

Retail Store Sales

Kaushal Khatiwada

2024-04-19

Data Wrangling

a)Data Collection b)Data Inspection c)Data Cleaning Handling Missing Values Removing Duplicates Standardizing Formats d)Exploratory Data Analysis (EDA)

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.0      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(tidyr)
```

A) Data Collection

Import Retail Store sales dataset in a command-seperated file.

```
sales<- read.csv("Sales.csv")
```

B) Data Inspection

Summary of the structure of a dataset.

```
str(sales)
```

```
## 'data.frame':   1000 obs. of  17 variables:
## $ Invoice.ID      : chr  "750-67-8428" "226-31-3081" "631-41-3108" "123-19-1176" ...
## $ Branch         : chr  "A" "C" "A" "A" ...
## $ City           : chr  "Yangon" "Naypyitaw" "Yangon" "Yangon" ...
## $ Customer.type  : chr  "Member" "Normal" "Normal" "Member" ...
```

```
## $ Gender           : chr  "Female" "Female" "Male" "Male" ...
## $ Product.line     : chr  "Health and beauty" "Electronic accessories" "Home and lifestyle" "
## $ Unit.price       : num  74.7 15.3 46.3 58.2 86.3 ...
## $ Quantity         : int   7 5 7 8 7 7 6 10 2 3 ...
## $ Tax.5.           : num  26.14 3.82 16.22 23.29 30.21 ...
## $ Total            : num  549 80.2 340.5 489 634.4 ...
## $ Date             : chr   "1/5/2019" "3/8/2019" "3/3/2019" "1/27/2019" ...
## $ Time             : chr   "13:08" "10:29" "13:23" "20:33" ...
## $ Payment          : chr   "Ewallet" "" "Credit card" "Ewallet" ...
## $ cogs             : num  522.8 76.4 324.3 465.8 604.2 ...
## $ gross.margin.percentage: num  4.76 4.76 4.76 4.76 4.76 ...
## $ gross.income      : num  26.14 3.82 16.22 23.29 30.21 ...
## $ Rating           : num  9.1 9.6 7.4 8.4 5.3 4.1 5.8 8 7.2 5.9 ...
```

Dimensions of dataset.

```
dim(sales)
```

```
## [1] 1000  17
```

Display few rows of dataset.

```
head(sales)
```

```
## Invoice.ID Branch City Customer.type Gender Product.line
## 1 750-67-8428 A Yangon Member Female Health and beauty
## 2 226-31-3081 C Naypyitaw Normal Female Electronic accessories
## 3 631-41-3108 A Yangon Normal Male Home and lifestyle
## 4 123-19-1176 A Yangon Member Male Health and beauty
## 5 373-73-7910 A Yangon Normal Male Sports and travel
## 6 699-14-3026 C Naypyitaw Normal Male Electronic accessories
## Unit.price Quantity Tax.5. Total Date Time Payment cogs
## 1 74.69 7 26.1415 548.9715 1/5/2019 13:08 Ewallet 522.83
## 2 15.28 5 3.8200 80.2200 3/8/2019 10:29 76.40
## 3 46.33 7 16.2155 340.5255 3/3/2019 13:23 Credit card 324.31
## 4 58.22 8 23.2880 489.0480 1/27/2019 20:33 Ewallet 465.76
## 5 86.31 7 30.2085 634.3785 2/8/2019 10:37 Ewallet 604.17
## 6 85.39 7 29.8865 627.6165 3/25/2019 18:30 Ewallet 597.73
## gross.margin.percentage gross.income Rating
## 1 4.761905 26.1415 9.1
## 2 4.761905 3.8200 9.6
## 3 4.761905 16.2155 7.4
## 4 4.761905 23.2880 8.4
## 5 4.761905 30.2085 5.3
## 6 4.761905 29.8865 4.1
```

```
colnames(sales)
```

```
## [1] "Invoice.ID" "Branch"
## [3] "City" "Customer.type"
## [5] "Gender" "Product.line"
## [7] "Unit.price" "Quantity"
```

```
## [9] "Tax.5." "Total"
## [11] "Date" "Time"
## [13] "Payment" "cogs"
## [15] "gross.margin.percentage" "gross.income"
## [17] "Rating"
```

1. Invoice id: Computer generated sales slip invoice identification number
2. Branch: Branch of supercenter (3 branches are available identified by A, B and C).
3. City: Location of supercenters
4. Customer type: Members for customers using member card and Normal for without member card.
5. Gender: Gender type of customer
6. Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel
7. Unit price: Price of each product in USD
8. Quantity: Number of products purchased by customer
9. Tax: 5% tax fee for customer buying
10. Total: Total price including tax
11. Date: Date of purchase (Record available from January 2019 to March 2019)
12. Time: Purchase time (10am to 9pm)
13. Payment: Payment Type – Cash, Credit card and Ewallet)
14. COGS: Cost of goods sold
15. Gross margin percentage: Gross margin percentage
16. Gross income: Gross income
17. Rating: Customer rating on shopping experience (On a scale of 1 to 10)

```
colSums(is.na(sales))
```

```
## Invoice.ID Branch City
## 0 0 0
## Customer.type Gender Product.line
## 0 0 0
## Unit.price Quantity Tax.5.
## 2 16 14
## Total Date Time
## 14 0 0
## Payment cogs gross.margin.percentage
## 0 0 29
## gross.income Rating
## 0 0
```

```
sum(duplicated(sales))
```

```
## [1] 0
```

There is no duplicate records.

c) Data Cleaning

Handling Missing Values For Unit Price

```
sales %>% filter(!complete.cases(sales$Unit.price))
```

```
##      Invoice.ID Branch      City Customer.type Gender      Product.line
## 1 315-22-5665      C Naypyitaw      Normal Female Home and lifestyle
## 2 372-26-1506      C Naypyitaw      Normal Female Fashion accessories
##      Unit.price Quantity Tax.5.   Total      Date Time Payment  cogs
## 1          NA      10 36.780 772.380 2/24/2019 11:38 Ewallet 735.6
## 2          NA       5  5.955 125.055 1/28/2019 19:24 Ewallet 119.1
##      gross.margin.percentage gross.income Rating
## 1                4.761905        36.780    8.0
## 2                4.761905         5.955    5.4
```

2 records are NA. So, fill the NA as Unit price = (Total - 5% Tax)/Quantity

