Winning Space Race with Data Science

<Andrea Bastari>
<2024-05-27>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

PROJECT BACKGROUND AND CONTEXT

PROBLEMS YOU WANT TO FIND ANSWERS

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

We used Rest API and web scraping to collection data for analysis

Perform data wrangling

We used method:

.fillna and .value_counts

Perform exploratory data analysis (EDA) using visualization and SQL

We used:

data visualization techniques, considering the following factors: payload, launch site, flight number and yearly trend

Perform interactive visual analytics using Folium and Plotly Dash

we used a map and marker to lock for a best site launch

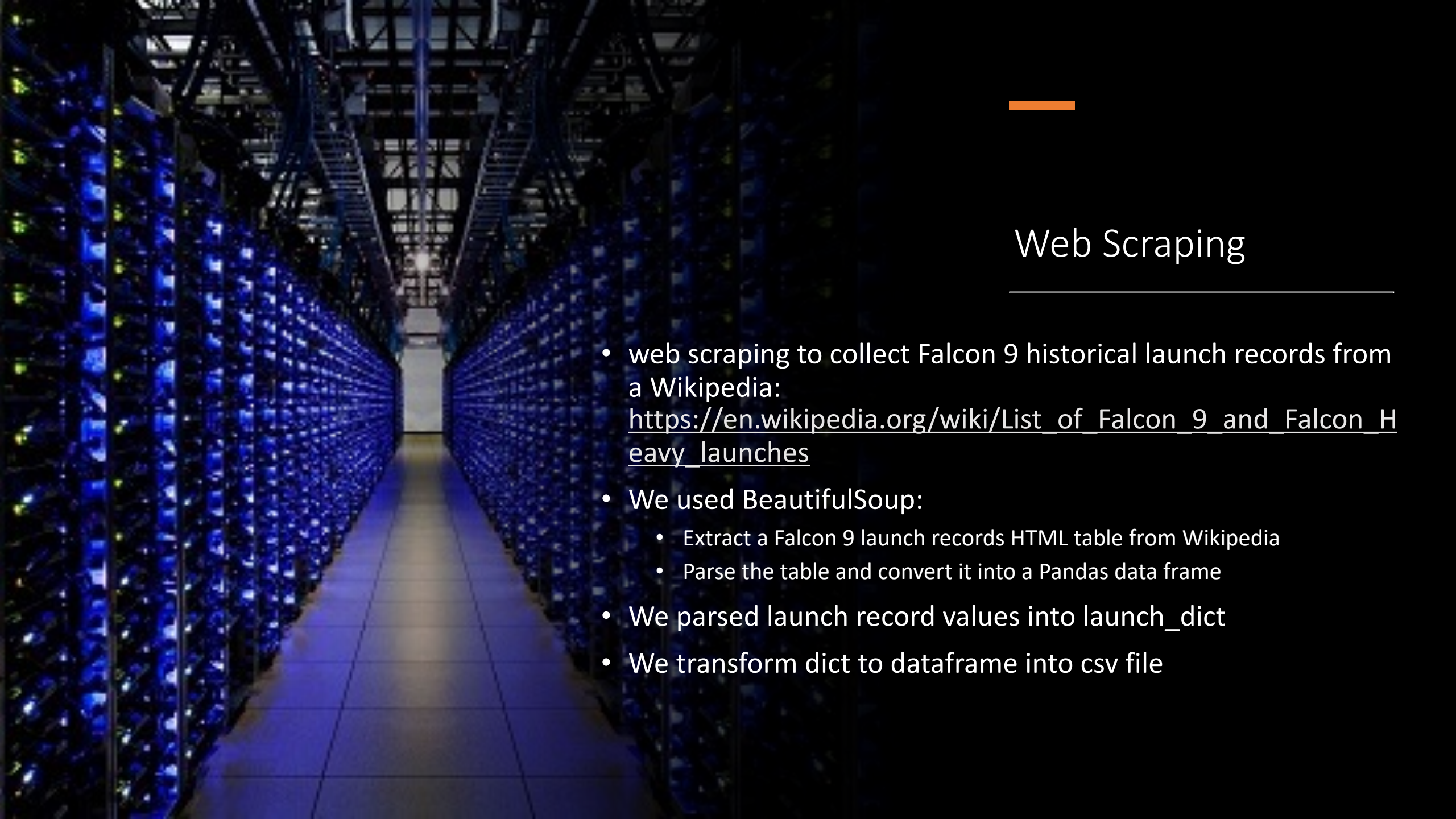
Perform predictive analysis using classification models

We predicted the outcomes of the various landings ultilizing:

- logistic regression;
- SVM;
- the decision tree;
- KNN.

Data Collection

- Request and parse the SpaceX launch data using the GET request from Space X API.
- Normalize json with .json_normalize
- Filter the dataframe to only include Falcon 9 launches
- Replace missing values of payload mass with calculated .mean()
- Export dataframe to csv file



Web Scraping

- web scraping to collect Falcon 9 historical launch records from a Wikipedia: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- We used BeautifulSoup:
  - Extract a Falcon 9 launch records HTML table from Wikipedia
  - Parse the table and convert it into a Pandas data frame
- We parsed launch record values into launch_dict
- We transform dict to dataframe into csv file

Data Collection – SpaceX API

- GITHUB LINK: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/108512667f832ade8f68cc9142a41d37867048e2/jupyter-labs-spacex-data-collection-api.ipynb

1) Requesting rocket launch data from SpaceX A
2) Normalize the response with .json_normalize()
3) creating a data dict and transform in dataframe
4) Filtered data for Falcon9 launches
5) Export data to csv file

Data Collection – Web Scraping

- GITHUB LINK: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/108512667f832ade8f68cc9142a41d37867048e2/jupyter-labs-webscraping.ipynb

1) Take launch data of Falcon 9 from Wikipedia
2) Use BeautifulSoup
3) Extracting name of all column from Table
4) Creating data into dict
5) Creating dataframe from dict
6) Export data to csv file

Data Wrangling

we will perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

1) Calculate numbers of launches from each site
2) Calculate number and occurrence of each orbit
3) Create a landing Outcome Label from colum: Outcome
4) Exportig dataset to csv file

- GITHUB LINK: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/108512667f832ade8f68cc9142a41d37867048e2/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Flight number Vs Payload Mass
- Number Flight Vs Launch site
- Payload vs Launch Site
- bar chart for the sucess rate of each orbit

Github link: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/108512667f832ade8f68cc9142a41d37867048e2/edadataviz.ipynb

EDA with SQL

- Query:
- 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: 45596
- Display average payload mass carried by booster version F9 v1.1: 2928.4
- List date of successfull fist launch
- Total number of successful and failur mission
- More query is in the Jupiter notebook, please visit github link
- Github link: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/7b45b9c8404d1da29216803524 5f73170b4d68cc/jupyter-labs-eda-sql-coursera_sqllite_.ipynb

Build an Interactive Map with Folium

- Target 1: Mark all launch sites on a map
- Target 2: Mark the success/failed launches for each site on the map
- Target 3: Calculate the distances between a launch site to its proximities

- Github link: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/7b45b9c8404d1da292168035245f73170b4d68cc/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Add a Launch Site Drop-down Input Component
- Add a callback function to render success-pie-chart based on selected site dropdown
- Add a Range Slider to Select Payload
- Add a callback function to render the success-payload-scatter-chart scatter plot

