

# EDA Board

## 1. Problem Analysis

Exploratory Data Analysis (EDA) board to streamline and expedite the often repetitive and time-consuming process of data exploration during data analysis. The objective is to create a basic online platform that significantly reduces the time required for EDA tasks.

Key functionalities of the EDA board include:

- **Identification of Dataset Characteristics:**
  - Extraction of the number of features and columns within the dataset.
  - Identification and visualization of columns containing missing values through a bar graph displaying the percentage of missing values.
- **Datatype Analysis:**
  - Categorization of data into continuous and categorical values.
  - Visual representation of data distribution using a pie chart.
- **Dataset Description:**
  - Description of dataset to provide users with a better understanding of the dataset.
- **Graphical Representations:**
  - Visualizations of the dataset, including:
    - Bar graphs illustrating missing values distribution.
    - Pie chart depicting the proportion of categorical and continuous values.
    - Histograms for a clear visualization of data distributions.
    - Count plot highlighting the distribution of the target variable.
    - Heatmap/Correlation matrix to get the inter-feature relationships.
    - Feature correlation analysis with the dependent variable.

## 2. Business Requirements

- a. **Efficiency:** Streamline and speed up data exploration during analysis using **EDA Board**
- b. **Automation:** Create a platform that significantly reduces the time required for Exploratory Data Analysis (EDA) tasks.
- c. **Comprehensiveness:** Provide a range of functionalities covering **dataset** characteristics identification, **data type** analysis, dataset **description**, and graphical representations.
- d. **User-Friendly Interface:** Design an intuitive online platform for ease of use.
- e. **Visualization:** Enable graphical representations to enhance understanding and interpretation of data based on user entered **features** and display graphs based on different visualizations like **pie chart**, **bar graph**, **histograms**, **count plot**, **heatmap**, **correlation matrix**.
- f. **Insight Generation:** Facilitate the identification of **dataset characteristics**, **missing values**, **data distributions** and correlations to **target variable**.

### Nouns

### Verbs

## 3. Collection of Nouns and Verbs

- a. **Nouns:** EDA board, dataset, visualization, description.
- b. **Verbs:** Features, missing values, data types, pie chart, bar graph, histograms, count plot, heatmap, correlation matrix, dataset characteristics, data distribution, target variable.

## 4. Aggregation of nouns and verbs by topic (Classes with their attributes)

### Classes:

- EDABoard
- Dataset
- Visualization
- Description

### Attributes:

- Dataset: features, missing values, data types
- Visualization: pie chart, bar graph, histograms, count plot, heatmap, correlation matrix
- Description: dataset characteristics, data distribution, target variable

### Associations:

- EDABoard has a Dataset.
- Dataset contains Features and Columns.
- Visualization represents data distribution and correlations.
- Description provides insights into dataset characteristics and distributions.

## 5. Target Audience

- **Students and Educators:** Students studying data science, statistics, or related fields, and educators teaching data analysis courses. They need accessible tools for hands-on learning and teaching exploratory data analysis concepts.
- **Data Analysts:** Professionals responsible for analyzing datasets to derive insights and make data-driven decisions. They require tools that streamline the process of exploring and understanding datasets efficiently.
- **Business Analysts:** Professionals in business roles who analyze data to identify trends, make forecasts, and support strategic decision-making. They require user-friendly tools to explore datasets and communicate insights effectively.

## 6. Rules

- The platform must accurately identify and visualize dataset characteristics.
- It should categorize data into continuous and categorical values effectively.
- The graphical representations should be clear and intuitive for easy interpretation.
- Dataset description must provide sufficient context for users to understand the dataset.
- Feature correlation analysis should be accurate to help users derive insights.
- The platform should be user-friendly and accessible online.

