COVER PAGE:

DATA SCIENCE TOOLBOX: PYTHON PROGRAMMINGPROJECT

(Project Semester January-April 2025)

Exploratory Analysis of Non-Workers and Main Activities in India

(Census Data)

Submitted by

K. Dhanush

12316514

Data Science – K23EP

INT375

Under the Guidance of

DR Tanima Thakur - 23532

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

PYTHON CA 2 PROJECT

Project Description

This project presents an exploratory data analysis (EDA) of the Indian population based on the **Census of India** dataset. The focus is on understanding the distribution and engagement of non-working individuals across various age groups, genders, and main activity types such as students, household duties, dependents, pensioners, and others. The insights aim to uncover social and demographic patterns relevant for policymaking and socio-economic studies.

Key Objectives:

- To analyze the **non-working population** across different age groups.
- To study the **student population** by **gender and age group**.
- To evaluate the distribution of individuals involved in **household duties** across genders.
- To summarize the **main activities** of non-working individuals at the national level.

- To **compare non-workers and students** to reveal overlaps and contrasts in age-group distribution.

1. **Objective 1: Analyze the non-working population across different age groups to understand which age brackets contribute most to the nonworking segment in India.**


```
code1.py - C:/Users/rohith kumar/OneDrive/Desktop/Dhanush/code1.py (3.13.2)
File  Edit  Format  Run  Options  Window  Help
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv("census.csv");

filtered = data[(data["Area Name"] == "India") &
                (data["Total/Rural/Urban"] == "Total") &
                (data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(12, 6))
sns.barplot(data=filtered, x="Age-Group", y="Non-Workers - Persons")
plt.xticks(rotation=45)
plt.title("Non-Workers by Age Group in India")
plt.ylabel("Number of Non-Workers")
plt.xlabel("Age Group")
plt.tight_layout()
plt.show()
```

import pandas as pd import

numpy as np import seaborn
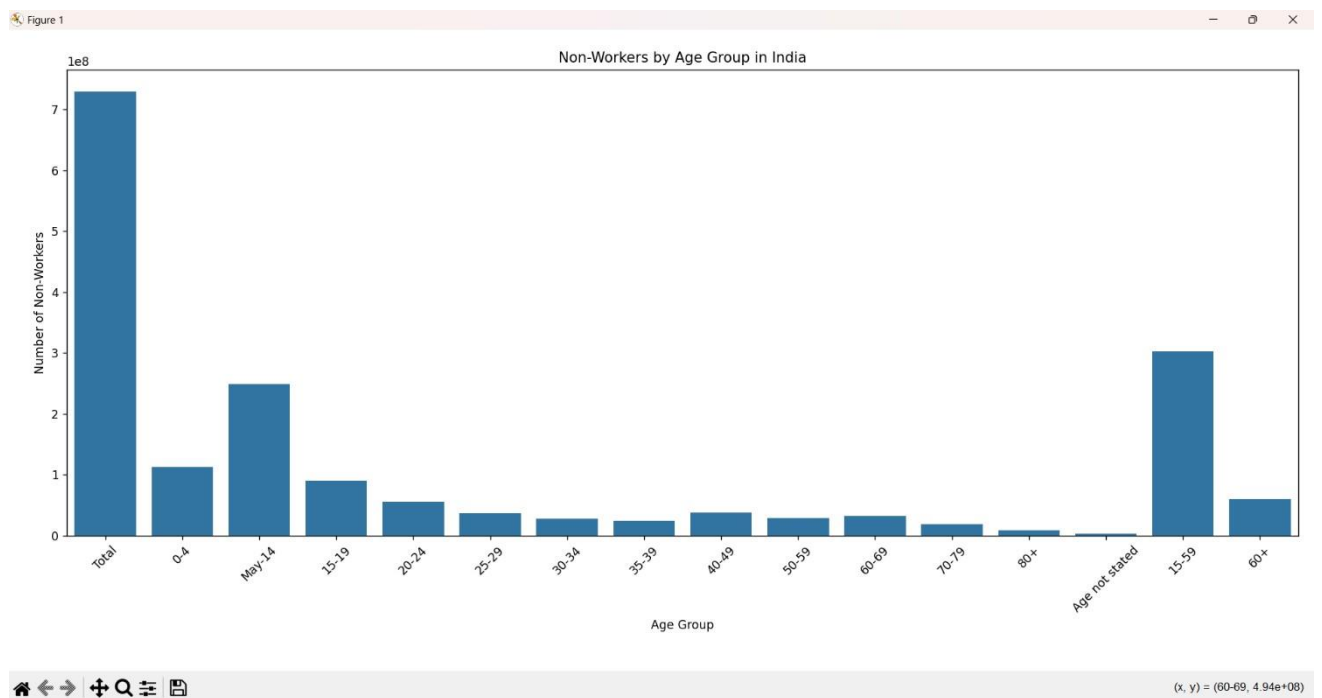
as sns import

