Correlation Analysis Report

Student Name : Ali Shoaib

Stutent Roll No: NIM-BSCS-2021-33

Intructor : Mohammad Bilal

Task No 3: Python Implementation

Import Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Read Data file and use short names for Columns

```
# Read the CSV file into a DataFrame
df = pd.read_csv("/content/1- mental-illnesses-prevalence.csv")

# Create a dictionary mapping original column names to short names
short_names = {
    'Schizophrenia disorders (share of population) - Sex: Both - Age: Age-standardized': 'Schizophrenia',
    'Depressive disorders (share of population) - Sex: Both - Age: Age-standardized': 'Depressive',
    'Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized': 'Anxiety',
    'Bipolar disorders (share of population) - Sex: Both - Age: Age-standardized': 'Bipolar',
    'Eating disorders (share of population) - Sex: Both - Age: Age-standardized': 'Eating'
}

# Rename columns in the DataFrame using short names
df.rename(columns=short_names, inplace=True)
```

Find Descriptive Statistics for all numeric Columns

```
# Rename columns in the DataFrame using short names
df.rename(columns=short_names, inplace=True)

# Iterate over each column in the DataFrame
for column in df.columns:
    # Check if the column contains numeric values
    if pd.api.types.is_numeric_dtype(df[column]):
        # Calculate descriptive statistics for the numeric column
        mean_value = df[column].mean()
        mode_value = df[column].mode().iloc[0]  # Mode might return multiple values, so we take the first one
        median_value = df[column].median()
        variance_value = df[column].var()
        std_dev_value = df[column].std()

        # Print the statistics for the column
        print(f"\nColumn: {column}")
        print(f"Mean: {mean_value}")
        print(f"Mode: {mode_value}")
        print(f"Median: {median_value}")
        print(f"Variance: {variance_value}")
        print(f"Standard Deviation: {std_dev_value}")
```

```
Column: Year
Mean: 2004.5
Mode: 1990
Median: 2004.5
Variance: 74.92833774731267
Standard Deviation: 8.656115626960668

Column: Schizophrenia
Mean: 0.26660410361214953
Mode: 0.2116415
Median: 0.27347716499999997
Variance: 0.0015510073366489547
Standard Deviation: 0.03938283048041309

Column: Depressive
```

```
Mean: 3.767035886853583
Mode: 2.9647393
Median: 3.6367716999999997
Variance: 0.8561538742698497
Standard Deviation: 0.9252858338210143

Column: Anxiety
Mean: 4.101839659922118
Mode: 3.531769
Median: 3.9395473
Variance: 1.10364080393327
Standard Deviation: 1.0505430995124712

Column: Bipolar
Mean: 0.636968395228972
Mode: 0.54847974
Median: 0.5793314
Variance: 0.05447124853970372
Standard Deviation: 0.2333907636126668

Column: Eating
Mean: 0.19566409069797508
Mode: 0.0658978
Median: 0.144150375
Variance: 0.01914908557740843
Standard Deviation: 0.13838022104841582
```

