Wholesale Customer Segmentation Technical Document

To find out the spending pattern of the business clients, we did the below clustering analysis and built up channel and region analysis upon it.

First, we use a database that contains 440 observations from the distributor's clients.

```
## Install required packages
# install.packages('corrplot')
# install.packages('tidyverse')

library(stats)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(cluster)
library(ggplot2)
# library(corrplot)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## v readr     2.1.4

## -- 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
```

```
# Read the data into a dataframe
wholesale = read.csv('Wholesale customers data.csv')

## Replace values for Channel
wholesale$Channel[wholesale$Channel == 1] <- 'Horeca'
```

```
wholesale$Channel[wholesale$Channel == 2] <- 'Retail'

## Replace values for Region
wholesale$Region[wholesale$Region == 1] <- 'Lisbon'
wholesale$Region[wholesale$Region == 2] <- 'Oporto'
wholesale$Region[wholesale$Region == 3] <- 'Other regions'

# Set k = the number of clusters we want
k=3
```

Performed a rudimentary exploratory data analysis on the data provided.

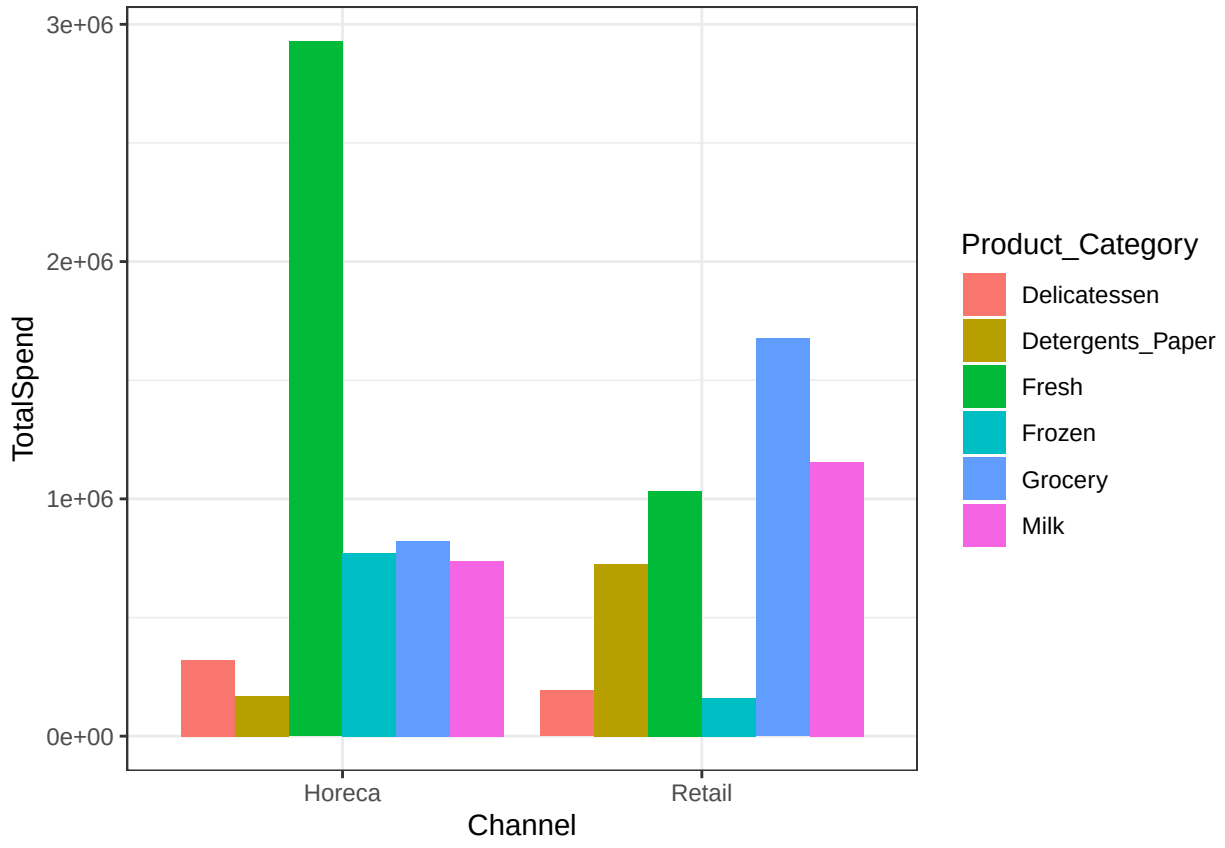
```
## CORRELATION ANALYSIS
corr_matrix <- cor(wholesale[,3:8])
round(corr_matrix, 2)
```

```
##                   Fresh Milk Grocery Frozen Detergents_Paper Delicatessen
## Fresh              1.00 0.10   -0.01   0.35            -0.10         0.24
## Milk               0.10 1.00    0.73   0.12             0.66         0.41
## Grocery           -0.01 0.73    1.00  -0.04             0.92         0.21
## Frozen             0.35 0.12   -0.04   1.00            -0.13         0.39
## Detergents_Paper -0.10 0.66    0.92  -0.13             1.00         0.07
## Delicatessen       0.24 0.41    0.21   0.39             0.07         1.00
```

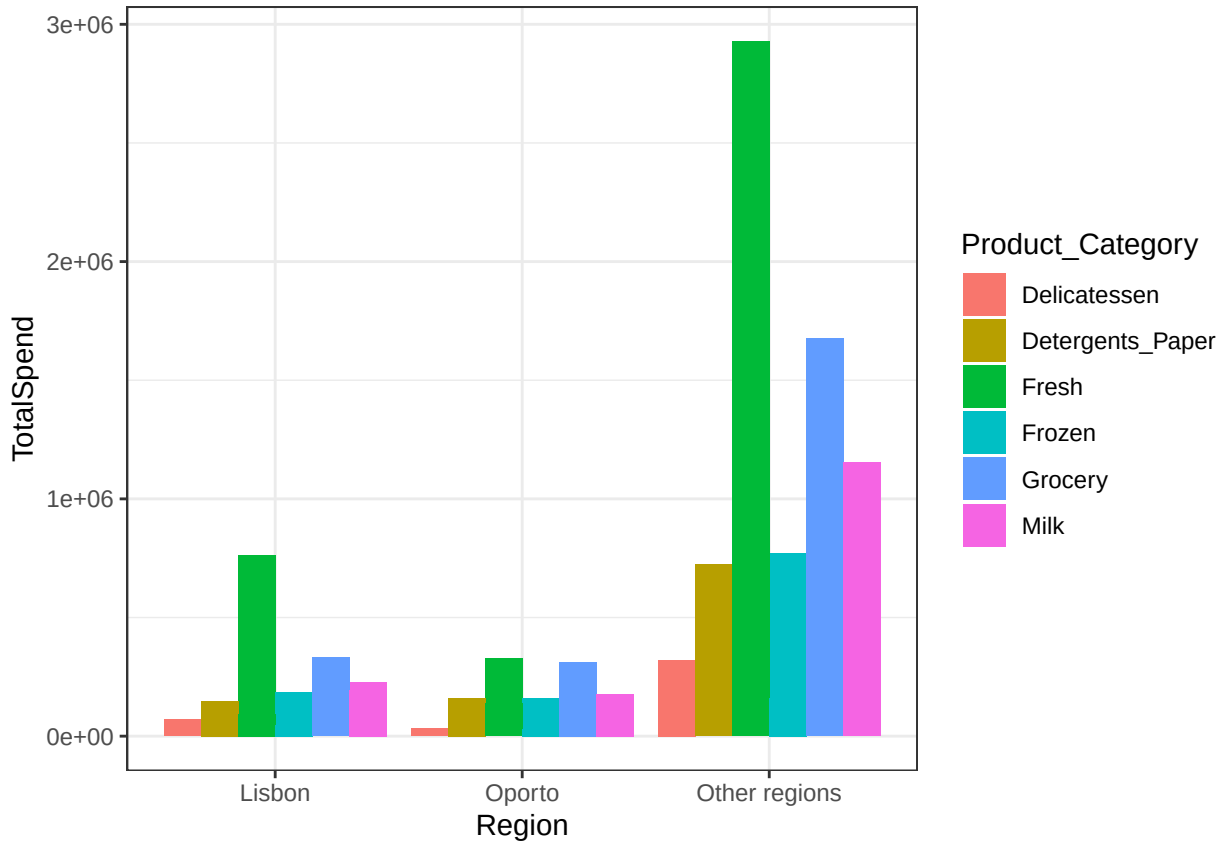
```
# Pivot data to create a single continuous/numerical factor using gather function
temp1 <- gather(data = wholesale, key = 'Product_Category', value = 'Spend', Fresh:Delicatessen)

# Summarise data for analysis
temp2 <- temp1 %>% group_by(Channel, Region, Product_Category) %>%
  summarise(TotalSpend = sum(Spend), AvgSpend = mean(Spend))
```

```
## `summarise()` has grouped output by 'Channel', 'Region'. You can override using
## the `.groups` argument.
```

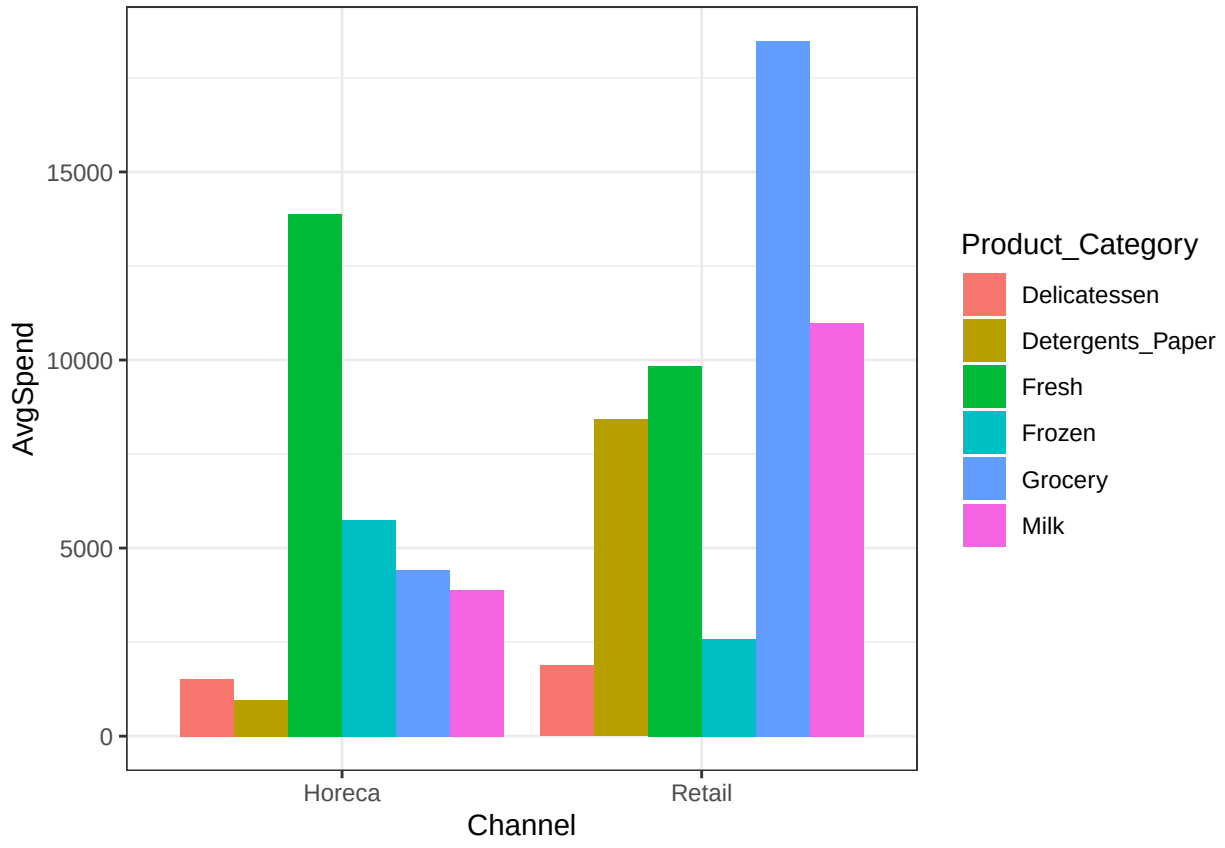
```
## TOTAL SPEND PATTERN
# Across Channels
ggplot(temp2, aes(Channel, TotalSpend, fill = Product_Category)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  theme_bw()
```



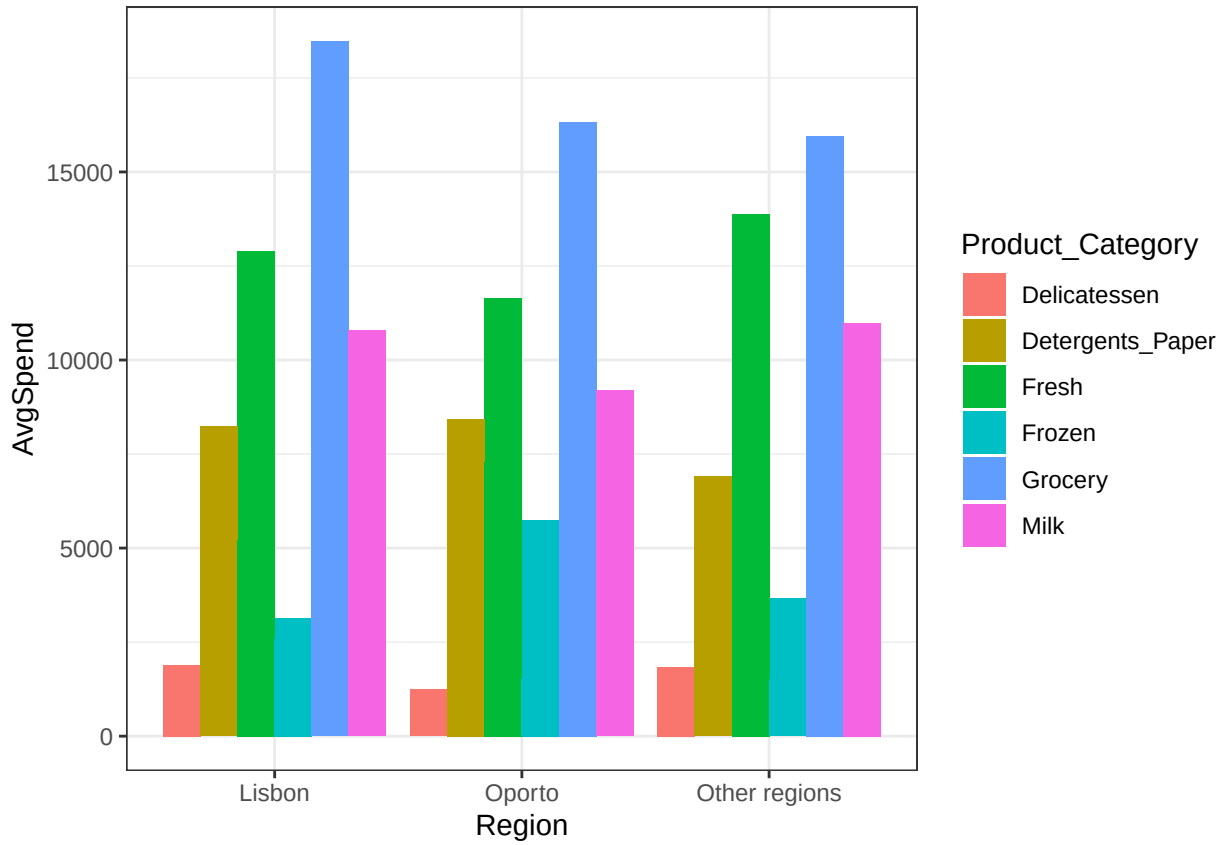

```
# Across Regions
ggplot(temp2, aes(Region, TotalSpend, fill = Product_Category)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  theme_bw()
```




```
## AVERAGE SPEND PATTERN
# Across Channels
ggplot(temp2, aes(Channel, AvgSpend, fill = Product_Category)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  theme_bw()
```




```
# Across Regions
ggplot(temp2, aes(Region, AvgSpend, fill = Product_Category)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  theme_bw()
```

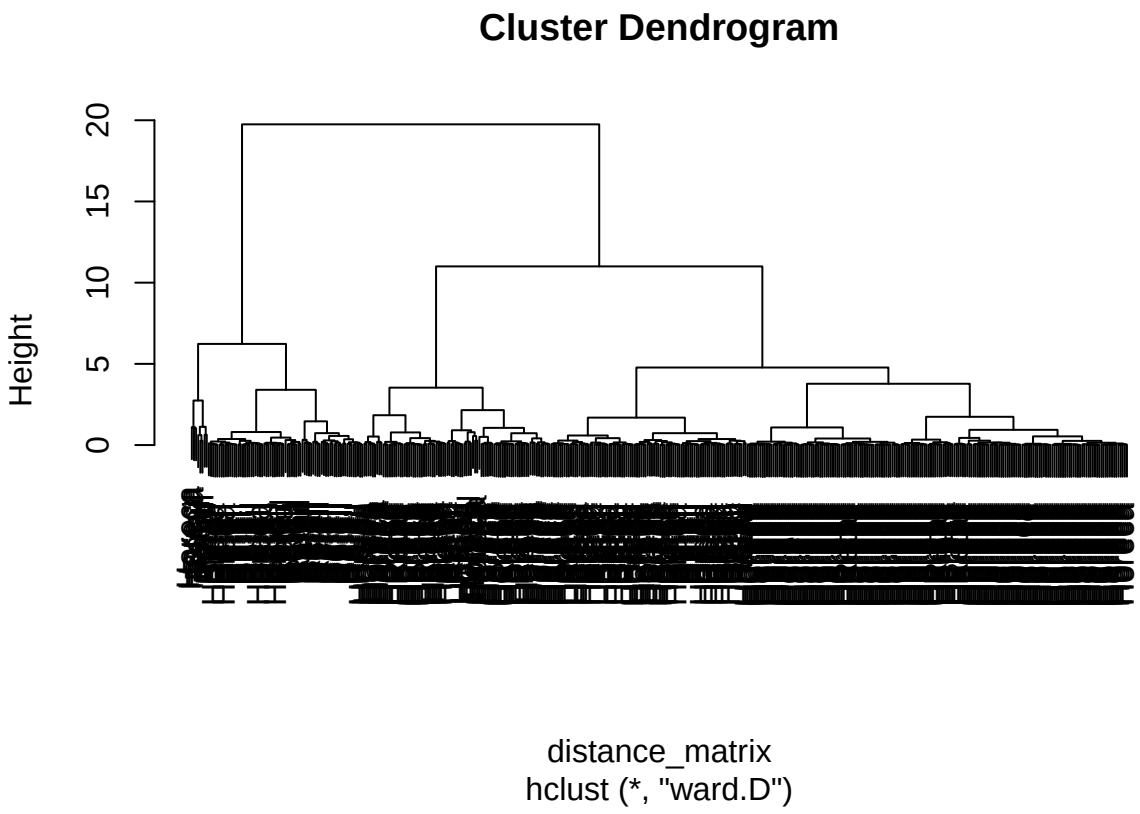


Although our data come with the same scale, they still have relatively large gaps in amounts. Therefore, we applied normalization to preprocess our data and calculated the distance matrix using euclidean method accordingly.

