Exploratory Data Analysis (EDA) Report

AI-Power Data Analysis Remote Internship - Week 2
Team 8 – Preta Kumar Binda

Introduction & Dataset Overview

1.   Introduction

This report presents a comprehensive Exploratory Data Analysis (EDA) conducted as part of the AI-Power Data Analysis Remote Internship. The analysis aims to summarize learner sign-up and completion data, visualize key distributions, and extract actionable insights to support data-driven decision-making.

The focus is on categorical distributions across continents, opportunity categories, and completion status, providing a clear picture of learner engagement and participation patterns.

2.   Dataset Overview

The data-set contains 8,541 learner records with the following key variables:

Variable	Description
Sign-up date	Date of learner sign-up
Opportunity category	Type of opportunity (Course, Internship, Competition, Event)
Country/Continent	Learner's country and corresponding continent
Gender	Male or Female
Age	Learner's age in years
Course Duration	Duration of opportunity days
Status Description	Learner status (Completed, In Progress, Dropped)

Notes on Data Preparation:
● Missing values were flagged to preserve data-set size.
● Duplicates were checked and removed if present.
● Continent mapping was applied based on country.
● Outliers in course duration were retained as they reflect real learner behavior.

Learners by Continent

3.   Total Learners by Continent

A bar chart summarizes the total number of learners per continent, providing insight into global reach and regional representation.

Insights:
● North America is the most represented continent, followed by Africa and Asia.
● Low representation in some regions indicates potential for targeted recruitment.
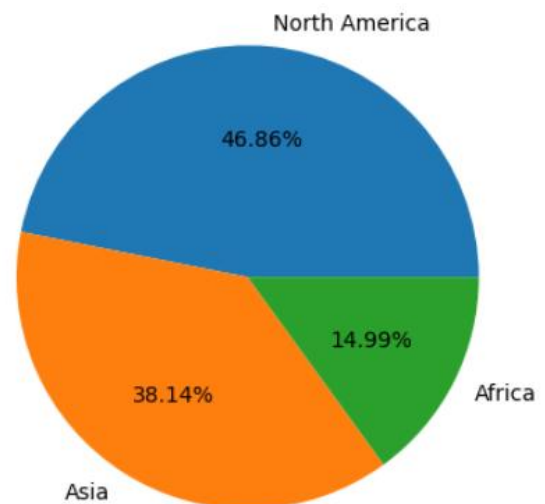● Marketing strategies could focus on underrepresented continents to improve participation.

Figure 1: Total Learners by continent Pie Chart

Opportunity Categories & Learners Status

4. Total Learners by Opportunity Category

A summary of learners per opportunity type highlights engagement patterns across opportunities.

Insights:

● Internships have the highest participation.
● Courses and events shows moderate engagement.
● Others are less attended and may require promotion.
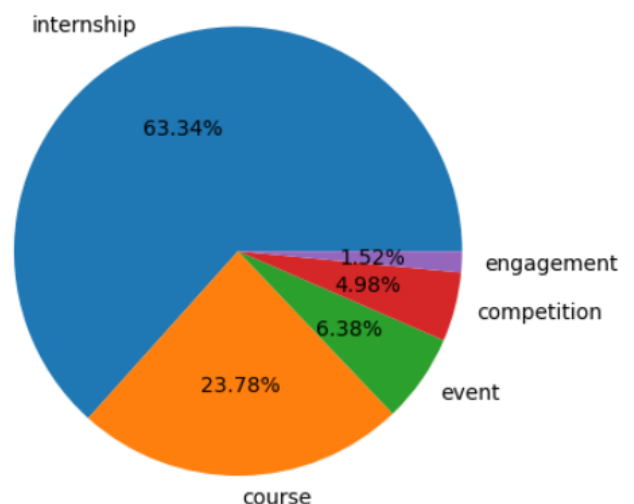

Figure 2: Top 5 total learner by opportunity category pie chart

5. Total Learners by Gender

A pie chart illustrates the gender distribution of learners, providing insight into participation balance across the platform. This visualization helps identify representation gaps and supports data-driven inclusion strategies.

