

Date - 17-10-2023

Team Id : 3896

Project Title : Assessment of Marginal Workers in Tamil Nadu-A

Sociaeconomic Analysis(ADS)

Importing the Required packages

```
In [2]: 1 import pandas as pd  
2 import numpy as np  
3 import matplotlib.pyplot as plt  
4 import seaborn as sns
```

Loading Dataset

```
In [3]: 1 dataset = pd.read_csv("D:\IBM Naan Mudhalvan\Dataset\Main\DDW_B06_330
```

Data Exploration

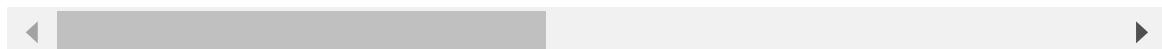
In [4]:

1 dataset

Out[4]:

Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	
						Persons	- Males	- Females	Persons	
0	B0706	33	`000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891
1	B0706	33	`000	State - TAMIL NADU	Total	`5-9	48238	24511	23727	2051
2	B0706	33	`000	State - TAMIL NADU	Total	`10- 14	76288	39191	37097	6993
3	B0706	33	`000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938
4	B0706	33	`000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036
...
1381	B0706	33	`633	District - Tiruppur	Urban	50-59	4965	2800	2165	901
1382	B0706	33	`633	District - Tiruppur	Urban	60-69	2827	1590	1237	578
1383	B0706	33	`633	District - Tiruppur	Urban	70-79	920	581	339	204
1384	B0706	33	`633	District - Tiruppur	Urban	80+	191	104	87	47
1385	B0706	33	`633	District - Tiruppur	Urban	Age not stated	31	23	8	9

1386 rows × 69 columns



```
In [5]: 1 dataset.head()
```

Out[5]:

	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	In C -	
0	B0706	33	000	TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	...	State
1	B0706	33	000	TAMIL NADU	Total	5-9	48238	24511	23727	2051	...	State
2	B0706	33	000	TAMIL NADU	Total	10- 14	76288	39191	37097	6993	...	State
3	B0706	33	000	TAMIL NADU	Total	15-19	257605	141262	116343	41938	...	State
4	B0706	33	000	TAMIL NADU	Total	20-24	478082	257149	220933	81036	...	State

5 rows × 69 columns



```
In [6]: 1 dataset.shape
```

Out[6]: (1386, 69)

In [30]: 1 dataset.columns

```
Out[30]: Index(['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',
       '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',
       'Industrial Category - A - Agricultural labourers - Males',
       'Industrial Category - A - Agricultural labourers - Females',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females',
       'Industrial Category - B - Persons', 'Industrial Category - B - Males',
       'Industrial Category - B - Females',
       'Industrial Category - C - HHI - Persons',
       'Industrial Category - C - HHI - Males',
       'Industrial Category - C - HHI - Females',
       'Industrial Category - C - Non HHI - Persons',
       'Industrial Category - C - Non HHI - Males',
       'Industrial Category - C - Non HHI - Females',
       'Industrial Category - D & E - Persons',
       'Industrial Category - D & E - Males',
       'Industrial Category - D & E - Females',
       'Industrial Category - F - Persons', 'Industrial Category - F - Males',
       'Industrial Category - F - Females',
       'Industrial Category - G - HHI - Persons',
       'Industrial Category - G - HHI - Males',
       'Industrial Category - G - HHI - Females',
       'Industrial Category - G - Non HHI - Persons',
       'Industrial Category - G - Non HHI - Males',
       'Industrial Category - G - Non HHI - Females',
       'Industrial Category - H - Persons', 'Industrial Category - H - Males',
       'Industrial Category - H - Females',
       'Industrial Category - I - Persons', 'Industrial Category - I - Males',
       'Industrial Category - I - Females',
       'Industrial Category - J - HHI - Persons',
       'Industrial Category - J - HHI - Males',
       'Industrial Category - J - HHI - Females',
       'Industrial Category - J - Non HHI - Persons',
       'Industrial Category - J - Non HHI - Males',
       'Industrial Category - J - Non HHI - Females',
       'Industrial Category - K to M - Persons',
       'Industrial Category - K to M - Males',
       'Industrial Category - K to M - Females',
       'Industrial Category - N to O - Persons',
       'Industrial Category - N to O - Males',
       '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'],
dtype='object')
```

Data Preprocessing

```
In [8]: 1 dataset["Age group"].value_counts()
```

```
Out[8]: Total      99
`5-9       99
`10-14     99
15-19      99
20-24      99
25-29      99
30-34      99
35-39      99
40-49      99
50-59      99
60-69      99
70-79      99
80+        99
Age not stated  99
Name: Age group, dtype: int64
```

```
In [9]: 1 dataset["Age group"].replace(['1','`5-9','`10-14','0'],['Total','5-9',
```

```
In [10]: 1 dataset["Age group"].value_counts()
```

```
Out[10]: Total      99
5-9       99
10-14     99
15-19      99
20-24      99
25-29      99
30-34      99
35-39      99
40-49      99
50-59      99
60-69      99
70-79      99
80+        99
Age not stated  99
Name: Age group, dtype: int64
```

```
In [11]: 1 dataset['Total/ Rural/ Urban'].value_counts()
```

```
Out[11]: Total    462
Rural    462
Urban    462
Name: Total/ Rural/ Urban, dtype: int64
```

```
In [32]: 1 dataset['Area Name'].value_counts()
```

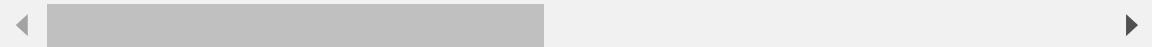
```
Out[32]: State - TAMIL NADU      42
District - Nagapattinam    42
District - Coimbatore      42
District - Krishnagiri      42
District - Dharmapuri      42
District - Kanniyakumari    42
District - Tirunelveli      42
District - Thoothukkudi     42
District - Ramanathapuram   42
District - Virudhunagar     42
District - Theni            42
District - Madurai           42
District - Sivaganga         42
District - Pudukkottai       42
District - Thanjavur         42
District - Thiruvarur        42
District - Cuddalore          42
District - Thiruvallur        42
District - Ariyalur           42
District - Perambalur         42
District - Tiruchirappalli    42
District - Karur              42
District - Dindigul            42
District - The Nilgiris        42
District - Erode                42
District - Namakkal             42
District - Salem                42
District - Viluppuram            42
District - Tiruvannamalai       42
District - Vellore              42
District - Kancheepuram          42
District - Chennai               42
District - Tiruppur              42
Name: Area Name, dtype: int64
```

```
In [12]: 1 dataset.isnull()
```

Out[12]:

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months	Worked for less than 3 months	...
							Persons	- Males	Females	Persons	...
0	False	False	False	False	False	False	False	False	False	False	...
1	False	False	False	False	False	False	False	False	False	False	...
2	False	False	False	False	False	False	False	False	False	False	...
3	False	False	False	False	False	False	False	False	False	False	...
4	False	False	False	False	False	False	False	False	False	False	...
...
1381	False	False	False	False	False	False	False	False	False	False	...
1382	False	False	False	False	False	False	False	False	False	False	...
1383	False	False	False	False	False	False	False	False	False	False	...
1384	False	False	False	False	False	False	False	False	False	False	...
1385	False	False	False	False	False	False	False	False	False	False	...

