



Department of Big data analytics

Report on Feedback of CURAJ basic amenities

Guided by: Dr. Manas Patra

Submitted by:

Pawan Bisht

Arkaprabha Majumdar

Pradyumna kumar Sahoo

Table of contents

CONTENTS:

- 2.1 TITLE.....
- 2.2 SUMMARY.....
- 2.3 METHODS.....
- 2.4 RESULTS...
- 2.5 APPENDIX.....

SUMMARY:

This report assess the results of a survey which was gather by means of a questionnaire titled “UNIVERSITY FEEDBACK FORM” in which the students of CURAJ were asked about various aspects of the university and their campus life.

The whole survey can be divided into 3 parts-

1.Data collection

2. Methodologies

3. Statistical analysis.

This report unfolds many curious assumptions and their statistical validations. We have employed chi-square test to test our assumptions as well as prepared some graphs to showcase the various findings.

FORMULAE:

Contingency Table-

	A	B
X	n_{11}	n_{12}
Y	n_{21}	n_{22}

Chi Square test-

O_i =Observed frequency

E_i =Expected frequency

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

DATA COLLECTION:

In order to collect reliable sample data for the survey, both offline and online(Google form)

Questionnaire was prepared and students were asked to submit their responses. University feedback form consists of 21 questions, spread across 4 major sub-headings, namely:

1. Hostel and mess facility

2. Academics

3. about the instructors and others

4. Security

Rating Parameters: Each question has 5 choices ranging from below average to excellent.

METHODS:

Significance Level (α)

The significance level, also denoted as alpha or α , is the probability of rejecting the null hypothesis when it is true.

A significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

We took **0.05** as a **significance level** for our project.

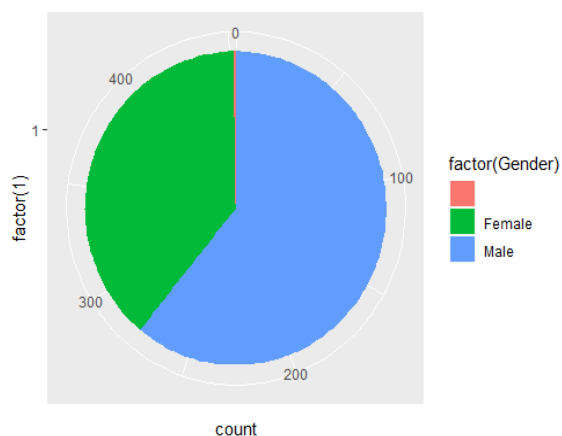
Data Visualization:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

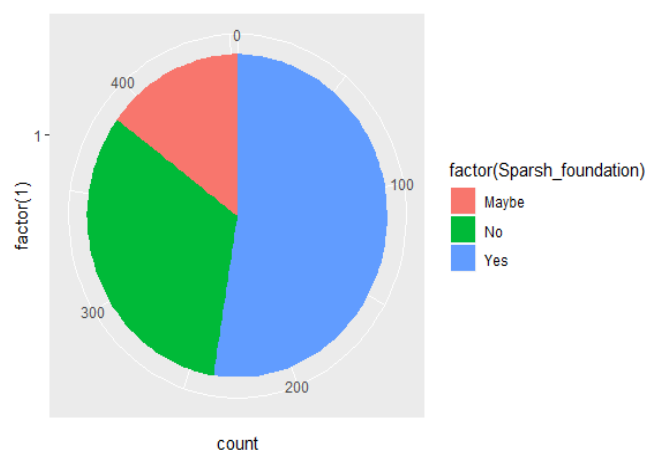
Chi Square Test:

A chi-squared test, also written as **χ^2 test**, is any statistical hypothesis test where the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true. Without other qualification, 'chi-squared test' often is used as short for Pearson's chi-squared test. The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.

