```
In [ ]:   # This Python 3 environment comes with many helpful analytics libraries installed
          # It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
          # For example, here's several helpful packages to load

          import numpy as np # linear algebra
          import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

          # Input data files are available in the read-only "../input/" directory
          # For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

          import os
          for dirname, _, filenames in os.walk('/kaggle/input'):
              for filename in filenames:
                  print(os.path.join(dirname, filename))

          # You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run A
          # You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

/kaggle/input/population-of-pakistan-dataset/sub-division_population_of_pakistan.csv

A). About the Author

Author: Muhammad Adnan

Date: 1/11/2023

Data: Pakistan Population from 1998 - 2017

For more information please follow me on the following accounts
Twitter
Github
LinkedIn
Gmail

B). Kernel Version Used

Python 3.12.0

C). About the DataSet

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.

D). Import libraries

- Pandas (Data manipulation and analysis library) \

Pandas version (2.1.1)

- Plolty (Data visualization library) \

Plotly version (5.17.0)

E). Purpose of the data analysis?

The primary objective of this analysis is to discern underlying patterns by leveraging built-in libraries. I aim to extract insights from the data by posing a series of self-generated questions and representing their answers through a set of accompanying visualizations. The list of questions and their corresponding graphical representations is provided below.

F). List of Questions

- Q1. How many province in the Pakistan?
- Q2. How many Divisions in the Pakistan?
- Q3. How many Districts in the Pakistan?
- Q4. How many Sub-Divisions in the Pakistan?
- Q5. How much AREA (sq.km)/Province has in Pakistan?
- Q6. How many divisions in each province of Pakistan?
- Q7. How many Sub-Divisions in each province of Pakistan?
- Q8. How many Sub Divisions in each province of Pakistan?
- Q9. What is the Gender Distribution in each province of Pakistan?
- Q10. What is the Average House Hold both in (Rural & Urban) in each province of Pakistan?
- Q11. Show population of Pakistan in 1998 (Rural & Urban).
- Q12. Show Population of Pakistan in 1998, broken down into (rural & urban), for each province of Pakistan.
- Q13. Show sex ratio of Pakistan in both (Rural & Urban).

- Q14. Show sex ratio (Rural & Urban) in five provinces of Pakistan.
- Q15. Show annual growth rate of Pakistan in both (Rural & Urban).
- Q16. Show average annual growth rate of pakistan both (Rural & Urban) in each province.
- Q17. Show average annual growth rate of Pakistan both (RURAL & URBAN) in each Division.

```python
# Import libraries
import pandas as pd # For data manipulations
import plotly.express as px # For data visualization
import plotly.graph_objs as go
```

```python
# Load the dataset
df = pd.read_csv("/kaggle/input/population-of-pakistan-dataset/sub-division_population_of_pakistan.csv")
```

Now check the Data Composition and so on for getting useful insights from the data.

```python
# Check the first few row of some data
df.head(4)
```

	PROVINCE	DIVISION	DISTRICT	SUB DIVISION	AREA (sq.km)	ALL SEXES (RURAL)	MALE (RURAL)	FEMALE (RURAL)	TRANSGENDER (RURAL)	SEX RATIO (RURAL)	...	POPULATION 1998 (RURAL)	ANNUAL GROWTH RATE (RURAL)	ALL SEXES (URBAN)	(UR
0	PUNJAB	BAHAWALPUR DIVISION	BAHAWALNAGAR DISTRICT	BAHAWALNAGAR TEHSIL	1729.0	619550	316864	302644	42	104.70	...	407768	2.22	193840	
1	PUNJAB	BAHAWALPUR DIVISION	BAHAWALNAGAR DISTRICT	CHISHTIAN TEHSIL	1500.0	540342	273788	266500	54	102.73	...	395983	1.65	149424	
2	PUNJAB	BAHAWALPUR DIVISION	BAHAWALNAGAR DISTRICT	FORT ABBAS TEHSIL	2536.0	361240	182655	178541	44	102.30	...	250959	1.93	61528	
3	PUNJAB	BAHAWALPUR DIVISION	BAHAWALNAGAR DISTRICT	HAROONABAD TEHSIL	1295.0	382115	192278	189808	29	101.30	...	297343	1.33	142600	

4 rows × 21 columns

```python
# Check how many columns and rows inside this data
print("The columns of the datasets are:", df.shape[0]) # 0  indicates columns in the dataset
print("The row of the datasets are:", df.shape[1]) # 1 indicates the rows of the dataset
```

```
The columns of the datasets are: 528
The row of the datasets are: 21
```

```python
# check the missing values in the dataset
df.isnull().sum()
```

```
PROVINCE                        0
DIVISION                        0
DISTRICT                        0
SUB DIVISION                    0
AREA (sq.km)                    0
ALL SEXES (RURAL)               0
MALE (RURAL)                    0
FEMALE (RURAL)                  0
TRANSGENDER (RURAL)             0
SEX RATIO (RURAL)               0
AVG HOUSEHOLD SIZE (RURAL)      0
POPULATION 1998 (RURAL)         0
ANNUAL GROWTH RATE (RURAL)      0
ALL SEXES (URBAN)               0
MALE (URBAN)                    0
FEMALE (URBAN)                  0
TRANSGENDER (URBAN)             0
SEX RATIO (URBAN)               0
AVG HOUSEHOLD SIZE (URBAN)      0
POPULATION 1998 (URBAN)         0
ANNUAL GROWTH RATE (URBAN)      0
dtype: int64
```

```python
# check the summary of the data, keep in mind the descirbe function will only work on numeric data
df.describe()
```

	AREA (sq.km)	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)	FEM/ (URB/
count	528.000000	5.280000e+02	5.280000e+02	5.280000e+02	528.000000	528.000000	528.000000	5.280000e+02	528.000000	5.280000e+02	5.280000e+02	5.280000e
mean	1492.005871	2.462780e+05	1.252757e+05	1.209841e+05	18.174242	98.982614	6.277064	1.674280e+05	3.124792	1.408635e+05	7.284339e+04	6.799787e
std	2039.453778	2.711898e+05	1.375630e+05	1.337169e+05	25.522248	26.812660	2.074947	1.783890e+05	9.577872	3.512463e+05	1.823492e+05	1.688725e
min	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.000000e+00	0.000000	0.000000e+00	0.000000e+00	0.000000e
25%	425.000000	5.093450e+04	2.712725e+04	2.397900e+04	1.000000	101.105000	5.700000	3.527350e+04	1.440000	0.000000e+00	0.000000e+00	0.000000e
50%	882.000000	1.652410e+05	8.413450e+04	8.204400e+04	10.000000	105.285000	6.310000	1.172065e+05	2.030000	4.325450e+04	2.198000e+04	2.099900e
75%	1734.250000	3.129112e+05	1.605022e+05	1.522190e+05	25.000000	108.347500	7.200000	2.130542e+05	2.812500	1.178148e+05	6.030175e+04	5.746525e
max	18374.000000	2.297375e+06	1.172995e+06	1.124167e+06	213.000000	139.380000	12.430000	1.044035e+06	100.000000	3.653616e+06	1.905921e+06	1.746900e

If you see we get a scroll bar to see all the columns, if you dont want to scroll it then just simply put T at the end of the code.

```
# Summary of the dataset
df.describe().T # T means transpose
```

	count	mean	std	min	25%	50%	75%	max
AREA (sq.km)	528.0	1492.005871	2039.453778	0.0	425.000	882.000	1734.2500	18374.00
ALL SEXES (RURAL)	528.0	246278.011364	271189.816559	0.0	50934.500	165241.000	312911.2500	2297375.00
MALE (RURAL)	528.0	125275.687500	137563.021458	0.0	27127.250	84134.500	160502.2500	1172995.00
FEMALE (RURAL)	528.0	120984.149621	133716.898296	0.0	23979.000	82044.000	152219.0000	1124167.00
TRANSGENDER (RURAL)	528.0	18.174242	25.522248	0.0	1.000	10.000	25.0000	213.00
SEX RATIO (RURAL)	528.0	98.982614	26.812660	0.0	101.105	105.285	108.3475	139.38
AVG HOUSEHOLD SIZE (RURAL)	528.0	6.277064	2.074947	0.0	5.700	6.310	7.2000	12.43
POPULATION 1998 (RURAL)	528.0	167427.994318	178388.976993	0.0	35273.500	117206.500	213054.2500	1044035.00
ANNUAL GROWTH RATE (RURAL)	528.0	3.124792	9.577872	0.0	1.440	2.030	2.8125	100.00
ALL SEXES (URBAN)	528.0	140863.528409	351246.279136	0.0	0.000	43254.500	117814.7500	3653616.00
MALE (URBAN)	528.0	72843.386364	182349.241379	0.0	0.000	21980.000	60301.7500	1905921.00
FEMALE (URBAN)	528.0	67997.865530	168872.528212	0.0	0.000	20999.000	57465.2500	1746900.00
TRANSGENDER (URBAN)	528.0	22.276515	66.068127	0.0	0.000	3.000	19.0000	795.00
SEX RATIO (URBAN)	528.0	75.411269	49.687341	0.0	0.000	103.195	107.0400	297.81
AVG HOUSEHOLD SIZE (URBAN)	528.0	4.406402	2.948336	0.0	0.000	5.755	6.3400	10.06
POPULATION 1998 (URBAN)	528.0	80144.566288	202312.019227	0.0	0.000	21298.000	65422.2500	2075867.00
ANNUAL GROWTH RATE (URBAN)	528.0	1.920814	2.098908	0.0	0.000	1.855	2.9850	19.78

Now see all the columns becomes rows and easy to see the results.

Visualize the dataset to see the patterns inside

### Question 1. How many province in the Pakistan?

```
# Create a new dataframe into a pandas dataframe
province = df['PROVINCE'].value_counts()

# Create a pie object
# create a Pie object
pie = go.Pie(labels=province.index, values=province.values)
# create a layout object
layout = go.Layout(title='Province distribution in Pakistan')
# create a Figure object that contains the pie chart and layout
fig = go.Figure(data=[pie], layout=layout)
# set the width and height
fig.update_layout(width=700, height=500)

# display the chart
fig.show()
```

Province distribution in Pakistan
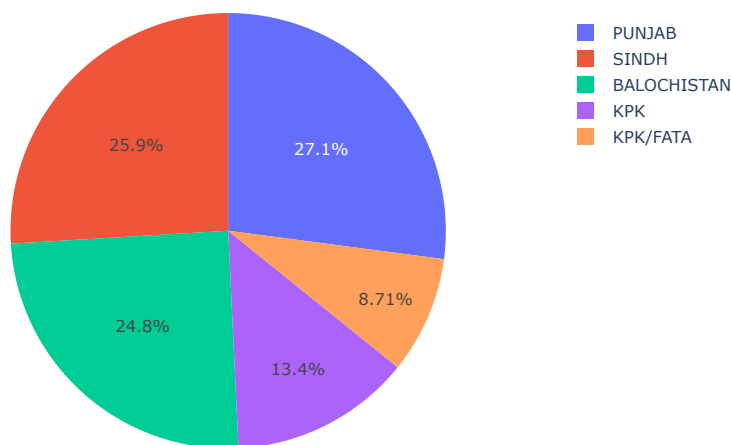
Interpretation

This pie chart and barplot presents comprehensive data on the regional divisions within Pakistan, encompassing its five constituent provinces: Balochistan, Punjab, Sindh, Khyber Pakhtunkhwa (KPK), and KPK/FATA. Among these, Balochistan stands as the largest province, covering a significant expanse, accounting for approximately 27.1% of Pakistan's total land area. The subsequent order of provinces, in terms of land area, includes Punjab at 25.9%, Sindh at 24.8%, KPK at 13.4%, and KPK/FATA at 8.71%, respectively.

Question 1. How many province in the Pakistan?

Answer. There are five provinces in Pakistan, Balochistan, Sindh, Punjab, KPK, & KPK/FATA.

Question 2. How many Divisions in the Pakistan?

```
In [ ]:  print("Show all the division names:",df['DIVISION'].unique())
         print("Show how many divisions in Pakistan?:", df['DIVISION'].nunique())
```

```
Show all the division names: ['BAHAWALPUR DIVISION' 'D.G.KHAN DIVISION' 'FAISALABAD DIVISION'
 'GUJRANWALA DIVISION' 'LAHORE DIVISION' 'MULTAN DIVISION'
 'RAWALPINDI DIVISION' 'SAHIWAL DIVISION' 'SARGODHA DIVISION'
 'Badin Division' 'Hyderabad Division' 'Karachi Division'
 'Larkana Division' 'Mirpurkhas Division' 'Shaheed Benazirabad Division'
 'Sukkur Division' 'Makran Division' 'Kalat Division'
 'Naseerabad Division' 'Quetta Division' 'Zhob Division' 'BANNU DIVISION'
 'DERA ISMAIL KHAN DIVISION' 'HAZARA DIVISION' 'KOHAT DIVISION'
 'MARDAN DIVISION' 'PESHAWAR DIVISION' 'MALAKAND DIVISION']
Show how many divisions in Pakistan?: 28
```
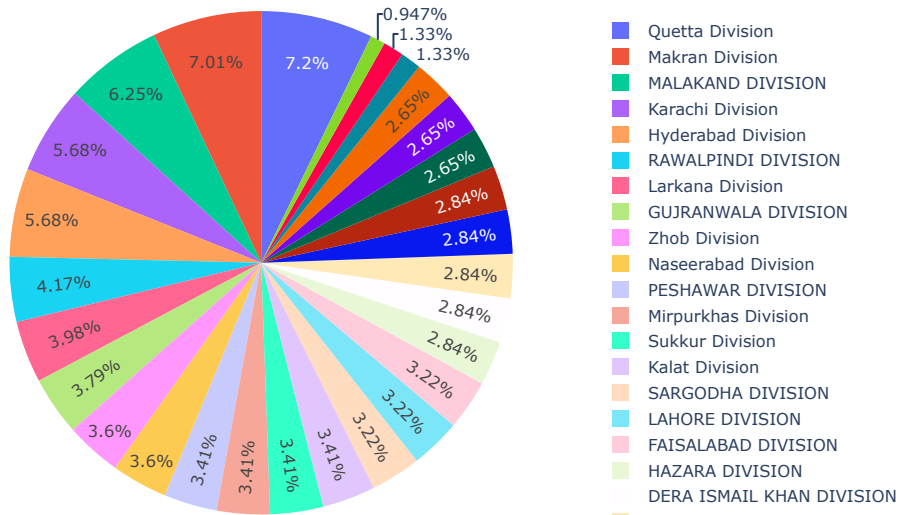
```
In [ ]:  # Create a new dataframe into a pandas dataframe
         div = df['DIVISION'].value_counts()

         # Create a pie object
         # create a Pie object
         pie = go.Pie(labels=div.index, values=div.values)
         # create a layout object
         layout = go.Layout(title='Total Divisions in Pakistan')
         # create a Figure object that contains the pie chart and layout
         fig = go.Figure(data=[pie], layout=layout)
         # set the width and height
         fig.update_layout(width=800, height=550)

         # display the chart
         fig.show()
```

Total Divisions in Pakistan

Legend:
- Quetta Division
- Makran Division
- MALAKAND DIVISION
- Karachi Division
- Hyderabad Division
- RAWALPINDI DIVISION
- Larkana Division
- GUJRANWALA DIVISION
- Zhob Division
- Naseerabad Division
- PESHAWAR DIVISION
- Mirpurkhas Division
- Sukkur Division
- Kalat Division
- SARGODHA DIVISION
- LAHORE DIVISION
- FAISALABAD DIVISION
- HAZARA DIVISION
- DERA ISMAIL KHAN DIVISION

Pie slice percentages: 7.2%, 0.947%, 1.33%, 1.33%, 2.65%, 2.65%, 2.65%, 2.84%, 2.84%, 2.84%, 2.84%, 2.84%, 3.22%, 3.22%, 3.22%, 3.41%, 3.41%, 3.41%, 3.41%, 3.6%, 3.6%, 3.79%, 3.98%, 4.17%, 5.68%, 5.68%, 6.25%, 7.01%

Question 2. How many Divisions in the Pakistan?

Answer. There are total 28 divisions in Pakistan.

Question 3. How many Districts in the Pakistan?

```
In [ ]:  print("Show the names of all the districts in pakistan:",df['DISTRICT'].unique())
         print("Show how many of districts in pakistan:", df['DISTRICT'].nunique())
```

```
Show the names of all the districts in pakistan: ['BAHAWALNAGAR DISTRICT' 'BAHAWALPUR DISTRICT' 'RAHIM YAR KHAN DISTRICT'
 'DERA GHAZI KHAN DISTRICT' 'LAYYAH DISTRICT' 'MUZAFFARGARH DISTRICT'
 'RAJANPUR DISTRICT' 'CHINIOT DISTRICT' 'FAISALABAD DISTRICT'
 'JHANG DISTRICT' 'TOBA TEK SINGH DISTRICT' 'GUJRANWALA DISTRICT'
 'GUJRAT DISTRICT' 'HAFIZABAD DISTRICT' 'MANDI BAHAUDDIN DISTRICT'
 'NAROWAL DISTRICT' 'SIALKOT DISTRICT' 'KASUR DISTRICT' 'LAHORE DISTRICT'
 'NANKANA SAHIB DISTRICT' 'SHEIKHUPURA DISTRICT' 'KHANEWAL DISTRICT'
 'LODHRAN DISTRICT' 'MULTAN DISTRICT' 'VEHARI DISTRICT' 'ATTOCK DISTRICT'
 'CHAKWAL DISTRICT' 'JHELUM DISTRICT' 'RAWALPINDI DISTRICT'
 'OKARA DISTRICT' 'PAKPATTAN DISTRICT' 'SAHIWAL DISTRICT'
 'BHAKKAR DISTRICT' 'KHUSHAB DISTRICT' 'MIANWALI DISTRICT'
 'SARGODHA DISTRICT' 'BADIN DISTRICT' 'DADU DISTRICT' 'HYDERABAD DISTRICT'
 'JAMSHORO DISTRICT' 'MATIARI DISTRICT' 'SUJAWAL DISTRICT'
 'TANDO ALLAHYAR DISTRICT' 'TANDO MUHAMMAD KHAN DISTRICT'
 'THATTA DISTRICT' 'KARACHI CENTRAL DISTRICT' 'KARACHI EAST DISTRICT'
 'KARACHI SOUTH DISTRICT' 'KARACHI WEST DISTRICT' 'KORANGI DISTRICT'
 'MALIR DISTRICT' 'JACOBABAD DISTRICT' 'KAMBAR SHAHDAD KOT DISTRICT'
 'KASHMORE DISTRICT' 'LARKANA DISTRICT' 'SHIKARPUR DISTRICT'
 'MIRPUR KHAS DISTRICT' 'THARPARKAR DISTRICT' 'UMER KOT DISTRICT'
 'NAUSHAHRO FEROZE DISTRICT' 'SANGHAR DISTRICT'
 'SHAHEED BENAZIRABAD DISTRICT' 'GHOTKI DISTRICT' 'KHAIRPUR DISTRICT'
 'SUKKUR DISTRICT' 'AWARAN DISTRICT' 'KALAT DISTRICT' 'KHARAN DISTRICT'
 'KHUZDAR DISTRICT' 'LASBELA DISTRICT' 'MASTUNG DISTRICT'
 'WASHUK DISTRICT' 'GWADAR DISTRICT' 'KECH DISTRICT' 'PANJGUR DISTRICT'
 'JAFFARABAD DISTRICT' 'JHAL MAGSI DISTRICT' 'KACHHI DISTRICT'
 'NASIRABAD DISTRICT' 'SOHBATPUR DISTRICT' 'CHAGAI DISTRICT'
 'KILLA ABDULLAH DISTRICT' 'NUSHKI DISTRICT' 'PISHIN DISTRICT'
 'QUETTA DISTRICT' 'DERA BUGTI DISTRICT' 'HARNAI DISTRICT'
 'KOHLU DISTRICT' 'SIBI DISTRICT' 'ZIARAT DISTRICT' 'BARKHAN DISTRICT'
 'KILLA SAIFULLAH DISTRICT' 'LORALAI DISTRICT' 'MUSAKHEL DISTRICT'
 'SHERANI DISTRICT' 'ZHOB DISTRICT' 'BANNU DISTRICT'
 'LAKKI MARWAT DISTRICT' 'DERA ISMAIL KHAN DISTRICT' 'TANK DISTRICT'
 'ABBOTTABAD DISTRICT' 'HARIPUR DISTRICT' 'MANSEHRA DISTRICT'
 'TORGHAR DISTRICT' 'KOHISTAN DISTRICT' 'HANGU DISTRICT' 'KOHAT DISTRICT'
 'KARAK DISTRICT' 'MARDAN DISTRICT' 'SWABI DISTRICT' 'CHARSADDA DISTRICT'
 'NOWSHERA DISTRICT' 'PESHAWAR DISTRICT' 'BUNER DISTRICT'
 'CHITRAL DISTRICT' 'SHANGLA DISTRICT' 'SWAT DISTRICT'
 'LOWER DIR DISTRICT' 'UPPER DIR DISTRICT' 'BAJAUR DISTRICT'
 'FR BANNU DISTRICT' 'FR D.I.KHAN DISTRICT' 'FR KOHAT DISTRICT'
 'FR LAKKI MARWAT DISTRICT' 'FR PESHAWAR DISTRICT' 'KHYBER DISTRICT'
 'KURRAM DISTRICT' 'MOHMAND DISTRICT' 'NORTH WAZIRISTAN DISTRICT'
 'ORAKZAI DISTRICT' 'SOUTH WAZIRISTAN DISTRICT']
Show how many of districts in pakistan: 131
```

Question 3. How many Districts in the Pakistan?

Answer. There are 131 districts in Pakistan, due to huge number we can't visualize it.

Question 4. How many Sub-Divisions in the Pakistan?

```python
print("Show all the sub-diviions in pakistan:", df['SUB DIVISION'].unique())
print("Show the total number of Sub-divisions in pakistan:", df['SUB DIVISION'].nunique())
```

Question 4. How many Sub-Divisions in the Pakistan?

```python
print("Show all the sub-diviions in pakistan:", df['SUB DIVISION'].unique())
print("Show the total number of Sub-divisions in pakistan:", df['SUB DIVISION'].nunique())
```

Show all the sub-diviions in pakistan: ['BAHAWALNAGAR TEHSIL' 'CHISHTIAN TEHSIL' 'FORT ABBAS TEHSIL'
 'HAROONABAD TEHSIL' 'MINCHINABAD TEHSIL' 'AHMADPUR EAST TEHSIL'
 'BAHAWALPUR CITY TEHSIL' 'BAHAWALPUR SADDAR TEHSIL' 'HASILPUR TEHSIL'
 'KHAIRPUR TAMEWALI TEHSIL' 'YAZMAN TEHSIL' 'KHANPUR TEHSIL'
 'LIAQUATPUR TEHSIL' 'RAHIM YAR KHAN TEHSIL' 'SADIQABAD TEHSIL'
 'DE-EXCLUDED AREA D.G KHAN' 'DERA GHAZI KHAN TEHSIL' 'KOT CHHUTTA TEHSIL'
 'TAUNSA TEHSIL' 'CHOUBARA TEHSIL' 'LAYYAH TEHSIL' 'ALIPUR TEHSIL'
 'JATOI TEHSIL' 'KOT ADDU TEHSIL' 'MUZAFFARGARH TEHSIL'
 'DE-EXCLUDED AREA RAJANPUR' 'JAMPUR TEHSIL' 'RAJANPUR TEHSIL'
 'ROJHAN TEHSIL' 'BHAWANA TEHSIL' 'CHINIOT TEHSIL' 'LALIAN TEHSIL'
 'CHAK JHUMRA TEHSIL' 'FAISALABAD CITY TEHSIL' 'FAISALABAD SADAR TEHSIL'
 'JARANWALA TEHSIL' 'SAMMUNDRI TEHSIL' 'TANDLIAN WALA TEHSIL'
 '18-HAZARI TEHSIL' 'AHMADPUR SIAL TEHSIL' 'JHANG TEHSIL' 'SHORKOT TEHSIL'
 'GOJRA TEHSIL' 'KAMALIA TEHSIL' 'PIRMAHAL TEHSIL' 'TOBA TEK SINGH TEHSIL'
 'GUJRANWALA CITY TEHSIL' 'GUJRANWALA SADDAR TEHSIL' 'KAMOKE TEHSIL'
 'NOWSHERA VIRKAN TEHSIL' 'WAZIRABAD TEHSIL' 'GUJRAT TEHSIL'
 'KHARIAN TEHSIL' 'SARAI ALAMGIR TEHSIL' 'HAFIZABAD TEHSIL'
 'PINDI BHATTIAN TEHSIL' 'MALAKWAL TEHSIL' 'MANDI BAHAUDDIN TEHSIL'
 'PHALIA TEHSIL' 'NAROWAL TEHSIL' 'SHAKARGARH TEHSIL' 'ZAFARWAL TEHSIL'
 'DASKA TEHSIL' 'PASRUR TEHSIL' 'SAMBRIAL TEHSIL' 'SIALKOT TEHSIL'
 'CHUNIAN TEHSIL' 'KASUR TEHSIL' 'KOT RADHA KISHAN TEHSIL'
 'PATTOKI TEHSIL' 'LAHORE CANTT TEHSIL' 'LAHORE CITY TEHSIL'
 'MODEL TOWN TEHSIL' 'RAIWIND TEHSIL' 'SHALIMAR TEHSIL'
 'NANKANA SAHIB TEHSIL' 'SANGLA HILL TEHSIL' 'SHAH KOT TEHSIL'
 'FEROZEWALA TEHSIL' 'MURIDKE TEHSIL' 'SAFDARABAD TEHSIL'
 'SHARAK PUR TEHSIL' 'SHEIKHUPURA TEHSIL' 'JAHANIAN TEHSIL'
 'KABIRWALA TEHSIL' 'KHANEWAL TEHSIL' 'MIAN CHANNU TEHSIL'
 'DUNYAPUR TEHSIL' 'KAHROR PACCA TEHSIL' 'LODHRAN TEHSIL'
 'JALALPUR PIRWALA TEHSIL' 'MULTAN CITY TEHSIL' 'MULTAN SADDAR TEHSIL'
 'SHUJABAD TEHSIL' 'BUREWALA TEHSIL' 'MAILSI TEHSIL' 'VEHARI TEHSIL'
 'ATTOCK TEHSIL' 'FATEH JANG TEHSIL' 'HASAN ABDAL TEHSIL' 'HAZRO TEHSIL'
 'JAND TEHSIL' 'PINDI GHEB TEHSIL' 'CHAKWAL TEHSIL'
 'CHOA SAIDAN SHAH TEHSIL' 'KALLAR KAHAR TEHSIL' 'LAWA TEHSIL'
 'TALA GANG TEHSIL' 'DINA TEHSIL' 'JHELUM TEHSIL' 'PIND DADAN KHAN TEHSIL'
 'SOHAWA TEHSIL' 'GUJAR KHAN TEHSIL' 'KAHUTA TEHSIL'
 'KALLAR SAYADDAN TEHSIL' 'KOTLI SATTIAN TEHSIL' 'MURREE TEHSIL'
 'RAWALPINDI TEHSIL' 'TAXILA TEHSIL' 'DEPALPUR TEHSIL' 'OKARA TEHSIL'
 'RENALA KHURD TEHSIL' 'ARIF WALA TEHSIL' 'PAKPATTAN TEHSIL'
 'CHICHAWATNI TEHSIL' 'SAHIWAL TEHSIL' 'BHAKKAR TEHSIL'
 'DARYA KHAN TEHSIL' 'KALUR KOT TEHSIL' 'MANKERA TEHSIL' 'KHUSHAB TEHSIL'
 'NOORPUR TEHSIL' 'NOWSHERA TEHSIL' 'QUAIDABAD TEHSIL' 'ISAKHEL TEHSIL'
 'MIANWALI TEHSIL' 'PIPLAN TEHSIL' 'BHALWAL TEHSIL' 'BHERA TEHSIL'
 'KOT MOMIN TEHSIL' 'SARGODHA TEHSIL' 'SHAHPUR TEHSIL' 'SILLANWALI TEHSIL'
 'BADIN TALUKA' 'GOLARCHI (S.F.RAHU) TALUKA' 'MATLI TALUKA'
 'TALHAR TALUKA' 'TANDO BAGO TALUKA' 'DADU TALUKA' 'JOHI TALUKA'
 'KHAIRPUR NATHAN SHAH TALUKA' 'MEHAR TALUKA' 'HYDERABAD CITY TALUKA'
 'HYDERABAD TALUKA' 'LATIFABAD TALUKA' 'QASIMABAD TALUKA' 'KOTRI TALUKA'
 'MANJHAND TALUKA' 'SEHWAN TALUKA' 'THANO BULA KHAN TALUKA' 'HALA TALUKA'
 'MATIARI TALUKA' 'SAEEDABAD TALUKA' 'JATI TALUKA' 'MIRPUR BATHORO TALUKA'
 'SHAH BUNDER TALUKA' 'SUJAWAL TALUKA' 'CHAMBER TALUKA'
 'JHANDO MARI TALUKA' 'TANDO ALLAHYAR TALUKA' 'BULRI SHAH KARIM TALUKA'
 'TANDO GHULAM HYDER TALUKA' 'TANDO MUHAMMAD KHAN TALUKA'
 'GHORABARI TALUKA' 'KETI BUNDER TALUKA' 'MIRPUR SAKRO TALUKA'
 'THATTA TALUKA' 'GULBERG TEHSIL' 'LIAQUATABAD TEHSIL' 'NAZIMABAD TEHSIL'
 'NEW KARACHI TEHSIL' 'NORTH NAZIMABAD TEHSIL' 'FEROZABAD TEHSIL'
 'GULSHAN-E-IQBAL TEHSIL' 'GULZAR-E-HIJRI TEHSIL'
 'JAMSHED QUARTERS TEHSIL' 'ARAM BAGH TEHSIL' 'CIVIL LINES TEHSIL'
 'GARDEN TEHSIL' 'LYARI TEHSIL' 'SADDAR TEHSIL' 'BALDIA TEHSIL'
 'HARBOUR TEHSIL' 'MANGHOPIR TEHSIL' 'MAURIPUR TEHSIL' 'MOMINABAD TEHSIL'
 'ORANGI TEHSIL' 'SITE TEHSIL' 'KORANGI TEHSIL' 'LANDHI TEHSIL'
 'MODEL COLONY TEHSIL' 'SHAH FAISAL TEHSIL' 'AIRPORT TEHSIL'
 'BIN QASIM TEHSIL' 'GADAP TEHSIL' 'IBRAHIM HYDRI TEHSIL'
 'MURAD MEMON TEHSIL' 'SHAH MUREED TEHSIL' 'GARHI KHAIRO TALUKA'
 'JACOBABAD TALUKA' 'THUL TALUKA' 'KAMBAR ALI KHAN TALUKA'
 'MIRO KHAN TALUKA' 'NASIRABAD TALUKA' 'QUBO SAEED KHAN TALUKA'
 'SHAHDAD KOT TALUKA' 'SUJAWAL JUNEJO TALUKA' 'WARAH TALUKA'
 'KANDHKOT TALUKA' 'KASHMORE TALUKA' 'TANGWANI TALUKA' 'BAKRANI TALUKA'
 'DOKRI TALUKA' 'LARKANA TALUKA' 'RATODERO TALUKA' 'GARHI YASIN TALUKA'
 'KHANPUR TALUKA' 'LAKHI TALUKA' 'SHIKARPUR TALUKA' 'DIGRI TALUKA'
 'HUSSAIN BUX MARI TALUKA' 'JHUDO TALUKA' 'KOT GHULAM MUHAMMAD TALUKA'
 'MIRPUR KHAS TALUKA' 'SHUJAABAD TALUKA' 'SINDHRI TALUKA' 'CHACHRO TALUKA'
 'DAHLI TALUKA' 'DIPLO TALUKA' 'ISLAMKOT TALUKA' 'KALOI TALUKA'
 'MITHI TALUKA' 'NAGAR PARKAR TALUKA' 'KUNRI TALUKA' 'PITHORO TALUKA'
 'SAMARO TALUKA' 'UMER KOT TALUKA' 'BHIRIA TALUKA' 'KANDIARO TALUKA'
 'MEHRABPUR TALUKA' 'MORO TALUKA' 'NAUSHAHRO FEROZE TALUKA'
 'JAM NAWAZ ALI TALUKA' 'KHIPRO TALUKA' 'SANGHAR TALUKA'
 'SHAHDADPUR TALUKA' 'SINJHORO TALUKA' 'TANDO ADAM TALUKA' 'DAUR TALUKA'
 'KAZI AHMED TALUKA' 'NAWABSHAH TALUKA' 'SAKRAND TALUKA' 'DAHARKI TALUKA'
 'GHOTKI TALUKA' 'KHANGARH TALUKA' 'MIRPUR MATHELO TALUKA' 'UBAURO TALUKA'
 'FAIZ GANJ TALUKA' 'GAMBAT TALUKA' 'KHAIRPUR TALUKA' 'KINGRI TALUKA'
 'KOT DIJI TALUKA' 'MIRWAH TALUKA' 'NARA TALUKA' 'SOBHO DERO TALUKA'
 'NEW SUKKUR TALUKA' 'PANO AQIL TALUKA' 'ROHRI TALUKA' 'SALEHPAT TALUKA'
 'SUKKUR CITY TALUKA' 'AWARAN TEHSIL' 'GISHKORE SUB-TEHSIL'
 'JHAL JAO SUB-TEHSIL' 'MASHKAI TEHSIL' 'GAZG SUB-TEHSIL'
 'JOHAN SUB-TEHSIL' 'KALAT TEHSIL' 'MANGOCHAR TEHSIL' 'SURAB TEHSIL'
 'KHARAN TEHSIL' 'SAR KHARAN TEHSIL' 'SUB TEHSIL TOHMULK'
 'ARANJI SUB-TEHSIL' 'KARAKH SUB-TEHSIL' 'KHUZDAR TEHSIL'
 'MOOLA SUB-TEHSIL' 'NAL SUB-TEHSIL' 'ORNACH SUB-TEHSIl'
 'SAROONA SUB-TEHSIL' 'WADH TEHSIL' 'ZEHRI TEHSIL' 'BELA TEHSIL'
 'DUREJI TEHSIL' 'GADDANI TEHSIL' 'HUB TEHSIL' 'LAKHRA TEHSIL'
 'LIARI TEHSIL' 'SONMIANI/WINDER TEHSIL' 'UTHAL TEHSIL' 'DASHT TEHSIL'
 'KHAD KOCHA SUB-TEHSIL' 'KIRDGAP SUB-TEHSIL' 'MASTUNG TEHSIL'
 'BESIMA SUB-TEHSIL' 'MASHKHEL SUB-TEHSIL' 'NAG SUB-TEHSIL'

```
 'SHAHGORI SUB-TEHSIL' 'WASHUK SUB-TEHSIL' 'GWADAR TEHSIL' 'JIWANI TEHSIL'
 'ORMARA TEHSIL' 'PASNI TEHSIL' 'SUNTSER SUB-TEHSIL'
 'BALNIGORE SUB-TEHSIL' 'BULAIDA SUB-TEHSIL' 'DASHT SUB-TEHSIL'
 'HOSHAB SUB-TEHSIL' 'KECH (TURBAT) TEHSIL' 'MAND SUB-TEHSIL'
 'TUMP SUB-TEHSIL' 'ZAMORAN SUB-TEHSIL' 'GICHK SUB-TEHSIL'
 'GOWARGO TEHSIL' 'PANJGUR TEHSIL' 'PAROME TEHSIL' 'GANDHAKA TEHSIL'
 'JHAT PAT TEHSIL' 'USTA MUHAMMAD TEHSIL' 'GANDAWA TEHSIL'
 'JHAL MAGSI TEHSIL' 'MIRPUR SUB-TEHSIL' 'BALANARI SUB-TEHSIL'
 'BHAG TEHSIL' 'DHADAR TEHSIL' 'KHATTAN SUB-TEHSIL' 'MACH SUB-TEHSIL'
 'SANNI SUB-TEHSIL' 'BABA KOT TEHSIL' 'CHATTAR SUB-TEHSIL'
 'DERA MURAD JAMALI TEHSIL' 'TAMBOO TEHSIL' 'FARIDABAD TEHSIL'
 'SANHRI TEHSIL' 'SOHBATPUR TEHSIL' 'CHAGAI SUB-TEHSIL' 'DALBANDIN TEHSIL'
 'NOKUNDI SUB-TEHSIL' 'TAFTAN SUB-TEHSIL' 'CHAMAN TEHSIL'
 'DOBANDI SUB-TEHSIL' 'GULISTAN TEHSIL' 'KILLA ABDULLAH TEHSIL'
 'NUSHKI TEHSIL' 'BARSHORE SUB-TEHSIL' 'HURRAM ZAI TEHSIL'
 'KAREZAT SUB-TEHSIL' 'PISHIN TEHSIL' 'SARANAN TEHSIL'
 'PANJ PAI SUB-TEHSIL' 'QUETTA CITY TEHSIL' 'QUETTA SADDAR TEHSIL'
 'BAIKER SUB-TEHSIL' 'DERA BUGTI TEHSIL' 'LOTI TEHSIL' 'MALAM SUB-TEHSIL'
 'PHELAWAGH TEHSIL' 'PIR KOH SUB-TEHSIL' 'SANGSILLAH SUB-TEHSIL'
 'SUI TEHSIL' 'HARNAI TEHSIL' 'KHOAST SUB-TEHSIL' 'SHAHRIG TEHSIL'
 'GRISANI SUB-TEHSIL' 'KAHAN TEHSIL' 'KOHLU TEHSIL' 'MAWAND TEHSIL'
 'KOT MANDAI SUB-TEHSIL' 'LEHRI TEHSIL' 'SANGAN SUB-TEHSIL' 'SIBI TEHSIL'
 'SINJAWI SUB-TEHSIL' 'ZIARAT TEHSIL' 'BARKHAN TEHSIL' 'BADINI SUB-TEHSIL'
 'KAN MEHTARZAI SUB-TEHSIL' 'KILLA SAIFULLAH TEHSIL' 'LOIBAND TEHSIL'
 'MUSLIM BAGH TEHSIL' 'SHINKI SUB-TEHSIL' 'DUKI TEHSIL'
 'LORALAI\\BORI TEHSIL' 'MEKHTAR SUB-TEHSIL' 'DRUG TEHSIL' 'KINGRI TEHSIL'
 'MUSAKHEL TEHSIL' 'SHERANI TEHSIL' 'ASHWAT SUB-TEHSIL'
 'KASHATOO SUB-TEHSIL' 'QAMAR DIN KAREZ TEHSIL' 'SAMBAZA SUB-TEHSIL'
 'ZHOB TEHSIL' 'BANNU TEHSIL' 'DOMEL TEHSIL' 'LAKKI MARWAT TEHSIL'
 'SARAI NAURANG TEHSIL' 'DARABAN TEHSIL' 'DERA ISMAIL KHAN TEHSIL'
 'KULACHI TEHSIL' 'PAHARPUR TEHSIL' 'PAROA TEHSIL' 'TANK TEHSIL'
 'ABBOTTABAD TEHSIL' 'HAVELIAN TEHSIL' 'ALLAI TEHSIL' 'BATAGRAM TEHSIL'
 'GHAZI TEHSIL' 'HARIPUR TEHSIL' 'BALA KOT TEHSIL' 'MANSEHRA TEHSIL'
 'OGHI TEHSIL' 'JUDBA TEHSIL' 'KHANDER TEHSIL' 'DASSU TEHSIL'
 'KANDIA TEHSIL' 'PALAS TEHSIL' 'PATTAN TEHSIL' 'HANGU TEHSIL'
 'TALL TEHSIL' 'KOHAT TEHSIL' 'LACHI TEHSIL' 'BANDA DAUD SHAH TEHSIL'
 'KARAK TEHSIL' 'TAKHT-E-NASRATI TEHSIL' 'KATLANG TEHSIL' 'MARDAN TEHSIL'
 'TAKHT BHAI TEHSIL' 'LAHOR TEHSIL' 'RAZAR TEHSIL' 'SWABI TEHSIL'
 'TOPI TEHSIL' 'CHARSADDA TEHSIL' 'SHABQADAR TEHSIL' 'TANGI TEHSIL'
 'JEHANGIRA TEHSIL' 'PABBI TEHSIL' 'PESHAWAR TEHSIL' 'DAGGAR TEHSIL'
 'GAGRA TEHSIL' 'KHADO KHEL TEHSIL' 'MANDANR TEHSIL' 'CHITRAL TEHSIL'
 'MASTUJ TEHSIL' 'ALPURI TEHSIL' 'BISHAM TEHSIL' 'PURAN TEHSIL'
 'BABUZAI TEHSIL' 'BARIKOT TEHSIL' 'BEHRAIN TEHSIL' 'CHARBAGH TEHSIL'
 'KABAL TEHSIL' 'KHAWAZAKHELA TEHSIL' 'MATTA TEHSIL' 'ADENZAI TEHSIL'
 'LAL QILA TEHSIL' 'SAMARBAGH TEHSIL' 'TEMERGARA TEHSIL' 'DIR TEHSIL'
 'SHARINGAL TEHSIL' 'WARI TEHSIL' 'MALAKAND PROTECTED AREA'
 'SAM RANI ZAI TEHSIL' 'SWAT RANI ZAI TEHSIL' 'BAR CHAMER KAND TEHSIL'
 'BARANG TEHSIL' 'KHAR BAJAUR TEHSIL' 'MAMUND TEHSIL' 'NAWAGAI TEHSIL'
 'SALARZAI TEHSIL' 'UTMAN KHEL TEHSIL' 'FR BANNU' 'FR D.I.KHAN' 'FR KOHAT'
 'FR LAKKI MARWAT' 'FR PESHAWAR' 'BARA TEHSIL' 'JAMRUD TEHSIL'
 'LANDI KOTAL TEHSIL' 'MULLA GORI TEHSIL' 'LOWER KURRAM TEHSIL'
 'UPPER KURRAM TEHSIL' 'AMBAR UTMAN KHEL TEHSIL' 'HALIM ZAI TEHSIL'
 'PINDIALI TEHSIL' 'PRANG GHAR TEHSIL' 'SAFI TEHSIL'
 'UPPER MOHMAND TEHSIL' 'YAKE GHUND TEHSIL' 'DATTA KHEL TEHSIL'
 'DOSSALI TEHSIL' 'GHARYUM TEHSIL' 'GHULAM KHAN TEHSIL' 'MIR ALI TEHSIL'
 'MIRAN SHAH TEHSIL' 'RAZMAK TEHSIL' 'SHEWA TEHSIL' 'SPINWAM TEHSIL'
 'CENTRAL TEHSIL' 'ISMAIL ZAI TEHSIL' 'LOWER TEHSIL' 'UPPER TEHSIL'
 'BIRMIL TEHSIL' 'LADHA TEHSIL' 'MAKIN TEHSIL' 'SARAROGHA TEHSIL'
 'SERWEKAI TEHSIL' 'TIARZA TEHSIL' 'TOI KHULLA TEHSIL' 'WANA TEHSIL']
Show the total number of Sub-divisions in pakistan: 528
```

Question 4. How many Sub-Divisions in the Pakistan?

Answer. There are 528 sub divisions in Pakistan.

Question 5. How much AREA (sq.km)/Province has in Pakistan?

