IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

William Letwaba
2024-09-30

Outline

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

Executive Summary

Summary of methodologies

1. We collect data about launch and landing specifications and outcomes, rockets used, payload delivered using Space X REST API
2. We obtain the rocket launches data using web scrabing
3. We process the data using Data Wrangling
4. We select and sort the data using SQL queries on the data loaded into a Database(Db2)
5. We do Exploratory Data analysis by writing Python codes manipulating the data into Panda Dataframes for further analysis and visuals using scatter plots, bar charts, Folium maps(using Folium Library) and Dashboards (uising the Plotly Dash Python library.)

Summary of all results

Jupiter Notebooks that has codes and Powerpoint Presentation

Introduction

- Project background and context

Space X advertises Falcon9 rocket launches on its website with a cost of 62 million dollars, other providers cost upwards costs of 165 million dollars each . The reason Space X can save this much is because they reuse the first stage.Therefore we act as Data Scientist to determine a prediction of whether Falcon 9 will make successful first stage landing so to determine the cost of the launch.

Problems you want to find answers

To determine the price of each launch. Whether SpaceX will reuse the first stage.

And train a machine learning model and use public information to predict if SpaceX

will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX REST API endpoints or URL like api.spacexdata.com/v4/launches/past. through a request library like get request and the result is viewed using JSon method in a list of JSon objects each representing a launch converted to a flat table using JSon Normalize.
  - Also Data collection was done from HTML tables in WIki related pages containing Falcon 9 past launch records(https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches) using web scraping through Python BeautifulSoup packages and parse the data converting them to Pandas Data frames.
- Perform data wrangling
  - Data was processed through Data wrangling using an API, Data Sampling and dealing with Nulls
  - And Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.

Data Collection

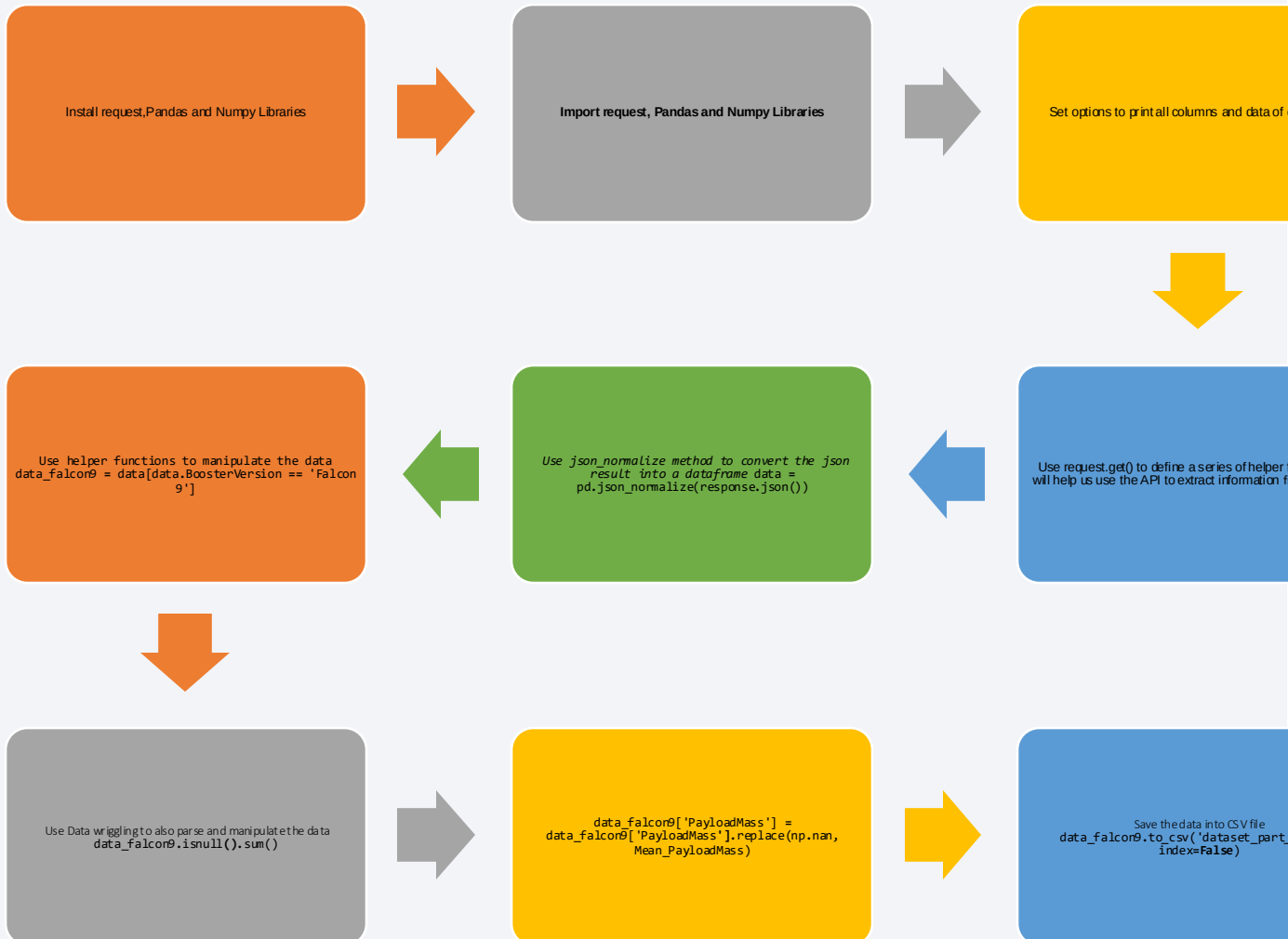
- Data was collected using SpaceX REST API endpoints or URL like api.spacexdata.com/v4/launches/past. through a request library like get request and the result is viewed using JSon method in a list of JSon objects each representing a launch converted to a flat table using JSon Normalize.
- Also Data collection was done from HTML tables in Wlki related pages containing Falcon 9 past launch records(https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches) using web scraping through Python BeautifulSoup packages and parse the data converting them to Pandas Data frames.

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

Data Collection – SpaceX API

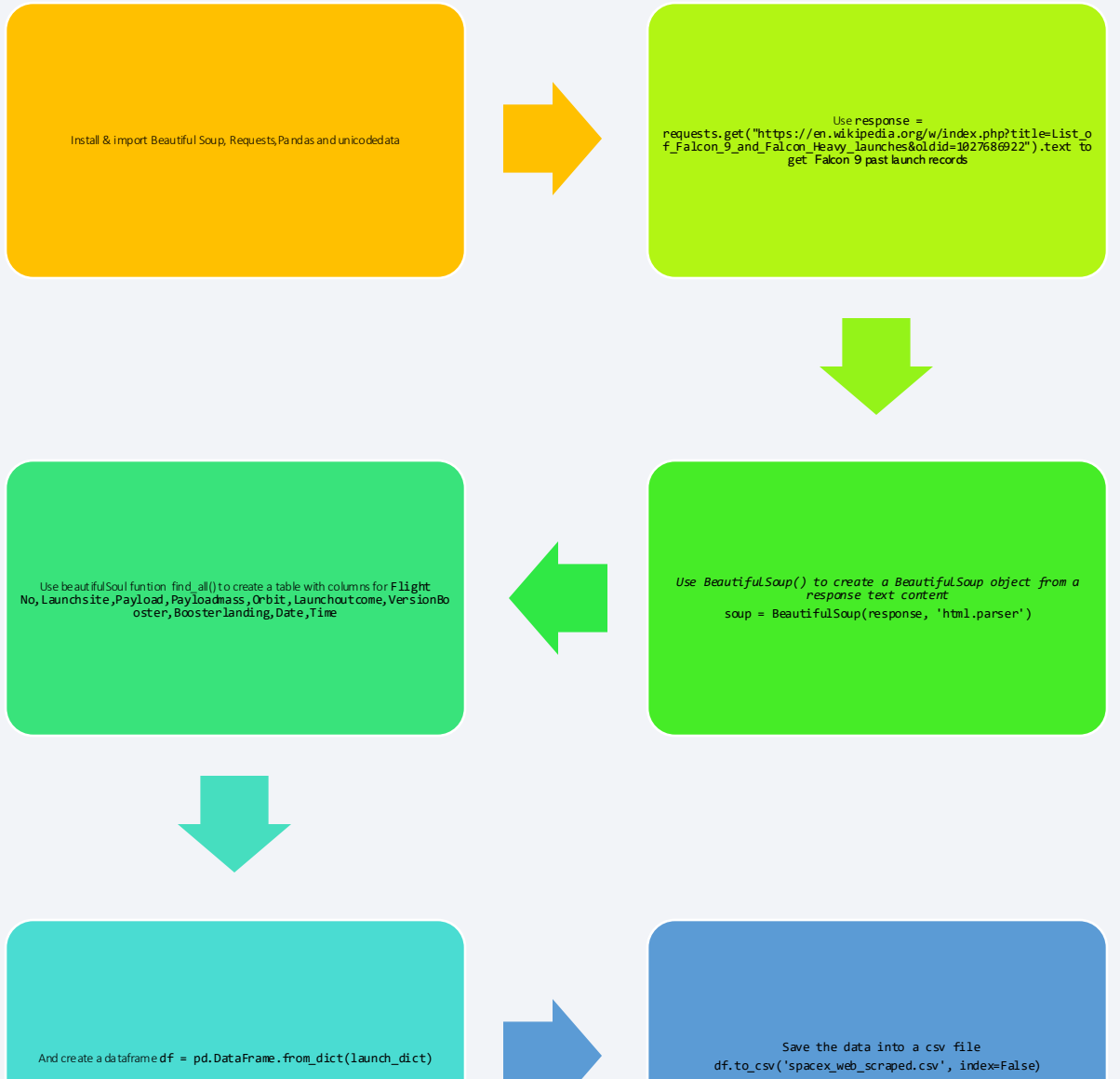
- Data collection with SpaceX REST is in the figure on the right with flowcharts