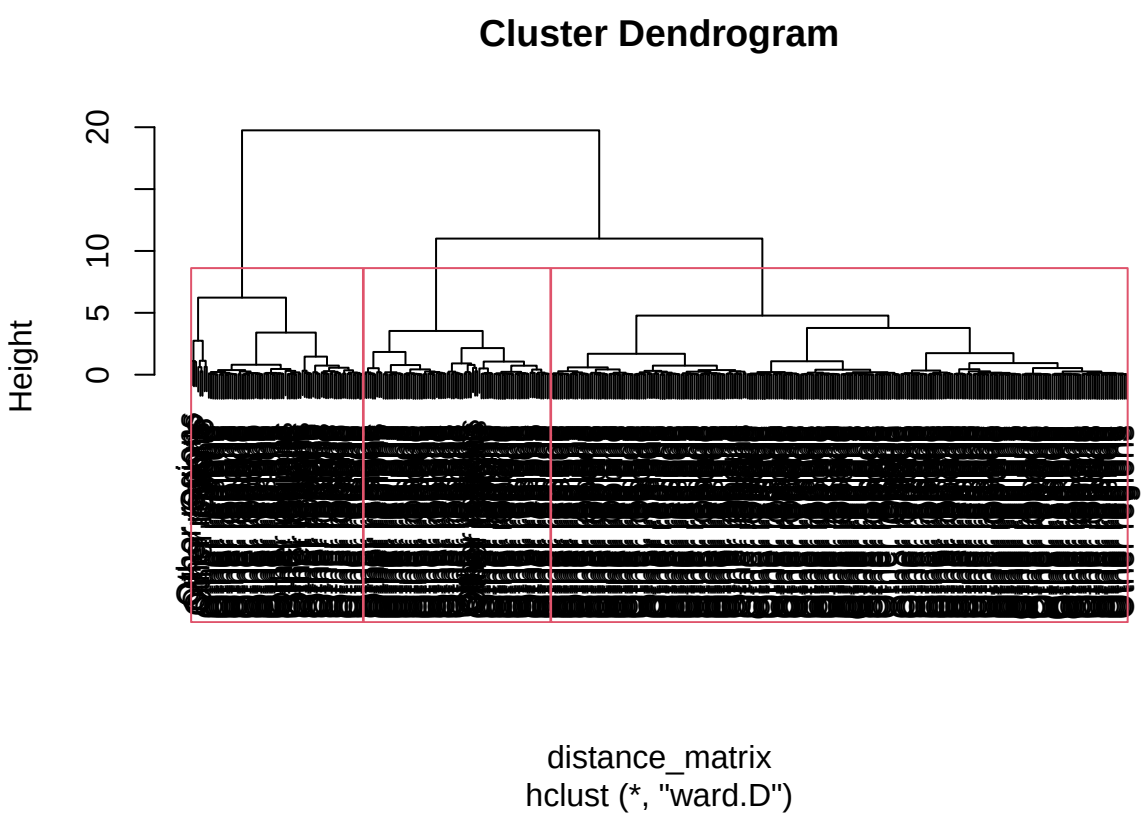
```
# Normalize the raw data

normalize = function(x){
  return ((x - min(x))/(max(x) - min(x)))}
wholesale_norm = wholesale %>% mutate_at(c(3:8), normalize)
distance_matrix = dist(wholesale_norm[,3:8], method = "euclidean")
```

To decide on the optimized clusters, we first used hierarchical clustering analysis with Ward distance measurement method to generate a dendogram. When deciding on the number of clusters, we first looked at the case where k = 5. As shown in the dendogram, there would be a very small cluster generated under 5-cluster scenario, which we considered less meaningful to analyze. Therefore, we eventually landed on 3 clusters.

```
# Calculate hierarchial clustering and display dendogram
h_clust = hclust(distance_matrix,method = "ward.D")
plot(h_clust,labels = (wholesale_norm$Channel))
```

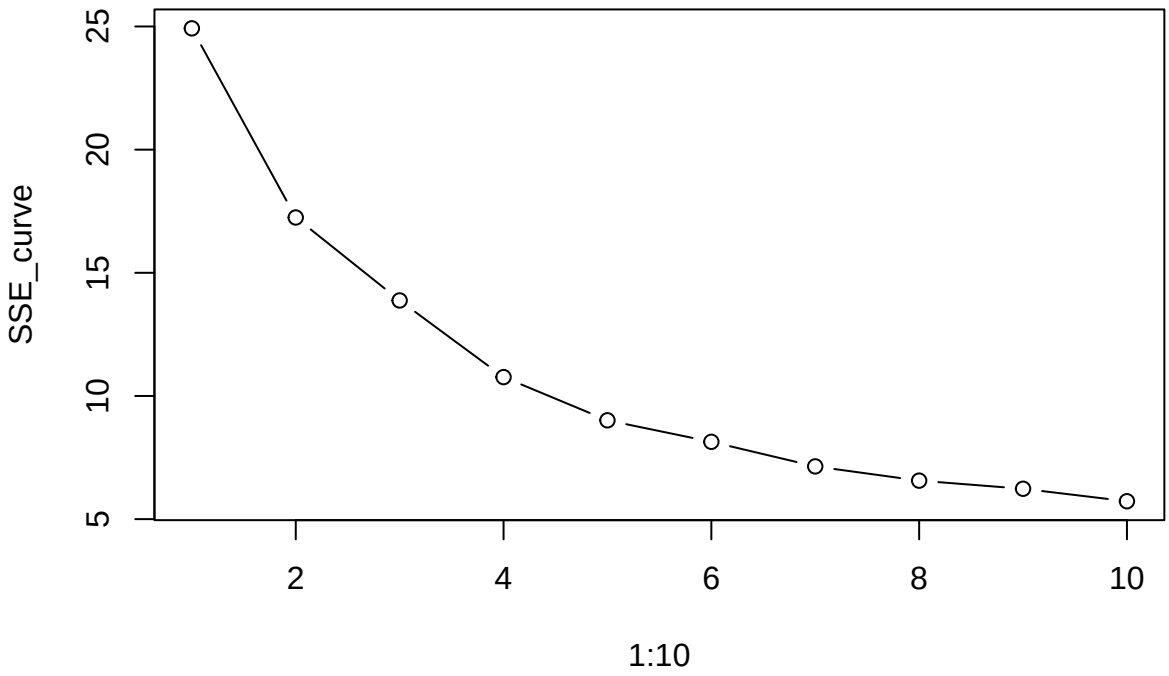



```
plot(h_clust,labels = (wholesale_norm$Region))
rect.hclust(h_clust, k = 3)
```



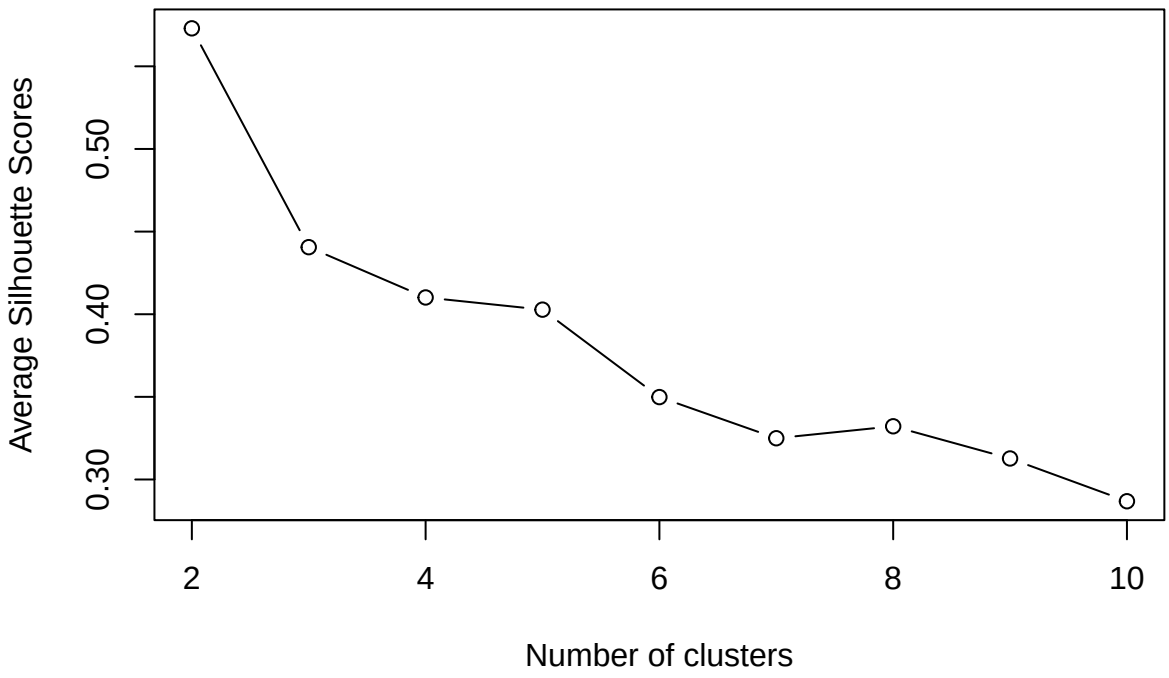
Then, we used the k-means method to do the clustering analysis. We plotted an SSE curve to look into the proper number of clusters. Looking into the plot, we noticed a relatively larger SSE gain when k = 3 and k = 5, similar to the findings in hierarchical clustering analysis.

```
SSE_curve <- c()
for (n in 1:10) {
  kcluster = kmeans(wholesale_norm[,3:8], n)
  sse = kcluster$tot.withinss
  SSE_curve[n] = sse
}
# plot SSE against number of clusters
plot(1:10, SSE_curve, type = "b")
```



Then, we tested through different k to calculate the mean silhouette score for each, as an additional support which shows the goodness of the clustering. It ends up that when k = 3, we can get a relatively good clustering result.

```
#test optimal k
silhouette_score <- function(k){
  kcluster <- kmeans(wholesale_norm[,3:8], centers = k, nstart = 25)
  ss <- silhouette(kcluster$cluster, dist = distance_matrix)
  mean(ss[, 3])
}
k_test = 2:10
avg_sil_score <- sapply(k_test, silhouette_score)
plot(k_test, type='b', avg_sil_score, xlab='Number of clusters', ylab='Average Silhouette Scores')
```



We then looked into the cluster distribution under the two different scenarios. Similar to hierarchical clustering analysis, we found some of the clusters under the 5-cluster scenario is too small to analyze. Therefore, we also landed on 3 clusters under the kmeans method.

```
## K-means with 5 clusters
kcluster1 = kmeans(wholesale_norm[,3:8], centers = 5)
cluster1 = wholesale_norm %>% mutate(Cluster = kcluster1$cluster)
cluster1 %>% group_by(Cluster) %>% summarise(count = n())
```

```
## # A tibble: 5 x 2
##   Cluster count
##     <int> <int>
## 1       1   101
## 2       2   257
## 3       3    28
## 4       4    49
## 5       5     5
```

```
# Observation - there are clusters with only 3 and 10 data points, which might not be helpful in genera

## K-means with 3 clusters
kcluster2 = kmeans(wholesale_norm[,3:8], centers = 3)
cluster2 = wholesale_norm %>% mutate(Cluster = kcluster2$cluster)
cluster2 %>% group_by(Cluster) %>% summarise(count = n())
```

```
## # A tibble: 3 x 2
```

```
##   Cluster count
##     <int> <int>
## 1       1   338
## 2       2    61
## 3       3    41
```

The next step would be choosing between the two clustering results from the two analysis methods. With a glimpse in the clustering centroids, we found that the two clustering results showed similar centroids, which we considered as a good sign.

```
# View h-clusters in normalized data
wholesale_norm$cluster = cutree(h_clust, k=3)
wholesale_norm %>% group_by(cluster) %>% summarise_at(c(3:8), mean) %>% mutate_if(is.numeric, round, di
```

```
## # A tibble: 3 x 7
##   cluster  Fresh   Milk Grocery Frozen Detergents_Paper Delicatessen
##     <dbl>  <dbl>  <dbl>   <dbl>  <dbl>            <dbl>        <dbl>
## 1       1 0.0727 0.0439  0.0471 0.0318           0.0309       0.0218
## 2       2 0.243  0.0740  0.0627 0.110            0.0252       0.0480
## 3       3 0.0737 0.198   0.240  0.0464           0.252        0.0473
```

```
# K-Means clustering (centroids)
kcluster2 = kmeans(wholesale_norm[,3:8], centers = 3)
kcluster2$centers
```

```
##        Fresh       Milk    Grocery     Frozen Detergents_Paper Delicatessen
## 1 0.07297318 0.05468315 0.06047771 0.03993169       0.04513443   0.02316561
## 2 0.31612924 0.08086923 0.07042464 0.11874307       0.02650803   0.06815213
## 3 0.07612741 0.26781048 0.31604316 0.03156052       0.34516819   0.04830937
```

We then calculated the Silhouette coefficients to evaluate the two clustering results. The kmeans method gives a better coefficient than the hclust. We then used the k-means clustering result for the following analysis.

```
# Silhouette coefficient - k means
library(cluster)
sc = silhouette(kcluster2$cluster, dist = distance_matrix)
summary(sc)
```

```
## Silhouette of 440 units in 3 clusters from silhouette.default(x = kcluster2$cluster, dist = distance
##  Cluster sizes and average silhouette widths:
##       338        61        41
## 0.5320006 0.0961455 0.1993003
## Individual silhouette widths:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.1810  0.2999  0.5179  0.4406  0.6217  0.6735
```

```
# Silhouette coefficient - hclust
library(cluster)
sc = silhouette(wholesale_norm$cluster, dist = distance_matrix)
summary(sc)
```

```
## Silhouette of 440 units in 3 clusters from silhouette.default(x = wholesale_norm$cluster, dist = dis
##  Cluster sizes and average silhouette widths:
##        271         88         81
## 0.53366522 0.03492611 0.03904118
## Individual silhouette widths:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.4102  0.1738  0.4277  0.3429  0.5934  0.6683
```

