# **Tamil Nadu Marginal Workers**

# **Development 2.0**

**Phase :04 Project Number: 03**

**Introduction:**

In this assignment we looking for the very important topic in the data analysis course. Demographic analysis is a very important aspect in data analysis category. It is a process of studying the characteristics of data. It is creating visualization like pictures and graph.

**Project Objectives:**

* Perform the Demographic analysis
* Perform visualization using TN-Marginal Dataset
* Calculate the distribution of marginal workers based on age, industrial category and sex using data aggregation and manipulation
* Create visualization using data visualization libraries (e.g., Matplotlib and Seaborn)

**Data Set:**

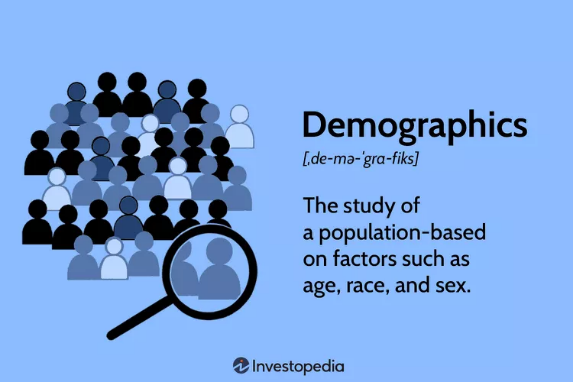
Here we are using the data set from the govt organization official site(tn.data.gov.in).

This dataset from: <https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil>

This dataset contains 71 columns and 595 rows.

**Demographic Analysis:**

Demographic analysis is a process of studying and understanding the characteristics of a population. It helps us learn more about the people who make up a particular group, community, or society. By examining various aspects such as age, gender, income, education, and more, we can gain valuable insights into a population's composition and how it is changing over time.



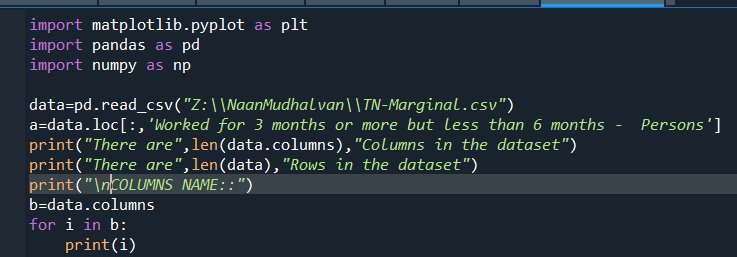
One important aspect of demographic analysis is creating visualizations. Visualizations are like pictures or graphs that represent data in a way that is easy to understand. They help us see patterns and trends within the population data. In this process, we'll use charts, graphs, and maps to illustrate the information we gather, making it simpler for everyone to grasp and interpret. This helps decision-makers, researchers, and the general public to make informed choices and understand the dynamics of a population.

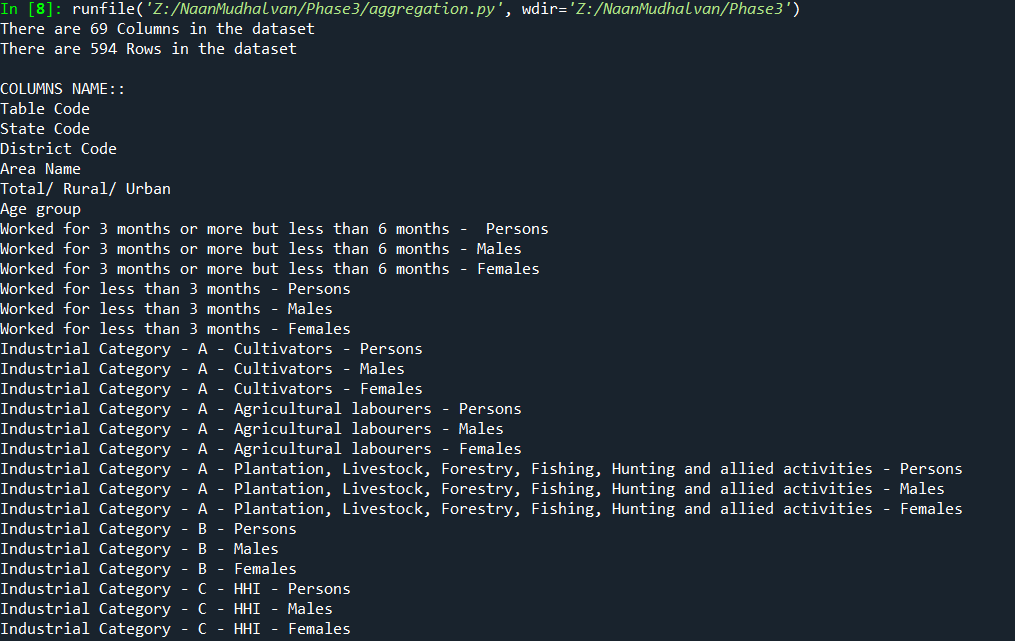
**“The Study of a Population based on factors such as age, language and sex”**

