World Bank Data

```
import numpy as np
import pandas as pd
import plotly.express as px
pd.set_option('display.float_format', lambda x: '%.2f' % x)
```

```
population=pd.read_csv('https://raw.githubusercontent.com/Vivega28/week-4-prepinsta/main/country_population.csv')

fertility=pd.read_csv('https://raw.githubusercontent.com/Vivega28/week-4-prepinsta/main/fertility_rate.csv')

country =pd.read_csv('https://raw.githubusercontent.com/Vivega28/week-4-prepinsta/main/metadata_country.csv')

life_expectancy=pd.read_csv('https://raw.githubusercontent.com/Vivega28/week-4-prepinsta/main/life_expectancy.csv')
```

Country Data

```
country.head()
```

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Unnamed: 5
0	ABW	Latin America & Caribbean	High income	SNA data for 2000-2011 are updated from offici...	Aruba	NaN
1	AFG	South Asia	Low income	Fiscal year end: March 20; reporting period fo...	Afghanistan	NaN
2	AGO	Sub-Saharan Africa	Lower middle income	NaN	Angola	NaN
3	ALB	Europe & Central Asia	Upper middle	NaN	Albania	NaN

```
country = country[["Country Code", "Region", "TableName"]]
country.shape
```

```
(263, 3)
```

```
country.head()
```

	Country Code	Region	TableName
0	ABW	Latin America & Caribbean	Aruba
1	AFG	South Asia	Afghanistan
2	AGO	Sub-Saharan Africa	Angola
3	ALB	Europe & Central Asia	Albania
4	AND	Europe & Central Asia	Andorra

```
country.rename(columns={'TableName': 'Country Name'}, inplace=True)
country.head()
```

```
<ipython-input-41-9b6f6c2ef81b>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/u
  country.rename(columns={'TableName': 'Country Name'}, inplace=True)
```

	Country Code	Region	Country Name
0	ABW	Latin America & Caribbean	Aruba
1	AFG	South Asia	Afghanistan
2	AGO	Sub-Saharan Africa	Angola
3	ALB	Europe & Central Asia	Albania
4	AND	Europe & Central Asia	Andorra

```
country.isna().sum()
```

```
Country Code     0
Region          46
Country Name     0
dtype: int64
```

Population Data

```
population.head()
```

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	
0	Aruba	ABW	Population, total	SP.POP.TOTL	54211.00	55438.00	56225.00	
1	Afghanistan	AFG	Population, total	SP.POP.TOTL	8996351.00	9166764.00	9345868.00	9
2	Angola	AGO	Population, total	SP.POP.TOTL	5643182.00	5753024.00	5866061.00	5
3	Albania	ALB	Population, total	SP.POP.TOTL	1608800.00	1659800.00	1711319.00	1
4	Andorra	AND	Population, total	SP.POP.TOTL	13411.00	14375.00	15370.00	

5 rows × 61 columns

```
population.isna().sum()
```

```
Country Name      0
Country Code      0
Indicator Name    0
Indicator Code    0
1960              4
                 ..
2012              2
2013              2
2014              2
2015              2
2016              2
Length: 61, dtype: int64
```

```
years = [str(i) for i in range(1960,2017)]
```

```
population.drop(["Indicator Name","Indicator Code"], axis=1, inplace=True)
```

```
population.head()
```

	Country Name	Country Code	1960	1961	1962	1963	1964	
0	Aruba	ABW	54211.00	55438.00	56225.00	56695.00	57032.00	5
1	Afghanistan	AFG	8996351.00	9166764.00	9345868.00	9533954.00	9731361.00	993
2	Angola	AGO	5643182.00	5753024.00	5866061.00	5980417.00	6093321.00	620
3	Albania	ALB	1608800.00	1659800.00	1711319.00	1762621.00	1814135.00	186
4	Andorra	AND	13411.00	14375.00	15370.00	16412.00	17469.00	1

5 rows × 59 columns

```
population.shape
```

```
(264, 59)
```

```
population.dropna(axis=0, inplace=True)
population.shape
```

```
(258, 59)
```

```
df_1 = pd.melt(population, id_vars='Country Code', value_vars= years, var_name='Year', value_name='Population' )
```

```
df_1.head()
```

