Statistical Analysis of University Survey

PROJECT REPORT

UNDER THE GUIDANCE OF:
DR. MANAS PATRA
HEAD OF THE DEPARTEMNT
DEPT. OF DATA SCIENCE AND ANALYTICS
Central University Of Rajasthan, Ajmer, INDIA

SUBMITTED BY:
DIVYA KUMARI (2018MSBDA020)
ANURAG SHARMA (2018MSBDA004)
ABHISHEK KUMAR (2018MSBDA001)
DEPT. OF DATA SCIENCE AND ANALYTICS

TABLE OF CONTENT:

- # INTRODUCTION:

- **THE AIM OF THIS REPORT IS TO EXAMINE AND ASSESS THE RESULTS OF A SURVEY WHICH WAS GATHERED BY MEANS OF A QUESTIONNAIRE TITLED "UNIVERSITY FEEDBACK FORM" IN WHICH STUDENTS OF CENTRAL UNIVERSITY OF RAJASTHAN WERE ASKED ABOUT THE ACADEMICS, EXTRA CURRICULAR ACTIVITIES AND CAMPUS LIFE IN GENERAL.**

- **THE WHOLE SURVEY CAN BE DIVIDED INTO THREE PARTS:**

1. **DATA COLLECTION**

2. **METHODOLOGIES**

3. **STATISTICAL ANALYSIS**

DATA COLLECTION:

IN ORDER TO COLLECT RELIABLE SAMPLE DATA FOR THE SURVEY, BOTH OFFLINE AND ONLINE (GOOGLE FORMS) QUESTIONNAIRE TITLED 'UNIVERSITY FEEDBACK FORM' WAS PREPARED AND STUDENTS WERE ASKED TO SUBMIT THEIR RESPONSES.

UNIVERSITY FEEDBACK FORM CONSISTED OF A TOTAL OF 21 QUESTIONS SPREAD ACROSS FOUR MAJOR SUBHEADINGS NAMELY, HOSTEL AND MESS FACILITIES, ACADEMICS, ABOUT THE INSTRUCTOR AND OTHERS.

- RATING PARAMETERS

EACH QUESTION HAD FIVE CHOICES RANGING FROM BELOW AVERAGE TO EXCELLENT.

1. BELOW AVERAGE

2. AVERAGE

3. GOOD

4. VERY GOOD

5. EXCELLENT

METHODOLOGIES

1. DATA CLEANING: OUT OF 467 STUDENTS SURVEYED, DATA GIVEN BY 421 STUDENTS WAS USED IN ACTUAL ANALYSIS AS THE REST WERE WEED OUT IN DATA CLEANING PROCESS.
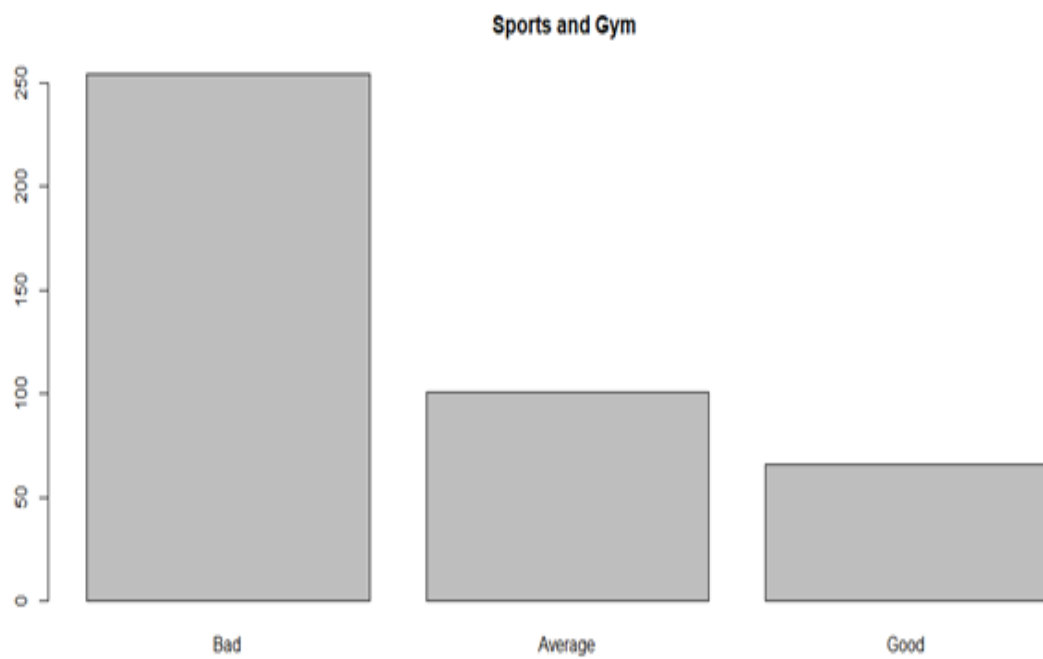
➢ PROCESS:

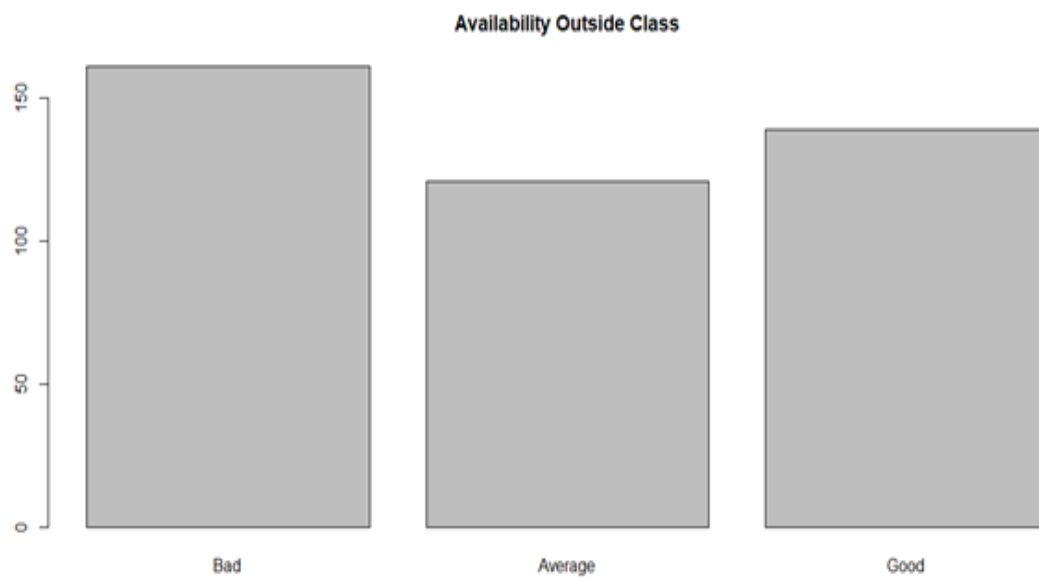MANUAL: DATA WAS CLEANED MANUALLY IN EXCEL FILE BY REMOVING NULL VALUES.
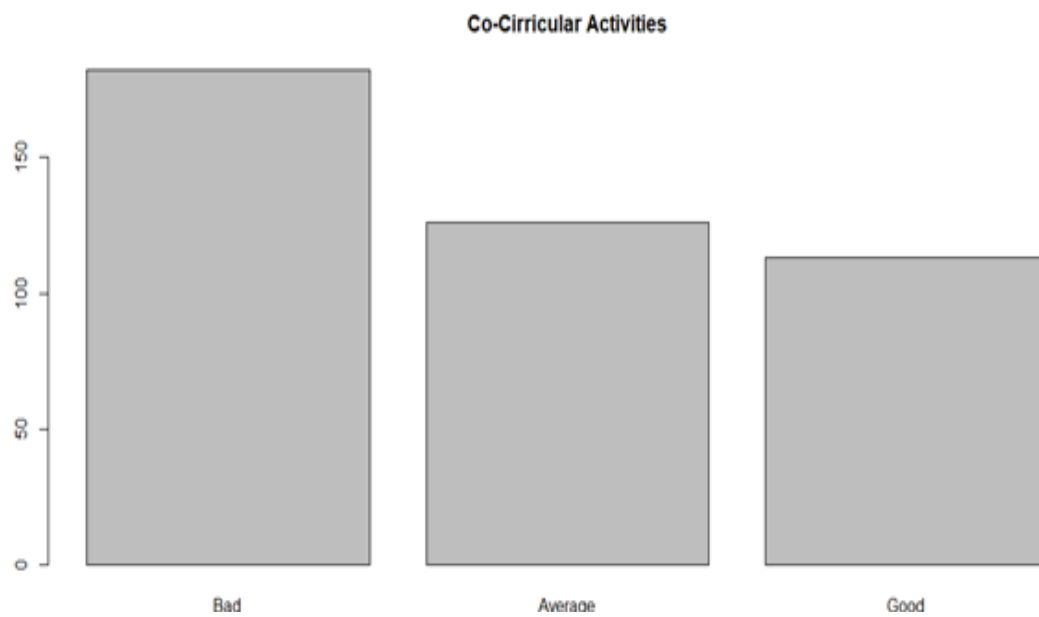
SOFTWARE: DUPLICATE DATA WERE REMOVED WITH HELP OF R SOFTWARE.

