

Problem Set 1

DSME 6756: Business Intelligence Techniques and Applications (Winter 2021)

Due at 9:30AM, Thursday, December 16, 2021

Instructions

Please read the Jupyter Notebook of Session 1 (1-Python_DA_Basics.ipynb, before the part on merging data) and finish the questions below. Submit a Jupyter Notebook of your solutions with code on Blackboard. The total achievable points are 5 for this problem set. Please name your Jupyter Notebook as

YourLastName_YourFirstName_PS1.ipynb (e.g., Zhang_Renyu_PS1.ipynb)

1. Playing with the WHO Data Set (5 points)

Please read the data set `WHO.csv` into Python and answer the following questions:

- (a) (1 point) **Missing data.** Which variables have at least THREE missing (i.e., `NA`) value?
- (b) (1 point) **Fertility rate.** Which country has the highest and lowest fertility rate?
- (c) (1 point) **Variations of GNI.** Which region has the minimum variation (measured by standard deviation) in Gross National Income (GNI)? What is the standard deviation of GNI in this region?
- (d) (1 point) **Child mortality of rich countries.** We define a country to be a rich country if its GNI exceeds \$20,000. What is the mean child mortality of the rich countries?
- (e) (1 point) **Correlation.** Demonstrate the relationship between income level vs. life expectancy through calculating their correlations.