Imported Panda Package

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

matplotlib is a library that is used for plotting data

seaborn is also a library that is used for visualization

A dataset containing various parameters is uploaded as 'df'

```
df=pd.read_excel('/content/smoking.xlsx')
```

```
df.shape
```

```
(4000, 11)
```

For finding the percentage, a function is created

```
def percentage(a,b):
  return (a/b)*100
```

The contents of the dataset are being showed

```
df.head(18)
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75.00	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
1	Female	49.00	No	No	Yes	Private	Village	108.8	26.7	smokes	No
2	Male	32.00	No	No	Yes	Private	City	64.1	23.4	smokes	No
3	Male	78.00	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
4	Male	39.00	No	No	Yes	Private	City	55.4	41.6	formerly smoked	No
5	Female	-3.00	No	No	Yes	Private	Village	83.2	29.3	smokes	No
6	Male	63.00	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
7	Male	4.00	No	No	No	children	City	94.1	16.1	Unknown	No
8	Female	45.00	No	No	Yes	Private	City	64.6	40.5	never smoked	No
9	Female	52.00	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
10	Female	45.00	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
11	Male	31.00	No	No	Yes	Private	Village	101.4	29.5	Unknown	No
12	Female	45.00	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
13	Female	57.00	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56.00	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
15	Female	20.00	No	No	No	Govt_job	Village	80.4	NaN	never smoked	No
16	Male	0.24	No	No	No	children	City	89.6	12.1	Unknown	No
17	Female	38.00	No	No	Yes	Govt_job	Village	77.1	25.5	never smoked	No

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.shape
```

```
(4000, 11)
```

In the following dataset, there exists some data columns which contains null values. correcting those errors. saving the dataset as 'part_correct_df'

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

```
Gender              0
Age                 0
HasTension          0
AnyHeartDisease     0
NeverMarried        0
Occupation          0
```

```
LivesIn             0
GlucoseLevel        0
BMI               149
SmokingStatus       0
HeartAttack         0
dtype: int64
```

```
part_correct_df=df.dropna()
part_correct_df.shape
```

```
(3851, 11)
```

Now, coverting this dataset's 'Age' values into integer values

```
part_correct_df['Age']=part_correct_df['Age'].astype(int)
part_correct_df.head(18)
```

```
<ipython-input-11-0aabd39dafbe>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df['Age']=part_correct_df['Age'].astype(int)
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
1	Female	49	No	No	Yes	Private	Village	108.8	26.7	smokes	No
2	Male	32	No	No	Yes	Private	City	64.1	23.4	smokes	No
3	Male	78	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
4	Male	39	No	No	Yes	Private	City	55.4	41.6	formerly smoked	No
5	Female	-3	No	No	Yes	Private	Village	83.2	29.3	smokes	No
6	Male	63	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
7	Male	4	No	No	No	children	City	94.1	16.1	Unknown	No
8	Female	45	No	No	Yes	Private	City	64.6	40.5	never smoked	No
9	Female	52	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
10	Female	45	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
11	Male	31	No	No	Yes	Private	Village	101.4	29.5	Unknown	No
12	Female	45	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
13	Female	57	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
16	Male	0	No	No	No	children	City	89.6	12.1	Unknown	No
17	Female	38	No	No	Yes	Govt_job	Village	77.1	25.5	never smoked	No
18	Male	45	No	No	Yes	Private	Village	56.6	22.3	smokes	No

Next steps: Generate code with part_correct_df | View recommended plots | New interactive sheet

```
part_correct_df.shape
```

```
(3851, 11)
```

Now, coverting Age section's negative values into positive values

```
part_correct_df['Age']=part_correct_df['Age'].abs()
part_correct_df.head(18)
```

```
<ipython-input-13-434efcf9acde>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df['Age']=part_correct_df['Age'].abs()
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
1	Female	49	No	No	Yes	Private	Village	108.8	26.7	smokes	No
2	Male	32	No	No	Yes	Private	City	64.1	23.4	smokes	No
3	Male	78	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
4	Male	39	No	No	Yes	Private	City	55.4	41.6	formerly smoked	No
5	Female	3	No	No	Yes	Private	Village	83.2	29.3	smokes	No
6	Male	63	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
7	Male	4	No	No	No	children	City	94.1	16.1	Unknown	No
8	Female	45	No	No	Yes	Private	City	64.6	40.5	never smoked	No
9	Female	52	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
10	Female	45	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
11	Male	31	No	No	Yes	Private	Village	101.4	29.5	Unknown	No
12	Female	45	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
13	Female	57	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
16	Male	0	No	No	No	children	City	89.6	12.1	Unknown	No
17	Female	38	No	No	Yes	Govt_job	Village	77.1	25.5	never smoked	No
18	Male	45	No	No	Yes	Private	Village	56.6	22.3	smokes	No

Next steps: Generate code with part_correct_df | View recommended plots | New interactive sheet

Again, in this dataset's 'GlucoseLevel' section some values are negative. correctiong those errors

```
part_correct_df['GlucoseLevel']=part_correct_df['GlucoseLevel'].abs()
```

```
<ipython-input-14-1fb487b0e5ae>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df['GlucoseLevel']=part_correct_df['GlucoseLevel'].abs()
```

Now, it is known that the lowest Glucose level of a human being is 20. So showing the dataset in which Glucose level are less than 20. Naming the dataset as 'Glucose_lessthan20'

```
Glucose_lessthan20=part_correct_df[part_correct_df['GlucoseLevel']<20]
Glucose_lessthan20
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
98	Male	1	No	No	No	children	City	2.0	20.4	Unknown	No
697	Male	14	No	No	No	Private	Village	2.0	26.2	Unknown	No
829	Female	66	No	No	Yes	Self-employed	Village	2.0	31.9	never smoked	No
1460	Female	55	No	No	Yes	Govt_job	City	2.0	28.2	formerly smoked	No
2428	Male	4	No	No	No	children	Village	2.0	18.8	Unknown	No
2543	Female	15	No	No	No	Private	Village	2.0	24.2	Unknown	No
3095	Female	73	No	No	Yes	Private	Village	2.0	31.2	Unknown	No
3370	Male	3	No	No	No	children	City	2.0	16.5	Unknown	No
3986	Male	11	No	No	No	children	Village	2.0	23.9	Unknown	No

Next steps: Generate code with Glucose_lessthan20 | View recommended plots | New interactive sheet

```
Glucose_lessthan20.shape
```

```
(9, 11)
```

```
part_correct_df.drop(Glucose_lessthan20.index, inplace=True)
part_correct_df.shape
```

```
<ipython-input-17-2ca10274b64a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df.drop(Glucose_lessthan20.index, inplace=True)
(3842, 11)
```

BMI cannont be less than 12 so finding those records whose bmi is less than 12 and removing them from the dataframe

```
bmi_lessthan12=part_correct_df[part_correct_df['BMI']<12]
bmi_lessthan12
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
2001	Female	79	No	No	Yes	Self-employed	City	78.5	9.6	Unknown	No
2505	Female	1	No	No	No	children	Village	121.9	10.9	Unknown	No

Next steps: Generate code with `bmi_lessthan12` View recommended plots New interactive sheet

```
part_correct_df.drop(bmi_lessthan12.index, inplace=True)
part_correct_df.shape
```

```
<ipython-input-19-66da93acba2a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df.drop(bmi_lessthan12.index, inplace=True)
(3840, 11)
```

Showing the dataset

```
part_correct_df.head(18)
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
1	Female	49	No	No	Yes	Private	Village	108.8	26.7	smokes	No
2	Male	32	No	No	Yes	Private	City	64.1	23.4	smokes	No
3	Male	78	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
4	Male	39	No	No	Yes	Private	City	55.4	41.6	formerly smoked	No
5	Female	3	No	No	Yes	Private	Village	83.2	29.3	smokes	No
6	Male	63	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
7	Male	4	No	No	No	children	City	94.1	16.1	Unknown	No
8	Female	45	No	No	Yes	Private	City	64.6	40.5	never smoked	No
9	Female	52	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
10	Female	45	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
11	Male	31	No	No	Yes	Private	Village	101.4	29.5	Unknown	No
12	Female	45	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
13	Female	57	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
16	Male	0	No	No	No	children	City	89.6	12.1	Unknown	No
17	Female	38	No	No	Yes	Govt_job	Village	77.1	25.5	never smoked	No
18	Male	45	No	No	Yes	Private	Village	56.6	22.3	smokes	No

Next steps: Generate code with part_correct_df | View recommended plots | New interactive sheet

Now, in this dataset's 'Age' section some values are in points. correcting those errors by converting the values into 'int' and showing the dataset

Now, droping those dataset where age=0

```
part_correct_df.drop(part_correct_df[part_correct_df['Age']==0].index,inplace=True)
part_correct_df.shape
```

```
<ipython-input-21-bade1089e938>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  part_correct_df.drop(part_correct_df[part_correct_df['Age']==0].index,inplace=True)
(3809, 11)
```

Now, in the dataset some people's gender are marked as 'Unknown' so dropping the values and naming the the updated dataset as 'correct_df_final'

```
Unknown_gender_df=part_correct_df[part_correct_df['Gender']=='Unknown']
Unknown_gender_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
170	Unknown	5	No	No	No	children	Village	67.5	18.2	Unknown	No
476	Unknown	77	Yes	No	Yes	Self-employed	Village	175.9	32.2	never smoked	No
718	Unknown	79	Yes	No	Yes	Self-employed	Village	173.1	26.5	never smoked	Yes
1346	Unknown	79	No	No	Yes	Private	Village	90.4	22.2	never smoked	No
2673	Unknown	53	No	No	Yes	Govt_job	City	83.6	43.7	Unknown	No
3573	Unknown	10	No	No	No	children	City	94.1	24.7	Unknown	No

Next steps: Generate code with Unknown_gender_df | View recommended plots | New interactive sheet

```
correct_df_final=part_correct_df.drop([170,476,718,1346,2673,3573])
correct_df_final.shape
```

```
(3803, 11)
```

It is found that the living area of some people in this dataset are marked as 'Unknown' dropping those values

```
unknown_lives_in=correct_df_final[correct_df_final['LivesIn']=='Unknown']
unknown_lives_in
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
1594	Female	37	No	No	Yes	Self-employed	Unknown	94.9	34.1	never smoked	No
1680	Female	15	No	No	No	Private	Unknown	68.6	26.7	never smoked	No
2506	Male	29	No	No	No	Private	Unknown	58.7	34.4	smokes	No
2559	Male	18	No	No	No	Private	Unknown	104.8	23.7	never smoked	No
3967	Male	34	No	No	Yes	Private	Unknown	109.4	23.9	Unknown	No

Next steps: Generate code with unknown_lives_in | View recommended plots | New interactive sheet

```
correct_df_final=correct_df_final.drop([1594,1680,2506,2559,3967])
correct_df_final.shape
```

```
(3798, 11)
```

Therefore, after all the filtration of data, the available dataset is stored as "correct_df_final" and all the analysis will done with this dataset

Finding correlation between Age and GlucoseLevel column

correlation can work with numeric data

```
correct_df_final['Age'].corr(correct_df_final['GlucoseLevel'])
```

```
0.2336740379173944
```

If correlation value will be 1 or -1 or near 1 or -1, then it is said that the parameters has a strong correlation when the value is near 0 it is said that the parameters has no correlation 0.2 is defined as weak correlation

Therefore, These two parameters has no correlations between them

now **checking correlation between age and bmi**

```
correct_df_final['Age'].corr(correct_df_final['BMI'])
```

```
0.31203945611867756
```

Therefore these two parameters has no correlation

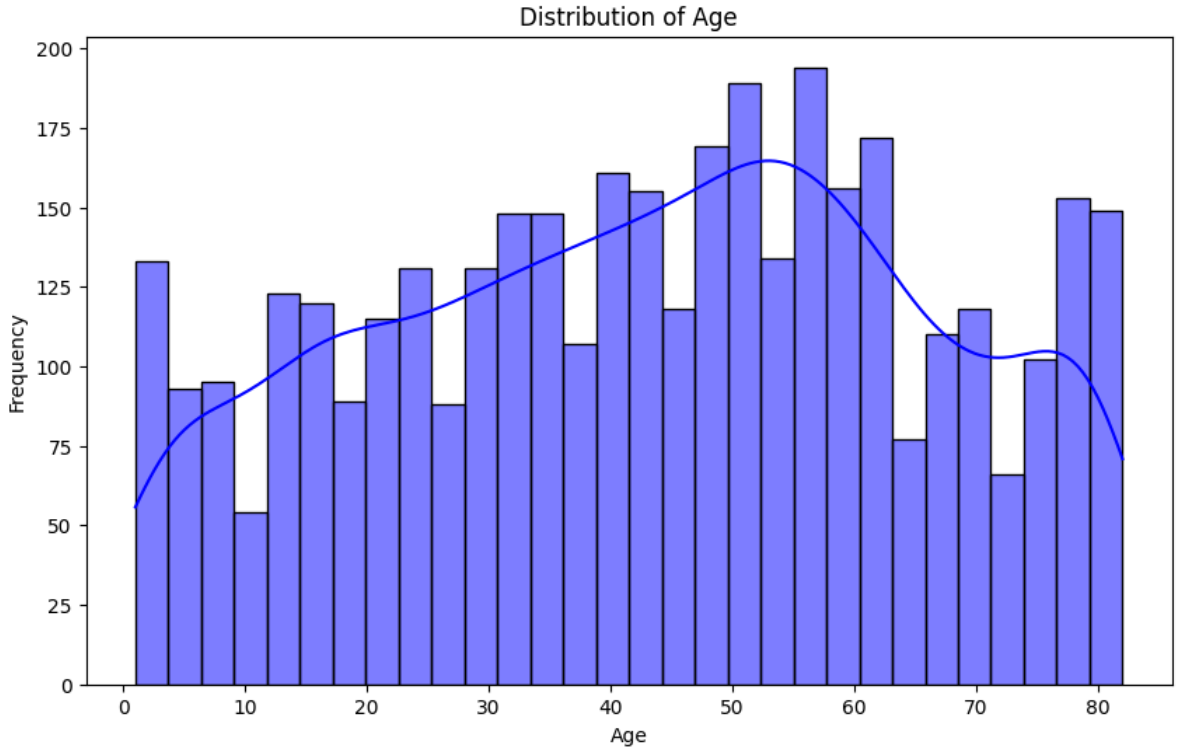
Now, **finding correlation between glucoselevel and bmi**

```
correct_df_final['GlucoseLevel'].corr(correct_df_final['BMI'])
```

```
0.17258774646944808
```

Therefore, there is no correlation between them

```
plt.figure(figsize=(10, 6))
sns.histplot(correct_df_final['Age'], kde=True, bins=30, color='blue')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

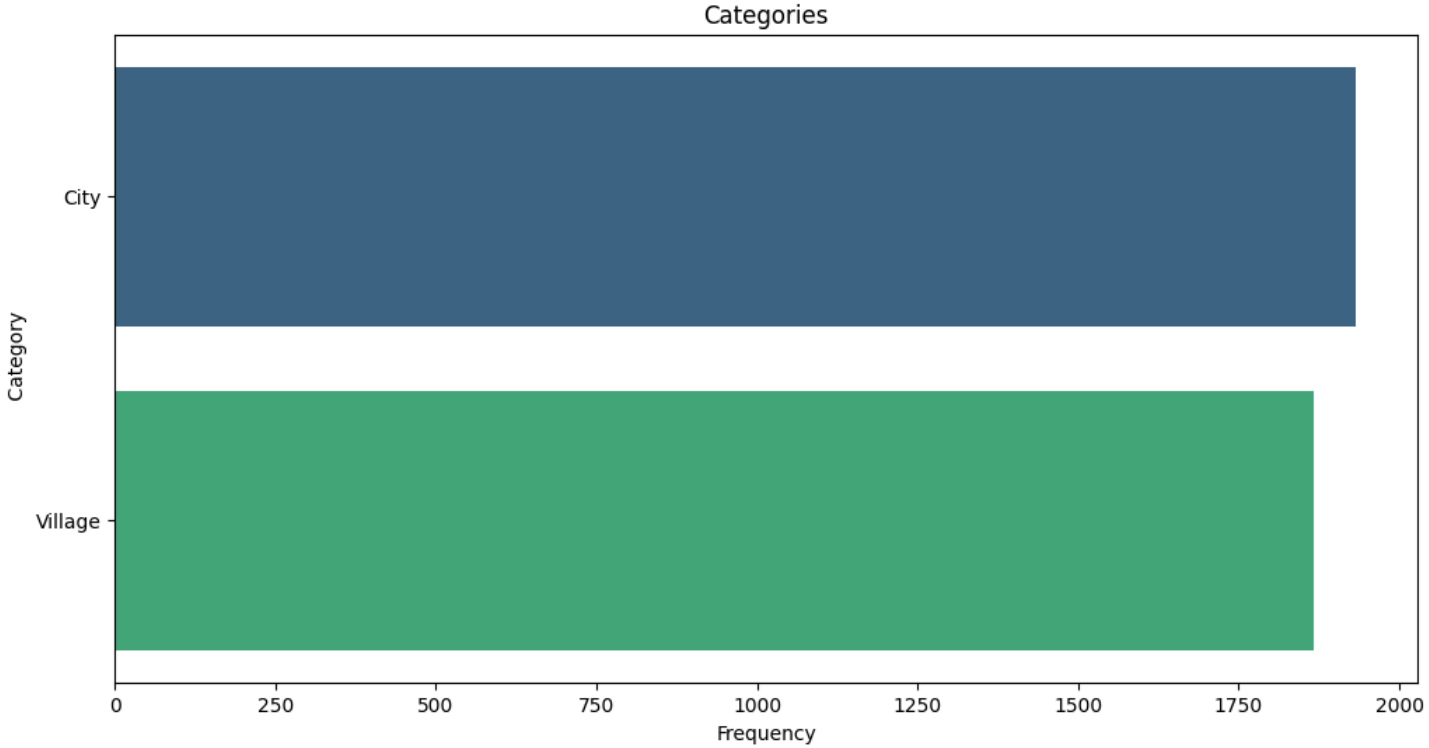



```
plt.figure(figsize=(12, 6))
sns.countplot(y=correct_df_final['LivesIn'], order=correct_df_final['LivesIn'].value_counts().index, palette='viridis')
plt.title('Categories')
plt.xlabel('Frequency')
plt.ylabel(' Category')
plt.show()
```

```
<ipython-input-30-c5392ee98c9f>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(y=correct_df_final['LivesIn'], order=correct_df_final['LivesIn'].value_counts().index, palette='viridis')
```

