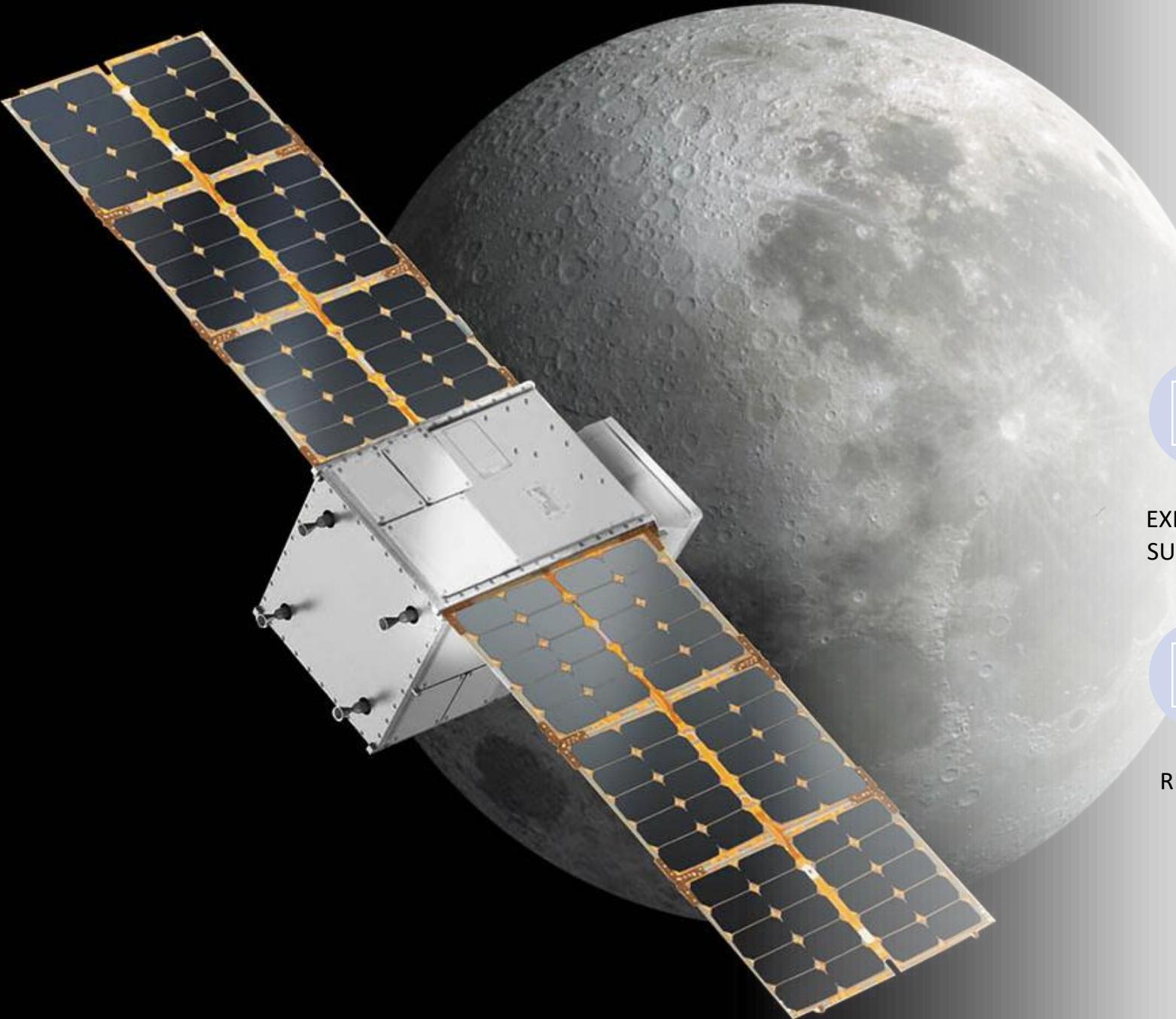


Winning Space Race with Data Science

Shokoufeh
Mohammadi
...---2023
GitHub : shokoufeh1992/capstone





Outline



EXECUTIVE
SUMMARY



INTRODUCTION



METHODOLOGY



RESULTS



CONCLUSION



APPENDIX

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- EDA results
- Interactive analytics
- Predictive analysis

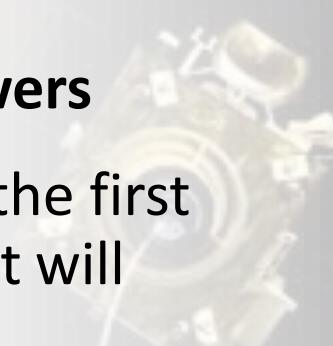
— Introduction

- 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

The project task is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully.



Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

SpaceX Rest API

Web Scrapping from 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 models have been built and evaluated for the best classifier

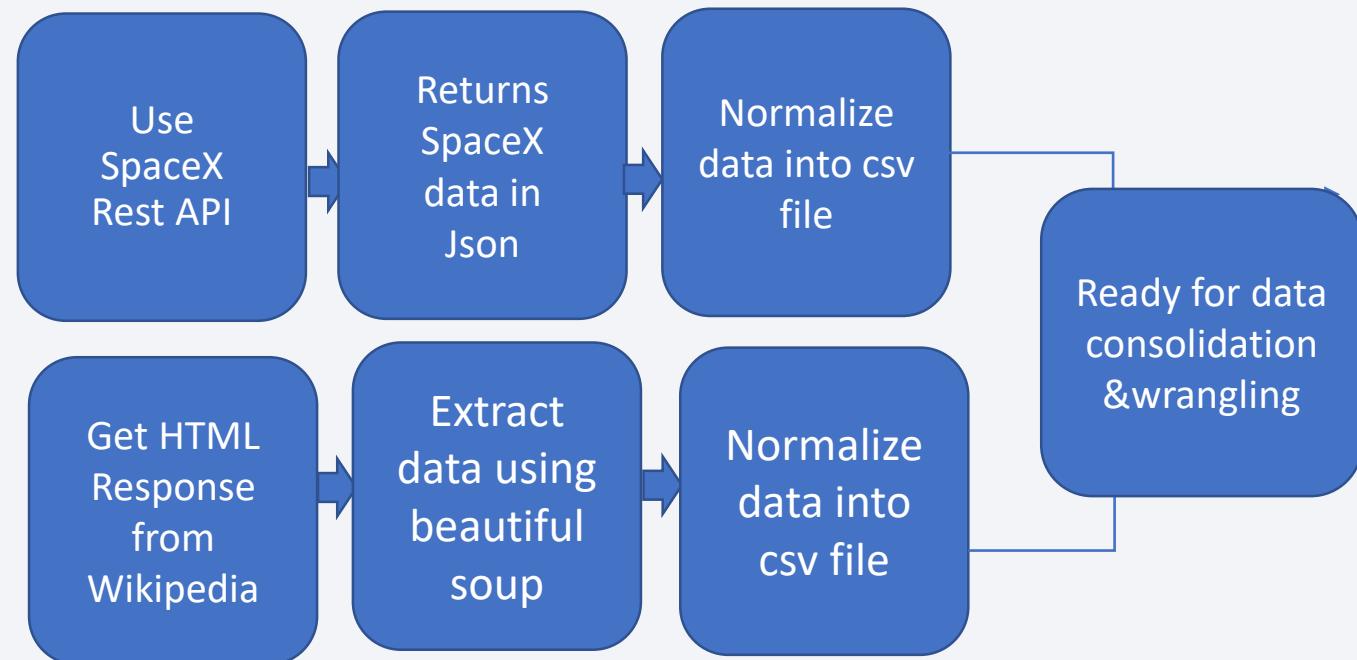
Data Collection

SpaceX launch data that is gathered from the SpaceX REST API.

This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.

Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using Beautiful Soup.