```
sales<- sales %>%
  mutate(Unit.price = ifelse(is.na(Unit.price), (Total - Tax.5.)/Quantity, Unit.price))
```

```
colSums(is.na(sales))
```

```
##      Invoice.ID      Branch      City
##           0           0           0
##      Customer.type      Gender      Product.line
##           0           0           0
##      Unit.price      Quantity      Tax.5.
##           0           16           14
##      Total      Date      Time
##          14           0           0
##      Payment      cogs gross.margin.percentage
##           0           0           29
##      gross.income      Rating
##           0           0
```

For Quantity

```
sales %>%
  filter(!complete.cases(Quantity))
```

```
##      Invoice.ID Branch      City Customer.type Gender      Product.line
## 1 418-05-0656      B Mandalay      Normal Female Fashion accessories
## 2 804-38-3935      A Yangon      Member Male Electronic accessories
## 3 866-70-2814      B Mandalay      Normal Female Electronic accessories
## 4 101-81-4070      C Naypyitaw      Member Female Health and beauty
## 5 851-98-3555      B Mandalay      Normal Female Health and beauty
## 6 186-71-5196      A Yangon      Member Female Food and beverages
## 7 181-82-6255      B Mandalay      Normal Female Home and lifestyle
## 8 235-06-8510      C Naypyitaw      Member Male Home and lifestyle
## 9 500-02-2261      C Naypyitaw      Normal Female Food and beverages
```

| ## | Invoice.ID | Branch | City | Customer.type | Gender | Product.line | | |
|-------|-------------------------|--------------|-----------|---------------|-----------|------------------------|-------------|--------|
| ## 10 | 577-34-7579 | C | Naypyitaw | Member | Male | Food and beverages | | |
| ## 11 | 518-71-6847 | B | Mandalay | Member | Male | Food and beverages | | |
| ## 12 | 156-20-0370 | B | Mandalay | Normal | Female | Electronic accessories | | |
| ## 13 | 552-44-5977 | B | Mandalay | Member | Male | Health and beauty | | |
| ## 14 | 602-16-6955 | B | Mandalay | Normal | Female | Sports and travel | | |
| ## 15 | 745-74-0715 | A | Yangon | Normal | Male | Electronic accessories | | |
| ## 16 | 690-01-6631 | B | Mandalay | Normal | Male | Fashion accessories | | |
| ## | Unit.price | Quantity | Tax.5. | Total | Date | Time | Payment | cogs |
| ## 1 | 25.56 | NA | 8.9460 | 187.8660 | 2/2/2019 | 20:42 | Cash | 178.92 |
| ## 2 | 93.78 | NA | 14.0670 | 295.4070 | 1/30/2019 | 11:32 | Credit card | 281.34 |
| ## 3 | 52.79 | NA | 26.3950 | 554.2950 | 2/25/2019 | 11:58 | Ewallet | 527.90 |
| ## 4 | 62.82 | NA | 6.2820 | 131.9220 | 1/17/2019 | 12:36 | Ewallet | 125.64 |
| ## 5 | 82.88 | NA | 20.7200 | 435.1200 | 3/24/2019 | 14:08 | Credit card | 414.40 |
| ## 6 | 79.54 | NA | 7.9540 | 167.0340 | 3/27/2019 | 16:30 | Ewallet | 159.08 |
| ## 7 | 16.37 | NA | 4.9110 | 103.1310 | 2/8/2019 | 10:58 | Cash | 98.22 |
| ## 8 | 85.72 | NA | 12.8580 | 270.0180 | 1/24/2019 | 20:59 | Ewallet | 257.16 |
| ## 9 | 57.29 | NA | 17.1870 | 360.9270 | 3/21/2019 | 17:04 | Ewallet | 343.74 |
| ## 10 | 50.49 | NA | 22.7205 | 477.1305 | 1/10/2019 | 17:16 | Cash | 454.41 |
| ## 11 | 26.60 | NA | 7.9800 | 167.5800 | 2/26/2019 | 15:10 | Ewallet | 159.60 |
| ## 12 | 25.45 | NA | 1.2725 | 26.7225 | 3/10/2019 | 18:10 | Credit card | 25.45 |
| ## 13 | 62.00 | NA | 24.8000 | 520.8000 | 1/3/2019 | 19:08 | Credit card | 496.00 |
| ## 14 | 76.60 | NA | 38.3000 | 804.3000 | 1/24/2019 | 18:10 | Ewallet | 766.00 |
| ## 15 | 58.03 | NA | 5.8030 | 121.8630 | 3/10/2019 | 20:46 | Ewallet | 116.06 |
| ## 16 | 17.49 | NA | 8.7450 | 183.6450 | 2/22/2019 | 18:35 | Ewallet | 174.90 |
| ## | gross.margin.percentage | gross.income | Rating | | | | | |
| ## 1 | 4.761905 | 8.9460 | 7.1 | | | | | |
| ## 2 | 4.761905 | 14.0670 | 5.9 | | | | | |
| ## 3 | 4.761905 | 26.3950 | 10.0 | | | | | |
| ## 4 | 4.761905 | 6.2820 | 4.9 | | | | | |
| ## 5 | NA | 20.7200 | 6.6 | | | | | |
| ## 6 | NA | 7.9540 | 6.2 | | | | | |
| ## 7 | 4.761905 | 4.9110 | 7.0 | | | | | |
| ## 8 | 4.761905 | 12.8580 | 5.1 | | | | | |
| ## 9 | 4.761905 | 17.1870 | 5.9 | | | | | |
| ## 10 | 4.761905 | 22.7205 | 5.4 | | | | | |
| ## 11 | 4.761905 | 7.9800 | 4.9 | | | | | |
| ## 12 | 4.761905 | 1.2725 | 5.1 | | | | | |
| ## 13 | 4.761905 | 24.8000 | 6.2 | | | | | |
| ## 14 | 4.761905 | 38.3000 | 6.0 | | | | | |
| ## 15 | 4.761905 | 5.8030 | 8.8 | | | | | |
| ## 16 | NA | 8.7450 | 6.6 | | | | | |

16 records are NA. So, fill the NA as

(Total - 5% Tax)/Unit price

