Project Plan: Employee Wellness Analysis

Phase 1: Project Setup and Initial Exploration (☑ Completed)

- **Objective**: Load the data and get a first look.
- **Tasks Done**:
    - Loaded the `employee_wellness_dataset.csv` into a pandas DataFrame.
    - Used `info()`, `describe()`, `head()`, and `isnull().sum()` to perform an initial inspection.
    - Identified issues with the `Age` and `Gender` columns and located missing values.

Phase 2: Data Cleaning and Preprocessing (☑ Completed)

- **Objective**: Clean and prepare the data for analysis.
- **Tasks Done**:
    - Removed invalid and outlier entries from the `Age` column.
    - Standardized the `Gender` column into three clear categories ('Male', 'Female', 'Other').
    - Dropped unnecessary columns (`S.No`, `comments`, `Timestamp`).
    - Developed and executed a strategy to handle all missing values.
    - **Outcome**: We now have a clean, reliable dataset with 1042 rows and 25 columns.

Phase 3: Exploratory Data Analysis (EDA) (♀ This is our next step)

- **Objective**: Analyze and visualize the data to uncover patterns, relationships, and insights. We will break this into three parts.
- **3.1: Univariate Analysis (Understanding Individual Columns)**
    - **Demographics**:
        - Visualize the distribution of employee **Age** using a histogram.
        - Visualize the **Gender** distribution using a bar chart.
        - Visualize the top 10 countries of employees using a bar chart.
    - **Workplace Profile**:
        - Visualize the distribution of company sizes (`no_employees`).
        - Analyze the proportion of employees working in tech companies.
    - **Mental Health Baseline**:
        - Visualize the breakdown of employees with a `family_history` of mental illness.

- **Crucially**, visualize the overall percentage of employees who are seeking `treatment`. This is our target variable.
- **3.2: Bivariate Analysis (Finding Relationships Between Columns)**
  - **Focus on the `treatment` variable.** Our goal is to see what factors are most correlated with an employee seeking treatment.
  - `treatment` **vs. Demographics**:
    - Does `Gender` affect the likelihood of seeking treatment?
    - Does `Age` play a role in seeking treatment?
  - `treatment` **vs. Workplace Support**:
    - How does the availability of `benefits` correlate with seeking treatment?
    - Does a company `wellness_program` impact the treatment rate?
    - Is there a link between `remote_work` and seeking treatment?
  - `treatment` **vs. Personal/Work Impact**:
    - How strongly does `family_history` influence the decision to get treatment?
    - How does the perceived interference with work (`work_interfere`) relate to seeking treatment?
- **3.3: Attitude & Stigma Analysis**
  - Analyze the perceived consequences of discussing mental health (`mental_health_consequence`).
  - Visualize employees' willingness to discuss mental health with `coworkers` and `supervisors`.
  - Explore whether employees feel their employer takes mental health as seriously as physical health (`mental_vs_physical`).

Phase 4: Insights and Interpretation

- **Objective**: Synthesize our findings from the EDA phase into actionable insights.
- **Tasks**:
  - For each visualization from Phase 3, we will write a clear, concise summary of what it tells us.
  - We will connect the dots between different findings to build a comprehensive picture.
  - We will identify the key factors that are most strongly associated with an employee needing treatment.

Phase 5: Presentation Report & Submission

- **Objective**: Compile the entire analysis into a well-documented Jupyter Notebook.
- **Tasks**:
    - Structure the notebook with clear headings for each phase (Introduction, Data Cleaning, EDA, Insights, Conclusion).
    - Ensure all code is clean and accompanied by markdown cells that explain the steps and, most importantly, the **insights** derived.
    - Conclude with a summary that directly addresses the project's goal: helping the organization understand factors contributing to employee health and identifying employees who may need support.

This detailed plan will be our guide. We'll now proceed step-by-step through **Phase 3**.

Ready to start with **Phase 3.1: Univariate Analysis**?