- Github link: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/7b45b9c8404d1da292168035245f73170b4d68cc/spacex_dash_app.py

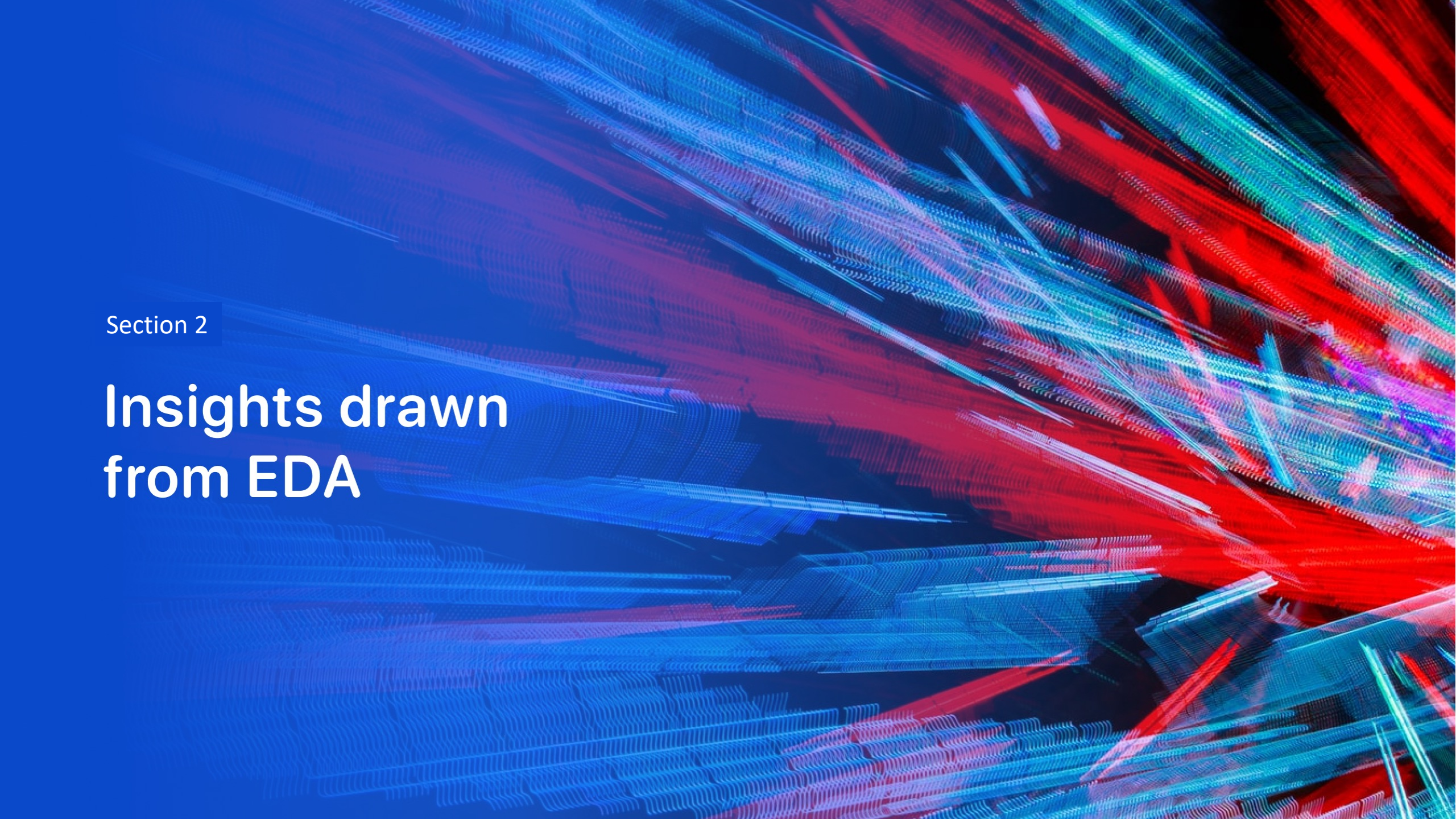
Predictive Analysis (Classification)

- Create a column for the class
- Standardize the data
- Split into training data and test data
- -Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

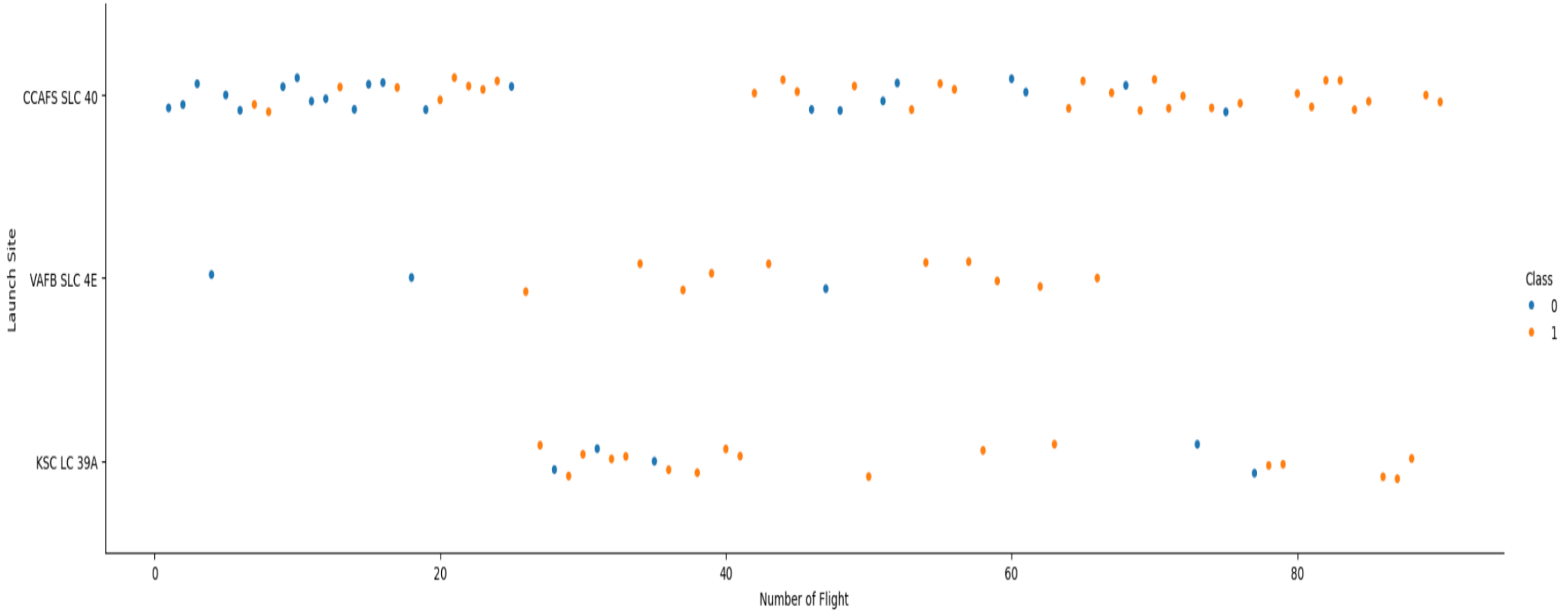
Github link: https://github.com/abastariacsi/Data-Science-Capstone-IBM/blob/7b45b9c8404d1da292168035245f73170b4d68cc/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

