IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rose Hashemi

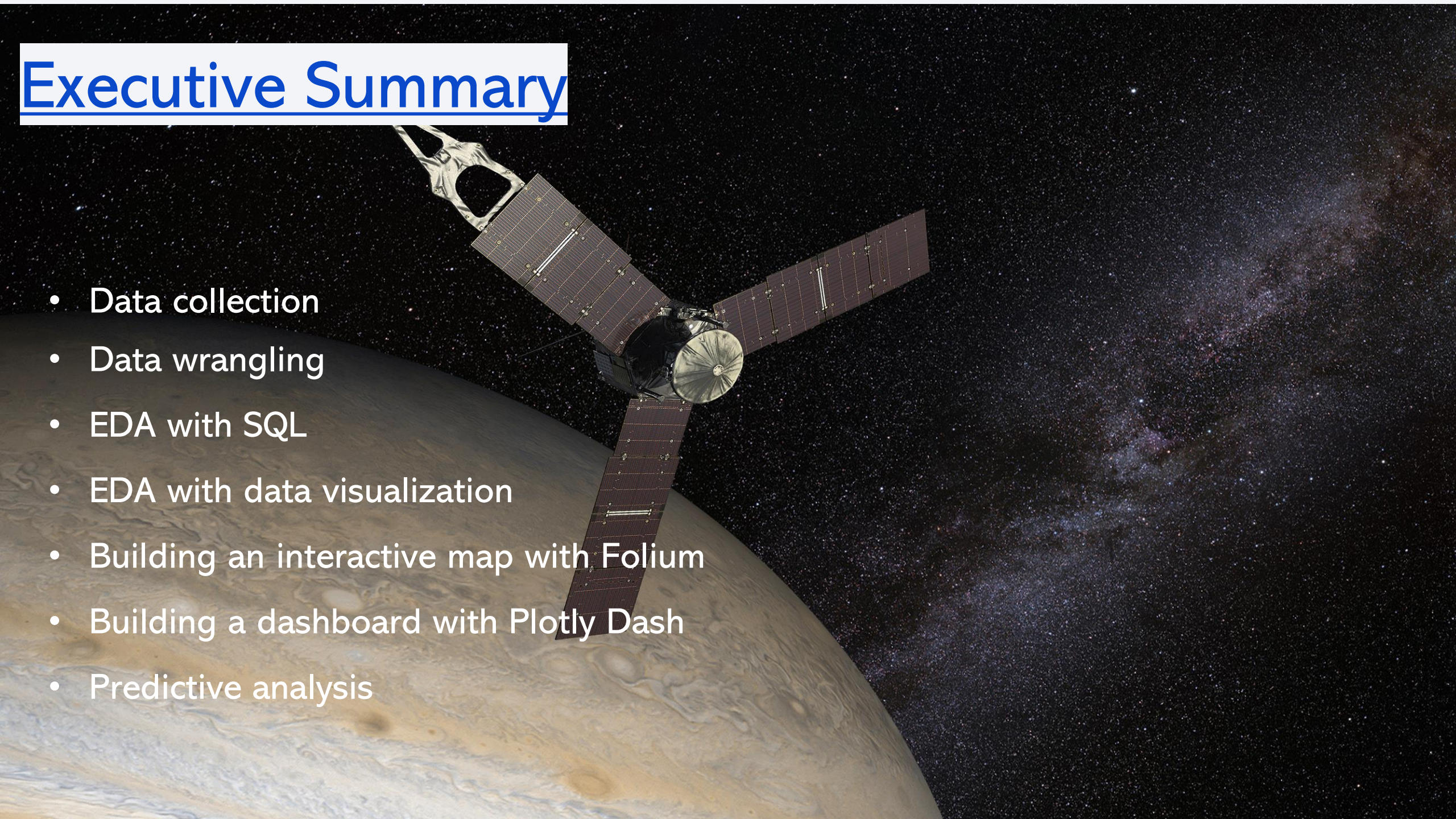
30-08-2023

GitHub :
rhashemi63/Capstone_Falcon9_IBM

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary



- Data collection
- Data wrangling
- EDA with SQL
- EDA with data visualization
- Building an interactive map with Folium
- Building a dashboard with Plotly Dash
- Predictive analysis

Introduction

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

- Problems you want to find answers
  - We will predict if the Falcon 9 first stage will land successfully. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

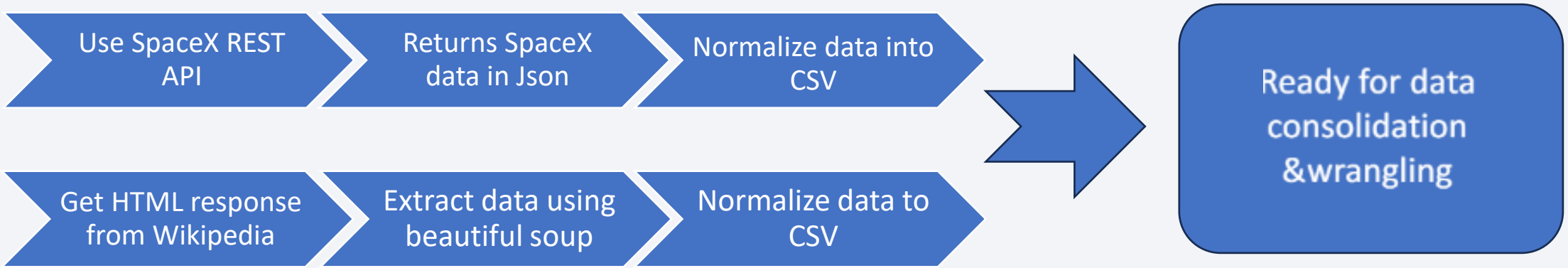
- Data collection methodology:
  - ✓ SpaceX REST API
  - ✓ Web Scrapping from Wikipedia
- Perform data wrangling
  - ✓ **Handling Missing values with a different strategy**
  - ✓ **Using One Hot Encoding method for the data field**
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - ✓Logistic Regression, K Nearest Neighbor, Support Vector Machine, Decision Tree

Data Collection

- Describe how data sets were collected:

✓ Gathering SpaceX launch data using SpaceX REST API

✓ Obtaining data via Web scrapping from Wikipedia using Beautiful soup



Data Collection – SpaceX API

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
import pandas as pd
import requests

