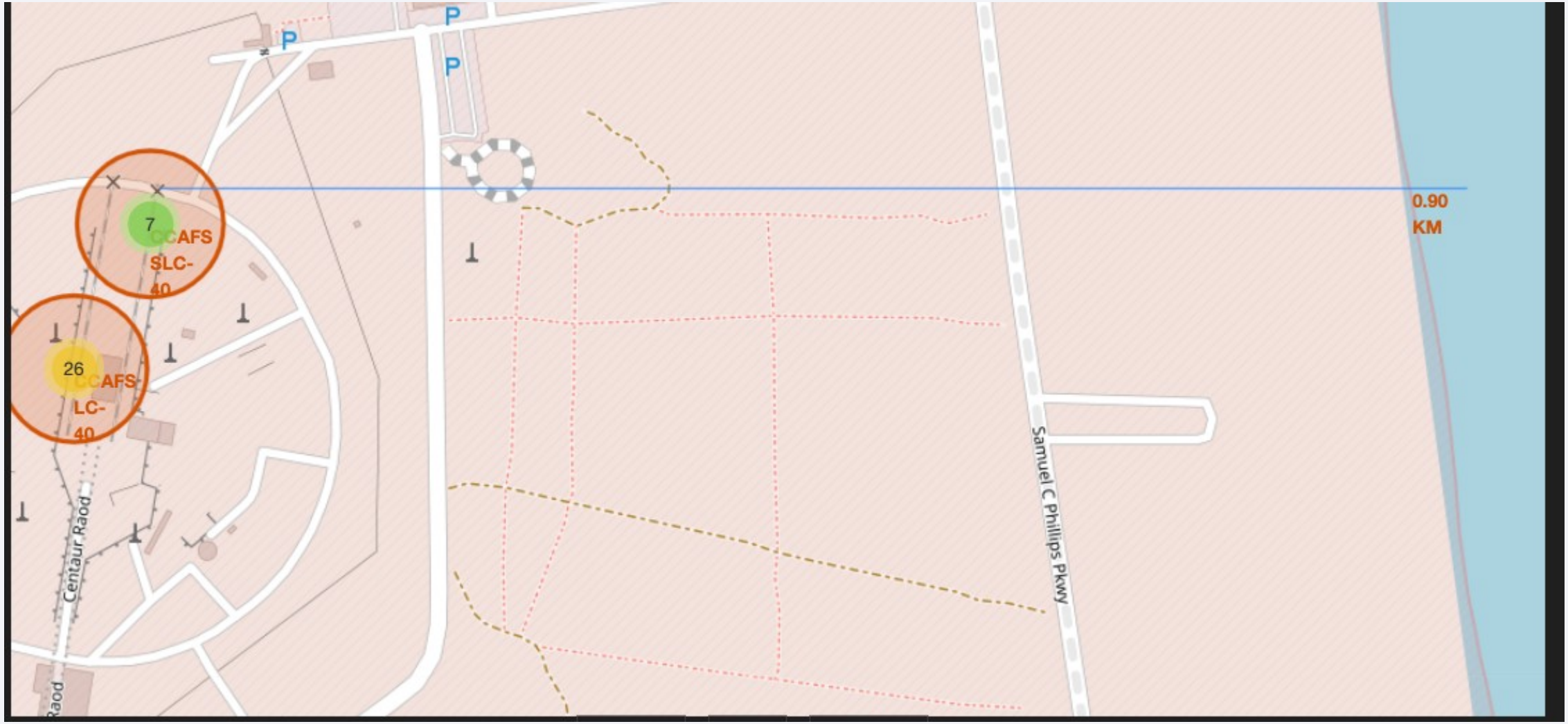
<Folium Map Screenshot 1>

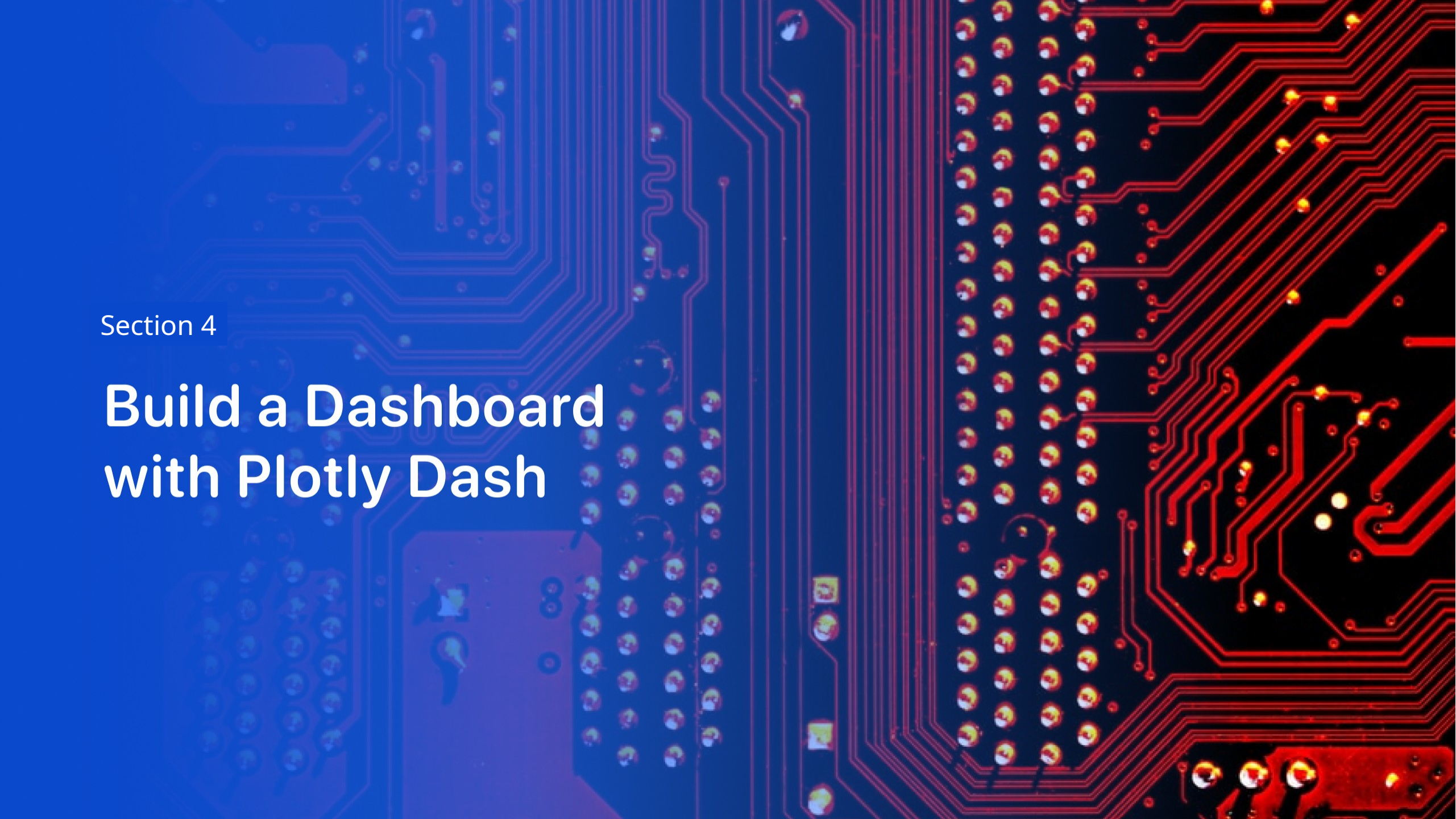


<Folium Map Screenshot 2>

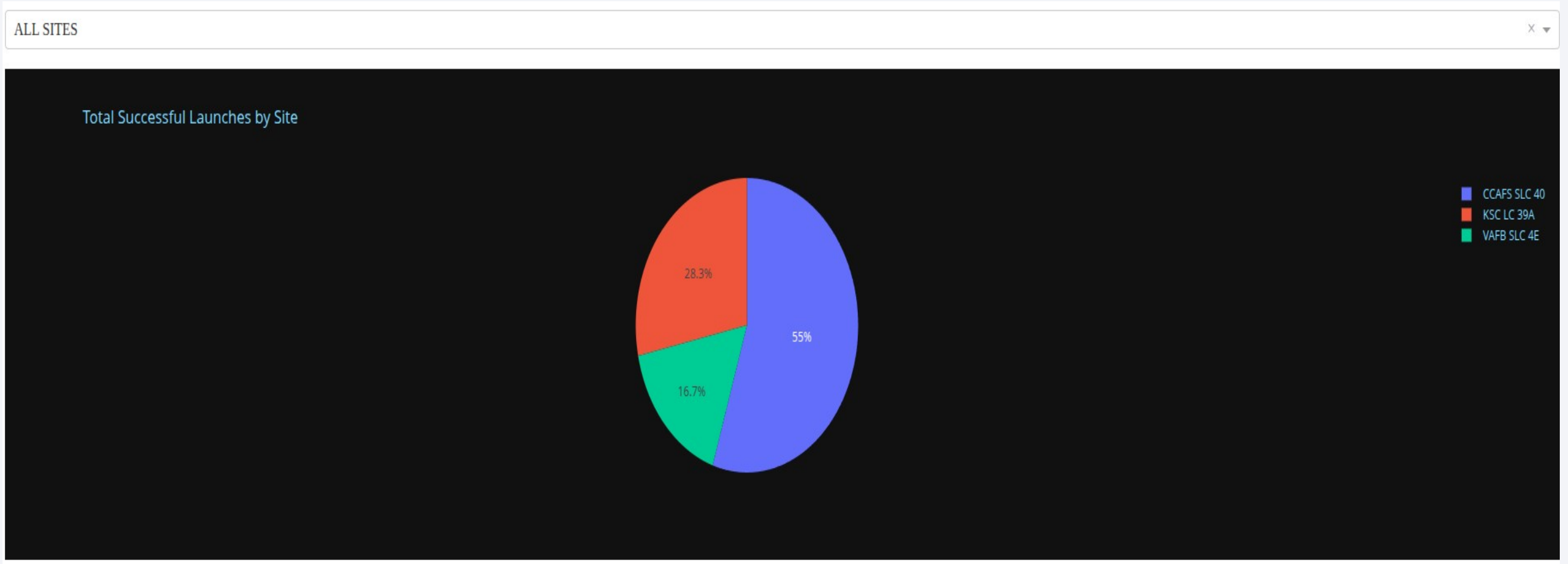


<Folium Map Screenshot 3>

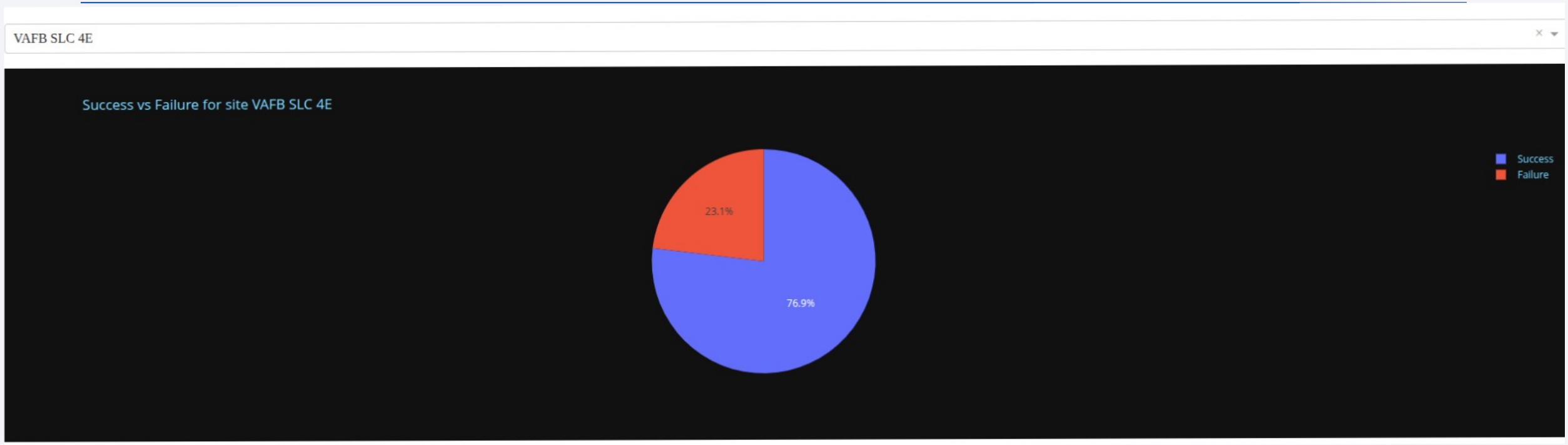



Section 4
Build a Dashboard with Plotly Dash

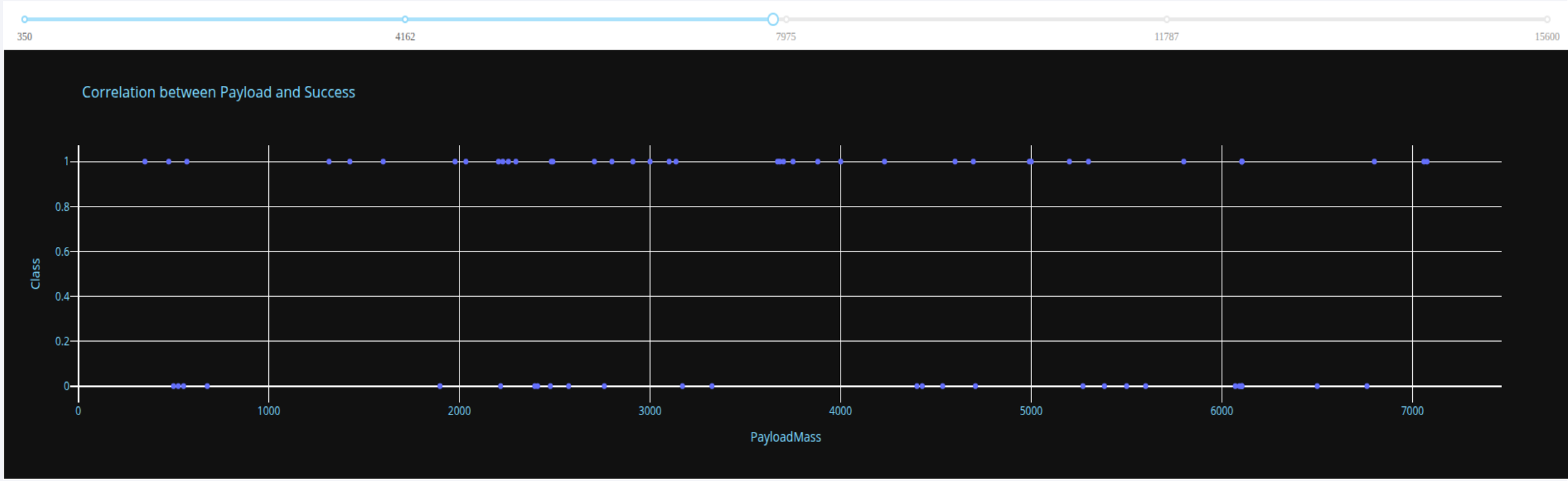
<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



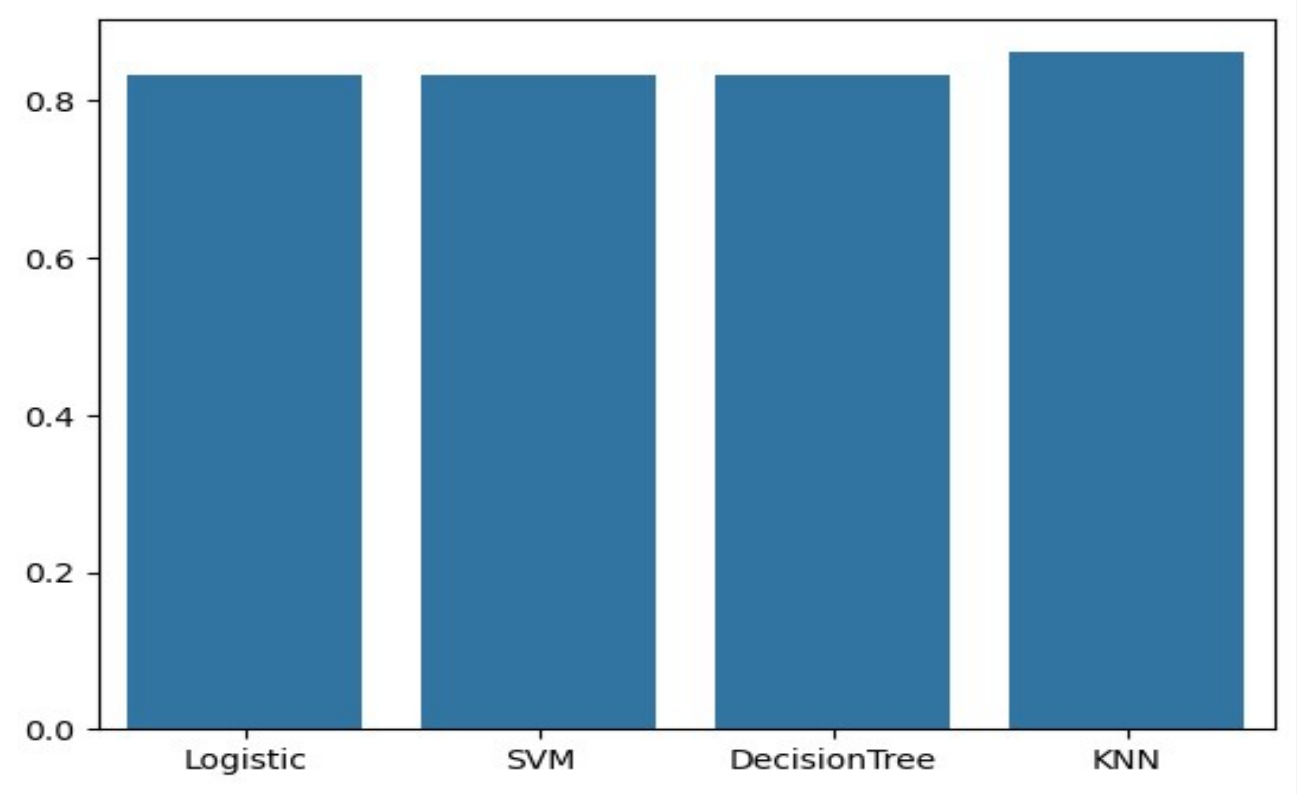
<Dashboard Screenshot 3>



Section 5

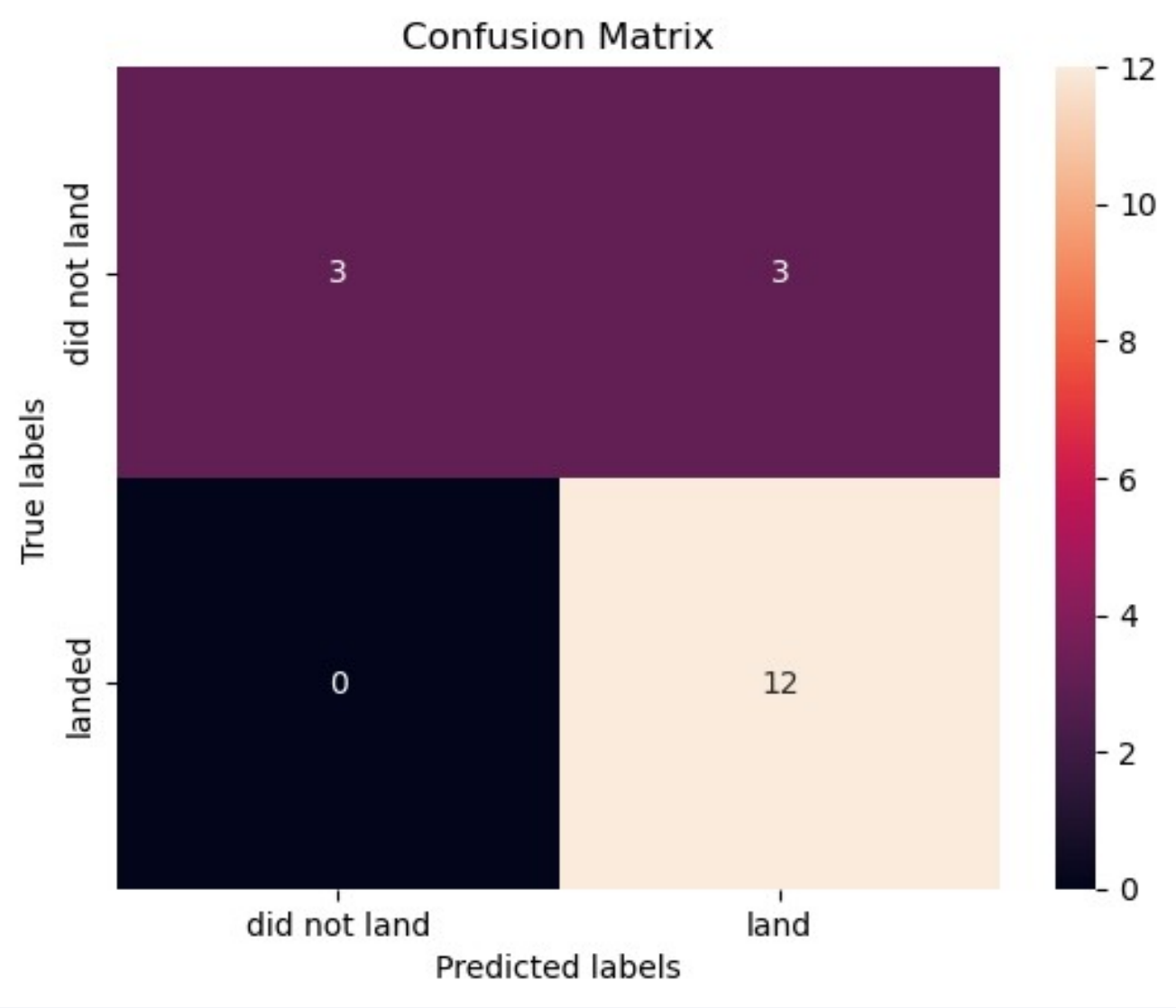
Predictive Analysis (Classification)

Classification Accuracy



KNN has highest Test Accuracy

Confusion Matrix



- Very Good Confusion Matrix S

Conclusions

- Classification Model can predict failure or success to a high degree

- KNN is best classifier out of four used in prediction

- Higher Accuracy can be achieved by tuning the hyperparameters

- Feature Engineering can greatly increase model ability to identify patterns and in turn higher accuracy

Appendix

Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!