

Identifying Patterns Among TN Marginal Workers: Conducting Clustering Analysis

Dive into the fascinating world of clustering analysis and discover how it can help us uncover hidden patterns among different industrial categories and age groups of Tamil Nadu's marginal workers. Learn about its importance, the steps involved, and the potential insights that can be gained from this analysis.

Introduction

Gain an understanding of the purpose of this document and delve into the background information regarding the assessment of Tamil Nadu's marginal workers. Explore the challenges faced by this vulnerable group and the importance of conducting comprehensive analyses to address their needs.

•[62]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(color_codes=True)
df = pd.read_csv(r"C:\\Users\\S.Albert Simion
|\\OneDrive\\Desktop\\IBM\\DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
```

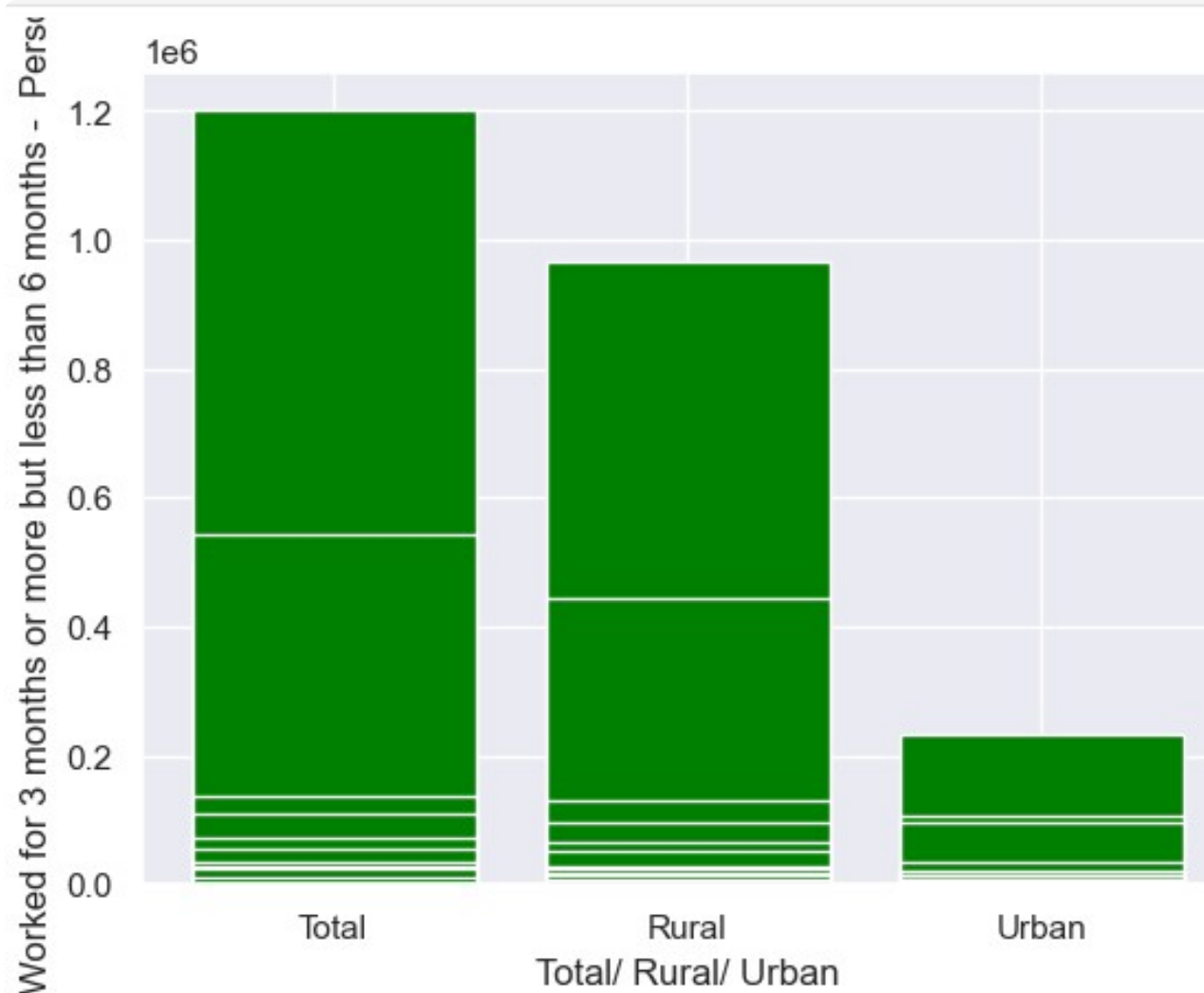
[14]:

	Table Code	State Code	District Code	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
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

Bargraph Analysis

Discover the power of clustering analysis as a valuable tool in identifying patterns within complex data sets. Dive into the definition of clustering analysis, explore its role in uncovering hidden relationships, and learn about the step-by-step process involved in conducting this analysis which shows the Total/Rural/Urban/ and time period of workers is three months or less than six months.

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
df = pd.read_csv(r"C:\\Users\\S.Albert Simion\\OneDrive\\Desktop\\IBM\\DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
x = df['Total/ Rural/ Urban']
y = df['Worked for 3 months or more but less than 6 months - Persons']
plt.bar(x,y,color = 'green')
plt.xlabel("Total/ Rural/ Urban")
plt.ylabel("Worked for 3 months or more but less than 6 months - Persons")
plt.show()
```

