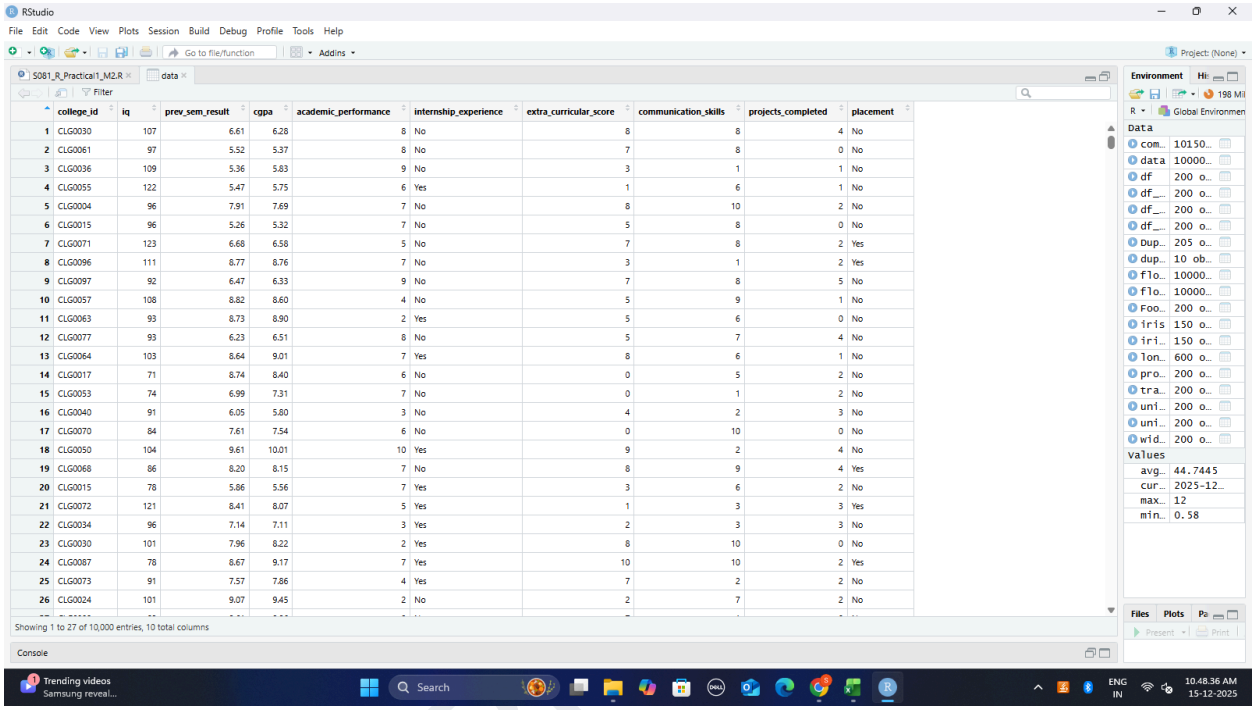
Sheth L.U.J. & Sir M.V. College

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

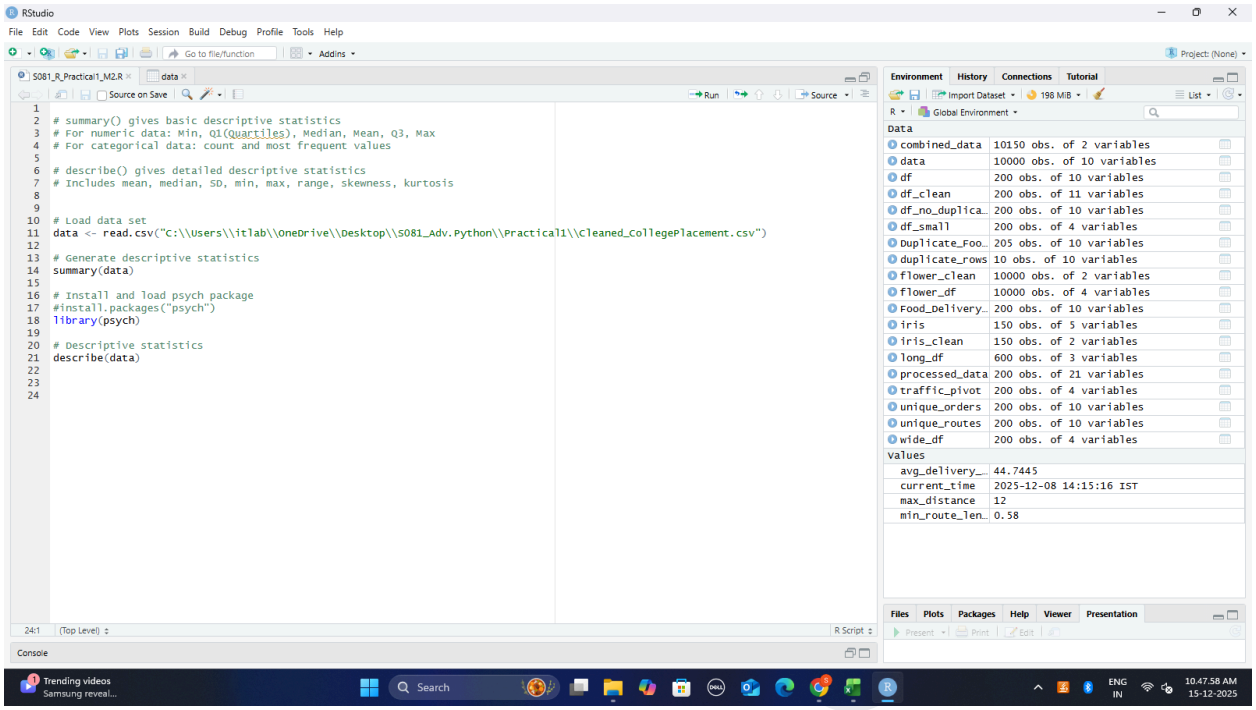


	college_id	iq	prev_sem_result	cgpa	academic_performance	internship_experience	extra_curricular_score	communication_skills	projects_completed	placement
1	CLG0030	107	6.61	6.28	8	No	8	8	4	No
2	CLG0061	97	5.52	5.37	8	No	7	8	0	No
3	CLG0036	109	5.36	5.83	9	No	3	1	1	No
4	CLG0055	122	5.47	5.75	6	Yes	1	6	1	No
5	CLG0004	96	7.91	7.69	7	No	8	10	2	No
6	CLG0015	96	5.26	5.32	7	No	5	8	0	No
7	CLG0071	123	6.68	6.58	5	No	7	8	2	Yes
8	CLG0096	111	8.77	8.76	7	No	3	1	2	Yes
9	CLG0097	92	6.47	6.33	9	No	7	8	5	No
10	CLG0057	108	8.82	8.60	4	No	5	9	1	No
11	CLG0063	93	8.73	8.90	2	Yes	5	6	0	No
12	CLG0077	93	6.23	6.51	8	No	5	7	4	No
13	CLG0064	103	8.64	9.01	7	Yes	8	6	1	No
14	CLG0017	71	8.74	8.40	6	No	0	5	2	No
15	CLG0053	74	6.99	7.31	7	No	0	1	2	No
16	CLG0040	91	6.05	5.80	3	No	4	2	3	No
17	CLG0070	84	7.61	7.54	6	No	0	10	0	No
18	CLG0050	104	9.61	10.01	10	Yes	9	2	4	No
19	CLG0068	86	8.20	8.15	7	No	8	9	4	Yes
20	CLG0015	78	5.86	5.56	7	Yes	3	6	2	No
21	CLG0072	121	8.41	8.07	5	Yes	1	3	3	Yes
22	CLG0034	96	7.14	7.11	3	Yes	2	3	3	No
23	CLG0030	101	7.96	8.22	2	Yes	8	10	0	No
24	CLG0087	78	8.67	9.17	7	Yes	10	10	2	Yes
25	CLG0073	91	7.57	7.86	4	Yes	7	2	2	No
26	CLG0024	101	9.07	9.45	2	No	2	7	2	No

