IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kough Teryila Gabriel

March, 2025

Outline

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

Executive Summary

- This report provides insights into SpaceX Falcon 9 launches, analyzing launch data using API collection, web scraping, data wrangling, SQL, exploratory data analysis (EDA), interactive visualizations, and machine learning classification models to predict landing outcomes.

Introduction

- The SpaceX Falcon 9 program aims to revolutionize space travel by enabling reusable rocket landings. This project explores historical launch data to analyze success factors and predict future landing outcomes using data-driven methods.

Section 1

Methodology

Methodology

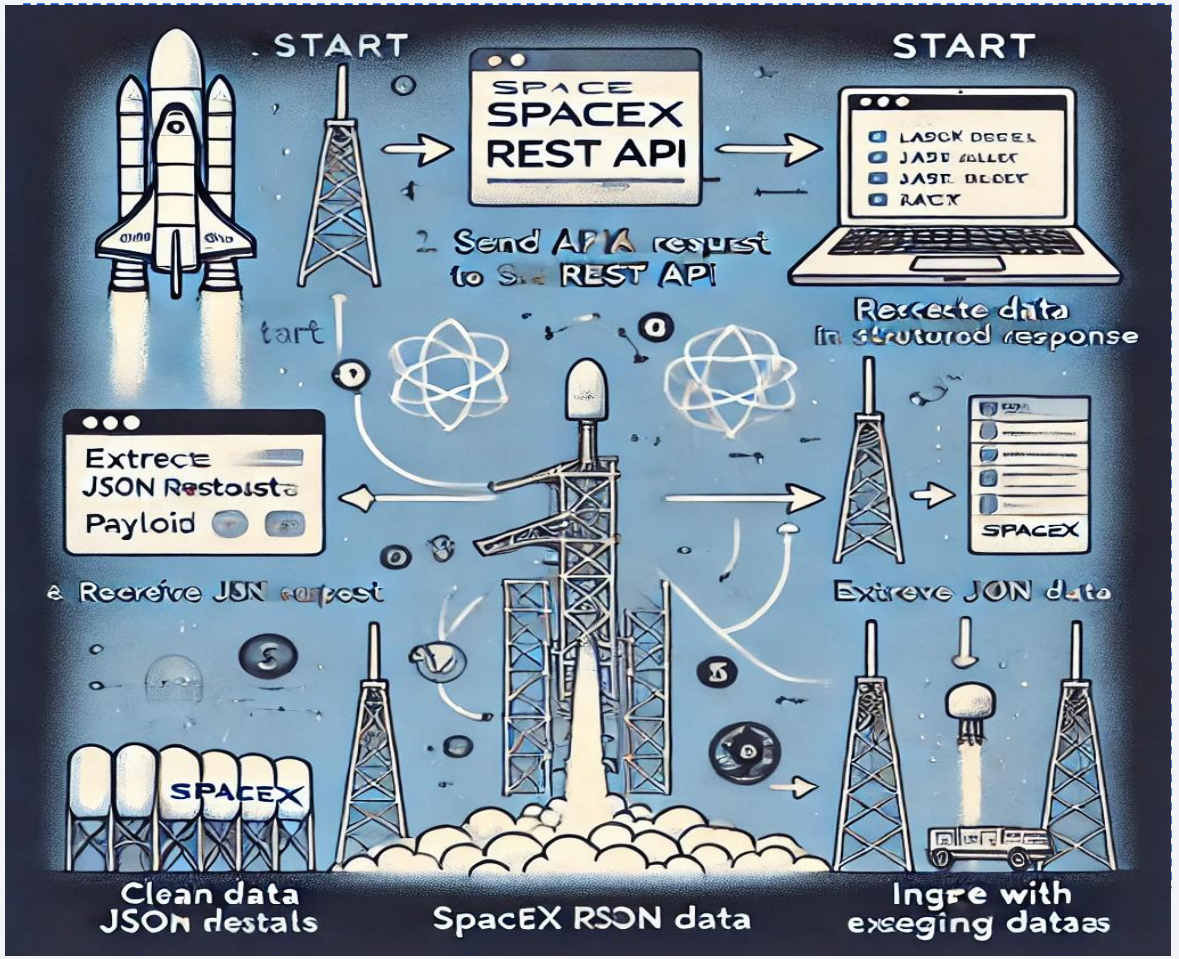
- **Data Collection:**
  - **API Calls:** SpaceX REST API was used to retrieve rocket, launchpad, payload, and core data.
  - **Web Scraping:** Additional historical launch data was extracted using BeautifulSoup.
  - **Data Wrangling:** Processed API and scraped data, handling missing values and formatting issues.
- **Exploratory Data Analysis (EDA) with SQL:**
  - **SQL Queries:** Analyzed launch success rates, payload mass effects, and orbit types.
  - **Data Visualizations:** Used histograms, scatter plots, and bar charts for insights.
- **Machine Learning:**
  - **Feature Engineering:** Created relevant predictors from launch data.
  - **Model Training:** Tested multiple classifiers to predict landing success.
  - **Model Evaluation:** Compared models using accuracy metrics.

