410821323

Vincent Evans Sebayang

Graphic Data of Population in Jakarta from 2016 to 2021 based on gender and age

1. Motivation

My motivation to doing this project to understand and improve my skill on understanding of data science by trying to process certain data and plot it to make it became a graphic that easily to understand how certain data is distributed

1. Purpose

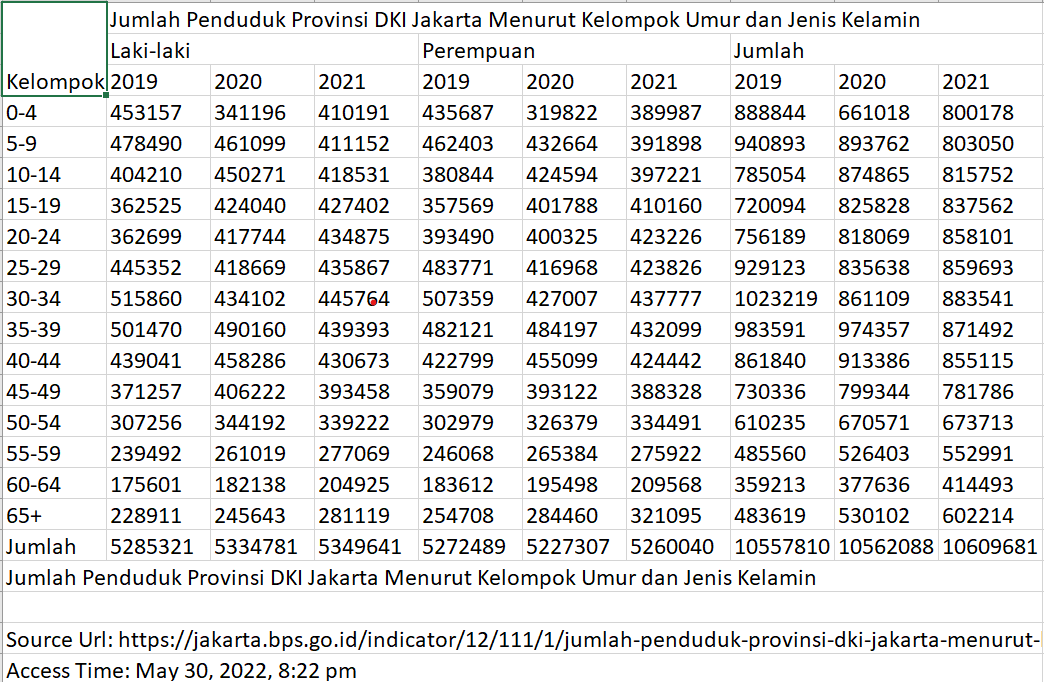
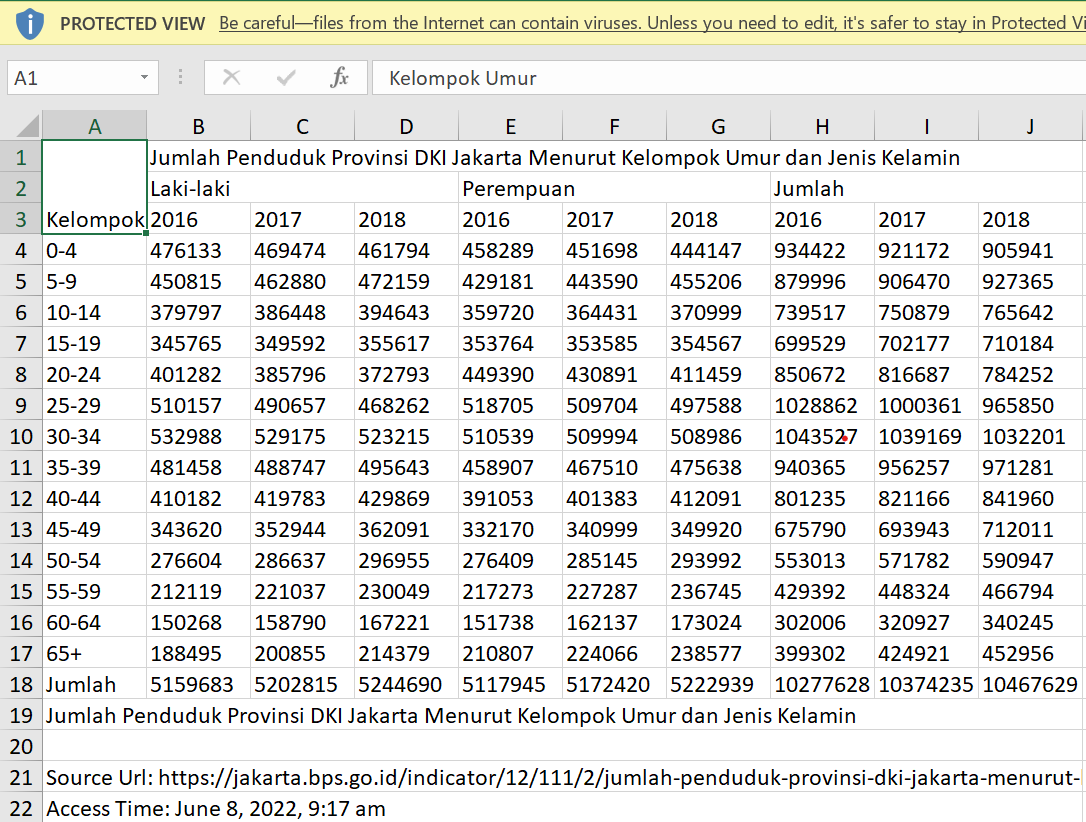
The reason why I pick this topic first of all because I came from Indonesia and when I was in Indonesia, I was living in Jakarta which is the capital city of Indonesia. Jakarta as capital city make a lot of people from all of Indonesia want to go to Jakarta to seek an opportunity to have a better life so in this project, I would like to try plot the data of Jakarta’s population that I get from central statistics agency (Badan Pusat Statistic) Jakarta and see how it is distributed based on age from 2016 to 2021

