IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Robert Muigai Ngechu
08/06/2025

# Outline

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

# Executive Summary

This project predicts the success of SpaceX's Falcon 9 first-stage landings, utilizing real-world launch data. By analysing factors influencing landing outcomes, the goal is to enhance launch cost efficiency and provide competitive insights.

**Methodology Summary**

- **Data Collection**: Utilized RESTful APIs and web scraping to gather Falcon 9 launch data.
- **Data Wrangling**: Processed and cleaned data for analysis.
- **Exploratory Data Analysis (EDA)**: Employed SQL and Python for data exploration and visualization.
- **Predictive Modeling**: Developed classification models (SVM, Decision Trees, Logistic Regression) to predict landing success.
- **Dashboard Development**: Created interactive dashboards using Plotly Dash and Folium for data visualization.

**Summary of Results**

- Cleaned and structured Falcon 9 launch data
- Identified key success factors (e.g., launch site, payload)
- Found trends: heavier payloads = lower success
- Built models with up to **87% accuracy**
- Delivered interactive dashboard for predictions

# Introduction

Predicting Falcon 9 first-stage landing success using SpaceX launch data.

**Project Background**

SpaceX aims to reduce launch costs through booster reuse. Landing success is key to this goal.

**Key Questions**

- What factors affect landing success?
- Can we predict landing outcomes accurately?

Section 1

# Methodology

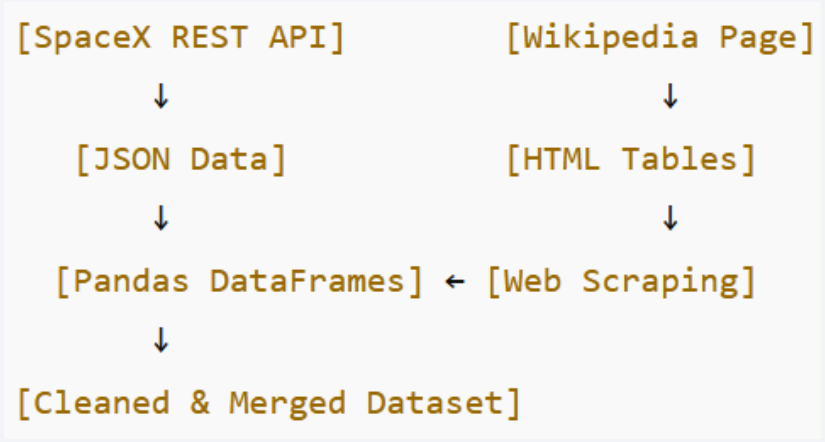
# Methodology

## Executive Summary

- Data collection methodology:
- 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