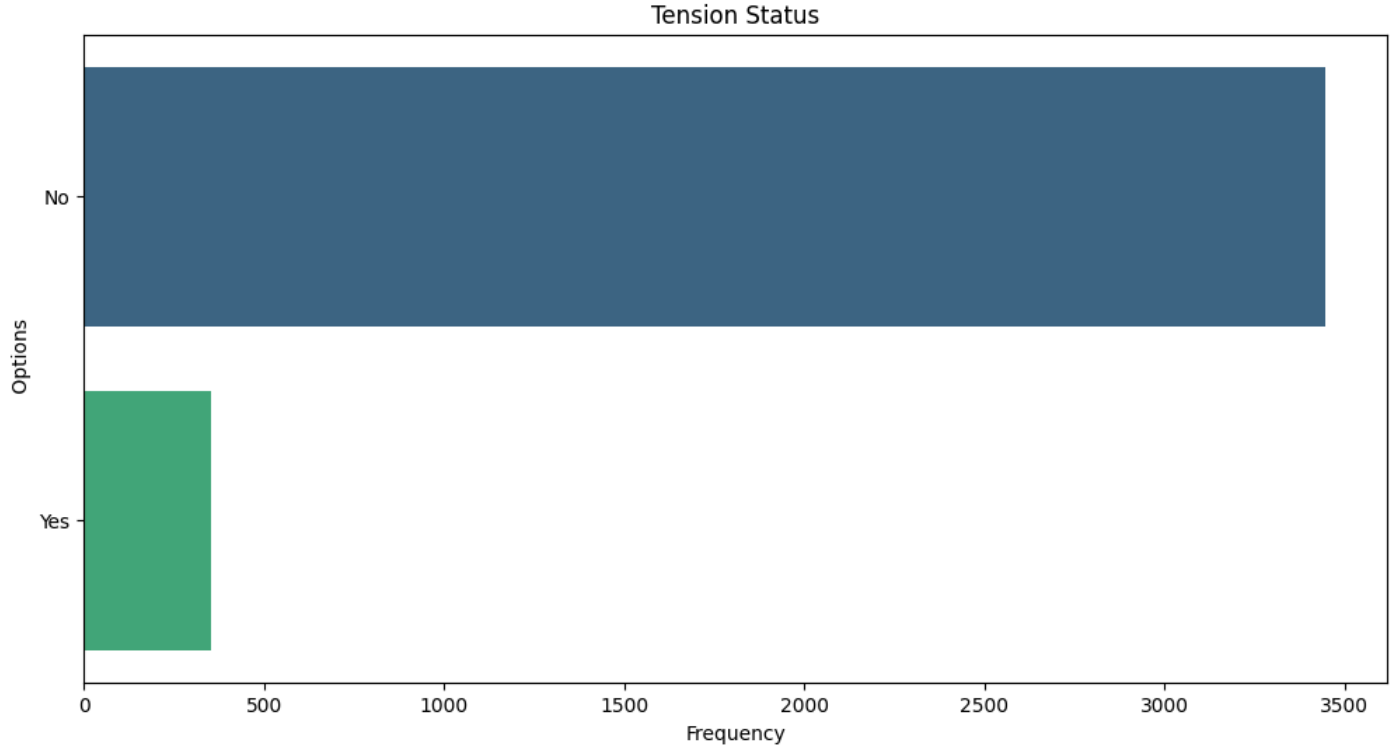



```
plt.figure(figsize=(12, 6))
sns.countplot(y=correct_df_final['HasTension'], order=correct_df_final['HasTension'].value_counts().index, palette='viridis')
plt.title('Tension Status')
plt.xlabel('Frequency')
plt.ylabel(' Options')
plt.show()
```

```
<ipython-input-31-497a52401e58>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(y=correct_df_final['HasTension'], order=correct_df_final['HasTension'].value_counts().index, palette='viridis')
```



Now, Clustering data for 1 to 10 age group. naming the dataframe as 'onetotendf'

```
onetotendf = correct_df_final[(correct_df_final['Age'] >= 1) & (correct_df_final['Age'] <= 10)]
onetotendf
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
5	Female	3	No	No	Yes	Private	Village	83.2	29.3	smokes	No
7	Male	4	No	No	No	children	City	94.1	16.1	Unknown	No
25	Male	7	No	No	No	children	City	78.1	17.0	Unknown	No
33	Female	7	No	No	No	children	City	62.6	23.0	Unknown	No
35	Male	2	No	No	No	children	Village	85.9	19.2	Unknown	No
...	...	...	...	...	...	...	...	...	...	...	...
3951	Male	5	No	No	No	children	Village	98.6	18.1	Unknown	No
3952	Female	3	No	No	Yes	Govt_job	City	57.2	24.5	smokes	No
3964	Male	4	No	No	No	children	City	98.8	17.3	Unknown	No
3972	Male	2	No	No	No	children	Village	66.0	14.7	Unknown	No
3973	Male	1	No	No	No	children	Village	86.6	20.6	Unknown	No

349 rows × 11 columns

Next steps: Generate code with onetotendf | View recommended plots | New interactive sheet

Now, in this dataframe, it can seen that many data contains meaningless information. For Example: in the dataframe many children has smoking habbit.It can be seen that in the 'occupation'section, those children's occupation are not marked as 'children' only those data's are incorrect. So clustering the dataset and naming the dataset as 'ZEROtoTENdfcorrected'

```
ONEtoTENdfcorrected=onetotendf[onetotendf['Occupation']=='children']
ONEtoTENdfcorrected
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
7	Male	4	No	No	No	children	City	94.1	16.1	Unknown	No
25	Male	7	No	No	No	children	City	78.1	17.0	Unknown	No
33	Female	7	No	No	No	children	City	62.6	23.0	Unknown	No
35	Male	2	No	No	No	children	Village	85.9	19.2	Unknown	No
40	Male	6	No	No	No	children	City	87.3	18.2	Unknown	No
...	...	...	...	...	...	...	...	...	...	...	...
3940	Male	5	No	No	No	children	Village	105.7	16.7	Unknown	No
3951	Male	5	No	No	No	children	Village	98.6	18.1	Unknown	No
3964	Male	4	No	No	No	children	City	98.8	17.3	Unknown	No
3972	Male	2	No	No	No	children	Village	66.0	14.7	Unknown	No
3973	Male	1	No	No	No	children	Village	86.6	20.6	Unknown	No

339 rows × 11 columns

Next steps: Generate code with ONEtoTENdfcorrected | View recommended plots | New interactive sheet

```
ONEtoTENdfcorrected.shape
```

```
(339, 11)
```

Now, it is obvious that no children in this age frame has the abilty to smoke. Therefore replacing the 'Unknown' status of 'SmokingStatus' section as 'Never smoked'

```
ONEtoTENdfcorrected['SmokingStatus'] = ONEtoTENdfcorrected['SmokingStatus'].replace('Unknown','Never smoked')
ONEtoTENdfcorrected
```

```
<ipython-input-35-527c70656bfa>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  ONEtoTENdfcorrected['SmokingStatus'] = ONEtoTENdfcorrected['SmokingStatus'].replace('Unknown','Never smoked')
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
7	Male	4	No	No	No	children	City	94.1	16.1	Never smoked	No
25	Male	7	No	No	No	children	City	78.1	17.0	Never smoked	No
33	Female	7	No	No	No	children	City	62.6	23.0	Never smoked	No
35	Male	2	No	No	No	children	Village	85.9	19.2	Never smoked	No
40	Male	6	No	No	No	children	City	87.3	18.2	Never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3940	Male	5	No	No	No	children	Village	105.7	16.7	Never smoked	No
3951	Male	5	No	No	No	children	Village	98.6	18.1	Never smoked	No
3964	Male	4	No	No	No	children	City	98.8	17.3	Never smoked	No
3972	Male	2	No	No	No	children	Village	66.0	14.7	Never smoked	No
3973	Male	1	No	No	No	children	Village	86.6	20.6	Never smoked	No

339 rows × 11 columns

Next steps: Generate code with ONEtoTENdfcorrected | View recommended plots | New interactive sheet

Now, finding what perecentage of this group are male and female

```
one_to_ten_male=ONEtoTENdfcorrected[ONEtoTENdfcorrected['Gender']=='Male']
a=percentage(one_to_ten_male.shape[0],ONEtoTENdfcorrected.shape[0])
print(a)
```

```
51.03244837758112
```

```
one_to_ten_female=ONEtoTENdfcorrected[ONEtoTENdfcorrected['Gender']=='Female']
a=percentage(one_to_ten_female.shape[0],ONEtoTENdfcorrected.shape[0])
print(a)
```

```
48.96755162241888
```

Therefore, there are about 51% male and 49% female in this group.

Now, we have to find out whether any children in this age group had heart-attack. Saving the database as 'ZEROtoTEN_heart_attackdf'

```
ONEtoTEN_heart_attackdf=ONEtoTENdfcorrected[ONEtoTENdfcorrected['HeartAttack']=='Yes']
ONEtoTEN_heart_attackdf
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack

Therefore, it can be said that all the children of this age group has a healthy heart condition

Now, we will see if any children in this age group has tension

```
ONEtoTEN_Tension=ONEtoTENdfcorrected[ONEtoTENdfcorrected['HasTension']=="Yes"]
ONEtoTEN_Tension
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack

Now for filtering the dataset more accurately, we will find if any children got married.

```
ONEtoTEN_Marriage=ONEtoTENdfcorrected[ONEtoTENdfcorrected['NeverMarried']=='Yes']
ONEtoTEN_Marriage
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack

All the data given in this dataset are accurate as no children got married.

Now, finding what percentage of these group has obesity and genderwise

```
one_to_ten_obese=ONEtoTENdfcorrected[ONEtoTENdfcorrected['BMI']>=30]
a=percentage(one_to_ten_obese.shape[0],ONEtoTENdfcorrected.shape[0])
print(a)
```

```
2.359882005899705
```

```
one_to_ten_obese_male=one_to_ten_obese[one_to_ten_obese['Gender']=='Male']
a=percentage(one_to_ten_obese_male.shape[0],one_to_ten_obese.shape[0])
print(a)
```

```
75.0
```

```
one_to_ten_obese_female=one_to_ten_obese[one_to_ten_obese['Gender']=='Female']
a=percentage(one_to_ten_obese_female.shape[0],one_to_ten_obese.shape[0])
print(a)
```

```
25.0
```

Therefore, about 2% of these group has obesity and among them, 75% are male and 25% are female

Now clustering data with parameter 11 to 20 years old age limit

```
ELEVEN_TO_TWENTY_df=correct_df_final[(correct_df_final['Age']>=11) & (correct_df_final['Age']<=20 )]
ELEVEN_TO_TWENTY_df.shape
```

```
(403, 11)
```

```
ELEVEN_TO_TWENTY_corrected=ELEVEN_TO_TWENTY_df
```

What percentage of male and female exists in this group

```
ELEVEN_TO_TWENTY_male=ELEVEN_TO_TWENTY_corrected[ELEVEN_TO_TWENTY_corrected['Gender']=='Male']
a=percentage(ELEVEN_TO_TWENTY_male.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
45.6575682382134
```

```
ELEVEN_TO_TWENTY_female=ELEVEN_TO_TWENTY_corrected[ELEVEN_TO_TWENTY_corrected['Gender']=='Female']
a=percentage(ELEVEN_TO_TWENTY_female.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
54.3424317617866
```

Therefore, about 46% are male and 54% are female

Now, gathering those data, where glucose level is less than 70.0. A blood glucose level of less than 70 mg/dL indicates hypoglycemia.

```
ELEVEN_TO_TWENTY_hypoglycemia=ELEVEN_TO_TWENTY_corrected[ELEVEN_TO_TWENTY_corrected['GlucoseLevel']<70.0]
a=percentage(ELEVEN_TO_TWENTY_hypoglycemia.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
15.88089330024814
```

now finding genderwise

```
ELEVEN_TO_TWENTY_hypoglycemia_male=ELEVEN_TO_TWENTY_hypoglycemia[ELEVEN_TO_TWENTY_hypoglycemia['Gender']=='Male']
a=percentage(ELEVEN_TO_TWENTY_hypoglycemia_male.shape[0],ELEVEN_TO_TWENTY_hypoglycemia.shape[0])
print(a)
```

```
51.5625
```

```
ELEVEN_TO_TWENTY_hypoglycemia_female=ELEVEN_TO_TWENTY_hypoglycemia[ELEVEN_TO_TWENTY_hypoglycemia['Gender']=='Female']
a=percentage(ELEVEN_TO_TWENTY_hypoglycemia_female.shape[0],ELEVEN_TO_TWENTY_hypoglycemia.shape[0])
print(a)
```

```
48.4375
```

Therefore, it is evident that 16% people in the specified age range has hypoglycemia

and among them 52% are male and 48% are female

Now finding what percentage of them has a healthy glucose level

```
ELEVEN_TO_TWENTY_healthy_glucose= ELEVEN_TO_TWENTY_corrected[(ELEVEN_TO_TWENTY_corrected['GlucoseLevel'] >70) & (ELEVEN_TO_TWENTY_corrected['GlucoseLevel']<100) ]
a = percentage(ELEVEN_TO_TWENTY_healthy_glucose.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
51.1166253101737
```

Now finding genderwise

```
ELEVEN_TO_TWENTY_healthy_glucose_male=ELEVEN_TO_TWENTY_healthy_glucose[ELEVEN_TO_TWENTY_healthy_glucose['Gender']=='Male']
a=percentage(ELEVEN_TO_TWENTY_healthy_glucose_male.shape[0],ELEVEN_TO_TWENTY_healthy_glucose.shape[0])
print(a)
```

```
40.29126213592233
```

```
ELEVEN_TO_TWENTY_healthy_glucose_female=ELEVEN_TO_TWENTY_healthy_glucose[ELEVEN_TO_TWENTY_healthy_glucose['Gender']=='Female']
a=percentage(ELEVEN_TO_TWENTY_healthy_glucose_female.shape[0],ELEVEN_TO_TWENTY_healthy_glucose.shape[0])
print(a)
```

```
59.70873786407766
```

about 51% people from the specified age range has a heatlhly glucose level and among them 40% are male and 60% are female

at the age of 18, children become adult and they will be at legal age to smoke. So, collecting those data where children below 18 years of age are smokers

```
teen_eleven_to_twenty=ELEVEN_TO_TWENTY_corrected[(ELEVEN_TO_TWENTY_corrected['Age']<18)]
a=percentage(teen_eleven_to_twenty.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
66.74937965260546
```

Therefore about 67% of this age group are teens

```
smokers_below18 = ELEVEN_TO_TWENTY_corrected[(ELEVEN_TO_TWENTY_corrected['Age'] < 18) & (ELEVEN_TO_TWENTY_corrected['SmokingStatus'] == 'smokes')]
a=percentage(smokers_below18.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
1.7369727047146404
```

Therefore, total 2% of teens are teenage smokers

Now, finding out the percentage of male teen smokers and female teen smokers

```
smokers_below18_male=smokers_below18[smokers_below18['Gender']=='Male']
a=percentage(smokers_below18_male.shape[0],smokers_below18.shape[0])
print(a)
```

```
28.57142857142857
```

```
smokers_below18_female=smokers_below18[smokers_below18['Gender']=='Female']
a=percentage(smokers_below18_female.shape[0],smokers_below18.shape[0])
print(a)
```

```
71.42857142857143
```

