# Introduction

The given dataset in the project is nutritional data with given attributes such as height, weight, BMI, etc which is used in the analysis of the eating habits causing obesity. The data-set has been taken from 207 random samples. It is a primitive dataset. The data-set is used in the analysis of factors influencing obesity in a significant way and how obesity is affecting nutrition.The data-set can be divided into two parts: Qualitative and Quantitative.

The following tests have been performed on the data set:

**Qualitative:**

1.Chi-squared Test

2.Factor Analysis Test

3.Reliability Test

**Quantitative**

1.Linear Regression2.F-test

3.T-test

# Quantitative Analysis

## Dependent Parameter: Weight

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | -131.8068847 | **Intercept** |
| w1 | 0.878242025 | **Height** |
| w2 | 1.648001084 | **BMI** |
| w3 | 0.432728042 | **VFAT** |
| w4 | -0.021956739 | **BP Systolic** |
| w5 | 0.016427966 | **BP Diastolic** |
| w6 | 0.077003483 | **Tummy** |
| w7 | 0.042658487 | **Hips** |
| w8 | 0.019403892 | **Waist** |

### RSquare Value: 0.97779768401941

### RSquare Adjusted: 0.974508452022286

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 5744.54587 | 8 | 718.0682 | 448.4293 | 2.1152 | Yes |
| Error | 86.47000 | 54 | 1.6013 |  |  |  |
| Total | 5831.01587 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Crticial value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Coefficient(w) | **Variable** | sw | T = w/sw | Decision |
| w0 | -131.8068847 | Intercept |  |  |  |
| w1 | 0.878242025 | Height | 0.03555 | 24.70606 | yes |
| w2 | 1.648001084 | BMI | 0.04878 | 33.78248 | yes |
| w3 | 0.432728042 | VFAT | 0.04109 | 10.53120 | yes |
| w4 | -0.021956739 | BP Systolic | 0.01099 | -1.99738 | no |
| w5 | 0.016427966 | BP Diastolic | 0.01660 | 0.98973 | no |
| w6 | 0.077003483 | Tummy | 0.02562 | 3.00549 | yes |
| w7 | 0.042658487 | Hips | 0.01930 | 2.21016 | yes |
| w8 | 0.019403892 | Waist | 0.02318 | 0.83723 | no |

### Conclusion from individual T-test:

Height, BMI, V. fat, Tummy, Hips are significant.

## Dependent Parameter: Height

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 142.608511036635 | **Intercept** |
| w1 | 0.833059526060548 | **Weight** |
| w2 | -1.29044887650336 | **BMI** |
| w3 | -0.458671512087079 | **VFAT** |
| w4 | 0.035601970742046 | **BP Systolic** |
| w5 | -0.040721363876095 | **BP Diastolic** |
| w6 | -0.047879589110741 | **Tummy** |
| w7 | 0.043337984932621 | **Hips** |
| w8 | -0.040051701824293 | **Waist** |

### RSquare Value: 0.669673220714483

### RSquare Adjusted: 0.620735920079592

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 844.44730 | 8 | 105.5559 | 40.0676 | 2.11522328 | Yes |
| Error | 142.26000 | 54 | 2.6344 |  |  |  |
| Total | 986.70730 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

### 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 142.608511036635 | Intercept |  |  |  |
| w1 | 0.833059526060548 | Weight | 0.02115 | 39.39031 | yes |
| w2 | -1.29044887650336 | BMI | 0.06257 | -20.62368 | yes |
| w3 | -0.458671512087079 | VFAT | 0.05270 | -8.70274 | yes |
| w4 | 0.035601970742046 | BP Systolic | 0.01410 | 2.52498 | yes |
| w5 | -0.040721363876095 | BP Diastolic | 0.02129 | -1.91269 | no |
| w6 | -0.047879589110741 | Tummy | 0.03286 | -1.45696 | no |
| w7 | 0.043337984932621 | Hips | 0.02476 | 1.75057 | no |
| w8 | -0.040051701824293 | Waist | 0.02973 | -1.34731 | no |

### 

### Conclusion from individual T-test:

Weight, BMI, V.Fat, BP Systolic are significant.

## Dependent Parameter: BMI

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 52.76564818 | **Intercept** |
| w1 | 0.389397053 | **Weight** |
| w2 | -0.335800547 | **Height** |
| w3 | 0.05737404 | **VFAT** |
| w4 | 0.003441228 | **BP Systolic** |
| w5 | -0.003624154 | **BP Diastolic** |
| w6 | -0.027785049 | **Tummy** |
| w7 | 0.0148731 | **Hips** |
| w8 | 0.011657467 | **Waist** |

### RSquare Value: 0.995039415769138

### RSquare Adjusted: 0.99173235961523

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 267.83429 | 8 | 33.4793 | 426.3871 | 2.11522328 | Yes |
| Error | 4.24000 | 54 | 0.0785 |  |  |  |
| Total | 272.07429 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 52.76564818 | **Intercept** |  |  |  |
| w1 | 0.389397053 | **Weight** | 0.01206 | 32.29881 | yes |
| w2 | -0.335800547 | **Height** | 0.02695 | -12.46133 | yes |
| w3 | 0.05737404 | **VFAT** | 0.03237 | 1.77217 | no |
| w4 | 0.003441228 | **BP Systolic** | 0.00849 | 0.40532 | no |
| w5 | -0.003624154 | **BP Diastolic** | 0.01281 | -0.28285 | no |
| w6 | -0.027785049 | **Tummy** | 0.02174 | -1.27784 | no |
| w7 | 0.0148731 | **Hips** | 0.01775 | 0.83813 | no |
| w8 | 0.011657467 | **Waist** | 0.02064 | 0.56485 | no |

### Conclusion from individual T-test:

Weight and Height are significant.

## 

## 

## Dependent Parameter: VFAT

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 22.48584851 |  |
| w1 | 0.280261385 | **Weight** |
| w2 | -0.313175783 | **Height** |
| w3 | 0.196263224 | **BMI** |
| w4 | 0.016718618 | **BP Systolic** |
| w5 | -0.000107881 | **BP Diastolic** |
| w6 | 0.039975126 | **Tummy** |
| w7 | 0.07151716 | **Hips** |
| w8 | -0.004404368 | **Waist** |

### RSquare Value: 0.82779819135346

### RSquare Adjusted: 0.802286812294713

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 781.70429 | 8 | 97.7130 | 34.8998 | 2.11522328 | Yes |
| Error | 151.19000 | 54 | 2.7998 |  |  |  |
| Total | 932.89429 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 22.48584851 |  |  |  |  |
| w1 | 0.280261385 | **Weight** | 0.02180 | 12.85454 | yes |
| w2 | -0.313175783 | **Height** | 0.04700 | -6.66268 | yes |
| w3 | 0.196263224 | **BMI** | 0.06451 | 3.04259 | yes |
| w4 | 0.016718618 | **BP Systolic** | 0.01454 | 1.15017 | no |
| w5 | -0.000107881 | **BP Diastolic** | 0.02195 | -0.00492 | no |
| w6 | 0.039975126 | **Tummy** | 0.03388 | 1.17996 | no |
| w7 | 0.07151716 | **Hips** | 0.02552 | 2.80220 | yes |
| w8 | -0.004404368 | **Waist** | 0.03065 | -0.14372 | no |

### Conclusion from individual T-test:

Weight, Height, BMI and Hips are significant.

## Dependent Parameter: BP Systolic

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | -69.5181221 |  |
| w1 | -0.465884708 | **Weight** |
| w2 | 0.796384575 | **Height** |
| w3 | 0.262436 | **BMI** |
| w4 | 0.547725275 | **V FAT** |
| w5 | 1.109902344 | **BP Diastolic** |
| w6 | -0.144356651 | **Tummy** |
| w7 | 0.08748796 | **Hips** |
| w8 | 0.012950619 | **Waist** |

### RSquare Value: 0.59818701594533

### RSquare Adjusted:0.53865916645574

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 7502.97429 | 8 | 937.8718 | 12.6383 | 2.11522328 | Yes |
| Error | 4007.26000 | 54 | 74.2085 |  |  |  |
| Total | 11510.23429 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | -69.5181221 |  |  |  |  |
| w1 | -0.465884708 | **Weight** | 0.11225 | -4.15059 | yes |
| w2 | 0.796384575 | **Height** | 0.24199 | 3.29095 | yes |
| w3 | 0.262436 | **BMI** | 0.33209 | 0.79025 | no |
| w4 | 0.547725275 | **V FAT** | 0.27972 | 1.95810 | no |
| w5 | 1.109902344 | **BP Diastolic** | 0.11300 | 9.82256 | yes |
| w6 | -0.144356651 | **Tummy** | 0.17442 | -0.82766 | no |
| w7 | 0.08748796 | **Hips** | 0.13139 | 0.66585 | no |
| w8 | 0.012950619 | **Waist** | 0.15777 | 0.08208 | no |

### Conclusion from individual T-test:

Weight, Height and BP diastolic are significant.

## Dependent Parameter: BP Diastolic

### Coefficients

|  |  |  |
| --- | --- | --- |
|  | **Coefficient(w)** | **Variable** |
| w0 | 62.67579877 | **Intercept** |
| w1 | 0.169597206 | **Weight** |
| w2 | -0.443195555 | **Height** |
| w3 | -0.167933661 | **BMI** |
| w4 | -0.00171961 | **V FAT** |
| w5 | 0.540019011 | **BP Systolic** |
| w6 | 0.218953476 | **Tummy** |
| w7 | -0.049807216 | **Hips** |
| w8 | 0.002335961 | **Waist** |

### RSquare Value: 0.797791199667911

### RSquare Adjusted: 0.767834340359453

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 4575.90238 | 8 | 571.9878 | 18.5293 | 2.11522328 | Yes |
| Error | 1666.95000 | 54 | 30.8694 |  |  |  |
| Total | 6242.85238 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 62.67579877 | **Intercept** |  |  |  |
| w1 | 0.169597206 | **Weight** | 0.07239 | 2.34268 | yes |
| w2 | -0.443195555 | **Height** | 0.15608 | -2.83959 | yes |
| w3 | -0.167933661 | **BMI** | 0.21419 | -0.78405 | no |
| w4 | -0.00171961 | **V FAT** | 0.18041 | -0.00953 | no |
| w5 | 0.540019011 | **BP Systolic** | 0.04827 | 11.18852 | yes |
| w6 | 0.218953476 | **Tummy** | 0.11249 | 1.94638 | no |
| w7 | -0.049807216 | **Hips** | 0.08474 | -0.58774 | no |
| w8 | 0.002335961 | **Waist** | 0.10176 | 0.02296 | no |

### Conclusion from individual T-test:

Weight, Height and BP systolic are significant.

## Dependent Parameter: Tummy

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 39.1688531 |  |
| w1 | 0.257781966 | **Weight** |
| w2 | -0.168978229 | **Height** |
| w3 | -0.448471381 | **BMI** |
| w4 | 0.206625432 | **V FAT** |
| w5 | -0.022775522 | **BP Systolic** |
| w6 | 0.071000123 | **BP Diastolic** |
| w7 | 0.181564754 | **Hips** |
| w8 | 0.501378187 | **Waist** |

### 

### RSquare Value: 0.965678274123577

### RSquare Adjusted: 0.960593573993737

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 2353.22000 | 8 | 294.1525 | 25.9453 | 2.11522328 | Yes |
| Error | 612.22000 | 54 | 11.3374 |  |  |  |
| Total | 2965.44000 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 39.1688531 |  |  |  |  |
| w1 | 0.257781966 | **Weight** | 0.04387 | 5.87562 | yes |
| w2 | -0.168978229 | **Height** | 0.09459 | -1.78648 | no |
| w3 | -0.448471381 | **BMI** | 0.12980 | -3.45500 | yes |
| w4 | 0.206625432 | **V FAT** | 0.10933 | 1.88984 | no |
| w5 | -0.022775522 | **BP Systolic** | 0.02925 | -0.77864 | no |
| w6 | 0.071000123 | **BP Diastolic** | 0.04417 | 1.60757 | no |
| w7 | 0.181564754 | **Hips** | 0.05136 | 3.53532 | yes |
| w8 | 0.501378187 | **Waist** | 0.06167 | 8.13019 | yes |

### Conclusion from individual T-test:

Weight, BMI, Hips and Waist are significant.

## Dependent Parameter: HIPS

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 16.63445208 |  |
| w1 | 0.14489848 | **Weight** |
| w2 | 0.155190553 | **Height** |
| w3 | 0.225738964 | **BMI** |
| w4 | 0.375076989 | **V FAT** |
| w5 | 0.014005417 | **BP Systolic** |
| w6 | -0.016387615 | **BP Diastolic** |
| w7 | 0.184224664 | **Tummy** |
| w8 | 0.26957973 | **Waist** |

### RSquare Value: 0.662730906353985

### RSquare Adjusted: 0.61276511470272

### 

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 2655.55222 | 8 | 331.9440 | 22.5889 | 2.11522328 | Yes |
| Error | 793.53000 | 54 | 14.6950 |  |  |  |
| Total | 3449.08222 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 16.63445208 |  |  |  |  |
| w1 | 0.14489848 | **Weight** | 0.04995 | 2.90093 | yes |
| w2 | 0.155190553 | **Height** | 0.10769 | 1.44114 | no |
| w3 | 0.225738964 | **BMI** | 0.14778 | 1.52754 | no |
| w4 | 0.375076989 | **V FAT** | 0.12448 | 3.01324 | yes |
| w5 | 0.014005417 | **BP Systolic** | 0.03330 | 0.42057 | no |
| w6 | -0.016387615 | **BP Diastolic** | 0.05028 | -0.32591 | no |
| w7 | 0.184224664 | **Tummy** | 0.07761 | 2.37358 | yes |
| w8 | 0.26957973 | **Waist** | 0.07021 | 3.83967 | yes |

### Conclusion from individual T-test:

Weight, V Fat, Tummy and Waist are significant.

## Dependent Parameter: Waist

### Coefficients

|  |  |  |
| --- | --- | --- |
| w0 | 39.38381534 |  |
| w1 | 0.066217551 | **Weight** |
| w2 | -0.144093175 | **Height** |
| w3 | 0.134844925 | **BMI** |
| w4 | -0.023207034 | **V FAT** |
| w5 | 0.00208288 | **BP Systolic** |
| w6 | 0.000772174 | **BP Diastolic** |
| w7 | 0.511101944 | **Tummy** |
| w8 | 0.270840187 | **HIPS** |

### RSquare Value: 0.736338924297107

### RSquare Adjusted: 0.697278024192974

### ANOVA Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of variation | Sum of Squares | DOF | Mean Square | F calc | F tab | Reject Null |
| Regression | 2130.91810 | 8 | 266.3648 | 20.6342 | 2.11522328 | Yes |
| Error | 697.08000 | 54 | 12.9089 |  |  |  |
| Total | 2827.99810 | 62 |  |  |  |  |

### T-Test:

DOF = 54

Critical value of t(talpha/2) = 2.005

If t > talpha/2 Reject Null Hypothesis, The variable is significant.

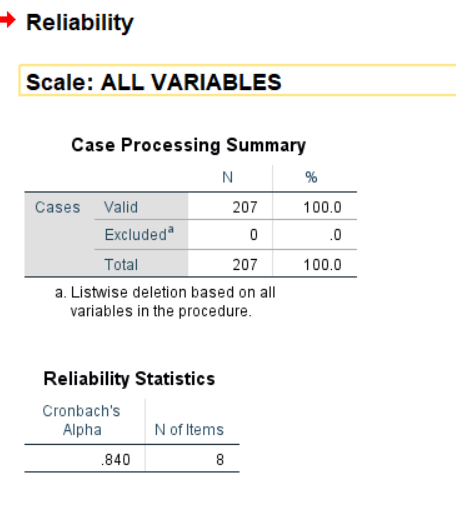
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coefficient(w)** | **Variable** | **sw** | **T = w/sw** | **Decision** |
| w0 | 39.38381534 |  |  |  |  |
| w1 | 0.066217551 | **Weight** | 0.04682 | 1.41445 | no |
| w2 | -0.144093175 | **Height** | 0.10093 | -1.42766 | no |
| w3 | 0.134844925 | **BMI** | 0.13851 | 0.97355 | no |
| w4 | -0.023207034 | **V FAT** | 0.11667 | -0.19892 | no |
| w5 | 0.00208288 | **BP Systolic** | 0.03121 | 0.06673 | no |
| w6 | 0.000772174 | **BP Diastolic** | 0.04713 | 0.01638 | no |
| w7 | 0.511101944 | **Tummy** | 0.07275 | 7.02592 | yes |
| w8 | 0.270840187 | **HIPS** | 0.05480 | 4.94223 | yes |

### Conclusion from individual T-test:

Tummy and Hips are significant.

**Reliability Analysis**

Cronbach’s alpha is a measure of internal consistency, that is, how closely related a set of items are as a group. It is considered to be a measure of scale reliability.



The alpha coefficient for the eight items(Ordinal attributes) is .840, suggesting that the items have relatively high internal consistency. (Note that a reliability coefficient of .70or higher is considered “acceptable” in most social science research situations.)

# **Qualitative Analysis**

## **Chi-squared test**

### **Occupation**

Number of Occupations=24

Degree of freedom=23

Chi-square alpha value=35.172

For a null Hypothesis to be accepted:

𝝌**2** <= 𝝌𝛂**2**

**The Chi-square test has been performed with following attributes:**

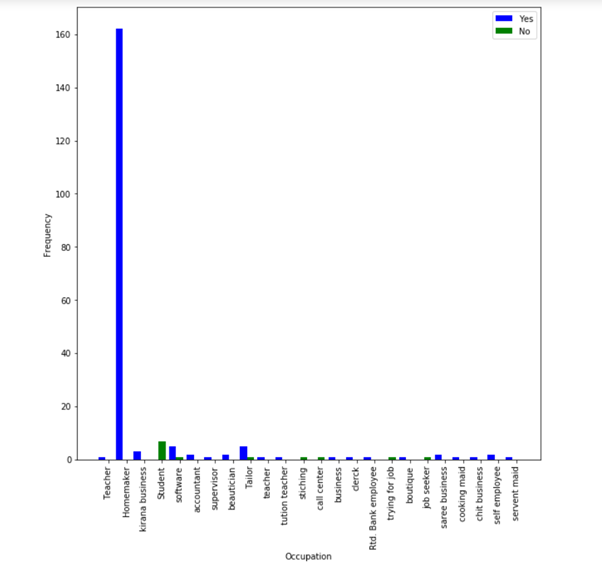
1. **Marital status**

H0:Occupation is not associated with the marital status of a person.

Ha:Occupation is associated with the marital status of a person.

𝝌**2** =178.68318 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis



Marital Status vs Occupation

**Conclusion 1**: Occupation is associated with marital status

**Conclusion 2**: Most of the people have the occupation as homeworkers and all of them are married.

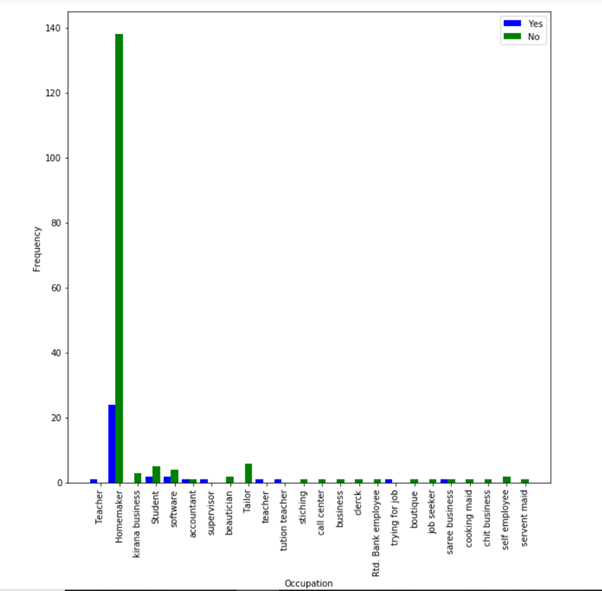
**(b)Family History of Heart Problem**

H0:Occupation is not associated with a person having a family history of heart problem.

Ha:Occupation is associated a person having a family history of heart problem.

𝝌**2** =34.704674 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.



Family History of heart problem vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

Heart problem.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of heart diseases.

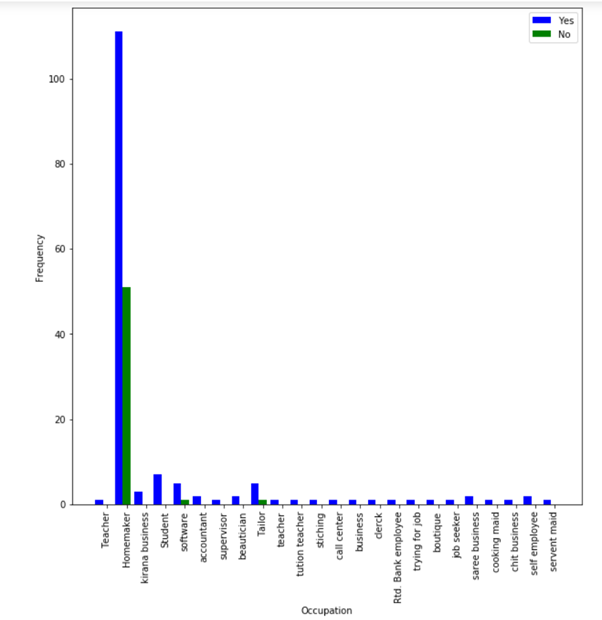
**(c)Family History of Hypertension**

H0:Occupation is not associated with a person having a family history of Hypertension.

Ha:Occupation is associated a person having a family history of Hypertension.

𝝌**2** =14.79839500 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.



Family History of Hypertension vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

Hypertension.

**Conclusion 2**: Most of the people are homeworkers and most of them have a

family history of Hypertension.

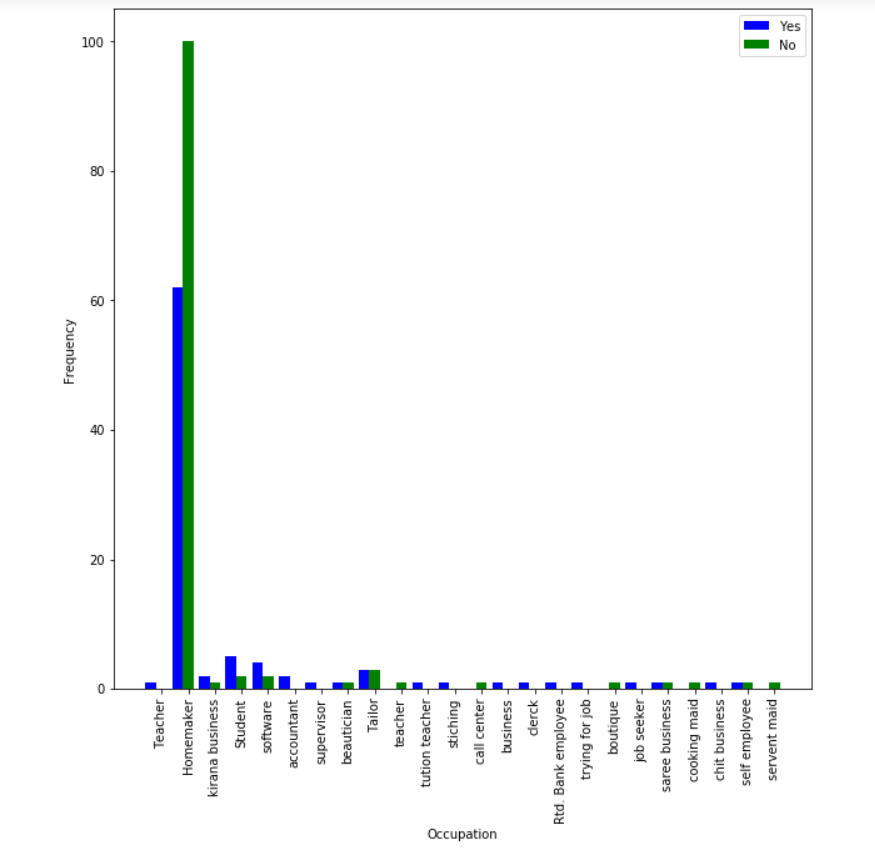
**(d)Family History of Diabetes**

H0:Occupation is not associated with a person having a family history of Diabetes.

Ha:Occupation is associated a person having a family history of Diabetes.

𝝌**2** =25.5526849 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.



Family History of Diabetes vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

Diabetes.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of Diabetes.

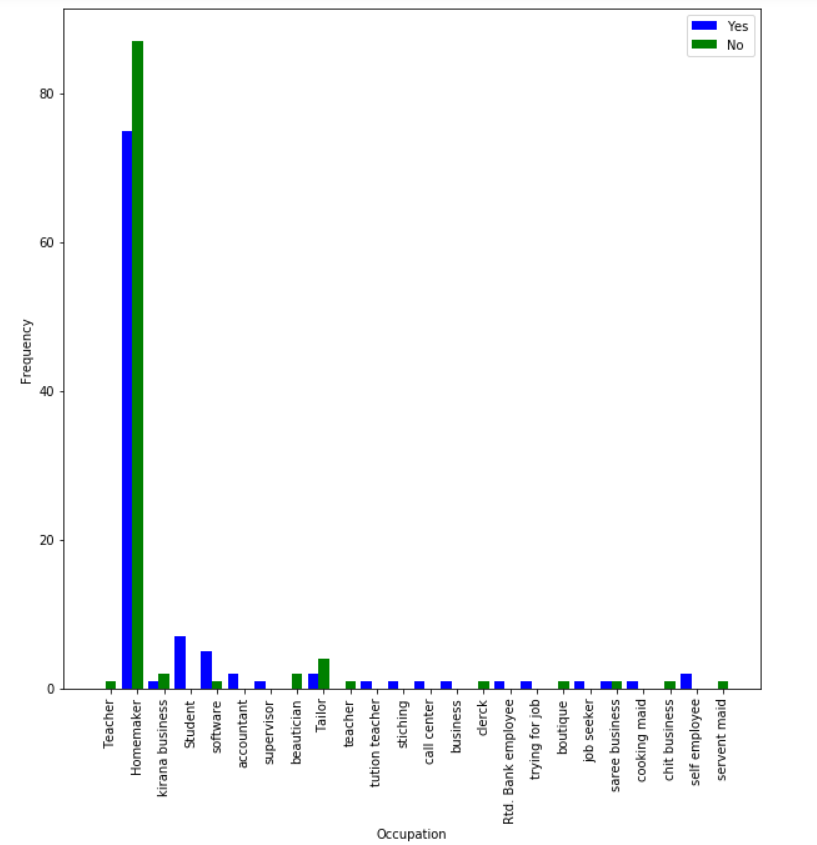
**(e)Family History of Obesity**

H0:Occupation is not associated with a person having a family history of Obestiy.

Ha:Occupation is associated a person having a family history of Obesity.

𝝌**2** =32.551484 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.

****

Family History of Obesity vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

Obesity.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of Obesity.

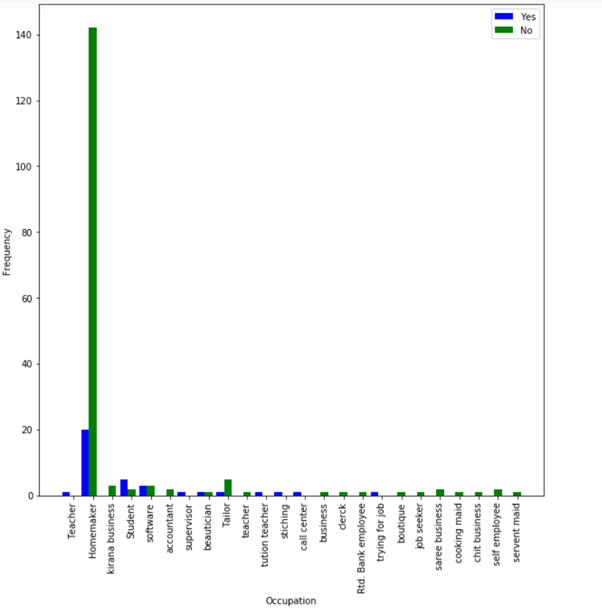
**(f)Family History of Pcos**

H0:Occupation is not associated with a person having a family history of Pcos.

Ha:Occupation is associated a person having a family history of Pcos.

𝝌**2** = 55.310858 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.



Family History of Pcos vs Occupation

**Conclusion 1**: Occupation is associated with a person having a family history of

Pcos.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of Pcos.

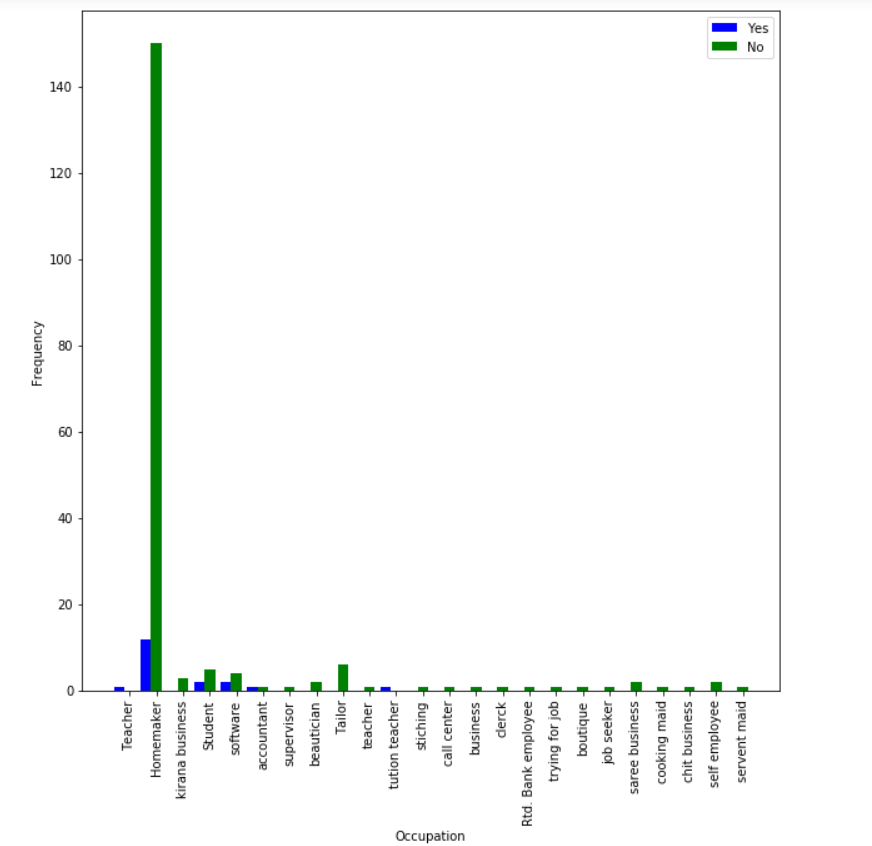
**(g)Family History of Hypothyroid**

H0:Occupation is not associated with a person having a family history of Hypothyroid.

Ha:Occupation is associated a person having a family history of Hypothyroid.

𝝌**2** =34.5841665 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.



Family History of Hypothyroid vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

Hypothyroid.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of Hypothyroid.

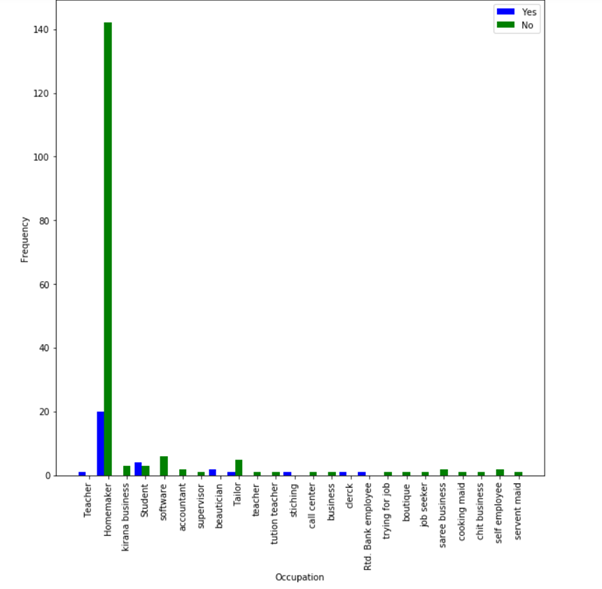
**(h)Family History of Cancer**

H0:Occupation is not associated with a person having a family history of Cancer.

Ha:Occupation is associated a person having a family history of Cancer.

𝝌**2** = 49.31251309 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.



Family History of Cancer vs Occupation

**Conclusion 1**: Occupation is associated with a person having a family history of

Cancer.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have a

family history of Cancer.

**(i) Family History of any other disease**

H0:Occupation is not associated with a person having a family history of any other

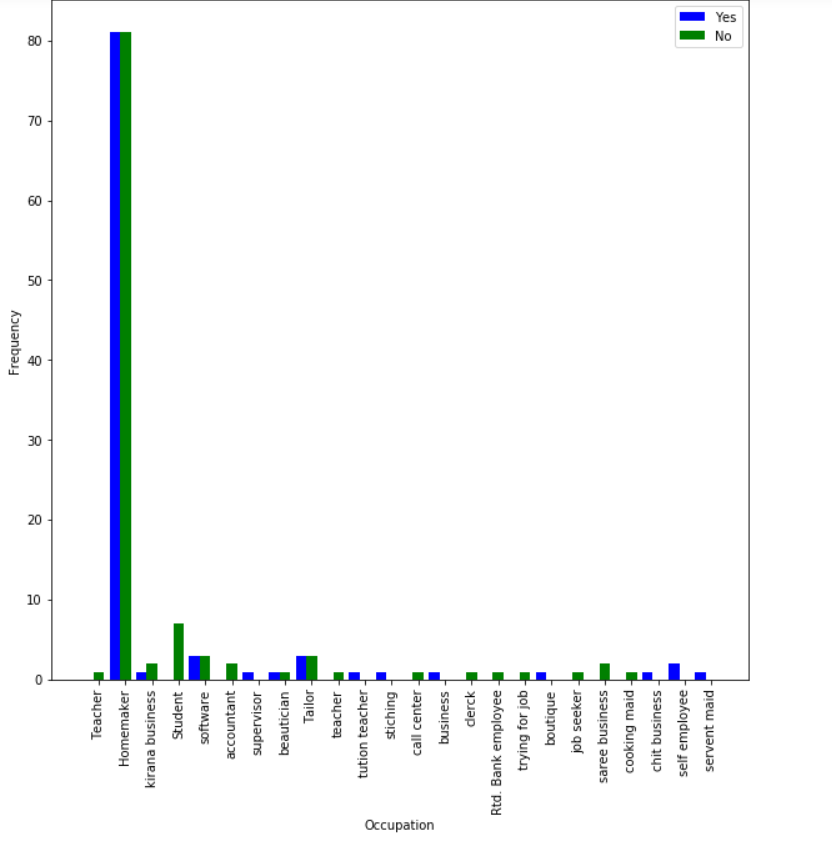
Disease.

Ha:Occupation is associated a person having a family history of any other

disease.

𝝌**2** =27.8273731 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we accept the null Hypothesis.



Family History of any other disease. vs Occupation

**Conclusion 1**: Occupation is not associated with a person having a family history of

any other disease.

**Conclusion 2**: Most of the people are homeworkers and the of them do not have a

family history of any other disease

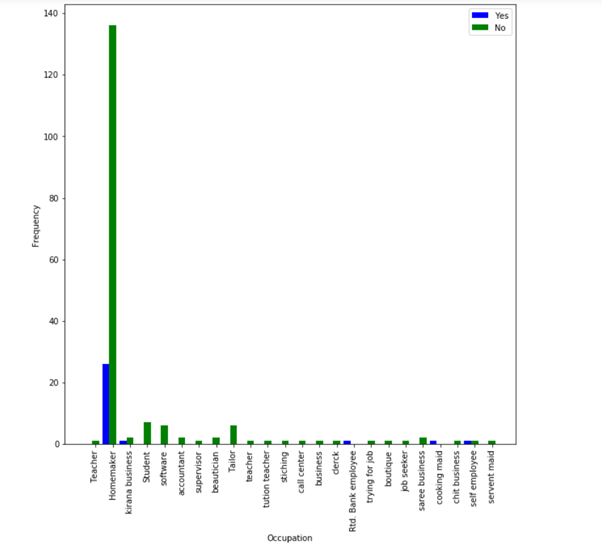
**(j) Person having Diabetes**

H0:Occupation is not associated with a person having Diabetes.

Ha:Occupation is associated a person having Diabetes.

𝝌**2** = 21.4515065 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person having Diabetes vs Occupation

**Conclusion 1**: Occupation is not associated with a person having Diabetes.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Diabetes.

.

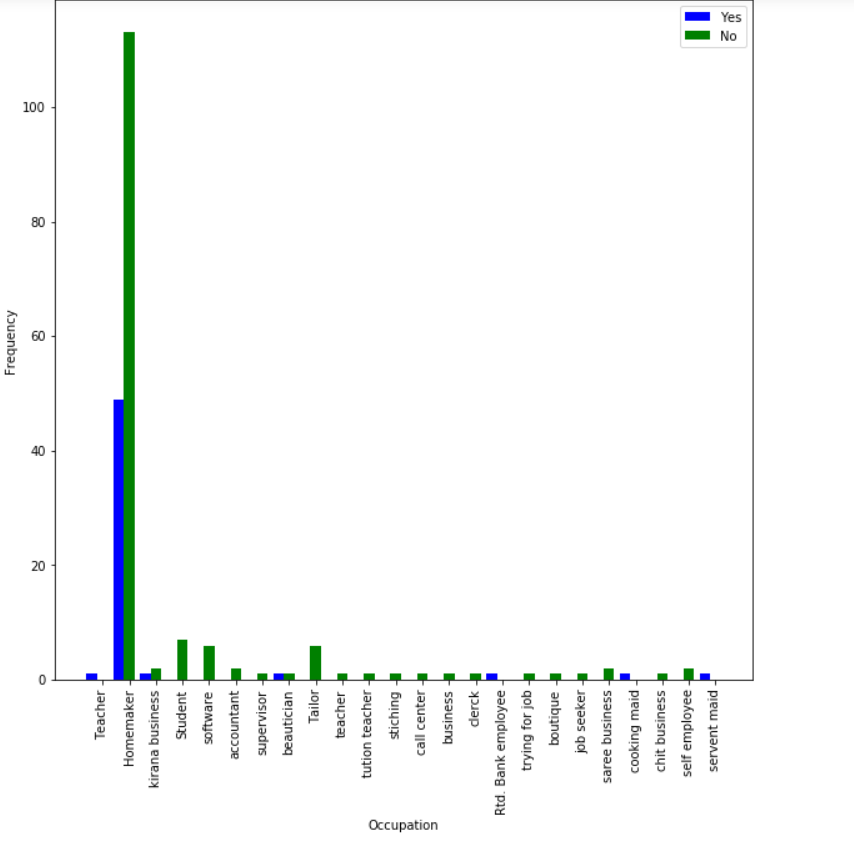
**(k) Person having HTN**

H0:Occupation is not associated with a person having HTN.

Ha:Occupation is associated a person having HTN.

𝝌**2** = 25.8364832 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having HTN vs Occupation

**Conclusion 1**: Occupation is not associated with a person having HTN.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have HTN.

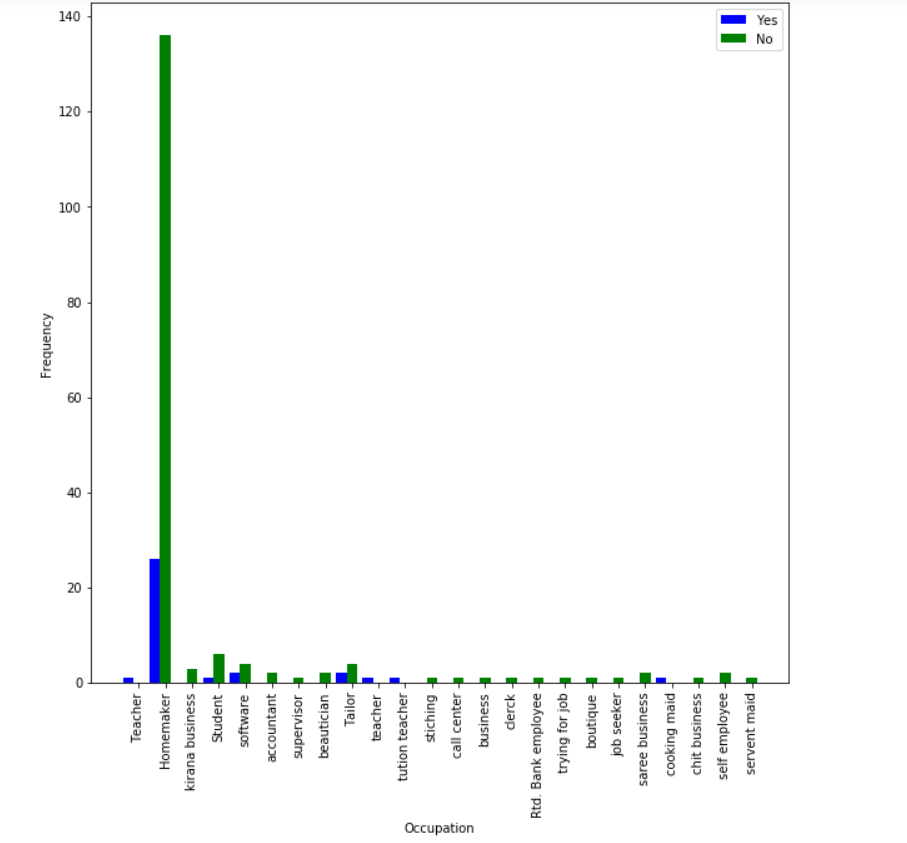
**(l) Person having Pcos**

H0:Occupation is not associated with a person having Pcos.

Ha:Occupation is associated a person having Pcos.

𝝌**2** = 26.55752254 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Pcos vs Occupation

**Conclusion 1**: Occupation is not associated with a person having Pcos.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Pcos.

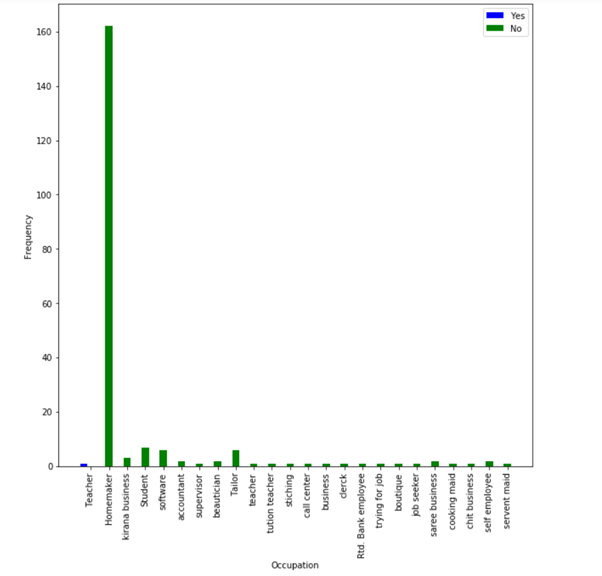
**(m) Person having CVD**

H0:Occupation is not associated with a person having CVD.

Ha:Occupation is associated a person having CVD.

𝝌**2** = 206.999999 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having CVD vs Occupation

**Conclusion 1**: Occupation is associated with a person having CVD.

**Conclusion 2**: Most of the people are homeworkers and all of them do not have CVD.

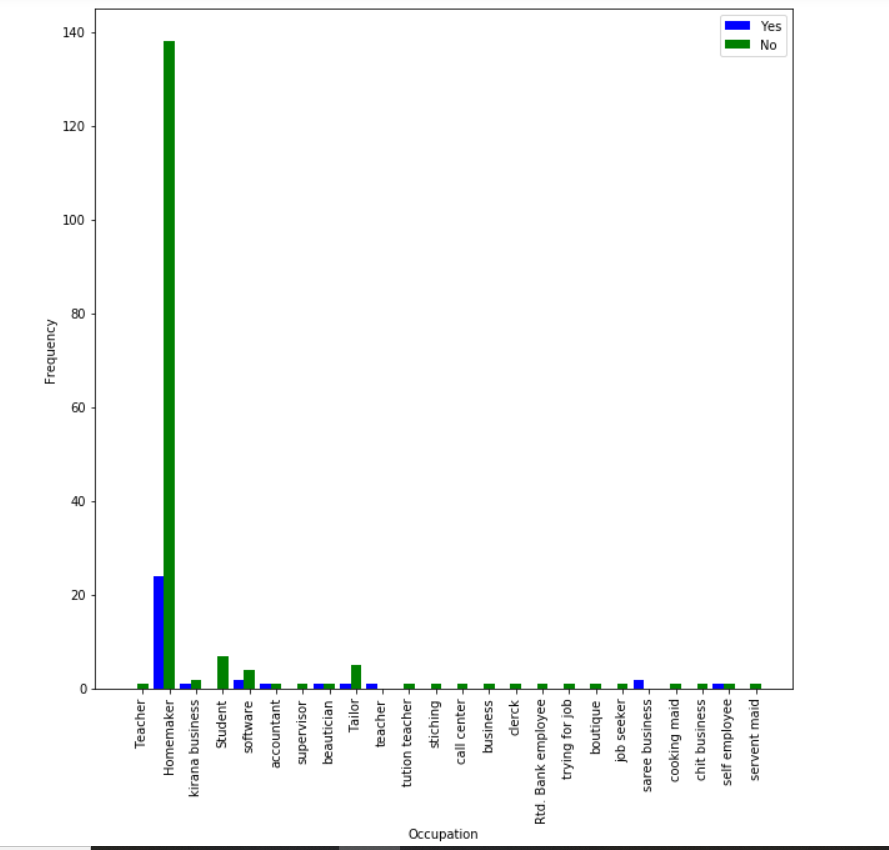
**(o) Person having Hypothyroid**

H0:Occupation is not associated with a person having Hypothyroid.

Ha:Occupation is associated a person having Hypothyroid.

𝝌**2** = 26.49965997 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person having Hypothyroid vs Occupation

**Conclusion 1**: Occupation is not associated with a person having Hypothyroid.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Hypothyroid.

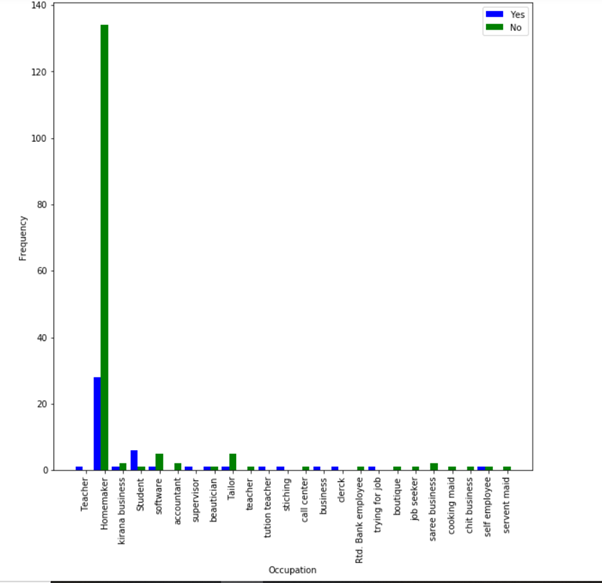
**(p)Person having Menstrual Irregularities**

H0:Occupation is not associated with a person having Menstrual Irregularities.

Ha:Occupation is associated a person having Menstrual Irregularities.

𝝌**2** = 48.7551024 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having Menstrual Irregularities vs Occupation

**Conclusion 1**: Occupation is associated with a person having Menstrual Irregularities.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Menstrual Irregularities.

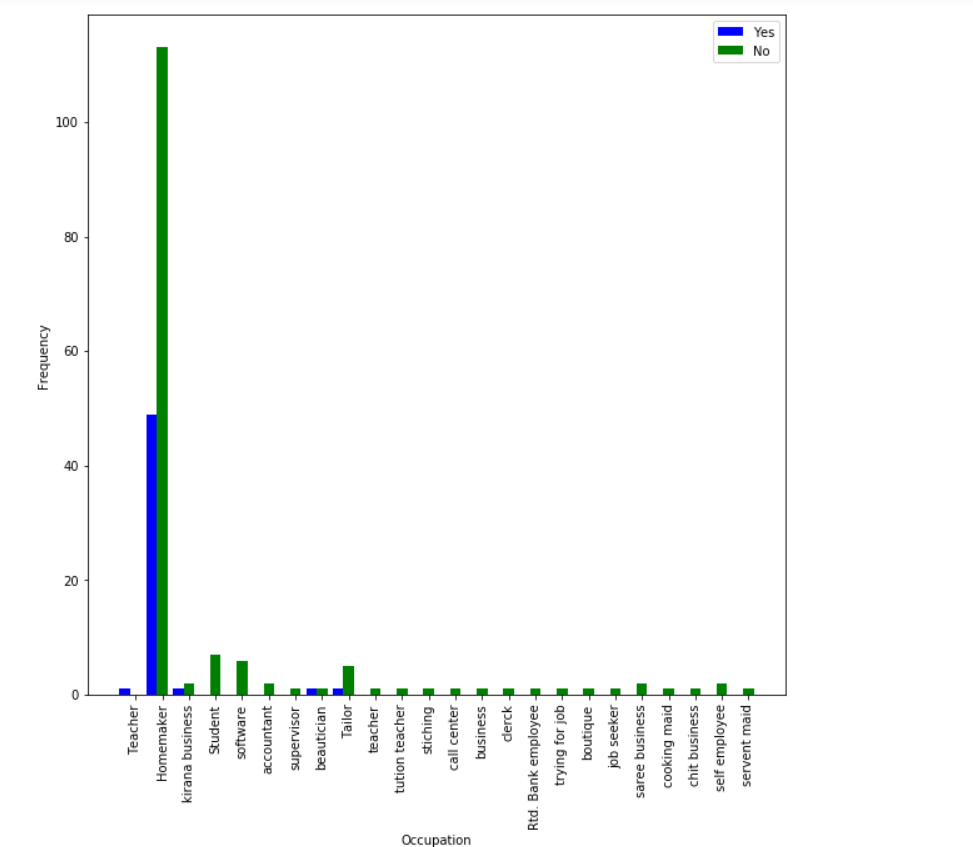
**(q) Person having Arthritis**

H0:Occupation is not associated with a person having Arthritis.

