ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU

Phase3

Download the data set

- **Download the dataset for Marginal workers in tamilnadu Using the following link**
- **Dataset link:**
- https://tn.data.gov.in/catalog/marginal-workers-classified-age-industrial-category-and-sex-census-2011-india-and-states

Insertion the dataset

- #import necessary libraries
- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- #loading dataset
- df= pd.read_csv(“marginal.csv”)
- df.head()

Output

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	for 3 months or more but less than 6 months - Persons	for 3 months or more but less than 6 months - Males	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	In Ca - I N - I
0	B0806SC	'33	'000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266	12567	
1	B0806SC	'33	'000	State - TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	51	427	169	258	
2	B0806SC	'33	'000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127	6219	
3	B0806SC	'33	'000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487	5104	
4	B0806SC	'33	'000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	118	1457	483	974	

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	Industrial Category - R to U - Non HHI - Males	Industrial Category - R to U - Non HHI - Females
State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266	12567	122088	55801	66287
State - TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	51	427	169	258	19305	9774	9531
State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127	6219	68929	32803	36126
State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487	5104	26498	9675	16823
State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	118	1457	483	974	7065	3394	3671

Missing values in dataset

- #find missing values in the dataset
- df.isnull().
- sum(axis=0)

```
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
```

Retrieve and Testing the dataset

- # Retrieve training and testing dataset
- X = df.iloc[:,:-1]
- Y = df.iloc[:,-1
- print(X)

Output

	Table Code	State Code	District Code	Area Name	\
0	B0806SC	`33	`000	State - TAMIL NADU	
1	B0806SC	`33	`000	State - TAMIL NADU	
2	B0806SC	`33	`000	State - TAMIL NADU	
3	B0806SC	`33	`000	State - TAMIL NADU	
4	B0806SC	`33	`000	State - TAMIL NADU	
..	...	...	...	...	
589	B0806SC	`33	`633	District - Tiruppur	
590	B0806SC	`33	`633	District - Tiruppur	
591	B0806SC	`33	`633	District - Tiruppur	
592	B0806SC	`33	`633	District - Tiruppur	
593	B0806SC	`33	`633	District - Tiruppur	

	Total/ Rural/ Urban	Age group	\
0	Total	Total	
1	Total	`5-14	
2	Total	15-34	
3	Total	35-59	
4	Total	60+	
..	...	...	
589	Urban	`5-14	
590	Urban	15-34	
591	Urban	35-59	
592	Urban	60+	
593	Urban	Age not stated	

	...
592	35
593	0

```
X = df.iloc[:,:-1]
Y = df.iloc[:,-1]
print(Y)
```

```
0      66287
1       9531
2      36126
3      16823
4       3671
       ...
589      124
590      428
591      176
592       46
593        0
Name: Industrial Category - R to U - Non HHI - Females, Length: 594, dtype: int64
```

Select relevant Columns For analysis

- # Select relevant columns for analysis
- selected_columns = [‘Age group’, ‘Industrial Category – A – Cultivators – Persons’,
- ‘Industrial Category – B – Persons’, ‘Industrial Category – C – HHI – Persons’,
- ‘Industrial Category – D & E – Persons’, ‘Industrial Category – F – Persons’,
- ‘Industrial Category – G – HHI – Persons’, ‘Industrial Category – H – Persons’,
- ‘Industrial Category – I – Persons’, ‘Industrial Category – J – HHI – Persons’,
- ‘Industrial Category – K to M – Persons’, ‘Industrial Category – N to O – Persons’,
- ‘Industrial Category – P to Q – Persons’, ‘Industrial Category – R to U – HHI – Persons’]

