RFM analysis for an Electronics webshop

Brief RFM description

Recency-Frequency-Monetary distribution analysis
- `Recency` - Nr of days since last order
- `Frequency` - Nr of sales
- `Monetary` - Average value of sales

The analysis is done based on browserid. K-means algorith is used for clustering.

Distribution of browserids based on frequency, recency and monetary values

```
csvfilepath <- paste("../input/csv", "/",  "rfm.csv", sep = "")
d_rfm <- read.csv(file=csvfilepath, header=TRUE, sep=",")
```

```
segments <- c("Hibernating", "Customers needing attention", "Potential Loyalist", "Loyal customers", "R
```

```
percents <- c(0, 0.25, 0.5, 0.75, 1)
title <- "Ratio of sales under given number of days"
x <- "% of sales under the given number of days"
y <- "Recency"
caption <- "(in the middle - count of sales under the given percentage and max nr of days since last or
binwidth = 0.4
barwidth = 0.8
y_text_position = 100
```

Recency

Standard scales

```
# 2. standard table summary
d_summary(d_rfm, d_rfm$recency)
```

```
##   mean  sd min max IQR median  total
## 1  181 116   1 438 181    152 212618
```

```
quantile(d_rfm$recency)
```

```
##   0%  25%  50%  75% 100%
##    1   88  152  269  438
```

```
# 4. all summaries in table format for standard percentage
d_rfm_r <- d_kpi_feature(percents, d_rfm$recency)
d_rfm_r
```

```
##   percent  count  range range_count value
## 1       0    259      0         259     1
## 2      25  53309   0-25       53050    88
## 3      50 106408  25-50       53099   152
```

```
## 4      75 159848  50-75       53440   269
## 5     100 212618 75-100       52770   438
```

Observed scales

In case of Recency no need of observed scales

```
percents <- c(0, 0.25, 0.5, 0.7, 0.9, 0.95, 0.99, 0.999, 1)

d_rfm_r <-d_kpi_feature(percents, d_rfm$recency)
d_rfm_r
```

```
##   percent  count    range range_count value
## 1     0.0    259        0         259     1
## 2    25.0  53309     0-25       53050    88
## 3    50.0 106408    25-50       53099   152
## 4    70.0 149194    50-70       42786   246
## 5    90.0 191739    70-90       42545   366
## 6    95.0 202429    90-95       10690   395
## 7    99.0 210522    95-99        8093   416
## 8    99.9 212453  99-99.9        1931   435
## 9   100.0 212618 99.9-100         165   438
```

Frequency

```
percents <- c(0, 0.25, 0.5, 0.75, 1)
title <- "Ratio of sales under given number of days"
x <- "Nr of Sales"
y <- "Frequency"
caption <- "(in the middle - )"
binwidth = 0.4
barwidth = 0.8
y_text_position = 4
```

Standard scales

```
# 2. standard table summary
d_summary(d_rfm, d_rfm$frequency)
```

```
##   mean sd min max IQR median  total
## 1    2  2   1 236   1      1 212618
```

```
quantile(d_rfm$frequency)
```

```
##   0%  25%  50%  75% 100%
##    1    1    1    2  236
```

```
# 4. all summaries in table format for standard percentage
d_rfm_f <- d_kpi_feature(percents, d_rfm$frequency)
d_rfm_f
```

```
##   percent  count  range range_count value
## 1       0 149177      0      149177     1
```

```
## 2      25 149177   0-25           0     1
## 3      50 149177  25-50           0     1
## 4      75 186764  50-75       37587     2
## 5     100 212618 75-100       25854   236
```

Observed scales

```
percents <- c(0, 0.9, 0.99, 0.999, 1)
d_rfm_r <-d_kpi_feature(percents, d_rfm$frequency)
d_rfm_r
```

```
##   percent  count    range range_count   value
## 1     0.0 149177        0      149177   1.000
## 2    90.0 199983     0-90       50806   3.000
## 3    99.0 210720    90-99       10737   7.000
## 4    99.9 212405  99-99.9        1685  18.383
## 5   100.0 212618 99.9-100         213 236.000
```

Monetary

```
percents <- c(0, 0.25, 0.5, 0.75, 1)
title <- "Ratio of sales under given values"
x <- "Nr of Sales"
y <- "Monetary"
caption <- "(in the middle - )"
binwidth = 0.4
barwidth = 0.8
y_text_position = 9000
```