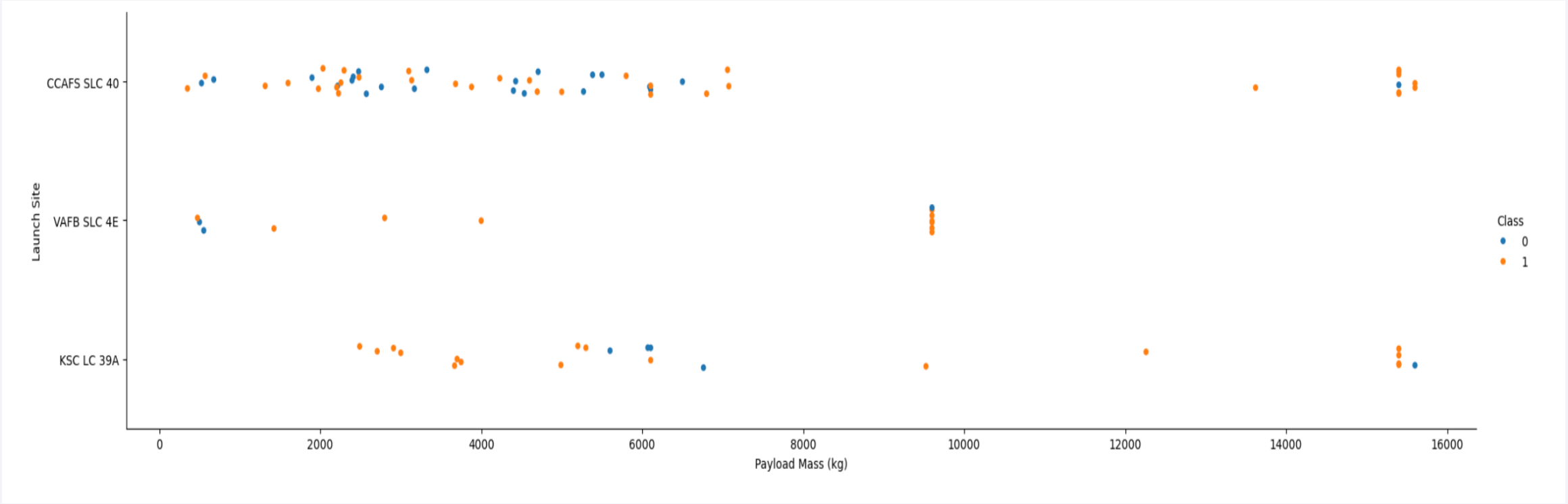
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