1386 rows × 69 columns



```
In [13]: 1 dataset.isnull().sum()
```

```
Out[13]: Table Code          0
          State Code         0
          District Code       0
          Area Name          0
          Total/ Rural/ Urban 0
          ...
          Industrial Category - R to U - HHI - Males      0
          Industrial Category - R to U - HHI - Females      0
          Industrial Category - R to U - Non HHI - Persons   0
          Industrial Category - R to U - Non HHI - Males      0
          Industrial Category - R to U - Non HHI - Females    0
Length: 69, dtype: int64
```

```
In [14]: 1 dataset.isnull().sum().sum()
```

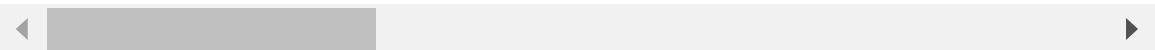
Out[14]: 0

```
In [15]: 1 dataset.describe()
```

Out[15]:

	State Code	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
count	1386.0	1.386000e+03	1.386000e+03	1.386000e+03	1386.000000	1386.000000
mean	33.0	2.435142e+04	1.233409e+04	1.201733e+04	4178.303030	1946.712843
std	0.0	1.530754e+05	7.669251e+04	7.656262e+04	26234.919027	12024.992364
min	33.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000
25%	33.0	8.372500e+02	4.637500e+02	3.792500e+02	123.000000	71.000000
50%	33.0	3.985000e+03	2.047500e+03	1.812000e+03	650.500000	315.500000
75%	33.0	1.251725e+04	6.273000e+03	6.255500e+03	2071.750000	955.250000
max	33.0	4.218884e+06	2.136881e+06	2.082003e+06	723891.000000	337268.000000

8 rows × 64 columns

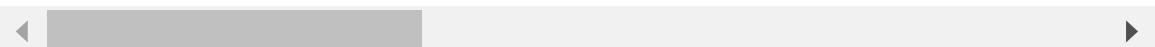


```
In [16]: 1 dataset.describe(include='all')
```

Out[16]:

	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
count	1386	1386.0	1386	1386	1386	1386	1.386000e+03	1.386000e+03	1.386000e+03
unique	1	NaN	33	33	3	14	NaN	NaN	NaN
top	B0706	NaN	'000	State TAMIL NADU	Total	Total	NaN	NaN	NaN
freq	1386	NaN	42	42	462	99	NaN	NaN	NaN
mean	NaN	33.0	NaN	NaN	NaN	NaN	2.435142e+04	1.233409e+04	1.201733e+04
std	NaN	0.0	NaN	NaN	NaN	NaN	1.530754e+05	7.669251e+04	7.656262e+04
min	NaN	33.0	NaN	NaN	NaN	NaN	0.000000e+00	0.000000e+00	0.000000e+00
25%	NaN	33.0	NaN	NaN	NaN	NaN	8.372500e+02	4.637500e+02	3.792500e+02
50%	NaN	33.0	NaN	NaN	NaN	NaN	3.985000e+03	2.047500e+03	1.812000e+03
75%	NaN	33.0	NaN	NaN	NaN	NaN	1.251725e+04	6.273000e+03	6.255500e+03
max	NaN	33.0	NaN	NaN	NaN	NaN	4.218884e+06	2.136881e+06	2.082003e+06

11 rows × 69 columns

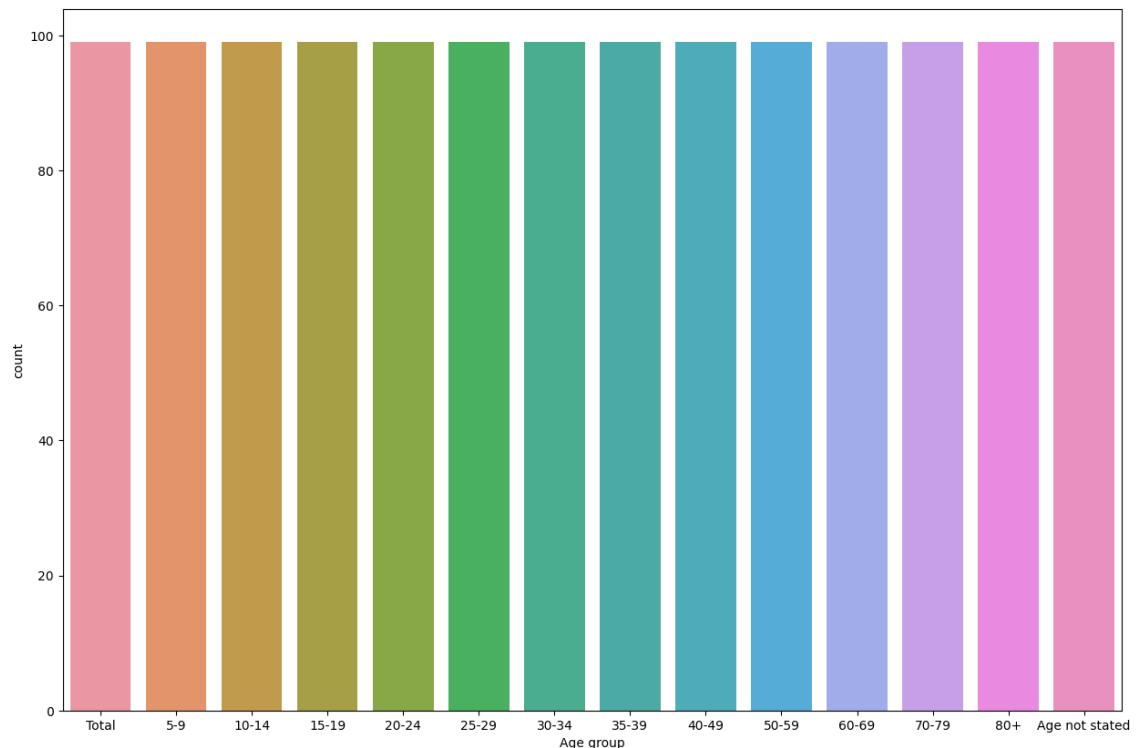


```
In [17]: 1 dataset.info()
          2
          3 1386 non-null object
          4 Total/ Rural/ Urban
          5 1386 non-null object
          6 Age group
          7 1386 non-null object
          8 Worked for 3 months or more but less than 6 months - Persons
          9 1386 non-null int64
          10 Worked for 3 months or more but less than 6 months - Males
          11 1386 non-null int64
          12 Worked for 3 months or more but less than 6 months - Females
          13 1386 non-null int64
          14 Worked for less than 3 months - Persons
          15 1386 non-null int64
          16 Worked for less than 3 months - Males
          17 1386 non-null int64
          18 Worked for less than 3 months - Females
          19 1386 non-null int64
          20 Industrial Category - A - Cultivators - Persons
          21 1386 non-null int64
          22 Industrial Category - A - Cultivators - Males
```

Data Visualization

```
In [21]: 1 plt.figure(figsize=(15,10))
          2 sns.countplot(x='Age group',data=dataset)
```