Data Collection – SpaceX API

Data collection with SpaceX REST calls

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

Web Scraping from Wikipedia

1 .Getting Response from HTML

```
page = requests.get(static_url)
```

2. Creating BeautifulSoup Object

```
soup = BeautifulSoup(page.text, 'html.parser')
```

3. Finding tables

```
html_tables = soup.find_all('table')
```

4. Getting column names

```
column_names = []
temp = soup.find_all('th')
for x in range(len(temp)):
    try:
        name = extract_column_from_header(temp[x])
        if (name is not None and len(name) > 0):
            column_names.append(name)
    except:
        pass
```

6. Appending data to keys (refer) to notebook block 12

```
In [12]: extracted_row = 0
#Extract each table
for table_number,table in enumerate(
    # get table row
    for rows in table.find_all("tr",
        #check to see if first table
        
```

5. Creation of dictionary

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelevant column
del launch_dict['Date and time ( )']

launch_dict['Flight No.']= []
launch_dict['Launch site']= []
launch_dict['Payload']= []
launch_dict['Payload mass']= []
launch_dict['Orbit']= []
launch_dict['Customer']= []
launch_dict['Launch outcome']= []
launch_dict['Version Booster']= []
launch_dict['Booster landing']= []
launch_dict['Date']= []
launch_dict['Time']= []
```

7. Converting dictionary to dataframe

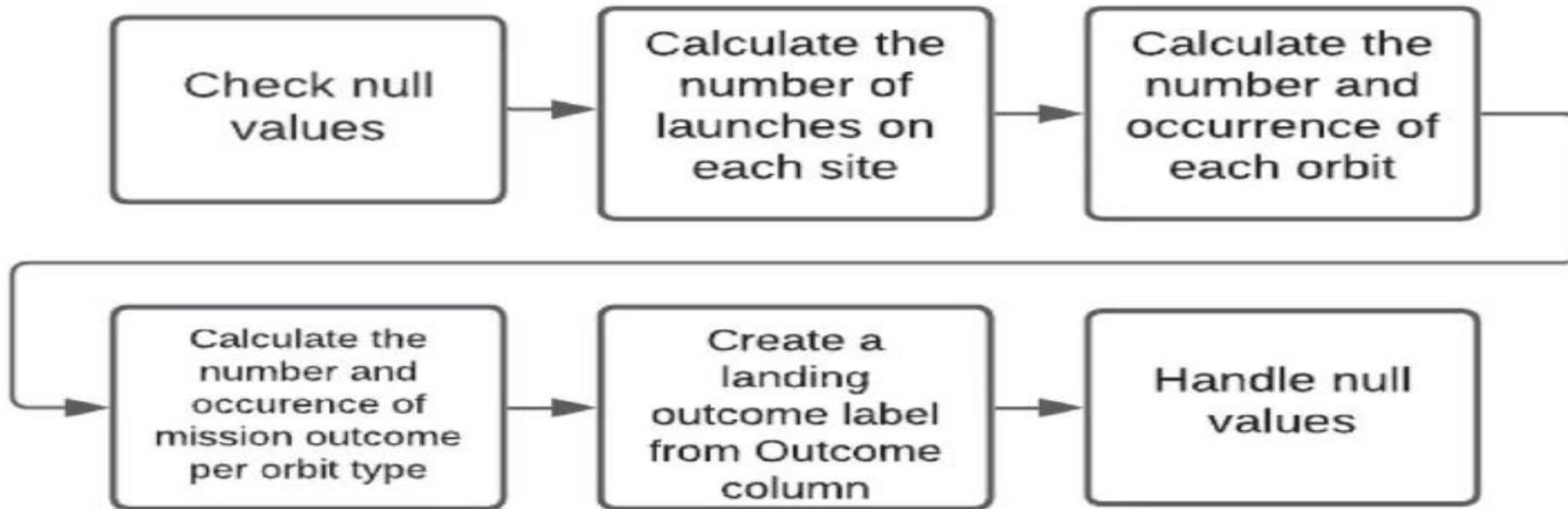
```
df = pd.DataFrame.from_dict(launch_dict)
```

8. Dataframe to .CSV

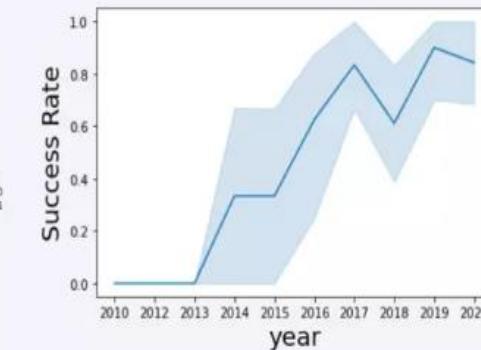
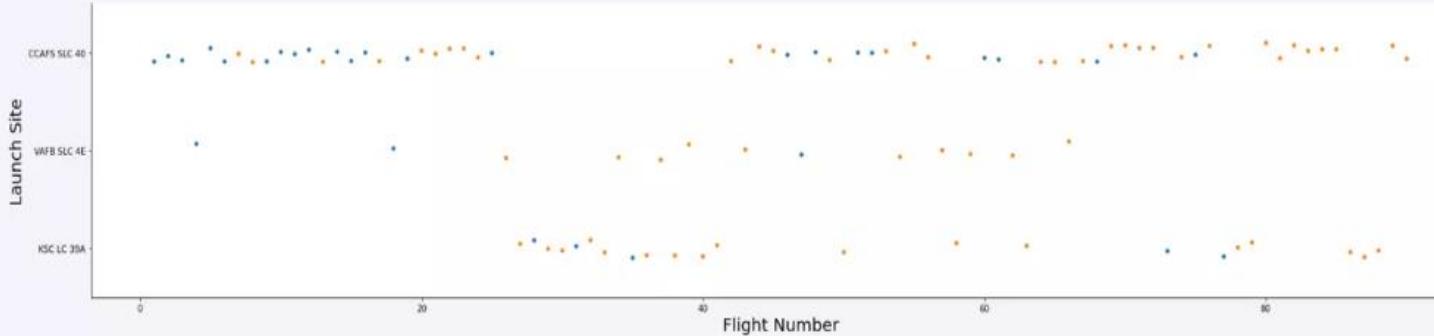
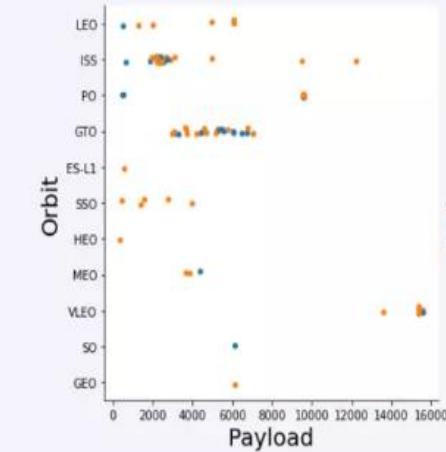
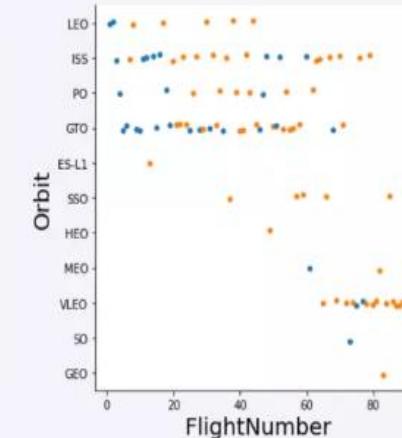
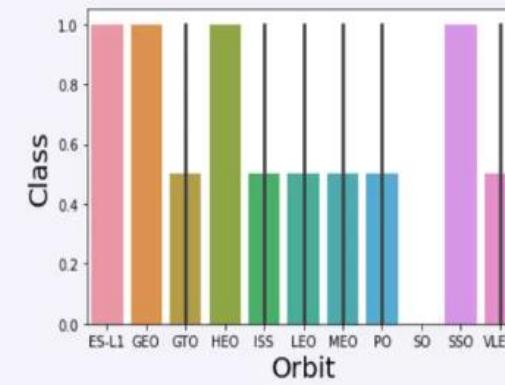
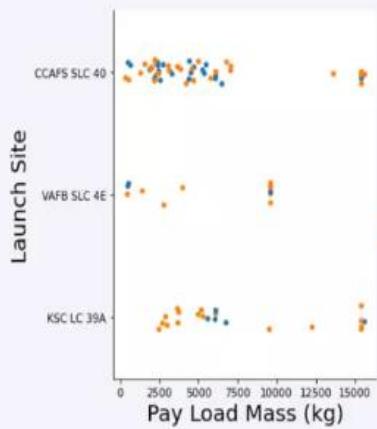
```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Wrangling

