Project Documentation

1.Project objective:

The main objective of the project is to analyze the given csv file big data of marginal workers of Tamil Nadu and to provide in a proper virtual understanding by data visualization techniques like scatter plot ,bar chart, histogram, pie chart etc.. can be used.

2. Analyze approach:

1. Clean and prepare the data.

   This may involve removing any duplicate rows, handling missing values, and converting the data into the appropriate format for your analysis.

2. Calculate the share of marginal workers in the total workforce.

   This can be done by dividing the number of marginal workers by the total number of workers in your dataset.

3.Virtualization types:

1. Create a pie chart.

   A pie chart is a type of chart that shows the proportional relationship between different categories. To create a pie chart for your marginal worker analysis, you will need to divide the pie into slices, each representing a different category of workers. For example, you could divide the pie into slices for marginal workers, non-marginal workers, and unemployed workers.

2. Label the pie slices and add a title.

   The pie slices should be labelled with the category of workers that they represent. The title of the pie chart should be descriptive and informative

3.Other virtualization types:

Some of the other virtualization types are pie chart, bar chart, histogram, pie plot, scatter plot, etc.

3. Code implementation:

We used Python code programming language for the big data analysis purpose to virtualize the given csv file

This code will create a pie chart that shows the distribution of virtualization types among marginal workers. The chart shows that the majority of marginal workers use desktop virtualization. This suggests that desktop virtualization is the most popular type of virtualization among marginal workers.

You can also use a pie chart to analyse the distribution of virtualization types across different industries or occupations. Simply change the categories in the pie chart to the industries or occupations that you want to analyse.

[1]:
```
#importing the libraries in python
import pandas as pd
import matplotlib.pyplot as plt
#import seaborn for using piechart
import seaborn as sns
import numpy as np
```

1 Importing csv file data sets

[2]:
```
# Load the dataset into a Pandas DataFrame
df = pd.read_csv("Downloads/DDW_B06SC_3300_State_TAMIL_NADU-2011.csv")
```

2 checking the data fully filled that is fully true

[3]:
```
#checking the dataset given is null or not
df.isnull()
```

[3]:

	Table Code	State Code	District Code	Area Name	Total/ Rural/ Urban	\
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
..	...	...	...	...	...	
589	False	False	False	False	False	
590	False	False	False	False	False	
591	False	False	False	False	False	
592	False	False	False	False	False	
593	False	False	False	False	False	

	Age group	Worked for 3 months or more but less than 6 months -	Persons	\

0	False	False
1	False	False
2	False	False
3	False	False

```
4        False                                                  False
..         ...                                                  ...
589      False                                                  False
590      False                                                  False
591      False                                                  False
592      False                                                  False
593      False                                                  False

     Worked for 3 months or more but less than 6 months - Males  \
0                                                        False
1                                                        False
2                                                        False
3                                                        False
4                                                        False
..                                                         ...
589                                                      False
590                                                      False
591                                                      False
592                                                      False
593                                                      False

     Worked for 3 months or more but less than 6 months - Females  \
0                                                        False
1                                                        False
2                                                        False
3                                                        False
4                                                        False
..                                                         ...
589                                                      False
590                                                      False
591                                                      False
592                                                      False
593                                                      False

     Worked for less than 3 months - Persons  ...  \
0                                      False  ...
1                                      False  ...
2                                      False  ...
3                                      False  ...
4                                      False  ...
..                                       ...  ...
589                                    False  ...
590                                    False  ...
591                                    False  ...
592                                    False  ...
593                                    False  ...
```

	Industrial Category - N to O - Females	\
0	False	
1	False	
2	False	
3	False	
4	False	
..	...	
589	False	
590	False	
591	False	
592	False	
593	False	

	Industrial Category - P to Q - Persons	\
0	False	
1	False	
2	False	
3	False	
4	False	
..	...	
589	False	
590	False	
591	False	
592	False	
593	False	