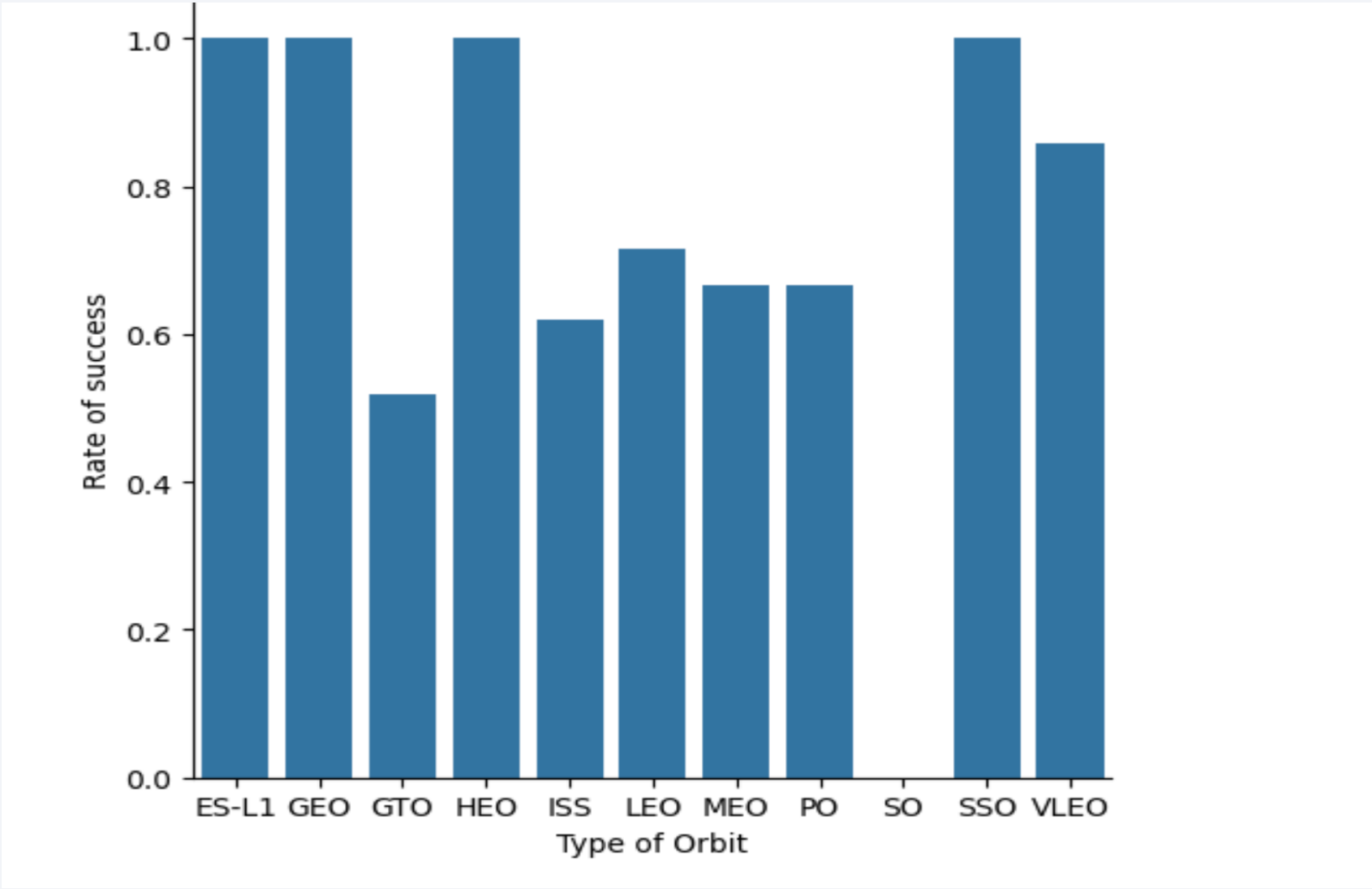
Flight Number vs. Launch Site



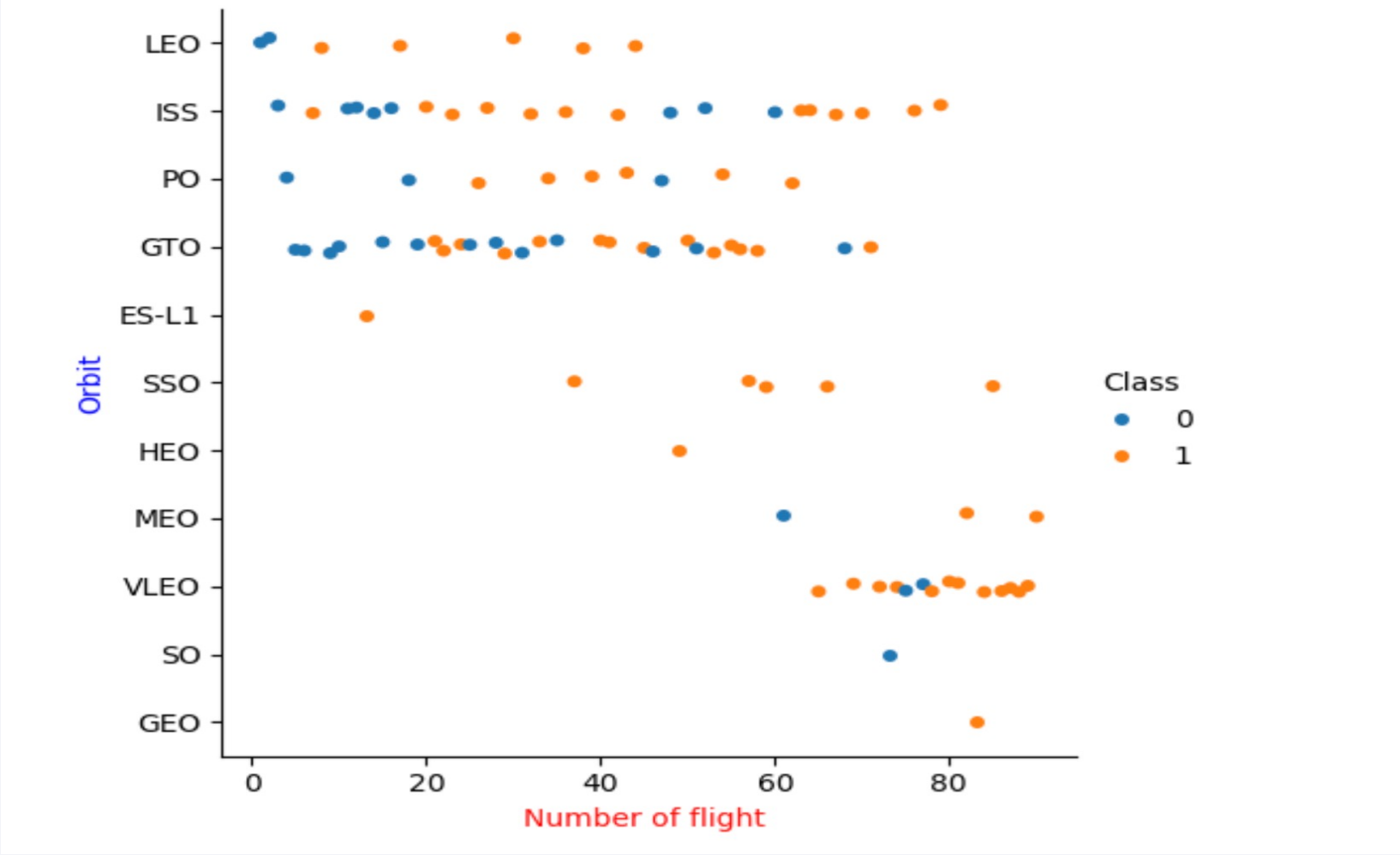
Payload vs. Launch Site



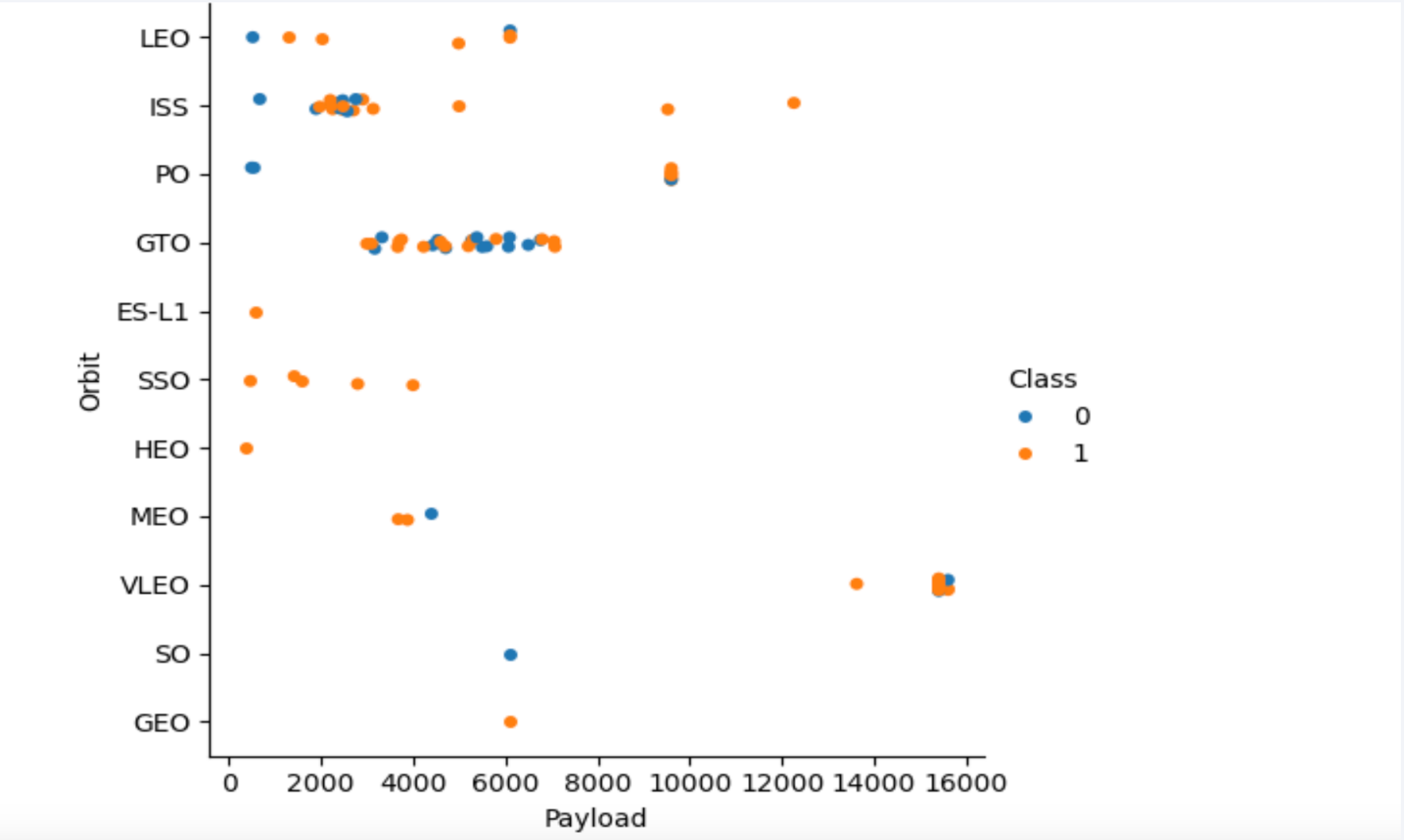
Success Rate vs. Orbit Type



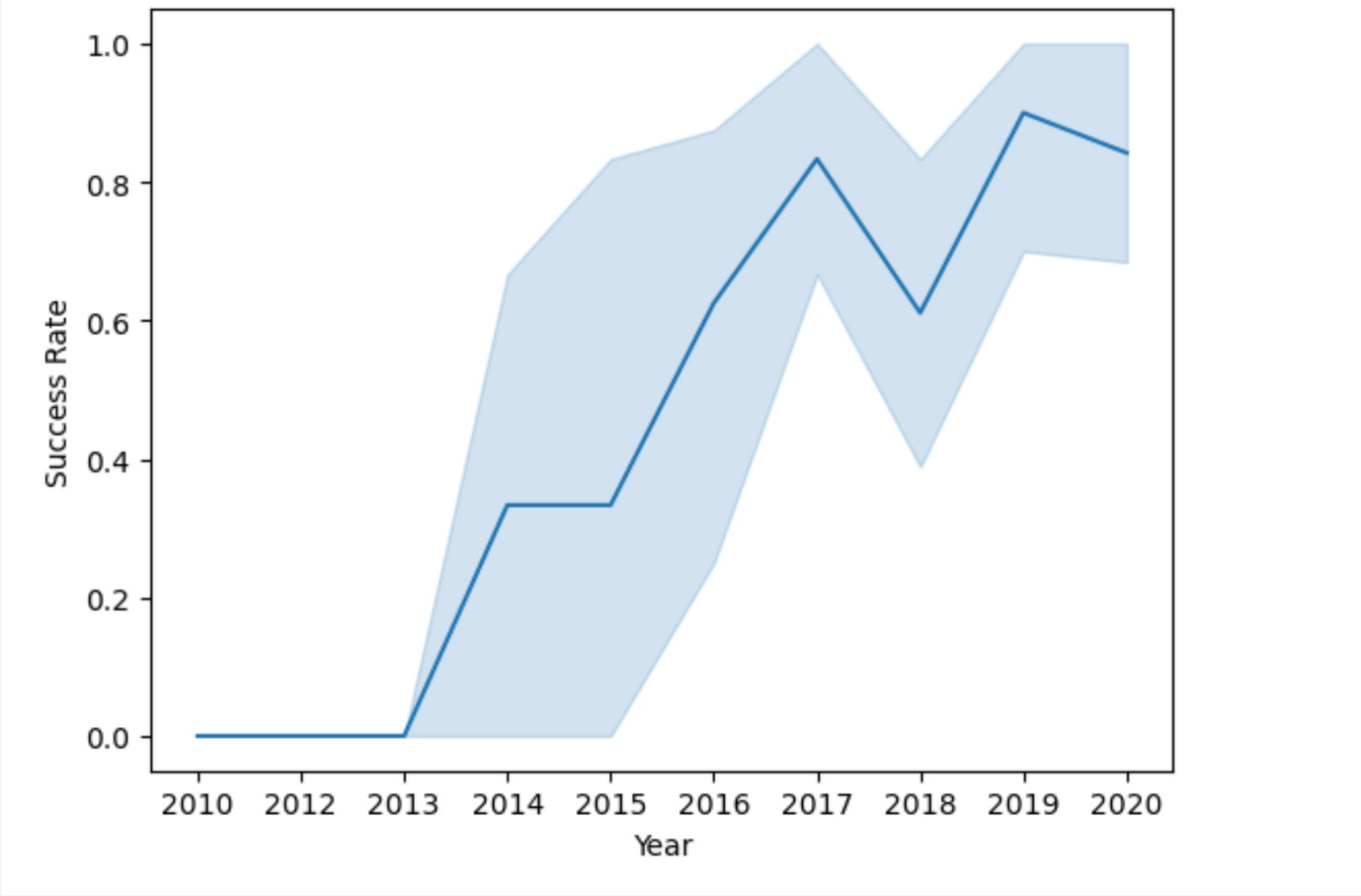
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Out[23]:

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

Launch Site Names Begin with 'CCA'

Out[10]:

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

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

1]:

TOTAL_PAYLOAD_MASS
45596

Average Payload Mass by F9 v1.1

AVERAGE_PAYLOAD_MASS
2928.4

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

Out[26]:

first_successful_landing
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

Out[29]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your

Out[15]:

Mission_Outcome	TOTAL_NUMBER
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Out[16]:

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

Out[30]:

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

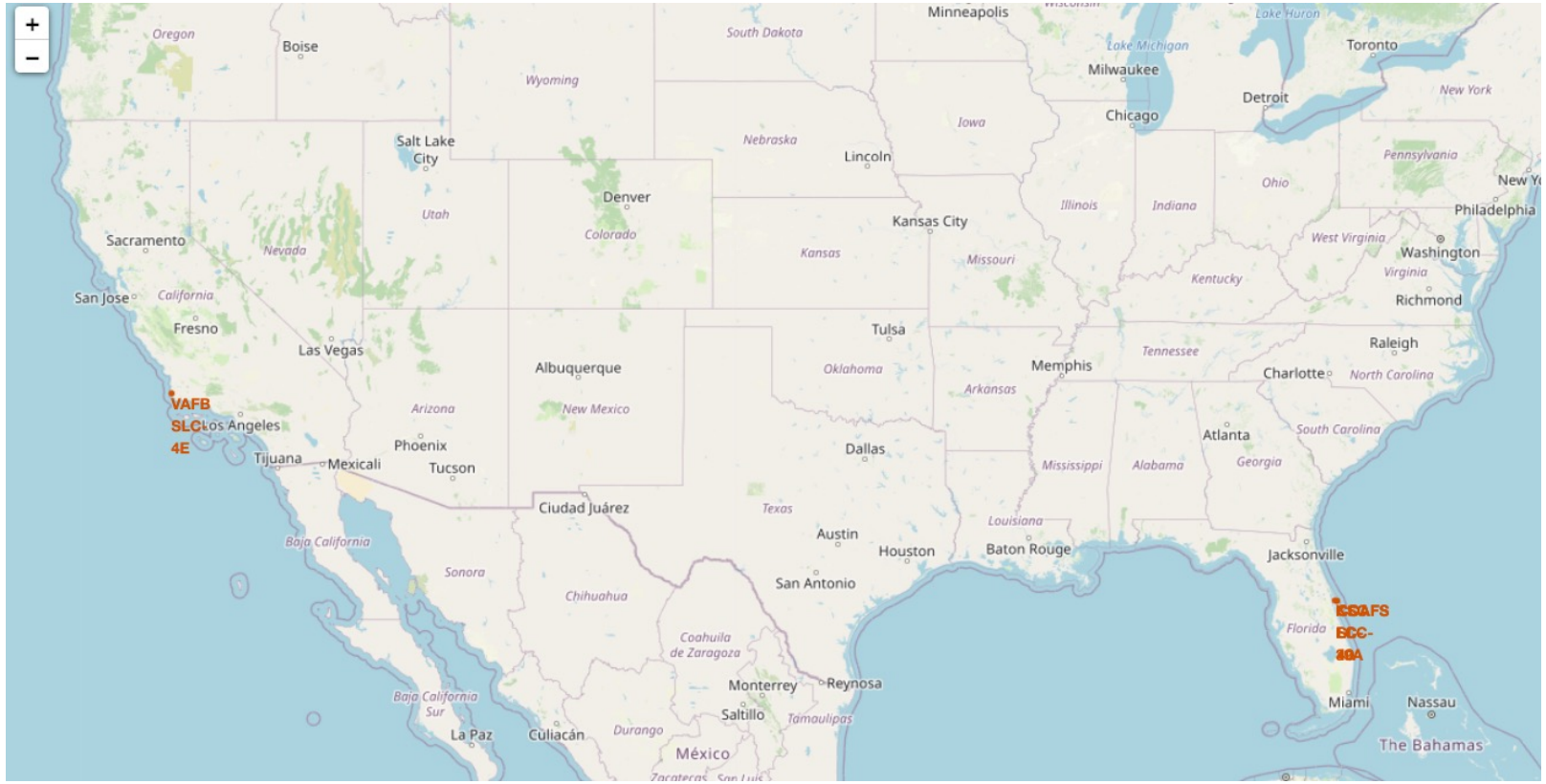
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Out[31]:

Landing_Outcome	qty
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

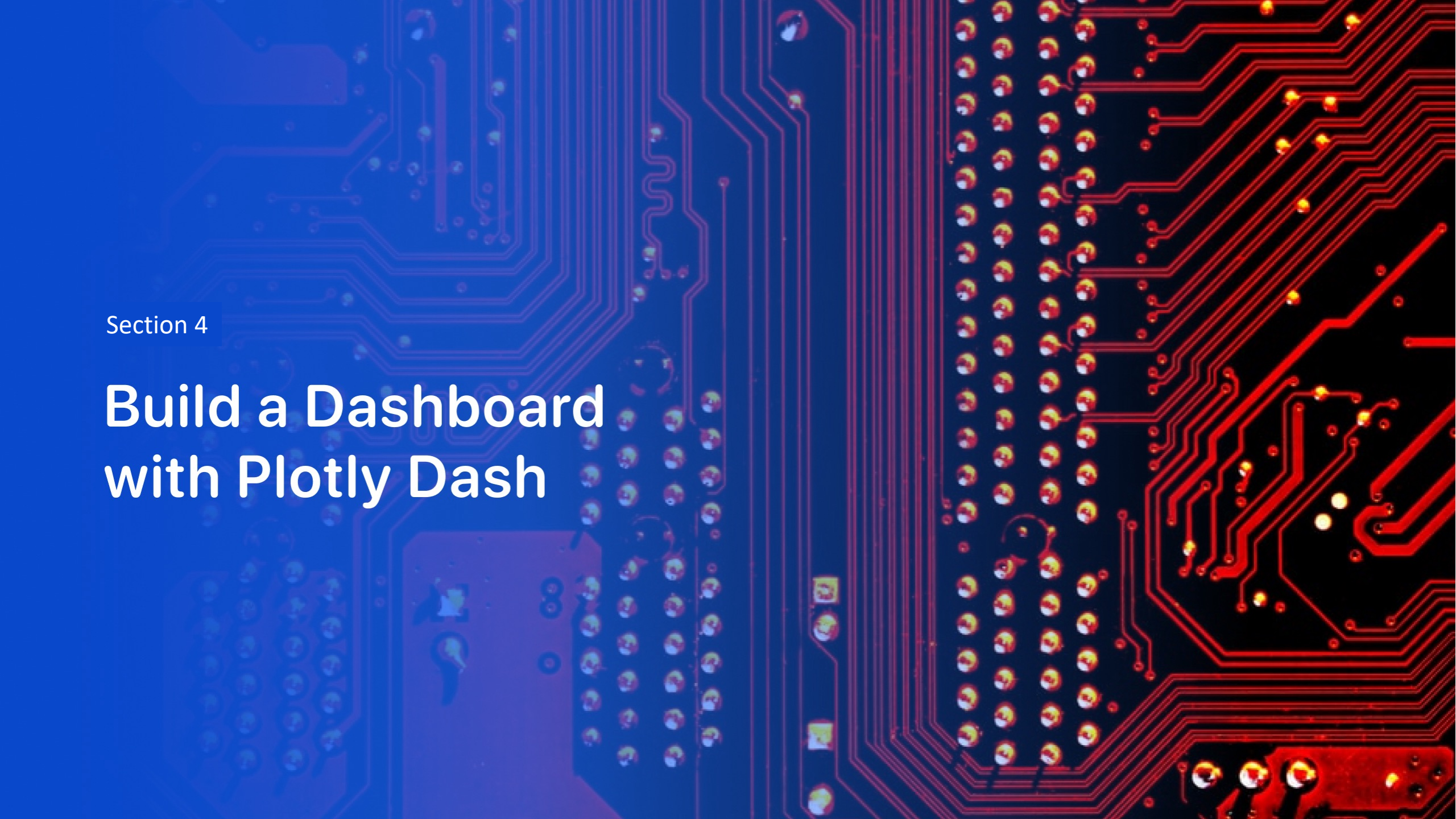

Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

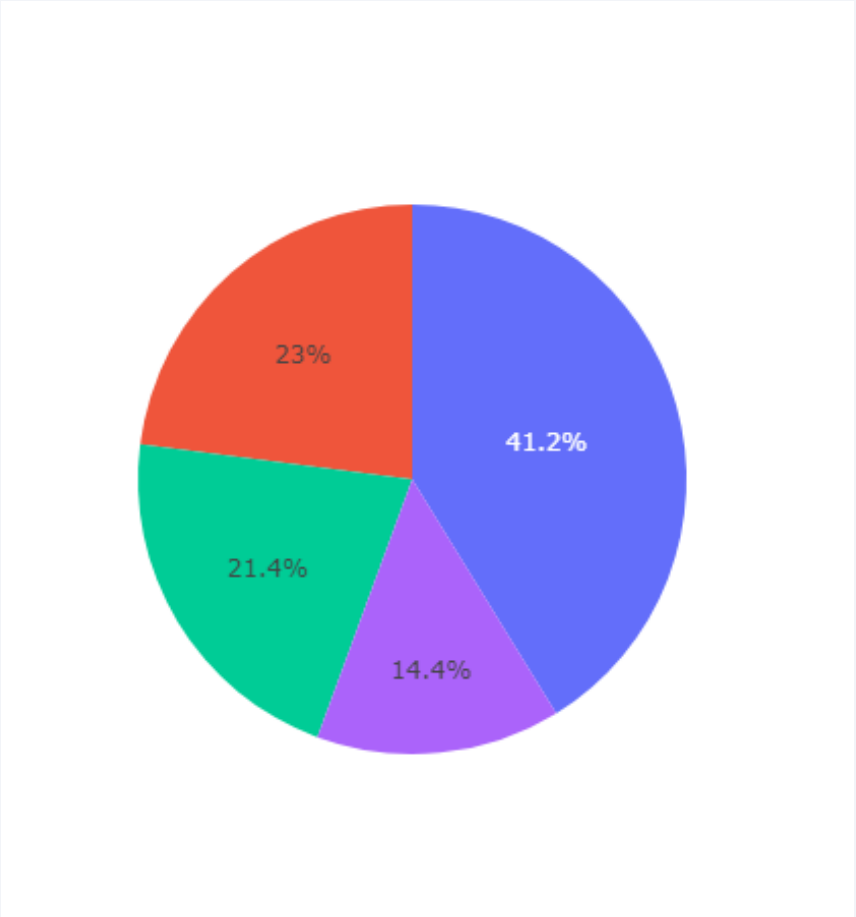