Standard scales

```
# standard table summary
d_summary(d_rfm, d_rfm$monetary)
```

```
##   mean    sd min      max  IQR median  total
## 1 1681 33770   0 12501064 1627    724 212618
```

```
quantile(d_rfm$monetary)
```

```
##       0%      25%      50%      75%     100%
##        0      245      724     1872 12501064
```

```
# all summaries in table format for standard percentage
d_rfm_m <- d_kpi_feature(percents, d_rfm$monetary)
d_rfm_m
```

```
##   percent  count  range range_count    value
## 1       0  17919      0       17919        0
## 2      25  53183   0-25       35264      245
## 3      50 106351  25-50       53168      724
## 4      75 159483  50-75       53132     1872
## 5     100 212618 75-100       53135 12501064
```

Observed scales

```
percents<- c(0, 0.25, 0.5, 0.75, 0.9, 0.99, 0.999, 1)

summary(d_rfm$monetary)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##        0      245      724     1681     1872 12501064
```

```
quantile(d_rfm$monetary)
```

```
##       0%      25%      50%      75%     100%
##        0      245      724     1872 12501064
```

```
d_rfm_r <-d_kpi_feature(percents, d_rfm$monetary)
d_rfm_r
```

```
##   percent  count    range range_count       value
## 1     0.0  17919        0       17919        0.00
## 2    25.0  53183     0-25       35264      245.00
## 3    50.0 106351    25-50       53168      724.00
## 4    75.0 159483    50-75       53132     1872.00
## 5    90.0 191366    75-90       31883     3898.00
## 6    99.0 210491    90-99       19125    10439.32
## 7    99.9 212405  99-99.9        1914    22638.38
## 8   100.0 212618 99.9-100         213 12501064.00
```

Quaterly changes in distribution of customers on recency, frequency and monetary

Profile(browserid) categories

Recency categories are:
recent (recency_1): NrOfDaysSinceLastOrder < 123 (70 % of sales are less than that value)
some_time_ago (recency_2): NrOfDaysSinceLastOrder >= 123 and NrOfDaysSinceLastOrder < 233
long_time_ago (recency_3): NrOfDaysSinceLastOrder >= 233

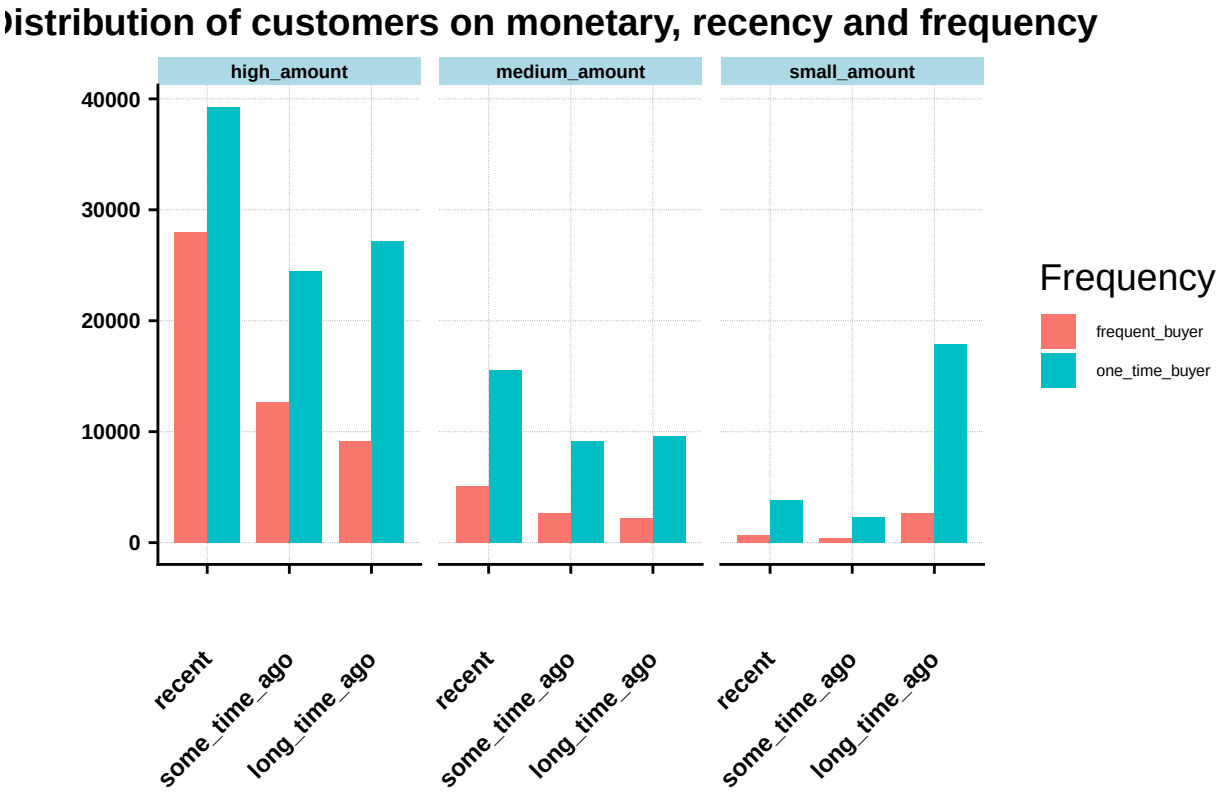
Frequency categories are:
one_time_buyer (frequency_1): NrOfSales = 1 (70 % of sales are less than that value)
frequent_buyer (frequency_2): NrOfSales > 1