EDA analysis



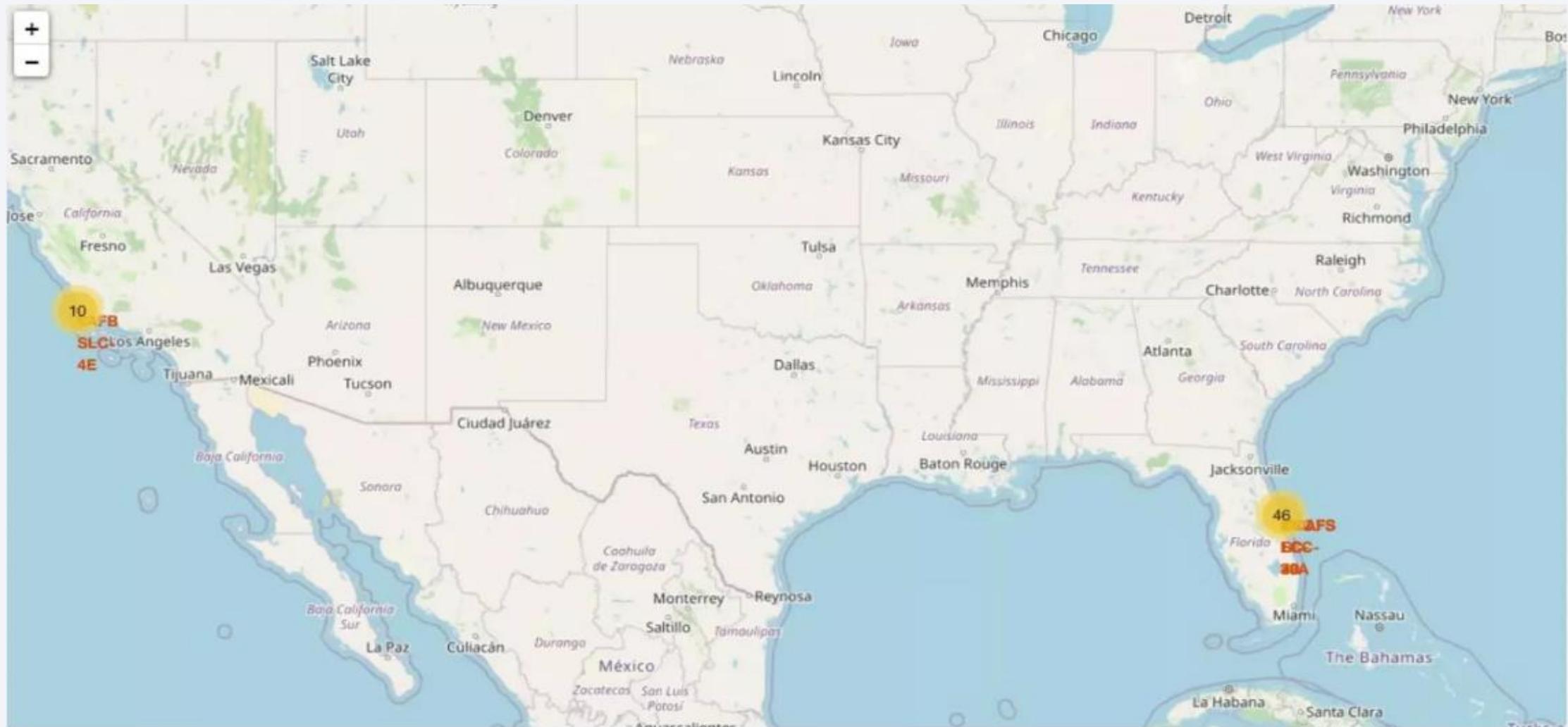
EDA with Data Visualization



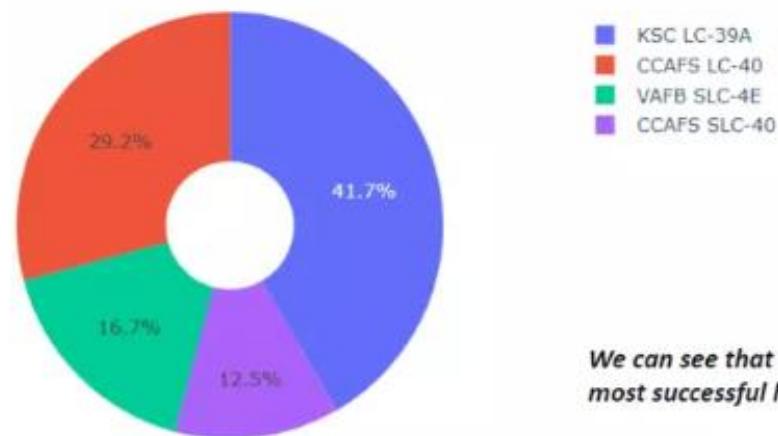
EDA with SQL

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

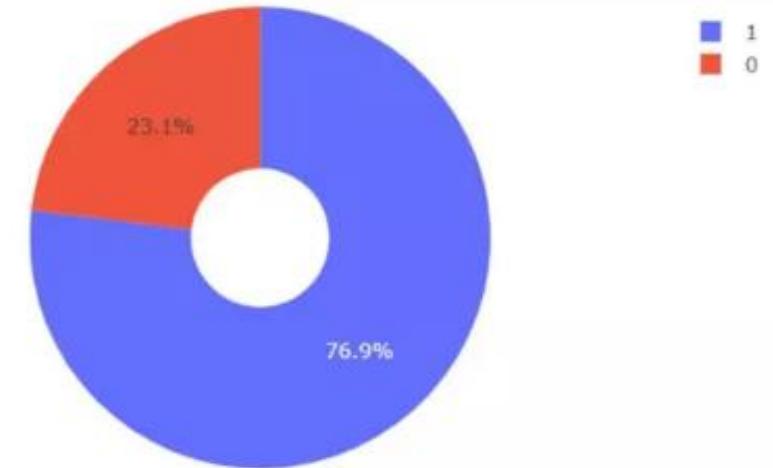
Build an Interactive Map with Folium



Total Success Launches By all sites

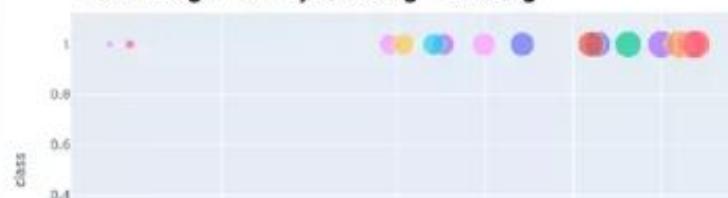


We can see that KSC LC-39A had the most successful launches from all the sites



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

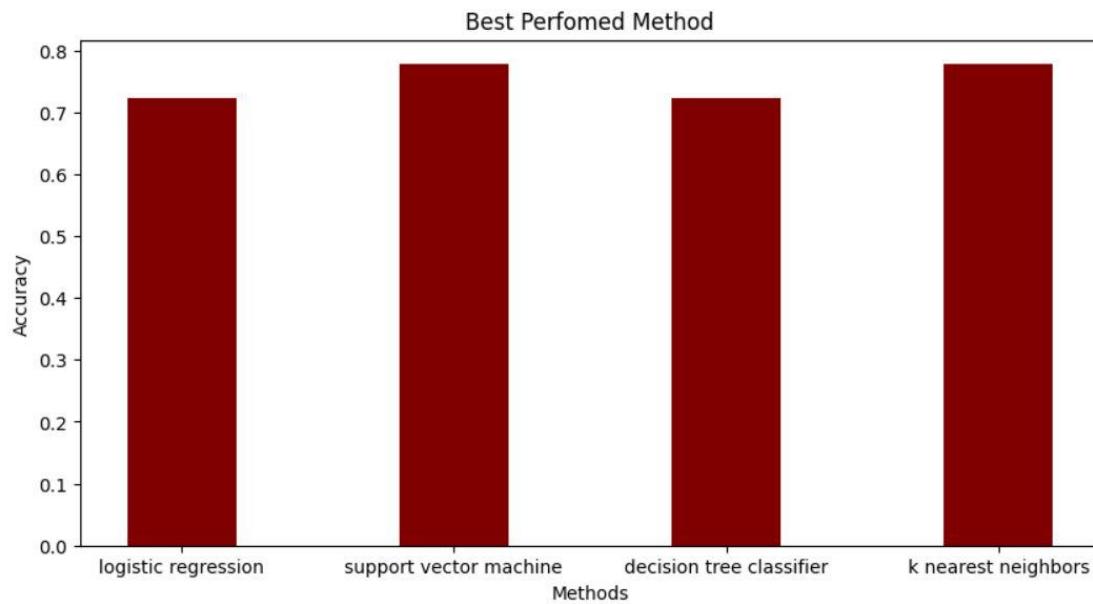
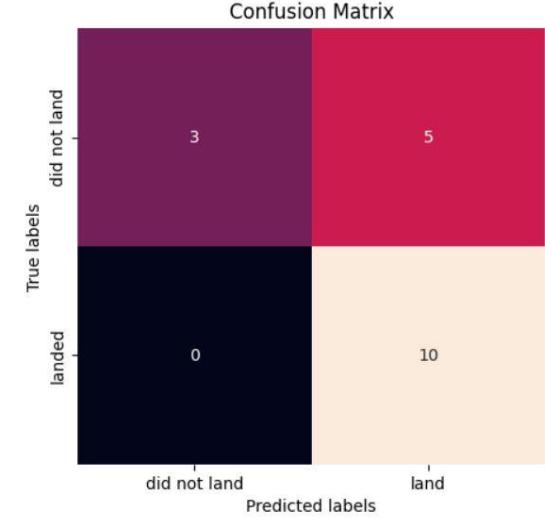
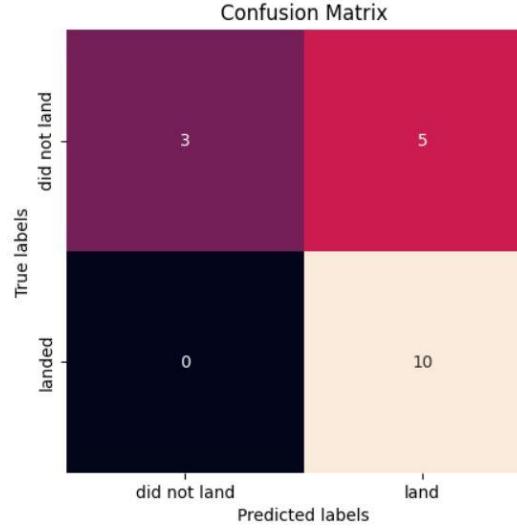
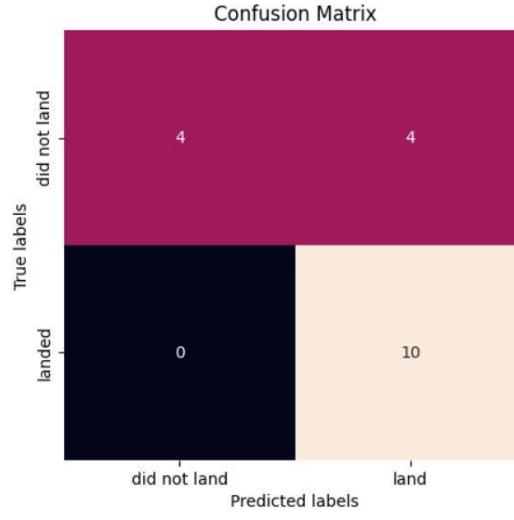
