PHASE – IV

Development Part - II

Date	**27-10-2023**
Team ID	**4570**
Project Name	**Assessment of Marginal Workers in Tamil Nadu**

[1]
```
import pandas as pd

# Create a pandas DataFrame from your dataset
data = {
    'Industrial Category - A - Cultivators - Persons': [64235, 190],
    'Industrial Category - A - Cultivators - Males': [34632, 107],
    'Industrial Category - A - Cultivators - Females': [29603, 83],
    'Industrial Category - A - Agricultural labourers - Persons': [907752,␣
↪2853],
    'Industrial Category - A - Agricultural labourers - Males': [404844, 1862],
    'Industrial Category - A - Agricultural labourers - Females': [502908, 991],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,␣
↪Hunting and allied activities - Persons': [29410, 3],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,␣
↪Hunting and allied activities - Males': [16268, 3],
    'Industrial Category - A - Plantation, Livestock, Forestry, Fishing,␣
↪Hunting and allied activities - Females': [13142, 0],
}

df = pd.DataFrame(data)
df
```

[1]:
```
   Industrial Category - A - Cultivators - Persons  \0                  64235
1                                                190

   Industrial Category - A - Cultivators - Males  \0              34632
1                                              107

   Industrial Category - A - Cultivators - Females  \0            29603
1                                                83

   Industrial Category - A - Agricultural labourers - Persons  \0    907752
1                                                2853
```

```
   Industrial Category - A - Agricultural labourers - Males  \
0                                                 404844
1                                                   1862

   Industrial Category - A - Agricultural labourers - Females  \
0                                                 502908
1                                                    991

   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Persons  \
0                                                  29410
1                                                      3

   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Males  \
0                                                  16268
1                                                      3

   Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting
and allied activities - Females
0                                                  13142
1                                                      0
```

[3]:
```
data = {
    'Age group': ['Total', '5-14'],
    'Worked for 3 months or more but less than 6 months - Persons': [1200828,␣
↪27791],
    'Worked for 3 months or more but less than 6 months - Males': [589003,␣
↪14125],
    'Worked for 3 months or more but less than 6 months - Females': [611825,␣
↪13666],
}

df = pd.DataFrame(data)
df
```

[3]:
```
  Age group  Worked for 3 months or more but less than 6 months - Persons  \
0     Total                                                 1200828
1      5-14                                                   27791

   Worked for 3 months or more but less than 6 months - Males  \
0                                                 589003
1                                                  14125

   Worked for 3 months or more but less than 6 months - Females
0                                                 611825
```

1 13666

[6]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
```

[7]:
```
dataset= pd.read_csv("merin.csv")
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 594 entries, 0 to 593
Data columns (total 69 columns):
 #    Column
Non-Null Count  Dtype
---  -------
-------------------------------
 0    Table Code
594 non-null     object
 1    State Code
594 non-null     object
 2    District Code
594 non-null     object
 3    Area Name
594 non-null     object
 4    Total/ Rural/ Urban
594 non-null     object
 5    Age group
594 non-null     object
 6    Worked for 3 months or more but less than 6 months -  Persons
594 non-null     int64
 7    Worked for 3 months or more but less than 6 months - Males
594 non-null     int64
 8    Worked for 3 months or more but less than 6 months - Females
594 non-null     int64
 9    Worked for less than 3 months - Persons
594 non-null     int64
 10   Worked for less than 3 months - Males
594 non-null     int64
 11   Worked for less than 3 months - Females
594 non-null     int64
```

12 Industrial Category - A - Cultivators - Persons
594 non-null int64
13 Industrial Category - A - Cultivators - Males
594 non-null int64
14 Industrial Category - A - Cultivators - Females
594 non-null int64
15 Industrial Category - A - Agricultural labourers - Persons
594 non-null int64
16 Industrial Category - A - Agricultural labourers - Males
594 non-null int64
17 Industrial Category - A - Agricultural labourers - Females
