

**REPORT**

On

**Descriptive statistics**

Submitted by

**EKANSH BARI**

**Registration No 11615973**

**(MGN909) Section KOE11 - A22**

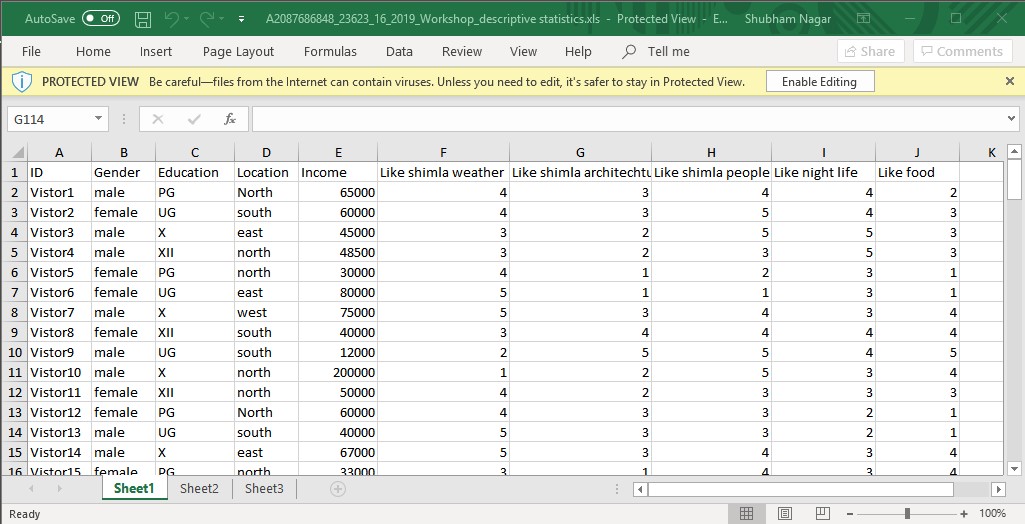
**Programme Name: B.Tech Computer Science**

**School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

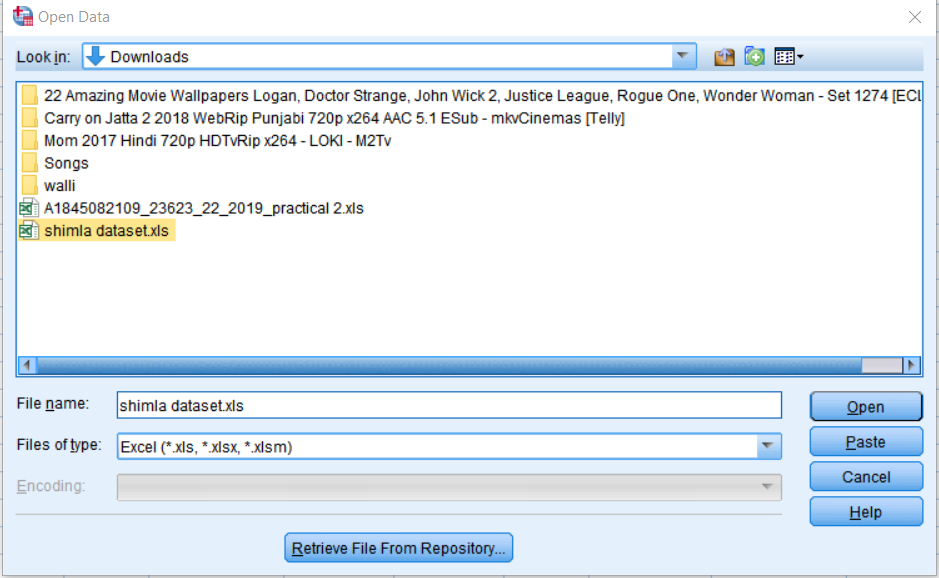
Assignment – **Descriptive statistics**

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.

