Pakistan Population Analysis

AUTHOR NAME: NOMAN khan
learning from: Codanics.com

About Dataset

the data set collected from the source is describe following

`Description` This dataset contains demographic information from the Pakistan Population Census conducted in 2017. It provides detailed population data at various administrative levels within Pakistan, including provinces, divisions, districts, and sub-divisions. The dataset also includes information on urban and rural populations, gender distribution, transgender individuals, sex ratios, population figures from the 1998 census, and annual growth rates.

`Features` Province: The administrative provinces or regions of Pakistan where the census data was collected.

Division: The divisions within each province. Divisions are the second level of administrative divisions in Pakistan.

District: Districts within each division, representing larger administrative units.

Sub-Division: Sub-divisions or tehsils within each district, providing more localized data.

Area: The land area (in square kilometers) of each sub-division.

Urban Population 2017: The population of urban areas within each sub-division for the year 2017.

Rural Population 2017: The population of rural areas within each sub-division for the year 2017.

Male Population 2017: The male population within each sub-division for the year 2017.

Female Population 2017: The female population within each sub-division for the year 2017.

Transgender Population 2017: The population of transgender individuals within each sub-division for the year 2017.

Sex Ratio 2017: The sex ratio, calculated as the number of females per 1000 males, within each sub-division for the year 2017.

Population in 1998: The total population of each sub-division as recorded in the 1998 census.

Annual Growth Rate: The annual growth rate of the population in each sub-division, calculated as the percentage increase from 1998 to 2017.

`Data Source` The data in this dataset was collected from official Pakistan Population Census reports and may include data from various government sources. It is essential to provide proper attribution and reference the original sources when using this dataset for analysis or research.

`Data Usage` Researchers and analysts can use this dataset to explore demographic trends, population growth, urbanization rates, gender distribution, and more within Pakistan at different administrative levels. Ensure

compliance with ethical and legal guidelines when using this data for research or public sharing.

Please note that this description is a template, and you should adapt it based on the actual data sources and specific details of your dataset when creating it for Kaggle or any other platform.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

> loading the Pakistan population datadset and making DSA `df=pd.read_csv('data2.csv')`

---

1. All columns are not visible so we use set option function for all columns

```
pd.set_option('display.max_columns',None)
```

2. #checking the first five rows of the dataset `df.head()`

3. information about the dataset `df.info()`

4. making statistatical analysis `df.describe()`

5. lets check all the column and make quetion but befor checking null values `df.isnull().sum()`

> the data is fully cleaned no errors no null values it contained.

6. `df.columns`

```
 'ALL SEXES (RURAL)', 'MALE (RURAL)', 'FEMALE (RURAL)',
 'TRANSGENDER (RURAL)', 'SEX RATIO (RURAL)',
 'AVG HOUSEHOLD SIZE (RURAL)', 'POPULATION 1998 (RURAL)',
 'ANNUAL GROWTH RATE (RURAL)', 'ALL SEXES (URBAN)', 'MALE (URBAN)',
 'FEMALE (URBAN)', 'TRANSGENDER (URBAN)', 'SEX RATIO (URBAN)',
 'AVG HOUSEHOLD SIZE (URBAN)', 'POPULATION 1998 (URBAN)',
 'ANNUAL GROWTH RATE (URBAN)'],
dtype='object')```
```

7. which division has the highest area in Pakistan?

```
area_i=df['AREA (sq.km)'].max()
area_i
```

```
area_l=df['AREA (sq.km)'].idxmax()
name=df.loc[area_l,'DIVISION']  # Use the index to get the division name

#the loc function is used to locate the name of division based on index number

print(f"The area {name} has the highest area of {area_i}.")
> The area BAHAWALPUR DIVISION has the highest area of 18374.0.
```

8. how much the population increased from 1998 t0 2017 in rural ?

```
pop_rural_1998=df['POPULATION 1998 (RURAL)'].sum()
pop_rural_2017=df['ALL SEXES (RURAL)'].sum()
pop_change_rural=pop_rural_2017 - pop_rural_1998
print(pop_change_rural)
pop_rural_percentage_change=(pop_change_rural/pop_rural_1998)*100
print(pop_rural_percentage_change)

> 73692421
56.37
```

9. which division has the lowest area in Pakistan?

```
area_i_m=df['AREA (sq.km)'].min()
area_l_m=df['AREA (sq.km)'].idxmin()
name=df.loc[area_l_m,'DIVISION']  # Use the index to get the division name
#the loc function is used to locate the name of division based on index number
print(f"The area {name} has the Lowest area of {area_i}.")
> The area Zhob Division has the Lowest area of 18374.0.
```

