

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

15. Generating basic summaries using str() or summary() (R).

The screenshot shows the RStudio interface with a dataset named 'Food_Delivery_Route_Efficiency_Dataset' loaded. The dataset is displayed in a table with 27 rows and 10 columns. The columns are: order_id, distance_km, delivery_time_min, traffic_level, route_length_km, delivery_mode, weather, order_time, restaurant_zone, and customer_zone. The data shows various delivery orders with details on distance, time, traffic, route length, mode of transport, weather, and location zones.

order_id	distance_km	delivery_time_min	traffic_level	route_length_km	delivery_mode	weather	order_time	restaurant_zone	customer_zone
1	7.97	63.8	High	9.75	Bicycle	Clear	2025-01-01 15:29	South	North
2	0.90	7.6	High	1.28	Car	Cloudy	2025-01-03 00:47	West	North
3	11.12	78.0	Medium	16.65	Bike	Rainy	2025-01-04 17:32	South	Central
4	4.90	24.8	Low	5.25	Scooter	Rainy	2025-01-01 14:12	Central	Central
5	10.04	56.0	High	11.34	Car	Rainy	2025-01-02 16:50	West	North
6	10.96	76.8	High	13.62	Car	Windy	2025-01-02 09:56	West	North
7	9.60	54.4	Low	13.32	Bike	Windy	2025-01-03 00:29	Central	North
8	6.24	52.9	Low	8.41	Bike	Cloudy	2025-01-02 21:36	South	Central
9	9.98	81.4	Medium	14.87	Bicycle	Rainy	2025-01-01 08:19	Central	West
10	2.48	17.1	High	3.38	Car	Windy	2025-01-01 09:28	South	North
11	7.23	47.7	Medium	7.63	Bike	Rainy	2025-01-03 15:24	West	Central
12	6.65	27.8	High	6.69	Bike	Rainy	2025-01-03 06:47	Central	West
13	11.42	83.4	Medium	13.97	Car	Rainy	2025-01-03 05:24	Central	South
14	1.26	9.9	Low	1.55	Bicycle	Windy	2025-01-03 16:37	South	North
15	9.88	56.6	Medium	12.45	Bike	Cloudy	2025-01-03 04:12	North	Central
16	4.69	23.2	Low	4.86	Scooter	Clear	2025-01-04 04:16	North	North
17	7.27	34.9	High	8.66	Scooter	Cloudy	2025-01-03 14:38	Central	West
18	10.19	60.2	High	10.24	Car	Windy	2025-01-03 02:11	North	South
19	2.11	9.2	Medium	2.23	Scooter	Rainy	2025-01-02 17:10	Central	North
20	3.12	18.0	High	3.41	Bicycle	Windy	2025-01-04 06:23	South	Central
21	8.22	50.4	Low	10.42	Bike	Windy	2025-01-01 08:26	East	South
22	8.13	46.0	High	9.43	Car	Windy	2025-01-03 11:13	Central	Central
23	0.54	2.4	Medium	0.58	Scooter	Rainy	2025-01-02 09:43	South	East
24	5.15	33.7	Medium	7.30	Bicycle	Windy	2025-01-04 16:58	South	West
25	0.53	2.9	Low	0.62	Bike	Cloudy	2025-01-04 04:12	West	Central
26	4.23	26.1	Medium	4.97	Car	Windy	2025-01-02 15:05	West	Central

The screenshot shows the RStudio interface with R code being executed. The code loads the dataset 'Food_Delivery_Route_Efficiency_Dataset' and generates basic summaries using str() and summary().

```
1 #
2 # 15. Basic Summaries using str() and summary()
3 # Dataset: Food Delivery Route Efficiency
4 #
5
6 library(dplyr)
7
8 #
9 # 1. LOAD YOUR DATA
10 #
11
12 df <- read.csv(
13   "c:\\users\\itlab\\onedrive\\desktop\\s081_r_studio\\Food_Delivery_Route_Efficiency_Dataset.csv",
14   na.strings = c("", "NA")
15 )
16
17 print("--- Data Loaded Successfully ---")
18
19 #
20 # 2. USING str() - Structure of Dataset
21 #
22
23 print("--- OUTPUT OF str() ---")
24 str(df)
25
26 #
27 # 3. USING summary() - Statistical Summary
28 #
29
30 print("--- OUTPUT OF summary() [original] ---")
31 summary(df)
32
33 #
34 # 4. IMPROVING SUMMARY USING FACTORS
35 # Convert character columns to factors for better summary
36 #
37
38 df$traffic_level <- as.factor(df$traffic_level)
39 df$delivery_mode <- as.factor(df$delivery_mode)
40 df$weather <- as.factor(df$weather)
41 df$restaurant_zone <- as.factor(df$restaurant_zone)
42 df$customer_zone <- as.factor(df$customer_zone)
43
44 # (Untitled) 2
```

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 'Food_Delivery_Route_Efficiency_Dataset.R'. The script contains the following code:

```
1/ print("---- Data Loaded Successfully ----")
18
19 # -----
20 # 2. USING str() -> Structure of Datasets
21 # -----
22
23 print("---- OUTPUT OF str() ----")
24 str(df)
25
26 # -----
27 # 3. USING summary() -> Statistical Summary
28 # -----
29
30 print("---- OUTPUT OF summary() [original] ----")
31 summary(df)
32
33 # -----
34 # 4. IMPROVING SUMMARY USING FACTORS
35 # Convert character columns to factors for better summary
36 # -----
37
38 df$traffic_level <- as.factor(df$traffic_level)
39 df$delivery_mode <- as.factor(df$delivery_mode)
40 df$weather <- as.factor(df$weather)
41 df$restaurant_zone <- as.factor(df$restaurant_zone)
42 df$customer_zone <- as.factor(df$customer_zone)
43
44 print("---- OUTPUT OF summary() [After Factor Conversion] ----")
45 summary(df)
46
47 # -----
48 # 5. Specific summaries (Examples)
49 # -----
50
51 avg_delivery_time <- mean(df$delivery_time_min, na.rm = TRUE)
52 max_distance <- max(df$distance_km, na.rm = TRUE)
53 min_route_length <- min(df$route_length_km, na.rm = TRUE)
54
55 print(paste("Average Delivery Time (min):", avg_delivery_time))
56 print(paste("Maximum Delivery Distance (km):", max_distance))
57 print(paste("Minimum Route Length (km):", min_route_length))
58
```

The Environment pane on the right shows the following objects:

Object	Size
combined_d...	10150 obs. of 2 variables
df	200 obs. of 10 variables
df_clean	200 obs. of 11 variables
df_no_dupl...	200 obs. of 10 variables
df_small	200 obs. of 4 variables
duplicate...	205 obs. of 10 variables
duplicate...	10 obs. of 10 variables
flower_cle...	10000 obs. of 2 variables
flower_df	10000 obs. of 4 variables
Food_Deliv...	200 obs. of 10 variables
iris	150 obs. of 5 variables
iris_clean	150 obs. of 2 variables
long_df	600 obs. of 3 variables
processed...	200 obs. of 21 variables
traffic_pi...	200 obs. of 4 variables
unique_ord...	200 obs. of 10 variables
unique_rou...	200 obs. of 10 variables
wide_df	200 obs. of 4 variables

The Console shows the output of the script, including the data structure and summary statistics.

