CUHK Business School

Problem Set 1

DSME 6756: Business Intelligence Techniques and Applications (Winter 2023)

Due at 9:30AM, Monday, December 11, 2023

Please read the Jupyter Notebook of Session 1 and finish the questions below. Submit a Jupyter Notebook of your solutions with code on Blackboard. The total achievable points are 6 for this problem set. Please name your Jupyter Notebook as

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

1. Writing Python Functions (3 points)

(a) (1 point) A sequence $\{A_n : n \geq 1\}$ satisfies that $A_1 = 1$, $A_2 = 1$, and $A_{n+2} = A_{n+1} + A_n + 2n$ for $n \geq 1$. Write a Python function to find the value of A_n for a given positive integer n. Compute A_50.

(b) (1 point) We define $\binom{n}{m}$ as the coefficient of binomial expansion for $(1+x)^n$ ($m = 0, 1, 2, \cdots, n$). Hence, $\binom{n}{m}$ is the coefficient of x^m in $(1+x)^n$, where $m = 0, 1, 2, \cdots, n$. We know from the property of binomial coefficients that $\binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}$ for all n and all $0 \leq m \leq n-1$. Please calculate $\binom{n}{m}$ for any $n \geq m \geq 0$. Compute $\binom{40}{20}$.

(c) (1 point) Find the least common multiple of (p, q, r) given that p, q, and r are positive integers. Compute the least common multiple of $(123, 234, 345)$.

2. Playing with the WHO Data Set (3 points)

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

(a) (0.6 point) **Missing data**. Which variables have at least THREE missing (i.e., `NA`) value?

(b) (0.6 point) **Fertility rate**. Which country has the highest and lowest fertility rate?

(c) (0.6 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) (0.6 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) (0.6 point) **Correlation**. Demonstrate the relationship between income level vs. life expectancy through calculating their correlations and visualization.