We matched the clusters back to the original dataset.

```
wholesale_clustered = wholesale %>% mutate(Cluster = kcluster2$cluster)
wholesale_clustered
```

```
##     Channel        Region  Fresh  Milk Grocery Frozen Detergents_Paper
## 1    Retail Other regions  12669  9656    7561    214              2674
## 2    Retail Other regions   7057  9810    9568   1762              3293
## 3    Retail Other regions   6353  8808    7684   2405              3516
## 4    Horeca Other regions  13265  1196    4221   6404               507
## 5    Retail Other regions  22615  5410    7198   3915              1777
## 6    Retail Other regions   9413  8259    5126    666              1795
## 7    Retail Other regions  12126  3199    6975    480              3140
## 8    Retail Other regions   7579  4956    9426   1669              3321
## 9    Horeca Other regions   5963  3648    6192    425              1716
## 10   Retail Other regions   6006 11093   18881   1159              7425
## 11   Retail Other regions   3366  5403   12974   4400              5977
## 12   Retail Other regions  13146  1124    4523   1420               549
## 13   Retail Other regions  31714 12319   11757    287              3881
## 14   Retail Other regions  21217  6208   14982   3095              6707
## 15   Retail Other regions  24653  9465   12091    294              5058
## 16   Horeca Other regions  10253  1114    3821    397               964
## 17   Retail Other regions   1020  8816   12121    134              4508
## 18   Horeca Other regions   5876  6157    2933    839               370
## 19   Retail Other regions  18601  6327   10099   2205              2767
## 20   Horeca Other regions   7780  2495    9464    669              2518
## 21   Retail Other regions  17546  4519    4602   1066              2259
## 22   Horeca Other regions   5567   871    2010   3383               375
## 23   Horeca Other regions  31276  1917    4469   9408              2381
## 24   Retail Other regions  26373 36423   22019   5154              4337
## 25   Retail Other regions  22647  9776   13792   2915              4482
## 26   Retail Other regions  16165  4230    7595    201              4003
## 27   Horeca Other regions   9898   961    2861   3151               242
## 28   Horeca Other regions  14276   803    3045    485               100
## 29   Retail Other regions   4113 20484   25957   1158              8604
## 30   Horeca Other regions  43088  2100    2609   1200              1107
## 31   Horeca Other regions  18815  3610   11107   1148              2134
## 32   Horeca Other regions   2612  4339    3133   2088               820
## 33   Horeca Other regions  21632  1318    2886    266               918
## 34   Horeca Other regions  29729  4786    7326   6130               361
## 35   Horeca Other regions   1502  1979    2262    425               483
## 36   Retail Other regions    688  5491   11091    833              4239
## 37   Horeca Other regions  29955  4362    5428   1729               862
## 38   Retail Other regions  15168 10556   12477   1920              6506
## 39   Retail Other regions   4591 15729   16709     33              6956
```

```
## 40   Horeca Other regions  56159   555     902  10002             212
## 41   Horeca Other regions  24025  4332    4757   9510            1145
## 42   Horeca Other regions  19176  3065    5956   2033            2575
## 43   Retail Other regions  10850  7555   14961    188            6899
## 44   Retail Other regions    630 11095   23998    787            9529
## 45   Retail Other regions   9670  7027   10471    541            4618
## 46   Retail Other regions   5181 22044   21531   1740            7353
## 47   Retail Other regions   3103 14069   21955   1668            6792
## 48   Retail Other regions  44466 54259   55571   7782           24171
## 49   Retail Other regions  11519  6152   10868    584            5121
## 50   Retail Other regions   4967 21412   28921   1798           13583
## 51   Horeca Other regions   6269  1095    1980   3860             609
## 52   Horeca Other regions   3347  4051    6996    239            1538
## 53   Retail Other regions  40721  3916    5876    532            2587
## 54   Retail Other regions    491 10473   11532    744            5611
## 55   Horeca Other regions  27329  1449    1947   2436             204
## 56   Horeca Other regions   5264  3683    5005   1057            2024
## 57   Retail Other regions   4098 29892   26866   2616           17740
## 58   Retail Other regions   5417  9933   10487     38            7572
## 59   Horeca Other regions  13779  1970    1648    596             227
## 60   Horeca Other regions   6137  5360    8040    129            3084
## 61   Retail Other regions   8590  3045    7854     96            4095
## 62   Retail Other regions  35942 38369   59598   3254           26701
## 63   Retail Other regions   7823  6245    6544   4154            4074
## 64   Retail Other regions   9396 11601   15775   2896            7677
## 65   Horeca Other regions   4760  1227    3250   3724            1247
## 66   Retail Other regions     85 20959   45828     36           24231
## 67   Horeca Other regions      9  1534    7417    175            3468
## 68   Retail Other regions  19913  6759   13462   1256            5141
## 69   Horeca Other regions   2446  7260    3993   5870             788
## 70   Horeca Other regions   8352  2820    1293    779             656
## 71   Horeca Other regions  16705  2037    3202  10643             116
## 72   Horeca Other regions  18291  1266   21042   5373            4173
## 73   Horeca Other regions   4420  5139    2661   8872            1321
## 74   Retail Other regions  19899  5332    8713   8132             764
## 75   Retail Other regions   8190  6343    9794   1285            1901
## 76   Horeca Other regions  20398  1137       3   4407               3
## 77   Horeca Other regions    717  3587    6532   7530             529
## 78   Retail Other regions  12205 12697   28540    869           12034
## 79   Horeca Other regions  10766  1175    2067   2096             301
## 80   Horeca Other regions   1640  3259    3655    868            1202
## 81   Horeca Other regions   7005   829    3009    430             610
## 82   Retail Other regions    219  9540   14403    283            7818
## 83   Retail Other regions  10362  9232   11009    737            3537
## 84   Horeca Other regions  20874  1563    1783   2320             550
## 85   Retail Other regions  11867  3327    4814   1178            3837
## 86   Retail Other regions  16117 46197   92780   1026           40827
## 87   Retail Other regions  22925 73498   32114    987           20070
## 88   Horeca Other regions  43265  5025    8117   6312            1579
## 89   Horeca Other regions   7864   542    4042   9735             165
## 90   Horeca Other regions  24904  3836    5330   3443             454
## 91   Horeca Other regions  11405   596    1638   3347              69
## 92   Horeca Other regions  12754  2762    2530   8693             627
## 93   Retail Other regions   9198 27472   32034   3232           18906
```

```
## 94   Horeca Other regions  11314  3090    2062  35009              71
## 95   Retail Other regions   5626 12220   11323    206            5038
## 96   Horeca Other regions      3  2920    6252    440             223
## 97   Retail Other regions     23  2616    8118    145            3874
## 98   Horeca Other regions    403   254     610    774              54
## 99   Horeca Other regions    503   112     778    895              56
## 100  Horeca Other regions   9658  2182    1909   5639             215
## 101  Retail Other regions  11594  7779   12144   3252            8035
## 102  Retail Other regions   1420 10810   16267   1593            6766
## 103  Retail Other regions   2932  6459    7677   2561            4573
## 104  Horeca Other regions  56082  3504    8906  18028            1480
## 105  Horeca Other regions  14100  2132    3445   1336            1491
## 106  Horeca Other regions  15587  1014    3970    910             139
## 107  Retail Other regions   1454  6337   10704    133            6830
## 108  Retail Other regions   8797 10646   14886   2471            8969
## 109  Retail Other regions   1531  8397    6981    247            2505
## 110  Retail Other regions   1406 16729   28986    673             836
## 111  Horeca Other regions  11818  1648    1694   2276             169
## 112  Retail Other regions  12579 11114   17569    805            6457
## 113  Horeca Other regions  19046  2770    2469   8853             483
## 114  Horeca Other regions  14438  2295    1733   3220             585
## 115  Horeca Other regions  18044  1080    2000   2555             118
## 116  Horeca Other regions  11134   793    2988   2715             276
## 117  Horeca Other regions  11173  2521    3355   1517             310
## 118  Horeca Other regions   6990  3880    5380   1647             319
## 119  Horeca Other regions  20049  1891    2362   5343             411
## 120  Horeca Other regions   8258  2344    2147   3896             266
## 121  Horeca Other regions  17160  1200    3412   2417             174
## 122  Horeca Other regions   4020  3234    1498   2395             264
## 123  Horeca Other regions  12212   201     245   1991              25
## 124  Retail Other regions  11170 10769    8814   2194            1976
## 125  Horeca Other regions  36050  1642    2961   4787             500
## 126  Horeca Other regions  76237  3473    7102  16538             778
## 127  Horeca Other regions  19219  1840    1658   8195             349
## 128  Retail Other regions  21465  7243   10685    880            2386
## 129  Horeca Other regions    140  8847    3823    142            1062
## 130  Horeca Other regions  42312   926    1510   1718             410
## 131  Horeca Other regions   7149  2428     699   6316             395
## 132  Horeca Other regions   2101   589     314    346              70
## 133  Horeca Other regions  14903  2032    2479    576             955
## 134  Horeca Other regions   9434  1042    1235    436             256
## 135  Horeca Other regions   7388  1882    2174    720              47
## 136  Horeca Other regions   6300  1289    2591   1170             199
## 137  Horeca Other regions   4625  8579    7030   4575            2447
## 138  Horeca Other regions   3087  8080    8282    661             721
## 139  Horeca Other regions  13537  4257    5034    155             249
## 140  Horeca Other regions   5387  4979    3343    825             637
## 141  Horeca Other regions  17623  4280    7305   2279             960
## 142  Horeca Other regions  30379 13252    5189    321              51
## 143  Horeca Other regions  37036  7152    8253   2995              20
## 144  Horeca Other regions  10405  1596    1096   8425             399
## 145  Horeca Other regions  18827  3677    1988    118             516
## 146  Retail Other regions  22039  8384   34792     42           12591
## 147  Horeca Other regions   7769  1936    2177    926              73
```

```
## 148  Horeca Other regions   9203  3373    2707   1286             1082
## 149  Horeca Other regions   5924   584     542   4052              283
## 150  Horeca Other regions  31812  1433    1651    800              113
## 151  Horeca Other regions  16225  1825    1765    853              170
## 152  Horeca Other regions   1289  3328    2022    531              255
## 153  Horeca Other regions  18840  1371    3135   3001              352
## 154  Horeca Other regions   3463  9250    2368    779              302
## 155  Horeca Other regions    622    55     137     75                7
## 156  Retail Other regions   1989 10690   19460    233            11577
## 157  Retail Other regions   3830  5291   14855    317             6694
## 158  Horeca Other regions  17773  1366    2474   3378              811
## 159  Retail Other regions   2861  6570    9618    930             4004
## 160  Retail Other regions    355  7704   14682    398             8077
## 161  Retail Other regions   1725  3651   12822    824             4424
## 162  Horeca Other regions  12434   540     283   1092                3
## 163  Horeca Other regions  15177  2024    3810   2665              232
## 164  Retail Other regions   5531 15726   26870   2367            13726
## 165  Retail Other regions   5224  7603    8584   2540             3674
## 166  Retail Other regions  15615 12653   19858   4425             7108
## 167  Retail Other regions   4822  6721    9170    993             4973
## 168  Horeca Other regions   2926  3195    3268    405             1680
## 169  Horeca Other regions   5809   735     803   1393               79
## 170  Horeca Other regions   5414   717    2155   2399               69
## 171  Retail Other regions    260  8675   13430   1116             7015
## 172  Retail Other regions    200 25862   19816    651             8773
## 173  Horeca Other regions    955  5479    6536    333             2840
## 174  Retail Other regions    514  7677   19805    937             9836
## 175  Horeca Other regions    286  1208    5241   2515              153
## 176  Retail Other regions   2343  7845   11874     52             4196
## 177  Horeca Other regions  45640  6958    6536   7368             1532
## 178  Horeca Other regions  12759  7330    4533   1752               20
## 179  Horeca Other regions  11002  7075    4945   1152              120
## 180  Horeca Other regions   3157  4888    2500   4477              273
## 181  Horeca Other regions  12356  6036    8887    402             1382
## 182  Horeca Other regions 112151 29627   18148  16745             4948
## 183  Horeca Other regions    694  8533   10518    443             6907
## 184  Horeca Other regions  36847 43950   20170  36534              239
## 185  Horeca Other regions    327   918    4710     74              334
## 186  Horeca Other regions   8170  6448    1139   2181               58
## 187  Horeca Other regions   3009   521     854   3470              949
## 188  Horeca Other regions   2438  8002    9819   6269             3459
## 189  Retail Other regions   8040  7639   11687   2758             6839
## 190  Retail Other regions    834 11577   11522    275             4027
## 191  Horeca Other regions  16936  6250    1981   7332              118
## 192  Horeca Other regions  13624   295    1381    890               43
## 193  Horeca Other regions   5509  1461    2251    547              187
## 194  Retail Other regions    180  3485   20292    959             5618
## 195  Horeca Other regions   7107  1012    2974    806              355
## 196  Horeca Other regions  17023  5139    5230   7888              330
## 197  Horeca        Lisbon  30624  7209    4897  18711              763
## 198  Retail        Lisbon   2427  7097   10391   1127             4314
## 199  Horeca        Lisbon  11686  2154    6824   3527              592
## 200  Horeca        Lisbon   9670  2280    2112    520              402
## 201  Retail        Lisbon   3067 13240   23127   3941             9959
```

```
## 202  Retail        Lisbon   4484 14399   24708   3549            14235
## 203  Horeca        Lisbon  25203 11487    9490   5065              284
## 204  Horeca        Lisbon    583   685    2216    469              954
## 205  Horeca        Lisbon   1956   891    5226   1383                5
## 206  Retail        Lisbon   1107 11711   23596    955             9265
## 207  Horeca        Lisbon   6373   780     950    878              288
## 208  Retail        Lisbon   2541  4737    6089   2946             5316
## 209  Horeca        Lisbon   1537  3748    5838   1859             3381
## 210  Retail        Lisbon   5550 12729   16767    864            12420
## 211  Horeca        Lisbon  18567  1895    1393   1801              244
## 212  Retail        Lisbon  12119 28326   39694   4736            19410
## 213  Horeca        Lisbon   7291  1012    2062   1291              240
## 214  Horeca        Lisbon   3317  6602    6861   1329             3961
## 215  Retail        Lisbon   2362  6551   11364    913             5957
## 216  Horeca        Lisbon   2806 10765   15538   1374             5828
## 217  Retail        Lisbon   2532 16599   36486    179            13308
## 218  Horeca        Lisbon  18044  1475    2046   2532              130
## 219  Retail        Lisbon     18  7504   15205   1285             4797
## 220  Horeca        Lisbon   4155   367    1390   2306               86
## 221  Horeca        Lisbon  14755   899    1382   1765               56
## 222  Horeca        Lisbon   5396  7503   10646     91             4167
## 223  Horeca        Lisbon   5041  1115    2856   7496              256
## 224  Retail        Lisbon   2790  2527    5265   5612              788
## 225  Horeca        Lisbon   7274   659    1499    784               70
## 226  Horeca        Lisbon  12680  3243    4157    660              761
## 227  Retail        Lisbon  20782  5921    9212   1759             2568
## 228  Horeca        Lisbon   4042  2204    1563   2286              263
## 229  Horeca        Lisbon   1869   577     572    950             4762
## 230  Horeca        Lisbon   8656  2746    2501   6845              694
## 231  Retail        Lisbon  11072  5989    5615   8321              955
## 232  Horeca        Lisbon   2344 10678    3828   1439             1566
## 233  Horeca        Lisbon  25962  1780    3838    638              284
## 234  Horeca        Lisbon    964  4984    3316    937              409
## 235  Horeca        Lisbon  15603  2703    3833   4260              325
## 236  Horeca        Lisbon   1838  6380    2824   1218             1216
## 237  Horeca        Lisbon   8635   820    3047   2312              415
## 238  Horeca        Lisbon  18692  3838     593   4634               28
## 239  Horeca        Lisbon   7363   475     585   1112               72
## 240  Horeca        Lisbon  47493  2567    3779   5243              828
## 241  Horeca        Lisbon  22096  3575    7041  11422              343
## 242  Horeca        Lisbon  24929  1801    2475   2216              412
## 243  Horeca        Lisbon  18226   659    2914   3752              586
## 244  Horeca        Lisbon  11210  3576    5119    561             1682
## 245  Horeca        Lisbon   6202  7775   10817   1183             3143
## 246  Retail        Lisbon   3062  6154   13916    230             8933
## 247  Horeca        Lisbon   8885  2428    1777   1777              430
## 248  Horeca        Lisbon  13569   346     489   2077               44
## 249  Horeca        Lisbon  15671  5279    2406    559              562
## 250  Horeca        Lisbon   8040  3795    2070   6340              918
## 251  Horeca        Lisbon   3191  1993    1799   1730              234
## 252  Retail        Lisbon   6134 23133   33586   6746            18594
## 253  Horeca        Lisbon   6623  1860    4740   7683              205
## 254  Horeca        Lisbon  29526  7961   16966    432              363
## 255  Horeca        Lisbon  10379 17972    4748   4686             1547
```

```
## 256  Horeca        Lisbon  31614   489    1495   3242             111
## 257  Horeca        Lisbon  11092  5008    5249    453             392
## 258  Horeca        Lisbon   8475  1931    1883   5004            3593
## 259  Horeca        Lisbon  56083  4563    2124   6422             730
## 260  Horeca        Lisbon  53205  4959    7336   3012             967
## 261  Horeca        Lisbon   9193  4885    2157    327             780
## 262  Horeca        Lisbon   7858  1110    1094   6818              49
## 263  Horeca        Lisbon  23257  1372    1677    982             429
## 264  Horeca        Lisbon   2153  1115    6684   4324            2894
## 265  Retail        Lisbon   1073  9679   15445     61            5980
## 266  Horeca        Lisbon   5909 23527   13699  10155             830
## 267  Retail        Lisbon    572  9763   22182   2221            4882
## 268  Horeca        Lisbon  20893  1222    2576   3975             737
## 269  Retail        Lisbon  11908  8053   19847   1069            6374
## 270  Horeca        Lisbon  15218   258    1138   2516             333
## 271  Horeca        Lisbon   4720  1032     975   5500             197
## 272  Horeca        Lisbon   2083  5007    1563   1120             147
## 273  Horeca        Lisbon    514  8323    6869    529              93
## 274  Horeca Other regions  36817  3045    1493   4802             210
## 275  Horeca Other regions    894  1703    1841    744             759
## 276  Horeca Other regions    680  1610     223    862              96
## 277  Horeca Other regions  27901  3749    6964   4479             603
## 278  Horeca Other regions   9061   829     683  16919             621
## 279  Horeca Other regions  11693  2317    2543   5845             274
## 280  Retail Other regions  17360  6200    9694   1293            3620
## 281  Horeca Other regions   3366  2884    2431    977             167
## 282  Retail Other regions  12238  7108    6235   1093            2328
## 283  Horeca Other regions  49063  3965    4252   5970            1041
## 284  Horeca Other regions  25767  3613    2013  10303             314
## 285  Horeca Other regions  68951  4411   12609   8692             751
## 286  Horeca Other regions  40254   640    3600   1042             436
## 287  Horeca Other regions   7149  2247    1242   1619            1226
## 288  Horeca Other regions  15354  2102    2828   8366             386
## 289  Horeca Other regions  16260   594    1296    848             445
## 290  Horeca Other regions  42786   286     471   1388              32
## 291  Horeca Other regions   2708  2160    2642    502             965
## 292  Horeca Other regions   6022  3354    3261   2507             212
## 293  Horeca Other regions   2838  3086    4329   3838             825
## 294  Retail        Oporto   3996 11103   12469    902            5952
## 295  Horeca        Oporto  21273  2013    6550    909             811
## 296  Retail        Oporto   7588  1897    5234    417            2208
## 297  Horeca        Oporto  19087  1304    3643   3045             710
## 298  Retail        Oporto   8090  3199    6986   1455            3712
## 299  Retail        Oporto   6758  4560    9965    934            4538
## 300  Horeca        Oporto    444   879    2060    264             290
## 301  Retail        Oporto  16448  6243    6360    824            2662
## 302  Retail        Oporto   5283 13316   20399   1809            8752
## 303  Retail        Oporto   2886  5302    9785    364            6236
## 304  Retail        Oporto   2599  3688   13829    492           10069
## 305  Retail        Oporto    161  7460   24773    617           11783
## 306  Retail        Oporto    243 12939    8852    799            3909
## 307  Retail        Oporto   6468 12867   21570   1840            7558
## 308  Horeca        Oporto  17327  2374    2842   1149             351
## 309  Horeca        Oporto   6987  1020    3007    416             257
```

```
## 310  Retail        Oporto    918 20655   13567   1465             6846
## 311  Horeca        Oporto   7034  1492    2405  12569              299
## 312  Horeca        Oporto  29635  2335    8280   3046              371
## 313  Retail        Oporto   2137  3737   19172   1274            17120
## 314  Horeca        Oporto   9784   925    2405   4447              183
## 315  Horeca        Oporto  10617  1795    7647   1483              857
## 316  Retail        Oporto   1479 14982   11924    662             3891
## 317  Horeca        Oporto   7127  1375    2201   2679               83
## 318  Horeca        Oporto   1182  3088    6114    978              821
## 319  Horeca        Oporto  11800  2713    3558   2121              706
## 320  Retail        Oporto   9759 25071   17645   1128            12408
## 321  Horeca        Oporto   1774  3696    2280    514              275
## 322  Horeca        Oporto   9155  1897    5167   2714              228
## 323  Horeca        Oporto  15881   713    3315   3703             1470
## 324  Horeca        Oporto  13360   944   11593    915             1679
## 325  Horeca        Oporto  25977  3587    2464   2369              140
## 326  Horeca        Oporto  32717 16784   13626  60869             1272
## 327  Horeca        Oporto   4414  1610    1431   3498              387
## 328  Horeca        Oporto    542   899    1664    414               88
## 329  Horeca        Oporto  16933  2209    3389   7849              210
## 330  Horeca        Oporto   5113  1486    4583   5127              492
## 331  Horeca        Oporto   9790  1786    5109   3570              182
## 332  Retail        Oporto  11223 14881   26839   1234             9606
## 333  Horeca        Oporto  22321  3216    1447   2208              178
## 334  Retail        Oporto   8565  4980   67298    131            38102
## 335  Retail        Oporto  16823   928    2743  11559              332
## 336  Retail        Oporto  27082  6817   10790   1365             4111
## 337  Horeca        Oporto  13970  1511    1330    650              146
## 338  Horeca        Oporto   9351  1347    2611   8170              442
## 339  Horeca        Oporto      3   333    7021  15601               15
## 340  Horeca        Oporto   2617  1188    5332   9584              573
## 341  Retail Other regions    381  4025    9670    388             7271
## 342  Retail Other regions   2320  5763   11238    767             5162
## 343  Horeca Other regions    255  5758    5923    349             4595
## 344  Retail Other regions   1689  6964   26316   1456            15469
## 345  Horeca Other regions   3043  1172    1763   2234              217
## 346  Horeca Other regions   1198  2602    8335    402             3843
## 347  Retail Other regions   2771  6939   15541   2693             6600
## 348  Retail Other regions  27380  7184   12311   2809             4621
## 349  Horeca Other regions   3428  2380    2028   1341             1184
## 350  Retail Other regions   5981 14641   20521   2005            12218
## 351  Horeca Other regions   3521  1099    1997   1796              173
## 352  Retail Other regions   1210 10044   22294   1741            12638
## 353  Horeca Other regions    608  1106    1533    830               90
## 354  Retail Other regions    117  6264   21203    228             8682
## 355  Horeca Other regions  14039  7393    2548   6386             1333
## 356  Horeca Other regions    190   727    2012    245              184
## 357  Horeca Other regions  22686   134     218   3157                9
## 358  Retail Other regions     37  1275   22272    137             6747
## 359  Horeca Other regions    759 18664    1660   6114              536
## 360  Horeca Other regions    796  5878    2109    340              232
## 361  Horeca Other regions  19746  2872    2006   2601              468
## 362  Horeca Other regions   4734   607     864   1206              159
## 363  Horeca Other regions   2121  1601    2453    560              179
```

```
## 364  Horeca Other regions   4627   997    4438    191             1335
## 365  Horeca Other regions   2615   873    1524   1103              514
## 366  Retail Other regions   4692  6128    8025   1619             4515
## 367  Horeca Other regions   9561  2217    1664   1173              222
## 368  Horeca Other regions   3477   894     534   1457              252
## 369  Horeca Other regions  22335  1196    2406   2046              101
## 370  Horeca Other regions   6211   337     683   1089               41
## 371  Retail Other regions  39679  3944    4955   1364              523
## 372  Horeca Other regions  20105  1887    1939   8164              716
## 373  Horeca Other regions   3884  3801    1641    876              397
## 374  Retail Other regions  15076  6257    7398   1504             1916
## 375  Horeca Other regions   6338  2256    1668   1492              311
## 376  Horeca Other regions   5841  1450    1162    597              476
## 377  Retail Other regions   3136  8630   13586   5641             4666
## 378  Horeca Other regions  38793  3154    2648   1034               96
## 379  Horeca Other regions   3225  3294    1902    282               68
## 380  Retail Other regions   4048  5164   10391    130              813
## 381  Horeca Other regions  28257   944    2146   3881              600
## 382  Horeca Other regions  17770  4591    1617   9927              246
## 383  Horeca Other regions  34454  7435    8469   2540             1711
## 384  Horeca Other regions   1821  1364    3450   4006              397
## 385  Horeca Other regions  10683 21858   15400   3635              282
## 386  Horeca Other regions  11635   922    1614   2583              192
## 387  Horeca Other regions   1206  3620    2857   1945              353
## 388  Horeca Other regions  20918  1916    1573   1960              231
## 389  Horeca Other regions   9785   848    1172   1677              200
## 390  Horeca Other regions   9385  1530    1422   3019              227
## 391  Horeca Other regions   3352  1181    1328   5502              311
## 392  Horeca Other regions   2647  2761    2313    907               95
## 393  Horeca Other regions    518  4180    3600    659              122
## 394  Horeca Other regions  23632  6730    3842   8620              385
## 395  Horeca Other regions  12377   865    3204   1398              149
## 396  Horeca Other regions   9602  1316    1263   2921              841
## 397  Retail Other regions   4515 11991    9345   2644             3378
## 398  Horeca Other regions  11535  1666    1428   6838               64
## 399  Horeca Other regions  11442  1032     582   5390               74
## 400  Horeca Other regions   9612   577     935   1601              469
## 401  Horeca Other regions   4446   906    1238   3576              153
## 402  Horeca Other regions  27167  2801    2128  13223               92
## 403  Horeca Other regions  26539  4753    5091    220               10
## 404  Horeca Other regions  25606 11006    4604    127              632
## 405  Horeca Other regions  18073  4613    3444   4324              914
## 406  Horeca Other regions   6884  1046    1167   2069              593
## 407  Horeca Other regions  25066  5010    5026   9806             1092
## 408  Retail Other regions   7362 12844   18683   2854             7883
## 409  Retail Other regions   8257  3880    6407   1646             2730
## 410  Horeca Other regions   8708  3634    6100   2349             2123
## 411  Horeca Other regions   6633  2096    4563   1389             1860
## 412  Horeca Other regions   2126  3289    3281   1535              235
## 413  Horeca Other regions     97  3605   12400     98             2970
## 414  Horeca Other regions   4983  4859    6633  17866              912
## 415  Horeca Other regions   5969  1990    3417   5679             1135
## 416  Retail Other regions   7842  6046    8552   1691             3540
## 417  Retail Other regions   4389 10940   10908    848             6728
```

```
## 418  Horeca Other regions   5065  5499   11055    364             3485
## 419  Retail Other regions    660  8494   18622    133             6740
## 420  Horeca Other regions   8861  3783    2223    633             1580
## 421  Horeca Other regions   4456  5266   13227     25             6818
## 422  Retail Other regions  17063  4847    9053   1031             3415
## 423  Horeca Other regions  26400  1377    4172    830              948
## 424  Retail Other regions  17565  3686    4657   1059             1803
## 425  Retail Other regions  16980  2884   12232    874             3213
## 426  Horeca Other regions  11243  2408    2593  15348              108
## 427  Horeca Other regions  13134  9347   14316   3141             5079
## 428  Horeca Other regions  31012 16687    5429  15082              439
## 429  Horeca Other regions   3047  5970    4910   2198              850
## 430  Horeca Other regions   8607  1750    3580     47               84
## 431  Horeca Other regions   3097  4230   16483    575              241
## 432  Horeca Other regions   8533  5506    5160  13486             1377
## 433  Horeca Other regions  21117  1162    4754    269             1328
## 434  Horeca Other regions   1982  3218    1493   1541              356
## 435  Horeca Other regions  16731  3922    7994    688             2371
## 436  Horeca Other regions  29703 12051   16027  13135              182
## 437  Horeca Other regions  39228  1431     764   4510               93
## 438  Retail Other regions  14531 15488   30243    437            14841
## 439  Horeca Other regions  10290  1981    2232   1038              168
## 440  Horeca Other regions   2787  1698    2510     65              477
##     Delicatessen Cluster
## 1           1338       1
## 2           1776       1
## 3           7844       1
## 4           1788       1
## 5           5185       2
## 6           1451       1
## 7            545       1
## 8           2566       1
## 9            750       1
## 10          2098       1
## 11          1744       1
## 12           497       1
## 13          2931       2
## 14           602       1
## 15          2168       2
## 16           412       1
## 17          1080       1
## 18          4478       1
## 19          3181       1
## 20           501       1
## 21          2124       1
## 22           569       1
## 23          4334       2
## 24         16523       3
## 25          5778       2
## 26            57       1
## 27           833       1
## 28           518       1
## 29          5206       3
## 30           823       2
```

```
## 31          2963       1
## 32           985       1
## 33           405       1
## 34          1083       2
## 35           395       1
## 36           436       1
## 37          4626       2
## 38           714       1
## 39           433       3
## 40          2916       2
## 41          5864       2
## 42          2802       1
## 43            46       1
## 44            72       3
## 45            65       1
## 46          4985       3
## 47          1452       3
## 48          6465       3
## 49          1476       1
## 50          1163       3
## 51          2162       1
## 52           301       1
## 53          1278       2
## 54           224       1
## 55          1333       2
## 56          1130       1
## 57          1340       3
## 58          1282       1
## 59           436       1
## 60          1603       1
## 61           225       1
## 62          2017       3
## 63           964       1
## 64          1295       1
## 65          1145       1
## 66          1423       3
## 67            27       1
## 68           834       1
## 69          3095       1
## 70           144       1
## 71          1365       1
## 72         14472       2
## 73           181       1
## 74           648       1
## 75          1780       1
## 76           975       1
## 77           894       1
## 78          1009       3
## 79           167       1
## 80          1653       1
## 81           529       1
## 82           156       1
## 83          2342       1
## 84           772       1
```

```
## 85           120       1
## 86          2944       3
## 87           903       3
## 88         14351       2
## 89            46       1
## 90          3178       2
## 91           360       1
## 92          1117       1
## 93          5130       3
## 94          2698       2
## 95           244       1
## 96           709       1
## 97           217       1
## 98            63       1
## 99           132       1
## 100          323       1
## 101         3029       1
## 102         1838       1
## 103         1386       1
## 104         2498       2
## 105          548       1
## 106         1378       1
## 107         1831       1
## 108         1438       1
## 109         1236       1
## 110            3       1
## 111         1647       1
## 112         1519       1
## 113         2708       1
## 114         1561       1
## 115         1266       1
## 116          610       1
## 117          222       1
## 118         1160       1
## 119          933       1
## 120          635       1
## 121         1136       1
## 122          255       1
## 123          860       1
## 124          143       1
## 125         1621       2
## 126          918       2
## 127          483       1
## 128         2749       1
## 129            3       1
## 130         1819       2
## 131          911       1
## 132          310       1
## 133          328       1
## 134          396       1
## 135          537       1
## 136          326       1
## 137         1542       1
## 138           36       1
```

```
## 139         3271       1
## 140          929       1
## 141         2616       1
## 142         1450       2
## 143            3       2
## 144          318       1
## 145          201       1
## 146         4430       3
## 147          520       1
## 148          526       1
## 149          434       1
## 150         1440       2
## 151         1067       1
## 152         1774       1
## 153          184       1
## 154         1627       1
## 155            8       1
## 156         2153       3
## 157         3182       1
## 158          418       1
## 159         1682       1
## 160          303       1
## 161         2157       1
## 162         2233       1
## 163          610       1
## 164          446       3
## 165          238       1
## 166         2379       3
## 167         3637       1
## 168          693       1
## 169          429       1
## 170          750       1
## 171          323       1
## 172         6250       3
## 173          707       1
## 174          716       3
## 175         1442       1
## 176         1697       1
## 177          230       2
## 178         2631       1
## 179          395       1
## 180         2165       1
## 181         2794       1
## 182         8550       2
## 183          156       1
## 184        47943       2
## 185           11       1
## 186          247       1
## 187          727       1
## 188            3       1
## 189          404       1
## 190         1856       1
## 191           64       1
## 192           84       1
```

```
## 193          409       1
## 194          666       1
## 195         1142       1
## 196         1755       1
## 197         2876       2
## 198         1468       1
## 199          697       1
## 200          347       1
## 201          731       3
## 202         1681       3
## 203         6854       2
## 204           18       1
## 205         1328       1
## 206          710       3
## 207          285       1
## 208          120       1
## 209          806       1
## 210          797       3
## 211         2100       1
## 212         2870       3
## 213         1775       1
## 214         1215       1
## 215          791       1
## 216         2388       1
## 217          674       3
## 218         1158       1
## 219         6372       1
## 220          130       1
## 221          749       1
## 222          239       1
## 223          375       1
## 224         1360       1
## 225          659       1
## 226          786       1
## 227         1553       1
## 228          689       1
## 229          203       1
## 230          980       1
## 231         2137       1
## 232          490       1
## 233          834       2
## 234            7       1
## 235         2563       1
## 236          295       1
## 237          225       1
## 238         1215       1
## 239          216       1
## 240         2253       2
## 241         2564       2
## 242         1047       2
## 243          578       1
## 244         2398       1
## 245         1970       1
## 246         2784       1
```

```
## 247          610       1
## 248          659       1
## 249          572       1
## 250          291       1
## 251          710       1
## 252         5121       3
## 253         1693       1
## 254         1391       2
## 255         3265       1
## 256          615       2
## 257          373       1
## 258          987       1
## 259         3321       2
## 260          818       2
## 261          548       1
## 262          287       1
## 263          655       1
## 264          411       1
## 265         1265       1
## 266         3636       1
## 267         2563       1
## 268         3628       1
## 269          698       1
## 270          204       1
## 271           56       1
## 272         1550       1
## 273         1040       1
## 274         1824       2
## 275         1153       1
## 276          379       1
## 277         2503       2
## 278          139       1
## 279         1409       1
## 280         1721       1
## 281         1104       1
## 282         2079       1
## 283         1404       2
## 284         1384       2
## 285         2406       2
## 286           18       2
## 287          128       1
## 288         1027       1
## 289          258       1
## 290           22       2
## 291         1522       1
## 292          686       1
## 293         1060       1
## 294          741       1
## 295         1854       1
## 296          254       1
## 297          898       1
## 298          531       1
## 299         1037       1
## 300          259       1
```

```
## 301         2005       1
## 302          172       3
## 303          555       1
## 304           59       1
## 305         2410       3
## 306          211       1
## 307         1543       3
## 308          925       1
## 309          656       1
## 310          806       3
## 311         1117       1
## 312          117       2
## 313          142       3
## 314          297       1
## 315         1233       1
## 316         3508       1
## 317         1059       1
## 318         1637       1
## 319           51       1
## 320         1625       3
## 321          834       1
## 322         1113       1
## 323          229       1
## 324          573       1
## 325         1092       2
## 326         5609       2
## 327          834       1
## 328          522       1
## 329         1534       1
## 330          739       1
## 331         1043       1
## 332         1102       3
## 333         2602       1
## 334         1215       3
## 335         3486       1
## 336         2139       2
## 337          778       1
## 338          868       1
## 339          550       1
## 340         1942       1
## 341         1371       1
## 342         2158       1
## 343         1328       1
## 344           37       3
## 345          379       1
## 346          303       1
## 347         1115       1
## 348         1022       2
## 349          665       1
## 350          445       3
## 351          995       1
## 352         3137       3
## 353          195       1
## 354         1111       1
```

```
## 355         2341       1
## 356          127       1
## 357          548       1
## 358          110       1
## 359         4100       1
## 360          776       1
## 361          503       1
## 362          405       1
## 363          712       1
## 364          314       1
## 365          468       1
## 366         3105       1
## 367          447       1
## 368          342       1
## 369          558       1
## 370          296       1
## 371         2235       2
## 372          790       1
## 373         4829       1
## 374         3113       1
## 375          686       1
## 376           70       1
## 377         1426       1
## 378         1242       2
## 379         1114       1
## 380          179       1
## 381          270       2
## 382          532       1
## 383         2893       2
## 384          361       1
## 385         5120       1
## 386         1068       1
## 387          967       1
## 388          961       1
## 389          406       1
## 390          684       1
## 391         1000       1
## 392         1827       1
## 393          654       1
## 394          819       2
## 395          452       1
## 396          290       1
## 397         2213       1
## 398          743       1
## 399          247       1
## 400          375       1
## 401         1014       1
## 402         1902       2
## 403          340       2
## 404          288       2
## 405          715       1
## 406          378       1
## 407          960       2
## 408          553       3
```

```
## 409          344       1
## 410         5137       1
## 411         1892       1
## 412         4365       1
## 413           62       1
## 414         2435       1
## 415          290       1
## 416         1874       1
## 417          993       1
## 418         1063       1
## 419          776       1
## 420         1521       1
## 421         1393       1
## 422         1784       1
## 423         1218       2
## 424          668       1
## 425          249       1
## 426         1886       1
## 427         1894       1
## 428         1163       2
## 429          317       1
## 430         2501       1
## 431         2080       1
## 432         1498       1
## 433          395       1
## 434         1449       1
## 435          838       1
## 436         2204       2
## 437         2346       2
## 438         1867       3
## 439         2125       1
## 440           52       1
```

