

Winning Space Race with Data Science

<Vagner Vale>
<12/02/2025>



Outline

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

Executive Summary

- Data were collected from various sources and data wrangling were performed. After, EAD techniques was used to get insights. Futhermore, predictive models were built, evaluated and refined for discovering more exciting insights.
- Summary of all results

Introduction

- The commercial space age is here, companies are making space travel affordable for everyone. The most successful is SpaceX. Space Y would like to compete with SpaceX. SpaceX takes cost advantage because can reuse the first stage
- Can we predict if the first stage will land? If so, we can determine the cost of a launch

Section 1

Methodology

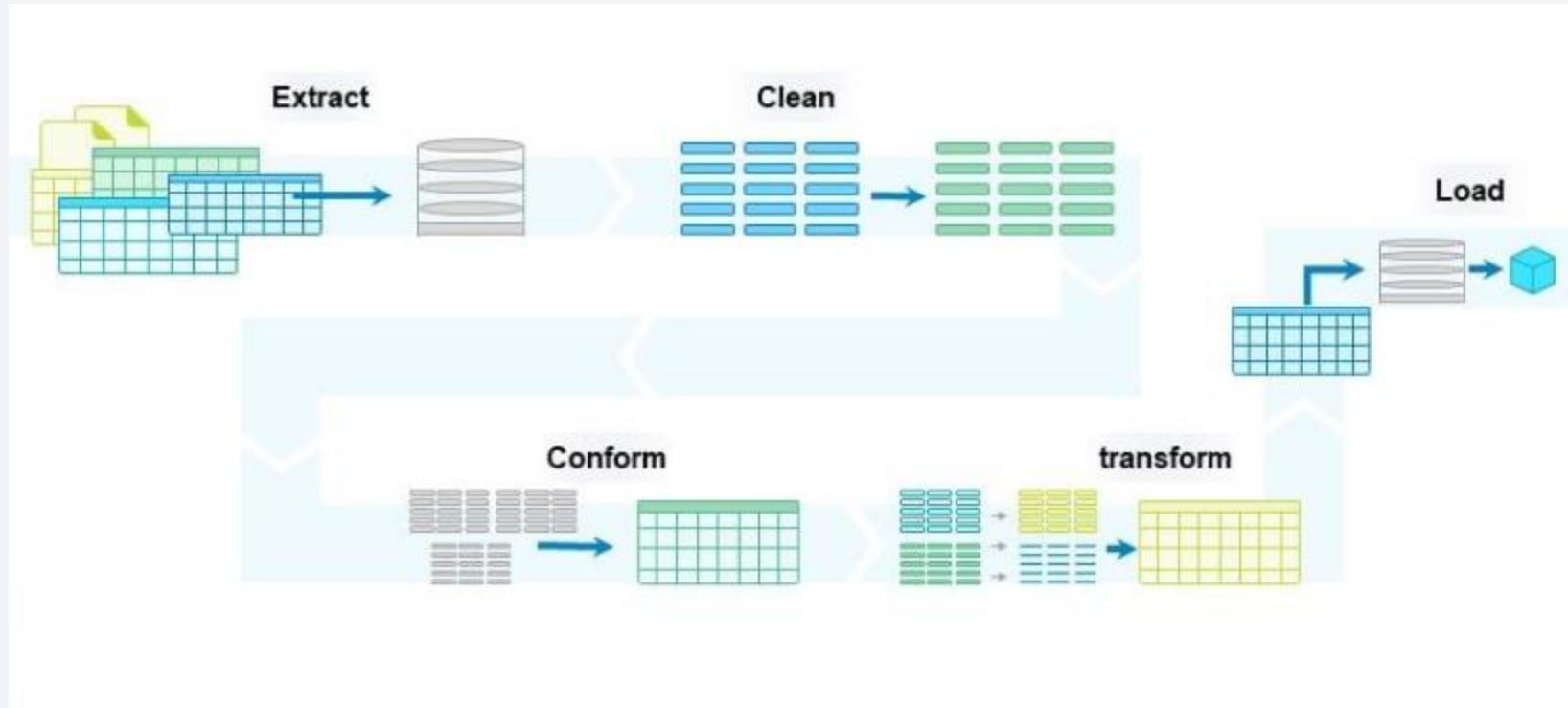
Methodology

Executive Summary

- Data collection methodology:
 - API get request and web scraping
- Perform data wrangling
 - New integer label created, called "Class"
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - Data was standardized, divided in training and test, hyperparameters used, cross validated in GridSearch and applied logistic regression estimator.