Data Collection

- **Web Scraping:** Additional historical launch data was extracted using BeautifulSoup.
- **Data Wrangling:** Processed API and scraped data, handling missing values and formatting issues.

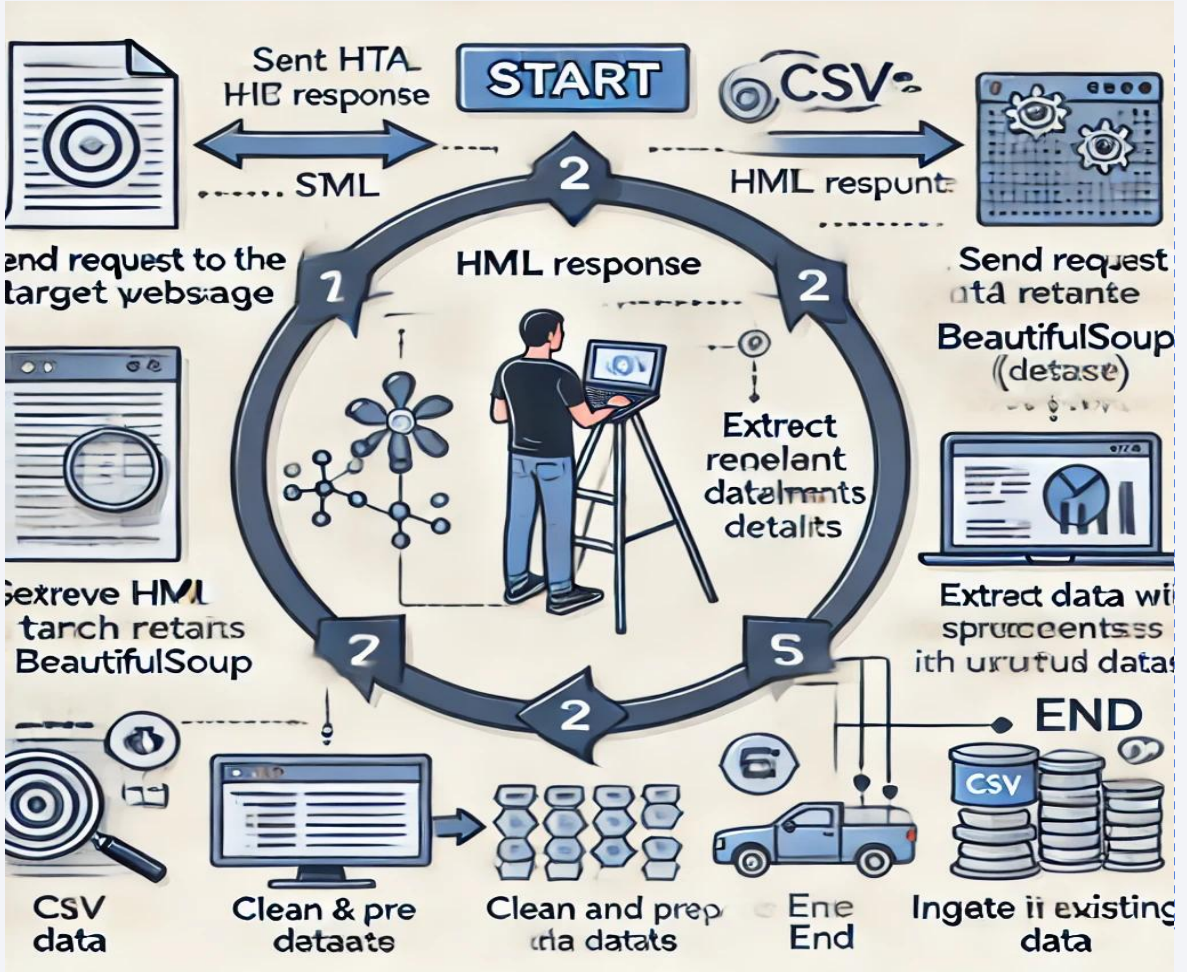
Data Collection – SpaceX API

- **API Calls:** SpaceX REST API was used to retrieve rocket, launchpad, payload, and core data.



Data Collection - Scraping

- **Step 1:** Send request to target web page.
- **Step 2:** Parse HTML using BeautifulSoup.
- **Step 3:** Extract relevant data elements (e.g., launch details).
- **Step 4:** Store data in structured format for further analysis.
- **Step 5:** Clean and integrate data with existing datasets.



- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review

Data Wrangling

- **Data Cleaning:**
  - Removed duplicate entries and null values.
  - Standardized column names and formats.
- **Feature Engineering:**
  - Extracted useful fields such as launch year, payload category, and booster type.
  - Created binary success/failure labels for classification tasks.
- **Data Transformation:**
  - Converted categorical values to numerical representations.
  - Normalized payload mass and other numerical attributes.
- **Integration:**
  - Merged datasets from API, web scraping, and existing historical data.
- https://github.com/Gabby210992/Machine-Learning/blob/main/SpaceX_data_wrangling.ipynb

EDA with Data Visualization

- **Flight Number vs. Launch Site:** Scatter plot analysis to show their relationship.
- **Payload vs. Launch Site:** Scatter plot visualization.
- **Orbit Type vs. Success Rate:** Bar chart of success rates.
- **Yearly Launch Success Trend:** Line chart of launch success over time.

EDA with SQL

- **Data Cleaning:**
  - Removed duplicate entries and null values.
