**Social networks in mild-to-moderate Alzheimer disease: longitudinal relationships with dementia severity, cognitive function, and adverse events**

Chaithanya N (22PD09)

Sujan S (22PD35)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**The Research Paper**

The study explores the impact of social networks on dementia progression, cognitive function, and adverse events in individuals with mild-to-moderate Alzheimer's disease (AD), focusing on how network size and quality may affect patient outcomes over time.

* **Social Network**: Assessed via the Lubben Social Network Scale, which measures engagement with family and friends.
* **Dementia Severity**: Evaluated using the Clinical Dementia Rating - Sum of Boxes (CDR-Sb), gauging the severity of AD symptoms.
* **Cognitive Function**: Measured by the Alzheimer’s Disease Assessment Scale (ADAS-Cog).
* **Adverse Events**: Defined as significant health complications or hospitalizations, tracked throughout the study.

Key findings include a correlation between poor social networks and higher dementia severity but not with cognitive impairment. Additionally, baseline dementia severity was found to predict a reduction in social network size over time, with poor networks linked to higher rates of adverse events​

### \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### 

### 

### **Introduction**

#### **Background**

Alzheimer’s disease (AD) is a debilitating neurodegenerative disorder that progressively impairs cognitive function, memory, and the ability to perform everyday tasks. Affecting millions of individuals worldwide, AD is the most common cause of dementia, with its prevalence steadily increasing due to the aging global population. As a chronic and incurable condition, Alzheimer's disease imposes a heavy burden not only on the individuals affected but also on their families, caregivers, and healthcare systems. The progressive nature of the disease leads to a gradual decline in cognitive abilities, leaving individuals increasingly dependent on their social and physical environments.

In the context of aging and dementia, the preservation of cognitive health is essential to maintaining independence and quality of life. Recent studies suggest that social networks—defined by the size, strength, and quality of social interactions—play a significant role in mitigating the effects of cognitive decline. For older adults, especially those diagnosed with Alzheimer’s disease, maintaining strong social connections may slow the progression of cognitive impairments. Engaging in meaningful social relationships has been shown to support mental well-being, encourage cognitive stimulation, and potentially contribute to building cognitive reserve, which could delay the onset or slow the progression of AD symptoms.

#### **Problem Statement**

Despite growing evidence highlighting the benefits of social networks in preventing or delaying cognitive decline, their impact on individuals with an established diagnosis of Alzheimer’s disease is less clear. People with mild-to-moderate Alzheimer’s disease are particularly vulnerable to social isolation as their condition progresses, potentially leading to accelerated cognitive deterioration and an increased risk of adverse health events such as hospitalizations, falls, or worsening symptoms.

A study on "Social Networks in Mild-to-Moderate Alzheimer’s Disease" explores the longitudinal relationships between social networks and dementia severity. It reveals that individuals with poorer social networks tend to experience faster dementia progression, though these relationships are complex. Social isolation may not directly cause cognitive decline, but rather, the worsening of cognitive and physical health may lead to diminished social engagement. Additionally, poor social networks have been associated with a higher likelihood of experiencing serious adverse health events, making this a critical issue for healthcare providers and caregivers to address.

This study is crucial because it sheds light on the interplay between social networks and cognitive health in AD patients, emphasizing the importance of social engagement as part of comprehensive dementia care. Understanding these dynamics can help develop interventions aimed at maintaining or strengthening social ties to improve outcomes for individuals with Alzheimer’s disease.

#### **Objectives**

The primary objective of this study is to investigate the relationships between social networks, cognitive function, dementia severity, and adverse events in individuals diagnosed with mild-to-moderate Alzheimer’s disease. Using data ,which assessed the efficacy of Nilvadipine in slowing AD progression, this analysis focuses on how social networks, measured through the Lubben Social Network Scale (LSNS), are associated with cognitive outcomes (using the Alzheimer’s Disease Assessment Scale-Cognitive Subscale [ADAS-Cog]) and dementia severity (measured by the Clinical Dementia Rating-Sum of Boxes [CDR-Sb]). Additionally, the study aims to evaluate how poor social networks contribute to the likelihood of adverse health events over time.

By examining these relationships, this research aims to provide valuable insights into how social networks may influence the course of Alzheimer’s disease and highlight the need for targeted social interventions to support patients and mitigate negative health outcomes.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **Methods**

#### **Study Design**

This study is an analysis of data, a phase-3, multi-center investigation aimed at assessing the efficacy of Nilvadipine, a calcium channel blocker, in slowing the progression of Alzheimer’s disease (AD). The trial involved 464 participants across nine European countries, all diagnosed with mild-to-moderate Alzheimer’s disease. Participants were eligible if they met specific inclusion criteria, including a Mini-Mental State Examination (MMSE) score between 12 and 26, indicating mild to moderate cognitive impairment. In addition to the cognitive decline, participants were required to have an informant, such as a family member or caregiver, who interacted with them face-to-face at least three times a week.

The trial collected data at baseline and again after 18 months of follow-up, providing a longitudinal view of dementia progression. Cognitive health was assessed using two key measures: the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) and the Clinical Dementia Rating-Sum of Boxes (CDR-Sb). These tools allowed researchers to evaluate changes in both cognitive performance and overall dementia severity over time. Social network data were also collected at both time points using the Lubben Social Network Scale-6 (LSNS-6), providing valuable insights into the participants’ social environments and how these might influence their disease trajectories.

#### **Social Networks Measurement**

The Lubben Social Network Scale (LSNS-6) was employed to measure the social networks of participants. This validated instrument is designed to assess the size, strength, and functionality of social networks, focusing on two primary components: family and friendship ties. The scale includes six questions, three related to family interactions and three to friendships. Questions addressed aspects such as the frequency of contact with family members or friends, the level of comfort in confiding in them, and whether participants could rely on them for help.

Each question was scored on a scale of 0 to 5, with higher scores indicating stronger social networks. A total LSNS score ranging from 0 to 30 was computed by summing all six responses. To capture more detailed social network data, participants’ scores were divided into quartiles, with those in the lowest quartile classified as having a poor social network. This classification allowed the study to explore how variations in social network size and quality impacted cognitive decline and other health outcomes in the context of Alzheimer’s disease.

#### **Data Analysis**

The data analysis employed a range of statistical methods to explore the relationships between social networks, cognitive function, dementia severity, and adverse health outcomes. Mixed-effects linear regression models were used to assess the association between social network scores and both cognitive function (ADAS-Cog scores) and dementia severity (CDR-Sb scores). These models accounted for repeated measures over time, controlling for individual differences and adjusting for potential confounding variables such as age, gender, education level, and the duration of dementia symptoms.

Additionally, Poisson models were utilized to examine the relationship between poor social networks and the likelihood of experiencing adverse events, such as hospitalizations or serious medical complications. These models allowed researchers to evaluate how social isolation may contribute to negative health outcomes in Alzheimer’s patients, adjusting for variables like comorbidities and medication use. By combining these statistical approaches, the study aimed to provide a comprehensive understanding of how social networks impact the progression of Alzheimer’s disease and the overall well-being of individuals diagnosed with the condition.

The analysis further explored whether baseline dementia severity predicted a decline in social networks over time. To assess this, changes in LSNS-6 scores between baseline and 18 months were examined, with dementia severity at baseline used as a key independent variable.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### **Results**

#### **Social Networks and Dementia Severity**

The analysis revealed a significant relationship between poor social networks and increased dementia severity at baseline. Participants with weaker social networks, as measured by the Lubben Social Network Scale (LSNS-6), exhibited higher Clinical Dementia Rating-Sum of Boxes (CDR-Sb) scores, indicating more advanced dementia. This association underscores the critical role that social engagement plays in the lives of individuals with Alzheimer’s disease (AD).

Baseline characteristics of the participants showed a broad demographic range. The mean age was approximately 73 years, with the majority being female. Participants with poor social networks were typically older, had been living with the symptoms of dementia for a longer period, and had more severe dementia at the time of enrollment. Notably, those in the lowest quartile for social network size were more likely to have higher CDR-Sb scores, reflecting greater functional impairments and cognitive deficits. This correlation suggests that individuals with diminished social interactions may be more vulnerable to faster dementia progression, possibly due to reduced opportunities for cognitive stimulation and emotional support.

#### **Cognitive Decline and Social Networks**

While poor social networks were strongly associated with increased dementia severity, they did not predict a direct decline in cognitive performance as measured by the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog). Over the 18-month follow-up period, cognitive function declined across the cohort, but participants with poor social networks did not experience significantly worse cognitive outcomes than those with more robust social ties.

