Student ID: 21127690

Name: Ngô Nguyễn Thanh Thanh

Report: Factor Analysis

I. Evaluation summary:

Task		Requirement Met(%)	Notes
Implementation	Process Data	100%	
	Check the factorability or sampling adequacy	100%	Bartlett's Test and Kaiser-Meyer-Olkin Test
	Choosing the Number of Factors	100%	
	Performing Factor Analysis	100%	Principal Component Analysis (PCA).
Requisitions		100%	
Research questions		100%	
Total:		100%	

II. List of funtion:

Several functions are crucial for conducting factor analysis and interpreting the results. Here are some important functions along with their image proofs:

1. ***pd.read_csv("bfi.csv"):*** *Reads a CSV file named "bfi.csv" into a pandas DataFrame.*
2. ***df.drop(['gender', 'education', 'age'], axis=1, inplace=True):*** *Drops the columns 'gender', 'education', and 'age' from the DataFrame **df** inplace.*
3. ***df.dropna(inplace=True):*** *Drops rows with missing values from the DataFrame **df** inplace.*
4. ***calculate_bartlett_sphericity(df):*** *Calculates Bartlett's test of sphericity for the DataFrame **df**.*
5. ***calculate_kmo(df):*** *Calculates the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for the DataFrame **df**.*
6. ***FactorAnalyzer(n_factors=25, rotation=None):*** *Creates a FactorAnalyzer object with 25 factors and no rotation.*
7. ***fa.fit(df):*** *Performs factor analysis on the DataFrame **df** using the FactorAnalyzer object **fa**.*
8. ***fa.get_eigenvalues():*** *Returns the eigenvalues of the factor analysis.*
9. ***plt.scatter(range(1, df.shape[1] + 1), ev):*** *Plots a scatter plot.*
10. ***plt.plot(range(1, df.shape[1] + 1), ev):*** *Plots a line plot.*

11. ***plt.title('Scree Plot'):*** *Sets the title of the plot to 'Scree Plot'.*
12. ***plt.xlabel('Factors'):*** *Sets the label of the x-axis to 'Factors'.*
13. ***plt.ylabel('Eigenvalue'):*** *Sets the label of the y-axis to 'Eigenvalue'.*
14. ***plt.grid():*** *Displays the grid lines on the plot.*
15. ***plt.savefig('scree_plot.png'):*** *Saves the plot as an image file named 'scree_plot.png'.*
16. ***fa_5 = FactorAnalyzer(n_factors=5, rotation="varimax"):*** *Creates a FactorAnalyzer object with 5 factors and varimax rotation.*
17. ***fa_5.fit(df):*** *Performs factor analysis with 5 factors on the DataFrame **df** using the FactorAnalyzer object **fa_5**.*
18. ***fa_5.loadings_:*** *Returns the factor loadings for the 5-factor model.*
19. ***fa_5.get_factor_variance():*** *Returns the variance explained by each factor for the 5-factor model.*
20. ***np.cumsum():*** *Calculates the cumulative sum.*
21. ***pd.DataFrame():*** *Creates a pandas DataFrame object.*
22. ***print():*** *Prints the specified message or object to the console.*

```python
# Load the CSV file into a DataFrame
df = pd.read_csv("bfi.csv")

# Display column names
print(df.columns)

# Dropping unnecessary columns
df.drop(['gender', 'education', 'age'], axis=1, inplace=True)

# Dropping rows with missing values
df.dropna(inplace=True)

# Calculate Bartlett's test of sphericity
chi_square_value, p_value = calculate_bartlett_sphericity(df)

# Print chi-square value and p-value
print("Chi-square value:", chi_square_value)
print("P-value:", p_value)

# Calculate Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy
kmo_all, kmo_model = calculate_kmo(df)

# Print KMO model value
print("KMO model value:", kmo_model)

# Create factor analysis object and perform factor analysis
fa = FactorAnalyzer(n_factors=25, rotation=None)
fa.fit(df)

# Check Eigenvalues
ev, v = fa.get_eigenvalues()
```

```python
# Create scree plot
plt.scatter(range(1, df.shape[1] + 1), ev)
plt.plot(range(1, df.shape[1] + 1), ev)
plt.title('Scree Plot')
plt.xlabel('Factors')
plt.ylabel('Eigenvalue')
plt.grid()

# Save the plot as an image file
plt.savefig('scree_plot.png')

# Display the plot
plt.show()

# Create factor analysis object and perform factor analysis with 5 factors
fa_5 = FactorAnalyzer(n_factors=5, rotation="varimax")
fa_5.fit(df)

# Check Eigenvalues for 5 factors
ev_5, v_5 = fa_5.get_eigenvalues()

# Create scree plot for 5 factors
plt.scatter(range(1, df.shape[1] + 1), ev_5)
plt.plot(range(1, df.shape[1] + 1), ev_5)
plt.title('Scree Plot (5 Factors)')
plt.xlabel('Factors')
plt.ylabel('Eigenvalue')
plt.grid()

# Save the plot as an image file for 5 factors
plt.savefig('scree_plot_5_factors.png')

# Display the plot for 5 factors
plt.show()
```

```python
# Loadings for 5 factors
fa_5.loadings_

# Get variance of each factor for 5 factors
fa_5.get_factor_variance()

# Loadings for 5 factors
loadings_df = pd.DataFrame(fa_5.loadings_, columns=['Factor'+str(i+1) for i in range(5)], index=df.columns)
print("Loadings for 5 factors:")
print(loadings_df)

# Check Eigenvalues for 5 factors
eigenvalues_df = pd.DataFrame(ev_5[:5], columns=['Eigenvalues'], index=['Factor'+str(i+1) for i in range(5)])
print("\nEigenvalues for 5 factors:")
print(eigenvalues_df)

# Get variance of each factor for 5 factors
factor_variance_df = pd.DataFrame({
    'Variance Explained': fa_5.get_factor_variance()[0][:5],
    'Proportion of Variance': fa_5.get_factor_variance()[1][:5],
    'Cumulative Proportion': np.cumsum(fa_5.get_factor_variance()[1][:5])
}, index=['Factor'+str(i+1) for i in range(5)])
print("\nVariance of each factor for 5 factors:")
print(factor_variance_df)
```