# Decode the response content as JSON
json_data = response.json()

# Convert JSON data to a pandas DataFrame using json_normalize
data= pd.json_normalize(json_data)

# Display the resulting DataFrame
data
```

Now let's start requesting rocket launch data from SpaceX API with the following URL:

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

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

Data Collection - Scraping

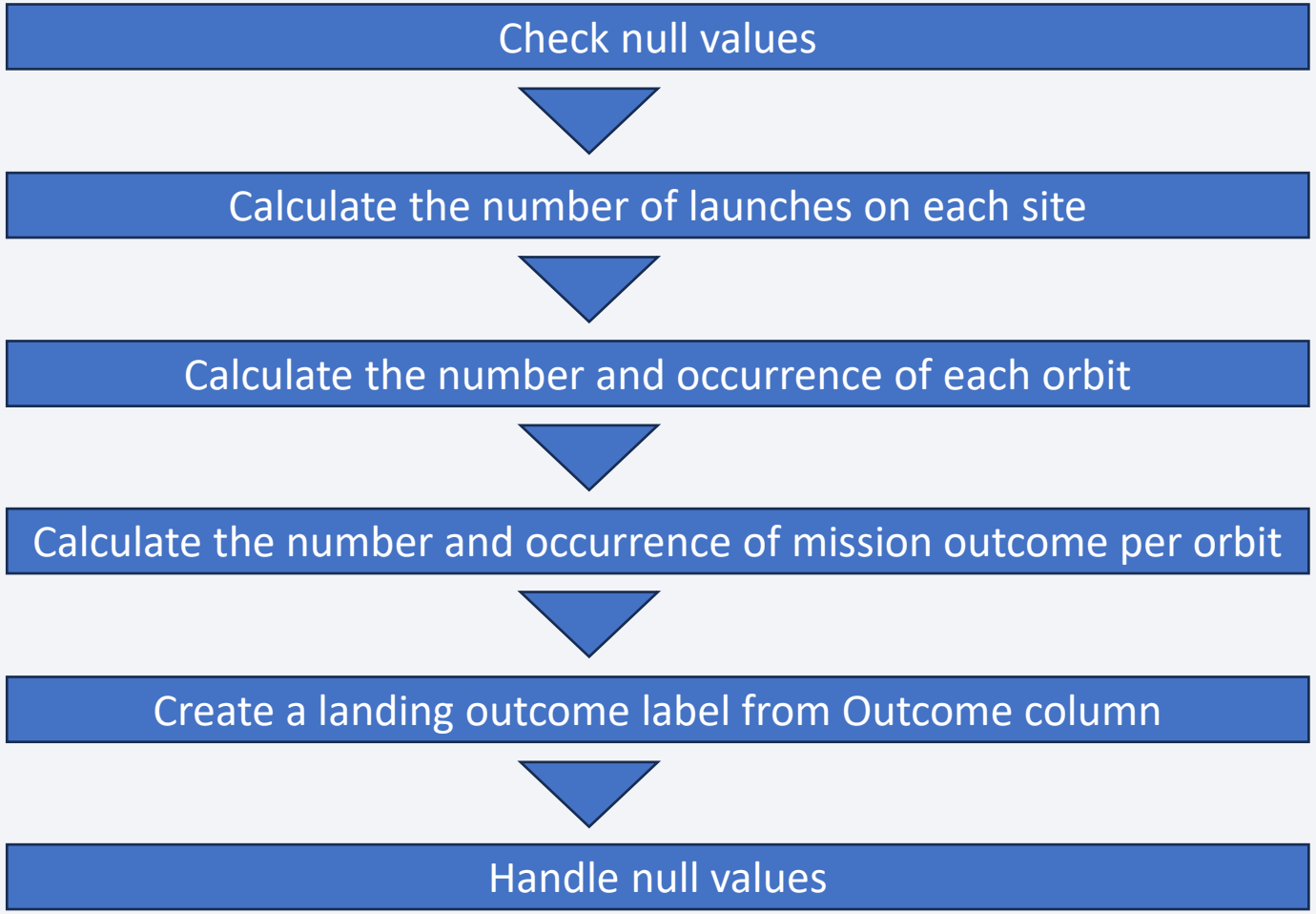
Data Collection and Web Scraping have been done based on following three tasks:

TASK 1: Request the Falcon9 Launch Wiki page from its URL

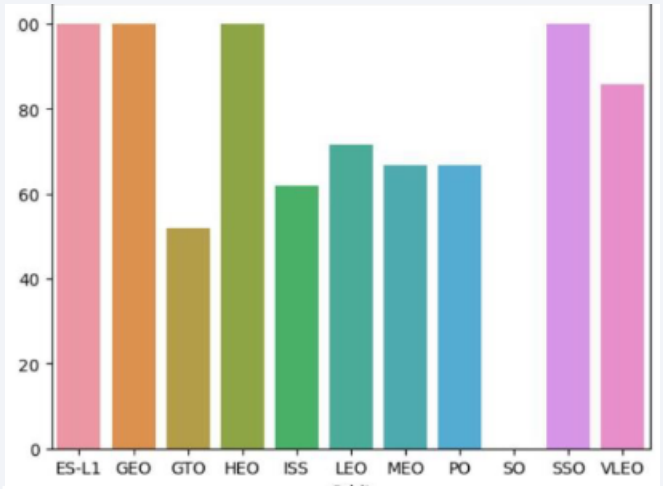
TASK 2: Extract all column/variable names from the HTML table header

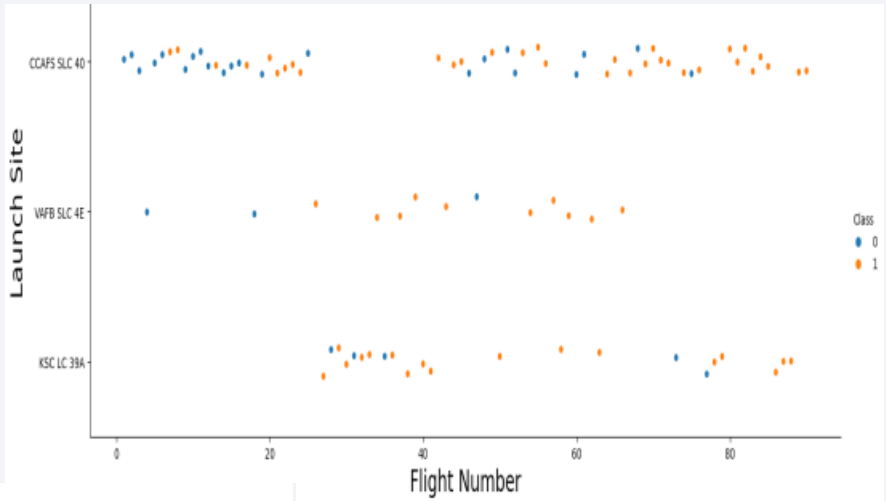
TASK 3: Create a data frame by parsing the launch HTML tables

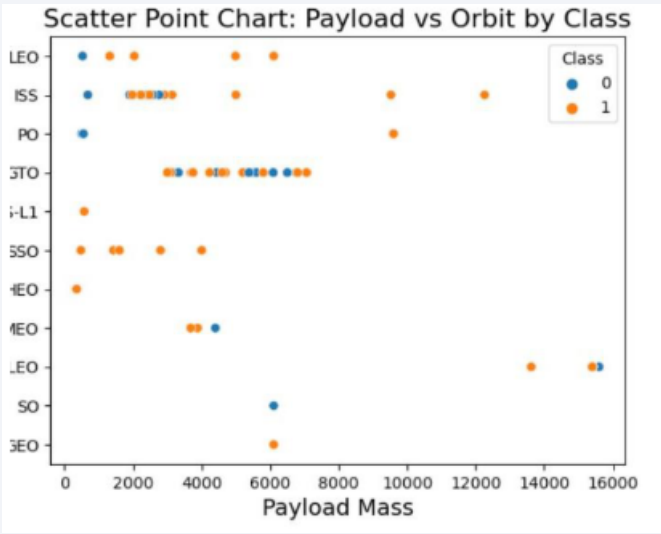
Data Wrangling

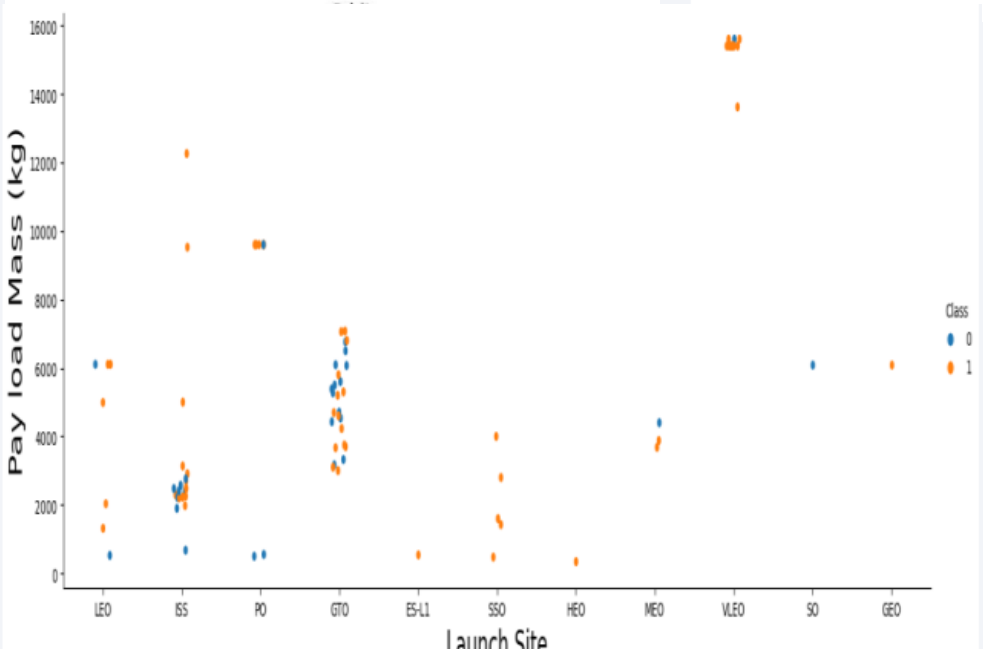


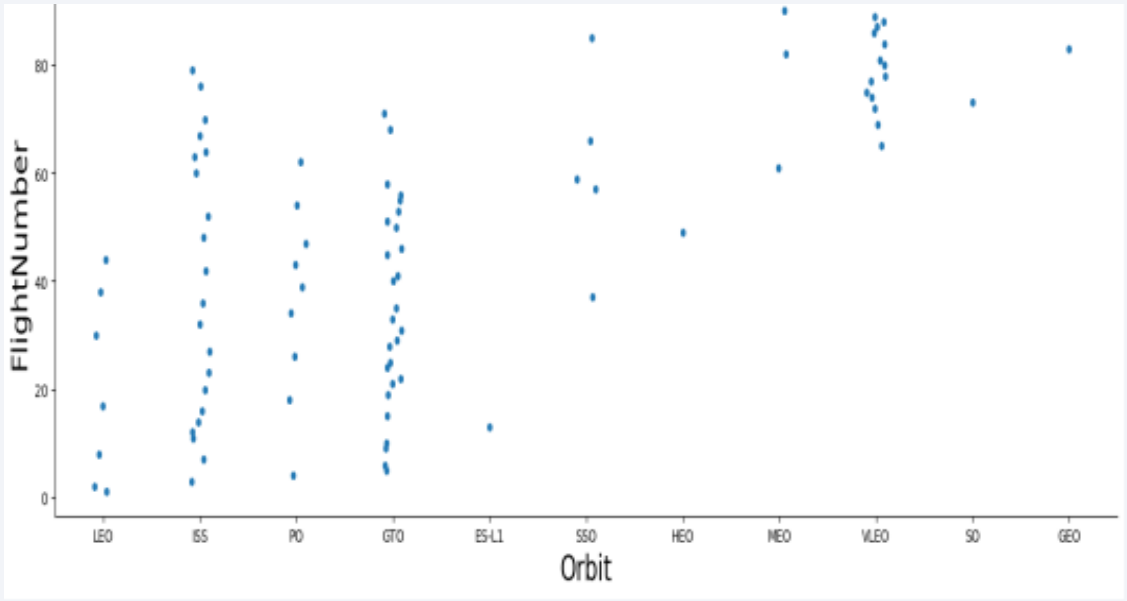
EDA with Data Visualization







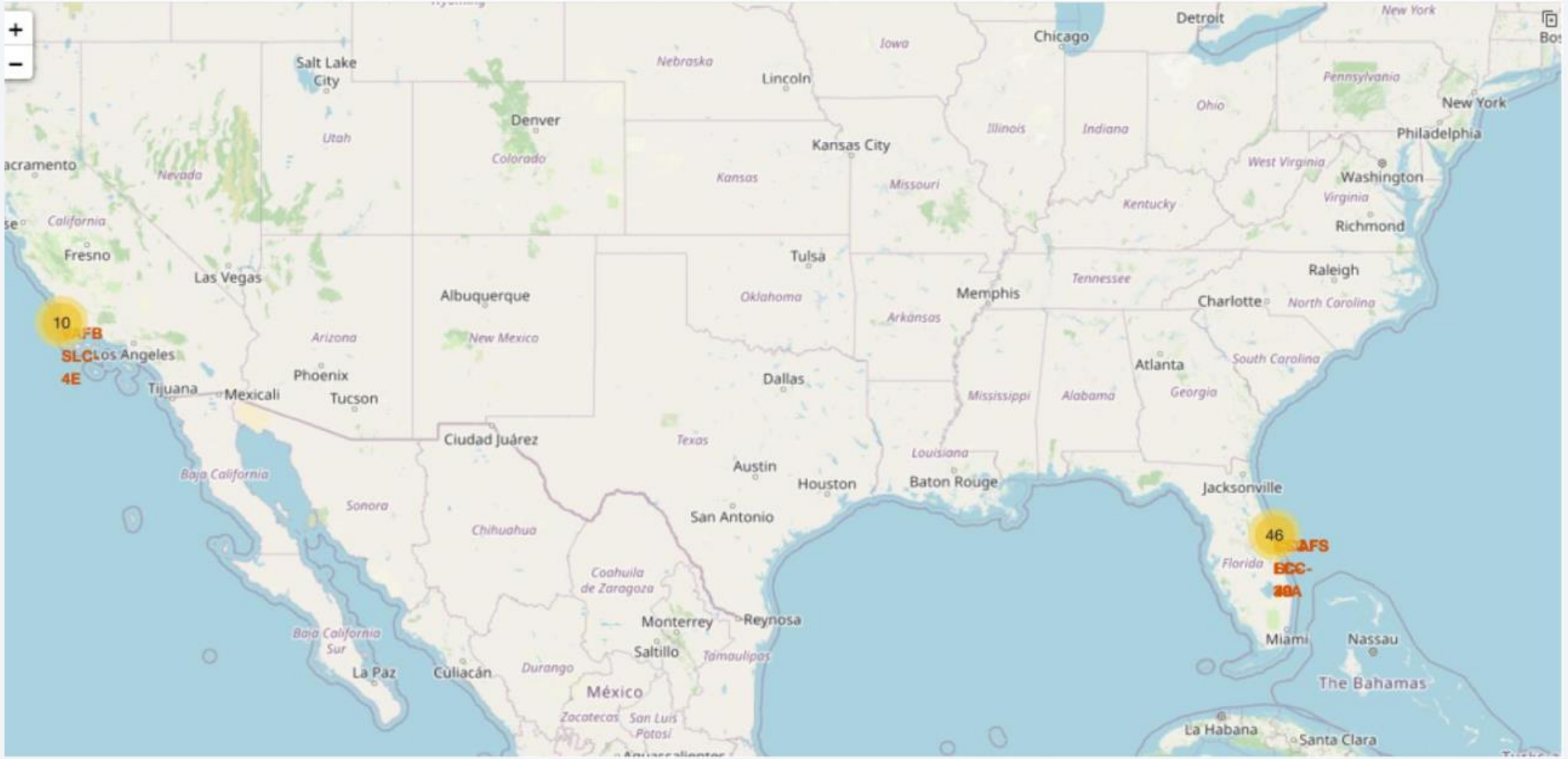




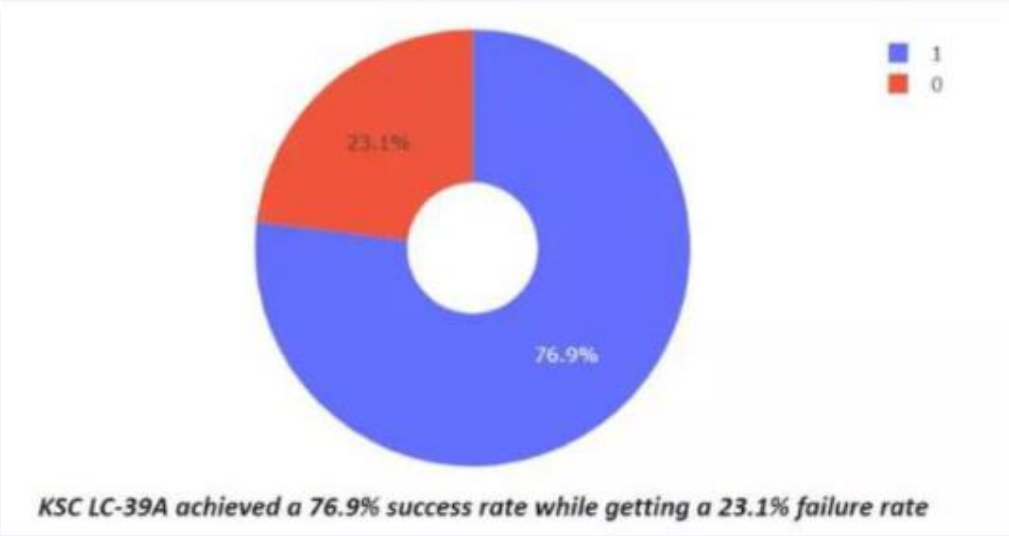
EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA’
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in the ground pad was achieved
- Listing the names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failed mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the records that will display the month names, failure landing outcomes in drone ship, booster versions, and launch site for the months in the year 2015
- Ranking the count of landing outcomes between the dates 2010-06-04 and 2017-03-20 in descending order

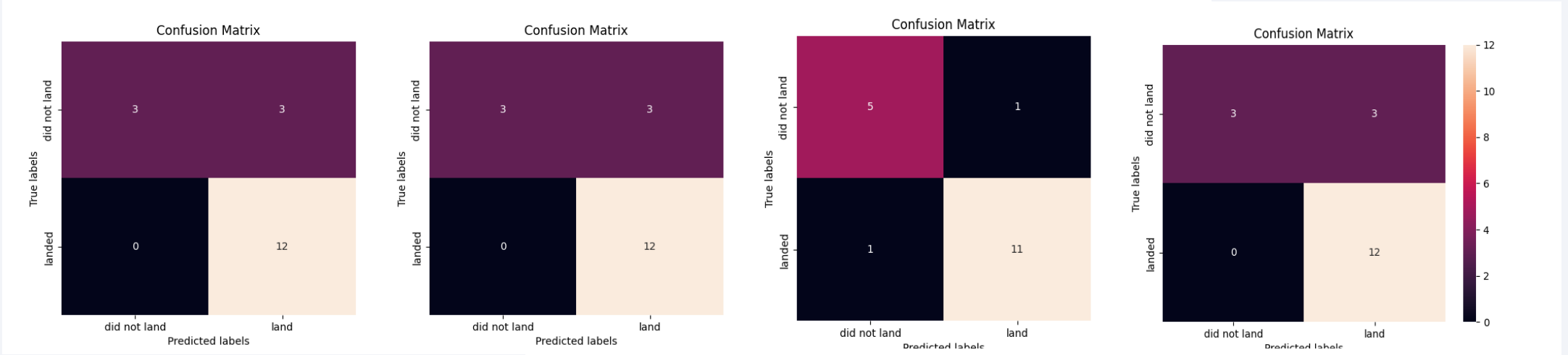
Build an Interactive Map with Folium



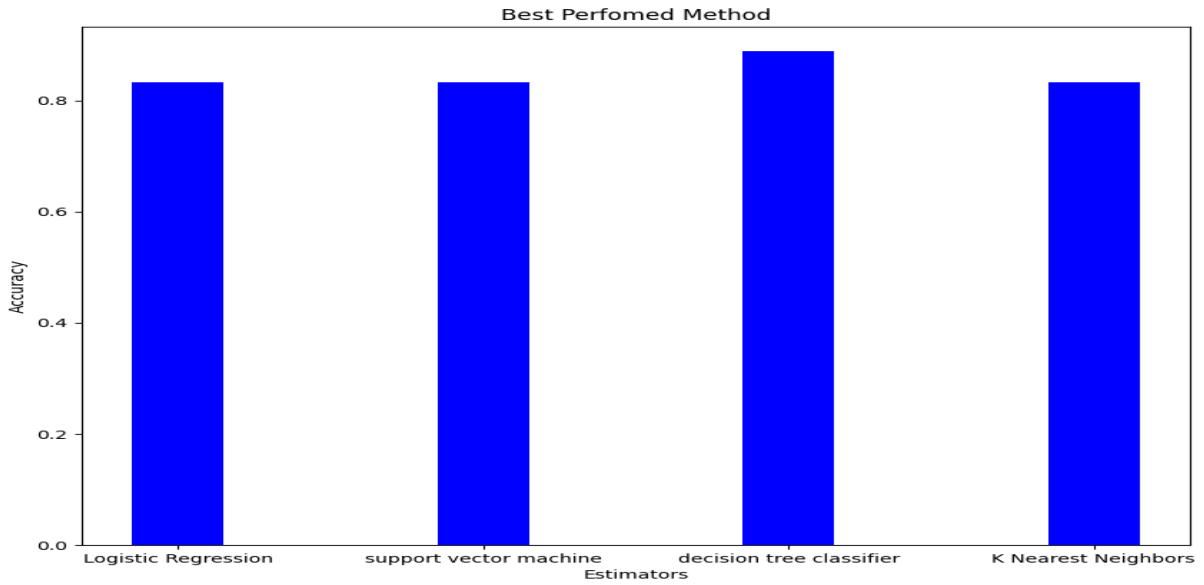
Build a Dashboard with Plotly Dash



Predictive Analysis (Classification)



- ✓ The SVM, KNN, and Logistic Regression models achieved the highest accuracy at 83.3 %, while the SVM performs the best in terms of Area Under the Curve.



Results

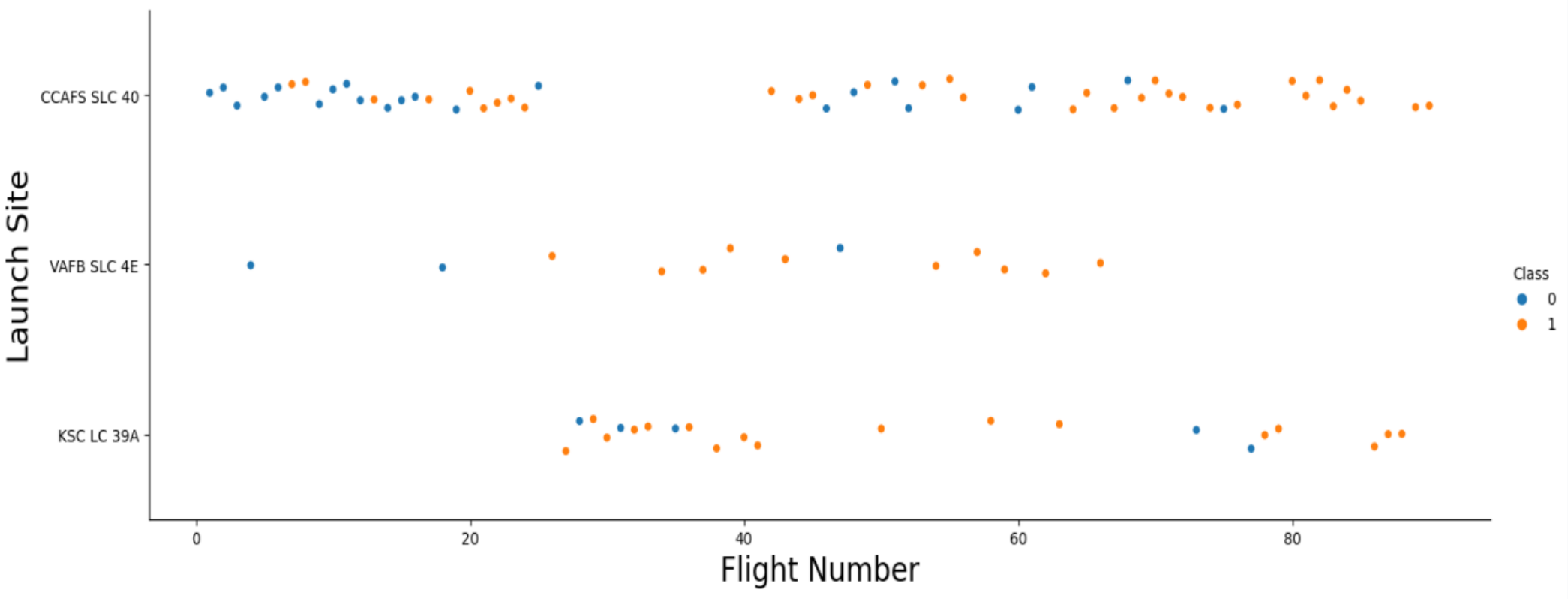
Factors associated with the success rate of the Falcon 9 rocket launches include the launch sites, payload mass, orbit type, and year of launch. The KSC LC-39A launch site, massive payload mass, ES-L1, GEO, HEO, and SSO orbit types are more likely to have a high success rate. These findings will help Space X reuse the first stage of Falcon 9 and save costs.

