

Demographic analysis

Explore how the virus affects different age groups

Objective of this report: We are going to work on 4 columns which are the three different age ranges which are from 0 to 14 , from 15 to 64 and finally from 65 to above and comparing each range of those columns to the number of death, seeing which range had the highest number of death. Now we will show the steps for this report.

1. Cleaning data:

Column	Unique	Null	Blank	Pattern Count	Mean
deaths	0.00% (0)	0.00% (0)			102.12
SP_POP_0014_TO	0.00% (0)	2.19% (5,482)			154.79
SP_POP_1564_TO	0.00% (0)	2.97% (7,525)			149.5...
SP_POP_65UP_TO	0.00% (0)	2.97% (7,525)			24.56...

- As we can see total death column has no missing (null) values but as for the three different age range there are empty (null) cells so we need to clean it by removing the empty cell as its percentage is not high so it won't affect our dataset fewer columns will be removed.

Column	Minimum Length	Maximum Length	Ordinal Position	Primary Key	Non-null
longitude			7	No	244,927
SP_POP_0014_TO			19	No	245,977
SP_POP_1564_TO			20	No	245,977
SP_POP_65UP_TO			21	No	245,977

Column	Operator
SP_POP_0014_TO	Not null
SP_POP_1564_TO	Not null
SP_POP_65UP_TO	Not null

- We used filter to remove the null values from the three age ranges we got.
- Now we will need to create new columns to see how many number of death in each age range as the column (total death) contain the number of total death of the three ranges together.

2. Adding new columns.

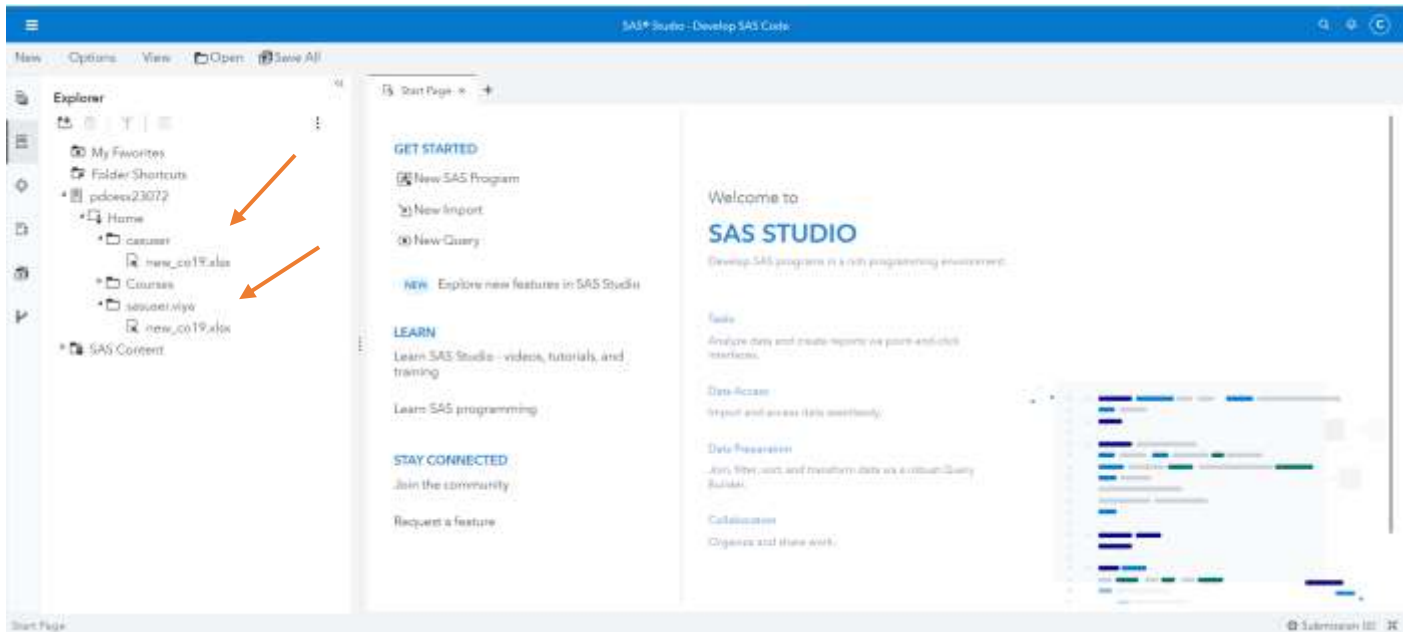
The screenshot shows the SAS Data Studio interface. The main window displays the 'NOVEL_CORONA_VIRUS (session)' table with columns: death..., reco..., death..., date, cou..., cou..., and state. The '3. Calculated Column' dialog box is open, showing the formula $(SP_POP_65UP_TO * deaths_total) / SP_POP_TOTL$ and the option to 'Create new column' named 'death(15-64)'. The 'Plan' panel on the right shows the data flow with three calculated columns.

