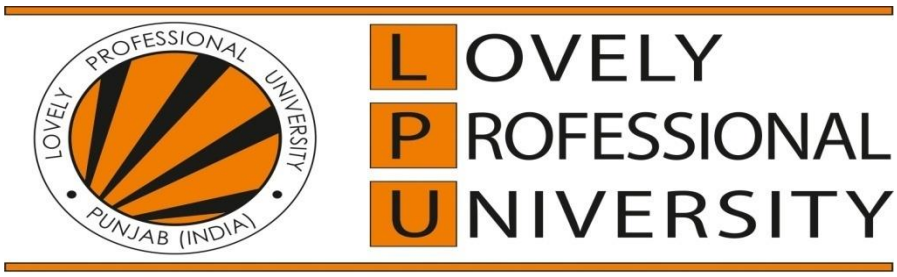
2 D

DATA MANAGEMENT PROJECT REPORT



ON

COUNTRY's GDP DASHBOARD

Submitted By:

ADITYA SAHU

11702583

RKM062B48

Under the guidance of:

Mrs. Savleen Kaur

School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Aditya Sahu** bearing Registration number **11702583** has completed **Data Management (INT217)** project titled, **"Country's GDP Project"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor
School of Computer Science and Engineering
Lovely Professional University
Date:

DECLARATION

I, **Aditya Sahu**, student of **Computer Science and Engineering** under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No. 11702583 Aditya Sahu

ACKNOWLEDGEMENT

I take this opportunity to present our votes of thanks to all those people who really acted as lightening pillars to enlighten my way throughout this Project that has led to successful and satisfactory completion of this Project. I am really grateful to **Lovely Professional University** for providing us with an opportunity to undertake this Project and providing us with all the facilities. I am highly thankful to All for their active support, valuable time and advice, whole-hearted guidance, sincere cooperation and painstaking involvement during the project and in completing the assignment of preparing the said project within the time stipulated. Lastly, I am thankful to all those, particularly the various friends, who have been instrumental in creating proper, healthy and conductive environment and including new and fresh innovative ideas for me during the project, without their help, it would have been extremely difficult for me to complete the project in a time bound framework.

INDEX

1. Introduction……1
    a. About Data Analysis
    b. Background of the project
    c. Data Definition

2. Scope of Analysis……...2
    a. Objectives of the project
    b. Scope

3. ETL Process……3
    a. What is ETL?
    b. ETL on the dataset

4. Analysis Of Dataset……4

5. Slicers……..…..14

6. Analyzing Results……15
    a. Dashboard I
    b. Dashboard II

7. References……17

8. Bibliography……17

INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

Using Country's GDP dataset, we want to design a Summery dashboard to analyze the countries based on various categories like gdp in a particular year, per capita, gender, suicides generations. Next, we want to analyses top 10 countries to in given above categories.

The Country's GDP data has the following data fields:

- Country
- Year
- Sex
- Age
- Suicides number
- Population
- Suicide/100k population
- GDP for year
- GDP per Capita
- Generation

SCOPE OF ANALYSIS

The analysis is done to produce useful information out of the datasheet. It will give a visualization of the analysis by graphs and chart which are interactive. The Analytics team wishes to answer the following objectives:

1. Population Age wise
2. Top 10 Highest gdp countries
3. Top 10 Highest per capita countries
4. Count of generation country wise
5. Suicide rate trends
6. Suicide per 100k pop
7. Per Capita Per Generation
8. Sum of suicides/100k pop by age
9. Sex wise suicide rate
10. Population of Generations

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively.

ETL PROCESS

ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL full-form is Extract, Transform and Load. Before ETL, the dataset looked like this. **This data was taken from Kaggle.**

Country	Year	Sex	Age	Sucieds No.	Population	suicides/100k pop	country-year	HDI for year	gdp_for_year ($)	gdp_per_capita ($)	Generation
Albania	1987	male	15-24 years	21	312900	6.71	Albania1987		2,15,66,24,900	796	Generation X
Albania	1987	male	35-54 years	16	308000	5.19	Albania1987		2,15,66,24,900	796	Silent
Albania	1987	female	15-24 years	14	289700	4.83	Albania1987		2,15,66,24,900	796	Generation X
Albania	1987	male	75+ years	1	21800	4.59	Albania1987		2,15,66,24,900	796	G.I. Generation
Albania	1987	male	25-34 years	9	274300	3.28	Albania1987		2,15,66,24,900	796	Boomers
Albania	1987	female	75+ years	1	35600	2.81	Albania1987		2,15,66,24,900	796	G.I. Generation
Albania	1987	female	35-54 years	6	278800	2.15	Albania1987		2,15,66,24,900	796	Silent
Albania	1987	female	25-34 years	4	257200	1.56	Albania1987		2,15,66,24,900	796	Boomers
Albania	1987	male	55-74 years	1	137500	0.73	Albania1987		2,15,66,24,900	796	G.I. Generation
Albania	1987	female	5-14 years	0	311000	0	Albania1987		2,15,66,24,900	796	Generation X
Albania	1987	female	55-74 years	0	144600	0	Albania1987		2,15,66,24,900	796	G.I. Generation
Albania	1987	male	5-14 years	0	338200	0	Albania1987		2,15,66,24,900	796	Generation X
Albania	1988	female	75+ years	2	36400	5.49	Albania1988		2,12,60,00,000	769	G.I. Generation
Albania	1988	male	15-24 years	17	319200	5.33	Albania1988		2,12,60,00,000	769	Generation X
Albania	1988	male	75+ years	1	22300	4.48	Albania1988		2,12,60,00,000	769	G.I. Generation
Albania	1988	male	35-54 years	14	314100	4.46	Albania1988		2,12,60,00,000	769	Silent
Albania	1988	male	55-74 years	4	140200	2.85	Albania1988		2,12,60,00,000	769	G.I. Generation
Albania	1988	female	15-24 years	8	295600	2.71	Albania1988		2,12,60,00,000	769	Generation X
Albania	1988	female	55-74 years	3	147500	2.03	Albania1988		2,12,60,00,000	769	G.I. Generation
Albania	1988	female	25-34 years	5	262400	1.91	Albania1988		2,12,60,00,000	769	Boomers
Albania	1988	male	25-34 years	5	279900	1.79	Albania1988		2,12,60,00,000	769	Boomers
Albania	1988	female	35-54 years	4	284500	1.41	Albania1988		2,12,60,00,000	769	Silent
Albania	1988	female	5-14 years	0	317200	0	Albania1988		2,12,60,00,000	769	Generation X
Albania	1988	male	5-14 years	0	345000	0	Albania1988		2,12,60,00,000	769	Generation X
Albania	1989	male	75+ years	2	22500	8.89	Albania1989		2,33,51,24,988	833	G.I. Generation
Albania	1989	male	25-34 years	18	283600	6.35	Albania1989		2,33,51,24,988	833	Boomers
Albania	1989	male	35-54 years	15	318400	4.71	Albania1989		2,33,51,24,988	833	Silent
Albania	1989	male	55-74 years	6	142100	4.22	Albania1989		2,33,51,24,988	833	G.I. Generation
Albania	1989	male	15-24 years	12	323500	3.71	Albania1989		2,33,51,24,988	833	Generation X
Albania	1989	female	35-54 years	7	288600	2.43	Albania1989		2,33,51,24,988	833	Silent
Albania	1989	female	15-24 years	5	299900	1.67	Albania1989		2,33,51,24,988	833	Generation X
Albania	1989	female	25-34 years	2	266300	0.75	Albania1989		2,33,51,24,988	833	Boomers
Albania	1989	female	55-74 years	1	149600	0.67	Albania1989		2,33,51,24,988	833	G.I. Generation
Albania	1989	female	5-14 years	0	321900	0	Albania1989		2,33,51,24,988	833	Generation X
Albania	1989	female	75+ years	0	37000	0	Albania1989		2,33,51,24,988	833	G.I. Generation

