```r
# Amazon Order

library(readxl)
amazon <- read_excel("orders_data.xlsx")

# View the Structure of Amazon Order Data
str(amazon)

# Quick Summary
summary(amazon)

# Check for Missing Value
colSums(is.na(amazon))


### Will do few analysis with Dataset
# How is the order take places?
table(amazon$cod)
prop.table(table(amazon$cod))
# 27% COD & 72% Online

# iS return of order is depends on mode of delivery
table(amazon$cod, amazon$order_status)
# both way having similar values so it's not really depends on mode of delivery



#install.packages("dplyr")
library(dplyr)
# Sales Trend (Sales Pattern over the Month)
sales_by_month = amazon %>%
  select(item_total,month) %>%
  group_by(month) %>%
  summarise(item_total=sum(item_total))

barplot(sales_by_month$item_total, names.arg = sales_by_month$month)


# More detailed
sales_by_month_year = amazon %>%
  select(item_total,month,year)%>%
  group_by(month,year) %>%
  summarise(sales=sum(item_total))


sales_by_month_year <- sales_by_month_year[order(sales_by_month_year$year,
                                                  sales_by_month_year$month),]

sales_by_month_year$MonthYear <- paste(sales_by_month_year$month,
                                       sales_by_month_year$year,
                                       sep = '-')

barplot(sales_by_month_year$sales,
        names.arg = sales_by_month_year$MonthYear,
        main="Sales Trend over the month",
        xlab = "Month-Year",
        ylab = "Sales",
        col=rainbow(length(sales_by_month_year$MonthYear)))
```

```r
# Which day of the month draw more sales?
dec_2021 <- subset(amazon, month==12 & year==2021)

sales_by_day = dec_2021 %>%
  select(day, item_total) %>%
  group_by(day) %>%
  summarise(sales=sum(item_total))

barplot(sales_by_day$sales,
        names.arg = sales_by_day$day,
        main="Sales Trend For Dec-2021",
        xlab = "Day",
        ylab = "Sales",
        col=rainbow(length(sales_by_day$day)))


# Find out from which state draw more sales?
sales_by_state <-
  amazon %>%
  select(ship_state, item_total) %>%
  group_by(ship_state) %>%
  summarise(sales=sum(item_total))

sales_by_state <-
  sales_by_state[order(sales_by_state$sales, decreasing = TRUE),]

# Top 5 States
head(sales_by_state,5)


#calculate the sales percentage
sales_by_state$sales_per <- (sales_by_state$sales / sum(sales_by_state$sales)) * 100
sales_by_state$sales_per <- round(sales_by_state$sales_per,2)
sales_by_state$state_with_per <- paste(sales_by_state$ship_state,
                                       sales_by_state$sales_per,
                                       sep = "-")

top_state = head(sales_by_state,5)

pie(top_state$sales_per,
    top_state$state_with_per,
    main="Top 5 States % of Sales",
    col=rainbow(5))


# How many orders got return in Maharastra
mah_sales <- subset(amazon, ship_state=="MAHARASHTRA")

table(mah_sales$order_status)

# What is the delivery mode of Return order
table(mah_sales$order_status, mah_sales$cod)



install.packages("jsonify")
library(jsonify)
```

```r
data <- from_json("sample.json")


install.packages("haven")
library(haven)
survey <- read_sav("survey.sav")



library(XML)
sample <- xmlParse("books.xml")
sampleRoot <- xmlRoot(sample)

sample_df <- xmlToDataFrame(sampleRoot)
```

In [ ]: