Project Report

Contents

Team members

- Ahmed Ashraf Mohamed
  - ID: 2022446758
- Abdelrhman Mohamed Abdelhady Hodid
  - ID : 2022513643
- Antonuose Gerges Nageh
  - ID : 20221903971

Project Description

Getting and preparing the data set

importing the library we are going to use

```
library(dplyr)
library(ggplot2)
library(forcats)
library(arules)
library(hrbrthemes)
```

Reading the data and checking the frist 10 rows of it

```
dataPath <- readline("Enter the path to the data set : ")
grc <- as_tibble(read.csv(dataPath,stringsAsFactors = FALSE))
# displaying first 10 rows of our data
print(grc,n = 10, width = 80)
```

```
## # A tibble: 9,835 x 8
##    items                      count total   rnd customer   age city   paymentType
##    <chr>                      <int> <int> <int> <chr>    <int> <chr>  <chr>
##  1 citrus fruit,semi-finish~     4  1612     9 Maged       60 Hurgh~ Cash
##  2 tropical fruit,yogurt,co~     3   509    12 Eman        23 Aswan  Cash
##  3 whole milk                    1  2084     8 Rania       37 Dakah~ Cash
##  4 pip fruit,yogurt,cream c~     4   788     8 Rania       37 Dakah~ Cash
##  5 other vegetables,whole m~     4  1182    14 Magdy       36 Sohag  Cash
##  6 whole milk,butter,yogurt~     5  1771     3 Ahmed       30 Giza   Credit
##  7 rolls/buns                    1  2196     7 Huda        39 Gharb~ Cash
##  8 other vegetables,UHT-mil~     5  1657     6 Walaa       29 Cairo  Cash
##  9 pot plants                    1  2373     2 Mohamed     25 Alexa~ Credit
## 10 whole milk,cereals            2   343     5 Shimaa      55 Port ~ Cash
## # ... with 9,825 more rows
```

Preparing the data for k-means and apriori algorithms

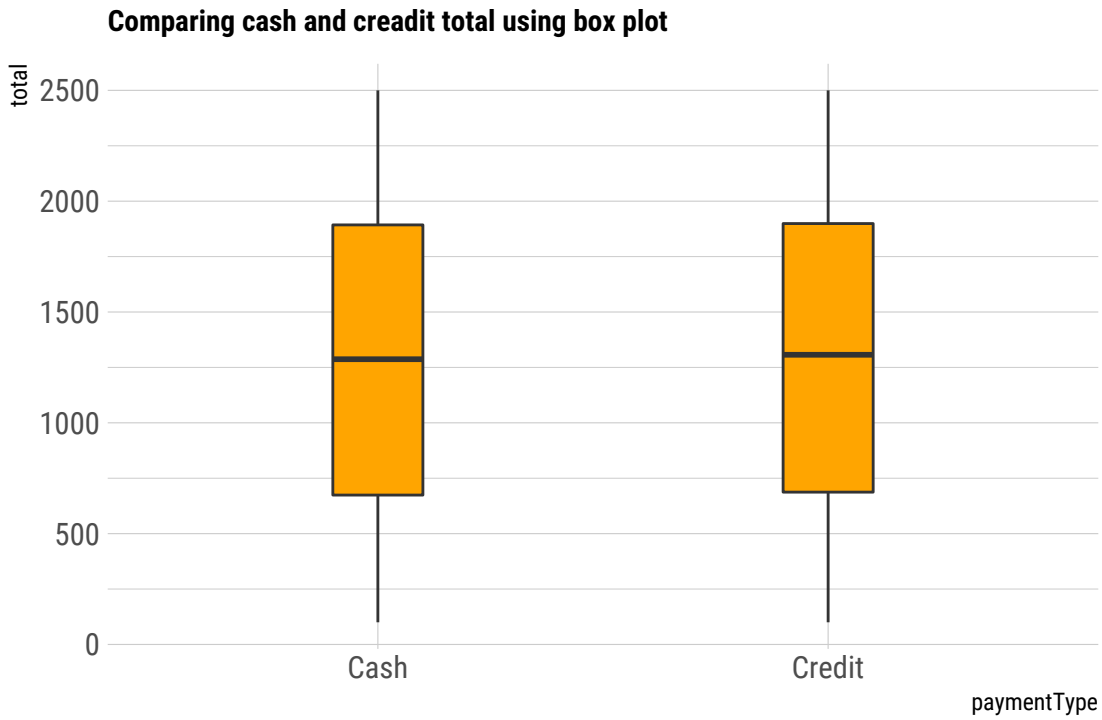
```
# creating a data framing containing customers, their ages and total spending
grc_customers <- grc %>%
  select(customer,age,total) %>%
  group_by(customer)%>%
  mutate(total = sum(total))%>%
  unique()
# making the data frame suitable for k-means
grc_k <- data.frame(grc_customers[,2:3],row.names = grc_customers$customer)
# splitting the items to suitable for apriori algorithms
tdata <- strsplit(as.vector(grc$items), ',')
tdata <- transactions(tdata)
```