Looking into the spending patterns within each cluster, we found some patterns regarding both average gross spending and average categorical spending. Cluster 1: Low spending clients Cluster 2: High spending clients who prefer Milk, Grocery and Detergents Paper Cluster 3: Middle spending customer who prefer Fresh and Frozen

```
# Pivot data to create a single continuous/numerical factor using gather function
wholesale_pivot <- gather(data = wholesale_clustered, key = 'Product_Category', value = 'Spend', Fresh:

cluster_avgspending_overall <- wholesale_pivot %>% group_by(Cluster) %>%
  summarise(TotalSpend = sum(Spend),AvgSpend = mean(Spend))
cluster_avgspending_overall
```

```
## # A tibble: 3 x 3
##   Cluster TotalSpend AvgSpend
##     <int>      <int>    <dbl>
## 1       1    7870520    3881.
## 2       2    3635143    9932.
## 3       3    3113837   12658.
```

```
cluster_avgspending_category <- wholesale_pivot %>% group_by(Cluster,Product_Category) %>%
  summarise(TotalSpend = sum(Spend),AvgSpend = mean(Spend))
```

```
## `summarise()` has grouped output by 'Cluster'. You can override using the
## `.groups` argument.
```

```
cluster_avgspending_category
```

```
## # A tibble: 18 x 4
## # Groups:   Cluster [3]
##    Cluster Product_Category TotalSpend AvgSpend
##      <int> <chr>                 <int>    <dbl>
##  1       1 Delicatessen         376383    1114.
##  2       1 Detergents_Paper     623802    1846.
##  3       1 Fresh               2767137    8187.
##  4       1 Frozen               829656    2455.
##  5       1 Grocery             1897512    5614.
##  6       1 Milk                1376030    4071.
##  7       2 Delicatessen         199483    3270.
##  8       2 Detergents_Paper      66195    1085.
##  9       2 Fresh               2162832   35456.
## 10       2 Frozen               442238    7250.
## 11       2 Grocery              398744    6537.
## 12       2 Milk                 365651    5994.
## 13       3 Delicatessen          95077    2319.
## 14       3 Detergents_Paper     577860   14094.
## 15       3 Fresh                350162    8541.
## 16       3 Frozen                79756    1945.
## 17       3 Grocery             1202306   29325.
## 18       3 Milk                 808676   19724.
```

```
# Grouping data by channel and region (normalized BEFORE grouping)
# All groupings normalized
wholesale_groups_norm = wholesale_norm %>% group_by(Channel,Region) %>% summarise(Fresh_mean_norm = mean
wholesale_groups_norm %>% mutate_if(is.numeric, round, digits=2)
```

```
## # A tibble: 6 x 8
##   Channel Region        Fresh_mean_norm Milk_mean_norm Grocery_mean_norm
##   <chr>   <chr>                   <dbl>          <dbl>             <dbl>
## 1 Horeca  Lisbon                   0.12           0.05              0.04
## 2 Horeca  Oporto                   0.1            0.03              0.05
## 3 Horeca  Other regions            0.12           0.05              0.04
## 4 Retail  Lisbon                   0.05           0.15              0.2
## 5 Retail  Oporto                   0.06           0.12              0.18
## 6 Retail  Other regions            0.09           0.15              0.17
## # i 3 more variables: Frozen_mean_norm <dbl>, Detergents_Paper_mean_norm <dbl>,
## #   Delicatessen_mean_norm <dbl>
```

```
# Group by Channel normalized
wholesale_Channel_norm = wholesale_groups_norm %>% group_by(Channel) %>% summarise(Fresh_mean_norm = mea
wholesale_Channel_norm %>% mutate_if(is.numeric, round, digits=2)
```

```
## # A tibble: 2 x 7
##   Channel Fresh_mean_norm Milk_mean_norm Grocery_mean_norm Frozen_mean_norm
##   <chr>             <dbl>          <dbl>             <dbl>            <dbl>
## 1 Horeca             0.11           0.04              0.04             0.07
## 2 Retail             0.07           0.14              0.18             0.03
## # i 2 more variables: Detergents_Paper_mean_norm <dbl>,
## #   Delicatessen_mean_norm <dbl>
```

```
# Group by Region normalized
wholesale_Region_norm = wholesale_groups_norm %>% group_by(Region) %>% summarise(Fresh_mean_norm = mean
wholesale_Region_norm %>% mutate_if(is.numeric, round, digits=2)
```

```
## # A tibble: 3 x 7
##   Region       Fresh_mean_norm Milk_mean_norm Grocery_mean_norm Frozen_mean_norm
##   <chr>                  <dbl>          <dbl>             <dbl>            <dbl>
## 1 Lisbon                  0.08           0.1               0.12             0.05
## 2 Oporto                  0.08           0.08              0.11             0.06
## 3 Other regio~            0.11           0.1               0.11             0.04
## # i 2 more variables: Detergents_Paper_mean_norm <dbl>,
## #   Delicatessen_mean_norm <dbl>
```

```
# Non-normalized results
# All groupings
wholesale_groups = wholesale %>% group_by(Channel,Region) %>% summarise(Fresh_mean = mean(Fresh),Milk_m
wholesale_groups %>% mutate_if(is.numeric, round, digits=2)
```

```
## # A tibble: 6 x 8
##   Channel Region        Fresh_mean Milk_mean Grocery_mean Frozen_mean
##   <chr>   <chr>              <dbl>     <dbl>        <dbl>       <dbl>
## 1 Horeca  Lisbon            12902.     3870.        4026.       3127.
## 2 Horeca  Oporto           11651.     2304.        4396.       5745.
## 3 Horeca  Other regions     13878.     3487.        3887.       3657.
## 4 Retail  Lisbon             5200     10784        18472.       2584.
## 5 Retail  Oporto             7290.     9191.       16326.       1541.
## 6 Retail  Other regions      9832.    10981.       15954.       1513.
## # i 2 more variables: Detergents_Paper_mean <dbl>, Delicatessen_mean <dbl>
```

```
# Group by Channel
wholesale_Channel = wholesale_groups %>% group_by(Channel) %>% summarise(Fresh_mean = mean(Fresh_mean),
wholesale_Channel %>% mutate_if(is.numeric, round, digits=2)
```

```
## # A tibble: 2 x 7
##   Channel Fresh_mean Milk_mean Grocery_mean Frozen_mean Detergents_Paper_mean
##   <chr>        <dbl>     <dbl>        <dbl>       <dbl>                 <dbl>
## 1 Horeca      12810.     3220.        4103.       4176.                  740.
## 2 Retail       7440.    10319.       16917.       1879.                 7845.
## # i 1 more variable: Delicatessen_mean <dbl>
```

```
# Group by Region
wholesale_Region = wholesale_groups %>% group_by(Region) %>% summarise(Fresh_mean = mean(Fresh_mean), M
wholesale_Region %>% mutate_if(is.numeric, round, digits=2)
```

