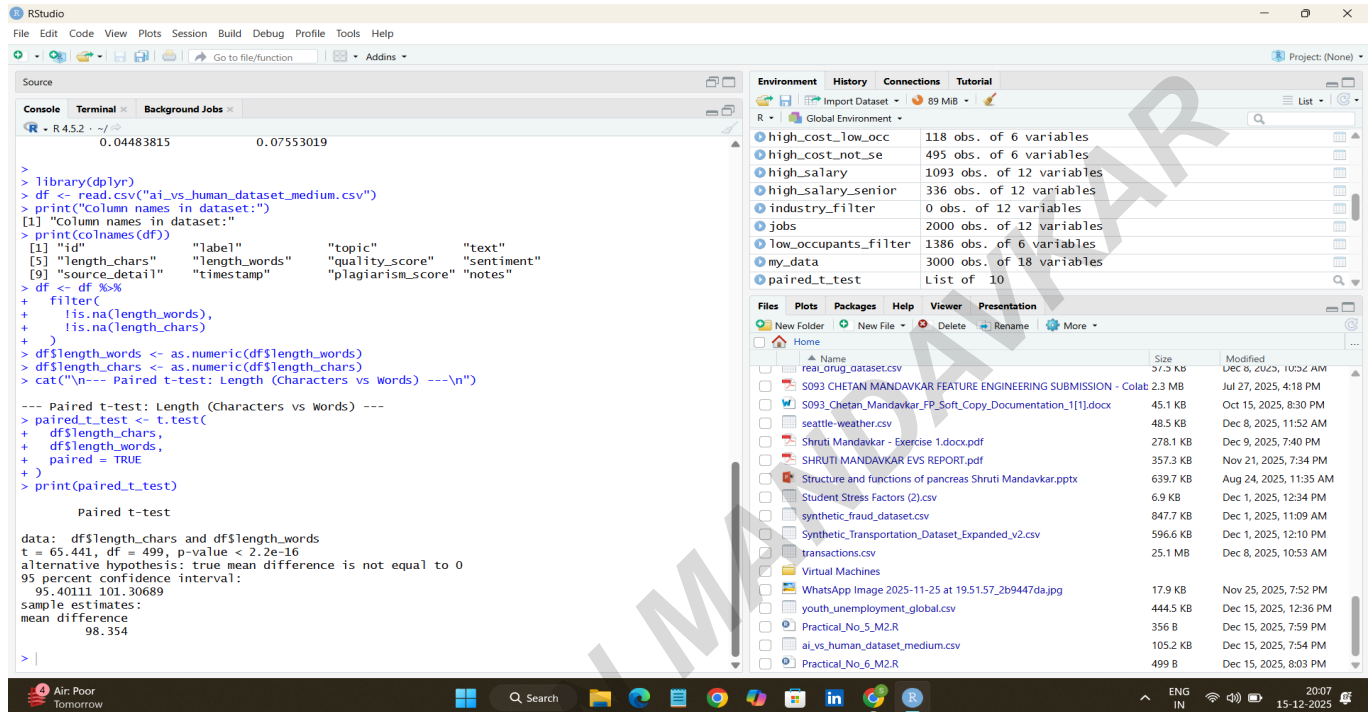


SHETH L.U.J. AND SIR M.V. COLLEGE

Aim:- Performing paired t-tests using `t.test(paired=TRUE)` (R).



```
> library(dplyr)
> df <- read.csv("ai_vs_human_dataset_medium.csv")
> print("Column names in dataset:")
[1] "Column names in dataset:"
> print(colnames(df))
[1] "id"          "label"        "topic"        "text"
[5] "length_chars" "length_words" "quality_score" "sentiment"
[9] "source_detail" "timestamp"    "plagiarism_score" "notes"
> df <- df %>%
+   filter(
+     !is.na(length_words),
+     !is.na(length_chars)
+   )
> df$length_words <- as.numeric(df$length_words)
> df$length_chars <- as.numeric(df$length_chars)
> cat("\n--- Paired t-test: Length (Characters vs Words) ---\n")

--- Paired t-test: Length (Characters vs Words) ---
> paired_t_test <- t.test(
+   df$length_chars,
+   df$length_words,
+   paired = TRUE
+ )
> print(paired_t_test)

Paired t-test

data:  df$length_chars and df$length_words
t = 65.441, df = 499, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 95.40111 101.30689
sample estimates:
mean difference
 98.354
```

The Environment pane shows the following datasets:

Dataset	Observations	Variables
high_cost_low_occ	118 obs.	6 variables
high_cost_not_se	495 obs.	6 variables
high_salary	1093 obs.	12 variables
high_salary_senior	336 obs.	12 variables
industry_filter	0 obs.	12 variables
jobs	2000 obs.	12 variables
low_occupants_filter	1386 obs.	6 variables
my_data	3000 obs.	18 variables
paired_t_test	List of 10	

The Files pane shows the following files:

Name	Size	Modified
real_time_dataset.csv	21.2 KB	Jul 27, 2025, 4:18 PM
S093 CHETAN MANDAVKAR FEATURE ENGINEERING SUBMISSION - Colab	2.3 MB	Oct 15, 2025, 8:30 PM
S093_Chetan_Mandavkar_FP_Soft_Copy_Documentation_1[1].docx	45.1 KB	Oct 15, 2025, 11:52 AM
seattle-weather.csv	48.5 KB	Dec 8, 2025, 7:40 PM
Shruti Mandavkar - Exercise 1.docx.pdf	278.1 KB	Dec 9, 2025, 7:34 PM
SHRUTI.MANDAVKAR EVS REPORT.pdf	357.3 KB	Nov 21, 2025, 7:34 PM
Structure and functions of pancreas Shruti Mandavkar.pptx	639.7 KB	Aug 24, 2025, 11:35 AM
Student Stress Factors (2).csv	6.9 KB	Dec 1, 2025, 12:34 PM
synthetic_fraud_dataset.csv	847.7 KB	Dec 1, 2025, 11:09 AM
Synthetic_Transportation_Dataset_Expanded_v2.csv	596.6 KB	Dec 1, 2025, 12:10 PM
transactions.csv	25.1 MB	Dec 8, 2025, 10:53 AM
Virtual Machines		
WhatsApp Image 2025-11-25 at 19.51.57_2b9447da.jpg	17.9 KB	Nov 25, 2025, 7:52 PM
youth_unemployment_global.csv	444.5 KB	Dec 15, 2025, 12:36 PM
Practical_No_5_M2.R	356 B	Dec 15, 2025, 7:59 PM
ai_vs_human_dataset_medium.csv	105.2 KB	Dec 15, 2025, 7:54 PM
Practical_No_6_M2.R	499 B	Dec 15, 2025, 8:03 PM

NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis With SAS / SPSS / R