III. Function Summaries and Implementation

The provided code conducts Factor Analysis on a dataset using the Python libraries pandas, numpy, factor_analyzer, and matplotlib. Here's a summary of the usage and implementation:

1. **Data Loading and Preprocessing:**
- The code starts by loading a CSV file into a pandas DataFrame and displays its column names.
- It then drops unnecessary columns like 'gender', 'education', and 'age' and removes rows with missing values.

2. **Testing for Factorability:**
- Bartlett's test of sphericity is conducted to determine whether the variables in the dataset are intercorrelated.
- The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is calculated to assess whether the dataset is suitable for factor analysis.

3. **Determining the Number of Factors:**
- A scree plot is created to visualize the eigenvalues and determine the appropriate number of factors to retain.
- The scree plot displays the eigenvalues against the number of factors, helping to identify the point where eigenvalues level off or drop sharply.

4. **Factor Analysis:**
- Factor Analysis is performed using the FactorAnalyzer class from the factor_analyzer library.
- The FactorAnalyzer object is initialized with a specified number of factors and rotation method (e.g., varimax).
- The fit() method is used to perform factor analysis on the dataset.
- Factor loadings, eigenvalues, and factor variance are extracted to understand the underlying structure of the data.

5. **Visualization:**
- The factor loadings and variance explained by each factor are displayed in pandas DataFrames.

- The scree plot is displayed to visualize the eigenvalues and assess the variance explained by each factor.