594 non-null int64
18 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons 594 non-null int64
19 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males 594 non-null int64
20 Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females 594 non-null int64
21 Industrial Category - B - Persons
594 non-null int64
22 Industrial Category - B - Males
594 non-null int64
23 Industrial Category - B - Females
594 non-null int64
24 Industrial Category - C - HHI - Persons
594 non-null int64
25 Industrial Category - C - HHI - Males
594 non-null int64
26 Industrial Category - C - HHI - Females
594 non-null int64
27 Industrial Category - C - Non HHI - Persons
594 non-null int64
28 Industrial Category - C - Non HHI - Males
594 non-null int64
29 Industrial Category - C - Non HHI - Females
594 non-null int64
30 Industrial Category - D & E - Persons
594 non-null int64
31 Industrial Category - D & E - Males
594 non-null int64
32 Industrial Category - D & E - Females
594 non-null int64
33 Industrial Category - F - Persons
594 non-null int64
34 Industrial Category - F - Males
594 non-null int64
35 Industrial Category - F - Females
594 non-null int64

 36 Industrial Category - G - HHI - Persons
594 non-null   int64
 37 Industrial Category - G - HHI - Males
594 non-null   int64
 38 Industrial Category - G - HHI - Females
594 non-null   int64
 39 Industrial Category - G - Non HHI - Persons
594 non-null   int64
 40 Industrial Category - G - Non HHI - Males
594 non-null   int64
 41 Industrial Category - G - Non HHI - Females
594 non-null   int64
 42 Industrial Category - H - Persons
594 non-null   int64
 43 Industrial Category - H - Males
594 non-null   int64
 44 Industrial Category - H - Females
594 non-null   int64
 45 Industrial Category - I - Persons
594 non-null   int64
 46 Industrial Category - I - Males
594 non-null   int64
 47 Industrial Category - I - Females
594 non-null   int64
 48 Industrial Category - J - HHI - Persons
594 non-null   int64
 49 Industrial Category - J - HHI - Males
594 non-null   int64
 50 Industrial Category - J - HHI - Females
594 non-null   int64
 51 Industrial Category - J - Non HHI - Persons
594 non-null   int64
 52 Industrial Category - J - Non HHI - Males
594 non-null   int64
 53 Industrial Category - J - Non HHI - Females
594 non-null   int64
 54 Industrial Category - K to M - Persons
594 non-null   int64
 55 Industrial Category - K to M - Males
594 non-null   int64
 56 Industrial Category - K to M - Females
594 non-null   int64
 57 Industrial Category - N to O - Persons
594 non-null   int64
 58 Industrial Category - N to O - Males
594 non-null   int64
 59 Industrial Category - N to O - Females
594 non-null   int64

```
 60  Industrial Category - P to Q - Persons
594 non-null    int64
 61  Industrial Category - P to Q - Males
594 non-null    int64
 62  Industrial Category - P to Q - Females
594 non-null    int64
 63  Industrial Category - R to U - HHI - Persons
594 non-null    int64
 64  Industrial Category - R to U - HHI - Males
594 non-null    int64
 65  Industrial Category - R to U - HHI - Females
594 non-null    int64
 66  Industrial Category - R to U - Non HHI - Persons
594 non-null    int64
 67  Industrial Category - R to U - Non HHI - Males
594 non-null    int64
 68  Industrial Category - R to U - Non HHI - Females
594 non-null    int64
dtypes: int64(63), object(6)
memory usage: 320.3+ KB
```

[8]:
```
dataset.describe()
```

[8]:

	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females
count	5.940000e+02	594.000000	594.000000
mean	1.617277e+04	7932.700337	8240.067340
std	7.607172e+04	36864.822704	39259.545337
min	0.000000e+00	0.000000	
25%	2.872500e+02	147.250000	
50%	2.225500e+03	1147.000000	
75%	9.628500e+03	4770.500000	
max	1.200828e+06	589003.000000	

min	0.000000
25%	144.000000
50%	1076.000000
75%	4887.500000
max	611825.000000

	Worked for less than 3 months - Persons \
count	594.000000
mean	2981.629630
std	13909.621137
min	0.000000
25%	27.000000
50%	430.000000
75%	1775.250000
max	221386.000000

	Worked for less than 3 months - Males \
count	594.000000
mean	1338.289562
std	6127.047670
min	0.000000
25%	14.250000
50%	198.500000
75%	774.250000
max	99368.000000

	Worked for less than 3 months - Females \
count	594.000000
mean	1643.340067
std	7808.832522
min	0.000000
25%	13.000000
50%	213.000000
75%	946.500000
max	122018.000000

	Industrial Category - A - Cultivators - Persons \
count	594.000000
mean	865.117845
std	4274.458077
min	0.000000
25%	9.000000
50%	69.500000
75%	466.000000
max	64235.000000

Industrial Category - A - Cultivators - Males \

	Industrial Category - A - Cultivators - Females \
count	594.000000
mean	398.693603
std	1978.682322
min	0.000000
25%	4.000000
50%	32.000000
75%	204.750000
max	29603.000000

	Industrial Category - A - Agricultural labourers - Persons ... \
count	594.000000 ...
