
SpaceX Falcon 9
First Stage
Landing Prediction
JG
IBM Data Science
Capstone Project
December 2024

Outline

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

https://github.com/FankaRoy/IBM-Data-Science-Professional-Certificate/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Executive Summary

(Updates with a few bullets of overall contents)

- Data collection and Webscraping with API
- Data wrangling/cleaning
- Data analysis with SQL and Visualization
- Interactive visual analytics and dashboarding
- Predictive Analytics with Machine Learning

Introduction

- Project background and context
- Problems you want to find answers

Introduction

- Space exploration offers unique opportunities for research & development innovations.
- However, it is prohibitively expensive.
- SpaceX advertises Falcon 9 rocket launches at ~37% of the costs compared to other providers.
- Reduced costs for SpaceX launches are generated from savings by reusing its rockets' first stage.

Introduction

Business Opportunity

- Other providers may offer competitive bids when reusing the first stage.
- To reuse the first stage, there must be a successful first stage landing.
- Therefore, the cost of a launch can be determined based on successful landing of the first stage.

Project Objective

This project will analyze publicly available Falcon 9 data to assess characteristics of first stage launches to predict successful landings.

Information gleaned from this analysis and prediction will inform proposals competing with SpaceX bids.

Section 1

Methodology

insert method image

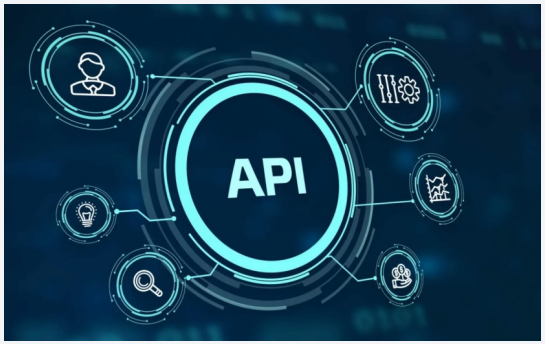
Methodology

Executive Summary

Methods will describe:

- Data collection
- Data wrangling processes
- Exploratory data analysis (EDA)
- Interactive visual analytics
- Predictive analysis using classification models

Data Collection

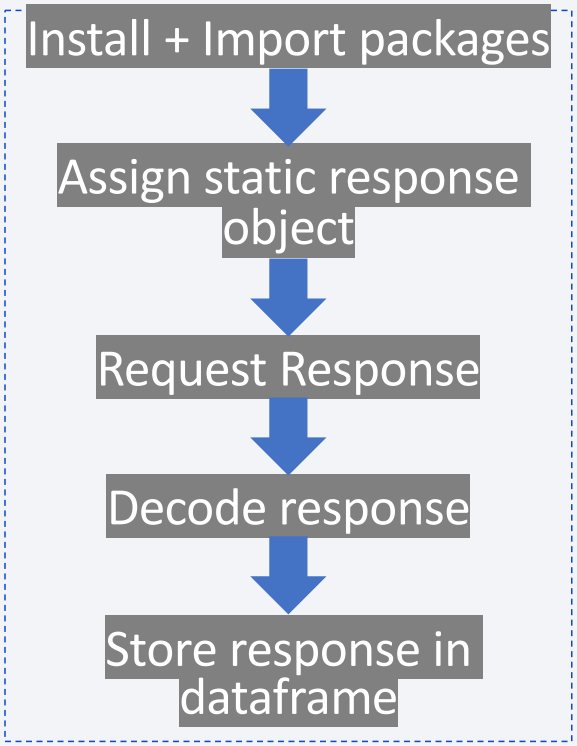


https://api.spacexdata.com/v4/launches/past

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

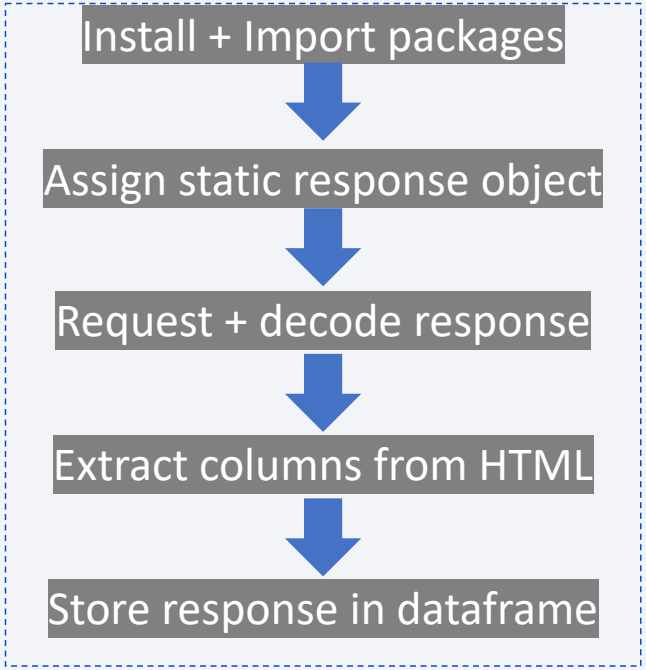
Data Collection – SpaceX API

- Import python packages (requests, pandas, numpy, datetime)
- Assign static response object to static_json_url
- Use requests.get() to request API response
- Use .json() to decode the response content
- Use json_normalize() to turn json data into dataframe
- See Github notebook for details: (must include completed code cell and outcome cell),



Data Collection - Scraping

- Install required packages
- Assign Falcon 9 website to static_url variable
- Use requests.get() method to request response from Falcon 9 website
- Use BeautifulSoup() to create an object from the response
- Extract column/variable names from Falcon 9 website
- Create data frame by parsing the Falcon 9 website tables
- See Github notebook for details: (must include completed code cell and outcome cell),



Data Wrangling

Data processing methods to assure data quality & prepare data for analysis

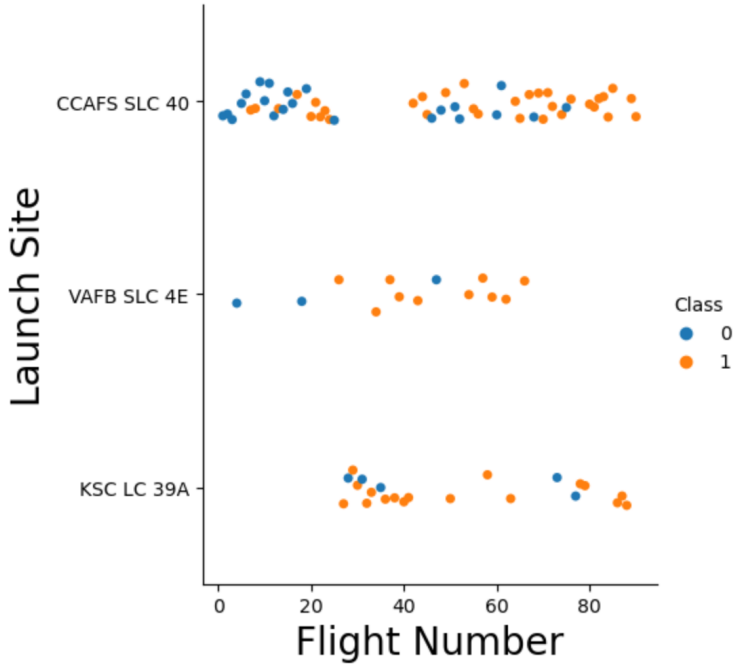
- Computed the percentage of the missing values in each attribute
  - df.isnull().sum()/len(df)*100
- Calculated the number of launches by launch site
  - launch_site_counts = df['LaunchSite'].value_counts()
- Calculated the number and occurrence of each orbit
  - orbit_counts = df['Orbit'].value_counts()
- Calculated the number & occurrence of mission outcome of the orbits
  - landing_outcomes = df['Outcome'].value_counts()
- Created binary landing outcome label (0/1) from Outcome column
  - landing_class=[0 if outcome in bad_outcome else 1 for outcome in df['Outcome']]
- GitHub URL of completed data wrangling

EDA with Data Visualization

The following visualizations were created:

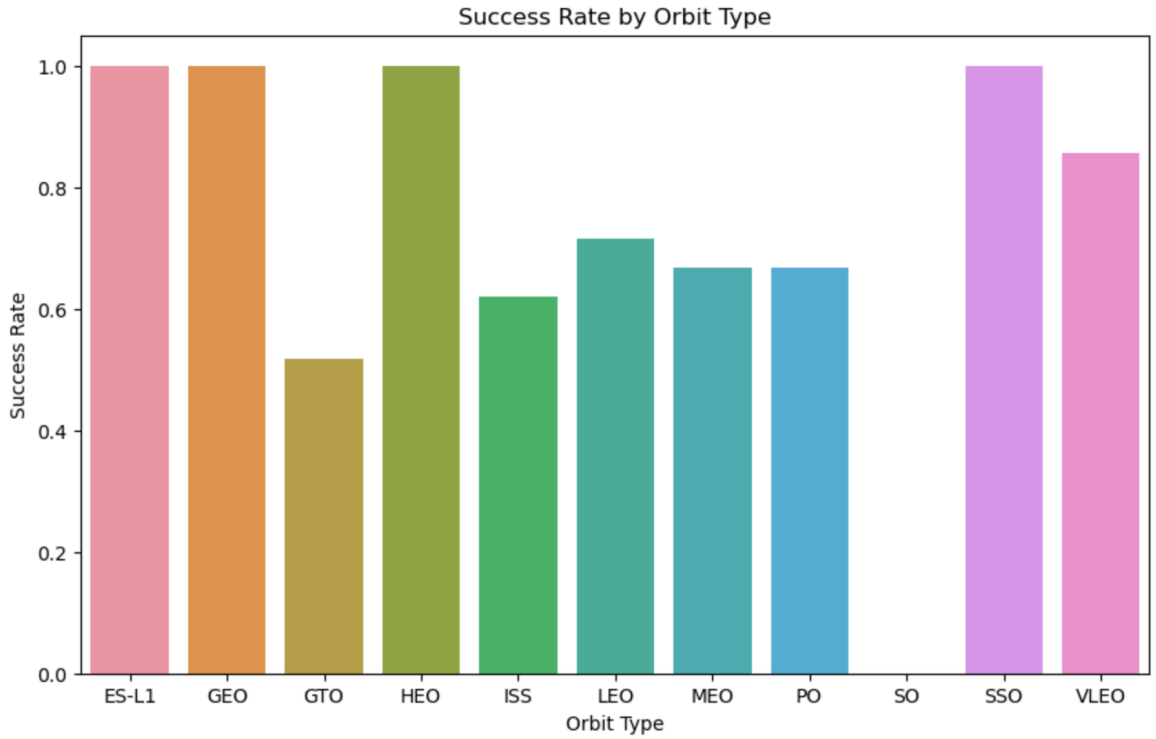
- Flight Number and Launch Site relationship
- Payload and Launch Site relationship
- Success rate of each orbit type
- Flight Number and Orbit type relationship
- Payload and Orbit type relationship
- Launch Success yearly trend

Relationship: Payload and Launch Site



- VAFB SLC 4E Appears that payloads < 10,000 Kg at launched at this site
- Unclear patterns for CCAFS SLC 40 & KSC LC 39A