9. which area has the largest population in pakistan?

```
#first compute the all sex urban and all sex rural into one columns

df['total population']=df[['ALL SEXES (RURAL)','ALL SEXES (URBAN)']].sum(axis=1)
highest=df['total population'].max()
highest_index=df['total population'].idxmax()
name_pop=df.loc[highest_index,'SUB DIVISION']
print(f"The area {name_pop} has the Lowest area of {highest}.")
> 'PESHAWAR TEHSIL'
```

10. what is the average rural popoulation?

```
df['ALL SEXES (RURAL)'].max()
> 2297375
```

11. what is the min urban population?

```
df['ALL SEXES (URBAN)'].min()
> 0
```

12. which area has the highest annual growth rate in urban?

```
urban_growth=df['ANNUAL GROWTH RATE (URBAN)'].idxmax()
name_urban=df.loc[urban_growth,'SUB DIVISION']
print("the ",name_urban," has highest annual growth rate")
> the  RAIWIND TEHSIL  has highest annual growth rate
```

13. which area has highest annual growth rate in pakistan?

```
df['highest annual growth rate']=df[['ANNUAL GROWTH RATE (RURAL)','ANNUAL GROWTH
RATE (URBAN)']].sum(axis=1)
highest_pop_annual_rate=df['highest annual growth rate'].max()
highest_pop_annual_rate_ind=df['highest annual growth rate'].idxmax()
highest_pop_annual_rate_name=df.loc[highest_pop_annual_rate_ind,'SUB DIVISION']
print("The highest Annual population rate in makistan is
",highest_pop_annual_rate," which name is ",highest_pop_annual_rate_name)
> The highest Annual population rate in makistan is  119.78  which name is
RAIWIND TEHSIL
```

14. how much population increased from 1998 to 2017?

```
#calculate count and percentage?
df['population_1998']=df[['POPULATION 1998 (URBAN)','POPULATION 1998
(RURAL)']].sum(axis=1)

pop_1998=df['population_1998'].sum()
pop_2017=df['total population'].sum()
pop_change=pop_2017 - pop_1998
print(pop_change)
pop_percentage_change=(pop_change/pop_1998)*100
print(round(pop_percentage_change,2))
>73692421
56.37
```

15. # how much the population increased from 1998 t0 2017 in rural ?

```
pop_rural_1998=df['POPULATION 1998 (RURAL)'].sum()
pop_rural_2017=df['ALL SEXES (RURAL)'].sum()
pop_change_rural=pop_rural_2017 - pop_rural_1998
print(pop_change_rural)
pop_rural_percentage_change=(pop_change_rural/pop_rural_1998)*100
print(pop_rural_percentage_change)
>41632809
47.094882409931515
```

16.
```
'Question': [
    '#which division has the highest area in Pakistan?',
    'which division has the lowest area in Pakistan?',
    'which area has the largest population in pakistan?',
    'which area has the lowest population in pakistan?',
    'what is the average  rural popoulation?',
    'what is the min  urban popoulation?',
    'which area has the highest annual growth rate in urban?',
    'which area has the lowest annual growth rate in Pakistan?',
    'how much population increased from 1998 to 2017?',
    'how much the population increased from 1998 t0 2017 in urban ?',
    'how much the population increased from 1998 t0 2017 in rural ?'

],
'Option A': ['BAHAWALPUR DIVISION', 'D.G.KHAN DIVISION', 'GUJRANWALA SADDAR
TEHSIL', 'KHARIAN TEHSIL', '236336','2','PINDI BHATTIAN TEHSIL','RAIWIND
TEHSIL','4567532','12456','41632809'],
'Option B': ['D.G.KHAN DIVISIO', 'FAISALABAD DIVISION', 'PESHAWAR TEHSIL',
'PINDI BHATTIAN TEHSIL', '363847','9090','RAIWIND TEHSIL','KHARIAN
TEHSIL','16327287','33338','242556'],
'Option C': ['FAISALABAD DIVISION', 'Zhob Division', 'WAZIRABAD TEHSIL',
```

```
  'SANGAN SUB-TEHSIL', '2297375','10','PESHAWAR TEHSIL','LAHORE CANTT
  TEHSIL','783939','32059612','127654'],
  'Option D': ['Zhob Division', 'GUJRANWALA DIVISION', 'KHARIAN TEHSIL',
  'GUJRANWALA SADDAR TEHSIL', '190922','0','KHARIAN TEHSIL','MODEL TOWN
  TEHSIL','73692421','283909','10938'],
  'Correct Answer': ['A', 'C', 'B', 'C', 'C','D','B','A','D','C','A']
```

}

mcq_df = pd.DataFrame(mcq_data) mcq_df ```

```
mcq_df.to_csv('quiz_QUESTions.csv',index=False)
```

```
        THE ENd
```