Now, according to above analysis, it can be seen that 29% are male and 71% are female

Now finding the same for adults

```
ELEVEN_TO_TWENTY_corrected['Gender'].value_counts()
```

Gender	count
Female	219
Male	184

dtype: int64

```
adult_eleven_to_twenty=ELEVEN_TO_TWENTY_corrected[(ELEVEN_TO_TWENTY_corrected['Age']>=18) & (ELEVEN_TO_TWENTY_corrected['Age']<=20)]
a=percentage(adult_eleven_to_twenty.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
33.25062034739454
```

We can see that about 33% people of this age group are adults

```
adult_eleven_to_twenty['SmokingStatus'].value_counts()
```

SmokingStatus	count
never smoked	66
Unknown	49
smokes	14
formerly smoked	5

dtype: int64

```
adults_smoke=adult_eleven_to_twenty[adult_eleven_to_twenty['SmokingStatus']=='smokes']
a=percentage(adults_smoke.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
10.44776119402985
```

Therefore, only 10% adults of this age group are smokers

```
adults_nonsmoke=adult_eleven_to_twenty[adult_eleven_to_twenty['SmokingStatus']=='Unknown']
a=percentage(adults_nonsmoke.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
36.56716417910448
```

```
adult_unknownsmoke=adult_eleven_to_twenty[adult_eleven_to_twenty['SmokingStatus']=='never smoked']
a=percentage(adult_unknownsmoke.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
49.25373134328358
```

```
adult_formersmoker= adult_eleven_to_twenty[adult_eleven_to_twenty['SmokingStatus']=='formerly smoked']
a=percentage(adult_formersmoker.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
3.731343283582089
```

It can be seen that 37% adults are non-smokers, 49% adult's smoking status are unknown and 4% adult are former smokers

Finding how many are males and how many are females

```
adults_smoke_male=adults_smoke[adults_smoke['Gender']=='Male']
a=percentage(adults_smoke_male.shape[0],adults_smoke.shape[0])
print(a)
```

```
35.714285714285715
```

```
adults_smoke_female=adults_smoke[adults_smoke['Gender']=='Female']
a=percentage(adults_smoke_female.shape[0],adults_smoke.shape[0])
print(a)
```

```
64.28571428571429
```

In this scenario, about 36% are male and 64% are female

Now for the adults, lets find if those people both smokes and have tension.

```
adults_smoke_tension=adults_smoke[adults_smoke['HasTension']=='Yes']
a=percentage(adults_smoke_tension.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
0.0
```

So its evident that out of total adult smokers, none of the smokers has tension. So they smokes for any reason but having tension.

Now lets see if teenagers who smokes ever had an heartattack or not

```
teen_smokers_heartattack = smokers_below18[smokers_below18['HeartAttack'] == 'Yes']
a=percentage(teen_smokers_heartattack.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
0.0
```

Therefore, none of the teenagers who smokes had a heartattack

Now, lets see if adult smokers ever had an heartattack or not

```
adult_smokers_heartattack=adults_smoke[adults_smoke['HeartAttack']=='Yes']
a=percentage(adult_smokers_heartattack.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
0.0
```

Therefore, none of the adults who smokes had a heartattack

Now finding out what percentage of them works

```
adult_eleven_to_twenty['Occupation'].value_counts()
```

Occupation	count
Private	118
Govt_job	7
Self-employed	5
Never_worked	4

dtype: int64

```
eleven_to_twenty_adult_works = adult_eleven_to_twenty[(adult_eleven_to_twenty['Occupation'] != 'Never_worked')]
a = percentage(eleven_to_twenty_adult_works.shape[0], adult_eleven_to_twenty.shape[0])
print(a)
```

```
97.01492537313433
```

```
eleven_to_twenty_adult_not_works= adult_eleven_to_twenty[adult_eleven_to_twenty['Occupation']=='Never_worked']
a=percentage(eleven_to_twenty_adult_not_works.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
2.9850746268656714
```

```
adult_eleven_to_twenty_child=adult_eleven_to_twenty[adult_eleven_to_twenty['Occupation']=='Children']
a=percentage(adult_eleven_to_twenty_child.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
0.0
```

Therefore, among the adults 97% adult works and 3% doesn't work

Therefore, 61%What percentage of them are male and female

```
eleven_to_twenty_adult_works_male=eleven_to_twenty_adult_works[eleven_to_twenty_adult_works['Gender']=='Male']
a=percentage(eleven_to_twenty_adult_works_male.shape[0],eleven_to_twenty_adult_works.shape[0])
print(a)
```

```
42.30769230769231
```

```
eleven_to_twenty_adult_works_female=eleven_to_twenty_adult_works[eleven_to_twenty_adult_works['Gender']=='Female']
a=percentage(eleven_to_twenty_adult_works_female.shape[0],eleven_to_twenty_adult_works.shape[0])
print(a)
```

```
57.692307692307686
```

In this group about 97% works and 42% of them are male and 58% are female

Now finding what percentage works in which sector

```
eleven_to_twenty_adult_works['Occupation'].value_counts()
```

Occupation	count
Private	118
Govt_job	7
Self-employed	5

dtype: int64

```
eleven_to_twenty_private=eleven_to_twenty_adult_works[eleven_to_twenty_adult_works['Occupation']=='Private']
a=percentage(eleven_to_twenty_private.shape[0],eleven_to_twenty_adult_works.shape[0])
print(a)
```

```
90.76923076923077
```

```
eleven_to_twenty_private_male=eleven_to_twenty_private[eleven_to_twenty_private['Gender']=='Male']
a=percentage(eleven_to_twenty_private_male.shape[0],eleven_to_twenty_private.shape[0])
print(a)
```

```
41.52542372881356
```

```
eleven_to_twenty_private_female=eleven_to_twenty_private[eleven_to_twenty_private['Gender']=='Female']
a=percentage(eleven_to_twenty_private_female.shape[0],eleven_to_twenty_private.shape[0])
print(a)
```

```
58.47457627118644
```

91% adult works in private sector and among them 42% are male and 58% are female

```
eleven_to_twenty_govt=eleven_to_twenty_adult_works[eleven_to_twenty_adult_works['Occupation']=='Govt_job']
a=percentage(eleven_to_twenty_govt.shape[0],eleven_to_twenty_adult_works.shape[0])
print(a)
```

```
5.384615384615385
```

```
eleven_to_twenty_govt_male=eleven_to_twenty_govt[eleven_to_twenty_govt['Gender']=='Male']
a=percentage(eleven_to_twenty_govt_male.shape[0],eleven_to_twenty_govt.shape[0])
print(a)
```

```
42.857142857142854
```

```
eleven_to_twenty_govt_female=eleven_to_twenty_govt[eleven_to_twenty_govt['Gender']=='Female']
a=percentage(eleven_to_twenty_govt_female.shape[0],eleven_to_twenty_govt.shape[0])
print(a)
```

```
57.14285714285714
```

5% adult works in govt sector and among them 43% are male and 57% are female

```
eleven_to_twenty_self=eleven_to_twenty_adult_works[eleven_to_twenty_adult_works['Occupation']=='Self-employed']
a=percentage(eleven_to_twenty_self.shape[0],eleven_to_twenty_adult_works.shape[0])
print(a)
```

```
3.8461538461538463
```

```
eleven_to_twenty_self_male=eleven_to_twenty_self[eleven_to_twenty_self['Gender']=='Male']
a=percentage(eleven_to_twenty_self_male.shape[0],eleven_to_twenty_self.shape[0])
print(a)
```

```
60.0
```

```
eleven_to_twenty_self_female=eleven_to_twenty_self[eleven_to_twenty_self['Gender']=='Female']
a=percentage(eleven_to_twenty_self_female.shape[0],eleven_to_twenty_self.shape[0])
print(a)
```

```
40.0
```

4% are self employed and out of them 60% are male and 40% are female

Finding what percentage of adults are from village and city (Genderwise)

```
adult_village=adult_eleven_to_twenty[adult_eleven_to_twenty['LivesIn']=='Village']
a=percentage(adult_village.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
48.507462686567166
```

```
adult_village_male=adult_village[adult_village['Gender']=='Male']
a=percentage(adult_village_male.shape[0],adult_village.shape[0])
print(a)
```

```
47.69230769230769
```

```
adult_village_female=adult_village[adult_village['Gender']=='Female']
a=percentage(adult_village_female.shape[0],adult_village.shape[0])
print(a)
```

```
52.307692307692314
```

Among the adults from village, 48% are male and 52% are female

```
adult_city=adult_eleven_to_twenty[adult_eleven_to_twenty['LivesIn']=='City']
a=percentage(adult_city.shape[0],adult_eleven_to_twenty.shape[0])
print(a)
```

```
51.49253731343284
```

```
adult_city_male=adult_city[adult_city['Gender']=='Male']
a=percentage(adult_city_male.shape[0],adult_city.shape[0])
print(a)
```

```
34.78260869565217
```

```
adult_city_female= adult_city[adult_city['Gender']=='Female']
a=percentage(adult_city_female.shape[0],adult_city.shape[0])
print(a)
```

```
65.21739130434783
```

Among the adults from city 35% are male and 65% are female

Therefore 49% adults are from village and 51% are from city

Now, finding how many adults are unemployed and from where ?

```
eleven_to_twenty_adult_not_works['LivesIn'].value_counts()
```

LivesIn	count
City	4

dtype: int64

```
adult_unemployed_city= eleven_to_twenty_adult_not_works[eleven_to_twenty_adult_not_works['LivesIn']=='City']
a=percentage(adult_unemployed_city.shape[0],eleven_to_twenty_adult_not_works.shape[0])
print(a)
```

```
100.0
```

```
adult_unemployed_city_male=adult_unemployed_city[adult_unemployed_city['Gender']=='Male']
a=percentage(adult_unemployed_city_male.shape[0],adult_unemployed_city.shape[0])
print(a)
```

```
0.0
```

```
adult_unemployed_city_female=adult_unemployed_city[adult_unemployed_city['Gender']=='Female']
a=percentage(adult_unemployed_city_female.shape[0],adult_unemployed_city.shape[0])
print(a)
```

```
100.0
```

Therefore, the adults who are unemployed are all from city and all of them are female

Now finding how many of them married ?

```
adult_unemployed_city_female['NeverMarried'].value_counts()
```

NeverMarried	count
No	4

dtype: int64

Therefore All of the unemployed adults are unmarried

now finding how many obese are from city and village

```
eleven_to_twenty_obese=ELEVEN_TO_TWENTY_corrected[ELEVEN_TO_TWENTY_corrected['BMI']>=30]
a=percentage(eleven_to_twenty_obese.shape[0],ELEVEN_TO_TWENTY_corrected.shape[0])
print(a)
```

```
18.610421836228287
```

```
eleven_to_twenty_obese_male=eleven_to_twenty_obese[eleven_to_twenty_obese['Gender']=='Male']
a=percentage(eleven_to_twenty_obese_male.shape[0],eleven_to_twenty_obese.shape[0])
print(a)
```

```
45.33333333333333
```

```
eleven_to_twenty_obese_female=eleven_to_twenty_obese[eleven_to_twenty_obese['Gender']=='Female']
a=percentage(eleven_to_twenty_obese_female.shape[0],eleven_to_twenty_obese.shape[0])
print(a)
```

```
54.666666666666664
```

Therefore 19% of this group has obesity and out of them 45% are male and 55% are female

```
eleven_to_twenty_obese_village=eleven_to_twenty_obese[eleven_to_twenty_obese['LivesIn']=='Village']
a=percentage(eleven_to_twenty_obese_village.shape[0],eleven_to_twenty_obese.shape[0])
print(a)
```

```
54.666666666666664
```

```
eleven_to_twenty_obese_village_male=eleven_to_twenty_obese_village[eleven_to_twenty_obese_village['Gender']=='Male']
a=percentage(eleven_to_twenty_obese_village_male.shape[0],eleven_to_twenty_obese_village.shape[0])
print(a)
```

```
46.34146341463415
```

```
eleven_to_twenty_obese_village_female= eleven_to_twenty_obese_village[eleven_to_twenty_obese_village['Gender']=='Female']
a=percentage(eleven_to_twenty_obese_village_female.shape[0],eleven_to_twenty_obese_village.shape[0])
print(a)
```

```
53.65853658536586
```

55% obese are from village and out of them 46% are male and 54% are female

```
eleven_to_twenty_obese_city=eleven_to_twenty_obese[eleven_to_twenty_obese['LivesIn']=='City']
a=percentage(eleven_to_twenty_obese_city.shape[0],eleven_to_twenty_obese.shape[0])
print(a)
```

```
45.33333333333333
```

```
eleven_to_twenty_obese_city_male=eleven_to_twenty_obese_city[eleven_to_twenty_obese_city['Gender']=='Male']
a=percentage(eleven_to_twenty_obese_city_male.shape[0],eleven_to_twenty_obese_city.shape[0])
print(a)
```

```
44.11764705882353
```

```
eleven_to_twenty_obese_city_female=eleven_to_twenty_obese_city[eleven_to_twenty_obese_city['Gender']=='Female']
a=percentage(eleven_to_twenty_obese_city_female.shape[0],eleven_to_twenty_obese_city.shape[0])
print(a)
```

```
55.88235294117647
```

45% obese are from city and out of them 44% are male and 56% are female

Now, clustering dataset where age limit is from 21 to 30

```
TWENTYONE_TO_THIRTY_DF=correct_df_final[(correct_df_final['Age']>=21) & (correct_df_final['Age']<=30)]
TWENTYONE_TO_THIRTY_DF
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
21	Female	29	No	No	No	Private	City	57.3	26.1	never smoked	No
29	Female	26	No	No	No	Private	City	116.4	18.8	formerly smoked	No
47	Female	29	No	No	Yes	Private	Village	91.7	24.3	never smoked	No
53	Female	27	No	No	No	Self-employed	City	73.0	24.3	Unknown	No
64	Female	25	No	No	Yes	Private	Village	82.7	39.4	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3976	Female	22	No	No	No	Private	City	140.0	20.6	never smoked	No
3988	Female	28	No	No	Yes	Private	City	111.2	26.2	Unknown	No
3994	Female	27	No	No	No	Private	Village	85.7	21.1	Unknown	No
3997	Male	27	No	No	Yes	Private	City	76.5	21.0	never smoked	No
3998	Female	28	No	No	No	Private	City	80.0	27.1	smokes	No

420 rows × 11 columns

Next steps: Generate code with TWENTYONE_TO_THIRTY_DF | View recommended plots | New interactive sheet

```
TWENTYONE_TO_THIRTY_corrected=TWENTYONE_TO_THIRTY_DF
```

Now finding how many are male and how many are female

```
TWENTYONE_TO_THIRTY_male=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_male.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
30.714285714285715
```

```
TWENTYONE_TO_THIRTY_female=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_female.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
69.28571428571428
```

Therefore, 31% of this age group are male and 69% are female

Its known that bmi level equal or above 30 indicates Obesity. so finding out people who has obesity

```
bmi_30=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['BMI']>=30]
a=percentage(bmi_30.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
30.0
```

About 30% of this age group has obesity

Now from this cluster searching how many are these are male and how many are female

```
bmi30_male=bmi_30[bmi_30['Gender']=='Male']
a=percentage(bmi30_male.shape[0],bmi_30.shape[0])
print(a)
```

```
29.365079365079367
```

```
bmi30_female=bmi_30[bmi_30['Gender']=='Female']
a=percentage(bmi30_female.shape[0],bmi_30.shape[0])
print(a)
```

```
70.63492063492063
```

So, out of total 30% of this group has obesity, and it can be seen that about 29% are male and about 71% are female

Now, obesity can cause heart-attack. so sorting out the obese people who have heart-attack

```
bmi30_heartattack=bmi_30[bmi_30['HeartAttack']=='Yes']
a=percentage(bmi30_heartattack.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
0.0
```

So, obese people in this dataset haven't had heartattack.

Now finding out if they have any heart disease or not.

```
bmi30_heartdisease=bmi_30[bmi_30['AnyHeartDisease']=='Yes']
a=percentage(bmi30_heartdisease.shape[0],bmi_30.shape[0])
print(a)
```

```
0.0
```

So, all the obese people do not have any heart disease either.

Now, it is known that if obese people smokes, the chances of having circulatory disease will increase. Therefore, checking out the parameter.