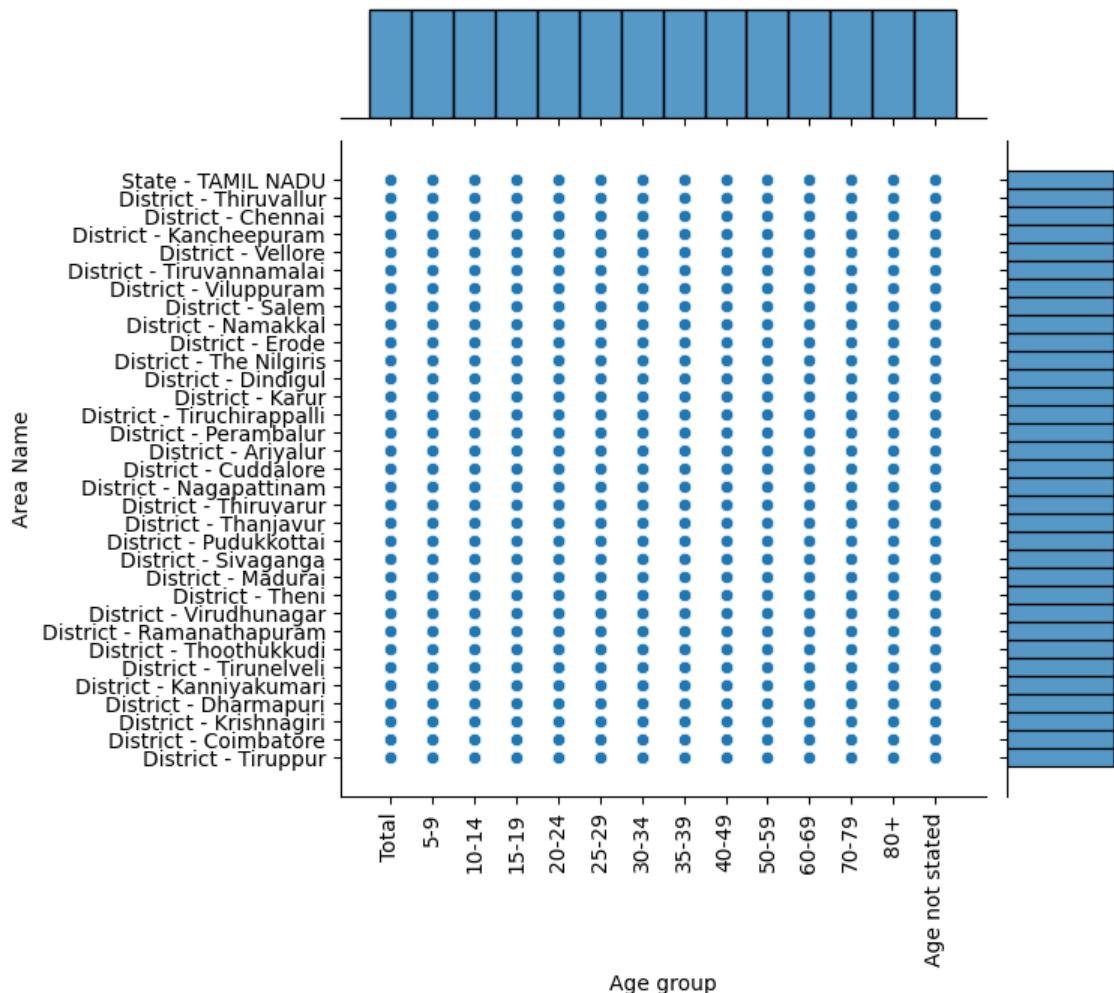
```
Out[21]: <AxesSubplot: xlabel='Age group', ylabel='count'>
```



```
In [168]: 1 plt.figure(figsize=(10,10))
2 sns.jointplot(dataset,x='Age group',y='Area Name')
3 plt.xticks(rotation=90)
```

```
Out[168]: ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],
 [Text(0, 0, 'Total'),
  Text(1, 0, '5-9'),
  Text(2, 0, '10-14'),
  Text(3, 0, '15-19'),
  Text(4, 0, '20-24'),
  Text(5, 0, '25-29'),
  Text(6, 0, '30-34'),
  Text(7, 0, '35-39'),
  Text(8, 0, '40-49'),
  Text(9, 0, '50-59'),
  Text(10, 0, '60-69'),
  Text(11, 0, '70-79'),
  Text(12, 0, '80+'),
  Text(13, 0, 'Age not stated')])
```

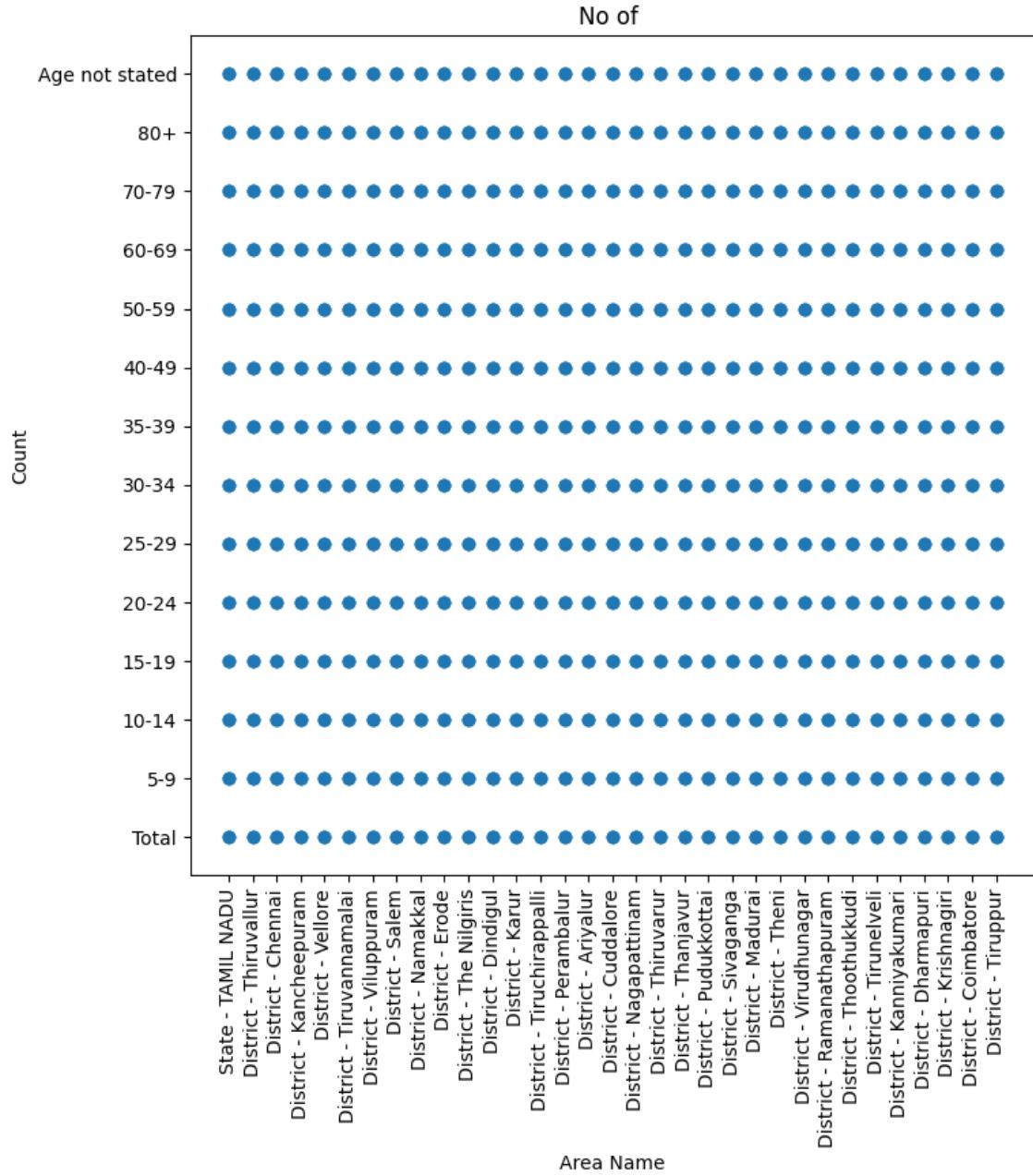
<Figure size 1000x1000 with 0 Axes>



In [163]:

```
1 plt.figure(figsize=(8,8))
2 plt.scatter(dataset[ 'Area Name' ],dataset[ 'Age group' ])
3 plt.xticks(rotation=90)
4 plt.xlabel("Area Name")
5 plt.ylabel("Count")
6
```

Out[163]: Text(0.5, 1.0, 'No of ')



Analysis by Age group and Industrial category by persons

```
In [19]: 1 IndustryCategorybyPerson = dataset[['Industrial Category - A - Cultiva  
2 'Industrial Category - A - Agricultu  
3 'Industrial Category - A - Plantatio  
4 'Industrial Category - B - Persons',  
5 'Industrial Category - C - HHI - Per  
6 'Industrial Category - C - Non HHI -  
7 'Industrial Category - D & E - Perso  
8 'Industrial Category - F - Persons',  
9 'Industrial Category - G - HHI - Per  
10 'Industrial Category - G - Non HHI -  
11 'Industrial Category - H - Persons',  
12 'Industrial Category - I - Persons',  
13 'Industrial Category - J - HHI - Per  
14 'Industrial Category - J - Non HHI -  
15 'Industrial Category - K to M - Pers  
16 'Industrial Category - N to O - Pers  
17 'Industrial Category - P to Q - Pers  
18 'Industrial Category - R to U - HHI  
19 'Industrial Category - R to U - Non
```

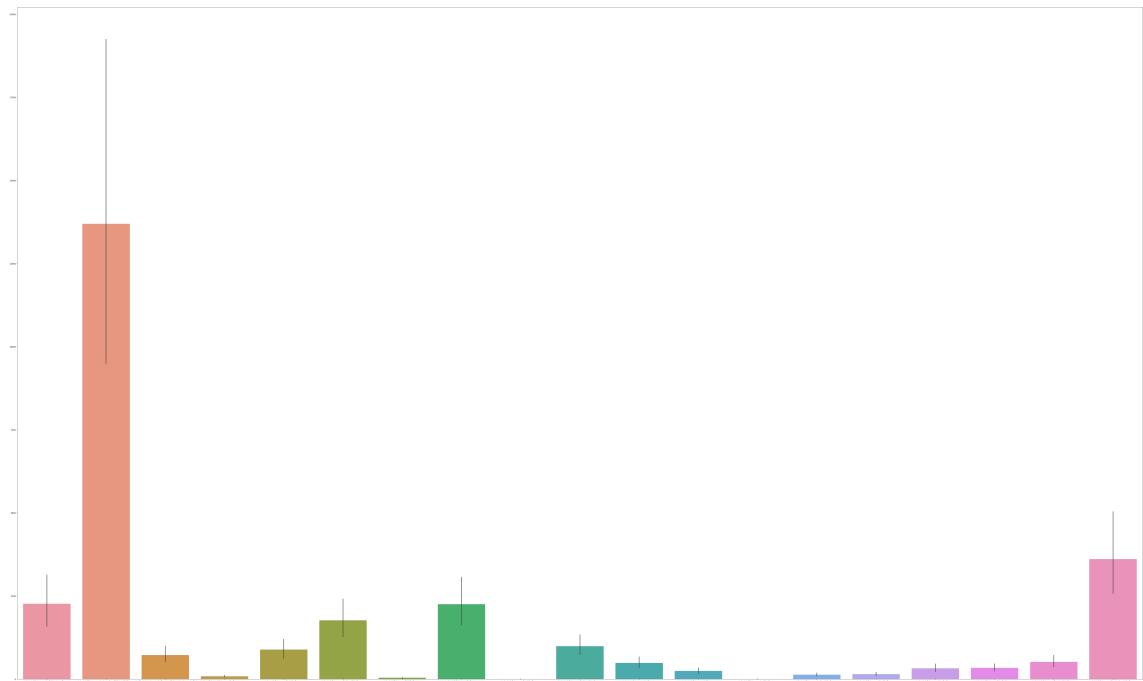
```
In [20]: 1 IndustryCategorybyPerson.head()
```

Out[20]:

	Industrial Category - A - Cultivators - Persons	Industrial Category - A - Agricultural labourers - Persons	Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons	Industrial Category - B - Persons	Industrial Category - C - HHI - Persons	Industrial Category - C - Non HHI - Persons	Industrial Category - D & E - Persons	Industrial Category - F - Persons
0	393082	2372446	125099	14979	154133	306528	7137	390275
1	3363	4169	198	6	228	246	6	264
2	5072	13939	834	70	996	2808	24	1522
3	17864	102106	4613	689	6642	28826	319	18661
4	33647	216966	9171	1490	14366	50791	1032	41762

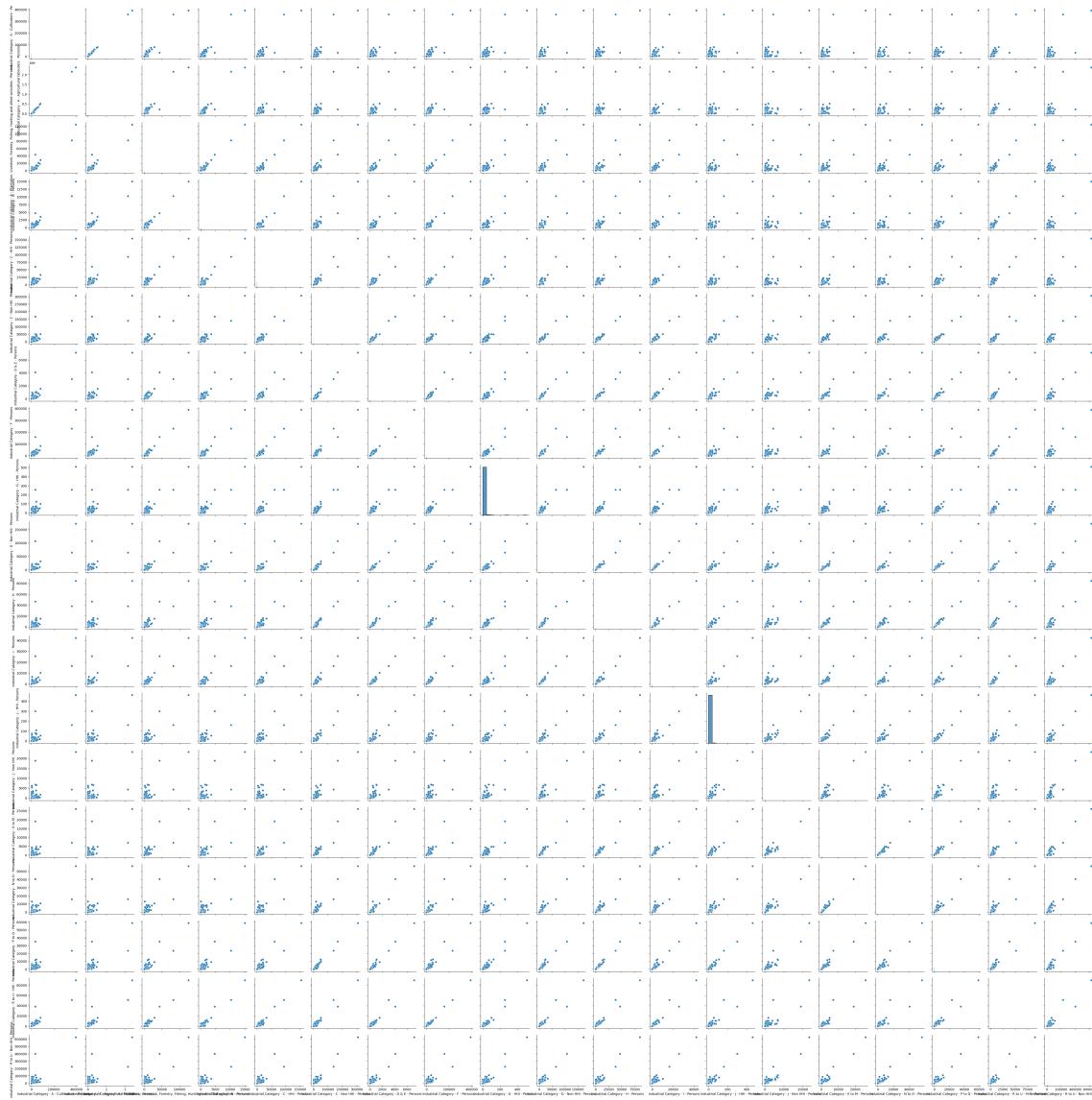
```
In [133]: 1 plt.figure(figsize=(100,60))  
2 sns.barplot(IndustryCategorybyPerson)
```

