**Each group should do the followings**

1. From the given CSV file, make a new data frame, which should include the following columns / variables:

['Country name', 'Regional indicator', 'Ladder score', 'Logged GDP per capita', 'Social support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', 'Perceptions of corruption']

2. Select one region from the list of
   [**Sub Saharan Africa, Central and Eastern Europe, Southeast Asia, East Asia**]
   and make a Data Frame based on your choices (Example: df_east_asia = Should include only east asia region's values)
3. Take descriptive info from your dataframe and without making any visual presentations, each member of the group, should make assumptions by looking the descriptive info of the '**Perceptions of corruption**'

4. Select Justify your assumptions by using different techniques which was discussed during the review.
5. During the analysis, if you find the outliers,
   Without making anything, just make an assumption, what will happen, it this outlier was not there?
6. Make a copy of your data frame, and for the sake of the review, remove the outliers from your dataframe and make analysis and justify your assumptions.
7. By removing outliers from your dataframe what observations you observe?
8. Return to original regional dataframe with outlier and make **Covariance matrix** and comment on it. And make covariance analysis between 'Ladder Score' and 'Perceptions of corruption' and comment on it.
9. Make **correlation matrix** and comment on it. Also make correlation analysis between 'Ladder Score' and 'Perceptions of corruption' and comment on it.
10. Make a **scatter plot** between 'Ladder Score' and 'Perceptions of corruption' and make comment on it.
11. Repeat number 8,9,10 without outlier in it and make a comparison between your results.

Hope it would be helpful to improve your analysis skills.