ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU.

A SOCIOECONOMICAN ANALYSIS (ADS).

TEAM MEMBER

732521104056 : VELMURUGAN P

**Phase 5 Submission Document**

**Project :** ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU.

**Introduction:**

Tamil Nadu, a diverse and vibrant state in southern India, is home to a significant population of marginalized workers who face various economic and social challenges. Marginal workers, in the context of this assessment, are those individuals who are engaged in low-paying, informal, and often precarious employment, and typically have limited access to social and economic resources. These workers play a crucial role in various sectors of the state's economy, including agriculture, construction, and informal services.

This assessment seeks to understand the conditions and dynamics surrounding marginal workers in Tamil Nadu, shedding light on the challenges they encounter and the opportunities for improving their well-being. By examining their demographic characteristics, living conditions, income levels, and access to social services, we aim to provide a comprehensive view of the socioeconomic status of this vulnerable group.

Through this assessment, we intend to identify potential areas for policy interventions, support systems, and empowerment initiatives that can uplift the living standards and social integration of marginal workers in Tamil Nadu. This research is essential not only to address immediate concerns but also to work towards a more equitable and inclusive society that ensures the well-being of all its residents.

The subsequent sections of this assessment will delve into the methodology, findings, and recommendations to provide a thorough understanding of the challenges and opportunities for marginalized workers in Tamil Nadu.

**Problem Definition:**

The project involves analyzing the demographic characteristics of marginal workers in Tamil Nadu based on their age, industrial category, and sex. The objective is to perform a socioeconomic analysis and create visualizations to represent the distribution of marginal workers across different categories. This project includes defining objectives, designing the analysis approach, selecting appropriate visualization types, and performing the analysis using Python and data visualization libraries.

In this I am going to put my design into innovation to solve the problem.

In this I am going to begin building your project by loading and preprocessing the dataset.

In this I am going to continue building the project by performing different activities like feature engineering, model training, evaluation etc as per the instructions in the project.

**Design Thinking:**

1. Project Objectives: Define objectives such as analyzing marginal worker demographics, understanding age and gender distribution, and exploring industrial categories.
2. Analysis Approach: Plan the steps to extract, clean, and analyze the dataset to derive insights.
3. Visualization Selection: Determine suitable visualization types (e.g., bar charts, pie charts, heatmaps) to represent demographic distributions effectively.

**THE FOUR MAJOR CATEGORIES:**

 Cultivators.

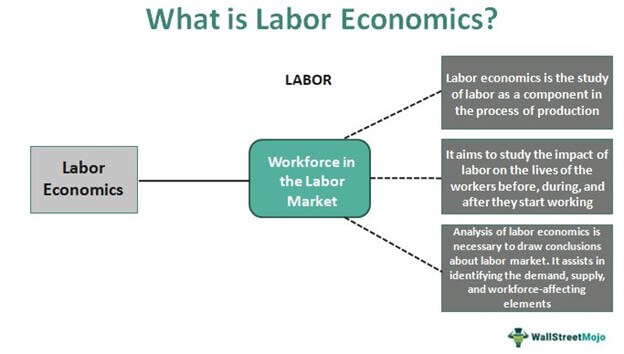
 Agricultural Labourers.

 Household Industrial Workers.

 Other Workers.

**THE FOUR MAJOR TYPES OF LABOR ARE:**

* professional,
* semi-skilled,
* skilled, and
* unskilled labor.



Assertion: The working population classified according to occupations is known as **occupational structure**. Reason: Occupational structure indicates development of a country.

**NEW CATEGORIES FOR CENSUS 2023:**

1. Gender: New category “Other” introduced in addition to Male and Female.   
2. Date of Birth question introduced along with Age.   
3. Current Marital Status: Separate codes Assigned for Separated and Divorced.   
4. New filter Question on SC/ST Introduced – “Is this person SC/ST?”

