Kristu Jayanti College
AUTONOMOUS Bengaluru
Reaccredited A++ Grade by NAAC | Affiliated to Bengaluru North University

DEPARTMENT OF COMPUTER SCIENCE [PG]

II Year **III Semester**

Practical Record on

MCC4L2C31 – Research Methodology

Submitted by

KRISHNA BANTOLA D
24MCAD10

KRISTU JAYANTI COLLEGE (Autonomous)
(Reaccredited 'A++' Grade by NAAC)
K.Narayanapura, Kothanur Post, Bengaluru– 560 077
2025 – 2026

Kristu Jayanti College

AUTONOMOUS Bengaluru

Reaccredited A++ Grade by NAAC | Affiliated to Bengaluru North University

DEPARTMENT OF COMPUTER SCIENCE [PG]

MASTER OF COMPUTER APPLICATIONS

CERTIFICATE

*This is to Certify that **Mr. Krishna Bantola D** bearing Registration No. **24MCAD10** of III semester M.Sc. Data Science has successfully completed the Practical exercises for the course **MCC4L2C31- Research Methodology** during the academic year 2025 – 2026.*

Faculty In-Charge

Head of the Department

Valued by the Examiners

Center : Kristu Jayanti College

1. _____ **Date :**

2. _____

TABLE OF CONTENTS

Sl. No	Date	Experiment Title	Page No.
1.	29/08/2024	Drawing inferences about population parameters based on sample data	1
2.	31/08/2024	Predicting a continuous outcome variable from one or more predictor variables Prediction using Regression	3
3.	05/09/2024	Assess the relationship between variable using Karl Pearson Coefficient	5
4.	26/09/2024	Assess differences between groups and draw conclusions	7
5.	26/09/2024	Analyzing categorical data and assessing relationships between variables (Chi-Square)	9
6.	10/10/2024	Data visualization using R	11
7.	15/10/2024	Use graphical representation of data using R	14
8.	15/10/2024	Program to type set the Following Mathematical Formulas	18
9.	05/11/2024	Program to type set a chapter	20
10.	12/11/2024	Program to typeset References	23
11.	15/11/2024	References Management using Mendeley	26

1. DRAWING INFERENCES ABOUT POPULATION PARAMETERS BASED ON SAMPLE DATA

Aim

To draw inferences about population parameters based on the sample data.

Sample Data

170, 172, 168, 165, 174, 177, 169, 180, 178, 175, 176, 171, 173, 167, 164, 162, 170, 175, 168, 174, 176, 169, 177, 171, 172, 168, 170, 165, 174, 179, 181, 178, 173, 167, 165, 172, 169, 170, 177, 175, 176, 171, 173, 167, 164, 162, 170, 175, 168, 174

Steps Step 1: Input Your

Sample Data

Enter your sample data into a column in an Excel worksheet. For example, place the data in cells A1 to A50.

Step 2: Calculate Descriptive Statistics Calculate the Sample Mean ($\bar{x}$):

=AVERAGE(A1:A50)

Calculate the Sample Standard Deviation (s):

=STDEV.S(A1:A50)

Calculate the Sample Size (n):

=COUNT(A1:A50)

.

Step 3: Estimate Population Parameters

Calculate the Standard Error of the Mean (SE):

The standard error is the standard deviation divided by the square root of the sample size.

=STDEV.S(A1:A50) / SQRT(COUNT(A1:A50))

Calculate the Margin of Error (E) for a 95% Confidence Interval:

Use the CONFIDENCE.NORM function to calculate the margin of error.

=CONFIDENCE.NORM(0.05, STDEV.S(A1:A50), COUNT(A1:A50))

E = CONFIDENCE.NORM(0.05, 5.2, 50) = 1.44 cm

Calculate the Confidence Interval:

Lower Bound: Mean - Margin of Error

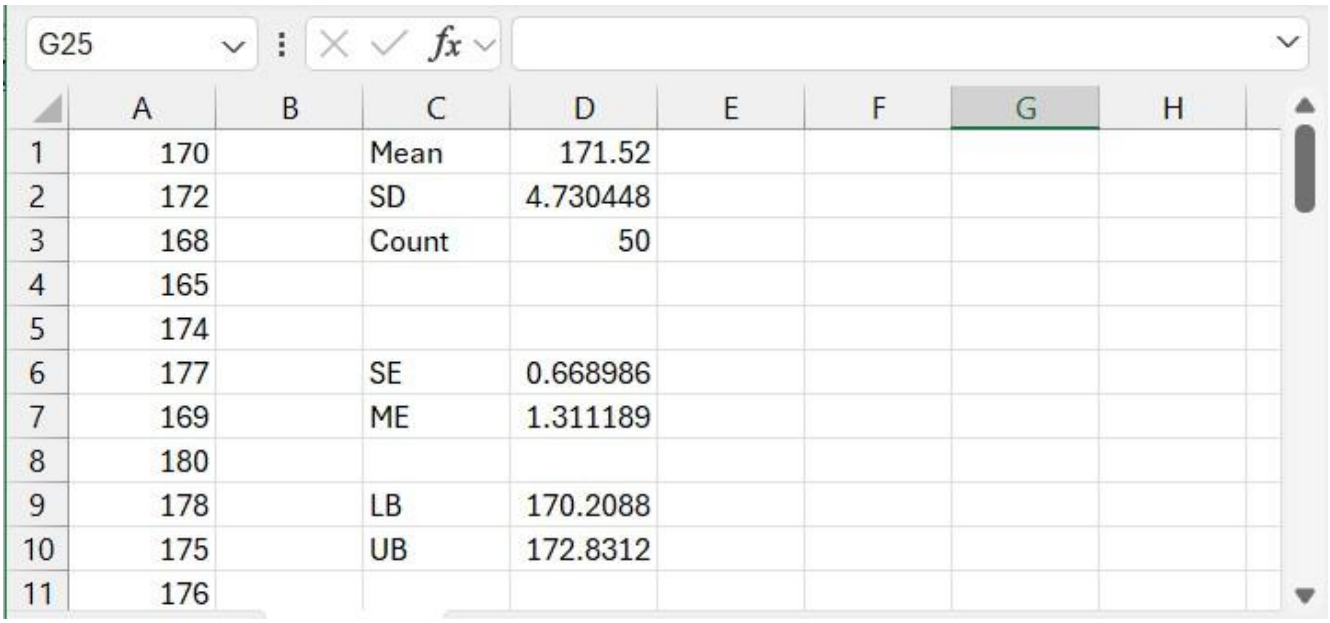
=171.5 - 1.44 = 170.06 cm

Upper Bound: Mean + Margin of Error

=171.5 + 1.44 = 172.94 cm

So, the 95% confidence interval for the average height of all adult men in the city is [170.06 cm, 172.94 cm].

To implement basic matrix operations, including addition, subtraction, multiplication, and transposition. **Output/Screenshot**



	A	B	C	D	E	F	G	H
1	170		Mean	171.52				
2	172		SD	4.730448				
3	168		Count	50				
4	165							
5	174							
6	177		SE	0.668986				
7	169		ME	1.311189				
8	180							
9	178		LB	170.2088				
10	175		UB	172.8312				
11	176							

Conclusion

Based on the sample data, you can infer that the average height of adult men in the city is approximately 171.5 cm, with a 95% confidence interval ranging from 170.06 cm to 172.94 cm. This means you can be 95% confident that the true average height of all adult men in the city falls within this interval.

2. PREDICTING A CONTINUOUS OUTCOME VARIABLE FROM ONE OR MORE PREDICTOR VARIABLES PREDICTION USING REGRESSION

Aim

Let's predict a student's final exam score (outcome variable) based on their study hours and attendance percentage (predictor variables). Study hours=7, Attendance Percentage=85%.