In my case datasheet is perfect so there is no need to clean the datasheet.

Just one column is need to delete because there is no use of that column in the analysis.

Country	Year	Sex	Age	Sucieds No.	Population	suicides/100k pop	country-year	gdp_for_year ($)	gdp_per_capita ($)	Generation
Albania	1987	male	15-24 years	21	312900	6.71	Albania1987	2,15,66,24,900	796	Generation X
Albania	1987	male	35-54 years	16	308000	5.19	Albania1987	2,15,66,24,900	796	Silent
Albania	1987	female	15-24 years	14	289700	4.83	Albania1987	2,15,66,24,900	796	Generation X
Albania	1987	male	75+ years	1	21800	4.59	Albania1987	2,15,66,24,900	796	G.I. Generation
Albania	1987	male	25-34 years	9	274300	3.28	Albania1987	2,15,66,24,900	796	Boomers
Albania	1987	female	75+ years	1	35600	2.81	Albania1987	2,15,66,24,900	796	G.I. Generation
Albania	1987	female	35-54 years	6	278800	2.15	Albania1987	2,15,66,24,900	796	Silent
Albania	1987	female	25-34 years	4	257200	1.56	Albania1987	2,15,66,24,900	796	Boomers
Albania	1987	male	55-74 years	1	137500	0.73	Albania1987	2,15,66,24,900	796	G.I. Generation
Albania	1987	female	5-14 years	0	311000	0	Albania1987	2,15,66,24,900	796	Generation X
Albania	1987	female	55-74 years	0	144600	0	Albania1987	2,15,66,24,900	796	G.I. Generation
Albania	1987	male	5-14 years	0	338200	0	Albania1987	2,15,66,24,900	796	Generation X
Albania	1988	female	75+ years	2	36400	5.49	Albania1988	2,12,60,00,000	769	G.I. Generation
Albania	1988	male	15-24 years	17	319200	5.33	Albania1988	2,12,60,00,000	769	Generation X
Albania	1988	male	75+ years	1	22300	4.48	Albania1988	2,12,60,00,000	769	G.I. Generation
Albania	1988	male	35-54 years	14	314100	4.46	Albania1988	2,12,60,00,000	769	Silent
Albania	1988	male	55-74 years	4	140200	2.85	Albania1988	2,12,60,00,000	769	G.I. Generation
Albania	1988	female	15-24 years	8	295600	2.71	Albania1988	2,12,60,00,000	769	Generation X
Albania	1988	female	55-74 years	3	147500	2.03	Albania1988	2,12,60,00,000	769	G.I. Generation
Albania	1988	female	25-34 years	5	262400	1.91	Albania1988	2,12,60,00,000	769	Boomers
Albania	1988	male	25-34 years	5	279900	1.79	Albania1988	2,12,60,00,000	769	Boomers
Albania	1988	female	35-54 years	4	284500	1.41	Albania1988	2,12,60,00,000	769	Silent
Albania	1988	female	5-14 years	0	317200	0	Albania1988	2,12,60,00,000	769	Generation X
Albania	1988	male	5-14 years	0	345000	0	Albania1988	2,12,60,00,000	769	Generation X
Albania	1989	male	75+ years	2	22500	8.89	Albania1989	2,33,51,24,988	833	G.I. Generation
Albania	1989	male	25-34 years	18	283600	6.35	Albania1989	2,33,51,24,988	833	Boomers
Albania	1989	male	35-54 years	15	318400	4.71	Albania1989	2,33,51,24,988	833	Silent
Albania	1989	male	55-74 years	6	142100	4.22	Albania1989	2,33,51,24,988	833	G.I. Generation
Albania	1989	male	15-24 years	12	323500	3.71	Albania1989	2,33,51,24,988	833	Generation X
Albania	1989	female	35-54 years	7	288600	2.43	Albania1989	2,33,51,24,988	833	Silent
Albania	1989	female	15-24 years	5	299900	1.67	Albania1989	2,33,51,24,988	833	Generation X
Albania	1989	female	25-34 years	2	266300	0.75	Albania1989	2,33,51,24,988	833	Boomers
Albania	1989	female	55-74 years	1	149600	0.67	Albania1989	2,33,51,24,988	833	G.I. Generation
Albania	1989	female	5-14 years	0	321900	0	Albania1989	2,33,51,24,988	833	Generation X
Albania	1989	female	75+ years	0	37000	0	Albania1989	2,33,51,24,988	833	G.I. Generation
Albania	1989	male	5-14 years	0	349700	0	Albania1989	2,33,51,24,988	833	Generation X

Proper Data Formatting

Without proper Data Formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format.