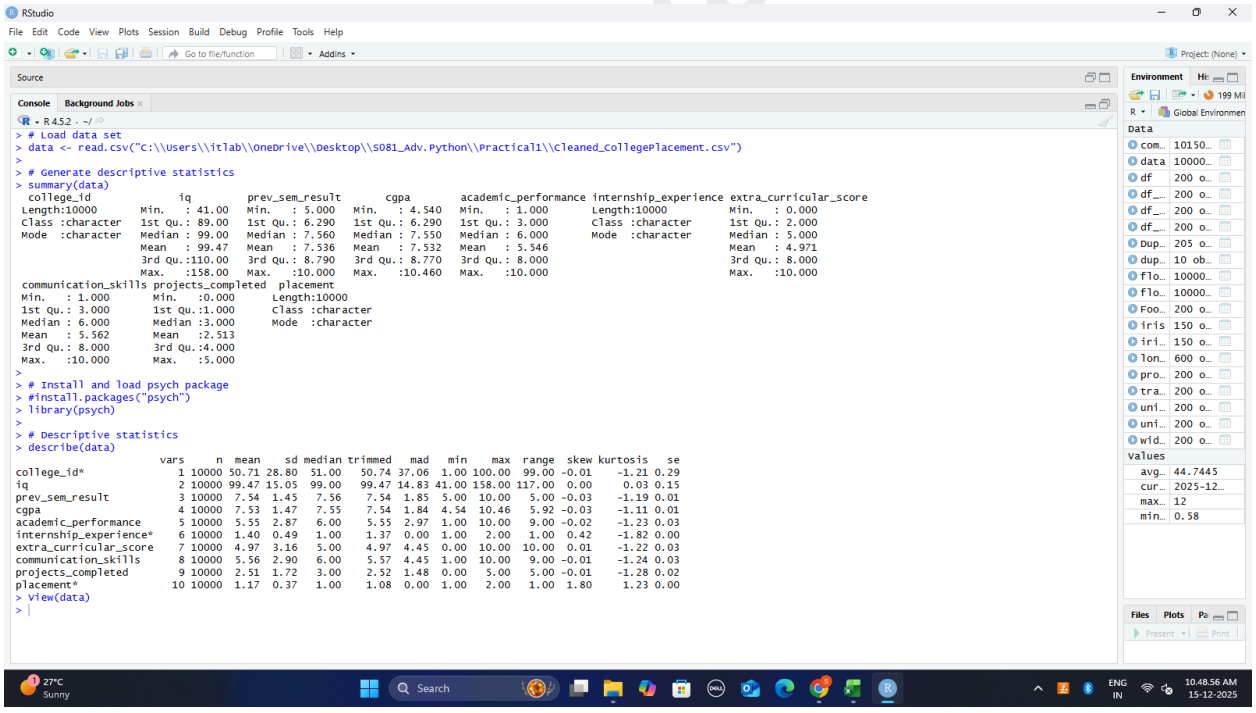
Name :- Priya Gupta

Roll No :- S081

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



```
# summary() gives basic descriptive statistics
# For numeric data: Min, Q1(Quartiles), Median, Mean, Q3, Max
# For categorical data: count and most frequent values

# describe() gives detailed descriptive statistics
# Includes mean, median, SD, min, max, range, skewness, kurtosis


# Load data set
data <- read.csv("C:\\Users\\itlab\\OneDrive\\Desktop\\S081_Adv.Python\\Practical1\\Cleaned_CollegePlacement.csv")

# Generate descriptive statistics
summary(data)

# Install and load psych package
#install.packages("psych")
library(psych)

# Descriptive statistics
describe(data)
```



