IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name> Christian Gollee
<Date> 07/04/2025

Outline

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

Executive Summary

- I used the SpaceX API and Web scraping to collect the data
- I used pandas to do some data wrangling
- I used sql to get some insights inside the data
- I used visual EDA to investigate correlations between different quantities
- I developed machine learning models to predict future launch successes

Introduction

In this capstone, I wanted to predict if the Falcon 9 first stage will land successfully. 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. Therefore if I can determine if the first stage will land, I can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this lab, I collected and made sure the data is in the correct format from an API.

Section 1

Methodology

Methodology

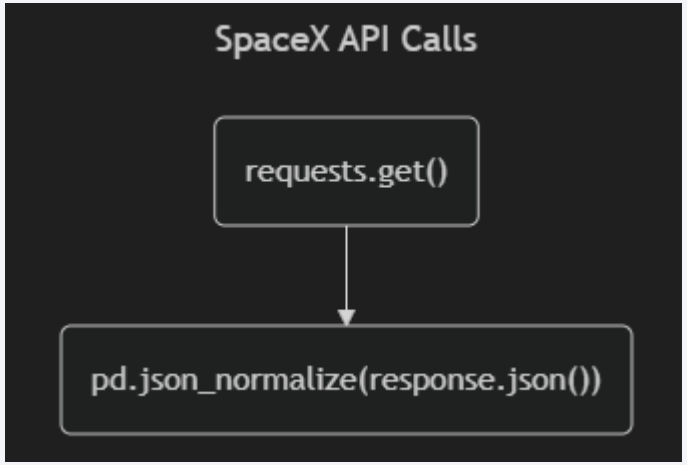
Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

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

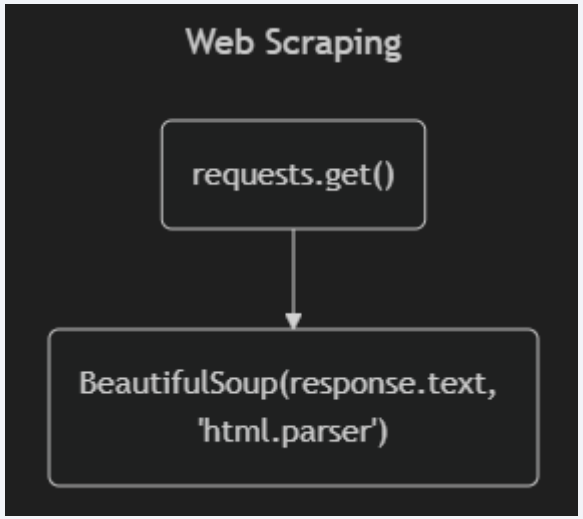


Coursera-Data-Science-Specialisation/Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · golleech/Coursera-Data-Science-Specialisation