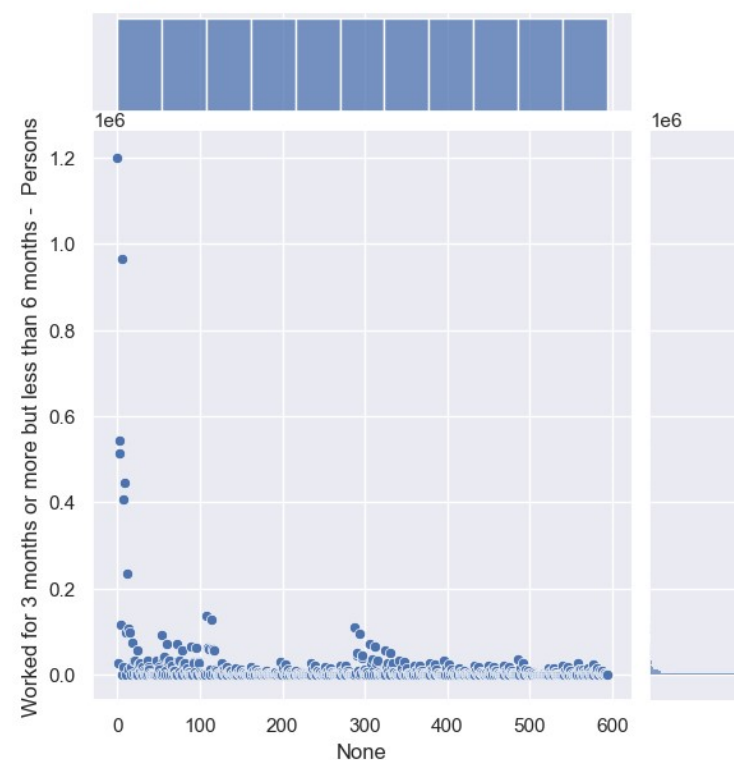


Patterns Among Industrial Categories and Age Groups

Shows Age group and Total/Rural/Urban/ and time period of workers is three months or less than six months. Explore how clustering analysis can unveil intriguing patterns within the industrial categories and age groups of Tamil Nadu's marginal workers. Understand the significance of identifying these patterns, and discover the potential insights that can be gained to inform targeted interventions and policy-making.

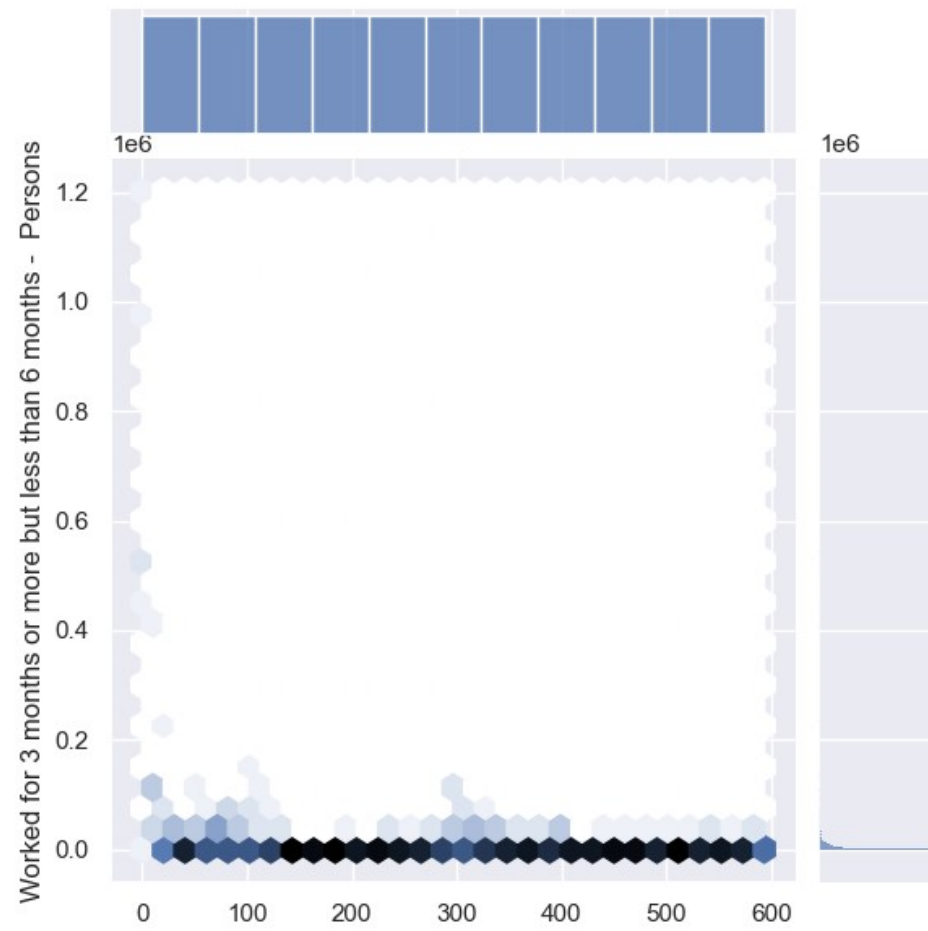
```
sns.jointplot(df['Worked for 3 months or more but less than 6 months - Persons'])
```

