Trabalho - Métodos Matriciais e Clusterização

Hugo Muniz Albuquerque

Entendimento do Problema

Neste trabalho o objetivo será identificar se existem padrões de agrupamentos entre os diferentes clientes de um distribuidor de produtos de atacado. A partir desta análise de clusters vou poder obter uma referência para o padrão de consumo de cada cluster, como a média de compras deles, o que fornece informações para o fornecedor comprar seus estoques e direcioná-los para seus pontos de vendas em quantidades mais adequadas a cada tipo de cliente obtido nos clusters, com base na sua região e padrão de consumo.

Entendimento dos Dados

A base de dados utilizada neste trabalho foi obtida neste link https://archive.ics.uci.edu/ml/datasets/Wholesale+customers. Esta base contém os gastos anuais dos clientes em diversas categorias de produtos fresh products (produtos frescos), milk products (produtos de leite), grocery (mercearia), frozen products (produtos congelados), detergents (detergente) and paper (papel) products e delicatessen. Além disso, a base possui duas variáveis discretas, Region (1 = Lisboa, 2 = Porto e 3 = other) e Channel (1 = HRC - Hotel, Restaurants or Cafe e 2 = Retail).

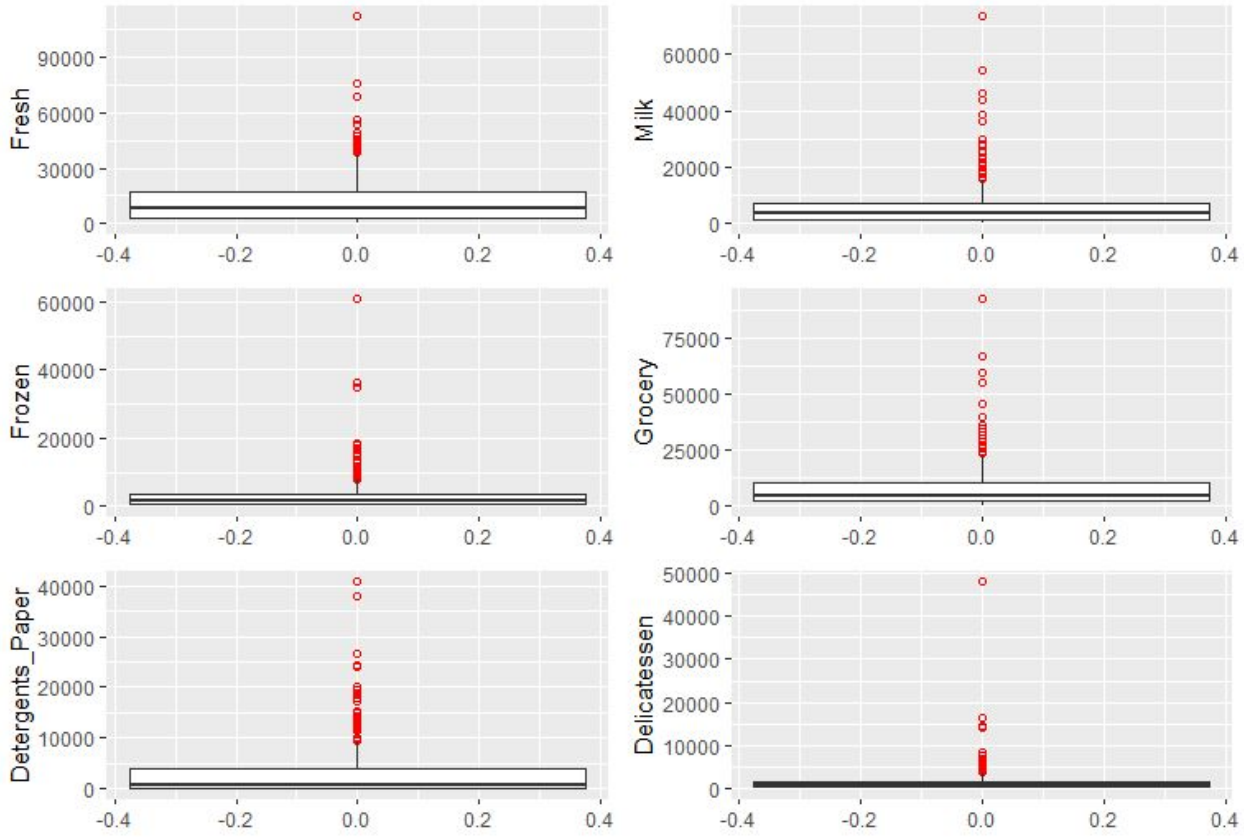
Descrições estatísticas:

Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Min. : 3	Min. : 55	Min. : 3	Min. : 25.0	Min. : 3.0	Min. : 3.0
1st Qu.: 3128	1st Qu.: 1533	1st Qu.: 2153	1st Qu.: 742.2	1st Qu.: 256.8	1st Qu.: 408.2
Median : 8504	Median : 3627	Median : 4756	Median : 1526.0	Median : 816.5	Median : 965.5
Mean : 12000	Mean : 5796	Mean : 7951	Mean : 3071.9	Mean : 2881.5	Mean : 1524.9
3rd Qu.: 16934	3rd Qu.: 7190	3rd Qu.:10656	3rd Qu.: 3554.2	3rd Qu.: 3922.0	3rd Qu.: 1820.2
Max. :112151	Max. :73498	Max. :92780	Max. :60869.0	Max. :40827.0	Max. :47943.0

De acordo com as estatísticas descritivas obtidas pela função summary, os valores mínimos e máximos estão muito discrepantes, o que pode gerar um viés nos dados. No caso da categoria Fresh, o mínimo é 3 e o máximo chega a 112151, para Milk o mínimo é 55 e o máximo 73498, e nas outras também é observado essa alta variabilidade. Para contornar esse possível problema foi realizado uma análise de outliers por meio do gráfico de boxplot. Assim, o resultado ficará mais fiel ao comportamento do cliente médio que é o padrão buscado neste trabalho.