Monetary categories are:
small_amount_spent (monetary_1): AOV (Average Value of Sales) < 85 (25 % of sales are less than that value)
medium_amount_spent (monetary_2): AOV >= 85 and AOV >= 737 (90% of the sales < 600)
high_amount_spent (monetary_3): AOV > 737

```
lrecency <- c("recent", "some_time_ago", "long_time_ago")
lfrequency <- c("frequent_buyer", "one_time_buyer")
lmonetary <- c("small_amount", "medium_amount", "high_amount")

d_rfm <- d_rfm %>%
 mutate(recency_distr =
  ifelse(recency >= 233, lrecency[3],
  ifelse(recency <= 123 , lrecency[1],
```

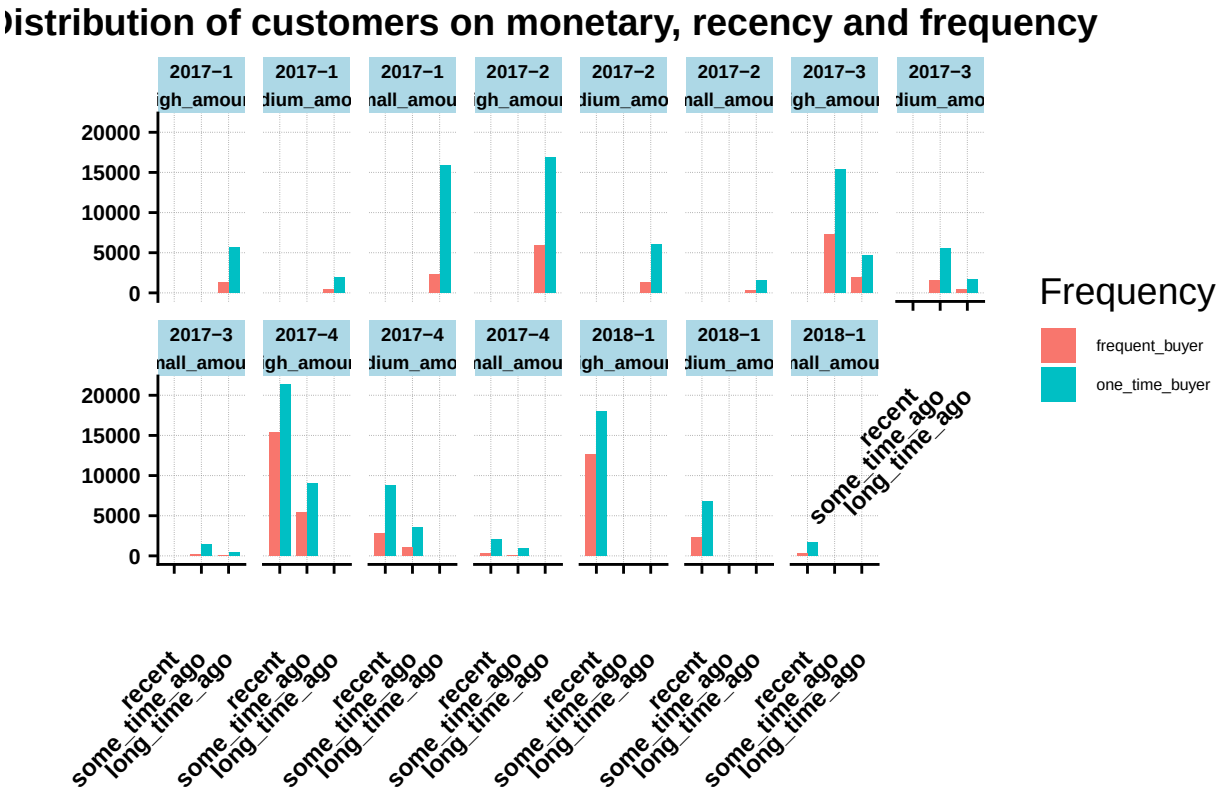
```
  lrecency[2]
  )))

d_rfm <- d_rfm %>%
 mutate(frequency_distr =
  ifelse(frequency == 1, lfrequency[2],
  lfrequency[1]
  ))

d_rfm <- d_rfm %>%
mutate(monetary_distr =
  ifelse(monetary >= 382, lmonetary[3],
  ifelse(monetary <= 85, lmonetary[1],
  lmonetary[2]
  )))
```

```
# custom order recency, frequency, monetary columns
d_rfm$recency_distr <- factor(d_rfm$recency_distr, levels=lrecency)
d_rfm$frequency_distr <- factor(d_rfm$frequency_distr, levels=lfrequency)

 p_base <- ggplot(d_rfm, aes(x = recency_distr, fill = frequency_distr)) +
  geom_bar(stat='count', position = "dodge", width = 0.8)+
  labs( title = "Distribution of customers on monetary, recency and frequency",
        x = "",
        y = "",
        caption = "",
        fill = "Frequency")+
  theme(plot.title = element_text(hjust = 0.6),
      legend.text = element_text(size = 6),
        axis.text.x=element_text(angle=45,hjust=1,vjust=0.5, size = 9, face = "bold"),

      strip.text = element_text(face = "bold", size=7,lineheight=5.0),
      strip.background = element_rect(fill="lightblue", colour="black",size=.1),
      axis.text.y=element_text(hjust=1,vjust=0.5, size = 8, face = "bold"),
      strip.text.x = element_text(margin = margin(.05, .05, 0.05, 0.05, "cm")),
      panel.grid.major = element_line(colour = "gray50", size = 0.1, linetype = "dotted"))

p1<-p_base + facet_wrap(~ monetary_distr, nrow = 1)
p1
```