2. BAR PLOTS

3. HYPOTHESIS TESTING USING PEARSON'S CHI-SQUARED TEST

STATISTICAL ANALYSIS

1. SOME BAR PLOTS

Sports and Gym

Co-Cirricular Activities

Availability Outside Class

Organisation and Coverage of Syllabus

Food Quality and Variety

Relevance and Real World Application

Level of Interaction During Lectures

Have you experiences Harrasment in any form

Gender Discrimination in University Campus

Security Issues
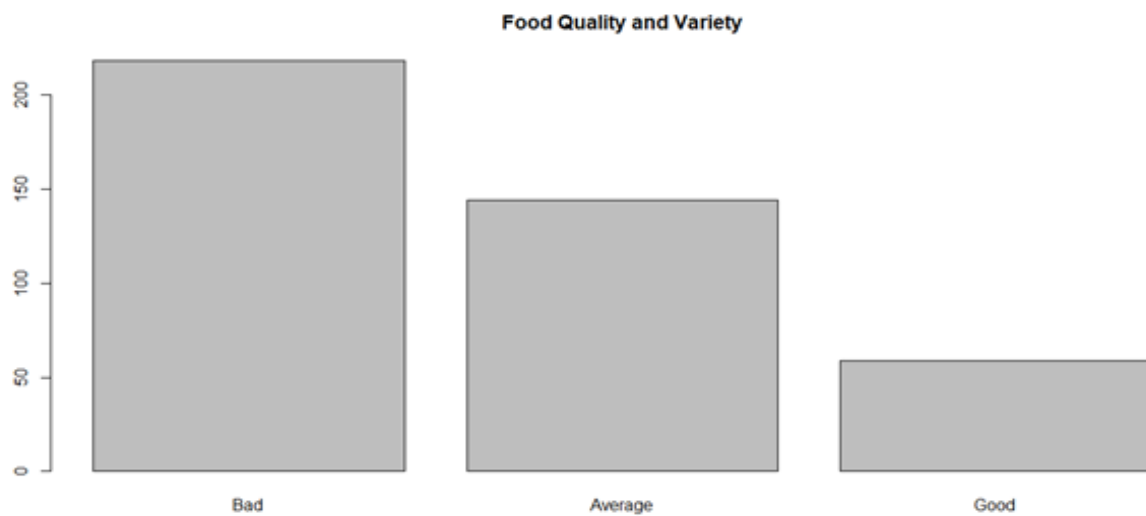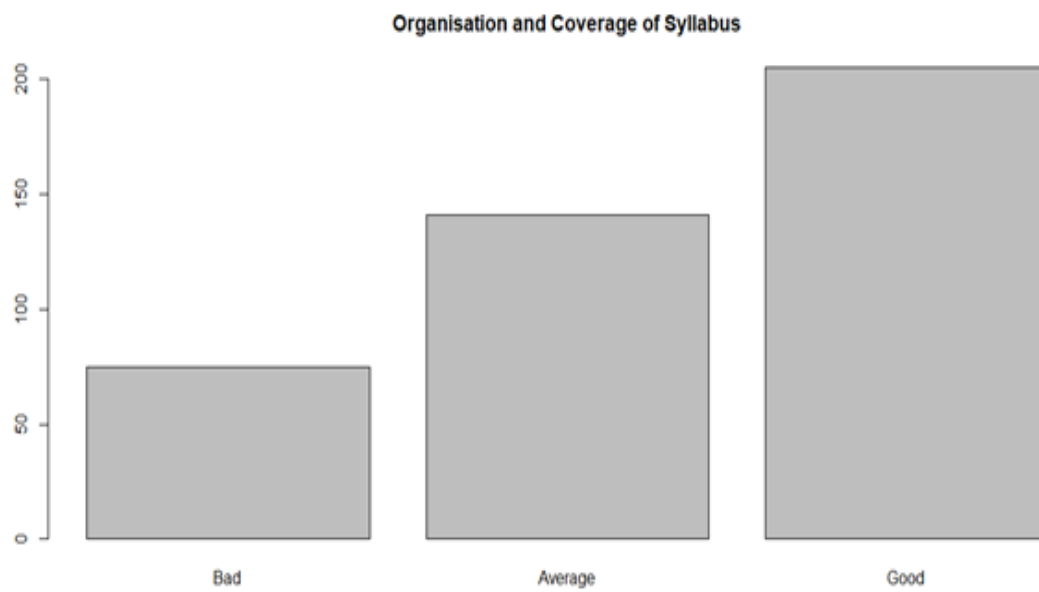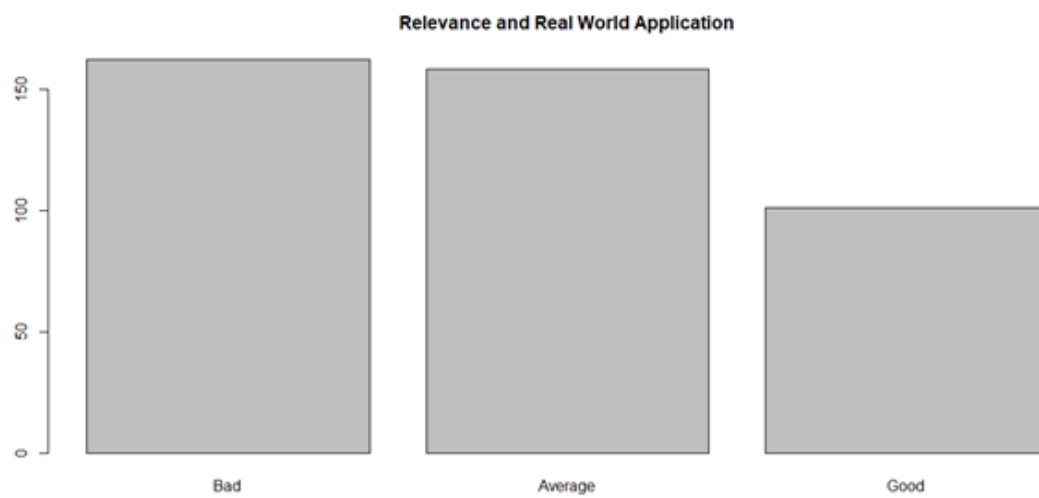
Summary:

```
> summary(file)
       X              ENROLLMENT_NUMBER    GENDER         Q1              Q2
 Min.   :  1.0    2013IMSBMT017:  1    Female:164   Min.   :1.000   Min.   :1.000
 1st Qu.:109.0    2013IMSBMT023:  1    Male  :257   1st Qu.:2.000   1st Qu.:2.000
 Median :223.0    2013imsbt005 :  1                 Median :2.000   Median :3.000
 Mean   :222.2    2013IMSST005 :  1                 Mean   :2.485   Mean   :2.751
