Phase 3

In this part you will begin building your project by loading and preprocessing the dataset.

Start the data analysis by loading and preprocessing the dataset

Load the dataset using Python and data manipulation libraries (e.g., pandas).

First we need to import pandas library by using the following code

```
import pandas as pd
```

Now we need to upload our dataset

```
df = pd.read_csv('/content/DDW_B06_3300_State_TAMIL_NADU-2011.csv')
```

+ Code + Text

```
print(df.head())
```

```
  Table Code State Code District Code           Area Name Total/ Rural/ Urban  \
0      B0706        `33          `000  State - TAMIL NADU               Total
1      B0706        `33          `000  State - TAMIL NADU               Total
2      B0706        `33          `000  State - TAMIL NADU               Total
3      B0706        `33          `000  State - TAMIL NADU               Total
4      B0706        `33          `000  State - TAMIL NADU               Total

  Age group  Worked for 3 months or more but less than 6 months -  Persons  \
0     Total                                                4218884
1      `5-9                                                  48238
2    `10-14                                                  76288
3     15-19                                                 257605
4     20-24                                                 478082

   Worked for 3 months or more but less than 6 months - Males  \
0                                            2136881
1                                              24511
2                                              39191
3                                             141262
4                                             257149

   Worked for 3 months or more but less than 6 months - Females  \
0                                            2082003
1                                              23727
2                                              37097
3                                             116343
4                                             220933

   Worked for less than 3 months - Persons  ...  \
0                                   723891  ...
1                                     2051  ...
2                                     6993  ...
3                                    41938  ...
4                                    81036  ...

   Industrial Category - N to O - Females  \
0                                   14495
1                                      20
2                                      44
3                                     768
4                                    2267

   Industrial Category - P to Q - Persons  \
0                                   58788
1                                     312
2                                     506
3                                    2114
4                                   11529

   Industrial Category - P to Q - Males  \
0                                 19892
1                                   169
2                                   256
3                                   695
4                                  2861

   Industrial Category - P to Q - Females  \
0                                   38896
```

Check the missing values and outliers

```
print(df.info())
```

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

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

```
       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
```

In our data set there is no missing values but in Age group column we have outliers so we need to handle that values by using dropna() function

```
import pandas as pd

# Assuming 'df' is your DataFrame
df =pd.read_csv('/content/DDW_B06_3300_State_TAMIL_NADU-2011.csv')
# Drop rows where a specific column meets a condition
df = df.drop(df[df['Age group'] == 'Total'].index)

# Drop rows based on multiple conditions
df = df.drop(df[(df['Age group'] == 'Age not stated')].index)
df = df.drop(df[df['Age group'] == '80+'].index)
df = df.drop(df[df['Age group'] == '`5-9'].index)
df = df.drop(df[df['Age group'] == '`10-14'].index)
print(df)
```

```
     Table Code State Code District Code            Area Name  \
3         B0706        `33          `000   State - TAMIL NADU
4         B0706        `33          `000   State - TAMIL NADU
5         B0706        `33          `000   State - TAMIL NADU
6         B0706        `33          `000   State - TAMIL NADU
7         B0706        `33          `000   State - TAMIL NADU
...         ...        ...           ...                  ...
1379      B0706        `33          `633  District - Tiruppur
1380      B0706        `33          `633  District - Tiruppur
1381      B0706        `33          `633  District - Tiruppur
1382      B0706        `33          `633  District - Tiruppur
1383      B0706        `33          `633  District - Tiruppur

     Total/ Rural/ Urban Age group  \
3                  Total     15-19
4                  Total     20-24
5                  Total     25-29
6                  Total     30-34
7                  Total     35-39
...                  ...       ...
1379               Urban     35-39
1380               Urban     40-49
1381               Urban     50-59
1382               Urban     60-69
1383               Urban     70-79

      Worked for 3 months or more but less than 6 months -  Persons  \
3                                                   257605
4                                                   478082
5                                                   554851
6                                                   483456
7                                                   502791
...                                                    ...
1379                                                  5043
1380                                                  8225
1381                                                  4965
```

```
1382                                          2827
1383                                           920

      Worked for 3 months or more but less than 6 months - Males  \
3                                           141262
4                                           257149
5                                           283442
6                                           240046
7                                           230695
...                                            ...
1379                                          2455
1380                                          4269
1381                                          2800
1382                                          1590
1383                                           581

      Worked for 3 months or more but less than 6 months - Females  \
3                                           116343
4                                           220933
5                                           271409
6                                           243410
7                                           272096
```

