Multivariate Data Analysis project

Students' info:

Name	Code
Haneen Ahmed Elgabry	**5200199**
Nour Mohamed Eletreby	**5200917**
Nada Hussien Mohamed	**5200699**
Mareez maher maawad	**5200524**

Submitted to: Dr. Sarah Osama

Title: Analyzing Character Strengths and Mental Health during Covid-19 Lockdown

Introduction:

The Covid-19 pandemic has not only posed a threat to physical health but has also led to significant psychological consequences due to prolonged periods of lockdown and self-isolation. This study aims to investigate the protective role of character strengths in sustaining mental health and self-efficacy during the lockdown period. The data for this analysis were collected from 944 Italian respondents, through an online survey conducted one month after the initiation of the lockdown.

Methodology:

1. Random Sampling:

- A random sample of 850 observations will be selected from the dataset to ensure a representative subset for analysis.

2. Descriptive Analysis:

- Before diving into more advanced analyses, a descriptive analysis will be conducted on the selected random sample. This will involve calculating summary statistics, such as mean, median, standard deviation, and percentiles, for key variables including character strengths, psychological distress measures (DASS21 and GHQ12), and Covid-19-related self-efficacy (SEC). Descriptive statistics will offer an initial understanding of the distribution and central tendencies of the variables.

3. Factor Analysis:

- Factor analysis will be performed on the selected random sample to identify latent factors within the character strengths data. This technique will help reveal the underlying structure and relationships among the observed variables. The results will be compared with the factors extracted in the original study.

4. Cluster Analysis:

- Utilizing the four strengths factors (transcendence, interpersonal, openness, and restraint), cluster analysis will be conducted on the entire dataset to identify distinct groups or clusters of respondents based on their character strengths. Descriptive statistics for each cluster will be examined to understand the characteristics of each group.

5. Multivariate Regression:

- Multivariate regression models will be employed to analyze the relationship between character strengths (transcendence, interpersonal, openness, restraint) and the three dependent measures: DASS21, GHQ12, and SEC (Self-efficacy for Covid-19). These regression models will help assess the impact of character strengths on psychological distress and self-efficacy during the pandemic.

First: random sampling

To create a representative subset for analysis, a random sample of 850 observations was drawn from the original dataset. The process involved setting a seed value to ensure the reproducibility of the results. By employing a random sampling technique without replacement, 850 observations were selected from the dataset. This subset, referred to as 'random_sample,' serves as a representative portion of the data, allowing for more efficient and focused analyses. The sampled data was then available for further exploration and examination.

Second: descriptive analysis:

We have in our dataset 41 variables defined as follow:

Numeric variables:

Participants: participant number

Four columns for the second-order strengths factors previously extracted via PCA:

Openness

Restraint

Transcendence

Interpersonal

The three dependent measures

DASS_21 (Depression Anxiety and Stress Scale)

GHQ_12 (General Health Questionnaire)

SEC (Self-efficacy for Covid-19)

demographic variables:

Age

Work (representing the perceived work change after lockdown)

Day (how many days passed when the participant responded since the day the survey was opened)

Character strength variables:

24-character strength variables which are: Appreciation_of_beauty Bravery Creativity Curiosity Fairness Forgiveness Gratitude Honesty Hope Humilty Humor Judgment Kindness Leadership Love Love_of_learning Perseverance Perspective Prudence Self_regulation Social_intelligence Spirituality Teamwork Zest

Categorical variables:

Gender

Student (being a student or not)

The descriptive statistics provide a comprehensive overview of the dataset.

1) Numeric variables.

When running the describe code we found that, the positively skewed distribution of age, and the variations in work experience and family size. Additionally, psychological trait variables exhibit slightly negatively skewed distributions, while mental health scores demonstrate right-skewed distributions.

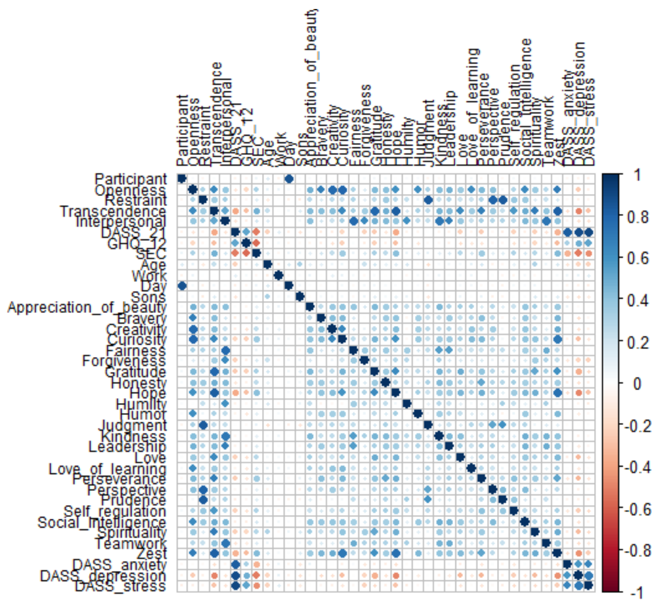

Figure 1: correlation plot

The plot is a correlation matrix represented as a heatmap. It’s a graphical representation of the correlation coefficients between different psychological and behavioral traits.

The cells are color-coded based on the correlation coefficient values, which range from -1 to 1. Positive correlations are indicated in blue shades, while negative correlations are in red shades. The dark blue diagonal line running from the top left corner to the bottom right corner represents a perfect positive correlation of each trait with itself, which is expected in a correlation matrix.

2) Categorical variables:

We have 2 categorical variables in the dataset which are gender and student, so we will make them a contingency table as follow :

Gender	**Student**		**Sum**
	Other	**Student**	
Female	505	128	633
Male	169	47	216
Sum	674	175	849

- The ‘Female’ category has a total of 633 students, with 505 in the ‘Gender’ column and 128 in the ‘Other’ column.

- The ‘Male’ category has a total of 216 students, with 169 in the ‘Gender’ column and 47 in the ‘Other’ column.

Third: Factor analysis

To achieve our objective to study interdependence among all the quantitative variables through finding indices, or we can call the factors, expressing this interdependence, we must use Factor Analysis.

Based on the data story and the focus on character strengths, psychological distress, and Covid-19-related self-efficacy, it seems appropriate to include the character strength variables in the principal component analysis (PCA). These include: (Appreciation_of_beauty , Bravery , Creativity , Curiosity , Fairness , Forgiveness

, Gratitude, Honesty , Hope ,Humility , Humor , Judgment , Kindness , Leadership , Love , Love_of_learning

, Perseverance, Perspective , Prudence , Self_regulation , Social_intelligence , Spirituality , Teamwork , Zest)

So we excluded specific variables (e.g., 'Participant,' 'Openness,' 'Restraint,' DASS_21, GHQ_12, SEC, Age, Gender, Work, Student, Day, Sons, DASS_anxiety, DASS_depression, DASS_stress) as they are not relevant to the focus on character strengths, psychological distress, and Covid-19-related self-efficacy.