Interestingly, it was found that greater baseline dementia severity, rather than poor social networks, predicted a decline in social network size over the study period. Participants with more severe dementia at baseline saw a greater reduction in their social engagement over 18 months, suggesting a reciprocal relationship between social interaction and cognitive health. As dementia progresses, individuals may withdraw from social activities or find it harder to maintain relationships, which, in turn, can exacerbate their cognitive and emotional decline. This finding highlights the complexity of the relationship between social networks and cognitive health in AD, where the disease itself may lead to a shrinking social circle, rather than the absence of social ties directly causing cognitive decline.

#### **Adverse Events**

A key finding from the analysis was the significant association between poor social networks and an increased likelihood of experiencing serious adverse events. Over the 18-month period, participants with weaker social networks were notably more likely to face medical complications such as hospitalizations, falls, and other serious health issues. The analysis indicated that social isolation may contribute to poorer overall health outcomes in individuals with Alzheimer’s disease.

Those with limited social networks were at a greater risk for adverse events, even after adjusting for factors like age, gender, comorbidities, and medication use. The findings suggest that individuals with poor social support systems may lack the necessary resources, supervision, or intervention that could prevent these events. Socially isolated individuals may also experience greater psychological distress, which could lead to worsened physical health. This underscores the need for comprehensive care strategies that include not only medical interventions but also social support systems to improve outcomes for patients with Alzheimer’s disease.

In summary, the results of this study emphasize the critical role social networks play in the health and well-being of individuals with Alzheimer’s disease. Poor social networks are associated with increased dementia severity, a higher likelihood of adverse events, and a decline in social engagement as the disease progresses. However, the relationship between social networks and cognitive decline is complex, with dementia severity itself contributing to social isolation over time. These findings highlight the importance of interventions that maintain or strengthen social connections as part of an integrated approach to Alzheimer’s care.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dataset**

**1. Demographics and Patient Information**

**-PatientID**: Unique identifier for each patient.

**- Age**: Age of the patient, typically used to understand the demographic spread and observe if age correlates with disease severity.

**- Gender**: Indicates the gender of each patient, often coded as binary (0 for male, 1 for female).

**- Diagnosis**: Binary classification (e.g., 0 for no Alzheimer's, 1 for Alzheimer’s) indicating whether the patient is diagnosed with Alzheimer's or a related dementia.

**2. Clinical and Cognitive Assessments**

**- MMSE (Mini-Mental State Examination):** A commonly used cognitive test where lower scores indicate cognitive impairment, with 30 being the highest (normal) and scores <24 typically indicating possible cognitive issues.

**- FunctionalAssessment:** A measure of how well a patient performs daily activities, often on a scale where lower scores indicate greater impairment.

**- ADL (Activities of Daily Living)**: Assessment of a patient’s ability to perform everyday tasks like eating, bathing, and dressing, often rated as a percentage or score.

**- MemoryComplaints**: Reports from the patient or caregiver regarding memory loss, which can be binary or scaled, capturing the subjective experience of memory decline.

**- BehavioralProblems:** Often rated on a scale or as binary data, indicating whether the patient displays behavioral issues, such as aggression or mood changes.

**3. Health and Physiological Metrics**

**- SystolicBP & DiastolicBP:** Blood pressure readings, recorded in mmHg, to track cardiovascular health and its potential impact on brain health.

**- CholesterolTotal:** Total cholesterol level, typically in mg/dL, to evaluate cardiovascular health since high cholesterol is sometimes associated with increased dementia risk.

**- HeartRate:** Resting heart rate in beats per minute (bpm), included as an indicator of cardiovascular health.

**- BMI (Body Mass Index):** Calculated from weight and height, this feature reflects overall health and fitness, with extremes linked to various health risks.

**4. Neurological and Neuropsychiatric Indicators**

**- CognitiveScores:** Additional cognitive test scores beyond MMSE, which may include specific memory, attention, or executive function tests.

**- NeuropsychiatricSymptoms:** May include ratings for symptoms like depression, anxiety, or apathy, typically seen in dementia patients.

**- Imaging or Biomarkers**: Features such as PET or MRI imaging scores, or biomarker levels from cerebrospinal fluid (CSF), may be included if the dataset has biological indicators.

**5. Treatment or Medication**

**- MedicationStatus:** Indicates whether the patient is on medications that could influence cognition, such as cholinesterase inhibitors or antidepressants.

**- TreatmentStatus:** Includes information on whether the patient has undergone any form of therapy or treatment related to cognitive decline.

These attributes provide insights into different aspects of a patient’s health and functionality, which can be used to study how Alzheimer's and related dementias progress and impact various aspects of life. You can also derive meaningful correlations, such as links between cognitive scores and lifestyle factors, which may help in understanding the factors that contribute to Alzheimer's.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Our Implementation**

**Import Libraries:** The code imports pandas, networkx, matplotlib.pyplot, seaborn, and numpy using the import statement

**Import Dataset:**  The Alzheimer's disease dataset is loaded into a pandas DataFrame called df using the pd.read\_csv() function, specifying the file path

**Data types:**

**object:** Represents strings or mixed data types. Columns like 'Gender', 'Ethnicity', 'Diagnosis', etc., which contain categorical values, are usually of type object.

**int64:** Represents integer values. Columns like 'PatientID', 'Age', 'EducationLevel', etc., which contain whole numbers, are likely of type int64.

**float64:** Represents floating-point numbers. Columns like 'BMI', 'AlcoholConsumption', 'SystolicBP', etc., which contain numerical values with decimals, are typically of type float64.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Numeric and Categorical Features:**

**Missing values and duplicates:**

**df.isnull().sum():** Calculates and displays the total number of missing values (NaN) in each column, highlighting potential data cleaning needs.

**df.duplicated().sum():** Checks for and prints the total number of duplicated rows, which is vital for ensuring data integrity and avoiding bias in analysis.

**Statistical analysis:**

**df.describe().T:** Generates descriptive statistics (mean, standard deviation, min, max, quartiles) for numerical columns, providing insights into the data's distribution and central tendencies. The .T transposes the output for better readability.

**Analyze On Diagnosis:**

This part specifically analyzes the dataset based on the 'Diagnosis' column.

mean\_stats = df.groupby('Diagnosis').mean(numeric\_only=True): Calculates the average values of numeric features for each diagnostic group (0 or 1).

mean\_stats.to\_csv("stats-by-diagnosis.csv"): Saves the calculated means to a CSV file for further analysis or sharing.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**EDA**

**1. Feature Categorization:**

The code defines two lists: categorical\_features and numeric\_features.

It explicitly lists the names of columns belonging to each category.

This categorization is crucial for applying appropriate visualizations and analyses to each feature type.

**2. Barplots (Countplots):**

The countplot() function is defined to create bar plots (countplots using Seaborn) for categorical features.

It takes a column name as input and generates a plot showing the frequency of each category within that column.

Countplots are called for several categorical features like 'Gender', 'Ethnicity', 'EducationLevel', 'Smoking', 'Diagnosis', and others.

These visualizations help in understanding the distribution of categorical variables, identifying dominant categories, and spotting potential imbalances.

**3. Distributions (Histograms):**

The code creates histograms to visualize the distributions of numeric features.

It iterates through the numeric\_features list, generating a histogram for each feature.

Histograms are helpful in understanding the shape of the data distribution, identifying central tendencies, and spotting potential outliers or skewness.

**4. Boxplots:**

Boxplots are created for numeric features to visualize their distributions and identify potential outliers.

They provide insights into the median, quartiles, and range of the data, highlighting any extreme values.

**5. Heatmap (Correlation Analysis):**

The code calculates and visualizes the correlation between numeric features and the target variable ('Diagnosis') using a heatmap.

It uses the corrwith() function to calculate correlations and then plots them using Seaborn's heatmap() function.

Heatmaps help in identifying potential relationships between variables and the target, aiding in feature selection and understanding feature importance.

**6. Swarmplots with respect to Diagnosis:**

Swarmplots are used to visualize the distribution of numeric features within different 'Diagnosis' categories.

This helps compare the distributions and identify potential differences or patterns between groups.

Swarmplots are created for features like 'FunctionalAssessment', 'ADL', 'MMSE', 'Age', 'Gender', 'MemoryComplaints', and 'BehavioralProblems'.

**7. Social Network Analysis (SNA):**

This section uses the networkx library to create and analyze social networks based on shared health attributes.

It focuses on relationships between patients based on similarity in attributes like 'SystolicBP', 'DiastolicBP', 'CholesterolTotal', and 'MMSE'.

Visualizations of the network graphs are created to highlight connections and clusters of patients.