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 891 entries, 3 to 1383
Data columns (total 69 columns):
 #   Column                                                                                                  Non-Null Count
---  ------                                                                                                  --------------
 0   Table Code                                                                                              891 non-null
 1   State Code                                                                                              891 non-null
 2   District Code                                                                                           891 non-null
 3   Area Name                                                                                               891 non-null
 4   Total/ Rural/ Urban                                                                                     891 non-null
 5   Age group                                                                                               891 non-null
 6   Worked for 3 months or more but less than 6 months -  Persons                                           891 non-null
 7   Worked for 3 months or more but less than 6 months - Males                                              891 non-null
 8   Worked for 3 months or more but less than 6 months - Females                                            891 non-null
 9   Worked for less than 3 months - Persons                                                                 891 non-null
 10  Worked for less than 3 months - Males                                                                   891 non-null
 11  Worked for less than 3 months - Females                                                                 891 non-null
 12  Industrial Category - A - Cultivators - Persons                                                         891 non-null
 13  Industrial Category - A - Cultivators - Males                                                           891 non-null
 14  Industrial Category - A - Cultivators - Females                                                         891 non-null
 15  Industrial Category - A - Agricultural labourers - Persons                                              891 non-null
 16  Industrial Category - A - Agricultural labourers - Males                                                891 non-null
 17  Industrial Category - A - Agricultural labourers - Females                                              891 non-null
 18  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons  891 non-null
 19  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males    891 non-null
 20  Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Females  891 non-null
 21  Industrial Category - B - Persons                                                                       891 non-null
 22  Industrial Category - B - Males                                                                         891 non-null
 23  Industrial Category - B - Females                                                                       891 non-null
 24  Industrial Category - C - HHI - Persons                                                                 891 non-null
 25  Industrial Category - C - HHI - Males                                                                   891 non-null
 26  Industrial Category - C - HHI - Females                                                                 891 non-null
 27  Industrial Category - C - Non HHI - Persons                                                             891 non-null
 28  Industrial Category - C - Non HHI - Males                                                               891 non-null
 29  Industrial Category - C - Non HHI - Females                                                             891 non-null
 30  Industrial Category - D & E - Persons                                                                   891 non-null
 31  Industrial Category - D & E - Males                                                                     891 non-null
 32  Industrial Category - D & E - Females                                                                   891 non-null
 33  Industrial Category - F - Persons                                                                       891 non-null
 34  Industrial Category - F - Males                                                                         891 non-null
 35  Industrial Category - F - Females                                                                       891 non-null
 36  Industrial Category - G - HHI - Persons                                                                 891 non-null
 37  Industrial Category - G - HHI - Males                                                                   891 non-null
 38  Industrial Category - G - HHI - Females                                                                 891 non-null
 39  Industrial Category - G - Non HHI - Persons                                                             891 non-null
 40  Industrial Category - G - Non HHI - Males                                                               891 non-null
 41  Industrial Category - G - Non HHI - Females                                                             891 non-null
 42  Industrial Category - H - Persons                                                                       891 non-null
 43  Industrial Category - H - Males                                                                         891 non-null
 44  Industrial Category - H - Females                                                                       891 non-null
 45  Industrial Category - I - Persons                                                                       891 non-null
 46  Industrial Category - I - Males                                                                         891 non-null
 47  Industrial Category - I - Females                                                                       891 non-null
 48  Industrial Category - J - HHI - Persons                                                                 891 non-null
 49  Industrial Category - J - HHI - Males                                                                   891 non-null
 50  Industrial Category - J - HHI - Females                                                                 891 non-null
 51  Industrial Category - J - Non HHI - Persons                                                             891 non-null
```

Now our dataset preprocessing is complete now we will use this dataset for Analysis