Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

Summary of Methodologies:

- **Data Collection**:

API requests, SQL queries, and web scraping

- **Data Wrangling**:

Cleaning and transforming datasets for consistency and analysis

- **Exploratory Data Analysis (EDA)**:
  - Univariate and bivariate analysis to identify trends and patterns
  - Geographic visualizations using maps
- **Machine Learning**:
  - Classification models for prediction
  - Feature engineering for improved model performance

Summary of All Results

- **Key Insights**:
  - Successful identification of patterns in launch site locations and performance
  - Predictive model with high accuracy for space launch outcomes
- **Visualization**:
  - Clear geographic mapping of launch data and insights
- **Final Outcome**:
  - Developed actionable insights for optimizing launch operations

Introduction

Project Background and Context:

- **Space Exploration Growth**:
  - Increased number of space missions in recent years
  - Demand for efficient and cost-effective launch strategies
- **Data-Driven Decisions**:
  - Leveraging data to optimize launch operations
  - Predicting launch success based on historical data

Problems I Want to Find Answers:

- **Launch Site Optimization**:
  - Which locations are most successful for launches?
  - How can launch performance be predicted based on historical data?
- **Predicting Success**:
  - Can machine learning accurately predict the outcome of a space mission?
  - What factors contribute the most to successful launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API data requests and web scraping for space mission data
  - SQL queries to manage and filter the datasets
- Perform data wrangling
  - Cleaned missing and inconsistent data
  - Feature engineering to prepare data for analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built classification models (e.g., Logistic Regression, Decision Trees)
  - Hyperparameter tuning for model optimization
  - Evaluated models with accuracy, precision, recall, and F1 score

Data Collection

- **API Requests**:
  - Collected mission data using API endpoints from space agencies
  - Automated script to extract data periodically
- **Web Scraping**:
  - Scraped supplementary data (e.g., launch costs, mission details) from space-related websites
  - Cleaned and organized scraped data
- **SQL Queries**:
  - Accessed large datasets stored in databases
  - Filtered and aggregated data for analysis using SQL

Flowchart:

1. **API Requests** → Gather mission data → Stored in local database
2. **Web Scraping** → Extract launch details → Clean and transform data
3. **SQL Queries** → Filter large datasets → Aggregate and process for analysis

Data Collection – SpaceX API

Data Collection Using SpaceX API:

- **REST API Calls**:
  - Utilized SpaceX public API for accessing mission data
  - Retrieved launch data, including launch site, payloads, and outcomes
- **Automated Requests**:
  - Implemented Python scripts to send periodic GET requests
  - Parsed JSON responses for relevant data
- **Data Storage**:
  - Stored extracted data in a local database for future analysis
  - Data organized by launch date, site, and outcome

Flowchart:

- REST API Request → GET mission data (launch, payload, outcome)
- Parse JSON → Extract key fields (launch site, rocket type, outcome)
- Store Data → Save in structured database for analysis

Data Collection - Scraping

- **Target Websites**:
  - Scraped space mission-related websites for launch details, costs, and payload information
- **Scraping Tools**:
  - Used BeautifulSoup and requests library to extract HTML data
  - Employed pandas to organize scraped data into structured formats (DataFrames)
- **Data Cleaning**:
  - Removed duplicates and irrelevant entries
  - Handled missing values and standardized formats for consistency

Flowchart:

HTTP Request → Send requests to target website
HTML Parsing → Extract key information (launch cost, mission details)
Data Structuring → Store data in a structured format (e.g., pandas DataFrame)
Data Cleaning → Handle missing and duplicate data, standardize formats

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

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

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
- 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

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>

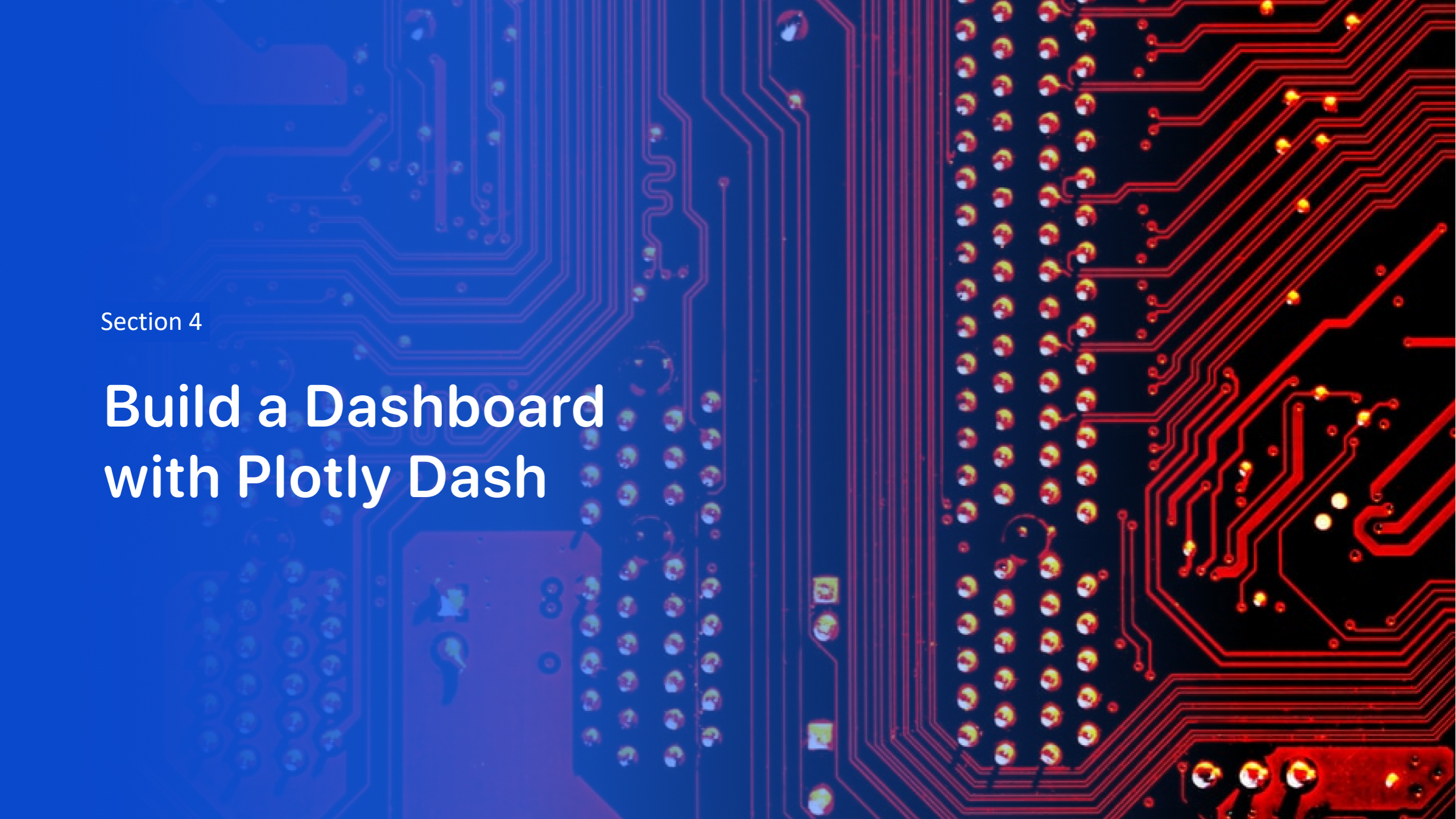
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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!