 3rd Qu.:336.0    2013phdcms01 :  1                 3rd Qu.:3.000   3rd Qu.:3.000
 Max.   :441.0    2013PHDES04  :  1                 Max.   :5.000   Max.   :5.000
                  (Other)      :415
       Q3              Q4              Q5              Q6             Q7
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.000
 1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.00   1st Qu.:3.000
 Median :2.000   Median :2.000   Median :3.000   Median :2.00   Median :4.000
 Mean   :2.399   Mean   :1.922   Mean   :2.696   Mean   :2.29   Mean   :3.487
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:3.00   3rd Qu.:4.000
 Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.00   Max.   :5.000


       Q8              Q9              Q10             Q11             Q12
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000
 Median :3.000   Median :3.000   Median :3.000   Median :3.000   Median :3.000
 Mean   :3.418   Mean   :2.784   Mean   :3.154   Mean   :2.703   Mean   :2.983
 3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
```
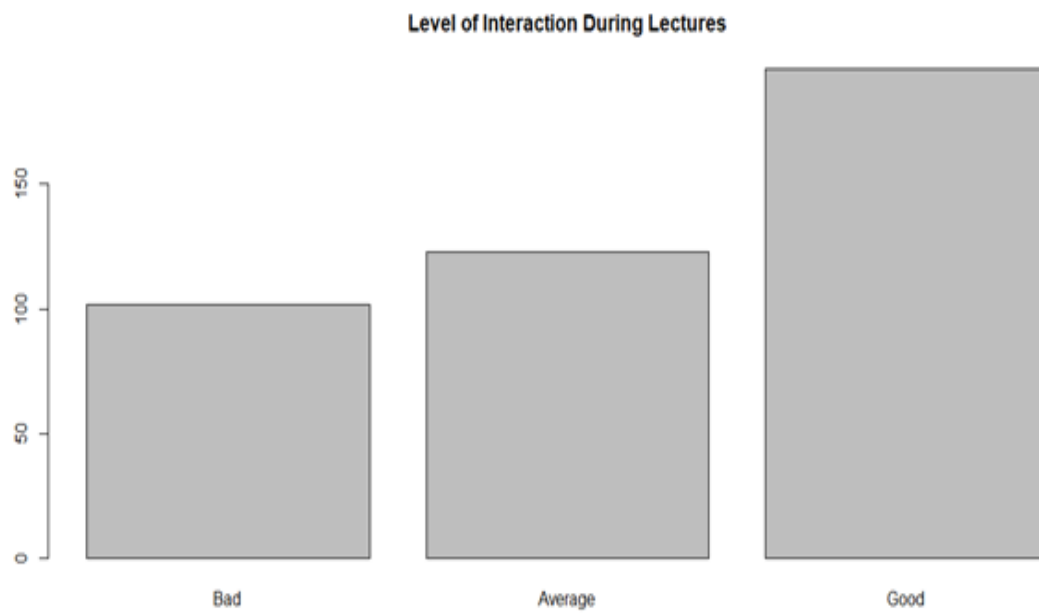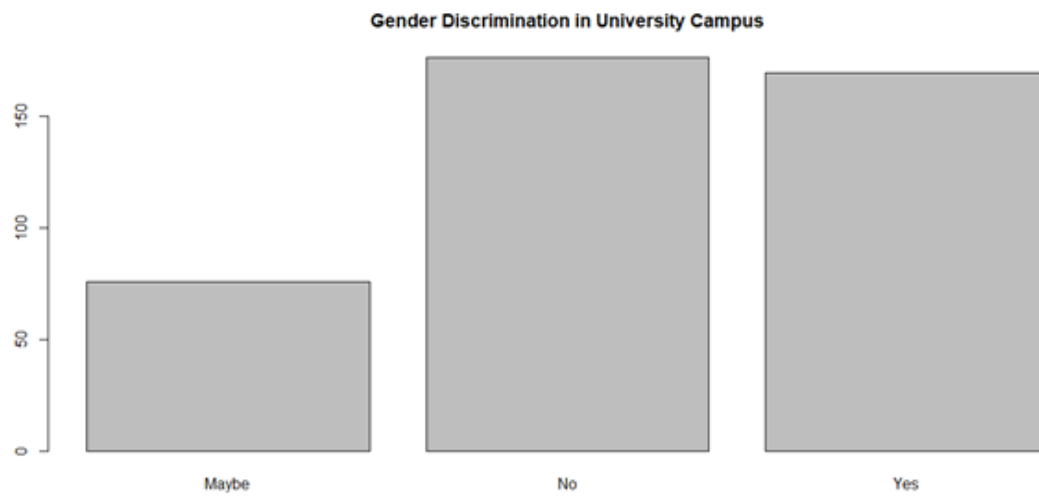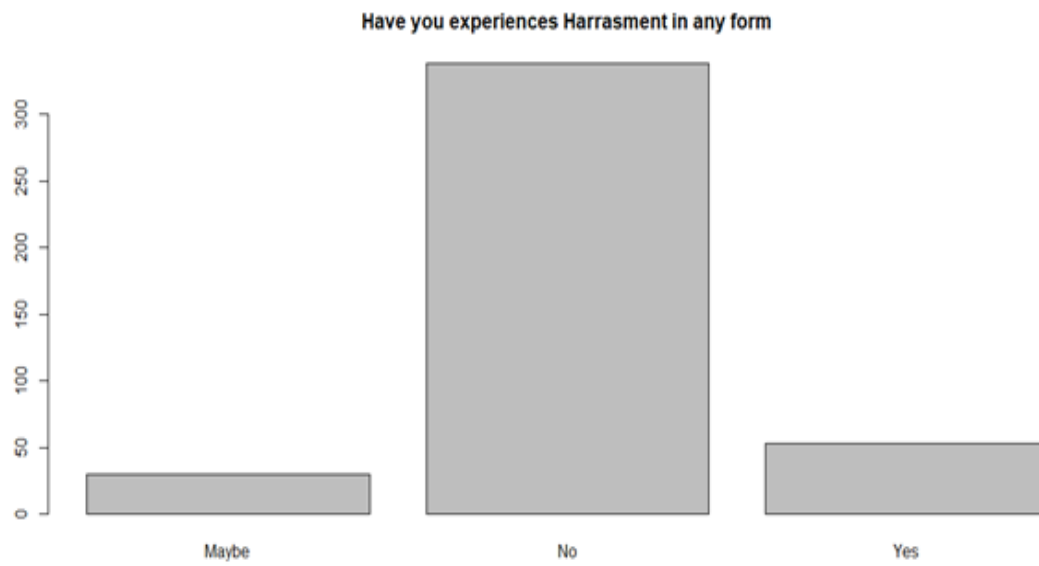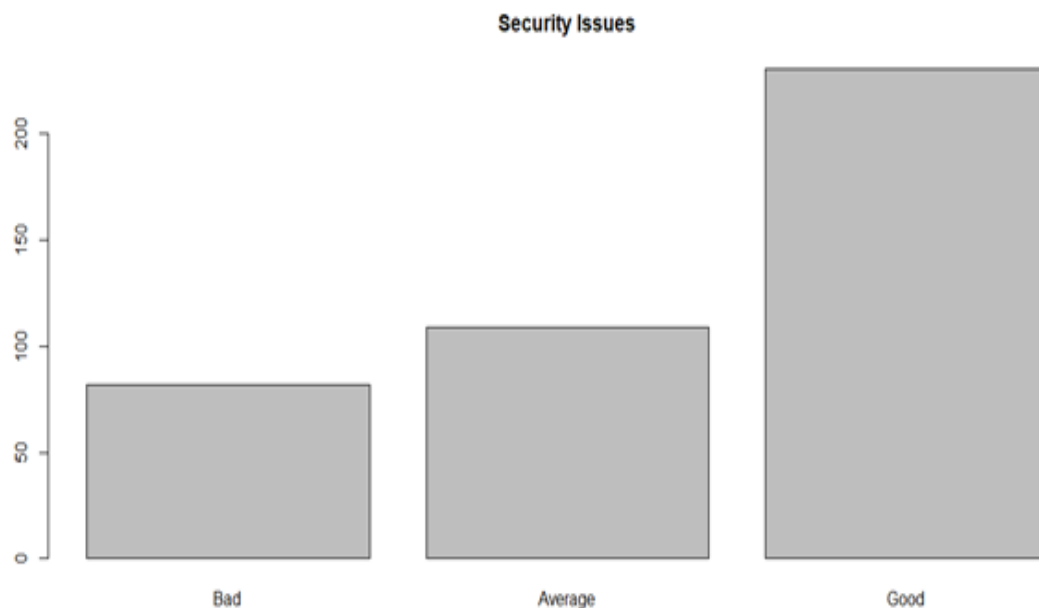
```
      Q13              Q14             Q15              Q16             Q17
 Min.   :1.000   Min.    :1.000   Min.    :1.000   Min.    :1.000   Min.    :1.000
 1st Qu.:3.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
 Median :3.000   Median :3.000   Median :3.000   Median :3.000   Median :3.000
 Mean   :3.292   Mean   :2.891   Mean   :2.979   Mean   :3.031   Mean   :2.696
 3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :5.000   Max.    :5.000   Max.    :5.000   Max.    :5.000   Max.    :5.000

    Q18             Q19             Q20             Q21
 Maybe: 62     Maybe: 76     Maybe: 30     Maybe: 75
 No    :135    No    :176    No    :338    No    :177
 Yes   :224    Yes   :169    Yes   : 53    Yes   :169
```

