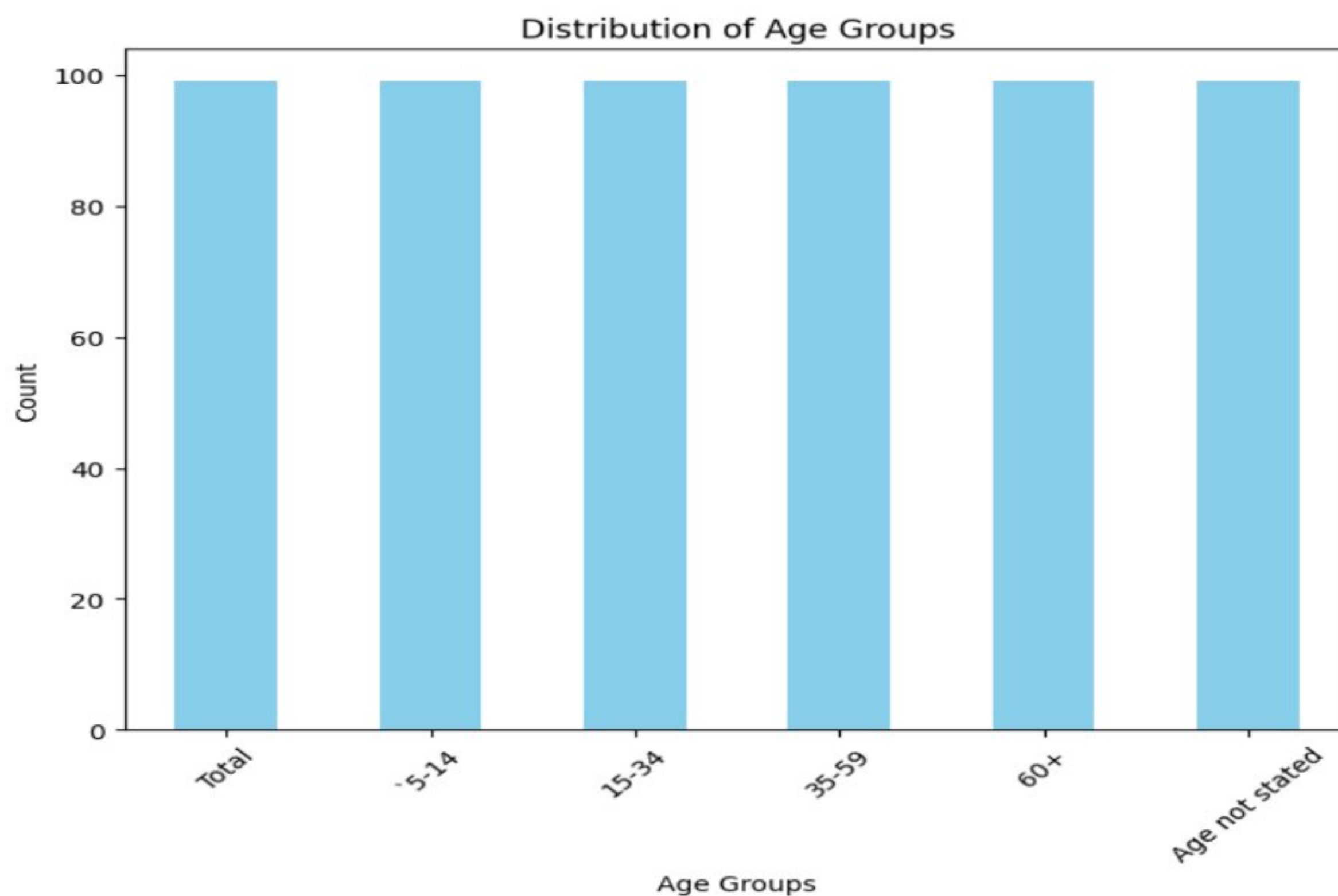


PHASE 4-DEVELOPMENT PART-2

ASSESSMENT OF MARGINAL WORKERS-A SOCIOECONOMIC ANALYSIS

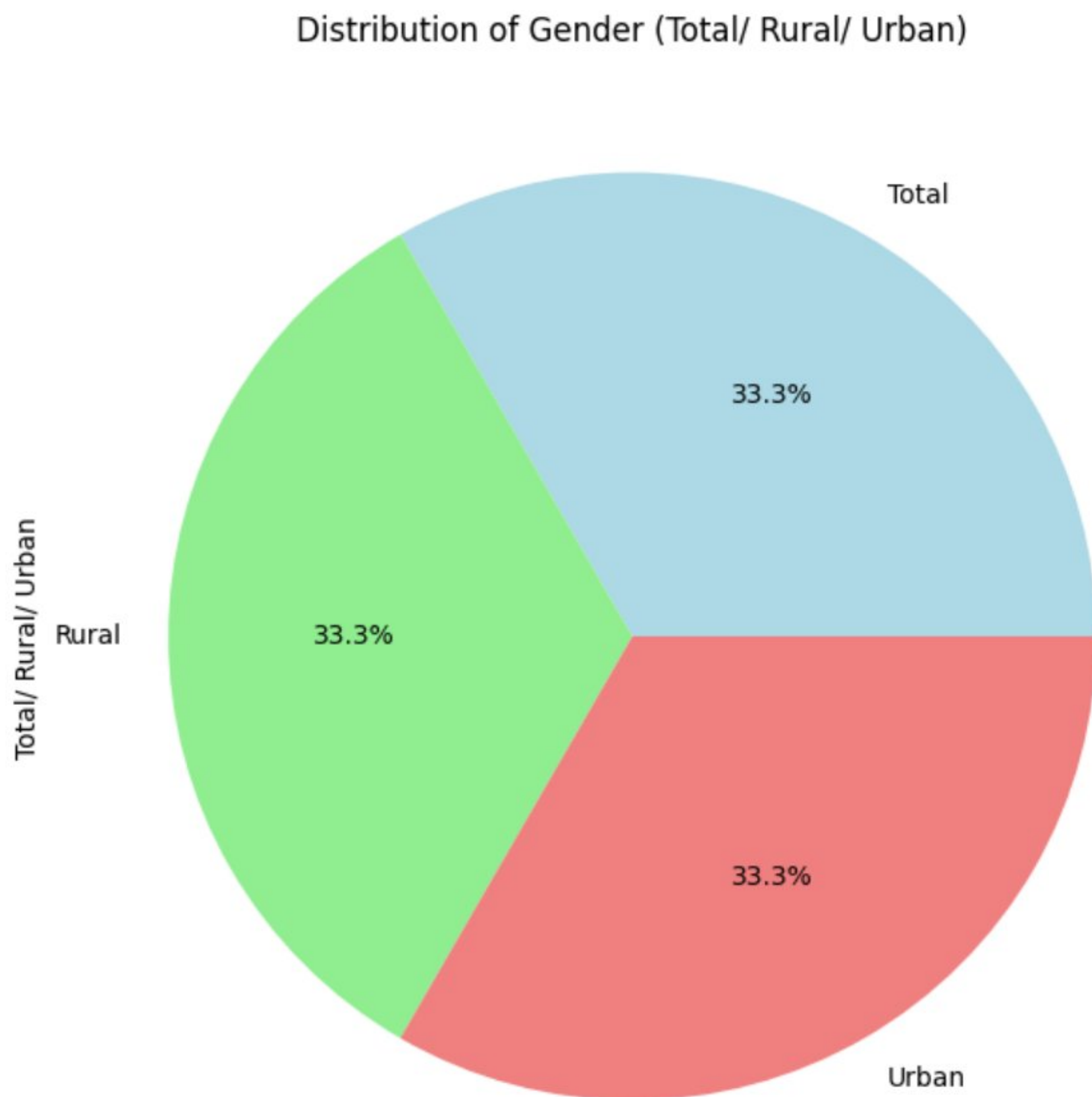
Demographic Analysis and Visualization

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('Marginal Workers.csv')
# Demographic Analysis: Age Groups
age_group_counts = df['Age group'].value_counts()
age_group_counts.plot(kind='bar', figsize=(8, 6), color='skyblue')
plt.title('Distribution of Age Groups')
plt.xlabel('Age Groups')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