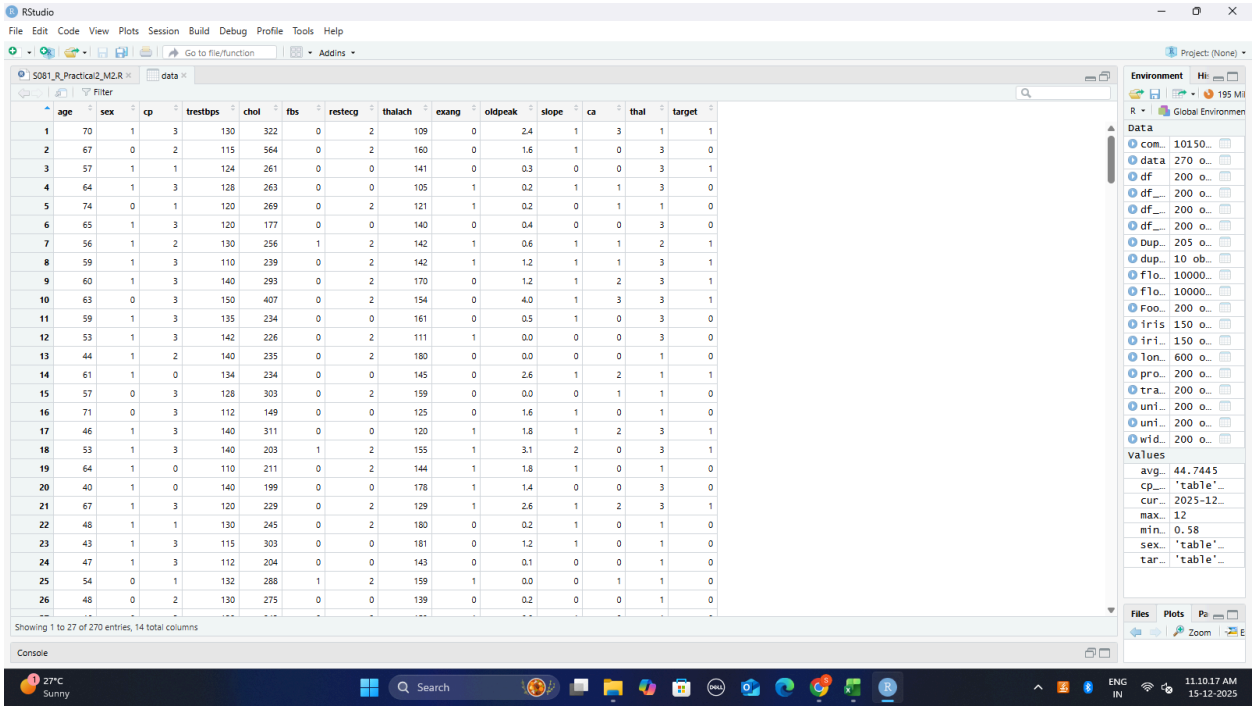
```
> # Load data set
> data <- read.csv("C:\\Users\\itlab\\OneDrive\\Desktop\\S081_Adv.Python\\Practical1\\Cleaned_CollegePlacement.csv")
>
> # Generate descriptive statistics
> summary(data)
  college_id              iq         prev_sem_result       cgpa        academic_performance internship_experience 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

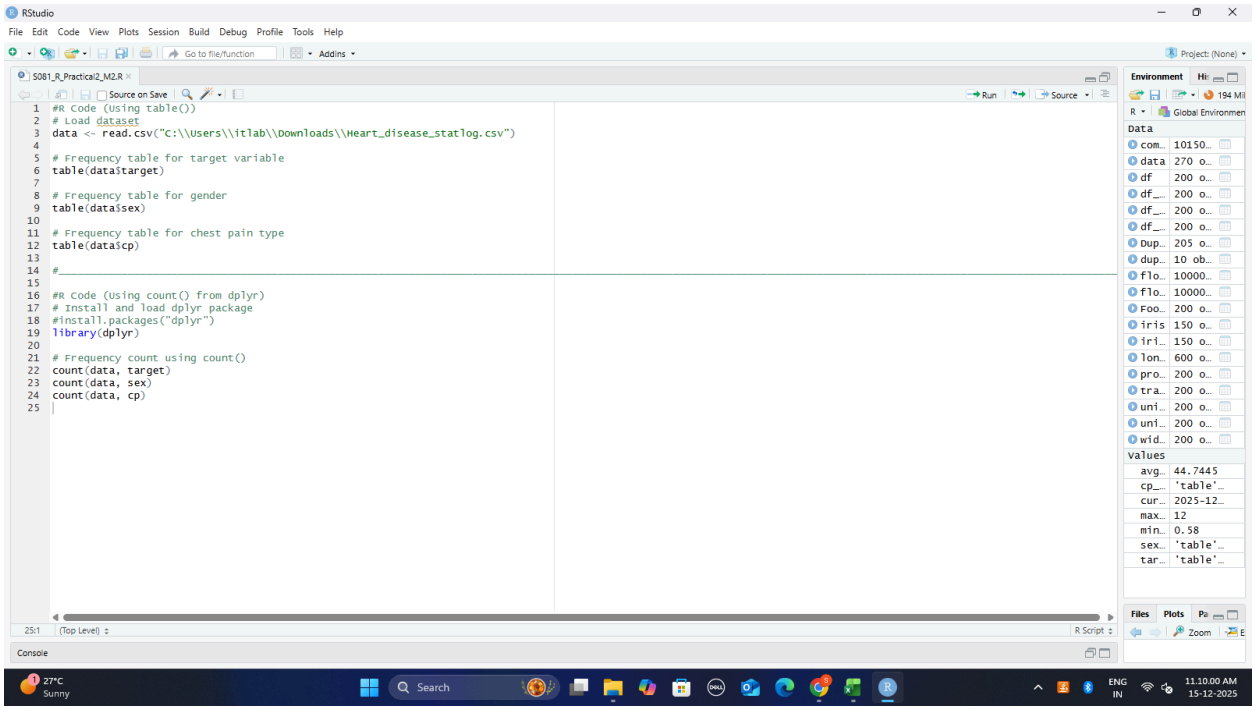
2. Generating frequency tables using table() or count() (R)

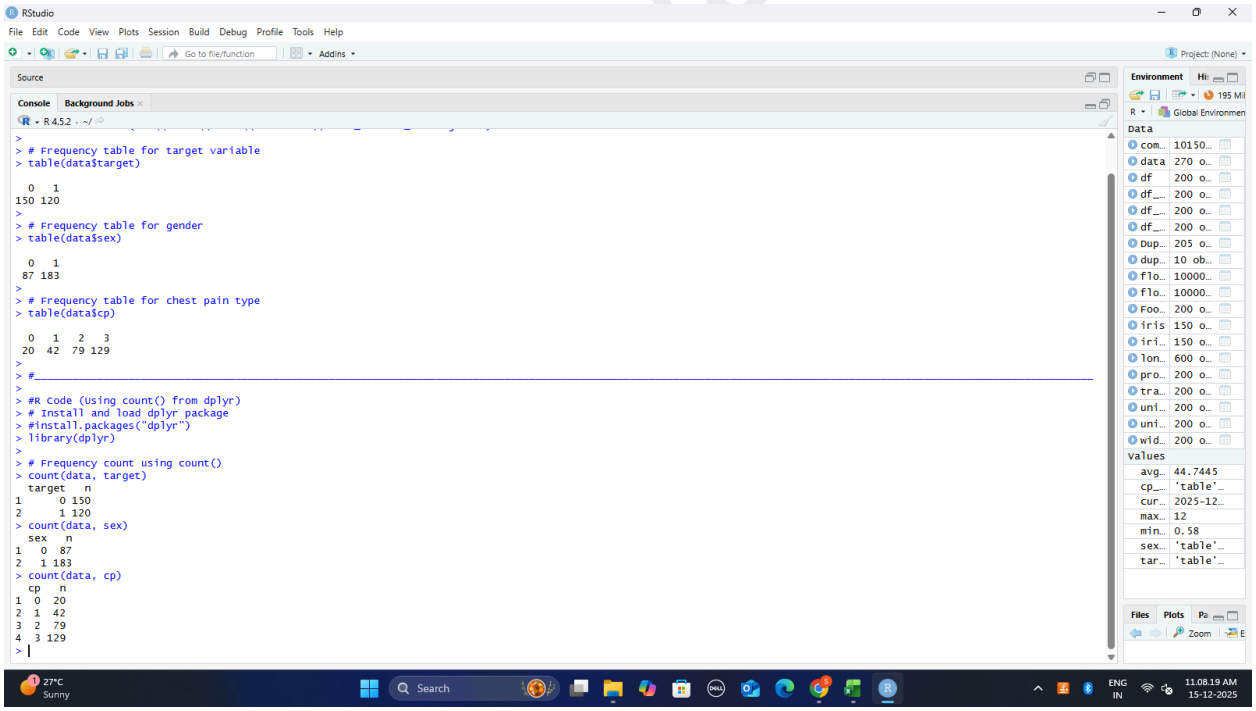


Name :- Priya Gupta

Roll No :- S081

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



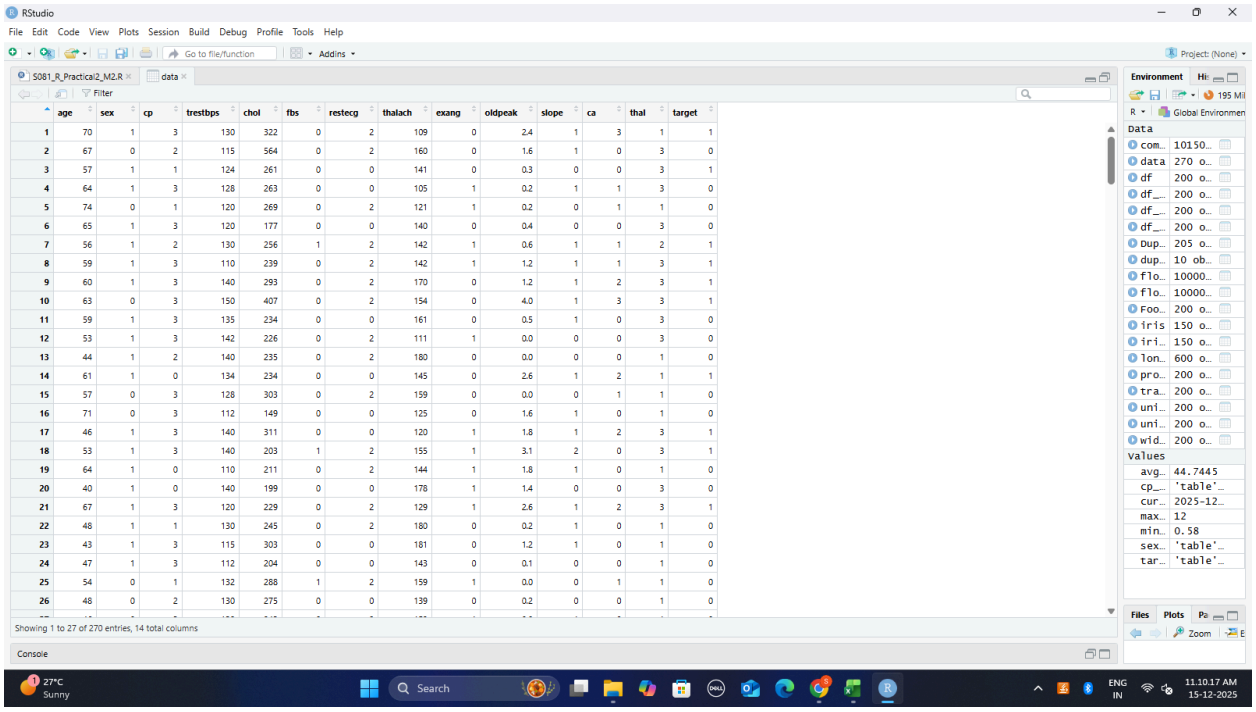


Name :- Priya Gupta

Roll No :- S081

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

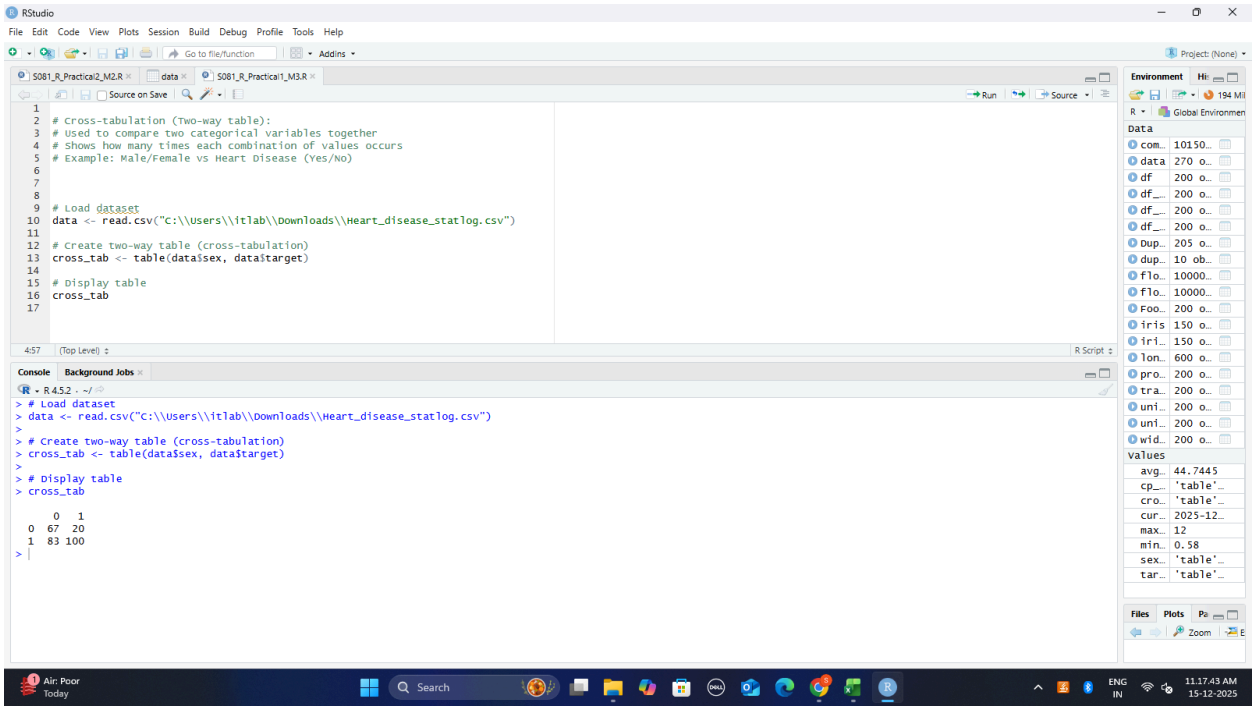
3. Creating cross-tabulations and two-way tables using table() (R).



