**Exercise 5**

The given script is an analysis of COVID-19 deaths in the United States, focusing on the age groups and comorbidities associated with these deaths. The script uses data from the 'covid\_comorbidities\_USsummary.csv' file and imports necessary libraries like Pandas, Seaborn, and Matplotlib for data manipulation and visualization.  
  
The first step in the analysis is to load the data and display the first ten rows to check the data's structure and contents. After this, the script creates a list of unique age groups present in the dataset and calculates the total number of deaths associated with each age group. The results are stored in a dictionary, which is then used to create a Pandas DataFrame with the age groups and their associated death counts. The DataFrame is then displayed to the console to show the number of people in each age group who died due to COVID-19.  
  
Next, the script creates two visualizations to represent the data in the DataFrame. The first visualization is a pie chart that displays the percentage of deaths in each age group. The pie chart uses random colors to represent each age group and displays the percentage of deaths associated with each age group in the legend. The second visualization is a bar plot that displays the number of deaths associated with each age group. The bar plot uses the same age group labels as the pie chart and displays the exact number of deaths associated with each age group.  
  
The script then calculates the correlation between the number of comorbidities and the number of COVID-19 deaths. To do this, the script creates a list of unique comorbidities present in the dataset and assigns a unique number to each comorbidity. The script then calculates the correlation between the number of comorbidities and the number of COVID-19 deaths using the Pandas DataFrame's corr() method. The resulting correlation coefficient is displayed to the console. The script also creates a heatmap visualization using the Seaborn library to display the correlation matrix.  
  
Finally, the script identifies the comorbidities associated with the highest number of deaths in two age groups - 0-24 and 25-34. For each age group, the script selects the row with the highest number of COVID-19 deaths, sorts the rows in ascending order by the number of COVID-19 deaths, and selects the second-highest row. The script then displays the comorbidity associated with this row to the console.  
  
In conclusion, the given script is a useful analysis of COVID-19 deaths in the United States, providing insights into the age groups and comorbidities associated with these deaths. The script uses various data manipulation and visualization techniques to present the data in a clear and concise manner, making it easy to understand and interpret. The script's use of Pandas, Seaborn, and Matplotlib libraries also demonstrates the power and flexibility of these tools in performing data analysis and visualization tasks. Overall, this script provides a valuable resource for anyone interested in understanding the impact of COVID-19 on different age groups and the comorbidities associated with these deaths.

Output of the program :

Chart

Description automatically generated

Count of people per class of age:

Age Groups Deaths

0 0-24 2865

1 25-34 10923

2 35-44 29371

3 45-54 84443

4 55-64 220269

5 65-74 408166

6 75-84 515797

7 85+ 547443

8 Not stated 35