Centrality measures like degree centrality, betweenness centrality, and closeness centrality are calculated to identify influential nodes in the network.

**8. Homophily Analysis:**

This part examines the tendency of similar individuals to form connections in the social network (homophily).

It calculates homophily ratios based on shared attributes like age, gender, and diagnosis.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Modeling**

**1. Splitting Dependent/Independent Variables:**

X = df.drop(['Diagnosis', 'PatientID'], axis=1).values: This line creates the feature matrix X by dropping the 'Diagnosis' (target variable) and 'PatientID' columns from the DataFrame. It's then converted to a NumPy array using .values.

y = df['Diagnosis'].values: This line creates the target variable array y containing the 'Diagnosis' values.

**2. PCA & Splitting Train/Test Sets:**

This part involves Principal Component Analysis (PCA) for dimensionality reduction and splitting the data into training and testing sets.

PCA:

pca = PCA(12): Initializes a PCA object with 12 components.

X\_pca = pca.fit\_transform(X): Applies PCA to the feature matrix X, reducing the number of features to 12.

Train-Test Split:

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X\_pca, y, test\_size=0.2, random\_state=42): Splits the data into training (80%) and testing (20%) sets using train\_test\_split. random\_state ensures reproducibility.

**3. Scale:**

Data scaling is performed using StandardScaler to standardize the features.

scaler = StandardScaler(): Initializes a StandardScaler object.

X\_train = scaler.fit\_transform(X\_train): Fits the scaler on the training data and transforms it.

X\_test = scaler.transform(X\_test): Transforms the testing data using the fitted scaler.

**4. Model Definition:**

Necessary libraries for model evaluation and selection are imported.

XGBoost:

xgb\_model = XGBClassifier(): Initializes an XGBoost classifier object.

param\_dist: A dictionary defining the hyperparameter grid for tuning the XGBoost model.

Hyperparameter tuning can be performed using techniques like RandomizedSearchCV or GridSearchCV (not explicitly shown in the provided code).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ROC and AUC**

**Prediction Probabilities**: After training your XGBoost model, it's used to predict the probabilities of a patient having Alzheimer's (Diagnosis = 1) on the test data.

**ROC Curve Calculation:** Using the predicted probabilities and the actual diagnosis from the test data, a ROC curve is generated. This curve visually represents the trade-off between correctly identifying patients with Alzheimer's (true positives) and incorrectly identifying patients without Alzheimer's (false positives) at different probability thresholds.

**AUC Calculation:** The area under the ROC curve (AUC) is then calculated. This single number summarizes the overall performance of the model. A higher AUC indicates better discrimination between the two classes.

**Interpreting ROC and AUC**

AUC = 0.5: The model is no better than random guessing.

AUC > 0.5: The model has some predictive power.

AUC = 1.0: The model is a perfect classifier (rarely achieved in practice).

The closer the ROC curve is to the top-left corner, the better the model's performance.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Social Networks Analysis**

**Nodes:** Represent individual entities in the network (patients in your case).

**Edges:** Represent connections or relationships between nodes (patients with similar health attributes).

**Network:** The overall structure formed by the nodes and edges, representing the relationships between patients.

**Centrality:** Measures the importance or influence of nodes within the network. Different centrality measures capture different aspects of importance (e.g., degree centrality, betweenness centrality).

**Homophily:** The tendency of similar individuals to form connections (e.g., patients with similar age or blood pressure).

**Data Preparation:**

Selects specific health attributes (e.g., SystolicBP, DiastolicBP, CholesterolTotal, MMSE) for network construction.

Normalizes these attributes to ensure they have a similar range and prevent features with larger values from dominating the analysis.

**Network Construction:**

Creates a network graph where patients are represented as nodes.

Connects patients (edges) based on similarity in the selected health attributes. If the distance between two patients' attribute values is below a certain threshold, an edge is created between them.

**Visualization:**

Generates visual representations of the network graphs to highlight connections and clusters of patients.

**Centrality Analysis:**

Calculates various centrality measures (e.g., degree centrality, betweenness centrality, closeness centrality) to identify influential patients within the network. Patients with higher centrality scores are considered more central or influential.

**Homophily Analysis:**

Examines the tendency of patients with similar characteristics (e.g., age, gender, diagnosis) to form connections in the network. This helps understand if certain attributes drive the formation of social structures.

**Insights from SNA**

**Social Structures:** By visualizing the network, you can identify potential clusters or subgroups of patients with similar health profiles.

**Influential Patients:** Centrality analysis reveals patients who are more central or influential within the network, potentially playing key roles in information dissemination or social support.

**Homophily Patterns:** Examining homophily can shed light on factors that drive social connections, such as shared health conditions, demographics, or risk factors.

Overall, the SNA in your code provides a framework for understanding the social relationships between patients based on their health attributes, offering insights into potential social influences and structures that could be relevant to Alzheimer's disease research.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Clustering Coefficient**

**Concept:** The clustering coefficient measures the degree to which nodes in a network tend to cluster together. It quantifies how close a node's neighbors are to forming a complete subgraph (a clique).

**Integration:** You would calculate the clustering coefficient after constructing your patient network within your construct\_network function. This would involve using the networkx library's functionality to calculate either the average clustering coefficient for the entire network or individual clustering coefficients for each node (patient).

**Interpretation within Your Analysis**

**Average Clustering Coefficient**: A higher value (closer to 1) indicates a more clustered network, where patients with similar health attributes are more likely to be connected to each other's neighbors, forming tightly-knit communities. A lower value suggests a more dispersed network with fewer local connections.

**Node-Level Clustering Coefficients (optional)**: These values provide insights into specific patients who are more central within their local neighborhoods or communities. This could highlight potential influential patients or social hubs within the network.

**Combining Clustering Coefficient with Other Analyses**

**Compare across diagnosis groups:** Calculate the average clustering coefficient separately for patients with and without Alzheimer's to see if their social network structures differ.

**Relate to centrality:** Explore if patients with higher centrality measures (e.g., degree centrality) also tend to have higher clustering coefficients, indicating that they are central both globally within the network and locally within their neighborhoods.

**Combine with homophily**: Investigate if homophily based on certain attributes (e.g., age, gender) is associated with higher clustering coefficients, suggesting that these attributes drive the formation of tightly-knit communities.

**Insights from Clustering Coefficient**

By considering the clustering coefficient in your analysis, you can gain insights into:

**Community Structure:** Identify potential communities or subgroups of patients with similar health attributes who are densely connected to each other.

**Information Flow:** Higher clustering might imply that information or social support flows more easily within these tightly-knit communities.

**Disease Progression:** Changes in clustering coefficient over time could potentially be linked to disease progression or the effectiveness of interventions.

In essence, by incorporating the clustering coefficient into your analysis (without explicitly modifying your code here), you can enhance your understanding of the local structure and community formation within your patient network, enriching your insights into the social aspects related to Alzheimer's disease.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Interpretation**

1. **Distribution of Categorical Features:**

**Barplots (Countplots)**: These visualizations revealed the frequency of different categories within categorical features like gender, ethnicity, education level, smoking habits, and medical history factors. You could observe things like the proportion of males and females, the dominant ethnicity, and the prevalence of certain medical conditions within your dataset.

**Diagnosis Distribution:** A pie chart or countplot specifically for the 'Diagnosis' feature would have shown the proportion of patients diagnosed with Alzheimer's disease versus those without the diagnosis. This provides an understanding of the class balance in your dataset.

**2. Distribution of Numeric Features:**

**Histograms:** These plots revealed the shape of the distribution for numeric features like age, BMI, alcohol consumption, blood pressure, cholesterol levels, and cognitive assessment scores (MMSE, Functional Assessment). You could identify potential outliers, skewness, and the overall range of values for each feature.

**Boxplots:** These visualizations provided a different perspective on the distribution of numeric features, highlighting the median, quartiles, and potential outliers more clearly.

**3. Correlation Analysis:**

Heatmap: The heatmap visualized the correlation between numeric features and the target variable ('Diagnosis'). This helped identify features that might be strongly associated with the presence or absence of Alzheimer's disease. You could observe positive or negative correlations and their strength.

**4. Swarmplots with respect to Diagnosis:**

These visualizations compared the distribution of numeric features across different 'Diagnosis' categories. You could identify potential differences or patterns in features like functional assessment, ADL, MMSE, age, gender, memory complaints, and behavioral problems between diagnosed and non-diagnosed groups. This provided insights into which features might be more discriminative for predicting Alzheimer's.

**5. Social Network Analysis (SNA):**

**Network Visualization:** The visualizations of the patient network graphs revealed potential clusters or subgroups of patients with similar health profiles. This could suggest social structures or communities based on shared health attributes.