Section 2

Insights drawn from EDA

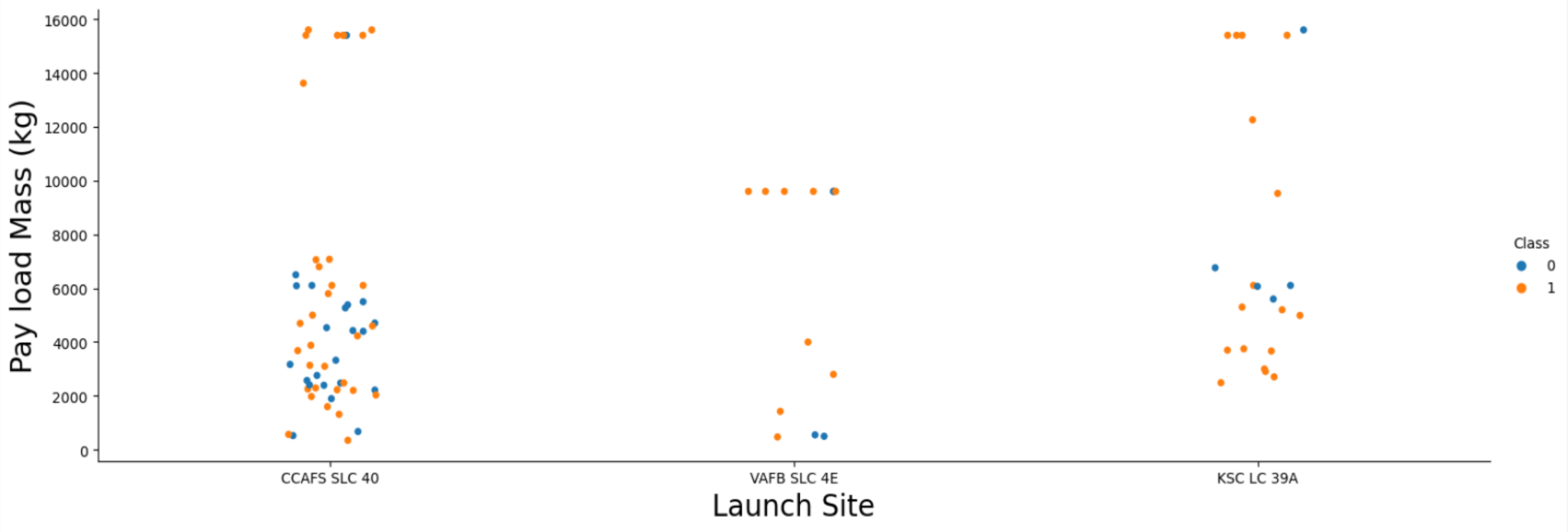
Flight Number vs. Launch Site

- A scatter plot of Flight Number vs. Launch Site



Payload vs. Launch Site

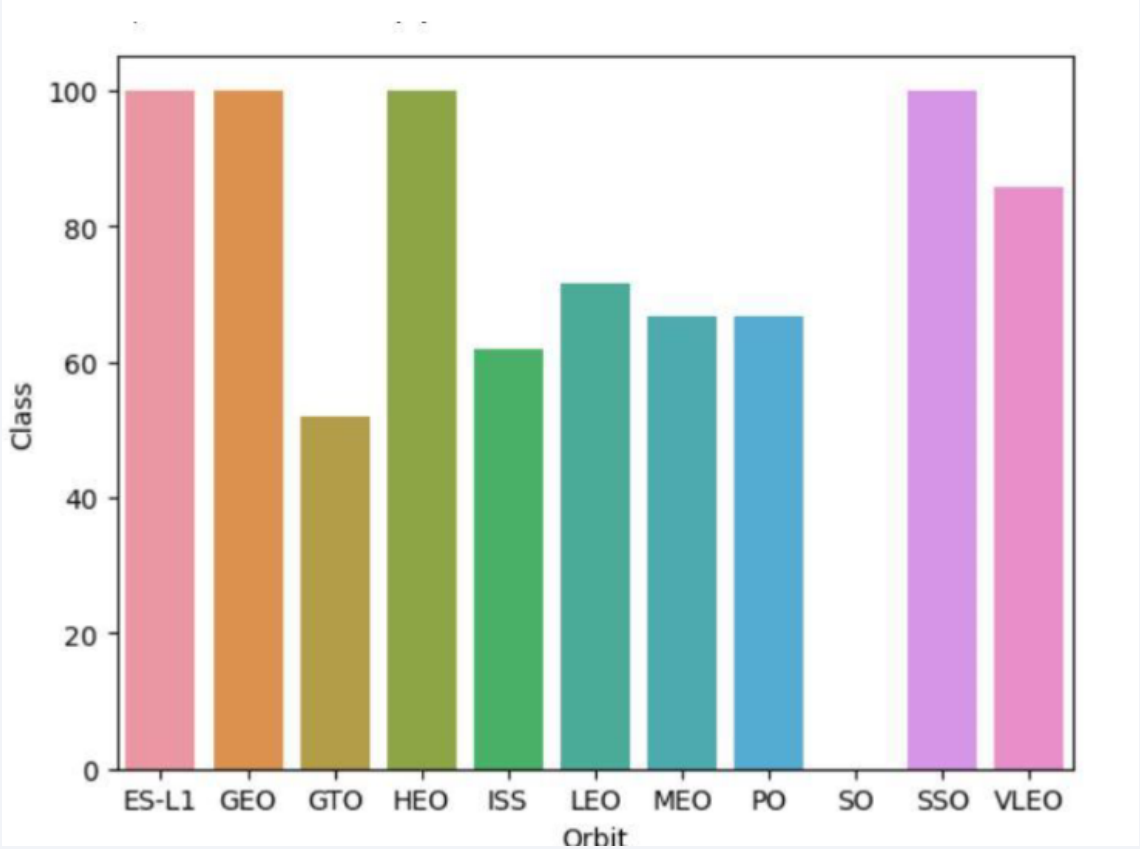
- A scatter plot of Payload vs. Launch Site



Success Rate vs. Orbit Type

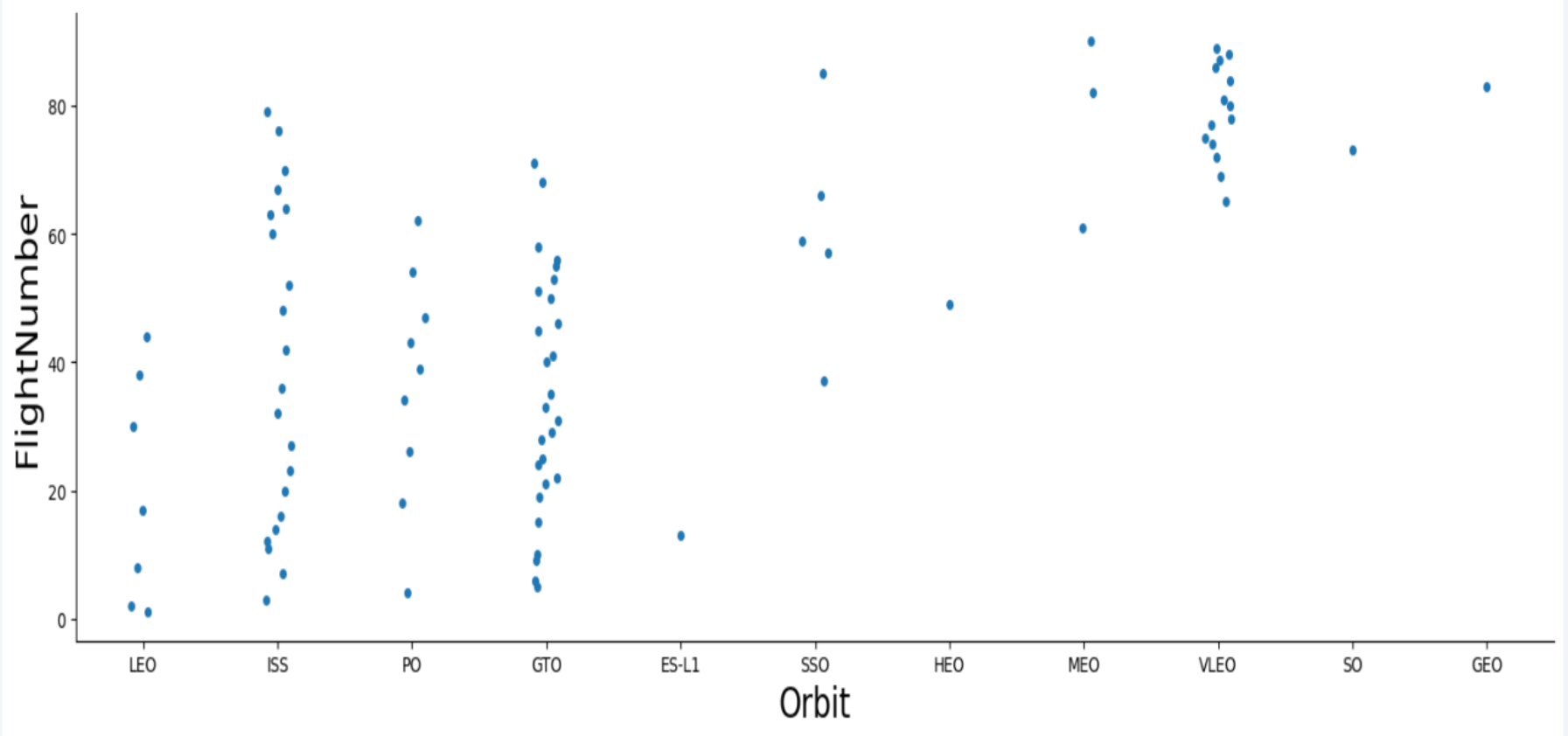
- A bar chart for the success rate of each orbit type

  - The orbit types of ES-L1, GEO, HEO, and SSO are among the highest success rates.



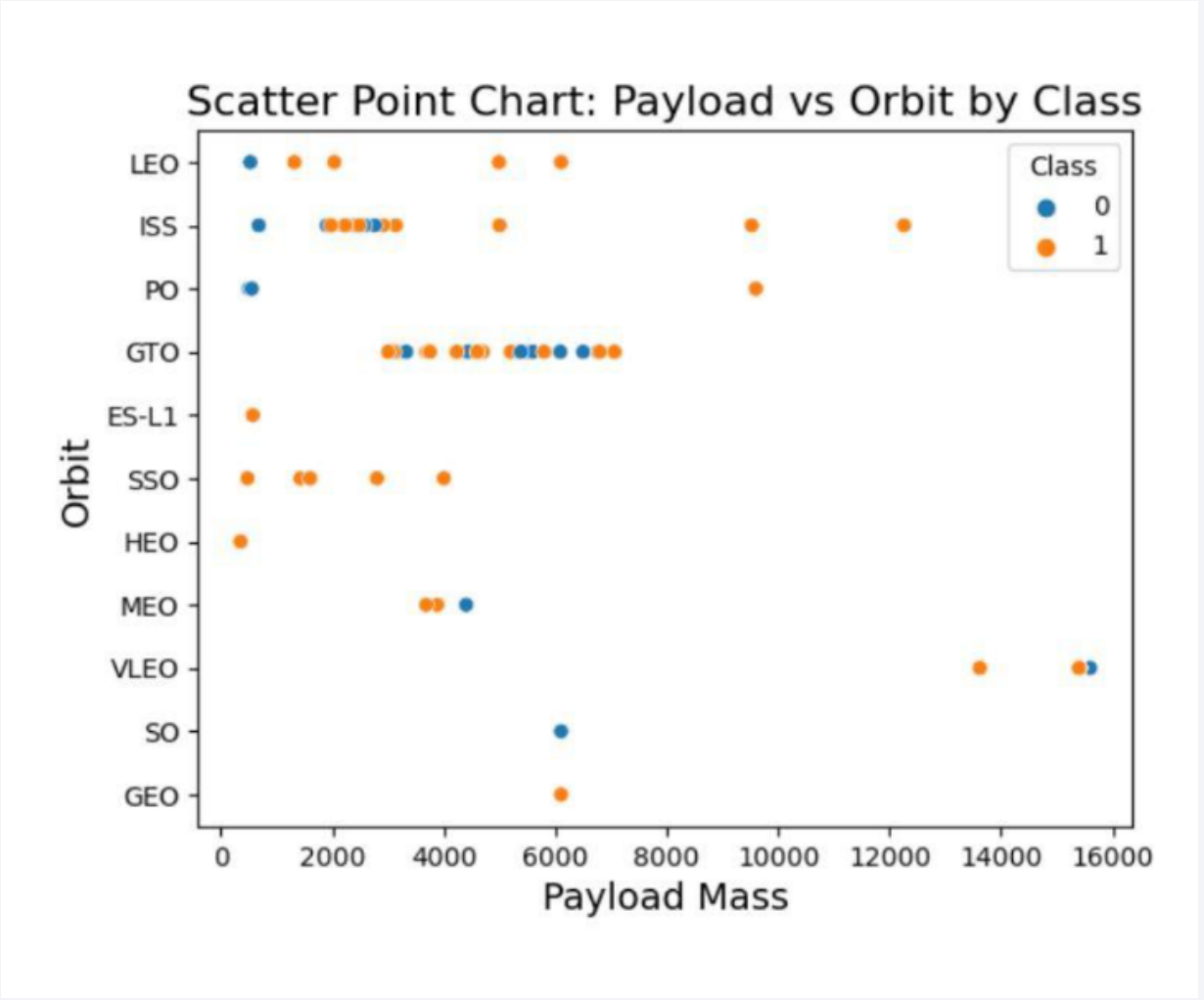
Flight Number vs. Orbit Type

- A scatter point of Flight number vs. Orbit type



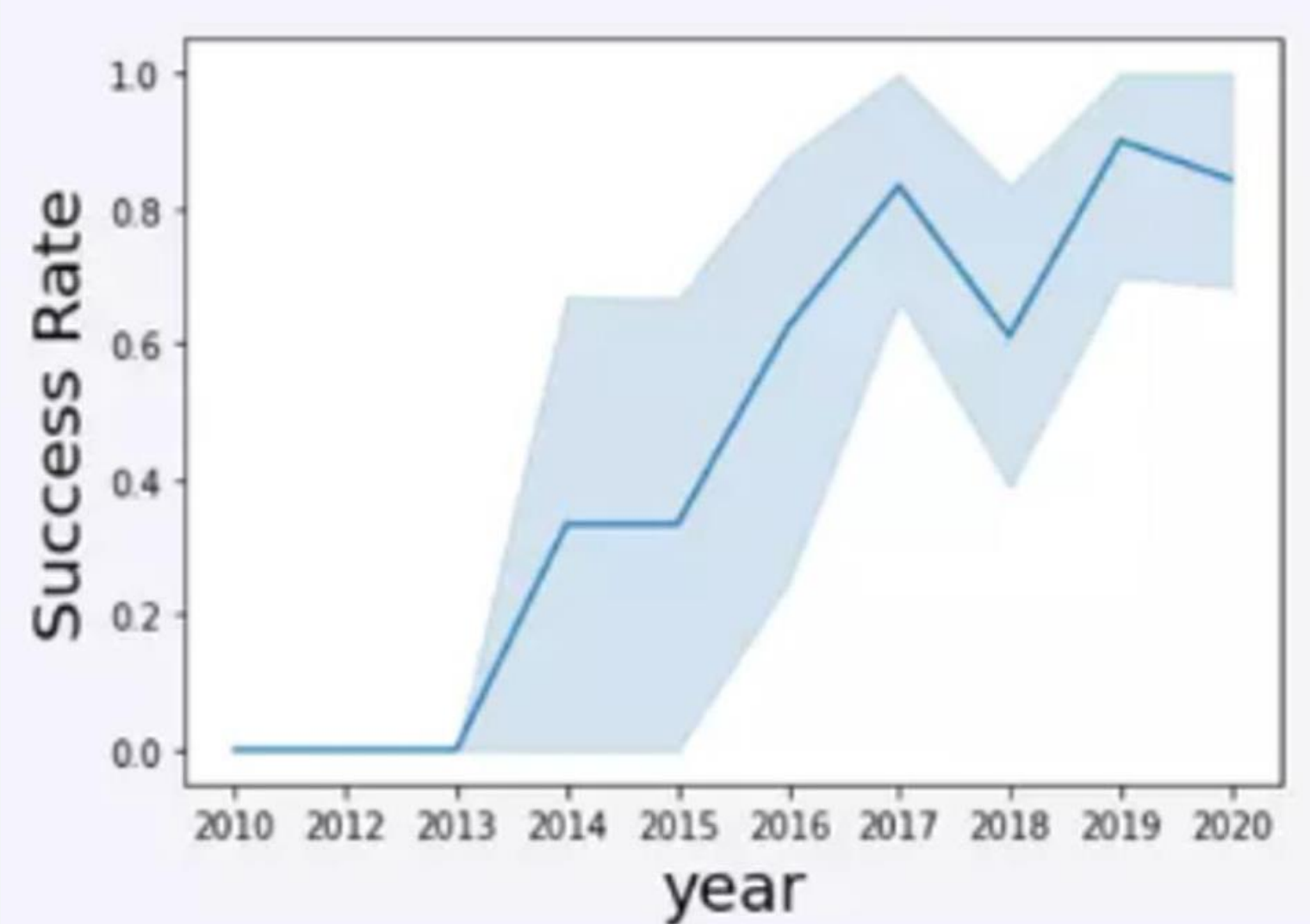
Payload vs. Orbit Type

- A scatter plot of payload vs. orbit type



Launch Success Yearly Trend

- A line chart of the yearly average success rate



All Launch Site Names

- The names of the unique launch sites:
  - ✓ CCAFS LC-40
  - ✓ VAFB SLC-4E
  - ✓ KSC LC-39A
  - ✓ CCAFS SLC-40

- Query:

```
%sql select DISTINCT "Launch_Site" from SPACEXTABLE;
```

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- `%sql SELECT *  from SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;`

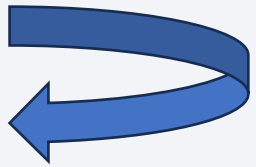
DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- `%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" == 'NASA (CRS)';`

- Output:

SUM(PAYLOAD_MASS__KG_)



  - 45596

Average Payload Mass by F9 v1.1

- `%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster Version" = 'F9 v1.1';`

Output:

✓ 2928.4

First Successful Ground Landing Date

- % sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Mission Outcome" = 'Success' AND "Landing Outcome" = 'Success (ground pad)';

✓ 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT "Booster Version" FROM SPACEXTABLE WHERE  "Landing Outcome" =
'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND
"PAYLOAD_MASS__KG_" < 6000;
```

- Booster Version

✓F9 FT B1022

✓F9 FT B1026

✓F9 FT B1021.2

✓F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- `%sql SELECT "Mission Outcome" , COUNT(*)  FROM SPACEXTABLE GROUP BY "Mission Outcome";`

Mission_Outcome	COUNT(*)
✓Failure (in flight)	1
✓Success	98
✓Success	1
✓Success (payload status unclear)	1

Boosters Carried Maximum Payload

- %sql SELECT "Booster Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);

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

%sql select * from SPACEXTBL where Landin Outcome like 'Success%' and (Date between '2015-01-01' and '2015-12-31') order by date

time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)

