Exercise 5 – There are 2 Parts, **please submit both**

**PART 1: General questions**

1. What is Agile software development and why it's relevant?
2. Why waterfall approach is rarely a great solution for software development?
3. Define Machine Learning and provide examples of its application.

**PART 2: Coding**

SPECIFICATION

Description: This program uses the covid\_comorbidities\_USsummary.csv data to print totals and create a bar and pie charts showing the number of people who died because of comorbidities in COVID-19 cases.

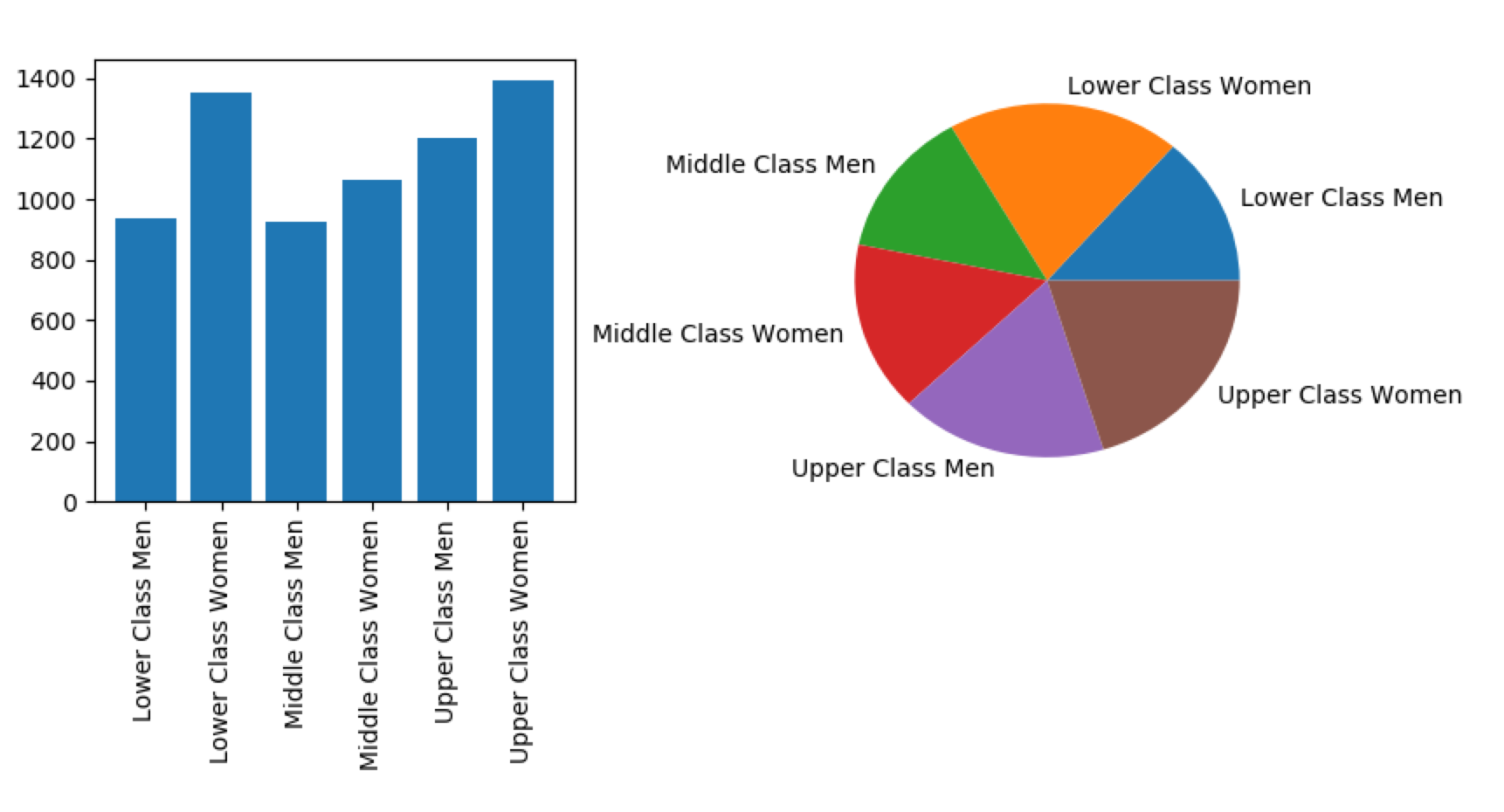
It will also explore correlations between the different variables.

Input:

covid\_comorbidities\_USsummary.csv

Output:

* Count of people per class of age ("Age Group")
* Bar chart with classes of age. The x-axis will be the class ages, y-axis the related number of deaths
* Pie chart on the same data, showing the distribution of deaths per age group (sample below – **NOT the actual graph**)
* Print the comorbidity with the highest number of "COVID-19 Deaths" for the population of less than 35 years of age



Procedure notes:

You will need to import matplotlib.pyplot. The preferred way to do this is: import matplotlib.pyplot as plt

You may want to create a list of counters, like: counts\_conditions = [0, 0, 0, 0, 0, 0, 0, 0] where the elements of the list would be the number of deaths per each one of the 8 age categories ['0-24', '25-34', '35-44', '45-54', '55-64', '65-74', '75-84', '85+']

You may also want to create a dictionary "counter\_dict" to hold the number of "COVID-19 Deaths" for the population of less than 35 years of age.

The program will use a function (to be created as part of this exercise) get\_index(parts) to process each relevant line of data and return an index pointing to the proper element of the list of counters.

It may be useful to use the variables "counts\_conditions" and "counter\_dict" as global variables.

You also have the option of reading the file into a pandas data structure and use pandas features to calculate the values needed to plot the charts and to get the number of deaths. In this case, you may not need "counts\_conditions" and "counter\_dict". Please note: the procedure below in NOT using pandas.

Procedure:

1. import matplotlib

**Function**

* Define a function named get\_index(parts) that accepts a list created from a row of the data file and returns an index that is used to update the list of counters. You would skip rows with ' 'COVID-19' as cause of death. With the same function, you may want to calculate the comorbidity with the highest number of "COVID-19 Deaths" for the population of less than 35 years of age.

**Main Program**

1. Create a list of counters with eight elements, initialized at zeros.
2. Create a dictionary to calculate the comorbidity with the highest number of "COVID-19 Deaths" for the population of less than 35 years of age, initialized at empty.
3. Open the data file for reading. Since this is a program specific to this file, there is no reason to use input to request the file name.
4. Read each line in the file. Skip any line with 'COVID-19' as cause of death(meaning no comorbidity). For the other lines:

* call the function to get the appropriate index for the *counters* list: index = get\_index(parts)
* update the appropriate count
* update the dictionary to trace the comorbidity/deaths for the population of less than 35 years of age

1. Create the charts, as subplots.
   * Subplot 221 (left): bar chart

* bar plot
* x = ['0-24', '25-34', '35-44', '45-54', '55-64', '65-74', '75-84', '85+']
* y = values from the counters list for each one of the 8 groups. Name y axis as “Count”
* do not overlap labels
  + Subplot 222 (right): pie chart. This is with the same y data as the bar chart
* pie plot
* label the wedges using the age categories
* Add percentage values as second label to your pie chart
  + Insert a proper title to the main plot

1. Print the comorbidity with the highest number of deaths for the population of less than 35 years of age.
2. Run correlation analysis to determine relationships between variables
3. Write a 2+ pages with an interpretation of your analysis. The report will contain the visualizations generated by your script.

**Submit your program .py and related narrative doc/pdf files via Canvas.**

**The PART 1 can be either in a separate doc/pdf file or added as comment to the .py file**