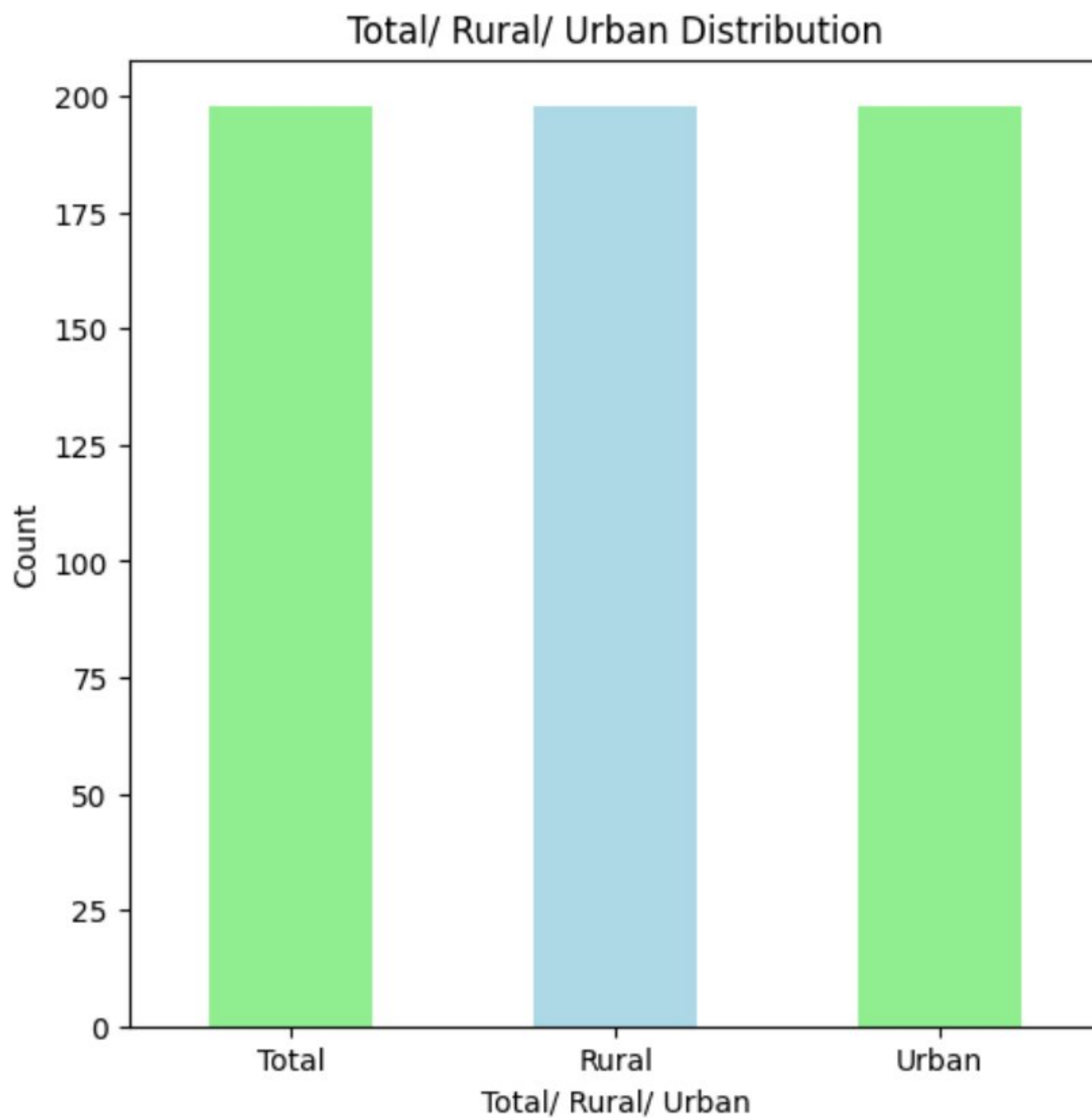
Demographic Analysis: Gender

```
gender_counts = df['Total/ Rural/ Urban'].value_counts()
gender_counts.plot(kind='pie', autopct='%1.1f%%', colors=['lightblue',
'lightgreen', 'lightcoral'], figsize=(8, 8))
plt.title('Distribution of Gender (Total/ Rural/ Urban)')
plt.show()
```



Demographic Analysis: Total/ Rural/ Urban Distribution

```
urban_rural_counts = df['Total/ Rural/ Urban'].value_counts()
urban_rural_counts.plot(kind='bar', color=['lightgreen', 'lightblue'],
figsize=(6,
6))
plt.title("Total/ Rural/ Urban Distribution")
plt.xlabel("Total/ Rural/ Urban")
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.show()
```



Data Aggregation

```
import pandas as pd

df = pd.read_csv('Marginal Workers.csv')

# Group by age group, industrial categories-a-cultivators-persons,and
worked for 3 months or more but less than 6 months - Persons then
aggregate the counts

marginal_workers_distribution = df.groupby(['Age group', 'Industrial
Category - A - Cultivators - Persons', 'Worked for 3 months or more but
less than 6 months - Persons'])['Industrial Category - A - Cultivators -
Males'].sum().reset_index()

# Print the resulting DataFrame (optional)

print(marginal_workers_distribution)

# To save the aggregated data to a new CSV file (optional)

#marginal_workers_distribution.to_csv('marginal_workers_distribution.csv', index=False)
```

```
Age group ... Industrial Category - A - Cultivators - Males

0      15-34 ...                0
1      15-34 ...                3
2      15-34 ...                6
3      15-34 ...                4
4      15-34 ...            8 ..      ... ...      ...
540    `5-14 ...            133
541    `5-14 ...            137
542    `5-14 ...            141
543    `5-14 ...            684
544    `5-14 ...            825

[545 rows x 4 columns]
```

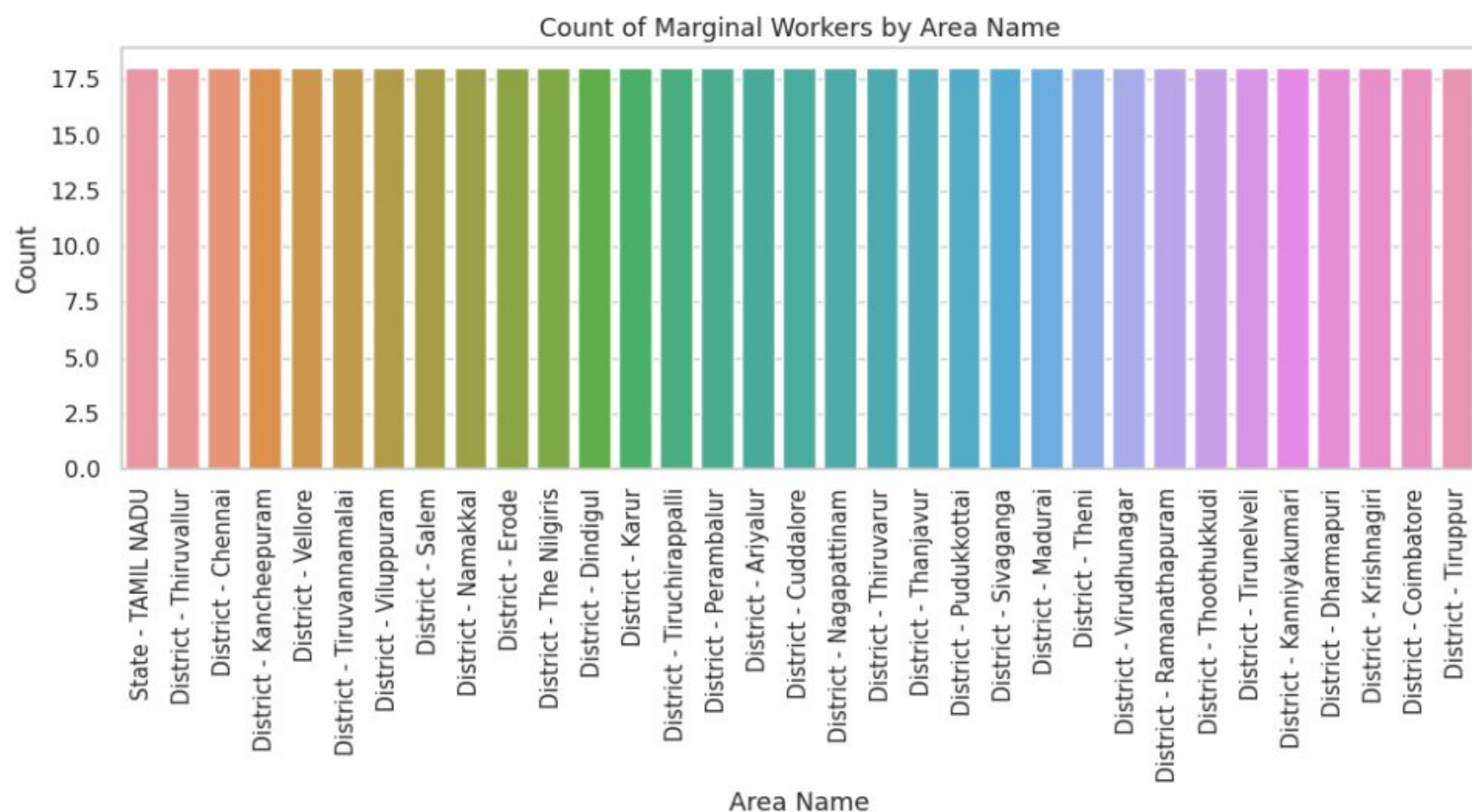
Data Visualization

Countplot:

#1

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

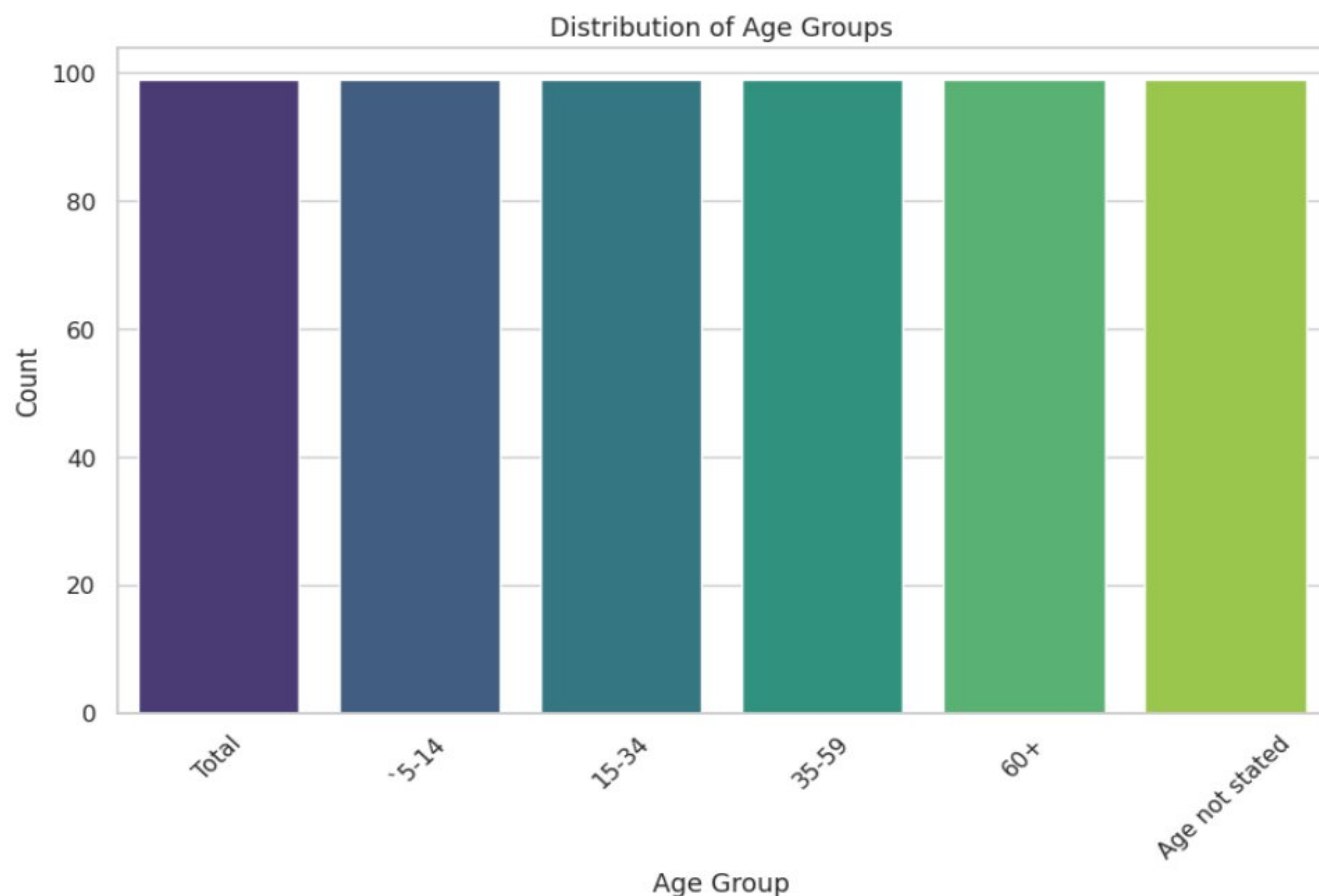
df = pd.read_csv("Marginal Workers.csv")
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x="Area Name")
plt.title("Count of Marginal Workers by Area Name")
plt.xlabel("Area Name")
plt.ylabel("Count")
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



#2

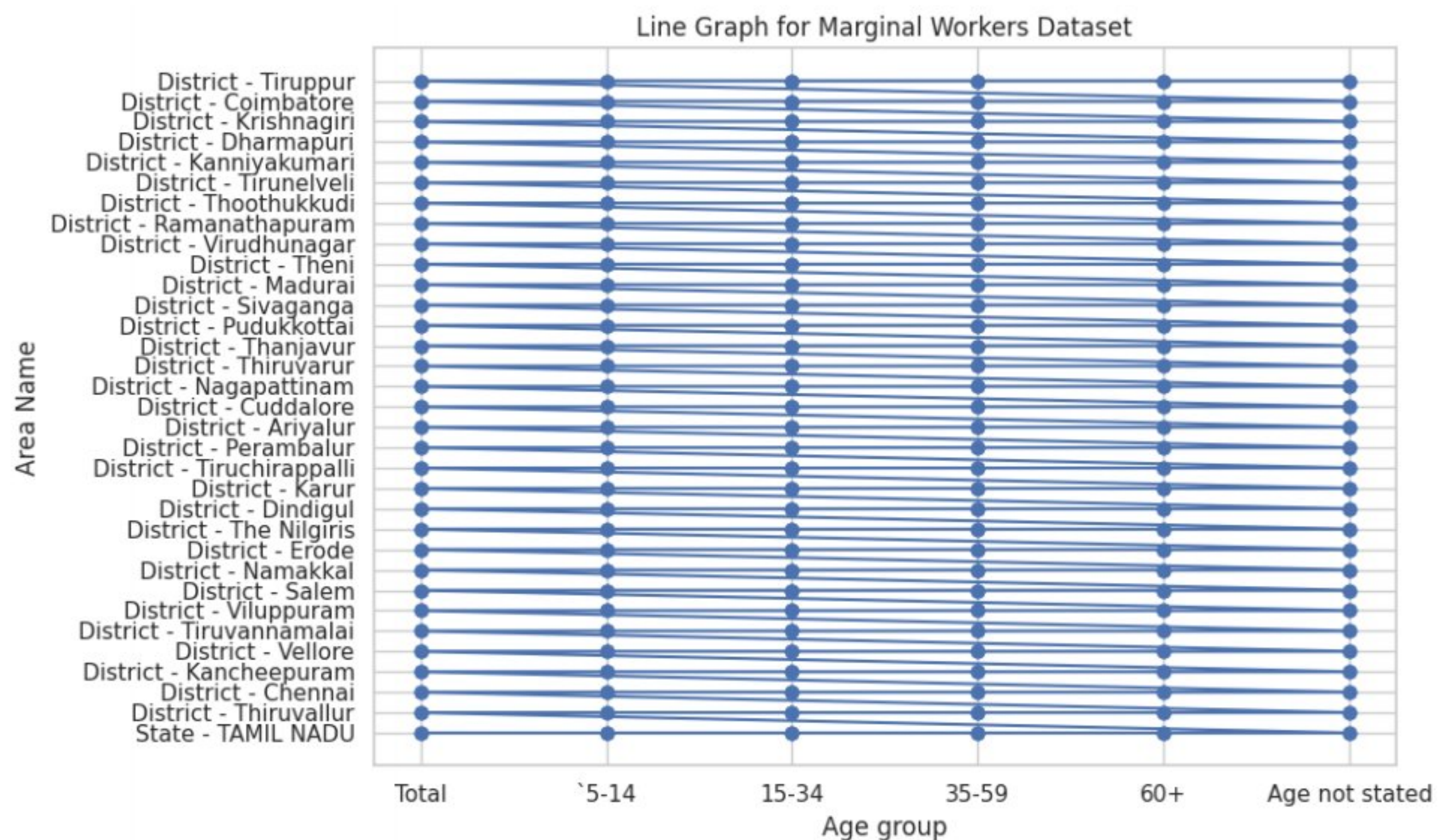
```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

# Assuming 'Age group' is the column representing age groups in your dataset
df = pd.read_csv("Marginal Workers.csv")
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.countplot(x='Age group', data=df, palette='viridis')
plt.title('Distribution of Age Groups')
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



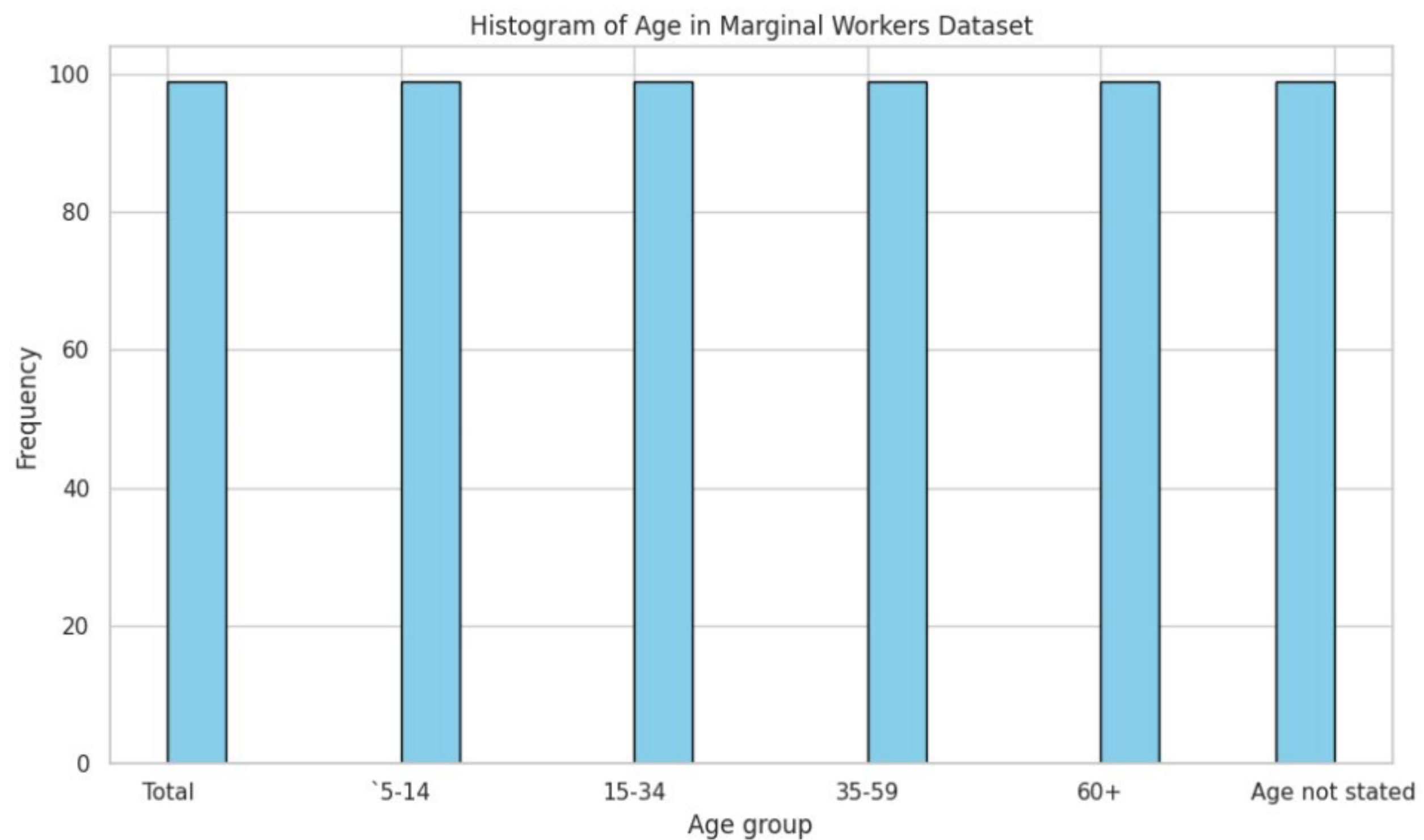
Line Graph:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Marginal Workers.csv")
x_values = df['Age group']
y_values = df['Area Name']
plt.figure(figsize=(10, 6))
plt.plot(x_values, y_values, marker='o', linestyle='-')
plt.title("Line Graph for Marginal Workers Dataset")
plt.xlabel("Age group")
plt.ylabel("Area Name")
plt.grid(True)
plt.tight_layout()
plt.show()
```



Histogram:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Marginal Workers.csv")
data = df['Age group']
plt.figure(figsize=(10, 6))
plt.hist(data, bins=20, color='skyblue', edgecolor='black')
plt.title("Histogram of Age in Marginal Workers Dataset")
plt.xlabel("Age group")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```



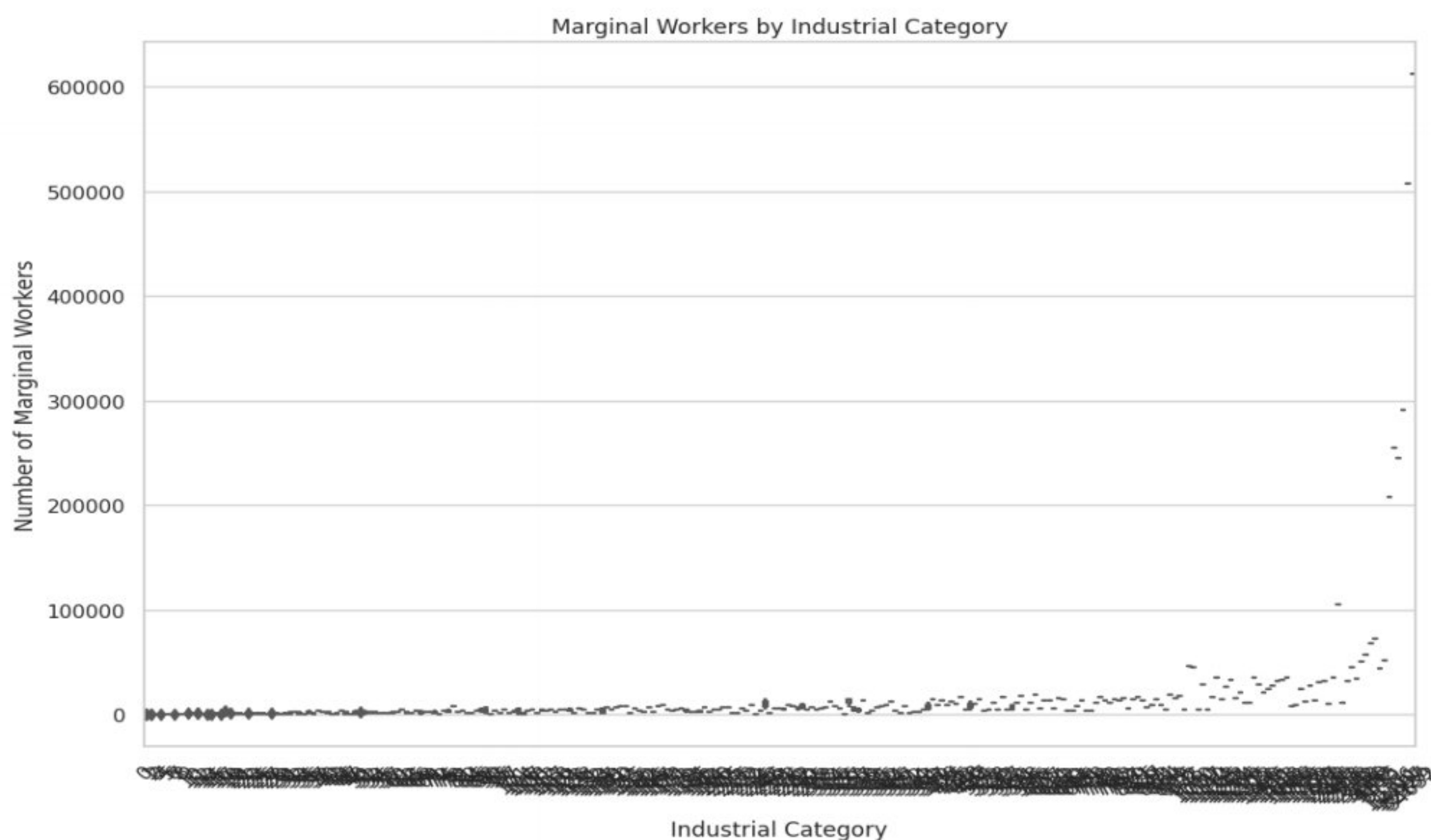
Scatter Plot:

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Marginal Workers.csv")
x_values = df['Worked for 3 months or more but less than 6 months - Persons']
y_values = df['Worked for 3 months or more but less than 6 months - Males']
plt.figure(figsize=(10, 6))
plt.scatter(x_values, y_values, c='blue', alpha=0.5, edgecolors='k')
plt.title("Scatter Plot for Marginal Workers Dataset")
plt.xlabel("Worked for 3 months or more but less than 6 months - Persons")
plt.ylabel("Worked for 3 months or more but less than 6 months - Males")
plt.tight_layout()
plt.show()
```



Boxplot:

```
import seaborn as sns
import matplotlib.pyplot
as plt
import pandas as pd
df = pd.read_csv('Marginal Workers.csv')
plt.figure(figsize=(12, 8))
sns.boxplot(x='Industrial Category - A - Cultivators - Males',
y='Worked for 3 months or more but less than 6 months -
Females', data=df, palette='Set2')
plt.title('Marginal Workers by Industrial Category')
plt.xlabel('Industrial Category')
plt.ylabel('Number of Marginal Workers')
plt.xticks(rotation=45)
plt.show()
```



Violinplot:

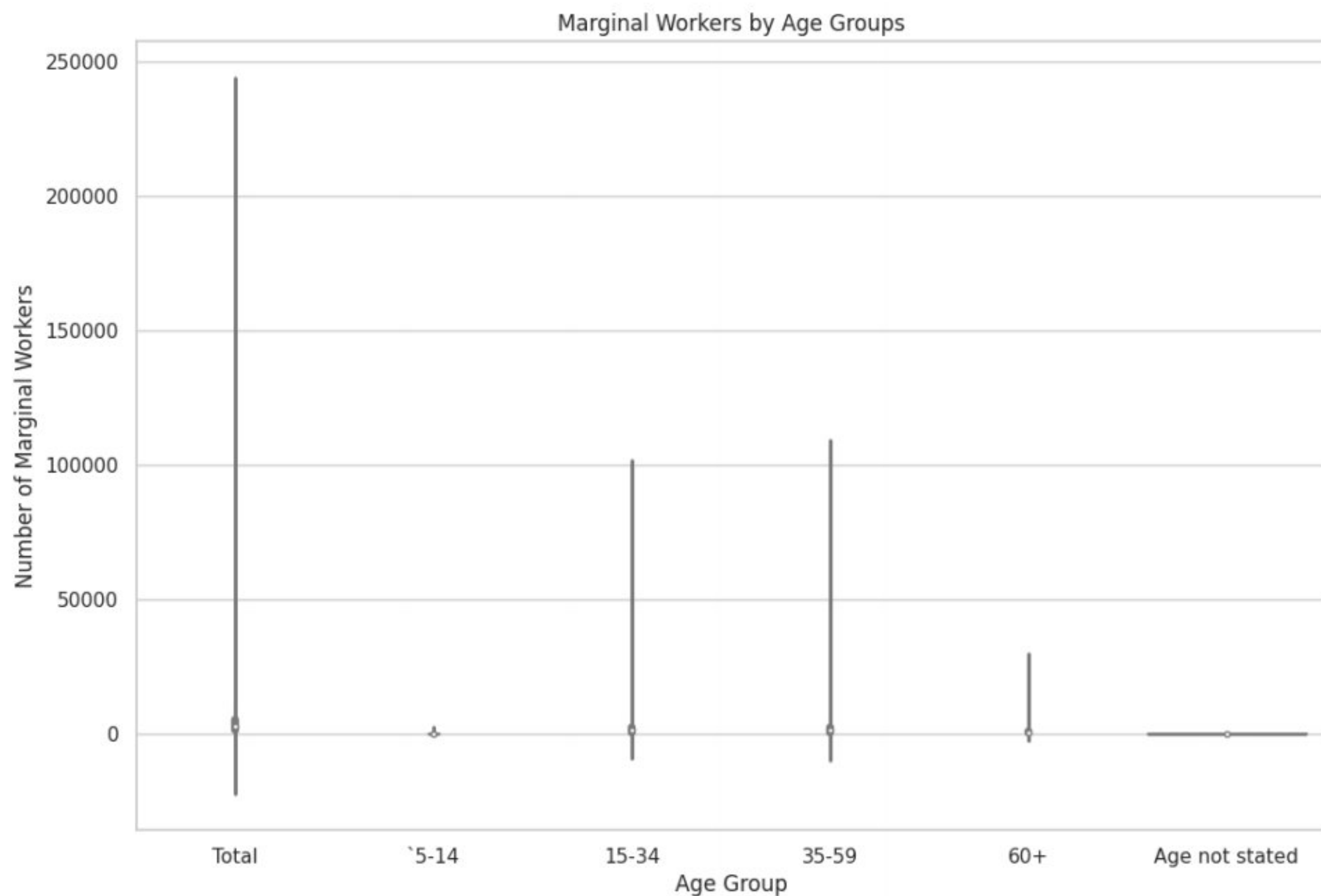
```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Marginal Workers.csv')

plt.figure(figsize=(12, 8))

sns.violinplot(x='Age group', y='Worked for less than 3 months -
Persons', data=df, palette='Pastel1')

plt.title('Marginal Workers by Age Groups')
plt.xlabel('Age Group')
plt.ylabel('Number of Marginal Workers')
plt.show()
```



Heatmap:

```
import seaborn as sns
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

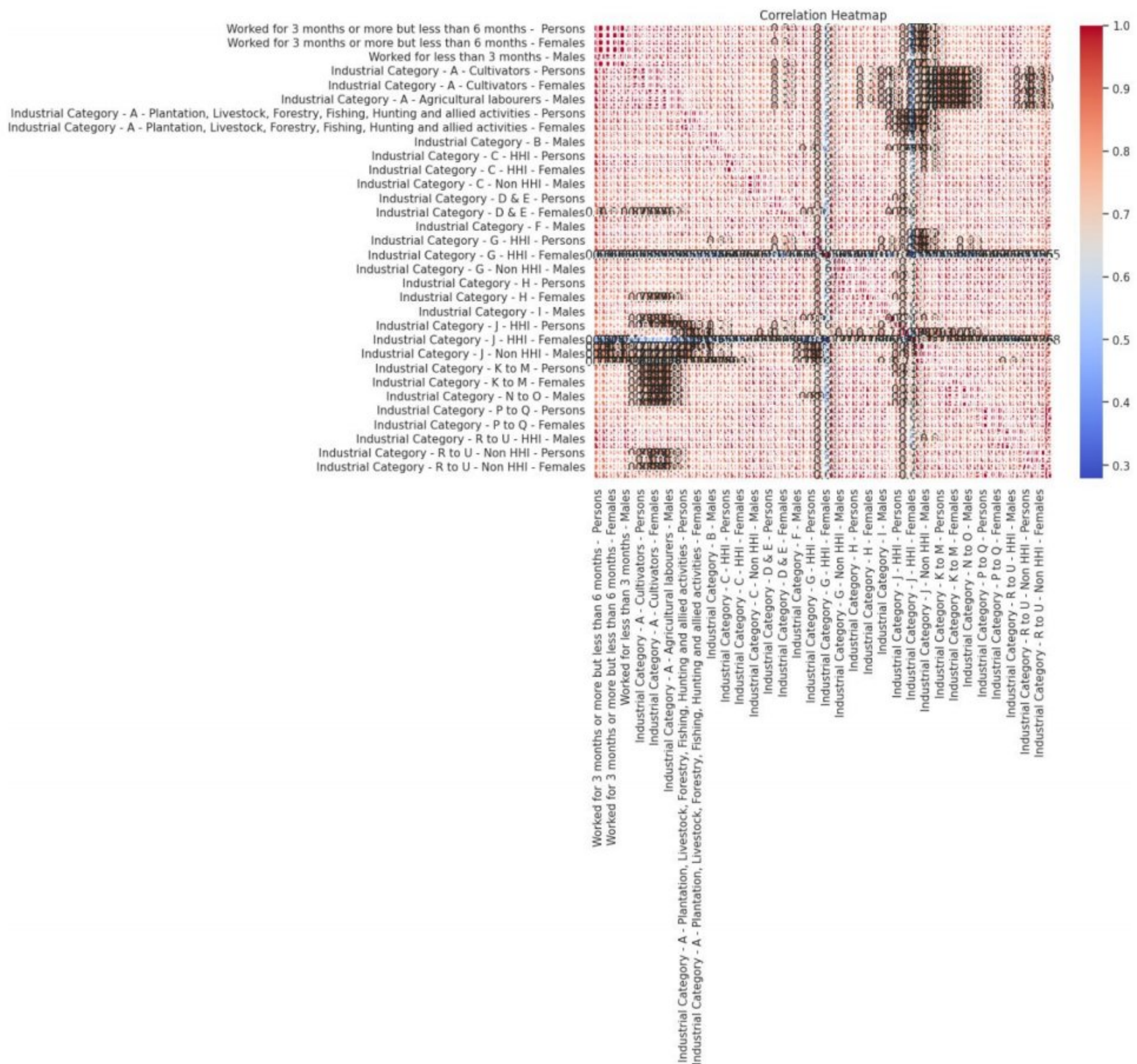
```
df = pd.read_csv('Marginal Workers.csv')
```

```
plt.figure(figsize=(10, 8))
```

```
sns.heatmap(df.corr(), annot=True, cmap='coolwarm',
```

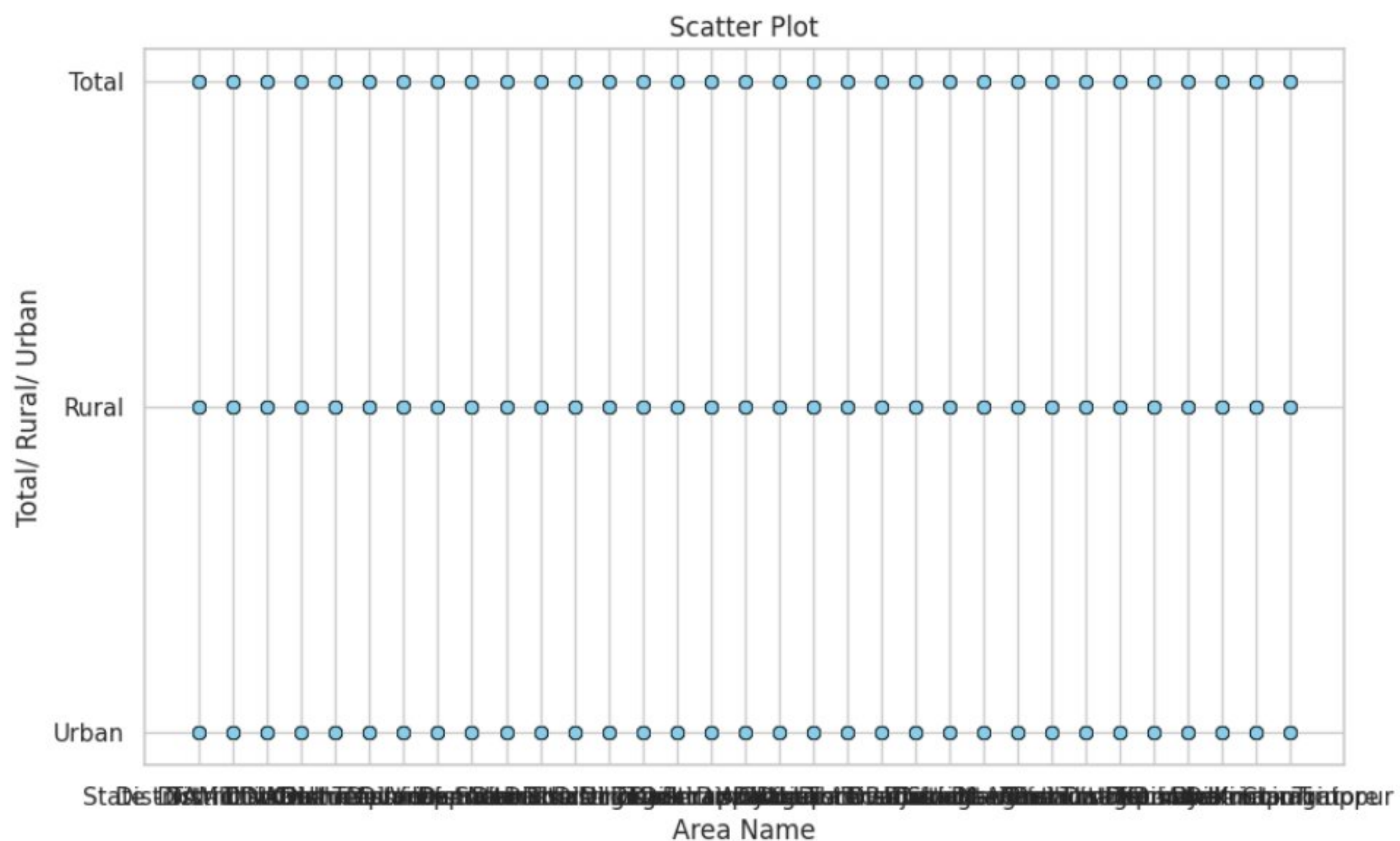
```
fmt=".2f") plt.title('Correlation Heatmap')
```

```
plt.show()
```