Gender Wise Data Analysis:

❖ For making the contingency tables there was a need to make categorical variables properly. So, we used some inbuilt functions in R to get the required data.

❖ Here are some excerpts from our code :
library(sqldf)
female=sqldf('Select * from p where Gender="Female"')
male=sqldf('Select* from p where Gender="Male" ')
fq1=table(female['Q1'])
mq1=table(male['Q1']

❖ Following are the counts for the Male and Female for Food Quality :

```
> fq1

  1   2   3   4   5
 14  49  73  27   1
> mq1

  1    2    3    4    5
 53  102   71   23   8
>
```

❖ Null Hypothesis : Food Quality is independent of Gender

tbl=matrix(c(14,53,49,102,73,71,27,23,1,8),nrow = 5,ncol = 2,byrow = T)
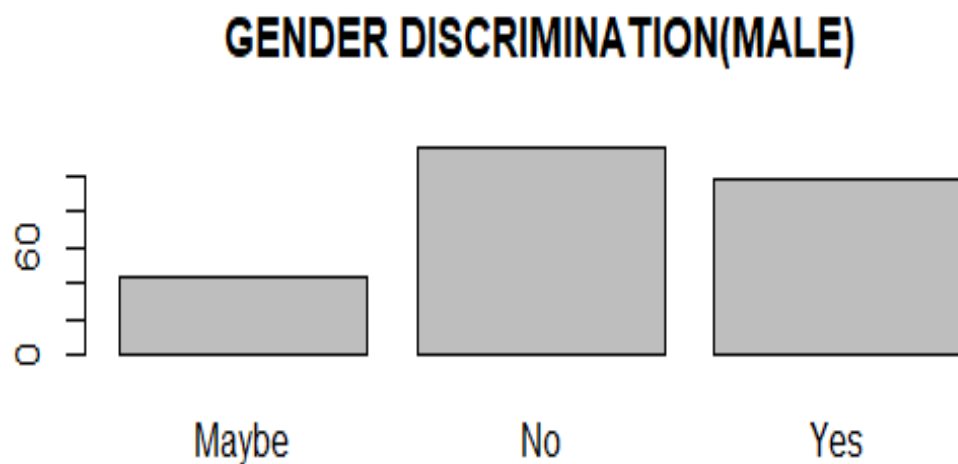
```
> chisq.test(tbl)

        Pearson's Chi-squared test

data:  tbl
X-squared = 209.23, df = 4, p-value < 2.2e-16
```