```
sales<- sales %>%
  mutate(Quantity = ifelse(is.na(Quantity),(Total - Tax.5.)/Unit.price,Quantity))
```

```
colSums(is.na(sales))
```

| ## | Invoice.ID | Branch | City |
|----|---------------|--------|--------------|
| ## | 0 | 0 | 0 |
| ## | Customer.type | Gender | Product.line |

```
##          0          0          0
##      Unit.price      Quantity      Tax.5.
##          0          0          14
##      Total          Date          Time
##          14          0          0
##      Payment      cogs gross.margin.percentage
##          0          0          29
##      gross.income      Rating
##          0          0
```

For 5% Tax

```
sales %>%
  filter(!complete.cases(Tax.5.))
```

```
##      Invoice.ID Branch      City Customer.type Gender      Product.line
## 1  299-46-1805      B Mandalay      Member Female Sports and travel
## 2  656-95-9349      A  Yangon      Member Female Health and beauty
## 3  765-26-6951      A  Yangon      Normal   Male   Sports and travel
## 4  393-65-2792      C Naypyitaw Normal   Male   Food and beverages
## 5  796-12-2025      C Naypyitaw Normal   Male   Fashion accessories
## 6  510-95-6347      B Mandalay      Member Female Food and beverages
## 7  766-85-7061      B Mandalay      Normal   Male   Health and beauty
## 8  659-36-1684      C Naypyitaw Member   Male   Sports and travel
## 9  575-30-8091      A  Yangon      Normal   Male   Sports and travel
## 10 894-41-5205      C Naypyitaw Normal Female Food and beverages
## 11 490-29-1201      A  Yangon      Normal Female Sports and travel
## 12 442-48-3607      A  Yangon      Member   Male   Food and beverages
## 13 525-88-7307      B Mandalay      Member   Male   Sports and travel
## 14 234-03-4040      B Mandalay      Member Female Food and beverages
##      Unit.price Quantity Tax.5.      Total      Date Time      Payment      cogs
## 1      93.72         6      NA 590.4360 1/15/2019 16:19      Cash 562.32
## 2      68.93         7      NA 506.6355 3/11/2019 11:03 Credit card 482.51
## 3      72.61         6      NA 457.4430 1/1/2019 10:39 Credit card 435.66
## 4      89.48        10      NA 939.5400 1/6/2019 12:46 Credit card 894.80
## 5      62.12        10      NA 652.2600 2/11/2019 16:19      Cash 621.20
## 6      48.52         3      NA 152.8380 3/5/2019 18:17      Ewallet 145.56
## 7      87.87        10      NA 922.6350 3/29/2019 10:25      Ewallet 878.70
## 8      57.12         7      NA 419.8320 1/12/2019 12:02 Credit card 399.84
## 9      72.50         8      NA 609.0000 3/16/2019 19:25      Ewallet 580.00
## 10     43.18         8      NA 362.7120 1/19/2019 19:39 Credit card 345.44
## 11     15.34         1      NA  16.1070 1/6/2019 11:09      Cash  15.34
## 12     23.48         2      NA  49.3080 3/14/2019 11:21 Credit card  46.96
## 13     75.82         1      NA  79.6110 1/31/2019 13:19      Cash  75.82
## 14     73.05        10      NA 767.0250 3/3/2019 12:25 Credit card 730.50
##      gross.margin.percentage gross.income Rating
## 1      4.761905      28.1160      4.5
## 2      4.761905      24.1255      4.6
## 3      4.761905      21.7830      6.9
## 4      4.761905      44.7400      9.6
## 5      4.761905      31.0600      5.9
## 6      4.761905       7.2780      4.0
## 7      4.761905      43.9350      5.1
```

```
## 8          4.761905      19.9920      6.5
## 9          4.761905      29.0000      9.2
## 10         4.761905      17.2720      8.3
## 11         4.761905       0.7670      6.5
## 12         4.761905       2.3480      7.9
## 13         4.761905       3.7910      5.8
## 14         4.761905      36.5250      8.7
```

14 records are NA. So, fill the NA as

Total - (Unit price * Quantity)

```
sales<- sales %>%
  mutate(Tax.5. = ifelse(is.na(Tax.5.),Total - (Unit.price * Quantity),Tax.5.))
```

```
colSums(is.na(sales))
```

```
##          Invoice.ID          Branch          City
##              0              0              0
##      Customer.type          Gender          Product.line
##              0              0              0
##          Unit.price          Quantity          Tax.5.
##              0              0              0
##              Total          Date          Time
##              14              0              0
##          Payment          cogs gross.margin.percentage
##              0              0              29
##      gross.income          Rating
##              0              0
```

For Total

```
sales %>%
  filter(!complete.cases(Total))
```

```
##      Invoice.ID Branch      City Customer.type Gender          Product.line
## 1  145-94-9061    B  Mandalay      Normal Female      Food and beverages
## 2  162-48-8011    A   Yangon      Member Female      Food and beverages
## 3  551-21-3069    C Naypyitaw      Normal Female      Electronic accessories
## 4  568-90-5112    C Naypyitaw      Normal   Male      Health and beauty
## 5  316-55-4634    B  Mandalay      Member   Male      Food and beverages
## 6              C Naypyitaw      Member Female      Fashion accessories
## 7  560-30-5617    B  Mandalay      Normal Female      Sports and travel
## 8  540-11-4336    A   Yangon      Normal   Male      Food and beverages
## 9  154-74-7179    B  Mandalay      Normal   Male      Sports and travel
## 10 283-79-9594    B  Mandalay      Normal Female      Food and beverages
## 11 174-36-3675    C Naypyitaw      Member   Male      Food and beverages
## 12 301-81-8610    B  Mandalay      Member   Male      Fashion accessories
## 13 746-19-0921    C Naypyitaw      Normal   Male      Food and beverages
## 14 628-90-8624    B  Mandalay      Member   Male      Health and beauty
##      Unit.price Quantity  Tax.5. Total      Date Time      Payment  cogs
## 1         88.36         5 22.0900    NA 1/25/2019 19:48      441.80
```