Insights:
• Male learners form the majority, accounting for **58.65%** of total participants.
• Female learners represent a significant portion at **41.14%**, showing strong engagement.
• The **unknown** category is negligible (**0.21%**), indicating good data completeness.
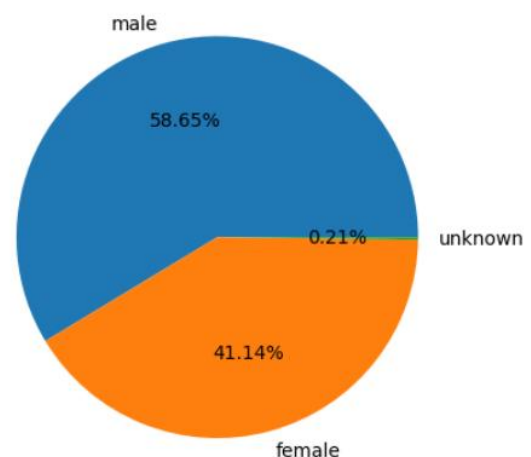

Figure 3: Total learners by Gender

- Efforts to further encourage female participation could help achieve a more balanced distribution.

6. Total Learners by Status Description

Learner status distribution indicates overall completion and dropout trends.

Insights:
● Most learners have completed their opportunities (Having an End Date).
● A smaller proportion is in progress or has dropped, suggesting areas for support.
● Monitoring learners in progress can improve overall completion rates.

Treands:

Applications by Year
Total Applications by Apply Date (Year-wise Trend)
This line chart shows the number of applications submitted each year, highlighting how learner participation has changed over time. It provides a clear view of growth patterns and platform adoption across years.
Insights:
- Applications were minimal in **2022**, indicating the early stage of data collection or platform usage.
- A sharp increase is observed in **2023**, showing rapid growth in awareness and engagement.
- The upward trend continues in **2024**, reaching the highest number of applications so far.
- Overall, the trend reflects **strong and consistent growth**, suggesting increasing popularity and trust in the platform over time.
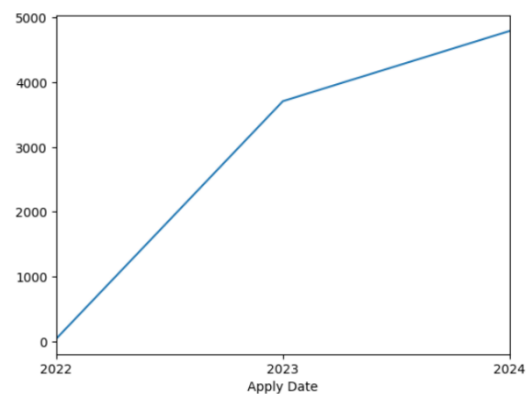