## 7. Challenge questions

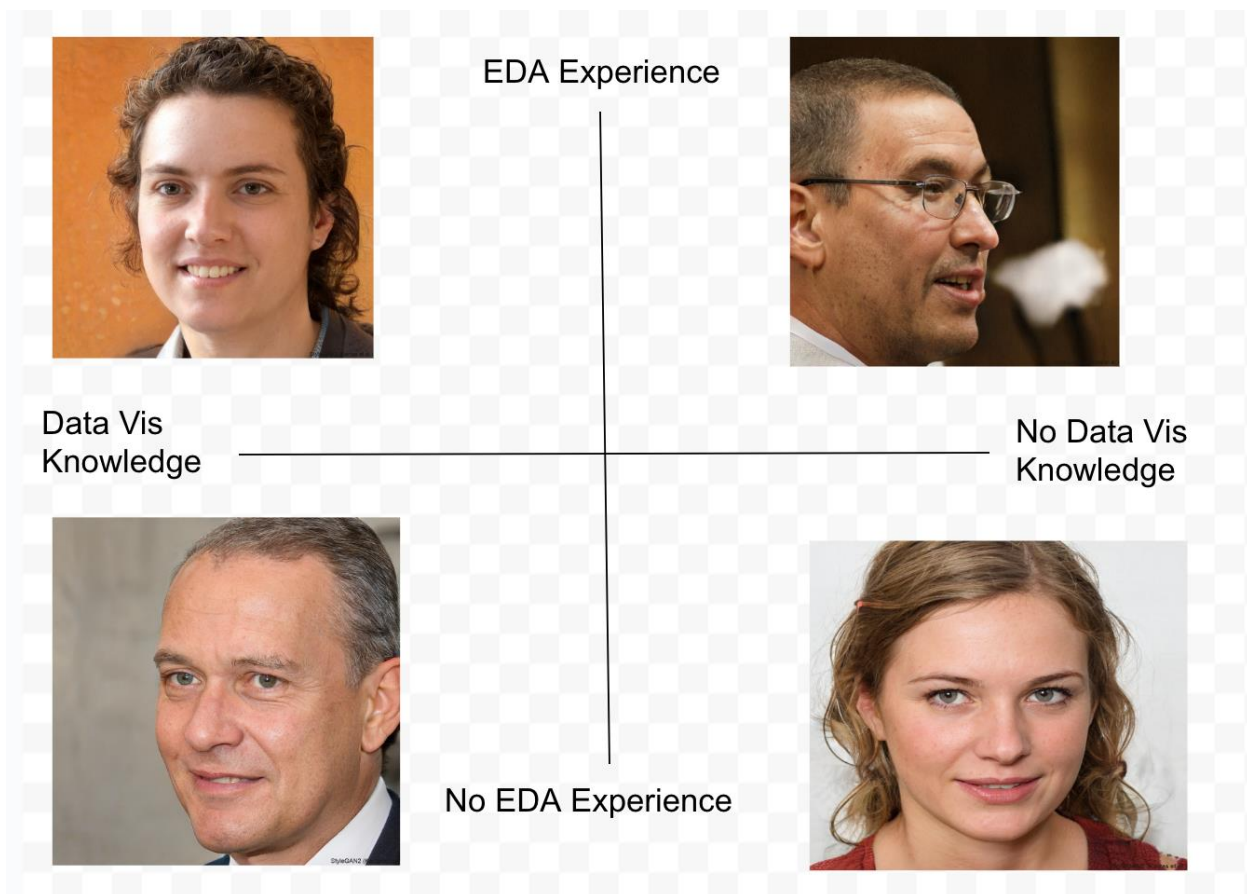
- How can we ensure the accuracy of feature correlation analysis with the dependent variable?
- What measures should be taken to handle large datasets efficiently?
- What visualization techniques can be employed to represent complex data distributions effectively?
- How do we ensure the platform remains user-friendly while incorporating advanced analytical functionalities?