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

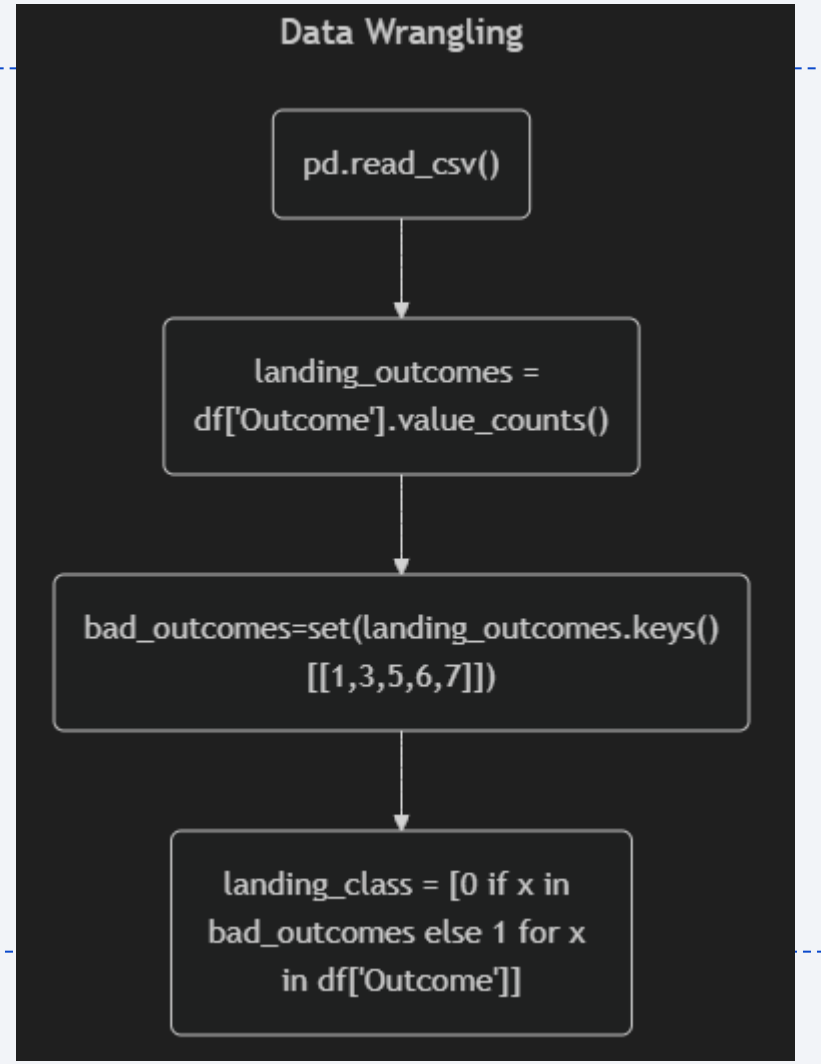


Coursera-Data-Science-Specialisation/Capstone/jupyter-labs-webscraping.ipynb at main · golleech/Coursera-Data-Science-Specialisation

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

[Coursera-Data-Science-Specialisation/Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · golleech/Coursera-Data-Science-Specialisation]



EDA with Data Visualization

- Scatter point chart for x = FlightNumber and y = Launch Site
- Scatter point chart for x = Payload Mass and y = Launch Site
- Bar chart for Orbit
- Scatter point chart for x = Flight Number and y = Orbit type
- Scatter point chart for x = Payload Mass and y = Orbit type
- Line chart for x = Year and y = average success rate

Coursera-Data-Science-Specialisation/Capstone/edadataviz.ipynb at main · golleech/Coursera-Data-Science-Specialisation

EDA with SQL

Coursera-Data-Science-Specialisation/Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · golleech/Coursera-Data-Science-Specialisation

- %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
- %sql select * from spacextable where Launch_Site like "CCA%" limit 16;
- %sql SELECT SUM(Payload) AS TotalPayloadMass FROM SPACEXTABLE WHERE Customer LIKE '%NASA (CRS)%';
- %sql select avg(payload_mass__kg_) as meanpayload from spacextable where booster_version like "F9%“
- %sql select min(date) from spacextable where mission_outcome = 'Success' and landing_outcome like '%ground pad%';
- %sql select * from spacextable where landing_outcome like '%drone ship%' and payload_mass__kg_ between 4000 and 6000;
- %sql select count(*) from spacextable where mission_outcome = "Success";
- %sql select booster_version from spacextable where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextable);
- %sql select substr(date, 6, 2) as month_name, landing_outcome, booster_version, launch_site from spacextable where landing_outcome like "%failure%drone ship%" and substr(date, 0, 5) = "2015";
- %sql select landing_outcome, count(landing_outcome) as outcome_count from spacextable where date between "2010-06-04" and "2017-03-20" group by landing_outcome order by date desc;

Build an Interactive Map with Folium

- I created markers and circles for the launch sites, lines for the shortest distance to railways, streets and coast lines
- I colored markers in a markers groups to highlight the success or missuccesses

- Coursera-Data-Science-Specialisation/Capstone/lab_jupyter_launch_site_location.ipynb at main · golleech/Coursera-Data-Science-Specialisation

Build a Dashboard with Plotly Dash

- I added dropdowns for the launch sites, a range slider for the payload mass to plot pie charts
- I added those widgets to investigate the success rate dependency from the launch site
- Coursera-Data-Science-Specialisation/Capstone/spacex-dash-app.py at main · golleech/Coursera-Data-Science-Specialisation

