



LOVELY  
PROFESSIONAL  
UNIVERSITY

*Transforming Education Transforming India*

## REPORT

On

### Descriptive statistics

Submitted by

**EKANSH BARI**

**Registration No 11615973**

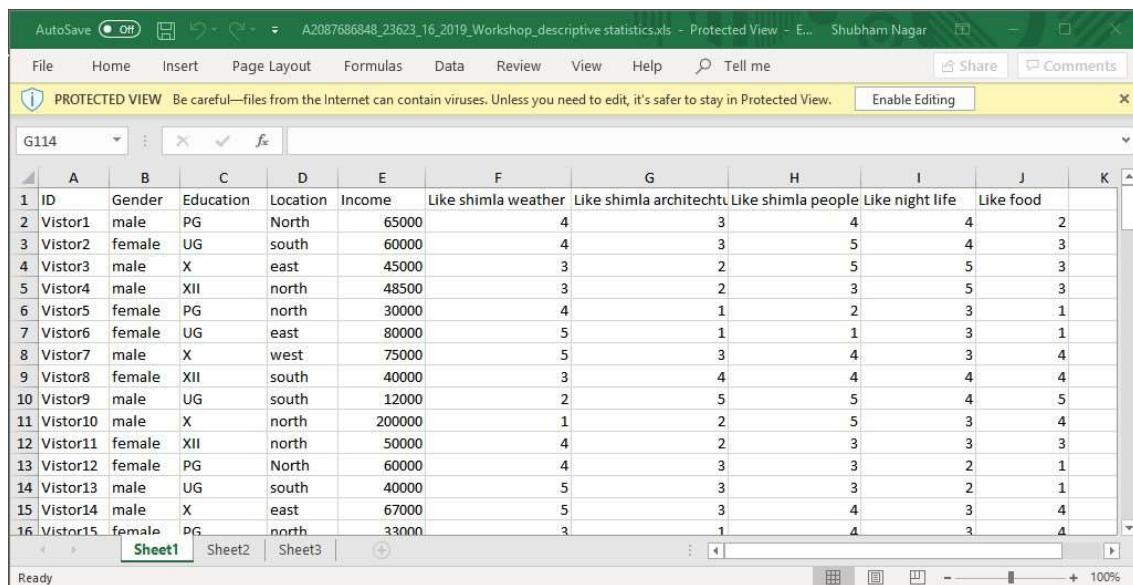
**(MGN909) Section KOE11 - A22**

**Programme Name: B.Tech Computer Science**

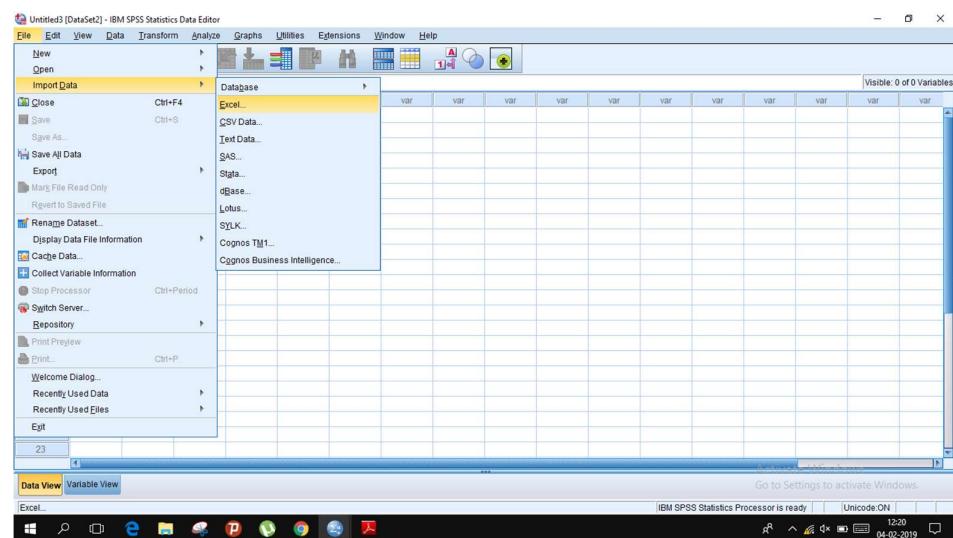
**School of Computer Science & Engineering  
Lovely Professional University, Phagwara**

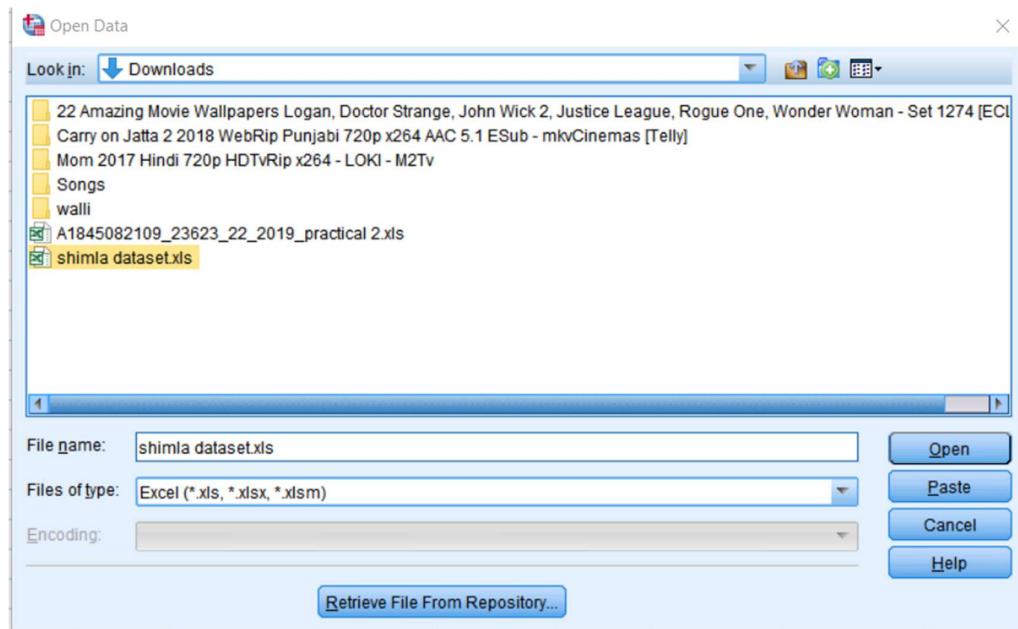
# Assignment – Descriptive statistics

1. **Import excel file:** For importing the excel file, we first go to the file tab and from the dropdown we go to the open and from there to data and from there we select the excel option. After that a panel opens from where we search our file that we have to import.



	A	B	C	D	E	F	G	H	I	J	K
1	ID	Gender	Education	Location	Income	Like shimla weather	Like shimla architechture	Like shimla people	Like night life	Like food	
2	Vistor1	male	PG	North	65000	4	3	4	4	2	
3	Vistor2	female	UG	south	60000	4	3	5	4	3	
4	Vistor3	male	X	east	45000	3	2	5	5	3	
5	Vistor4	male	XII	north	48500	3	2	3	5	3	
6	Vistor5	female	PG	north	30000	4	1	2	3	1	
7	Vistor6	female	UG	east	80000	5	1	1	3	1	
8	Vistor7	male	X	west	75000	5	3	4	3	4	
9	Vistor8	female	XII	south	40000	3	4	4	4	4	
10	Vistor9	male	UG	south	12000	2	5	5	4	5	
11	Vistor10	male	X	north	200000	1	2	5	3	4	
12	Vistor11	female	XII	north	50000	4	2	3	3	3	
13	Vistor12	female	PG	North	60000	4	3	3	2	1	
14	Vistor13	male	UG	south	40000	5	3	3	2	1	
15	Vistor14	male	X	east	67000	5	3	4	3	4	
16	Vistor15	female	PG	north	33000	3	1	4	3	4	

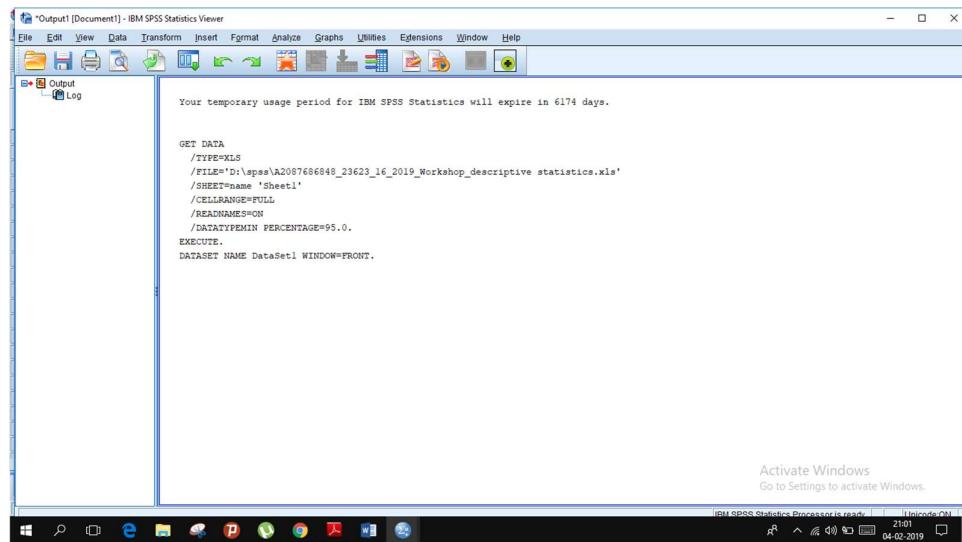




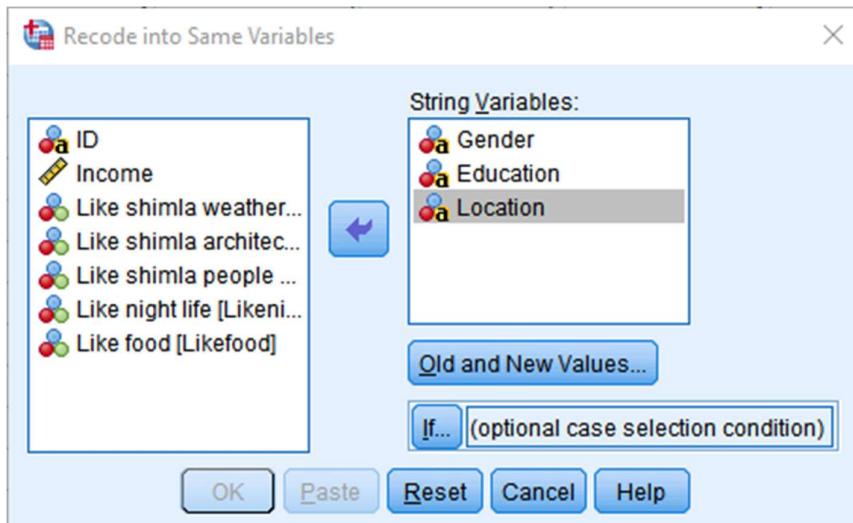
Select the range of row and column.

	ID	Gender	Educati...	Location	Income	Likeshi...	L
1	Vistor1	male	PG	North	65000	4	3
2	Vistor2	female	UG	south	60000	4	3
3	Vistor3	male	X	east	45000	3	2
4	Vistor4	male	XII	north	48500	3	2
5	Vistor5	female	PG	north	30000	4	1
6	Vistor6	female	UG	east	80000	5	1

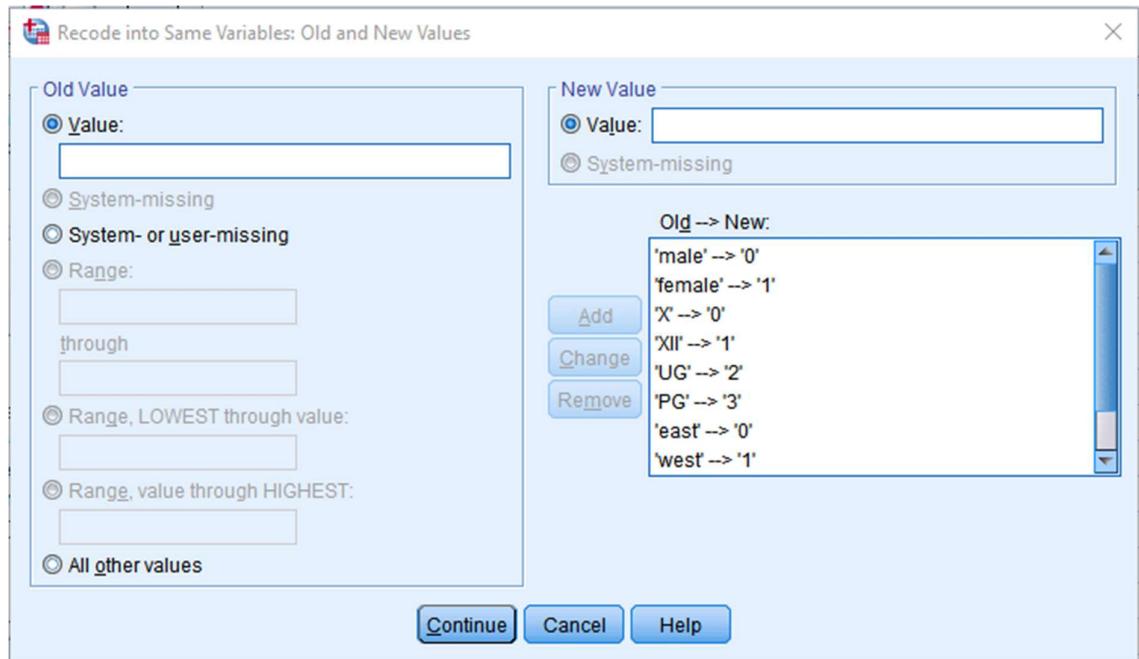
After selecting the range press ok then the dataset is imported.



2. **Recode text into numeric such as – gender, education and location etc :** For this we need to transform our dataset the we have imported. For transforming go to transform on the ribbon then select recode into same variable. Then select the variable name for which you want to transform and recode the dataset for gender, education and location.

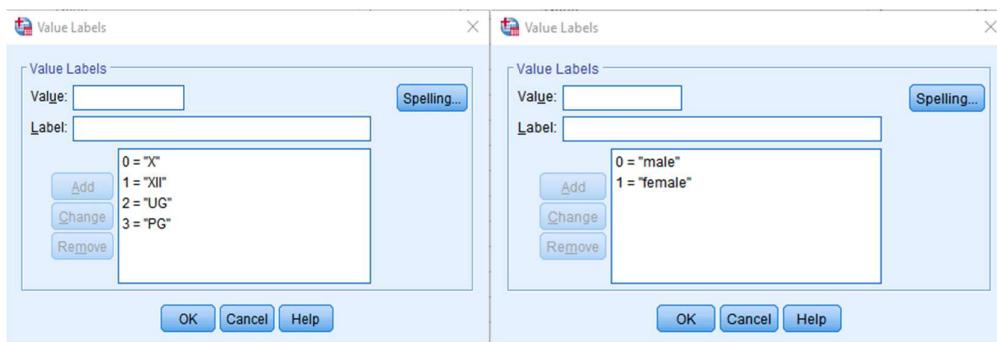


Assign values for strings.



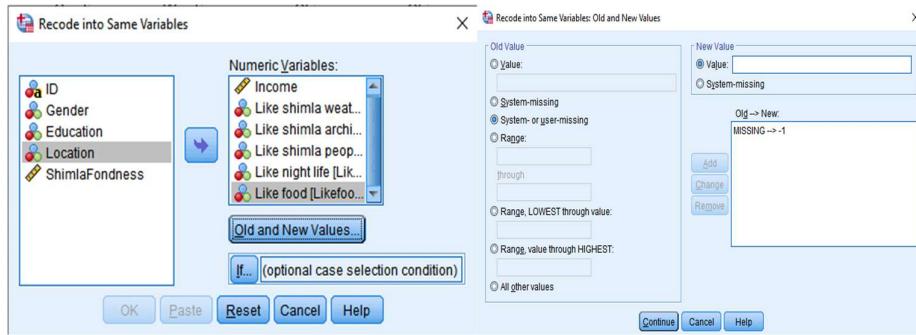
After assigning values switch to variable view and change width & type of variable and assign values to variable.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
ID	String	9	0		None	None	9	Left	Nominal	Input
Gender	Numeric	1	0		{0, male}...	None	6	Right	Nominal	Input
Education	Numeric	1	0		{0, X}...	None	3	Right	Nominal	Input
Location	Numeric	1	0		{0, Rest}...	None	5	Right	Nominal	Input
Income	Numeric	11	0		None	-1	11	Right	Scale	Input
Likeshimlaw...	Numeric	1	0	Like shimla we...	{1, Worst}...	-1	11	Right	Nominal	Input
Likeshimlaa...	Numeric	1	0	Like shimla arc...	{1, Worst}...	-1	11	Right	Nominal	Input
Likeshimlap...	Numeric	1	0	Like shimla peo...	{1, Worst}...	-1	11	Right	Nominal	Input
Likenightlife	Numeric	1	0	Like night life	{1, Worst}...	-1	11	Right	Nominal	Input
Likefood	Numeric	1	0	Like food	{1, Worst}...	-1	11	Right	Nominal	Input
ShimlaFond...	Numeric	8	2		None	-1.00	16	Right	Scale	Input



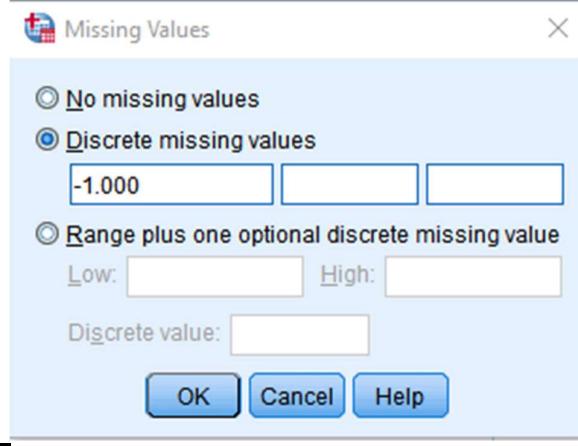


- 3. Assign missing code on – income , like weather, like architecture, like food, like night life, like people :** In this we will assign some value. It can be done by using same process we used in step 2.



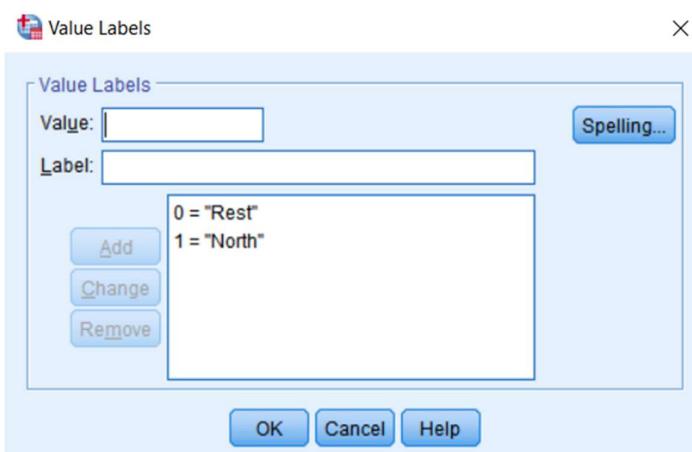
In variable view assign missing column otherwise these values will be used in calculations.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
ID	String	9	0		None	None	9	Left	Nominal	Input
Gender	Numeric	1	0		{0, male}...	None	6	Right	Nominal	Input
Education	Numeric	1	0		{0, X}...	None	3	Right	Nominal	Input
Location	Numeric	1	0		{0, Rest}...	None	5	Right	Nominal	Input
Income	Numeric	11	0		None	-1	11	Right	Scale	Input
Likeshimlaw...	Numeric	1	0	Like shimla we...	{1, Worst}...	-1	11	Right	Nominal	Input
Likeshimlaa...	Numeric	1	0	Like shimla arc...	{1, Worst}...	-1	11	Right	Nominal	Input
Likeshimlap...	Numeric	1	0	Like shimla peo...	{1, Worst}...	-1	11	Right	Nominal	Input
Likenightlife	Numeric	1	0	Like night life	{1, Worst}...	-1	11	Right	Nominal	Input
Likefood	Numeric	1	0	Like food	{1, Worst}...	-1	11	Right	Nominal	Input
ShimlaFond...	Numeric	8	2		None	-1.00	16	Right	Scale	Input



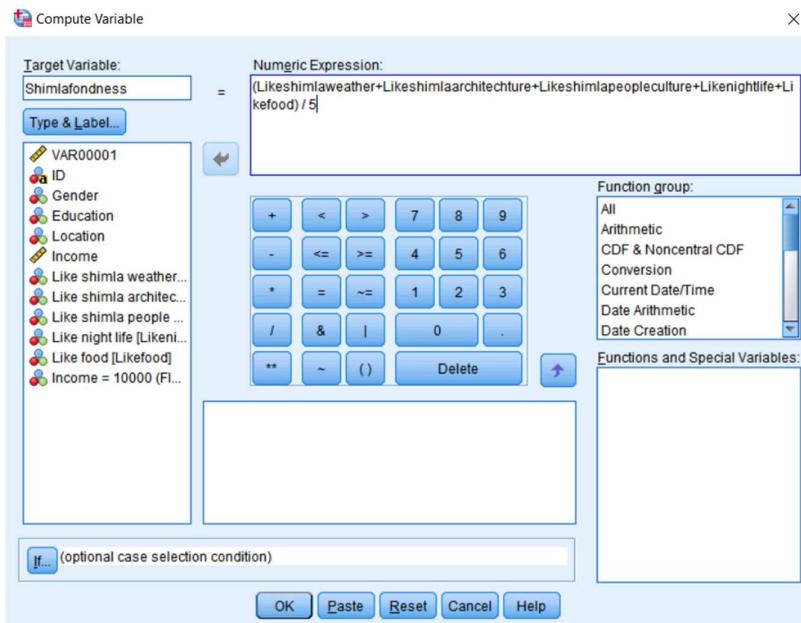
- 4. Recode location : North Vs Rest ( 2 category) :** In this we will code the location by the attribute value on the variable view and code the value as 1 = North and 0 = Rest.

The SPSS Data Editor window shows a dataset with 22 rows and 11 columns. The columns are labeled ID, Gender, E\_u, Location, Income, Likeshimpawather, Likeshimparchitecture, Likeshimpapeculpture, Likenightsife, Likefood, and shimpafondness. A 'Recode into Same Variables: Old and New Values' dialog box is open over the data view. It contains two main sections: 'Old Value' and 'New Value'. Under 'Old Value', there are radio buttons for 'System-missing', 'System- or user-missing', 'Range', 'Range, LOWEST through value', 'Range, value through HIGHEST', and 'All other values'. Under 'New Value', there are radio buttons for 'Value:' and 'System-missing'. A 'Change...' button is present. A preview section shows the mapping: 'Old --> New: North -> 1 ELSE -> 2'. Buttons at the bottom include 'Add', 'Change...', 'Remove', 'Continue', 'Cancel', and 'Help'. The status bar at the bottom right shows 'IBM SPSS Statistics Processor is ready' and the date '04-02-2019'.



5. **Compute variable : Shimla fondness** : This is done by the option Recode into different variable from the compute variable tab at menu ribbon giving it a new variable name as Shimla fondness and the formulation goes as  

$$\text{shimlafondness} = (\text{Likeshimlaweather} + \text{Likeshimlaarchitechture} + \text{Likeshimlapeopleculture} + \text{Likenightlife} + \text{Likefood}) / 5$$



6. **Compute frequency** : In this by the help of Analyse tab and descriptive statistics and then to frequencies we can calculate frequency.

- a. **Architecture** : Here the snapshot attached below depicts the frequency of architecture's review by the people as worst, bad, ok, good, best and also it's percentile (without missing value) with valid percentile (with missing value included) and cumulative percentile. By this one can infer that how is the architecture by this review of people

## → Frequencies

### Statistics

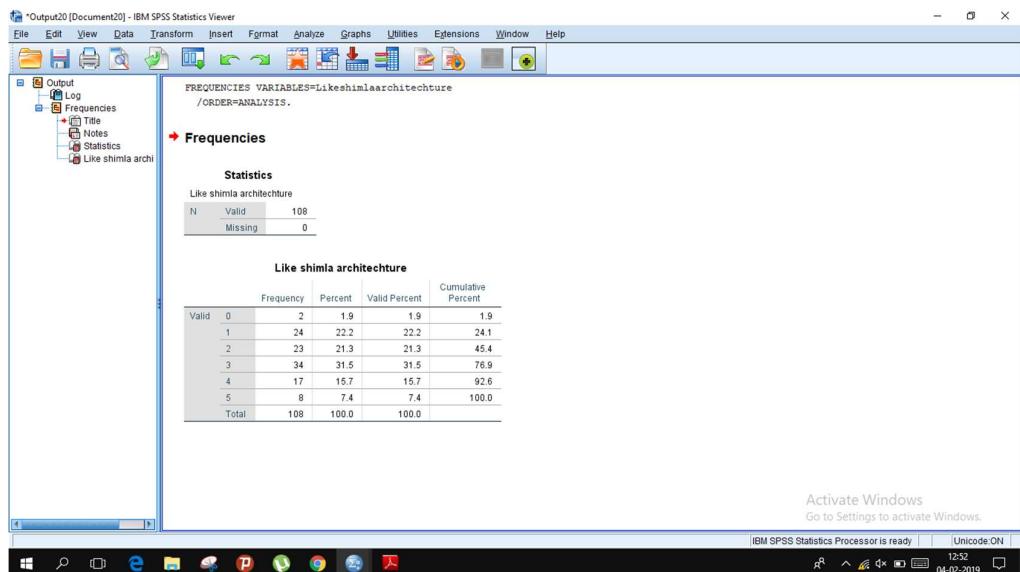
Like shimla architechture

N	Valid	106
	Missing	2
Mode		3
Range		4
Minimum		1
Maximum		5

## Frequency Table

Like shimla architecture					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Worst	24	22.2	22.6	22.6
	Bad	23	21.3	21.7	44.3
	Ok	34	31.5	32.1	76.4
	Good	17	15.7	16.0	92.5
	Best	8	7.4	7.5	100.0
	Total	106	98.1	100.0	
Missing	-1	2	1.9		
	Total	108	100.0		

Around 55.7 % people like Shimla architecture and around 44.3 % doesn't like Shimla architecture; nearly half of the people like Shimla architecture and half do not.



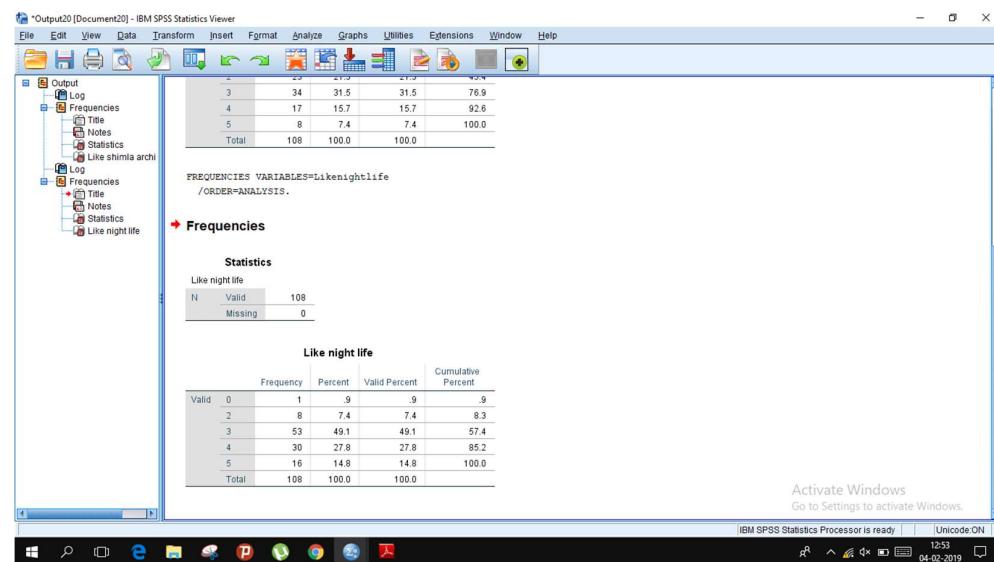
- b. **Nightlife** :As per above here also the snapshot attached below depicts the frequency of nightlife's review by the people as bad, ok, good, best and also it's percentile (without missing value) with valid percentile(with missing value included) and cumulative percentile. By this one can infer that how is the nightlife by this review of people

## Frequencies

### Statistics

Like night life

N	Valid	107
	Missing	1
Mode		3
Range		3
Minimum		2
Maximum		5



Like night life					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Bad	8	7.4	7.5	7.5
	Ok	53	49.1	49.5	57.0
	Good	30	27.8	28.0	85.0
	Best	16	14.8	15.0	100.0
	Total	107	99.1	100.0	
	Missing	-1	.9		
Total		108	100.0		

Maximum number of people like Shimla's night life.

7. **Cross tab frequency : Location and fondness**: The below picture states the frequency is a cross tab i.e. horizontally with one attribute as a row compulsorily. Cross tabulations are simply data tables that present the results of the entire group of respondents as well as results from sub-groups of survey respondents. Cross tabulations enable you to examine

relationships within the data that might not be readily apparent when analysing total survey responses.

Here the cross tab gives relation between location and fondness.

		Location * ShimlaFondness Crosstabulation															
		ShimlaFondness															
		1.80	2.20	2.40	2.60	2.80	3.00	3.20	3.40	3.60	3.80	4.00	4.20	4.40	4.60	Total	
Location	Rest	Count	1	2	3	5	5	10	4	5	5	13	4	2	3	0	62
		Expected Count	.6	3.6	4.2	4.8	4.8	7.8	4.8	6.0	6.0	10.8	3.6	2.4	1.8	.6	62.0
		% within Location	1.6%	3.2%	4.8%	8.1%	8.1%	16.1%	6.5%	8.1%	8.1%	21.0%	6.5%	3.2%	4.8%	0.0%	100.0%
		% within ShimlaFondness	100.0%	33.3%	42.9%	62.5%	62.5%	76.9%	50.0%	50.0%	50.0%	72.2%	66.7%	50.0%	100.0%	0.0%	60.2%
		% of Total	1.0%	1.9%	2.9%	4.9%	4.9%	9.7%	3.9%	4.9%	4.9%	12.6%	3.9%	1.9%	2.9%	0.0%	60.2%
North	Count	0	4	4	3	3	3	4	5	5	5	2	2	0	1	41	
		Expected Count	.4	2.4	2.8	3.2	3.2	5.2	3.2	4.0	4.0	7.2	2.4	1.6	1.2	.4	41.0
		% within Location	0.0%	9.8%	9.8%	7.3%	7.3%	7.3%	9.8%	12.2%	12.2%	12.2%	4.9%	4.9%	0.0%	2.4%	100.0%
		% within ShimlaFondness	0.0%	66.7%	57.1%	37.5%	37.5%	23.1%	50.0%	50.0%	50.0%	27.8%	33.3%	50.0%	0.0%	100.0%	39.8%
		% of Total	0.0%	3.9%	3.9%	2.9%	2.9%	2.9%	3.9%	4.9%	4.9%	4.9%	1.9%	1.9%	0.0%	1.0%	39.8%
Total	Count	1	6	7	8	8	13	8	10	10	18	6	4	3	1	103	
		Expected Count	1.0	6.0	7.0	8.0	8.0	13.0	8.0	10.0	10.0	18.0	6.0	4.0	3.0	1.0	103.0
		% within Location	1.0%	5.8%	6.8%	7.8%	7.8%	12.6%	7.8%	9.7%	9.7%	17.5%	5.8%	3.9%	2.9%	1.0%	100.0%
		% within ShimlaFondness	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	1.0%	5.8%	6.8%	7.8%	7.8%	12.6%	7.8%	9.7%	9.7%	17.5%	5.8%	3.9%	2.9%	1.0%	100.0%