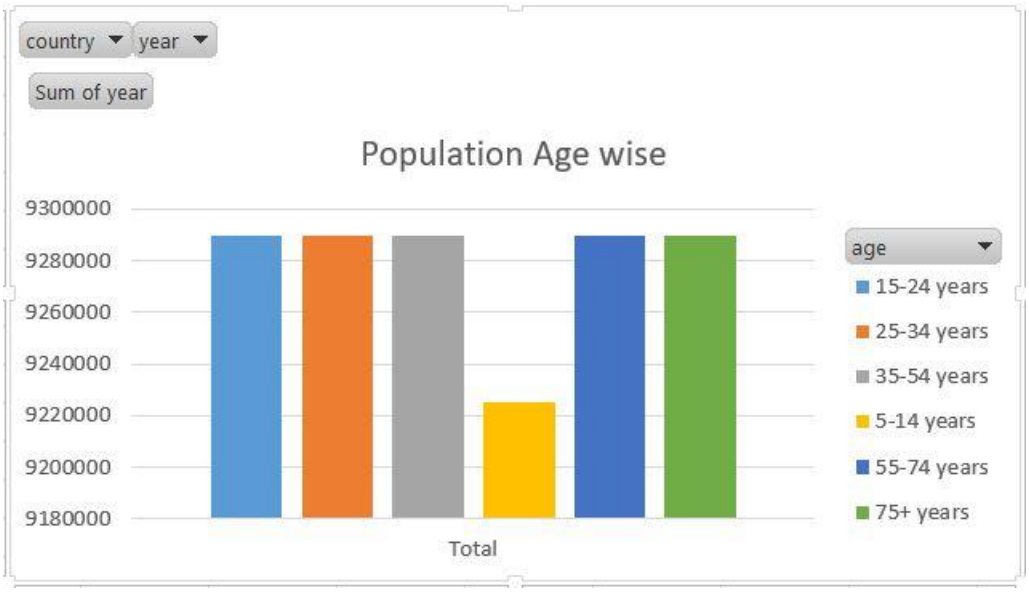
ANALYSIS OF DATASET

1. Population Age wise

Country, Age, Year are column are used to determine the data. In this country and year are put in the filter so that we can view it in the more descriptive form.

country	(All)						
year	(All)						
	Column Labels						
	15-24 years	25-34 years	35-54 years	5-14 years	55-74 years	75+ years	Grand Total
Sum of year	9289920	9289920	9289920	9225408	9289920	9289920	55675008

Visualization

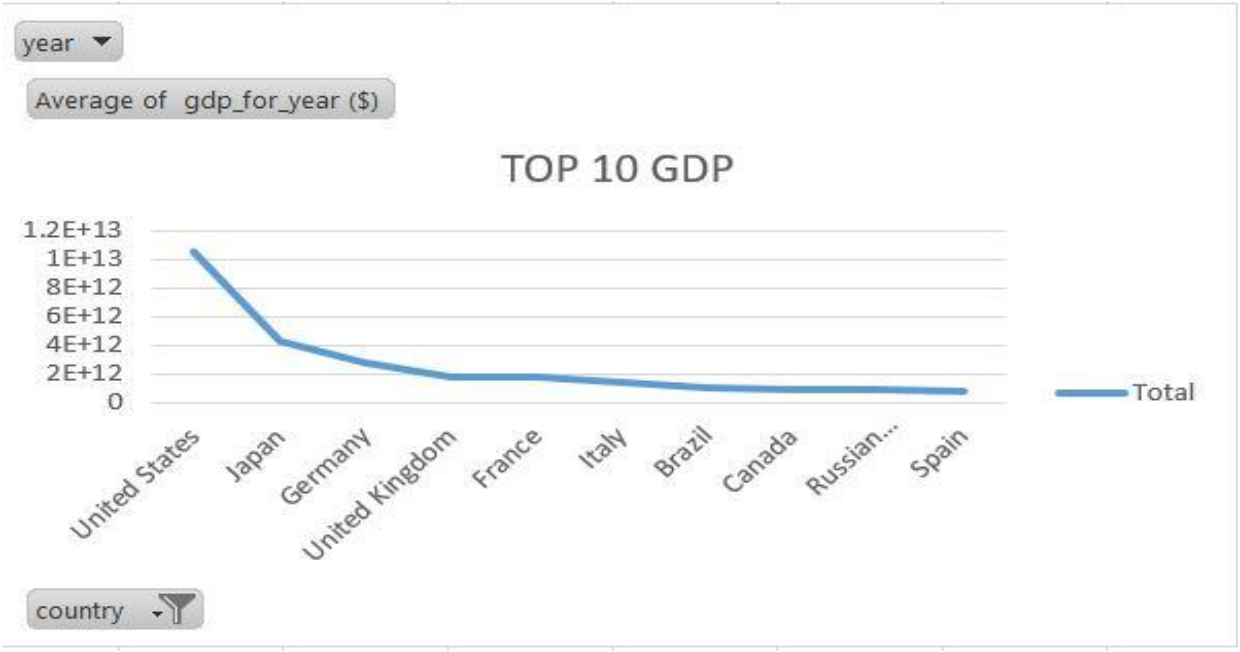


2. Top 10 Highest gdp countries

Country, Year, Gdp for year is used to manipulate data. After that the graph is done into Top 10 highest Gdp countries. This is Line graph.

year	(All)
Row Labels	**Average of gdp_for_year ($)**
United States	1.05107E+13
Japan	4.33922E+12
Germany	2.74223E+12
United Kingdom	1.81607E+12
France	1.78119E+12
Italy	1.48175E+12
Brazil	1.02256E+12
Canada	9.13188E+11
Russian Federation	8.84323E+11
Spain	8.56568E+11
Grand Total	**2.67089E+12**

