

```
install.packages(c(
  "tidyverse",
  "tibble",
  "tidyr",
  "purrr",
  "dplyr",
  "stringr",
  "plotly"
))

library(tidyverse)
library(plotly)
library(GGally)
library(corrplot)

store_data <- read.table(text = "
Store_ID Region Month Total_Sales Number_Of_Customers Average_Transaction_Value
1 North January 10000 500 20
2 South January 15000 700 21.43
3 East January 12000 600 20
4 West January 13000 650 20
5 North February 9000 450 20
6 South February 16000 800 20
7 East February 14000 700 20
8 West February 11000 550 20
9 North March 11000 550 20
10 South March 17000 850 20
11 East March 13000 650 20
12 West March 15000 750 20
13 North April 12000 600 20
14 South April 18000 900 20
15 East April 16000 800 20")
```

16 West April 14000 700 20
17 North May 13000 650 20
18 South May 19000 950 20
19 East May 15000 750 20
20 West May 16000 800 20
21 North June 14000 700 20
22 South June 20000 1000 20
23 East June 17000 850 20
24 West June 15000 750 20
25 North July 15000 750 20
26 South July 21000 1050 20
27 East July 16000 800 20
28 West July 14000 700 20
29 North August 16000 800 20
30 South August 22000 1100 20
31 East August 13000 650 20
32 West August 15000 750 20
33 North September 12000 600 20
34 South September 18000 900 20
35 East September 16000 800 20
36 West September 14000 700 20
37 North October 13000 650 20
38 South October 19000 950 20
39 East October 15000 750 20
40 West October 16000 800 20
41 North November 14000 700 20
42 South November 20000 1000 20
43 East November 17000 850 20
44 West November 15000 750 20
45 North December 15000 750 20
46 South December 21000 1050 20

```
47 East December 16000 800 20
```

```
48 West December 14000 700 20
```

```
", header = TRUE)
```

```
month_levels <-  
c("January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December")
```

```
store_data$Region <- factor(store_data$Region)
```

```
store_data$Month <- factor(store_data$Month, levels = month_levels, ordered = TRUE)
```

```
store_data$Month_Index <- as.numeric(store_data$Month)
```

```
na_counts <- sapply(store_data, function(x) sum(is.na(x)))
```

```
store_data <- distinct(store_data)
```

```
region_sales <- store_data %>%
```

```
group_by(Region) %>%
```

```
summarise(Total_Sales_Region = sum(Total_Sales), Total_Customers_Region =  
sum(Number_Of_Customers))
```

```
monthly_sales <- store_data %>%
```

```
group_by(Month, Month_Index) %>%
```

```
summarise(Total_Sales_Month = sum(Total_Sales), Total_Customers_Month =  
sum(Number_Of_Customers))
```

```
region_month <- store_data %>%
```

```
group_by(Region, Month, Month_Index) %>%
```

```
summarise(Total_Sales = sum(Total_Sales), Customers = sum(Number_Of_Customers))
```

```
top_sales <- store_data %>%
```

```
arrange(desc(Total_Sales))
```

```
customer_month_analysis <- monthly_sales %>%
```

```
arrange(Month_Index)

atv_stats <- store_data %>%
  summarise(Mean_ATV = mean(Average_Transaction_Value),
            Median_ATV = median(Average_Transaction_Value),
            SD_ATV = sd(Average_Transaction_Value))

print(region_sales)
print(monthly_sales)
print(top_sales)
print(customer_month_analysis)
print(atv_stats)

p1 <- ggplot(monthly_sales, aes(x = Month, y = Total_Sales_Month, group = 1)) + geom_line() +
  geom_point()
print(p1)