5. Disability: The question on disability canvassed at the Census 2001 has been modified. Household Schedule attempts to collect information on eight types of disabilities as against five included in the Household Schedule of Census of India 2001. The information is being collected on disabilities namely, disability ‘In Seeing’, ‘In Hearing’, ‘In Speech’, ‘In Movement’, ‘Mental retardation’, ‘Mental Illness’, ‘Any Other’ and ‘Multiple Disability’.   
6. Literacy Status for “Other” sex added in addition to existing Male and Female.   
7. New Codes under Status of Attendance in Educational Institutions introduced for Not Attending viz., (i) Attended before and (ii) Never attended.   
8. Work: In the previous censuses, workers were categorized as ‘Main workers’ and ‘Marginal workers’. Those who worked for more than 6 months during last year were categorized as ‘Main workers’ whereas those who worked less than 6 months were categorized as ‘Marginal Workers’. At the Census 2011, for better capturing and analysis of Census data, ‘Marginal workers have been classified into two categories viz., (i) worked for 3 months or more but less than 6 months (ii) worked for less than 3 months. The definition of ‘Main worker’ remains the same.   
9. A separate code-5 has been included under Non-economic activity for rentiers.   
10. Migration – Provision to specify the present name of the Village/Town of the Birth Place as well as the Place of Last Residence introduced.   
11. Name of the Institutional Household is also being recorded.   
  
 This was stated by the Minister of State in the Ministry of Home Affairs, Shri Gurudas Kamat in written reply to a question in Rajya Sabha today.

**A SOCIOECONOMICAN ANALYSIS (ADS):**

**INTRODUCTION:**

One of the main objectives of socioeconomic assessment is **to identify various socioeconomic positive and negative impacts**. The socioeconomic impacts assessment should highlights possible relationships between assessed socioeconomic variables and the environmental quality in the study site.

**SOCIO-ECONOMIC ANALYSES:**

**Socio-economic analyses can help us assess the benefits and costs associated with climate change adaptation measures.**

* socio-economic analyses to calculate the value of advantages and disadvantages for society of different climate change adaptation measures. In this way we can find the measures that will be of greatest use for society or which are the most cost-effective.

Socio-economic screening of climate change adaptation:

The Ministry of Climate and Energy has published a cross-sector, national socio-economic screening of climate change adaptation, June 2010.  An English summary of the report can be downloaded from the publication list.

* The screening looks at climate adaptation across the 14 sectors dealt with in the government's climate change adaptation strategy from 2008.
* The screening indicates that coastal protection, buildings, roads/railways and sewerage are especially relevant candidates for more in-depth analyses. Potential damage costs are high in these sectors, and the example calculations indicate that these sectors have the greatest potential for limiting damage costs in a cost-effective manner through adaptation measures.
* The sectors mentioned are characterised by long-term investments, and this calls for early incorporation of climate change adaption. Within the sector of coastal protection however, it is possible to implement adaptation