Visualization

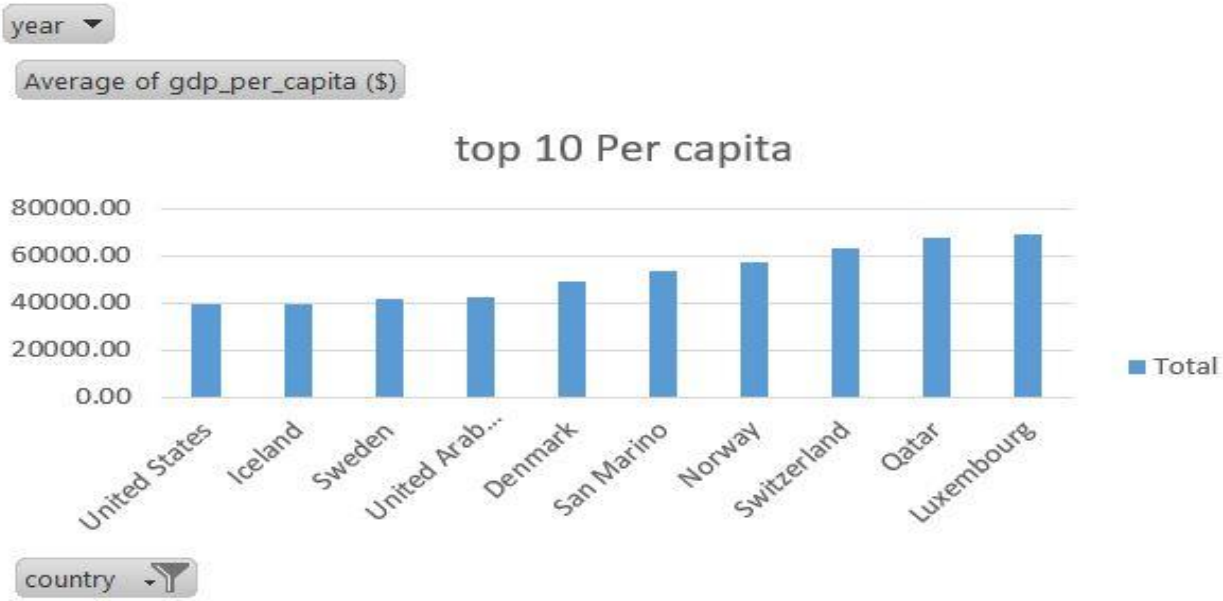


3. Top 10 Highest per capita countries

In this the graph is showing the top 10 countries with highest per capita.

year	(All)
Row Labels	**Average of gdp_per_capita ($)**
United States	39269.61
Iceland	39274.75
Sweden	41357.58
United Arab Emirates	42162.00
Denmark	49299.91
San Marino	53663.67
Norway	57319.60
Switzerland	62981.76
Qatar	67756.45
Luxembourg	68798.39
Grand Total	**51610.02**

Visualization

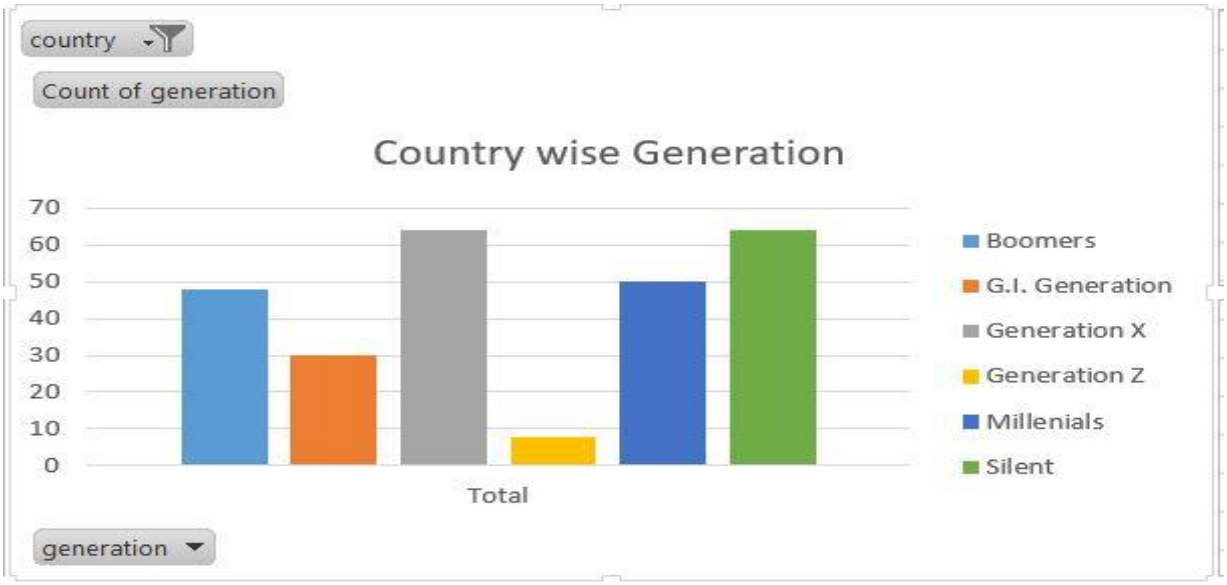


4. **Count of generation country wise**

In this you will get the total of all the types of generations and also you can make a filter of a country to check the particular generation.

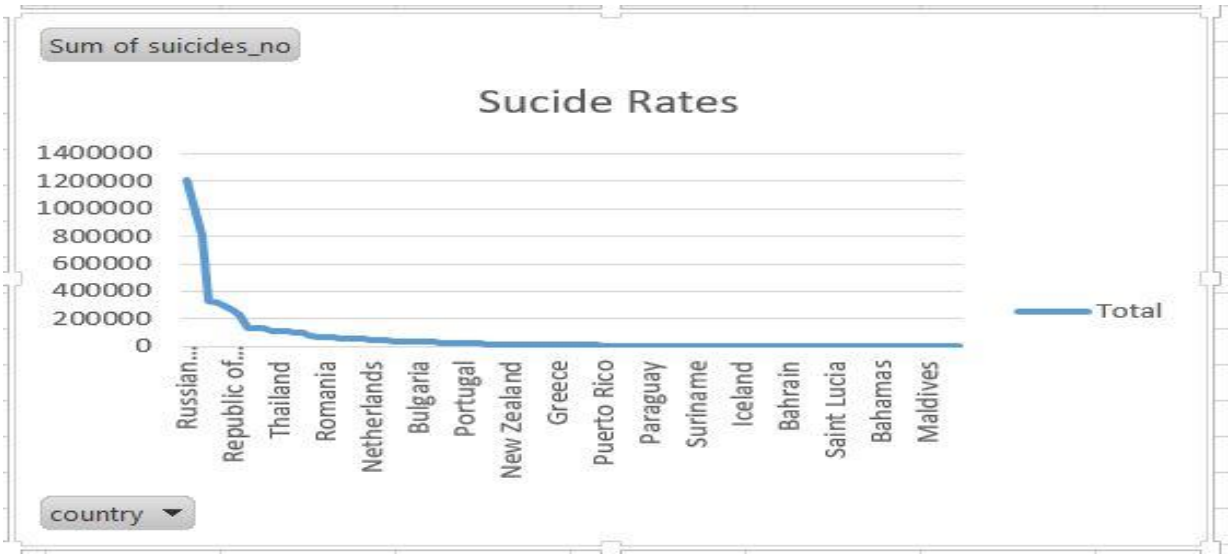
country	Albania						
	Column Labels						
	Boomers	G.I. Generation	Generation X	Generation Z	Millenials	Silent	Grand Total
Count of generation	48	30	64	8	50	64	264