Name :- Priya Gupta

Roll No :- S081

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

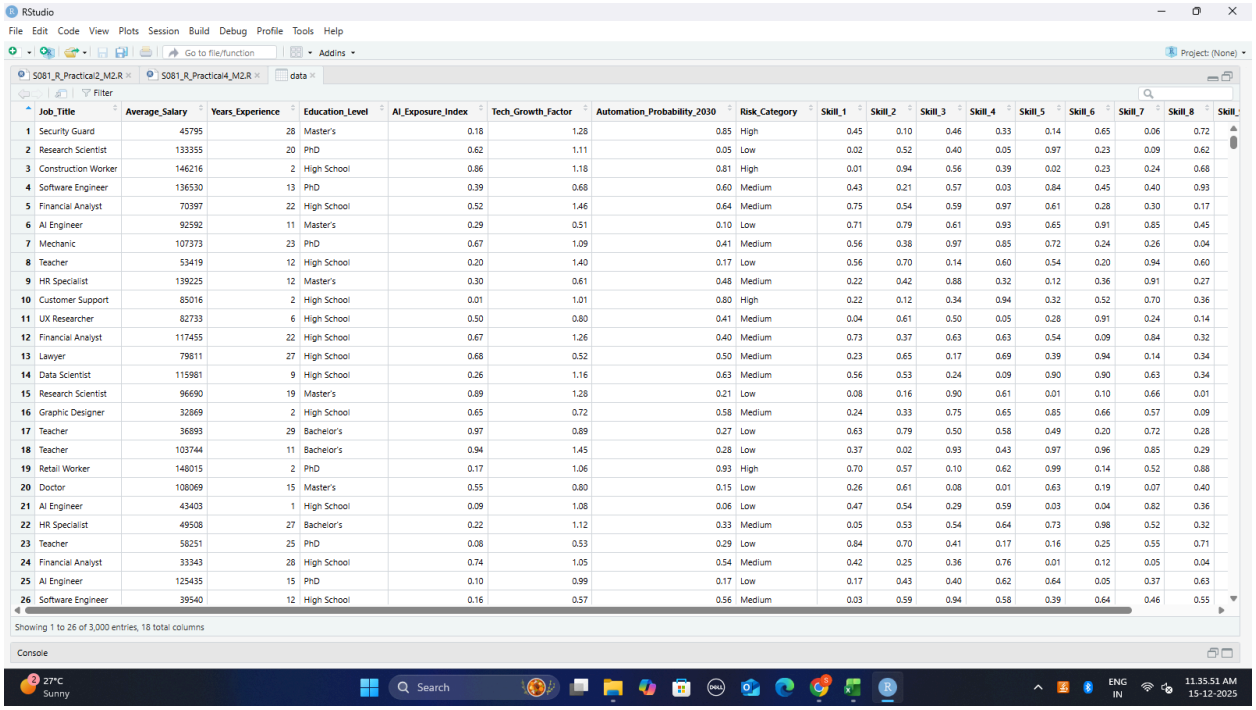


Name :- Priya Gupta

Roll No :- S081

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

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



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

S081_R_Practical2_M2.R | S081_R_Practical4_M2.R | data

Filter

	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
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	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

Showing 1 to 26 of 3,000 entries, 18 total columns

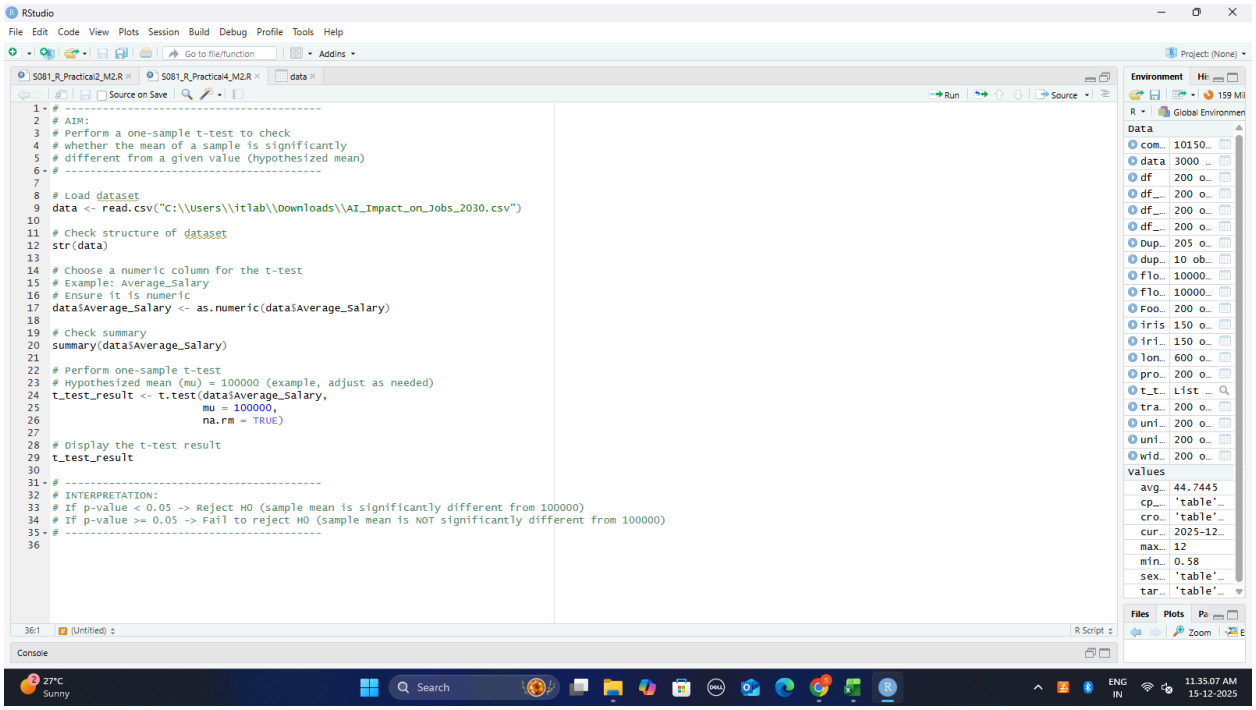
Console

27°C Sunny | Search | ENG IN | 11:35:51 AM 15-12-2025

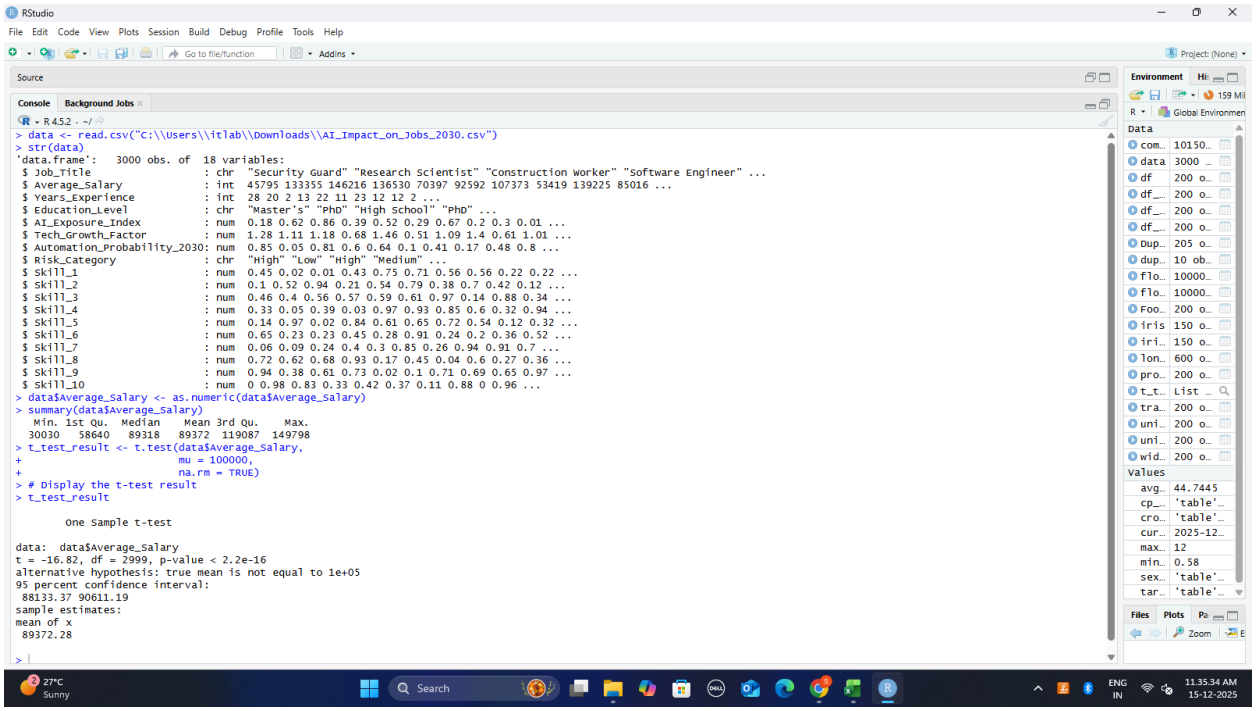
Name :- Priya Gupta

Roll No :- S081

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