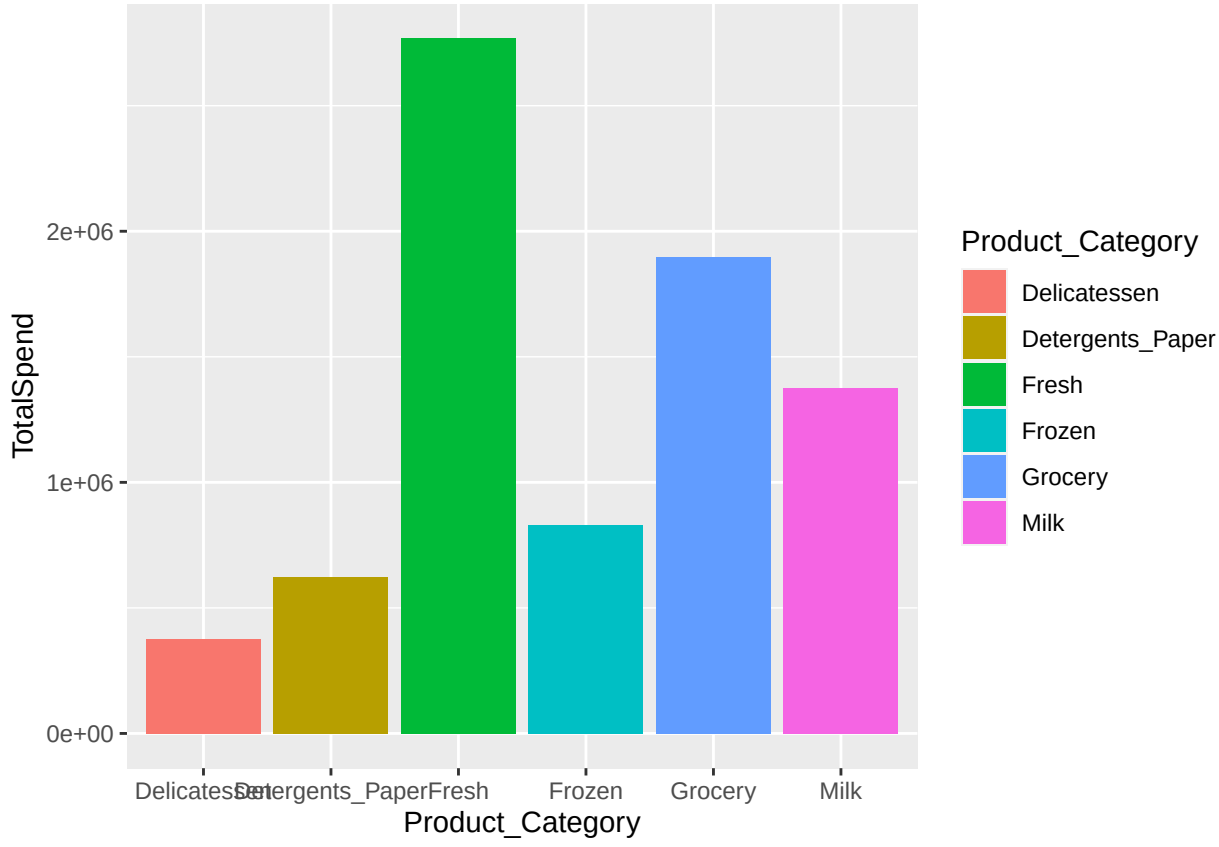
```
## # A tibble: 3 x 7
##   Region     Fresh_mean Milk_mean Grocery_mean Frozen_mean Detergents_Paper_mean
##   <chr>           <dbl>     <dbl>        <dbl>       <dbl>                 <dbl>
## 1 Lisbon          9051.     7327.       11249.       2856.                 4588.
## 2 Oporto          9470.     5748.       10361.       3643.                 4446.
## 3 Other reg~     11855.     7234         9920.       2585.                 3843.
## # i 1 more variable: Delicatessen_mean <dbl>
```

Analyse spending patterns across different clusters using visualisations! CLUSTER 1 summary

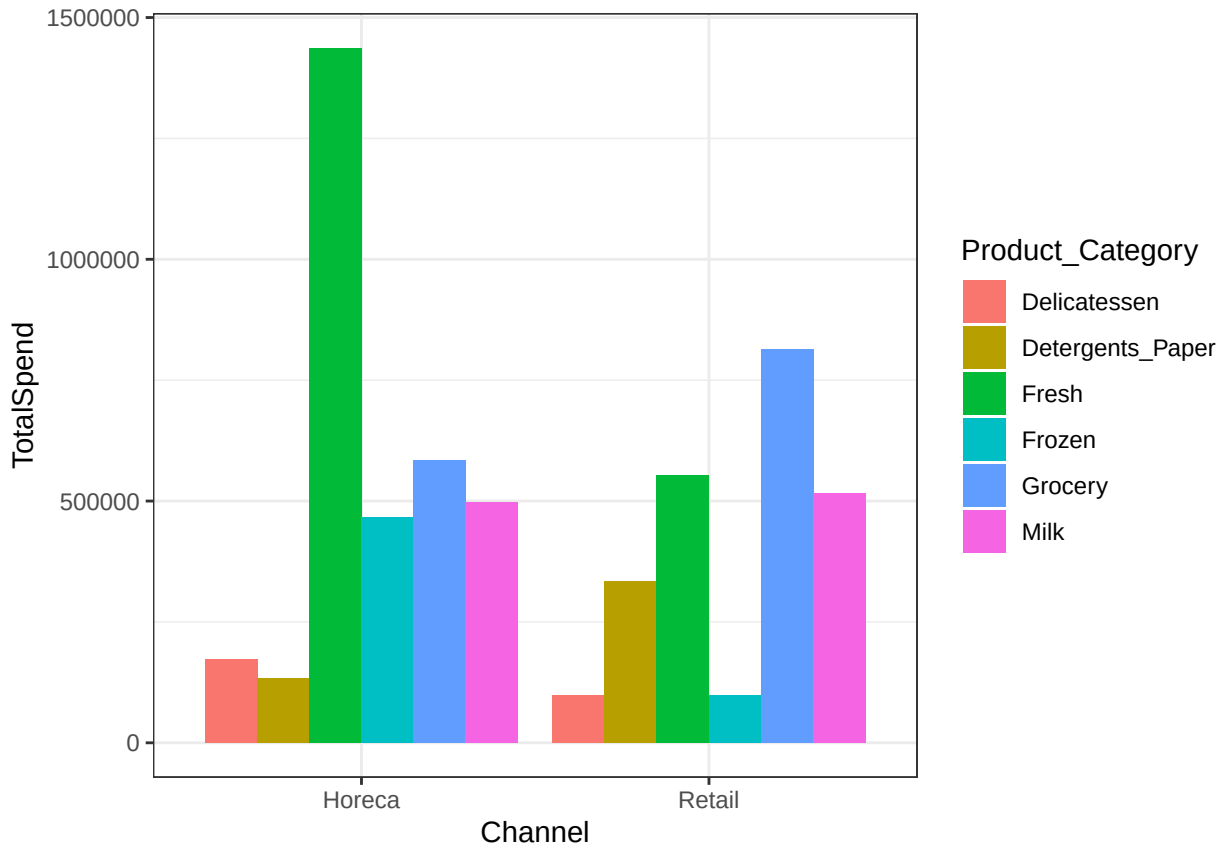
```
## Cluster 1

cluster1 <- wholesale_pivot[ which(wholesale_pivot$Cluster==1), ]
cluster1 <- subset(cluster1, select = -c(Cluster))

cluster1_summary <- cluster1 %>% group_by(Channel, Region, Product_Category) %>%
  summarise(TotalSpend = sum(Spend), AvgSpend = mean(Spend))
```

```
## `summarise()` has grouped output by 'Channel', 'Region'. You can override using
## the `.groups` argument.
```

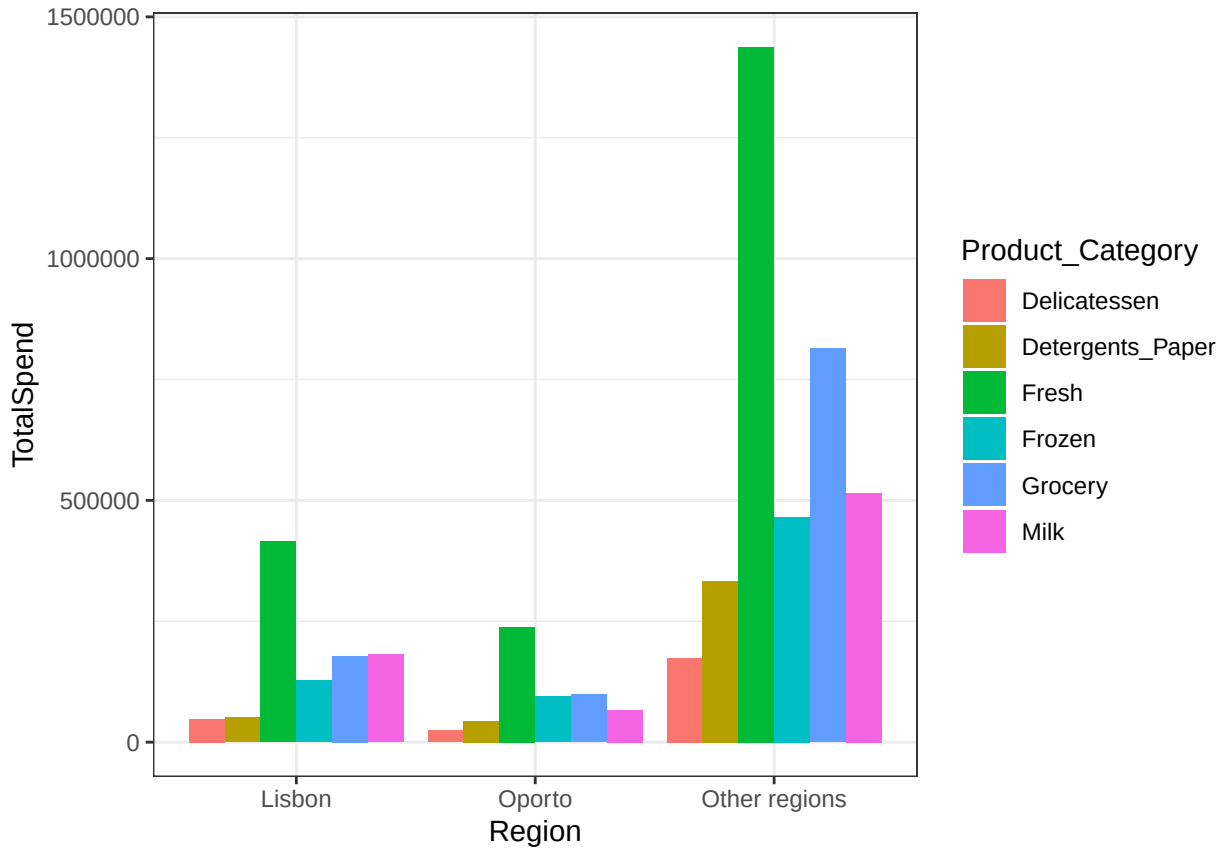
```
# Plot 1 (Total Spend across categories)
cluster1_summary %>% ggplot(aes(x = Product_Category, y = TotalSpend, fill = Product_Category)) + geom_
```