matplotlib.pyplot as plt

data = pd.read_csv("census.csv");

filtered = data[(data["Area Name"] == "India") &

```
            (data["Total/Rural/Urban"] == "Total") &
            (data["Religion"] == "All Religious Communities")]
```

plt.figure(figsize=(12, 6)) sns.barplot(data=filtered, x="Age-Group",

y="Non-Workers - Persons") plt.xticks(rotation=45) plt.title("Non-

Workers by Age Group in India") plt.ylabel("Number of Non-Workers")

plt.xlabel("Age Group") plt.tight_layout() plt.show() Output:

2. **Objective 2: Examine the student population by age group and gender to identify
   agewise trends and assess gender disparities in student engagement**.

```
import pandas as pd import
numpy as np import seaborn
as sns import
matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
        (data["Total/Rural/Urban"] == "Total") &
        (data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(12, 6)) sns.barplot(data=filtered, x="Age-Group", y="Main
Activity - Students - Persons", color="lightblue", label="Total")
sns.barplot(data=filtered, x="Age-Group", y="Main Activity - Students - Males",
color="blue", label="Males") sns.barplot(data=filtered, x="Age-Group", y="Main
Activity - Students - Females", color="pink", label="Females", alpha=0.6)
plt.xticks(rotation=45) plt.title("Students by Age Group and Gender") plt.legend()
plt.tight_layout() plt.show() Output :
```
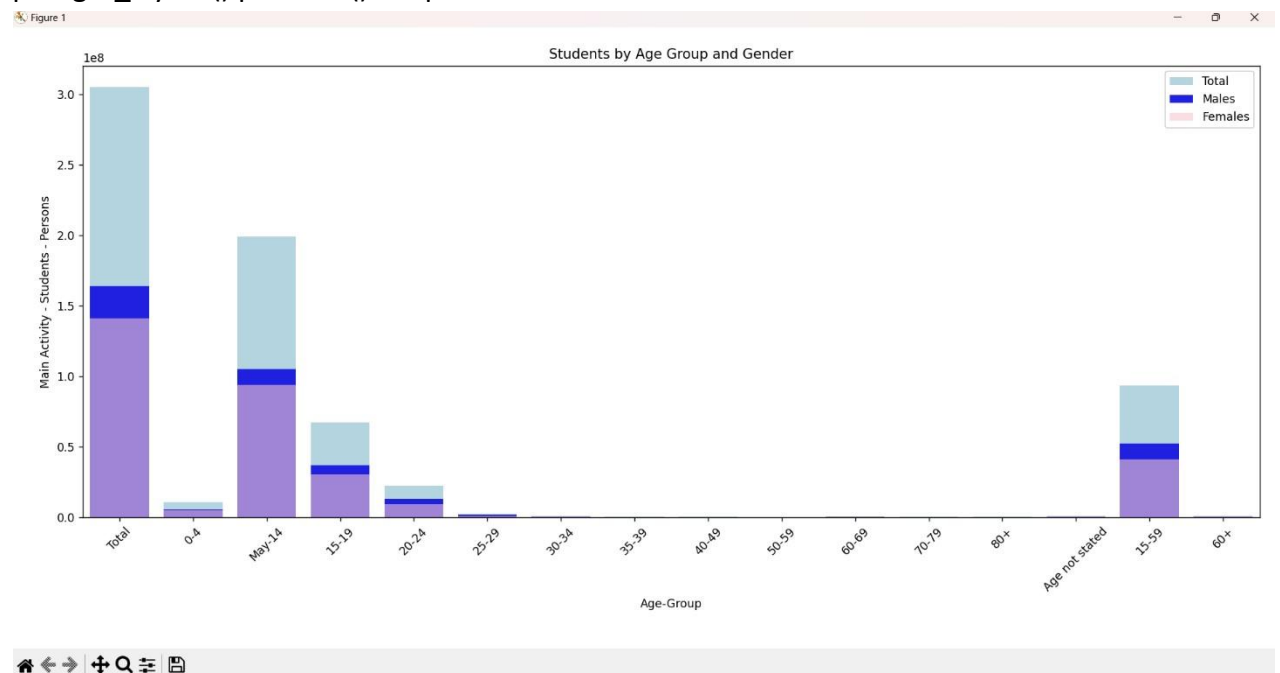

