CARE COLLEGE OF ENGINEERING

Approved by AICTE, New Delhi | Affiliated to Anna University, Chennai
Accredited by NAAC with 'A' Grade
#27, Thayanur, Tiruchirappalli - 620009

COLLEGE CODE : 8107

COURSE :DATA ANALYTICS WITH COGNOS

PHASE III: PROJECT SUBMISSION

PROJECT TITLE: Assessment of marginal workers in Tamil Nadu – A socioeconomic Analysis

TEAM MEMBERS DETAILS:

JAYASHREE. M
QUINY CHRISTABEL. R
SHINORA THOMAS. T
SHREENITHE. R

Code implementation steps:

Step1: Import libraries.

Step2: Load the given dataset.

Step3: Preprocessing the data:

- Data cleaning
- Data processing
- Data transforming

Program:

Step1

```
>>import pandas as pd
df=pd.read_csv(r"/content/nm.csv")
print(df)
```

```
590                                                 6
591                                                 9
592                                                 3
593                                                 0

     Industrial Category - R to U - HHI - Females   \
0                                              12567
1                                                258
2                                               6219
3                                               5104
4                                                974
..                                               ...
589                                                0
590                                               56
591                                               27
592                                                7
593                                                0

     Industrial Category - R to U - Non HHI - Persons   \
0                                                122088
1                                                 19305
2                                                 68929
3                                                 26498
4                                                  7065
..                                                  ...
589                                                 228
590                                                 675
591                                                 279
592                                                  81
593                                                   0

     Industrial Category - R to U - Non HHI - Males   \
0                                              55801
1                                               9774
2                                              32803
3                                               9675
4                                               3394
..                                               ...
```

Step2.1

>>df.head()

	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	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	Industr Categor R to Non HH Ma
0	B0806SC	'33	'000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266	12567	122088	5
1	B0806SC	'33	'000	State - TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	51	427	169	258	19305	9
2	B0806SC	'33	'000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127	6219	68929	3
3	B0806SC	'33	'000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487	5104	26498	9
4	B0806SC	'33	'000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	118	1457	483	974	7065	3

5 rows × 69 columns

0s completed at 4:08 PM

Step 2.2

>>df.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 
```

0s completed at 4:08 PM

Step 2.3

```
>>df.describe()
```

	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 - N to O - Females	Industrial Category - P to Q - Persons	Industrial Category - P to Q - Males	I C
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	594.000000	...	594.000000	594.000000	594.000000	5
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562	1643.340067	865.117845	466.424242	398.693603	12225.616162	...	48.013468	149.225589	54.127946	
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670	7808.832522	4274.458077	2298.072295	1978.682322	60458.382586	...	222.553500	696.553730	253.067862	4
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	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	4.000000	79.250000	...	0.000000	0.000000	0.000000	
50%	2.225500e+03	1147.000000	1076.000000	430.000000	198.500000	213.000000	69.500000	35.500000	32.000000	1094.000000	...	2.000000	14.500000	6.000000	
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	774.250000	946.500000	466.000000	244.250000	204.750000	6279.750000	...	18.000000	99.750000	35.750000	
max	1.200828e+06	589003.000000	611825.000000	221386.000000	99368.000000	122018.000000	64235.000000	34632.000000	29603.000000	907752.000000	...	3565.000000	11080.000000	4019.000000	7(

8 rows × 63 columns

Step 2.4

```
>>df.isnull()
```

	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	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	Industri Category R to U Non HHI Mal
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
589	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
590	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
591	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
592	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal
593	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	Fal

594 rows × 69 columns

Step 3.1

```
>>missingvalues=df.isna()

print(missingvalues)
```

```
593                                 False

     Industrial Category - R to U - HHI - Females  \
0                                    False
1                                    False
2                                    False
3                                    False
4                                    False
..                                     ...
589                                  False
590                                  False
591                                  False
592                                  False
593                                  False

     Industrial Category - R to U - Non HHI - Persons  \
0                                              False
1                                              False
2                                              False
3                                              False
4                                              False
..                                               ...
589                                            False
590                                            False
591                                            False
592                                            False
593                                            False

     Industrial Category - R to U - Non HHI - Males  \
0                                           False
1                                           False
2                                           False
3                                           False
4                                           False
..                                            ...
589                                         False
590                                         False
591                                         False
592                                         False
593                                         False

