PHASE 5: PROJECT DOCUMENTATION

PROJECT TITLE:

ASSESSMENT OF MARGINAL WORKERS IN TAMIL NADU-A SOCIO-ECONOMIC ANALYSIS.

TEAM MEMBERS:

SIVA GURU K

NITHISH

MANIBARATHI

KARTHIKEYAN

OBJECTIVES:

- Analyze the demographic characteristics of marginal workers in the state of Tamil Nadu, India.
- Identify trends and patterns among marginal workers, including their age, gender, education level, and employment status.
- Understand the challenges faced by marginal workers in the state.
- Provide insights into the development, job creation, and financial education needs of marginal workers.

ANALYSIS APPROACH:

The following steps were taken to analyze the dataset using IBM COGNOS:

1. The dataset was cleaned and preprocessed to remove any errors or inconsistencies.
2. Descriptive statistics were calculated to summarize the demographic characteristics of marginal workers.
3. Exploratory data analysis (EDA) techniques were used to identify trends and patterns in the data.
4. Inferential statistics were used to test hyp otheses about the data.

VISUALIZATION TYPES:

The following types of visualizations were used to present the results of the analysis:

- Bar charts and
- Line charts

DATA SET :

Link:(https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil)

CODE IMPLEMENTATION:

The project was implemented in Python using the following libraries:

- Pandas
- NumPy
- Matplotlib
- Seaborn

PYTHON CODES:

#Importing nesscary pakages

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
#Loading the data set(DATA EXPLORATION)
```

```
df = pd.read_csv("/content/DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
df
```

OUTPUT:

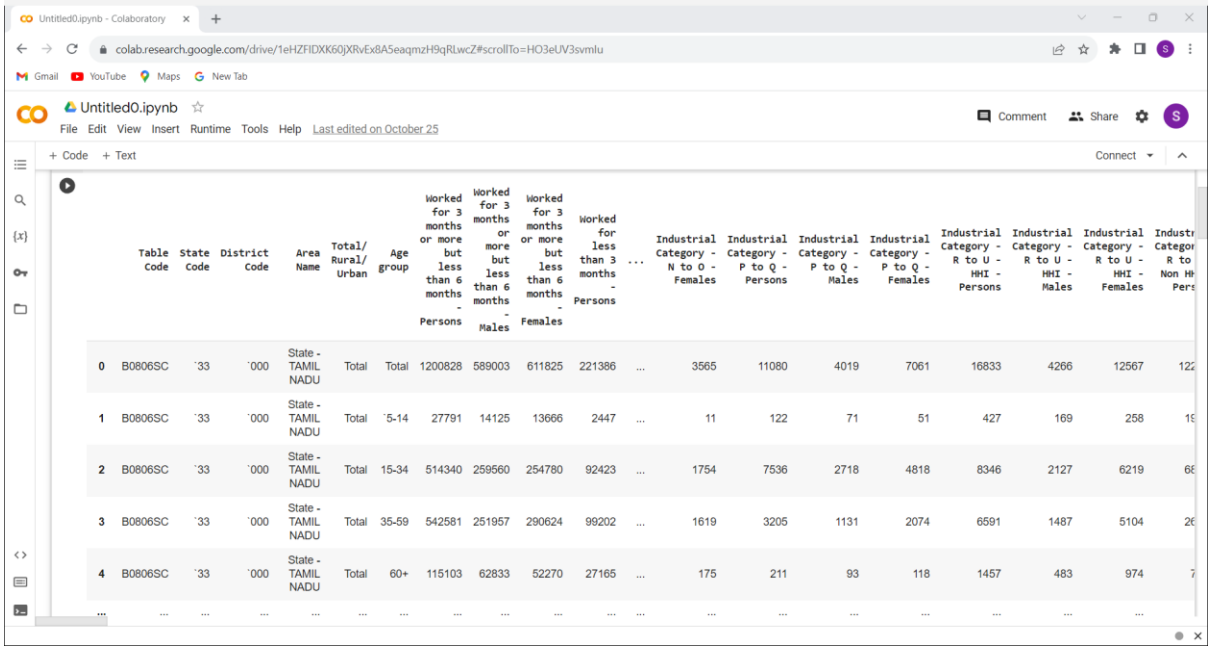
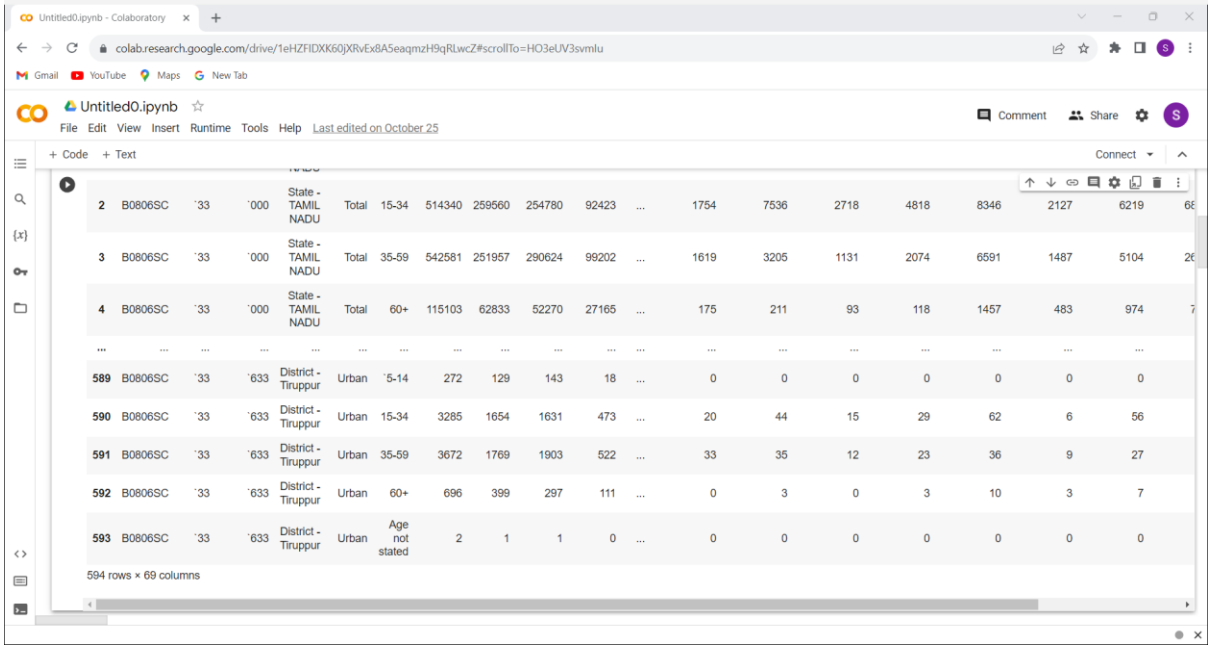


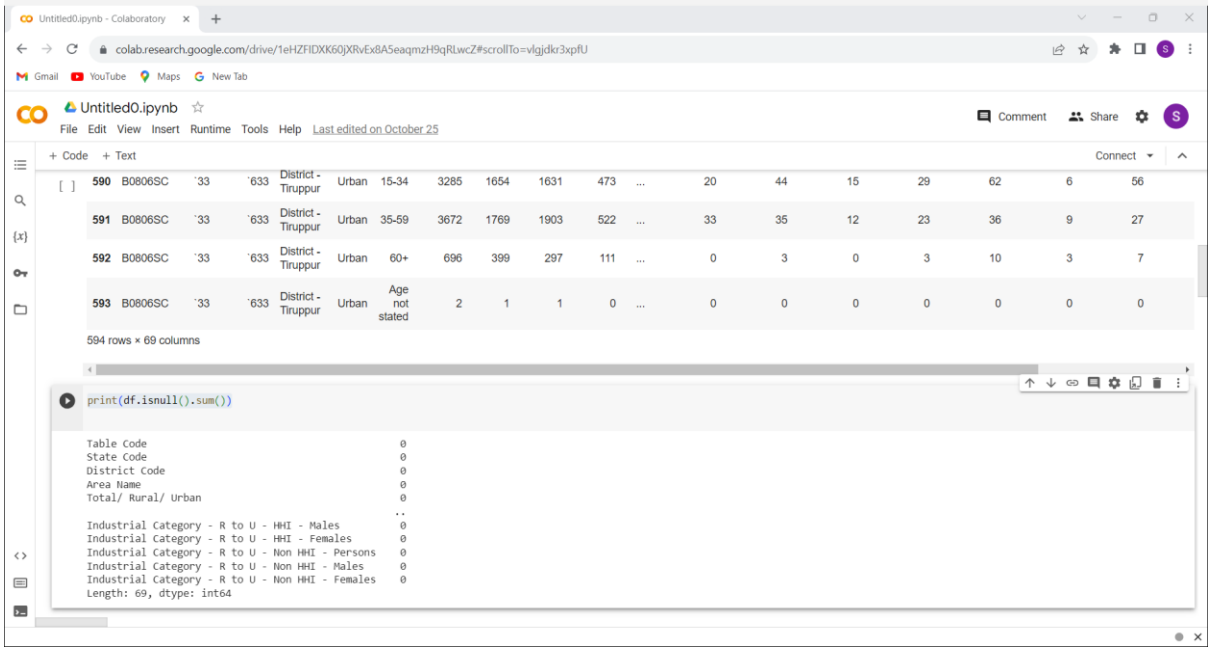
	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	Age group	Worked for 3 months or more but less than 6 months - Persons	Worked for 3 months or more but less than 6 months - Males	Worked for 3 months or more but less than 6 months - Females	Worked for less than 3 months - Persons	...	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	Industr Categor R to Non HH Pers
0	B0806SC	'33	'000	State - TAMIL NADU	Total	Total	1200828	589003	611825	221386	...	3565	11080	4019	7061	16833	4266	12567	122
1	B0806SC	'33	'000	State - TAMIL NADU	Total	'5-14	27791	14125	13666	2447	...	11	122	71	51	427	169	258	19
2	B0806SC	'33	'000	State - TAMIL NADU	Total	15-34	514340	259560	254780	92423	...	1754	7536	2718	4818	8346	2127	6219	68
3	B0806SC	'33	'000	State - TAMIL NADU	Total	35-59	542581	251957	290624	99202	...	1619	3205	1131	2074	6591	1487	5104	26
4	B0806SC	'33	'000	State - TAMIL NADU	Total	60+	115103	62833	52270	27165	...	175	211	93	118	1457	483	974	7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	



#Checking for missing values:

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

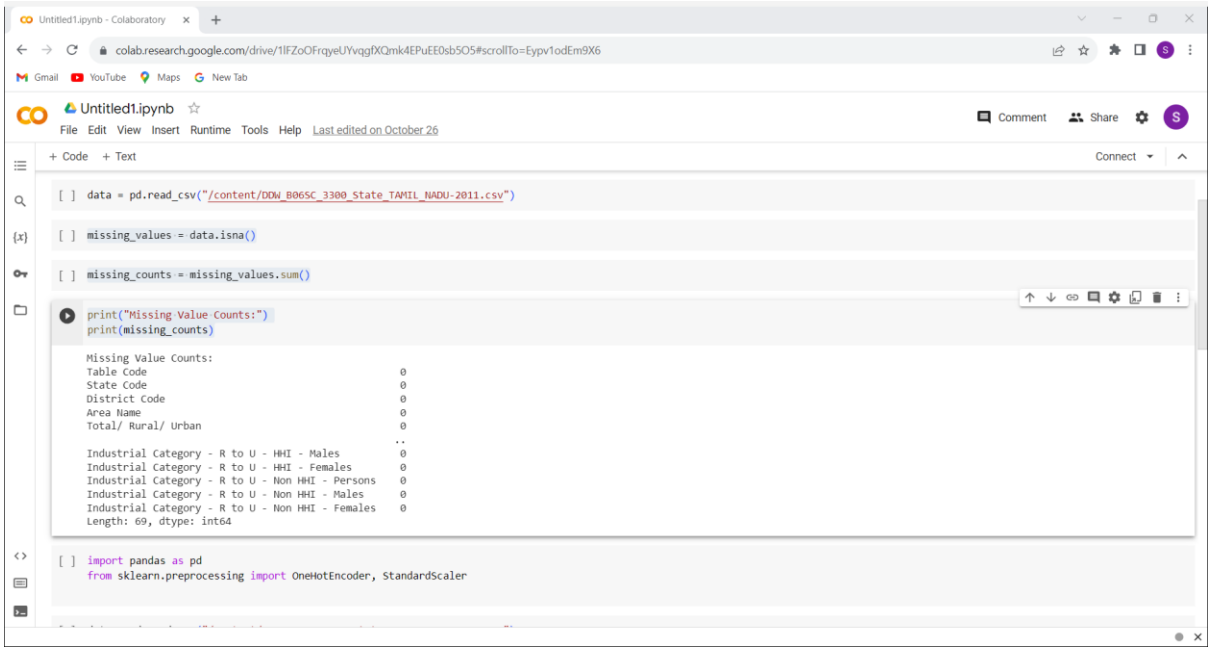
OUTPUT:



#Handling Missing Values:

```
missing_values = data.isna()
missing_counts = missing_values.sum()
print("Missing Value Counts:")
print(missing_counts)
```

OUTPUT:



#Demographic Analysis(Marginal workers based on Age)

```
age_counts = df['Age group'].value_counts()
print(age_counts)
```

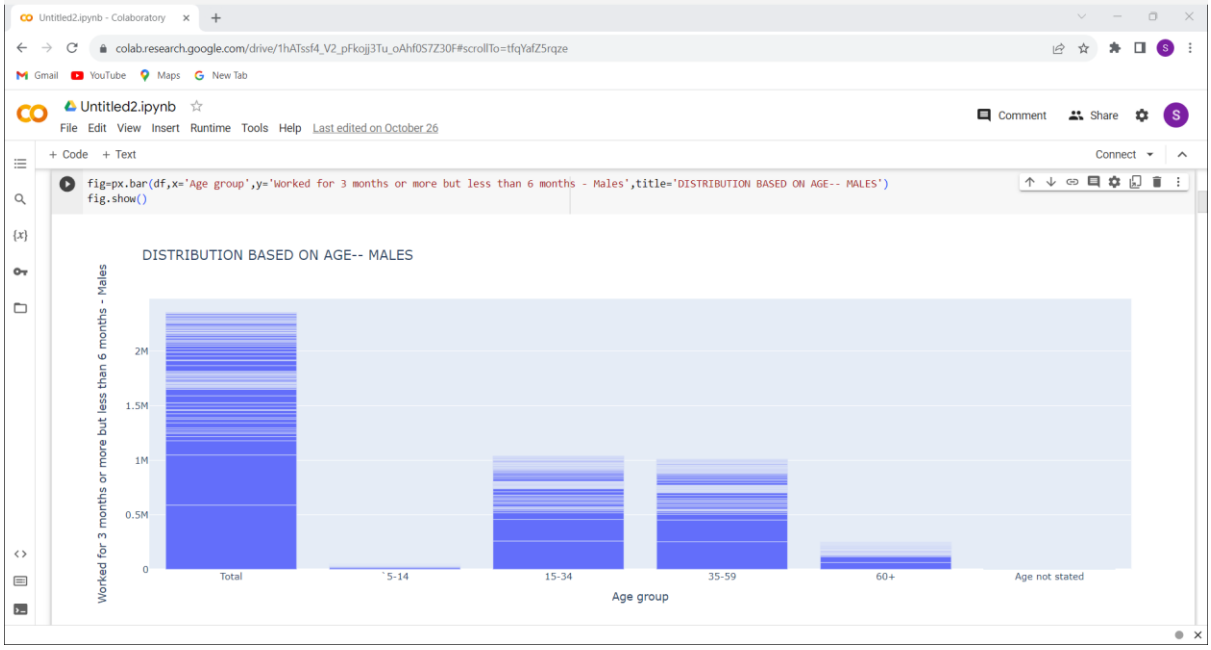
OUTPUT:

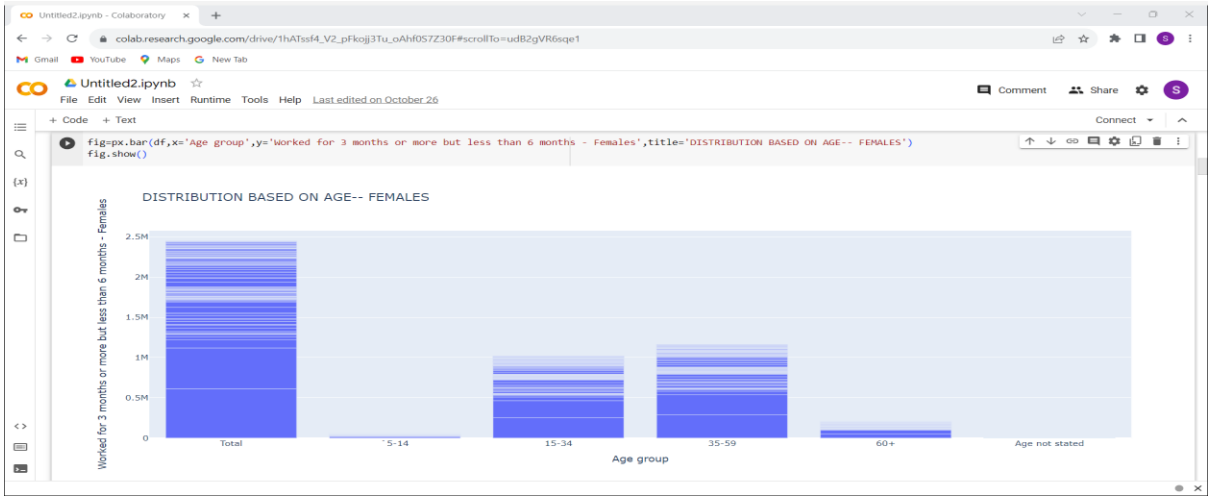
```
Total              99
`5-14              99
15-34              99
35-59              99
60+                99
Age not stated     99

Name: Age group, dtype: int64
```

VISUALIZATION:

```
fig=px.bar(df,x='Age group',y='Worked for 3 months or more but less than 6 months -  Persons',title='DISTRIBUTION BASED ON AGE-- PERSONS')
fig.show()
```

Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/1hATssf4_V2_pFkojj3Tu_oAhf0S7Z30F#scrollTo=D22q6M21p6a1
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last edited on October 26
Comment Share
+ Code + Text
Connect
fig=px.bar(df,x='Age group',y='Worked for 3 months or more but less than 6 months - Persons',title='DISTRIBUTION BASED ON AGE-- PERSONS')
fig.show()
DISTRIBUTION BASED ON AGE-- PERSONS
Worked for 3 months or more but less than 6 months - Persons
5M
4M
3M
2M
1M
0
Total
`5-14
15-34
35-59
60+
Age not stated
Age group


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/1hATssf4_V2_pFkojj3Tu_oAhf0S7Z30F#scrollTo=tfqYafZ5rqze
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last edited on October 26
Comment Share
+ Code + Text
Connect
fig=px.bar(df,x='Age group',y='Worked for 3 months or more but less than 6 months - Males',title='DISTRIBUTION BASED ON AGE-- MALES')
fig.show()
DISTRIBUTION BASED ON AGE-- MALES
Worked for 3 months or more but less than 6 months - Males
2M
1.5M
1M
0.5M
0
Total
`5-14
15-34
35-59
60+
Age not stated
Age group


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/1hATssf4_V2_pFkojj3Tu_oAhf0S7Z30F#scrollTo=udB2gVR6sqe1
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last edited on October 26
Comment Share
+ Code + Text
Connect
fig=px.bar(df,x='Age group',y='Worked for 3 months or more but less than 6 months - Females',title='DISTRIBUTION BASED ON AGE-- FEMALES')
fig.show()
DISTRIBUTION BASED ON AGE-- FEMALES
Worked for 3 months or more but less than 6 months - Females
2.5M
2M
1.5M
1M
0.5M
0
Total
`5-14
15-34
35-59
60+
Age not stated
Age group

#DEMOGRAPHIC ANALYSIS(INDUSTRIAL WORKER A):

```
industrial_category_counts = df['Industrial Category - A - Cultivators - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      84
2      23
1      22
11     10
6      10
       ..
117     1
443     1
203     1
167     1
119     1

Name: Industrial Category - A - Cultivators - Males, Length: 273,
dtype: int64
```

```
industrial_category_counts = df['Industrial Category - A - Cultivators - Females'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      93
1      25
3      14
11     13
4      12
       ..
210     1
398     1
170     1
168     1
140     1

Name: Industrial Category - A - Cultivators - Females, Length: 266,
dtype: int64
```