A seguir foi realizado a análise de outliers:

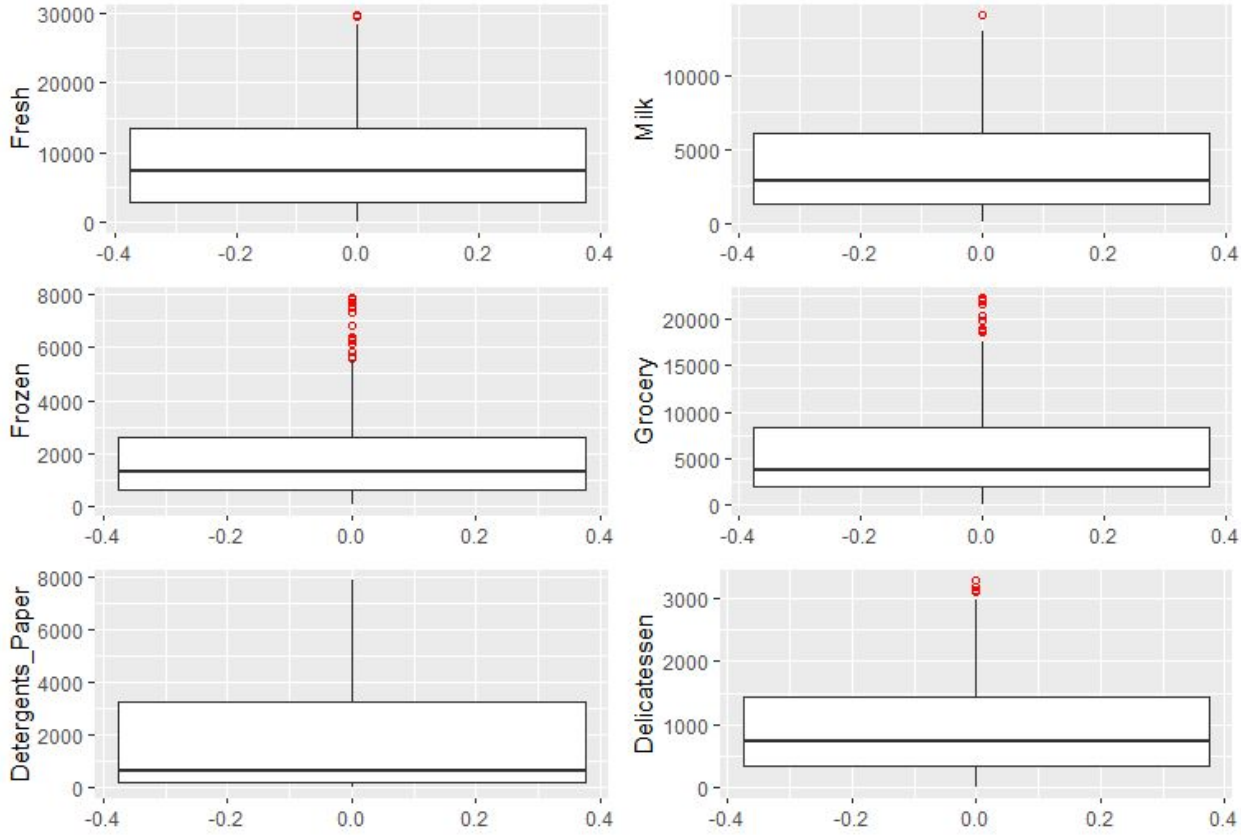
Gráfico boxplot de cada variável



De acordo com o gráfico acima, todas as variáveis possuem outliers. Para a variável Fresh foram removidos todas as observações acima de 30.000, para a variável Milk todas acima de 15.000, Frozen todas acima de 8.000, Grocery todas acima de 23.000, Detergents_paper todas acima de 8.000 e Delicatessen todas acima de 3.500.

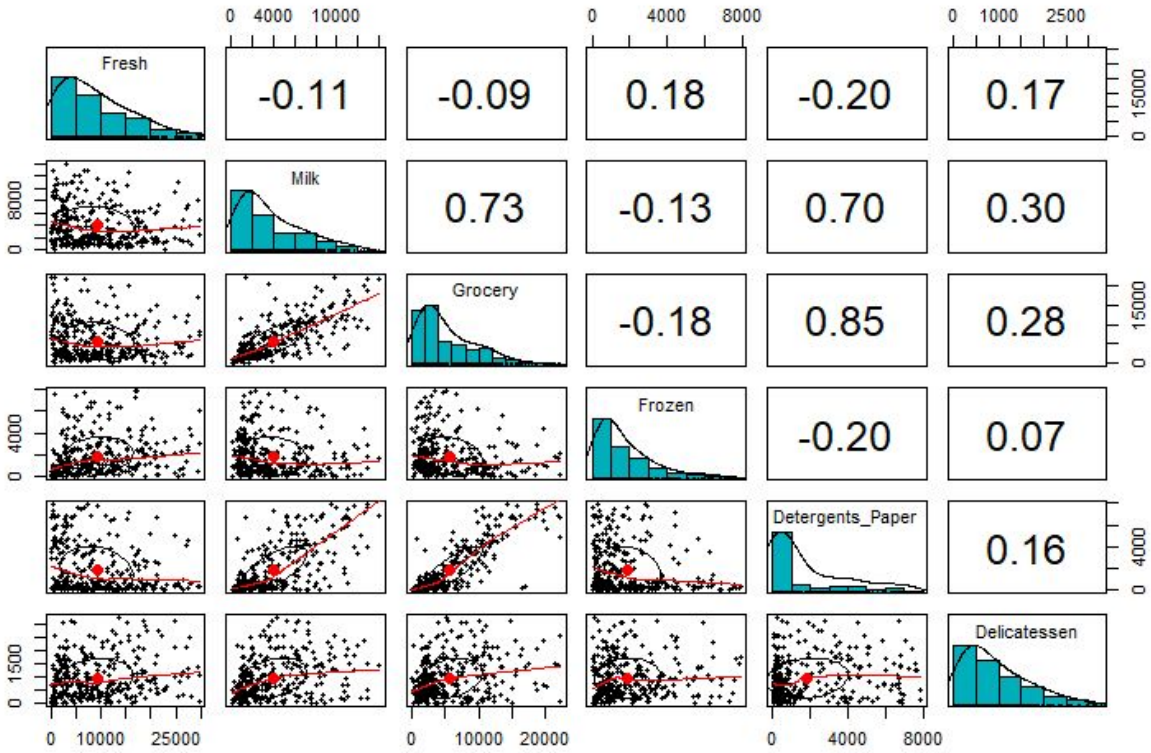
O próximo gráfico boxplot foi feito sobre a base com os outliers removidos.