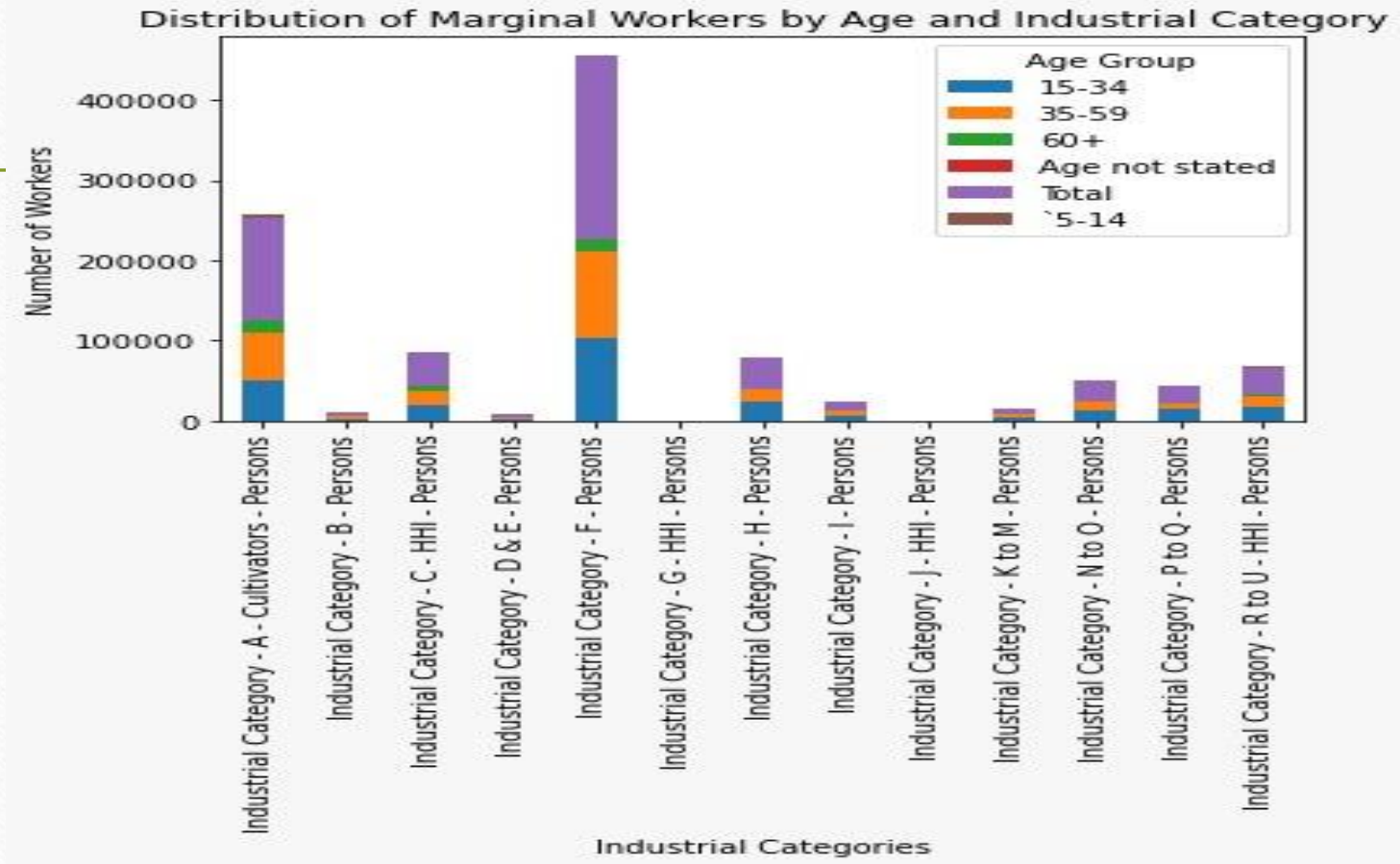
Filter the dataframe

- # Filter the DataFrame for marginal workers and selected columns
- marginal_workers_df = df[df['Total/ Rural/ Urban'] == 'Total'][selected_columns]
- # Group by age group and sum the counts for each industrial category
- age_group_data = marginal_workers_df.groupby('Age group').sum()

Plotting points

- # Plotting
- plt.figure(figsize=(15, 8))
- age_group_data.T.plot(kind='bar', stacked=True)
- plt.title('Distribution of Marginal Workers by Age and Industrial Category')
- plt.xlabel('Industrial Categories')
- plt.ylabel('Number of Workers')
- plt.legend(title='Age Group')
- plt.show()

Output:


Distribution of Marginal Workers by Age and Industrial Category
Number of Workers
400000
300000
200000
100000
0
Age Group
15-34
35-59
60+
Age not stated
Total
`5-14
Industrial Category - A - Cultivators - Persons
Industrial Category - B - Persons
Industrial Category - C - HHI - Persons
Industrial Category - D & E - Persons
Industrial Category - F - Persons
Industrial Category - G - HHI - Persons
Industrial Category - H - Persons
Industrial Category - I - Persons
Industrial Category - J - HHI - Persons
Industrial Category - K to M - Persons
Industrial Category - N to O - Persons
Industrial Category - P to Q - Persons
Industrial Category - R to U - HHI - Persons
Industrial Categories

Conclusion:

- Tamil Nadu marginal workers were classified and various processings were done using the given dataset