```
industrial_category_counts = df['Industrial Category - A - Agricultural labourers - Persons'].value_counts()
print(industrial_category_counts)
```

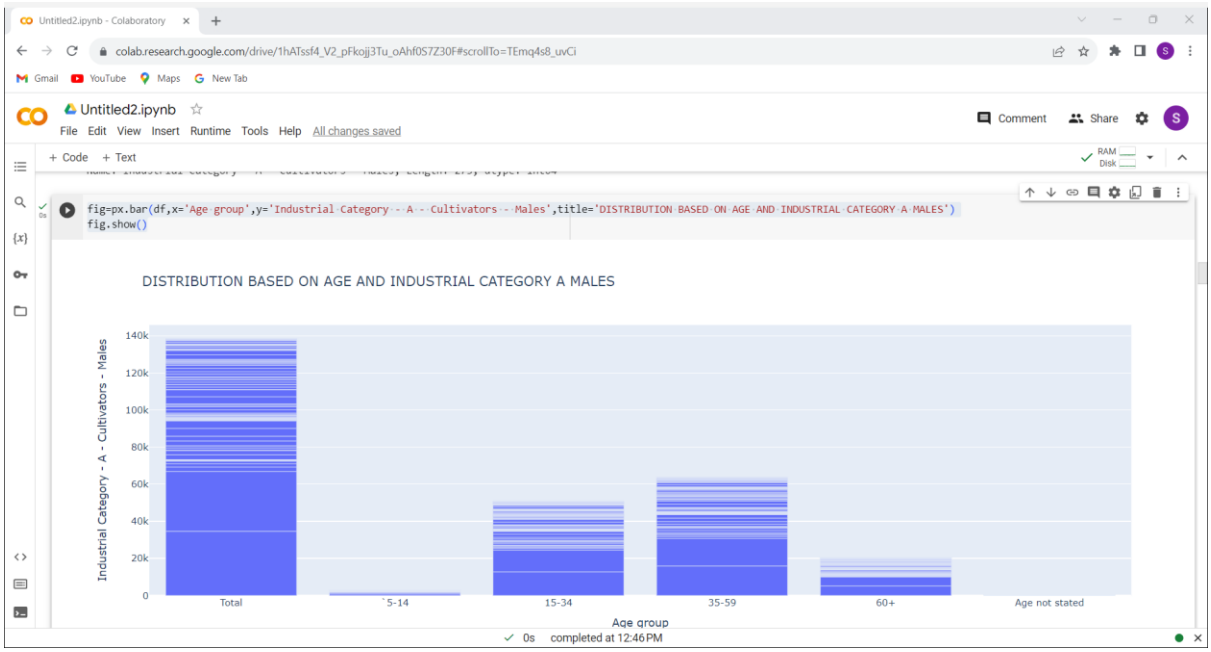
OUTPUT:

```
0       24
1       12
2       11
3        9
4        9
        ..
142      1
859      1
```

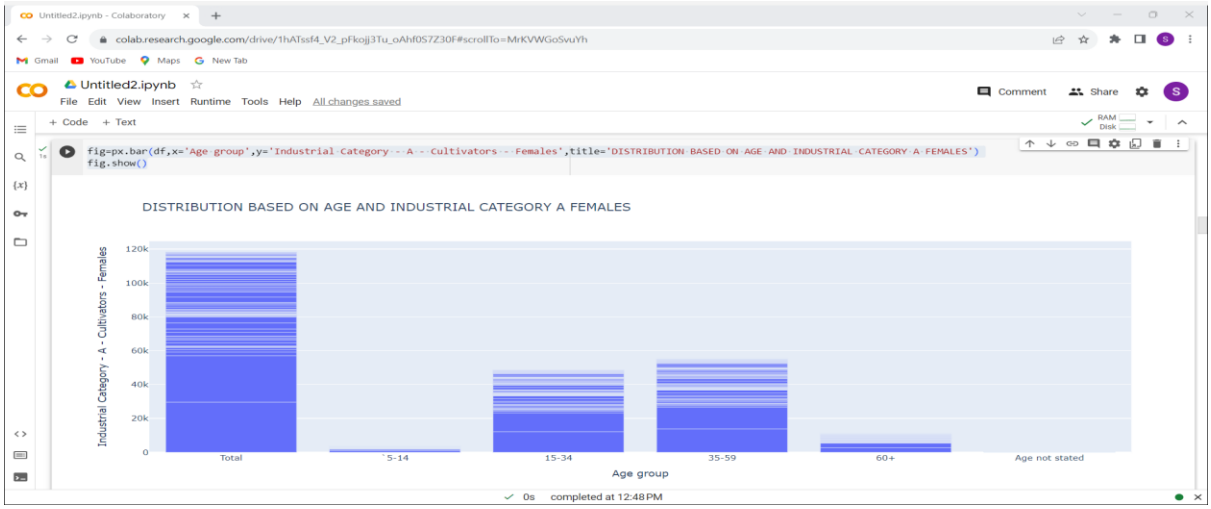
```
572       1
1577      1
2063      1

Name: Industrial Category - A - Agricultural labourers - Persons,
Length: 471, dtype: int64
```

VISUALIZATION:

```
fig=px.bar(df,x='Age group',y='Industrial Category - A - Cultivators - Males',title='DISTRIBUTION BASED ON AGE AND INDUSTRIAL CATEGORY A MALES')
fig.show()
```




```
fig=px.bar(df,x='Age group',y='Industrial Category - A - Cultivators - Females',title='DISTRIBUTION BASED ON AGE AND INDUSTRIAL CATEGORY A FEMALES')
fig.show()
```



#Demographic Analysis:

```
industrial_category_counts = df['Industrial Category - A - Agricultural labourers - Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0       24
1       12
2       11
3        9
4        9
        ..
142      1
859      1
572      1
1577     1
2063     1

Name: Industrial Category - A - Agricultural labourers - Persons,
Length: 471, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - A - Agricultural labourers - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      30
1      24
2      17
3      15
4       7
       ..
443     1
331     1
62      1
232     1
935     1

Name: Industrial Category - A - Agricultural labourers - Males, Length:
437, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - A - Agricultural labourers - Females'].value_counts()
print(industrial_category_counts)
```

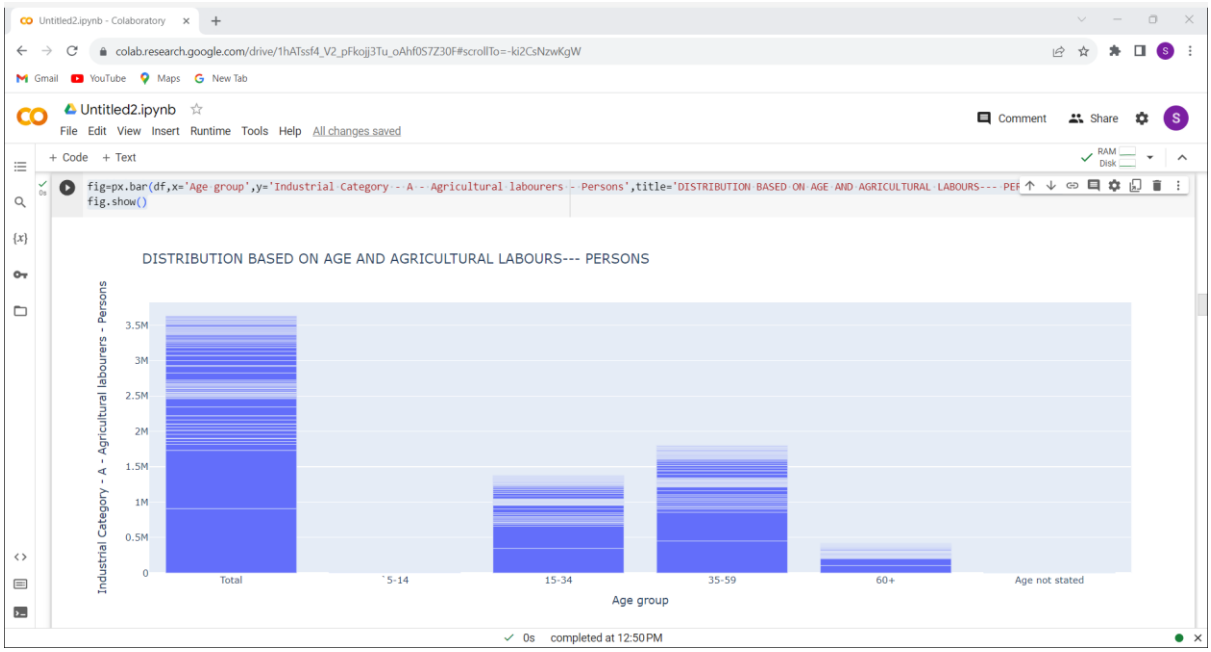
OUTPUT:

```
0        40

2        15
6        12
1        12
3        11
         ..
13216     1
416       1
```

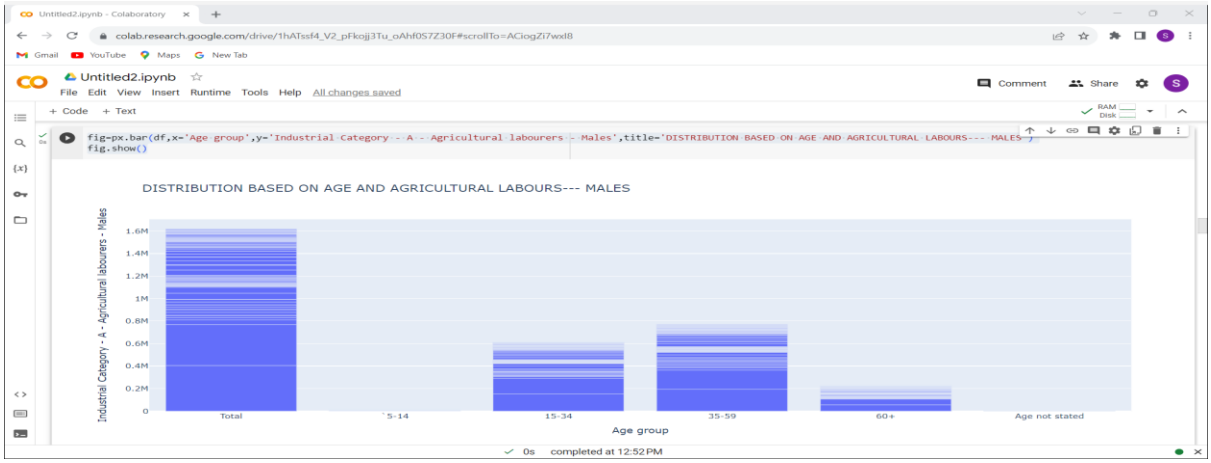
```
241        1
733        1
198        1

Name: Industrial Category - A - Agricultural labourers - Females,
Length: 432, dtype: int64
```

VISUALIZATION:

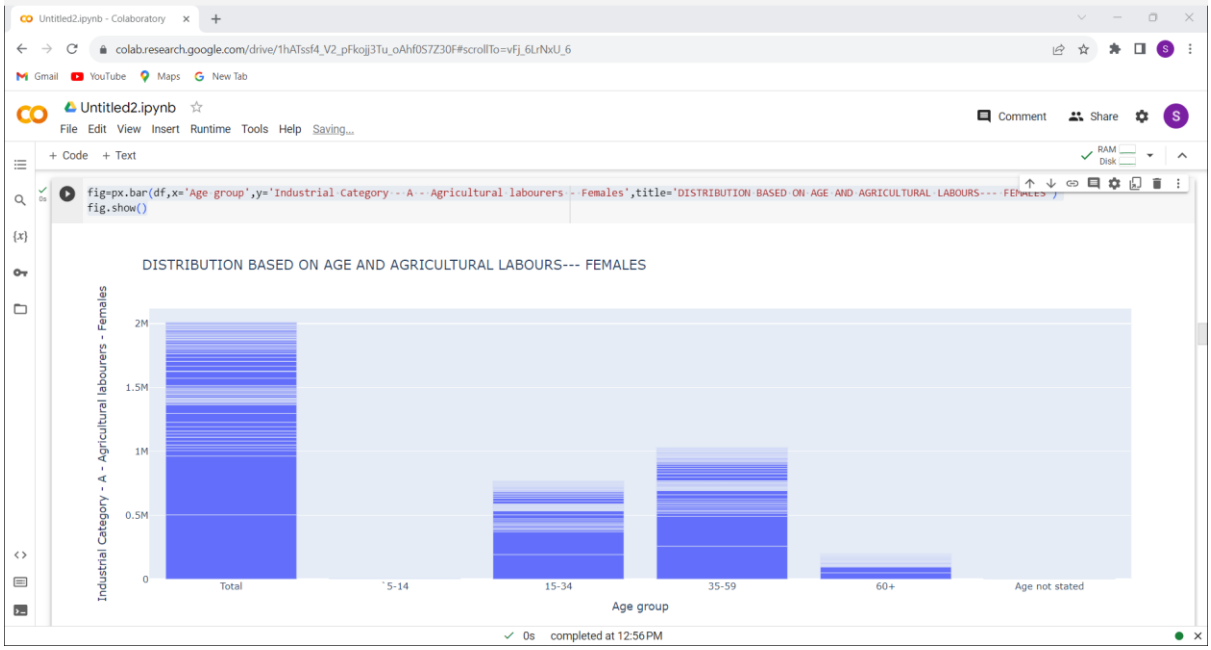
```
fig=px.bar(df,x='Age group',y='Industrial Category - A - Agricultural labourers - Persons',title='DISTRIBUTION BASED ON AGE AND AGRICULTURAL LABOURS--- PERSONS')
fig.show()
```




