VIRTUAL GLOBAL APPRENTICESHIP

Company Introduction

Careerera was founded in 2014 by founder Mr Vivek Kumar Singh and Mr Alok Kumar Singh. Careerera is an online education provider company of professional certification training, based in Herndon, Virginia, USA. Careerera is one of the leading providers of Higher Education Professional Certification Training, Test Preparation, K -12 Educations, Language Training, and other skill training for Adults and kids in the field of IT, Management, Software Development, Project Management, Quality Assurance, and the list goes on.

Being a partner to some of the top universities and certification bodies, the organization aims at facilitating quality training to professionals worldwide. Careerera has its online learners in 60 countries including America, Canada, Europe, and the Asia Pacific region. It has a track record of training thousands of professionals successfully via classroom and online training. Careerera welcomes you to join one of the largest live online education systems.

Website Link:
https://www.careerera.com/

Theme: Data Analytics and Power BI

Project Task 2: EDA and Summarization

Background Information

You will be given the opportunity to do EDA on a cleaned dataset from the previous task. This knowledge will surely help you to grow further and do the data analysis on real-time datasets when you work with the organization.

Task Explanation

EDA - EDA is an approach to analyze data sets to summarize their main characteristics often using statistical graphics and other data visualization methods.

To keep your work structured and impactive, you can follow 9 step workflow as follows:

1. Introduction - setting context, some research about the dataset/content over the internet
2. Problem Statement - what is the end Goal?
3. Installing and Importing Libraries
4. Data Acquisition and Description - Strategies heading towards end Goal
5. Data Preprofiling - Detecting the issues - missing data - inconsistent data
6. Data Cleaning - appropriate measures
7. Data Postprofiling - Second check for issues if there are any more
8. EDA - Question - way to proceed the same - appropriate data visualization techniques
9. Summarization - Conclusion and Actionable Insights for the company

In the previous task, you have already worked upon the dataset and reached a stage with a cleaned dataset. Here, in this task, you are expected to perform EDA on a cleaned dataset and decide upon conclusions and actionable insights for the same.

CareerEra Dataset (on which you will work) - Survey_Dataset.xls

Resources -

1) https://www.analyticsvidhya.com/blog/2021/05/exploratory-data-analysis-eda-a-step-by-step-guide/
2) https://towardsdatascience.com/an-extensive-guide-to-exploratory-data-analysis-ddd99a03199e

Detailed Sample Solution for Sample Dataset

Sample Dataset - Mental Health in Tech workspace Dataset (shared in the folder itself names sample dataset.csv)

8. Exploratory Data Analysis

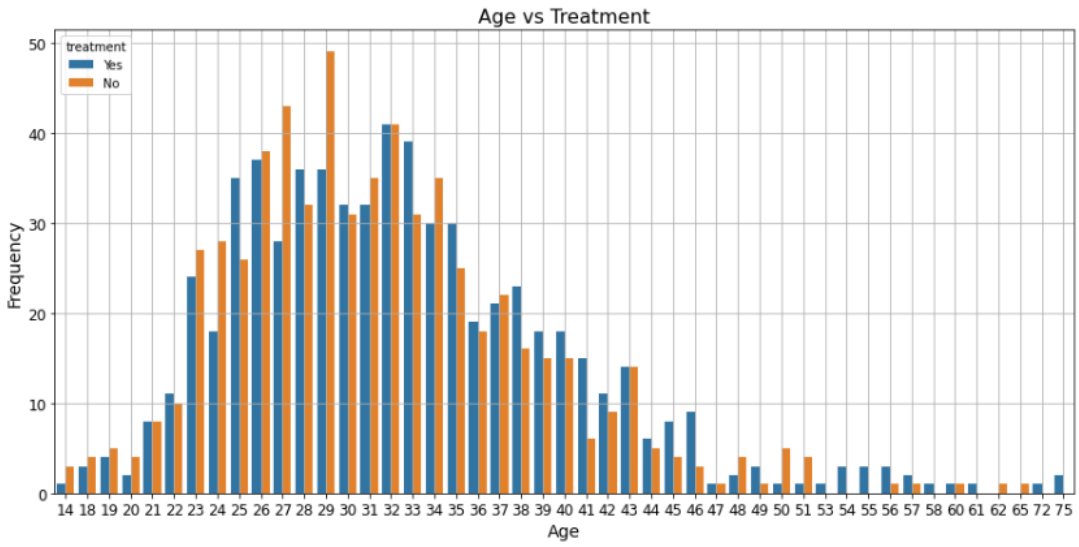
Question: How does age relate to various behaviors and/or their awareness of their employer's attitude toward mental health?

```
# Initialize figure of size 15 X 7
fig = plt.figure(figsize=(15, 7))

# Plot countplot of age concerning treatment
sns.countplot(x='Age', hue='treatment', data=data)

# Add some cosmetics
plt.title(label='Age vs Treatment', size=16)
plt.xlabel(xlabel='Age', size=14)
plt.ylabel(ylabel='Frequency', size=14)
plt.xticks(size=12)
plt.yticks(size=12)
plt.grid(b=True)

# Display the plot
plt.show()
```



Observation:

- We can observe that people of **age 14 - age 22** are **mildly conscious** while people of **age 23 - age 37** are **less conscious** for **treatment**.
- People of **age 38 and above** are **highly conscious** and are **up for treatment**.