	Industrial Category - P to Q - Males	\
0	False	
1	False	
2	False	
3	False	
4	False	
..	...	
589	False	
590	False	
591	False	
592	False	
593	False	

	Industrial Category - P to Q - Females	\
0	False	
1	False	
2	False	
3	False	
4	False	
..	...	
589	False	

```
590                                    False
591                                    False
592                                    False
593                                    False

     Industrial Category - R to U - HHI - Persons  \
0                                          False
1                                          False
2                                          False
3                                          False
4                                          False
..                                           ...
589                                        False
590                                        False
591                                        False
592                                        False
593                                        False

     Industrial Category - R to U - HHI - Males  \
0                                        False
1                                        False
2                                        False
3                                        False
4                                        False
..                                         ...
589                                      False
590                                      False
591                                      False
592                                      False
593                                      False

     Industrial Category - R to U - HHI - Females  \
0                                          False
1                                          False
2                                          False
3                                          False
4                                          False
..                                           ...
589                                        False
590                                        False
591                                        False
592                                        False
593                                        False

     Industrial Category - R to U - Non HHI - Persons  \
0                                              False
1                                              False
```

```
2                                                    False
3                                                    False
4                                                    False
..                                                     ...
589                                                  False
590                                                  False
591                                                  False
592                                                  False
593                                                  False

     Industrial Category - R to U - Non HHI - Males  \
0                                                    False
1                                                    False
2                                                    False
3                                                    False
4                                                    False
..                                                     ...
589                                                  False
590                                                  False
591                                                  False
592                                                  False
593                                                  False

     Industrial Category - R to U - Non HHI - Females
0                                                    False
1                                                    False
2                                                    False
3                                                    False
4                                                    False
..                                                     ...
589                                                  False
590                                                  False
591                                                  False
592                                                  False
593                                                  False

[594 rows x 69 columns]
```

3 Fetching and describe the data

[4]:
```
l1=tuple([df['Worked for 3 months or more but less than 6 months -␣
↪Females'],df['Worked for 3 months or more but less than 6 months - Males']])
```

[5]:
```
l2=tuple([df['Industrial Category - N to O - Females'],df['Industrial Category␣
↪- P to Q - Persons']])
```

[6]:
```
df.describe()
```

[6]:

	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	Worked for less than 3 months - Males
count	5.940000e+02	594.000000	594.000000	594.000000	594.000000
mean	1.617277e+04	7932.700337	8240.067340	2981.629630	1338.289562
std	7.607172e+04	36864.822704	39259.545337	13909.621137	6127.047670
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000
25%	2.872500e+02	147.250000	144.000000	27.000000	
50%	2.225500e+03	1147.000000	1076.000000	430.000000	
75%	9.628500e+03	4770.500000	4887.500000	1775.250000	
max	1.200828e+06	589003.000000	611825.000000	221386.000000	

	Worked for less than 3 months - Females \
count	594.000000
mean	1643.340067
std	7808.832522
min	0.000000
25%	13.000000
50%	213.000000
75%	946.500000
max	122018.000000

	Industrial Category - A - Cultivators - Persons \
count	594.000000
mean	865.117845
std	4274.458077
min	0.000000
25%	9.000000
50%	69.500000
75%	466.000000
max	64235.000000

	Industrial Category - A - Cultivators - Males \
count	594.000000
mean	466.424242
std	2298.072295
min	0.000000
25%	5.000000
50%	35.500000
75%	244.250000
max	34632.000000

	Industrial Category - A - Cultivators - Females \
count	594.000000
mean	398.693603
std	1978.682322
min	0.000000
25%	4.000000
50%	32.000000
75%	204.750000
max	29603.000000

	Industrial Category - A - Agricultural labourers - Persons ... \
count	594.000000 ...

mean	12225.616162	…
std	60458.382586	…
min	0.000000	…
25%	79.250000	…
50%	1094.000000	…
75%	6279.750000	…
max	907752.000000	…

	Industrial Category - N to O - Females \
count	594.000000
mean	48.013468
std	222.553500
min	0.000000
25%	0.000000
50%	2.000000
75%	18.000000
max	3565.000000

