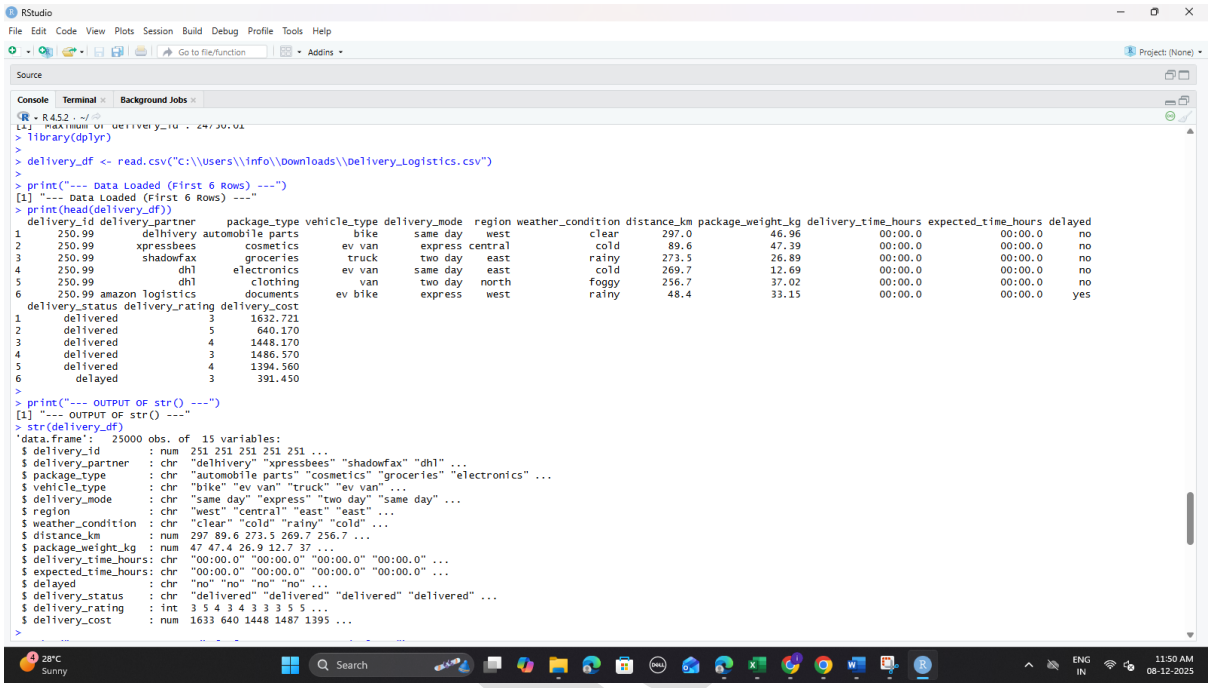
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