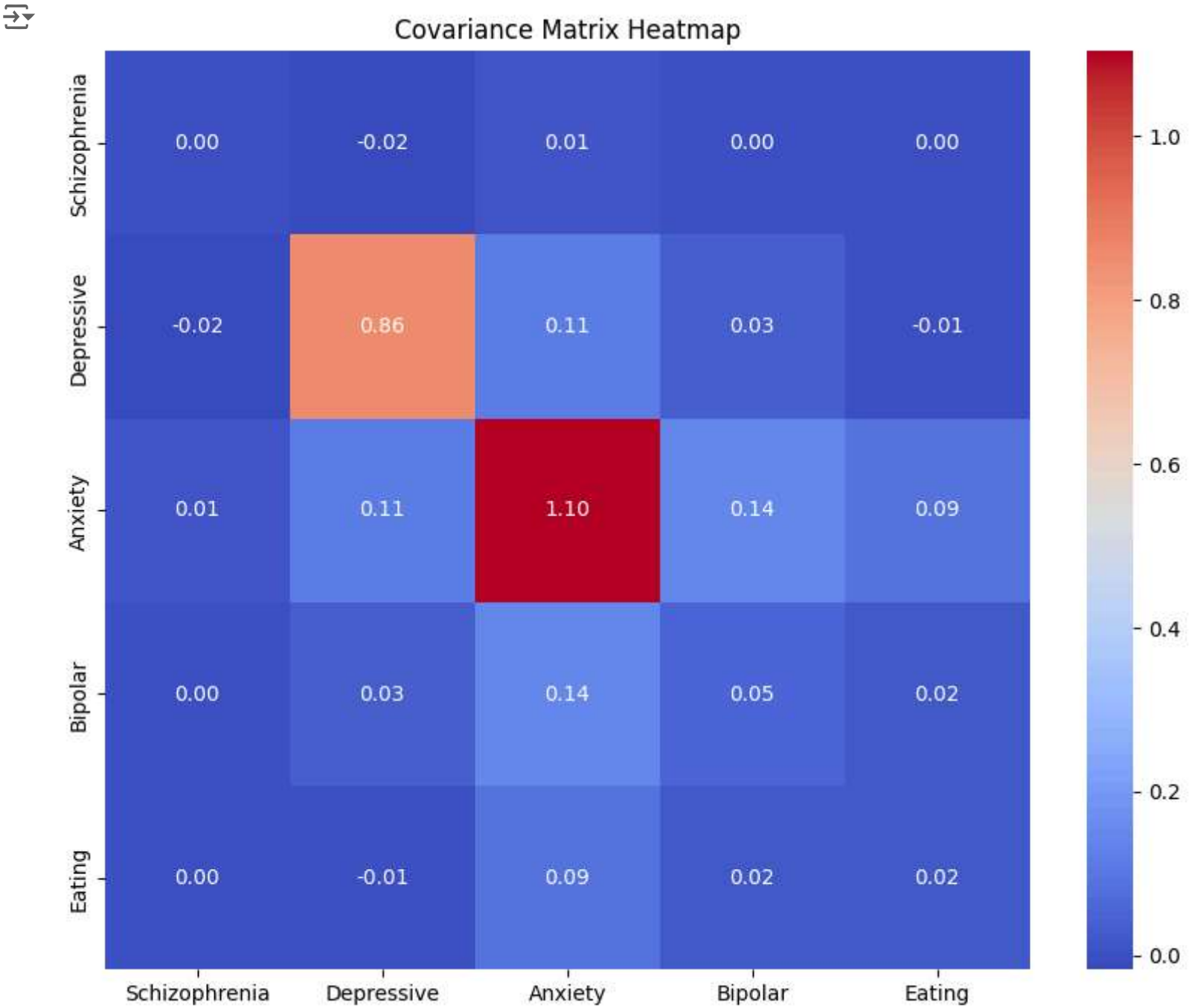
Covariance of each pair of Column

```
# Create a list of numeric columns with short names
numeric_columns = list(short_names.values())

# Create a DataFrame containing only numeric columns
numeric_df = df[numeric_columns]

# Calculate the covariance matrix
covariance_matrix = numeric_df.cov()

print("This is Covariance Matrix:")
print(covariance_matrix)
```

```
This is Covariance Matrix:
               Schizophrenia  Depressive   Anxiety   Bipolar    Eating
Schizophrenia       0.001551   -0.017309  0.012584  0.001117  0.002728
Depressive         -0.017309    0.856154  0.111231  0.033049 -0.006667
Anxiety             0.012584    0.111231  1.103641  0.141284  0.086427
Bipolar             0.001117    0.033049  0.141284  0.054471  0.021895
Eating              0.002728   -0.006667  0.086427  0.021895  0.019149
```

Representation of Covariance using Heatmap

```
# Plot the heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(covariance_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Covariance Matrix Heatmap')
plt.show()
```


Covariance Matrix Heatmap
Schizophrenia
Depressive
Anxiety
Bipolar
Eating
0.00 -0.02 0.01 0.00 0.00
-0.02 0.86 0.11 0.03 -0.01
0.01 0.11 1.10 0.14 0.09
0.00 0.03 0.14 0.05 0.02
0.00 -0.01 0.09 0.02 0.02
Schizophrenia
Depressive
Anxiety
Bipolar
Eating
1.0
0.8
0.6
0.4
0.2
0.0

