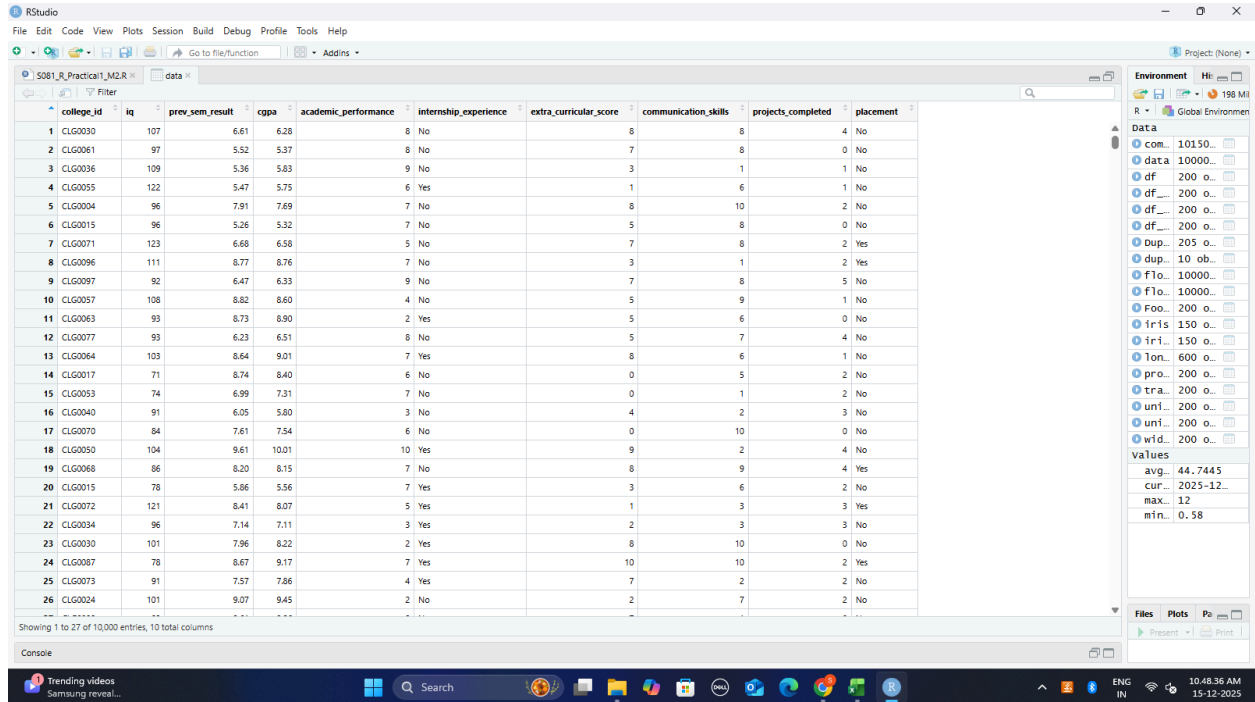


# Sheth L.U.J. & Sir M.V. College

## 1. Generating descriptive statistics using `summary()` or `describe()` (R)

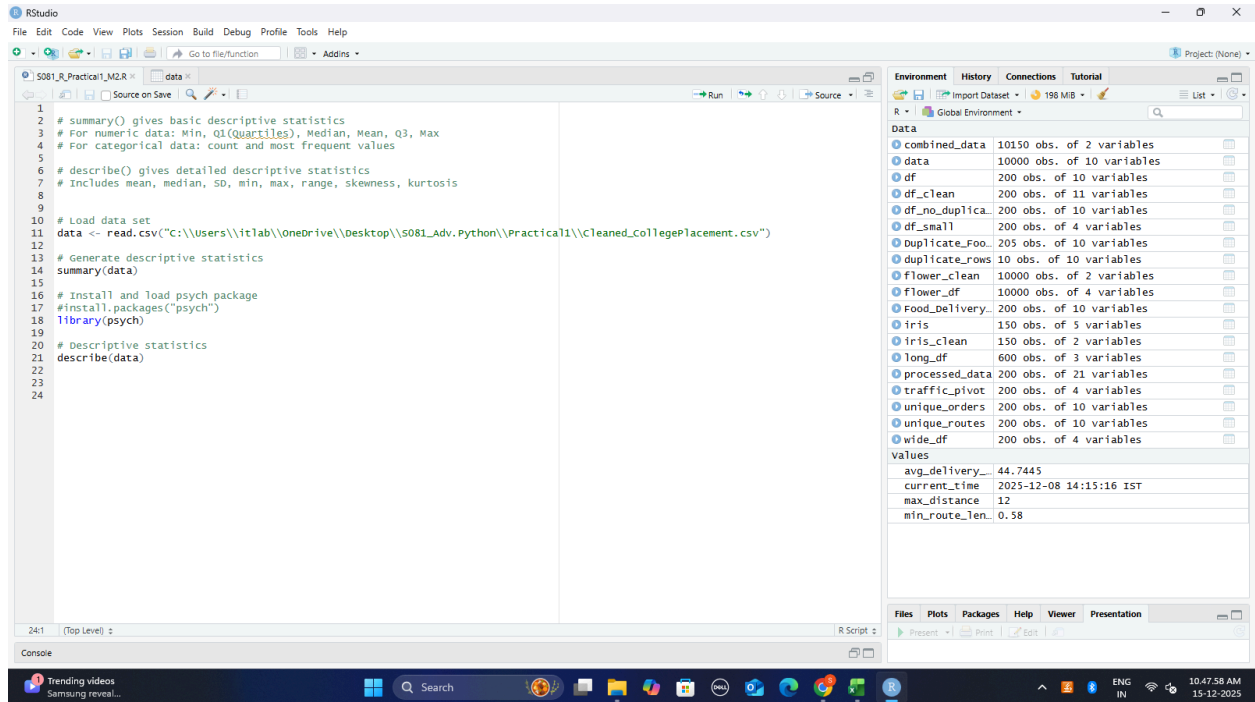


Summary of data:

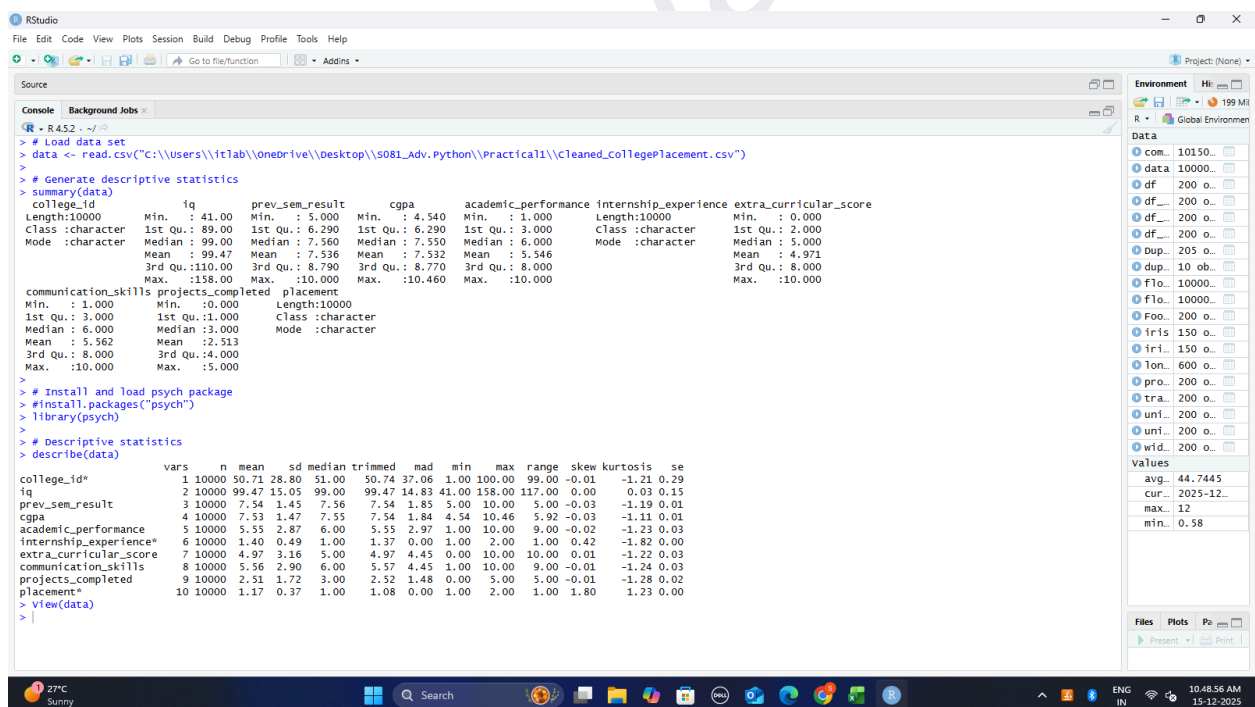
Variable	Summary
college_id	1 to 26
iq	71 to 107
prev_sem_result	5.26 to 7.91
cgpa	5.32 to 6.28
academic_performance	6 to 9
internship_experience	No, Yes
extra_curricular_score	1 to 10
communication_skills	3 to 10
projects_completed	0 to 10
placement	No, Yes