p2 <- ggplot(region_sales, aes(x = Region, y = Total_Sales_Region)) + geom_col()
print(p2)

p3 <- ggplot(region_month, aes(x = Month, y = Total_Sales, color = Region, group = Region)) +
  geom_line() + geom_point()
print(p3)

p4 <- ggplot(monthly_sales, aes(x = Month, y = Total_Customers_Month, group = 1)) + geom_line() +
  geom_point()
print(p4)

p5 <- ggplot(region_sales, aes(x = Region, y = Total_Customers_Region)) + geom_col()
print(p5)
```

```
p6 <- ggplot(store_data, aes(x = Number_of_Customers, y = Total_Sales, color = Region)) +  
  geom_point()  
  
print(p6)  
  
p7 <- ggplot(store_data, aes(x = Average_Transaction_Value, y = Total_Sales, color = Region)) +  
  geom_point()  
  
print(p7)  
  
p8 <- ggplot(store_data, aes(x = Region, y = Total_Sales)) + geom_boxplot()  
  
print(p8)  
  
p9 <- ggplot(store_data, aes(x = Region, y = Number_of_Customers)) + geom_boxplot()  
  
print(p9)  
  
region_month_sales <- region_month  
  
p10 <- ggplot(region_month_sales, aes(x = Month, y = Region, fill = Total_Sales)) + geom_tile()  
  
print(p10)  
  
p11 <- plot_ly(store_data, x = ~Month_Index, y = ~Total_Sales, z = ~Number_of_Customers, color =  
  ~Region, type = "scatter3d", mode = "markers")  
  
p11  
  
p12 <- plot_ly(region_month, x = ~Month_Index, y = ~Total_Sales, z = ~Customers, color = ~Region,  
  type = "scatter3d", mode = "lines+markers")  
  
p12  
  
p13 <- ggplot(monthly_sales, aes(x = Month_Index, y = Total_Sales_Month)) + geom_line() +  
  geom_smooth(method = "lm", se = FALSE)  
  
print(p13)  
  
p14 <- ggplot(region_month, aes(x = Month_Index, y = Total_Sales, color = Region)) + geom_point() +  
  geom_smooth(method = "lm", se = FALSE)  
  
print(p14)
```

```
p15 <- ggplot(region_month, aes(x = Month_Index, y = Total_Sales, group = Region, color = Region)) +  
  geom_line() + facet_wrap(~Region)
```

```
print(p15)
```

```
p16 <- ggplot(store_data, aes(x = Total_Sales)) + geom_histogram(bins = 10)
```

```
print(p16)
```

```
p17 <- ggplot(store_data, aes(x = Total_Sales)) + geom_density()
```

```
print(p17)
```

```
p18 <- ggplot(store_data, aes(x = Number_Of_Customers)) + geom_histogram(bins = 10)
```

```
print(p18)
```

```
p19 <- ggplot(store_data, aes(x = Average_Transaction_Value)) + geom_histogram(bins = 10)
```

```
print(p19)
```

```
region_share <- store_data %>%  
  group_by(Region) %>%  
  summarise(Total_Sales_Region = sum(Total_Sales))
```

```
p20 <- ggplot(region_share, aes(x = "", y = Total_Sales_Region, fill = Region)) + geom_col(width = 1) +  
  coord_polar(theta = "y")
```

```
print(p20)
```

```
p21 <- ggplot(region_month, aes(x = Month, y = Total_Sales, fill = Region)) + geom_col()
```

```
print(p21)
```

```
p22 <- ggplot(region_month, aes(x = Month, y = Customers, color = Region, group = Region)) +  
  geom_line() + geom_point()
```

```
print(p22)
```

```
p23 <- ggplot(store_data, aes(x = Average_Transaction_Value, y = Number_Of_Customers, color = Month)) + geom_point()  
print(p23)
```

```
numeric_data <- store_data %>%  
  select(Total_Sales, Number_Of_Customers, Average_Transaction_Value, Month_Index)
```

```
p24 <- ggpairs(numeric_data)
```

```
print(p24)
```

```
corr_matrix <- cor(numeric_data)
```

```
p25 <- corrplot(corr_matrix, method = "color")
```