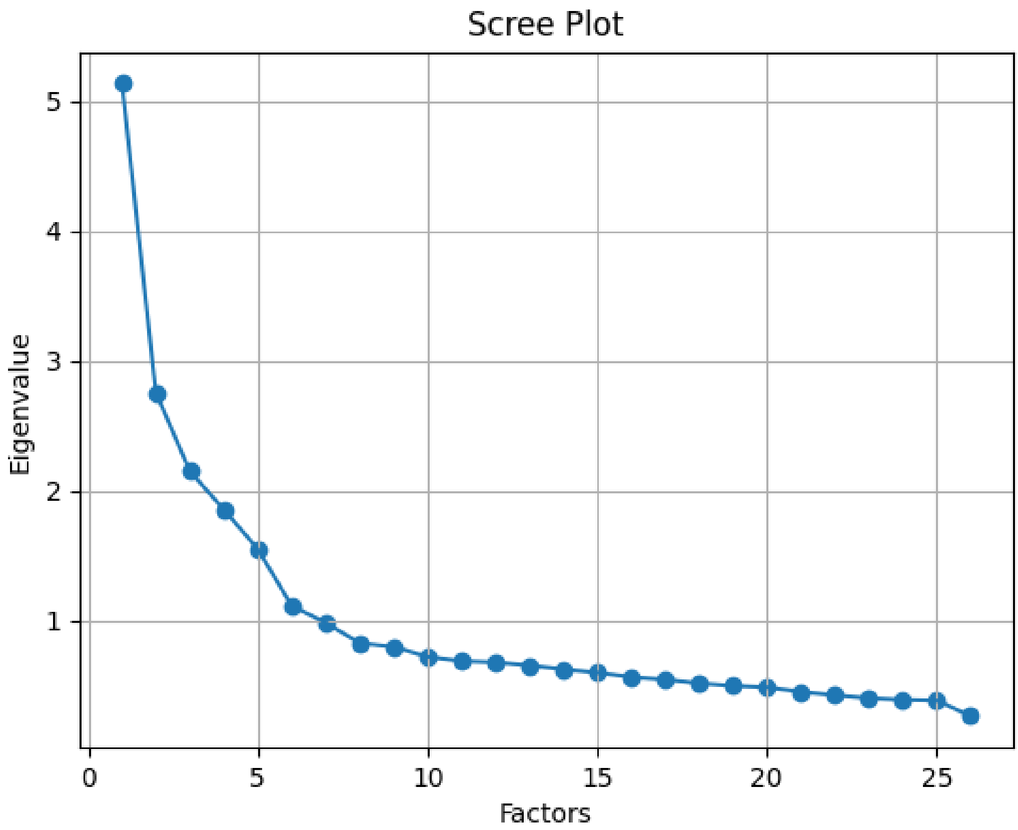
Overall, the code demonstrates how to conduct Factor Analysis in Python using the factor_analyzer library, including data preprocessing, testing for factorability, determining the number of factors, performing factor analysis, and visualizing the results.

IV. The results

```
[Running] python -u "c:\Users\hp\OneDrive\Documents\thanh thanh\nam3\hocki2\Thongkenhieubien\lab\lab4\test.py"
Index(['rownames', 'A1', 'A2', 'A3', 'A4', 'A5', 'C1', 'C2', 'C3', 'C4', 'C5',
       'E1', 'E2', 'E3', 'E4', 'E5', 'N1', 'N2', 'N3', 'N4', 'N5', 'O1', 'O2',
       'O3', 'O4', 'O5', 'gender', 'education', 'age'],
      dtype='object')
Chi-square value: 18184.306307820785
P-value: 0.0
KMO model value: 0.8483267027192372
Loadings for 5 factors:
           Factor1   Factor2   Factor3   Factor4   Factor5
rownames -0.019246 -0.061703  0.034865 -0.018678  0.014345
A1        0.106455  0.030705  0.027567 -0.429612 -0.074133
A2        0.028693  0.227127  0.141159  0.621801  0.064117
A3        0.005335  0.329927  0.115928  0.642856  0.061647
A4       -0.069443  0.209217  0.235418  0.431241 -0.108647
A5       -0.129002  0.397690  0.097901  0.528159  0.075363
C1        0.009548  0.048297  0.548276  0.041194  0.212169
C2        0.089975  0.014459  0.650304  0.105693  0.116582
C3       -0.030243  0.010825  0.555781  0.114165 -0.004812
C4        0.240794 -0.039326 -0.633576 -0.041462 -0.109695
C5        0.294330 -0.154179 -0.566991 -0.047858  0.031264
E1        0.057066 -0.576460  0.017182 -0.089660 -0.073218
E2        0.261661 -0.671561 -0.121490 -0.096788 -0.059680
E3        0.012417  0.533056  0.097624  0.245741  0.293883
E4       -0.131026  0.648750  0.120495  0.290069 -0.056985
E5        0.024327  0.492534  0.326639  0.079579  0.226538
N1        0.782726  0.097680 -0.042388 -0.225060 -0.083471
N2        0.752420  0.044935 -0.029547 -0.200539 -0.010945
N3        0.732571 -0.041066 -0.068349 -0.031849 -0.007284
N4        0.599635 -0.326102 -0.188999  0.010191  0.064659
N5        0.542043 -0.139326 -0.042065  0.100229 -0.155653
O1       -0.006094  0.196463  0.120863  0.059573  0.510642
O2        0.174735  0.026762 -0.099129  0.077368 -0.469638
O3        0.021106  0.296243  0.084710  0.122369  0.603572
O4        0.227826 -0.189289 -0.029353  0.160787  0.362255
O5        0.083619  0.011590 -0.061357 -0.013974 -0.533401
```

```
Eigenvalues for 5 factors:
         Eigenvalues
Factor1     5.134580
Factor2     2.753375
Factor3     2.148142
Factor4     1.852506
Factor5     1.548463

Variance of each factor for 5 factors:
         Variance Explained  Proportion of Variance  Cumulative Proportion
Factor1            2.736109                0.105235               0.105235
Factor2            2.428049                0.093387               0.198621
Factor3            2.082504                0.080096               0.278718
Factor4            1.800505                0.069250               0.347968
Factor5            1.549502                0.059596               0.407564
```