```
# ----------------------------------------
# AIM:
# Perform a one-sample t-test to check
# whether the mean of a sample is significantly
# different from a given value (hypothesized mean)
# ----------------------------------------

# Load dataset
data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")

# Check structure of dataset
str(data)

# Choose a numeric column for the t-test
# Example: Average_Salary
# Ensure it is numeric
data$Average_Salary <- as.numeric(data$Average_Salary)

# Check summary
summary(data$Average_Salary)

# Perform one-sample t-test
# Hypothesized mean (mu) = 100000 (example, adjust as needed)
t_test_result <- t.test(data$Average_Salary,
                        mu = 100000,
                        na.rm = TRUE)

# Display the t-test result
t_test_result

# ----------------------------------------
# INTERPRETATION:
# If p-value < 0.05 -> Reject H0 (Sample mean is significantly different from 100000)
# If p-value >= 0.05 -> Fail to reject H0 (Sample mean is NOT significantly different from 100000)
# ----------------------------------------
```



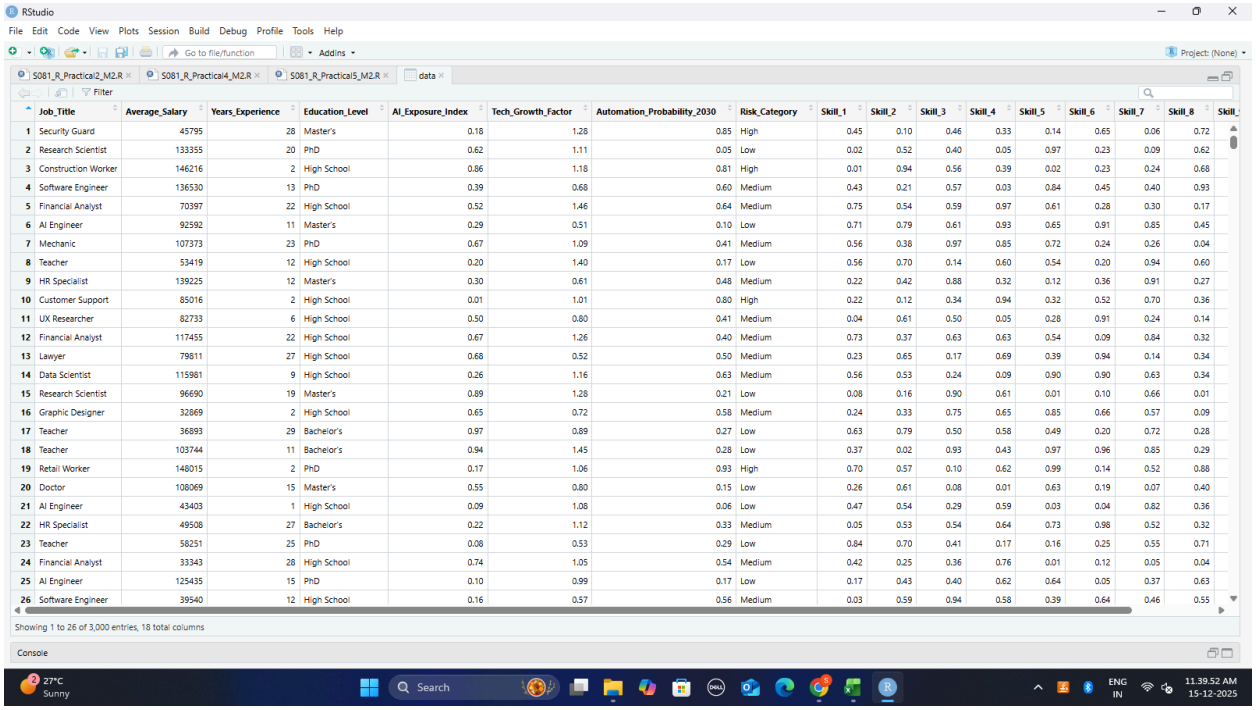
```
> data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
> str(data)
'data.frame':   3000 obs. of  18 variables:
 $ Job_Title                : chr  "Security Guard" "Research Scientist" "Construction Worker" "Software Engineer" ...
 $ Average_Salary           : int  45795 133355 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 ...
 $ 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 ...
> 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
```

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).



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

S081_R_Practical2_M2.R | S081_R_Practical4_M2.R | S081_R_Practical5_M2.R | data

Filter

	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
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	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

Showing 1 to 26 of 3,000 entries, 18 total columns

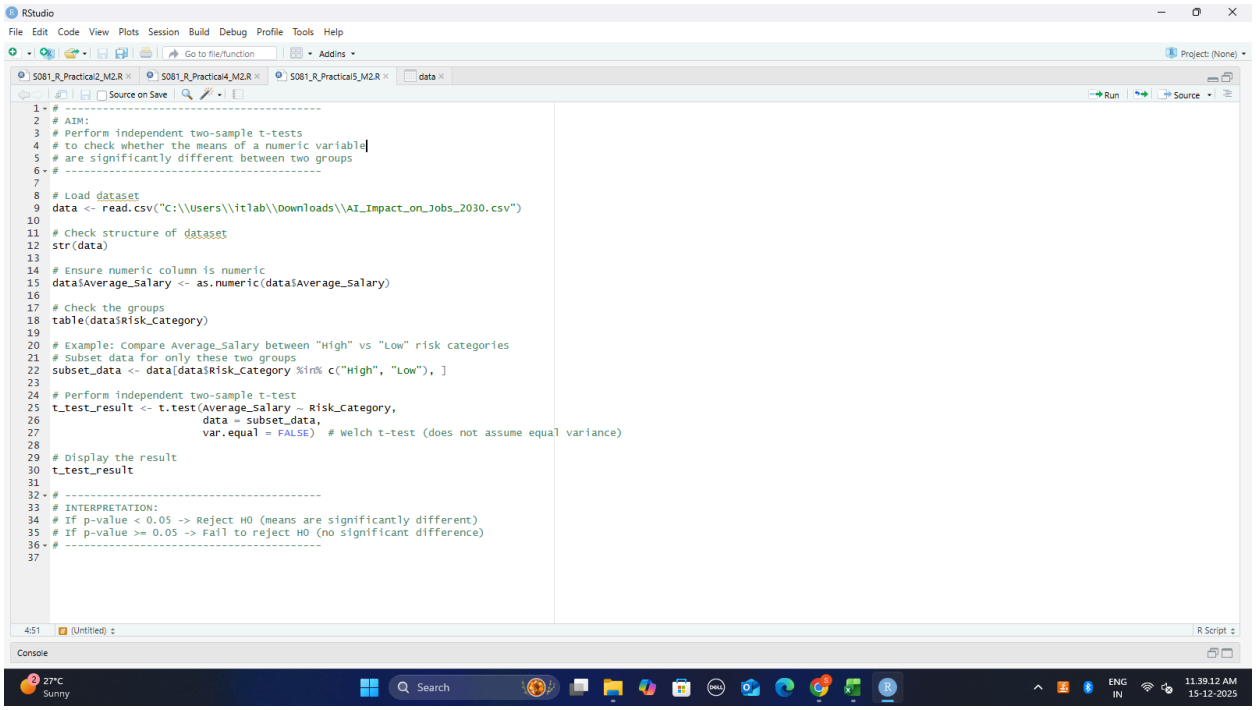
Console

27°C Sunny | Search | ENG IN | 11:39:52 AM 15-12-2025

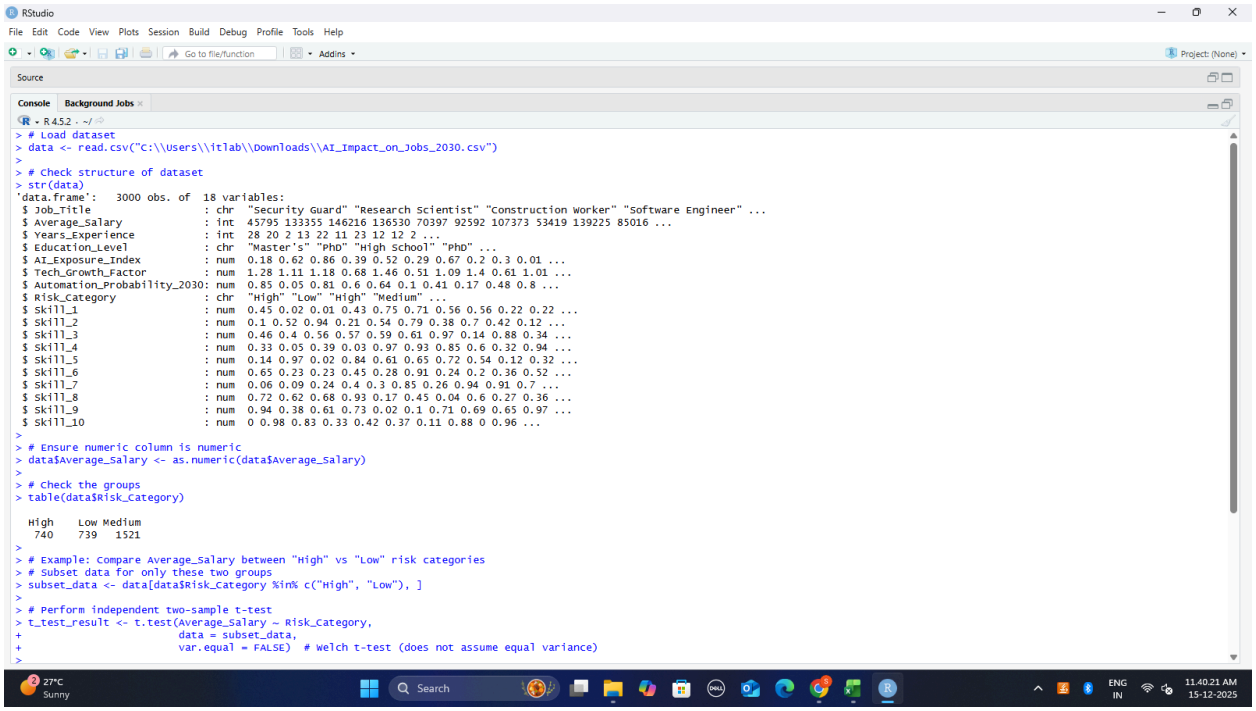
Name :- Priya Gupta

Roll No :- S081

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