**Aggregation functions :**

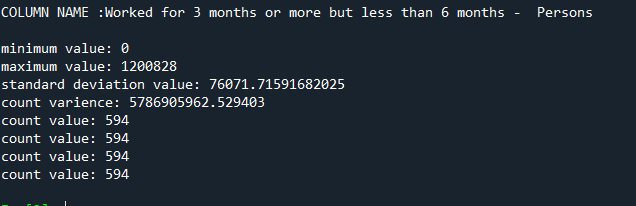
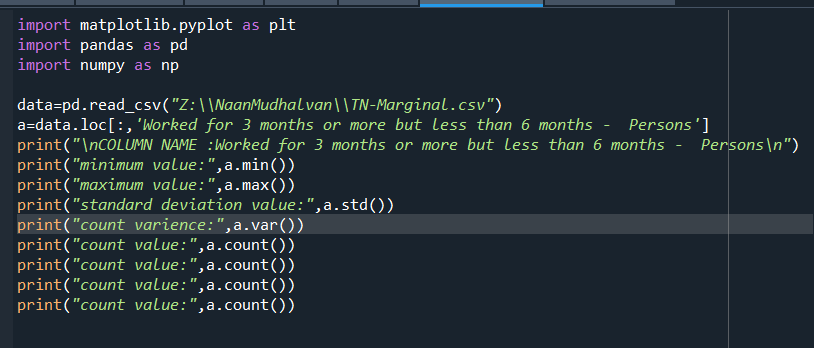
Pandas is a powerful Python library for data manipulation and analysis. It provides data structures and functions for working with structured data, making it an essential tool for data scientists and analysts. Here are some of the key features and concepts related to Pandas:

The detail about dataset:





Common aggregation functions (min , max):



Statistical analysis:

print(df.describe())

output:

Worked for 3 months or more but less than 6 months - Persons ... Industrial Category - R to U - Non HHI - Females

count 5.940000e+02 ... 594.000000

mean 1.617277e+04 ... 892.754209

std 7.607172e+04 ... 3988.125301

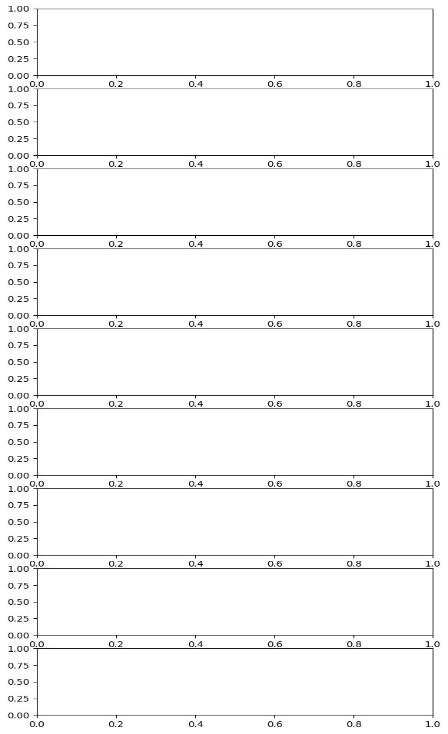
min 0.000000e+00 ... 0.000000

25% 2.872500e+02 ... 30.500000

50% 2.225500e+03 ... 135.000000

75% 9.628500e+03 ... 500.000000

max 1.200828e+06 ... 66287.000000

[8 rows x 63 columns]

Check the outliers:

fig, axs = plt.subplots(9,1,dpi=95, figsize=(7,17))

i = 0

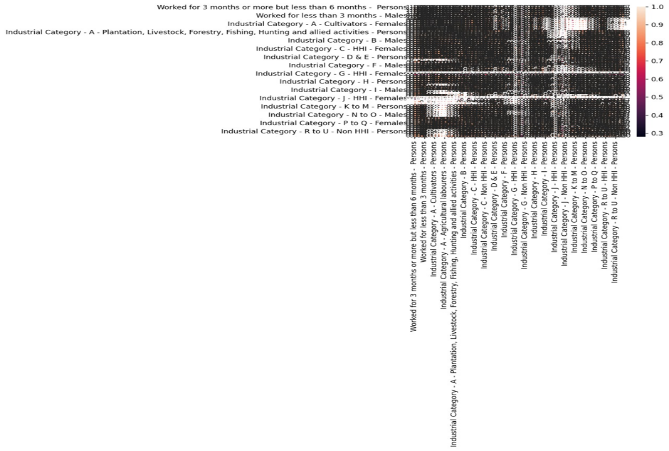
for col in df.columns:

axs[i].boxplot(df[col], vert=False)

axs[i].set\_ylabel(col)

i+=1

plt.show()



Correlation:

corr = df.corr()

plt.figure(dpi=130)

sns.heatmap(df.corr(), annot=True, fmt= '.2f')

plt.show()

Missing value:

print(df.isnull().sum())

output:

Table Code 0

State Code 0

District Code 0

Area Name 0

Total/ Rural/ Urban 0

...

Industrial Category - R to U - HHI - Males 0

Industrial Category - R to U - HHI - Females 0

Industrial Category - R to U - Non HHI - Persons 0

Industrial Category - R to U - Non HHI - Males 0

Industrial Category - R to U - Non HHI - Females 0

Length: 69, dtype: int64

**Visualization using matplotlib and seaborn:**

Here we use the TN-Marginal workers dataset, further analyzing the data we show some examples of visualization following.

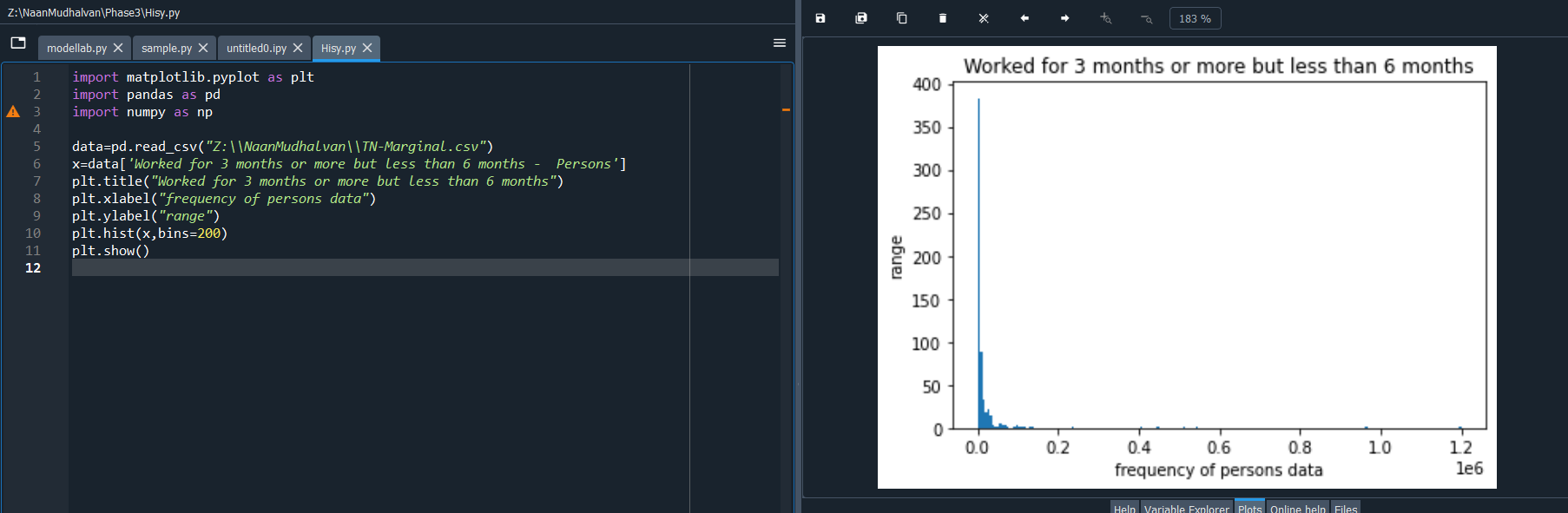
Worked Range Analysis:

That category satisfying more visualization tool, but here we choose the histogram plot,

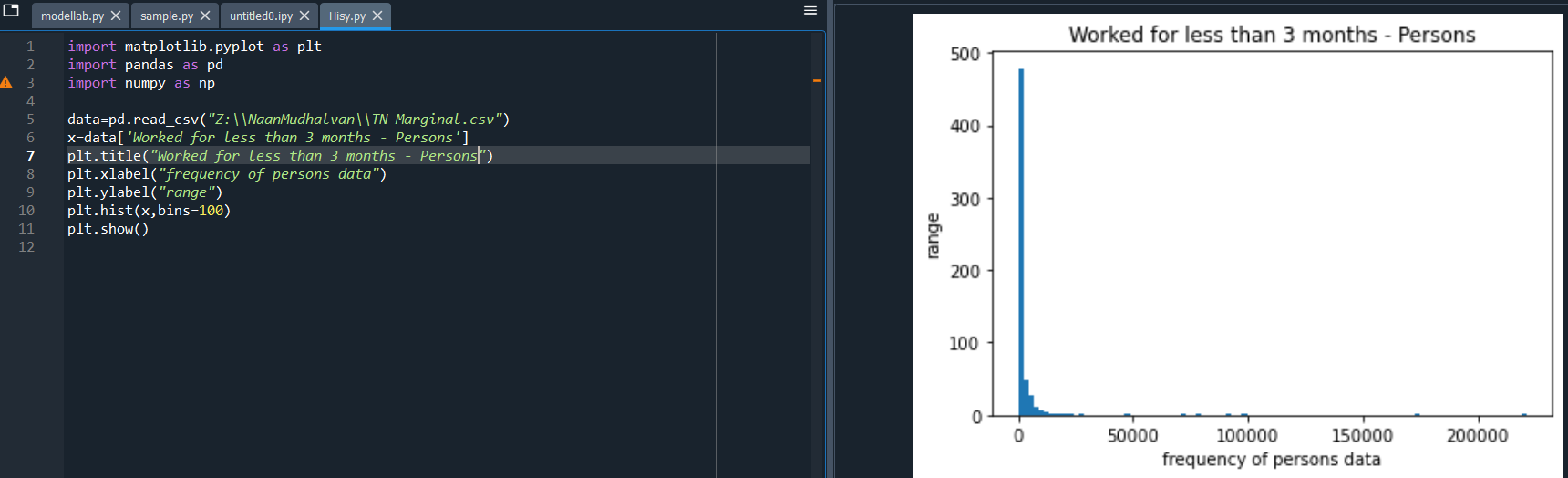
The hist is a function of matplotlib library, it require two mandatory data