- Find here the link for the GIThub with the Data collection jupyter notebook https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/William%20Letwaba-jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

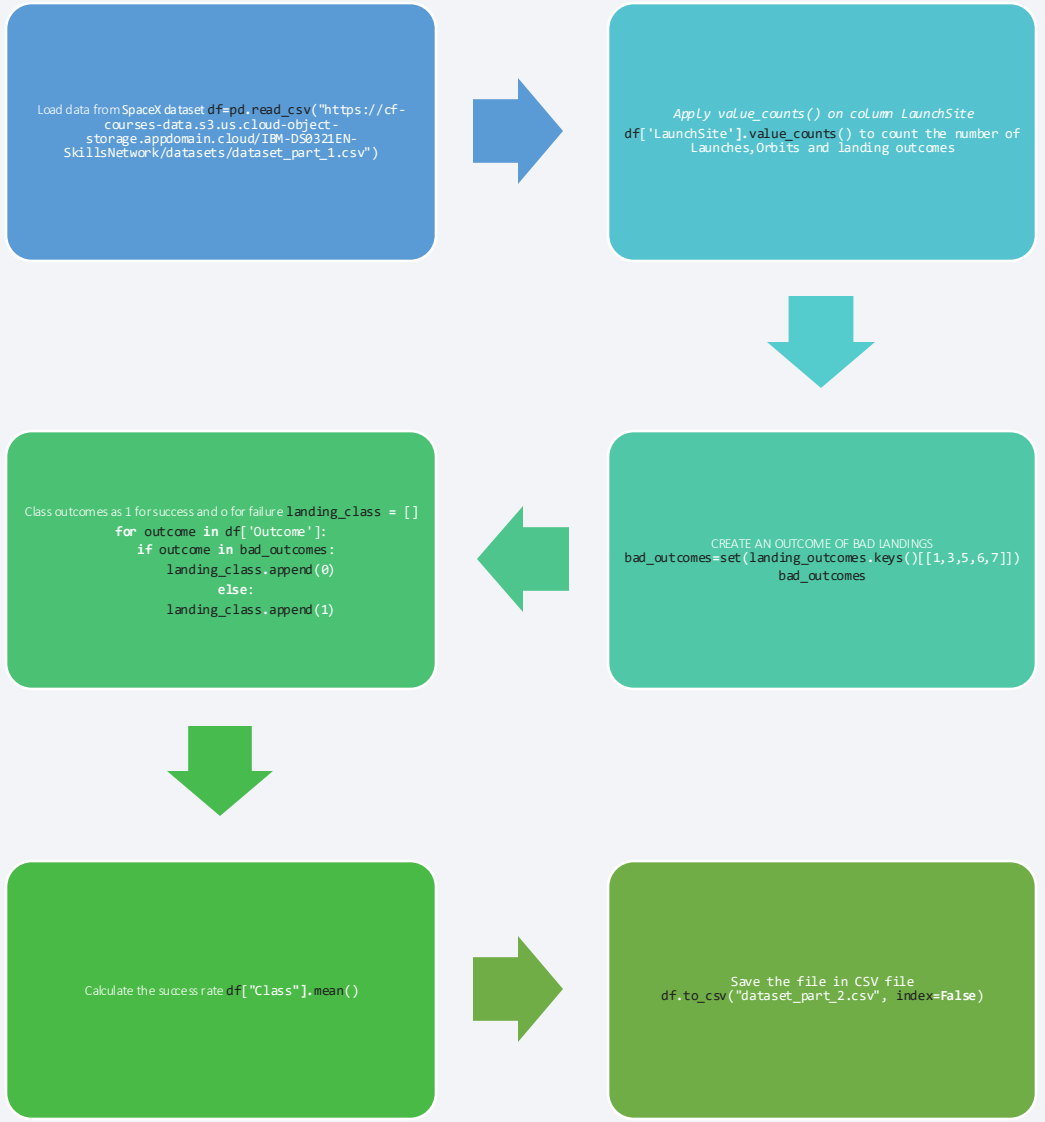
- The figure on the right shows the collection of data from past Falcon 9 rocket past launches through webscraping
- Find here the link for the GIThub with the Data collection jupyter notebook
- https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/WilliamLetwaba-jupyter-labs-webscraping.ipynb