- We calculated three columns so that each range we will know how many death were in this range the equations we used were
 - As for ages from 0-14 $total(0-14) * total\ death / total\ population$ this will give us approximate number of death in the range of 0-14 and store it in a new column.
 - As for ages from 15-64 $total(16-64) * total\ death / total\ population$ this will give us approximate number of death in the range of 16-64 and store it in a new column.
 - As for ages from 65-up $total(65-up) * total\ death / total\ population$ this will give us approximate number of death in the range of 65-UP and store it in a new column.
- Now we got a new table with clean and extra columns so we can download it in our device to upload it on sas as following.

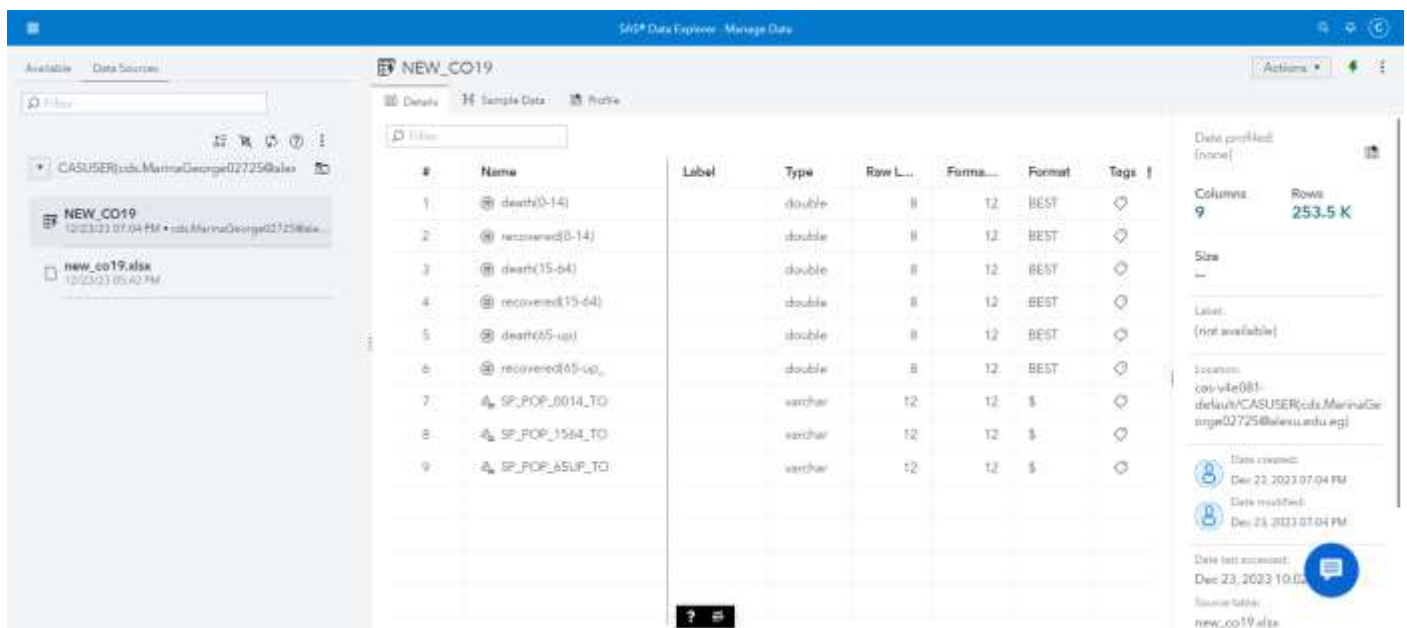
The screenshot shows a Microsoft Excel spreadsheet with columns A through F. Column A is labeled 'death(0-14)', column B is 'death(15-64)', and column C is 'death(65-up)'. Columns D, E, and F are labeled 'SP_POP_0014_TO', 'SP_POP_1564_TO', and 'SP_POP_65UP_TO' respectively. The data is organized in rows, with the first row containing headers and subsequent rows containing numerical values.

