IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

<Name>
<Date>

# Outline

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

# Executive Summary

## Summary of methodologies

1. Data collection via API
2. Data collection by web scraping
3. Exploratory data analysis using SQL
4. Exploratory data analysis with data visualization
5. Interactive maps with Folium
6. Machine learning predictions

## Summary of all results

1. Results of exploratory data analysis
2. Machine learning model predictions

# Introduction

## Project background and context

SpaceX advertises Falcon 9 rocket launches with a cost of 62 millions dollars unlike other providers who cost upwards of 165 million each. Much of these savings are because SpaceX can reuse the first stage. So the aim of the project was to analyze SpaceX data and determine whether the first stage will land successfully or not so that we can determine the cost of each launch.

## Problems you want to find answers

1. What features are needed to determine if a launch will be successful or not
2. What conditions are needed to ensure a successful landing

Section 1

# Methodology

# Methodology

- Data collection methodology:
  - Data was collected using SpaceX API and webscraping from Wikipedia
- Perform data wrangling
  - One hot encoding was applied, data was cleaned using Pandas library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different models like KNN, SVM were used and accuracy was compared

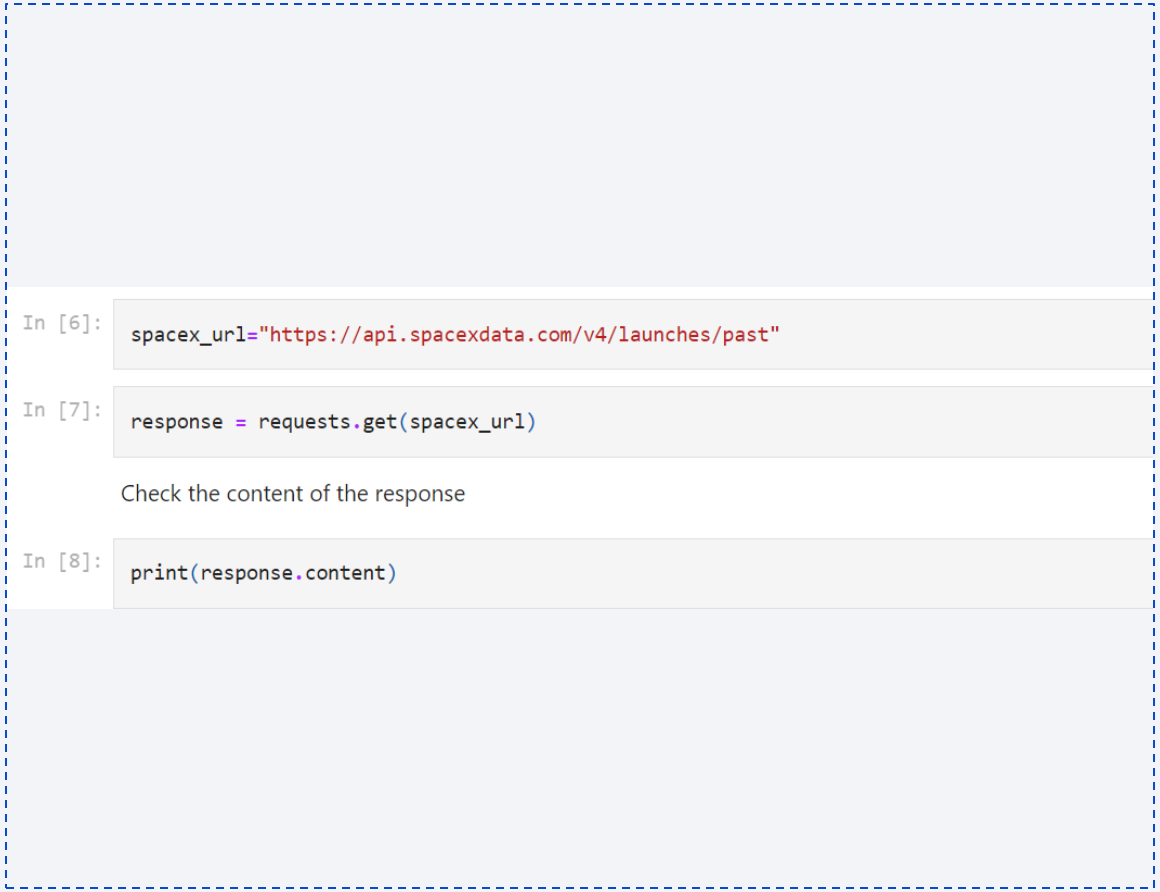
# Data Collection

1. Data was collected using requests library on the SpaceX API.
2. We read the response content as a JSON file and converted it to a dataframe using .json_normalize().
3. We also performed webscraping on a Wikipedia page about Falcon 9 landings using the BeautifulSoup library.
4. We parsed the HTML content, read table contents, and converted it to a dataframe for future use.

# Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/Data%20Collection%20API.ipynb



# Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/Data%20Collection%20API.ipynb

```
# use requests.get() method with the provided static_url
# assign the response to a object
html_data = requests.get(static_url)
html_data.status_code
```

```
200
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(html_data.text, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

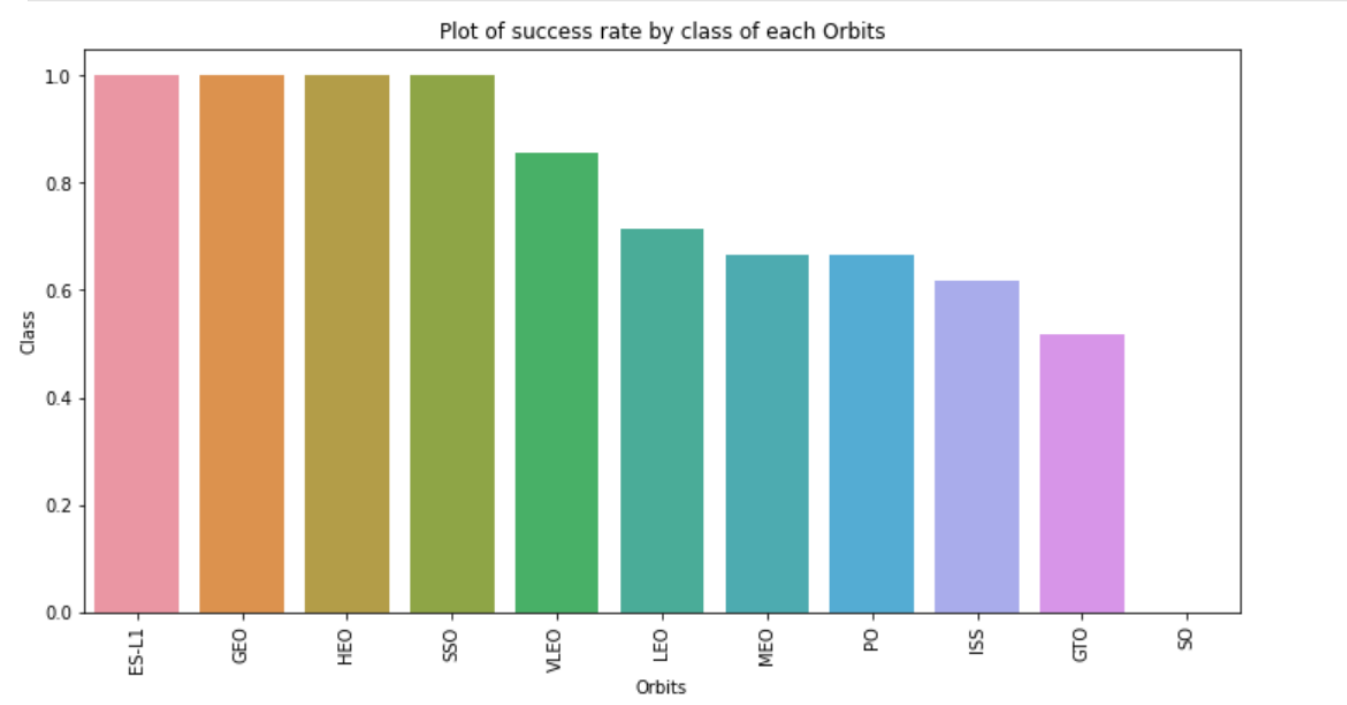
```
# Use soup.title attribute
soup.title
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

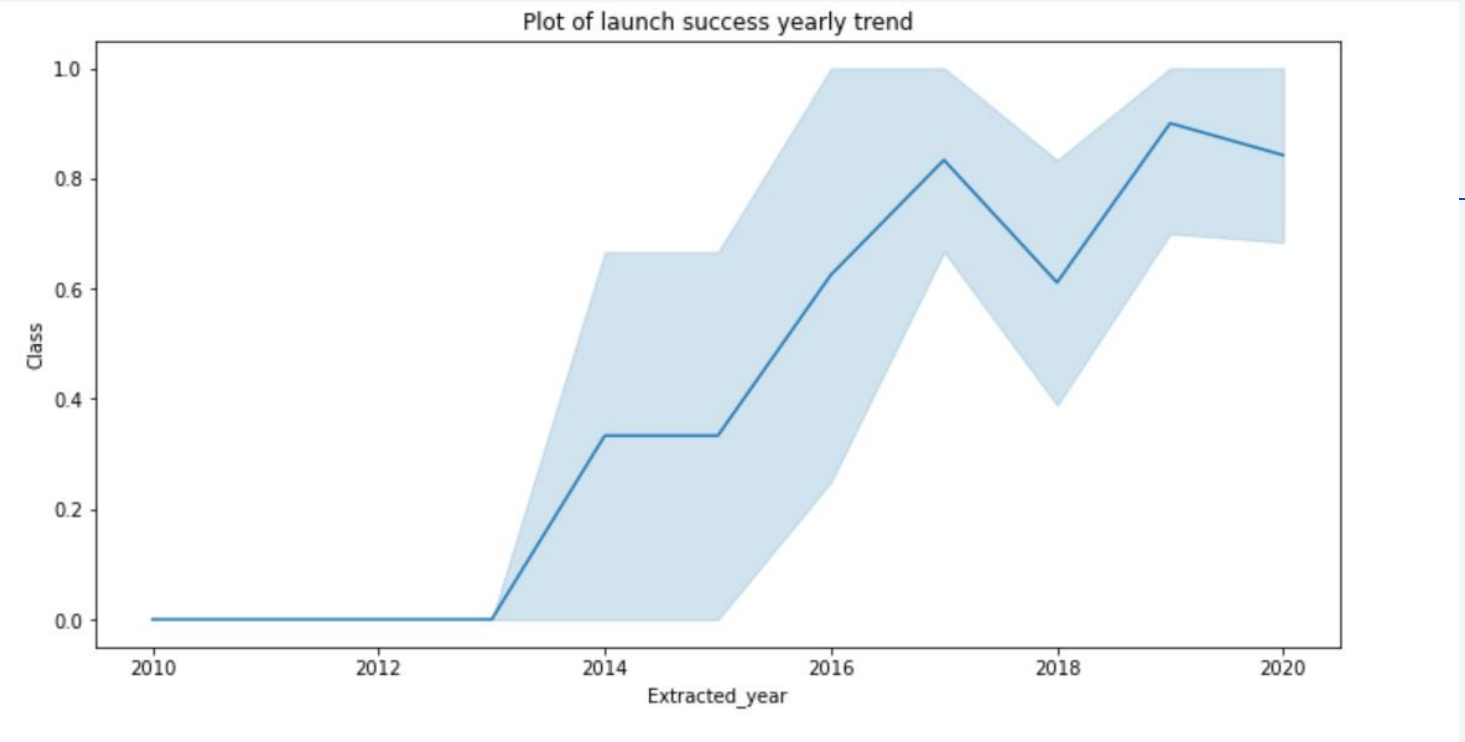
# Data Wrangling

- The data was cleaned by removing null or invalid values, formatting the data for consistency, and columns were adjusted.
- We performed exploratory data analysis using pandas functions and determined the features important for training.
- We calculated the number of launches at each site and the number of each orbits by using groupby function on rows.
- We created a landing outcome column.
- https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/Data%20Wrangling.ipynb

