Task 5: Exploratory Data Analysis (EDA)

Objective:-

The goal of this task was to extract insights from the employee dataset using visual and statistical exploration. The analysis was conducted in Python using **Pandas, Matplotlib, and Seaborn**.

Data Overview:-

- The dataset contains 4653 records and 10 columns.
- Data types: a mix of numerical (e.g., Age, PaymentTier, Experience) and categorical (e.g., Education, City, Gender).
- No missing values or duplicates were found.
- .describe() showed that the average age is ~29 years and average domain experience ~3 years.

Univariate Analysis:-

- Age (Histogram): Most employees are aged 24–30 years, with fewer in higher age groups.
- Experience in Current Domain (Boxplot): The majority have 2–4 years of experience, though some outliers indicate highly experienced professionals.
- Education & Payment Tier (Count Plots): Most employees hold a Bachelor's degree and are in Payment Tier 3.

Bivariate & Multivariate Analysis:-

- Pairplot: Highlighted relationships between numerical variables, showing clustering of employees by age and experience.
- Correlation Heatmap: Confirmed positive correlations between Age and Experience, while other variables had weaker relationships.
- Scatterplots: Revealed clear grouping patterns among employees across cities and education levels.

Key Insights:-

- The dataset is dominated by younger employees in early stages of their careers.
- Domain experience is limited for most, though a small group of senior employees exists as outliers.
- Strong relationships (e.g., Age ↔ Experience) suggest logical consistency in the dataset.

- Workforce distribution shows higher representation in Tier 3 payment groups and among Bachelor's graduates.

Conclusion:-

The EDA provided valuable insights into workforce demographics, experience levels, and variable relationships. The data reflects a young workforce with moderate experience, complemented by a minority of senior professionals. These findings establish a strong foundation for predictive modeling, employee attrition studies, and HR analytics.