IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Pauline Mutemeri>
<02 July 2023>

Outline

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

Executive Summary

- Summary of methodologies

Data was collected using Rest API and a Get request was used and output was a Json which was converted to a dataframe using pd.json.normalize()

BeautifulSoup was also used for webscrapping into a html data set, there after turned into a dataframe for further analysis

Exploratory data analysis (EDA) using visualization and SQL

Visual analytics using Folium and Plotly Dash

- Summary of all results

90 Falcon launches were observed

26 Missing Values were obtained in Landingpad.

37.5% success launches

Introduction

- Project background and context

Space X would like to analyse and determine the cost of a launch by observing Falcon 9 first stage landing. By understanding the first stage landing, a cost analysis can be recommended.

- Problems you want to find answers

The Void to be filled is to find out the success rate of Falcon 9 first stage landing

Cost implication for both failed and successfully based on the launchsite

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The Space X Rest API was used to give data about the Launches. A get request was implemented using the request library and the results were viewed by calling an API. The output was a Json which was converted to a dataframe using pd.json.normalize()
- Perform data wrangling
  - Percentage of missing values of the total values was obtained is see if the data is in a usable state. 26 LandingPad were found to be missing and had Null values
  - BeautifulSoup and Data Parsing
  - Data types were also observed and
  - Value counts were done on Launchsite, Obits to see the number of launches were done per site
  - Bad Outcomes were also calculated using the set and a if and else statement was used there

Data Collection

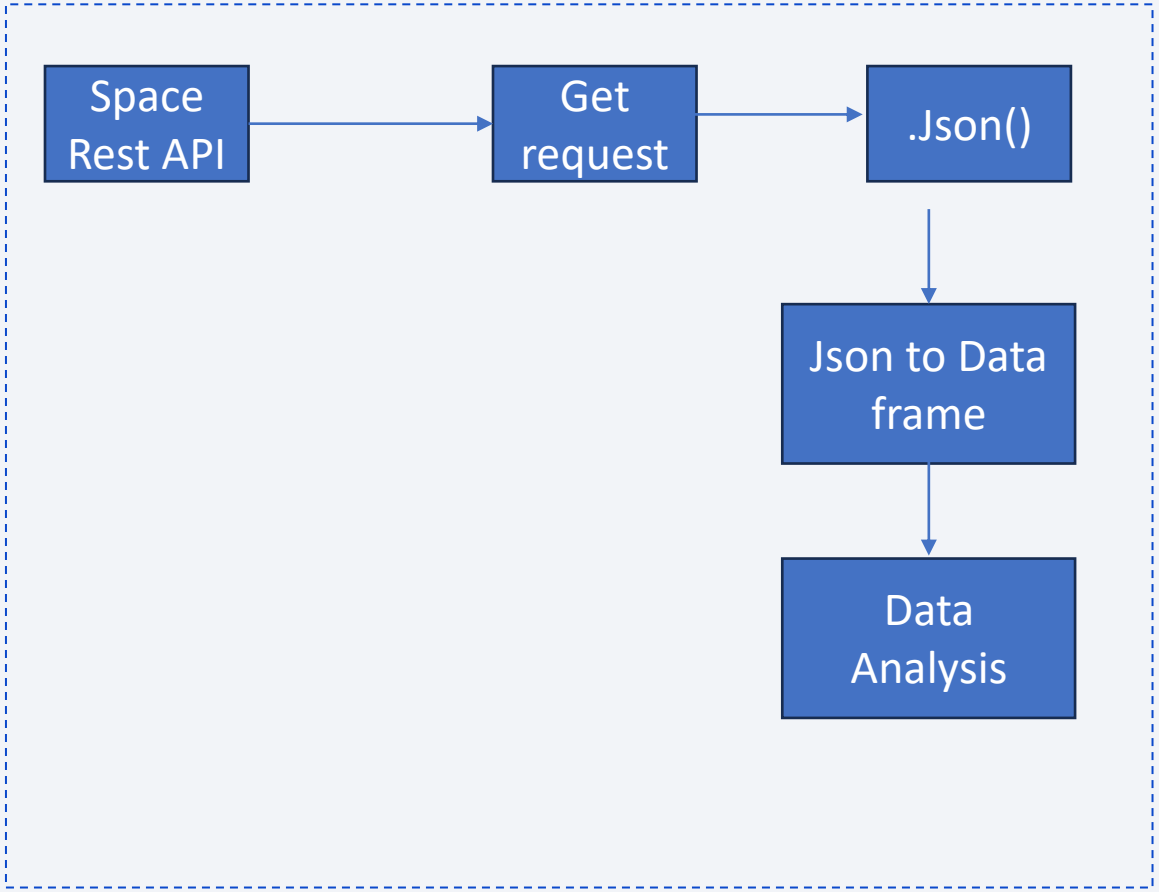
- Describe how data sets were collected.