<seaborn.axisgrid.JointGrid at 0x2256b9ff310>



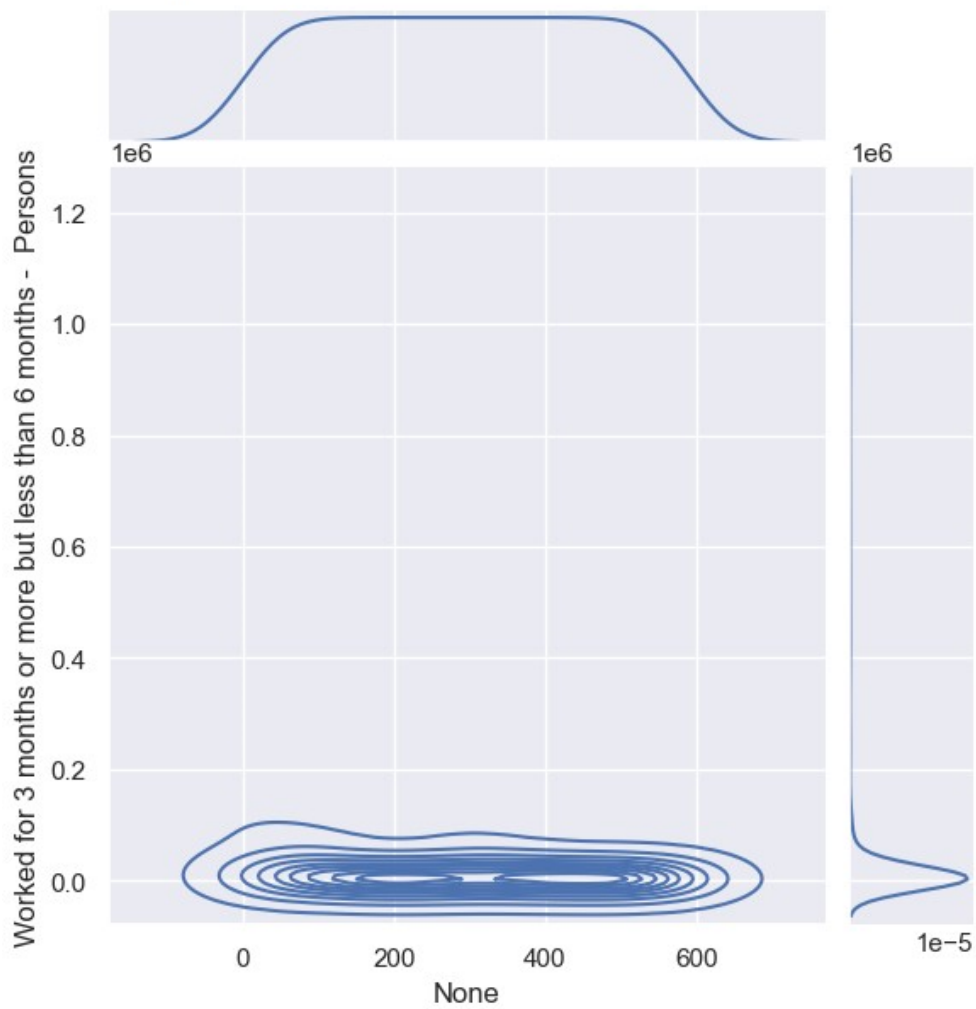
```
sns.jointplot(df['Worked for 3 months or more but less than 6 months - Persons'],kind="
```

