```
# Project - Amazon Sales Analysis

# Name - Nabeel Ghalib M A Q

# install and load packages

install.packages("tidyverse")
install.packagees("lubridate")

library(tidyverse)
library(lubridate)
library(dplyr)
library(ggplot2)
library(janitor)

# Importing dataset

amazondatacpy = read.csv("F:/Data Analytics/Unified mentor projects datasets/Projects to
work/Amazonsalesdata_dmy.csv")

view(amazondatacpy)


# Converting date to different format and separating the date in order date into three separate columns
such as year, month ,day so we can group and find data for each year, month etc..

amazondatacpy$‘Order Date‘ <- format(as.Date(amazondatacpy$‘Order Date‘,
format="%d/%m/%Y"),"%Y/%m/%d")
amazondatacpy$‘Ship Date‘  <- format(as.Date(amazondatacpy$‘Ship Date‘,
format="%d/%m/%Y"),"%Y/%m/%d")

amazondatacpy$Year= format(as.Date(amazondatacpy$‘Order Date‘), "%Y")
amazondatacpy$month= format(as.Date(amazondatacpy$‘Order Date‘), "%m")
amazondatacpy$day= format(as.Date(amazondatacpy$‘Order Date‘), "%d")


# Finding distinct values
# Checking for null, NA , duplicates etc

nrow(amazondatacpy)

ncol(amazondatacpy)

is.null(amazondatacpy)

is.na(amazondatacpy)

is.na(amazondatacpy$‘Order Date‘)

n_distinct(amazondatacpy)

n_distinct(amazondatacpy$Region)

n_distinct(amazondatacpy$Country)
```

```
n_distinct(amazondatacpy$‘Item Type‘)

n_distinct(amazondatacpy$Year)

n_distinct(amazondatacpy$‘Order Priority‘)

# Finding total profit, sales, cost, units sold, revenue

amazondatacpy %>%
  select(Region,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  view()

# Finding total profit , sales, cost ,units sold, revenue country and region wise

# Region wise

amazondatacpy %>%
  select(Region,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(Region) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-(total_profit)) %>%
  view()

# Country wise

amazondatacpy %>%
  select(Country,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(Country) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-(total_profit)) %>%
  view()

# Region and Country wise

amazondatacpy %>%
  select(Region, Country, ‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(Region,Country) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-(total_profit)) %>%
  view()


# Finding total profit , sales, cost ,units sold, revenue  region wise and sales channel wise

amazondatacpy %>%
  select(Region,‘Sales Channel‘,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(Region,‘Sales Channel‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(‘Sales Channel‘, -total_profit) %>%
```

```
  view()


# Finding region and item type wise total profit , sales, cost ,units sold, revenue

amazondatacpy %>%
  select(Region,‘Item Type‘,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(Region,‘Item Type‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(Region,-total_profit) %>%
  view()


# Finding region ,sales channel and item type wise total profit , sales, cost ,units sold, revenue

amazondatacpy %>%
  select(Region,‘Sales Channel‘,‘Item Type‘,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit
Cost‘) %>%
  group_by(Region,‘Sales Channel‘,‘Item Type‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(‘Sales Channel‘, -total_profit) %>%
  view()


# Finding year wise, monthly wise, yearly-monthly wise profit, revenue, units sold, total cost

# Year wise

amazondatacpy %>%
  select(Year,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(Year) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-total_profit) %>%
  view()


# month wise

amazondatacpy %>%
  select(month,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(month) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-total_profit) %>%
  view()


# Yearly-month wise

amazondatacpy %>%
  select(Year,month,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(Year,month) %>%
```

```
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-total_profit) %>%
  view()


# Sales channel

amazondatacpy %>%
  select(‘Sales Channel‘,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(‘Sales Channel‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-total_profit) %>%
  view()


# Order Priority wise

amazondatacpy %>%
  select(‘Order Priority‘,‘Total Revenue‘,‘Total Profit‘,‘Total Cost‘,‘Units Sold‘,‘Unit Cost‘) %>%
  group_by(‘Order Priority‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-(total_profit)) %>%
  view()


# Item type

amazondatacpy %>%
  select(‘Item Type‘,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(‘Item Type‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(-total_profit) %>%
  view()


# Finding item type and sales channel wise profit, tota cost, units sold, total revenue

amazondatacpy %>%
  select(‘Sales Channel‘,‘Item Type‘,‘Total Revenue‘,‘Total Cost‘,‘Total Profit‘,‘Units Sold‘) %>%
  group_by(‘Sales Channel‘,‘Item Type‘) %>%
  summarise(total_units_sold=sum(‘Units Sold‘),total_cost=sum(‘Total Cost‘),total_revenue=sum(‘Total
Revenue‘),total_profit=sum(‘Total Profit‘)) %>%
  arrange(‘Sales Channel‘,-total_profit) %>%
  view()


                                                                    # END #
```