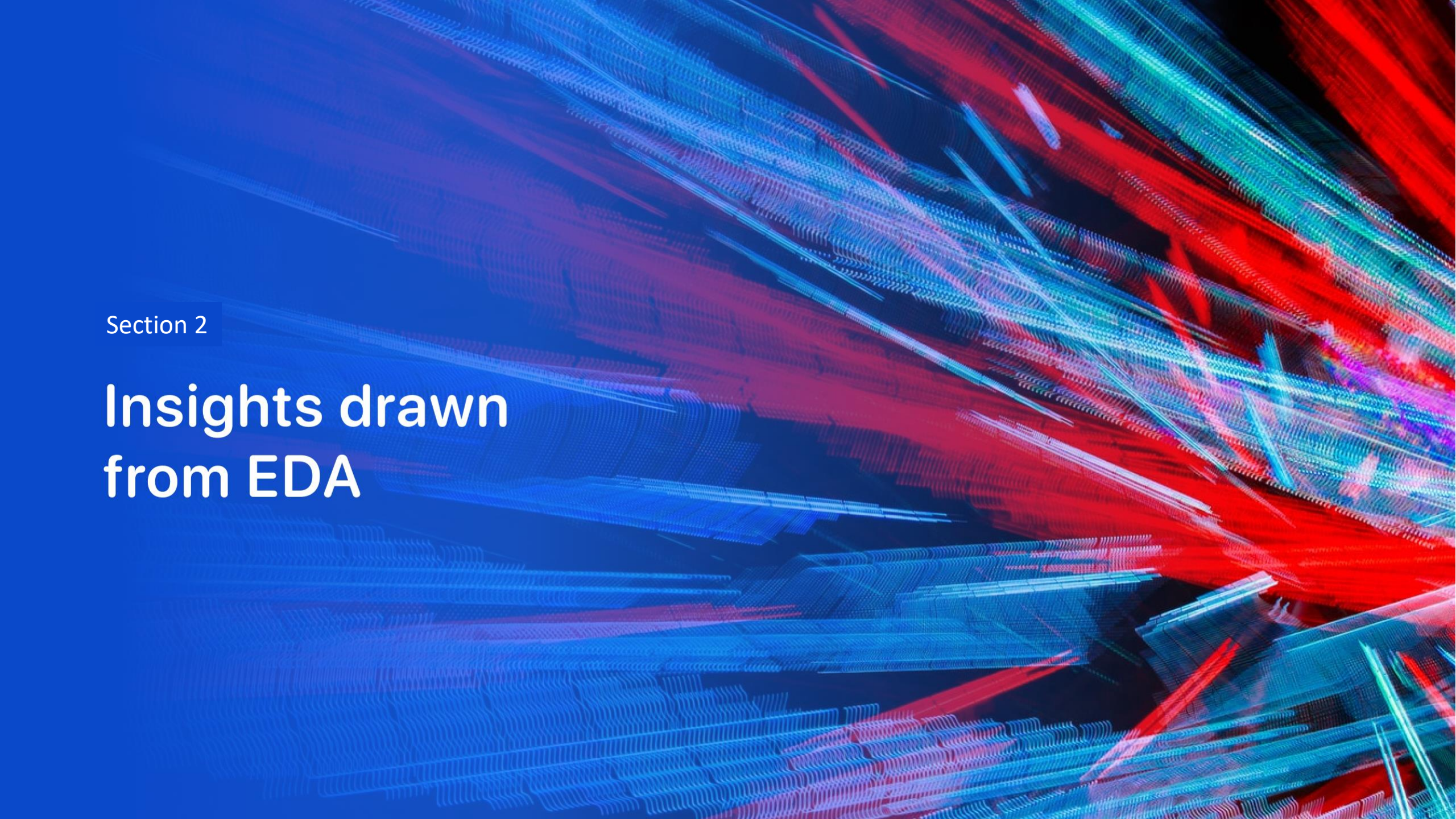
Predictive Analysis (Classification)