Visualizing our Data

Comparison between cash and creadit total spending using box plot

```
boxplot_cashCreadit <- ggplot(
  grc,
  aes(x = paymentType, y = total, fill = paymentType)
) +
  geom_boxplot(width = .2,
               fill = "orange",
               outlier.color = "orange",
               outlier.size = 2)+
  theme_ipsum_rc() +
  theme(
    legend.position="none",
    plot.title = element_text(size=11)
  ) +
```

```
  ggtitle("Comparing cash and creadit total using box plot")
boxplot_cashCreadit
```



Observations

After brief moments of seeing the figure, it is quite easy to see that people nearly equally pay with Cash or credit money.

Compare each age and sum of total spending.

Before visualizing

let's create a contingency table using the data prepared and look at a table containing customers, their ages and their total individual spending

```
table(grc_customers$age)
```

```
##
## 22 23 25 29 30 35 36 37 39 50 55 60
##  2  1  1  1  1  1  1  2  1  1  2  1
```

```
arrange(grc_customers,desc(total))
```

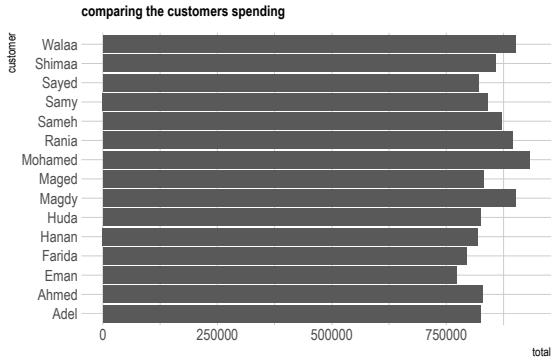
```
## # A tibble: 15 x 3
## # Groups:   customer [15]
##    customer   age  total
##    <chr>    <int>  <int>
##  1 Mohamed     25 932250
```

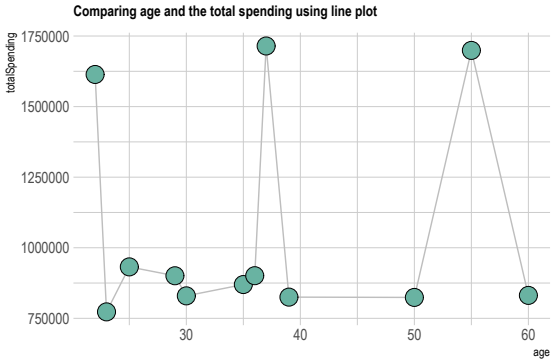
```
##  2 Magdy       36 901010
##  3 Walaa       29 900797
##  4 Rania       37 893789
##  5 Sameh       35 869668
##  6 Shimaa      55 857901
##  7 Samy        55 841167
##  8 Maged       60 831272
##  9 Ahmed       30 829587
## 10 Huda        39 825147
## 11 Adel        50 824064
## 12 Sayed       37 820900
## 13 Hanan       22 819231
## 14 Farida      22 794570
## 15 Eman        23 772871
```