  - Standardized column names and formats.
- **Feature Engineering:**
  - Extracted useful fields such as launch year, payload category, and booster type.
  - Created binary success/failure labels for classification tasks.
- **Data Transformation:**
  - Converted categorical values to numerical representations.
  - Normalized payload mass and other numerical attributes.
- **Integration:**
  - Merged datasets from API, web scraping, and existing historical data.
- *Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose*

Interactive Map with Folium

- **Folium Maps:**
  - Global launch site markers.
  - Launch outcomes color-coded.
  - Launch site proximity analysis.
- https://github.com/Gabby210992/Machine-Learning/blob/main/spaceX_map.ipynb

Build a Dashboard with Plotly Dash

- **Plotly Dash Dashboards:**
  - Pie charts of launch success rates.
  - Payload vs. Launch Outcome scatter plots.
- *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)

- **Classification Models:** Built machine learning models to predict landing success.
- **Model Evaluation:** Accuracy comparison across models.
- **Best Model:** Displaying accuracy and confusion matrix.
- *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

- The SpaceX Falcon 9 Landing Prediction project combined data collection, exploratory analysis, and machine learning to derive meaningful insights about launch success factors. The key results are as follows:
- The increasing trend in successful landings highlights the effectiveness of SpaceX's iterative design improvements.
- Certain launch conditions, payload characteristics, and booster versions significantly influence landing success.
- Predictive modeling provides a reliable approach to assessing future landing probabilities.