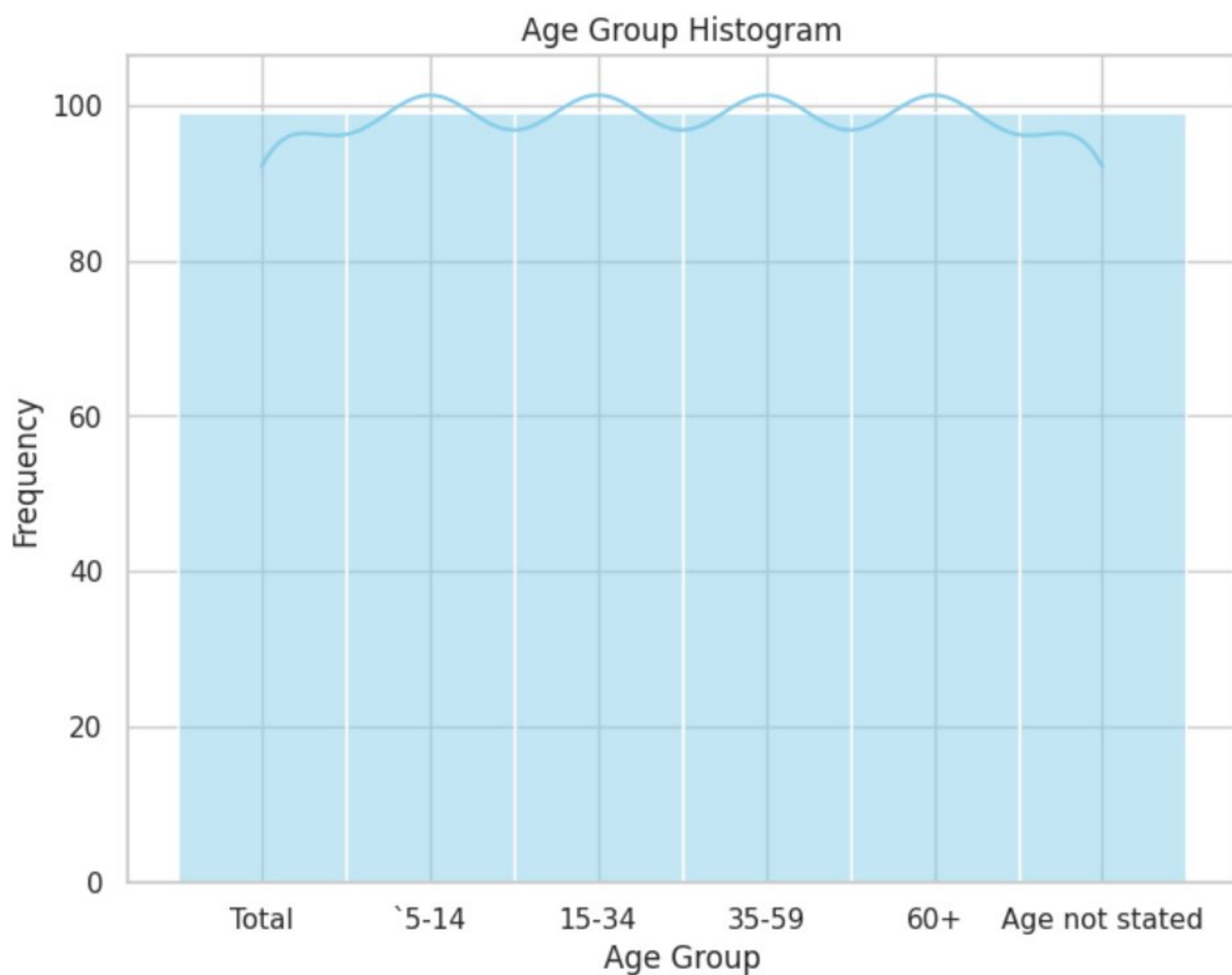
Scatterplot:

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('Marginal Workers.csv')
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Area Name', y='Total/ Rural/ Urban',
data=df, color='skyblue', marker='o', edgecolor='black')
plt.title('Scatter Plot')
plt.xlabel('Area Name')
plt.ylabel('Total/ Rural/ Urban')
plt.show()
```

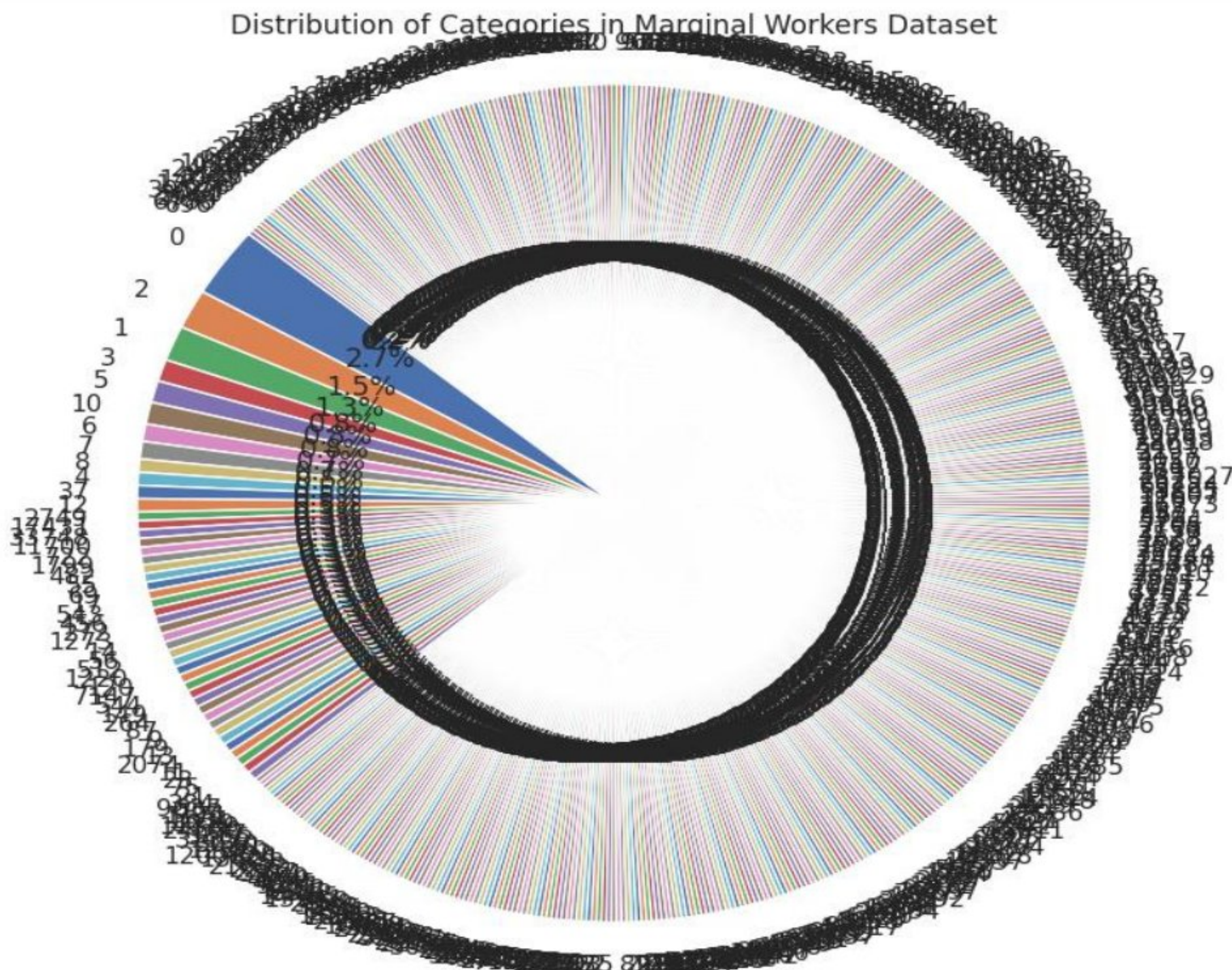


Histplot:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('Marginal Workers.csv')
plt.figure(figsize=(8, 6))
sns.histplot(df['Age group'], bins=20, kde=True,
color='skyblue')
plt.title('Age Group Histogram')
plt.xlabel('Age Group')
plt.ylabel('Frequency')
plt.show()
```