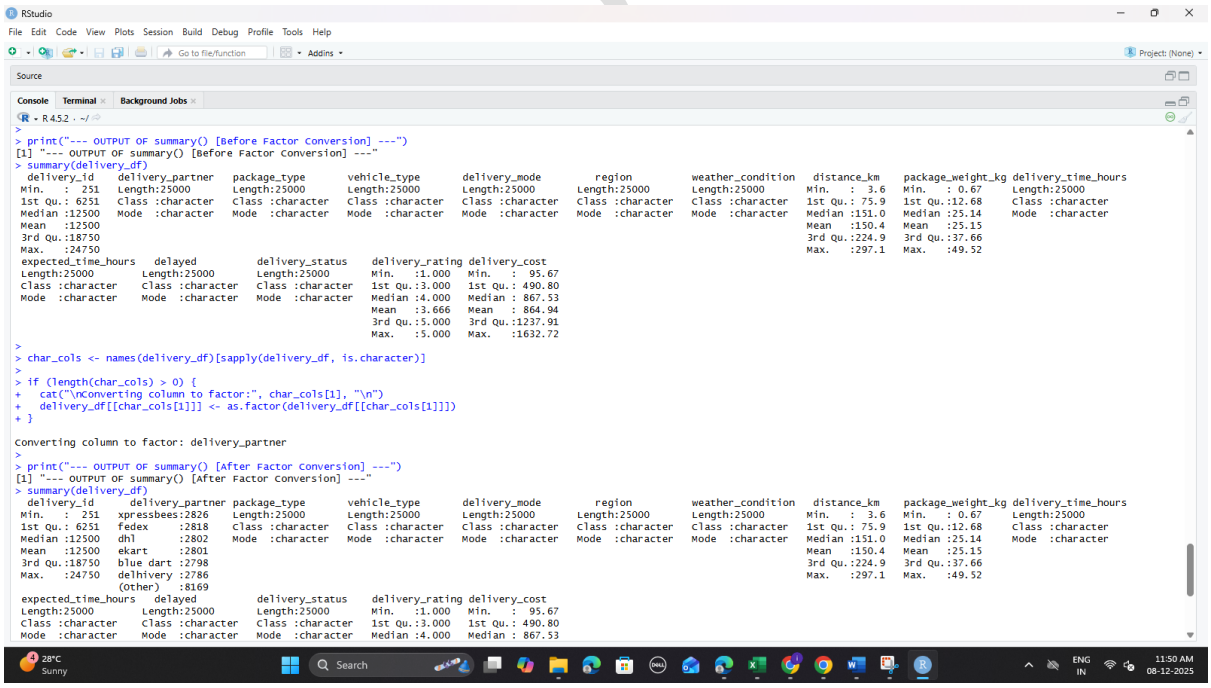
PRACTICAL – 15

AIM:- Generating basic summaries using str() or summary() (R).

OUTPUT:-



```
> library(dplyr)
>
> delivery_df <- read.csv("C:\\Users\\info\\Downloads\\Delivery_Logistics.csv")
>
> print("--- Data Loaded (First 6 Rows) ---")
[1] "--- Data Loaded (First 6 Rows) ---"
> print(head(delivery_df))
  delivery_id delivery_partner     package_type vehicle_type delivery_mode  region weather_condition distance_km package_weight_kg delivery_time_hours expected_time_hours delayed
1       250.99       delhivery automobile parts         bike      same day    west             clear       297.0             46.96             00:00.0             00:00.0      no
2       250.99      xpressbees        cosmetics       ev van       express central              cold        89.6             47.39             00:00.0             00:00.0      no
3       250.99       shadowfax        groceries        truck      two day     east             rainy       273.5             26.89             00:00.0             00:00.0      no
4       250.99             dhl      electronics       ev van      same day     east              cold       269.7             12.69             00:00.0             00:00.0      no
5       250.99             dhl         clothing          van      two day    north             foggy       256.7             37.02             00:00.0             00:00.0      no
6       250.99 amazon logistics        documents       ev bike      express     west             rainy        48.4             33.15             00:00.0             00:00.0     yes
  delivery_status delivery_rating delivery_cost
1       delivered               3       1632.721
2       delivered               5        640.170
3       delivered               4       1448.170
4       delivered               3       1486.570
5       delivered               4       1394.560
6         delayed               3        391.450
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(delivery_df)
'data.frame':   25000 obs. of  15 variables:
 $ delivery_id       : num  251 251 251 251 251 ...
 $ delivery_partner  : chr  "delhivery" "xpressbees" "shadowfax" "dhl" ...
 $ package_type      : chr  "automobile parts" "cosmetics" "groceries" "electronics" ...
 $ vehicle_type      : chr  "bike" "ev van" "truck" "ev van" ...
 $ delivery_mode     : chr  "same day" "express" "two day" "same day" ...
 $ region            : chr  "west" "central" "east" "east" ...
 $ weather_condition : chr  "clear" "cold" "rainy" "cold" ...
 $ distance_km       : num  297 89.6 273.5 269.7 256.7 ...
 $ package_weight_kg : num  47 47.4 26.9 12.7 37 ...
 $ delivery_time_hours: chr  "00:00.0" "00:00.0" "00:00.0" "00:00.0" ...
 $ expected_time_hours: chr  "00:00.0" "00:00.0" "00:00.0" "00:00.0" ...
 $ delayed           : chr  "no" "no" "no" "no" ...
 $ delivery_status   : chr  "delivered" "delivered" "delivered" "delivered" ...
 $ delivery_rating   : int  3 5 4 3 4 3 3 3 5 5 ...
 $ delivery_cost     : num  1633 640 1448 1487 1395 ...
>
```



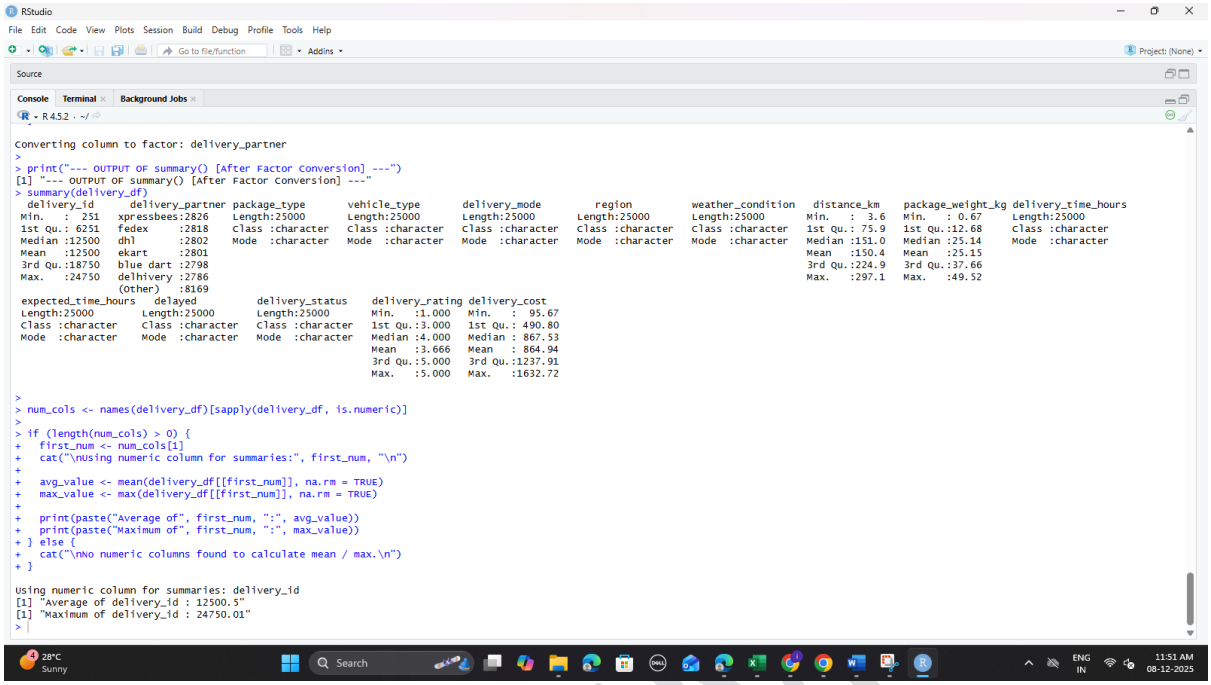
```
>
> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(delivery_df)
  delivery_id    delivery_partner   package_type       vehicle_type       delivery_mode        region          weather_condition   distance_km    package_weight_kg delivery_time_hours
 Min.   :  251   Length:25000       Length:25000       Length:25000       Length:25000       Length:25000       Length:25000       Min.   :  3.6   Min.   : 0.67     Length:25000
 1st Qu.: 6251   Class :character   Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.: 75.9   1st Qu.:12.68     Class :character
 Median :12500   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :151.0   Median :25.14     Mode  :character
 Mean   :12500                                                                                                                    Mean   :150.4   Mean   :25.15
 3rd Qu.:18750                                                                                                                    3rd Qu.:224.9   3rd Qu.:37.66
 Max.   :24750                                                                                                                    Max.   :297.1   Max.   :49.52
 expected_time_hours   delayed          delivery_status    delivery_rating delivery_cost
 Length:25000        Length:25000       Length:25000       Min.   :1.000   Min.   :  95.67
 Class :character    Class :character   Class :character   1st Qu.:3.000   1st Qu.: 490.80
 Mode  :character    Mode  :character   Mode  :character   Median :4.000   Median : 867.53
                                                           Mean   :3.666   Mean   : 864.94
                                                           3rd Qu.:5.000   3rd Qu.:1237.91
                                                           Max.   :5.000   Max.   :1632.72
>
> char_cols <- names(delivery_df)[sapply(delivery_df, is.character)]
>
> if (length(char_cols) > 0) {
+   cat("\nConverting column to factor:", char_cols[1], "\n")
+   delivery_df[[char_cols[1]]] <- as.factor(delivery_df[[char_cols[1]]])
+ }

Converting column to factor: delivery_partner
>
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(delivery_df)
  delivery_id    delivery_partner package_type       vehicle_type       delivery_mode        region          weather_condition   distance_km    package_weight_kg delivery_time_hours
 Min.   :  251   xpressbees:2826  Length:25000       Length:25000       Length:25000       Length:25000       Length:25000       Min.   :  3.6   Min.   : 0.67     Length:25000
 1st Qu.: 6251   fedex     :2818  Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.: 75.9   1st Qu.:12.68     Class :character
 Median :12500   dhl       :2802  Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :151.0   Median :25.14     Mode  :character
 Mean   :12500   ekart     :2801                                                                                                  Mean   :150.4   Mean   :25.15
 3rd Qu.:18750   blue dart :2798                                                                                                  3rd Qu.:224.9   3rd Qu.:37.66
 Max.   :24750   delhivery :2786                                                                                                  Max.   :297.1   Max.   :49.52
                 (Other)   :8169
 expected_time_hours   delayed          delivery_status    delivery_rating delivery_cost
 Length:25000        Length:25000       Length:25000       Min.   :1.000   Min.   :  95.67
 Class :character    Class :character   Class :character   1st Qu.:3.000   1st Qu.: 490.80
 Mode  :character    Mode  :character   Mode  :character   Median :4.000   Median : 867.53
```