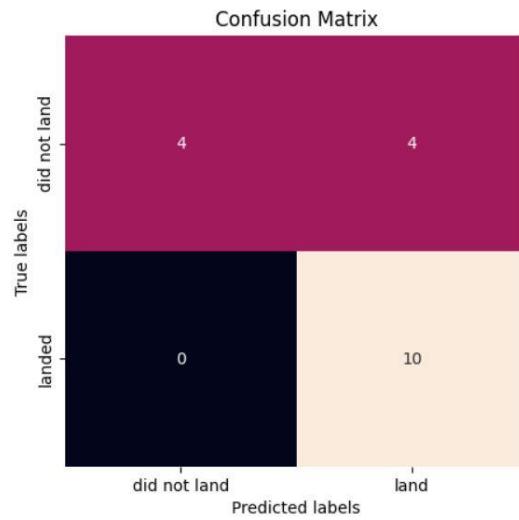
Low Weighted Payload 0kg – 4000kg



Heavy Weighted Payload 4000kg – 10000kg



Dashboard with Plotly Dash



Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.

Results

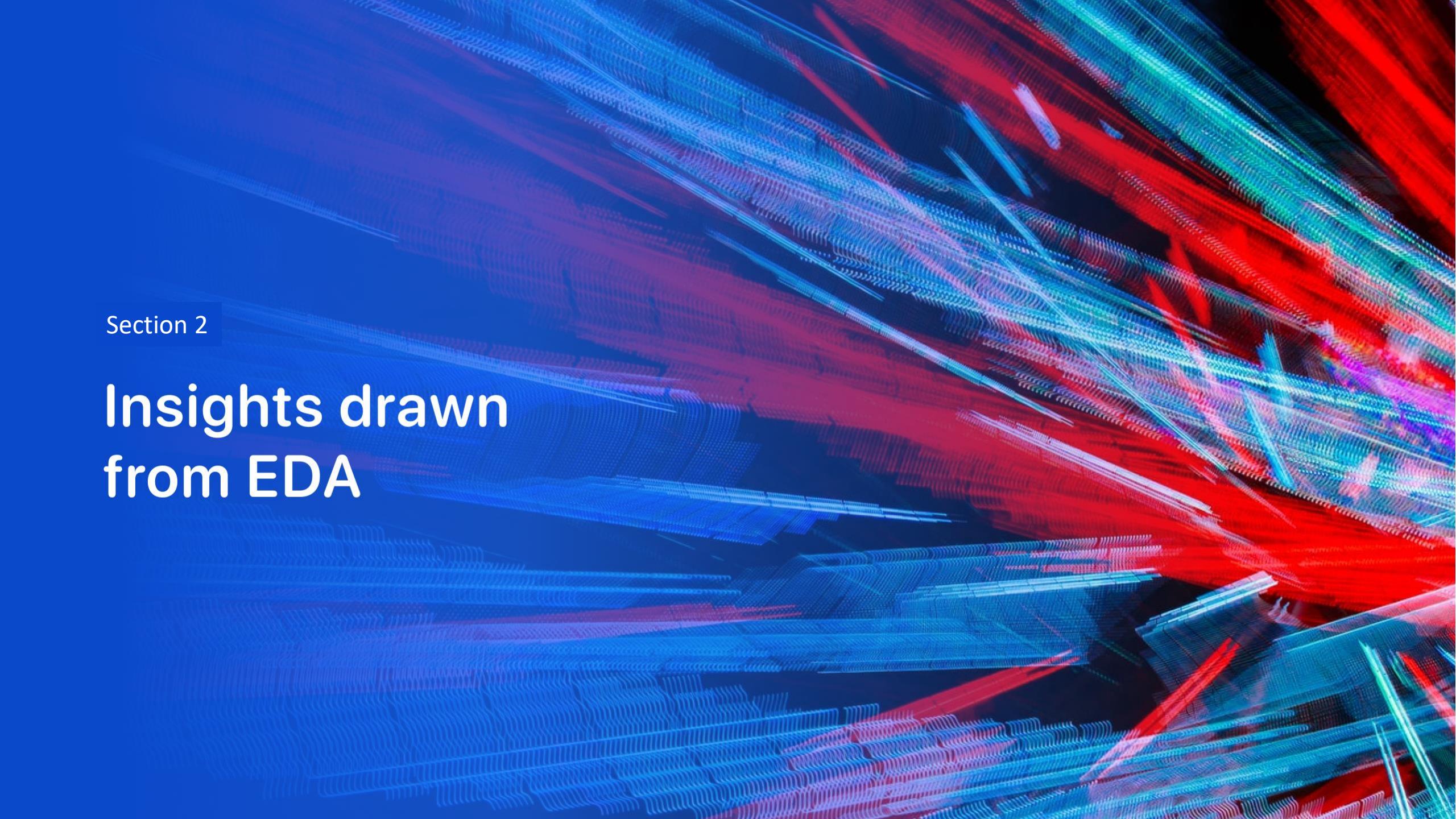
The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.

- Low weighted payloads perform better than the heavier payloads.

The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.

KSC LC 39A had the most successful launches from all the sites.

- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a three-dimensional space or a network of data points. The overall effect is futuristic and dynamic.

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

Payload vs. Launch Site

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

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

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

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

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

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

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

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

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

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

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

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

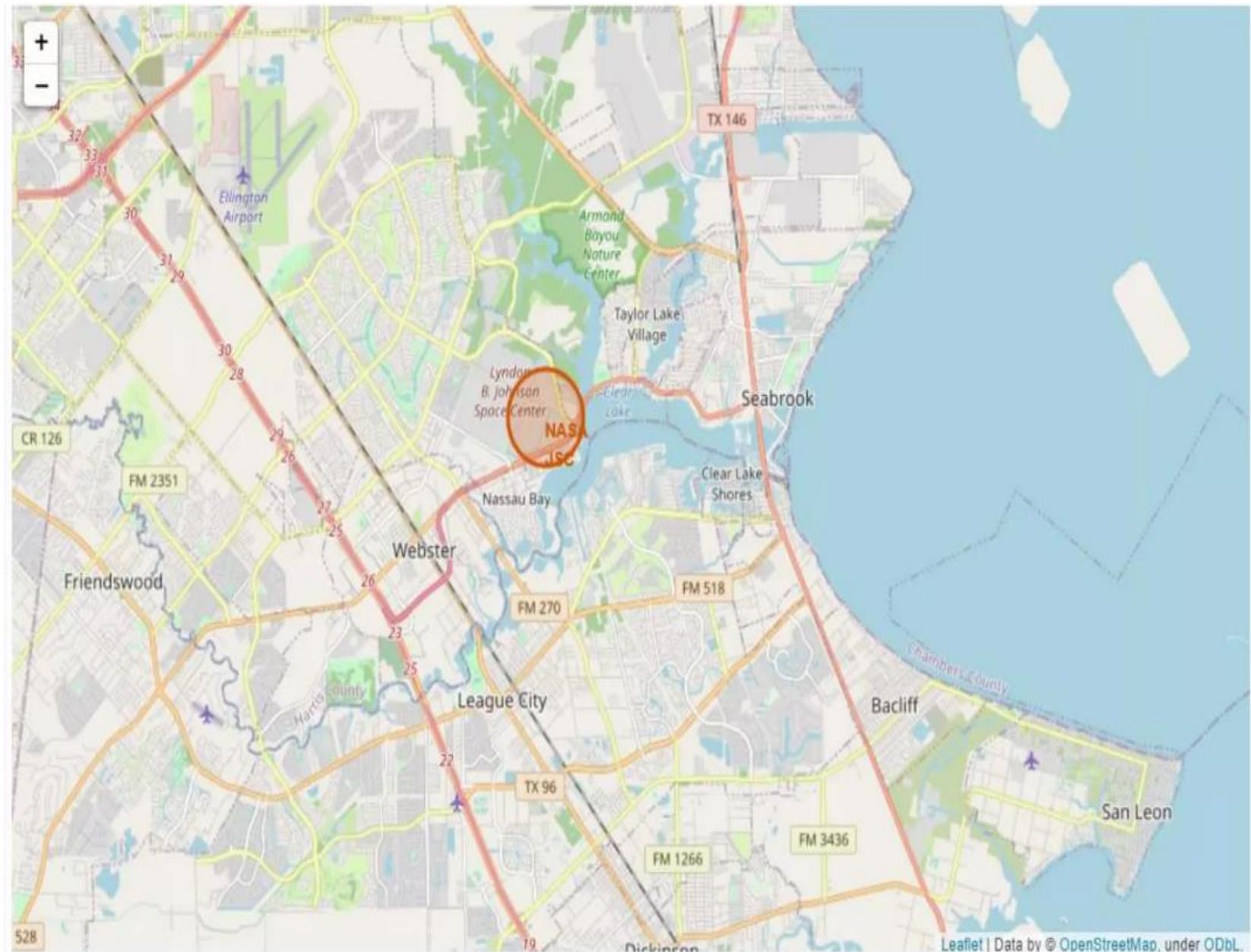
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- Present your query result with a short explanation here

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth's horizon against a dark blue sky. Numerous glowing yellow and white points represent city lights, concentrated in coastal and urban areas. In the upper right quadrant, there are bright green and yellowish bands of light, likely the Aurora Borealis or Australis. The overall atmosphere is dark and mysterious.

