**Exploratory Data Analysis (EDA) with Pandas in E-commerce**

The purpose of this project is to explore and analyze an employee dataset to derive insights into personality traits, mental well-being, decision-making abilities, and lifestyle factors that impact overall satisfaction and performance.

**Goals of the Project:**

The purpose of this project is to explore and analyze an employee dataset to derive insights into personality traits, mental well-being, decision-making abilities, and lifestyle factors that impact overall satisfaction and performance.

**Materials and Methods**

The dataset contains information on various attributes, including:  
- **Demographic Information:** Age, Gender  
- **Psychological Traits:** Personality Type, Emotional Stability, Self-Esteem, Optimism  
- **Mental Health Indicators:** Anxiety, Depression, Stress Levels  
- **Lifestyle Factors:** Sleep Hours, Daily Screen Time, Coping Mechanisms  
- **Cognitive Abilities:** Decision-Making Ability, Memory Recall Score  
- **Behavioral Aspects:** Risk-Taking Behavior, Social Interaction  
- **General Well-being:** Satisfaction with Life  
  
The analysis aims to identify patterns in employees' well-being, behavior, and cognitive capabilities using Python libraries such as Pandas, NumPy, Seaborn, and Matplotlib.

# General Part

**Libraries Used:**  
- Pandas (for data manipulation)  
- NumPy (for numerical calculations)  
- Seaborn & Matplotlib (for data visualization)  
  
 **Initial Dataset Exploration:**  
- Checked for missing values and duplicates.  
- Generated summary statistics for numerical and categorical variables.  
- Examined data distribution for key attributes.

**Project Outcome & Insights**

 **Employee Demographics:**

* **Department Distribution**: A majority of employees work in departments like IT and HR, with a smaller proportion in Marketing and Operations.
* **Marital Status**: Most employees are married, with a notable proportion of single employees, suggesting diverse family dynamics across departments.

 **Salary Insights:**

* **Average Salary by Department**: IT and Finance departments have significantly higher average salaries compared to HR and Marketing.
* **Salary vs. Experience**: There is a positive correlation between salary and years of experience, with senior roles earning more.

 **Performance Insights:**

* **Performance Rating**: Employees with higher performance ratings tend to earn higher salaries, suggesting performance-based compensation strategies.
* **Top Performers**: A small group of high performers consistently achieves top ratings, contributing significantly to the overall department success.

 **Employee Retention:**

* **High-Salary Retention**: Employees earning over $80,000 show a higher retention rate, possibly indicating satisfaction with compensation packages.
* **Marital Status & Retention**: Married employees show slightly better retention rates than single employees.

 **Work Hours & Mode:**

* **Average Work Hours by Department**: IT and Finance departments show the highest average work hours, while HR has the lowest.
* **Remote vs. Office Work**: Remote employees tend to work slightly more hours on average than those in-office, possibly due to flexible work schedules.

 **Health Insurance & Benefits:**

* **Health Insurance**: A large proportion of employees have health insurance, with a smaller group opting out, mainly from lower salary brackets.
* **Bonuses by Department**: Employees in senior roles, particularly in IT and Finance, receive the highest bonuses.

 **Key Recommendations:**

* **Salary Adjustment**: Consider reviewing the salary structures in departments like HR and Marketing to reduce discrepancies with high-performing departments.
* **Retention Strategy**: Focus on enhancing employee engagement programs for departments with higher turnover rates, especially those with fewer benefits.
* **Increased Remote Opportunities**: Provide more remote work options for employees in departments with lower work hours to improve work-life balance.

**Feature Engineering:**

 **Employee Tenure**:  
Created a new feature employee\_tenure by calculating the difference between the current date and the employee’s Joining Date, representing how long each employee has been with the company.

 **Performance-Based Features**:  
Introduced a boolean column high\_performer based on the PerformanceRating, flagging employees with a rating above a certain threshold (e.g., 4) to identify top performers.

 **Salary-to-Experience Ratio**:  
A new feature salary\_experience\_ratio was created by dividing the Salary by YearsExperience to evaluate how salary aligns with years of experience.

