REG NO : 421221243033
COLLEGE CODE : 4212

DATA ANALYTICS WITH COGOS-GROUP:2 ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU: PHASE4 IBM NAAN MUDHULVAN

In the previous phases we have discussed about the step-by-step process, Design thinking and at the phase3 we have discussed about the data preprocessing techniques and many more in the last steps and in this step we have given some problem statements to solve in the IBM COGNOS ANALYTICS. In this part we will continue building our project, Building the analysis by creating visualizations using libraries.

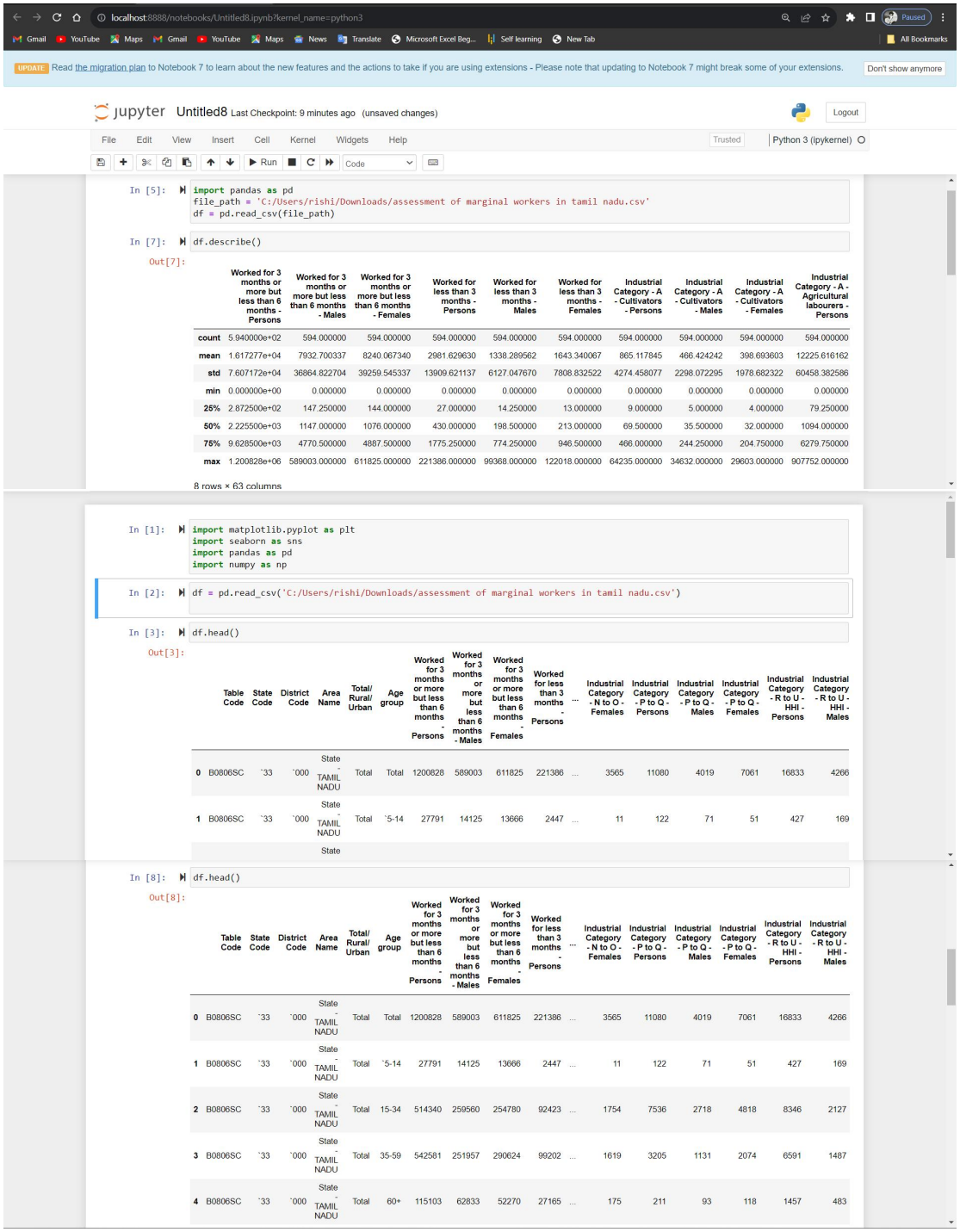
PROBLEM 1: Perform the demographic analysis and create visualizations.

PROBLEM 2: Calculate the distribution of marginal workers based on age, industrial category, and sex using data aggregation and manipulation.