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

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

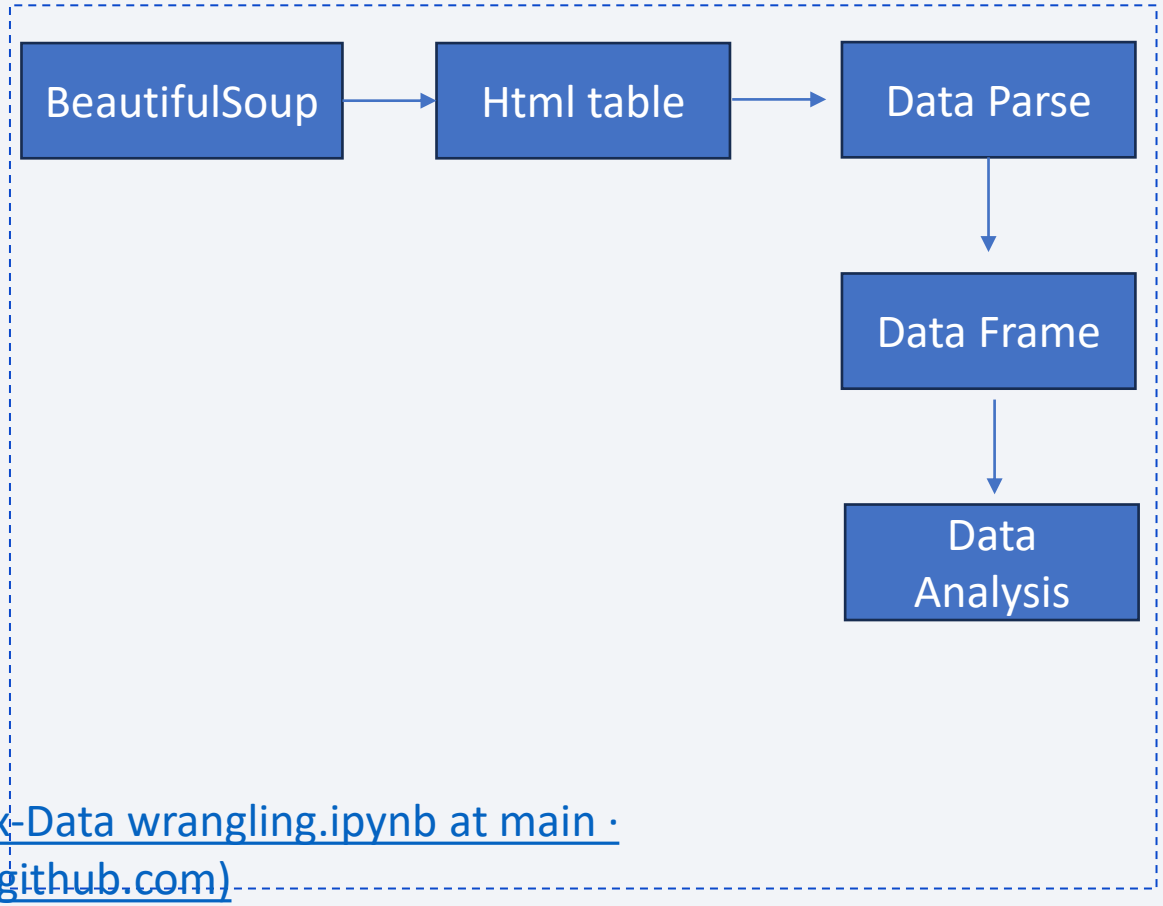
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



testrepoo/labs-jupyter-spacex-Data wrangling.ipynb at main · mutemeripauline/testrepoo (github.com)

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts



- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

testrepoo/labs-jupyter-spacex-Data wrangling.ipynb at main · mutemeripauline/testrepoo (github.com)

Data Wrangling

- Describe how data were processed
- Percentage of missing values of the total values was obtained is see if the data is in a usable state. 26 LandingPad were found to be missing and had Null values
- BeautifulSoup and Data Parsing
- Data types were also observed and
- Value counts were done on Launchsite, Obits to see the number of launches were done per site
- Bad Outcomes were also calculated using the set and a if and else statement was used there

- You need to present your data wrangling process using key phrases and flowcharts

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- A Seaborn Catplot -This plot was used as it provides access to several axes-level functions that show the relationship between Flight Number and PayloadMass categorical variable.
- A Seaborn Catplot -This plot was used as it provides access to several axes-level functions that show the relationship between LaunchSite and PayloadMass categorical variable.
- A Seaborn Catplot -This plot was used as it provides access to several axes-level functions that show the relationship between Flight Number and LaunchSite categorical variable.

- testrepoo/jupyter-labs-eda-dataviz.ipynb at main · mutemeripauline/testrepoo (github.com)

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- SELECT DISTINCT Launch_Site FROM SPACEXTBL
- SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE ('CCA%') LIMIT 5
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer IN ('NASA (CRS)')
- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version IN ('F9 v1.1')
- SELECT DISTINCT Mission_Outcome , COUNT(*) FROM SPACEXTBL GROUP BY Mission_Outcome
- SELECT MIN (Date) FROM SPACEXTBL where Mission_Outcome = 'Success'
- SELECT DISTINCT FROM SPACEXTBL WHERE Landing_Outcome IN ('Success

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model

Finding a relationship between payload and Orbit

Finding a relationship between payload and Launchsite

Finding a relationship between payload and flight number

Feature Engineering get_dummies and assigning the Feature One Hot encoding

- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose
- testrepoo/jupyter-labs-eda-dataviz.ipynb at main · mutemeripauline/testrepoo (github.com)

