**Data Wrangling**

1. Import to Python
2. Join World Health 2020 data sets excel sheets for each variable– using “country”
3. Join World Happiness 2020 – using “country”

Health data not all 2020, different dates

**Evaluation Questions**

1. Identify how happiness differs by country.
   * **Variables**: Happiness Score (continuous DV), Country (categorical IV)
   * **Statistical Test**: ANOVA
   * *Possible to add population as co-variate.*
   * *Possible to run analysis by country type – developing, developed…*
2. Identify health metrics that influence happiness.
   * **Variables**: Happiness Score (continuous - DV), infant mortality rate (continuous - IV), life expectancy at birth (continuous - IV), suicide rates (continuous - IV), road traffic deaths (continuous - IV)
   * **Statistical Test**: Multiple linear regression
   * *At least 20 rows per independent variable – can add a few more variables*
3. Identify the environmental and lifestyle indicators that influence happiness.
   * **Variables**: Happiness Score (continuous - DV), clean fuel and tech (continuous), basic sanitation services (continuous), basic drinking water services (continuous), alcohol substance abuse (continuous)
   * **Statistical Test**: Multiple linear regression