	Industrial Category - P to Q - Persons \
count	594.000000
mean	149.225589
std	696.553730
min	0.000000
25%	0.000000
50%	14.500000
75%	99.750000
max	11080.000000

	Industrial Category - P to Q - Males \
count	594.000000
mean	54.127946
std	253.067862
min	0.000000
25%	0.000000
50%	6.000000
75%	35.750000
max	4019.000000

	Industrial Category - P to Q - Females \
count	594.000000
mean	95.097643
std	444.011425
min	0.000000
25%	0.000000
50%	6.500000
75%	64.000000
max	7061.000000

	Industrial Category - R to U - HHI - Persons	\
count	594.000000	
mean	226.707071	
std	1039.953069	
min	0.000000	
25%	0.000000	
50%	27.000000	
75%	126.750000	
max	16833.000000	

	Industrial Category - R to U - HHI - Males	\
count	594.000000	
mean	57.454545	
std	265.230865	
min	0.000000	
25%	0.000000	
50%	7.500000	
75%	32.000000	
max	4266.000000	

	Industrial Category - R to U - HHI - Females	\
count	594.000000	
mean	169.252525	
std	776.206806	
min	0.000000	
25%	0.000000	
50%	20.000000	
75%	97.500000	
max	12567.000000	

	Industrial Category - R to U - Non HHI - Persons	\
count	594.000000	
mean	1644.282828	
std	7325.241597	
min	0.000000	
25%	64.500000	
50%	263.500000	
75%	994.000000	
max	122088.000000	

	Industrial Category - R to U - Non HHI - Males	\
count	594.000000	
mean	751.528620	
std	3352.811737	
min	0.000000	
25%	34.000000	

	Industrial Category - R to U - Non HHI - Females
count	594.000000
mean	892.754209
std	3988.125301
min	0.000000
25%	30.500000
50%	135.000000
75%	500.000000
max	66287.000000

[8 rows x 63 columns]