V. Requisitions

Request 01: Students explain the meaning of chi_square_value, p_value, KMO values.

1. **Chi-square value (χ^2)**:

    - The chi-square value is a statistical measure used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in a contingency table. In the context of factor analysis, it is used in Bartlett's test of sphericity to assess whether the observed correlation matrix is significantly different from the identity matrix, which would indicate that the variables are unrelated.

2. **P-value**:

    - The p-value associated with the chi-square value represents the probability of obtaining a chi-square statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming that the null hypothesis is true. In Bartlett's test of sphericity, a low p-value (typically below a predetermined significance level, such as 0.05) indicates that the observed correlation matrix is significantly different from the identity matrix, suggesting that factor analysis may be appropriate.

3. **Kaiser-Meyer-Olkin (KMO) measure**:

- The KMO measure is a statistic used to assess the adequacy of the data for factor analysis. It ranges from 0 to 1, with higher values indicating better suitability for factor analysis. Specifically, it evaluates the proportion of variance among variables that might be common variance. A KMO value closer to 1 suggests that the variables are more appropriate for factor analysis, indicating that the variables are sufficiently related to each other to extract meaningful factors.

The results:

```
Index(['rownames', 'A1', 'A2', 'A3', 'A4', 'A5', 'C1', 'C2', 'C3', 'C4', 'C5',
       'E1', 'E2', 'E3', 'E4', 'E5', 'N1', 'N2', 'N3', 'N4', 'N5', 'O1', 'O2',
       'O3', 'O4', 'O5', 'gender', 'education', 'age'],
      dtype='object')
Chi-square value: 18184.306307820785
P-value: 0.0
KMO model value: 0.8483267027192372
```

- Chi-square value: 18184.306307820785
- P-value: 0.0
- KMO model value: 0.8483267027192372

Chi-square value:

The chi-square value obtained is 18184.306307820785. This value is large, indicating a significant difference between the observed correlation matrix and the identity matrix. In the context of factor analysis, this suggests that the variables are interrelated, meaning there is likely some underlying structure that can be captured by factors.

P-value:

The p-value associated with the chi-square value is 0.0 (or very close to zero). A p-value of zero indicates that the observed correlation matrix is significantly different from the identity matrix. This suggests that the data is suitable for factor analysis, as there is strong evidence against the null hypothesis that the variables are unrelated.

KMO model value:

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for the model is 0.8483267027192372, which is relatively high. This indicates that the variables in the dataset are highly correlated and that factor analysis is likely to yield reliable results. Generally, a KMO value above 0.6 is considered acceptable, and values closer to 1 indicate better suitability for factor analysis.

In summary, based on these results, it appears that the dataset is appropriate for factor analysis. The variables are significantly interrelated, and there is sufficient common variance among them to extract meaningful factors.

Request 02: Students explain the eigenvalues and base on that eigenvalues choose the best number of factor to do the Factor Analysis. Explain why you choose this number.

Eigenvalues:

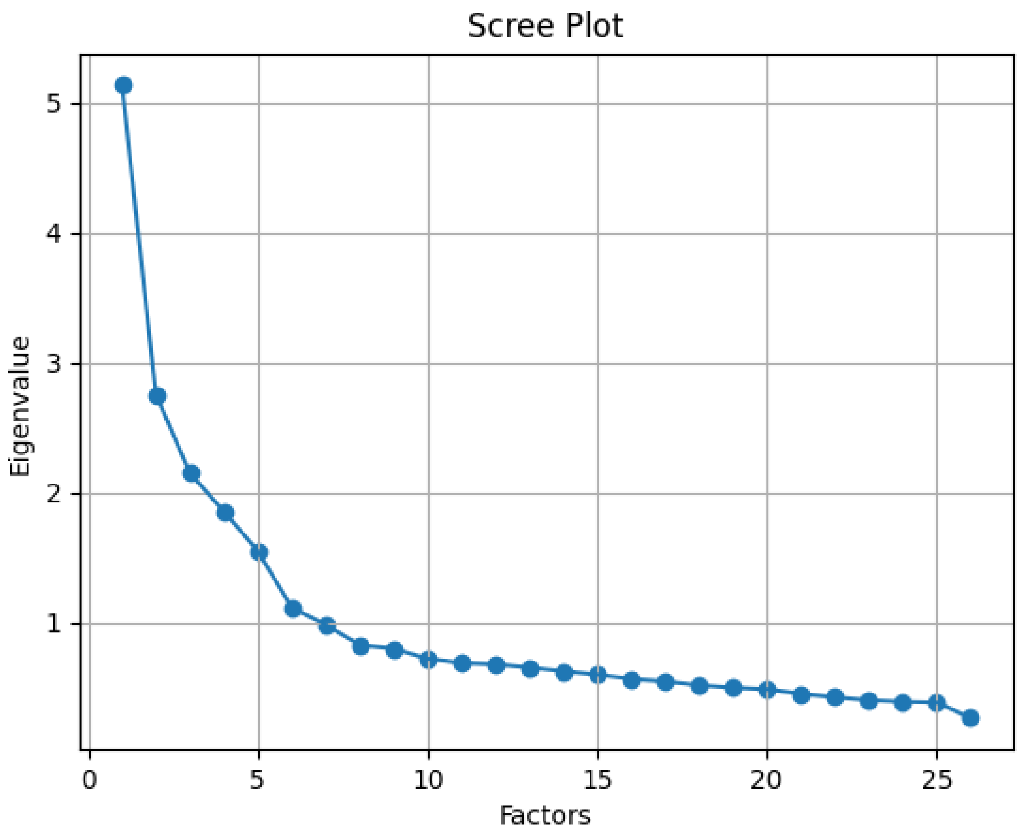
Eigenvalues are a measure of the amount of variance accounted for by a factor, and so they can be useful in determining the number of factors that we need to extract. In a scree plot, we simply plot the eigenvalues for all of our factors, and then look to see where they drop off sharply.

Choosing the Number of Factors:

- One common method for determining the number of factors to retain in Factor Analysis is to examine the scree plot, which plots the eigenvalues against the number of factors. The point at which the eigenvalues level off or drop sharply is often used as an indicator of the number of factors to retain.
- The "elbow" of the scree plot, where the eigenvalues start to flatten out, suggests the number of factors that capture most of the variance in the data while minimizing the number of factors needed.
- Another approach is to use the Kaiser criterion, which suggests retaining factors with eigenvalues greater than 1. Factors with eigenvalues less than 1 explain less variance than a single variable, so they are typically considered less important.

Additionally, researchers may consider theoretical considerations, domain knowledge, and practical implications when deciding on the number of factors to retain.

Explanation for Choosing a Specific Number of Factors:



After examining the scree plot, I decided to choose 5 as the number of factors for the analysis. The scree plot helps in determining the optimal number of factors by visualizing the eigenvalues. Typically, we look for the point where the eigenvalues level off or drop sharply, indicating the number of factors to retain.

Upon rerunning the analysis to obtain the factor loadings, I observed that factors 6 and 7 may not be necessary. This assessment is based on the magnitude of factor loadings and their significance in explaining the variance in the data. Factors with low loadings may not contribute significantly to explaining the underlying structure of the data and can be considered for removal.