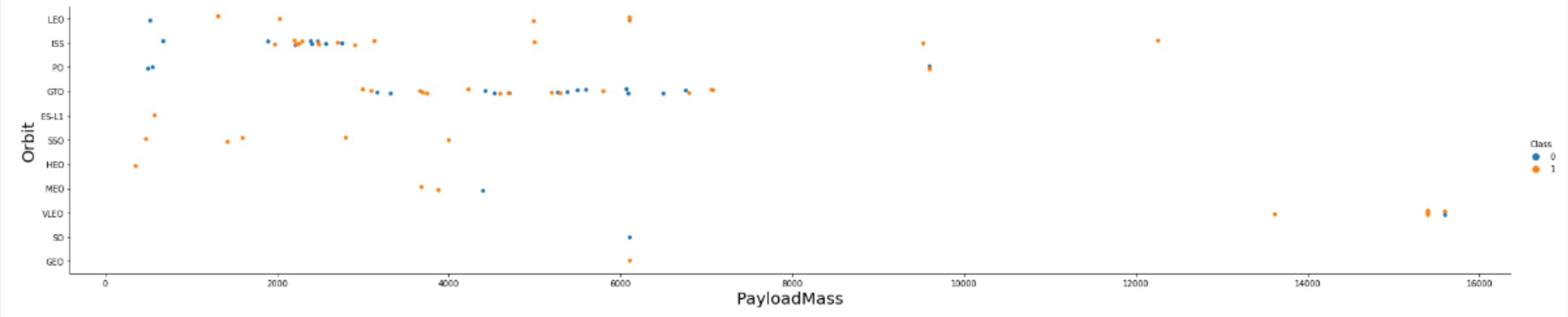
# EDA with Data Visualization




We visualised the relationship between several features including flight number vs launch site, payload vs launch site, success rate vs orbit, flight number vs orbit, etc.




https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/EDA%20with%20Data%20Visualization.ipynb



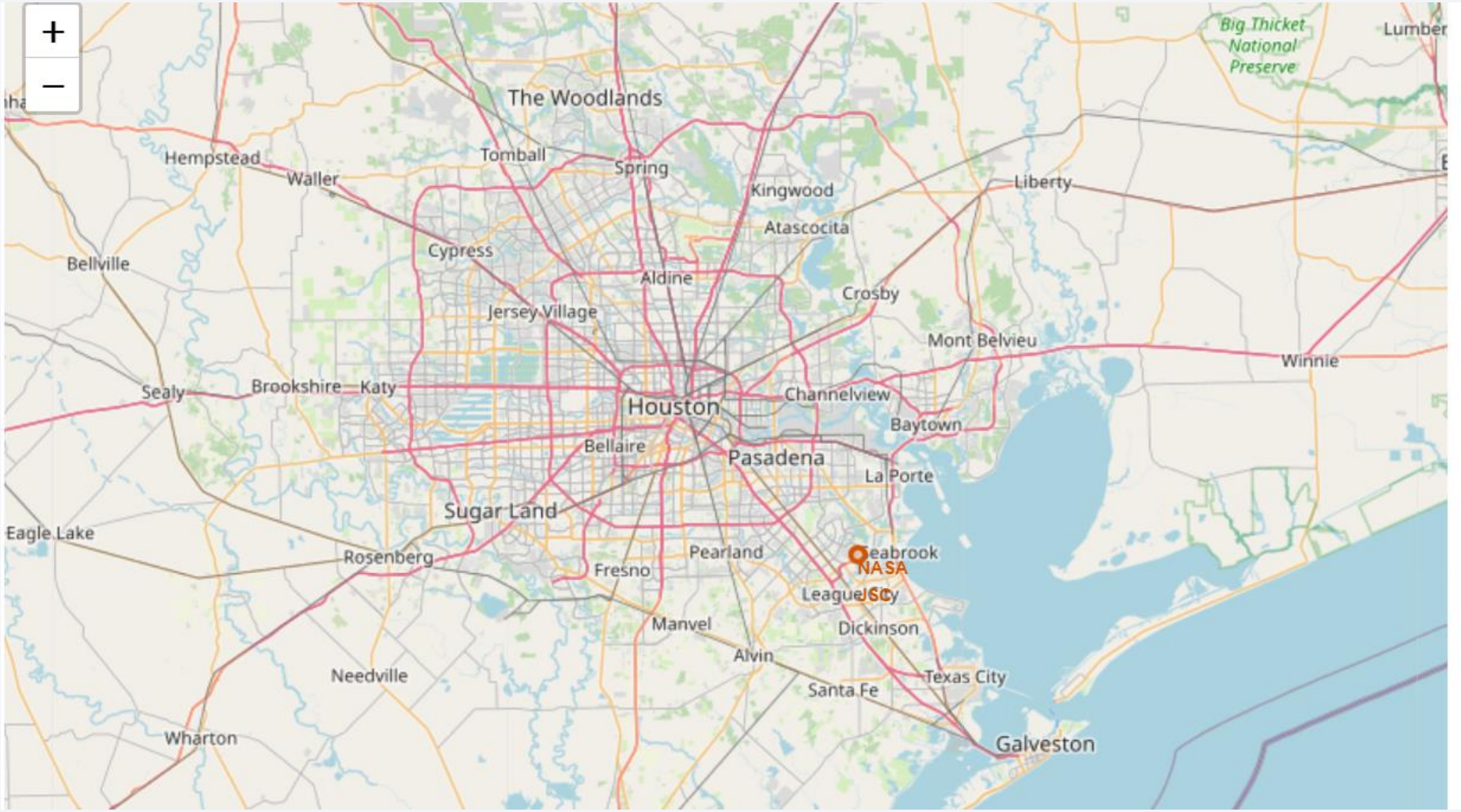
# EDA with SQL

We loaded the data into a database and established a connection. Then we wrote queries using magic SQL for-

1. Finding unique launch sites in the data
2. Finding total payload mass carried by NASA (CRS)
3. Finding the dates when successful landing on ground pads was achieved
4. Finding names of the boosters with weight within particular range which have success in drone ship
5. Listing total number of successes and failures

https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/EDA%20with%20SQL.ipynb

# Build an Interactive Map with Folium



# Predictive Analysis (Classification)

The data was split into training and testing sets after cleaning and analysis.

