NAAN MUDHALVAN PROJECT PHASE 3: DEVELOPMENT PART 1

PROJECT TITTLE: TN MARGINAL WORKERS - SOCIOECONOMIC ANALYSIS

DATA ANALYTICS

Import Libraries:

pythonCopy code

import pandas as pd

Load the Dataset:

If the dataset is in a CSV file:

pythonCopy code

df = pd.read_csv('your_dataset.csv')

If the dataset is in an Excel file:

pythonCopy code

df = pd.read_excel('your_dataset.xlsx')

For other formats, pandas provides similar functions (e.g., pd.read_sql() for databases, pd.read_json() for JSON files, etc.).

Explore the Dataset:

After loading, it's a good practice to take a look at the first few rows of the dataset using df.head() and check the data types and null values using df.info().

Handle Missing Data:

If there are missing values, you can choose to drop them or fill them with appropriate values using df.dropna() or df.fillna().

Data Cleaning and Transformation:

Perform any necessary data cleaning steps (e.g., removing duplicates, correcting data types, etc.) and transformations (e.g., feature scaling, encoding categorical variables, etc.).

Exploratory Data Analysis (EDA):

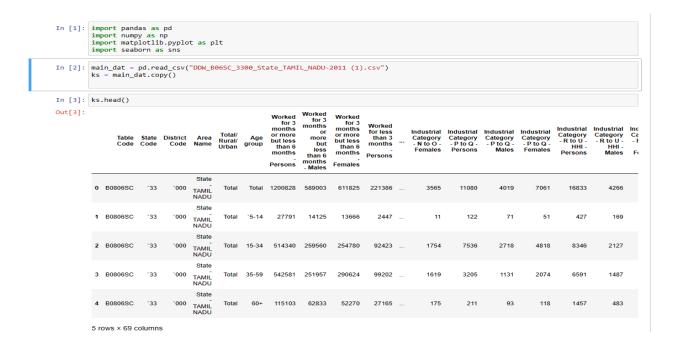
Conduct exploratory data analysis to understand the characteristics and relationships within the dataset. This may include summary statistics. visualizations. etc.

Feature Engineering:

If needed, create new features or modify existing ones to enhance the performance of your analysis or machine learning model.

Split Data (if applicable):

If you're doing predictive modeling, split the dataset into training and testing sets using techniques like cross-validation.

