DATA ANALYTICS

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns  # For better visuals

from google.colab import files
uploaded = files.upload()  # Opens file upload dialog
```

<IPython.core.display.HTML object>

Saving file.xlsx.csv to file.xlsx.csv

```python
import pandas as pd
df = pd.read_csv("file.xlsx.csv")  # Replace with your actual file
name
print(df.head())  # Check if data is loaded correctly
```

Sl No	State/UT	Population[50]	Percent (%)	Male	Female
0 1	Uttar Pradesh	199812341	16.50	104480510	95331831
1 2	Maharashtra	112374333	9.28	58243056	54131277
2 3	Bihar	104099452	8.60	54278157	49821295
3 4	West Bengal	91276115	7.54	46809027	44467088
4 5	Madhya Pradesh	72626809	6.00	37612306	35014503

	Difference between male and female	Sex ratio	Rural[51]	Urban[51]
0	9148679	930	155111022	44470455
1	4111779	929	61545441	50827531
2	4456862	918	92075028	11729609
3	2341939	950	62213676	29134060
4	2597803	931	52537899	20059666

	Area[52] (km2)	Density (per km2)
0	240928	828
1	307713	365
2	94163	1102
3	88752	1030
4	308245	236

```
plt.figure(figsize=(12, 6))

<Figure size 1200x600 with 0 Axes>

<Figure size 1200x600 with 0 Axes>

# Plot bar chart with Male and Female Population
df[['State/UT', 'Male',
'Female']].set_index('State/UT').plot(kind='bar', stacked=True,
figsize=(12, 6), colormap="coolwarm")

plt.title("Male vs Female Population by State", fontsize=14)
plt.xlabel("State/UT", fontsize=12)
plt.ylabel("Population", fontsize=12)
plt.xticks(rotation=90)
plt.legend(["Male", "Female"])
plt.grid(axis="y", linestyle="--", alpha=0.7)

