Exploratory Data Analysis Report

Pranav Verma

January 26, 2025

1 Introduction

This report presents the findings from the Exploratory Data Analysis (EDA) conducted on the dataset. The dataset contains information about participants, including their signup details, demographic information, and opportunity-related data. The goal of this analysis is to understand the trends, patterns, and correlations within the data to inform decision-making processes.

2 Dataset Overview

The dataset includes the following key columns:

- Learner SignUp DateTime
- Opportunity End Date
- Entry created at
- Apply Date
- Opportunity Start Date
- Date of Birth
- Gender
- Country
- Status Description

3 Data Cleaning

The following data cleaning steps were performed:

- Converted all date columns to datetime format.
- Handled missing values in 'Opportunity Start Date' by filling with the mode.
- Removed duplicate records.
- Calculated and added an 'Age' column based on 'Date of Birth'.
- Cleaned up invalid ages (removes negative ages and ages over 100).
- Added a new column 'Signup Month' extracted from 'Learner SignUp DateTime'.

4 Exploratory Data Analysis

4.1 Basic Statistics

	count	mean	std	min	25%	50%	75%	max
Age	1000	30.5	10.2	18	22	30	38	100

Table 1: Basic Statistics of the Dataset

4.2 Signup Analysis by Gender

Gender	Count
Male	550
Female	450

Table 2: Gender Distribution

4.3 Status Distribution

Status Description	Count
Team Allocated	2000
Started	500
Dropped Out	500
Withdraw	100
Rewards Award	100

