Name: Han Hong Tuck from EP0302_01

Title of Data Analysis: The demand for the number of medical practitioners per 10,000 total population increases with time.

Questions to answer to gain deeper insights into the chosen datasets

Question 1: Is there an increasing or decreasing trend in the number of medical practitioners (specifically doctors, nurses and dentists) over the years? ¶

Question 2: Are all the data available/present for number of medical practitioners every 10000 total population from 1960 to 2019

Question 3: How many data points should we plot to show a consistent trend for the number of medical practitioners per 10000 total population? / In other words, from which year to which year should we extract the data out of the dataset and plot to display the trend?

Url of dataset used: [https://data.gov.sg/dataset/doctors-per-10-000-total-population (https://data.gov.sg/dataset/doctors-per-10-000-total-population)](https://data.gov.sg/dataset/doctors-per-10-000-total-population)

Write Python code that uses the Pandas package to extract useful statistical or summary information about the data

In [2]:
```
import pandas as pd

df_medical_pract = pd.read_csv('medical-practitioner-per-10-000-total-population.

#to see the first five rows of the pandas dataframe
print(f"First Five Sets of dataset: \n {display(df_medical_pract.head())} \n\n")

#to see the last five rows of the pandas dataframe
print(f"Last Five Sets of dataset: \n{df_medical_pract.tail()} \n\n")

#to get details/info about the pandas dataframe
print(f"\n Dataframe Info: \n{df_medical_pract.info(verbose=bool)}\n")

#to get info on the number of rows and columns about the pandas dataframe
print(f"\n Number of rows and columns: \n{df_medical_pract.shape}\n\n")

#to get summary statistics for all the data as a whole
print(f"Summary statistics for all data: \n\n{df_medical_pract.describe()}\n\n")

#to get summary statistics for doctors,dentists and nurses individually
df_medical_pract_stats = df_medical_pract.groupby(["level_1"])[["value"]].describ
print(f"Summary Statistics for doctors,dentists and nurses individually: \n\n{df_
```

	level_1	value
year		
1960	Doctors Per 10,000 Total Population	4.0
1960	Dentists Per 10,000 Total Population	2.0
1960	Nurses Per 10,000 Total Population	NaN
1961	Doctors Per 10,000 Total Population	4.0
1961	Dentists Per 10,000 Total Population	2.0