```
TWENTYONE_TO_THIRTY_obese_smokes=bmi_30[bmi_30['SmokingStatus']=='smokes']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes.shape[0],bmi_30.shape[0])
print(a)
```

```
21.428571428571427
```

So about 21% of total obese people of this group smokes. Now, finding out how many are male and how many are female

```
TWENTYONE_TO_THIRTY_obese_smokes_male=TWENTYONE_TO_THIRTY_obese_smokes[TWENTYONE_TO_THIRTY_obese_smokes['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes_male.shape[0],TWENTYONE_TO_THIRTY_obese_smokes.shape[0])
print(a)
```

```
25.925925925925924
```

```
TWENTYONE_TO_THIRTY_obese_smokes_female=TWENTYONE_TO_THIRTY_obese_smokes[TWENTYONE_TO_THIRTY_obese_smokes['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes_female.shape[0],TWENTYONE_TO_THIRTY_obese_smokes.shape[0])
print(a)
```

```
74.07407407407408
```

Therefore, there exists about 21% who has obesity and has smoking habbit and out of them 74% females and about 26% males who has obesity and also has a smoking habbit.

Now, searching for those obese males and females who got rid of their smoking habbits

```
TWENTYONE_TO_THIRTY_obese_formersmokes=bmi_30[bmi_30['SmokingStatus']=='formerly smoked']
a=percentage(TWENTYONE_TO_THIRTY_obese_formersmokes.shape[0],bmi_30.shape[0])
print(a)
```

```
12.698412698412698
```

```
TWENTYONE_TO_THIRTY_obese_formersmokes_male = TWENTYONE_TO_THIRTY_obese_formersmokes[(TWENTYONE_TO_THIRTY_obese_formersmokes['Gender'] == 'Male')]
a=percentage(TWENTYONE_TO_THIRTY_obese_formersmokes_male.shape[0],TWENTYONE_TO_THIRTY_obese_formersmokes.shape[0])
print(a)
```

```
43.75
```

```
TWENTYONE_TO_THIRTY_obese_formersmokes_female = TWENTYONE_TO_THIRTY_obese_formersmokes[(TWENTYONE_TO_THIRTY_obese_formersmokes['Gender'] == 'Female')]
a=percentage(TWENTYONE_TO_THIRTY_obese_formersmokes_female.shape[0], TWENTYONE_TO_THIRTY_obese_formersmokes.shape[0])
print(a)
```

```
56.25
```

so, about 13% obese people has got rid of their smoking habbits

and among them, 44% are male and 56% are female

Now searching for obese males and females who smokes and has tension

```
TWENTYONE_TO_THIRTY_obese_smokes_tension=TWENTYONE_TO_THIRTY_obese_smokes[TWENTYONE_TO_THIRTY_obese_smokes['HasTension']=='Yes']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes_tension.shape[0],bmi_30.shape[0])
print(a)
```

```
3.1746031746031744
```

```
TWENTYONE_TO_THIRTY_obese_smokes_tension_male =TWENTYONE_TO_THIRTY_obese_smokes_tension[TWENTYONE_TO_THIRTY_obese_smokes_tension['Gender'] == 'Male']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes_tension_male.shape[0],TWENTYONE_TO_THIRTY_obese_smokes_tension.shape[0])
print(a)
```

```
25.0
```

```
TWENTYONE_TO_THIRTY_obese_smokes_tension_female= TWENTYONE_TO_THIRTY_obese_smokes_tension[TWENTYONE_TO_THIRTY_obese_smokes_tension['Gender'] == 'Female']
a=percentage(TWENTYONE_TO_THIRTY_obese_smokes_tension_female.shape[0],TWENTYONE_TO_THIRTY_obese_smokes_tension.shape[0])
print(a)
```

```
75.0
```

So, 3% of total obese who smokes also has tensions, and out of them 25% of them are male and 75% are female

Now finding how many does work in govt, private sectors and etc

```
TWENTYONE_TO_THIRTY_corrected['Occupation'].value_counts()
```

Occupation	count
Private	354
Govt_job	39
Self-employed	26
Never_worked	1

dtype: int64

```
TWENTYONE_TO_THIRTY_govt=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Occupation']=='Govt_job']
a=percentage(TWENTYONE_TO_THIRTY_govt.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
9.285714285714286
```

```
TWENTYONE_TO_THIRTY_govt_male=TWENTYONE_TO_THIRTY_govt[TWENTYONE_TO_THIRTY_govt['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_govt_male.shape[0],TWENTYONE_TO_THIRTY_govt.shape[0])
print(a)
```

```
28.205128205128204
```

```
TWENTYONE_TO_THIRTY_govt_female=TWENTYONE_TO_THIRTY_govt[TWENTYONE_TO_THIRTY_govt['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_govt_female.shape[0],TWENTYONE_TO_THIRTY_govt.shape[0])
print(a)
```

```
71.7948717948718
```

9% works in govt sector and among them 28% are male and 72% are female

```
TWENTYONE_TO_THIRTY_private=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Occupation']=='Private']
a=percentage(TWENTYONE_TO_THIRTY_private.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
84.28571428571429
```

```
TWENTYONE_TO_THIRTY_private_male=TWENTYONE_TO_THIRTY_private[TWENTYONE_TO_THIRTY_private['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_private_male.shape[0],TWENTYONE_TO_THIRTY_private.shape[0])
print(a)
```

```
29.66101694915254
```

```
TWENTYONE_TO_THIRTY_private_female=TWENTYONE_TO_THIRTY_private[TWENTYONE_TO_THIRTY_private['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_private_female.shape[0],TWENTYONE_TO_THIRTY_private.shape[0])
print(a)
```

```
70.33898305084746
```

84% works in private sector, among them 30% are male and 70% are female

```
TWENTYONE_TO_THIRTY_none=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Occupation']=='Never_worked']
a=percentage(TWENTYONE_TO_THIRTY_none.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
0.2380952380952381
```

```
TWENTYONE_TO_THIRTY_none_male= TWENTYONE_TO_THIRTY_none[TWENTYONE_TO_THIRTY_none['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_none_male.shape[0],TWENTYONE_TO_THIRTY_none.shape[0])
print(a)
```

```
100.0
```

```
TWENTYONE_TO_THIRTY_none_female=TWENTYONE_TO_THIRTY_none[TWENTYONE_TO_THIRTY_none['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_none_female.shape[0],TWENTYONE_TO_THIRTY_none.shape[0])
```

```
print(a)
```

```
0.0
```

Only few(near to 0) are unemployed and among them all are male

```
TWENTYONE_TO_THIRTY_self=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Occupation']=='Self-employed']
a=percentage(TWENTYONE_TO_THIRTY_self.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
6.190476190476191
```

```
TWENTYONE_TO_THIRTY_self_male=TWENTYONE_TO_THIRTY_self[TWENTYONE_TO_THIRTY_self['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_self_male.shape[0],TWENTYONE_TO_THIRTY_self.shape[0])
print(a)
```

```
46.15384615384615
```

```
TWENTYONE_TO_THIRTY_self_female=TWENTYONE_TO_THIRTY_self[TWENTYONE_TO_THIRTY_self['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_self_female.shape[0],TWENTYONE_TO_THIRTY_self.shape[0])
print(a)
```

```
53.84615384615385
```

About 6% are self employed and among them 46% are male and 54% are female

Now finding out what percentage of them has tension genderwise

```
TWENTYONE_TO_THIRTY_govt_tension=TWENTYONE_TO_THIRTY_govt[TWENTYONE_TO_THIRTY_govt['HasTension']=='Yes']
a=percentage(TWENTYONE_TO_THIRTY_govt_tension.shape[0],TWENTYONE_TO_THIRTY_govt.shape[0])
print(a)
```

```
2.564102564102564
```

```
TWENTYONE_TO_THIRTY_govt_tension_male=TWENTYONE_TO_THIRTY_govt_tension[TWENTYONE_TO_THIRTY_govt_tension['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_govt_tension_male.shape[0],TWENTYONE_TO_THIRTY_govt_tension.shape[0])
print(a)
```

```
0.0
```

```
TWENTYONE_TO_THIRTY_govt_tension_female=TWENTYONE_TO_THIRTY_govt_tension[TWENTYONE_TO_THIRTY_govt_tension['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_govt_tension_female.shape[0],TWENTYONE_TO_THIRTY_govt_tension.shape[0])
print(a)
```

```
100.0
```

3% govt employees has tension and out of them all are female

```
TWENTYONE_TO_THIRTY_private_tension=TWENTYONE_TO_THIRTY_private[TWENTYONE_TO_THIRTY_private['HasTension']=='Yes']
a=percentage(TWENTYONE_TO_THIRTY_private_tension.shape[0],TWENTYONE_TO_THIRTY_private.shape[0])
print(a)
```

```
1.977401129943503
```

```
TWENTYONE_TO_THIRTY_private_tension_male=TWENTYONE_TO_THIRTY_private_tension[TWENTYONE_TO_THIRTY_private_tension['Gender']=='Male']
a=percentage(TWENTYONE_TO_THIRTY_private_tension_male.shape[0],TWENTYONE_TO_THIRTY_private_tension.shape[0])
print(a)
```

```
28.57142857142857
```

```
TWENTYONE_TO_THIRTY_private_tension_female=TWENTYONE_TO_THIRTY_private_tension[TWENTYONE_TO_THIRTY_private_tension['Gender']=='Female']
a=percentage(TWENTYONE_TO_THIRTY_private_tension_female.shape[0],TWENTYONE_TO_THIRTY_private_tension.shape[0])
print(a)
```

```
71.42857142857143
```

2% private employees has tension and out of them 29% are male and 71% are female

```
TWENTYONE_TO_THIRTY_self_tension=TWENTYONE_TO_THIRTY_self[TWENTYONE_TO_THIRTY_self['HasTension']=='Yes']
a=percentage(TWENTYONE_TO_THIRTY_self_tension.shape[0],TWENTYONE_TO_THIRTY_self.shape[0])
print(a)
```

```
0.0
```

```
TWENTYONE_TO_THIRTY_none_tension=TWENTYONE_TO_THIRTY_none[TWENTYONE_TO_THIRTY_none['HasTension']=='Yes']
a=percentage(TWENTYONE_TO_THIRTY_none_tension.shape[0],TWENTYONE_TO_THIRTY_none.shape[0])
print(a)
```

```
0.0
```

Therefore,No self-employed person and no unemployed person has tension

Finding how many unemployed people are from city and village and genderwise

```
twentyone_to_thirty_unemployed=TWENTYONE_TO_THIRTY_corrected[TWENTYONE_TO_THIRTY_corrected['Occupation']=='Never_worked']
a=percentage(twentyone_to_thirty_unemployed.shape[0],TWENTYONE_TO_THIRTY_corrected.shape[0])
print(a)
```

```
0.2380952380952381
```

```
twentyone_to_thirty_unemployed_city=twentyone_to_thirty_unemployed[twentyone_to_thirty_unemployed['LivesIn']=='City']
a=percentage(twentyone_to_thirty_unemployed_city.shape[0],twentyone_to_thirty_unemployed.shape[0])
print(a)
```

```
100.0
```

```
twentyone_to_thirty_unemployed_city_male=twentyone_to_thirty_unemployed_city[twentyone_to_thirty_unemployed_city['Gender']=='Male']
a=percentage(twentyone_to_thirty_unemployed_city_male.shape[0],twentyone_to_thirty_unemployed_city.shape[0])
print(a)
```

```
100.0
```

Therefore, it can be seen that the 1 person who is unemployed in this group is from city and the person is a male now finding if he has tension or not

```
twentyone_to_thirty_unemployed_city_male_tension=twentyone_to_thirty_unemployed_city_male[twentyone_to_thirty_unemployed_city_male['HasTension']=='Yes']
a=percentage(twentyone_to_thirty_unemployed_city_male_tension.shape[0],twentyone_to_thirty_unemployed_city_male.shape[0])
print(a)
```

```
0.0
```

It means that he has no tension

Now, clustering data with age parameter 31 to 40

```
THIRTYONE_TO_FORTY_df=correct_df_final[(correct_df_final['Age']>=31) & (correct_df_final['Age']<=40)]
THIRTYONE_TO_FORTY_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
2	Male	32	No	No	Yes	Private	City	64.1	23.4	smokes	No
4	Male	39	No	No	Yes	Private	City	55.4	41.6	formerly smoked	No
11	Male	31	No	No	Yes	Private	Village	101.4	29.5	Unknown	No
17	Female	38	No	No	Yes	Govt_job	Village	77.1	25.5	never smoked	No
20	Female	34	No	No	Yes	Private	Village	231.8	44.7	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3970	Male	36	No	No	Yes	Private	Village	119.6	39.8	never smoked	No
3975	Female	39	No	No	Yes	Private	Village	69.7	27.9	Unknown	No
3985	Female	38	No	No	Yes	Private	City	70.7	22.4	Unknown	No
3989	Male	37	No	No	Yes	Govt_job	Village	72.5	34.9	never smoked	No
3995	Female	40	No	No	Yes	Private	City	88.4	36.5	smokes	No

510 rows × 11 columns

Next steps: Generate code with THIRTYONE_TO_FORTY_df | View recommended plots | New interactive sheet

```
THIRTYONE_TO_FORTY_corrected=THIRTYONE_TO_FORTY_df
```

What percentage of this group are male and female

```
THIRTYONE_TO_FORTY_male=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_male.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
32.94117647058823
```

```
THIRTYONE_TO_FORTY_female=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_female.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
67.05882352941175
```

About 33% of this group are male and 67% are female

Finding how many has obese

```
THIRTYONE_TO_FORTY_bmi30=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['BMI']>=30]
a=percentage(THIRTYONE_TO_FORTY_bmi30.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
```

```
print(a)
```

```
49.6078431372549
```

About 50% of this age group has obese

Finding genderwise

```
THIRTYONE_TO_FORTY_obese_male=THIRTYONE_TO_FORTY_bmi30[THIRTYONE_TO_FORTY_bmi30['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_obese_male.shape[0],THIRTYONE_TO_FORTY_bmi30.shape[0])
print(a)
```

```
35.96837944664031
```

```
THIRTYONE_TO_FORTY_obese_female=THIRTYONE_TO_FORTY_bmi30[THIRTYONE_TO_FORTY_bmi30['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_obese_female.shape[0],THIRTYONE_TO_FORTY_bmi30.shape[0])
print(a)
```

```
64.03162055335969
```

Therefore, about 36% of total obese people are male and rest 64% are female.

Finding whether they had heartattacks, genderwise

```
THIRTYONE_TO_FORTY_obese_heartattack=THIRTYONE_TO_FORTY_bmi30[THIRTYONE_TO_FORTY_bmi30['HeartAttack']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_obese_heartattack.shape[0],THIRTYONE_TO_FORTY_bmi30.shape[0])
print(a)
```

```
0.0
```

Therefore, no obese smokers have had a heartattack

now searching if they had any heart disease, genderwise

```
THIRTYONE_TO_FORTY_obese_heartdisease=THIRTYONE_TO_FORTY_bmi30[THIRTYONE_TO_FORTY_bmi30['AnyHeartDisease']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_obese_heartdisease.shape[0],THIRTYONE_TO_FORTY_bmi30.shape[0])
print(a)
```

```
0.0
```

Therefore, no obese people had any heart disease

Now, Finding obese people who also smokes, genderwise

```
THIRTYONE_TO_FORTY_obese_smokes=THIRTYONE_TO_FORTY_bmi30[THIRTYONE_TO_FORTY_bmi30['SmokingStatus']=='smokes']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes.shape[0],THIRTYONE_TO_FORTY_bmi30.shape[0])
print(a)
```

```
20.55335968379447
```

```
THIRTYONE_TO_FORTY_obese_smokes_male= THIRTYONE_TO_FORTY_obese_smokes[THIRTYONE_TO_FORTY_obese_smokes['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes_male.shape[0],THIRTYONE_TO_FORTY_obese_smokes.shape[0])
print(a)
```

```
36.53846153846153
```

```
THIRTYONE_TO_FORTY_obese_smokes_female= THIRTYONE_TO_FORTY_obese_smokes[THIRTYONE_TO_FORTY_obese_smokes['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes_female.shape[0],THIRTYONE_TO_FORTY_obese_smokes.shape[0])
print(a)
```

```
63.46153846153846
```

Therefore, out of total obese about 21% smokes and out of them 37% are male and 63% are female

now finding out Obese smokers who also got tension

```
THIRTYONE_TO_FORTY_obese_smokes_tension=THIRTYONE_TO_FORTY_obese_smokes[THIRTYONE_TO_FORTY_obese_smokes['HasTension']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes_tension.shape[0],THIRTYONE_TO_FORTY_obese_smokes.shape[0])
print(a)
```

```
7.6923076923076925
```

```
THIRTYONE_TO_FORTY_obese_smokes_male_tension=THIRTYONE_TO_FORTY_obese_smokes_tension[THIRTYONE_TO_FORTY_obese_smokes_tension['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes_male_tension.shape[0],THIRTYONE_TO_FORTY_obese_smokes_tension.shape[0])
print(a)
```

```
50.0
```

```
THIRTYONE_TO_FORTY_obese_smokes_female_tension=THIRTYONE_TO_FORTY_obese_smokes_tension[THIRTYONE_TO_FORTY_obese_smokes_tension['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_obese_smokes_female_tension.shape[0],THIRTYONE_TO_FORTY_obese_smokes_tension.shape[0])
print(a)
```

```
50.0
```