```
fig=px.bar(df,x='Age group',y='Industrial Category - A - Agricultural labourers - Males',title='DISTRIBUTION BASED ON AGE AND AGRICULTURAL LABOURS--- MALES')
fig.show()
```



```
fig=px.bar(df,x='Age group',y='Industrial Category - A - Agricultural labourers - Females',title='DISTRIBUTION BASED ON AGE AND AGRICULTURAL LABOURS--- FEMALES')
fig.show()
```



#Demographic analysis(ALLIED ACTIVITIES Labours)

```
industrial_category_counts = df['Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT: 0 154

```
6        17
12        9
8         9
14        9
       ...
1224      1
2332      1
1586      1
870       1
35        1

Name: Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Persons, Length: 256, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - A - Plantation, Livestock, Forestry, Fishing, Hunting and allied activities - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
 0      168
```

```
6       22
12      13
18      11
8       10
      ...
49       1
102      1
254      1
132      1
84       1

Name: Industrial Category - A - Plantation, Livestock, Forestry,
Fishing, Hunting and allied activities - Males, Length: 224, dtype:
int64
```

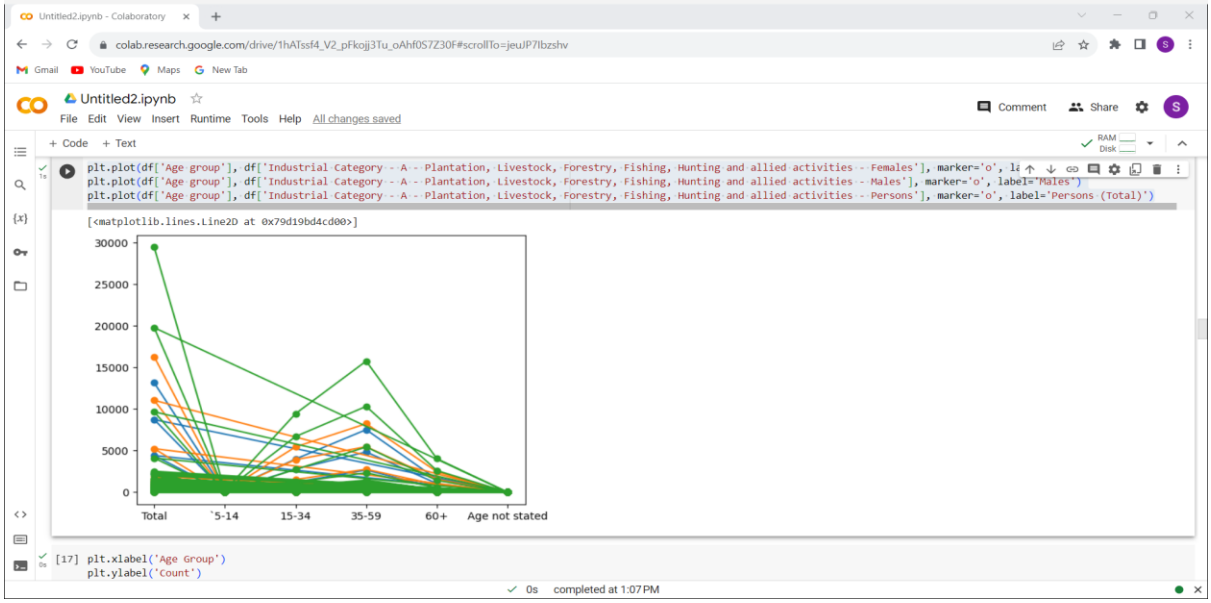
```
industrial_category_counts = df['Industrial Category - A - Plantation,
Livestock, Forestry, Fishing, Hunting and allied activities -
Females'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
 0      186

12      17
2       17
34      12
8       11
      ...
349      1
566      1
364      1
640      1
29       1

Name: Industrial Category - A - Plantation, Livestock, Forestry,
Fishing, Hunting and allied activities - Females, Length: 196, dtype:
int64
```

VISUALIZATION:

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


plt.plot(df['Age group'], df['Industrial Category - A - Plantation,
Livestock, Forestry, Fishing, Hunting and allied activities -
Females'], marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - A - Plantation,
Livestock, Forestry, Fishing, Hunting and allied activities - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - A - Plantation,
Livestock, Forestry, Fishing, Hunting and allied activities -
Persons'], marker='o', label='Persons (Total)')
```



#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY-B)

```
industrial_category_counts = df['Industrial Category - B - Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
 0       322

6        46
3        18
12       14
2        12
       ...
57        1
28        1
294       1
161       1
46        1

Name: Industrial Category - B - Persons, Length: 98, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - B - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
 0       344

6        54
12       23
3        18
9         9
       ...
16        1
7         1
```

```
36        1
130       1
31        1

Name: Industrial Category - B - Males, Length: 77, dtype: int64
```

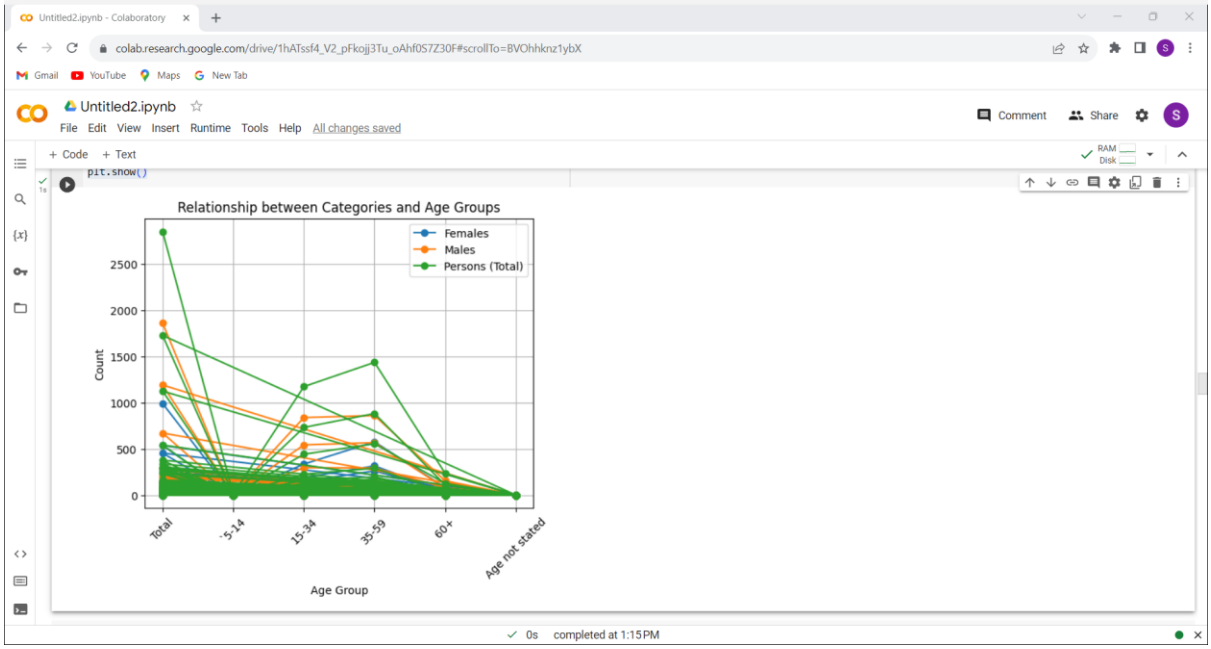
```
industrial_category_counts = df['Industrial Category - B - Females'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0       401

3        24
2        23
6        16
8        14
       ...
15        1
144       1
78        1
114       1
11        1

Name: Industrial Category - B - Females, Length: 65, dtype: int64
```

#VISUALIZATION:

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


plt.plot(df['Age group'], df['Industrial Category - B - Females'], marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - B - Males'], marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - B - Persons'], marker='o', label='Persons (Total)')


plt.xlabel('Age Group')
plt.ylabel('Count')

plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)

plt.show()
```



#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY C HHI)

```
industrial_category_counts = df['Industrial Category - C - HHI - Persons'].value_counts()
print(industrial_category_counts)
OUTPUT:
0       157
6        15
2        11
3        10
12       10
       ...
169       1
232       1
79        1
47        1
171       1
Name: Industrial Category - C - HHI - Persons, Length: 236, dtype: int64
```