```python
# Calculate the mean age for each sex
area_sqr = df.groupby('PROVINCE')['AREA (sq.km)'].mean().reset_index()
area_sqr
```

|	PROVINCE	AREA (sq.km)
0	BALOCHISTAN	2637.129771
1	KPK	1055.733803
2	KPK/FATA	533.239130
3	PUNJAB	1417.433566
4	SINDH	1022.890511

```python
# Calculate the mean area (sq.km) for each province
area_sqr = df.groupby('PROVINCE')['AREA (sq.km)'].mean().reset_index()

# Create a bar chart
fig = px.bar(area_sqr, x='PROVINCE', y='AREA (sq.km)', color = 'PROVINCE', title='Area Square (KM) of each Province',
             color_discrete_sequence=['#fa5807', '#fad107', '#07fad1', '#07a9fa', '#e207fa'])
fig.update_traces(width=0.6)
fig.update_layout(width=700, height=500)
# shows descending in order
```

```
fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories
fig.show()
```

Area Square (KM) of each Province
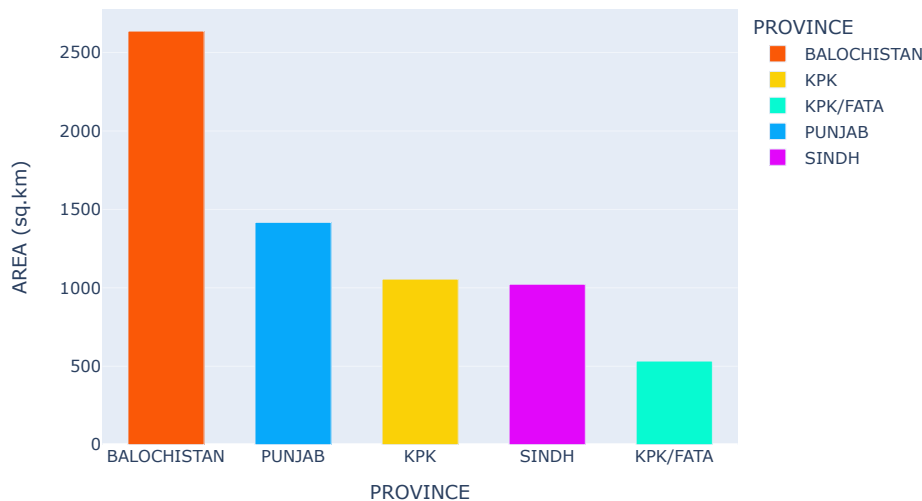
Interpretation

The provided graph illustrates the land area in square kilometers per province in Pakistan. Based on the bar plot and the computed mean values, it is evident that Balochistan boasts the largest land area, totaling 2637.12 square kilometers. Following closely is the province of Punjab with an area of 1417.43 square kilometers, while KPK, Sindh, and KPK/FATA exhibit land areas of 1055.73, 1022.89, and 533.23 square kilometers, respectively.

Question 5. How much AREA (sq.km)/Province has in Pakistan?

Answer. Balochistan (2637.12), Punjab (1417.43), KPK (1055.73), Sindh (1022.89), & KPK/FATA (533.23).

Question 6. How many divisions in each province of Pakistan?

```
In [ ]:   #Divisions in each province
          divisions=df.groupby('PROVINCE')['DIVISION'].unique()
          prov_div=dict(zip(divisions.index,list(divisions.values)))
```

```
In [ ]:   for prov,div in prov_div.items():
              print(f"There are {len(div)} divisions in {prov}")
          print(f"There are 33 division in Pakistan")

          There are 5 divisions in BALOCHISTAN
          There are 7 divisions in KPK
          There are 5 divisions in KPK/FATA
          There are 9 divisions in PUNJAB
          There are 7 divisions in SINDH
          There are 33 division in Pakistan
```

```
In [ ]:   division = pd.DataFrame({ 'Punjab':[9],
                              'KPK':[7],
                               'Sindh':[7],
                               'Balochistan':[5],
                              'KPK/FATA':[5]})
          division
```

```
Out[ ]:
```

|	Punjab	KPK	Sindh	Balochistan	KPK/FATA
0	9	7	7	5	5

```
In [ ]:   # Convert into dataframe with div_df variable
          div_df = pd.DataFrame(division)
          # use the melt function to reshape the data
          div_df = pd.melt(div_df, var_name='Province', value_name='Divisions')
          # Display the reshape Dataframe
          print(div_df)

                Province  Divisions
          0        Punjab          9
          1           KPK          7
          2         Sindh          7
          3   Balochistan          5
          4      KPK/FATA          5
```

```
In [ ]:   # Visualize the division in each provinces
          fig = px.bar(div_df, x='Province', y='Divisions', color='Province')
          fig.update_traces(width=0.5)
```

```
fig.update_layout(width=700, height=500)
# x axis ticks in the 90 degree angle
fig.update_layout(xaxis_tickangle=0)
# graph title name
fig.update_layout(title_text='Pakistan total Divisions (33), divisions in each Province')

fig.show()
```

Pakistan total Divisions (33), divisions in each Province
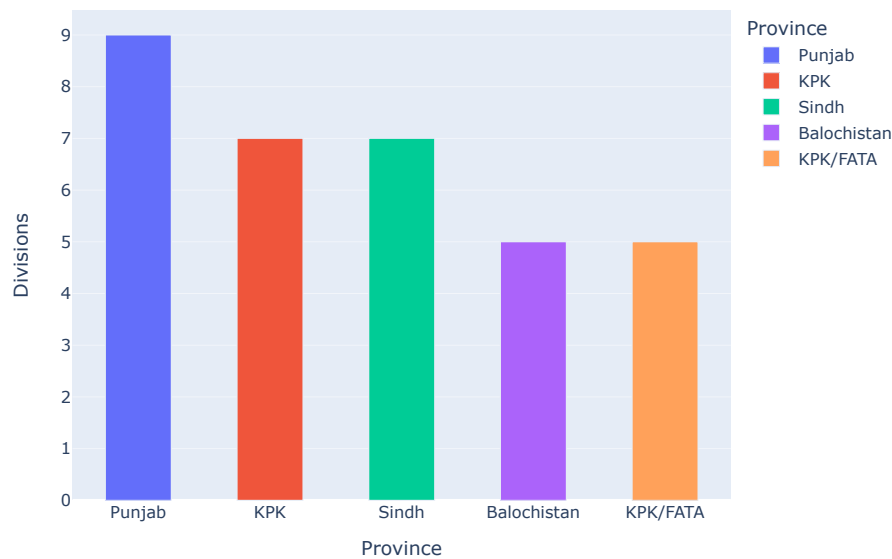
Interpretation

The provided bar graph illustrates the distribution of divisions within each province of Pakistan. Upon a thorough analysis of the graph and the computed data, it becomes evident that Pakistan is comprised of a total of 33 divisions. Punjab emerges as the province with the highest number of divisions, boasting 9 in total. Following closely, both Khyber Pakhtunkhwa (KPK) and Sindh each contain 7 divisions, while Balochistan and KPK/FATA are found to have 5 divisions within their respective boundaries.

Question 6. How many divisions in each province of Pakistan?

Answer. There are in total 33 divisions in pakistan, Punjab (9), KPK (7), Sindh (7), Balochistan (5), & KPK/FATA (5).

Question 7. How many Districts in each province of Pakistan?

```
In [ ]: #Districts in each province
        districts=df.groupby('PROVINCE')['DISTRICT'].unique()
        prov_dis=dict(zip(districts.index,list(districts.values)))
```

```
In [ ]: for prov,div in prov_dis.items():
            print(f"There are {len(div)} divisions in {prov}")
        print(f"There are 33 division in Pakistan")

        There are 31 divisions in BALOCHISTAN
        There are 23 divisions in KPK
        There are 12 divisions in KPK/FATA
        There are 36 divisions in PUNJAB
        There are 29 divisions in SINDH
        There are 33 division in Pakistan
```

```
In [ ]: districts = pd.DataFrame({'Punjab':[36],
                                  'Balochistan':[31],
                                  'Sindh':[29],
                                  'KPK':[23],
                                  'KPK/FATA':[12]})
        districts
```

Out[ ]:

	Punjab	Balochistan	Sindh	KPK	KPK/FATA
0	36	31	29	23	12

```
In [ ]: # Convert it into dataframe with dis_df varibales
        dis_df = pd.DataFrame(districts)
        # use the melt function to reshape the data
        dis_df = pd.melt(dis_df, var_name='Province', value_name='Districts')
        # Display the reshape Dataframe
        print(dis_df)
```

```
      Province  Districts
0      Punjab        36
1  Balochistan       31
2       Sindh        29
3        KPK         23
4     KPK/FATA       12
```

```python
# make plot on melt_df by using plotly library
import plotly.express as px

fig = px.bar(dis_df, x='Province', y='Districts', color='Province')
fig.update_traces(width=0.5)
fig.update_layout(width=700, height=500)
# x axis ticks in the 90 degree angle
fig.update_layout(xaxis_tickangle=0)
# graph title name
fig.update_layout(title_text='Pakistan total Districts is (131), districts in each Province')

fig.show()
```

Pakistan total Districts is (131), districts in each Province

Interpretation

The preceding bar graph provides an overview of the total number of districts within each province of Pakistan. It is discernible from the graph that there are a combined 131 districts across all provinces. Punjab, the most extensive province in terms of districts, houses a total of 36 districts. Following this, Balochistan features 31 districts, Sindh comprises 29, Khyber Pakhtunkhwa (KPK) consists of 23, and the KPK/FATA region encompasses 12 districts, as revealed by the data.

Question 7. How many Districts in each province of Pakistan?

Answer. There are in total 131 Districts in pakistan, Punjab (36), Balochistan (31), Sindh (29), KPK (23), & KPK/FATA (12).

Question 8. How many Sub-Divisions in each province of Pakistan?