The screenshot shows the RStudio interface with the same script file. The Console pane shows the output of the script, including the data structure and summary statistics.

```
> # 15. Basic Summaries using str() and summary()
> # Dataset: Food Delivery Route Efficiency
>
> library(dplyr)
>
> # 1. LOAD YOUR DATA
> # -----
>
> df <- read.csv(
+ "c:\\Users\\ftlab\\OneDrive\\Desktop\\S081_R_Studio\\Food_Delivery_Route_Efficiency_Dataset.csv",
+ na.strings = c("", "NA")
+ )
>
> print("---- Data Loaded Successfully ----")
[1] "---- Data Loaded Successfully ----"
>
> # 2. USING str() -> Structure of Dataset
> # -----
>
> print("---- OUTPUT OF str() ----")
[1] "---- OUTPUT OF str() ----"
> str(df)
'data.frame': 200 obs. of 10 variables:
 $ order_id : int 1 2 3 4 5 6 7 8 9 10 ...
 $ distance_km : num 7.97 0.9 11.12 4.9 10.04 ...
 $ delivery_time_min: num 63.8 7.6 78.24 8.56 76.8 54.4 52.9 81.4 17.1 ...
 $ traffic_level : chr "High" "High" "Medium" "Low" ...
 $ route_length_km : num 9.75 1.28 16.65 5.25 11.34 ...
 $ delivery_mode : chr "Bicycle" "Car" "Bike" "Scooter" ...
 $ weather : chr "Clean" "Cloudy" "Rainy" "Rainy" ...
 $ order_time : chr "2025-01-01 15:29" "2025-01-03 00:47" "2025-01-04 17:32" "2025-01-01 14:12" ...
 $ restaurant_zone : chr "South" "West" "South" "Central" ...
 $ customer_zone : chr "North" "North" "Central" "Central" ...
>
> # -----
> # 3. USING summary() -> Statistical Summary
> # -----
>
> print("---- OUTPUT OF summary() [original] ----")
```

The Environment pane on the right shows the same objects as the first screenshot.

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
Source
Console Background Jobs
R - R 4.5.2 - ~/
> # =====
> # 3. USING summary() - Statistical summary
> # =====
> print("---- OUTPUT OF summary() [original] ----")
[1] "---- OUTPUT OF summary() [original] ----"
> summary(df)
  order_id      distance_km  delivery_time_min traffic_level
Min.   : 1.00   Min.   : 0.530   Min.   : 2.40   Length:200
1st Qu.: 50.75   1st Qu.: 3.688   1st Qu.: 23.73   Class :character
Median :100.50   Median : 6.845   Median : 44.65   Mode  :character
Mean   :100.50   Mean   : 6.625   Mean   : 44.74
3rd Qu.:150.25   3rd Qu.: 9.797   3rd Qu.: 63.90
Max.   :200.00   Max.   :12.000   Max.   :108.80
route_length_km  delivery_mode      weather      order_time
Min.   : 0.580   Length:200   Length:200   Length:200
1st Qu.: 4.420   Class :character   Class :character   Class :character
Median : 8.520   Mode  :character   Mode  :character   Mode  :character
Mean   : 8.156
3rd Qu.:11.800
Max.   :17.290
restaurant_zone  customer_zone
Length:200       Length:200
Class :character   Class :character
Mode  :character   Mode  :character

>
> # =====
> # 4. IMPROVING SUMMARY USING FACTORS
> # Convert character columns to factors for better summary
> # =====
> df$traffic_level <- as.factor(df$traffic_level)
> df$delivery_mode <- as.factor(df$delivery_mode)
> df$weather <- as.factor(df$weather)
> df$restaurant_zone <- as.factor(df$restaurant_zone)
> df$customer_zone <- as.factor(df$customer_zone)
>
> print("---- OUTPUT OF summary() [After Factor Conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Background Jobs
R - R 4.5.2 - ~/
> # Convert character columns to factors for better summary
> # =====
> df$traffic_level <- as.factor(df$traffic_level)
> df$delivery_mode <- as.factor(df$delivery_mode)
> df$weather <- as.factor(df$weather)
> df$restaurant_zone <- as.factor(df$restaurant_zone)
> df$customer_zone <- as.factor(df$customer_zone)
>
> print("---- OUTPUT OF summary() [After Factor Conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
> summary(df)
  order_id      distance_km  delivery_time_min traffic_level
Min.   : 1.00   Min.   : 0.530   Min.   : 2.40   High :68
1st Qu.: 50.75   1st Qu.: 3.688   1st Qu.: 23.73   Low  :65
Median :100.50   Median : 6.845   Median : 44.65   medium:67
Mean   :100.50   Mean   : 6.625   Mean   : 44.74
3rd Qu.:150.25   3rd Qu.: 9.797   3rd Qu.: 63.90
Max.   :200.00   Max.   :12.000   Max.   :108.80
route_length_km  delivery_mode      weather      order_time      restaurant_zone
Min.   : 0.580   Bicycle:52   Clear :40   Length:200   Central:47
1st Qu.: 4.420   Bike :49   Cloudy:55   Class :character   East :28
Median : 8.520   Car :47   Rainy :55   Mode  :character   North :30
Mean   : 8.156   Scooter:52   windy :50
3rd Qu.:11.800
Max.   :17.290
customer_zone
Central :38
East :45
North :46
South :36
west :35