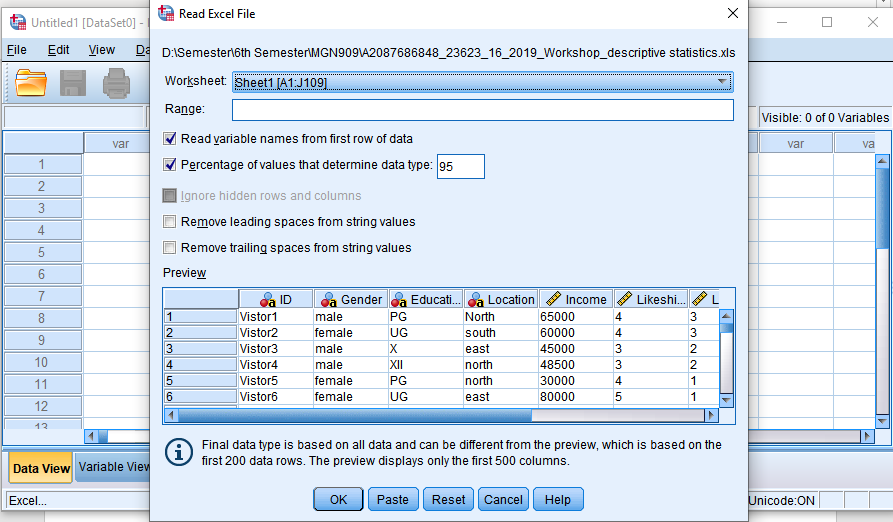


A screenshot of a computer

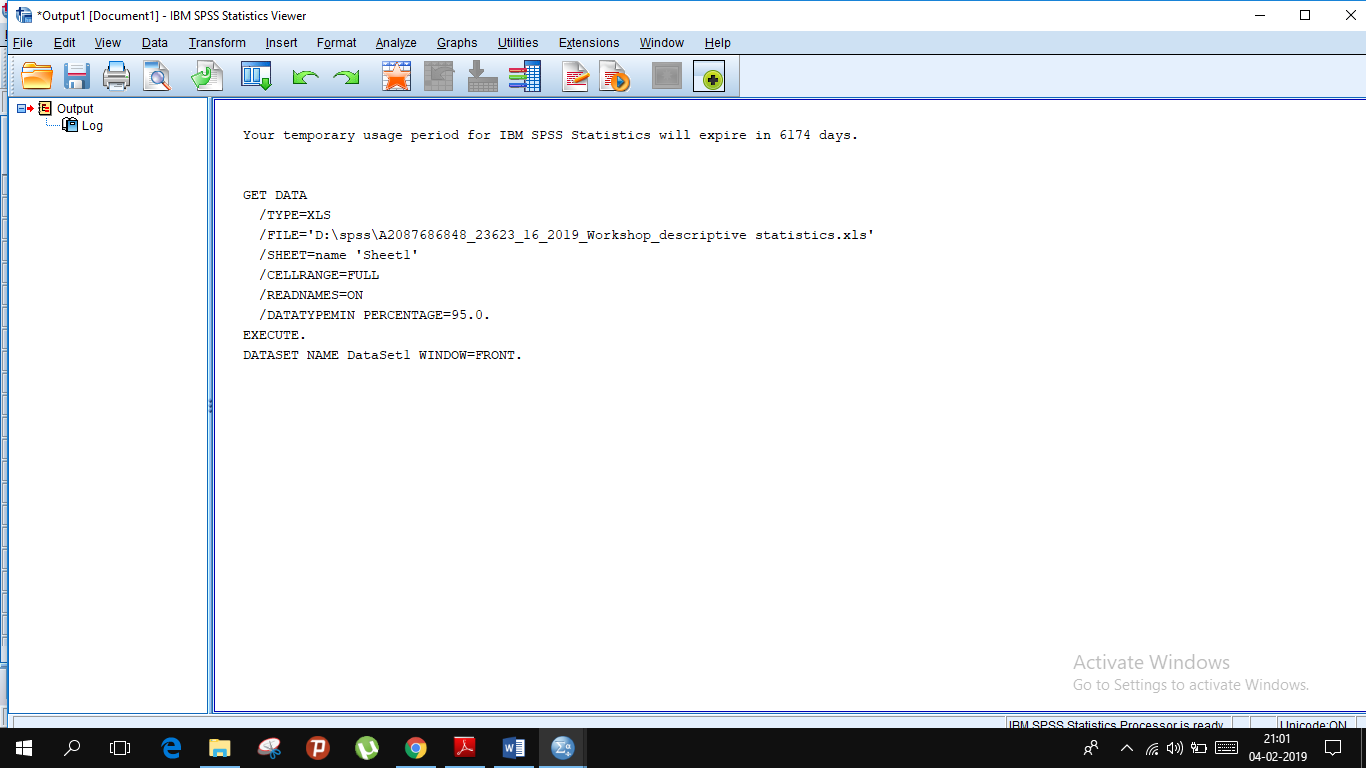
Description automatically generated



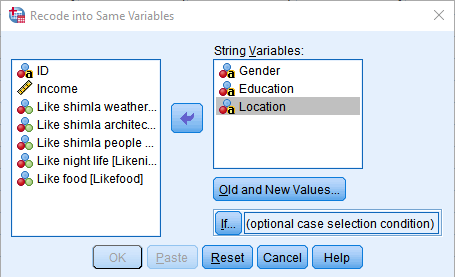
Select the range of row and column.



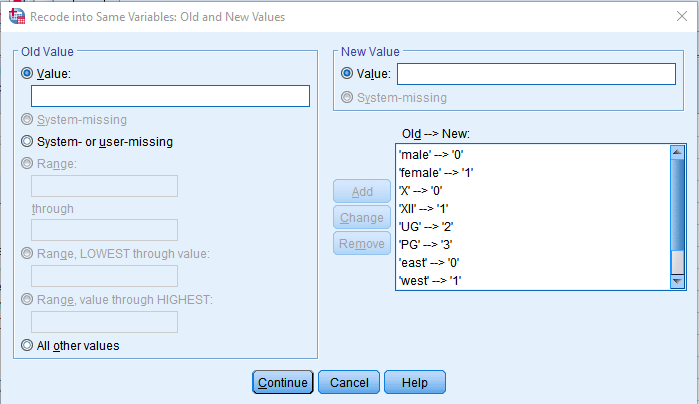
After selecting the range press ok then the dataset is imported.