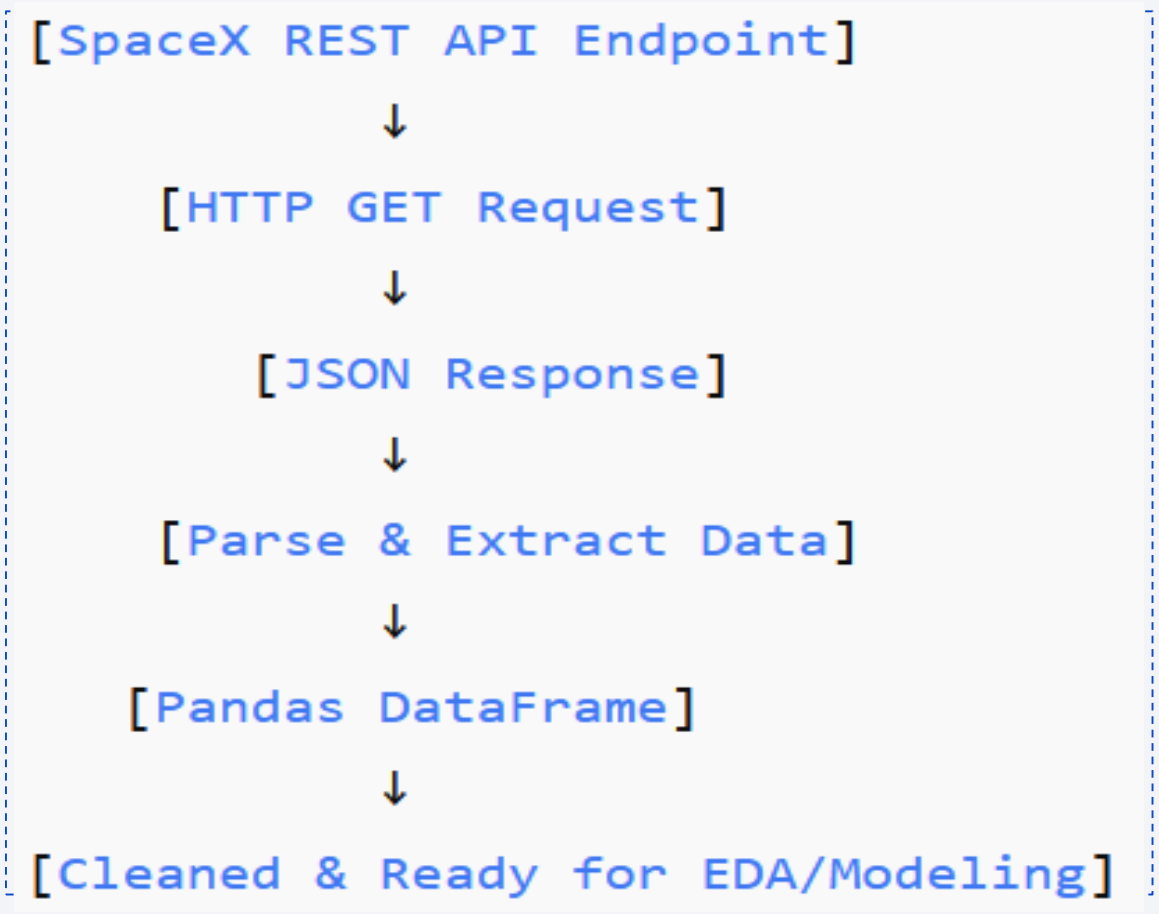
# Data Collection

- Data was collected using **SpaceX Launch API** and **web scraping** from Wikipedia. Launch records were stored, cleaned, and structured for analysis.
- **Data Collection Process**
  - **Key Phrases:**
  - **REST API Access** – Retrieved SpaceX launch data from public API
  - **Web Scraping (Wikipedia)** – Collected supplemental data on launches
  - **JSON to DataFrame** – Converted raw data into structured format
  - **Data Storage** – Stored and processed using Pandas for analysis
- Flowchart



# Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- **Key Phrases:**
  - **API Endpoint Access** – Retrieved launch data from SpaceX RESTful API
  - **HTTP GET Requests** – Fetched JSON responses for each launch
  - **Data Parsing** – Extracted relevant fields (launch date, site, payload, outcome)
  - **DataFrame Conversion** – Transformed JSON data into Pandas DataFrame for analysis



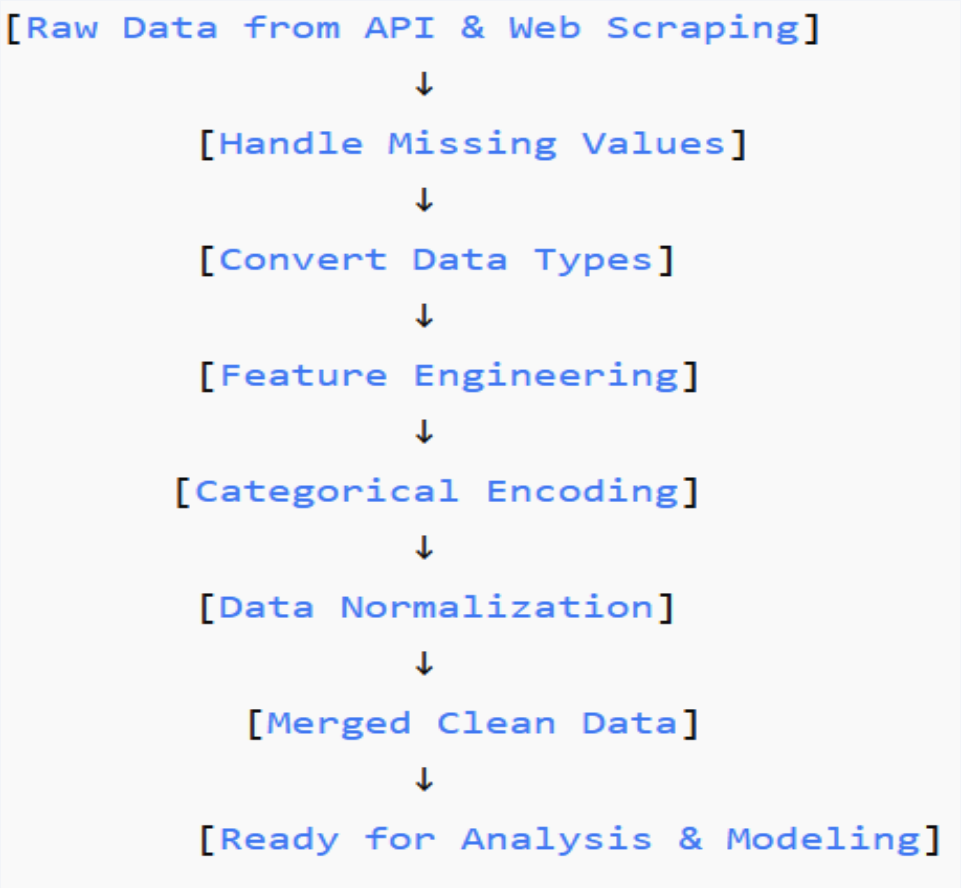
# Data Collection - Scraping

Web Scraping Process

- **Key Phrases:**
  - **Target URL Identification** – Selected SpaceX launch data page on Wikipedia
  - **HTTP Request** – Sent request to retrieve HTML content
  - **HTML Parsing** – Used BeautifulSoup to parse HTML tables
  - **Data Extraction** – Extracted launch records and relevant fields
  - **Data Cleaning** – Removed inconsistencies and formatted data
  - **DataFrame Conversion** – Stored cleaned data into Pandas DataFrame

# Data Wrangling

- **Key Phrases:**
  - **Handling Missing Values** – Imputed or removed incomplete data
  - **Data Type Conversion** – Converted columns to appropriate types (e.g., dates, categories)
  - **Feature Engineering** – Created new variables (e.g., success flag, payload groups)
  - **Categorical Encoding** – Transformed categorical data into numerical format
  - **Data Normalization** – Scaled features for consistent range
  - **Merged Datasets** – Combined API and scraped data into unified DataFrame



.

# EDA with Data Visualization

## Charts Plotted & Their Purpose

- **Bar Charts**

To compare landing success rates across different launch sites and booster versions.

- **Histograms**

  To visualize the distribution of payload masses and identify common ranges.

- **Scatter Plots**

  To explore relationships between payload mass and landing outcomes.

- **Pie Charts**

  To show the proportion of successful vs. failed landings.

- **Box Plots**

  To detect outliers and compare payload mass across different orbit types.

- **Heatmaps**

  To visualize correlations between features affecting landing success

# EDA with SQL

- **SQL Queries Summary**
  - Retrieved total launches and success counts by launch site
  - Filtered launches by orbit type to analyze success rates
  - Grouped payload mass ranges and calculated average landing outcomes
  - Joined multiple tables to combine launch details with booster info
  - Selected launches with reused boosters to compare success vs. new boosters
  - Calculated success rate trends over time (by year or mission number)

# Build an Interactive Map with Folium

## Folium Map Objects & Purpose

- **Markers**

  Placed at each launch site to identify locations of Falcon 9 launches.
- **Circles**
- Used to show launch site areas with sizes representing number of launches or success rates.
- **Circle Markers**

  Colored by landing outcome (success/failure) for quick visual distinction.
- **Lines / Polylines**

  Connected launch sites to landing zones to illustrate flight paths and booster recovery routes.
- **Popups and Tooltips**

  Added interactive info (e.g., launch date, payload mass) on markers for user engagement.

These objects help visualize spatial data, making patterns and relationships easier to understand.

