1.1 Data Attributes

(To be done over Week #2)

(1) For the type of data below, state which would be considered continuous or discrete data types:

    (a) Number of patients in the wards
    (b) The blood pressure of the patient
    (c) The pulse of the patient
    (d) The size of the patient's tumour
    (e) The emergency room waiting time rounded to the nearest minute

Discrete: a, c, e; Continuous: b, d

(2) Which of the **NOIR** (Nominal, Ordinal, Interval, Ratio) scale would describe each of the data attributes below:

O (a) Rank in the Singapore Army
R (b) Number of students registered for SD6105
I (c) Shoe sizes
N (d) Jersey numbers in the Singapore national football team

	A	B	C	D	E	F	G
1	Student ID	Gender	Year of Birth	Attendance (weeks)	Score (%)	Grade	Completion Time (sec)
2	1	M	2002	12	85.5	A+	1600.31
3	2	F	2004	2	20	F	600.13
4	3	F	2003	13	55.5	C+	1800.00
5	4	M	1999	10	73.5	B+	1800.00
6	5	M	2000	11	65	B	1500.23
7	6	F	2004	8	70	B+	1700.00
8	7	F	2005	13	90.5	A+	900.34

Table 1 – Excel Table of Students Particulars and their performance in an online test

(3) Which **NOIR** scale measure best represent each of the data category (column) in the Table 1?

(4) Which measure(s) in the Table 1 (if any) can be considered a continuous data variable? Provide suitable justification for your answer.
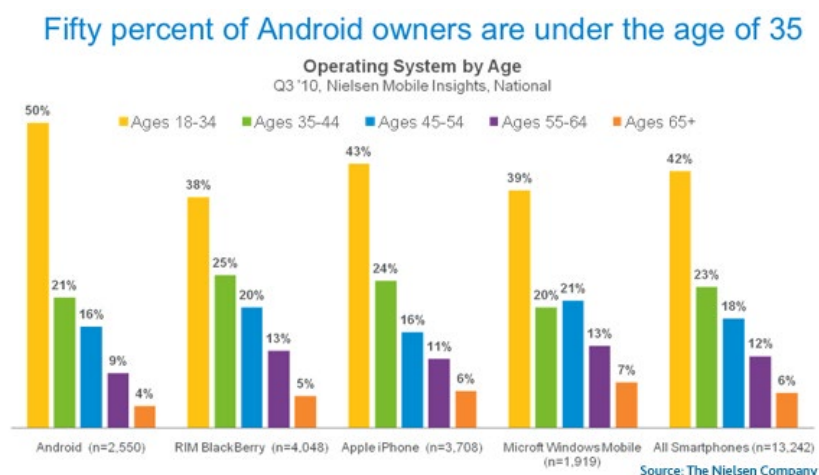

Figure 1 – Mobile OS by Age (From https://www.nielsen.com/us/en/insights/article/2010/mobile-snapshot-smartphones-now-28-of-u-s-cellphone-market/)

(5) How many data dimensions have been simultaneously visualised in the clustered bar (column) chart shown in Figure 1? Draw a possible table (*partial table will do*) for this dataset and state which **NOIR** scale measure best represent each data dimension in your table.
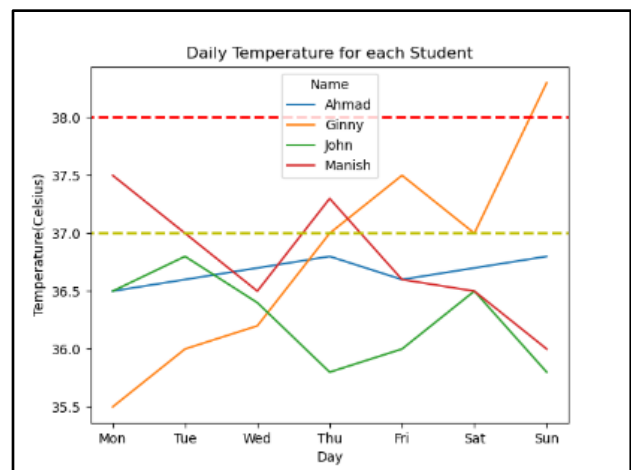
1.2 Data Wrangling

(1) **Reshaping dataframes**. Figure 2(a) shows the table that can be extracted from the datafile "*Daily_Temperature.csv*" (available in NTULearn). Using the template file "*1-2-Daily Temperature.py*", plot the line chart shown in Figure 2(c). In order to get to the stage where you can directly plot the line chart shown, the dataframe created by the table shown in Figure 2(a) must be reshaped to that shown in Figure 2(b). Use the pandas **melt** and **pivot** functions to reshape the dataframe.