	Country Code	Year	Population
0	ABW	1960	54211.00
1	AFG	1960	8996351.00
2	AGO	1960	5643182.00
3	ALB	1960	1608800.00
4	AND	1960	13411.00

```
df_1.isna().sum()
```

```
Country Code    0
Year            0
Population      0
dtype: int64
```

```
df_1.tail()
```

	Country Code	Year	Population
14701	XKX	2016	1816200.00
14702	YEM	2016	27584213.00
14703	ZAF	2016	56015473.00
14704	ZMB	2016	16591390.00
14705	ZWE	2016	16150362.00

```
df_1['Country Code'].value_counts()
```

```
ABW    57
PYF    57
MWI    57
MYS    57
NAC    57
       ..
GUY    57
HIC    57
HKG    57
HND    57
ZWE    57
Name: Country Code, Length: 258, dtype: int64
```

```
df_1.shape
```

```
(14706, 3)
```

```
country.head()
```

	Country Code	Region	Country Name
0	ABW	Latin America & Caribbean	Aruba
1	AFG	South Asia	Afghanistan
2	AGO	Sub-Saharan Africa	Angola
3	ALB	Europe & Central Asia	Albania
4	AND	Europe & Central Asia	Andorra

```
df_merged = pd.merge(country, df_1, how='left', on='Country Code')
```

```
df_merged.head()
```

	Country Code	Region	Country Name	Year	Population
0	ABW	Latin America & Caribbean	Aruba	1960	54211.00
1	ABW	Latin America & Caribbean	Aruba	1961	55438.00
2	ABW	Latin America & Caribbean	Aruba	1962	56225.00
3	ABW	Latin America & Caribbean	Aruba	1963	56695.00
4	ABW	Latin America & Caribbean	Aruba	1964	57032.00

```
df_merged.tail()
```

	Country Code	Region	Country Name	Year	Population
14706	ZWE	Sub-Saharan Africa	Zimbabwe	2012	14710826.00
14707	ZWE	Sub-Saharan Africa	Zimbabwe	2013	15054506.00
14708	ZWE	Sub-Saharan Africa	Zimbabwe	2014	15411675.00
14709	ZWE	Sub-Saharan Africa	Zimbabwe	2015	15777451.00
14710	ZWE	Sub-Saharan Africa	Zimbabwe	2016	16150362.00

```
df_merged.shape
```

```
(14711, 5)
```

Fertility Data

```
fertility.head()
```

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965
0	Aruba	ABW	Fertility rate, total (births per woman)	SP.DYN.TFRT.IN	4.82	4.66	4.47	4.27	4.06	3.84
1	Afghanistan	AFG	Fertility rate, total (births per woman)	SP.DYN.TFRT.IN	7.45	7.45	7.45	7.45	7.45	7.45
2	Angola	AGO	Fertility rate, total (births per woman)	SP.DYN.TFRT.IN	7.48	7.52	7.56	7.59	7.61	7.62

```
fertility.drop(["Indicator Name","Indicator Code"], axis=1, inplace=True)
```

```
fertility.head()
```

	Country Name	Country Code	1960	1961	1962	1963	1964	1965	1966	1967	...	2007	200
0	Aruba	ABW	4.82	4.66	4.47	4.27	4.06	3.84	3.62	3.42	...	1.76	1.7
1	Afghanistan	AFG	7.45	7.45	7.45	7.45	7.45	7.45	7.45	7.45	...	6.46	6.2
2	Angola	AGO	7.48	7.52	7.56	7.59	7.61	7.62	7.62	7.61	...	6.37	6.3
3	Albania	ALB	6.49	6.40	6.28	6.13	5.96	5.77	5.58	5.39	...	1.67	1.6
4	Andorra	AND	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1.18	1.2

```
fertility.dropna(axis=0, inplace=True)
```

```
df_2 = pd.melt(fertility,
              id_vars='Country Code',
              value_vars= years,
              var_name='Year',
              value_name='Fertility')
df_2.head()
```

	Country Code	Year	Fertility
0	ABW	1960	4.82
1	AFG	1960	7.45
2	AGO	1960	7.48
3	ALB	1960	6.49
4	ARB	1960	6.95

```
df_2.shape
```

```
(13395, 3)
```

```
df_merged.head()
```