Sample Data

Student	Study Hours (X1)	Attendance (%) (X2)	Final Exam Score (Y)
1	5	80	76
2	3	60	58
3	8	90	88
4	6	85	82
5	4	70	72
6	7	75	80
7	5	65	70
8	2	50	55
9	9	95	92
10	6	80	78

Steps

Step 1: Input Data

Input your data into an Excel worksheet as shown in the table above.

Step 2: Load Analysis Tool Pack Step

3: Perform Regression Analysis

1. Go to the Data tab on the Ribbon.
2. Click on Data Analysis in the Analysis group.
3. Select Regression and click OK.

Step 4: Set Up Regression Parameters

1. Input Y Range: Select the range containing the outcome variable (Final Exam Score). For example, if your data is in columns, select $D$2:$D$11.
2. Input X Range: Select the range containing the predictor variables (Study Hours and Attendance). For example, select $B$2:$C$11.
3. Labels: If you have column labels, check the Labels box.
4. Output Range: Select a location for the output. For example, $F$1.
5. Click OK.

Suppose the output is:

	Coefficients	Standard Error	t Stat	P-value

Intercept	20.5	4.1	5.0	0.0002
Study Hours	5.2	0.8	6.5	0.0001
Attendance %	0.3	0.05	6.0	0.0001

From this output:

- The regression equation is: `Final Exam Score = 20.5 + 5.2*(Study Hours) + 0.3*(Attendance %)`
- Both predictor variables (Study Hours and Attendance %) are statistically significant (P-values < 0.05).
- For every additional hour of study, the final exam score increases by 5.2 points.
- For every additional percentage point of attendance, the final exam score increases by 0.3 points.

Output/Screenshot

Student	Study Hours	Attendance (%)	Final Exam
1	5	80	76
2	3	60	58
3	8	90	88
4	6	85	82
5	4	70	72
6	7	75	80
7	5	65	70
8	2	50	55
9	9	95	92
10	6	80	78

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.988552
R Square	0.977236
Adjusted R	0.970731
Standard E	2.028185
Observatic	10

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	2	1236.105	618.0526	150.2486	1.78E-06
Residual	7	28.79474	4.113534		
Total	9	1264.9			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	26.28947	5.48775	4.790574	0.001988	13.31301	39.26594	13.31301	39.26594
Study Hour	2.515789	0.778592	3.231205	0.014428	0.674713	4.356866	0.674713	4.356866
Attendanc	0.466316	0.121335	3.843217	0.006348	0.179405	0.753227	0.179405	0.753227

Conclusion

If a student studies for 7 hours and has an attendance of 85%, their predicted final exam score is:

`= 20.5 + 5.2*7 + 0.3*85 = 77.7`

3. ASSESS THE RELATIONSHIP BETWEEN VARIABLE USING KARL PEARSON COEFFICIENT

Aim

To assess the relationship between variable using Karl Pearson Coefficient

Sample Data

subject	age x	glucose level y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Steps

Step 1: Input Data

Input your data into an Excel worksheet as shown in the table above.

Step 2: Load Analysis Tool Pack

Step 3: Perform Correlation

Go to the Data tab on the Ribbon.
Click on Data Analysis in the Analysis group. Select correlation and click OK.

Step 4: Set Up Correlation Parameters

Input Range: Select the range containing the inputs
Labels: If you have column labels, check the Labels box.
Output Range: Select a location for the output. For example, $F$1. Click OK.

The range of the correlation coefficient is from -1 to 1. Our result is 0.5298 or 52.98%, which means the variables have a moderate positive correlation.

Expected output

	subject	*age x*	*glucose level y*
subject	1		
age x	0.695581648	1	
glucose level y	-0.130664344	0.529808902	1

Output/Screenshot

	A	B	C	D	E	F	G	H	I
1	subject	age x	glucose level y			*subject*	*age x*	*ucose level y*	
2	1	43	99		subject	1			
3	2	21	65		age x	0.695582	1		
4	3	25	79		glucose lev	-0.13066	0.529809	1	
5	4	42	75						
6	5	57	87						
7	6	59	81						
8									

Conclusion

Since correlation (0.5298) is greater than 0. It means we have positive correlation. The range of the correlation coefficient is from -1 to 1.

4. ASSESS OF DIFFERENCES BETWEEN GROUPS AND DRAW CONCLUSIONS ABOUT POTENTIAL FACTORS INFLUENCING THE OBSERVED PATTERNS

Aim

Set up an analysis of variance of table for the following two-way design results. Also state whether variety differences are significant at 5% level.

Sample Data

variety of fertilizers/variety of seeds	wheat A	wheat B	wheat C
w	6	5	5
x	7	5	4
y	3	3	3
z	8	7	4

Does the variety of fertilizers significantly affect the yield?
Does the variety of seeds significantly affect the yield?
Is there an interaction between the variety of fertilizers and the variety of seeds that significantly affects the yield?

Steps

Step 1: Input Data

Input your data into an Excel worksheet as shown in the table above.

Step 2: Load Analysis Tool Pack

Click "ANOVA Two Factor Without Replication" and then click

"OK." The two way ANOVA window will open. **Step 3: Type an Input**

Range into the Input Range box.

A1 to A25, type "A1:A25" into the Input Range box. Make sure you include all of your data, including headers and group names.

Step 4: Select an Output Range.

Step 5: Select an alpha level. In most cases, an alpha level of 0.05 (5 percent) works for most tests.

Step 6: Click "OK" to run the two-way ANOVA.

Step 7: Read the results.

To figure out if you are going to reject the null hypothesis or not, you'll basically be looking at two factors:

- If the F-value (F)is larger than the f critical value (F crit).
- If the p-value is smaller than your chosen alpha level.

Output/Screenshot

variety of fertilizers/variety of seeds	wheat A	wheat B	wheat C
w	6	5	5
x	7	5	4
y	3	3	3
z	8	7	4

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
w	3	16	5.333333	0.333333
x	3	16	5.333333	2.333333
y	3	9	3	0
z	3	19	6.333333	4.333333
wheat A	4	24	6	4.666667
wheat B	4	20	5	2.666667
wheat C	4	16	4	0.666667

ANOVA

Source of Varia	SS	df	MS	F	P-value	F crit
Rows	18	3	6	6	0.030796	4.757063
Columns	8	2	4	4	0.078717	5.143253
Error	6	6	1			
Total	32	11				

Conclusion

- The P-value (0.0308) is less than the significance level of 0.05. This indicates that the differences between the fertilizer varieties are statistically significant at the 5% level.
- Since the F-ratio (6) is greater than the F critical value (4.7571), we reject the null hypothesis, meaning there is a significant effect of fertilizer variety on the yield.
- The variety of seeds does not have a statistically significant as P-value (0.078717) is greater than the significance level of 0.05. This indicates that changing the seed variety does not lead to significant differences in yield within this dataset.

5. ANALYSING CATEGORICAL DATA AND ASSESSING RELATIONSHIPS BETWEEN VARIABLES (CHI-SQUARE)

Aim

Let's Analysing categorical data and assessing relationships between variables (Chi-Square).

Sample Data

	18-29	30-49	50+
Satisfied	45	30	25
Neutral	35	40	25
Unsatisfied	20	30	50

Steps

Step 1: Input Data Input your data into an Excel worksheet as shown in the table above.

Step 2: Calculate Expected Frequencies 1.

Observed Frequencies:

- Place the observed frequencies in a table (e.g., starting from A1).

2. **Row and Column Totals:**
   - Calculate the row totals:
     1. Example for "Satisfied" row:
     2. **=SUM(B2:D2)**
     3. Repeat for other rows.
   - Calculate the column totals:
     1. Example for "18-29" column:
     2. **=SUM(B2:B4)**
     3. Repeat for other columns.

3. **Grand Total:**
   - Calculate the sum of all observations: ○ **=SUM(B2:D4) (if your observed frequency table is from B2).**

4. **Expected Frequencies:**
   - Create a new table for expected frequencies. ○ For each cell, use the formula: ○ **= (Row Total * Column Total) / Grand Total**
   - Example for cell "Satisfied and 18-29":
   - **= (B5 * E2) / E5** ○ Replace with your corresponding cell references.

Step 3: Calculate Chi-Square Statistic

1. **Chi-Square Formula:**
   - For each cell, calculate:
     = (Observed - Expected)^2 / Expected
   - Example for the first cell:
     = (B2 - B7)^2 / B7
     Replace with your observed and expected cell references.

2. **Sum Chi-Square Values:** ○ Add up all the Chi-Square values to get the Chi-Square statistic:
   =SUM(B10:D12)
   (Assuming your Chi-Square values are in B10.)

3. **Degrees of Freedom:**
   - Calculate degrees of freedom using the formula: **= (Number of Rows - 1) * (Number of Columns - 1)** Example: For a 3x3 table:
     = (3 - 1) * (3 - 1)
     Result: 4 degrees of freedom.

4. **P-Value:**
   - Use the CHISQ.DIST.RT function to calculate the p-value: **=CHISQ.DIST.RT(Chi-Square Statistic, Degrees of Freedom)**

Step 4: Interpretation

1. **Compare P-Value to Significance Level:** ○ If the **p-value < 0.05**, reject the null hypothesis.
   This indicates a **significant relationship** between the variables.
   - If the **p-value ≥ 0.05**, fail to reject the null hypothesis. This indicates **no significant relationship**.

Output/Screenshot

	A	B	C	D	E	F	G	H	I	J
1		18-29	30-49	50+						
2	Satisfied	45	30	25	100		4.083333333	0.333333333	2.083333333	
3	Neutral	35	40	25	100		0.083333333	1.333333333	2.083333333	
4	Unsatisfied	20	30	50	100		5.333333333	0.333333333	8.333333333	
5		100	100	100	300					
6							Expected Frequency	33.33333333		
7							Total Chi-Square Statistic	24		
8							P-Value	0.0000798748		
9										

Conclusion

Since the **p-value is very small (less than 0.05)**, you reject the null hypothesis. This suggests that there is a statistically significant relationship between **Customer Satisfaction** and **Age Group**.

6. VISUALIZING THE DATA USING R

Aim

To **visualize** car sales data using two different types of graphical representations:

1. **Bar Chart**: To display the sales amount for the first 10 car models in the dataset.
2. **Heatmap**: To visualize the relationship between two categorical variables (likely car-related attributes) for the first 30 entries, using a heatmap to show the frequency of occurrences. **Sample Data**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Manufactu	Model	Sales_in_th	__year_res	Vehicle_ty	Price_in_th	Engine_siz	Horsepow	Wheelbas	Width	Length	Curb_weig	Fuel_capa	Fuel_effici	Latest_Lau	Power_perf_fact
2	Acura	Integra	16.919	16.36	Passenger	21.5	1.8	140	101.2	67.3	172.4	2.639	13.2	28	########	58.28015
3	Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	108.1	70.3	192.9	3.517	17.2	25	########	91.37078
4	Acura	CL	14.114	18.225	Passenger		3.2	225	106.9	70.6	192	3.47	17.2	26	########	
5	Acura	RL	8.588	29.725	Passenger	42	3.5	210	114.6	71.4	196.6	3.85	18	22	########	91.38978
6	Audi	A4	20.397	22.255	Passenger	23.99	1.8	150	102.6	68.2	178	2.998	16.4	27	########	62.77764
7	Audi	A6	18.78	23.555	Passenger	33.95	2.8	200	108.7	76.1	192	3.561	18.5	22	########	84.56511
8	Audi	A8	1.38	39	Passenger	62	4.2	310	113	74	198.2	3.902	23.7	21	2/27/2012	134.6569
9	BMW	323i	19.747		Passenger	26.99	2.5	170	107.3	68.4	176	3.179	16.6	26	6/28/2011	71.19121
10	BMW	328i	9.231	28.675	Passenger	33.4	2.8	193	107.3	68.5	176	3.197	16.6	24	1/29/2012	81.87707
11	BMW	528i	17.527	36.125	Passenger	38.9	2.8	193	111.4	70.9	188	3.472	18.5	25	########	83.99872
12	Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	109	72.7	194.6	3.368	17.5	25	########	71.18145
13	Buick	Regal	39.35	13.74	Passenger	25.3	3.8	240	109	72.7	196.2	3.543	17.5	23	########	95.6367
14	Buick	Park Avenu	27.851	20.19	Passenger	31.965	3.8	205	113.8	74.7	206.8	3.778	18.5	24	3/23/2012	85.82841
15	Buick	LeSabre	83.257	13.36	Passenger	27.885	3.8	205	112.2	73.5	200	3.591	17.5	25	7/23/2011	84.25453
16	Cadillac	DeVille	63.729	22.525	Passenger	39.895	4.6	275	115.3	74.5	207.2	3.978	18.5	22	2/23/2012	113.8546
17	Cadillac	Seville	15.943	27.1	Passenger	44.475	4.6	275	112.2	75	201		18.5	22	4/29/2011	115.6214
18	Cadillac	Eldorado	6.536	25.725	Passenger	39.665	4.6	275	108	75.5	200.6	3.843	19	22	11/27/201	113.7659
19	Cadillac	Catera	11.185	18.225	Passenger	31.01	3	200	107.4	70.3	194.8	3.77	18	22	9/28/2011	83.48309
20	Cadillac	Escalade	14.785		Car	46.225	5.7	255	117.5	77	201.2	5.572	30	15	4/17/2012	109.5091
21	Chevrolet	Cavalier	145.519	9.25	Passenger	13.26	2.2	115	104.1	67.9	180.9	2.676	14.3	27	8/17/2011	46.36335