(1) Before conducting factor analysis, we need to prepare the data by scaling the variables to have a mean of zero and a standard deviation of one. This is important because factor analysis is sensitive to differences in scale between variables.
(2) The next step is to determine the number of factors to extract from the data. This can be done using a variety of methods, we’ll use scree plot method:

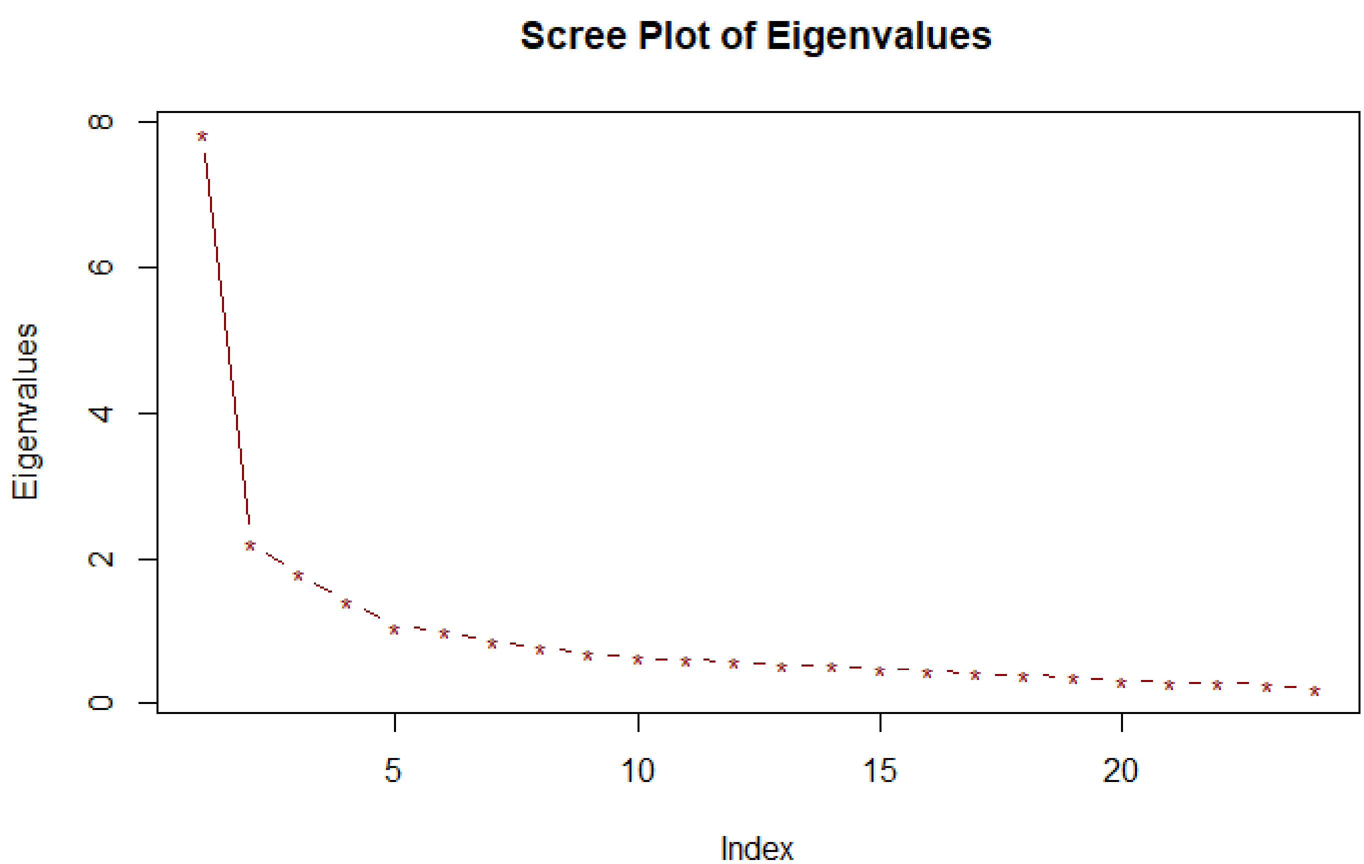

Figure 2:scree plot

We observe from figure (1) that the factor analysis is suitable in our data since it has elbow shape and we find that it has 2 elbows, corresponding to the second and fifth eigenvalues, so we will either retain 1 or 4 factors. To support our decision, we will look on the cumulative proportions of variance explained by both factors, the first 1 has only a cumulative proportion of about 33% while the fourth has a cumulative proportion of about 55% which is more satisfying. So, our decision is to retain 4 factors.

```
[1] 7.8385108 [2]2.2119184 [3]1.7984444 [4]1.4213153 [5]1.0620349 [6]1.0087484 [7]0.87214
24 [8]0.7648405 [9]0.6963540 [10] 0.6350515 [11]0.6105471 [12]0.5834624 [13]0.5420792
[14]0.5282837 [15]0.4700058 [16]0.4588001 [17]0.4124095 [18]0.3916712 [19]0.3576649
[20]0.3192546[21] 0.2831635 [22] 0.2788286 [23]0.2676669 [24]0.1868020
```

These are the eigenvalues of all 24 factors which represent the amount of variance that can be explained by the corresponding factor. If the retain is depending on kaiser methods it would suggest retaining 6 factors, since they are the factors having an eigenvalue greater than 1. However, we will depend on more reliable approach, the scree plot.

(3) Extracting the factors and estimating their loadings on the variables: We have two methods for extracting the factors and estimating their loadings which are: (Principle Factor Method - Principle Component Method) We are going to compare the two methods, if the two methods lead to the same results we will choose any of them and if there is a difference between them, we will choose the more logical. In the following, there are two tables about the results of the two methods.

Table (1): Principle Factor Method

Rotation	**No rotation**			
variable	**PA1**	**PA2**	**PA3**	**PA4**
Appreciation_of_beauty	0.57	0.05	-0.05	0.13
Bravery	0.51	-0.29	0.18	0.19
Creativity	0.56	-0.30	0.25	0.21
Curiosity	0.71	-0.38	0.12	0.01
Fairness	0.55	0.35	-0.28	0.23
Forgiveness	0.36	0.17	-0.29	-0.18
Gratitude	0.69	-0.04	-0.17	-0.28
Honesty	0.59	0.17	0.07	-0.01
Hope	0.74	-0.26	0.00	-0.26
Humilty	0.26	0.47	-0.19	-0.10
Humor	0.42	-0.25	-0.09	0.27
Judgment	0.42	0.36	0.52	0.09
Kindness	0.63	0.17	-0.35	0.20
Leadership	0.61	0.14	-0.14	0.27
Love	0.54	-0.02	-0.12	-0.08
Love_of_learning	0.44	-0.12	0.17	0.04
Perseverance	0.59	-0.03	0.17	-0.27
Perspective	0.50	0.20	0.36	0.14
Prudence	0.32	0.61	0.38	-0.15
Self_regulation	0.45	0.15	0.20	-0.27
Social_intelligence	0.64	0.01	0.02	0.22
Spirituality	0.55	-0.09	-0.16	-0.27
Teamwork	0.51	0.26	-0.31	0.06
Zest	0.77	-0.39	-0.07	-0.16