Gráfico boxplot sem outliers



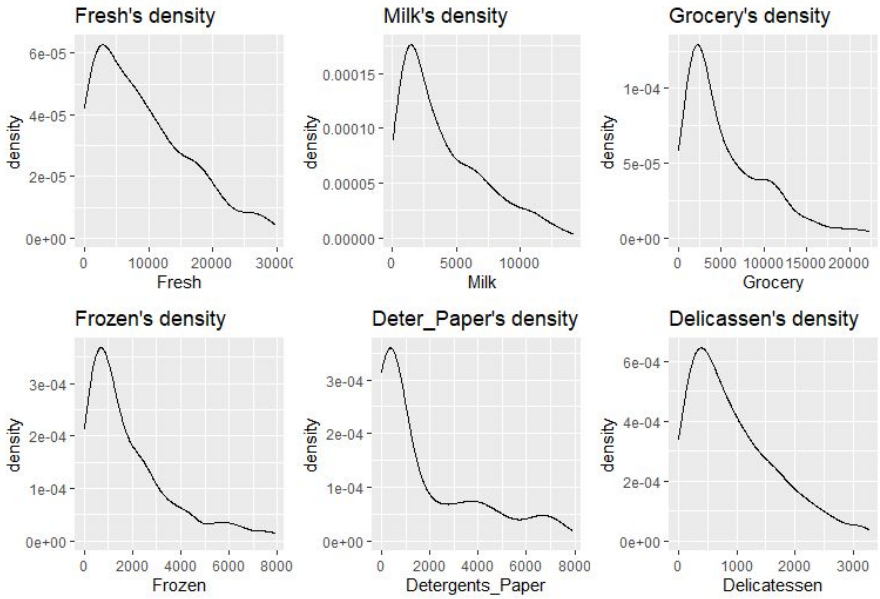
Ao todo foram removidos 123 observações (Fresh = 24, Milk = 23, Frozen = 29, Grocery = 10, Detergents_paper = 13 e Delicatessen = 14). Mesmo após a remoção dos outliers, ainda é possível ver que existem pontos representados como outliers, porém não foram removidos para que o modelo não leve em consideração apenas as observações próximas da média e mediana.

Gráfico das correlações entre as variáveis



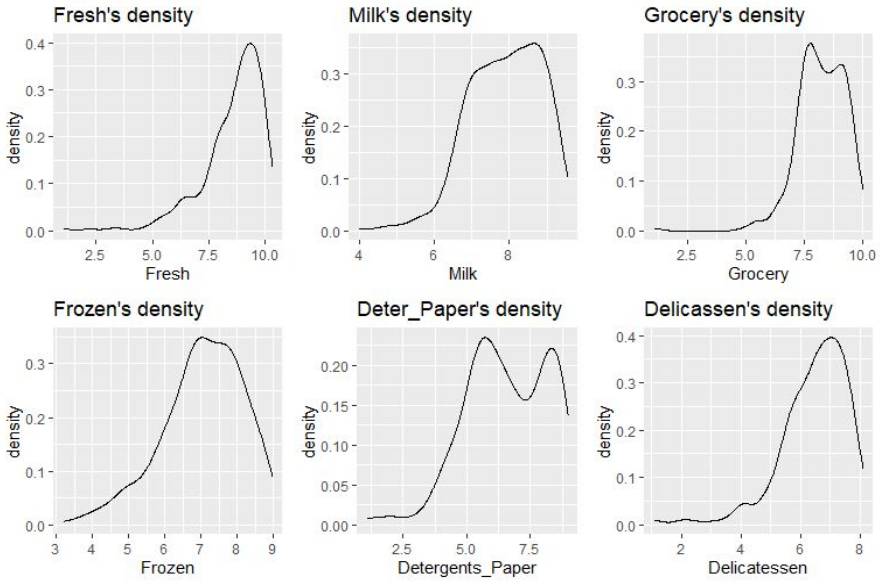
As variáveis que possuem alta correlação entre si são: Milk e Grocery (0.73), Milk e Detergents_paper (0.70), Grocery e Detergents_paper (0.85), os três fazem parte das compras feitas por clientes do canal 2. Uma média taxa de correlação é vista entre Milk e Delicassen (0.30) e entre Grocery e Delicatessen (0.28). Assim, a correlação aparece como um indicador de que os produtos com correlação forte entre si são comprados juntos por esses grupos específicos.

