# **ASSIGNMENT 1**

**Question 1:**

**Produce a line plot showing multiple lines with proper labels and legend. Describe what conclusions you can draw from this plot.**

**Solution:**

**Line Graph:**

**Description:**

The line graph visualizes the fluctuation of different entities in a data set with the passage of time.

**Reason of selection:**

The reason of selecting a line graph is that it can easily handle the large bulk of data. The user can easily visualize its data and can easily track changes in its dataset time to time.

**Code:**

**#imporitng libraries**

import pandas as pd

import matplotlib.pyplot as plt  
**#reading data from file**

**file = pd.read\_excel(r'C:\Users\DELL\Desktop\ADS1 Assignment\Covid deaths recorded from march-September 2022.xlsx')**

**#sperating the exel data and saving it inside variables**

months = file['Week ending']

months

**#now let's plot the graph**

plt.figure(figsize= (10,4))

plt.plot(months,file.loc[:,"Hospital"], color="blue", label="total deaths in hospital")

plt.plot(months,file.loc[:,"Care home"],color = "red" , label="total deaths in care home ")

plt.plot(months,file.loc[:,"Home"],color="green", label="total deaths in home")

plt.plot(months,file.loc[:,"Other"],color="brown", label="other deaths")

plt.grid(linewidth=1)

plt.legend()

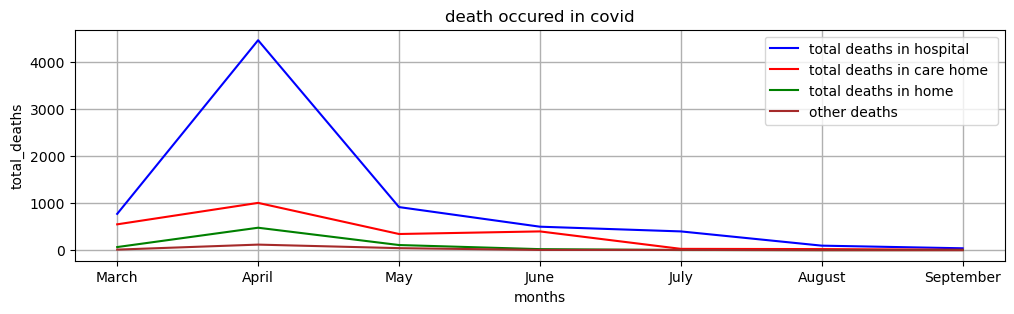
plt.xlabel("months")

plt.ylabel("total\_deaths")

plt.title("death occured in covid")

plt.show()

**Output:**



**Link:** https://data.london.gov.uk/dataset/coronavirus--covid-19--deaths

**Conclusion:**

When a user takes a good look in to the graph he can clearly analyze the deaths caused by the covid-19 and can easily visualize the ratio of where most of the deaths have occurs. The blue line of this graph analyzes the no of deaths caused in a hospital, the red line visualizes the death caused in a care home, the green line visualizes the deaths caused in homes and the brown line visualizes the deaths caused in other places. As we can clearly visualized that the most deaths caused in the hospitals due to covid-19 thus we need some extra care on those spots to save the life. On x-axis we can see the no of months and on y-axis we can see the total number of deaths occur during this session.

**Question 2: Produce graphs using two other visualisation methods. Explain why you picked this type of graph and describe what conclusions you can draw.**

**Solution:**

**Part A:**

**Scatter plot:**

**Description:**

A scatter plot shows the correlation between two related entities thus forming a group of dots.

The group of dots are usually known as clusters.

**Reason of selection:**

The reason I selected this graph is because it really help us to analyze the relation between two entities likewise if we combine the two entities would that mean something or not.