Table (2): Principle Component Method				
Rotation		**No rotation**		
	PC1	**PC2**	**PC3**	**PC4**
Appreciation_of_beauty	0.59	0.04	-0.06	0.15
Bravery	0.53	-0.38	0.20	0.20
Creativity	0.58	-0.37	0.26	0.19
Curiosity	0.71	-0.41	0.09	-0.04
Fairness	0.57	0.40	-0.29	0.30
Forgiveness	0.38	0.25	-0.38	-0.27
Gratitude	0.71	-0.04	-0.19	-0.31
Honesty	0.61	0.18	0.11	-0.01
Hope	0.74	-0.27	-0.02	-0.28
Humilty	0.28	0.61	-0.21	-0.09
Humor	0.45	-0.33	-0.13	0.39
Judgment	0.43	0.35	0.62	0.13
Kindness	0.65	0.18	-0.37	0.26
Leadership	0.63	0.15	-0.14	0.35
Love	0.57	-0.03	-0.16	-0.10
Love_of_learning	0.47	-0.17	0.22	0.01
Perseverance	0.61	-0.04	0.19	-0.34
Perspective	0.52	0.19	0.46	0.19
Prudence	0.33	0.65	0.47	-0.12
Self_regulation	0.48	0.19	0.28	-0.40
Social_intelligence	0.67	-0.02	0.03	0.28
Spirituality	0.57	-0.09	-0.22	-0.39
Teamwork	0.54	0.32	-0.35	0.12
Zest	0.77	-0.39	-0.10	-0.18

We observe from table (1) and table (2), there are differences between the results of the two methods and we can see also that the factors in the second method “Principal Component Method” is better than those in the first method “Principal Factor Method” because they have more loadings > 0.6 i.e. they capture more variables. And we can also see that in table (1) that most factor 1 loads on most of the variables which is insufficient, So we will depend on Principle Component Method.

Principle component model in details:

Table (3)	PC1	PC2	PC3	PC4
Eigen values	7.84	2.21	1.80	1.42
Proportion Var	0.33	0.09	0.07	0.06
Cumulative Var	0.33	0.42	0.49	0.55

After deciding to retain 4 factors we got these results. Our four-retained factors are capturing variances of 7.84, 2.21, 1.8 and 1.42 respectively. Factor 1 is capturing about 33% of the variability in data, then each of the following factors is adding a descending percentage to this proportion.

Table (4): Principal component method with no rotation

	PC1	PC2	PC3	PC4	Uniqueness
Appreciation_of_beauty	0.59	0.04	-0.06	0.15	0.622
Bravery	0.53	-0.38	0.20	0.20	0.493
Creativity	0.58	-0.37	0.26	0.19	0.418
Curiosity	0.71	-0.41	0.09	-0.04	0.313
Fairness	0.57	0.40	-0.29	0.30	0.353
Forgiveness	0.38	0.25	-0.38	-0.27	0.574
Gratitude	0.71	-0.04	-0.19	-0.31	0.366
Honesty	0.61	0.18	0.11	-0.01	0.580
Hope	0.74	-0.27	-0.02	-0.28	0.292
Humilty	0.28	0.61	-0.21	-0.09	0.496
Humor	0.45	-0.33	-0.13	0.39	0.521
Judgment	0.43	0.35	0.62	0.13	0.296
Kindness	0.65	0.18	-0.37	0.26	0.340
Leadership	0.63	0.15	-0.14	0.35	0.437
Love	0.57	-0.03	-0.16	-0.10	0.639
Love_of_learning	0.47	-0.17	0.22	0.01	0.700
Perseverance	0.61	-0.04	0.19	-0.34	0.471
Perspective	0.52	0.19	0.46	0.19	0.442
Prudence	0.33	0.65	0.47	-0.12	0.243
Self_regulation	0.48	0.19	0.28	-0.40	0.499
Social_intelligence	0.67	-0.02	0.03	0.28	0.476
Spirituality	0.57	-0.09	-0.22	-0.39	0.467
Teamwork	0.54	0.32	-0.35	0.12	0.471
Zest	0.77	-0.39	-0.10	-0.18	0.219

Table (4) represented the factor loadings in each of the original variables.in other words do this variable strongly or weakly affect this factor? It also represents the uniqueness which is a part of the variance that is unique only for this variable.

Interpretation of uniqueness: 62.2% of the variation of appreciation of beauty comes from the unique factor so this variable share with the common factor with 38%.

Interpretation of factor loading: From table (3) we notice that factor 1 loads most on these variables: zest, preservence, honesty, Hope, Gratitude, curiosity, social intelligence, Kindness, and leadership. The 2nd factor loads only on humility and prudence and the 3rd factor only on judgment and the 4th factor have very weak loadings on almost all variables, which makes them inefficient and give no clear interpretation So we can apply a rotation method on our retained factors.

(4) Rotation of factor loading:

Table 5: Correlations among factors				
	TC1	TC2	TC3	TC4
TC1	1.00	0.43	0.22	0.23
TC2	0.43	1.00	0.23	0.08
TC3	0.22	0.23	1.00	0.02
TC4	0.23	0.08	0.02	1.00

We can observe from table (5) that the correlation coefficients among the three factors are greater than 0.2 for some factors which means that correlations among them are strong. Accordingly, we are going to depend on the rotated factors by obliman rotation method in the principal component method.

Table 6: oblimian rotation

	TC1	TC2	TC3	TC4	Uniqueness	Communalities
Appreciation_of_beauty	0.20	0.41	0.11	0.16	0.622	0.378
Bravery	0.28	0.08	0.08	0.55	0.493	0.507
Creativity	0.31	0.05	0.15	0.58	0.418	0.582
Curiosity	0.62	0.03	0.02	0.40	0.313	0.687
Fairness	-0.08	0.81	0.10	-0.06	0.353	0.647
Forgiveness	0.38	0.33	-0.10	-0.42	0.574	0.426
Gratitude	0.72	0.19	-0.01	-0.13	0.366	0.634
Honesty	0.28	0.26	0.34	0.03	0.580	0.420
Hope	0.81	0.01	0.00	0.10	0.292	0.708
Humilty	0.00	0.46	0.22	-0.48	0.496	0.504
Humor	0.07	0.40	-0.19	0.50	0.521	0.479
Judgment	-0.05	0.05	0.81	0.19	0.296	0.704
Kindness	0.10	0.77	-0.06	0.03	0.340	0.660
Leadership	0.01	0.67	0.11	0.21	0.437	0.563
Love	0.45	0.26	-0.01	-0.01	0.639	0.361
Love_of_learning	0.33	-0.01	0.20	0.29	0.700	0.300
Perseverance	0.66	-0.11	0.28	-0.02	0.471	0.529
Perspective	0.02	0.15	0.62	0.30	0.442	0.558
Prudence	-0.01	0.06	0.83	-0.24	0.243	0.757
Self_regulation	0.53	-0.16	0.45	-0.19	0.499	0.501
Social_intelligence	0.15	0.46	0.16	0.34	0.476	0.524
Spirituality	0.73	0.07	-0.10	-0.19	0.467	0.533
Teamwork	0.10	0.68	0.00	-0.17	0.471	0.529
Zest	0.79	0.09	-0.12	0.22	0.219	0.781

From table 6 we can see that the variables: curiosity, gratitude, hope, perseverance, spirituality, and zest are the variables most affecting factor 1. While factor 2 loads on the variables: fairness, kindness, leadership, and teamwork. Factor 3 loads on judgment, perspective, and prudence variables. The fourth factor is a weak factor, it has no variables with high loadings. The most affecting variable for this factor is creativity with a loading factor equal to 0.58 which is moderate.

Naming the factors: We should take into consideration that oblimin rotation accepts correlation between factors which make it harder to name them.

Factor 1: transcendence

Variables: Curiosity, Gratitude, Hope, Perseverance, Spirituality, Zest

Interpretation: This factor seems to capture a positive and optimistic outlook on life, encompassing qualities such as curiosity, gratitude, hope, perseverance, spirituality, and zest.

Factor 2: interpersonal

Variables: Fairness, Kindness, Leadership, Teamwork

Interpretation: This factor may reflect a social and interpersonal engagement, including qualities related to fairness, kindness, leadership, and teamwork.

Factor 3: oppenness

Variables: Judgment, Perspective, Prudence

Interpretation: This factor appears to capture cognitive aspects related to judgment, perspective, and prudence, suggesting a thoughtful and analytical dimension.

Factor 4: restraint

Variables: Creativity

Interpretation: This factor appears to have a weak theme.

(5) Factor results comparison with the result of the study:

1. Our 1st factor is loading on Curiosity, Gratitude, Hope, Perseverance, Spirituality, Zest but researcher factor is loading on hope, curiosity, love, forgiveness, spirituality, zest, gratitude, perseverance, self-regulation.
2. Our 2nd factor is loading on fairness, kindness, teamwork, and leadership. But the researcher's 2nd factor is loading on fairness, kindness, humor, social intelligent, teamwork, leadership, forgiveness, humility, appreciation of beauty.
3. Our 3rd factor is loading on judgment, prudence, and perspectives. But the researcher's 3rd factor is loading on hope, perspective, zest, humility, love of learning, creativity, bravery, curiosity, humor, and social intelligence.
4. Our 4th factor is loading on creativity, but researcher 4th factor Is loading on self-regulation, judgment, perspective, honesty, prudence.

(6) Constructing factor scores to get the indices.

We observe from the previous steps that the factor loadings measure the contribution of the factors to all variables only, but we need to construct indices, so we must construct factor scores which measure contribution of each variable to the factor which represents the index.

We are going to describe the 4 indices by plotting their histograms.

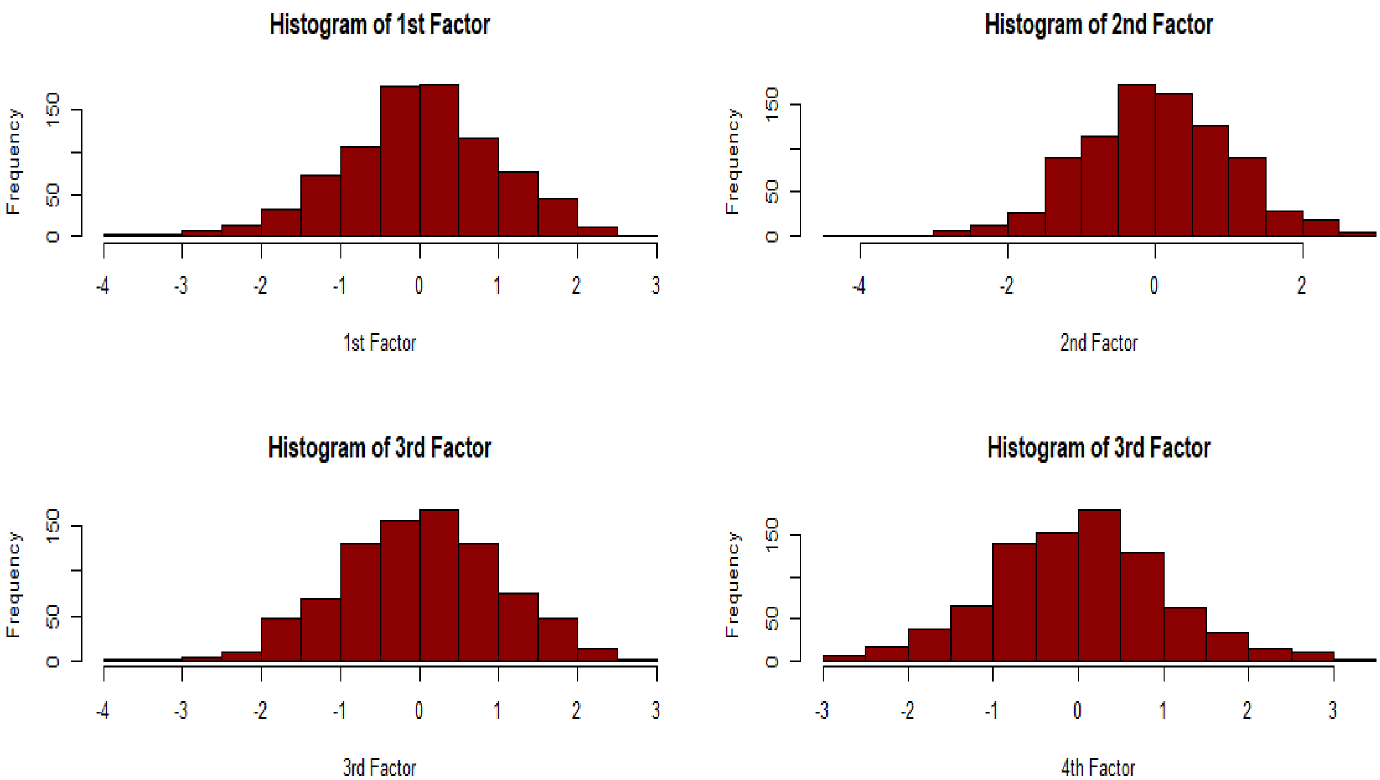

Figure 3: histogram of the 4 indices respectively

We observe from figure (2) that:

-the distribution of the transcendence factor/index is symmetric around the mean which means that the most observations have values around the mean of this index i.e. most observations have medium transcendence.

-the distribution of the interpersonal factor/index is symmetric around the mean which means that the most observations have values around the mean of this index i.e. most observations have medium interpersonal.

-the distribution of the openness factor/index is symmetric around the mean which means that the most observations have values around the mean of this index i.e. most observations have medium openness.

- the distribution of the restraint factor/index is symmetric around the mean which means that the most observations have values around the mean of this index i.e. most observations have medium restraint.

Fourth: cluster analysis

Steps for Cluster Analysis:

1. Variable Selection:

The analysis focuses on specific variables, including "transcendence," "interpersonal," "openness," and "restraint" factors that extracted using factor analysis.

2. Data Scaling:

Standardize or normalize the data to ensure that variables are on a similar scale. This is important because cluster analysis is sensitive to the scale of variables.

3. Cluster Algorithm, Determine the Number of Clusters:

We’ll use the kmeans function from the stats package to perform K-means clustering on the scaled data.