Ha:Occupation is associated a person having Arthritis.

𝝌**2** =17.0668341 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Arthritis vs Occupation

**Conclusion 1**: Occupation is not associated with a person having Arthritis.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Arthritis.

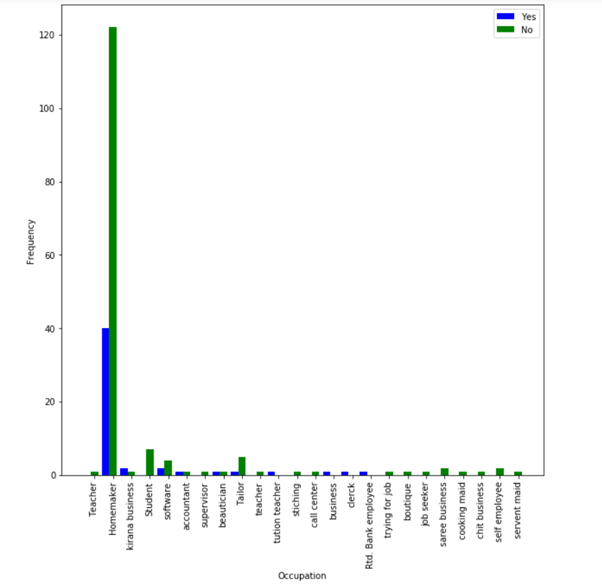
**(r)Person having Vertigo**

H0:Occupation is not associated with a person having Vertigo.

Ha:Occupation is associated a person having Vertigo.

𝝌**2** = 24.1173328 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Vertigo vs Occupation

**Conclusion 1**: Occupation is not associated with a person having Vertigo.

**Conclusion 2**: Most of the people are homeworkers and most of them do not have Vertigo.

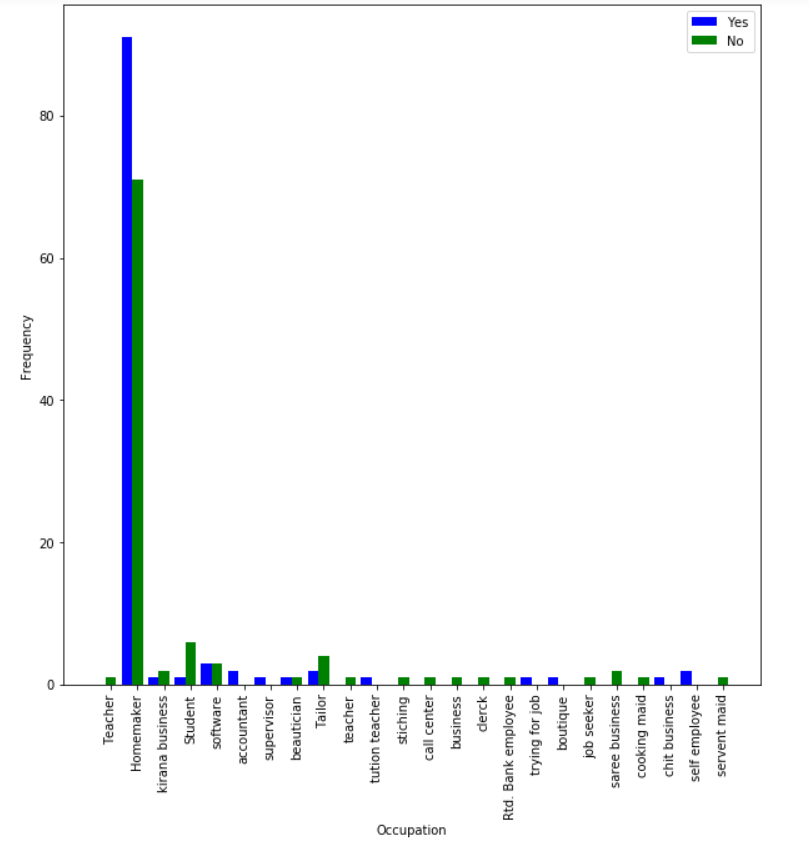
**(s) Person having any other medical problem**

H0:Occupation is not associated with a person having any other medical problem.

Ha:Occupation is associated a person having any other medical problem.

𝝌**2** = 27.70162604 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having any other medical problem vs Occupation

**Conclusion 1**: Occupation is not associated with a person having any other medical problem.

**Conclusion 2**: Most of the people are homeworkers and most of them have any other medical problem.

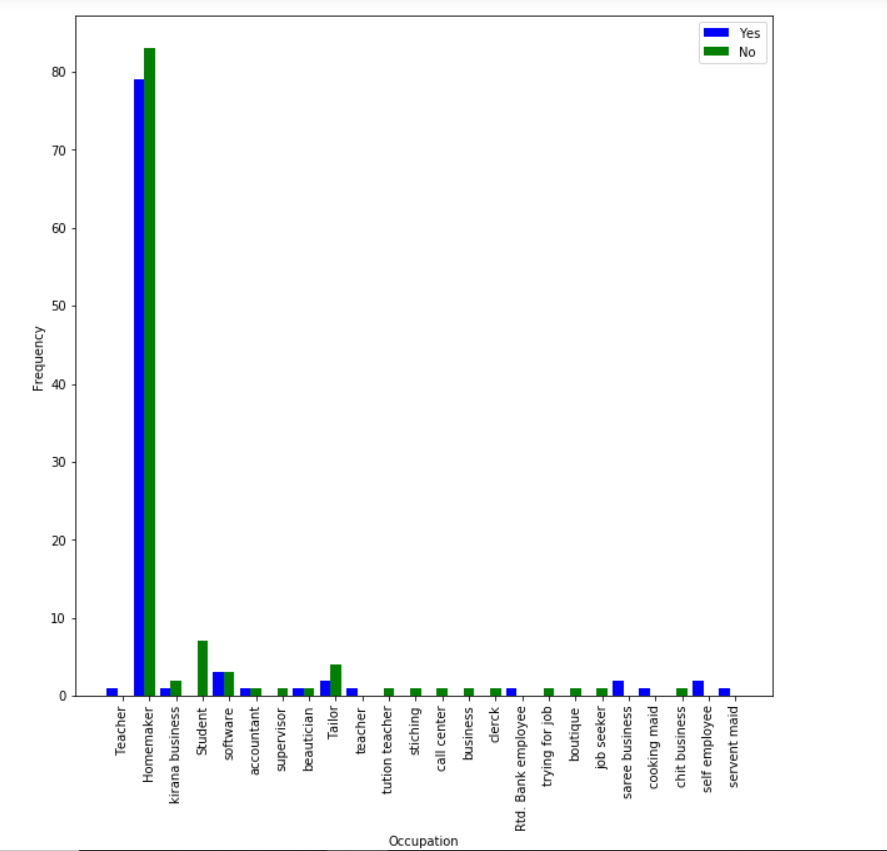
**(t) Person using any medication/supplement**

H0:Occupation is not associated with a person using any medication/supplement.

Ha:Occupation is associated a person using any medication/supplement.

𝝌**2** = 26.149117867 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person using any medication/supplement. vs Occupation

**Conclusion 1**: Occupation is not associated with a person using any

medication/supplement.

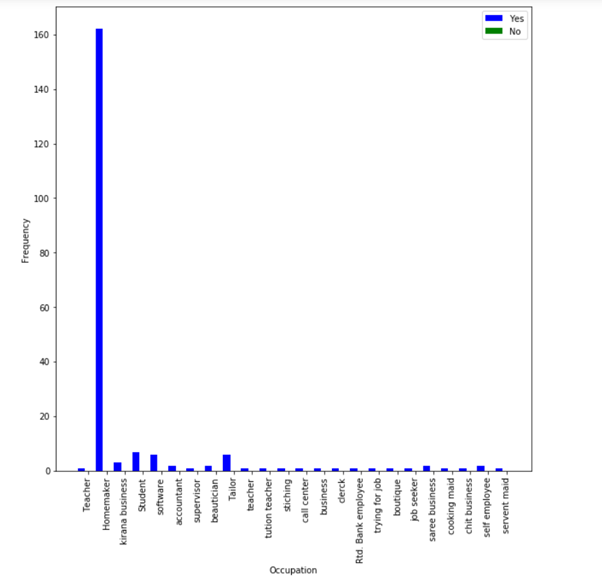
**Conclusion 2**: Most of the people are homeworkers and most of them do not use any medication/supplement.

**(u)Person has interest in Weight Management**

H0:Occupation is not associated with a person having Vertigo.

Ha:Occupation is associated a person having Vertigo.

𝝌**2** = NaN and 𝝌𝛂**2** = 35.172

****

Person willing to have weight management vs Occupation

**Conclusion 1**: Occupation is not associated with a person willing to have weight

management.

**Conclusion 2**: Weight Management is always Yes.

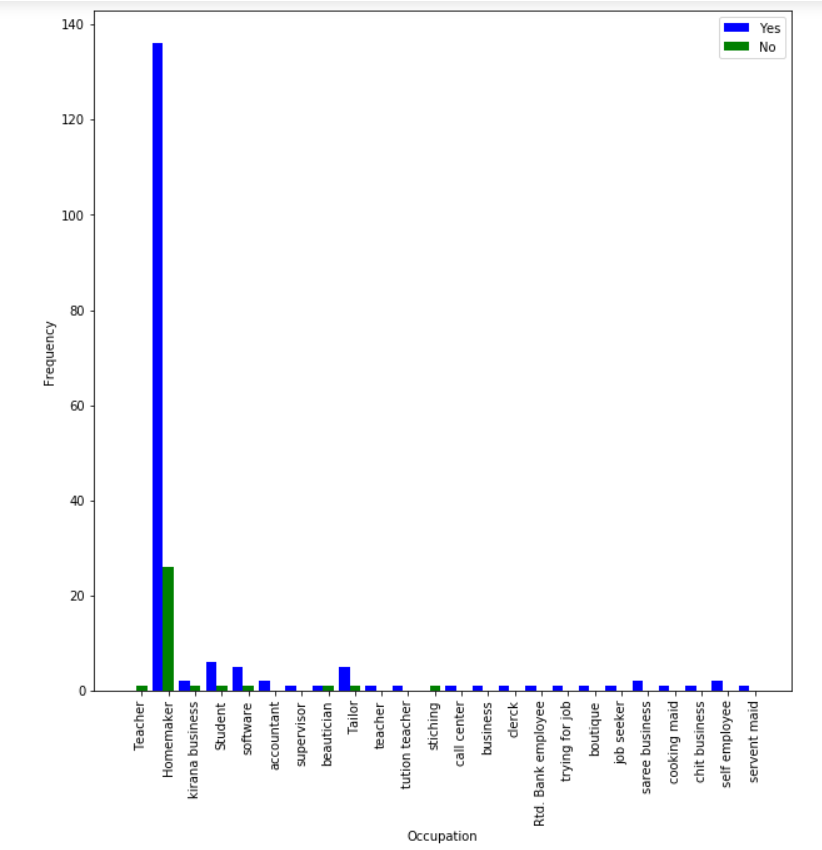
**(v)Person associated with improved fitness**

H0:Occupation is not associated with a person associated with improved fitness.

Ha:Occupation is associated a person associated with improved fitness.

𝝌**2** = 16.5576359 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person associated with improved fitness. vs Occupation

**Conclusion 1**: Occupation is not associated with a person associated with improved

fitness.

**Conclusion 2**: Most of the people are homeworkers and most of them are associated with improved fitness.

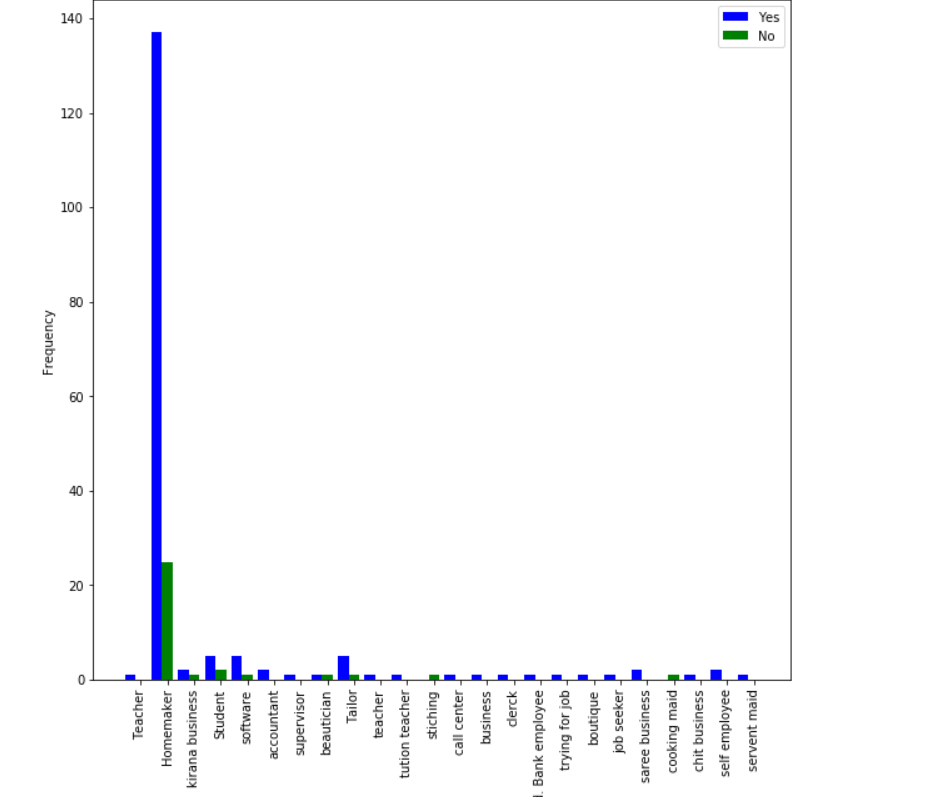
**(w)Person associated with Back Strengthening**

H0:Occupation is not associated with a person associated with Back Strengthening.

Ha:Occupation is associated a person associated with Back Strengthening.

𝝌**2** = 17.42653132 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person associated with Back Strengthening. vs Occupation

**Conclusion 1**: Occupation is not associated with a person associated with Back

Strengthening

**Conclusion 2**: Most of the people are homeworkers and most of them are associated with Back Strengthening.

**(x)Person willing to have Nutrition and Disease prevention**

H0:Occupation is not associated with a person willing to have Nutrition and Disease

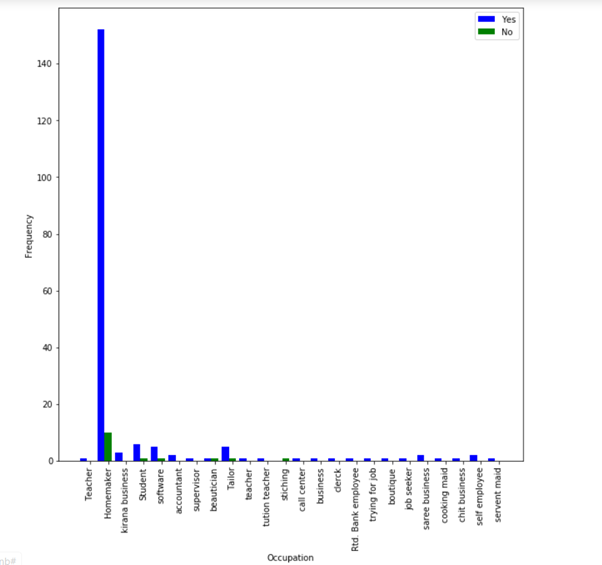
Prevention.

Ha:Occupation is associated with a person willing to have Nutrition and Disease

Prevention.

𝝌**2** = 22.414617 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person person willing to have Nutrition and Disease prevention vs Occupation

**Conclusion 1**: Occupation is not associated with a person willing to have Nutrition and

Disease prevention

**Conclusion 2**: Most of the people are homeworkers and most of them are willing to have Nutrition and Disease prevention.

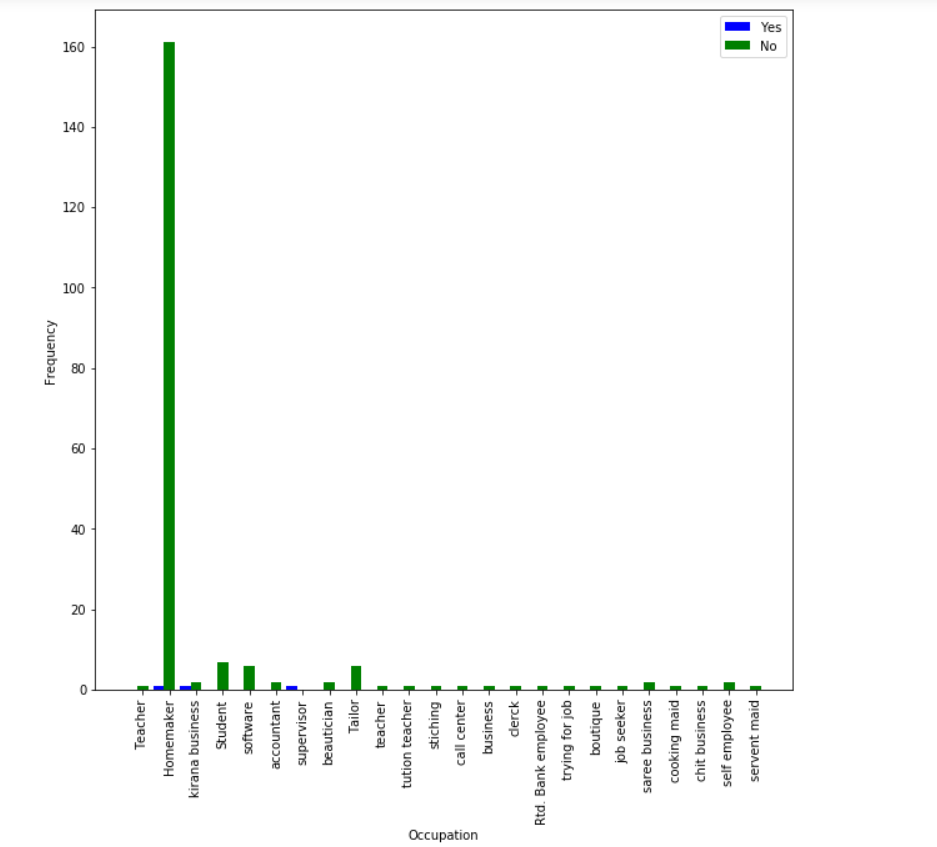
**(y)Person being Vegetarian**

H0:Occupation is not associated with a person being Vegetarian.

Ha:Occupation is associated a person being Vegetarian.

𝝌**2** = 90.74101307 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person being Vegetarian vs Occupation

**Conclusion 1**: Occupation is associated with a person being Vegetarian

**Conclusion 2**: Most of the people are homeworkers and most of them are not Vegetarian.

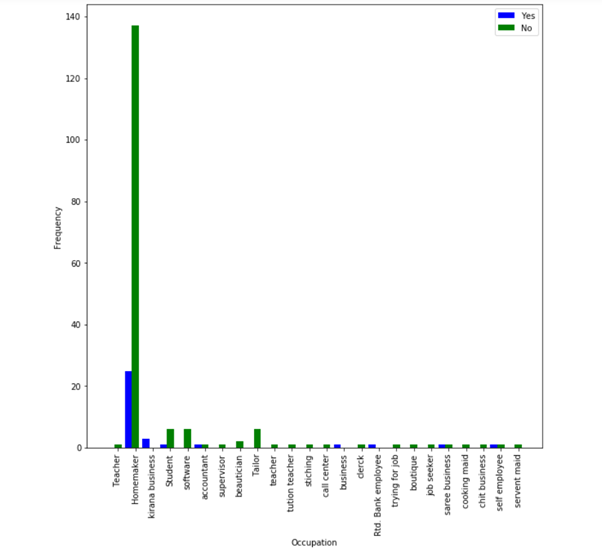
**(z)Person being Lacto-Vegetarian**

H0:Occupation is not associated with a person being Lacto-Vegetarian.

Ha:Occupation is associated a person being Lacto-Vegetarian.

𝝌**2** = 35.8143974 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person being Lacto-Vegetarian vs Occupation

**Conclusion 1**: Occupation is associated with a person being Lacto-Vegetarian

**Conclusion 2**: Most of the people are homeworkers and most of them are not Lacto-Vegetarian.

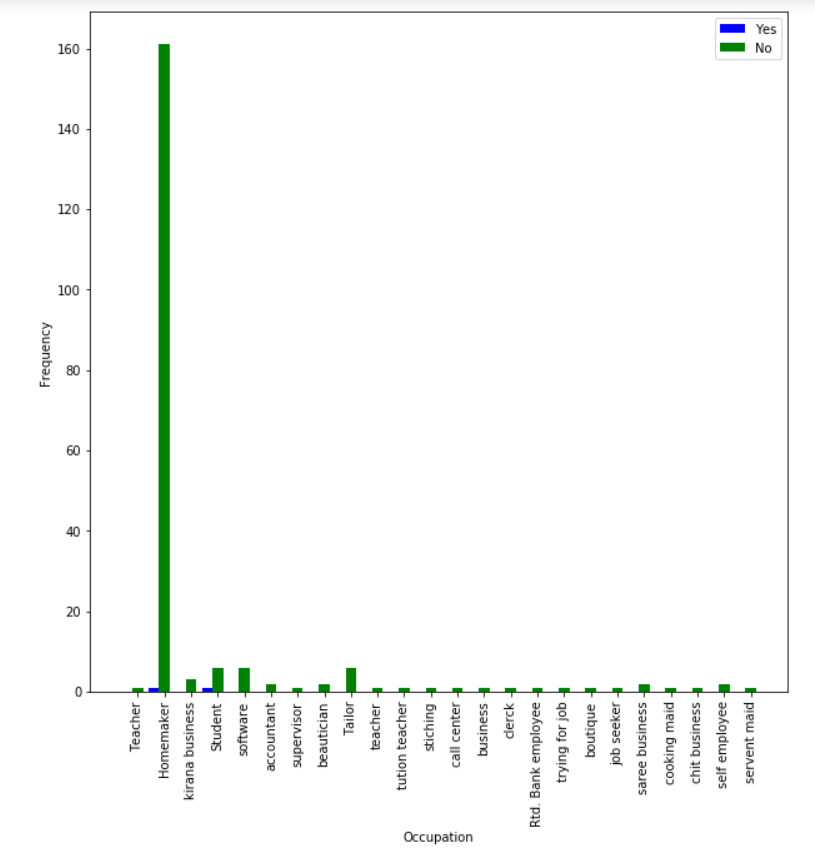
**(aa)Person being Ovo Vegetarian**

H0:Occupation is not associated with a person being Ovo Vegetarian.

Ha:Occupation is associated a person being Ovo Vegetarian.

𝝌**2** = 13.5555749 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person being Ovo Vegetarian vs Occupation

**Conclusion 1**: Occupation is not associated with a person being Ovo Vegetarian

**Conclusion 2**: Most of the people are homeworkers and most of them are not Ovo Vegetarian.

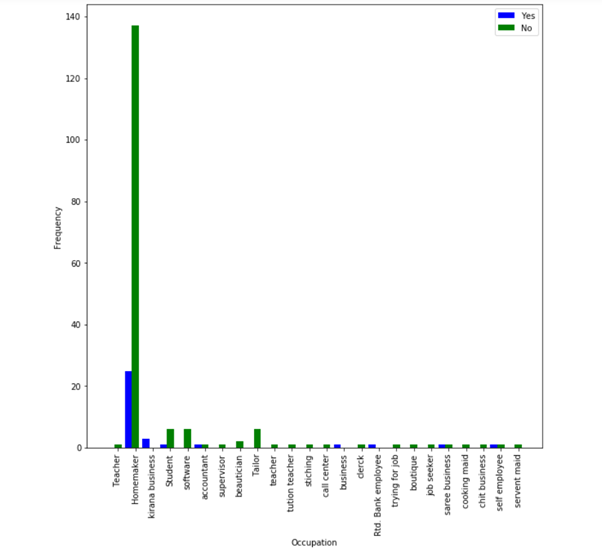
**(ab)Person being Lacto-Vegetarian**

H0:Occupation is not associated with a person being Lacto-Ovo-Vegetarian.