Data Wrangling

- We do EXploratory data analysis on the data

- The figure on the right shows
- Below is the GitHub URL of the completed data wrangling https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/WilliamLetwaba-labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- With Exploratory Data analysis we write Python codes manipulating the data into Panda Dataframes for further analysis and visuals using scatter plots , bar charts and line charts
- The analysis done with scatter plot is to visualize the launch outcomes from relationships of Payloadmass,flight number and launchsite and Orbit type
- The analysis done with bar charts is to visualize the success rate of orbit types
- The GitHub URL https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/EDA%20with%20Visualization%20lab.ipynb

EDA with SQL

- Displayed the names of unique launch sites and launchsites with string 'KSC'
- Displayed sum of payloadMass of NASA(CRC) and average payloadMass of booster version F9 v1.1
- Displayed the date where Landingoutcome was successful and booster versions of a certain payload Mass with successful landing outcomes
- Displayed a list of successful & failure landing outcomes , dates and payloadmass that are in certain dates
- The GitHub URL https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/jupyter-labs-eda-sql-edx_sqllite.ipynb

Build an Interactive Map with Folium

The Folium Map Objects that we use are :

- Folium.Circle,Folium.PolyLine(), Folium.Map.Marker,MarkerCluster and MousePosition functions

All the above mentioned are used to :

- Add a circle area of an intial centre location on the map
- create a series of connected line segments on the map marking the distance between launchsites and and major cities etc..
- Create a marker at a specific launch location
- Create cluster markers for succeful and failed launches
- Display the latitude and longitude coordinates of the on the map on a mouse cursor

The GitHub URL of interactive map with Folium map,https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/WilliamLetwaba-lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

For the Dashboard we used :