1. Relative Work

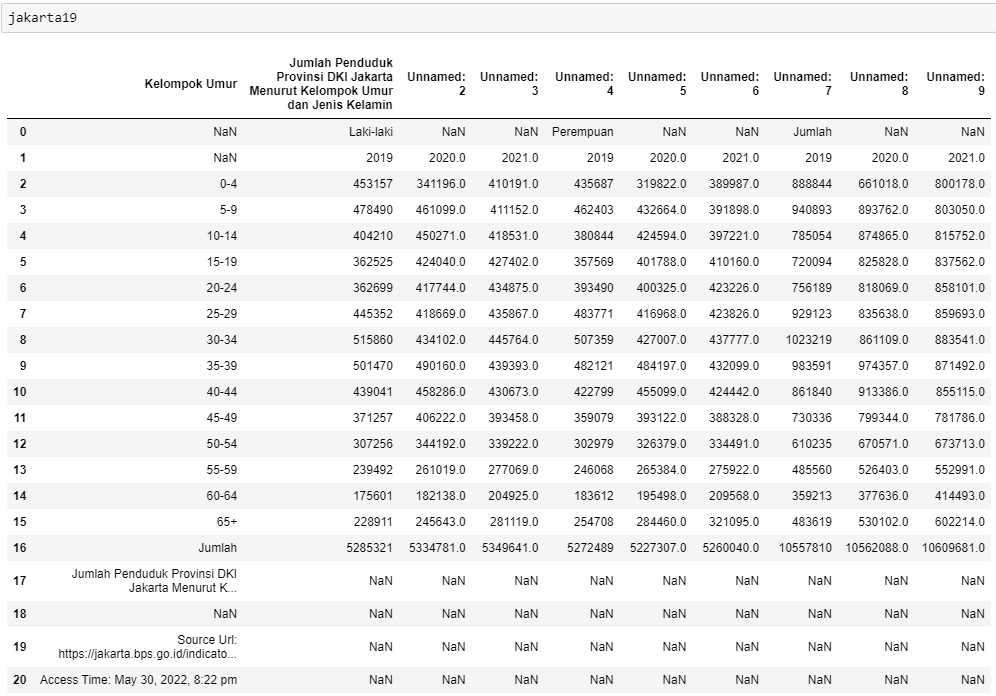
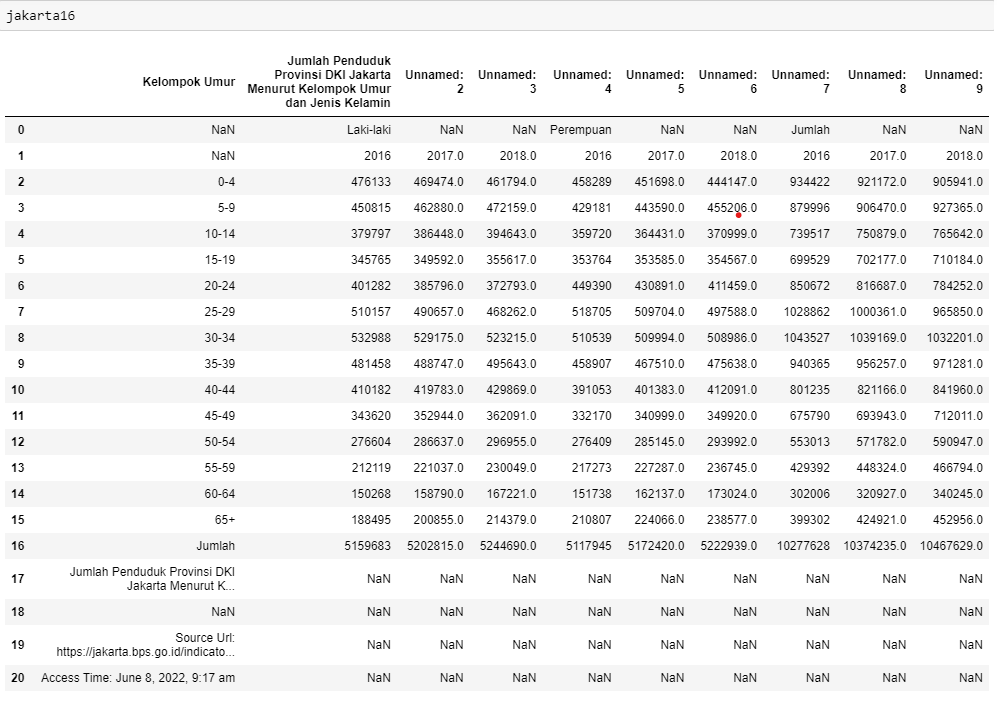
This type of activity is very important for every field to gather data and make it into a graph that easier and faster to understand and make a certain decision