PROBLEM 3: Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

ORIGINAL

PROBLEM 1



localhost:8888/notebooks/Untitled8.ipynb?kernel_name=python3

jupyter Untitled8 Last Checkpoint: 9 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

In [5]:
```
import pandas as pd
file_path = 'C:/Users/rishi/Downloads/assessment of marginal workers in tamil nadu.csv'
df = pd.read_csv(file_path)
```

In [7]:
```
df.describe()
```

Out[7]:

	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	Worked for less than 3 months - Males	Worked for less than 3 months - Females	Industrial Category - A - Cultivators - Persons	Industrial Category - A - Cultivators - Males	Industrial Category - A - Cultivators - Females	Industrial Category - A - Agricultural labourers - Persons
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562	1643.340067	865.117845	466.424242	398.693603	12225.616162
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670	7808.832522	4274.458077	2298.072295	1978.682322	60458.382586
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.872500e+02	147.250000	144.000000	27.000000	14.250000	13.000000	9.000000	5.000000	4.000000	79.250000
50%	2.225500e+03	1147.000000	1076.000000	430.000000	198.500000	213.000000	69.500000	35.500000	32.000000	1094.000000
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	774.250000	946.500000	466.000000	244.250000	204.750000	6279.750000
max	1.200828e+06	589003.000000	611825.000000	221386.000000	99368.000000	122018.000000	64235.000000	34632.000000	29603.000000	907752.000000

8 rows × 63 columns

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

In [2]:
```
df = pd.read_csv('C:/Users/rishi/Downloads/assessment of marginal workers in tamil nadu.csv')
```

In [3]:
```
df.head()
```

Out[3]:

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industrial Category - R to U - HHI - Persons	Industrial Category - R to U - HHI - Males
0	B0806SC	`33	`000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266
1	B0806SC	`33	`000	State - TAMIL NADU	Total	`5-14	27791	14125	13666	2447	...	11	122	71	51	427	169
				State													

In [8]:
```
df.head()
```

Out[8]:

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industrial Category - R to U - HHI - Persons	Industrial Category - R to U - HHI - Males
0	B0806SC	`33	`000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266
1	B0806SC	`33	`000	State - TAMIL NADU	Total	`5-14	27791	14125	13666	2447	...	11	122	71	51	427	169
2	B0806SC	`33	`000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127
3	B0806SC	`33	`000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487
4	B0806SC	`33	`000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	118	1457	483

In [10]:
```
df.tail()
```

Out[10]:

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industrial Category - R to U - HHI - Persons	Indus Cate - R t I M
589	B0806SC	`33	`633	District - Tiruppur	Urban	`5-14	272	129	143	18	...	0	0	0	0	0	
590	B0806SC	`33	`633	District - Tiruppur	Urban	15-34	3285	1654	1631	473	...	20	44	15	29	62	
591	B0806SC	`33	`633	District - Tiruppur	Urban	35-59	3672	1769	1903	522	...	33	35	12	23	36	
592	B0806SC	`33	`633	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0	3	10	
593	B0806SC	`33	`633	District - Tiruppur	Urban	Age not stated	2	1	1	0	...	0	0	0	0	0	

5 rows × 69 columns

In [12]:
```
print(df.info())  # To check the columns and data types
print(df.describe())  # Summary statistics of numerical columns
```

```
<class 'pandas.core.frame.DataFrame'>
```

In [12]:
```
print(df.info())  # To check the columns and data types
print(df.describe())  # Summary statistics of numerical columns
```

```
------  -----
 0   Table Code                                                 594 non-
null    object
 1   State Code                                                 594 non-
null    object
 2   District Code                                              594 non-
null    object
 3   Area Name                                                  594 non-
null    object
 4   Total/ Rural/ Urban                                        594 non-
null    object
 5   Age group                                                  594 non-
null    object
 6   Worked for 3 months or more but less than 6 months -  Persons  594 non-
null    int64
 7   Worked for 3 months or more but less than 6 months - Males  594 non-
null    int64
 8   Worked for 3 months or more but less than 6 months - Females  594 non-
null    int64
```

In [ ]:

In [18]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='Urban_Rural')
plt.title('Distribution of Marginal Workers in Urban and Rural Areas')
plt.xlabel('Location')
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

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

In [2]:
```
df = pd.read_csv('C:/Users/rishi/Downloads/assessment of marginal workers in tamil nadu.csv')
```

In [3]:
```
df.head()
```

Out[3]:

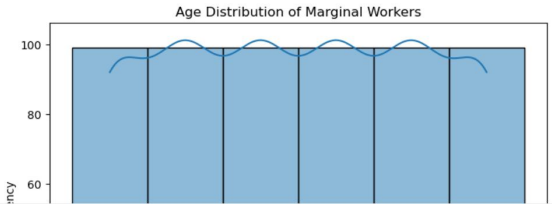
Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industrial Category - R to U - HHI - Persons	Industrial Category - R to U - HHI - Males	Industrial Category - R to U - HHI - Females	Industrial Category - R to U - Non HHI - Persons	Industrial Category - R to U - Non HHI - Males	Industrial Category - R to U - Non HHI - Females
Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266	12567	122088	55801	66287
`5-14	27791	14125	13666	2447	...	11	122	71	51	427	169	258	19305	9774	9531
15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127	6219	68929	32803	36126
35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487	5104	26498	9675	16823
60+	115103	62833	52270	27165	...	175	211	93	118	1457	483	974	7065	3394	3671