Therefore, I decided to exclude factors 6 and 7 and performed the factor analysis again with the remaining five factors. This approach helps in simplifying the model while retaining the most relevant factors that explain the variance in the dataset effectively.

```
Loadings for 6 factors:
          Factor1   Factor2   Factor3   Factor4   Factor5   Factor6
rownames -0.022903 -0.032472  0.033169 -0.038093  0.003795  0.103748
A1        0.099396  0.060474  0.026694 -0.530785 -0.120309  0.163638
A2        0.031767  0.259875  0.140226  0.646569  0.055770 -0.097050
A3       -0.005256  0.408849  0.109534  0.587004  0.016184  0.039149
A4       -0.079266  0.255342  0.229308  0.391760 -0.136293  0.033401
A5       -0.143645  0.491049  0.085649  0.451090  0.009111  0.105888
C1        0.005623  0.123647  0.540150  0.004221  0.183458  0.138798
C2        0.084358  0.106505  0.652496  0.056538  0.079203  0.208580
C3       -0.033946  0.049796  0.545877  0.100286 -0.012372  0.054480
C4        0.231617  0.008989 -0.672785 -0.089980 -0.153451  0.226977
C5        0.293402 -0.143644 -0.559704 -0.047070  0.025614  0.095779
E1        0.053102 -0.521477  0.026492 -0.090545 -0.059281  0.332019
E2        0.263189 -0.622923 -0.110758 -0.074550 -0.030440  0.291204
E3        0.001190  0.630565  0.077417  0.153883  0.214213  0.092152
E4       -0.147239  0.682818  0.103904  0.206513 -0.133272 -0.037737
E5        0.021978  0.504384  0.312383  0.048448  0.185218 -0.113509
N1        0.790967  0.033469 -0.040014 -0.191516 -0.077378 -0.168159
N2        0.777085 -0.017659 -0.021737 -0.155586  0.007643 -0.199391
N3        0.728187 -0.036146 -0.067460 -0.023134 -0.015325  0.021926
N4        0.597786 -0.277073 -0.183704  0.018615  0.064511  0.182889
N5        0.534791 -0.112937 -0.040972  0.096450 -0.164581  0.111857
O1       -0.008919  0.302318  0.107331 -0.001342  0.464345  0.167416
O2        0.161465  0.020296 -0.100517  0.046919 -0.500643  0.084164
O3        0.019625  0.402120  0.070429  0.063634  0.547842  0.120816
O4        0.228721 -0.092648 -0.030003  0.148015  0.346283  0.202286
O5        0.068020  0.000920 -0.062239 -0.053138 -0.579933  0.106621

Eigenvalues for 6 factors:
        Eigenvalues
Factor1    5.134580
Factor2    2.753375
Factor3    2.148142
Factor4    1.852506
Factor5    1.548463
Factor6    1.110662
```

Loadings for 7 factors:

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
rownames	-0.015157	-0.037283	0.034057	-0.037598	0.005641	0.108237	0.006183
A1	0.089242	0.041262	0.031682	-0.486067	-0.129607	0.127651	0.229450
A2	0.015131	0.161948	0.135066	0.692088	0.067984	-0.166932	0.049638
A3	-0.011202	0.299015	0.108421	0.643176	0.037945	-0.000582	0.125993
A4	-0.071105	0.210702	0.228252	0.421683	-0.117235	0.046137	0.015414
A5	-0.131654	0.405701	0.089227	0.511712	0.039322	0.117499	0.115993
C1	0.013639	0.082042	0.543486	0.026618	0.189620	0.135899	0.069074
C2	0.094925	0.050328	0.655941	0.083295	0.086922	0.196939	0.080877
C3	-0.043196	0.001652	0.544670	0.121580	-0.015321	0.016680	0.068728
C4	0.221335	-0.052152	-0.676839	-0.059543	-0.155898	0.156621	0.208496
C5	0.290474	-0.156422	-0.558643	-0.060958	0.022000	0.053225	0.032735
E1	0.029280	-0.602118	0.022212	-0.109039	-0.090177	0.208863	0.102109
E2	0.243559	-0.677065	-0.118419	-0.116520	-0.060866	0.158377	0.032111
E3	0.002895	0.531109	0.086094	0.240073	0.240243	0.086717	0.256652
E4	-0.117159	0.685010	0.111956	0.275199	-0.097414	0.075563	0.063136
E5	-0.010044	0.415599	0.320663	0.124290	0.190452	-0.189109	0.298438
N1	0.756845	0.008989	-0.039717	-0.177801	-0.092304	-0.264009	0.163912
N2	0.743804	-0.045746	-0.021074	-0.151321	-0.010336	-0.320656	0.149091
N3	0.754130	-0.017749	-0.063656	-0.040556	-0.005759	0.032397	-0.046718
N4	0.616765	-0.285632	-0.184860	-0.016444	0.067471	0.160197	-0.060434
N5	0.569589	-0.095056	-0.038570	0.077538	-0.157337	0.141978	-0.103891
O1	-0.027556	0.180903	0.111509	0.058009	0.472142	0.089002	0.296191
O2	0.152677	-0.006841	-0.103057	0.073979	-0.503982	0.043161	0.098521
O3	0.024468	0.318293	0.078641	0.112861	0.564411	0.110451	0.181449
O4	0.223223	-0.171973	-0.030362	0.144983	0.343191	0.124454	0.081531
O5	0.064661	-0.006609	-0.064025	-0.027070	-0.581624	0.088062	0.083604