Out[133]: <AxesSubplot: >



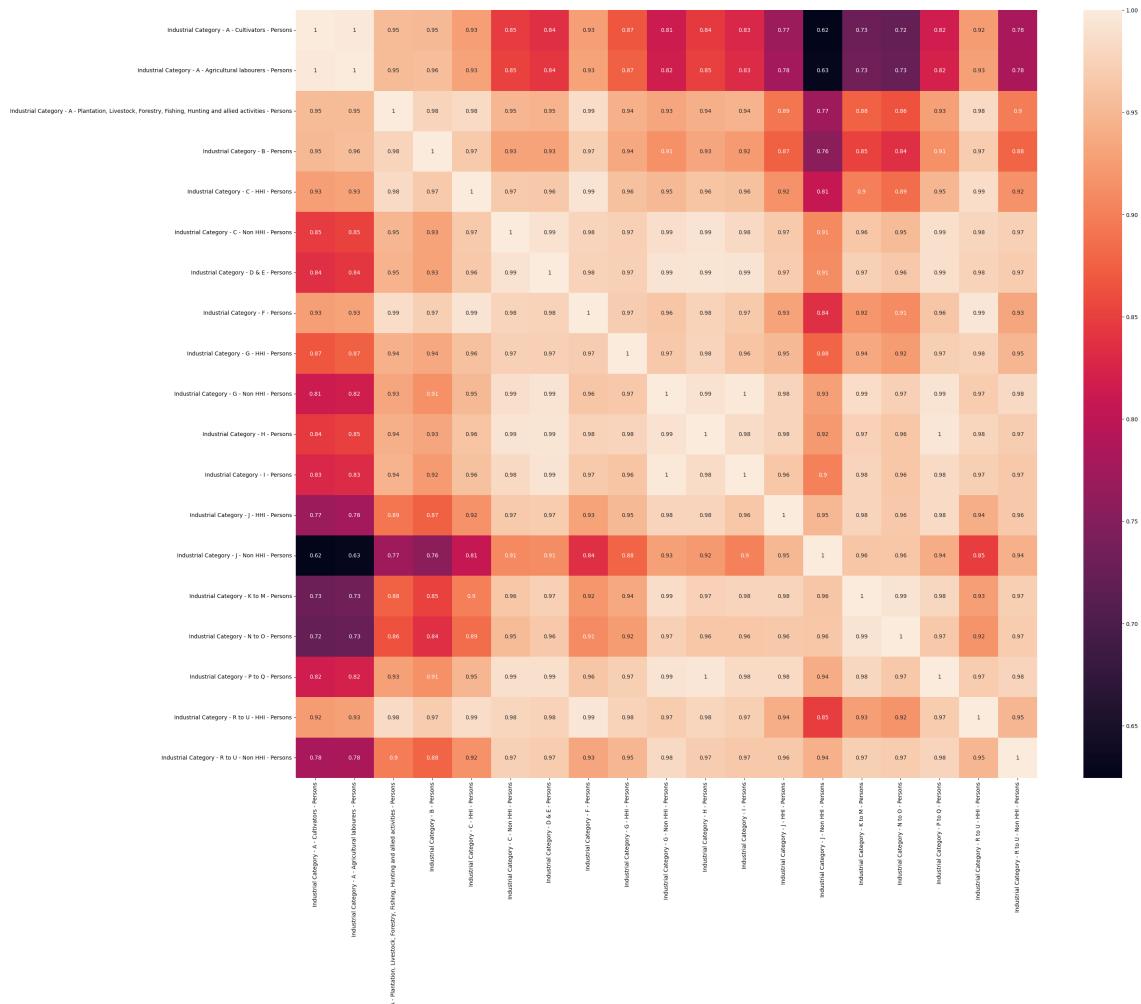
```
In [164]: 1 sns.pairplot(IndustryCategorybyPerson)
```

```
Out[164]: <seaborn.axisgrid.PairGrid at 0x23575f32850>
```



```
In [80]: 1 plt.figure(figsize=(30,25))
2 sns.heatmap(IndustryCategorybyPerson.corr(), annot=True)
```

Out[80]: <AxesSubplot: >



Visualization of Industrial Category by Age group and Area Name

Industrial Category A

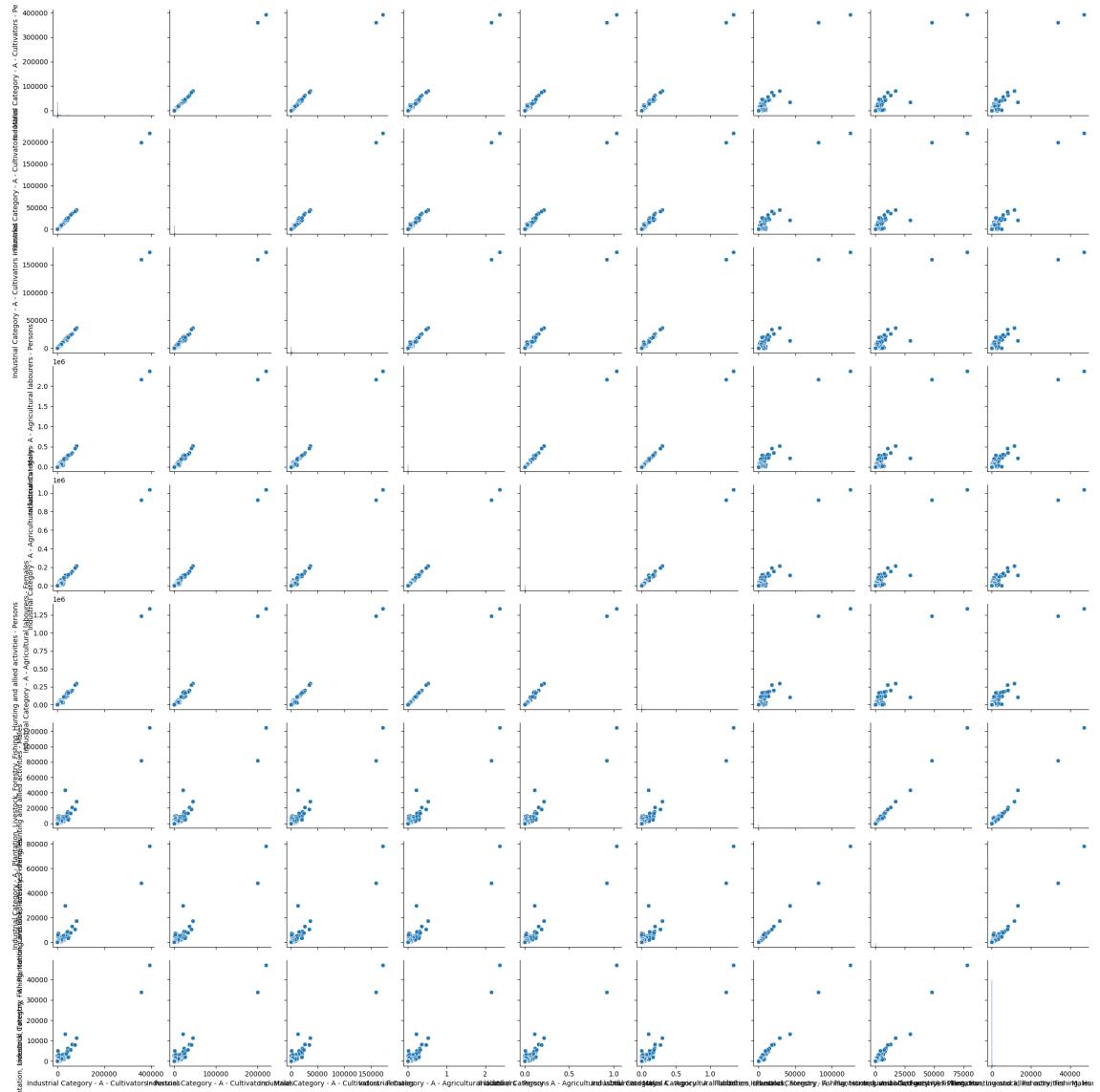
```
In [48]: 1 IndustrialCategoryA = dataset[['Industrial Category - A - Cultivators
2           'Industrial Category - A - Cultivators - Males',
3           'Industrial Category - A - Cultivators - Females',
4           'Industrial Category - A - Agricultural labourers - Persons',
5           'Industrial Category - A - Agricultural labourers - Males',
6           'Industrial Category - A - Agricultural labourers - Females',
7           'Industrial Category - A - Plantation, Livestock, Forestry, Fis
8           'Industrial Category - A - Plantation, Livestock, Forestry, Fis
9           'Industrial Category - A - Plantation, Livestock, Forestry, Fis
```