```
# ----------------------------------------
# AIM:
# Perform independent two-sample t-tests
# to check whether the means of a numeric variable
# are significantly different between two groups
# ----------------------------------------

# Load dataset
data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")

# Check structure of dataset
str(data)

# Ensure numeric column is numeric
data$Average_Salary <- as.numeric(data$Average_Salary)

# Check the groups
table(data$Risk_Category)

# 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

# ----------------------------------------
# INTERPRETATION:
# If p-value < 0.05 -> Reject H0 (means are significantly different)
# If p-value >= 0.05 -> Fail to reject H0 (no significant difference)
# ----------------------------------------
```

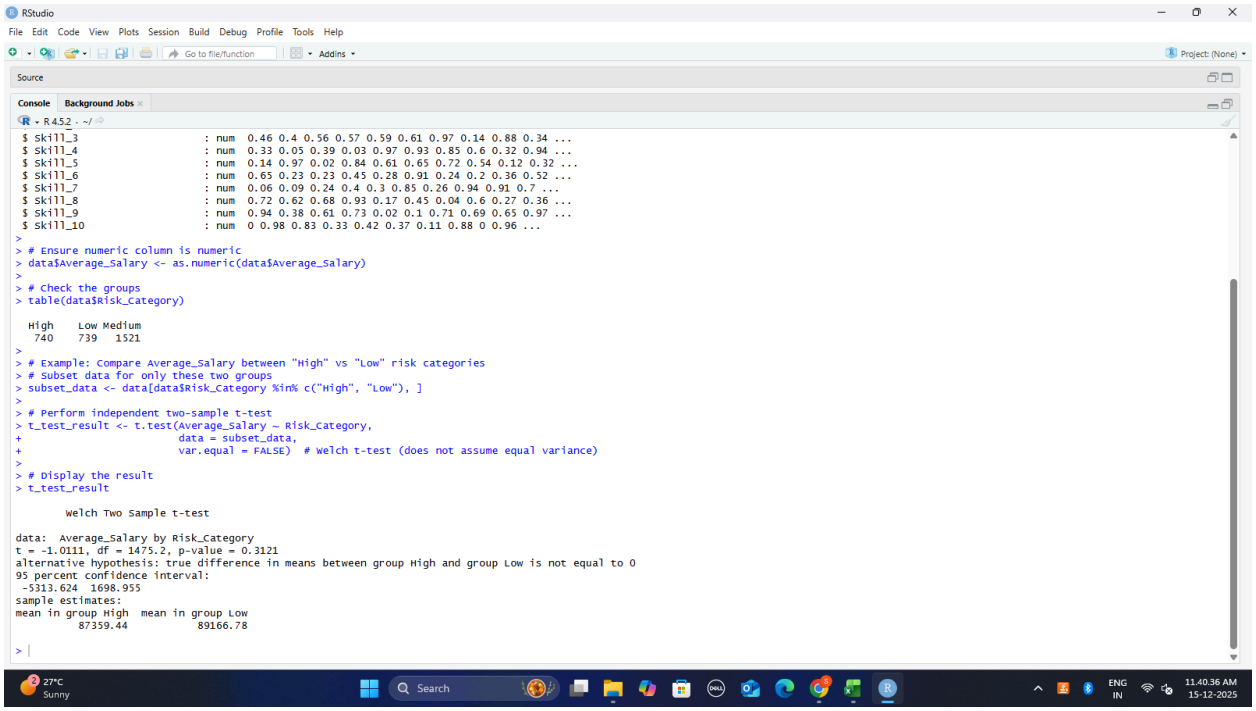