SUMEET JITENDRA YADAV
S124

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

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R



```
RStudio
File  Edit  Code  View  Plots  Session  Build  Debug  Profile  Tools  Help
Source
Console  Terminal  Background Jobs
R 4.5.2 · ~/

Converting column to factor: delivery_partner
>
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(delivery_df)
  delivery_id      delivery_partner package_type       vehicle_type       delivery_mode        region          weather_condition   distance_km    package_weight_kg delivery_time_hours
 Min.    :  251   xpressbees:2826    Length:25000       Length:25000       Length:25000       Length:25000       Length:25000       Min.    :  3.6   Min.    : 0.67    Length:25000
 1st Qu.: 6251   fedex     :2818    Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.: 75.9   1st Qu.:12.68    Class :character
 Median :12500   dhl       :2802    Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :151.0   Median :25.14    Mode  :character
 Mean   :12500   ekart     :2801                                                                                                  Mean   :150.4   Mean   :25.15
 3rd Qu.:18750   blue dart :2798                                                                                                  3rd Qu.:224.9   3rd Qu.:37.66
 Max.   :24750   delhivery :2786                                                                                                  Max.   :297.1   Max.   :49.52
                 (Other)   :8169
 expected_time_hours   delayed          delivery_status    delivery_rating delivery_cost
 Length:25000        Length:25000       Length:25000       Min.   :1.000   Min.   :  95.67
 Class :character    Class :character   Class :character   1st Qu.:3.000   1st Qu.: 490.80
 Mode  :character    Mode  :character   Mode  :character   Median :4.000   Median : 867.53
                                                           Mean   :3.666   Mean   : 864.94
                                                           3rd Qu.:5.000   3rd Qu.:1237.91
                                                           Max.   :5.000   Max.   :1632.72

>
> num_cols <- names(delivery_df)[sapply(delivery_df, is.numeric)]
>
> if (length(num_cols) > 0) {
+   first_num <- num_cols[1]
+   cat("\nUsing numeric column for summaries:", first_num, "\n")
+
+   avg_value <- mean(delivery_df[[first_num]], na.rm = TRUE)
+   max_value <- max(delivery_df[[first_num]], na.rm = TRUE)
+
+   print(paste("Average of", first_num, ":", avg_value))
+   print(paste("Maximum of", first_num, ":", max_value))
+ } else {
+   cat("\nNo numeric columns found to calculate mean / max.\n")
+ }

Using numeric column for summaries: delivery_id
[1] "Average of delivery_id : 12500.5"
[1] "Maximum of delivery_id : 24750.01"
>
```


SUMEET JITENDRA YADAV

S124