Name :- Priya Gupta  
Roll No :- S081

# Sheth L.U.J. & Sir M.V. College



```
1 # summary() gives basic descriptive statistics
2 # For numeric data: Min, Q1(Quartiles), Median, Mean, Q3, Max
3 # For categorical data: count and most frequent values
4
5 # describe() gives detailed descriptive statistics
6 # Includes mean, median, sd, min, max, range, skewness, kurtosis
7
8
9
10 # Load data set
11 data <- read.csv("C:\\Users\\itlab\\OneDrive\\Desktop\\S081_Adv.Python\\Practical1\\Cleaned.collegePlacement.csv")
12
13 # Generate descriptive statistics
14 summary(data)
15
16 # Install and load psych package
17 #install.packages("psych")
18 library(psych)
19
20 # Descriptive statistics
21 describe(data)
22
23
24
```



```
> # Load data set
> data <- read.csv("C:\\Users\\itlab\\OneDrive\\Desktop\\S081_Adv.Python\\Practical1\\Cleaned.collegePlacement.csv")
>
> # Generate descriptive statistics
> summary(data)
 college_id      iq      prev_sem_result      cgpa      academic_performance      internship_experience      extra_curricular_score
Length:10000    Min.   : 41.00    Min.   : 5.000    Min.   : 4.540    Min.   : 1.000    Length:10000    Min.   : 0.000
Class :character 1st Qu.: 89.00    1st Qu.: 6.290    1st Qu.: 6.290    1st Qu.: 3.000    Class :character 1st Qu.: 2.000
Mode :character  Median : 99.00    Median : 7.560    Median : 7.550    Median : 6.000    Mode :character  Median : 5.000
              Mean : 99.47    Mean : 7.536    Mean : 7.532    Mean : 5.546    Mean : 4.971
              3rd Qu.:110.00  3rd Qu.: 8.790    3rd Qu.: 8.770    3rd Qu.: 8.000    3rd Qu.: 8.000
              Max.   :158.00    Max.   :10.000    Max.   :10.460    Max.   :10.000    Max.   :10.000
communication_skills projects_completed placement
Min.   : 1.000    Min.   :0.000    Length:10000
1st Qu.: 3.000    1st Qu.:1.000    class:character
Median : 6.000    Median :3.000    Mode :character
Mean : 5.562    Mean :2.513
3rd Qu.: 8.000    3rd Qu.:4.000
Max.   :10.000    Max.   :5.000

> # Install and load psych package
> #install.packages("psych")
> library(psych)
>
> # Descriptive statistics
> describe(data)
      vars      n mean sd median trimmed mad min max range skew kurtosis se
college_id* 1 10000 50.71 28.80 51.00 50.74 37.06 1.00 100.00 99.00 -0.01 -1.21 0.29
iq           2 10000 99.47 15.05 99.00 99.47 14.83 41.00 158.00 117.00 0.00 0.03 0.15
prev_sem_result* 3 10000 7.54 1.45 7.56 7.54 1.85 5.00 10.00 5.00 -0.03 -1.19 0.01
cgpa          4 10000 7.53 1.47 7.55 7.54 1.84 4.54 10.46 5.92 -0.03 -1.11 0.01
academic_performance 5 10000 5.55 2.87 6.00 5.55 2.97 1.00 10.00 9.00 -0.02 -1.23 0.03
internship_experience* 6 10000 1.40 0.49 1.00 1.37 0.00 1.00 2.00 1.00 0.42 -1.82 0.00
extra_curricular_score 7 10000 4.97 3.16 5.00 4.97 4.45 0.00 10.00 10.00 0.01 -1.22 0.03
communication_skills 8 10000 5.56 2.90 6.00 5.57 4.45 1.00 10.00 9.00 -0.01 -1.24 0.03
projects_completed 9 10000 2.51 1.72 3.00 2.52 1.48 0.00 5.00 5.00 -0.01 -1.28 0.02
placement*     10 10000 1.17 0.37 1.00 1.08 0.00 1.00 2.00 1.00 1.80 1.23 0.00

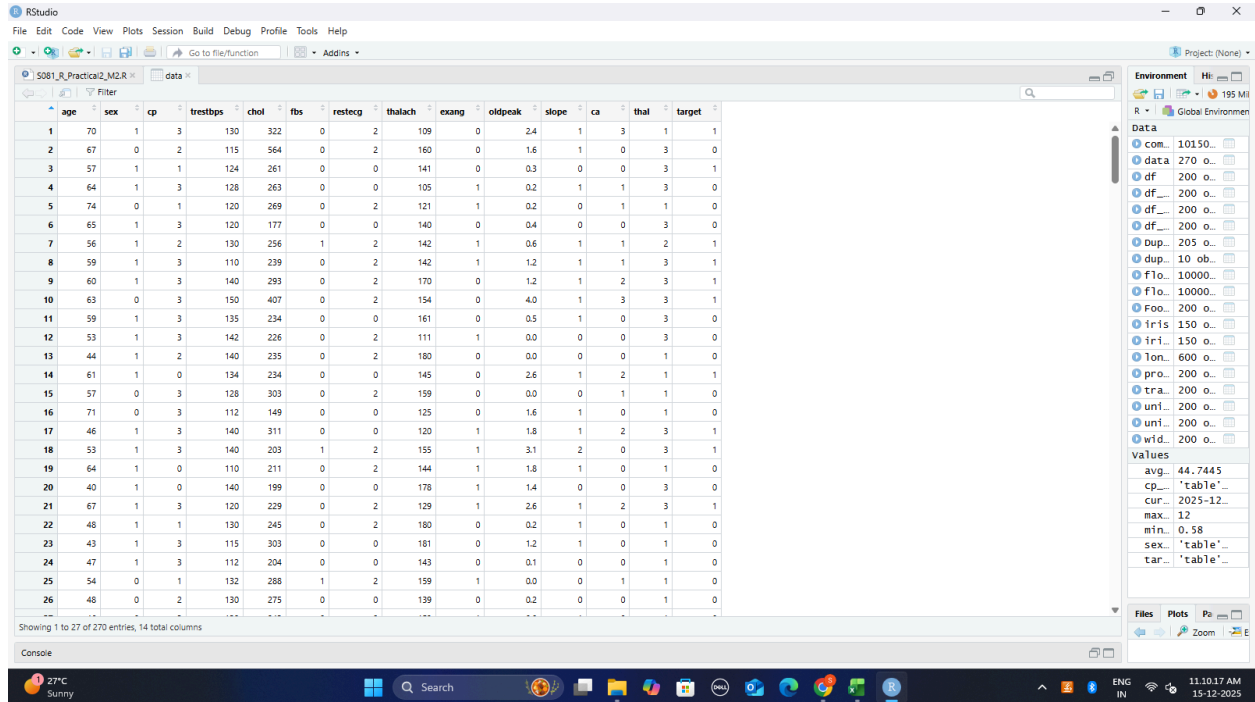
> view(data)
> |
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

## 2. Generating frequency tables using `table()` or `count()` (R)



Showing 1 to 27 of 270 entries, 14 total columns

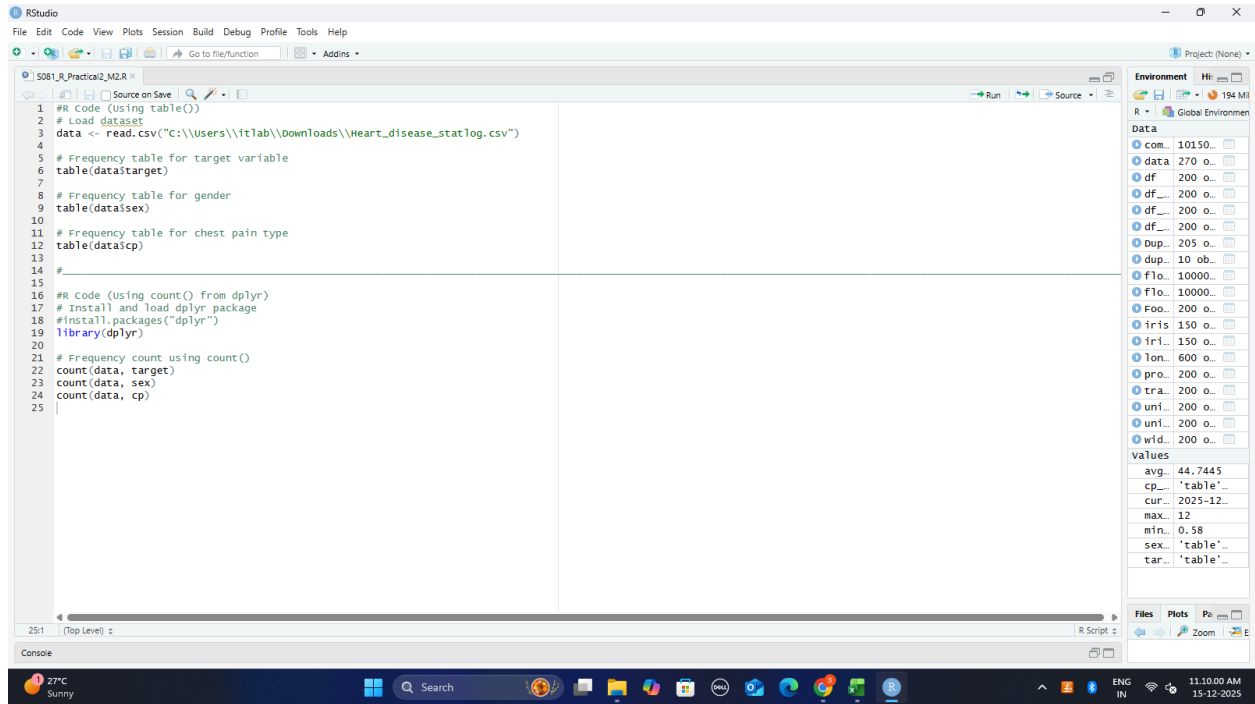
Console