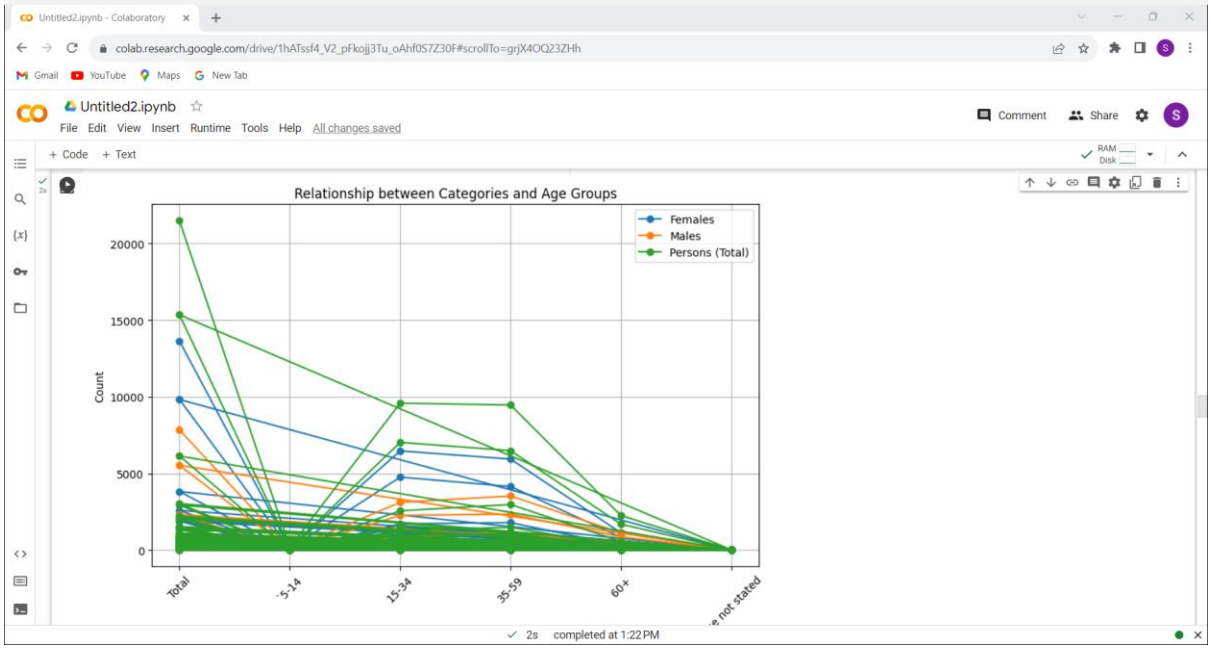
```
industrial_category_counts = df['Industrial Category - C - HHI - Males'].value_counts()
print(industrial_category_counts)
OUTPUT:
0       196
6        33
12       20
```

```
18       12
8        10
       ...
263       1
7         1
732       1
588       1
166       1

Name: Industrial Category - C - HHI - Males, Length: 155, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - C - HHI - Females'].value_counts()
print(industrial_category_counts)
OUTPUT:
0       177
6        21
14       14
2        13
8        12
       ...
230       1
426       1
176       1
470       1
83        1

Name: Industrial Category - C - HHI - Females, Length: 187, dtype: int64
```

VISUALIZATION:

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

plt.plot(df['Age group'], df['Industrial Category - C - HHI - Females'], marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - C - HHI - Males'], marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - C - HHI - Persons'], marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')

plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True
)plt.show()
```



#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY C NON HHI):

```
industrial_category_counts = df['Industrial Category - C - Non HHI - Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0       99
6       26
2       16
18      12
3       10
        ..
496      1
515      1
304      1
177      1
626      1

Name: Industrial Category - C - Non HHI - Persons, Length: 303, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - C - Non HHI - Males'].value_counts()
print(industrial_category_counts)
OUTPUT: 0       99

6       26
2       16
18      12
3       10
        ..
496      1
515      1
```

```
304     1
177     1
626     1
Name: Industrial Category - C - Non HHI - Persons, Length: 303, dtype:
int64
```

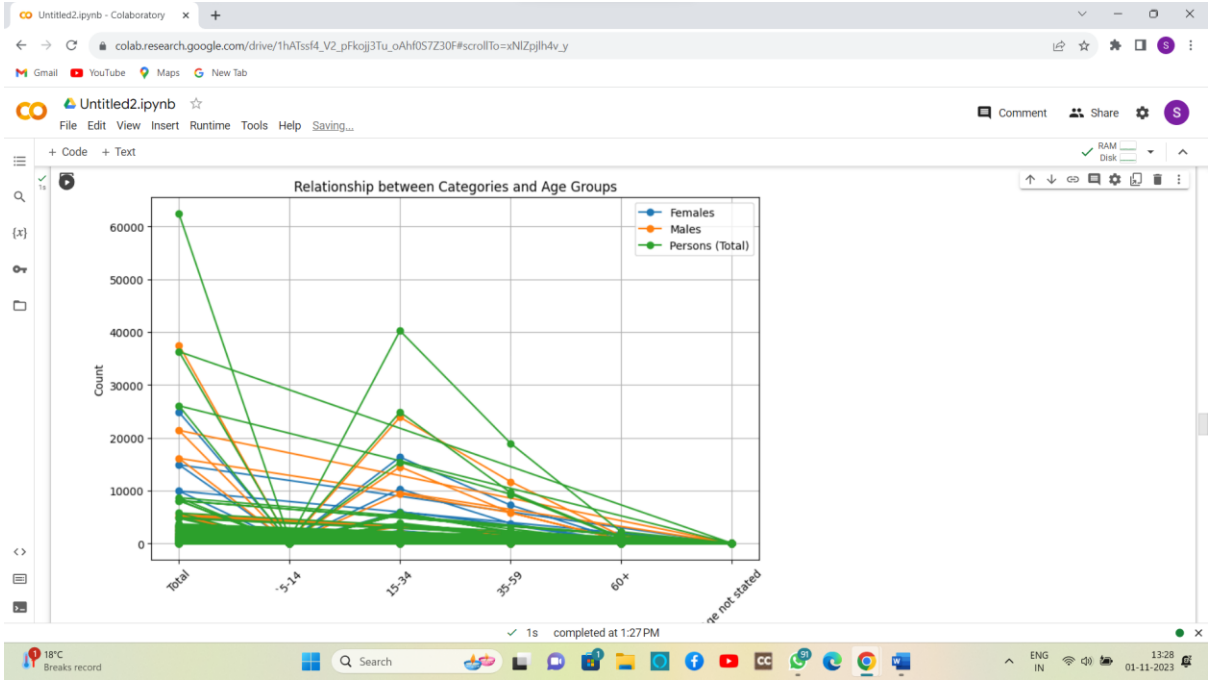
```
industrial_category_counts = df['Industrial Category - C - Non HHI -
Females'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      155
6       20
18      15
20      13
12      11
      ...
107      1
422      1
213      1
195      1
287      1

Name: Industrial Category - C - Non HHI - Females, Length: 230, dtype:
int64
```

#VISUALIZATION:

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

plt.plot(df['Age group'], df['Industrial Category - C - Non HHI -
Females'], marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - C - Non HHI -
Males'], marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - C - Non HHI -
Persons'], marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()

plt.xticks(rotation=45)
plt.grid(True)

plt.show()
```



#DEMOGRAHIC ANALYSIS(INDUSTRIAL CATEGORY D &E) 0 300

```
industrial_category_counts = df['Industrial Category - D & E - Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0       300
6        40
2        23
12       21
14       17
       ...
105       1
67        1
58        1
91        1
19        1

Name: Industrial Category - D & E - Persons, Length: 89, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - D & E - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      330
6       45
3       25
12      19
8       15
      ...
91       1
61       1
```

```
74      1
58      1
64      1

Name: Industrial Category - D & E - Males, Length: 73, dtype: int64
```

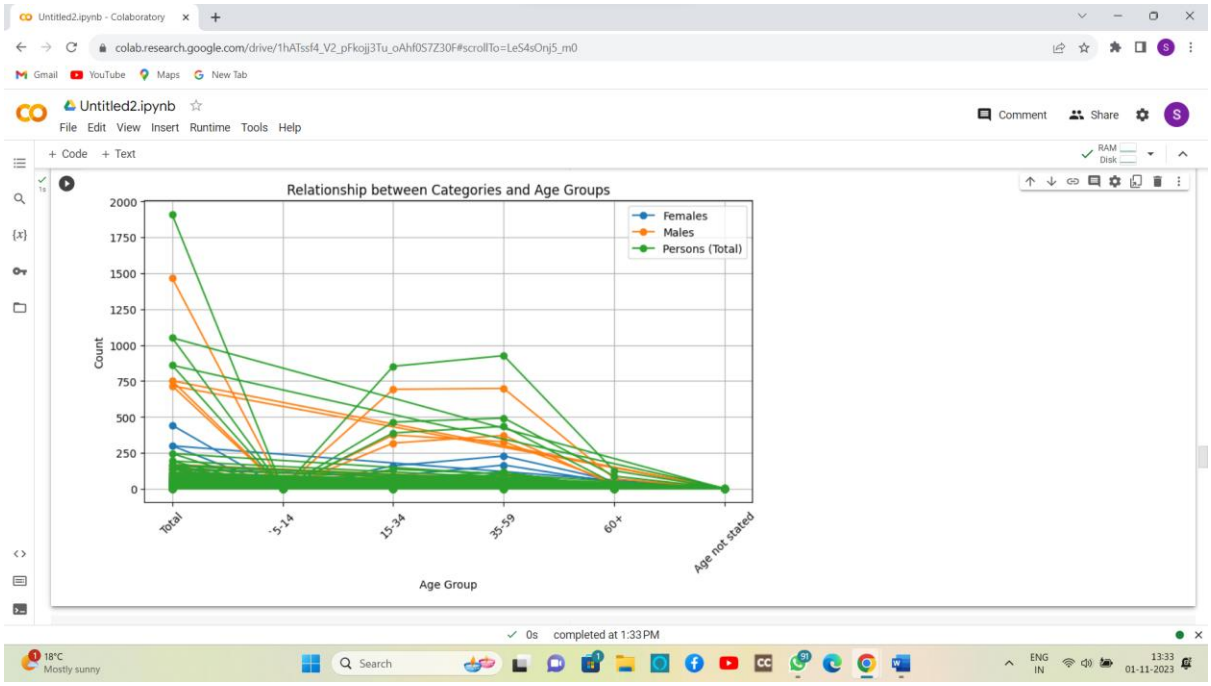
```
industrial_category_counts = df['Industrial Category - D & E - Females'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      393
2       58
6       27
12      13
5       10
3       10
7        7
4        6
10       6
14       6
8        5
9        4
22       4
16       4
27       3
11       3
19       3
18       3
24       2
34       2
28       2
26       2
15       1
21       1
17       1
13       1
442      1
58       1
62       1
33       1
43       1
76       1
44       1
165      1
90       1
299      1
63       1
70       1
143      1
54       1
228      1
160      1
32       1

Name: Industrial Category - D & E - Females, dtype: int64
```

#VISUALIZATION:

```
plt.figure(figsize=(10, 5))

plt.plot(df['Age group'], df['Industrial Category - D & E - Females'],
marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - D & E - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - D & E - Persons'],
marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY F)

```
industrial_category_counts = df['Industrial Category - F -
Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0        99
6        18
2        16
```

```
8         8
12        8
         ..
1263      1
116       1
172       1
494       1
95        1

Name: Industrial Category - F - Persons, Length: 353, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - F - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0        112
6         27
2         14
12        11
20         9
        ...
358        1
1579       1
15         1
736        1
47         1