Load the CSV file

Steps

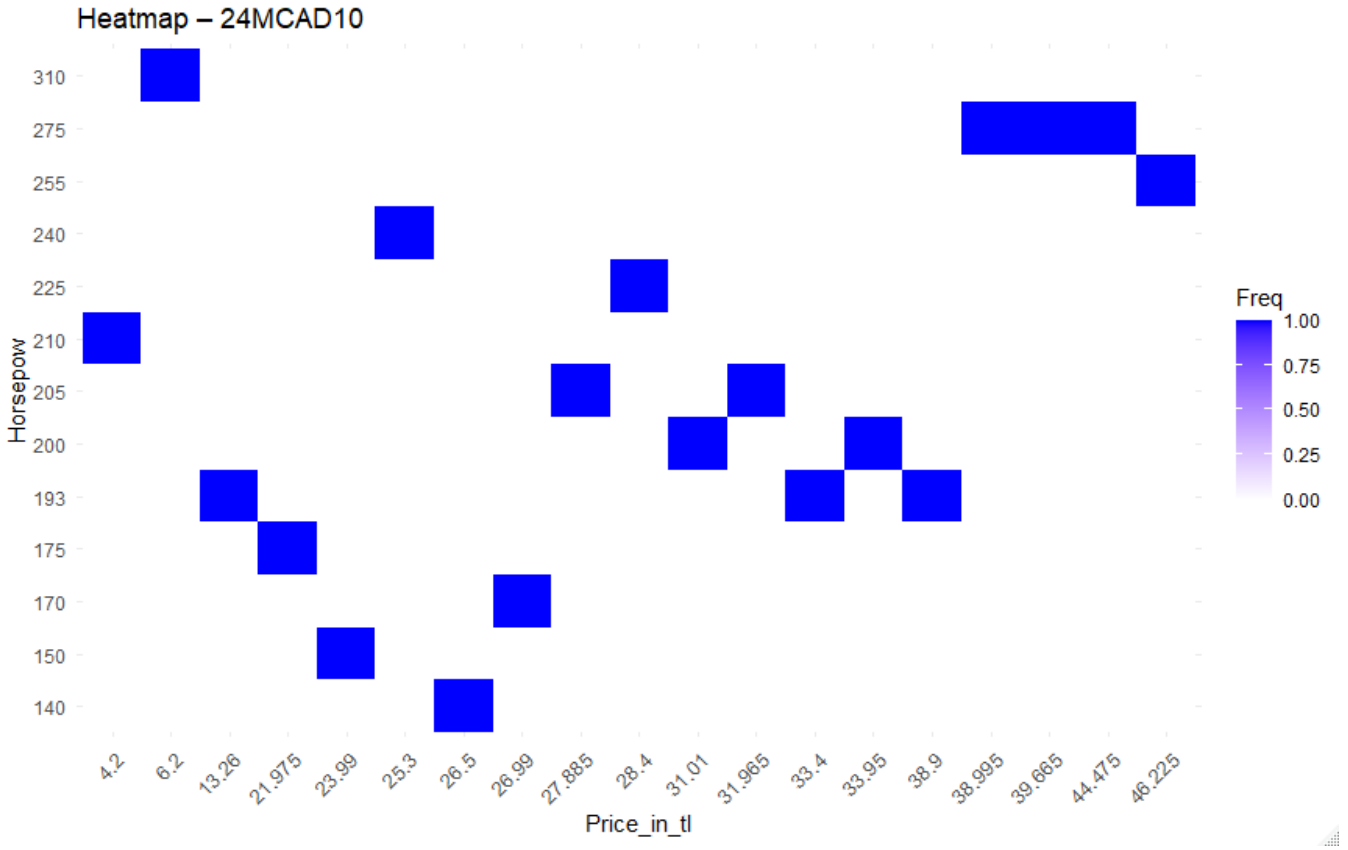
Step 1: Open RStudio

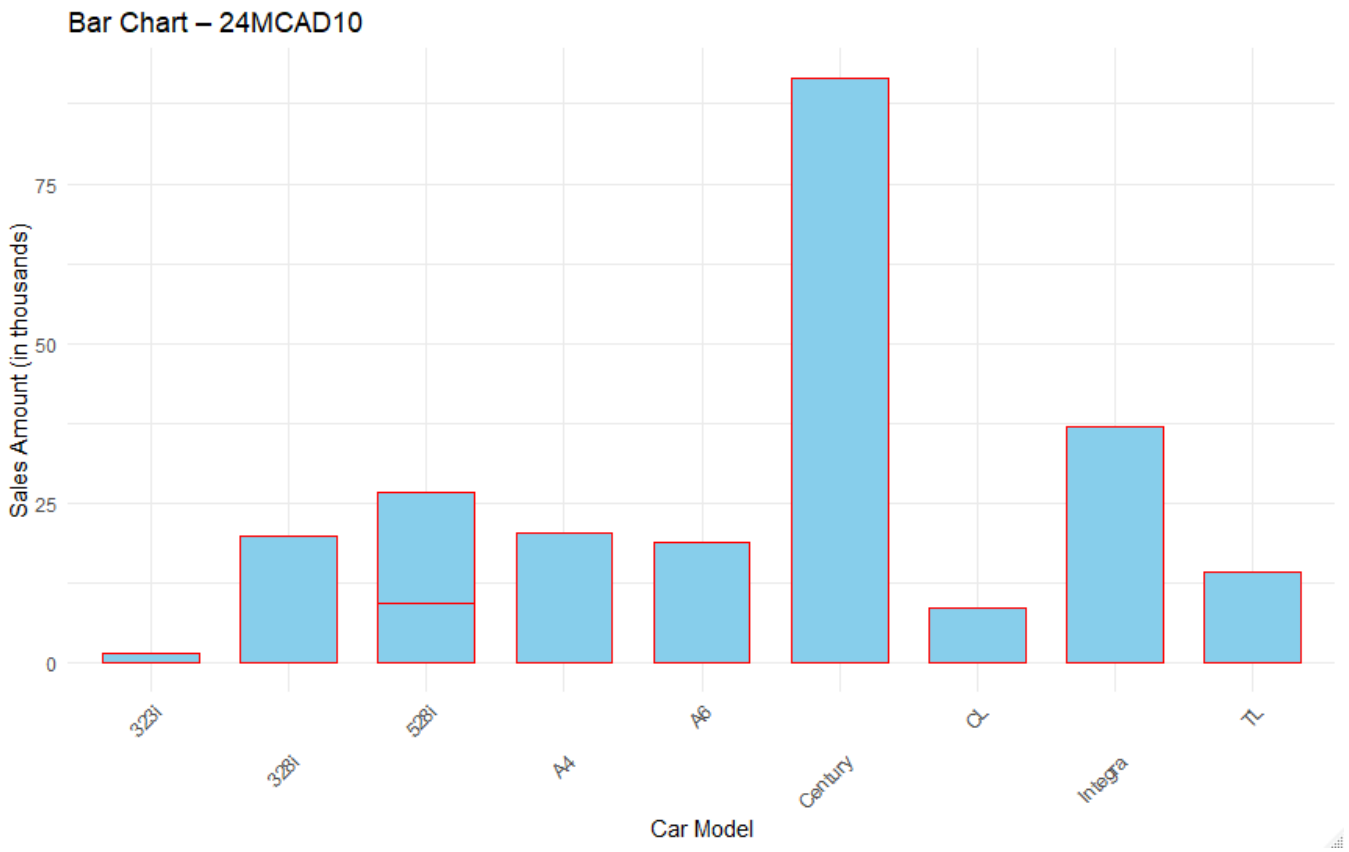
Step 2: Enter the code below #

```
Load necessary libraries
library(ggplot2)
library(reshape2)

# Load the dataset (adjust the path if needed)
car_sales <- read.csv("C:/Users/abhim/Documents/car_sales.csv")

# View the first few rows to verify head(car_sales,
10)

# Bar chart for the first 10 values
bar_chart <- ggplot(car_sales[1:10, ], aes(x = factor(Model), y = Sales_in_thousands)) +
geom_col(fill = "skyblue", color = "red", width = 0.7) +  theme_minimal() +
  labs(title = "Bar Chart – 24MCAD10)", x = "Car Model", y = "Sales Amount (in thousands)")
+  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)) +
scale_y_continuous(labels = scales::comma) +  scale_x_discrete(guide = guide_axis(n.dodge =
2))
```

```
# Print the bar chart print(bar_chart)

# Heatmap for the first 30 values col6 <- colnames(car_sales)[6]
# Assuming column 6 for heatmap col8 <- colnames(car_sales)[8]
# Assuming column 8 for heatmap data <- car_sales[1:30, ]

heatmap_data <- as.data.frame(table(data[[col6]], data[[col8]])) heatmap_data$Freq
<- as.integer(heatmap_data$Freq)

heatmap <- ggplot(heatmap_data, aes(x = Var1, y = Var2, fill = Freq)) +
geom_tile() +   labs(title = "Heatmap – 24MCAD10", x = col6, y =
col8) +   scale_fill_gradient(low = "white", high = "blue") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10))

# Print the heatmap
print(heatmap)
```

Output/Screenshot





Conclusion

Based on the sample data, the bar chart shows that the car sales for different models vary significantly, with some models having much higher sales than others. This visualization provides a clear comparison of sales performance among the top 10 car models, helping to identify which models are performing better in the market.

