AUCA Innovation Center:

Data science course

Employee Data Analysis Project

Project Overview

You will analyze an employee dataset to gain insights into workforce demographics, satisfaction levels, and compensation patterns. The dataset contains information about employee age, environment satisfaction, gender, marital status, and monthly income.

Learning Objectives

- Perform data cleaning and preprocessing
- Apply statistical analysis techniques
- Create meaningful data visualizations
- Draw business insights from data analysis
- Practice Python programming with real-world data

Part 1: Data Exploration and Cleaning

Tasks:

1. **Load and examine the dataset**
    - How many records and features are in the dataset?
    - What are the data types of each column?
    - Are there any missing values?
2. **Data Quality Assessment**
    - Identify potential outliers in the MonthlyIncome column
    - Check for any unusual or impossible values in the Age column
    - Examine the distribution of categorical variables (Gender, MaritalStatus)
3. **Data Cleaning**
    - Handle any outliers you identified (explain your chosen method)
    - Create age groups (e.g., 20-30, 31-40, etc.)

- Ensure all categorical variables are properly formatted

Part 2: Descriptive Statistics and Visualization

Tasks:

1. **Basic Statistics**
   - Calculate mean, median, and standard deviation of MonthlyIncome
   - Find the average age of employees
   - Determine the most common MaritalStatus
2. **Data Visualization**

   Create the following visualizations with proper titles, labels, and legends:

   - Histogram of Age distribution
   - Box plot of MonthlyIncome by Gender
   - Bar chart of EnvironmentSatisfaction levels
   - Combined visualization showing income distribution across different MaritalStatus categories

Part 3: Business Insights and Recommendations

Tasks:

1. **Key Findings**
   - Identify three key insights from your analysis
   - Support each insight with specific data points and visualizations
2. **Recommendations**
   - Provide two actionable recommendations based on your analysis
   - Explain how these recommendations could benefit the organization

Deliverables

1. Python code with comments explaining your analysis
2. A report (maximum 2000 words) containing:
   - Executive summary
   - Methodology
   - Results and visualizations
   - Key findings and recommendations
3. Presentation slides (5-7 slides)

Evaluation Criteria

- Code quality and documentation
- Proper application of statistical methods
- Quality of visualizations
- Depth of analysis and insights
- Clarity of presentation

Bonus Challenges

Choose one:

1. Create an interactive dashboard using Streamlit or Dash
2. Perform predictive analysis for MonthlyIncome based on other variables
3. Conduct a detailed satisfaction analysis with recommendations for improvement

Good luck !