

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Q1: Create a bar chart or histogram to visualize the distribution of a categorical or continuous variable, such as the distribution of ages or genders in a population. 📊

```
In [2]: df=pd.read_csv("population.csv")
df
```

```
Out[2]:
```

| | Country | Total | 65+ | 25-64 years | 15-24 years | 5-14 years | 0-4 years |
|-----|-------------------|----------|---------|-------------|-------------|------------|-----------|
| 0 | Afghanistan | 40099460 | 964016 | 12995277 | 8727160 | 10922454 | 6490554 |
| 1 | Albania | 2854710 | 463442 | 1527360 | 399113 | 323402 | 141393 |
| 2 | Algeria | 44177964 | 2731079 | 21871300 | 6007614 | 8697129 | 4870841 |
| 3 | American Samoa | 45056 | 3088 | 22433 | 7141 | 8365 | 4029 |
| 4 | Andorra | 79057 | 11499 | 48651 | 8570 | 7797 | 2540 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 230 | Wallis and Futuna | 11654 | 1521 | 5767 | 1652 | 1961 | 753 |
| 231 | Western Sahara | 565590 | 31718 | 314200 | 80572 | 92147 | 46953 |
| 232 | Yemen | 32981644 | 886890 | 12210659 | 6733620 | 8436417 | 4714059 |
| 233 | Zambia | 19473132 | 338628 | 6734346 | 3975624 | 5351636 | 3072899 |
| 234 | Zimbabwe | 15993525 | 537915 | 5572892 | 3342436 | 4240570 | 2299712 |

235 rows × 7 columns

```
In [4]: df.shape
```

```
Out[4]: (235, 7)
```

```
In [6]: df.ndim
```

```
Out[6]: 2
```

In [7]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 235 entries, 0 to 234
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Country                235 non-null   object 
 1   Total                  235 non-null   int64  
 2   65+                    235 non-null   int64  
 3   25-64 years            235 non-null   int64  
 4   15-24 years            235 non-null   int64  
 5   5-14 years             235 non-null   int64  
 6   0-4 years              235 non-null   int64  
dtypes: int64(6), object(1)
memory usage: 13.0+ KB
```

In [8]: df.columns

Out[8]: Index(['Country', 'Total', '65+', '25-64 years', '15-24 years', '5-14 years', '0-4 years'], dtype='object')

In [10]: print(df["65+"].unique())
print(df["25-64 years"].unique())
print(df["15-24 years"].unique())
print(df["5-14 years"].unique())
print(df["0-4 years"].unique())

```
[ 964016  463442  2731079   3088   11499   895191   1614
    9508  5352580  355723   16564  4294906  1734558  692682
    34956   51052  9868579   44197  1611927  2254595   19833
    399839   12643   47716   593121    3605   593456   92992
    20534732    2953   25885  1543888   562955   310795   917741
    734810   7066799   31995    5340   136388   347091  2471958
    187501470  4492094   35364   156274    1929   542728   660329
    892052   1763913   27755   180164  2149717  2840766  1186587
    49708    6758   796941  1360115  5216774   514645   51457
    145084   270679   47617  3774984    9477    411   52878
    1267270  13760207   17201   29278   91259   64317   547043
    18491898  1141530    6684  2351054    5327   12277   77716
    19545    862895   10460   454099   57952   49604   513974
    431477   1469182  1982728   55278  95749030  18559456  6484409
    1484419    739454   18578  1061792  14028334   204616  37118444
    17587   415122  1526067  1507360    4829   164730   189459
    286261   323523   405051   538125    96002   173512   324799
    7370    573773   94274   84606   954231   536684  2434270
    23702   533925   99381    1800    81619   150426   159695
    6070   10005730    6666   116553   100000   117000   100000]
```

```
In [13]: df=pd.read_csv("population.csv")

median1=df["65+"].median()
median2=df["25-64 years"].median()
median3=df["15-24 years"].median()
median4=df["5-14 years"].median()
median5=df["0-4 years"].median()