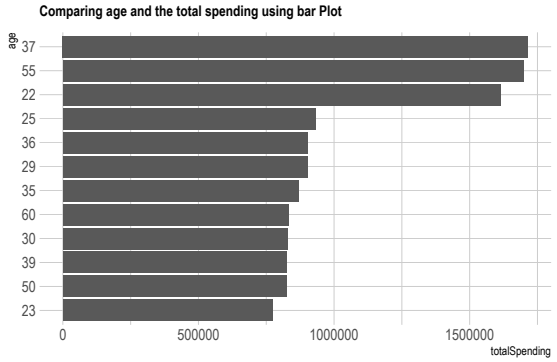
It becomes clear that there are more people aged (22,37 and 55) than the other age groups

```
customersBarPlot<- grc_customers %>%
  ggplot(
    aes(x = customer, y = total)
  ) +
  geom_col() +
  coord_flip()+
  theme_ipsum() +
  theme(
    plot.title = element_text(size=11)
  )+
  ggtitle("comparing the customers spending")
grc_age <- select(grc,age,total)
grc_age <- grc_age %>%
  group_by(age) %>%
  summarise(totalSpending = sum(total))
linePlotAgeSum <- ggplot(
  grc_age,
  aes(x = age, y = totalSpending)) +
  geom_line( color="gray") +
  geom_point(shape=21, color="black", fill="#69b3a2", size=6) +
  theme_ipsum() +
  theme(
    plot.title = element_text(size=11)
  )+
  ggtitle("Comparing age and the total spending using line plot")
grc_age <- mutate(grc_age,age = fct_reorder(as.factor(age),totalSpending))
barPlotAgeSum<-ggplot(
  grc_age,
  aes(x = age, y = totalSpending)) +
  geom_col() +
  coord_flip() +
  theme_ipsum() +
  theme(
    plot.title = element_text(size=11)
  )+
  ggtitle("Comparing age and the total spending using bar Plot")
customersBarPlot
linePlotAgeSum
```

```
barPlotAgeSum
```