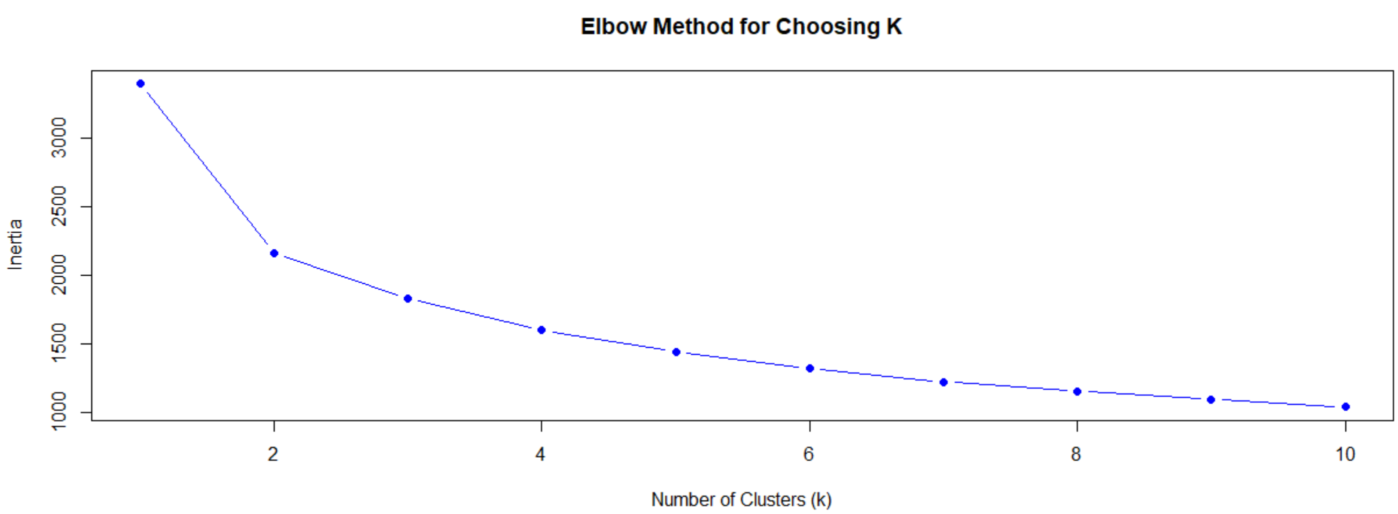

Figure 4: optimal k for cluster analysis

Plot (3) is generated with the x-axis representing the number of clusters (k) and the y-axis representing the inertia for each k.

The "Elbow Method" helps identify a point on the plot where adding more clusters ceases to significantly reduce inertia. in this case, we can consider 2 or 3 clusters.

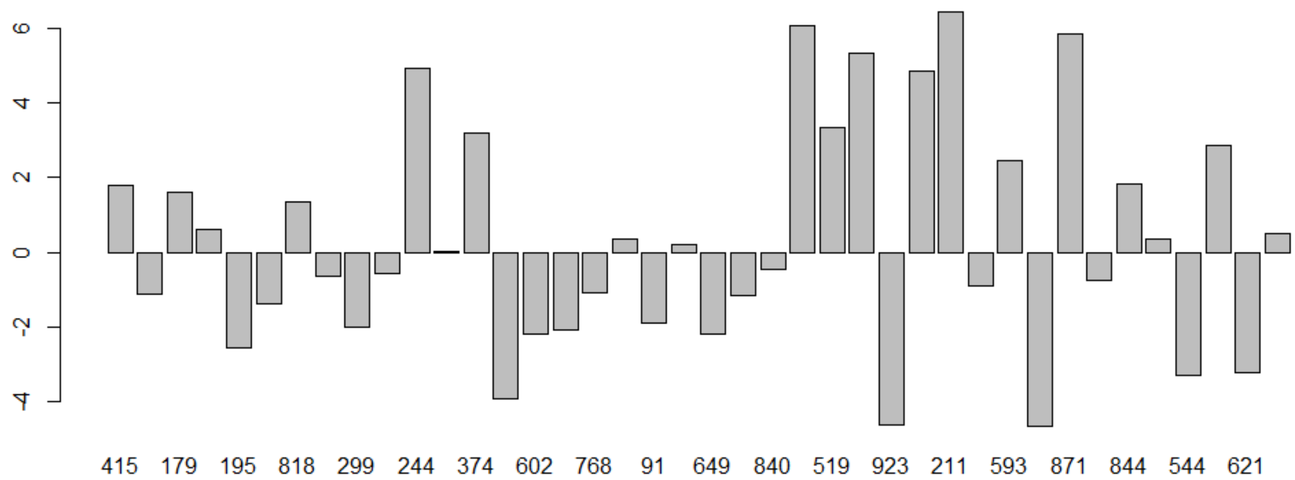

Figure 5: bar plot of the variable's vs first principal component

From figure 4, If we apply a classification to the previous bar plot, we could potentially classify the bars into three groups:

Large Positive values: Bars that extend a large distance upwards from the x-axis.

Small positive values: Bars that extend a short distance upwards from the x-axis.

negative values: Bars that extend downwards from the x-axis.

4. Cluster Analysis and interpretation:

Cluser	Within cluster sum of square
1	727.5758
2	793.1598
3	608.4321

Table 7: within cluster for 3 cluster

Betweenness = 1266.832
Total = 2129.168
Betweenness/total = 59.5%

Cluster	Within cluster sum of square
1	1047.871
2	1439.709

Table 8:within cluster for 2 cluster

Betweenness = 908.4198
Total = 2487.58
Betweenness/total = 36.5%

We tried both and found 3 clusters are more efficient as it has higher between-clusters variance and lower within-cluster variance (it maximizes the distance between clusters and minimizes it within the cluster). They are also more real since people's characteristics differ. If two people have the same characteristic, it differs in the level of strength between them.

Cluster:	1	2	3
Transcendence	0.8204762	-0.65870881	-0.2337048
Interpersonal	0.7990311	-0.50349474	-0.4061118
Openness	0.5840104	0.06899489	-0.8621348
Restraint	0.2669945	-0.78910763	0.6674614

Table 8: cluster means for 3 clusters.

The output of table 8 shows for each cluster and for each factor (Transcendence, Interpersonal, Openness, Restraint), the output provides the mean values. These mean values represent the central tendency of each factor within each cluster.

For example, in Cluster 1, the mean value for transcendence is 0.82, interpersonal is 0.80, openness is 0.58, and restraint is 0.27.

Interpretation:

Cluster 1 seems to have higher values for Transcendence and Interpersonal, moderate values for openness, and a lower value for restraint.

Cluster 2 has lower values for Transcendence and Interpersonal and openness, and a higher value for restraint.

Cluster 3 has lower values for openness and Restraint, a very low value for Transcendence, and a higher value for Interpersonal.

