

R Data Visualization Project

Step-by-Step Code Explanation (Beginner Level)

Project Introduction

Is project mein hum ne supermarket ke sales data par kaam kiya. Is ka maqsad yeh sikhna hai ke R language ka use karke:

- Dataset kaise load hota hai
- Data clean kaise hota hai
- Graphs kaise banaye jate hain
- Graphs ko image file mein kaise save kiya jata hai

Yeh explanation bilkul beginner student ke liye likhi gayi hai.

STEP 1: Dataset Load Karna

Sab se pehla kaam hota hai yeh check karna ke R kis folder mein kaam kar raha hai.

Working Directory Check

```
getwd()
```

Explanation: `getwd()` ka matlab hai *Get Working Directory*. Yeh command batati hai ke R abhi kis folder mein kaam kar raha hai. Dataset file isi folder ke andar honi chahiye warna R usay read nahi karega.

Dataset Load Karna

```
data <- read.csv("data/SuperMarketAnalysis.csv")
```

Explanation:

- `read.csv()` CSV file ko read karta hai
- `data <-` ka matlab hai ke hum dataset ko `data` naam ke variable mein store kar rahe hain
- Ab poora dataset R ke andar load ho chuka hai

Dataset ka First Look

```
head(data)
str(data)
```

Explanation:

- `head(data)` dataset ki pehli 6 rows dikhata hai
- `str(data)` dataset ka structure dikhata hai

Is step se hume pata chalta hai ke data mein kaun se columns hain aur unka type kya hai.

STEP 2: Data Cleaning

Data cleaning ka matlab hai data ko analysis ke liye ready karna.

Date Column Format Change

```
data$date <- as.Date(data$date, format = "%m/%d/%Y")
```

Explanation:

- Dataset mein Date pehle text format mein hoti hai
- `as.Date()` use karke hum date ko proper date bana dete hain
- Is ke baghair time-based graph nahi ban sakta

Missing Values Check

```
colSums(is.na(data))
```

Explanation:

- `is.na()` missing values check karta hai

- `colSums()` har column ka total batata hai
- Agar result 0 aaye to matlab data complete hai

STEP 3: Libraries Install aur Load Karna

Graphs aur analysis ke liye packages zaroori hote hain.

Packages Install Karna

```
install.packages("ggplot2")
install.packages("dplyr")
```

Explanation:

- `ggplot2` graphs ke liye use hota hai
- `dplyr` data ko group aur summarize karne ke liye
- Install sirf ek dafa karna hota hai

Libraries Load Karna

```
library(ggplot2)
library(dplyr)
```

Explanation: Library load karne ke baad hi hum unke functions use kar sakte hain.

STEP 4: Visualizations

Is step mein hum different graphs banate hain.

Plot 1: Sales Trend Over Time

```
sales_time <- data %>%
  group_by(Date) %>%
  summarise(Total_Sales = sum(Sales))
```

Explanation:

- `group_by(Date)` same date wali sales ko group karta hai

- `sum(Sales)` har din ki total sales nikalta hai

```
p1 <- ggplot(sales_time, aes(x = Date, y = Total_Sales)) +
  geom_line(color = "blue", linewidth = 1) +
  labs(title = "Sales\u20a6Trend\u20a6Over\u20a6Time",
       x = "Date",
       y = "Total\u20a6Sales") +
  theme_minimal()
```

Explanation:

- `ggplot()` graph ka base hai
- `aes()` x aur y axis define karta hai
- `geom_line()` line graph banata hai

```
ggsave("01_sales_trend_over_time.png", p1, width = 8, height = 5)
```

Explanation: Graph ko image file mein save karta hai.

Baaki Graphs ka Simple Concept

Isi tarah:

- Bar chart → product comparison
- Box plot → city aur customer comparison
- Histogram → sales distribution
- Scatter plot → sales aur profit ka relationship

Sab graphs ka structure same hota hai:

```
ggplot() + geom_*( ) + labs() + ggsave()
```

Final Learning Outcome

Is project ke baad student:

- Dataset load karna seekh jata hai
- Data clean karna samajh jata hai

- ggplot2 ka basic structure samajh leta hai
- Real-world data par graphs banana seekh jata hai

Yeh project R Data Visualization seekhne ke liye best starting point hai.