Visualization



5. Suicide rate trends

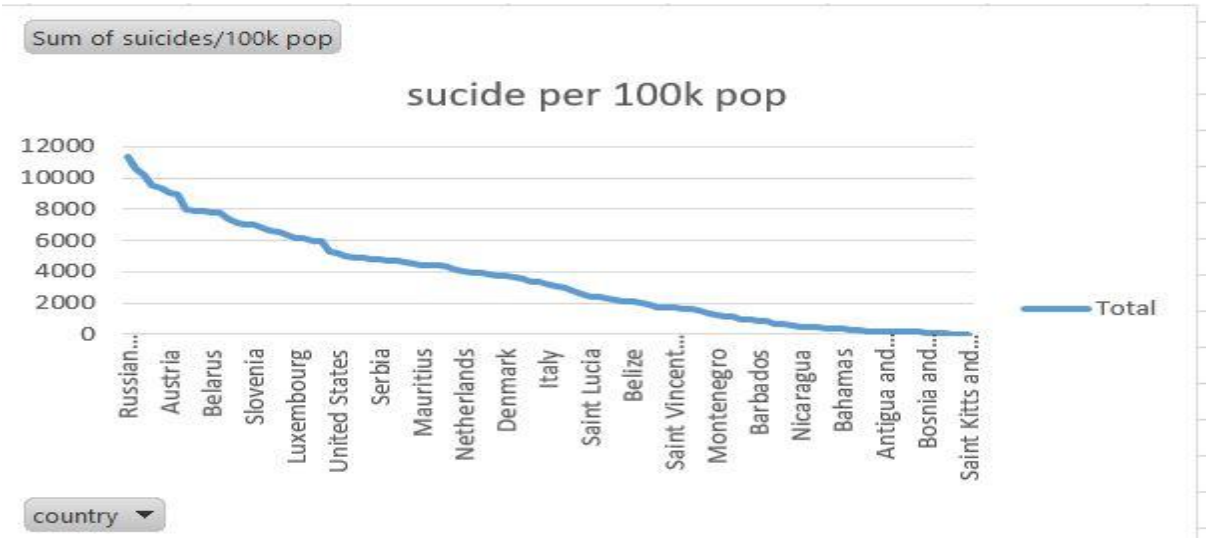
This graph is the suicides trends over the year and which country is having the highest rate of suicides.



Visualization

Row Labels	Sum of suicides_no
Russian Federation	1209742
United States	1034013
Japan	806902
France	329127
Ukraine	319950
Germany	291262
Republic of Korea	261730
Brazil	226613
Poland	139098
United Kingdom	136805
Italy	132060
Mexico	111139
Thailand	110643
Canada	107561
Kazakhstan	101546
Spain	100202
Argentina	82219
Hungary	73891
Romania	72777
Australia	70111
Belgium	62761
Belarus	59892
Sri Lanka	55641
Colombia	53080
Netherlands	50833

6. Suicide per 100k pop



Visualization

Row Labels	Sum of suicides/100k pop
Russian Federation	11305.13
Lithuania	10588.88
Hungary	10156.07
Kazakhstan	9519.52
Republic of Korea	9350.45
Austria	9076.23
Ukraine	8931.66
Japan	8025.23
Finland	7924.11
Belgium	7900.5
Belarus	7831.13
France	7803.25
Latvia	7373.35
Suriname	7162.32
Bulgaria	7016.08
Slovenia	7012.62
Estonia	6873.78
Guyana	6655.92
Uruguay	6538.96
Singapore	6340.98
Luxembourg	6156.56
Cuba	6111.95
Croatia	5982.84
Czech Republic	5952.99
Sweden	5247.72
United States	5140.97

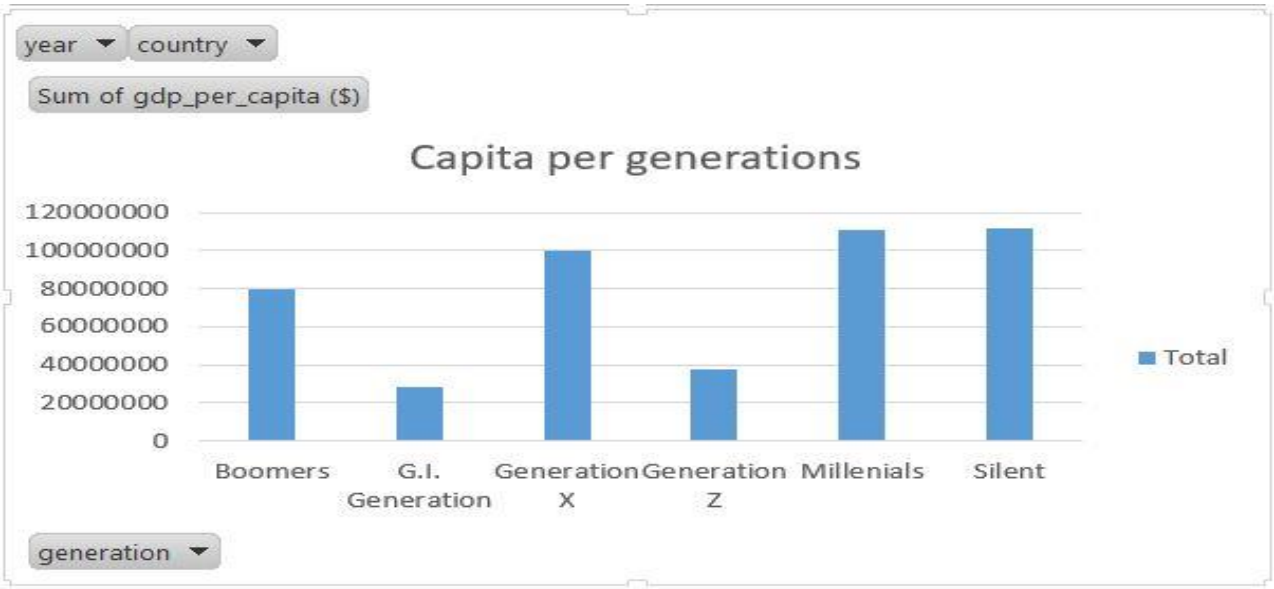
7. Per Capita Per Generation

This graph is between Generations and per capita and filters between year and country.