```python
#Sub-Divisions in each province
sub_div=df.groupby('PROVINCE')['SUB DIVISION'].unique()
prov_sub_div=dict(zip(sub_div.index,list(sub_div.values)))
```

```python
for prov,div in prov_sub_div.items():
    print(f"There are {len(div)} divisions in {prov}")
print(f"There are 33 division in Pakistan")
```

```
There are 131 divisions in BALOCHISTAN
There are 71 divisions in KPK
There are 46 divisions in KPK/FATA
There are 143 divisions in PUNJAB
There are 137 divisions in SINDH
There are 33 division in Pakistan
```

```python
sub_div = pd.DataFrame({'Punjab':[143],
                        'Sindh':[137],
                        'Balochistan':[131],
                        'KPK':[71],
                        'KPK/FATA':[46]})
sub_div
```

	Punjab	Sindh	Balochistan	KPK	KPK/FATA
0	143	137	131	71	46

```python
# Create a new dataset with sub_div_df variable
sub_div_df = pd.DataFrame(sub_div)
# use the melt function to reshape the data
div_df = pd.melt(sub_div_df, var_name='Province', value_name='Sub-Division')
# Display the reshape Dataframe
print(div_df)
```

```
     Province  Sub-Division
0      Punjab           143
1       Sindh           137
2  Balochistan          131
3         KPK            71
4     KPK/FATA           46
```

```python
# make plot on melt_df by using plotly library
import plotly.express as px

fig = px.bar(div_df, x='Province', y='Sub-Division', color='Province')
fig.update_traces(width=0.5)
fig.update_layout(width=700, height=500)
# x axis ticks in the 90 degree angle
fig.update_layout(xaxis_tickangle=0)
# graph title name
fig.update_layout(title_text='Pakistan total Sub-Divisions are (528), Sub-divisions in each Province')

fig.show()
```

Pakistan total Sub-Divisions are (528), Sub-divisions in each Province

Interpretation

The presented bar graph provides a comprehensive representation of the sub-divisions within Pakistan. The data indicates that the nation, as a whole, comprises a total of 528 sub-divisions. Notably, the highest concentration of sub-divisions is observed within the Punjab province, with 143 sub-divisions. Sindh follows closely with 137, while Balochistan houses 131 sub-divisions. In contrast, Khyber Pakhtunkhwa (KPK) contains 71 sub-divisions, and the KPK/FATA region features 46 sub-divisions, as revealed by the data.

Question 8. How many Sub Divisions in each province of Pakistan?

Answer. There are in total 528 sub-divisions in pakistan, Punjab (143), Sindh (137), Balochistan (131), KPK (71), & KPK/FATA (46).
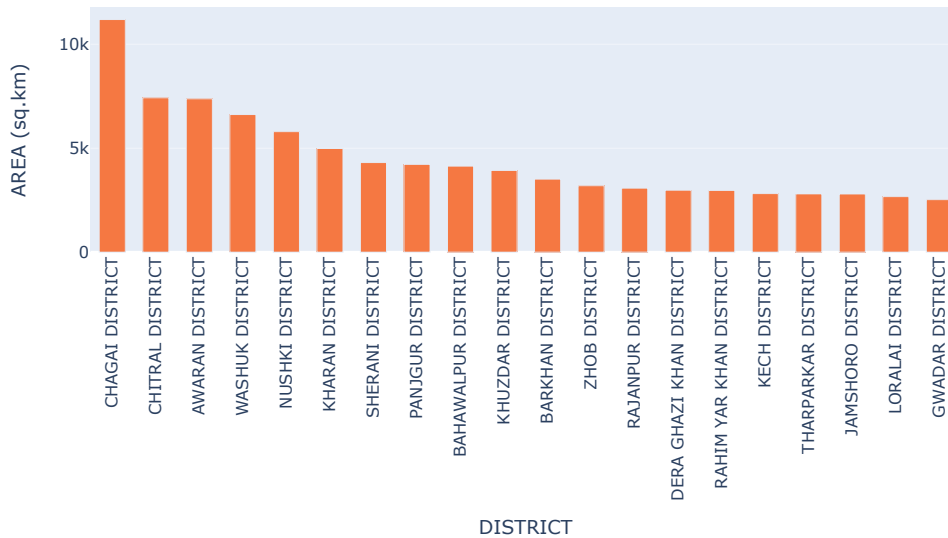
Show top 20 Districts AREA square (km) in provinces of Pakistan

```python
# Calculate the area square of each district and select the top 10
top_10_districts = df.groupby('DISTRICT')['AREA (sq.km)'].mean().reset_index().nlargest(20, 'AREA (sq.km)')

fig = px.bar(top_10_districts, x='DISTRICT', y='AREA (sq.km)',
             title='Top 10 Districts by Area Square (KM)',
             color_discrete_sequence=['#f57842'])
fig.update_traces(width=0.6)
fig.update_layout(width=800, height=500)
# x axis ticks in the 90 degree angle
fig.update_layout(xaxis_tickangle=-90)
```

```
fig.update_xaxes(categoryorder='total descending')   # Optional: Sort the x-axis categories
fig.show()
```

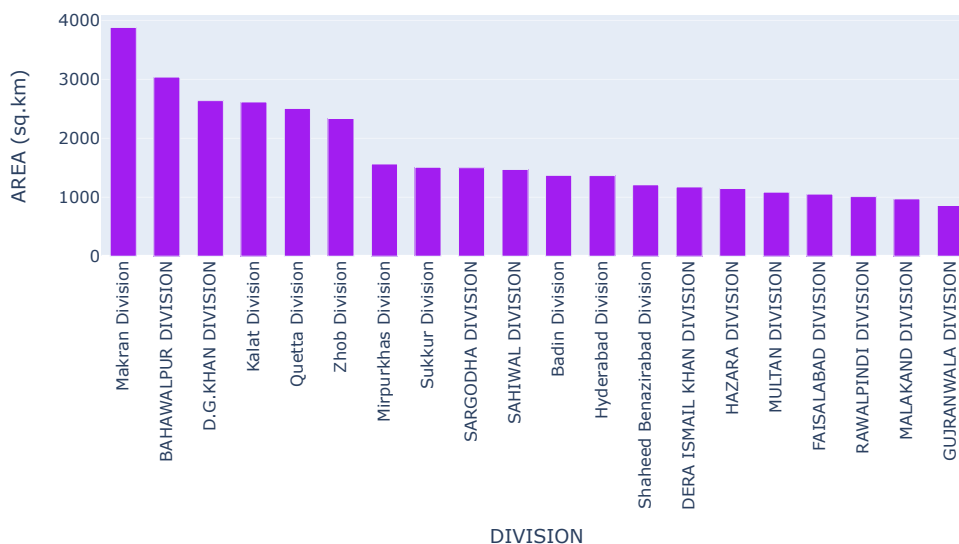Top 10 Districts by Area Square (KM)

## Show top 20 Division based on Area square (km) in provinces of Pakistan

```
In [ ]:  # Calculate the area square of each district and select the top 10
         top_10_divisions = df.groupby('DIVISION')['AREA (sq.km)'].mean().reset_index().nlargest(20, 'AREA (sq.km)')

         fig = px.bar(top_10_divisions, x='DIVISION', y='AREA (sq.km)',
                      title='Top 10 Divisions by Area Square (KM)',
                      color_discrete_sequence=['#a21df0'])
         fig.update_traces(width=0.6)
         fig.update_layout(width=800, height=500)
         # x axis ticks in the 90 degree angle
         fig.update_layout(xaxis_tickangle=-90)

         fig.update_xaxes(categoryorder='total descending')   # Optional: Sort the x-axis categories
         fig.show()
```

Top 10 Divisions by Area Square (KM)

Show top 20 Sub-Divisions based on Area square (km) in provinces of Pakistan

```
In [ ]:  # Calculate the area square of each district and select the top 10
         top_10_sub_divisions = df.groupby('SUB DIVISION')['AREA (sq.km)'].mean().reset_index().nlargest(20, 'AREA (sq.km)')

         fig = px.bar(top_10_sub_divisions, x='SUB DIVISION', y='AREA (sq.km)',
                      title='Top 10 SUB-Divisions by Area Square (KM)',
                      color_discrete_sequence=['#fa0791'])
         fig.update_traces(width=0.6)
         fig.update_layout(width=800, height=500)
         # x axis ticks in the 90 degree angle
```

```
fig.update_layout(xaxis_tickangle=-90)

fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories
fig.show()
```

Top 10 SUB-Divisions by Area Square (KM)
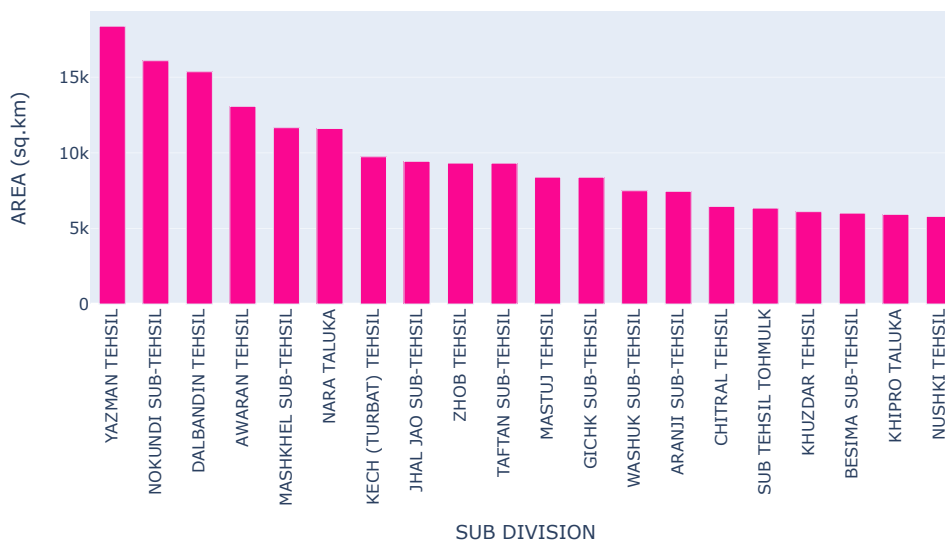
Question 9. What is the Gender Distribution in each province of Pakistan?

```
In [ ]: import pandas as pd

        # Your existing code
        df_gender = df[['MALE (RURAL)','MALE (URBAN)', 'FEMALE (RURAL)','FEMALE (URBAN)','TRANSGENDER (RURAL)','TRANSGENDER (URBAN)']]
        df_gender.head()
        df_sex = pd.DataFrame(df_gender)

        # Add the "PROVINCE" column to the DataFrame
        df_sex['Province'] = df['PROVINCE']

        # Use the melt function to reshape the data
        melted_df = pd.melt(df_sex, id_vars='Province', var_name='Gender', value_name='value')

        # Display the reshaped DataFrame
        print(melted_df)
             Province            Gender   value
        0      PUNJAB      MALE (RURAL)  316864
        1      PUNJAB      MALE (RURAL)  273788
        2      PUNJAB      MALE (RURAL)  182655
        3      PUNJAB      MALE (RURAL)  192278
        4      PUNJAB      MALE (RURAL)  231506
        ...       ...               ...     ...
        3163  KPK/FATA  TRANSGENDER (URBAN)       0
        3164  KPK/FATA  TRANSGENDER (URBAN)       0
        3165  KPK/FATA  TRANSGENDER (URBAN)       0
        3166  KPK/FATA  TRANSGENDER (URBAN)       0
        3167  KPK/FATA  TRANSGENDER (URBAN)       0

        [3168 rows x 3 columns]
```

```
In [ ]: import plotly.express as px

        # Calculate the mean values for 'value' by 'sex'
        mean_values = melted_df.groupby(['Gender', 'Province'])['value'].mean().reset_index()

        # Define custom colors for 'Rural' and 'Urban'
        colors = {'PUNJAB': '#de1f4f','BALOCHISTAN': '#f79011',  'SINDH': '#07fad1', 'KPK': '#07a9fa', 'KPK/FATA': '#e207fa'}

        # Create a bar plot using Plotly Express with facet_col
        # Create a grouped bar chart with custom colors and text labels
        fig = px.bar(mean_values, x="Province", y="value", color="Gender", barmode="group",
                     color_discrete_map=colors,
                     title="The average values for gender distribution in each province, encompassing both rural and urban demographics")

        # bar width size
        fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories
        fig.show()
```
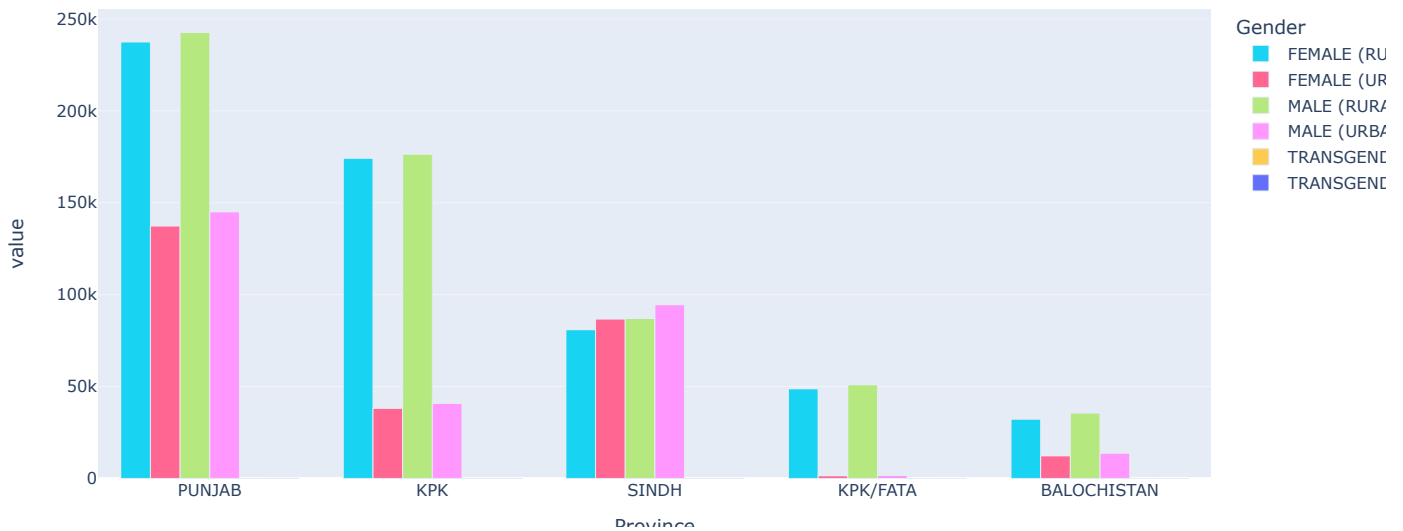
The average values for gender distribution in each province, encompassing both rural and urban demographics

Interpretation

The provided bar graph illustrates gender distribution within the provinces of Pakistan. It encompasses data from five distinct provinces and delineates the representation of three gender categories: Male, Female, and Transgender, across two distinct settings: Rural and Urban.

A discernible pattern emerges, revealing a notable concentration of both Male and Female populations in rural areas, as compared to their urban counterparts. This discrepancy suggests a potential correlation between rural residence and limited access to educational resources, particularly in the context of family planning. Thus, the graph serves as an effective visual representation of the prevailing conditions in Pakistan.

Conversely, the Transgender group exhibits a less pronounced concentration, rendering it less prominent in the graphical depiction. Further analysis of the data indicates that the highest density of Male and Female individuals is found in the rural areas of Punjab and Khyber Pakhtunkhwa (KPK) provinces, while in Sindh, the distribution of Male and Female individuals across both rural and urban regions remains relatively uniform. Additionally, within KPK/Federally Administered Tribal Areas (FATA), rural areas exhibit a higher concentration of Male and Female populations, followed by Balochistan.

Question 9. What is the Gender Distribution in each province of Pakistan?

Answer. see the above interpretation section.

Question 10. What is the Average House Hold both in (Rural & Urban) in each province of Pakistan?

```python
import pandas as pd

# Define the provinces and categories
provinces = ['PUNJAB', 'SINDH', 'BALOCHISTAN', 'KPK', 'KPK/FATA']
categories = ['Rural', 'Urban']

# Create an empty DataFrame
data = []

# Define values for each province and category
values = {
    'PUNJAB': {'Rural': 6.15, 'Urban': 6.13},
    'SINDH': {'Rural': 4.36, 'Urban': 5.28},
    'BALOCHISTAN': {'Rural': 6.60, 'Urban': 2.92},
    'KPK': {'Rural': 8.09, 'Urban': 4.30},
    'KPK/FATA': {'Rural': 8.68, 'Urban': 0.86}
}

# Populate the DataFrame with values
for province in provinces:
    for category in categories:
        value = values[province][category]
        data.append([province, category, value])

# Create the DataFrame
avg_house_prov = pd.DataFrame(data, columns=['Province', 'Category', 'Average Household/Province'])

# Print the DataFrame
print(avg_house_prov)
```

```
        Province Category  Average Household/Province
0       PUNJAB    Rural                          6.15
1       PUNJAB    Urban                          6.13
2        SINDH    Rural                          4.36
3        SINDH    Urban                          5.28
4   BALOCHISTAN   Rural                          6.60
5   BALOCHISTAN   Urban                          2.92
6          KPK    Rural                          8.09
7          KPK    Urban                          4.30
8     KPK/FATA    Rural                          8.68
9     KPK/FATA    Urban                          0.86
```

```python
In [ ]:  import plotly.express as px

         # Use the DataFrame 'df' created in the previous code

         # Define custom colors for 'Rural' and 'Urban'
         colors = {'Rural': '#fa5807', 'Urban': '#fad107'}

         # Create a grouped bar chart with custom colors and text labels
         fig = px.bar(avg_house_prov, x="Province", y="Average Household/Province", color="Category", barmode="group",
                     color_discrete_map=colors,
                     title="Province-wise Avg. house hold values for Rural & Urban Categories")

         # Add text labels on top of each bar
         fig.update_traces(text=avg_house_prov['Average Household/Province'], texttemplate='%{text:.2f}', textposition='outside')

         # Customize the appearance of the chart
         fig.update_traces(width=0.4)
         fig.update_layout(width=700, height=500, xaxis_title="Province", yaxis_title="Average Household/Province")
         fig.show()
```

Province-wise Avg. house hold values for Rural & Urban Categories
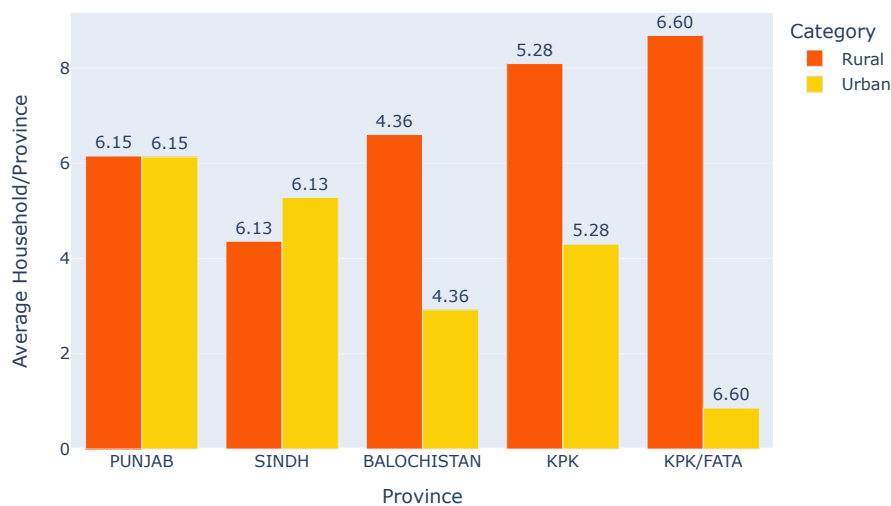
Interpretation

The provided bar graph presents data concerning the average household size in both rural and urban areas within each province of Pakistan. The x-axis denotes the five distinct provinces of Pakistan, while the y-axis represents the average household size.

Evidently, a conspicuous trend emerges, illustrating that rural households, on average, accommodate more individuals than their urban counterparts. This contrast is particularly striking, with the exception of Punjab province, where both rural and urban households share an identical average size of 6.15 persons per household. In contrast, a marginal decrease in household size is observed in Sindh in comparison to urban areas.

Conversely, a notable divergence is observed in the remaining three provinces, namely Balochistan, Khyber Pakhtunkhwa (KPK), and KPK/Federally Administered Tribal Areas (FATA), where rural households exhibit a substantial increase in size, while urban households tend to be smaller by comparison.

Question 10. What is the Average House Hold both in (Rural & Urban) in each province of Pakistan?

Answer. see the above interpretation section.

Question 11. Show population of Pakistan in 1998 (Rural & Urban)

```python
In [ ]:  import pandas as pd

         # Your existing code
         pak_pop_prov = df[['POPULATION 1998 (RURAL)', 'POPULATION 1998 (URBAN)']]
         pak_pop_prov = pd.DataFrame(pak_pop_prov)

         # Add the "PROVINCE" column to the DataFrame
```

```
pak_pop_prov['Province'] = df['PROVINCE']

# Use the melt function to reshape the data
melted_pak_pop_prov = pd.melt(df_sex, id_vars='Province', var_name='Pakistan Population', value_name='value')

# Display the reshaped DataFrame
print(melted_pak_pop_prov)
```

```
      Province  Pakistan Population     value
0       PUNJAB         MALE (RURAL)    316864
1       PUNJAB         MALE (RURAL)    273788
2       PUNJAB         MALE (RURAL)    182655
3       PUNJAB         MALE (RURAL)    192278
4       PUNJAB         MALE (RURAL)    231506
...        ...                  ...       ...
3163  KPK/FATA  TRANSGENDER (URBAN)        0
3164  KPK/FATA  TRANSGENDER (URBAN)        0
3165  KPK/FATA  TRANSGENDER (URBAN)        0
3166  KPK/FATA  TRANSGENDER (URBAN)        0
3167  KPK/FATA  TRANSGENDER (URBAN)        0

[3168 rows x 3 columns]
```

In [ ]:
```
# calculate mean of the population in 1998 rural vs urban
print("Population of rural in 1998:", df['POPULATION 1998 (RURAL)'].mean())
print("Population of urban in 1998:", df['POPULATION 1998 (URBAN)'].mean())
```

```
Population of rural in 1998: 167427.99431818182
Population of urban in 1998: 80144.56628787878
```

In [ ]:
```
pak_pop = pd.DataFrame({'Rural Population (1998)':[167427.99],
                        'Urban population (1998)':[80144.56]})
pak_pop
```

Out[ ]:

	Rural Population (1998)	Urban population (1998)
0	167427.99	80144.56

In [ ]:
```
pak_pop = pd.DataFrame(pak_pop)
# use the melt function to reshape the data
pak_pop = pd.melt(pak_pop, var_name='Population', value_name='value')
# Display the reshape Dataframe
print(pak_pop)
```

```
               Population     value
0  Rural Population (1998)  167427.99
1  Urban population (1998)   80144.56
```

In [ ]:
```
import plotly.express as px

# Create a bar chart using Plotly Express
fig = px.bar(pak_pop, x='Population', y='value', title='Pakistan population (1998), both in Rural & Urban')

# Customize the appearance of the bar chart
fig.update_traces(marker=dict(color=['#07eb93', '#f7dd11']), width=0.3)
fig.update_layout(width=600, height=400)
fig.update_layout(xaxis_tickangle=0)

fig.show()
```


Pakistan population (1998), both in Rural & Urban
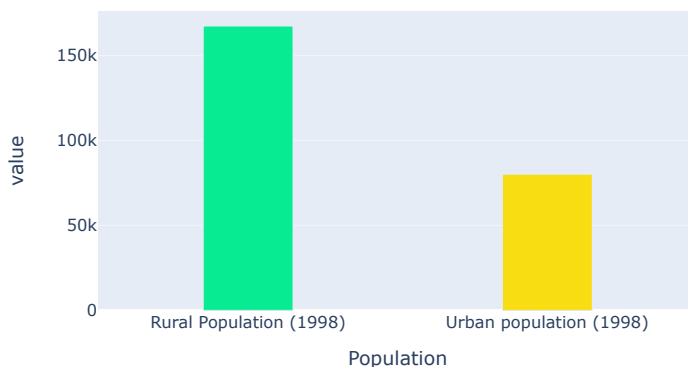
Question 12. Show Population of Pakistan in 1998, broken down into (rural & urban), for each province of Pakistan.

In [ ]:
```
import pandas as pd

# Your existing code
pop_pak_prov = df[['POPULATION 1998 (RURAL)', 'POPULATION 1998 (URBAN)']]
pop_pak_prov = pd.DataFrame(pop_pak_prov)
```

```python
# Add the "PROVINCE" column to the DataFrame
pop_pak_prov['Province'] = df['PROVINCE']

# Use the melt function to reshape the data
melt_pop_pak_prov = pd.melt(pop_pak_prov, id_vars='Province', var_name='Population', value_name='Mean value')

# Display the reshaped DataFrame
print(melt_pop_pak_prov)
```

```
      Province              Population  Mean value
0       PUNJAB  POPULATION 1998 (RURAL)      407768
1       PUNJAB  POPULATION 1998 (RURAL)      395983
2       PUNJAB  POPULATION 1998 (RURAL)      250959
3       PUNJAB  POPULATION 1998 (RURAL)      297343
4       PUNJAB  POPULATION 1998 (RURAL)      316593
...        ...                      ...         ...
1051  KPK/FATA  POPULATION 1998 (URBAN)           0
1052  KPK/FATA  POPULATION 1998 (URBAN)           0
1053  KPK/FATA  POPULATION 1998 (URBAN)           0
1054  KPK/FATA  POPULATION 1998 (URBAN)           0
1055  KPK/FATA  POPULATION 1998 (URBAN)           0

[1056 rows x 3 columns]
```

In [ ]:
```python
# Group by the two categorical columns and calculate the mean
result = melt_pop_pak_prov.groupby(['Province', 'Population'])['Mean value'].mean().reset_index()
result
```

Out[ ]:

	Province	Population	Mean value
0	BALOCHISTAN	POPULATION 1998 (RURAL)	37944.450382
1	BALOCHISTAN	POPULATION 1998 (URBAN)	11975.419847
2	KPK	POPULATION 1998 (RURAL)	209638.816901
3	KPK	POPULATION 1998 (URBAN)	40732.830986
4	KPK/FATA	POPULATION 1998 (RURAL)	62657.086957
5	KPK/FATA	POPULATION 1998 (URBAN)	1858.108696
6	PUNJAB	POPULATION 1998 (RURAL)	350172.034965
7	PUNJAB	POPULATION 1998 (URBAN)	160350.944056
8	SINDH	POPULATION 1998 (RURAL)	113796.167883
9	SINDH	POPULATION 1998 (URBAN)	108320.160584

In [ ]:
```python
import plotly.express as px

# Define custom colors for 'Rural' and 'Urban'
colors = {'POPULATION 1998 (RURAL)': '#45f5dd', 'POPULATION 1998 (URBAN)': '#f545d4'}