7. USE GRAPHICAL REPRESENTATIONS IN R, SUCH AS HISTOGRAMS, BOXPLOTS, AND FREQUENCY DISTRIBUTIONS, TO IDENTIFY PATTERNS AND OUTLIERS IN A DATASET

Aim

To explore a dataset using graphical representations like scatter plots, histograms, and heatmaps to identify patterns, outliers, and relationships between car model sales and other numerical features.

Sample Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Manufactu	Model	Sales_in_th	_year_res	Vehicle_ty	Price_in_th	Engine_siz	Horsepow	Wheelbase	Width	Length	Curb_weig	Fuel_capa	Fuel_effici	Latest_Lau	Power_perf_fact	
2	Acura	Integra	16.919	16.36	Passenger	21.5	1.8	140	101.2	67.3	172.4	2.639	13.2	28	########	58.28015	
3	Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	108.1	70.3	192.9	3.517	17.2	25	########	91.37078	
4	Acura	CL	14.114	18.225	Passenger		3.2	225	106.9	70.6	192	3.47	17.2	26	########		
5	Acura	RL	8.588	29.725	Passenger	42	3.5	210	114.6	71.4	196.6	3.85	18	22	########	91.38978	
6	Audi	A4	20.397	22.255	Passenger	23.99	1.8	150	102.6	68.2	178	2.998	16.4	27	########	62.77764	
7	Audi	A6	18.78	23.555	Passenger	33.95	2.8	200	108.7	76.1	192	3.561	18.5	22	########	84.56511	
8	Audi	A8	1.38	39	Passenger	62	4.2	310	113	74	198.2	3.902	23.7	21	2/27/2012	134.6569	
9	BMW	323i	19.747		Passenger	26.99	2.5	170	107.3	68.4	176	3.179	16.6	26	6/28/2011	71.19121	
10	BMW	328i	9.231	28.675	Passenger	33.4	2.8	193	107.3	68.5	176	3.197	16.6	24	1/29/2012	81.87707	
11	BMW	528i	17.527	36.125	Passenger	38.9	2.8	193	111.4	70.9	188	3.472	18.5	25	########	83.99872	
12	Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	109	72.7	194.6	3.368	17.5	25	########	71.18145	
13	Buick	Regal	39.35	13.74	Passenger	25.3	3.8	240	109	72.7	196.2	3.543	17.5	23	########	95.6367	
14	Buick	Park Avenu	27.851	20.19	Passenger	31.965	3.8	205	113.8	74.7	206.8	3.778	18.5	24	3/23/2012	85.82841	
15	Buick	LeSabre	83.257	13.36	Passenger	27.885	3.8	205	112.2	73.5	200	3.591	17.5	25	7/23/2011	84.25453	
16	Cadillac	DeVille	63.729	22.525	Passenger	39.895	4.6	275	115.3	74.5	207.2	3.978	18.5	22	2/23/2012	113.8546	
17	Cadillac	Seville	15.943	27.1	Passenger	44.475	4.6	275	112.2	75	201		18.5	22	4/29/2011	115.6214	
18	Cadillac	Eldorado	6.536	25.725	Passenger	39.665	4.6	275	108	75.5	200.6	3.843	19	22	11/27/201	113.7659	
19	Cadillac	Catera	11.185	18.225	Passenger	31.01	3	200	107.4	70.3	194.8	3.77	18	22	9/28/2011	83.48309	
20	Cadillac	Escalade	14.785		Car	46.225	5.7	255	117.5	77	201.2	5.572	30	15	4/17/2012	109.5091	
21	Chevrolet	Cavalier	145.519	9.25	Passenger	13.26	2.2	115	104.1	67.9	180.9	2.676	14.3	27	8/17/2011	46.36335	