```
table()
```

Name :- Priya Gupta

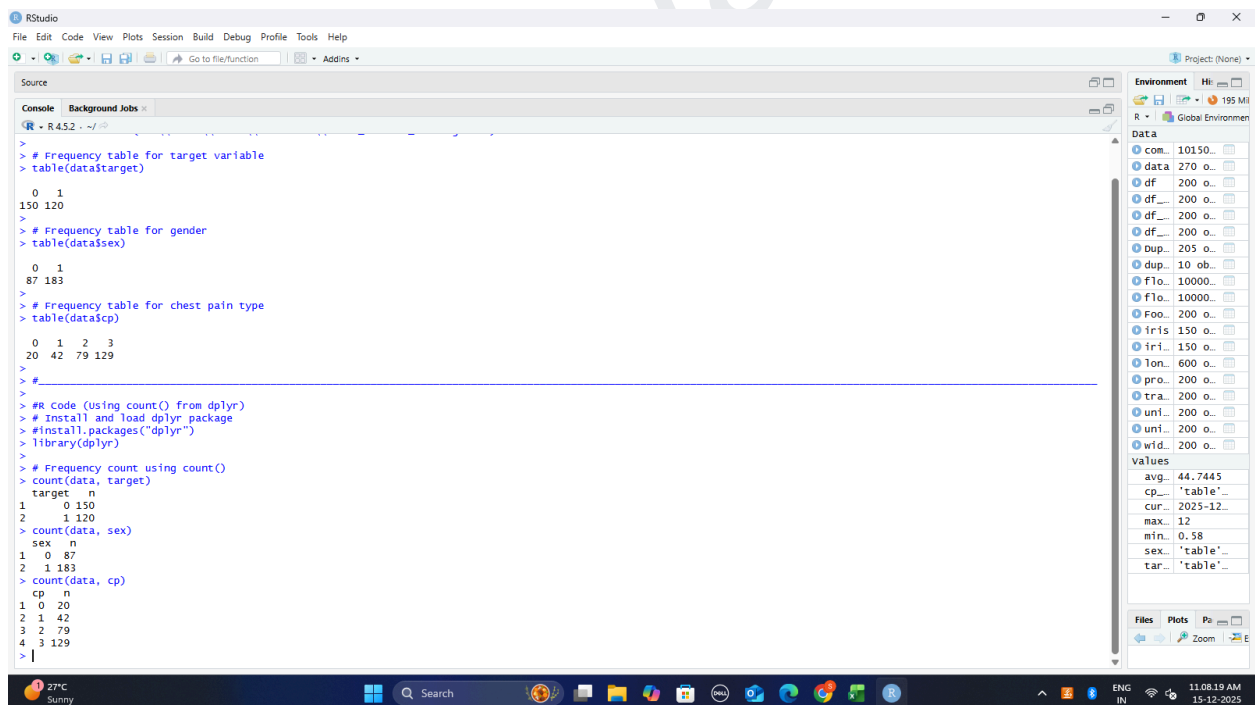
Roll No :- S081

# Sheth L.U.J. & Sir M.V. College



The screenshot shows the RStudio interface with a script file named 'S081\_R\_Practical2\_M2.R'. The code includes comments and functions for loading data, creating frequency tables, and installing the dplyr package. The Environment pane on the right lists various objects like 'com', 'data', 'df', 'dup', 'f1o', 'f1o...', 'Foo', 'iris', 'lon', 'pro', 'tra', 'uni', 'uni...', 'wid', and 'wid...'. The Console pane at the bottom shows the output of the code.

```
1 #R code (using table())
2 # Load dataset
3 data <- read.csv("c:\\users\\itlab\\Downloads\\Heart_disease_statlog.csv")
4
5 # Frequency table for target variable
6 table(data$target)
7
8 # Frequency table for gender
9 table(data$sex)
10
11 # Frequency table for chest pain type
12 table(data$cp)
13
14 #
15
16 #R code (using count() from dplyr)
17 # Install and load dplyr package
18 #install.packages("dplyr")
19 library(dplyr)
20
21 # Frequency count using count()
22 count(data, target)
23 count(data, sex)
24 count(data, cp)
25
```



The screenshot shows the RStudio interface with the same script file. The Console pane now displays the output of the code, showing frequency tables for 'target', 'sex', and 'cp'. The Environment pane on the right is also visible, showing the same objects as in the first screenshot.

```
> # Frequency table for target variable
> table(data$target)
0 1
150 120
> # Frequency table for gender
> table(data$sex)
0 1
87 183
> # Frequency table for chest pain type
> table(data$cp)
0 1 2 3
20 42 79 129
> #
> #R code (using count() from dplyr)
> # Install and load dplyr package
> #install.packages("dplyr")
> library(dplyr)
> # Frequency count using count()
> count(data, target)
target n
1 0 150
2 1 120
> count(data, sex)
sex n
1 0 87
2 1 183
> count(data, cp)
cp n
1 0 20
2 1 42
3 2 79
4 3 129
>
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

## 3. Creating cross-tabulations and two-way tables using `table()` (R).

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Environment Hi 195 MB

R Global Environment

Data

- com... 10150...
- data 270 o...
- df 200 o...
- df\_ 200 o...
- df\_ 200 o...
- df\_ 200 o...
- dup... 205 o...
- dup... 10 ob...
- flo... 10000...
- flo... 10000...
- Foo... 200 o...
- iris 150 o...
- iri... 150 o...
- Ion... 600 o...
- pro... 200 o...
- tra... 200 o...
- uni... 200 o...
- uni... 200 o...
- wid... 200 o...

values

- avg... 44.7445
- cp... 'table'...
- cur... 2025-12...
- max... 12
- min... 0.58
- sex... 'table'...
- tar... 'table'...

Files Plots Pa

Console

Showing 1 to 27 of 270 entries, 14 total columns

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1	70	1	3	130	322	0	2	109	0	2.4	1	3	1	1
2	67	0	2	115	564	0	2	160	0	1.6	1	0	3	0
3	57	1	1	124	261	0	0	141	0	0.3	0	0	3	1
4	64	1	3	128	263	0	0	105	1	0.2	1	1	3	0
5	74	0	1	120	269	0	2	121	1	0.2	0	1	1	0
6	65	1	3	120	177	0	0	140	0	0.4	0	0	3	0
7	56	1	2	130	256	1	2	142	1	0.6	1	1	2	1
8	59	1	3	110	239	0	2	142	1	1.2	1	1	3	1
9	60	1	3	140	293	0	2	170	0	1.2	1	2	3	1
10	63	0	3	150	407	0	2	154	0	4.0	1	3	3	1
11	59	1	3	135	234	0	0	161	0	0.5	1	0	3	0
12	53	1	3	142	226	0	2	111	1	0.0	0	0	3	0
13	44	1	2	140	235	0	2	180	0	0.0	0	0	1	0
14	61	1	0	134	234	0	0	145	0	2.6	1	2	1	1
15	57	0	3	128	303	0	2	159	0	0.0	0	1	1	0
16	71	0	3	112	149	0	0	125	0	1.6	1	0	1	0
17	46	1	3	140	311	0	0	120	1	1.8	1	2	3	1
18	53	1	3	140	203	1	2	155	1	3.1	2	0	3	1
19	64	1	0	110	211	0	2	144	1	1.8	1	0	1	0
20	40	1	0	140	199	0	0	178	1	1.4	0	0	3	0
21	67	1	3	120	229	0	2	129	1	2.6	1	2	3	1
22	48	1	1	130	245	0	2	180	0	0.2	1	0	1	0
23	43	1	3	115	303	0	0	161	0	1.2	1	0	1	0
24	47	1	3	112	204	0	0	143	0	0.1	0	0	1	0
25	54	0	1	132	288	1	2	159	1	0.0	0	1	1	0
26	48	0	2	130	275	0	0	139	0	0.2	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

