Mental health and substance use disorders with unsupervised outlier detection

Problem and Context:

Mental health and substance use disorders affect our life in all directions in the world. The prevalence of mental health and substance disorders includes depression, anxiety, bipolar, eating disorders, alcohol or drug use disorders, and schizophrenia. They involve changes in thinking, mood, and/or behavior. The data showed mental health or substance use disorders are common everywhere but with different rates for different countries and with different rates by gender and age. The goal of this project is to show people how mental health and substance use disorders impact our life, how it behaves differently by ages and genders, and detect outlier countries with gender rates data.

Target Clients:

The primary clients this project targets are society organizations, government, and general people like us. Hopefully, this project makes people know how important and close to people's lives, and all of us need to pay attention to it and help people who are suffering from mental health and substance use disorders.

Data:

The datasets used for this project were produced by the Institute for Health Metrics and Evaluation and reported in their flagship Global Burden of Disease study. The original blog was published on Our World in Data (https://ourworldindata.org/mental-health#all-charts-preview). A few datasets were downloaded from Our World in Data chart sources. This project with being interested in the following datasets:

- death-rates-from-mental-health-and-substance-use-disorders.csv
- GDP-per-capita-worldbank.csv
- mental-and-substance-use-as-share-of-disease.csv
- prevalence-by-mental-and-substance-use-disorder.csv
- prevalence-of-mental-and-substance-disorders-by-age.csv
- share-with-mental-and-substance-disorders.csv
- share-with-mental-or-substance-disorders-by-sex.csv
- suicide-rates-vs-prevalence-of-mental-and-substance-use-disorders.csv

The datasets include diagnosis rates from 1990 to 2017 in 231 entities (countries or regions) for different focuses.

Approach:

Multiple steps will be taken to show the information of data, check the difference among groups, and modeling for outliers detection.

1. The multiple CSV files consisting of the data will be imported, merged, and cleaned via Python 3.
2. Datasets will be explored visually via graphs and charts and comparisons among groups via statistics.
3. A few unsupervised methods will be applied to detect outliers for the gender dataset.

Deliverables:

The final draft of the project will be presented via PowerPoint slides. Jupyter Notebooks will be delivered detailing each step is taken and code written for the analysis of the project. A Github repository for the project will be created as well.