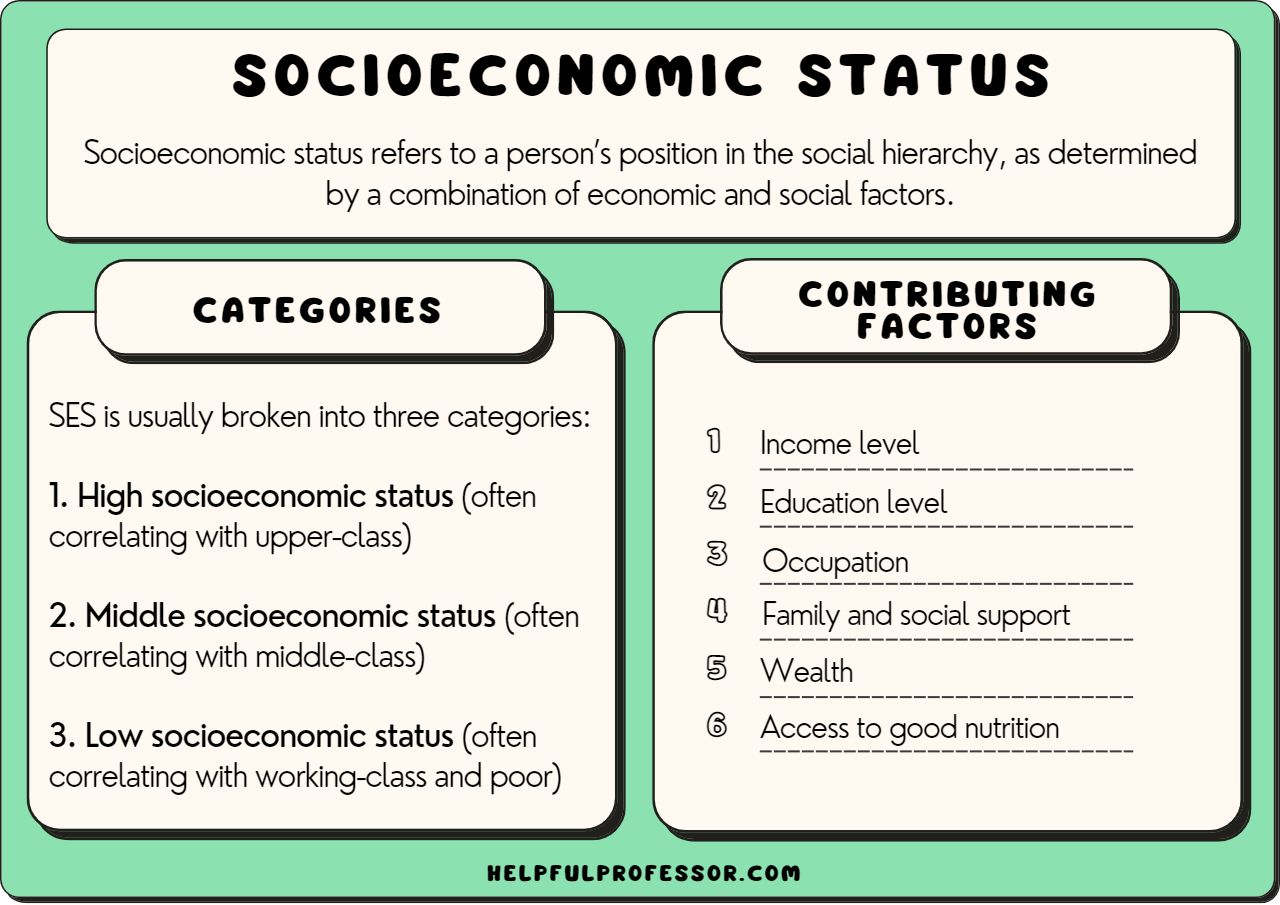
**PARTIAL ANALYSES:**

* Socio-economic assessments of climate change adaptation measures are often carried out as partial analyses (as is also the case in other areas). This means that only the effects that can be directly or indirectly linked to the measure are quantified and valued.
* This partial approach has the advantage of being well defined, the results are simple and easy to interpret, and it opens up for numerous different methods of calculation of consequences. The disadvantage is that we cannot be sure that what is optimal in the analysis of a measure is also optimal in general. This makes it more difficult to compare results across measures.

**SOCIO-ECONOMIC ANALYSIS IN FEASIBILITY STUDY**:

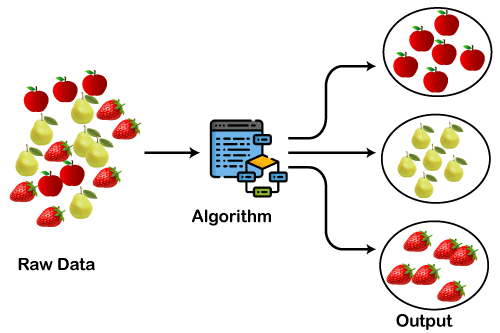
The socio-economic aspect of a project feasibility study involves **analyzing the broader social and economic implications of the project**. It aims to evaluate how the project will contribute to the socio-economic development of the area, address social needs, and enhance the quality of life for the community.

**SOCIO-ECONOMIC CHALLENGES:**

* Some of the most significant examples of socio-economic issues would include things like: **Income levels within a community**. The kind of educational opportunities that exist. The employment situation of a community. Safety within a community.
* The climate change adaptation area raises certain methodological issues.  These occur because calculations are made for very long time horizons, and because there is substantial uncertainty linked to the future effects. Socio-economic analyses in the climate change adaptation area are relatively new, both in Denmark and elsewhere. Methods are under continuous debate and development.  
  