1. System description

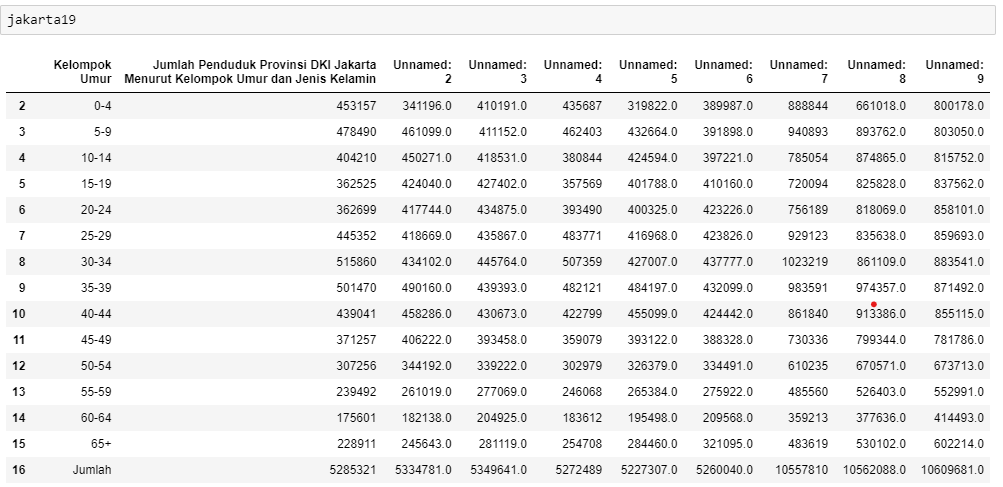
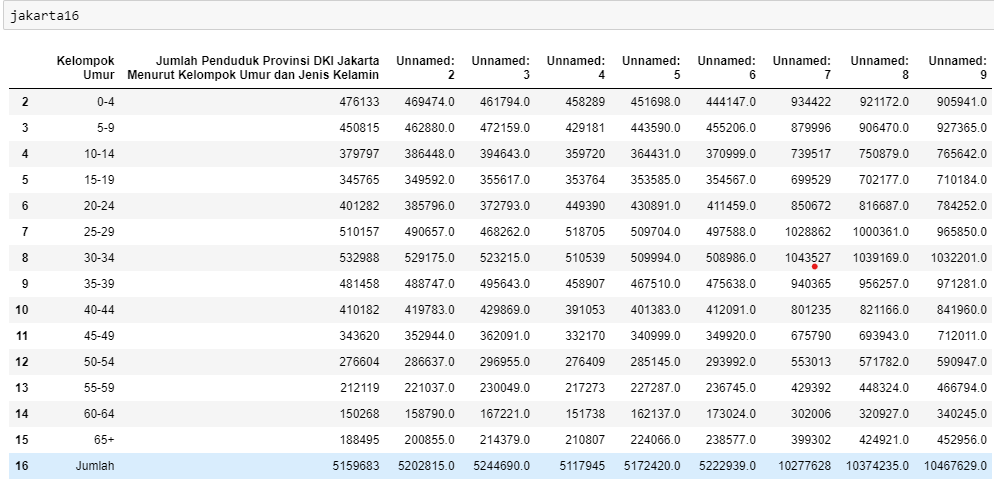
To processing data, I use python because it is the most popular and powerful tools because the ease syntax and it also has many other tools to help processing the data such as pandas which is very good to process the data or table that we would like to manipulate. To make the graph I use another tool which is matplotlib



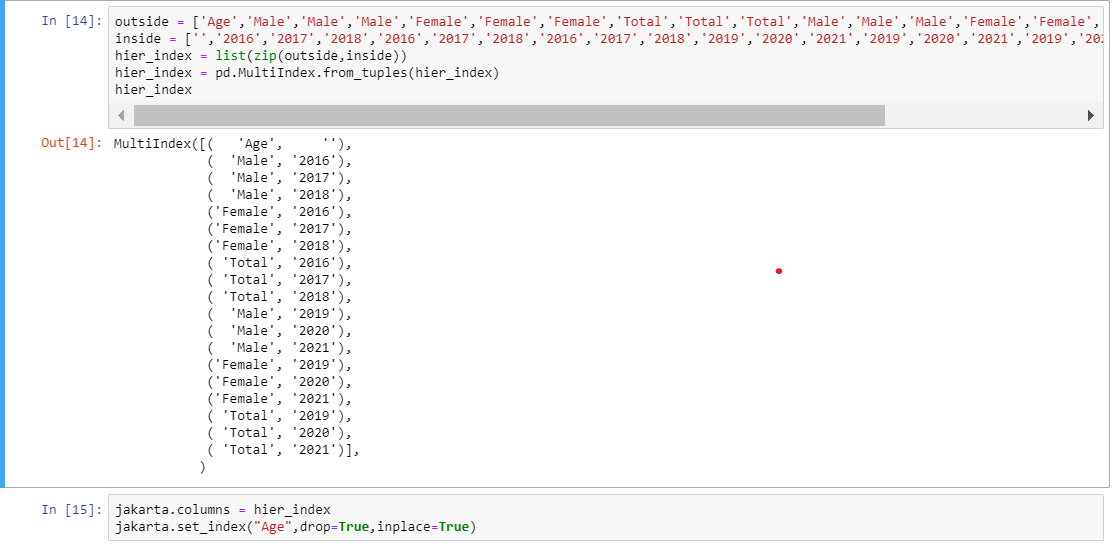
This is the data that I got from BPS Jakarta from 2016-2018 and 2019-2021, the word of Laki-laki means mans, Perempuan mean girls and for Jumlah means total.

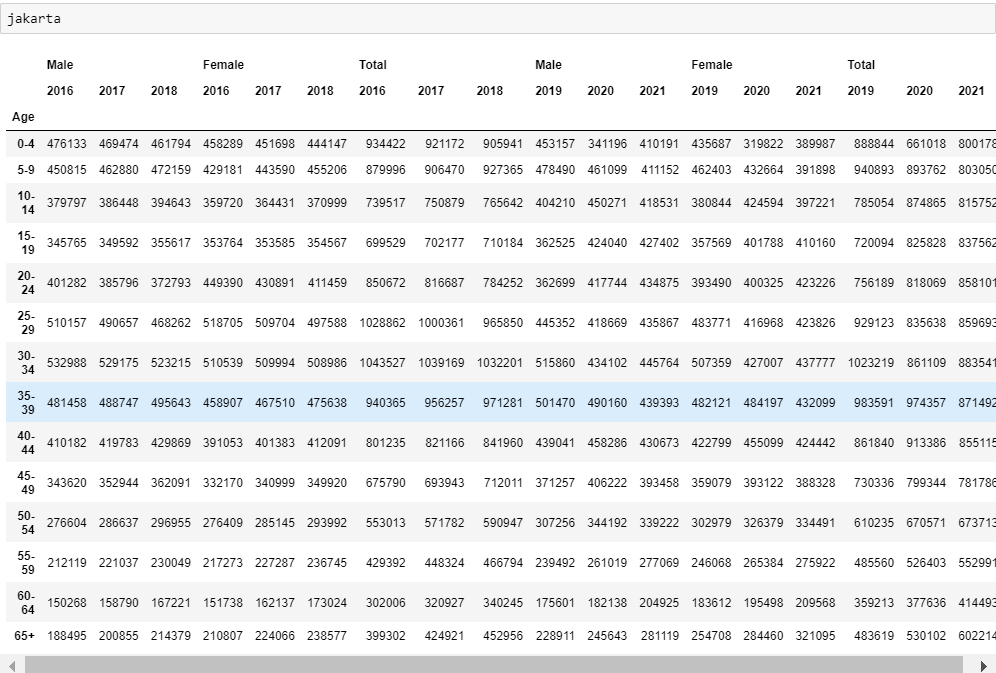


As we can see the data is still quite not in proper position and there are many NaN value which isn’t needed so I will first drop the NaN

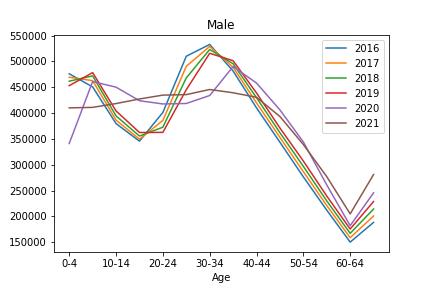
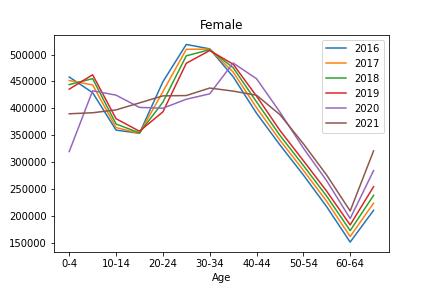


And this the table that I have after I drop the NaN but now the problem is that the names of the columns is still not proper so what I did first is merger these table and after that I fixed the name’s column but Before I directly fixed the columns, I prepared the columns first by making Age, Male, Female, Total as primary columns and the year become the second columns





And this is the table after columns name is fixed and I can plot the table base on every gender



1. Conclusion and Reflection

First of for the graph it self now it can be easily to understand that most of the majority people who live in Jakarta either male or female are 30 to 34 years old. The challenge that I faced was how to clean the data, merged the data since before it came from two different data and another challenge how to set the columns since there are columns for gender and the second one is for the year and make it the second columns is nested to the first columns