ASSEESSMENT OF MARGINAL WORKERS IN TAMIL NADU

Date 31-10-2023

Team ID 4498

Project Name ASSEESSMENT OF MARGINAL WORKERS IN TAMIL NADU

Import depends

In [51]:
```
import pandas as pd
import numpy as np
import xgboost as xg
```

In [52]:
```
import os
print("Current working directory:", os.getcwd())

file_path = 'datafile.csv'
if os.path.exists(file_path):
    print("The file exists.")
```

```
Current working directory: C:\Users\VIJAYRAJ R
The file exists.
```

In [60]:
```
dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
```

In [54]:
```
dataset
```

Out[54]:

	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	In C - F
0	B0806SC	`33	`000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	
1	B0806SC	`33	`000	State - TAMIL NADU	Total	`5-14	27791	14125	13666	2447	...	11	122	71	
2	B0806SC	`33	`000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	
3	B0806SC	`33	`000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	
4	B0806SC	`33	`000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
589	B0806SC	`33	`633	District - Tiruppur	Urban	`5-14	272	129	143	18	...	0	0	0	
590	B0806SC	`33	`633	District - Tiruppur	Urban	15-34	3285	1654	1631	473	...	20	44	15	
591	B0806SC	`33	`633	District - Tiruppur	Urban	35-59	3672	1769	1903	522	...	33	35	12	
592	B0806SC	`33	`633	District - Tiruppur	Urban	60+	696	399	297	111	...	0	3	0	
593	B0806SC	`33	`633	District - Tiruppur	Urban	Age not stated	2	1	1	0	...	0	0	0	

594 rows × 69 columns

In [55]:
```
dataset.head(5)
```

Out[55]:

	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	Indust Categ - P to Fema
0	B0806SC	`33	`000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7
1	B0806SC	`33	`000	State - TAMIL NADU	Total	`5-14	27791	14125	13666	2447	...	11	122	71	
2	B0806SC	`33	`000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4
3	B0806SC	`33	`000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2
4	B0806SC	`33	`000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	

5 rows × 69 columns

In [43]:
```
dataset.columns
```

Out[43]:
```
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 activitie
s - Persons',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activitie
s - Males',
       'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activitie
s - 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 [56]:
```
dataset.info()
```