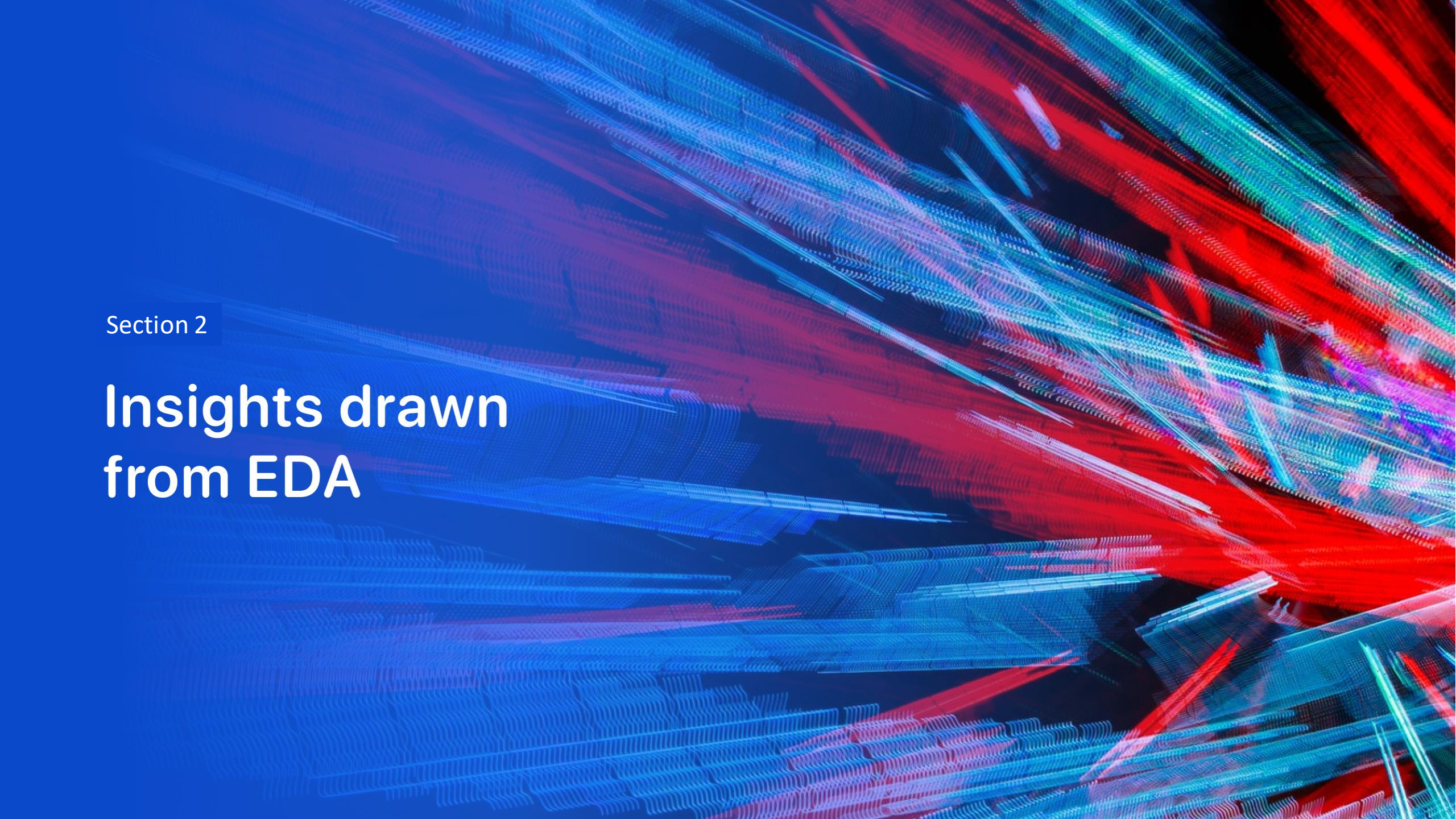
We tuned the hyperparameters using GridSearchCV.

Using different models such as KNN, SVM, logistic regression, etc. , the rate of accuracy was determined for every one of them so we can decide which one performs the best.

https://github.com/saniyavijayvargiya/ibm-assignments/blob/main/Machine%20Learning%20Prediction.ipynb