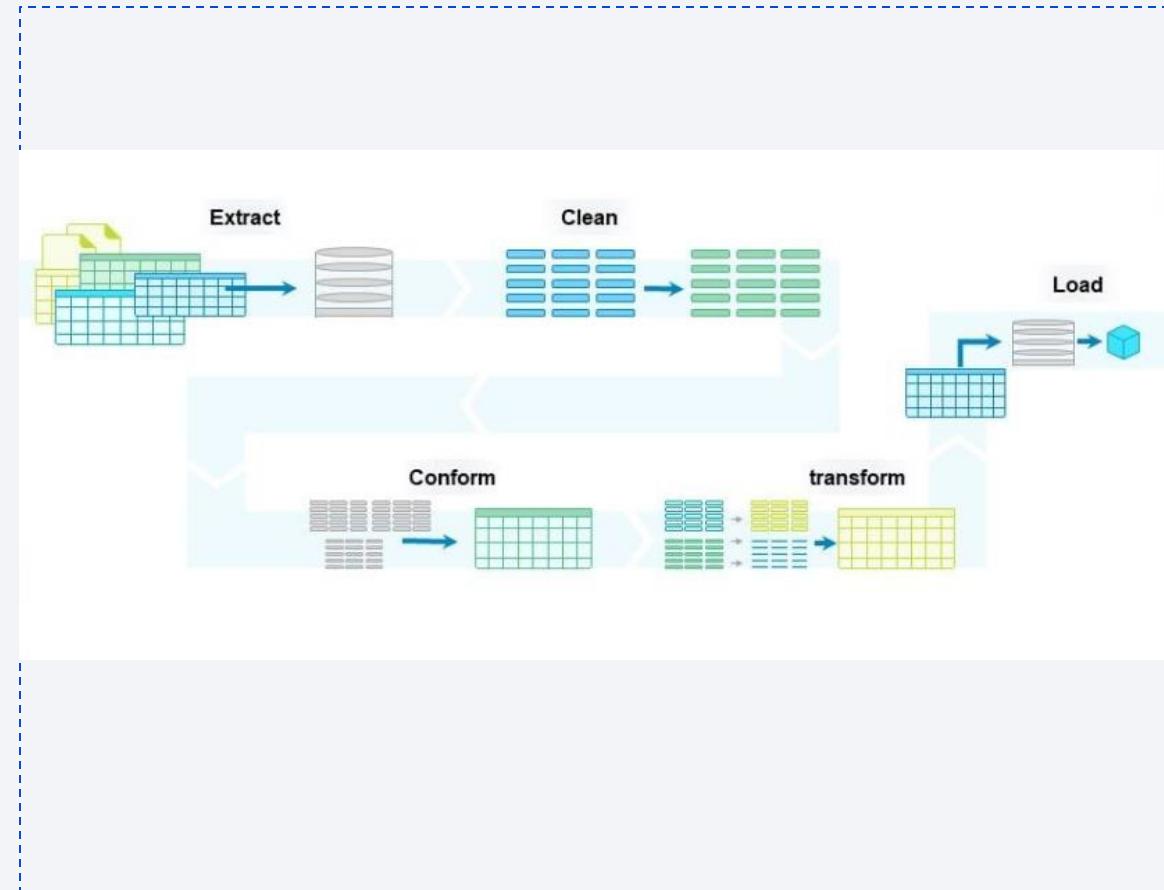
Data Collection

- Data were collected from SpaceX API in "<https://api.spacexdata.com/v4/launches/past>".