Pair Plot

```
In [70]: 1 plt.figure(figsize=(10,10))
2 sns.pairplot(IndustrialCategoryA)
```

Out[70]: <seaborn.axisgrid.PairGrid at 0x2349f3e7f10>

<Figure size 1000x1000 with 0 Axes>



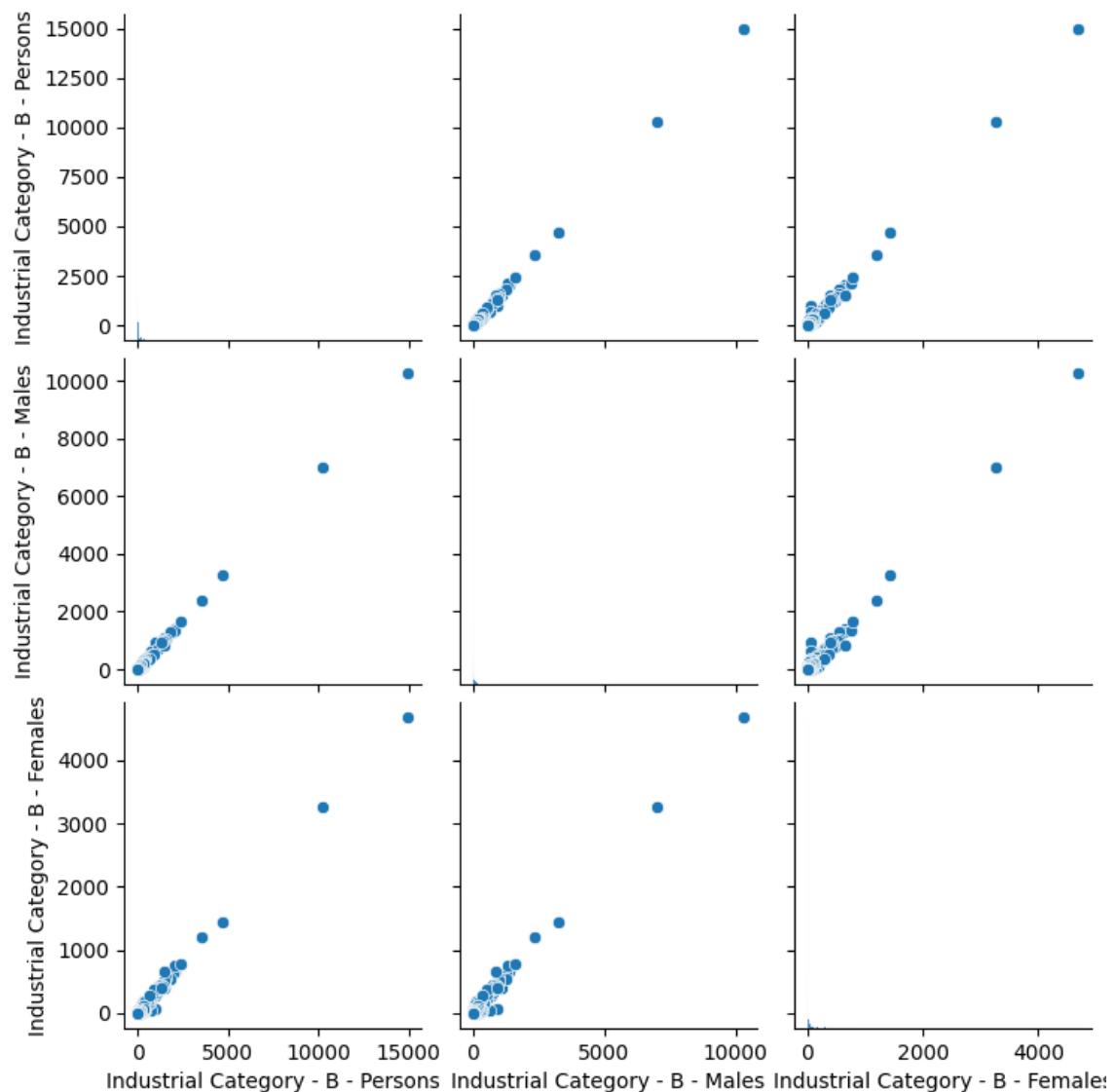
Industrial Category B

```
In [ ]: 1 IndustrialCategoryB = dataset[['Industrial Category - B - Persons', 'I
2 'Industrial Category - B - Females',]]
```

Pair Plot

```
In [60]: 1 sns.pairplot(IndustrialCategoryB)
          2 plt.xticks(rotation=30)
```