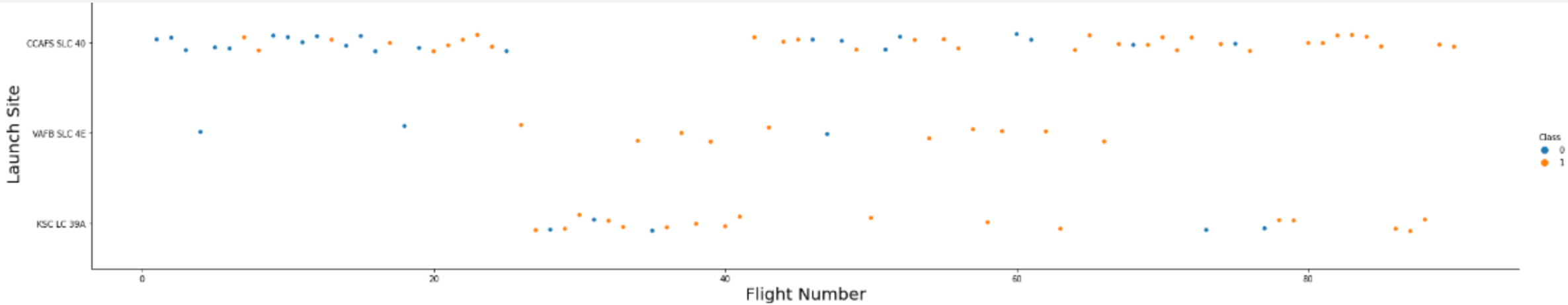
Section 2

# Insights drawn from EDA



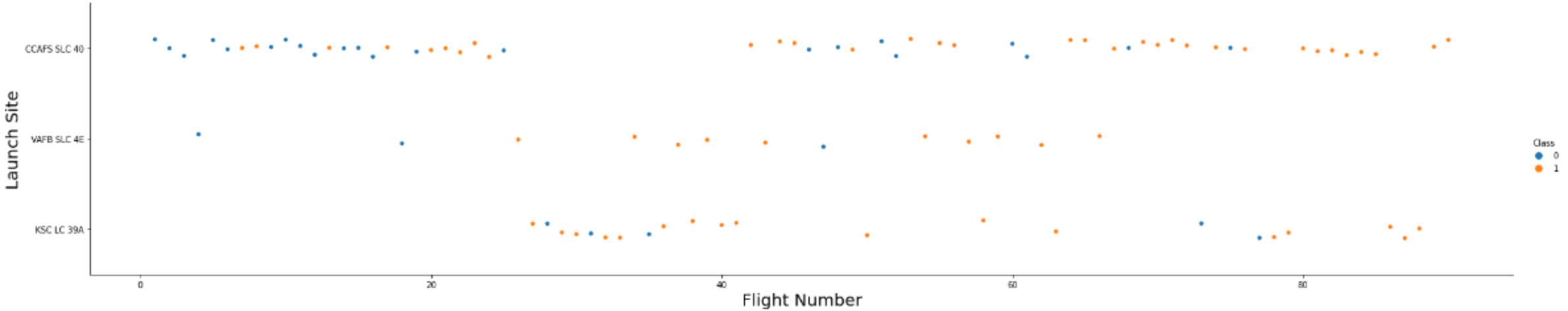
# Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site



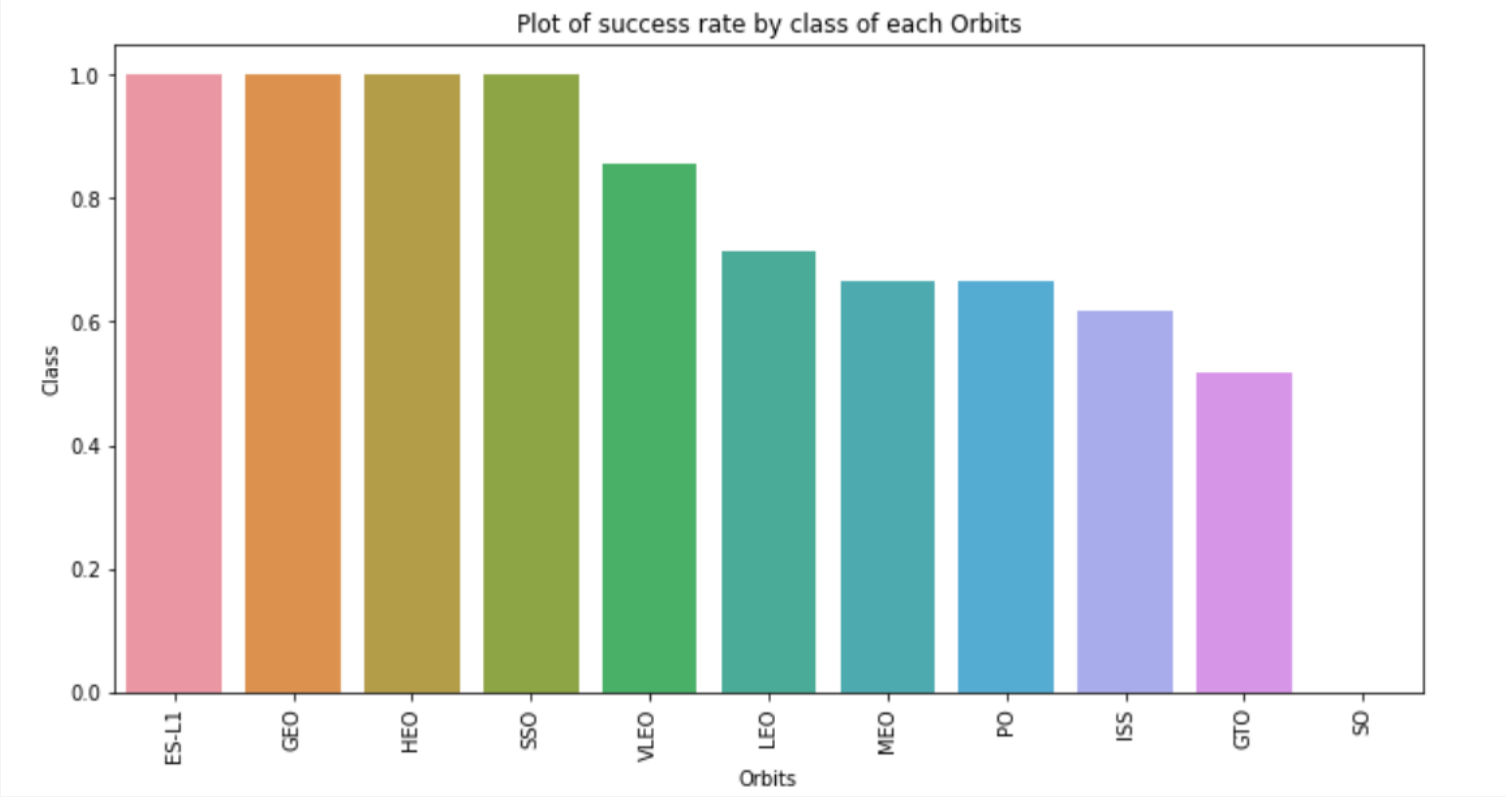
# Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



