IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Brief overview of project goals: Predict Falcon 9 first stage landing success.
- Summary of methodologies: Data collection, wrangling, EDA, visualization, predictive modeling.
- Summary of results: Key insights from EDA, geospatial analysis, interactive dashboards, model performance.

Introduction

- SpaceX reduces launch costs by reusing Falcon 9 first stage.
- Launch cost comparison: $62M vs. $165M+ for competitors.
- Importance of predicting landing success for cost and competitive bidding.

- Problem: Can we predict if Falcon 9 first stage will land successfully?
- Objectives: Analyze launch data, identify key factors, build classification models.
- Questions: Impact of payload, launch site, orbit, flight number on landing success?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Interactive Visual Analytics
  - Predictive Analysis (Classification)

Data Collection

- Data sources: SpaceX REST API and Wikipedia web scraping.
- Tools: Python requests, BeautifulSoup, Pandas.
- Flowchart illustrating data retrieval and dataset creation.
- GitHub URL for API calls notebook.

Data Collection – SpaceX API

- Steps: GET requests, JSON to DataFrame conversion, filtering Falcon 9 launches, handling missing values.
- Key variables collected: flight number, payload mass, launch site, orbit, landing outcome.
- Flowchart of API data extraction process

Data Collection - Scraping

- Scraped Falcon 9 launch records from Wikipedia.
- Steps: Request HTML, parse tables with BeautifulSoup, convert to DataFrame.
- Flowchart of scraping and parsing process.
- GitHub URL for scraping notebook.

Place your flowchart of web scraping here

Data Wrangling

- Cleaning: Handling missing values with fillna(), removing duplicates.
- Feature engineering: Creating binary landing outcome label (0 = fail, 1 = success).
- Encoding categorical variables for modeling.

EDA with Data Visualization

- Tools: Pandas, Matplotlib, Seaborn, SQL queries.
- Goals: Understand variable distributions, relationships, and patterns.
- Scatter plot shows launch frequency across sites over time.
- Insight: Some sites have more launches, indicating operational preference.
- Scatter plot of payload mass by orbit type.
- Insight: Payload mass varies significantly with orbit destination.
- Line chart showing average landing success rate per year.
- Insight: Success rate improves over time, indicating technological progress.

EDA with SQL

- Query: List all unique launch sites.
- Result: Names of sites (e.g., CCAFS SLC-40, VAFB SLC-4E).
- Query: Find launch sites starting with 'CCA'.
- Result: Five records matching criteria.
- Query: Sum of payload mass for NASA missions.
- Result: Total payload mass value.
- Query: Average payload for booster version F9 v1.1.
- Result: Average value.
- Query: Date of first successful ground pad landing.
- Result: Specific date.
- Query: Boosters with successful drone ship landings and payload between 4000-6000 kg.
- Result: List of booster names.
- Query: Count of successful vs. failed missions.
- Result: Numbers for each category.
- Query: Boosters that carried maximum payload mass.
- Result: Booster names and payload values.
- Query: Failed drone ship landings in 2015 with booster and site info.
- Result: List of records.
- Query: Count and rank of landing outcomes in date range.
- Result: Ranked list of outcomes.

Build an Interactive Map with Folium

- Map showing all launch site locations globally.
- Map with color-coded markers for successful and failed landings.
- Map showing distances from launch site to nearby infrastructure: railway, highway, coastline.

Build a Dashboard with Plotly Dash

- Pie chart showing success counts for all launch sites.
- Pie chart focusing on site with highest success ratio.
- Scatter plot with interactive payload range slider.

Predictive Analysis (Classification)