**Cluster analysis**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more [similar](https://en.wikipedia.org/wiki/Similarity_measure) (in some sense) to each other than to those in other groups (clusters). It is a main task of [exploratory data analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis), and a common technique for [statistical](https://en.wikipedia.org/wiki/Statistics) [data analysis](https://en.wikipedia.org/wiki/Data_analysis), used in many fields, including [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), [image analysis](https://en.wikipedia.org/wiki/Image_analysis), [information retrieval](https://en.wikipedia.org/wiki/Information_retrieval), [bioinformatics](https://en.wikipedia.org/wiki/Bioinformatics), [data compression](https://en.wikipedia.org/wiki/Data_compression), [computer graphics](https://en.wikipedia.org/wiki/Computer_graphics) and [machine learning](https://en.wikipedia.org/wiki/Machine_learning).



**TYPES OF CLUSTERING** :

* The clustering methods are broadly divided into **Hard clustering** (datapoint belongs to only one group) and **Soft Clustering** (data points can belong to another group also). But there are also other various approaches of Clustering exist.
* Below are the main clustering methods used in Machine learning:

1. **Partitioning Clustering**
2. **Density-Based Clustering**
3. **Distribution Model-Based Clustering**
4. **Hierarchical Clustering**
5. **Fuzzy Clustering**

**Clustering Algorithms**

The Clustering algorithms can be divided based on their models that are explained above. There are different types of clustering algorithms published, but only a few are commonly used. The clustering algorithm is based on the kind of data that we are using. Such as, some algorithms need to guess the number of clusters in the given dataset, whereas some are required to find the minimum distance between the observation of the dataset.

Here we are discussing mainly popular Clustering algorithms that are widely used in machine learning:

1. **K-Means algorithm:** The k-means algorithm is one of the most popular clustering algorithms. It classifies the dataset by dividing the samples into different clusters of equal variances. The number of clusters must be specified in this algorithm. It is fast with fewer computations required, with the linear complexity of **O(n).**
2. **Mean-shift algorithm:** Mean-shift algorithm tries to find the dense areas in the smooth density of data points. It is an example of a centroid-based model, that works on updating the candidates for centroid to be the center of the points within a given region.
3. **DBSCAN Algorithm:** It stands **for Density-Based Spatial Clustering of Applications with Noise**. It is an example of a density-based model similar to the mean-shift, but with some remarkable advantages. In this algorithm, the areas of high density are separated by the areas of low density. Because of this, the clusters can be found in any arbitrary shape.
4. **Expectation-Maximization Clustering using GMM:** This algorithm can be used as an alternative for the k-means algorithm or for those cases where K-means can be failed. In GMM, it is assumed that the data points are Gaussian distributed.
5. **Agglomerative Hierarchical algorithm:** The Agglomerative hierarchical algorithm performs the bottom-up hierarchical clustering. In this, each data point is treated as a single cluster at the outset and then successively merged. The cluster hierarchy can be represented as a tree-structure.
6. **Affinity Propagation:** It is different from other clustering algorithms as it does not require to specify the number of clusters. In this, each data point sends a message between the pair of data points until convergence. It has O(N2T) time complexity, which is the main drawback of this algorithm.

**Applications of Clustering**

Below are some commonly known applications of clustering technique in Machine Learning:

* **In Identification of Cancer Cells:** The clustering algorithms are widely used for the identification of cancerous cells. It divides the cancerous and non-cancerous data sets into different groups.
* **In Search Engines:** Search engines also work on the clustering technique. The search result appears based on the closest object to the search query. It does it by grouping similar data objects in one group that is far from the other dissimilar objects. The accurate result of a query depends on the quality of the clustering algorithm used.
* **Customer Segmentation:** It is used in market research to segment the customers based on their choice and preferences.
* **In Biology:** It is used in the biology stream to classify different species of plants and animals using the image recognition technique.
* **In Land Use:** The clustering technique is used in identifying the area of similar lands use in the GIS database. This can be very useful to find that for what purpose the particular land should be used, that means for which purpose it is more suitable.