Name: Industrial Category - F - Males, Length: 318, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - F - Females'].value_counts()
print(industrial_category_counts)
```

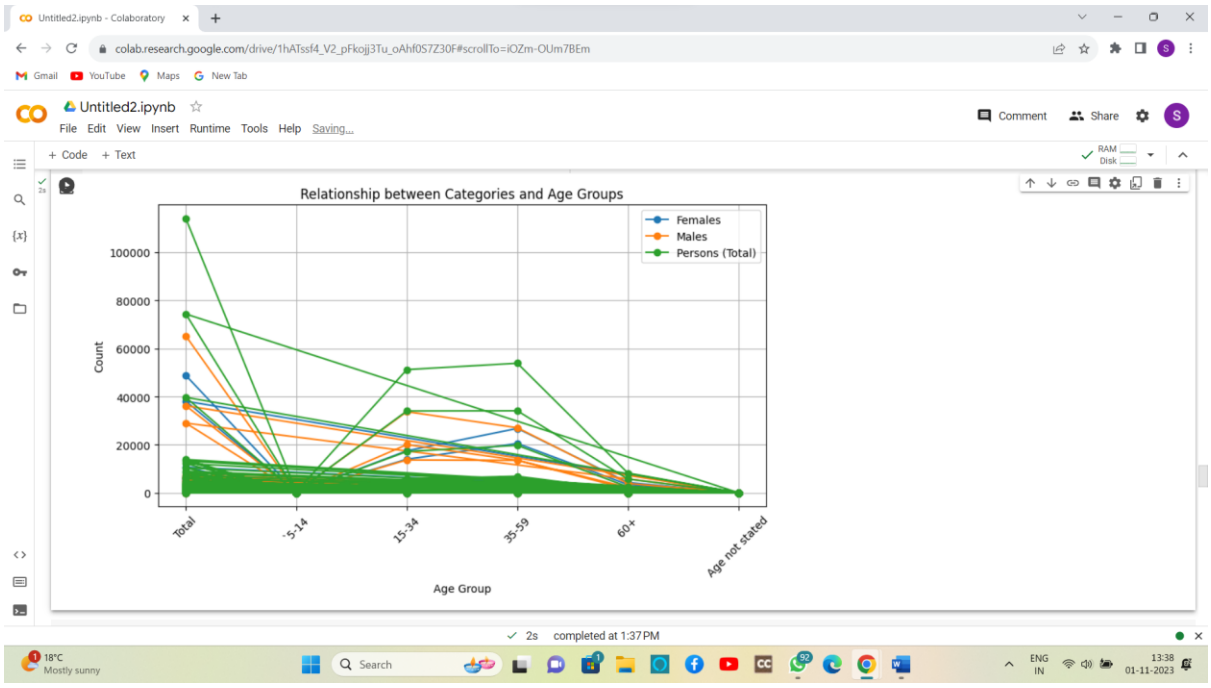
OUTPUT:

```
0        146

2         19
6         14
4          9
3          9
        ...
264        1
136        1
210        1
1931       1
277        1

Name: Industrial Category - F - Females, Length: 288, dtype: int64
```

#VISUALIZATIONS:

```
plt.figure(figsize=(10, 5))

plt.plot(df['Age group'], df['Industrial Category - F - Females'], marker='o', label='Females')
```

```
plt.plot(df['Age group'], df['Industrial Category - F - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - F - Persons'],
marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY H)

```
industrial_category_counts = df['Industrial Category - H -
Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0       199
6        41
3        17
12       10
2         9
       ...
221       1
108       1
126       1
113       1
116       1
```

```
Name: Industrial Category - H - Persons, Length: 217, dtype: int64
```

```python
industrial_category_counts = df['Industrial Category - H - Males'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0      204
6       47
3       17
12      13
18       7
      ...
300      1
124      1
136      1
131      1
29       1

Name: Industrial Category - H - Males, Length: 207, dtype: int64
```

```python
industrial_category_counts = df['Industrial Category - H - Females'].value_counts()
print(industrial_category_counts)
```

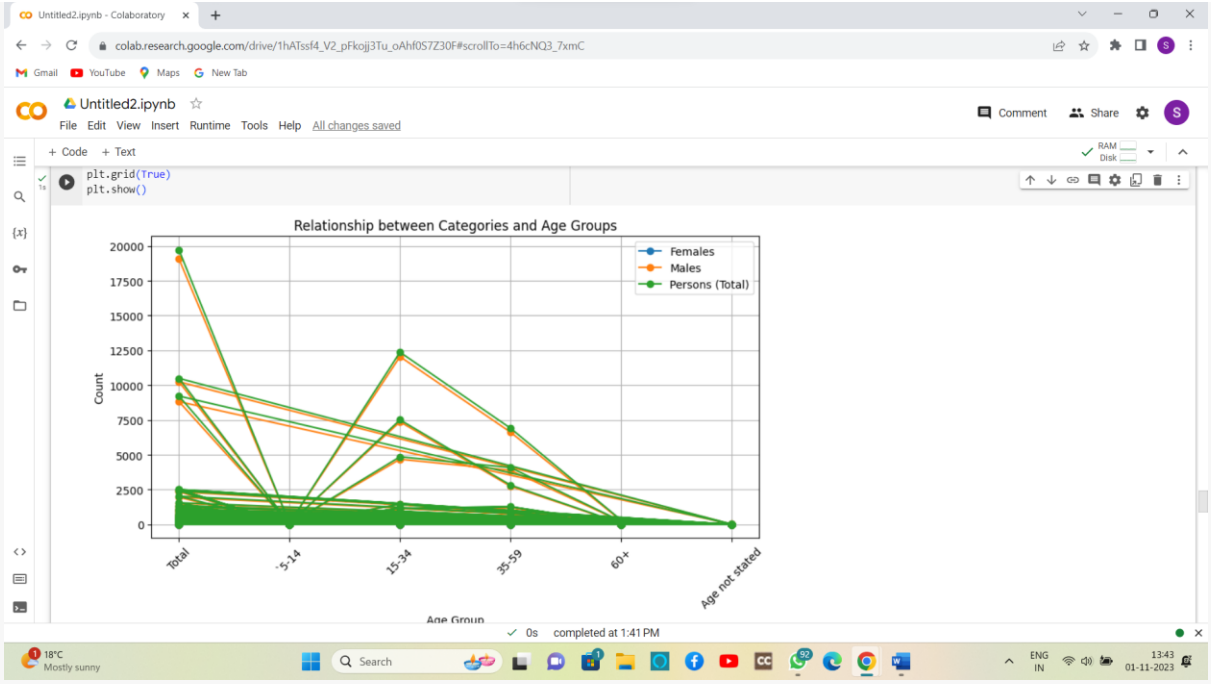
OUTPUT:

```
0      375
6       41
2       41
8       18
16      13
12       9
18       9
3        8
4        7
5        5
10       5
14       5
11       4
9        4
24       4
17       3
110      2
46       2
13       2
7        2
28       2
22       2
41       2
```

```
58        2
27        1
40        1
19        1
20        1
32        1
25        1
36        1
23        1
645       1
43        1
64        1
318       1
288       1
37        1
247       1
142       1
87        1
398       1
176       1
201       1
21        1
79        1
31        1
51        1
33        1
15        1
38        1

Name: Industrial Category - H - Females, dtype: int64
```

#VISUALIZATION:

```
plt.figure(figsize=(10, 5))

plt.plot(df['Age group'], df['Industrial Category - H - Females'],
marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - H - Males'],
marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - H - Persons'],
marker='o', label='Persons (Total)')
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)plt.show()
```




```
#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY I)
industrial_category_counts = df['Industrial Category - I - Persons'].value_counts()
print(industrial_category_counts)
```

```
OUTPUT:
0       230
6        24
8        16
2        13
3        13
       ...
233       1
276       1
533       1
147       1
53        1
Name: Industrial Category - I - Persons, Length: 144, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - I - Males'].value_counts()
print(industrial_category_counts)
```

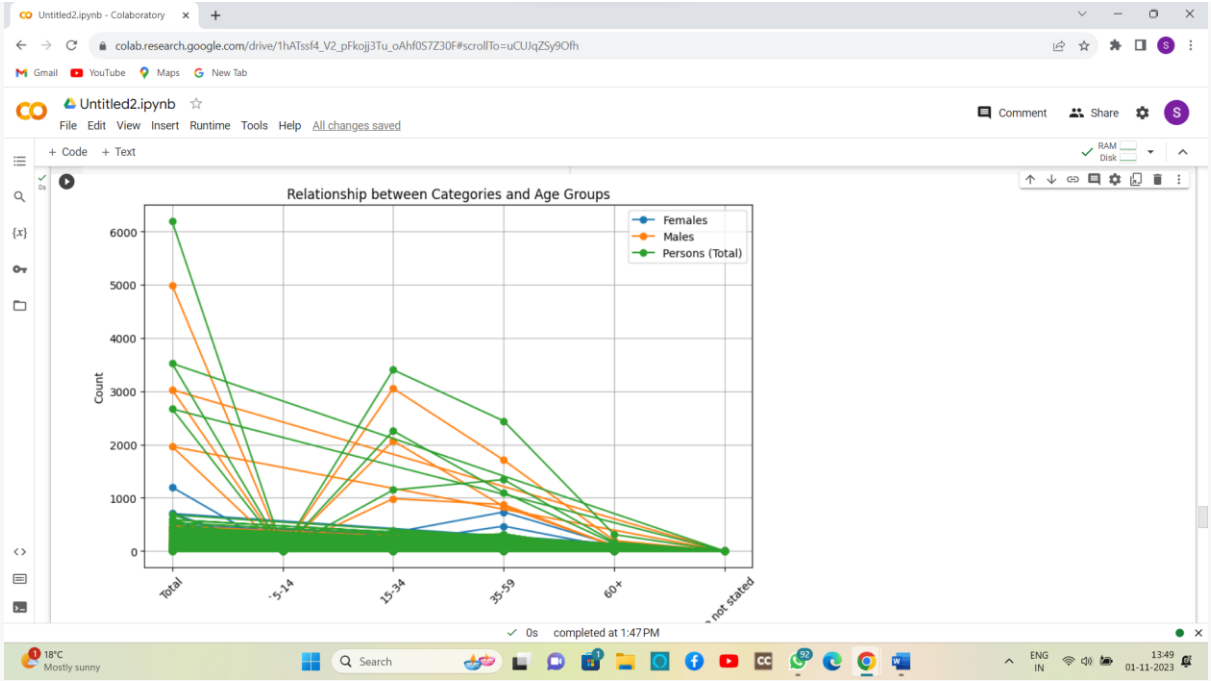
```
OUTPUT:
0       256
6        42
12       19
```

```
3       16
18      12
      ...
71       1
157      1
91       1
191      1
7        1
Name: Industrial Category - I - Males, Length: 133, dtype: int64
```

```
industrial_category_counts = df['Industrial Category - I -
Females'].value_counts()
print(industrial_category_counts)

OUTPUT:
```

```
0      301
6       53
2       37
4       25
8       19
      ...
104      1
106      1
69       1
88       1
41       1
Name: Industrial Category - I - Females, Length: 64, dtype: int64
```

```
VISUALIZATION:
plt.figure(figsize=(10, 6))

plt.plot(df['Age group'], df['Industrial Category - I - Females'],
marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - I - Males'],
marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - I - Persons'],
marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')

plt.title('Relationship between Categories and Age Groups')

plt.legend()
plt.xticks(rotation=45)
plt.grid(True)

plt.show()
```




```
#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY K TO M):

industrial_category_counts = df['Industrial Category - K to M - Persons'].value_counts()
print(industrial_category_counts)

OUTPUT:
0       265
6        32
12       24
18       21
24       14
       ...
139       1
187       1
337       1
13        1
41        1

Name: Industrial Category - K to M - Persons, Length: 114, dtype: int64
industrial_category_counts = df['Industrial Category - K to M - Males'].value_counts()
print(industrial_category_counts)

OUTPUT:
0       268
6        39
12       26
18       24
3        16
       ...
80        1
```

```
187      1
88       1
118      1
23       1
Name: Industrial Category - K to M - Males, Length: 107, dtype: int64
```

