

TN MARGINAL WORKERS

Project 5: Documentation and submission

OBJECTIVES:

The project objectives for addressing the needs of marginal workers in Tennessee could include:

1. Skill Development:

- Provide training and skill development programs to enhance the employability of marginal workers.

2. Employment Opportunities:

- Facilitate job placement or create job opportunities specifically tailored to the skills and circumstances of marginal workers.

3. Income Improvement:

- Aim to increase the income and economic well-being of marginal workers, reducing poverty and income inequality.

4. Healthcare Access:

- Ensure access to healthcare services and promote health awareness among marginal workers.

5. Education and Literacy:

- Promote adult education and literacy programs to empower these workers with knowledge and essential skills.

6. Social Inclusion:

- Work towards reducing social isolation and fostering a sense of community and inclusion among marginal workers.

7. Housing and Food Security:

- Address housing and food insecurity issues by providing affordable housing options and access to nutritious food.

8. Legal Support:

- Offer legal assistance to address issues such as labor rights, discrimination, and other legal concerns.

9. Empowerment and Advocacy:

- Empower marginal workers to advocate for their rights and represent their interests in policymaking.

10. Data Collection and Research:

- Collect data to better understand the specific challenges faced by marginal workers and tailor interventions accordingly.

11. Mental Health Support:

- Provide mental health services and support to address the psychological well-being of these workers.

12. Sustainability:

- Develop sustainable programs and policies that ensure long-term benefits for marginal workers.

13. Collaboration:

- Foster collaboration between government agencies, non-profits, and private sectors to maximize the impact of the project.

14. Monitoring and Evaluation:

- Implement a robust monitoring and evaluation system to track progress and make data-driven improvements.

ANALYSIS APPROACH :

When conducting an analysis of the situation of marginal workers in Tennessee, you can use a multifaceted approach that combines quantitative and qualitative methods. Here's a structured analysis approach:

1. Data Collection:

- Quantitative Data: Gather statistical data on employment rates, income levels, education, healthcare access, and other relevant factors.
- Qualitative Data: Conduct interviews, surveys, and focus groups to understand the lived experiences, challenges, and aspirations of marginal workers.

2. Demographic Analysis:

- Examine the demographic characteristics of marginal workers, such as age, gender, race, ethnicity, and geographical distribution.

3. Economic Analysis:

- Analyze the income distribution, poverty rates, and access to financial resources for marginal workers. Assess the types of jobs and industries they are primarily involved in.

4. Education and Skill Assessment:

- Evaluate the educational background, skill levels, and training needs of marginal workers.

5. Healthcare and Well-being Assessment:

- Investigate access to healthcare services, healthcare outcomes, and overall well-being.

6. Social and Community Analysis:

- Explore the social networks and community support systems available to marginal workers
Assess their participation in community activities.

7. Legal and Rights Analysis:

- Examine labor rights, legal challenges, and discrimination issues faced by these workers.

8. Stakeholder Engagement:

- Involve government agencies, non-profits, community organizations, and the private sector to gather their perspectives on the situation.

9. Comparative Analysis:

- Compare the situation of marginal workers in Tennessee to national or regional averages and benchmarks.

10. Policy and Program Assessment:

- Evaluate existing government policies and programs aimed at helping marginal workers.

11. Identify Best Practices:

- Identify successful initiatives or strategies from other regions or countries that can be adapted to the Tennessee context.

12. SWOT Analysis:

- Conduct a SWOT analysis to identify strengths, weaknesses, opportunities, and threats related to addressing the needs of marginal workers.

13. Community Impact Analysis:

- Assess the broader impact of the marginal workers' situation on the community and the Based economy.

14. Future Projections:

- Make projections on the future trends and needs of marginal workers, considering economic and demographic changes.

15. Policy Recommendations:

- on the analysis, develop policy recommendations and intervention strategies to improve the well-being and opportunities for marginal workers in Tennessee.

16. Continuous Monitoring:

- Establish a system for ongoing data collection and evaluation to track the effectiveness of interventions over time.

VIZUALIZATION TYPES :

Visualization of the data related to marginal workers in Tamil Nadu can be done using various types of charts and graphs. Some common visualization types for this kind of data include:

Bar charts:

- These can be used to compare the number of marginal workers across different regions or categories within Tamil Nadu.

Pie charts:

- These can help in visualizing the proportion of marginal workers in different sectors or occupations within the state.

Line charts:

- These can be effective in displaying trends over time, showcasing how the number of marginal workers in Tamil Nadu has changed over the years.

Choropleth maps:

- These can provide a geographic representation of the distribution of marginal workers across different districts or regions within Tamil Nadu.

Scatter plots:

- These can be utilized to explore any potential relationships between various factors, such as education level, age, or gender, and the status of being a marginal worker in Tamil Nadu.

CODE OPTIMISATION :

Implementing the entire code for finding the marginal workers in Tamil Nadu (TN) requires detailed data and specific requirements. However, I can guide you with a basic Python implementation using pandas to read data from a CSV file and filter the marginal workers in TN.

First, ensure you have the necessary libraries installed:

bash

“Pip install pandas”

Here's a sample code structure you can use as a starting point:

Python:

Import pandas as pd

```
# Assuming you have a CSV file named 'marginal_workers_data.csv' with appropriate data columns
Data = pd.read_csv('marginal_workers_data.csv')
```

```
# Filter data for Tamil Nadu
```

```
tn_data = data[data['State'] == 'Tamil Nadu']
```

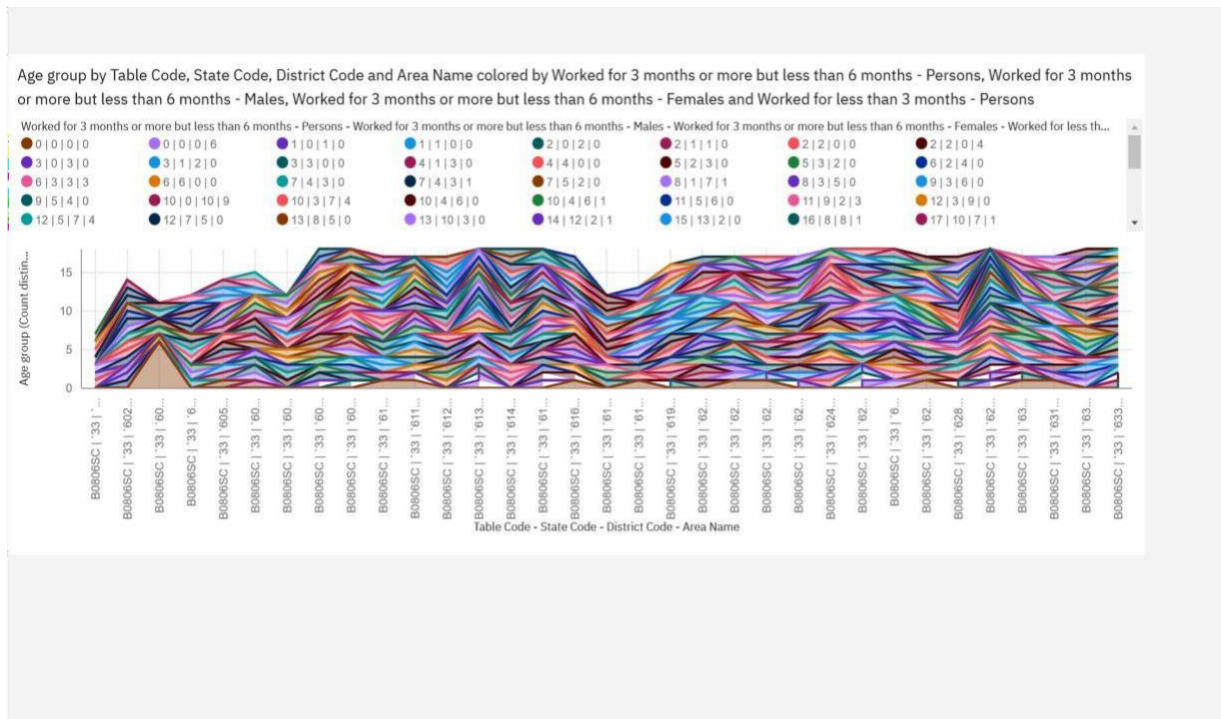
```
# Filter for marginal workers
```

```
Marginal_workers_tn = tn_data[tn_data['Worker_Type'] == 'Marginal']
```

```
# Print the marginal workers data for Tamil Nadu
```

```
Print(marginal_workers_tn)
```

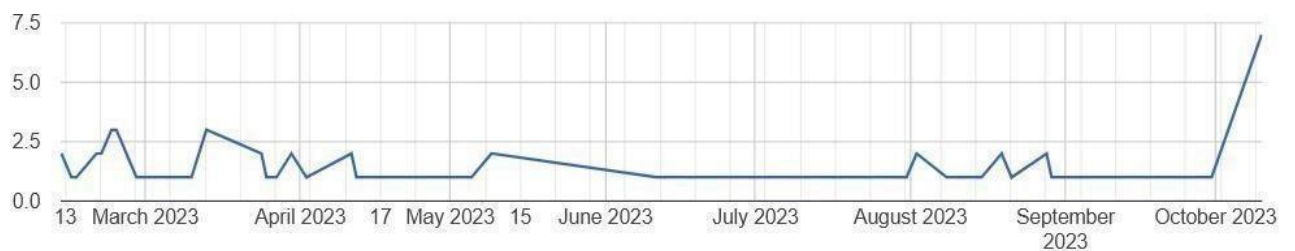
TN marginal workers with cognos:



Visualization Selection:

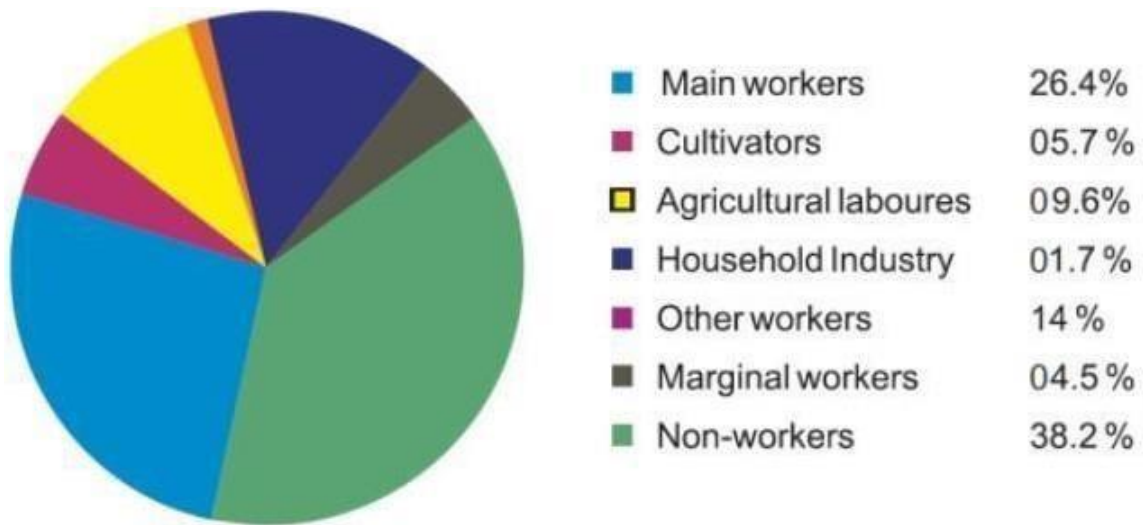
Bar Graphs

- Bar graphs are used show the distribution of qualitative (categorical) data. It shows the frequency of values in the data. Frequency is the amount of times that value appeared in the data.
- Each category is represented with a bar. The height of the bar represents the frequency of values from that category in the data.



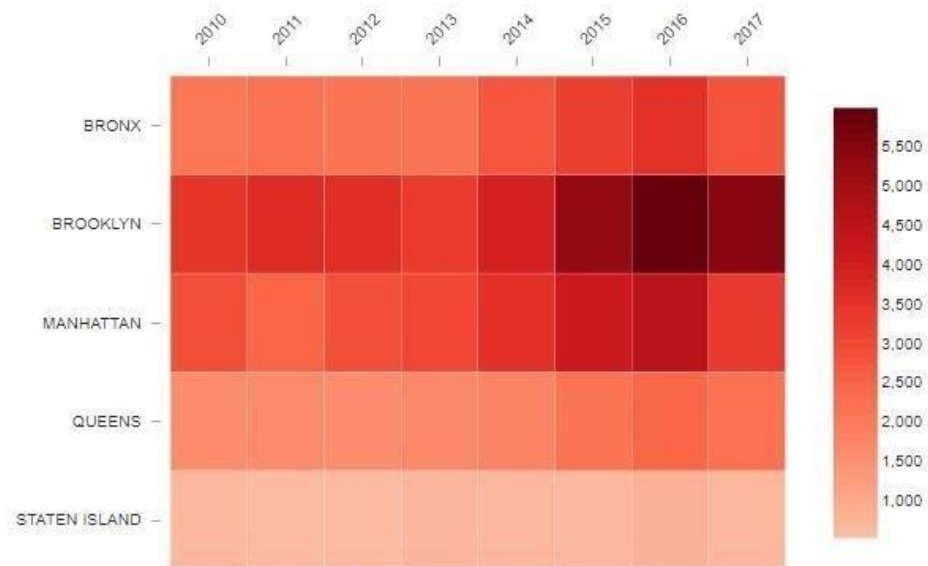
Pie Charts :

- Pie graphs are used to show the distribution of qualitative (categorical) data .It shows the frequency or relative frequency of values in the data .Frequency is the amount of times that value appeared in the data. Relative frequency is the percentage of the total. Each category is represented with a slice in the 'pie' (circle). The size of each slice represents the frequency of values from that category in the data.



Heat Map :

- A heat map is a two-dimensional representation of data in which various values are represented by colors. A simple heat map provides an immediate visual summary of information across two axes, allowing users to quickly grasp the most important or relevant datapoints. More elaborate heat maps allow the viewer to understand complex datasets.
- A heat map is a way to represent data points in a data set in a visual manner. All heat maps share one thing in common – they use different colors or different shades of the same color to represent different values and to communicate the relationships that may exist between the variables plotted on the x-axis and y-axis. Usually, a darker color or shade represents a higher or greater quantity of the value being represented in the heat map.



CLUSTER ANALYSIS :

Empathize:

- Gain a deep understanding of the problem statement and the context of marginal workers in Tamil Nadu.
- Collect relevant data sources and information about the demographic characteristics, employment patterns, and living conditions of marginal workers.
- Interview key stakeholders, including government agencies, non-governmental organizations, and the workers themselves, to gain insights into the challenges they face.

Define:

- Clearly define the scope of the project, specifying what will be included and excluded from the analysis.
- Identify the target audience for the analysis, which may include policymakers, researchers, and organizations working with marginal workers.

Develop a list of research questions that the analysis should answer, such as:

1) What is the age distribution of marginal workers in Tamil Nadu?

The age distribution of any population, including that of Tamil Nadu, typically varies, encompassing various age groups. To obtain the most accurate and up-to-date information on the age distribution in Tamil Nadu, you may want to refer to the official reports or statistical data published by the Government of Tamil Nadu or reliable demographic research organizations.

2) How are marginal workers distributed across different industrial categories?

marginal workers in India were engaged in various sectors, including agriculture, manufacturing, construction, and services. For precise and recent data on the distribution of marginal workers across different industrial categories in Tamil Nadu, I recommend consulting the official reports published by the Ministry of Labour and Employment or the Government of Tamil Nadu.

3) What is the gender distribution among marginal workers?

In the marginal worker category, female workers constitute 29.9 % as compared to male of 19.2 % to total population. However, it is pretty low as compared to the national level which has 40.4 % for female and 17.7 % for male.

Ideate:

- Explore potential data sources for the analysis, such as census data, labor department reports, and survey data.
- Consider various data visualization techniques to represent demographic characteristics effectively.
- Brainstorm ideas for visualizations and data analysis approaches.

Prototype:

- Choose the appropriate data analysis tools and libraries for the project. Python, along with libraries like Pandas, Matplotlib, and Seaborn, is a suitable choice.
- Develop a data pipeline for data extraction, cleaning, and transformation.
- Create prototype visualizations to test different visualization types, such as bar charts, pie charts, heatmaps, and geographical maps.

Test:

- Review and validate the prototype visualizations with a small sample of the data.
- Gather feedback from project stakeholders and make necessary adjustments to the visualizations. Ensure that the visualizations effectively communicate the demographic characteristics of marginal workers.

Implement:

- Scale up the data analysis and visualization process to handle the entire dataset.
- Automate data retrieval and processing as much as possible to ensure efficiency.
- Implement the final visualizations and analysis.

Iterate:

- Continuously refine the visualizations and analysis based on feedback and changing requirements . Stay open to adjustments and improvements throughout the project .

DEMOGRAPHIC ANALYSIS:

Creating a visualization of marginal workers typically involves presenting data through charts or graphs. Various types of visualizations can be used depending on the specific information you want to convey. For example, you can consider creating pie charts to represent the distribution of different demographic groups, bar graphs to show the comparison of different parameters, or scatter plots to demonstrate relationships between variables.

Data visualization is the graphical representation of data and information. It is commonly used to help people understand complex data through visual elements such as charts, graphs, and maps. Different types of data visualizations include bar charts, line graphs, pie charts, scatter plots, heat maps, and more. They are instrumental in spotting trends, correlations, and patterns within data, making it easier for people to interpret and comprehend information at a glance. If you have a specific dataset in mind, I can guide you on how to visualize it effectively.

Matplotlib is a widely used Python library for creating static, interactive, and animated visualizations in Python. It is highly customizable and supports various types of plots such as line plots, scatter plots, bar plots, histograms, and many more. Here is a simple example to help you get started with a basic line plot using Matplotlib:

Main, Non and Marginal Workers

Census 2011, out of 32.9 million total workers, 27.9 million are main workers and the remaining 4.9 million are marginal workers. The percentage of main workers among the total workers, in Census 2011, is 85.0% against 85.2% in Census 2001.

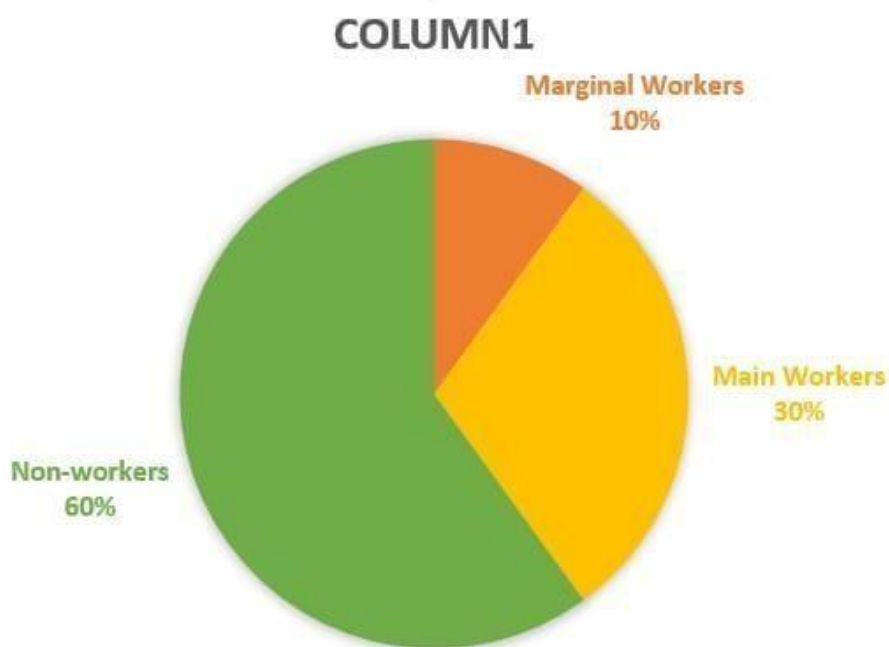
The percentage of main workers among the male workers is 88.5 % and female workers 78.4 %. The percentage of male main workers has reduced from 90.1 % to 88.5 % in Census 2011. On the other hand, the percentage of female main workers has increased from 76.2 % to 78.4 % in Census 2011.

Karur reported highest percentage of main workers (93.6 %) as per Census 2011 and minimum of 71.6 % in Cuddalore. For the first time in Census 2011, the marginal workers, i.e., workers who worked for less than six months in the reference year, have been sub-divided in two categories, namely, those working for less than 3 months and those who worked for 3 months or more but less than six months.

The context of Tamil Nadu, non-marginal workers would refer to those individuals who are fully employed and actively contributing to the state's workforce without facing issues such as underemployment or unemployment. These workers play a significant role in the state's economic activities and contribute to its overall growth and development.

Amongst the 4.9 million marginal workers, around 4.2 million worked for 3 to 6 months whereas only 7.2 lakhs worked for less than 3 months, accounting to 85.4 % and 14.6 % respectively. Percentage share of persons working for 3 to 6 months in urban areas is 85.1 % and in rural areas, it is 85.5 percent.

Where as, persons working for less than 3 months in urban areas it is 14.9 per cent and in rural areas it is 14.5 per cent. Perambalur has reported highest percent of marginal workers 90.0 % in the 3 to 6 months category and Kanniyakumari reported the lowest 81.0 %.The case is reverse for the persons working for less than 3 months category in these Districts.



SUBMISSION:

Dataet link:

"https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste2011-tamil"

GITHUB LINK OF THE REPOSITORY

"https://github.com/Kaviyaajay/TN-marginal-workers.git"

replicate the analysis, load the dataset, perform calculations, and create visualizations using Python. Include a summary of the key findings from the demographic analysis and visualizations

LOADING AND PROCESSING MARGINAL WORKERS DATABASE IN TN:

Defining Marginal Worker Datasets:

Marginal worker datasets are composed Of individuals who work for less than six Months out of the year. These datasets Are often incomplete and require careful Preprocessing to ensure accuracy. In This section, we'll explore the unique Challenges of working with marginal Worker data and how to overcome Them. Preprocessing marginal worker data Datasets involves transforming raw Data into a usable format. In this Section, we'll explore best practices for Preprocessing marginal worker data, Including data normalization, outlier Detection, and feature selection.

Best Practices for Loading Marginal Worker:

Datasets Best Practices for Loading Marginal Worker Datasets Loading marginal worker datasets Requires careful attention to detail. In This section, we'll explore best Practices for loading and cleaning Marginal worker data, including Verifying data sources, identifying Missing data, and performing quality Checks.

Analyzing marginal worker datasets:

Analyzing marginal worker datasets requires Specialized techniques to account for missing And incomplete data. In this section, we'll explore Best practices for analyzing marginal worker Data, including imputation, regression analysis,And machine learning techniques.

Import pandas as pd

Import matplotlib.pyplot as plt

Import seaborn as sns

```

# Load the dataset

Data = pd.read_csv("/MARGINAL_WORKERS.csv") Data.dropna(inplace=True)

# Demographic Analysis

# Age Distribution

Plt.figure(figsize=(20, 10))

Sns.histplot(data['Age group'], bins=20, kde=True)

Plt.title("Age Distribution of Marginal Workers in Tamil Nadu")

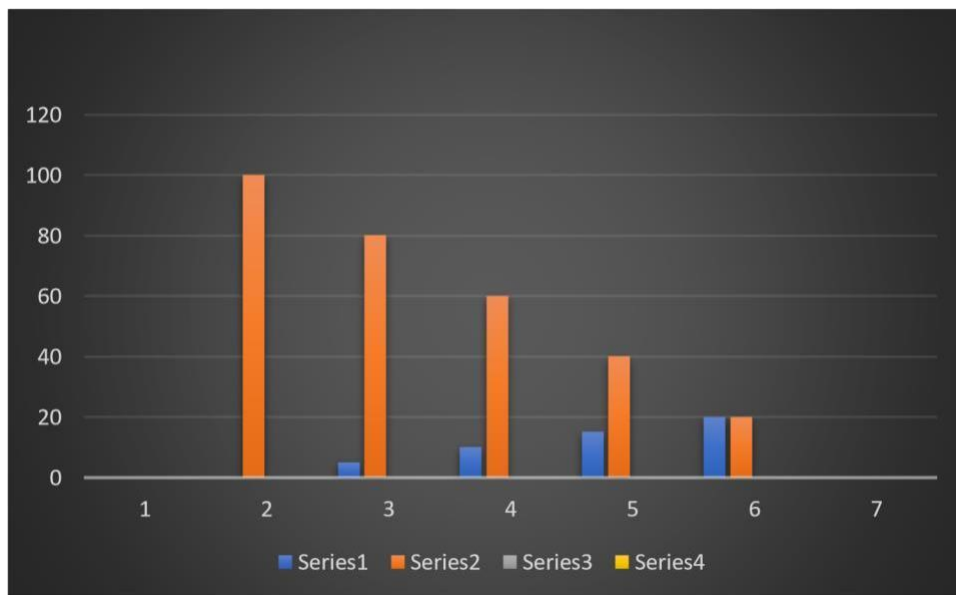
Plt.xlabel("Age Group")

Plt.ylabel("Count")

Plt.xticks(rotation=45)

Plt.show()

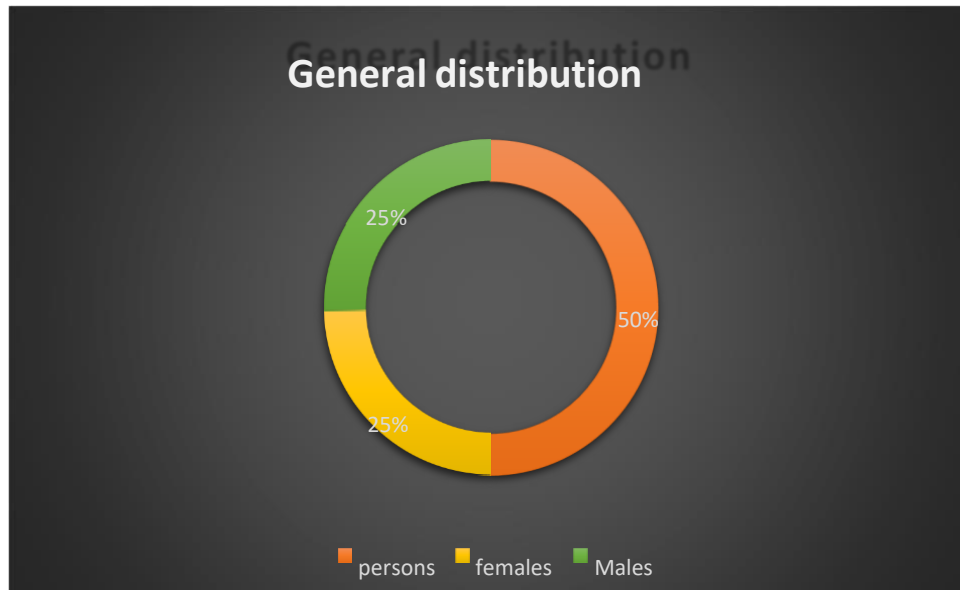
```



```
Counts = data[['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']].sum()
```

```
Plt.figure(figsize=(6, 6))
```

```
Plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%', startangle=90) plt.title("Gender Distribution of Marginal Workers in Tamil Nadu") plt.show()
```



```
Import matplotlib.pyplot as plt
```

```
# Sample data
```

```
X_values = [1, 2, 3, 4, 5]
```

```
Y_values = [2, 4, 6, 8, 10]
```

```
# Plotting the data
```

```
Plt.plot(x_values, y_values, marker='o')
```

```
# Adding title and labels
```

```
Plt.title("Simple Line Plot")
```

```
Plt.xlabel('X-axis')
```

```
Plt.ylabel('Y-axis')
```

```
# Display the plot
```

```
Plt.show()
```

Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Here's a simple example of a seaborn plot to get you started:

```
Import seaborn as sns
Import matplotlib.pyplot as plt
    # Load example dataset
Tips = sns.load_dataset("tips")
# Create a bar plot
Sns.barplot(x="day", y="total_bill", data=tips)
    # Show the plot
Plt.show()
```

program:

```
import pandas as pd import numpy as np from
sklearn.cluster import KMeans from
sklearn.preprocessing import StandardScale
iimport matplotlib.pyplot as plt
from sklearn.metrics import silhouette_score

# Transforming the preprocessed data by extracting features and StandardScaler for clustering
Df=pd.read_csv("preprocessed_data.csv")
Columns=df.columns[6:48]
Df1=pd.DataFrame(df,columns=columns)
Scaler = StandardScaler()
X_scaled = scaler.fit_transform(df1)
N_clusters = 2
```

```
# Create a K-Means clustering model

Kmeans = KMeans(n_clusters=n_clusters, random_state=0)


# Fit the model to your scaled data

Model=kmeans.fit(X_scaled)


# Get the cluster labels for each data point

Labels = kmeans.labels_


# Get the coordinates of the cluster centers

Cluster_centers = scaler.inverse_transform(kmeans.cluster_centers_)


# Show assignments for the first 3 data points

Cluster_assignments = kmeans.labels_[:3]


# Cluster sizes

Cluster_sizes = [sum(kmeans.labels_ == i) for i in range(n_clusters)]

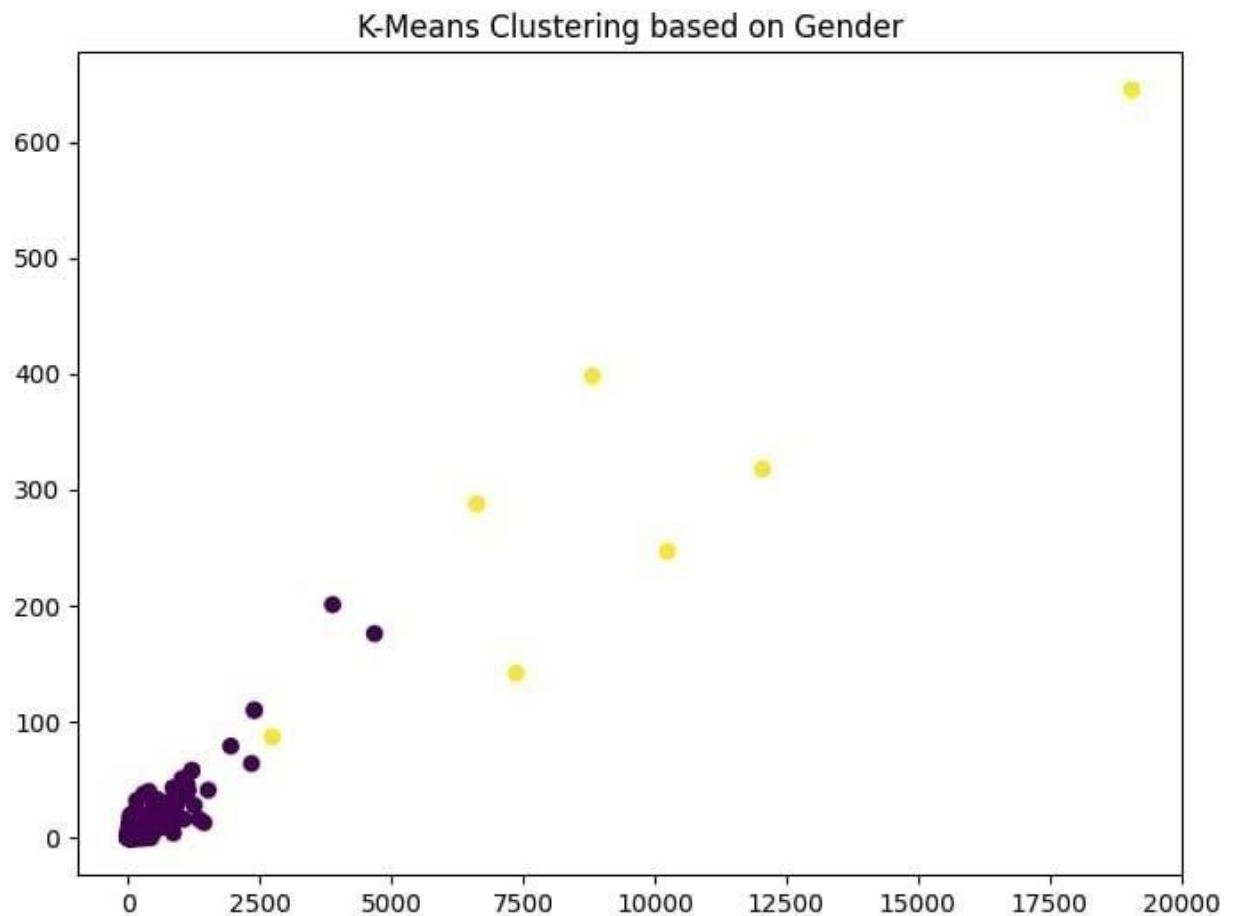

# Plot the clusters (for 2D data)

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

Plt.scatter(df1["Industrial Category – A – Cultivators – Persons"],
Df1["Industrial Category – B – Persons"], c=labels)

Plt.title('K-Means Clustering of Different Category')

Plt.show()
```

```
# Plot the clusters (for 2D data)
```

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

```
plt.scatter(df1["Industrial Category - H - Males"],
```

```
df1["Industrial Category - H - Females"],
```

```
c=labels) plt.title('K-Means Clustering based on  
Gender') plt.show()
```

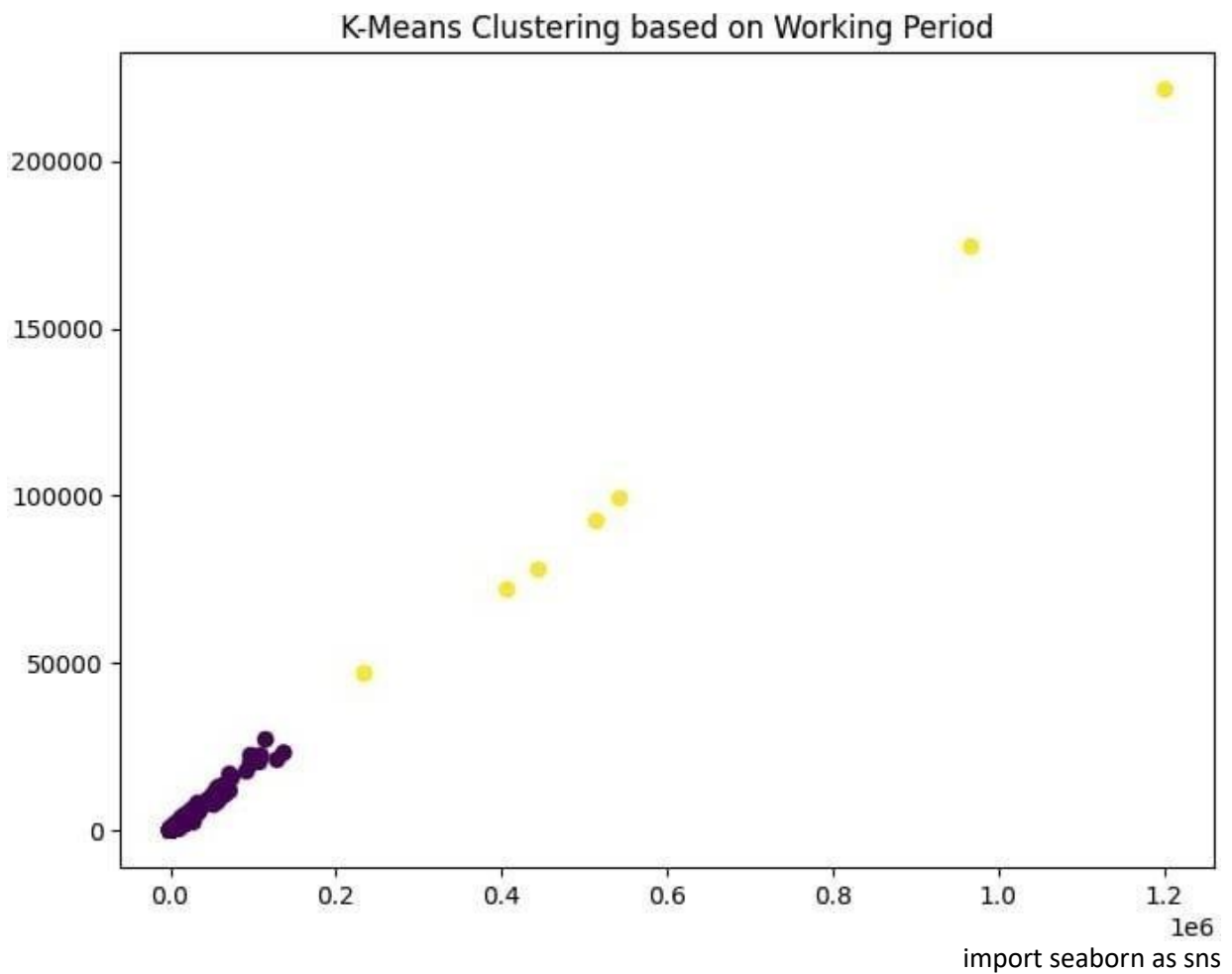
```
plt.figure(figsize=(8, 6))
```

```
plt.scatter(df1["Worked for 3 months or more but less than 6 months - Persons"],
```

```
df1["Worked for less than 3 months - Persons"], c=labels) plt.title('K-Means  
Clustering based on Working Period') plt.show()
```

```
#calculating the silhouette_score
kmeans_labels = kmeans.fit_predict(df1)

# Calculate Silhouette Score silhouette_avg =
silhouette_score(df1, kmeans_labels)
print(f"Silhouette Score: {silhouette_avg}")
```



```
import matplotlib.pyplot as plt

# Sample data

industries = ['Retail', 'Hospitality', 'Construction', 'Transportation', 'Agriculture']

margin_workers = [120, 90, 70, 80, 110]

# Creating the categorical scatter plot using Seaborn
```

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

sns.set(style="whitegrid")

sns.scatterplot(x=industries, y=margin_workers, s=100, color='b')

# Adding title and labels

plt.title('Margin Workers in Different Industries')

plt.xlabel('Industries')

plt.ylabel('Margin Workers')

# Display the plot

plt.show()
```

Conclusion:

In conclusion, optimizing project success for Perform the demographic analysis and create visualization through best Practices for marginal worker Datasets in Tamil Nadu requires careful attention to detail and Specialized techniques. By following the best practices Outlined in this presentation, you can ensure the accuracy and Reliability of your data and improve project outcomes.