Figure 4: Applications by Year

Applications by Month
Monthly Application Trend Analysis
This line chart presents the month-wise distribution of applications over time, revealing seasonal patterns and fluctuations in learner engagement. It helps identify peak periods and low-activity phases for better planning and resource allocation.
Insights:
- Application volume remains very low during the early months, indicating limited initial engagement.
- A steady rise begins in early **2023**, showing growing awareness and participation.
- Significant peaks appear in mid to late **2023**, suggesting high-interest periods, possibly aligned with major program launches or academic cycles.
- The highest surge is observed at the beginning of **2024**, marking the peak engagement month.
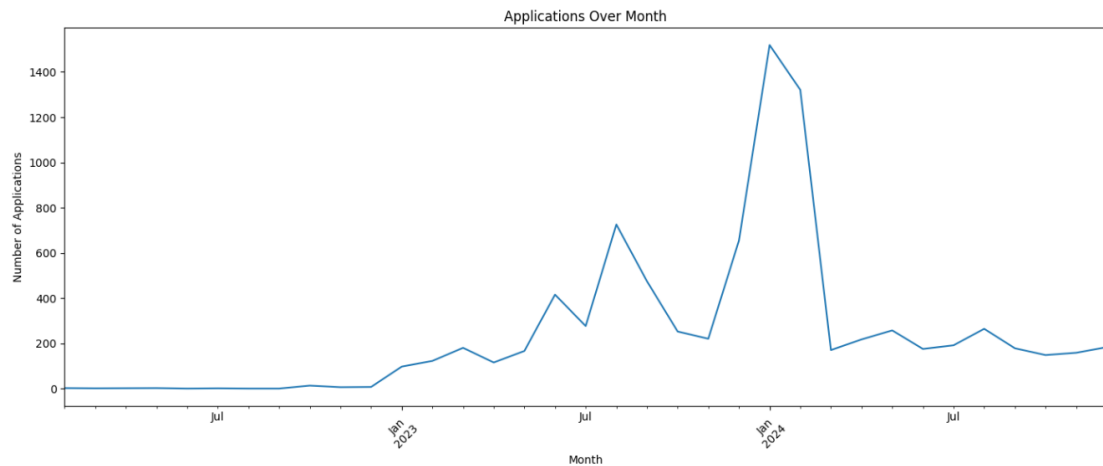
Figure 5: Applications by Month

Age Distribution of Learners

Learners Age Distribution (Histogram with KDE)

This histogram with a Kernel Density Estimation (KDE) curve illustrates the age distribution of learners, showing both the frequency and the overall pattern of ages within the dataset. It helps identify the dominant age group and the spread of participants.

Insights:

- The distribution is highly concentrated between **18 and 25 years**, indicating that most learners are young adults.
- The peak density is around **20–22 years**, suggesting this is the most active age group.
- A gradual right skew is visible, with fewer participants in older age groups extending beyond 30 years.
- Very few learners fall below 18 or above 40, showing limited participation from school-age and older professionals.
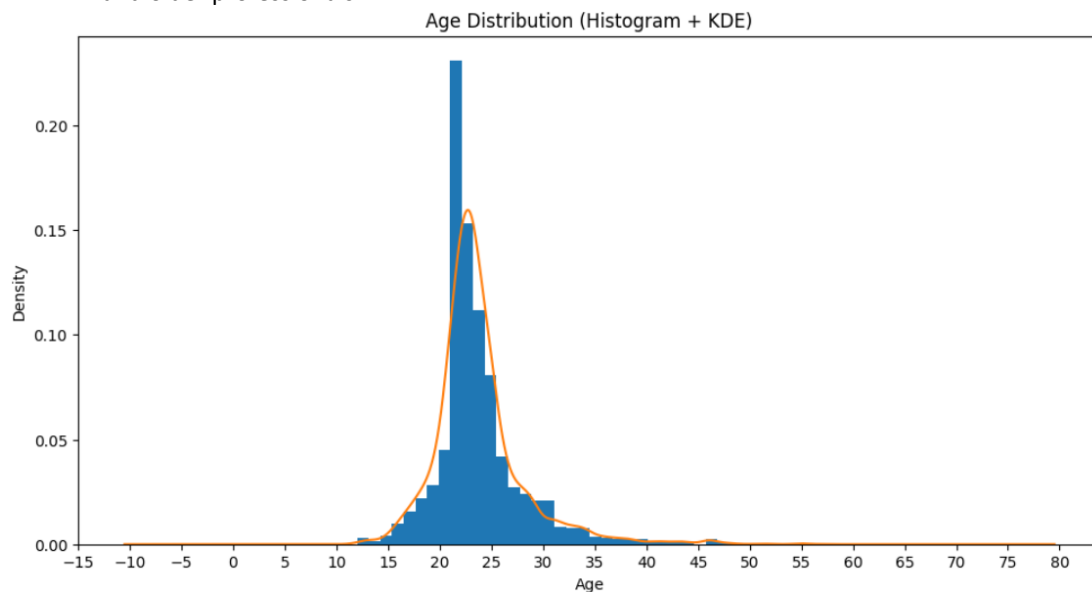