year	(All)
country	(All)

Row Labels	Sum of gdp_per_capita ($)
Boomers	79750168
G.I. Generation	28519510
Generation X	99996594
Generation Z	37678558
Millenials	111331886
Silent	111948324
Grand Total	**469225040**

Visualization

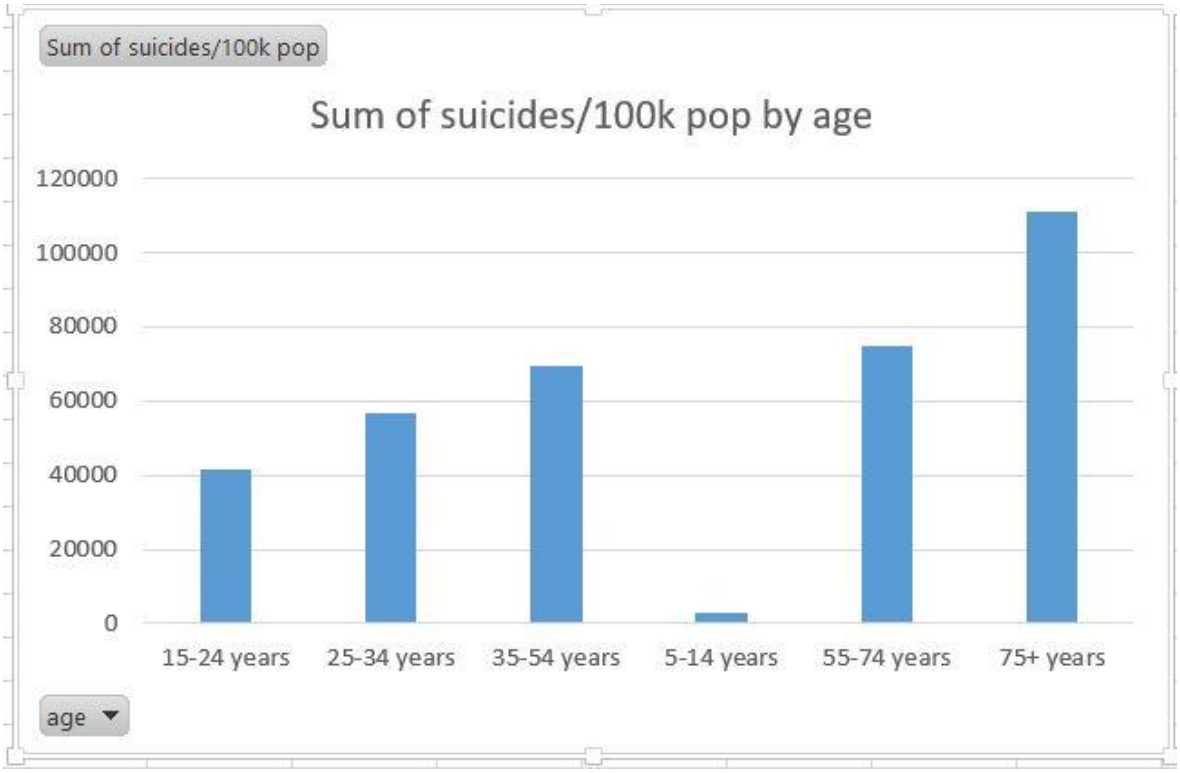


8. Sum of suicides/100k pop by age

The graph is between sums of suicide/100k population by age.

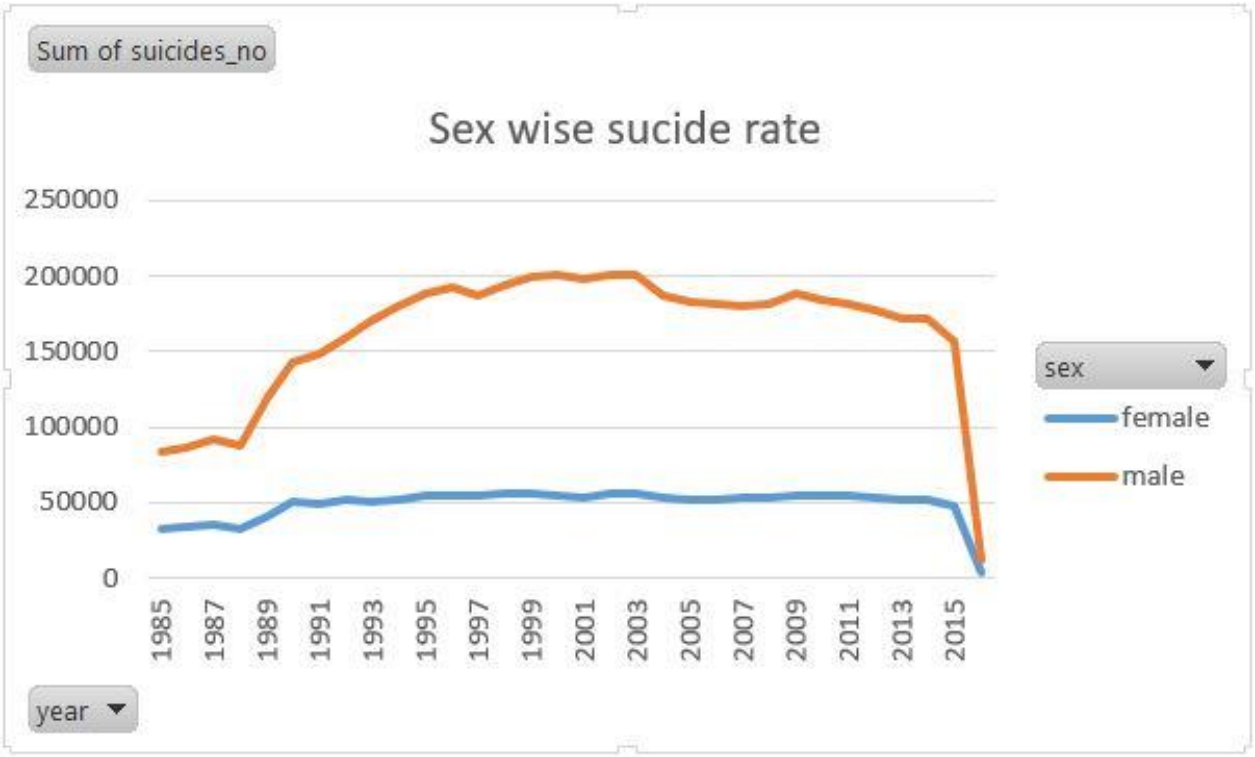
age	Sum of suicides/100k pop
15-24 years	41532.69
25-34 years	56571.52
35-54 years	69386.02
5-14 years	2858.39
55-74 years	74994.2
75+ years	111201.01