# Build a Dashboard with Plotly Dash

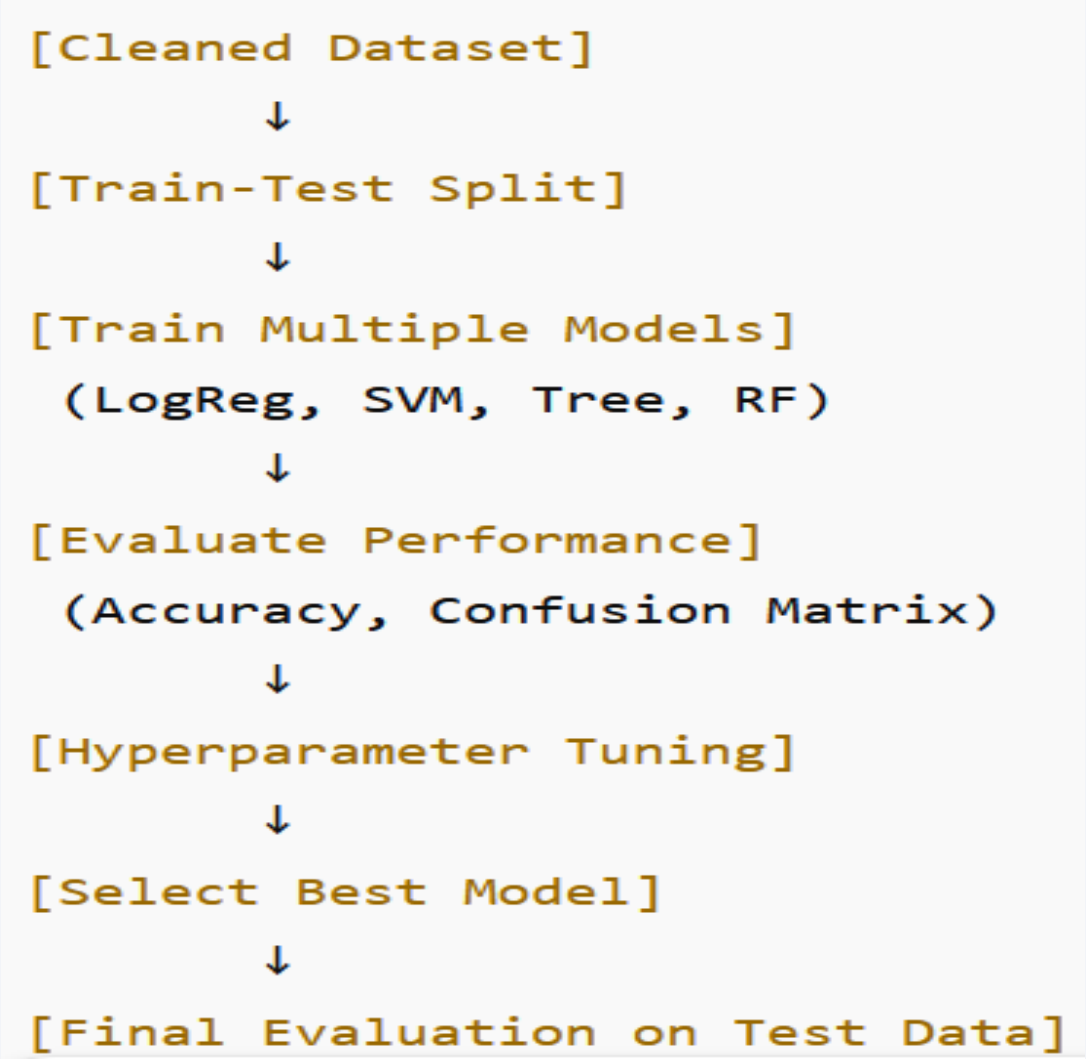
**Dashboard Plots & Interactions**

- **Landing Success Rate Bar Chart**
  Visualizes success rates by launch site for quick comparison.
- **Payload Mass Scatter Plot**
  Shows relationship between payload mass and landing outcome.
- **Orbit Type Pie Chart**
  Displays distribution of launches by orbit category.
- **Launch Timeline Line Graph**
  Tracks success trends over time.
- **Interactive Filters** (Dropdowns, sliders)
  Allow users to select launch site, orbit type, and payload range to customize views.
- **Hover Tooltips and Clickable Elements**
  Provide detailed info on data points for deeper insights.