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [4]:
```
df.tail()
```

ict de	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industrial Category - R to U - HHI - Persons	Industrial Category - R to U - HHI - Males	Industrial Category - R to U - HHI - Females	Industrial Category - R to U - Non HHI - Persons	Industrial Category - R to U - Non HHI - Males	Industrial Category - R to U - Non HHI - Females
33	District - Tiruppur	Urban	5-14	272	129	143	18	...	0	0	0	0	0	0	0	228	104	124
33	District - Tiruppur	Urban	15-34	3285	1654	1631	473	...	20	44	15	29	62	6	56	675	247	428
33	District - Tiruppur	Urban	35-59	3672	1769	1903	522	...	33	35	12	23	36	9	27	279	103	176
33	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0	3	10	3	7	81	35	46

In [8]:
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x='Age group', bins=20, kde=True)
plt.title('Age Distribution of Marginal Workers')
plt.xlabel('Age group')
plt.ylabel('Frequency')
plt.show()
```

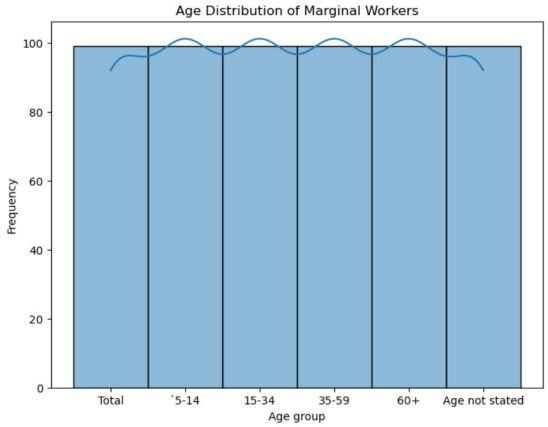


ORIGINAL

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

83	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0	3	10	3	7	81	35	46

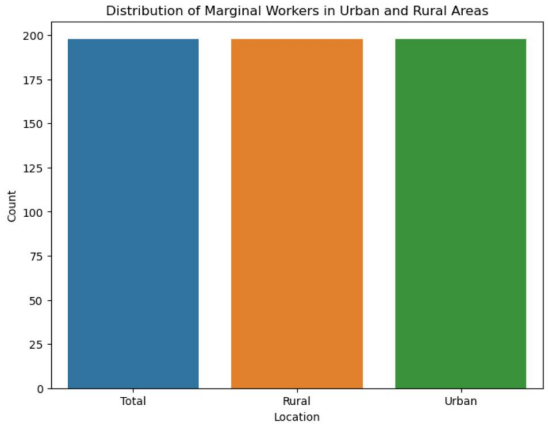
In [8]:
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x='Age group', bins=20, kde=True)
plt.title('Age Distribution of Marginal Workers')
plt.xlabel('Age group')
plt.ylabel('Frequency')
plt.show()
```



File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Age group

In [11]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='Total/ Rural/ Urban')
plt.title('Distribution of Marginal Workers in Urban and Rural Areas')
plt.xlabel('Location')
plt.ylabel('Count')
plt.show()
```

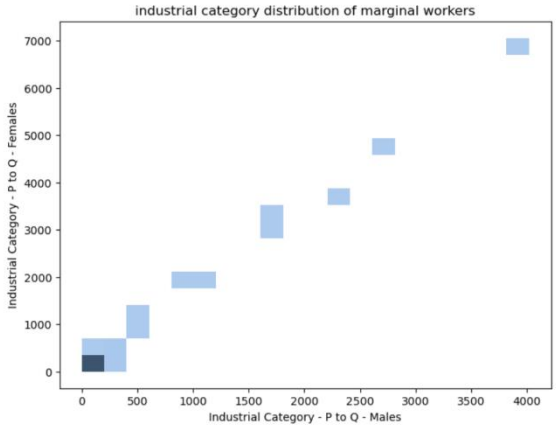


In [17]:
```
plt.figure(figsize=(8, 6))
```

PROBLEM 2

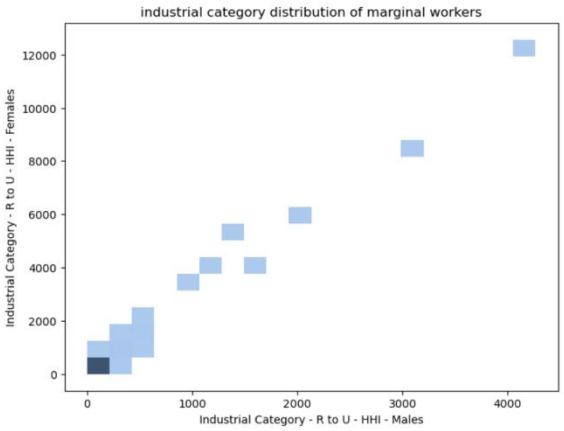
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [17]:
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x='Industrial Category - P to Q - Males', y='Industrial Category - P to Q - Females', bins=20, kde=True
plt.title('industrial category distribution of marginal workers')
plt.xlabel('Industrial Category - P to Q - Males')
plt.ylabel('Industrial Category - P to Q - Females')
plt.show()
```