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>

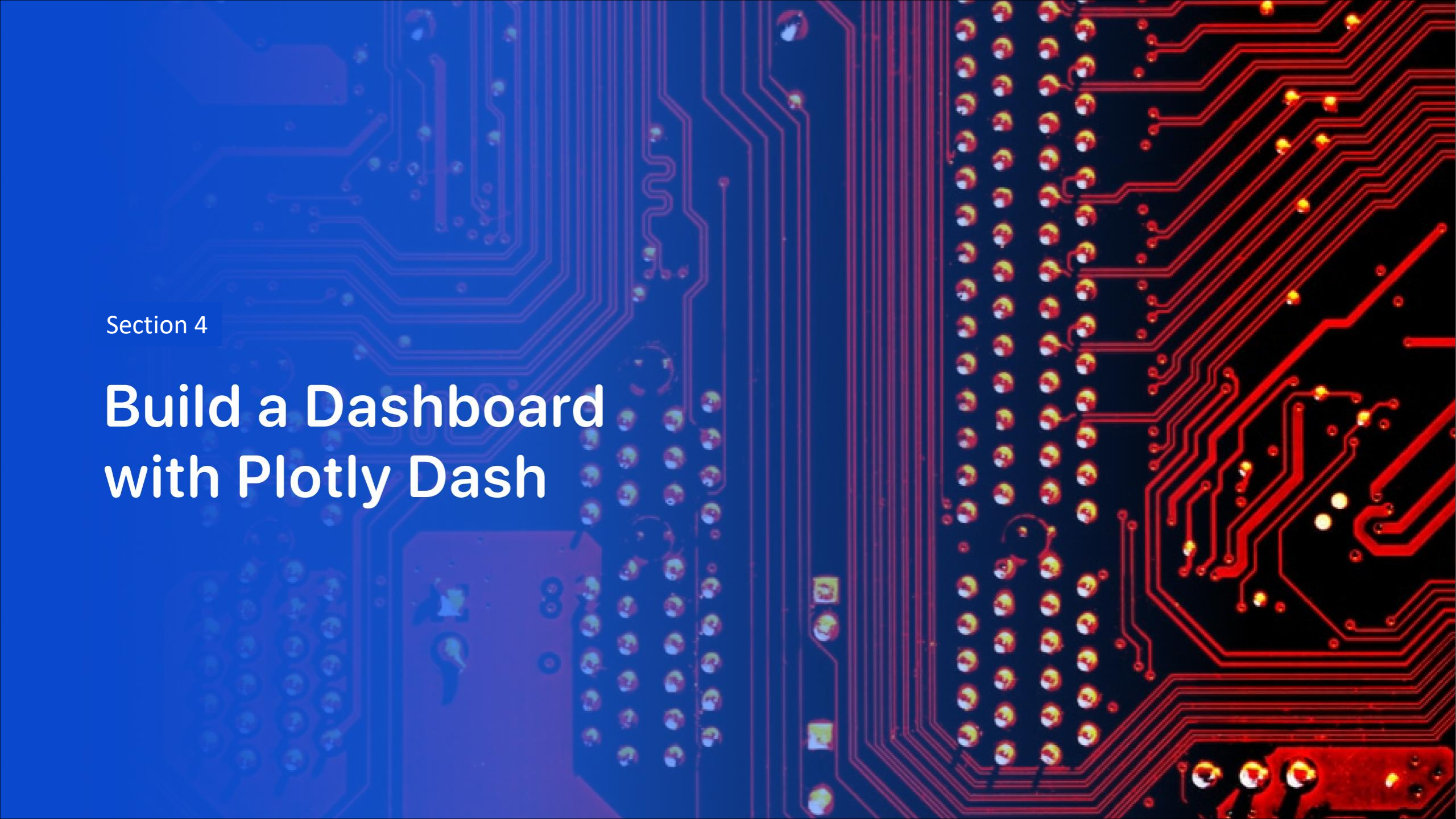
- Replace <Folium map screenshot 3> title with an appropriate title