Name	Mon	Tue	Wed	Thu	Fri	Sat	Sun
John	36.5	36.8	36.4	35.8	36.0	36.5	35.8
Ahmad	36.5	36.6	36.7	36.8	36.6	36.7	36.8
Ginny	35.5	36.0	36.2	37.0	37.5	37.0	38.3
Manish	37.5	37.0	36.5	37.3	36.6	36.5	36.0

(a)

Name	Ahmad	Ginny	John	Manish
Day				
Mon	36.5	35.5	36.5	37.5
Tue	36.6	36.0	36.8	37.0
Wed	36.7	36.2	36.4	36.5
Thu	36.8	37.0	35.8	37.3
Fri	36.6	37.5	36.0	36.6
Sat	36.7	37.0	36.5	36.5
Sun	36.8	38.3	35.8	36.0

(b)

(c)

Figure 2 – (a) Daily temperature table from "*Daily_Temperature.csv*". (b) The required dataframe format to plot the line chart shown in (c).

(2) **Data wrangling exercise**. Figures 3(a) and 3(b) show the tables in the datafiles "*Temp_week_1.csv*" and "*Temp_week_2.csv*" respectively (available in NTULearn). Apply the appropriate data wrangling techniques (in pandas) to produce the two line charts shown in Figures 3(c) and 3(d). Fill in the missing temperature data using the weekly mean temperature.
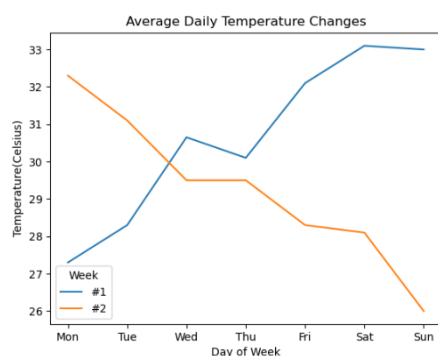
Challenge (Optional): If you can, fill in the missing data with a smoother interpolation from the two temporally nearest available data points.

(a)

Day	Week	Temp
Mon	#1	27.3
Tue	#1	28.3
Wed	#1	
Thu	#1	30.1
Fri	#1	32.1
Sat	#1	33.1
Sun	#1	33.0

(b)

Day	Week	Temp
Mon	#2	32.3
Tue	#2	31.1
Wed	#2	29.5
Thu	#2	29.5
Fri	#2	28.3
Sat	#2	28.1
Sun	#2	26.0

(c)
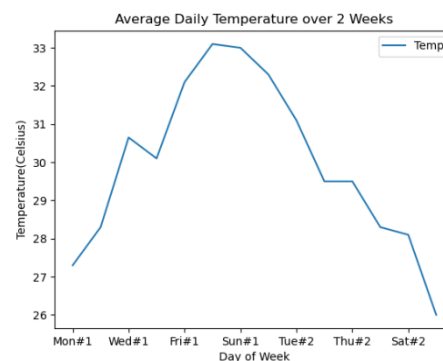
(d)

Figure 3 – (a) Temperature tables for week #1 and (b) week #2. (c) Weekly temperature changes plotted as two different lines. (d) Single line plot of temperature changes over 2 weeks.

Appendix A – Listing of 1-2-Daily Temperature Python source file

```python
# -*- coding: utf-8 -*-
"""
Created on Mon Aug  2 13:17:07 2021

CZ4124 Data Visualisation (Tutorial 1 Template)
@author: Put your name here


"""

import matplotlib.pyplot as plt
import pandas as pd

PlotWithPandas = True  # you can plot either with Pandas or Matplotlib
#----------------------------------------------------------------------
# Read in Daily Temperature datafile into dataframe Temp
#----------------------------------------------------------------------
Temp  = pd.read_csv('C:/Datasets/Daily_Temperature.csv')   #  change  to  your
directory
print('\nTable read in\n',Temp,'\n')

#----------------------------------------------------------------------
# Use MELT to covert to long form and apply column names
#----------------------------------------------------------------------
# put in your code here



#----------------------------------------------------------------------
# Use PIVOT to convert to wide form with Names in each column
#----------------------------------------------------------------------
# put in your code here



#-----------------------------
# Use Pandas to plot line chart
#-----------------------------
if(PlotWithPandas):
    print('\nPlotting with Pandas')
    # put your Pandas plotting code here



#-------------------------------
# Use Matplotlib to plot line chart
#-------------------------------
else:
    print('\nPlotting with Matplotlib')
    # Alternatively, you can do the plot using Matplotlib or
    # any other Python plotting library
```