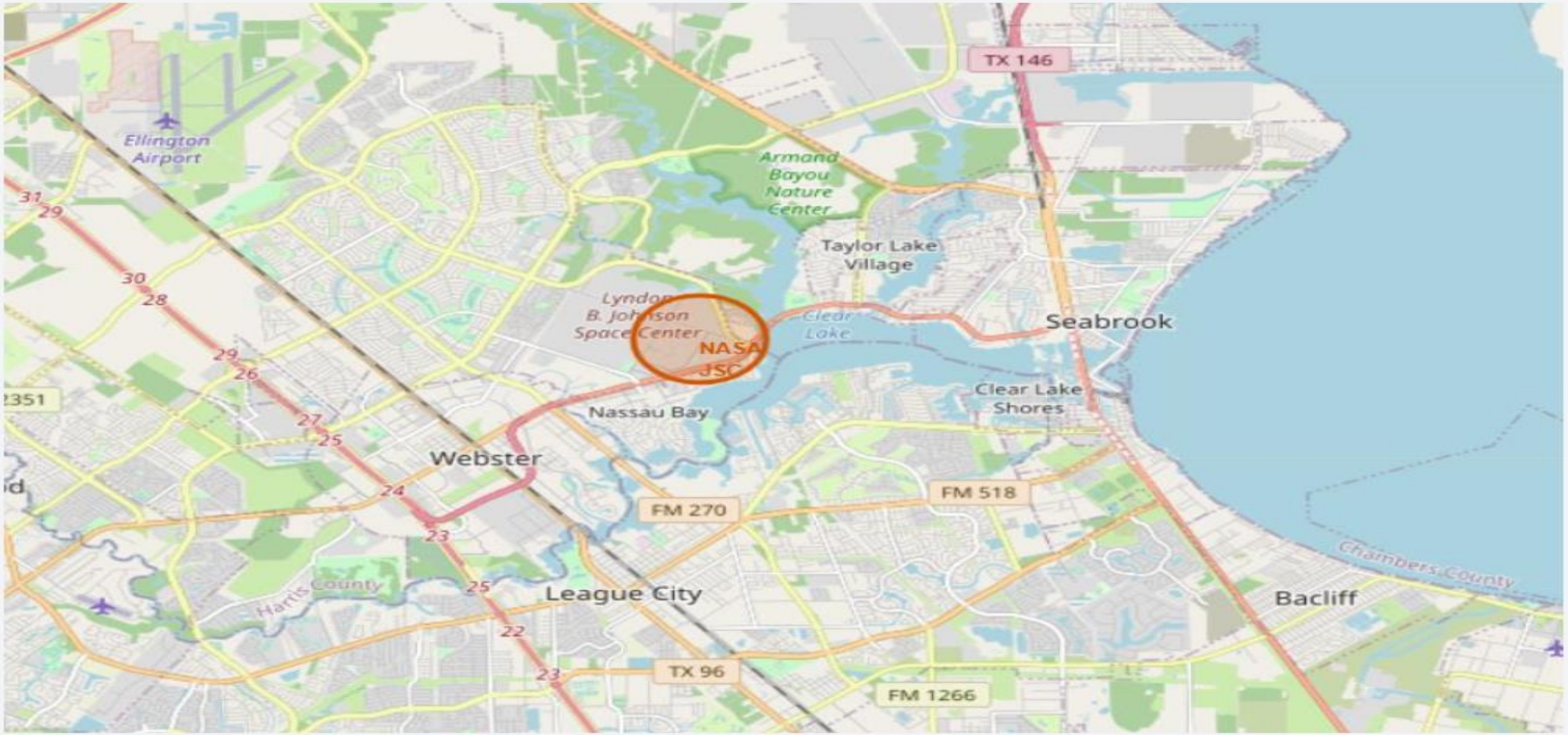
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- %sql select count(*) from SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' or Landing_Outcome = 'Success (ground pad)' and (DATE between '2010-06-04' and '2017-03-20') order by date desc;