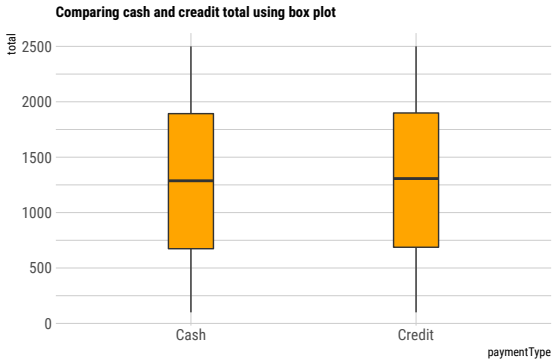


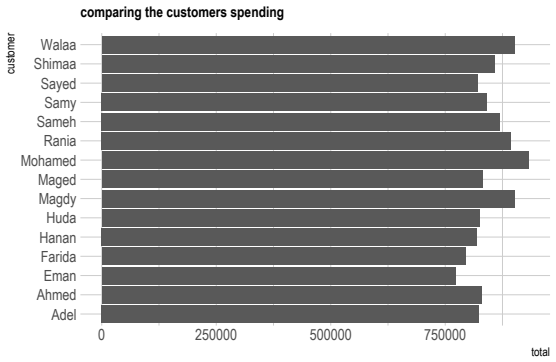


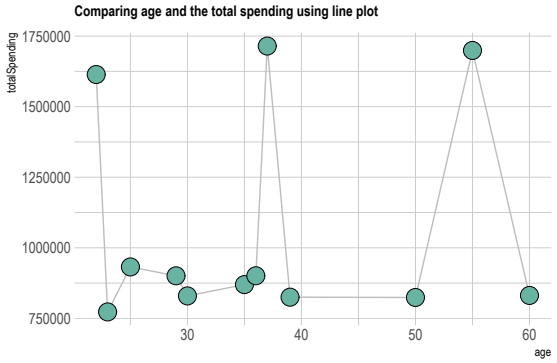
Observations

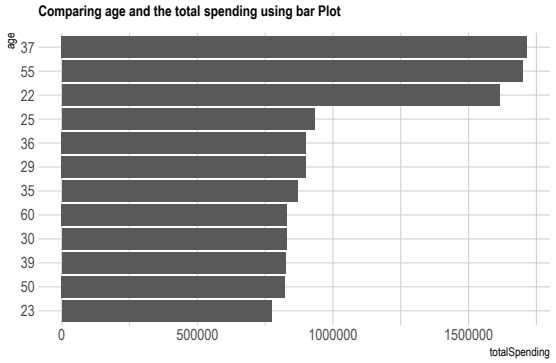
Since there are more customers aged (22, 37 and 55), it makes sense that the total spending of these groups is higher than the rest. But, looking at the data again we can see that if we look at the customers and their individual spending, we find that customers aged between 20 and 30 -even though there are single individual of these groups- are the highest spending.

Dashboard









Generating of association rules

Brief explanition of Apriori algorithm for generating the rules

Apriori algorithm is an iterative approach for discovering the most frequent item sets. The frequent item sets generated by the algorithm can be used to determine association rules that highlight general trends in the data-set, it is especially useful in the analysis of super-market items in our data set

Implementing the alogrithm

Reading both minimum support and minimum confidance from the user

```
min_support <- as.numeric(readline("Enter the minimum Support : "))
min_conf <- as.numeric(readline("Enter the minimum Confidance : "))
```