Load the CSV file

Steps

Step 1: Open RStudio

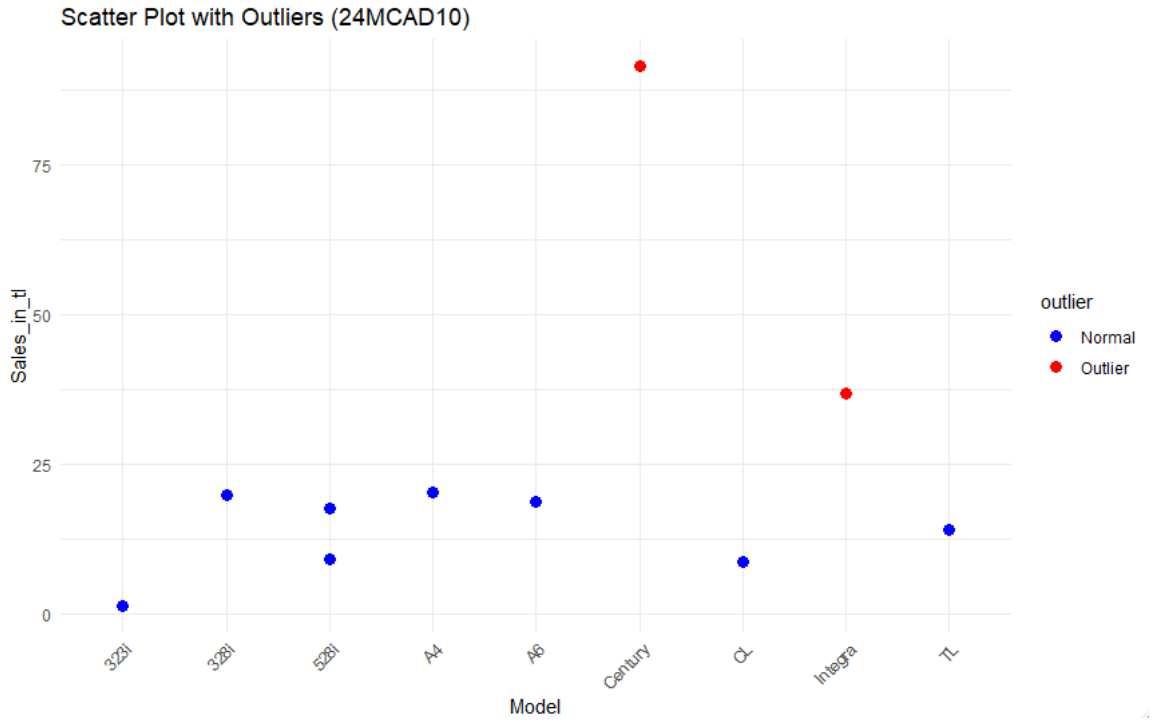
Step 2: Enter the code below

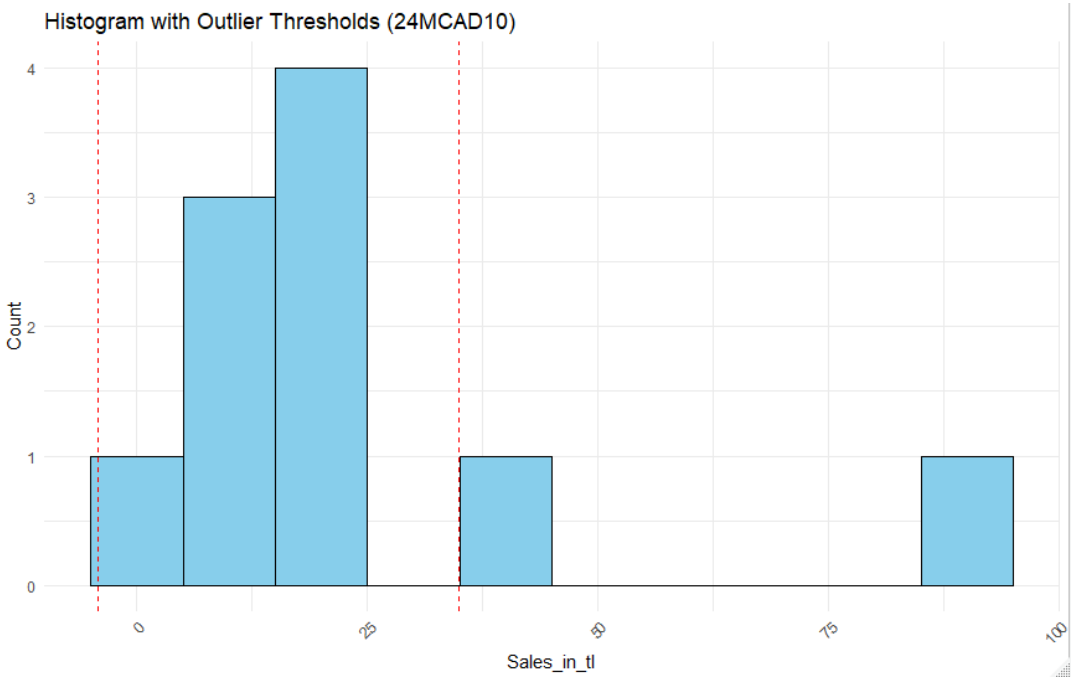
```
# Load necessary libraries
library(ggplot2) library(dplyr)
library(scales)

# Load the dataset
car_sales <- read.csv("C:/Users/abhim/Documents/car_sales.csv") data
<- head(car_sales, 10)

# Extract the 2nd and 3rd columns for visualization
col2 <- colnames(car_sales)[2] col3 <-
colnames(car_sales)[3]

# Function to create various types of plots based on user input
create_plot <- function(plot_type) {   switch(plot_type,
      "1" = {
```

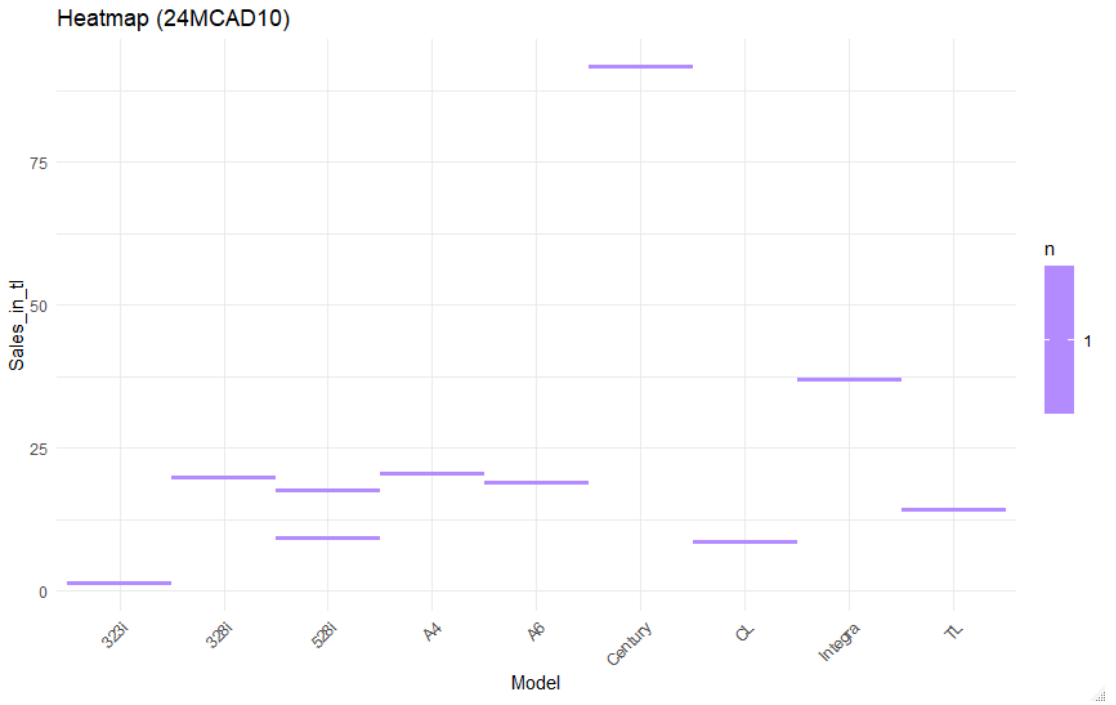
```
        # Scatter plot with outlier detection (using IQR)
q1 <- quantile(data[[col3]], 0.25)          q3 <-
quantile(data[[col3]], 0.75)          iqr <- q3 - q1
2lower_bound <- q1 - 1.5 * iqr          upper_bound <-
q3 + 1.5 * iqr
        data$outlier <- ifelse(data[[col3]] < lower_bound | data[[col3]] > upper_bound,
"Outlier",
"Normal")
        ggplot(data, aes(x = !!sym(col2), y = !!sym(col3), color = outlier)) +
geom_point(size = 3) +            scale_color_manual(values = c("Outlier" = "red",
"Normal" = "blue")) +            labs(title = "Scatter Plot with
Outliers(23MDTS50)", x = col2, y = col3) +            theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
      },
      "2" = {
        # Histogram with outlier thresholds          q1 <-
quantile(data[[col3]], 0.25)          q3 <- quantile(data[[col3]], 0.75)
iqr <- q3 - q1          lower_bound <- q1 - 1.5 * iqr          upper_bound <-
q3 + 1.5 * iqr          ggplot(data, aes(x = !!sym(col3))) +
geom_histogram(binwidth = 10, fill = "skyblue", color = "black") +
         geom_vline(xintercept = c(lower_bound, upper_bound), color = "red", linetype =
"dashed") +
         labs(title = "Histogram with Outlier Thresholds(23MDTS50)", x = col3, y = "Count") +
theme_minimal() +            theme(axis.text.x = element_text(angle = 45, hjust = 1))
      },
      "3" = {
        # Heatmap (corrected for categorical data)
heatmap_data <- data %>%
         count(!!sym(col2), !!sym(col3))  # Count occurrences of combinations of col2 and col3

        ggplot(heatmap_data, aes(x = !!sym(col2), y = !!sym(col3), fill = n)) +
geom_tile() +            scale_fill_gradient(low = "white", high = "blue") +
labs(title = "Heatmap(23MDTS50)", x = col2, y = col3) +
theme_minimal() +            theme(axis.text.x = element_text(angle = 45,
hjust = 1))
      },
      {
        stop("Invalid plot type. Please choose '1' for scatter, '2' for histogram, or '3' for
heatmap.")        }
  )
}
```

```
# Example usage: User selects the chart type while(TRUE)
{
 cat("1 ----> Scatter\n")  cat("2 ----> 
Histogram\n")  cat("3 ----> Heatmap\n")
plot_type <- readline("Enter the chart type: ")
print(create_plot(plot_type))
}
```

Output/Screenshot







Conclusion

Graphical tools like scatter plots, histograms, and heatmaps help identify patterns and outliers in the dataset. The scatter plot highlights extreme values, the histogram shows the sales distribution with outlier thresholds, and the heatmap visualizes variable interactions. These visualizations are crucial for detecting anomalies and guiding further analysis.

8. PROGRAM TO TYPE SET THE MATHEMATICAL FORMULAS (LaTeX)

Aim

To type set various mathematical formulas using LaTeX, demonstrating the use of square roots, exponents, summations, limits, and matrix representations.