3. **Objective 3: Study the distribution of individuals engaged in household duties, focusing on gender differences and how these duties vary across age groups**.

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
                (data["Total/Rural/Urban"] == "Total") &
                (data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(12, 6))
sns.lineplot(data=filtered, x="Age-Group", y="Main Activity - Household Duties - Males", label="Males", marker='o')
sns.lineplot(data=filtered, x="Age-Group", y="Main Activity - Household Duties - Females", label="Females", marker='o')
plt.xticks(rotation=45)
plt.title("Household Duties by Gender and Age Group")
plt.tight_layout()
plt.show()
```

import pandas as pd import
numpy as np import seaborn
as sns import
matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
          (data["Total/Rural/Urban"] == "Total") &
          (data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(12, 6)) sns.lineplot(data=filtered, x="Age-Group", y="Main Activity - Household Duties - Males", label="Males", marker='o')
sns.lineplot(data=filtered, x="Age-Group", y="Main Activity - Household Duties - Females", label="Females", marker='o') plt.xticks(rotation=45)
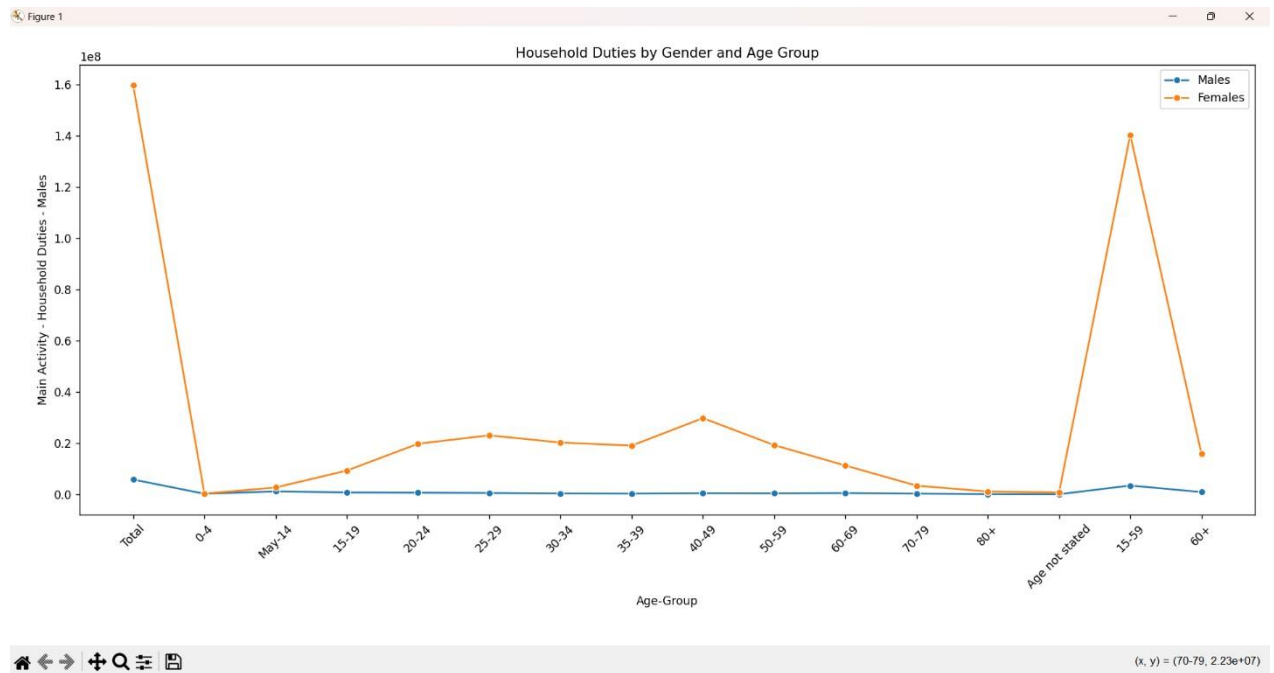plt.title("Household Duties by Gender and Age Group") plt.tight_layout()
plt.show()

Output :

Objective 4: Summarize and visualize the overall distribution of main activities (students, household duties, dependents, pensioners, rentiers, beggars, others) to capture the broader landscape of non-working roles in India.