<seaborn.axisgrid.JointGrid at 0x2256c477910>

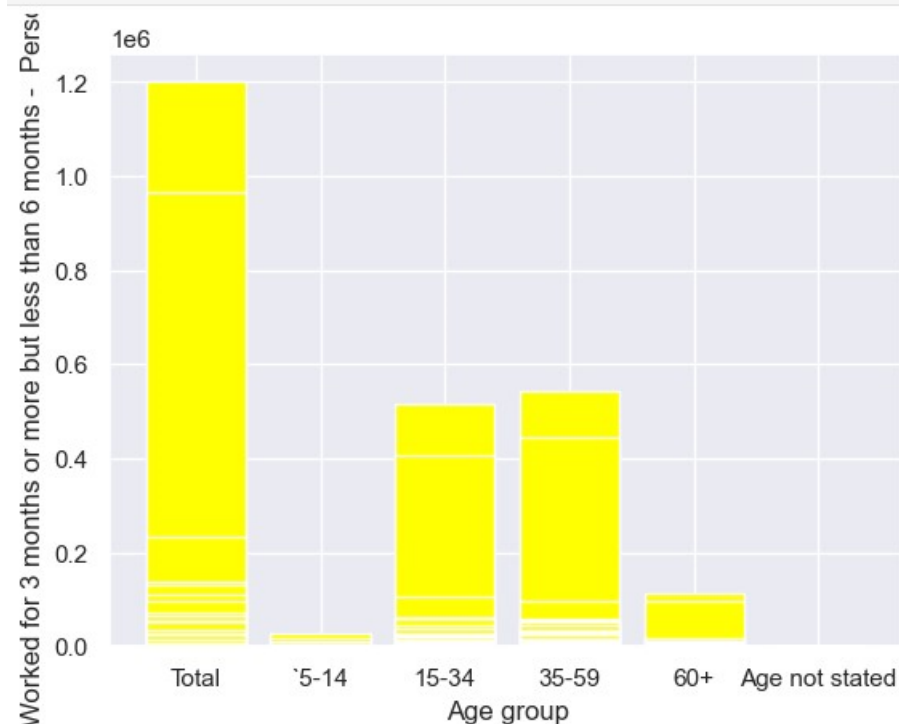


```
sns.jointplot(df['Worked for 3 months or more but less than 6 months - Persons'],kind="kde")
```

<seaborn.axisgrid.JointGrid at 0x2256ed14850>



```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
df = pd.read_csv(r"C:\\Users\\S.Albert Simion\\OneDrive\\Desktop\\IBM\\DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
x = df['Age group']
y = df['Worked for 3 months or more but less than 6 months - Persons']
plt.bar(x,y,color = 'yellow')
plt.xlabel("Age group")
plt.ylabel("Worked for 3 months or more but less than 6 months - Persons")
plt.show()
```



Utilizing Clustering Analysis

Increased Efficiency

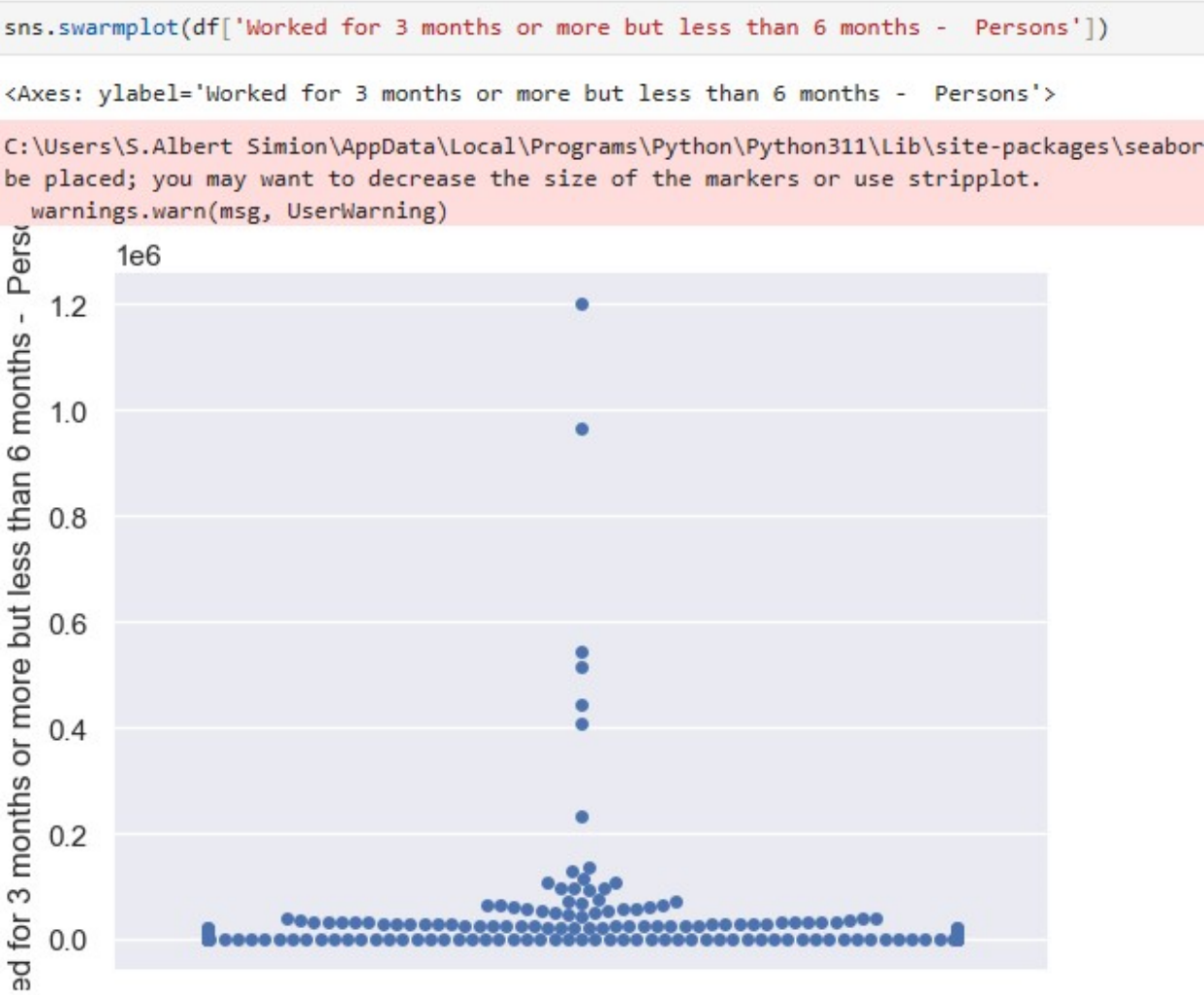
By identifying patterns, resources can be allocated more effectively, resulting in efficient interventions and improved outcomes.