1. X -data
2. Bins-range/interval

* Worked for 3 months or more but less than 6 months



* Worked for less than 3 months – Persons

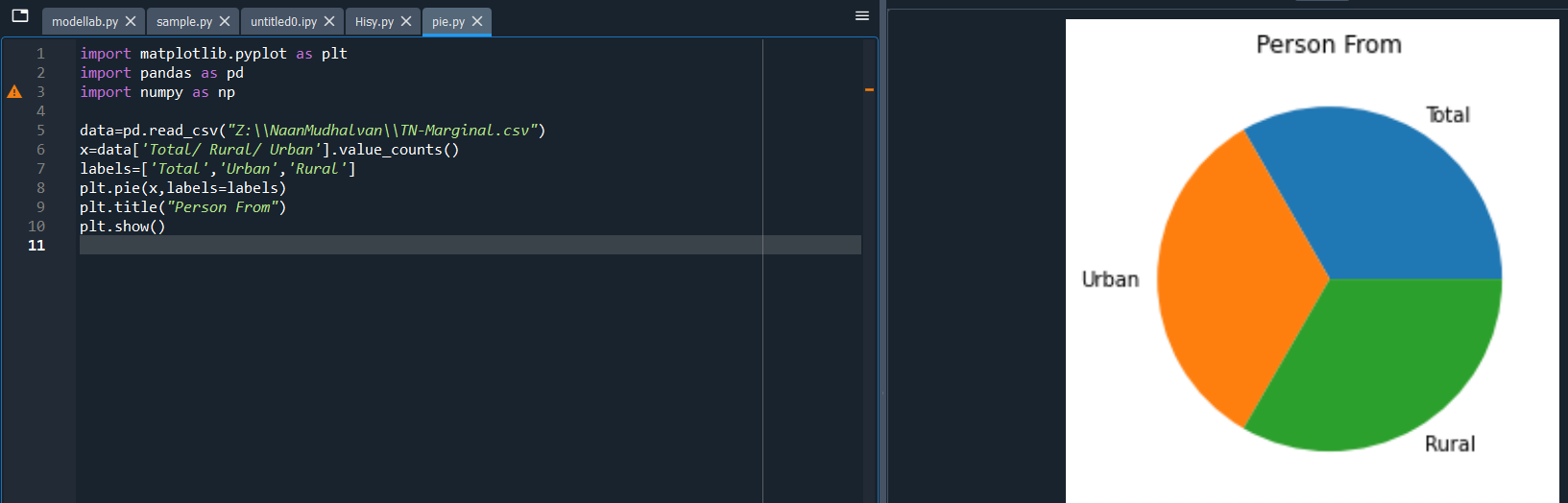


Person From:

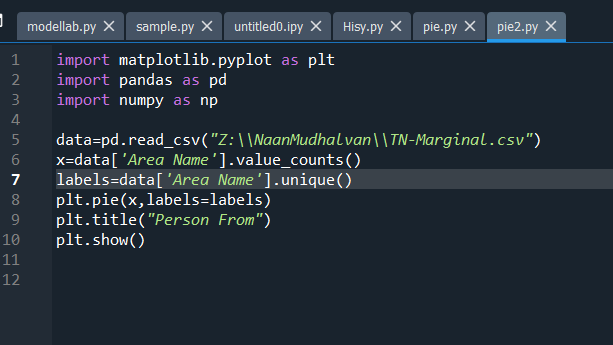
That category satisfying more visualization tool, but here we choose the Pie Chart,

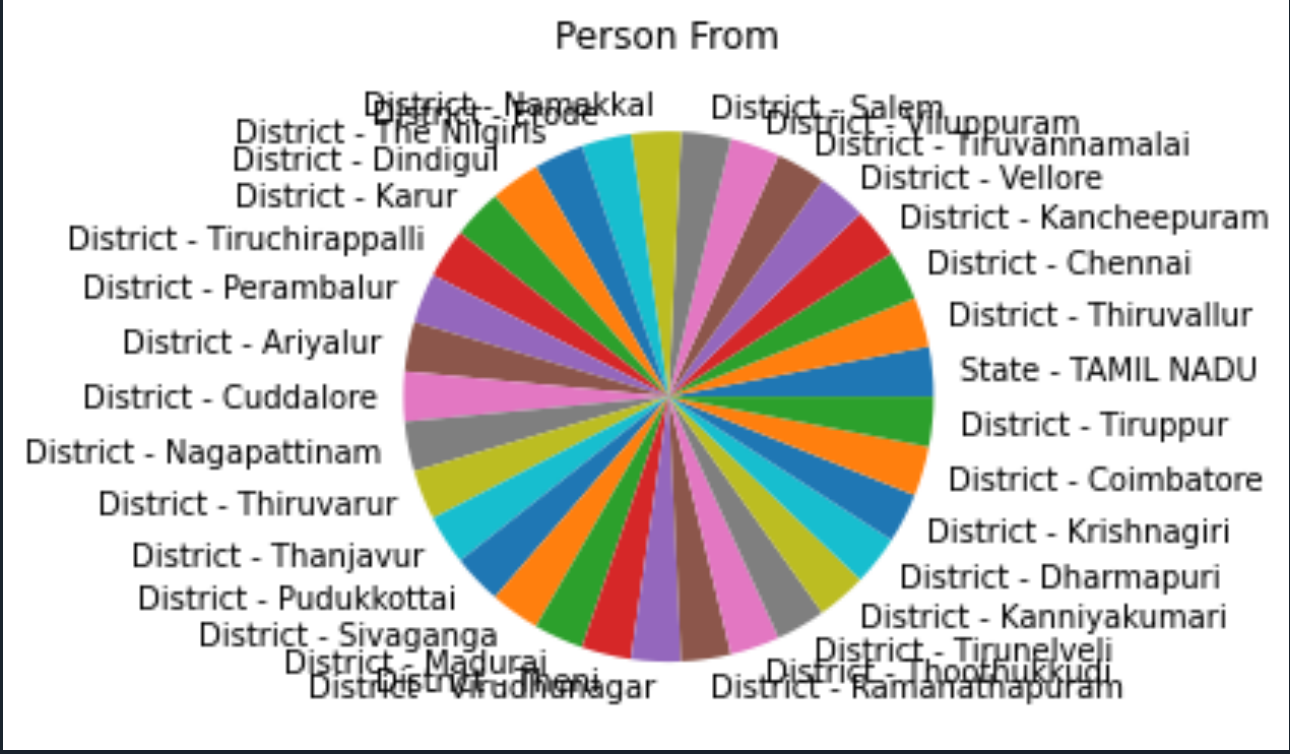
The pie chart requires one mandatory parameter data.

