**Report:**

**2.1: Cause of death in the US:**

**Algorithm for finding out the major cause of death in a year:**

1. Read the json and csv files of a year using open and read\_csv function of pandas module.
2. Read the column ‘358\_cause\_recode’ from the csv file and find out the unique values and element count of this column using the unique method of NumPy module.
3. Note the index of the maximum value of element count and use the same index for pulling out the corresponding unique value. This gives the code of major cause of death and is used as a key in the json file to find out its equivalent vale i.e., major cause of death. (Used index method of a list, max function of NumPy module, indexing method of a dictionary)
4. It is to be noted that codes of major causes in csv files are 1,2,3 etc. and the same codes in json file are 01, 02, 03etc. So, while reading json file at the end, codes are to be restructured accordingly. (Use string concatenation)
5. Using a for loop, repeat the same steps for all the years to find out the major cause of death across all the years.

**Result:** Considering 358\_cause\_recode, the major cause of death is “215”: “All other forms of chronic iso-chemic heart disease (I20, I25.1-I25.9)” and is observed to be the same over the years 2005 to 2015.

Text

Description automatically generated

**Trend of major cause of death:**

* Maximum value of element count obtained in step-2 of above algorithm is the number of deaths for a year. Similar value is recorded over all the years and a graph is plotted as shown below.
* Significantly decreasing trend is observed i.e., number of deaths due to major cause is decreasing from 2005 to 2015.

Chart, line chart

Description automatically generated

**Algorithm for finding major Causes of death for different ages:**

1. Read the json and csv files of a year using open and read\_csv function of pandas module.
2. Using groupby function, csv file is grouped and sorted based on ages i.e., ‘age\_recode\_52’ column.
3. An empty list is taken and all the rows with an age group(obtained above) are appended as individual elements of the list.
4. Running a for loop over the length of the above list (i.e, 52 age groups), record the causes of death for each age group and find out the unique values and element count of this causes of death column using the unique method of NumPy module.
5. Note the index of the maximum value of element count and use the same index for pulling out the corresponding unique value. This gives the code of major cause of death and is used as a key in the json file to find out its equivalent vale i.e., major cause of death. (Used index method of a list, max function of NumPy module, indexing method of a dictionary)
6. It is to be noted that codes of major causes in csv files are 1,2,3 etc. and the same codes in json file are 01, 02, 03etc. So, while reading json file at the end, codes are to be restructured accordingly. (Use string concatenation)
7. In a year, the causes of death for various age groups can be printed and visualized for a better understanding.
8. Plot a bar graph for number of deaths in a particular year for different ages groups.
9. Using a for loop, major causes of death for young people and old people are printed.
10. Using a for loop, repeat the same steps for all the years to find out the major cause of death across all the years.

Text

Description automatically generatedText

Description automatically generated**Result for major cause of death across all age groups:**

**Text

Description automatically generatedResults of major causes of death among young and old people are:**

**Chart, histogram

Description automatically generatedGraph for the number of deaths for different age groups in year 2005 is:**

Similar graphs are obtained for years up to 2015.

**Observations:** It is inferred from these graphs that the major cause of death lies in the age group 43 which means people of age 85-89 are the most affected ones and the major cause of death among them is found out again to be “215”: “All other forms of chronic iso-chemic heart disease (I20, I25.1-I25.9)” which aligns well with our conclusions in first part.