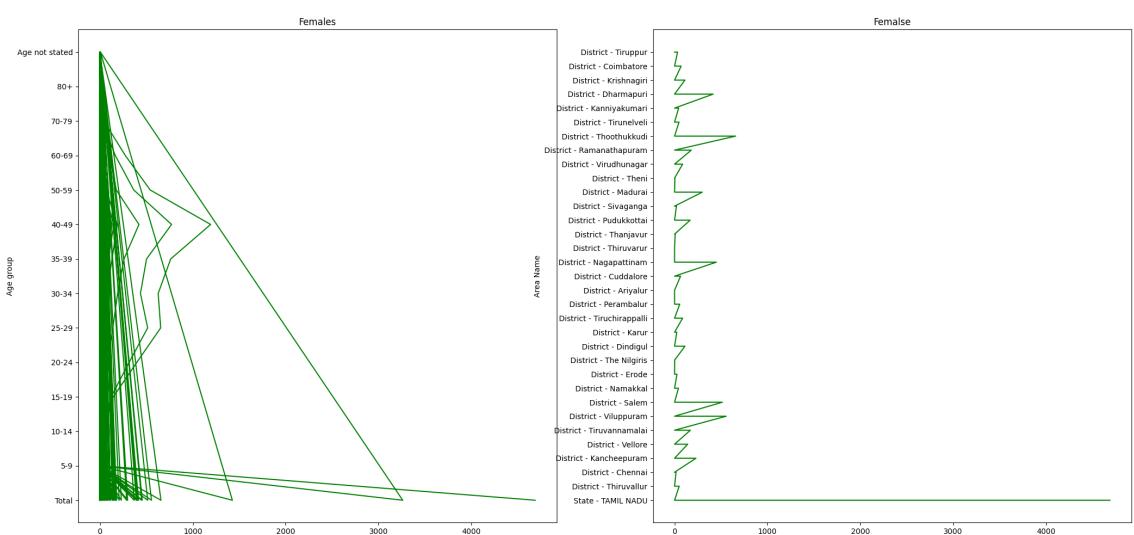
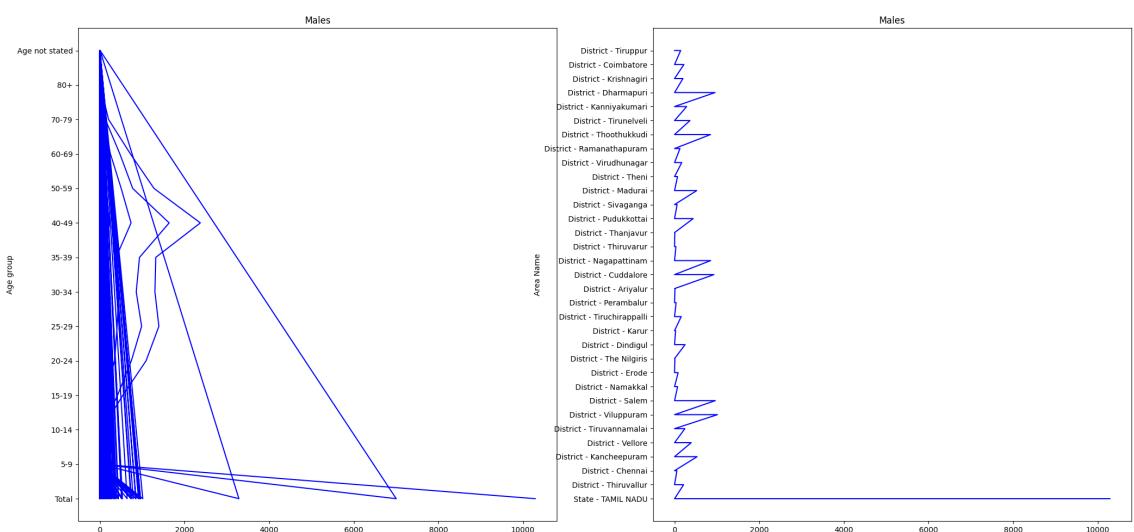
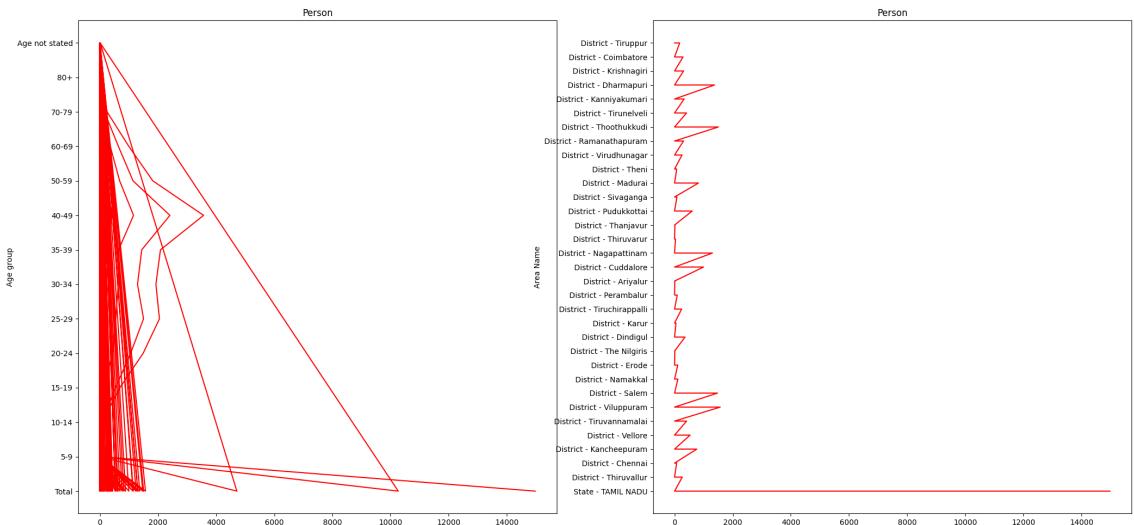
```
Out[60]: (array([-2000.,      0.,  2000.,  4000.,  6000.]),
 [Text(-2000.0, 0, '-2000'),
 Text(0.0, 0, '0'),
 Text(2000.0, 0, '2000'),
 Text(4000.0, 0, '4000'),
 Text(6000.0, 0, '6000')])
```



SUB plot Comparing Age group and Area Name

```
In [112]: 1 fig, ax = plt.subplots(figsize=(25,40),nrows=3, ncols=2)
2 ax[0, 0].plot(IndustrialCategoryB['Industrial Category - B - Persons'])
3 ax[0, 1].plot(IndustrialCategoryB['Industrial Category - B - Persons'])
4 ax[1, 0].plot(IndustrialCategoryB['Industrial Category - B - Males'],
5 ax[1, 1].plot(IndustrialCategoryB['Industrial Category - B - Males']),
6 ax[2, 0].plot(IndustrialCategoryB['Industrial Category - B - Females'])
7 ax[2, 1].plot(IndustrialCategoryB['Industrial Category - B - Females'])
8
9 ax[0,0].set_ylabel("Age group")
10 ax[0,1].set_ylabel("Area Name")
11 ax[0,0].set_title("Person")
12 ax[0,1].set_title("Person")
13
14
15 ax[1,0].set_ylabel("Age group")
16 ax[1,1].set_ylabel("Area Name")
17 ax[1,0].set_title("Males")
18 ax[1,1].set_title("Males")
19
20
21 ax[2,0].set_ylabel("Age group")
22 ax[2,1].set_ylabel("Area Name")
23 ax[2,0].set_title("Females")
24 ax[2,1].set_title("Femalse")
25
26
27
```

Out[112]: Text(0.5, 1.0, 'Femalse')



Industrial Category c

In [64]:

```
1 IndustrialCategoryC = dataset[['Industrial Category - C - HHI - Person'
2     'Industrial Category - C - HHI - Males',
3     'Industrial Category - C - HHI - Females',
4     'Industrial Category - C - Non HHI - Persons',
5     'Industrial Category - C - Non HHI - Males',
6     'Industrial Category - C - Non HHI - Females']]
```