# Create a grouped bar chart with custom colors and text labels
fig = px.bar(result, x="Province", y="Mean value", color="Population", barmode="group",
             color_discrete_sequence=[colors['POPULATION 1998 (RURAL)'], colors['POPULATION 1998 (URBAN)']],
             title="A cumulative Pakistan population in (1998), (Rural & Urban) in each Province")

fig.update_layout(height=500, width=900)
fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories

fig.show()
```
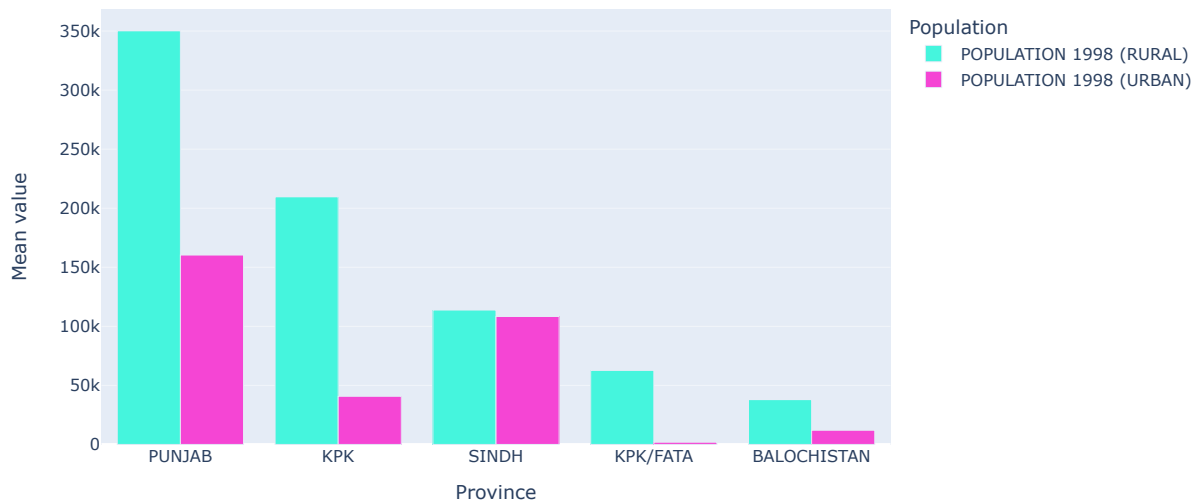
A cumulative Pakistan population in (1998), (Rural & Urban) in each Province

Question 13. Show sex ratio of Pakistan in both (Rural & Urban)

```
In [ ]:  # calculate mean of the population in 1998 rural vs urban
         print("Sex ratio in Rural:", df['SEX RATIO (RURAL)'].mean())
         print("Sex ratio in urban:", df['SEX RATIO (URBAN)'].mean())
```

```
Sex ratio in Rural: 98.98261363636364
Sex ratio in urban: 75.41126893939395
```

```
In [ ]:  sex_ratio = pd.DataFrame({'Sex Ratio (Rural)':[98.98],
                                   'Sex Ration (Urban)':[75.41]})
         sex_ratio
```

Out[ ]:

	Sex Ratio (Rural)	Sex Ration (Urban)
0	98.98	75.41

```
In [ ]:  sex_ratio = pd.DataFrame(sex_ratio)
         # use the melt function to reshape the data
         sex_ratio = pd.melt(sex_ratio, var_name='Sex Ratio', value_name='value')
         # Display the reshape Dataframe
         print(sex_ratio)
```

```
            Sex Ratio  value
0   Sex Ratio (Rural)  98.98
1   Sex Ration (Urban)  75.41
```

```
In [ ]:  import plotly.express as px

         # Create a bar chart using Plotly Express
         fig = px.bar(sex_ratio, x='Sex Ratio', y='value', title='Annual Growth Rate in Rural and Urban Areas')

         # Customize the appearance of the bar chart
         fig.update_traces(marker=dict(color=['#07c1eb', '#f7dd11']), width=0.3)
         fig.update_layout(width=600, height=400)
         fig.update_layout(xaxis_tickangle=0)

         fig.show()
```

Annual Growth Rate in Rural and Urban Areas
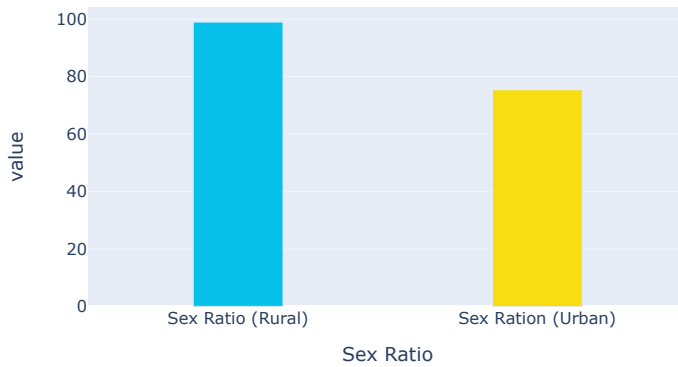
Question 14. Show sex ratio (Rural & Urban) in five provinces of Pakistan.

```python
import pandas as pd

# Your existing code
ann_sex_ratio_pro = df[['SEX RATIO (RURAL)', 'SEX RATIO (URBAN)']]
ann_sex_ratio_pro = pd.DataFrame(ann_sex_ratio_pro)

# Add the "PROVINCE" column to the DataFrame
ann_sex_ratio_pro['Province'] = df['PROVINCE']

# Use the melt function to reshape the data
melt_ann_sex_ratio_pro = pd.melt(ann_sex_ratio_pro, id_vars='Province', var_name='sex ratio', value_name='value')

# Display the reshaped DataFrame
print(melt_ann_sex_ratio_pro)
```

```
      Province        sex ratio   value
0       PUNJAB  SEX RATIO (RURAL)  104.70
1       PUNJAB  SEX RATIO (RURAL)  102.73
2       PUNJAB  SEX RATIO (RURAL)  102.30
3       PUNJAB  SEX RATIO (RURAL)  101.30
4       PUNJAB  SEX RATIO (RURAL)  104.67
...        ...                ...     ...
1051  KPK/FATA  SEX RATIO (URBAN)    0.00
1052  KPK/FATA  SEX RATIO (URBAN)    0.00
1053  KPK/FATA  SEX RATIO (URBAN)    0.00
1054  KPK/FATA  SEX RATIO (URBAN)    0.00
1055  KPK/FATA  SEX RATIO (URBAN)    0.00

[1056 rows x 3 columns]
```

```python
# Group by the two categorical columns and calculate the mean
result = melt_ann_sex_ratio_pro.groupby(['Province', 'sex ratio'])['value'].mean().reset_index()
result
```

Out[ ]:

|	Province	sex ratio	value
0	BALOCHISTAN	SEX RATIO (RURAL)	111.259466
1	BALOCHISTAN	SEX RATIO (URBAN)	47.571069
2	KPK	SEX RATIO (RURAL)	101.510423
3	KPK	SEX RATIO (URBAN)	59.512254
4	KPK/FATA	SEX RATIO (RURAL)	105.427609
5	KPK/FATA	SEX RATIO (URBAN)	15.611739
6	PUNJAB	SEX RATIO (RURAL)	97.235874
7	PUNJAB	SEX RATIO (URBAN)	101.496364
8	SINDH	SEX RATIO (RURAL)	85.592628
9	SINDH	SEX RATIO (URBAN)	103.122993

```python
import plotly.express as px

# Define custom colors for 'Rural' and 'Urban'
colors = {'Rural': '#6845f5', 'Urban': '#f5e045'}

# Create a grouped bar chart with custom colors and text labels
fig = px.bar(result, x="Province", y="value", color="sex ratio", barmode="group",
             color_discrete_sequence=[colors['Rural'], colors['Urban']],
             title="Average cumulative sex ratio (Rural & Urban) in each Province of Pakistan")

fig.update_layout(height=500, width=900)
fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories
```

```
fig.show()
```

Average cumulative sex ratio (Rural & Urban) in each Province of Pakistan
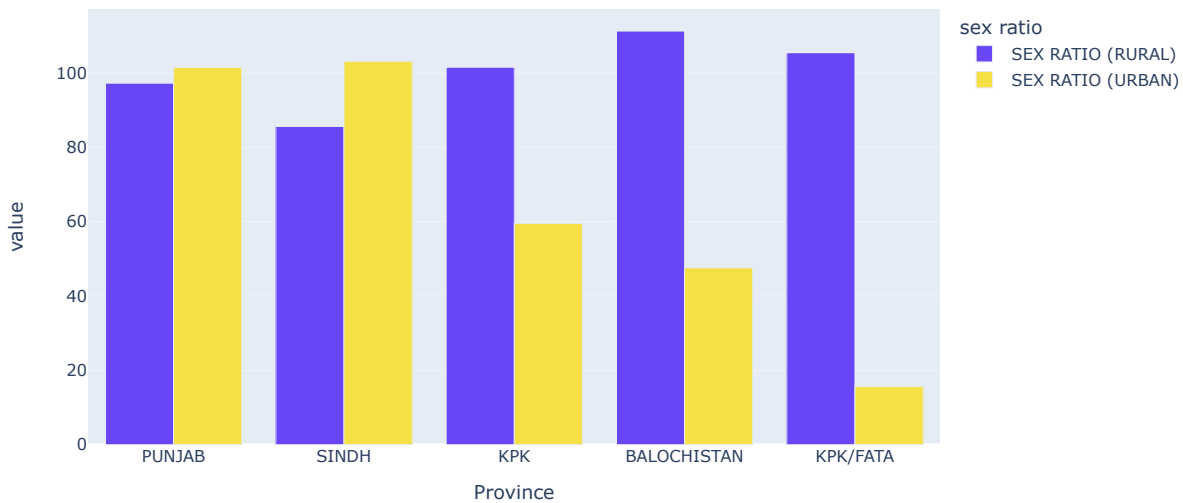
Question 15. Show annual growth rate of Pakistan in both (Rural & Urban)

```python
In [ ]:  # calculate mean of the population in 1998 rural vs urban
         print("Sex ratio in Rural:", df['ANNUAL GROWTH RATE (RURAL)'].mean())
         print("Sex ratio in urban:", df['ANNUAL GROWTH RATE (URBAN)'].mean())
```

```
Sex ratio in Rural: 3.1247916666666664
Sex ratio in urban: 1.920814393939394
```

```python
In [ ]:  ann_grw = pd.DataFrame({'ANNUAL GROWTH RATE (RURAL)':[3.12],
                                 'ANNUAL GROWTH RATE (URBAN)':[1.92]})
         ann_grw
```

Out[ ]:

	ANNUAL GROWTH RATE (RURAL)	ANNUAL GROWTH RATE (URBAN)
0	3.12	1.92

```python
In [ ]:  ann_grw = pd.DataFrame(ann_grw)
         # use the melt function to reshape the data
         ann_grw = pd.melt(ann_grw, var_name='Annual Growth', value_name='value')
         # Display the reshape Dataframe
         print(ann_grw)
```

```
            Annual Growth  value
0  ANNUAL GROWTH RATE (RURAL)   3.12
1  ANNUAL GROWTH RATE (URBAN)   1.92
```

```python
In [ ]:  import plotly.express as px

         # Create a bar chart using Plotly Express
         fig = px.bar(ann_grw, x='Annual Growth', y='value', title='Cumulative Average Annual Growth Rate in Rural & Urban Areas')

         # Customize the appearance of the bar chart
         fig.update_traces(marker=dict(color=['#eb07d4', '#f7dd11']), width=0.3)
         fig.update_layout(width=600, height=400)
         fig.update_layout(xaxis_tickangle=0)

         fig.show()
```

Cumulative Average Annual Growth Rate in Rural & Urban Areas
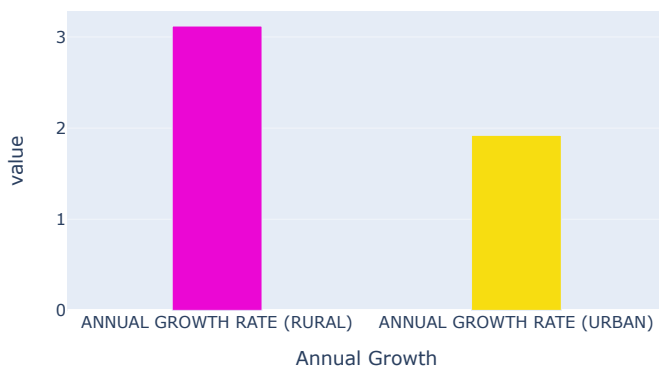
Question 16. Show average annual growth rate of pakistan both (Rural & Urban) in each province

```
In [ ]: import pandas as pd

        # Your existing code
        ann_pak_pop_prov = df[['ANNUAL GROWTH RATE (RURAL)', 'ANNUAL GROWTH RATE (URBAN)']]
        ann_pak_pop_prov = pd.DataFrame(ann_pak_pop_prov)

        # Add the "PROVINCE" column to the DataFrame
        ann_pak_pop_prov['Province'] = df['PROVINCE']

        # Use the melt function to reshape the data
        melted_ann_pak_pop_prov = pd.melt(ann_pak_pop_prov, id_vars='Province', var_name='Annual Growth', value_name='value')

        # Display the reshaped DataFrame
        print(melted_ann_pak_pop_prov)

             Province           Annual Growth  value
        0      PUNJAB  ANNUAL GROWTH RATE (RURAL)   2.22
        1      PUNJAB  ANNUAL GROWTH RATE (RURAL)   1.65
        2      PUNJAB  ANNUAL GROWTH RATE (RURAL)   1.93
        3      PUNJAB  ANNUAL GROWTH RATE (RURAL)   1.33
        4      PUNJAB  ANNUAL GROWTH RATE (RURAL)   1.90
        ...       ...                      ...    ...
        1051  KPK/FATA  ANNUAL GROWTH RATE (URBAN)   0.00
        1052  KPK/FATA  ANNUAL GROWTH RATE (URBAN)   0.00
        1053  KPK/FATA  ANNUAL GROWTH RATE (URBAN)   0.00
        1054  KPK/FATA  ANNUAL GROWTH RATE (URBAN)   0.00
        1055  KPK/FATA  ANNUAL GROWTH RATE (URBAN)   0.00

        [1056 rows x 3 columns]
```

```
In [ ]: # Group by the two categorical columns and calculate the mean
        result = melted_ann_pak_pop_prov.groupby(['Province', 'Annual Growth'])['value'].mean().reset_index()
        result
```

Out[ ]:

	Province	Annual Growth	value
0	BALOCHISTAN	ANNUAL GROWTH RATE (RURAL)	2.649389
1	BALOCHISTAN	ANNUAL GROWTH RATE (URBAN)	1.071603
2	KPK	ANNUAL GROWTH RATE (RURAL)	2.599296
3	KPK	ANNUAL GROWTH RATE (URBAN)	1.329014
4	KPK/FATA	ANNUAL GROWTH RATE (RURAL)	2.884565
5	KPK/FATA	ANNUAL GROWTH RATE (URBAN)	0.280652
6	PUNJAB	ANNUAL GROWTH RATE (RURAL)	5.267762
7	PUNJAB	ANNUAL GROWTH RATE (URBAN)	2.724126
8	SINDH	ANNUAL GROWTH RATE (RURAL)	1.695547
9	SINDH	ANNUAL GROWTH RATE (URBAN)	2.751752