	Country Code	Region	Country Name	Year	Population
0	ABW	Latin America & Caribbean	Aruba	1960	54211.00
1	ABW	Latin America & Caribbean	Aruba	1961	55438.00
2	ABW	Latin America & Caribbean	Aruba	1962	56225.00
3	ABW	Latin America & Caribbean	Aruba	1963	56695.00
4	ABW	Latin America & Caribbean	Aruba	1964	57032.00

```
df_merged_2 = pd.merge(df_merged, df_2, how='left', on=['Country Code','Year'])
df_merged_2.head()
```

	Country Code	Region	Country Name	Year	Population	Fertility
0	ABW	Latin America & Caribbean	Aruba	1960	54211.00	4.82
1	ABW	Latin America & Caribbean	Aruba	1961	55438.00	4.66
2	ABW	Latin America & Caribbean	Aruba	1962	56225.00	4.47

```
df_merged_2.shape
```

```
(14711, 6)
```

Life expectancy Data

[ ] ↳ *11 cells hidden*

Final Data

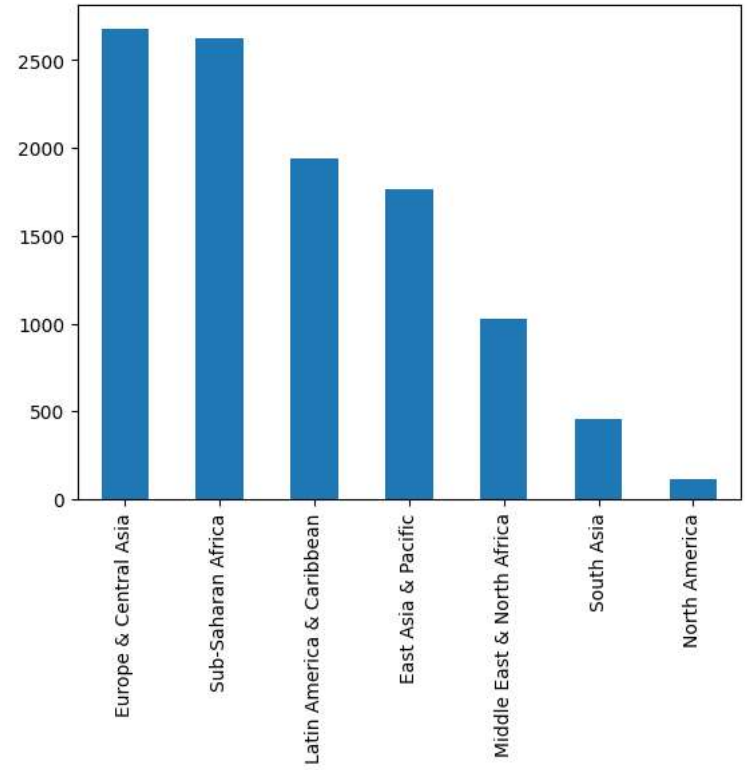
[ ] ↳ *2 cells hidden*

Data Visualisation

Pandas library

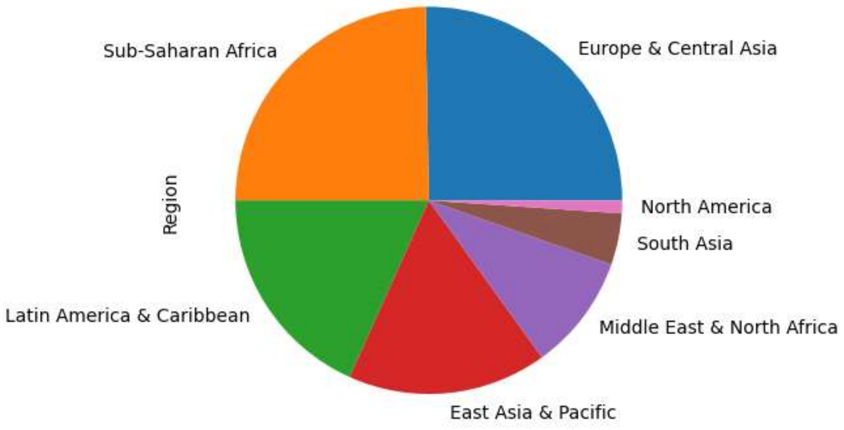
```
df_final['Region'].value_counts().plot(kind='bar')
```

```
<Axes: >
```



```
df_final['Region'].value_counts().plot(kind = 'pie')
```

```
<Axes: ylabel='Region'>
```