     Industrial Category - R to U - Non HHI - Females
0                                             False
1                                             False
```

Step 3.2

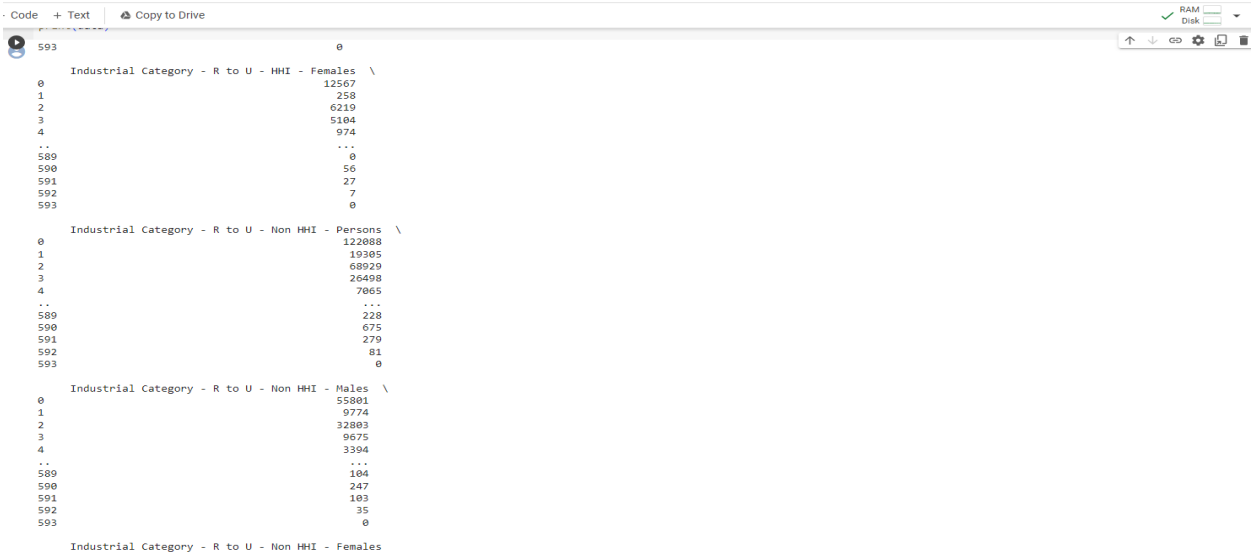
```
>>missingvalues.sum()
```

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

Step 3.3

```
>>data=df.fillna(df.mean())

print(data)
```



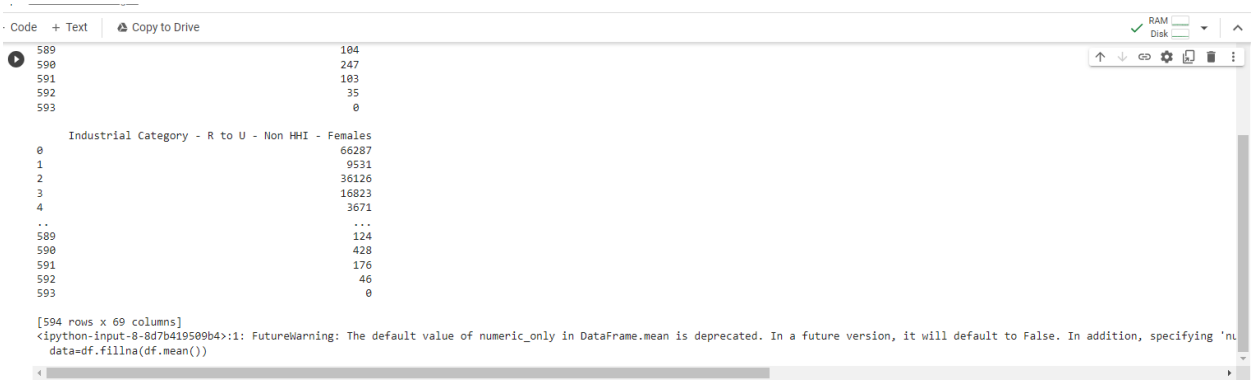
```
593                                        0

     Industrial Category - R to U - HHI - Females  \
0                                          12567
1                                            258
2                                           6219
3                                           5104
4                                            974
..                                           ...
589                                            0
590                                           56
591                                           27
592                                            7
593                                            0

     Industrial Category - R to U - Non HHI - Persons  \
0                                           122088
1                                            19305
2                                            68929
3                                            26498
4                                             7065
..                                             ...
589                                            228
590                                            675
591                                            279
592                                             81
593                                              0

     Industrial Category - R to U - Non HHI - Males  \
0                                           55801
1                                            9774
2                                           32803
3                                            9675
4                                            3394
..                                            ...
589                                           104
590                                           247
591                                           103
592                                            35
593                                             0

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



```
589                                        104
590                                        247
591                                        103
592                                         35
593                                          0

     Industrial Category - R to U - Non HHI - Females
0                                          66287
1                                           9531
2                                          36126
3                                          16823
4                                           3671
..                                           ...
589                                          124
590                                          428
591                                          176
592                                           46
593                                            0

[594 rows x 69 columns]
<ipython-input-8-8d7b419509b4>:1: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'nu
  data=df.fillna(df.mean())
```