## 8. User stories

- Defining user personas (target users)
  - Students
    - 🔗 As a student like Sam, I want to utilize the EDA platform to analyze datasets for academic projects or research purposes efficiently, considering factors like dataset characteristics, data types, and correlations.
  - Data Analysts
    - 🔗 As a data analyst like Mike, I want to utilize the EDA platform to analyze datasets for various projects or tasks efficiently, focusing on factors like dataset characteristics, data types, and correlations to derive meaningful insights.
  - Business Analysts
    - 🔗 As a business analyst like Dan, I want to leverage the EDA platform to explore datasets related to market trends or consumer behavior, using visualizations like histograms and pie charts to identify patterns and inform business strategies.
  - Educators:
    - 🔗 As an educator like Karen, I want to use the EDA platform to teach students about data analysis concepts effectively, providing hands-on experience with datasets and visualizations to enhance their learning and understanding.
- User persona dimensions

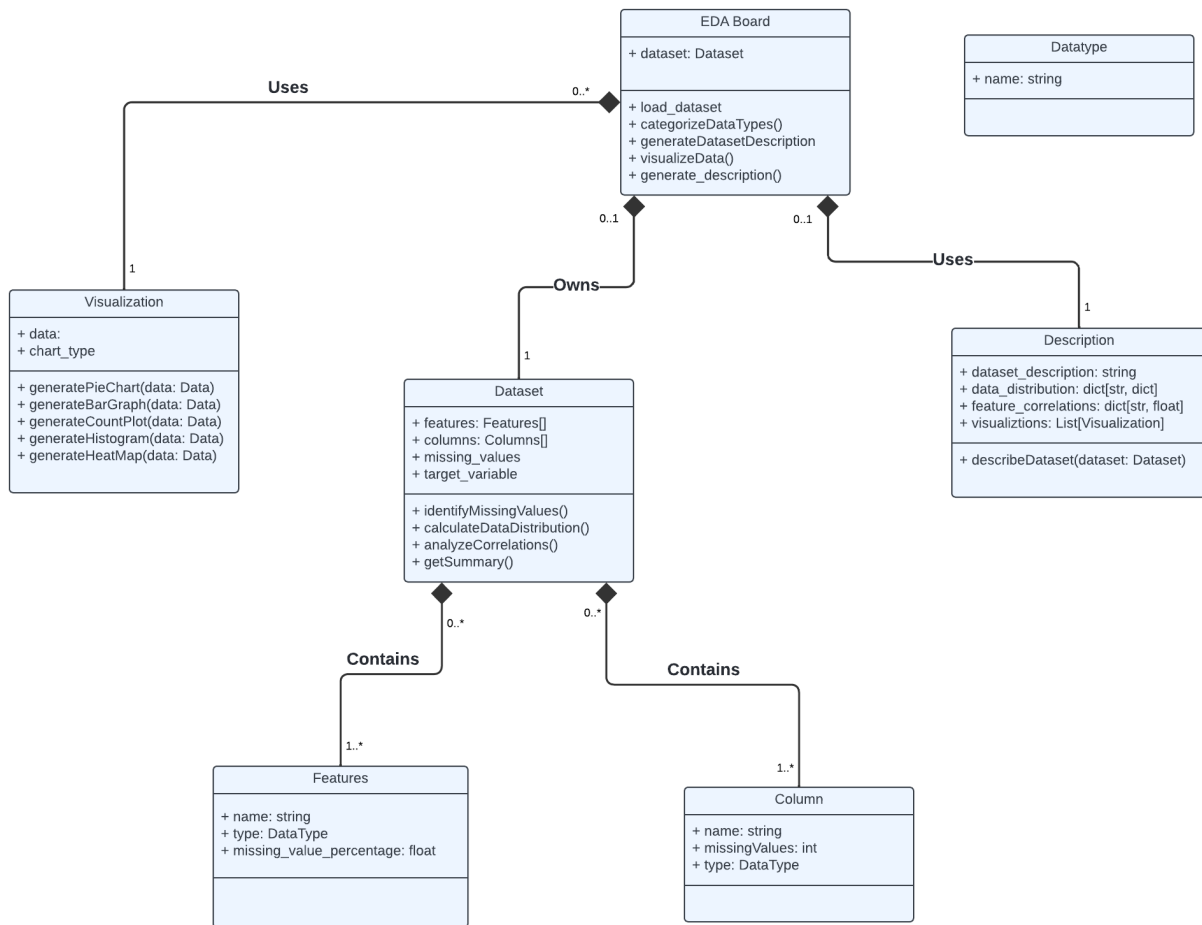
- i. Technical Vs Non-technical
  - ii. Programmer vs non-programmers
  - iii. Employed vs unemployed
  - iv. Highly educated vs no formal education
  - v. Knowledge about Data visualization vs No knowledge about Data Visualization
  - vi. Experience with EDA vs No experience with EDA
- c. Ranking the dimensions
  - i. Experience with EDA vs No experience with EDA
  - ii. Knowledge about Data visualization vs No knowledge about Data visualization