Visualization



9. Sex wise suicide rate

The graph is about sex wise suicide over the year



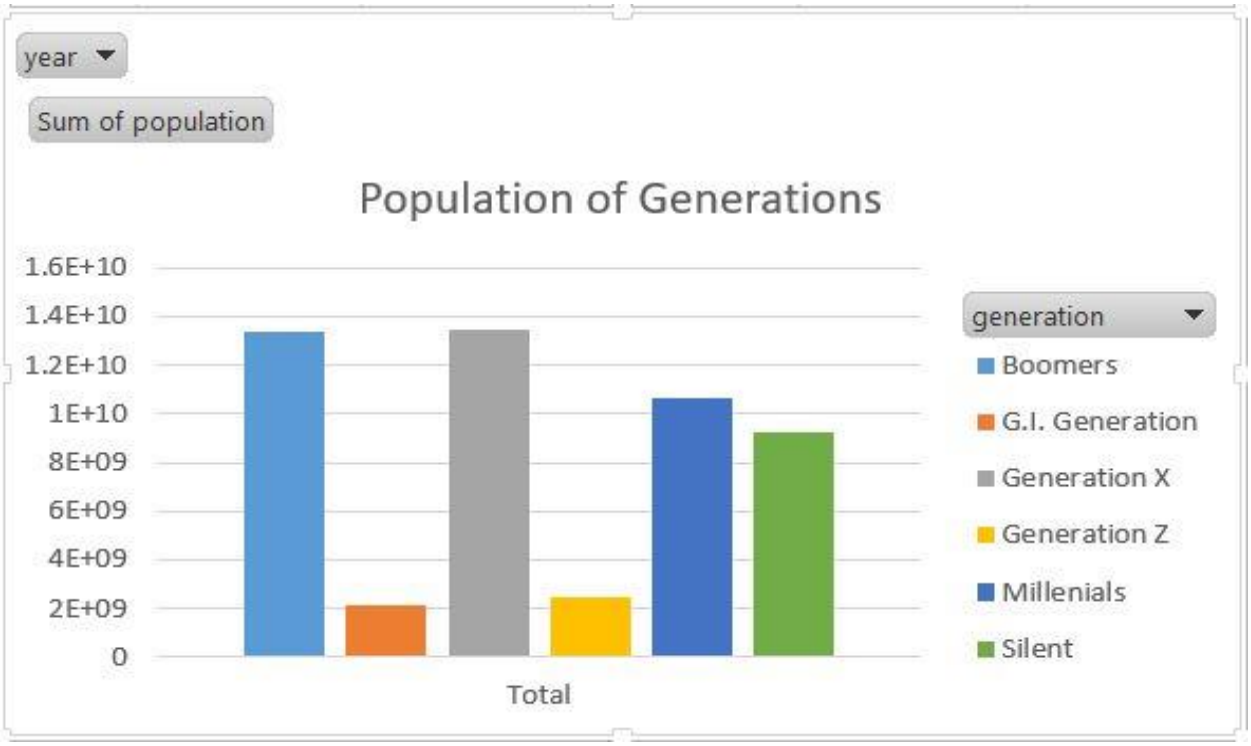
Sum of suicides_no	Column Labels		
Row Labels	female	male	Grand Total
1985	32479	83584	116063
1986	33852	86818	120670
1987	35006	91836	126842
1988	33015	88011	121026
1989	41361	118883	160244
1990	50118	143243	193361
1991	49622	148398	198020
1992	51567	159906	211473
1993	51331	170234	221565
1994	51532	180531	232063
1995	54504	189040	243544
1996	54583	192142	246725
1997	54126	186619	240745
1998	55631	193960	249591
1999	56215	199904	256119
2000	55254	200578	255832
2001	52999	197653	250652
2002	55549	200546	256095
2003	55627	200452	256079
2004	53232	187629	240861
2005	52035	182340	234375
2006	52039	181322	233361
2007	53324	180084	233408
2008	53973	181474	235447
2009	54920	188567	243487

10. Population of Generations

The graph is between sum of population and generation. Year is put in filter, we can check the population year wise.

year	(All)						
	Column Labels						
	Boomers	G.I. Generation	Generation X	Generation Z	Millenials	Silent	Grand Total
Sum of population	13350511729	2126202724	13472109292	2503541842	10649461202	9220331647	5132215843

Visualization



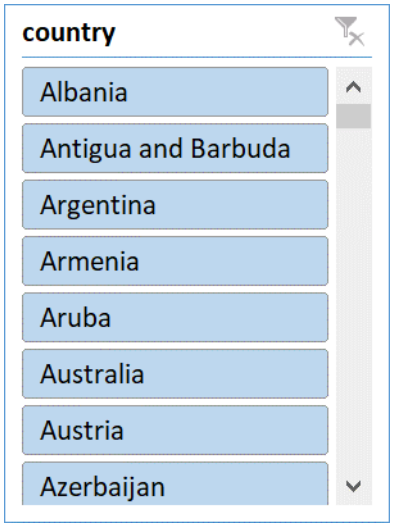
Slicers

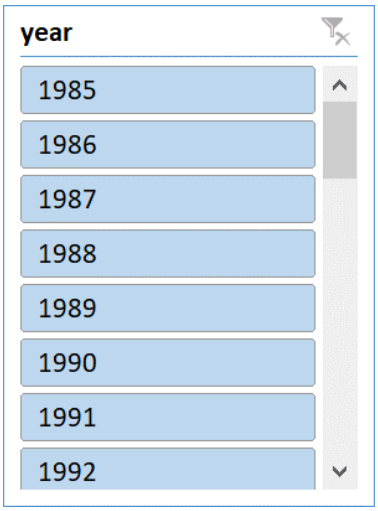
Slicers are *visual filters*. Using a slicer, you can filter your data (or pivot table, pivot chart) by clicking on the type of data you want.