27°C Sunny

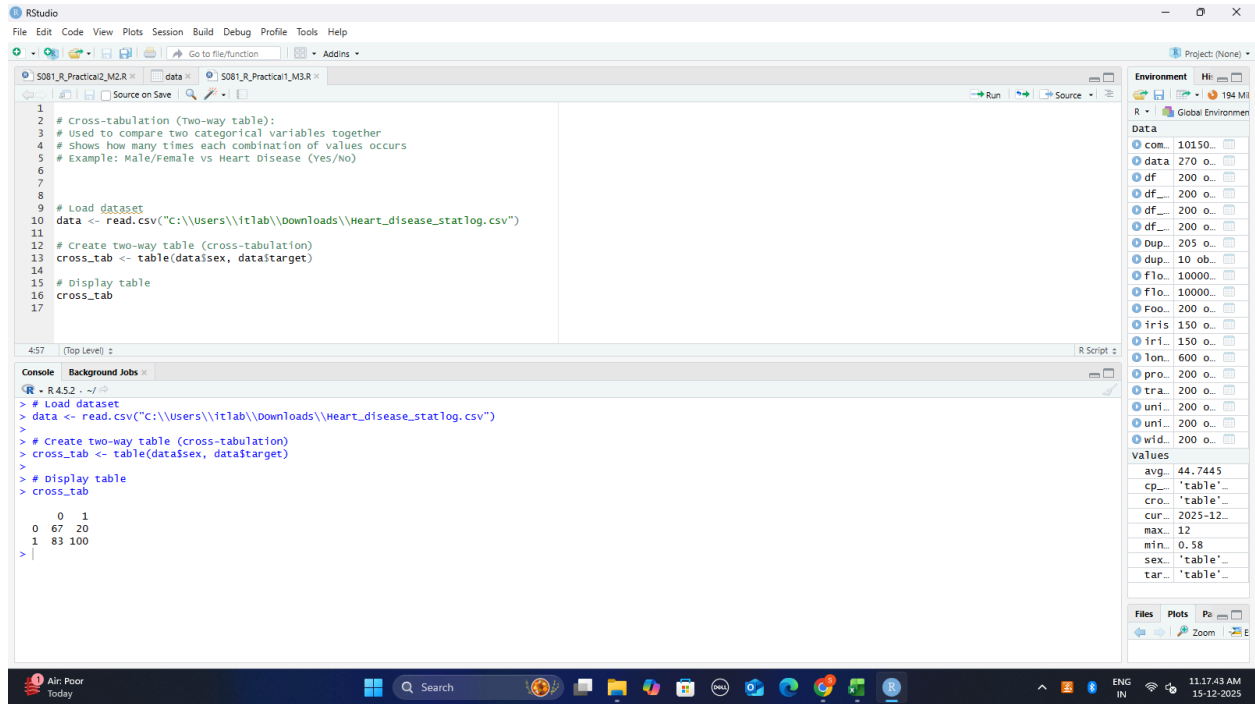
Search

ENG IN 11:10:17 AM 15-12-2025

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College



The screenshot displays the RStudio interface with a script editor, console, and environment pane. The script defines a cross-tabulation for heart disease status by sex. The console shows the execution of the script, and the environment pane lists the objects created.

```
1 # Cross-tabulation (Two-way table):  
2 # Used to compare two categorical variables together  
3 # Shows how many times each combination of values occurs  
4 # Example: Male/Female vs Heart Disease (Yes/No)  
5  
6  
7  
8  
9 # Load dataset  
10 data <- read.csv("c:\\Users\\itlab\\Downloads\\Heart_disease_statlog.csv")  
11  
12 # Create two-way table (cross-tabulation)  
13 cross_tab <- table(data$sex, data$target)  
14  
15 # Display table  
16 cross_tab  
17
```

```
> # Load dataset  
> data <- read.csv("c:\\Users\\itlab\\Downloads\\Heart_disease_statlog.csv")  
>  
> # Create two-way table (cross-tabulation)  
> cross_tab <- table(data$sex, data$target)  
>  
> # Display table  
> cross_tab  
      0  1  
0  67  20  
1  83 100  
> |
```

Environment pane objects:

- com... 10150...
- data 270 0...
- df 200 0...
- df\_ 200 0...
- df\_ 200 0...
- df\_ 200 0...
- Dup... 205 0...
- dup... 10 ob...
- f1o... 10000...
- f1o... 10000...
- Foo... 200 0...
- iris 150 0...
- irf... 150 0...
- lon... 600 0...
- pro... 200 0...
- tra... 200 0...
- uni... 200 0...
- uni... 200 0...
- wid... 200 0...

values

- avg... 44.7445
- cp... 'table'...
- cro... 'table'...
- cur... 2025-12...
- max... 12
- min... 0.58
- sex... 'table'...
- tar... 'table'...

**Name :- Priya Gupta**

**Roll No :- S081**

# Sheth L.U.J. & Sir M.V. College

## 4. Performing one-sample $t$ -tests using $t.test()$ (R).

	Job_Title	Average_Salary	Years_Experience	Education_Level	AI_Exposure_Index	Tech_Growth_Factor	Automation_Probability_2030	Risk_Category	Skill_1	Skill_2	Skill_3	Skill_4	Skill_5	Skill_6	Skill_7	Skill_8	Skill_9
1	Security Guard	45795	28	Master's	0.18	1.28	0.85	High	0.45	0.10	0.46	0.33	0.14	0.65	0.06	0.72	
2	Research Scientist	133355	20	PhD	0.62	1.11	0.05	Low	0.02	0.52	0.40	0.05	0.97	0.23	0.09	0.62	
3	Construction Worker	146216	2	High School	0.86	1.18	0.81	High	0.01	0.94	0.56	0.39	0.02	0.23	0.24	0.68	
4	Software Engineer	136530	13	PhD	0.39	0.68	0.60	Medium	0.43	0.21	0.57	0.03	0.84	0.45	0.40	0.93	
5	Financial Analyst	70397	22	High School	0.52	1.46	0.64	Medium	0.75	0.54	0.59	0.97	0.61	0.28	0.30	0.17	
6	AI Engineer	92592	11	Master's	0.29	0.51	0.10	Low	0.71	0.79	0.61	0.93	0.65	0.91	0.85	0.45	
7	Mechanic	107373	23	PhD	0.67	1.09	0.41	Medium	0.56	0.38	0.97	0.85	0.72	0.34	0.26	0.04	
8	Teacher	53419	12	High School	0.20	1.40	0.17	Low	0.56	0.70	0.14	0.60	0.54	0.20	0.94	0.60	
9	HR Specialist	139225	12	Master's	0.30	0.61	0.48	Medium	0.22	0.42	0.88	0.32	0.12	0.36	0.91	0.27	
10	Customer Support	85016	2	High School	0.01	1.01	0.80	High	0.22	0.12	0.34	0.94	0.32	0.52	0.70	0.36	
11	UX Researcher	82733	6	High School	0.50	0.80	0.41	Medium	0.04	0.61	0.50	0.05	0.28	0.91	0.24	0.14	
12	Financial Analyst	117455	22	High School	0.67	1.26	0.40	Medium	0.73	0.37	0.63	0.63	0.54	0.09	0.84	0.32	
13	Lawyer	79811	27	High School	0.68	0.52	0.50	Medium	0.23	0.65	0.17	0.69	0.39	0.94	0.14	0.34	
14	Data Scientist	115981	9	High School	0.26	1.16	0.63	Medium	0.56	0.53	0.24	0.09	0.90	0.90	0.63	0.34	
15	Research Scientist	96690	19	Master's	0.89	1.28	0.21	Low	0.08	0.16	0.90	0.61	0.01	0.10	0.66	0.01	
16	Graphic Designer	32869	2	High School	0.65	0.72	0.58	Medium	0.24	0.33	0.75	0.65	0.85	0.66	0.57	0.09	
17	Teacher	36893	29	Bachelor's	0.97	0.89	0.27	Low	0.63	0.79	0.50	0.58	0.49	0.20	0.72	0.28	
18	Teacher	103744	11	Bachelor's	0.94	1.45	0.28	Low	0.37	0.02	0.93	0.43	0.97	0.96	0.85	0.29	
19	Retail Worker	148015	2	PhD	0.17	1.06	0.93	High	0.70	0.57	0.10	0.62	0.99	0.14	0.52	0.88	
20	Doctor	108069	15	Master's	0.55	0.80	0.15	Low	0.26	0.61	0.08	0.01	0.63	0.19	0.07	0.40	
21	AI Engineer	43403	1	High School	0.09	1.08	0.06	Low	0.47	0.54	0.29	0.59	0.03	0.04	0.82	0.36	
22	HR Specialist	49508	27	Bachelor's	0.22	1.12	0.33	Medium	0.05	0.53	0.54	0.64	0.73	0.98	0.52	0.32	
23	Teacher	58251	25	PhD	0.08	0.53	0.29	Low	0.84	0.70	0.41	0.17	0.16	0.25	0.55	0.71	
24	Financial Analyst	33343	28	High School	0.74	1.05	0.54	Medium	0.42	0.25	0.36	0.76	0.01	0.12	0.05	0.04	
25	AI Engineer	125435	15	PhD	0.10	0.99	0.17	Low	0.17	0.43	0.40	0.62	0.64	0.05	0.37	0.63	
26	Software Engineer	39540	12	High School	0.16	0.57	0.56	Medium	0.03	0.59	0.94	0.58	0.39	0.64	0.46	0.55	