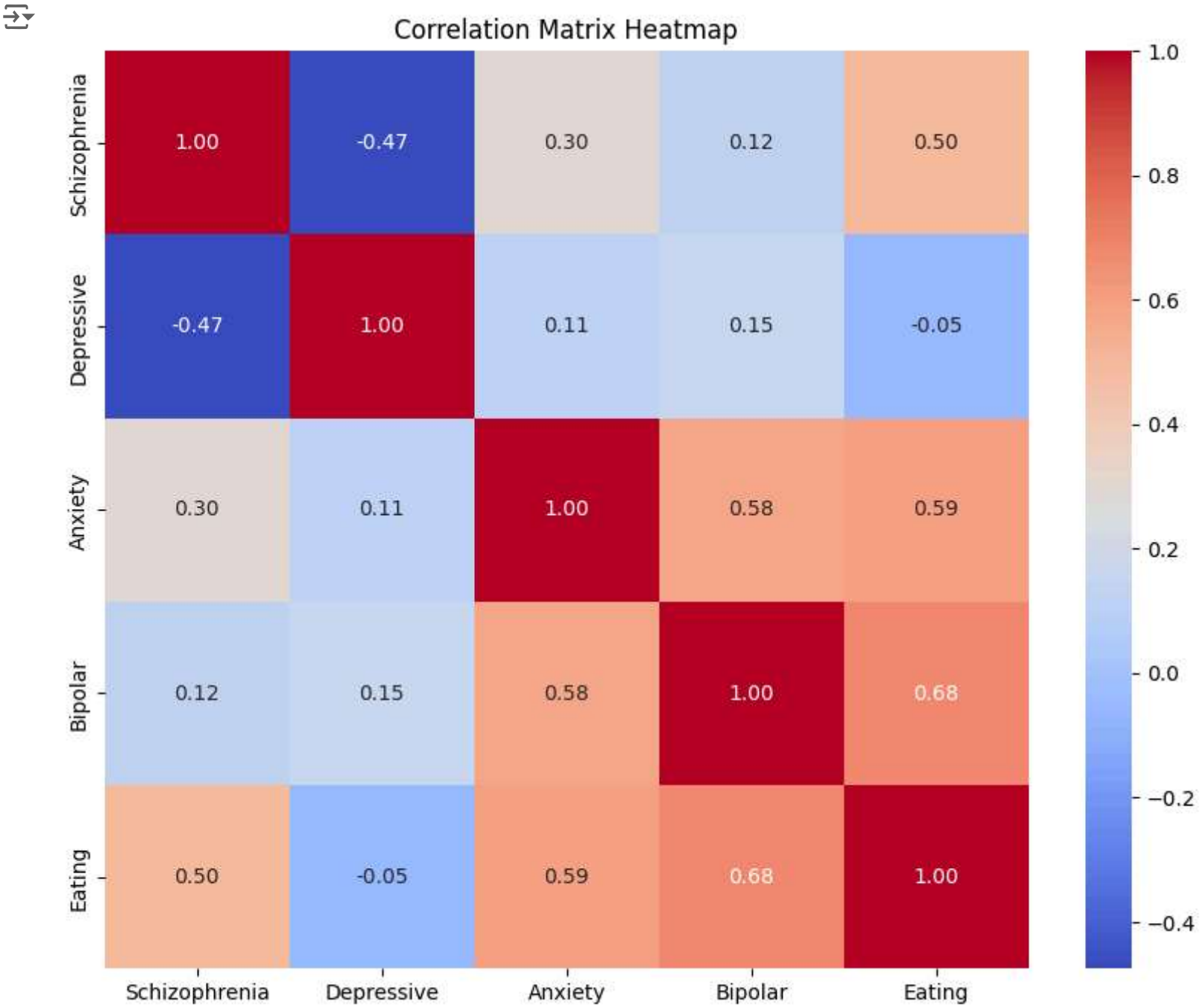
Corelation Coefficient for each pair

```
# Calculate the correlation matrix
correlation_matrix = numeric_df.corr()
print(" This is Correlation Matrix:")
print(correlation_matrix)
```

```
 This is Correlation Matrix:
                Schizophrenia  Depressive   Anxiety   Bipolar    Eating
Schizophrenia       1.000000   -0.474994  0.304149  0.121542  0.500656
Depressive         -0.474994    1.000000  0.114429  0.153039 -0.052067
Anxiety             0.304149    0.114429  1.000000  0.576230  0.594511
Bipolar             0.121542    0.153039  0.576230  1.000000  0.677927
Eating              0.500656   -0.052067  0.594511  0.677927  1.000000
```

Representation of Correlation using Heatmap

```
# Plot the heatmap for correlation
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix Heatmap')
plt.show()
```


Correlation Matrix Heatmap
Schizophrenia
Depressive
Anxiety
Bipolar
Eating
1.00
-0.47
0.30
0.12
0.50
-0.47
1.00
0.11
0.15
-0.05
0.30
0.11
1.00
0.58
0.59
0.12
0.15
0.58
1.00
0.68
0.50
-0.05
0.59
0.68
1.00
Schizophrenia
Depressive
Anxiety
Bipolar
Eating
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Analysis

Significant correlation results:

Eating disorders and anxiety disorders: Moderate to strong positive correlation (0.59).

Implication: Individuals with anxiety disorders may have an increased risk of developing eating disorders, and vice versa. Screening for both conditions can aid in early detection and intervention. Bipolar disorders and eating disorders: Moderately strong positive correlation (0.68).

Implication: Individuals with bipolar disorders may be more prone to developing eating disorders, and vice versa. Comprehensive treatment plans should address both conditions simultaneously for improved outcomes. Depressive disorders and schizophrenia: Moderate negative correlation (-0.47).