Gráficos da Densidade da Distribuição de cada produto sem Log



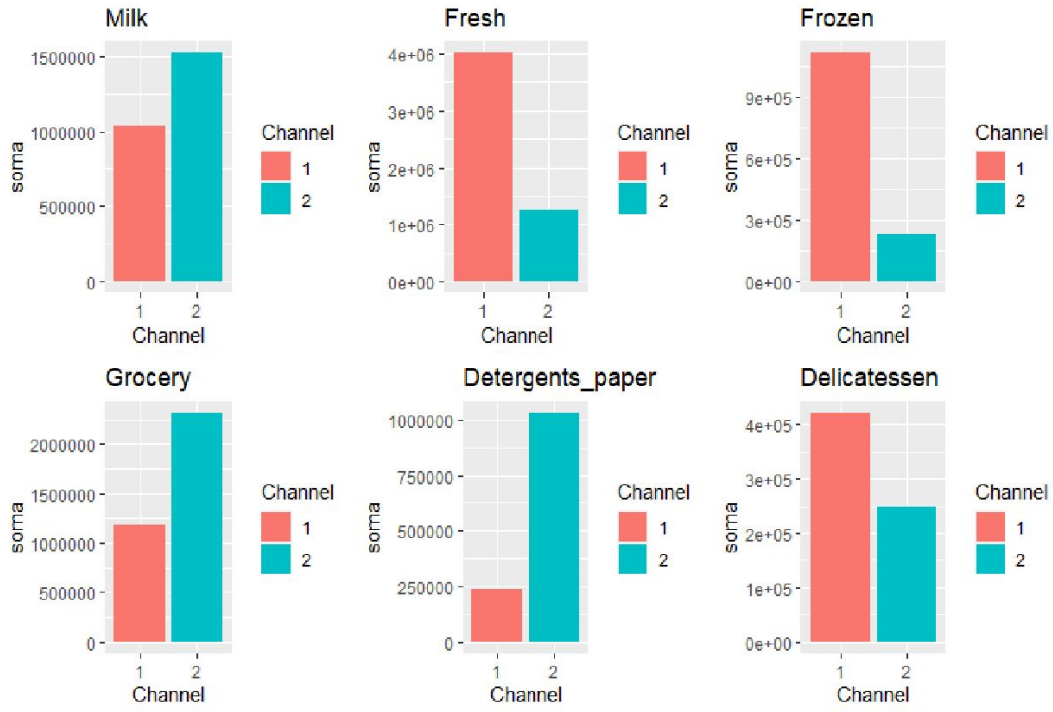
O gráfico de densidade da distribuição revela que as variáveis distribuições estão assimétricas à esquerda, o que pode enviesar a análise. Portanto, foi aplicado o log para tentar contornar o problema.

Gráficos da Densidade da Distribuição de cada produto com Log



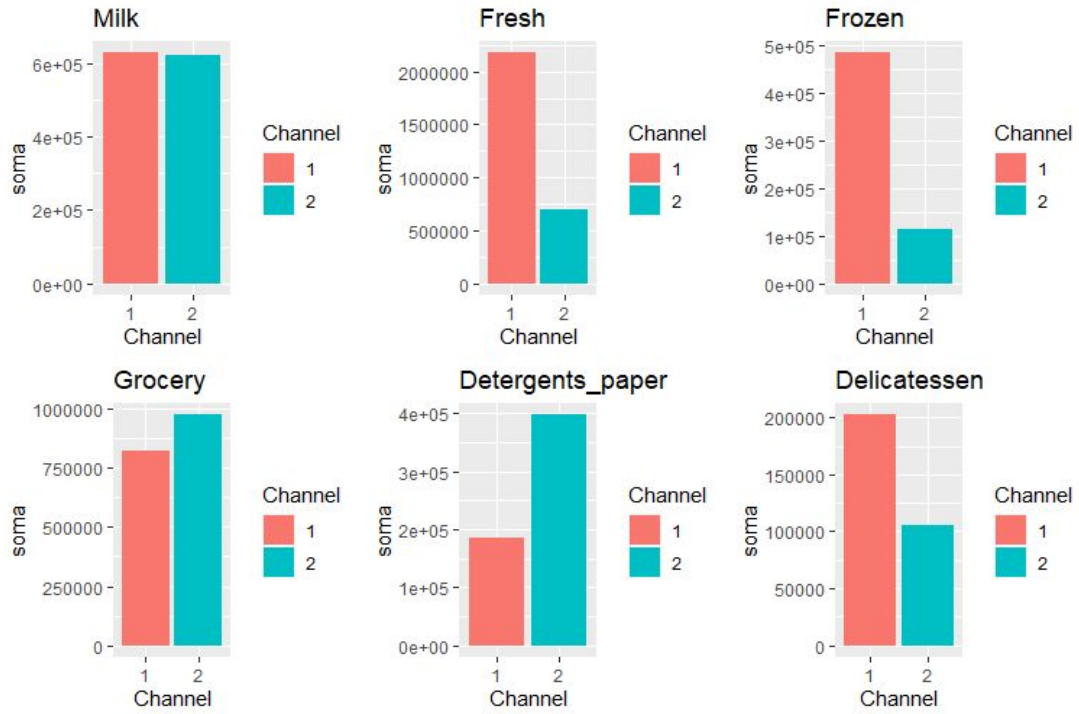
Após a aplicação do log a densidade da distribuição ficou mais simétrica, apesar de agora parecer que está mais para a direita. Neste gráfico pode-se perceber que os produtos Fresh e Delicatessen possuem maior magnitude.

Gráficos da quantidade de produtos vendidos por Canais de Venda antes da remoção de outliers



Antes da remoção dos outliers as maiores quantidades para o canal 1 foram para as categorias Fresh, Frozen e Delicatessen, e para o canal 2, Grocery, Detergents_paper e Milk.

Gráficos da quantidade de produtos vendidos por Canais de Venda depois da remoção de outliers



Após a remoção dos outliers os gráficos mostraram quase o mesmo padrão, o canal 1 apresentou maiores quantidades das categorias Fresh, Frozen e Delicatessen, enquanto que o canal 2, Grocery e Detergents_paper. A categoria Milk ficou praticamente empatada para os dois canais. No geral, os dados ficaram levemente mais balanceados para cada categoria de produtos.