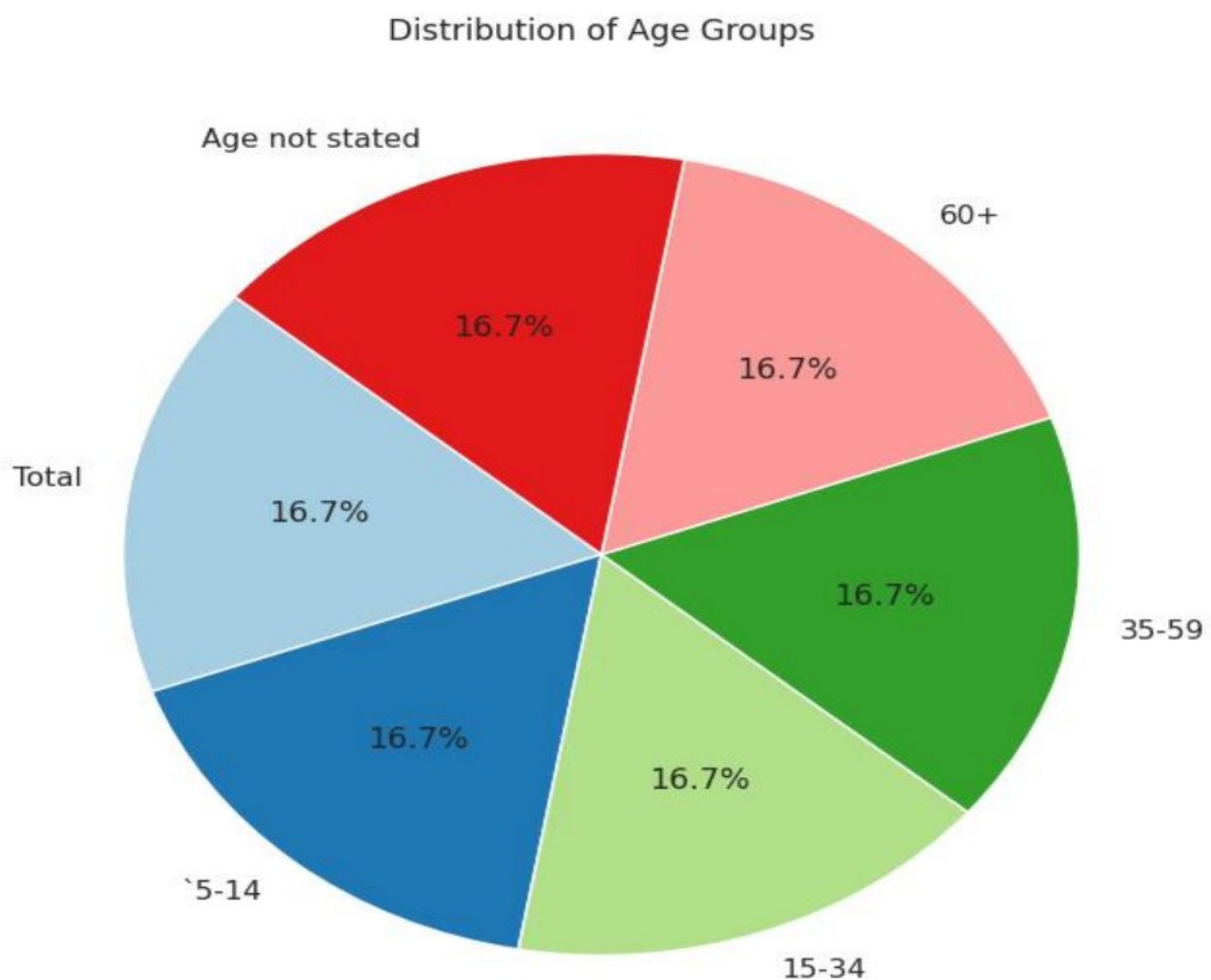


#1



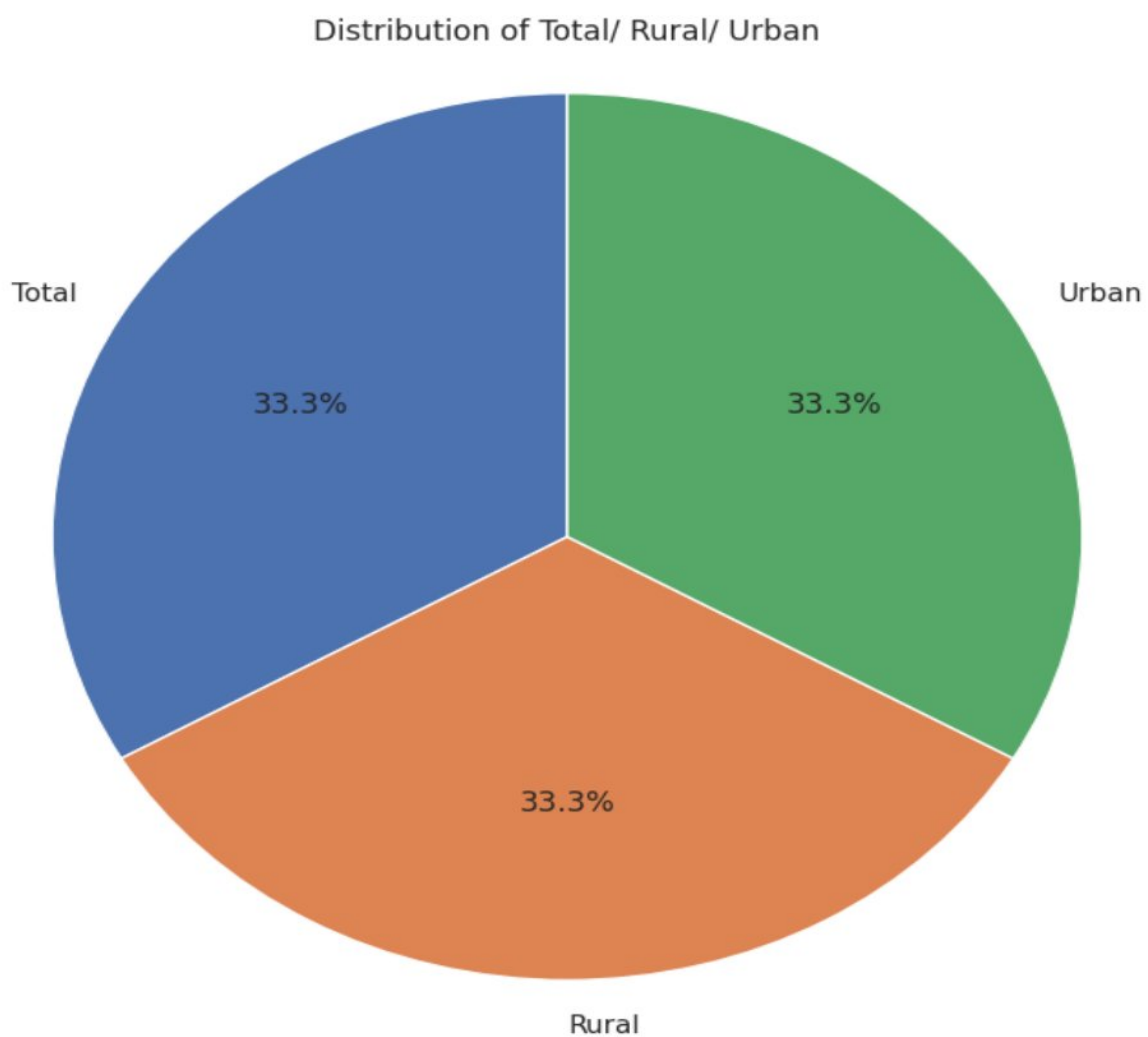
#2

```
import matplotlib.pyplot as plt
import pandas as pd
df=pd.read_csv('Marginal Workers.csv')
age_group_counts = df['Age group'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(age_group_counts, labels=age_group_counts.index,
autopct='%1.1f%%', startangle=140,
colors=plt.cm.Paired.colors)
plt.title('Distribution of Age Groups')
plt.show()
```