**Program:**

import pandas as pd

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

# Load your dataset

data = pd.read\_csv('your\_dataset.csv')

# Select relevant features

X = data[['industrial\_category', 'age']]

# Choose the number of clusters (k)

k = 3

# Apply K-Means clustering

kmeans = KMeans(n\_clusters=k)

data['cluster'] = kmeans.fit\_predict(X)

# Visualize the results

plt.scatter(data['industrial\_category'], data['age'], c=data['cluster'])

plt.xlabel('Industrial Category')

plt.ylabel('Age')

plt.title('Clustering Results')

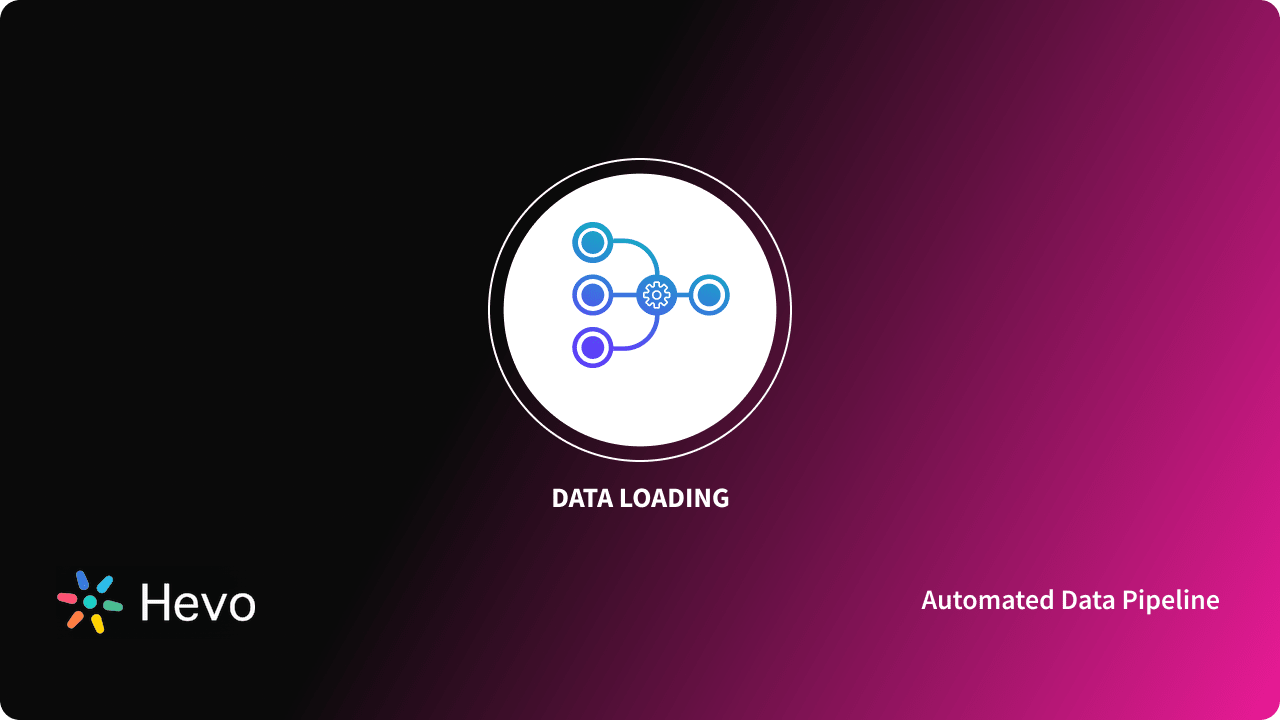
plt.show()

|  |  |  |
| --- | --- | --- |
| Industrial Category | Age | Clustering Results |
| LABOURS | 51 | 32 |
| WORKERS | 55 | 99 |

O

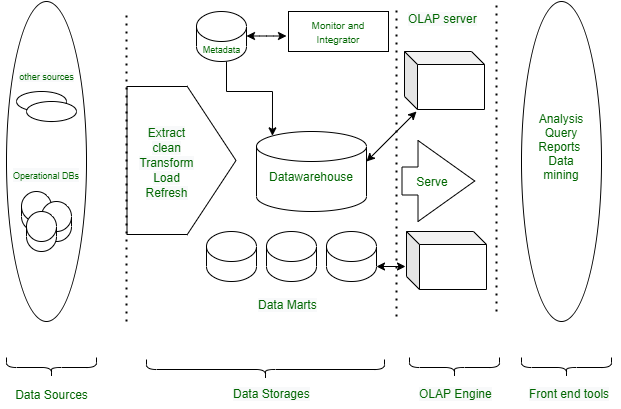
**DATA LOADING IN DATA SET:**

Data Loading is defined as **copying data from one electronic file or database into another**. Data loading implies converting from one format into another; for example, from one type of production database into a decision support database from a different vendor.