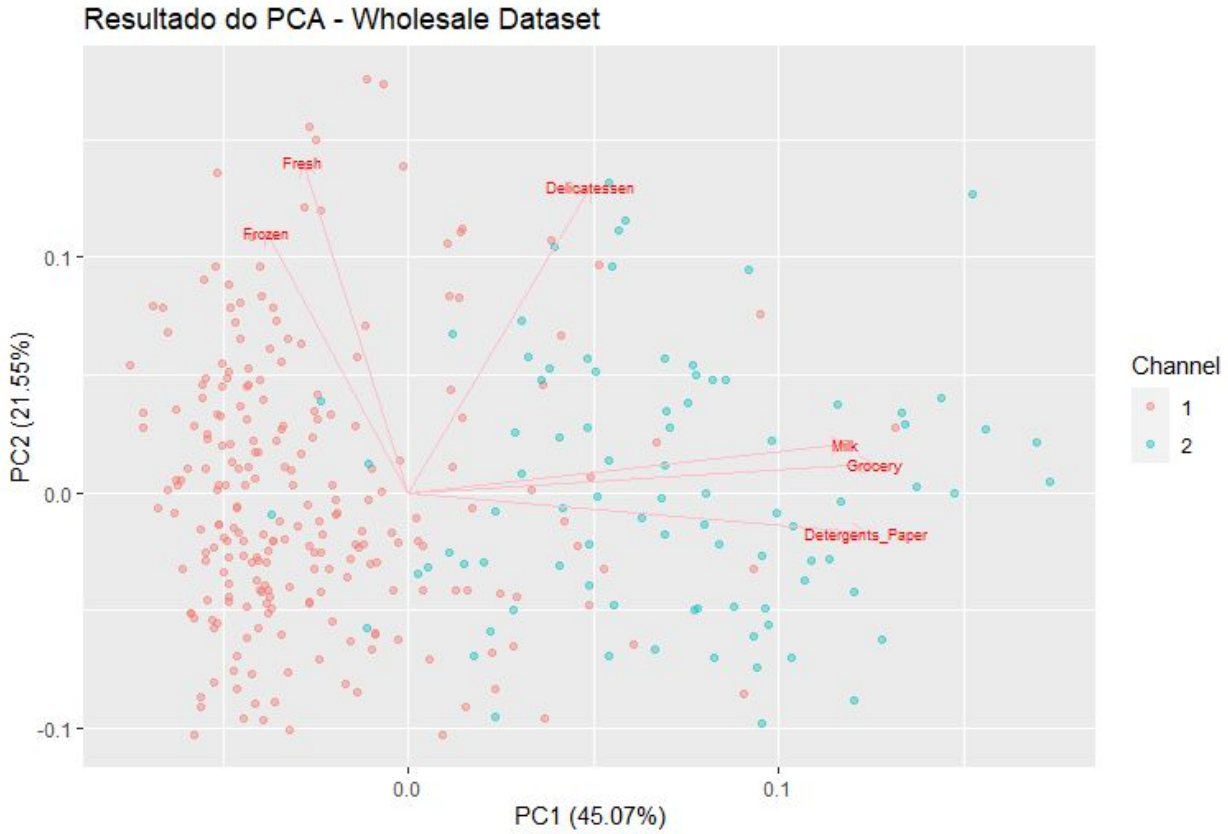
Nestes gráficos fica evidente que os produtos Milk, Grocery e Detergents_paper são preferidos pelo Channel 2 (Retail) e os produtos Fresh, Frozen e Delicatessen são preferidos do Channel 1 (HRC). Este padrão pode refletir no agrupamento dos clusters, já que as preferências são fortes. Vale ressaltar que o produto Milk é o produto com maior participação em ambos os canais de vendas. O produto Grocery também possui uma significativa participação em ambos os canais podendo ser colocado em segundo lugar neste quesito.

Preparação dos dados:

Nesta etapa foi realizado o escalonamento e centralização dos dados através dos argumentos da função prcomp quando aplicada para a obtenção dos componentes principais do dataset.

O dataset antes da aplicação do PCA possuía variáveis contínuas em cada coluna e observações em cada linha. Após a aplicação do PCA, as variáveis passaram para as linhas e nas colunas ficaram os principais componentes de cada variável.