```
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
```

In [57]:
```
dataset.describe()
```

Out[57]:

	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	Cat - C
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	59
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562	1643.340067	865.117845	466.424242	39
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670	7808.832522	4274.458077	2298.072295	197
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	2.872500e+02	147.250000	144.000000	27.000000	14.250000	13.000000	9.000000	5.000000	
50%	2.225500e+03	1147.000000	1076.000000	430.000000	198.500000	213.000000	69.500000	35.500000	3
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	774.250000	946.500000	466.000000	244.250000	20
max	1.200828e+06	589003.000000	611825.000000	221386.000000	99368.000000	122018.000000	64235.000000	34632.000000	296

8 rows × 63 columns

Preprossing the data

In [63]:
```
dataset = dataset.select_dtypes(include=[np.number])
correlation_matrix = dataset.corr()
```

In [64]:
```
dataset.corr(numeric_only=True)
```

Out[64]:

	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	...	Industr Catego - N to (Femal
Worked for 3 months or more but less than 6 months - Persons	1.000000	0.999263	0.999351	0.998865	0.994883	0.998631	0.987308	0.985738	0.987987	0.991821	...	0.8862
Worked for 3 months or more but less than 6 months - Males	0.999263	1.000000	0.997232	0.999020	0.997622	0.996757	0.982657	0.981185	0.983225	0.987097	...	0.8979
Worked for 3 months or more but less than 6 months - Females	0.999351	0.997232	1.000000	0.997381	0.990976	0.999052	0.990352	0.988691	0.991132	0.994927	...	0.8740
Worked for less than 3 months - Persons	0.998865	0.999020	0.997381	1.000000	0.997598	0.998522	0.984360	0.983491	0.984228	0.988923	...	0.8941
Worked for less than 3 months - Males	0.994883	0.997622	0.990976	0.997598	1.000000	0.992359	0.973140	0.972236	0.973061	0.977145	...	0.9156
...	...	...	...	...	...	...	...	...	...	...	...	
Industrial Category - R to U - HHI - Males	0.987514	0.991627	0.982327	0.988538	0.993430	0.981375	0.963460	0.961010	0.965188	0.965973	...	0.9101
Industrial Category - R to U - HHI - Females	0.983226	0.987635	0.977768	0.985253	0.991781	0.976818	0.952677	0.949565	0.955187	0.955424	...	0.9402
Industrial Category - R to U - Non HHI - Persons	0.883359	0.898087	0.868345	0.887969	0.913979	0.864574	0.820028	0.815232	0.824646	0.819429	...	0.9620
Industrial Category - R to U - Non HHI - Males	0.890343	0.904637	0.875727	0.893847	0.918710	0.871332	0.832068	0.827014	0.836972	0.829783	...	0.9454
Industrial Category - R to U - Non HHI - Females	0.874011	0.889046	0.858721	0.879533	0.906405	0.855491	0.806678	0.802119	0.811038	0.807500	...	0.9722

63 rows × 63 columns

In [58]:
```
cleandata=dataset.isnull().sum()
```

In [38]:
```
print(dataset.isna())
```

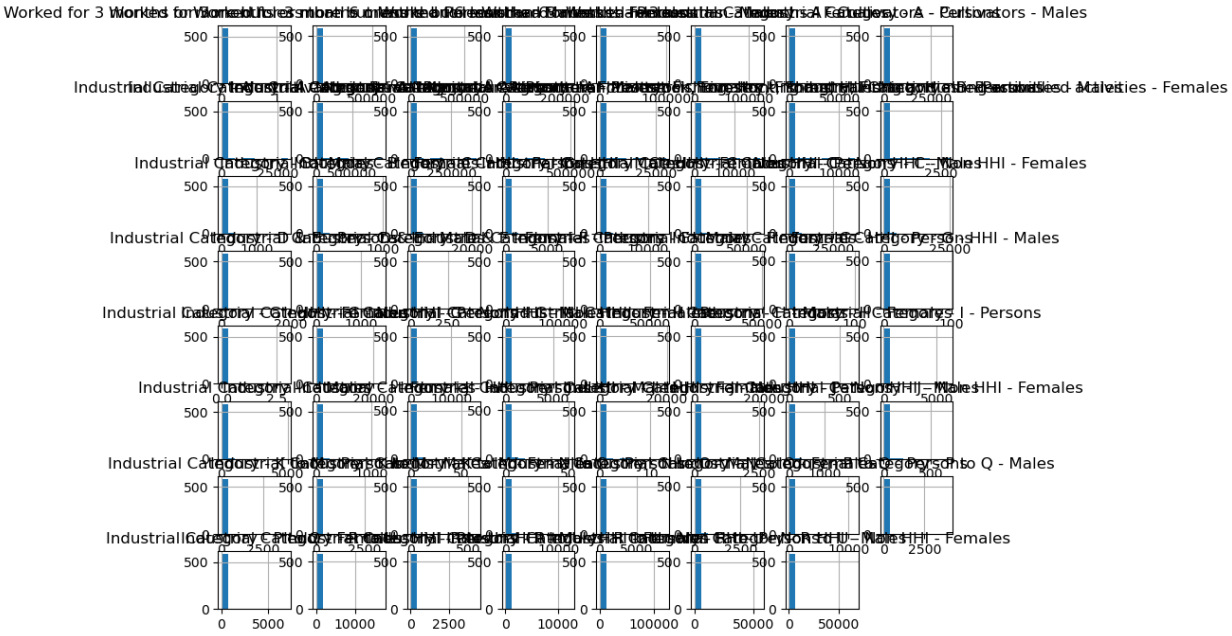
```
1                                                False
2                                                False
3                                                False
4                                                False
..                                                 ...
589                                              False
590                                              False
591                                              False
592                                              False
593                                              False