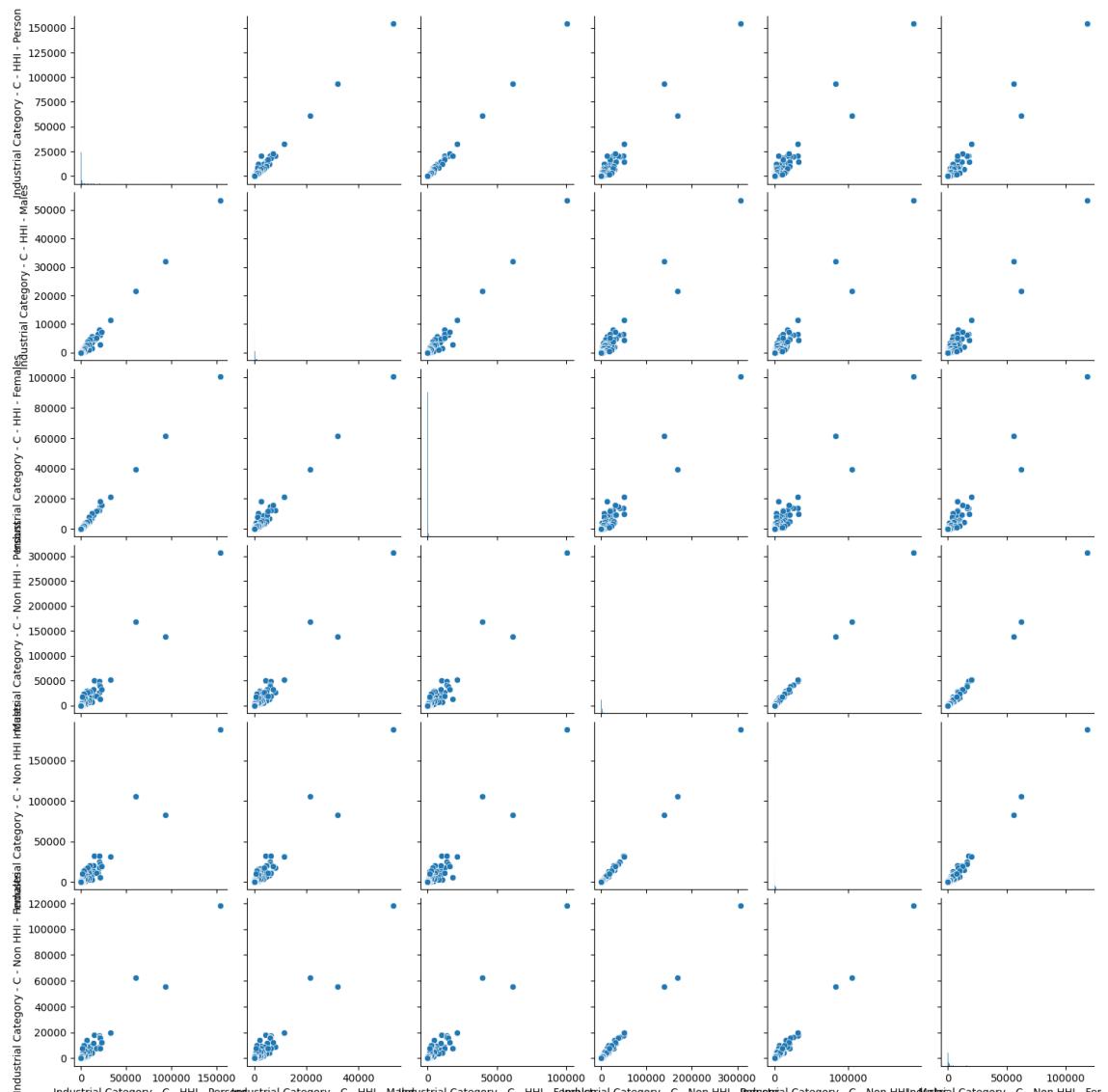
Pair Plot

In [165]:

```
1 sns.pairplot(IndustrialCategoryC)
2 plt.xticks(rotation=90)
```

Out[165]: (array([-50000., 0., 50000., 100000., 150000.]),

```
[Text(-50000.0, 0, '-50000'),
Text(0.0, 0, '0'),
Text(50000.0, 0, '50000'),
Text(100000.0, 0, '100000'),
Text(150000.0, 0, '150000')])
```



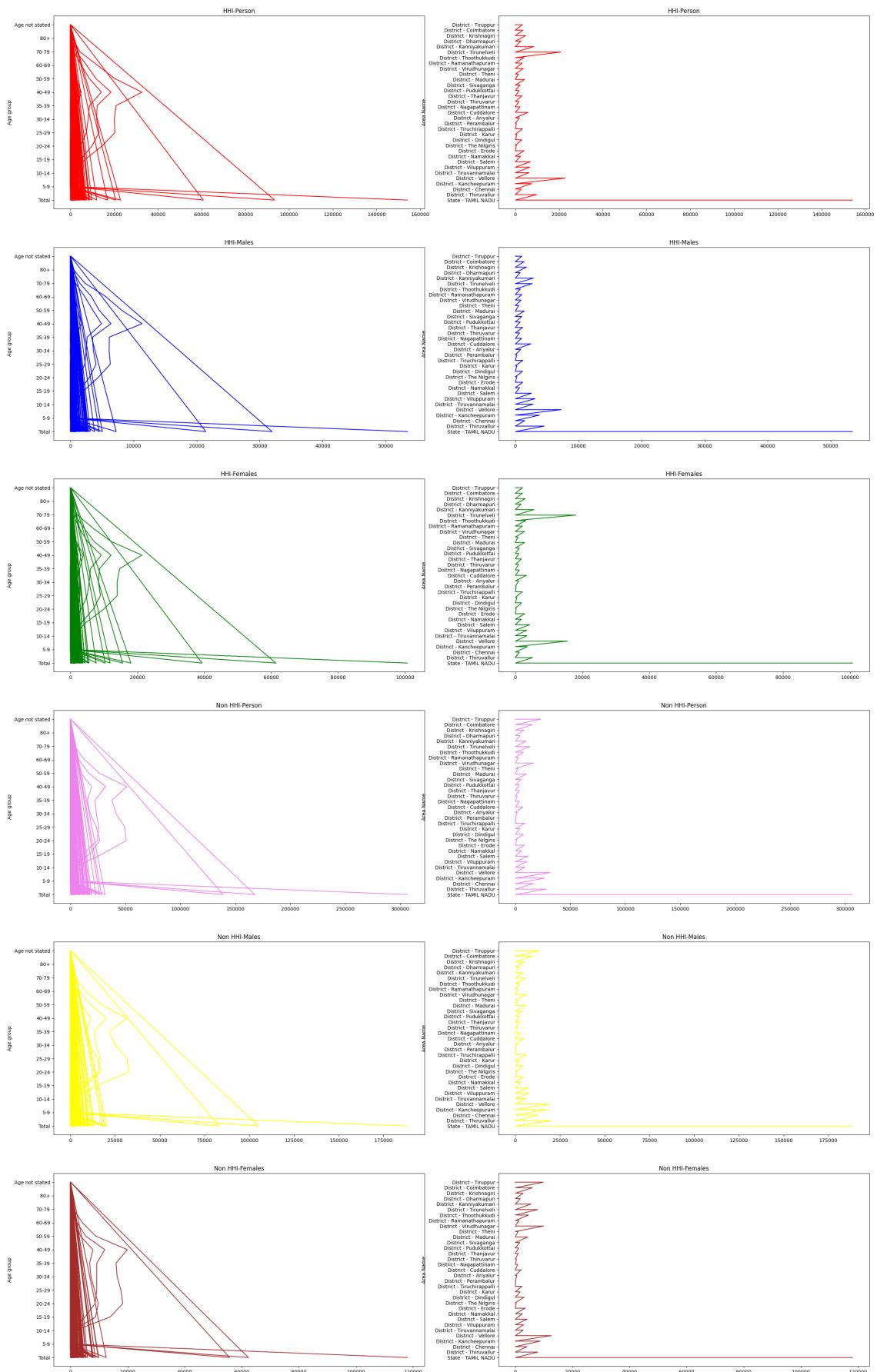
Sub plot

```

In [132]: 1 fig , ax =plt.subplots(figsize=(30,50),nrows=6,ncols=2)
2 ax[0, 0].plot(dataset['Industrial Category - C - HHI - Persons'], data
3 ax[0, 1].plot(dataset['Industrial Category - C - HHI - Persons'], data
4
5 ax[0,0].set_ylabel("Age group")
6 ax[0,1].set_ylabel("Area Name")
7 ax[0,0].set_title("HHI-Person")
8 ax[0,1].set_title("HHI-Person")
9
10 ax[1, 0].plot(dataset['Industrial Category - C - HHI - Males'], data
11 ax[1, 1].plot(dataset['Industrial Category - C - HHI - Males'], data
12 ax[1,0].set_ylabel("Age group")
13 ax[1,1].set_ylabel("Area Name")
14 ax[1,0].set_title("HHI-Males")
15 ax[1,1].set_title("HHI-Males")
16
17
18 ax[2, 0].plot(dataset['Industrial Category - C - HHI - Females'], data
19 ax[2, 1].plot(dataset['Industrial Category - C - HHI - Females'], data
20 ax[2,0].set_ylabel("Age group")
21 ax[2,1].set_ylabel("Area Name")
22 ax[2,0].set_title("HHI-Females")
23 ax[2,1].set_title("HHI-Females")
24
25 ax[3, 0].plot(dataset['Industrial Category - C - Non HHI - Persons'], data
26 ax[3, 1].plot(dataset['Industrial Category - C - Non HHI - Persons'], data
27 ax[3,0].set_ylabel("Age group")
28 ax[3,1].set_ylabel("Area Name")
29 ax[3,0].set_title("Non HHI-Person")
30 ax[3,1].set_title("Non HHI-Person")
31
32 ax[4, 0].plot(dataset['Industrial Category - C - Non HHI - Males'], data
33 ax[4, 1].plot(dataset['Industrial Category - C - Non HHI - Males'], data
34 ax[4,0].set_ylabel("Age group")
35 ax[4,1].set_ylabel("Area Name")
36 ax[4,0].set_title("Non HHI-Males")
37 ax[4,1].set_title("Non HHI-Males")
38
39 ax[5, 0].plot(dataset['Industrial Category - C - Non HHI - Females'], data
40 ax[5, 1].plot(dataset['Industrial Category - C - Non HHI - Females'], data
41 ax[5,0].set_ylabel("Age group")
42 ax[5,1].set_ylabel("Area Name")
43 ax[5,0].set_title("Non HHI-Females")
44 ax[5,1].set_title("Non HHI-Females")

```

Out[132]: Text(0.5, 1.0, 'Non HHI-Females')



Like wise visual all the industrial category