	A	B	C	D	E	F
1	death(0-14)	death(15-64)	death(65-up)	SP_POP_0014_TO	SP_POP_1564_TO	SP_POP_65UP_TO
2	0	0	0	16065209	75571759	34892133
3	0	0	0	16065209	75571759	34892133
4	10.9192903	51.36503203	23.71567835	16065209	75571759	34892133
5	0.141731412	41.00328263	19.85498653	16065209	75571759	34892133
6	1.269684918	5.972678143	2.757637018	16065209	75571759	34892133
7	13.07775466	61.51858487	28.40366128	16065209	75571759	34892133
8	0.380905475	1.791803443	0.827291105	16065209	75571759	34892133
9	0.126968492	0.597267814	0.275763702	16065209	75571759	34892133
10	0.380905475	1.791803443	0.827291105	16065209	75571759	34892133
11	0.380905475	1.791803443	0.827291105	16065209	75571759	34892133
12	0	0	0	16065209	75571759	34892133
13	0.126968492	0.597267814	0.275763702	16065209	75571759	34892133
14	0.126968492	0.597267814	0.275763702	16065209	75571759	34892133
15	2.79310642	13.13889191	0.066801438	16065209	75571759	34892133
16	1.269684918	5.972678143	2.757637018	16065209	75571759	34892133
17	9.141731412	41.00328263	19.85498653	16065209	75571759	34892133
18	0	0	0	16065209	75571759	34892133
19	0	0	0	16065209	75571759	34892133
20	0	0	0	16065209	75571759	34892133
21	41.64566552	195.9038431	90.45049419	16065209	75571759	34892133
22	0	0	0	16065209	75571759	34892133
23	0.126968492	0.597267814	0.275763702	16065209	75571759	34892133
24	0.126968492	0.597267814	0.275763702	16065209	75571759	34892133
25	0	0	0	16065209	75571759	34892133
26	10.9192903	51.36503203	23.71567835	16065209	75571759	34892133
27	0.888778443	4.1808747	1.930345813	16065209	75571759	34892133
28	0	0	0	16065209	75571759	34892133
29	0.253936884	1.194515629	0.551527408	16065209	75571759	34892133

3. Uploading new data set on sas.

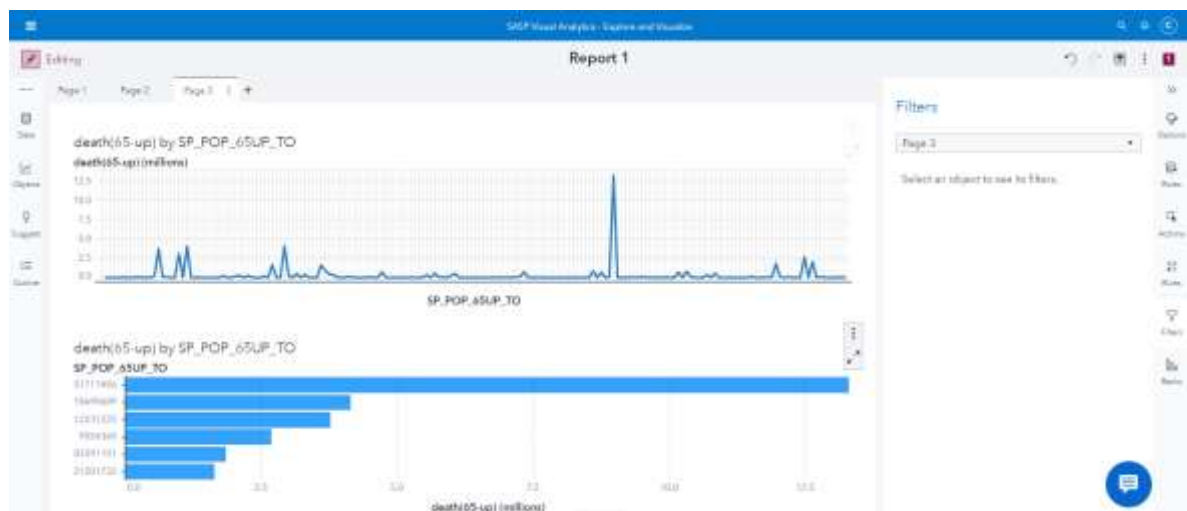
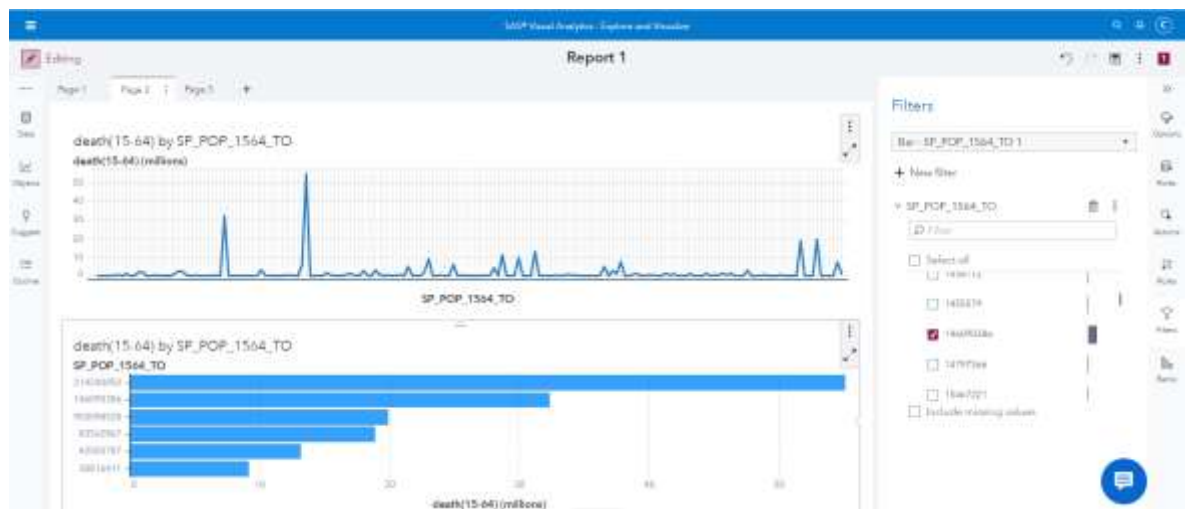
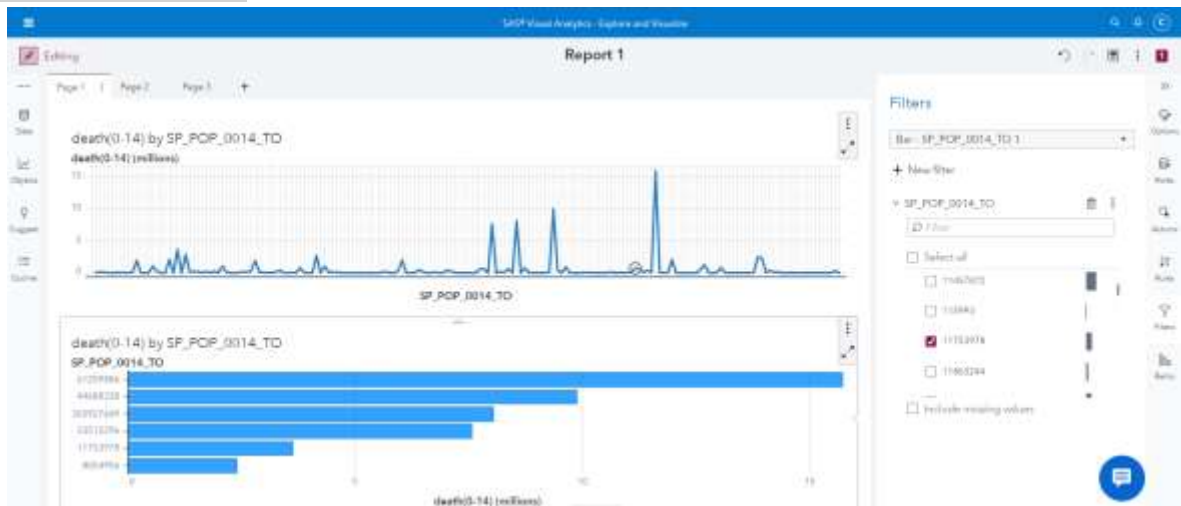


- first in sas develop sas code we will upload our new dataset in sasuser.viya and it will automatically be also uploaded in causer in which we can use in data visualization next steps.