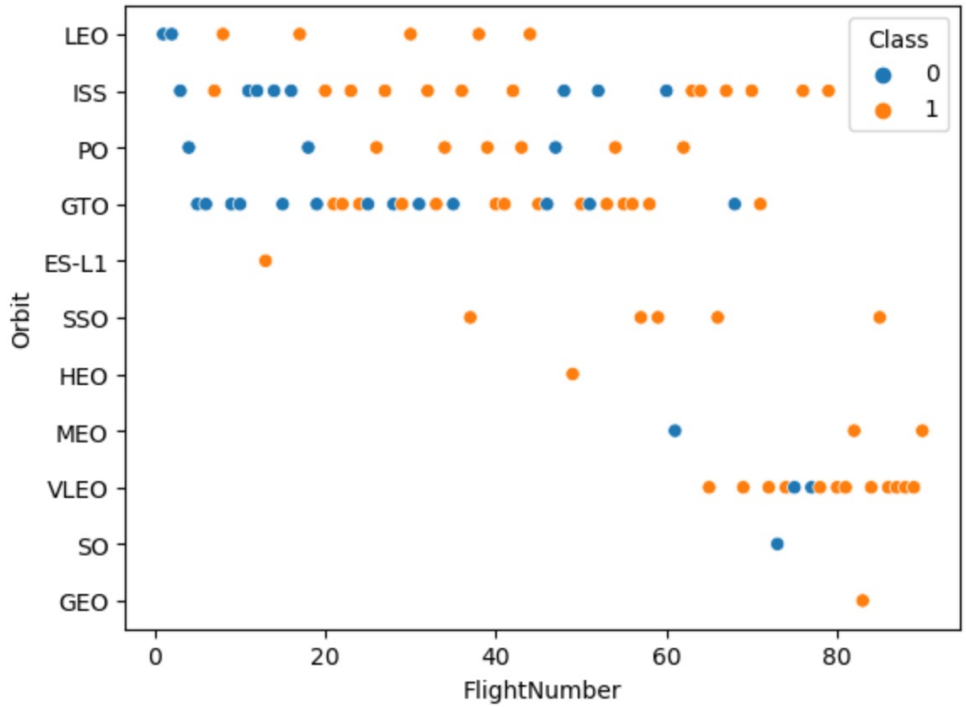
Relationship: Success rate of each orbit type



Higher success rates observed for orbits:

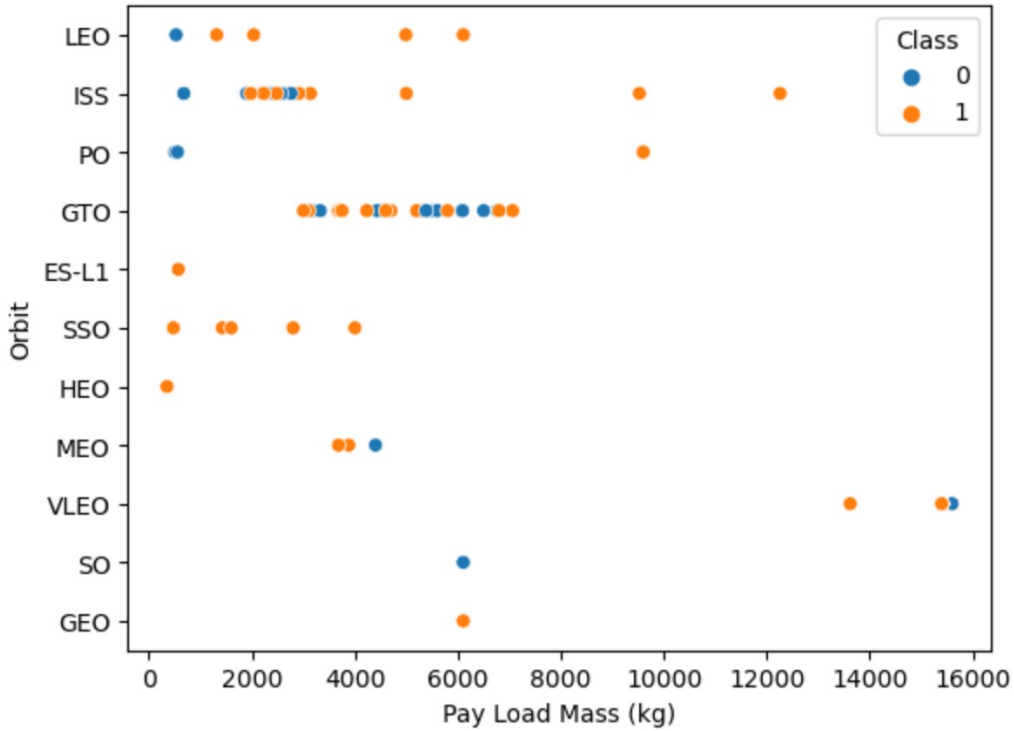
ES-L1
GEO
HEO
SSO
VLEO

Relationship: Flight Number and Orbit type



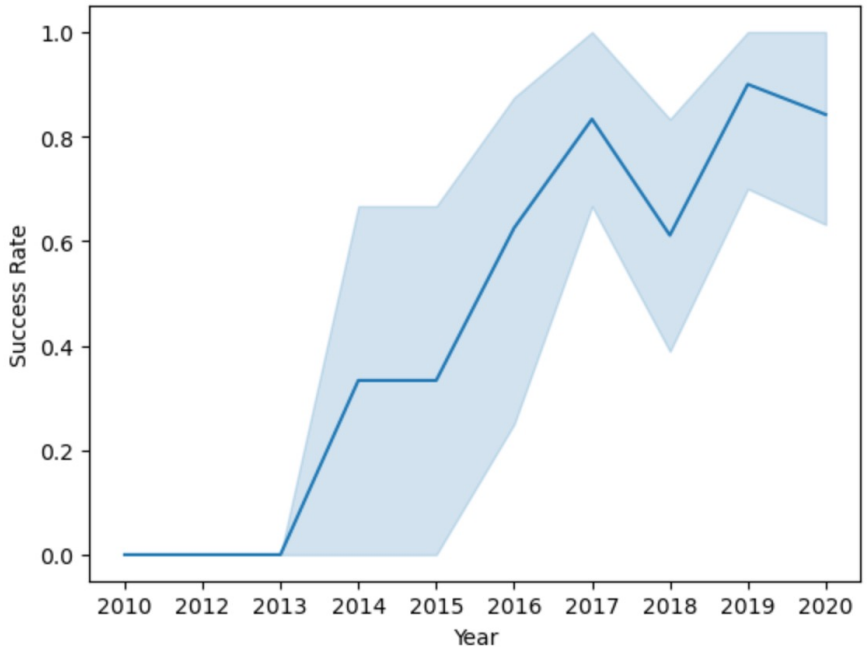
- LEO: Higher flight number, greater successful landings
- GTO: indeterminable pattern

Relationship: Payload and Orbit type



- LEO and ISS have successful landing with heavier payloads.
- Undetermined relationship for GTO

Launch success yearly trend



Generally upward trend of success rate following the year 2015

EDA with SQL

Summary of the SQL queries conducted

- Names of the unique launch sites in the space mission
- Launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date of first successful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship & payload mass greater than 4000 but less than 6000
- Names of the booster_versions which have carried the maximum payload mass. Use a subquery
- 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
- Github:

Build an Interactive Map with Folium

- Marked all launch sites on a Folium map
- Marked the success/failed launches for each site
- Calculated the distances between a launch site to its proximities
- Added a launch site drop-down input component
- Added a callback function to render the required success outcome pie chart
- Added a range slider to choose payload
- Added callback function to render the payload-outcome scatter plot
- Add the GitHub URL