Eigenvalues for 7 factors:

	Eigenvalues
Factor1	5.134580
Factor2	2.753375
Factor3	2.148142
Factor4	1.852506
Factor5	1.548463
Factor6	1.110662
Factor7	0.980677

Based on the loading table, it appears that Factor 6 and Factor 7 do not have high loadings on any attribute, and Factor 7 does not have any loading coefficient exceeding 0.5. Therefore, considering removing these two factors and conducting factor analysis again with the remaining factors could be warranted.

Request 03: Students look at the loadings table explain the significant of each factor versus each property. If there are factor(s) that has no "high loading" value, you can remove these and perform Factor Analysis again with the remain factor. Otherwise, explain the Factor Variance Table

Loadings for 5 factors:

	Factor1	Factor2	Factor3	Factor4	Factor5
rownames	-0.019246	-0.061703	0.034865	-0.018678	0.014345
A1	0.106455	0.030705	0.027567	-0.429612	-0.074133
A2	0.028693	0.227127	0.141159	0.621801	0.064117
A3	0.005335	0.329927	0.115928	0.642856	0.061647
A4	-0.069443	0.209217	0.235418	0.431241	-0.108647
A5	-0.129002	0.397690	0.097901	0.528159	0.075363
C1	0.009548	0.048297	0.548276	0.041194	0.212169
C2	0.089975	0.014459	0.650304	0.105693	0.116582
C3	-0.030243	0.010825	0.555781	0.114165	-0.004812
C4	0.240794	-0.039326	-0.633576	-0.041462	-0.109695
C5	0.294330	-0.154179	-0.566991	-0.047858	0.031264
E1	0.057066	-0.576460	0.017182	-0.089660	-0.073218
E2	0.261661	-0.671561	-0.121490	-0.096788	-0.059680
E3	0.012417	0.533056	0.097624	0.245741	0.293883
E4	-0.131026	0.648750	0.120495	0.290069	-0.056985
E5	0.024327	0.492534	0.326639	0.079579	0.226538
N1	0.782726	0.097680	-0.042388	-0.225060	-0.083471
N2	0.752420	0.044935	-0.029547	-0.200539	-0.010945
N3	0.732571	-0.041066	-0.068349	-0.031849	-0.007284
N4	0.599635	-0.326102	-0.188999	0.010191	0.064659
N5	0.542043	-0.139326	-0.042065	0.100229	-0.155653
O1	-0.006094	0.196463	0.120863	0.059573	0.510642
O2	0.174735	0.026762	-0.099129	0.077368	-0.469638
O3	0.021106	0.296243	0.084710	0.122369	0.603572
O4	0.227826	-0.189289	-0.029353	0.160787	0.362255
O5	0.083619	0.011590	-0.061357	-0.013974	-0.533401

Eigenvalues for 5 factors:

	Eigenvalues
Factor1	5.134580
Factor2	2.753375
Factor3	2.148142
Factor4	1.852506
Factor5	1.548463

```
Variance of each factor for 5 factors:
         Variance Explained  Proportion of Variance  Cumulative Proportion
Factor1            2.736109                0.105235               0.105235
Factor2            2.428049                0.093387               0.198621
Factor3            2.082504                0.080096               0.278718
Factor4            1.800505                0.069250               0.347968
Factor5            1.549502                0.059596               0.407564
```

The Factor Variance Table provides an overview of each factor's ability to explain the variance of the data. Below are explanations for each column in the table:

1. **Variance Explained**: This is the total variance explained by each factor. It indicates the extent to which the factor accounts for variability in the data. Higher values suggest greater importance of the factor in explaining variation.
2. **Proportion of Variance**: This is the percentage of the total dataset variance explained by each factor. It reflects the importance of each factor relative to the total variance of the data. Higher values indicate greater contribution of the factor to data variation.
3. **Cumulative Proportion**: This is the cumulative percentage of variance explained by the preceding factors, up to the current factor. It indicates the total benefit gained when adding a new factor to the model. Higher values indicate increasing cumulative explanatory power of the factors.