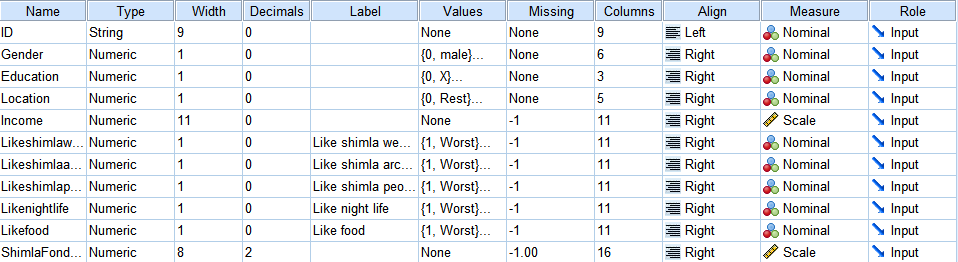
1. **Recode text into numeric such as – gender, education and location etc :** For this we need to transform our dataset the we have imported. For transforming go to transform on the ribbon then select recode into same variable. Then select the variable name for which you want to transform and recode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

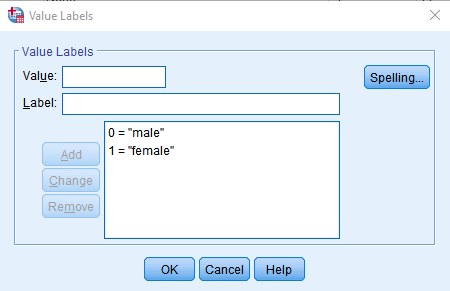
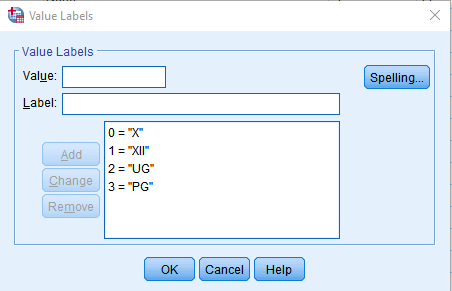


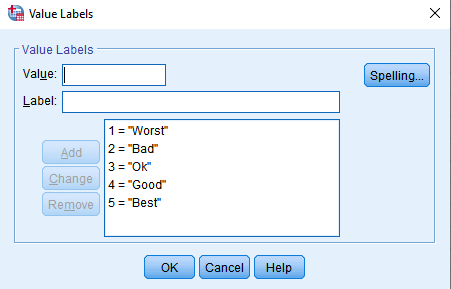
Assign values for strings.



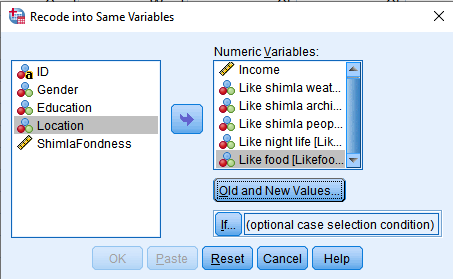
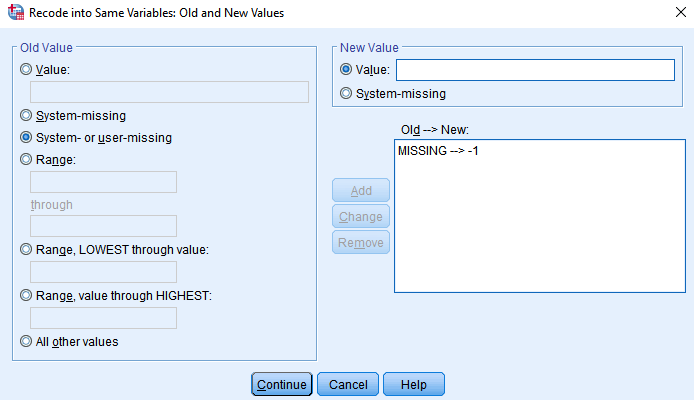
After assigning values switch to variable view and change width & type of variable and assign values to variable.



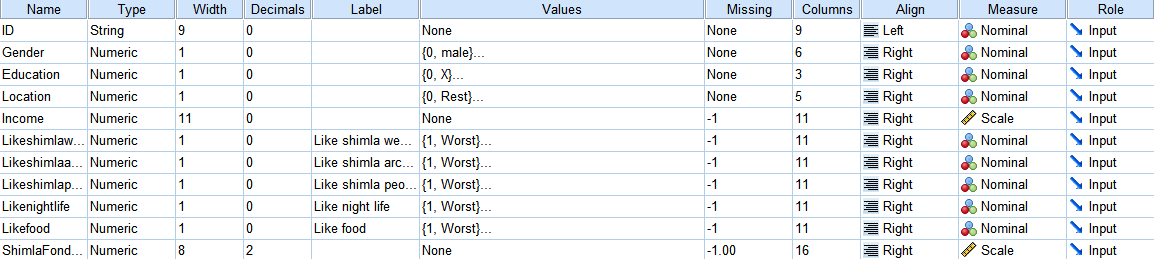


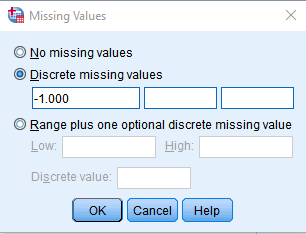


1. **Assign missing code on – income , like weather, like architecture, like food, like night life, like people** : In this we will assign some value. It can be done by using same process we used in step 2.