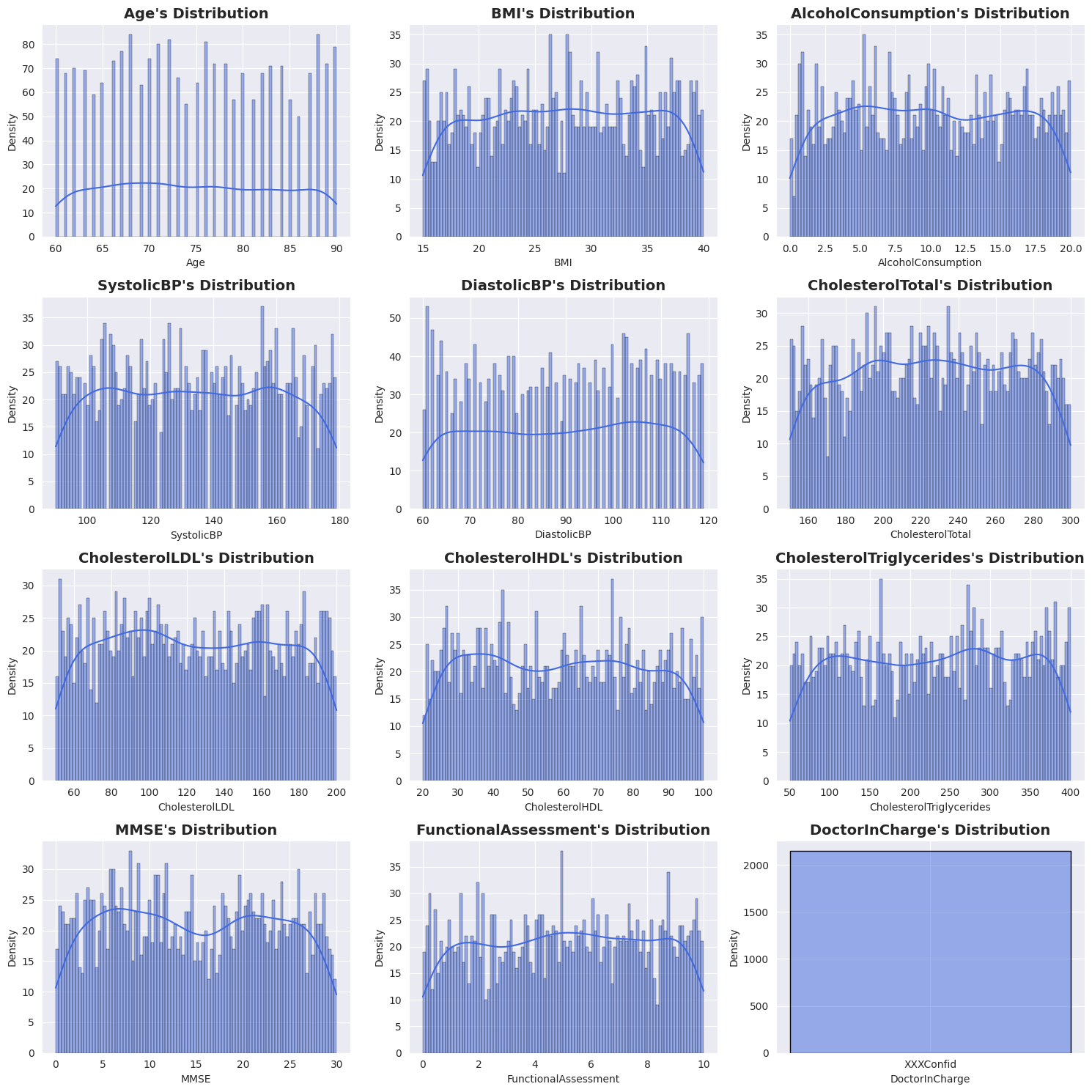
**Centrality Measures:** Calculating centrality measures like degree centrality, betweenness centrality, and closeness centrality helped identify influential patients within the network. These patients might play important roles in information dissemination or social support.

**Homophily Analysis**: Examining homophily patterns revealed whether patients with similar characteristics (e.g., age, gender, diagnosis) tended to form connections in the network. This could suggest factors that drive social connections and potentially influence disease risk or progression.

**Clustering Coefficient:** Analyzing the clustering coefficient provided insights into the density of connections within local neighborhoods of the patient network. Higher clustering coefficients indicated tightly-knit communities, suggesting that patients with similar health attributes are more likely to be connected to each other's neighbors.

Overall, the EDA results provide a comprehensive overview of the data's characteristics, potential relationships between features and the target variable, and insights into the social structure within the patient network.

**Distributions**



**Inference**

Age Distribution: The distribution of ages is relatively uniform across the older adult age range (60-90), suggesting a diverse sample of patients without a clear age bias.

BMI and Alcohol Consumption: The BMI and alcohol consumption distributions show varying levels of density, indicating some common BMI ranges and alcohol consumption levels among the patients. The peaks suggest certain BMI values may be more prevalent.

Blood Pressure Measurements: The distributions for systolic and diastolic blood pressure indicate that there may be patients with higher blood pressure readings, which is critical since hypertension is a risk factor for cognitive decline.

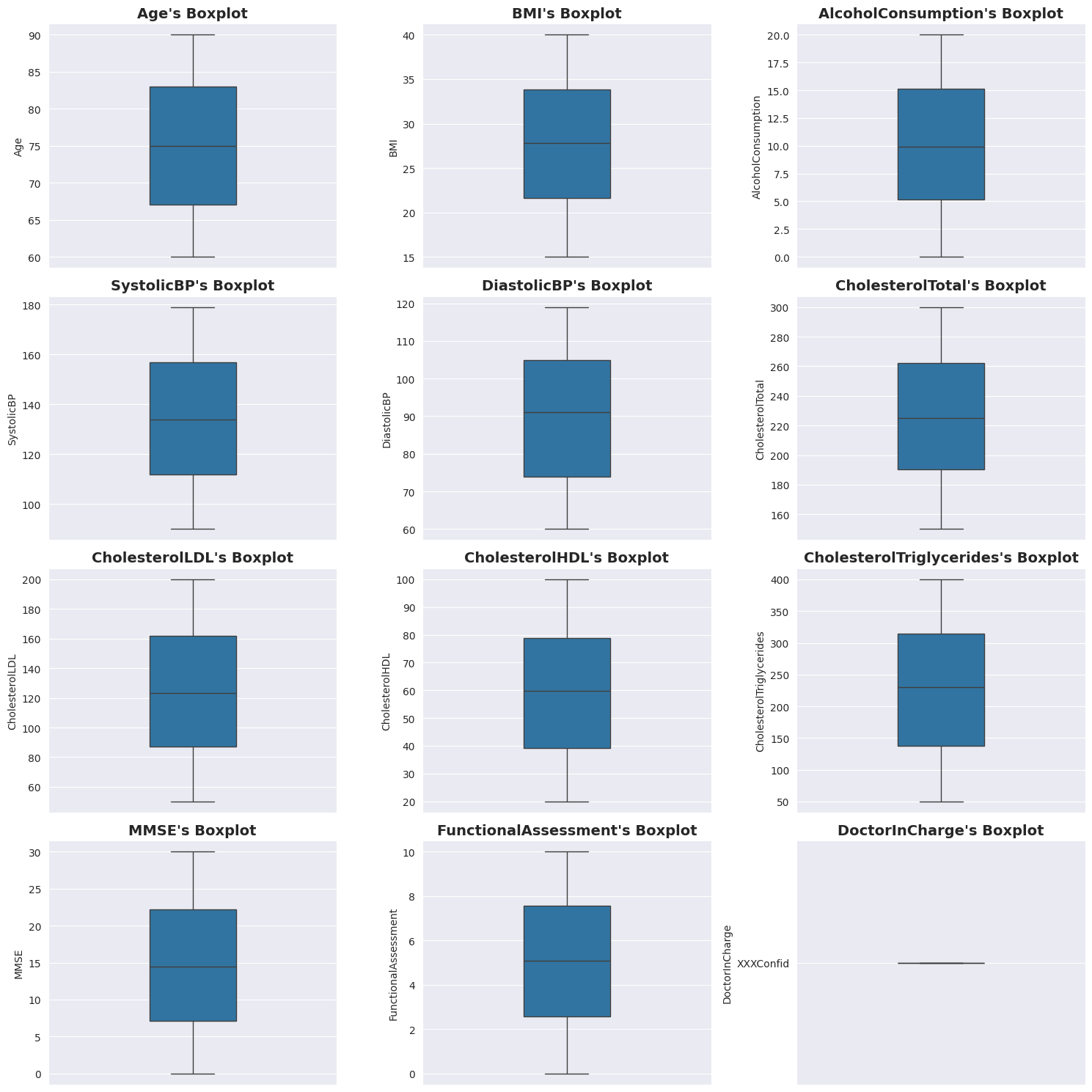
Cholesterol Levels: The cholesterol levels (total, HDL, LDL, and triglycerides) exhibit some peaks and variances. Monitoring these levels is essential for assessing cardiovascular health, which can influence Alzheimer’s progression.

