MENTAL HEALTH ANALYSIS - NHANES DATASET

Team Members:

Neha Shetty
Shubhangi Kishore
Melita Saldanha
Ronak Khandelwal

System used to run the code: Google Cloud (Configurations same as Assignment 3)

Data:

We have considered 18 questionnaires with data from 1999-2018. The questionnaires are as follows:

1. Mental Health - Depression
2. Early Childhood
3. Food Security
4. Occupation
5. Sleep Disorders
6. Housing Characteristics
7. Sexual Behavior
8. Diet Behavior and Nutrition
9. Hospital Utilization and Access to Care
10. Consumer Behavior
11. Oral Health
12. Body Measures
13. Alcohol Consumption
14. Income
15. Household People Count
16. Cholesterol Level
17. Blood Count
18. Blood Pressure

Data Pre-processing:

We merged the data for all the years and cleaned each questionnaire data individually, based on its specific requirements.
Filename: data_preprocessing.py
Usage: ***python data_preprocessing.py***

Hypothesis Testing:

1. **Finding associations between depression questionnaire data and select features**
   Description - For pre-selected general datasets (Questionnaire, Laboratory, Examination) from NHANES we find the correlation to investigate the associations and derive Beta Values by performing regression, We also perform hypothesis testing using two tailed T Tests and calculating the resulting P value. We used a size of the test equal to $\alpha = 0.05$ level and performed multi-test correction.
   Filename - correlation_all_features.py
   Usage - ***python3 correlation_all_features.py***

2. **Regression Analysis**
   Description - Performing Multivariate Linear regression on given datasets with the target column using tensorflow (making use of momentum vector).
   Filename - regression_analysis.py
   Usage - ***regression_analysis.py "feature_filename" "target_filename" "target_col"***

Recommendation System:

Frameworks and concepts used: Spark, HDFS, Similarity and Collaborative filtering
Filename: Recommendation_system.py
Usage: ***spark-submit Recommendation_system.py <target_column> <space_separated_list_of_csv_files>***

Mediation Analysis:

As all the features do not directly have an effect on mental health, we performed mediation analysis by introducing a mediator to check the effect of an independent variable on mental health features.
Filename: mediationAnalysis.py
Usage: ***python mediationAnalysis.py***