Therefore, In this age group about 8% of total obese smokers got tension and out of them 50% are male and 50% are female

Now finding how many people of this group smokes and also got tension genderwise

```
THIRTYONE_TO_FORTY_corrected['SmokingStatus'].value_counts()
```

SmokingStatus	count
never smoked	231
Unknown	111
smokes	99
formerly smoked	69

dtype: int64

```
THIRTYONE_TO_FORTY_smokes= THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['SmokingStatus']=='smokes']
a=percentage(THIRTYONE_TO_FORTY_smokes.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
19.411764705882355
```

```
THIRTYONE_TO_FORTY_smokes_male=THIRTYONE_TO_FORTY_smokes[THIRTYONE_TO_FORTY_smokes['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_smokes_male.shape[0],THIRTYONE_TO_FORTY_smokes.shape[0])
print(a)
```

```
36.36363636363637
```

```
THIRTYONE_TO_FORTY_smokes_female=THIRTYONE_TO_FORTY_smokes[THIRTYONE_TO_FORTY_smokes['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_smokes_female.shape[0],THIRTYONE_TO_FORTY_smokes.shape[0])
print(a)
```

```
63.63636363636363
```

Therefore, 19% of this group smokes and out of them, 36% are male and 64% are female

now finding how many smokers got tension

```
THIRTYONE_TO_FORTY_smokes_tension= THIRTYONE_TO_FORTY_smokes[THIRTYONE_TO_FORTY_smokes['HasTension']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_smokes_tension.shape[0],THIRTYONE_TO_FORTY_smokes.shape[0])
print(a)
```

```
5.05050505050505
```

```
THIRTYONE_TO_FORTY_smokes_tension_male=THIRTYONE_TO_FORTY_smokes_tension[THIRTYONE_TO_FORTY_smokes_tension['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_smokes_tension_male.shape[0],THIRTYONE_TO_FORTY_smokes_tension.shape[0])
print(a)
```

```
60.0
```

```
THIRTYONE_TO_FORTY_smokes_tension_female=THIRTYONE_TO_FORTY_smokes_tension[THIRTYONE_TO_FORTY_smokes_tension['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_smokes_tension_female.shape[0],THIRTYONE_TO_FORTY_smokes_tension.shape[0])
```

```
print(a)
```

```
40.0
```

Therefore in this group, about 5% smokers got tension and out of them 60% are male and 40% are female

Now, lets find out how many obese men and women lives in city and village

```
THIRTYONE_TO_FORTY_obese_male_city=THIRTYONE_TO_FORTY_obese_male[THIRTYONE_TO_FORTY_obese_male['LivesIn']=='City']
a=percentage(THIRTYONE_TO_FORTY_obese_male_city.shape[0],THIRTYONE_TO_FORTY_obese_male.shape[0])
print(a)
```

```
49.45054945054945
```

```
THIRTYONE_TO_FORTY_obese_male_village=THIRTYONE_TO_FORTY_obese_male[THIRTYONE_TO_FORTY_obese_male['LivesIn']=='Village']
a=percentage(THIRTYONE_TO_FORTY_obese_male_village.shape[0],THIRTYONE_TO_FORTY_obese_male.shape[0])
print(a)
```

```
50.54945054945055
```

Therefore, almost 49% of obese male are from village and other 51% from city

```
THIRTYONE_TO_FORTY_obese_female_city=THIRTYONE_TO_FORTY_obese_female[THIRTYONE_TO_FORTY_obese_female['LivesIn']=='City']
a=percentage(THIRTYONE_TO_FORTY_obese_female_city.shape[0],THIRTYONE_TO_FORTY_obese_female.shape[0])
print(a)
```

```
48.148148148148145
```

```
THIRTYONE_TO_FORTY_obese_female_village=THIRTYONE_TO_FORTY_obese_female[THIRTYONE_TO_FORTY_obese_female['LivesIn']=='Village']
a=percentage(THIRTYONE_TO_FORTY_obese_female_village.shape[0],THIRTYONE_TO_FORTY_obese_female.shape[0])
print(a)
```

```
51.85185185185185
```

Therefore, 48% obese male are from city and 52% are from village

What percentage of them works at which sector (genderwise)

```
THIRTYONE_TO_FORTY_corrected.Occupation.value_counts()
```

Occupation	count
Private	383
Govt_job	71
Self-employed	56

dtype: int64

```
THIRTYONE_TO_FORTY_corrected.shape
```

```
(510, 11)
```

```
THIRTYONE_TO_FORTY_govt=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Occupation']=='Govt_job']
a=percentage(THIRTYONE_TO_FORTY_govt.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
13.92156862745098
```

```
THIRTYONE_TO_FORTY_govt_male=THIRTYONE_TO_FORTY_govt[THIRTYONE_TO_FORTY_govt['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_govt_male.shape[0],THIRTYONE_TO_FORTY_govt.shape[0])
print(a)
```

```
30.985915492957744
```

```
THIRTYONE_TO_FORTY_govt_female=THIRTYONE_TO_FORTY_govt[THIRTYONE_TO_FORTY_govt['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_govt_female.shape[0],THIRTYONE_TO_FORTY_govt.shape[0])
print(a)
```

```
69.01408450704226
```

TErefor 14% works in govt sector and out of them 31% are male and 69% are female

```
THIRTYONE_TO_FORTY_private=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Occupation']=='Private']
a=percentage(THIRTYONE_TO_FORTY_private.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
75.09803921568627
```

```
THIRTYONE_TO_FORTY_private_male=THIRTYONE_TO_FORTY_private[THIRTYONE_TO_FORTY_private['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_private_male.shape[0],THIRTYONE_TO_FORTY_private.shape[0])
print(a)
```

```
33.68146214099217
```

```
THIRTYONE_TO_FORTY_private_female=THIRTYONE_TO_FORTY_private[THIRTYONE_TO_FORTY_private['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_private_female.shape[0],THIRTYONE_TO_FORTY_private.shape[0])
print(a)
```

```
66.31853785900783
```

Therefore, 75% works in private sector and out of them 34% are male and 66% are female

```
THIRTYONE_TO_FORTY_self=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Occupation']=='Self-employed']
a=percentage(THIRTYONE_TO_FORTY_self.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
10.980392156862745
```

```
THIRTYONE_TO_FORTY_self_male= THIRTYONE_TO_FORTY_self[THIRTYONE_TO_FORTY_self['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_self_male.shape[0],THIRTYONE_TO_FORTY_self.shape[0])
print(a)
```

```
30.357142857142854
```

```
THIRTYONE_TO_FORTY_self_female=THIRTYONE_TO_FORTY_self[THIRTYONE_TO_FORTY_self['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_self_female.shape[0],THIRTYONE_TO_FORTY_self.shape[0])
print(a)
```

```
69.64285714285714
```

11% are self employed and out of them 30% are male and 70% are female

```
THIRTYONE_TO_FORTY_none=THIRTYONE_TO_FORTY_corrected[THIRTYONE_TO_FORTY_corrected['Occupation']=='Never_worked']
a=percentage(THIRTYONE_TO_FORTY_none.shape[0],THIRTYONE_TO_FORTY_corrected.shape[0])
print(a)
```

```
0.0
```

Therefore, None of them are unemployed

Now finding out how many of them has tension

```
THIRTYONE_TO_FORTY_govt_tension=THIRTYONE_TO_FORTY_govt[THIRTYONE_TO_FORTY_govt['HasTension']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_govt_tension.shape[0],THIRTYONE_TO_FORTY_govt.shape[0])
print(a)
```

```
2.8169014084507045
```

```
THIRTYONE_TO_FORTY_govt_tension_male=THIRTYONE_TO_FORTY_govt_tension[THIRTYONE_TO_FORTY_govt_tension['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_govt_tension_male.shape[0],THIRTYONE_TO_FORTY_govt_tension.shape[0])
print(a)
```

```
100.0
```

```
THIRTYONE_TO_FORTY_govt_tension_female=THIRTYONE_TO_FORTY_govt_tension[THIRTYONE_TO_FORTY_govt_tension['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_govt_tension_female.shape[0],THIRTYONE_TO_FORTY_govt_tension.shape[0])
print(a)
```

```
0.0
```

3% govt employees has tension and out of them all are male

```
THIRTYONE_TO_FORTY_private_tension=THIRTYONE_TO_FORTY_private[THIRTYONE_TO_FORTY_private['HasTension']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_private_tension.shape[0],THIRTYONE_TO_FORTY_private.shape[0])
print(a)
```

```
3.91644908616188
```

```
THIRTYONE_TO_FORTY_private_tension_male=THIRTYONE_TO_FORTY_private_tension[THIRTYONE_TO_FORTY_private_tension['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_private_tension_male.shape[0],THIRTYONE_TO_FORTY_private_tension.shape[0])
print(a)
```

```
40.0
```

```
THIRTYONE_TO_FORTY_private_tension_female=THIRTYONE_TO_FORTY_private_tension[THIRTYONE_TO_FORTY_private_tension['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_private_tension_female.shape[0],THIRTYONE_TO_FORTY_private_tension.shape[0])
print(a)
```

```
60.0
```

4% private workers got tension and 40% of them are male and 60% are female

```
THIRTYONE_TO_FORTY_self_tension=THIRTYONE_TO_FORTY_self[THIRTYONE_TO_FORTY_self['HasTension']=='Yes']
a=percentage(THIRTYONE_TO_FORTY_self_tension.shape[0],THIRTYONE_TO_FORTY_self.shape[0])
print(a)
```

```
7.142857142857142
```

```
THIRTYONE_TO_FORTY_self_tension_male=THIRTYONE_TO_FORTY_self_tension[THIRTYONE_TO_FORTY_self_tension['Gender']=='Male']
a=percentage(THIRTYONE_TO_FORTY_self_tension_male.shape[0],THIRTYONE_TO_FORTY_self_tension.shape[0])
print(a)
```

```
50.0
```

```
THIRTYONE_TO_FORTY_self_tension_female=THIRTYONE_TO_FORTY_self_tension[THIRTYONE_TO_FORTY_self_tension['Gender']=='Female']
a=percentage(THIRTYONE_TO_FORTY_self_tension_female.shape[0],THIRTYONE_TO_FORTY_self_tension.shape[0])
```

```
print(a)
```

```
50.0
```

Therefore,

7% Self employed people has tension and out of them 50% are male and 50% are female

No one is unemployed, so finding out non employed people percentage of village and city is absurd

And, finding out unemployed people percentage who got tension is also absurd

Now clustering data for age group 41 to 50

```
FORTYONE_TO_FIFTY_df=correct_df_final[(correct_df_final['Age']>=41) & (correct_df_final['Age']<=50)]
FORTYONE_TO_FIFTY_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
1	Female	49	No	No	Yes	Private	Village	108.8	26.7	smokes	No
8	Female	45	No	No	Yes	Private	City	64.6	40.5	never smoked	No
10	Female	45	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
12	Female	45	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
18	Male	45	No	No	Yes	Private	Village	56.6	22.3	smokes	No
...	...	...	...	...	...	...	...	...	...	...	...
3981	Female	48	No	No	Yes	Govt_job	Village	77.0	26.2	Unknown	No
3982	Male	47	No	No	Yes	Govt_job	City	66.3	32.3	never smoked	No
3984	Female	43	No	No	Yes	Govt_job	Village	80.4	24.7	never smoked	No
3990	Female	45	No	No	Yes	Private	Village	89.4	21.3	formerly smoked	No
3992	Female	46	No	No	Yes	Private	City	62.2	27.3	Unknown	No

558 rows × 11 columns

Next steps: Generate code with FORTYONE_TO_FIFTY_df | View recommended plots | New interactive sheet

So, there are 558 people who are in the age range of 41-50

Now, filtering the data

Dropping the incorrected data

```
FORTYONE_TO_FIFTY_correct_df_final = FORTYONE_TO_FIFTY_df
FORTYONE_TO_FIFTY_correct_df_final
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
1	Female	49	No	No	Yes	Private	Village	108.8	26.7	smokes	No
8	Female	45	No	No	Yes	Private	City	64.6	40.5	never smoked	No
10	Female	45	No	No	Yes	Self-employed	City	110.6	29.9	never smoked	No
12	Female	45	No	No	Yes	Govt_job	City	81.4	40.7	never smoked	No
18	Male	45	No	No	Yes	Private	Village	56.6	22.3	smokes	No
...	...	...	...	...	...	...	...	...	...	...	...
3981	Female	48	No	No	Yes	Govt_job	Village	77.0	26.2	Unknown	No
3982	Male	47	No	No	Yes	Govt_job	City	66.3	32.3	never smoked	No
3984	Female	43	No	No	Yes	Govt_job	Village	80.4	24.7	never smoked	No
3990	Female	45	No	No	Yes	Private	Village	89.4	21.3	formerly smoked	No
3992	Female	46	No	No	Yes	Private	City	62.2	27.3	Unknown	No

558 rows × 11 columns

Next steps: Generate code with FORTYONE_TO_FIFTY_df | View recommended plots | New interactive sheet

What percentage of them are male and female

```
FORTYONE_TO_FIFTY_male=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_male.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
37.27598566308244
```

```
FORTYONE_TO_FIFTY_female= FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_female.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
62.72401433691756
```

Therefore 37% of this group are male and 63% are female

Finding out how many has obese, genderwise

```
FORTYONE_TO_FIFTY_obese=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['BMI']>=30]
a=percentage(FORTYONE_TO_FIFTY_obese.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
50.0
```

```
FORTYONE_TO_FIFTY_male_obese = FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_male_obese.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
42.29390681003584
```

```
FORTYONE_TO_FIFTY_female_obese= FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_female_obese.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
57.70609318996416
```

Therefore, 50% of this group has obesity and out of them 42% are male and 58 % are female

Now finding out obese smoke, genderwise

```
FORTYONE_TO_FIFTY_obese_smokes=FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['SmokingStatus']=='smokes']
a=percentage(FORTYONE_TO_FIFTY_obese_smokes.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
21.50537634408602
```

```
FORTYONE_TO_FIFTY_male_obese_smoke=FORTYONE_TO_FIFTY_obese_smokes[FORTYONE_TO_FIFTY_obese_smokes['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_male_obese_smoke.shape[0],FORTYONE_TO_FIFTY_obese_smokes.shape[0])
print(a)
```

```
41.66666666666667
```

```
FORTYONE_TO_FIFTY_female_obese_smoke=FORTYONE_TO_FIFTY_female_obese[FORTYONE_TO_FIFTY_female_obese['SmokingStatus']=='smokes']
a=percentage(FORTYONE_TO_FIFTY_female_obese_smoke.shape[0],FORTYONE_TO_FIFTY_obese_smokes.shape[0])
print(a)
```

```
58.333333333333336
```

Therefore,22% of total obese people smokes and out of them 42% are male and 58% are female

now finding out if they have Hypoglycemia(low glucose level)

```
FORTYONE_TO_FIFTY_low_glucose= FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['GlucoseLevel']<70]
a=percentage(FORTYONE_TO_FIFTY_low_glucose.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
14.874551971326163
```

```
FORTYONE_TO_FIFTY_male_low_glucose=FORTYONE_TO_FIFTY_low_glucose[FORTYONE_TO_FIFTY_low_glucose['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_male_low_glucose.shape[0],FORTYONE_TO_FIFTY_low_glucose.shape[0])
print(a)
```

```
40.963855421686745
```

```
FORTYONE_TO_FIFTY_female_low_glucose=FORTYONE_TO_FIFTY_low_glucose[FORTYONE_TO_FIFTY_low_glucose['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_female_low_glucose.shape[0],FORTYONE_TO_FIFTY_low_glucose.shape[0])
print(a)
```

```
59.036144578313255
```

Therefore, in this age group,15% has hypoglycemia and out of them, 41% male and about 59% female has Hypoglycemia

Finding those people who smokes and has tension, genderwise

```
FORTYONE_TO_FIFTY_smokes=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['SmokingStatus']=='smokes']
a=percentage(FORTYONE_TO_FIFTY_smokes.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
19.892473118279568
```

```
FORTYONE_TO_FIFTY_smokes_male=FORTYONE_TO_FIFTY_smokes[FORTYONE_TO_FIFTY_smokes['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_smokes_male.shape[0],FORTYONE_TO_FIFTY_smokes.shape[0])
print(a)
```

```
40.54054054054054
```

```
FORTYONE_TO_FIFTY_smokes_female=FORTYONE_TO_FIFTY_smokes[FORTYONE_TO_FIFTY_smokes['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_smokes_female.shape[0],FORTYONE_TO_FIFTY_smokes.shape[0])
print(a)
```

```
59.45945945945946
```