- d. Identify user personas (diagram)



## 9. UML Diagram

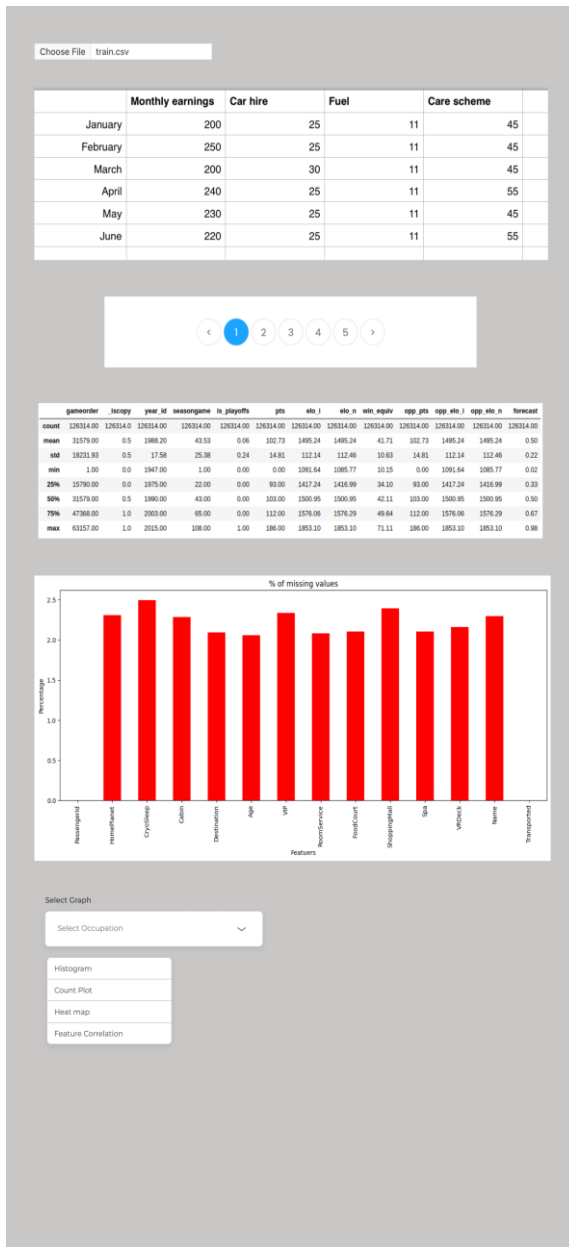
- **EDABoard**: Represents the main controller class responsible for coordinating different functionalities.
- **Dataset**: Represents the dataset being analyzed, containing features and columns.
- **Feature** and **Column**: Represent individual features and columns in the dataset, including their characteristics such as name and data type.
- **Visualization**: Handles the generation of various graphical representations like pie charts, bar graphs, histograms, etc.
- **Description**: Provides methods to generate descriptive information about the dataset.
- **Datatype**: Represents the data types that can be assigned to features and columns.
- **Data**: Represents the data used for visualization and description.