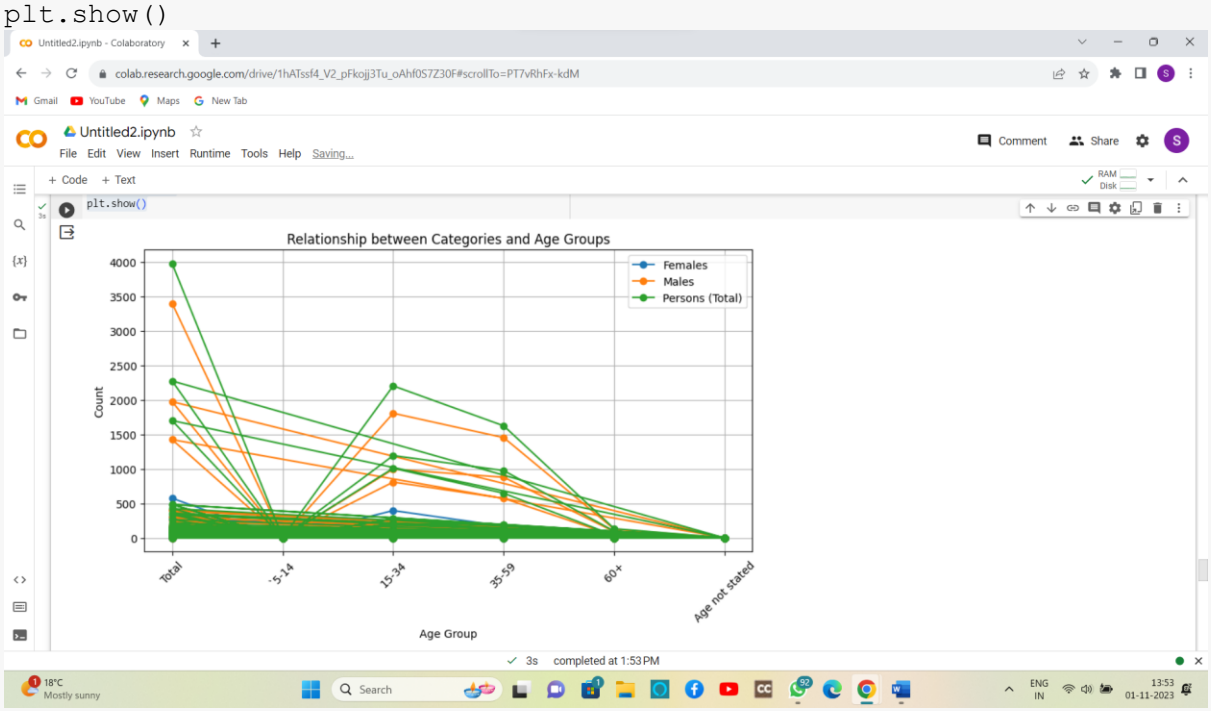
```
industrial_category_counts = df['Industrial Category - K to M - Females'].value_counts()
print(industrial_category_counts)
```

```
OUTPUT:
0      405
6       31
12      23
2       20
8       14
3       11
14      10
9        9
18       9
15       5
5        4
22       4
10       4
26       4
20       3
53       2
11       2
24       2
31       2
79       2
21       2
30       2
7        1
29       1
36       1
33       1
27       1
38       1
23       1
17       1
4        1
28       1
576      1
19       1
40       1
57       1
34       1
97       1
197      1
300      1
74       1
200      1
276      1
171      1
397      1
13       1
```

```
Name: Industrial Category - K to M - Females, dtype: int64
```

VISUALIZATION:

```
plt.figure(figsize=(10, 5))

plt.plot(df['Age group'], df['Industrial Category - K to M - Females'],
marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - K to M - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - K to M - Persons'],
marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```




```
#DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY N TO O)
industrial_category_counts = df['Industrial Category - N to O -
Persons'].value_counts()
print(industrial_category_counts)
```

OUTPUT:

```
0    204
6     28
8     16
2     15
36    11
   ...
444    1
```

```
434    1
935    1
374    1
82     1
Name: Industrial Category - N to O - Persons, Length: 168, dtype: int64
```

[ ]

```
industrial_category_counts = df['Industrial Category - N to O - Males'].value_counts()
print(industrial_category_counts)
```

output

```
0    215
6     53
18    21
12    15
2     14
   ...
307    1
663    1
288    1
370    1
71     1
Name: Industrial Category - N to O - Males, Length: 147, dtype: int64
```

[ ]

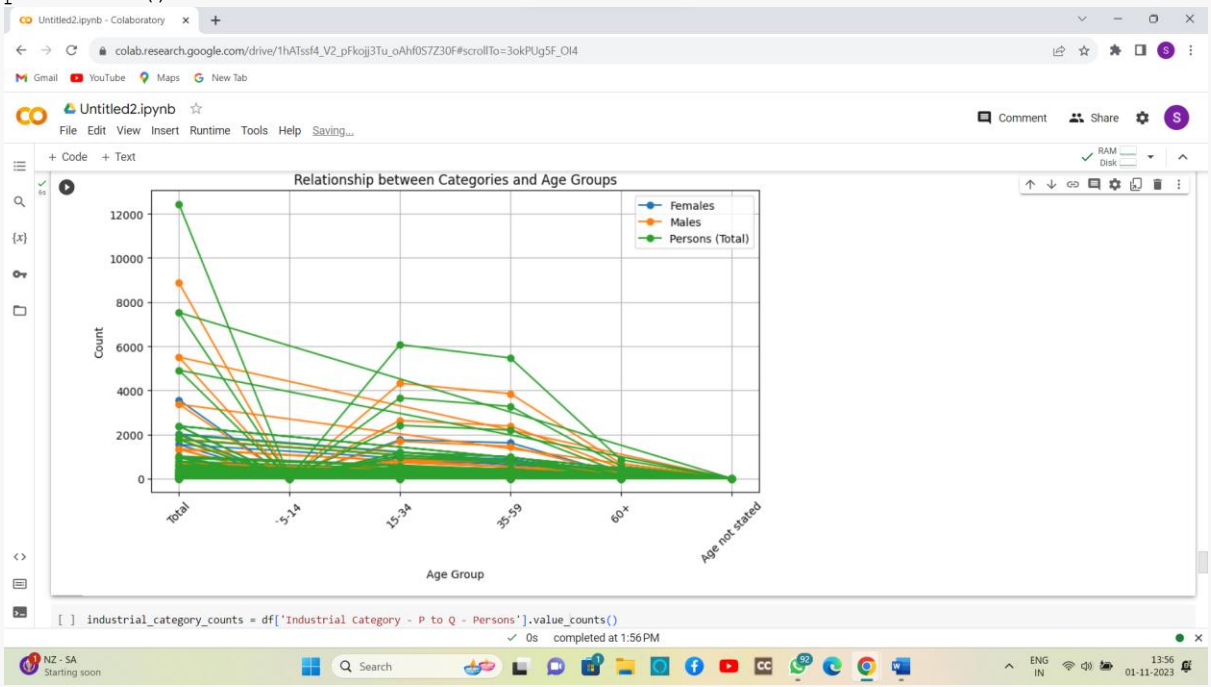
```
industrial_category_counts = df['Industrial Category - N to O -
 Females'].value_counts()
print(industrial_category_counts)
```

output

```
0    265
6     38
2     33
18    16
8     16
   ...
138    1
232    1
194    1
398    1
33     1
Name: Industrial Category - N to O - Females, Length: 104, dtype: int64
```

```
VISUALIZATION:
plt.figure(figsize=(10, 5))

plt.plot(df['Age group'], df['Industrial Category - N to O - Females'],
marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - N to O - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - N to O - Persons'],
marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
```

```
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```




```
DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY P TO Q)
 industrial_category_counts = df['Industrial Category - P to Q - Persons'].value_counts()
 print(industrial_category_counts)
```

output

```
0    205
2     22
6     22
8     11
3     10
   ...
782    1
72     1
100    1
276    1
35     1
Name: Industrial Category - P to Q - Persons, Length: 181, dtype: int64
```

[ ]

```
 industrial_category_counts = df['Industrial Category - P to Q - Males'].value_counts()
 print(industrial_category_counts)
```

output

```
0    262
6     34
12    16
18    15
20    15
   ...
```

187 1
33 1
113 1
149 1
39 1
Name: Industrial Category - P to Q - Males, Length: 111, dtype: int64

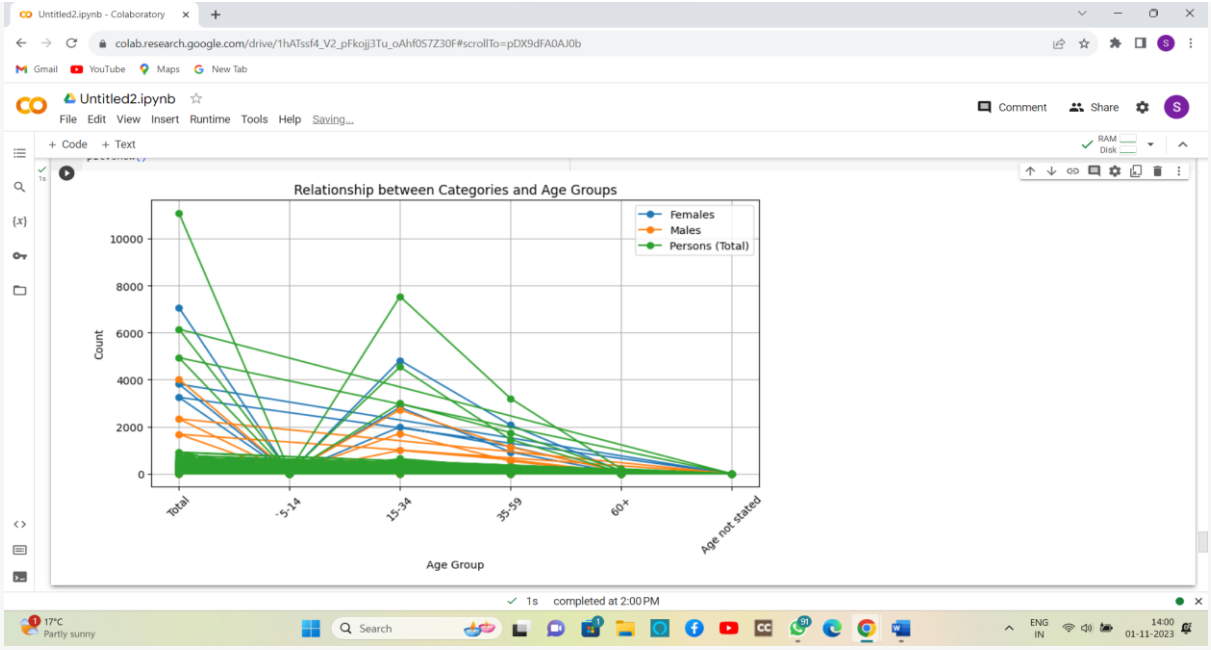
[ ]

```
industrial_category_counts = df['Industrial Category - P to Q - Females'].value_counts()
print(industrial_category_counts)
```

output

0 233
2 25
6 21
18 12
3 10
...
332 1
456 1
250 1
246 1
23 1
Name: Industrial Category - P to Q - Females, Length: 158, dtype: int64

```
VISUALIZATION:
plt.figure(figsize=(10, 5))
plt.plot(df['Age group'], df['Industrial Category - P to Q - Females'],
marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - P to Q - Males'],
marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - P to Q - Persons'],
marker='o', label='Persons (Total)')
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```