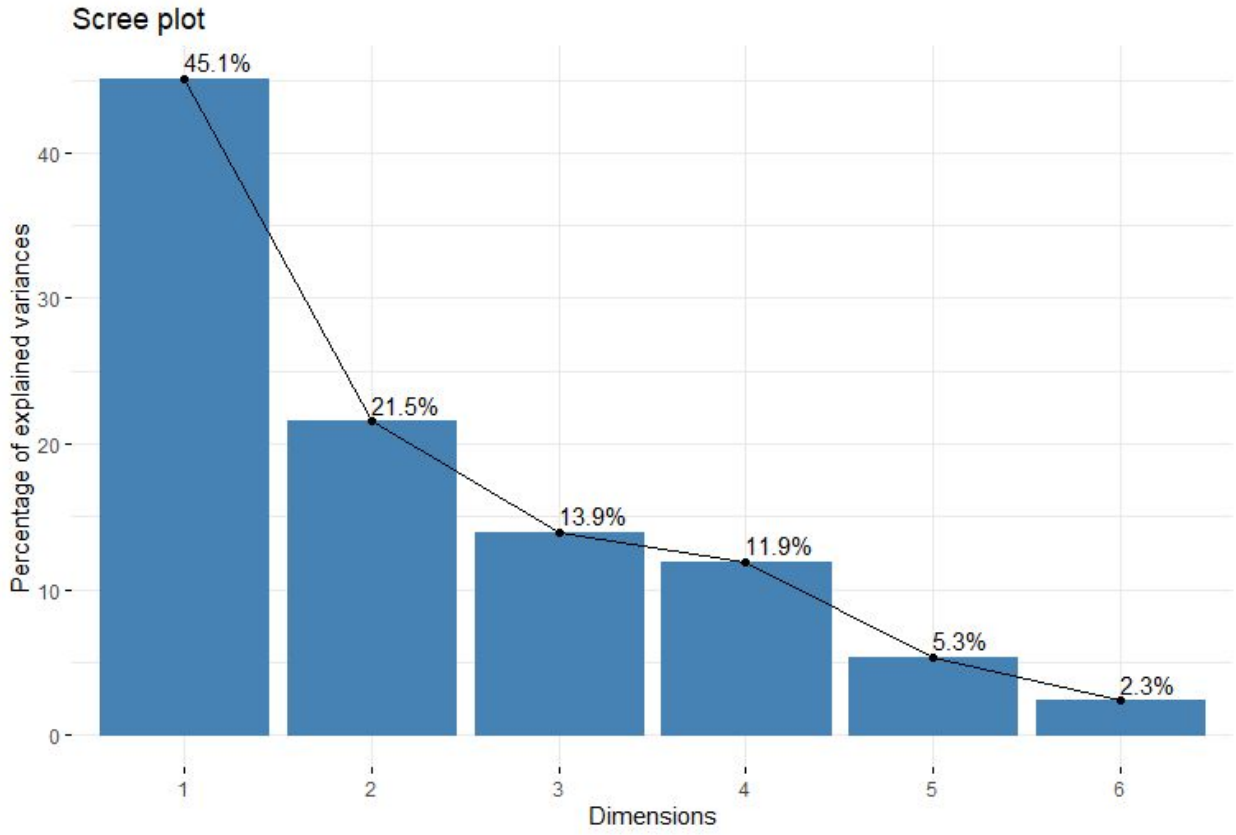
Gráfico do PCA PC1 e PC2 com outliers removidos



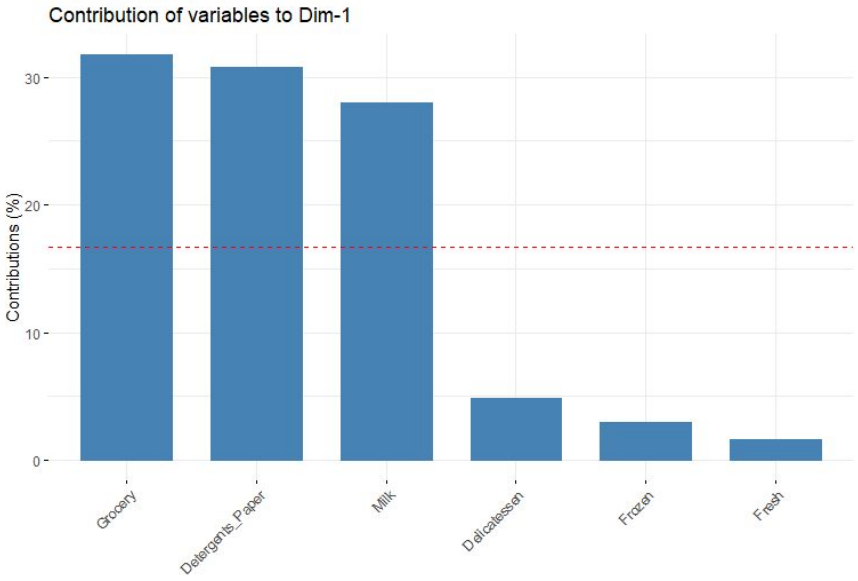
Neste gráfico podemos observar os dois principais componentes deste dataset em cada eixo. Os produtos preferidos de cada canal de venda estão mais próximos um do outro, como notado no gráfico de quantidade vendida por canal de venda. Detergents_paper, Grocery e Milk estão mais próximos entre si refletindo um padrão de agrupamento para o canal 2 e os produtos Delicatessen, Fresh e Frozen estão mais próximos entre si refletindo outro padrão de agrupamento para o canal 1. Nota-se que

Milk e Delicatessen são os produtos que estão mais perto do grupo que não pertencem, indicando serem produtos complementares destes canais de vendas. Os dois primeiros componentes principais explicam 66.62% da variância desta base de dados.

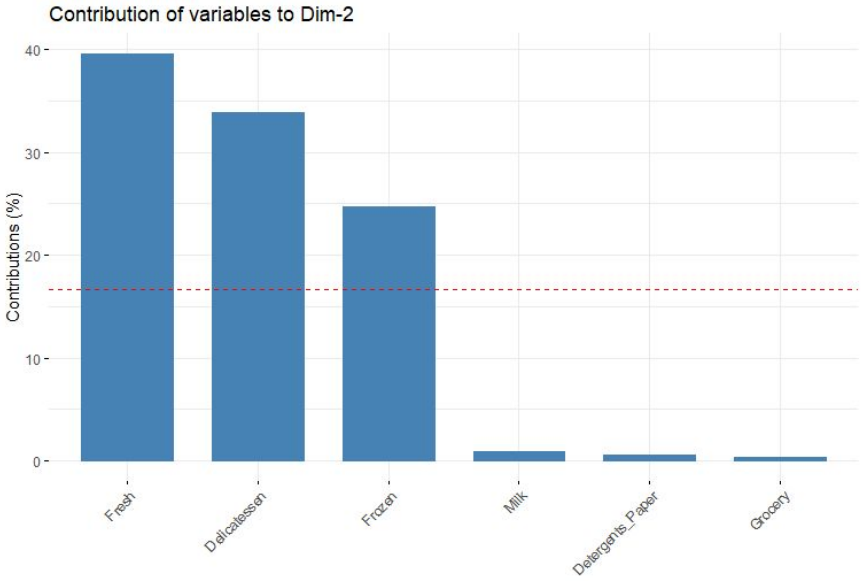
Gráfico do percentual de variância explicada por cada dimensão