Ha:Occupation is associated a person being Lacto-Ovo-Vegetarian.

𝝌**2** = 26.5735084 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person being Lacto-Ovo- Vegetarian vs Occupation

**Conclusion 1**: Occupation is not associated with a person being Lacto-Vegetarian

**Conclusion 2**: Most of the people are homeworkers and most of them are not Lacto-Ovo-Vegetarian.

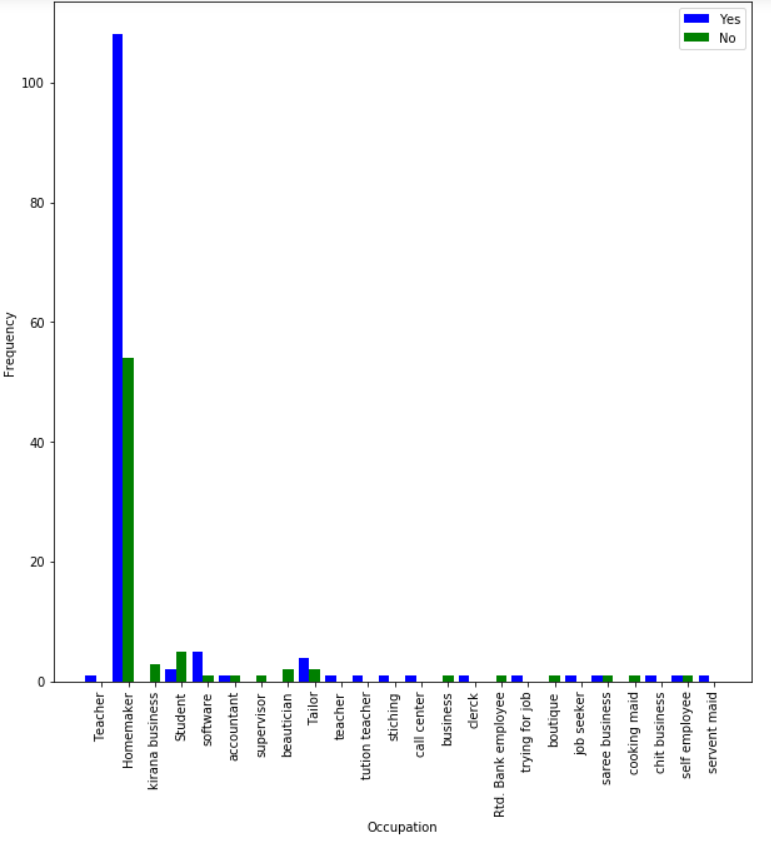
**(ac)Person being Non-Vegetarian**

H0:Occupation is not associated with a person being Non-Vegetarian.

Ha:Occupation is associated a person being Non-Vegetarian.

𝝌**2** = 29.13233766 and 𝝌𝛂**2** = 35.172

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.



Person being Non- Vegetarian vs Occupation

**Conclusion 1**: Occupation is not associated with a person being Non-Vegetarian

**Conclusion 2**: Most of the people are homeworkers and most of them are Non-Vegetarian.

**Occupation of the person was found to be associated with:**

1.Marital Status of Person

2.Family History with Pcos

3.Family History with Cancer

4.Person having CVD

5.Person having Menstrual Irregularities

6.Person being Vegetarian

7.Person being Lacto-Vegetarian

### 

### **Number of Children**

Maximum number of children=8

Degree of freedom=7

Chi-square alpha value=14.067

For a null Hypothesis to be accepted:

𝝌**2** <= 𝝌𝛂**2**

**The Chi-square test has been performed with following attributes:**

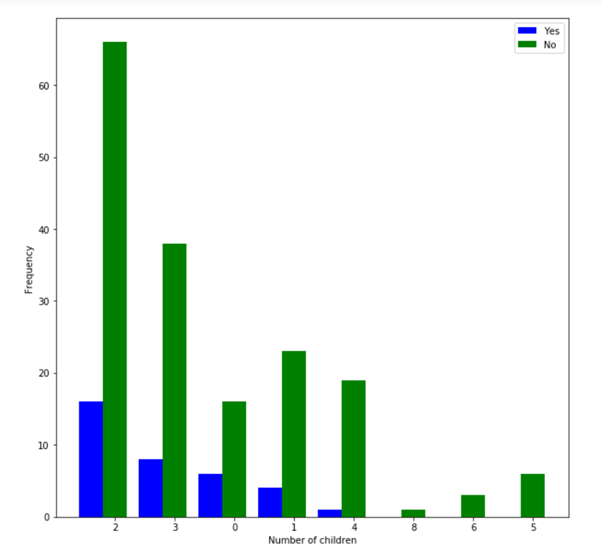
1. **Family History of Heart Problem**

H0:Number of Children a person has is not associated with a person having a family history of Heart Problem.

Ha:Number of children a person has is associated with a person having a family history of Heart Problem.

𝝌**2** =6.2233360 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis



Family History of Heart Problem vs Number of children

**Conclusion 1**: Number of children a person has is not associated with a person having a family history of Heart Problem.

**Conclusion 2**: Most of the people have 2 children and do not have a family history of heart problem.

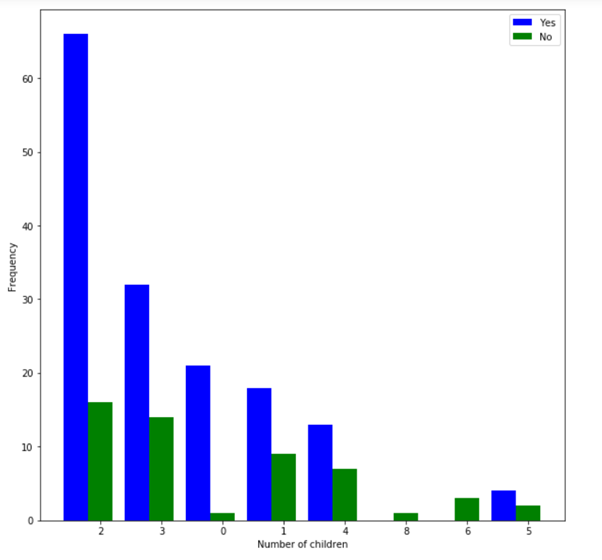
1. **Family History of Hypertension**

H0:Number of Children a person has is not associated with a person having a family history of Hypertension.

Ha:Number of children a person has is associated with a person having a family history of Heart Problem.

𝝌**2** = 20.86746136 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.



Family History of Heart Problem vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having a family

history of Heart Problem.

**Conclusion 2**: Most of the people have 2 children and have a family history of Hypertension.

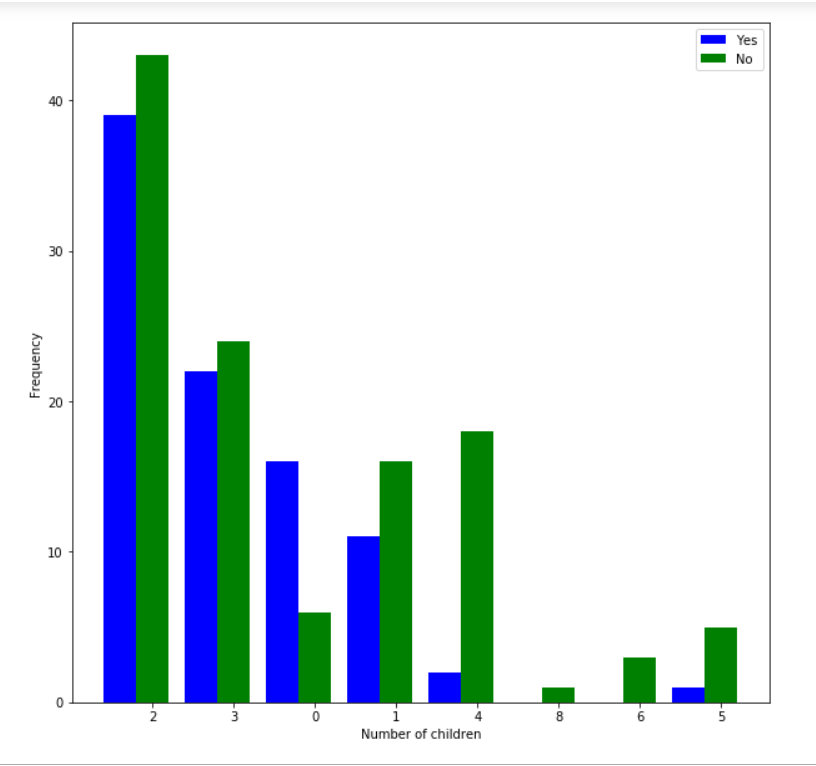
1. **Family History of Diabetes**

H0:Number of Children a person has is not associated with a person having a family history of Diabetes.

Ha:Number of children a person has is associated with a person having a family history of Diabetes.

𝝌**2** =22.52941839and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Family History of Diabetes vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having a family

history of Diabetes.

`

**Conclusion 2**: Most of the people have 2 children and do not have a family history of Diabetes.

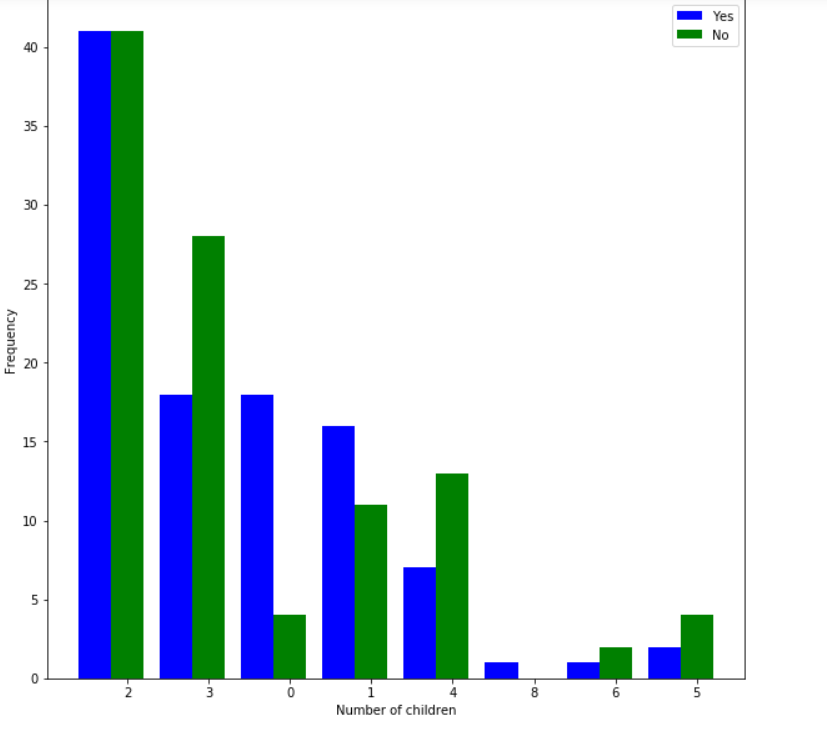
1. **Family History of Obesity**

H0:Number of Children a person has is not associated with a person having a family history of Obesity.

Ha:Number of children a person has is associated with a person having a family history of Obesity.

𝝌**2** =15.804467801 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Family History of Obesity vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having a family

history of Obesity.

`

**Conclusion 2**: Most of the people have 2 children.

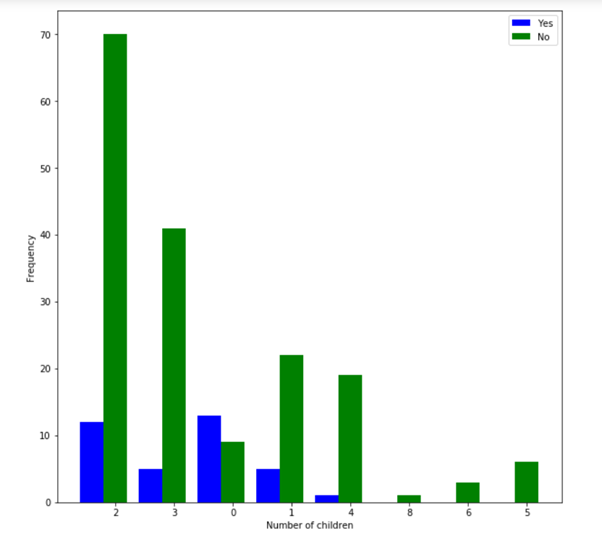
1. **Family History of Pcos**

H0:Number of Children a person has is not associated with a person having a family history of Pcos.

Ha:Number of children a person has is associated with a person having a family history of Pcos.

𝝌**2** =32.68976635 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Family History of Pcos vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having a family

history of Pcos.

`

**Conclusion 2**: Most of the people have 2 children and do not have a family history of Pcos.

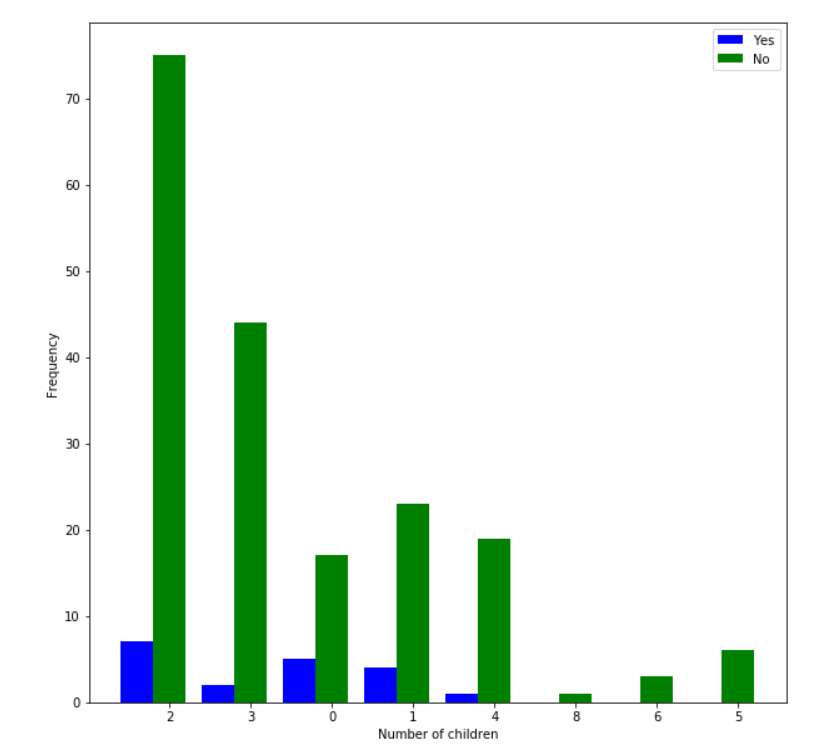
1. **Family History of Hypothyroid**

H0:Number of Children a person has is not associated with a person having a family history of Hypothyroid.

Ha:Number of children a person has is associated with a person having a family history of Hypothyroid.

𝝌**2** =8.631127015 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Family History of Hypothyroid vs Number of children

**Conclusion 1**:Number of children a person has is not associated with a person having a family

history of Hypothyroid.

`

**Conclusion 2**: Most of the people have 2 children and do not have a family history of Hypothyroid.

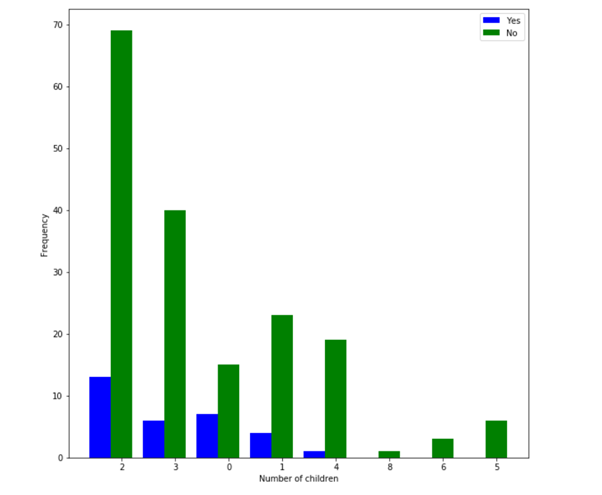
1. **Family History of Cancer**

H0:Number of Children a person has is not associated with a person having a family history of Cancer.

Ha:Number of children a person has is associated with a person having a family history of Cancer.

𝝌**2** =8.4106670405 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Family History of Cancer vs Number of children

**Conclusion 1**:Number of children a person has is not associated with a person having a family

history of Cancer.

`

**Conclusion 2**: Most of the people have 2 children and do not have a family history of Cancer.

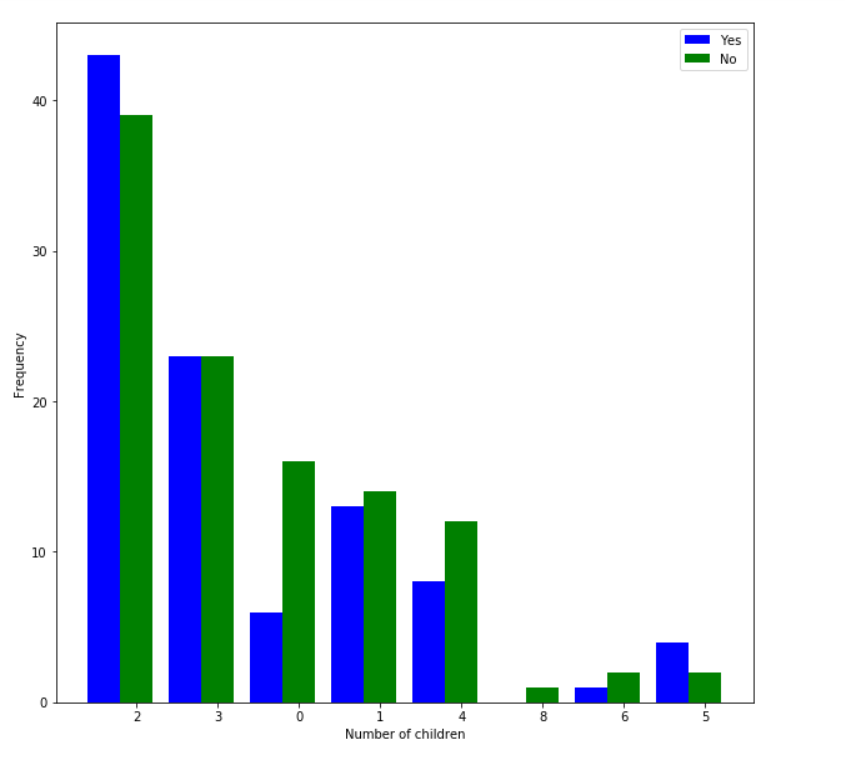
1. **Family History of any other disease**

H0:Number of Children a person has is not associated with a person having a family history of any other disease.

Ha:Number of children a person has is associated with a person having a family history of any other disease.

𝝌**2** =7.01287591 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Family History of any other disease vs Number of children

**Conclusion 1**:Number of children a person has is not associated with a person having a family

history of any other disease.

`

**Conclusion 2**: Most of the people have 2 children and do not have a family history of any other disease.

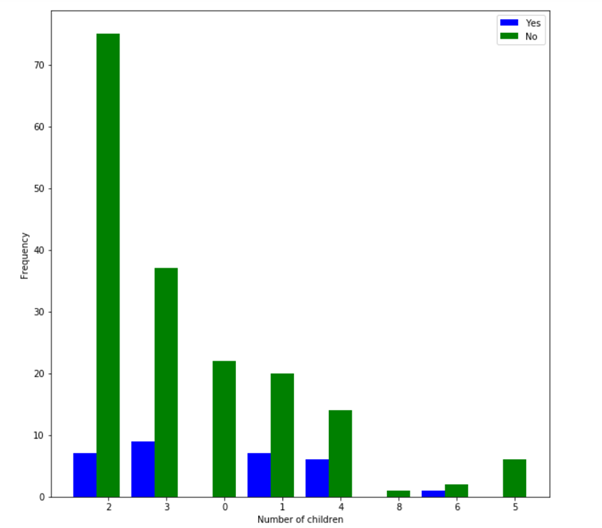
**(9)Person having Diabetes**

H0:Number of Children a person has is not associated with a person having Diabetes.

Ha:Number of children a person has is associated with a person having Diabetes

𝝌**2** =15.8061375 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having Diabetes vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having Diabetes.

`

**Conclusion 2**: Most of the people have 2 children and do not have Diabetes.

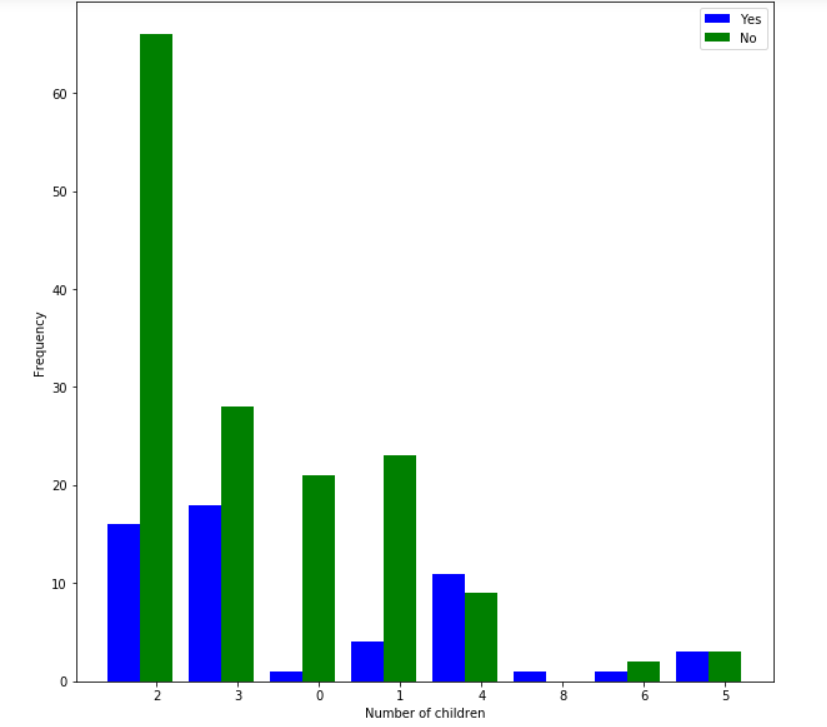
**(10)Person having HTN**

H0:Number of Children a person has is not associated with a person having HTN.

Ha:Number of children a person has is associated with a person having HTN

𝝌**2** =26.003004739 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having HTN vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person having HTN.

`

**Conclusion 2**: Most of the people have 2 children and do not have HTN.

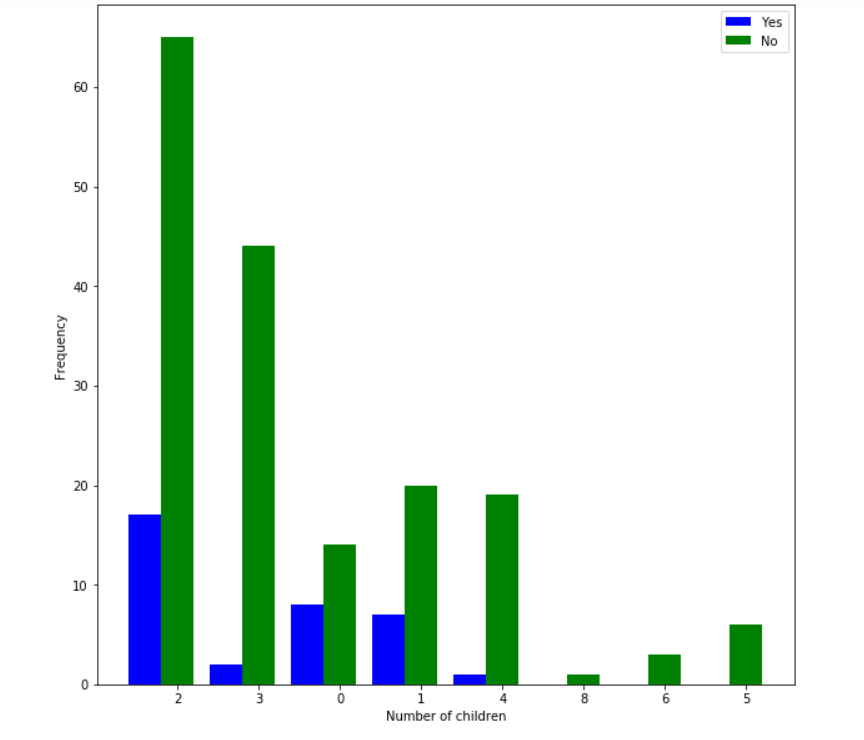
**(11)Person having Pcos**

H0:Number of Children a person has is not associated with a person having Pcos.

Ha:Number of children a person has is associated with a person having Pcos

𝝌**2** =17.562239854 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having Pcos vs Number of children

**Conclusion 1**:Number of children a person has is associated with having Pcos.

`

**Conclusion 2**: Most of the people have 2 children and do not have Pcos.

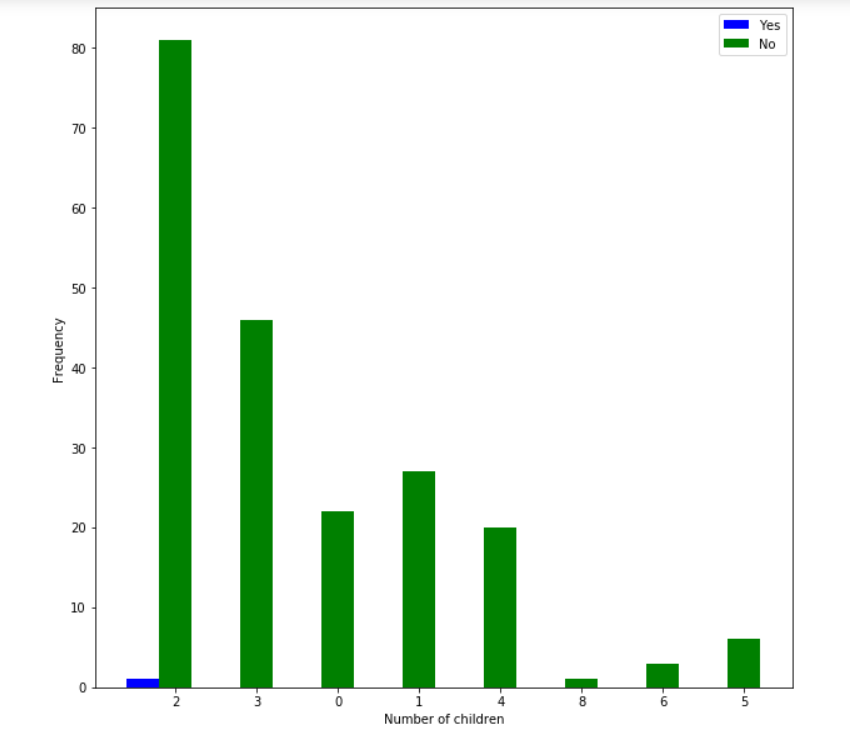
**(12)Person having CVD**

H0:Number of Children a person has is not associated with a person having CVD.

Ha:Number of children a person has is associated with a person having CVD

𝝌**2** =1.53179019 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having CVD vs Number of children

**Conclusion 1**:Number of children a person has is not associated with a person having CVD.

`

**Conclusion 2**: Most of the people have 2 children and do not have CVD.

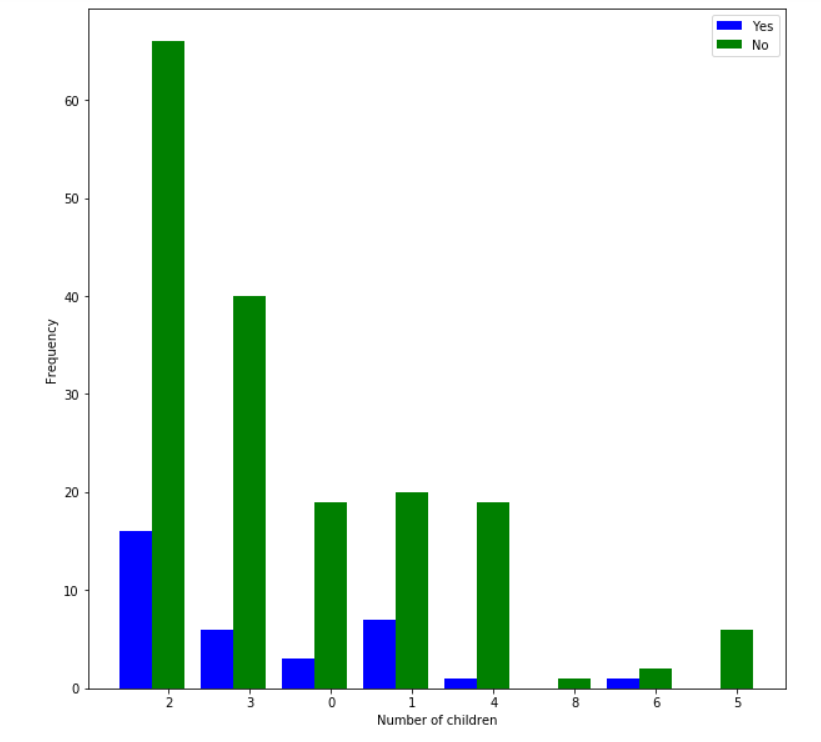
**(13)Person having Hypothyroid**

H0:Number of Children a person has is not associated with a person having Hypothyroid.

Ha:Number of children a person has is associated with a person having Hypothyroid

𝝌**2** =6.7548579 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Hypothyroid vs Number of children

**Conclusion 1**:Number of children a person has is not associated with having Hypothyroid.

`

**Conclusion 2**: Most of the people have 2 children and do not have Hypothyroid.

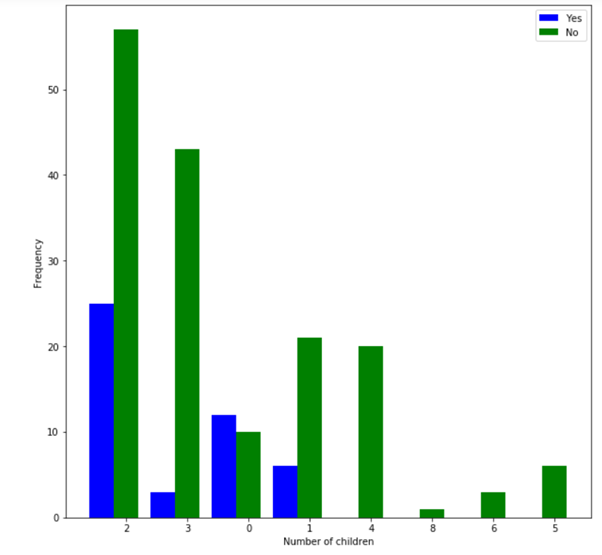
**(14)Person having Menstrual Irregularities**

H0:Number of Children a person has is not associated with a person having Menstrual Irregularities.

Ha:Number of children a person has is associated with a person having Menstrual Irregularities.

𝝌**2** =31.671978074 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person having Menstrual Irregularities vs Number of children

**Conclusion 1**:Number of children a person has is associated with having Menstrual Irregularities.

`

**Conclusion 2**: Most of the people have 2 children and do not have Menstrual Irregularities.

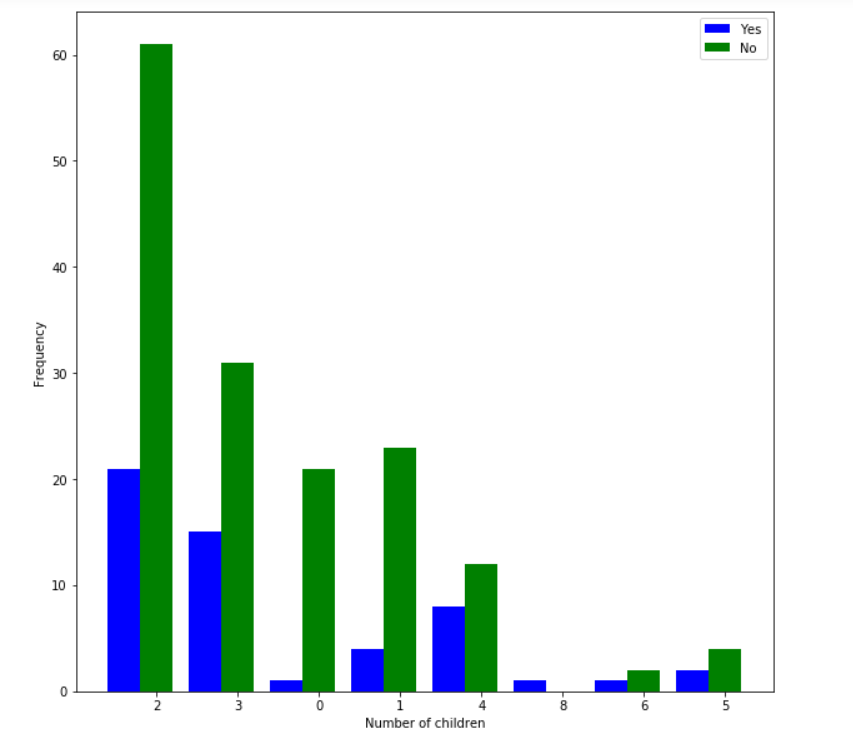
**(15)Person having Arthritis**

H0:Number of Children a person has is not associated with a person having Arthritis.

Ha:Number of children a person has is associated with a person having Arthritis

𝝌**2** =13.320631464 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Arthritis vs Number of children

**Conclusion 1**:Number of children a person has is not associated with a person having Arthritis.

`

**Conclusion 2**: Most of the people have 2 children and do not have Arthritis.

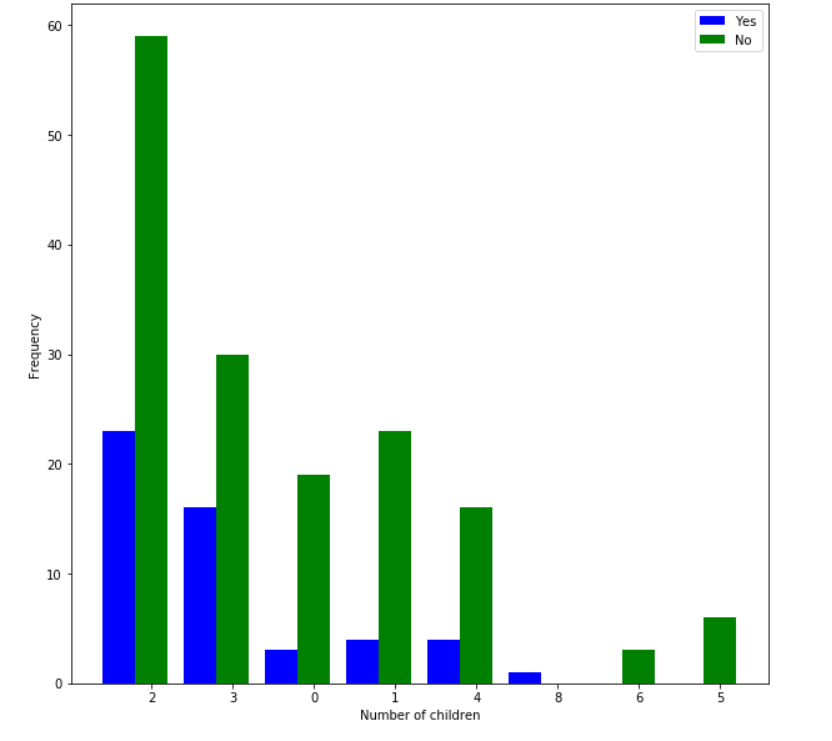
**(16)Person having Vertigo**

H0:Number of Children a person has is not associated with a person having Vertigo.

Ha:Number of children a person has is associated with a person having Vertigo

𝝌**2** =12.1335647 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person having Vertigo vs Number of children

**Conclusion 1**:Number of children a person has is not associated with having Vertigo.

`

**Conclusion 2**: Most of the people have 2 children and do not have Vertigo.

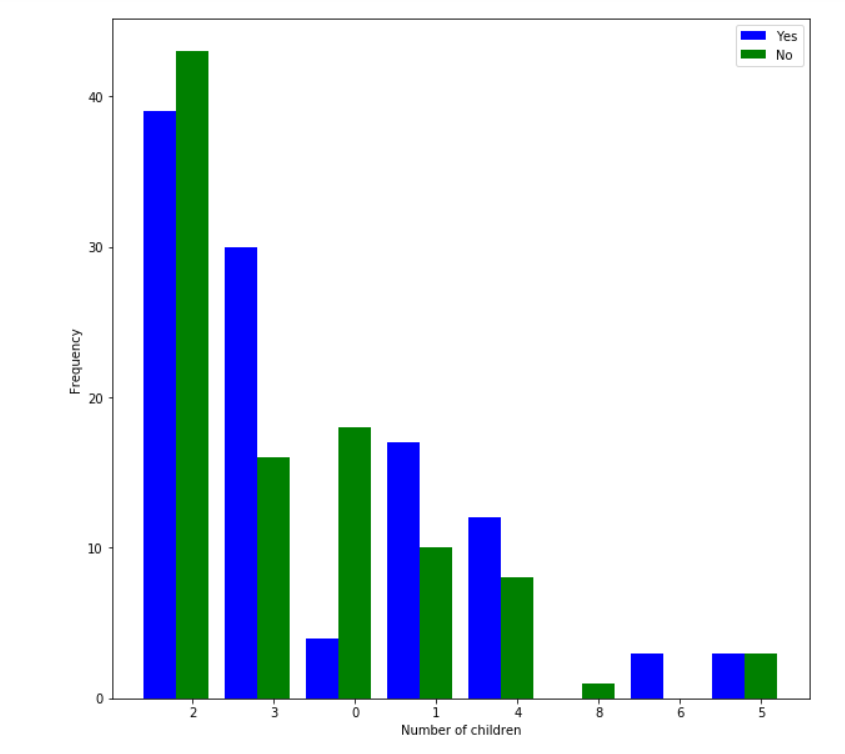
**(17)Person having other medical problem**

H0:Number of Children a person has is not associated with a person having Other medical problem.

Ha:Number of children a person has is associated with a person having Other medical problem

𝝌**2** =27.70162604 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.



Person having Other medical problem vs Number of children

**Conclusion 1**:Number of children a person has is associated with having Other medical problem.

`

**Conclusion 2**: Most of the people have 2 children and do not have Other medical problem.

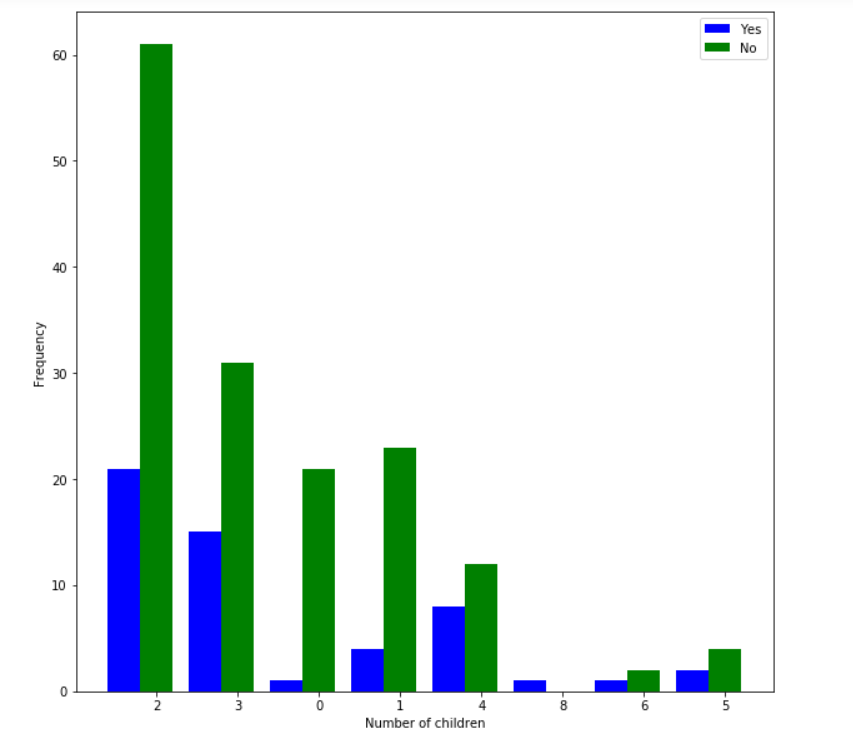
**(18) Person taking medication/supplements**

H0:Number of Children a person has is not associated with a person taking medication/supplements.

Ha:Number of children a person has is associated with a person having Arthritis

𝝌**2** =28.6416842 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** > 𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person taking medication/supplements vs Number of children

**Conclusion 1**:Number of children a person has is associated with a person taking medication/supplements.

`

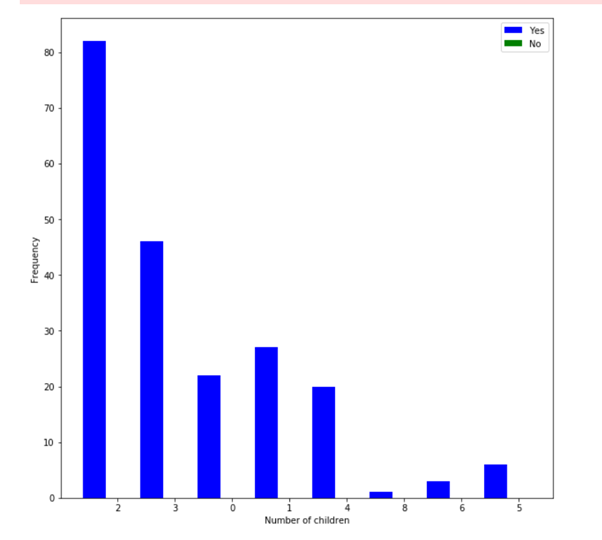
**Conclusion 2**: Most of the people have 2 children and do not take medication/supplements.

**(19) Person has weight management goals**

H0:Number of Children a person has is not associated with a person having weight management goals.

Ha:Number of children a person has is associated with a person having weight management goals.

𝝌**2** =NaN and 𝝌𝛂**2** = 14.067

****

Person taking medication/supplements vs Number of children

**Conclusion 1**:Weight management goal is always yes thud number of children a person has is not associated with it.

`

**Conclusion 2**: Most of the people have 2 children and all of them have interest in weight management .

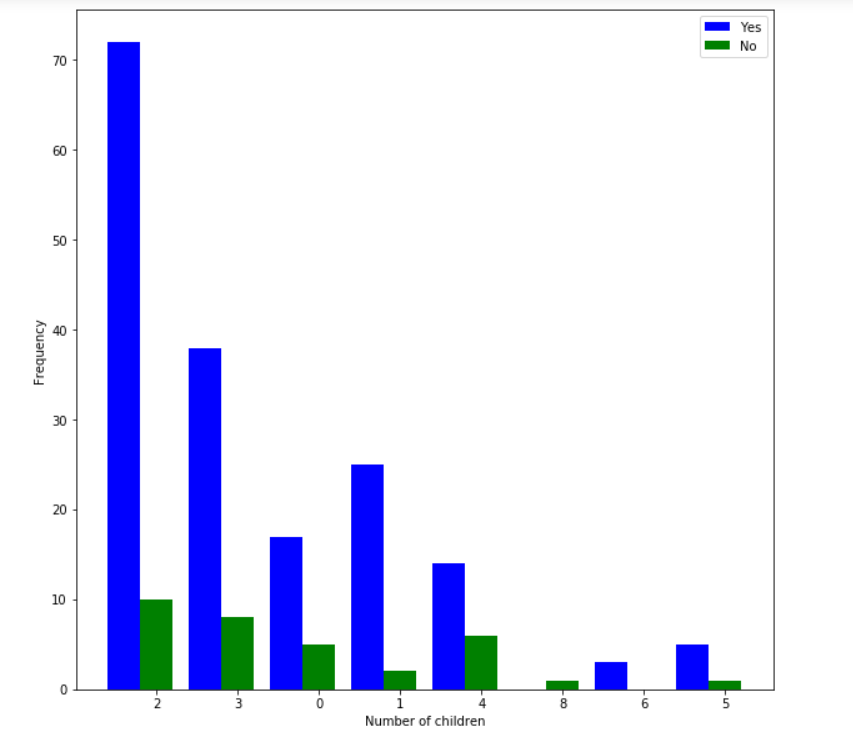
**(20) Person associated with improved fitness goals**

H0:Number of Children a person has is not associated with a person associated with improved fitness goals

Ha:Number of children a person has is associated with a person associated with improved fitness goals .

𝝌**2** =11.9482102931 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** < 𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Person associated with improved fitness

vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person associated with improved fitness

**Conclusion 2**: Most of the people have 2 children and most of them associated with Improved Fitness .

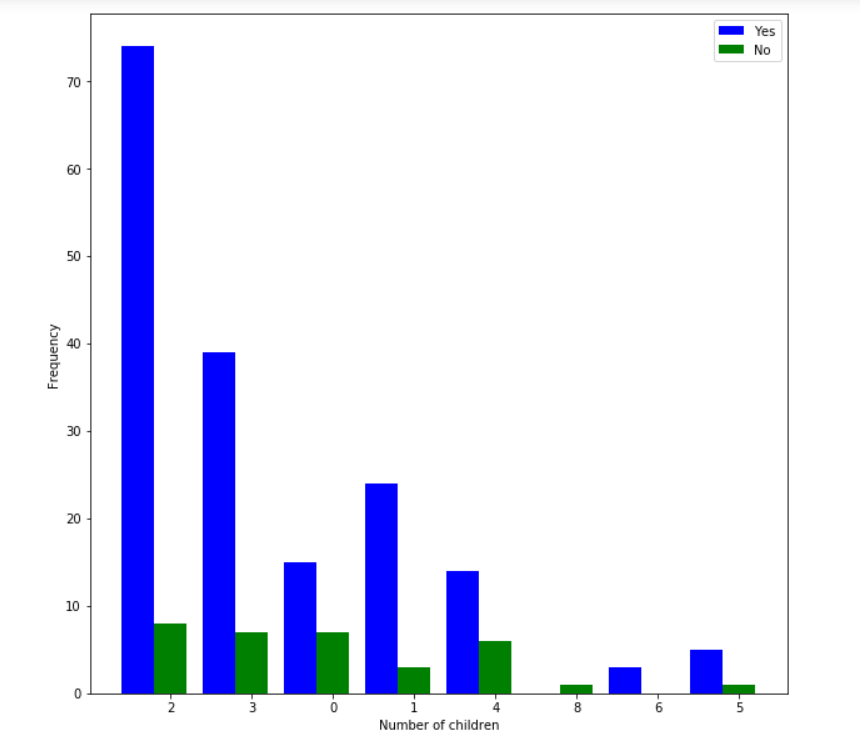
**(21) Person associated with Back Strengthening Goals**

H0:Number of Children a person has is not associated with a person associated with Back Strengthening

Ha:Number of children a person has is associated with a person associated with Back Strengthening.

𝝌**2** =15.761360334 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** >𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person associated with Back Strengthening

vs Number of children

**Conclusion 1**:`Number of Children a person has is associated with a person associated with Back Strengthenings

**Conclusion 2**: Most of the people have 2 children and most of them are associated with Back Strengthening.

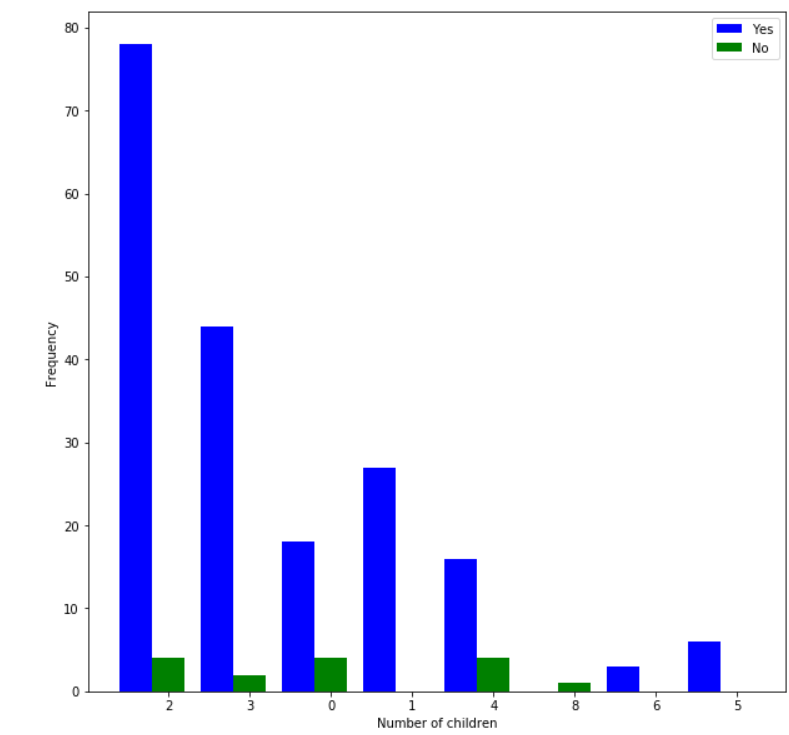
**(22) Person associated with Nutrition and Disease prevention**

H0:Number of Children a person has is not associated with a person associated with Nutrition and Disease prevention

Ha:Number of children a person has is associated with a person associated with Nutrition and Disease prevention

𝝌**2** =25.626003 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** >𝝌𝛂**2** ,we reject the null Hypothesis.