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

Predictive Analysis (Classification)

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

- 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>

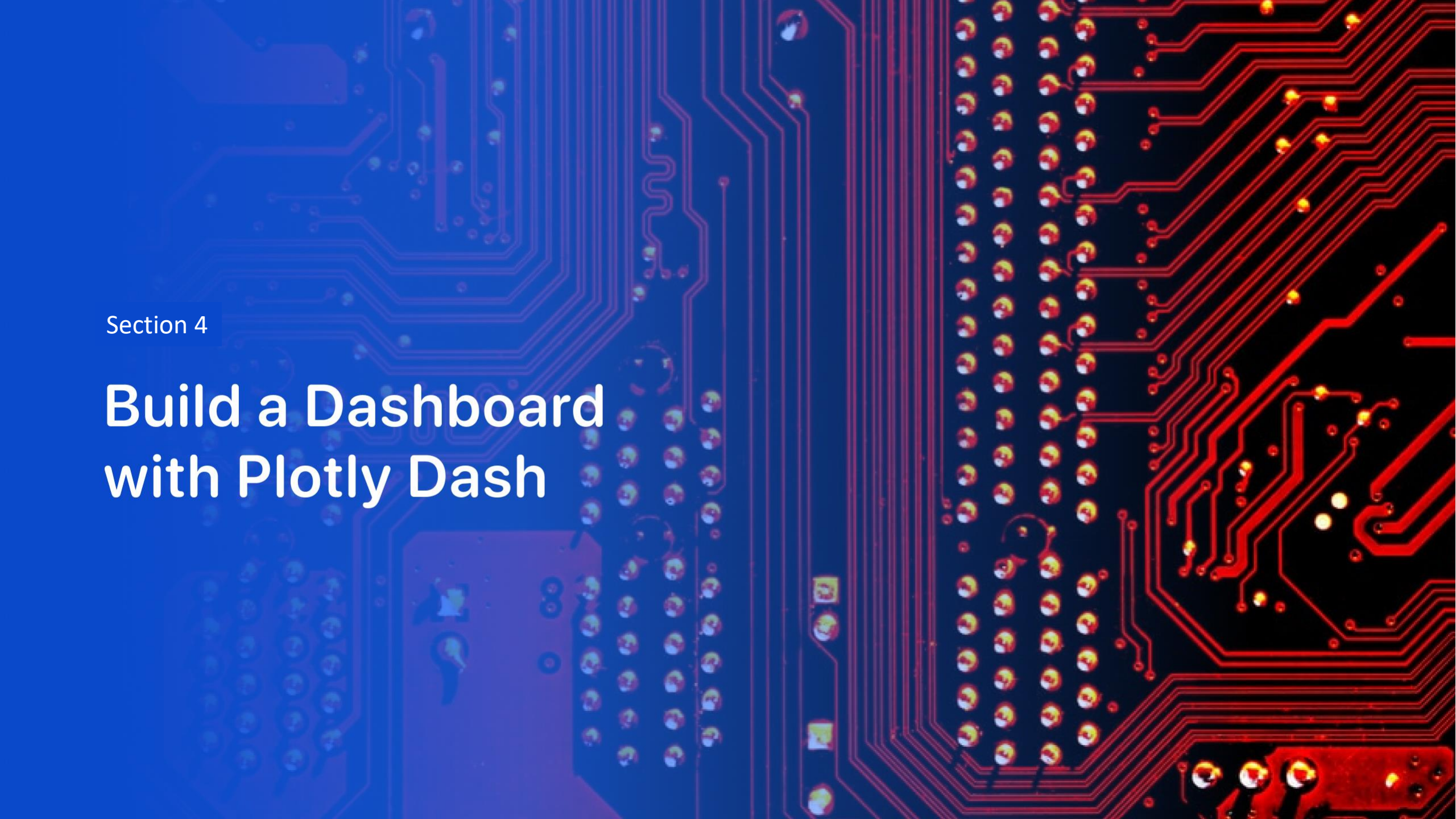
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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

- 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!