

INNOVATION

PROJECT TITLE : Assessment of marginal workers in tamilnadu -A socioeconomic analysis

Summary :

1. **Definition:** Marginal workers are those who work for a relatively short duration in a year and have limited access to regular employment.
2. **Demographics:** Analyze the age, gender, and educational background of marginal workers in Tamil Nadu.
3. **Employment Types:** Examine the various sectors where marginal workers are employed, such as agriculture, construction, or informal labor.
4. **Income Levels:** Assess the income disparities among marginal workers and compare them to the state's minimum wage standards.
5. **Livelihood Challenges:** Highlight the challenges they face, including job insecurity, lack of benefits, and vulnerability to economic shocks.
6. **Social Welfare Programs:** Discuss government initiatives aimed at improving the livelihoods of marginal workers, such as skill development and social safety nets.
7. **Policy Recommendations:** Provide suggestions for policies that can uplift the living conditions and job security of marginal workers in Tamil Nadu.

This summary can serve as a starting point for a comprehensive assessment report on marginal workers in the state.

Dataset and its details:

We used the below dataset for our project:

Dataset Link: <https://tn.data.gov.in/catalog/marginal-workers-classified-age-industrial-category-and-sex-census-2011-india-and-states>

Here are explanations for the columns you mentioned:

1. ***Table Code:*** This column may indicate the unique identifier or code for the specific table or dataset being referred to.
2. ***State Code:*** This column likely represents a unique identifier or code for different states or regions within a country.
3. ***District Code:*** It's probably an identifier or code for various districts or subdivisions within the states.
4. ***Area Name:*** This column likely contains the names or labels for specific areas, which could be districts, towns, or regions.
5. ***Total/ Rural/ Urban:*** This is likely a categorical column that classifies the data into three categories: "Total," "Rural," and "Urban." It may represent different population segments or areas.
6. ***Age group:*** This column may contain information about different age groups or ranges, used for demographic analysis.
7. ***Worked for 3 months or more but less than 6 months - Persons/Males/Females:*** These columns likely contain data about the number of persons, males, and females who have worked for a specific duration (between 3 and 6 months).

8. *Worked for less than 3 months - Persons/Males/Females:* Similar to the previous columns, these represent data about the number of persons, males, and females who have worked for less than 3 months.
9. *Industrial Category - A - Cultivators - Persons/Males/Females:* These columns seem to pertain to a specific industrial category (Category A - Cultivators) and provide data on the number of persons, males, and females in this category.
10. *Industrial Category - A - Agricultural laborers - Persons/Males/Females:* Similarly, these columns appear to be related to the agricultural laborers within Category A, providing data on persons, males, and females.
11. *Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting, and allied activities - Persons:* This column appears to contain data on the number of persons involved in activities related to plantations, livestock, forestry, fishing, hunting, and allied activities within Category A.
12. *Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males/Females:* These columns seem to pertain to Category A and provide data on the number of males and females involved in activities related to plantation, livestock, forestry, fishing, hunting, and allied activities.
13. *Industrial Category - B - Persons/Males/Females:* These columns appear to be related to Category B and provide data on the number of persons, males, and females within this specific industrial category.
14. *Industrial Category - C - HHI (Household Industries) - Persons/Males/Females:* These columns represent data for Category C in the subcategory of Household Industries (HHI) and provide counts of persons, males, and females involved in these industries.
15. *Industrial Category - C - Non HHI (Non-Household Industries) - Persons/Males/Females:* Similar to the previous columns, these represent data for Category C in the subcategory of Non-Household Industries, providing counts of persons, males, and females.

16. *Industrial Category - D & E - Persons/Males/Females:* These columns are related to Category D & E, providing data on the number of persons, males, and females in these industrial categories.
17. *Industrial Category - F - Persons/Males/Females:* These columns pertain to Category F and provide data on the number of persons, males, and females involved in this specific industrial category.
18. *Industrial Category - G - HHI (Household Industries) - Persons/Males/Females:* These columns are similar to the Category C columns but for Category G in the Household Industries subcategory.
19. *Industrial Category - G - Non HHI (Non-Household Industries) - Persons/Males/Females:* These columns represent Category G but for Non-Household Industries.
20. *Industrial Category - H - Persons/Males/Females:* These columns represent data for Category H and provide counts of persons, males, and females in this industrial category.
21. *Industrial Category - I - Persons/Males/Females:* These columns are related to Category I, providing data on the number of persons, males, and females in this industrial category.
22. *Industrial Category - J - HHI (Household Industries) - Persons/Males/Females:* Similar to Category G and C columns, these represent Category J in the Household Industries subcategory.
23. *Industrial Category - J - Non HHI (Non-Household Industries) - Persons/Males/Females:* These columns are related to Category J but for Non-Household Industries.
24. *Industrial Category - K to M - Persons/Males/Females:* These columns pertain to Categories K to M and provide data on the number of persons, males, and females within these categories.