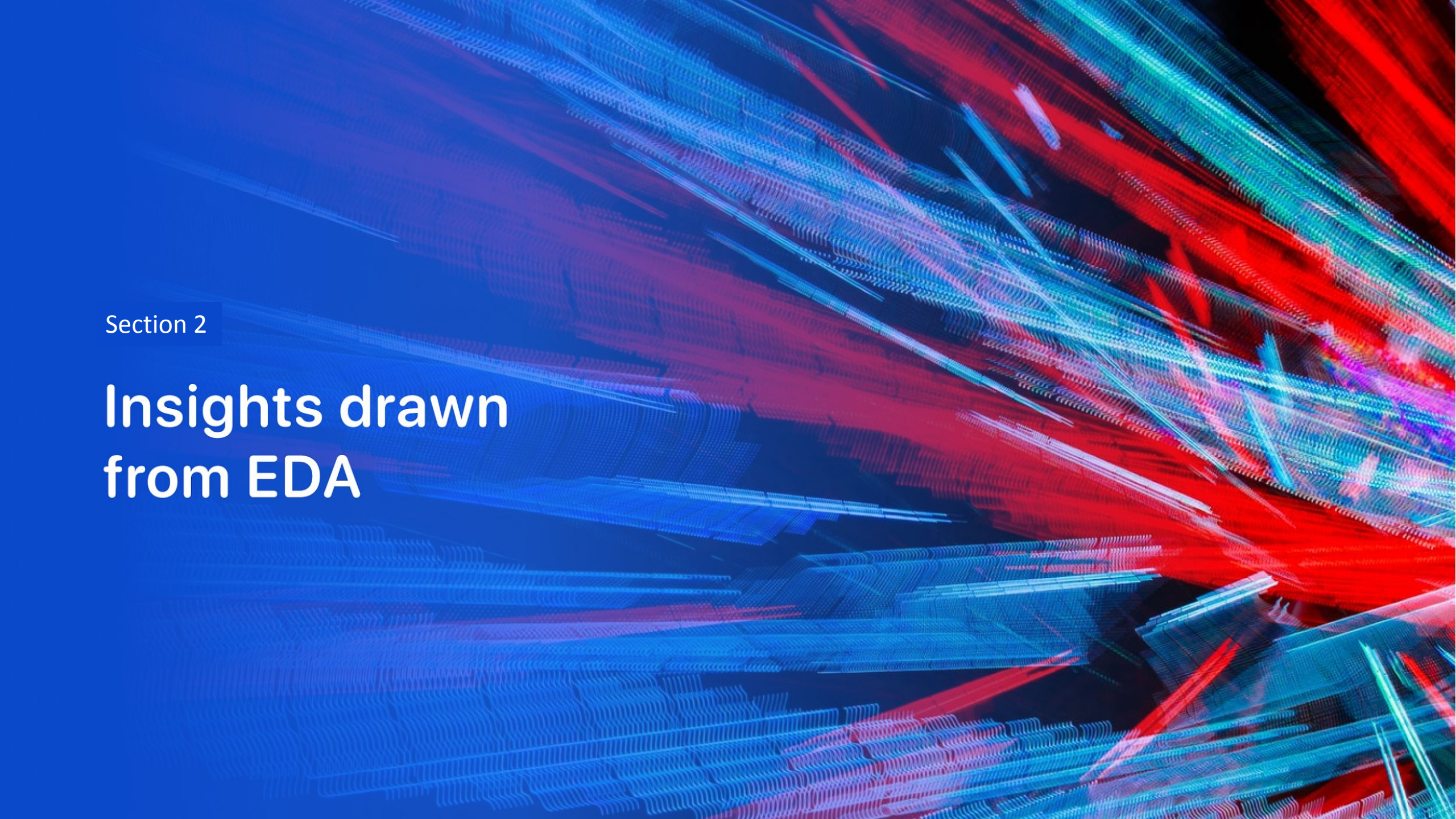
- Models used: Logistic Regression, Decision Trees, SVM, KNN.
- Data preprocessing: Standardization, train-test split.
- Hyperparameter tuning with GridSearchCV.
- Bar chart comparing accuracy of all models.
- Highlight best performing model.
- Confusion matrix visualization.
- Explanation of true positives, false positives, etc.
- Interpretation of model performance.
- Key features influencing landing success (e.g., payload, orbit).
- Model strengths and limitations.
- Potential for improvement with more data.
- Recap: Data collection and wrangling enabled robust analysis.
- EDA revealed important patterns influencing landing success.
- Interactive maps and dashboards enhanced data exploration.
- Predictive models achieved good accuracy for classification.

Results

- Predicting landing success aids cost estimation and competitive bidding.
- Insights can guide launch planning and risk management.