VI. Research questions

1. What are factors in Factor Analysis, and why are they important?

- In Factor Analysis, factors represent underlying latent variables or constructs that cannot be directly observed but can be inferred from observed variables. These latent factors capture the common variance among observed variables and help simplify the data structure by reducing the dimensionality. Factors are important because they provide a way to uncover the underlying structure in the data and identify meaningful patterns or relationships among variables.

2. Explain the significance of eigenvalues and eigenvectors in Factor Analysis.

Eigenvalues in Factor Analysis:

Eigenvalues represent the amount of variance explained by each factor in Factor Analysis. Higher eigenvalues indicate that the corresponding factor explains more variance in the data. Eigenvalues are crucial in determining the significance of each factor and deciding how many factors to retain in the analysis. For instance, the Kaiser criterion suggests retaining only factors with eigenvalues greater than 1.

Eigenvectors in Factor Analysis:

Eigenvectors represent the direction or pattern of the factor loading for each observed variable. They provide information about how strongly each variable contributes to each factor. By examining eigenvectors, analysts can interpret the structure of factors and identify which variables are most strongly associated with each factor. This helps in understanding the underlying constructs represented by the factors and aids in the interpretation of the Factor Analysis results.

In summary, eigenvalues are used to determine the significance of factors, while eigenvectors help interpret the structure of factors and identify the variables contributing most strongly to each factor.

3. Compare Factor Analysis Vs. Principle Component Analysis.

- Factor Analysis (FA) and Principal Component Analysis (PCA) are both dimensionality reduction techniques, but they have different underlying assumptions and objectives.
    - FA assumes that observed variables are influenced by a smaller number of underlying latent factors, and it seeks to uncover these factors. It focuses on explaining covariance among observed variables.
    - PCA, on the other hand, seeks to capture the maximum variance in the data by transforming the observed variables into a new set of uncorrelated variables called principal components. It does not make assumptions about the underlying structure of the data and does not differentiate between common and unique variance.
- In summary, FA is more suitable when there is a theoretical interest in uncovering latent factors that explain covariance among observed variables, while PCA is more suitable for data reduction when the focus is on capturing maximum variance without regard to the underlying structure.

Feature	Factor Analysis (FA)	Principal Component Analysis (PCA)
Objective	Uncover latent factors explaining covariance	Capture maximum variance
Underlying Assumptions	Assumes latent factors	No assumptions about underlying structure
Focus	Explains covariance among observed variables	Captures variance without regard to structure
Components/Factors	Represent underlying latent variables	Linear combinations of observed variables
Error Terms	Considers unexplained variance (error terms)	Does not differentiate between common and unique variance
Suitability	When there's a theoretical interest in latent factors	For data reduction focused on maximum variance

4. Provide examples of real-world applications where Factor Analysis can be useful.

1. **Customer Satisfaction Surveys:** In the hospitality industry, Factor Analysis can be applied to customer satisfaction survey data to identify underlying factors that contribute to overall satisfaction, such as service quality, cleanliness, amenities, and location. This helps hotel managers focus their resources on areas that have the greatest impact on customer satisfaction.
2. **Educational Assessment:** Factor Analysis is used in educational research to analyze test scores and identify underlying factors that contribute to academic performance, such as reading comprehension, mathematical ability, and critical thinking skills. This information can help educators tailor teaching methods and interventions to address specific areas of weakness in students.
3. **Human Resources:** In organizational psychology, Factor Analysis can be applied to employee performance evaluations to identify underlying factors that contribute to job performance,

such as technical skills, interpersonal communication, and problem-solving abilities. This helps HR professionals develop training programs and performance appraisal systems that are aligned with organizational goals.

4. **Product Development:** In the manufacturing industry, Factor Analysis can be used to analyze customer feedback and identify underlying product attributes that drive consumer preferences, such as durability, ease of use, and aesthetic appeal. This information can inform product design decisions and help companies develop new products that better meet the needs of their target market.
5. **Health Sciences:** Factor Analysis is used in medical research to identify underlying factors contributing to health outcomes or disease risk factors, such as lifestyle factors, genetic predispositions, or environmental exposures. For example, researchers may use Factor Analysis to analyze data from large-scale population studies to identify clusters of risk factors associated with chronic diseases like heart disease, diabetes, or cancer. This information can help healthcare professionals develop targeted prevention and intervention strategies to reduce disease burden and improve public health outcomes.