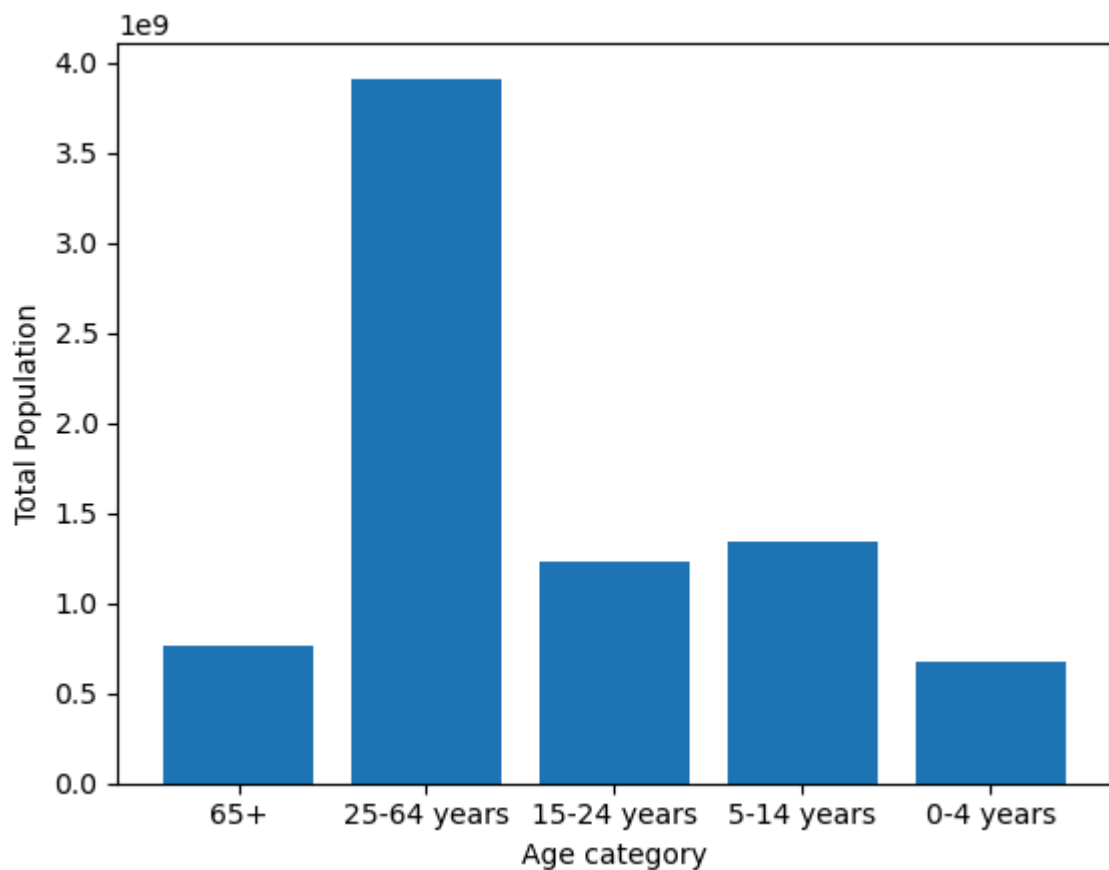
df["65+"].fillna(median1,inplace=True)
df["25-64 years"].fillna(median2,inplace=True)
df["15-24 years"].fillna(median3,inplace=True)
df["5-14 years"].fillna(median4,inplace=True)
df["0-4 years"].fillna(median5,inplace=True)

age_categories=["65+", "25-64 years", "15-24 years", "5-14 years", "0-4 years"]

df_age=df[["65+", "25-64 years", "15-24 years", "5-14 years", "0-4 years"]].copy()
df_age.columns=age_categories

total_population=df_age.sum()

plt.bar(age_categories,total_population)
plt.xlabel("Age category")
plt.ylabel("Total Population")
plt.show()
```



In []: