

Assessment of Marginal Workers in Tamil Nadu – A Socioeconomic Analysis

Team Member

Name : HARIHARAN R

Register Number : **211521243059**

Applied Data Science Phase-4 document

Team Members :

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

Phase 4: Development Part 2

Problem Statement:

In this part you will continue building your project.

- Perform the demographic analysis**
- Calculate the distribution of marginal workers based on age, industrial category, and sex using data aggregation and manipulation.**

Create visualizations.

- Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).**

ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU

Applied Data Science **Phase-4** document

Team Members : • A R HRUDAYABHIRAM 211521243019 • LAKSHMI KANTH R 211521243095 • ASWIN S 211521243026 • DINESH S 211521243049 • HARIHARAN R 211521243059

Google collab :

STEP 1 : Demographic Analysis

<https://colab.research.google.com/drive/1yI7FN8GHvhWCTr5EzqdhMWYtGkJwq4x>

In [3]: #Load the dataset

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

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	In C	
0	B0706	'33	'000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	...	14495	58788	19892	38896
1	B0706	'33	'000	State - TAMIL NADU	Total	'5-9	48238	24511	23727	2051	...	20	312	169	143
2	B0706	'33	'000	State - TAMIL NADU	Total	'10-14	76288	39191	37097	6993	...	44	506	256	250
3	B0706	'33	'000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	...	768	2114	695	1419
4	B0706	'33	'000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	...	2267	11529	2861	8668
...
1381	B0706	'33	'633	District - Tiruppur	Urban	50-59	4965	2800	2165	901	...	25	111	51	60
1382	B0706	'33	'633	District - Tiruppur	Urban	60-69	2827	1590	1237	578	...	7	21	6	15
1383	B0706	'33	'633	District - Tiruppur	Urban	70-79	920	581	339	204	...	2	6	6	0
1384	B0706	'33	'633	District - Tiruppur	Urban	80+	191	104	87	47	...	0	2	0	2
1385	B0706	'33	'633	District - Tiruppur	Urban	Age not stated	31	23	8	9	...	0	0	0	0

1386 rows × 69 columns

In [5]: # Clean the dataset by removing '' from the columns of Age group, state code ,Dsitriict 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

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	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	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	In C	
						- Persons	- Males	- Females	- Persons	
0	B0706	33	000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	...	14495	58788	19892	38896
1	B0706	33	000	State - TAMIL NADU	Total	5-9	48238	24511	23727	2051	...	20	312	169	143
2	B0706	33	000	State - TAMIL NADU	Total	10-14	76288	39191	37097	6993	...	44	506	256	250
3	B0706	33	000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	...	768	2114	695	1419
4	B0706	33	000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	...	2267	11529	2861	8668
...
1381	B0706	33	633	District - Tiruppur	Urban	50-59	4965	2800	2165	901	...	25	111	51	60
1382	B0706	33	633	District - Tiruppur	Urban	60-69	2827	1590	1237	578	...	7	21	6	15
1383	B0706	33	633	District - Tiruppur	Urban	70-79	920	581	339	204	...	2	6	6	0
1384	B0706	33	633	District - Tiruppur	Urban	80+	191	104	87	47	...	0	2	0	2
1385	B0706	33	633	District - Tiruppur	Urban	Age not stated	31	23	8	9	...	0	0	0	0

1386 rows × 69 columns

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

Out[]:

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	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	Industri Categ - R to U HHI Person	
						- Persons	- Males	- Females	- Persons	
0	B0706	33	000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	...	14495	58788	19892	38896
1	B0706	33	000	State - TAMIL NADU	Total	5-9	48238	24511	23727	2051	...	20	312	169	143
2	B0706	33	000	State - TAMIL NADU	Total	10-14	76288	39191	37097	6993	...	44	506	256	250
3	B0706	33	000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	...	768	2114	695	1419
4	B0706	33	000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	...	2267	11529	2861	8668

5 rows × 69 columns

In []: `import pandas as pd`

```
# Assuming 'df' is your cleaned DataFrame
df.to_csv('cleaned_dataset.csv', index=False)
```

In []: `# Check for missing values`
`missing_values = df.isnull().sum()`
`print(missing_values)`

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 []: `df.columns`

```
Out[ ]: 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')
```

In []: `df.isnull()`

Out[]:

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	Indus- cate - R to t Per	
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
...
1381	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
1382	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
1383	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
1384	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False
1385	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False

1386 rows × 69 columns

Unique Values in Categorical Columns

```
In [ ]: # 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

```
In [ ]: # finding correlation
correlation_matrix = df.corr()
print(correlation_matrix)
```

	Worked for 3 months or more but less than 6 mon...	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-10-edeeab342ee8>:2: 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()

Demographic analysis

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

```
print(df.describe())
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1386 entries, 0 to 1385
Data columns (total 69 columns):
 #   Column          Dtype   Nulls on Axis 1
 0   Table Code      object   1386
 1   State Code     object   1386
 2   District Code  object   1386
 3   Area Name     object   1386
 4   Total/ Rural/ Urban  object   1386
 5   Age group     object   1386
 6   Worked for 3 months or more but less than 6 months - Persons  int64   1386
 7   Worked for 3 months or more but less than 6 months - Males    int64   1386
 8   Worked for 3 months or more but less than 6 months - Females  int64   1386
 9   Worked for less than 3 months - Persons   int64   1386
10   Worked for less than 3 months - Males    int64   1386
11   Worked for less than 3 months - Females  int64   1386
12   Industrial Category - A - Cultivators - Persons  int64   1386
13   Industrial Category - A - Cultivators - Males   int64   1386
14   Industrial Category - A - Cultivators - Females  int64   1386
15   Industrial Category - A - Agricultural labourers - Persons  int64   1386
16   Industrial Category - A - Agricultural labourers - Males   int64   1386
17   Industrial Category - A - Agricultural labourers - Females  int64   1386
18   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons  int64   1386
19   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males   int64   1386
20   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females  int64   1386
21   Industrial Category - B - Persons   int64   1386
22   Industrial Category - B - Males    int64   1386
23   Industrial Category - B - Females  int64   1386
24   Industrial Category - C - HHI - Persons  int64   1386
25   Industrial Category - C - HHI - Males   int64   1386
26   Industrial Category - C - HHI - Females  int64   1386
27   Industrial Category - C - Non HHI - Persons  int64   1386
28   Industrial Category - C - Non HHI - Males   int64   1386
29   Industrial Category - C - Non HHI - Females  int64   1386
30   Industrial Category - D & E - Persons   int64   1386
31   Industrial Category - D & E - Males    int64   1386
32   Industrial Category - D & E - Females  int64   1386
33   Industrial Category - F - Persons   int64   1386
34   Industrial Category - F - Males    int64   1386
35   Industrial Category - F - Females  int64   1386
36   Industrial Category - G - HHI - Persons  int64   1386
37   Industrial Category - G - HHI - Males   int64   1386
38   Industrial Category - G - HHI - Females  int64   1386
39   Industrial Category - G - Non HHI - Persons  int64   1386
40   Industrial Category - G - Non HHI - Males   int64   1386
41   Industrial Category - G - Non HHI - Females  int64   1386
42   Industrial Category - H - Persons   int64   1386
43   Industrial Category - H - Males    int64   1386

```

```

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
    Worked for 3 months or more but less than 6 months - Persons \
count                  1.386000e+03
mean                  2.435142e+04
std                   1.530754e+05
min                   0.000000e+00
25%                   8.372500e+02
50%                   3.985000e+03
75%                   1.251725e+04
max                   4.218884e+06

    Worked for 3 months or more but less than 6 months - Males \
count                  1.386000e+03
mean                  1.233409e+04
std                   7.669251e+04
min                   0.000000e+00
25%                   4.637500e+02
50%                   2.047500e+03
75%                   6.273000e+03
max                   2.136881e+06

    Worked for 3 months or more but less than 6 months - Females \
count                  1.386000e+03
mean                  1.201733e+04
std                   7.656262e+04
min                   0.000000e+00
25%                   3.792500e+02
50%                   1.812000e+03
75%                   6.255500e+03
max                   2.082003e+06

    Worked for less than 3 months - Persons \
count                  1386.000000
mean                  4178.303030
std                   26234.919027
min                   0.000000
25%                   123.000000
50%                   650.500000
75%                   2071.750000
max                   723891.000000

    Worked for less than 3 months - Males \
count                  1386.000000

```

mean	1946.712843
std	12024.992364
min	0.000000
25%	71.000000
50%	315.500000
75%	955.250000
max	337268.000000

Worked for less than 3 months - Females \

count	1386.000000
mean	2231.590188
std	14281.201871
min	0.000000
25%	51.250000
50%	337.500000
75%	1091.250000
max	386623.000000

Industrial Category - A - Cultivators - Persons \

count	1386.000000
mean	2268.871573
std	15445.653849
min	0.000000
25%	56.000000
50%	215.500000
75%	1068.000000
max	393082.000000

Industrial Category - A - Cultivators - Males \

count	1386.000000
mean	1271.653680
std	8627.700716
min	0.000000
25%	32.000000
50%	129.500000
75%	584.250000
max	220314.000000

Industrial Category - A - Cultivators - Females \

count	1386.000000
mean	997.217893
std	6827.658762
min	0.000000
25%	21.000000
50%	88.500000
75%	491.500000
max	172768.000000

Industrial Category - A - Agricultural labourers - Persons ... \

count	1.386000e+03
mean	1.369377e+04
std	9.330282e+04
min	0.000000e+00
25%	2.135000e+02
50%	1.282000e+03
75%	6.713750e+03
max	2.372446e+06

Industrial Category - N to O - Females \

count	1386.000000
mean	83.665224
std	543.170274
min	0.000000
25%	0.000000
50%	6.000000
75%	29.000000
max	14495.000000

Industrial Category - P to Q - Persons \

count	1386.000000
mean	339.324675
std	2114.109688
min	0.000000
25%	5.000000
50%	44.000000
75%	166.000000
max	58788.000000

Industrial Category - P to Q - Males \

count	1386.000000
mean	114.816739
std	710.763665
min	0.000000
25%	0.000000
50%	18.000000
75%	60.000000
max	19892.000000

Industrial Category - P to Q - Females \

count	1386.000000
mean	224.507937
std	1405.839106
min	0.000000
25%	0.000000
50%	26.000000

```
75%          112.000000  
max         38896.000000  
  
Industrial Category - R to U - HHI - Persons \  
count        1386.000000  
mean         517.766234  
std          3177.844267  
min          0.000000  
25%          18.000000  
50%          80.000000  
75%          243.500000  
max         89703.000000  
  
Industrial Category - R to U - HHI - Males \  
count        1386.000000  
mean         123.324675  
std          756.489766  
min          0.000000  
25%          6.000000  
50%          21.000000  
75%          62.000000  
max         21366.000000  
  
Industrial Category - R to U - HHI - Females \  
count        1386.000000  
mean         394.441558  
std          2425.750623  
min          0.000000  
25%          11.000000  
50%          57.500000  
75%          187.000000  
max         68337.000000  
  
Industrial Category - R to U - Non HHI - Persons \  
count        1386.000000  
mean         3609.523810  
std          22377.933258  
min          0.000000  
25%          208.500000  
50%          593.000000  
75%          1548.000000  
max         625350.000000  
  
Industrial Category - R to U - Non HHI - Males \  
count        1386.000000  
mean         1586.210678  
std          9787.231574  
min          0.000000  
25%          96.000000  
50%          260.000000  
75%          695.750000  
max         274811.000000  
  
Industrial Category - R to U - Non HHI - Females  
count        1386.000000  
mean         2023.313131  
std          12641.139629  
min          0.000000  
25%          104.000000  
50%          317.500000  
75%          820.000000  
max         350539.000000
```

[8 rows x 63 columns]

In []: df.columns

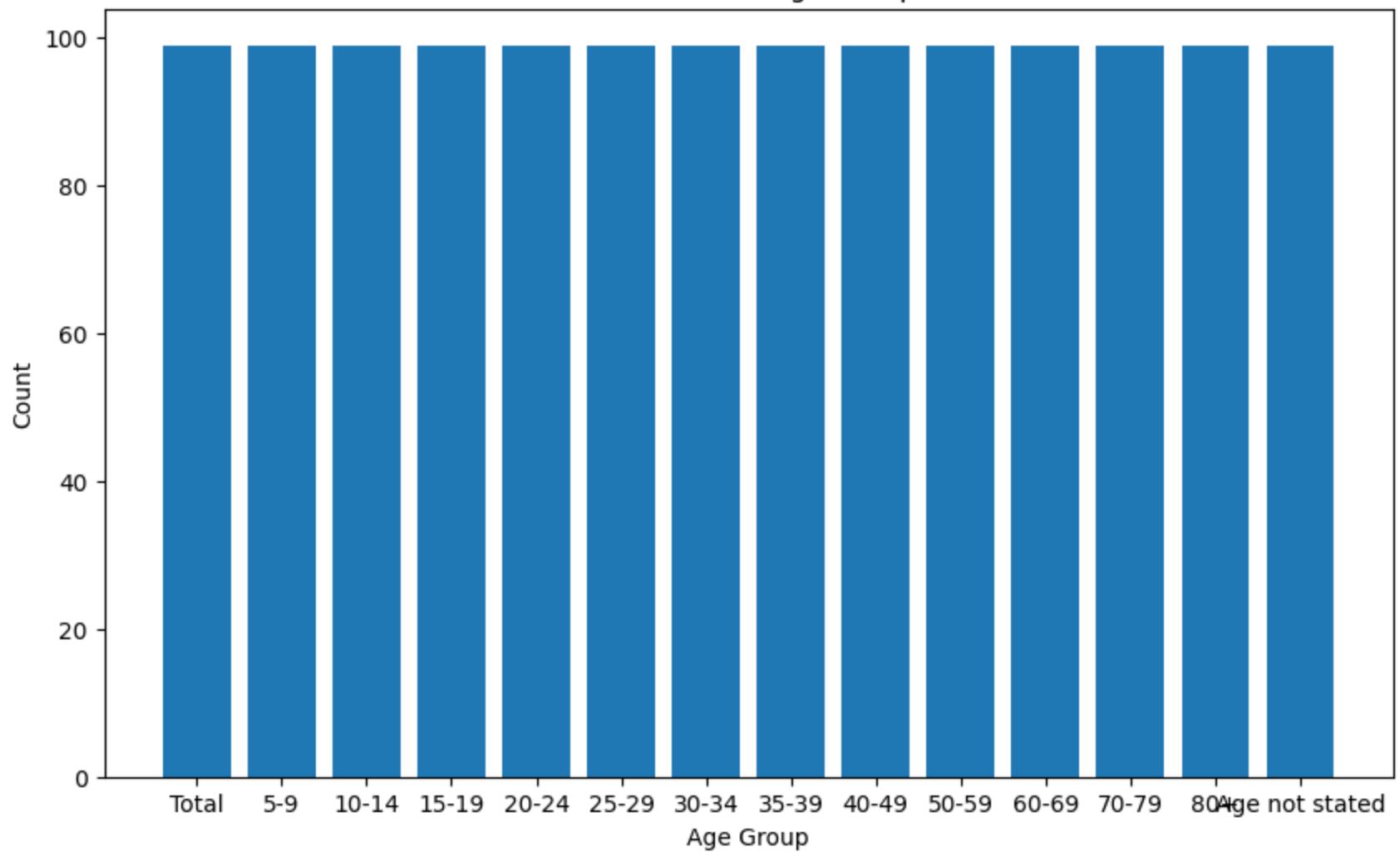
```
Out[ ]: Index(['Table Code', 'State Code', 'District Code', 'Area Name',
   'Total/ Rural/ Urban', 'Age group',
   'Worked for 3 months or more but less than 6 months - Persons',
   'Worked for 3 months or more but less than 6 months - Males',
   'Worked for 3 months or more but less than 6 months - 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')
```

Distribution of Age Groups:

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

# Assuming 'Age group' is a column in your DataFrame
age_distribution = df['Age group'].value_counts()
plt.figure(figsize=(10, 6))
plt.bar(age_distribution.index, age_distribution.values)
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Distribution of Age Groups')
plt.show()
```

Distribution of Age Groups



```
In [ ]: age_stats = df.groupby('Age group')['Area Name'].describe()
print(age_stats)
```

Age group	count	unique	top	freq
10-14	99	33	State - TAMIL NADU	3
15-19	99	33	State - TAMIL NADU	3
20-24	99	33	State - TAMIL NADU	3
25-29	99	33	State - TAMIL NADU	3
30-34	99	33	State - TAMIL NADU	3
35-39	99	33	State - TAMIL NADU	3
40-49	99	33	State - TAMIL NADU	3
5-9	99	33	State - TAMIL NADU	3
50-59	99	33	State - TAMIL NADU	3
60-69	99	33	State - TAMIL NADU	3
70-79	99	33	State - TAMIL NADU	3
80+	99	33	State - TAMIL NADU	3
Age not stated	99	33	State - TAMIL NADU	3
Total	99	33	State - TAMIL NADU	3

District wise analysis

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

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

Out[]:

	Area Name	Age group	Total/ Rural/ Urban	Industrial Category - A - Cultivators - Persons
0	District - Ariyalur	10-14	Rural	68
1	District - Ariyalur	10-14	Total	74
2	District - Ariyalur	10-14	Urban	6
3	District - Ariyalur	15-19	Rural	411
4	District - Ariyalur	15-19	Total	425
5	District - Ariyalur	15-19	Urban	14
6	District - Ariyalur	20-24	Rural	926
7	District - Ariyalur	20-24	Total	950
8	District - Ariyalur	20-24	Urban	24
9	District - Ariyalur	25-29	Rural	1358
10	District - Ariyalur	25-29	Total	1402
11	District - Ariyalur	25-29	Urban	44
12	District - Ariyalur	30-34	Rural	1308
13	District - Ariyalur	30-34	Total	1346
14	District - Ariyalur	30-34	Urban	38
15	District - Ariyalur	35-39	Rural	1412
16	District - Ariyalur	35-39	Total	1470
17	District - Ariyalur	35-39	Urban	58
18	District - Ariyalur	40-49	Rural	2617
19	District - Ariyalur	40-49	Total	2756
20	District - Ariyalur	40-49	Urban	139
21	District - Ariyalur	5-9	Rural	33
22	District - Ariyalur	5-9	Total	34
23	District - Ariyalur	5-9	Urban	1
24	District - Ariyalur	50-59	Rural	1867
25	District - Ariyalur	50-59	Total	1969
26	District - Ariyalur	50-59	Urban	102
27	District - Ariyalur	60-69	Rural	1303
28	District - Ariyalur	60-69	Total	1365
29	District - Ariyalur	60-69	Urban	62

Visualization District wise

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

# 'Area Name' represents the districts, 'Age group' represents the age groups, 'Total/ Rural/ Urban' represents rural or urban areas
# '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/ 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()
```

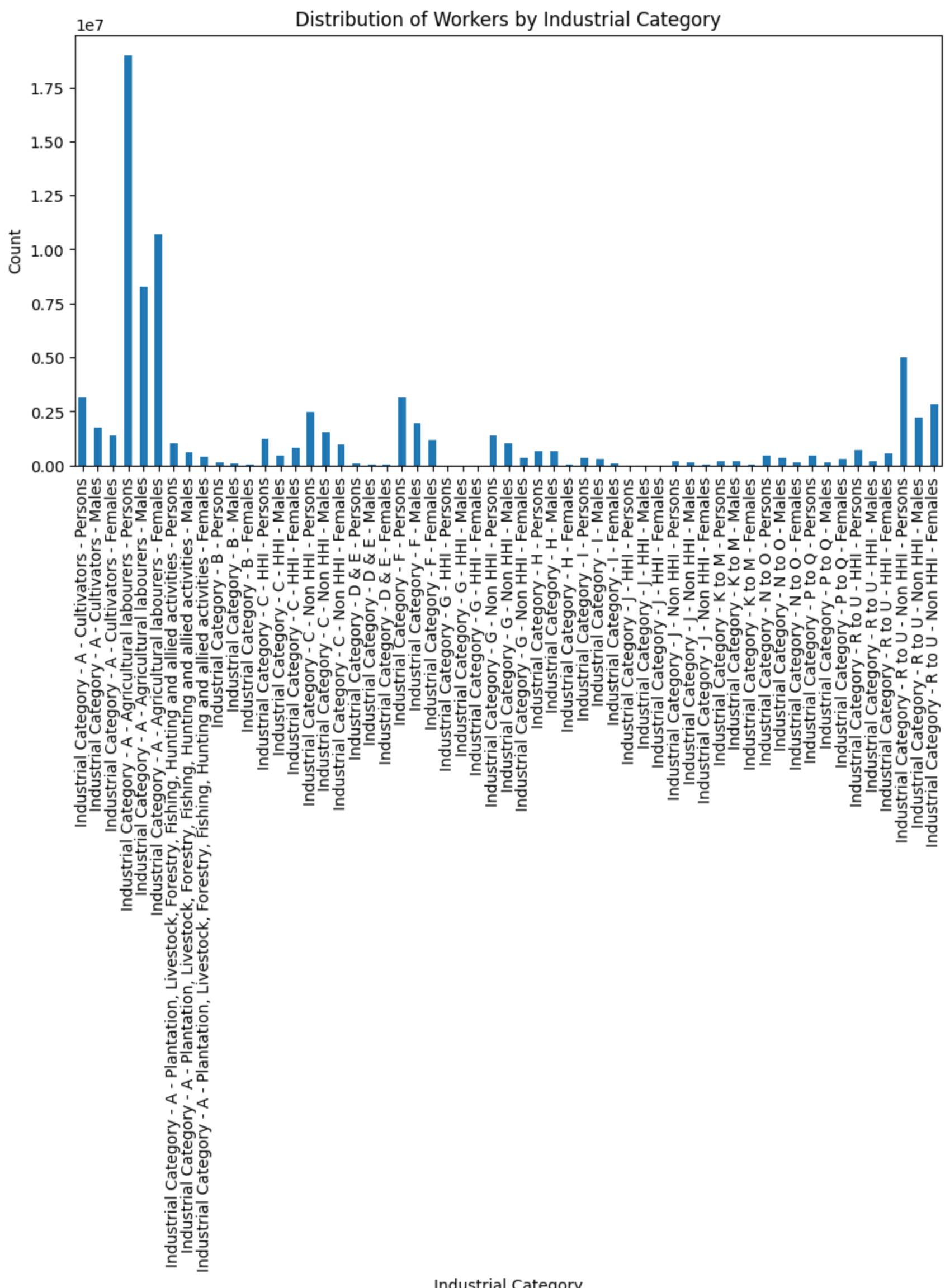
```
In [ ]: import matplotlib.pyplot as plt

# Combine all the relevant columns for industrial categories
industrial_columns = ['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']

# Sum the counts across all the industrial category columns
industrial_counts = df[industrial_columns].sum()

# Create a bar chart to visualize the distribution of workers in each industrial category
plt.figure(figsize=(10, 5))
industrial_counts.plot(kind='bar', title='Distribution of Workers by Industrial Category')
plt.xlabel('Industrial Category')
plt.ylabel('Count')
plt.show()
```



```
In [ ]: # CLUSTERING OF AGE GROUPS AND INDUSTRIAL GROUPS INCLUDING MALES AND FEMALES CATEGORIES
```

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Load your dataset into a DataFrame (replace 'your_dataset_path' with the actual path)
df = pd.read_csv('marginal_workers_tamil_nadu.csv.csv')

# Data Cleaning and Exploration
numeric_columns = ['Industrial Category - A - Cultivators - Persons', 'Industrial Category - A - Cultivators - Males',
'IIndustrial Category - A - Cultivators - Females', 'Industrial Category - B - Persons',
'IIndustrial Category - B - Males', 'Industrial Category - B - Females',
'IIndustrial Category - C - HHI - Males',
'IIndustrial Category - C - HHI - Females',
'IIndustrial Category - D & E - Males',
'IIndustrial Category - D & E - Females',
'IIndustrial Category - F - Persons',
'IIndustrial Category - F - Males',
'IIndustrial Category - F - Females',
'IIndustrial 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'
# List of your numeric column names

]

# 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']

# Standardizing Data
scaler = StandardScaler()
features = numeric_columns[1:] # Use all the industrial categories (starting from the 2nd column)

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

# Applying K-Means Algorithm
n_clusters = 3 # Assuming you've determined the optimal number of clusters
kmeans = KMeans(n_clusters=n_clusters, random_state=0)
df['Cluster'] = kmeans.fit_predict(df[features])

# Scatter Plots for All Industrial Categories
for feature in features:
    plt.figure(figsize=(15, 8))

    for cluster in range(n_clusters):
        cluster_df = df[df['Cluster'] == cluster]
        plt.scatter(cluster_df['Age group'], cluster_df[feature], label=f'Cluster {cluster}')

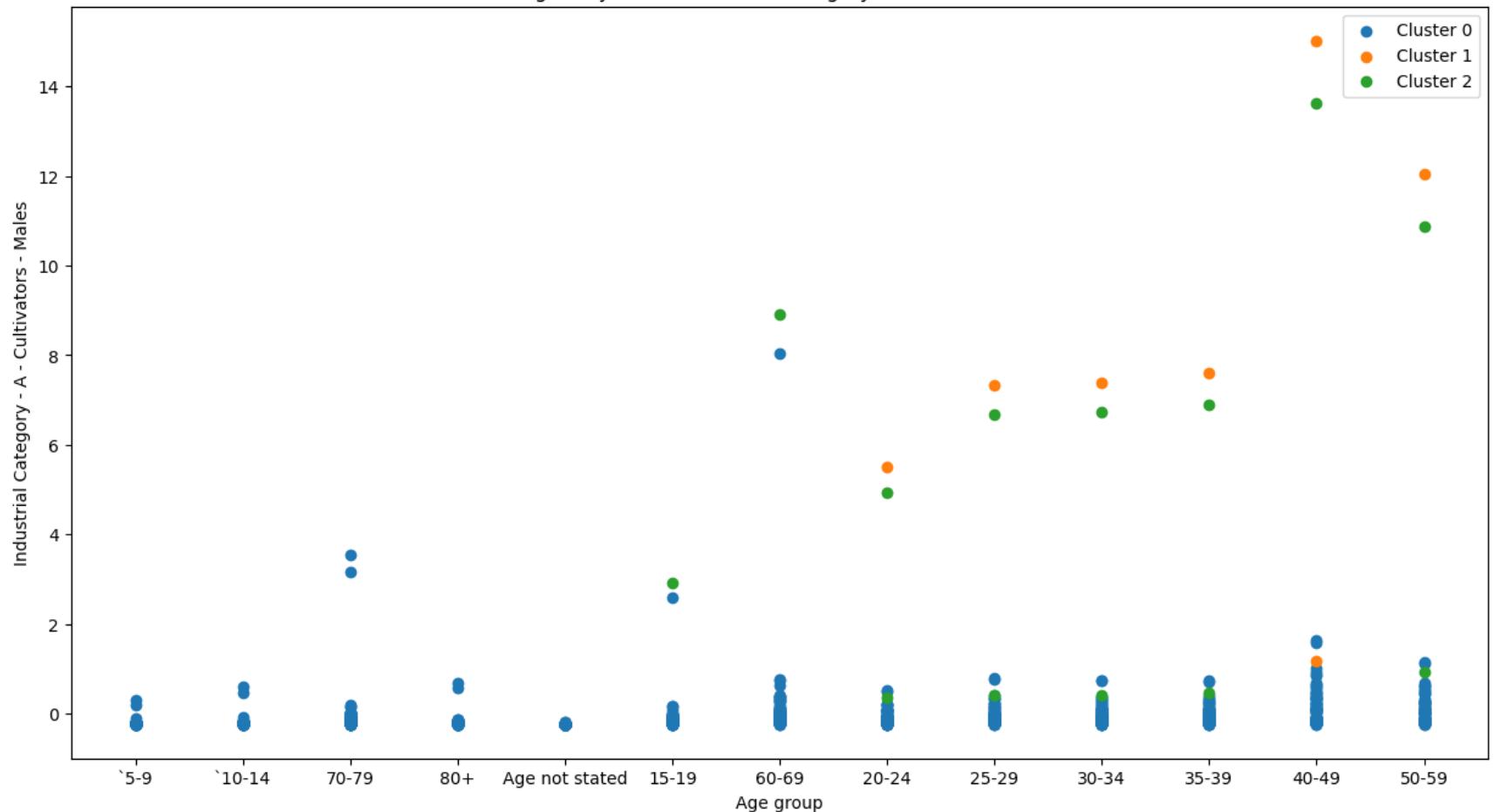
    plt.xlabel('Age group')
    plt.ylabel(feature)
    plt.title(f'Clustering Analysis for {feature}')
    plt.legend()
    plt.show()

# Further steps for analysis and recommendations can be done here

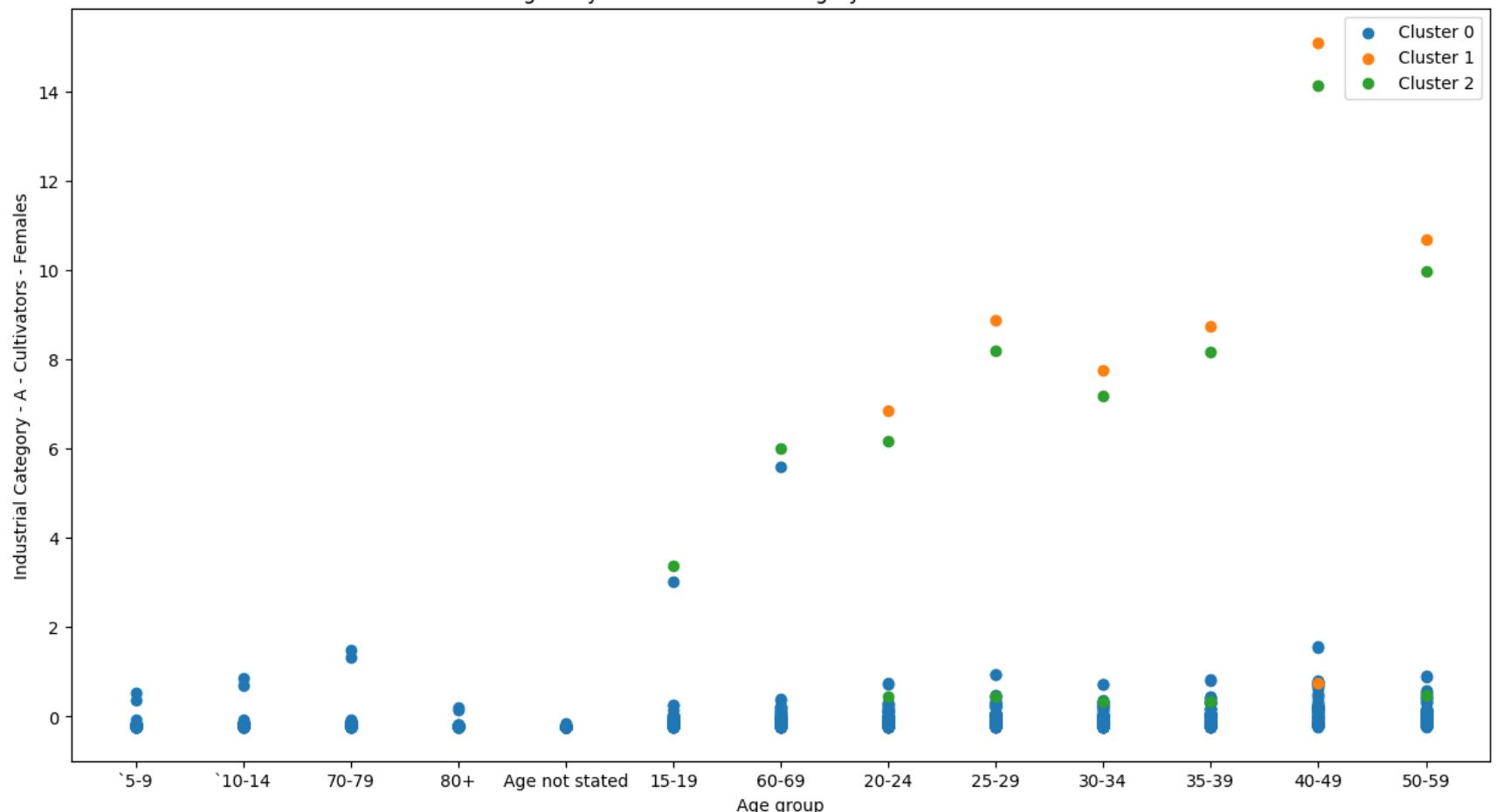
```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(

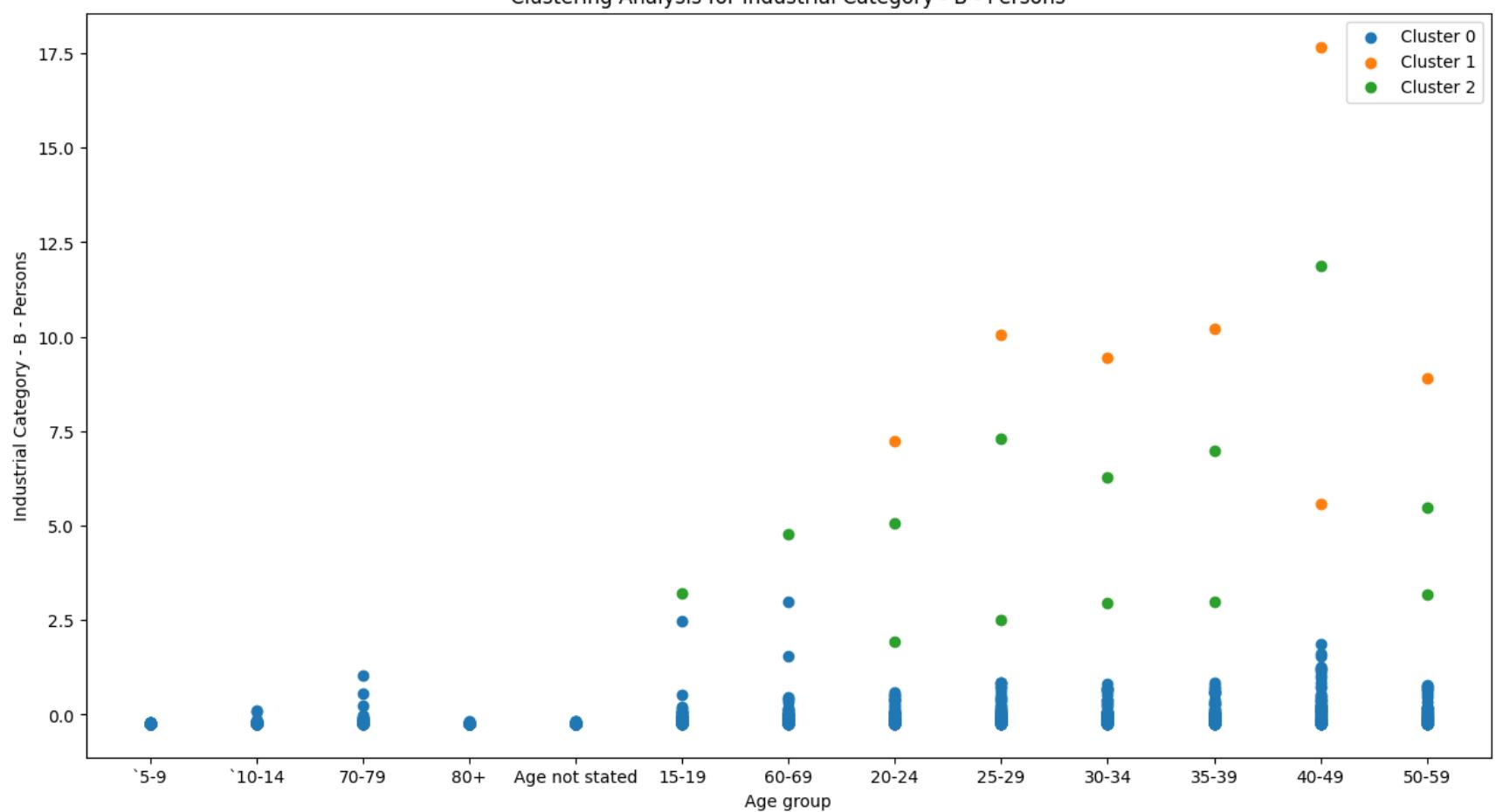
Clustering Analysis for Industrial Category - A - Cultivators - Males



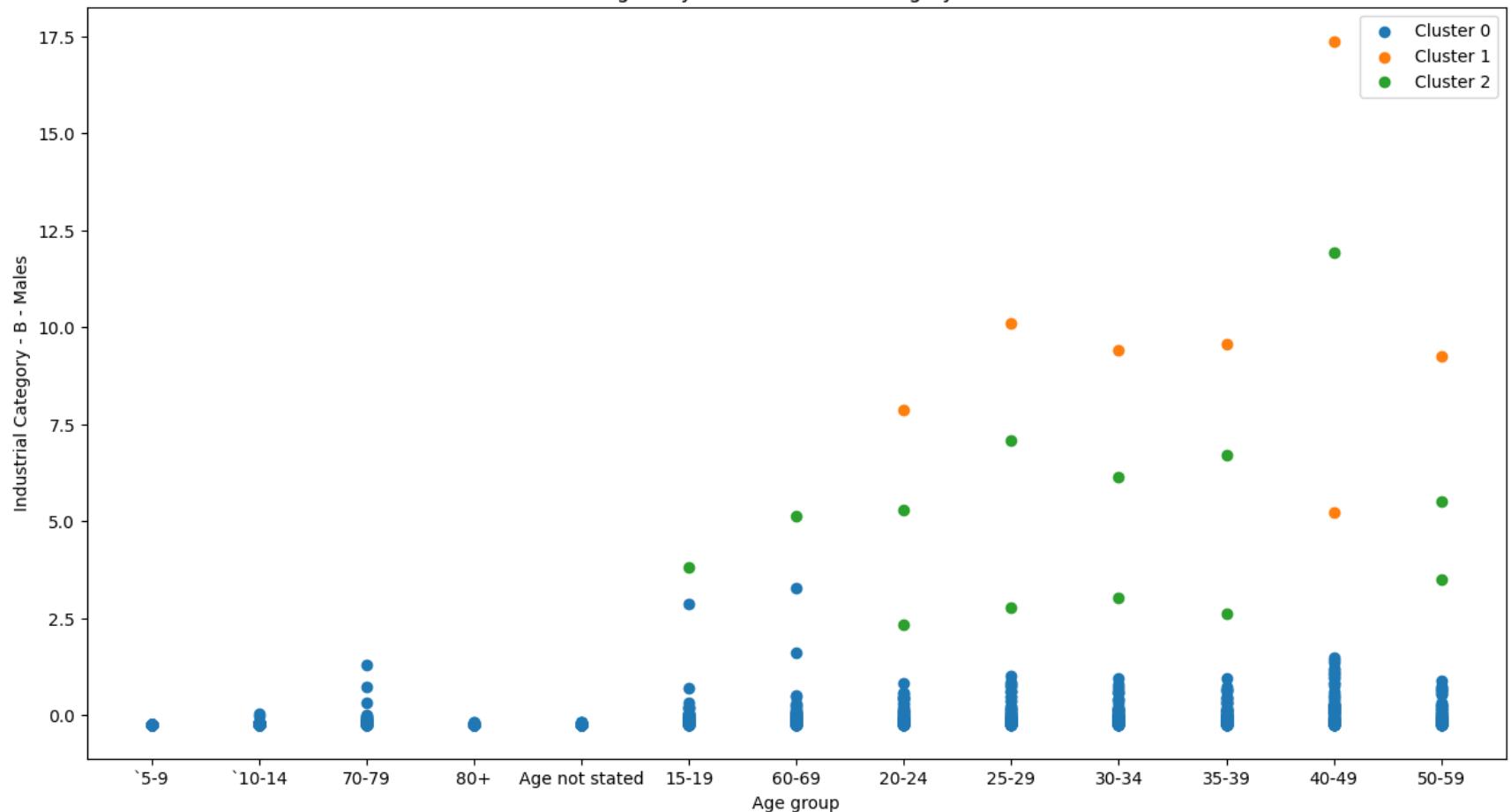
Clustering Analysis for Industrial Category - A - Cultivators - Females



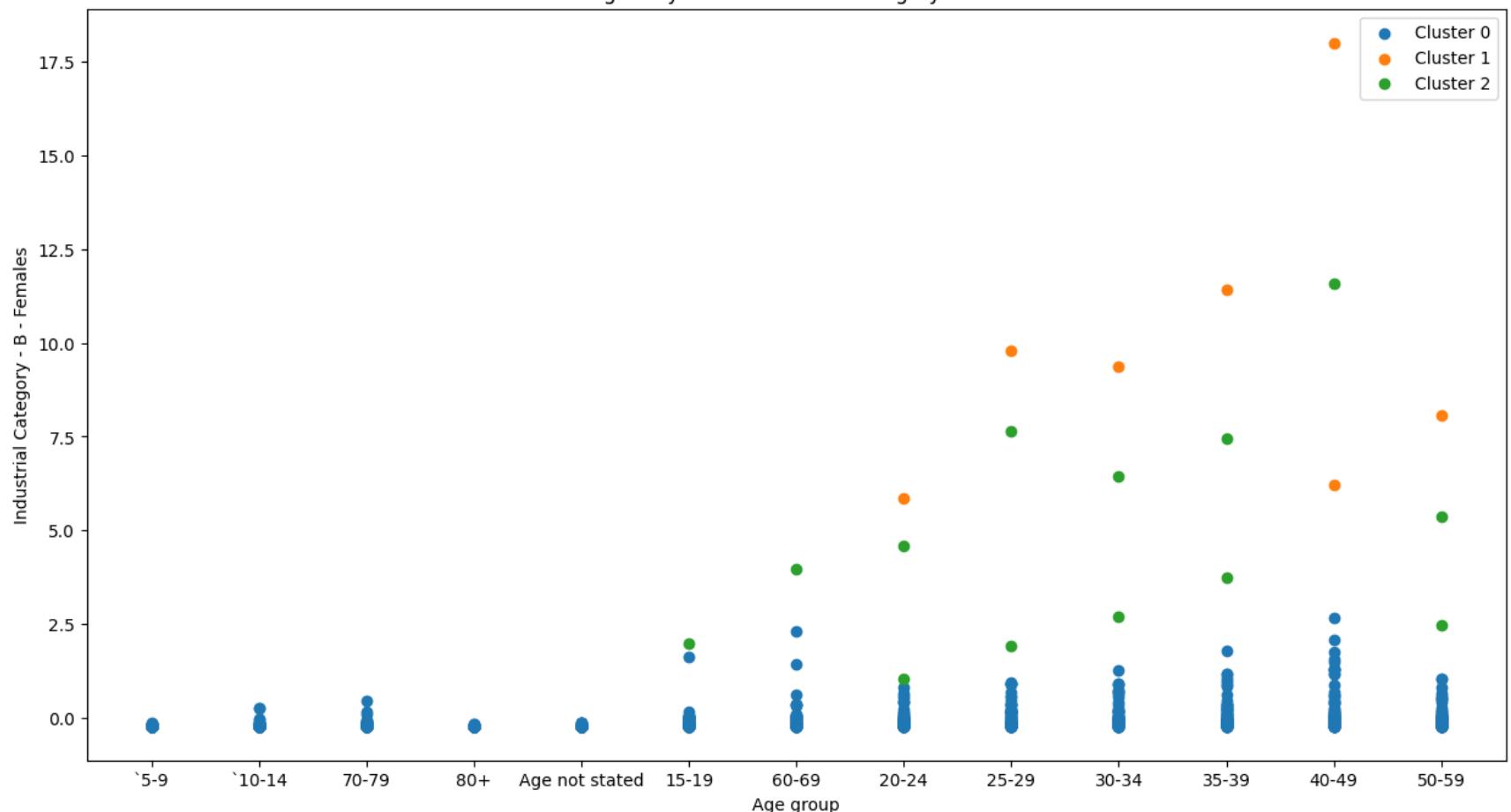
Clustering Analysis for Industrial Category - B - Persons



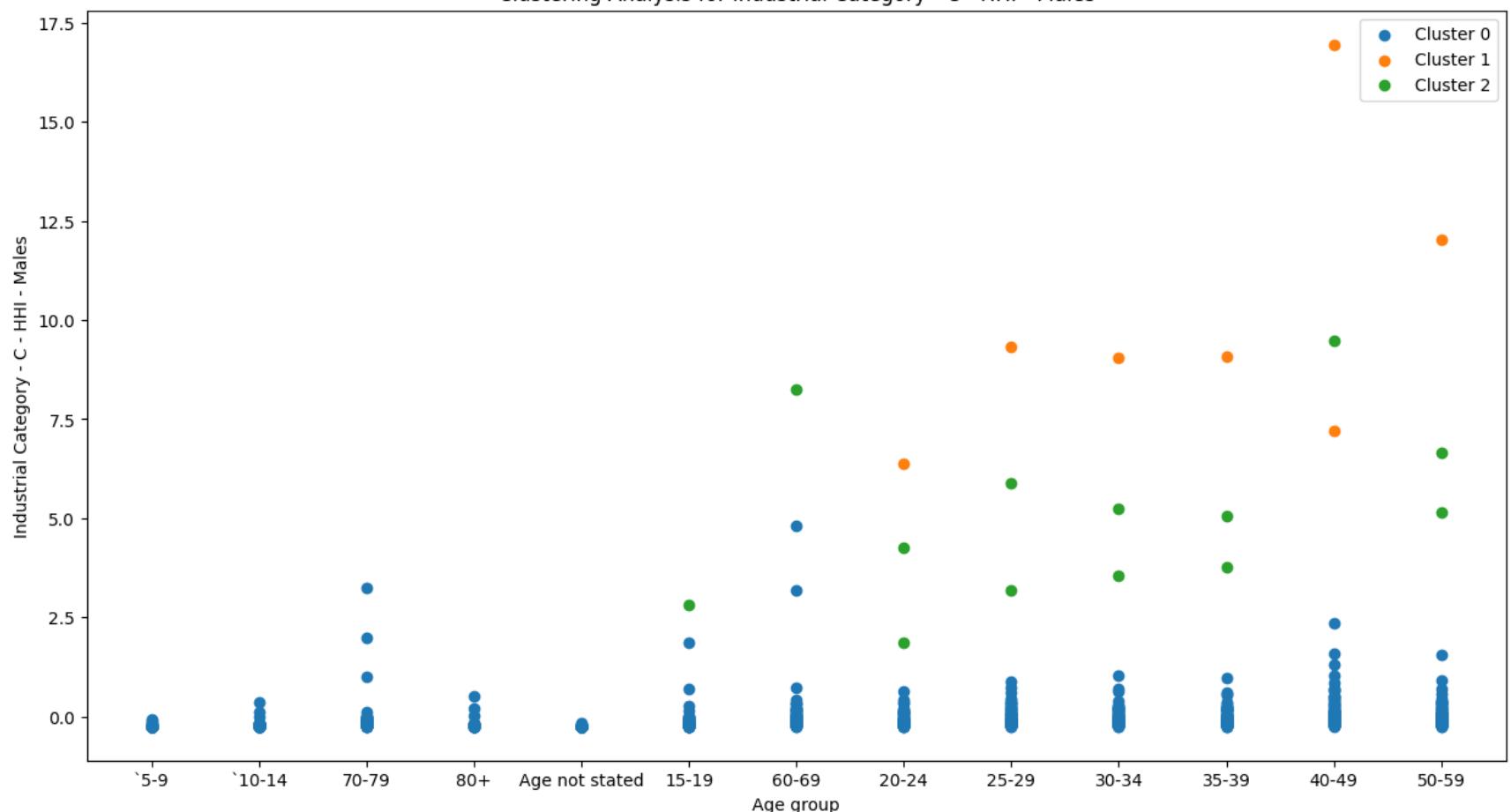
Clustering Analysis for Industrial Category - B - Males



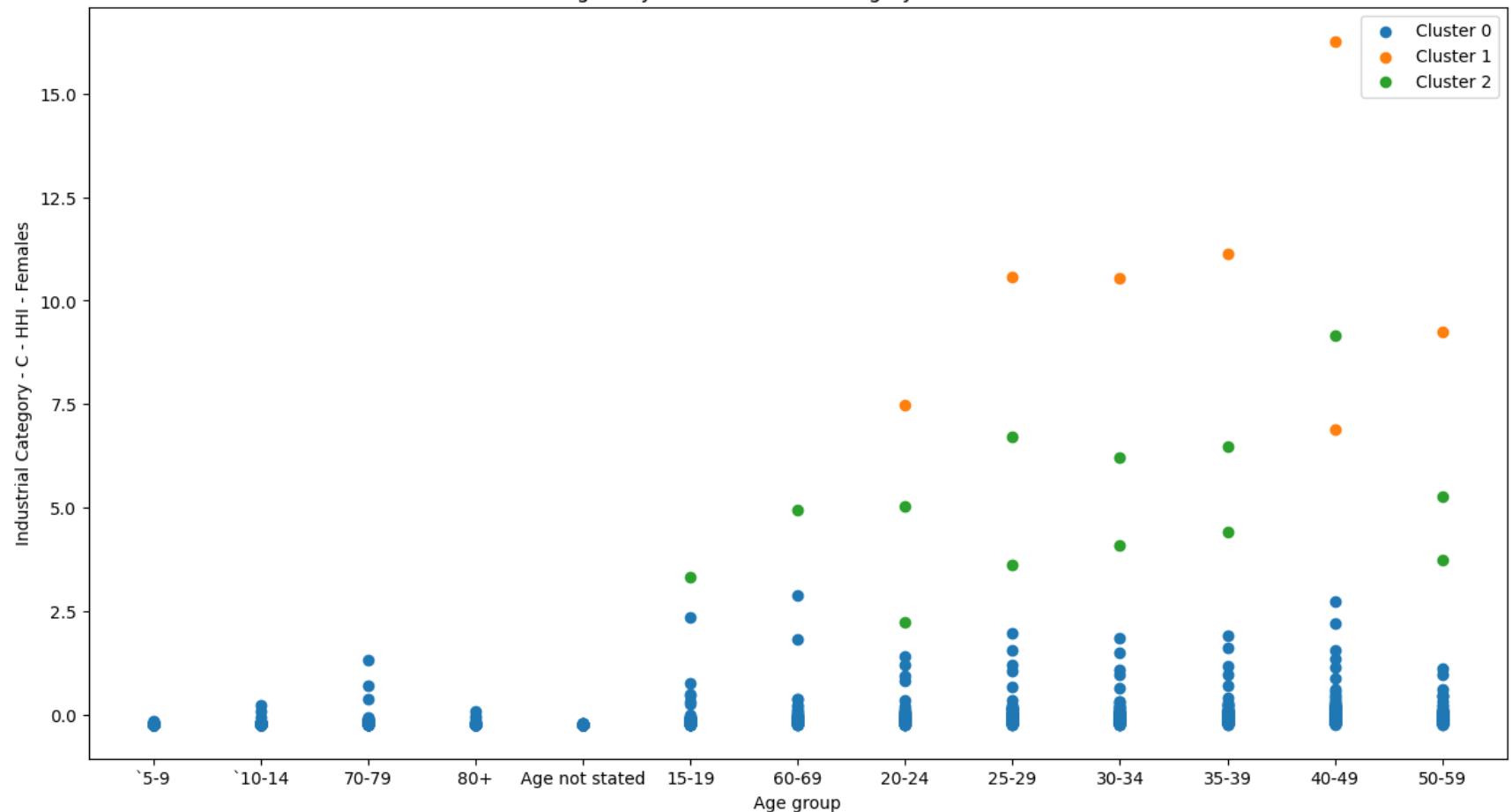
Clustering Analysis for Industrial Category - B - Females



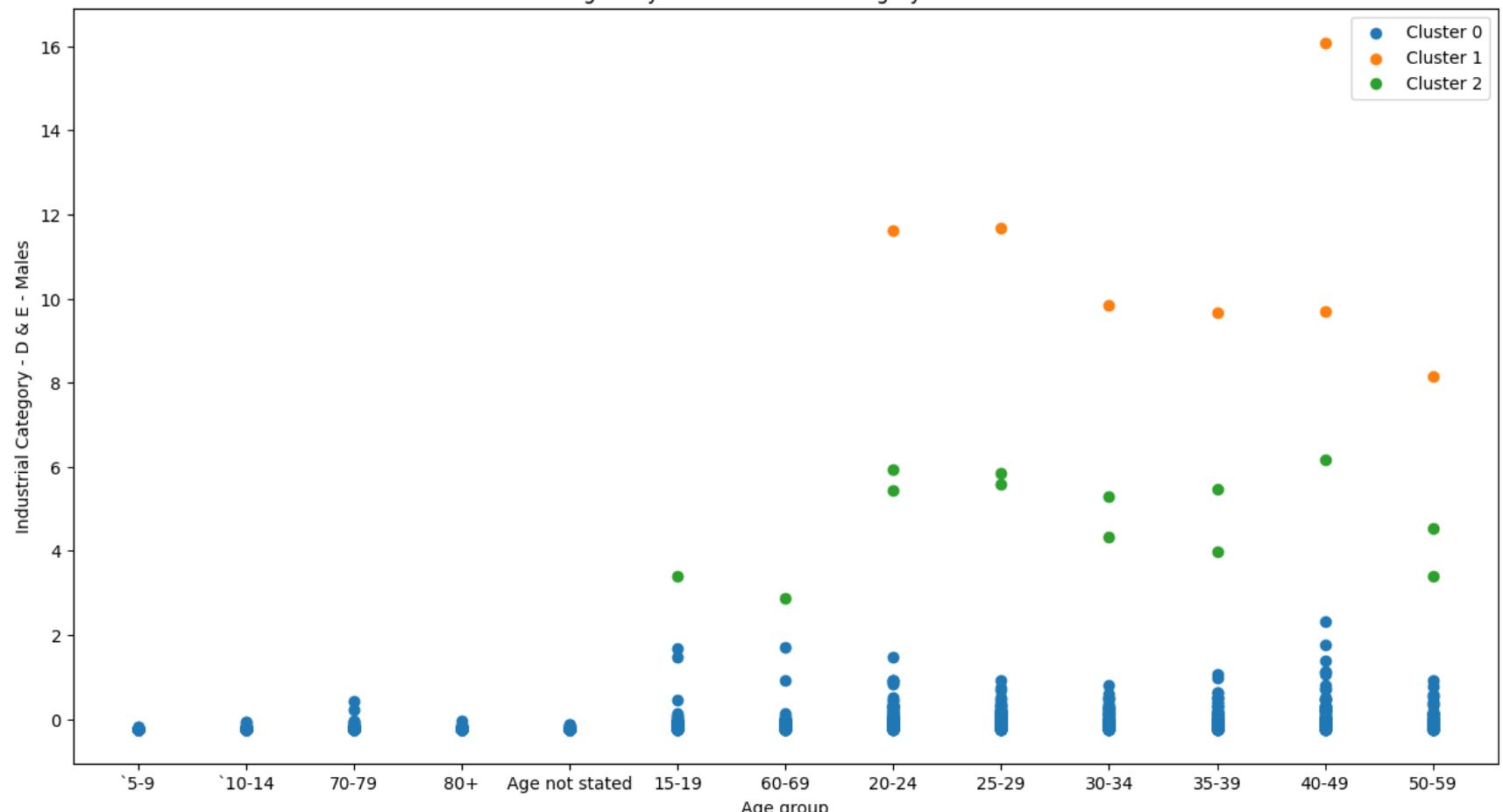
Clustering Analysis for Industrial Category - C - HHI - Males



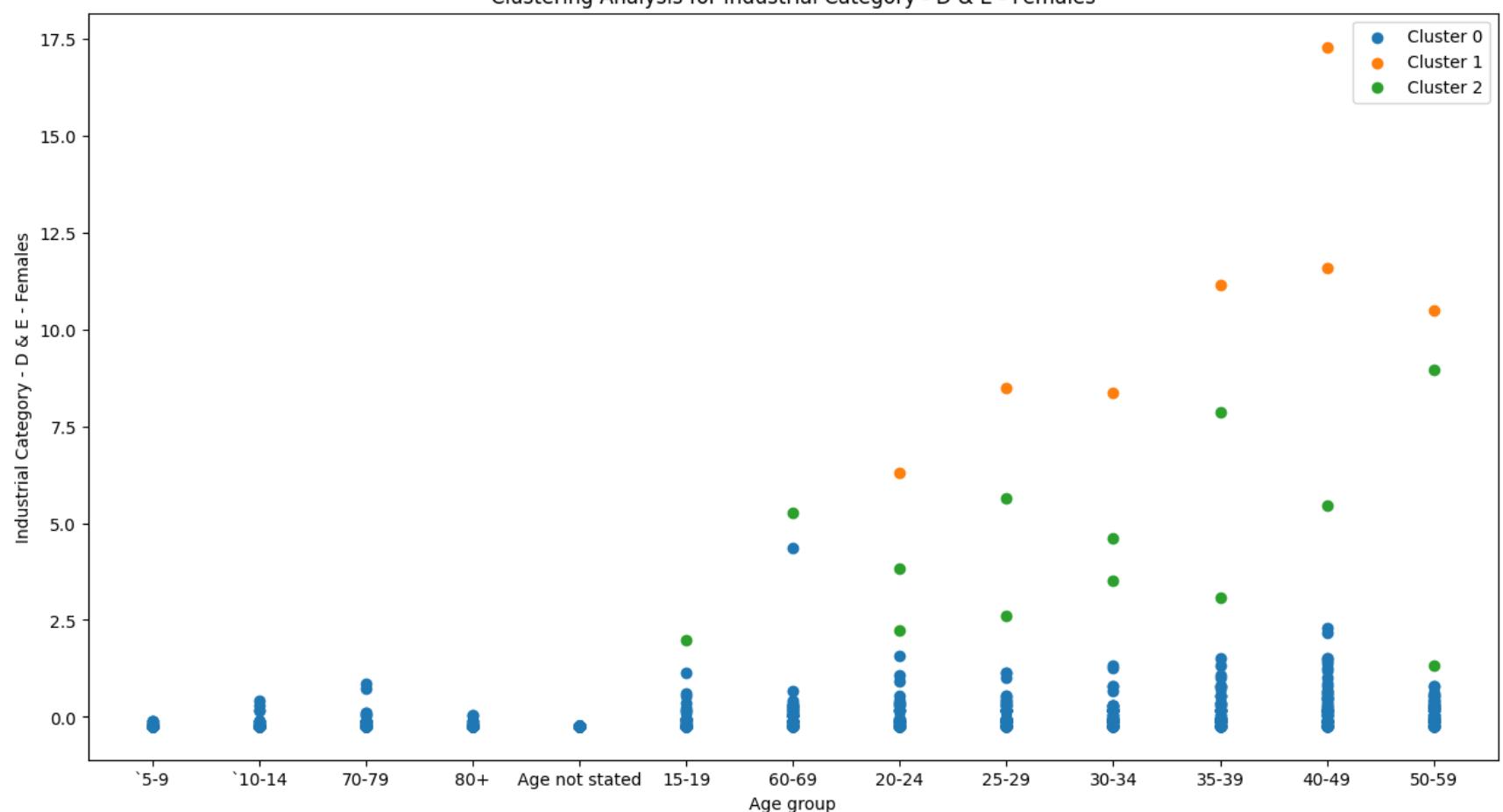
Clustering Analysis for Industrial Category - C - HHI - Females



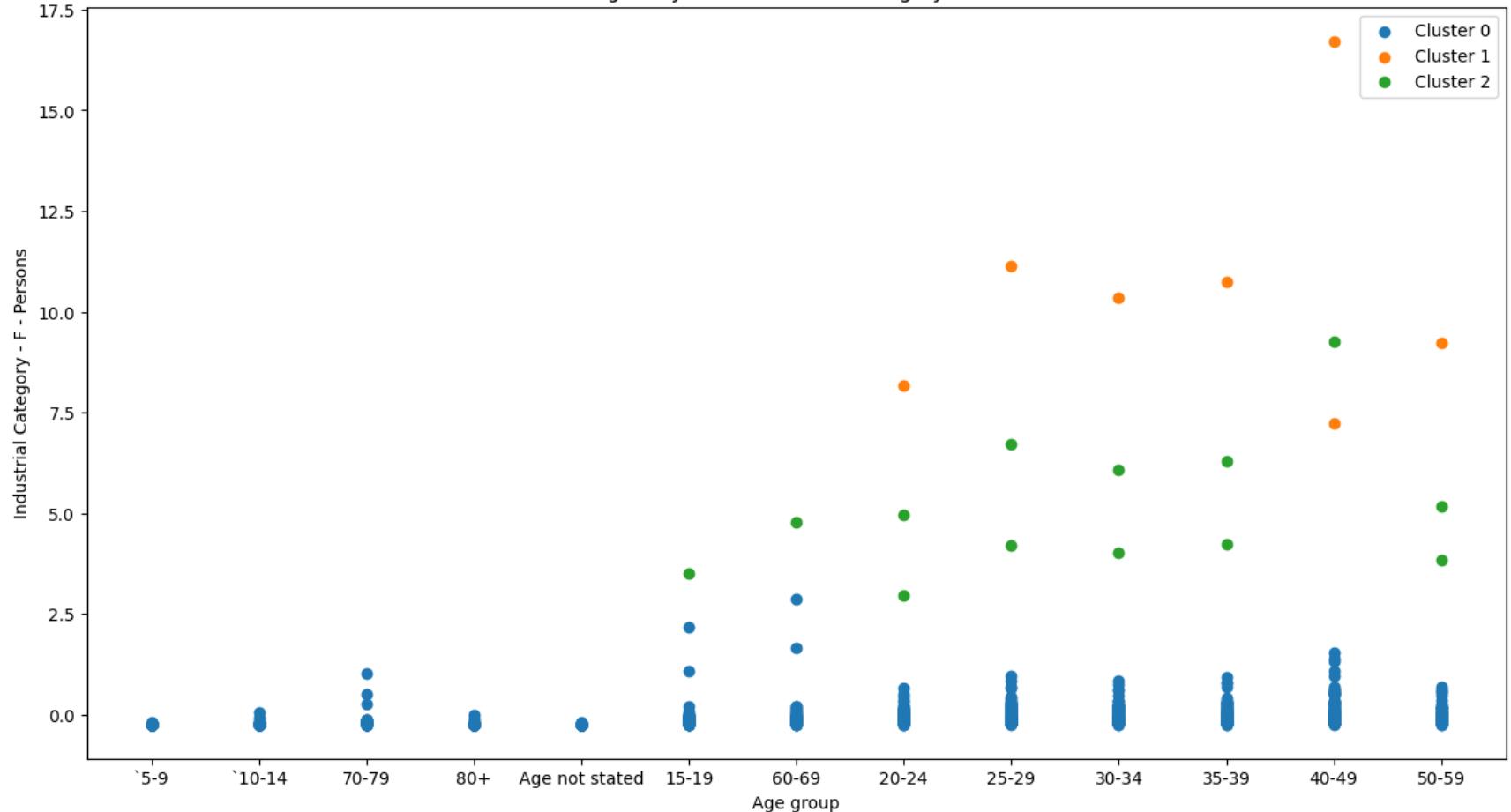
Clustering Analysis for Industrial Category - D & E - Males



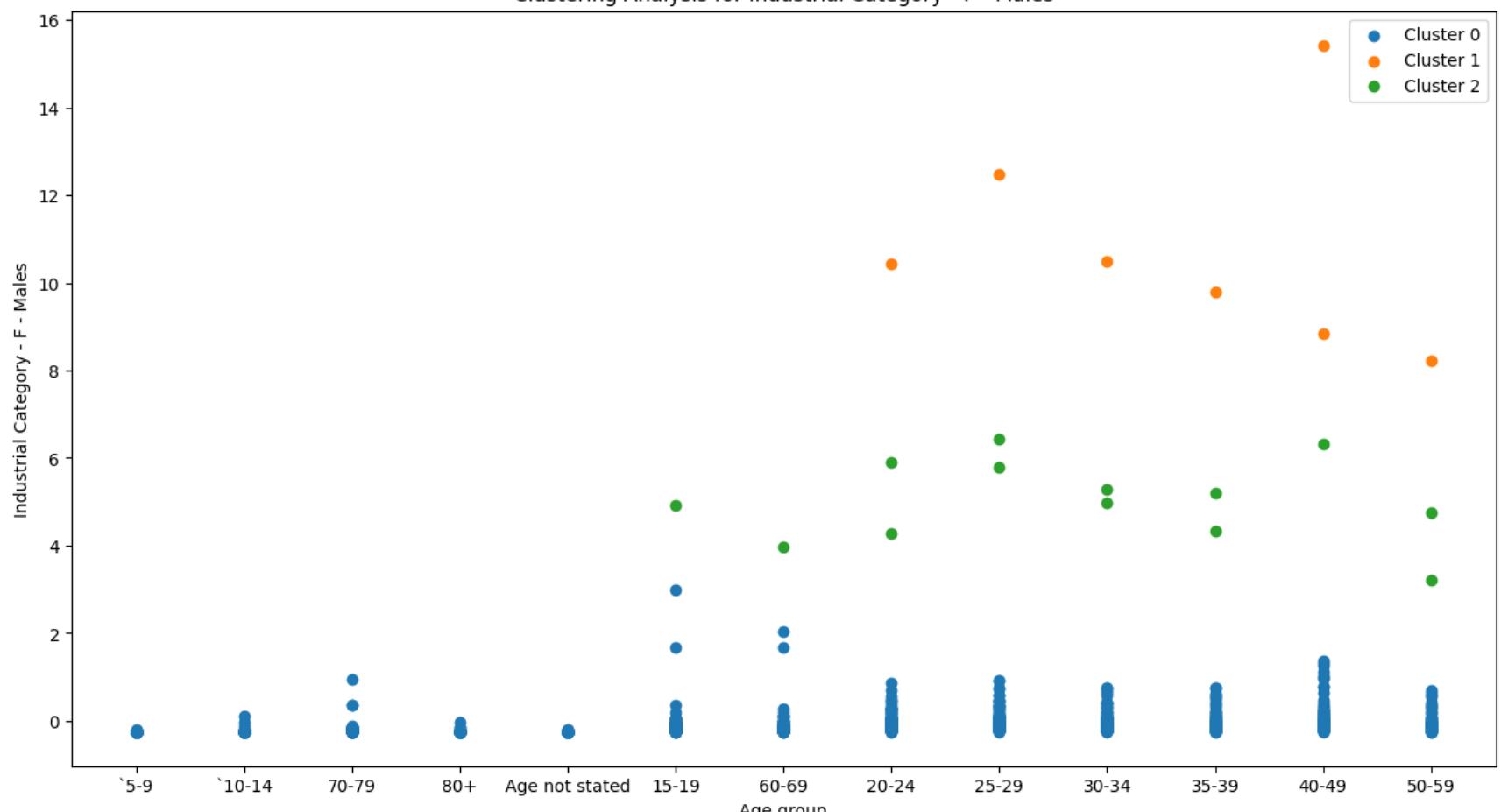
Clustering Analysis for Industrial Category - D & E - Females



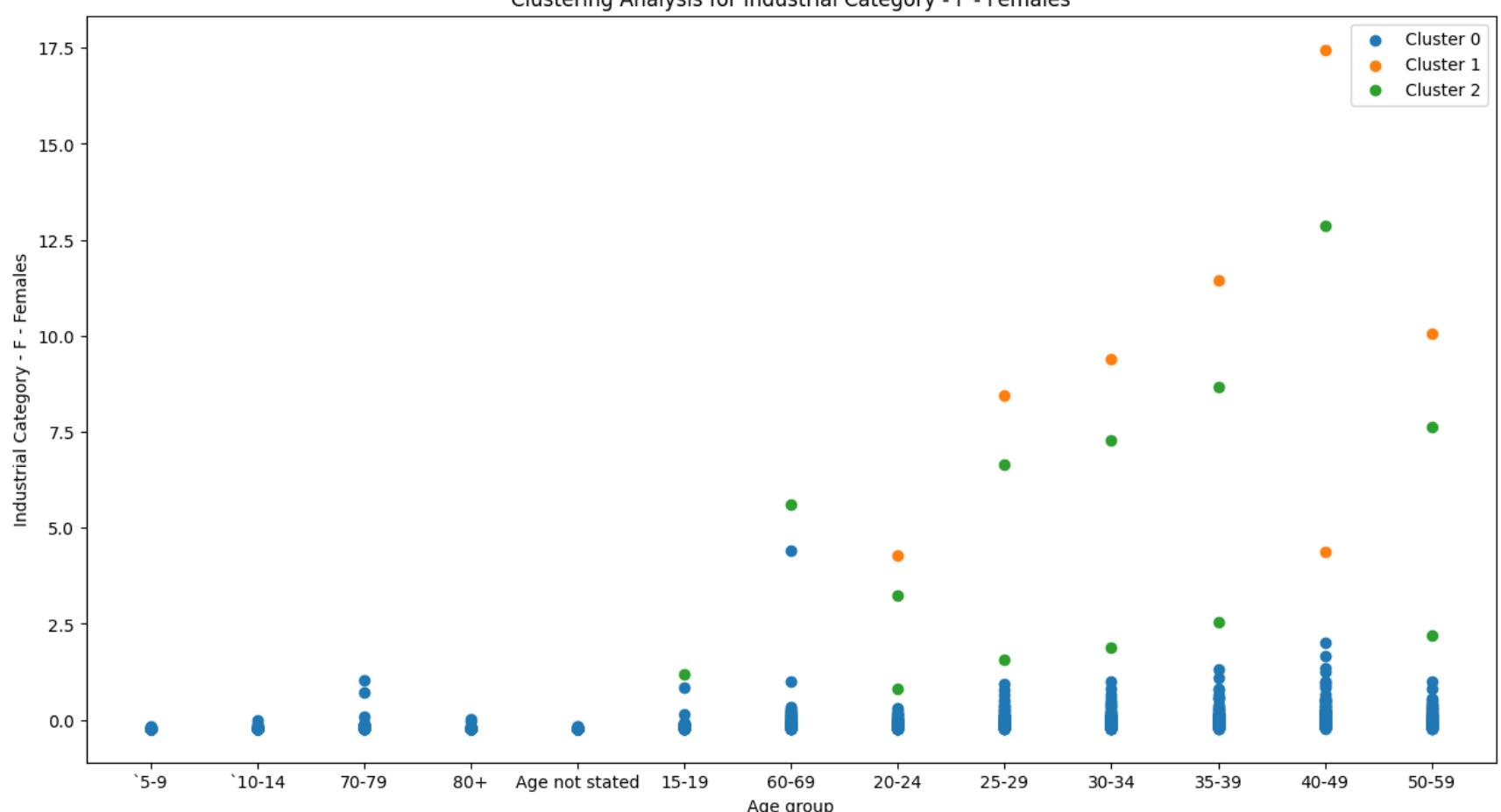
Clustering Analysis for Industrial Category - F - Persons



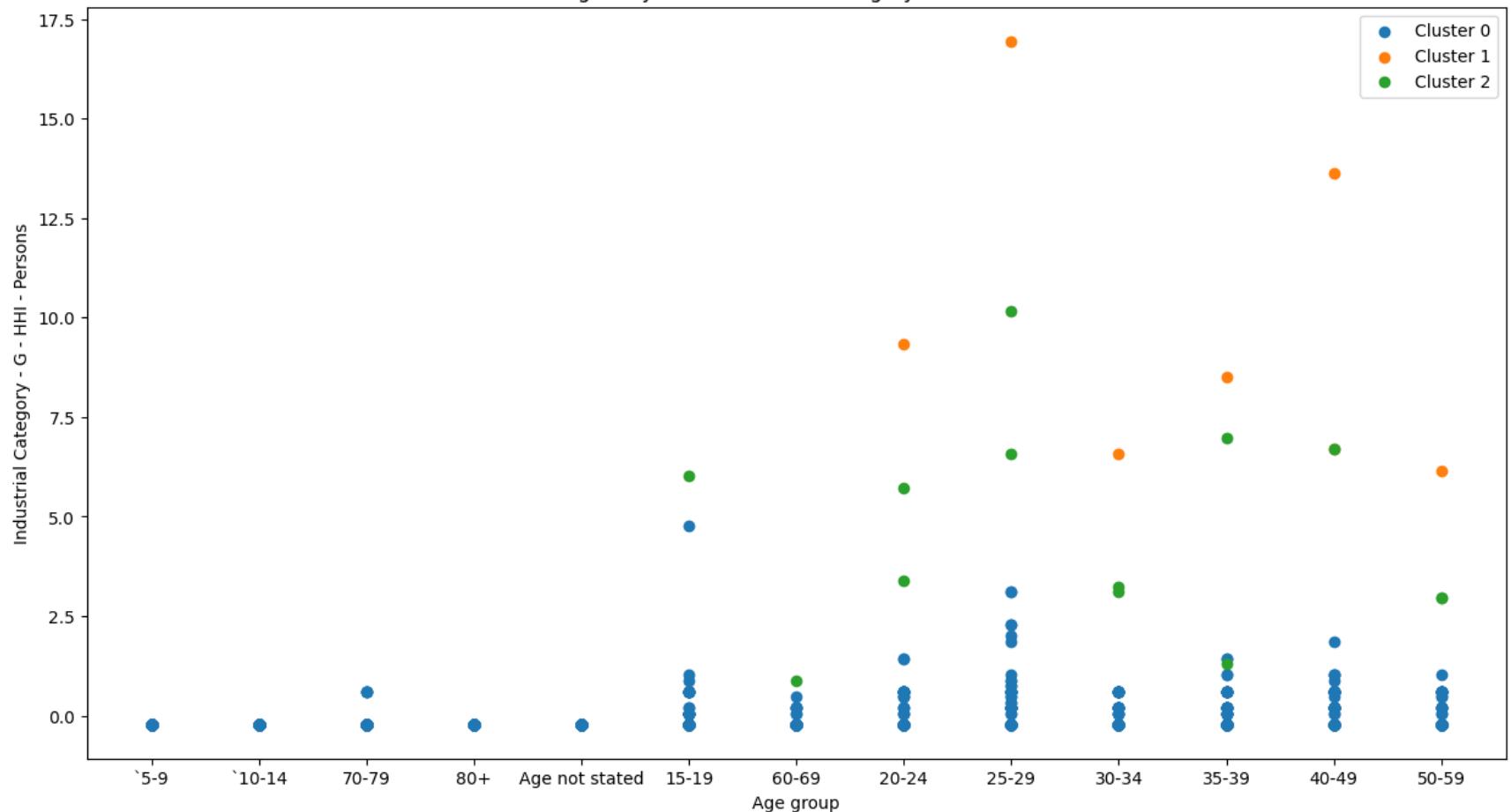
Clustering Analysis for Industrial Category - F - Males



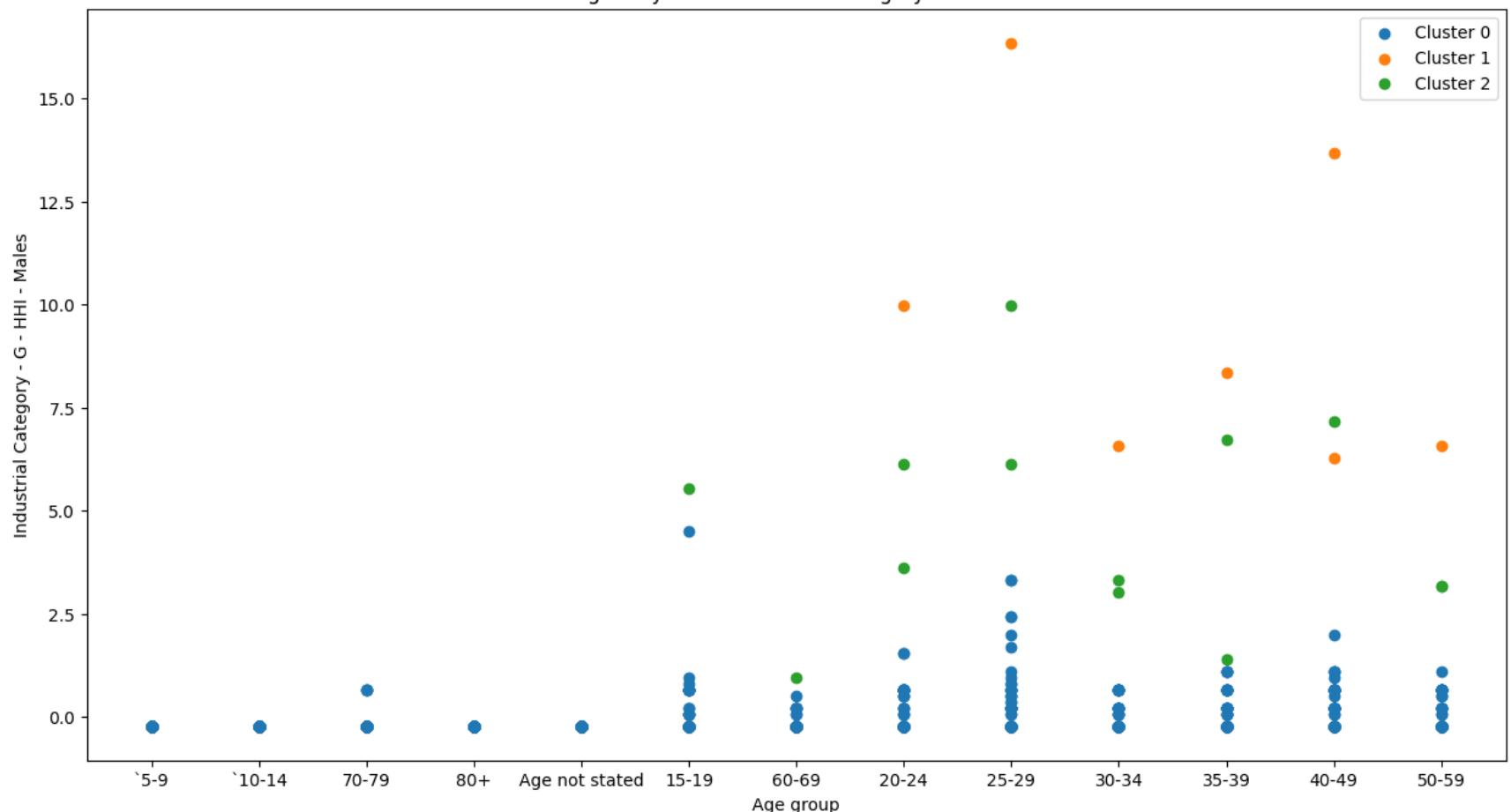
Clustering Analysis for Industrial Category - F - Females



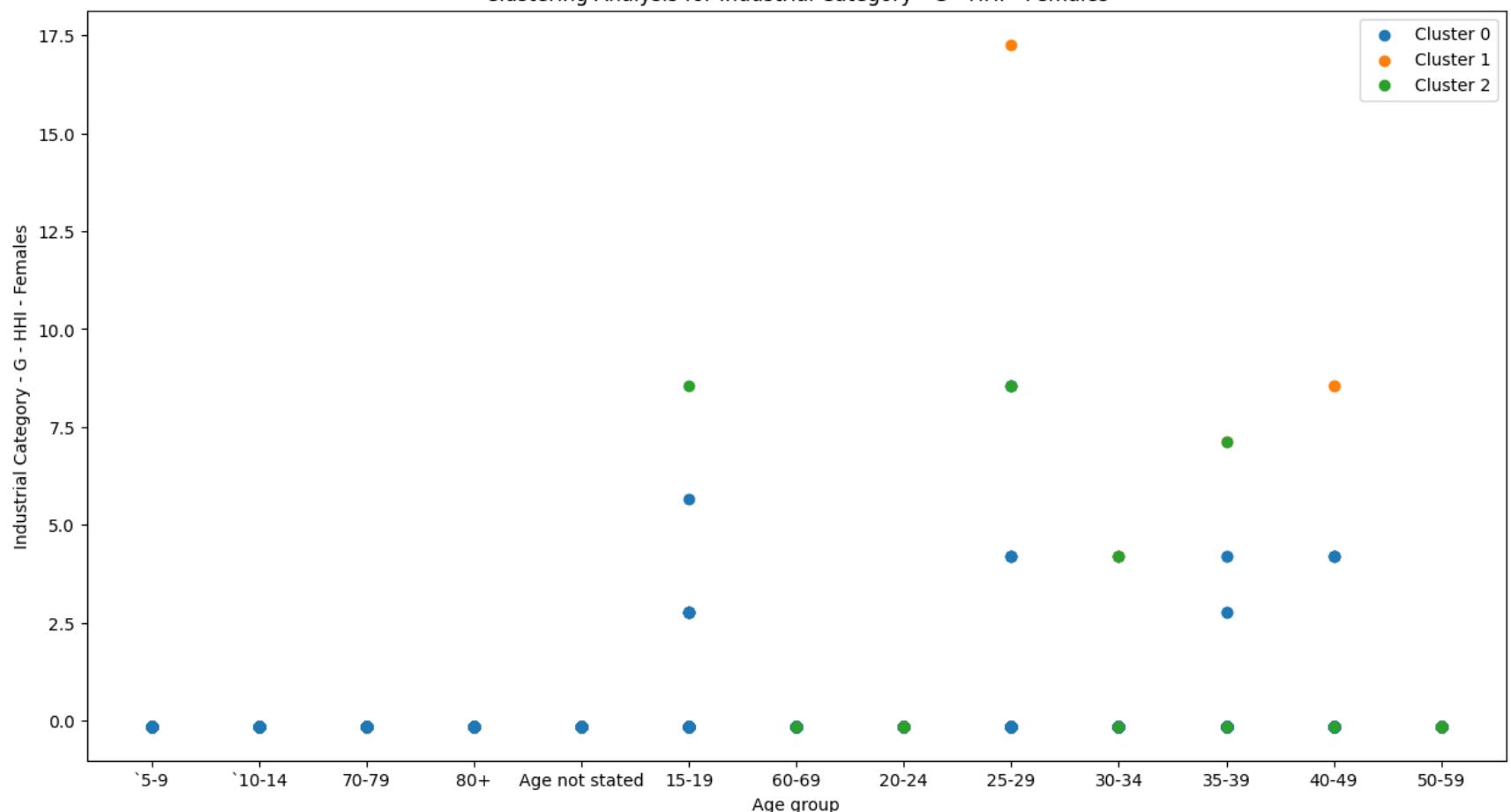
Clustering Analysis for Industrial Category - G - HHI - Persons



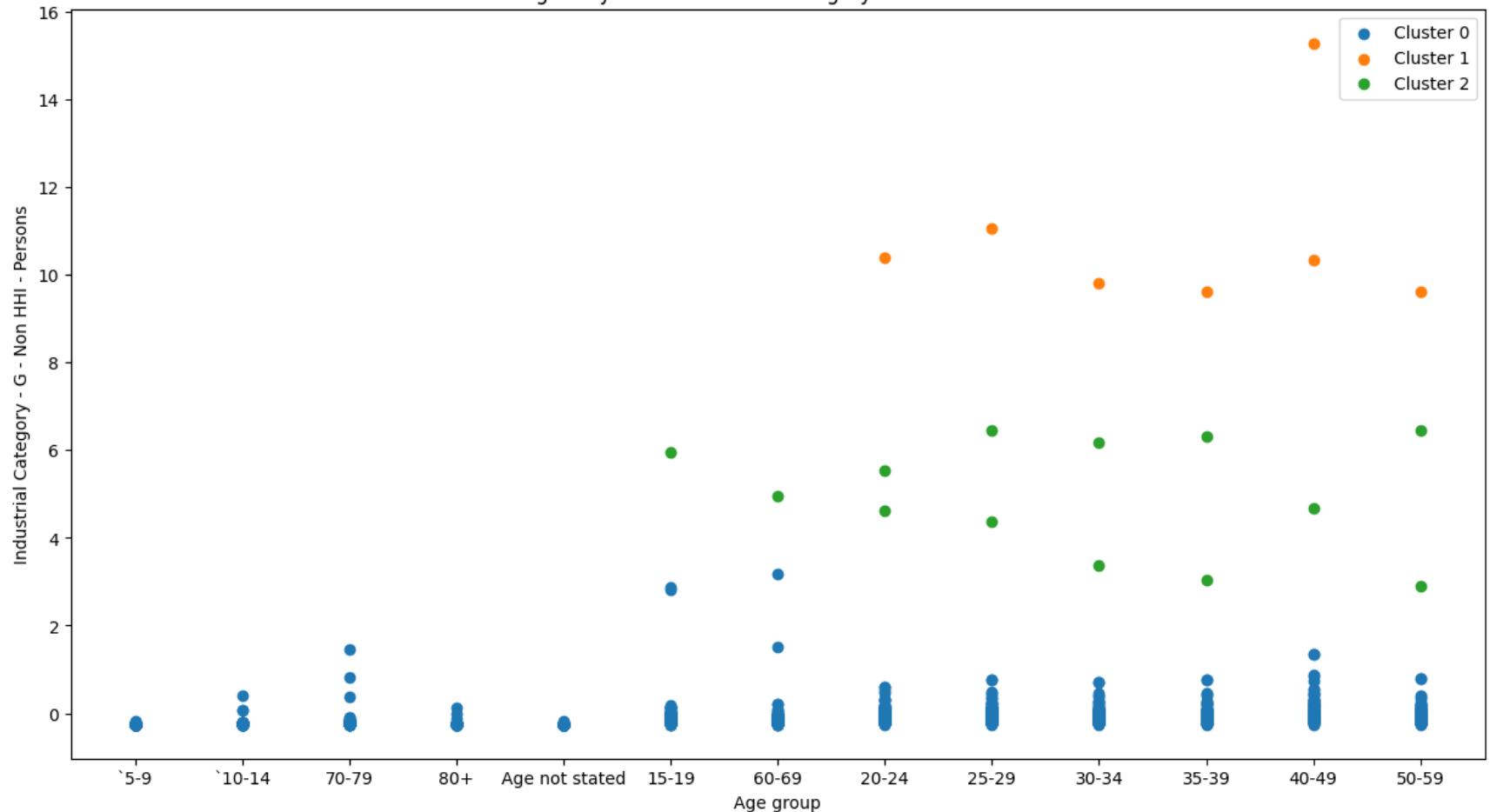
Clustering Analysis for Industrial Category - G - HHI - Males



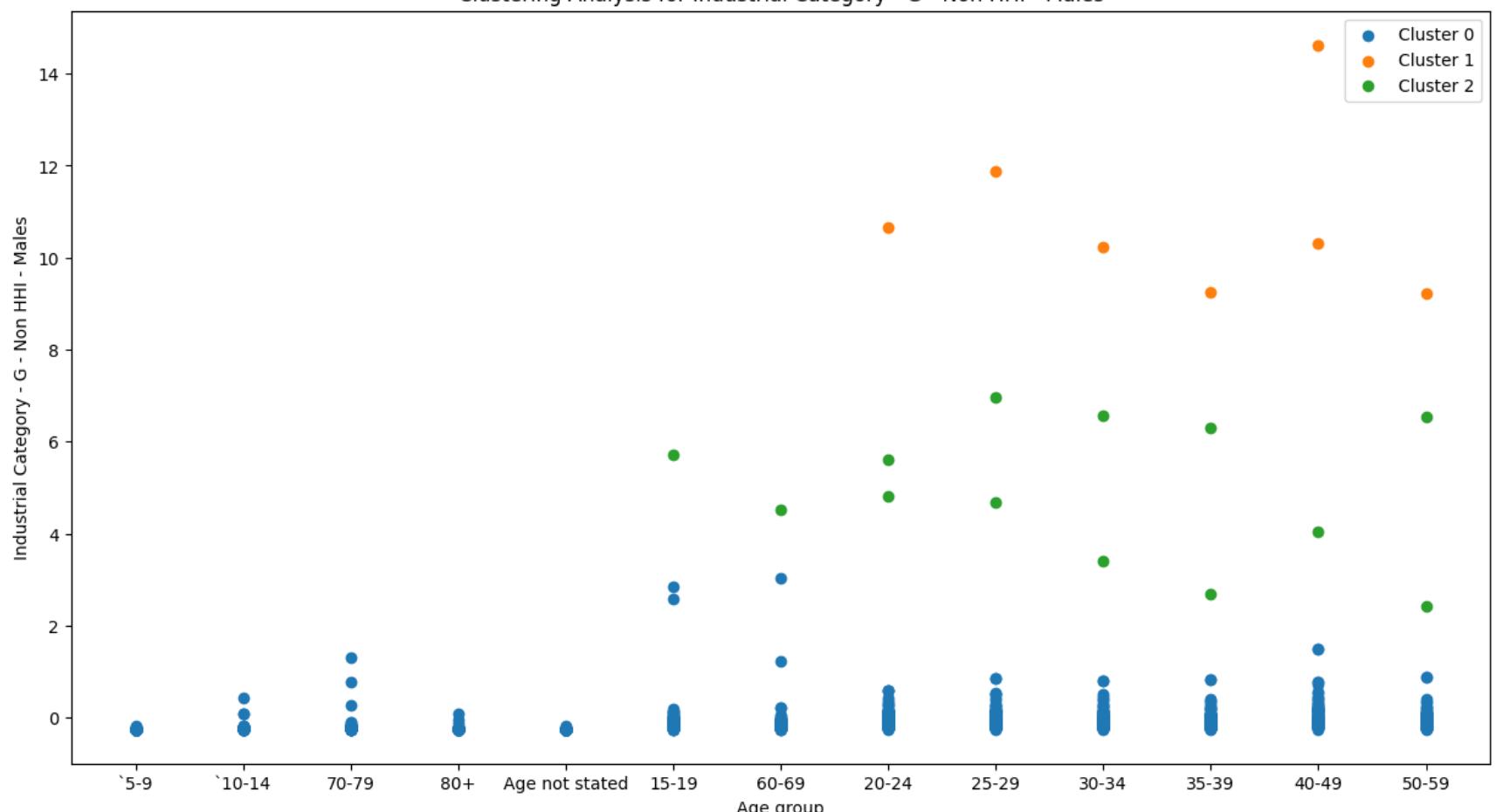
Clustering Analysis for Industrial Category - G - HHI - Females



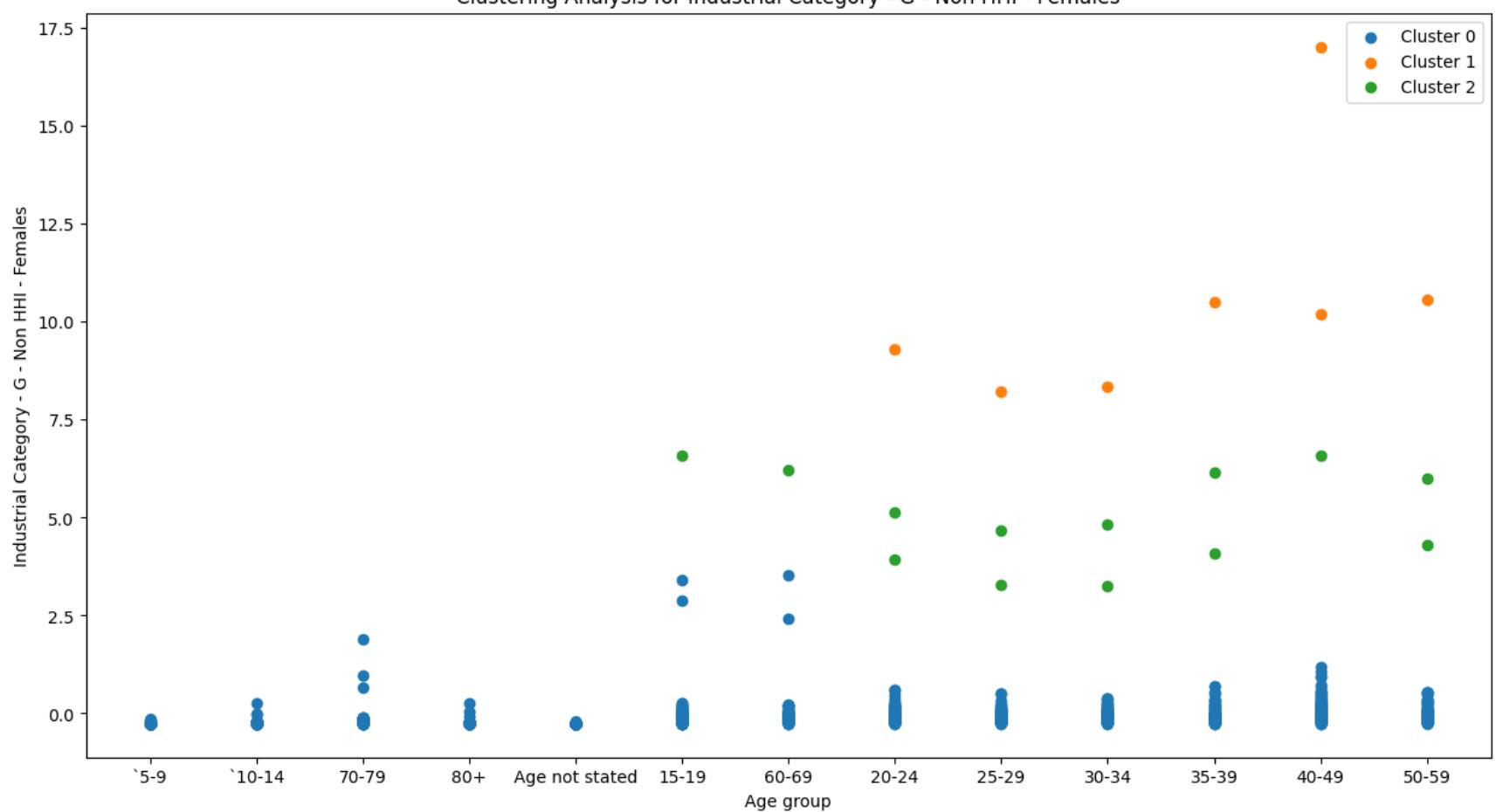
Clustering Analysis for Industrial Category - G - Non HHI - Persons



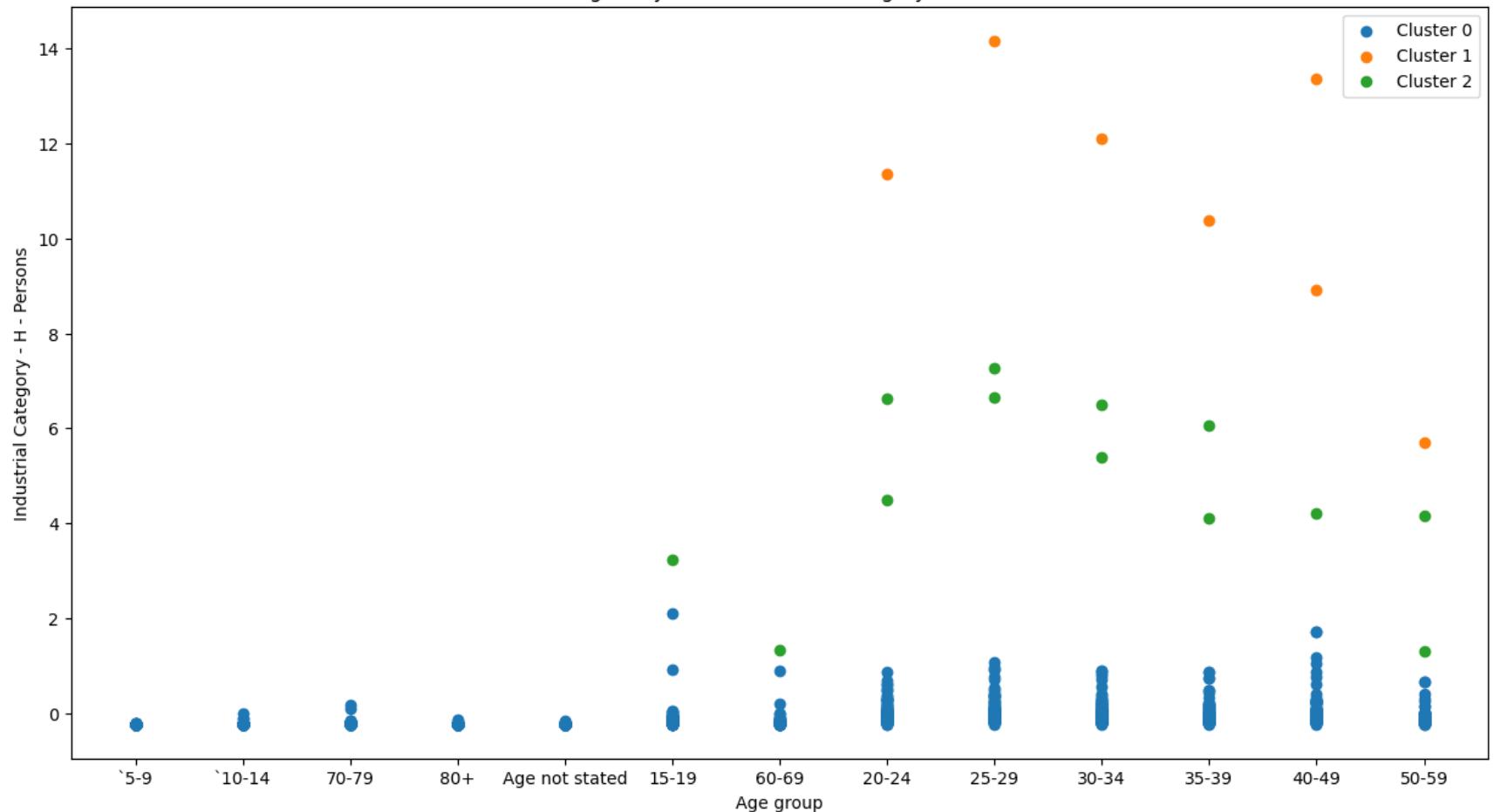
Clustering Analysis for Industrial Category - G - Non HHI - Males



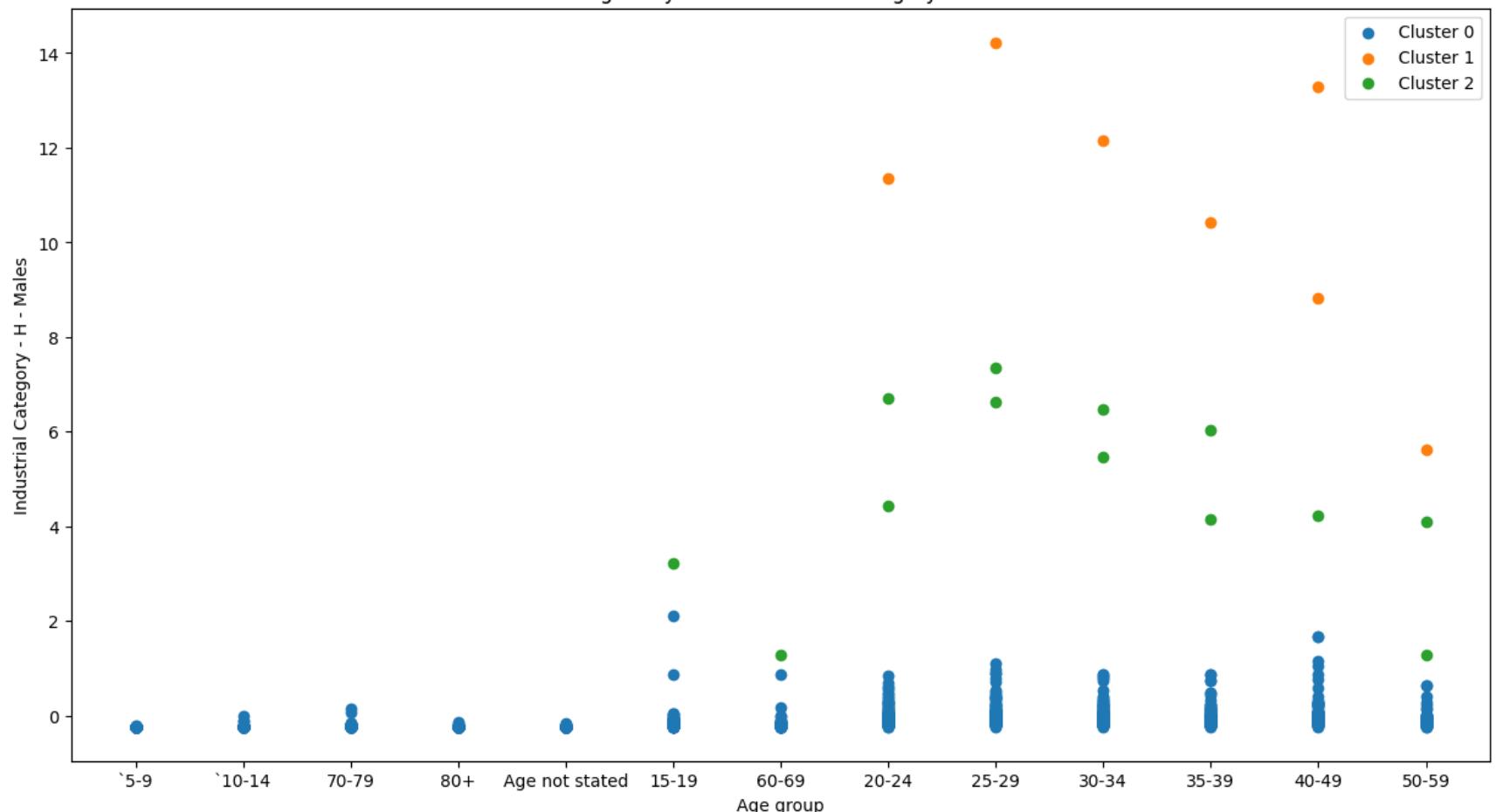
Clustering Analysis for Industrial Category - G - Non HHI - Females



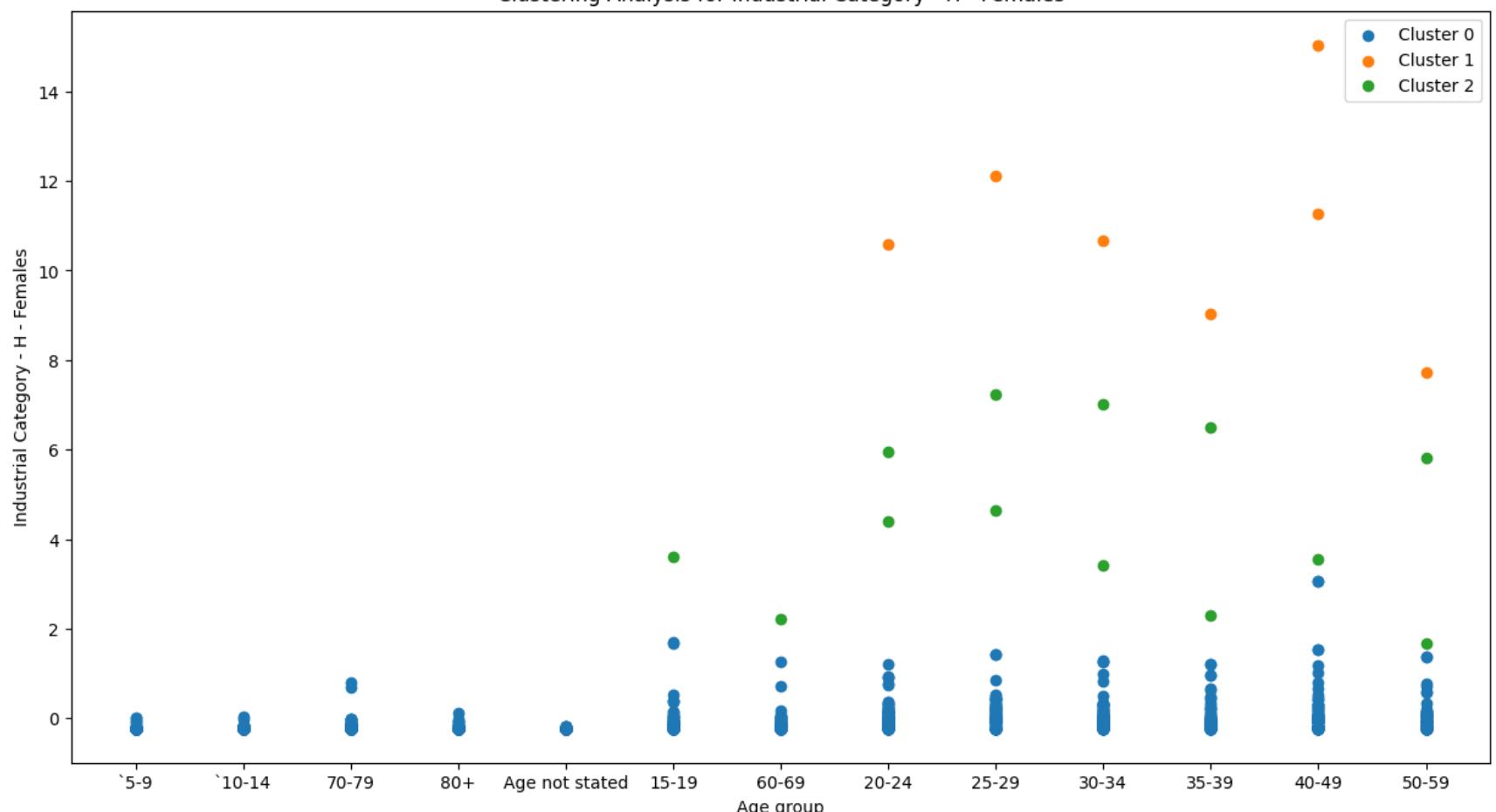
Clustering Analysis for Industrial Category - H - Persons



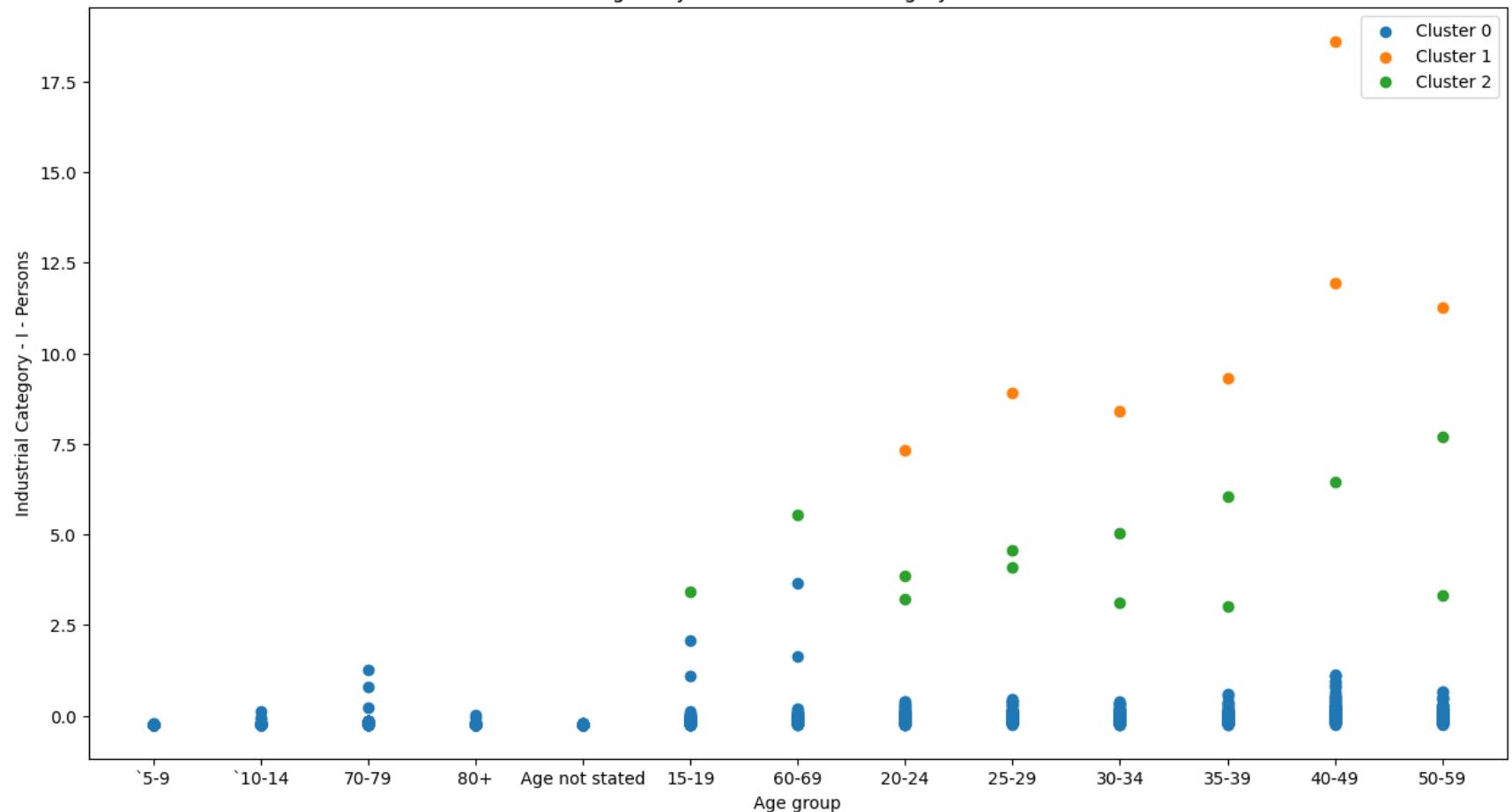
Clustering Analysis for Industrial Category - H - Males



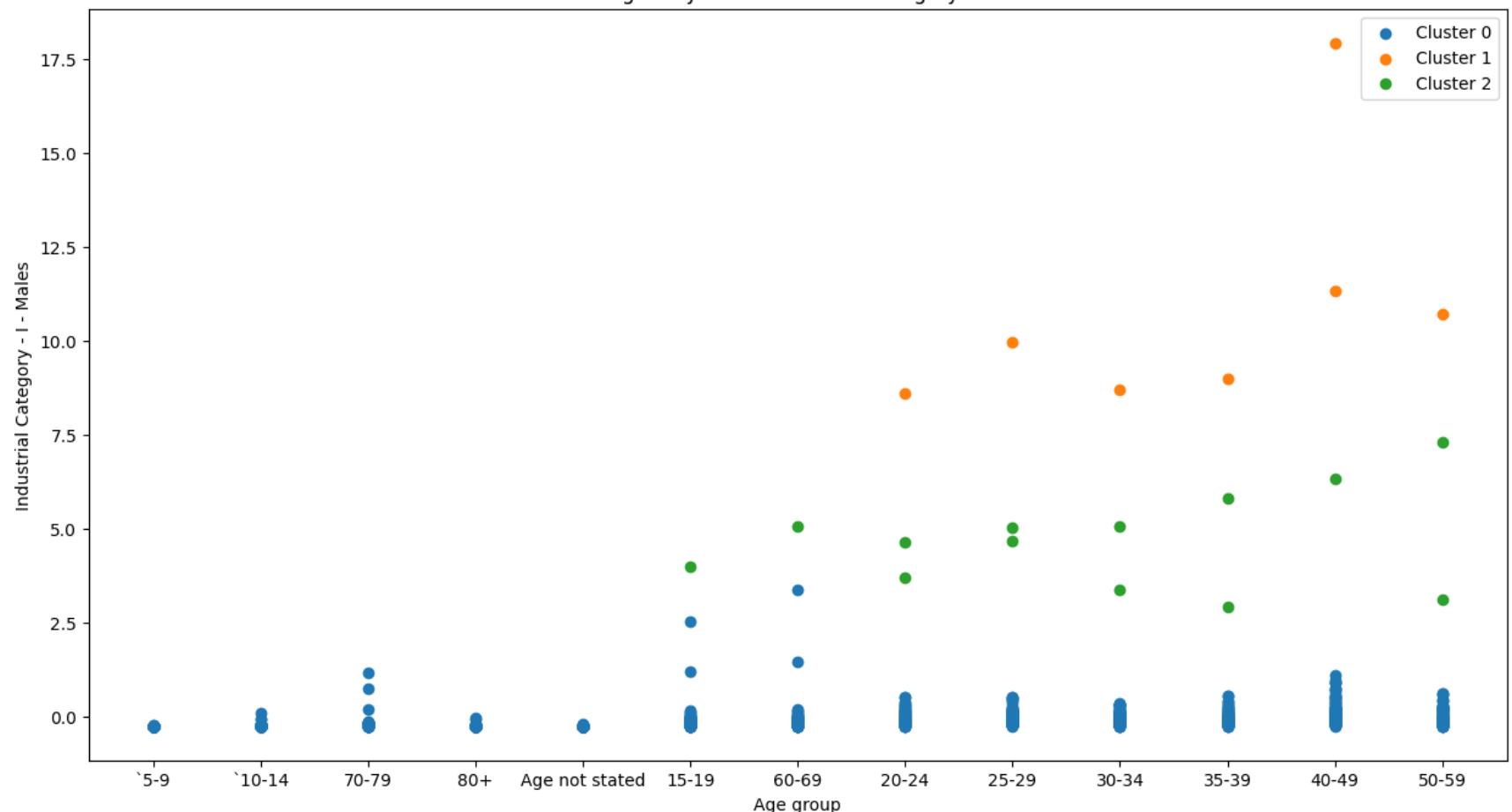
Clustering Analysis for Industrial Category - H - Females



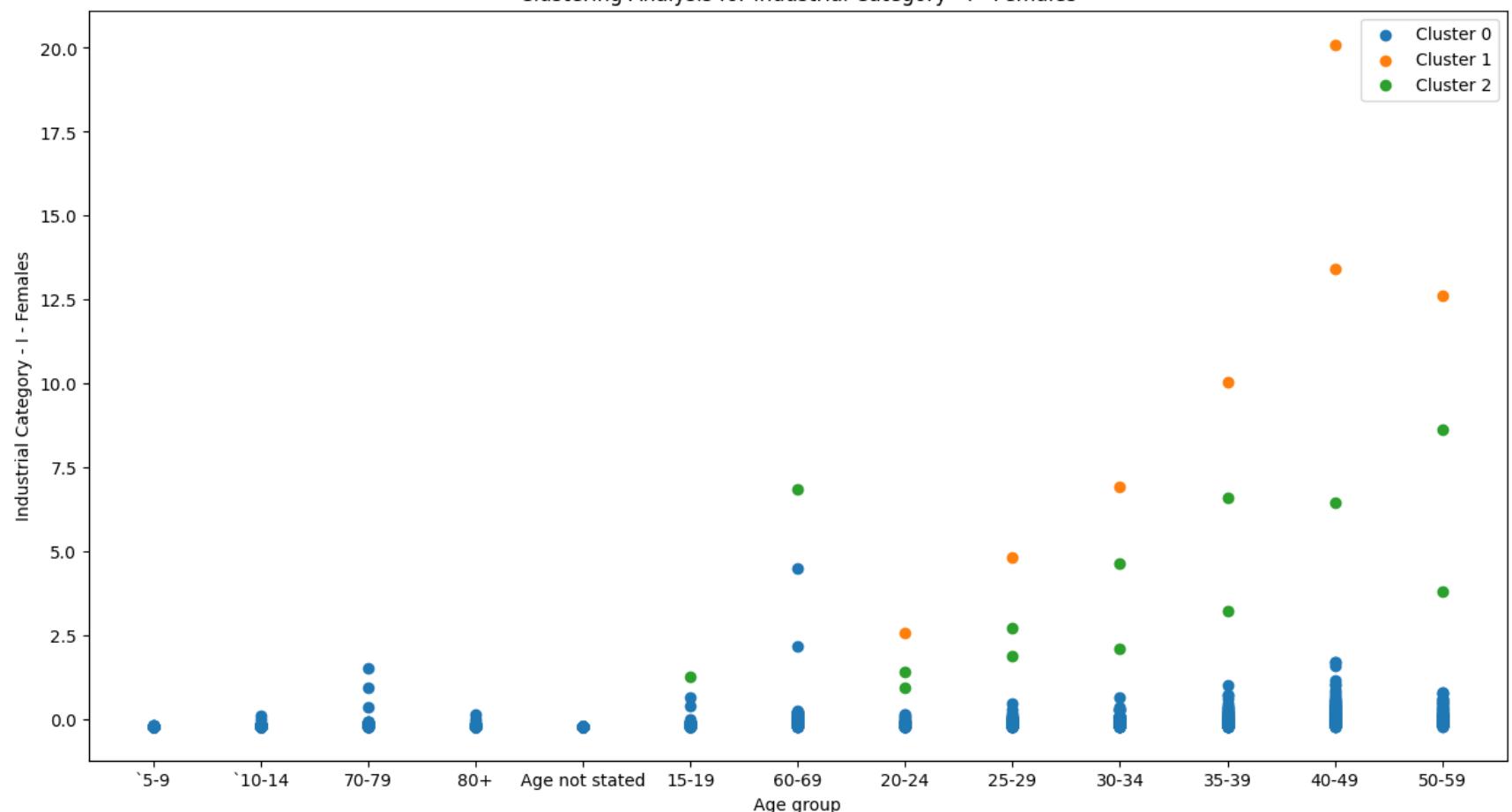
Clustering Analysis for Industrial Category - I - Persons



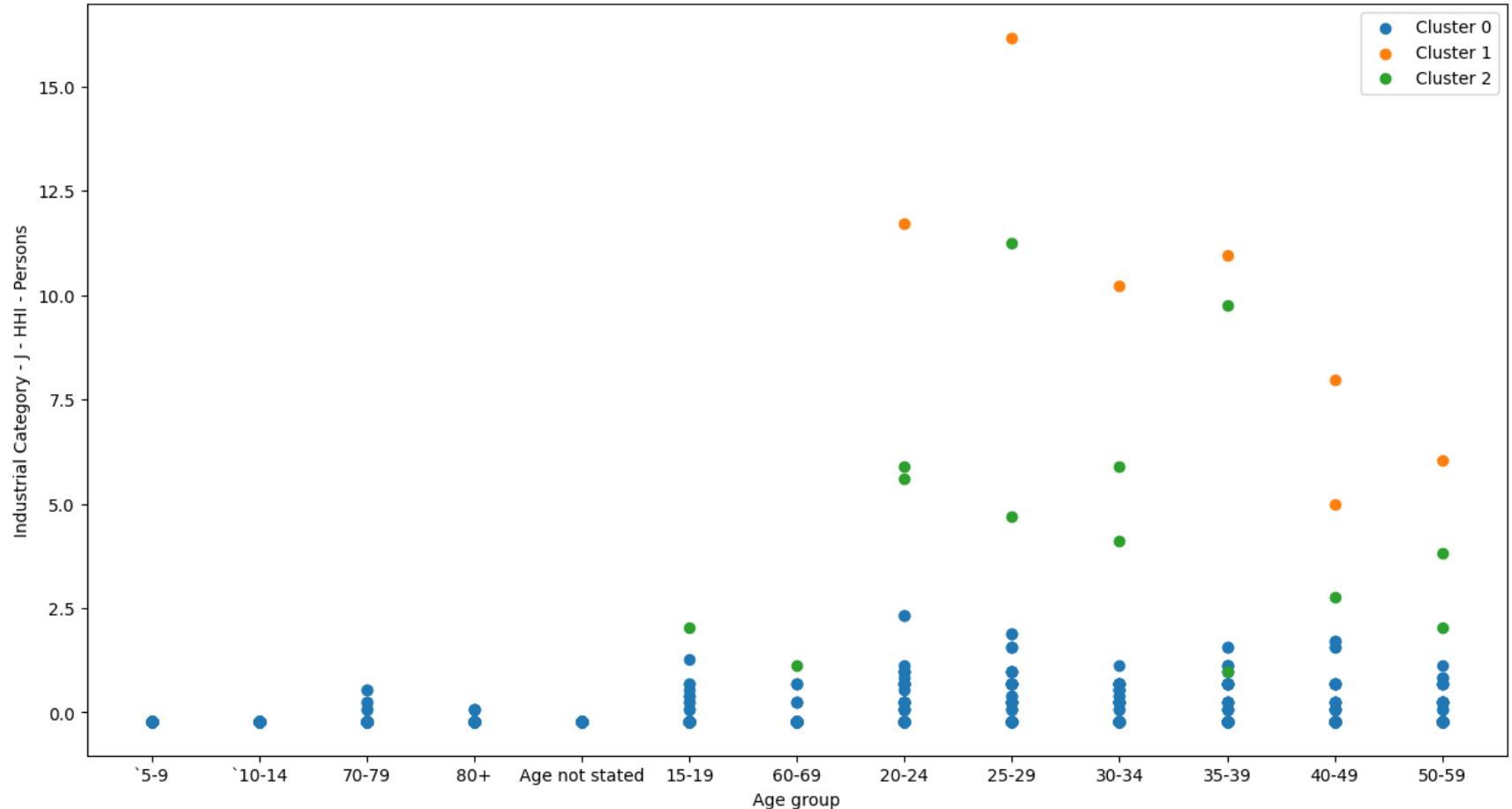
Clustering Analysis for Industrial Category - I - Males



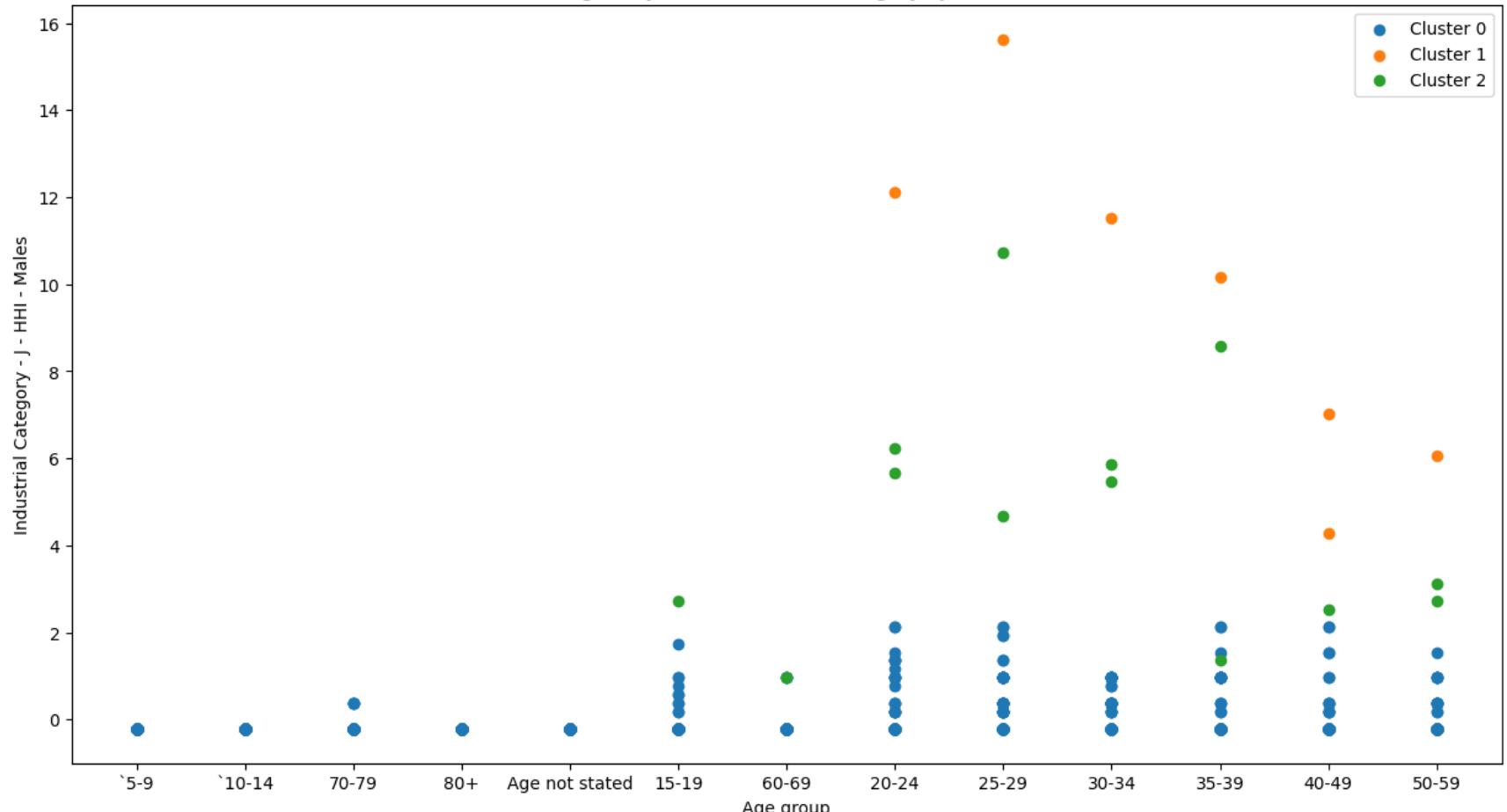
Clustering Analysis for Industrial Category - I - Females



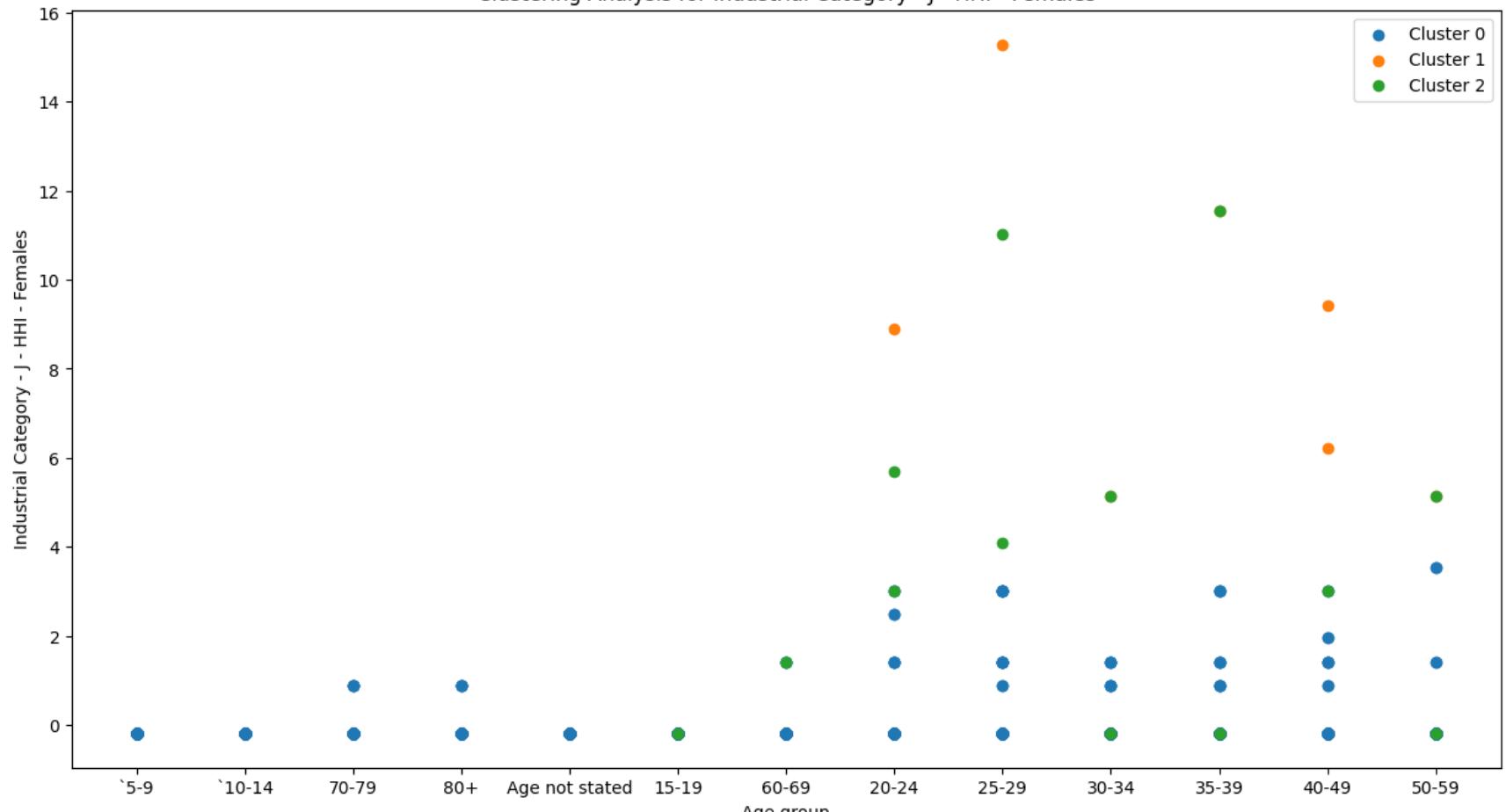
Clustering Analysis for Industrial Category - J - HHI - Persons



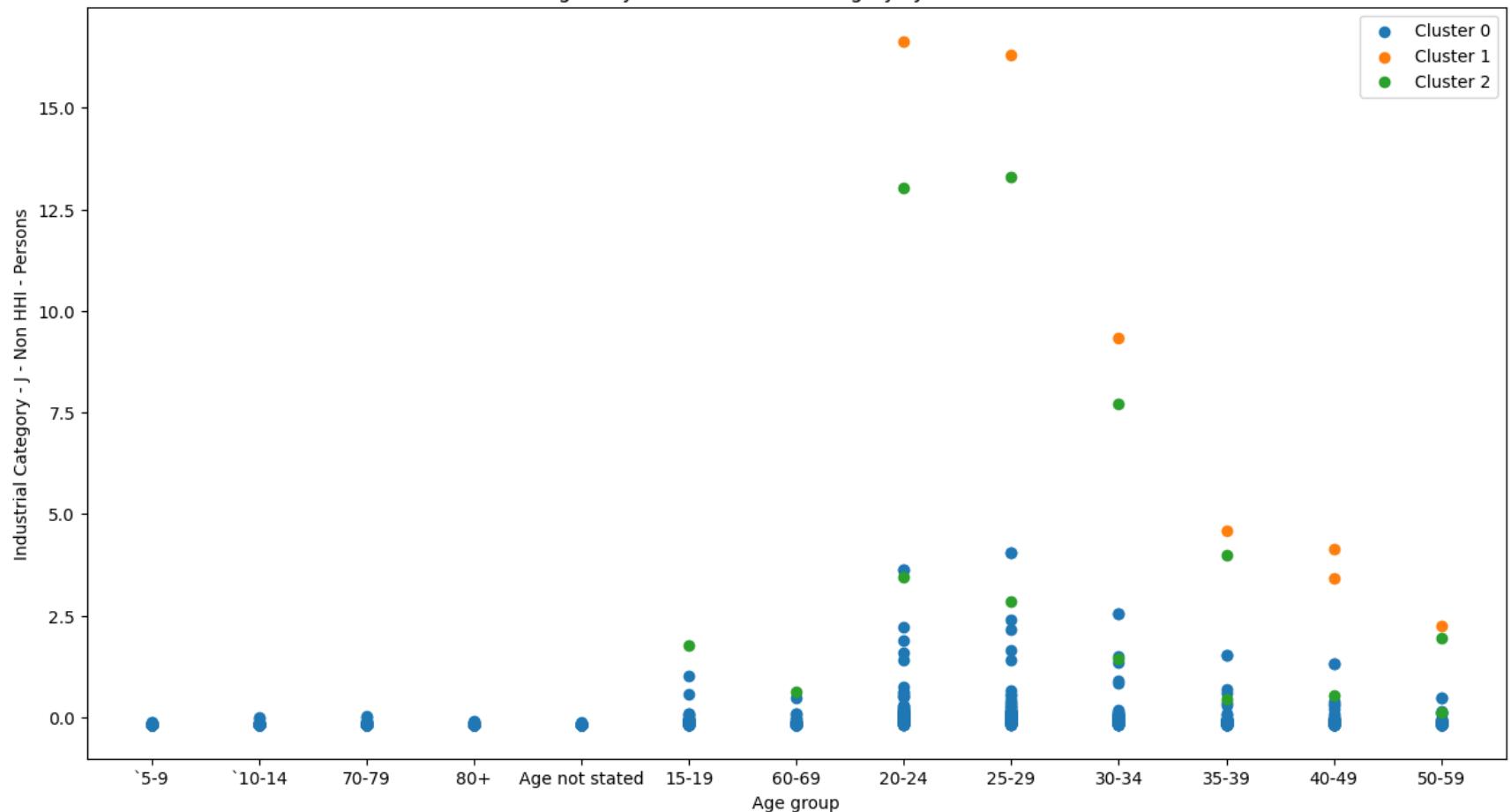
Clustering Analysis for Industrial Category - J - HHI - Males



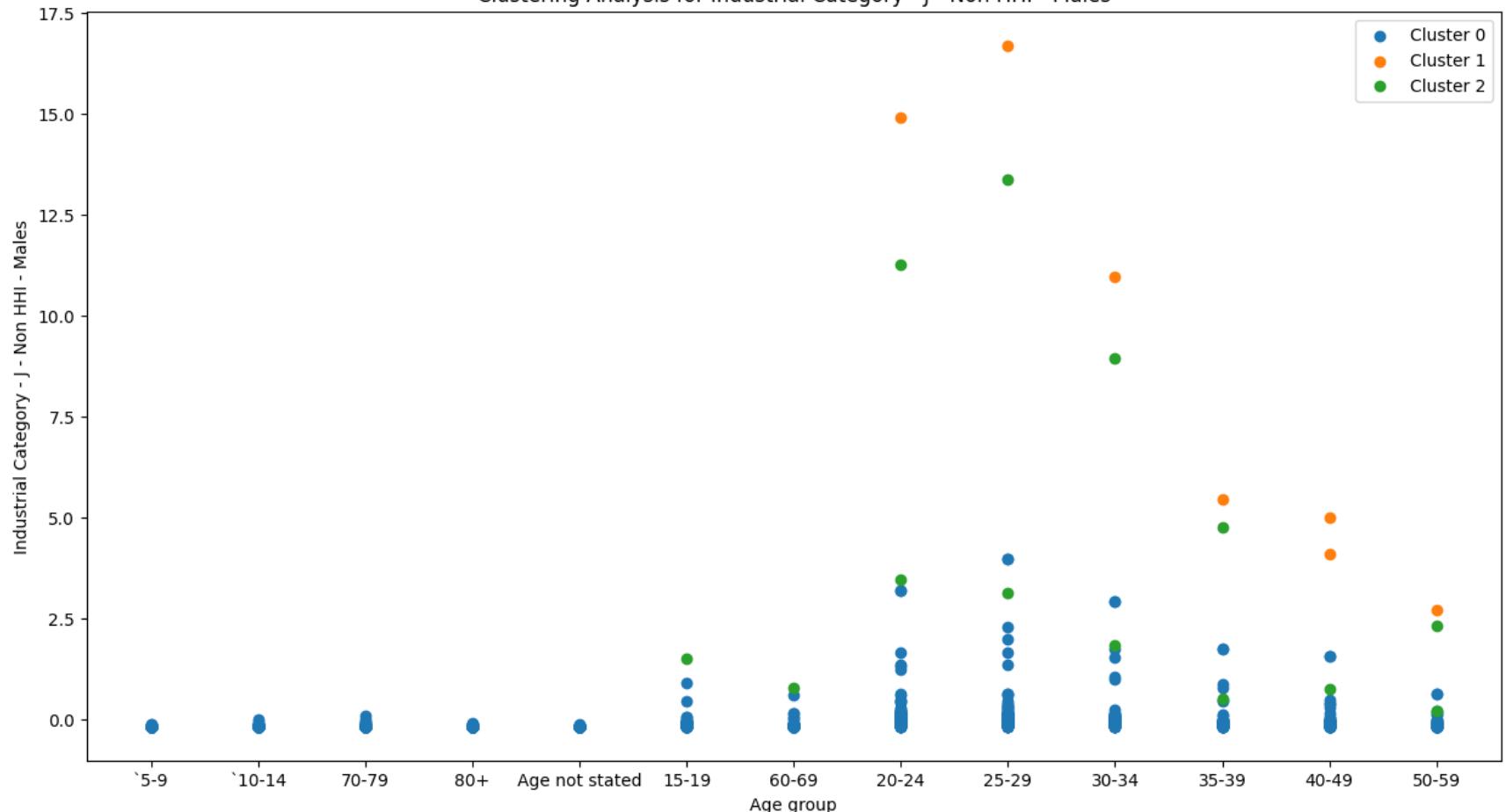
Clustering Analysis for Industrial Category - J - HHI - Females



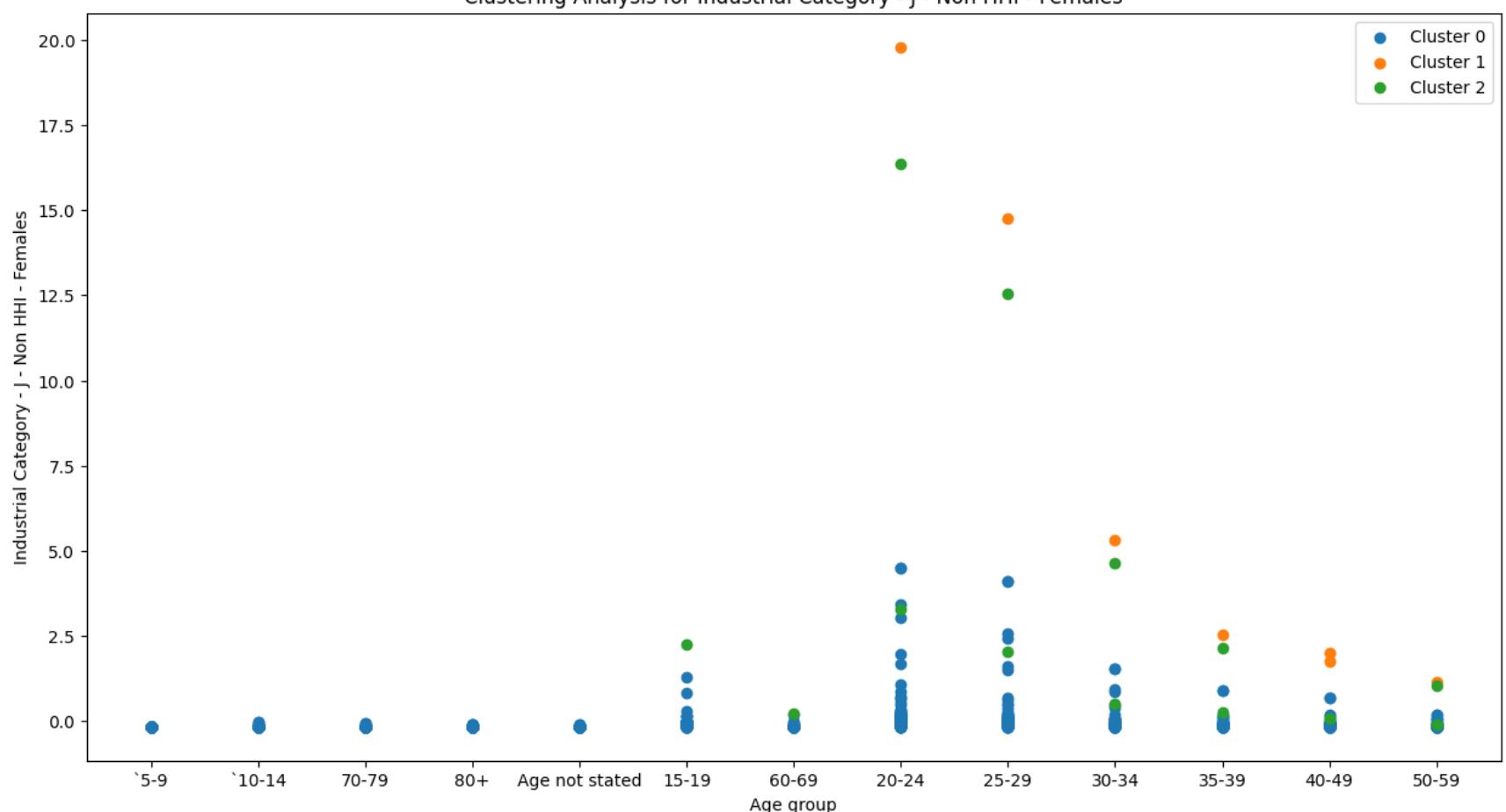
Clustering Analysis for Industrial Category - J - Non HHI - Persons



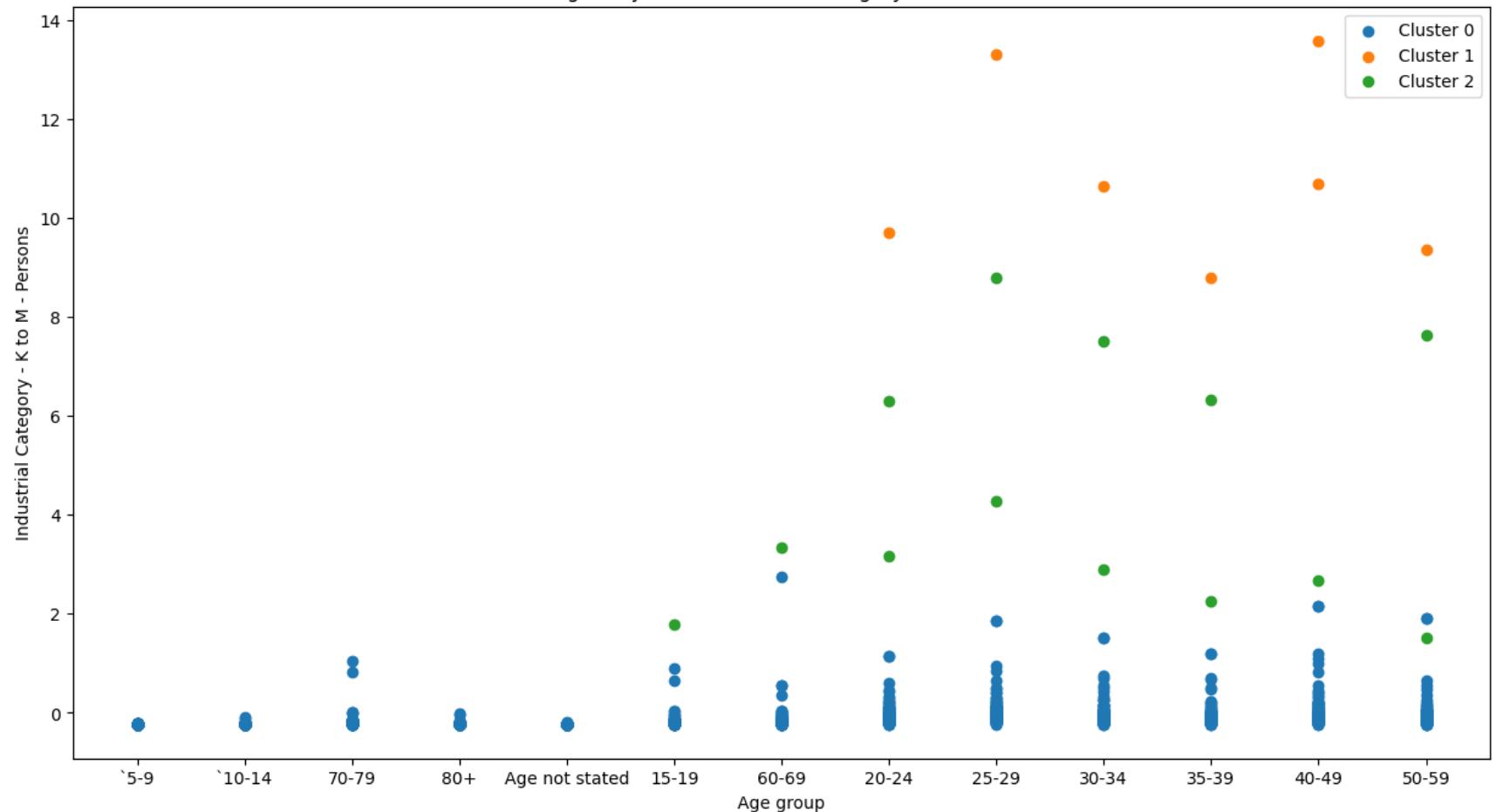
Clustering Analysis for Industrial Category - J - Non HHI - Males



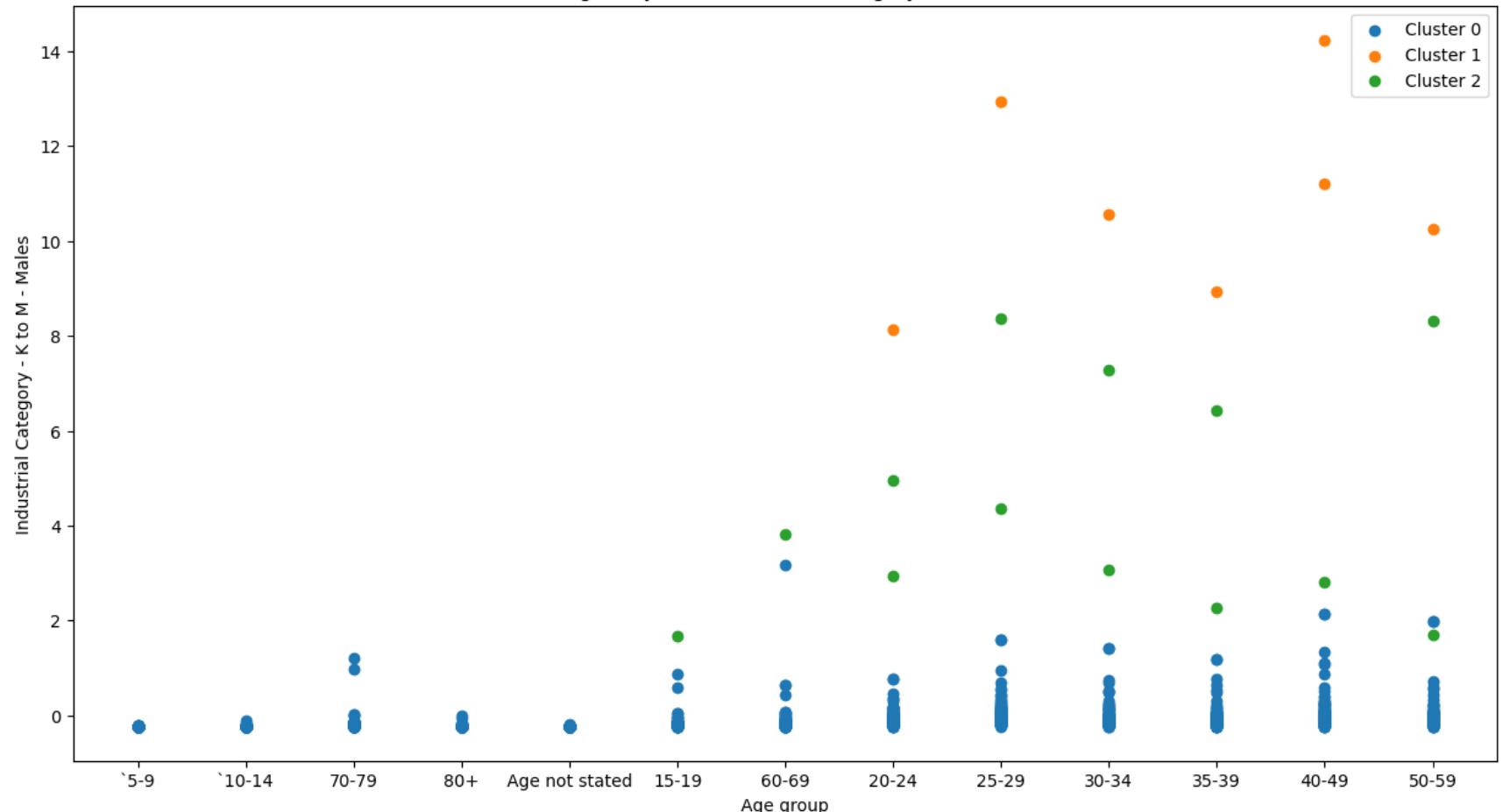
Clustering Analysis for Industrial Category - J - Non HHI - Females



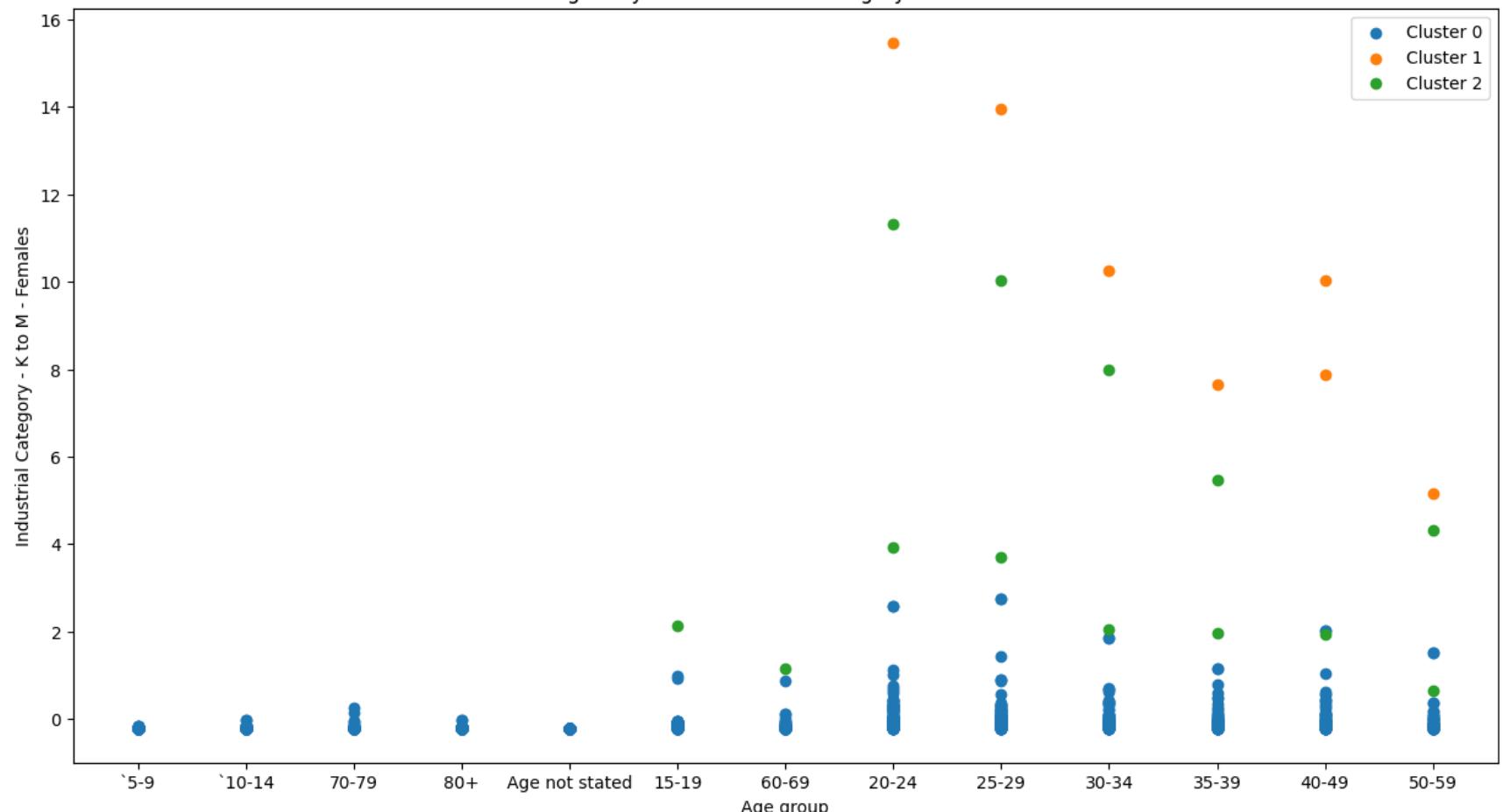
Clustering Analysis for Industrial Category - K to M - Persons



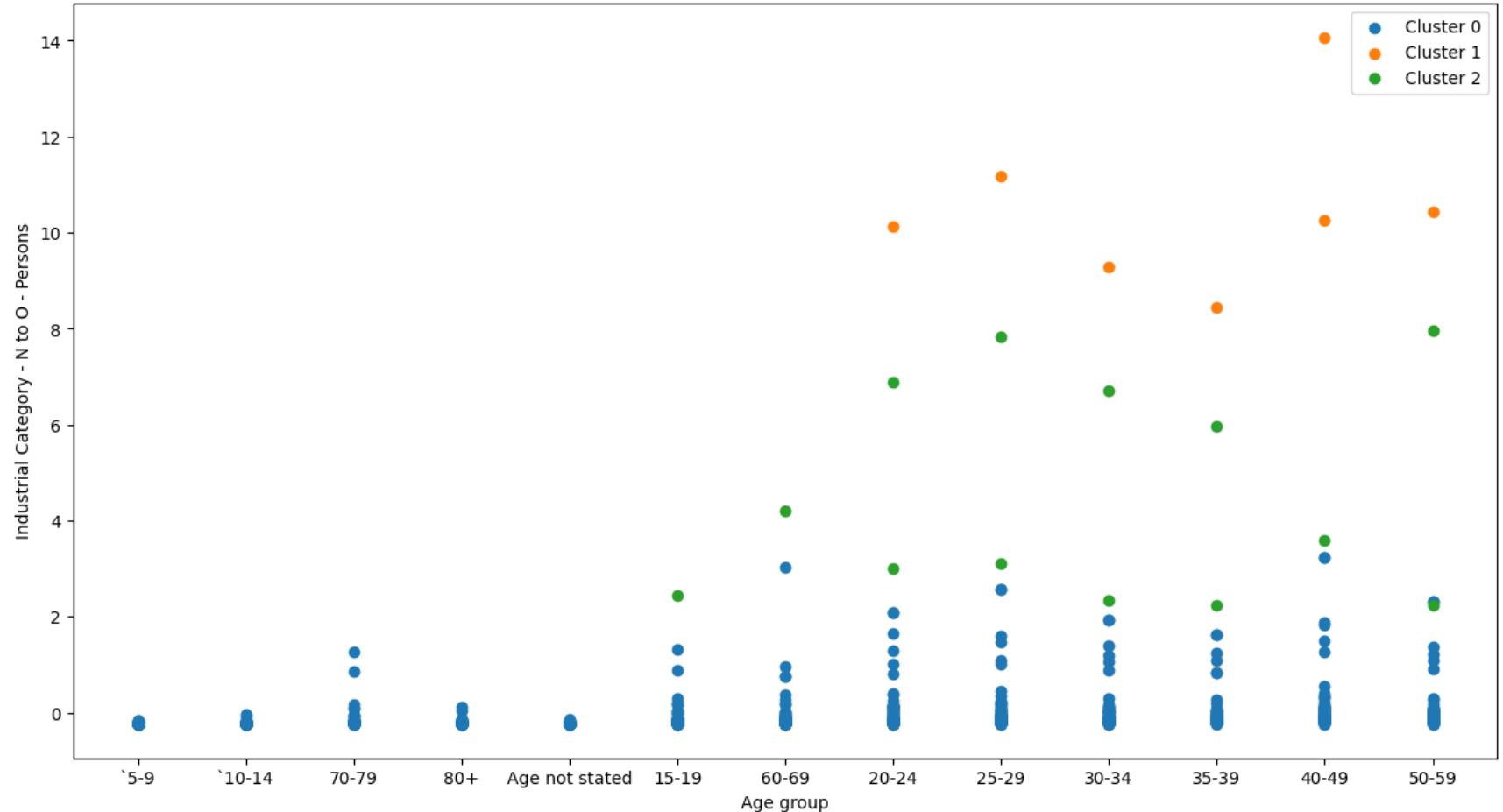
Clustering Analysis for Industrial Category - K to M - Males



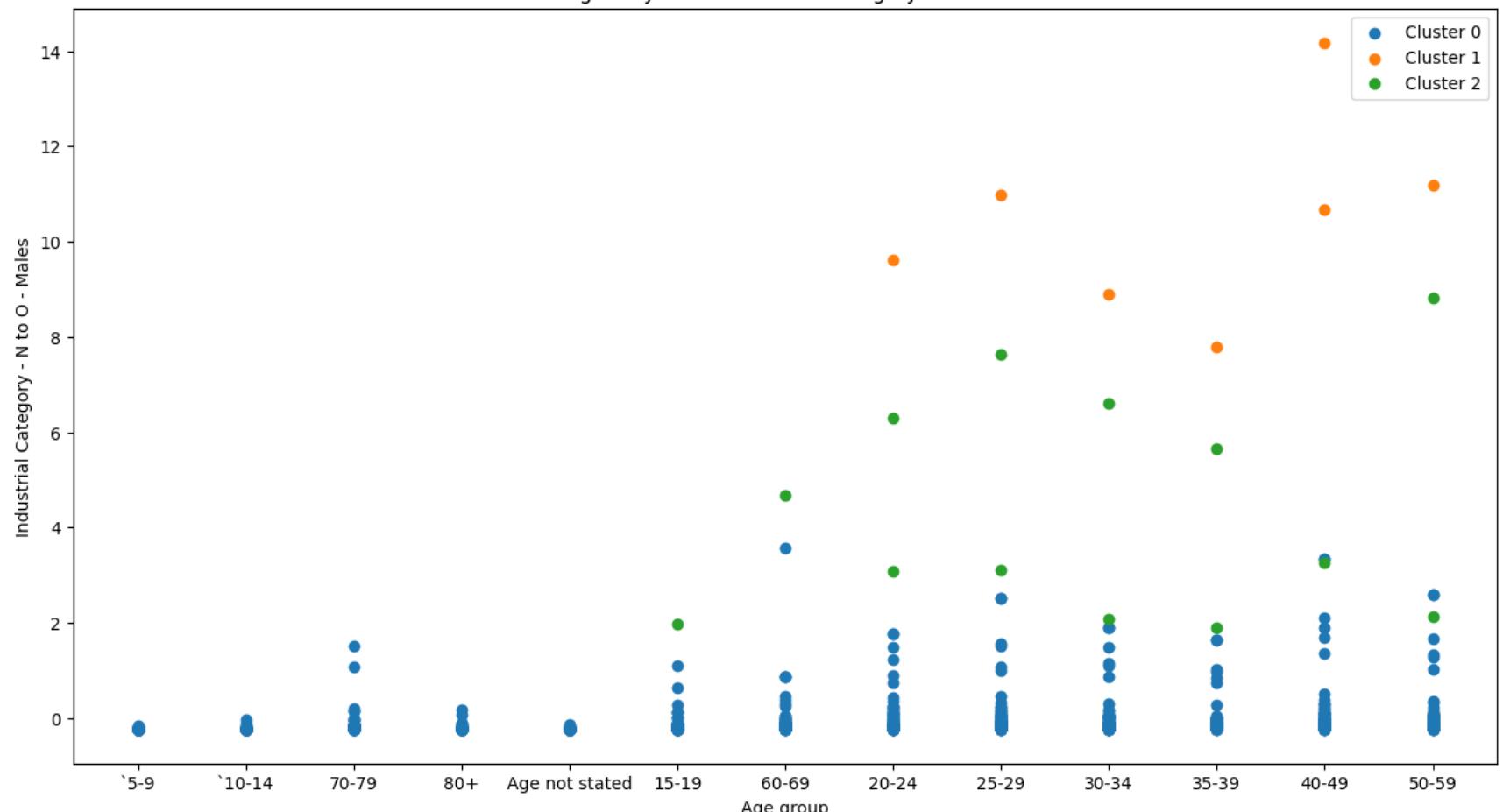
Clustering Analysis for Industrial Category - K to M - Females



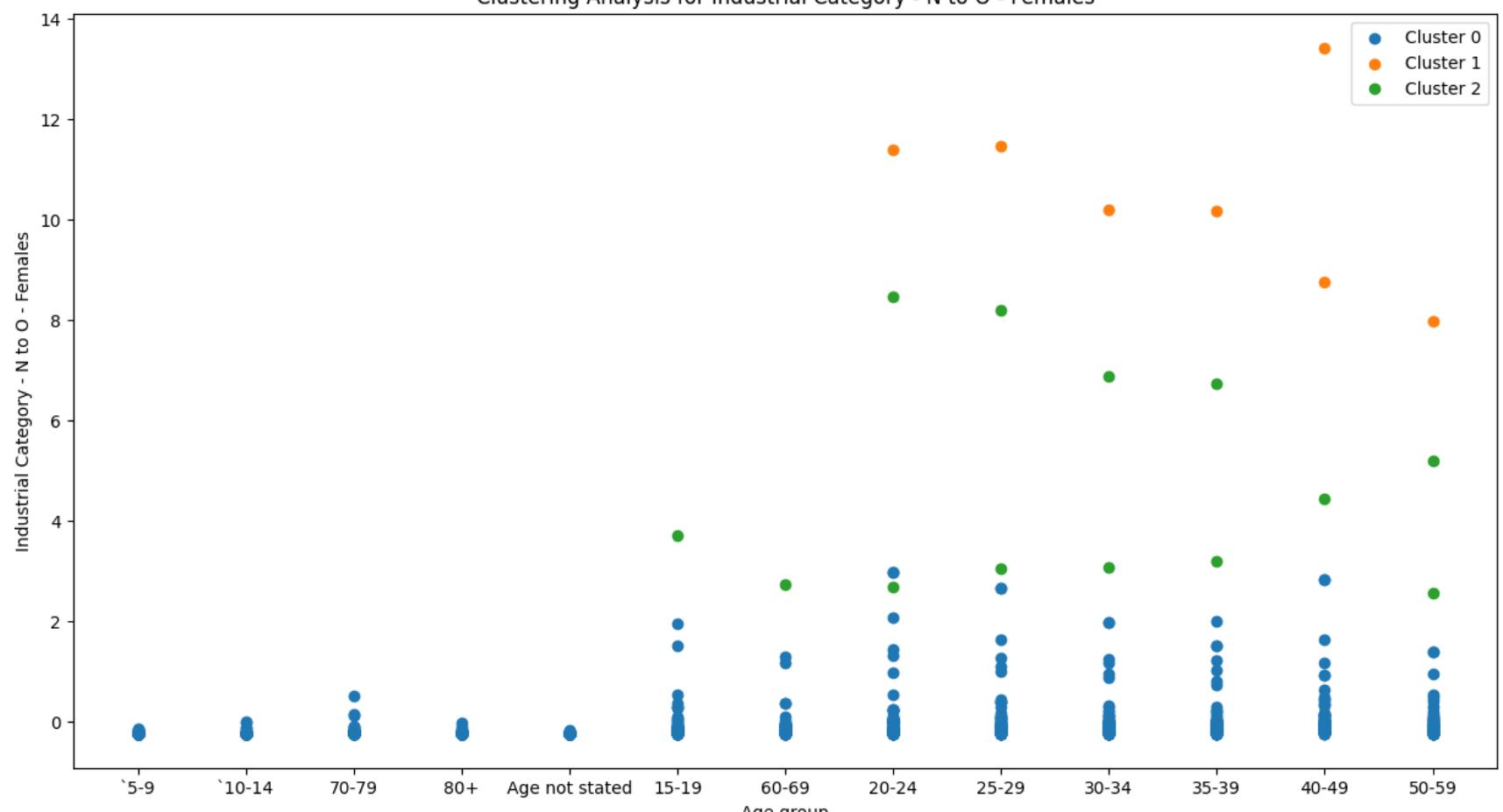
Clustering Analysis for Industrial Category - N to O - Persons



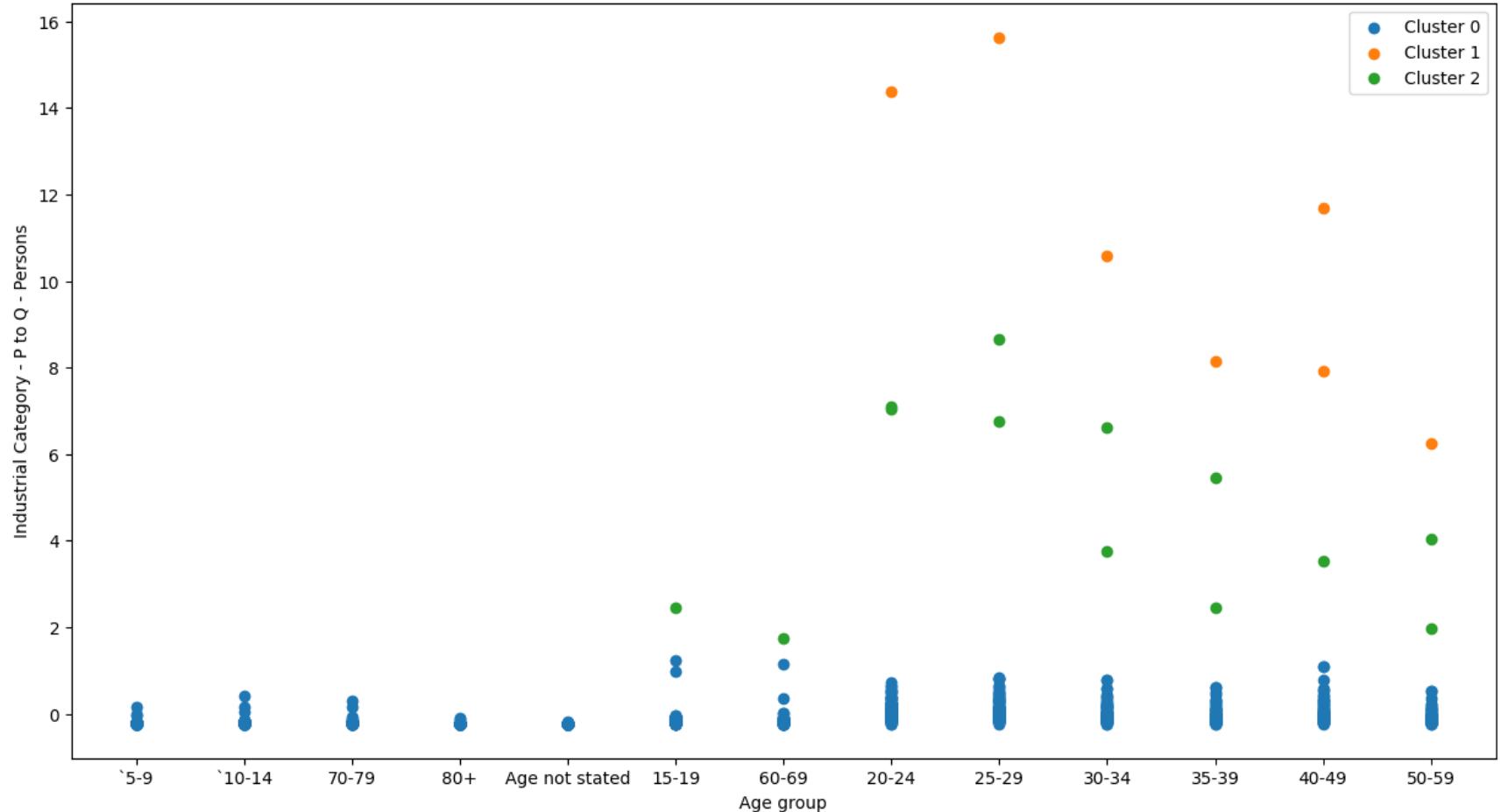
Clustering Analysis for Industrial Category - N to O - Males



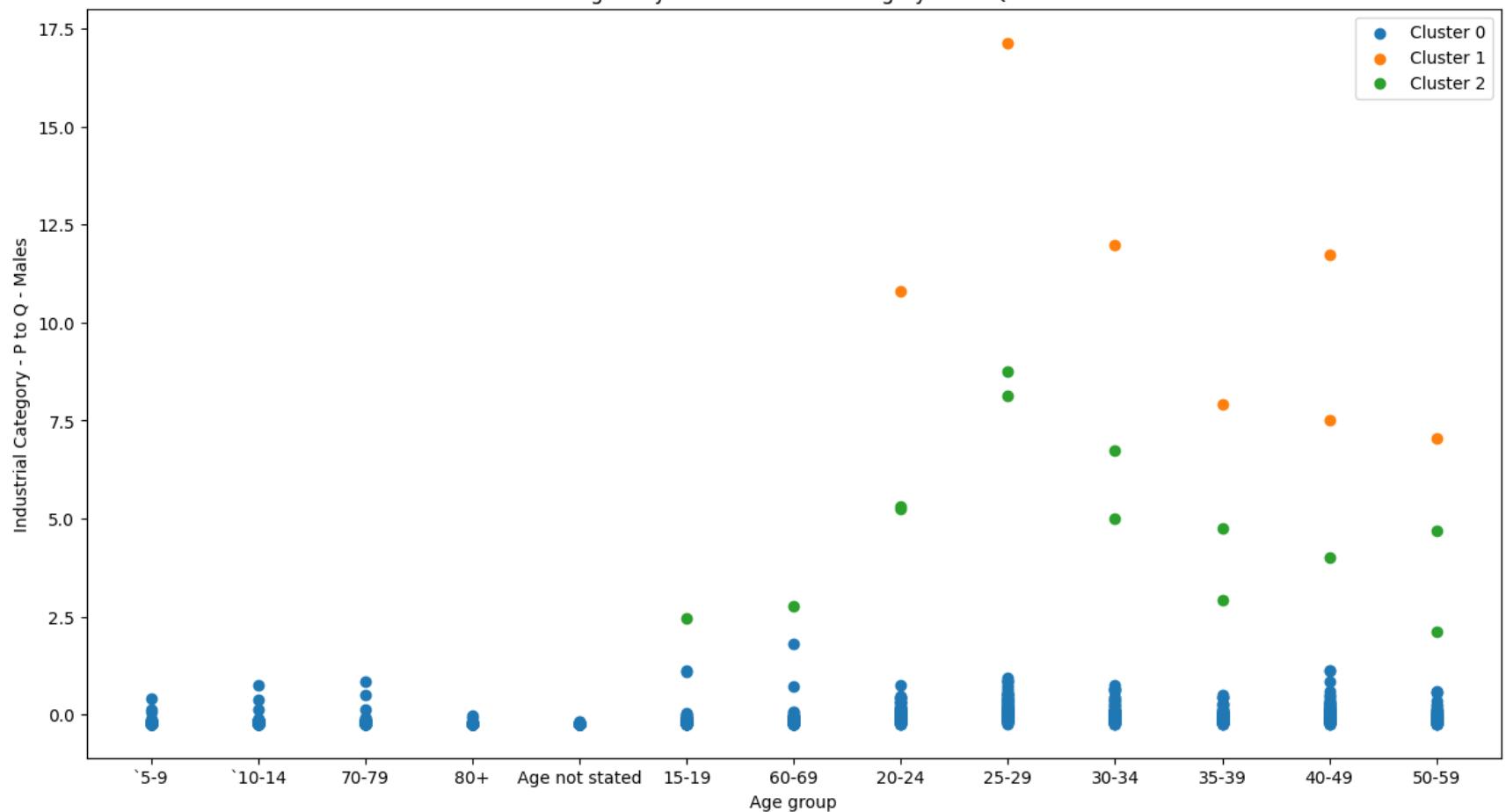
Clustering Analysis for Industrial Category - N to O - Females



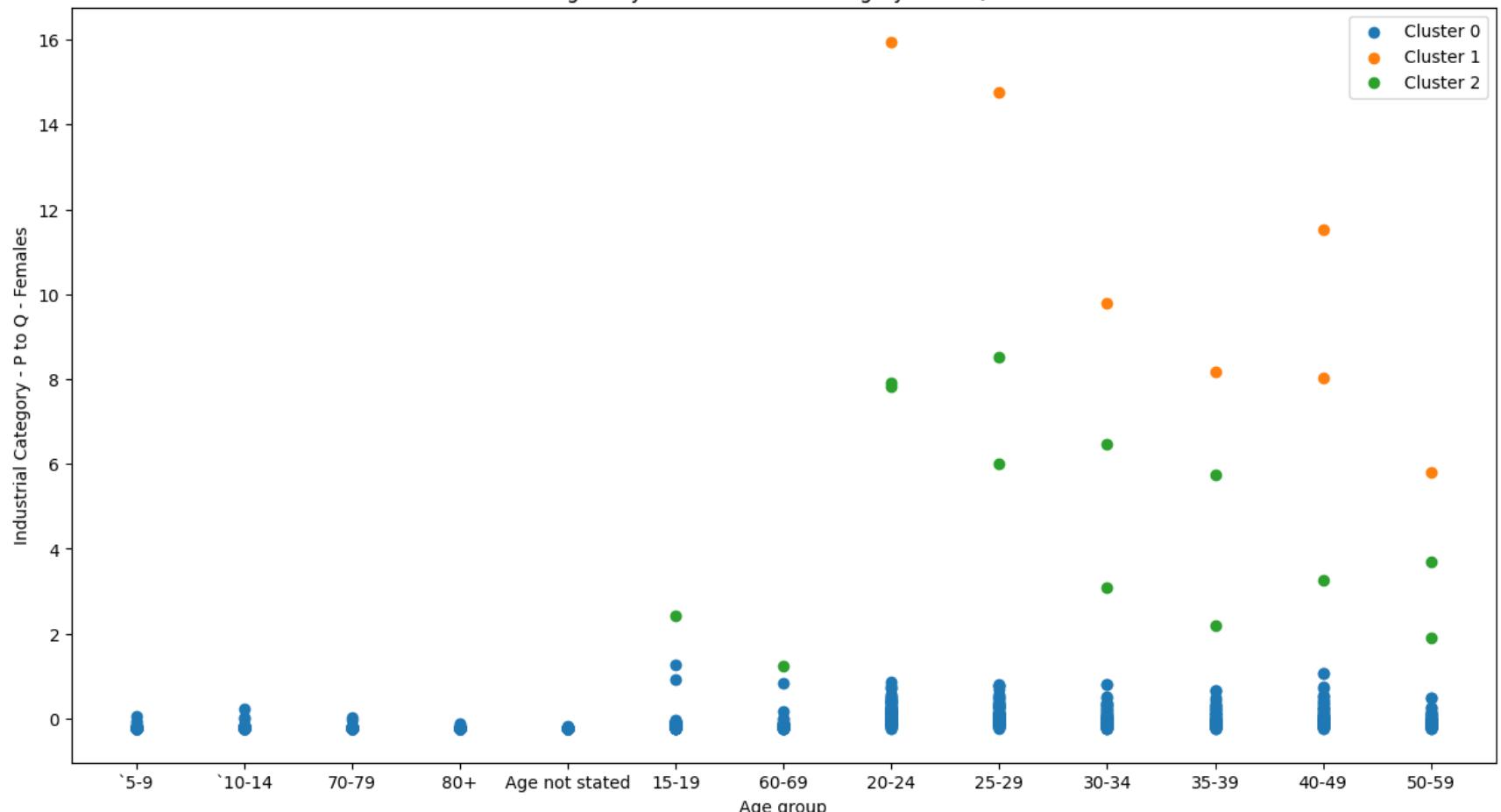
Clustering Analysis for Industrial Category - P to Q - Persons



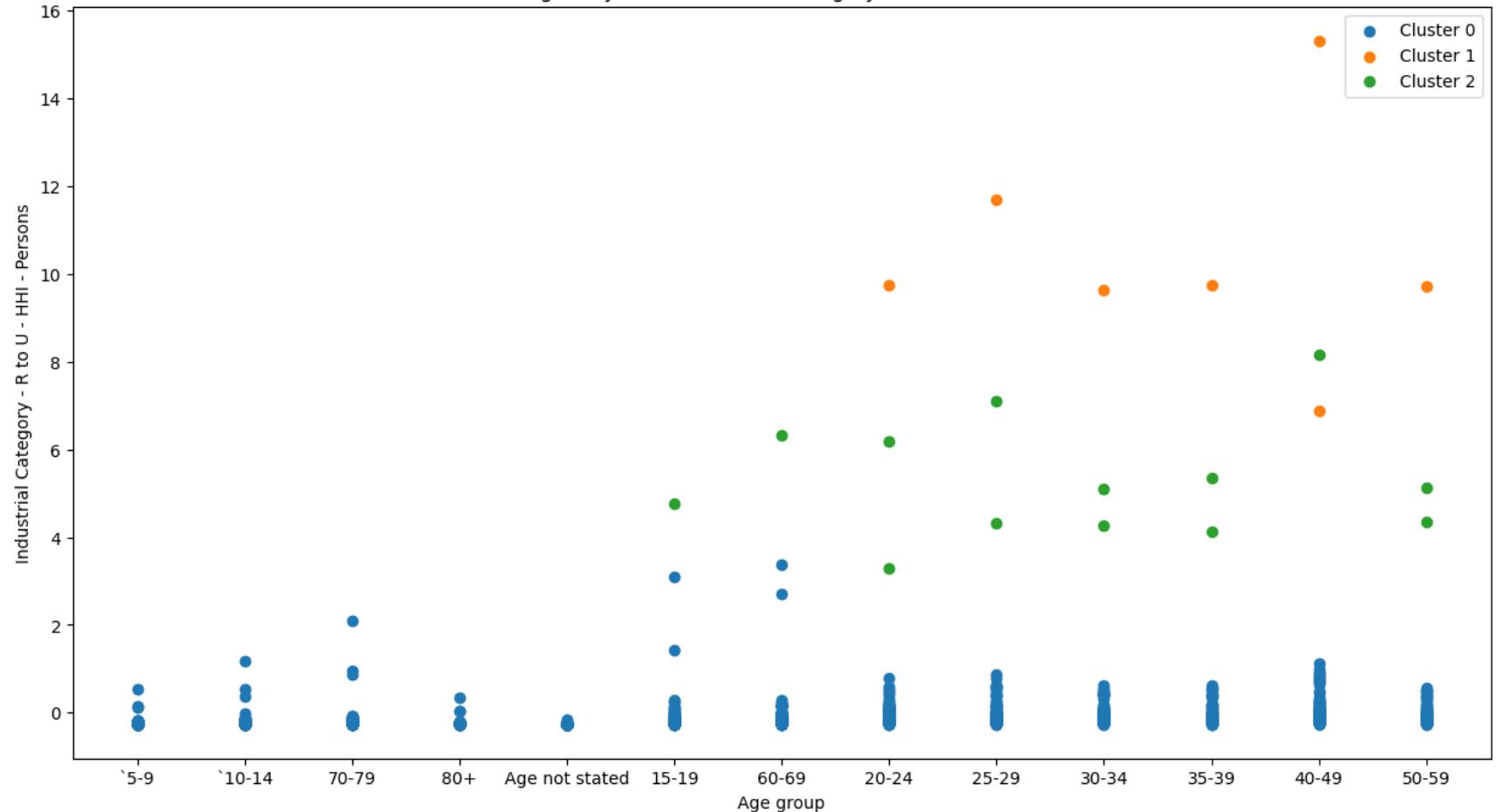
Clustering Analysis for Industrial Category - P to Q - Males



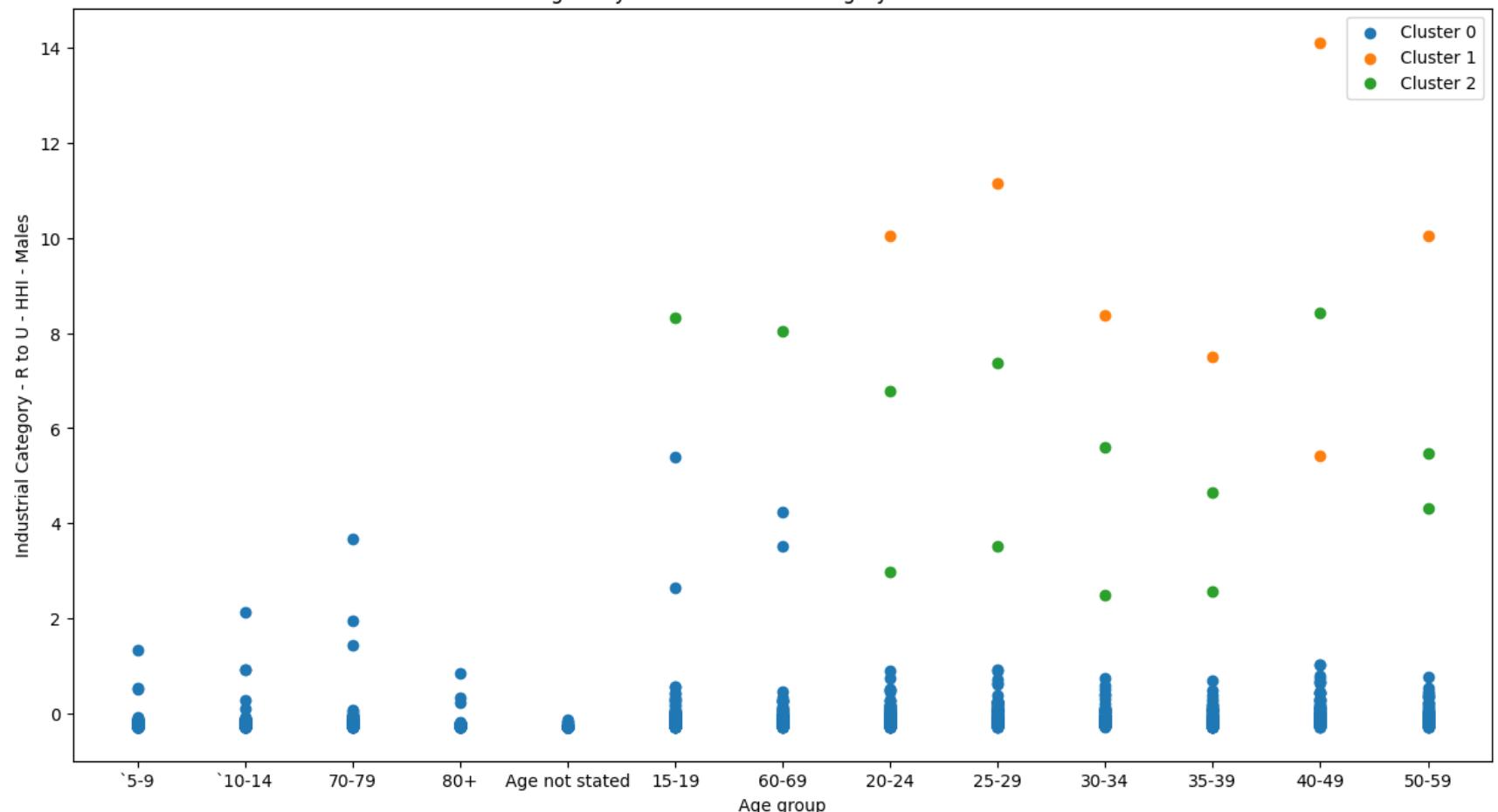
Clustering Analysis for Industrial Category - P to Q - Females



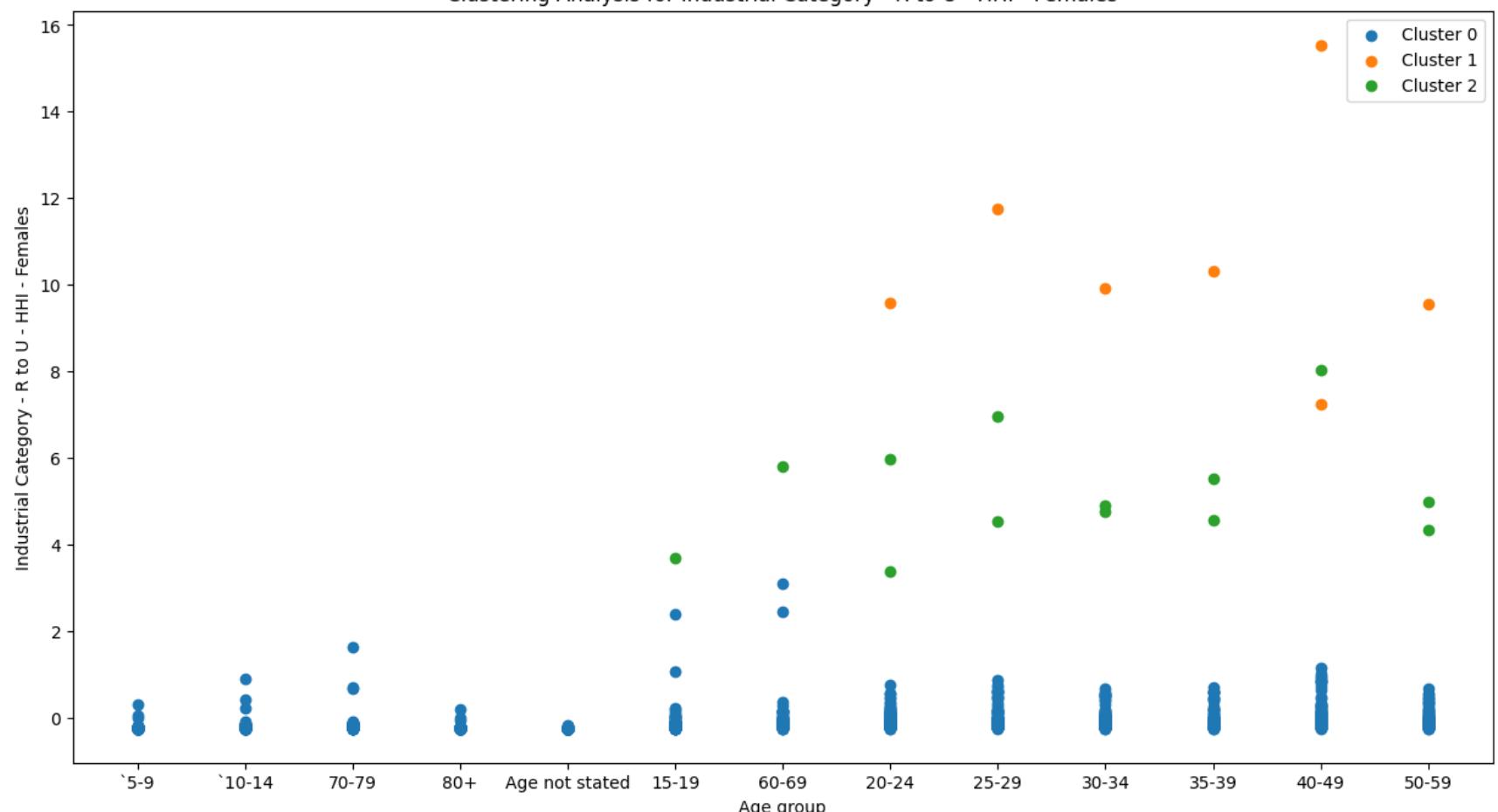
Clustering Analysis for Industrial Category - R to U - HHI - Persons



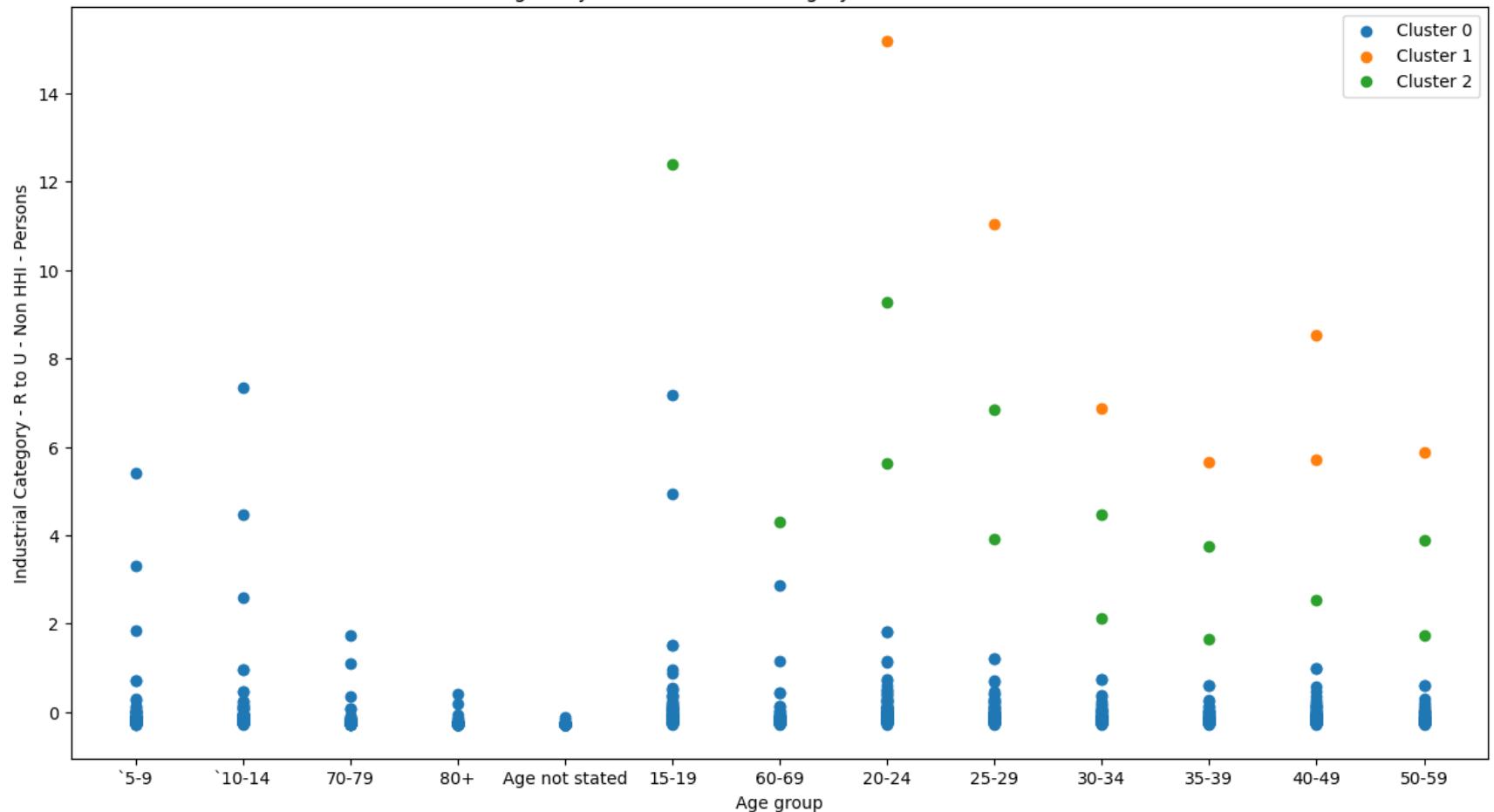
Clustering Analysis for Industrial Category - R to U - HHI - Males



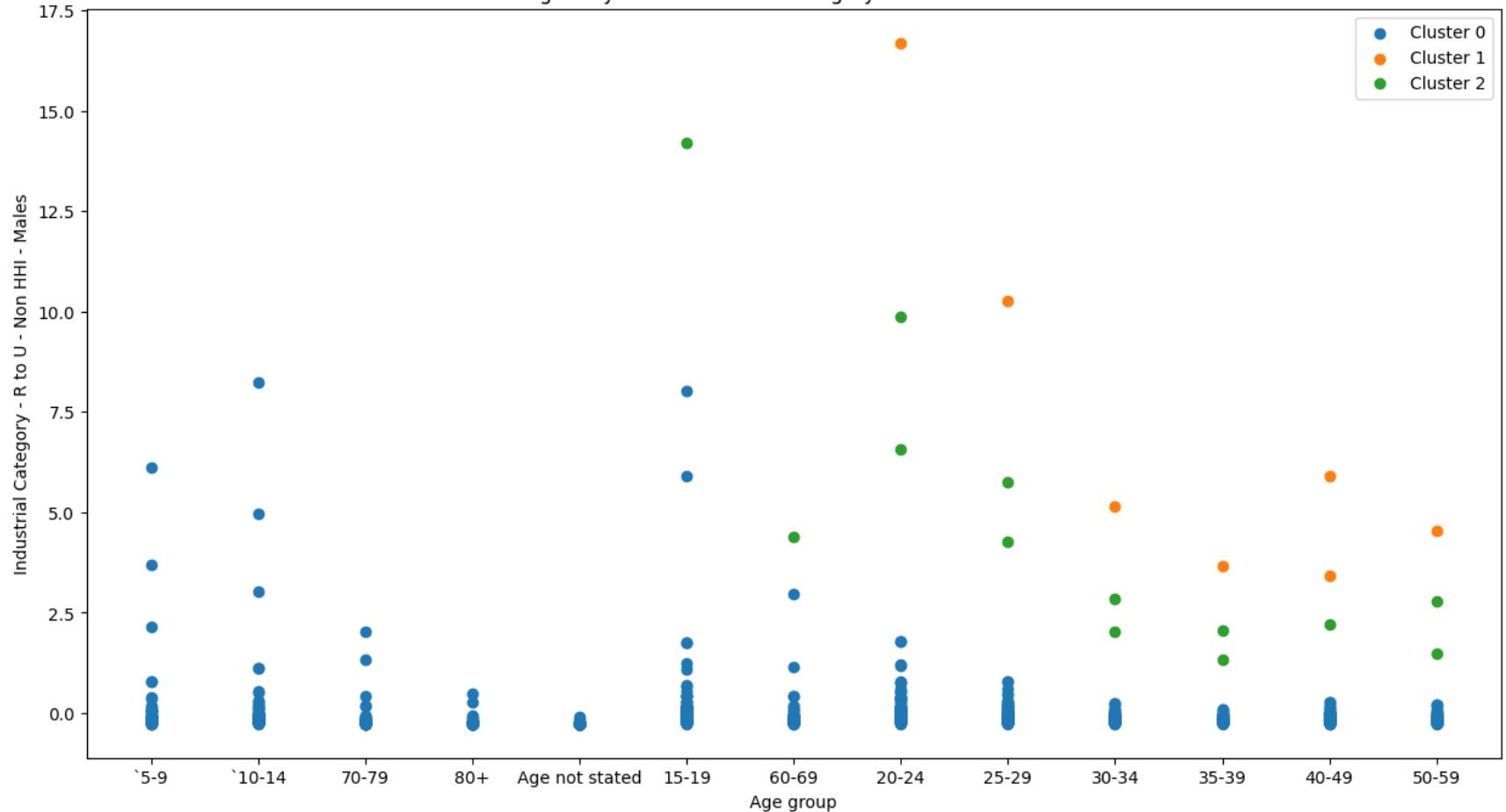
Clustering Analysis for Industrial Category - R to U - HHI - Females



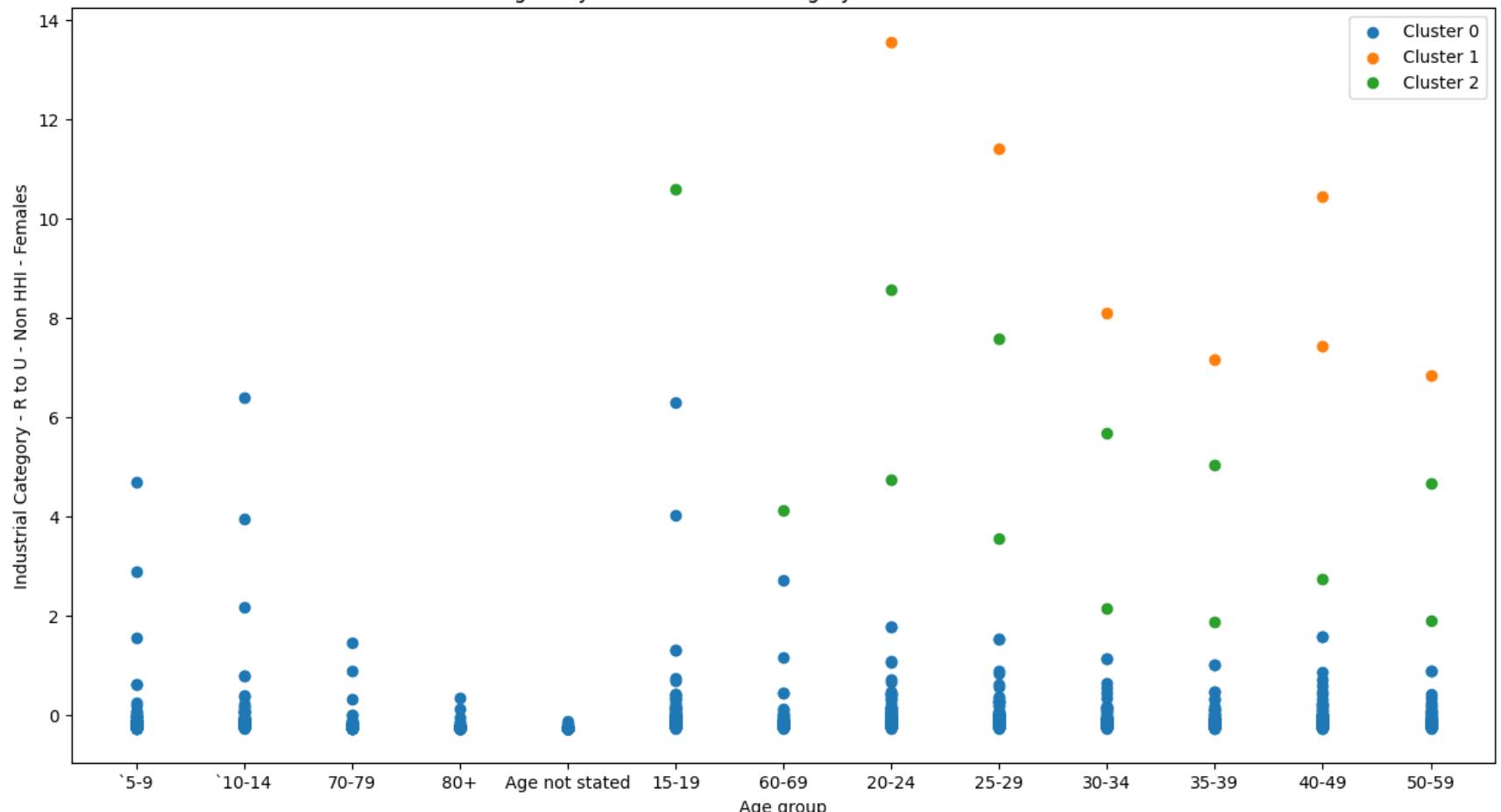
Clustering Analysis for Industrial Category - R to U - Non HHI - Persons



Clustering Analysis for Industrial Category - R to U - Non HHI - Males



Clustering Analysis for Industrial Category - R to U - Non HHI - Females



```
In [ ]: 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 - Kanniakumari' 'District - Dharmapuri'
'District - Krishnagiri' 'District - Coimbatore' 'District - Tiruppur']
```

In []: df.columns

```
Out[ ]: 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', 'Cluster'],
dtype='object')
```

In [6]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Assuming df is your DataFrame after data cleaning and clustering

```
# List of industrial category columns
industrial_categories = [
    'Industrial Category - A - Cultivators - Persons',
    'Industrial Category - A - Cultivators - Males',
    'Industrial Category - A - Cultivators - Females',
    # Add all other industrial category columns here
]
```

```
# List of district names
districts = df['Area Name'].unique()
```

```
# Loop through districts
```

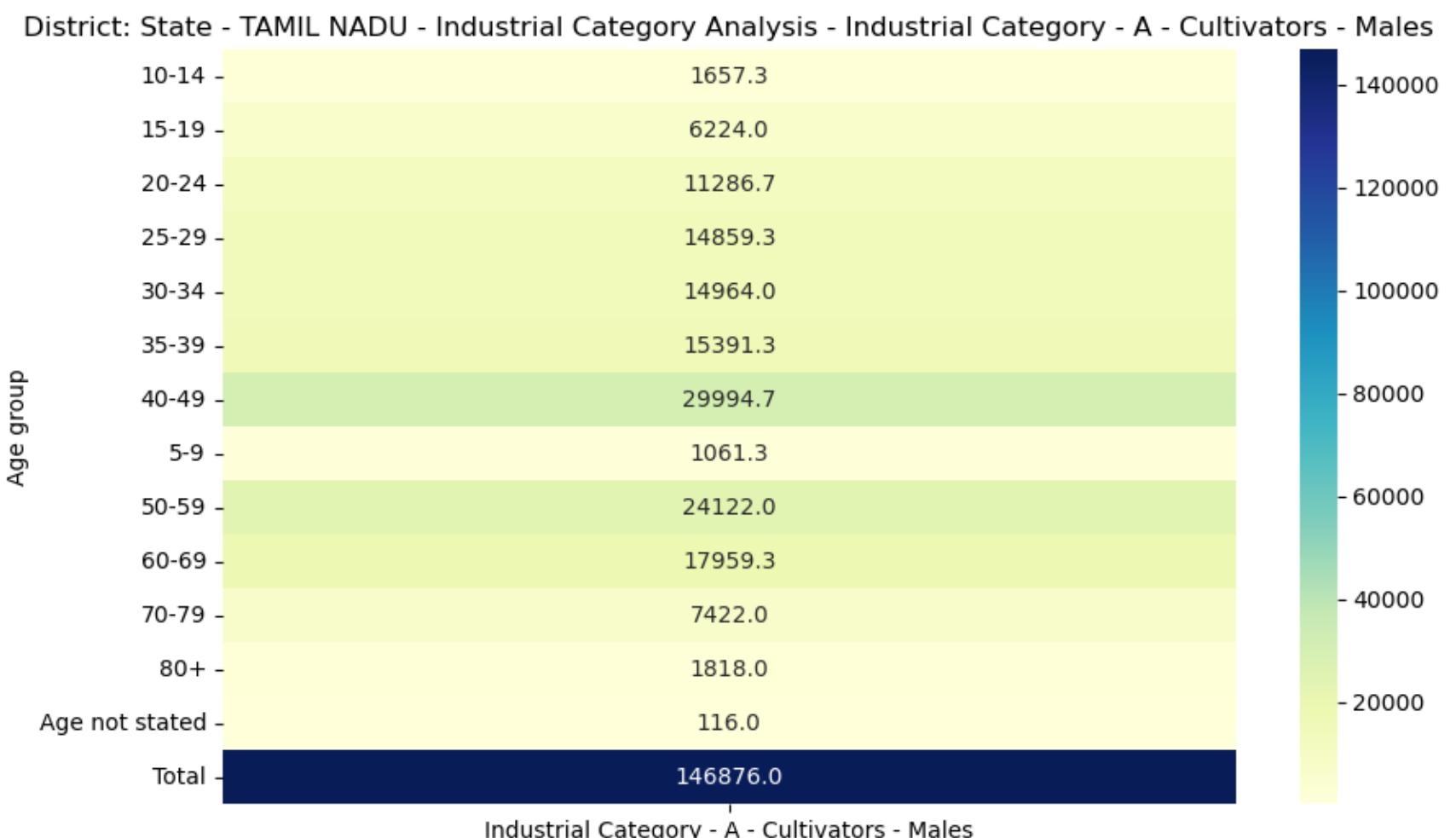
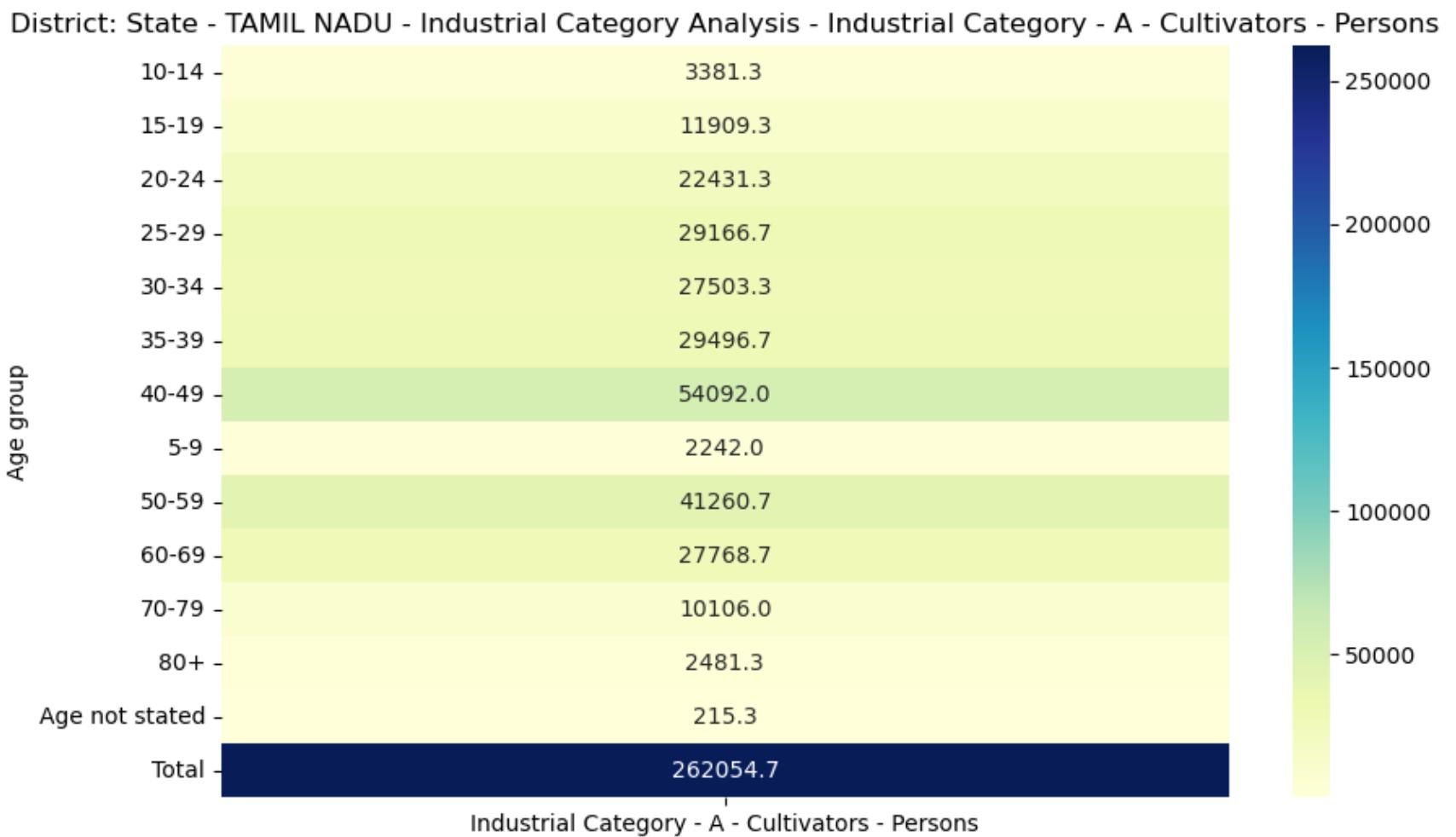
```

for district in districts:
    district_df = df[df['Area Name'] == district]

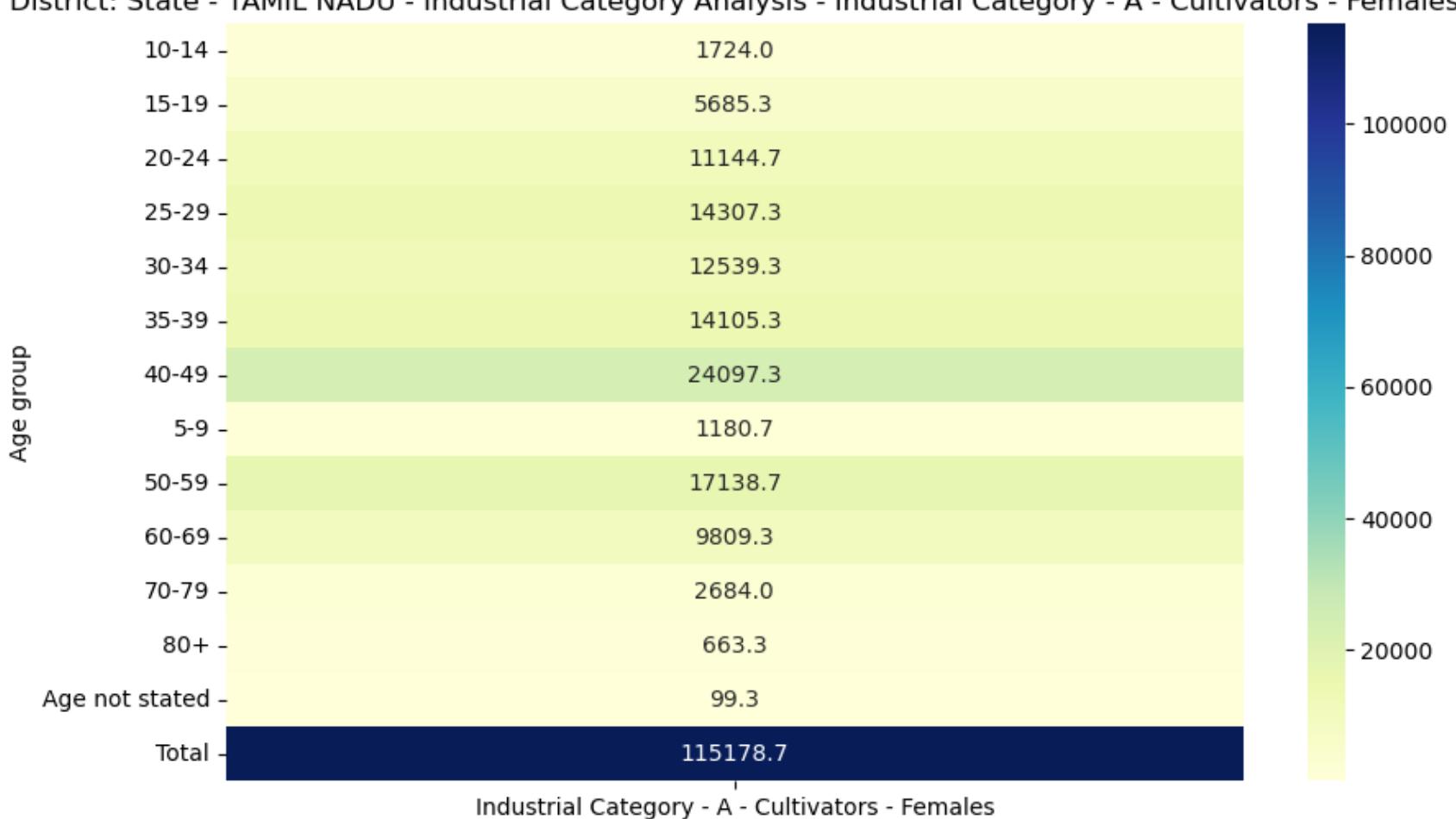
    # Loop through industrial categories
    for category in industrial_categories:
        # Create a pivot table for the specific category in the district
        pivot_table = district_df.pivot_table(index='Age group', values=category, aggfunc='mean')

        # Create a heatmap
        plt.figure(figsize=(10, 6))
        sns.heatmap(pivot_table, annot=True, fmt=".1f", cmap='YlGnBu')
        plt.title(f'District: {district} - Industrial Category Analysis - {category}')
        plt.show()

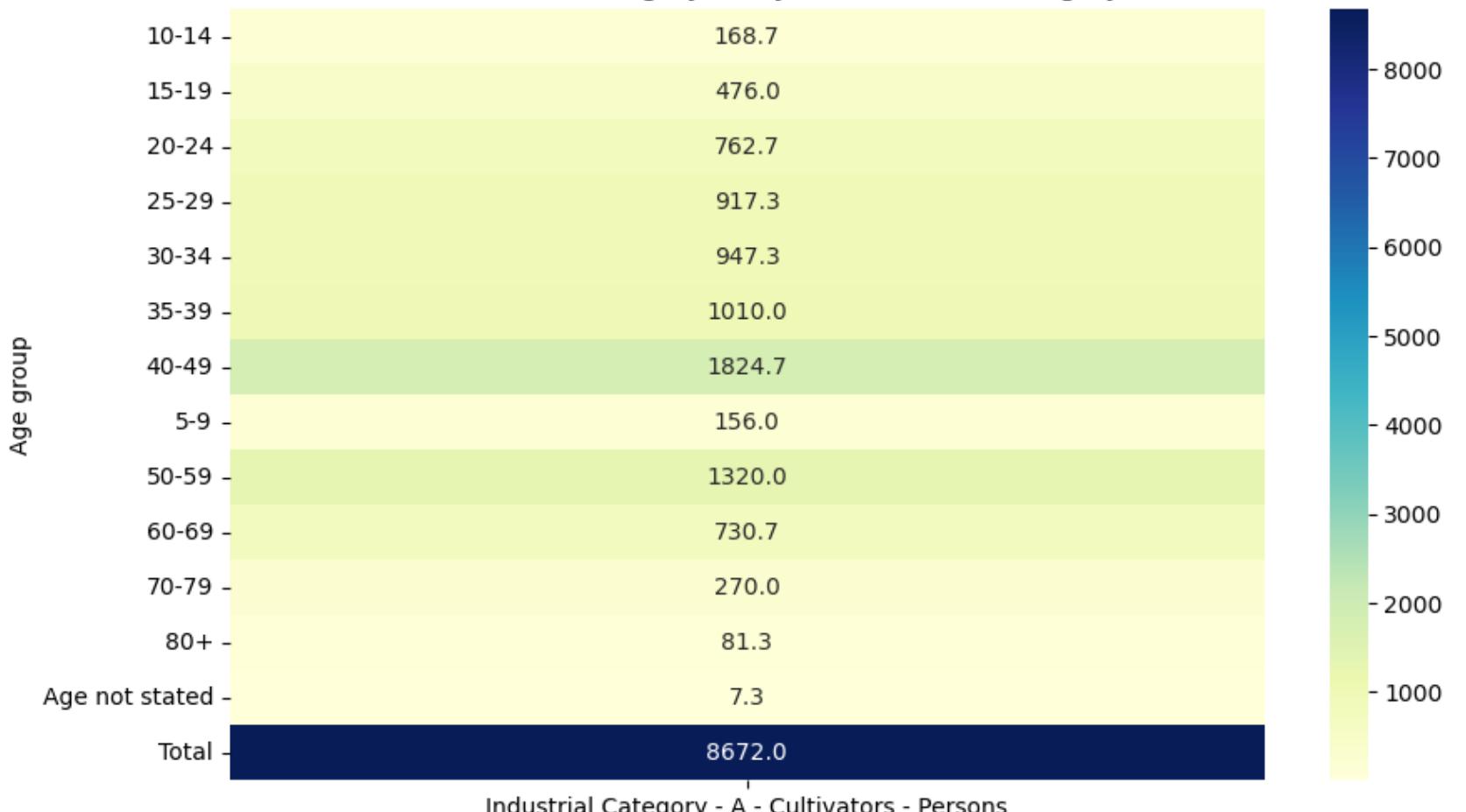
```



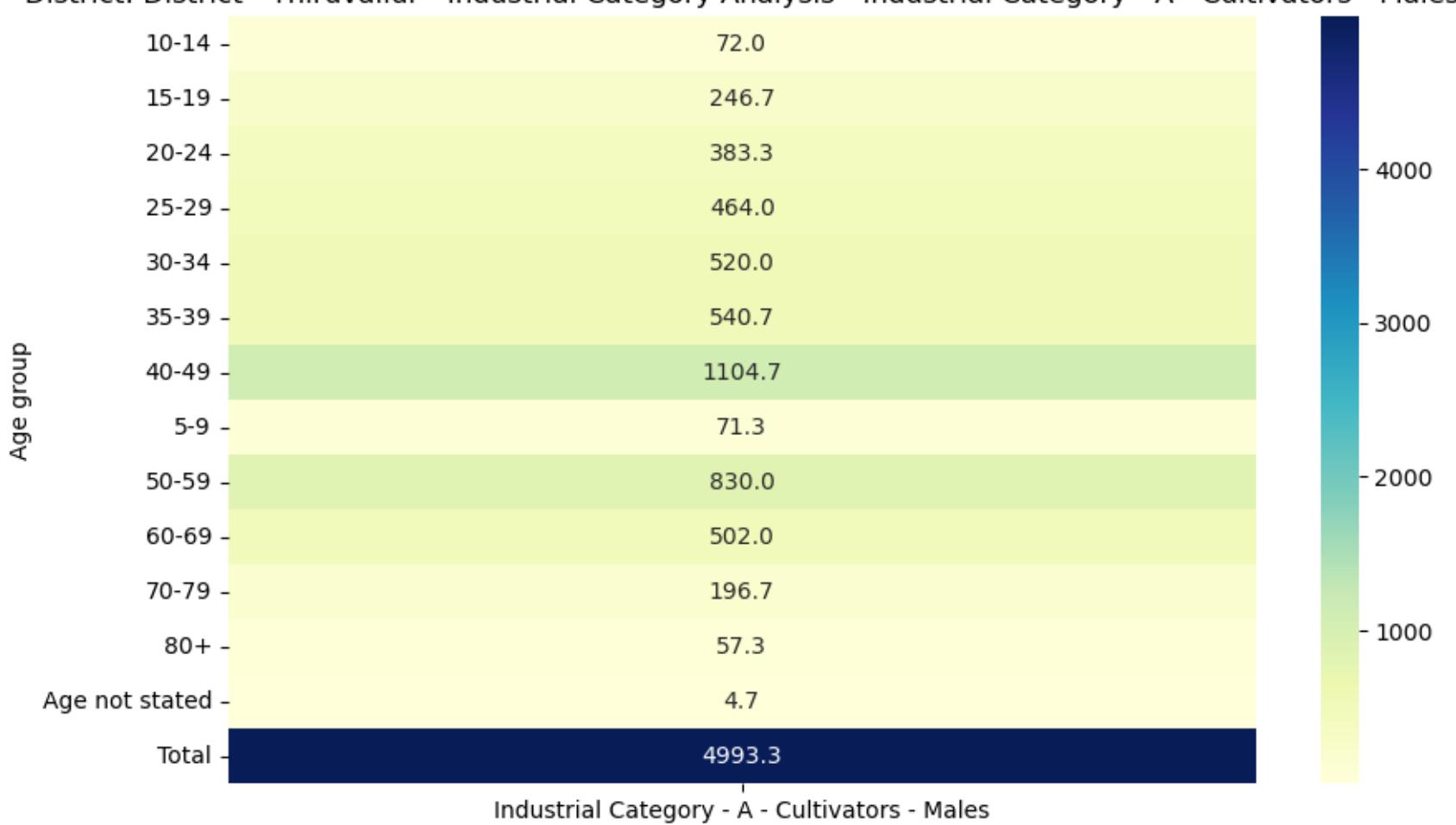
District: State - TAMIL NADU - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



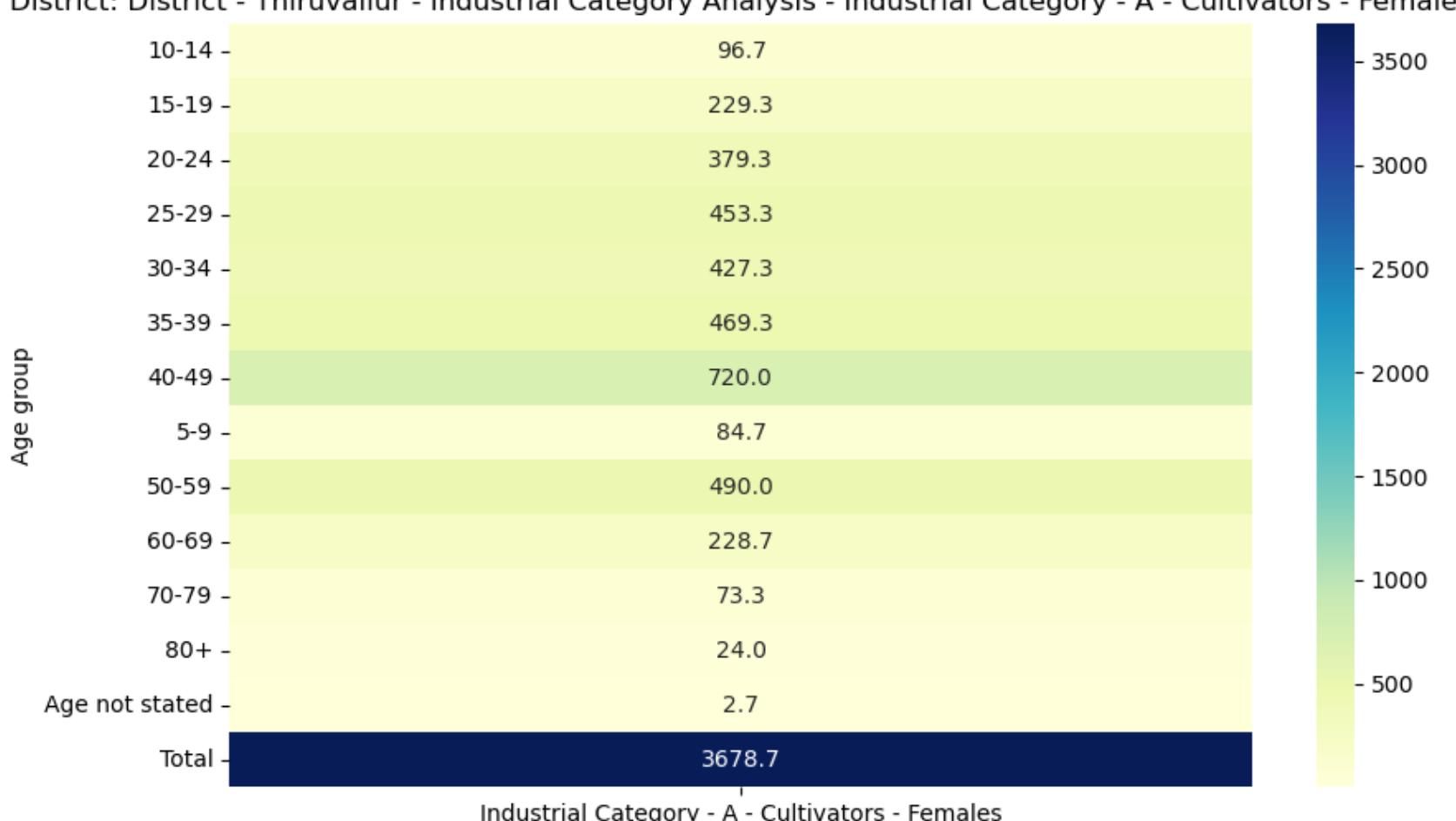
District: District - Thiruvallur - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



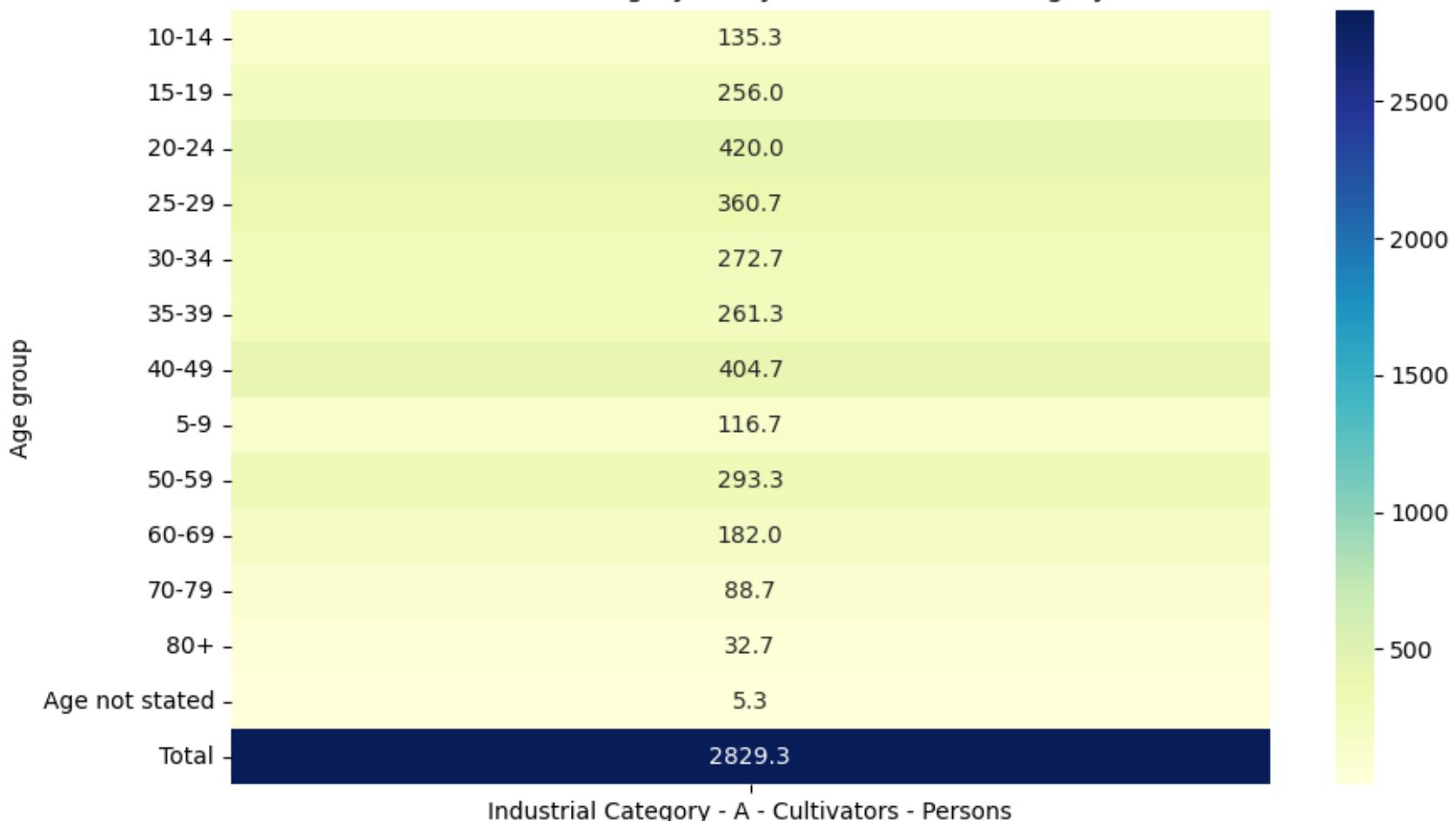
District: District - Thiruvallur - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



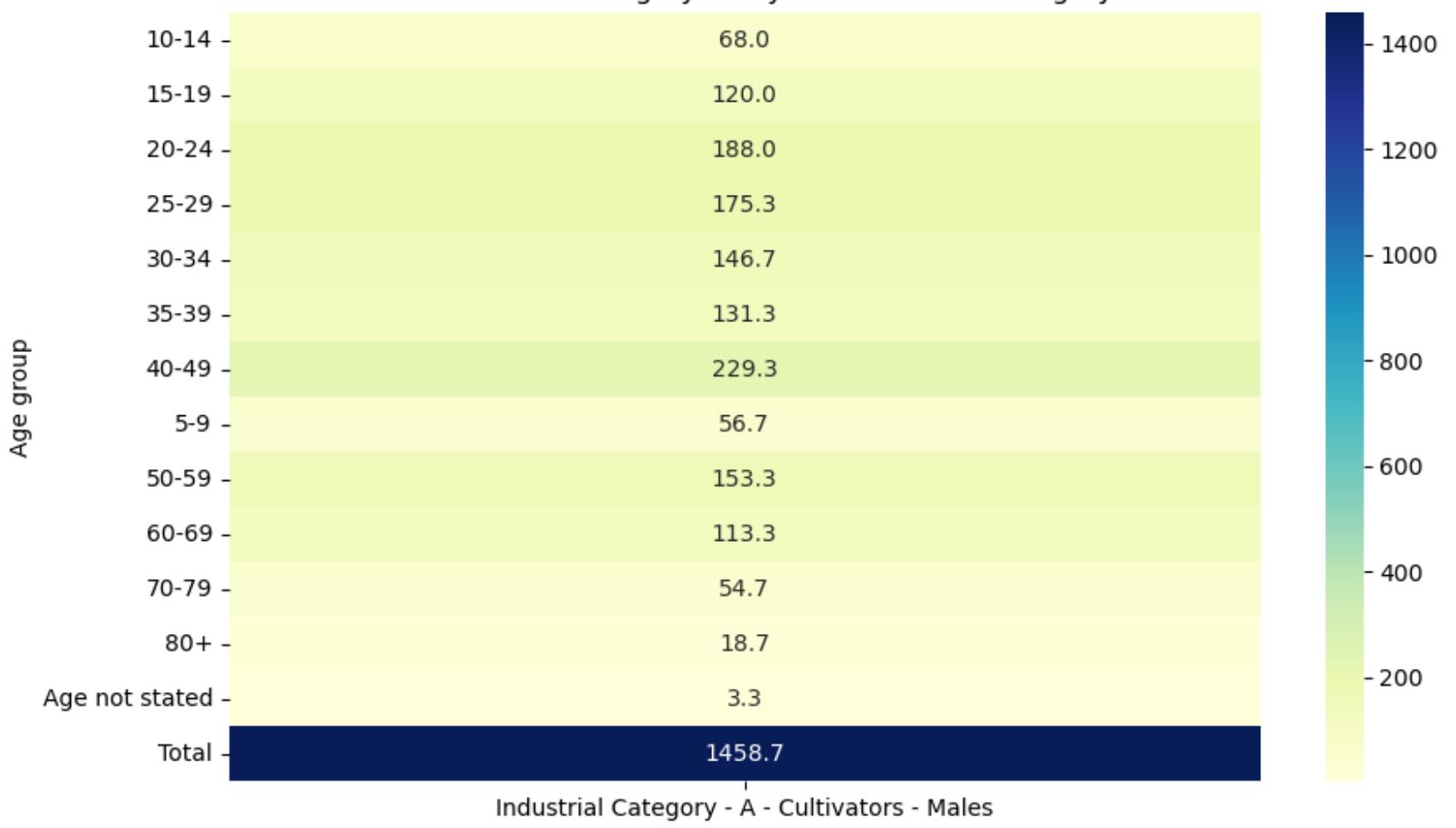
District: District - Thiruvallur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females

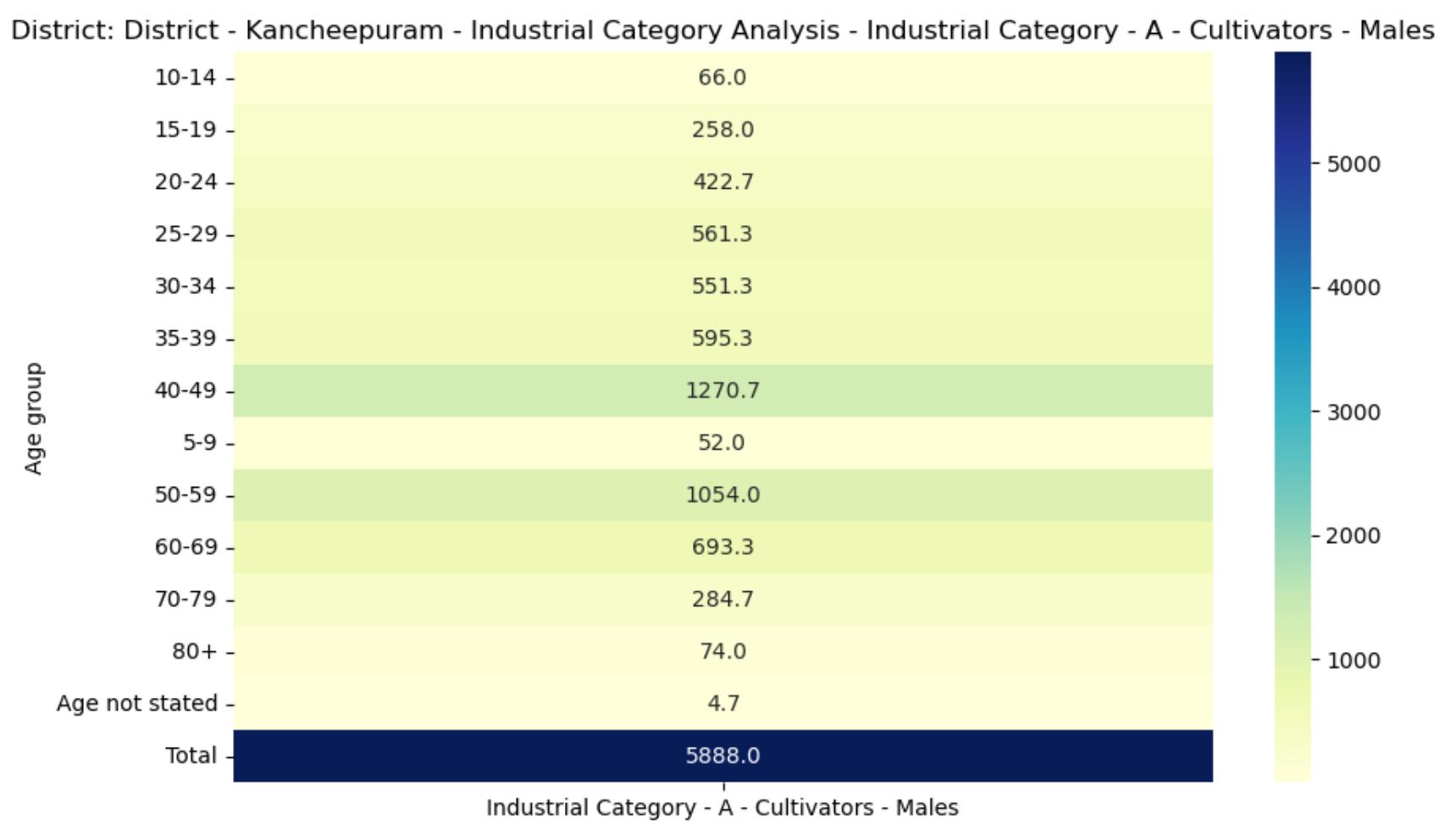
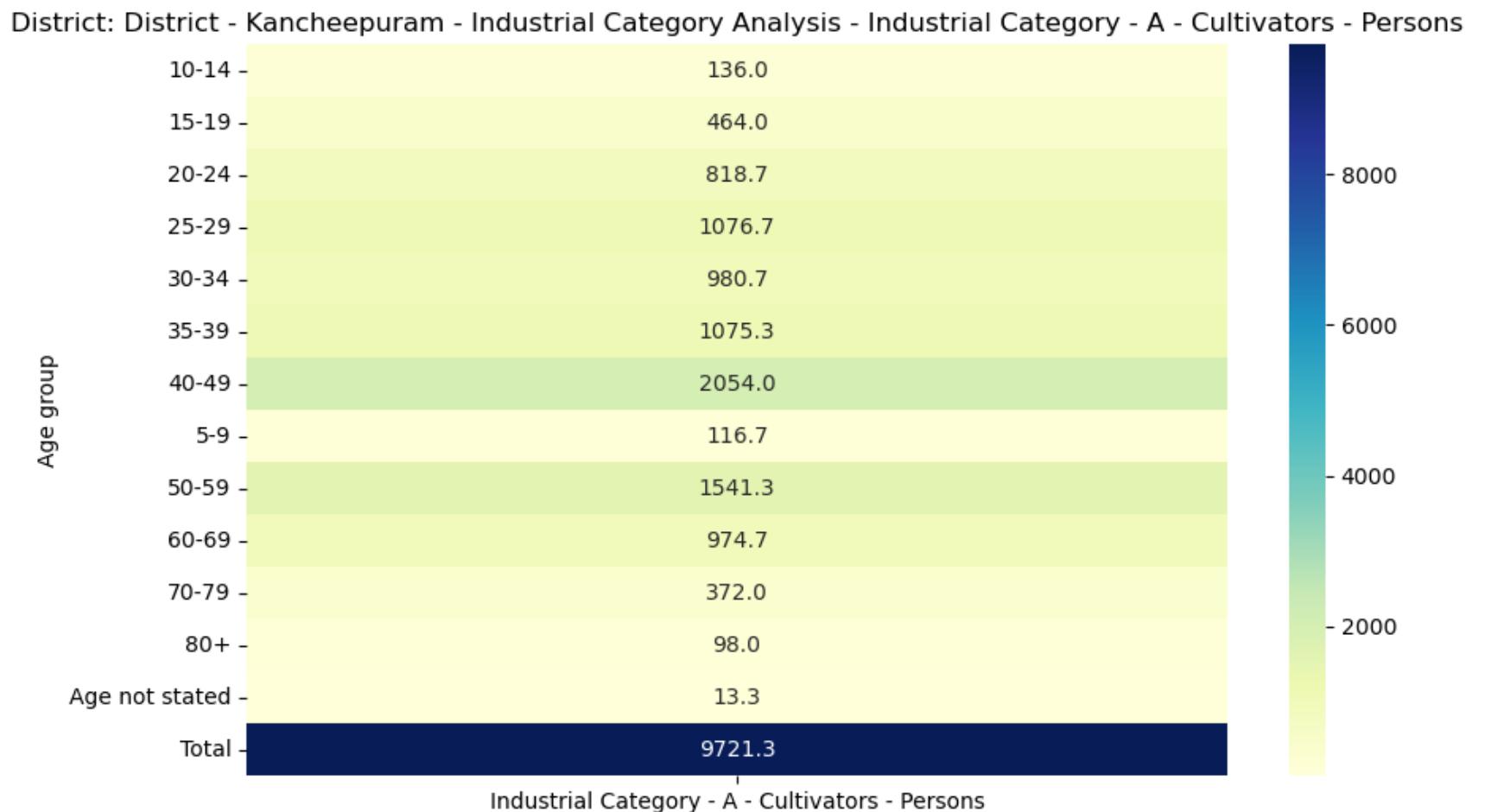
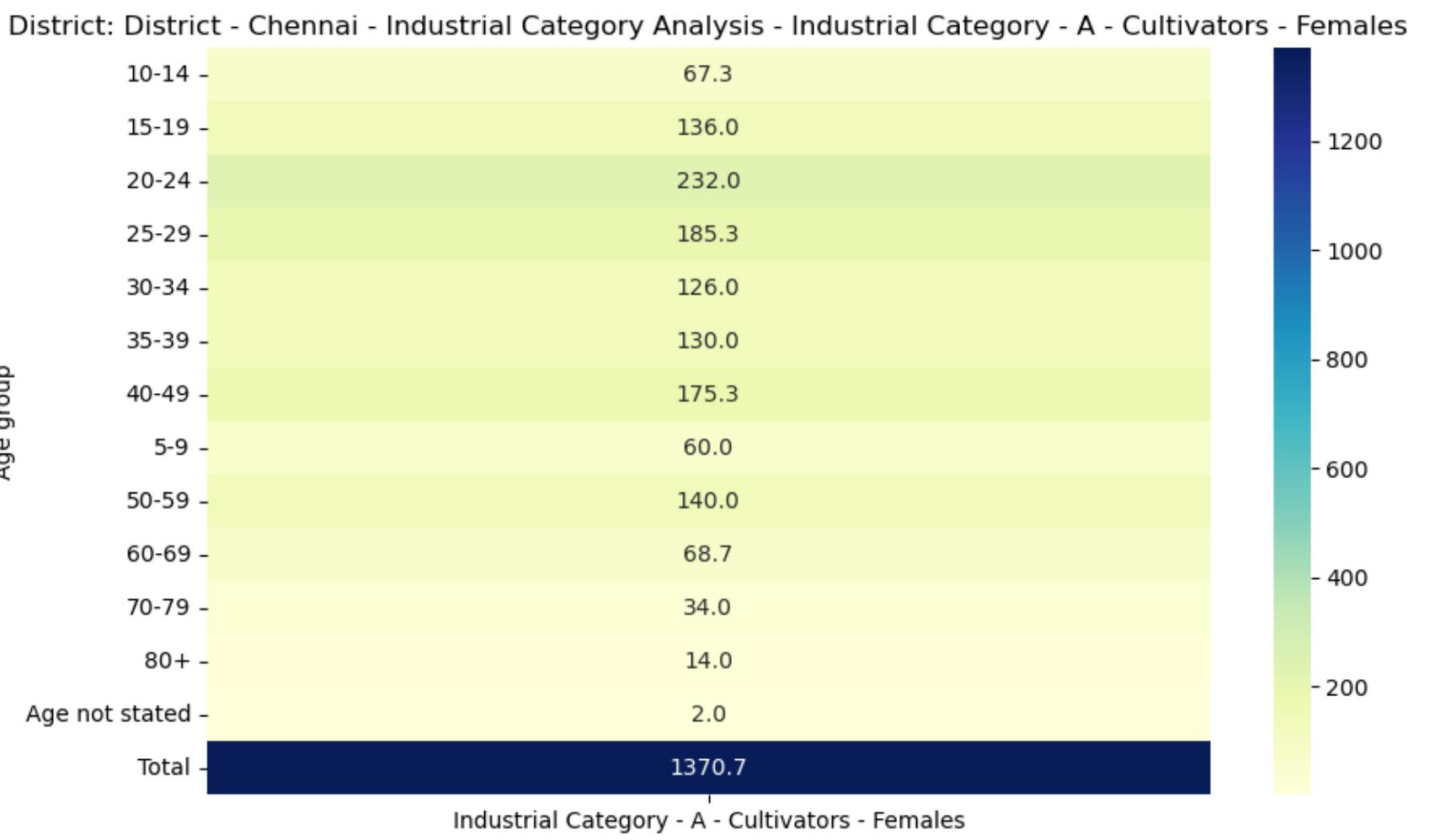


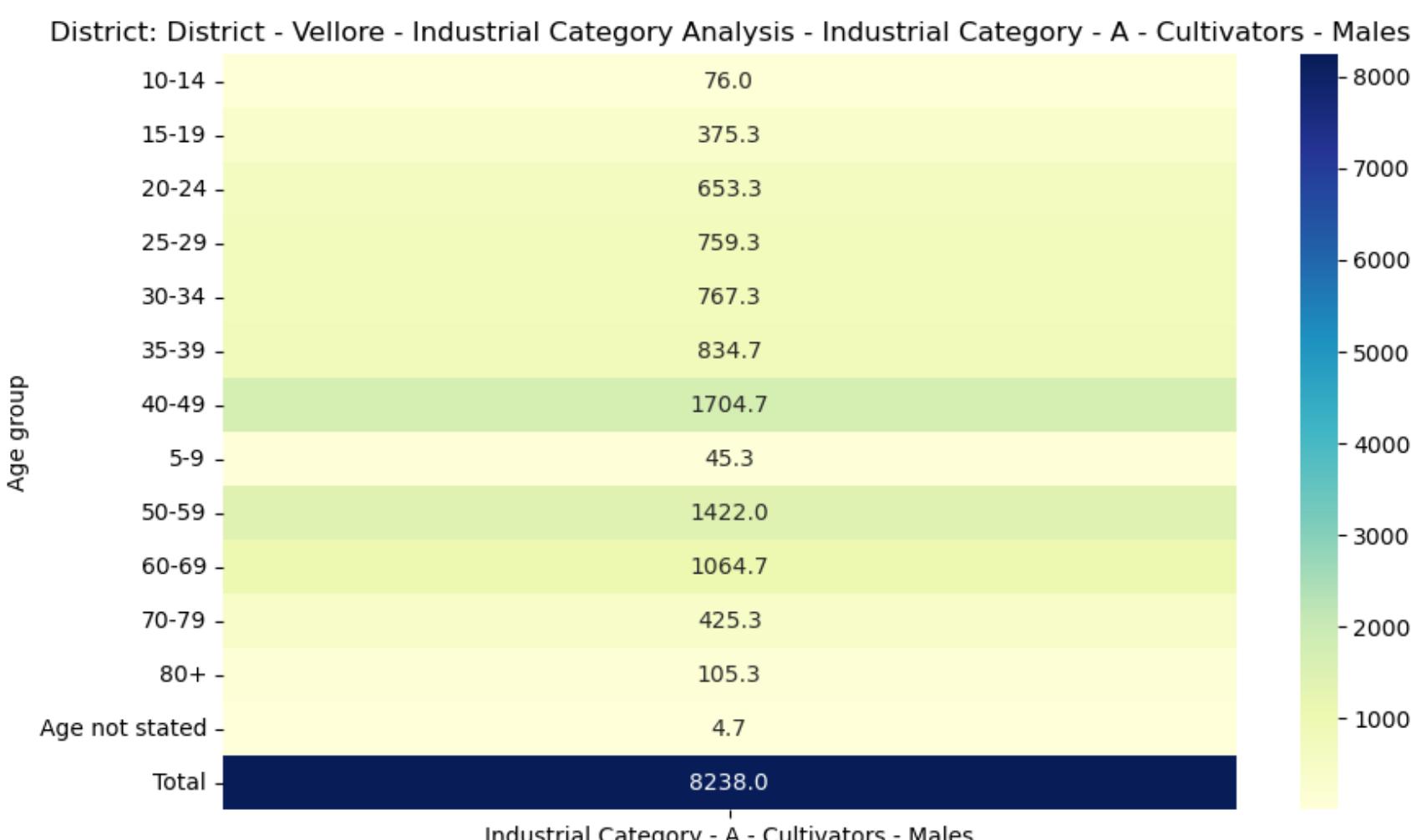
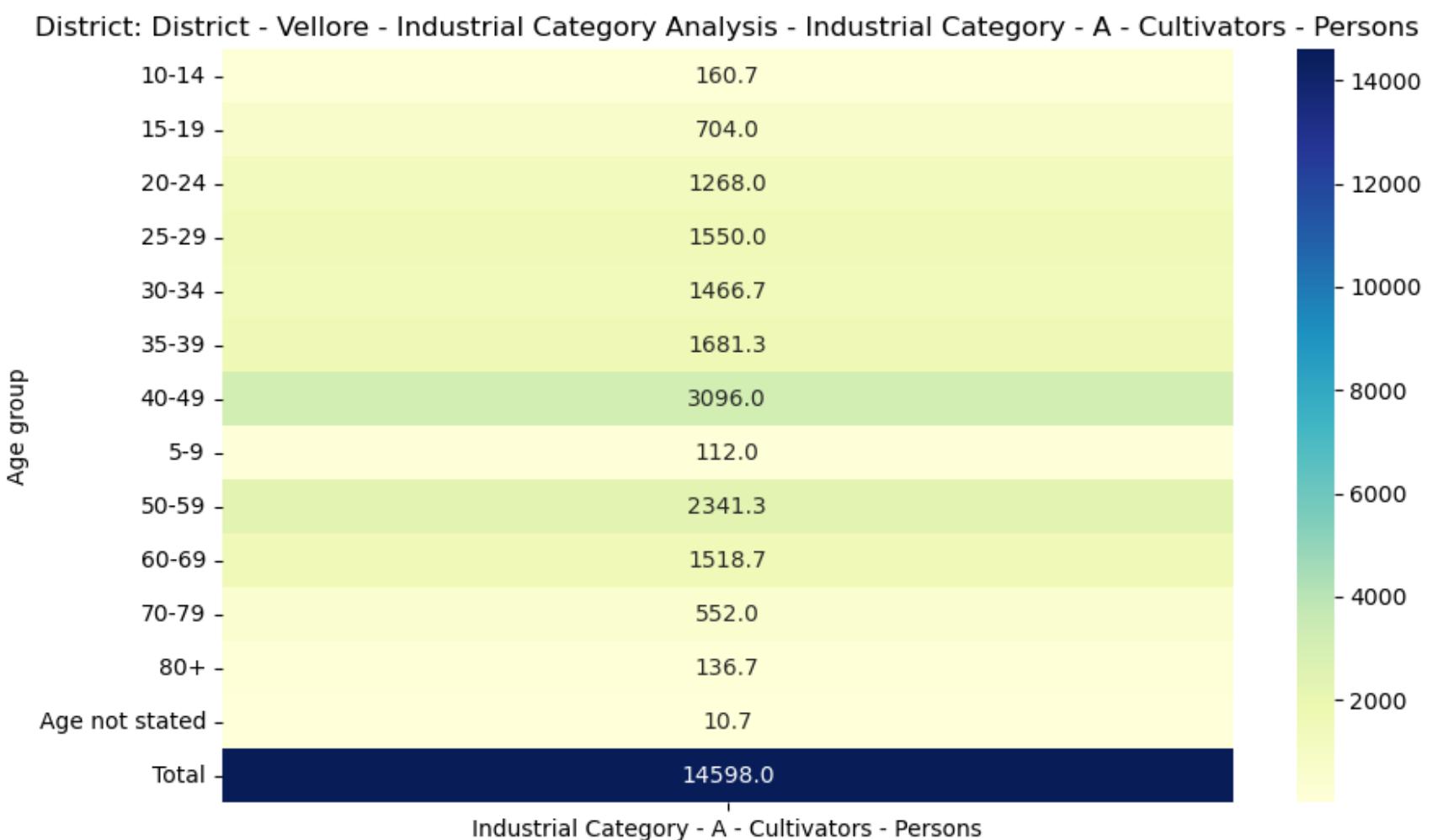
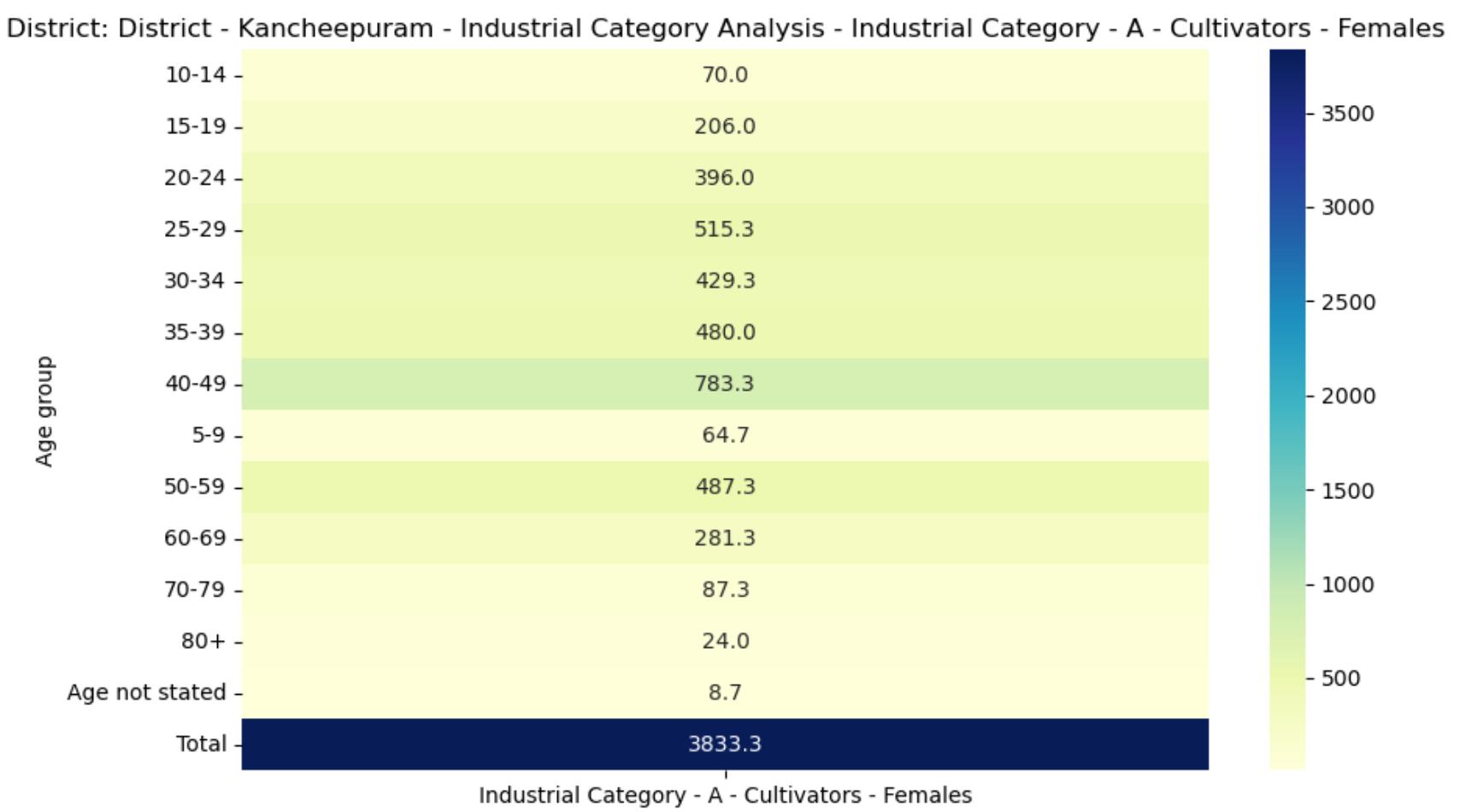
District: District - Chennai - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



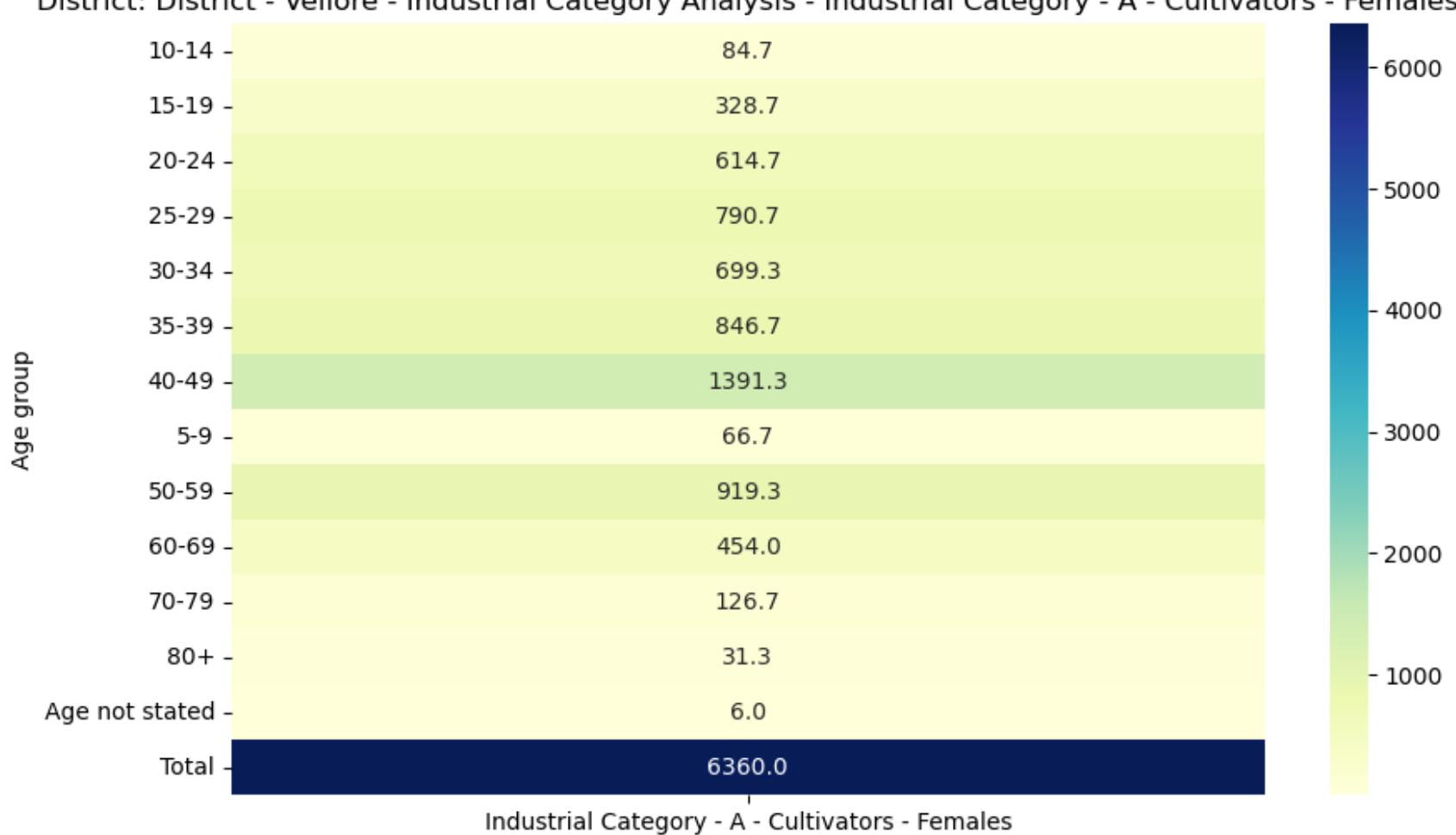
District: District - Chennai - Industrial Category Analysis - Industrial Category - A - Cultivators - Males





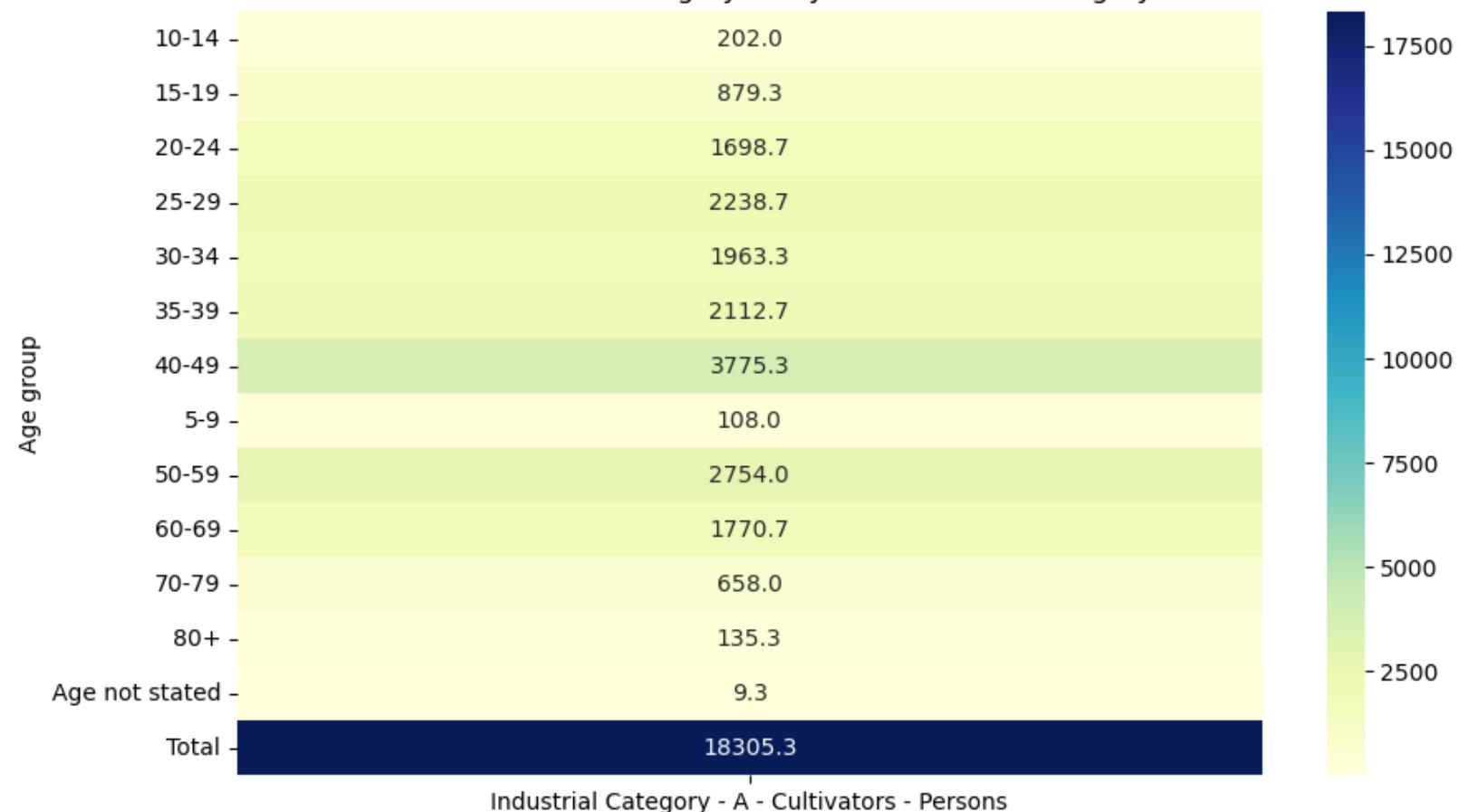


District: District - Vellore - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



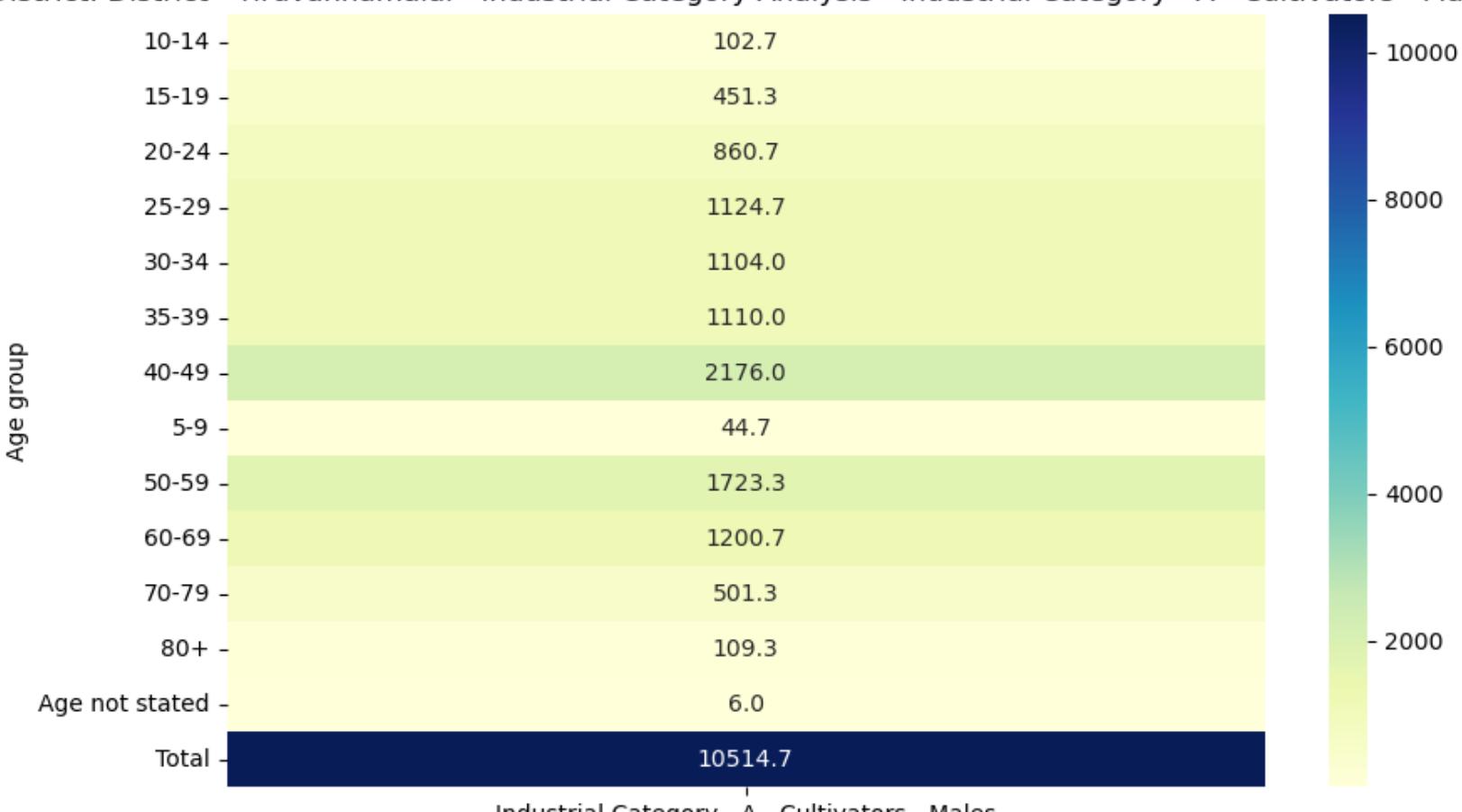
Industrial Category - A - Cultivators - Females

District: District - Tiruvannamalai - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons

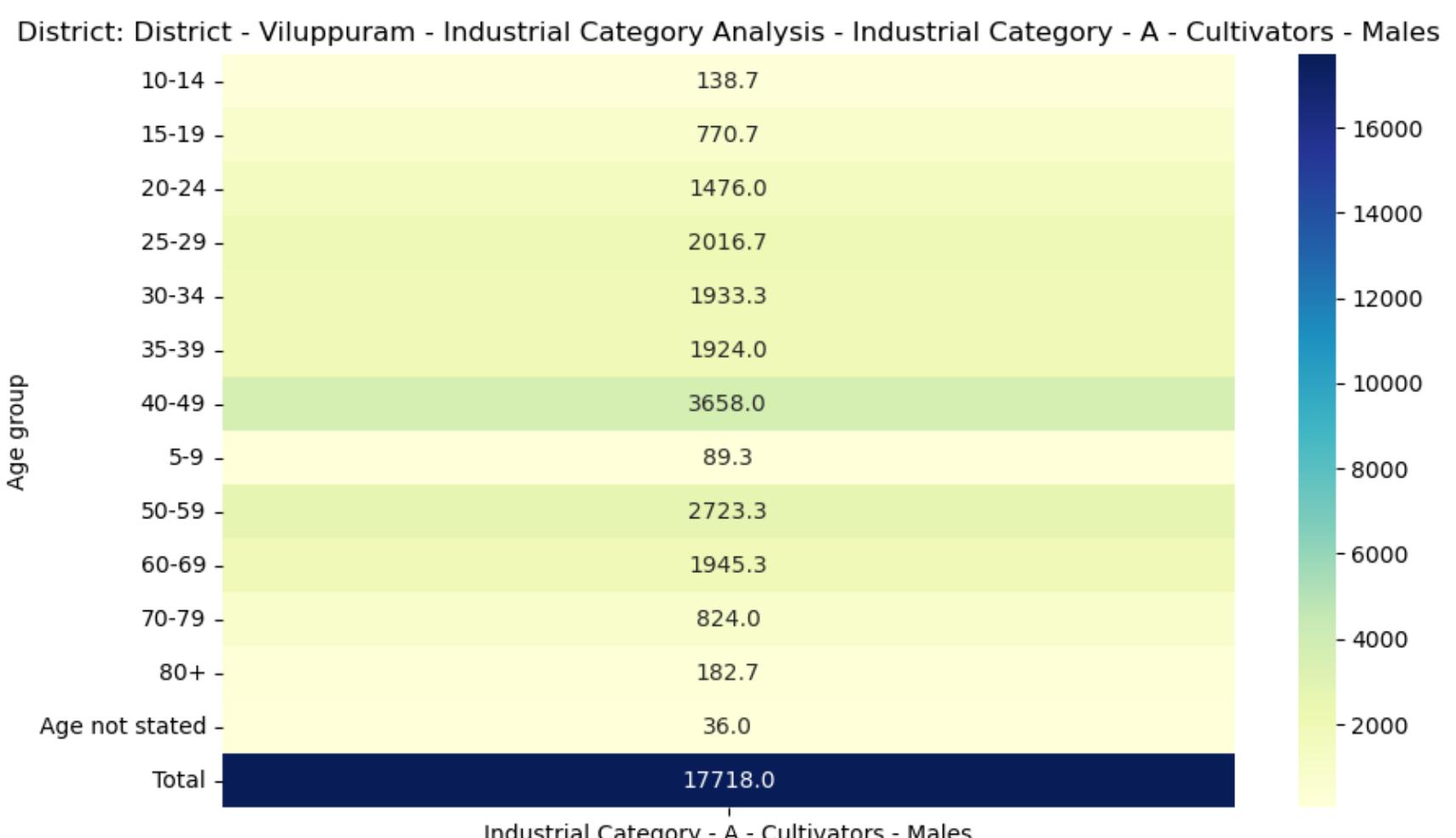
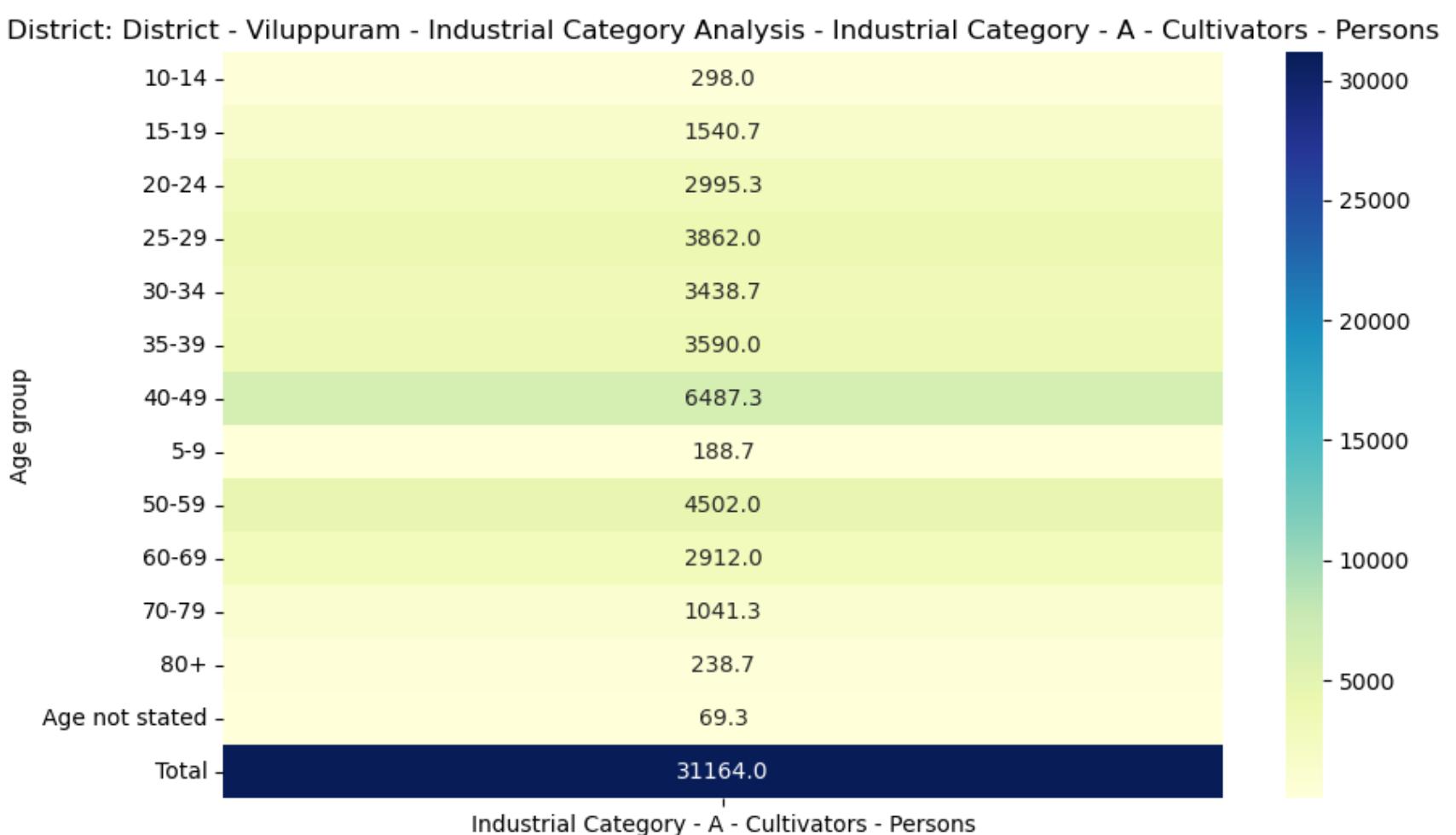
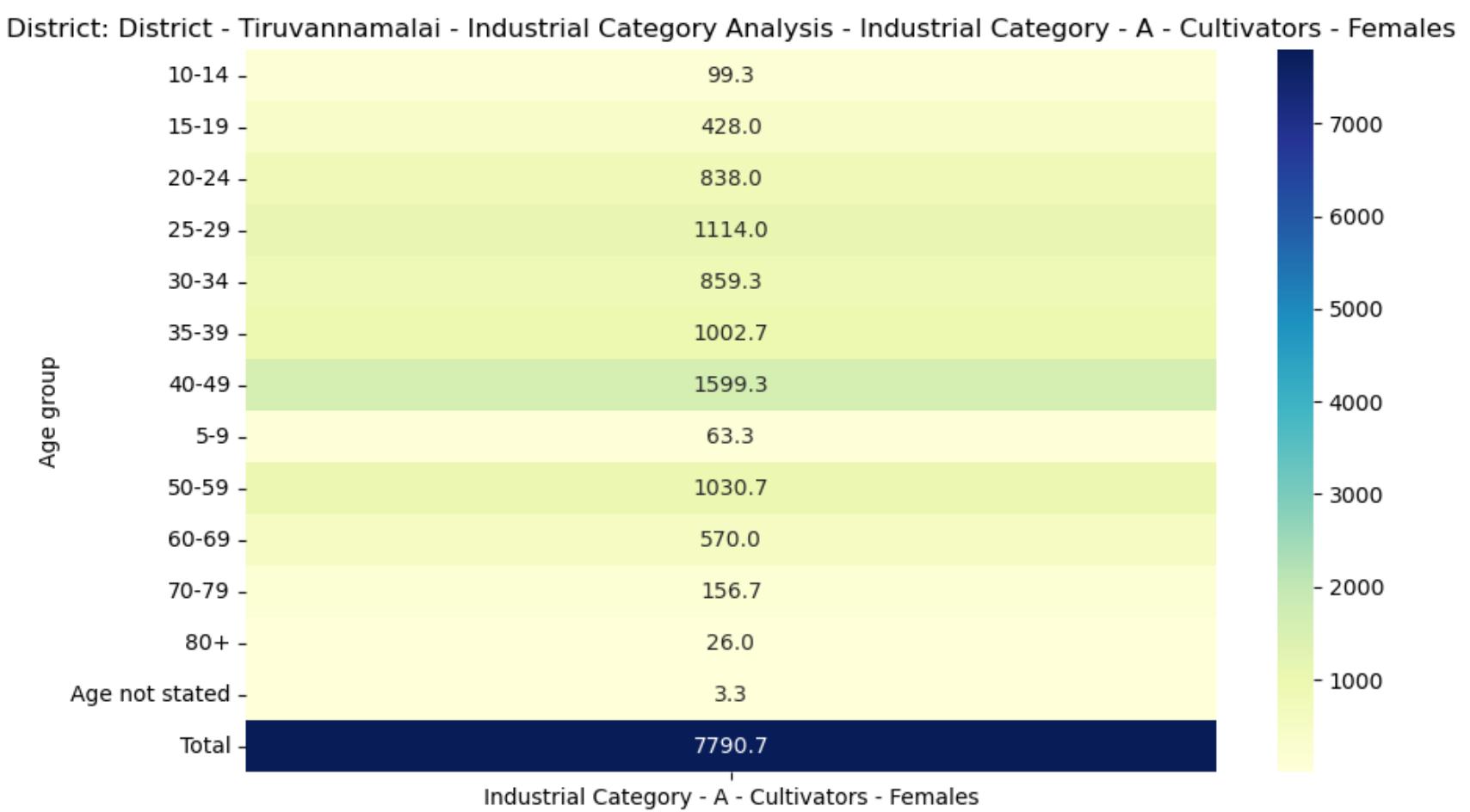


Industrial Category - A - Cultivators - Persons

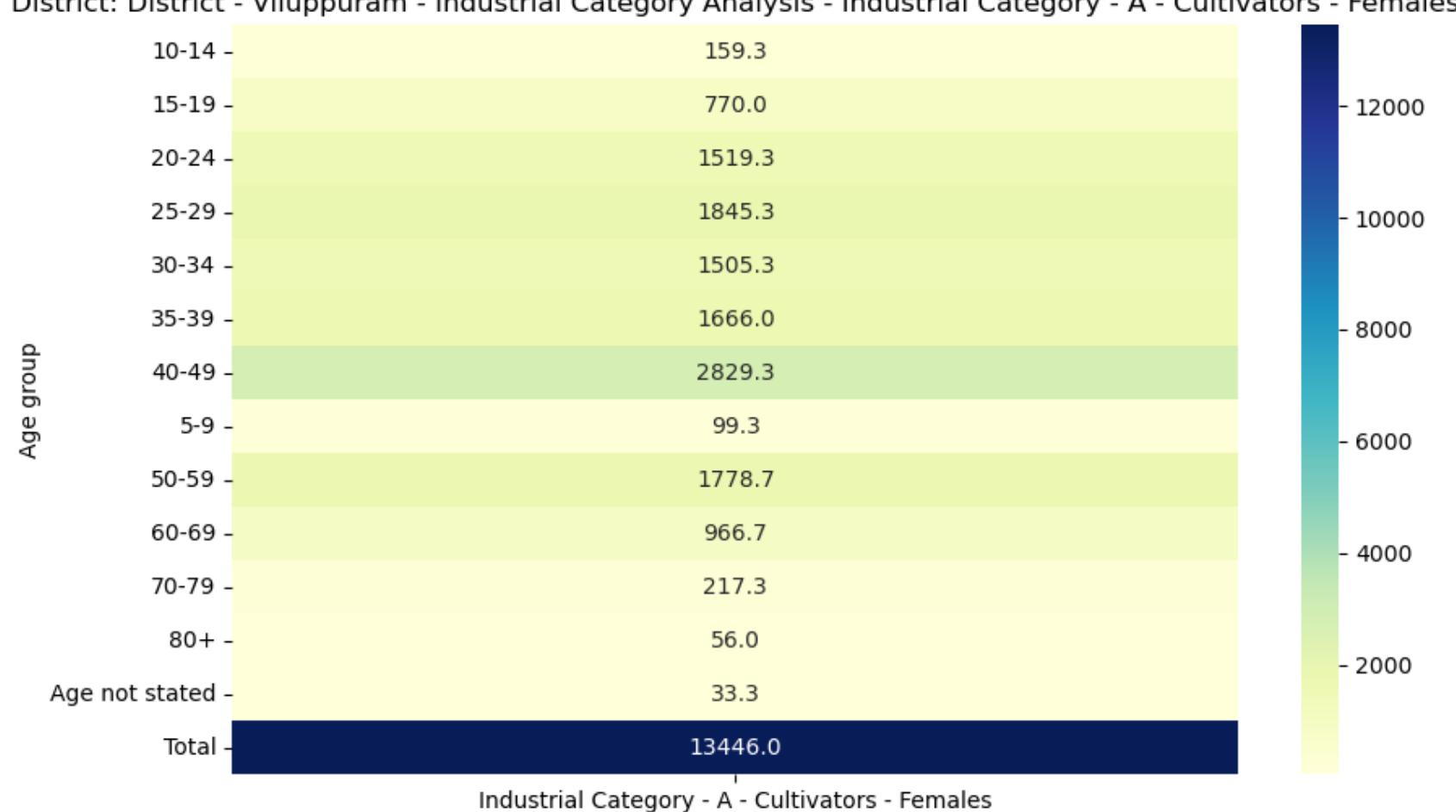
District: District - Tiruvannamalai - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



Industrial Category - A - Cultivators - Males

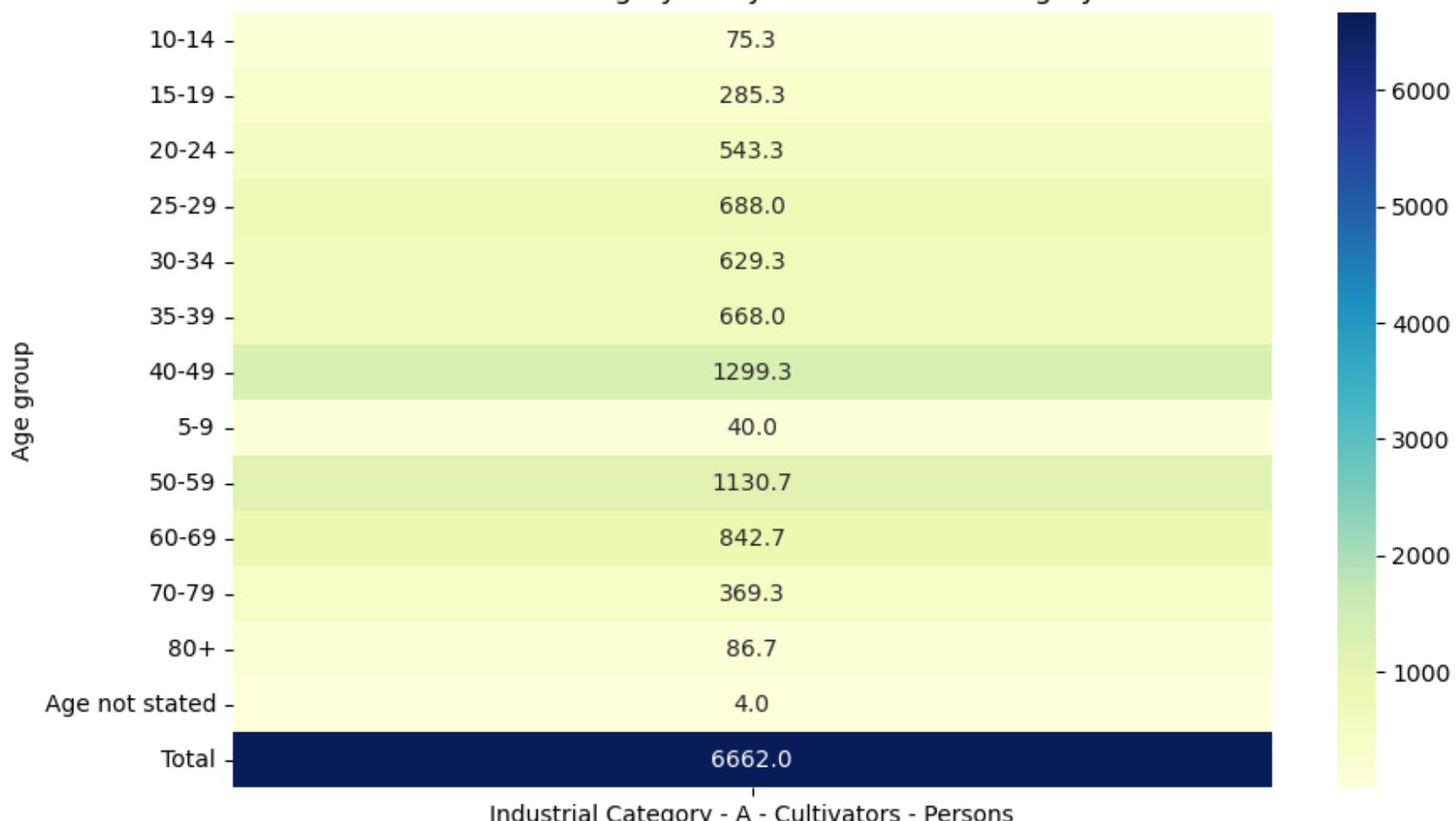


District: District - Viluppuram - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



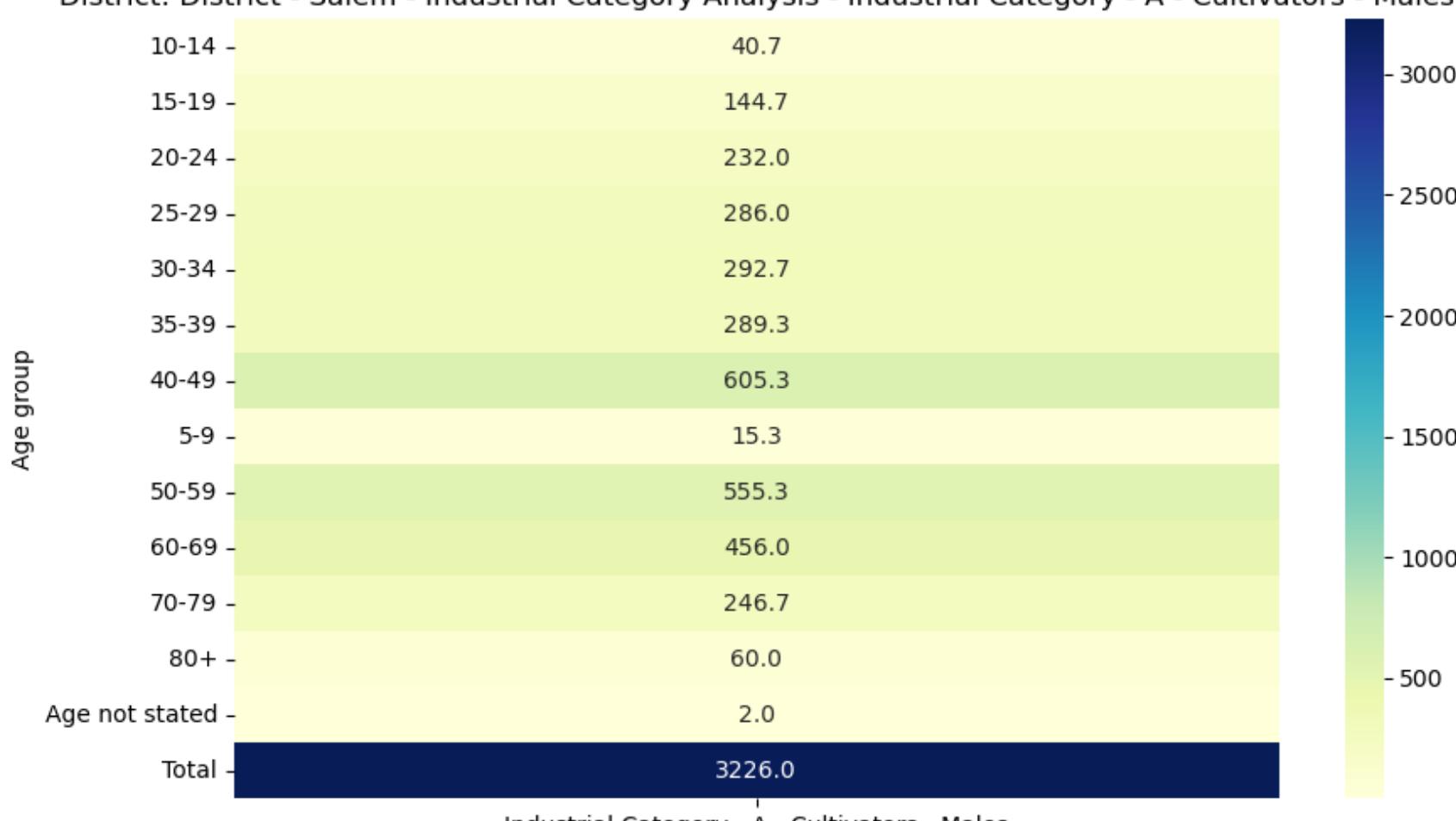
Industrial Category - A - Cultivators - Females

District: District - Salem - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



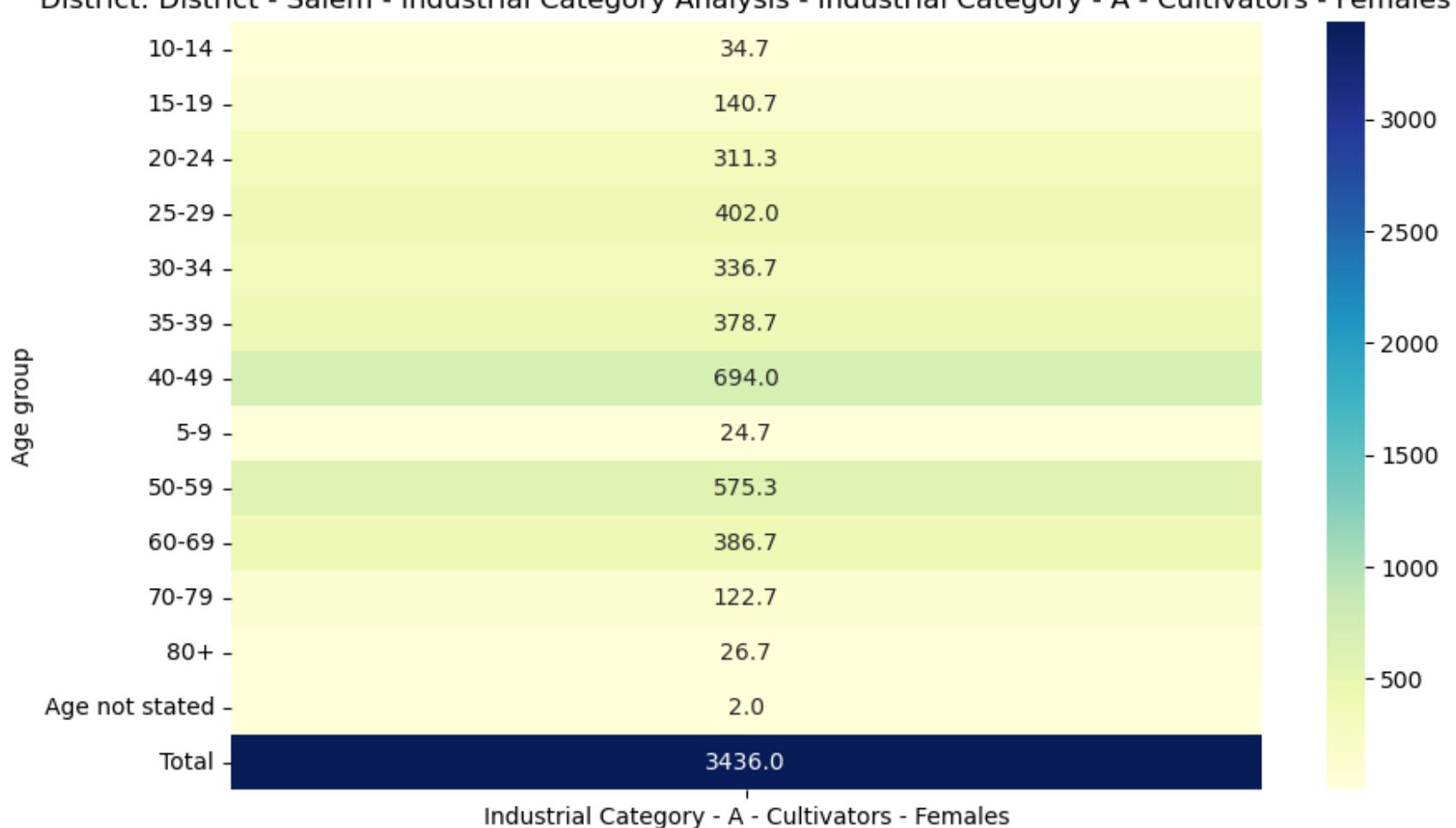
Industrial Category - A - Cultivators - Persons

District: District - Salem - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



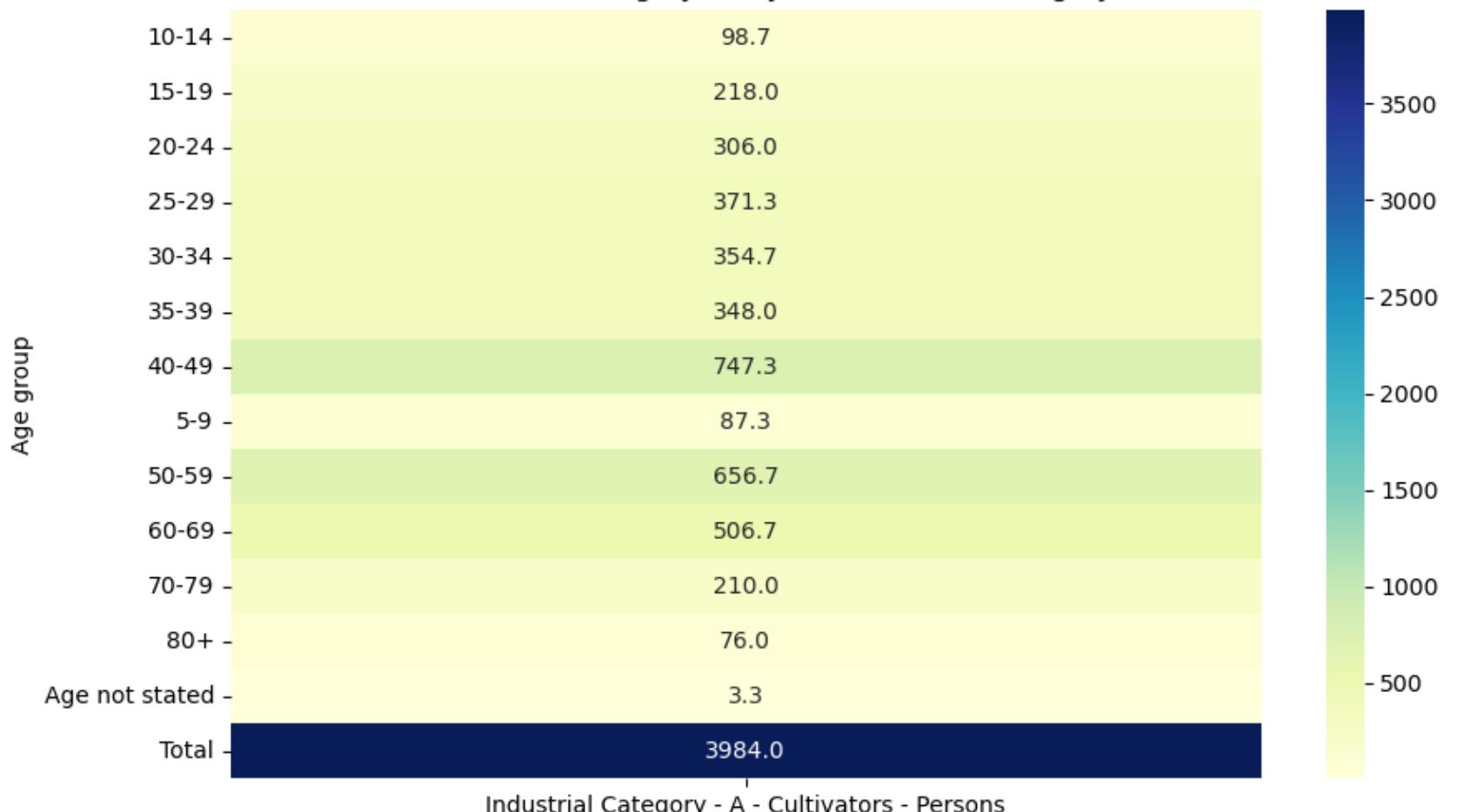
Industrial Category - A - Cultivators - Males

District: District - Salem - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



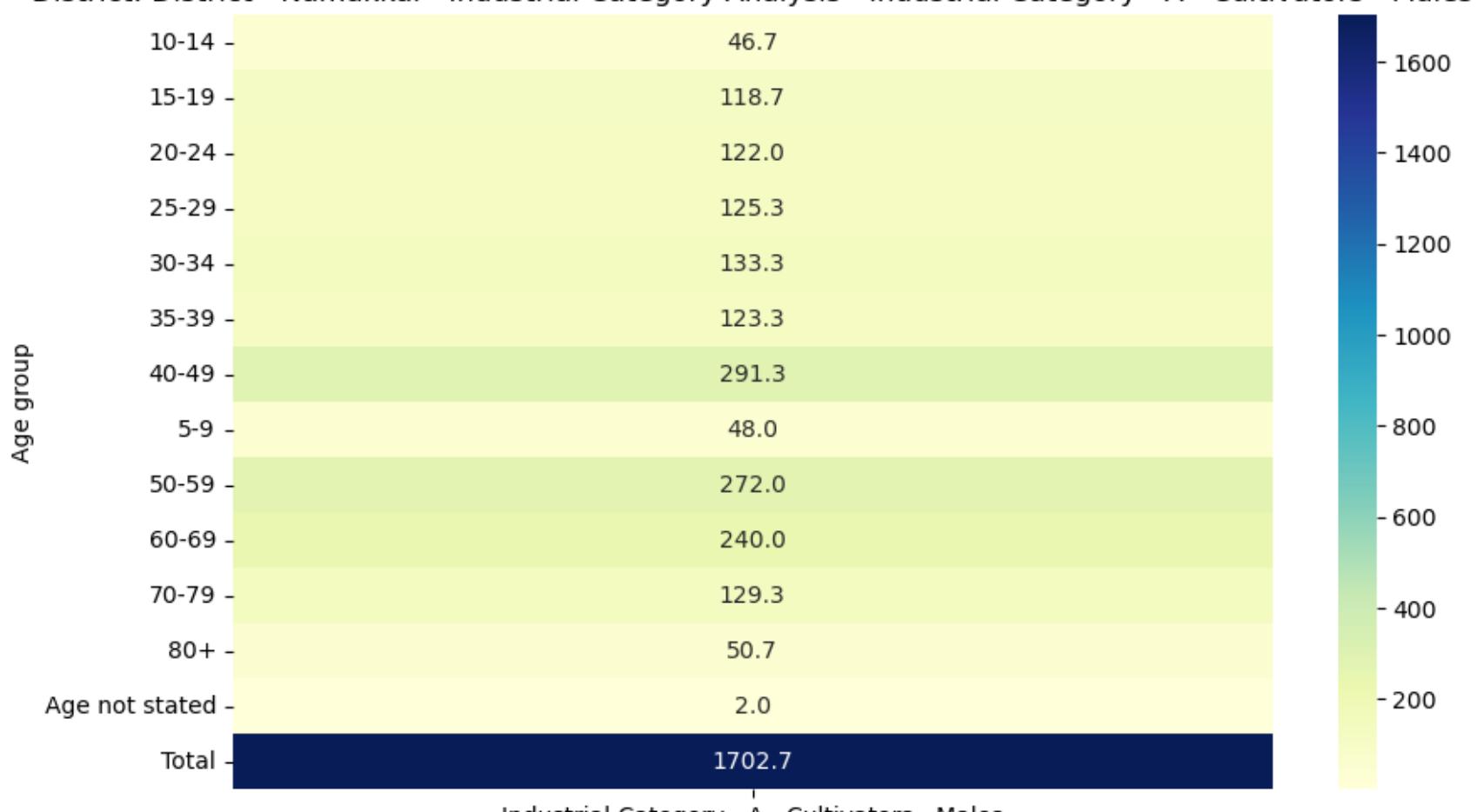
Industrial Category - A - Cultivators - Females

District: District - Namakkal - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



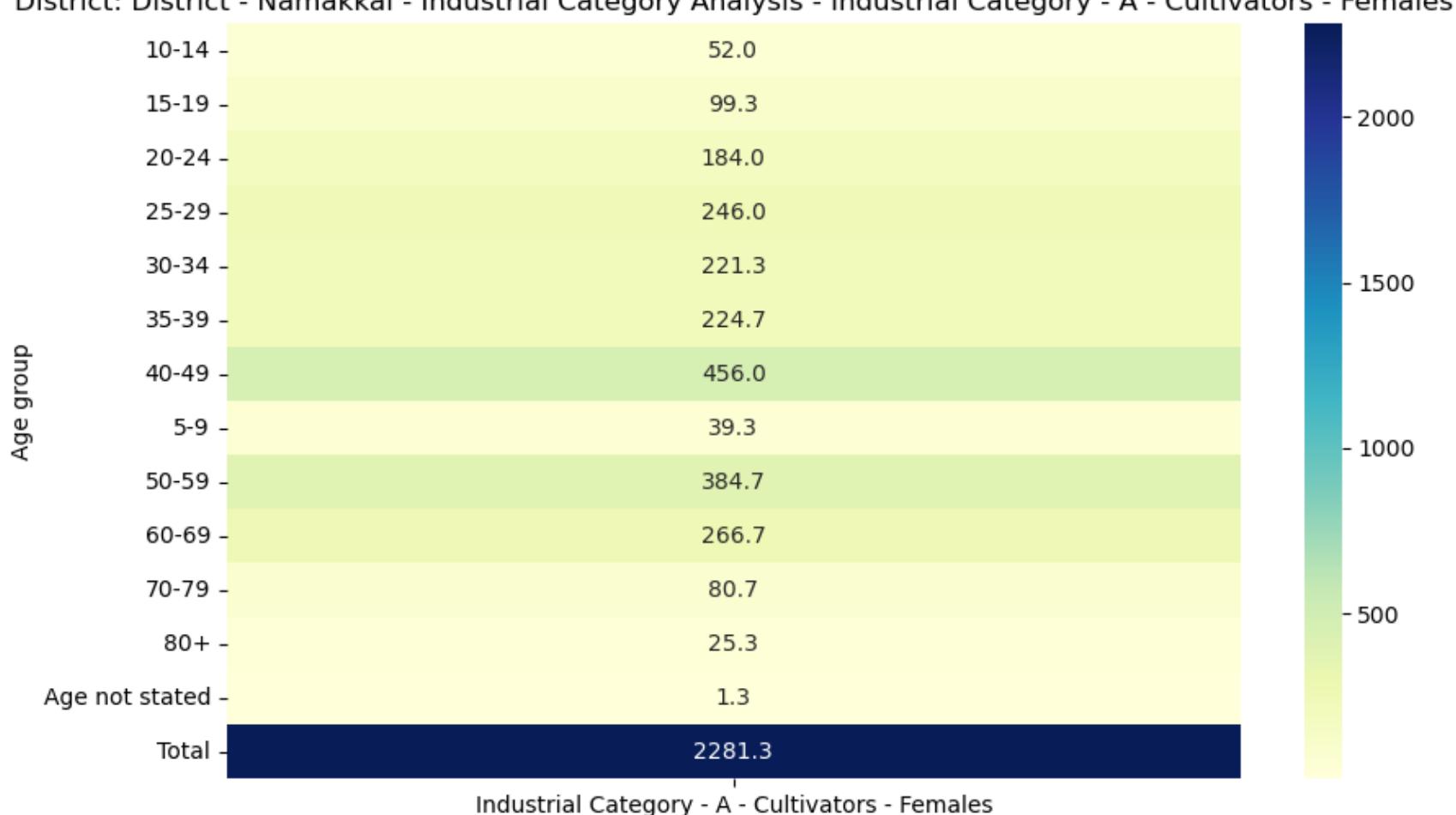
Industrial Category - A - Cultivators - Persons

District: District - Namakkal - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

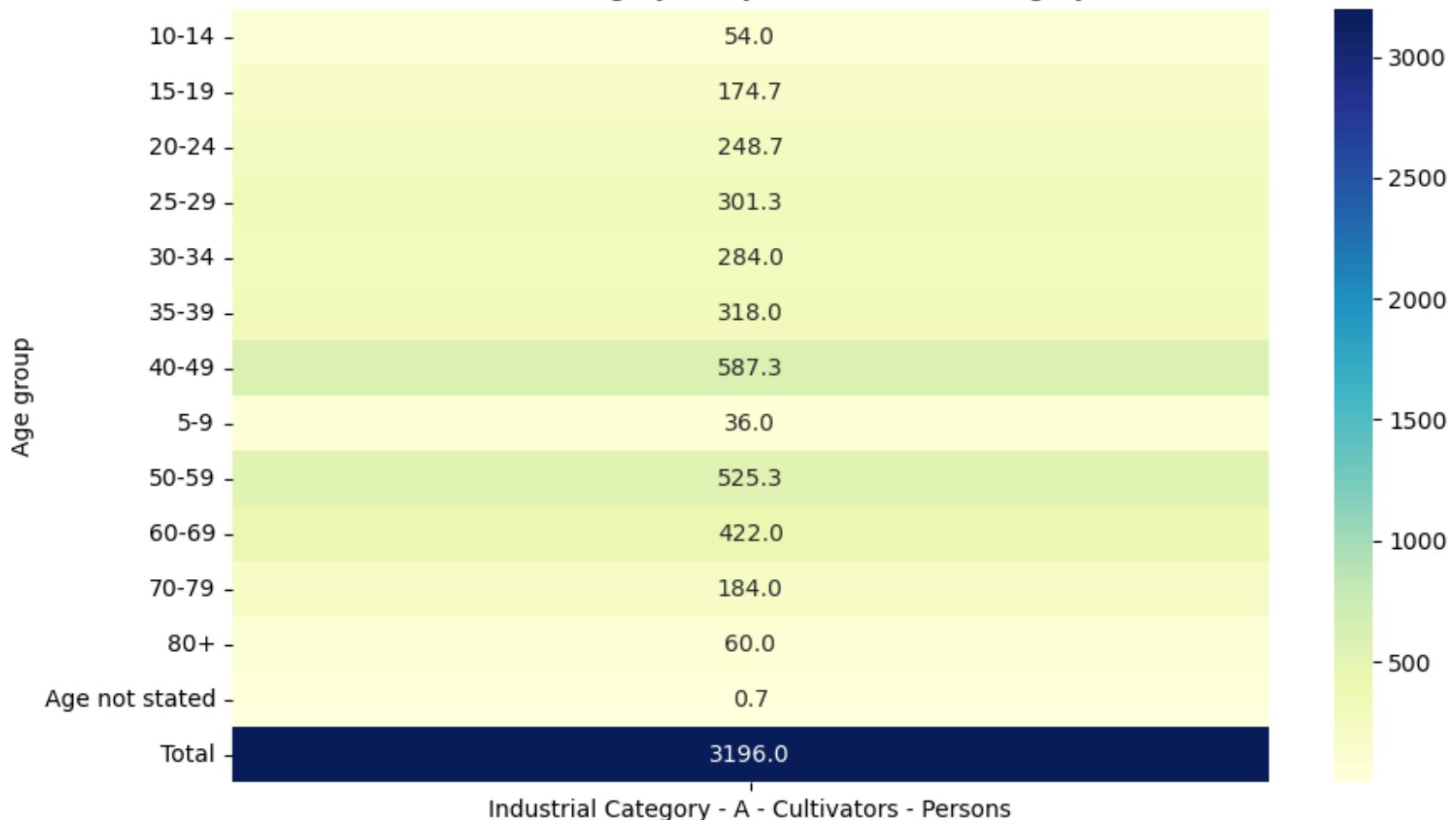


Industrial Category - A - Cultivators - Males

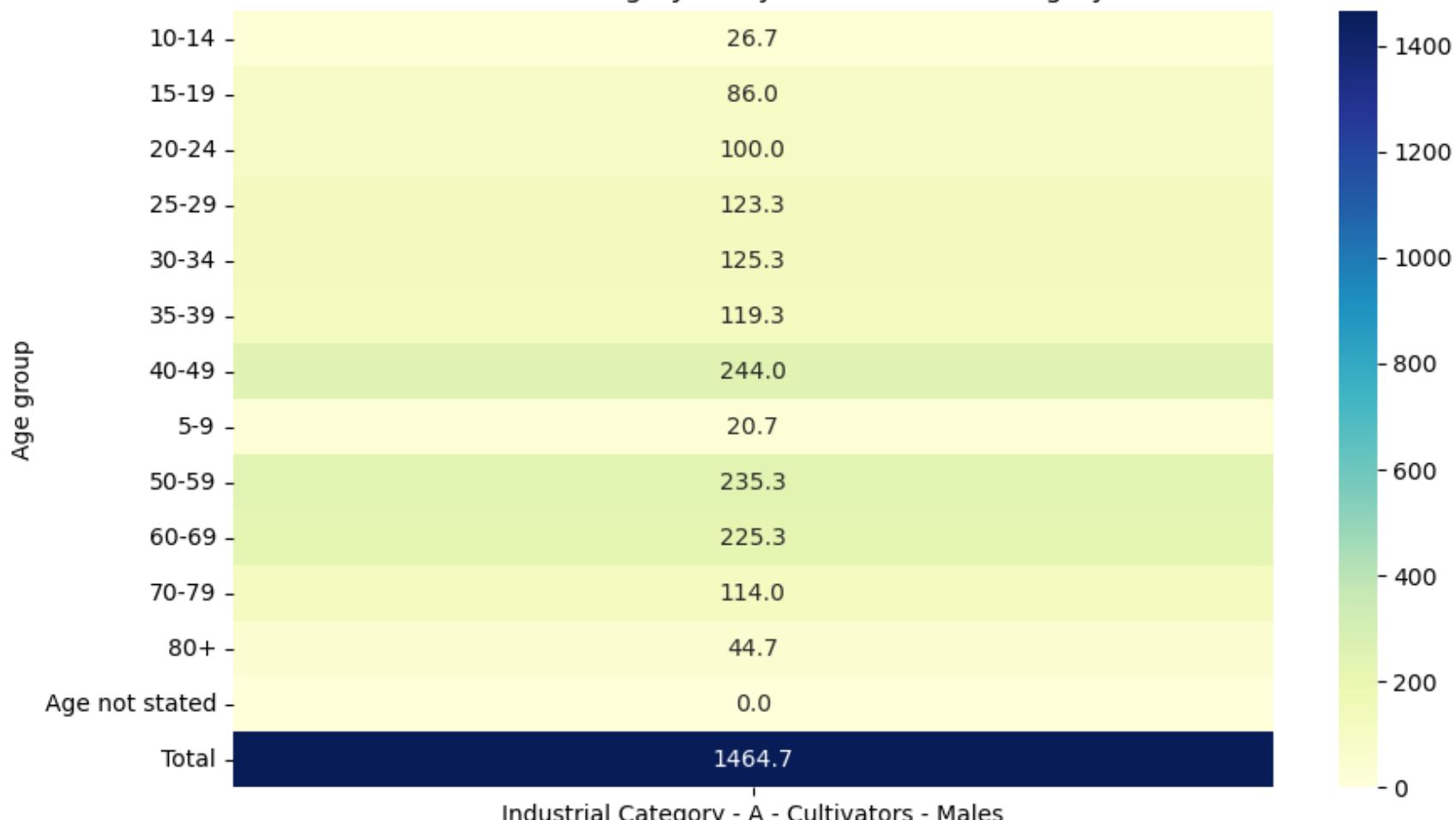
District: District - Namakkal - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



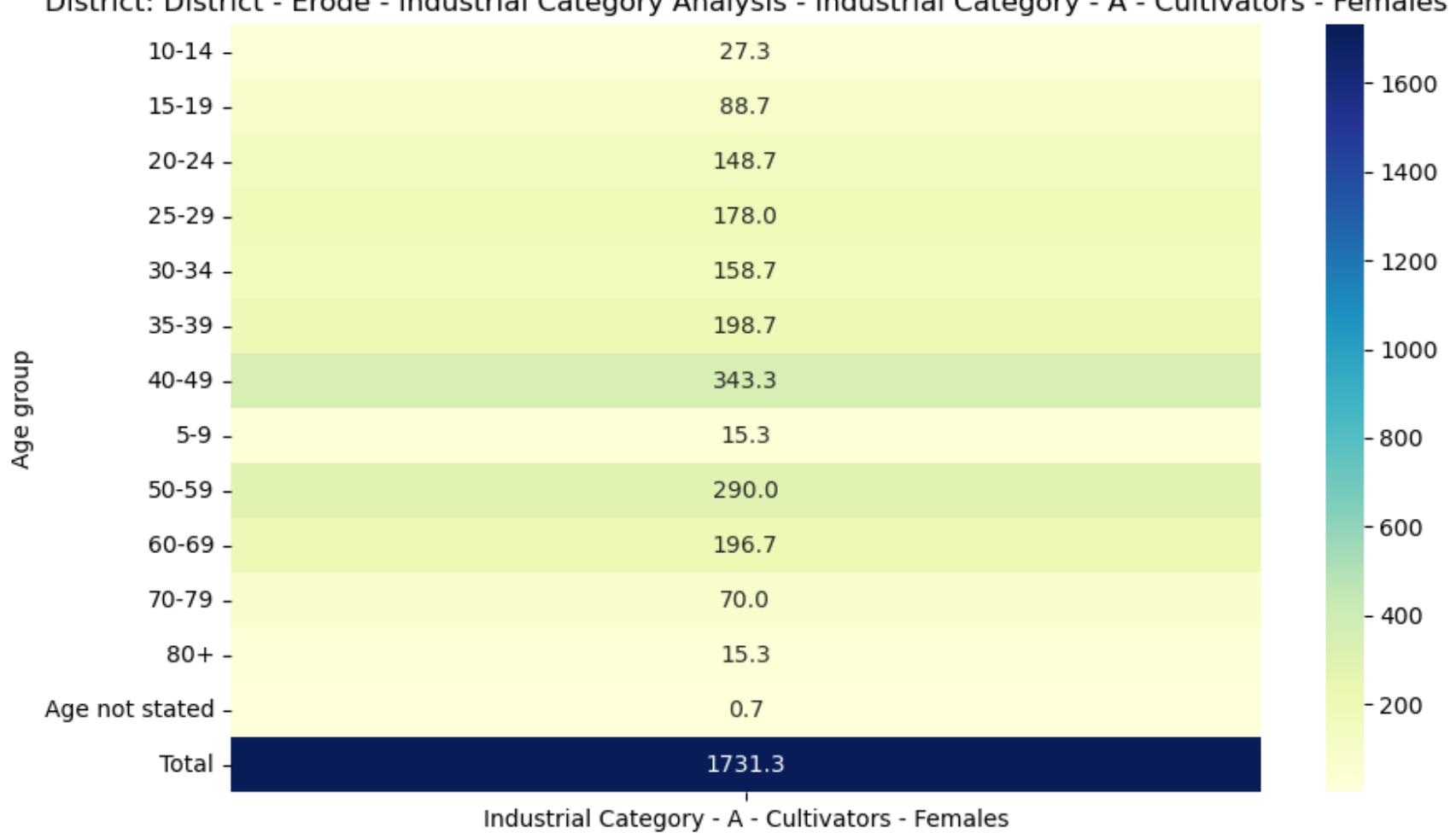
District: District - Erode - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



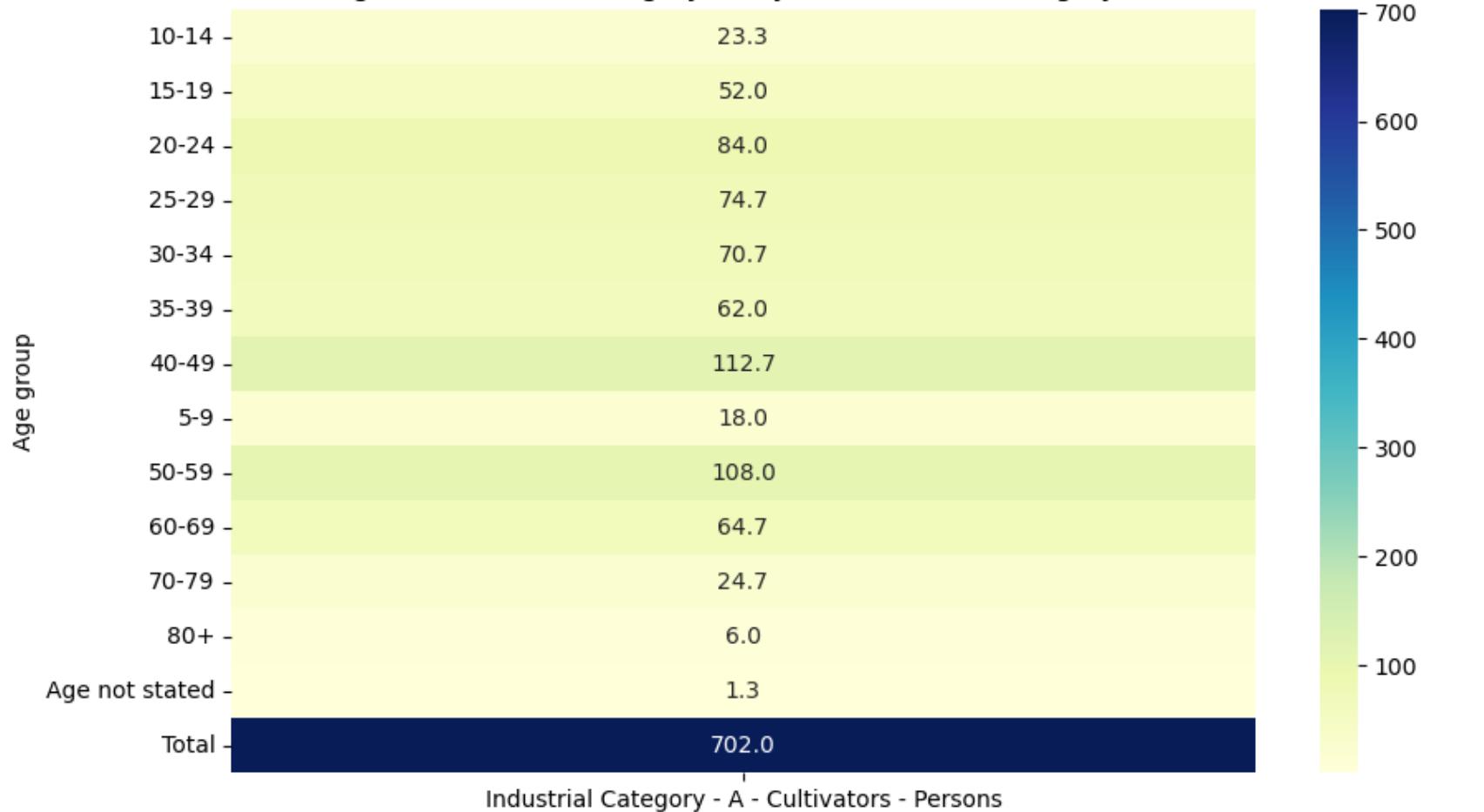
District: District - Erode - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



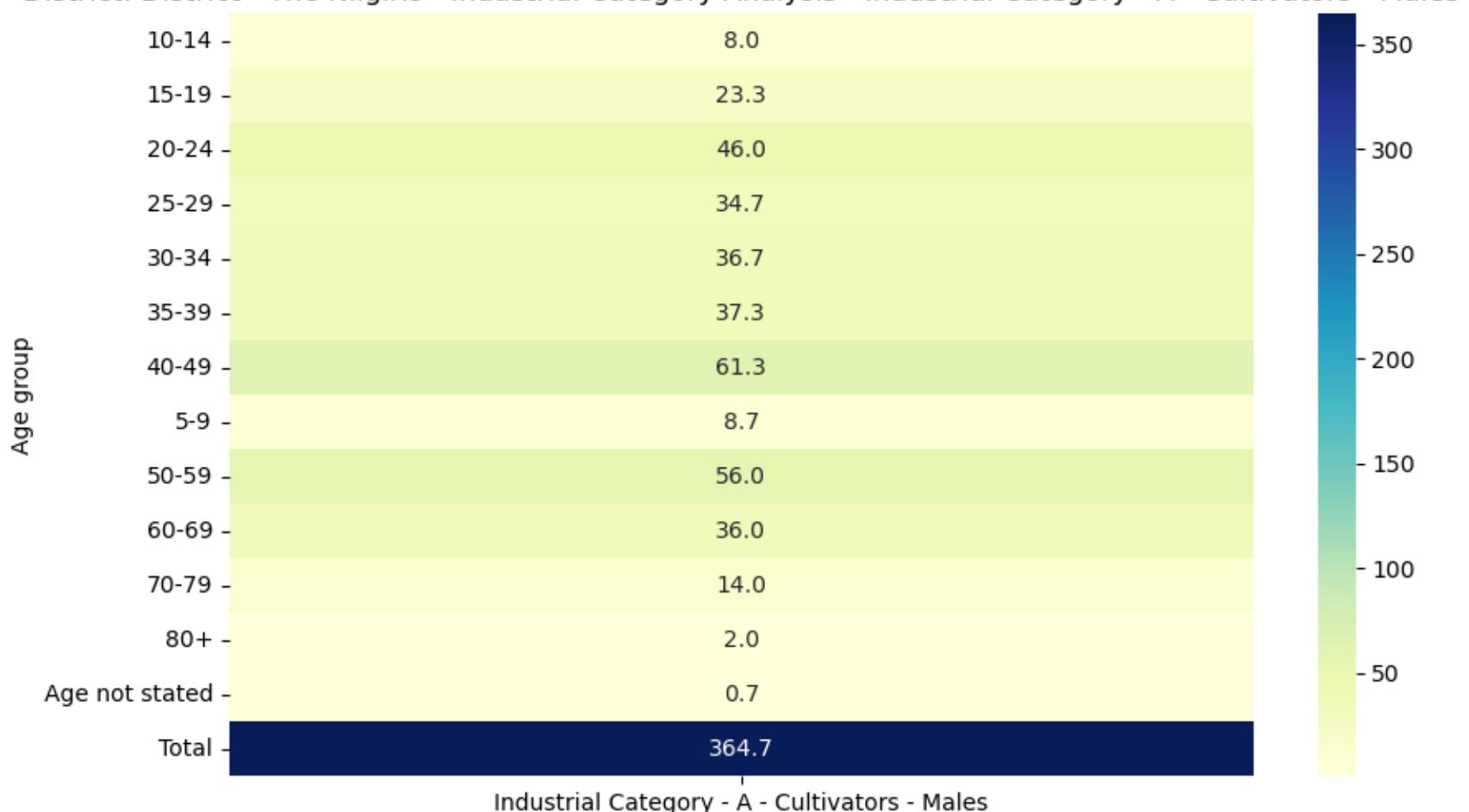
District: District - Erode - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



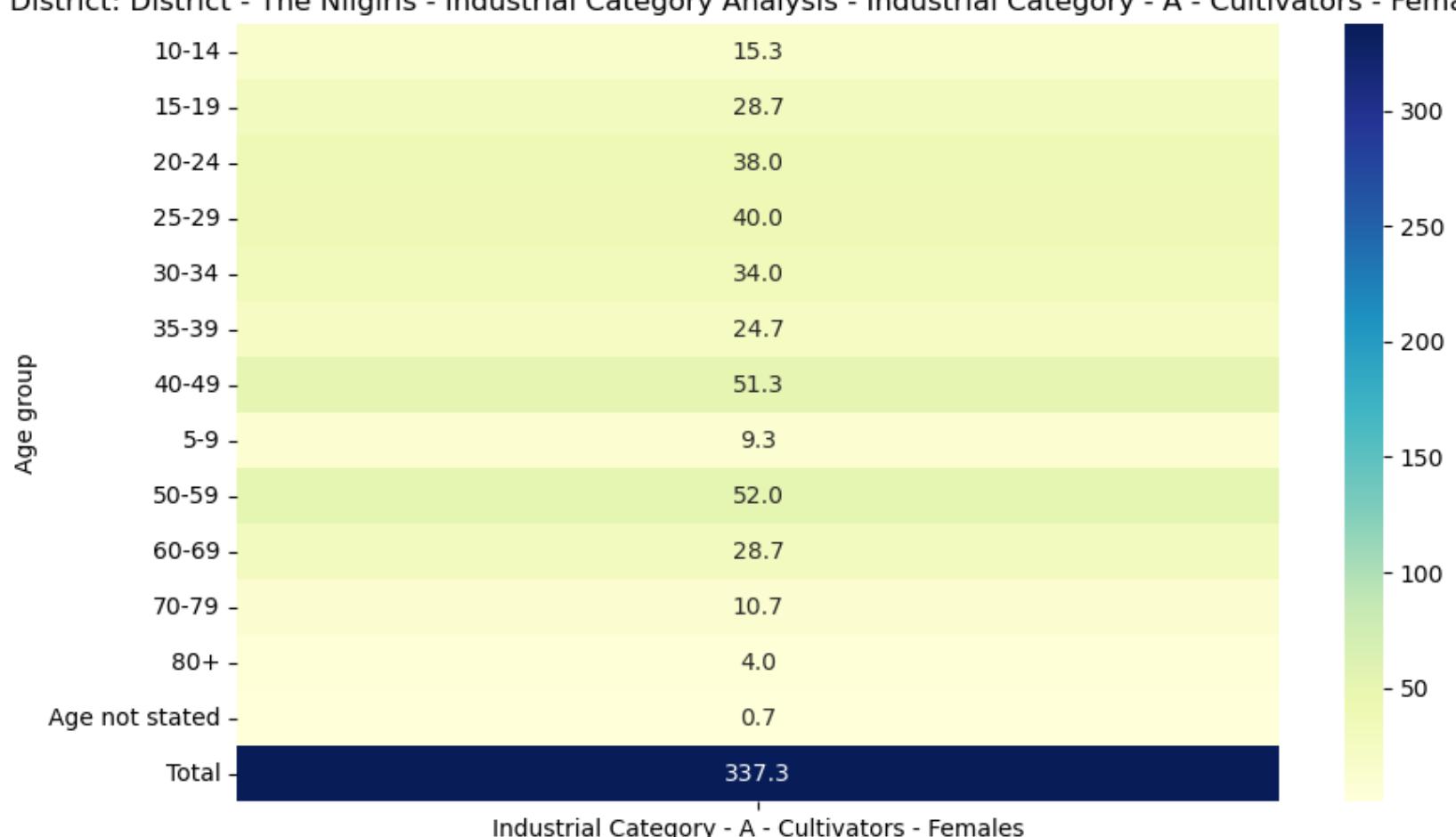
District: District - The Nilgiris - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



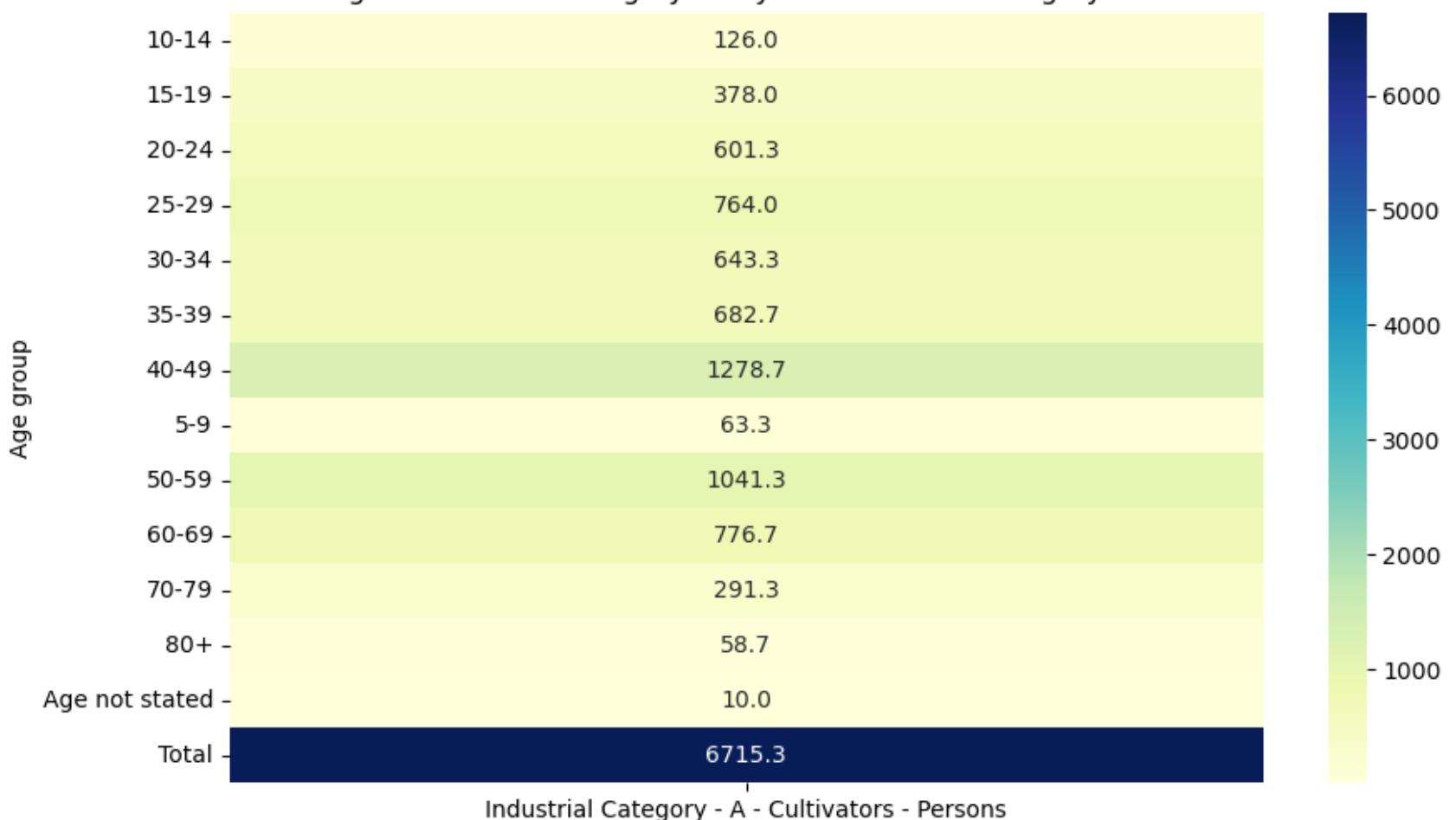
District: District - The Nilgiris - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



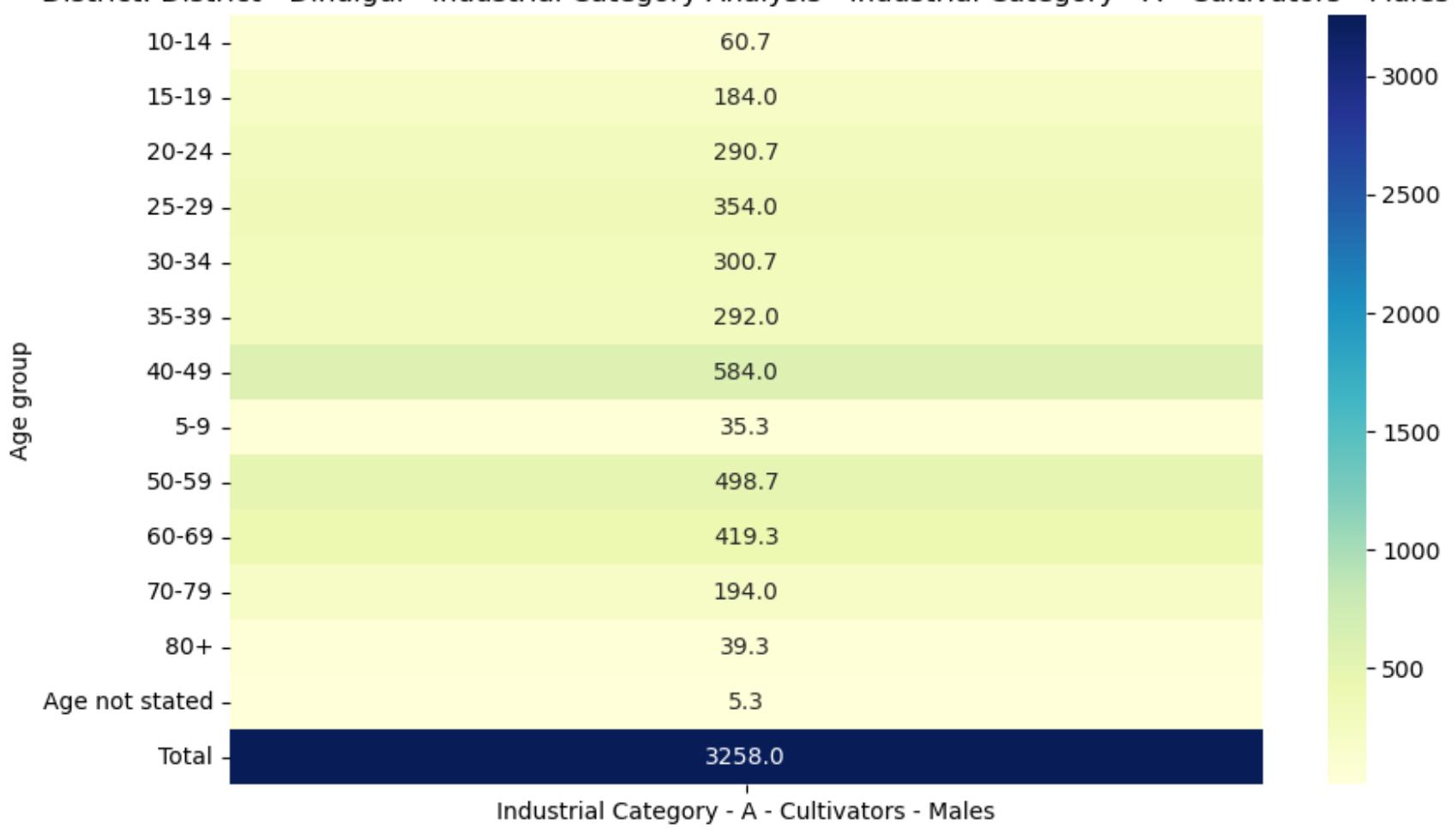
District: District - The Nilgiris - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



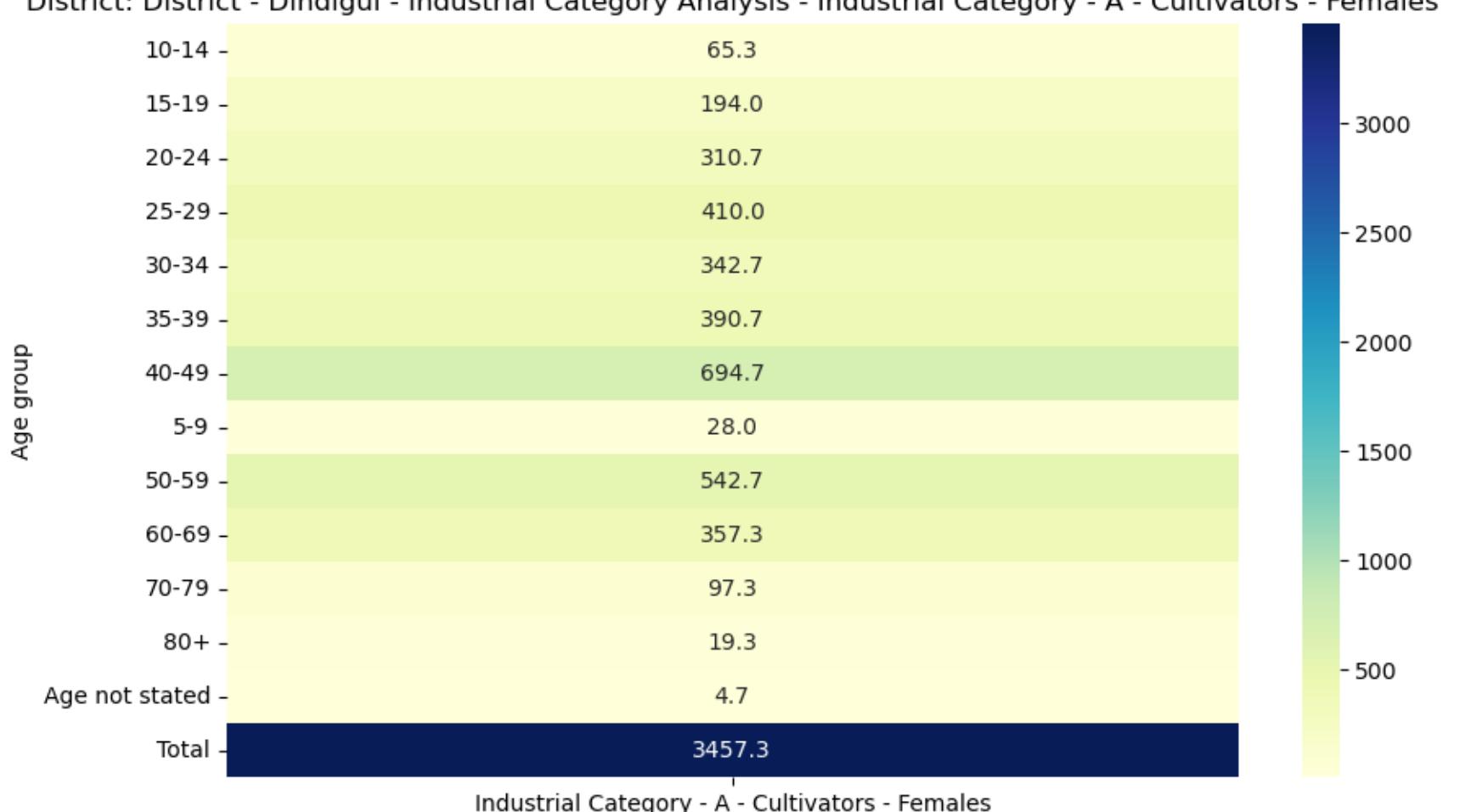
District: District - Dindigul - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



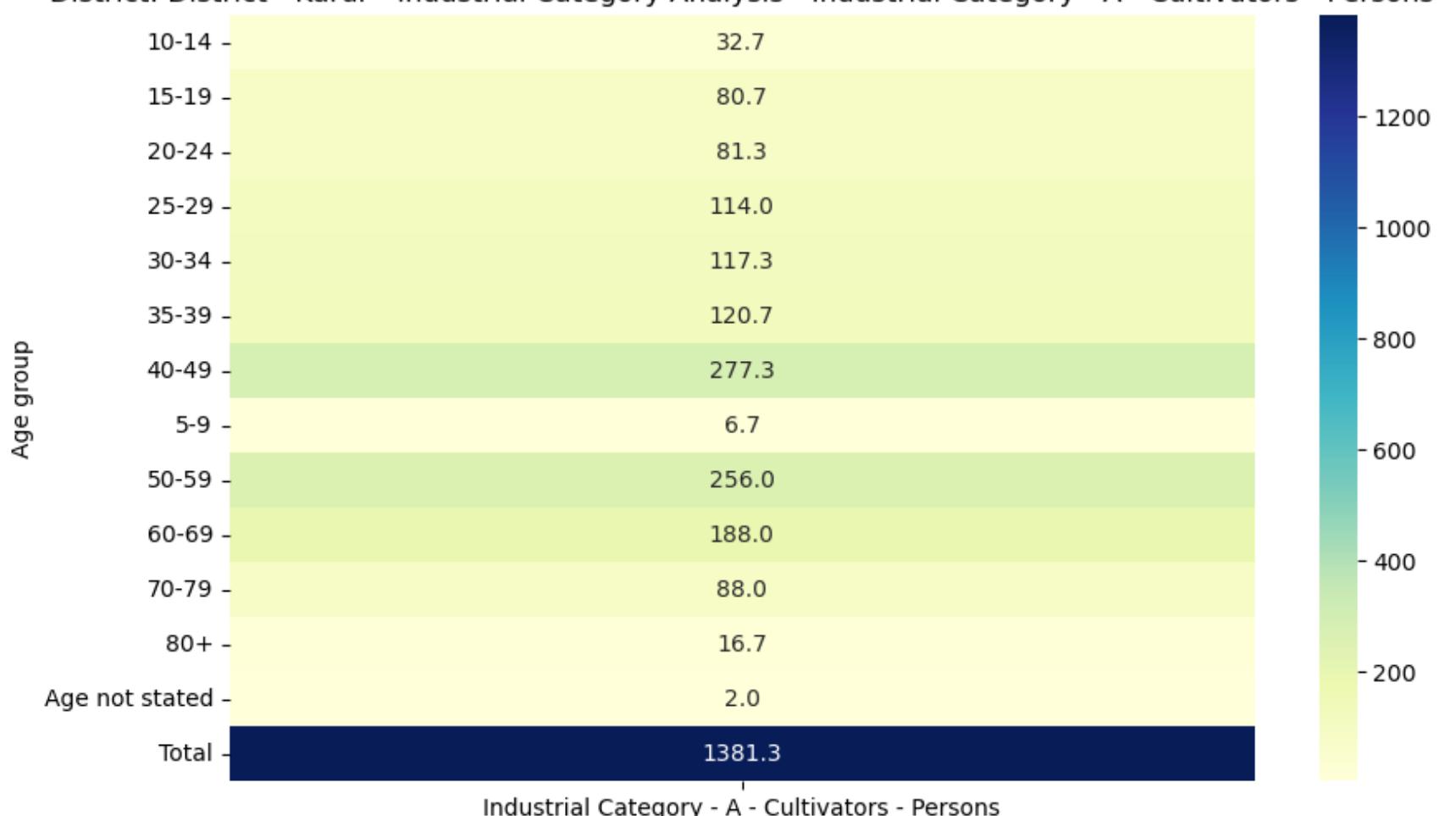
District: District - Dindigul - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



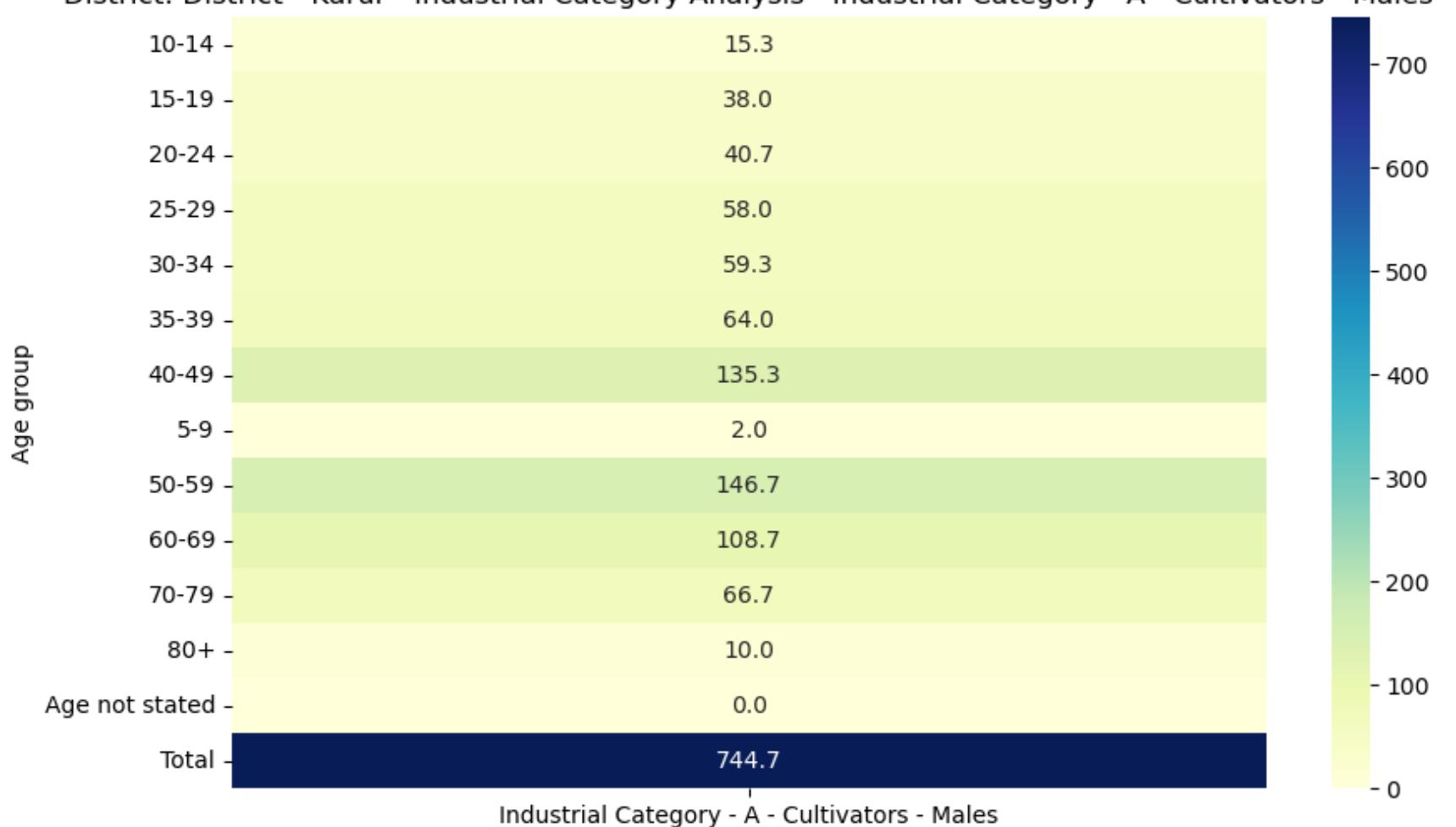
District: District - Dindigul - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



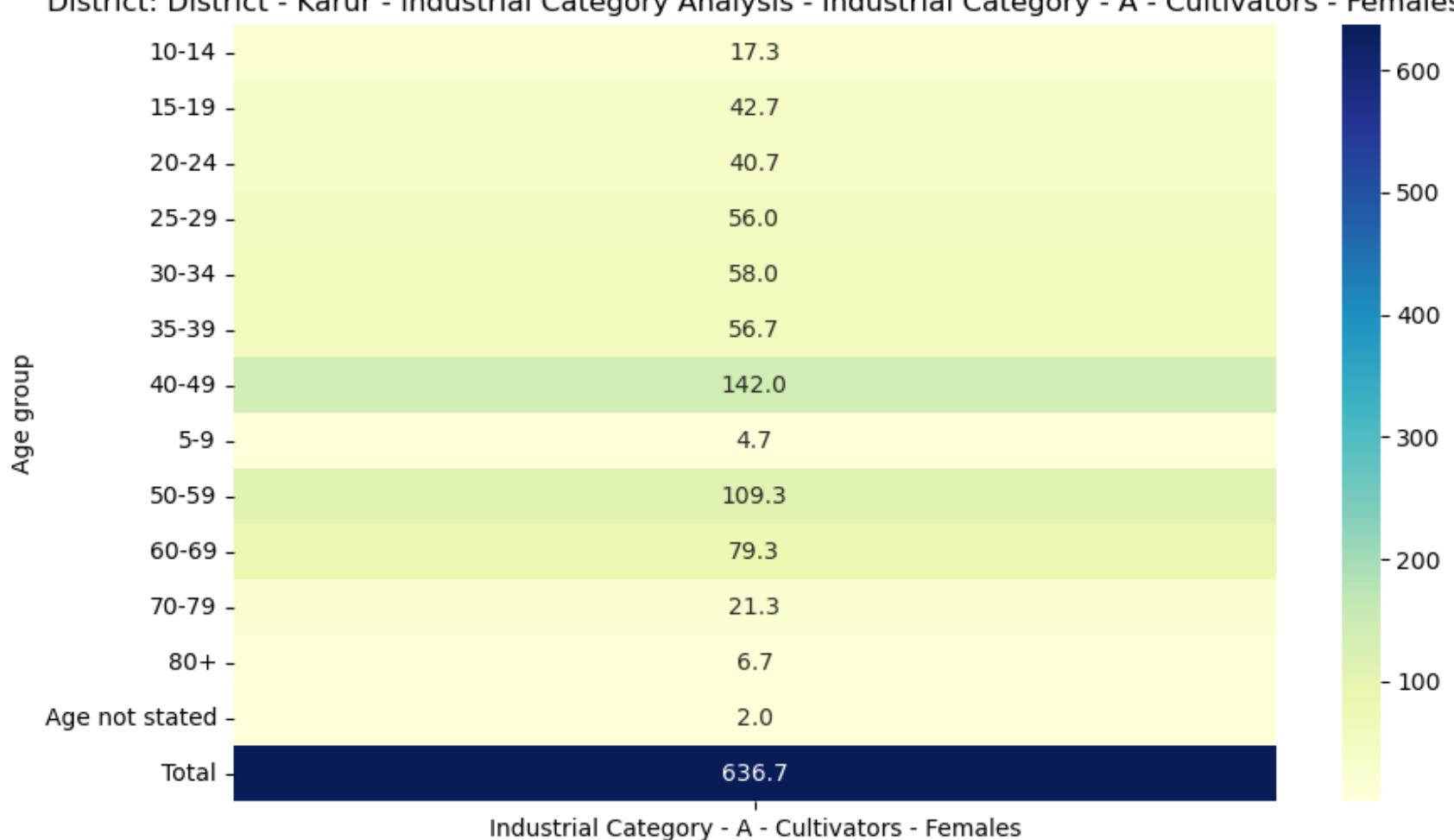
District: District - Karur - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



District: District - Karur - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

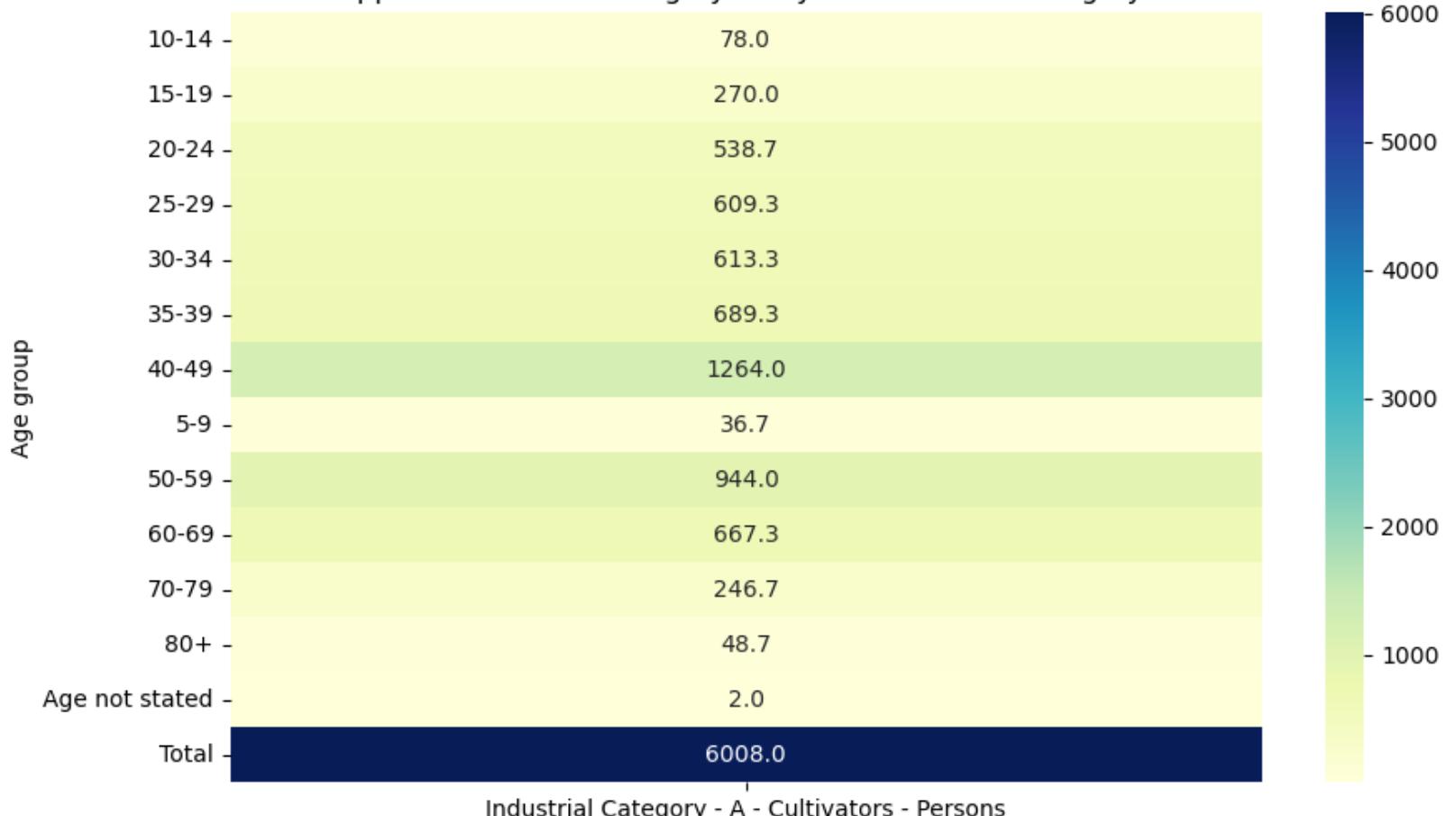


District: District - Karur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



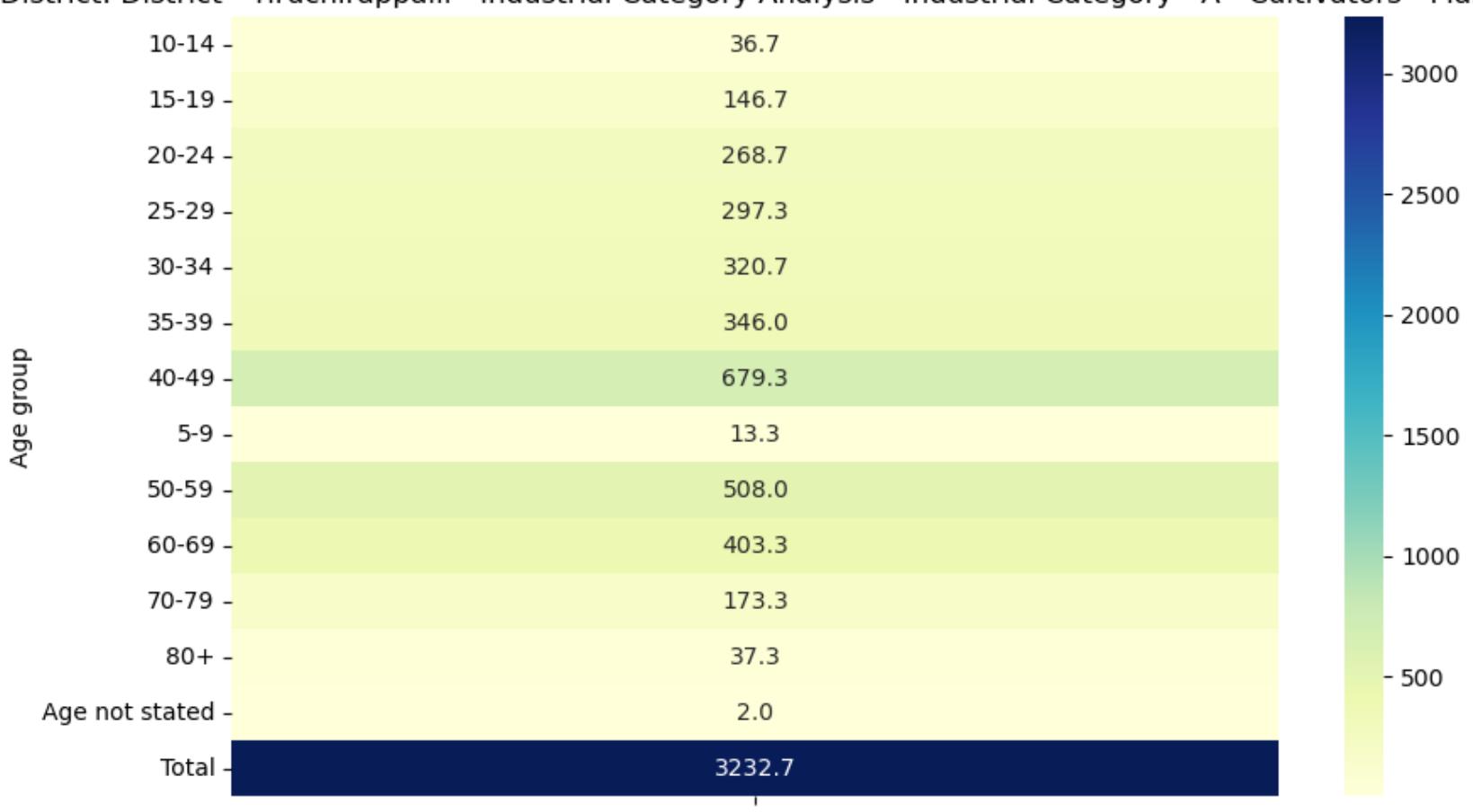
Industrial Category - A - Cultivators - Females

District: District - Tiruchirappalli - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons

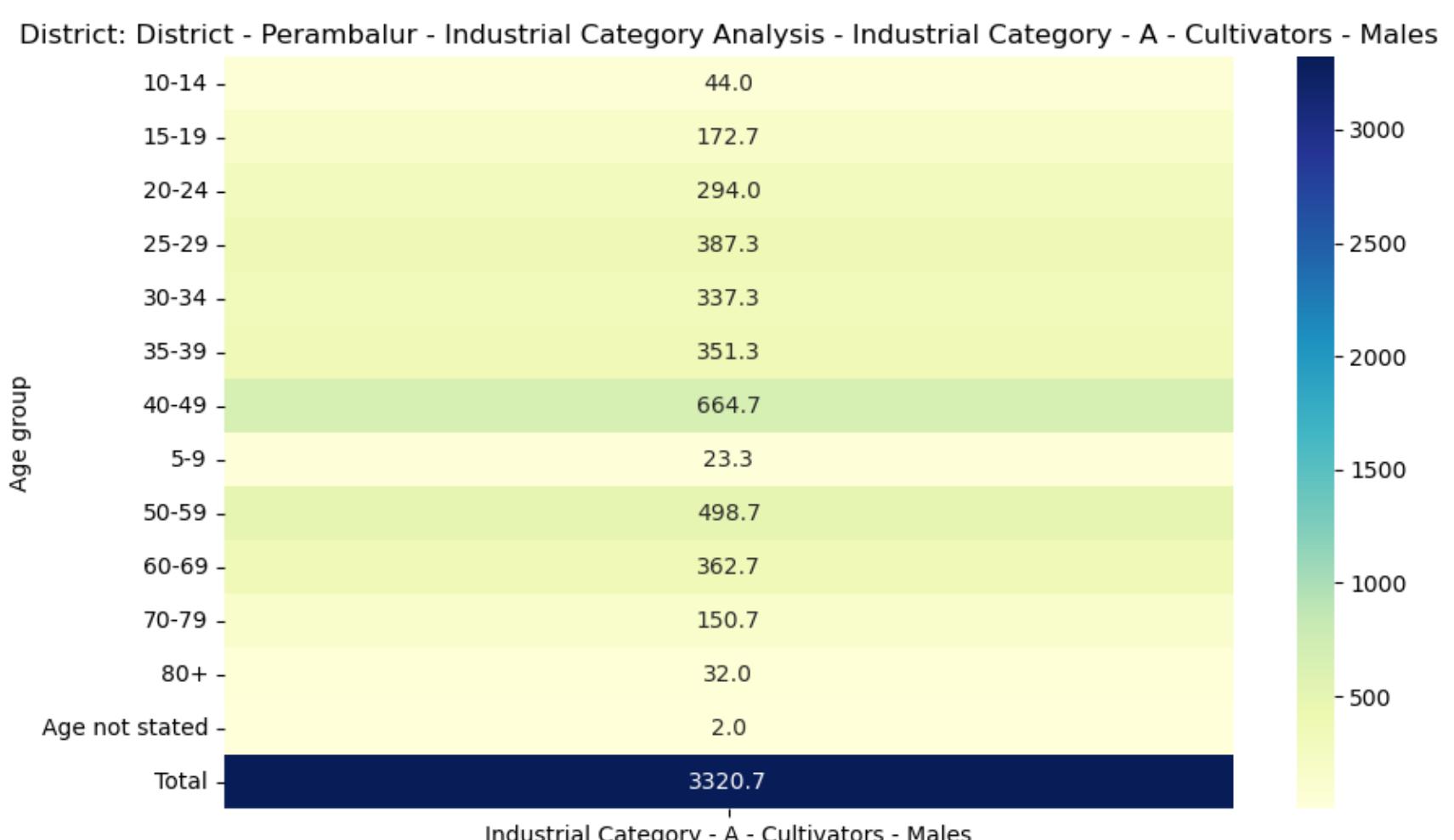
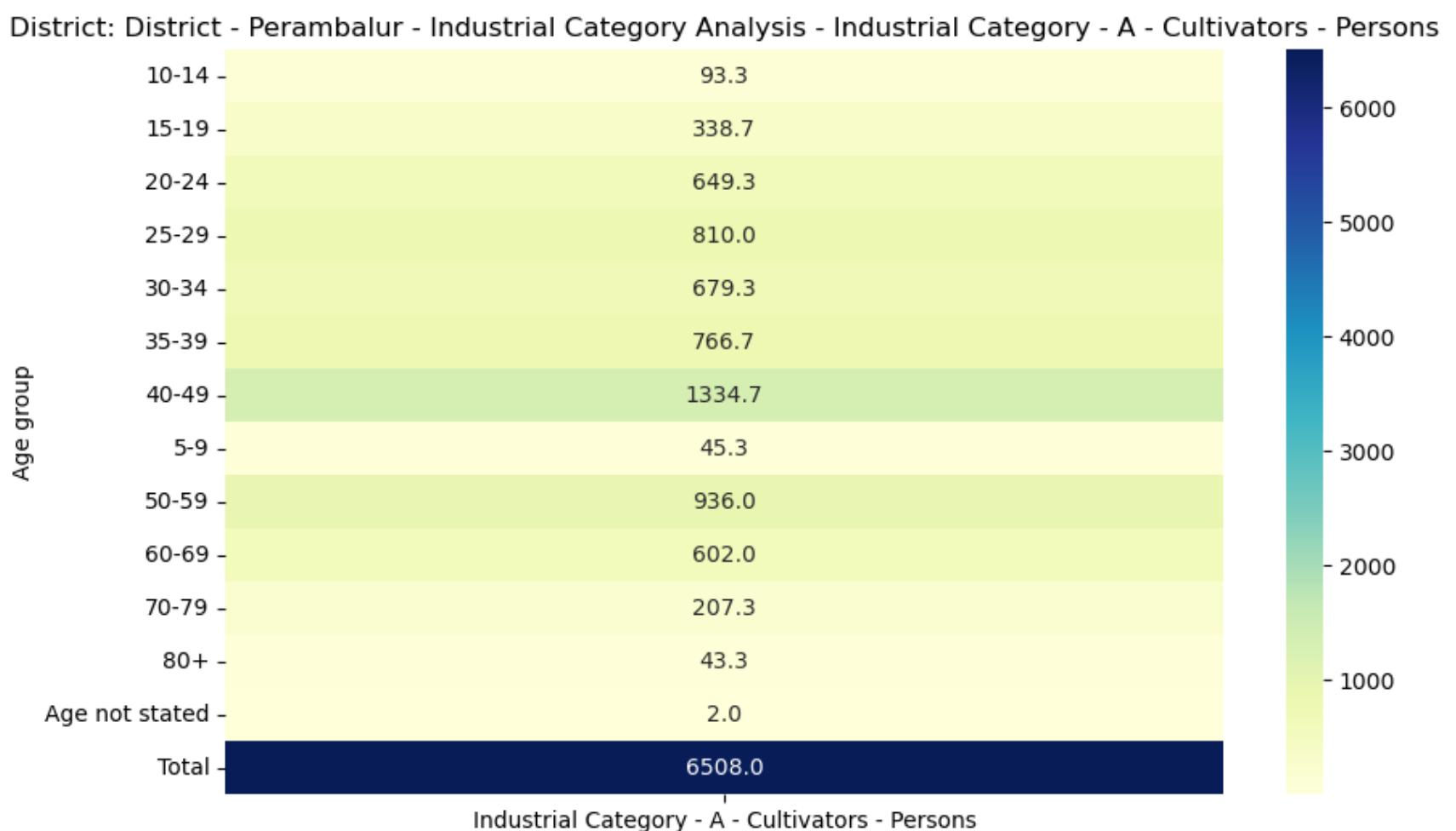
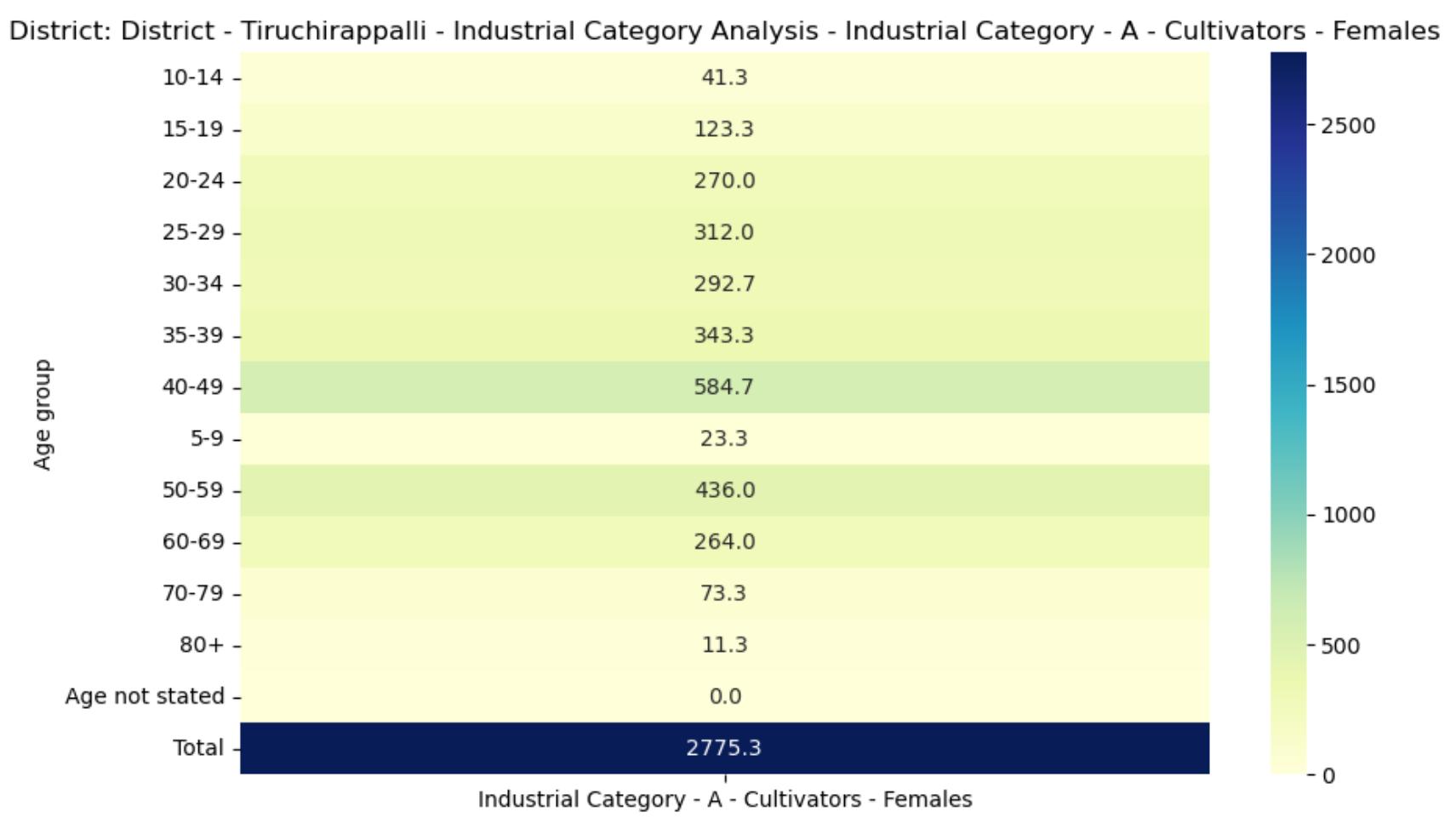


Industrial Category - A - Cultivators - Persons

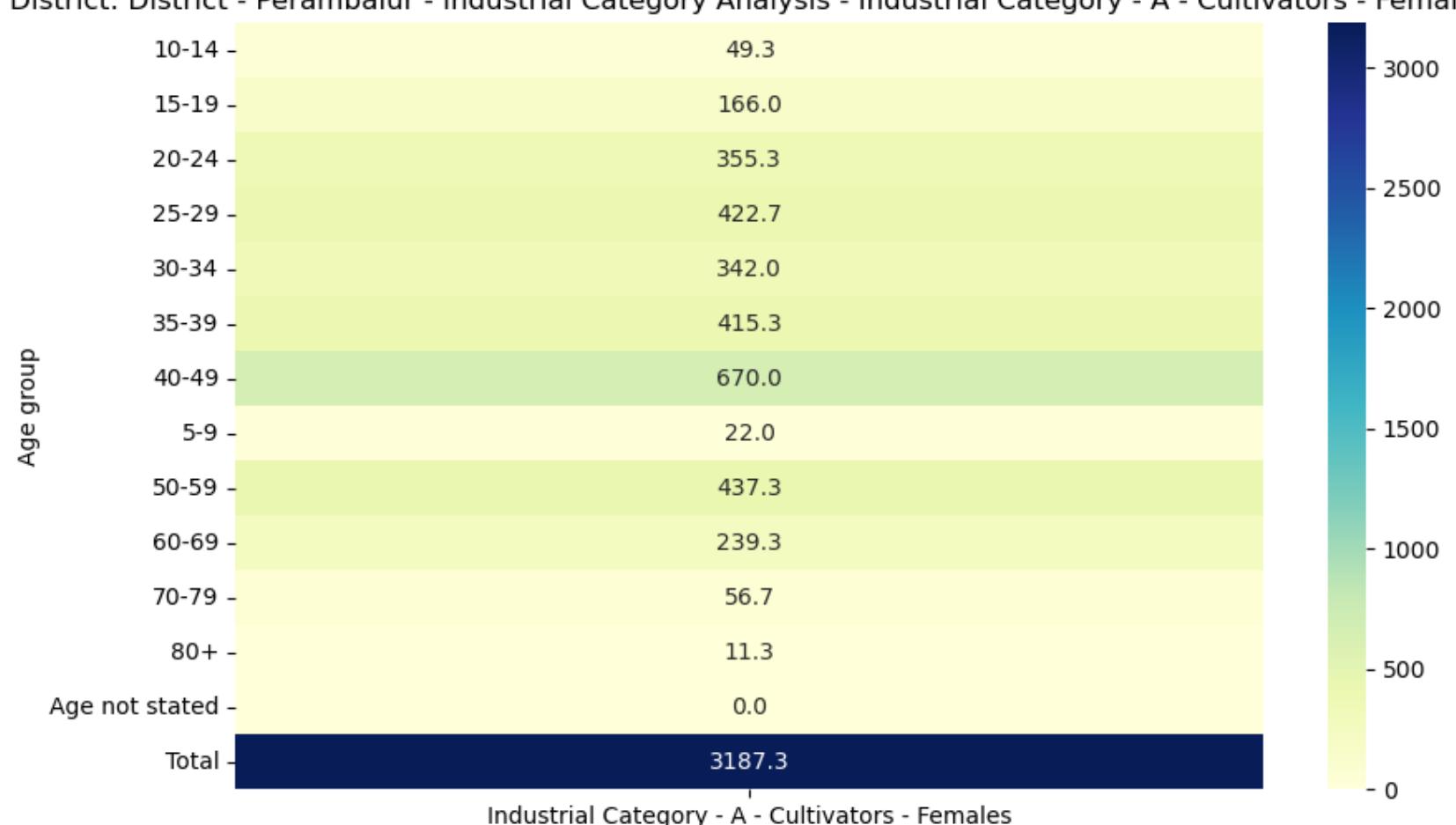
District: District - Tiruchirappalli - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



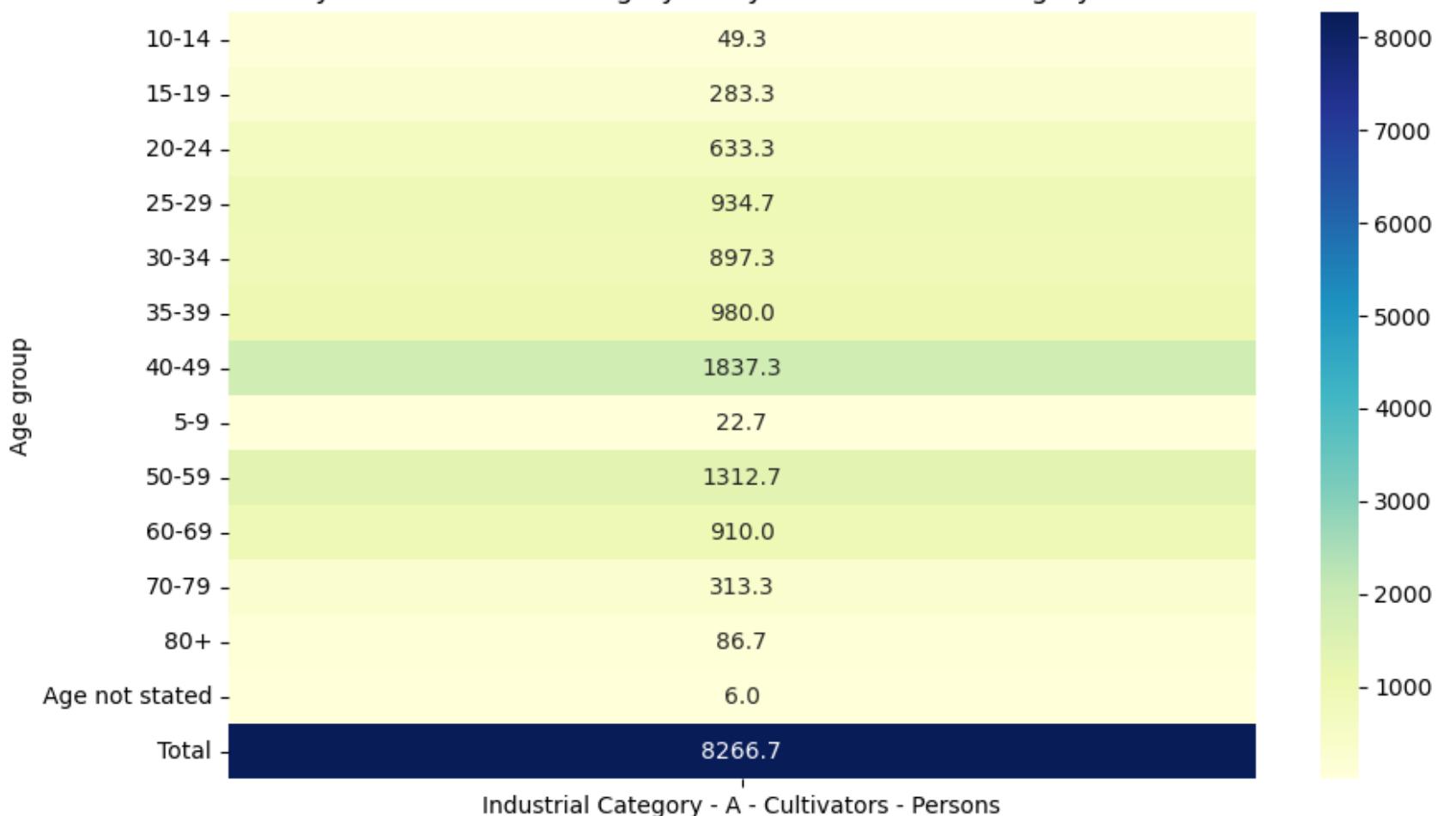
Industrial Category - A - Cultivators - Males



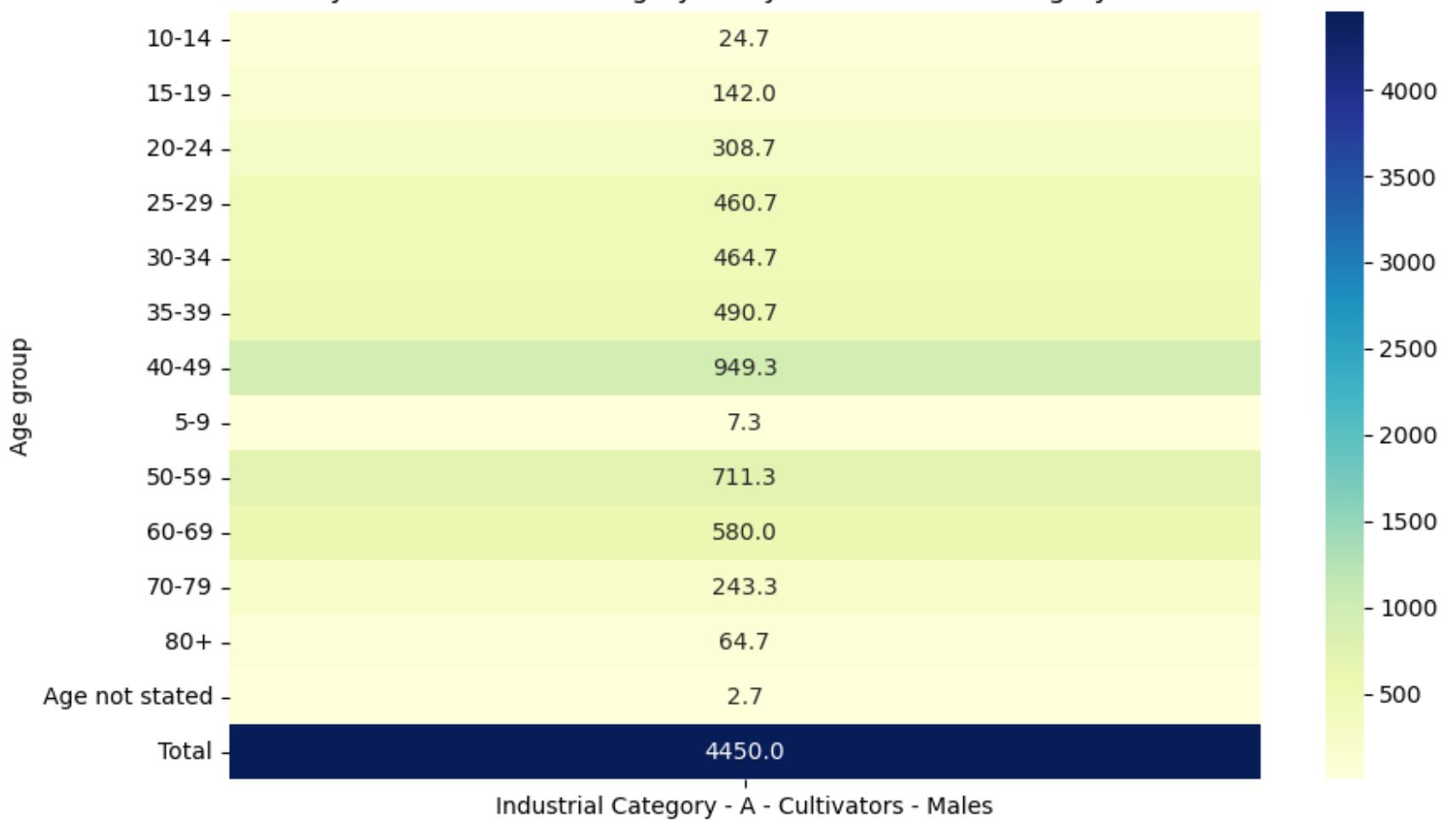
District: District - Perambalur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



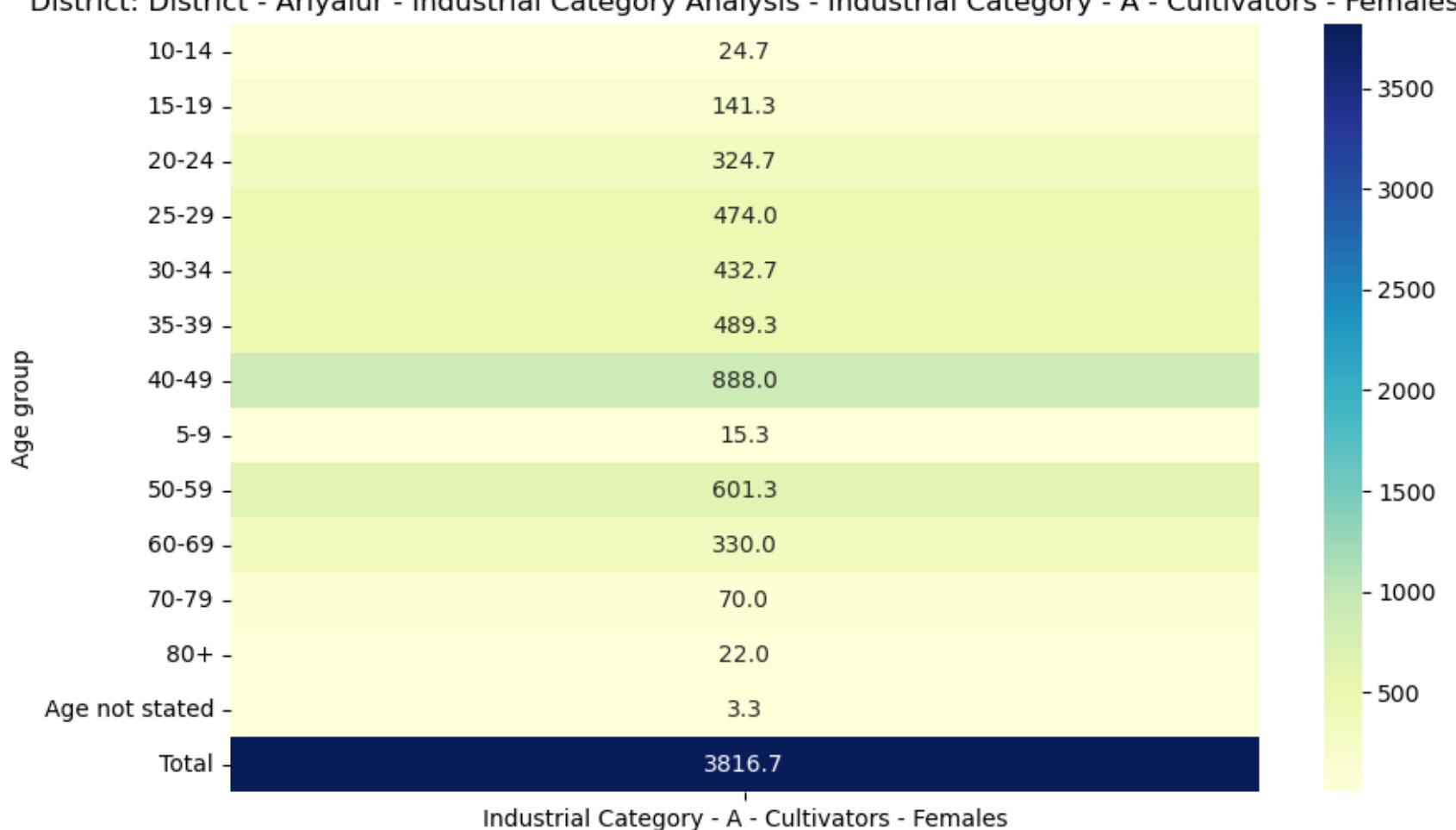
District: District - Ariyalur - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



District: District - Ariyalur - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

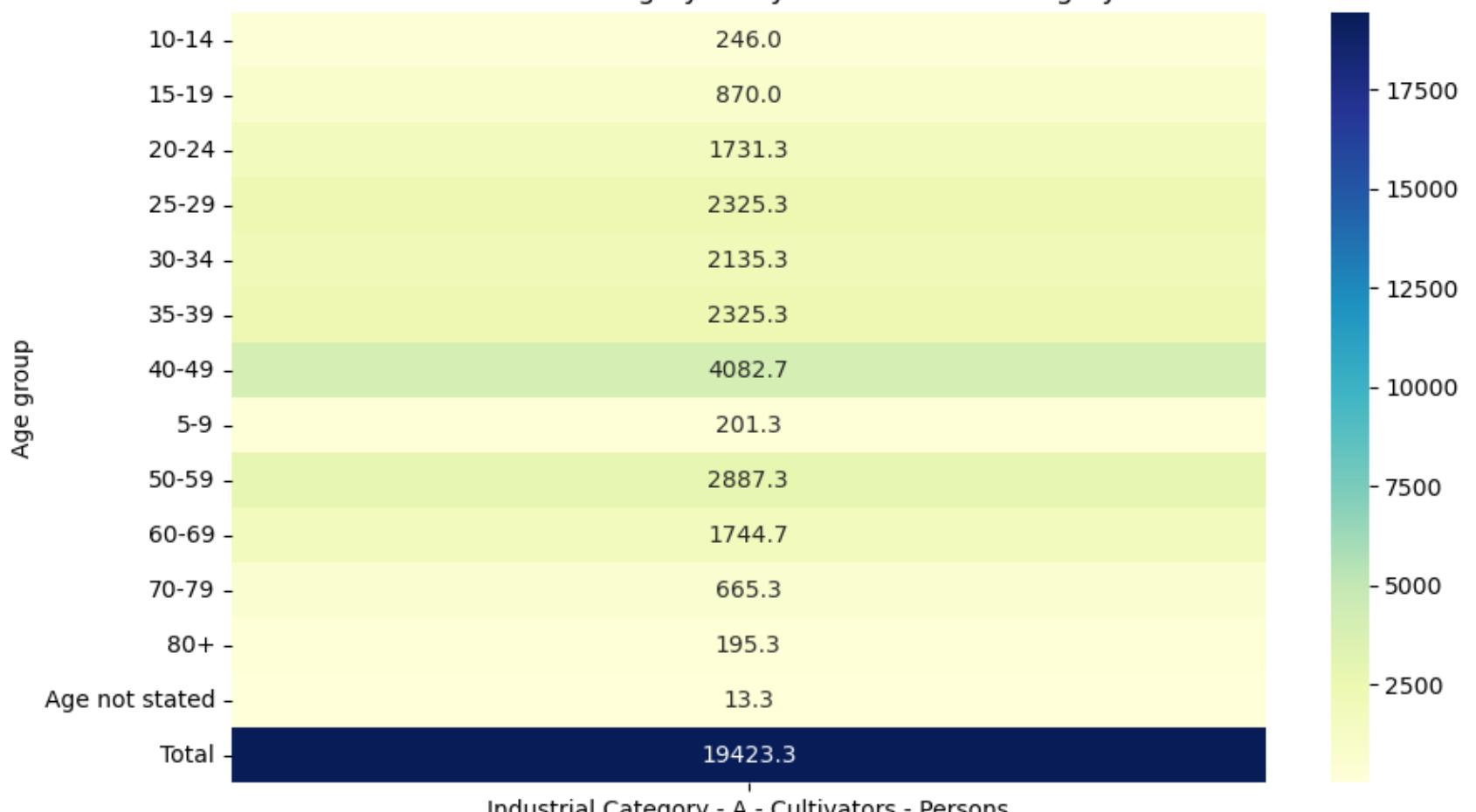


District: District - Ariyalur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



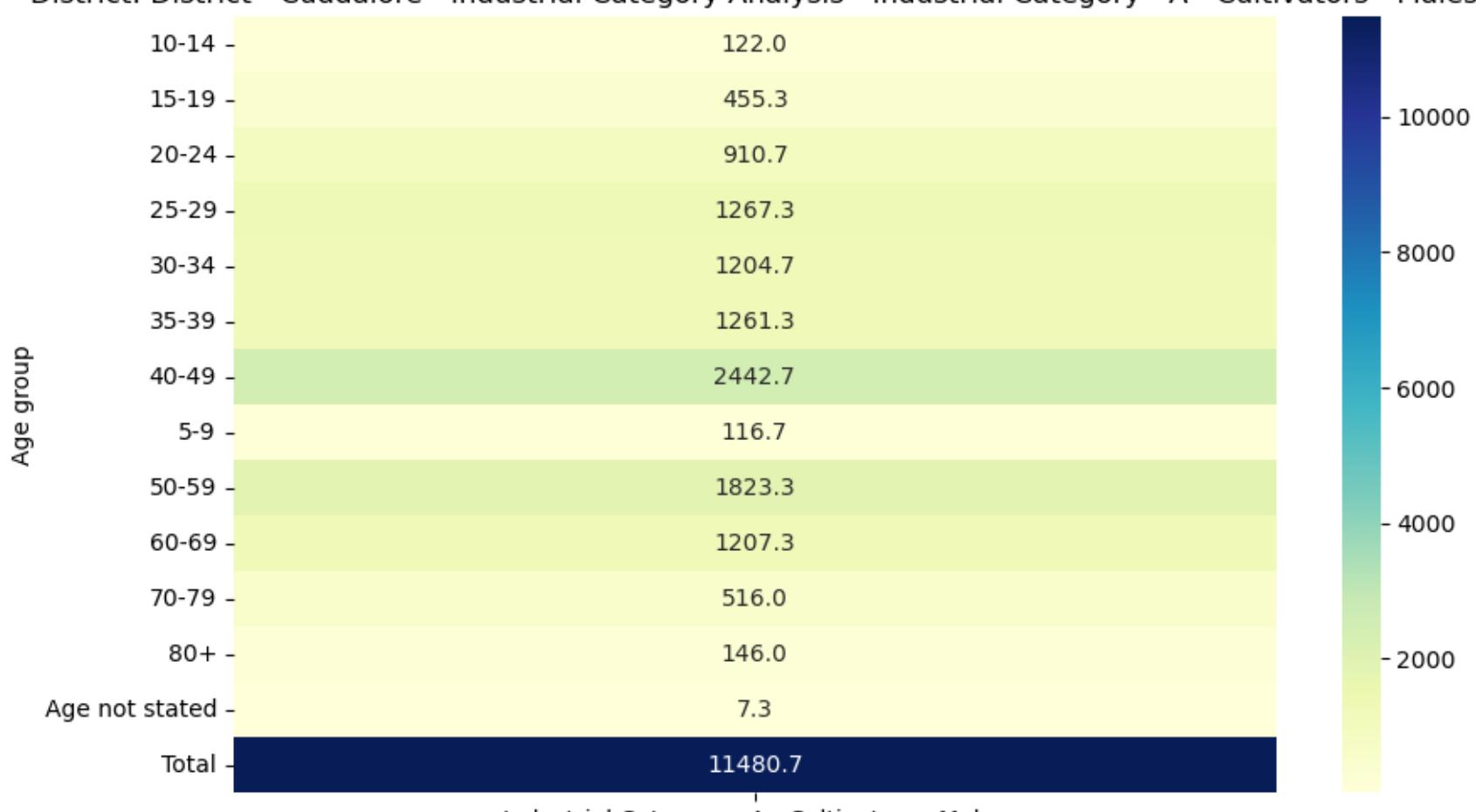
Industrial Category - A - Cultivators - Females

District: District - Cuddalore - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



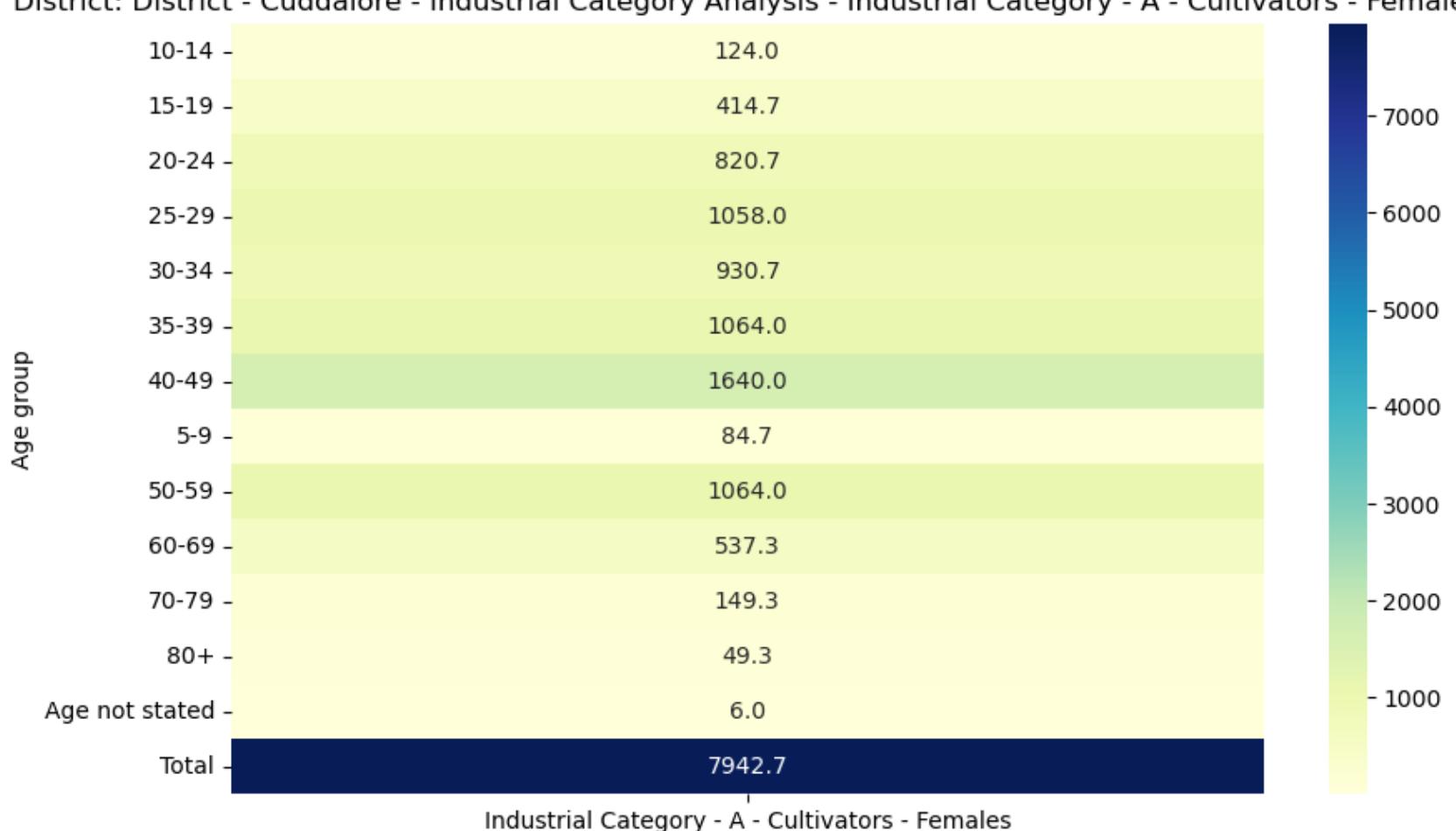
Industrial Category - A - Cultivators - Persons

District: District - Cuddalore - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



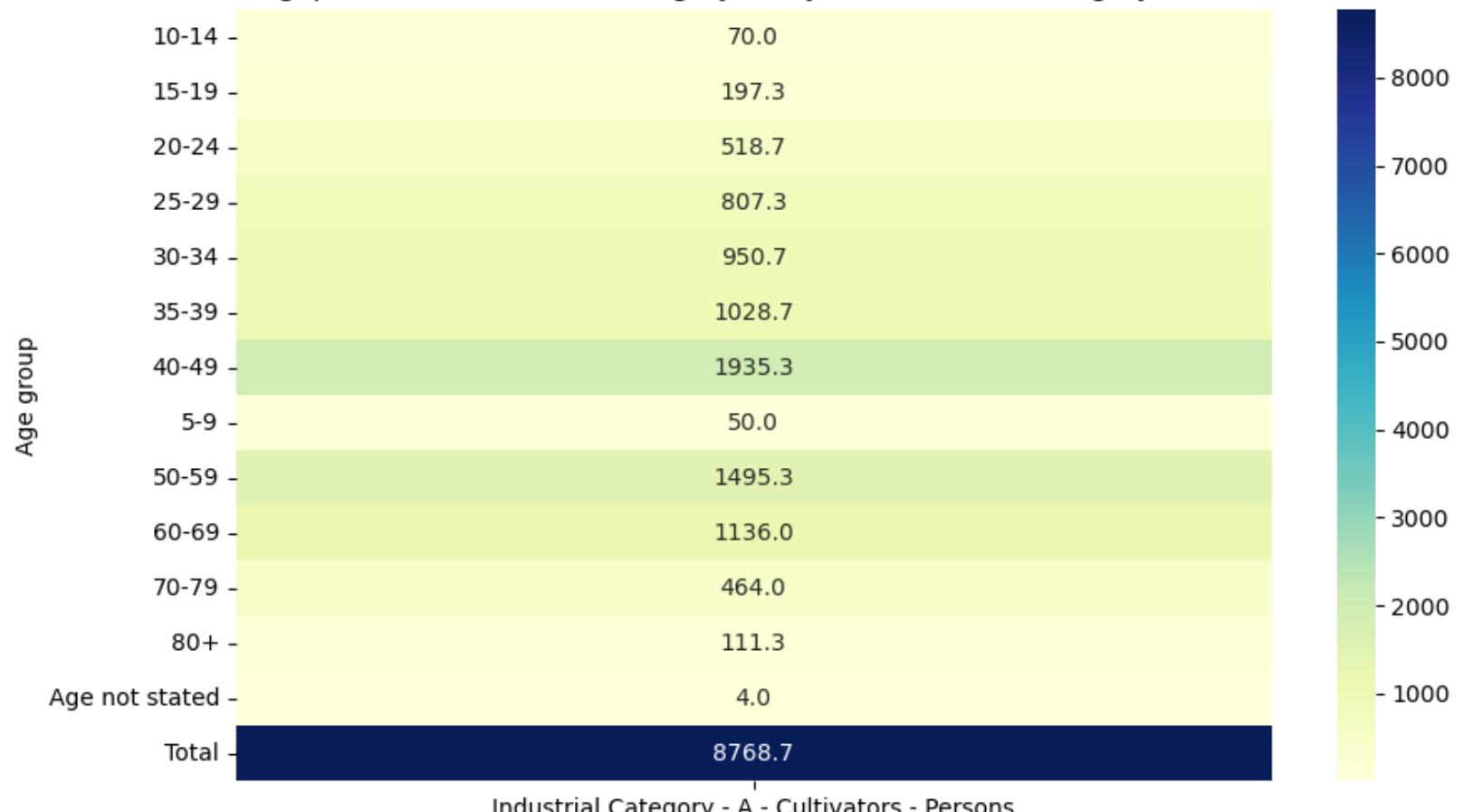
Industrial Category - A - Cultivators - Males

District: District - Cuddalore - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



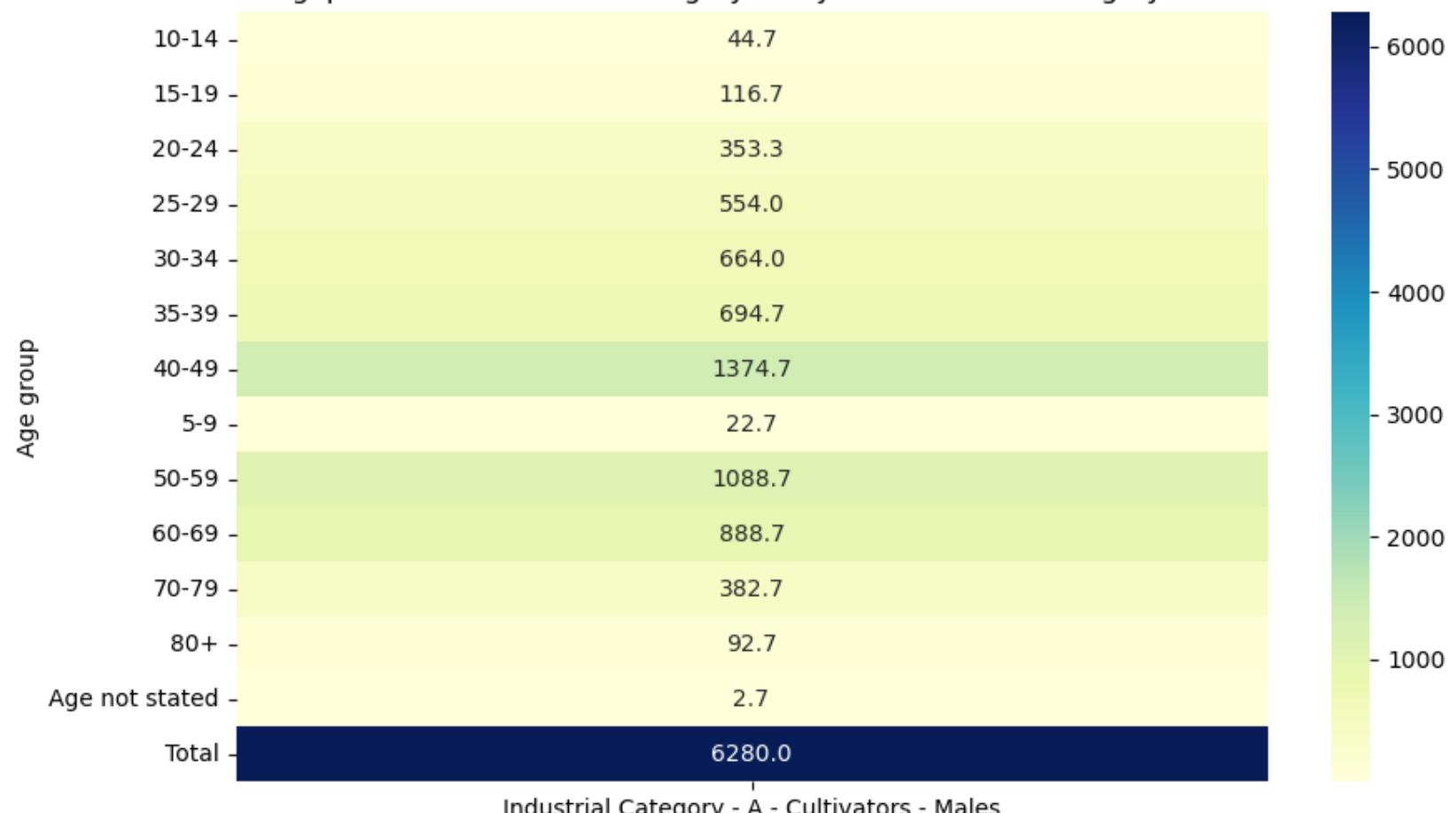
Industrial Category - A - Cultivators - Females

District: District - Nagapattinam - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons

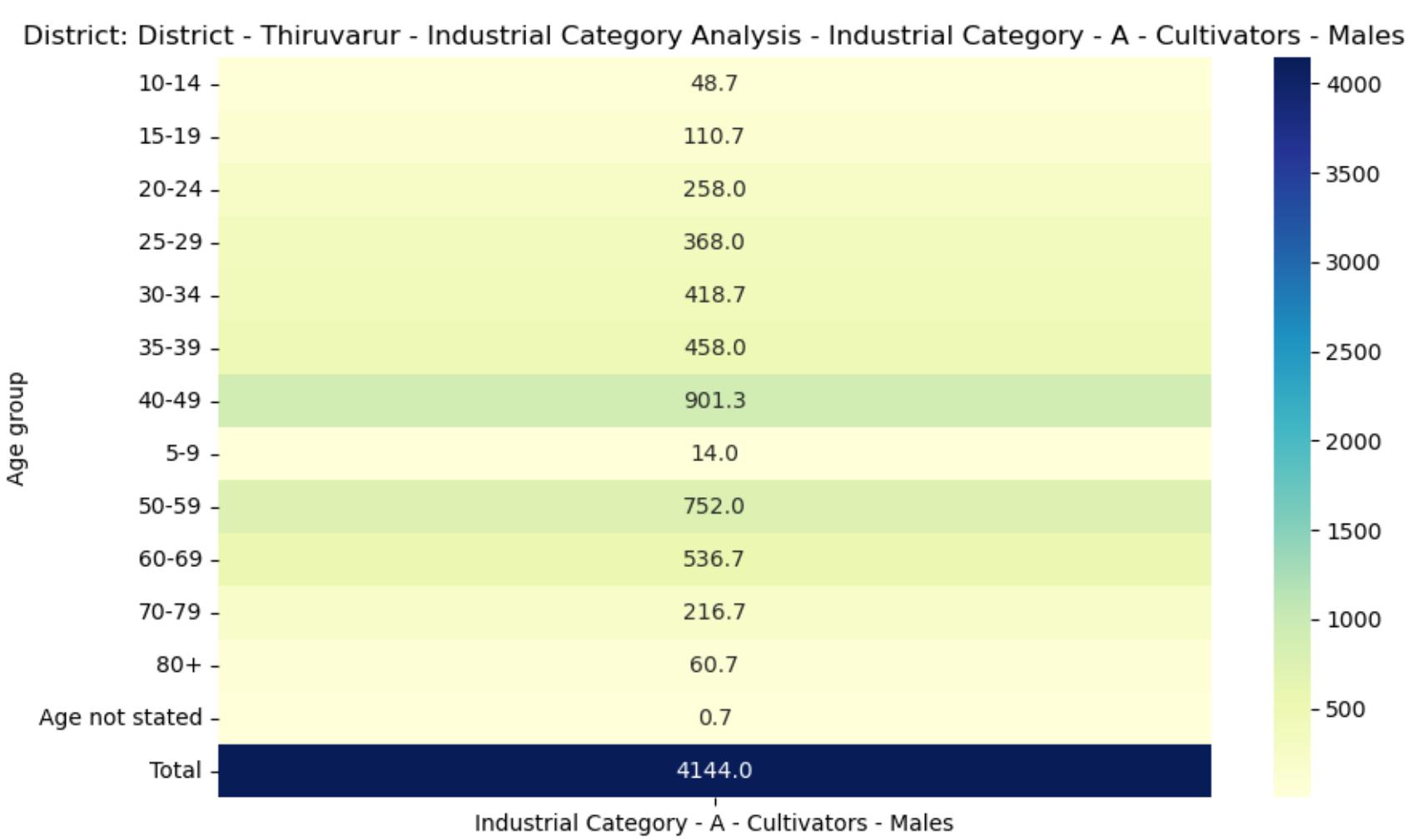
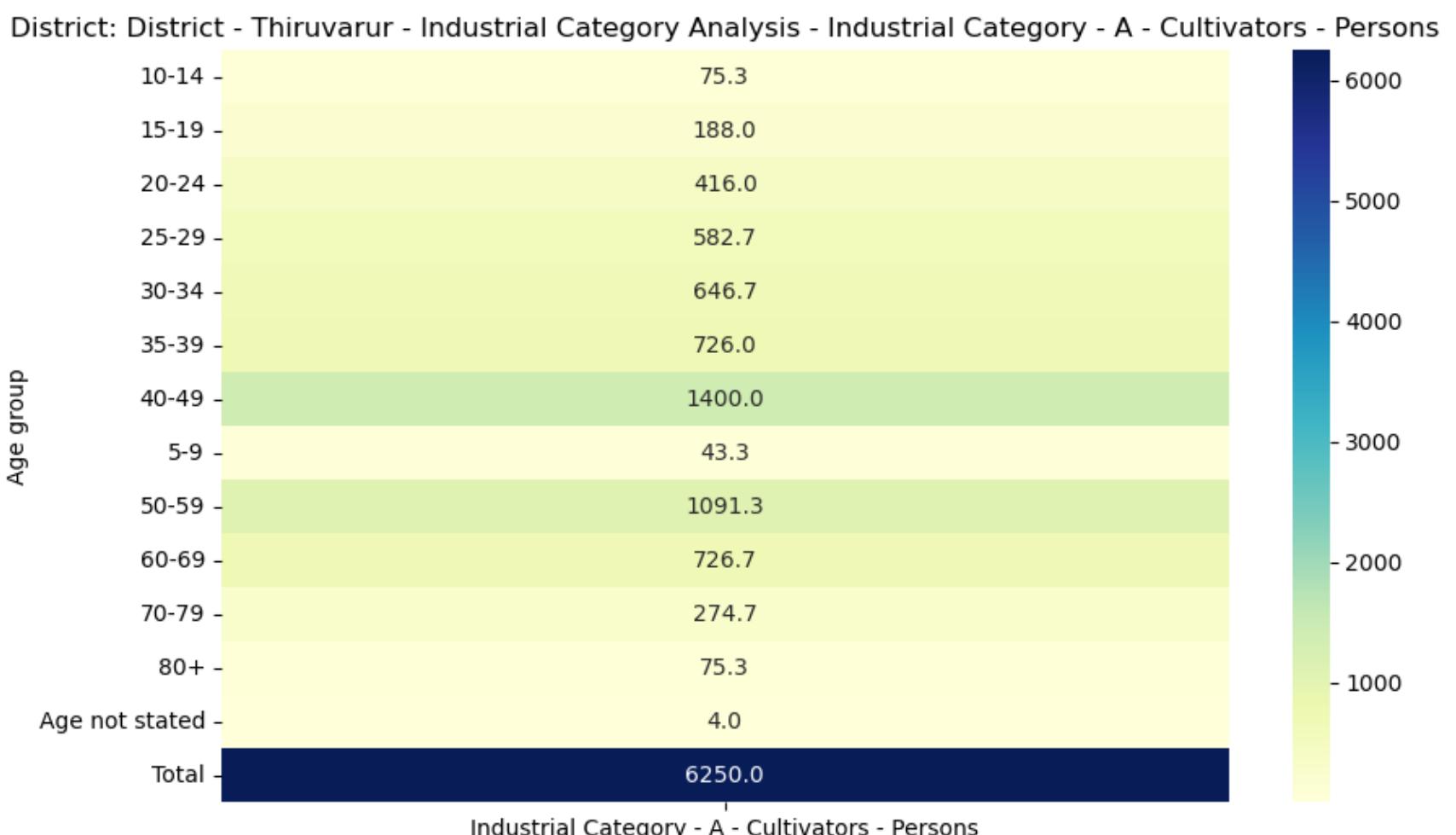
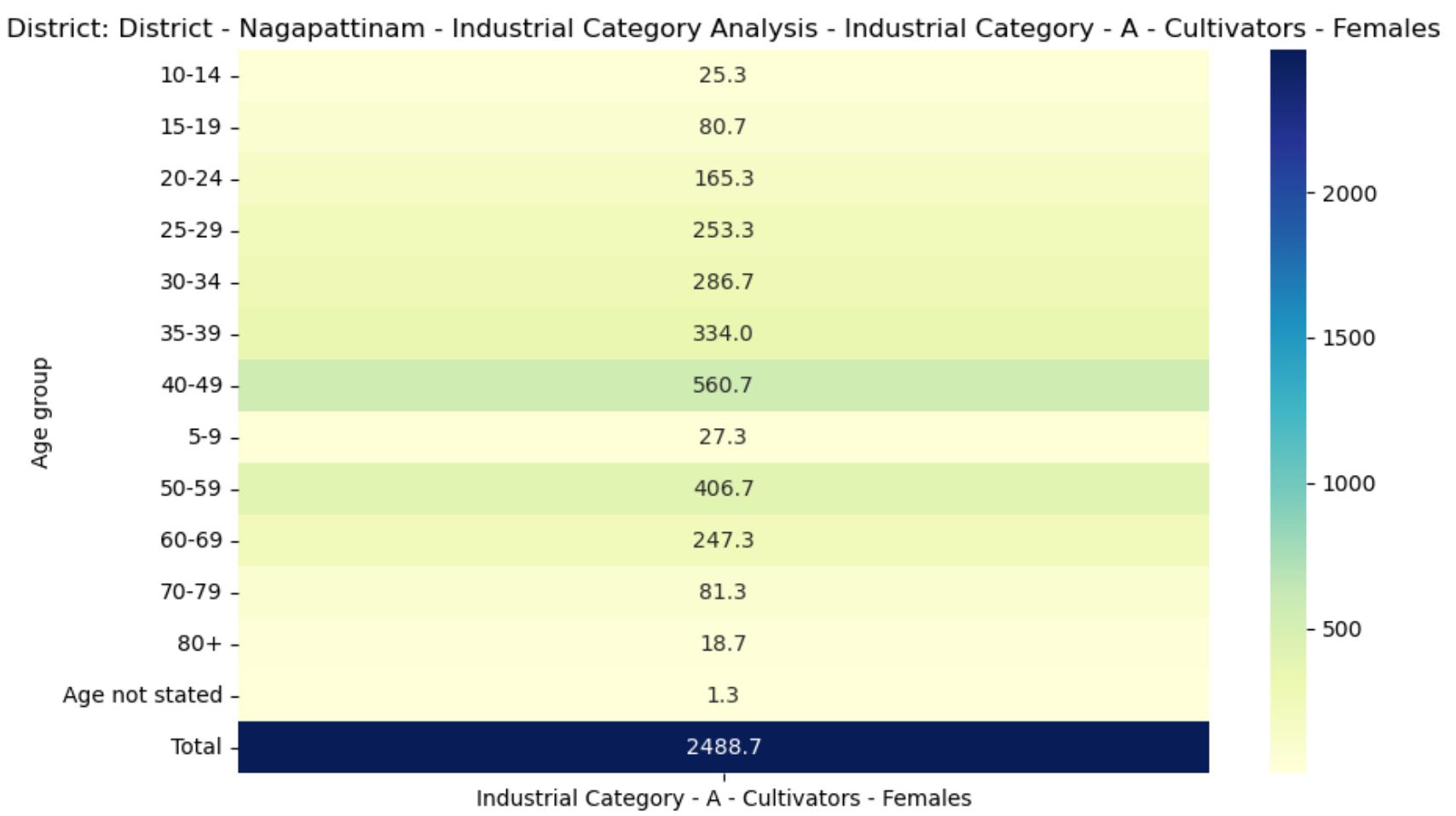


Industrial Category - A - Cultivators - Persons

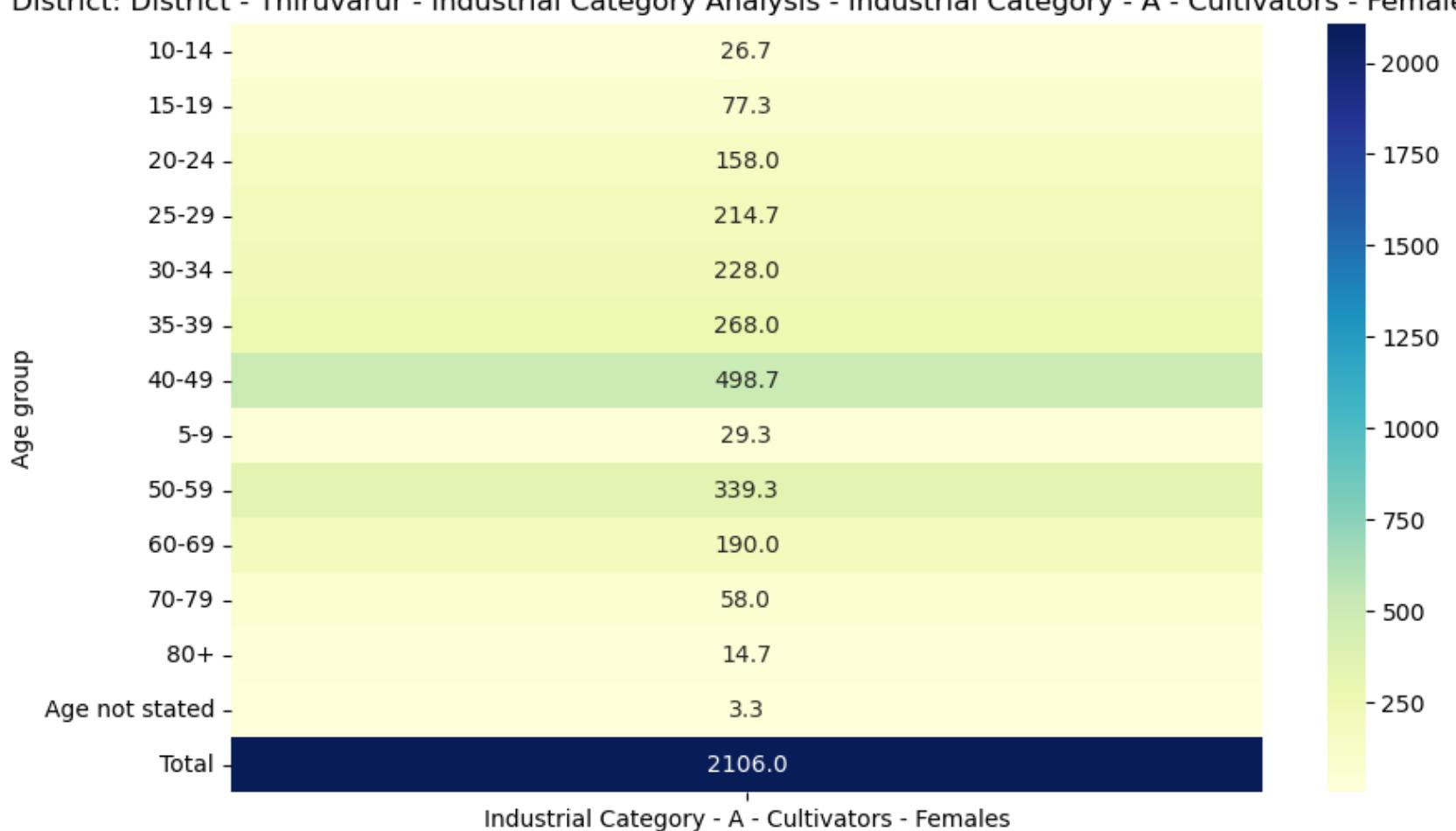
District: District - Nagapattinam - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



Industrial Category - A - Cultivators - Males

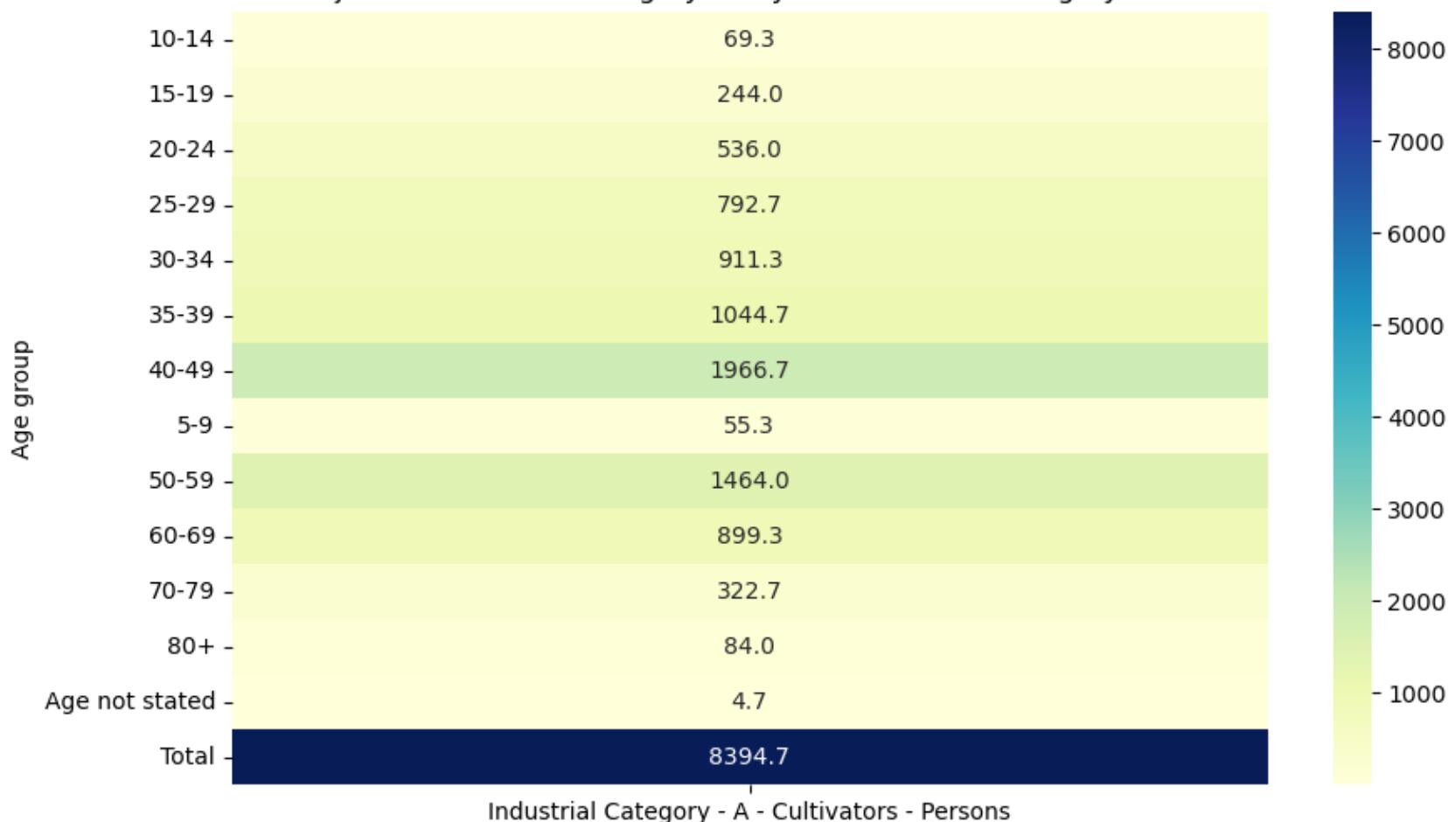


District: District - Thiruvarur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



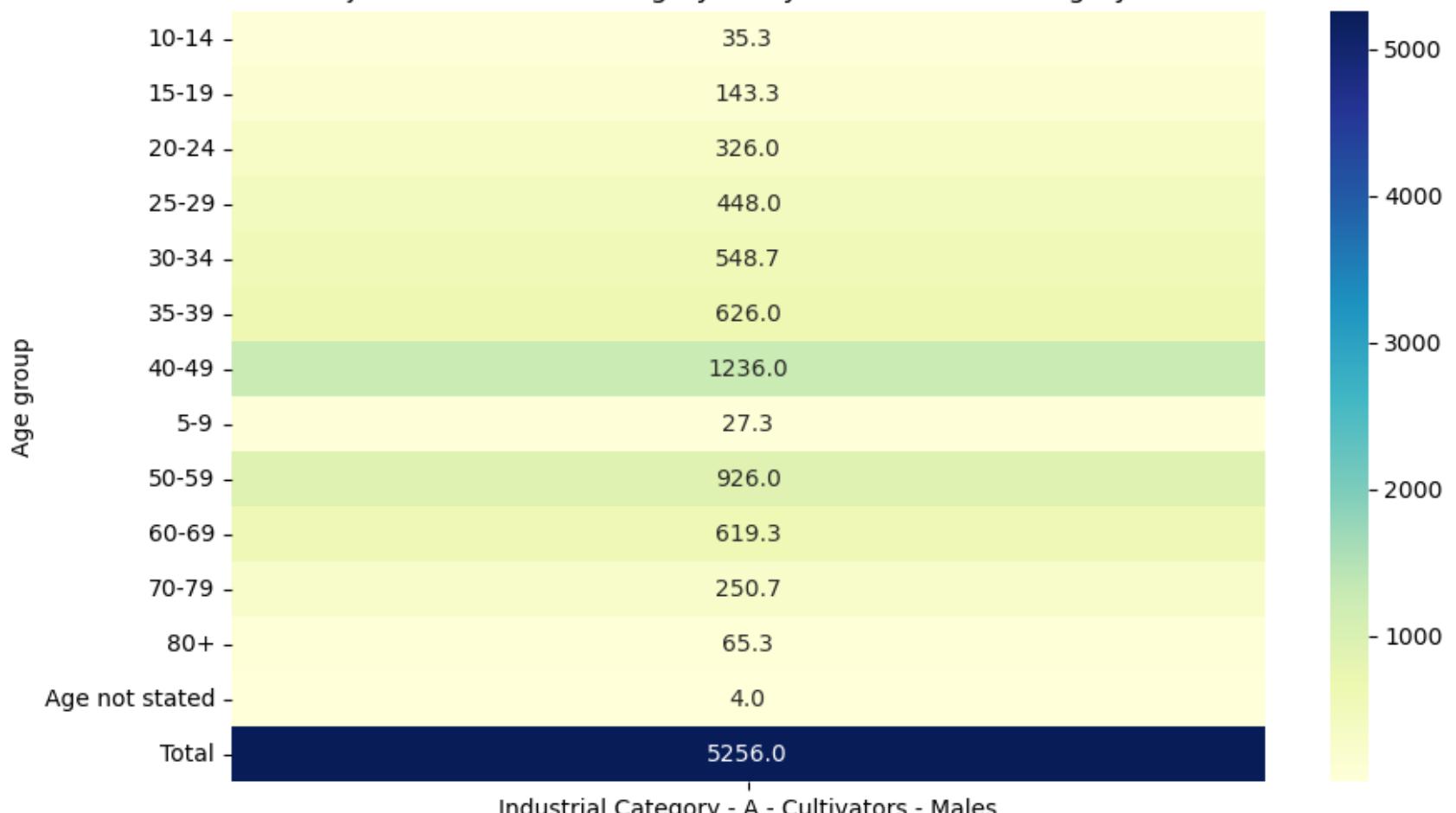
Industrial Category - A - Cultivators - Females

District: District - Thanjavur - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



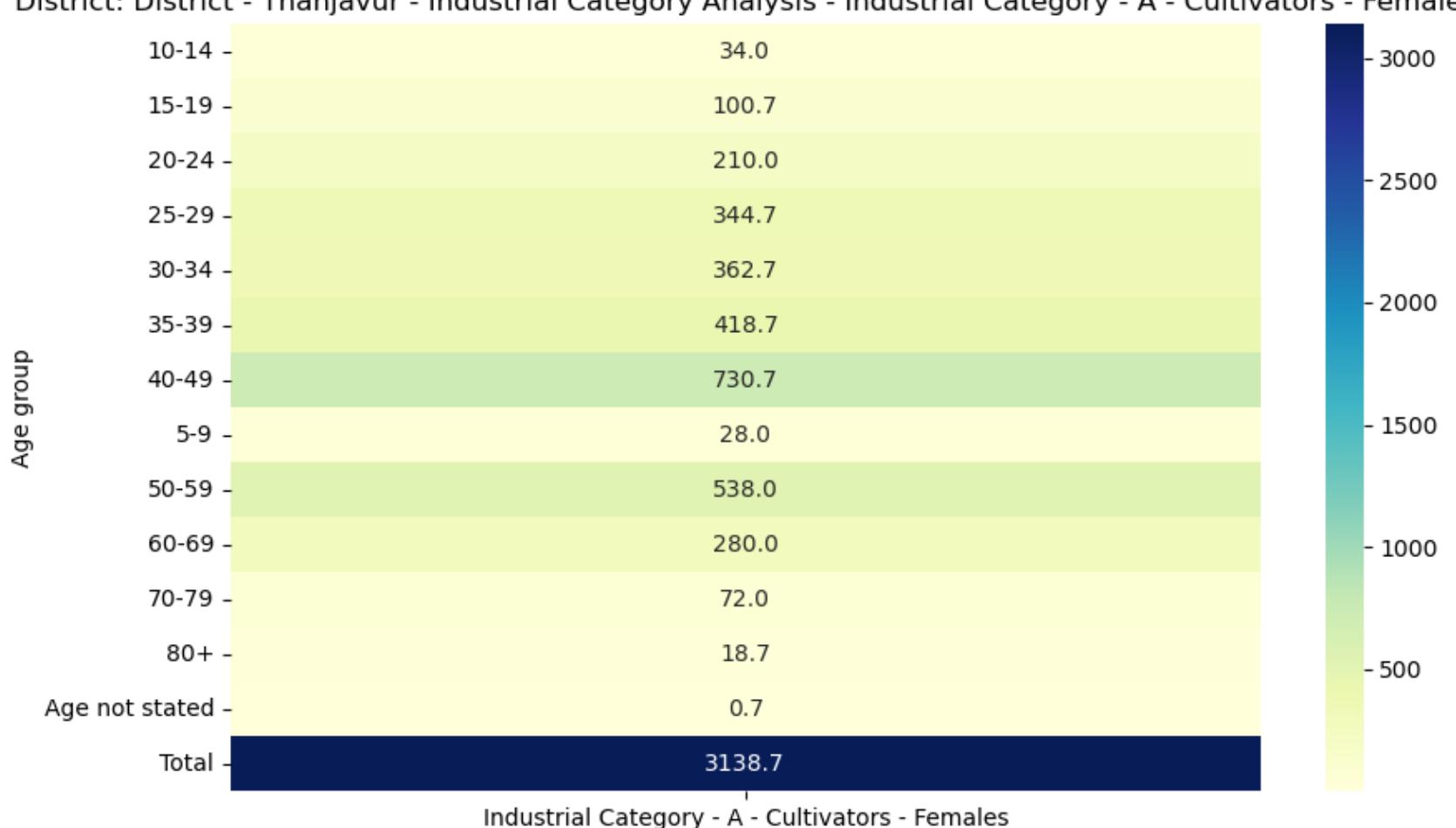
Industrial Category - A - Cultivators - Persons

District: District - Thanjavur - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

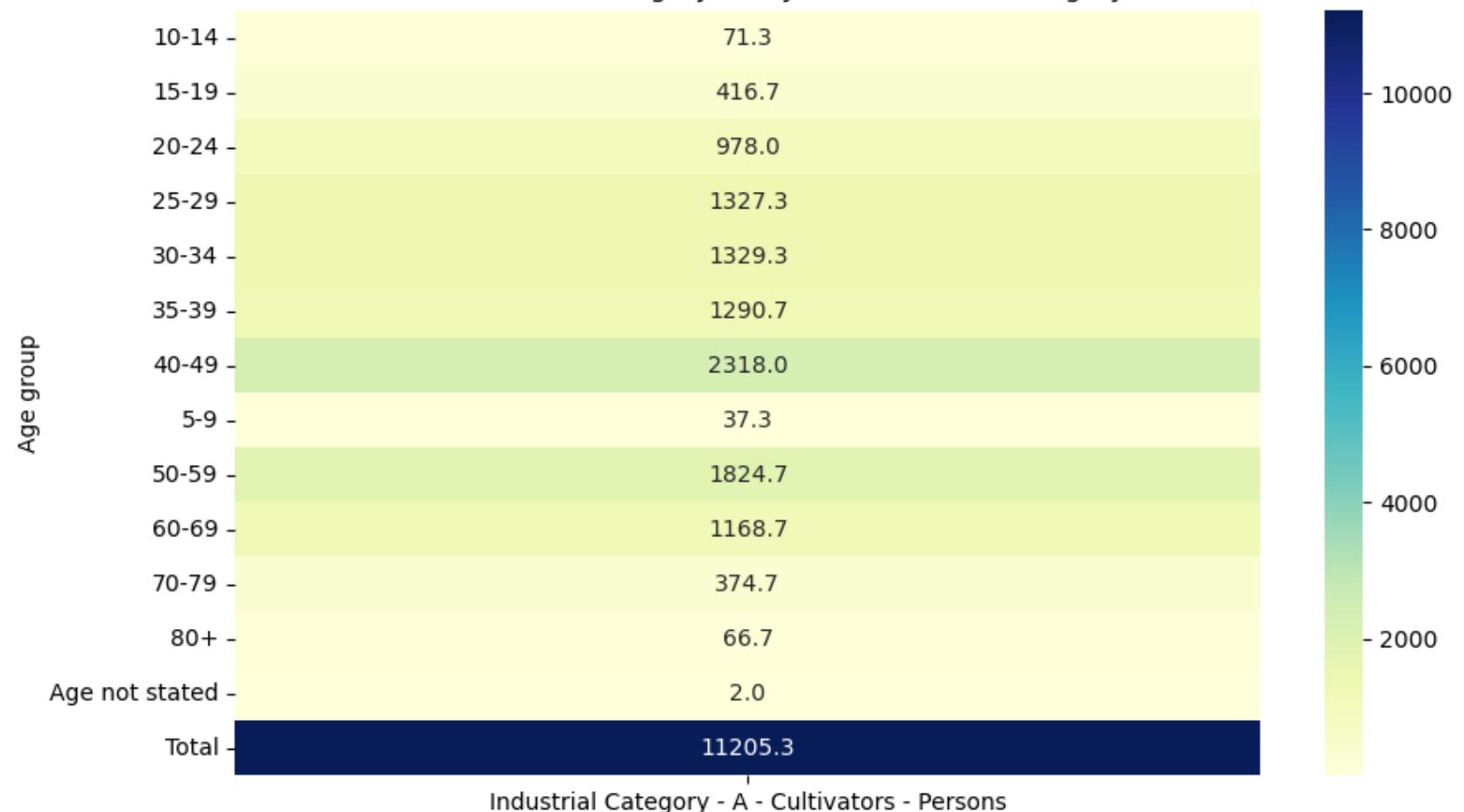


Industrial Category - A - Cultivators - Males

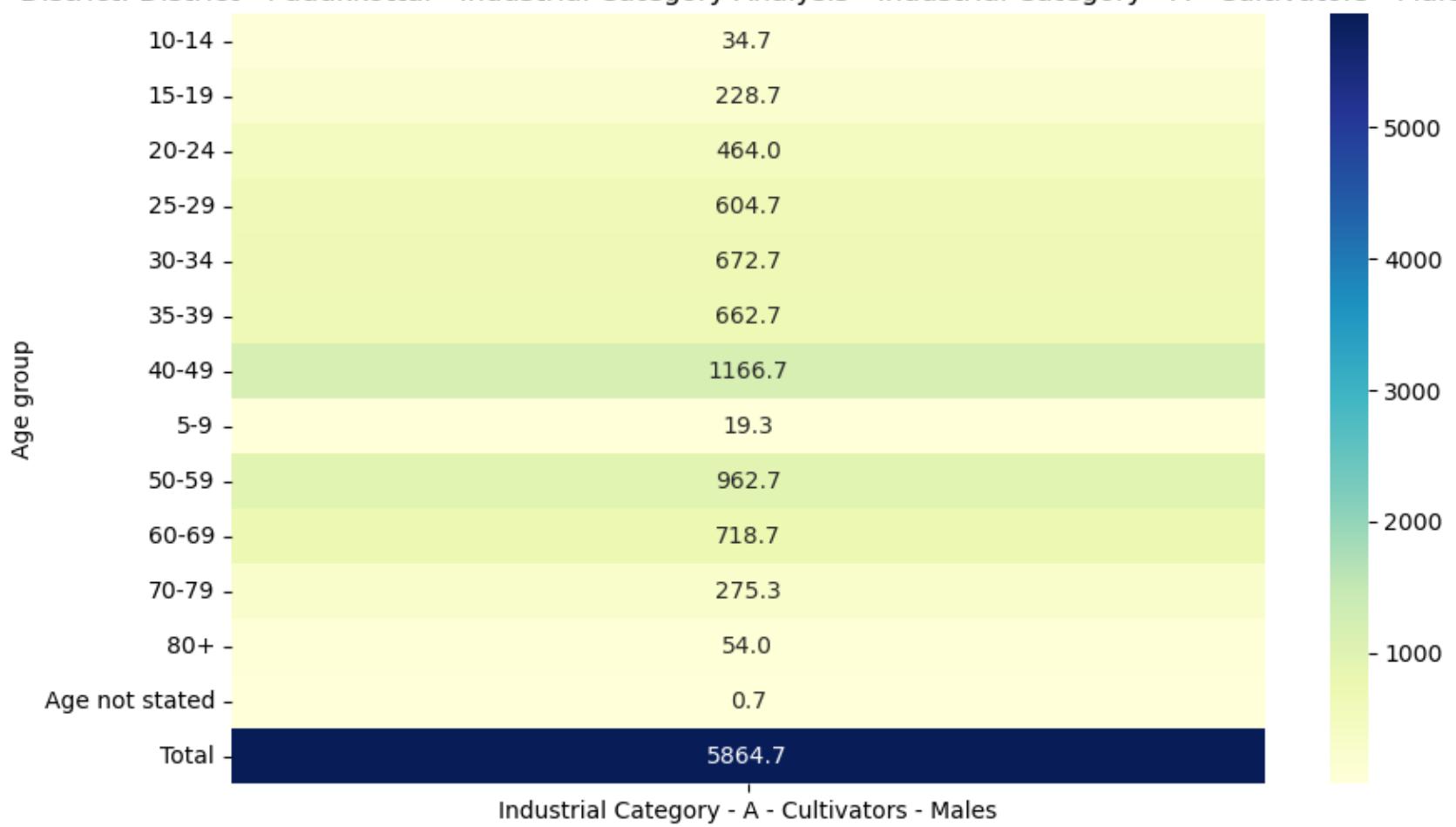
District: District - Thanjavur - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



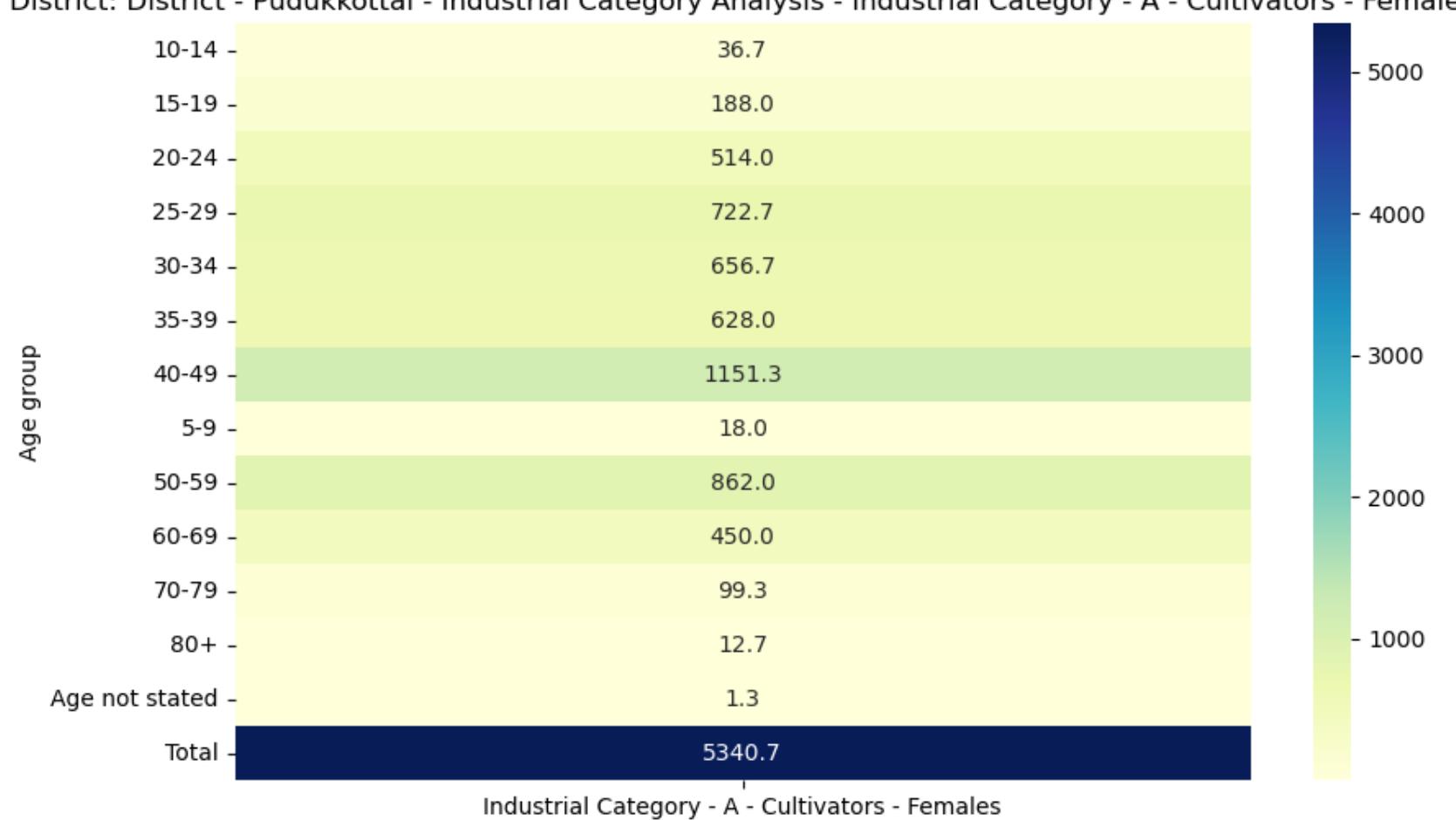
District: District - Pudukkottai - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



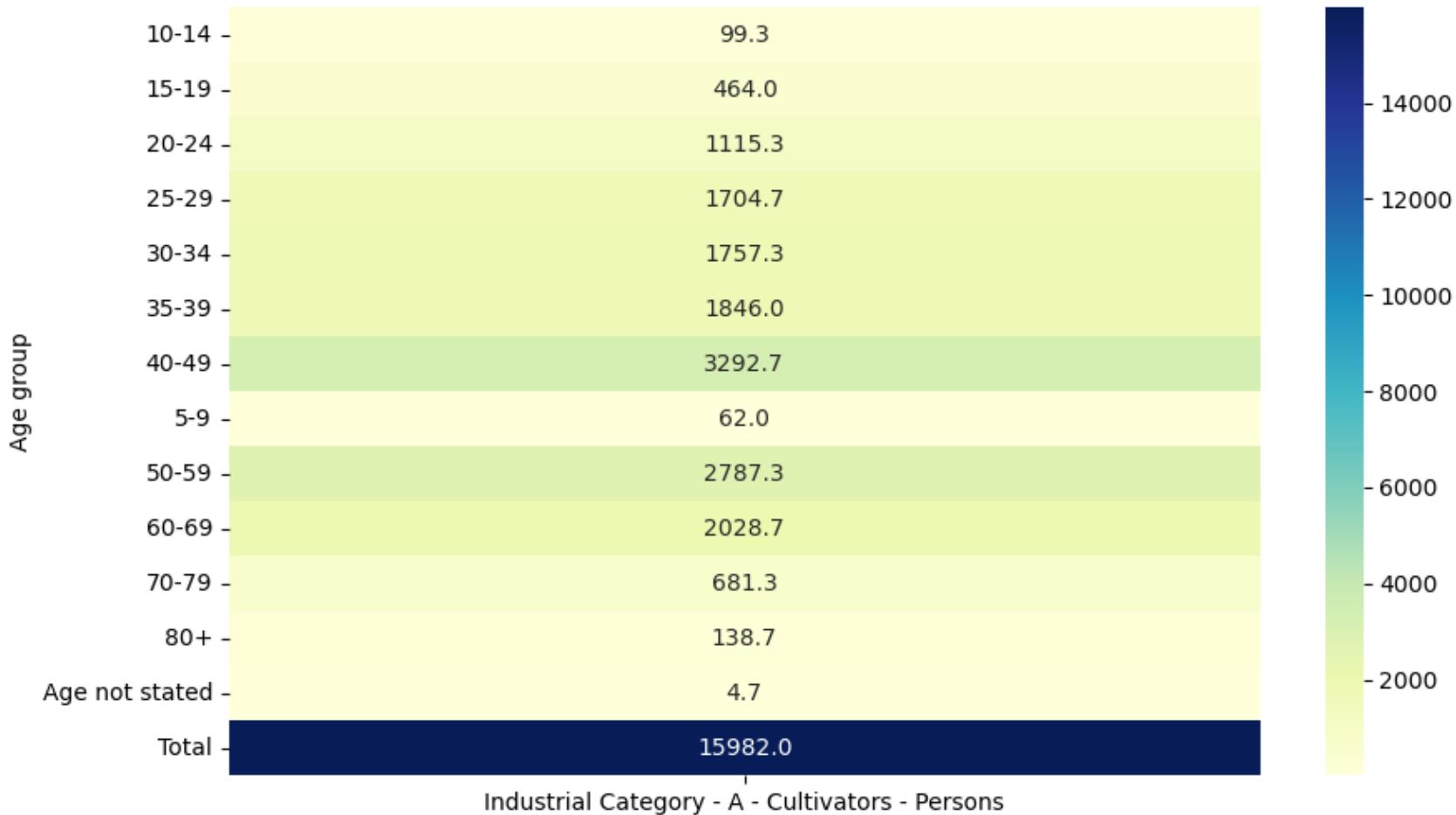
District: District - Pudukkottai - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



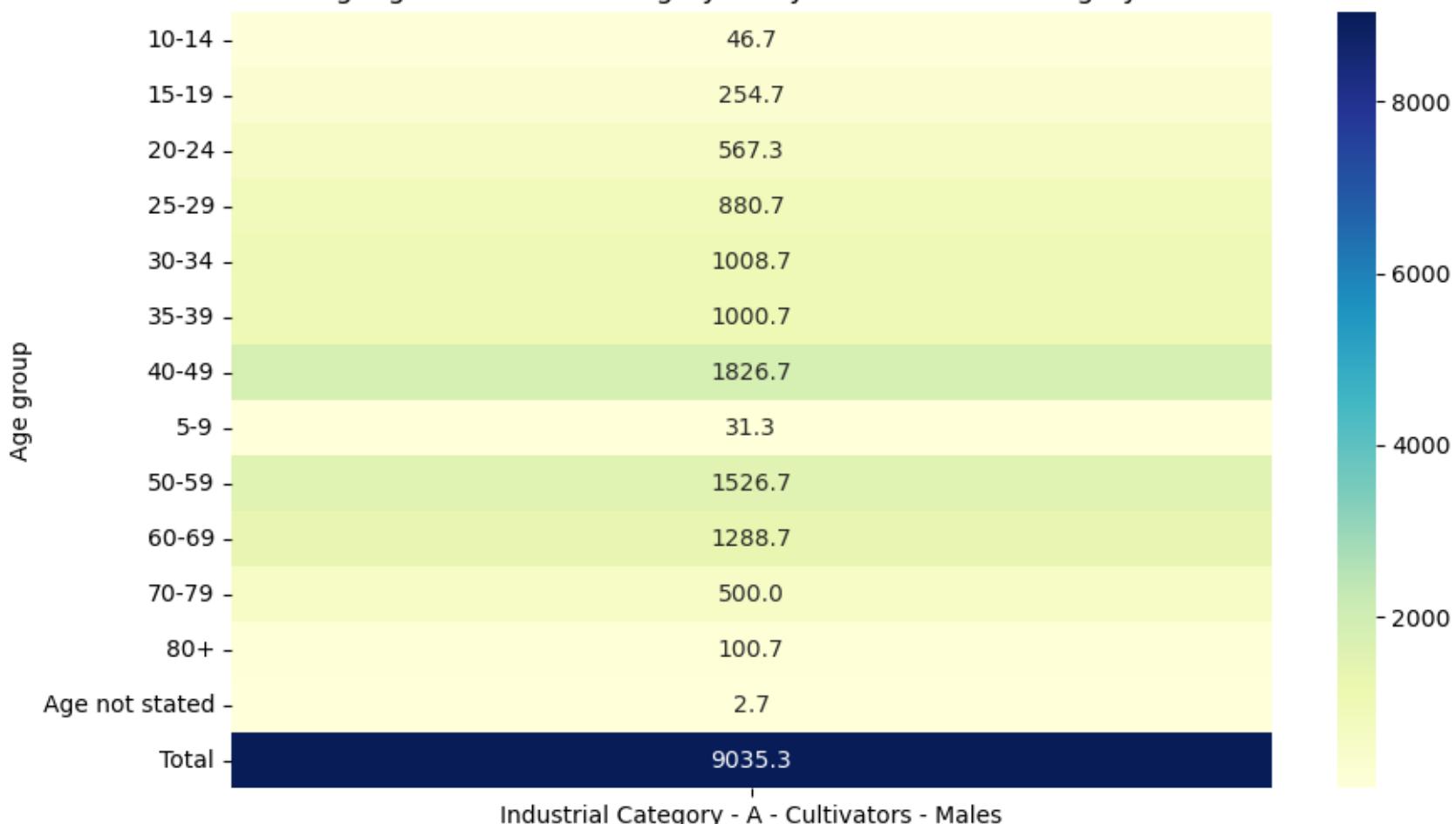
District: District - Pudukkottai - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



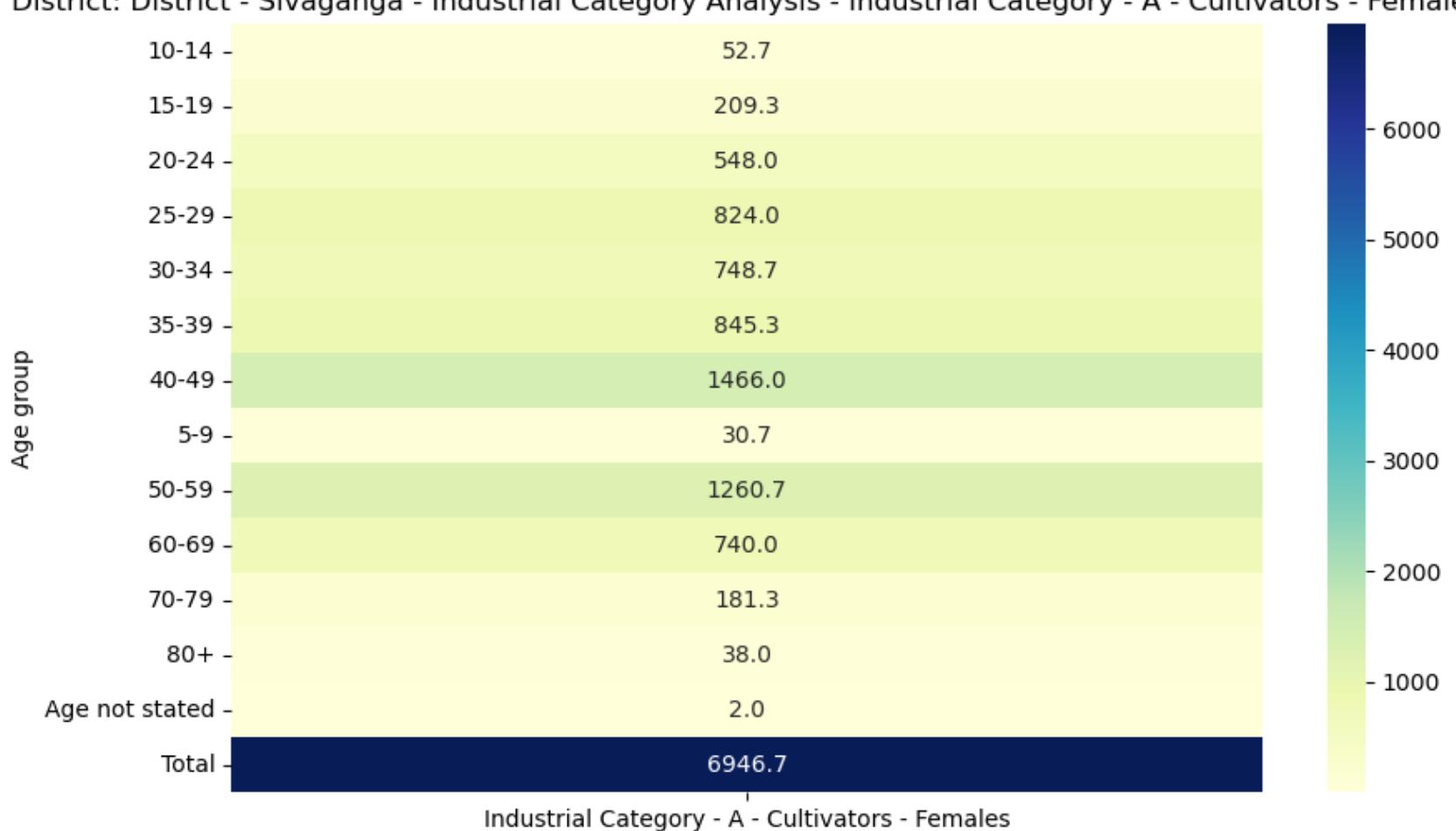
District: District - Sivaganga - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



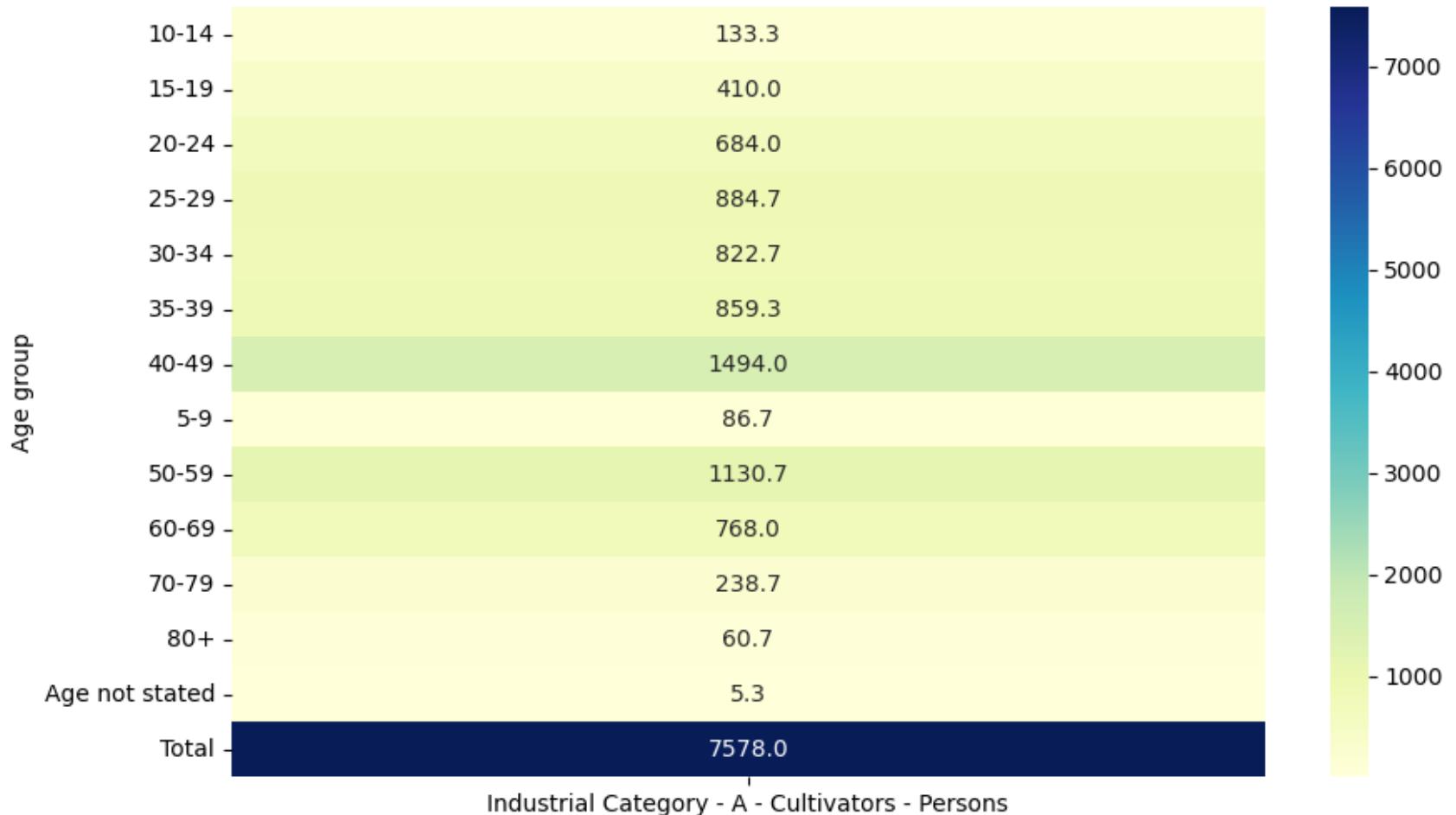
District: District - Sivaganga - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



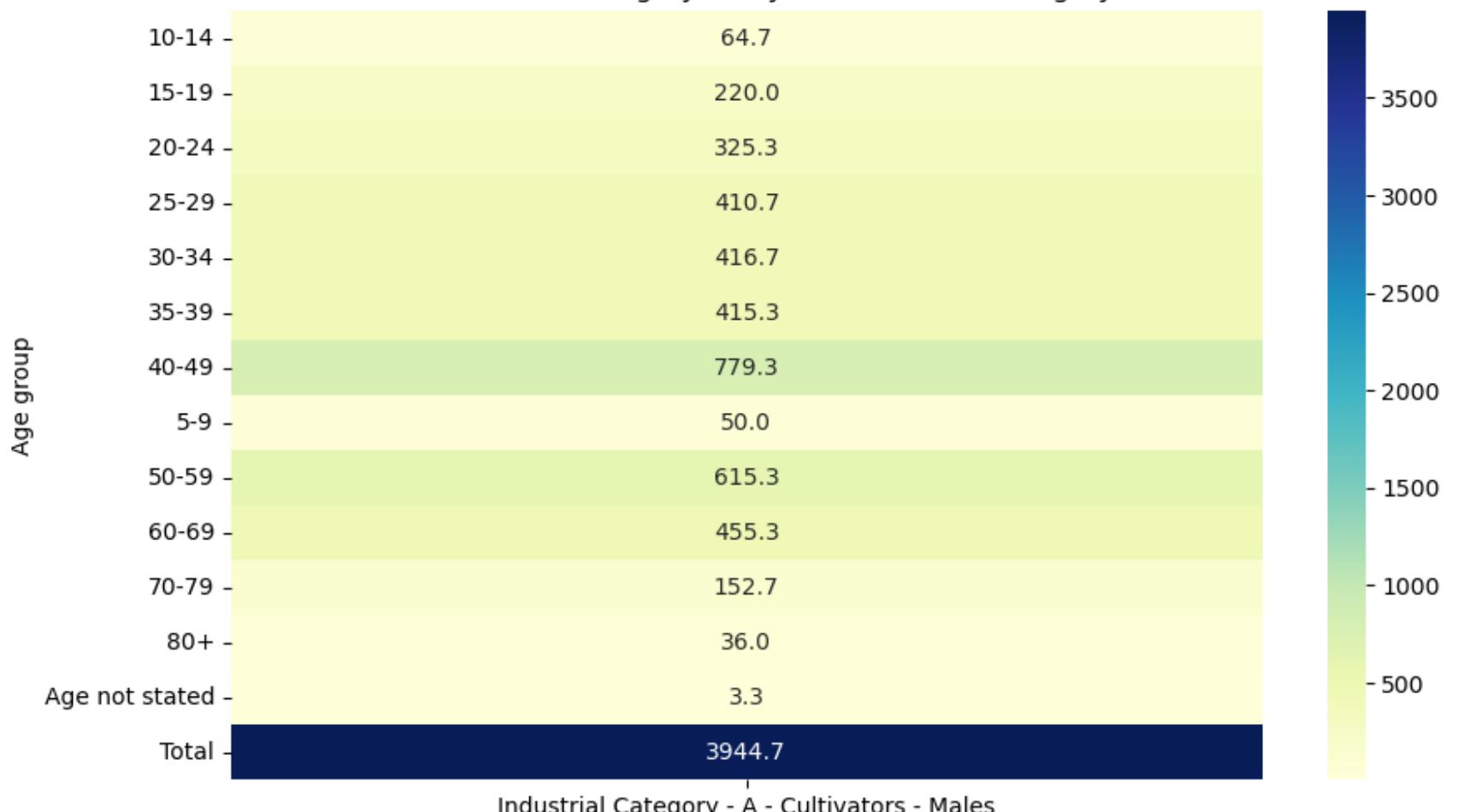
District: District - Sivaganga - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



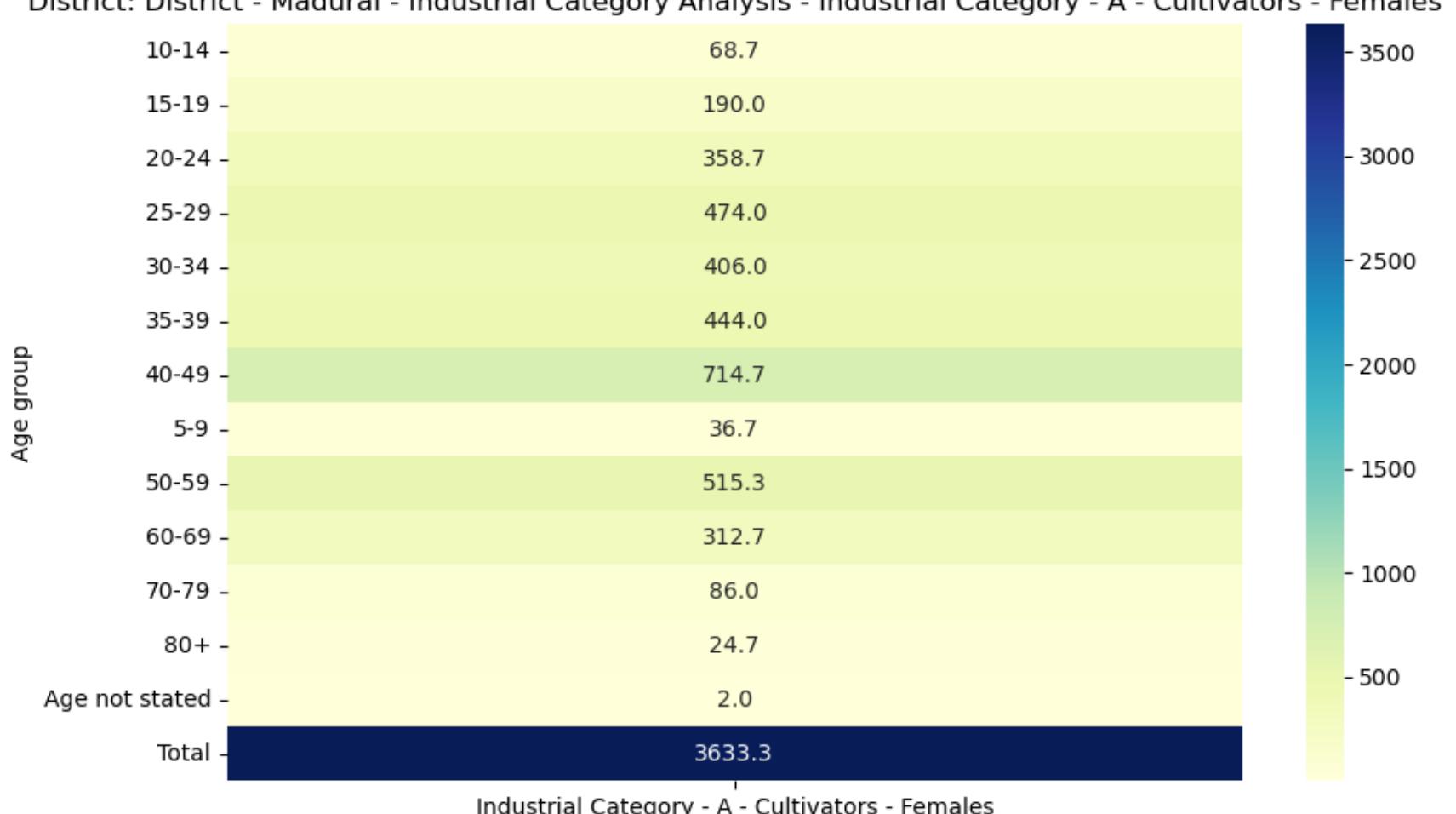
District: District - Madurai - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



District: District - Madurai - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

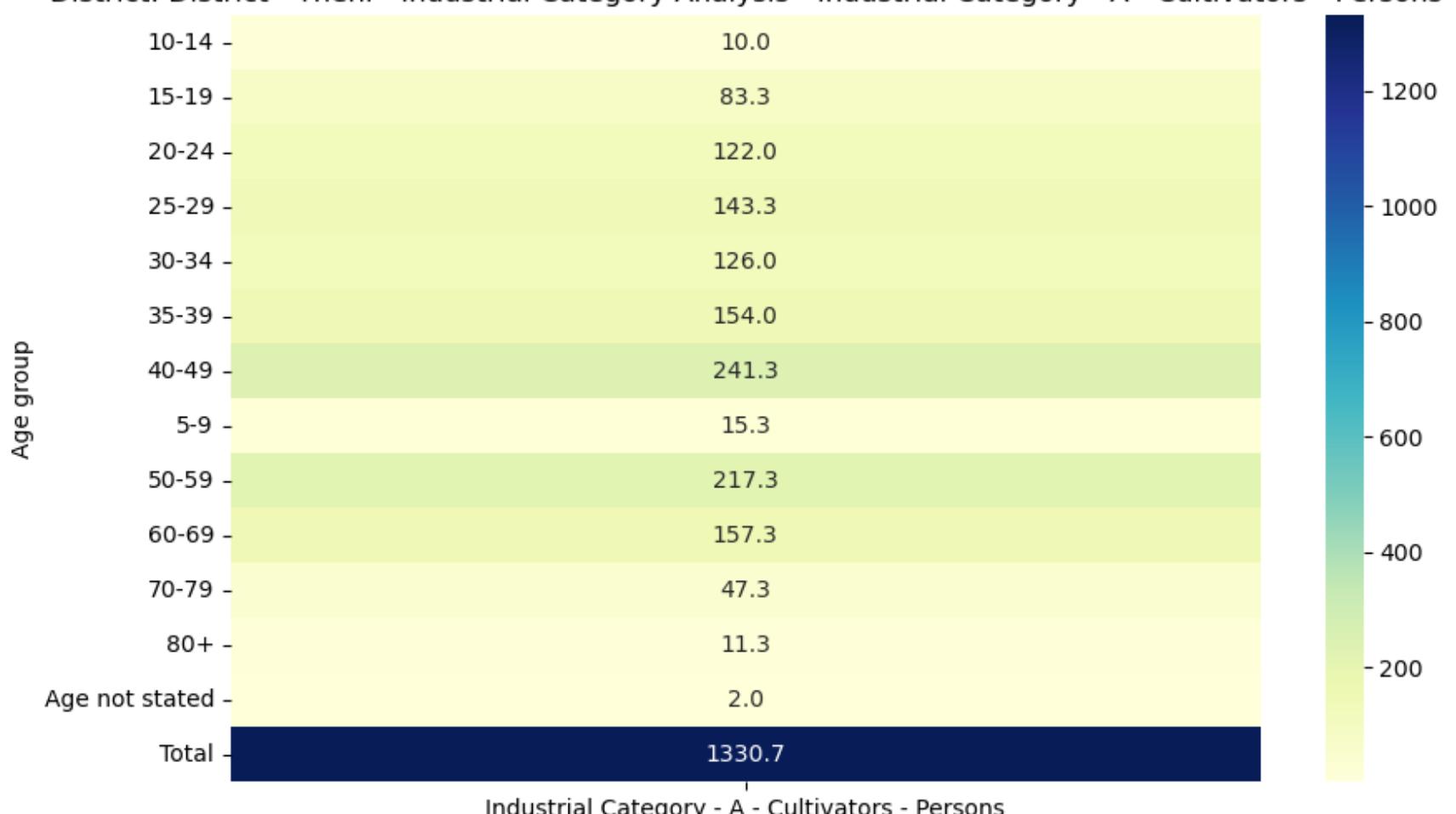


District: District - Madurai - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



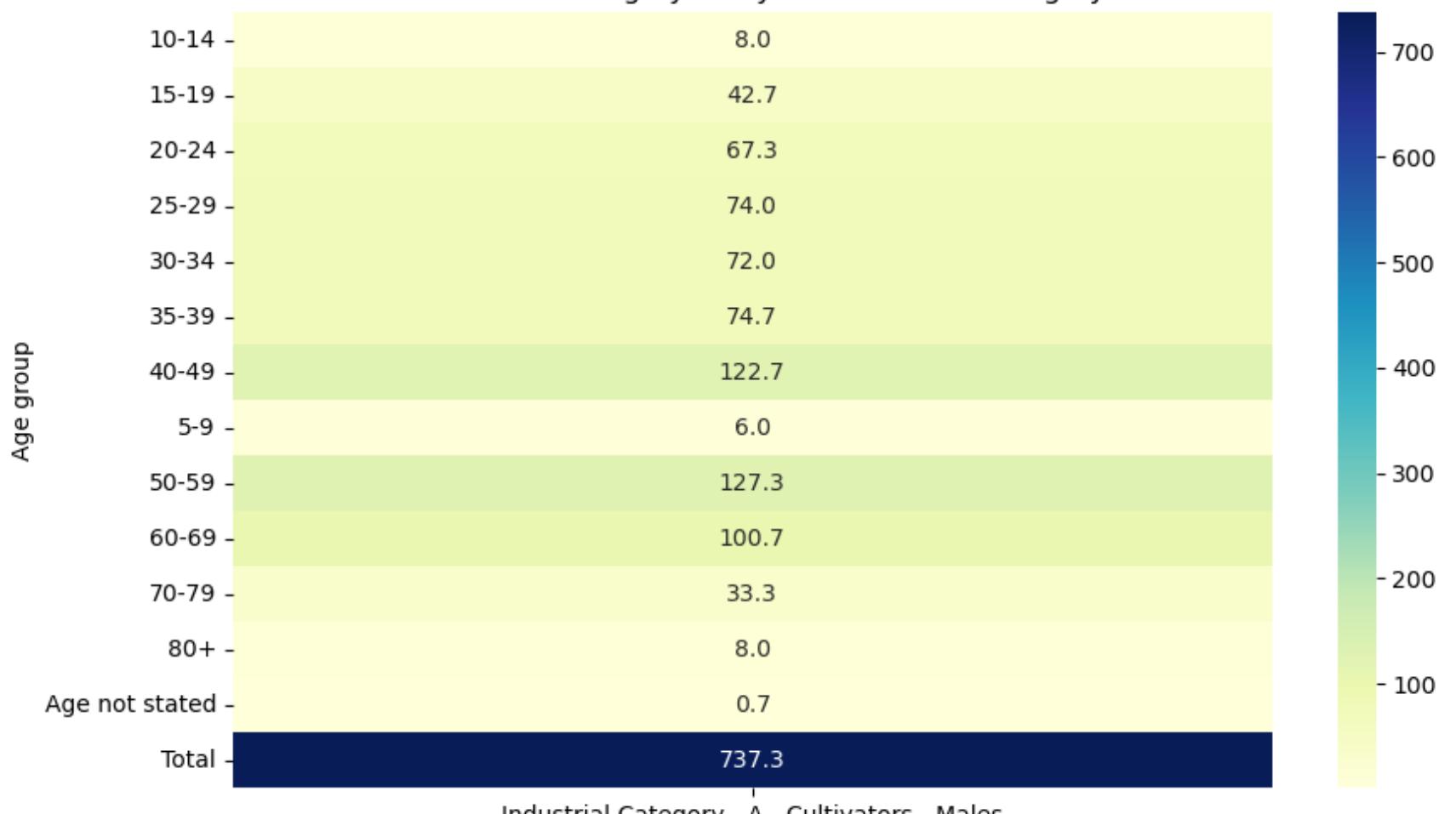
Industrial Category - A - Cultivators - Females

District: District - Theni - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



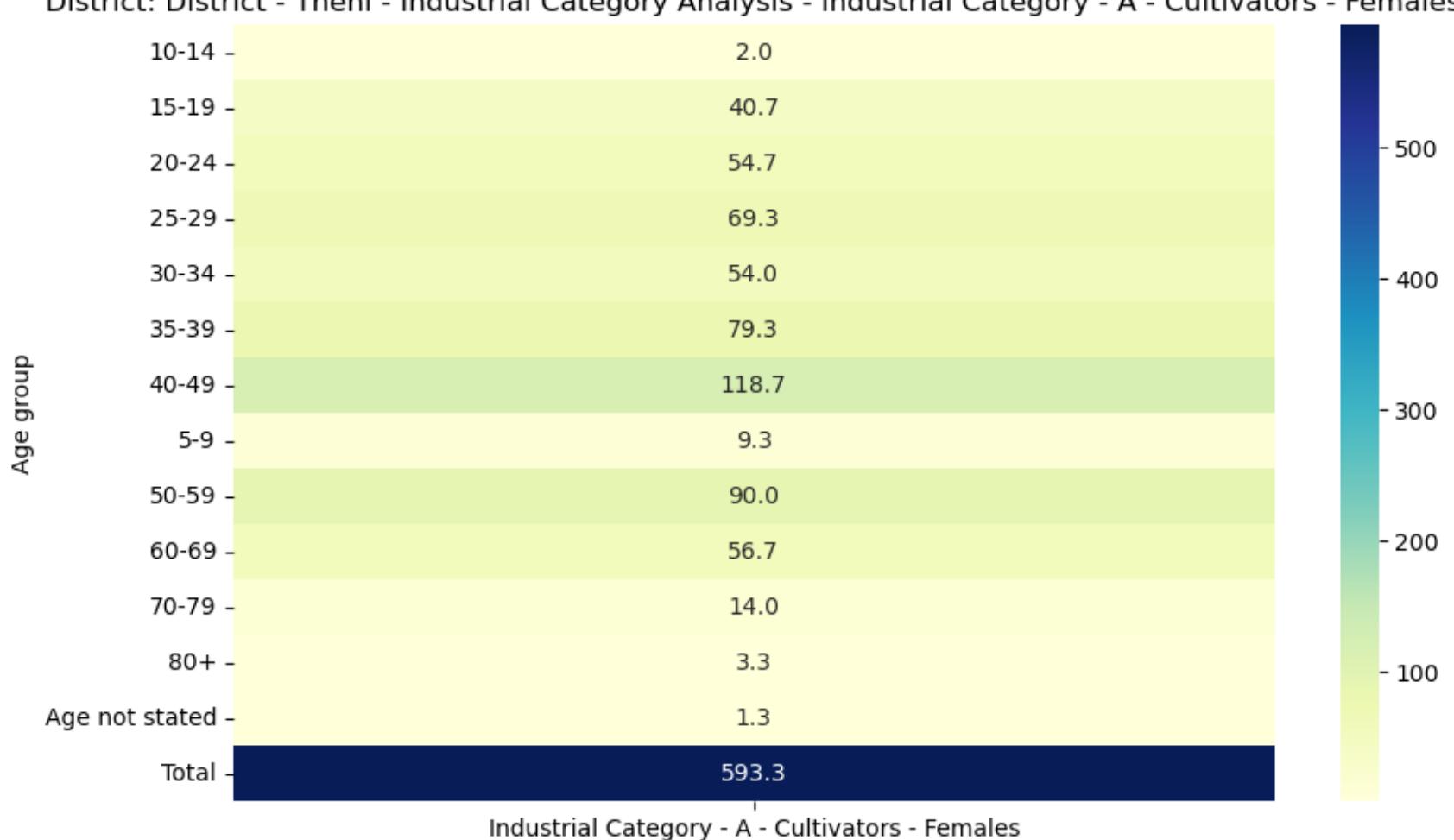
Industrial Category - A - Cultivators - Persons

District: District - Theni - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

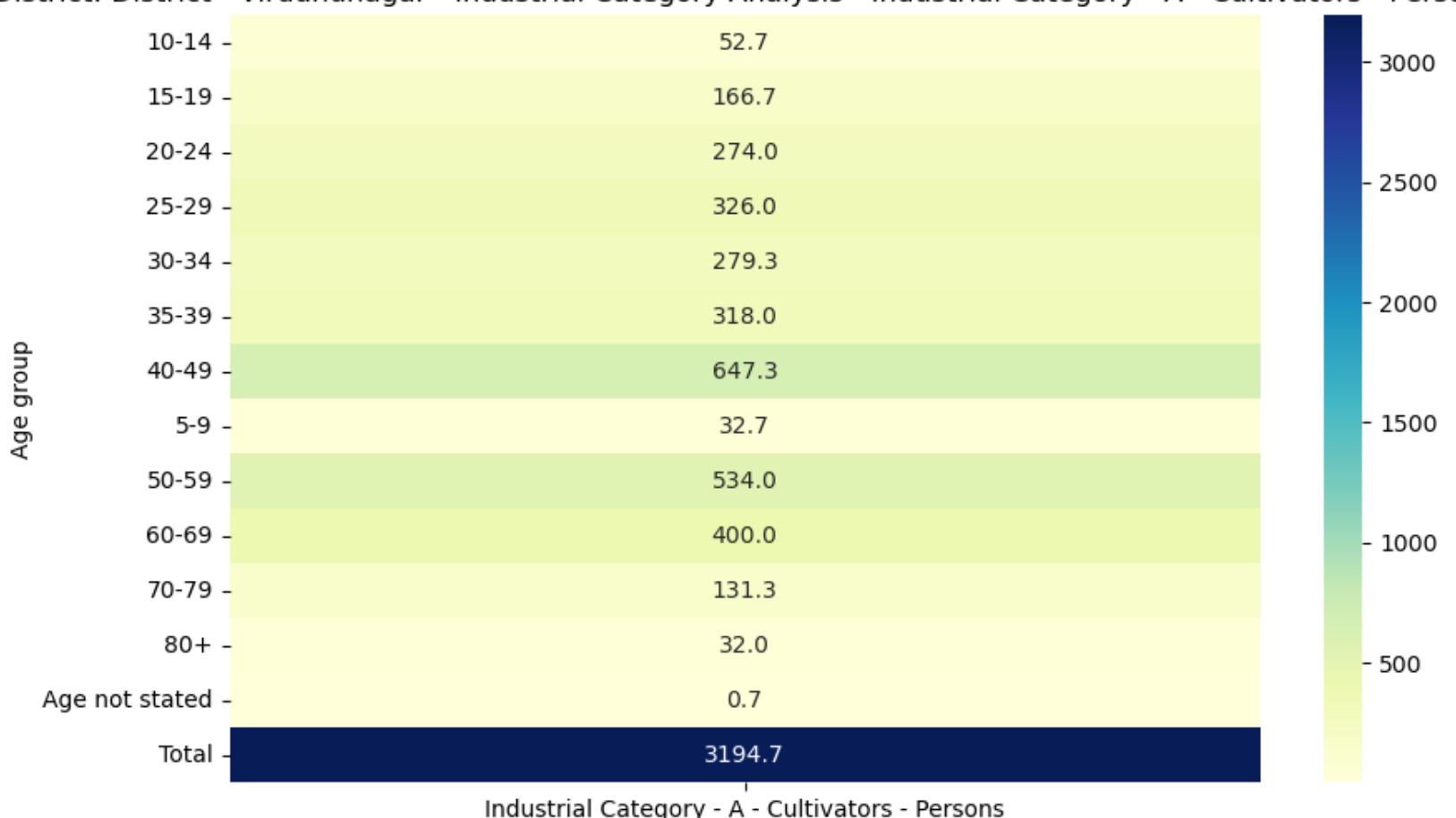


Industrial Category - A - Cultivators - Males

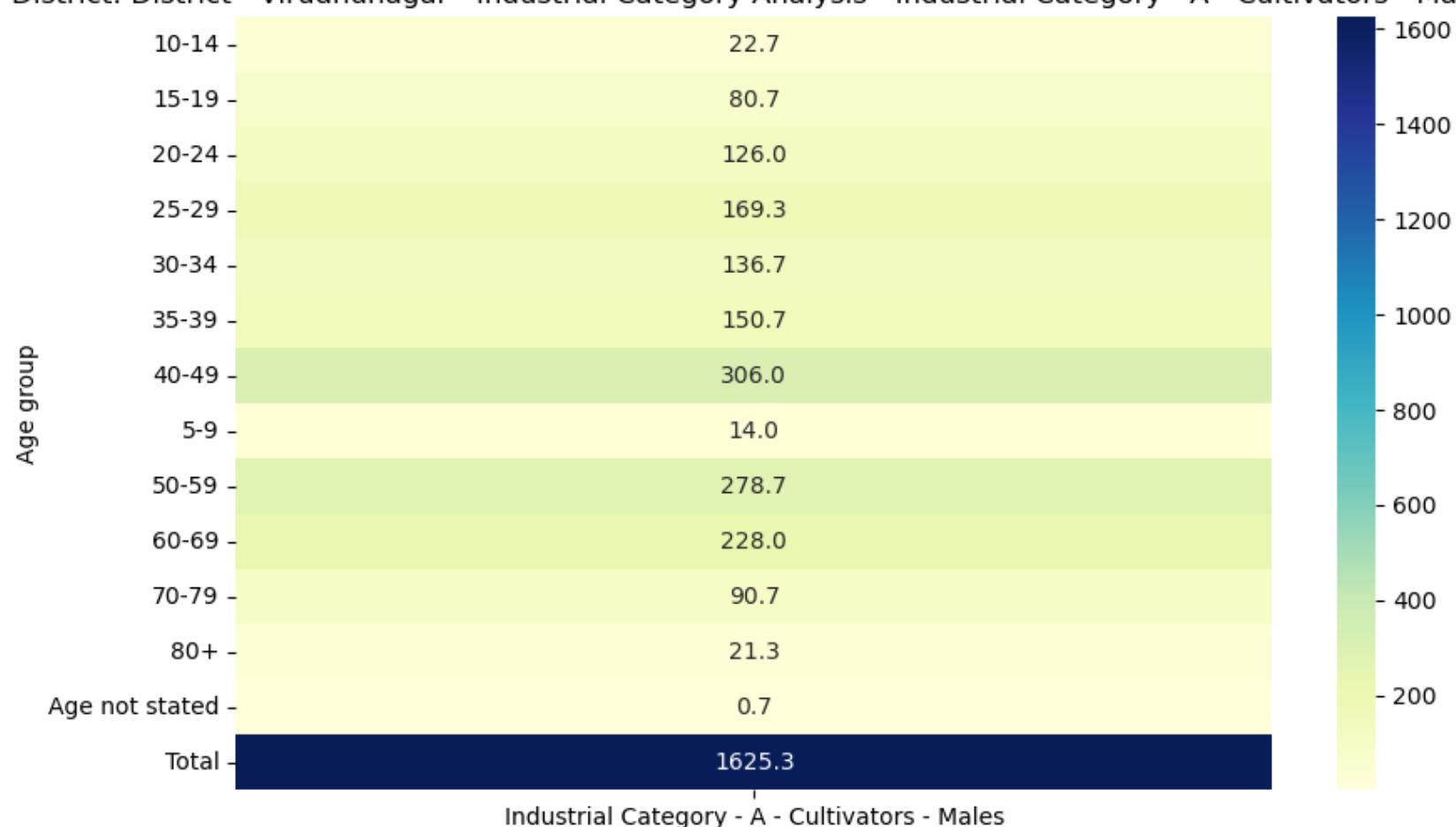
District: District - Theni - Industrial Category Analysis - Industrial Category - A - Cultivators - Females

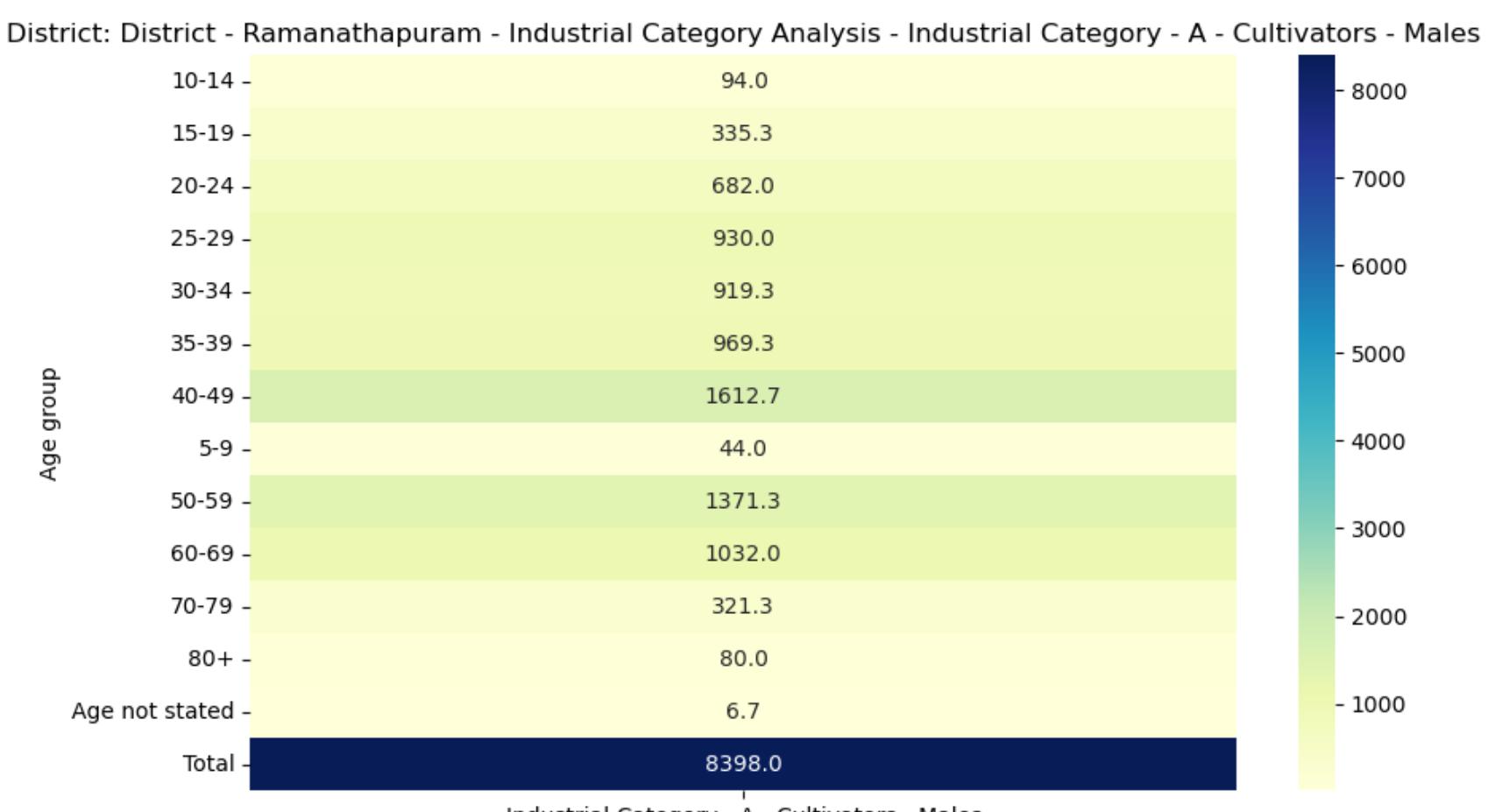
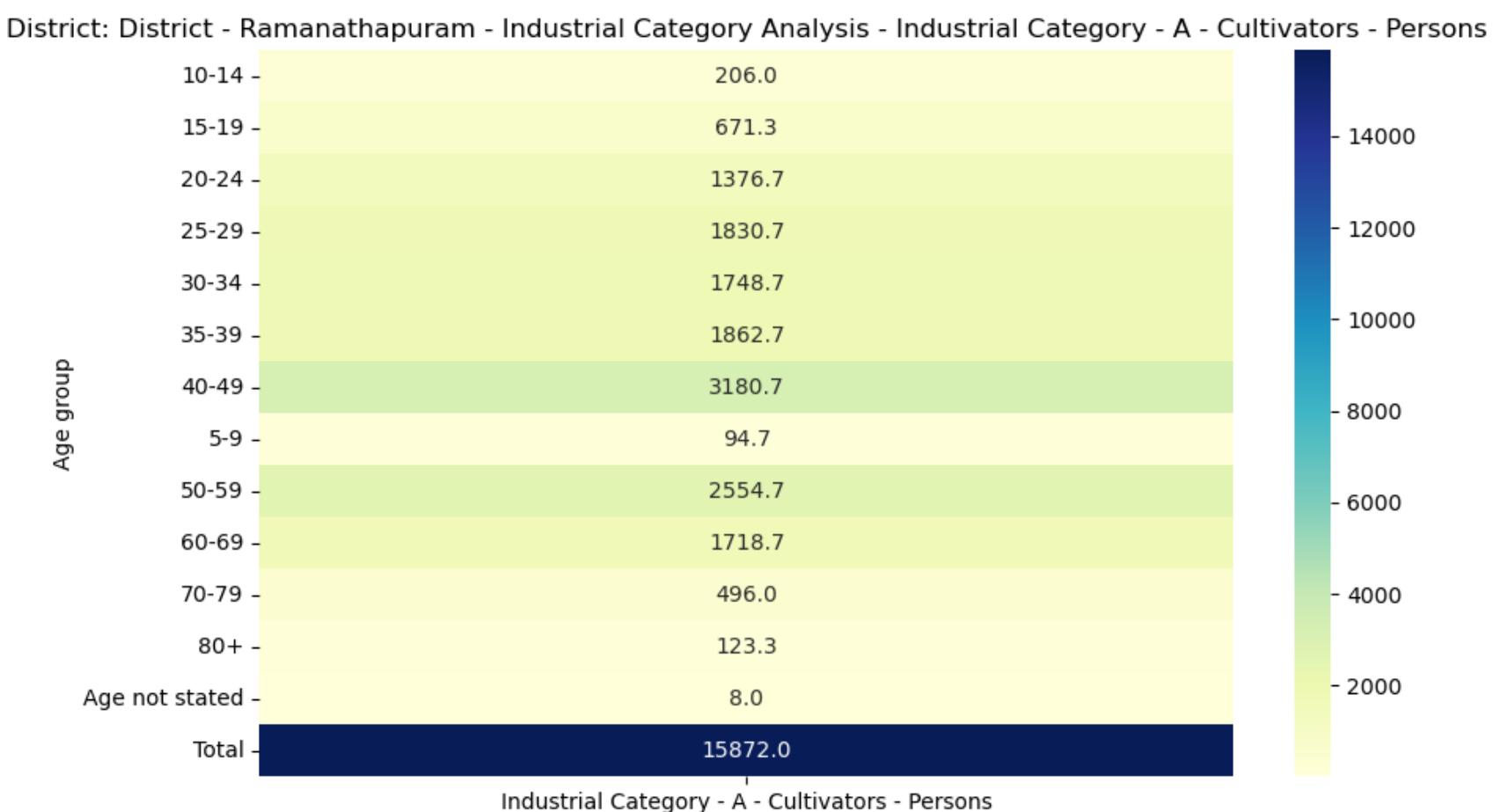
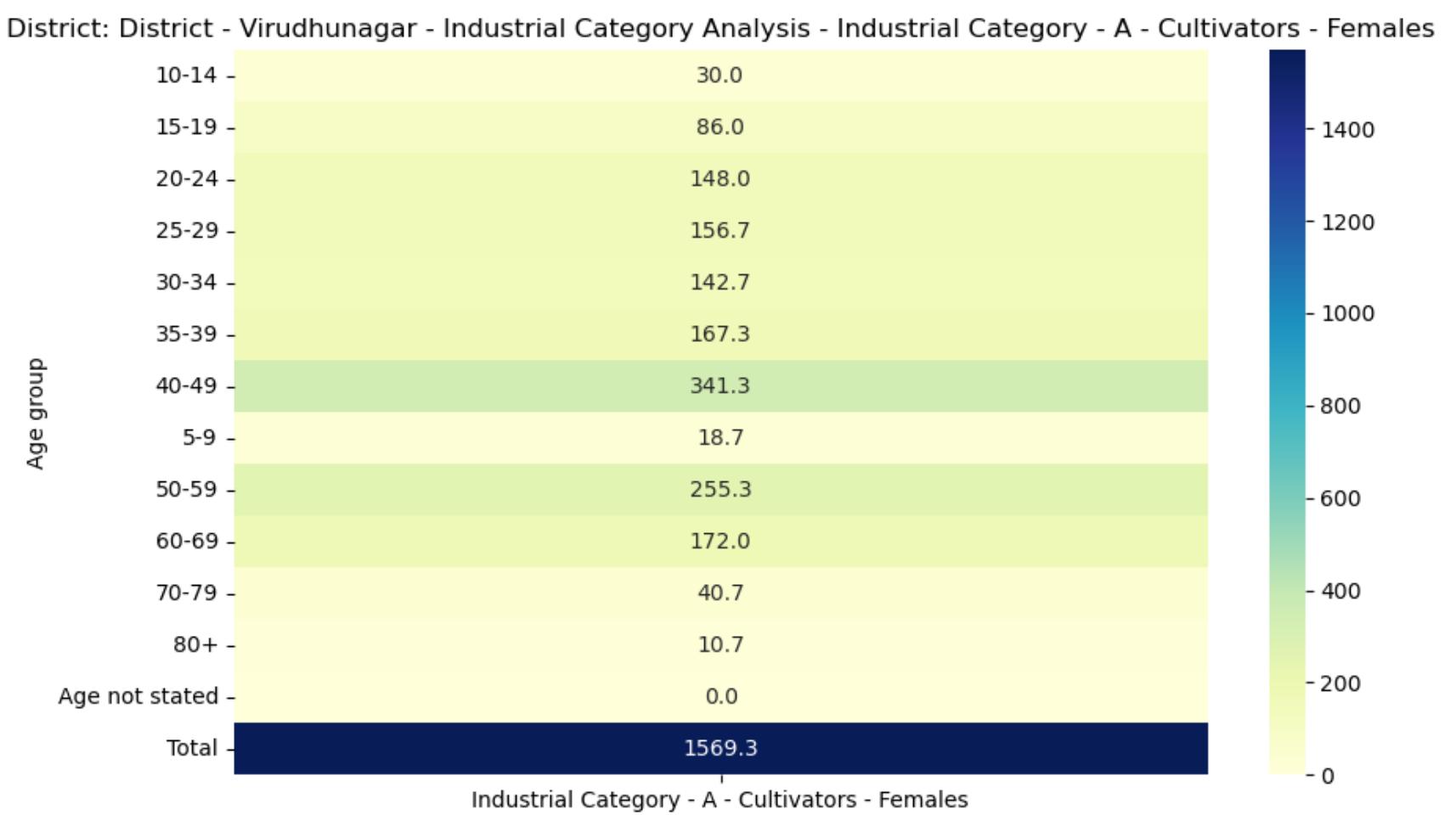


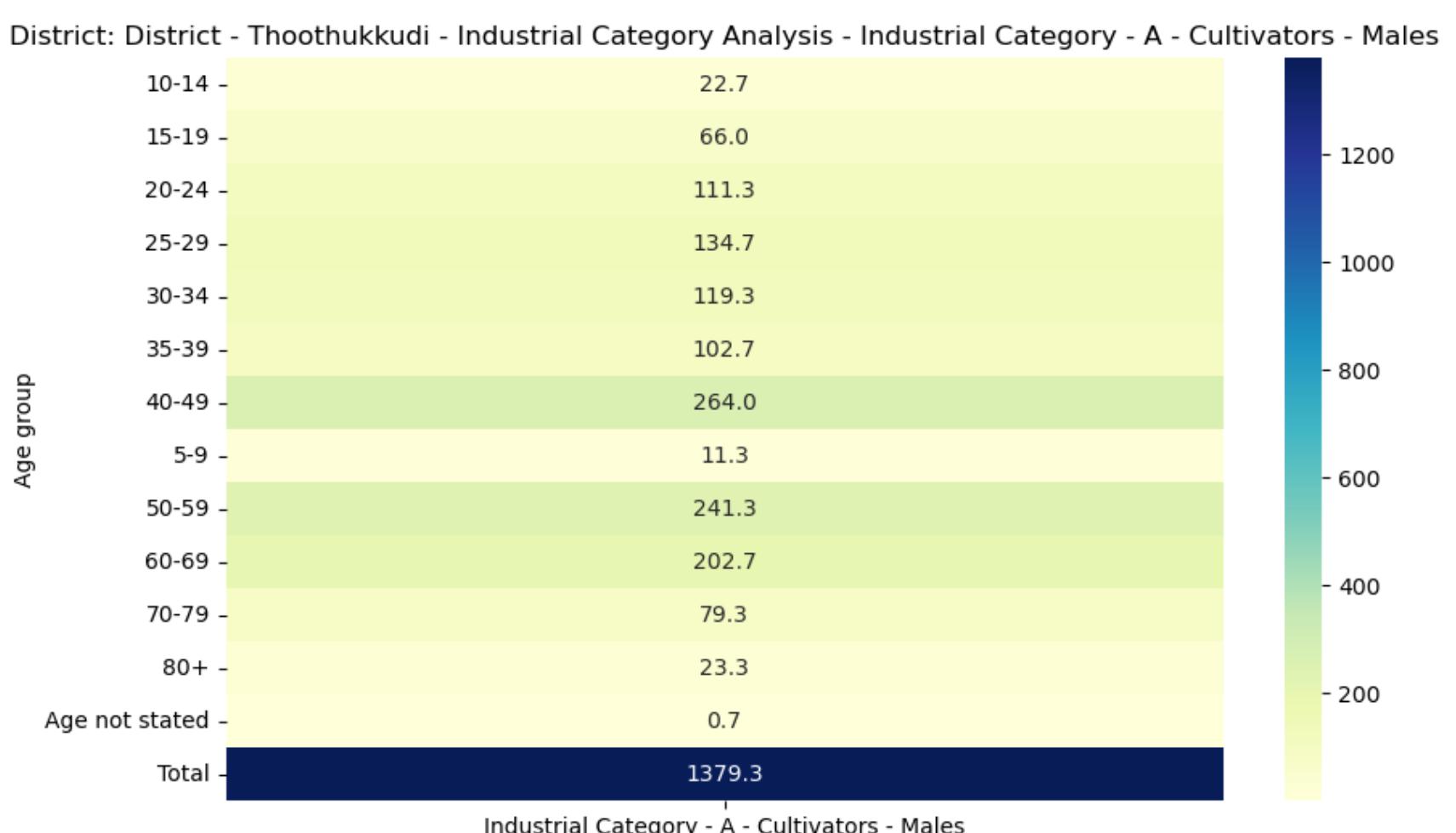
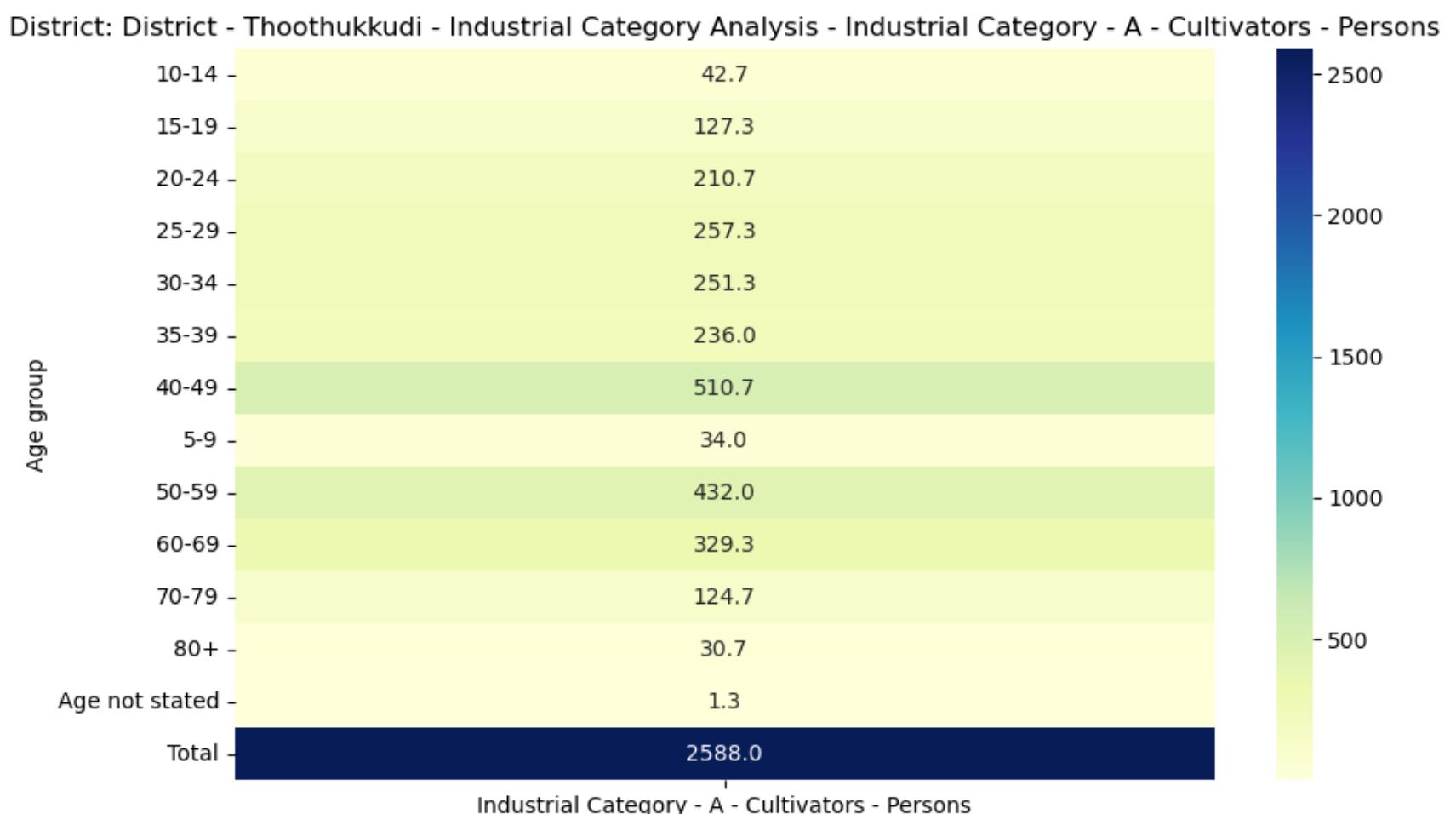
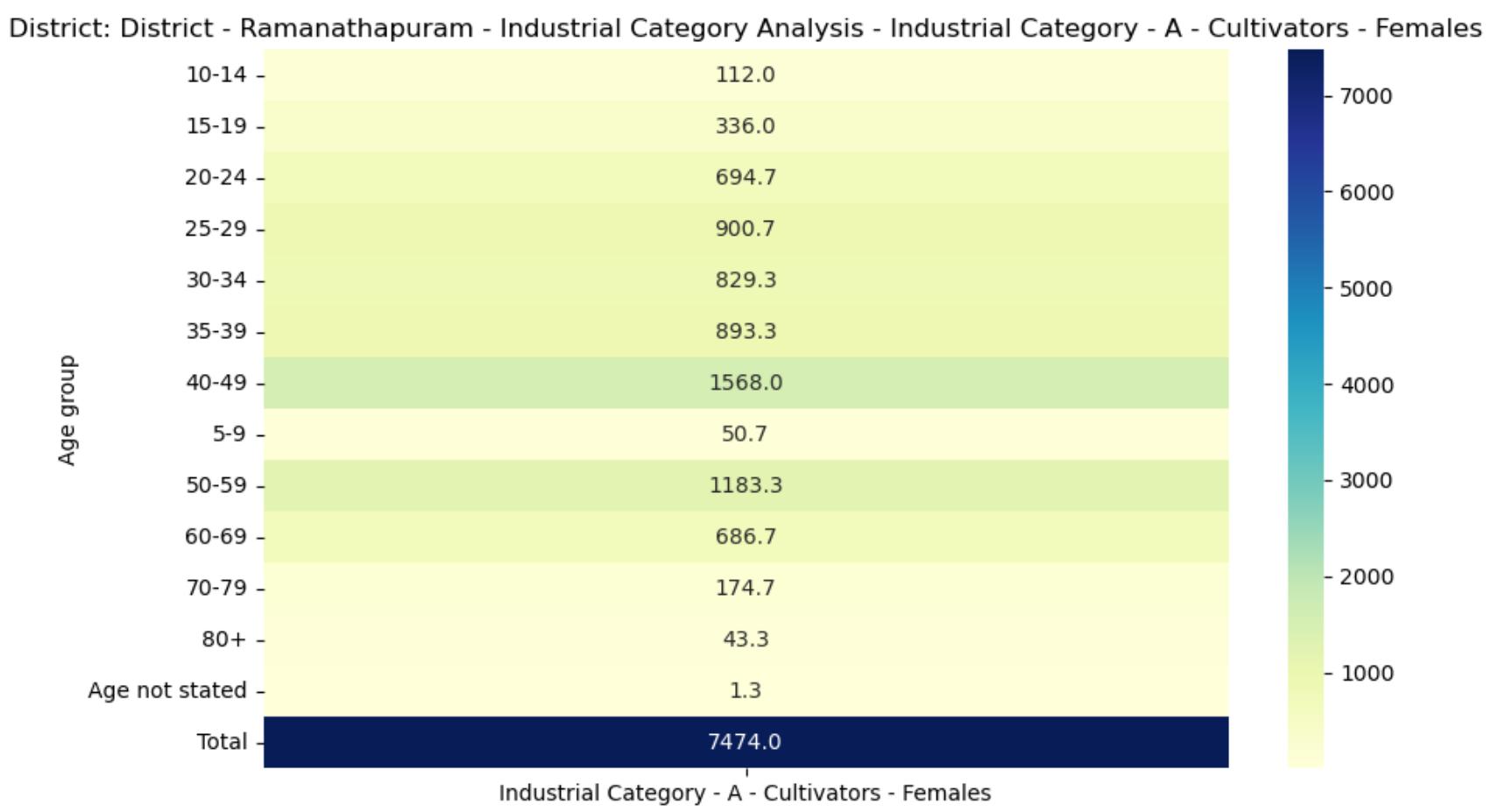
District: District - Virudhunagar - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons

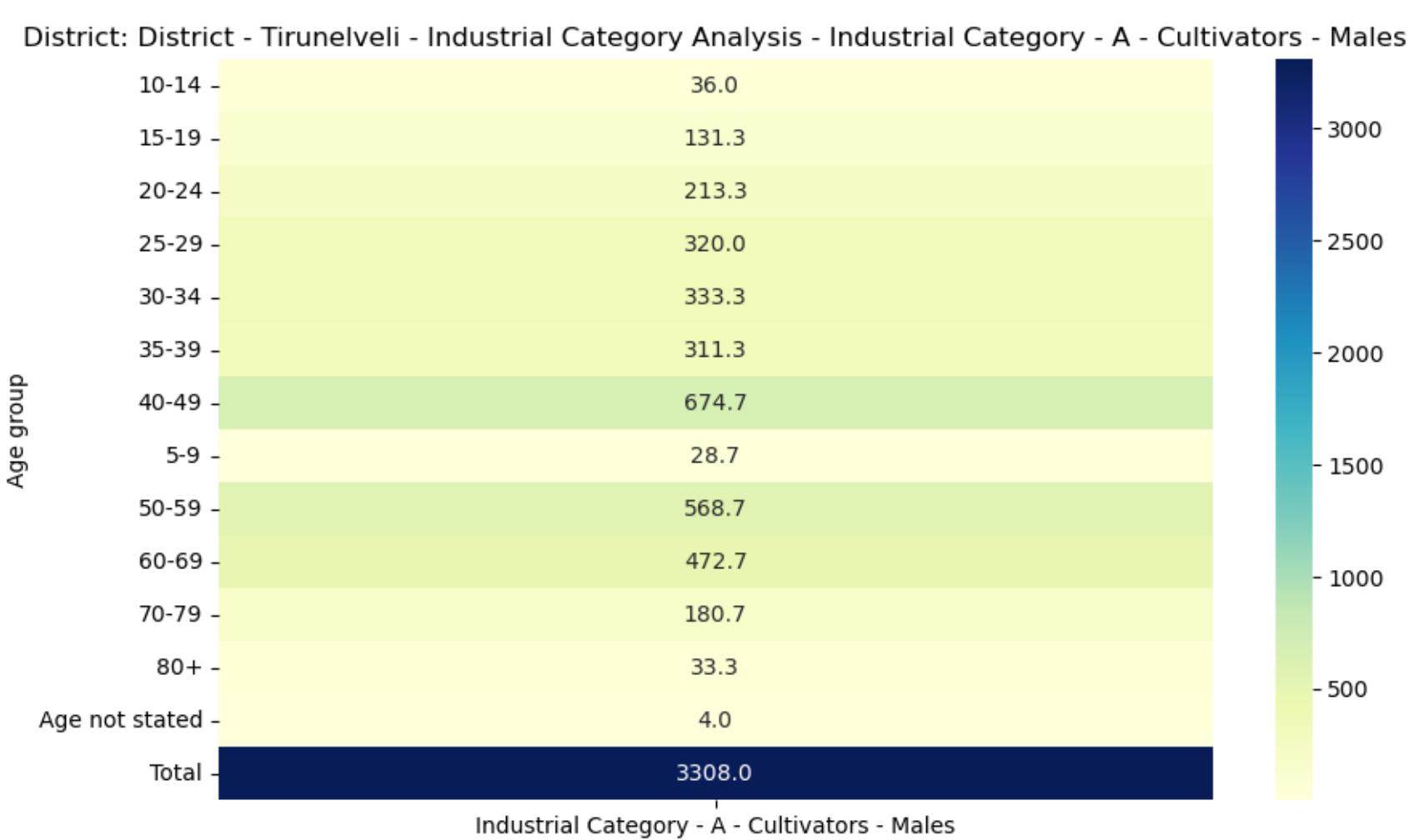
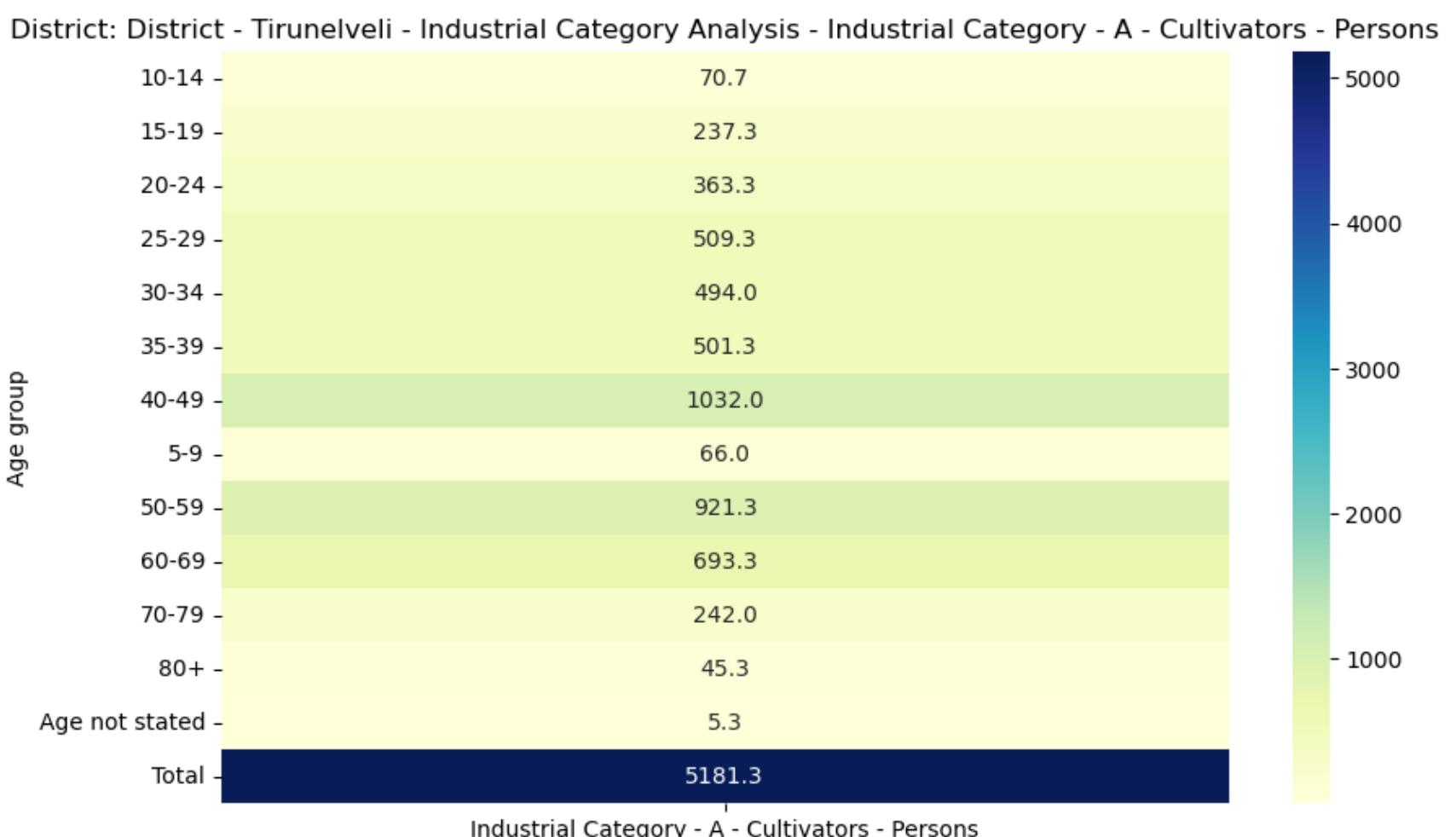
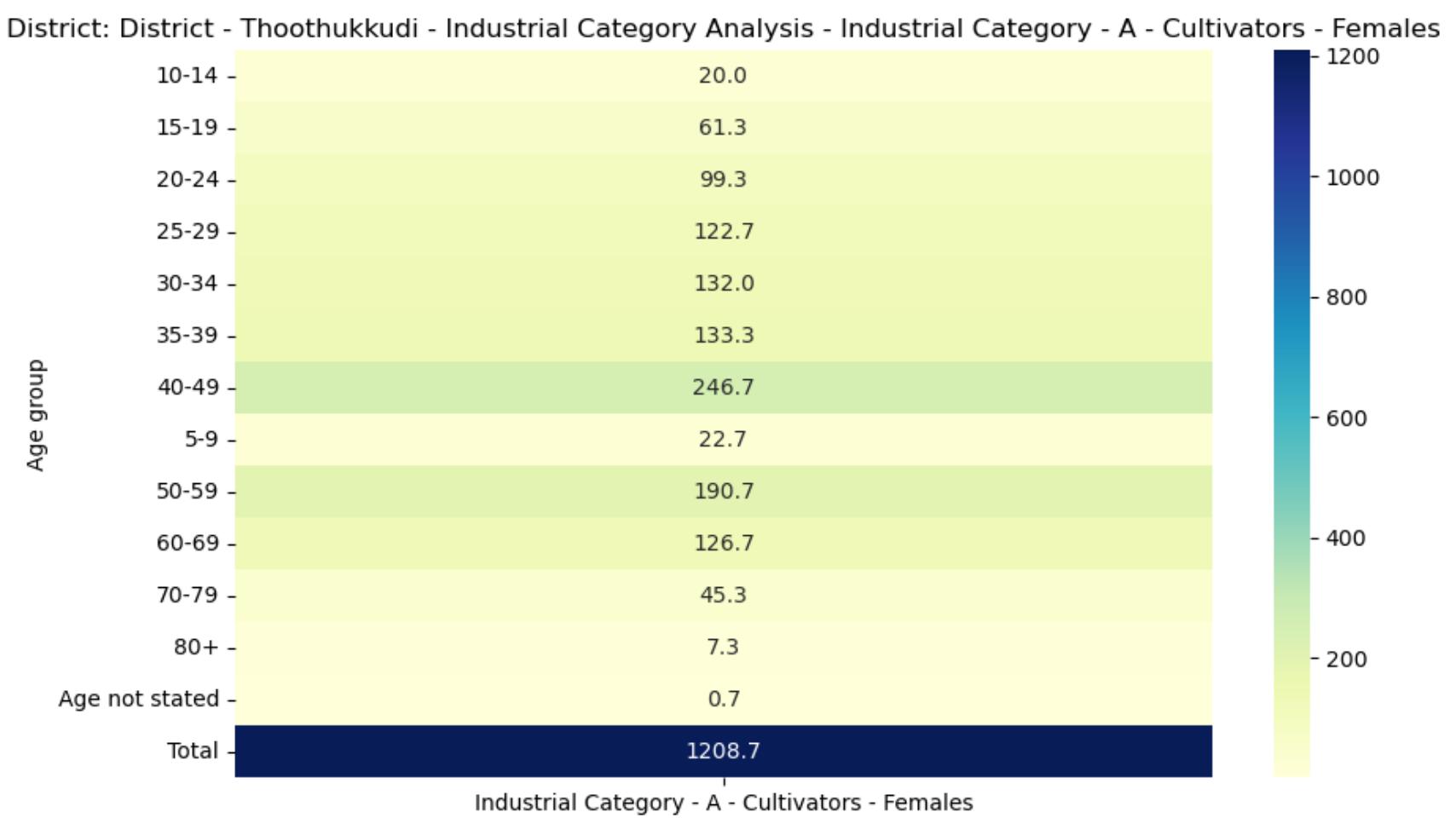


District: District - Virudhunagar - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

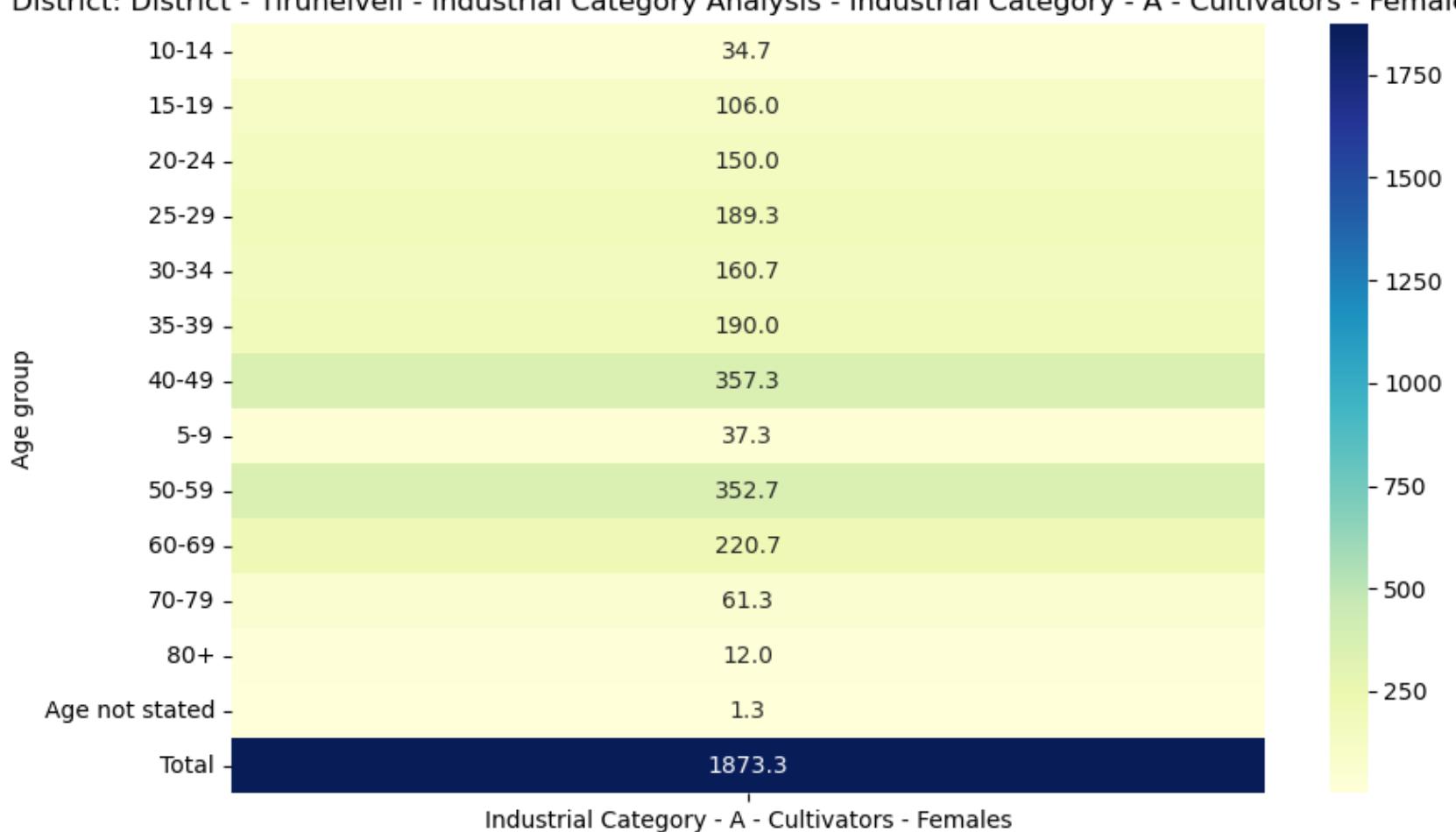






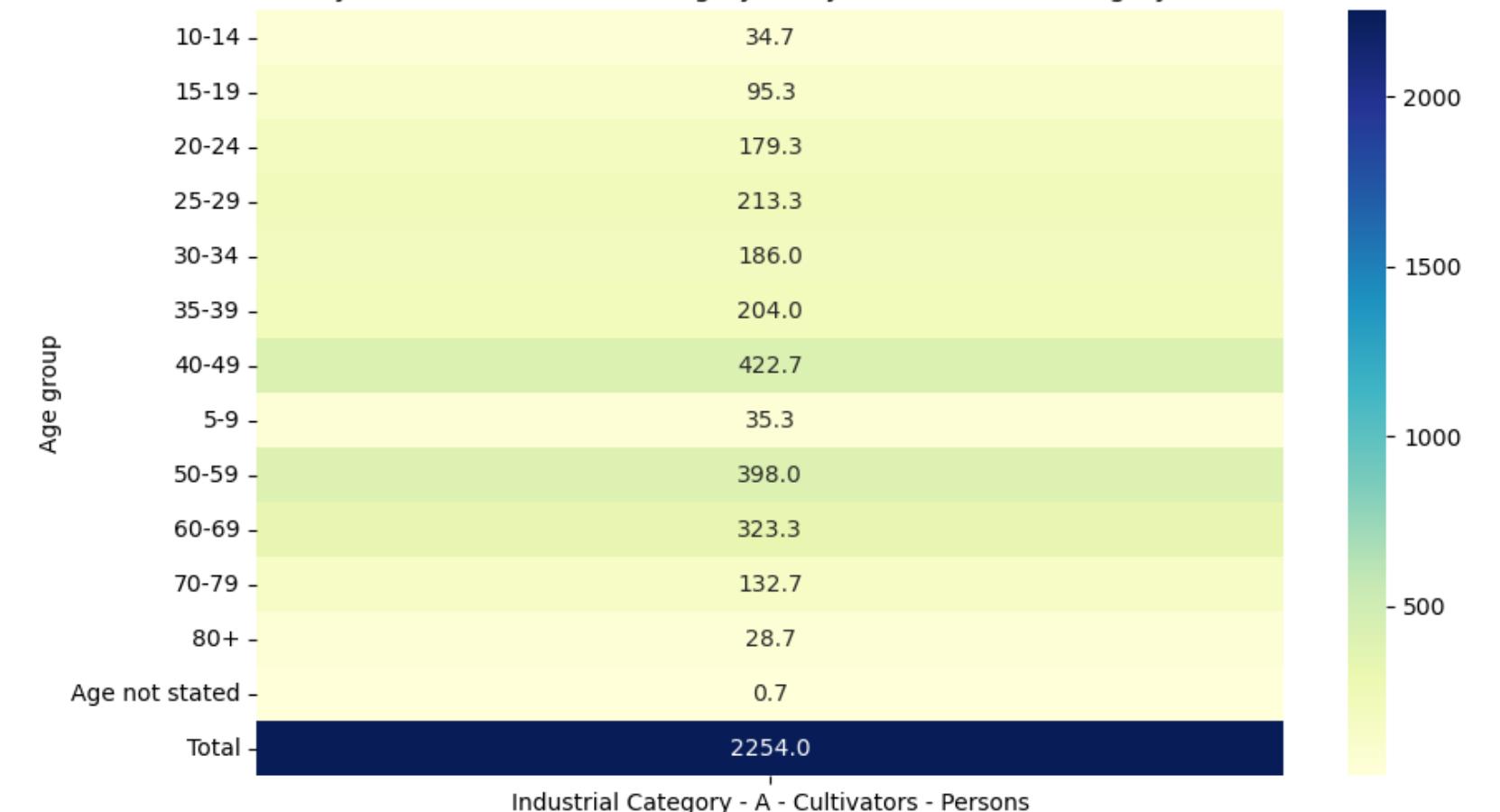


District: District - Tirunelveli - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



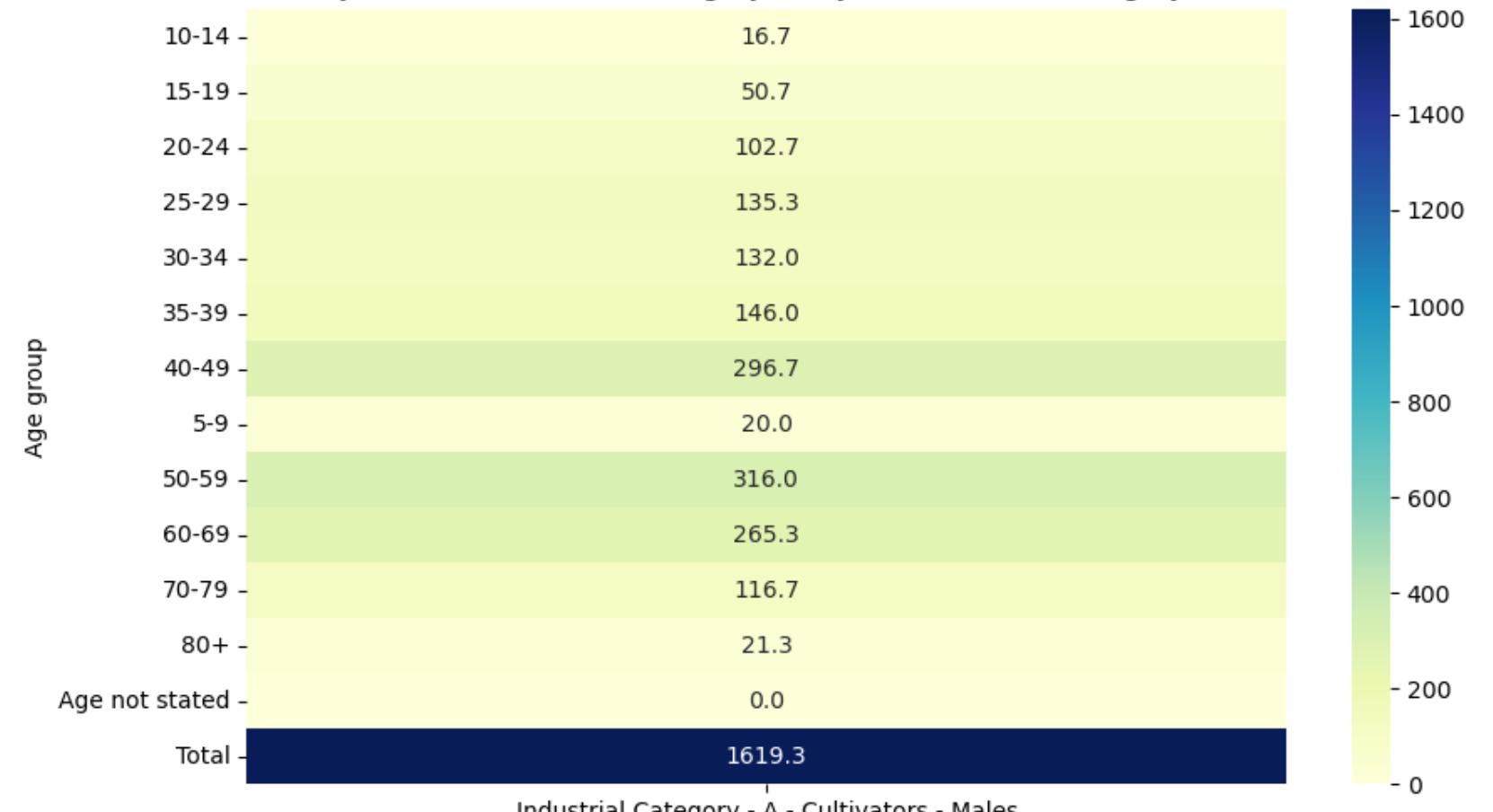
Industrial Category - A - Cultivators - Females

District: District - Kanyakumari - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons

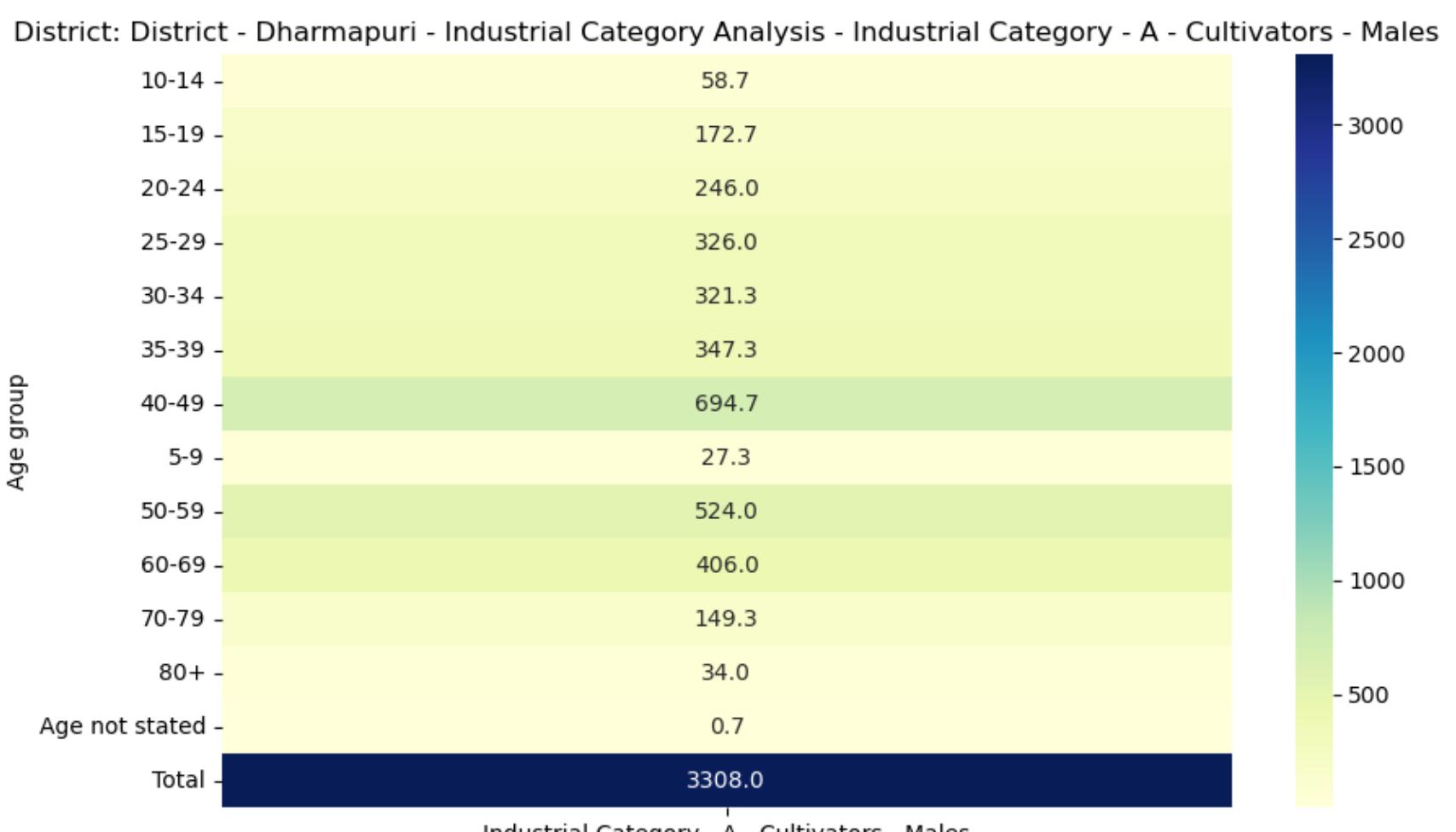
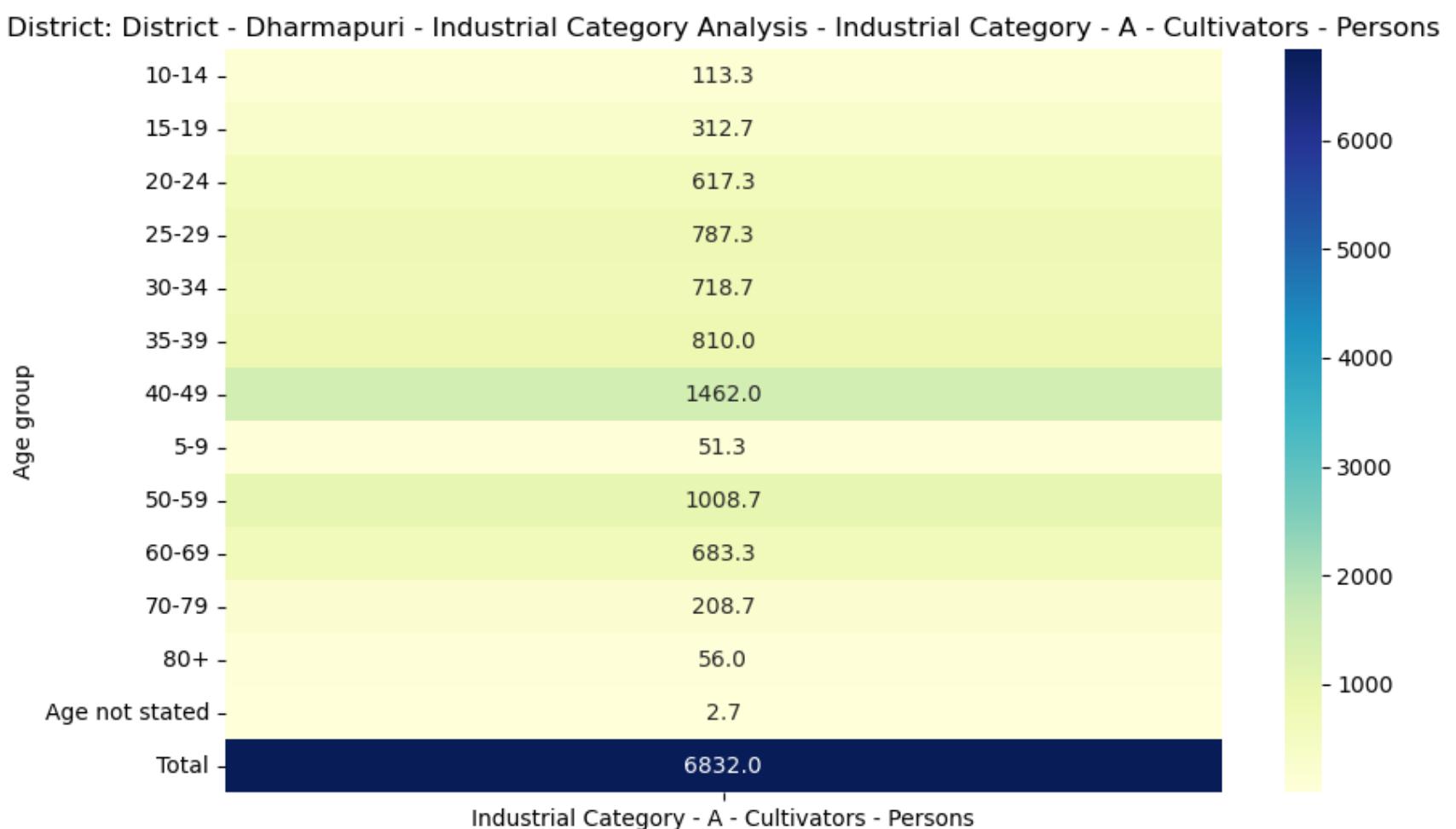
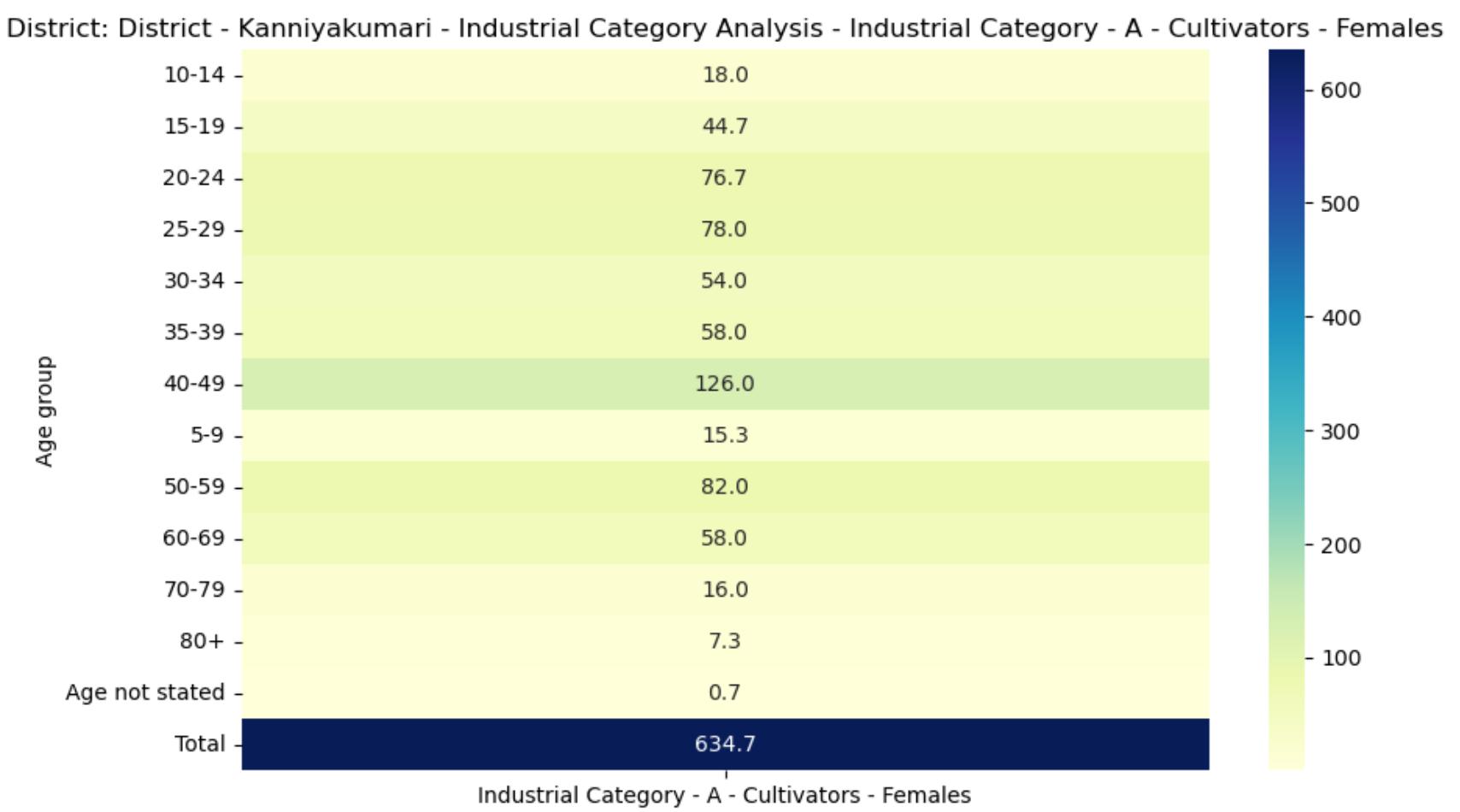


Industrial Category - A - Cultivators - Persons

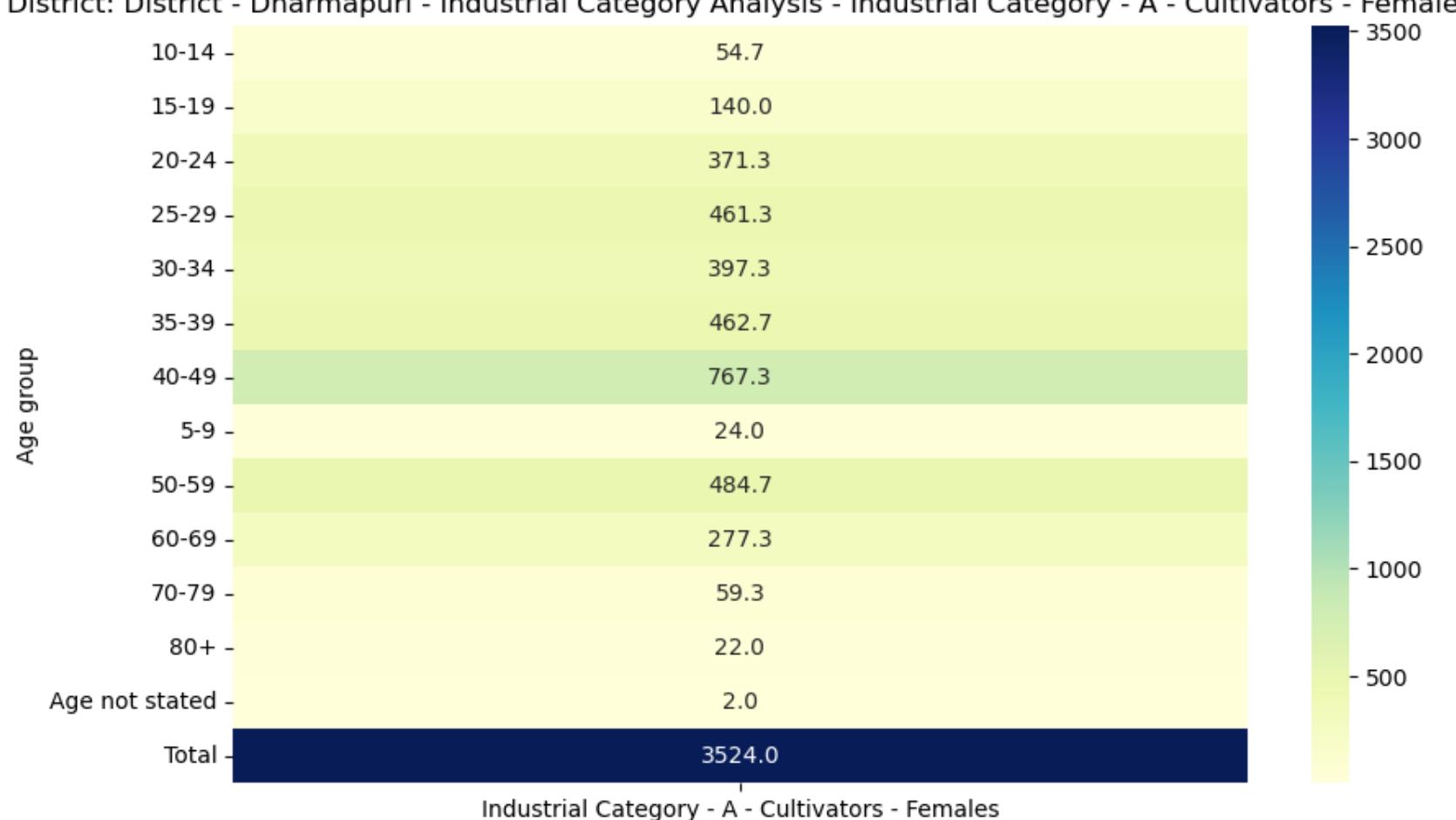
District: District - Kanyakumari - Industrial Category Analysis - Industrial Category - A - Cultivators - Males



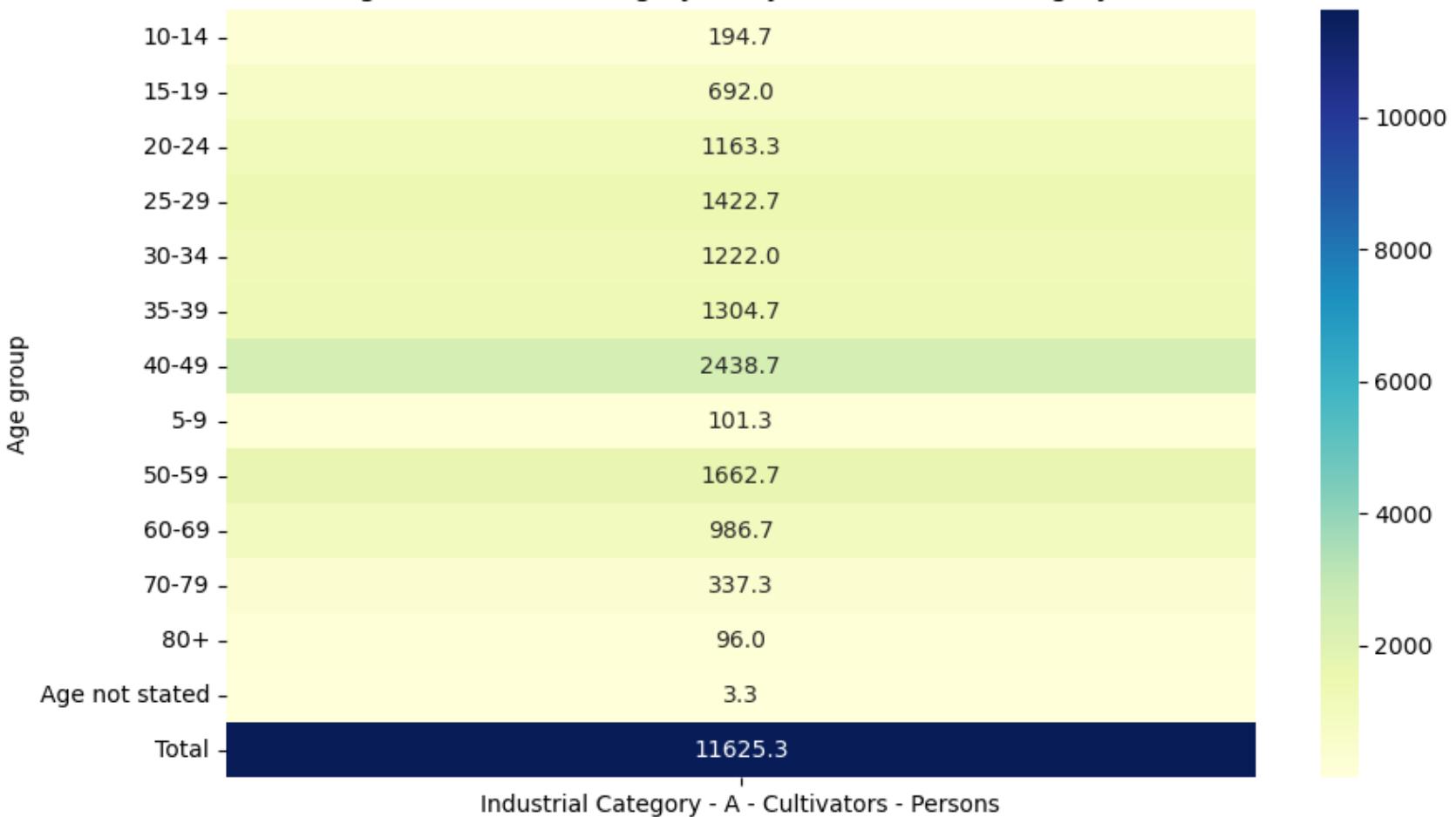
Industrial Category - A - Cultivators - Males



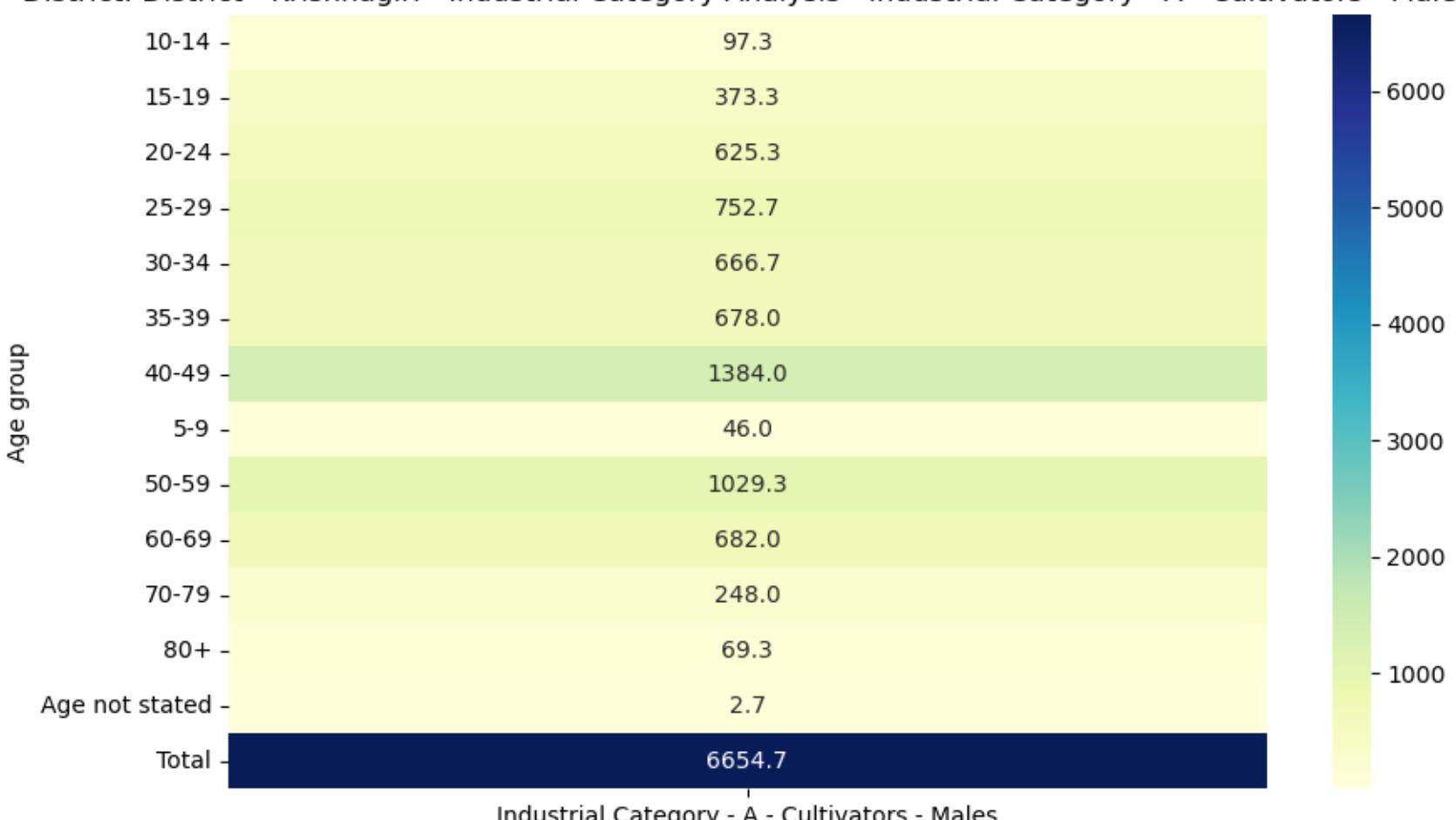
District: District - Dharmapuri - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



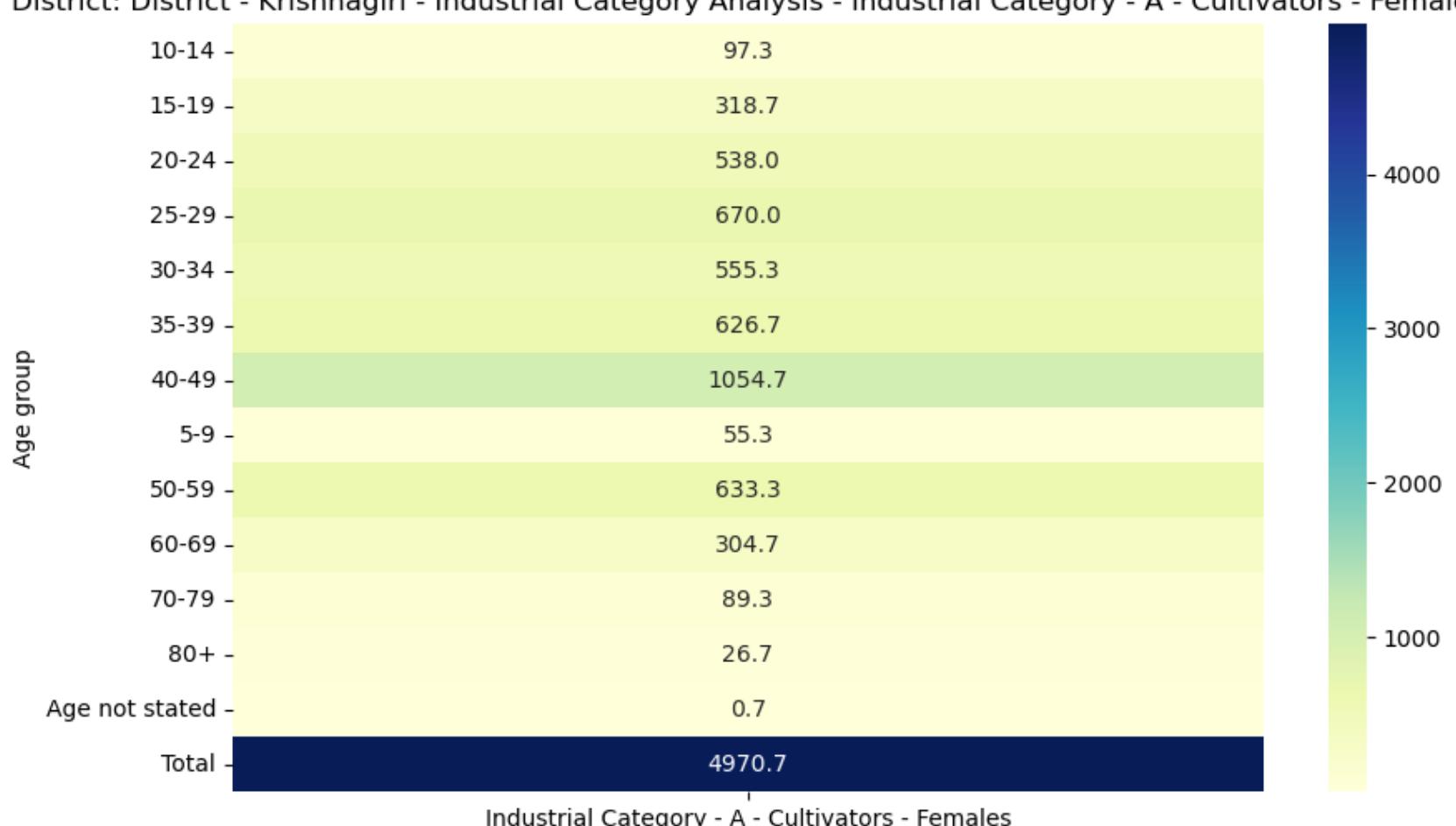
District: District - Krishnagiri - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



District: District - Krishnagiri - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

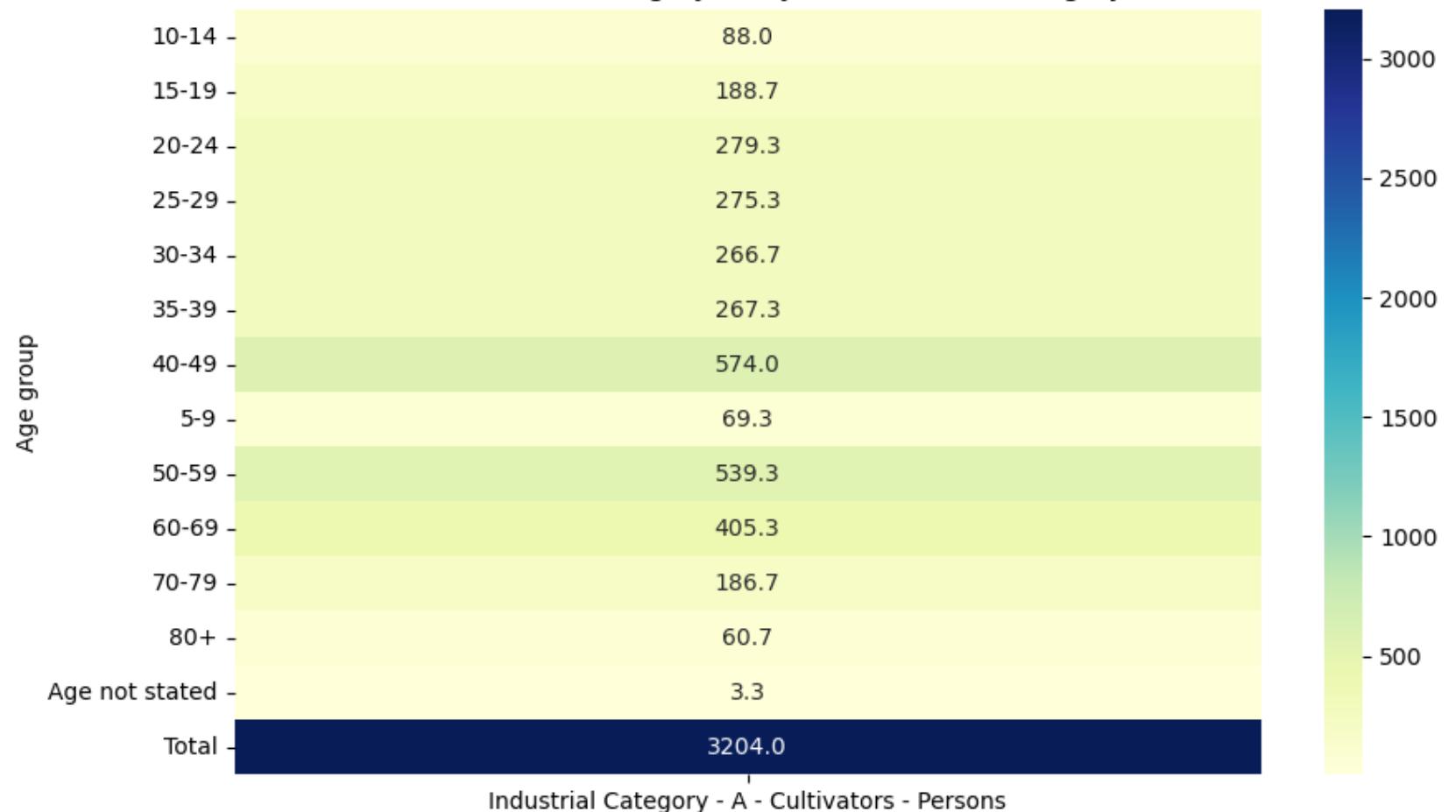


District: District - Krishnagiri - Industrial Category Analysis - Industrial Category - A - Cultivators - Females



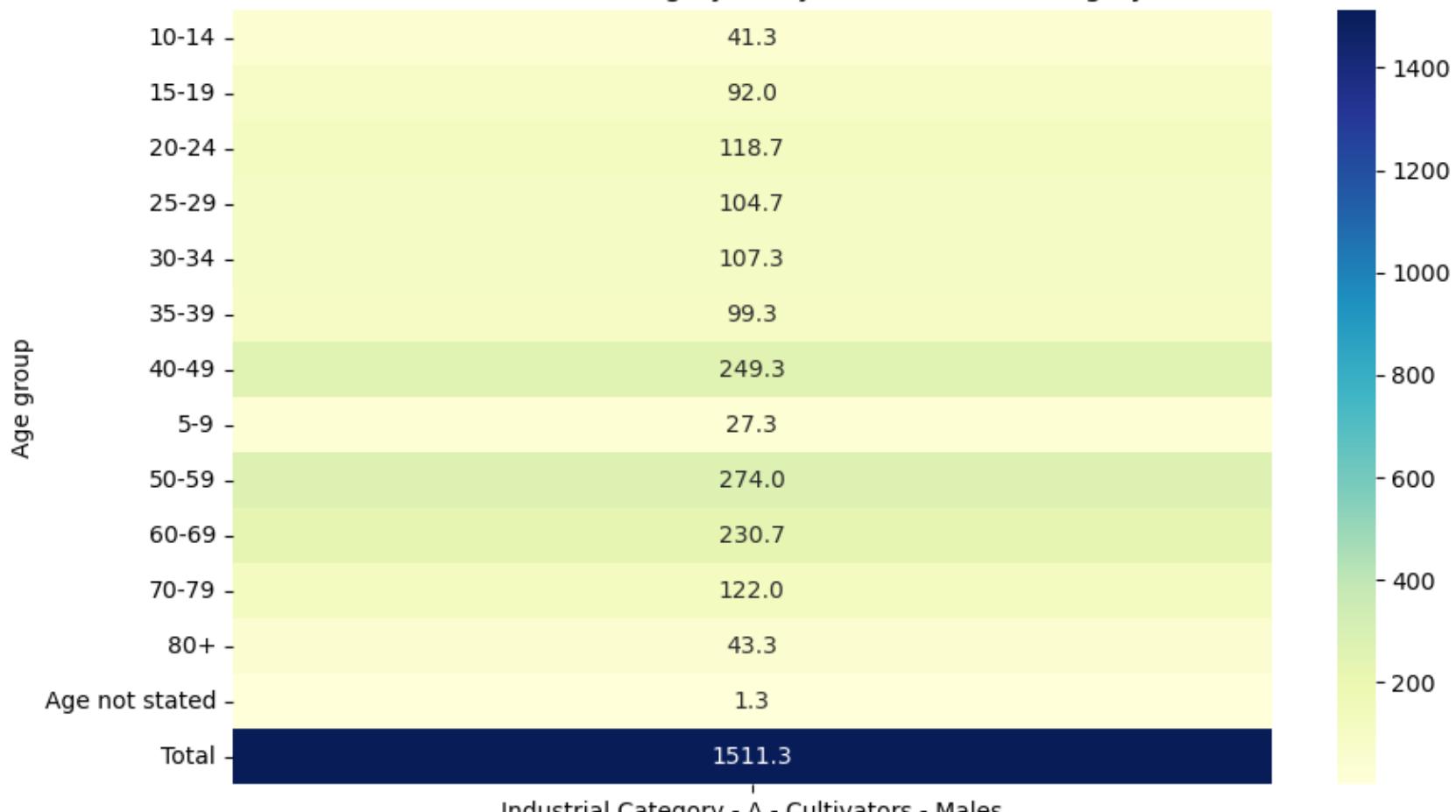
Industrial Category - A - Cultivators - Females

District: District - Coimbatore - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



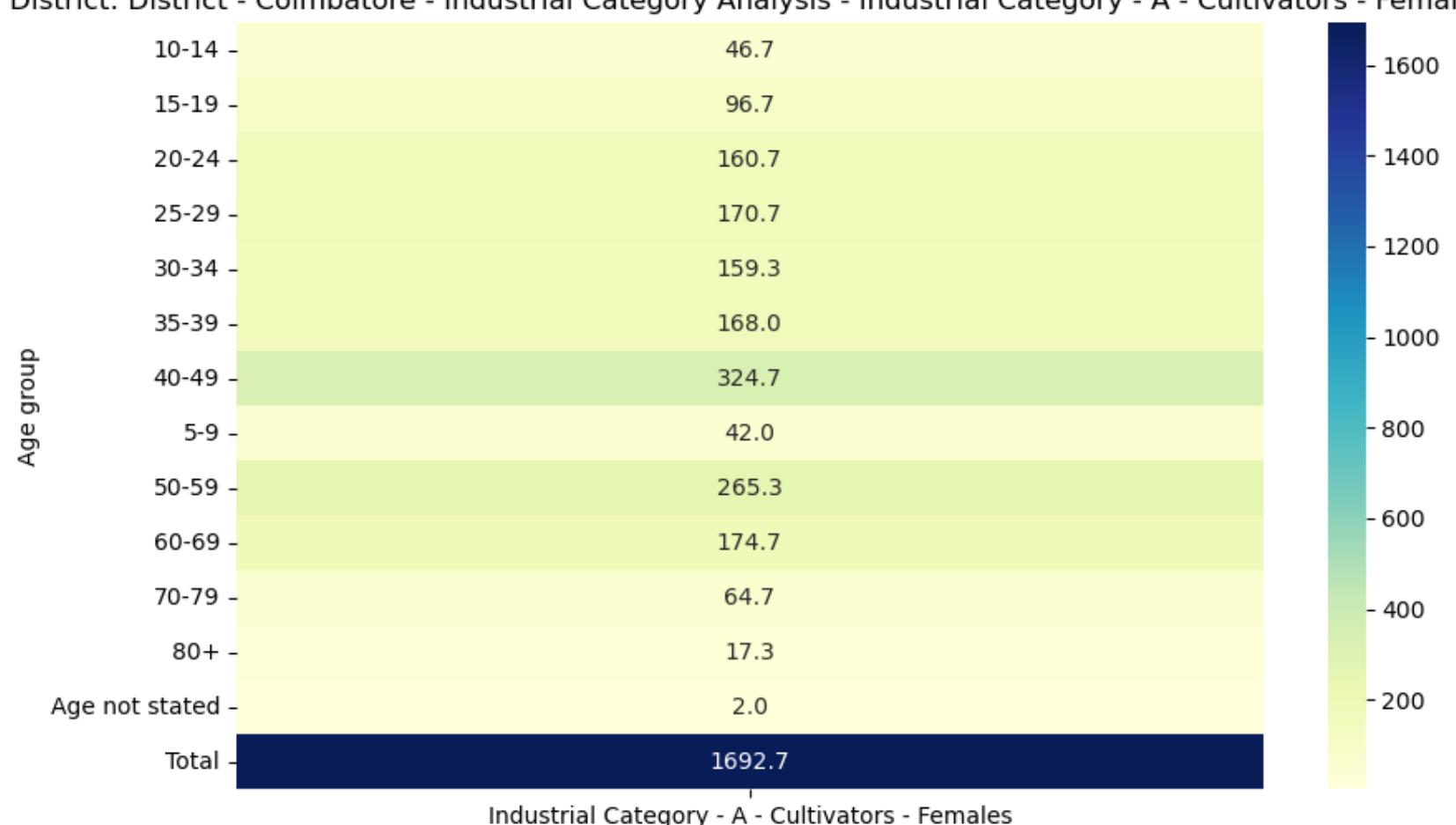
Industrial Category - A - Cultivators - Persons

District: District - Coimbatore - Industrial Category Analysis - Industrial Category - A - Cultivators - Males

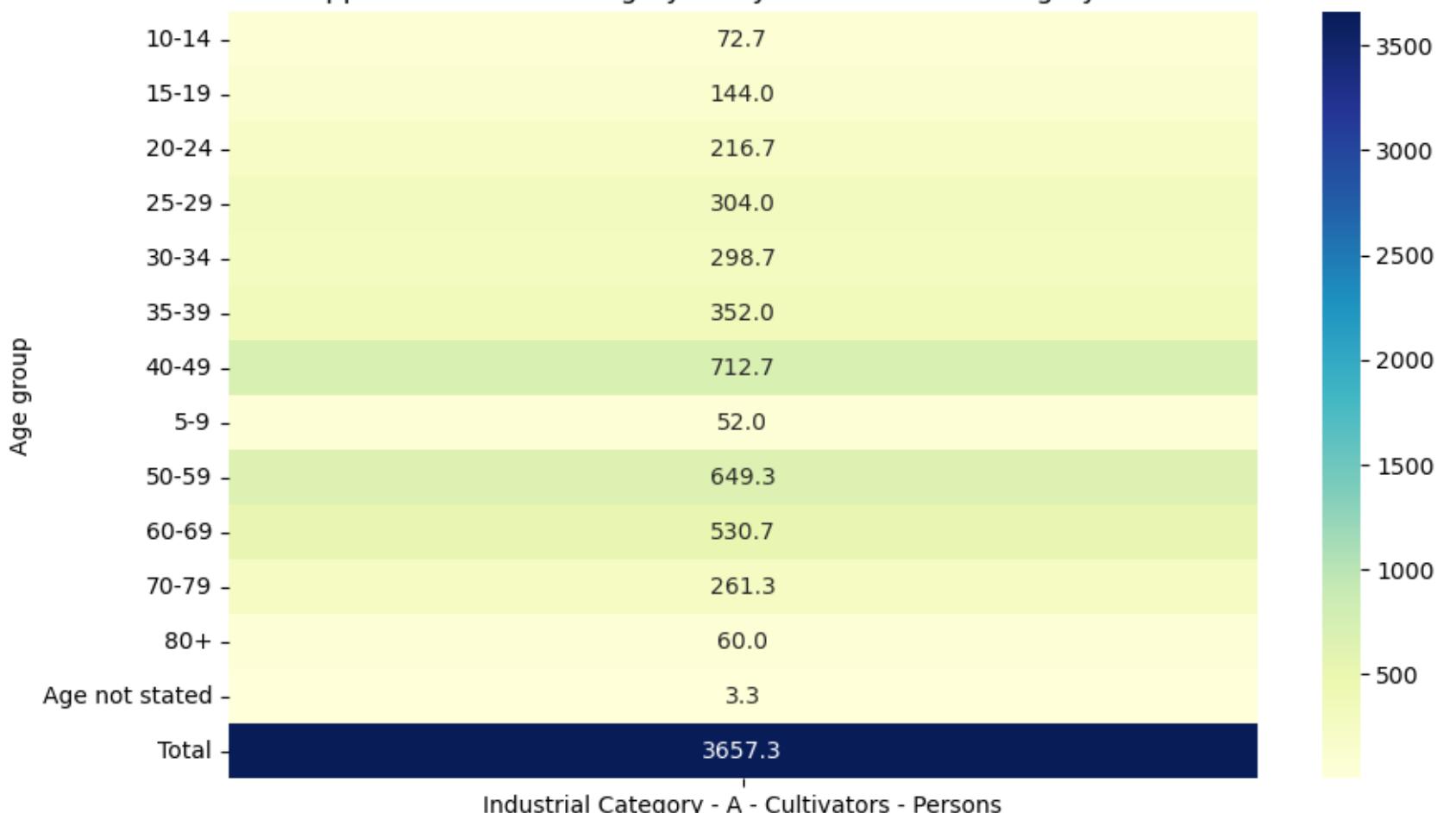


Industrial Category - A - Cultivators - Males

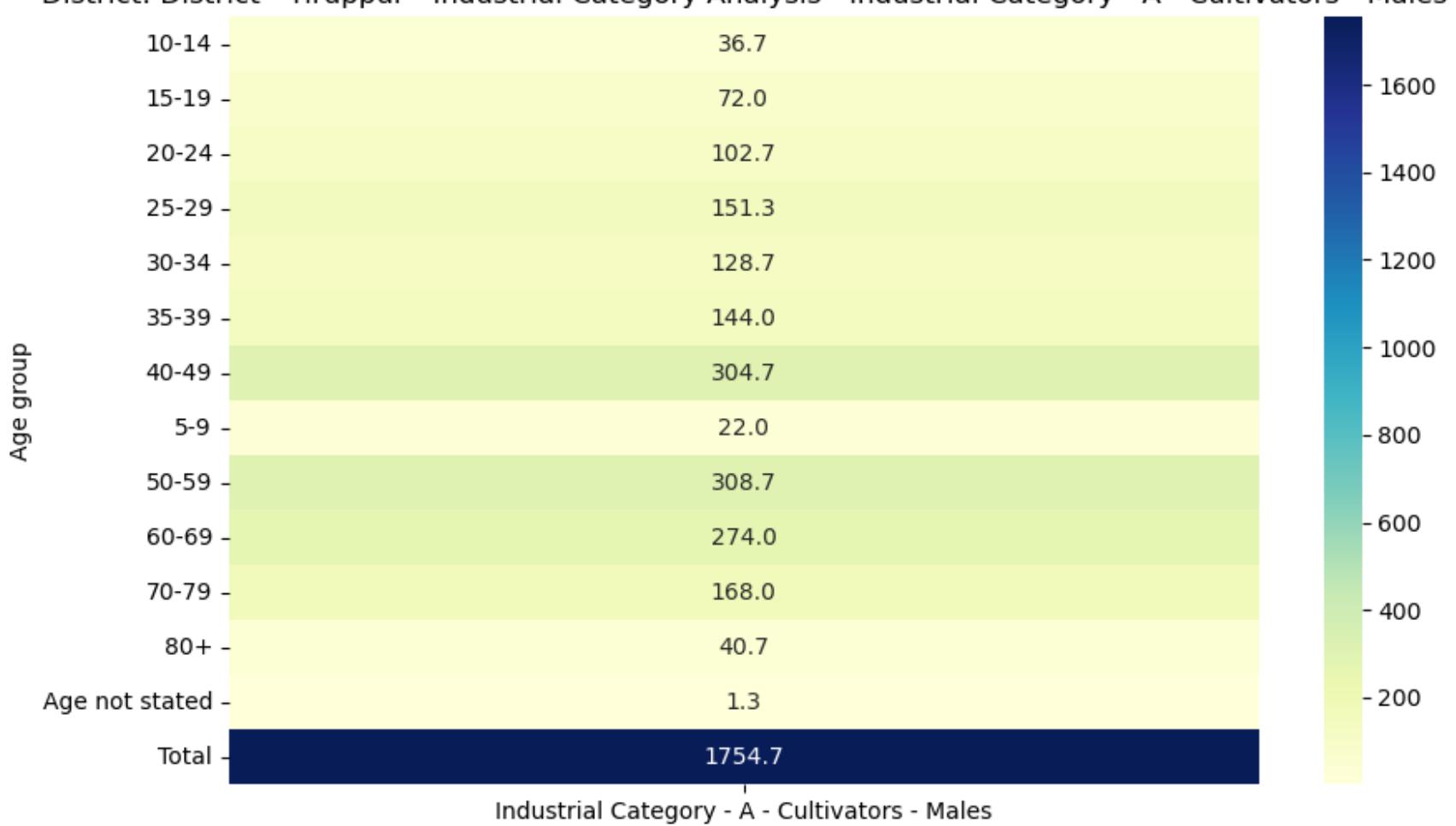
District: District - Coimbatore - Industrial Category Analysis - Industrial Category - A - Cultivators - Females

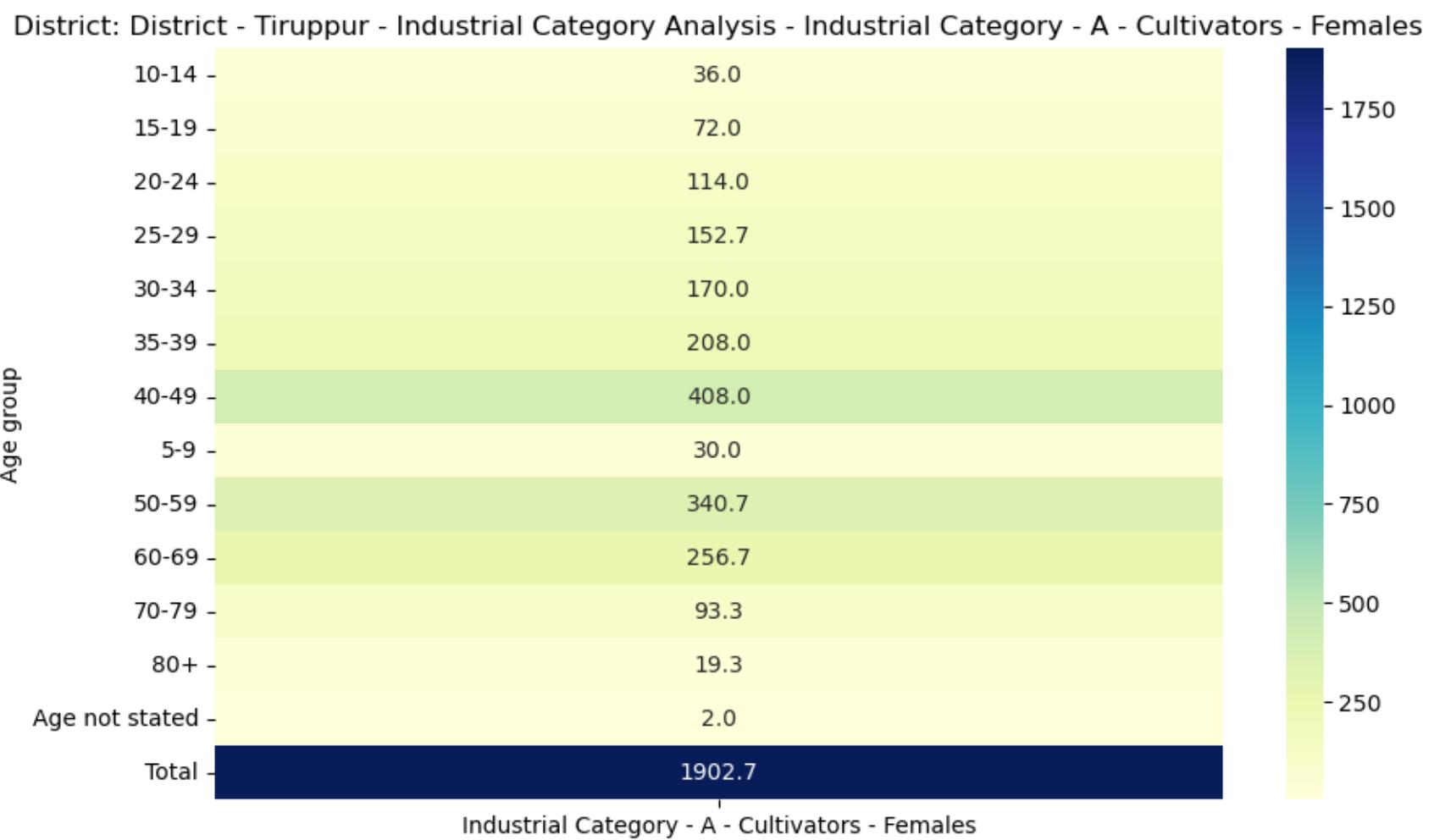


District: District - Tiruppur - Industrial Category Analysis - Industrial Category - A - Cultivators - Persons



District: District - Tiruppur - Industrial Category Analysis - Industrial Category - A - Cultivators - Males





Full State

```
In [ ]: # 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
# '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']

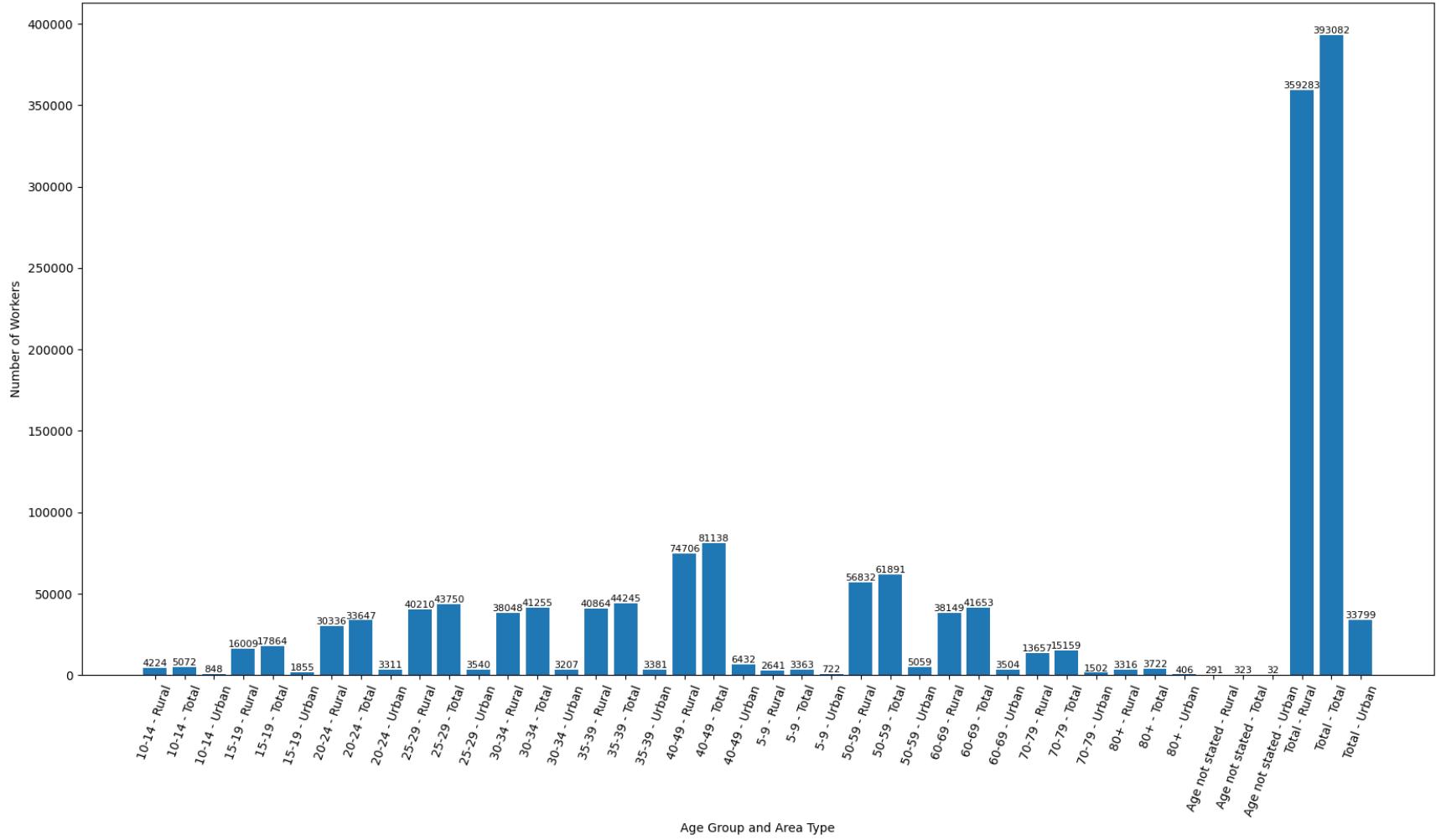
# Grouping by 'Age group', 'Total/ Rural/ Urban' and summing up the number of workers
grouped_data = state_data.groupby(['Age group', 'Total/ Rural/ Urban'])['Industrial Category - A - Cultivators - Person']

# Create the bar chart
plt.figure(figsize=(20, 10))
bars = plt.bar(grouped_data['Age group'] + ' - ' + grouped_data['Total/ Rural/ Urban'], grouped_data['Industrial Catego

# 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()
```

Distribution of workers in State - Tamil Nadu

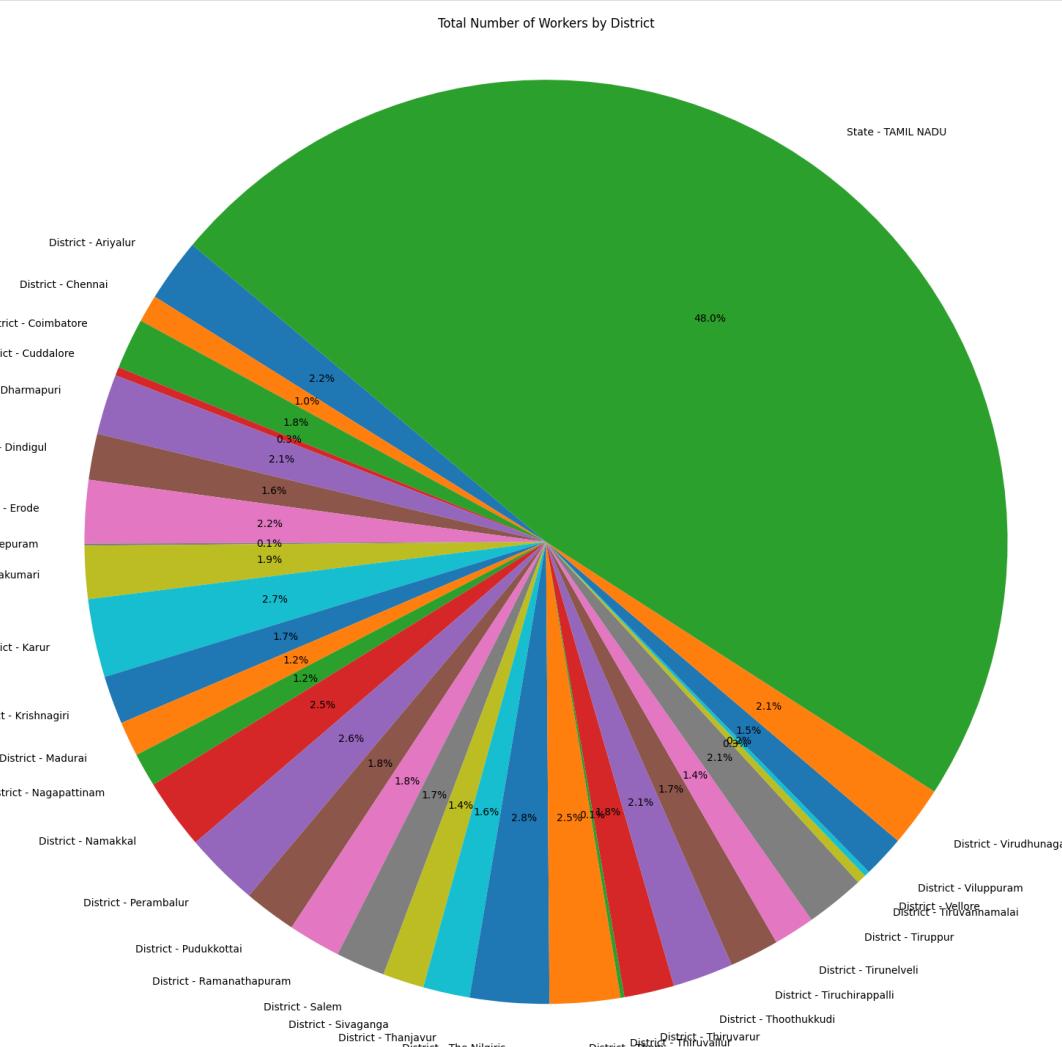


```
In [ ]: import matplotlib.pyplot as plt

# Take the absolute values of district workers
district_workers = df.groupby('Area Name')[['Worked for 3 months or more but less than 6 months - Persons']].sum().abs()

# Create a pie chart
plt.figure(figsize=(30, 18))
plt.pie(district_workers, labels=district_workers.index, autopct='%1.1f%%', startangle=140)
plt.title('Total Number of Workers by District')
plt.axis('equal') # Equal aspect ratio ensures the pie chart is circular.

# Show the plot
plt.show()
```



STEP 2 : DISTRIBUTION OF MARGINAL WORKERS

```
In [ ]: df.columns
```

```
Out[ ]: 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')
```

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

```
# 'Area Name' represents the districts, 'Age group' represents the age groups, 'Total/ Rural/ Urban' represents rural or
# '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()
print(grouped_data)

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

	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	
...	
1381	State - TAMIL NADU	Age not stated	Total	
1382	State - TAMIL NADU	Age not stated	Urban	
1383	State - TAMIL NADU	Total	Rural	
1384	State - TAMIL NADU	Total	Total	
1385	State - TAMIL NADU	Total	Urban	
Industrial Category - A - Cultivators - Persons				
0		68		
1		74		
2		6		
3		411		
4		425		
...		...		
1381		323		
1382		32		
1383		359283		
1384		393082		
1385		33799		

[1386 rows x 4 columns]

```
In [ ]: # List of industrial category columns
industrial_categories = [
    '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'
]

# Initialize an empty list to store the grouped data
grouped_data_list = []

# Loop through industrial categories
```

```
for category in industrial_categories:  
    # Group the data and aggregate the counts of workers  
    grouped_data = df.groupby(['Area Name', 'Age group', 'Total/ Rural/ Urban'])[category].sum().reset_index()  
    # Add the grouped data to the list  
    grouped_data_list.append(grouped_data)  
  
# Now, grouped_data_list contains the grouped data for all industrial categories  
print(grouped_data_list)
```

	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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

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

0		565
1		640
2		75
3		3157
4		3294
...		...
1381		3605
1382		1279
1383		3009302
1384		4218884
1385		1209582

	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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

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

0		309
1		352
2		43
3		1596
4		1669
...		...
1381		1926
1382		758
1383		1443929
1384		2136881
1385		692952

	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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

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

0		256
1		288
2		32
3		1561
4		1625
...		...
1381		1679
1382		521
1383		1565373
1384		2082003
1385		516630

	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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Worked for less than 3 months - Persons

0		76
1		90
2		14

3		650
4		673
...		...
1381		483
1382		144
1383		510909
1384		723891
1385		212982

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Worked for less than 3 months - Males

0		44
1		50
2		6
3		307
4		325
...		...
1381		237
1382		80
1383		216805
1384		337268
1385		120463

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Worked for less than 3 months - Females

0		32
1		40
2		8
3		343
4		348
...		...
1381		246
1382		64
1383		294104
1384		386623
1385		92519

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - A - Cultivators - Persons

0		68
1		74
2		6
3		411
4		425
...		...
1381		323
1382		32
1383		359283
1384		393082
1385		33799

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total

1382 State - TAMIL NADU Age not stated Urban
1383 State - TAMIL NADU Total Rural
1384 State - TAMIL NADU Total Total
1385 State - TAMIL NADU Total Urban

Industrial Category - A - Cultivators - Males

0	35
1	37
2	2
3	203
4	213
...	...
1381	174
1382	23
1383	199440
1384	220314
1385	20874

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - A - Cultivators - Females

0	33
1	37
2	4
3	208
4	212
...	...
1381	149
1382	9
1383	159843
1384	172768
1385	12925

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - A - Agricultural labourers - Persons

0	319
1	332
2	13
3	2440
4	2486
...	...
1381	1444
1382	123
1383	2155158
1384	2372446
1385	217288

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - A - Agricultural labourers - Males

0	158
1	163
2	5
3	1140
4	1165
...	...
1381	645
1382	65
1383	921652
1384	1034184
1385	112532

```
[1386 rows x 4 columns],
   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
...
       ...           ...
1381  State - TAMIL NADU  Age not stated      Total
1382  State - TAMIL NADU  Age not stated      Urban
1383  State - TAMIL NADU        Total          Rural
1384  State - TAMIL NADU        Total          Total
1385  State - TAMIL NADU        Total          Urban
```

Industrial Category - A - Agricultural labourers - Females

0	161
1	169
2	8
3	1300
4	1321
...	...
1381	799
1382	58
1383	1233506
1384	1338262
1385	104756

```
[1386 rows x 4 columns],           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
...
...
1381    State - TAMIL NADU  Age not stated    Total
1382    State - TAMIL NADU  Age not stated    Urban
1383    State - TAMIL NADU        Total        Rural
1384    State - TAMIL NADU        Total        Total
1385    State - TAMIL NADU        Total        Urban
```

Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons

0	0
1	0
2	0
3	22
4	22
...	...
1381	88
1382	17
1383	81901
1384	125099
1385	43198

```
[1386 rows x 4 columns],           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
...
          ...                      ...
1381    State - TAMIL NADU    Age not stated    Total
1382    State - TAMIL NADU    Age not stated    Urban
1383    State - TAMIL NADU        Total        Rural
1384    State - TAMIL NADU        Total        Total
1385    State - TAMIL NADU        Total        Urban
```

Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males

	Number of categories	Number of observations
0		0
1		0
2		0
3		6
4		6
...		...
1381		68
1382		11
1383		48169
1384		78052
1385		29883

```
[1386 rows x 4 columns],
   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
...
          ...
          ...
1381  State - TAMIL NADU  Age not stated      Total
1382  State - TAMIL NADU  Age not stated      Urban
1383  State - TAMIL NADU      Total        Rural
1384  State - TAMIL NADU      Total        Total
1385  State - TAMIL NADU      Total        Urban
```

Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females

Industrial category A plantation, LIVESTOCK,

2		0
3		16
4		16
...		...
1381		20
1382		6
1383		33732
1384		47047
1385		13315

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - B - Persons	
0	0
1	0
2	0
3	0
4	0
...	...
1381	12
1382	0
1383	10266
1384	14979
1385	4713

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - B - Males	
0	0
1	0
2	0
3	0
4	0
...	...
1381	6
1382	0
1383	7004
1384	10290
1385	3286

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - B - Females	
0	0
1	0
2	0
3	0
4	0
...	...
1381	6
1382	0
1383	3262
1384	4689
1385	1427

[1386 rows x 4 columns],		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
...

1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - C - HHI - Persons

0		6	
1		6	
2		0	
3		60	
4		64	
...		...	
1381		88	
1382		24	
1383		93344	
1384		154133	
1385		60789	

[1386 rows x 4 columns], 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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - C - HHI - Males

0		6	
1		6	
2		0	
3		26	
4		30	
...		...	
1381		54	
1382		11	
1383		31962	
1384		53418	
1385		21456	

[1386 rows x 4 columns], 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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - C - HHI - Females

0		0	
1		0	
2		0	
3		34	
4		34	
...		...	
1381		34	
1382		13	
1383		61382	
1384		100715	
1385		39333	

[1386 rows x 4 columns], 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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - C - Non HHI - Persons

0		16	
1		16	
2		0	
3		128	
4		140	
...		...	
1381		231	
1382		114	
1383		138835	
1384		306528	

1385		167693		
[1386 rows x 4 columns],	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		
...		
1381 State - TAMIL NADU	Age not stated	Total		
1382 State - TAMIL NADU	Age not stated	Urban		
1383 State - TAMIL NADU	Total	Rural		
1384 State - TAMIL NADU	Total	Total		
1385 State - TAMIL NADU	Total	Urban		
Industrial Category - C - Non HHI - Males				
0	0			
1	0			
2	0			
3	52			
4	60			
...	...			
1381	147			
1382	75			
1383	83121			
1384	188464			
1385	105343			
[1386 rows x 4 columns],	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		
...		
1381 State - TAMIL NADU	Age not stated	Total		
1382 State - TAMIL NADU	Age not stated	Urban		
1383 State - TAMIL NADU	Total	Rural		
1384 State - TAMIL NADU	Total	Total		
1385 State - TAMIL NADU	Total	Urban		
Industrial Category - C - Non HHI - Females				
0	16			
1	16			
2	0			
3	76			
4	80			
...	...			
1381	84			
1382	39			
1383	55714			
1384	118064			
1385	62350			
[1386 rows x 4 columns],	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		
...		
1381 State - TAMIL NADU	Age not stated	Total		
1382 State - TAMIL NADU	Age not stated	Urban		
1383 State - TAMIL NADU	Total	Rural		
1384 State - TAMIL NADU	Total	Total		
1385 State - TAMIL NADU	Total	Urban		
Industrial Category - D & E - Persons				
0	0			
1	0			
2	0			
3	0			
4	0			
...	...			
1381	9			
1382	9			
1383	3038			
1384	7137			
1385	4099			
[1386 rows x 4 columns],	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		
...		
1381 State - TAMIL NADU	Age not stated	Total		
1382 State - TAMIL NADU	Age not stated	Urban		
1383 State - TAMIL NADU	Total	Rural		
1384 State - TAMIL NADU	Total	Total		
1385 State - TAMIL NADU	Total	Urban		
Industrial Category - D & E - Males				
0	0			

1		0
2		0
3		0
4		0
...		...
1381		9
1382		9
1383		2704
1384		6003
1385		3299

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - D & E - Females			
0		0	
1		0	
2		0	
3		0	
4		0	
...		...	
1381		0	
1382		0	
1383		334	
1384		1134	
1385		800	

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - F - Persons			
0		0	
1		2	
2		2	
3		70	
4		84	
...		...	
1381		297	
1382		147	
1383		231542	
1384		390275	
1385		158733	

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - F - Males			
0		0	
1		2	
2		2	
3		62	
4		74	
...		...	
1381		170	
1382		106	
1383		117854	
1384		241619	
1385		123765	

[1386 rows x 4 columns],		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

...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - F - Females

0	0
1	0
2	0
3	8
4	10
...	...
1381	127
1382	41
1383	113688
1384	148656
1385	34968

[1386 rows x 4 columns],	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
...
1381	State - TAMIL NADU	Age not stated Total
1382	State - TAMIL NADU	Age not stated Urban
1383	State - TAMIL NADU	Total Rural
1384	State - TAMIL NADU	Total Total
1385	State - TAMIL NADU	Total Urban

Industrial Category - G - HHI - Persons

0	0
1	0
2	0
3	0
4	0
...	...
1381	0
1382	0
1383	255
1384	510
1385	255

[1386 rows x 4 columns],	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
...
1381	State - TAMIL NADU	Age not stated Total
1382	State - TAMIL NADU	Age not stated Urban
1383	State - TAMIL NADU	Total Rural
1384	State - TAMIL NADU	Total Total
1385	State - TAMIL NADU	Total Urban

Industrial Category - G - HHI - Males

0	0
1	0
2	0
3	0
4	0
...	...
1381	0
1382	0
1383	242
1384	478
1385	236

[1386 rows x 4 columns],	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
...
1381	State - TAMIL NADU	Age not stated Total
1382	State - TAMIL NADU	Age not stated Urban
1383	State - TAMIL NADU	Total Rural
1384	State - TAMIL NADU	Total Total
1385	State - TAMIL NADU	Total Urban

Industrial Category - G - HHI - Females

0	0
1	0
2	0
3	0
4	0
...	...
1381	0
1382	0
1383	13

1384		32
1385		19

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - G - Non HHI - Persons

0	12
1	14
2	2
3	30
4	36
...	...
1381	134
1382	83
1383	64318
1384	171440
1385	107122

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - G - Non HHI - Males

0	12
1	14
2	2
3	30
4	34
...	...
1381	108
1382	71
1383	46693
1384	129234
1385	82541

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - G - Non HHI - Females

0	0
1	0
2	0
3	0
4	2
...	...
1381	26
1382	12
1383	17625
1384	42206
1385	24581

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - H - Persons

0		0
1		6
2		6
3		62
4		68
...		...
1381		78
1382		40
1383		37949
1384		84686
1385		46737

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - H - Males		
0		0
1		6
2		6
3		62
4		68
...		...
1381		76
1382		38
1383		36972
1384		81715
1385		44743

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - H - Females		
0		0
1		0
2		0
3		0
4		0
...		...
1381		2
1382		2
1383		977
1384		2971
1385		1994

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - I - Persons		
0		0
1		0
2		0
3		24
4		24
...		...
1381		22
1382		14
1383		16644
1384		42321
1385		25677

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - I - Males

0		0
1		0
2		0
3		24
4		24
...		...
1381		22
1382		14
1383		13736
1384		33725
1385		19989

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - I - Females

0		0
1		0
2		0
3		0
4		0
...		...
1381		0
1382		0
1383		2908
1384		8596
1385		5688

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - HHI - Persons

0		0
1		0
2		0
3		0
4		0
...		...
1381		0
1382		0
1383		162
1384		463
1385		301

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - HHI - Males

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

1383		140
1384		350
1385		210

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - HHI - Females

0	0
1	0
2	0
3	0
4	0
...	...
1381	0
1382	0
1383	22
1384	113
1385	91

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - Non HHI - Persons

0	0
1	0
2	0
3	0
4	0
...	...
1381	21
1382	13
1383	4407
1384	23293
1385	18886

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - Non HHI - Males

0	0
1	0
2	0
3	0
4	0
...	...
1381	14
1382	6
1383	3386
1384	16984
1385	13598

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - J - Non HHI - Females

0		0
1		0
2		0
3		0
4		0
...		...
1381		7
1382		7
1383		1021
1384		6309
1385		5288

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - K to M - Persons

0		0
1		0
2		0
3		0
4		0
...		...
1381		15
1382		9
1383		7010
1384		26047
1385		19037

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - K to M - Males

0		0
1		0
2		0
3		0
4		0
...		...
1381		15
1382		9
1383		5817
1384		21250
1385		15433

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - K to M - Females

0		0
1		0
2		0
3		0
4		0
...		...
1381		0
1382		0
1383		1193
1384		4797
1385		3604

		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - N to O - Persons

0	0
1	0
2	0
3	0
4	0
...	...
1381	66
1382	51
1383	15844
1384	56495
1385	40651

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - N to O - Males

0	0
1	0
2	0
3	0
4	0
...	...
1381	55
1382	46
1383	11143
1384	42000
1385	30857

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - N to O - Females

0	0
1	0
2	0
3	0
4	0
...	...
1381	11
1382	5
1383	4701
1384	14495
1385	9794

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - P to Q - Persons

0	0
1	0
2	0
3	12
4	12
...	...
1381	41

1382		17
1383		23473
1384		58788
1385		35315

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - P to Q - Males		
0		0
1		0
2		0
3		6
4		6
...		...
1381		17
1382		5
1383		8436
1384		19892
1385		11456

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - P to Q - Females		
0		0
1		0
2		0
3		6
4		6
...		...
1381		24
1382		12
1383		15037
1384		38896
1385		23859

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - R to U - HHI - Persons		
0		0
1		10
2		10
3		72
4		74
...		...
1381		98
1382		39
1383		51408
1384		89703
1385		38295

[1386 rows x 4 columns],		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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - R to U - HHI - Males

0		0
1		6
2		6
3		34
4		34
...		...
1381		34
1382		9
1383		13231
1384		21366
1385		8135

[1386 rows x 4 columns],			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	
...	
1381	State - TAMIL NADU	Age not stated	Total	
1382	State - TAMIL NADU	Age not stated	Urban	
1383	State - TAMIL NADU	Total	Rural	
1384	State - TAMIL NADU	Total	Total	
1385	State - TAMIL NADU	Total	Urban	

Industrial Category - R to U - HHI - Females

0		0
1		4
2		4
3		38
4		40
...		...
1381		64
1382		30
1383		38177
1384		68337
1385		30160

[1386 rows x 4 columns],			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	
...	
1381	State - TAMIL NADU	Age not stated	Total	
1382	State - TAMIL NADU	Age not stated	Urban	
1383	State - TAMIL NADU	Total	Rural	
1384	State - TAMIL NADU	Total	Total	
1385	State - TAMIL NADU	Total	Urban	

Industrial Category - R to U - Non HHI - Persons

0		220
1		270
2		50
3		476
4		532
...		...
1381		1121
1382		691
1383		225374
1384		625350
1385		399976

[1386 rows x 4 columns],			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	
...	
1381	State - TAMIL NADU	Age not stated	Total	
1382	State - TAMIL NADU	Age not stated	Urban	
1383	State - TAMIL NADU	Total	Rural	
1384	State - TAMIL NADU	Total	Total	
1385	State - TAMIL NADU	Total	Urban	

Industrial Category - R to U - Non HHI - Males

0		142
1		168
2		26
3		258
4		280
...		...
1381		549
1382		340
1383		109032
1384		274811
1385		165779

[1386 rows x 4 columns],			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
...
1381	State - TAMIL NADU	Age not stated	Total
1382	State - TAMIL NADU	Age not stated	Urban
1383	State - TAMIL NADU	Total	Rural
1384	State - TAMIL NADU	Total	Total
1385	State - TAMIL NADU	Total	Urban

Industrial Category - R to U - Non HHI - Females			
0		78	
1		102	
2		24	
3		218	
4		252	
...		...	
1381		572	
1382		351	
1383		116342	
1384		350539	
1385		234197	

[1386 rows x 4 columns]]

```
In [ ]: # Define the list of industrial category columns
industrial_categories = [
    'Industrial Category - A - Cultivators - Persons',
    'Industrial Category - A - Cultivators - Males',
    'Industrial Category - A - Cultivators - Females',
    # Add all other industrial category columns here
]

# Create an empty dictionary to store the aggregated data
aggregated_data = {}

# Loop through industrial categories
for category in industrial_categories:
    # Group the data by 'Age group' and 'Total/ Rural/ Urban' and aggregate the counts of marginal workers
    grouped_data = df.groupby(['Age group', 'Total/ Rural/ Urban'])[category].sum().reset_index()

    # Store the grouped and aggregated data in the dictionary
    aggregated_data[category] = grouped_data

# Now, 'aggregated_data' contains the aggregated data for each industrial category
print(aggregated_data)
```

```

{'Industrial Category - A - Cultivators - Persons': Age group Total/ Rural/ Urban \
0      10-14          Rural
1      10-14          Total
2      10-14          Urban
3      15-19          Rural
4      15-19          Total
5      15-19          Urban
6      20-24          Rural
7      20-24          Total
8      20-24          Urban
9      25-29          Rural
10     25-29          Total
11     25-29          Urban
12     30-34          Rural
13     30-34          Total
14     30-34          Urban
15     35-39          Rural
16     35-39          Total
17     35-39          Urban
18     40-49          Rural
19     40-49          Total
20     40-49          Urban
21      5-9           Rural
22      5-9           Total
23      5-9           Urban
24     50-59          Rural
25     50-59          Total
26     50-59          Urban
27     60-69          Rural
28     60-69          Total
29     60-69          Urban
30     70-79          Rural
31     70-79          Total
32     70-79          Urban
33      80+           Rural
34      80+           Total
35      80+           Urban
36  Age not stated   Rural
37  Age not stated   Total
38  Age not stated   Urban
39      Total          Rural
40      Total          Total
41      Total          Urban

Industrial Category - A - Cultivators - Persons
0                      8448
1                     10144
2                      1696
3                     32018
4                     35728
5                      3710
6                     60672
7                     67294
8                      6622
9                     80420
10                    87500
11                    7080
12                    76096
13                    82510
14                    6414
15                    81728
16                    88490
17                    6762
18                   149412
19                   162276
20                    12864
21                    5282
22                    6726
23                    1444
24                   113664
25                   123782
26                    10118
27                    76298
28                   83306
29                    7008
30                   27314
31                   30318
32                    3004
33                    6632
34                    7444
35                     812
36                     582
37                     646
38                       64
39                   718566
40                   786164
41                   67598 , 'Industrial Category - A - Cultivators - Males': Age gr
oup Total/ Rural/ Urban \
0      10-14          Rural
1      10-14          Total
2      10-14          Urban
3      15-19          Rural
4      15-19          Total
5      15-19          Urban
6      20-24          Rural

```

7	20-24	Total
8	20-24	Urban
9	25-29	Rural
10	25-29	Total
11	25-29	Urban
12	30-34	Rural
13	30-34	Total
14	30-34	Urban
15	35-39	Rural
16	35-39	Total
17	35-39	Urban
18	40-49	Rural
19	40-49	Total
20	40-49	Urban
21	5-9	Rural
22	5-9	Total
23	5-9	Urban
24	50-59	Rural
25	50-59	Total
26	50-59	Urban
27	60-69	Rural
28	60-69	Total
29	60-69	Urban
30	70-79	Rural
31	70-79	Total
32	70-79	Urban
33	80+	Rural
34	80+	Total
35	80+	Urban
36	Age not stated	Rural
37	Age not stated	Total
38	Age not stated	Urban
39	Total	Rural
40	Total	Total
41	Total	Urban

Industrial Category - A - Cultivators - Males

0	4072
1	4972
2	900
3	16714
4	18672
5	1958
6	30446
7	33860
8	3414
9	40726
10	44578
11	3852
12	41124
13	44892
14	3768
15	42092
16	46174
17	4082
18	81734
19	89984
20	8250
21	2458
22	3184
23	726
24	65570
25	72366
26	6796
27	48798
28	53878
29	5080
30	19988
31	22266
32	2278
33	4856
34	5454
35	598
36	302
37	348
38	46
39	398880
40	440628
41	41748

, 'Industrial Category - A - Cultivators - Females':

Age gr

0	10-14	Rural
1	10-14	Total
2	10-14	Urban
3	15-19	Rural
4	15-19	Total
5	15-19	Urban
6	20-24	Rural
7	20-24	Total
8	20-24	Urban
9	25-29	Rural
10	25-29	Total
11	25-29	Urban
12	30-34	Rural
13	30-34	Total
14	30-34	Urban

```

15      35-39          Rural
16      35-39          Total
17      35-39          Urban
18      40-49          Rural
19      40-49          Total
20      40-49          Urban
21      5-9            Rural
22      5-9            Total
23      5-9            Urban
24      50-59          Rural
25      50-59          Total
26      50-59          Urban
27      60-69          Rural
28      60-69          Total
29      60-69          Urban
30      70-79          Rural
31      70-79          Total
32      70-79          Urban
33      80+             Rural
34      80+             Total
35      80+             Urban
36  Age not stated   Rural
37  Age not stated   Total
38  Age not stated   Urban
39      Total           Rural
40      Total           Total
41      Total           Urban

```

```

Industrial Category - A - Cultivators - Females
0                      4376
1                      5172
2                      796
3                     15304
4                     17056
5                     1752
6                     30226
7                     33434
8                     3208
9                     39694
10                    42922
11                    3228
12                    34972
13                    37618
14                    2646
15                    39636
16                    42316
17                    2680
18                    67678
19                    72292
20                    4614
21                    2824
22                    3542
23                    718
24                    48094
25                    51416
26                    3322
27                    27500
28                    29428
29                    1928
30                    7326
31                    8052
32                    726
33                    1776
34                    1990
35                    214
36                    280
37                    298
38                     18
39                  319686
40                  345536
41                  25850  }

```

```

In [ ]: # Define the list of industrial category columns
industrial_categories = [
    'Industrial Category - A - Cultivators - Persons',
    'Industrial Category - A - Cultivators - Males',
    'Industrial Category - A - Cultivators - Females',
    # Add all other industrial category columns here
]

# Create an empty dictionary to store the aggregated data
aggregated_data = {}

# Loop through industrial categories
for category in industrial_categories:
    # Group the data by 'Age group' and 'Total/ Rural/ Urban' and aggregate the counts of marginal workers
    grouped_data = df.groupby(['Age group', 'Total/ Rural/ Urban'])[category].mean().reset_index()

    # Store the grouped and aggregated data in the dictionary
    aggregated_data[category] = grouped_data

# Now, 'aggregated_data' contains the aggregated data for each industrial category
print(aggregated_data)

```

```

{'Industrial Category - A - Cultivators - Persons':      Age group Total/ Rural/ Urban \n
0          10-14           Rural
1          10-14           Total
2          10-14           Urban
3          15-19           Rural
4          15-19           Total
5          15-19           Urban
6          20-24           Rural
7          20-24           Total
8          20-24           Urban
9          25-29           Rural
10         25-29          Total
11         25-29          Urban
12         30-34           Rural
13         30-34           Total
14         30-34           Urban
15         35-39           Rural
16         35-39           Total
17         35-39           Urban
18         40-49           Rural
19         40-49           Total
20         40-49           Urban
21         5-9             Rural
22         5-9             Total
23         5-9             Urban
24         50-59           Rural
25         50-59           Total
26         50-59           Urban
27         60-69           Rural
28         60-69           Total
29         60-69           Urban
30         70-79           Rural
31         70-79           Total
32         70-79           Urban
33         80+             Rural
34         80+             Total
35         80+             Urban
36  Age not stated     Rural
37  Age not stated     Total
38  Age not stated     Urban
39          Total         Rural
40          Total         Total
41          Total         Urban

Industrial Category - A - Cultivators - Persons
0                      256.000000
1                      307.393939
2                      51.393939
3                     970.242424
4                    1082.666667
5                     112.424242
6                    1838.545455
7                    2039.212121
8                     200.666667
9                    2436.969697
10                   2651.515152
11                   214.545455
12                   2305.939394
13                   2500.303030
14                   194.363636
15                   2476.606061
16                   2681.515152
17                   204.909091
18                   4527.636364
19                   4917.454545
20                   389.818182
21                   160.060606
22                   203.818182
23                   43.757576
24                   3444.363636
25                   3750.969697
26                   306.606061
27                   2312.060606
28                   2524.424242
29                   212.363636
30                   827.696970
31                   918.727273
32                   91.030303
33                   200.969697
34                   225.575758
35                   24.606061
36                   17.636364
37                   19.575758
38                   1.939394
39                   21774.727273
40                   23823.151515
41                   2048.424242 , 'Industrial Category - A - Cultivators - Males': Age gr
oup Total/ Rural/ Urban \
0          10-14           Rural
1          10-14           Total
2          10-14           Urban
3          15-19           Rural
4          15-19           Total
5          15-19           Urban
6          20-24           Rural

```

7	20-24	Total
8	20-24	Urban
9	25-29	Rural
10	25-29	Total
11	25-29	Urban
12	30-34	Rural
13	30-34	Total
14	30-34	Urban
15	35-39	Rural
16	35-39	Total
17	35-39	Urban
18	40-49	Rural
19	40-49	Total
20	40-49	Urban
21	5-9	Rural
22	5-9	Total
23	5-9	Urban
24	50-59	Rural
25	50-59	Total
26	50-59	Urban
27	60-69	Rural
28	60-69	Total
29	60-69	Urban
30	70-79	Rural
31	70-79	Total
32	70-79	Urban
33	80+	Rural
34	80+	Total
35	80+	Urban
36	Age not stated	Rural
37	Age not stated	Total
38	Age not stated	Urban
39	Total	Rural
40	Total	Total
41	Total	Urban

Industrial Category - A - Cultivators - Males

0	123.393939
1	150.666667
2	27.272727
3	506.484848
4	565.818182
5	59.333333
6	922.606061
7	1026.060606
8	103.454545
9	1234.121212
10	1350.848485
11	116.727273
12	1246.181818
13	1360.363636
14	114.181818
15	1275.515152
16	1399.212121
17	123.696970
18	2476.787879
19	2726.787879
20	250.000000
21	74.484848
22	96.484848
23	22.000000
24	1986.969697
25	2192.909091
26	205.939394
27	1478.727273
28	1632.666667
29	153.939394
30	605.696970
31	674.727273
32	69.030303
33	147.151515
34	165.272727
35	18.121212
36	9.151515
37	10.545455
38	1.393939
39	12087.272727
40	13352.363636
41	1265.090909

, 'Industrial Category - A - Cultivators - Females':

Age gr

oup Total/	Rural/	Urban \
0	10-14	Rural
1	10-14	Total
2	10-14	Urban
3	15-19	Rural
4	15-19	Total
5	15-19	Urban
6	20-24	Rural
7	20-24	Total
8	20-24	Urban
9	25-29	Rural
10	25-29	Total
11	25-29	Urban
12	30-34	Rural
13	30-34	Total
14	30-34	Urban

15	35-39	Rural
16	35-39	Total
17	35-39	Urban
18	40-49	Rural
19	40-49	Total
20	40-49	Urban
21	5-9	Rural
22	5-9	Total
23	5-9	Urban
24	50-59	Rural
25	50-59	Total
26	50-59	Urban
27	60-69	Rural
28	60-69	Total
29	60-69	Urban
30	70-79	Rural
31	70-79	Total
32	70-79	Urban
33	80+	Rural
34	80+	Total
35	80+	Urban
36	Age not stated	Rural
37	Age not stated	Total
38	Age not stated	Urban
39	Total	Rural
40	Total	Total
41	Total	Urban

Industrial Category - A - Cultivators - Females

0	132.606061
1	156.727273
2	24.121212
3	463.757576
4	516.848485
5	53.090909
6	915.939394
7	1013.151515
8	97.212121
9	1202.848485
10	1300.666667
11	97.818182
12	1059.757576
13	1139.939394
14	80.181818
15	1201.090909
16	1282.303030
17	81.212121
18	2050.848485
19	2190.666667
20	139.818182
21	85.575758
22	107.333333
23	21.757576
24	1457.393939
25	1558.060606
26	100.666667
27	833.333333
28	891.757576
29	58.424242
30	222.000000
31	244.000000
32	22.000000
33	53.818182
34	60.303030
35	6.484848
36	8.484848
37	9.030303
38	0.545455
39	9687.454545
40	10470.787879
41	783.333333 }

```
In [ ]: # Apply one-hot encoding for 'Table Code'
df_encoded_table_code = pd.get_dummies(df, columns=['Table Code'], prefix=['Table_Code'])
df_encoded_table_code
```

Out[]:

	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	Worked for less than 3 months	...	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
						- Persons	- Males	- Females	- Persons	- Males	...				
0	33	000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	337268	...	58788	19892	38896	89703
1	33	000	State - TAMIL NADU	Total	5-9	48238	24511	23727	2051	1021	...	312	169	143	842
2	33	000	State - TAMIL NADU	Total	10-14	76288	39191	37097	6993	3716	...	506	256	250	1523
3	33	000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	23489	...	2114	695	1419	5349
4	33	000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	42916	...	11529	2861	8668	10653
...
1381	33	633	District - Tiruppur	Urban	50-59	4965	2800	2165	901	462	...	111	51	60	119
1382	33	633	District - Tiruppur	Urban	60-69	2827	1590	1237	578	307	...	21	6	15	71
1383	33	633	District - Tiruppur	Urban	70-79	920	581	339	204	124	...	6	6	0	22
1384	33	633	District - Tiruppur	Urban	80+	191	104	87	47	32	...	2	0	2	13
1385	33	633	District - Tiruppur	Urban	Age not stated	31	23	8	9	5	...	0	0	0	3

1386 rows × 69 columns

STEP 3 Feature Engineering

In []: #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])
```

In []: # Apply one-hot encoding for 'State Code'
df_encoded_state_code = pd.get_dummies(df, columns=['State Code'], prefix=['State_Code'])
df_encoded_state_code

Out[]:

Table 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 less than 3 months	Worked for less than 3 months	...	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		
					- Persons	- Males	- Females	- Persons	... - Males	... - Females	... - Males	... - Females	... - Persons		
0	B0706	000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	1.000000	1.000000	...	58788	19892	38896	89703
1	B0706	000	State - TAMIL NADU	Total	5-9	48238	24511	23727	0.002833	0.003027	...	312	169	143	842
2	B0706	000	State - TAMIL NADU	Total	10-14	76288	39191	37097	0.009660	0.011018	...	506	256	250	1523
3	B0706	000	State - TAMIL NADU	Total	15-19	257605	141262	116343	0.057934	0.069645	...	2114	695	1419	5349
4	B0706	000	State - TAMIL NADU	Total	20-24	478082	257149	220933	0.111945	0.127246	...	11529	2861	8668	10653
...	
1381	B0706	633	District - Tiruppur	Urban	50-59	4965	2800	2165	0.001245	0.001370	...	111	51	60	119
1382	B0706	633	District - Tiruppur	Urban	60-69	2827	1590	1237	0.000798	0.000910	...	21	6	15	71
1383	B0706	633	District - Tiruppur	Urban	70-79	920	581	339	0.000282	0.000368	...	6	6	0	22
1384	B0706	633	District - Tiruppur	Urban	80+	191	104	87	0.000065	0.000095	...	2	0	2	13
1385	B0706	633	District - Tiruppur	Urban	Age not stated	31	23	8	0.000012	0.000015	...	0	0	0	3

1386 rows × 69 columns

In []: # Apply one-hot encoding for 'District Code'
df_encoded_district_code = pd.get_dummies(df, columns=['District Code'], prefix=['District_Code'])
df_encoded_district_code

Out[]:

Table Code	State 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 less than 3 months		... Persons	District_Code_624	District_Code_625	District_
					Persons	Males	Females	Persons	Males	Persons				
0	B0706	33	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	337268	...	0	0	0
1	B0706	33	State - TAMIL NADU	Total	5-9	48238	24511	23727	2051	1021	...	0	0	0
2	B0706	33	State - TAMIL NADU	Total	10-14	76288	39191	37097	6993	3716	...	0	0	0
3	B0706	33	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	23489	...	0	0	0
4	B0706	33	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	42916	...	0	0	0
...
1381	B0706	33	District - Tiruppur	Urban	50-59	4965	2800	2165	901	462	...	0	0	0
1382	B0706	33	District - Tiruppur	Urban	60-69	2827	1590	1237	578	307	...	0	0	0
1383	B0706	33	District - Tiruppur	Urban	70-79	920	581	339	204	124	...	0	0	0
1384	B0706	33	District - Tiruppur	Urban	80+	191	104	87	47	32	...	0	0	0
1385	B0706	33	District - Tiruppur	Urban	Age not stated	31	23	8	9	5	...	0	0	0

1386 rows × 101 columns

In []: # Apply one-hot encoding for 'Total/ Rural/ Urban'
df_encoded_total_rural_urban = pd.get_dummies(df, columns=['Total/ Rural/ Urban'], prefix=['Total_Rural_Urban'])
df_encoded_total_rural_urban

Out[]:

Table Code	State Code	District Code	Area Name	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	...	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
0	B0706	33	000	State - TAMIL NADU	Total	4218884	2136881	2082003	723891	337268	...	38896	89703	21366	68337
1	B0706	33	000	State - TAMIL NADU	5-9	48238	24511	23727	2051	1021	...	143	842	386	456
2	B0706	33	000	State - TAMIL NADU	10-14	76288	39191	37097	6993	3716	...	250	1523	576	947
3	B0706	33	000	State - TAMIL NADU	15-19	257605	141262	116343	41938	23489	...	1419	5349	2065	3284
4	B0706	33	000	State - TAMIL NADU	20-24	478082	257149	220933	81036	42916	...	8668	10653	2478	8175
...
1381	B0706	33	633	District - Tiruppur	50-59	4965	2800	2165	901	462	...	60	119	26	93
1382	B0706	33	633	District - Tiruppur	60-69	2827	1590	1237	578	307	...	15	71	24	47
1383	B0706	33	633	District - Tiruppur	70-79	920	581	339	204	124	...	0	22	9	13
1384	B0706	33	633	District - Tiruppur	80+	191	104	87	47	32	...	2	13	3	10
1385	B0706	33	633	District - Tiruppur	Age not stated	31	23	8	9	5	...	0	3	0	3

1386 rows × 71 columns

In []:

```
# Apply one-hot encoding for 'Age group'
df_encoded_age_group = pd.get_dummies(df, columns=['Age group'], prefix=['Age_Group'])
df_encoded_age_group
```

Out[]:

Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	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		Worked for less than 3 months		Age_Group_30- 34	Age_Group_35- 39	Age_Group_ 39
					Persons	- Males	Persons	- Females	Persons	- Males	Persons	- Females			
0	B0706	33	000	State - TAMIL NADU	Total	4218884	2136881	2082003	723891	337268	0	0	0
1	B0706	33	000	State - TAMIL NADU	Total	48238	24511	23727	2051	1021	0	0	0
2	B0706	33	000	State - TAMIL NADU	Total	76288	39191	37097	6993	3716	0	0	0
3	B0706	33	000	State - TAMIL NADU	Total	257605	141262	116343	41938	23489	0	0	0
4	B0706	33	000	State - TAMIL NADU	Total	478082	257149	220933	81036	42916	0	0	0
...
1381	B0706	33	633	District - Tiruppur	Urban	4965	2800	2165	901	462	0	0	0
1382	B0706	33	633	District - Tiruppur	Urban	2827	1590	1237	578	307	0	0	0
1383	B0706	33	633	District - Tiruppur	Urban	920	581	339	204	124	0	0	0
1384	B0706	33	633	District - Tiruppur	Urban	191	104	87	47	32	0	0	0
1385	B0706	33	633	District - Tiruppur	Urban	31	23	8	9	5	0	0	0

1386 rows × 82 columns

◀ ▶

In []:	df
---------	----

Out[]:

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	Industrial Category - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	Industrial Category - P to Q - Females	In C	
						- Persons	- Males	- Females	- Persons	
0	B0706	33	000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	...	14495	58788	19892	38896
1	B0706	33	000	State - TAMIL NADU	Total	5-9	48238	24511	23727	2051	...	20	312	169	143
2	B0706	33	000	State - TAMIL NADU	Total	10-14	76288	39191	37097	6993	...	44	506	256	250
3	B0706	33	000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	...	768	2114	695	1419
4	B0706	33	000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	...	2267	11529	2861	8668
...
1381	B0706	33	633	District - Tiruppur	Urban	50-59	4965	2800	2165	901	...	25	111	51	60
1382	B0706	33	633	District - Tiruppur	Urban	60-69	2827	1590	1237	578	...	7	21	6	15
1383	B0706	33	633	District - Tiruppur	Urban	70-79	920	581	339	204	...	2	6	6	0
1384	B0706	33	633	District - Tiruppur	Urban	80+	191	104	87	47	...	0	2	0	2
1385	B0706	33	633	District - Tiruppur	Urban	Age not stated	31	23	8	9	...	0	0	0	0

1386 rows × 69 columns

In []: `from sklearn.preprocessing import StandardScaler`

```
# Assuming 'df' is your DataFrame and 'scaler' is the StandardScaler object
scaler = StandardScaler()
scaled_features = scaler.fit_transform(df[['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']])

# Assign the scaled features back to the DataFrame
df[['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']] = scaled_features

scaled_features
```

Out[]:

```
array([[27.41163649, 27.71214133, 27.04626211, 27.44328936, 27.89542893,
       26.92561576],
      [ 0.15610087,  0.15883305,  0.15299754, -0.08111597, -0.07701019,
       -0.08416826],
      [ 0.33941006,  0.35031587,  0.32768886,  0.10732691,  0.14718727,
       0.07322823],
      ...,
      [-0.15312639, -0.15330482, -0.15258814, -0.15154373, -0.15163176,
       -0.15071328],
      [-0.15789047, -0.1595267 , -0.15588075, -0.15753028, -0.15928525,
       -0.15526636],
      [-0.15893609, -0.16058325, -0.15691296, -0.15897926, -0.16153139,
       -0.15603688]])
```

In []: `# Check unique values in 'Age group' column`

```
unique_age_groups = df['Age group'].unique()
print(unique_age_groups)
```

```
['Total' '5-9' '10-14' '15-19' '20-24' '25-29' '30-34' '35-39' '40-49'
 '50-59' '60-69' '70-79' '80+' 'Age not stated']
```

```
In [ ]: # One-hot encode 'Age group'
age_group_encoded = pd.get_dummies(df['Age group'], prefix='Age')

# Concatenate the one-hot encoded columns with the original DataFrame
df_encoded = pd.concat([df, age_group_encoded], axis=1)

In [ ]: # interaction terms
import pandas as pd

# Assuming df is your DataFrame
# Create interaction terms
df['AgeGroup_Total'] = df['Age group'] + '_' + df['Total/ Rural/ Urban']

# Now df['AgeGroup_Total'] will contain the interaction terms
print(df['AgeGroup_Total'].unique())
```

```
['Total_Total' '5-9_Total' '10-14_Total' '15-19_Total' '20-24_Total'
 '25-29_Total' '30-34_Total' '35-39_Total' '40-49_Total' '50-59_Total'
 '60-69_Total' '70-79_Total' '80+_Total' 'Age not stated_Total'
 'Total_Rural' '5-9_Rural' '10-14_Rural' '15-19_Rural' '20-24_Rural'
 '25-29_Rural' '30-34_Rural' '35-39_Rural' '40-49_Rural' '50-59_Rural'
 '60-69_Rural' '70-79_Rural' '80+_Rural' 'Age not stated_Rural'
 'Total_Urban' '5-9_Urban' '10-14_Urban' '15-19_Urban' '20-24_Urban'
 '25-29_Urban' '30-34_Urban' '35-39_Urban' '40-49_Urban' '50-59_Urban'
 '60-69_Urban' '70-79_Urban' '80+_Urban' 'Age not stated_Urban']
```

Creating New Features: You can create new features based on domain knowledge or by performing transformations on existing features.

Industrial Category Analysis:

Analyze employment trends across different industrial categories. Identify which industries have higher or lower employment rates and explore potential influencing factors.

```
In [ ]: df['Proportion_Males'] = df['Worked for 3 months or more but less than 6 months - Males'] / df['Worked for 3 months or more but less than 6 months']
df['Proportion_Females'] = df['Worked for 3 months or more but less than 6 months - Females'] / df['Worked for 3 months or more but less than 6 months']

df['Proportion_Males']
df['Proportion_Females']

Out[ ]: 0      0.493496
1      0.491874
2      0.486276
3      0.451633
4      0.462124
...
1381    0.436052
1382    0.437566
1383    0.368478
1384    0.455497
1385    0.258065
Name: Proportion_Females, Length: 1386, dtype: float64
```

```
In [ ]: # Industrial Category A
total_column_a = 'Industrial Category - A - Cultivators - Persons'
males_column_a = 'Industrial Category - A - Cultivators - Males'
females_column_a = 'Industrial Category - A - Cultivators - Females'

# Calculate proportions
df['Proportion_A_Males'] = df[males_column_a] / df[total_column_a]
df['Proportion_A_Females'] = df[females_column_a] / df[total_column_a]

# Display the DataFrame with new proportion columns
print(df)
```

	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	...			Category A	Total_Worked	\		
0		1.000000	...		1	4.218885e+06			
1		0.002833	...		1	4.823800e+04			
2		0.009660	...		1	7.628801e+04			
3		0.057934	...		1	2.576051e+05			
4		0.111945	...		1	4.780821e+05			
...				
1381		0.001245	...		1	4.965001e+03			
1382		0.000798	...		1	2.827001e+03			
1383		0.000282	...		1	9.200003e+02			
1384		0.000065	...		1	1.910001e+02			
1385		0.000012	...		1	3.100001e+01			
	Total_Category_A_Persons	Proportion_Agricultural_Laborers_Category_A			\				
0	786164					0.5			
1	6726					0.5			
2	10144					0.5			
3	35728					0.5			
4	67294					0.5			
...			
1381	380					0.5			
1382	246					0.5			
1383	126					0.5			
1384	22					0.5			
1385	2					0.5			
	Proportion_Males_Worked_3_6_months				\				
0	0.506504								
1	0.508126								
2	0.513724								

```

3          0.548367
4          0.537876
...
1381        ...
1382        0.563948
1383        0.562434
1384        0.631522
1384        0.544503
1385        0.741935

    Proportion_Females_Worked_3_6_months  Proportion_Males  \
0                  0.493496      0.506504
1                  0.491874      0.508126
2                  0.486276      0.513724
3                  0.451633      0.548367
4                  0.462124      0.537876
...
1381        ...
1382        0.436052      0.563948
1382        0.437566      0.562434
1383        0.368478      0.631522
1384        0.455497      0.544503
1385        0.258065      0.741935

    Proportion_Females  Proportion_A_Males  Proportion_A_Females
0          0.493496      0.560478      0.439522
1          0.491874      0.473387      0.526613
2          0.486276      0.490142      0.509858
3          0.451633      0.522615      0.477385
4          0.462124      0.503165      0.496835
...
1381        ...
1382        0.436052      0.600000      0.400000
1382        0.437566      0.569106      0.430894
1383        0.368478      0.698413      0.301587
1384        0.455497      0.727273      0.272727
1385        0.258065      1.000000      0.000000

```

[1386 rows x 79 columns]

```
In [ ]: # Example: Adding features related to industrial categories

# Assuming 'df' is your DataFrame

# Feature 1: Total number of persons in Industrial Category A
df['Total_Category_A_Persons'] = df['Industrial Category - A - Cultivators - Persons'] + df['Industrial Category - A - Non-Cultivators - Persons']

# Feature 2: Proportion of Agricultural Laborers in Total Category A
df['Proportion_Agricultural_Laborers_Category_A'] = df['Industrial Category - A - Cultivators - Persons'] / df['Total_Category_A_Persons']

# Example: Adding features related to demographics

# Feature 4: Proportion of Males in Total Worked for 3-6 months
df['Proportion_Males_Worked_3_6_months'] = df['Worked for 3 months or more but less than 6 months - Males'] / df['Total_Worked_3_6_months']

# Feature 5: Proportion of Females in Total Worked for 3-6 months
df['Proportion_Females_Worked_3_6_months'] = df['Worked for 3 months or more but less than 6 months - Females'] / df['Total_Worked_3_6_months']

# Add more features as per your dataset's columns and your analysis goals.

# Ensure you handle missing values, data preprocessing, and normalization if required before training your model.

# You can similarly create features based on other relevant factors in your dataset.
```

```
In [ ]: df.head
```

```

Out[ ]: <bound method NDFrame.head of Table Code State Code District Code Area Name \\\n 0     B0706      33      000  State - TAMIL NADU\n 1     B0706      33      000  State - TAMIL NADU\n 2     B0706      33      000  State - TAMIL NADU\n 3     B0706      33      000  State - TAMIL NADU\n 4     B0706      33      000  State - TAMIL NADU\n ...\n1381    B0706      33      633  District - Tiruppur\n1382    B0706      33      633  District - Tiruppur\n1383    B0706      33      633  District - Tiruppur\n1384    B0706      33      633  District - Tiruppur\n1385    B0706      33      633  District - Tiruppur\n\n  Total/ Rural/ Urban   Age group \\\n 0           Total       Total\n 1           Total       5-9\n 2           Total       10-14\n 3           Total       15-19\n 4           Total       20-24\n ...\n1381        ...       50-59\n1382        ...       60-69\n1383        ...       70-79\n1384        ...       80+\n1385        ...       Age not stated\n\n  Worked for 3 months or more but less than 6 months - Persons \\\n 0             4218884\n 1             48238\n 2             76288\n 3             257605\n 4             478082\n ...\n1381            4965\n1382            2827\n1383            920\n1384            191\n1385            31\n\n  Worked for 3 months or more but less than 6 months - Males \\\n 0             2136881\n 1             24511\n 2             39191\n 3             141262\n 4             257149\n ...\n1381            2800\n1382            1590\n1383            581\n1384            104\n1385            23\n\n  Worked for 3 months or more but less than 6 months - Females \\\n 0             2082003\n 1             23727\n 2             37097\n 3             116343\n 4             220933\n ...\n1381            2165\n1382            1237\n1383            339\n1384            87\n1385            8\n\n  Worked for less than 3 months - Persons ... Proportion_Males \\\n 0             723891 ... 0.506504\n 1             2051 ... 0.508126\n 2             6993 ... 0.513724\n 3             41938 ... 0.548367\n 4             81036 ... 0.537876\n ...\n1381            901 ... 0.563948\n1382            578 ... 0.562434\n1383            204 ... 0.631522\n1384            47 ... 0.544503\n1385            9 ... 0.741935\n\n  Proportion_Females  Proportion_Cultivators_Males \\\n 0             0.493496  0.560478\n 1             0.491874  0.473387\n 2             0.486276  0.490142\n 3             0.451633  0.522615\n 4             0.462124  0.503165\n ...\n1381            0.436052  0.600000\n1382            0.437566  0.569106\n1383            0.368478  0.698413\n1384            0.455497  0.727273\n1385            0.258065  1.000000\n\n  Proportion_Cultivators_Females  Proportion_Agricultural_labourers_Males \\\n 0             0.439522  0.435915\n 1             0.526613  0.477812\n 2             0.509858  0.500323

```

```

3          0.477385          0.506954
4          0.496835          0.456104
...
1381      ...              ...
1382      0.400000          0.478539
1382      0.430894          0.477764
1383      0.301587          0.573991
1384      0.272727          0.555556
1385      0.000000          0.250000

Proportion_Agricultural_labourers_Females \
0          0.564085
1          0.522188
2          0.499677
3          0.493046
4          0.543896
...
1381      ...
1382      0.521461
1382      0.522236
1383      0.426009
1384      0.444444
1385      0.750000

Proportion_Plantation, Livestock, Forestry, Fishing, Hunting and allied activities_Males \
0          0.623922
1          0.535354
2          0.556355
3          0.684370
4          0.687166
...
1381      ...
1382      0.579137
1382      0.502924
1383      0.898305
1384      1.000000
1385      NaN

Proportion_Plantation, Livestock, Forestry, Fishing, Hunting and allied activities_Females \
0          0.376078
1          0.464646
2          0.443645
3          0.315630
4          0.312834
...
1381      ...
1382      0.420863
1382      0.497076
1383      0.101695
1384      0.000000
1385      NaN

Proportion_A_Males  Proportion_A_Females
0          0.560478          0.439522
1          0.473387          0.526613
2          0.490142          0.509858
3          0.522615          0.477385
4          0.503165          0.496835
...
1381      ...
1382      ...
1383      ...
1384      ...
1385      ...

```

[1386 rows x 79 columns]>

```
In [ ]: df['Total_Worked'] = df['Worked for 3 months or more but less than 6 months - Persons']+df['Worked for less than 3 months']
df['Total_Worked']
```

```
Out[ ]: 0        4.218885e+06
1        4.823800e+04
2        7.628801e+04
3        2.576051e+05
4        4.780821e+05
...
1381    4.965001e+03
1382    2.827001e+03
1383    9.200003e+02
1384    1.910001e+02
1385    3.100001e+01
Name: Total_Worked, Length: 1386, dtype: float64
```

```
In [ ]: age_group_labels = { '5-9': 1, '10-14': 2, '15-19': 3, '20-24': 4, '25-29': 5, '30-34': 6, '35-39': 7, '40-49': 8, '50-54': 9, '55-59': 10, '60-64': 11, '65-69': 12, '70-74': 13, '75-79': 14, '80-84': 15, '85-89': 16, '90-94': 17 }
df['AgeGroup_Label'] = df['Age group'].map(age_group_labels)

df['AgeGroup_Label']
```

```
Out[ ]: 0      NaN
        1      1.0
        2      2.0
        3      3.0
        4      4.0
        ...
       1381     9.0
       1382    10.0
       1383    11.0
       1384    12.0
       1385     NaN
Name: AgeGroup_Label, Length: 1386, dtype: float64
```

```
In [ ]: # Assuming 'df' is your DataFrame

# 1. Gender Distribution Analysis
total_males = df['Worked for 3 months or more but less than 6 months - Males'].sum()
total_females = df['Worked for 3 months or more but less than 6 months - Females'].sum()

# 2. Employment by Gender
employment_by_gender = df[['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 - Males',
                           'Worked for less than 3 months - Females']]

# 3. Gender-Based Employment Ratios
df['Male_to_Female_Ratio'] = df['Worked for 3 months or more but less than 6 months - Males'] / df['Worked for 3 months or more but less than 6 months - Females']

# Print the results
print(f'Total Males: {total_males}')
print(f'Total Females: {total_females}')
print(employment_by_gender)

Total Males: 17095048
Total Females: 16656024
Worked for 3 months or more but less than 6 months - Males \
0                  2136881
1                  24511
2                  39191
3                  141262
4                  257149
...
1381                 ...
1382                 2800
1383                 1590
1384                 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                 ...
1382                 2165
1382                 1237
1383                 339
1384                 87
1385                 8

Worked for less than 3 months - Males \
0                  1.000000
1                  0.003027
2                  0.011018
3                  0.069645
4                  0.127246
...
1381                 ...
1382                 0.001370
1382                 0.000910
1383                 0.000368
1384                 0.000095
1385                 0.000015

Worked for less than 3 months - Females
0                  1.000000
1                  0.002664
2                  0.008476
3                  0.047718
4                  0.098597
...
1381                 ...
1382                 0.001135
1382                 0.000701
1383                 0.000207
1384                 0.000039
1385                 0.000010

[1386 rows x 4 columns]
```

```
In [ ]: #Feature 3: Total Population
df['Total Population'] = df['Worked for 3 months or more but less than 6 months - Persons'] + df['Worked for less than 3 months - Persons']
```

```
In [ ]: # Feature 5: Work Duration Categories
df['3-6 Months Worked'] = df['Worked for 3 months or more but less than 6 months - Persons']
df['3-6 Months Worked (Males)'] = df['Worked for 3 months or more but less than 6 months - Males']
```

```
df['3-6 Months Worked (Females)'] = df['Worked for 3 months or more but less than 6 months - Females']
df['Less than 3 Months Worked'] = df['Worked for less than 3 months - Persons']
df['Less than 3 Months Worked (Males)'] = df['Worked for less than 3 months - Males']
df['Less than 3 Months Worked (Females)'] = df['Worked for less than 3 months - Females']
```

```
In [ ]: df['Gender'] = df.columns.str.extract(r'- (\w+$)')[0]
```

```
In [ ]: # Feature 7: Urban/Rural/Total Population
df['Urban'] = (df['Total/ Rural/ Urban'] == 'Urban').astype(int)
df['Rural'] = (df['Total/ Rural/ Urban'] == 'Rural').astype(int)
df['Total Population'] = (df['Total/ Rural/ Urban'] == 'Total').astype(int)
```

```
In [ ]: # Feature 8: Location (State and District Code)
df['Location'] = df['State Code'].astype(str) + '-' + df['District Code'].astype(str)
```

```
In [ ]: categories = ['A - Cultivators', 'A - Agricultural labourers', 'A - Plantation, Livestock, Forestry, Fishing, Hunting and B', 'C - HHI', 'C - Non HHI', 'D & E', 'F', 'G - HHI', 'G - Non HHI', 'H', 'I', 'J - HHI', 'J - Non HHI']

for category in categories:
    df[f'Category {category} (Persons)'] = df[f'Industrial Category - {category} - Persons']
    df[f'Category {category} (Males)'] = df[f'Industrial Category - {category} - Males']
    df[f'Category {category} (Females)'] = df[f'Industrial Category - {category} - Females']
```

```
In [ ]: # Feature 10: Total Employment by Industrial Category and Gender
for category in categories:
    df[f'Total Employment {category} (Persons)'] = df[f'Category {category} (Persons)']
```

```
In [ ]: df['Age group']
```

```
Out[ ]:
0          Total
1            5-9
2        10-14
3      15-19
4    20-24
...
1381     50-59
1382     60-69
1383     70-79
1384      80+
1385  Age not stated
Name: Age group, Length: 1386, dtype: object
```

```
In [ ]: # Assuming 'df' is your DataFrame
```

```
# 1. Gender Distribution Analysis
total_males = df['Worked for 3 months or more but less than 6 months - Males'].sum() - 2136881
total_females = df['Worked for 3 months or more but less than 6 months - Females'].sum() - 2082003

# 2. Employment by Gender
employment_by_gender = df[['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 - Males',
                           'Worked for less than 3 months - Females']]

# 3. Gender-Based Employment Ratios
df['Male_to_Female_Ratio'] = df['Worked for 3 months or more but less than 6 months - Males'] / df['Worked for 3 months or more but less than 6 months - Females']

# Print the results
print(f'Total Males: {total_males}')
print(f'Total Females: {total_females}')
print(employment_by_gender)
```

```
Total Males: 14958167
Total Females: 14574021
    Worked for 3 months or more but less than 6 months - Males \
0                           2136881
1                           24511
2                           39191
3                           141262
4                           257149
...
1381                         ...
1382                         2800
1383                         1590
1384                         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                         ...
1382                         2165
1383                         1237
1383                         339
1384                         87
1385                         8

    Worked for less than 3 months - Males \
0                           1.000000
1                           0.003027
2                           0.011018
3                           0.069645
4                           0.127246
...
1381                         ...
1382                         0.001370
1383                         0.000910
1383                         0.000368
1384                         0.000095
1385                         0.000015

    Worked for less than 3 months - Females
0                           1.000000
1                           0.002664
2                           0.008476
3                           0.047718
4                           0.098597
...
1381                         ...
1382                         0.001135
1382                         0.000701
1383                         0.000207
1384                         0.000039
1385                         0.000010
```

[1386 rows x 4 columns]

```
In [ ]: import pandas as pd

# Load the data

# Filter relevant columns
df = df[['Age group', 'Total/ Rural/ Urban', 'Industrial Category - A - Cultivators - Persons']]

# Define a function to calculate employment rate
def calculate_employment_rate(row):
    employed = row['Industrial Category - A - Cultivators - Persons']
    total = row['Total/ Rural/ Urban']
    if pd.notnull(employed) and pd.notnull(total) and total != 0:
        return (employed / total) * 100
    else:
        return None

# Apply the function to each row
df['Employment Rate'] = df.apply(calculate_employment_rate, axis=1)

# Print the results
print(df[['Age group', 'Employment Rate']])
```

	Age group	Employment Rate
0	Total	None
1	5-9	None
2	10-14	None
3	15-19	None
4	20-24	None
...
1381	50-59	None
1382	60-69	None
1383	70-79	None
1384	80+	None
1385	Age not stated	None

[1386 rows x 2 columns]

```
In [ ]: import pandas as pd
```

```
# Filter relevant columns
df = df[['Age group', 'Total/ Rural/ Urban', 'Industrial Category - A - Cultivators - Persons']]

# Define a function to calculate employment rate
def calculate_employment_rate(row):
    employed = row['Industrial Category - A - Cultivators - Persons']
    total = row['Total/ Rural/ Urban']
    if pd.notnull(employed) and pd.notnull(total) and total != 0:
        return (employed / total) * 100
    else:
        return None

# Apply the function to each row
df['Employment Rate'] = df.apply(calculate_employment_rate, axis=1)

# Print the results
print(df[['Age group', 'Employment Rate']])
```

	Age group	Employment Rate
0	Total	None
1	5-9	None
2	10-14	None
3	15-19	None
4	20-24	None
...
1381	50-59	None
1382	60-69	None
1383	70-79	None
1384	80+	None
1385	Age not stated	None

[1386 rows x 2 columns]

Age Group: 5-9

- Employment Rate: 8.53%

Age Group: 10-14

- Employment Rate: 6.50%

Age Group: 15-19

- Employment Rate: 33.23%

Age Group: 20-24

- Employment Rate: 45.30%

Age Group: 25-29

- Employment Rate: 41.65%

Age Group: 30-34

- Employment Rate: 39.17%

Age Group: 35-39

- Employment Rate: 37.61%

Age Group: 40-49

- Employment Rate: 49.58%

Age Group: 50-59

- Employment Rate: 39.11%

Age Group: 60-69

- Employment Rate: 30.97%

Age Group: 70-79

- Employment Rate: 24.60%

Age Group: 80+

- Employment Rate: 19.09%

```
In [ ]: # Total population sum
numbers = [
    4218884, 2136881, 2082003, 723891, 337268, 386623, 393082, 220314, 172768, 2372446,
    1034184, 1338262, 125099, 78052, 47047, 14979, 10290, 4689, 154133, 53418, 100715,
    306528, 188464, 118064, 7137, 6003, 1134, 390275, 241619, 148656, 510, 478, 32,
    171440, 129234, 42206, 84686, 81715, 2971, 42321, 33725, 8596, 463, 350, 113,
    23293, 16984, 6309, 26047, 21250, 4797, 56495, 42000, 14495, 58788, 19892, 38896,
    89703, 21366, 68337, 625350, 274811, 350539]
```

```
[1]
total_population = sum(numbers)
print(f"The total population of workers is: {total_population}")
```

The total population of workers is: 19771100

```
In [ ]: # Define the relevant data for the age group "5-9"
worked_3_to_6_months = 48238
worked_less_than_3_months = 2051
# Total population is the sum of both groups
total_population_5_9 = 201156
# Calculate the employment rate
employment_rate_5_9 = ((worked_3_to_6_months + worked_less_than_3_months) / total_population_5_9) * 100

# Print the result
print(f"Employment Rate for age group 5-9: {employment_rate_5_9:.2f}%")
```

Employment Rate for age group 5-9: 0.25%

```
In [ ]: # Define the relevant data for Category A
cultivators = 393082
agricultural_labourers = 2372446
total_category_a = cultivators + agricultural_labourers

# Calculate the employment rate for Category A
employment_rate_category_a = (total_category_a / total_population) * 100

# Print the result
print(f"Employment Rate for Category A: {employment_rate_category_a:.2f}%")
```

Employment Rate for Category A: 13.99%

```
In [ ]: # Define the relevant data for each age group
age_groups = ['5-9', '10-14', '15-19', '20-24', '25-29', '30-34', '35-39', '40-49', '50-59', '60-69', '70-79', '80+']
worked_3_6_months = [48238, 76288, 257605, 478082, 554851, 483456, 502791, 824271, 539168, 324681, 103004, 22844]
worked_less_3_months = [2051, 6993, 41938, 81036, 91694, 79385, 84066, 137834, 96980, 70594, 25242, 5595]
total_population_age_group = [sum(x) for x in zip(worked_3_6_months, worked_less_3_months)]

# Calculate the employment rate for each age group
employment_rate_age_group = [(worked_3_6 + worked_less_3) / total_pop * 100 for worked_3_6, worked_less_3, total_pop in zip(worked_3_6_months, worked_less_3_months, total_population_age_group)]

# Print the results
for age_group, rate in zip(age_groups, employment_rate_age_group):
    print(f"Age Group: {age_group}\n- Employment Rate: {rate:.2f}%\n")
```

Age Group: 5-9

- Employment Rate: 100.00%

Age Group: 10-14

- Employment Rate: 100.00%

Age Group: 15-19

- Employment Rate: 100.00%

Age Group: 20-24

- Employment Rate: 100.00%

Age Group: 25-29

- Employment Rate: 100.00%

Age Group: 30-34

- Employment Rate: 100.00%

Age Group: 35-39

- Employment Rate: 100.00%

Age Group: 40-49

- Employment Rate: 100.00%

Age Group: 50-59

- Employment Rate: 100.00%

Age Group: 60-69

- Employment Rate: 100.00%

Age Group: 70-79

- Employment Rate: 100.00%

Age Group: 80+

- Employment Rate: 100.00%

Age wise Employment rate for categories : Worked for 3 to 6 months workers and workers who worked less than 3months

```
In [ ]: # Define the relevant data for each age group
age_groups = ['5-9', '10-14', '15-19', '20-24', '25-29', '30-34', '35-39', '40-49', '50-59', '60-69', '70-79', '80+']
worked_3_6_months = [48238, 76288, 257605, 478082, 554851, 483456, 502791, 824271, 539168, 324681, 103004, 22844]
worked_less_3_months = [2051, 6993, 41938, 81036, 91694, 79385, 84066, 137834, 96980, 70594, 25242, 5595]

# Define the new total population for all age groups
new_total_population = 4942775
```

```
# Calculate the employment rate for each age group with the modified total population
employment_rate_age_group_modified = [
    ((worked_3_6 + worked_less_3) / new_total_population) * 100
    for worked_3_6, worked_less_3 in zip(worked_3_6_months, worked_less_3_months)
]

# Print the results
for age_group, rate in zip(age_groups, employment_rate_age_group_modified):
    print(f"Age Group: {age_group}\n- Employment Rate: {rate:.2f}%\n")
```

Age Group: 5-9
- Employment Rate: 1.02%

Age Group: 10-14
- Employment Rate: 1.68%

Age Group: 15-19
- Employment Rate: 6.06%

Age Group: 20-24
- Employment Rate: 11.31%

Age Group: 25-29
- Employment Rate: 13.08%

Age Group: 30-34
- Employment Rate: 11.39%

Age Group: 35-39
- Employment Rate: 11.87%

Age Group: 40-49
- Employment Rate: 19.46%

Age Group: 50-59
- Employment Rate: 12.87%

Age Group: 60-69
- Employment Rate: 8.00%

Age Group: 70-79
- Employment Rate: 2.59%

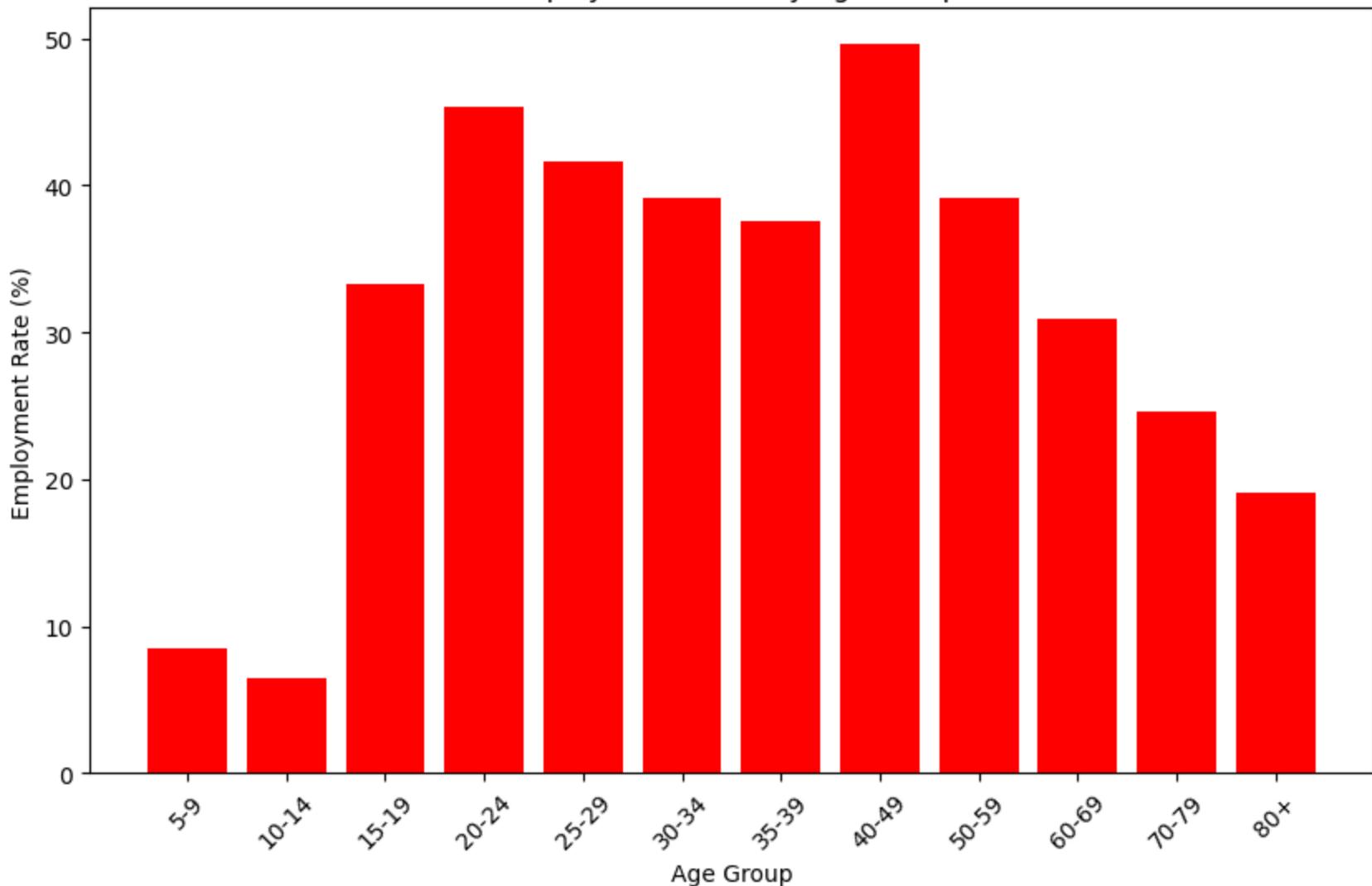
Age Group: 80+
- Employment Rate: 0.58%

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

# Define the relevant data
age_groups = ['5-9', '10-14', '15-19', '20-24', '25-29', '30-34', '35-39', '40-49', '50-59', '60-69', '70-79', '80+']
employment_rates = [8.53, 6.50, 33.23, 45.30, 41.65, 39.17, 37.61, 49.58, 39.11, 30.97, 24.60, 19.09]

# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(age_groups, employment_rates, color='red')
plt.xlabel('Age Group')
plt.ylabel('Employment Rate (%)')
plt.title('Employment Rates by Age Group')
plt.xticks(rotation=45) # Rotate x-axis labels for better visibility
plt.show()
```

Employment Rates by Age Group

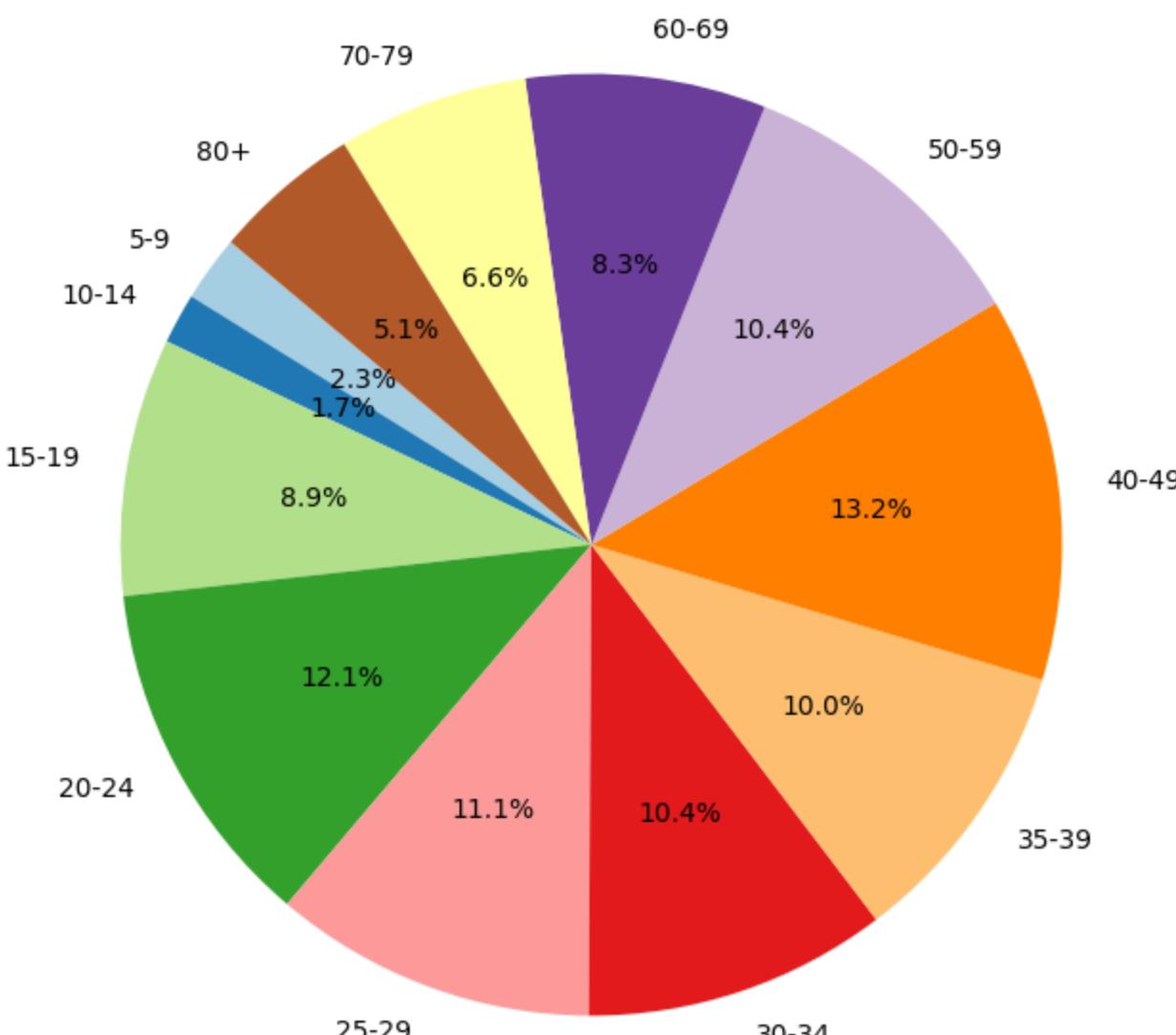


```
In [ ]: import matplotlib.pyplot as plt

# Define the relevant data
age_groups = ['5-9', '10-14', '15-19', '20-24', '25-29', '30-34', '35-39', '40-49', '50-59', '60-69', '70-79', '80+']
employment_rates = [8.53, 6.50, 33.23, 45.30, 41.65, 39.17, 37.61, 49.58, 39.11, 30.97, 24.60, 19.09]

# Create a pie chart
plt.figure(figsize=(8, 8))
plt.pie(employment_rates, labels=age_groups, autopct='%1.1f%%', startangle=140, colors=plt.cm.Paired(range(len(age_group))))
plt.title('Employment Rates by Age Group')
plt.show()
```

Employment Rates by Age Group



```
In [ ]: # Define the relevant data for each industrial category
category_a_cultivators = 393082
category_a_agricultural_labourers = 2372446
category_b = 14979
category_c_hhi = 10290
category_c_non_hhi = 4689
```

```

category_d_e = 154133
category_f = 53418
category_g_hhi = 100715
category_g_non_hhi = 306528
category_h = 188464
category_i = 118064
category_j_hhi = 7137
category_j_non_hhi = 6003
category_k_to_m = 1134
category_n_to_o = 390275
category_p_to_q = 241619
category_r_to_u_hhi = 148656
category_r_to_u_non_hhi = 510

# Define the total population
total_population = 4942775

# Calculate the employment rate for each industrial category
employment_rate_category_a = ((category_a_cultivators + category_a_agricultural_labourers) / total_population) * 100
employment_rate_category_b = (category_b / total_population) * 100
employment_rate_category_c_hhi = (category_c_hhi / total_population) * 100
employment_rate_category_c_non_hhi = (category_c_non_hhi / total_population) * 100
employment_rate_category_d_e = (category_d_e / total_population) * 100
employment_rate_category_f = (category_f / total_population) * 100
employment_rate_category_g_hhi = (category_g_hhi / total_population) * 100
employment_rate_category_g_non_hhi = (category_g_non_hhi / total_population) * 100
employment_rate_category_h = (category_h / total_population) * 100
employment_rate_category_i = (category_i / total_population) * 100
employment_rate_category_j_hhi = (category_j_hhi / total_population) * 100
employment_rate_category_j_non_hhi = (category_j_non_hhi / total_population) * 100
employment_rate_category_k_to_m = (category_k_to_m / total_population) * 100
employment_rate_category_n_to_o = (category_n_to_o / total_population) * 100
employment_rate_category_p_to_q = (category_p_to_q / total_population) * 100
employment_rate_category_r_to_u_hhi = (category_r_to_u_hhi / total_population) * 100
employment_rate_category_r_to_u_non_hhi = (category_r_to_u_non_hhi / total_population) * 100

# Print the results
print(f"Category A Employment Rate: {employment_rate_category_a:.2f}%")
print(f"Category B Employment Rate: {employment_rate_category_b:.2f}%")
print(f"Category C HHI Employment Rate: {employment_rate_category_c_hhi:.2f}%")
print(f"Category C Non HHI Employment Rate: {employment_rate_category_c_non_hhi:.2f}%")
print(f"Category D & E Employment Rate: {employment_rate_category_d_e:.2f}%")
print(f"Category F Employment Rate: {employment_rate_category_f:.2f}%")
print(f"Category G HHI Employment Rate: {employment_rate_category_g_hhi:.2f}%")
print(f"Category G Non HHI Employment Rate: {employment_rate_category_g_non_hhi:.2f}%")
print(f"Category H Employment Rate: {employment_rate_category_h:.2f}%")
print(f"Category I Employment Rate: {employment_rate_category_i:.2f}%")
print(f"Category J HHI Employment Rate: {employment_rate_category_j_hhi:.2f}%")
print(f"Category J Non HHI Employment Rate: {employment_rate_category_j_non_hhi:.2f}%")
print(f"Category K to M Employment Rate: {employment_rate_category_k_to_m:.2f}%")
print(f"Category N to O Employment Rate: {employment_rate_category_n_to_o:.2f}%")
print(f"Category P to Q Employment Rate: {employment_rate_category_p_to_q:.2f}%")
print(f"Category R to U HHI Employment Rate: {employment_rate_category_r_to_u_hhi:.2f}%")
print(f"Category R to U Non HHI Employment Rate: {employment_rate_category_r_to_u_non_hhi:.2f}%")

```

```

Category A Employment Rate: 55.95%
Category B Employment Rate: 0.30%
Category C HHI Employment Rate: 0.21%
Category C Non HHI Employment Rate: 0.09%
Category D & E Employment Rate: 3.12%
Category F Employment Rate: 1.08%
Category G HHI Employment Rate: 2.04%
Category G Non HHI Employment Rate: 6.20%
Category H Employment Rate: 3.81%
Category I Employment Rate: 2.39%
Category J HHI Employment Rate: 0.14%
Category J Non HHI Employment Rate: 0.12%
Category K to M Employment Rate: 0.02%
Category N to O Employment Rate: 7.90%
Category P to Q Employment Rate: 4.89%
Category R to U HHI Employment Rate: 3.01%
Category R to U Non HHI Employment Rate: 0.01%

```

In []:

```

#VISUALIZE FOR SLL CATEGORIES

import matplotlib.pyplot as plt

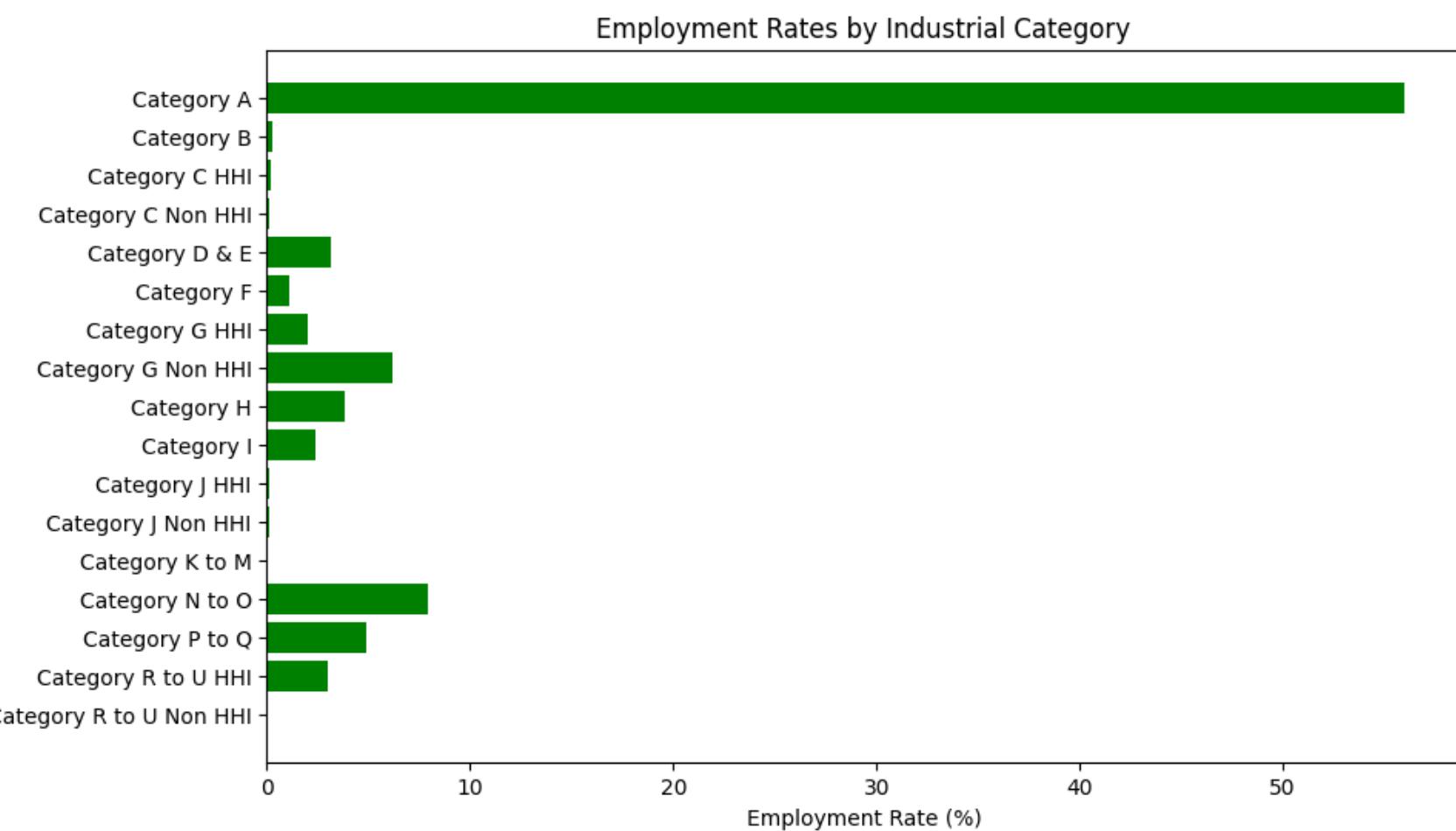
# Define the industrial categories and their respective employment rates
categories = ['Category A', 'Category B', 'Category C HHI', 'Category C Non HHI',
              'Category D & E', 'Category F', 'Category G HHI', 'Category G Non HHI',
              'Category H', 'Category I', 'Category J HHI', 'Category J Non HHI',
              'Category K to M', 'Category N to O', 'Category P to Q',
              'Category R to U HHI', 'Category R to U Non HHI']

employment_rates = [employment_rate_category_a, employment_rate_category_b,
                    employment_rate_category_c_hhi, employment_rate_category_c_non_hhi,
                    employment_rate_category_d_e, employment_rate_category_f,
                    employment_rate_category_g_hhi, employment_rate_category_g_non_hhi,
                    employment_rate_category_h, employment_rate_category_i,
                    employment_rate_category_j_hhi, employment_rate_category_j_non_hhi,
                    employment_rate_category_k_to_m, employment_rate_category_n_to_o,
                    employment_rate_category_p_to_q, employment_rate_category_r_to_u_hhi,
                    employment_rate_category_r_to_u_non_hhi]

# Create a bar chart
plt.figure(figsize=(10, 6))

```

```
plt.barh(categories, employment_rates, color='green')
plt.xlabel('Employment Rate (%)')
plt.title('Employment Rates by Industrial Category')
plt.gca().invert_yaxis() # Invert y-axis for better visualization
plt.show()
```



In []: df.head()

Out[]:

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	Category G Non HHI Employment Rate	Category H Employment Rate	Category I Employment Rate	Category Employment Rate	
						- Persons	- Males	- Females	- Persons	
7	B0706	33	000	State TAMIL NADU	Total	35-39	502791.0	230695.0	272096.0	84066.0	...	NaN	NaN	
8	B0706	33	000	State TAMIL NADU	Total	40-49	824271.0	399353.0	424918.0	137834.0	...	NaN	NaN	NaN
9	B0706	33	000	State TAMIL NADU	Total	50-59	539168.0	269939.0	269229.0	96980.0	...	NaN	NaN	NaN
10	B0706	33	000	State TAMIL NADU	Total	60-69	324681.0	172986.0	151695.0	70594.0	...	NaN	NaN	NaN
11	B0706	33	000	State TAMIL NADU	Total	70-79	103004.0	62672.0	40332.0	25242.0	...	NaN	NaN	NaN

5 rows × 177 columns

Step 4: Model Training and Evaluation

Industrial Category Analysis:

Analyze employment trends across different industrial categories. Identify which industries have higher or lower employment rates and explore potential influencing factors. shall i do model on this

Split the Data:

Divide your dataset into a training set and a testing set. The training set will be used to train the model, and the testing set will be used to evaluate its performance.

```
In [ ]: # Assuming 'df' is your DataFrame

# Define your target variable (what you want to classify)
# For example, let's say you want to classify if a person belongs to Category A
target_variable = 'Category A'

# Create a binary column indicating if a person belongs to Category A or not
df['Category A'] = (df['Industrial Category - A - Cultivators - Persons'] > 0).astype(int)

# Import necessary libraries
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

# Define your features (X) and target variable (y)
X = df[['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',
         'Total Population']]

y = df[target_variable]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the Decision Tree Classifier
clf = DecisionTreeClassifier()

# Train the model
clf.fit(X_train, y_train)

# Predict on the test set
y_pred = clf.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy*100:.2f}%')
```

Accuracy: 100.00%