Based on the previous characteristics of each cluster we will name our clusters as follow:

Cluster 1: "Transcendent Connectors"

Cluster 2: "Restrained Observers"

Cluster 3: "Interpersonal Explorers"

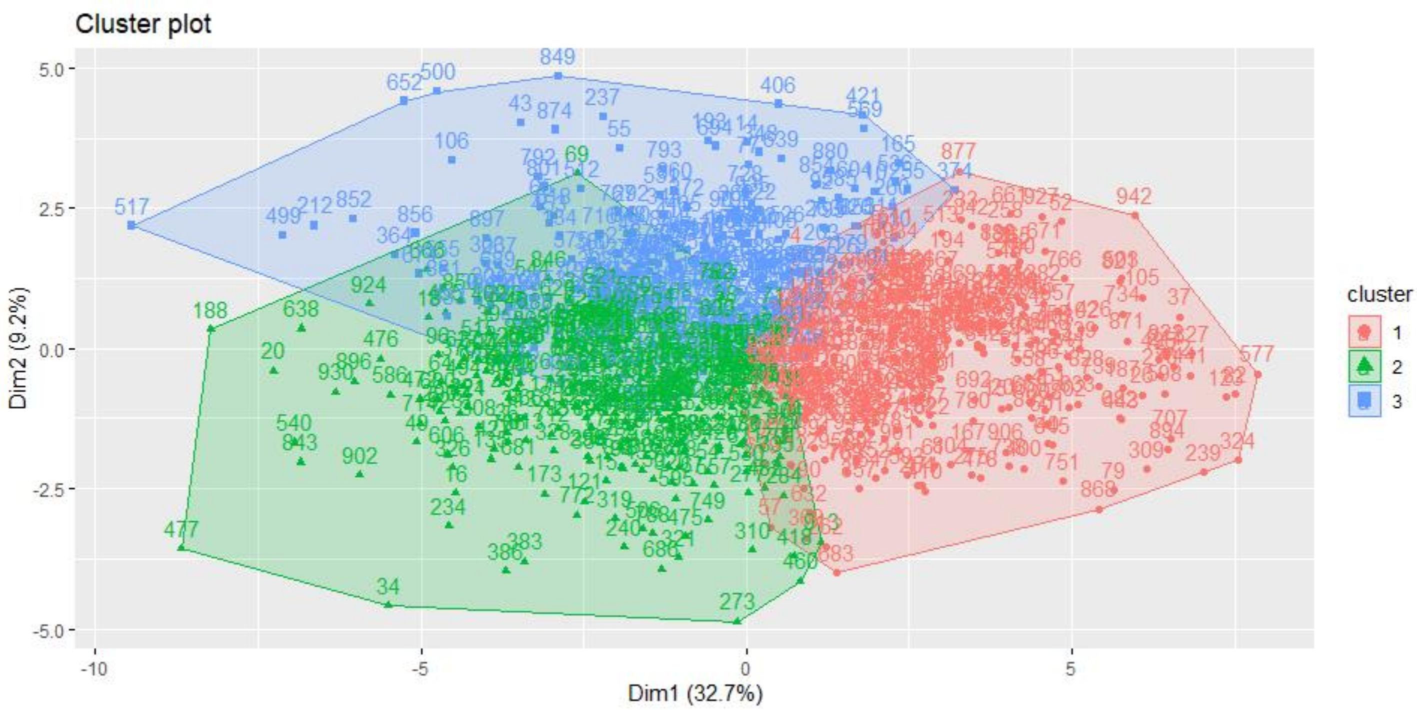

Figure 6: cluster plot

Figure 5 displays three distinct clusters of data points, each represented by different colors: blue for cluster 1, green for cluster 2, and red for cluster 3.

The x-axis is labeled “Dim1 (32.7%)” and the y-axis is labeled “Dim2 (9.2%)”, indicating the dimensions and the percentage of variance explained by each. the plot is a result of a dimensionality reduction technique Principal Component Analysis (PCA), where Dim1 and Dim2 are the first two principal components.

The clusters are well-separated, suggesting that the clustering algorithm has done a good job of grouping similar data points together.

Discriminant analysis:

Validation to our cluster analysis results. we tried to check three assumptions of discriminant nalysis which are first, normality assumption, covariance between groups are equal, different means between groups.

Firstly, normality:

For the first group :

```
$multivariateNormality
             Test        Statistic               p value Result
1 Mardia Skewness 191.771603747323 4.73861835946685e-30     NO
2 Mardia Kurtosis  7.3432522479014 2.08499884024604e-13     NO
3             MVN             <NA>                  <NA>     NO
```

We reject H_0 ,multivariate normality assumption violated.

For the second group:

```
$multivariateNormality
             Test        Statistic              p value Result
1 Mardia Skewness  93.0379155874069 2.17148146341283e-11     NO
2 Mardia Kurtosis 0.701719872288802    0.482853877486065    YES
3             MVN              <NA>                 <NA>     NO
```

We reject H_0 , multivariate normality assumption is violated.

For the third group:

```
$multivariateNormality
             Test        Statistic              p value Result
1 Mardia Skewness 121.313484341308 1.62784066292436e-16     NO
2 Mardia Kurtosis 3.99316110883744 6.51982368602422e-05     NO
3             MVN             <NA>                 <NA>     NO
```

We reject H_0 , multivariate normality assumption is violated.

But since we have large sample size by CLT we can say that we have MVN

Secondly, for equality of covariances among 3 groups.

$H_0:\Sigma_1= \Sigma_2= \Sigma_3$

```
       Box's M Test

Chi-Squared Value = 84.2024 , df = 20  and p-value: 7.51e-10
```

Reject H_0, covariances are not same for the 3 groups.

The assumption of equality of covariance violated.

For equality of means among 3 groups:

$H_0:\mu_1= \mu_2= \mu_3$

```
$HT2
         [,1]
[1,] 1258.646

$F
         [,1]
[1,] 313.2182

$df
[1]   4 651

$p.value
             [,1]
[1,] 4.278766e-150
```

p-value<0.05

reject H_0,we have enough evidence to say that means of the three groups are not equal.

Based on these findings, it can be concluded that **quadratic** discriminant analysis is more appropriate since the assumption of equal covariance is not met

We split data into train and test and try to do classification table.

actual / predicted	1	2	3
1	77	0	1
2	2	86	2
3	1	2	83

We can say that probability of correctly classifying an observation in cluster 1 is 98.717%

We can say that probability of correctly classifying an observation in cluster 2 is 95.55%

We can say that probability of correctly classifying an observation in cluster 3 is 96.5%

The overall percentage of correct classification is 96.85% which indicates that cluster analysis have high ability of correct classification.

Conclusion:

From factor analysis we decide to retain

For multivariate regression:

For DASS_21:

Coefficients	Estimate	Std.Error	T value	p-value
Intercept	29.90601	3.7981	7.874	1.06e-14*
Age	-0.06171	0.02852	-2.164	0.030754*
Gender	2.852522	0.74825	3.776	0.000171*
Work	0.72935	0.21568	3.382	0.000754*
Student	0.60739	0.94232	0.645	0.519
Sons	-1.243	0.72218	-1.721	0.085588
Openness	0.10007	0.03133	3.194	0.0014*
Restraint	0.06041	0.04828	1.251	0.211155
Transcendence	-0.2836	0.02666	-10.637	2e-16*
Interpersonal	0.03416	0.03177	1.075	0.282586

Adjusted R^2 =0.1999

βo as all explanatory variables are equal zero the DASS_21 (Depression Anxiety and Stress Scale) will be 29.90601 on average.

β_1:as age increase by one unit the DASS_21 (Depression Anxiety and Stress Scale)will decrease by 0.06171 on average holding other variables constant.