Labels and explode are not necessary

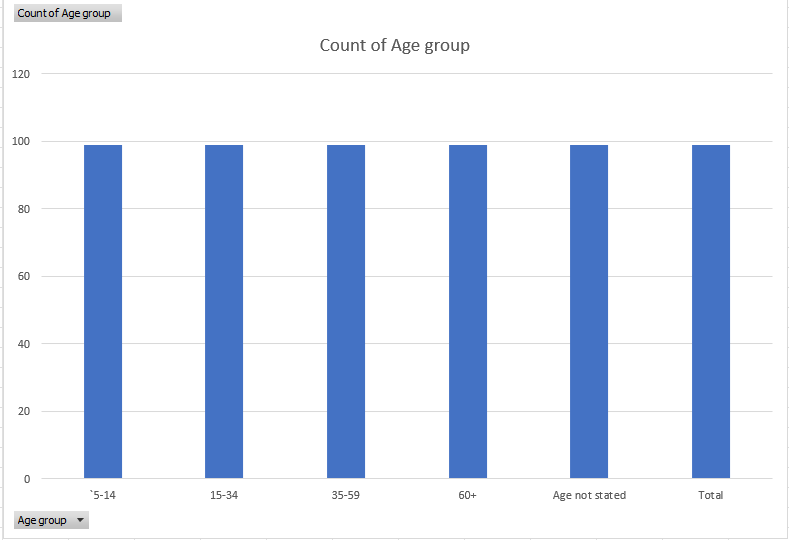
****

Person from (district vise):