Quaterly changes in distribution of customers on recency, frequency and monetary

```
p2<-p_base +
    facet_wrap(Q ~ monetary_distr, nrow = 2)
p2
```



Distribution of customers on recency, frequency cluster and monetary

Firstly, a cluster comprised of recency and frequency was done using a clustering algoritm. Then a distribution on monetary values and this cluster is done.

```
d_rfm <- d_rfm %>%
  filter(Q == params$quarter)
```

```
set.seed(100)

d_clusters <- select(d_rfm, recency, frequency, monetary)
d_clusters <- as.data.frame(scale(d_clusters))

#d_clusters$recency<-scale(d_clusters$recency)

clusters <- kmeans(d_clusters, centers = 6, nstart = 12)
#clusters$cluster
#clusters$size
#clusters$centers
#clusters$betweens/clusters$totss
#clusters$tot.withinss/clusters$totss
#dunn_km <- dunn(clusters = clusters$cluster, Data = d_clusters, method = "euclidian")
#dunn_km
```

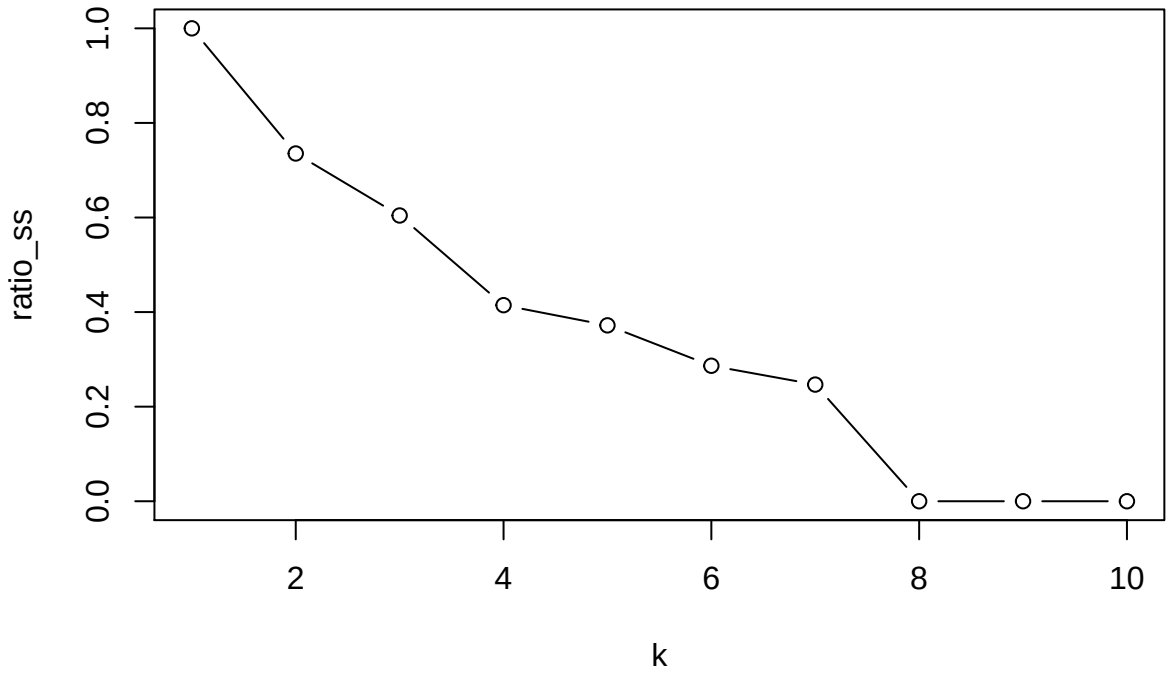
```
# lengths(segments) should be equal to cluster number
df_rank <- naming_segments(clusters, segments, 6)
```

```
df_rank
```

```
##   cluster                     segment  size value     recency   frequency
## 1       6                 Hibernating 16354   141  0.93317814 -0.05025941
## 2       2 Customers needing attention  1686   254  0.51197970  2.64708861
## 3       1          Potential Loyalist    25   365  0.22996362 23.85454421
## 4       5             Loyal customers 12947   433 -0.06524579 -0.14096694
## 5       4            Recent Customers 10544   522 -1.44905387 -0.22869339
## 6       3                   Champions     4   616 -1.64522450 -0.24263769
##       monetary r_rank f_rank m_rank
## 1 -0.052112110      1      4      1
## 2  0.230615798      2      5      4
## 3  0.408713137      3      6      5
## 4  0.015774641      4      3      3
## 5 -0.006397579      5      2      2
## 6 79.106794655      6      1      6
```

```
## hierarchical clustering
#Dist <- dist(d_clusters,method="euclidean")
#clusterObj <- hclust(Dist, method="average")
#nc <- 5 ## number of clusters
#clusters <- cutree(clusterObj,nc)
#dunn(Dist, cluster)
```

```
#clmethods <- c("hierarchical","kmeans")
#intern <- clValid(d_clusters, nClust = 5,
#              clMethods = clmethods, validation = "internal", maxitems = 19000, metric = #"euclidean")
# Summary
#summary(intern)
```