- ❖ Since , we have taken Significant level , s=0.01
- ❖ And p value from our Chi-squared test, p=2.2e-16 < s , hence it highly signifies that we can reject our null hypothesis.
- ❖ We can say that Food Quality is not Independent of Gender.

GENDER DISCRIMINATION(FEMALE)

GENDER DISCRIMINATION(MALE)
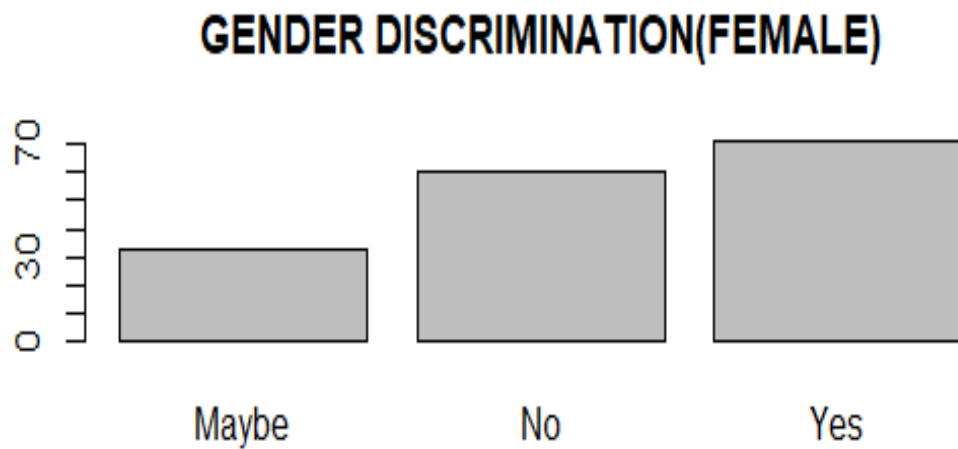
Hypothesis 2 : Security Issues are GENDER INDEPENDENT

 We will perform chi-Squared Test , to check the validity of our Hypothesis .
Following are the code excerpts :

```
> fq7=table(female["Q7"])
> mq7=table(male["Q7"])
> fq7

 1  2  3  4  5
 5  9 47 66 37
> mq7

 1  2  3  4  5
29 39 62 73 54
> tbl7=matrix(c(5,29,9,39,47,62,66,73,37,54),5,2)
> chisq.test(tbl7,correct=F)

        Pearson's Chi-squared test

data:   tbl7
X-squared = 59.15, df = 4, p-value = 4.377e-12
```

Since, p = 4.377e-12 is very less than 0.01, hence it highly signifies that we can reject our Null Hypothesis.

❖ So we can say that, thoughts of Students on Security issued also differ by Gender.
❖ Note: Since chi-Squared approximation may be incorrect, we may not always get the desired result.

Year Wise Data Analysis:

In order to know whether students from different year had different reviews about the questions data was analyzed on basis of year:

Null Hypothesis 1 : Organization and coverage of Syllabus is independent Year Wise .

Following is the Table for year wise rating:

	[18]	[17]	[16]	[15]
[1,]	10	6	5	2
[2,]	19	14	4	6
[3,]	43	46	21	17

| [4,] | 70 | 31 | 11 | 13 |
| [5,] | 40 | 17 | 8 | 3 |

chisq.test(tbl8)

Pearson's Chi-squared test

data: tbl8

X-squared = 21.322, df = 12, p-value = 0.04586

Warning message:

In chisq.test(tbl8) : Chi-squared approximation may be incorrect

- Sine p value is > 0.01 , hence we cannot reject our hypothesis
- So we conclude that organization and coverage of Syllabus is Same throughout previous four years
- **<u>Hypothesis 2: Availability of Books, EBooks, and resources in library does not changes with the year.</u>**

-    [18] [17] [16] [15]

- [1,]   21   12   2   5

- [2,]   33   20   14   7

- [3,]   55   49   19   18

- [4,]   51   23   7   7

- [5,]   22   10   7   4

- > chisq.test(tbl16)

- Pearson's Chi-squared test

- data: tbl16

- X-squared = 14.713, df = 12, p-value = 0.2575

- **Since p value is greater than 0.01 , hence we cannot reject our Hypothesis.**

- **We can conclude that , Availability of Library resources does not changes with year**

Independence Test among Questions:

Null Hypothesis 1: Organization and coverage of syllabus is independent Of relevance and real world application.