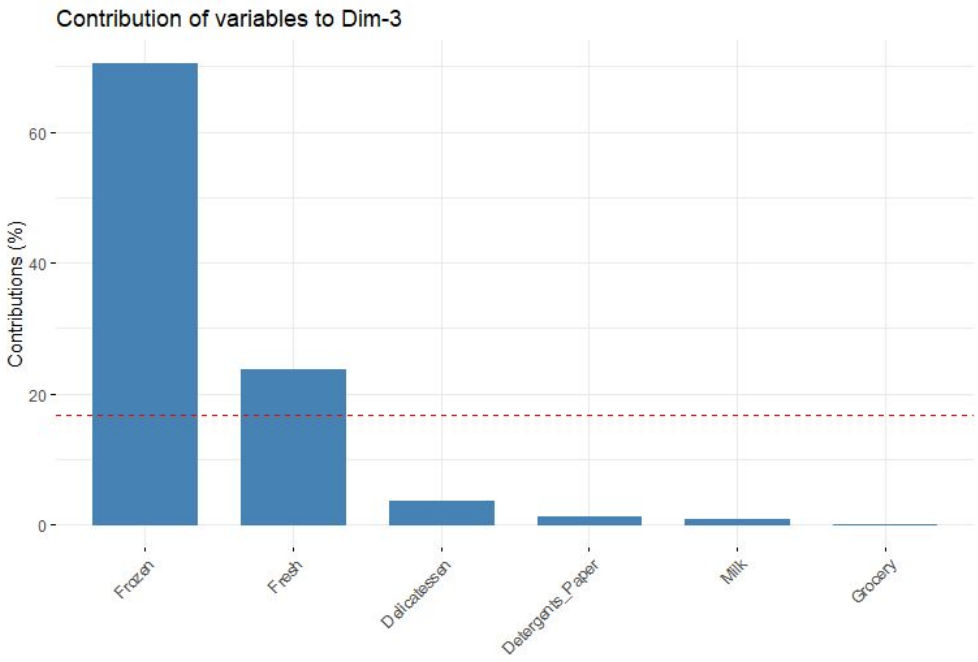
No gráfico acima pode-se ver que as quatro primeiras dimensões representam 92.42% da variância explicada, logo foi utilizado o dataset com PCA e com 4 componentes principais. Em seguida irei plotar a contribuição das variáveis para cada dimensão e para o acumlado das duas primeiras.



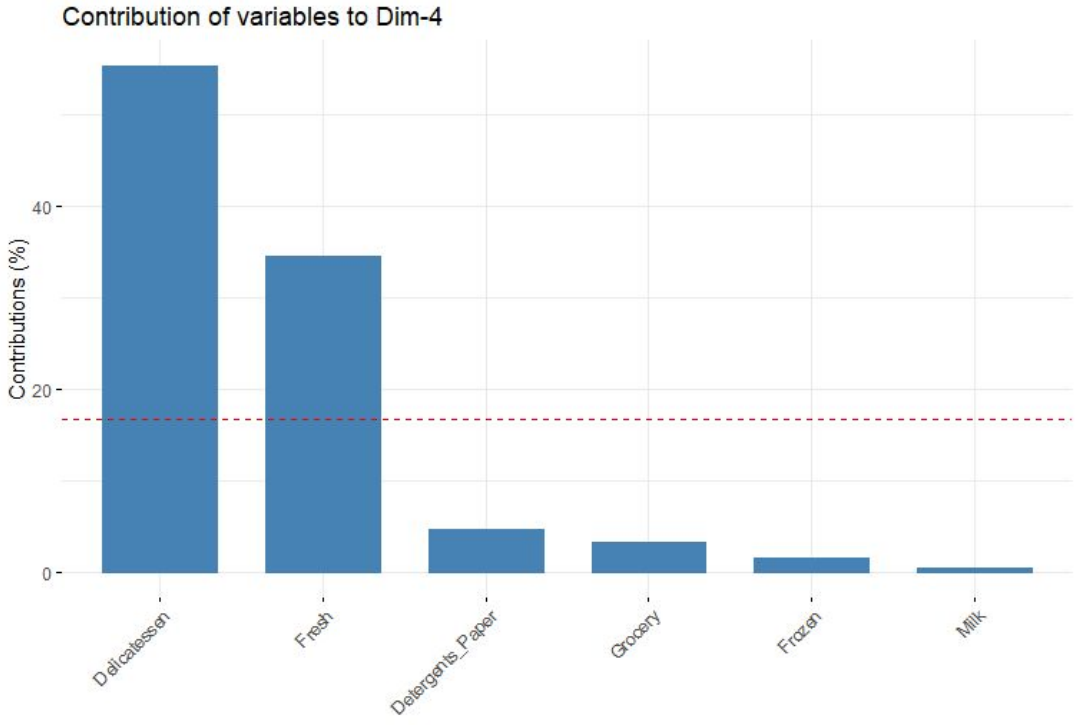
A primeira dimensão possui uma contribuição significativa dos produtos Grocery, Detergents_paper e Milk, enquanto que Delicatessen, Frozen e Fresh não.