Cognitive Assessment (MMSE): The MMSE scores show a concentration around certain values, indicating varying cognitive statuses among patients. This may reflect the progression of the disease.

Functional Assessment: The functional assessment scores appear to have a significant spread, suggesting that the patients experience different levels of functional impairment.

Doctor In Charge: The last plot appears to be empty, indicating that there may be a lack of variance in this categorical variable or an issue in the dataset.

**Boxplot**

****

**Inference**

Age:

The median age is around 77 years, with a range from approximately 60 to 90 years.

The interquartile range (IQR) is about 10-15 years, suggesting a relatively consistent age distribution among the patients.

There may be some outliers at the higher end, indicating a few patients are significantly older than the rest.

BMI:

The median BMI is around 28, suggesting that many patients are overweight or borderline obese.

The IQR is around 25-32, indicating that most patients have BMIs within this range.

There are a few outliers on the higher end, indicating some patients may have very high BMI values.

Alcohol Consumption:

The median alcohol consumption is about 7.5, with a wide range (0 to 20), indicating some patients do not consume alcohol, while others consume high amounts.

The IQR suggests that most patients consume between 3 and 12 units of alcohol.

Systolic Blood Pressure (BP):

The median systolic BP is around 140 mmHg, which is on the higher end of the normal range (120-130 mmHg) and may indicate some patients are at risk for hypertension.

The IQR is around 130-150 mmHg, showing that most patients have systolic BP readings within this range.

Diastolic BP:

The median diastolic BP is approximately 90 mmHg, which is near the threshold for hypertension.

The IQR suggests most patients have diastolic BP values between 80 and 100 mmHg, indicating potential health risks.

Cholesterol Levels (HDL):

The median HDL cholesterol level appears to be around 50 mg/dL, which is generally considered acceptable.

The IQR is fairly narrow, indicating a consistent range among most patients.

Total Cholesterol:

The median total cholesterol is about 240 mg/dL, which may be concerning as it is at the borderline high level.

There are some outliers indicating that some patients have significantly higher total cholesterol levels.

Triglycerides:

The median triglyceride level is around 200 mg/dL, which is at the upper end of the normal range.

The IQR suggests a reasonable spread, indicating variability in triglyceride levels among patients.

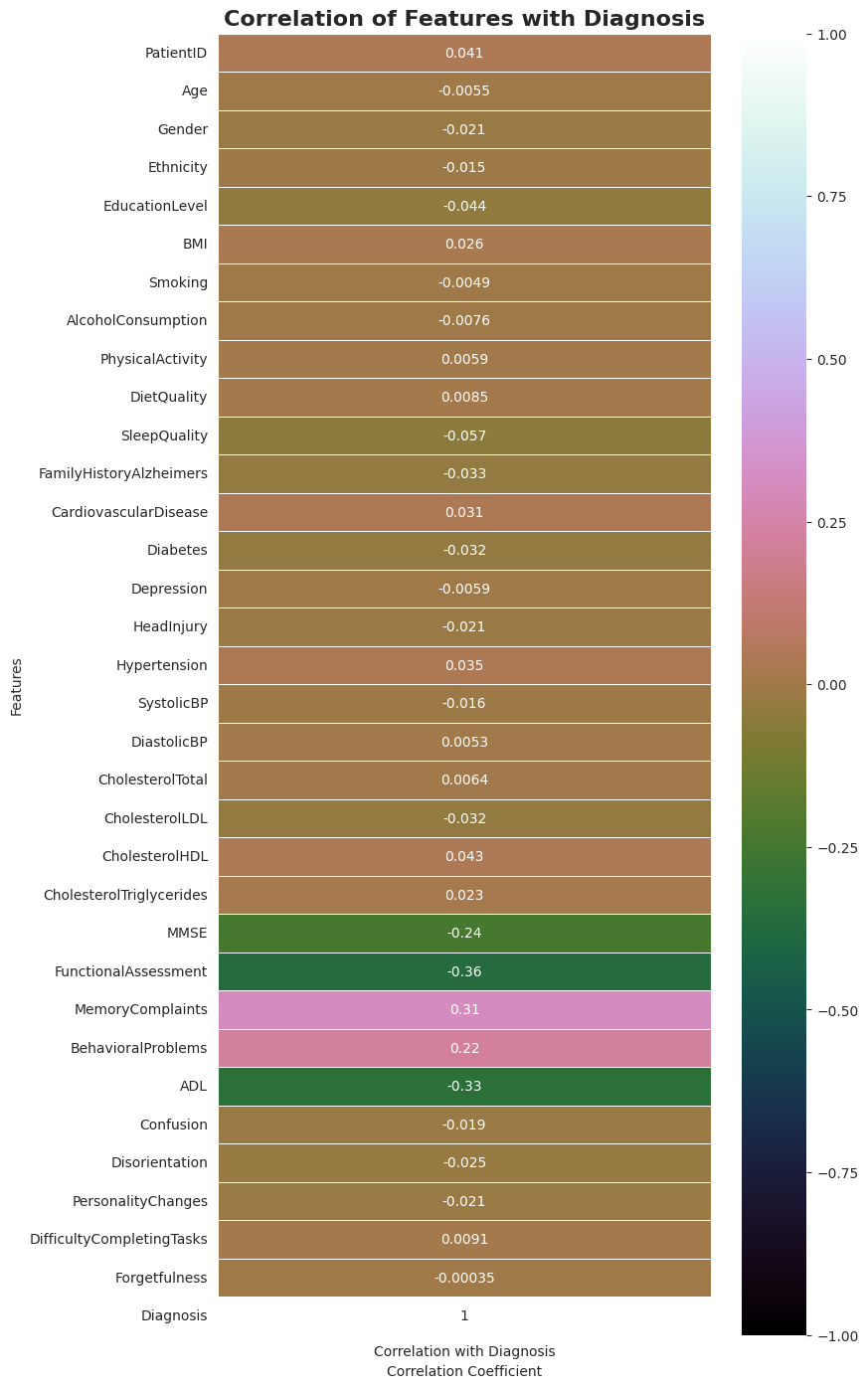
General Inferences:

Health Risks: Several measures indicate potential health risks, such as high BMI, elevated blood pressure, and borderline cholesterol levels, which are critical for the management of Alzheimer's disease.

Variability: The presence of outliers in BMI and cholesterol levels indicates a subset of patients with particularly high values, warranting closer monitoring.

Public Health Implications: The data suggest that interventions focused on weight management, blood pressure control, and cholesterol management could be beneficial for this patient population.