```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
                (data["Total/Rural/Urban"] == "Total") &
                (data["Religion"] == "All Religious Communities")]

activities = {
    "Students": filtered["Main Activity - Students - Persons"].sum(),
    "Household Duties": filtered["Main Activity - Household Duties - Persons"].sum(),
    "Dependents": filtered["Main Activity - Dependents - Persons"].sum(),
    "Pensioners": filtered["Main Activity - Pensioners - Persons"].sum(),
    "Rentiers": filtered["Main Activity - Rentiers - Persons"].sum(),
    "Beggars, Vagrants": filtered["Main Activity - Beggars, Vagrants etc. - Persons"].sum(),
    "Others": filtered["Main Activity - Others - Persons"].sum()
}

plt.figure(figsize=(8, 8))
plt.pie(activities.values(), labels=activities.keys(), autopct='%1.1f%%', startangle=140)
plt.title("Distribution of Main Activities - India Total")
plt.tight_layout()
plt.show()
```
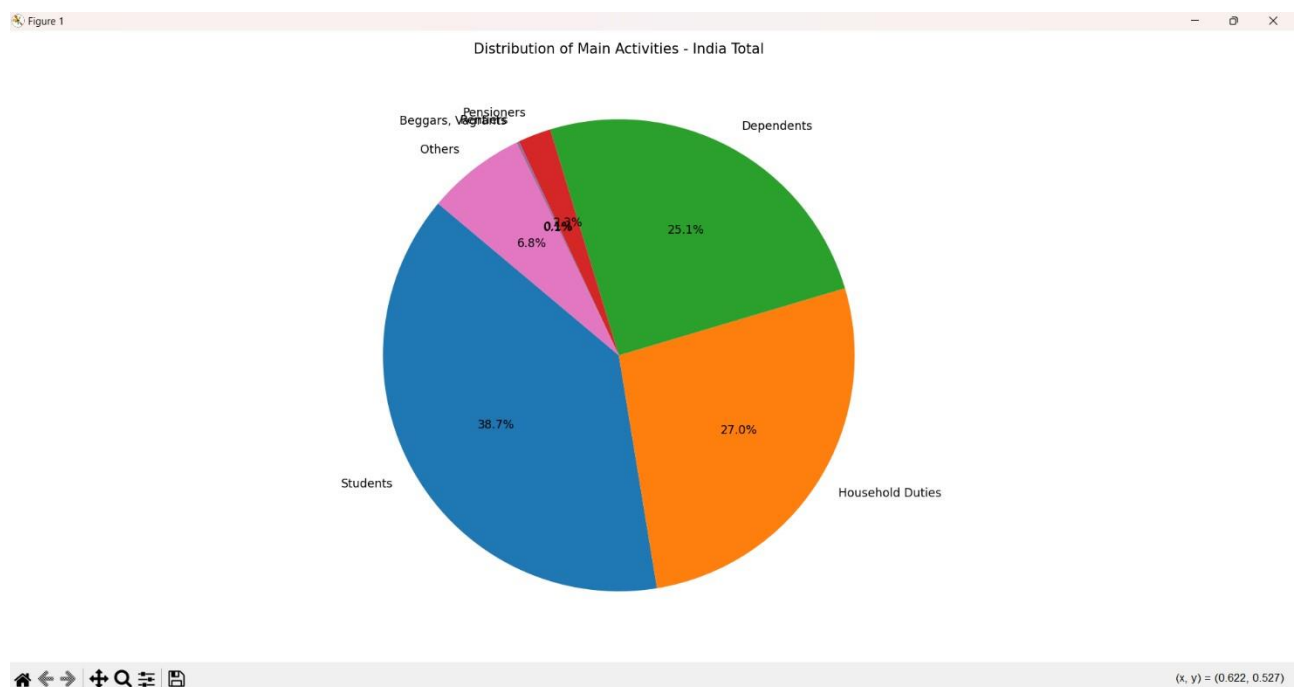
import pandas as pd import
numpy as np import seaborn