mean	12225.616162 ...
std	60458.382586 ...
min	0.000000 ...
25%	79.250000 ...
50%	1094.000000 ...
75%	6279.750000 ...
max	907752.000000 ...

	Industrial Category - N to O - Females \
count	594.000000
mean	48.013468
std	222.553500
min	0.000000
25%	0.000000
50%	2.000000
75%	18.000000
max	3565.000000

	Industrial Category - P to Q - Persons \
count	594.000000
mean	149.225589
std	696.553730
min	0.000000
25%	0.000000
50%	14.500000
75%	99.750000

	Industrial Category - P to Q - Males \
count	594.000000
mean	54.127946
std	253.067862
min	0.000000
25%	0.000000
50%	6.000000
75%	35.750000
max	4019.000000

	Industrial Category - P to Q - Females \
count	594.000000
mean	95.097643
std	444.011425
min	0.000000
25%	0.000000
50%	6.500000
75%	64.000000
max	7061.000000

	Industrial Category - R to U - HHI - Persons \
count	594.000000
mean	226.707071
std	1039.953069
min	0.000000
25%	0.000000
50%	27.000000
75%	126.750000
max	16833.000000

	Industrial Category - R to U - HHI - Males \
count	594.000000
mean	57.454545
std	265.230865
min	0.000000
25%	0.000000
50%	7.500000
75%	32.000000
max	4266.000000

	Industrial Category - R to U - HHI - Females \
count	594.000000
mean	169.252525
std	776.206806
min	0.000000

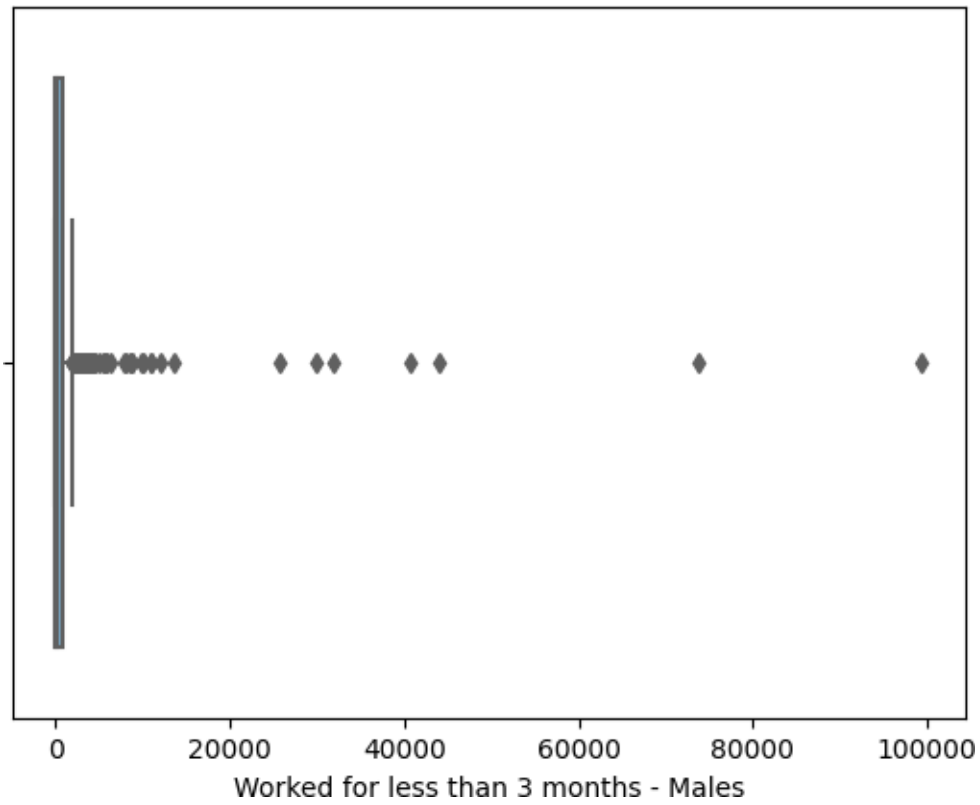
	Industrial Category - R to U - Non HHI - Persons \	Industrial Category - R to U - Non HHI - Males \	Industrial Category - R to U - Non HHI - Females
count	594.000000	594.000000	594.000000
mean	1644.282828	751.528620	892.754209
std	7325.241597	3352.811737	3988.125301
min	0.000000	0.000000	0.000000
25%	64.500000	34.000000	30.500000
50%	263.500000	123.000000	135.000000
75%	994.000000	447.750000	500.000000
max	122088.000000	55801.000000	66287.000000

[8 rows x 63 columns]

[9]:
```
sns.boxplot(dataset, x="Worked for less than 3 months - Males",