plt.show()
```
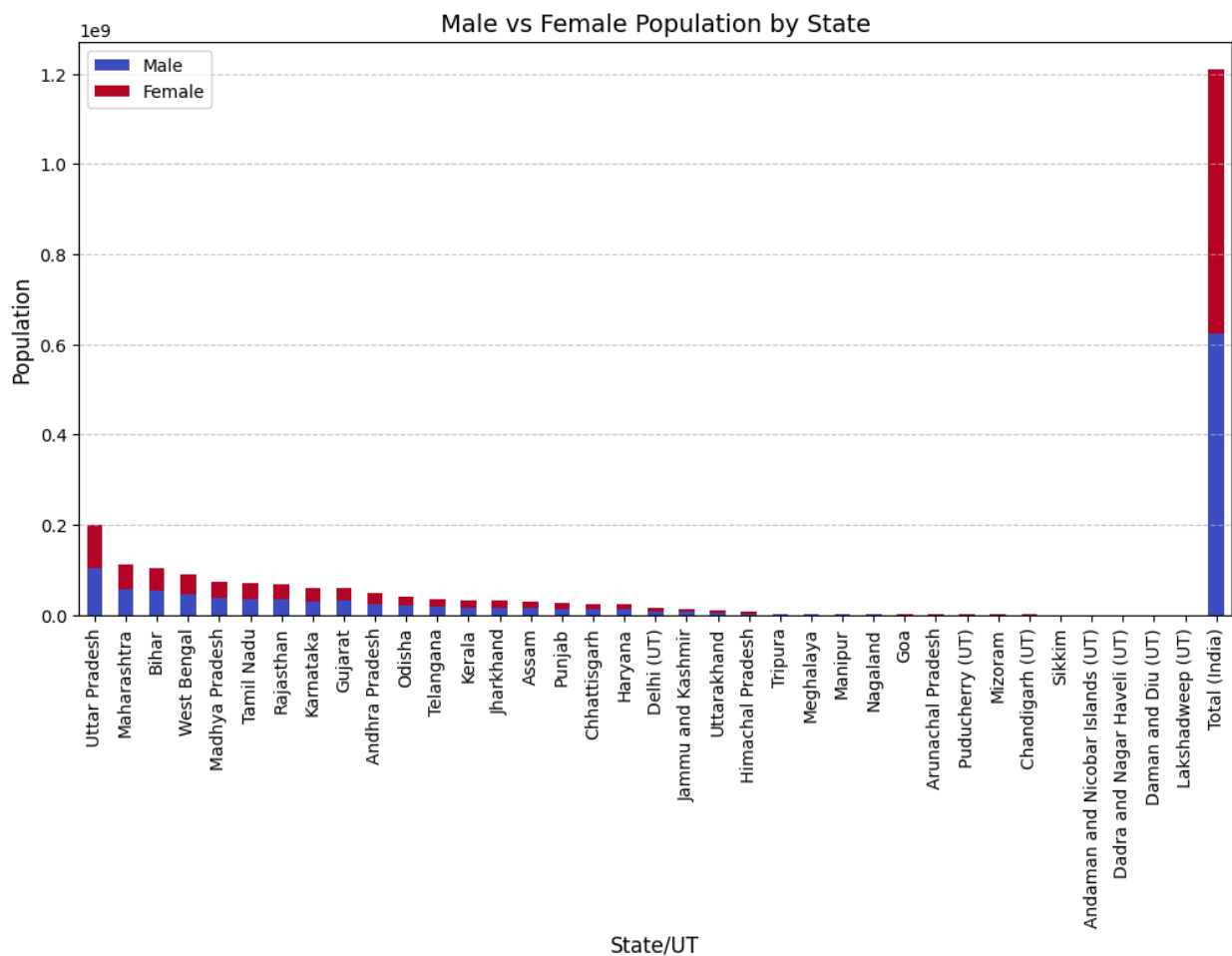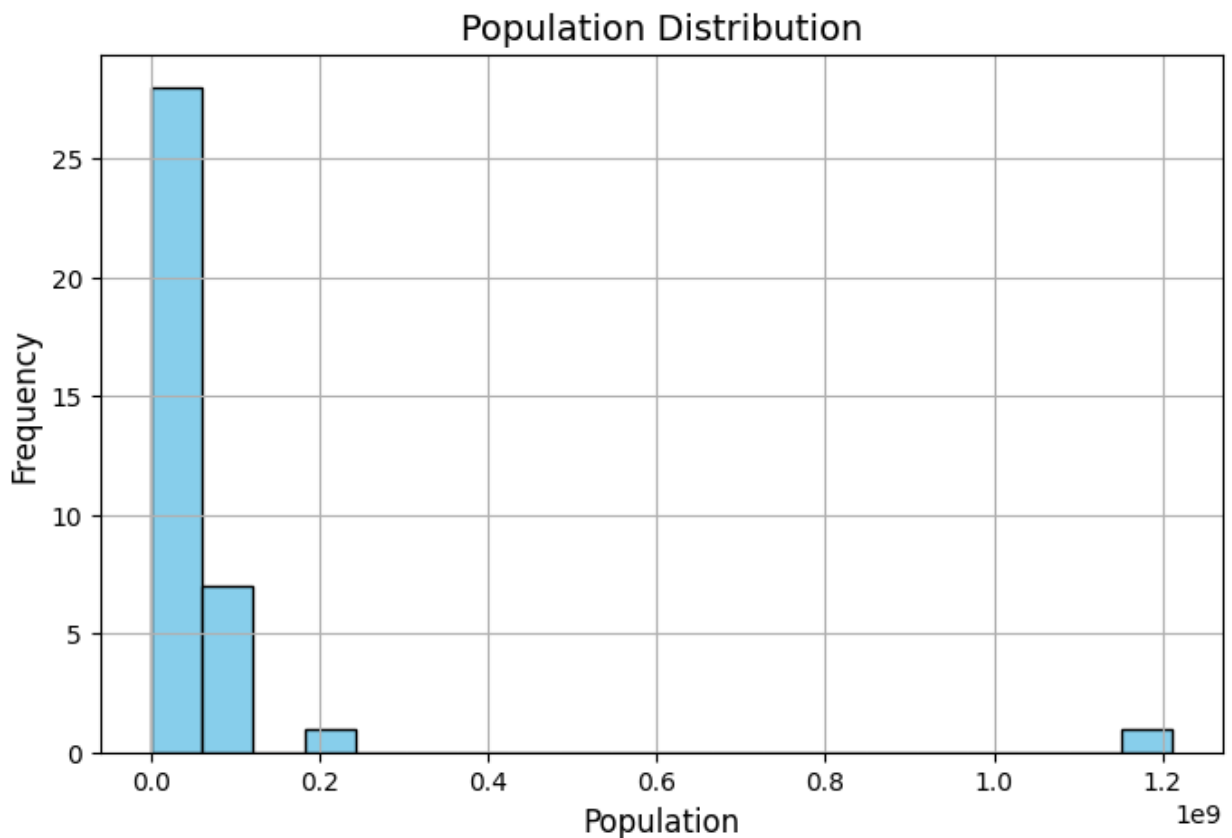
```
plt.figure(figsize=(8, 5))