Name :- Priya Gupta  
Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
S081_R_Practical2_M2.R S081_R_Practical4_M2.R data
Run Source
1 # -----
2 # AIM:
3 # Perform a one-sample t-test to check
4 # whether the mean of a sample is significantly
5 # different from a given value (hypothesized mean)
6 # -----
7 # Load dataset
8 data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
9
10 # Check structure of dataset
11 str(data)
12
13 # Choose a numeric column for the t-test
14 # Example: Average_Salary
15 # Ensure it is numeric
16 data$Average_Salary <- as.numeric(data$Average_Salary)
17
18 # Check summary
19 summary(data$Average_Salary)
20
21 # Perform one-sample t-test
22 # Hypothesized mean (mu) = 100000 (example, adjust as needed)
23 t_test_result <- t.test(data$Average_Salary,
24                          mu = 100000,
25                          na.rm = TRUE)
26
27 # Display the t-test result
28 t_test_result
29
30 # -----
31 # INTERPRETATION:
32 # If p-value < 0.05 -> Reject H0 (sample mean is significantly different from 100000)
33 # If p-value >= 0.05 -> Fail to reject H0 (sample mean is NOT significantly different from 100000)
34 # -----
35
36
361 (Untitled)
R Script
Console
27°C Sunny
Search
11:35:07 AM 15-12-2025
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Background Jobs
R - R452 - ~/
data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
> str(data)
'data.frame': 3000 obs. of 18 variables:
 $ Job_title: chr "Security Guard" "Research scientist" "Construction worker" "Software Engineer" ...
 $ Average_Salary: int 45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
 $ Years_Experience: int 28 20 2 13 22 11 23 12 12 2 ...
 $ Education_Level: chr "Master's" "Phd" "High School" "Phd" ...
 $ AI_Exposure_Index: num 0.18 0.62 0.86 0.39 0.52 0.29 0.67 0.2 0.3 0.01 ...
 $ TechGrowth_Factor: num 1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
 $ Automation_Probability_2030: num 0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
 $ Risk_Category: chr "High" "Low" "High" "Medium" ...
 $ sk11l_1: num 0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
 $ sk11l_2: num 0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
 $ sk11l_3: num 0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ sk11l_4: num 0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ sk11l_5: num 0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ sk11l_6: num 0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
 $ sk11l_7: num 0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
 $ sk11l_8: num 0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
 $ sk11l_9: num 0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
 $ sk11l_10: num 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
> data$Average_Salary <- as.numeric(data$Average_Salary)
> summary(data$Average_Salary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
30030   58640   89318   89372 119087 149798
> t_test_result <- t.test(data$Average_Salary,
+                          mu = 100000,
+                          na.rm = TRUE)
> # Display the t-test result
> t_test_result