>
> # =====
> # 5. Specific Summaries (Examples)
> # =====
> avg_delivery_time <- mean(df$delivery_time_min, na.rm = TRUE)
> max_distance <- max(df$distance_km, na.rm = TRUE)
> min_route_length <- min(df$route_length_km, na.rm = TRUE)
>
> print(paste("Average delivery time (min):", avg_delivery_time))
```

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
Source
Console Background Jobs
R - R4.5.2 - ~/
> df$customer_zone <- as.factor(df$customer_zone)
> print("---- OUTPUT OF summary() [After Factor Conversion] ----")
[1] "---- OUTPUT OF summary() [After Factor conversion] ----"
> summary(df)
  order_id      distance_km  delivery_time_min traffic_level
Min.   : 1.00      Min.   : 0.530      Min.   : 2.40      High :68
1st Qu.:50.75      1st Qu.: 3.688      1st Qu.: 23.73     Low  :65
Median :100.50     Median : 6.845      Median : 44.65     medium:67
Mean   :100.50     Mean   : 6.625      Mean   : 44.74
3rd Qu.:150.25     3rd Qu.: 9.797      3rd Qu.: 63.90
Max.   :200.00     Max.   :12.000      Max.   :108.80
route_length_km  delivery_mode  weather  order_time  restaurant_zone
Min.   : 0.580      Bicycle:52      Clear :40      Length:200      Central:47
1st Qu.: 4.420      Bike :49        Cloudy:55     Class :character  East :28
Median : 8.320      Car :47         Rainy :35     Mode :character   North:30
Mean   : 8.156      Scooter:52      windy :50
3rd Qu.:11.800
Max.   :17.290
customer_zone
Central:38
East :45
North :46
South :36
West :35

>
> # =====
> # 5. Specific Summaries (Examples)
> # =====
>
> avg_delivery_time <- mean(df$delivery_time_min, na.rm = TRUE)
> max_distance <- max(df$distance_km, na.rm = TRUE)
> min_route_length <- min(df$route_length_km, na.rm = TRUE)
>
> print(paste("Average Delivery Time (min):", avg_delivery_time))
[1] "Average Delivery Time (min): 44.7445"
> print(paste("Maximum Delivery Distance (km):", max_distance))
[1] "Maximum Delivery Distance (km): 12"
> print(paste("Minimum Route Length (km):", min_route_length))
[1] "Minimum Route Length (km): 0.58"
>

Environment History Connections Tutorial
R - Global Environment
Data
combined_d... 10150 obs. of 2 variables
df            200 obs. of 10 variables
df_clean      200 obs. of 11 variables
df_no_dupl... 200 obs. of 10 variables
df_small      200 obs. of 4 variables
duplicate...  205 obs. of 10 variables
duplicate...  10 obs. of 10 variables
flower_cle... 10000 obs. of 2 variables
flower_df     10000 obs. of 4 variables
Food_Deliv... 200 obs. of 5 variables
iris          150 obs. of 5 variables
iris_clean    150 obs. of 2 variables
long_df       600 obs. of 3 variables
processed...  200 obs. of 21 variables
traffic_pi... 200 obs. of 4 variables
unique_ord... 200 obs. of 10 variables
unique_rou... 200 obs. of 10 variables
wide_df       200 obs. of 4 variables
Values
avg_delive... 44.7445
current_ti... 2025-12-08 14:15:16 IST
max_distan... 12
min_route... 0.58
Files Plots Packages Help Viewer Presentation
Present < Print < Edit <
```

Name :- Priya Gupta

Roll no. :- S081