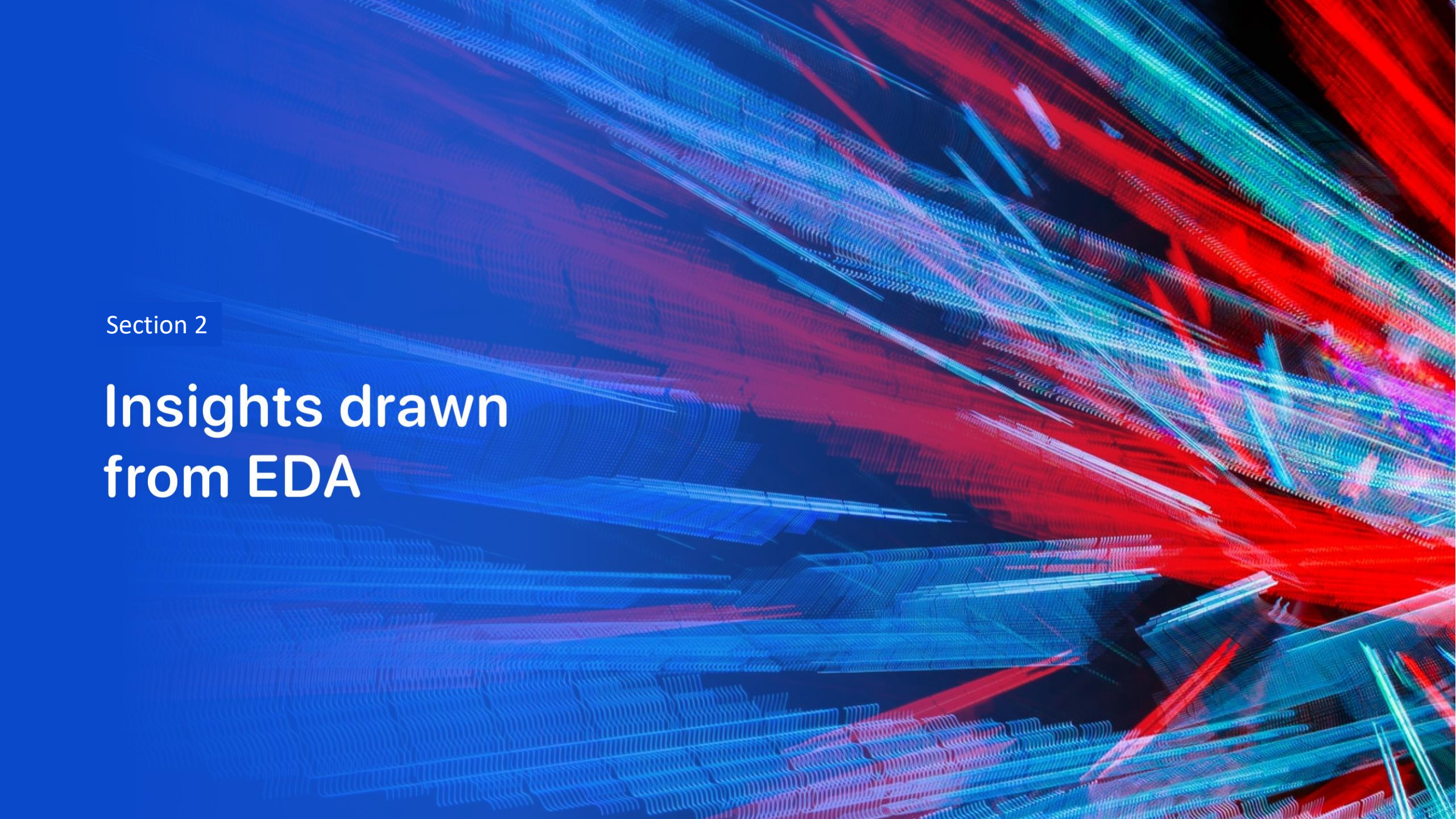
Results

- Exploratory data analysis results

CCAFS SLC 40 success rate if 60% below 10000kg but 100% if more the 10000kg

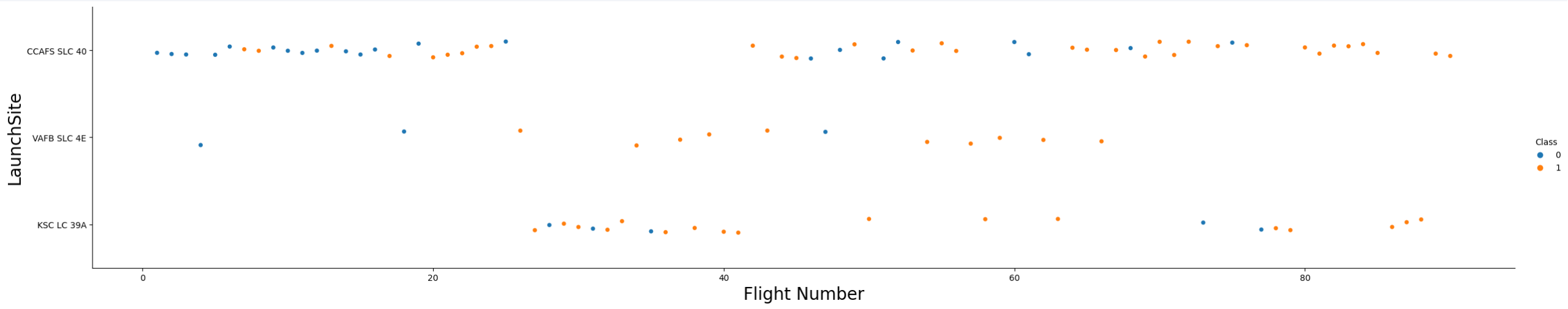
KSC LC 39A success rate of 77% as well as VAFB SLC 4E

- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

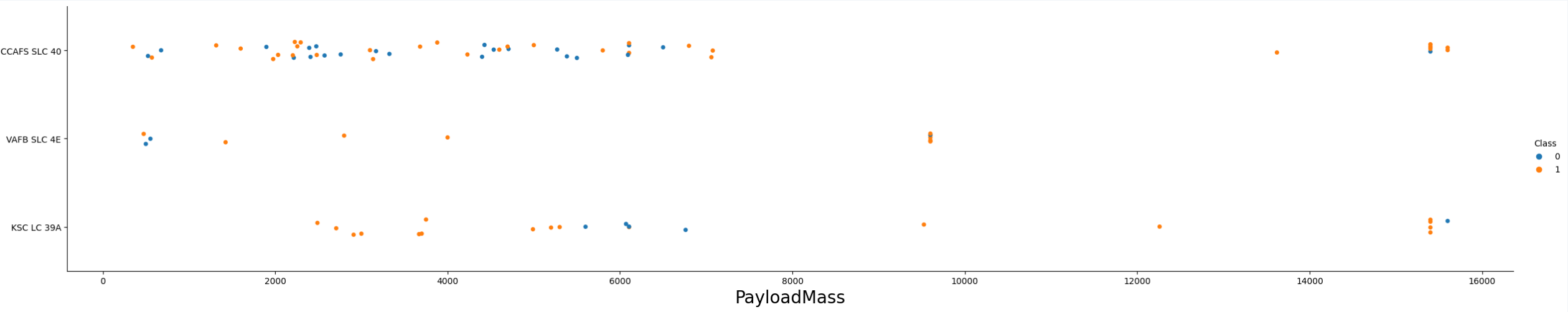
- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



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

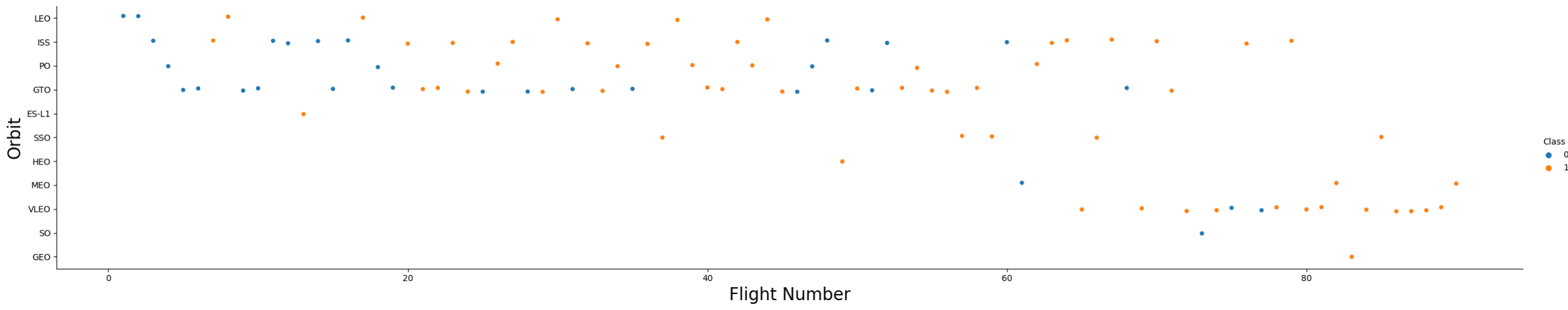
- Show a bar chart for the success rate of each orbit type