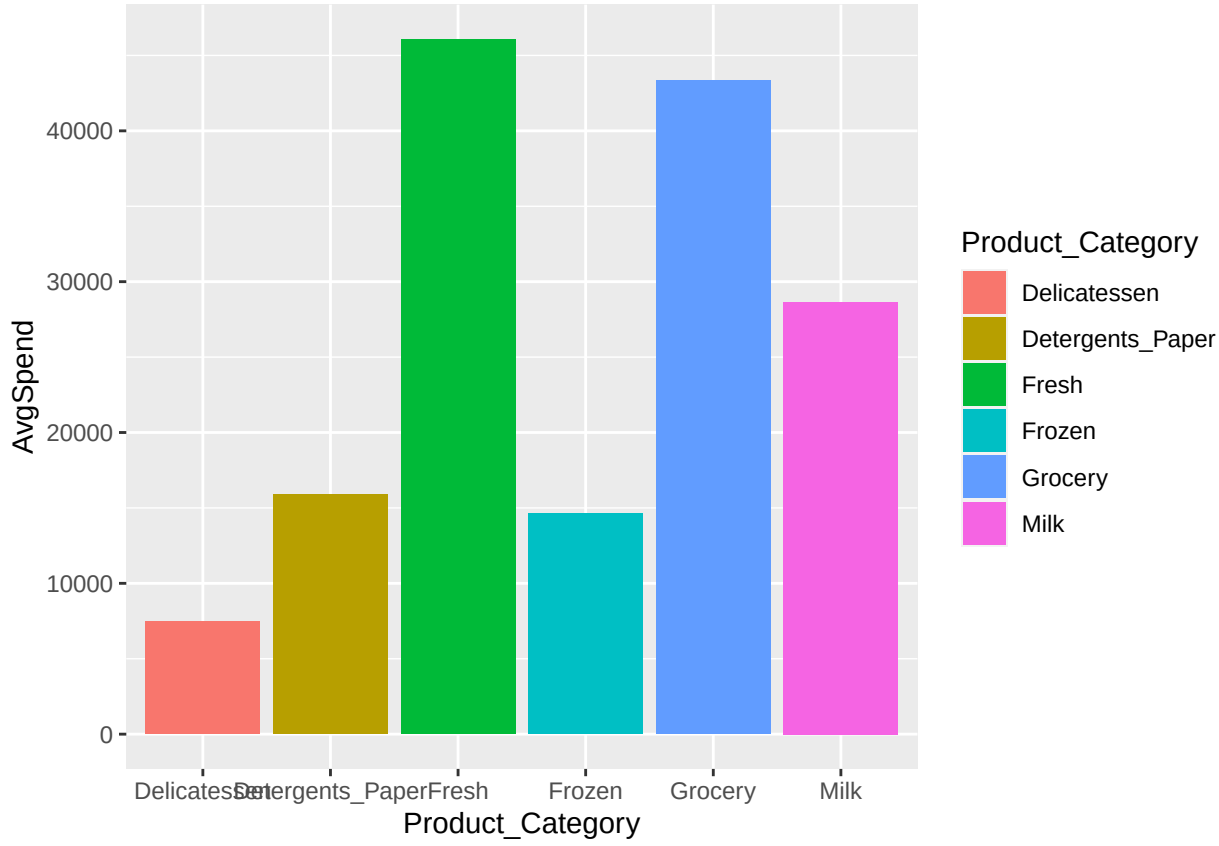
```r
# Plot 2 (Total Spend across channels and categories)
cluster1_summary %>% ggplot(aes(Channel, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'ident
```



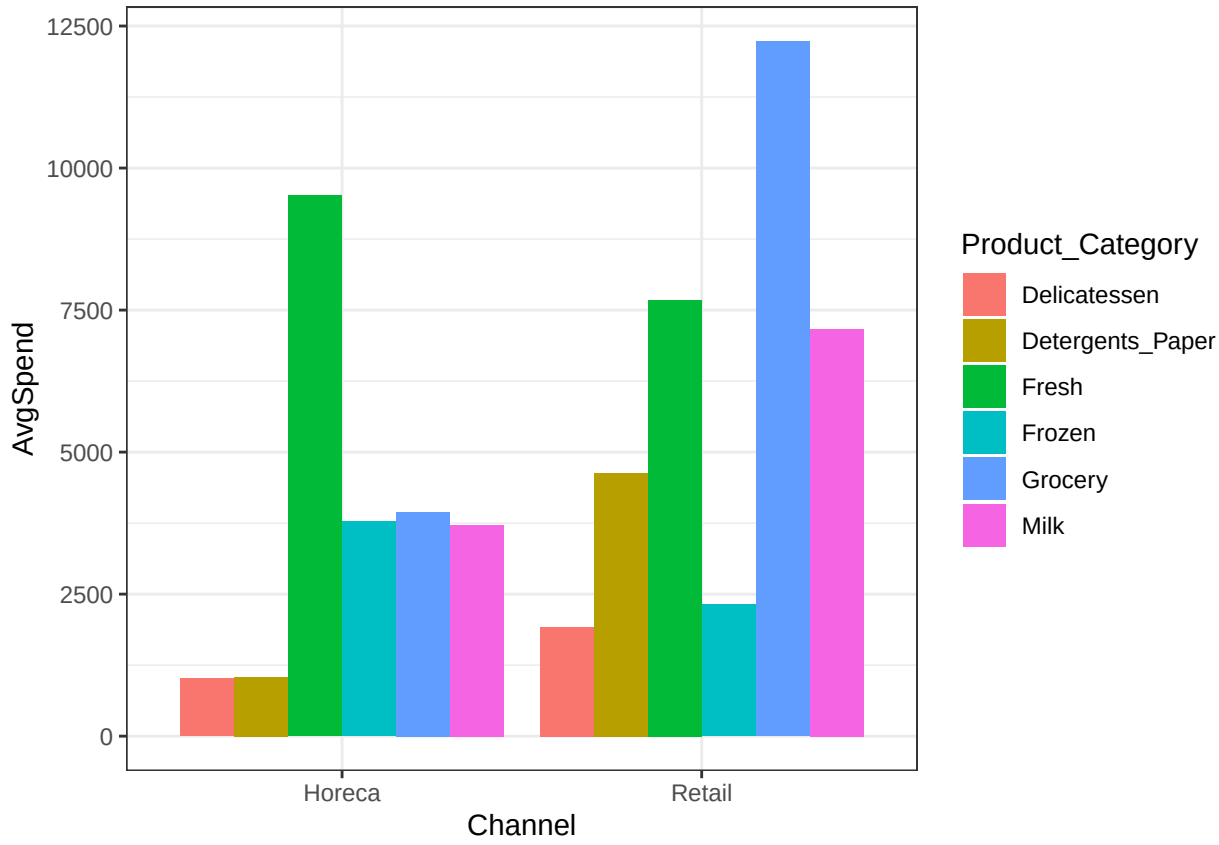

```r
# Plot 3 (Total Spend across regions and categories)
cluster1_summary %>% ggplot(aes(Region, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'identi
```



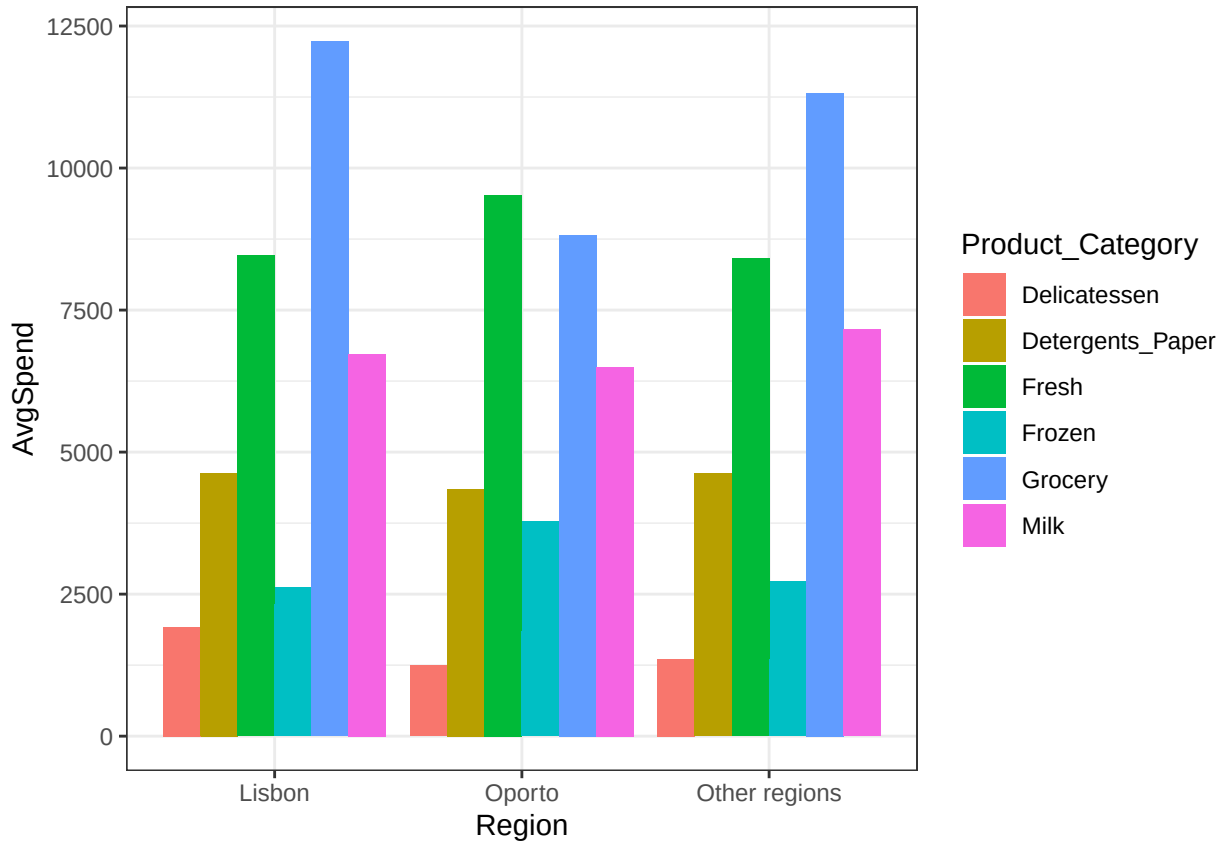

```
# Plot 4 (Average Spend across categories)
cluster1_summary %>% ggplot(aes(x = Product_Category, y = AvgSpend, fill = Product_Category)) + geom_col
```




```
# Plot 5 (Average Spend across channels and categories)
cluster1_summary %>% ggplot(aes(Channel, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```




```
# Plot 6 (Average Spend across regions and categories)
cluster1_summary %>% ggplot(aes(Region, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```

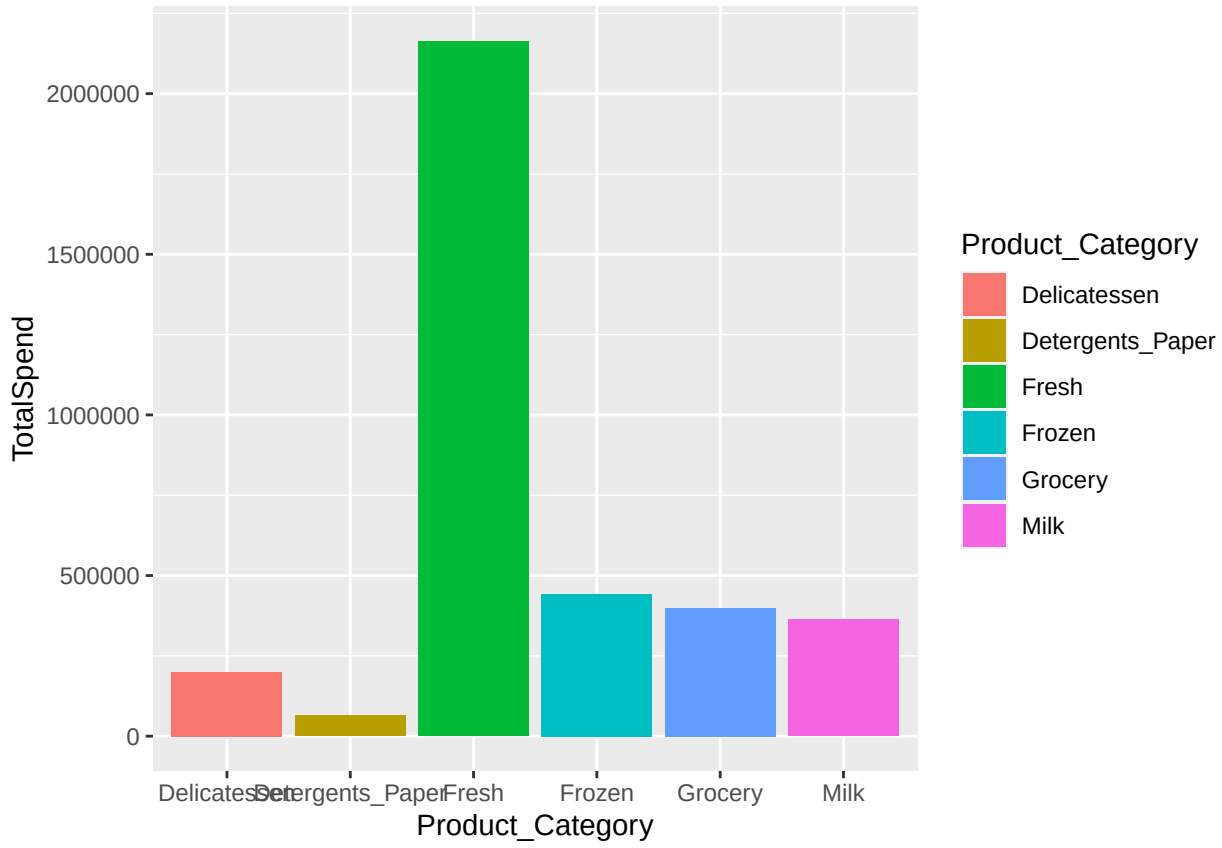


CLUSTER 2 summary

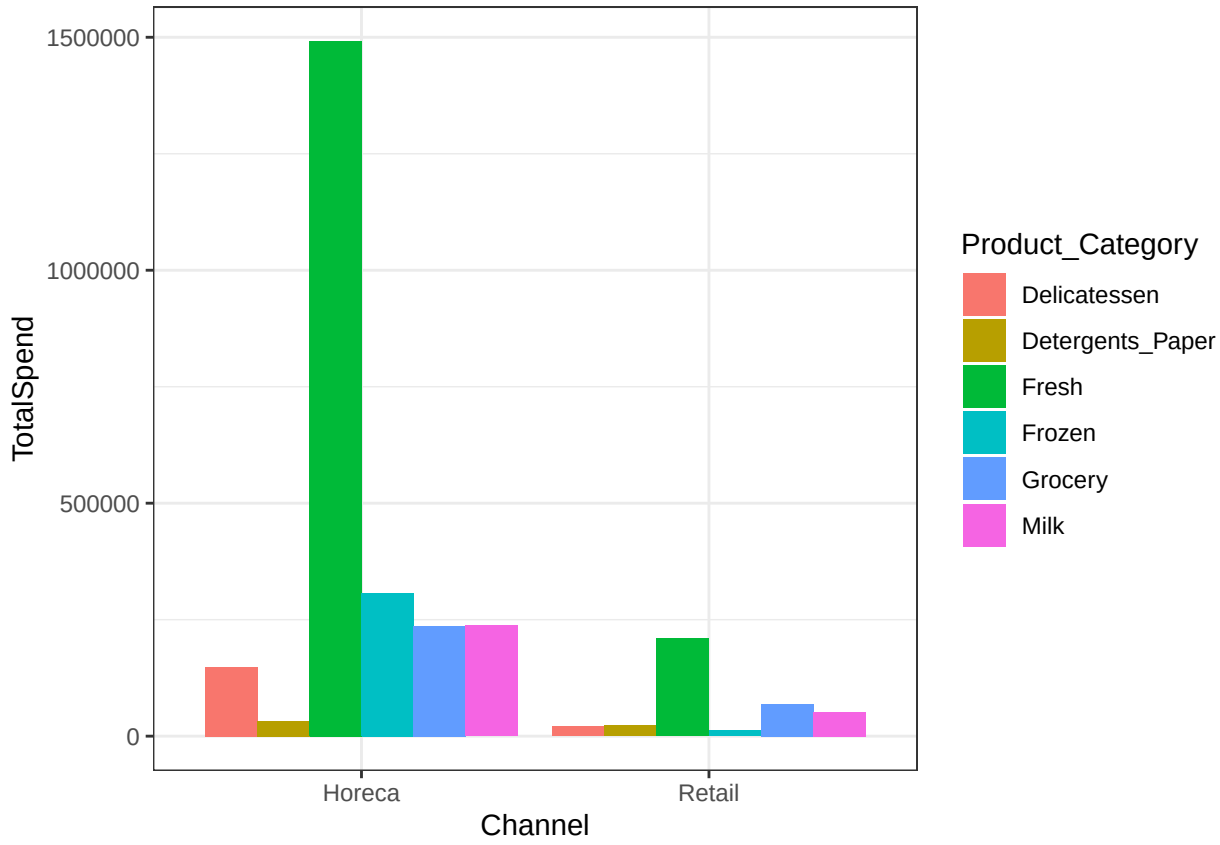
```
## Cluster 2

cluster2 <- wholesale_pivot[ which(wholesale_pivot$Cluster==2), ]
cluster2 <- subset(cluster2, select = -c(Cluster))

cluster2_summary <- cluster2 %>% group_by(Channel, Region, Product_Category) %>%
  summarise(TotalSpend = sum(Spend), AvgSpend = mean(Spend))
```

```
## `summarise()` has grouped output by 'Channel', 'Region'. You can override using
## the `.groups` argument.
```