```
> # Load dataset
> data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
>
> # Check structure of dataset
> str(data)
'data.frame':   3000 obs. of  18 variables:
 $ Job_Title                  : chr  "Security Guard" "Research Scientist" "Construction Worker" "Software Engineer" ...
 $ 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



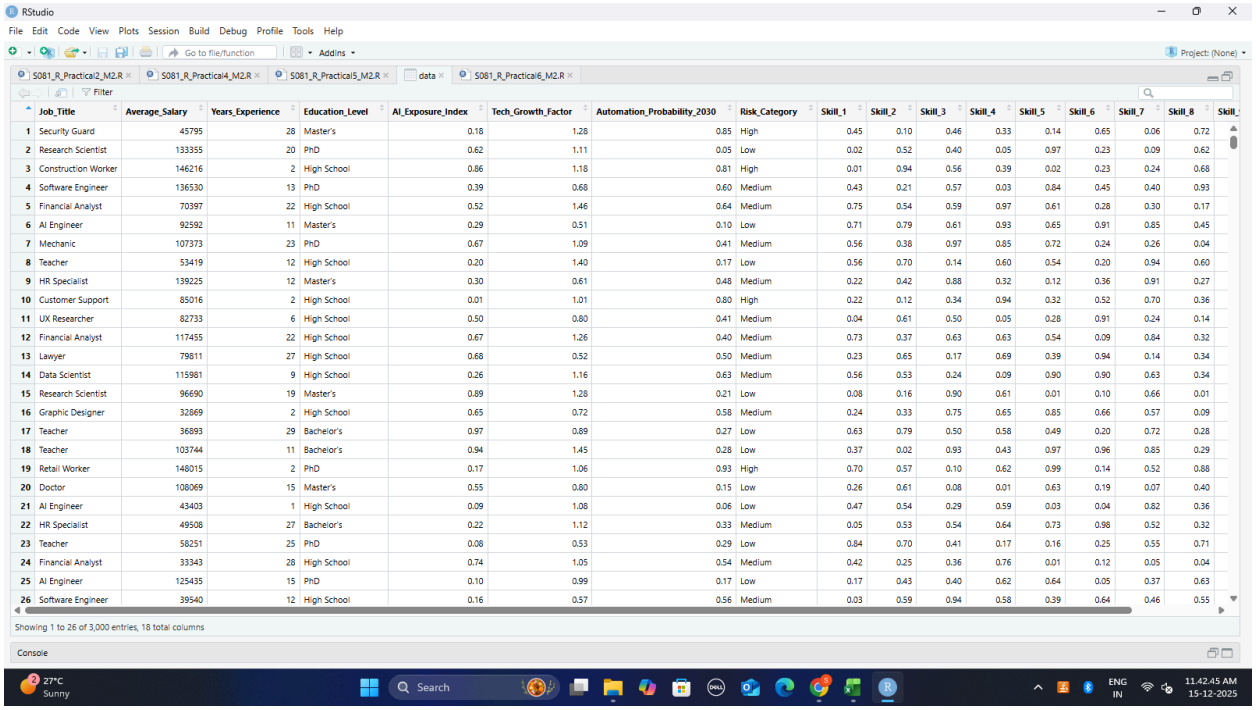
```
 $ Skill_3                  : num  0.46 0.4 0.56 0.57 0.59 0.61 0.97 0.14 0.88 0.34 ...
 $ Skill_4                  : num  0.33 0.05 0.39 0.03 0.97 0.93 0.85 0.6 0.32 0.94 ...
 $ Skill_5                  : num  0.14 0.97 0.02 0.84 0.61 0.65 0.72 0.54 0.12 0.32 ...
 $ Skill_6                  : num  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)
>
> # 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).



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

S081_R_Practical2_M2.R | S081_R_Practical4_M2.R | S081_R_Practical5_M2.R | data | S081_R_Practical6_M2.R

Filter

	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
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	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

Showing 1 to 26 of 3,000 entries, 18 total columns

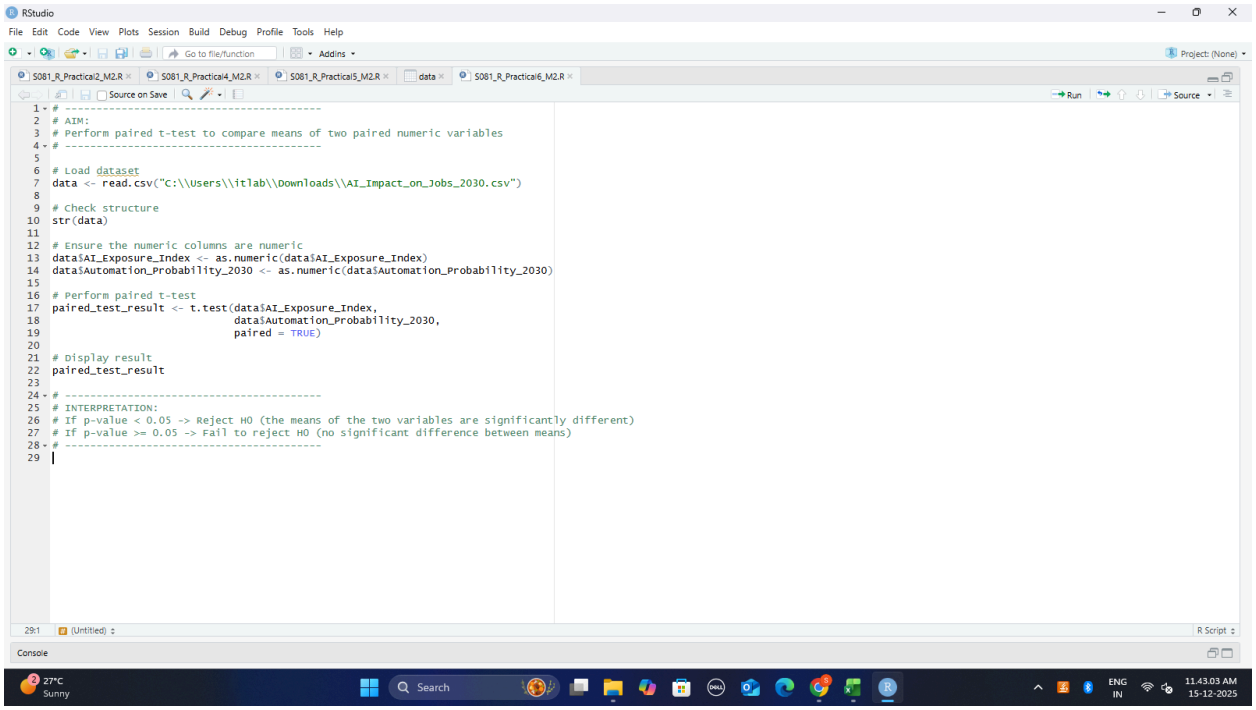
Console

27°C Sunny | Search | ENG IN | 11:42:45 AM 15-12-2025

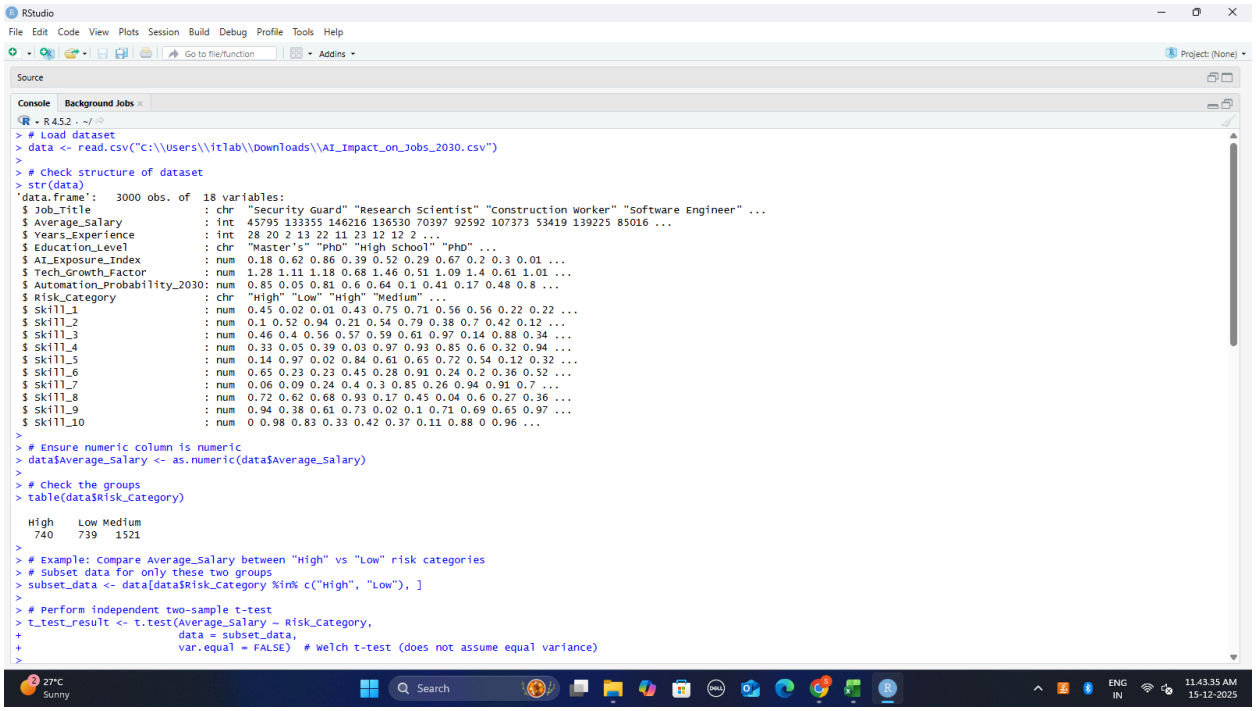
Name :- Priya Gupta

Roll No :- S081

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