25. ***Industrial Category - N to O - Persons/Males/Females:*** These columns are related to Categories N to O, providing data on the number of persons, males, and females involved in these industrial categories.

26. ***Industrial Category - P to Q - Persons/Males/Females:*** These columns represent Categories P to Q and provide counts of persons, males, and females in these industrial categories.

27. ***Industrial Category - R to U - HHI (Household Industries) - Persons/Males/Females:*** These columns are similar to the Category G, C, and J columns but for Categories R to U in the Household Industries subcategory.

28. ***Industrial Category - R to U - Non HHI (Non-Household Industries) - Persons/Males/Females:*** These columns represent Categories R to U but for Non-Household Industries.

Libraries to be installed and used:

The Python code provided uses several libraries for data analysis and visualization. Here's an explanation of the libraries used in the code:

1. ***Pandas (`import pandas as pd`):*** Pandas is a powerful library for data manipulation and analysis. It provides data structures like dataframes and tools for cleaning, transforming, and analyzing data. In the code, Pandas is used to load the dataset from a CSV file and perform various data-related operations, such as filtering and calculating statistics.

2. ***Matplotlib (`import matplotlib.pyplot as plt`):*** Matplotlib is a popular library for creating static, animated, or interactive visualizations in Python. In the code, Matplotlib is used for creating various types of plots, including histograms and bar charts, to visualize data distributions and relationships.

3. ***Seaborn (`import seaborn as sns`):*** Seaborn is a data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. In the code, Seaborn is used for creating more visually appealing and informative plots, such as histograms and bar plots.

These libraries are essential for working with data, performing data analysis, and creating data visualizations. They are commonly used in data science and analysis tasks to explore, understand, and communicate insights from datasets.

Testing and training of dataset:

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
# Load your socio-economic data into a DataFrame (replace 'your_data.csv' with  
your data file)
```

```
data = pd.read_csv('your_data.csv')
```

```
# Explore the data
```

```
print(data.head()) # Display the first few rows of the data
```

```
print(data.info()) # Check data types and missing values
```

```
# Filter the data for marginal workers in Tamil Nadu (assuming 'State Code' or  
'Area Name' columns)
```

```
tn_marginal_workers = data[(data['State Code'] == 'TN') & (data['Category'] ==  
'Marginal Worker')]
```

```
# Basic statistical summary
```

```
marginal_stats = tn_marginal_workers.describe()
```

```
# Data visualization
```

```
# Example: Create a bar plot showing the distribution of marginal workers by age  
group
```

```
sns.countplot(x='Age group', data=tn_marginal_workers)
```

```
plt.title('Distribution of Marginal Workers by Age Group in Tamil Nadu')
```

```
plt.xticks(rotation=45)
```

```
plt.show()
```

```
# Example: Create a pie chart to show the gender distribution of marginal workers
```

```
gender_distribution = tn_marginal_workers['Gender'].value_counts()
```

```
plt.pie(gender_distribution, labels=gender_distribution.index, autopct='%1.1f%%',  
startangle=140)
```

```
plt.axis('equal')
```

```
plt.title('Gender Distribution of Marginal Workers in Tamil Nadu')
```

```
plt.show()
```

You can perform more in-depth analysis and visualization based on your specific research questions and data.

Metrics used for accuracy check:

1. ***Descriptive Statistics:*** For basic data exploration and understanding the characteristics of marginal workers, you can use descriptive statistics like mean, median, standard deviation, and quartiles for quantitative variables (e.g., income) and frequency tables for categorical variables (e.g., employment type).
2. ***Percentage Distribution:*** Calculate the percentage distribution of different categories of marginal workers, such as by gender, age group, education level, and employment type. This can help you understand the composition of this group.
3. ***Average Monthly Income:*** If your analysis involves predicting or comparing monthly income, metrics like mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) can be used to assess the accuracy of income predictions.
4. ***Employment Type Classification:*** If you're classifying marginal workers into specific employment types, classification metrics like accuracy, precision, recall, F1-score, and confusion matrices can help evaluate the model's performance.
5. ***Correlation Analysis:*** Determine the correlations between different variables to understand relationships. You can use correlation coefficients like Pearson's r for quantitative variables.
6. ***Visualization:*** Data visualizations, such as bar charts, histograms, and scatter plots, can provide valuable insights into the characteristics and distribution of marginal workers.
7. ***Geospatial Analysis:*** If your dataset includes geographical information, you might consider geospatial metrics and visualization techniques to understand the geographic distribution of marginal workers.

8. ***Qualitative Assessment:*** In some cases, qualitative assessment methods like interviews, surveys, or focus groups may be used to gather subjective feedback and insights from marginal workers themselves.