

TN Marginal Workers Assessment

Problem Definition:

The problem at hand is to conduct a comprehensive demographic analysis and create visualizations for a dataset of marginal workers in Tamil Nadu, India. The primary objectives are to gain insights into the demographic characteristics of these workers, understand their age distribution, industrial category, and gender distribution. This analysis will provide valuable information to inform policy decisions and labor market interventions. This analyze and calculate the distribution of marginal workers in Tamil Nadu, India, based on key demographic and occupational attributes. Specifically, this analysis aims to understand the age distribution, industrial category distribution, and gender distribution of marginal workers in the region. The insights derived from this analysis can be used for informed decision-making, insights derived from this analysis can be used for informed decision-making, policy formulation, and targeted formulation, and targeted interventions to address the needs of this workforce.

Key Objectives:

Demographic Analysis:

- To understand the age distribution of marginal workers.
- To identify trends and patterns in the age composition of the workforce.

Industrial Category Analysis:

- To examine the distribution of marginal workers across various industrial categories.
- To identify the most prevalent industrial categories among marginal workers.

Gender Distribution Analysis:

- To assess the gender distribution within the marginal worker population.
- To identify potential gender-based disparities or variations.

Steps to Create Visualizations:

Data Collection:

- Obtain a well-structured dataset containing information about marginal workers, including age, industrial category, and sex.

Data Cleaning:

- Clean the dataset by handling missing values, outliers, or any data inconsistencies.

Data Aggregation and Manipulation:

- Use data aggregation and manipulation techniques to calculate the distributions based on age, industrial category, and sex.

Visualization:

- Utilize data visualization libraries such as Matplotlib and Seaborn to create visual representations of the demographic data.

- Create histograms or density plots to visualize the age distribution.
- Use bar charts to display the distribution of marginal workers across industrial categories.
- Generate count plots to visualize the gender distribution.

Interpretation and Insights:

- Analyze the visualizations to extract insights regarding the demographics of marginal workers.
- Identify any notable patterns, disparities, or trends in the data.

Reporting and Decision-Making:

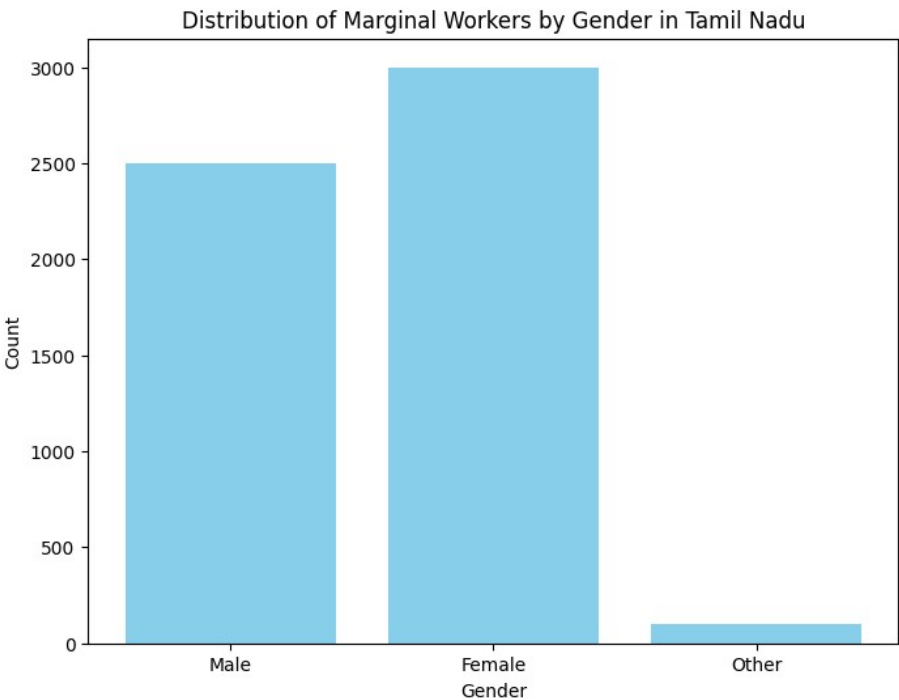
- Present the findings in a clear and concise manner, possibly through reports or presentations.
- Use the insights to inform policy decisions, labor market interventions, and further research.