- 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

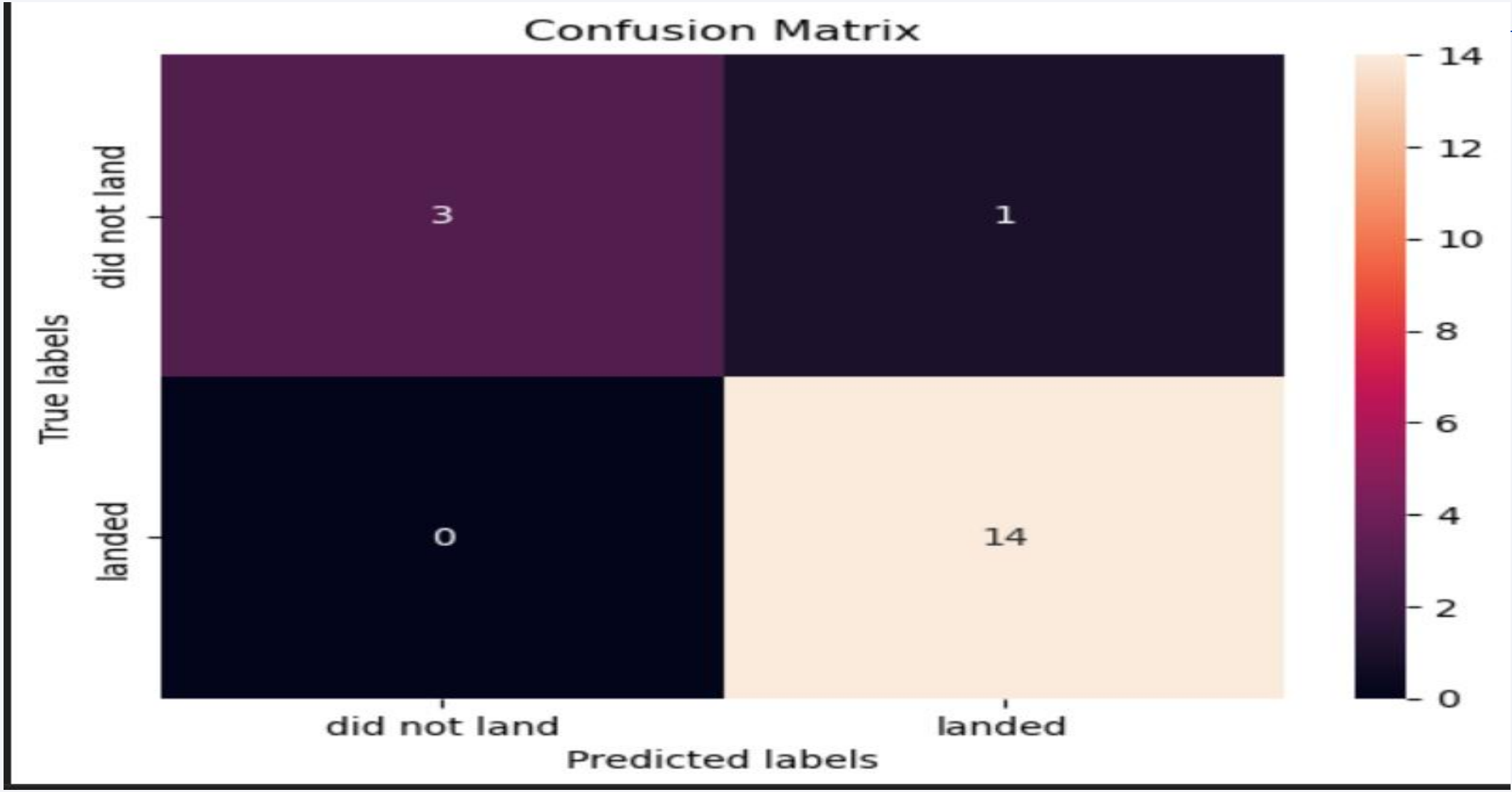
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- *Find which model has the highest classification accuracy*

KNN Confusion Matrix



Conclusions

- This project provided valuable insights into SpaceX Falcon 9 landings using a combination of web scraping, SQL analysis, and machine learning.
- The results highlighted key success factors such as launch site performance, payload impact, and orbit type.
- The predictive model demonstrated promising accuracy in forecasting landing success, which could help optimize future launches.
- Further improvements can be made by integrating more real-time data sources and refining feature engineering techniques.

Appendix

- Python code snippets
- SQL queries
- Data visualizations
- Model outputs

Thank you!