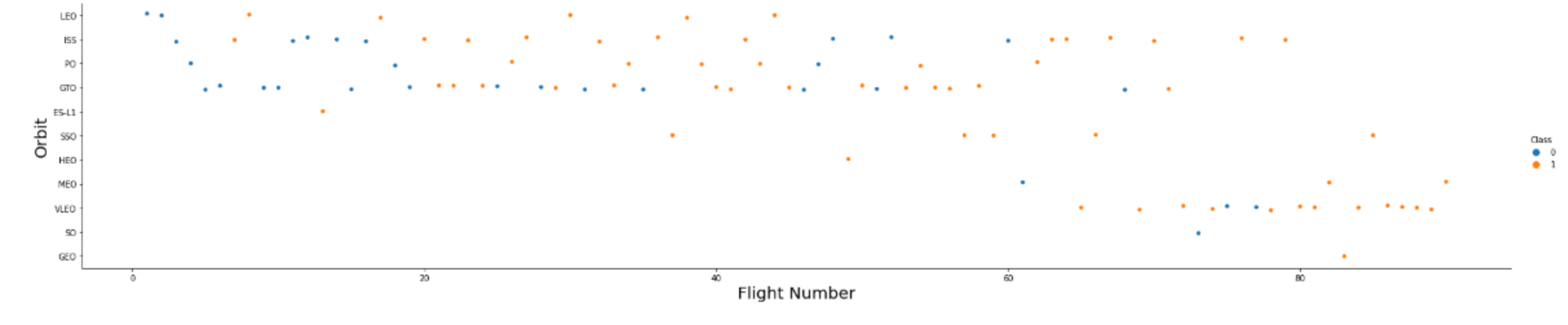
# Success Rate vs. Orbit Type

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



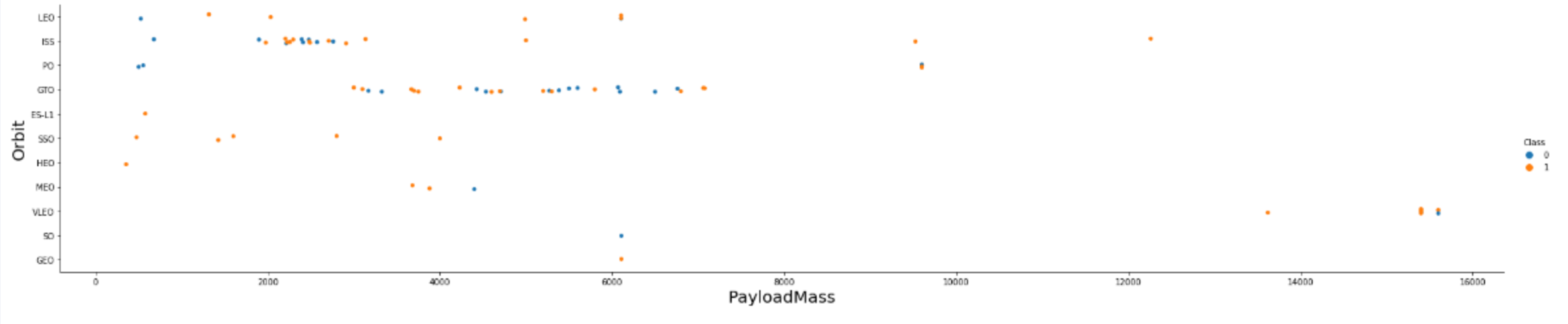
# Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type



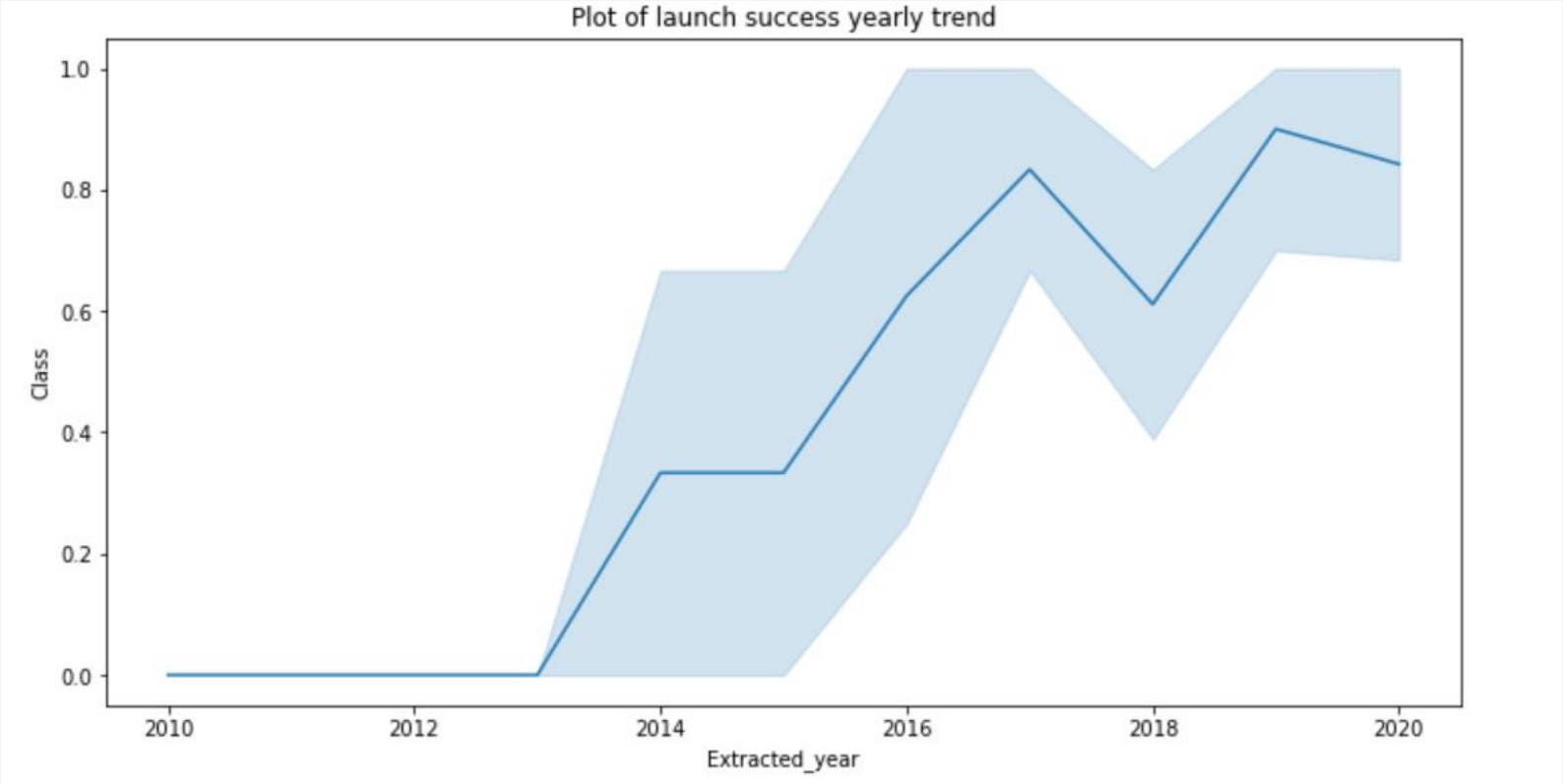
# Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