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

- DONE IN EARLIER SLIDES

Launch Site Names Begin with 'CCA'

- DONE IN EARLIER SLIDES

Total Payload Mass

- DONE IN EARLIER SLIDES

Average Payload Mass by F9 v1.1

- DONE IN EARLIER SLIDES

First Successful Ground Landing Date

- DONE IN EARLIER SLIDES

Successful Drone Ship Landing with Payload between 4000 and 6000

- DONE IN EARLIER SLIDES

Total Number of Successful and Failure Mission Outcomes

- DONE IN EARLIER SLIDES

Boosters Carried Maximum Payload

- DONE IN EARLIER SLIDES

2015 Launch Records

- DONE IN EARLIER SLIDES

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

- DONE IN EARLIER SLIDES


Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

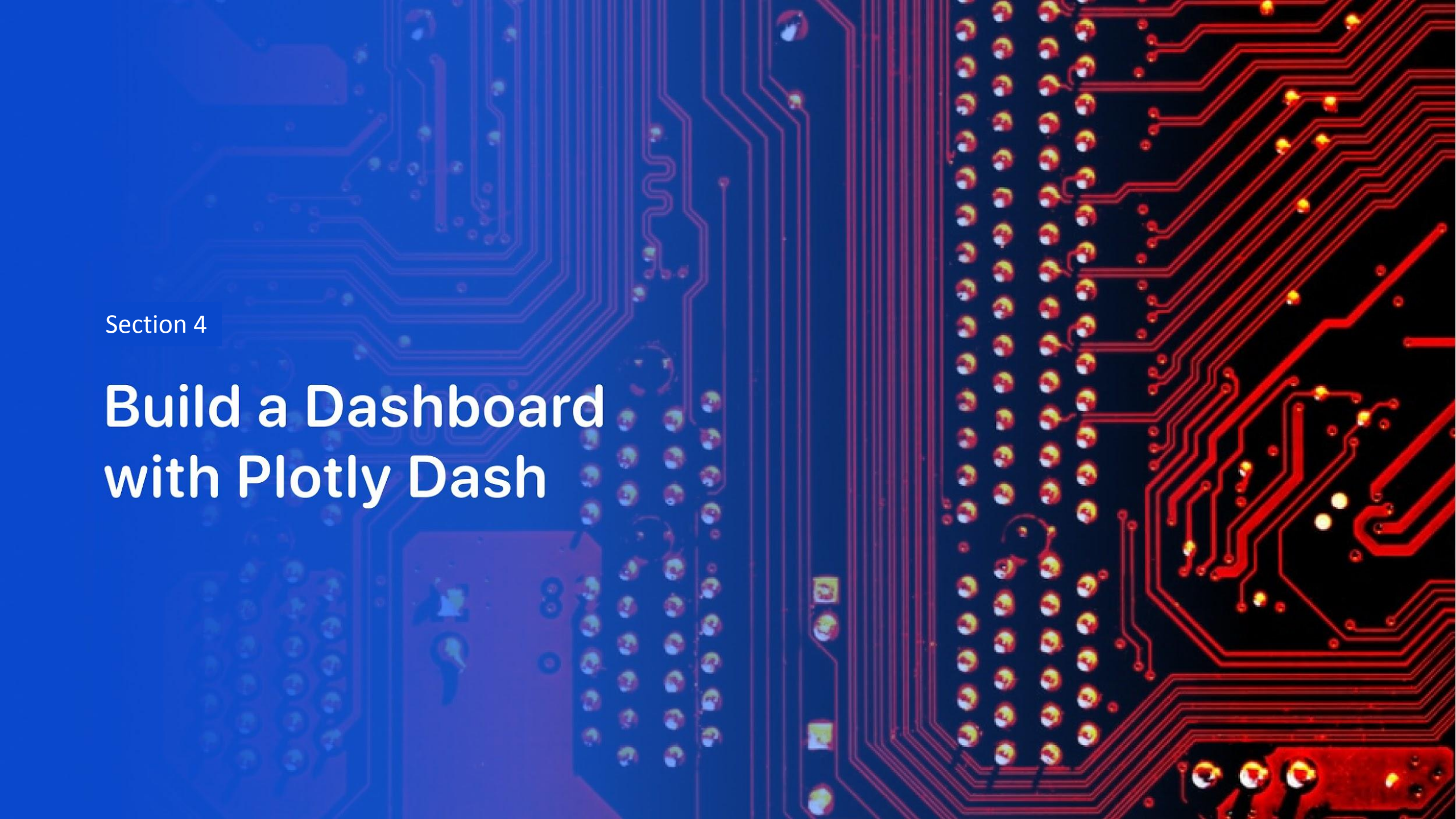
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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!