Data Collection – SpaceX API

- Data were collected from SpaceX API and converted to a pandas dataframe
- <https://github.com/vagnervale/Data-Science-Course/blob/eef4dab75e10753789342ed989a0137cdcfb2b10/jupyter-labs-spacex-data-collection-api.ipynb>



Data Collection - Scraping

- Data scraped from "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
- <https://github.com/vagnervale/Data-Science-Course/blob/b1b316861bbb6de4f0b77b9426e6f2029f32ecb9/jupyter-labs-spacex-data-collection-api.ipynb>



Data Wrangling

- Payload Mass null data were completed with the mean
- <https://github.com/vagnervale/Data-Science-Course/blob/eef4dab75e10753789342ed989a0137cdcfb2b10/jupyter-labs-spacex-data-collection-api.ipynb>

EDA with Data Visualization

- We plot some charts to get insights about the relationship between some labels and the success of landing
- Charts Plotted:
 - Flight Number Vs Launch Site
 - Payload Mass Vs Launch Site
 - Success rate Vs orbit type
 - FlightNumber Vs Orbit type
 - Payload Mass Vs Orbit type
 - Launch success yearly trend
- <https://github.com/vagnervale/Data-Science-Course/blob/e3e88662a8135eec5b22112e38ac44a055ec5821/edadataviz.ipynb>

EDA with SQL

- SQL queries performed
 - Display the names of the unique launch sites in the space mission
 - Display 5 records where launch sites begin with the string 'CCA'
 - Display the total payload mass carried by boosters launched by NASA (CRS)
 - Display average payload mass carried by booster version F9 v1.1
 - Date when the first successful landing outcome in ground pad was achieved
 - names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
 - Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.
- <https://github.com/vagnervale/Data-Science-Course/blob/94078d7b963bc20c5f1889a91c44092596d4a873/jupyter->

Build an Interactive Map with Folium

- Markers, circles, icon, clusters and lines were added to the map to illustrate launch site positions and distances
- Those objects were added to illustrate launch site positions and distances
- https://github.com/vagnervale/Data-Science-Course/blob/76cd6dc10c8933c1a84584c40093f4be71542fa6/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