> table(unlist(file['Q8']),unlist(file['Q9']))
```
    1  2  3  4  5
 1 16  8  0  0  4
 2 13 23 10  1  0
 3 16 53 61 10  1
 4  9 19 59 39  5
 5  1  4 28 26 15
```
tbl89=table(unlist(file['Q8']),unlist(file['Q9']))
> chisq.test(tbl89)
Pearson's Chi-squared test
Data: tbl89
X-squared = 196.75, df = 16, p-value < 2.2e-16
Since p value is too small, we can reject our Hypothesis. Hence we can say that responses for both the Questions are Dependent

<u>Null Hypothesis 2: Drinking water quality and security provide in the campus is independent.</u>
a=table(unlist(file['Q4']),unlist(file['Q7']))

> a

```
     1   2   3   4    5
 1  24  27  48  46   35
 2   4  13  41  47   25
 3   5   4  15  38   22
 4   0   1   4   8    5
 5   1   3   1   0    4
```

> chisq.test(a)

        Pearson's Chi-squared test

data: a

X-squared = 41.592, df = 16, p-value = 0.0004539

Since p value is less than 0.01 , Hence it highly signifies that we can reject our null hypothesis . So we can conclude that, Security Issues and Drinking Water Quality are Interdependent <u>Null Hypothesis 3: Sports and gym facilities and co-curricular activities are independent.</u>

> a=table(unlist(file['Q6']),unlist(file['Q11']))

> a

```
    1  2  3  4  5
 1 53 24 30 14  5
 2 28 35 38 19  8
```

```
 3  7 23 44 23  4
 4  4  5 12 19 11
 5  0  3  2  2  8
```
> chisq.test (a)

   Pearson's Chi-squared test

Data: a

X-squared = 124.77, df = 16, p-value < 2.2e-16

- **Since p value is incredibly small, so we can easily reject our null Hypothesis.**
- **So we conclude that Sports and Gym facilities are not independent of Co-curricular activities.**
- **Null Hypothesis 4: Level of interaction of Instructor during lectures with students is independent with their availability outside classes**

```
> a=table(unlist(file['Q6']),unlist(file['Q11']))
> a

      1   2   3   4   5
  1  53  24  30  14   5
  2  28  35  38  19   8
  3   7  23  44  23   4
  4   4   5  12  19  11
  5   0   3   2   2   8
> chisq.test(a)

         Pearson's Chi-squared test

 data:   a
 X-squared = 124.77, df = 16, p-value < 2.2e-16
```

Since p value is very small ,we can reject our null hypothesis

- **So we can conclude that Level of interaction of Instructor during lectures with the students is dependent on their availability outside classrooms.**
- **Null hypothesis 5: Gender Discrimination (if any) is independent of the harassment experiences (of any form).**

```
> a=table(unlist(file['Q19']),unlist(file['Q20']))
> a

        Maybe  No Yes
  Maybe     6  66   4
  No        7 160   9
  Yes      17 112  40
> chisq.test(a)

        Pearson's Chi-squared test

data:  a
X-squared = 39.099, df = 4, p-value = 6.647e-08
```
-

  Since p value is less than 0.01 , hence we can reject our null hypothesis .

- **We can conclude that , Gender Discrimination and Harassment are dependent**

Conclusion:

TO CONCLUDE, ON THE BASIS OF THESE FINDINGS, THE SURVEY CLEARLY SHOWS THAT:

- Of the 421 students surveyed majority of them felt that food quality was average or below average.

- A large no. of students felt that sports and gym facilities need to be improved in the university.
- Majority of the students felt secure in the campus.
- A significant percentage of those surveyed expressed contentment with level of interaction during classes.

ON THE BASIS OF GENDER BASED ANALYSIS IT WOULD SEEM THAT:
- FOOD QUALITY IS DEPENDENT ON GENDER.
- SECURITY IS GENDER DEPENDENT.

ON THE BASIS OF YEAR WISE ANALYSIS THIS SURVEY INDICATES THAT:
- STUDENT'S OPINION ON ORGANISATION AND COVERAGE OF SYLLABUS IS INDEPENDENT OF YEAR.
- STUDENT'S OPINION ON AVAILABILITY OF BOOKS, EBOOKS, RESOURCES IN LIBRARY HASN'T CHANGED WITH YEAR.