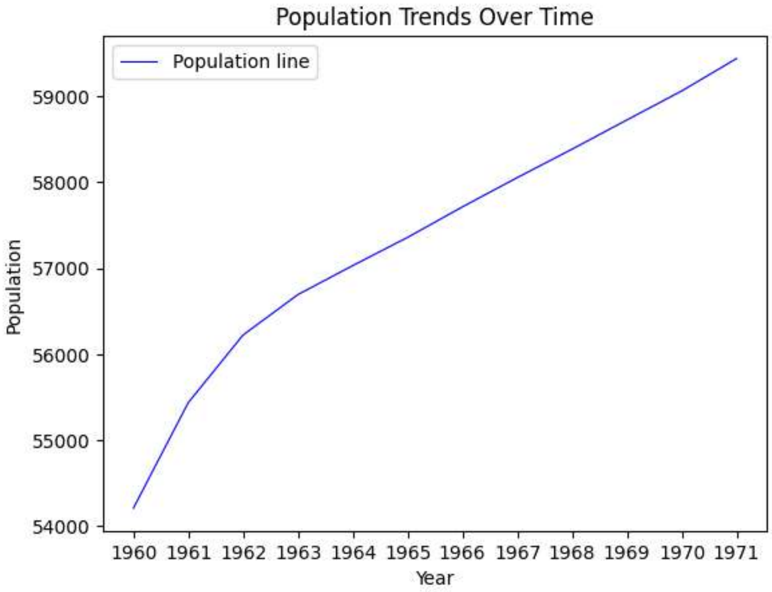


Matplotlib

```
import matplotlib.pyplot as plt
```

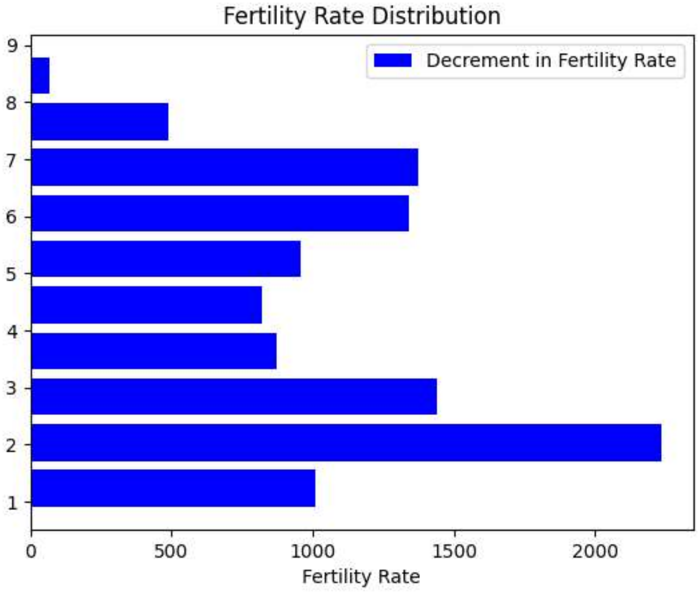
```
max_pop = df_final['Population'].max()
year = df_final['Year'][:12]
pop = df_final['Population'][:12]
```

```
plt.plot(year, pop, label='Population line', linewidth = 0.8, linestyle  = "-", color = 'b')
plt.title('Population Trends Over Time')
plt.ylabel('Population')
plt.xlabel('Year')
plt.legend()
plt.show()
```