palette="Blues")
```

	Worked for less than 3 months - Males
25%	
50%	20.000000
75%	97.500000
max	12567.000000

[9]: <Axes: xlabel='Worked for less than 3 months - Males'>

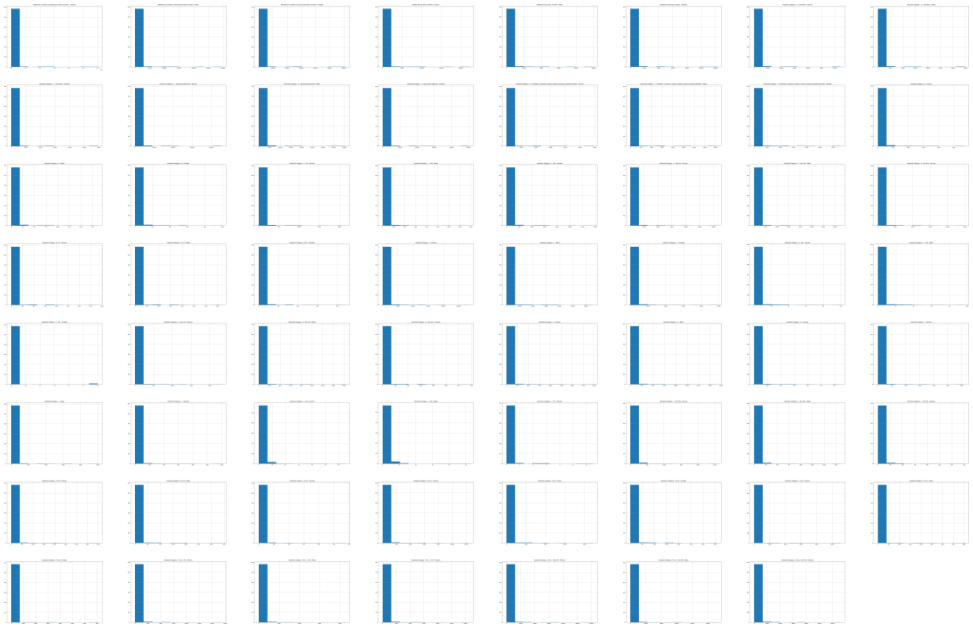


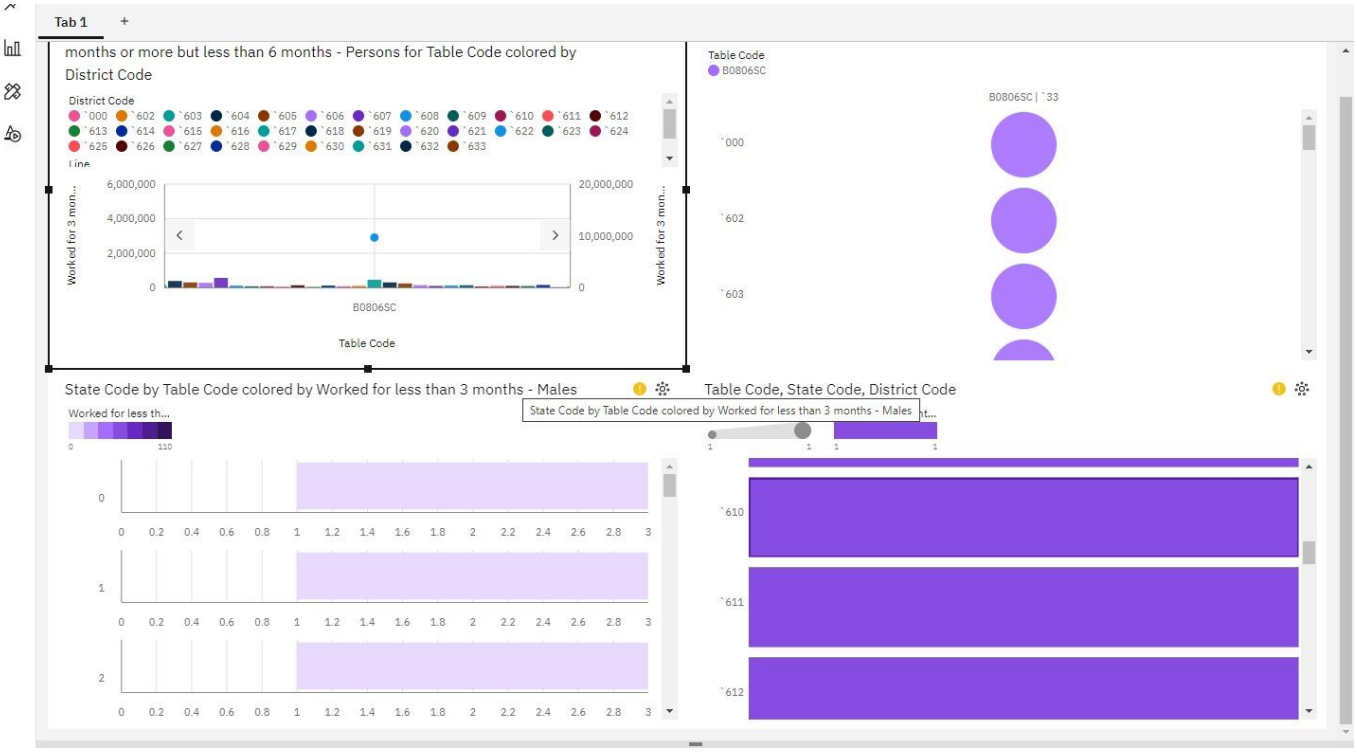
[11]:
```
dataset.hist(figsize=(155,100))
```

[11]:
```
array([[<Axes: title={'center': 'Worked for 3 months or more but less than 6
months -  Persons'}>,
        <Axes: title={'center': 'Worked for 3 months or more but less than 6
months - Males'}>,
        <Axes: title={'center': 'Worked for 3 months or more but less than 6
months - Females'}>,
        <Axes: title={'center': 'Worked for less than 3 months - Persons'}>,
        <Axes: title={'center': 'Worked for less than 3 months - Males'}>,
        <Axes: title={'center': 'Worked for less than 3 months - Females'}>,
        <Axes: title={'center': 'Industrial Category - A - Cultivators -
Persons'}>,
        <Axes: title={'center': 'Industrial Category - A - Cultivators -
Males'}>],
       [<Axes: title={'center': 'Industrial Category - A - Cultivators -
Females'}>,
        <Axes: title={'center': 'Industrial Category - A - Agricultural
labourers - Persons'}>,
        <Axes: title={'center': 'Industrial Category - A - Agricultural
labourers - Males'}>,
```

<Axes: title={'center': 'Industrial Category - A - Agricultural labourers - Females'}>,
<Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons'}>,
<Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males'}>,
<Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females'}>,
<Axes: title={'center': 'Industrial Category - B - Persons'}>],
[<Axes: title={'center': 'Industrial Category - B - Males'}>,
<Axes: title={'center': 'Industrial Category - B - Females'}>,
