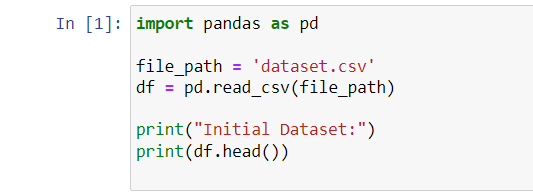
TN Marginal Workers Assessment

Phase 3: Development Part 1

Loading and Pre-processing the Dataset

Step 1: we have use the pandas for read the data set



Initial Dataset:

Table Code State Code District Code Area Name Total/ Rural/ Urban \

0 B0806SC `33 `000 State - TAMIL NADU Total

1 B0806SC `33 `000 State - TAMIL NADU Total

2 B0806SC `33 `000 State - TAMIL NADU Total

3 B0806SC `33 `000 State - TAMIL NADU Total

4 B0806SC `33 `000 State - TAMIL NADU Total

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

0 Total 1200828

1 `5-14 27791

2 15-34 514340

3 35-59 542581

4 60+ 115103

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

0 589003

1 14125

2 259560

3 251957

4 62833

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

0 611825

1 13666

2 254780

3 290624

4 52270

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

0 221386 ...

1 2447 ...

2 92423 ...

3 99202 ...

4 27165 ...

Industrial Category - N to O - Females \

0 3565

1 11

2 1754

3 1619

4 175

Industrial Category - P to Q - Persons \

0 11080

1 122

2 7536

3 3205

4 211

Industrial Category - P to Q - Males \

0 4019

1 71

2 2718

3 1131

4 93

Industrial Category - P to Q - Females \

0 7061

1 51

2 4818

3 2074

4 118

Industrial Category - R to U - HHI - Persons \

0 16833

1 427

2 8346

3 6591

4 1457

Industrial Category - R to U - HHI - Males \

0 4266

1 169

2 2127

3 1487

4 483

Industrial Category - R to U - HHI - Females \

0 12567

1 258

2 6219

3 5104

4 974

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

0 122088

1 19305

2 68929

3 26498

4 7065

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

0 55801

1 9774

2 32803

3 9675

4 3394

Industrial Category - R to U - Non HHI - Females

0 66287

1 9531

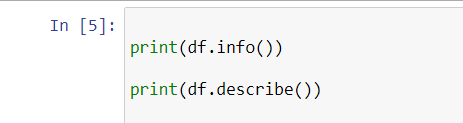
2 36126

3 16823

4 3671

[5 rows x 69 columns]

Step 2:Explore the Dataset



<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

None

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

count 5.940000e+02

mean 1.617277e+04

std 7.607172e+04

min 0.000000e+00

25% 2.872500e+02

50% 2.225500e+03

75% 9.628500e+03

max 1.200828e+06

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

count 594.000000

mean 7932.700337

std 36864.822704

min 0.000000

25% 147.250000

50% 1147.000000

75% 4770.500000

max 589003.000000

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

count 594.000000

mean 8240.067340

std 39259.545337

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 \

count 594.000000

mean 466.424242

std 2298.072295

min 0.000000

25% 5.000000

50% 35.500000

75% 244.250000

max 34632.000000

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

max 11080.000000

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

25% 0.000000

50% 20.000000

75% 97.500000

max 12567.000000

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

count 594.000000

mean 1644.282828

std 7325.241597

min 0.000000

25% 64.500000

50% 263.500000

75% 994.000000

max 122088.000000

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

count 594.000000

mean 751.528620

std 3352.811737

min 0.000000

25% 34.000000

50% 123.000000

75% 447.750000

max 55801.000000

Industrial Category - R to U - Non HHI - Females

count 594.000000

mean 892.754209

std 3988.125301

min 0.000000

25% 30.500000

50% 135.000000

75% 500.000000

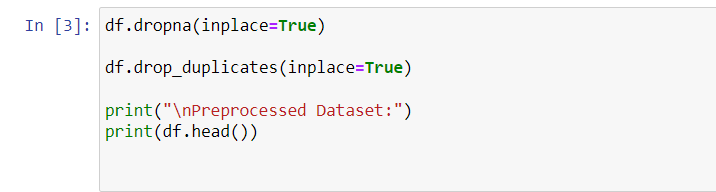
max 66287.000000

[8 rows x 63 columns]

Step 3: Data Pre-processing

1. Handling Missing Values

2. Removing Duplicates



Preprocessed Dataset:

Table Code State Code District Code Area Name Total/ Rural/ Urban \

0 B0806SC `33 `000 State - TAMIL NADU Total

1 B0806SC `33 `000 State - TAMIL NADU Total

2 B0806SC `33 `000 State - TAMIL NADU Total

3 B0806SC `33 `000 State - TAMIL NADU Total

4 B0806SC `33 `000 State - TAMIL NADU Total

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

0 Total 1200828

1 `5-14 27791

2 15-34 514340

3 35-59 542581

4 60+ 115103

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

0 589003

1 14125

2 259560

3 251957

4 62833

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

0 611825

1 13666

2 254780

3 290624

4 52270

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

0 221386 ...

1 2447 ...

2 92423 ...

3 99202 ...

4 27165 ...

Industrial Category - N to O - Females \

0 3565

1 11

2 1754

3 1619

4 175

Industrial Category - P to Q - Persons \

0 11080

1 122

2 7536

3 3205

4 211

Industrial Category - P to Q - Males \

0 4019

1 71

2 2718

3 1131

4 93

Industrial Category - P to Q - Females \

0 7061

1 51

2 4818

3 2074

4 118

Industrial Category - R to U - HHI - Persons \

0 16833

1 427

2 8346

3 6591

4 1457

Industrial Category - R to U - HHI - Males \

0 4266

1 169

2 2127

3 1487

4 483

Industrial Category - R to U - HHI - Females \

0 12567

1 258

2 6219

3 5104

4 974

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

0 122088

1 19305

2 68929

3 26498

4 7065

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

0 55801

1 9774

2 32803

3 9675

4 339++4

Industrial Category - R to U - Non HHI - Females

0 66287

1 9531

2 36126

3 16823

4 3671

[5 rows x 69 columns]