implementing the algorithm using the built-in function

```
apriori_rules <- apriori(
  tdata,
  parameter = list(supp = min_support, conf = min_conf, minlen = 2))
```

```
## Apriori
##
## Parameter specification:
```

```
##  confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target  ext
##         0.5    0.1    1 none FALSE            TRUE       5   0.001      2     10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.01s].
## writing ... [5668 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
# Displaying at most 100 rows of the rules
as_tibble(DATAFRAME(apriori_rules,separate = TRUE, setStart = "", setEnd = "")) %>%
  print(n = 100, width = 90)
```

```
## # A tibble: 5,668 x 7
##     LHS                                RHS           support confidence coverage  lift count
##     <fct>                              <fct>           <dbl>      <dbl>    <dbl> <dbl> <int>
##   1 honey                              whole milk    0.00112      0.733  0.00153  2.87    11
##   2 tidbits                            rolls/buns    0.00122      0.522  0.00234  2.84    12
##   3 cocoa drinks                       whole milk    0.00132      0.591  0.00224  2.31    13
##   4 pudding powder                     whole milk    0.00132      0.565  0.00234  2.21    13
##   5 cooking chocolate                  whole milk    0.00132      0.52   0.00254  2.04    13
##   6 cereals                            whole milk    0.00366      0.643  0.00569  2.52    36
##   7 jam                                whole milk    0.00295      0.547  0.00539  2.14    29
##   8 specialty cheese                   other vegeta~ 0.00427      0.5    0.00854  2.58    42
##   9 rice                               other vegeta~ 0.00397      0.52   0.00763  2.69    39
##  10 rice                               whole milk    0.00468      0.613  0.00763  2.40    46
##  11 baking powder                      whole milk    0.00925      0.523  0.0177   2.05    91
##  12 liver loaf,yogurt                  whole milk    0.00102      0.667  0.00153  2.61    10
##  13 curd cheese,tropical fruit         other vegeta~ 0.00102      0.667  0.00153  3.45    10
##  14 curd cheese,rolls/buns             whole milk    0.00102      0.625  0.00163  2.45    10
##  15 curd cheese,other vegetables       whole milk    0.00122      0.571  0.00214  2.24    12
##  16 curd cheese,whole milk             other vegeta~ 0.00122      0.522  0.00234  2.70    12
##  17 cleaner,other vegetables           whole milk    0.00102      0.625  0.00163  2.45    10
##  18 liquor,red/blush wine              bottled beer  0.00193      0.905  0.00214 11.2     19
##  19 liquor,soda                        bottled beer  0.00122      0.571  0.00214  7.10    12
##  20 cereals,curd                       whole milk    0.00102      0.909  0.00112  3.56    10
##  21 cereals,root vegetables            whole milk    0.00102      0.769  0.00132  3.01    10
##  22 cereals,yogurt                     whole milk    0.00173      0.810  0.00214  3.17    17
##  23 cereals,other vegetables           whole milk    0.00132      0.65   0.00203  2.54    13
##  24 butter,jam                         whole milk    0.00102      0.833  0.00122  3.26    10
##  25 jam,root vegetables                other vegeta~ 0.00102      0.526  0.00193  2.72    10
##  26 jam,other vegetables               root vegetab~ 0.00102      0.556  0.00183  5.10    10
##  27 jam,root vegetables                whole milk    0.00132      0.684  0.00193  2.68    13
##  28 jam,other vegetables               whole milk    0.00132      0.722  0.00183  2.83    13
##  29 instant coffee,whipped/sour cre~ other vegeta~ 0.00102      0.769  0.00132  3.98    10
##  30 instant coffee,other vegetables  whipped/sour~ 0.00102      0.526  0.00193  7.34    10
##  31 instant coffee,other vegetables  soda          0.00102      0.526  0.00193  3.02    10
```

```
##  32 instant coffee,other vegetables  whole milk    0.00102      0.526  0.00193  2.06    10
##  33 vinegar,yogurt                   other vegeta~ 0.00102      0.556  0.00183  2.87    10
##  34 other vegetables,vinegar         whole milk    0.00132      0.542  0.00244  2.12    13
##  35 vinegar,whole milk               other vegeta~ 0.00132      0.5    0.00264  2.58    13
##  36 popcorn,salty snack              soda          0.00122      0.545  0.00224  3.13    12
##  37 popcorn,soda                     salty snack   0.00122      0.632  0.00193 16.7     12
##  38 candles,tropical fruit           whole milk    0.00102      0.667  0.00153  2.61    10
##  39 candles,root vegetables          other vegeta~ 0.00102      0.588  0.00173  3.04    10
##  40 bottled beer,soups               whole milk    0.00112      0.917  0.00122  3.59    11
##  41 soups,whipped/sour cream         other vegeta~ 0.00102      0.667  0.00153  3.45    10
##  42 pip fruit,soups                  other vegeta~ 0.00102      0.714  0.00142  3.69    10
##  43 root vegetables,soups            other vegeta~ 0.00122      0.706  0.00173  3.65    12
##  44 root vegetables,soups            whole milk    0.00112      0.647  0.00173  2.53    11
##  45 other vegetables,soups           whole milk    0.00183      0.581  0.00315  2.27    18
##  46 