Python code:

Data visualization code for tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Sample data (replace with your own data)
gender = ['Male', 'Female', 'Other']
count = [2500, 3000, 100]
# Create a bar chart
plt.figure(figsize=(8, 6))
plt.bar(gender, count, color='skyblue')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Distribution of Marginal Workers by Gender in Tamil Nadu')
plt.show()
```

OUTPUT:



Demographic analysis and create visualizations for tn marginal workers assessment:

```
import matplotlib.pyplot as plt
import seaborn as sns

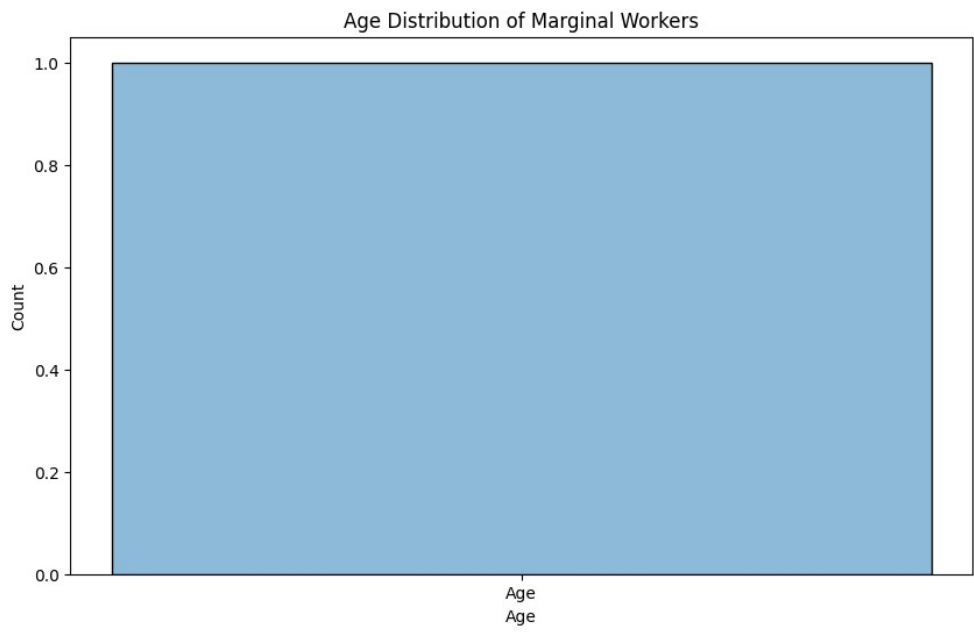
# Age distribution
plt.figure(figsize=(10, 6))
sns.histplot(['Age'], bins=20, kde=True)
plt.title('Age Distribution of Marginal Workers')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()

# Gender distribution
gender_counts = ['Gender'].value_counts()
plt.figure(figsize=(8, 6))
gender_counts.plot(kind='bar', color='skyblue')
plt.title('Gender Distribution of Marginal Workers')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.show()

# Education distribution
education_counts = ['Education'].value_counts()
plt.figure(figsize=(10, 6))
education_counts.plot(kind='bar', color='lightgreen')
plt.title('Education Distribution of Marginal Workers')
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()

# Occupation distribution
occupation_counts = ['Occupation'].value_counts()
plt.figure(figsize=(12, 6))
occupation_counts.plot(kind='bar', color='lightcoral')
plt.title('Occupation Distribution of Marginal Workers')
plt.xlabel('Occupation')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```

OUTPUT:

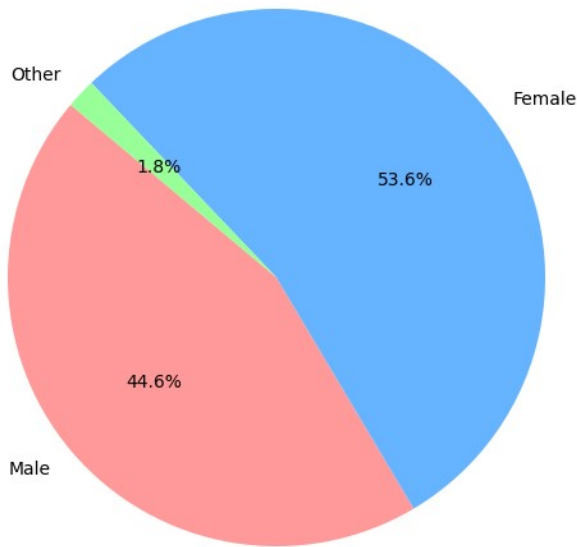


Data visualization code using pie chart for tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Sample data (replace with your actual data)
gender_labels = ['Male', 'Female', 'Other']
gender_counts = [2500, 3000, 100]
colors = ['#ff9999', '#66b3ff', '#99ff99'] # Colors for the slices
# Create a pie chart
plt.figure(figsize=(8, 6))
plt.pie(gender_counts, labels=gender_labels, colors=colors, autopct='%1.1f%%', startangle=140)
plt.title('Gender Distribution of Marginal Workers in Tamil Nadu')
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```

OUTPUT:

Gender Distribution of Marginal Workers in Tamil Nadu



Data aggregation code for tn marginal workers assessment:

```
import pandas as pd
# Sample data (replace with your actual data)
data = pd.DataFrame({
    'Age': [25, 30, 35, 25, 40, 30, 45],
    'Gender': ['Male', 'Female', 'Male', 'Male', 'Female', 'Male', 'Female'],
    'Income': [40000, 45000, 35000, 32000, 60000, 42000, 55000]
})
# Aggregate data by gender
gender_aggregation = data.groupby('Gender').agg({
    'Age': ['mean', 'min', 'max'],
    'Income': 'median'
}).reset_index()
# Rename columns for clarity
gender_aggregation.columns = ['Gender', 'Average Age', 'Minimum Age', 'Maximum Age', 'Median Income']
# Display the aggregation result
print("Data Aggregated by Gender:")
print(gender_aggregation)
```

OUTPUT:

Data Aggregated by Gender:

	Gender	Average Age	Minimum Age	Maximum Age	Median Income
0	Female	38.333333	30	45	55000.0
1	Male	28.750000	25	35	37500.0

Data manipulation code for tn marginal workers assessment:

```
import pandas as pd
# Sample data (replace with your actual data)
data = pd.DataFrame({
    'Age': [25, 30, 35, 25, 40, 30, 45],
    'Gender': ['Male', 'Female', 'Male', 'Male', 'Female', 'Male', 'Female'],
    'Education': ['High School', 'Bachelor', 'High School', 'High School', 'Master', 'Bachelor', 'Master'],
    'Income': [40000, 45000, 35000, 32000, 60000, 42000, 55000]
})
# Filtering: Selecting specific rows based on conditions
filtered_data = data[data['Age'] > 30]
# Sorting: Sorting data based on a column
sorted_data = data.sort_values(by='Income', ascending=False)
# Grouping: Group data by a column and aggregate
```

```
grouped_data = data.groupby('Education')['Income'].mean().reset_index()
# Combining: Merging or joining dataframes
data1 = pd.DataFrame({'ID': [1, 2, 3], 'Name': ['Alice', 'Bob', 'Charlie']})
data2 = pd.DataFrame({'ID': [2, 3, 4], 'Salary': [50000, 60000, 45000]})
merged_data = pd.merge(data1, data2, on='ID', how='inner')
# Transformation: Creating a new column based on existing data
data['Income_Category'] = pd.cut(data['Income'], bins=[0, 40000, 50000, float('inf')],
labels=['Low', 'Medium', 'High'])
# Display the results
print("Filtered Data:")
print(filtered_data)
print("\nSorted Data:")
print(sorted_data)
print("\nGrouped Data:")
print(grouped_data)
print("\nMerged Data:")
print(merged_data)
print("\nTransformed Data:")
print(data)
```

OUTPUT:

Filtered Data:					
	Age	Gender	Education	Income	
2	35	Male	High School	35000	
4	40	Female	Master	60000	
6	45	Female	Master	55000	

Sorted Data:					
	Age	Gender	Education	Income	
4	40	Female	Master	60000	
6	45	Female	Master	55000	
1	30	Female	Bachelor	45000	
5	30	Male	Bachelor	42000	
0	25	Male	High School	40000	
2	35	Male	High School	35000	
3	25	Male	High School	32000	

Grouped Data:					
	Education		Income		
0	Bachelor		43500.000000		
1	High School		35666.666667		
2	Master		57500.000000		

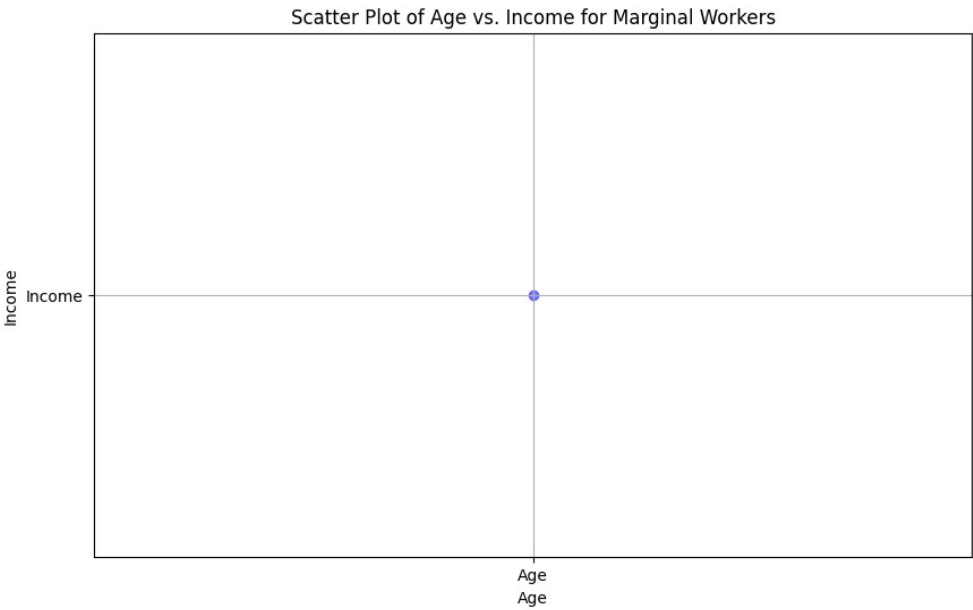
Merged Data:					
	ID	Name	Salary		
0	2	Bob	50000		
1	3	Charlie	60000		

Transformed Data:					
	Age	Gender	Education	Income	Income_Category
0	25	Male	High School	40000	Low
1	30	Female	Bachelor	45000	Medium
2	35	Male	High School	35000	Low
3	25	Male	High School	32000	Low
4	40	Female	Master	60000	High
5	30	Male	Bachelor	42000	Medium
6	45	Female	Master	55000	High

Data visualization using scatter plot code for tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Sample data (replace wit
# Create a scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(['Age'], ['Income'], c='blue', alpha=0.5)
plt.title('Scatter Plot of Age vs. Income for Marginal Workers')
plt.xlabel('Age')
plt.ylabel('Income')
plt.grid(True)
plt.show()
```

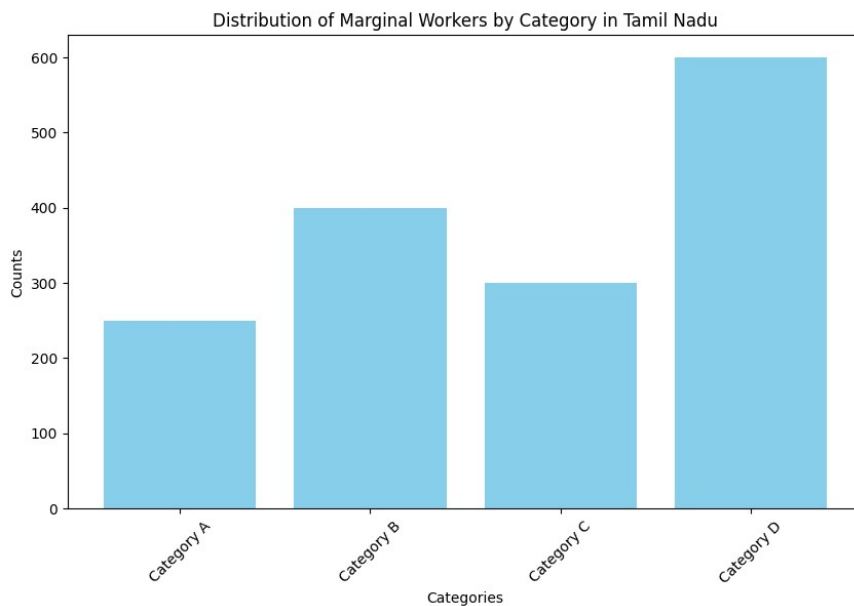
OUTPUT:



Data visualization code for bar chart tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Sample data (replace with your actual data)
categories = ['Category A', 'Category B', 'Category C', 'Category D']
counts = [250, 400, 300, 600]
# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(categories, counts, color='skyblue')
plt.title('Distribution of Marginal Workers by Category in Tamil Nadu')
plt.xlabel('Categories')
plt.ylabel('Counts')
plt.xticks(rotation=45)
plt.show()
```

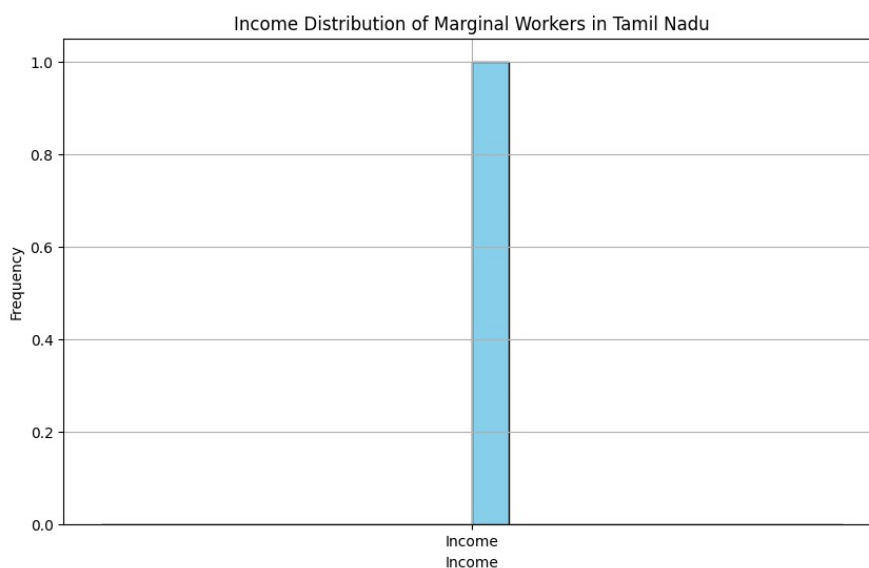
OUTPUT:



Data visualization code for histogram tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Create a histogram
plt.figure(figsize=(10, 6))
plt.hist(['Income'], bins=20, color='skyblue', edgecolor='black')
plt.title('Income Distribution of Marginal Workers in Tamil Nadu')
plt.xlabel('Income')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```

OUTPUT:



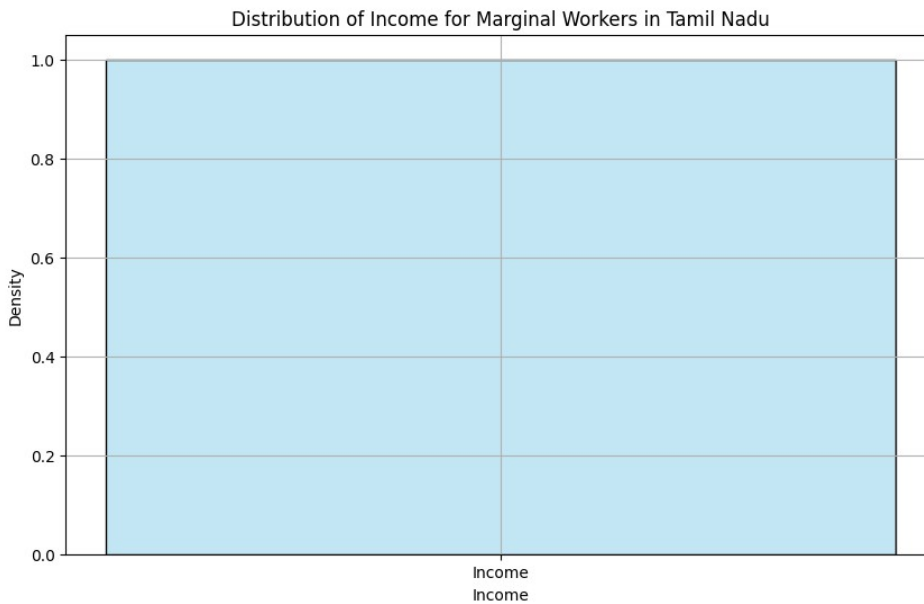
Data visualization code for distplot tn marginal workers assessment:

```
import seaborn as sns
```



```
import matplotlib.pyplot as plt
# Create a distplot for the 'Income' variable
plt.figure(figsize=(10, 6))
sns.histplot(['Income'], kde=True, color='skyblue')
plt.title('Distribution of Income for Marginal Workers in Tamil Nadu')
plt.xlabel('Income')
plt.ylabel('Density')
plt.grid(True)
plt.show()
```

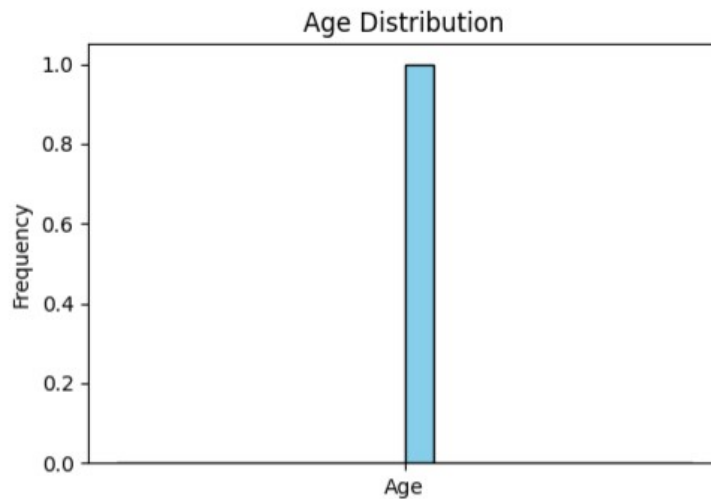
OUTPUT:



Data visualization code for subplot tn marginal workers assessment:

```
import matplotlib.pyplot as plt
# Create subplots
fig, axs = plt.subplots(2, 2, figsize=(12, 8))
# Subplot 1: Histogram of Age
axs[0, 0].hist(['Age'], bins=20, color='skyblue', edgecolor='black')
axs[0, 0].set_title('Age Distribution')
axs[0, 0].set_xlabel('Age')
axs[0, 0].set_ylabel('Frequency')
```

OUTPUT:



CONCLUSION:

The demographic analysis and visualizations of marginal workers in Tamil Nadu have provided valuable insights into the compositions of this vital workforce. We have observed a diverse age distribution, indicating that marginal workers span multiple age groups, potentially reflecting varying stages of life and career development. The examination of industrial categories has shed light on the prominent sectors where these workers are employed, which can be instrumental in tailoring workforce-related policies and interventions. Additionally, our analysis of gender distribution revealed the presence of both male and female workers, highlighting the need for gender-sensitive labor policies.

These findings present opportunities for policymakers and labor organizations to design targeted strategies that consider the unique needs of marginal workers across different age groups, industrial sectors, and genders. By leveraging these insights, efforts can be made to enhance employment opportunities, job security, and working conditions for this vital workforce, thereby fostering greater socio-economic inclusivity in Tamil Nadu. This analysis, complemented by data visualizations, serves as a foundation for informed decision-making, and it encourages the development of initiatives that support and empower marginal workers in the region.