**TYPES OF LOADING IN DATA WAREHOUSE:**

* Initial Load: For the very first time loading all the data warehouse tables.
* Incremental Load: Periodically applying ongoing changes as per the requirement. ...
* Full Refresh: Deleting the contents of a table and reloading it with fresh data.
* Data are observations or measurements (unprocessed or processed) represented as text, numbers, or multimedia. A dataset is a structured collection of data generally associated with a unique body of work.
* The entire dataset is dumped, or loaded, and is then completely replaced (i.e. deleted and replaced) with the new, updated dataset.
* Truncating the target feature class table, thus removing all features but leaving the schema intact, followed by loading (or reloading, as the case may be) the data back into the table.



**DATA LOADING EXAMPLE PROGRAM:**

with open(‘sample.csv', 'r') as f:

# Read the column names

column\_names = f.readline().strip().split(',')

# Initialize a list to store the data

# It will hold row values

data = []

# Loop over the remaining lines in the file

for j in f:

{

j = j.strip()

values = j.split(',')

# Append the values to the data list

data.append(values)

}

# Create a dictionary with a list comprehension to hold the data

df = {column\_names[i]: [row[i] for row in data] for i in range(len(column\_names))}

# Create a dataframe from the dictionary

df = pd.DataFrame(df)

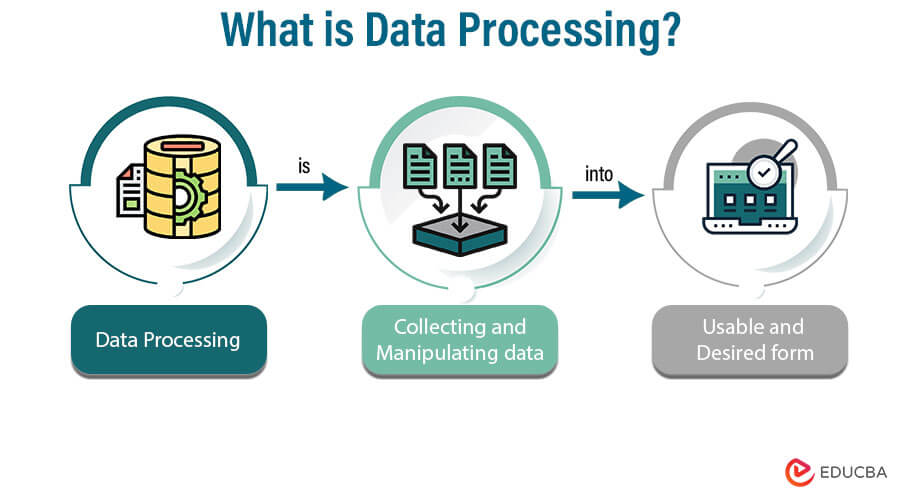
**OUTPUT:**

**DATA PROCESSING:**

* Data processing **occurs when data is collected and translated into usable information**.
* Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output.

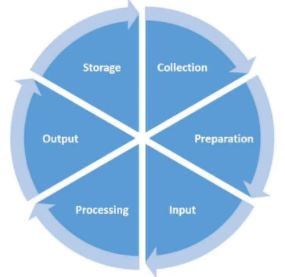
**STAGES OF DATA PROCESSING:**

* Data collection.
* Data input.
* Data processing.
* Data output.

****

**DATA PROCESSING CYCLE:**

The data processing cycle consists of **a series of steps where raw data (input) is fed into a system to produce actionable insights (output)**. Each step is taken in a specific order, but the entire process is repeated in a cyclic manner**.**



**TYPES OF DATA PROCESSING:**

* Manual Data Processing.
* Mechanical Data Processing.
* Electronic Data Processing.
* Batch Data Processing.
* Real-time Data Processing.
* Online Data Processing.
* Automatic Data Processing.