soups,whole milk                 other vegeta~ 0.00183      0.621  0.00295  3.21    18
##  47 dog food,tropical fruit          yogurt        0.00102      0.625  0.00163  4.48    10
##  48 dog food,yogurt                  tropical fru~ 0.00102      0.526  0.00193  5.02    10
##  49 dog food,tropical fruit          whole milk    0.00102      0.625  0.00163  2.45    10
##  50 dog food,root vegetables         whole milk    0.00102      0.714  0.00142  2.80    10
##  51 dog food,yogurt                  whole milk    0.00122      0.632  0.00193  2.47    12
##  52 dog food,other vegetables        whole milk    0.00112      0.5    0.00224  1.96    11
##  53 Instant food products,soda       hamburger me~ 0.00122      0.632  0.00193 19.0     12
##  54 hamburger meat,Instant food pro~ whole milk    0.00153      0.5    0.00305  1.96    15
##  55 Instant food products,whole milk hamburger me~ 0.00153      0.5    0.00305 15.0     15
##  56 Instant food products,root vege~ other vegeta~ 0.00102      0.526  0.00193  2.72    10
##  57 Instant food products,yogurt     whole milk    0.00112      0.786  0.00142  3.08    11
##  58 Instant food products,other veg~ whole milk    0.00153      0.556  0.00275  2.17    15
##  59 Instant food products,whole milk other vegeta~ 0.00153      0.5    0.00305  2.58    15
##  60 specialty cheese,whipped/sour c~ other vegeta~ 0.00102      0.714  0.00142  3.69    10
##  61 citrus fruit,specialty cheese    whole milk    0.00102      0.714  0.00142  2.80    10
##  62 bottled water,specialty cheese   other vegeta~ 0.00102      0.588  0.00173  3.04    10
##  63 root vegetables,specialty cheese yogurt        0.00112      0.524  0.00214  3.75    11
##  64 root vegetables,specialty cheese other vegeta~ 0.00142      0.667  0.00214  3.45    14
##  65 root vegetables,specialty cheese whole milk    0.00132      0.619  0.00214  2.42    13
##  66 specialty cheese,yogurt          other vegeta~ 0.00163      0.571  0.00285  2.95    16
##  67 specialty cheese,yogurt          whole milk    0.00203      0.714  0.00285  2.80    20
##  68 specialty cheese,whole milk      yogurt        0.00203      0.541  0.00376  3.87    20
##  69 other vegetables,specialty chee~ whole milk    0.00224      0.524  0.00427  2.05    22
##  70 specialty cheese,whole milk      other vegeta~ 0.00224      0.595  0.00376  3.07    22
##  71 butter,flower (seeds)            whole milk    0.00102      0.625  0.00163  2.45    10
##  72 flower (seeds),root vegetables   other vegeta~ 0.00112      0.579  0.00193  2.99    11
##  73 flower (seeds),root vegetables   whole milk    0.00102      0.526  0.00193  2.06    10
##  74 frozen potato products,fruit/ve~ whole milk    0.00112      0.688  0.00163  2.69    11
##  75 frozen potato products,yogurt    whole milk    0.00153      0.682  0.00224  2.67    15
##  76 frozen potato products,rolls/bu~ whole milk    0.00112      0.55   0.00203  2.15    11
##  77 frozen potato products,other ve~ whole milk    0.00183      0.692  0.00264  2.71    18
##  78 frozen potato products,whole mi~ other vegeta~ 0.00183      0.529  0.00346  2.74    18
##  79 house keeping products,napkins   whole milk    0.00132      0.812  0.00163  3.18    13
##  80 house keeping products,whipped/~ whole milk    0.00122      0.923  0.00132  3.61    12
##  81 house keeping products,sausage   whole milk    0.00112      0.688  0.00163  2.69    11
##  82 house keeping products,tropical~ whole milk    0.00112      0.688  0.00163  2.69    11
##  83 house keeping products,root veg~ whole milk    0.00132      0.765  0.00173  2.99    13
##  84 house keeping products,yogurt    other vegeta~ 0.00102      0.556  0.00183  2.87    10
##  85 house keeping products,yogurt    whole milk    0.00102      0.556  0.00183  2.17    10
```

```
##  86 house keeping products,rolls/bu~ whole milk    0.00112      0.647  0.00173  2.53    11
##  87 house keeping products,other ve~ whole milk    0.00173      0.630  0.00275  2.46    17
##  88 sweet spreads,white bread        whole milk    0.00102      0.667  0.00153  2.61    10
##  89 pastry,sweet spreads             whole milk    0.00102      0.909  0.00112  3.56    10
##  90 root vegetables,sweet spreads    whole milk    0.00122      0.6    0.00203  2.35    12
##  91 sweet spreads,yogurt             soda          0.00122      0.545  0.00224  3.13    12
##  92 soda,sweet spreads               whole milk    0.00142      0.56   0.00254  2.19    14
##  93 sweet spreads,yogurt             whole milk    0.00122      0.545  0.00224  2.13    12
##  94 curd,turkey                      other vegeta~ 0.00122      0.8    0.00153  4.13    12
##  95 curd,turkey                      whole milk    0.00102      0.667  0.00153  2.61    10
##  96 butter,turkey                    whole milk    0.00102      0.667  0.00153  2.61    10
##  97 turkey,whipped/sour cream        other vegeta~ 0.00102      0.769  0.00132  3.98    10
##  98 pastry,turkey                    whole milk    0.00112      0.579  0.00193  2.27    11
##  99 bottled water,turkey             whole milk    0.00122      0.75   0.00163  2.94    12
## 100 tropical fruit,turkey            root vegetab~ 0.00153      0.577  0.00264  5.29    15
## # ... with 5,568 more rows
```