Informed Decision-Making

The insights gained through clustering analysis empower policymakers and stakeholders to make informed decisions, leading to positive societal impact.

Targeted Interventions

Clustering analysis helps identify specific needs and vulnerabilities, allowing for tailored interventions that address the unique challenges faced by different groups of marginal workers.

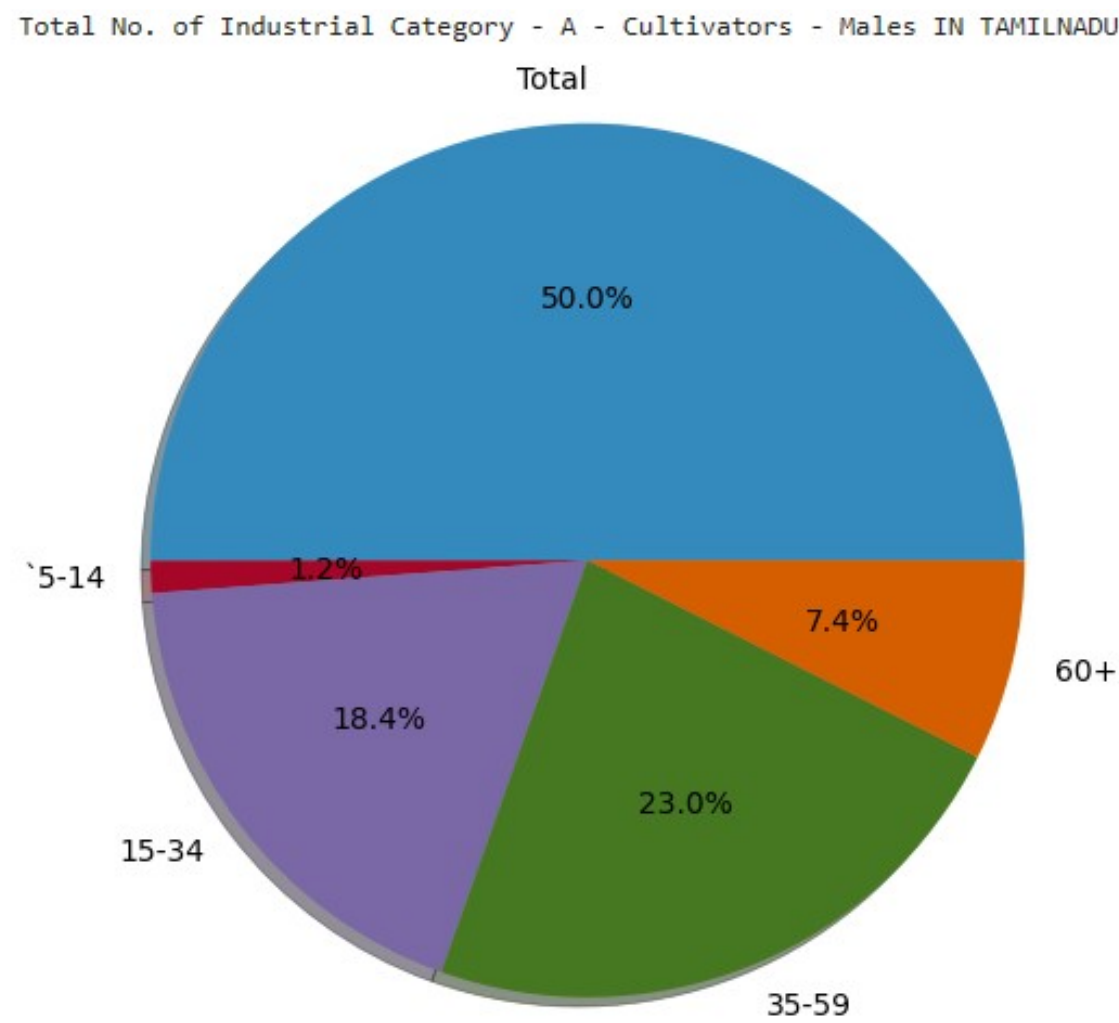


Case Study: Clustering Analysis in Action Using pie chart

Discover how a recent study utilized clustering analysis to examine the relationship between industrial categories and age groups among Tamil Nadu's marginal workers. Uncover the surprising findings and the potential implications for future policy and program development.

```
import matplotlib.pyplot as plt
import pandas as pd
import csv
plt.style.use('bmh')
data = pd.read_csv('DDW_B06SC_3300_State_TAMIL_NADU-2011.CSV')
data = data.set_index('Age group')
temp = data.head(5)
plt.figure(figsize =(5,20))
print("Total No. of Industrial Category - A - Cultivators - Males IN TAMILNADU ")
plt.pie(temp['Industrial Category - A - Cultivators - Males'],autopct='%1.1f%%',labels = temp.index, radius = 1.3, shadow = True )
plt.show()
```