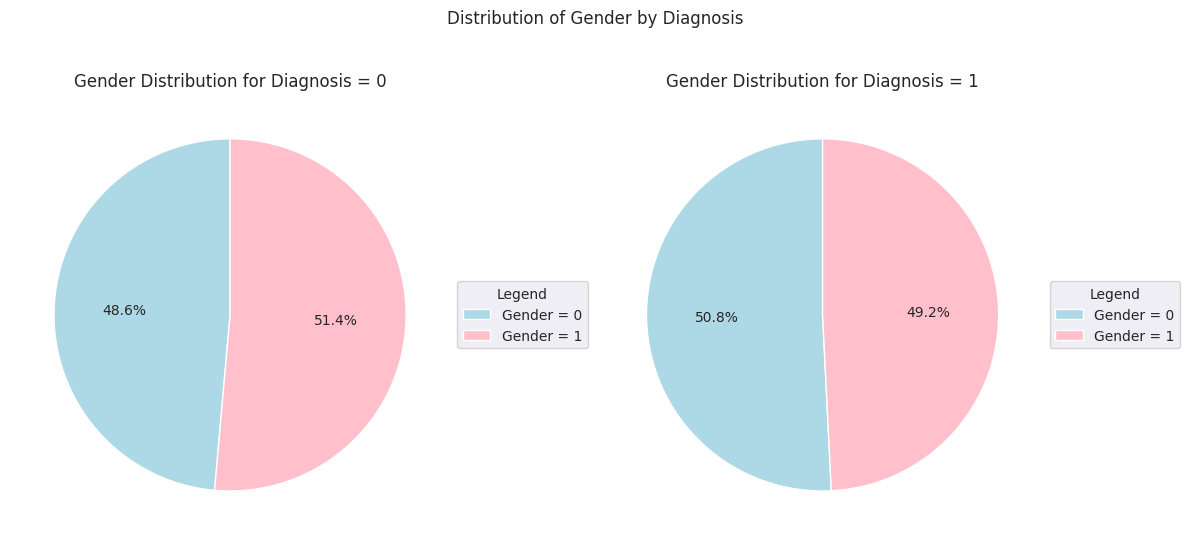
**Heatmap**

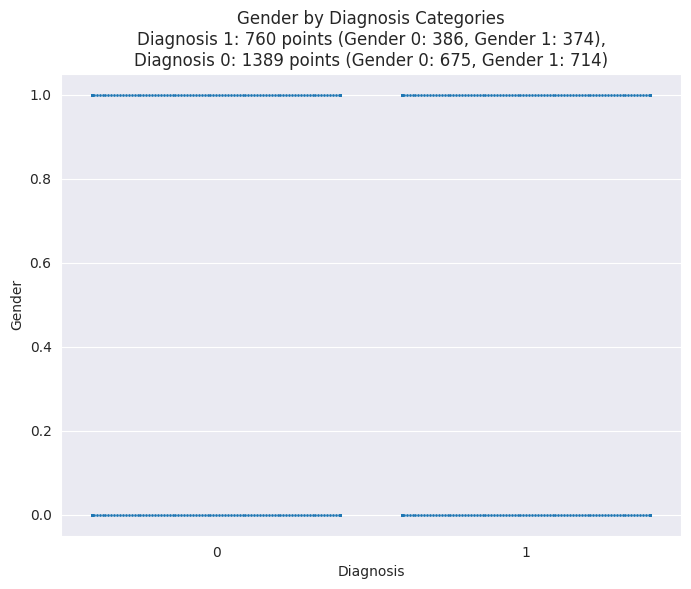
****

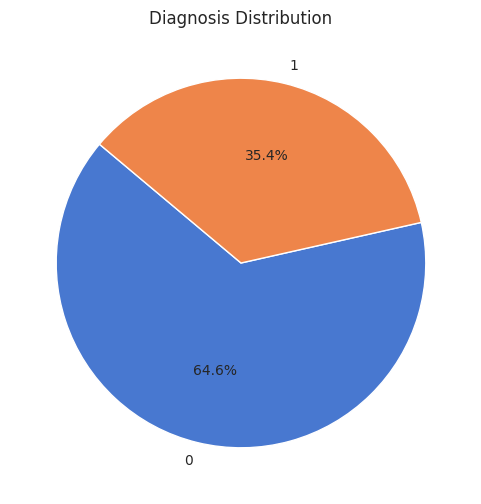
**Inference**

The heatmap reveals that the features do not have any strong correlations among themselves. However, there are five columns that show a correlation with the target variable.