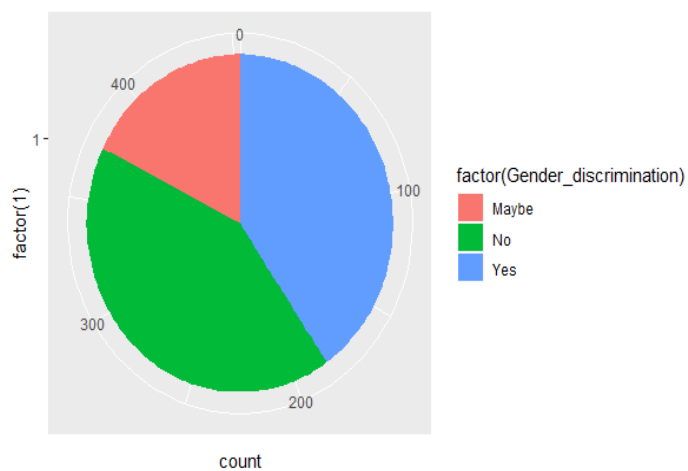
Data visualization:



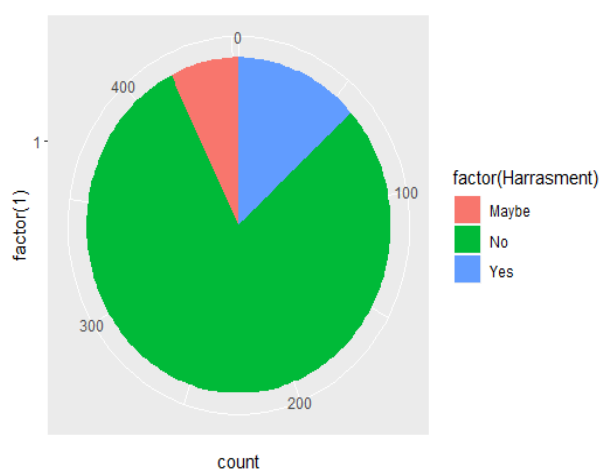
GENDER



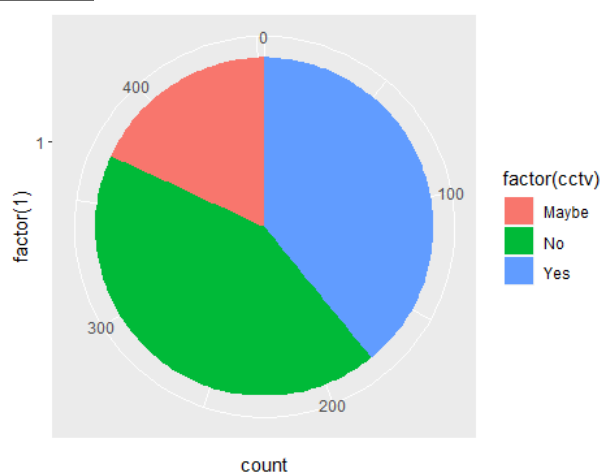
**Awareness about
SPARSH**



Gender Discrimination

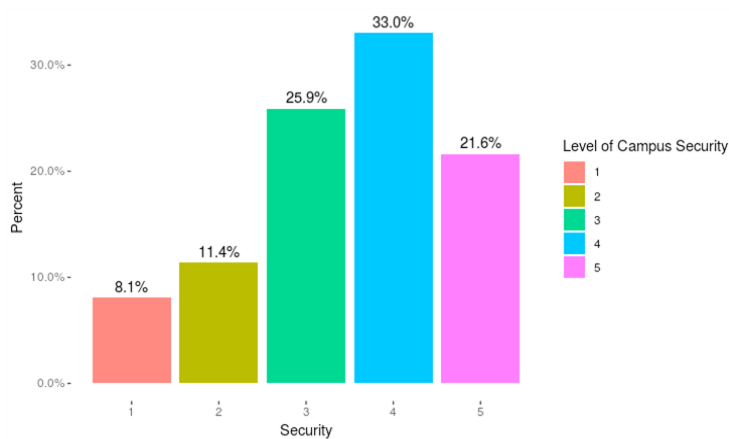


Harassment

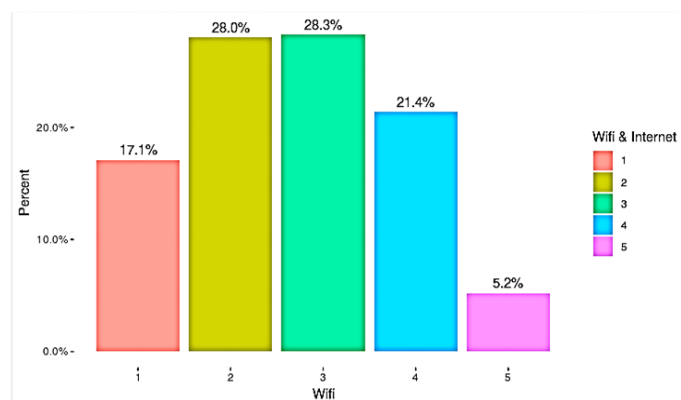


CCTV

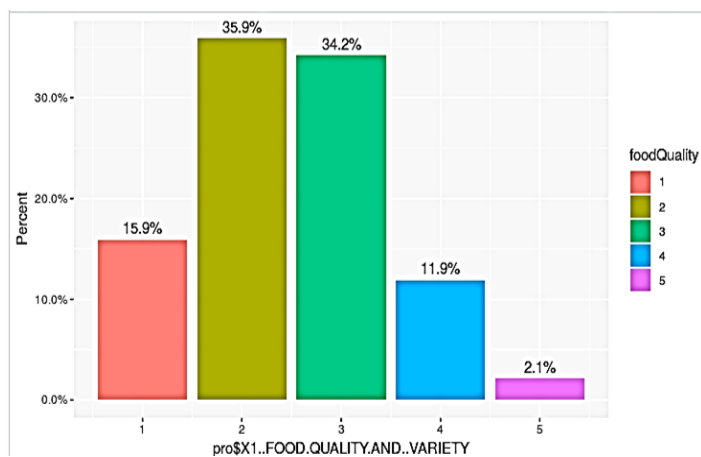
¹ (REF: Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling" (PDF). Philosophical Magazine. Series 5.50: 157–175. doi:10.1080/14786440009463897.)



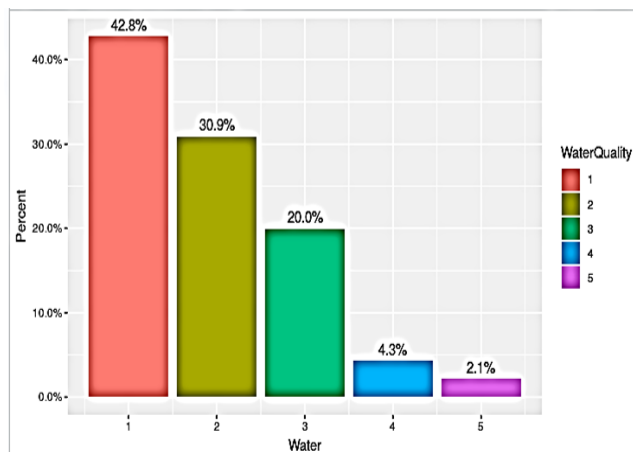
Campus Security



WIFI and Internet services



Food Quality and Services



Water Quality

Analysis of the Fields :

- Food vs Water :

Null Hypothesis : There exist independence between the two variables.

```
> foodWater<-table(food,water)
> foodWater
      water
food  1  2  3  4  5
  1  57 18  9  0  0
  2  76 57 24  6  0
  3  49 49 40  9  5
  4  17 11 18  5  1
  5   3  3  1  0  4
> chisq.test(foodWater)

      Pearson's Chi-squared test