[7]:
```
1
```

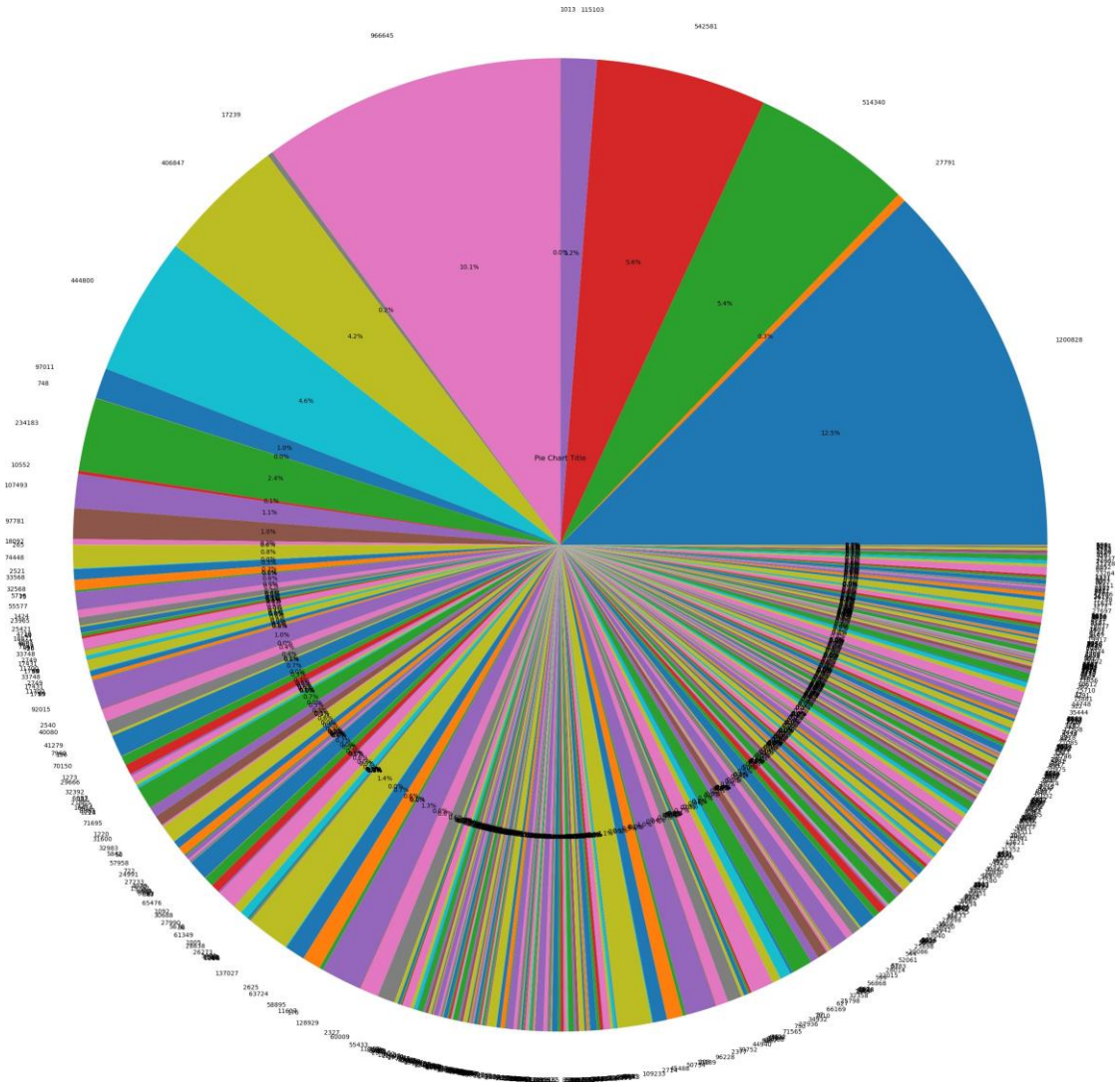
50%	123.000000
75%	447.750000
max	55801.000000

[7]: (0 611825
1 13666
2 254780
3 290624
4 52270
...
589 143
590 1631
591 1903
592 297
593 1
Name: Worked for 3 months or more but less than 6 months - Females, Length: 594, dtype: int64,
0 589003
1 14125
2 259560
3 251957
4 62833
...
589 129
590 1654
591 1769
592 399
593 1
Name: Worked for 3 months or more but less than 6 months - Males, Length: 594, dtype: int64)

[8]:
```
2
```

[8]:
```
(0       3565
 1         11
 2       1754
 3       1619
 4        175
        ...
 589        0
 590       20
 591       33
 592        0
 593        0
 Name: Industrial Category - N to O - Females, Length: 594, dtype: int64,
 0       11080
 1         122
 2        7536
 3        3205
 4         211
        ...
 589        0
 590       44
 591       35
 592        3
 593        0
 Name: Industrial Category - P to Q - Persons, Length: 594, dtype: int64)
```

[9]:
```
# assigning the csv data to variable of piechart
pie_chart_data = df['Worked for 3 months or more but less than 6 months - Persons']
#asigning values to the pie chart
plt.pie(pie_chart_data, labels=df['Worked for 3 months or more but less than 6 months -  Persons'], autopct='%1.1f%%',radius=7.5)

plt.title('Pie Chart Title')
#printing the pie chart
plt.show()
```