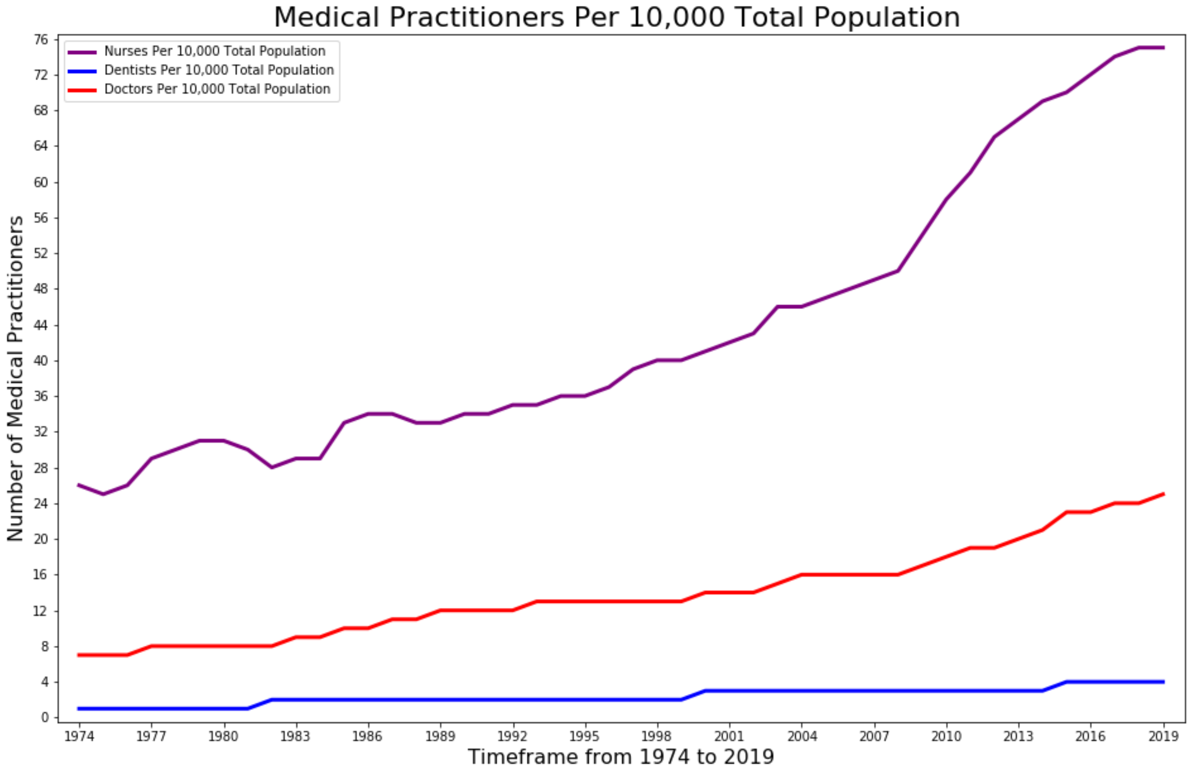
```
First Five Sets of dataset: 
 None 


Last Five Sets of dataset: 
                                  level_1  value
year                                            
2018  Dentists Per 10,000 Total Population    4.0
2018    Nurses Per 10,000 Total Population   75.0
2019   Doctors Per 10,000 Total Population   25.0
2019  Dentists Per 10,000 Total Population    4.0
2019    Nurses Per 10,000 Total Population   75.0 


<class 'pandas.core.frame.DataFrame'>
Int64Index: 180 entries, 1960 to 2019
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   level_1  180 non-null    object 
 1   value    166 non-null    float64
```

```
dtypes: float64(1), object(1)
memory usage: 4.2+ KB

 Dataframe Info:
None


 Number of rows and columns:
(180, 2)


Summary statistics for all data:

            value
count  166.000000
mean    17.180723
std     18.919379
min      1.000000
25%      3.000000
50%      8.500000
75%     28.750000
max     75.000000


Summary Statistics for doctors,dentists and nurses individually:

level_1      Dentists Per 10,000 Total Population  \
value count                             60.000000
      mean                               2.283333
      std                                0.804472
      min                                1.000000
      25%                                2.000000
      50%                                2.000000
      75%                                3.000000
      max                                4.000000

level_1      Doctors Per 10,000 Total Population  \
value count                            60.000000
      mean                             11.933333
      std                               5.689320
      min                               4.000000
      25%                               7.000000
      50%                              12.000000
      75%                              16.000000
      max                              25.000000

level_1      Nurses Per 10,000 Total Population
value count                           46.000000
      mean                            43.456522
      std                             15.191674
      min                             25.000000
      25%                             33.000000
      50%                             38.000000
      75%                             49.750000
      max                             75.000000
```

Write Python code that uses Matplotlib package to produce useful data visualizations that explain the data.

In [2]:

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker

#read from file to get dataset
df_medical_pract = pd.read_csv('medical-practitioner-per-10-000-total-population.

#declare list to store data in
medical_pract = ["df_nurses","df_dentists","df_doctors"]

#store indexes of doctors, nurses and dentists in a list
column_names = ["Nurses Per 10,000 Total Population","Dentists Per 10,000 Total P

#extracting doctors,nurses and dentists per 10000 total population from year 1974
#using the previously declared lists -> medical pract and column names
for i in range(len(medical_pract)):
    medical_pract[i] = df_medical_pract[df_medical_pract.level_1==column_names[i]

#declare fig and ax object for plotting
fig, ax = plt.subplots(figsize=(16,10))

#set margin of x-axis and y-axis
ax.set_xmargin(0.02), ax.set_ymargin(0.02)