β_2:for females the DASS_21 (Depression Anxiety and Stress Scale) is more than the males by 2.82522 on average holding other variables constant.

β_3:as work increase by one unit the DASS_21 (Depression Anxiety and Stress Scale)will increase by 0.72935 on average holding other variables constant.

β_6:as openness increase by one unit the DASS_21 (Depression Anxiety and Stress Scale)will increase by 0.10007 on average holding other variables constant.

β_8:as transcendence increase by one unit the DASS_21 (Depression Anxiety and Stress Scale)will decrease by 0.28360 on average holding other variables constant.

From R^2 we can say that 20% of variation in DASS_21 (Depression Anxiety and Stress Scale) explained by these explanatory variables. Which is very low percentage.

GHQ_12:

Coefficients	Estimate	Std.Error	T value	p-value
Intercept	22.07622	1.84797	11.946	2e-16*
Age	0.02371	0.01388	1.709	0.08798
Gender	0.82384	0.36406	2.263	0.0239
Work	0.17768	0.10494	1.693	0.09088
Student	0.65542	0.45848	1.43	0.1532
Sons	0.14399	0.35138	0.41	0.6821
Openness	-0.01122	0.01524	-0.736	0.4619
Restraint	0.04856	0.02349	2.067	0.039*
Transcendence	-0.09996	0.01297	-7.706	3.66e-14*
Interpersonal	0.03125	0.01546	2.022	0.0435*

Adjusted R^2:0.1044

β_0:as all explanatory variables are zero the GHQ_12 (General Health Questionnaire) will be 22.07622on average.

β_2: for females the GHQ_12 (General Health Questionnaire) is more than the males by 0.82384 on average holding other variables constant.

β_7:as restraint increase by one unit the GHQ_12 (General Health Questionnaire) will increase by 0.04856 on average holding other variables constant.

β_8:as transcendence increase by one unit the GHQ_12 (General Health Questionnaire) will increase by 0.110807 on average holding other variables constant.

β_9:as interpersonal increase by one unit the GHQ_12 (General Health Questionnaire) will decrease by 0.023424 on average holding other variables constant.

From R^2 we can say that 10.44% of variation in GHQ_12 (General Health Questionnaire) explained by these explanatory variables. Which is very low percentage.

For SEC:

Coefficients	Estimate	Std.Error	T value	p-value
Intercept	3.808815	1.416404	2.689	0.00731*
Age	0.013814	0.010635	1.299	0.19433
Gender	-0.886683	0.27904	-3.178	0.00154*
Work	-0.126116	0.080433	-1.568	0.11726
Student	-0.238537	0.3514	-0.679	0.49746
Sons	-0.051236	0.269320	-0.19	0.84917

Openness	0.008959	0.011684	0.767	0.44343
Restraint	-0.003411	0.018004	-0.189	0.84980
Transcendence	0.110807	0.009942	11.145	2e-16*
Interpersonal	-0.023424	0.011848	-1.977	0.04837*

Adjusted R^2:0.2387

β_0:as all explanatory variables are zero the SEC (Self-efficacy for Covid-19) will be 3.808815 on average.

β_2: for females the SEC (Self-efficacy for Covid-19) is less than the males by 0.886683 on average holding other variables constant.

β_8:as transcendence increase by one unit the SEC (Self-efficacy for Covid-19) will increase by 0.110807 on average holding other variables constant.

β_9:as interpersonal increase by one unit SEC (Self-efficacy for Covid-19) will decrease by 0.023424 on average holding other variables constant.

From R^2 we can say that 23.87% of variation in GHQ_12 (General Health Questionnaire) explained by these explanatory variables. Which is very low percentage.

Comparing to results of study:

Transcendence are always significant for the 3 dependent variables.

Regressing DASS_21,GHQ_12,SEC on 24 individual character strengths:

	$\hat{\beta}_{DASS21}$	$\hat{\beta}_{GHQ12}$	$\hat{\beta}_{SEC}$
Age	-0.063446*	0.025579	0.01429
Gender	1.818198*	0.484284	-0.77745
Student	1.094040	0.743936	-0.16991
Day of survey	-0.078865	-0.025944	0.03675
Work change	0.616777*	0.146818	-0.10943
Having a child	-0.681278	0.325302	-0.16414
Appreciation of beauty	0.506525*	0.106563	-0.09869
Bravery	0.027294	0.066038	-0.05431
Creativity	-0.009117	-0.037182	0.04904
Curiosity	0.049563	-0.049908	0.06523
fairness	0.231447	0.195580*	-0.04161
forgiveness	-0.324650*	-0.038051	0.02063
Gratitude	0.021320	-0.106011	0.02650
Honesty	-0.220056	-0.034020	0.15234*
Hope	-0.694194*	-0.118030	0.14138*
Humor	0.138125	-0.058501	-0.03479
Judgement	0.141732	0.042484	-0.04336
Kindness	0.245439	0.208385*	-0.12964
Leadership	0.175385	0.001635	-0.01656
Love	-0.392133*	-0.083336	0.13689*
Love of learning	0.091786	-0.069807	0.04174
Perseverance	0.078468	0.115499	0.07875
Perspective	0.169529	0.101585	-0.06631

prudence	-0.287432*	-0.081536	0.08808
Self-regulation	-0.202212	-0.084076	0.10206*
Social intelligence	0.180847	0.031511	-0.01754
Spirituality	0.014073	-0.089469	0.02069
Teamwork	-0.163200	-0.096662	-0.01670
Zest	-0.674444*	-0.291106*	0.25212*
R^2	0.2557	0.1321	0.2621

From results work,gender,age,appreciation of beauty,forgiveness,hope,love prudence and zest have a significant effect on DASS_21 (Depression Anxiety and Stress Scale) . fairness,kindness and zest have a significant effect on GHQ_12 (General Health Questionnaire). Honesty,hope,love,self-regulation and zest have a significant effect on SEC (Self-efficacy for Covid-19).

Comparing with the results of study appreciation of beauty,forgiveness,hope,love,prudence and zest have significant effect on DASS_21 (Depression Anxiety and Stress Scale). Zest is the only one that have a significant effect on GHQ_12 (General Health Questionnaire). Gender,love and zest have a significant effect on SEC (Self-efficacy for Covid-19).