Therefore 20% of this group smokes and out of them 41% are male and 59% are female

```
FORTYONE_TO_FIFTY_smokes_tension=FORTYONE_TO_FIFTY_smokes[FORTYONE_TO_FIFTY_smokes['HasTension']=='Yes']
a=percentage(FORTYONE_TO_FIFTY_smokes_tension.shape[0],FORTYONE_TO_FIFTY_smokes.shape[0])
print(a)
```

```
3.6036036036036037
```

```
FORTYONE_TO_FIFTY_male_smokes_tension=FORTYONE_TO_FIFTY_smokes_tension[FORTYONE_TO_FIFTY_smokes_tension['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_male_smokes_tension.shape[0],FORTYONE_TO_FIFTY_smokes_tension.shape[0])
print(a)
```

```
75.0
```

```
FORTYONE_TO_FIFTY_female_smokes_tension= FORTYONE_TO_FIFTY_smokes_tension[FORTYONE_TO_FIFTY_smokes_tension['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_female_smokes_tension.shape[0],FORTYONE_TO_FIFTY_smokes_tension.shape[0])
print(a)
```

```
25.0
```

Therefore, 4% smokers got tension and out of them 75% are male and 25% are female

Now, calculating what percentage of govt employees, private employees,self-employed and unemployed persons are there

```
FORTYONE_TO_FIFTY_govt=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Occupation']=='Govt_job']
a=percentage(FORTYONE_TO_FIFTY_govt.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
19.892473118279568
```

```
FORTYONE_TO_FIFTY_govt_male=FORTYONE_TO_FIFTY_govt[FORTYONE_TO_FIFTY_govt['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_govt_male.shape[0],FORTYONE_TO_FIFTY_govt.shape[0])
print(a)
```

```
36.93693693693694
```

```
FORTYONE_TO_FIFTY_govt_female=FORTYONE_TO_FIFTY_govt[FORTYONE_TO_FIFTY_govt['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_govt_female.shape[0],FORTYONE_TO_FIFTY_govt.shape[0])
print(a)
```

```
63.06306306306306
```

20% works in govt sector and out of them 37% are male and 63% are female

```
FORTYONE_TO_FIFTY_private=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Occupation']=='Private']
a=percentage(FORTYONE_TO_FIFTY_private.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
65.05376344086021
```

```
FORTYONE_TO_FIFTY_private_male=FORTYONE_TO_FIFTY_private[FORTYONE_TO_FIFTY_private['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_private_male.shape[0],FORTYONE_TO_FIFTY_private.shape[0])
print(a)
```

```
37.74104683195592
```

```
FORTYONE_TO_FIFTY_private_female=FORTYONE_TO_FIFTY_private[FORTYONE_TO_FIFTY_private['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_private_female.shape[0],FORTYONE_TO_FIFTY_private.shape[0])
print(a)
```

```
62.25895316804407
```

65% works in private sector and out of them 38% are male and 62% are female

```
FORTYONE_TO_FIFTY_self=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Occupation']=='Self-employed']
a=percentage(FORTYONE_TO_FIFTY_self.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
15.053763440860216
```

```
FORTYONE_TO_FIFTY_self_male=FORTYONE_TO_FIFTY_self[FORTYONE_TO_FIFTY_self['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_self_male.shape[0],FORTYONE_TO_FIFTY_self.shape[0])
print(a)
```

```
35.714285714285715
```

```
FORTYONE_TO_FIFTY_self_female=FORTYONE_TO_FIFTY_self[FORTYONE_TO_FIFTY_self['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_self_female.shape[0],FORTYONE_TO_FIFTY_self.shape[0])
a
```

```
64.28571428571429
```

15% are self employed and out of them 36% are male and 64% are female

```
FORTYONE_TO_FIFTY_correct_df_final['Occupation'].value_counts()
```

Occupation	count
Private	363
Govt_job	111
Self-employed	84

dtype: int64

```
FORTYONE_TO_FIFTY_none=FORTYONE_TO_FIFTY_correct_df_final[FORTYONE_TO_FIFTY_correct_df_final['Occupation']=='Never worked']
a=percentage(FORTYONE_TO_FIFTY_none.shape[0],FORTYONE_TO_FIFTY_correct_df_final.shape[0])
print(a)
```

```
0.0
```

Therefore, none are unemployed

Now finding what percentage of them has tension

```
FORTYONE_TO_FIFTY_govt_tension=FORTYONE_TO_FIFTY_govt[FORTYONE_TO_FIFTY_govt['HasTension']=='Yes']
a= percentage(FORTYONE_TO_FIFTY_govt_tension.shape[0],FORTYONE_TO_FIFTY_govt.shape[0])
print(a)
```

```
9.00900900900901
```

```
FORTYONE_TO_FIFTY_govt_tension_male=FORTYONE_TO_FIFTY_govt_tension[FORTYONE_TO_FIFTY_govt_tension['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_govt_tension_male.shape[0],FORTYONE_TO_FIFTY_govt_tension.shape[0])
print(a)
```

```
30.0
```

```
FORTYONE_TO_FIFTY_govt_tension_female=FORTYONE_TO_FIFTY_govt_tension[FORTYONE_TO_FIFTY_govt_tension['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_govt_tension_female.shape[0],FORTYONE_TO_FIFTY_govt_tension.shape[0])
print(a)
```

```
70.0
```

9% govt employee has tension and out of them 30% are male and 70% are female

```
FORTYONE_TO_FIFTY_private_tension=FORTYONE_TO_FIFTY_private[FORTYONE_TO_FIFTY_private['HasTension']=='Yes']
a=percentage(FORTYONE_TO_FIFTY_private_tension.shape[0],FORTYONE_TO_FIFTY_private.shape[0])
print(a)
```

```
3.8567493112947657
```

```
FORTYONE_TO_FIFTY_private_tension_male=FORTYONE_TO_FIFTY_private_tension[FORTYONE_TO_FIFTY_private_tension['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_private_tension_male.shape[0],FORTYONE_TO_FIFTY_private_tension.shape[0])
print(a)
```

```
57.14285714285714
```

```
FORTYONE_TO_FIFTY_private_tension_female=FORTYONE_TO_FIFTY_private_tension[FORTYONE_TO_FIFTY_private_tension['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_private_tension_female.shape[0],FORTYONE_TO_FIFTY_private_tension.shape[0])
print(a)
```

```
42.857142857142854
```

4% private sector workers has tension and out of them 57% are male and 43% are female

```
FORTYONE_TO_FIFTY_self_tension=FORTYONE_TO_FIFTY_self[FORTYONE_TO_FIFTY_self['HasTension']=='Yes']
a=percentage(FORTYONE_TO_FIFTY_self_tension.shape[0],FORTYONE_TO_FIFTY_self.shape[0])
print(a)
```

```
14.285714285714285
```

```
FORTYONE_TO_FIFTY_self_tension_male=FORTYONE_TO_FIFTY_self_tension[FORTYONE_TO_FIFTY_self_tension['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_self_tension_male.shape[0],FORTYONE_TO_FIFTY_self_tension.shape[0])
print(a)
```

```
83.33333333333334
```

```
FORTYONE_TO_FIFTY_self_tension_female=FORTYONE_TO_FIFTY_self_tension[FORTYONE_TO_FIFTY_self_tension['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_self_tension_female.shape[0],FORTYONE_TO_FIFTY_self_tension.shape[0])
print(a)
```

```
16.666666666666664
```

Therefore

14% self employed people has tension and out of them 83% are male and 17% are female

What percentage of obese people had heartattack genderwise

```
FORTYONE_TO_FIFTY_obese_heartattack=FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['HeartAttack']=='Yes']
a=percentage(FORTYONE_TO_FIFTY_obese_heartattack.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
3.225806451612903
```

```
FORTYONE_TO_FIFTY_obese_heartattack_male=FORTYONE_TO_FIFTY_obese_heartattack[FORTYONE_TO_FIFTY_obese_heartattack['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_obese_heartattack_male.shape[0],FORTYONE_TO_FIFTY_obese_heartattack.shape[0])
print(a)
```

```
33.33333333333333
```

```
FORTYONE_TO_FIFTY_obese_heartattack_female=FORTYONE_TO_FIFTY_obese_heartattack[FORTYONE_TO_FIFTY_obese_heartattack['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_obese_heartattack_female.shape[0],FORTYONE_TO_FIFTY_obese_heartattack.shape[0])
print(a)
```

```
66.66666666666666
```

Therefore about 3% obese people have had and heartattack and out of them 33% are male and 67% are female

Now finding out how many obese are from where

```
FORTYONE_TO_FIFTY_obese_city=FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['LivesIn']=='City']
a=percentage(FORTYONE_TO_FIFTY_obese_city.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
50.89605734767025
```

```
FORTYONE_TO_FIFTY_obese_city_male=FORTYONE_TO_FIFTY_obese_city[FORTYONE_TO_FIFTY_obese_city['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_obese_city_male.shape[0],FORTYONE_TO_FIFTY_obese_city.shape[0])
```

```
print(a)
```

```
45.07042253521127
```

```
FORTYONE_TO_FIFTY_obese_city_female=FORTYONE_TO_FIFTY_obese_city[FORTYONE_TO_FIFTY_obese_city['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_obese_city_female.shape[0],FORTYONE_TO_FIFTY_obese_city.shape[0])
print(a)
```

```
54.929577464788736
```

Therefore, 51% obese are from city and out of them 45% are male and 55% are female

```
FORTYONE_TO_FIFTY_obese_village=FORTYONE_TO_FIFTY_obese[FORTYONE_TO_FIFTY_obese['LivesIn']=='Village']
a=percentage(FORTYONE_TO_FIFTY_obese_village.shape[0],FORTYONE_TO_FIFTY_obese.shape[0])
print(a)
```

```
49.10394265232975
```

```
FORTYONE_TO_FIFTY_obese_village_male=FORTYONE_TO_FIFTY_obese_village[FORTYONE_TO_FIFTY_obese_village['Gender']=='Male']
a=percentage(FORTYONE_TO_FIFTY_obese_village_male.shape[0],FORTYONE_TO_FIFTY_obese_village.shape[0])
print(a)
```

```
39.416058394160586
```

```
FORTYONE_TO_FIFTY_obese_village_female=FORTYONE_TO_FIFTY_obese_village[FORTYONE_TO_FIFTY_obese_village['Gender']=='Female']
a=percentage(FORTYONE_TO_FIFTY_obese_village_female.shape[0],FORTYONE_TO_FIFTY_obese_village.shape[0])
print(a)
```

```
60.58394160583942
```

Therefore 49% obese are from village and out of them 39% are male and 61% are female

Now, in this cluster, none are unemployed. So finding out what percentage lives in city/village is absurd

and finding out how many unemployed has tension is also absurd

Now clustering data of 51 to 60 year age range

```
FIFTYONE_TO_SIXTY_df=correct_df_final[(correct_df_final['Age']>=51) & (correct_df_final['Age']<=60)]
FIFTYONE_TO_SIXTY_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
9	Female	52	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
13	Female	57	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
28	Female	51	No	No	No	Private	Village	66.7	29.0	never smoked	No
32	Female	53	No	No	Yes	Self-employed	City	84.0	25.3	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3977	Male	54	No	No	Yes	Private	City	66.6	22.7	never smoked	No
3978	Male	52	No	No	Yes	Govt_job	City	207.7	35.6	formerly smoked	No
3987	Female	53	No	No	Yes	Private	Village	101.7	33.2	never smoked	No
3991	Female	55	No	No	Yes	Private	City	83.4	22.8	Unknown	No
3999	Female	58	No	No	Yes	Private	Village	59.2	27.9	formerly smoked	Yes

611 rows × 11 columns

Next steps: Generate code with FIFTYONE_TO_SIXTY_df | View recommended plots | New interactive sheet

```
FIFTYONE_TO_SIXTY_corrected=FIFTYONE_TO_SIXTY_df
FIFTYONE_TO_SIXTY_corrected
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
9	Female	52	No	No	Yes	Private	Village	67.5	15.8	Unknown	No
13	Female	57	Yes	No	Yes	Self-employed	Village	63.5	36.6	smokes	No
14	Male	56	No	No	Yes	Private	Village	92.8	31.5	never smoked	No
28	Female	51	No	No	No	Private	Village	66.7	29.0	never smoked	No
32	Female	53	No	No	Yes	Self-employed	City	84.0	25.3	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3977	Male	54	No	No	Yes	Private	City	66.6	22.7	never smoked	No
3978	Male	52	No	No	Yes	Govt_job	City	207.7	35.6	formerly smoked	No
3987	Female	53	No	No	Yes	Private	Village	101.7	33.2	never smoked	No
3991	Female	55	No	No	Yes	Private	City	83.4	22.8	Unknown	No
3999	Female	58	No	No	Yes	Private	Village	59.2	27.9	formerly smoked	Yes

611 rows × 11 columns

Next steps: Generate code with FIFTYONE_TO_SIXTY_df | View recommended plots | New interactive sheet

```
FIFTYONE_TO_SIXTY_male=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_male.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
45.335515548281506
```

```
FIFTYONE_TO_SIXTY_female=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_female.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
54.6644844517185
```

Therefore, in this specific age group, there exists 611 data and out of them about 45% are male and 55% are female

Now finding what percentage works in which sector.

```
FIFTYONE_TO_SIXTY_govt=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Occupation']=='Govt_job']
a=percentage(FIFTYONE_TO_SIXTY_govt.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
18.494271685761046
```

```
FIFTYONE_TO_SIXTY_govt_male=FIFTYONE_TO_SIXTY_govt[FIFTYONE_TO_SIXTY_govt['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_govt_male.shape[0],FIFTYONE_TO_SIXTY_govt.shape[0])
print(a)
```

```
46.902654867256636
```

```
FIFTYONE_TO_SIXTY_govt_female=FIFTYONE_TO_SIXTY_govt[FIFTYONE_TO_SIXTY_govt['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_govt_female.shape[0],FIFTYONE_TO_SIXTY_govt.shape[0])
print(a)
```

```
53.09734513274337
```

Therefore, 18% works in govt sector and out of them, 47% are male and 53% are female

```
FIFTYONE_TO_SIXTY_private=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Occupation']=='Private']
a=percentage(FIFTYONE_TO_SIXTY_private.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
64.97545008183306
```

```
FIFTYONE_TO_SIXTY_private_male=FIFTYONE_TO_SIXTY_private[FIFTYONE_TO_SIXTY_private['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_private_male.shape[0],FIFTYONE_TO_SIXTY_private.shape[0])
print(a)
```

```
46.85138539042821
```

```
FIFTYONE_TO_SIXTY_private_female=FIFTYONE_TO_SIXTY_private[FIFTYONE_TO_SIXTY_private['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_private_female.shape[0],FIFTYONE_TO_SIXTY_private.shape[0])
print(a)
```

```
53.14861460957179
```

Therefore, 65% works in private sector and out of them 47% are male and 53% are female

```
FIFTYONE_TO_SIXTY_self=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Occupation']=='Self-employed']
a=percentage(FIFTYONE_TO_SIXTY_self.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
16.53027823240589
```

```
FIFTYONE_TO_SIXTY_self_male=FIFTYONE_TO_SIXTY_self[FIFTYONE_TO_SIXTY_self['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_self_male.shape[0],FIFTYONE_TO_SIXTY_self.shape[0])
print(a)
```

```
37.62376237623762
```

```
FIFTYONE_TO_SIXTY_self_female=FIFTYONE_TO_SIXTY_self[FIFTYONE_TO_SIXTY_self['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_self_female.shape[0],FIFTYONE_TO_SIXTY_self.shape[0])
print(a)
```

```
62.37623762376238
```

17% are self employed and out of them 38% are male and 62% are female

```
FIFTYONE_TO_SIXTY_none=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['Occupation']=='Never_worked']
a=percentage(FIFTYONE_TO_SIXTY_none.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
0.0
```

So,

and none of them are unemployed or retired yet.