```
## 2      44.59      5 11.1475      NA 2/10/2019 15:10      Cash 222.95
## 3      23.07      9 10.3815      NA 2/1/2019 11:27      Cash 207.63
## 4      66.14      4 13.2280      NA 3/19/2019 12:46 Credit card 264.56
## 5      80.05      5 20.0125      NA 1/26/2019 12:45 Credit card 400.25
## 6      74.29      1 3.7145      NA 1/13/2019 19:30      Cash 74.29
## 7      24.77      5 6.1925      NA 3/24/2019 18:27      Cash 123.85
## 8      24.94      9 11.2230      NA 1/11/2019 16:49 Credit card 224.46
## 9      54.45      1 2.7225      NA 2/26/2019 19:24      Ewallet 54.45
## 10     48.51      7 16.9785      NA 1/25/2019 13:30 Credit card 339.57
## 11     99.37      2 9.9370      NA 2/14/2019 17:29      Cash 198.74
## 12     25.42      8 10.1680      NA 3/19/2019 19:42 Credit card 203.36
## 13     21.58      1 1.0790      NA 2/9/2019 10:02      Ewallet 21.58
## 14     82.58     10 41.2900      NA 3/14/2019 14:41      Cash 825.80
##      gross.margin.percentage gross.income Rating
## 1      4.761905      22.0900 9.6
## 2      4.761905      11.1475 8.5
## 3      4.761905      10.3815 4.9
## 4      4.761905      13.2280 5.6
## 5      4.761905      20.0125 9.4
## 6      4.761905      3.7145 5.0
## 7      4.761905      6.1925 8.5
## 8      4.761905      11.2230 5.6
## 9      4.761905      2.7225 7.9
## 10     4.761905      16.9785 5.2
## 11     4.761905      9.9370 5.2
## 12     4.761905      10.1680 6.7
## 13     4.761905      1.0790 7.2
## 14     4.761905      41.2900 5.0
```

14 records are NA. So, fill the NA as

(Unit price * Quantity) + 5% Tax

```
sales<- sales %>%
  mutate(Total = ifelse(is.na(Total),(Unit.price * Quantity) + Tax.5.,Total))
```

```
colSums(is.na(sales))
```

```
##      Invoice.ID      Branch      City
##      0      0      0
##      Customer.type      Gender      Product.line
##      0      0      0
##      Unit.price      Quantity      Tax.5.
##      0      0      0
##      Total      Date      Time
##      0      0      0
##      Payment      cogs gross.margin.percentage
##      0      0      29
##      gross.income      Rating
##      0      0
```

For Gross Margin Percentage