#set different colors for doctors, nurses and dentists individually
colors=["purple","blue","red"]

#to plot the data for doctors, dentists and nurses using a loop
for i in range(len(medical_pract)):
    ax.plot(medical_pract[i].index,medical_pract[i]["value"],label=column_names[i

#to set title and label for x-axis and y-axis on the graph
ax.set_title("Medical Practitioners Per 10,000 Total Population ",fontsize=22)
ax.set_xlabel("Timeframe from 1974 to 2019",fontsize=16), ax.set_ylabel("Number o

#set the frequency of the xticks and yticks on the x-axis and y-axis
ax.xaxis.set_major_locator(ticker.MultipleLocator(3)), ax.yaxis.set_major_locator

#display the legend
ax.legend()

plt.show()
```



For each dataset, explain the nature of that dataset (i.e. what is in that dataset) or any pecularities about it you wish to highlight and explain the process you went through to analyse that dataset, . Where possible, you should specifically mention how you used the Pandas or Matplotlib functions to achieve a certain outcome e.g. to transform the data or to produce a certain visualization:

Pecularities to highlight:

From the year 1960 to 1973, the number of nurses per 10,000 total population is not available. However, during that period, there is still a record for the number of doctors and dentists. When the data for the numnber of nurses is available, it is significantly higher than the number of doctors and nurses.

Nature of dataset:

The nature of the dataset consists of the number of medical practitioners from their respective professions (doctors, dentists and nurses from 1960 to 2019). The number of medical practitioners is in relation to the population size of every 10,000. During the process of analysising the dataset, I found out that some of the data for the number of nurses is not available. Hence, I need to decide whether I should include the timeframe from the year 1960 to 1973. Upon further consideration, I find that including this subset of data is misleading as it will lead to false impression that there is a sudden increase in the number of nurses from 1973 (0 nurses) to 1974 (26 nurses), where the data was not yet available. This leaves me with the dataset from 1974 to 2019. Since I am trying to find a relationship for the number of medical practitioners over the years, I would need to include as many data into the graph to show a consistent trend. Therefore, I have decided to stick with the dataset from 1974 to 2019, without limiting the number of data to extract from the dataset.

Process of using Pandas or Matplotlib functions to transform the data:

The dataset consists of the columns: year, level_1 (specification of data for doctors, dentists and nurses) and value of number of each specification. Firstly, I declare a list to store the data in called "df_nurses", "df_dentists" and "df_doctors" and another list to store indexes of doctors, nurses and dentists. Then,I extract doctors,nurses and dentists per 10000 total population from year 1974 to 2019 using using the previously declared lists (medical pract and column names) within a for loop. Then, I declare fig and ax object for plotting and set the margin of x-axis and y-axis so that the xticks and yticks will stretch across the entire x-axis and y-axis. Afterwards, I store the colors of the doctors, nurses and dentists in a list so that when I plot the data for doctors, dentists and nurses using a loop, the line graph will show different colors for different job professions. Then, I set the title and label for x-axis and y-axis on the graph and also set the frequency of the xticks and yticks on the x-axis and y-axis. Finally, I display the legend and call plt.show() to display the graph.

For each dataset, highlight the insights you have gained from analysing the data and any conclusions or recommendations you want to make as a result of the analysis:

After plotting the graph, I am able to tell that the number of doctors, dentists and nurses increases with time, dentists showing only a very slight increase, followed by the doctors showing a gradual, steady incease and then followed by nurses with the fastest increase in numbers from 1974 to 2019. From this, I am able to conclude that the the demand for the number of nurses is the highest. Aside from that, I am also able to conclude that the demand for the number of nurses increases with time. From the graph, we are able to tell that from the year 2008-2009 to year 2016, the increase in the number of nurses becomes steeper compared to the steady increasing trend from 1974 to 2008. However, the increase in the number of nurses reduces when it is approaching 2019 where the increase becomes gradual. To conclude, this supports the title of the data analysis as there is increasing trend in the number of medical practitioners (doctors, nurses and dentists) from 1974 to 2019.