```
DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY  R TO U HHI)
```

```
industrial_category_counts = df['Industrial Category - R to U - HHI - Persons'].value_counts()
print(industrial_category_counts)
```

```
output:
0    150
6     21
8     15
14    11
26    11
    ...
340    1
746    1
185    1
400    1
154    1
Name: Industrial Category - R to U - HHI - Persons, Length: 204, dtype: int64
```

[ ]

```
industrial_category_counts = df['Industrial Category - R to U - HHI - Males'].value_counts()
print(industrial_category_counts)
```

```
output
0    214
6     55
12    27
14    15
18    15
    ...
131    1
293    1
98     1
154    1
73     1
Name: Industrial Category - R to U - HHI - Males, Length: 120, dtype: int64
```

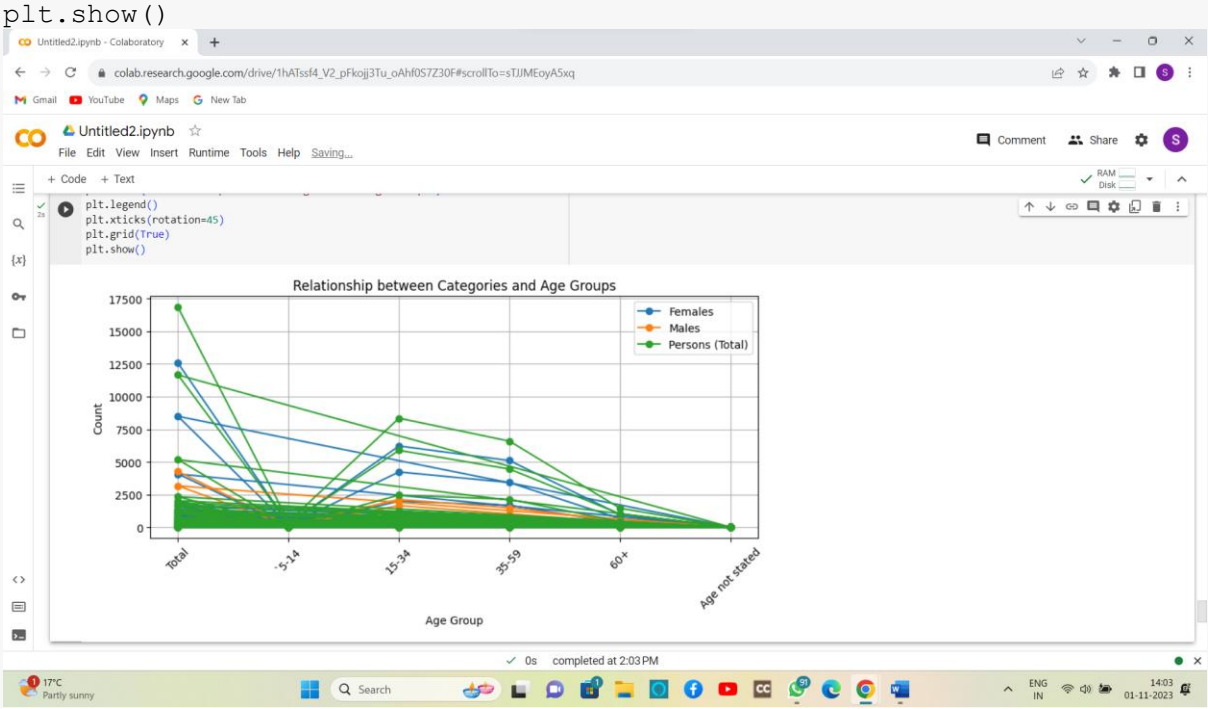
[ ]

```
industrial_category_counts = df['Industrial Category - R to U - HHI - Females'].value_counts()
print(industrial_category_counts)
```

output

```
0    158
6     24
8     21
2     15
3     13
   ...
540    1
139    1
157    1
336    1
118    1
Name: Industrial Category - R to U - HHI - Females, Length: 187, dtype: int64
```

```
VISUALIZATION:
plt.figure(figsize=(10, 4))

plt.plot(df['Age group'], df['Industrial Category - R to U - HHI - Females'], marker='o', label='Females')
plt.plot(df['Age group'], df['Industrial Category - R to U - HHI - Males'], marker='o', label='Males')
plt.plot(df['Age group'], df['Industrial Category - R to U - HHI - Persons'], marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



```
DEMOGRAPHIC ANALYSIS(INDUSTRIAL CATEGORY R TO U NON HHI)
```

```
industrial_category_counts = df['Industrial Category - R to U - Non HHI - Persons'].value_counts()
print(industrial_category_counts)
```

```
output:
    62
6     14
2     10
68     5
12     5
     ..
88     1
500    1
258    1
1154   1
81     1
Name: Industrial Category - R to U - Non HHI - Persons, Length: 397, dtype: int64
```

[ ]

```
industrial_category_counts = df['Industrial Category - R to U - Non HHI - Males'].value_counts()
print(industrial_category_counts)
```

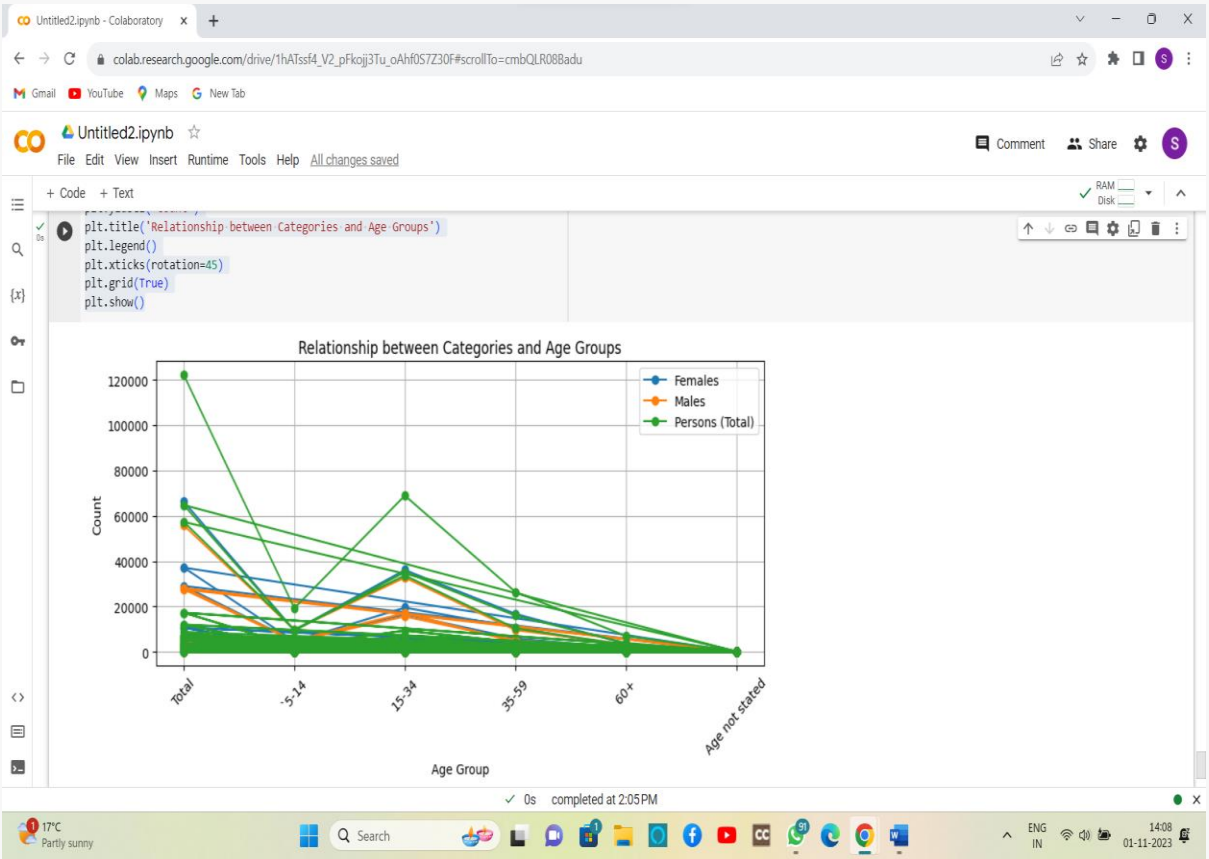
```
output
0     81
12    10
6     10
3      7
60     7
     ..
102    1
317    1
69     1
886    1
247    1
Name: Industrial Category - R to U - Non HHI - Males, Length: 314, dtype: int64
```

[ ]

```
industrial_category_counts = df['Industrial Category - R to U - Non HHI - Females'].value_counts()
print(industrial_category_counts)
```

```
output
0     71
6     12
2     11
32     8
30     8
     ..
99     1
39     1
962    1
548    1
774    1
Name: Industrial Category - R to U - Non HHI - Females, Length: 342, dtype: int64
```

```
VISUALIZATION:
plt.figure(figsize=(10, 4))

plt.plot(df['Age group'], df['Industrial Category - R to U - Non HHI - Females'], marker='o', label='Females')

plt.plot(df['Age group'], df['Industrial Category - R to U - Non HHI - Males'], marker='o', label='Males')

plt.plot(df['Age group'], df['Industrial Category - R to U - Non HHI - Persons'], marker='o', label='Persons (Total)')

plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Relationship between Categories and Age Groups')

plt.legend()
plt.xticks(rotation=45)
plt.grid(True)

plt.show()
```



CONCLUSION:

The TN Marginal Workers Assessment project has found that marginal workers in Tamil Nadu are mostly male, over 35 years old, have an educational level of below secondary school, and are employed in the agricultural sector. Scheduled caste marginal workers are more likely to be female and have a lower educational level than non-scheduled caste marginal workers.

The government can support marginal workers in Tamil Nadu by providing vocational training and job placement assistance, financial assistance, education, healthcare, and social safety nets.