- 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

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

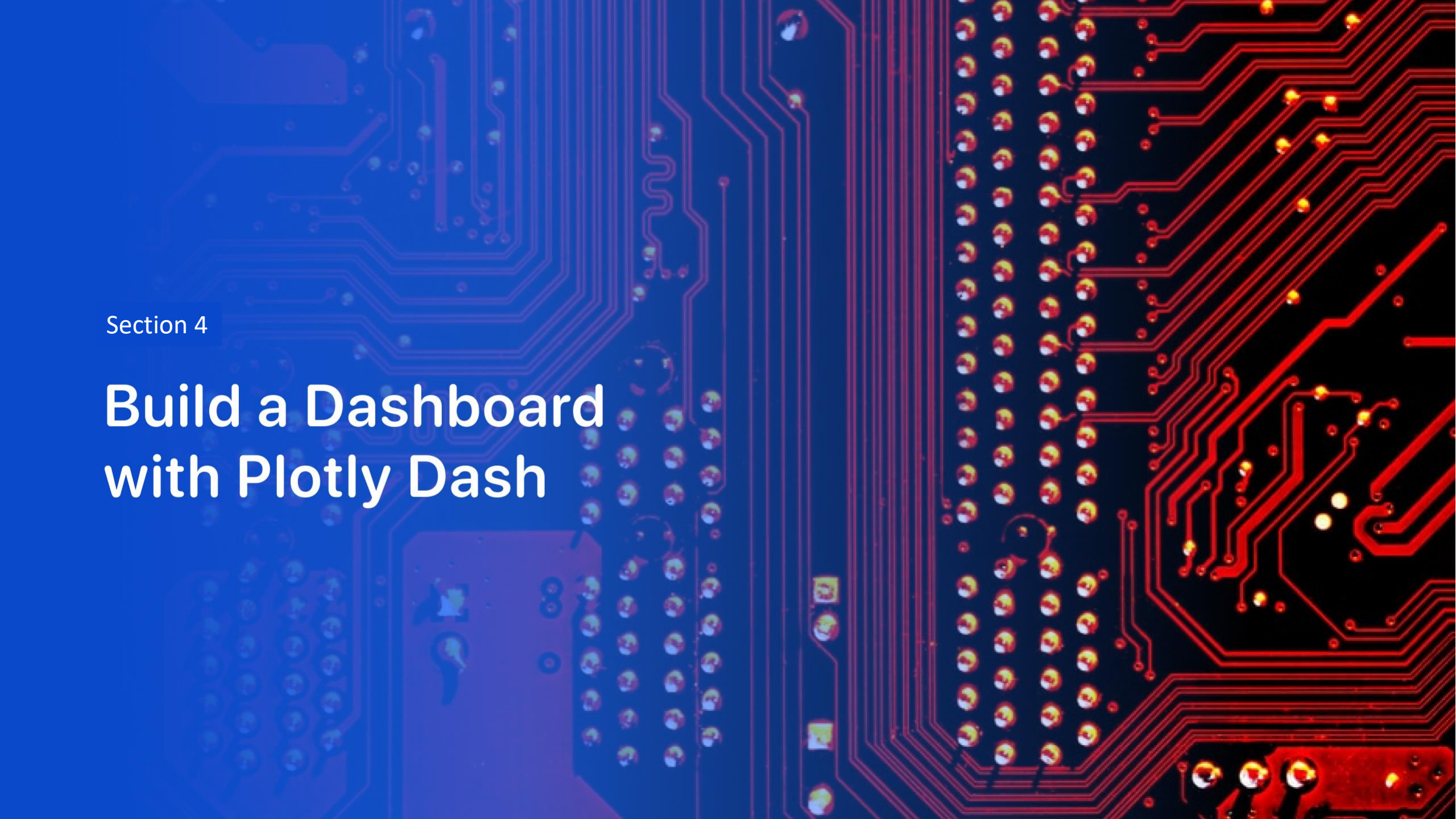
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- CCAFS SLC-40 Should be invested in as is has 57% success of landing based on the SQL data
- VAFB SLC-4E should not be implemented at all as it has the least % of landing outcome

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!