<Folium Map Screenshot 2>

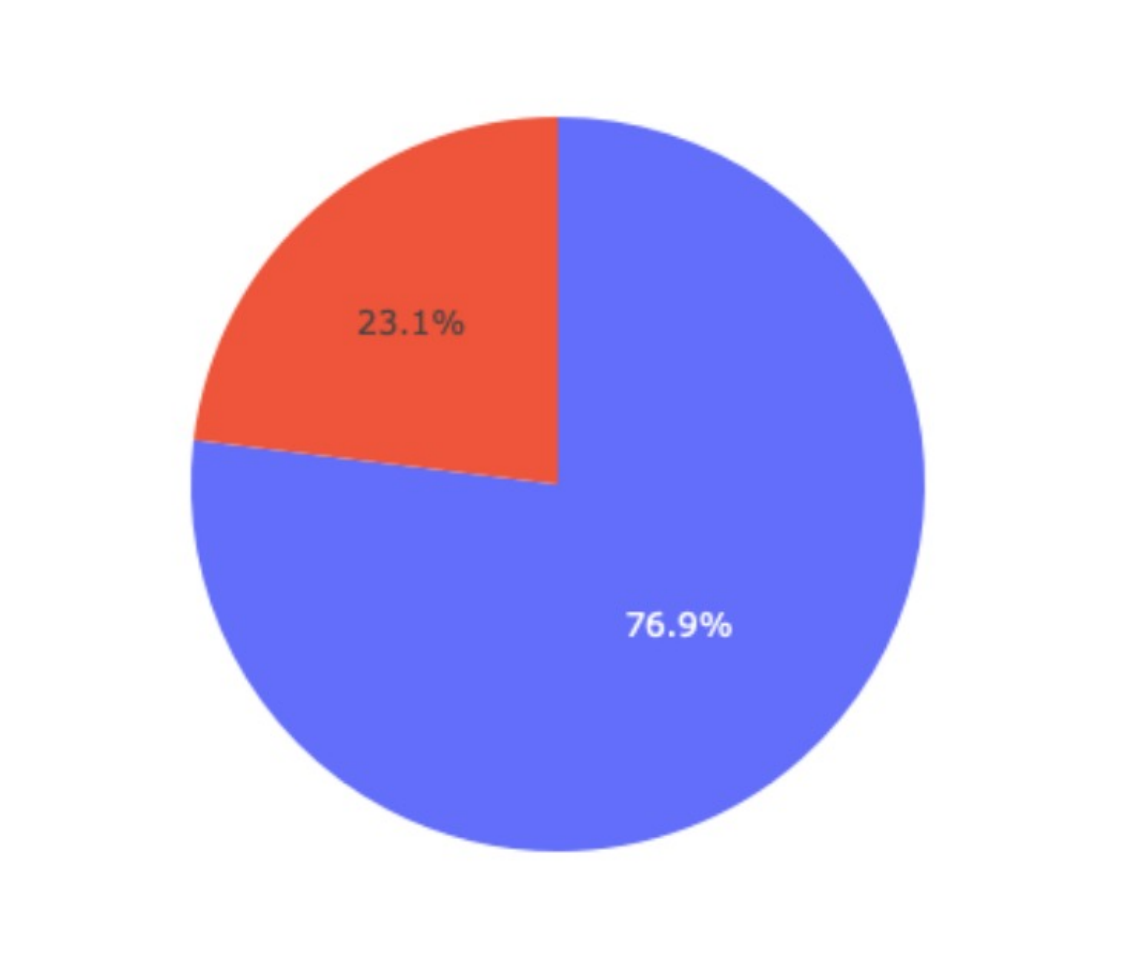
<Folium Map Screenshot 3>

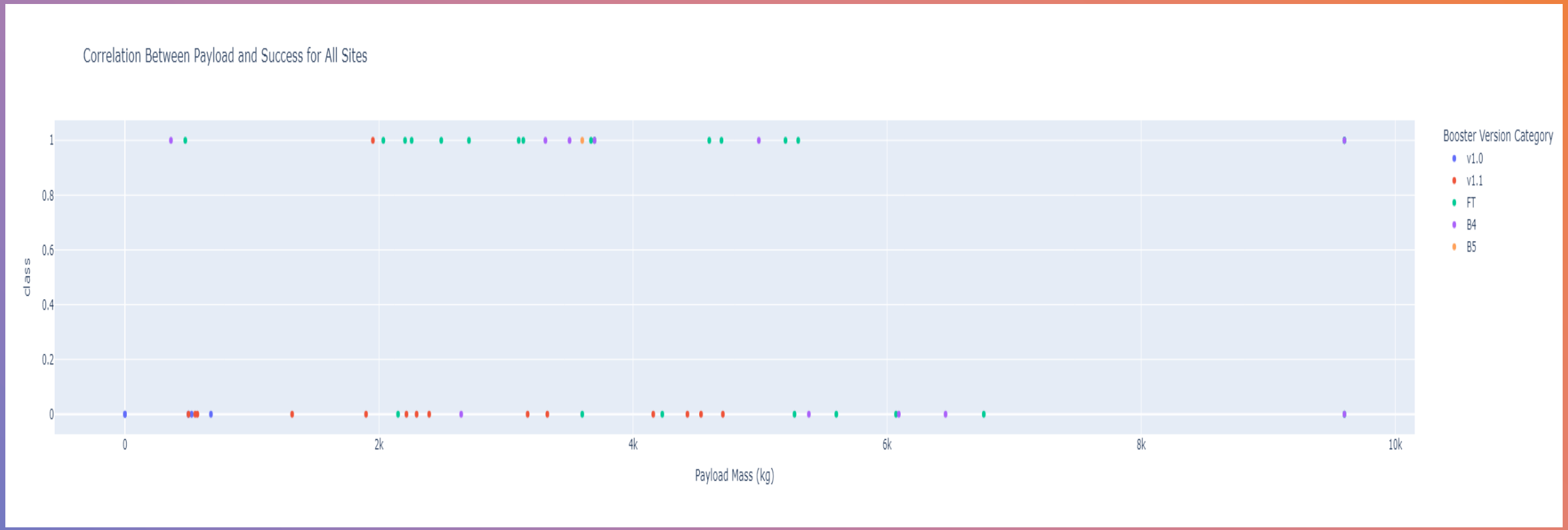

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all sites