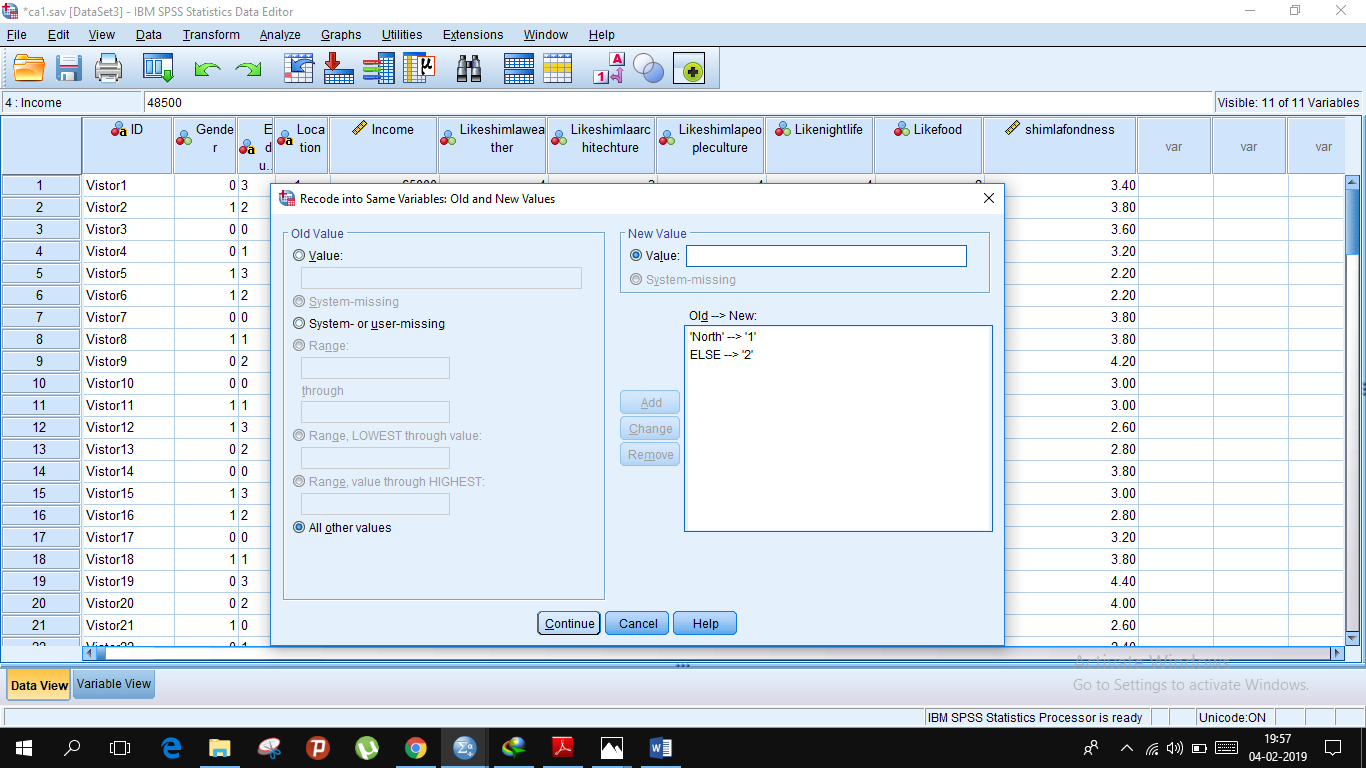
****

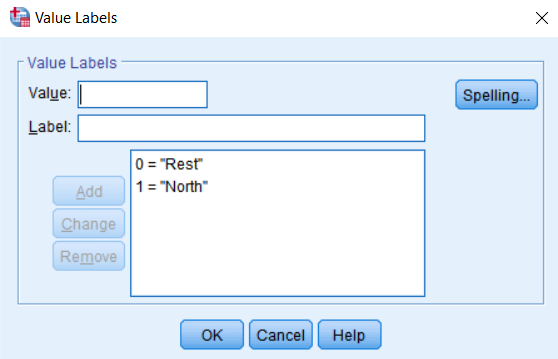
In variable view assign missing column otherwise these values will be used in calculations.





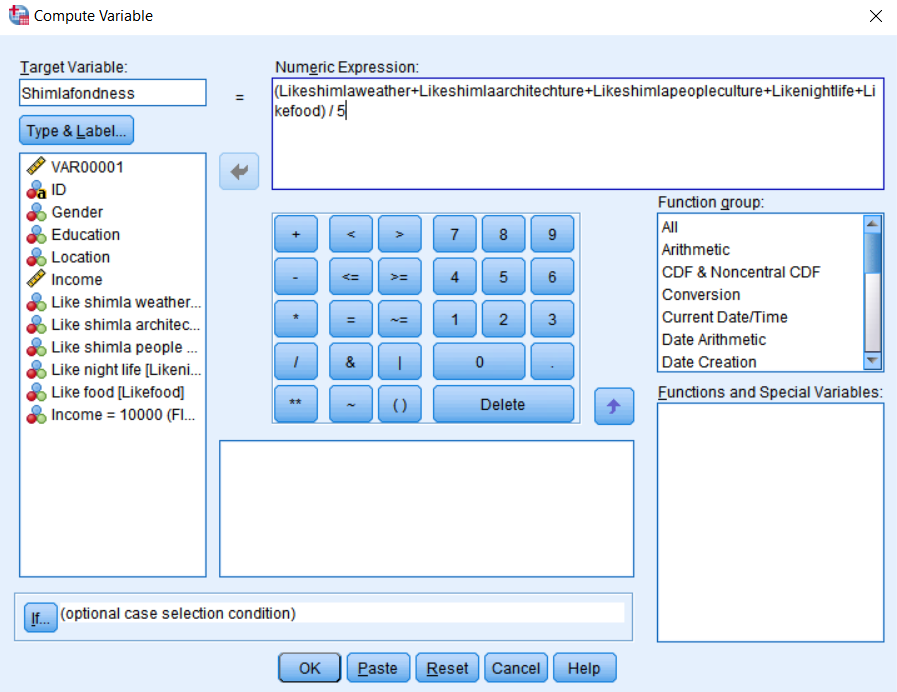
1. **Recode location : North Vs Rest ( 2 category) :** In this we will code the location by the attribute value on the variable view and code the value as 1 = North and 0 = Rest.



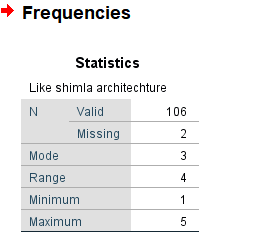


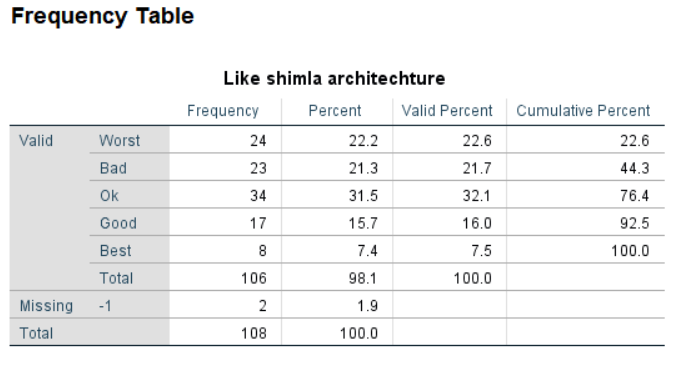
1. **Compute variable : Shimla fondness** : This is done by the option Recode into different variable from the compute variable tab at menu ribbon giving it a new variable name as Shimla fondness and the formulation goes as

shimlafondess=(Likeshimlaweather + Likeshimlaarchitechture + Likeshimlapeopleculture + Likenightlife + Likefood)/



1. **Compute frequency** : In this by the help of Analyse tab and descriptive statistics and then to frequencies we can calculate frequency.
2. **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



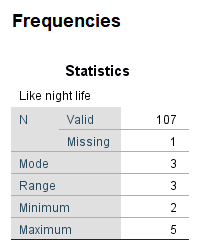


Around 55.7 % people like Shimla architecture and around 44.3 % doesn’t like Shimla architecture; nearly half of the people like Shimla architecture and half do not.

A screenshot of a computer

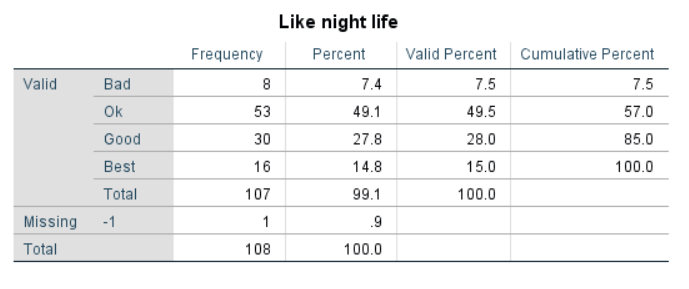
Description automatically generated

1. **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



A screenshot of a computer

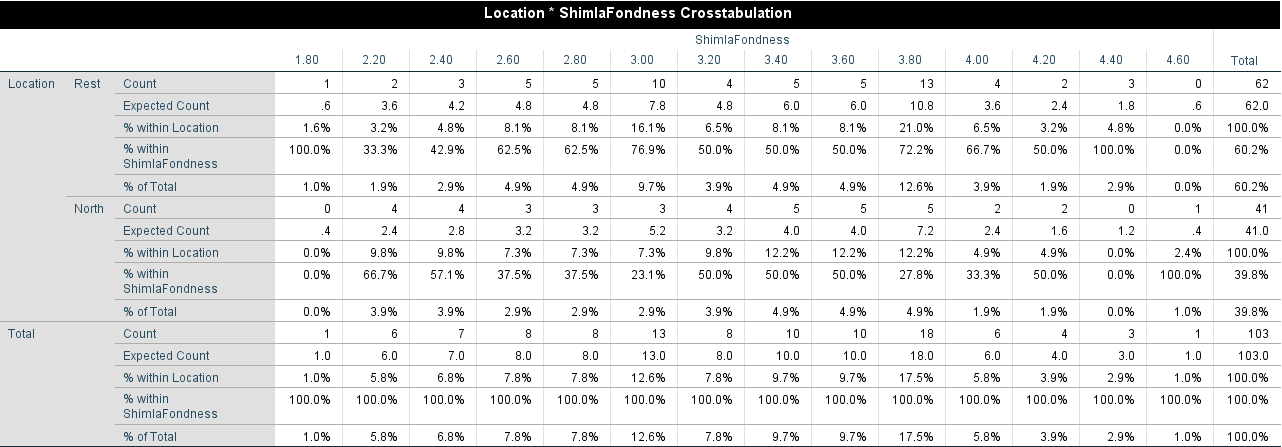
Description automatically generated



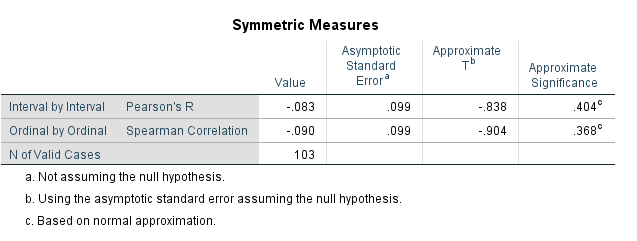
Maximum number of people like Shimla’s night life.

1. **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. Crosstabulations 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.

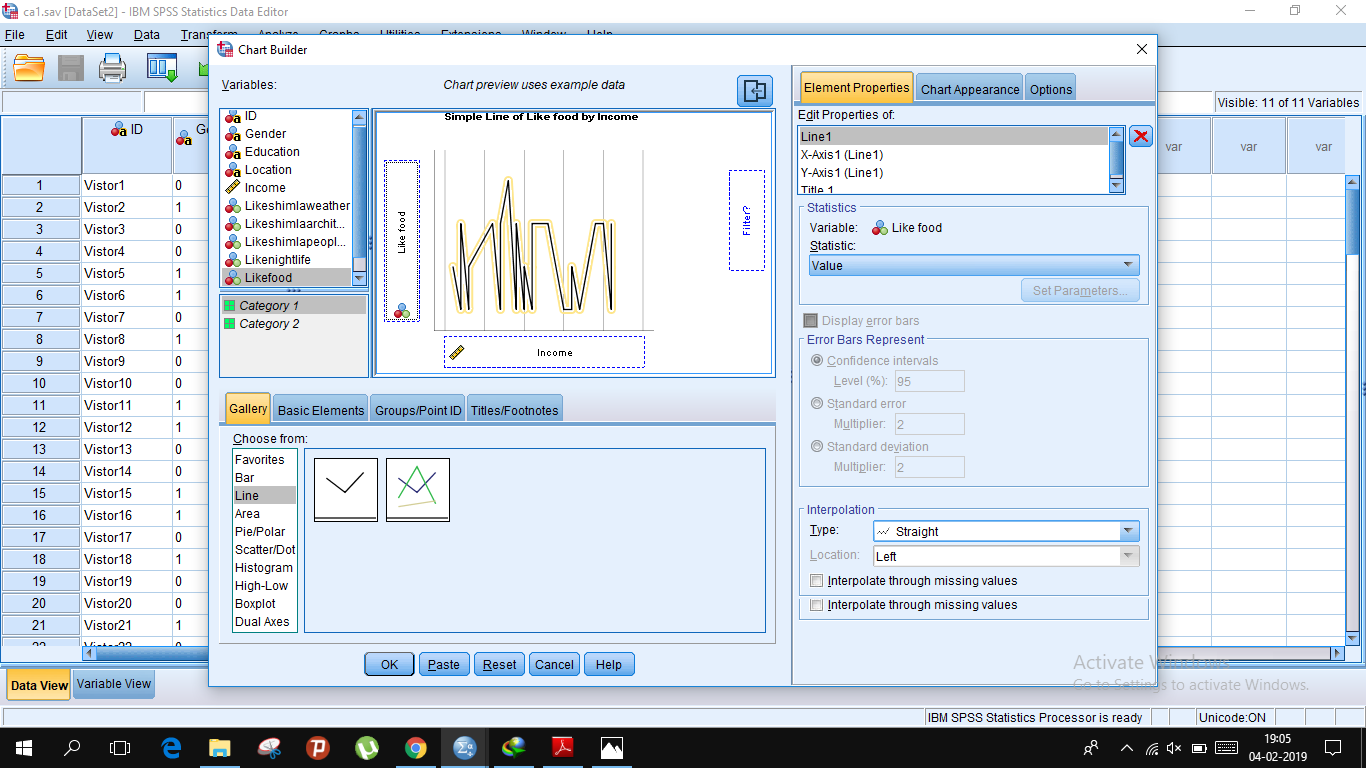


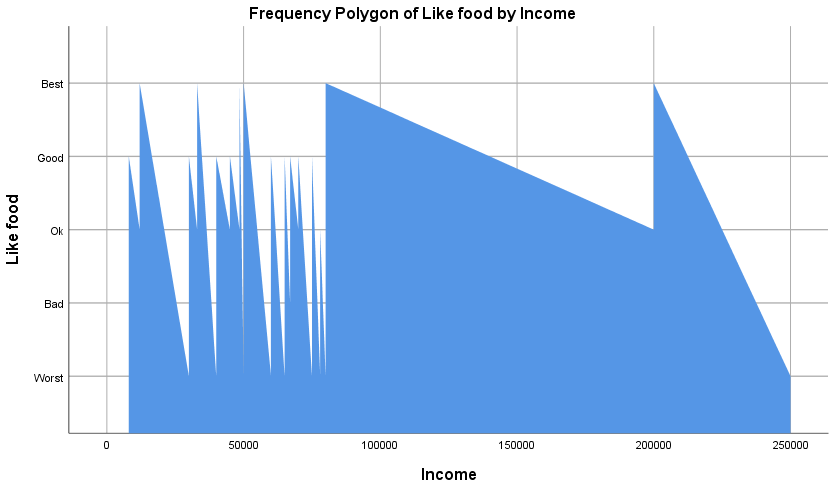
39.8 % of people are form north and rest are from east, west and south. Majorly people like Shimla



There is strong relationship between location and fondness.