     Worked for 3 months or more but less than 6 months - Females  \
0                                                False
1                                                False
2                                                False
3                                                False
4                                                False
..                                                 ...
589                                              False
590                                              False
```

In [37]:
```
cleandata
```

Out[37]:
```
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 [41]:
```
dataset.hist(figsize=(10,8))
```

Out[41]:
```
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, Hun
ting and allied activities - Persons'}>,
        <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hun
ting and allied activities - Males'}>,
        <Axes: title={'center': 'Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hun
ting 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)
```

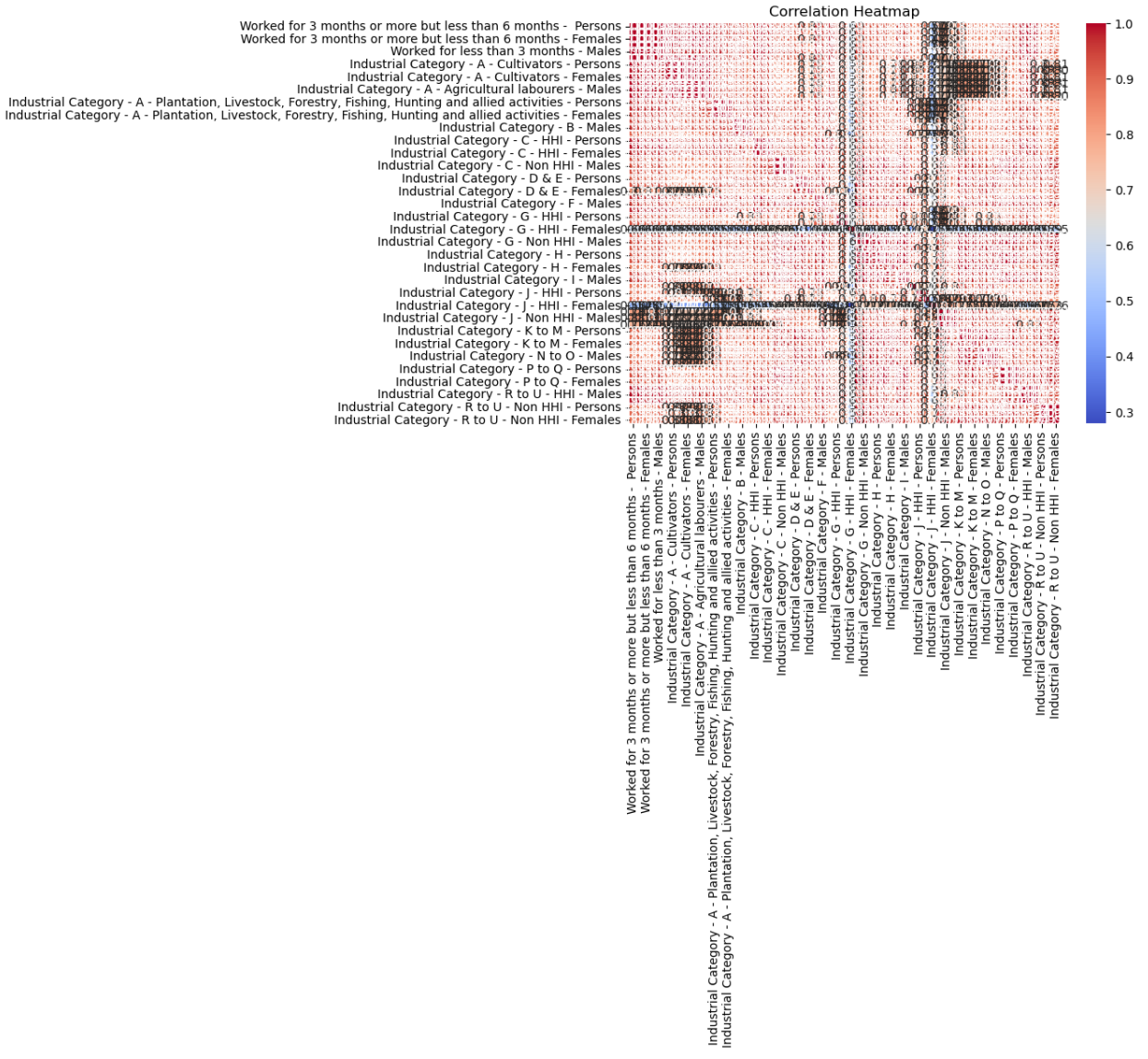


In [42]:
```
dataset.dtypes
```

Out[42]:
```
Table Code                                        object
State Code                                        object
District Code                                     object
Area Name                                         object
Total/ Rural/ Urban                               object
                                                   ...