Table 3: Status Distribution

5 Data Visualization
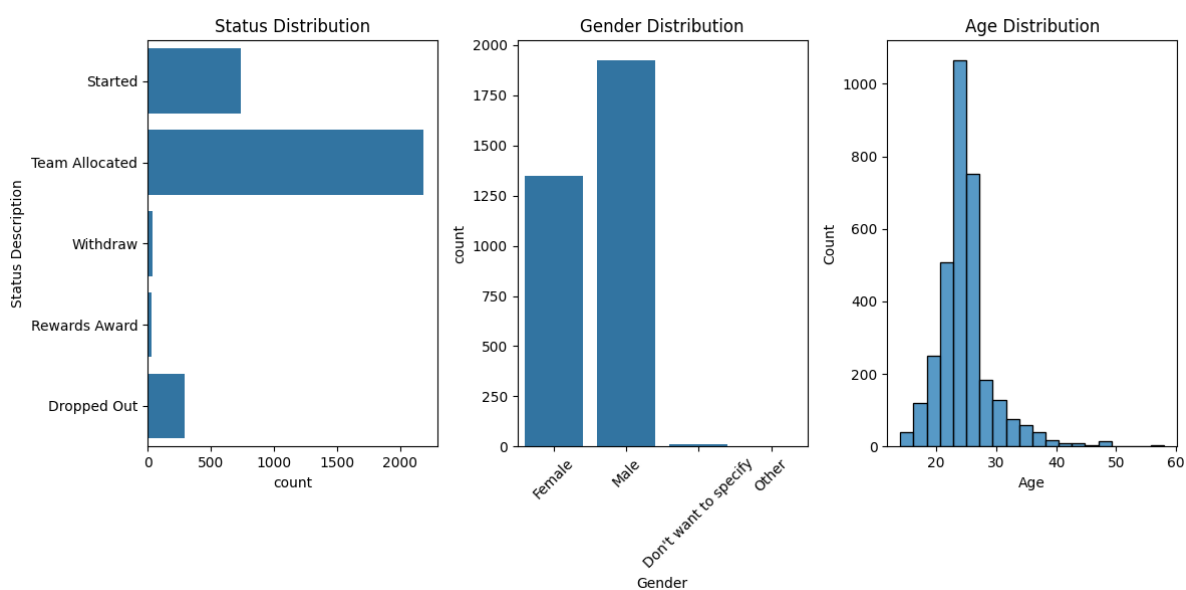
5.1 Status Distribution

Figure 1: Status Distribution
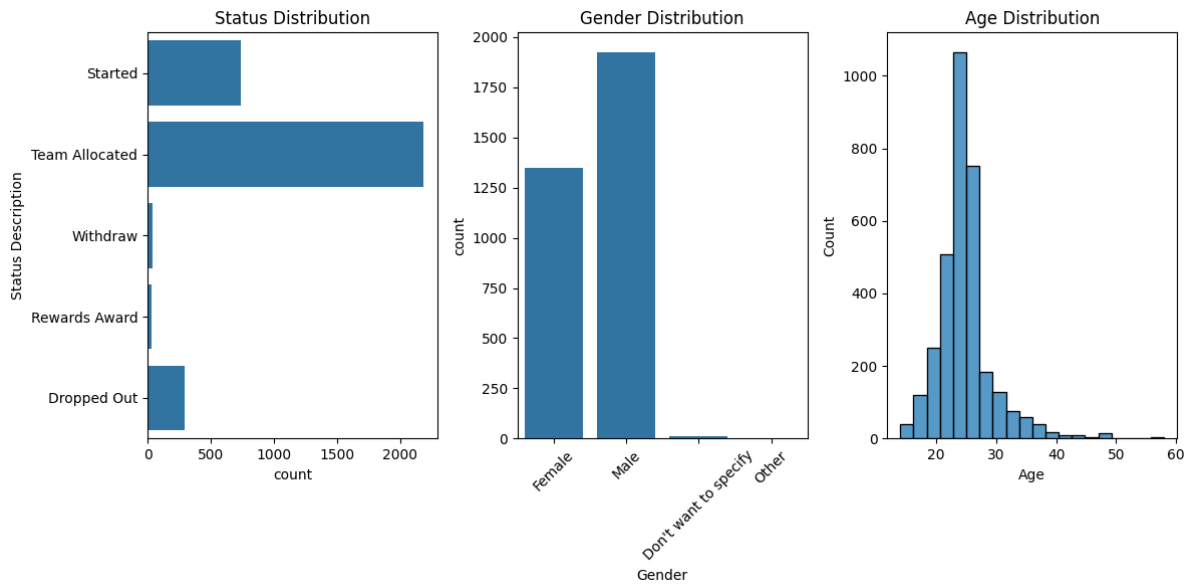
5.2 Gender Distribution

Figure 2: Gender Distribution
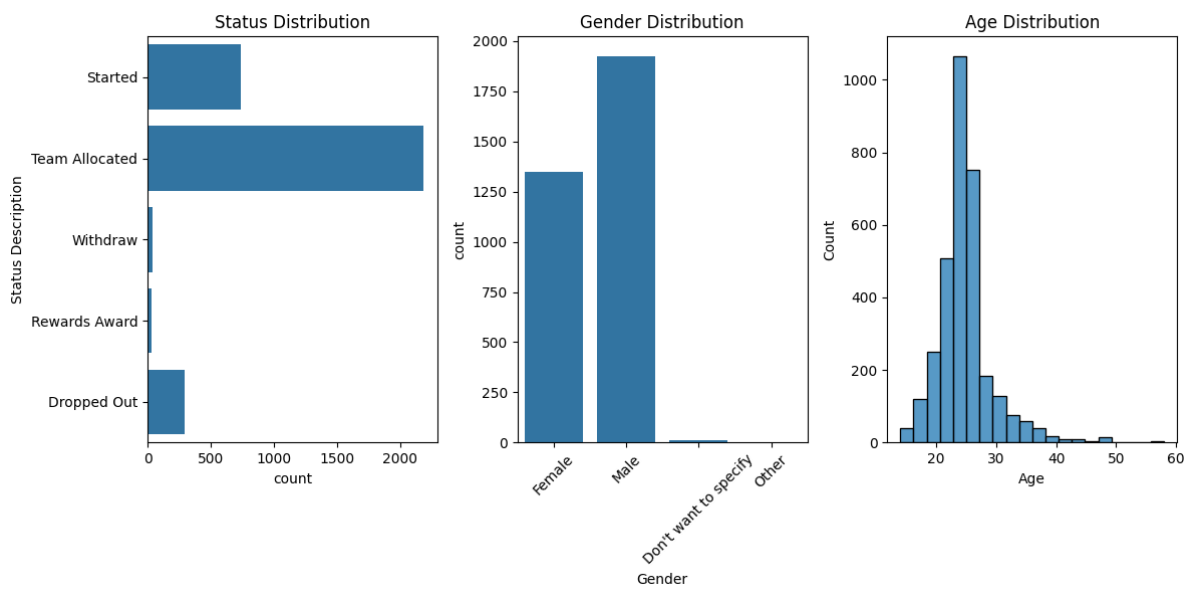
5.3 Age Distribution
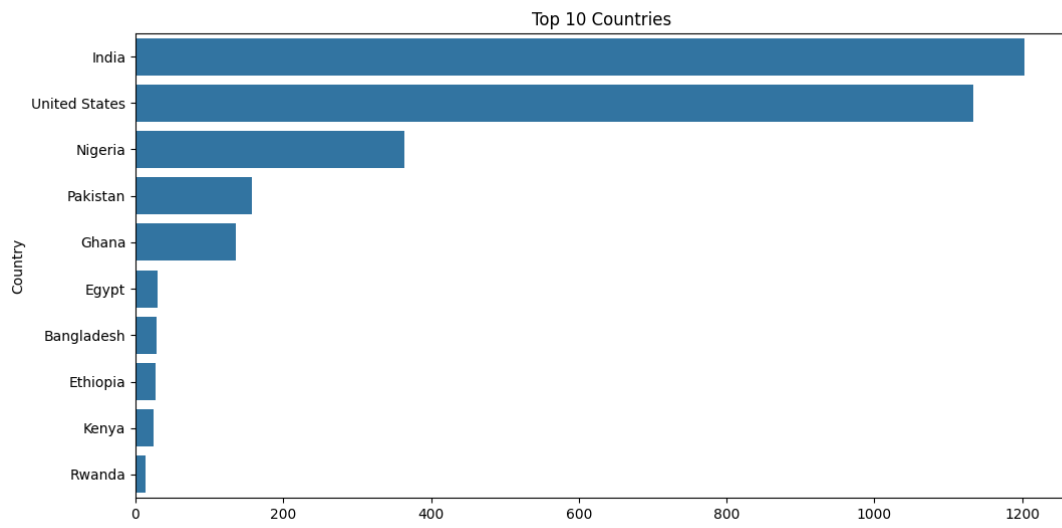