****

Person associated with Nutrition and Disease prevention

vs Number of children

**Conclusion 1**:`Number of Children a person has is associated with a person associated with Nutrition and Disease prevention

**Conclusion 2**: Most of the people have 2 children and most of them are associated with Nutrition and Disease prevention.

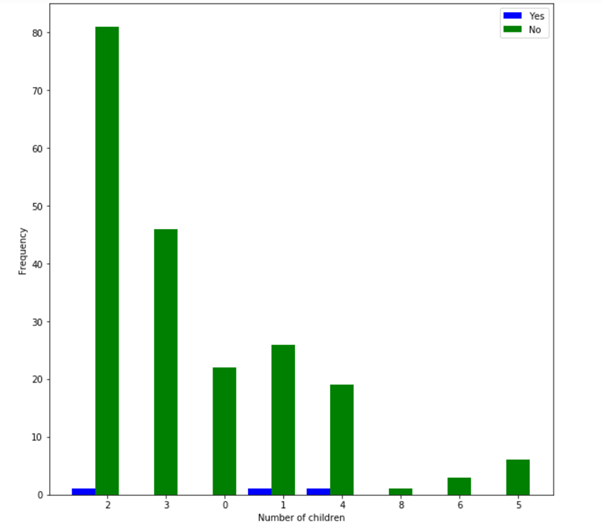
**(23) Person being Vegetarian**

H0:Number of Children a person has is not associated with a person being vegetarian.

Ha:Number of children a person has is associated with a person being Vegetarian.

𝝌**2** =3.90359277 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** <𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Number of Vegetarians vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person being vegetarian.

**Conclusion 2**: Most of the people have 2 children and most of them are not vegetarian.

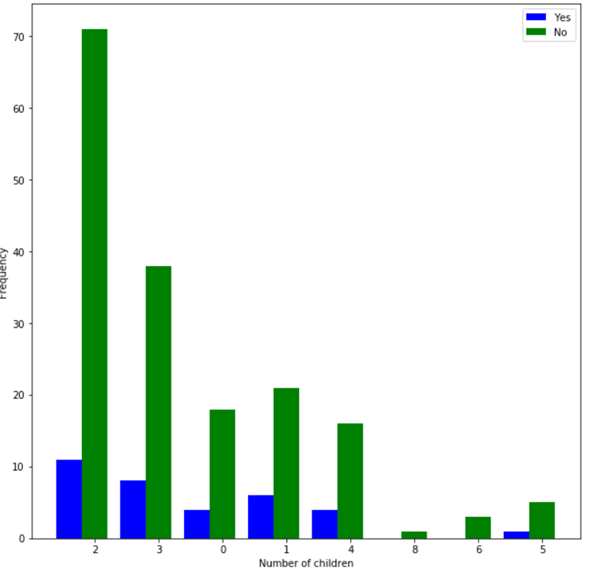
**(24) Person being Lacto-Vegetarian**

H0:Number of Children a person has is not associated with a person being lacto-vegetarian.

Ha:Number of children a person has is associated with a person being lacto-egetarian.

𝝌**2** =2.25569730 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** <𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Number of lacto-Vegetarians vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person being lacto-vegetarian.

**Conclusion 2**: Most of the people have 2 children and most of them are not lacto-vegetarian.

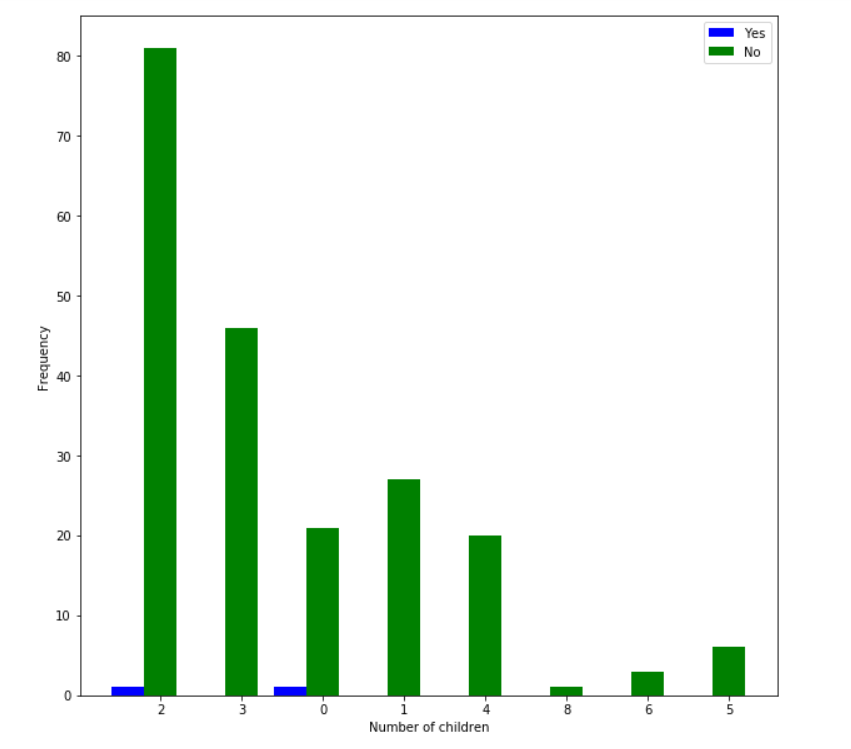
**(25) Person being Ovo Vegetarian**

H0:Number of Children a person has is not associated with a person being Ovo vegetarian.

Ha:Number of children a person has is associated with a person being Ovo Vegetarian.

𝝌**2** =4.00544048 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** <𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Number of Ovo Vegetarians vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person being Ovo vegetarian.

**Conclusion 2**: Most of the people have 2 children and most of them are not Ovo vegetarian.

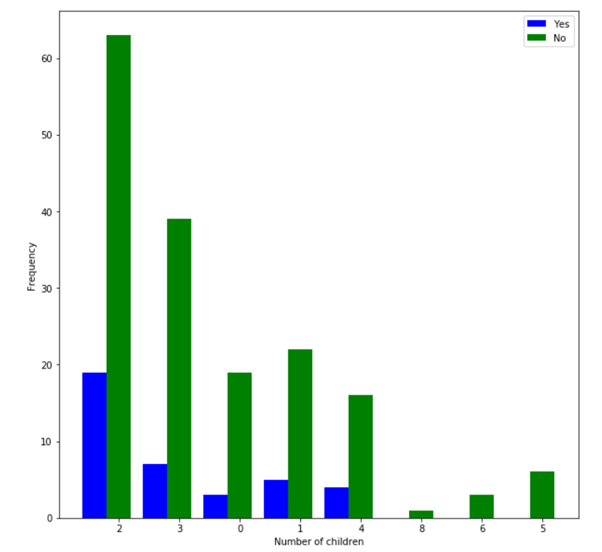
**(26) Person being Lacto-Ovo-Vegetarian**

H0:Number of Children a person has is not associated with a person being lacto-ovo-vegetarian.

Ha:Number of children a person has is associated with a person being lacto-ovo-vegetarian.

𝝌**2** =4.18233355 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** <𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Number of lacto-ovo-Vegetarians vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person being lacto-ovo-vegetarian.

**Conclusion 2**: Most of the people have 2 children and most of them are not lacto-ovo-vegetarian.

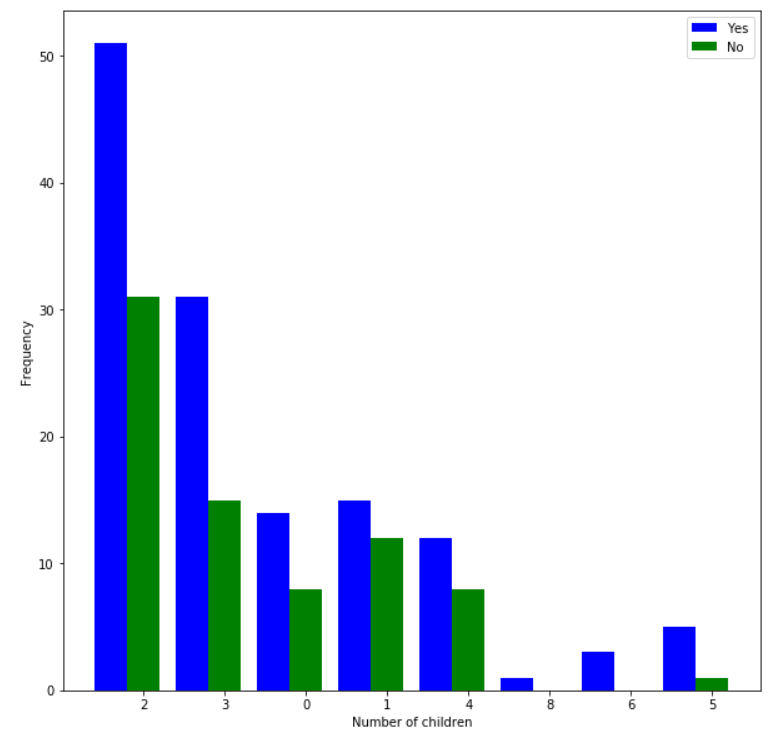
**(27) Person being Non- Vegetarian**

H0:Number of Children a person has is not associated with a person being Non- vegetarian.

Ha:Number of children a person has is associated with a person being Non-Vegetarian.

𝝌**2** =4.00544048 and 𝝌𝛂**2** = 14.067

Since 𝝌**2** <𝝌𝛂**2** ,we do not reject the null Hypothesis.

****

Number of Non-Vegetarians vs Number of children

**Conclusion 1**:`Number of Children a person has is not associated with a person being Non- vegetarian.

**Conclusion 2**: Most of the people have 2 children and most of them are not Non-vegetarian.

**Number of Children was found to be associated with:**

1. Family History of Hypertension
2. Family History of Diabetes
3. Family History of Obesity
4. Family History of Pcos
5. Person having Diabetes
6. Person having HTN
7. Person having Pcos
8. Person having Menstrual Irregularities
9. Person having other medical problem
10. Person taking medication/supplements
11. Person associated with Back Strengthening Goals
12. Person associated with Nutrition and Disease prevention

# FACTOR ANALYSIS

## Part 1

**Ready to Eat Bakery Items(Pastries, Puffs, Cookies, Ice-Cream,Cakes, Bakery Items, Sandwich)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptive Statistics** | | | |
|  | Mean | Std. Deviation | Analysis N |
| Pastries | 1.63 | 0.616 | 207 |
| Puffs | 2.52 | 0.667 | 207 |
| Cookies | 1.26 | 0.574 | 207 |
| Ice-cream | 1.72 | 0.620 | 207 |
| Cakes | 1.29 | 0.542 | 207 |
| Bakery\_Items | 1.43 | 0.707 | 207 |
| Sandwich | 2.04 | 0.852 | 207 |

The variance (/standard deviation) of Sandwiches is the maximum.

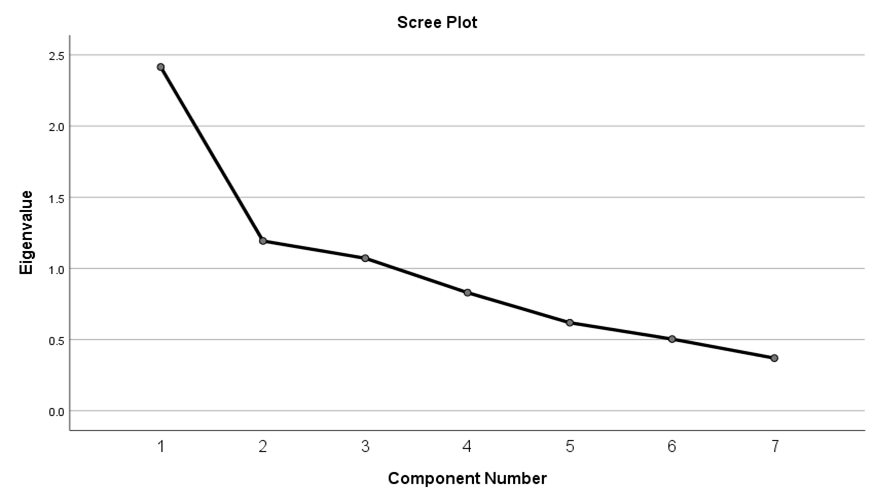
|  |  |  |
| --- | --- | --- |
| **KMO and Bartlett's Test** | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.656 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 247.913 |
| df | 21 |
| Sig. | 0.000 |

As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| **Communalities** | | |
|  | Initial | Extraction |
| Pastries | 1.000 | 0.586 |
| Puffs | 1.000 | 0.766 |
| Cookies | 1.000 | 0.769 |
| Ice-cream | 1.000 | 0.314 |
| Cakes | 1.000 | 0.795 |
| Bakery\_Items | 1.000 | 0.691 |
| Sandwich | 1.000 | 0.758 |

The major reason for obesity is Cakes followed by Cookies and Puffs. Also the most consumed item is Cakes.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Total Variance Explained** | | | | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.415 | 34.500 | 34.500 | 2.415 | 34.500 | 34.500 | 1.868 | 26.692 | 26.692 |
| 2 | 1.193 | 17.038 | 51.538 | 1.193 | 17.038 | 51.538 | 1.701 | 24.306 | 50.998 |
| 3 | 1.071 | 15.304 | 66.842 | 1.071 | 15.304 | 66.842 | 1.109 | 15.844 | 66.842 |
| 4 | 0.830 | 11.853 | 78.695 |  |  |  |  |  |  |
| 5 | 0.618 | 8.833 | 87.528 |  |  |  |  |  |  |
| 6 | 0.503 | 7.191 | 94.719 |  |  |  |  |  |  |
| 7 | 0.370 | 5.281 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by 3 components .

This also shows 3 components are taken as there is bend after value 3 on x axis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Component Matrix** | | | |
|  | Component | | |
| 1 | 2 | 3 |
| Pastries | 0.723 | 0.221 | 0.120 |
| Puffs | 0.414 | 0.589 | 0.497 |
| Cookies | 0.433 | -0.409 | 0.643 |
| Ice-cream | 0.522 | 0.197 | 0.053 |
| Cakes | 0.735 | -0.375 | -0.338 |
| Bakery\_Items | 0.691 | -0.439 | -0.144 |
| Sandwich | 0.494 | 0.506 | -0.507 |

|  |  |  |  |
| --- | --- | --- | --- |
| Rotated Component Matrix | | | |
|  | Component | | |
| 1 | 2 | 3 |
| Pastries | 0.379 | 0.665 | -0.012 |
| Puffs | -0.194 | 0.849 | 0.092 |
| Cookies | 0.306 | 0.290 | 0.769 |
| Ice-cream | 0.267 | 0.489 | -0.058 |
| Cakes | 0.883 | 0.107 | -0.062 |
| Bakery\_Items | 0.812 | 0.119 | 0.135 |
| Sandwich | 0.289 | 0.437 | -0.695 |

|  |  |  |
| --- | --- | --- |
| Component 1 | Component 2 | Component 3 |
| Cakes  Bakery\_Items | Pastries  Puffs  Ice-cream | Cookies  Sandwich |

|  |  |  |
| --- | --- | --- |
| Bakery\_Items | Others(Pastries, Puffs, Ice-cream) | Snacks |
| Cakes  Bakery\_Items | Pastries  Puffs  Ice-cream | Cookies  Sandwich |

Thus, after rotation, Bakery Items 1 accounts for 26.692% of the variance; Others(Pastries, Puffs and Ice-cream) accounts for 24.306% of the variance; Snacks accounts for 15.844% of the variance. All the 3 factors together explain for 66.842% of the variance of Ready to eat bakery items.

## Part 2

**Ready to Eat Snacks (Pizza,Veg-Roll,Burger,French-Fries)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Pizza | 1.70 | 0.621 | 207 |
| Burger | 1.29 | 0.515 | 207 |
| Veg\_Roll | 1.85 | 0.825 | 207 |
| French\_fries | 1.23 | 0.544 | 207 |

The variance (/standard deviation) of Veg\_Roll is the maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.686 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 100.103 |
| df | 6 |
| Sig. | 0.000 |

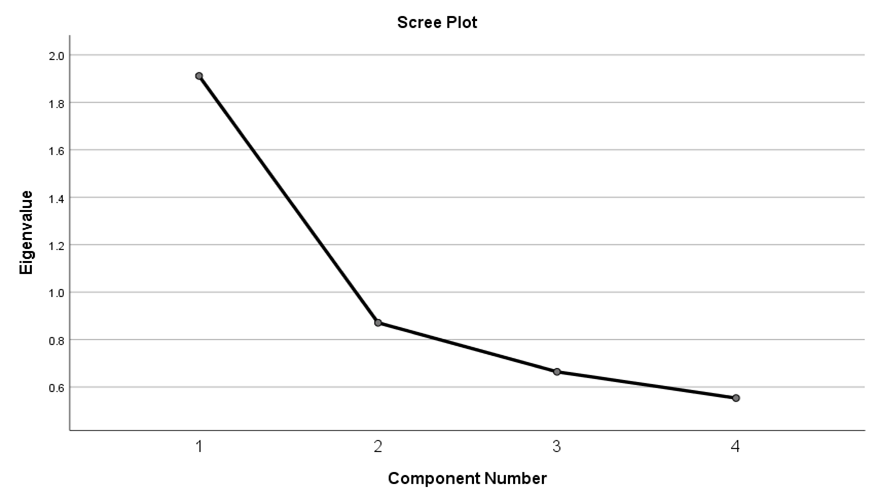
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Pizza | 1.000 | 0.520 |
| Burger | 1.000 | 0.600 |
| Veg\_Roll | 1.000 | 0.274 |
| French\_fries | 1.000 | 0.517 |

The major reason for obesity is Burger followed by Pizza. Also the most consumed item is Burger.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.911 | 47.786 | 47.786 | 1.911 | 47.786 | 47.786 |
| 2 | 0.871 | 21.772 | 69.558 |  |  |  |
| 3 | 0.664 | 16.604 | 86.162 |  |  |  |
| 4 | 0.554 | 13.838 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only.



There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Pizza | 0.721 |
| Burger | 0.774 |
| Veg\_Roll | 0.524 |
| French\_fries | 0.719 |

As only one component is extracted. The solution cannot be rotated.

|  |
| --- |
| Component 1(Ready to eat Snacks) |
| Pizza  Burger  Veg Roll  French Fries |

The only component explains 47.786% of the variance.

## Part 3

**Ready to Eat Packet Snacks (Potato chips, Kurkure, Lays, Haldiram Snacks)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Potato\_chips | 2.26 | 1.169 | 207 |
| Kurkure | 1.67 | 0.788 | 207 |
| Lays | 3.32 | 0.840 | 207 |
| Haldiram\_snacks | 2.05 | 0.910 | 207 |

The variance(/standard deviation) of Potato\_Chips is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.461 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 55.308 |
| df | 6 |
| Sig. | 0.000 |

As KMO < 0.5 Factor Analysis cannot be done as the sample is not adequate.

## Part 4

**Fried Snacks (Samosa, Bajjies, Bondas, Punugulu, Any Fried Snack)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Samosa | 2.11 | 0.614 | 207 |
| Bajjies | 2.27 | 0.740 | 207 |
| Bondas | 1.91 | 0.810 | 207 |
| Punugulu | 1.65 | 0.761 | 207 |
| Any\_fried\_snack | 4.29 | 0.751 | 207 |

The variance(/standard deviation) of Bondas is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.607 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 138.752 |
| df | 10 |
| Sig. | 0.000 |

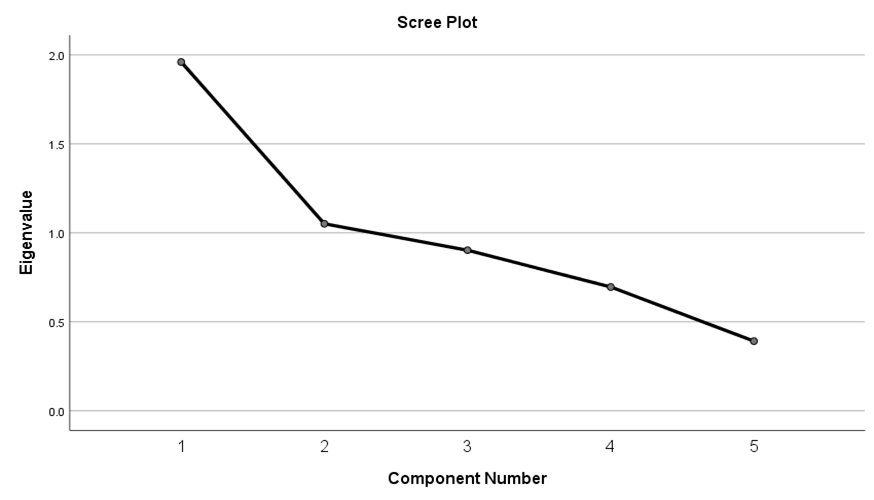
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Samosa | 1.000 | 0.877 |
| Bajjies | 1.000 | 0.596 |
| Bondas | 1.000 | 0.757 |
| Punugulu | 1.000 | 0.506 |
| Any\_fried\_snack | 1.000 | 0.275 |

The major reason for obesity is Samosa followed by Bondas.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.960 | 39.202 | 39.202 | 1.960 | 39.202 | 39.202 | 1.954 | 39.080 | 39.080 |
| 2 | 1.051 | 21.018 | 60.220 | 1.051 | 21.018 | 60.220 | 1.057 | 21.139 | 60.220 |
| 3 | 0.902 | 18.046 | 78.265 |  |  |  |  |  |  |
| 4 | 0.696 | 13.911 | 92.176 |  |  |  |  |  |  |
| 5 | 0.391 | 7.824 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by 2 components .



The data can be expressed by 2 components as the bend is seen after component 2.

|  |  |  |
| --- | --- | --- |
| Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Samosa | -0.015 | 0.937 |
| Bajjies | 0.767 | -0.088 |
| Bondas | 0.850 | -0.186 |
| Punugulu | 0.695 | 0.152 |
| Any\_fried\_snack | 0.408 | 0.329 |

|  |  |  |
| --- | --- | --- |
| Rotated Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Samosa | -0.091 | 0.932 |
| Bajjies | 0.772 | -0.025 |
| Bondas | 0.862 | -0.116 |
| Punugulu | 0.680 | 0.208 |
| Any\_fried\_snack | 0.380 | 0.361 |

|  |  |
| --- | --- |
| Component 1 | Component 2 |
| Bajjies  Bondas  Punugulu  Any\_fried\_snacks | Samosa |

|  |  |
| --- | --- |
| Fried Snacks except Samosa | Samosa |
| Bajjies  Bondas  Punugulu  Any\_fried\_snacks | Samosa |

Thus, after rotation, Fried Snacks(except Samosa) accounts for 39.080% of the variance; Samosa accounts for 21.139% of the variance; Both the 2 factors together explain for 60.220 % of the variance of Ready to eat bakery items.