```
In [ ]: import plotly.express as px
        # Define custom colors for 'Rural' and 'Urban'
        colors = {'Rural': '#fa5807', 'Urban': '#fad107'}

        # Create a grouped bar chart with custom colors and text labels
        fig = px.bar(result, x="Province", y="value", color="Annual Growth", barmode="group",
                     color_discrete_map=colors,
                     title="Average cumulative Annual Growth rate (Rural & Urban) in each province")

        fig.update_layout(height=500, width=900)
        fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories
```
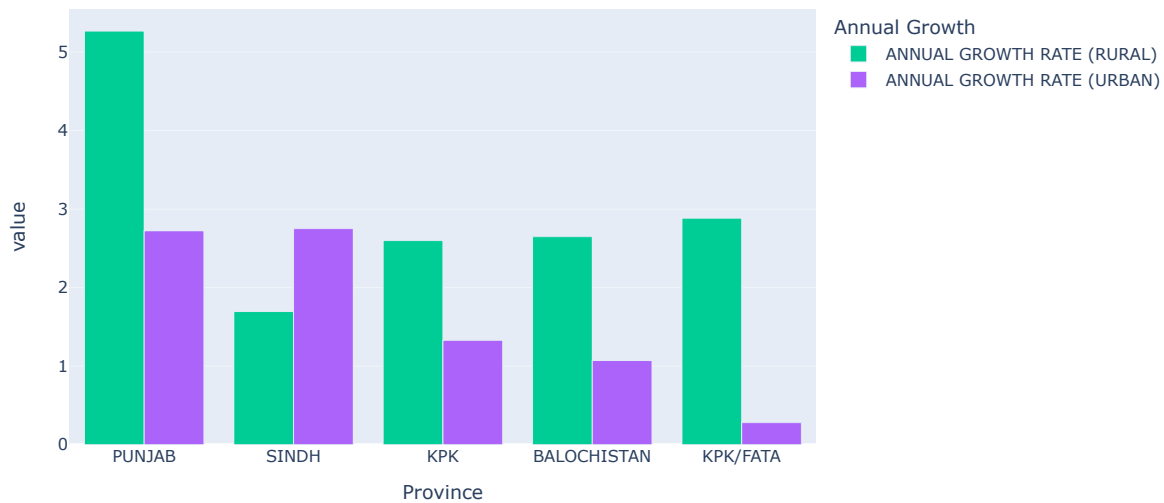
```
fig.show()
```

Average cumulative Annual Growth rate (Rural & Urban) in each province

Question 17. Show average annual growth rate of Pakistan both (RURAL & URBAN) in each Division

```python
import pandas as pd

# Your existing code
ann_pak_pop_div = df[['ANNUAL GROWTH RATE (RURAL)', 'ANNUAL GROWTH RATE (URBAN)']]
ann_pak_pop_div = pd.DataFrame(ann_pak_pop_div)

# Add the "PROVINCE" column to the DataFrame
ann_pak_pop_div['Division'] = df['DIVISION']

# Use the melt function to reshape the data
melted_ann_pak_pop_div = pd.melt(ann_pak_pop_div, id_vars='Division', var_name='Annual Growth', value_name='value')

# Display the reshaped DataFrame
print(melted_ann_pak_pop_div)
```

```
                   Division         Annual Growth  value
0         BAHAWALPUR DIVISION  ANNUAL GROWTH RATE (RURAL)   2.22
1         BAHAWALPUR DIVISION  ANNUAL GROWTH RATE (RURAL)   1.65
2         BAHAWALPUR DIVISION  ANNUAL GROWTH RATE (RURAL)   1.93
3         BAHAWALPUR DIVISION  ANNUAL GROWTH RATE (RURAL)   1.33
4         BAHAWALPUR DIVISION  ANNUAL GROWTH RATE (RURAL)   1.90
...                       ...                      ...    ...
1051  DERA ISMAIL KHAN DIVISION  ANNUAL GROWTH RATE (URBAN)   0.00
1052  DERA ISMAIL KHAN DIVISION  ANNUAL GROWTH RATE (URBAN)   0.00
1053  DERA ISMAIL KHAN DIVISION  ANNUAL GROWTH RATE (URBAN)   0.00
1054  DERA ISMAIL KHAN DIVISION  ANNUAL GROWTH RATE (URBAN)   0.00
1055  DERA ISMAIL KHAN DIVISION  ANNUAL GROWTH RATE (URBAN)   0.00

[1056 rows x 3 columns]
```

```python
# Group by the two categorical columns and calculate the mean
result = melted_ann_pak_pop_div.groupby(['Division', 'Annual Growth'])['value'].mean().reset_index()
result
```

	Division	Annual Growth	value
0	BAHAWALPUR DIVISION	ANNUAL GROWTH RATE (RURAL)	1.824667
1	BAHAWALPUR DIVISION	ANNUAL GROWTH RATE (URBAN)	2.728667
2	BANNU DIVISION	ANNUAL GROWTH RATE (RURAL)	3.859333
3	BANNU DIVISION	ANNUAL GROWTH RATE (URBAN)	0.576667
4	Badin Division	ANNUAL GROWTH RATE (RURAL)	2.272000
5	Badin Division	ANNUAL GROWTH RATE (URBAN)	4.420000
6	D.G.KHAN DIVISION	ANNUAL GROWTH RATE (RURAL)	2.714286
7	D.G.KHAN DIVISION	ANNUAL GROWTH RATE (URBAN)	2.614286
8	DERA ISMAIL KHAN DIVISION	ANNUAL GROWTH RATE (RURAL)	2.598667
9	DERA ISMAIL KHAN DIVISION	ANNUAL GROWTH RATE (URBAN)	1.109333
10	FAISALABAD DIVISION	ANNUAL GROWTH RATE (RURAL)	1.908824
11	FAISALABAD DIVISION	ANNUAL GROWTH RATE (URBAN)	2.384706
12	GUJRANWALA DIVISION	ANNUAL GROWTH RATE (RURAL)	1.616500
13	GUJRANWALA DIVISION	ANNUAL GROWTH RATE (URBAN)	2.686500
14	HAZARA DIVISION	ANNUAL GROWTH RATE (RURAL)	1.999333
15	HAZARA DIVISION	ANNUAL GROWTH RATE (URBAN)	0.864000
16	Hyderabad Division	ANNUAL GROWTH RATE (RURAL)	1.617333
17	Hyderabad Division	ANNUAL GROWTH RATE (URBAN)	2.937667
18	KOHAT DIVISION	ANNUAL GROWTH RATE (RURAL)	2.177143
19	KOHAT DIVISION	ANNUAL GROWTH RATE (URBAN)	1.384286
20	Kalat Division	ANNUAL GROWTH RATE (RURAL)	2.779444
21	Kalat Division	ANNUAL GROWTH RATE (URBAN)	1.260556
22	Karachi Division	ANNUAL GROWTH RATE (RURAL)	0.959667
23	Karachi Division	ANNUAL GROWTH RATE (URBAN)	2.113333
24	LAHORE DIVISION	ANNUAL GROWTH RATE (RURAL)	30.595882
25	LAHORE DIVISION	ANNUAL GROWTH RATE (URBAN)	3.884706
26	Larkana Division	ANNUAL GROWTH RATE (RURAL)	1.619048
27	Larkana Division	ANNUAL GROWTH RATE (URBAN)	3.129524
28	MALAKAND DIVISION	ANNUAL GROWTH RATE (RURAL)	2.683939
29	MALAKAND DIVISION	ANNUAL GROWTH RATE (URBAN)	0.449091
30	MARDAN DIVISION	ANNUAL GROWTH RATE (RURAL)	2.560000
31	MARDAN DIVISION	ANNUAL GROWTH RATE (URBAN)	1.984286
32	MULTAN DIVISION	ANNUAL GROWTH RATE (RURAL)	1.807857
33	MULTAN DIVISION	ANNUAL GROWTH RATE (URBAN)	2.476429
34	Makran Division	ANNUAL GROWTH RATE (RURAL)	2.281351
35	Makran Division	ANNUAL GROWTH RATE (URBAN)	1.182432
36	Mirpurkhas Division	ANNUAL GROWTH RATE (RURAL)	2.540556
37	Mirpurkhas Division	ANNUAL GROWTH RATE (URBAN)	1.837222
38	Naseerabad Division	ANNUAL GROWTH RATE (RURAL)	2.207368
39	Naseerabad Division	ANNUAL GROWTH RATE (URBAN)	1.009474
40	PESHAWAR DIVISION	ANNUAL GROWTH RATE (RURAL)	2.967222
41	PESHAWAR DIVISION	ANNUAL GROWTH RATE (URBAN)	1.162778
42	Quetta Division	ANNUAL GROWTH RATE (RURAL)	3.339474
43	Quetta Division	ANNUAL GROWTH RATE (URBAN)	0.881316
44	RAWALPINDI DIVISION	ANNUAL GROWTH RATE (RURAL)	1.628636
45	RAWALPINDI DIVISION	ANNUAL GROWTH RATE (URBAN)	2.847727
46	SAHIWAL DIVISION	ANNUAL GROWTH RATE (RURAL)	1.462857
47	SAHIWAL DIVISION	ANNUAL GROWTH RATE (URBAN)	2.484286
48	SARGODHA DIVISION	ANNUAL GROWTH RATE (RURAL)	1.860588
49	SARGODHA DIVISION	ANNUAL GROWTH RATE (URBAN)	2.176471
50	Shaheed Benazirabad Division	ANNUAL GROWTH RATE (RURAL)	1.906000
51	Shaheed Benazirabad Division	ANNUAL GROWTH RATE (URBAN)	3.429333
52	Sukkur Division	ANNUAL GROWTH RATE (RURAL)	1.961111
53	Sukkur Division	ANNUAL GROWTH RATE (URBAN)	2.951667
54	Zhob Division	ANNUAL GROWTH RATE (RURAL)	2.304737

|	Division	Annual Growth	value
55	Zhob Division	ANNUAL GROWTH RATE (URBAN)	1.119474

```python
import plotly.express as px
# Define custom colors for 'Rural' and 'Urban'
colors = {'Rural': '#fa5807', 'Urban': '#fad107'}

# Create a grouped bar chart with custom colors and text labels
fig = px.bar(result, x="Division", y="value", color="Annual Growth", barmode="group",
             color_discrete_map=colors,
             title="Average cumulative Annual Growth rate (Rural & Urban) in each Divisions")

fig.update_layout(height=500)
fig.update_xaxes(categoryorder='total descending')  # Optional: Sort the x-axis categories

fig.show()
```

Average cumulative Annual Growth rate (Rural & Urban) in each Divisions
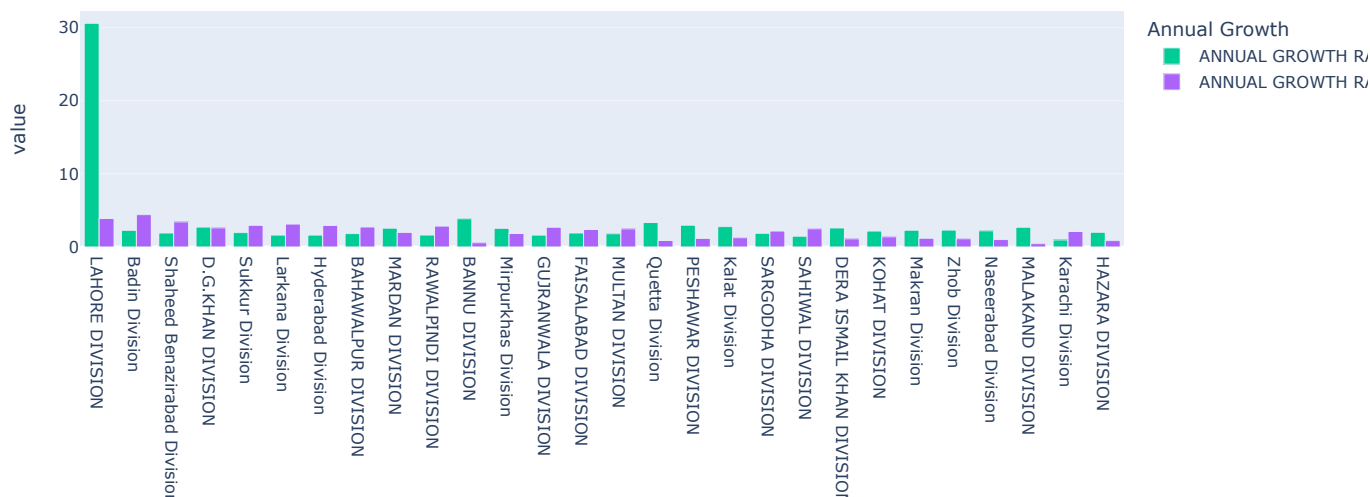
I have successfully completed an extensive data analysis project focused on Pakistan Population from 1998 - 2017. kindly requests your support through an upvote. Your positive feedback would greatly encourage and validate the valuable insights I have uncovered during this endeavor. Your support is highly appreciated and will serve as a motivating factor for his future contributions to the field of data analysis. Thank you for your time and consideration.

For more information please follow me on the following accounts

Twitter

Github

LinkedIn

Gmail