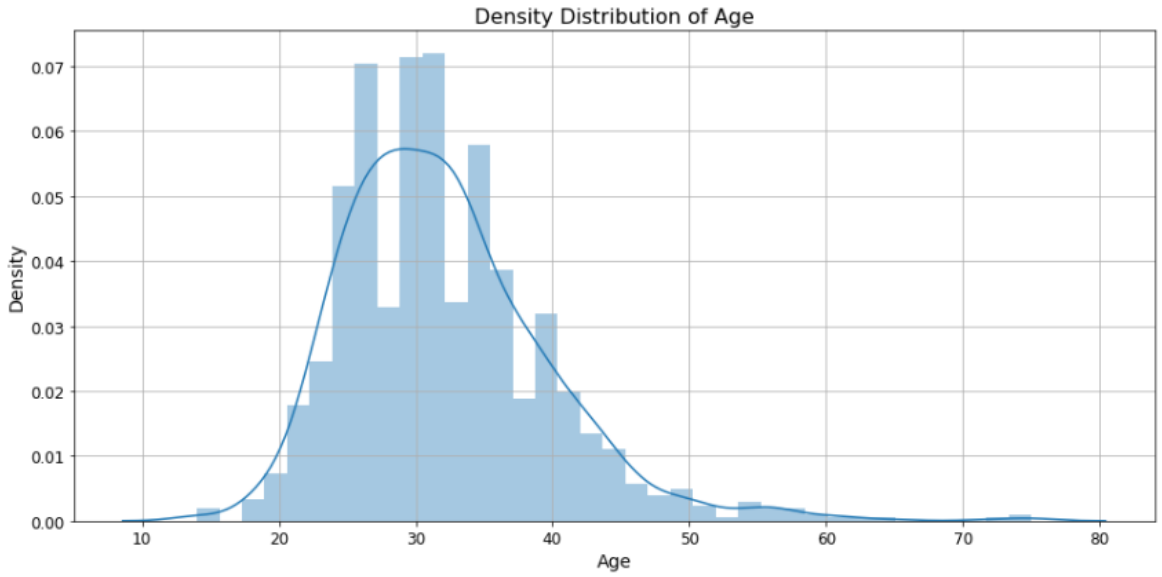
Question: What is the density distribution of Age feature?

```
[ ]  # Initialize figure of size 15 X 7
     fig = plt.figure(figsize=(15, 7))

     # Plot density distribution of age
     sns.distplot(a=data['Age'], kde=True)

     # Add some cosmetics
     plt.title(label='Density Distribution of Age', size=16)
     plt.xlabel(xlabel='Age', size=14)
     plt.ylabel(ylabel='Density', size=14)
     plt.xticks(size=12)
     plt.yticks(size=12)
     plt.grid(b=True)

     # Display the plot
     plt.show()
```



Observation:

- We can **observe a peak** between **mid-20s to about mid-30s**.
- This implies that **majority of people** are from **mid 20s to mid 30s**.

Likewise, figure out all the different possible questions you can think of for exploratory data analysis.

Format for all questions should be

- Question
- Code
- Output (display)
- Observation

9. **Summarization**

- **Conclusion**
  - The mental health survey has **helped** us to **understand** the **mental condition of employees** working in tech firms across countries.
  - A total of **1259 entries were recorded** during the survey out of which **1007 were recorded** from the **top 3 countries**.
  - The **United States leads the chart** in terms of participation in the survey **followed by** the **United Kingdom** and **Canada**.
  - From a **state point of view**, **California leads the chart** when run down the analysis.
  - **48.1%** of **males**, **70%** of **females**, and **88%** of **trans** were found to have **sought treatment** concerning the overall survey.
  - The following set of **parameters** are found to be **affecting mental health** the most and thus requires treatment:
    - Age
    - Family history,
    - Work Interference,
    - Number of employees working in a company,
- **Actionable Insights**
  - There should be an **awareness program** about mental health and its effects.
  - Relationship **Managers should be supportive** with the right guidance towards their employees.
  - Managers should be **unbiased** concerning the work and the employees.
  - There should be **appropriate measures** and **support** for the employees suffering from mental health.
  - It is **good to give** an **appreciation** at work **regularly**.

Task Submission

Summarise your work in a document (even Jupyter Notebook will do) that is easy to understand, Use listed numbers or tables to display your information clearly if required. Submit your document which is either in .docx or .pdf format and the file size is not larger than 2MB. To submit this task you can go to the Project Task section of your dashboard and find Company Task 2 listed there. Click on choose file and upload.

Deadline

The assignment should be submitted by 20th January 2022