One Sample t-test

data:  data$Average_Salary
t = -16.82, df = 2999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 1e+05
95 percent confidence interval:
 88133.37 90611.19
sample estimates:
mean of x
89372.28
> |
Environment
Data
com_ 10150...
data 3000 ...
df 200 o...
df 200 o...
df 200 o...
df 200 o...
df 200 o...
Dup_ 10 ob...
Dup_ 205 o...
flo_ 10000...
flo_ 10000...
Foo_ 200 o...
iris 150 o...
iri_ 150 o...
lon_ 600 o...
pro_ 200 o...
t_t_ List ...
tra_ 200 o...
uni_ 200 o...
uni_ 200 o...
wid_ 200 o...
values
avg_ 44.7445
cp_ 'table'...
cro_ 'table'...
cur_ 2025-12...
max_ 12
min_ 0.58
sex_ 'table'...
tar_ 'table'...
Files Plots Pa
Zoom
```

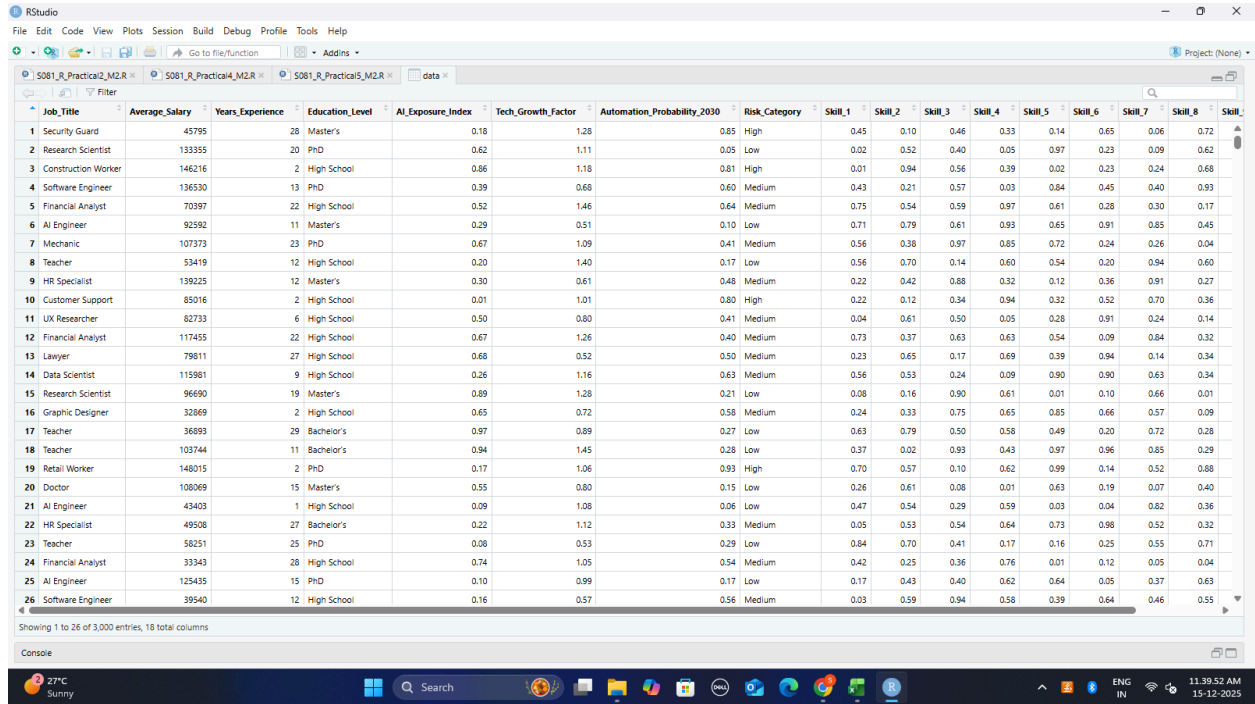
Name :- Priya Gupta

Roll No :- S081



# Sheth L.U.J. & Sir M.V. College

## 5. Performing independent two-sample t-tests using `t.test()` with grouping (R).



	Job Title	Average Salary	Years Experience	Education Level	AI Exposure Index	Tech Growth Factor	Automation Probability 2030	Risk Category	Skill_1	Skill_2	Skill_3	Skill_4	Skill_5	Skill_6	Skill_7	Skill_8	Skill_9
1	Security Guard	45795	28	Master's	0.18	1.28	0.85	High	0.45	0.10	0.46	0.33	0.14	0.65	0.06	0.72	
2	Research Scientist	133355	20	PhD	0.62	1.11	0.05	Low	0.02	0.52	0.40	0.05	0.97	0.23	0.09	0.62	
3	Construction Worker	146216	2	High School	0.86	1.18	0.81	High	0.01	0.94	0.56	0.39	0.02	0.23	0.24	0.68	
4	Software Engineer	136530	13	PhD	0.39	0.68	0.60	Medium	0.43	0.21	0.57	0.03	0.84	0.45	0.40	0.93	
5	Financial Analyst	70397	22	High School	0.52	1.46	0.64	Medium	0.75	0.54	0.59	0.97	0.61	0.28	0.30	0.17	
6	AI Engineer	92592	11	Master's	0.29	0.51	0.10	Low	0.71	0.79	0.61	0.93	0.65	0.91	0.85	0.45	
7	Mechanic	107373	23	PhD	0.67	1.09	0.41	Medium	0.56	0.38	0.97	0.85	0.72	0.24	0.26	0.04	
8	Teacher	53419	12	High School	0.20	1.40	0.17	Low	0.56	0.70	0.14	0.60	0.54	0.20	0.94	0.60	
9	HR Specialist	139225	12	Master's	0.30	0.61	0.48	Medium	0.22	0.42	0.88	0.32	0.12	0.36	0.91	0.27	
10	Customer Support	85016	2	High School	0.01	1.01	0.80	High	0.22	0.12	0.34	0.94	0.32	0.52	0.70	0.36	
11	UX Researcher	82733	6	High School	0.50	0.80	0.41	Medium	0.04	0.61	0.50	0.05	0.28	0.91	0.24	0.14	
12	Financial Analyst	117455	22	High School	0.67	1.26	0.40	Medium	0.73	0.37	0.63	0.63	0.54	0.09	0.84	0.32	
13	Lawyer	79611	27	High School	0.68	0.52	0.50	Medium	0.23	0.65	0.17	0.69	0.39	0.94	0.14	0.34	
14	Data Scientist	115981	9	High School	0.26	1.16	0.63	Medium	0.56	0.53	0.24	0.09	0.90	0.90	0.63	0.34	
15	Research Scientist	96690	19	Master's	0.89	1.28	0.21	Low	0.08	0.16	0.90	0.61	0.01	0.10	0.66	0.01	
16	Graphic Designer	32869	2	High School	0.65	0.72	0.58	Medium	0.24	0.33	0.75	0.65	0.85	0.66	0.57	0.09	
17	Teacher	36893	29	Bachelor's	0.97	0.89	0.27	Low	0.63	0.79	0.50	0.58	0.49	0.20	0.72	0.28	
18	Teacher	103744	11	Bachelor's	0.94	1.45	0.37	Low	0.37	0.02	0.93	0.43	0.97	0.96	0.85	0.29	
19	Retail Worker	148015	2	PhD	0.17	1.06	0.93	High	0.70	0.57	0.10	0.62	0.99	0.14	0.52	0.88	
20	Doctor	108069	15	Master's	0.55	0.80	0.15	Low	0.26	0.61	0.08	0.01	0.63	0.19	0.07	0.40	
21	AI Engineer	43403	1	High School	0.09	1.08	0.06	Low	0.47	0.54	0.29	0.59	0.03	0.04	0.82	0.36	
22	HR Specialist	49508	27	Bachelor's	0.22	1.12	0.33	Medium	0.05	0.53	0.54	0.64	0.73	0.98	0.52	0.32	
23	Teacher	58251	25	PhD	0.08	0.53	0.29	Low	0.84	0.70	0.41	0.17	0.16	0.25	0.55	0.71	
24	Financial Analyst	33343	28	High School	0.74	1.05	0.54	Medium	0.42	0.25	0.36	0.76	0.01	0.12	0.05	0.04	
25	AI Engineer	125435	15	PhD	0.10	0.99	0.17	Low	0.17	0.43	0.40	0.62	0.64	0.05	0.37	0.63	
26	Software Engineer	39540	12	High School	0.16	0.57	0.56	Medium	0.03	0.59	0.94	0.58	0.39	0.64	0.46	0.55	

Name :- Priya Gupta  
Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)
S081_R_Practical2_M2.R S081_R_Practical4_M2.R S081_R_Practical5_M2.R data
1 # -----
2 # AIM:
3 # Perform independent two-sample t-tests
4 # to check whether the means of a numeric variable
5 # are significantly different between two groups
6 # -----
7
8 # Load dataset
9 data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
10
11 # Check structure of dataset
12 str(data)
13
14 # Ensure numeric column is numeric
15 data$Average_Salary <- as.numeric(data$Average_Salary)
16
17 # Check the groups
18 table(data$Risk_Category)
19
20 # Example: Compare Average_Salary between "High" vs "Low" risk categories
21 # Subset data for only these two groups
22 subset_data <- data[data$Risk_Category %in% c("High", "Low"), ]
23
24 # Perform independent two-sample t-test
25 t_test_result <- t.test(Average_Salary ~ Risk_Category,
26                         data = subset_data,
27                         var.equal = FALSE) # Welch t-test (does not assume equal variance)
28
29 # Display the result
30 t_test_result
31
32 # -----
33 # INTERPRETATION:
34 # If p-value < 0.05 -> Reject H0 (means are significantly different)
35 # If p-value >= 0.05 -> Fail to reject H0 (no significant difference)
36 # -----
37
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)
Source
Console Background Jobs
R - R452 - ~/
> # Load dataset
> data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
>
> # Check structure of dataset
> str(data)
'data.frame': 3000 obs. of 18 variables:
 $ Job_Title      : chr "Security Guard" "Research Scientist" "Construction worker" "Software Engineer" ...
 $ Average_Salary : int 45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
 $ Years_Experience : int 28 20 2 13 22 11 23 12 12 2 ...
 $ Education_Level : chr "Master's" "PhD" "High School" "phd" ...
 $ AI_Exposure_Index : num 0.18 0.62 0.86 0.39 0.52 0.29 0.67 0.2 0.3 0.01 ...
 $ Tech_Growth_Factor : num 1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
 $ Automation_Probability_2030 : num 0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
 $ Risk_Category   : chr "High" "Low" "High" "Medium" ...
 $ Skill1_1       : num 0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
 $ Skill1_2       : num 0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
 $ Skill1_3       : num 0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ Skill1_4       : num 0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ Skill1_5       : num 0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ Skill1_6       : num 0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
 $ Skill1_7       : num 0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
 $ Skill1_8       : num 0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
 $ Skill1_9       : num 0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
 $ Skill1_10      : num 0 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
>
> # Ensure numeric column is numeric
> data$Average_Salary <- as.numeric(data$Average_Salary)
>
> # Check the groups
> table(data$Risk_Category)
High Low Medium
740 739 1521
>
> # Example: Compare Average_Salary between "High" vs "Low" risk categories
> # Subset data for only these two groups
> subset_data <- data[data$Risk_Category %in% c("High", "Low"), ]
>
> # Perform independent two-sample t-test
> t_test_result <- t.test(Average_Salary ~ Risk_Category,
+                         data = subset_data,
+                         var.equal = FALSE) # Welch t-test (does not assume equal variance)
>
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Background Jobs

R - R452 - ~/
$ ski11_3 : num 0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
$ ski11_4 : num 0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
$ ski11_5 : num 0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
$ ski11_6 : num 0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
$ ski11_7 : num 0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
$ ski11_8 : num 0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
$ ski11_9 : num 0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
$ ski11_10 : num 0 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
>
> # Ensure numeric column is numeric
> data$Average_Salary <- as.numeric(data$Average_Salary)
>
> # Check the groups
> table(data$Risk_Category)

    High    Low Medium 
    740    739   1521 
>
> # Example: Compare Average_Salary between "High" vs "Low" risk categories
> # Subset data for only these two groups
> subset_data <- data[data$Risk_Category %in% c("High", "Low"), ]
>
> # Perform independent two-sample t-test
> t_test_result <- t.test(Average_Salary ~ Risk_Category,
+ data = subset_data,
+ var.equal = FALSE) # Welch t-test (does not assume equal variance)
>
> # Display the result
> t_test_result

welch Two Sample t-test

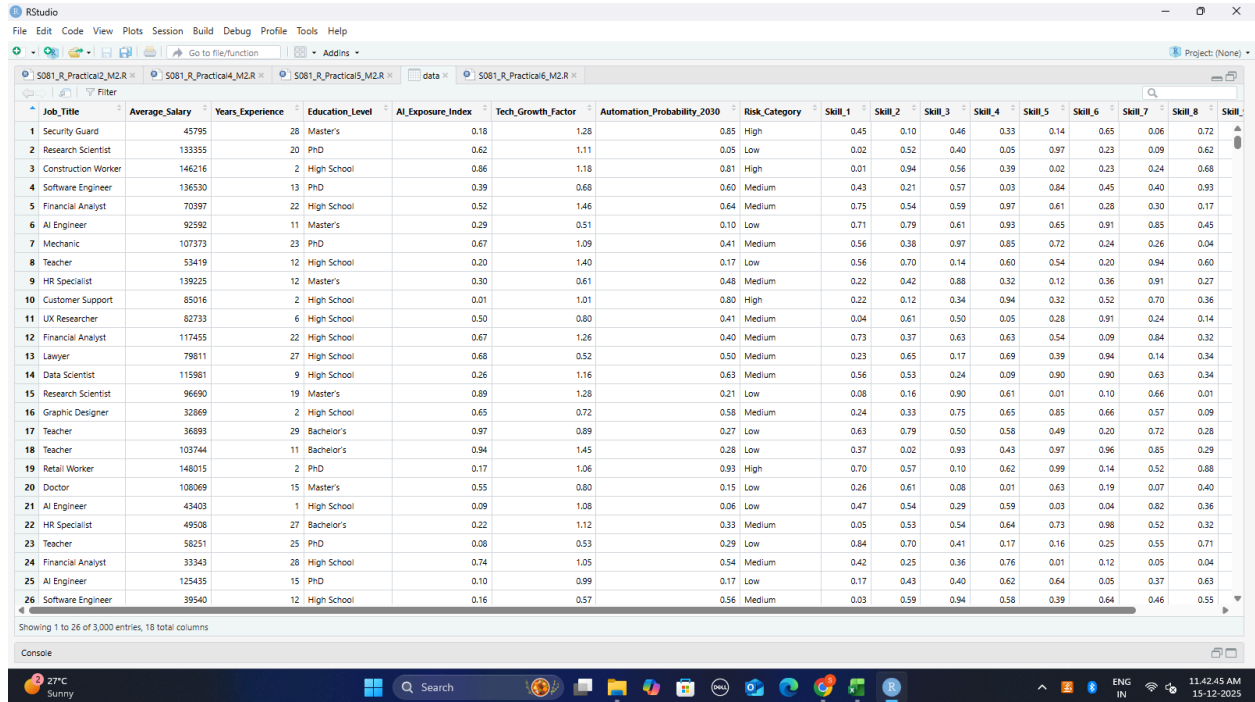
data: Average_Salary by Risk_Category
t = -1.0111, df = 1475.2, p-value = 0.3121
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
 -5313.624 1698.955
sample estimates:
mean in group High mean in group Low
      87359.44      89166.78
> |
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

## 6. Performing paired t-tests using `t.test(paired=TRUE)` (R).

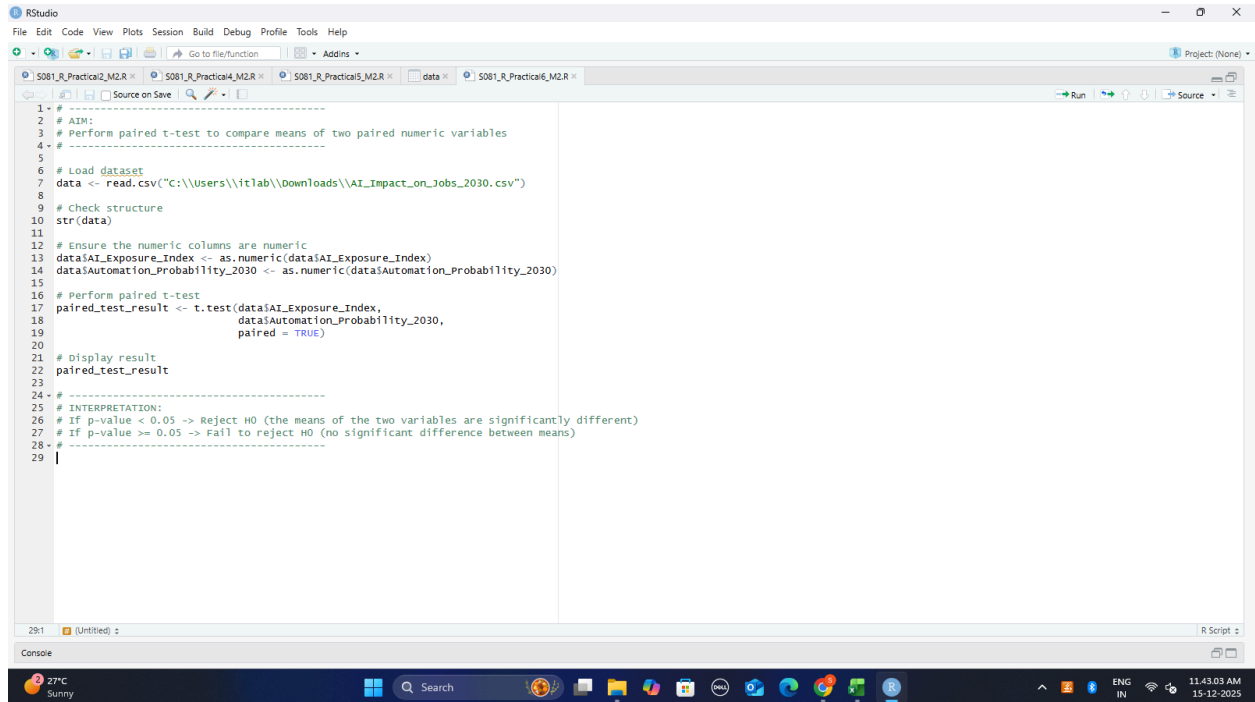


	Job_Title	Average_Salary	Years_Experience	Education_Level	AI_Exposure_Index	Tech_Growth_Factor	Automation_Probability_2030	Risk_Category	Skill_1	Skill_2	Skill_3	Skill_4	Skill_5	Skill_6	Skill_7	Skill_8	Skill_9
1	Security Guard	45795	28	Master's	0.18	1.28	0.85	High	0.45	0.10	0.46	0.33	0.14	0.65	0.06	0.72	
2	Research Scientist	133355	20	PhD	0.62	1.11	0.05	Low	0.02	0.52	0.40	0.05	0.97	0.23	0.09	0.62	
3	Construction Worker	146216	2	High School	0.86	1.18	0.81	High	0.01	0.94	0.56	0.39	0.02	0.23	0.24	0.68	
4	Software Engineer	136530	13	PhD	0.39	0.68	0.60	Medium	0.43	0.21	0.57	0.03	0.84	0.45	0.40	0.93	
5	Financial Analyst	70397	22	High School	0.52	1.46	0.64	Medium	0.75	0.54	0.59	0.97	0.61	0.28	0.30	0.17	
6	AI Engineer	92592	11	Master's	0.29	0.51	0.10	Low	0.71	0.79	0.61	0.93	0.65	0.91	0.85	0.45	
7	Mechanic	107373	23	PhD	0.67	1.09	0.41	Medium	0.56	0.38	0.97	0.85	0.72	0.34	0.26	0.04	
8	Teacher	53419	12	High School	0.20	1.40	0.17	Low	0.56	0.70	0.14	0.60	0.54	0.20	0.94	0.60	
9	HR Specialist	139225	12	Master's	0.30	0.61	0.48	Medium	0.22	0.42	0.88	0.32	0.12	0.36	0.91	0.27	
10	Customer Support	85016	2	High School	0.01	1.01	0.80	High	0.22	0.12	0.34	0.94	0.32	0.52	0.70	0.36	
11	UX Researcher	82733	6	High School	0.50	0.80	0.41	Medium	0.04	0.61	0.50	0.05	0.28	0.91	0.24	0.14	
12	Financial Analyst	117455	22	High School	0.67	1.26	0.40	Medium	0.73	0.37	0.63	0.63	0.54	0.09	0.84	0.32	
13	Lawyer	79811	27	High School	0.68	0.52	0.50	Medium	0.23	0.65	0.17	0.69	0.39	0.94	0.14	0.34	
14	Data Scientist	115981	9	High School	0.26	1.16	0.63	Medium	0.56	0.53	0.24	0.09	0.90	0.90	0.63	0.34	
15	Research Scientist	96690	19	Master's	0.89	1.28	0.21	Low	0.08	0.16	0.90	0.61	0.01	0.10	0.66	0.01	
16	Graphic Designer	32869	2	High School	0.65	0.72	0.58	Medium	0.24	0.33	0.75	0.65	0.85	0.66	0.57	0.09	
17	Teacher	36893	29	Bachelor's	0.97	0.89	0.27	Low	0.63	0.79	0.50	0.58	0.49	0.20	0.72	0.28	
18	Teacher	103744	11	Bachelor's	0.94	1.45	0.28	Low	0.37	0.02	0.93	0.43	0.97	0.96	0.85	0.29	
19	Retail Worker	148015	2	PhD	0.17	1.06	0.93	High	0.70	0.57	0.10	0.62	0.99	0.14	0.52	0.88	
20	Doctor	108069	15	Master's	0.55	0.80	0.15	Low	0.26	0.61	0.08	0.01	0.63	0.19	0.07	0.40	
21	AI Engineer	43403	1	High School	0.09	1.08	0.06	Low	0.47	0.54	0.29	0.59	0.03	0.04	0.82	0.36	
22	HR Specialist	49508	27	Bachelor's	0.22	1.12	0.33	Medium	0.05	0.53	0.54	0.64	0.73	0.98	0.52	0.32	
23	Teacher	58251	25	PhD	0.08	0.53	0.29	Low	0.84	0.70	0.41	0.17	0.16	0.25	0.55	0.71	
24	Financial Analyst	33343	28	High School	0.74	1.05	0.54	Medium	0.42	0.25	0.36	0.76	0.01	0.12	0.05	0.04	
25	AI Engineer	125435	15	PhD	0.10	0.99	0.17	Low	0.17	0.43	0.40	0.62	0.64	0.05	0.37	0.63	
26	Software Engineer	39540	12	High School	0.16	0.57	0.56	Medium	0.03	0.59	0.94	0.58	0.39	0.64	0.46	0.55	

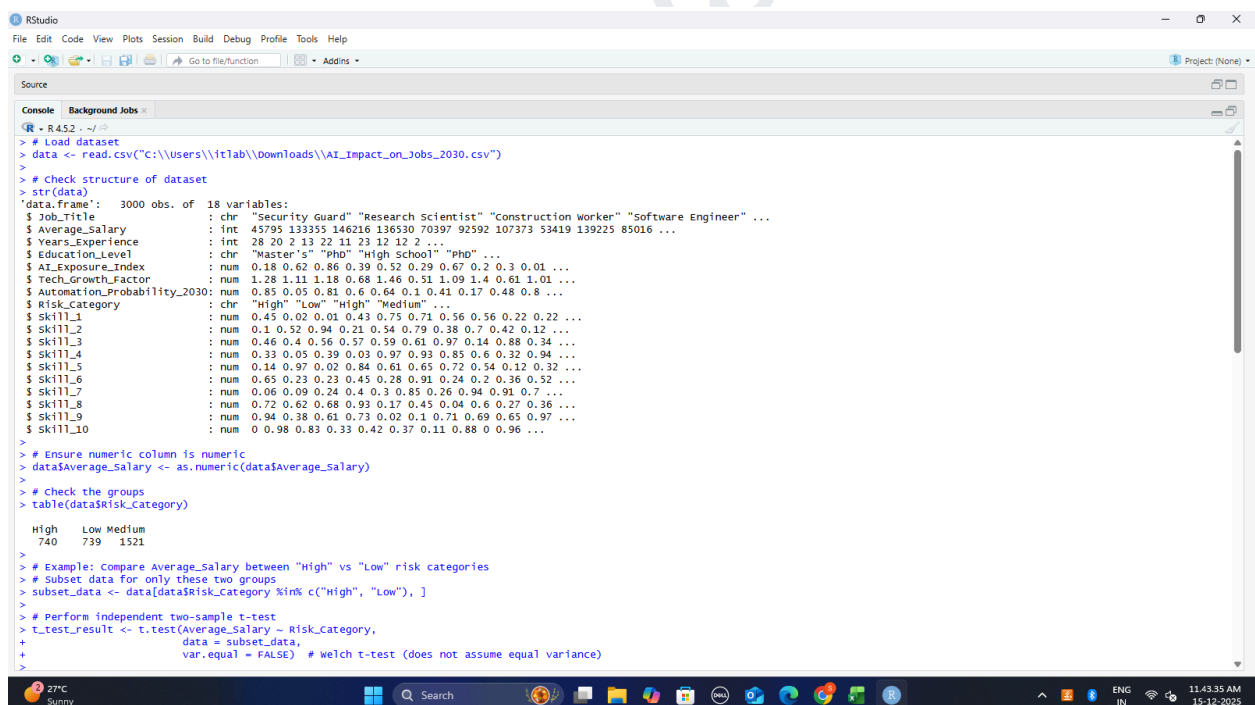
Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College