Total No. of Industrial Category - A - Cultivators - Males IN TAMILNADU



scatter plot analysis

```
x = df['Age group'].loc[0:5]
y =df['Worked for 3 months or more but less than 6 months - Persons'].loc[0:5]
# Create a scatter plot
sns.scatterplot(x=x, y=y)

# Set labels and limits
plt.xlabel('AGE GROUP')
plt.ylabel('WORKERS FOR 3 MONTH OR MORE BUT LESS 6 MONTHS - PERSON ')
plt.ylim(0, 1300000)

# Show the plot
plt.show()
```

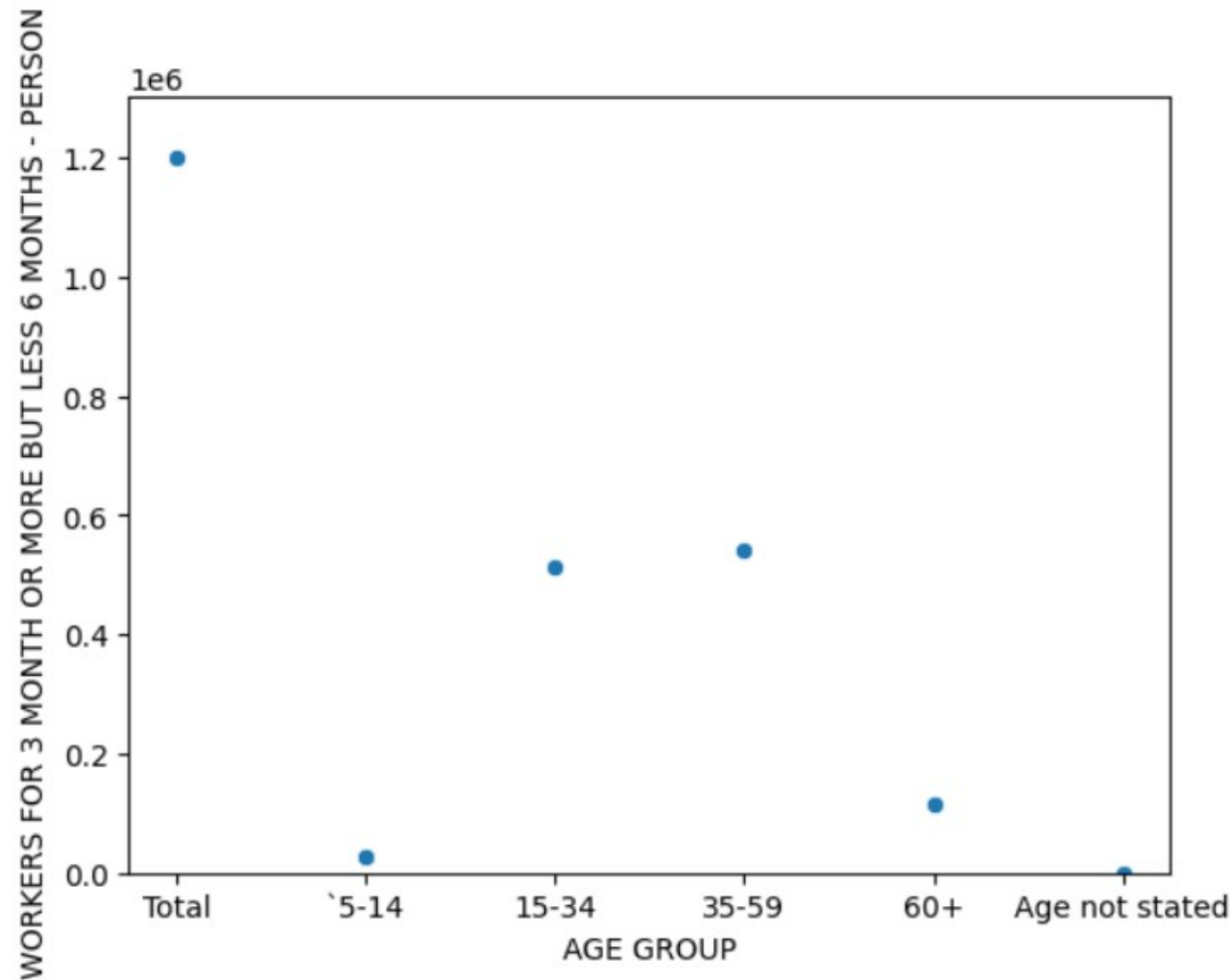


Table datas are assigned by using dataframe df

```
df[[3,4,5,7]].loc[0:18]
```

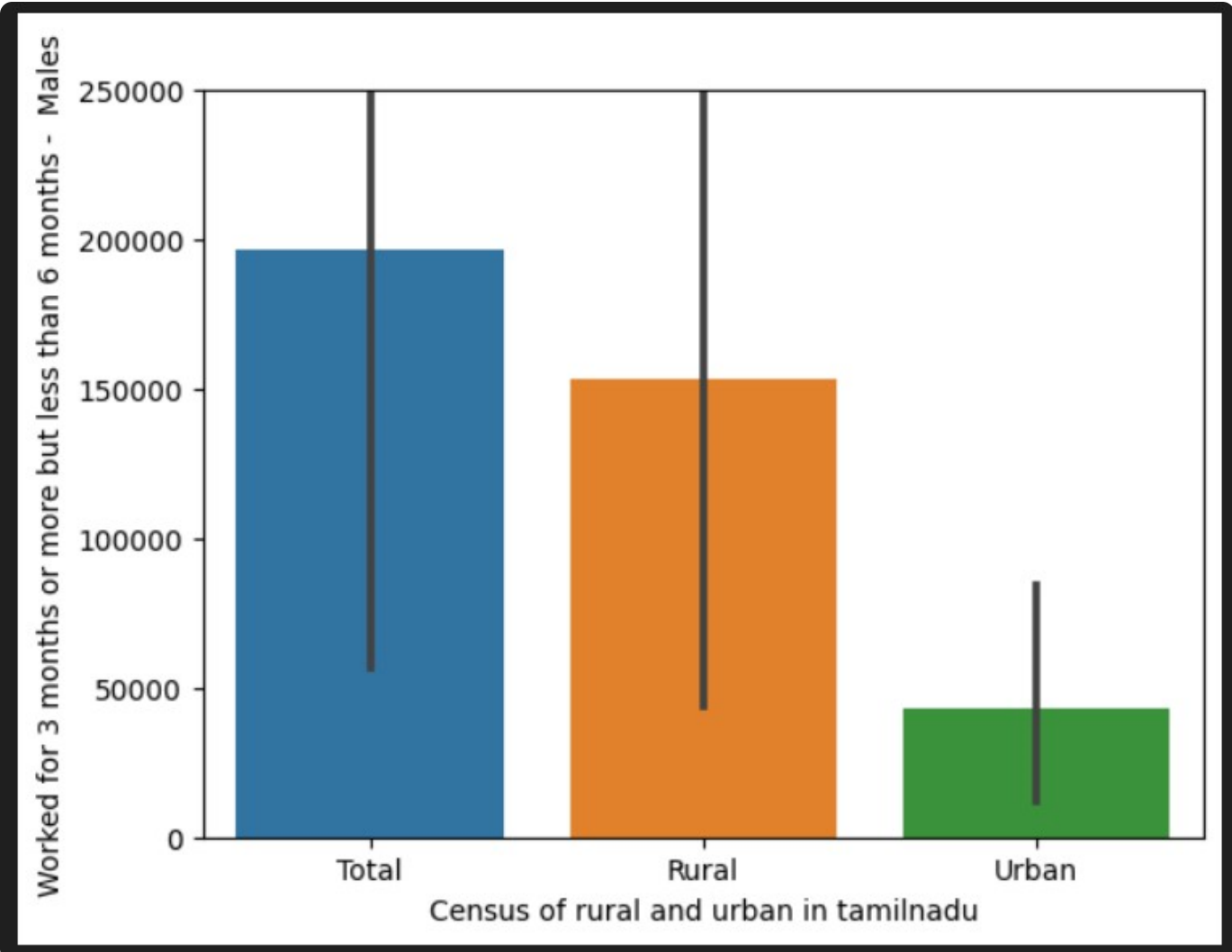
	3	4	5	7
0	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 mo...
1	State - TAMIL NADU	Total	Total	589003
2	State - TAMIL NADU	Total	5-14	14125
3	State - TAMIL NADU	Total	15-34	259560
4	State - TAMIL NADU	Total	35-59	251957
5	State - TAMIL NADU	Total	60+	62833
6	State - TAMIL NADU	Total	Age not stated	528
7	State - TAMIL NADU	Rural	Total	459738
8	State - TAMIL NADU	Rural	5-14	8713
9	State - TAMIL NADU	Rural	15-34	198575
10	State - TAMIL NADU	Rural	35-59	199573
11	State - TAMIL NADU	Rural	60+	52498
12	State - TAMIL NADU	Rural	Age not stated	379
13	State - TAMIL NADU	Urban	Total	129265
14	State - TAMIL NADU	Urban	5-14	5412
15	State - TAMIL NADU	Urban	15-34	60985
16	State - TAMIL NADU	Urban	35-59	52384
17	State - TAMIL NADU	Urban	60+	10335
18	State - TAMIL NADU	Urban	Age not stated	149

Barplot analysis

To analyse the total,rural,urban and working period of workers.