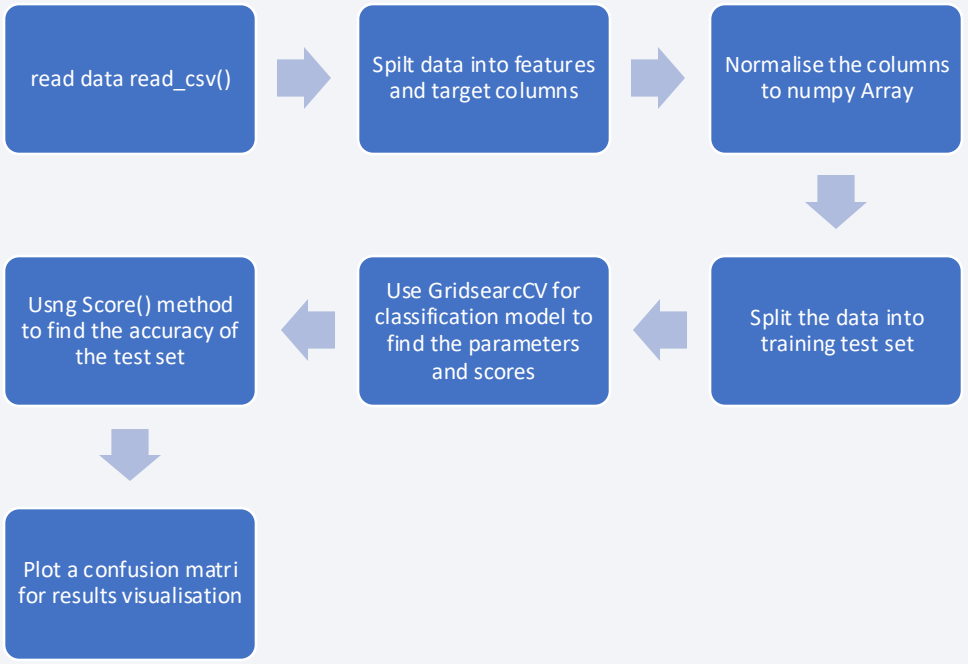
- Dropdown input of lauchsites
- A pie chart for showing success rates from the launch sites input
- RangerSlider for selecting payload Mass
- Scatter plots for showing correlations of payload and successful launchsites

- The GitHub URL of your Plotly Dash lab
  https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/WilliamLetwaba_dash_interactivity.ipynb

Predictive Analysis (Classification)

To perform a classification model we:

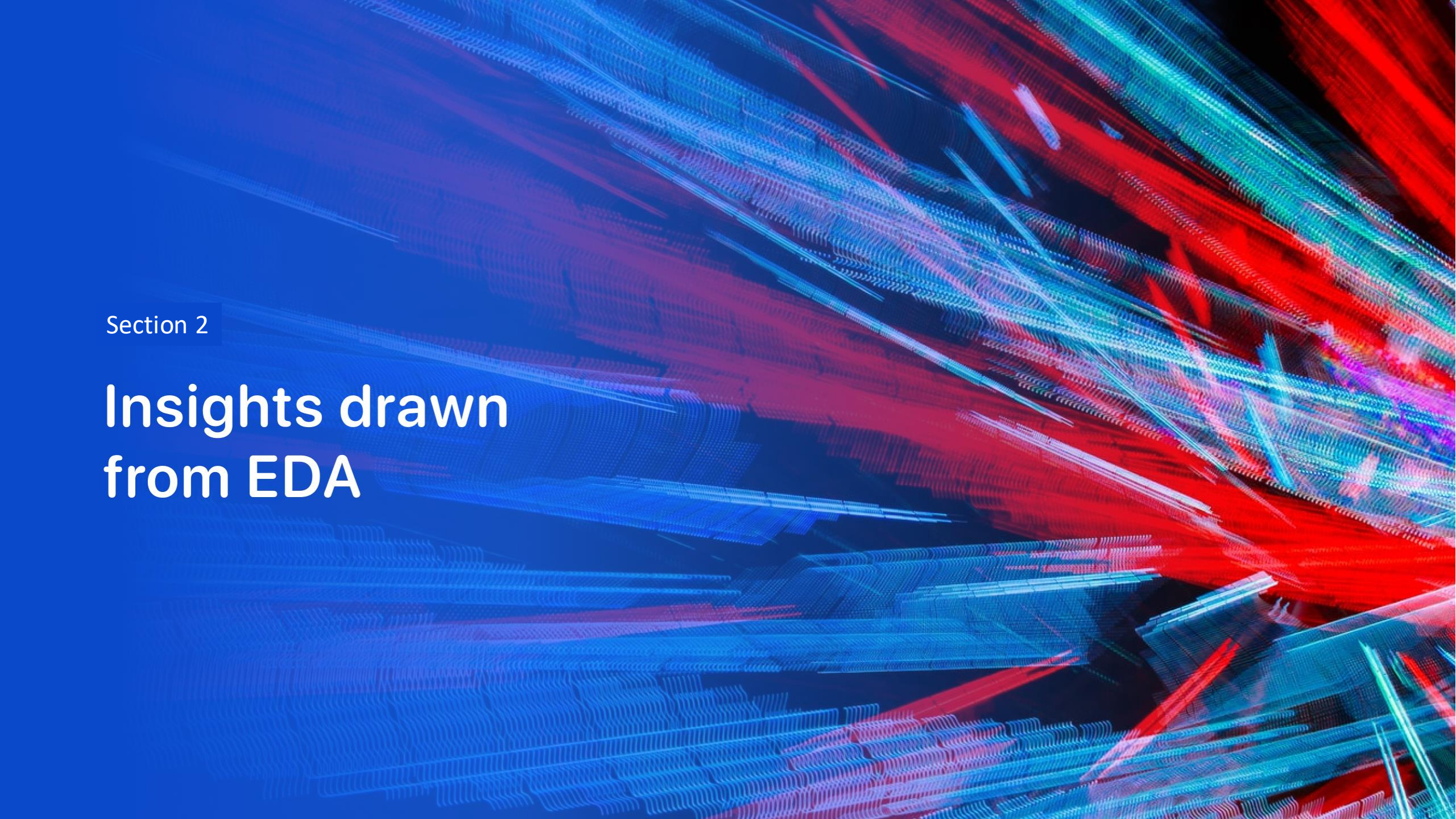
- Spilt data into features and target columns
- Normalise the columns to numpy Arrays
- Split the data into training test
- Do classification model to find the parameters and scores
- Usng Score() method to find the accuracy of the test set
- Plot a confusion matri for results visualisation



- The GitHub URL of your predictive analysis lab
  https://github.com/letwabaw/WLetwaba-Data-Science-Capstone-Project/blob/main/WilliamLetwaba_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

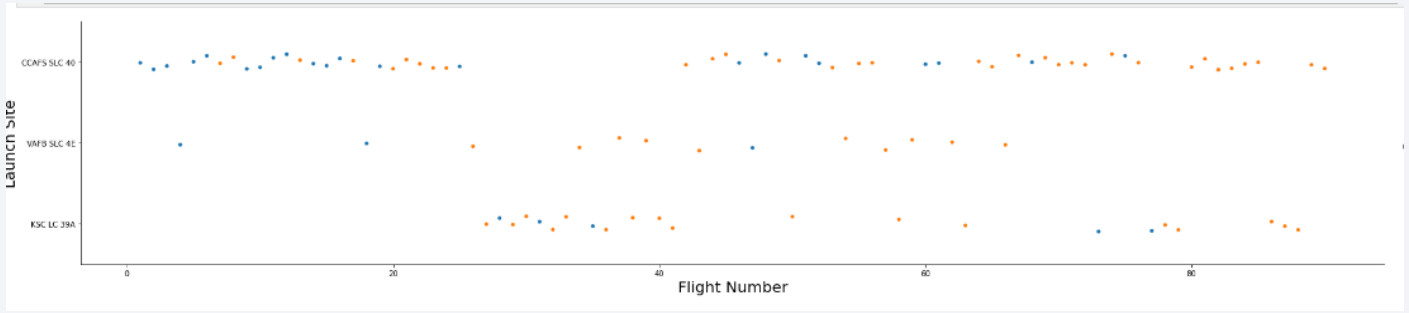
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