# Plot histogram for Population Distribution
plt.hist(df["Population[50]"], bins=20, color='skyblue',
edgecolor='black')

plt.title("Population Distribution", fontsize=14)
plt.xlabel("Population", fontsize=12)
plt.ylabel("Frequency", fontsize=12)
plt.grid(True)

plt.show()
```



```
from google.colab import files
uploaded = files.upload()

<IPython.core.display.HTML object>

Saving bar.xlsx.csv to bar.xlsx.csv

# Read the dataset (Modify filename accordingly)
df = pd.read_csv("bar.xlsx.csv")
```

```python
# Display first 5 rows
df.head()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 500,\n \"fields\": [\n   {\n   \"column\": \"ID\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 144,\n   \"min\": 1,\n   \"max\": 500,\n   \"num_unique_values\": 500,\n   \"samples\": [\n   362,\n   74,\n   375\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 18,\n   \"min\": 18,\n   \"max\": 79,\n   \"num_unique_values\": 62,\n   \"samples\": [\n   51,\n   18,\n   56\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Gender\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 3,\n   \"samples\": [\n   \"Female\",\n   \"Other\",\n   \"Male\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"State\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 10,\n   \"samples\": [\n   \"Madhya Pradesh\",\n   \"Tamil Nadu\",\n   \"Bihar\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Occupation\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 6,\n   \"samples\": [\n   \"Unemployed\",\n   \"Engineer\",\n   \"Doctor\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Income (INR)\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 54100,\n   \"min\": 10207,\n   \"max\": 199648,\n   \"num_unique_values\": 499,\n   \"samples\": [\n   78619,\n   180515,\n   29508\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Education Level\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 5,\n   \"samples\": [\n   \"Secondary\",\n   \"Primary\",\n   \"Graduate\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Marital Status\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 4,\n   \"samples\": [\n   \"Divorced\",\n   \"Married\",\n   \"Widowed\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Number of Children\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 1,\n   \"min\": 0,\n   \"max\": 4,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   4,\n   1,\n   2\n   ],\n

\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

```python
# Create a bar chart for Age distribution
plt.figure(figsize=(10, 5))
df["Age"].value_counts().sort_index().plot(kind="bar",
color="skyblue")

plt.xlabel("Age")
plt.ylabel("Count")
plt.title("Age Distribution in the Dataset")
plt.grid(axis="y", linestyle="--", alpha=0.7)

plt.show()
```


Age Distribution in the Dataset