```
sales %>%
  filter(!complete.cases(gross.margin.percentage))
```

| ## | Invoice.ID | Branch | City | Customer.type | Gender | Product.line | | |
|-------|-------------|----------|-----------|---------------|-----------|------------------------|-------------|--------|
| ## 1 | 862-29-5914 | C | Naypyitaw | Normal | Female | Sports and travel | | |
| ## 2 | 770-42-8960 | B | Mandalay | Normal | Male | Food and beverages | | |
| ## 3 | 761-49-0439 | B | Mandalay | Member | Female | Electronic accessories | | |
| ## 4 | 490-95-0021 | B | Mandalay | Member | Female | Food and beverages | | |
| ## 5 | 115-38-7388 | C | Naypyitaw | Member | Female | Fashion accessories | | |
| ## 6 | 311-13-6971 | B | Mandalay | Member | Male | Sports and travel | | |
| ## 7 | 291-55-6563 | A | Yangon | Member | Female | Home and lifestyle | | |
| ## 8 | 548-48-3156 | A | Yangon | Member | Female | Food and beverages | | |
| ## 9 | 460-93-5834 | A | Yangon | Normal | Male | Sports and travel | | |
| ## 10 | 118-62-1812 | C | Naypyitaw | Member | Female | Home and lifestyle | | |
| ## 11 | 450-42-3339 | C | Naypyitaw | Normal | Male | Health and beauty | | |
| ## 12 | 851-98-3555 | B | Mandalay | Normal | Female | Health and beauty | | |
| ## 13 | 186-71-5196 | A | Yangon | Member | Female | Food and beverages | | |
| ## 14 | 624-01-8356 | B | Mandalay | Normal | Female | Home and lifestyle | | |
| ## 15 | 313-66-9943 | B | Mandalay | Member | Female | Food and beverages | | |
| ## 16 | 151-27-8496 | C | Naypyitaw | Normal | Female | Electronic accessories | | |
| ## 17 | 453-33-6436 | A | Yangon | Normal | Female | Home and lifestyle | | |
| ## 18 | 522-57-8364 | A | Yangon | Member | Male | Fashion accessories | | |
| ## 19 | 459-45-2396 | A | Yangon | Member | Female | Food and beverages | | |
| ## 20 | 717-96-4189 | C | Naypyitaw | Normal | Female | Electronic accessories | | |
| ## 21 | 722-13-2115 | C | Naypyitaw | Member | Male | Sports and travel | | |
| ## 22 | 749-81-8133 | A | Yangon | Normal | Female | Fashion accessories | | |
| ## 23 | 690-01-6631 | B | Mandalay | Normal | Male | Fashion accessories | | |
| ## 24 | 652-49-6720 | C | Naypyitaw | Member | Female | Electronic accessories | | |
| ## 25 | 233-67-5758 | C | Naypyitaw | Normal | Male | Health and beauty | | |
| ## 26 | 303-96-2227 | B | Mandalay | Normal | Female | Home and lifestyle | | |
| ## 27 | 727-02-1313 | A | Yangon | Member | Male | Food and beverages | | |
| ## 28 | 347-56-2442 | A | Yangon | Normal | Male | Home and lifestyle | | |
| ## 29 | 849-09-3807 | A | Yangon | Member | Female | Fashion accessories | | |
| ## | Unit.price | Quantity | Tax.5. | Total | Date | Time | Payment | cogs |
| ## 1 | 22.38 | 1 | 1.1190 | 23.4990 | 1/30/2019 | 17:08 | Credit card | 22.38 |
| ## 2 | 21.12 | 8 | 8.4480 | 177.4080 | 1/1/2019 | 19:31 | Cash | 168.96 |
| ## 3 | 12.10 | 8 | 4.8400 | 101.6400 | 1/19/2019 | 10:17 | Ewallet | 96.80 |
| ## 4 | 33.21 | 10 | 16.6050 | 348.7050 | 1/8/2019 | 14:25 | Ewallet | 332.10 |
| ## 5 | 10.18 | 8 | 4.0720 | 85.5120 | 3/30/2019 | 12:51 | Credit card | 81.44 |
| ## 6 | 31.99 | 10 | 15.9950 | 335.8950 | 2/20/2019 | 15:18 | Credit card | 319.90 |
| ## 7 | 34.42 | 6 | 10.3260 | 216.8460 | 3/30/2019 | 12:45 | Ewallet | 206.52 |
| ## 8 | 83.34 | 2 | 8.3340 | 175.0140 | 3/19/2019 | 13:37 | Cash | 166.68 |
| ## 9 | 45.58 | 7 | 15.9530 | 335.0130 | 1/13/2019 | 10:03 | Cash | 319.06 |
| ## 10 | 78.38 | 4 | 15.6760 | 329.1960 | 3/24/2019 | 17:56 | Cash | 313.52 |
| ## 11 | 84.61 | 10 | 42.3050 | 888.4050 | 2/9/2019 | 18:58 | Credit card | 846.10 |
| ## 12 | 82.88 | 5 | 20.7200 | 435.1200 | 3/24/2019 | 14:08 | Credit card | 414.40 |
| ## 13 | 79.54 | 2 | 7.9540 | 167.0340 | 3/27/2019 | 16:30 | Ewallet | 159.08 |
| ## 14 | 49.01 | 10 | 24.5050 | 514.6050 | 1/27/2019 | 10:44 | Credit card | 490.10 |
| ## 15 | 29.15 | 3 | 4.3725 | 91.8225 | 3/27/2019 | 20:29 | Credit card | 87.45 |
| ## 16 | 56.13 | 4 | 11.2260 | 235.7460 | 1/19/2019 | 11:43 | Ewallet | 224.52 |
| ## 17 | 93.12 | 8 | 37.2480 | 782.2080 | 2/7/2019 | 10:09 | Cash | 744.96 |
| ## 18 | 51.34 | 8 | 20.5360 | 431.2560 | 1/31/2019 | 10:00 | Ewallet | 410.72 |
| ## 19 | 99.60 | 3 | 14.9400 | 313.7400 | 2/25/2019 | 18:45 | Cash | 298.80 |

```
## 20      35.49      6 10.6470 223.5870 2/2/2019 12:40      Cash 212.94
## 21      42.85      1  2.1425  44.9925 3/14/2019 15:36 Credit card 42.85
## 22      94.67      4 18.9340 397.6140 3/11/2019 12:04      Cash 378.68
## 23      17.49     10  8.7450 183.6450 2/22/2019 18:35      Ewallet 174.90
## 24      60.95      1  3.0475  63.9975 2/18/2019 11:40      Ewallet  60.95
## 25      40.35      1  2.0175  42.3675 1/29/2019 13:46      Ewallet  40.35
## 26      97.38     10 48.6900 1022.4900 3/2/2019 17:16      Ewallet 973.80
## 27      31.84      1  1.5920  33.4320 2/9/2019      Cash    31.84
## 28      65.82      1  3.2910  69.1110 2/22/2019 15:33      Cash    65.82
## 29      88.34      7 30.9190 649.2990 2/18/2019 13:28      Cash   618.38
##      gross.margin.percentage gross.income Rating
## 1              NA      1.1190      8.6
## 2              NA      8.4480      6.3
## 3              NA      4.8400      8.6
## 4              NA     16.6050      6.0
## 5              NA      4.0720      9.5
## 6              NA     15.9950      9.9
## 7              NA     10.3260      7.5
## 8              NA      8.3340      7.6
## 9              NA     15.9530      5.0
## 10             NA     15.6760      7.9
## 11             NA     42.3050      8.8
## 12             NA     20.7200      6.6
## 13             NA      7.9540      6.2
## 14             NA     24.5050      4.2
## 15             NA      4.3725      7.3
## 16             NA     11.2260      8.6
## 17             NA     37.2480      6.8
## 18             NA     20.5360      7.6
## 19             NA     14.9400      5.8
## 20             NA     10.6470      4.1
## 21             NA      2.1425      9.3
## 22             NA     18.9340      6.8
## 23             NA      8.7450      6.6
## 24             NA      3.0475      5.9
## 25             NA      2.0175      6.2
## 26             NA     48.6900      4.4
## 27             NA      1.5920      7.7
## 28             NA      3.2910      4.1
## 29             NA     30.9190      6.6
```

```
table(sales$gross.margin.percentage)
```

```
##
## 4.761904762
##      971
```

Here, Gross Margin Percentage is 4.761904762, So, fill NA as 4.761904762

```
sales<- sales %>%
  mutate(gross.margin.percentage = ifelse(is.na(gross.margin.percentage),4.761904762,gross.margin.percentage))
```

```
colSums(is.na(sales))
```

```
##          Invoice.ID          Branch          City
##              0              0              0
##    Customer.type    Gender    Product.line
##              0              0              0
##      Unit.price    Quantity    Tax.5.
##              0              0              0
##      Total      Date      Time
##              0              0              0
##      Payment    cogs gross.margin.percentage
##              0              0              0
##    gross.income    Rating
##              0              0
```

EDA

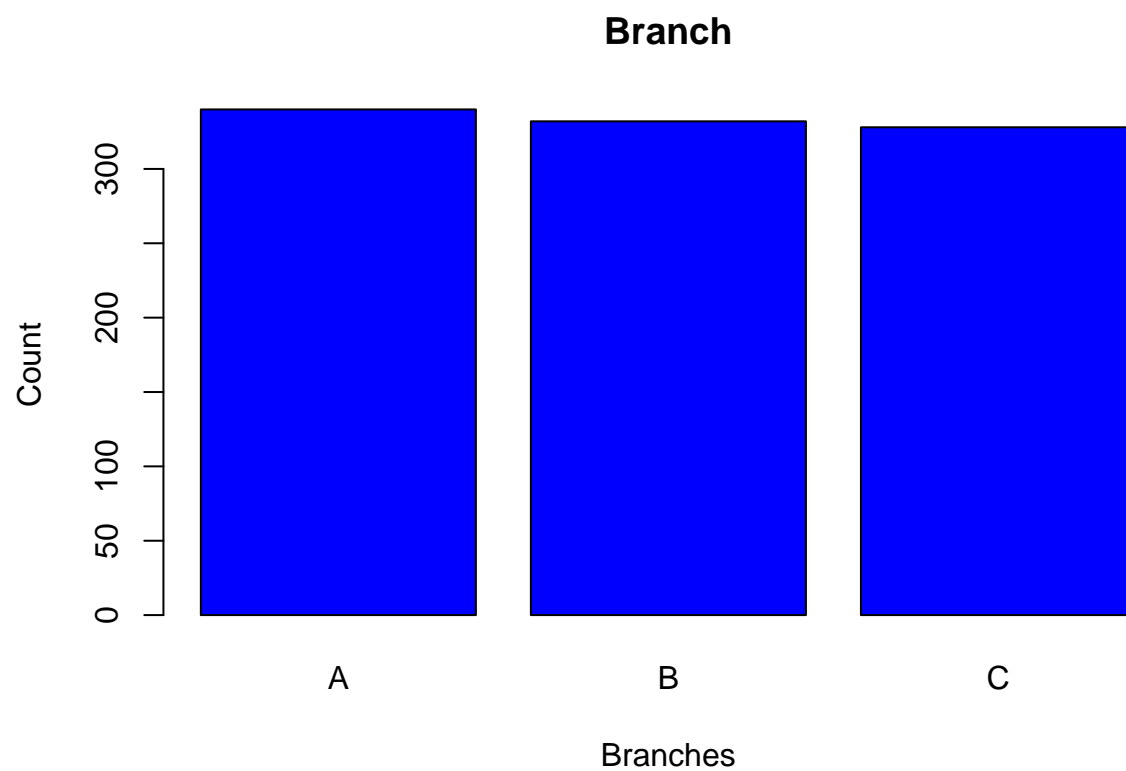
- Univariate Analysis: Studying one variable at a time
- Bivariate Analysis: Studying two variables at a time
- Multivariate Analysis: Studying multiple variables at a time

Univariate Analysis

```
branches <- table(sales$Branch)
branches
```

```
##
##  A  B  C
## 340 332 328
```

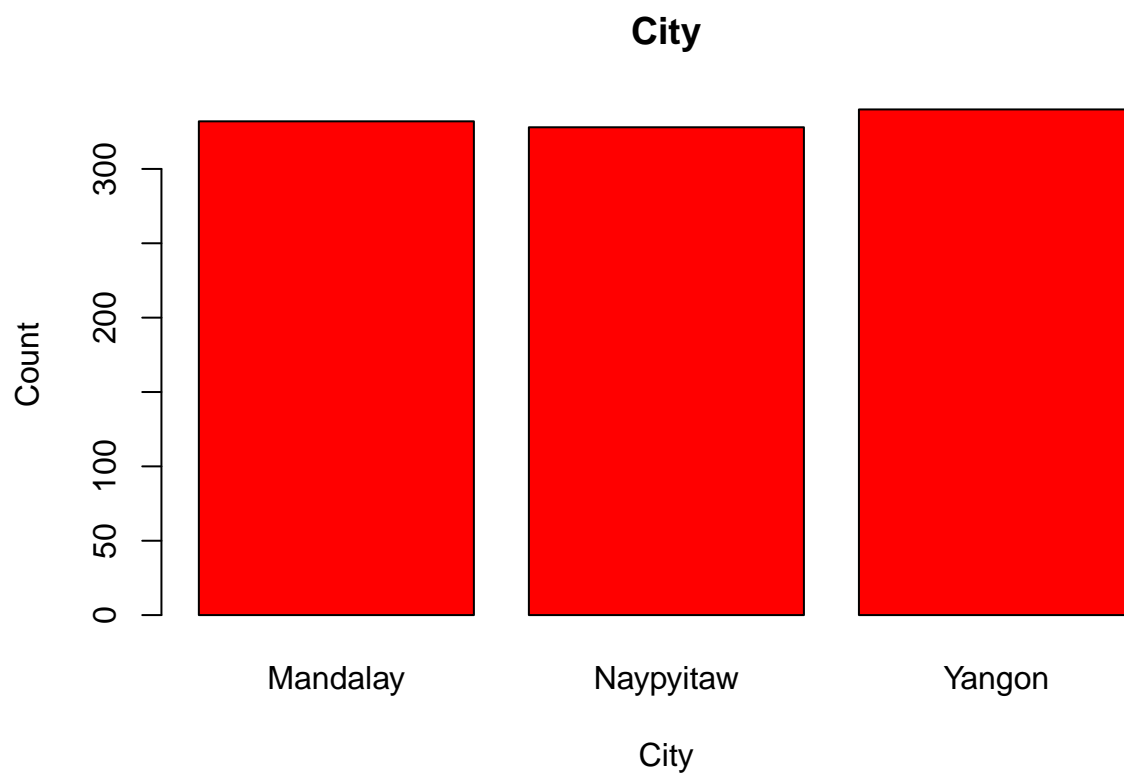
```
barplot(branches,
        main = "Branch",
        xlab = "Branches",
        ylab = "Count",
        col = "blue")
```



```
cities <- table(sales$City)
cities
```

```
##
## Mandalay Naypyitaw Yangon
##      332      328      340
```

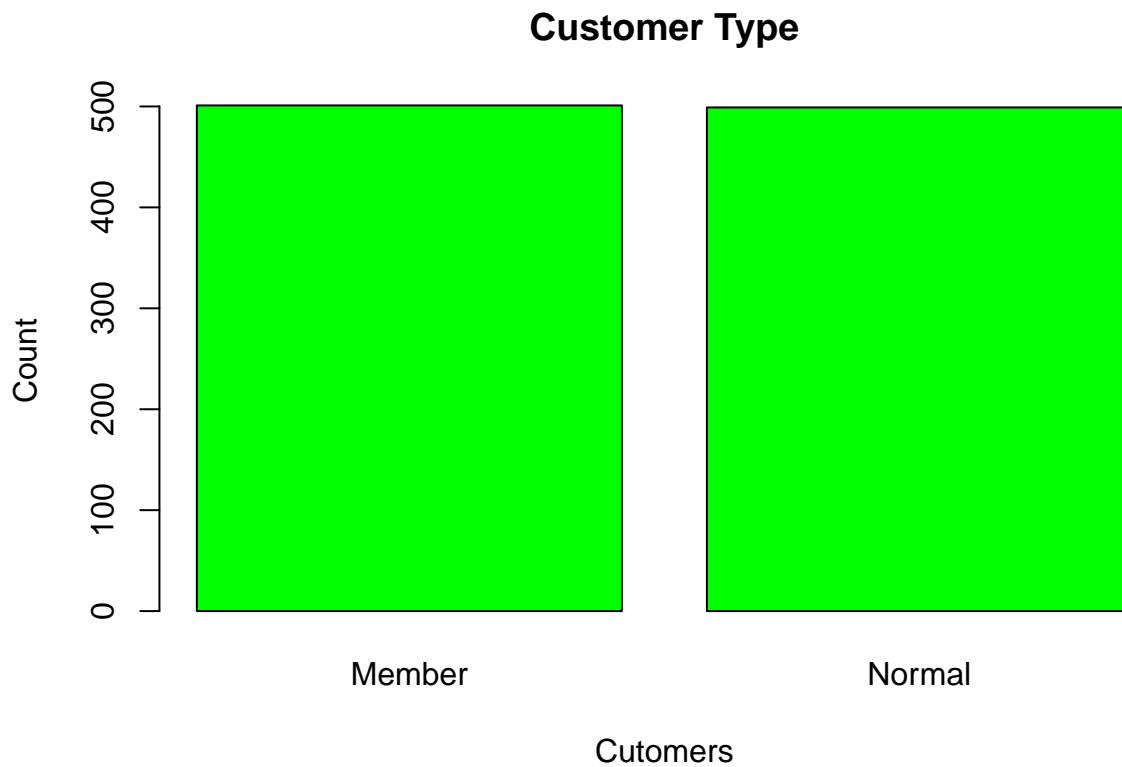
```
barplot(cities,
  main = "City",
  xlab = "City",
  ylab = "Count",
  col = "red") # Bar color
```



```
cust_type <- table(sales$Customer.type)
cust_type
```

```
##
## Member Normal
##    501    499
```

```
barplot(cust_type,
        main = "Customer Type",
        xlab = "Cutomers",
        ylab = "Count",
        col = "green")
```

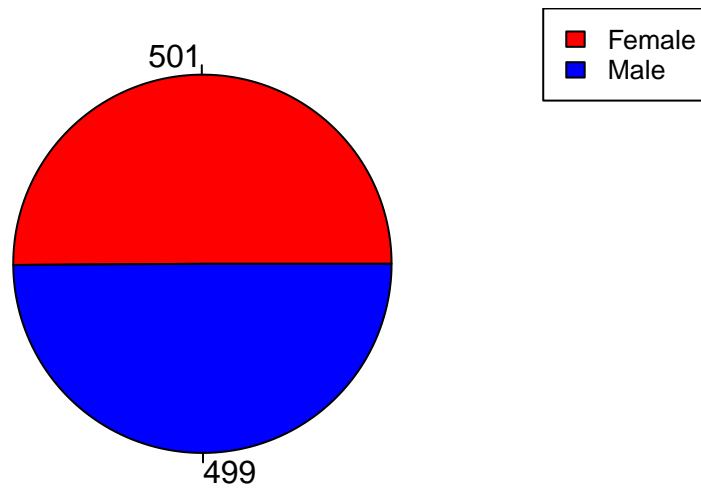


```
gender_type <- table(sales$Gender)
gender_type
```

```
##
## Female    Male
##     501     499
```

```
gender_lbl <- paste(gender_type)
pie(gender_type,
    main = "Gender",
    col = c("red", "blue"),
    label = gender_lbl)
legend("topright", names(gender_type),
    cex = 0.8, fill = c("red", "blue"))
```

Gender

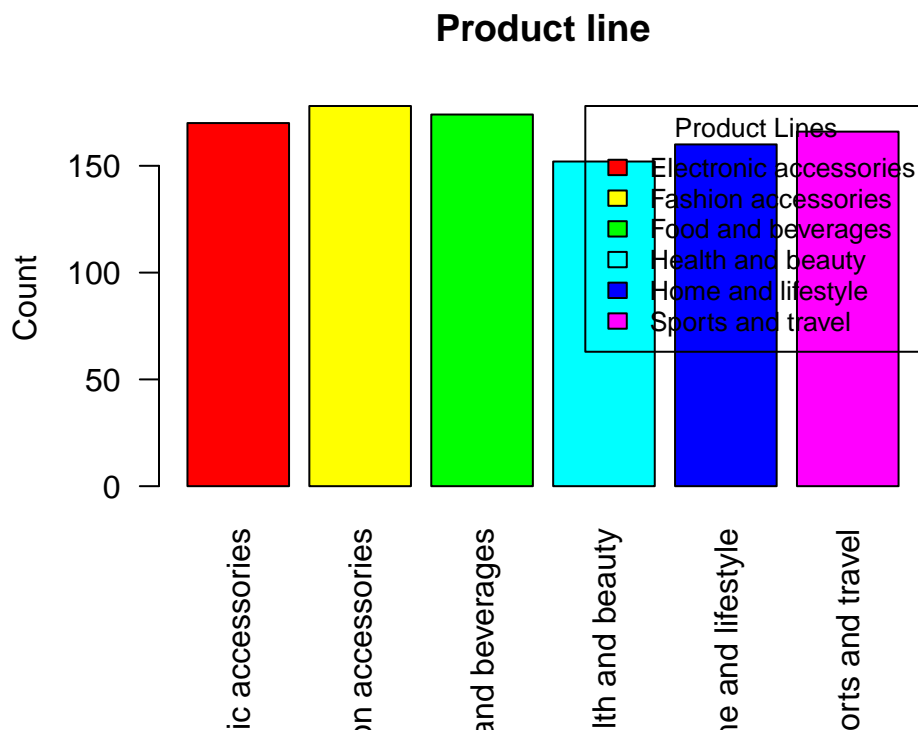


```
product_line <- table(sales$Product.line)
product_line
```

```
##
## Electronic accessories    Fashion accessories    Food and beverages
##           170             178             174
##   Health and beauty    Home and lifestyle    Sports and travel
##           152             160             166
```

```
par(pin = c(4, 2))
barplot(product_line,
        main = "Product line",
        ylab = "Count",
        las= 2,
        col = rainbow(6))

legend("topright",
      legend = names(product_line),
      fill = rainbow(6),
      title = "Product Lines",
      cex = 0.8,)
```



```
payment_type <- table(sales$Payment)
payment_type
```

```
##
##           Cash Credit card      Ewallet
##           3           341           311           345
```

3 records are not categorized under any category. SO, fill with mode categorical value.

```
mode_cat <- names(payment_type)[which.max(payment_type)]
mode_cat
```

```
## [1] "Ewallet"
```

Replace the null value with "Ewallet."

```
sales$Payment[sales$Payment == ""] <- mode_cat
```

```
payment_type <- table(sales$Payment)
payment_type
```

```
##
##           Cash Credit card      Ewallet
##           341           311           348
```

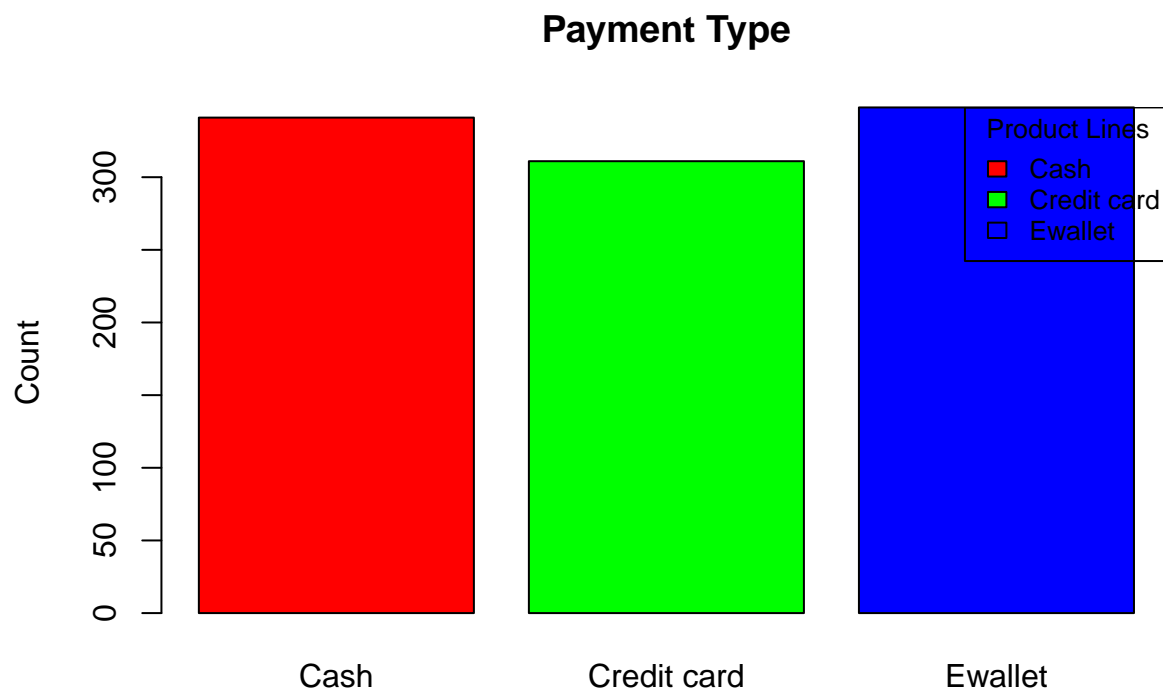


```

barplot(payment_type,
        main = "Payment Type",
        ylab = "Count",
        col = rainbow(3))

legend("topright",
      legend = names(payment_type),
      fill = rainbow(3),
      title = "Product Lines",
      cex = 0.8)

```



For Invoice ID

```
length(unique(sales$Invoice.ID))
```

```
## [1] 991
```

9 out of 1000 record have missing Invoice ID.