Industrial Category - R to U - HHI - Males         int64
Industrial Category - R to U - HHI - Females       int64
Industrial Category - R to U - Non HHI - Persons   int64
Industrial Category - R to U - Non HHI - Males     int64
Industrial Category - R to U - Non HHI - Females   int64
Length: 69, dtype: object
```

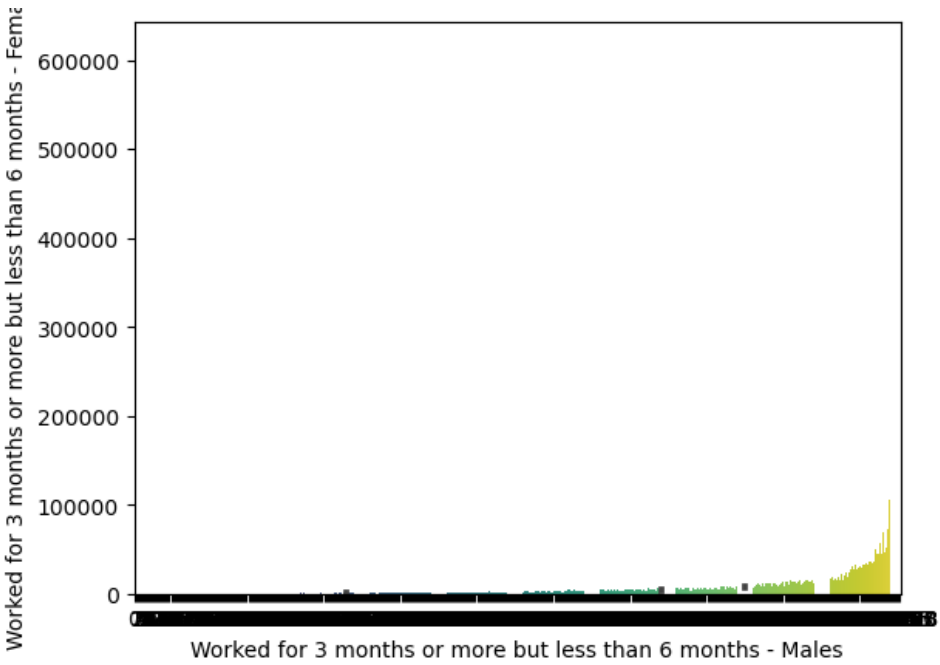
In [ ]:
```

```

In [48]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

def visualize_correlation(dataset):

    numeric_columns = dataset.select_dtypes(include='number')

    correlation_matrix = numeric_columns.corr()

    plt.figure(figsize=(8, 6))
    sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
    plt.title('Correlation Heatmap')
    plt.show()

your_dataset = pd.DataFrame(dataset)

visualize_correlation(your_dataset)
```



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

In [50]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

dataset = pd.read_csv("DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")

X = dataset['Worked for 3 months or more but less than 6 months - Males']
Y = dataset['Worked for 3 months or more but less than 6 months - Females']

sns.barplot(x=X, y=Y, data=dataset, palette='viridis')
plt.show()
```



In [*]:
```
plt.figure(figsize=(12,8))
sns.pairplot(dataset)
```

In [*]:
```
import seaborn as sns
sns.histplot(dataset, x='Worked for 3 months or more but less than 6 months - People', bins=50, color='r'
```

In [ ]:
```

```