



Mahatma Education Society's  
**Pillai College of Arts, Commerce & Science**  
(Autonomous)  
Affiliated to University of Mumbai  
NAAC Accredited 'A' grade (3 cycles)  
Best College Award by University of Mumbai  
ISO 9001:2015 Certified



MAHATMA EDUCATION SOCIETY'S  
PILLAI COLLEGE OF ARTS, COMMERCE & SCIENCE  
(Autonomous)  
NEW PANVEL

PROJECT REPORT ON  
**“Analysing Student Behaviour”**

IN PARTIAL FULFILLMENT OF  
**MASTERS OF DATA ANALYTICS**

SEMESTER I – 2023-24

PROJECT GUIDE  
**Name: Prof. Sabitha Praveen**

**SUBMITTED BY: SURAJ SHETE**  
**ROLL NO:3143**

# CA-2 Project

## ABOUT THE DATASET

The “**Analysing Student Behaviour**” Dataset contains **19 columns** that provides information about the student department, percentage, stress level, working or not and their financial condition

It contains **84 rows**.

The data is been manually collected with the help of survey.

### **1. Data Pre-processing:**

- Loading the dataset:  
`DF= read.csv("D:\\Statistics\\stats_data.csv", header=T, na.strings="")`  
`head(DF)`
- Missing values were handled using the `na.omit()` function.  
`dim(DF)`  
  
`colnames(DF)`  
  
`#handling missing values`  
`DF_clean <- na.omit(DF)`  
`DF_clean`
- Duplicate values were removed from the dataset.  
`#remove duplicate values`  
`DF_clean <- DF[!duplicated(DF), ]`  
`DF_clean`
- The summary statistics of the dataset were examined.  
`summary(DF)`

```

> dim(DF)
[1] 84 19
> colnames(DF)
[1] "CertificationCourses"
[2] "Gender"
[3] "Department"
[4] "Height"
[5] "Weight"
[6] "X10thMarks"
[7] "X12thMarks"
[8] "GraduationMarks"
[9] "Select.your.hobbies.or.interests"
[10] "How.many.hours.do.you.typically.spend.on.studying.each.day"
[11] "When.do.you.prefer.to.study"
[12] "What.is.your.salary.expectation.for.a.potential.job"
[13] "Do.You.Like.Your.Degree"
[14] "Willingness.to.Pursue.a.Career.Based.on.Your.Degree"
[15] "How.much.time.are.you.spending.on.social.media"
[16] "How.much.time.are.you.willing.to.spend.on.daily.commuting.or
traveling.to.college"
[17] "Select.your.current.stress.level"
[18] "How.would.you.describe.your.current.financial.status"
[19] "Are.you.currently.working.or.open.to.a.part.time.job.while.s
tudying"