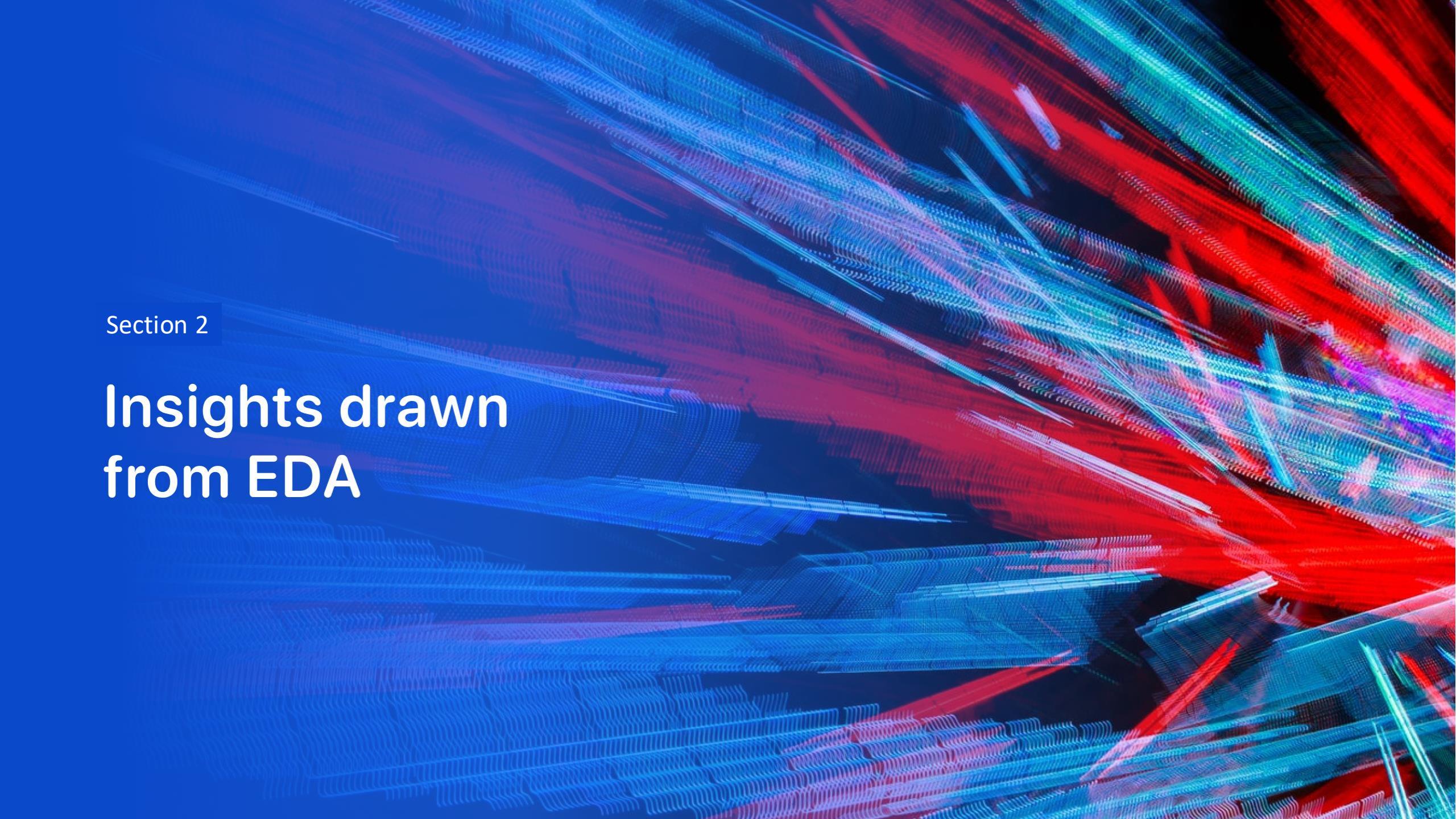
- Two dynamic graphs were added with results depending on the launch site and payload mass selected.
- The first graph was a pie chart representing the success launches. The second was a scatter graph correlation the payload mass and success launch.
- <https://github.com/vagnervale/Data-Science-Course/blob/bdddcb2b2c4afb18de9904546fd94eb48276ca98/spacex-dash-app.py>

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- A target variable Y was created according to the launch success, the features column were standardized (X) and both were split in training and testing sets.
- A GridSearchCV was designed with a estimator, some hyperparameters to find the best and the cross validation was done.
- The accuracies of four estimators (logistic regression, SVM, decision tree and KNN) were compared and their confusion matrix were plotted.
- https://github.com/vagnervale/Data-Science-Course/blob/cafbc2d1bfb6fe595b2072dd2f51bc9d1b2af91/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a 3D wireframe or a network of data points. The overall effect is futuristic and dynamic, suggesting concepts like data flow, digital communication, or complex systems.

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

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth's horizon against a dark blue sky. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper left quadrant, the green and yellow glow of the Aurora Borealis (Northern Lights) is visible.

