

NAME: MOHD HAFIZ AZIZI BIN MOHD RADZI

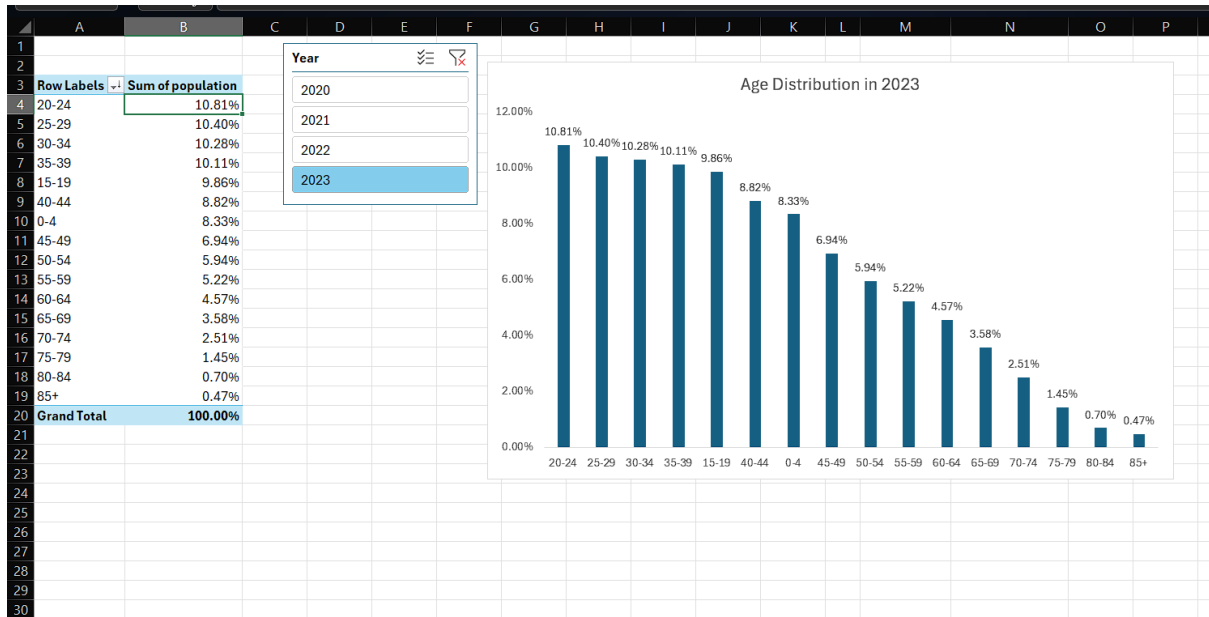
NO. IC: 960410125673

- The data is taken from Population Table: Administrative Districts.  
([https://open.dosm.gov.my/data-catalogue/population\\_district](https://open.dosm.gov.my/data-catalogue/population_district))
- The variable and definition in the table are as follows:

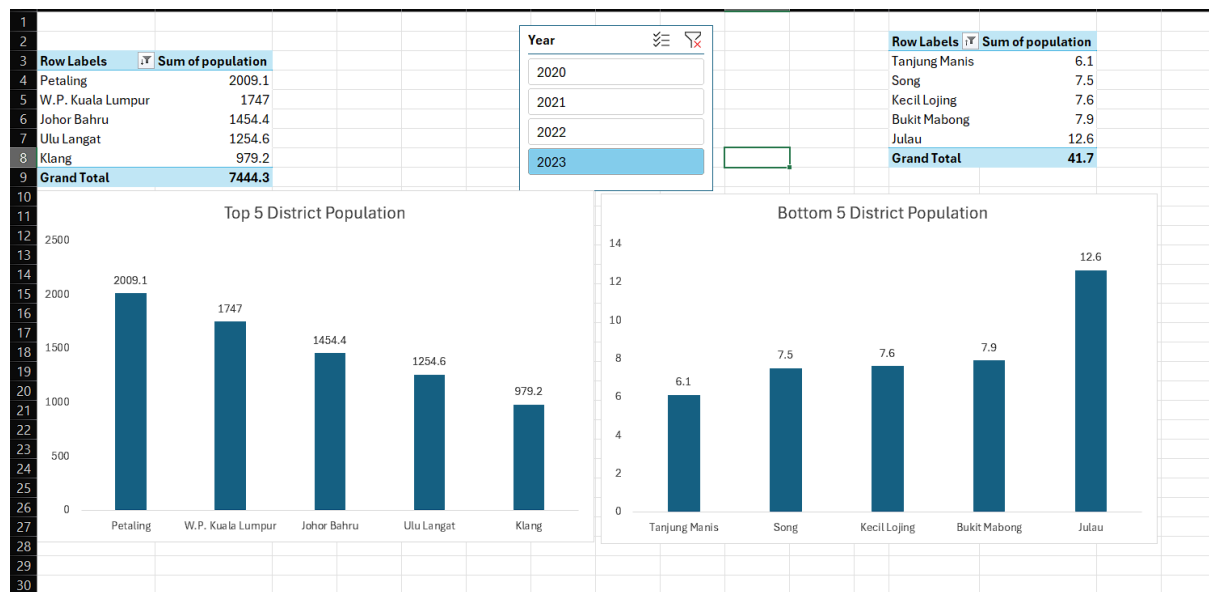
Name in dataset	Variable	Definition
date (Date)	Date	The date in YYYY-MM-DD format, with MM-DD set to 01-01 as this is annual data
state (Categorical)	State	One of 16 states
district (Categorical)	District	One of 160 administrative districts, including the 4 states (Perlis and the 3 Federal Territories) which do not have subdivisions
sex (Categorical)	Sex	Either both sexes ('both'), male ('male') or female ('female')
age (Categorical)	Age Group	Either all age groups ('overall') or five-year age groups e.g. 0-4, 5-9, 10-14, etc. 85+ is the oldest category.
ethnicity (Categorical)	Ethnicity	All ethnic groups ('overall'), Malay ('bumi_malay'), other Bumiputera ('bumi_other'), Chinese ('chinese'), Indian ('indian'), other citizens ('other_citizen'), or non-citizen residents ('other_noncitizen').
population (Float)	Population ('000)	Thousands of people ('000 people)