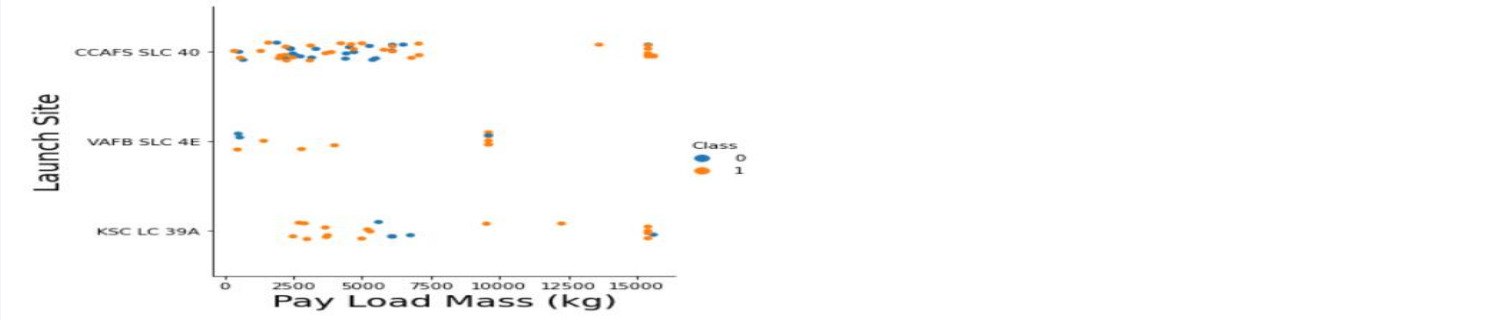
- Show a scatter plot of Flight Number vs. Launch Site



- The site CCAFS SLC 40 has the high number of rockets launhed than other sites
- Site VAFB SLC 4E has more successful launches

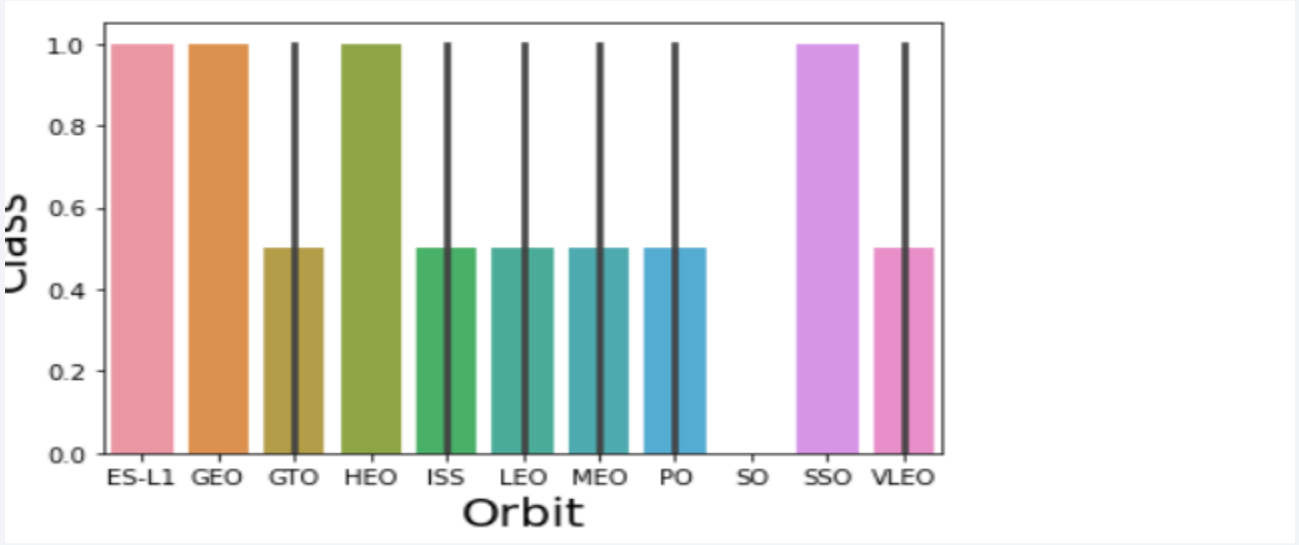
Payload vs. Launch Site

- The rockets lauched in all launch sites have payload mass less than 9000kg



Success Rate vs. Orbit Type

- The Orbit SO has least success rate



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 'KSC'