**Code:**

**#importing all libraries**

import pandas as pd

import matplotlib.pyplot as plt

**#reading data from an exel sheet**

file = pd.read\_csv(r'C:\Users\DELL\Desktop\ADS1 Assignment\2019-20-enforcement-data-sampling.csv')

**# Separating data from excel sheet**

microbiology = file['AnalysisType-Microbiologicalcontamination']

microbiology

other\_contamination = file['AnalysisType-Othercontamination']

other\_contamination

Analysis\_type\_composition = file['AnalysisType-Composition']

total\_samples = file['Totalnumberofsamplestaken']

total\_samples

**# Plotting the Graph**

plt.scatter(total\_samples,microbiology,color='green',s=100,label='sampling',marker='\*')

plt.scatter(other\_contamination,Analysis\_type\_composition,color='yellow',s=100,label='composed samples',marker='\*')

**#labelling and giving title to the graph**

plt.grid()

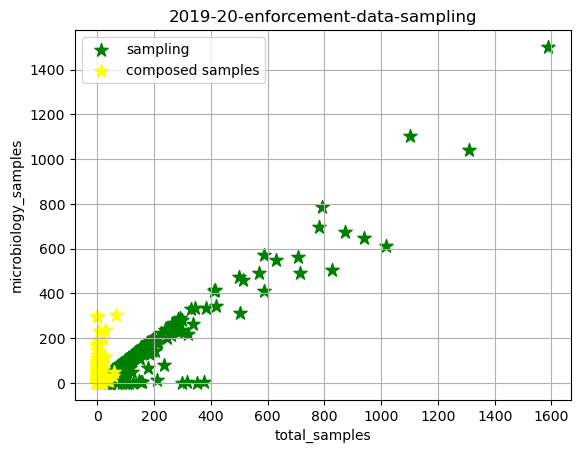
plt.xlabel('total\_samples')

plt.ylabel('microbiology\_samples')

plt.legend()

plt.title('2019-20-enforcement-data-sampling')

**Output:**

****

**link:** [Results for "Health" - data.gov.uk](https://www.data.gov.uk/search?filters%5Btopic%5D=Health)

**Conclusion:**

When a user takes an analysis of this graph the user can easily understand the correlation between different entities. Such as, we can clearly visualize the relation between total no of samples present and the total microbiology samples by green stars as well as the relation between contamination and composition by yellow dots. On x-axis we have the total number of samples and on y-axis we have the microbiology samples.

**Part B:**

**Area Chart:**

**Description:**

An area chart is the chart that show changes how one or more group values changes over the progression of time of a second variable, typically with that of time.

**Reason of selection:**

The reason of using the area charts is that is easily understandable and very good in comparing the datasets.

**Code:**

**#importing all libraries**

import pandas as pd

import matplotlib.pyplot as plt

**#reading data from an excel sheet**

file = pd.read\_excel(r'C:\Users\DELL\Desktop\ADS1 Assignment\Ages.xlsx')

file.head()

**#creating a new file to separate the job title column and to count all categories**

file\_1 = file.sort\_values(by=['All ages'])

file\_1

all\_ages = file\_1['All ages']

all\_ages

age\_04= file\_1['Age 0 to 4']

age\_04

age\_57= file\_1['Age 5 to 7']

age\_above85 = file\_1['Age 85 and over']

age\_above85

plt.fill\_between(file\_1['All ages'],age\_04,label=['Age 0 to 4'],color = 'red')

plt.fill\_between(file\_1['All ages'],age\_57,label=['Age 5 to 7'],color = 'green')

plt.fill\_between(file\_1['All ages'],age\_above85,label=['Age above 85'],color = 'blue')

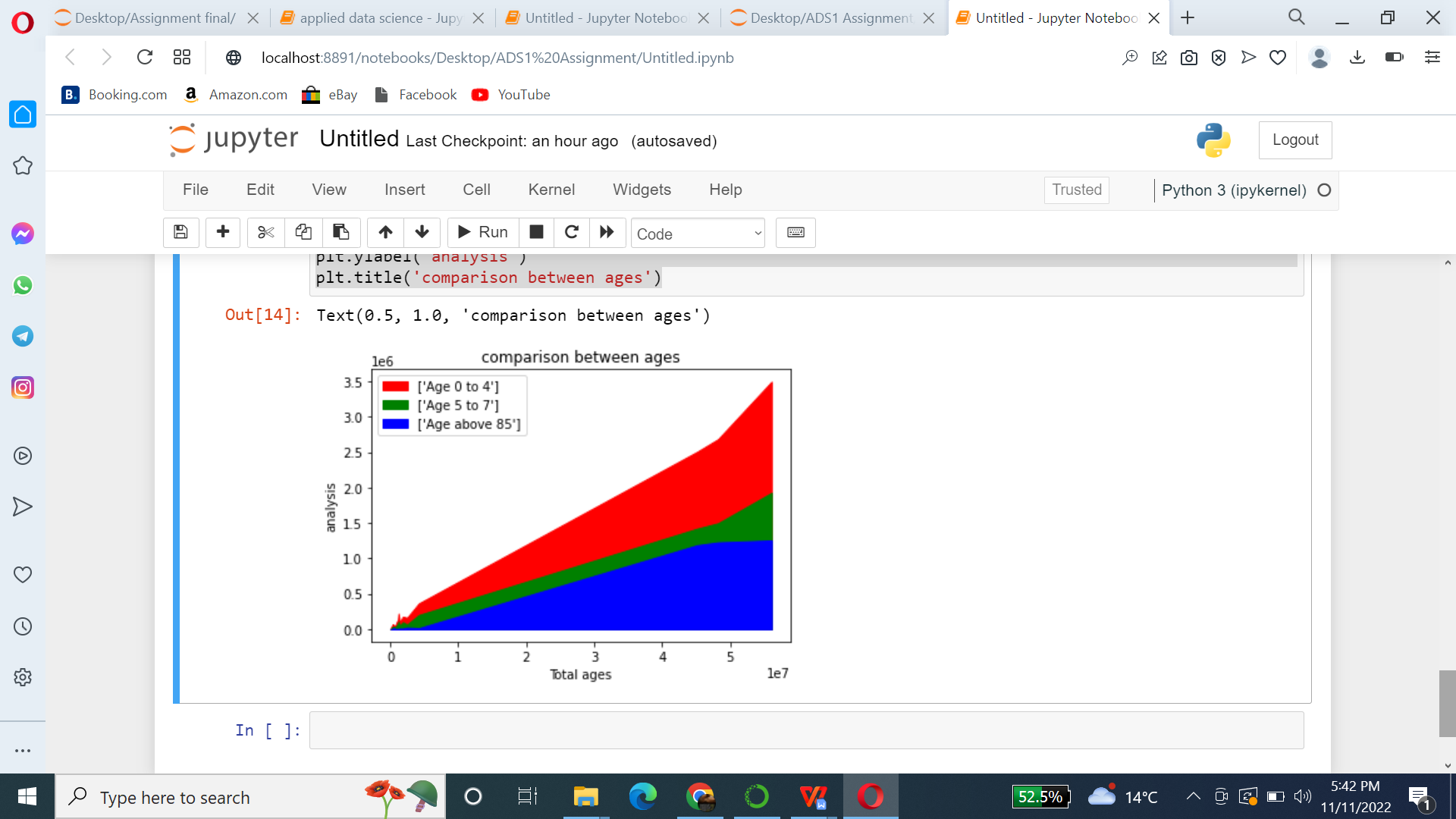
plt.legend(loc = 'upper left')

plt.xlabel('Total ages')

plt.ylabel('analysis')

plt.title('comparison between ages')

**Output:**



**link:** [Average age of the population by ethnicity - data.gov.uk](https://www.data.gov.uk/dataset/76bda24c-7793-407c-8430-7cb69039e9b7/average-age-of-the-population-by-ethnicity)

**Conclusion:**

When a user takes an analysis of this graph the user can easily compare the different data sets in the graph as we can easily compare the no of people with ages 0 to 4, ages 5 to 7, and ages 55 to 59 by easily visualizing the graph. The red area shows the no of people from ages 0 to 4, while the green area shows the no of people from 5 to 7 and the blue area is for the number of people from 55 to 59 age. On the x-axis, we can see the total ages of people present, while on the y-axis we can see the analysis change in no of people.