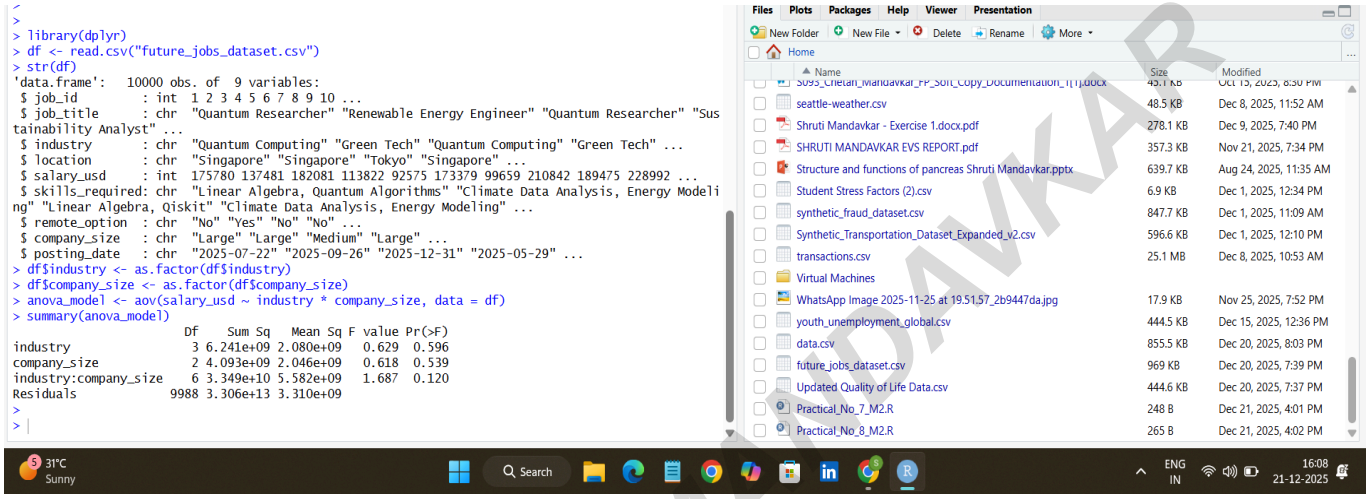
SHETH L.U.J. AND SIR M.V. COLLEGE

Aim:- Performing two-way ANOVA using aov() (R).



```
>
> library(dplyr)
> df <- read.csv("future_jobs_dataset.csv")
> str(df)
'data.frame':   10000 obs. of  9 variables:
 $ job_id         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ job_title      : chr  "Quantum Researcher" "Renewable Energy Engineer" "Quantum Researcher" "Sus
tainability Analyst" ...
 $ industry       : chr  "Quantum Computing" "Green Tech" "Quantum Computing" "Green Tech" ...
 $ location       : chr  "Singapore" "Singapore" "Tokyo" "Singapore" ...
 $ salary_usd     : int  175780 137481 182081 113822 92575 173379 99659 210842 189475 228992 ...
 $ skills_required: chr  "Linear Algebra, Quantum Algorithms" "Climate Data Analysis, Energy Modeli
ng" "Linear Algebra, Qiskit" "Climate Data Analysis, Energy Modeling" ...
 $ remote_option  : chr  "No" "Yes" "No" "No" ...
 $ company_size   : chr  "Large" "Large" "Medium" "Large" ...
 $ posting_date   : chr  "2025-07-22" "2025-09-26" "2025-12-31" "2025-05-29" ...
> df$industry <- as.factor(df$industry)
> df$company_size <- as.factor(df$company_size)
> anova_model <- aov(salary_usd ~ industry * company_size, data = df)
> summary(anova_model)
                     Df    Sum Sq   Mean Sq F value Pr(>F)
industry              3 6.241e+09 2.080e+09   0.629  0.596
company_size          2 4.093e+09 2.046e+09   0.618  0.539
industry:company_size    6 3.349e+10 5.582e+09   1.687  0.120
Residuals          9988 3.306e+13 3.310e+09
>
>
```


NAME:- CHETAN MANDAVKAR

ROLL NO. S093

SUBJECT:- Data Analysis With SAS / SPSS / R