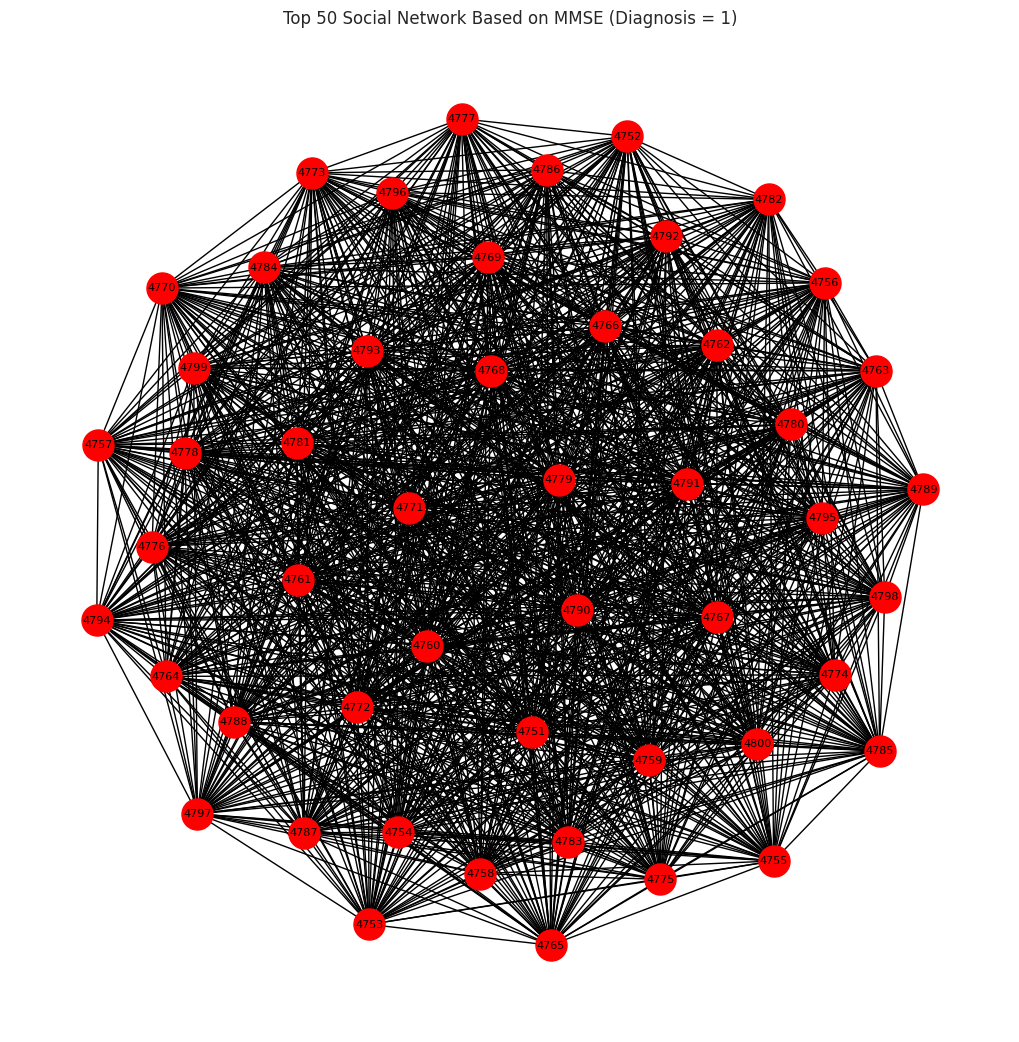
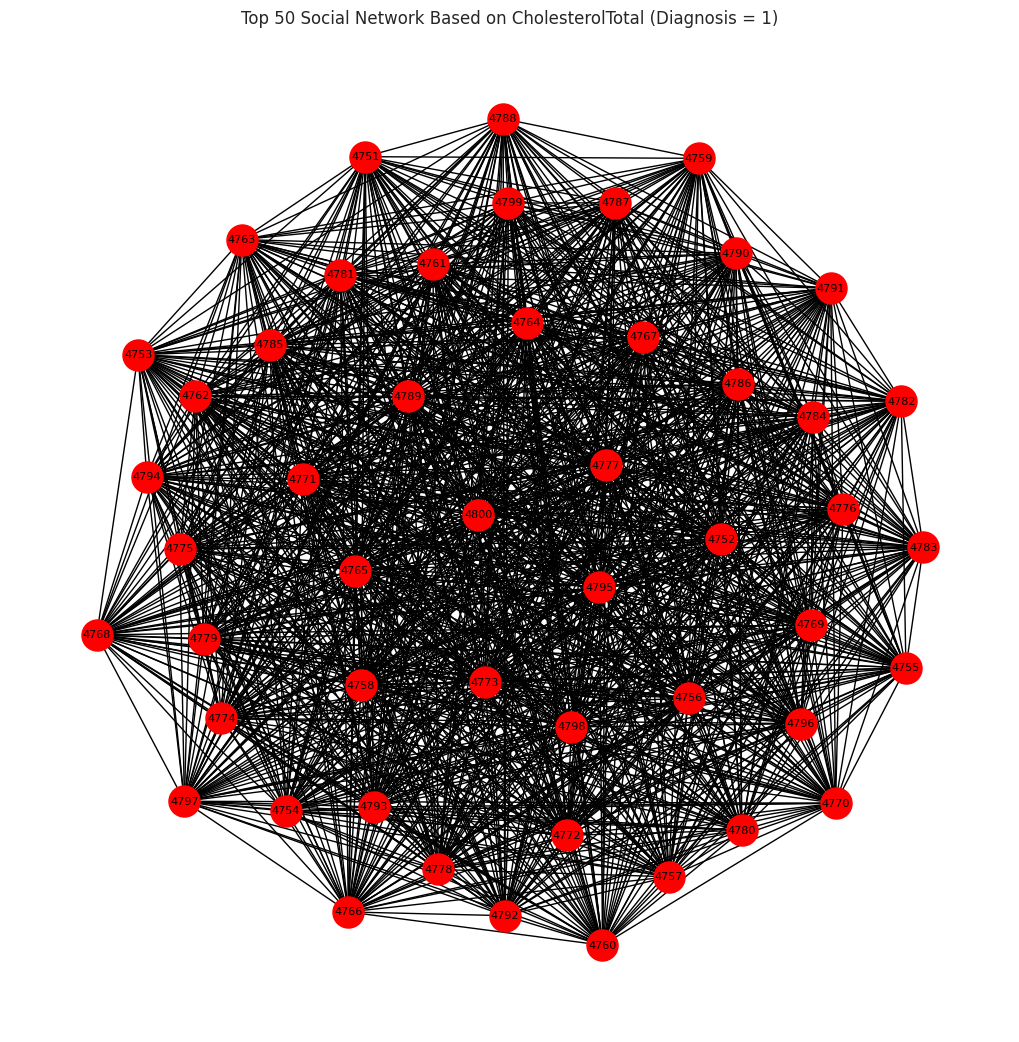
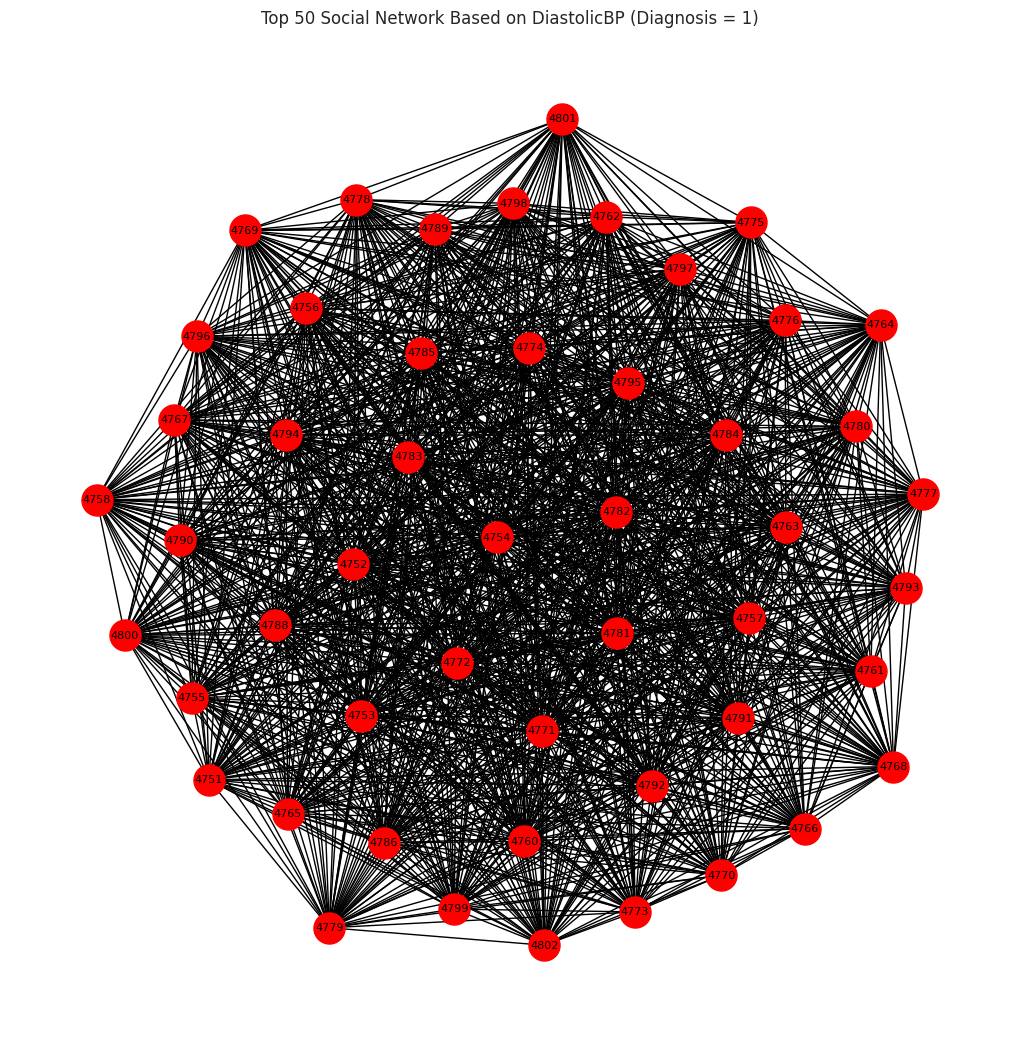
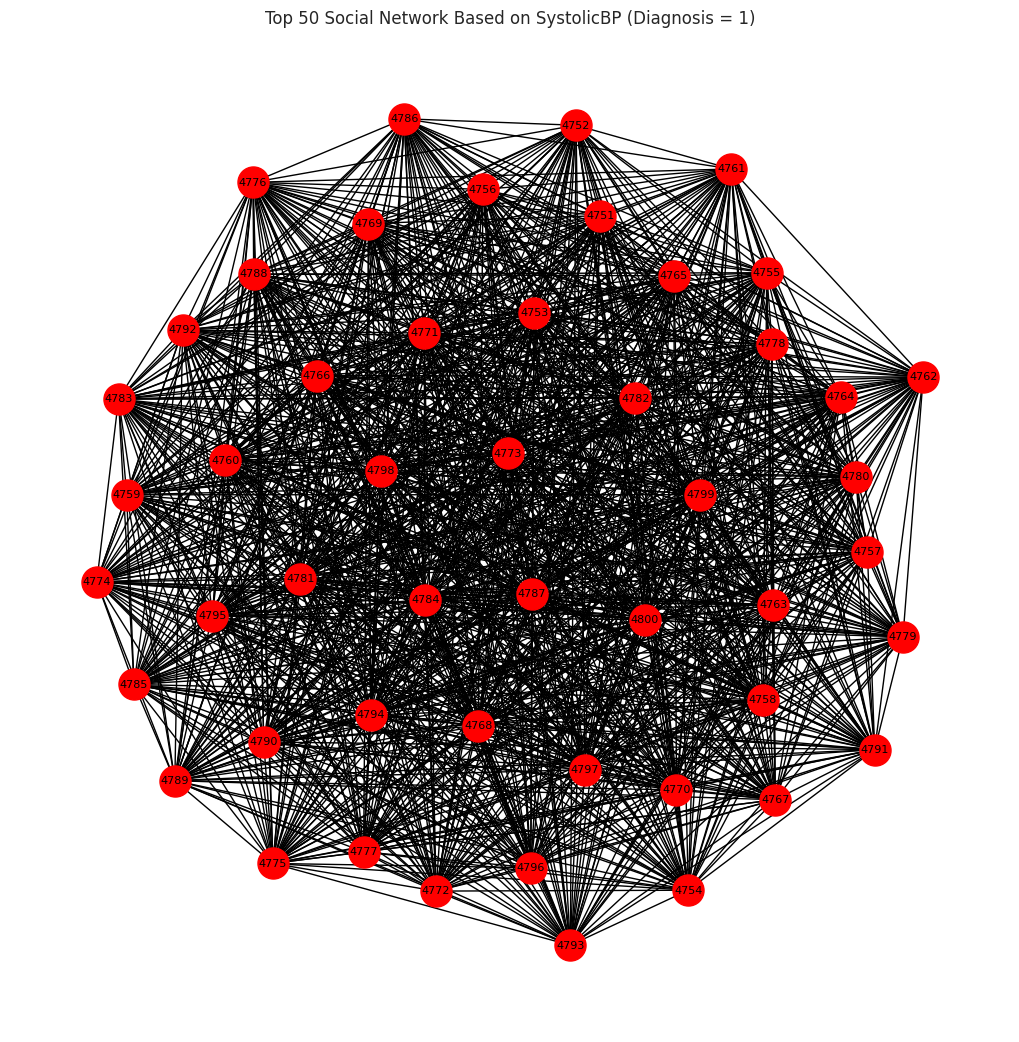
calculating Pearson correlation coefficient, also known as Pearson's r. It is a measure of the linear relationship between two variables. It quantifies the degree to which a pair of variables are linearly related, ranging from -1 to 1.