KSC LC-39° - highest launch success ratio



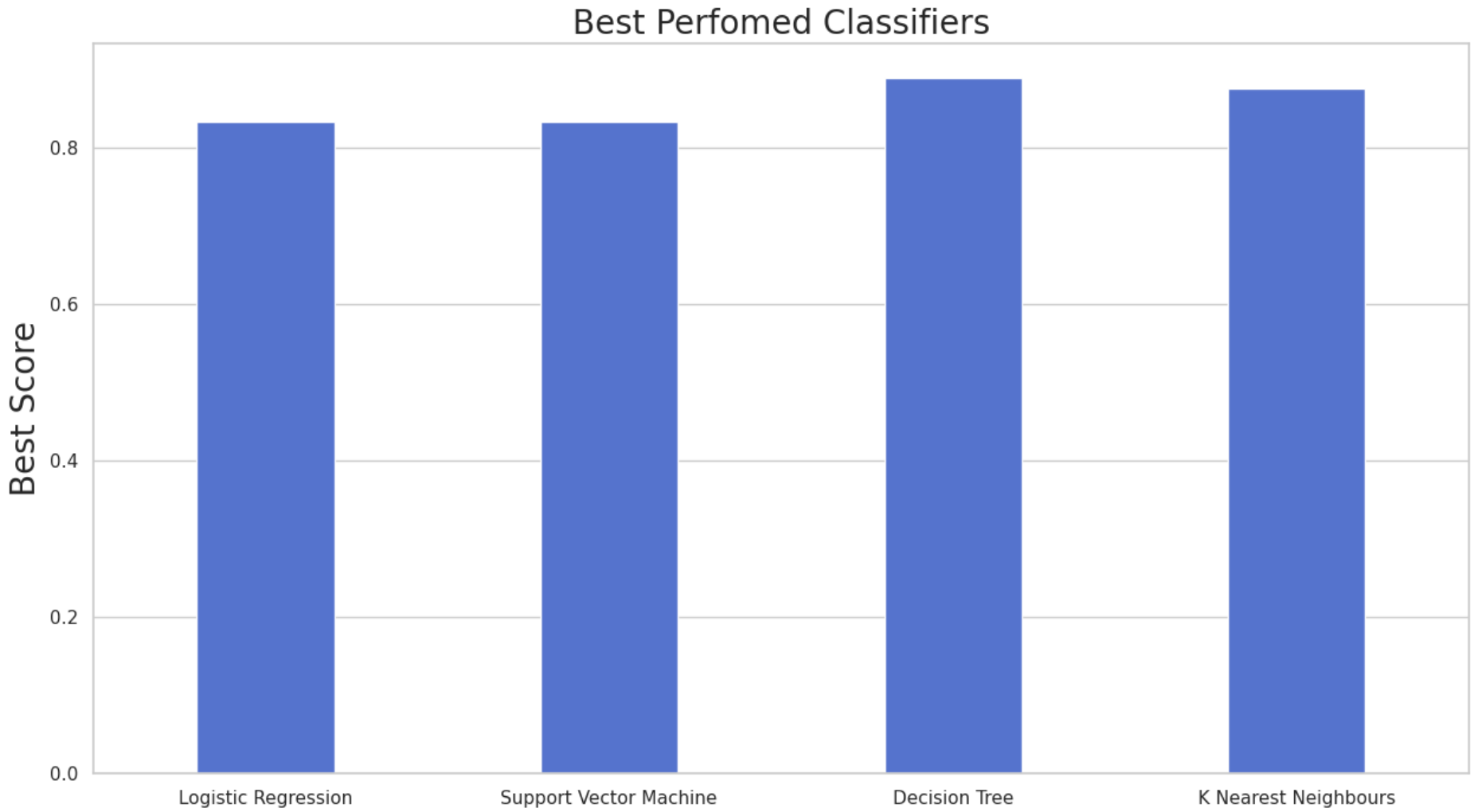


Payload Mass vs Class

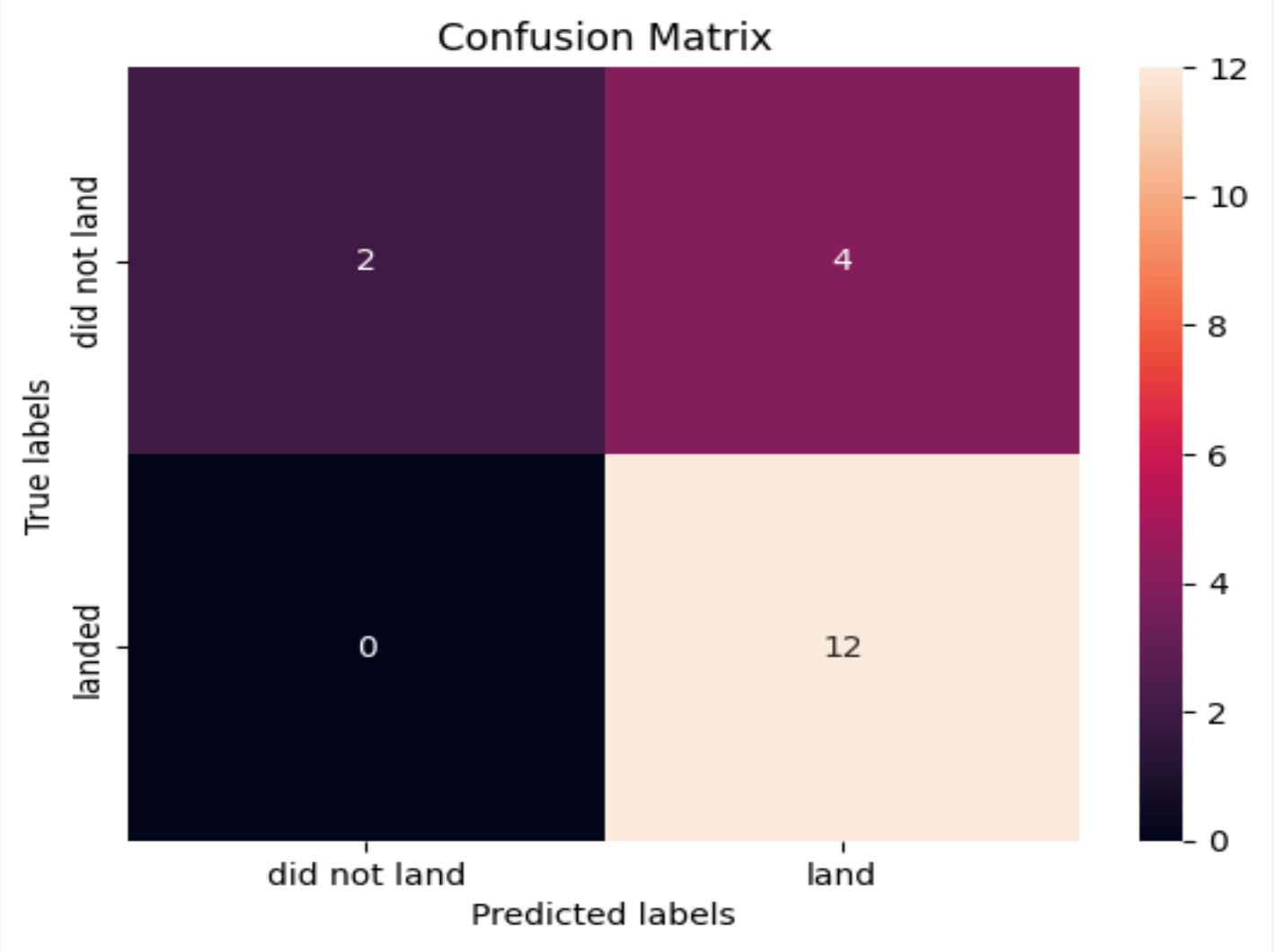
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Thank you!