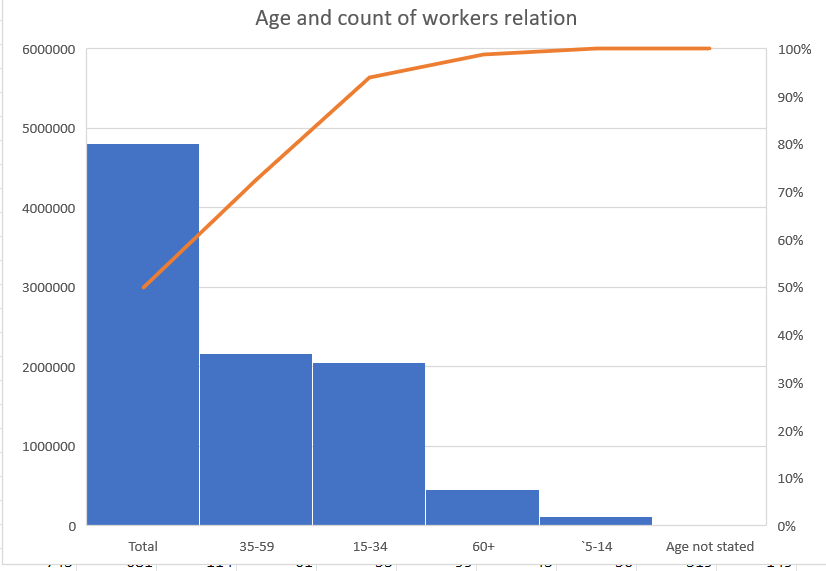


****

* frequency of age-group

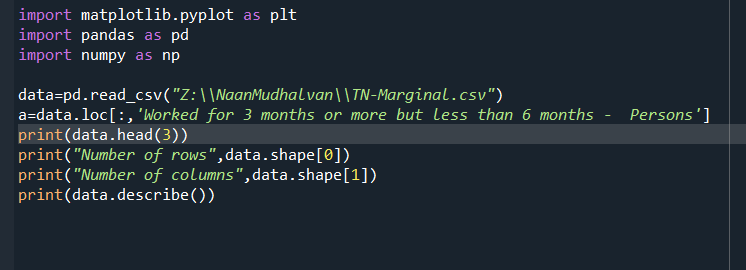
****

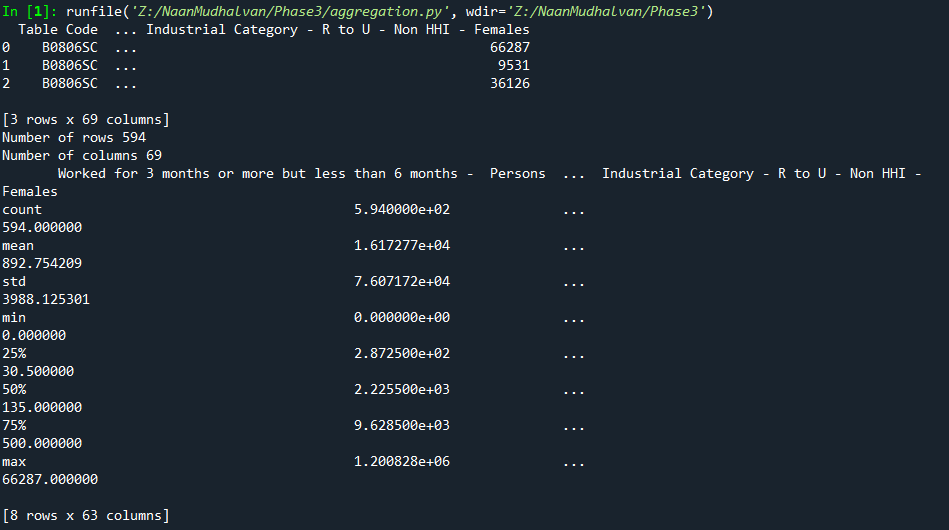
* Age and frequent of person worked within 6 months

****

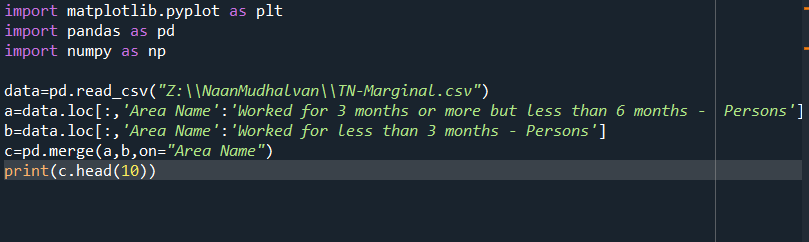
**Data Manipulation:**

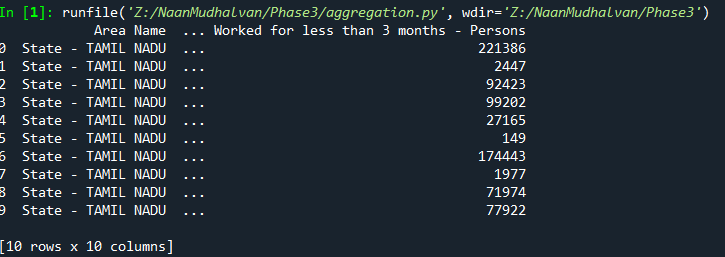
The data in the real world is very unpleasant & unordered so by performing certain operations we can make data understandable based on one’s requirements, this process of converting unordered data into meaningful information can be done by data manipulation.

****

****

Merge function:

****

****

**Conclusion:**

In conclusion, performing a demographic analysis means studying and understanding the characteristics of a specific group of people, like age, gender, income, education, and more. This helps us gain insights into the makeup of a population.