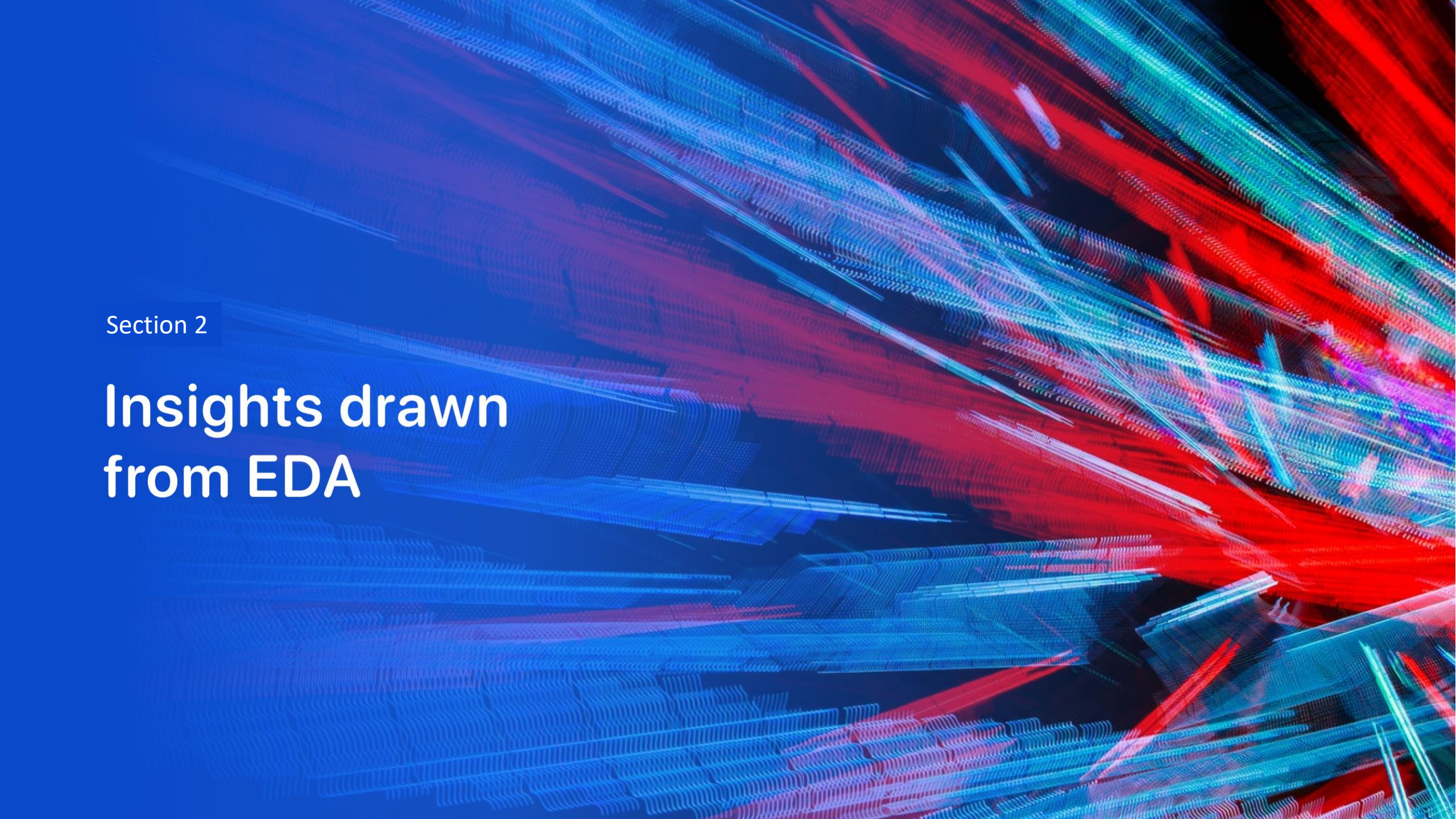
These elements enhance user engagement, making it easy to explore key factors influencing Falcon 9 landing success.