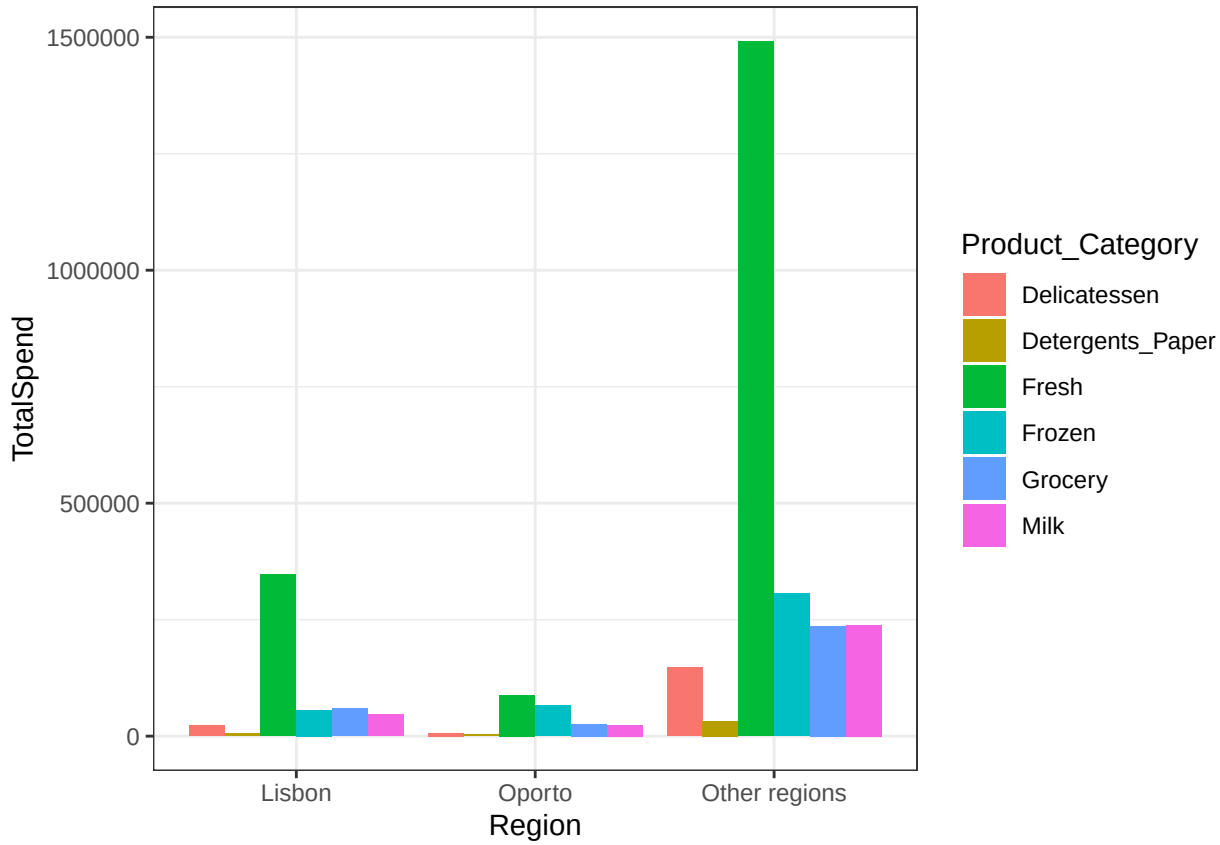
```
# Plot 1 (Total Spend across categories)
cluster2_summary %>% ggplot(aes(x = Product_Category, y = TotalSpend, fill = Product_Category)) + geom_
```



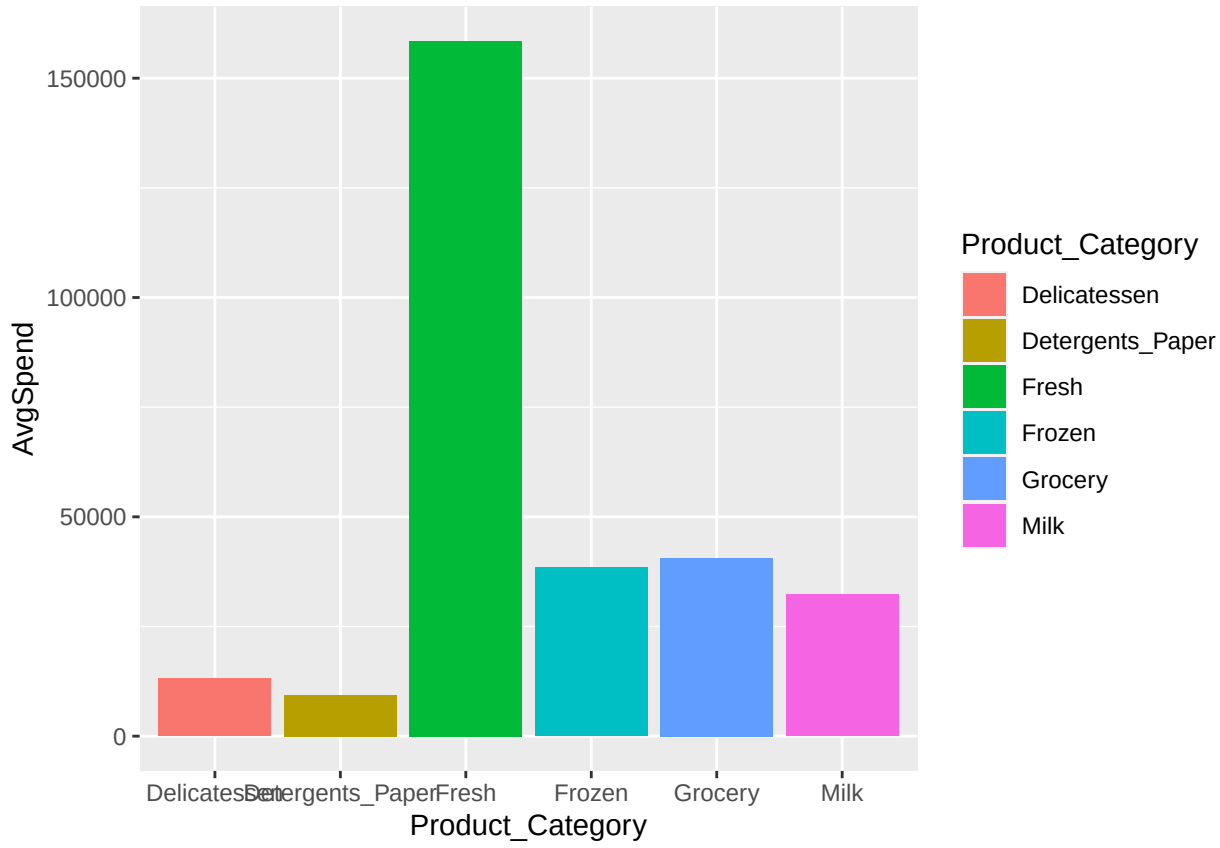

```
# Plot 2 (Total Spend across channels and categories)
cluster2_summary %>% ggplot(aes(Channel, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'ident
```



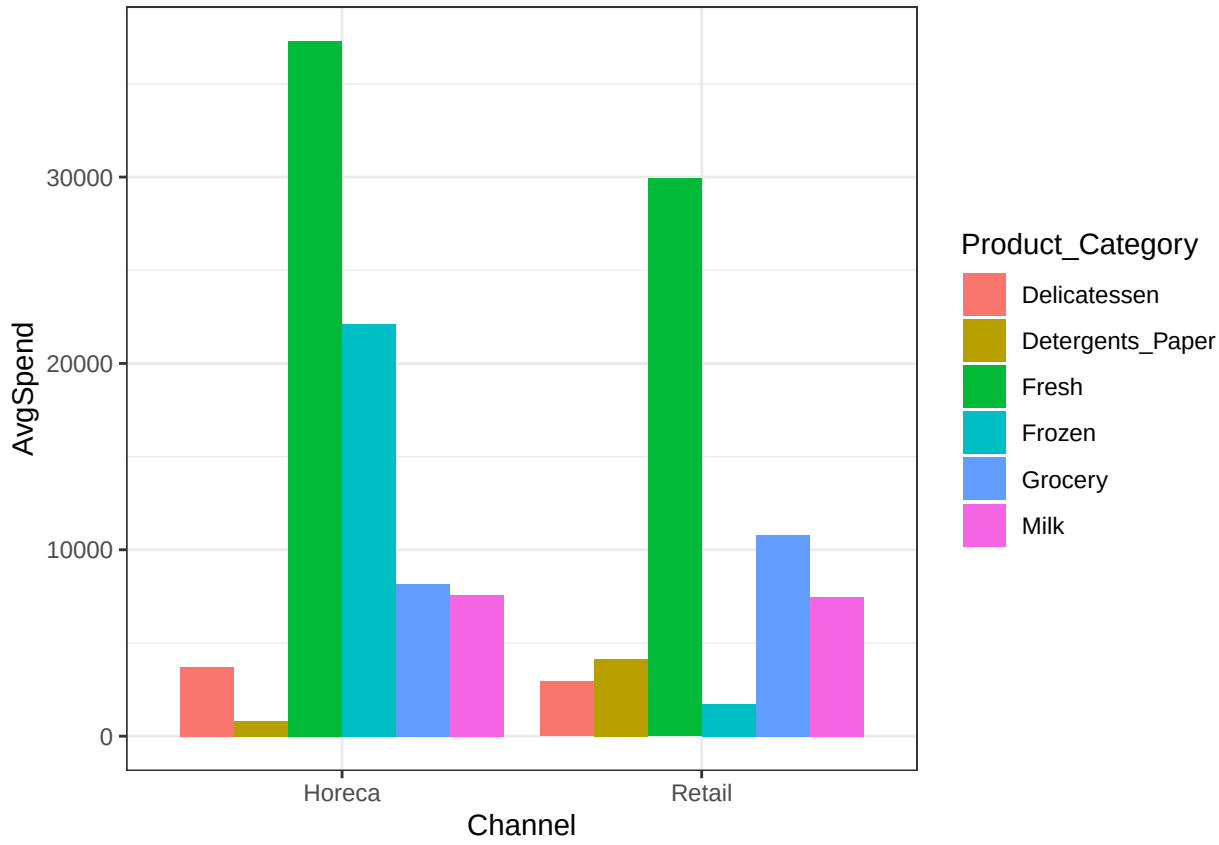

```
# Plot 3 (Total Spend across regions and categories)
cluster2_summary %>% ggplot(aes(Region, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'identi
```



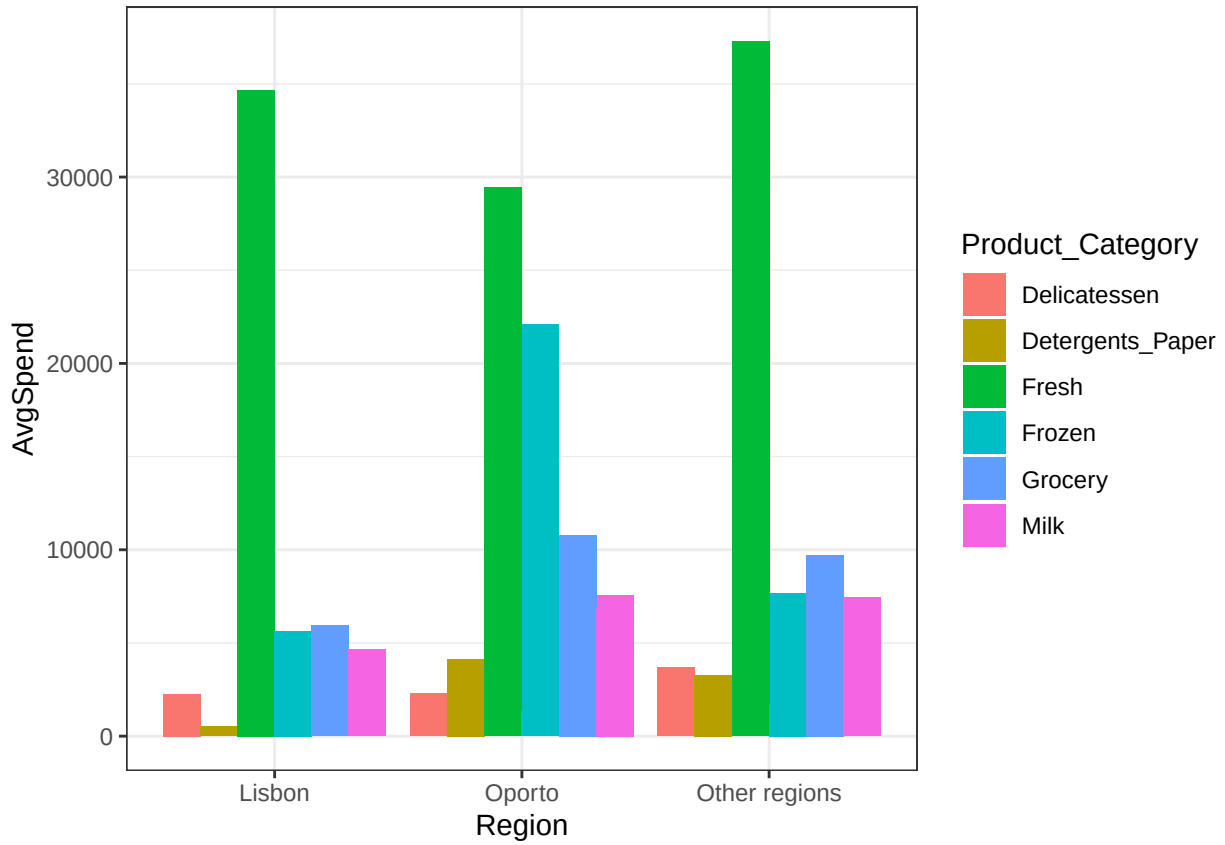

```
# Plot 4 (Average Spend across categories)
cluster2_summary %>% ggplot(aes(x = Product_Category, y = AvgSpend, fill = Product_Category)) + geom_col
```




```
# Plot 5 (Average Spend across channels and categories)
cluster2_summary %>% ggplot(aes(Channel, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```




```
# Plot 6 (Average Spend across regions and categories)
cluster2_summary %>% ggplot(aes(Region, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```

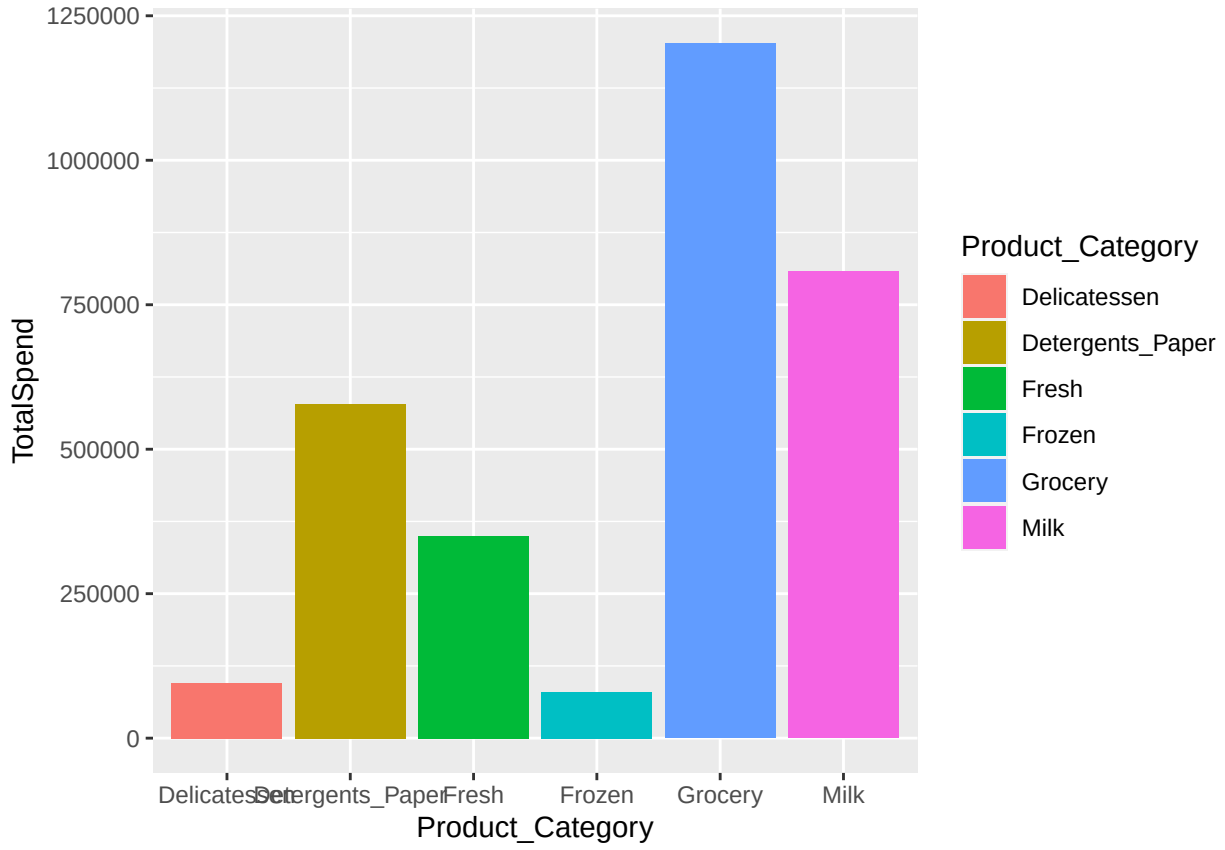


CLUSTER 3 summary

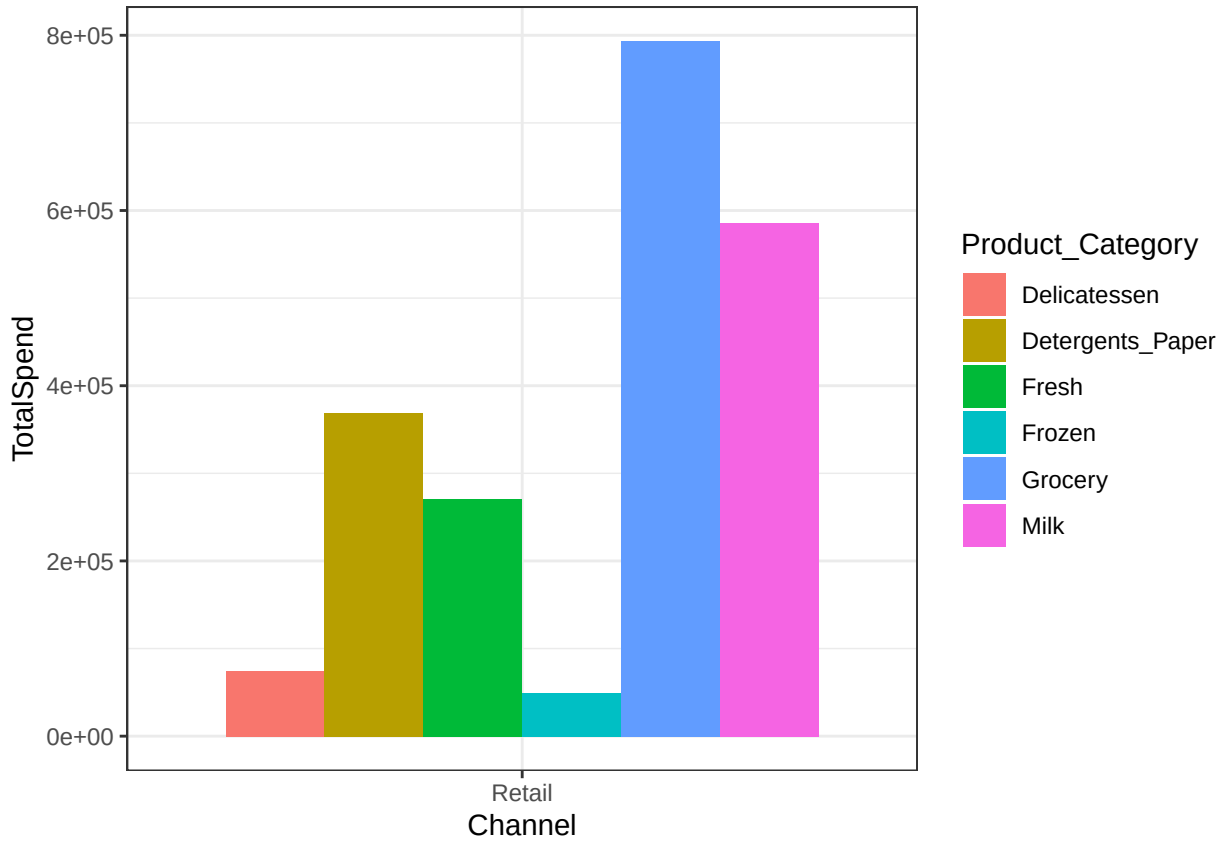
```
## Cluster 3

cluster3 <- wholesale_pivot[ which(wholesale_pivot$Cluster==3), ]
cluster3 <- subset(cluster3, select = -c(Cluster))

cluster3_summary <- cluster3 %>% group_by(Channel, Region, Product_Category) %>%
  summarise(TotalSpend = sum(Spend), AvgSpend = mean(Spend))
```

```
## `summarise()` has grouped output by 'Channel', 'Region'. You can override using
## the `.groups` argument.
```