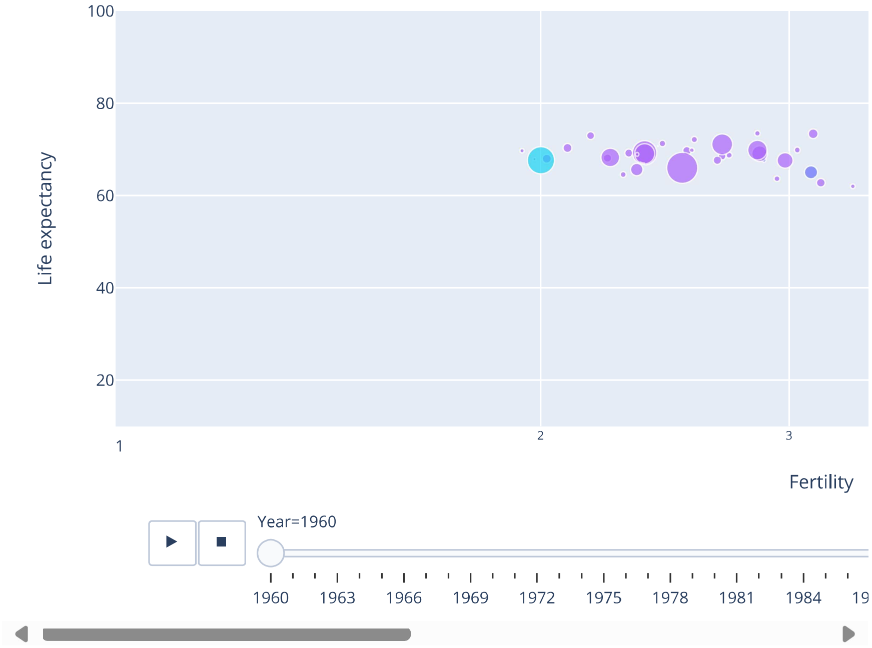
```
f_rate = df_final['Fertility']
```

```
plt.hist(f_rate, label='Decrement in Fertility Rate', rwidth= 0.8, histtype = 'bar', orientation = 'horizontal', color = 'b')
plt.title("Fertility Rate Distribution")
plt.xlabel("Fertility Rate")
plt.legend()
plt.show()
```

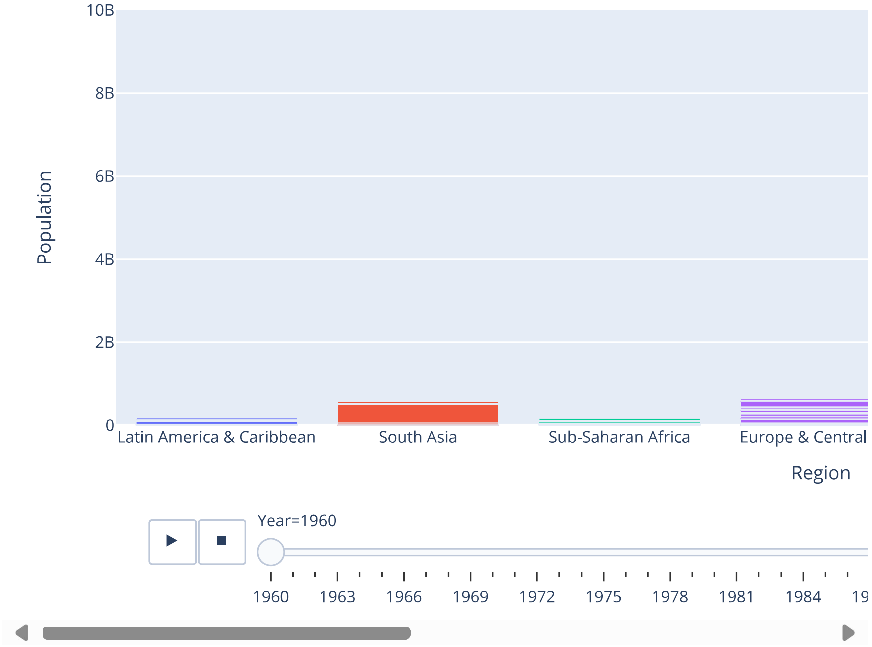


```
import plotly.express as px
```

```
px.scatter(df_final,
           x="Fertility",
           y="Life expectancy",
           animation_frame="Year",
           animation_group="Country Code",
           size="Population",
           color="Region",
           hover_name="Country Code",
           log_x=True,
           size_max=55,
           range_x=[1,10],
           range_y=[10,100])
```



```
px.bar(df_final,
       x="Region",
       y="Population",
       animation_frame="Year",
       animation_group="Country Code",
       color="Region",
       range_y=[0,10000000000])
```


10B
8B
6B
4B
2B
Population
Latin America & Caribbean
South Asia
Sub-Saharan Africa
Europe & Central
Region
Year=1960