```
as sns import
matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
        (data["Total/Rural/Urban"] == "Total") &
        (data["Religion"] == "All Religious Communities")]

activities = {
    "Students": filtered["Main Activity - Students - Persons"].sum(),
    "Household Duties": filtered["Main Activity - Household Duties - Persons"].sum(),
    "Dependents": filtered["Main Activity - Dependents - Persons"].sum(),
    "Pensioners": filtered["Main Activity - Pensioners - Persons"].sum(),
    "Rentiers": filtered["Main Activity - Rentiers - Persons"].sum(),
    "Beggars, Vagrants": filtered["Main Activity - Beggars, Vagrants etc. - Persons"].sum(),
"Others": filtered["Main Activity - Others - Persons"].sum()
}

plt.figure(figsize=(8, 8)) plt.pie(activities.values(), labels=activities.keys(),
autopct='%1.1f%%', startangle=140) plt.title("Distribution of Main Activities - India
Total") plt.tight_layout() plt.show() Output :
```

Objective 5: Compare non-workers and students across age groups to highlight overlapping trends, differences, and key insights regarding educational engagement among the non-working population.


```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &
                (data["Total/Rural/Urban"] == "Total") &
                (data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(10, 6))
sns.barplot(data=filtered, x="Age-Group", y="Non-Workers - Persons", color="gray", label="Non-Workers")
sns.barplot(data=filtered, x="Age-Group", y="Main Activity - Students - Persons", color="green", label="Students")
plt.xticks(rotation=45)
plt.title("Non-Workers vs. Students by Age Group")
plt.legend()
plt.tight_layout()
plt.show()
```

import pandas as pd

import numpy as np import

seaborn as sns import
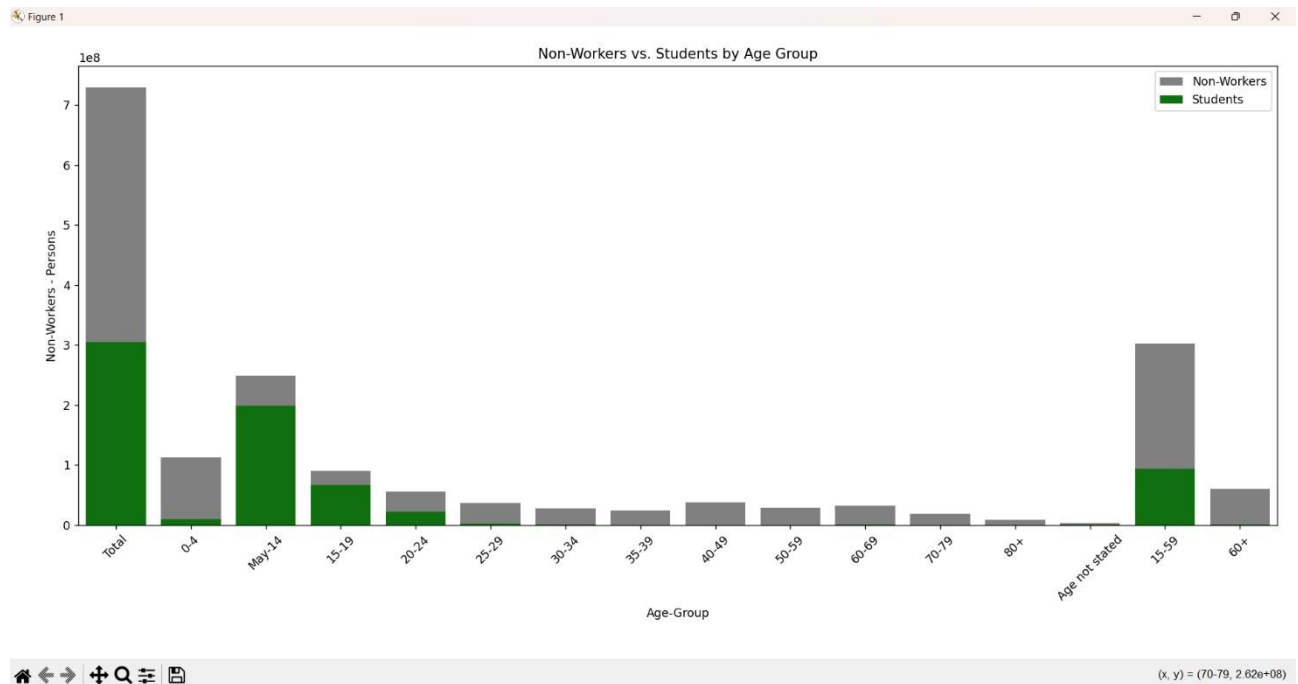
matplotlib.pyplot as plt

data = pd.read_csv("census.csv")

filtered = data[(data["Area Name"] == "India") &

(data["Total/Rural/Urban"] == "Total") &

(data["Religion"] == "All Religious Communities")]

plt.figure(figsize=(10, 6))

sns.barplot(data=filtered, x="Age-Group", y="Non-Workers - Persons", color="gray", label="Non-Workers")

sns.barplot(data=filtered, x="Age-Group", y="Main Activity - Students - Persons",

color="green", label="Students") plt.xticks(rotation=45) plt.title("Non-Workers

vs. Students by Age Group") plt.legend() plt.tight_layout() plt.show() Output :

Non-Workers vs. Students by Age Group

Conclusion:

This project provides valuable insights into the non-working population and their engagement in various main activities in India. Through visualizations such as bar charts, line plots, and pie charts, we were able to observe distinct patterns related to age, gender, and occupation. The analysis revealed the following key findings:

- Non-working Population: Significant age groups, particularly the elderly and younger individuals, contribute the most to the non-working segment. These insights can help inform policies for job creation and social security.

- Gender Differences: Gender disparities were observed in various categories, especially with respect to students and individuals involved in household duties. Females tend to dominate the household duties category, while the student population shows a more balanced gender distribution.

- Main Activities: A substantial portion of the non-working population is engaged in activities such as being dependents, pensioners, or rentiers, which may reflect India's socio-economic structure.

The visualizations provided a clear understanding of these trends, and the use of various analytical methods allowed for a comprehensive view of India's demographic behavior.

Future Scope:

While this analysis has provided a strong foundation for understanding India's nonworking population, several avenues can be explored further:

1. **Regional Analysis:** Expanding the study to include specific states or regions within India could offer insights into regional differences and better inform localized policy decisions.

2. **Time Series Analysis:** Incorporating time-series data to study how the non-working population and main activities have evolved over the years could uncover important trends related to socio-economic changes, such as the impact of education policies or economic reforms.

3. **Impact of Technology:** Future studies could explore how the rise of automation and digital platforms affects non-working individuals, particularly in urban areas.

4. **Socio-Economic Factors:** Delving deeper into the socio-economic background of the population, such as income, education level, and family structure, could help understand the factors that contribute to individuals being non-workers.

References:

1.Census 2011 INDIA and States - https://www.data.gov.in/catalog/non-workers-main-activity-age-and-sex-and-religious-community-census-2011-india-and-states