- Find 5 records where launch sites' names start with `KSC`
- 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 drone ship.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 records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

- 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>

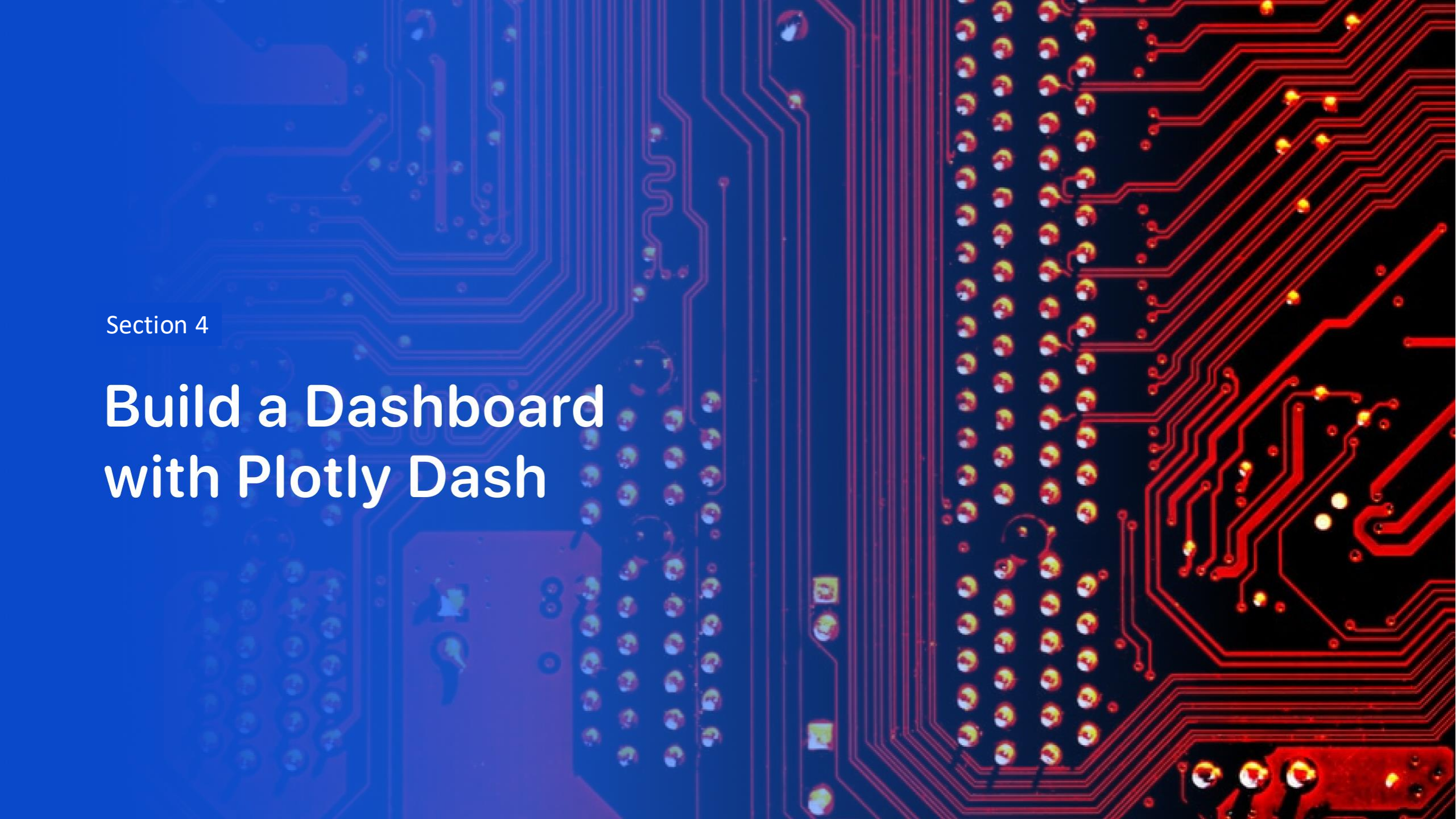
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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

- Point 1
- Point 2
- Point 3
- Point 4
- …

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!