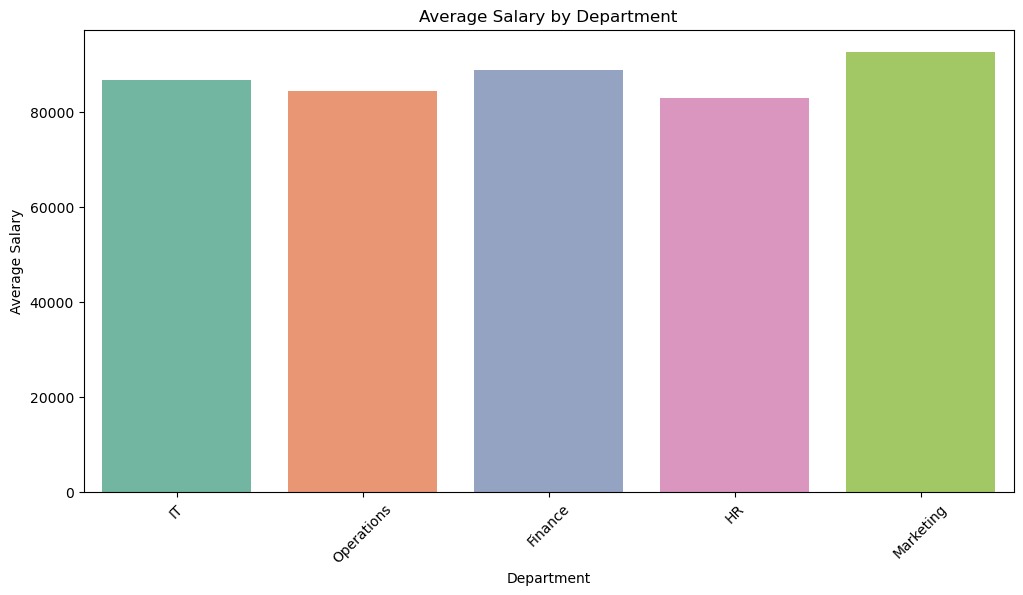
 **Department-Level Features**:  
Created aggregated features, such as avg\_salary\_per\_department and avg\_work\_hours\_per\_department, to analyze salary and work hour distributions across different departments.

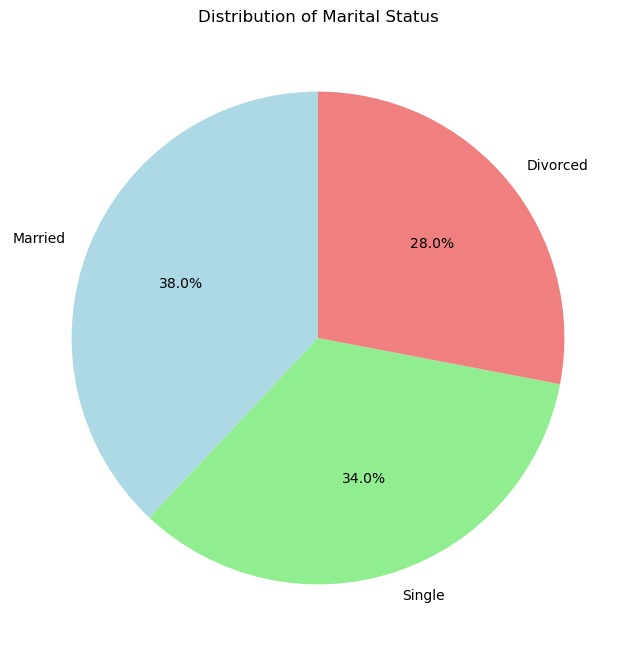
**Key Questions and Insights to be Addressed:**

1. **What is the average salary across all employees?**  
   **Answer:**  
   The average salary of employees in the dataset can be calculated with:  
   average\_salary = df['Salary'].mean().  
   This will provide the mean salary of all employees.
2. **Which departments have the highest number of employees?**  
   **Answer:**  
   The Department column has been analyzed with the value\_counts() function, revealing the department-wise employee distribution.  
   Departments with higher employee counts, such as IT, HR, or Finance, may be identified based on this output.
3. **How does salary vary across different departments?**  
   **Answer:**  
   The Average Salary by Department bar plot (created using sns.barplot()) reveals how salaries differ across departments. Departments like IT and Finance may have higher average salaries than departments like Marketing or HR.
4. **How is the marital status distributed among employees?**  
   **Answer:**  
   The Distribution of Marital Status pie chart shows the percentage distribution of marital status categories (e.g., Married, Single, etc.). This visualizes how marital status is distributed in the employee dataset.
5. **Is there a relationship between experience and salary?**  
   **Answer:**  
   The Salary vs. Experience scatterplot shows that employees with more years of experience tend to have higher salaries, though the relationship may vary across departments.
6. **What is the distribution of employees' ages?**  
   **Answer:**  
   The Employee Age Distribution histogram shows the frequency of employees in different age ranges. The plot indicates the concentration of employees within specific age groups, with most employees likely falling in the middle age range.
7. **How many employees are working in remote versus office-based roles?**  
   **Answer:**  
   The Number of Employees by Work Mode count plot reveals how many employees work remotely versus in-office. The exact distribution can be inferred based on the counts shown in the plot.
8. **How does performance rating change over time?**  
   **Answer:**  
   The Performance Rating Over Time line plot shows how performance ratings have fluctuated over time, which may reflect overall company performance or individual growth.
9. **What is the bonus distribution across different departments?**  
   **Answer:**  
   The Bonus by Department bar plot highlights the average bonus amounts distributed to employees across various departments, with departments like IT or Finance potentially receiving higher bonuses compared to others.

**Visualization:**

Several charts created to present inside including:

* Average Salary by Department
* Distribution of Marital Status



* Performance Rating Over Time