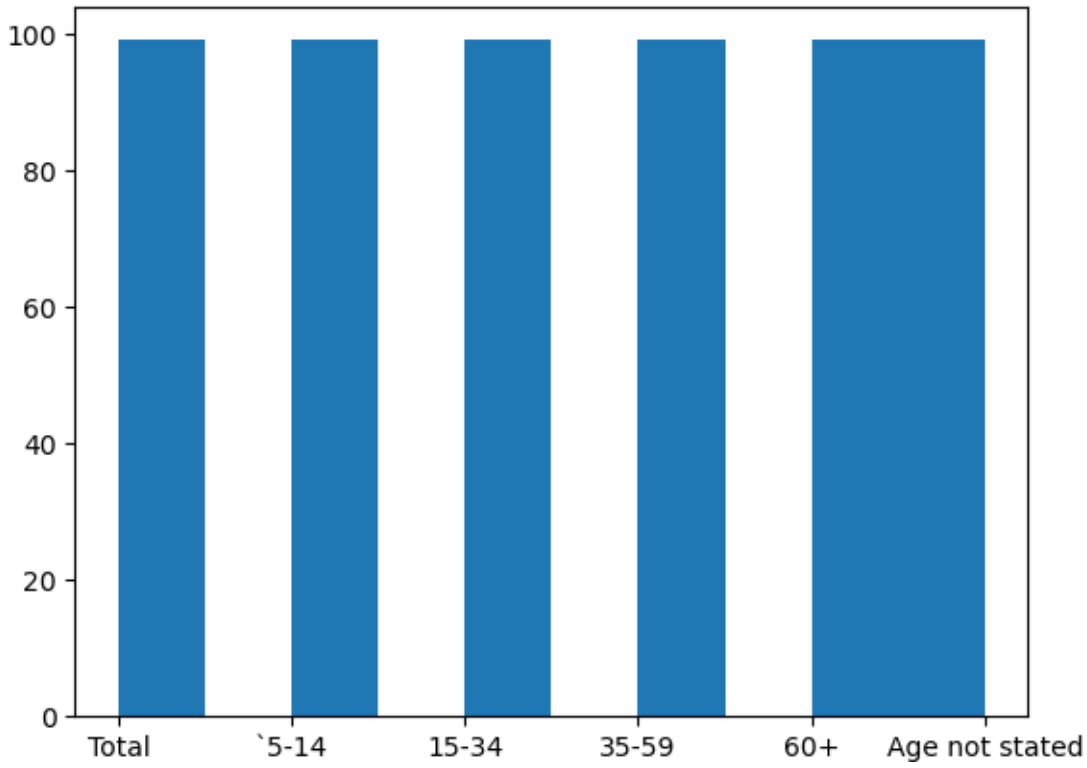
4 printing the pie chart using the given csv file data sets