1. **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 range1,00,000-1,50,000 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)

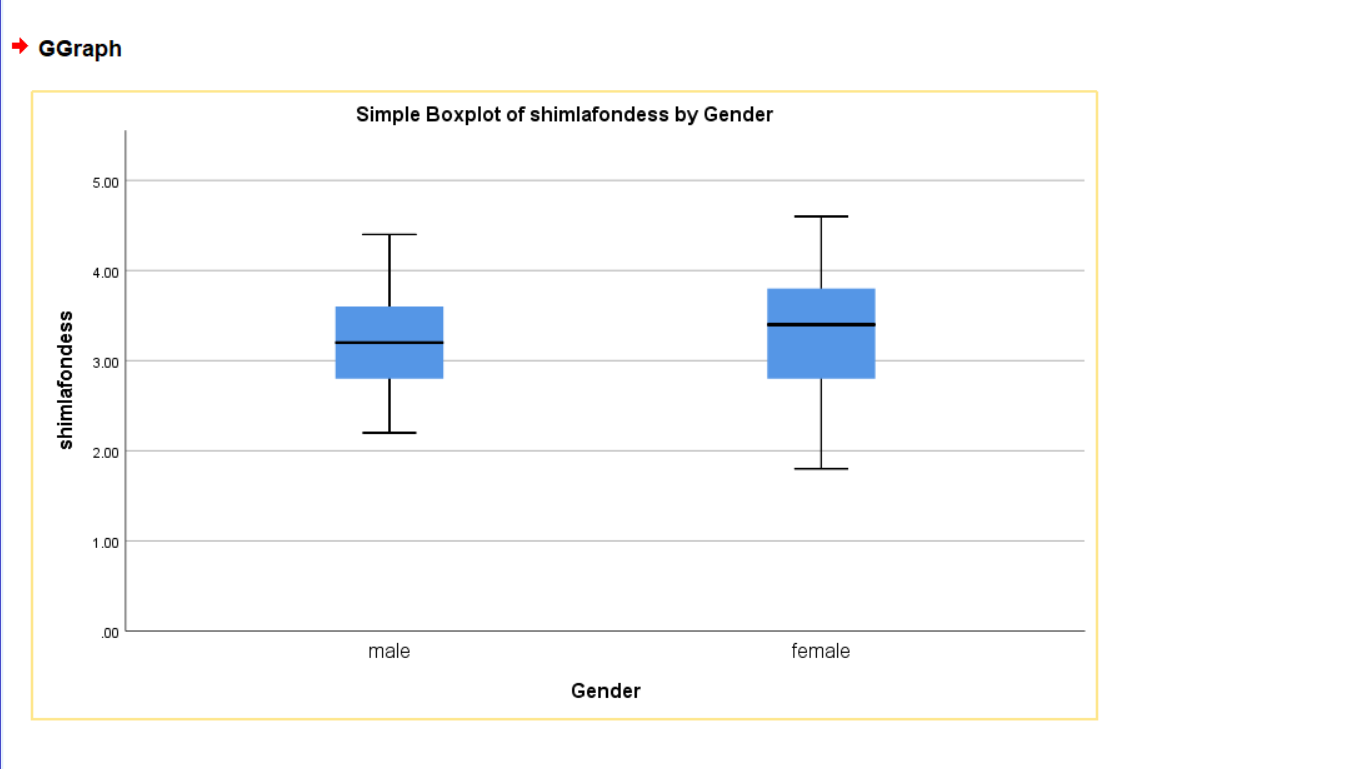
1. **Box plot – Gender, fondness**  : A boxplot can show whether a data set is symmetric (roughly the same on each side when cut down the middle) or skewed (lopsided). A symmetric data set shows the median roughly in the middle of the box. The median, part of the five-number summary, is shown by the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

A screenshot of a computer

Description automatically generated

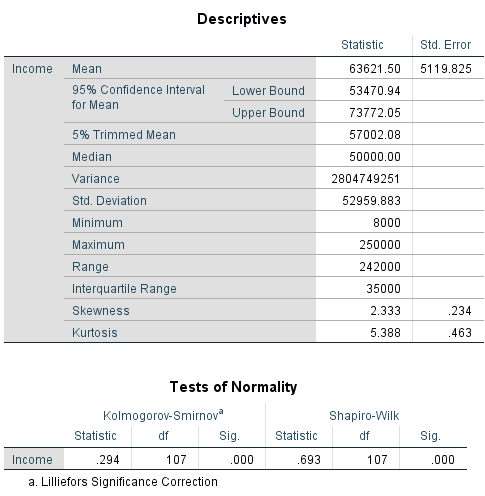
A screenshot of a computer

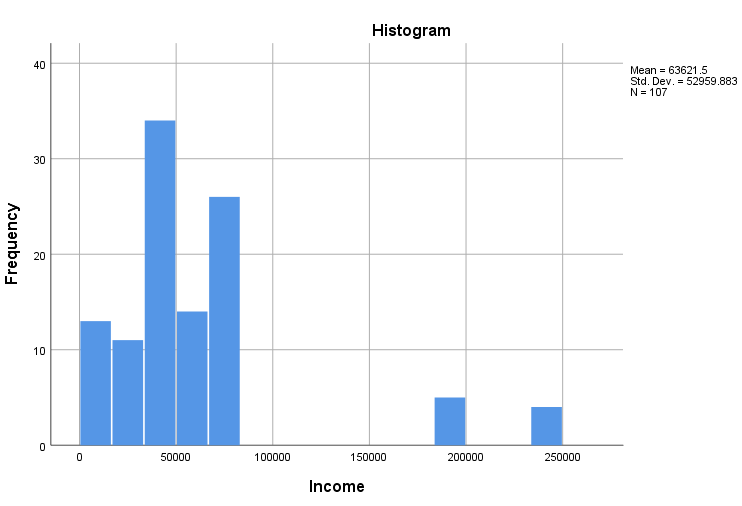
Description automatically generated



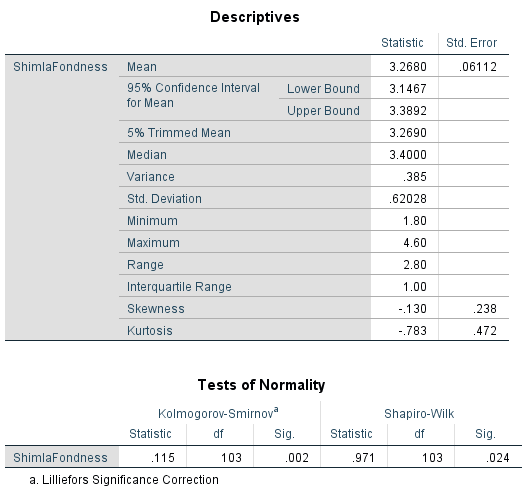
1. **Normality – fondness, income, like Shimla people** : In assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: graphically and numerically.

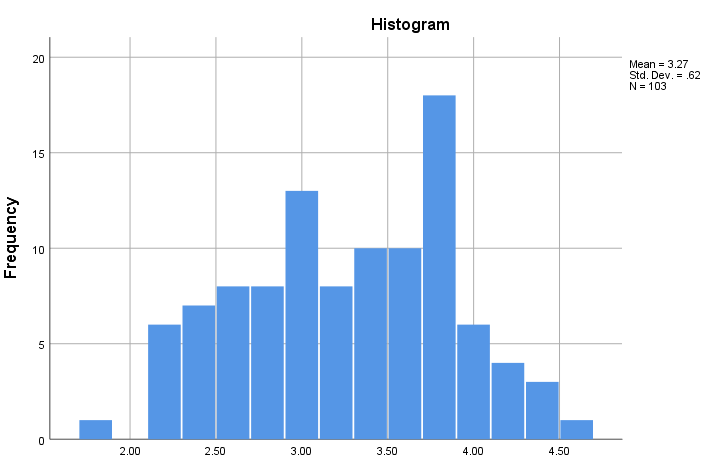
* 1. **Income:**

****

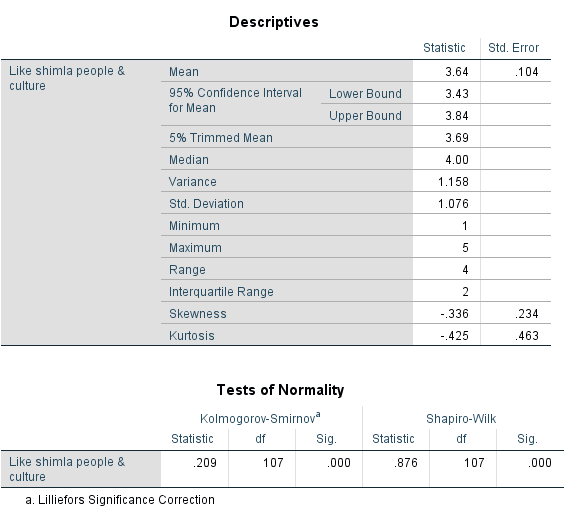
****

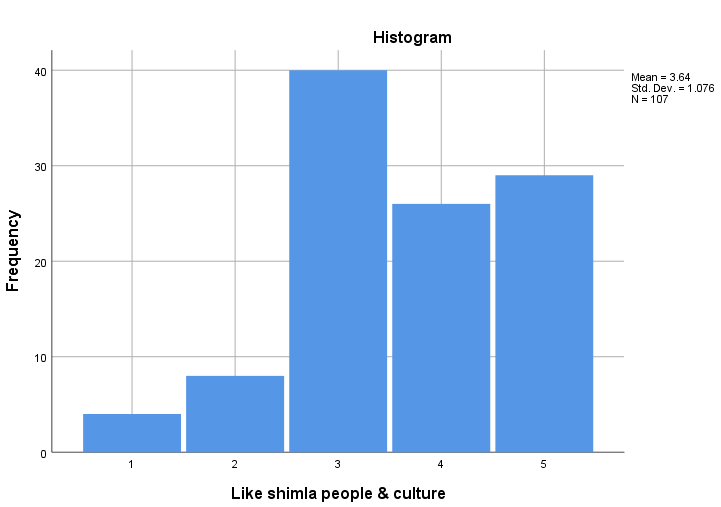
**b. Fondness:**

****



**c. Like Shimla People:**

****

****