39.8 % of people are from north and rest are from east, west and south. Majorly people like Shimla

### Symmetric Measures

		Value	Asymptotic Standard Error <sup>a</sup>	Approximate T <sup>b</sup>	Approximate Significance
Interval by Interval	Pearson's R	-.083	.099	-.838	.404 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	-.090	.099	-.904	.368 <sup>c</sup>
N of Valid Cases		103			

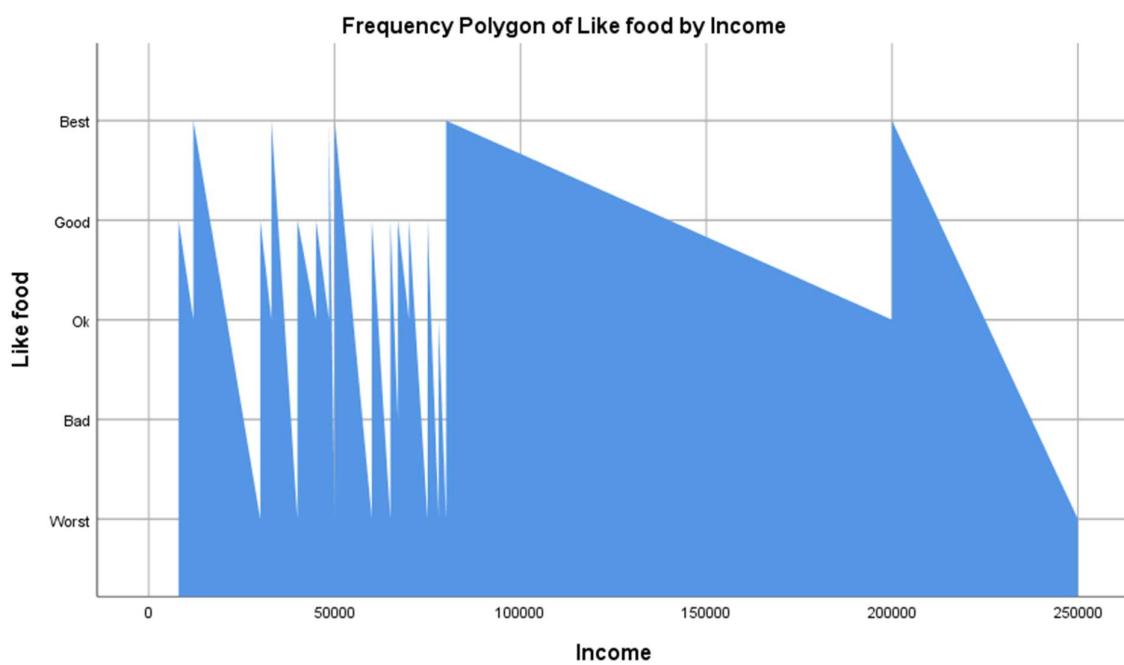
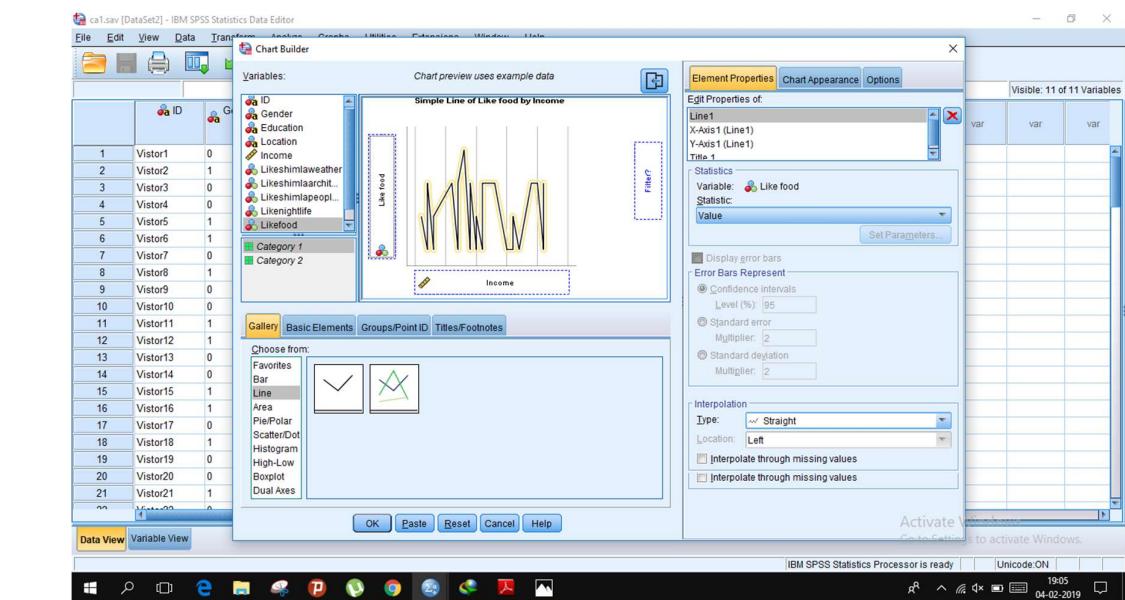
a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

There is strong relationship between location and fondness.

8. **Draw frequency polygon of income , food :** This is the polygon frequency graph between income v/s food by which we can infer that how the income influences the choices of people. The plot here tells or gives us an idea how the fluctuation takes place.

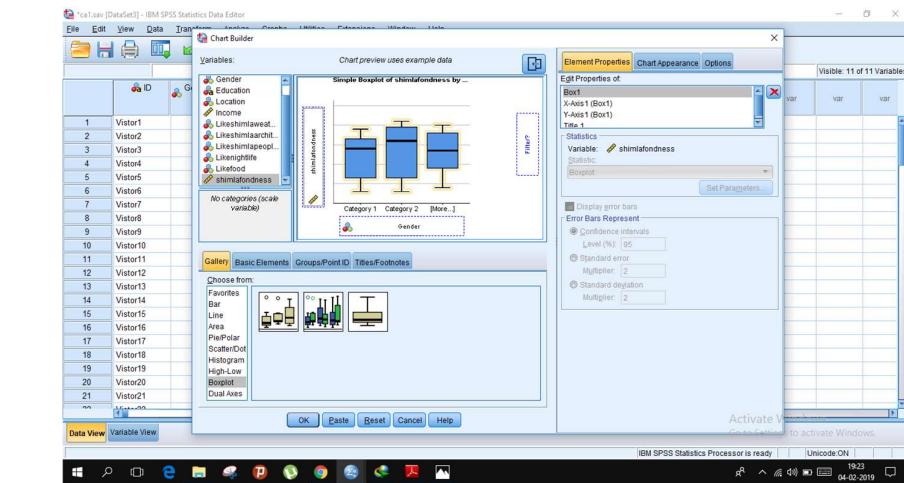
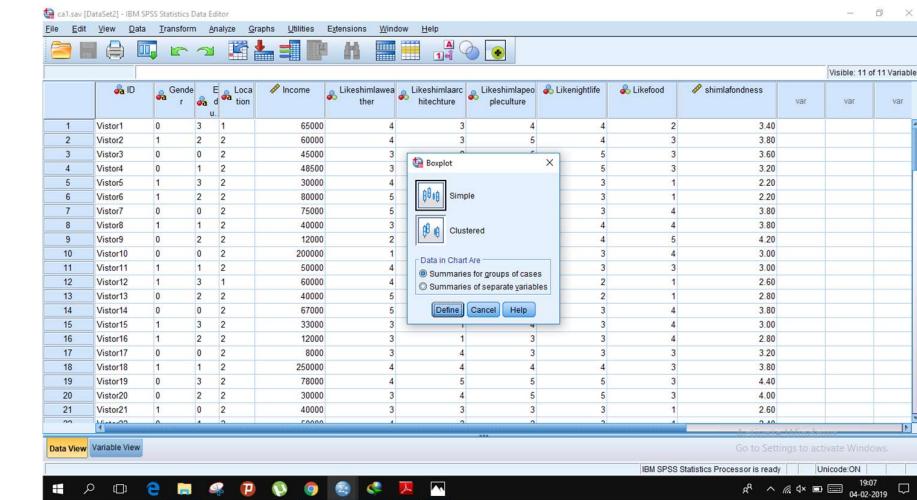


From the above graph, we can analyse relation between income and likefood as follows

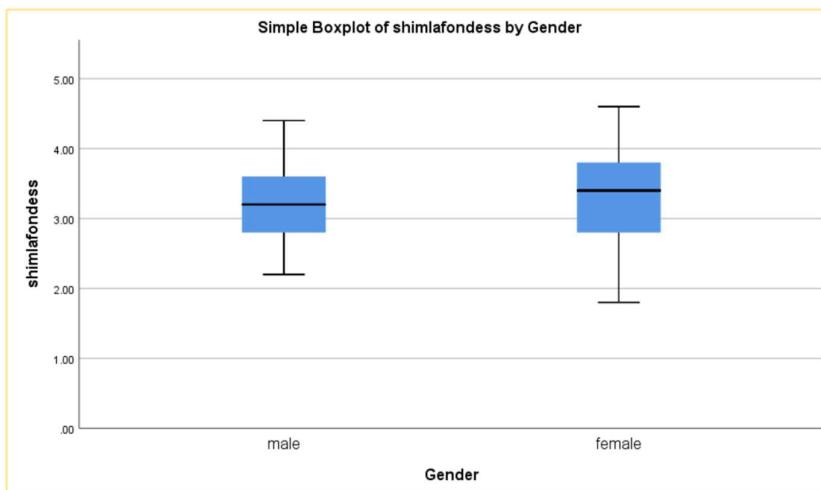
- Income range 0-50,000 , people liked food(Agree-4).
- Income range 50,000-1,00,000 people liked food(Normal-3)
- Income range 1,00,000-1,50,000 people liked food(Agree->4).

- Income range 1,50,000-2,00,000 people liked food(Agree-nearly 4)
- Income range 2,00,000-2,50,000 people liked food(Normal-3)

**9. Box plot – Gender, fondness** : A boxplot can show whether a data set is symmetric (roughly the same on each side when cut down the middle) or skewed (lopsided). A symmetric data set shows the median roughly in the middle of the box. The median, part of the five-number summary, is shown by the line that cuts through the box in the boxplot.



→ GGraph



- 10. Normality – fondness, income, like Shimla people :** In assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: graphically and numerically.

**a. Income:**

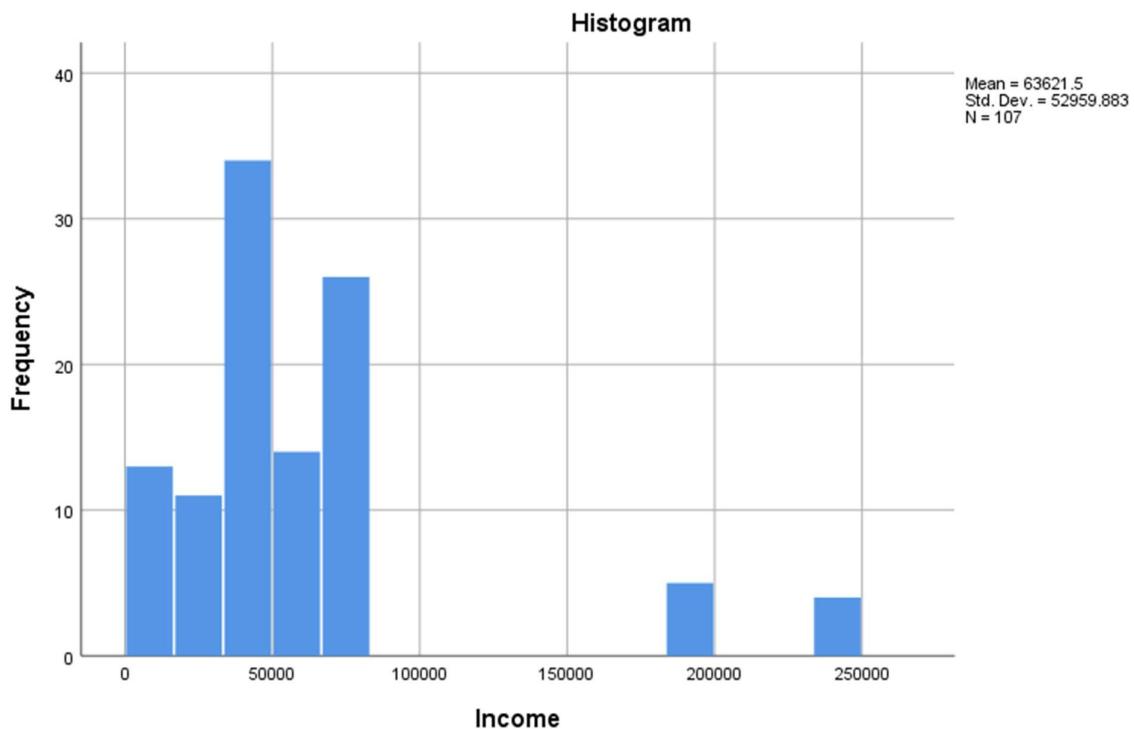
**Descriptives**

		Statistic	Std. Error
Income	Mean	63621.50	5119.825
	95% Confidence Interval for Mean		
	Lower Bound	53470.94	
	Upper Bound	73772.05	
	5% Trimmed Mean	57002.08	
	Median	50000.00	
	Variance	2804749251	
	Std. Deviation	52959.883	
	Minimum	8000	
	Maximum	250000	
	Range	242000	
	Interquartile Range	35000	
	Skewness	2.333	.234
	Kurtosis	5.388	.463

**Tests of Normality**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Income	.294	107	.000	.693	107	.000

a. Lilliefors Significance Correction



### b. Fondness:

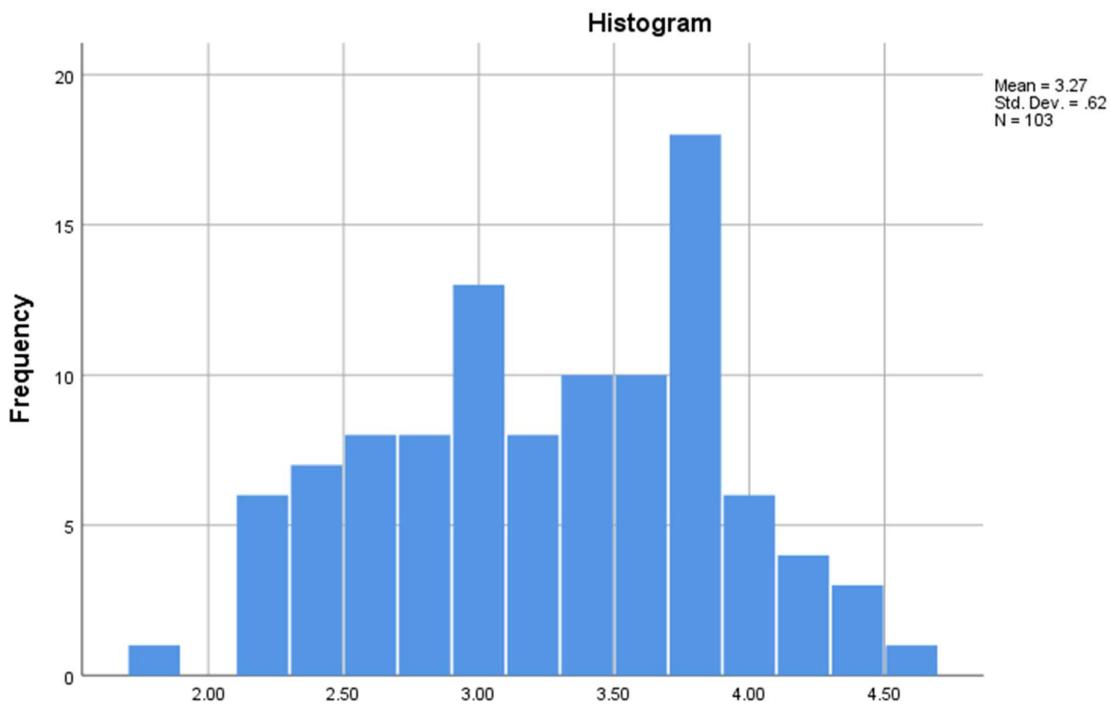
#### Descriptives

		Statistic	Std. Error
ShimlaFondness	Mean	3.2680	.06112
	95% Confidence Interval for Mean		
	Lower Bound	3.1467	
	Upper Bound	3.3892	
	5% Trimmed Mean	3.2690	
	Median	3.4000	
	Variance	.385	
	Std. Deviation	.62028	
	Minimum	1.80	
	Maximum	4.60	
	Range	2.80	
	Interquartile Range	1.00	
	Skewness	-.130	.238
	Kurtosis	-.783	.472

#### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ShimlaFondness	.115	103	.002	.971	103	.024

a. Lilliefors Significance Correction



### c. Like Shimla People:

#### Descriptives

		Statistic	Std. Error
Like shimla people & culture	Mean	3.64	.104
	95% Confidence Interval for Mean		
	Lower Bound	3.43	
	Upper Bound	3.84	
	5% Trimmed Mean	3.69	
	Median	4.00	
	Variance	1.158	
	Std. Deviation	1.076	
	Minimum	1	
	Maximum	5	
	Range	4	
	Interquartile Range	2	
	Skewness	-.336	.234
	Kurtosis	-.425	.463

#### Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Like shimla people & culture	.209	107	.000	.876	107	.000

a. Lilliefors Significance Correction

