

# **TN Marginal Workers Assessment**

## **Team Member**

Name : **HARIHARAN R**

Register Number : **211521243059**

## **Applied Data Science Phase-3 document**

### **Team Members :**

- 1. A R HRUDAYABHIRAM**
- 2. ASWIN S**
- 3. DINESH S**
- 4. LAKSHMI KANTH**
- 5. HARIHARAN R**

## **Phase 3: Development Part 1**

### **Problem Statement:**

In this part you will begin building your project by loading and preprocessing the dataset.

Start the data analysis by loading and preprocessing the dataset

Load the dataset using Python and data manipulation libraries (e.g., pandas).

# phase3-nm-1

October 17, 2023

```
[2]: #Load the dataset

import pandas as pd
df=pd.read_csv('marginal_workers_tamil_nadu.csv.csv')
df
```

```
[2]:      Table Code State Code District Code      Area Name \
0      B0706      `33      `000  State - TAMIL NADU
1      B0706      `33      `000  State - TAMIL NADU
2      B0706      `33      `000  State - TAMIL NADU
3      B0706      `33      `000  State - TAMIL NADU
4      B0706      `33      `000  State - TAMIL NADU
...
1381    B0706      `33      `633  District - Tiruppur
1382    B0706      `33      `633  District - Tiruppur
1383    B0706      `33      `633  District - Tiruppur
1384    B0706      `33      `633  District - Tiruppur
1385    B0706      `33      `633  District - Tiruppur
```

```
      Total/ Rural/ Urban      Age group \
0      Total      Total
1      Total      `5-9
2      Total      `10-14
3      Total      15-19
4      Total      20-24
...
1381    Urban      50-59
1382    Urban      60-69
1383    Urban      70-79
1384    Urban      80+
1385    Urban  Age not stated
```

```
      Worked for 3 months or more but less than 6 months - Persons \
0      4218884
1      48238
2      76288
3      257605
```

4	478082
...	...
1381	4965
1382	2827
1383	920
1384	191
1385	31

Worked for 3 months or more but less than 6 months - Males \

0	2136881
1	24511
2	39191
3	141262
4	257149
...	...
1381	2800
1382	1590
1383	581
1384	104
1385	23

Worked for 3 months or more but less than 6 months - Females \

0	2082003
1	23727
2	37097
3	116343
4	220933
...	...
1381	2165
1382	1237
1383	339
1384	87
1385	8

Worked for less than 3 months - Persons ... \

0	723891	...
1	2051	...
2	6993	...
3	41938	...
4	81036	...
...	...	...
1381	901	...
1382	578	...
1383	204	...
1384	47	...
1385	9	...

Industrial Category - N to O - Females \	
0	14495
1	20
2	44
3	768
4	2267
...	...
1381	25
1382	7
1383	2
1384	0
1385	0

Industrial Category - P to Q - Persons \	
0	58788
1	312
2	506
3	2114
4	11529
...	...
1381	111
1382	21
1383	6
1384	2
1385	0

Industrial Category - P to Q - Males \	
0	19892
1	169
2	256
3	695
4	2861
...	...
1381	51
1382	6
1383	6
1384	0
1385	0

Industrial Category - P to Q - Females \	
0	38896
1	143
2	250
3	1419
4	8668
...	...
1381	60

1382	15
1383	0
1384	2
1385	0

	Industrial Category - R to U - HHI - Persons \
0	89703
1	842
2	1523
3	5349
4	10653
...	...
1381	119
1382	71
1383	22
1384	13
1385	3

	Industrial Category - R to U - HHI - Males \
0	21366
1	386
2	576
3	2065
4	2478
...	...
1381	26
1382	24
1383	9
1384	3
1385	0

	Industrial Category - R to U - HHI - Females \
0	68337
1	456
2	947
3	3284
4	8175
...	...
1381	93
1382	47
1383	13
1384	10
1385	3

	Industrial Category - R to U - Non HHI - Persons \
0	625350
1	40358

2	53900
3	89780
4	109641
...	...
1381	857
1382	668
1383	319
1384	72
1385	12

	Industrial Category - R to U - Non HHI - Males \
0	274811
1	20664
2	27501
3	46737
4	54832
...	...
1381	257
1382	274
1383	168
1384	30
1385	9

	Industrial Category - R to U - Non HHI - Females
0	350539
1	19694
2	26399
3	43043
4	54809
...	...
1381	600
1382	394
1383	151
1384	42
1385	3

[1386 rows x 69 columns]

```
[3]: # Clean the dataset by removing '' from the columns of Age group, state code,
      ↪ District Code
df['Age group'] = df['Age group'].str.replace('', '')
df['State Code'] = df['State Code'].str.replace('', '')
df['District Code'] = df['District Code'].str.replace('', '')
df
```

```
[3]: Table Code State Code District Code Area Name \
0      B0706      33      000 State - TAMIL NADU
```

1	B0706	33	000	State - TAMIL NADU
2	B0706	33	000	State - TAMIL NADU
3	B0706	33	000	State - TAMIL NADU
4	B0706	33	000	State - TAMIL NADU
...	...	...	...	...
1381	B0706	33	633	District - Tiruppur
1382	B0706	33	633	District - Tiruppur
1383	B0706	33	633	District - Tiruppur
1384	B0706	33	633	District - Tiruppur
1385	B0706	33	633	District - Tiruppur

	Total/ Rural/ Urban	Age group \
0	Total	Total
1	Total	5-9
2	Total	10-14
3	Total	15-19
4	Total	20-24
...	...	...
1381	Urban	50-59
1382	Urban	60-69
1383	Urban	70-79
1384	Urban	80+
1385	Urban	Age not stated

	Worked for 3 months or more but less than 6 months - Persons \
0	4218884
1	48238
2	76288
3	257605
4	478082
...	...
1381	4965
1382	2827
1383	920
1384	191
1385	31

	Worked for 3 months or more but less than 6 months - Males \
0	2136881
1	24511
2	39191
3	141262
4	257149
...	...
1381	2800
1382	1590
1383	581

1384	104
1385	23

	Worked for 3 months or more but less than 6 months - Females \
0	2082003
1	23727
2	37097
3	116343
4	220933
...	...
1381	2165
1382	1237
1383	339
1384	87
1385	8

	Worked for less than 3 months - Persons ... \
0	723891 ...
1	2051 ...
2	6993 ...
3	41938 ...
4	81036 ...
...	... ...
1381	901 ...
1382	578 ...
1383	204 ...
1384	47 ...
1385	9 ...

	Industrial Category - N to O - Females \
0	14495
1	20
2	44
3	768
4	2267
...	...
1381	25
1382	7
1383	2
1384	0
1385	0

	Industrial Category - P to Q - Persons \
0	58788
1	312
2	506
3	2114



4	11529
...	...
1381	111
1382	21
1383	6
1384	2
1385	0

	Industrial Category - P to Q - Males \
0	19892
1	169
2	256
3	695
4	2861
...	...
1381	51
1382	6
1383	6
1384	0
1385	0

	Industrial Category - P to Q - Females \
0	38896
1	143
2	250
3	1419
4	8668
...	...
1381	60
1382	15
1383	0
1384	2
1385	0

	Industrial Category - R to U - HHI - Persons \
0	89703
1	842
2	1523
3	5349
4	10653
...	...
1381	119
1382	71
1383	22
1384	13
1385	3

	Industrial Category - R to U - HHI - Males \
0	21366
1	386
2	576
3	2065
4	2478
...	...
1381	26
1382	24
1383	9
1384	3
1385	0

	Industrial Category - R to U - HHI - Females \
0	68337
1	456
2	947
3	3284
4	8175
...	...
1381	93
1382	47
1383	13
1384	10
1385	3

	Industrial Category - R to U - Non HHI - Persons \
0	625350
1	40358
2	53900
3	89780
4	109641
...	...
1381	857
1382	668
1383	319
1384	72
1385	12

	Industrial Category - R to U - Non HHI - Males \
0	274811
1	20664
2	27501
3	46737
4	54832
...	...
1381	257

1382	274
1383	168
1384	30
1385	9

	Industrial Category - R to U - Non HHI - Females
0	350539
1	19694
2	26399
3	43043
4	54809
...	...
1381	600
1382	394
1383	151
1384	42
1385	3

[1386 rows x 69 columns]

```
[ ]: # first few rows
df.head()
```

[ ]:	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban \
0	B0706	33	000	State - TAMIL NADU	Total
1	B0706	33	000	State - TAMIL NADU	Total
2	B0706	33	000	State - TAMIL NADU	Total
3	B0706	33	000	State - TAMIL NADU	Total
4	B0706	33	000	State - TAMIL NADU	Total

	Age group	Worked for 3 months or more but less than 6 months - Persons \
0	Total	4218884
1	5-9	48238
2	10-14	76288
3	15-19	257605
4	20-24	478082

	Worked for 3 months or more but less than 6 months - Males \
0	2136881
1	24511
2	39191
3	141262
4	257149

	Worked for 3 months or more but less than 6 months - Females \
0	2082003
1	23727

2	37097
3	116343
4	220933

	Worked for less than 3 months - Persons	...	\
0	723891	...	
1	2051	...	
2	6993	...	
3	41938	...	
4	81036	...	

	Industrial Category - N to O - Females	\
0	14495	
1	20	
2	44	
3	768	
4	2267	

	Industrial Category - P to Q - Persons	\
0	58788	
1	312	
2	506	
3	2114	
4	11529	

	Industrial Category - P to Q - Males	\
0	19892	
1	169	
2	256	
3	695	
4	2861	

	Industrial Category - P to Q - Females	\
0	38896	
1	143	
2	250	
3	1419	
4	8668	

	Industrial Category - R to U - HHI - Persons	\
0	89703	
1	842	
2	1523	
3	5349	
4	10653	

	Industrial Category - R to U - HHI - Males	\
--	--	---

0	21366
1	386
2	576
3	2065
4	2478

Industrial Category - R to U - HHI - Females \	
0	68337
1	456
2	947
3	3284
4	8175

Industrial Category - R to U - Non HHI - Persons \	
0	625350
1	40358
2	53900
3	89780
4	109641

Industrial Category - R to U - Non HHI - Males \	
0	274811
1	20664
2	27501
3	46737
4	54832

Industrial Category - R to U - Non HHI - Females	
0	350539
1	19694
2	26399
3	43043
4	54809

[5 rows x 69 columns]

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

[ ]:	Table Code	State Code	District Code	Area Name \
1381	B0706	33	633	District - Tiruppur
1382	B0706	33	633	District - Tiruppur
1383	B0706	33	633	District - Tiruppur
1384	B0706	33	633	District - Tiruppur
1385	B0706	33	633	District - Tiruppur

Total/ Rural/ Urban	Age group \
1381 Urban	50-59

1382	Urban	60-69
1383	Urban	70-79
1384	Urban	80+
1385	Urban	Age not stated

	Worked for 3 months or more but less than 6 months - Persons \
1381	4965
1382	2827
1383	920
1384	191
1385	31

	Worked for 3 months or more but less than 6 months - Males \
1381	2800
1382	1590
1383	581
1384	104
1385	23

	Worked for 3 months or more but less than 6 months - Females \
1381	2165
1382	1237
1383	339
1384	87
1385	8

	Worked for less than 3 months - Persons ... \
1381	901 ...
1382	578 ...
1383	204 ...
1384	47 ...
1385	9 ...

	Industrial Category - N to O - Females \
1381	25
1382	7
1383	2
1384	0
1385	0

	Industrial Category - P to Q - Persons \
1381	111
1382	21
1383	6
1384	2
1385	0

	Industrial Category - P to Q - Males \
1381	51
1382	6
1383	6
1384	0
1385	0

	Industrial Category - P to Q - Females \
1381	60
1382	15
1383	0
1384	2
1385	0

	Industrial Category - R to U - HHI - Persons \
1381	119
1382	71
1383	22
1384	13
1385	3

	Industrial Category - R to U - HHI - Males \
1381	26
1382	24
1383	9
1384	3
1385	0

	Industrial Category - R to U - HHI - Females \
1381	93
1382	47
1383	13
1384	10
1385	3

	Industrial Category - R to U - Non HHI - Persons \
1381	857
1382	668
1383	319
1384	72
1385	12

	Industrial Category - R to U - Non HHI - Males \
1381	257
1382	274
1383	168
1384	30

1385

9

Industrial Category - R to U - Non HHI - Females

1381	600
1382	394
1383	151
1384	42
1385	3

[5 rows x 69 columns]

```
[ ]: # List the column names
print(df.columns)
```

```
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')

```

```

[ ]: #info
      print(df.info())

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1386 entries, 0 to 1385
Data columns (total 69 columns):
 #   Column
Non-Null Count  Dtype
---  -
-----
0    Table Code
1386 non-null   object
1    State Code
1386 non-null   object
2    District Code

```

1386 non-null object  
 3 Area Name  
 1386 non-null object  
 4 Total/ Rural/ Urban  
 1386 non-null object  
 5 Age group  
 1386 non-null object  
 6 Worked for 3 months or more but less than 6 months - Persons  
 1386 non-null int64  
 7 Worked for 3 months or more but less than 6 months - Males  
 1386 non-null int64  
 8 Worked for 3 months or more but less than 6 months - Females  
 1386 non-null int64  
 9 Worked for less than 3 months - Persons  
 1386 non-null int64  
 10 Worked for less than 3 months - Males  
 1386 non-null int64  
 11 Worked for less than 3 months - Females  
 1386 non-null int64  
 12 Industrial Category - A - Cultivators - Persons  
 1386 non-null int64  
 13 Industrial Category - A - Cultivators - Males  
 1386 non-null int64  
 14 Industrial Category - A - Cultivators - Females  
 1386 non-null int64  
 15 Industrial Category - A - Agricultural labourers - Persons  
 1386 non-null int64  
 16 Industrial Category - A - Agricultural labourers - Males  
 1386 non-null int64  
 17 Industrial Category - A - Agricultural labourers - Females  
 1386 non-null int64  
 18 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting  
 and allied activities - Persons 1386 non-null int64  
 19 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting  
 and allied activities - Males 1386 non-null int64  
 20 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting  
 and allied activities - Females 1386 non-null int64  
 21 Industrial Category - B - Persons  
 1386 non-null int64  
 22 Industrial Category - B - Males  
 1386 non-null int64  
 23 Industrial Category - B - Females  
 1386 non-null int64  
 24 Industrial Category - C - HHI - Persons  
 1386 non-null int64  
 25 Industrial Category - C - HHI - Males  
 1386 non-null int64  
 26 Industrial Category - C - HHI - Females

1386 non-null int64  
27 Industrial Category - C - Non HHI - Persons  
1386 non-null int64  
28 Industrial Category - C - Non HHI - Males  
1386 non-null int64  
29 Industrial Category - C - Non HHI - Females  
1386 non-null int64  
30 Industrial Category - D & E - Persons  
1386 non-null int64  
31 Industrial Category - D & E - Males  
1386 non-null int64  
32 Industrial Category - D & E - Females  
1386 non-null int64  
33 Industrial Category - F - Persons  
1386 non-null int64  
34 Industrial Category - F - Males  
1386 non-null int64  
35 Industrial Category - F - Females  
1386 non-null int64  
36 Industrial Category - G - HHI - Persons  
1386 non-null int64  
37 Industrial Category - G - HHI - Males  
1386 non-null int64  
38 Industrial Category - G - HHI - Females  
1386 non-null int64  
39 Industrial Category - G - Non HHI - Persons  
1386 non-null int64  
40 Industrial Category - G - Non HHI - Males  
1386 non-null int64  
41 Industrial Category - G - Non HHI - Females  
1386 non-null int64  
42 Industrial Category - H - Persons  
1386 non-null int64  
43 Industrial Category - H - Males  
1386 non-null int64  
44 Industrial Category - H - Females  
1386 non-null int64  
45 Industrial Category - I - Persons  
1386 non-null int64  
46 Industrial Category - I - Males  
1386 non-null int64  
47 Industrial Category - I - Females  
1386 non-null int64  
48 Industrial Category - J - HHI - Persons  
1386 non-null int64  
49 Industrial Category - J - HHI - Males  
1386 non-null int64  
50 Industrial Category - J - HHI - Females

```

1386 non-null    int64
51  Industrial Category - J - Non HHI - Persons
1386 non-null    int64
52  Industrial Category - J - Non HHI - Males
1386 non-null    int64
53  Industrial Category - J - Non HHI - Females
1386 non-null    int64
54  Industrial Category - K to M - Persons
1386 non-null    int64
55  Industrial Category - K to M - Males
1386 non-null    int64
56  Industrial Category - K to M - Females
1386 non-null    int64
57  Industrial Category - N to O - Persons
1386 non-null    int64
58  Industrial Category - N to O - Males
1386 non-null    int64
59  Industrial Category - N to O - Females
1386 non-null    int64
60  Industrial Category - P to Q - Persons
1386 non-null    int64
61  Industrial Category - P to Q - Males
1386 non-null    int64
62  Industrial Category - P to Q - Females
1386 non-null    int64
63  Industrial Category - R to U - HHI - Persons
1386 non-null    int64
64  Industrial Category - R to U - HHI - Males
1386 non-null    int64
65  Industrial Category - R to U - HHI - Females
1386 non-null    int64
66  Industrial Category - R to U - Non HHI - Persons
1386 non-null    int64
67  Industrial Category - R to U - Non HHI - Males
1386 non-null    int64
68  Industrial Category - R to U - Non HHI - Females
1386 non-null    int64
dtypes: int64(63), object(6)
memory usage: 747.3+ KB
None

```

```

[ ]: # Check for missing values
df.isnull()

```

```

[ ]:
   Table Code  State Code  District Code  Area Name  Total/ Rural/ Urban \
0         False         False         False         False         False
1         False         False         False         False         False

```

2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
1381	False	False	False	False	False
1382	False	False	False	False	False
1383	False	False	False	False	False
1384	False	False	False	False	False
1385	False	False	False	False	False

	Age group \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Worked for 3 months or more but less than 6 months - Persons \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Worked for 3 months or more but less than 6 months - Males \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False

1385 False

Worked for 3 months or more but less than 6 months - Females \

0 False

1 False

2 False

3 False

4 False

...

1381 False

1382 False

1383 False

1384 False

1385 False

Worked for less than 3 months - Persons ... \

0 False ...

1 False ...

2 False ...

3 False ...

4 False ...

...

1381 False ...

1382 False ...

1383 False ...

1384 False ...

1385 False ...

Industrial Category - N to O - Females \

0 False

1 False

2 False

3 False

4 False

...

1381 False

1382 False

1383 False

1384 False

1385 False

Industrial Category - P to Q - Persons \

0 False

1 False

2 False

3 False

4 False

...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Industrial Category - P to Q - Males \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Industrial Category - P to Q - Females \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Industrial Category - R to U - HHI - Persons \
0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

	Industrial Category - R to U - HHI - Males \
--	--

0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

Industrial Category - R to U - HHI - Females \

0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

Industrial Category - R to U - Non HHI - Persons \

0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False
1383	False
1384	False
1385	False

Industrial Category - R to U - Non HHI - Males \

0	False
1	False
2	False
3	False
4	False
...	...
1381	False
1382	False



```

1383                                     False
1384                                     False
1385                                     False

      Industrial Category - R to U - Non HHI - Females
0                                     False
1                                     False
2                                     False
3                                     False
4                                     False
...                                  ...
1381                                    False
1382                                    False
1383                                    False
1384                                    False
1385                                    False

```

[1386 rows x 69 columns]

```
[ ]: # Sum of missing values
      print(df.isnull().sum())
```

```

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

```

```
[ ]: # Data types of columns
      print(df.dtypes)
```

```

Table Code                                object
State Code                               object
District Code                             object
Area Name                                object
Total/ Rural/ Urban                       object
..
Industrial Category - R to U - HHI - Males  int64
Industrial Category - R to U - HHI - Females  int64
Industrial Category - R to U - Non HHI - Persons  int64
Industrial Category - R to U - Non HHI - Males  int64

```

Industrial Category - R to U - Non HHI - Females      int64  
Length: 69, dtype: object

```
[ ]: # Convert columns to numeric, handling errors as NaN
numeric_columns = ['Industrial Category - A - Cultivators - Persons',
↳ 'Industrial Category - A - Cultivators - Males',
    'Industrial Category - A - Cultivators - Females', 'Industrial Category - B
↳ - Persons',
    'Industrial Category - B - Males', 'Industrial Category - B - Females',
    'Industrial Category - C - HHI - Males',
    'Industrial Category - C - HHI - Females',
    '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']
# Convert columns to numeric, handling errors as NaN
df[numeric_columns] = df[numeric_columns].apply(pd.to_numeric, errors='coerce')

# Drop rows with NaN values in numeric columns
df.dropna(subset=numeric_columns, inplace=True)

```

```

[ ]: # Exclude rows where 'Age group' is "Total"
df = df[df['Age group'] != 'Total']
df

```

```

[ ]:      Table Code State Code District Code      Area Name \
1          B0706      33      000 State - TAMIL NADU
2          B0706      33      000 State - TAMIL NADU
3          B0706      33      000 State - TAMIL NADU
4          B0706      33      000 State - TAMIL NADU
5          B0706      33      000 State - TAMIL NADU
...
1381        B0706      33      633 District - Tiruppur
1382        B0706      33      633 District - Tiruppur
1383        B0706      33      633 District - Tiruppur
1384        B0706      33      633 District - Tiruppur
1385        B0706      33      633 District - Tiruppur

```

```

      Total/ Rural/ Urban      Age group \
1          Total      5-9
2          Total     10-14
3          Total     15-19
4          Total     20-24
5          Total     25-29
...
1381        Urban     50-59
1382        Urban     60-69
1383        Urban     70-79
1384        Urban      80+
1385        Urban Age not stated

```

```

      Worked for 3 months or more but less than 6 months - Persons \
1          48238
2          76288
3          257605
4          478082
5          554851
...
1381        4965
1382        2827

```

1383	920
1384	191
1385	31

	Worked for 3 months or more but less than 6 months - Males \
1	24511
2	39191
3	141262
4	257149
5	283442
...	...
1381	2800
1382	1590
1383	581
1384	104
1385	23

	Worked for 3 months or more but less than 6 months - Females \
1	23727
2	37097
3	116343
4	220933
5	271409
...	...
1381	2165
1382	1237
1383	339
1384	87
1385	8

	Worked for less than 3 months - Persons ... \
1	2051 ...
2	6993 ...
3	41938 ...
4	81036 ...
5	91694 ...
...	... ...
1381	901 ...
1382	578 ...
1383	204 ...
1384	47 ...
1385	9 ...

	Industrial Category - N to O - Females \
1	20
2	44
3	768

4	2267
5	2285
...	...
1381	25
1382	7
1383	2
1384	0
1385	0

	Industrial Category - P to Q - Persons \
1	312
2	506
3	2114
4	11529
5	12528
...	...
1381	111
1382	21
1383	6
1384	2
1385	0

	Industrial Category - P to Q - Males \
1	169
2	256
3	695
4	2861
5	4506
...	...
1381	51
1382	6
1383	6
1384	0
1385	0

	Industrial Category - P to Q - Females \
1	143
2	250
3	1419
4	8668
5	8022
...	...
1381	60
1382	15
1383	0
1384	2
1385	0

	Industrial Category - R to U - HHI - Persons \
1	842
2	1523
3	5349
4	10653
5	12713
...	...
1381	119
1382	71
1383	22
1384	13
1385	3

	Industrial Category - R to U - HHI - Males \
1	386
2	576
3	2065
4	2478
5	2745
...	...
1381	26
1382	24
1383	9
1384	3
1385	0

	Industrial Category - R to U - HHI - Females \
1	456
2	947
3	3284
4	8175
5	9968
...	...
1381	93
1382	47
1383	13
1384	10
1385	3

	Industrial Category - R to U - Non HHI - Persons \
1	40358
2	53900
3	89780
4	109641
5	80299
...	...

1381	857
1382	668
1383	319
1384	72
1385	12

	Industrial Category - R to U - Non HHI - Males \
1	20664
2	27501
3	46737
4	54832
5	34058
...	...
1381	257
1382	274
1383	168
1384	30
1385	9

	Industrial Category - R to U - Non HHI - Females
1	19694
2	26399
3	43043
4	54809
5	46241
...	...
1381	600
1382	394
1383	151
1384	42
1385	3

[1287 rows x 69 columns]

```
[ ]: # Standardizing Data
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
numeric_columns = ['Industrial Category - A - Cultivators - Persons',
                    'Industrial Category - A - Cultivators - Males',
                    'Industrial Category - A - Cultivators - Females', 'Industrial Category - B -
                    Persons',
                    'Industrial Category - B - Males', 'Industrial Category - B - Females',
                    'Industrial Category - C - HHI - Males',
                    'Industrial Category - C - HHI - Females',
                    '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']

features = numeric_columns[:] # Use all the industrial categories (starting
↪from the 2nd column)

df[features] = scaler.fit_transform(df[features])
print(df[features])

```

```

Industrial Category - A - Cultivators - Persons \
1                                         0.405030
2                                         0.728290
3                                         3.147914
4                                         6.133290

```



5	8.044287
...	...
1381	-0.195147
1382	-0.207821
1383	-0.219170
1384	-0.229006
1385	-0.230897

Industrial Category - A - Cultivators - Males \

1	0.307553
2	0.610609
3	2.932686
4	5.506971
5	7.323614
...	...
1381	-0.193474
1382	-0.208389
1383	-0.217203
1384	-0.229406
1385	-0.231779

Industrial Category - A - Cultivators - Females \

1	0.523249
2	0.868821
3	3.388315
4	6.860569
5	8.872094
...	...
1381	-0.195455
1382	-0.205208
1383	-0.219624
1384	-0.226408
1385	-0.227680

Industrial Category - B - Persons    Industrial Category - B - Males \

1	-0.203482	-0.237190
2	0.117635	0.052054
3	3.223442	3.797399
4	7.242425	7.861654
5	10.052201	10.108861
...	...	...
1381	-0.188430	-0.170442
1382	-0.218534	-0.214941
1383	-0.233587	-0.237190
1384	-0.233587	-0.237190
1385	-0.233587	-0.237190

Industrial Category - B - Females \

1	-0.130993
2	0.250980
3	1.992775
4	5.858338
5	9.785017
...	...
1381	-0.222666
1382	-0.222666
1383	-0.222666
1384	-0.222666
1385	-0.222666

Industrial Category - C - HHI - Males \

1	-0.080377
2	0.345587
3	2.821027
4	6.378810
5	9.330239
...	...
1381	-0.090988
1382	-0.118274
1383	-0.195585
1384	-0.238029
1385	-0.247125

Industrial Category - C - HHI - Females \

1	-0.153221
2	0.223598
3	3.328684
4	7.489173
5	10.578784
...	...
1381	-0.099832
1382	-0.158638
1383	-0.214348
1384	-0.231371
1385	-0.242203

Industrial Category - D & E - Males \

1	-0.186511
2	-0.059264
3	3.389144
4	11.609332
5	11.672956
...	...
1381	-0.161062
1382	-0.084713
1383	-0.211961

1384	-0.237410
1385	-0.237410

	Industrial Category - D & E - Females	...	\
1	-0.098491	...	
2	0.418363	...	
3	1.968925	...	
4	6.297577	...	
5	8.494206	...	
...	...	...	
1381	0.030722	...	
1382	-0.098491	...	
1383	-0.227705	...	
1384	-0.227705	...	
1385	-0.227705	...	

	Industrial Category - N to O - Females	\
1	-0.128224	
2	-0.005377	
3	3.700506	
4	11.373321	
5	11.465456	
...	...	
1381	-0.102631	
1382	-0.194766	
1383	-0.220359	
1384	-0.230597	
1385	-0.230597	

	Industrial Category - P to Q - Persons	\
1	0.163706	
2	0.409353	
3	2.445439	
4	14.366926	
5	15.631883	
...	...	
1381	-0.090805	
1382	-0.204765	
1383	-0.223758	
1384	-0.228823	
1385	-0.231356	

	Industrial Category - P to Q - Males	\
1	0.413473	
2	0.749111	
3	2.442731	
4	10.798952	
5	17.145205	

...	...
1381	-0.041759
1382	-0.215365
1383	-0.215365
1384	-0.238513
1385	-0.238513

	Industrial Category - P to Q - Females \
1	0.041268
2	0.240973
3	2.422791
4	15.952306
5	14.746613
...	...
1381	-0.113643
1382	-0.197631
1383	-0.225627
1384	-0.221894
1385	-0.225627

	Industrial Category - R to U - HHI - Persons \
1	0.529568
2	1.169897
3	4.767403
4	9.754640
5	11.691614
...	...
1381	-0.150254
1382	-0.195387
1383	-0.241461
1384	-0.249923
1385	-0.259326

	Industrial Category - R to U - HHI - Males \
1	1.331669
2	2.123351
3	8.327638
4	10.048505
5	11.161026
...	...
1381	-0.168360
1382	-0.176693
1383	-0.239195
1384	-0.264195
1385	-0.276695

	Industrial Category - R to U - HHI - Females \
1	0.293391

2	0.884728
3	3.699301
4	9.589792
5	11.749198
...	...
1381	-0.143790
1382	-0.199190
1383	-0.240138
1384	-0.243751
1385	-0.252181

Industrial Category - R to U - Non HHI - Persons \

1	5.417652
2	7.327504
3	12.387723
4	15.188754
5	11.050601
...	...
1381	-0.153244
1382	-0.179899
1383	-0.229119
1384	-0.263954
1385	-0.272415

Industrial Category - R to U - Non HHI - Males \

1	6.125273
2	8.239293
3	14.187119
4	16.690116
5	10.266736
...	...
1381	-0.184629
1382	-0.179373
1383	-0.212148
1384	-0.254818
1385	-0.261311

Industrial Category - R to U - Non HHI - Females

1	4.696718
2	6.389397
3	10.591180
4	13.561511
5	11.398516
...	...
1381	-0.123568
1382	-0.175573
1383	-0.236919
1384	-0.264436

1385

-0.274281

[1287 rows x 46 columns]

```
[ ]: # Unique Values in Categorical Columns
for column in df.select_dtypes(include=['object']):
    print(f"Unique values in {column}: {df[column].nunique()}")
```

```
Unique values in Table Code: 1
Unique values in State Code: 1
Unique values in District Code: 33
Unique values in Area Name: 33
Unique values in Total/ Rural/ Urban: 3
Unique values in Age group: 14
```

```
[ ]: correlation_matrix = df.corr()
print(correlation_matrix)
```

Worked for 3 months or more

```
but less than 6 months - Persons \
Worked for 3 months or more but less than 6 mon...
1.000000
Worked for 3 months or more but less than 6 mon...
0.998829
Worked for 3 months or more but less than 6 mon...
0.998825
Worked for less than 3 months - Persons
0.999108
Worked for less than 3 months - Males
0.994988
...
...
Industrial Category - R to U - HHI - Males
0.990432
Industrial Category - R to U - HHI - Females
0.980585
Industrial Category - R to U - Non HHI - Persons
0.894684
Industrial Category - R to U - Non HHI - Males
0.902864
Industrial Category - R to U - Non HHI - Females
0.884780
```

Worked for 3 months or more

```
but less than 6 months - Males \
Worked for 3 months or more but less than 6 mon...
0.998829
Worked for 3 months or more but less than 6 mon...
```

1.000000  
 Worked for 3 months or more but less than 6 mon...  
 0.995311  
 Worked for less than 3 months - Persons  
 0.998599  
 Worked for less than 3 months - Males  
 0.998280  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.995125  
 Industrial Category - R to U - HHI - Females  
 0.987377  
 Industrial Category - R to U - Non HHI - Persons  
 0.913023  
 Industrial Category - R to U - Non HHI - Males  
 0.919946  
 Industrial Category - R to U - Non HHI - Females  
 0.904020

Worked for 3 months or more

but less than 6 months - Females \  
 Worked for 3 months or more but less than 6 mon...  
 0.998825  
 Worked for 3 months or more but less than 6 mon...  
 0.995311  
 Worked for 3 months or more but less than 6 mon...  
 1.000000  
 Worked for less than 3 months - Persons  
 0.997273  
 Worked for less than 3 months - Males  
 0.989355  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.983406  
 Industrial Category - R to U - HHI - Females  
 0.971479  
 Industrial Category - R to U - Non HHI - Persons  
 0.874213  
 Industrial Category - R to U - Non HHI - Males  
 0.883634  
 Industrial Category - R to U - Non HHI - Females  
 0.863430

Worked for less than 3

months - Persons \  
 Worked for 3 months or more but less than 6 mon...

0.999108  
 Worked for 3 months or more but less than 6 mon...  
 0.998599  
 Worked for 3 months or more but less than 6 mon...  
 0.997273  
 Worked for less than 3 months - Persons  
 1.000000  
 Worked for less than 3 months - Males  
 0.996781  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.991516  
 Industrial Category - R to U - HHI - Females  
 0.982091  
 Industrial Category - R to U - Non HHI - Persons  
 0.894641  
 Industrial Category - R to U - Non HHI - Males  
 0.902575  
 Industrial Category - R to U - Non HHI - Females  
 0.884929

Worked for less than 3

months - Males \
 Worked for 3 months or more but less than 6 mon...  
 0.994988  
 Worked for 3 months or more but less than 6 mon...  
 0.998280  
 Worked for 3 months or more but less than 6 mon...  
 0.989355  
 Worked for less than 3 months - Persons  
 0.996781  
 Worked for less than 3 months - Males  
 1.000000  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.997250  
 Industrial Category - R to U - HHI - Females  
 0.991624  
 Industrial Category - R to U - Non HHI - Persons  
 0.924417  
 Industrial Category - R to U - Non HHI - Males  
 0.930383  
 Industrial Category - R to U - Non HHI - Females  
 0.916108

Worked for less than 3



months - Females \  
 Worked for 3 months or more but less than 6 mon...  
 0.997591  
 Worked for 3 months or more but less than 6 mon...  
 0.993884  
 Worked for 3 months or more but less than 6 mon...  
 0.998962  
 Worked for less than 3 months - Persons  
 0.997719  
 Worked for less than 3 months - Males  
 0.989097  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.981738  
 Industrial Category - R to U - HHI - Females  
 0.969162  
 Industrial Category - R to U - Non HHI - Persons  
 0.865105  
 Industrial Category - R to U - Non HHI - Males  
 0.874656  
 Industrial Category - R to U - Non HHI - Females  
 0.854260

Industrial Category - A -

Cultivators - Persons \  
 Worked for 3 months or more but less than 6 mon...  
 0.974229  
 Worked for 3 months or more but less than 6 mon...  
 0.963718  
 Worked for 3 months or more but less than 6 mon...  
 0.982470  
 Worked for less than 3 months - Persons  
 0.972156  
 Worked for less than 3 months - Males  
 0.952209  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.942469  
 Industrial Category - R to U - HHI - Females  
 0.916917  
 Industrial Category - R to U - Non HHI - Persons  
 0.781018  
 Industrial Category - R to U - Non HHI - Males  
 0.797561  
 Industrial Category - R to U - Non HHI - Females  
 0.765093

Industrial Category - A -

Cultivators - Males \

Worked for 3 months or more but less than 6 mon...	0.974465
Worked for 3 months or more but less than 6 mon...	0.964569
Worked for 3 months or more but less than 6 mon...	0.982091
Worked for less than 3 months - Persons	0.973186
Worked for less than 3 months - Males	0.953968
...	
...	
Industrial Category - R to U - HHI - Males	0.944032
Industrial Category - R to U - HHI - Females	0.918081
Industrial Category - R to U - Non HHI - Persons	0.783215
Industrial Category - R to U - Non HHI - Males	0.799391
Industrial Category - R to U - Non HHI - Females	0.767566

Industrial Category - A -

Cultivators - Females \

Worked for 3 months or more but less than 6 mon...	0.972546
Worked for 3 months or more but less than 6 mon...	0.961274
Worked for 3 months or more but less than 6 mon...	0.981553
Worked for less than 3 months - Persons	0.969472
Worked for less than 3 months - Males	0.948633
...	
...	
Industrial Category - R to U - HHI - Males	0.939155
Industrial Category - R to U - HHI - Females	0.914142
Industrial Category - R to U - Non HHI - Persons	0.777132
Industrial Category - R to U - Non HHI - Males	0.794115

Industrial Category - R to U - Non HHI - Females  
0.760882

Industrial Category - A -

Agricultural labourers - Persons \  
Worked for 3 months or more but less than 6 mon...  
0.976938  
Worked for 3 months or more but less than 6 mon...  
0.966549  
Worked for 3 months or more but less than 6 mon...  
0.985051  
Worked for less than 3 months - Persons  
0.975133  
Worked for less than 3 months - Males  
0.955123

...

...

Industrial Category - R to U - HHI - Males  
0.944333  
Industrial Category - R to U - HHI - Females  
0.920040  
Industrial Category - R to U - Non HHI - Persons  
0.782905  
Industrial Category - R to U - Non HHI - Males  
0.798867  
Industrial Category - R to U - Non HHI - Females  
0.767424

... \  
...

Worked for 3 months or more but less than 6 mon...  
Worked for 3 months or more but less than 6 mon...  
Worked for 3 months or more but less than 6 mon...  
Worked for less than 3 months - Persons  
Worked for less than 3 months - Males

...

...

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

Industrial Category - N to O

- Females \  
Worked for 3 months or more but less than 6 mon...  
0.872107  
Worked for 3 months or more but less than 6 mon...  
0.890596  
Worked for 3 months or more but less than 6 mon...

0.851539  
 Worked for less than 3 months - Persons  
 0.872193  
 Worked for less than 3 months - Males  
 0.901865  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.908918  
 Industrial Category - R to U - HHI - Females  
 0.937359  
 Industrial Category - R to U - Non HHI - Persons  
 0.974178  
 Industrial Category - R to U - Non HHI - Males  
 0.957497  
 Industrial Category - R to U - Non HHI - Females  
 0.983205

Industrial Category - P to Q

- Persons \  
 Worked for 3 months or more but less than 6 mon...  
 0.921899  
 Worked for 3 months or more but less than 6 mon...  
 0.937061  
 Worked for 3 months or more but less than 6 mon...  
 0.904547  
 Worked for less than 3 months - Persons  
 0.922799  
 Worked for less than 3 months - Males  
 0.946852  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.950798  
 Industrial Category - R to U - HHI - Females  
 0.972449  
 Industrial Category - R to U - Non HHI - Persons  
 0.977079  
 Industrial Category - R to U - Non HHI - Males  
 0.969548  
 Industrial Category - R to U - Non HHI - Females  
 0.979011

Industrial Category - P to Q

- Males \  
 Worked for 3 months or more but less than 6 mon...  
 0.934037  
 Worked for 3 months or more but less than 6 mon...

0.947751  
 Worked for 3 months or more but less than 6 mon...  
 0.918106  
 Worked for less than 3 months - Persons  
 0.934927  
 Worked for less than 3 months - Males  
 0.956228  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.960045  
 Industrial Category - R to U - HHI - Females  
 0.978942  
 Industrial Category - R to U - Non HHI - Persons  
 0.973794  
 Industrial Category - R to U - Non HHI - Males  
 0.966616  
 Industrial Category - R to U - Non HHI - Females  
 0.975466

Industrial Category - P to Q

- Females \  
 Worked for 3 months or more but less than 6 mon...  
 0.914128  
 Worked for 3 months or more but less than 6 mon...  
 0.929994  
 Worked for 3 months or more but less than 6 mon...  
 0.896089  
 Worked for less than 3 months - Persons  
 0.915031  
 Worked for less than 3 months - Males  
 0.940432  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.944437  
 Industrial Category - R to U - HHI - Females  
 0.967442  
 Industrial Category - R to U - Non HHI - Persons  
 0.977007  
 Industrial Category - R to U - Non HHI - Males  
 0.969311  
 Industrial Category - R to U - Non HHI - Females  
 0.979068

Industrial Category - R to U

- HHI - Persons \  
 Worked for 3 months or more but less than 6 mon...

0.984285  
 Worked for 3 months or more but less than 6 mon...  
 0.990587  
 Worked for 3 months or more but less than 6 mon...  
 0.975662  
 Worked for less than 3 months - Persons  
 0.985693  
 Worked for less than 3 months - Males  
 0.994335  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.995568  
 Industrial Category - R to U - HHI - Females  
 0.999570  
 Industrial Category - R to U - Non HHI - Persons  
 0.947750  
 Industrial Category - R to U - Non HHI - Males  
 0.948669  
 Industrial Category - R to U - Non HHI - Females  
 0.943257

Industrial Category - R to U

- HHI - Males \  
 Worked for 3 months or more but less than 6 mon...  
 0.990432  
 Worked for 3 months or more but less than 6 mon...  
 0.995125  
 Worked for 3 months or more but less than 6 mon...  
 0.983406  
 Worked for less than 3 months - Persons  
 0.991516  
 Worked for less than 3 months - Males  
 0.997250  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 1.000000  
 Industrial Category - R to U - HHI - Females  
 0.992382  
 Industrial Category - R to U - Non HHI - Persons  
 0.937578  
 Industrial Category - R to U - Non HHI - Males  
 0.943996  
 Industrial Category - R to U - Non HHI - Females  
 0.928868

Industrial Category - R to U

- HHI - Females \  
 Worked for 3 months or more but less than 6 mon...  
 0.980585  
 Worked for 3 months or more but less than 6 mon...  
 0.987377  
 Worked for 3 months or more but less than 6 mon...  
 0.971479  
 Worked for less than 3 months - Persons  
 0.982091  
 Worked for less than 3 months - Males  
 0.991624  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.992382  
 Industrial Category - R to U - HHI - Females  
 1.000000  
 Industrial Category - R to U - Non HHI - Persons  
 0.949204  
 Industrial Category - R to U - Non HHI - Males  
 0.948407  
 Industrial Category - R to U - Non HHI - Females  
 0.946034

Industrial Category - R to U

- Non HHI - Persons \  
 Worked for 3 months or more but less than 6 mon...  
 0.894684  
 Worked for 3 months or more but less than 6 mon...  
 0.913023  
 Worked for 3 months or more but less than 6 mon...  
 0.874213  
 Worked for less than 3 months - Persons  
 0.894641  
 Worked for less than 3 months - Males  
 0.924417  
 ...  
 ...  
 Industrial Category - R to U - HHI - Males  
 0.937578  
 Industrial Category - R to U - HHI - Females  
 0.949204  
 Industrial Category - R to U - Non HHI - Persons  
 1.000000  
 Industrial Category - R to U - Non HHI - Males  
 0.997095  
 Industrial Category - R to U - Non HHI - Females  
 0.998259

Industrial Category - R to U

- Non HHI - Males \  
Worked for 3 months or more but less than 6 mon...  
0.902864  
Worked for 3 months or more but less than 6 mon...  
0.919946  
Worked for 3 months or more but less than 6 mon...  
0.883634  
Worked for less than 3 months - Persons  
0.902575  
Worked for less than 3 months - Males  
0.930383  
...  
...  
Industrial Category - R to U - HHI - Males  
0.943996  
Industrial Category - R to U - HHI - Females  
0.948407  
Industrial Category - R to U - Non HHI - Persons  
0.997095  
Industrial Category - R to U - Non HHI - Males  
1.000000  
Industrial Category - R to U - Non HHI - Females  
0.990867

Industrial Category - R to U

- Non HHI - Females  
Worked for 3 months or more but less than 6 mon...  
0.884780  
Worked for 3 months or more but less than 6 mon...  
0.904020  
Worked for 3 months or more but less than 6 mon...  
0.863430  
Worked for less than 3 months - Persons  
0.884929  
Worked for less than 3 months - Males  
0.916108  
...  
...  
Industrial Category - R to U - HHI - Males  
0.928868  
Industrial Category - R to U - HHI - Females  
0.946034  
Industrial Category - R to U - Non HHI - Persons  
0.998259  
Industrial Category - R to U - Non HHI - Males  
0.990867



Industrial Category - R to U - Non HHI - Females  
1.000000

[63 rows x 63 columns]

<ipython-input-22-becb50483c66>:1: FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
correlation_matrix = df.corr()
```

```
[ ]: #One-Hot Encoding (for nominal variables) sample :
from sklearn.preprocessing import MinMaxScaler

columns_to_normalize = ['Worked for less than 3 months - Persons',
                        'Worked for less than 3 months - Males',
                        'Worked for less than 3 months - Females',

                        ]

scaler = MinMaxScaler()
df[columns_to_normalize] = scaler.fit_transform(df[columns_to_normalize])
```

```
[ ]: # Before dropping rows
print(f"Number of rows before dropping: {df.shape[0]}")

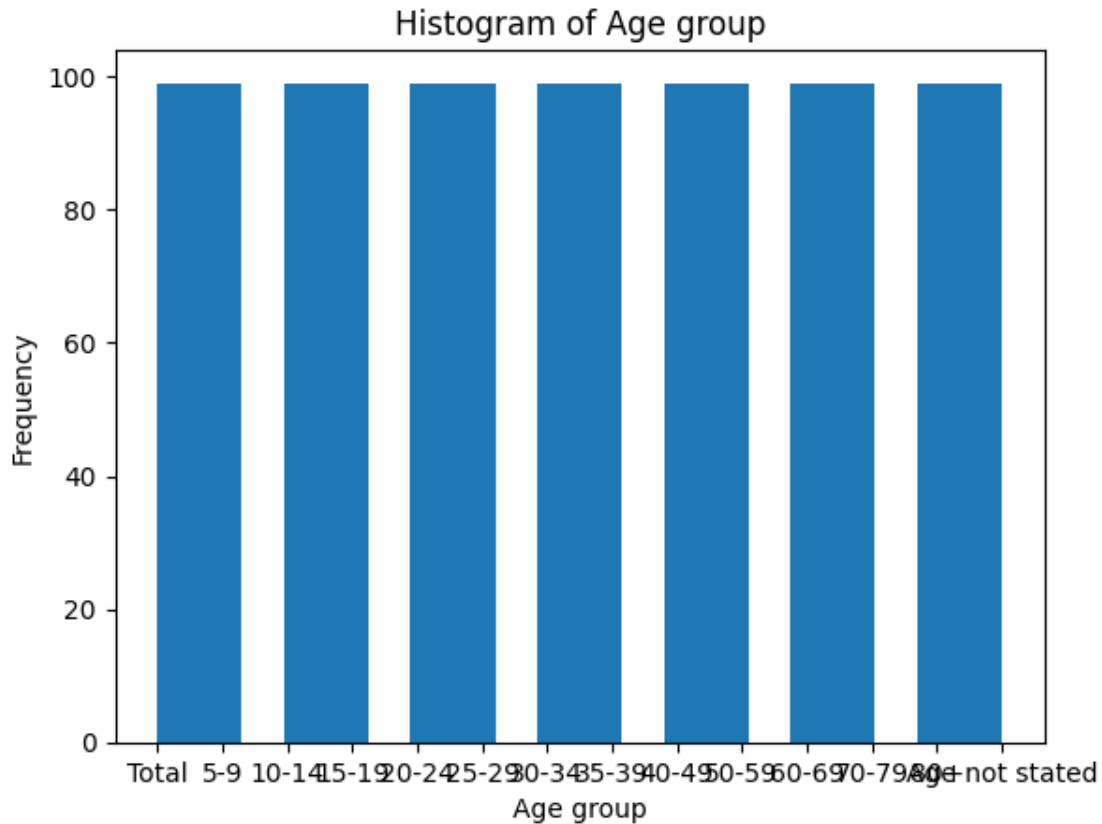
# Drop rows with missing values
df.dropna(inplace=True)

# After dropping rows
print(f"Number of rows after dropping: {df.shape[0]}")
```

Number of rows before dropping: 1287

Number of rows after dropping: 1287

```
[8]: # Example: Creating a histogram for a column 'Age group'
import matplotlib.pyplot as plt
plt.hist(df['Age group'], bins=20)
plt.xlabel('Age group')
plt.ylabel('Frequency')
plt.title('Histogram of Age group')
plt.show()
```



```
[ ]: districts = df['Area Name'].unique()
      print(districts)
```

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

```
[ ]: import matplotlib.pyplot as plt
```

```
# Assuming 'Area Name' contains district names and 'Total Workers' contains the
↳ total number of workers
```



3	B0706	33	000	State - TAMIL NADU	Total
4	B0706	33	000	State - TAMIL NADU	Total

Age group	Worked for 3 months or more but less than 6 months - Persons \
0	Total 4218884
1	5-9 48238
2	10-14 76288
3	15-19 257605
4	20-24 478082

	Worked for 3 months or more but less than 6 months - Males \
0	2136881
1	24511
2	39191
3	141262
4	257149

	Worked for 3 months or more but less than 6 months - Females \
0	2082003
1	23727
2	37097
3	116343
4	220933

	Worked for less than 3 months - Persons ... \
0	723891 ...
1	2051 ...
2	6993 ...
3	41938 ...
4	81036 ...

	Industrial Category - N to O - Females \
0	14495
1	20
2	44
3	768
4	2267

	Industrial Category - P to Q - Persons \
0	58788
1	312
2	506
3	2114
4	11529

	Industrial Category - P to Q - Males \
0	19892

1	169
2	256
3	695
4	2861

	Industrial Category - P to Q - Females \
0	38896
1	143
2	250
3	1419
4	8668

	Industrial Category - R to U - HHI - Persons \
0	89703
1	842
2	1523
3	5349
4	10653

	Industrial Category - R to U - HHI - Males \
0	21366
1	386
2	576
3	2065
4	2478

	Industrial Category - R to U - HHI - Females \
0	68337
1	456
2	947
3	3284
4	8175

	Industrial Category - R to U - Non HHI - Persons \
0	625350
1	40358
2	53900
3	89780
4	109641

	Industrial Category - R to U - Non HHI - Males \
0	274811
1	20664
2	27501
3	46737
4	54832

	Industrial Category - R to U - Non HHI - Females
0	350539
1	19694
2	26399
3	43043
4	54809

[5 rows x 69 columns]

```
[ ]: grouped_data = df.groupby(['Area Name', 'Age group', 'Total/ Rural/Urban'])['Industrial Category - A - Cultivators - Persons'].sum().reset_index()

# Create a separate plot for each district
districts = grouped_data['Area Name'].unique()
grouped_data.head(30)
```

```
[ ]:
   Area Name Age group Total/ Rural/ Urban \
0 District - Ariyalur 10-14 Rural
1 District - Ariyalur 10-14 Total
2 District - Ariyalur 10-14 Urban
3 District - Ariyalur 15-19 Rural
4 District - Ariyalur 15-19 Total
5 District - Ariyalur 15-19 Urban
6 District - Ariyalur 20-24 Rural
7 District - Ariyalur 20-24 Total
8 District - Ariyalur 20-24 Urban
9 District - Ariyalur 25-29 Rural
10 District - Ariyalur 25-29 Total
11 District - Ariyalur 25-29 Urban
12 District - Ariyalur 30-34 Rural
13 District - Ariyalur 30-34 Total
14 District - Ariyalur 30-34 Urban
15 District - Ariyalur 35-39 Rural
16 District - Ariyalur 35-39 Total
17 District - Ariyalur 35-39 Urban
18 District - Ariyalur 40-49 Rural
19 District - Ariyalur 40-49 Total
20 District - Ariyalur 40-49 Urban
21 District - Ariyalur 5-9 Rural
22 District - Ariyalur 5-9 Total
23 District - Ariyalur 5-9 Urban
24 District - Ariyalur 50-59 Rural
25 District - Ariyalur 50-59 Total
26 District - Ariyalur 50-59 Urban
27 District - Ariyalur 60-69 Rural
28 District - Ariyalur 60-69 Total
```

29 District - Ariyalur      60-69      Urban

	Industrial Category - A - Cultivators - Persons
0	68
1	74
2	6
3	411
4	425
5	14
6	926
7	950
8	24
9	1358
10	1402
11	44
12	1308
13	1346
14	38
15	1412
16	1470
17	58
18	2617
19	2756
20	139
21	33
22	34
23	1
24	1867
25	1969
26	102
27	1303
28	1365
29	62

```
[ ]: import matplotlib.pyplot as plt

# 'Area Name' represents the districts, 'Age group' represents the age groups,
# 'Total/ Rural/ Urban' represents rural or urban
# 'Industrial Category - A - Cultivators - Persons' represents the number of
# workers taken as sample'

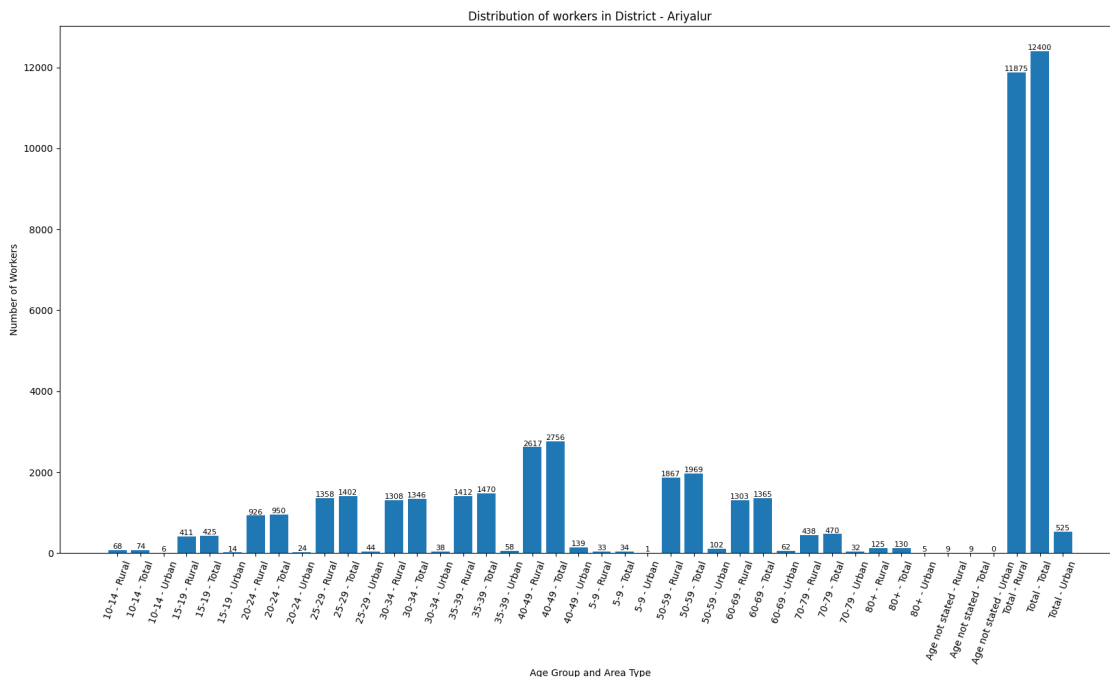
# Grouping by 'Area Name', 'Age group', 'Total/ Rural/ Urban' and summing up
# the number of workers
grouped_data = df.groupby(['Area Name', 'Age group', 'Total/ Rural/
Urban'])['Industrial Category - A - Cultivators - Persons'].sum().
reset_index()
```

```
# Create a separate plot for each district
districts = grouped_data['Area Name'].unique()

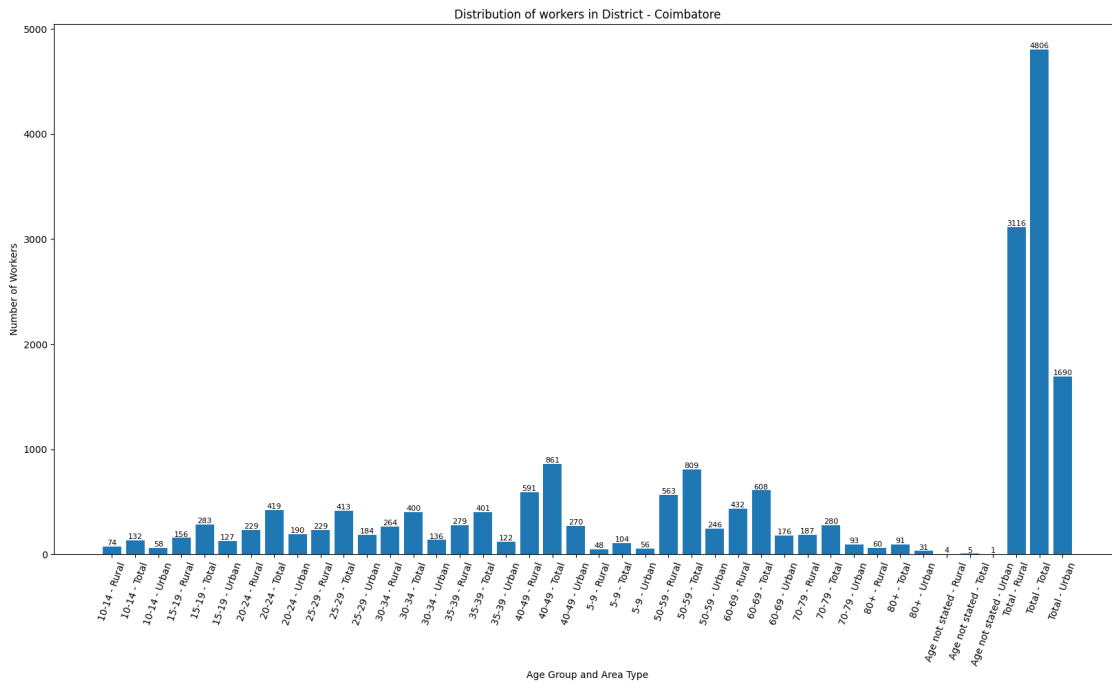
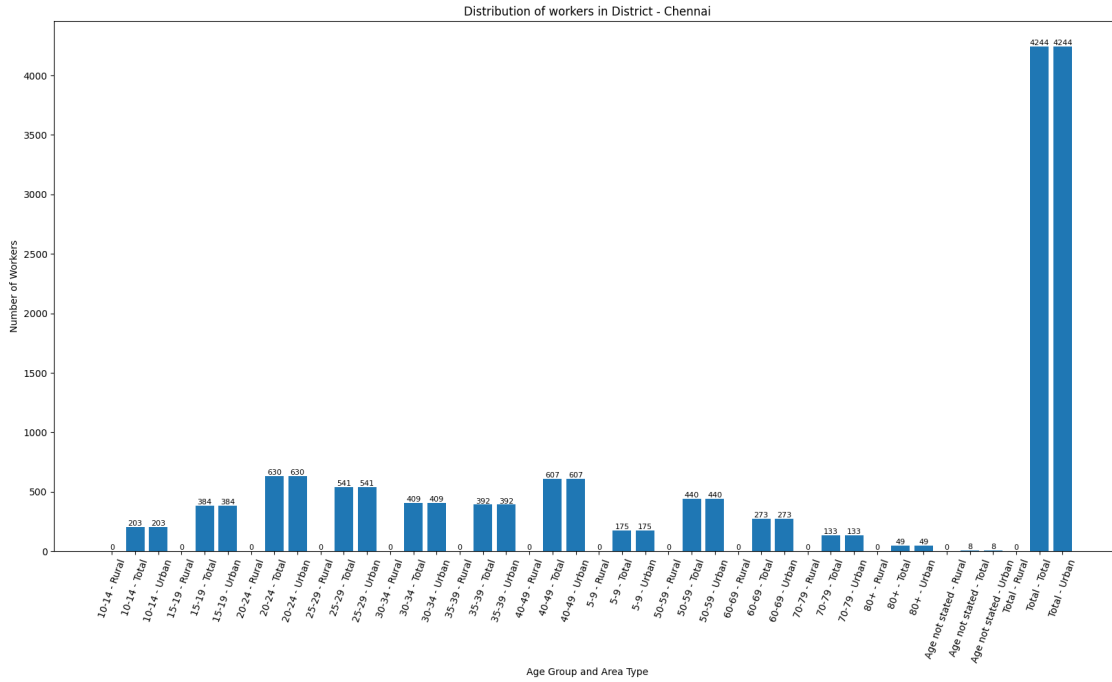
for district in districts:
    district_data = grouped_data[grouped_data['Area Name'] == district]
    plt.figure(figsize=(20, 10))
    bars = plt.bar(district_data['Age group'] + ' - ' + district_data['Total/Urban/ Rural/ Urban'], district_data['Industrial Category - A - Cultivators - Persons'])

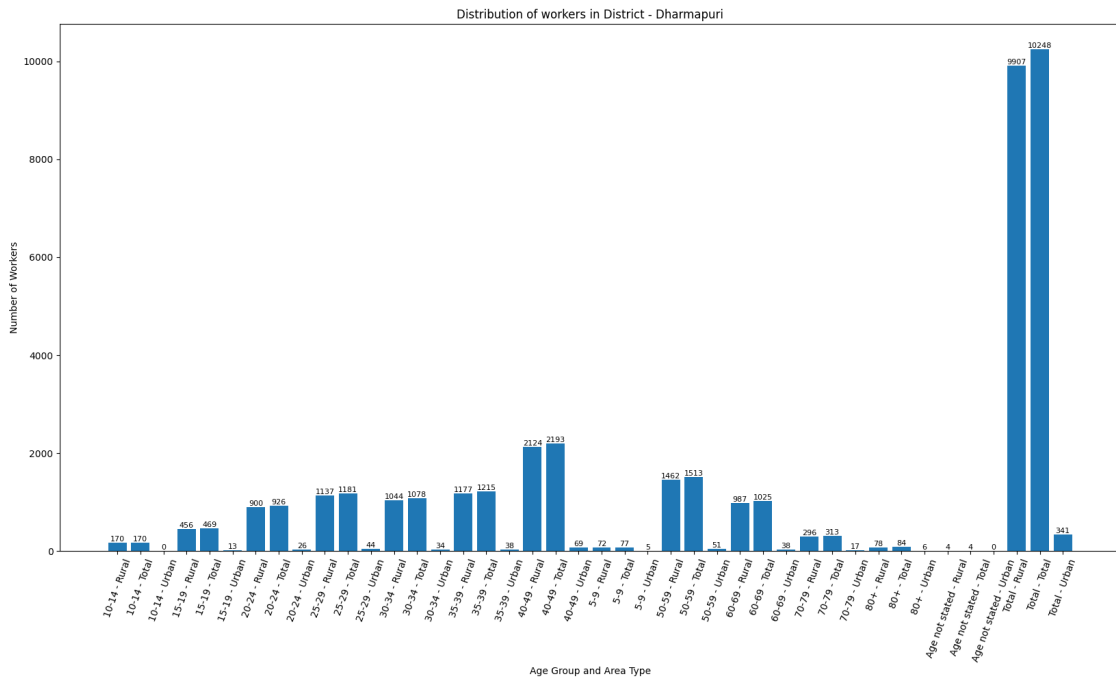
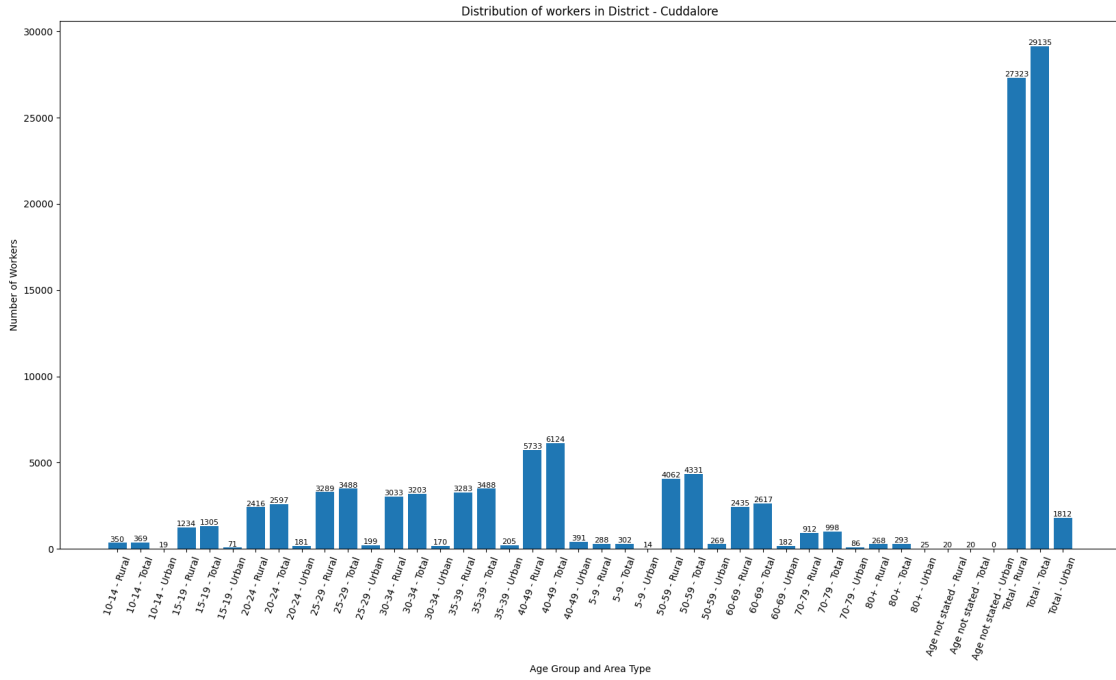
    # Adding numbers on top of the bars
    for bar in bars:
        yval = bar.get_height()
        plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval), va='bottom', ha='center', fontsize=8, color='black')

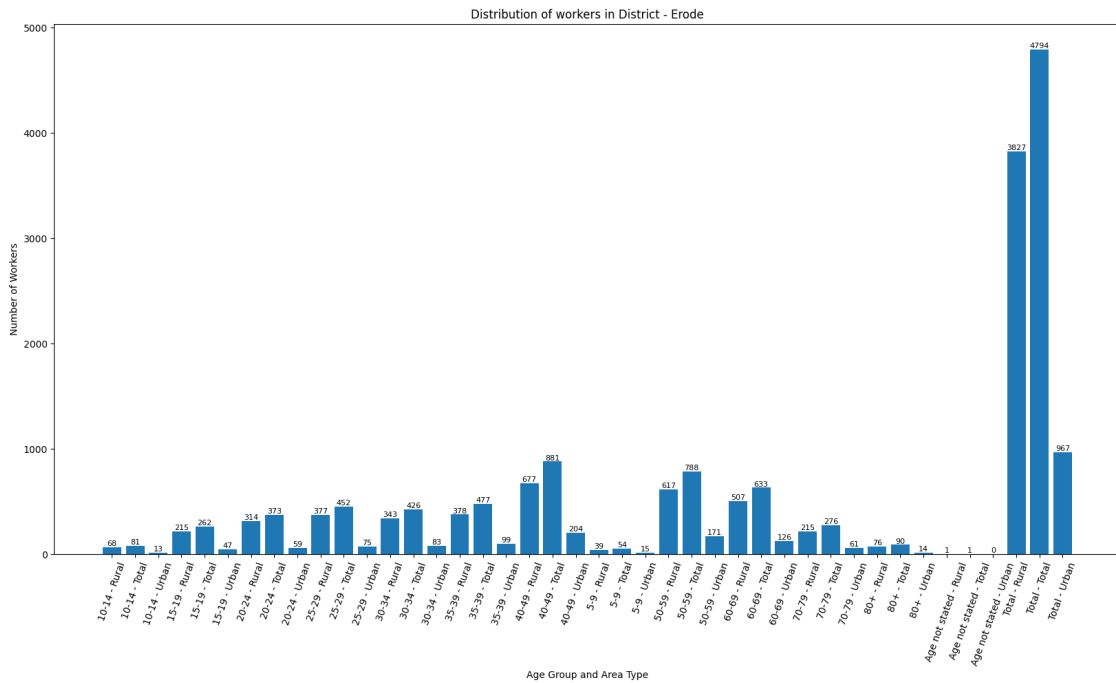
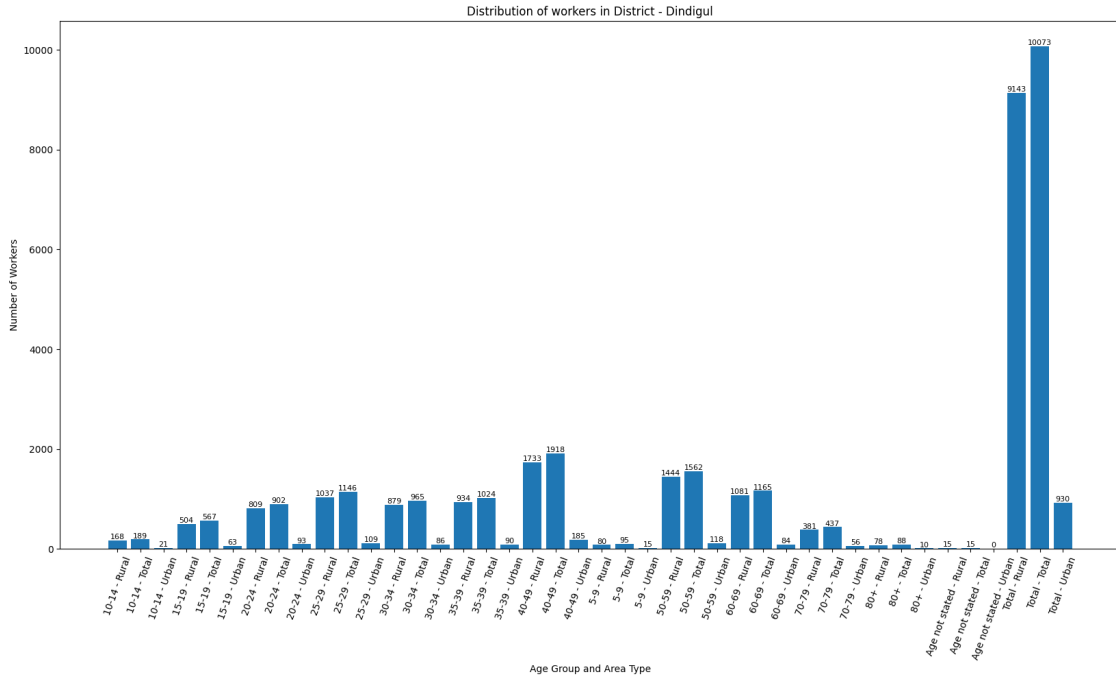
    plt.title(f'Distribution of workers in {district}')
    plt.xlabel('Age Group and Area Type')
    plt.ylabel('Number of Workers')
    plt.xticks(rotation=70)
    plt.show()
```

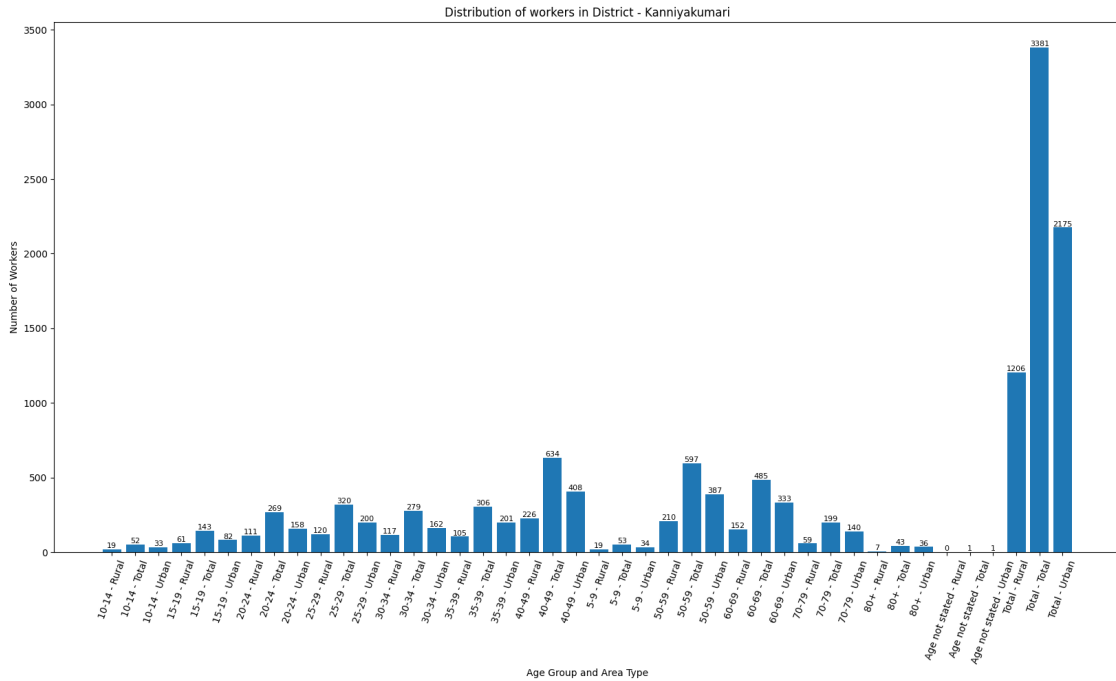
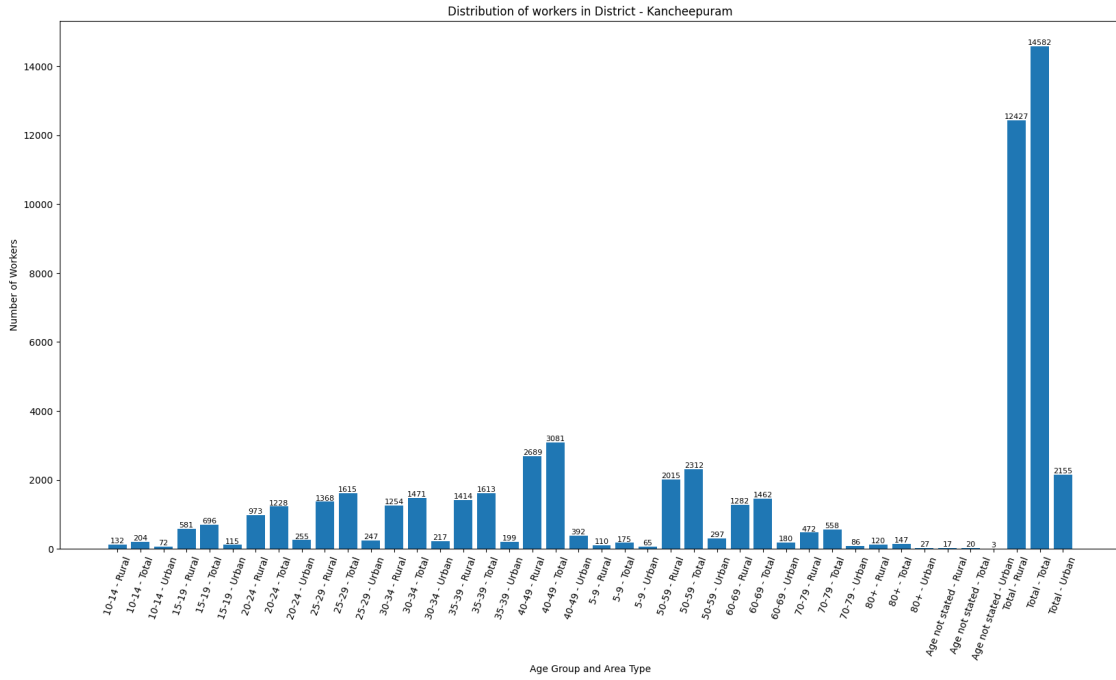


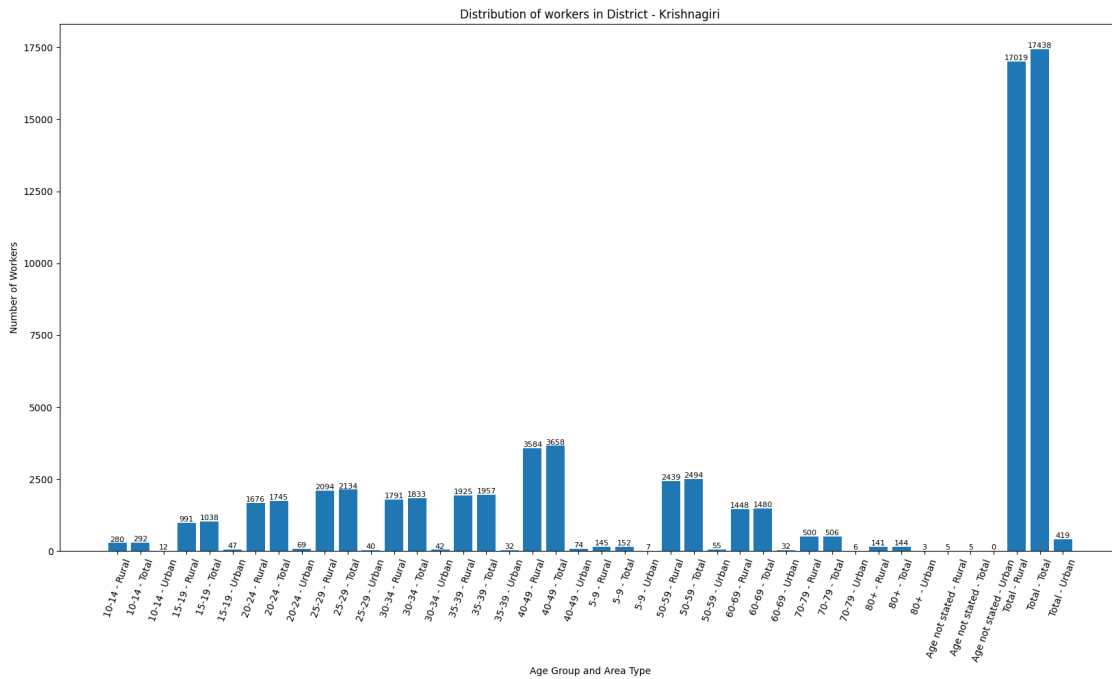
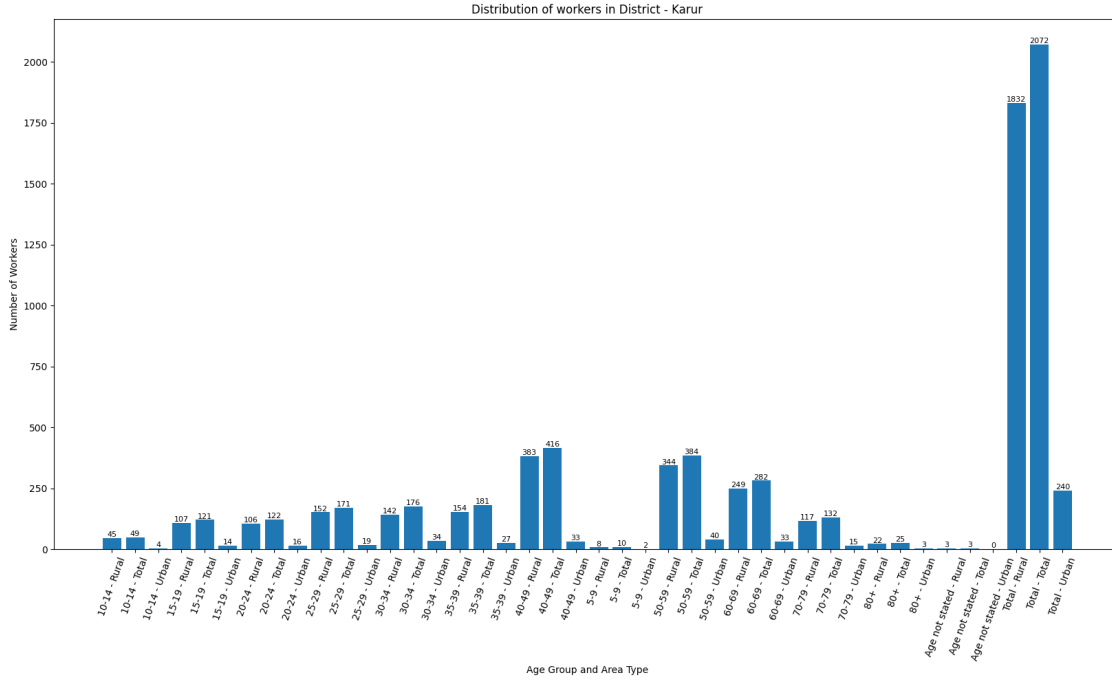


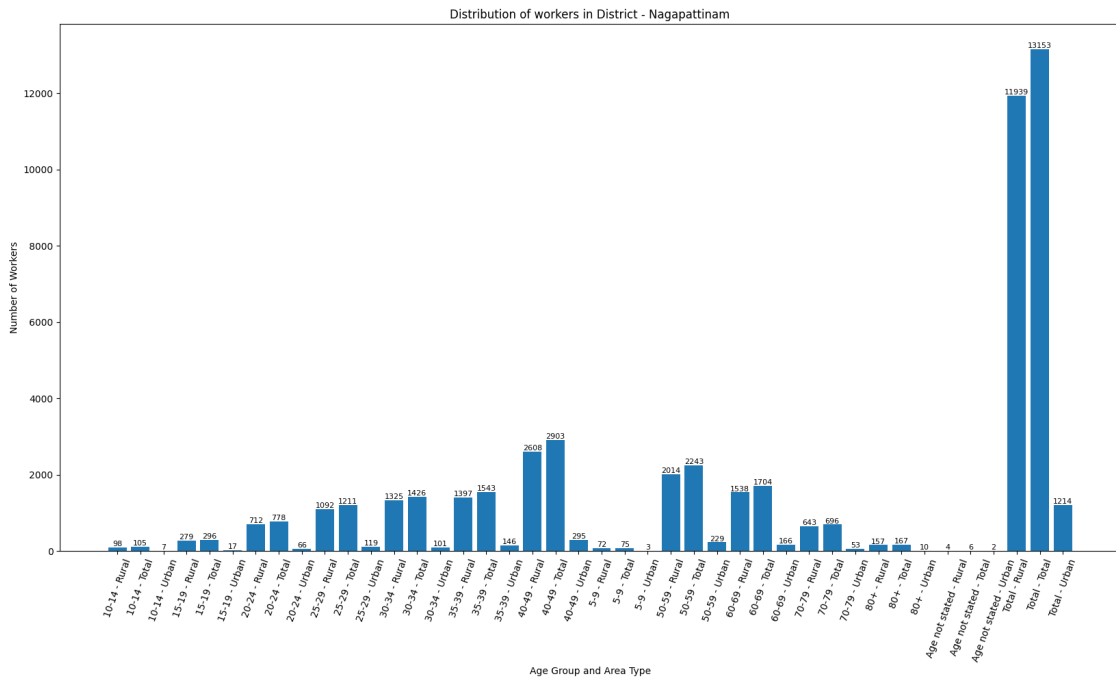
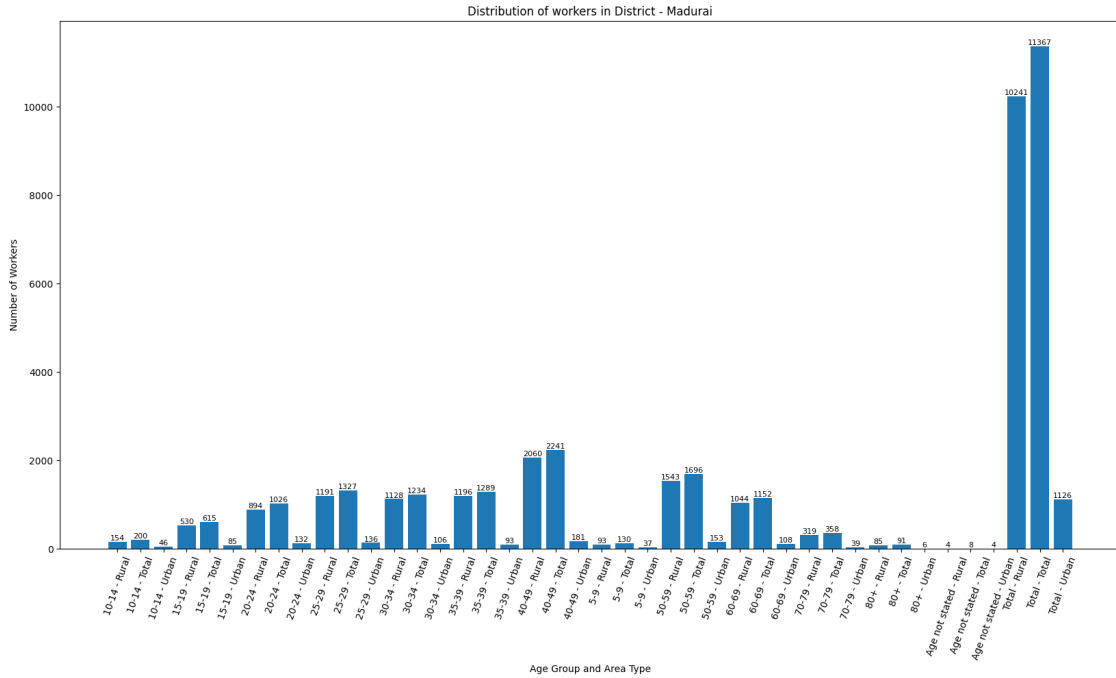


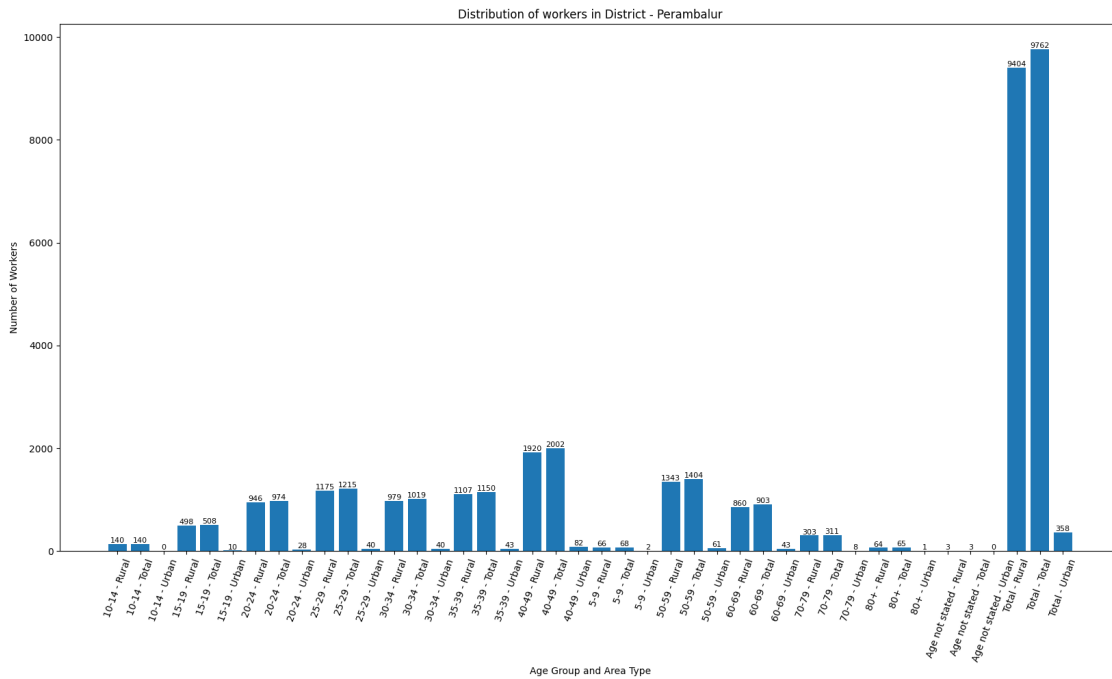
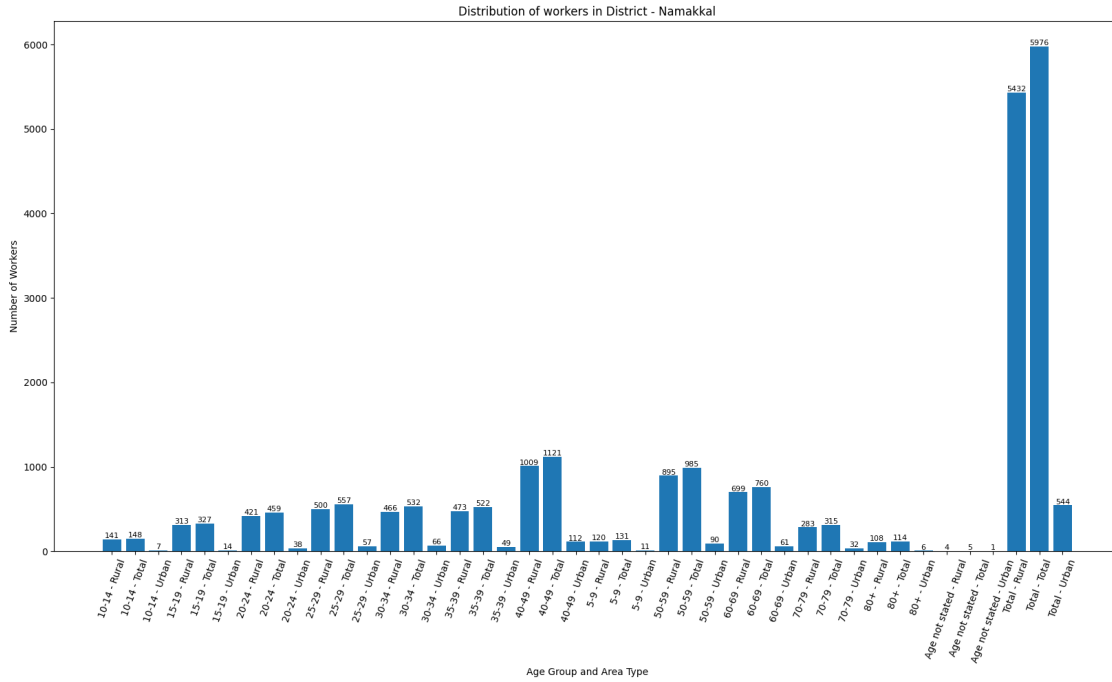


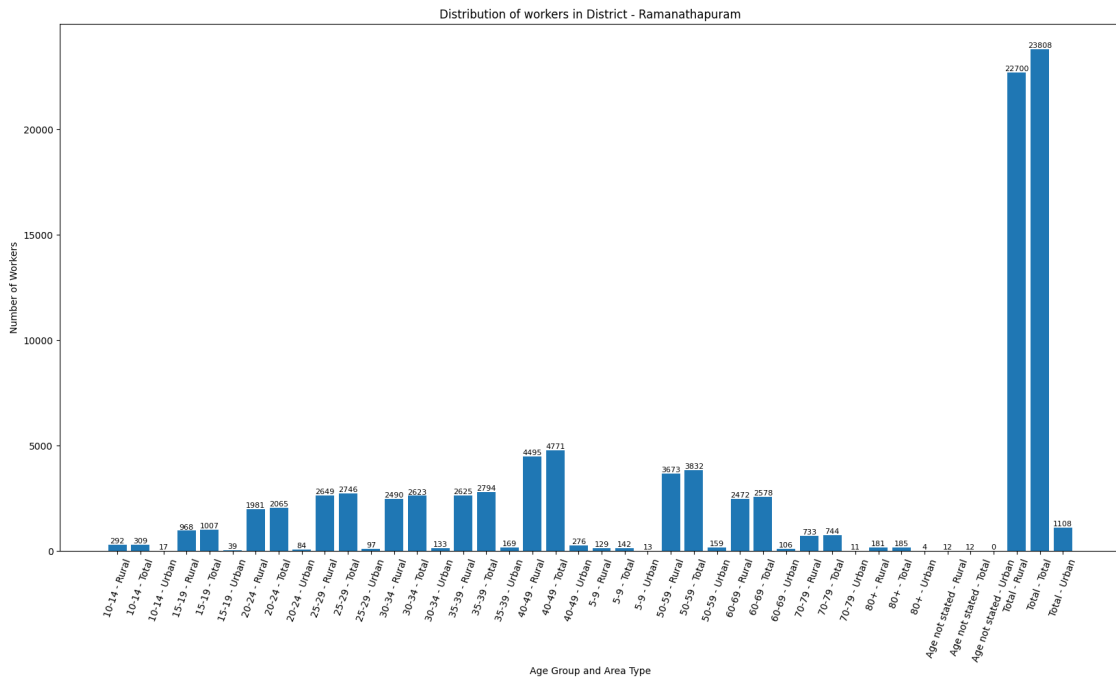
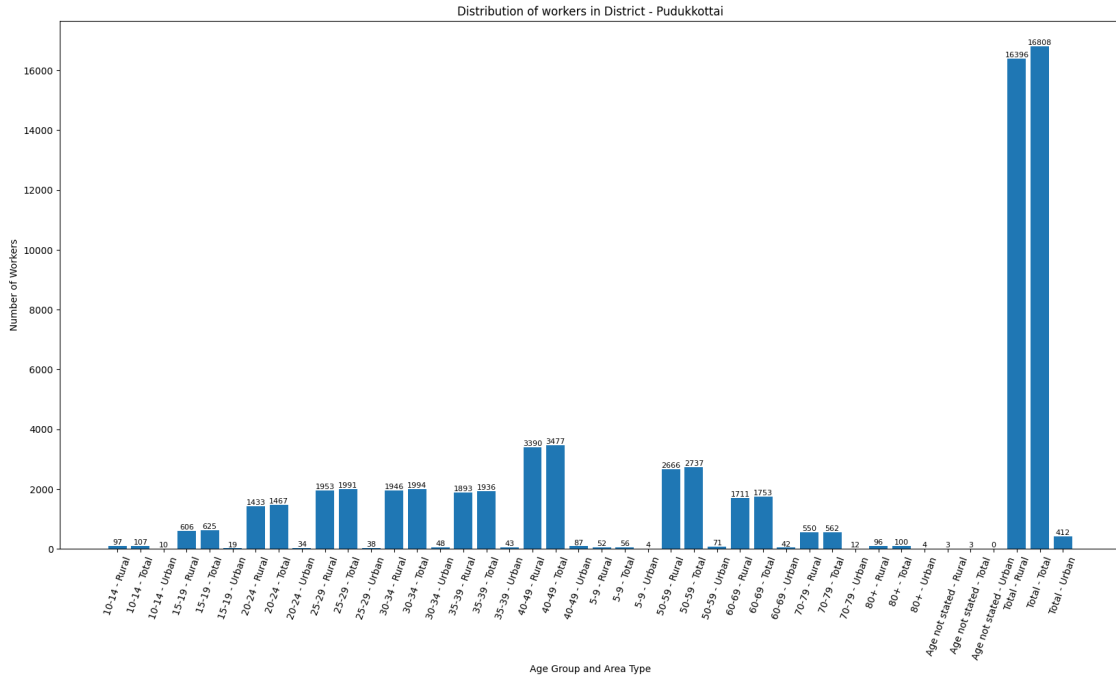




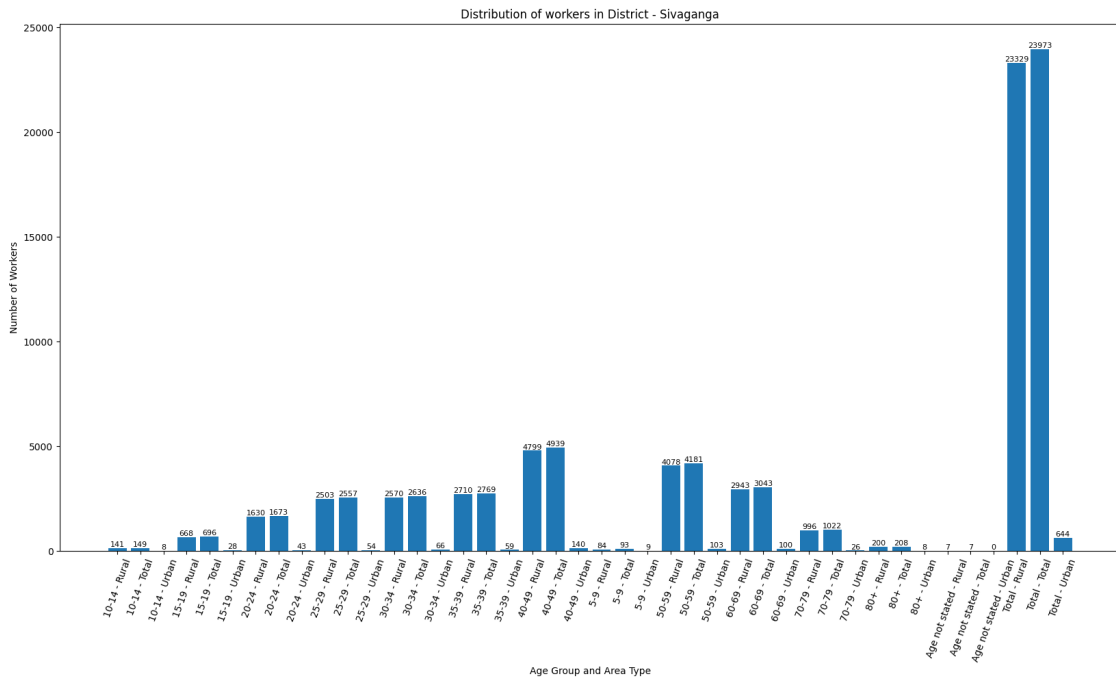
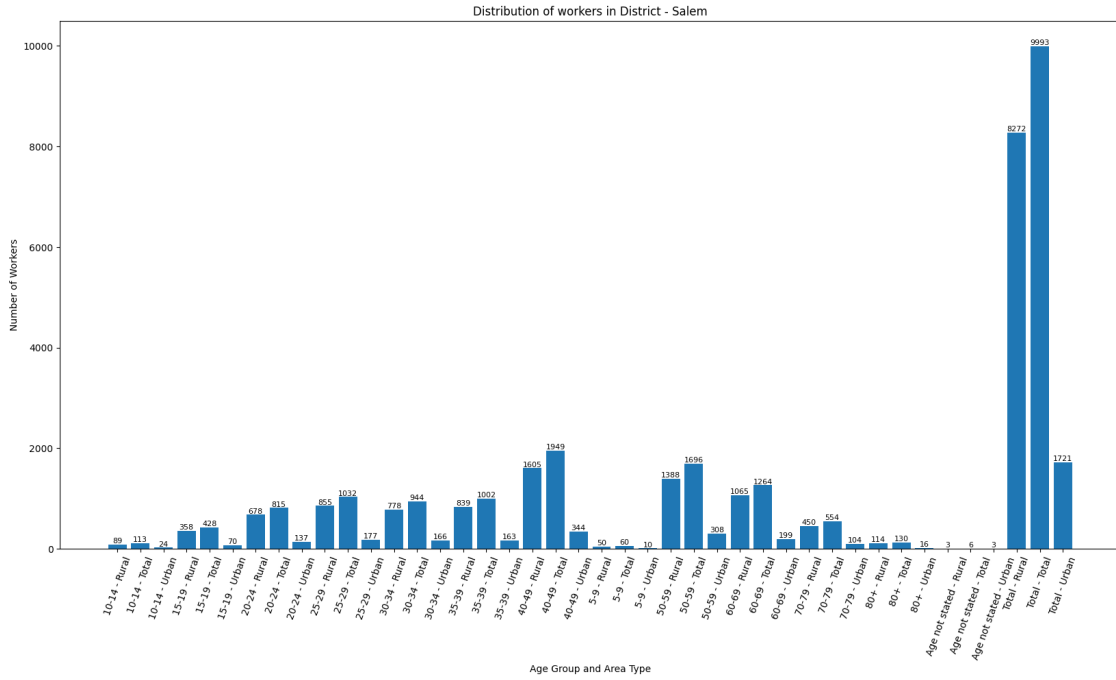




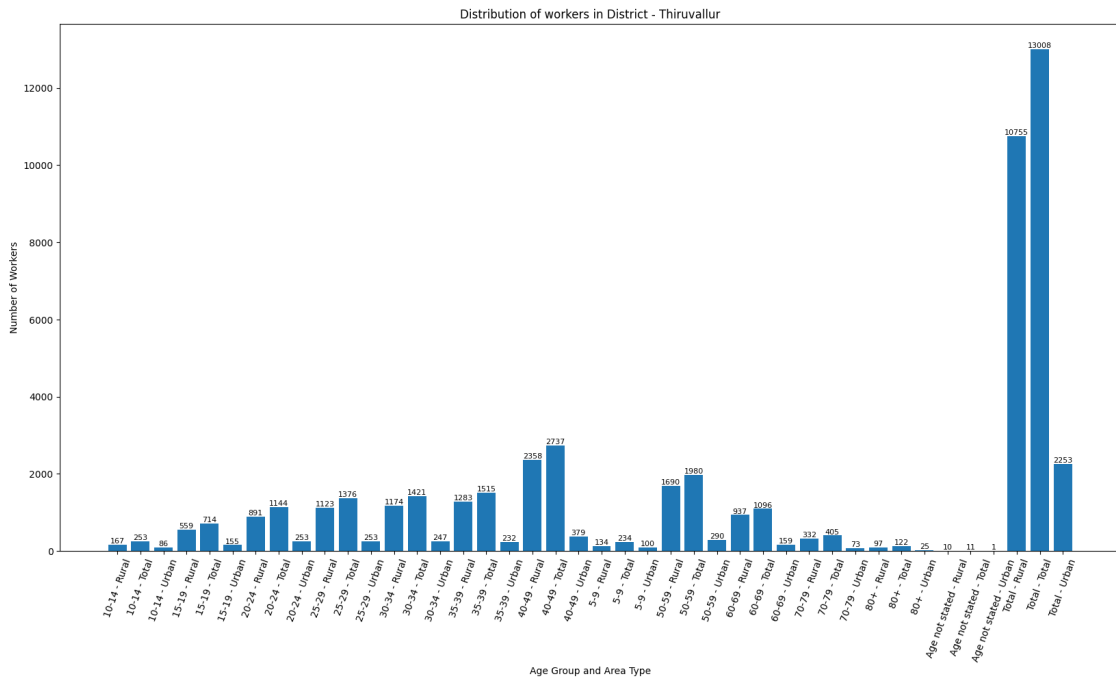
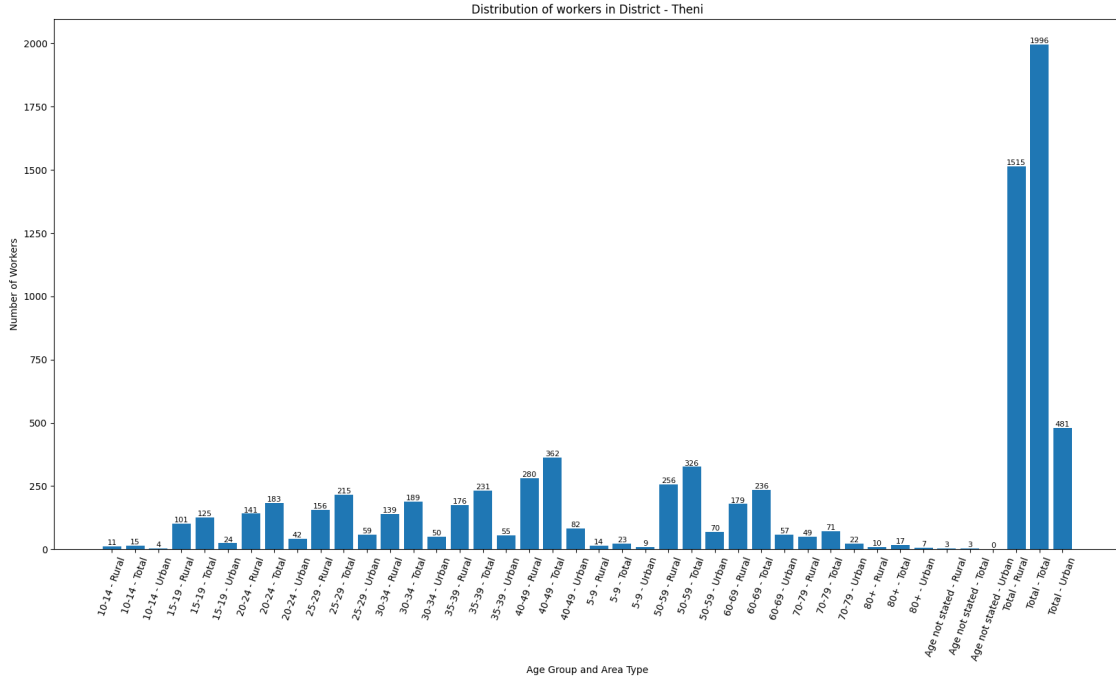


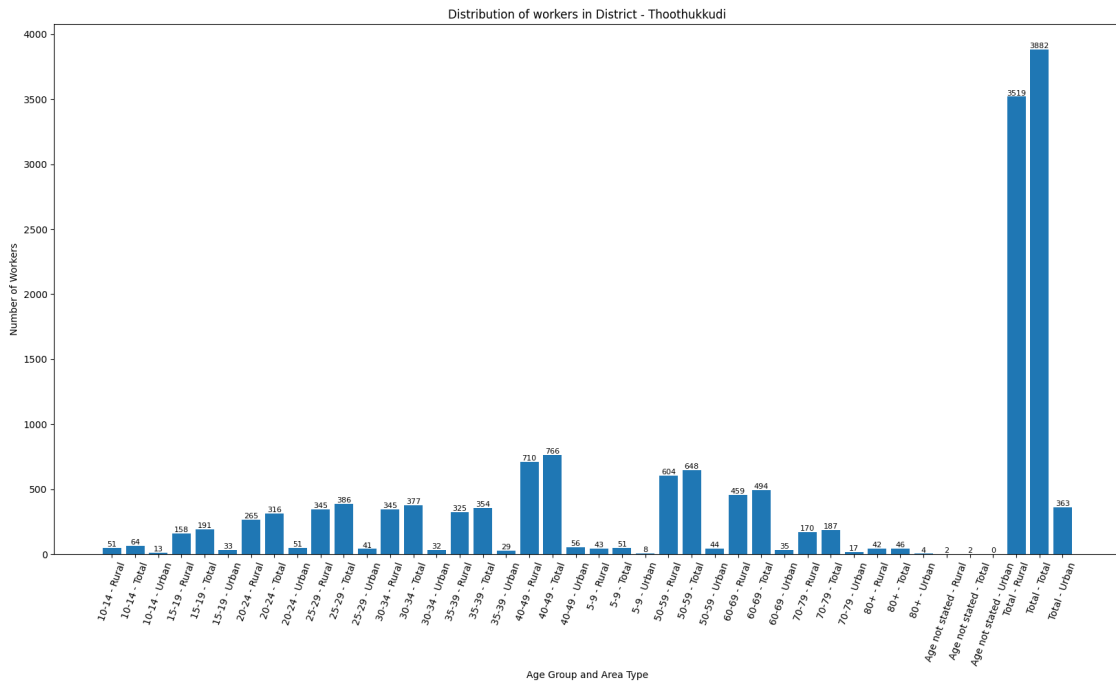
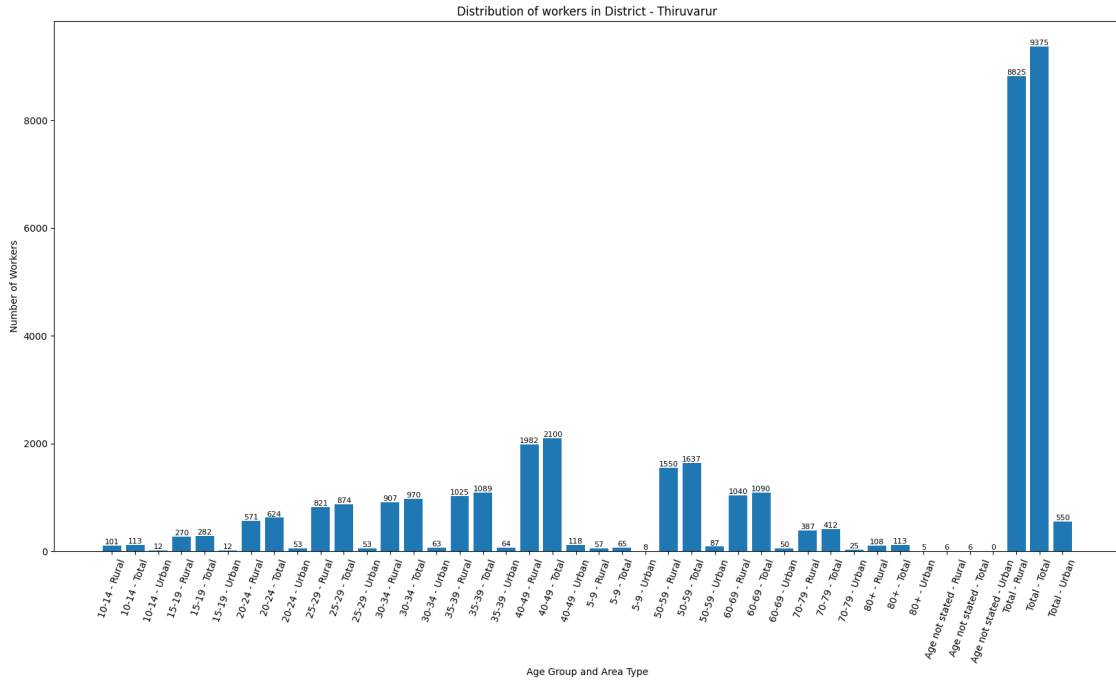


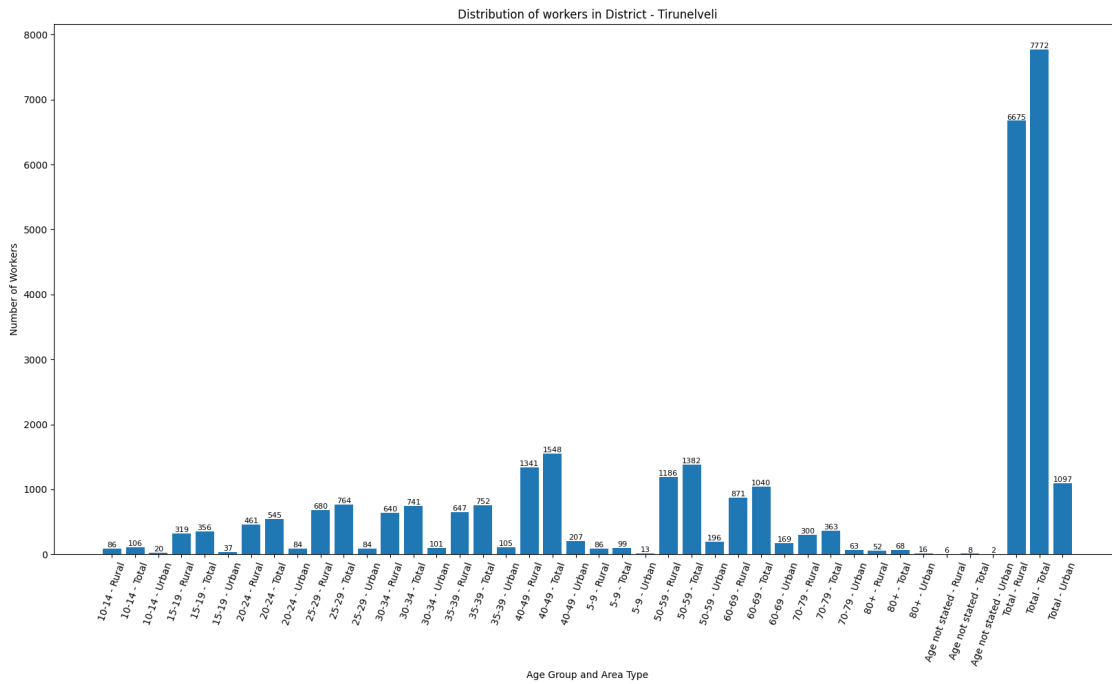
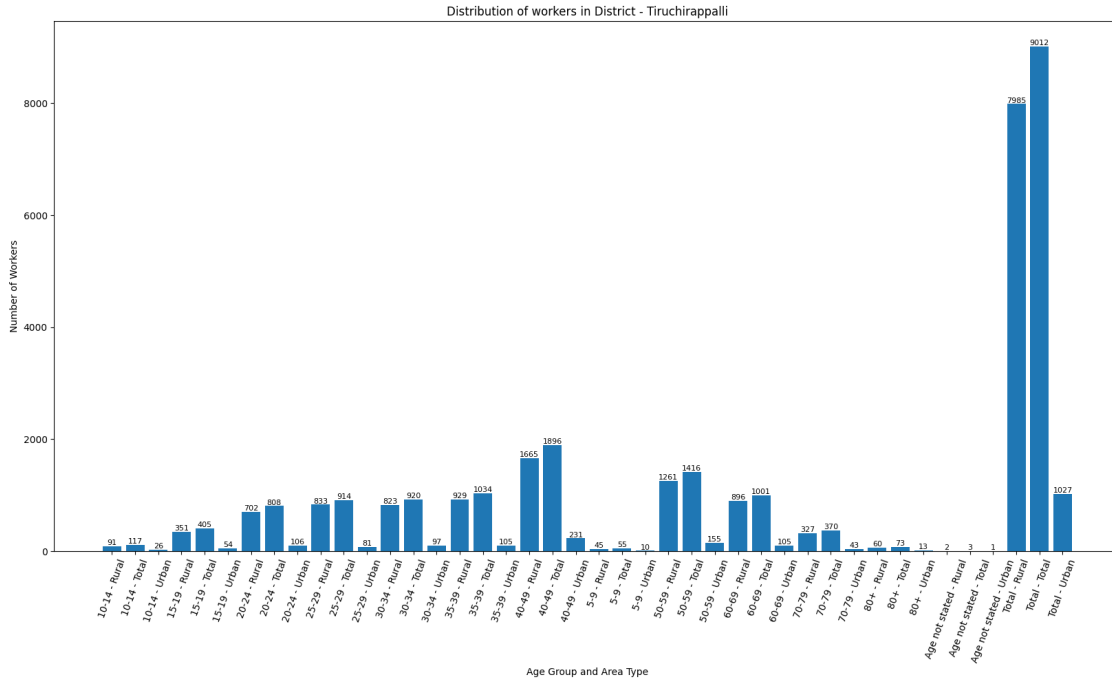


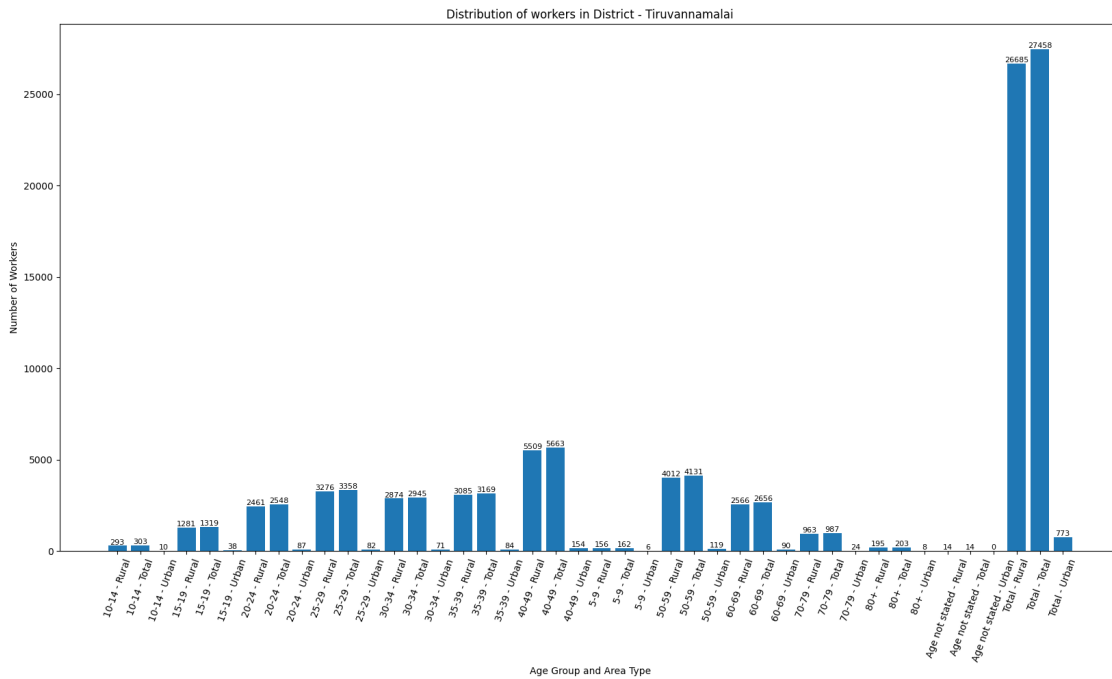
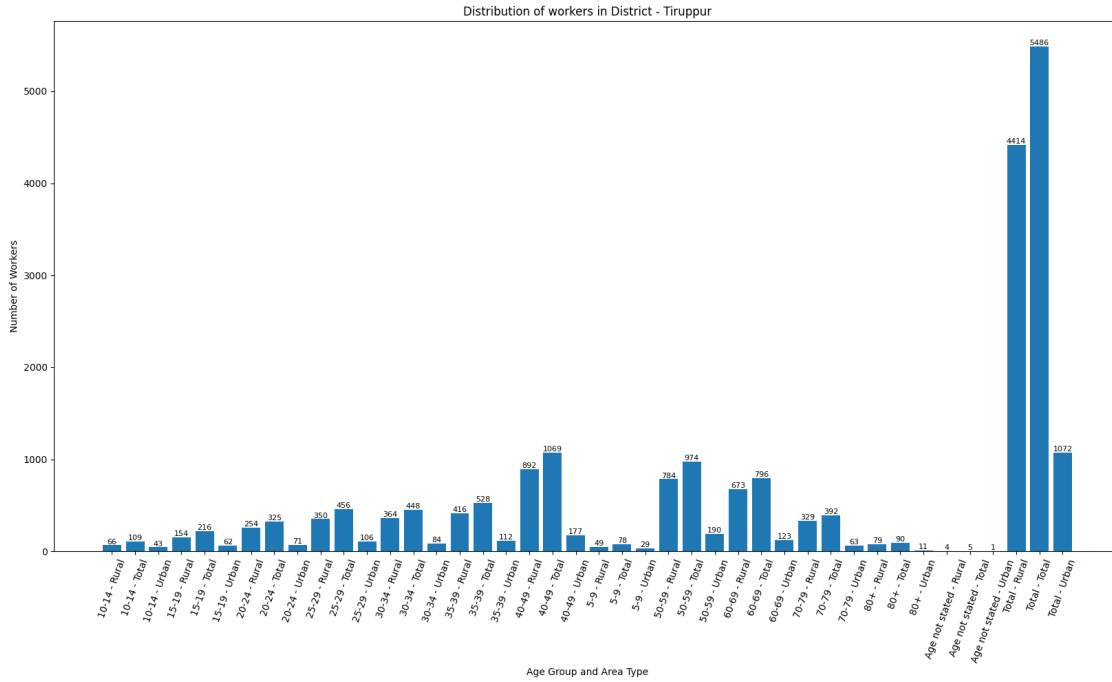


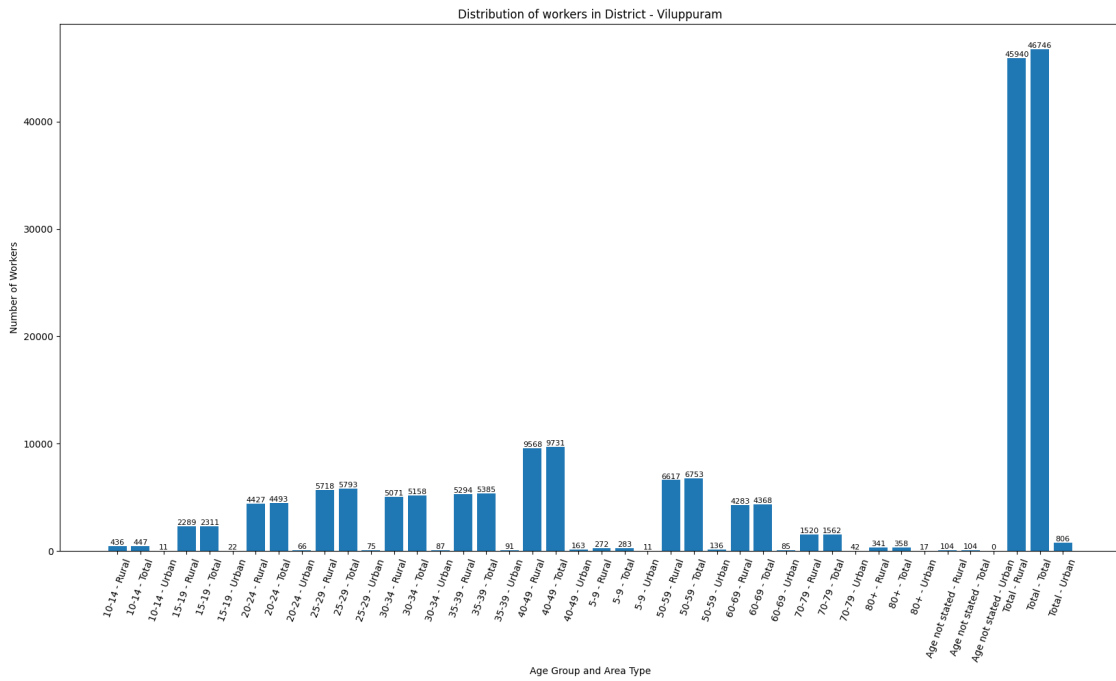
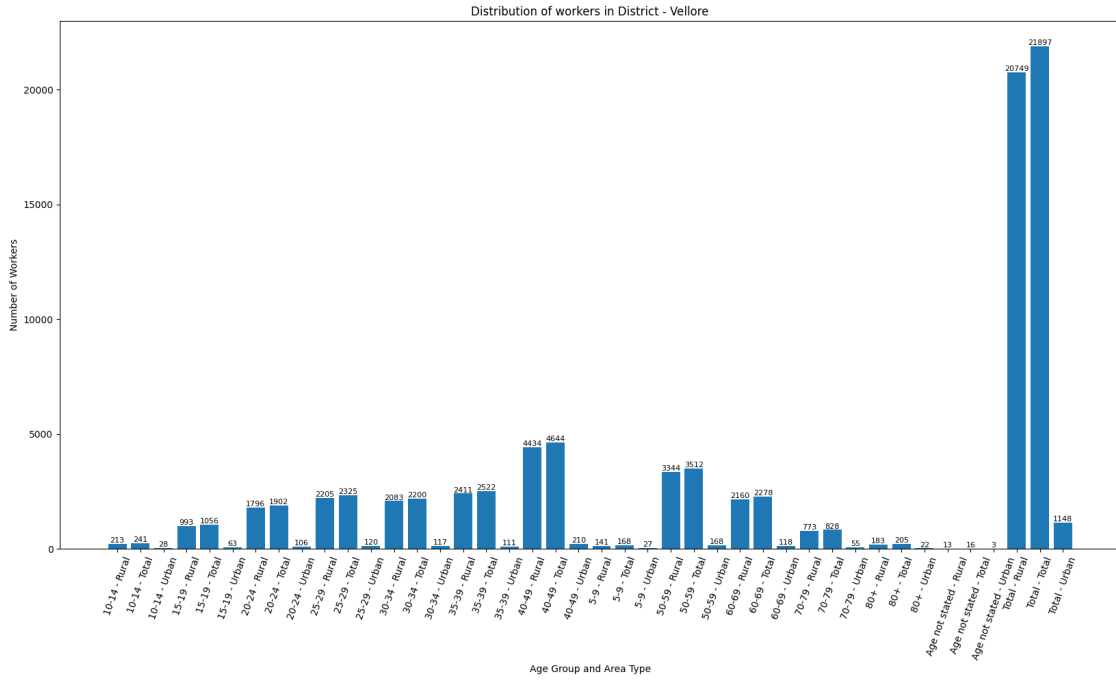


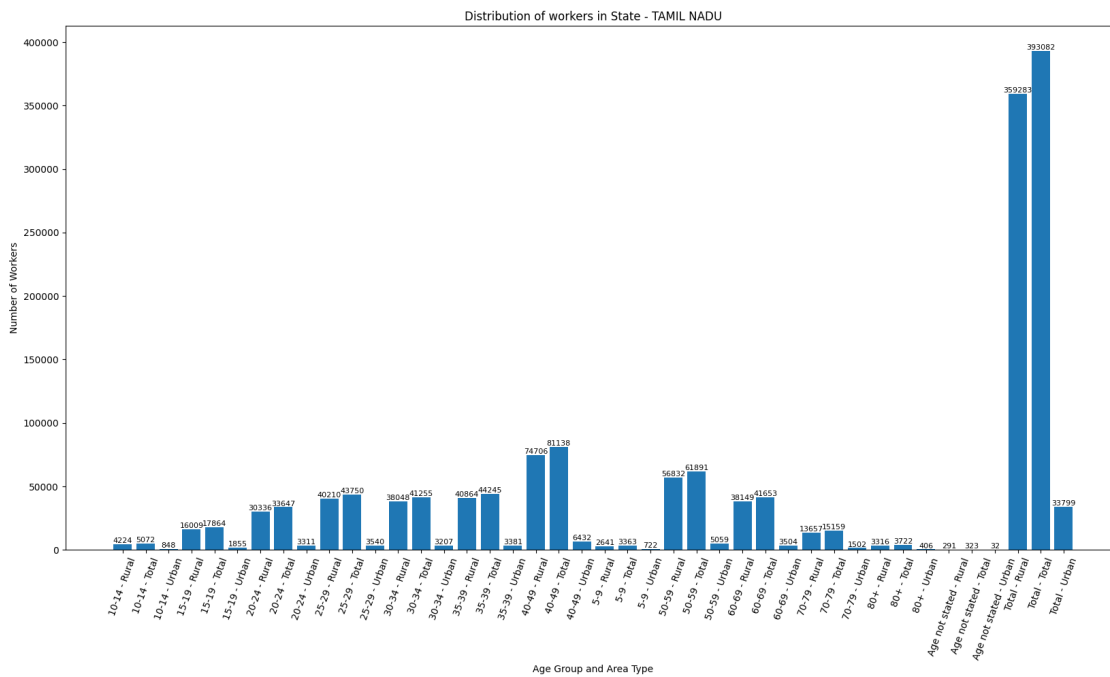
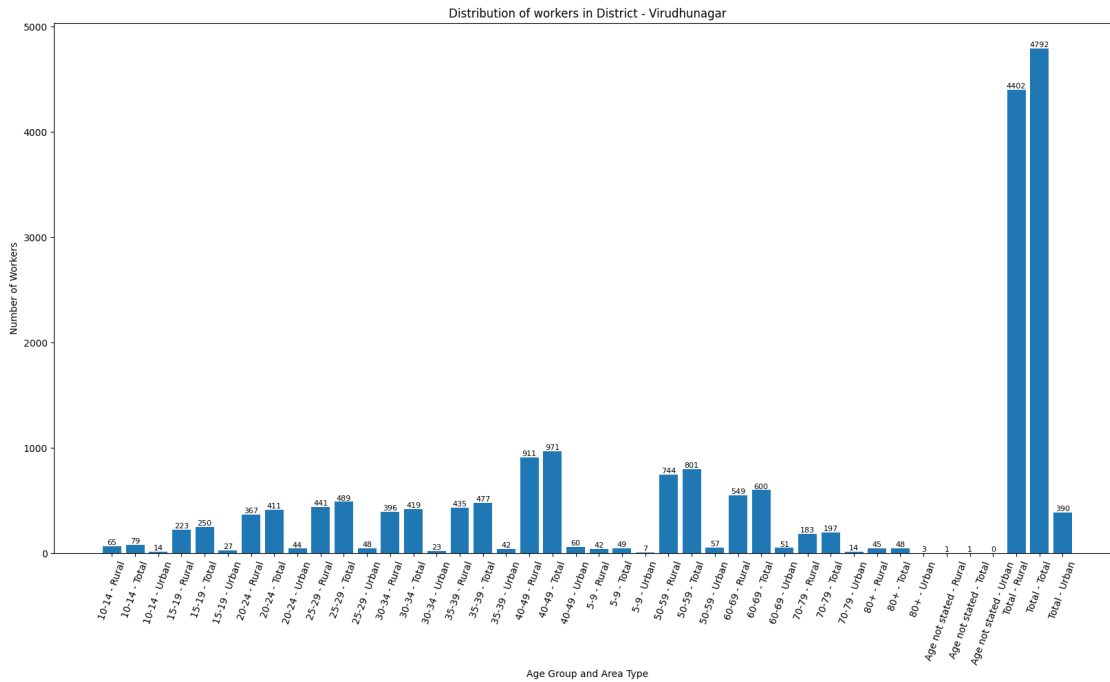