Now, finding what percentage of people working in these sectors have tension.

```
FIFTYONE_TO_SIXTY_govt_tension=FIFTYONE_TO_SIXTY_govt[FIFTYONE_TO_SIXTY_govt['HasTension']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_govt_tension.shape[0],FIFTYONE_TO_SIXTY_govt.shape[0])
print(a)
```

```
13.274336283185843
```

```
FIFTYONE_TO_SIXTY_govt_tension_male=FIFTYONE_TO_SIXTY_govt_tension[FIFTYONE_TO_SIXTY_govt_tension['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_govt_tension_male.shape[0],FIFTYONE_TO_SIXTY_govt_tension.shape[0])
print(a)
```

```
53.333333333333336
```

```
FIFTYONE_TO_SIXTY_govt_tension_female=FIFTYONE_TO_SIXTY_govt_tension[FIFTYONE_TO_SIXTY_govt_tension['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_govt_tension_female.shape[0],FIFTYONE_TO_SIXTY_govt_tension.shape[0])
print(a)
```

```
46.666666666666664
```

Therefore, 13% govt workers has tension and out of them 53% are male and 47% are female

```
FIFTYONE_TO_SIXTY_private_tension=FIFTYONE_TO_SIXTY_private[FIFTYONE_TO_SIXTY_private['HasTension']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_private_tension.shape[0],FIFTYONE_TO_SIXTY_private.shape[0])
print(a)
```

```
16.120906801007557
```

```
FIFTYONE_TO_SIXTY_private_tension_male=FIFTYONE_TO_SIXTY_private_tension[FIFTYONE_TO_SIXTY_private_tension['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_private_tension_male.shape[0],FIFTYONE_TO_SIXTY_private_tension.shape[0])
print(a)
```

```
39.0625
```

```
FIFTYONE_TO_SIXTY_private_tension_female=FIFTYONE_TO_SIXTY_private_tension[FIFTYONE_TO_SIXTY_private_tension['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_private_tension_female.shape[0],FIFTYONE_TO_SIXTY_private_tension.shape[0])
print(a)
```

```
60.9375
```

Therefore, 16% private employee has tension and out of them 39% are male and 61% are female

```
FIFTYONE_TO_SIXTY_self_tension=FIFTYONE_TO_SIXTY_self[FIFTYONE_TO_SIXTY_self['HasTension']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_self_tension.shape[0],FIFTYONE_TO_SIXTY_self.shape[0])
print(a)
```

```
10.891089108910892
```

```
FIFTYONE_TO_SIXTY_self_tension_male=FIFTYONE_TO_SIXTY_self_tension[FIFTYONE_TO_SIXTY_self_tension['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_self_tension_male.shape[0],FIFTYONE_TO_SIXTY_self_tension.shape[0])
print(a)
```

```
54.54545454545454
```

```
FIFTYONE_TO_SIXTY_self_tension_female=FIFTYONE_TO_SIXTY_self_tension[FIFTYONE_TO_SIXTY_self_tension['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_self_tension_female.shape[0],FIFTYONE_TO_SIXTY_self_tension.shape[0])
```

```
print(a)
```

```
45.45454545454545
```

Therefore, 11% self employeed person has tension and 55% of them are male and 45% are female

Now finding what percentage of people has obesity genderwise

```
FIFTYONE_TO_SIXTY_obese=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['BMI']>=30]
a=percentage(FIFTYONE_TO_SIXTY_obese.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
55.810147299509005
```

```
FIFTYONE_TO_SIXTY_obese_male=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_male.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
46.9208211143695
```

```
FIFTYONE_TO_SIXTY_obese_female=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_female.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
53.0791788856305
```

Therefore, 56% people has obesity and out of them 47% are male and 53% are female

Now finding out what percentage of obese people smokes genderwise

```
FIFTYONE_TO_SIXTY_obese_smokes=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['SmokingStatus']=='smokes']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
17.595307917888565
```

```
FIFTYONE_TO_SIXTY_obese_smokes_male=FIFTYONE_TO_SIXTY_obese_smokes[FIFTYONE_TO_SIXTY_obese_smokes['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes_male.shape[0],FIFTYONE_TO_SIXTY_obese_smokes.shape[0])
print(a)
```

```
48.333333333333336
```

```
FIFTYONE_TO_SIXTY_obese_smokes_female=FIFTYONE_TO_SIXTY_obese_smokes[FIFTYONE_TO_SIXTY_obese_smokes['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes_female.shape[0],FIFTYONE_TO_SIXTY_obese_smokes.shape[0])
print(a)
```

```
51.66666666666667
```

Therefore, 18% obese people smokes and out of them 48% are male and 52% are female

Now, finding what percentage of obese people had heartattack genderwise

```
FIFTYONE_TO_SIXTY_obese_heart_attack=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['HeartAttack']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_obese_heart_attack.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
6.451612903225806
```

```
FIFTYONE_TO_SIXTY_obese_heart_attack_male=FIFTYONE_TO_SIXTY_obese_heart_attack[FIFTYONE_TO_SIXTY_obese_heart_attack['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_heart_attack_male.shape[0],FIFTYONE_TO_SIXTY_obese_heart_attack.shape[0])
print(a)
```

```
50.0
```

```
FIFTYONE_TO_SIXTY_obese_heart_attack_female=FIFTYONE_TO_SIXTY_obese_heart_attack[FIFTYONE_TO_SIXTY_obese_heart_attack['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_heart_attack_female.shape[0],FIFTYONE_TO_SIXTY_obese_heart_attack.shape[0])
print(a)
```

```
50.0
```

So, 6% obese people had heartattack and out of them 50% are male and 50% are female

Now finding what percentage of people has low glucose genderwise

```
FIFTYONE_TO_SIXTY_low_glucose=FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['GlucoseLevel']<70]
a=percentage(FIFTYONE_TO_SIXTY_low_glucose.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
13.584288052373159
```

```
FIFTYONE_TO_SIXTY_male_low_glucose= FIFTYONE_TO_SIXTY_low_glucose[FIFTYONE_TO_SIXTY_low_glucose['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_male_low_glucose.shape[0],FIFTYONE_TO_SIXTY_low_glucose.shape[0])
print(a)
```

```
36.144578313253014
```

```
FIFTYONE_TO_SIXTY_female_low_glucose=FIFTYONE_TO_SIXTY_low_glucose[FIFTYONE_TO_SIXTY_low_glucose['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_female_low_glucose.shape[0],FIFTYONE_TO_SIXTY_low_glucose.shape[0])
print(a)
```

```
63.85542168674698
```

Therefore, 14% people has low low glucose level and out of them 36% are male and 64% are female

now finding what percentage of the obese people smokes and has tension and genderwise

```
FIFTYONE_TO_SIXTY_obese_smokes_tension=FIFTYONE_TO_SIXTY_obese_smokes[FIFTYONE_TO_SIXTY_obese_smokes['HasTension']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes_tension.shape[0],FIFTYONE_TO_SIXTY_obese_smokes.shape[0])
print(a)
```

```
26.666666666666668
```

```
FIFTYONE_TO_SIXTY_obese_smokes_tension_male=FIFTYONE_TO_SIXTY_obese_smokes_tension[FIFTYONE_TO_SIXTY_obese_smokes_tension['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes_tension_male.shape[0],FIFTYONE_TO_SIXTY_obese_smokes_tension.shape[0])
print(a)
```

```
50.0
```

```
FIFTYONE_TO_SIXTY_obese_smokes_tension_female=FIFTYONE_TO_SIXTY_obese_smokes_tension[FIFTYONE_TO_SIXTY_obese_smokes_tension['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_smokes_tension_female.shape[0],FIFTYONE_TO_SIXTY_obese_smokes_tension.shape[0])
print(a)
```

```
50.0
```

Therfore 27% obese smokers has tension and out of them 50% are male and 50% are female

Finding out how many smokes

```
FIFTYONE_TO_SIXTY_smokes= FIFTYONE_TO_SIXTY_corrected[FIFTYONE_TO_SIXTY_corrected['SmokingStatus']=='smokes']
a=percentage(FIFTYONE_TO_SIXTY_smokes.shape[0],FIFTYONE_TO_SIXTY_corrected.shape[0])
print(a)
```

```
18.00327332242226
```

```
FIFTYONE_TO_SIXTY_smokes_male=FIFTYONE_TO_SIXTY_smokes[FIFTYONE_TO_SIXTY_smokes['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_smokes_male.shape[0],FIFTYONE_TO_SIXTY_smokes.shape[0])
print(a)
```

```
49.09090909090909
```

```
FIFTYONE_TO_SIXTY_smokes_female=FIFTYONE_TO_SIXTY_smokes[FIFTYONE_TO_SIXTY_smokes['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_smokes_female.shape[0],FIFTYONE_TO_SIXTY_smokes.shape[0])
print(a)
```

```
50.90909090909091
```

Therefore, 18% of this clustered dataframe smokes and out of them 49% are male and 51% are female

finding out how many smokers has tension

```
FIFTYONE_TO_SIXTY_smokes_tension=FIFTYONE_TO_SIXTY_smokes[FIFTYONE_TO_SIXTY_smokes['HasTension']=='Yes']
a=percentage(FIFTYONE_TO_SIXTY_smokes_tension.shape[0],FIFTYONE_TO_SIXTY_smokes.shape[0])
print(a)
```

```
24.545454545454547
```

```
FIFTYONE_TO_SIXTY_smokes_tension_male=FIFTYONE_TO_SIXTY_smokes_tension[FIFTYONE_TO_SIXTY_smokes_tension['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_smokes_tension_male.shape[0],FIFTYONE_TO_SIXTY_smokes_tension.shape[0])
print(a)
```

```
51.85185185185185
```

```
FIFTYONE_TO_SIXTY_smokes_tension_female=FIFTYONE_TO_SIXTY_smokes_tension[FIFTYONE_TO_SIXTY_smokes_tension['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_smokes_tension_female.shape[0],FIFTYONE_TO_SIXTY_smokes_tension.shape[0])
print(a)
```

```
48.148148148148145
```

Therefore, 26% smokers has tension and out of them 52% are male and 48% are female

finding out how many obese lives in village/city

```
FIFTYONE_TO_SIXTY_obese_city=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['LivesIn']=='City']
a=percentage(FIFTYONE_TO_SIXTY_obese_city.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
51.90615835777126
```

```
FIFTYONE_TO_SIXTY_obese_city_male= FIFTYONE_TO_SIXTY_obese_city[FIFTYONE_TO_SIXTY_obese_city['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_city_male.shape[0],FIFTYONE_TO_SIXTY_obese_city.shape[0])
print(a)
```

```
45.19774011299435
```

```
FIFTYONE_TO_SIXTY_obese_city_female=FIFTYONE_TO_SIXTY_obese_city[FIFTYONE_TO_SIXTY_obese_city['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_city_female.shape[0],FIFTYONE_TO_SIXTY_obese_city.shape[0])
print(a)
```

```
54.80225988700565
```

Therefore, 52% obese lives in city and out of them 45% are male and 55% are female

```
FIFTYONE_TO_SIXTY_obese_village=FIFTYONE_TO_SIXTY_obese[FIFTYONE_TO_SIXTY_obese['LivesIn']=='Village']
a=percentage(FIFTYONE_TO_SIXTY_obese_village.shape[0],FIFTYONE_TO_SIXTY_obese.shape[0])
print(a)
```

```
48.09384164222874
```

```
FIFTYONE_TO_SIXTY_obese_village_male=FIFTYONE_TO_SIXTY_obese_village[FIFTYONE_TO_SIXTY_obese_village['Gender']=='Male']
a=percentage(FIFTYONE_TO_SIXTY_obese_village_male.shape[0],FIFTYONE_TO_SIXTY_obese_village.shape[0])
print(a)
```

```
48.78048780487805
```

```
FIFTYONE_TO_SIXTY_obese_village_female=FIFTYONE_TO_SIXTY_obese_village[FIFTYONE_TO_SIXTY_obese_village['Gender']=='Female']
a=percentage(FIFTYONE_TO_SIXTY_obese_village_female.shape[0],FIFTYONE_TO_SIXTY_obese_village.shape[0])
print(a)
```

```
51.21951219512195
```

Therefore, 48% obese are from village and out of them 49% are male and 51% are female

```
FIFTYONE_TO_SIXTY_corrected['Occupation'].unique()
```

```
array(['Private', 'Self-employed', 'Govt_job'], dtype=object)
```

Therefore, having 0 unemployed, It will be absurd to find out how many of unemployed are from city/village

also, it will be absurd to find out how many unemployed has tension

Now clustering data for sixtyone to seventy age group

```
sixtyone_to_seventy_df=correct_df_final[(correct_df_final['Age']>=61) & (correct_df_final['Age']<=70)]
sixtyone_to_seventy_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
6	Male	63	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
22	Male	63	No	No	Yes	Self-employed	Village	105.0	18.8	formerly smoked	No
42	Female	65	No	No	No	Self-employed	Village	87.2	32.3	never smoked	No
48	Male	63	No	No	Yes	Private	Village	79.6	29.9	Unknown	No
63	Female	65	No	No	Yes	Private	City	104.0	26.5	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3937	Male	63	No	No	Yes	Private	City	90.1	35.9	Unknown	No
3938	Male	68	No	Yes	Yes	Private	City	218.9	24.8	Unknown	No
3949	Male	70	No	No	Yes	Self-employed	Village	193.0	24.5	Unknown	No
3963	Female	70	Yes	No	Yes	Self-employed	City	99.5	33.4	formerly smoked	No
3980	Female	63	No	Yes	Yes	Private	Village	85.8	40.1	never smoked	No

435 rows × 11 columns

Next steps: Generate code with sixtyone_to_seventy_df View recommended plots New interactive sheet

Removing those data who are marked as govt employee because after 60 they will be retired

```
sixtyone_to_seventy_govt=sixtyone_to_seventy_df[sixtyone_to_seventy_df['Occupation']=='Govt_job']
sixtyone_to_seventy_govt
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
118	Female	64	No	No	Yes	Govt_job	City	80.2	38.9	formerly smoked	No
142	Male	64	No	No	Yes	Govt_job	Village	114.1	23.8	never smoked	No
207	Female	69	No	Yes	No	Govt_job	City	202.9	34.2	Unknown	No
300	Female	63	No	No	Yes	Govt_job	City	123.9	35.9	Unknown	No
397	Male	61	Yes	Yes	Yes	Govt_job	Village	86.9	34.1	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3808	Female	62	Yes	No	Yes	Govt_job	Village	75.8	34.5	never smoked	No
3892	Female	63	Yes	Yes	No	Govt_job	City	175.6	23.2	never smoked	No
3925	Male	67	No	No	Yes	Govt_job	City	66.0	35.3	smokes	No
3930	Male	62	No	No	No	Govt_job	Village	76.9	32.3	never smoked	No
3931	Female	69	Yes	No	Yes	Govt_job	City	206.5	32.7	formerly smoked	No

64 rows × 11 columns

Next steps: Generate code with sixtyone_to_seventy_govt View recommended plots New interactive sheet

```
sixtyone_to_seventy_corrected= sixtyone_to_seventy_df.drop(sixtyone_to_seventy_govt.index)
sixtyone_to_seventy_corrected
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
6	Male	63	No	No	Yes	Private	City	104.1	37.1	formerly smoked	No
22	Male	63	No	No	Yes	Self-employed	Village	105.0	18.8	formerly smoked	No
42	Female	65	No	No	No	Self-employed	Village	87.2	32.3	never smoked	No
48	Male	63	No	No	Yes	Private	Village	79.6	29.9	Unknown	No
63	Female	65	No	No	Yes	Private	City	104.0	26.5	never smoked	No
...	...	...	...	...	...	...	...	...	...	...	...
3937	Male	63	No	No	Yes	Private	City	90.1	35.9	Unknown	No
3938	Male	68	No	Yes	Yes	Private	City	218.9	24.8	Unknown	No
3949	Male	70	No	No	Yes	Self-employed	Village	193.0	24.5	Unknown	No
3963	Female	70	Yes	No	Yes	Self-employed	City	99.5	33.4	formerly smoked	No
3980	Female	63	No	Yes	Yes	Private	Village	85.8	40.1	never smoked	No

371 rows × 11 columns

Next steps: Generate code with sixtyone_to_seventy_corrected | View recommended plots | New interactive sheet

Now finding out what percentage of them are male and female

```
sixtyone_to_seventy__male=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['Gender']=='Male']
a=percentage(sixtyone_to_seventy__male.shape[0],sixtyone_to_seventy_corrected.shape[0])
print(a)
```

```
42.04851752021563
```

```
sixtyone_to_seventy__female=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['Gender']=='Female']
a=percentage(sixtyone_to_seventy__female.shape[0],sixtyone_to_seventy_corrected.shape[0])
print (a)
```

```
57.95148247978437
```

Therefore, in this group 42% are male and 58% are female

Now finding how many of them has obesity genderwise

```
sixtyone_to_seventy_obese=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['BMI']>=30]
a=percentage(sixtyone_to_seventy_obese.shape[0],sixtyone_to_seventy_corrected.shape[0])
print(a)
```

```
49.32614555256065
```

```
sixtyone_to_seventy_obese_male=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['Gender']=='Male']
a=percentage(sixtyone_to_seventy_obese_male.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
45.3551912568306
```

```
sixtyone_to_seventy_obese_female=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['Gender']=='Female']
a=percentage(sixtyone_to_seventy_obese_female.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
54.644808743169406
```

Therefore, 49% has obese and out of them 45% are male and 55% are female

Now finding how many obese smokes genderwise

```
sixtyone_to_seventy_obese_smokes=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['SmokingStatus']=='Smokes']
a=percentage(sixtyone_to_seventy_obese_smokes.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
0.0
```

Therefore, no obese smokes.

now finding out if they ever had heartaatack

```
sixtyone_to_seventy_obese_heartattack=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['HeartAttack']=='Yes']
a=percentage(sixtyone_to_seventy_obese_heartattack.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
8.743169398907105
```