## Part 5

**Chat Counter Snacks (Chat,Panipuri,Cutlet,Popcorn,Sweetcorn)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Chat | 1.91 | 0.816 | 207 |
| Panipuri | 3.44 | 0.953 | 207 |
| Cutlet | 1.64 | 0.723 | 207 |
| Popcorn | 1.16 | 0.481 | 207 |
| SweetCorn | 1.78 | 0.901 | 207 |

The variance(/standard deviation) of Panipuri is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.605 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 65.567 |
| df | 10 |
| Sig. | 0.000 |

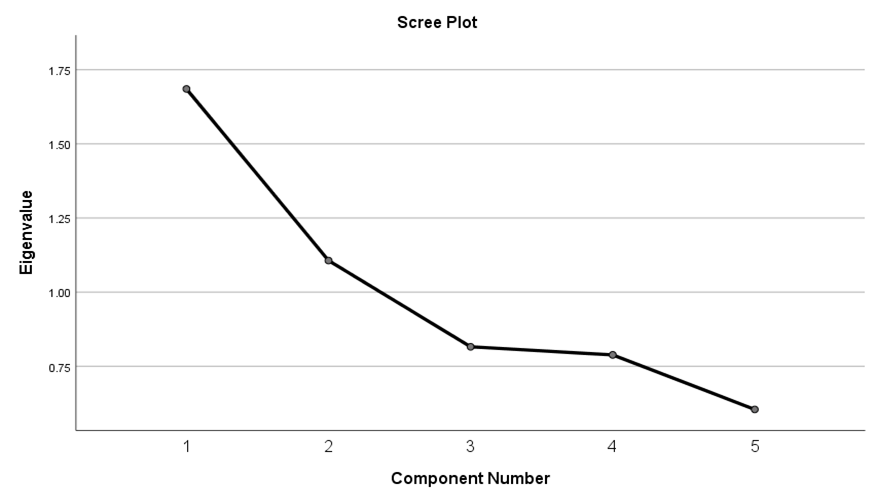
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Chat | 1.000 | 0.518 |
| Panipuri | 1.000 | 0.667 |
| Cutlet | 1.000 | 0.577 |
| Popcorn | 1.000 | 0.518 |
| SweetCorn | 1.000 | 0.511 |

Panipuri is preferred by most people followed by Cutlet.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.685 | 33.706 | 33.706 | 1.685 | 33.706 | 33.706 | 1.527 | 30.544 | 30.544 |
| 2 | 1.106 | 22.128 | 55.834 | 1.106 | 22.128 | 55.834 | 1.264 | 25.290 | 55.834 |
| 3 | 0.816 | 16.313 | 72.147 |  |  |  |  |  |  |
| 4 | 0.788 | 15.766 | 87.913 |  |  |  |  |  |  |
| 5 | 0.604 | 12.087 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by 2 components .



The data can be expressed by 2 components as the bend is seen after component 2.

|  |  |  |
| --- | --- | --- |
| Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Chat | 0.681 | -0.232 |
| Panipuri | 0.395 | 0.715 |
| Cutlet | 0.742 | -0.162 |
| Popcorn | 0.504 | -0.514 |
| SweetCorn | 0.510 | 0.501 |

|  |  |  |
| --- | --- | --- |
| Rotated Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Chat | 0.702 | 0.158 |
| Panipuri | -0.037 | 0.816 |
| Cutlet | 0.718 | 0.249 |
| Popcorn | 0.698 | -0.175 |
| SweetCorn | 0.173 | 0.693 |

|  |  |
| --- | --- |
| Component 1 | Component 2 |
| Chat  Cutlet  Popcorn | Panipuri  Sweetcorn |

|  |  |
| --- | --- |
| Chat | Panipuri & Sweetcorn |
| Chat  Cutlet  Popcorn | Panipuri  Sweetcorn |

Thus, after rotation, Chat accounts for 30.544% of the variance; Panipuri & Sweetcorn accounts for 25.29% of the variance; Both the 2 factors together explain for 55.834 % of the variance of Chat Snacks.

## Part 6

**Fruit and Milk Items(Faluda, Fruit salad, Fruit juice(f),Fruit juice(p), Milkshakes)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Faluda | 2.00 | 0.711 | 207 |
| Fruit\_salad | 1.67 | 0.654 | 207 |
| Fruit\_juice\_f | 1.80 | 0.702 | 207 |
| Fruit\_juice\_p | 2.48 | 0.924 | 207 |
| Milkshakes | 1.90 | 0.713 | 207 |

The variance(/standard deviation) of Fruit juice(p) is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.522 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 140.626 |
| df | 10 |
| Sig. | 0.000 |

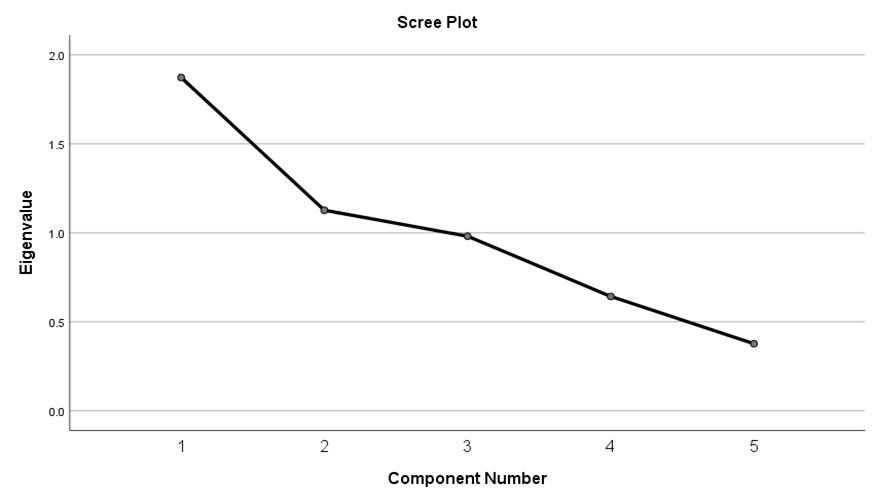
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Faluda | 1.000 | 0.323 |
| Fruit\_salad | 1.000 | 0.748 |
| Fruit\_juice\_f | 1.000 | 0.522 |
| Fruit\_juice\_p | 1.000 | 0.687 |
| Milkshakes | 1.000 | 0.721 |

Fruit salad is preferred by most people followed by Milkshakes.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.873 | 37.457 | 37.457 | 1.873 | 37.457 | 37.457 | 1.708 | 34.155 | 34.155 |
| 2 | 1.127 | 22.539 | 59.996 | 1.127 | 22.539 | 59.996 | 1.292 | 25.841 | 59.996 |
| 3 | 0.981 | 19.621 | 79.617 |  |  |  |  |  |  |
| 4 | 0.643 | 12.851 | 92.468 |  |  |  |  |  |  |
| 5 | 0.377 | 7.532 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by 2 components .



The data can be expressed by 2 components as the bend is seen after component 2.

|  |  |  |
| --- | --- | --- |
| Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Faluda | 0.359 | -0.440 |
| Fruit\_salad | 0.850 | -0.157 |
| Fruit\_juice\_f | 0.660 | -0.294 |
| Fruit\_juice\_p | 0.215 | 0.800 |
| Milkshakes | 0.734 | 0.427 |

|  |  |  |
| --- | --- | --- |
| Rotated Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Faluda | 0.524 | -0.219 |
| Fruit\_salad | 0.824 | 0.262 |
| Fruit\_juice\_f | 0.721 | 0.051 |
| Fruit\_juice\_p | -0.187 | 0.807 |
| Milkshakes | 0.447 | 0.722 |

|  |  |
| --- | --- |
| Component 1 | Component 2 |
| Faluda  Fruit\_salad  Fruit\_juice\_f | Fruit\_juice\_p  Milkshakes |

|  |  |
| --- | --- |
| Fruit Items | Drinks |
| Faluda  Fruit\_salad  Fruit\_juice\_f | Fruit\_juice\_p  Milkshakes |

Thus, after rotation, Fruit Items accounts for 34.155% of the variance; Drinks accounts for 25.841% of the variance; Both the 2 factors together explain for 59.996 % of the variance of Fruit and Milk Items.

## Part 7

**Drinks(Cool drinks, Diet drinks, Energy drinks, Hard drinks)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Cool\_drinks | 3.67 | 0.944 | 207 |
| Diet\_drinks | 1.56 | 1.031 | 207 |
| Energy\_drinks | 2.64 | 1.797 | 207 |
| Hard\_drinks | 1.10 | 0.369 | 207 |

The variance(/standard deviation) of Cool drinks is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.557 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 125.100 |
| df | 6 |
| Sig. | 0.000 |

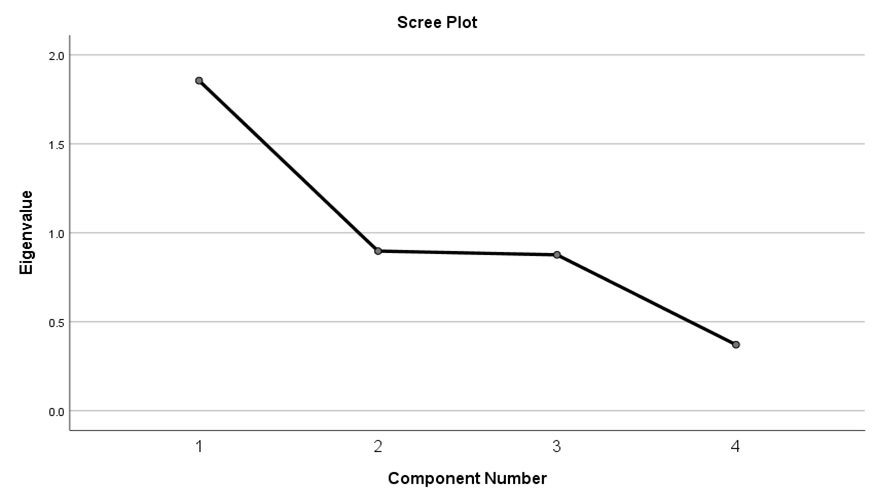
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Cool\_drinks | 1.000 | 0.579 |
| Diet\_drinks | 1.000 | 0.768 |
| Energy\_drinks | 1.000 | 0.271 |
| Hard\_drinks | 1.000 | 0.237 |

Diet Drinks is preferred by most people followed by Cool Drinks.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.855 | 46.385 | 46.385 | 1.855 | 46.385 | 46.385 |
| 2 | 0.897 | 22.430 | 68.815 |  |  |  |
| 3 | 0.876 | 21.908 | 90.723 |  |  |  |
| 4 | 0.371 | 9.277 | 100.000 |  |  |  |

This shows that the data can be expressed by only 1 components .



The data can be expressed by just 1 component.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Cool\_drinks | -0.761 |
| Diet\_drinks | 0.876 |
| Energy\_drinks | 0.521 |
| Hard\_drinks | 0.487 |

As only one component is extracted. The solution cannot be rotated.

|  |
| --- |
| Drinks |
| Cool drinks  Diet Drinks  Energy Drinks  Hard Drinks |

## Part 8

**Sweets(Gummy Candies, Hard Candies, Chocolates, Chocolate Bars)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Gummy\_Candies | 1.07 | 0.311 | 207 |
| Hard\_Candies | 1.16 | 0.481 | 207 |
| Chocolates | 2.02 | 1.088 | 207 |
| Chocolatebars | 1.56 | 0.906 | 207 |

The variance(/standard deviation) of Chocolates is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.618 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 197.651 |
| df | 6 |
| Sig. | 0.000 |

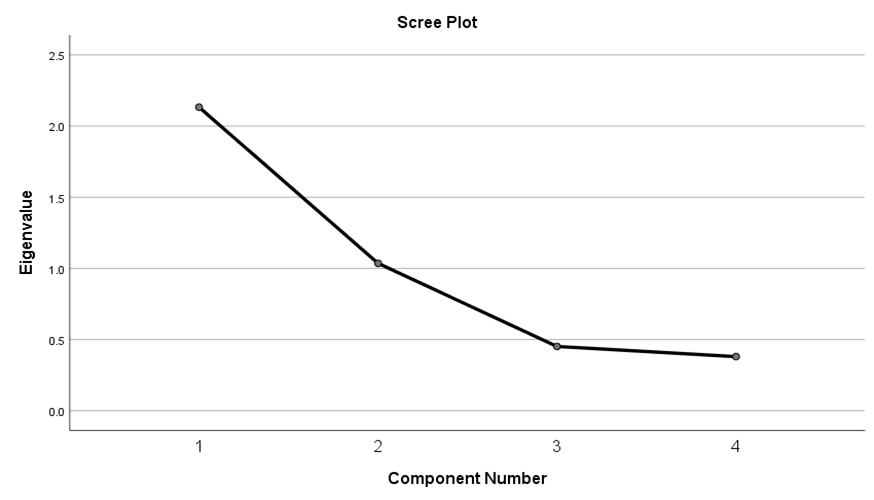
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Gummy\_Candies | 1.000 | 0.829 |
| Hard\_Candies | 1.000 | 0.783 |
| Chocolates | 1.000 | 0.817 |
| Chocolatebars | 1.000 | 0.740 |

Gummy Candies are the most preferred by people.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.132 | 53.312 | 53.312 | 2.132 | 53.312 | 53.312 | 1.632 | 40.808 | 40.808 |
| 2 | 1.036 | 25.894 | 79.206 | 1.036 | 25.894 | 79.206 | 1.536 | 38.398 | 79.206 |
| 3 | 0.451 | 11.286 | 90.493 |  |  |  |  |  |  |
| 4 | 0.380 | 9.507 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by 2 components .



The bend is after 2 component number and hence 2 components are taken into consideration.

|  |  |  |
| --- | --- | --- |
| Component Matrixa | | |
|  | Component | |
| 1 | 2 |
| Gummy\_Candies | 0.710 | -0.570 |
| Hard\_Candies | 0.799 | -0.380 |
| Chocolates | 0.635 | 0.643 |
| Chocolatebars | 0.766 | 0.391 |

|  |  |  |
| --- | --- | --- |
| Rotated Component Matrixa | | |
|  | Component | |
| 1 | 2 |
| Gummy\_Candies | 0.908 | 0.059 |
| Hard\_Candies | 0.846 | 0.260 |
| Chocolates | 0.034 | 0.903 |
| Chocolatebars | 0.301 | 0.806 |

|  |  |
| --- | --- |
| Component 1 | Component 2 |
| Gummy Candies  Hard Candies | Chocolates  Chocolate Bars |

|  |  |
| --- | --- |
| Candies | Chocolates |
| Gummy Candies  Hard Candies | Chocolates  Chocolate Bars |

Thus, after rotation, Candies accounts for 40.808% of the variance; Chocolates accounts for 38.398% of the variance; Both the 2 factors together explain for 79.206 % of the variance of Sweets.

## Part 9

**Maggie Products(Noodles,Maggie,Pasta, Cup Noodles)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Noodles | 1.55 | 0.761 | 207 |
| Maggie | 3.67 | 0.661 | 207 |
| Pasta | 1.88 | 0.924 | 207 |
| Cup\_Noodles | 1.31 | 0.683 | 207 |

The variance(/standard deviation) of Noodles is maximum.

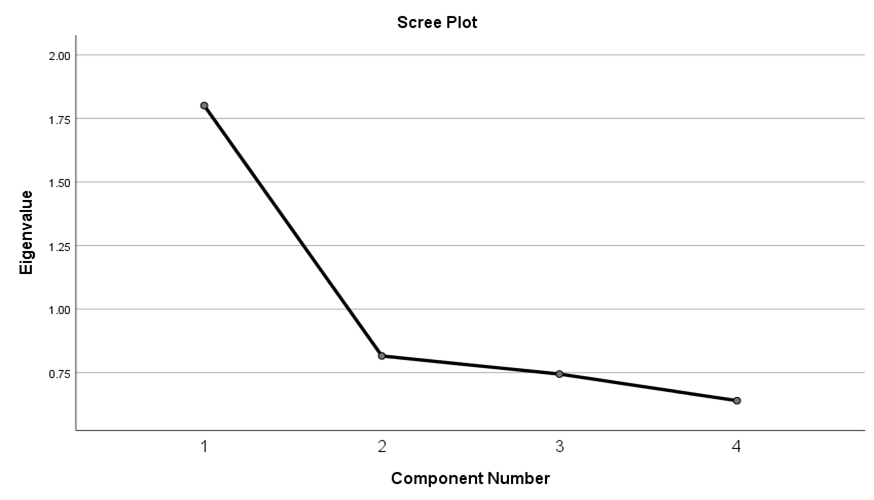
|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.688 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 72.972 |
| df | 6 |
| Sig. | 0.000 |

As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Noodles | 1.000 | 0.426 |
| Maggie | 1.000 | 0.370 |
| Pasta | 1.000 | 0.446 |
| Cup\_Noodles | 1.000 | 0.559 |

Cup Noodles are the most preferred by people.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.801 | 45.014 | 45.014 | 1.801 | 45.014 | 45.014 |
| 2 | 0.815 | 20.387 | 65.401 |  |  |  |
| 3 | 0.744 | 18.608 | 84.009 |  |  |  |
| 4 | 0.640 | 15.991 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only**.**

There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Noodles | 0.653 |
| Maggie | 0.608 |
| Pasta | 0.667 |
| Cup\_Noodles | 0.748 |

As only one component is extracted. The solution cannot be rotated.

The only component explains 45.014% of the variance.

|  |
| --- |
| Maggie Products |
| Noodles  Maggie  Pasta  Cup Noodles |

## Part 10

**Soup(Readymade Soup,Soup at Rest,Soup at Home)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Readymade\_soup | 1.50 | 0.985 | 207 |
| Soup\_rest | 1.39 | 0.792 | 207 |
| Soup\_home | 1.04 | 0.266 | 207 |

The variance(/standard deviation) of Readymade Soup is maximum.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.540 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 10.733 |
| df | 3 |
| Sig. | 0.013 |

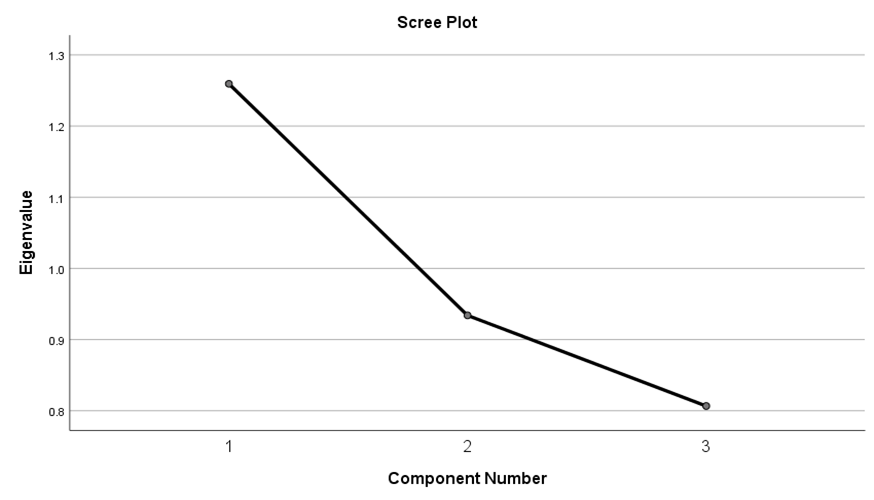
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Readymade\_soup | 1.000 | 0.482 |
| Soup\_rest | 1.000 | 0.516 |
| Soup\_home | 1.000 | 0.261 |

Soup\_rest are the most preferred by people.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.259 | 41.980 | 41.980 | 1.259 | 41.980 | 41.980 |
| 2 | 0.934 | 31.132 | 73.113 |  |  |  |
| 3 | 0.807 | 26.887 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only**.**



There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Readymade\_soup | 0.695 |
| Soup\_rest | 0.719 |
| Soup\_home | 0.510 |

As only one component is extracted. The solution cannot be rotated.

The only component explains 41.98% of the variance.

|  |
| --- |
| Soup |
| Readymade Soup  Soup@rest  Soup@home |

## Part 11

**Breakfast(Cereals,Oats,Chocoes)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Cereals | 1.97 | 1.094 | 207 |
| Oats | 3.18 | 0.877 | 207 |
| Chocoes | 1.50 | 0.769 | 207 |

Cereals depict the maximum variance.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.518 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 5.287 |
| df | 3 |
| Sig. | 0.152 |

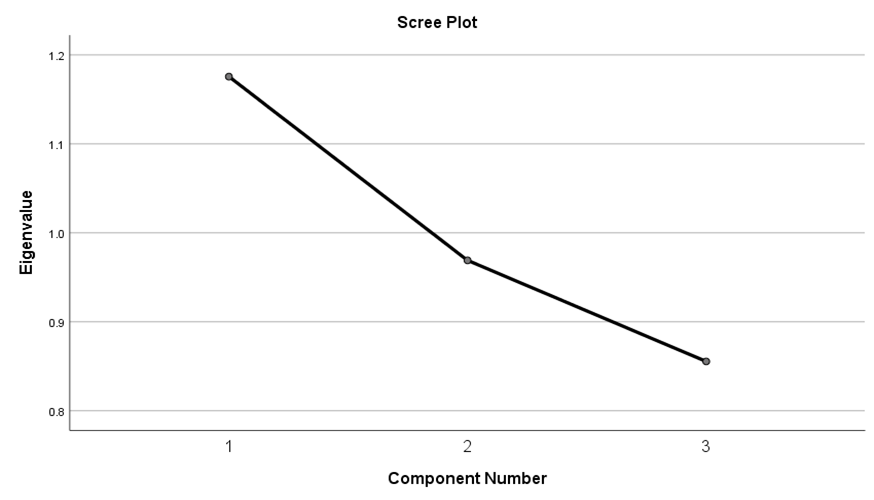
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Cereals | 1.000 | 0.530 |
| Oats | 1.000 | 0.275 |
| Chocoes | 1.000 | 0.371 |

Cereals are preferred the most followed by Chocoes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.176 | 39.186 | 39.186 | 1.176 | 39.186 | 39.186 |
| 2 | 0.969 | 32.301 | 71.487 |  |  |  |
| 3 | 0.855 | 28.513 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only**.**



There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Cereals | 0.728 |
| Oats | -0.524 |
| Chocoes | 0.609 |

As only one component is extracted. The solution cannot be rotated.

The only component explains 39.186 of the variance.

|  |
| --- |
| Breakfast |
| Cereals  Oats  Chocoes |

## Part 12

**Biscuits(Biscuits,Rusks,Cookies,Karachi Biscuits)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Biscuits | 4.44 | 0.694 | 207 |
| Rusks | 2.59 | 1.132 | 207 |
| CookiesBiscuits | 1.30 | 0.660 | 207 |
| Karachi Biscuits | 1.37 | 0.684 | 207 |

Rusks depict the maximum variance.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.644 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 172.103 |
| df | 6 |
| Sig. | 0.000 |

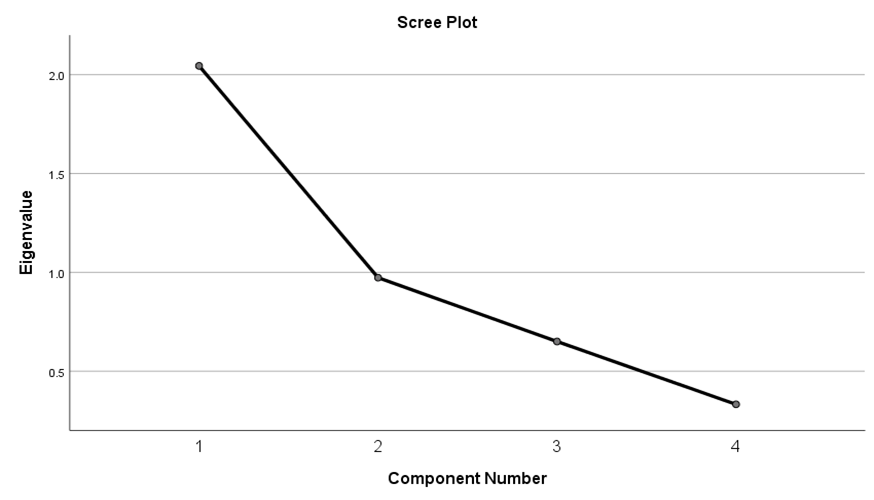
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Biscuits | 1.000 | 0.066 |
| Rusks | 1.000 | 0.513 |
| CookiesBiscuits | 1.000 | 0.719 |
| Karachi Biscuits | 1.000 | 0.748 |

Karachi Biscuits are preferred the most by people.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.044 | 51.112 | 51.112 | 2.044 | 51.112 | 51.112 |
| 2 | 0.973 | 24.324 | 75.437 |  |  |  |
| 3 | 0.650 | 16.255 | 91.691 |  |  |  |
| 4 | 0.332 | 8.309 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only**.**



There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Biscuits | 0.256 |
| Rusks | 0.716 |
| CookiesBiscuits | 0.848 |
| Karachi Biscuits | 0.865 |

As only one component is extracted. The solution cannot be rotated.