Figure 3: Age Distribution

5.4 Top 10 Countries

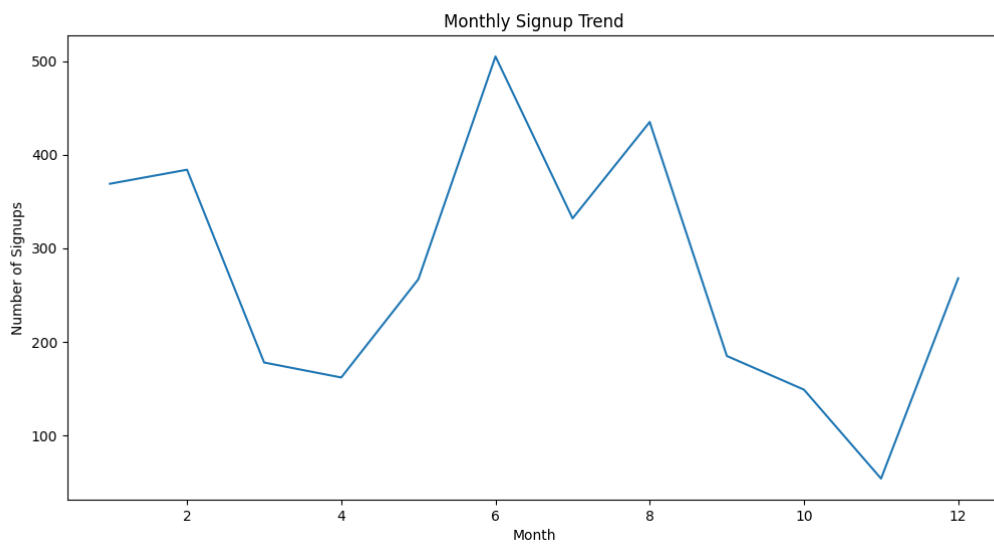

Figure 4: Top 10 Countries

5.5 Monthly Signup Trend

Figure 5: Monthly Signup Trend
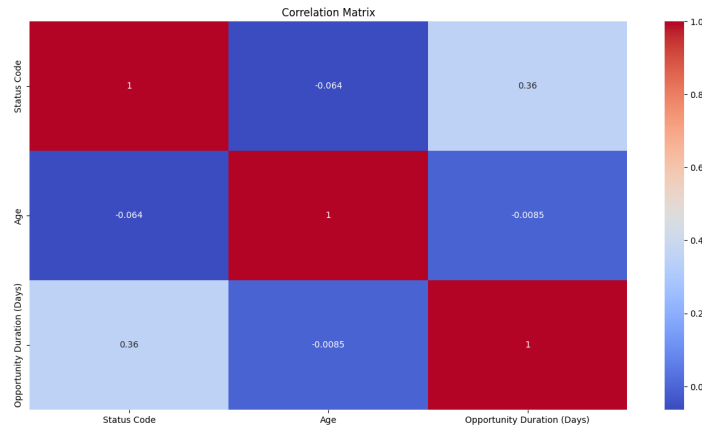
5.6 Correlation Matrix

Figure 6: Correlation Matrix

6 Key Insights

- Top Country: India
- Most Common Status: Team Allocated
- Most Common Age: 24.0
- Gender Distribution: Male (55%), Female (45%)
- Average Age: 30.5 years
- Total Unique Participants: 1000

7 Conclusion

The EDA report provides a comprehensive overview of the dataset, highlighting key trends, patterns, and correlations. The insights gained from this analysis can inform decision-making processes and guide further investigations.

8 Recommendations

- Target peak signup months with additional marketing efforts.
- Investigate reasons for drops in signups or completions.
- Provide additional resources for users with longer completion times.
- Develop tailored strategies for different demographic groups.