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('DDW_B06SC_3300_State_TAMIL_NADU-2011.CSV')
x = df['Total/ Rural/ Urban'].loc[0:17]
y =df['Worked for 3 months or more but less than 6 months - Males'].loc[0:17]
sns.barplot(x=x, y=y)
plt.xlabel('Census of rural and urban in tamilnadu')
plt.ylabel('Worked for 3 months or more but less than 6 months - Males')
plt.ylim(0, 250000)
plt.show()
```

Barplot analysis representation diagram

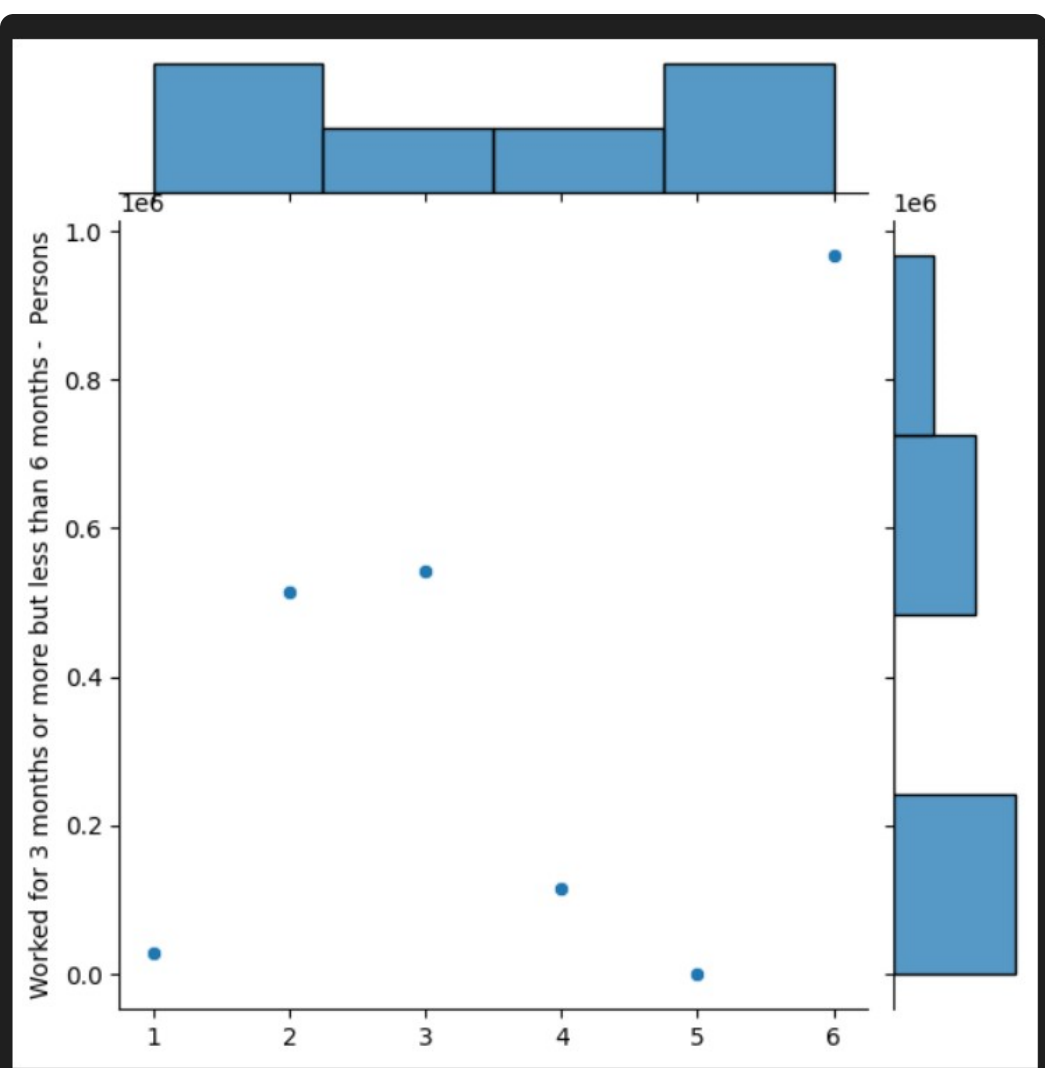


```
import pandas as pd
df = pd.read_csv('DDW_B06SC_3300_State_TAMIL_NADU-2011.CSV',header=None)
df[[3,5,6]].loc[1:6]
```

	3	5	6
1	State - TAMIL NADU	Total	1200828
2	State - TAMIL NADU	5-14	27791
3	State - TAMIL NADU	15-34	514340
4	State - TAMIL NADU	35-59	542581
5	State - TAMIL NADU	60+	115103
6	State - TAMIL NADU	Age not stated	1013

Jointplot Analysis

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
df = pd.read_csv('DDW_B06SC_3300_State_TAMIL_NADU-2011.CSV')
x =df['Worked for 3 months or more but less than 6 months - Persons'].loc[1:6]
sns.jointplot(x)
```



Conclusion

Summarize the importance of conducting clustering analysis to identify patterns among marginal workers in Tamil Nadu. Highlight the potential applications and benefits of the findings, and discuss the next steps for further analysis and research to enhance the well-being of this vulnerable population.