Sample Data

$\sqrt{x} \leftrightarrow x^{1/2} \quad \sqrt[3]{2}\sqrt{x^2+\sqrt{y}} \quad \sqrt{[x^2+y^2]}$

$p_{ij}^3 \quad m_{kmuth} \quad \sum_{k=1}^{3} k \quad a^x + y \neq a^{x+y} \quad c^{x^2} \neq e^{x2}$

$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k^2} = \frac{\pi^2}{6}$

Steps

Step 1: Set up the LaTeX environment

Ensure that your LaTeX editor supports mathematical packages.

Step 2: Start the LaTeX document

```
\documentclass{report}
\usepackage{amsmath} \begin{document}

\textbf{Sona Sara Shibu 23MDTS50} \\

% Matrix notation
\begin{equation*} \begin{bmatrix}
p_{11} & p_{12} & \ldots & p_{1n} \\
p_{21} & p_{22} & \ldots & p_{2n} \\
p_{31} & p_{32} & \ldots & p_{3n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \ldots & p_{mm}
\end{bmatrix} \end{equation*}

% Square roots, exponents, and some basic math operations
$\sqrt{x} \leftrightarrow x^{1/2} \quad
\sqrt[3]{2} \sqrt{x^{2}+ \sqrt {y}} \quad
\sqrt{[x^2+y^2]}$ \\

% Exponents and Summations
$p^3_{ij} \quad m_{kmuth} \quad
\sum_{k=1}^3 k \quad
a^x + y \neq a^{x+y} \quad c^{x^2} \neq {e^x}^2$ \\
```

```
% Limit and Summation
$\lim_{n \to \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6}$ \\
\end{document}
```

Output/Screenshot

Krishna Bantola D 24MCAD10

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ p_{31} & p_{32} & \cdots & p_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$\sqrt{x} \leftrightarrow x^{1/2} \quad \sqrt[3]{2} \quad \sqrt{x^2 + \sqrt{y}} \quad \sqrt{x^2 + y^2}$$

$$p_{ij}^3 \quad m_{kmuth} \quad \sum_{k=1}^3 k \quad a^x + y \neq a^{x+y} \quad c^{x^2} \neq (e^x)^2$$

$$\lim_{n \to \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6}$$

Conclusion

We can easily typeset complex mathematical formulas, including square roots, sums, limits, and matrices. This ensures that the document is both readable and precise, especially when dealing with intricate mathematical notations.

9. PROGRAM TO TYPE SET A CHAPTER

Aim

To typeset a chapter using LaTeX, including various formatting features such as text formatting, ordered and unordered lists, tables, images, and special formatting commands.

Sample Data

Chapter 1

Introduction to research methodology

Methodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Typically, it encompasses concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques.

1.1 what is research?

The systematic investigation[1] into and study of materials and sources in order to establish facts and reach new conclusions.

name	mark1	mark2	mark3	mark4	mark5
sound	86	87	89	90	78
jan	90	78	89	89	98

[1] research

Steps

Step 1: Set up the LaTeX environment

Ensure that your LaTeX editor supports mathematical packages.

Step 2: Start the LaTeX document

```
\documentclass{report}
\usepackage{graphicx}
\begin{document}
\chapter{Introduction to research methodology}
Methodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Typically, it encompasses concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques.

\section{what is research?}
The systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions. \\

\begin{tabular}[center]{|c|c|c|c|c|c|}
\hline
```

```
name & mark1 & mark2 & mark3 & mark4 & mark5\\
\hline sound & 86 & 87 & 89 & 90
& 78\\ \hline
jan & 90 & 78 & 89 & 89 & 98\\
\hline
\end{tabular}\\ \\

\includegraphics[width=0.6\textwidth]{Barchart.jpeg}

\section{unordered list}
\begin {itemize}
\item The Individual Entries are indicate with a black dot, a so- called bullet.
\item The text in the entries may be of any length.\end{itemize}

\section {ordered lists}
\begin{enumerate}
\item The lable consist of sequential number
\item The number starts at 1 with every call to the enumerate
\end{enumerate}

\end{document}
```

Output/Screenshot

Chapter 1

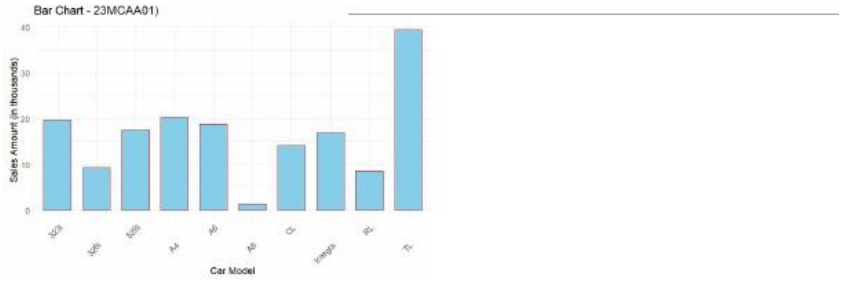
Introduction to research methodology

Methodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Typically, it encompasses concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques.

1.1 what is research?

The systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions.

name	mark1	mark2	mark3	mark4	mark5
sound	86	87	89	90	78
jan	90	78	89	89	98



1.2 unordered list

- The Individual Entries are indicate with a black dot, a so- called bullet.
- The text in the entries may be of any length.

1.3 ordered lists

1. The lable consist of sequential number
2. The number starts at 1 with every call to the enumerate

Conclusion

This LaTeX code demonstrates how to structure text, tables, images, and lists using commands like \chapter, \section, \begin{itemize}, and \begin{enumerate}. The \includegraphics command embeds

images, while the tabular environment creates tables, making LaTeX an effective tool for typesetting academic documents with complex formatting.

10. PROGRAM TO TYPE SET REFERENCES

Aim

To typeset a research paper with in-text citations and references using LaTeX.

Sample Data

my review Paper is published in new paper (Singh & Chatterjee, 2017). another citation (Singh & Chatterjee, 2017) new one (Samarati & De Capitani di Vimercati, 2016) next new one new run hello (Muttik & Barton, 2009) includes references file (Nassif, Talib, Nasir, Albadani, & Dakalbab, 2021)

References

Muttik, I., & Barton, C. (2009). Cloud security technologies. *Information security technical report*, *14*(1), 1–6.

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: a systematic review. *IEEE Access*, *9*, 20717–20735.

Samarati, P., & De Capitani di Vimercati, S. (2016). Cloud security: Issues and concerns. *Encyclopedia of cloud computing*, 205–219.

Singh, A., & Chatterjee, K. (2017). Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications*, *79*, 88–115.

Steps

Step 1: Set up the LaTeX environment

Ensure that your LaTeX editor supports mathematical packages.