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

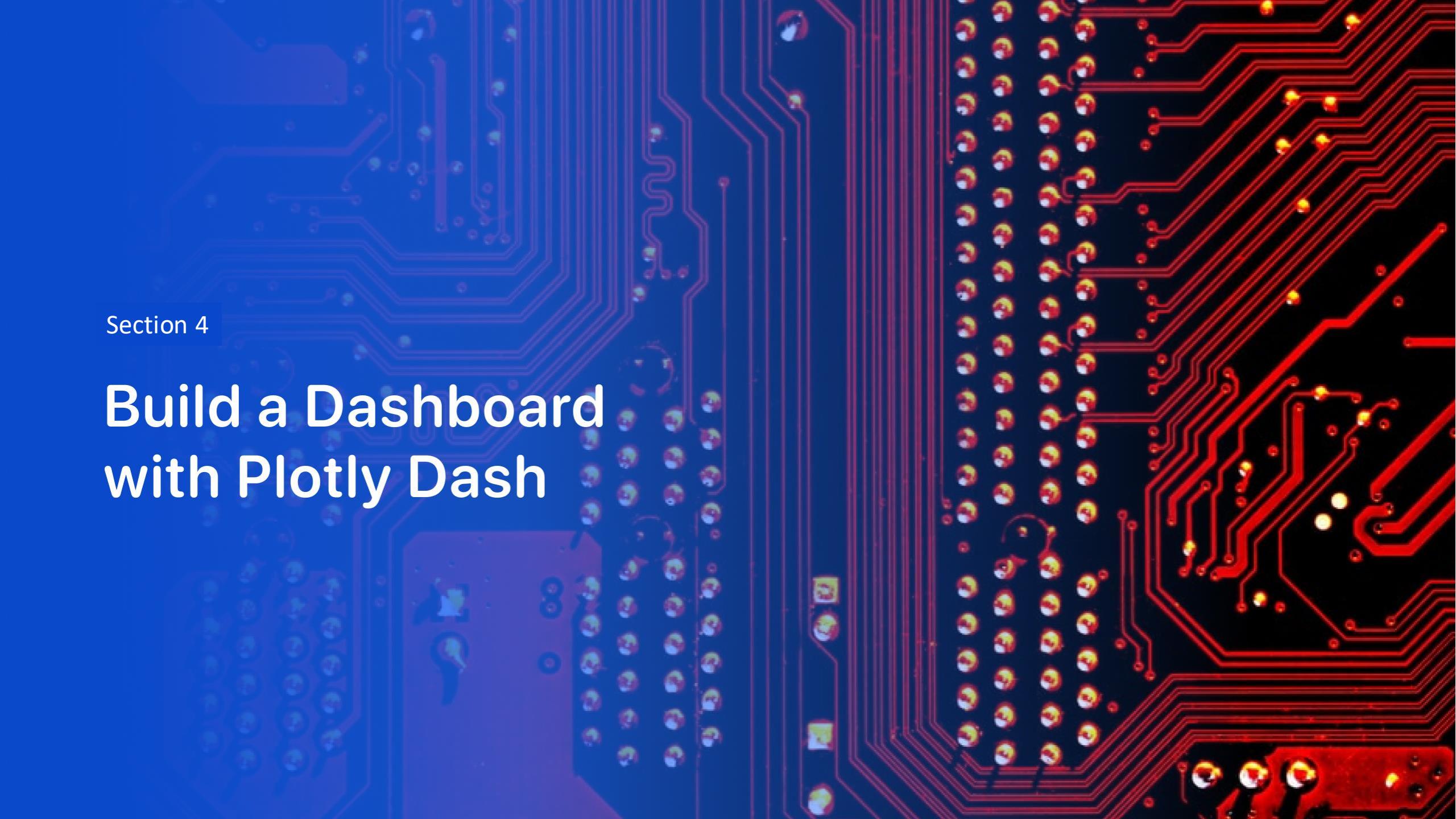
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<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

The background of the slide features a close-up photograph of a printed circuit board (PCB). The left side of the image has a blue color overlay, while the right side has a red color overlay. The PCB itself is dark grey or black, with numerous red and blue printed circuit lines (traces) connecting various components. Components visible include a large blue integrated circuit package at the top left, several smaller yellow and orange components, and a grid of surface-mount resistors on the left edge.

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.

The background of the slide features a dynamic, abstract design. It consists of several thick, curved lines in shades of blue and yellow, creating a sense of motion and depth. The lines curve from the bottom left towards the top right, with some lines being more prominent than others. The overall effect is reminiscent of a tunnel or a high-speed journey through a digital space.

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!