```
[ ]: import matplotlib.pyplot as plt
import pandas as pd
```



```

# Assuming 'df' is your DataFrame

# Filter data for Tamil Nadu, Total Persons
tn_data = df[(df['Area Name'] == 'State - TAMIL NADU') & (df['Total/ Rural/ Urban'] == 'Total')]

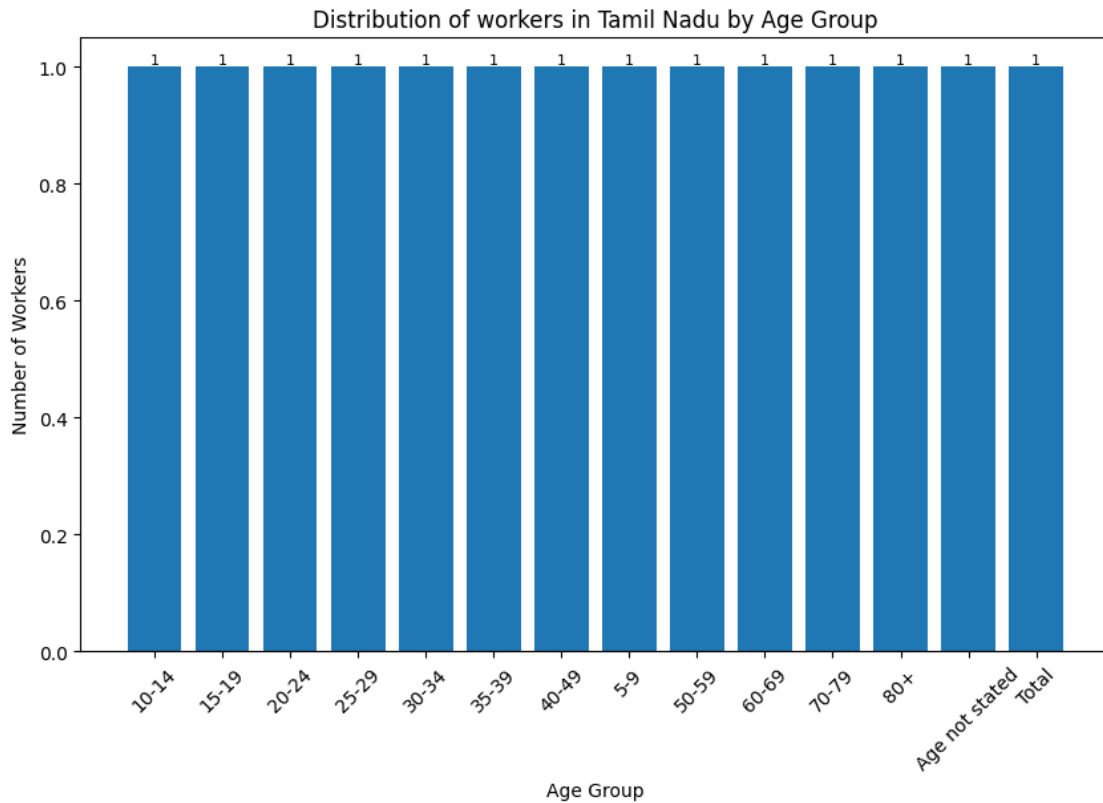
# Grouping by 'Age group' and summing up the number of workers
grouped_data = tn_data.groupby(['Age group'])['Total/ Rural/ Urban'].count().reset_index()

# Create a bar chart
plt.figure(figsize=(10, 6))
bars = plt.bar(grouped_data['Age group'], grouped_data['Total/ Rural/ Urban'])

# Adding numbers on top of the bars
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval), va='bottom', ha='center', fontsize=8, color='black')

plt.title(f'Distribution of workers in Tamil Nadu by Age Group')
plt.xlabel('Age Group')
plt.ylabel('Number of Workers')
plt.xticks(rotation=45)
plt.show()

```



```
[15]: # Entire State

import matplotlib.pyplot as plt

# Assuming 'df' is your DataFrame
# 'Area Name' represents the districts, 'Age group' represents the age groups,
# 'Total/ Rural/ Urban' represents rural or urban
# 'Industrial Category - A - Cultivators - Persons' represents the number of
# workers

# Filter data for State - Tamil Nadu
state_data = df[df['Area Name'] == 'State - TAMIL NADU']

# Grouping by 'Age group', 'Total/ Rural/ Urban' and summing up the number of
# workers
grouped_data = state_data.groupby(['Age group', 'Total/ Rural/
# Urban'])['Industrial Category - A - Cultivators - Persons'].sum().
# reset_index()

# Create the bar chart
plt.figure(figsize=(20, 10))
```

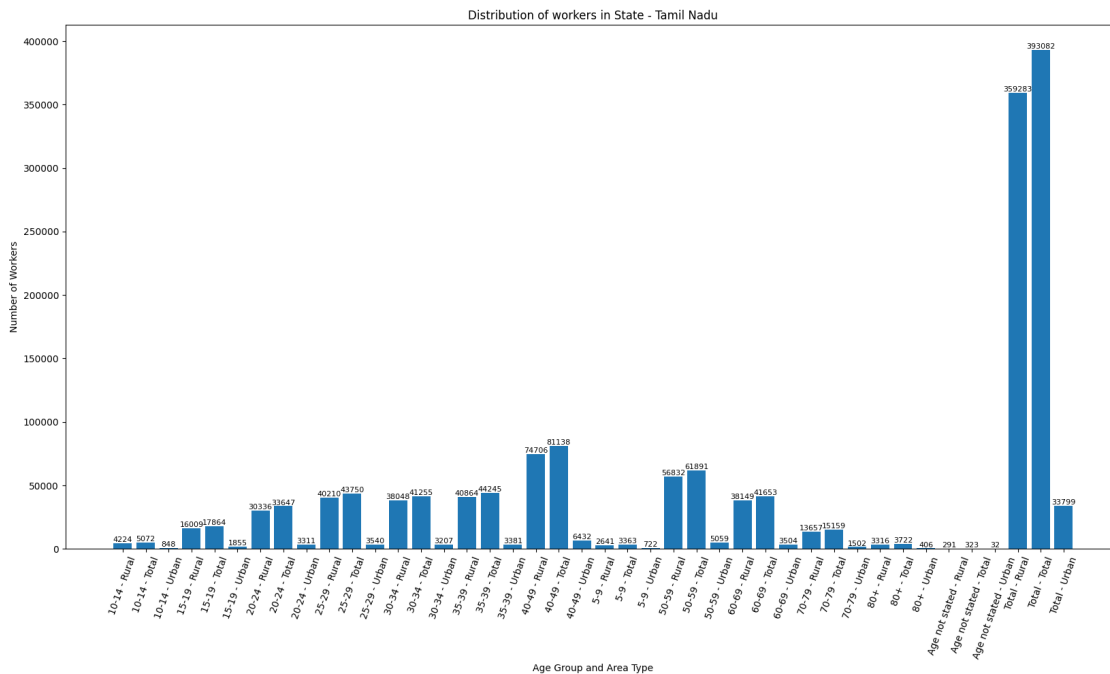
```

bars = plt.bar(grouped_data['Age group'] + ' - ' + grouped_data['Total/ Rural/Urban'], grouped_data['Industrial Category - A - Cultivators - Persons'])

# Adding numbers on top of the bars
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval), va='bottom', ha='center', fontsize=8, color='black')

plt.title('Distribution of workers in State - Tamil Nadu')
plt.xlabel('Age Group and Area Type')
plt.ylabel('Number of Workers')
plt.xticks(rotation=70)
plt.show()

```



```

[10]: import pandas as pd
import matplotlib.pyplot as plt

# Filter for Tamil Nadu state
df = df[df['Area Name'] == 'State - TAMIL NADU']

# Select columns related to Industrial Category A
cols = [col for col in df.columns if 'Industrial Category - A' in col]

```

```

df = df[['Area Name', 'Age group', 'Total/ Rural/ Urban'] + cols]

# Melt the dataframe to bring it in a tidy format
df_melted = df.melt(id_vars=['Area Name', 'Age group', 'Total/ Rural/ Urban'],
                    var_name='Category', value_name='Count')

# Plot
plt.figure(figsize=(10,6))
for category in df_melted['Category'].unique():
    data = df_melted[df_melted['Category'] == category]
    plt.bar(data['Age group'], data['Count'], label=category)

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Industrial Category A across different age groups in Tamil Nadu')
plt.legend()
plt.show()

```