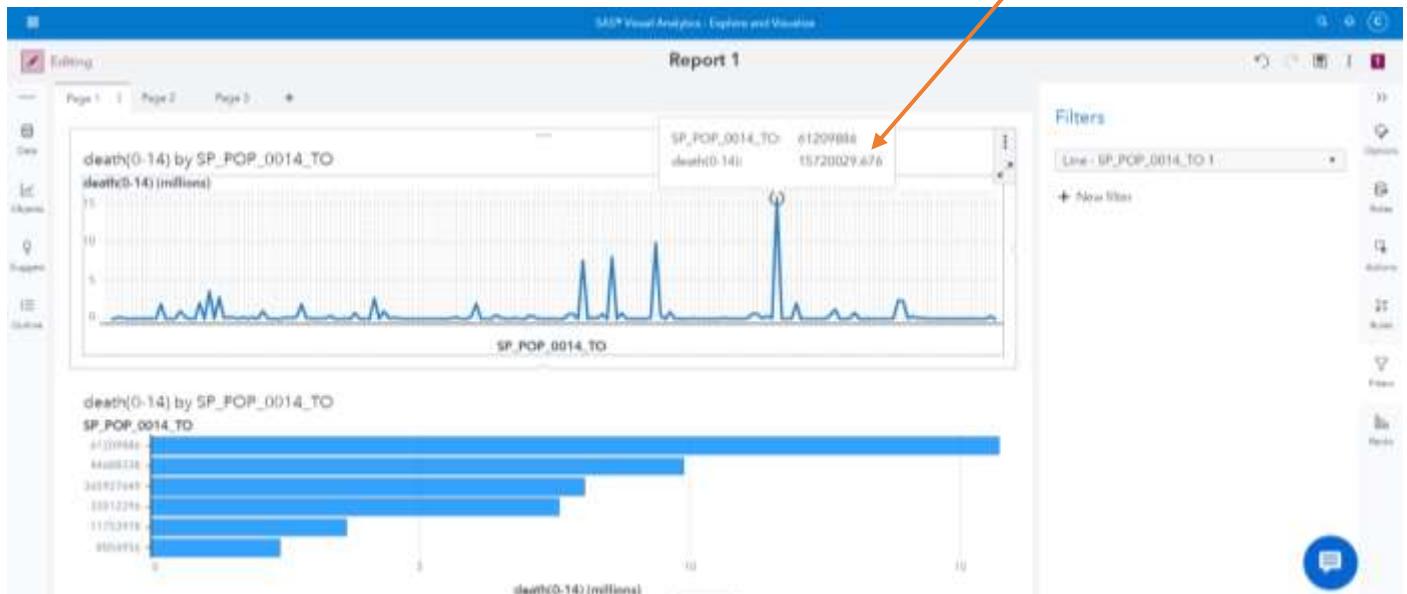


- second we will open our data source from data explorer manage data and import the dataset we just imported.
- Now we finished preparing our data by cleaning it and adding columns to do our visual analytics which we will do in the next steps.

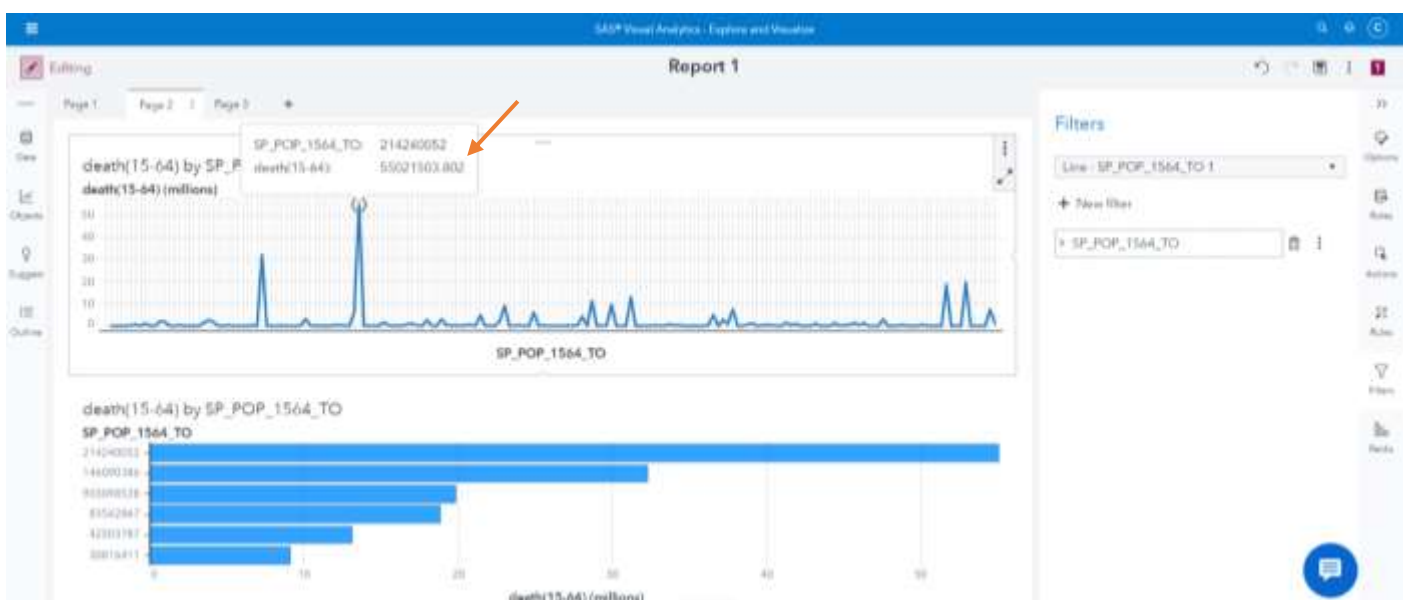
4. Sas visual analytics:

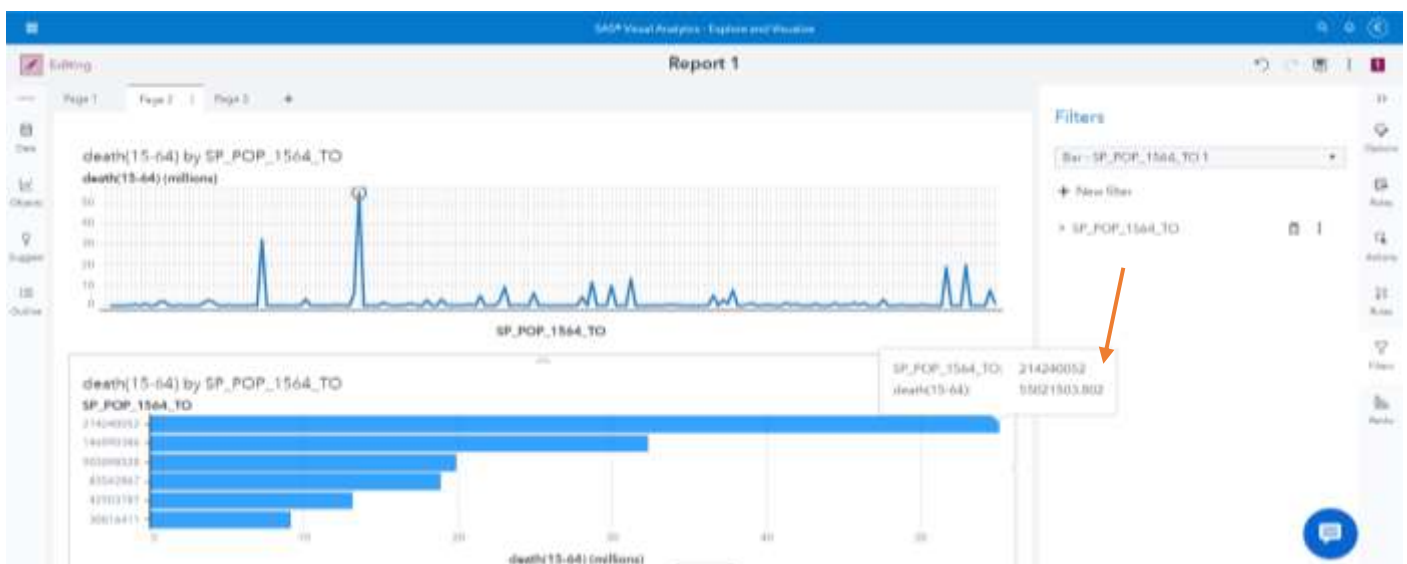


- First we will do line chart between number of death of each age range and population of each range we will see the numbers at the peaks which mean the highest population with highest number of death and then remove the other numbers (filter) to reduce complexity of the graph and draw the bar plot.
- By seeing the highest number in line char or bar chart we can determine the highest number of death of the three different age range.



- as we can see the total number of death for the highest number of group in age range from 0 to 14 is 15,720,029 people who died cause of coronavirus in this group of people in this range.





- as we can see the total number of death for the highest number of group in age range from 15 to 65 is 55,021,503 people who died cause of coronavirus in this group of people in this range.



- as we can see the total number of death for the highest number of group in age range from 65 to above is 13,282,177 people who died cause of coronavirus in this group of people in this range

5. conclusion:

- the highest number of death cause of coronavirus was in the age range of 15 to 64 with total number of death 55,021,503. The next patient(victim) probably will be in that range too.