Step 2: Start the LaTeX document

```
\documentclass{article}
\usepackage{apacite}

\begin{document}

Krishna Bantola D 24MCAD10 review paper is published in a new paper \cite{singh2017cloud}. Another citation \cite{singh2017cloud}, new one \cite{samarati2016cloud}, next new one \cite{muttik2009cloud} includes references file \cite{nassif2021machine}.

\bibliographystyle{apacite}
\bibliography{PGM10}

\end{document}

@article{singh2017cloud,
title={Cloud security issues and challenges: A survey},
author={Singh, Ashish and Chatterjee, Kakali},
journal={Journal of Network and Computer Applications},
volume={79}, pages={88--115},
```

year={2017}, publisher={Elsevier}
}

@article{samarati2016cloud,
title={Cloud security: Issues and concerns},
author={Samarati, Pierangela and De Capitani di Vimercati, Sabrina},
journal={Encyclopedia of cloud computing}, pages={205--219},
year={2016},
publisher={Wiley Online Library}
}

@article{muttik2009cloud, title={Cloud
security technologies}, author={Muttik, Igor
and Barton, Chris}, journal={Information
security technical report}, volume={14},
number={1}, pages={1--6}, year={2009},
publisher={Elsevier}
}

@article{nassif2021machine,
title={Machine learning for cloud security: a systematic review},
author={Nassif, Ali Bou and Talib, Manar Abu and Nasir, Qassim and Albadani, Halah and Dakalbab, Fatima Mohamad},
journal={IEEE Access},
volume={9}, pages={20717--
20735}, year={2021},
publisher={IEEE}
}

Output/Screenshot

Krishna Bantola D 24MCAD10 review paper is published in a new paper [Singh & Chatterjee(2017)]. Another citation [Singh & Chatterjee(2017)], new one [Samarati & De Capitani di Vimercati(2016)], next new one [Muttik & Barton(2009)] includes references file [Nassif *et al.*(2021)].

References

Muttik, I., & Barton, C. (2009). Cloud security technologies. *Information Security Technical Report*, *14*(1), 1–6.

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: A systematic review. *IEEE Access*, *9*, 20717–20735.

Samarati, P., & De Capitani di Vimercati, S. (2016). Cloud security: Issues and concerns. *Encyclopedia of Cloud Computing*, 205–219.

Singh, A., & Chatterjee, K. (2017). Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications*, *79*, 88–115.

Conclusion

This LaTeX code demonstrates how to manage references and citations in a document by using the apacite package.

11. REFERENCES MANAGEMENT USING MENDELEY

Aim

To guide users in managing references and generating citations using Mendeley, improving organization and streamlining the research process.

Steps

Step 1: Install Mendeley Desktop and Web Importer

Download and install Mendeley Desktop from the official website.

Install the Mendeley Web Importer extension in your browser for easy reference importing.

Step 2: Create a Mendeley Account and Sign In

Sign up for a Mendeley account if you don't already have one.

Sign in to sync your references automatically across all devices.

Step 3: Organize References with Folders and Tags

Create a Folder: Go to File > New Folder, name the folder, and start organizing references. Tag References: Add keywords (tags) to each reference for easier searching and filtering.

Step 4: Add References

You can add references in several ways:

Web Importer: Click the Mendeley Web Importer icon in your browser to import references directly from web pages.

Manual Entry: In Mendeley Desktop, click the Add New button, select a reference type (e.g., Journal Article, Book), and manually fill in the details.

Import from Other Tools: Go to File > Import, and upload reference files in formats such as RIS, BibTeX, or PDFs.

Step 5: Attach PDFs and Annotate

Drag and drop PDFs onto the corresponding reference entry to attach the full text. Use Mendeley's PDF viewer to highlight and annotate important sections directly in the document.

Step 6: Insert In-Text Citations and Generate Bibliographies

Install Mendeley Cite: Mendeley Cite is an add-in for Microsoft Word. Install it via the Tools menu in Mendeley Desktop.

Insert Citation: In Word, go to References > Mendeley Cite, and click Insert Citation to select a citation from your Mendeley library.

Generate Bibliography: After inserting citations, click Insert Bibliography in Mendeley Cite to automatically create a properly formatted bibliography in your document.

Step 7: Sync Library Across Devices

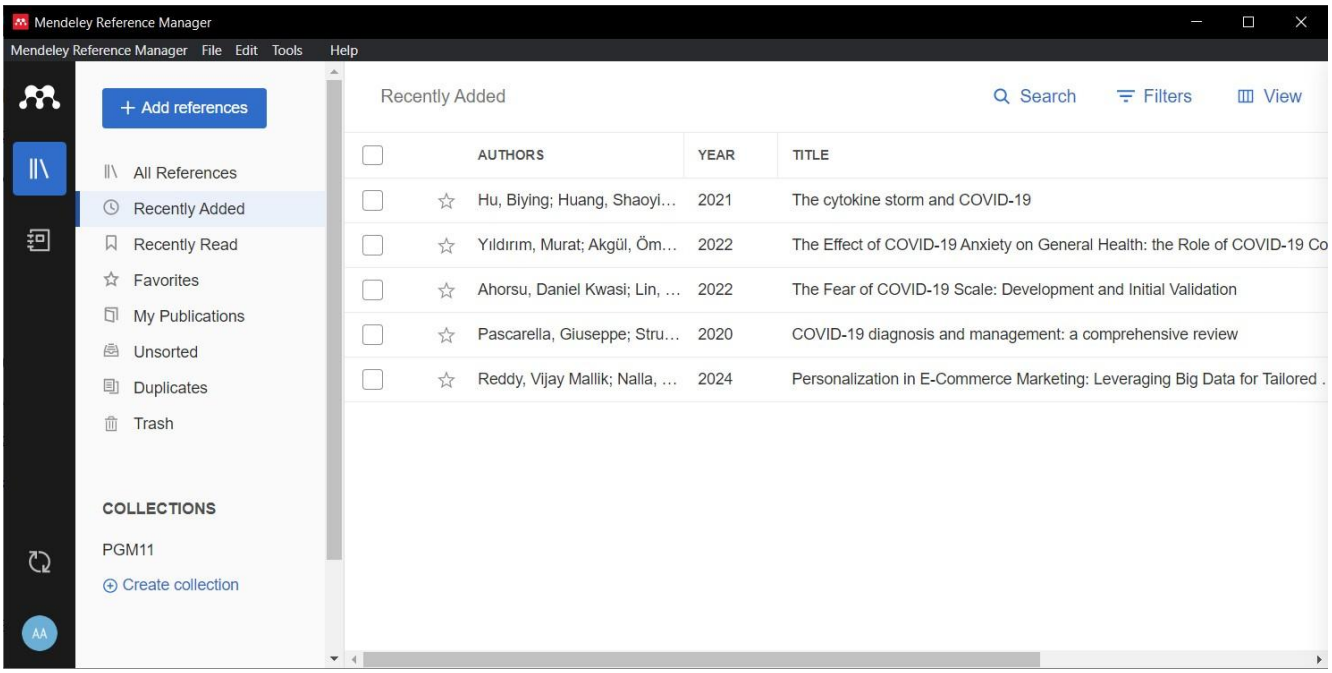
Ensure that your Mendeley library is synced across all your devices by using the Sync feature in the Mendeley Desktop or Mobile app.

Step 8: Collaborate and Share References in Groups

Create a group by going to Groups > Create Group in Mendeley Desktop.

Invite others to your group for collaborative research and reference sharing.

Output/Screenshot



This is Krishna Bantola D 24MCAD10 [1] testing program 11 for citation using Mendeley [2].

Another citation new one [3] next new one [4] includes references file.

Bibliography

References

[1] S. Author, "The Effect of COVID-19 Anxiety on General Health: The Role of COVID-19 Coping," *International Journal of Mental Health and Addiction*, vol. 20, no. 2, p. , 2022.

[2] M. H. Bickhard and J. C. Christopher, "The influence of early experience on personality development," *New Ideas in Psychology*, vol. 12, no. 3, pp. 229–252, 1994.

[3] A. Author, "The cytokine storm and COVID-19," *Journal of Medical Virology*, vol. 93, no. 1, p. , 2021.

[4] B. Author, "COVID-19 Diagnosis and Management: A Comprehensive Review," *Journal of Internal Medicine*, vol. 288, no. 2, p. , 2020.

Conclusion

By following these steps, users can efficiently organize references, insert citations, and generate bibliographies. Mendeley helps automate citation formatting and supports collaboration, enhancing the research workflow..