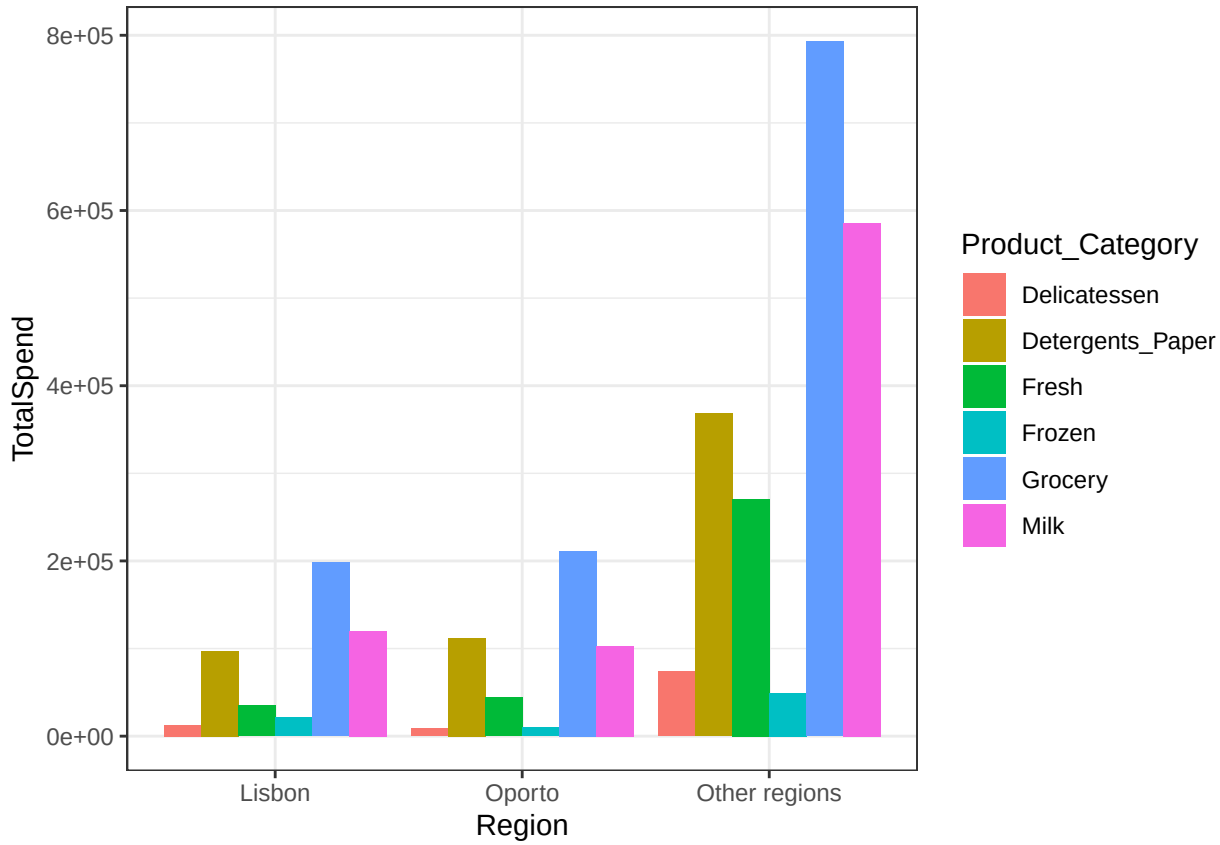
```
# Plot 1 (Total Spend across categories)
cluster3_summary %>% ggplot(aes(x = Product_Category, y = TotalSpend, fill = Product_Category)) + geom_
```



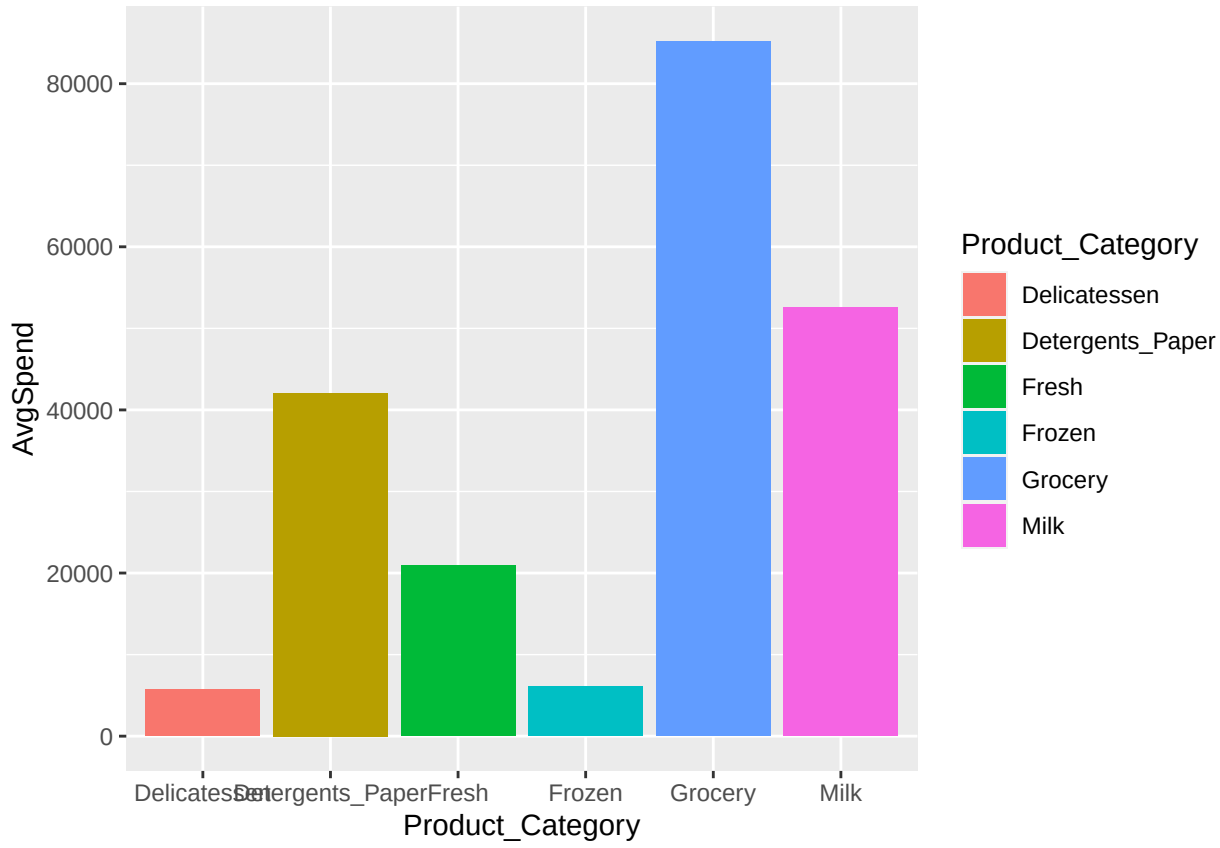

```
# Plot 2 (Total Spend across channels and categories)
cluster3_summary %>% ggplot(aes(Channel, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'ident
```



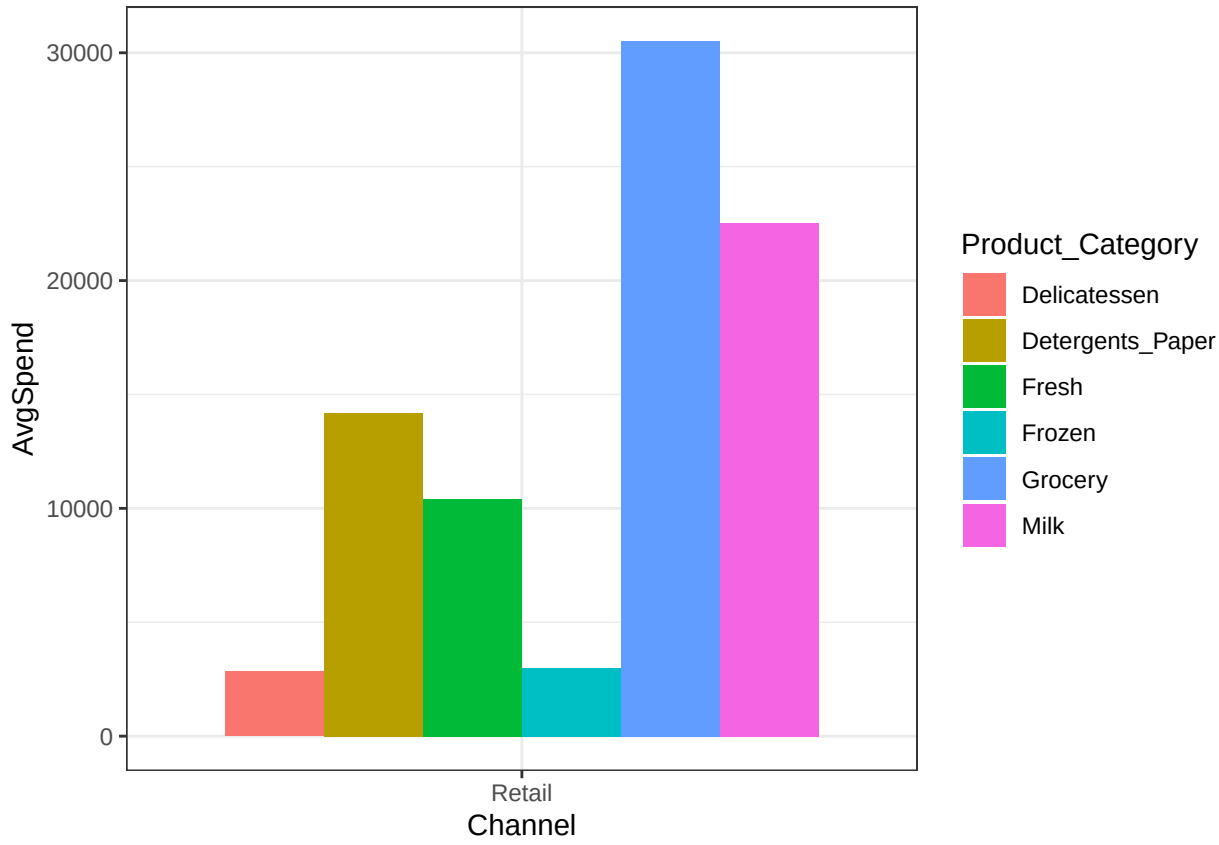

```
# Plot 3 (Total Spend across regions and categories)
cluster3_summary %>% ggplot(aes(Region, TotalSpend, fill = Product_Category)) + geom_bar(stat = 'identi
```



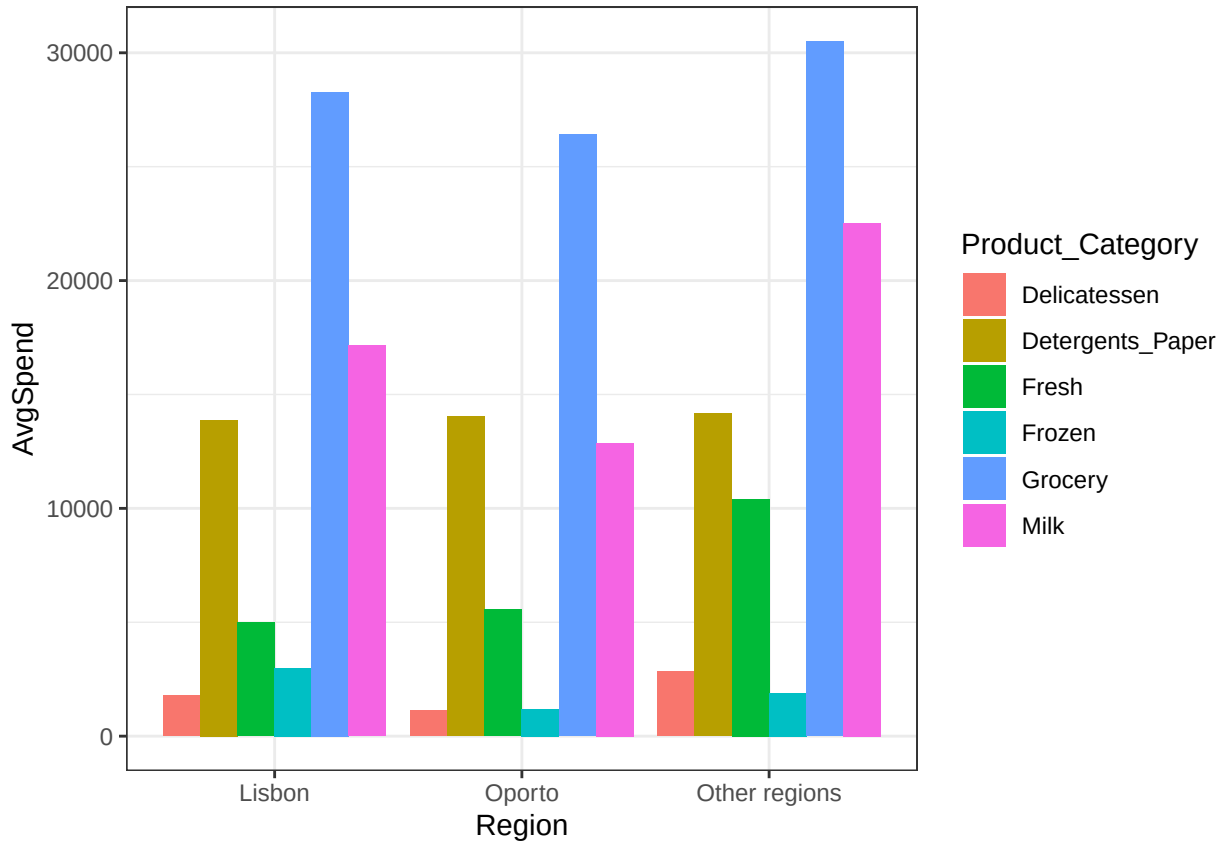

```
# Plot 4 (Average Spend across categories)
cluster3_summary %>% ggplot(aes(x = Product_Category, y = AvgSpend, fill = Product_Category)) + geom_col
```




```
# Plot 5 (Average Spend across channels and categories)
cluster3_summary %>% ggplot(aes(Channel, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```




```
# Plot 6 (Average Spend across regions and categories)
cluster3_summary %>% ggplot(aes(Region, AvgSpend, fill = Product_Category)) + geom_bar(stat = 'identity
```



Then we looked into the distribution of clusters across channels and regions. ## We can change the tables into better visualizations

```
# Composition counts for each cluster
cluster_composition = table(wholesale_clustered$Channel, wholesale_clustered$Region, wholesale_clustere
cluster_composition
```

```
## , ,  = 1
##
##
##          Lisbon Oporto Other regions
##   Horeca     49     25           171
##   Retail     11     10            72
##
## , ,  = 2
##
##
##          Lisbon Oporto Other regions
##   Horeca     10      3            40
##   Retail      0      1             7
##
## , ,  = 3
##
##
##          Lisbon Oporto Other regions
##   Horeca      0      0             0
```

```
##   Retail      7      8           26
```

```
percentages = cluster_composition
# Composition percentages for each cluster
for (i in 1:k){
  total = sum(percentages[,,i])
  percentages[,,i] = t(round(t(percentages[,,i])/total,2))
}
percentages
```

```
## , ,  = 1
##
##
##          Lisbon Oporto Other regions
##   Horeca   0.14   0.07          0.51
##   Retail   0.03   0.03          0.21
##
## , ,  = 2
##
##
##          Lisbon Oporto Other regions
##   Horeca   0.16   0.05          0.66
##   Retail   0.00   0.02          0.11
##
## , ,  = 3
##
##
##          Lisbon Oporto Other regions
##   Horeca   0.00   0.00          0.00
##   Retail   0.17   0.20          0.63
```

The data suggests the following:

CLUSTER 1 - This cluster constitutes of the clients majorly from Horeca sector (i.e. Hotels/Restaurants/Cafes) approx. 72%. These clients demonstarte high spending in the Fresh category, while the remaining 28% of the clients from Retail channel spend more on Grocery. Clients from both channels put together have spends across these 3 categories majorly - Fresh, Grocery and Milk.

CLUSTER 2 - This cluster constitutes of the clients only from Retail channel, majorly from the Other regions (63%). Their spending pattern shows huge spends in the Grocery, followed by Milk and Detergents & Paper.

CLUSTER 3 - This cluster constitutes of clients from both Horeca and Retail, majorly from Other regions (66%). These clients spend mostly in the Fresh category. The spend across other categories is minimal.