```
1- #-----
2 # AIM:
3 # Perform paired t-test to compare means of two paired numeric variables
4 #-----
5
6 # Load dataset
7 data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
8
9 # Check structure
10 str(data)
11
12 # Ensure the numeric columns are numeric
13 data$AI_Exposure_Index <- as.numeric(data$AI_Exposure_Index)
14 data$Automation_Probability_2030 <- as.numeric(data$Automation_Probability_2030)
15
16 # Perform paired t-test
17 paired_test_result <- t.test(data$AI_Exposure_Index,
18                             data$Automation_Probability_2030,
19                             paired = TRUE)
20
21 # display result
22 paired_test_result
23
24 #-----
25 # INTERPRETATION:
26 # If p-value < 0.05 -> Reject H0 (the means of the two variables are significantly different)
27 # If p-value >= 0.05 -> Fail to reject H0 (no significant difference between means)
28 #-----
29 |
```



```
> # Load dataset
> data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
>
> # Check structure of dataset
> str(data)
'data.frame': 3000 obs. of 18 variables:
 $ Job_Title      : chr  "Security Guard" "Research Scientist" "Construction worker" "Software Engineer" ...
 $ Average_Salary : int   45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
 $ Years_Experience : int   28 20 2 13 22 11 23 12 12 2 ...
 $ Education_Level : chr   "Master's" "PhD" "High School" "phd" ...
 $ AI_Exposure_Index : num  0.18 0.62 0.86 0.39 0.52 0.29 0.67 0.2 0.3 0.01 ...
 $ Tech_Growth_Factor : num  1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
 $ Automation_Probability_2030 : num  0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
 $ Risk_Category   : chr   "High" "Low" "High" "Medium" ...
 $ Skill_1         : num  0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
 $ Skill_2         : num  0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
 $ Skill_3         : num  0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ Skill_4         : num  0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ Skill_5         : num  0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ Skill_6         : num  0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
 $ Skill_7         : num  0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
 $ Skill_8         : num  0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
 $ Skill_9         : num  0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
 $ Skill_10        : num  0 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
>
> # Ensure numeric column is numeric
> data$Average_Salary <- as.numeric(data$Average_Salary)
>
> # Check the groups
> table(data$Risk_Category)
   High   Low Medium
   740   739  1521
>
> # Example: compare Average_Salary between "High" vs "Low" risk categories
> # Subset data for only these two groups
> subset_data <- data[data$Risk_Category %in% c("High", "Low"), ]
>
> # Perform independent two-sample t-test
> t_test_result <- t.test(Average_Salary ~ Risk_Category,
+                         data = subset_data,
+                         var.equal = FALSE) # Welch t-test (does not assume equal variance)
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Background Jobs

R 4.5.2 ~ /
> t_test_result <- t.test(Average_Salary ~ Risk_Category,
+ data = subset_data,
+ var.equal = FALSE) # Welch t-test (does not assume equal variance)
>
> # display the result
> t_test_result

welch two sample t-test

data: Average_Salary by Risk_Category
t = -1.0111, df = 1475.2, p-value = 0.3121
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
-5313.624 1698.955
sample estimates:
mean in group High mean in group Low
87359.44 89166.78

> # Load dataset
> data <- read.csv("c:\\users\\itlab\\downloads\\AI_Impact_on_Jobs_2030.csv")
>
> # Check structure
> str(data)
'data.frame': 3000 obs. of 18 variables:
 $ Job_Title : chr "Security Guard" "Research Scientist" "Construction Worker" "Software Engineer" ...
 $ Average_Salary : int 45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
 $ Years_Experience : int 28 20 2 13 22 11 23 12 12 2 ...
 $ Education_Level : chr "Master's" "PhD" "High School" "phd" ...
 $ AI_Exposure_Index : num 1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
 $ Tech_Growth_Factor : num 0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
 $ Automation_Probability_2030 : chr "High" "Low" "High" "Medium" ...
 $ Risk_Category : num 0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
 $ sk111_1 : num 0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
 $ sk111_2 : num 0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ sk111_3 : num 0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ sk111_4 : num 0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ sk111_5 : num 0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
 $ sk111_6 : num 0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
 $ sk111_7 : num 0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
 $ sk111_8 : num 0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
 $ sk111_9 : num 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
 $ sk111_10 : num 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
```

Name :- Priya Gupta

Roll No :- S081

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Background Jobs
R - R 4.5.2 - ~/
$ Average_Salary      : int  45795 133355 146216 136530 70397 92592 107373 53419 139225 85016 ...
$ Years_Experience    : int  28 20 2 13 22 11 23 12 12 2 ...
$ Education_Level     : chr   "Master's" "PhD" "High School" "PhD" ...
$ AI_Exposure_Index   : num   0.18 0.62 0.86 0.39 0.52 0.29 0.67 0.2 0.3 0.01 ...
$ Tech_Growth_Factor : num   1.28 1.11 1.18 0.68 1.46 0.51 1.09 1.4 0.61 1.01 ...
$ Automation_Probability_2030 : num  0.85 0.05 0.81 0.6 0.64 0.1 0.41 0.17 0.48 0.8 ...
$ Risk_Category       : chr   "High" "Low" "High" "Medium" ...
$ Skill_1            : num   0.45 0.02 0.01 0.43 0.75 0.71 0.56 0.56 0.22 0.22 ...
$ Skill_2            : num   0.1 0.52 0.94 0.21 0.54 0.79 0.38 0.7 0.42 0.12 ...
$ Skill_3            : num   0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
$ Skill_4            : num   0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
$ Skill_5            : num   0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
$ Skill_6            : num   0.65 0.23 0.23 0.45 0.28 0.91 0.24 0.2 0.36 0.52 ...
$ Skill_7            : num   0.06 0.09 0.24 0.4 0.3 0.85 0.26 0.94 0.91 0.7 ...
$ Skill_8            : num   0.72 0.62 0.68 0.93 0.17 0.45 0.04 0.6 0.27 0.36 ...
$ Skill_9            : num   0.94 0.38 0.61 0.73 0.02 0.1 0.71 0.69 0.65 0.97 ...
$ Skill_10           : num   0 0.98 0.83 0.33 0.42 0.37 0.11 0.88 0 0.96 ...
>
> # Ensure the numeric columns are numeric
> data$AI_Exposure_Index <- as.numeric(data$AI_Exposure_Index)
> data$Automation_Probability_2030 <- as.numeric(data$Automation_Probability_2030)
>
> # Perform paired t-test
> paired_test_result <- t.test(data$AI_Exposure_Index,
+                             data$Automation_Probability_2030,
+                             paired = TRUE)
>
> # Display result
> paired_test_result

Paired t-test

data:  data$AI_Exposure_Index and data$Automation_Probability_2030
t = -0.032195, df = 2999, p-value = 0.9743
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.01361869 0.01317869
sample estimates:
mean difference
 -0.00022
> |
```

Name :- Priya Gupta

Roll No :- S081