data:  foodWater
X-squared = 112.67, df = 16, p-value < 2.2e-16
```

The value of $p < \text{significance } (0.05)$, so we reject the hypothesis.

- Gender vs Harassment:

Null hypothesis: There exist independence between gender and harassment.

```
> chisq.test(data$GENDER,data$X3..HAVE.YOU.EXPERIENCED.HARASSMENT.IN.ANY.FORM..EX..PHYSICAL..VERBAL..ETC..IN.CURAJ.)

      Pearson's Chi-squared test

data:  data$GENDER and data$X3..HAVE.YOU.EXPERIENCED.HARASSMENT.IN.ANY.FORM..EX..PHYSICAL..VERBAL..ETC..IN.CURAJ.
X-squared = 1.96, df = 2, p-value = 0.3753
```

The value of $p > \text{significance } (0.05)$, so we do not reject the hypothesis.

- Gender vs Knowledge of Sparsh

Null Hypothesis: Both genders know equally about SPARSH foundation and what it does.

```
> chisq.test(data$GENDER,data$X1..ARE.YOU.AWARE.ABOUT.SPARSH.FOUNDATION.AND.I  
TS.ACTIVITIES.)
```

Pearson's Chi-squared test

data: data\$GENDER and data\$X1..ARE.YOU.AWARE.ABOUT.SPARSH.FOUNDATION.AND.ITS
.ACTIVITIES.

X-squared = 7.7863, df = 2, p-value = 0.02038

Considering $\alpha=0.05$, we observe that $p (=0.02038) < \alpha$.

Therefore, we reject the hypothesis.

- Security vs CCTV monitoring:

Null hypothesis: The security in campus does not depend on CCTV monitoring.

```
> g<-ggplot(pro, aes(x=pro$X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS., y=pro$X4..ARE.YOU.  
SATISFIED.WITH.THE.CCTV.MONITORING.IN..CURAJ.)) + geom_point()  
> g  
> chisq.test(pro$X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS.,pro$X4..ARE.YOU.SATISFIED.WIT  
H.THE.CCTV.MONITORING.IN..CURAJ.)
```

Pearson's Chi-squared test

data: pro\$X7..HOW.SECURE.DO.YOU.FEEL.IN.CAMPUS. and pro\$X4..ARE.YOU.SATISFIED.WITH.
THE.CCTV.MONITORING.IN..CURAJ.

X-squared = 37.103, df = 8, p-value = 1.102e-05

Considering $\alpha=0.05$, we observe that $p (=1.102e-05) < \alpha$

Therefore we reject the hypothesis.

- Gender Discrimination vs security:

Null Hypothesis: There exists independence between gender discrimination and security in the campus.

```
Chi-squared approximation may be incorrect
> chisq.test(pro$X2..DO.YOU.FEEL.GENDER.DISCRIMINATION.EXISTS.IN.CURAJ.,pro$X3.QUALITY.OF.GUEST.LECTURES..WORKSHOPS..SEMINARS..ETC)

Pearson's Chi-squared test

data:  pro$X2..DO.YOU.FEEL.GENDER.DISCRIMINATION.EXISTS.IN.CURAJ. and pro$X3.QUALITY.OF.GUEST.LECTURES..WORKSHOPS..SEMINARS..ETC
X-squared = 7.5043, df = 10, p-value = 0.6771

Warning message:
In chisq.test(pro$X2..DO.YOU.FEEL.GENDER.DISCRIMINATION.EXISTS.IN.CURAJ., :
Chi-squared approximation may be incorrect
```

Considering $\alpha=0.05$, we observe that $p (=0.6771) > \alpha$

Therefore, we do not reject the hypothesis.

- Washrooms vs Hygiene

Null hypothesis: There exist independence between the cleanliness of washrooms and hygiene .

```
> chisq.test(pro$X3..CLEANLINESS.OF.WASHROOMS,pro$X2.HYGIENE...MAINTENANCE..OF.MESS)

Pearson's Chi-squared test

data:  pro$X3..CLEANLINESS.OF.WASHROOMS and pro$X2.HYGIENE...MAINTENANCE..OF.MESS
X-squared = 157.1, df = 16, p-value < 2.2e-16

Warning message:
In chisq.test(pro$X3..CLEANLINESS.OF.WASHROOMS, pro$X2.HYGIENE...MAINTENANCE..OF.MESS) :
Chi-squared approximation may be incorrect
> |
```

The value of $p < \alpha (=0.05)$, so we reject the hypothesis.

- Gender vs Wi-Fi and internet services:

Null hypothesis: There is no discrimination in providing internet services on basis of gender.

```
> chisq.test(pro$GENDER,pro$X5..WI.FI.AND.INTERNET.SERVICES)

Pearson's Chi-squared test

data:  pro$GENDER and pro$X5..WI.FI.AND.INTERNET.SERVICES
X-squared = 6.2632, df = 4, p-value = 0.1803
```

The value of $p > \text{significance}$, thus we do not reject the hypothesis.

- Gender vs Washrooms

Null hypothesis: There is no difference in the cleanliness of washrooms for different genders.

```
> chisq.test(pro$GENDER,pro$X3..CLEANLINESS.OF.WASHROOMS)

Pearson's Chi-squared test

data:  pro$GENDER and pro$X3..CLEANLINESS.OF.WASHROOMS
X-squared = 7.4913, df = 8, p-value = 0.4847
```

The value of $p > \text{significance}$,

Therefore, we do not reject the hypothesis.

APPENDIX:

- Chi-Square Test 4
- Hypotheses:
 - Null, Alternate 6
- P-Value 6
- Significance Level 10

THANK YOU