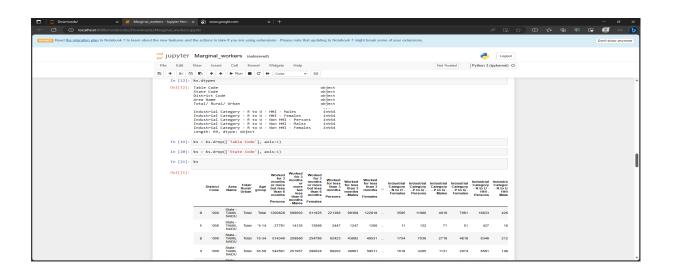


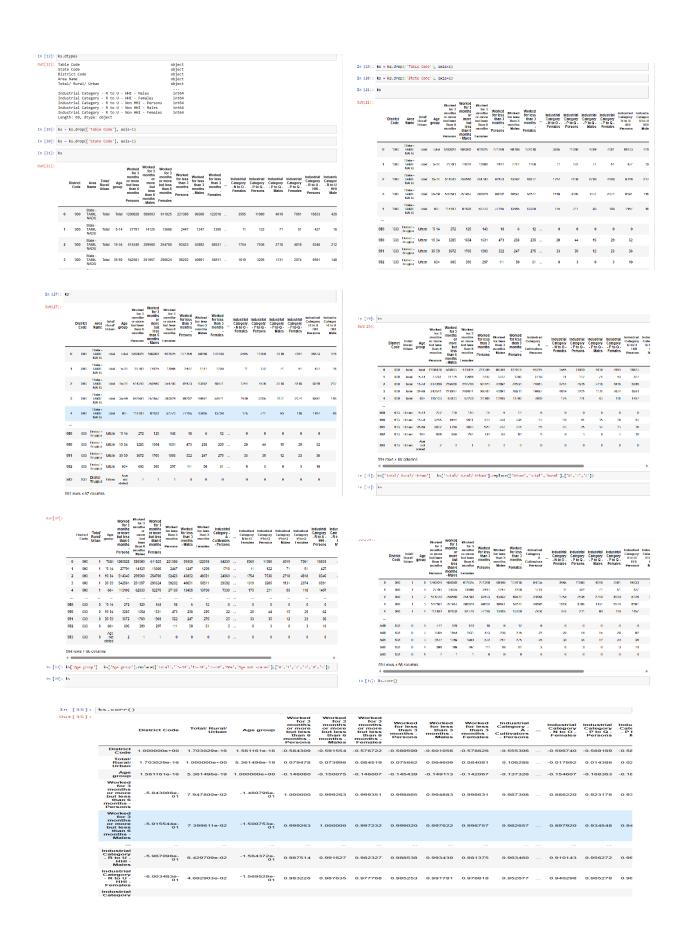
: ks	.sample(5)															
]:	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months	Worked for less than 3 months 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	Industri Catego - R to U HH Person
42	6 B0806SC	.33	`624	District - Theni	Urban	Total	5214	3003	2211	1151		18	62	12	50	:
25	2 B0806SC	.33	`615	District - Perambalur	Total	Total	16797	7835	8962	1741		12	140	36	104	1
57	5 B0806SC	.33	`632	District - Coimbatore	Urban	Age not stated	13	10	3	0		0	0	0	0	
33	6 B0806SC	.33	`619	District - Thiruvarur	Urban	Total	4807	2585	2222	788		0	20	0	20	
46	5 B0806SC	.33	`626	District - Ramanathapuram	Urban	35-59	923	440	483	253		0	26	0	26	
5 n	ows × 69 co	lumns														
4																

```
In [6]: ks.columns
                                                                                                                       Industrial Category - F - Persons', 'Industrial Category - F - Males',
'Industrial Category - F - Females - Persons',
'Industrial Category - G - HHI - Males',
'Industrial Category - G - HHI - Females',
'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 - G - Non HHI - Females',
'Industrial Category - G - Non HHI - Females',
'Industrial Category - I - Persons', 'Industrial Category - H - Males',
'Industrial Category - I - Persons', 'Industrial Category - I - Males',
'Industrial Category - I - Females',
'Industrial Category - J - HHI - Persons',
'Industrial Category - J - HHI - Persons',
'Industrial Category - J - Non HHI - Females',
'Industrial Category - J - Non HHI - Persons',
'Industrial Category - Non HHI - Females',
'Industrial Category - K to M - Persons',
'Industrial Category - K to M - Males',
'Industrial Category - N to O - Persons',
'Industrial Category - N to O - Persons',
'Industrial Category - P to Q - HHI - Hales',
'Industrial Category - R to U - HHI - Hales',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Category - R to U - Non HHI - Females',
'Industrial Catego
                                     In [7]: pd.isnull(ks).sum()
                                      Out[7]: Table Code
                                                                                                                                                                                                                                                                                                                                                                            0 0 0
                                                                                   District Code
                                                                                    Area Name
Total/ Rural/ Urban
                                                                                  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
Length: 69, dtype: int64
                                                                                                                                                                                                                                                                                                                                                                             0
                                     In [8]: ks.dropna(inplace=True)
                                                                                  pd.isnull(ks).sum()
                                     Out[8]: Table Code
                                                                                                                                                                                                                                                                                                                                                                            0 0 0
                                                                                  State Code
District Code
                                                                                   Area Name
Total/ Rural/ Urban
                                                                                  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
Length: 69, dtype: int64
                                                                                                                                                                                                                                                                                                                                                                            0
```

```
In [8]: ks.dropna(inplace-True)
  pd.isnull(ks).sum()
Out[8]: Table Code
State Code
District Code
Area Name
Total/ Rural/ Urban
             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
Length: 69, dtype: int64
In [9]: ks.describe()
                       Worked for 3
months or
more but
less than 6
months -
Persons

Worked for 3
more but less
than 6 months
- Males
                                                                                     Worked for
less than 3
months -
Persons
                                                                                                                           594.000000
                                                                                                                                             594.000000
                                                                                                                                                                                                     594.000000
              count 5.940000e+02
                                              594.000000
                                                                 594.000000
                                                                                     594.000000
                                                                                                       594.000000
                                                                                                                                                               594.000000
                                                                                                                                                                                 594.000000
               mean 1.617277e+04
                                             7932.700337
                                                                8240.067340
                                                                                    2981.629630
                                                                                                      1338.289562
                                                                                                                          1643.340067
                                                                                                                                             865.117845
                                                                                                                                                               466.424242
                                                                                                                                                                                  398.693603
                                                                                                                                                                                                  12225.616162
                std 7.607172e+04 36864.822704 39259.545337
                                                                                   13909.621137
                                                                                                    6127.047670
                                                                                                                         7808.832522 4274.458077
                                                                                                                                                              2298.072295
                                                                                                                                                                                1978.682322
                                                                                                                                                                                                  60458.382586
                                                             0.000000
144.000000
                 min 0.000000e+00
                                                0.000000
                                                                                                                                               0.000000
                                                                                                                                                                  0.000000
              25% 2.872500e+02 147.250000
                                                                                    27.000000
                                                                                                    14.250000
                                                                                                                         13.000000
                                                                                                                                                                                4.000000
                                                                                                                                                               5.000000
                                                                                                                                                                                                    79.250000
                                                                                                                                             9.000000
                 50% 2.225500e+03
                                         1147.000000 1076.000000
4770.500000 4887.500000
                                                                                     430.000000
                                                                                                    198.500000
774.250000
                                                                                                                          213.000000 69.500000
946.500000 466.000000
                                                                                                                                                                35.500000
                                                                                                                                                                                   32.000000
                                                                                                                                                                                                    1094.000000
                                                                                  1775.250000
                                                                                                                                                              244.250000
                                                                                                                                                                                204.750000
                                                                                                                                                                                                   6279.750000
              75% 9.628500e+03
                max 1.200828e+06 589003.00000 611825.000000 221386.000000 99368.00000 122018.000000 64235.000000 34632.000000 29603.000000 907752.000000
             8 rows × 63 columns
```





In [43]: ks Out[43]: 0 1200828 1 27791 51 1 2 514340 259560 92423 43892 1903 247 275 0 3 3672 1769 5 ... 633 0 594 rows × 66 columns In [44]: ks.to_csv('cleaned_data.csv', index-False)