<Axes: title={'center': 'Industrial Category - C - HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - C - HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - C - HHI - Females'}>,
<Axes: title={'center': 'Industrial Category - C - Non HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - C - Non HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - C - Non HHI - Females'}>],
[<Axes: title={'center': 'Industrial Category - D & E - Persons'}>,
<Axes: title={'center': 'Industrial Category - D & E - Males'}>,
<Axes: title={'center': 'Industrial Category - D & E - Females'}>,
<Axes: title={'center': 'Industrial Category - F - Persons'}>,
<Axes: title={'center': 'Industrial Category - F - Males'}>,
<Axes: title={'center': 'Industrial Category - F - Females'}>,
<Axes: title={'center': 'Industrial Category - G - HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - G - HHI - Males'}>],
[<Axes: title={'center': 'Industrial Category - G - HHI - Females'}>,
<Axes: title={'center': 'Industrial Category - G - Non HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - G - Non HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - G - Non HHI - Females'}>,
<Axes: title={'center': 'Industrial Category - H - Persons'}>,
<Axes: title={'center': 'Industrial Category - H - Males'}>,
<Axes: title={'center': 'Industrial Category - H - Females'}>,
<Axes: title={'center': 'Industrial Category - I - Persons'}>],
[<Axes: title={'center': 'Industrial Category - I - Males'}>,
<Axes: title={'center': 'Industrial Category - I - Females'}>,
<Axes: title={'center': 'Industrial Category - J - HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - J - HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - J - HHI - Females'}>,