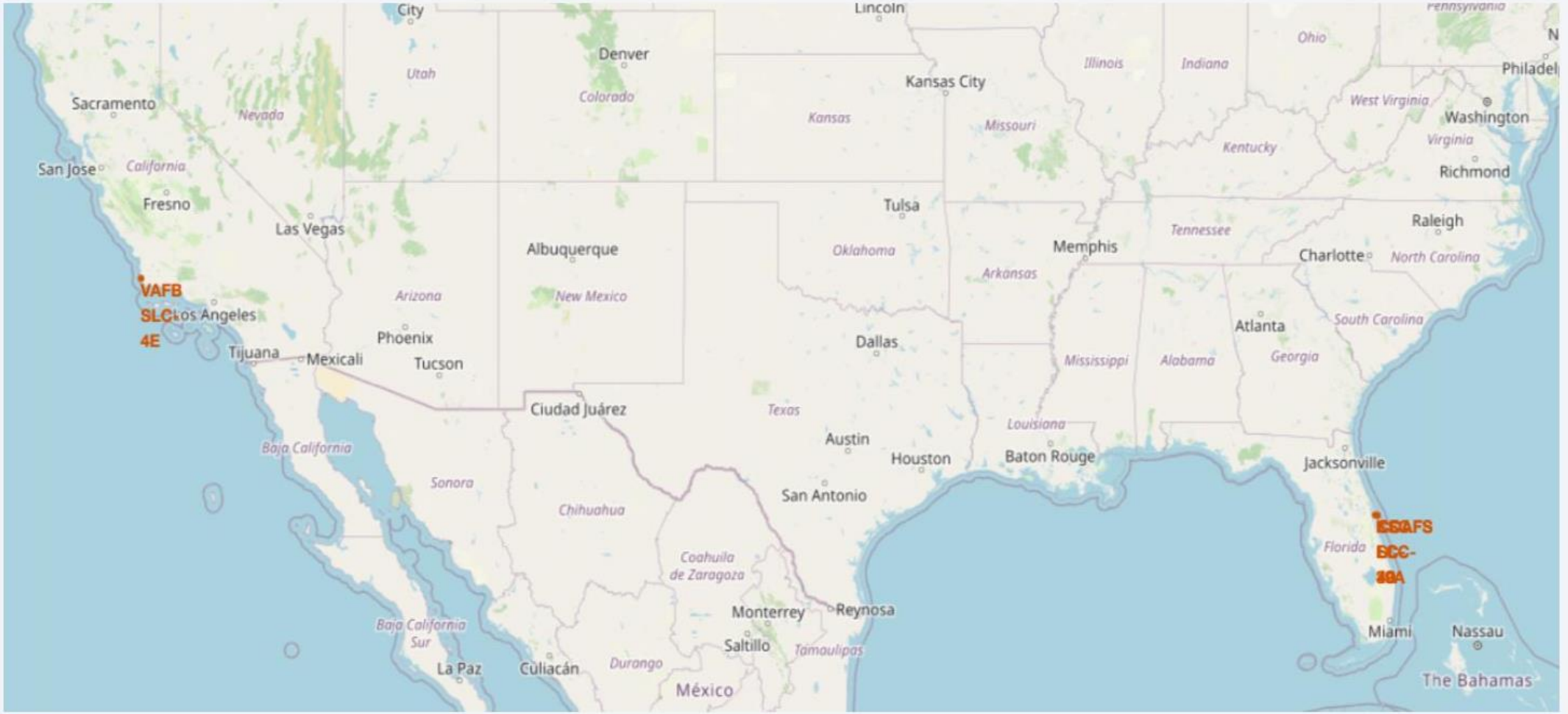
✓10


Section 3
Launch Sites
Proximities Analysis

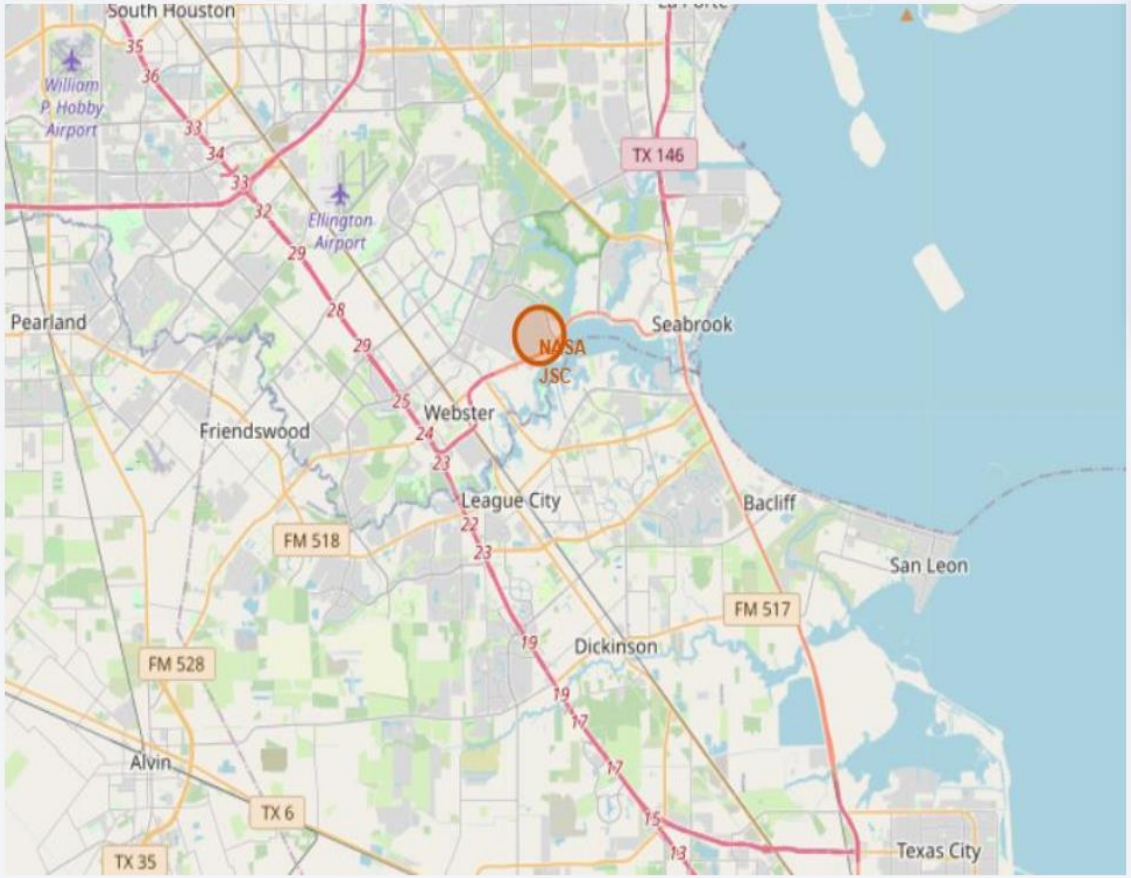
<Folium Map Screenshot 1>

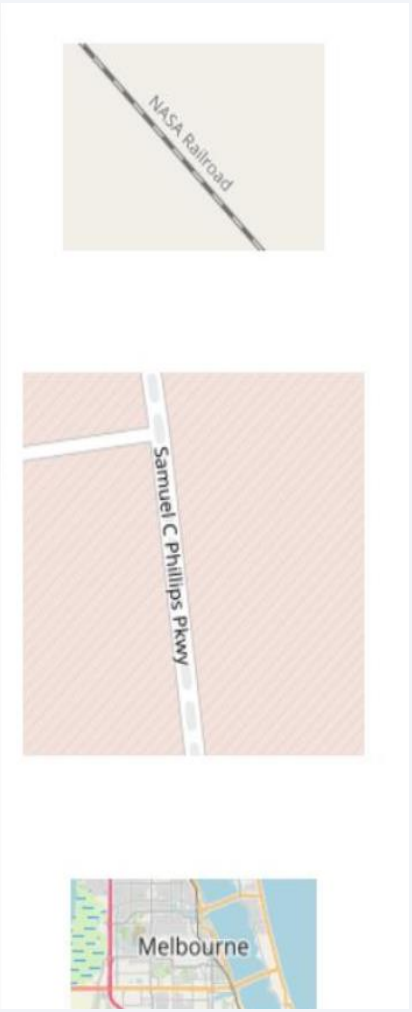


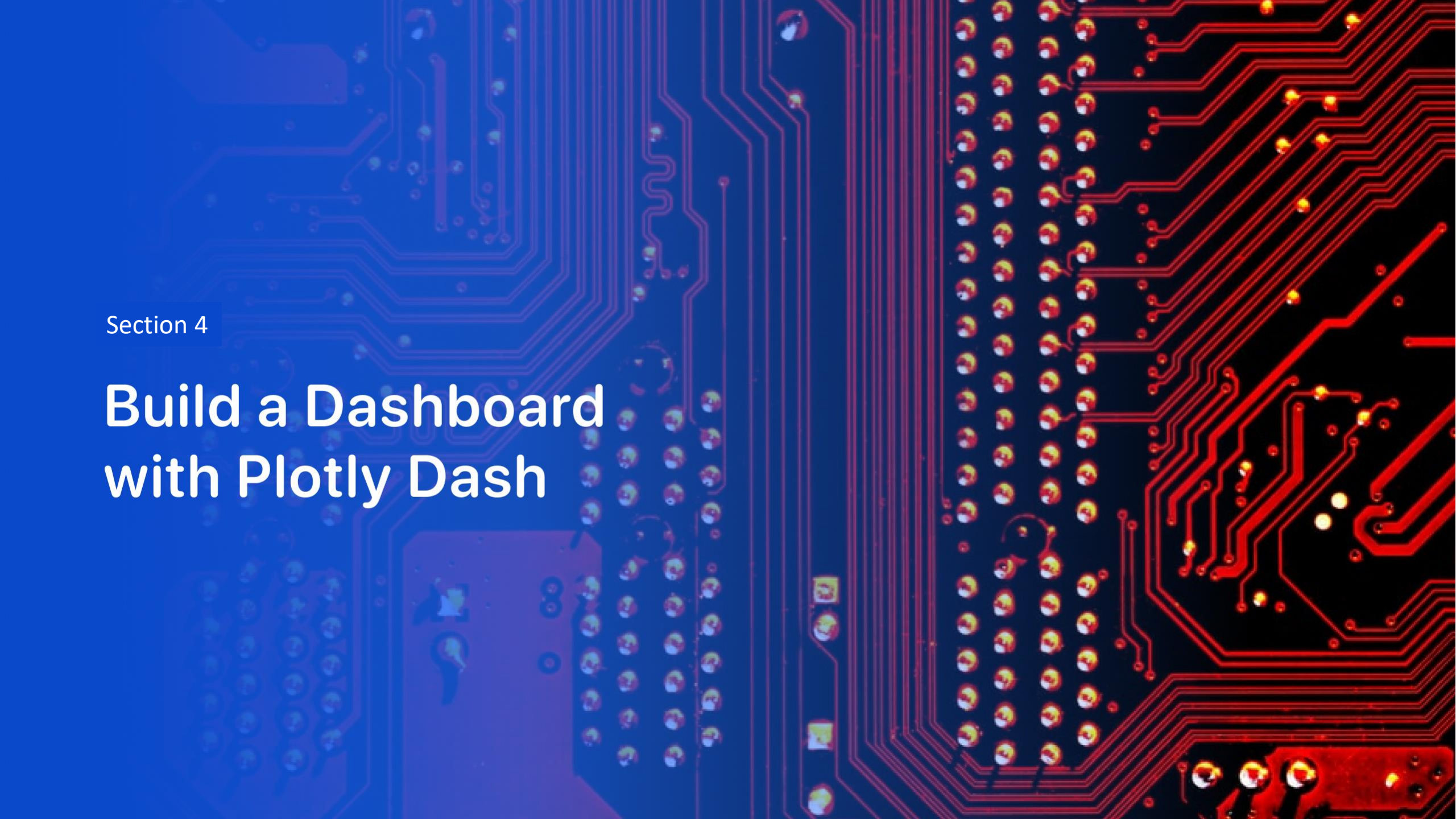
<Folium Map Screenshot 2>



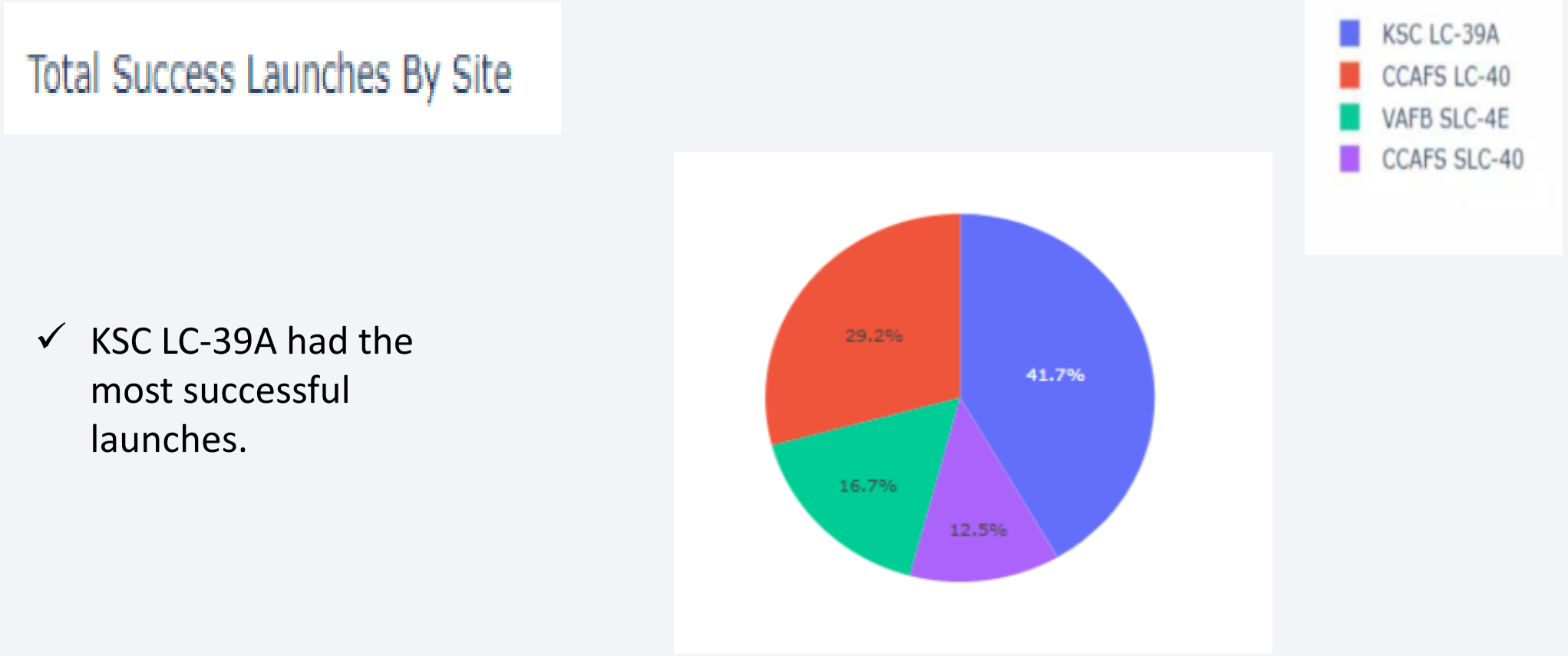
<Folium Map Screenshot 3>






Section 4
Build a Dashboard with Plotly Dash

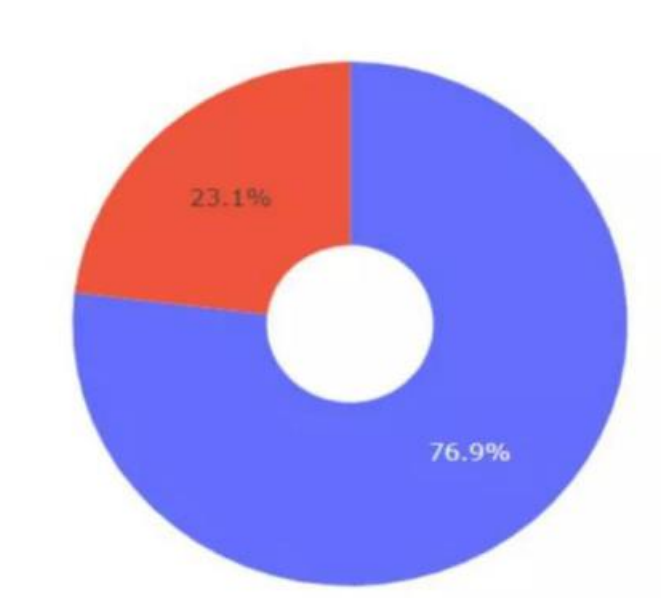
Dashboard Screenshot 1



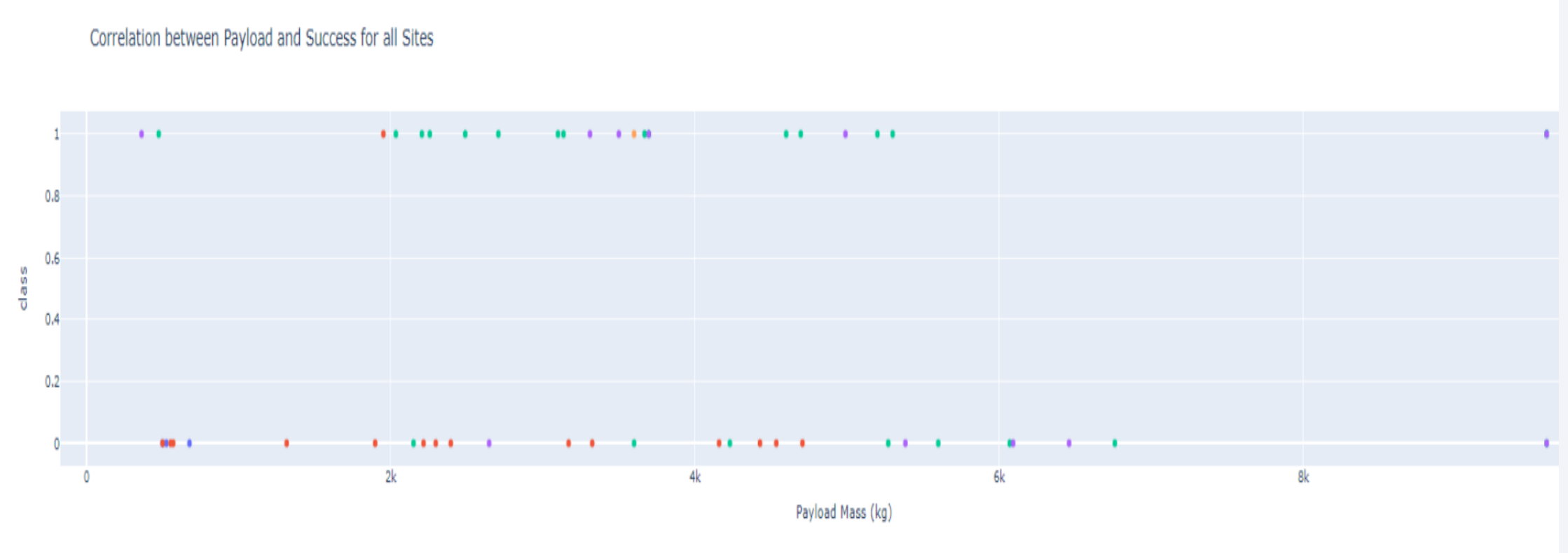
- ✓ KSC LC-39A had the most successful launches.