```
# ----------------------------------------
# AIM:
# Perform paired t-test to compare means of two paired numeric variables
# ----------------------------------------

# Load dataset
data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")

# Check structure
str(data)

# 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

# ----------------------------------------
# INTERPRETATION:
# If p-value < 0.05 -> Reject H0 (the means of the two variables are significantly different)
# If p-value >= 0.05 -> Fail to reject H0 (no significant difference between means)
# ----------------------------------------
```

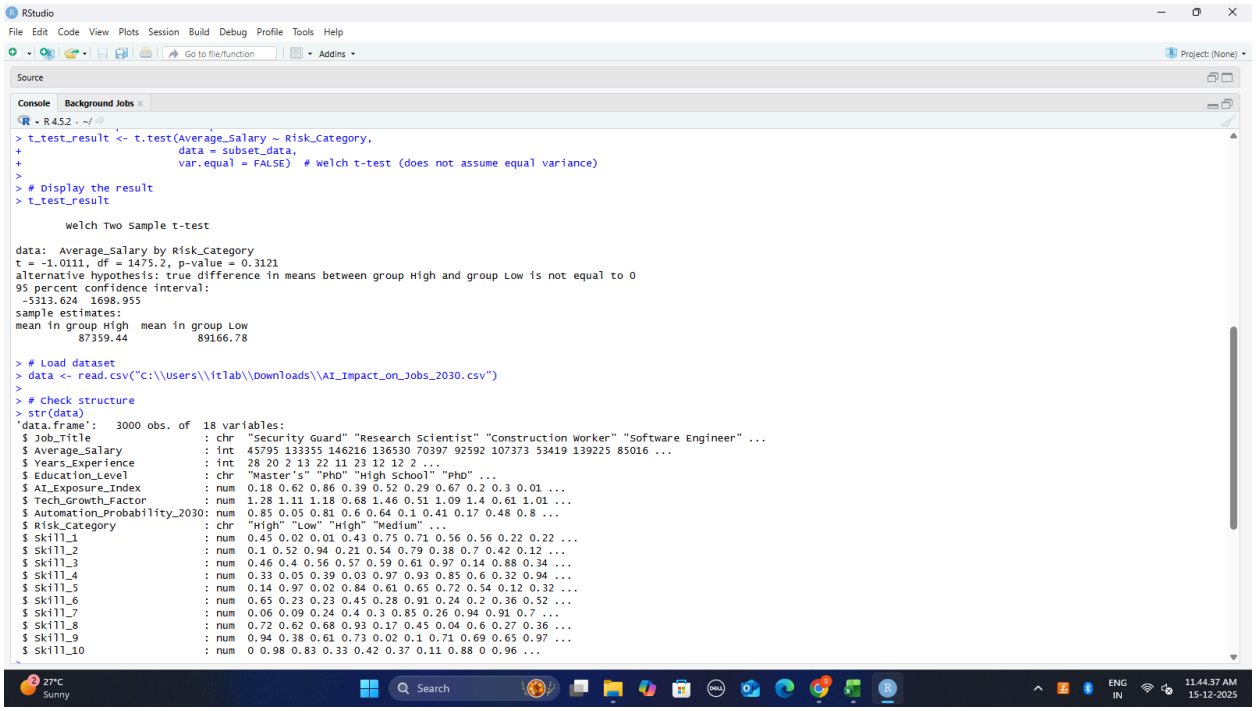


```
> # Load dataset
> data <- read.csv("C:\\Users\\itlab\\Downloads\\AI_Impact_on_Jobs_2030.csv")
>
> # Check structure of dataset
> str(data)
'data.frame':   3000 obs. of  18 variables:
 $ Job_Title                : chr  "Security Guard" "Research Scientist" "Construction Worker" "Software Engineer" ...
 $ 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

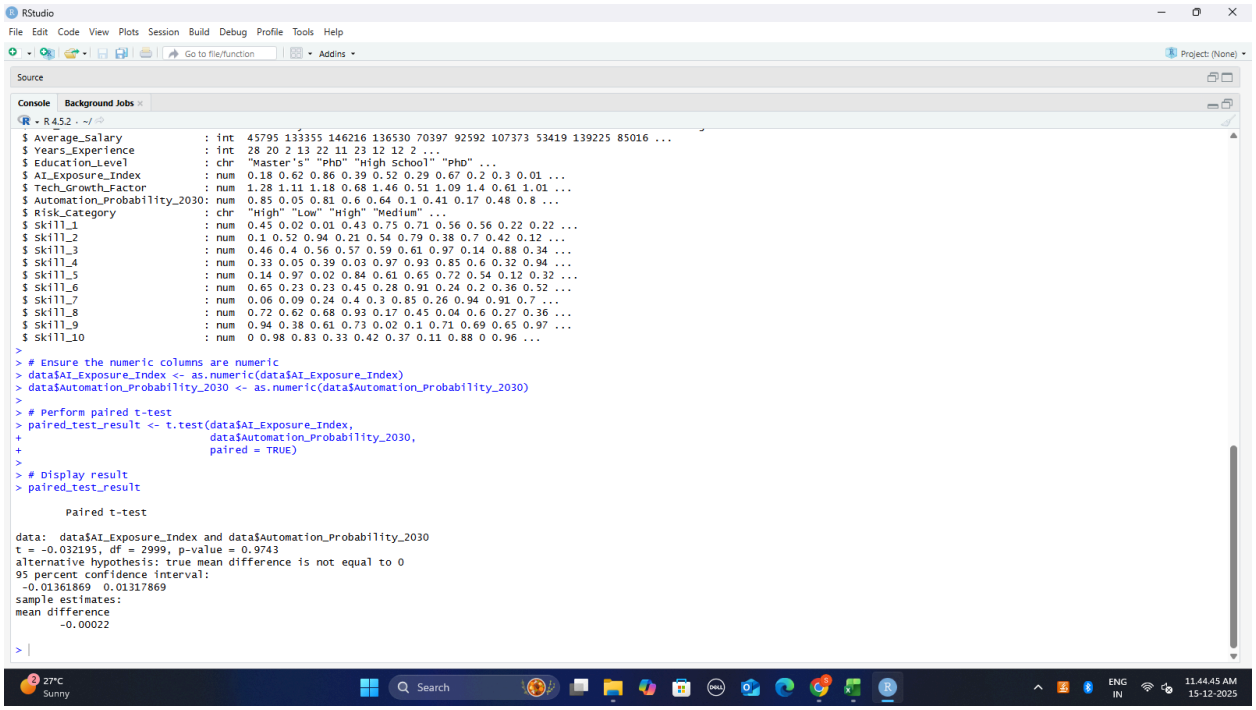
Sheth L.U.J. & Sir M.V. College



Name :- Priya Gupta

Roll No :- S081

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



Name :- Priya Gupta

Roll No :- S081