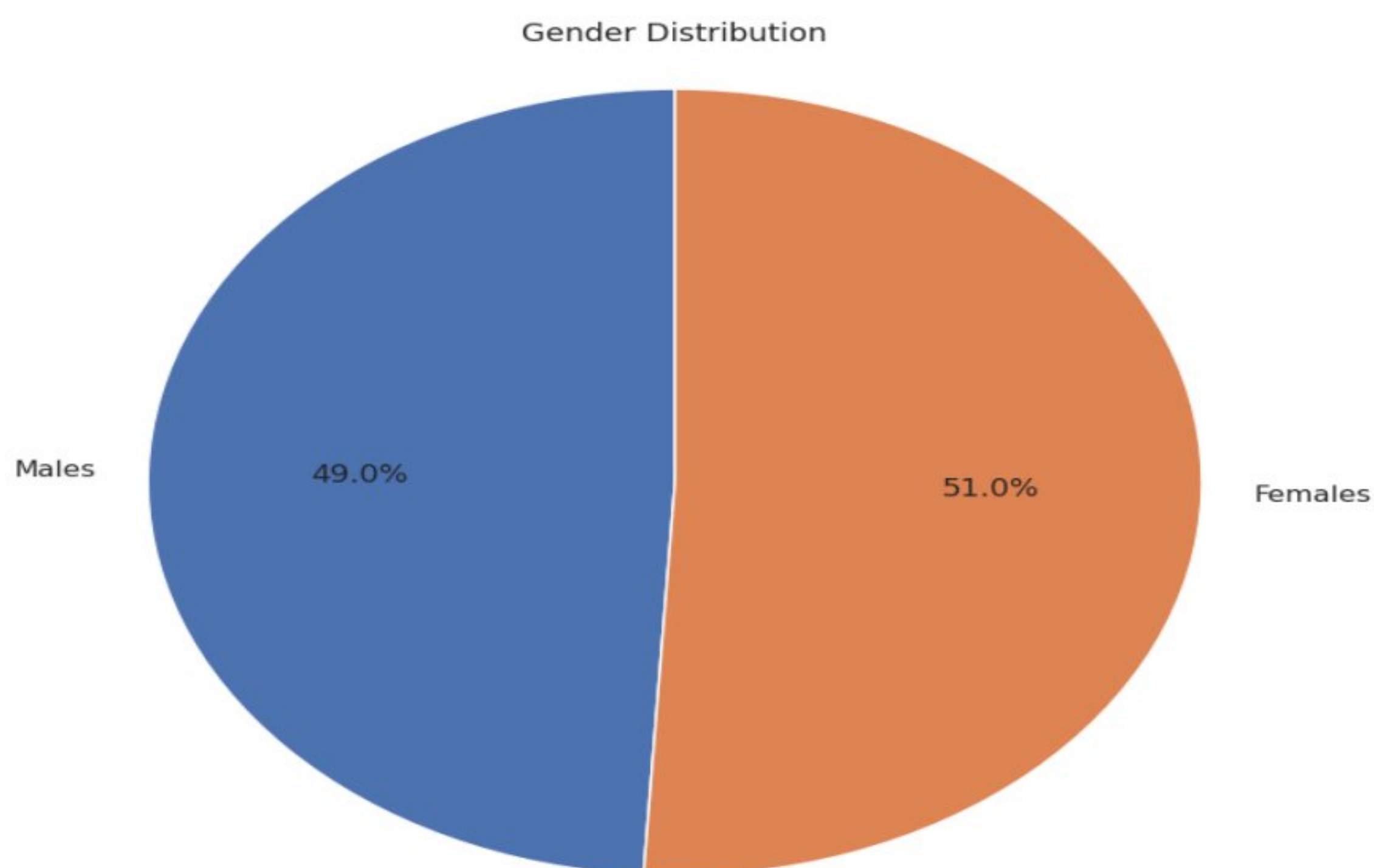
#3

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('Marginal Workers.csv')
urban_counts = df['Total/ Rural/ Urban'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(urban_counts, labels=urban_counts.index,
autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Total/ Rural/ Urban')
plt.axis('equal')
plt.show()
```



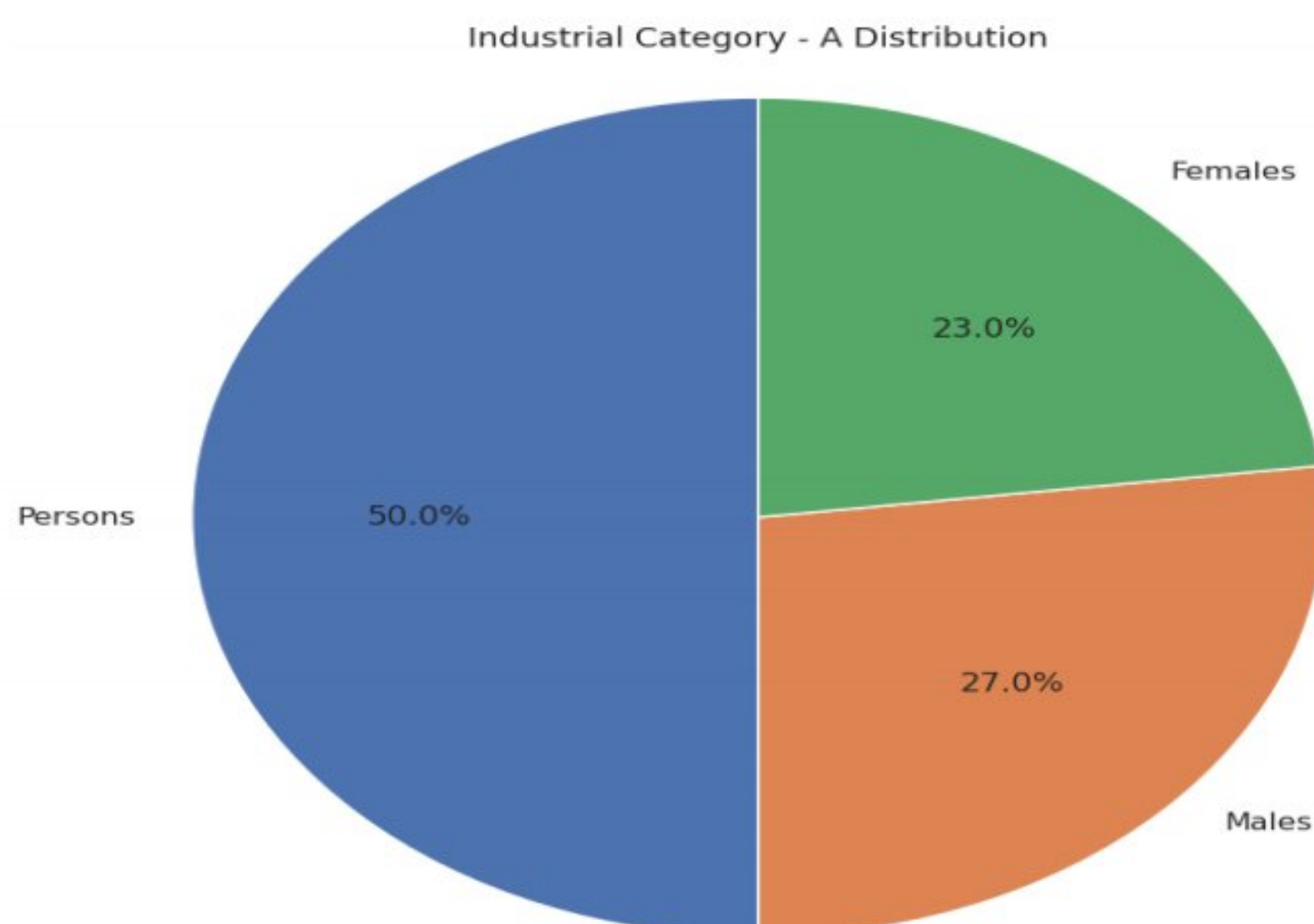
#4

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('Marginal Workers.csv')
male_count = df['Worked for 3 months or more but less than 6
months - Males'].sum()
female_count = df['Worked for 3 months or more but less than
6 months - Females'].sum()
gender_counts = [male_count, female_count]
plt.figure(figsize=(8, 8))
plt.pie(gender_counts, labels=['Males', 'Females'],
autopct='%1.1f%%', startangle=90)
plt.title('Gender Distribution')
plt.axis('equal')
plt.show()
```



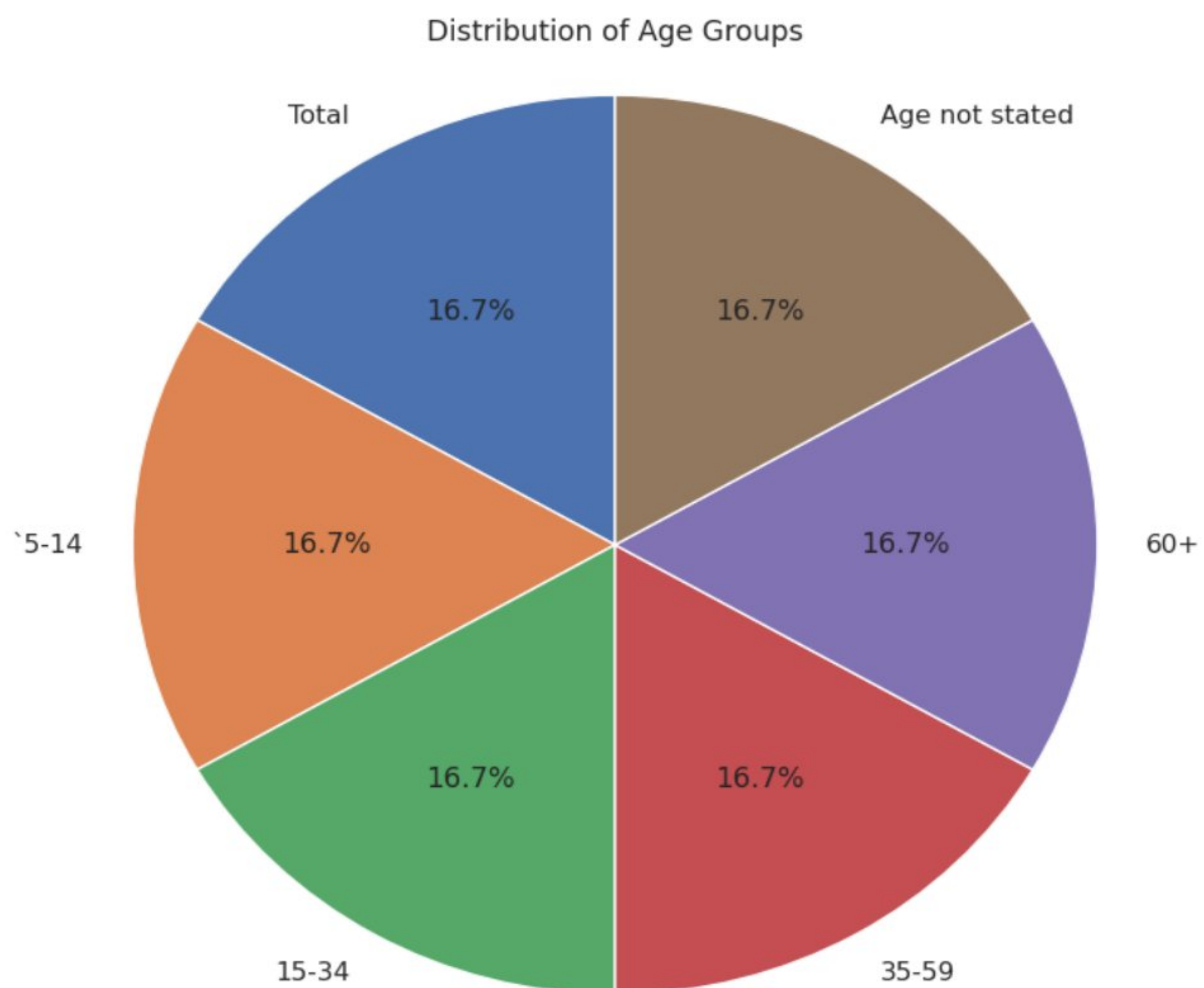
#5

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('Marginal Workers.csv')
persons_count = df['Industrial Category - A - Cultivators -
Persons'].sum()
males_count = df['Industrial Category - A - Cultivators -
Males'].sum()
females_count = df['Industrial Category - A - Cultivators -
Females'].sum()
category_counts = [persons_count, males_count, females_count]
plt.figure(figsize=(8, 8))
plt.pie(category_counts, labels=['Persons', 'Males', 'Females'],
autopct='%1.1f%%', startangle=90)
plt.title('Industrial Category - A Distribution')
plt.axis('equal')
plt.show()
```



#6

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('Marginal Workers.csv')
age_group_counts = df['Age group'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(age_group_counts, labels=age_group_counts.index,
autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Age Groups')
plt.axis('equal')
plt.show()
```




```

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read_csv('Marginal Workers.csv')

print(df.head())

urban_counts = df['Total/ Rural/
Urban'].value_counts()

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

urban_counts.plot(kind='bar', color='skyblue')

plt.title('Distribution of Total/ Rural/ Urban')

plt.xlabel('Category')

plt.ylabel('Count')

plt.xticks(rotation=45)

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.show()

```