Customer segments

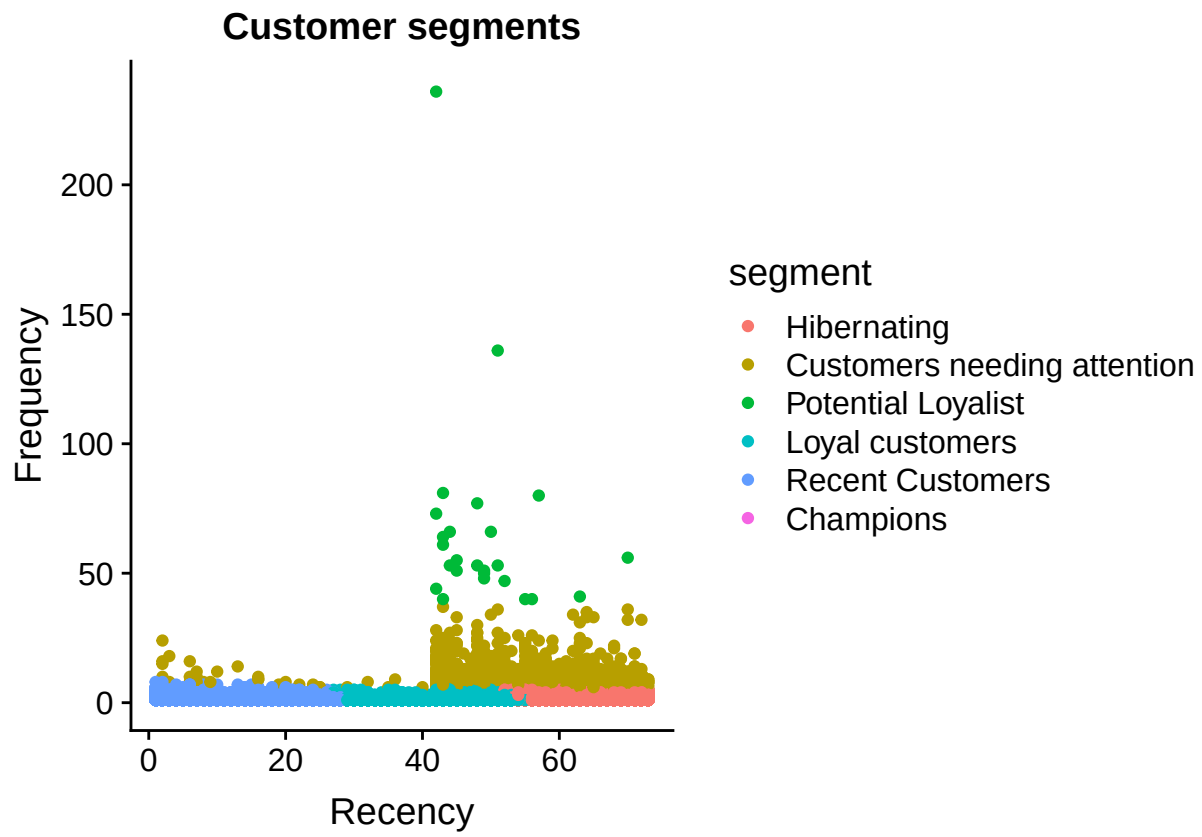
Champions :: Bought recently, buy often and spend the most!
Loyal Customers :: Spend good money with us often. Responsive to promotions.
Potential Loyalist :: Recent customers, but spent a good amount and bought more than once.
Recent Customers :: Bought most recently, but not often.
Promising :: Recent shoppers, but havenâ€™t spent much.
Customers Needing Attention :: Above average recency, frequency and monetary values.May not have bought very recently though.
Hibernating :: Last purchase was long back, low spenders and low number of orders.

```
d_result <- cbind(d_rfm, clusters$cluster)
#rename last column
names(d_result)[length(names(d_result))]<-"cluster"

# add segment names from df_rank
d_result <- add_column(d_result, segment = "")
d_result$segment <- as.character(d_result$cluster)

for (i in 1:nrow(d_result)) {
  # match cluster from results and ranking data
  position <- match(d_result$cluster[i], df_rank$cluster)
  d_result$segment[i] <- as.character(df_rank$segment[position])
}

d_result$segment <- factor(d_result$segment, levels = segments)

d_result_summary <- d_result  %>%
```

```
  select(segment) %>%
  group_by(segment) %>%
  summarise( size  = n())

kable(d_result_summary)
```

segment	size
Hibernating	16354