Dashboard Screenshot 2

KSL LC-39A recorded a success rate of 76.9% and a failure rate of 23.1%.



Dashboard Screenshot 3

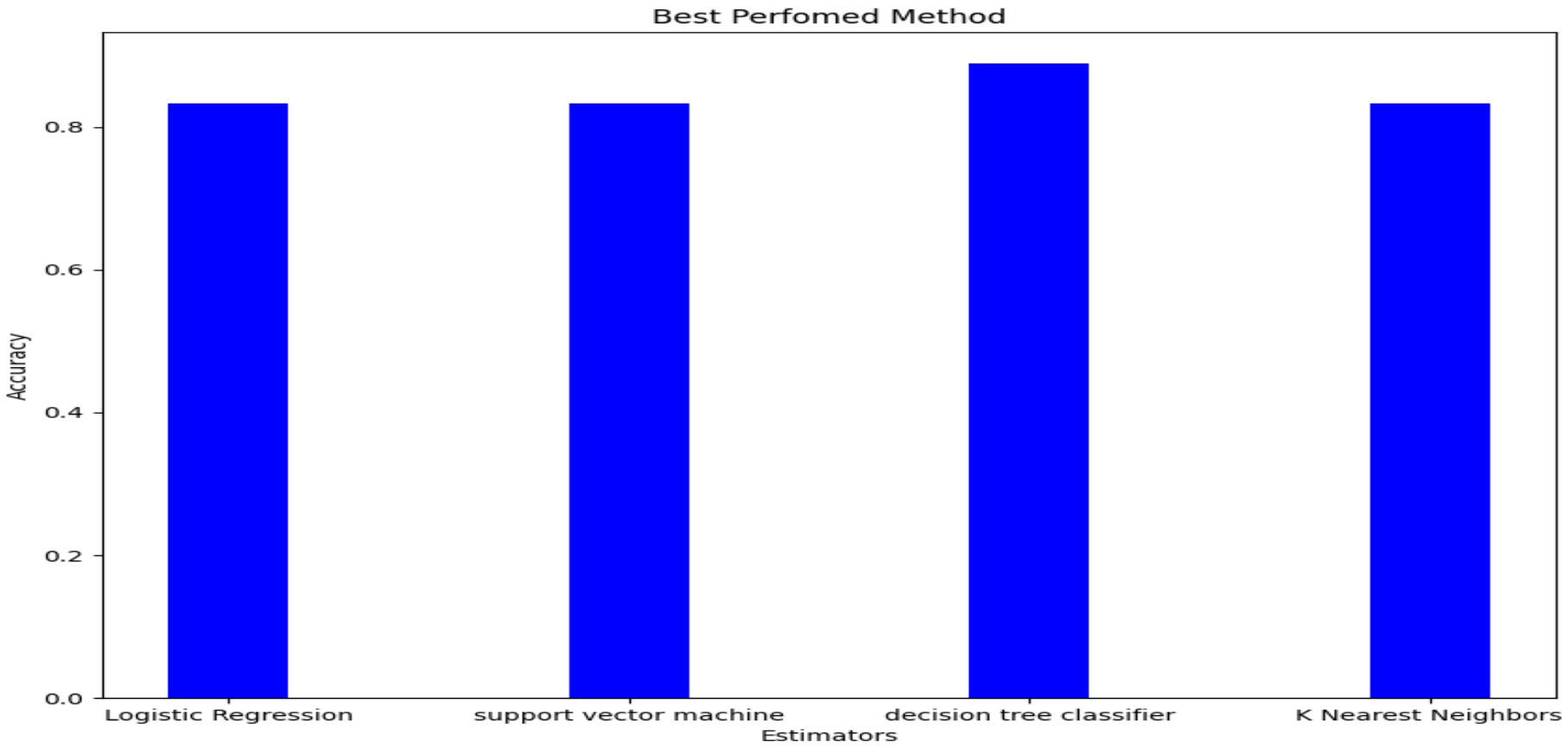


- ✓ The success rate for light-weight payloads is higher than that for heavy-weight payloads.

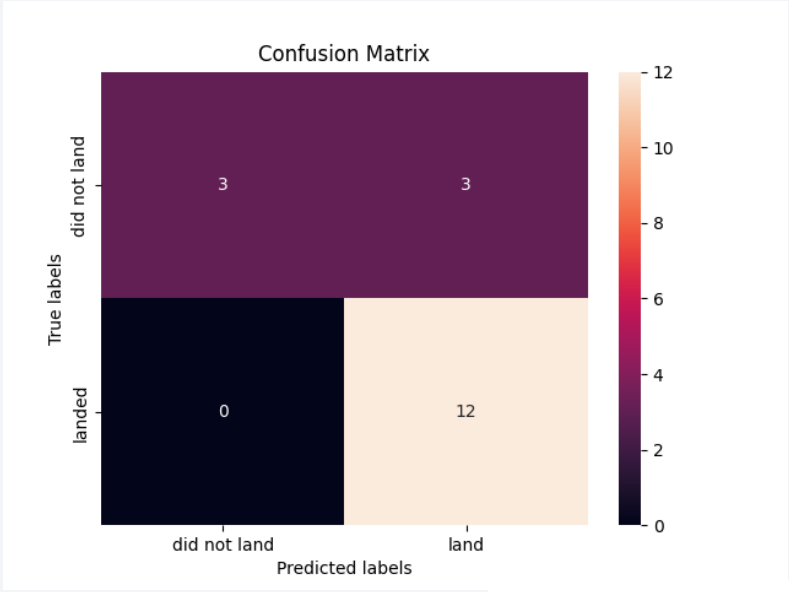
Section 5

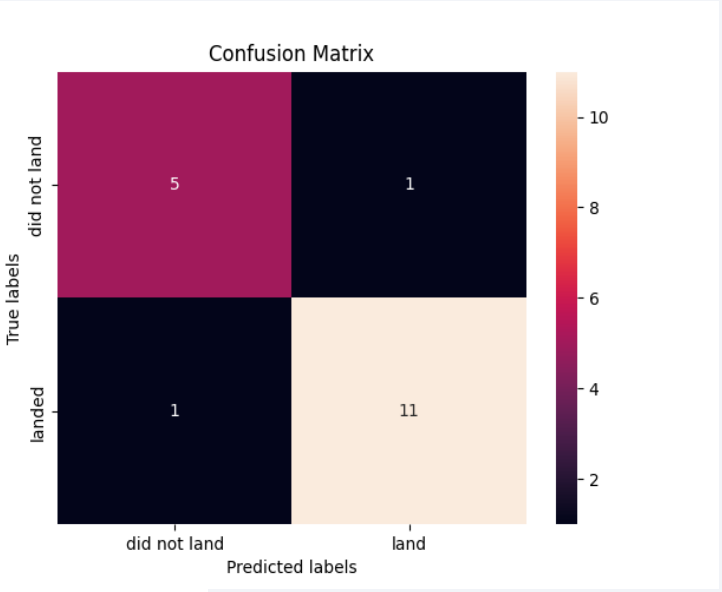
Predictive Analysis (Classification)

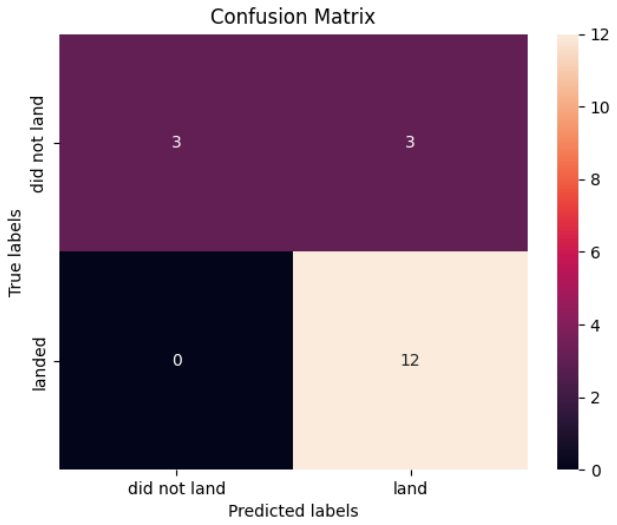
Classification Accuracy

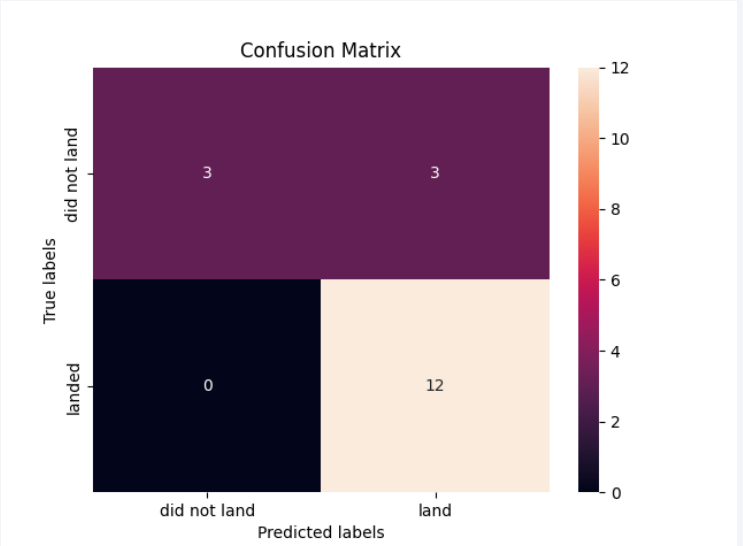


Confusion Matrix









Thank you!