[10]:
```
price = df["Age group"]
```

5 visualizing the data sets column in the form of of histogram

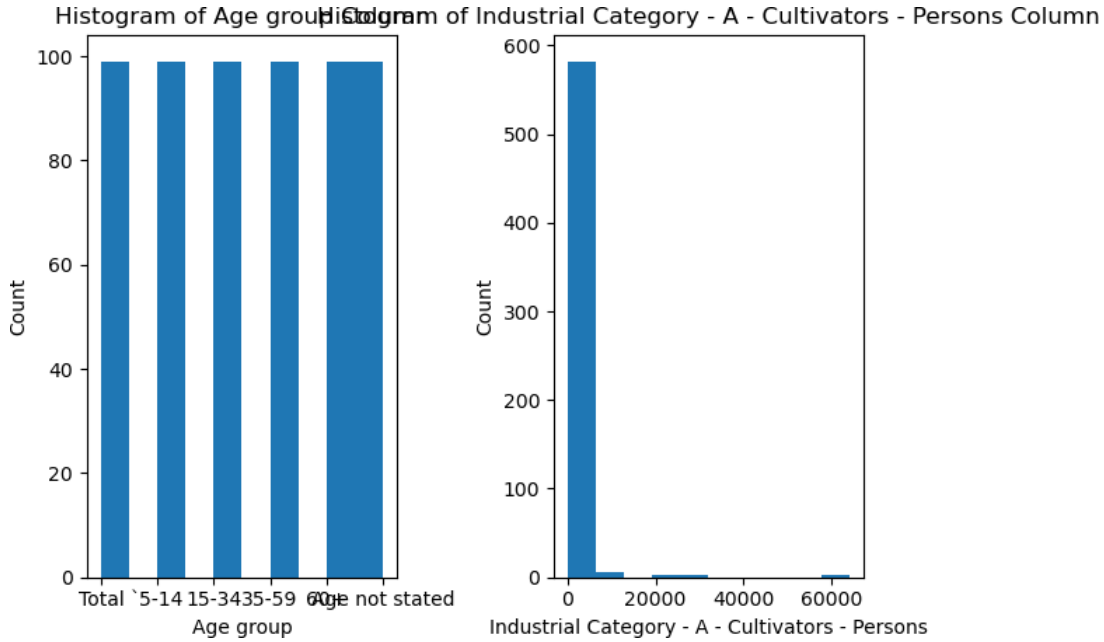
[11]:
```
plt.hist(price)
```

[11]:
```
(array([99.,  0., 99.,  0., 99.,  0., 99.,  0., 99., 99.]),
 array([0. , 0.5, 1. , 1.5, 2. , 2.5, 3. , 3.5, 4. , 4.5, 5. ]),
 <BarContainer object of 10 artists>)
```



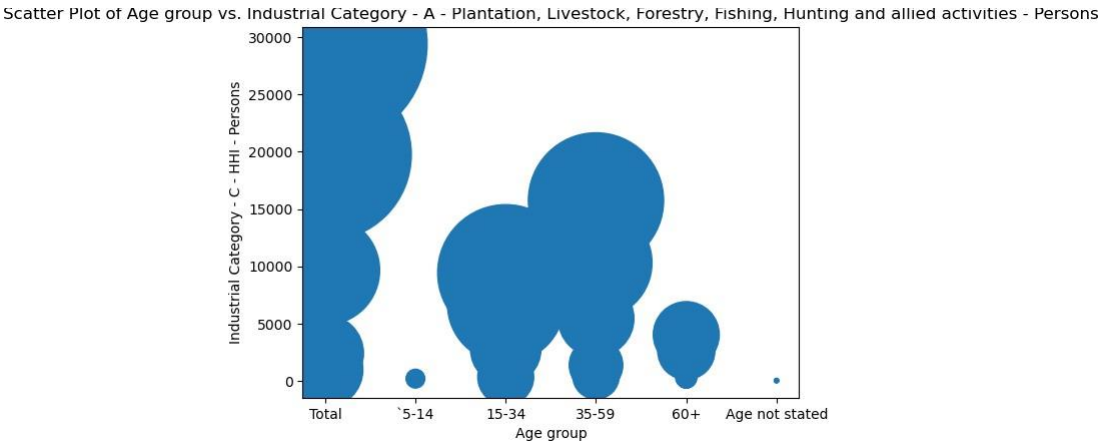
[16]:
```
column_1 = df["Age group"]
column_2 = df["Industrial Category - A - Cultivators - Persons"]

# Create the histogram
fig, axs = plt.subplots(1, 2)

axs[0].hist(column_1)
axs[1].hist(column_2)

# Add a title and axis labels for each subplot
axs[0].set_title("Histogram of {} Column".format(column_1.name))
axs[1].set_title("Histogram of {} Column".format(column_2.name))

axs[0].set_xlabel(column_1.name)
axs[1].set_xlabel(column_2.name)

axs[0].set_ylabel("Count")
axs[1].set_ylabel("Count")

# Adjust the subplot layout
plt.tight_layout()
```

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



6 visualizing three different columns by using scatter plot

[26]:
```
x_column = df["Age group"]
y_column = df["Industrial Category - A - Plantation, Livestock, Forestry,␣
↪Fishing, Hunting and allied activities - Persons"]
z_column = df["Industrial Category - C - HHI - Persons"]
# Create the scatter plot
plt.scatter(x_column, y_column, z_column)

# Add a title and axis labels
plt.title("Scatter Plot of {} vs. {}".format(x_column.name, y_column.name,␣
↪z_column.name))
plt.xlabel(x_column.name)
plt.ylabel(y_column.name)
plt.ylabel(z_column.name)

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



7 output for the scatter plot for the gib=ven three columns