Section 3

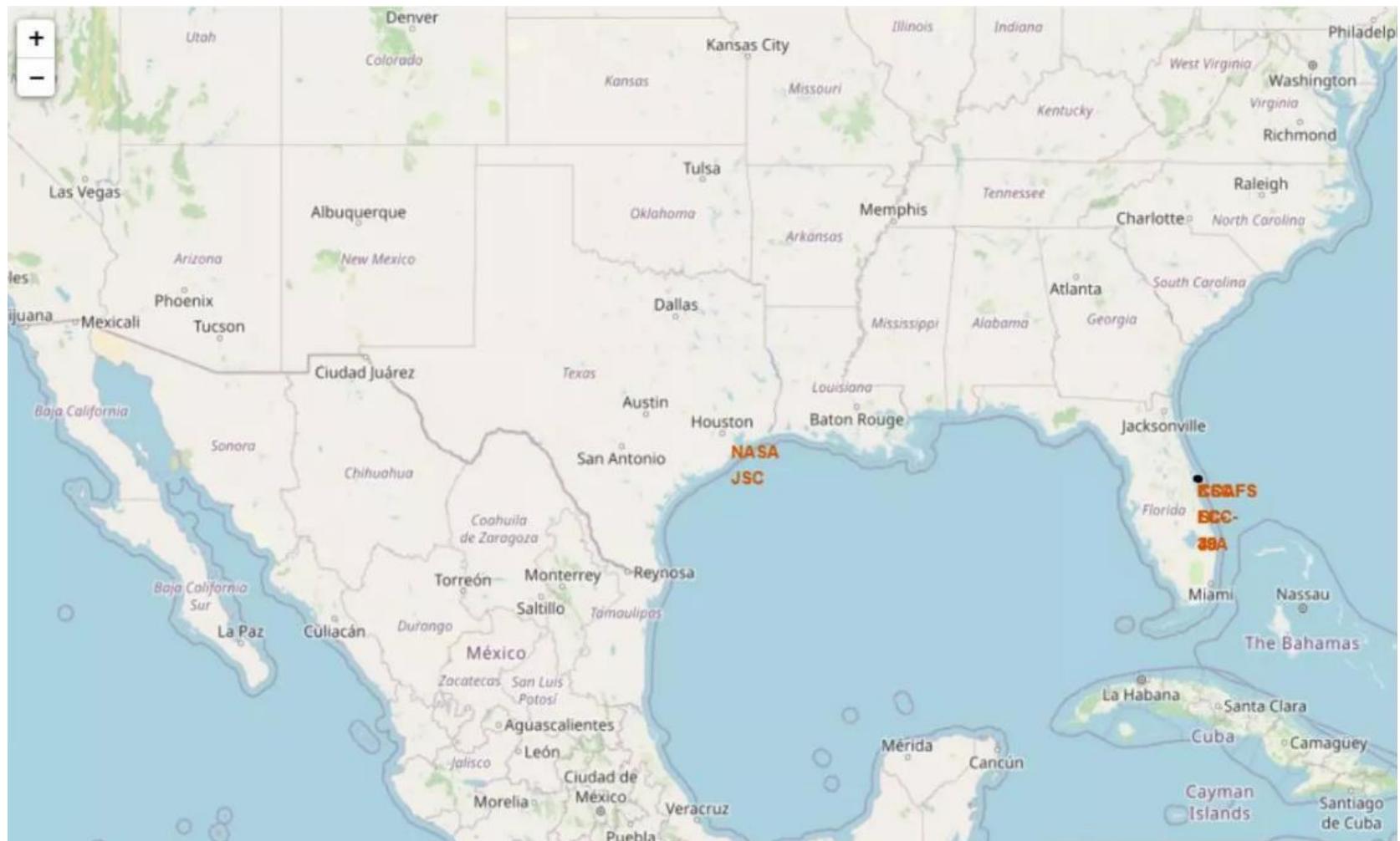
Launch Sites Proximities Analysis

Folium Map Screenshot 1

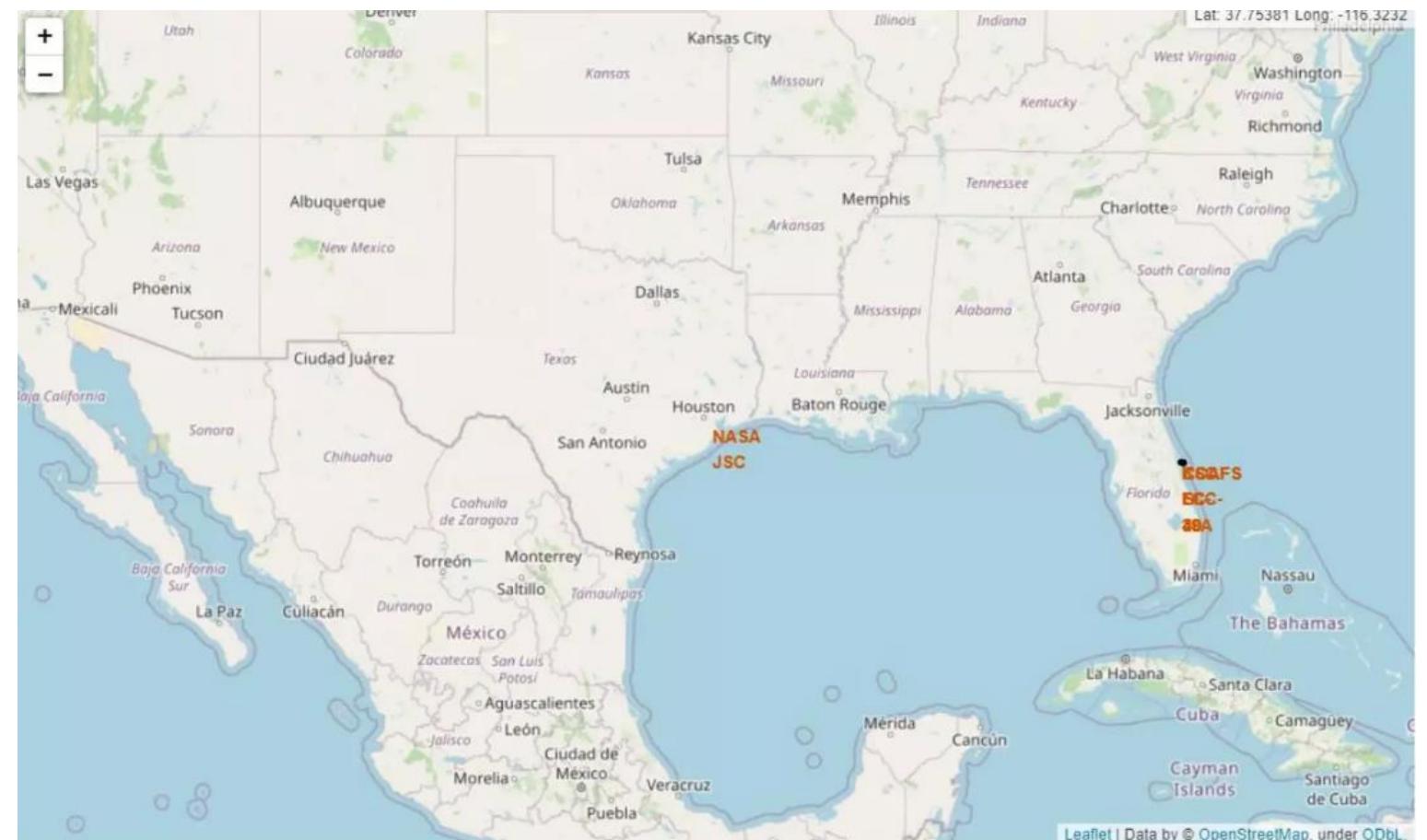


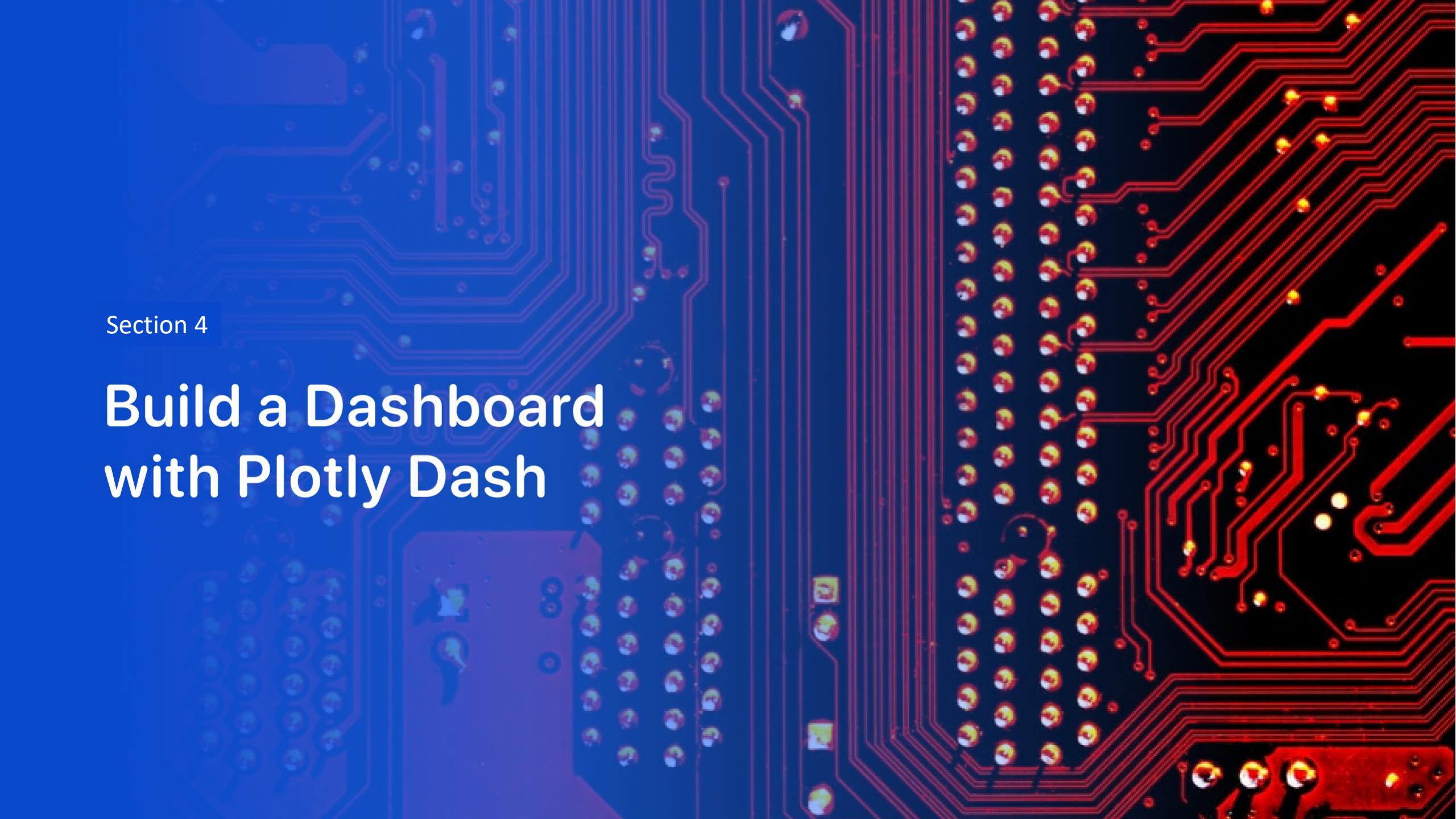
Leaflet | Data by © OpenStreetMap, under ODbL