# Launch Success Yearly Trend

- Show a line chart of yearly average success rate



# All Launch Site Names

- Find the names of the unique launch sites

| | launchsite |
|---|---|
| 0 | KSC LC-39A |
| 1 | CCAFS LC-40 |
| 2 | CCAFS SLC-40 |
| 3 | VAFB SLC-4E |

# Launch Site Names Begin with 'CCA'

| | date | time | boosterversion | launchsite |
|---|---|---|---|---|
| **0** | 2010-04-06 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 |
| **1** | 2010-08-12 | 15:43:00 | F9 v1.0 B0004 | CCAFS LC-40 |
| **2** | 2012-05-22 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 |
| **3** | 2012-08-10 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 |
| **4** | 2013-01-03 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 |

# Total Payload Mass

- Calculate the total payload carried by boosters from NASA

| | total_payloadmass |
|---|---|
| 0 | 45596 |

# Average Payload Mass by F9 v1.1

| | avg_payloadmass |
|---|---|
| 0 | 2928.4 |

# First Successful Ground Landing Date

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

| | firstsuccessfull_landing_date |
|---|---|
| 0 | 2015-12-22 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

| | boosterversion |
|---|---|
| 0 | F9 FT B1022 |
| 1 | F9 FT B1026 |
| 2 | F9 FT B1021.2 |
| 3 | F9 FT B1031.2 |

# Total Number of Successful and Failure Mission Outcomes

| | successoutcome |
|---|---|
| **0** | 100 |

| | failureoutcome |
|---|---|
| **0** | 1 |

# Boosters Carried Maximum Payload

| | boosterversion | payloadmasskg |
|---|---|---|
| 0 | F9 B5 B1048.4 | 15600 |
| 1 | F9 B5 B1048.5 | 15600 |
| 2 | F9 B5 B1049.4 | 15600 |
| 3 | F9 B5 B1049.5 | 15600 |
| 4 | F9 B5 B1049.7 | 15600 |
| 5 | F9 B5 B1051.3 | 15600 |

# 2015 Launch Records

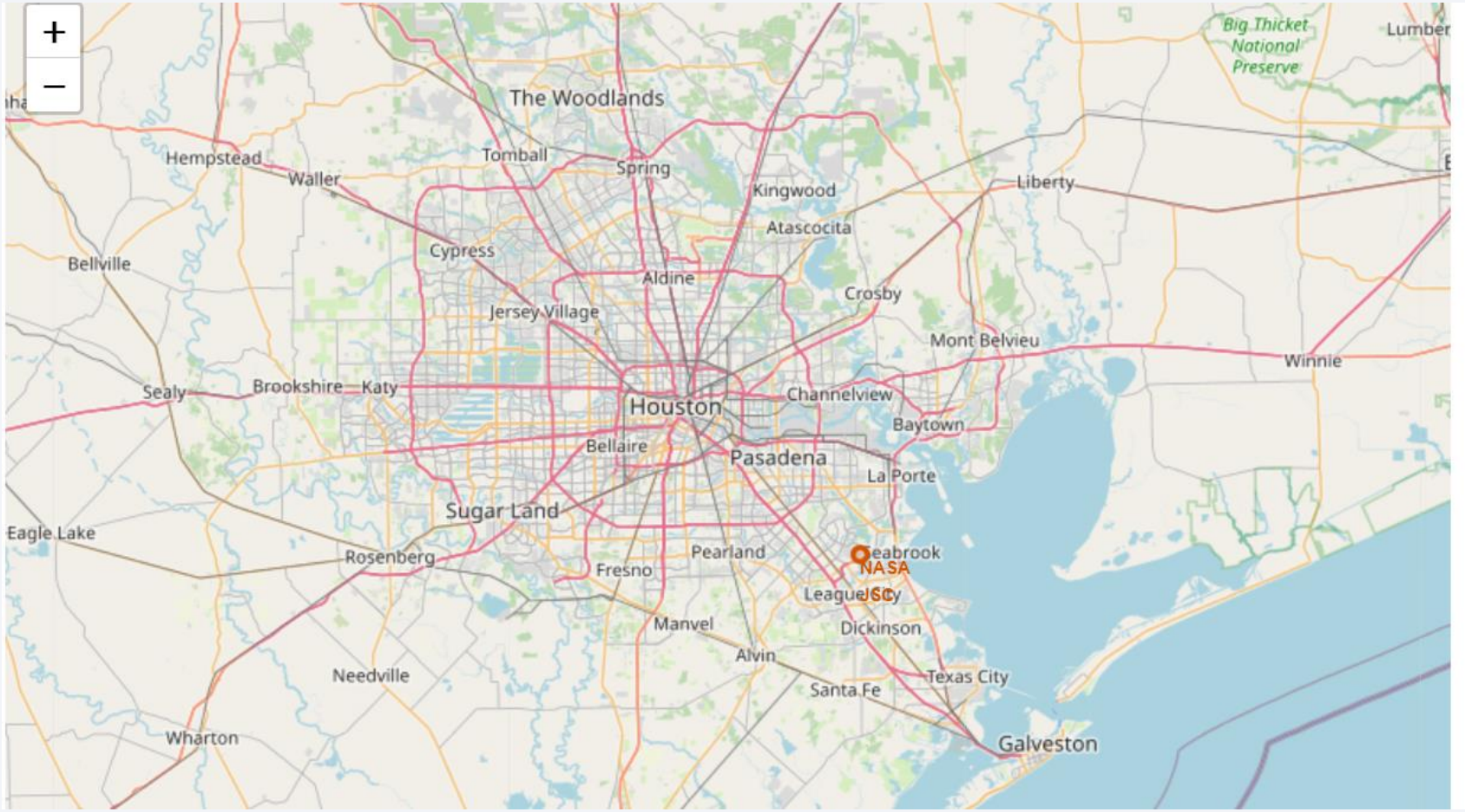
| | boosterversion | launchsite | landingoutcome |
|---|---|---|---|
| **0** | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| **1** | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