```

## 2. Descriptive Statistics:

- Mean, Median, and Measures of Dispersion:
- Mean Graduation Marks:  $(\bar{x} = \text{Percentage Mean})$
- Median Graduation Marks:  $(\text{Median} = \text{Percentage Median})$
- Standard Deviation:  $(\sigma = \text{Percentage standard deviation})$
- Variance:  $(\text{Percentage variance})$
- Coefficient of Variation:  $(\text{Coefficient of Variation} = \frac{\sigma}{\bar{x}})$
- Interquartile Range:  $(\text{Percentage interquartile range})$

```

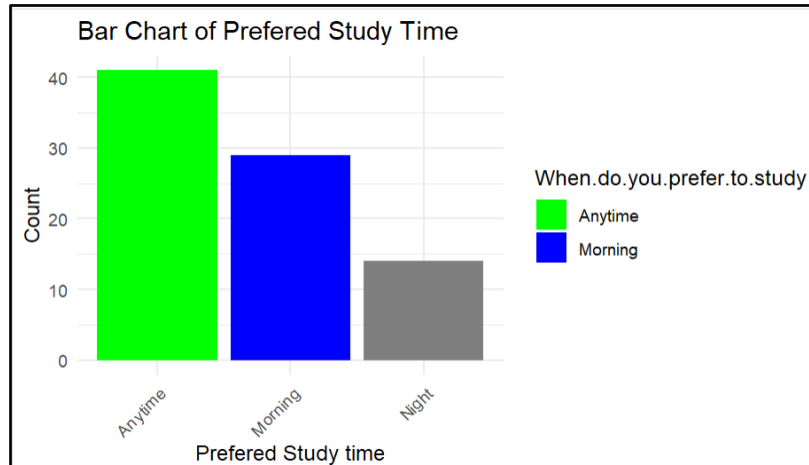
> Percentage_mean<- mean(DF$GraduationMarks, na.rm = TRUE)
> cat("Percentage Mean =", Percentage_mean, "\n")
Percentage Mean = 69.37262
> Percentage_median<- median(DF$GraduationMarks, na.rm = TRUE)
> cat("Percentage Median =", Percentage_median, "\n")
Percentage Median = 70
> Percentage_sd <- sd(DF$GraduationMarks, na.rm = TRUE)
> cat("Percentage standard deviation: ", Percentage_sd, "\n")
Percentage standard deviation: 16.37064
> Percentage_var <- var(DF$GraduationMarks, na.rm = TRUE)
> cat("Percentage variance: ", Percentage_var, "\n")
Percentage variance: 267.9979
> cv_vals <- Percentage_sd / Percentage_mean
> cat("Coefficient of Variation:\n", cv_vals, "\n")
Coefficient of Variation:
0.2359813
> percentage_iqr <- IQR(DF$GraduationMarks, na.rm = TRUE)
> cat("Percentage interquartile range: ", percentage_iqr, "\n")
Percentage interquartile range: 20
>

```

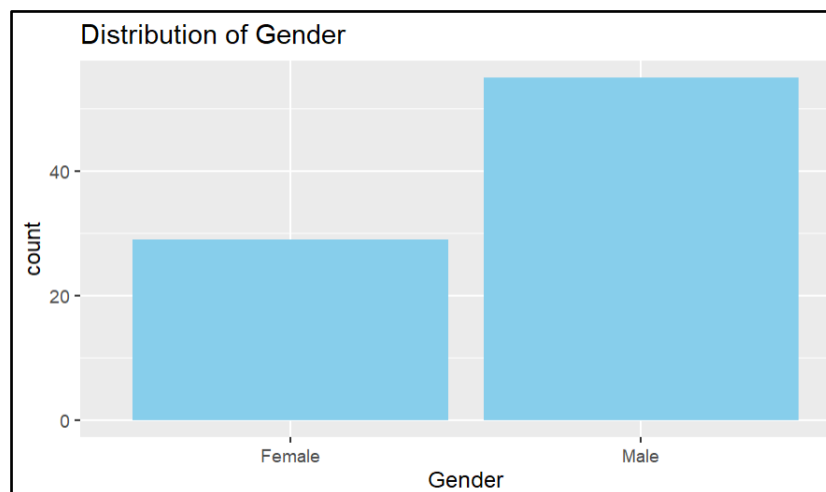
### 3. Data Visualization:

- Bar Chart of Preferred Study Time:

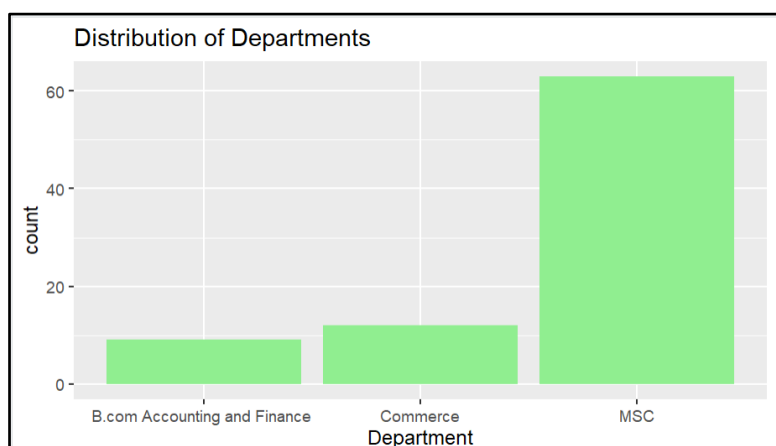
- A bar chart displaying the distribution of preferred study time.
- Majority prefer studying in the morning, followed by anytime and night.



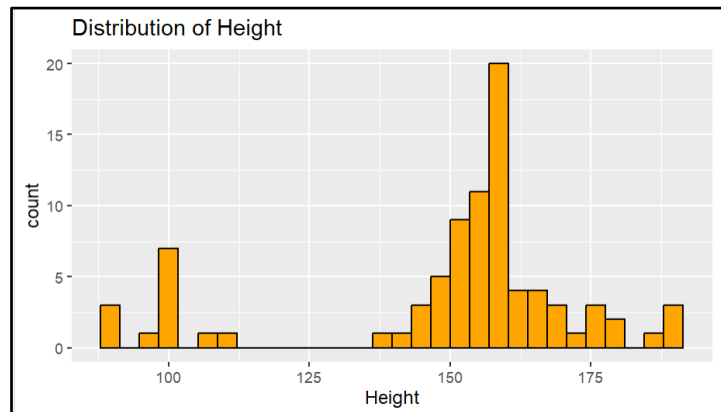
- Bar plot for distribution of Gender:



- Bar Plot showing distribution of departments.



- Histogram Plot for distribution of Height.



#### 4. Hypothesis Testing:

- **Chi-Square Test:**

-Null Hypothesis (H0): There is no significant association between Certification Courses and Willingness to Pursue a Career Based on Your Degree.

-Alternative Hypothesis (H1): There is a significant association between Certification Courses and Willingness to Pursue a Career Based on Your Degree.

```
> print(chi_square_test)

Pearson's Chi-squared test

data:  contingency_table
X-squared = 1.2171, df = 3, p-value = 0.7489
```

**-Result:**  $p\text{-value} = 0.7489 > 0.05$ , accept the null hypothesis suggesting no significant association between Certification Courses and Willingness to Pursue a Career Based on Your Degree.

- **Two-Sample T-Test:**

- Null Hypothesis (H0): There is no difference in the mean salary expectations between male and female respondents.

-Alternative Hypothesis: There is a difference in the mean salary expectations between male and female respondents.

```
Welch Two Sample t-test

data: group1 and group2
t = -0.866, df = 28.103, p-value = 0.3938
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
 -149161.33  60506.15
sample estimates:
mean of x mean of y
 24327.58  68655.17
```

- **Result:** Accept the null hypothesis  $p = 0.3938$  ( $p > 0.05$ ), suggesting no significant difference in the mean salary expectations between male and female respondents.

- **Kruskal-Wallis Test:**

- Null Hypothesis ( $H_0$ ): There is no difference in the median graduation marks among the different departments.

- Alternative Hypothesis ( $H_1$ ): There is a difference in the median graduation marks among the different departments.

```
> print(kruskal_test_result)

Kruskal-Wallis rank sum test

data: DF$GraduationMarks by DF$Department
Kruskal-Wallis chi-squared = 9.894, df = 2,
p-value = 0.007105
```

-**Result:**  $p\text{-value} = 0.007105 < 0.05$ , we reject the null hypothesis so there is a difference in the median graduation marks among the different departments.

## 5. Conclusion:

- The analysis reveals significant associations between various factors such as salary expectations, stress levels, department, and graduation marks.

- Different statistical tests provide insights into the dataset, aiding in understanding patterns and relationships among variables.