Conclusion:

Firstly we perform factor analysis and we retain 5 factors compared to the study which retain only 4 factors,

Secondly we perform cluster analysis and we classify our data to 2 clusters and 3 clusters and see that 3 clusters are better than 2 so we do discriminant analysis to validate the results of cluster analysis and we checked assumptions normality violated,equality of covariancies violated finally we have not w equal means for the 3 groups. We reached the step of classification table and see that the each percentage of correct classification is greater than 70% and overall percentage of correct classification is above 90% which indicates high ability of our cluster analysis to classify the observations finally we do multivariate regression analysis on fewer variables that the study use them see the effect of 3 dependent variables depression,mental health,efficacy on the explanatory variables.

appendix:

```
> anova(mv_reg2)
Analysis of Variance Table

                   Df  Pillai approx F num Df den Df   Pr(>F)    
(Intercept)         1 0.98537  18796.9      3    837 < 2e-16 ***
Age                 1 0.08750     26.8      3    837 < 2e-16 ***
factor(Gender)      1 0.00597      1.7      3    837 0.17093    
Work                1 0.01275      3.6      3    837 0.01319 *  
factor(Student)     1 0.00555      1.6      3    837 0.19845    
Sons                1 0.01113      3.1      3    837 0.02469 *  
Openness            1 0.10960     34.3      3    837 < 2e-16 ***
Restraint           1 0.01340      3.8      3    837 0.01022 *  
Transcendence       1 0.16997     57.1      3    837 < 2e-16 ***
Interpersonal       1 0.00633      1.8      3    837 0.14998    
Residuals         839                                           
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R script :

```
############mv proj#######################
data <- read.csv("DB (1).csv")

########data cleaning#########
##Check for missing values in the entire dataset
missing_values <- colSums(is.na(data))
missing_values
data_no_missing <- na.omit(data)
# Randomly sample 850 observations
set.seed(123)
random_sample <- data[sample(1:nrow(data_no_missing), 850, replace = FALSE), ]
view(random_sample)

############descriptive analysis ########
##1)numeric
numeric <- random_sample %>%
  select(where(is.numeric)) %>%
  select(where(~ all(!is.na(.))))
describe(numeric)
```

```
cor= round(cor(numeric, use="complete.obs"),2)
cor
corrplot(cor(numeric, use="complete.obs"), order = "hclust", tl.col='black', tl.cex=.75)
corrplot(cor(numeric), order = "original", tl.col='black', tl.cex=.75)

##2) categ
table1 <- xtabs(~ data_no_outliers$sex+data_no_outliers$cp) |> addmargins()
tab <- xtabs(~ random_sample$Gender+random_sample$Student)|> addmargins()
tab
###########Factor analysis#############
##Recode 'Gender' (0 male, 1 female)
random_sample$Gender <- ifelse(random_sample$Gender == "Male", 0, 1)
##Recode 'Student' (0 other, 1 student)
random_sample$Student <- ifelse(random_sample$Student == "Other", 0, 1)
install.packages(c("psych", "FactoMineR"))
library(psych)
library(FactoMineR)
##Scale the data
data_scaled <- scale(random_sample[,1:41])
data <- data_scaled[,-c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,39,40,41)] |> as.data.frame()  #removing unwated
variables
##scree plot
principal(data)$values
plot(principal(data)$values, type = "b", col="darkred", pch="*", ylab =
     "Eigenvalues", main = "Scree Plot of Eigenvalues")
##fa using principle factor method (norotation )
fa <- fa(data, nfactors= 4, rotate = 'none', fm = 'pa', max.iter= 1)
fa


##fa using principle component method (norotation)
pc <- principal(data, nfactors = 4, rotate = 'none')
pc
```

```
pc$values  ##eigen values
pc$uniquenesses
pc.ob <- principal(r =data, nfactors = 4, rotate = "oblimin")  #oblimin rotation
pc.ob
pc.ob$uniquenesses
pc.ob$communality

##pc scores
pc.ob$scores
factors <- data.frame(pc.ob$scores)
View(factors)
attach(factors)
par(mfrow=c(2,2))
hist(TC1, xlab = "1st Factor", main = "Histogram of 1st Factor", col = "darkred")
hist(TC2 , xlab = "2nd Factor", main = "Histogram of 2nd Factor", col = "darkred")
hist(TC3 , xlab = "3rd Factor", main = "Histogram of 3rd Factor", col = "darkred")
hist(TC4 , xlab = "4th Factor", main = "Histogram of 3rd Factor", col = "darkred")


###########cluster analysis################
library(factoextra)
attach(random_sample)
pca= prcomp(random_sample[,-c(1:14,39:41)], scale= TRUE)
par(mfrow=c(1,1))
barplot(pca$x[1:40,1])
##2clusters
set.seed(123)
clusters_2<- kmeans(factors,centers=2,nstart = 25)
means= clusters_2$centers
t(means)
fviz_cluster(clusters_2, data =factors)
```

```
# 3 clusters
set.seed(123)
clusters_3<- kmeans(factors,centers=3, nstart = 25)
means= clusters_3$centers #cluster center
t(means) #transpose the cluster center
fviz_cluster(clusters_3, data = factors)

######################descriminant analysis (validation of clusters)###################
library(caret)
set.seed(123)
clusters <- clusters_3$cluster
factors$cluster <- clusters
factors$cluster <- as.numeric(factors$cluster)
true_labels <- factors$cluster
library(MVN)
mvn(data = factors[,1:4],mvnTest = 'mardia')
factors
#the results show that MVN assumption is not satisfied among the 3 groups
group1 <- as.data.frame(factors[factor$cluster==1,1:4])
group2 <- as.data.frame(factors[factor$cluster==2,1:4])
group3 <- as.data.frame(factors[factor$cluster==3,1:4])

mvn(data = group1[,c(1,2,3,4)],mvnTest = 'mardia')
mvn(data = group2[,c(1,2,3,4)],mvnTest = 'mardia')
mvn(data = group3[,c(1,2,3,4)],mvnTest = 'mardia')
# MVN is not satisfied
library(candisc)
library(heplots)
library(MVTests)
factors = data.frame(factors,clusters_3$cluster)
#check equality of covariancies among groups
```

```
varcovequal <- BoxM(data = factors[, c(1:4)], group = factors$clusters_3.cluster)
summary(varcovequal) #small p-value so reject H0
#check equality of means
TwoSamplesHT2(factors[,c(1:4)], factors$clusters_3.cluster)
#p-value is small REJECT H0
library(caTools)
#solitting data into train,test
split_index <- createDataPartition(factor$cluster, p = 0.8, list = FALSE) #80% training
train_data <- factors[split_index, ]
test_data <- factors[-split_index, ]
set.seed(123)
library(MASS)
library(caret)
factor_3 <- factor(factors$clusters_3.cluster)
class(factors$clusters_3.cluster)
attach(factors)
qda_model <- qda(clusters_3.cluster ~ Openness+Restraint+Transcendence+Interpersonal, data = train_data)
cluster3_factor <- as.factor(factors$clusters_3.cluster)
predicted <- predict(qda_model, newdata = test_data)
predicted$class
#classification table
xtabs(~predicted$class + test_data$clusters_3.cluster)

#########multivariate regression analysis############
mv_reg2<- lm(cbind(DASS_21,GHQ_12,SEC) ~ Age+factor(Gender)+Work+ factor(Student)
        +Sons+Openness+Restraint+Transcendence+Interpersonal,data=random_sample)
summary(mv_reg2)
anova(mv_reg2)
my_reg1<-lm(cbind(DASS_21,GHQ_12,SEC)~ Age+factor(Gender)+
factor(Student)+Day+Work+Sons+Appreciation_of_beauty+Bravery+Creativity

+Curiosity+Fairness+Forgiveness+Gratitude+Honesty+Hope+Humor+Judgment+Kindness+Leadership+Love+
```

```
Love_of_learning+Perseverance+Perspective+Prudence+Self_regulation+Social_intelligence+Spirituality+Tea
mwork+Zest,data=random_sample)

summary(my_reg1)
```