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

| | landingoutcome | count |
|---|---|---|
| **0** | No attempt | 10 |
| **1** | Success (drone ship) | 6 |
| **2** | Failure (drone ship) | 5 |
| **3** | Success (ground pad) | 5 |
| **4** | Controlled (ocean) | 3 |
| **5** | Uncontrolled (ocean) | 2 |
| **6** | Precluded (drone ship) | 1 |
| **7** | Failure (parachute) | 1 |


Section 3
Launch Sites
Proximities Analysis

# <Folium Map Screenshot 1>

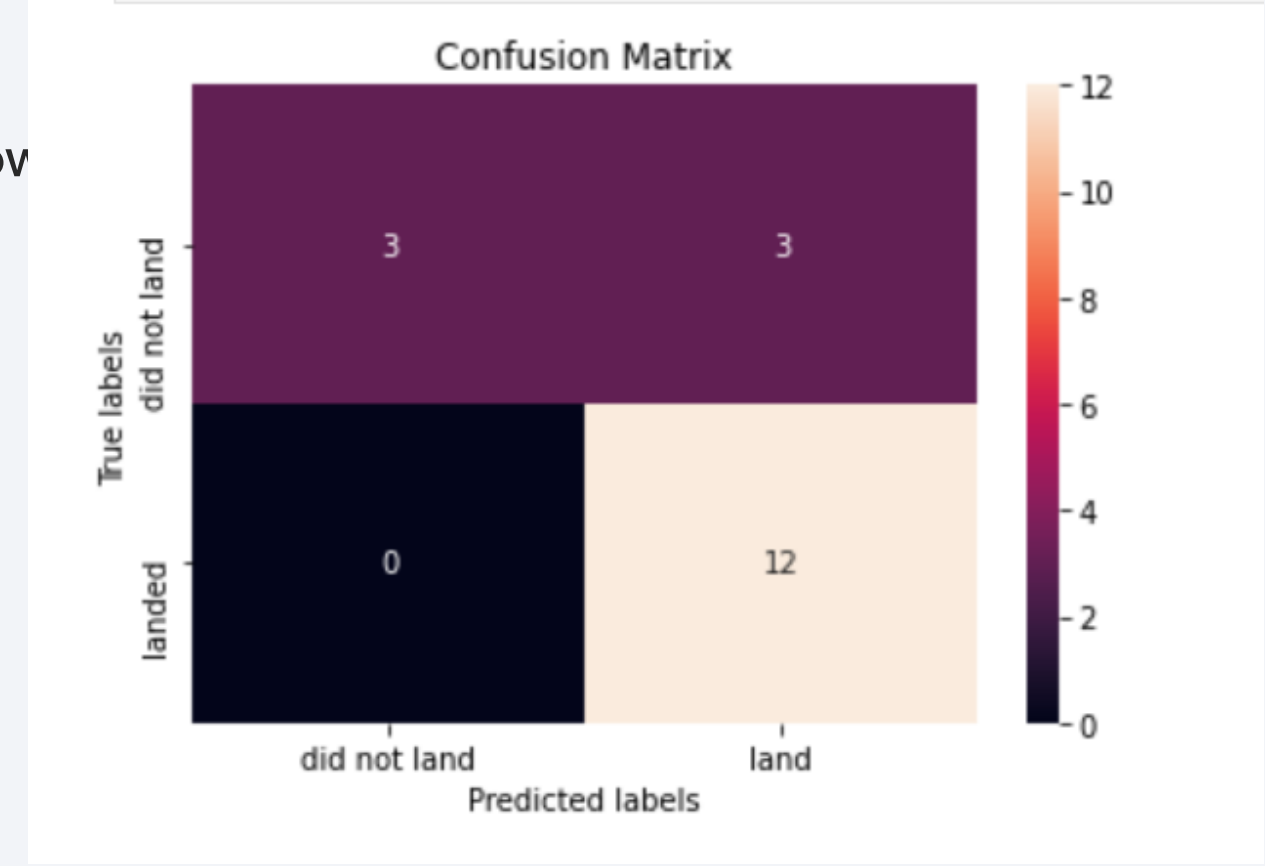


Section 5

# Predictive Analysis (Classification)

# Confusion Matrix

Best accuracy shov



# Conclusions

- Launch success rate started to increase after 2013 till 2020.
- Orbits ES-L1, GEO, HEO, VLEO, SSO had the most success rate.
- KSC LC-39A had the most successful launches of any launch site.
- The decision tree classifier is the best machine learning algorithm for this task.

# Thank you!