Build a Dashboard with Plotly Dash

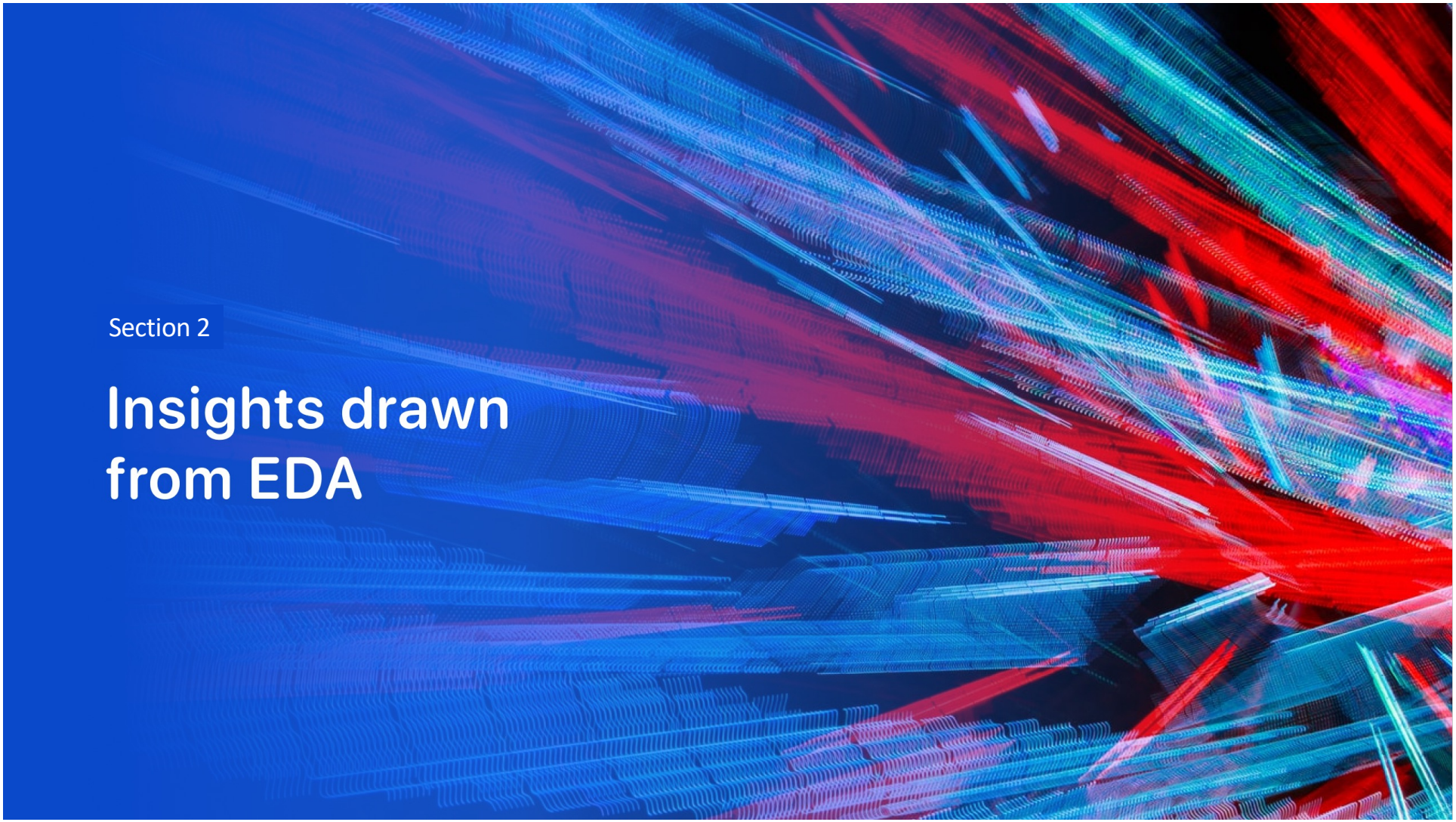
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL

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>

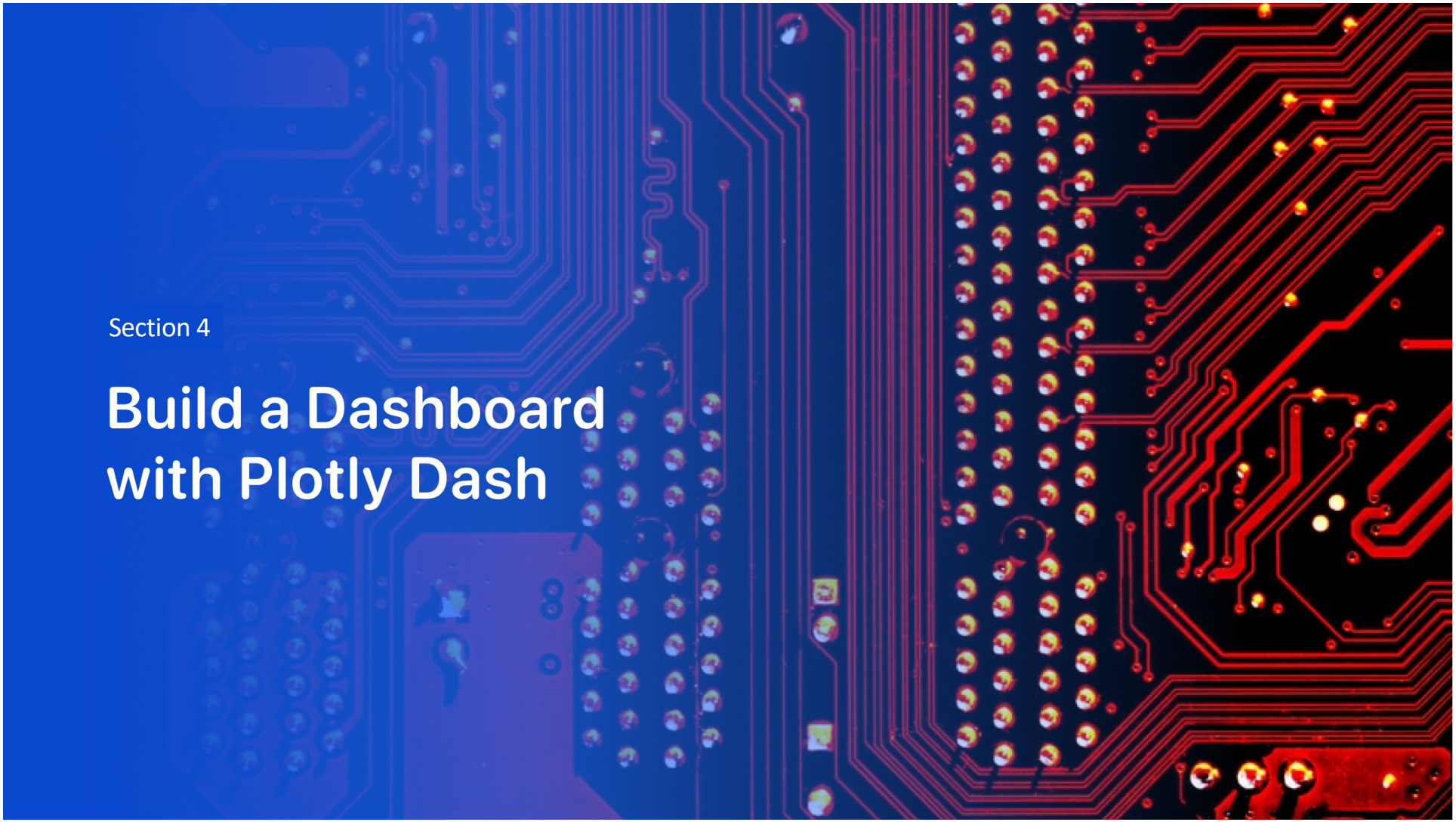
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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

- deliver your data-driven insights to determine if the first stage of Falcon 9 will land successfully.
- 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!