# Predictive Analysis (Classification)

- **Model Development Process**
- **Key Phrases:**
  - **Data Split** – Divided dataset into training and test sets
  - **Model Selection** – Tried multiple classifiers: Logistic Regression, SVM, Decision Tree, Random Forest
  - **Training** – Fit models on training data
  - **Evaluation** – Used accuracy, precision, recall, and confusion matrix
  - **Hyperparameter Tuning** – Applied GridSearchCV to improve model performance
  - **Best Model Selection** – Chose model with highest accuracy and generalization



# 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>

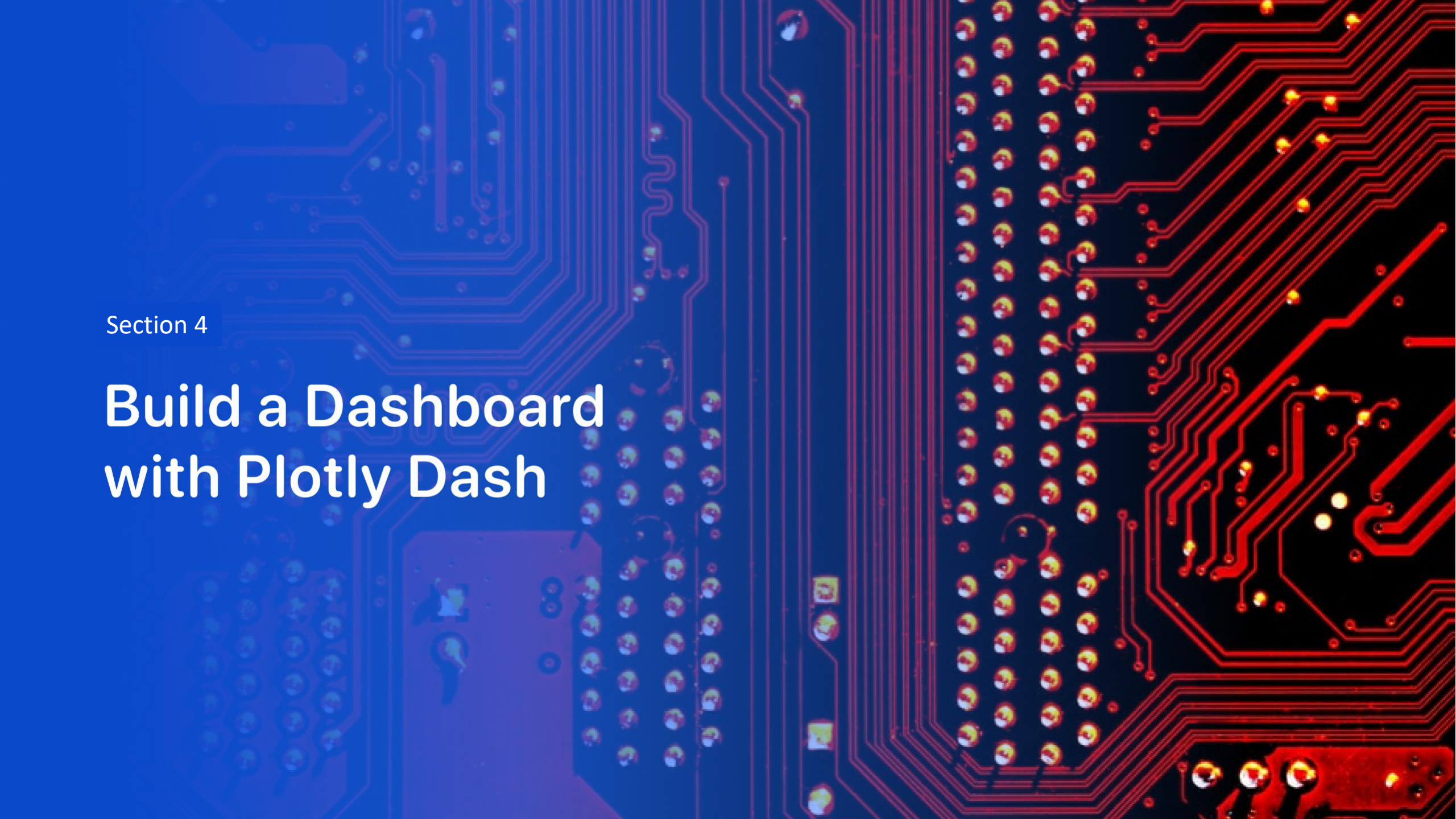
- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

# <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!