**7PAM2000 Applied Data Science 1**

**Assignment 1: Visualisation**

**Name:** **Ganesh Maloth**

**Student ID: 21082570**

**The data source is presented below:**

<https://www.kaggle.com/datasets/eliasturk/world-happiness-based-on-cpi-20152020>

**GitHub Link is presented below:**

https://github.com/ganeshmaloth9/applieddatascience

For the assignment I have created 3 visualisations:

1. 1.line plot
2. box plot
3. scatter plot

**1.Visualisation: line plot**

Chart, line chart

Description automatically generated

Figure 1:line plot

The line graph above shows the changes in happiness scores for each country from 2015 to 2020. For the graph, I have used the years, happiness score, and country features from the dataset. In the above graph, I have observed that the China and Russia happiness scores have come down since 2015 and are lower than the UK and Germany. The United Kingdom has shown an increase in happiness since 2015, while Germany remained almost the same. China has the lowest score among the four countries.

For plotting the line graph, I have defined a function for plotting line plot of each country from 2015 to 2020. The function takes df: data frame and countries: [list of countries] as the input parameters.

**2. Visualisation: Boxplot**

Chart, box and whisker chart

Description automatically generated

Figure 2: Boxplot

The box plot represents the corruption perception index (the higher the better) of four countries for five years. If you have numerical data and want to see how it's distributed, especially across multiple groups, a box plot is the way to go. So, in this plot, I wanted to see how the CPI is distributed among countries over a five-year time span. It is observed that Sweden and Denmark have higher CPI values compared to Norway and Germany. Germany is in last place with an average mean score of around 81. The mean score of Norway is around 88, Sweden is around 85, and Norway is around 84. Sweden and Denmark have a larger change in distribution compared to the other two countries.

For the plot I have used cpi\_score, year and country columns.

**3.Visualisation: scatter plot.**

Chart, scatter chart

Description automatically generated

Figure 3: scatter plot

Determine the existence of a correlation between two variables the scatter plot is used. The scatter plot is used to study the effect of the CPI score on the happiness score. It is observed that the happiness score increases as the CPI score increases, i.e., the happiness score is directly proportional to the CPI score or positively correlated. In the graph, I have observed that Australia, Europe, and North America have the higher cpi and happiness scores, and Africa and Asia have the lower cpi and happiness scores.

I have used the continent, happiness\_score, and cpi\_score features for the scatter plot.