<Axes: title={'center': 'Industrial Category - J - Non HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - J - Non HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - J - Non HHI - Females'}>],
[<Axes: title={'center': 'Industrial Category - K to M - Persons'}>,
<Axes: title={'center': 'Industrial Category - K to M - Males'}>,
<Axes: title={'center': 'Industrial Category - K to M - Females'}>,
<Axes: title={'center': 'Industrial Category - N to O - Persons'}>,

<Axes: title={'center': 'Industrial Category - N to O - Males'}>,
<Axes: title={'center': 'Industrial Category - N to O - Females'}>,
<Axes: title={'center': 'Industrial Category - P to Q - Persons'}>,
<Axes: title={'center': 'Industrial Category - P to Q - Males'}>],
[<Axes: title={'center': 'Industrial Category - P to Q - Females'}>,
<Axes: title={'center': 'Industrial Category - R to U - HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - R to U - HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - R to U - HHI - Females'}>,
<Axes: title={'center': 'Industrial Category - R to U - Non HHI - Persons'}>,
<Axes: title={'center': 'Industrial Category - R to U - Non HHI - Males'}>,
<Axes: title={'center': 'Industrial Category - R to U - Non HHI - Females'}>,
<Axes: >]], dtype=object)




Tab 1
months or more but less than 6 months - Persons for Table Code colored by District Code
District Code
Line
Worked for 3 mon...
6,000,000
4,000,000
2,000,000
0
20,000,000
10,000,000
0
B0806SC
Table Code
Table Code
B0806SC
B0806SC | `33
`000
`602
`603
State Code by Table Code colored by Worked for less than 3 months - Males
Worked for less th...
0
110
State Code by Table Code colored by Worked for less than 3 months - Males
0
1
2
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
Table Code, State Code, District Code
`610
`611
`612