The only component explains 51.12 of the variance.

|  |
| --- |
| Breakfast |
| Biscuits  Rusks  Cookies  Karachi Biscuits |

## Part 13

**Healthy food(Vegetables, Fruits, Rice, Ragi, Wheat, Jowar, Barley, Millets, Legumes, Dals,Soya, Soya\_Products, Full\_fat\_milk, Toned\_milk, Full\_fat\_yogurt, Packed\_yogurt, Paneer, Curd,Cashew\_Nuts, Raisins, Almond, Pumkin seeds, Anjeer, Walnuts, Peanuts, Dry apricots, Other\_dry\_fruits)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Vegetables | 3.30 | 0.557 | 207 |
| Fruits | 1.77 | 0.578 | 207 |
| Rice | 4.40 | 0.629 | 207 |
| Ragi | 1.24 | 0.555 | 207 |
| Wheat | 1.86 | 0.766 | 207 |
| Jowar | 1.69 | 0.633 | 207 |
| Barley | 1.01 | 0.120 | 207 |
| Millets | 1.01 | 0.098 | 207 |
| Legumes | 1.36 | 0.492 | 207 |
| Dals | 2.55 | 0.628 | 207 |
| Soya | 1.02 | 0.170 | 207 |
| Soya\_Products | 1.30 | 0.490 | 207 |
| Full\_fat\_milk | 1.64 | 1.431 | 207 |
| Toned\_milk | 4.37 | 1.391 | 207 |
| Full\_fat\_yogurt | 2.15 | 1.673 | 207 |
| Packed\_yogurt | 3.14 | 1.720 | 207 |
| Paneer | 1.40 | 0.709 | 207 |
| Curd | 1.74 | 1.371 | 207 |
| Cashew\_Nuts | 1.56 | 1.012 | 207 |
| Raisins | 1.49 | 0.965 | 207 |
| Almond | 1.64 | 1.056 | 207 |
| Pumpkin seeds | 1.06 | 0.395 | 207 |
| Anjeer | 1.19 | 0.684 | 207 |
| Walnuts | 1.20 | 0.693 | 207 |
| Peanuts | 4.28 | 1.060 | 207 |
| Dry apricots | 1.09 | 0.414 | 207 |
| Other\_dry\_fruits | 2.24 | 0.933 | 207 |

Packed Yogurt has maximum variance(/standard deviation)

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.624 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2539.541 |
| df | 351 |
| Sig. | 0.000 |

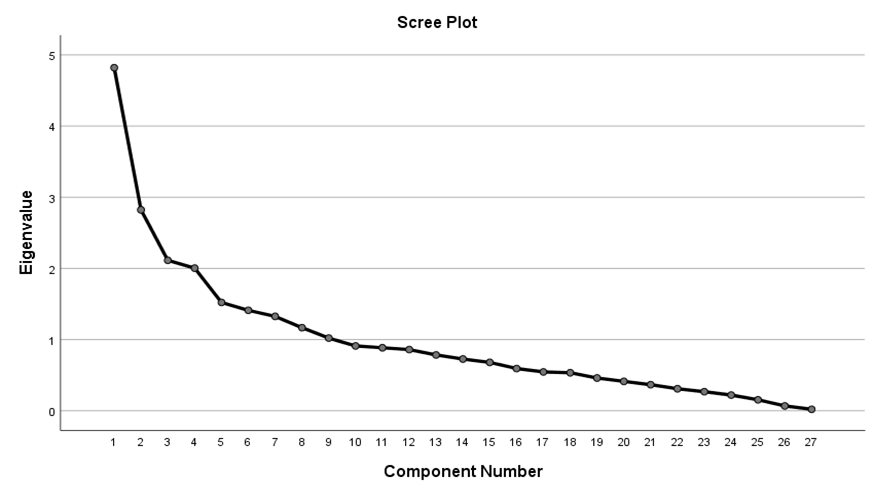
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Vegetables | 1.000 | 0.523 |
| Fruits | 1.000 | 0.531 |
| Rice | 1.000 | 0.714 |
| Ragi | 1.000 | 0.664 |
| Wheat | 1.000 | 0.629 |
| Jowar | 1.000 | 0.587 |
| Barley | 1.000 | 0.518 |
| Millets | 1.000 | 0.766 |
| Legumes | 1.000 | 0.676 |
| Dals | 1.000 | 0.545 |
| Soya | 1.000 | 0.618 |
| Soya\_Products | 1.000 | 0.560 |
| Full\_fat\_milk | 1.000 | 0.916 |
| Toned\_milk | 1.000 | 0.900 |
| Full\_fat\_yogurt | 1.000 | 0.734 |
| Packed\_yogurt | 1.000 | 0.785 |
| Paneer | 1.000 | 0.380 |
| Curd | 1.000 | 0.750 |
| Cashew\_Nuts | 1.000 | 0.777 |
| Raisins | 1.000 | 0.778 |
| Almond | 1.000 | 0.737 |
| Pumpkin Seeds | 1.000 | 0.783 |
| Anjeer | 1.000 | 0.692 |
| Walnuts | 1.000 | 0.571 |
| Peanuts | 1.000 | 0.680 |
| Dry Apricots | 1.000 | 0.609 |
| Other\_dry\_fruits | 1.000 | 0.783 |

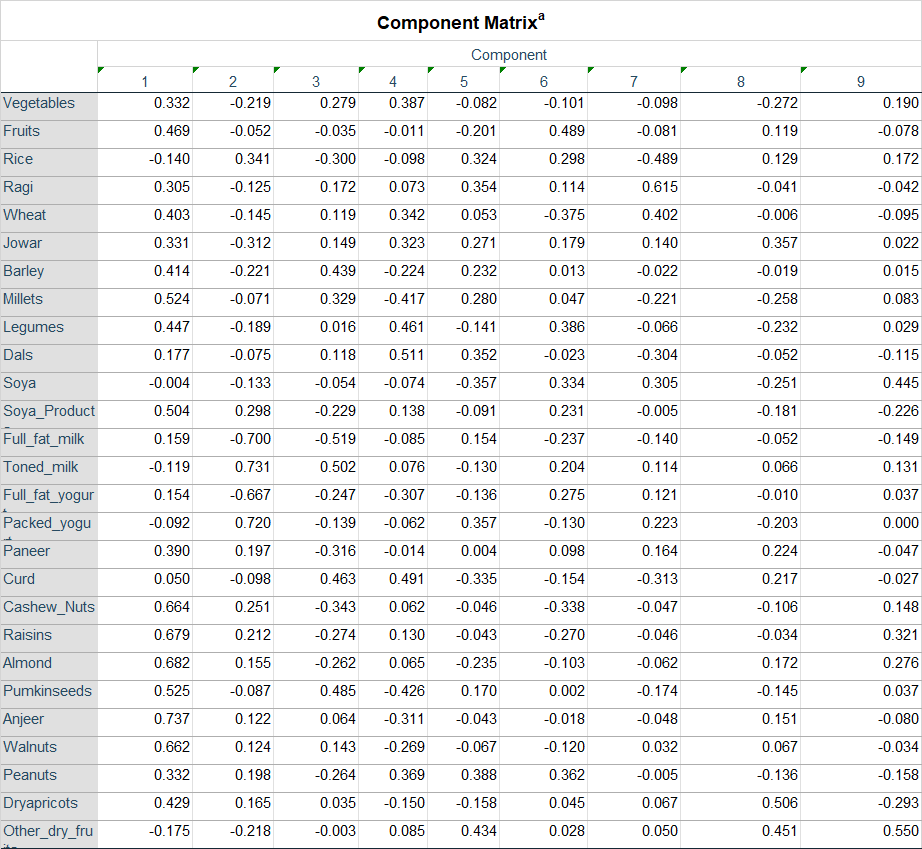
Full Fat milk and toned milk are widely preferred.

Full Fat milk is the most preferred.





The variables can be put in 9 components.



**Rotated Component Matrix**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Component | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Vegetables | 0.214 | 0.213 | 0.007 | 0.220 | 0.480 | -0.291 | 0.250 | 0.080 | -0.004 |
| Fruits | 0.101 | 0.188 | 0.064 | 0.375 | 0.159 | 0.464 | -0.114 | 0.289 | -0.061 |
| Rice | 0.056 | -0.039 | -0.100 | 0.215 | -0.237 | -0.014 | -0.723 | -0.162 | 0.219 |
| Ragi | -0.060 | 0.223 | -0.036 | 0.272 | -0.266 | 0.084 | 0.628 | 0.105 | 0.230 |
| Wheat | 0.279 | 0.031 | 0.057 | 0.081 | 0.130 | 0.011 | 0.707 | -0.148 | 0.039 |
| Jowar | -0.023 | 0.117 | 0.121 | 0.338 | 0.219 | 0.250 | 0.306 | -0.053 | 0.486 |
| Barley | -0.031 | 0.670 | 0.048 | 0.032 | 0.094 | 0.074 | 0.187 | -0.022 | 0.125 |
| Millets | 0.130 | 0.853 | 0.050 | 0.091 | -0.058 | -0.045 | -0.061 | 0.005 | -0.037 |
| Legumes | 0.154 | 0.071 | 0.079 | 0.630 | 0.383 | 0.005 | 0.114 | 0.281 | -0.072 |
| Dals | 0.033 | 0.074 | 0.033 | 0.469 | 0.299 | -0.188 | 0.045 | -0.414 | 0.139 |
| Soya | 0.045 | -0.048 | -0.033 | 0.001 | -0.011 | -0.121 | 0.043 | 0.772 | 0.009 |
| Soya\_Products | 0.347 | 0.037 | -0.071 | 0.496 | -0.087 | 0.246 | 0.002 | 0.037 | -0.342 |
| Full\_fat\_milk | 0.113 | -0.002 | 0.938 | 0.017 | -0.037 | -0.069 | 0.045 | -0.115 | 0.027 |
| Toned\_milk | -0.068 | 0.021 | -0.937 | -0.015 | 0.022 | 0.088 | -0.056 | 0.063 | -0.034 |
| Full\_fat\_yogurt | -0.120 | 0.140 | 0.656 | 0.002 | -0.029 | 0.177 | 0.032 | 0.483 | 0.059 |
| Packed\_yogurt | 0.192 | -0.089 | -0.491 | 0.069 | -0.627 | -0.180 | 0.002 | -0.248 | -0.092 |
| Paneer | 0.356 | -0.064 | 0.017 | 0.187 | -0.214 | 0.404 | 0.043 | 0.026 | 0.043 |
| Curd | -0.009 | -0.048 | -0.186 | -0.030 | 0.819 | 0.036 | 0.079 | -0.182 | 0.012 |
| Cashew\_Nuts | 0.843 | 0.103 | 0.063 | 0.116 | -0.072 | 0.042 | 0.075 | -0.090 | -0.135 |
| Raisins | 0.860 | 0.118 | 0.009 | 0.135 | 0.023 | 0.025 | 0.064 | 0.004 | 0.035 |
| Almond | 0.775 | 0.077 | 0.030 | 0.089 | 0.131 | 0.290 | -0.018 | 0.132 | 0.046 |
| Pumpkin Seeds | 0.074 | 0.875 | -0.019 | -0.025 | 0.060 | 0.065 | 0.028 | 0.001 | -0.043 |
| Anjeer | 0.429 | 0.516 | 0.017 | 0.041 | -0.026 | 0.477 | 0.043 | -0.021 | -0.102 |
| Walnuts | 0.413 | 0.490 | -0.046 | -0.036 | -0.006 | 0.346 | 0.150 | -0.007 | -0.123 |
| Peanuts | 0.151 | -0.033 | -0.027 | 0.771 | -0.198 | 0.048 | 0.000 | -0.134 | 0.040 |
| Dry Apricots | 0.170 | 0.116 | -0.067 | -0.033 | 0.035 | 0.732 | 0.075 | -0.127 | -0.031 |
| Other\_dry\_fruits | -0.030 | -0.047 | 0.048 | -0.090 | -0.032 | -0.103 | -0.064 | 0.010 | 0.868 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 |
| Cashews  Almonds  Raisins  Cashew\_nuts  Raisins  Almond | Barley  Millets  Pumpkin\_seeds  Anjeer  Walnuts | Full-fat-milk  Tonned\_milk  Full\_fat\_yogurt | Legumes  Dal  Soya-Products  Peanuts | Vegetables  Packed\_yogurt  Curd | Fruits  Paneer  Dry\_apricots | Rice  Ragi  Wheat | Soya | Jowar  Other\_dry\_fruits |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nuts | C2 | Milk Products | C4 | C5 | C6 | C7 | Soya | C9 |
| Cashews  Almonds  Raisins  Cashew\_nuts  Raisins  Almond | Barley  Millets  Pumpkin\_seeds  Anjeer  Walnuts | Full-fat-milk  Tonned\_milk  Full\_fat\_yogurt | Legumes  Dal  Soya-Products  Peanuts | Vegetables  Packed\_yogurt  Curd | Fruits  Paneer  Dry\_apricots | Rice  Ragi  Wheat | Soya | Jowar  Other\_dry\_fruits |

Thus, after rotation, Component 1 accounts for 10.985% of the variance;

Component 2 accounts for 9.959% of the variance; Component 3 accounts for 9.375% of the variance ; Component 4 accounts for 7.374% of the variance ; Component 5 accounts for 6.975% of the variance ; Component 6 accounts for 6.356% of the variance ; Component 7 accounts for 6.300% of the variance ; Component 8 accounts for 5.210% of the variance ;

Component 9 accounts for 4.898% of the variance

Both the 9 factors together explain for 67.431 % of the variance of Sweets.

## Part 14

**Non Veg Food (Poultry, Red Meat, Fish, Prawns, Beef)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Poultry | 2.27 | 1.049 | 207 |
| Redmeat | 1.52 | 0.749 | 207 |
| Fish | 1.11 | 0.358 | 207 |
| Prawns | 1.41 | 0.710 | 207 |
| Beef | 1.30 | 0.769 | 207 |

Poultry has the maximum standard deviation.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.722 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 240.287 |
| df | 10 |
| Sig. | 0.000 |

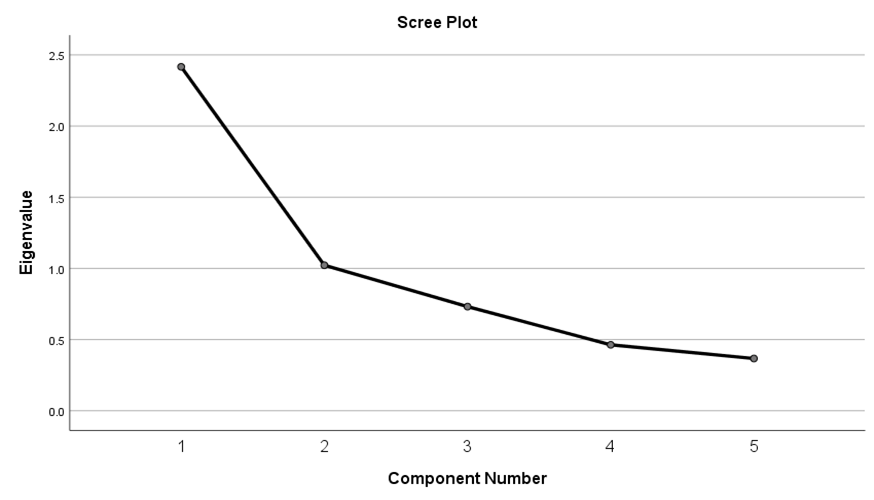
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Poultry | 1.000 | 0.655 |
| Redmeat | 1.000 | 0.732 |
| Fish | 1.000 | 0.449 |
| Prawns | 1.000 | 0.660 |
| Beef | 1.000 | 0.943 |

Beef is the most widely preferred.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.416 | 48.328 | 48.328 | 2.416 | 48.328 | 48.328 | 2.339 | 46.774 | 46.774 |
| 2 | 1.022 | 20.439 | 68.767 | 1.022 | 20.439 | 68.767 | 1.100 | 21.993 | 68.767 |
| 3 | 0.732 | 14.631 | 83.398 |  |  |  |  |  |  |
| 4 | 0.463 | 9.262 | 92.659 |  |  |  |  |  |  |
| 5 | 0.367 | 7.341 | 100.000 |  |  |  |  |  |  |

This shows that the data can be expressed by a two components.



There is a bend after 2 and hence two components are taken.

|  |  |  |
| --- | --- | --- |
| Component Matrixa | | |
|  | Component | |
| 1 | 2 |
| Poultry | 0.805 | 0.082 |
| Redmeat | 0.855 | 0.047 |
| Fish | 0.647 | -0.174 |
| Prawns | 0.735 | -0.345 |
| Beef | 0.281 | 0.929 |

|  |  |  |
| --- | --- | --- |
| Rotated Component Matrix | | |
|  | Component | |
| 1 | 2 |
| Poultry | 0.763 | 0.269 |
| Redmeat | 0.819 | 0.247 |
| Fish | 0.670 | -0.016 |
| Prawns | 0.796 | -0.162 |
| Beef | 0.053 | 0.969 |

|  |  |
| --- | --- |
| Component 1 | Component 2 |
| Poultry  Read Meat  Fish  Prawns | Beef |

|  |  |
| --- | --- |
| Non Veg Products (except Beef) | Beef |
| Poultry  Read Meat  Fish  Prawns | Beef |

Thus, after rotation, Non Veg Products (except Beef) accounts for 46.774% of the variance;

Beef accounts for 21,993% of the variance

Both the 2 factors together explain for 68.767% of the variance of Non veg products.

## Part 15

**Fat Products(Ghee,Butter,Cheese)**

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics | | | |
|  | Mean | Std. Deviation | Analysis N |
| Ghee | 1.84 | 1.150 | 207 |
| Butter | 1.50 | 0.858 | 207 |
| Cheese | 1.43 | 0.803 | 207 |

Ghee has the maximum variance/standard deviation.

|  |  |  |
| --- | --- | --- |
| KMO and Bartlett's Test | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.591 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 177.762 |
| df | 3 |
| Sig. | 0.000 |

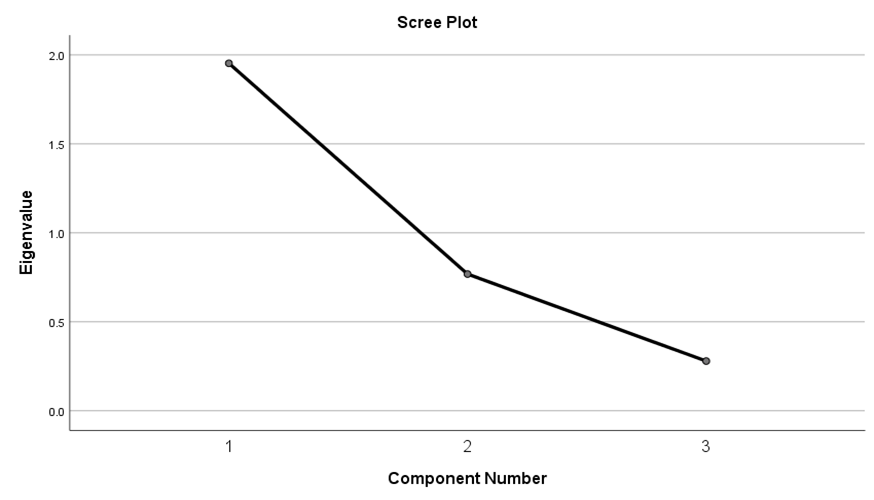
As KMO > 0.5 Factor Analysis can be done

|  |  |  |
| --- | --- | --- |
| Communalities | | |
|  | Initial | Extraction |
| Ghee | 1.000 | 0.382 |
| Butter | 1.000 | 0.780 |
| Cheese | 1.000 | 0.790 |

Cheese is preferred by most of the people.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total Variance Explained | | | | | | |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 1.953 | 65.090 | 65.090 | 1.953 | 65.090 | 65.090 |
| 2 | 0.768 | 25.607 | 90.697 |  |  |  |
| 3 | 0.279 | 9.303 | 100.000 |  |  |  |

This shows that the data can be expressed by a single component only.



There is a bend after 1 and hence only one component is sufficient.

|  |  |
| --- | --- |
| Component Matrixa | |
|  | Component |
| 1 |
| Ghee | 0.618 |
| Butter | 0.883 |
| Cheese | 0.889 |

As only one component is extracted. The solution cannot be rotated.

The only component explains 65.090 of the variance.

**Conclusions:**

Factor Analysis helped to reduce dimensions without reducing the number of variables. With each test the number of components was found. The most preferred food was also analysed. It was also seen that consumption of some food products led to obesity.

## Appendix

**Code for Linear Regression**

import pandas as pd

import numpy as np

from sklearn import preprocessing

from numpy.linalg import inv

from sklearn.linear\_model import LinearRegression

data=pd.read\_excel("Questionnaire \_data.xlsx",header=None)

X=data[[124,125,127,128,129,130,131,132]].iloc[:][1:]

y=data[126].iloc[:][1:]

X.reset\_index(inplace = True, drop = True)

y.reset\_index(inplace = True, drop = True)

X.columns=[0,1,2,3,4,5,6,7]

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.1, random\_state=1)

x\_train.reset\_index(inplace = True, drop = True)

x\_test.reset\_index(inplace = True, drop = True)

y\_train.reset\_index(inplace = True, drop = True)

y\_test.reset\_index(inplace = True, drop = True)

n=x\_train[0].count()

from sklearn.linear\_model import LinearRegression

reg = LinearRegression().fit(x\_train, y\_train)

a=reg.coef\_

b=reg.intercept\_

W=np.array([b,a[0],a[1],a[2],a[3],a[4],a[5],a[6],a[7]])

W=pd.DataFrame(W)

W.to\_csv('Results.csv', mode='a', index = False, header=None)

y\_pred=w0+x\_test[0]\*w1+x\_test[1]\*w2+x\_test[2]\*w3+x\_test[3]\*w4+x\_test[4]\*w5+x\_test[5]\*w6+x\_test[6]\*w7+x\_test[7]\*w8

ybar=y\_test.mean()

n=y\_test.count()

p=8

for i in range(0,n):

y\_pred[i]=round(y\_pred[i],1)

y\_test[i]=round(y\_test[i],1)

SSR=0

SST=0

SSE=0

for i in range (0,n):

SST=SST+(y\_test[i]-ybar)\*(y\_test[i]-ybar)

SSR=SSR+(y\_pred[i]-ybar)\*(y\_pred[i]-ybar)

SSE=SSE+(y\_test[i]-y\_pred[i])\*(y\_test[i]-y\_pred[i])

RR=SSR/SST

Adjusted\_RR=1-((1-RR)\*(n-1))/(n-p-1)

MSR=SSR/p

MSE=SSE/(n-p-1)

F=MSR/MSE

xw1bar=x\_train[0].mean()

xw2bar=x\_train[1].mean()

xw3bar=x\_train[2].mean()

xw4bar=x\_train[3].mean()

xw5bar=x\_train[4].mean()

xw6bar=x\_train[5].mean()

xw7bar=x\_train[6].mean()

xw8bar=x\_train[7].mean()

sumx1=0

sumx2=0

sumx3=0

sumx4=0

sumx5=0

sumx6=0

sumx7=0

sumx8=0

for i in range(0,n):

sumx1=sumx1+(x\_test[0][i]-xw1bar)\*(x\_test[0][i]-xw1bar)

sumx2=sumx2+(x\_test[1][i]-xw2bar)\*(x\_test[1][i]-xw2bar)

sumx3=sumx3+(x\_test[2][i]-xw3bar)\*(x\_test[2][i]-xw3bar)

sumx4=sumx4+(x\_test[3][i]-xw4bar)\*(x\_test[3][i]-xw4bar)

sumx5=sumx5+(x\_test[4][i]-xw5bar)\*(x\_test[4][i]-xw5bar)

sumx6=sumx6+(x\_test[5][i]-xw6bar)\*(x\_test[5][i]-xw6bar)

sumx7=sumx7+(x\_test[6][i]-xw7bar)\*(x\_test[6][i]-xw7bar)

sumx8=sumx8+(x\_test[7][i]-xw8bar)\*(x\_test[7][i]-xw8bar)

sw1=s/np.sqrt(sumx1)

t=w1/sw1

**Code for Chi square**

total=0

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

plt.style.use('seaborn-deep')

array1=data[4][1:].unique()

list1=[]

list2=[]

col=array1.size

chisq=0

r1=data[data[21]==1][0].count()

r2=data[data[21]==2][0].count()

n=data[0].count()

for i in range (0,col):

observedy=data[(data[4]==array1[i]) & (data[21]==1)][0].count()

list1.append(observedy)

observedn=data[(data[4]==array1[i]) & (data[21]==2)][0].count()

list2.append(observedn)

total=total+observedy+observedn

m=data[data[4]==array1[i]][0].count()

expectedy=(m\*r1)/n

expectedn=(m\*r2)/n

chisq=chisq+(((observedy-expectedy)\*\*2)/expectedy)+(((observedn-expectedn)\*\*2)/expectedn)

bar=[2, 3, 0, 1, 4, 8, 6, 5]

means\_frank = (list1)

means\_guido = (list2)

print(total)

print(chisq)

# create plot

fig, ax = plt.subplots()

fig.set\_size\_inches(10, 10)

index = np.arange(8)

bar\_width = 0.4

opacity = 1

rects1 = plt.bar(index, means\_frank, bar\_width,

alpha=opacity,

color='b',

label='Yes')

rects2 = plt.bar(index + bar\_width, means\_guido, bar\_width,

alpha=opacity,

color='g',

label='No')

#plt.xticks(rotation='vertical')

plt.xlabel('Number of children')

plt.ylabel('Frequency')

plt.xticks(index + bar\_width, (2, 3, 0, 1, 4, 8, 6, 5))

plt.legend()

plt.show()