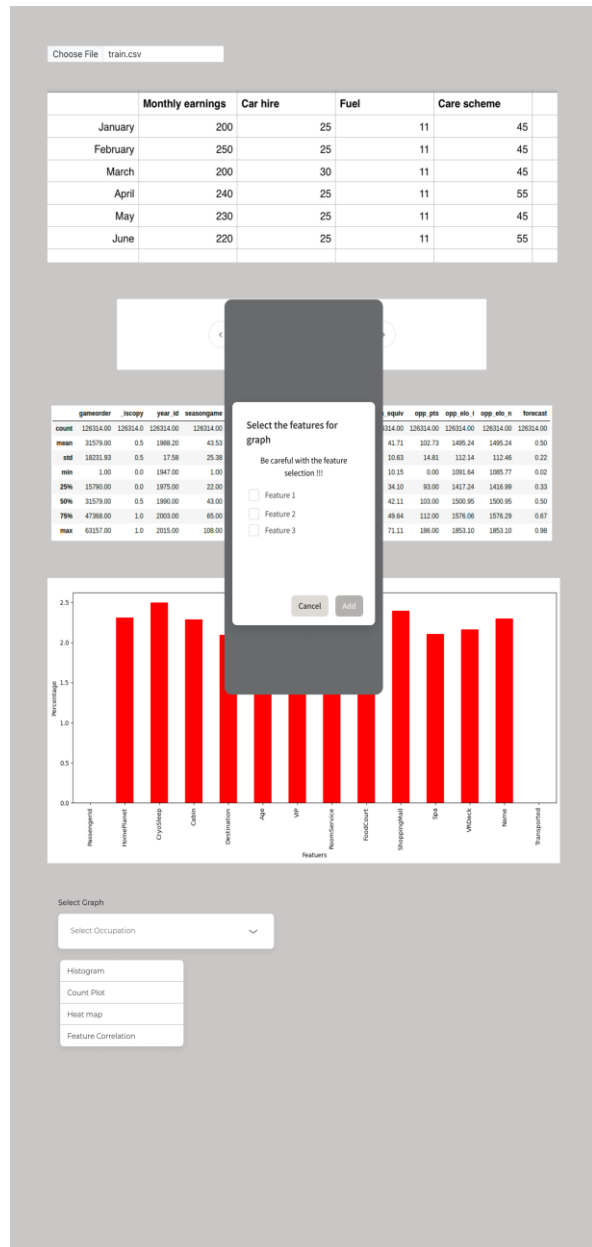
## 10. Low level mockups:

Link to Mockup: <https://www.figma.com/file/tlFJYLh2gmCJEQnfqkYNm/EDA-Board?type=design&node-id=19%3A521&mode=design&t=lkimlAJB37vg3si2-1>

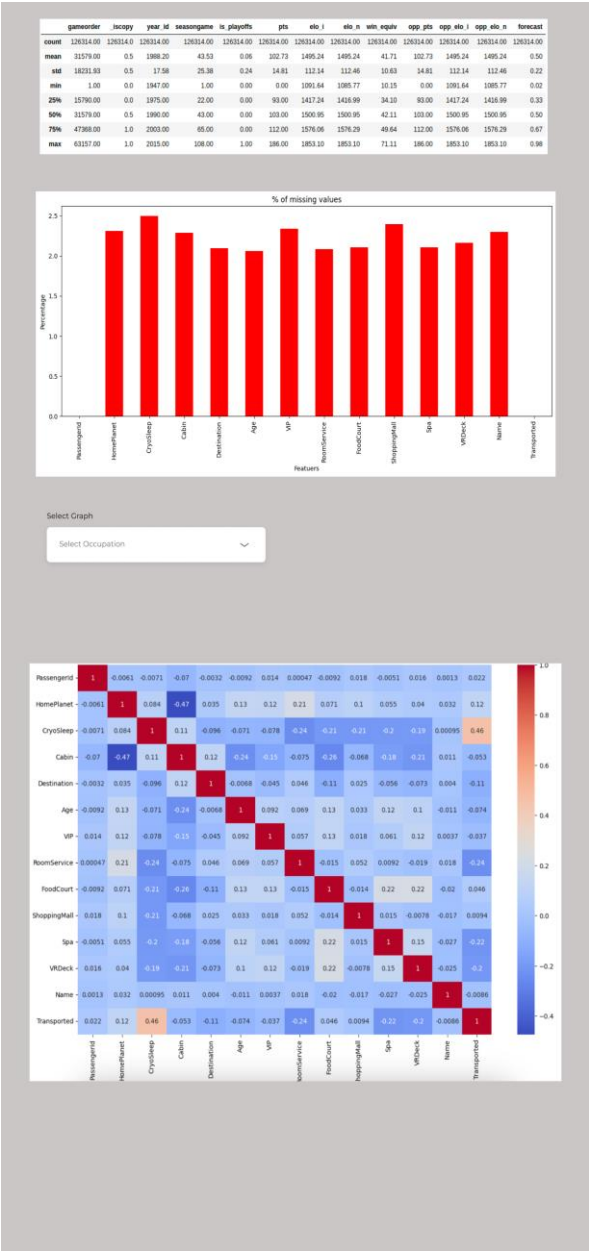
Frame 1:



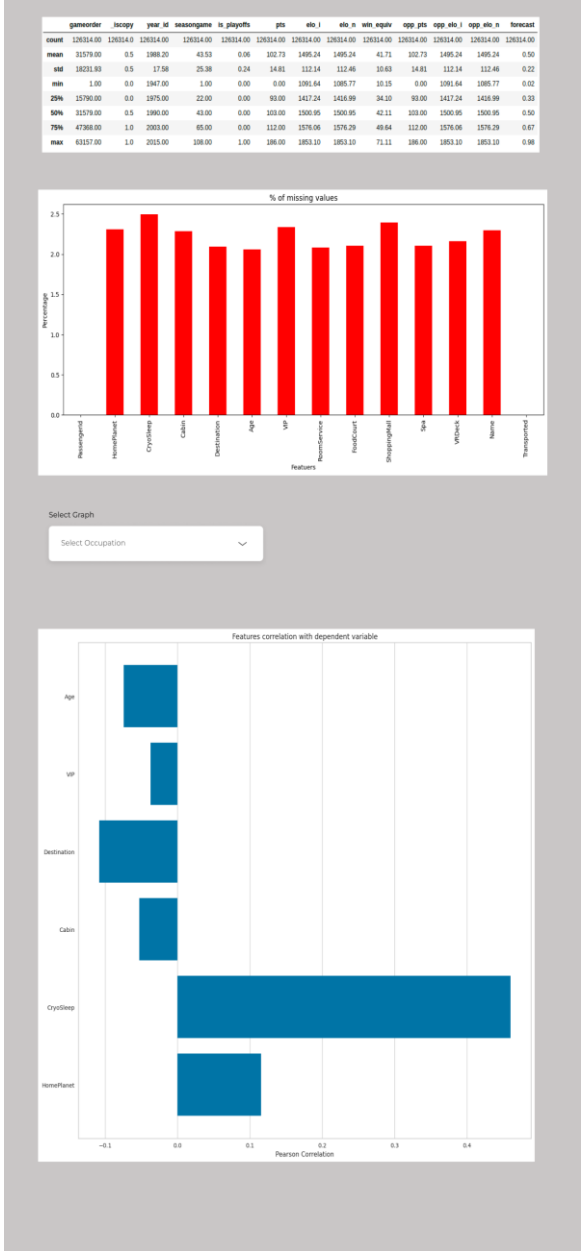
Frame 2:



Frame 3:



Frame 4:



#### Notes:

- The initial screen will only have the choose file button.
- Once the user upload a csv file of the dataset the initial analysis will be performed.
- After the initial analysis the the result as seen in Frame 1 will be displayed.
- The user also has the option to select various different graphs from the drop down menu.
- Once the user selects a graphs a dialog box will be displayed as shown in Frame 2.
- Based on the graph selected by the user the dialog box will show features to select for the particular graph.
- Once the user selects the features and clicks add the an interactive graph will be displayed as shown in Frame 3 and Frame 4.