```
sixtyone_to_seventy_obese_heartattack_male=sixtyone_to_seventy_obese_heartattack[sixtyone_to_seventy_obese_heartattack['Gender']=='Male']
a=percentage(sixtyone_to_seventy_obese_heartattack_male.shape[0],sixtyone_to_seventy_obese_heartattack.shape[0])
print(a)
```

```
62.5
```

```
sixtyone_to_seventy_obese_heartattack_female=sixtyone_to_seventy_obese_heartattack[sixtyone_to_seventy_obese_heartattack['Gender']=='Female']
a=percentage(sixtyone_to_seventy_obese_heartattack_female.shape[0],sixtyone_to_seventy_obese_heartattack.shape[0])
print(a)
```

```
37.5
```

Therefore 9% obese people had heartattack and out of them 62.5% are male and 37.5% are female

Now finding how many has low glucose genderwise (hypoglycemia)

```
sixtyone_to_seventy_low_glucose=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['GlucoseLevel']<70]
a=percentage(sixtyone_to_seventy_low_glucose.shape[0],sixtyone_to_seventy_corrected.shape[0])
print(a)
```

```
14.016172506738545
```

```
sixtyone_to_seventy_low_glucose_male=sixtyone_to_seventy_low_glucose[sixtyone_to_seventy_low_glucose['Gender']=='Male']
a=percentage(sixtyone_to_seventy_low_glucose_male.shape[0],sixtyone_to_seventy_low_glucose.shape[0])
print(a)
```

```
50.0
```

```
sixtyone_to_seventy_low_glucose_female=sixtyone_to_seventy_low_glucose[sixtyone_to_seventy_low_glucose['Gender']=='Female']
a=percentage(sixtyone_to_seventy_low_glucose_female.shape[0],sixtyone_to_seventy_low_glucose.shape[0])
print(a)
```

```
50.0
```

Therefore, 14% people has low glucose and out of them 50% are male and 50% are female

Now showing how many of them still works genderwise

```
sixtyone_to_seventy_self=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['Occupation']=='Self-employed']
a=percentage(sixtyone_to_seventy_self.shape[0],sixtyone_to_seventy_corrected.shape[0])
print(a)
```

```
31.536388140161726
```

```
sixtyone_to_seventy_self_male=sixtyone_to_seventy_self[sixtyone_to_seventy_self['Gender']=='Male']
a=percentage(sixtyone_to_seventy_self_male.shape[0],sixtyone_to_seventy_self.shape[0])
print(a)
```

```
34.18803418803419
```

```
sixtyone_to_seventy_self_female=sixtyone_to_seventy_self[sixtyone_to_seventy_self['Gender']=='Female']
a=percentage(sixtyone_to_seventy_self_female.shape[0],sixtyone_to_seventy_self.shape[0])
print(a)
```

```
65.8119658119658
```

Therefore, 32% are self employed and out of them, 34% are male and 66% are female

```
sixtyone_to_seventy_private=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['Occupation']=='Private']
a=percentage(sixtyone_to_seventy_private.shape[0],sixtyone_to_seventy_corrected.shape[0])
a
```

```
68.46361185983828
```

```
sixtyone_to_seventy_private_male=sixtyone_to_seventy_private[sixtyone_to_seventy_private['Gender']=='Male']
a=percentage(sixtyone_to_seventy_private_male.shape[0],sixtyone_to_seventy_private.shape[0])
print(a)
```

```
45.66929133858268
```

```
sixtyone_to_seventy_private_female= sixtyone_to_seventy_private[sixtyone_to_seventy_private['Gender']=='Female']
a=percentage(sixtyone_to_seventy_private_female.shape[0],sixtyone_to_seventy_private.shape[0])
print(a)
```

```
54.330708661417326
```

Therefore, 68% works in private sector and out of them 46% are male and 54% are female

```
sixtyone_to_seventy_corrected['Occupation'].value_counts()
```

Occupation	count
Private	254
Self-employed	117

dtype: int64

Now finding how many of them has tension genderwise

```
sixtyone_to_seventy_self_tension=sixtyone_to_seventy_self[sixtyone_to_seventy_self['HasTension']=='Yes']
a=percentage(sixtyone_to_seventy_self_tension.shape[0],sixtyone_to_seventy_self.shape[0])
print(a)
```

```
20.51282051282051
```

```
sixtyone_to_seventy_self_tension_male=sixtyone_to_seventy_self_tension[sixtyone_to_seventy_self_tension['Gender']=='Male']
a=percentage(sixtyone_to_seventy_self_tension_male.shape[0],sixtyone_to_seventy_self_tension.shape[0])
print(a)
```

```
33.33333333333333
```

```
sixtyone_to_seventy_self_tension_female=sixtyone_to_seventy_self_tension[sixtyone_to_seventy_self_tension['Gender']=='Female']
a=percentage(sixtyone_to_seventy_self_tension_female.shape[0],sixtyone_to_seventy_self_tension.shape[0])
```

```
print(a)
```

```
66.66666666666666
```

Therefore, 21% self employed has tension and out of them 33% are male and 67% are female

```
sixtyone_to_seventy_private_tension=sixtyone_to_seventy_private[sixtyone_to_seventy_private['HasTension']=='Yes']
a=percentage(sixtyone_to_seventy_private_tension.shape[0],sixtyone_to_seventy_private.shape[0])
print(a)
```

```
18.503937007874015
```

```
sixtyone_to_seventy_private_tension_male=sixtyone_to_seventy_private_tension[sixtyone_to_seventy_private_tension['Gender']=='Male']
a=percentage(sixtyone_to_seventy_private_tension_male.shape[0],sixtyone_to_seventy_private_tension.shape[0])
print(a)
```

```
59.57446808510638
```

```
sixtyone_to_seventy_private_tension_female=sixtyone_to_seventy_private_tension[sixtyone_to_seventy_private_tension['Gender']=='Female']
a=percentage(sixtyone_to_seventy_private_tension_female.shape[0],sixtyone_to_seventy_private_tension.shape[0])
print(a)
```

```
40.42553191489361
```

Therefore, 19% of private employee has tension and out of them 60% are male and 40% are female

Since none are unemployed, it is absurd to find how many unemployed are from village/city

and it is also absurd to find how many unemloyed has tension

now finding out how many of them smokes

```
sixtyone_to_seventy_smokes=sixtyone_to_seventy_corrected[sixtyone_to_seventy_corrected['SmokingStatus']=='smokes']
a=percentage(sixtyone_to_seventy_smokes.shape[0],sixtyone_to_seventy_corrected.shape[0])
print(a)
```

```
17.25067385444744
```

```
sixtyone_to_seventy_smokes_male= sixtyone_to_seventy_smokes[sixtyone_to_seventy_smokes['Gender']=='Male']
a=percentage(sixtyone_to_seventy_smokes_male.shape[0],sixtyone_to_seventy_smokes.shape[0])
print(a)
```

```
51.5625
```

```
sixtyone_to_seventy_smokes_female=sixtyone_to_seventy_smokes[sixtyone_to_seventy_smokes['Gender']=='Female']
a=percentage(sixtyone_to_seventy_smokes_female.shape[0],sixtyone_to_seventy_smokes.shape[0])
print(a)
```

```
48.4375
```

Therefore, 17% of this clustered group are smokers and out of them 52% are male and 48% are female

now, finding out how many smokers has tension

```
sixtyone_to_seventy_smokes_tension=sixtyone_to_seventy_smokes[sixtyone_to_seventy_smokes['HasTension']=='Yes']
a=percentage(sixtyone_to_seventy_smokes_tension.shape[0],sixtyone_to_seventy_smokes.shape[0])
print(a)
```

```
20.3125
```

```
sixtyone_to_seventy_smokes_tension_male=sixtyone_to_seventy_smokes_tension[sixtyone_to_seventy_smokes_tension['Gender']=='Male']
a=percentage(sixtyone_to_seventy_smokes_tension_male.shape[0],sixtyone_to_seventy_smokes_tension.shape[0])
print(a)
```

```
69.23076923076923
```

```
sixtyone_to_seventy_smokes_tension_female=sixtyone_to_seventy_smokes_tension[sixtyone_to_seventy_smokes_tension['Gender']=='Female']
a=percentage(sixtyone_to_seventy_smokes_tension_female.shape[0],sixtyone_to_seventy_smokes_tension.shape[0])
print(a)
```

```
30.76923076923077
```

Therefore, 20% of smokers has tension and out of them 69% are male and 31% are female

now finding out how many obese are from city/village

```
sixtyone_to_seventy_obese_village=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['LivesIn']=='Village']
a=percentage(sixtyone_to_seventy_obese_village.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
53.00546448087432
```

```
sixtyone_to_seventy_obese_village_male=sixtyone_to_seventy_obese_village[sixtyone_to_seventy_obese_village['Gender']=='Male']
a=percentage(sixtyone_to_seventy_obese_village_male.shape[0],sixtyone_to_seventy_obese_village.shape[0])
print(a)
```

```
47.42268041237113
```

```
sixtyone_to_seventy_obese_village_female=sixtyone_to_seventy_obese_village[sixtyone_to_seventy_obese_village['Gender']=='Female']
a=percentage(sixtyone_to_seventy_obese_village_female.shape[0],sixtyone_to_seventy_obese_village.shape[0])
print(a)
```

```
52.57731958762887
```

Therefore, 53% obese are from village and out of them 47% are male and 53% are female

```
sixtyone_to_seventy_obese_city=sixtyone_to_seventy_obese[sixtyone_to_seventy_obese['LivesIn']=='City']
a=percentage(sixtyone_to_seventy_obese_city.shape[0],sixtyone_to_seventy_obese.shape[0])
print(a)
```

```
46.994535519125684
```

```
sixtyone_to_seventy_obese_city_male=sixtyone_to_seventy_obese_city[sixtyone_to_seventy_obese_city['Gender']=='Male']
a=percentage(sixtyone_to_seventy_obese_city_male.shape[0],sixtyone_to_seventy_obese_city.shape[0])
print(a)
```

```
43.02325581395349
```

```
sixtyone_to_seventy_obese_city_female=sixtyone_to_seventy_obese_city[sixtyone_to_seventy_obese_city['Gender']=='Female']
a=percentage(sixtyone_to_seventy_obese_city_female.shape[0],sixtyone_to_seventy_obese_city.shape[0])
print(a)
```

```
56.97674418604651
```

Therefore 47% obese are from city and out of them 43% are male and 57% are female

Now clustering data for age group 71 to last

```
correct_df_final['Age'].max()
```

```
82
```

Therefore clustering data for age group 71 to 82

```
seventyone_eightytwo_df=correct_df_final[(correct_df_final['Age']>=71) & (correct_df_final['Age']<=82)]
seventyone_eightytwo_df
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
3	Male	78	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
19	Female	82	No	No	Yes	Private	Village	95.8	27.1	Unknown	No
24	Female	75	No	No	Yes	Self-employed	Village	70.8	24.8	formerly smoked	No
39	Female	79	No	No	Yes	Private	Village	92.5	24.5	never smoked	Yes
...	...	...	...	...	...	...	...	...	...	...	...
3966	Male	76	No	No	Yes	Private	City	139.7	30.4	formerly smoked	Yes
3969	Female	81	No	No	Yes	Private	City	77.8	35.3	Unknown	No
3979	Male	78	No	No	Yes	Private	City	103.4	30.4	never smoked	No
3983	Male	77	No	No	Yes	Private	City	223.0	29.6	formerly smoked	No
3993	Male	71	No	No	Yes	Private	City	85.4	28.3	never smoked	No

512 rows × 11 columns

Next steps: Generate code with `seventyone_eightytwo_df` View recommended plots New interactive sheet

Now, filtering the dataset

```
seventyone_eightytwo_govt=seventyone_eightytwo_df[seventyone_eightytwo_df['Occupation']=='Govt_job']
seventyone_eightytwo_govt
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
50	Female	73	No	No	No	Govt_job	City	63.2	25.7	formerly smoked	No
59	Female	76	No	No	Yes	Govt_job	City	96.2	26.1	smokes	No
60	Male	71	No	No	Yes	Govt_job	City	57.8	23.6	Unknown	No
212	Female	78	No	No	Yes	Govt_job	City	59.8	35.4	Unknown	No
274	Female	82	Yes	Yes	Yes	Govt_job	City	216.2	29.5	formerly smoked	Yes
...	...	...	...	...	...	...	...	...	...	...	...
3730	Female	75	No	No	Yes	Govt_job	City	89.3	38.4	never smoked	No
3830	Male	77	No	No	Yes	Govt_job	City	64.3	28.3	never smoked	No
3903	Female	75	No	No	Yes	Govt_job	Village	57.5	24.9	never smoked	No
3956	Male	75	No	No	Yes	Govt_job	City	95.4	34.7	Unknown	No
3962	Female	79	No	No	Yes	Govt_job	City	77.5	36.3	Unknown	No

62 rows × 11 columns

Next steps: Generate code with seventyone_eightytwo_govt View recommended plots New interactive sheet

```
seventyone_eightytwo_df_corrected= seventyone_eightytwo_df.drop(seventyone_eightytwo_govt.index)
seventyone_eightytwo_df_corrected
```

	Gender	Age	HasTension	AnyHeartDisease	NeverMarried	Occupation	LivesIn	GlucoseLevel	BMI	SmokingStatus	HeartAttack
0	Female	75	Yes	No	Yes	Self-employed	City	54.6	35.1	never smoked	No
3	Male	78	No	No	Yes	Self-employed	City	219.2	27.4	Unknown	Yes
19	Female	82	No	No	Yes	Private	Village	95.8	27.1	Unknown	No
24	Female	75	No	No	Yes	Self-employed	Village	70.8	24.8	formerly smoked	No
39	Female	79	No	No	Yes	Private	Village	92.5	24.5	never smoked	Yes
...	...	...	...	...	...	...	...	...	...	...	...
3966	Male	76	No	No	Yes	Private	City	139.7	30.4	formerly smoked	Yes
3969	Female	81	No	No	Yes	Private	City	77.8	35.3	Unknown	No
3979	Male	78	No	No	Yes	Private	City	103.4	30.4	never smoked	No
3983	Male	77	No	No	Yes	Private	City	223.0	29.6	formerly smoked	No
3993	Male	71	No	No	Yes	Private	City	85.4	28.3	never smoked	No

450 rows × 11 columns

Next steps: Generate code with seventyone_eightytwo_df_corrected View recommended plots New interactive sheet

Now, what percentage of them are male and female

```
seventyone_eightytwo_male=seventyone_eightytwo_df_corrected[seventyone_eightytwo_df_corrected['Gender']=='Male']
a=percentage(seventyone_eightytwo_male.shape[0],seventyone_eightytwo_df_corrected.shape[0])
print(a)
```

```
40.88888888888889
```

```
seventyone_eightytwo_female=seventyone_eightytwo_df_corrected[seventyone_eightytwo_df_corrected['Gender']=='Female']
a=percentage(seventyone_eightytwo_female.shape[0],seventyone_eightytwo_df_corrected.shape[0])
print(a)
```

```
59.111111111111114
```

Therefore, 41% are male and 59% are female

Now what percentage of them has obesity genderwise

```
seventyone_eightytwo_obese=seventyone_eightytwo_df_corrected[seventyone_eightytwo_df_corrected['BMI']>=30]
a=percentage(seventyone_eightytwo_obese.shape[0],seventyone_eightytwo_df_corrected.shape[0])
print(a)
```

```
40.0
```

```
seventyone_eightytwo_obese_male=seventyone_eightytwo_obese[seventyone_eightytwo_obese['Gender']=='Male']
a=percentage(seventyone_eightytwo_obese_male.shape[0],seventyone_eightytwo_obese.shape[0])
print(a)
```

```
36.11111111111111
```

```
seventyone_eightytwo_obese_female=seventyone_eightytwo_obese[seventyone_eightytwo_obese['Gender']=='Female']
a=percentage(seventyone_eightytwo_obese_female.shape[0],seventyone_eightytwo_obese.shape[0])
print(a)
```

```
63.888888888888886
```

Therefore, 40% of people has obesity and out of them 36% are male and 64% are female

Now Finding how many obese people smokes genderwise

```
seventyone_eightytwo_obese_smokes=seventyone_eightytwo_obese[seventyone_eightytwo_obese['SmokingStatus']=='Smokes']
a=percentage(seventyone_eightytwo_obese_smokes.shape[0],seventyone_eightytwo_df_corrected.shape[0])
print(a)
```

```
0.0
```

Therefore, No obese people smokes. Now, finding how many obese people had heartattack genderwise

```
seventyone_eightytwo_obese_heartattack=seventyone_eightytwo_obese[seventyone_eightytwo_obese['HeartAttack']=='Yes']
a=percentage(seventyone_eightytwo_obese_heartattack.shape[0],seventyone_eightytwo_df_corrected.shape[0])
print(a)
```