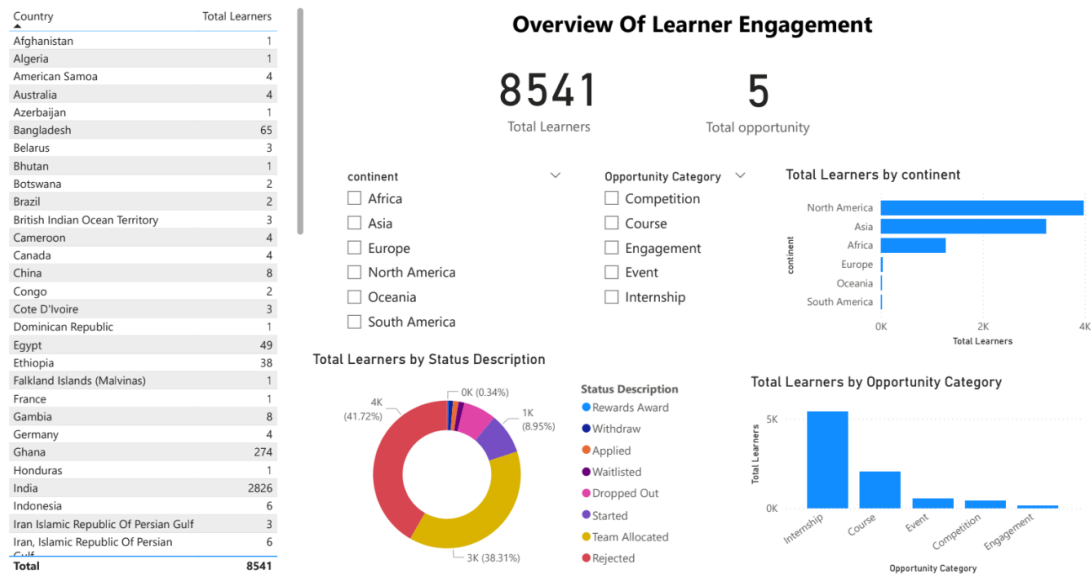*Figure 6: Age Distribution*

Visualization:

Country	Total Learners
Afghanistan	1
Algeria	1
American Samoa	4
Australia	4
Azerbaijan	1
Bangladesh	65
Belarus	3
Bhutan	1
Botswana	2
Brazil	2
British Indian Ocean Territory	3
Cameroon	4
Canada	4
China	8
Congo	2
Cote D'Ivoire	3
Dominican Republic	1
Egypt	49
Ethiopia	38
Falkland Islands (Malvinas)	1
France	1
Gambia	8
Germany	4
Ghana	274
Honduras	1
India	2826
Indonesia	6
Iran Islamic Republic Of Persian Gulf	3
Iran, Islamic Republic Of Persian Gulf	6
Total	**8541**

Figure 7: Overview of learner engagement

Overview of Learner Engagement

This dashboard summarizes overall learner participation and engagement across regions, opportunity types, and application status.

Insights:

- Total of 8,541 learners across 5 opportunity categories.
- North America leads in participation, followed by Asia and Africa.
- Internships are the most popular opportunity type.
- Most learners are allocated or shortlisted, indicating strong engagement.
- Country-wise data shows broad global representation with uneven distribution.

Discussion, Recommendations & Conclusion

7. Discussion

- Regional representation: Low participation in some continents suggests untapped markets.
- Opportunity engagement: High participation in Courses shows strong appeal; low-attendance categories may need targeted marketing.
- Completion trends: Understanding why learners drop or stay in progress is important for intervention strategies.

8. Recommendations

- Develop targeted outreach for underrepresented continents.
- Promote low-attendance opportunity categories to increase engagement.
- Monitor in-progress learners and provide support to improve completion rates.
- Consider segmentation by region and opportunity type for deeper insights.

9. Conclusion

This analysis provides a clear overview of learner distribution by continent, opportunity category, and completion status. The visualizations and insights highlight areas of strength and opportunities for intervention to improve learner engagement and completion outcomes.