**DATA MANIPULATION:**

There are some python libraries that are useful for data scientists to do Data Manipulation, Machine Learning, Data Visualization, and Statistical Analysis. Libraries like **NumPy and pandas** offer powerful tools for manipulating data in CSV or Excel. Matplotlib offers charts and plots for visualization.

**DATA MANIPULATION LIBRARIES PYTHON IN PANDAS:**

 Installing Pandas.

 Creating DataFrame.

 Adding data in DataFrame using Append Function.

 Getting Shape and information of the data.

 Getting Statistical Analysis of Data.

 Dropping Columns from Data.

 Dropping Rows from Data

**PROGRAM:**

# Importing the pandas library

import pandas as pd

# creating a dataframe object

student\_register = pd.DataFrame()

# assigning values to the

# rows and columns of the dataframe

student\_register['Name'] = ['Abhijit','Smriti','Akash', 'Roshni']

student\_register['Age'] = [20, 19, 20, 14]

student\_register['Student'] = [False, True,True, False]

print(student\_register)

**OUTPUT:**

|  |  |  |  |
| --- | --- | --- | --- |
| 0  1  2  3 | Name  Abhijit  Smriti  Akash  Roshni | Age  20  19  20  14 | STUDENT  False  True  True  False |

**Demographic Analysis in Assessment of Marginal Workers in TamilNadu :**

Demographic analysis can be crucial when assessing the population of marginal workers in Tamil Nadu, India. To perform such an analysis, you would typically consider factors like:

1. Age: Analyzing the age distribution of marginal workers can reveal trends related to young or aging labor forces.

2. Gender: Understanding the gender distribution among marginal workers can help in identifying gender-based employment disparities.

3. Educational Level: Assessing the education levels of marginal workers can provide insights into skill levels and potential for job mobility.

4. Rural vs. Urban Distribution: Examining whether marginal workers are predominantly from rural or urban areas can inform policy and employment initiatives.

5. Income: Analyzing income levels of marginal workers can highlight income inequality and living standards.

6. Occupational Distribution: Studying the types of jobs marginal workers are engaged in can help tailor vocational training and support programs.

7. Migration Patterns: Investigating if there is significant migration of workers within or outside Tamil Nadu can inform labor market dynamics.

These demographic insights can guide policymakers, organizations, and government agencies in developing targeted programs to uplift and support marginal workers in Tamil Nadu. Data sources for this analysis typically include census data, labor surveys, and employment records.

**How to perform demographic analysis in Assessment of Marginal Workers in TamilNadu :**

Performing a demographic analysis to assess marginal workers in Tamil Nadu involves several steps. Here's a general guide on how to go about it:

1. Define Your Research Objectives:

   Clearly outline what you aim to achieve with your demographic analysis. Are you interested in understanding employment trends, improving job opportunities, or addressing specific issues faced by marginal workers?

2. Data Collection:

   Gather relevant data from various sources, including government census data, labor surveys, and research reports. Ensure the data is up-to-date and specific to Tamil Nadu.

3. Select Key Demographic Factors:

   Identify the demographic factors you want to analyze. As mentioned earlier, these can include age, gender, education, income, occupation, location (rural or urban), and more.

4. Data Cleaning and Preparation:

   Clean and organize the data. This may involve dealing with missing values, outliers, and ensuring data consistency.

5. Data Analysis:

   Utilize statistical and data analysis tools to examine the data. Create visualizations such as graphs and charts to present your findings effectively.

6. Interpret the Results:

   Analyze the data to draw meaningful insights. For example, you might find that a significant proportion of marginal workers in Tamil Nadu are young, female, and engaged in low-paying agricultural jobs.

7. Comparative Analysis:

   Compare your findings with historical data or with data from other regions to identify trends and anomalies.

8. Policy Implications:

   Use your analysis to suggest policy changes or interventions that could benefit marginal workers in Tamil Nadu. This might include educational programs, skills training, and employment initiatives.

9. Disseminate Findings:

   Share your results with relevant stakeholders, including government bodies, non-profit organizations, and academic institutions. Effective communication can help drive change and improvement.

10. Review and Update:

   Demographic analysis is an ongoing process. Continuously monitor and update your analysis to keep it relevant and effective in addressing the evolving needs of marginal workers.

It's important to note that this process may require expertise in data analysis and statistical tools. Collaborating with researchers or experts in the field can be beneficial. Additionally, adhering to ethical guidelines and ensuring data privacy is crucial when working with sensitive demographic data.

**Calculate the distribution of marginal workers based on age, industrial category, and sex using data aggregation and manipulation.**

To calculate the distribution of marginal workers based on age, industrial category, and sex, you can use a programming language like Python and a data manipulation library such as Pandas. Here's a simplified example program that demonstrates how to perform this task:

```python

import pandas as pd

# Sample data (replace with your actual data source)

data = {

    'Age': [25, 35, 45, 28, 40, 50, 30],

    'Industrial\_Category': ['Agriculture', 'Manufacturing', 'Services', 'Agriculture', 'Manufacturing', 'Services', 'Agriculture'],

    'Sex': ['Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male'],

}

# Create a DataFrame from the data

df = pd.DataFrame(data)

# Perform data aggregation

result = df.groupby(['Age', 'Industrial\_Category', 'Sex']).size().reset\_index(name='Count')

# Print the distribution

print(result)

```

In this program:

1. You define a sample dataset with columns for Age, Industrial Category, and Sex. Replace this with your actual dataset.

2. Create a Pandas DataFrame from the data.

3. Use the `groupby` method to group the data based on Age, Industrial Category, and Sex.

4. The `size()` method counts the number of occurrences in each group, and `reset\_index` resets the index to make it more readable.

5. Finally, you print the distribution, which will display the count of marginal workers by age, industrial category, and sex.

Make sure to replace the sample data with your actual dataset for a meaningful analysis. Additionally, you might want to load data from a CSV file, a database, or another source depending on your data availability.

**Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).**

To perform a demographic analysis of marginal workers in Tamil Nadu and create visualizations, you can use Python with libraries like Pandas for data manipulation and Matplotlib or Seaborn for creating visualizations. Here's a simplified program:

```python

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Sample data (replace with your actual data source)

data = {

    'Age': [25, 35, 45, 28, 40, 50, 30],

    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male'],

    'Education': ['High School', 'College', 'High School', 'High School', 'College', 'College', 'High School'],

    'Income': [3000, 4500, 2500, 3200, 4800, 4000, 2700],

}

# Create a DataFrame from the data

df = pd.DataFrame(data)

# Visualization 1: Age Distribution

plt.figure(figsize=(8, 6))

sns.histplot(data=df, x='Age', kde=True)

plt.title('Age Distribution of Marginal Workers')

plt.xlabel('Age')

plt.ylabel('Frequency')

plt.show()

# Visualization 2: Gender Distribution

plt.figure(figsize=(8, 6))

sns.countplot(data=df, x='Gender')

plt.title('Gender Distribution of Marginal Workers')

plt.xlabel('Gender')

plt.ylabel('Count')

plt.show()

# Visualization 3: Education Level

plt.figure(figsize=(8, 6))

sns.countplot(data=df, x='Education')

plt.title('Education Level of Marginal Workers')

plt.xlabel('Education')

plt.ylabel('Count')

plt.xticks(rotation=45)

plt.show()

# Visualization 4: Income Distribution

plt.figure(figsize=(8, 6))

sns.histplot(data=df, x='Income', kde=True)

plt.title('Income Distribution of Marginal Workers')

plt.xlabel('Income')

plt.ylabel('Frequency')

plt.show()

```

This program provides four visualizations based on age, gender, education level, and income of marginal workers. Make sure to replace the sample data with your actual dataset. You may want to load your data from a CSV file or another source.

Ensure you have the required Python libraries installed by running `pip install pandas matplotlib seaborn` if you haven't already.

**Conclusion :**

The assessment of marginal workers in Tamil Nadu reveals important insights into the labor force dynamics of the state. It highlights the significance of understanding and addressing the challenges faced by this vulnerable group. To conclude, policymakers should focus on implementing targeted strategies to improve the socio-economic conditions, access to education, and job opportunities for marginal workers in Tamil Nadu. This will not only uplift their standard of living but also contribute to the overall development of the state.