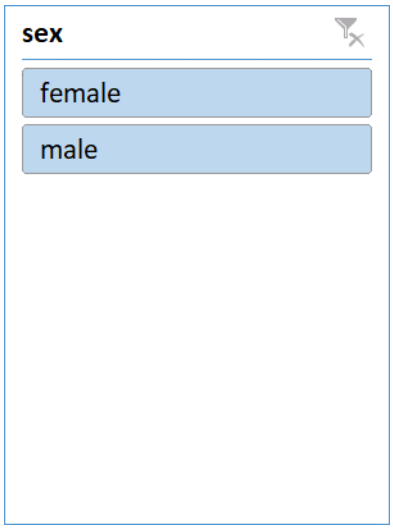
For example, let's say you are looking at sales by customer profession in a pivot report. And you want to see how the sales are for a particular region. There are 2 options for you do drill down to an individual region level.

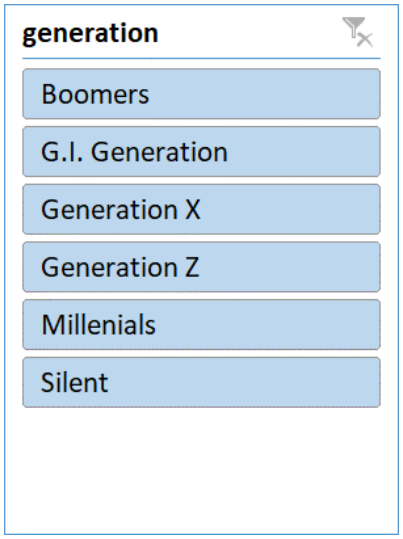
1. Add region as report filter and filter for the region you want.
2. Add a slicer on region and click on the region you want.

With a report filter (or any other filter), you will have to click several times to pick one store. **With slicers, it is a matter of simple click**.









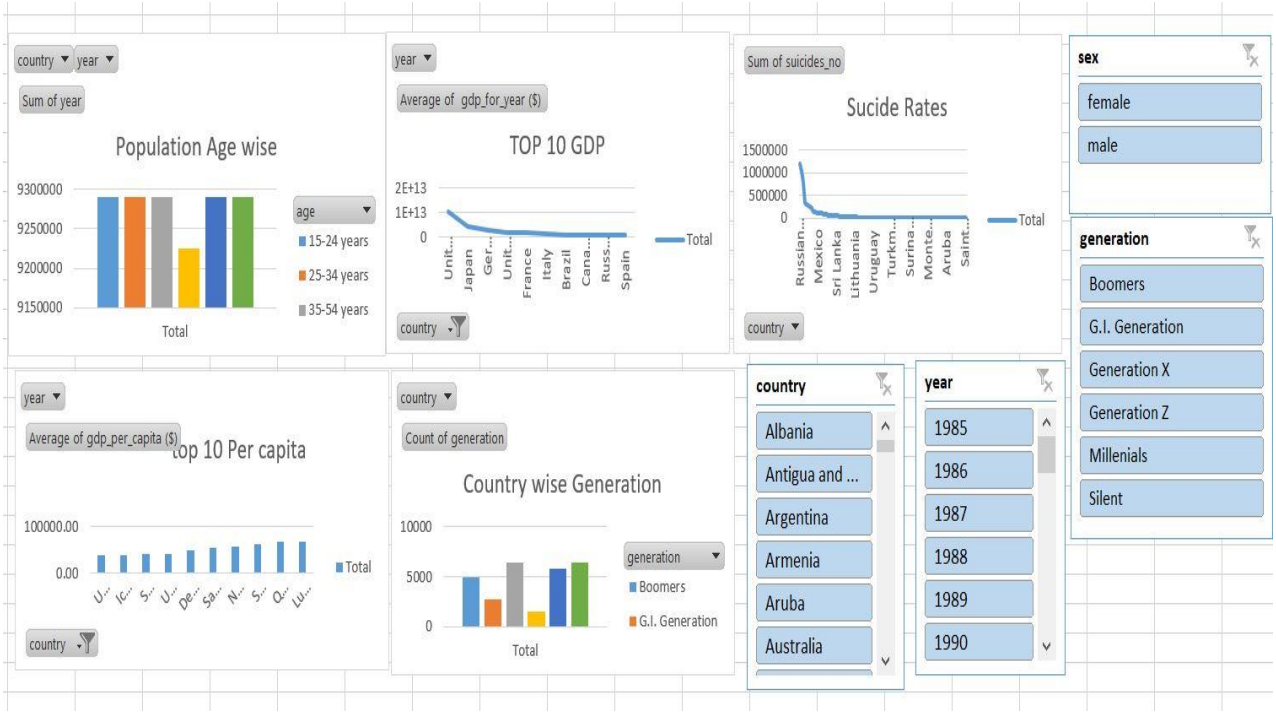
ANALYSIS RESULTS

Dashboard

An **Excel dashboard** is one pager (mostly, but not always necessary) that helps managers and business leaders in tracking key KPIs or metrics and take a decision based on it. It contains charts/tables/views that are backed by data. A **dashboard** is often called a report, however, not all reports are **dashboards**.

In this dashboard we did analysis by adding slicers, The whole **point** of the based **dashboard** is that it lets you visualize the Key Performance Indicators and other strategic data for your organization at a glance. It is the **dashboard** tool that presents management with the information for the practical.

Dashboard – 1



2 D

Dashboard – 2

REFERENCES AND BIBLIOGRAPHY

Wikipedia - https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita

Kaggle - https://www.kaggle.com/search?q=gdp+per+capita

Reddit - https://www.reddit.com/r/dataisbeautiful/comments/bdvazr/top_countries_by_gdp_per_capita_over_the_past_200/

Google - https://www.goskills.com/Excel/Articles/How-to-create-excel-dashboard

2 D