Customers needing attention	1686
Potential Loyalist	25
Loyal customers	12947
Recent Customers	10544
Champions	4

```
#custom order of segment
d_result$segment <- factor(d_result$segment, levels= segments)

ggplot(d_result, aes(x = recency, y = frequency, color = segment )) +
  geom_point() +
  labs(title = 'Customer segments', x='Recency', y='Frequency')
```




```
# custom order recency, frequency, monetary columns


 p_base_cluster <- ggplot(d_result, aes(x = segment, fill = segment)) +
```

```
  geom_bar(stat='count', position = "dodge", width = 0.8)+
  labs( title = "Customer segments on monetary values",
        x = "",
        y = "",
        caption = "")+
  theme(plot.title = element_text(hjust = 0.6),
      legend.text = element_text(size = 6),
        axis.text.x=element_text(angle=45,hjust=1,vjust=0.5, size = 9, face = "bold"),

      strip.text = element_text(face = "bold", size=7,lineheight=5.0),
      strip.background = element_rect(fill="lightblue", colour="black",size=.1),
      axis.text.y=element_text(hjust=1,vjust=0.5, size = 8, face = "bold"),
      strip.text.x = element_text(margin = margin(.05, .05, 0.05, 0.05, "cm")),
      panel.grid.major = element_line(colour = "gray50", size = 0.1, linetype = "dotted"))

p_base_cluster+facet_wrap(~ monetary_distr, nrow = 1)
```