- Using Power Query in Excel to do data cleaning, name of the worksheet is Table1,
  - Rename the 'date' column to 'Year' and show only the year value (2020,2021,2022,2023),
  - renaming 'population' column to 'population ('000)', which is population of 1000 people,
  - remove unneeded values in the age column ('overall' and 'date' value) and 'overall' value in 'ethnicity' column.
  - Remove 'both' value in 'sex' column.
- Using Pivot Table to analyse the population by district (as of 2023);
  - Theres a 2.49% increase in population when compared to 2022 population.

- For the population age group distribution, age 20-24 represents most of the population with 10.81% (3035.8 of 1000 people), followed by age 25-29 (10.40%), 30-34 (10.28%), 35-39 (10.11%) and 15-19 (9.86%),



- Top 5 district with the highest population of 1000 people are Petaling with 2009.1, W.P. Kuala Lumpur with 1747, Johor Bahru (1454.4), Ulu Langat (1254.6) and Klang (979.2),



- Bottom 5 district with the lowest population of 1000 people, 4 of them are in Sarawak, which is Tanjung Manis (6.1), Song (7.5), Bukit Mabong (7.9) and Julau (12.6), and in Kecil Lojing, Kelantan with 7.6,

- When sorting through the district with the largest population of certain ethnicity, below are the results:

1		
2		
3	Row Labels	Sum of population
4	Petaling	951.9
5	W.P. Kuala Lumpur	721.1
6	Ulu Langat	707.4
7	Johor Bahru	665
8	Gombak	488
9	Grand Total	3533.4
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

**ethnicity**

- bumi\_malay
- bumi\_other
- chinese
- indian**
- other\_citizen
- other\_noncitizen

**Year**

- 2020
- 2021
- 2022
- 2023**

District Pop by Ethnic
Top & Bottom District Pop
Age Distribution
Table1
population\_district

- Bumi\_malay – Petaling (951.9), W.P. Kuala Lumpur (721.1), Ulu Langat (707.4), Johor Bahru (665), Gombak (488),
- Bumi\_other – Kota Kinabalu (227), Sandakan (165.8), Tawau (162.2), Kuching (132.3), Lahad Datu (103.1)
- Chinese – Petaling (651.6), W.P. Kuala Lumpur (636.4), Ulu Langat (309), Johor Bahru (503.6), Timur Laut (282.7),
- Indian – Petaling (181.7), Klang (163.4), W.P.Kuala Lumpur (154.3), Johor Bahru (118.8), Kinta (94.2),
- Other\_citizen – Petaling (14.1), Gombak(12.9), W.P. Kuala Lumpur(10.9), Johor Bahru(9.6), Tawau(8.7),
- Other\_noncitizen – W.P. Kuala Lumpur (206.9), Petaling (190.7), Sandakan (141), Johor Bahru (130.9), Ulu Langat (125.1).

You can view the Excel File on my Github page.

<https://github.com/MHafizAzizi/DOSM-MyStep-41-Preliminary-Test>