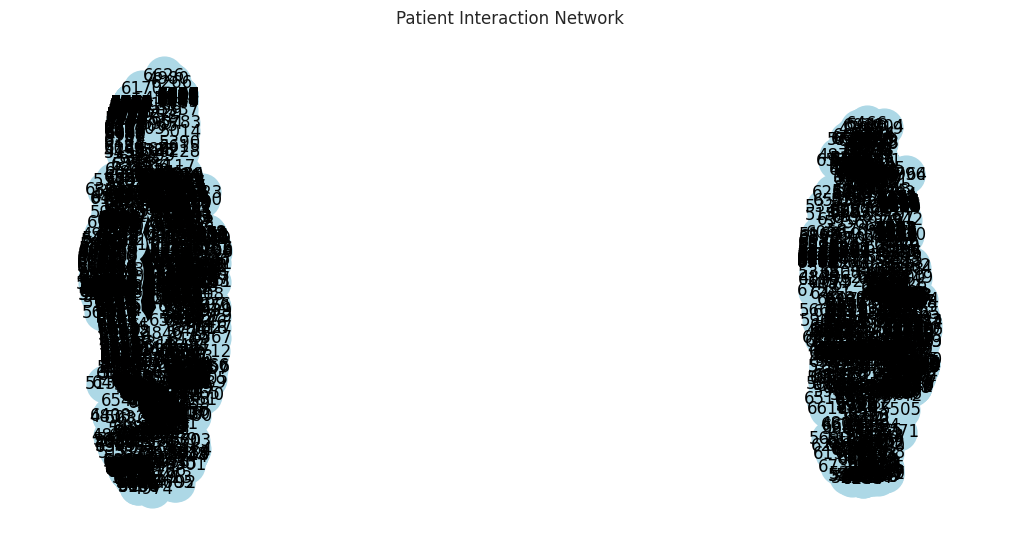




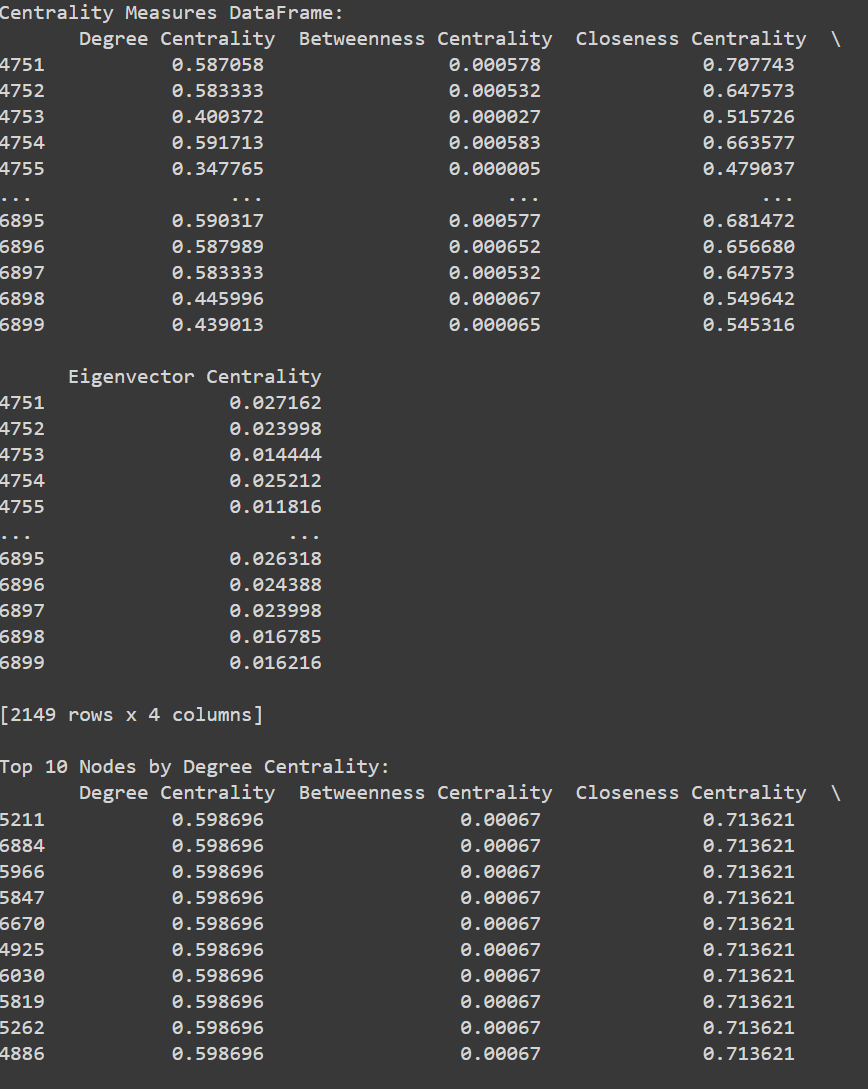


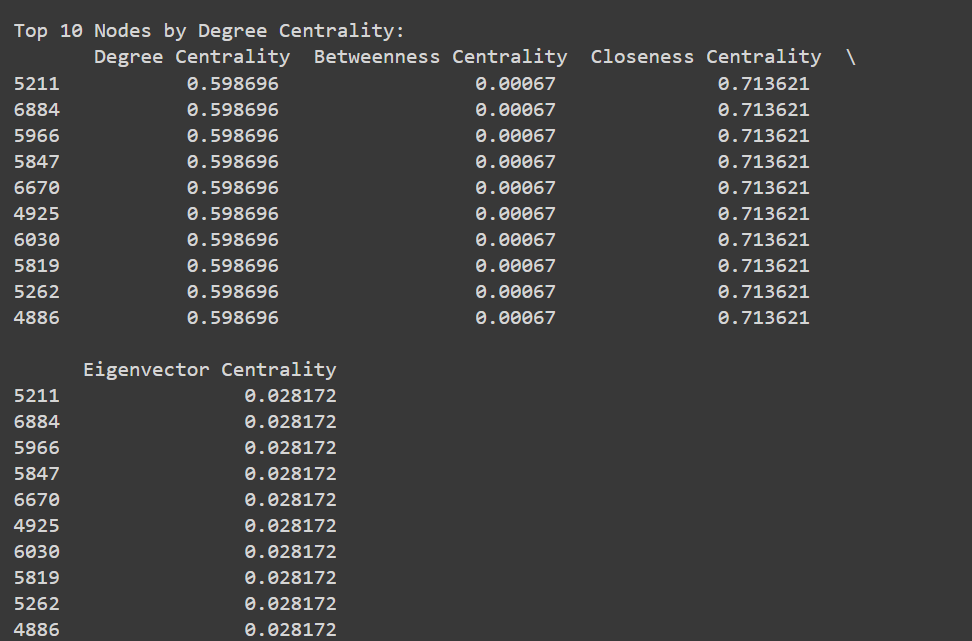
**Network**

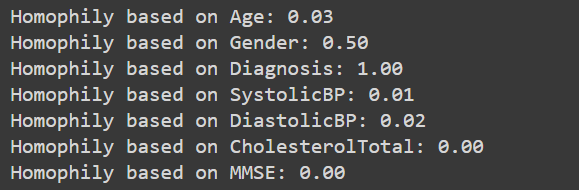




**Centrality measures**

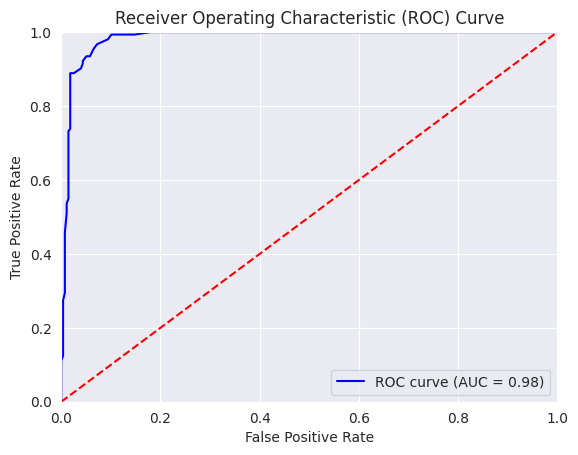
****

****

****

**Modeling**

****

****

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dataset :** [**https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset**](https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset)

**Reference research paper :** [**https://doi.org/10.1080/13607863.2020.1745146**](https://doi.org/10.1080/13607863.2020.1745146)

**Code :** [**https://colab.research.google.com/drive/14ankPsT23z0r2weoxS6B-OLiR7jMxJGs?usp=sharing**](https://colab.research.google.com/drive/14ankPsT23z0r2weoxS6B-OLiR7jMxJGs?usp=sharing)