File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

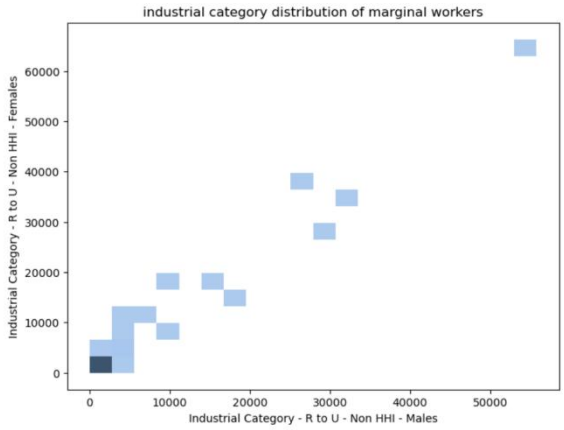
In [19]:
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x='Industrial Category - R to U - HHI - Males', y='Industrial Category - R to U - HHI - Females', bins=
plt.title('industrial category distribution of marginal workers')
plt.xlabel('Industrial Category - R to U - HHI - Males')
plt.ylabel('Industrial Category - R to U - HHI - Females')
plt.show()
```



File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Industrial Category - R to U - HHI - Males

In [20]:
```
lt.figure(figsize=(8, 6))
ns.histplot(data=df, x='Industrial Category - R to U - Non HHI - Males', y='Industrial Category - R to U - Non HHI - Females'
lt.title('industrial category distribution of marginal workers')
lt.xlabel('Industrial Category - R to U - Non HHI - Males')
lt.ylabel('Industrial Category - R to U - Non HHI - Females')
lt.show()
```

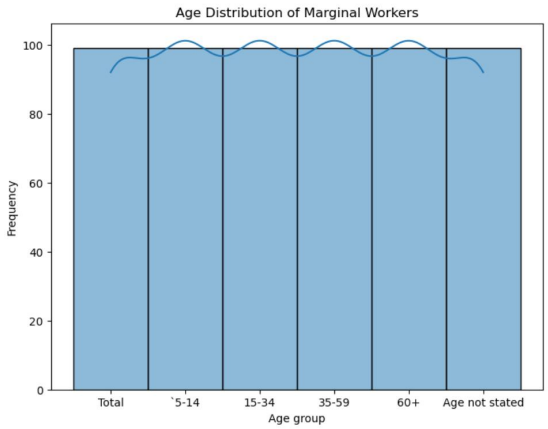


PROBLEM 3

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

33	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0	3	10	3	7	81	35	46

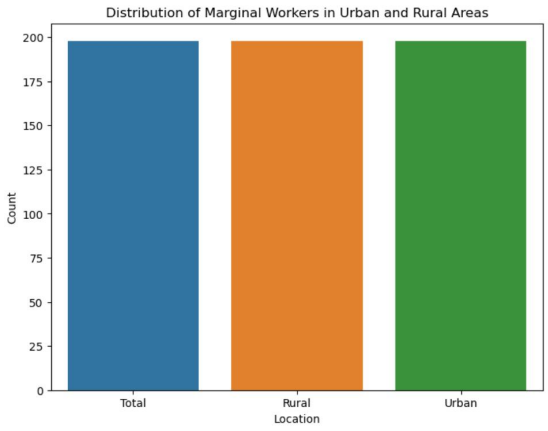
In [8]:
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x='Age group', bins=20, kde=True)
plt.title('Age Distribution of Marginal Workers')
plt.xlabel('Age group')
plt.ylabel('Frequency')
plt.show()
```



File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Age group

In [11]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x='Total/ Rural/ Urban')
plt.title('Distribution of Marginal Workers in Urban and Rural Areas')
plt.xlabel('Location')
plt.ylabel('Count')
plt.show()
```



In [17]:
```
plt.figure(figsize=(8, 6))
```