Folium Map Screenshot 2



Folium Map Screenshot 3



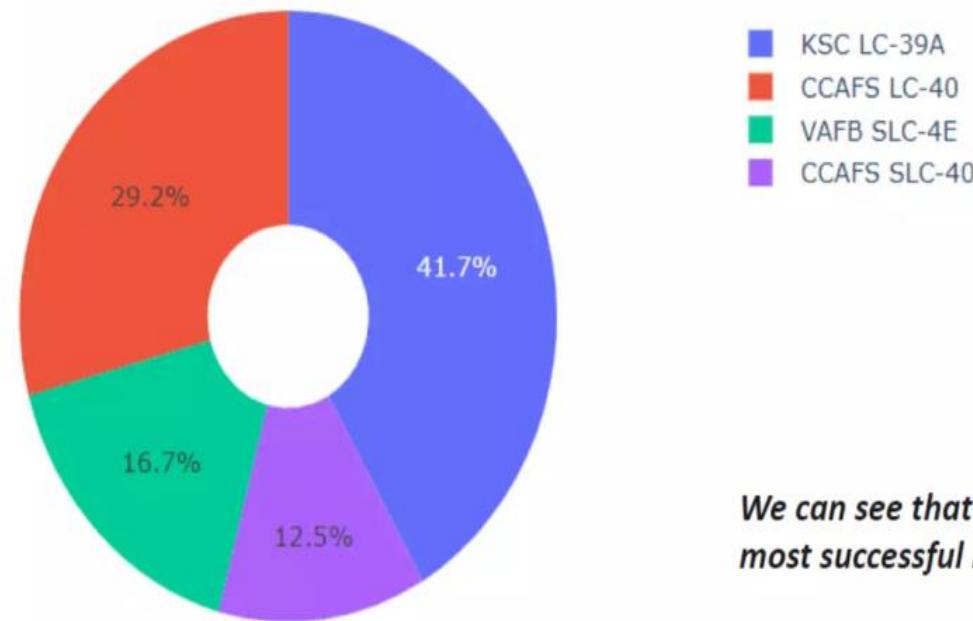


Section 4

Build a Dashboard with Plotly Dash

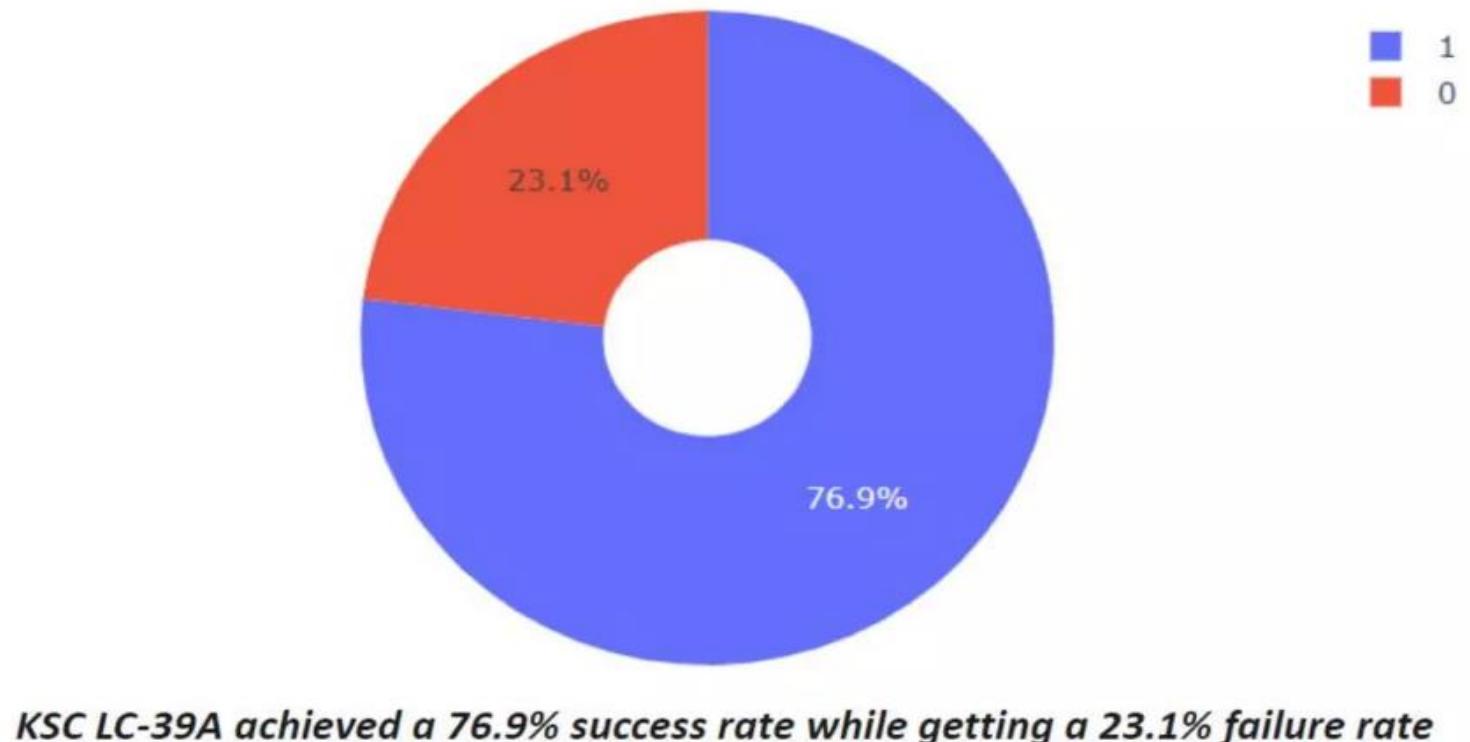
Dashboard Screenshot 1

Total Success Launches By all sites

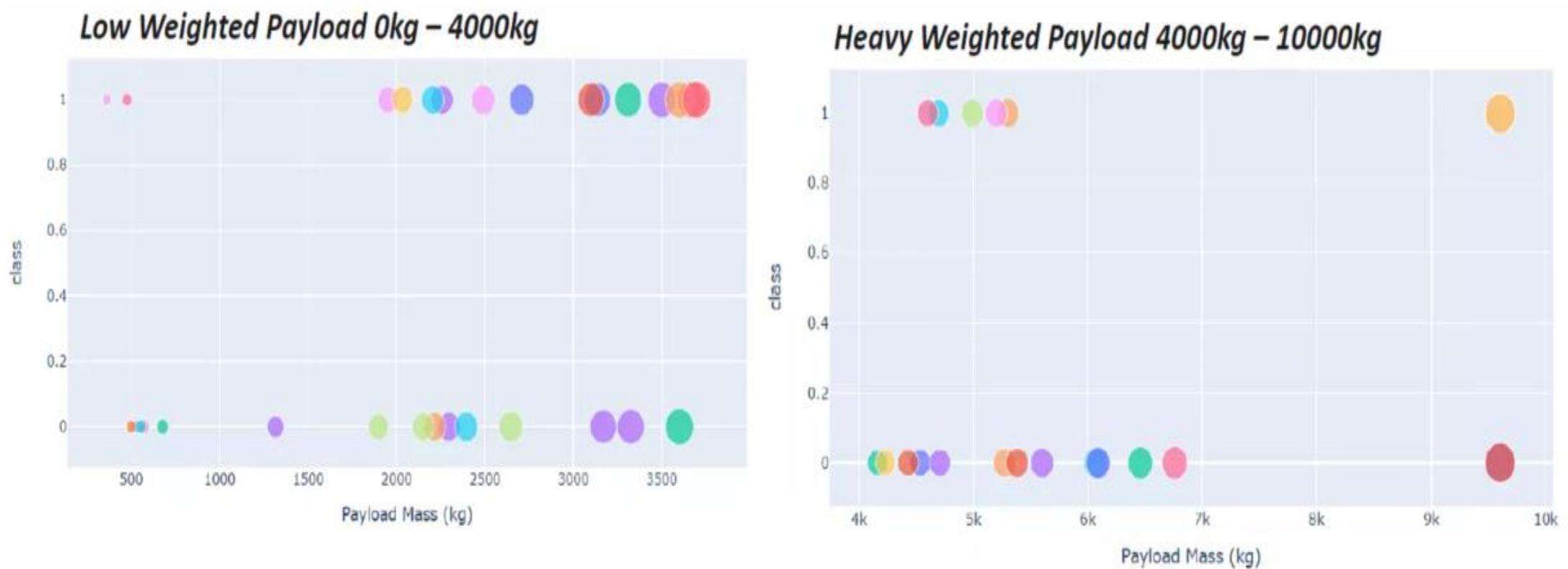


We can see that KSC LC-39A had the most successful launches from all the sites

Dashboard Screenshot 2



Dashboard Screenshot 3



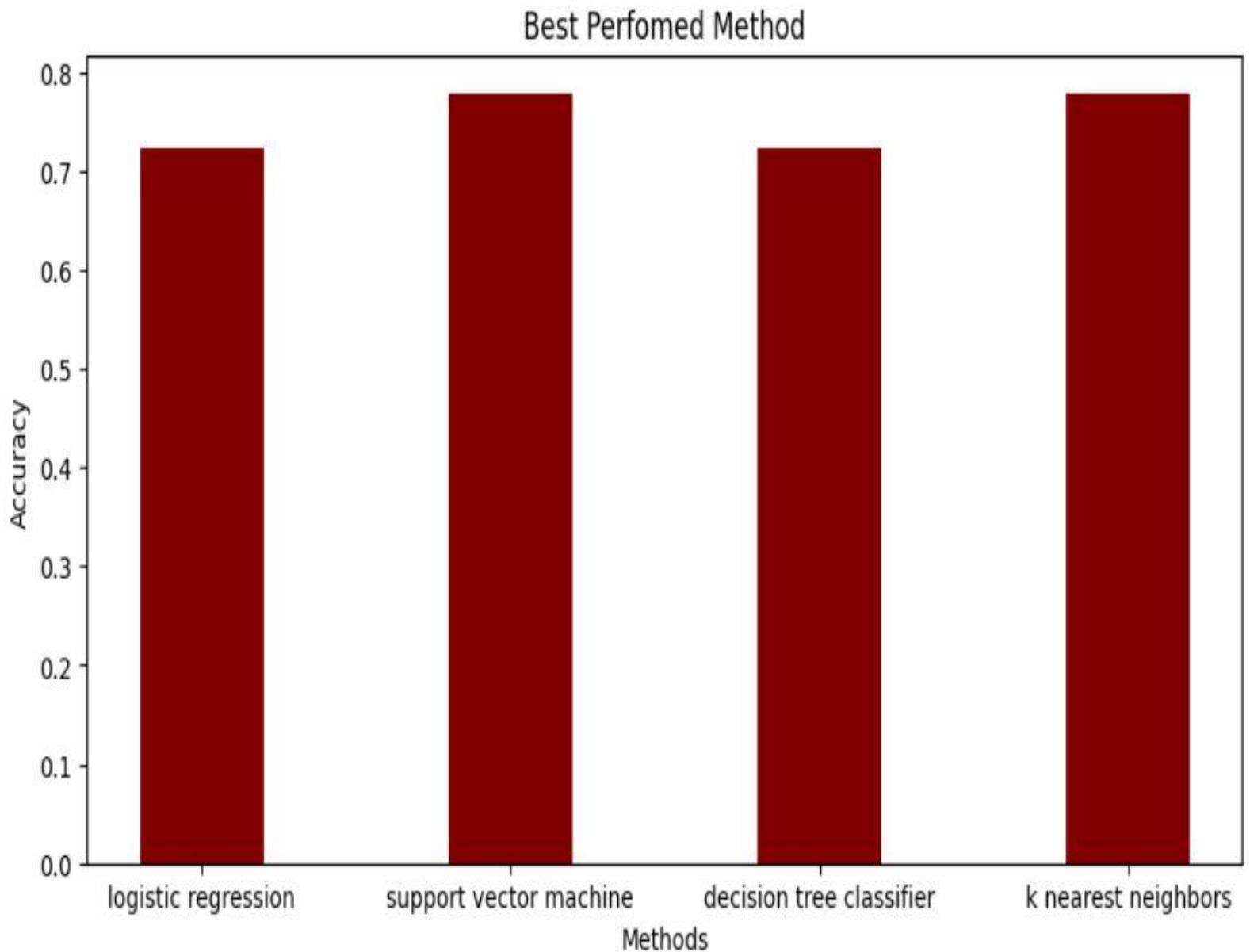
We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

The background of the slide features a dynamic, abstract design. It consists of several curved, overlapping bands of color. A prominent band on the left is a bright blue, while another on the right is a warm yellow. These colors transition into lighter, more diffused tones towards the edges of the frame. The overall effect is one of motion and depth.

Section 5

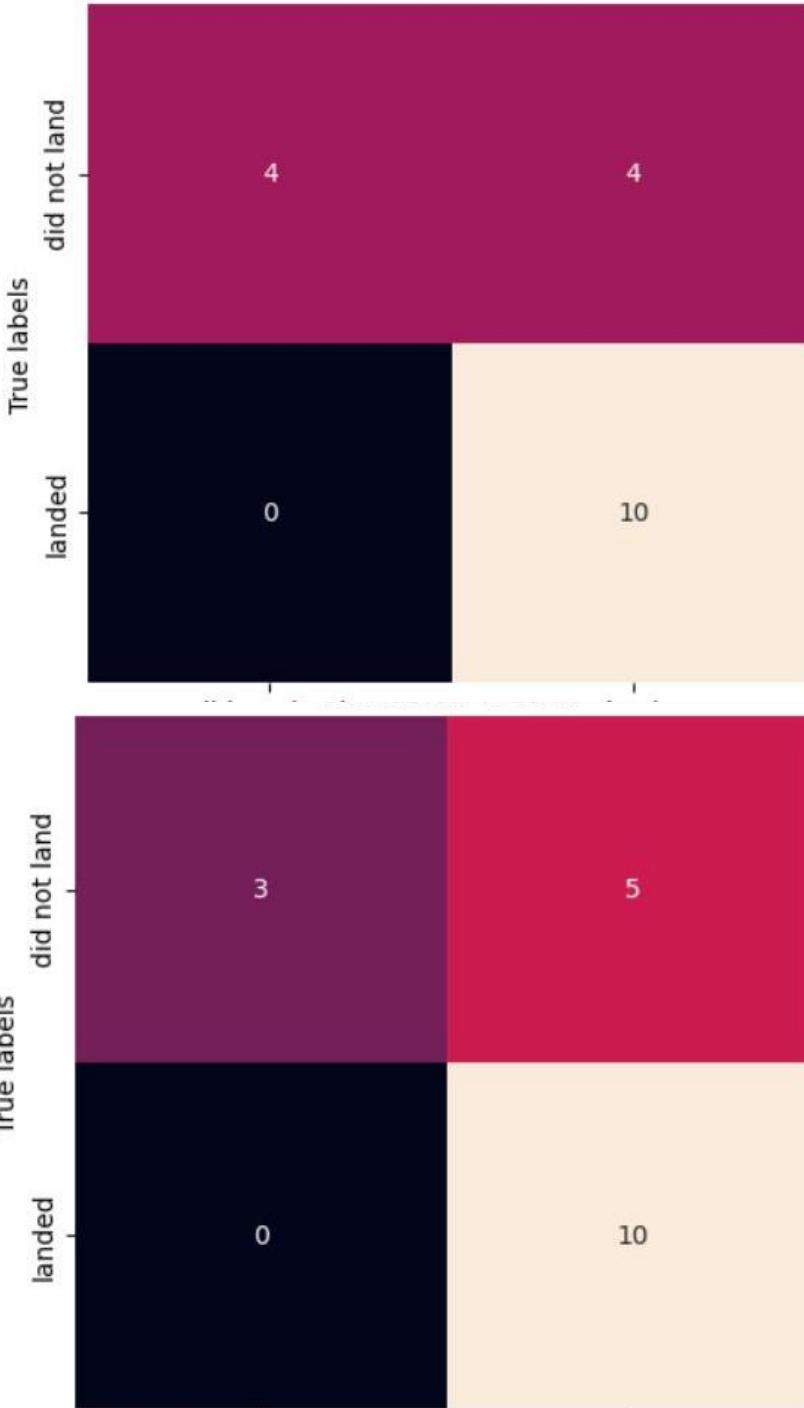
Predictive Analysis (Classification)

Classification Accuracy

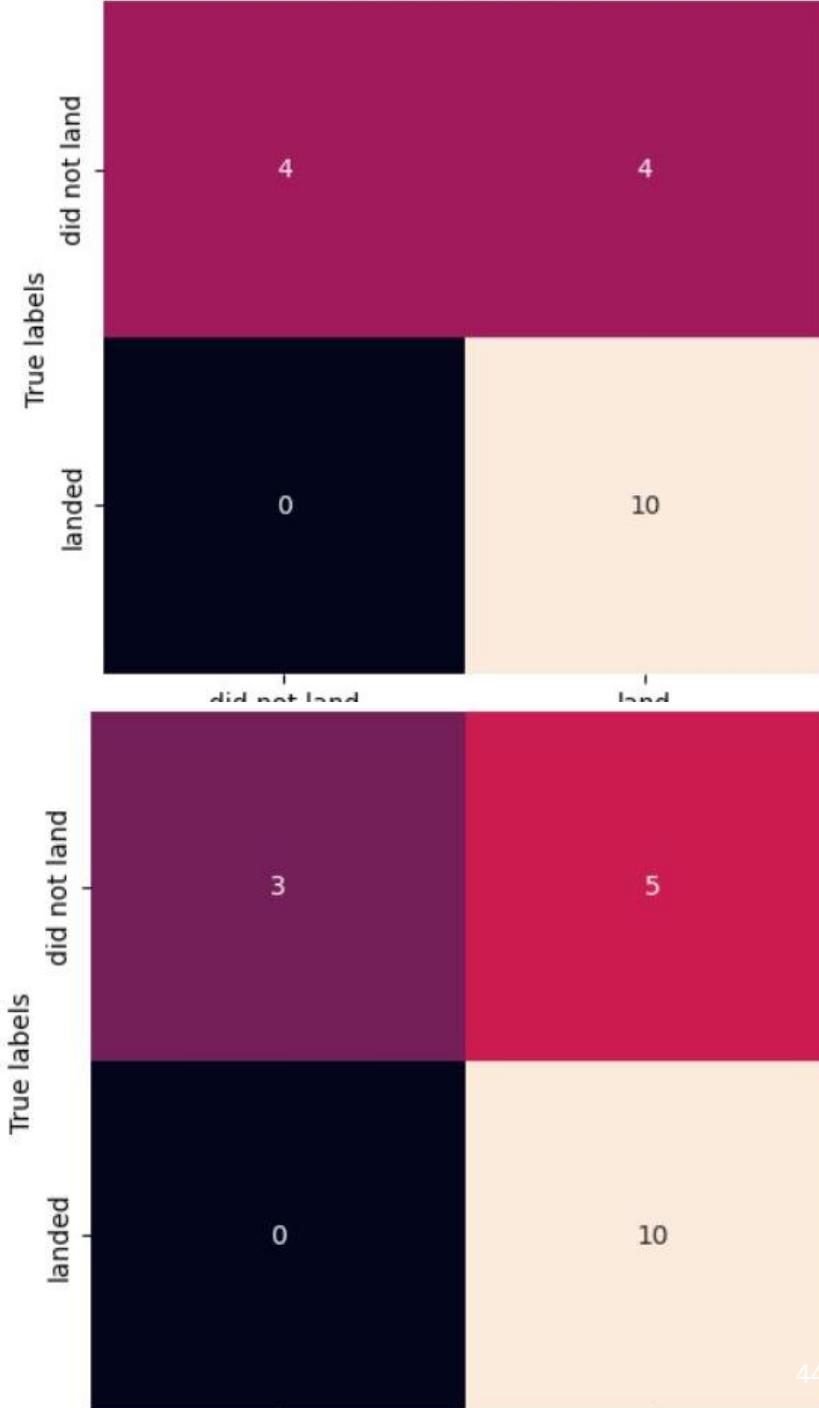


Confusion Matrix

Confusion Matrix



Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!