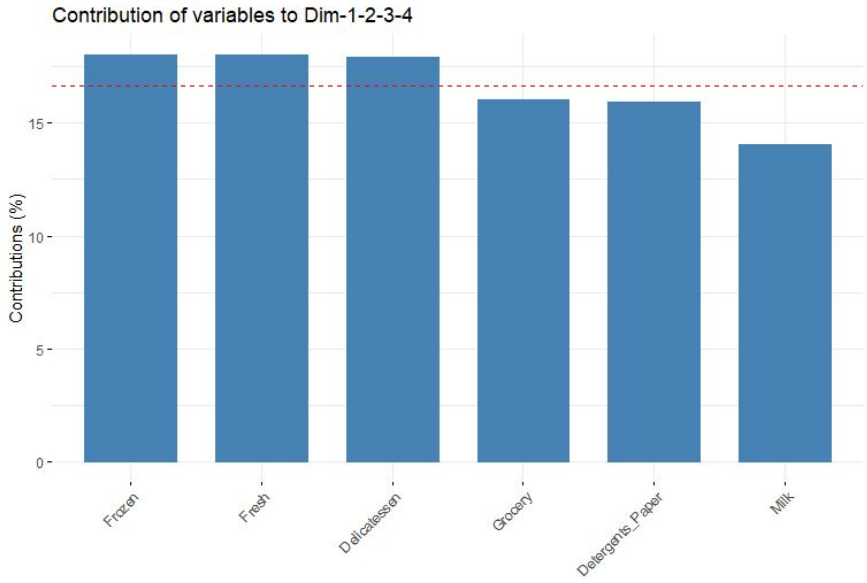
Na dimensão 2, os produtos Fresh, Delicatessen e Frozen são os mais significativos.



Na dimensão 3, os produtos significativos são Frozen e Fresh.



Na dimensão 4, os produtos significativos são Delicatessen e Fresh.



Ao analisar a contribuição de cada variável para o acumulado de 4 dimensões o padrão muda. Nesta situação, os produtos Frozen, Fresh e Delisatessen são os mais significativos, indicando serem os produtos que melhor satisfazem um cliente intermediário entre as quatro dimensões. Porém, os outros três produtos também estão próximo do nível de significância indicado.

Modelagem:

Os dois algoritmos escolhidos para a modelagem dos dados serão a clusterização por k-means e por hierarquia. Para aplicar o modelo k-means foi necessário avaliar o número de clusters ótimo para usar no modelo. O método utilizado foi o método da silhueta e do cotovelo. O método da silhueta é um método de avaliação interna que avalia a estrutura geométrica dos clusters, se são grupos compactos e separados entre si. O método do cotovelo consiste em observar o ponto onde a curva forma um cotovelo, o que indica não haver mais um ganho significativo de variância explicada a partir deste ponto.

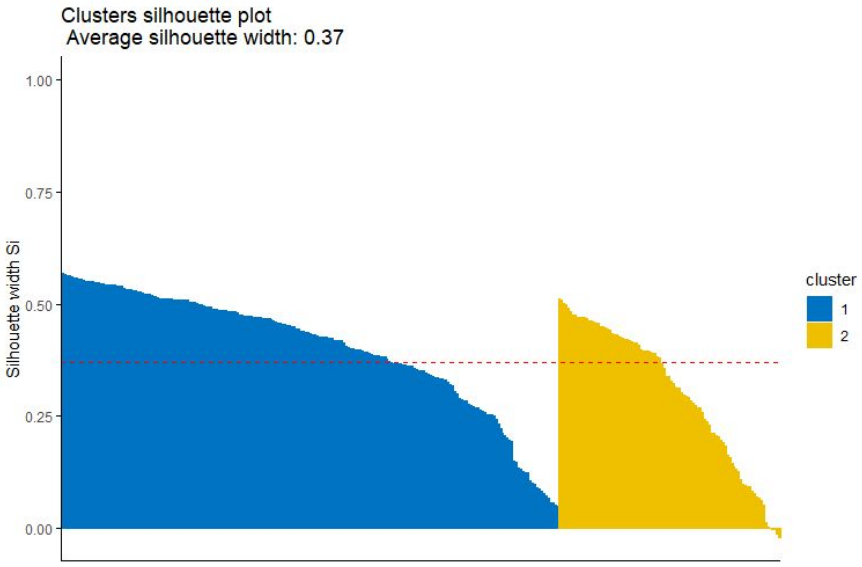
O modelo de clusterização hierárquica identifica agrupamentos e o provável número de grupos por meio de fusões ou divisões sucessivas. Os dois componentes principais mais próximos (medido pela similaridade) são agrupados em um nó do dendograma, posteriormente outro componente principal que estiver próximo é

agrupado a estes, e assim o processo é feito para todos os componentes até que todos estejam em um único cluster.

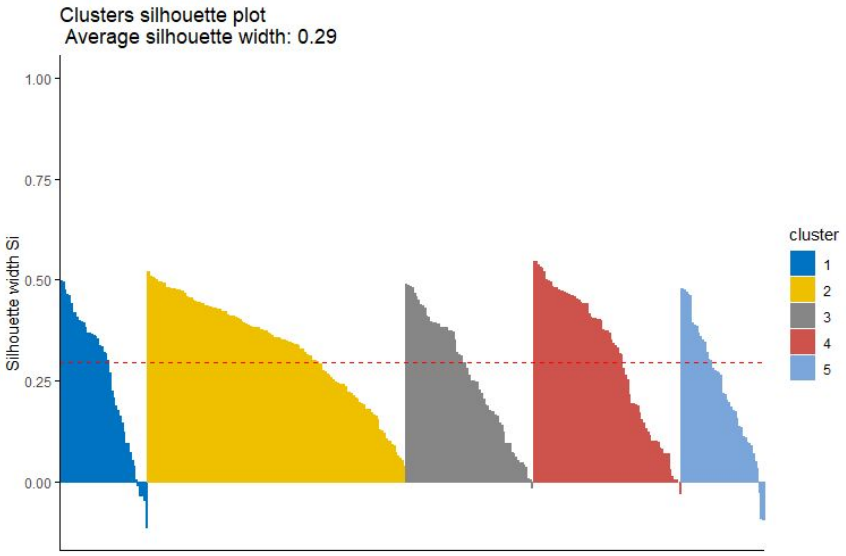
No modelo de clusterização hierárquica é possível alterar o tipo de métrica de distância utilizada pelo modelo, o que pode influenciar no resultado dependendo da distribuição dos dados.

Avaliação:

Escolha do número de clusters:

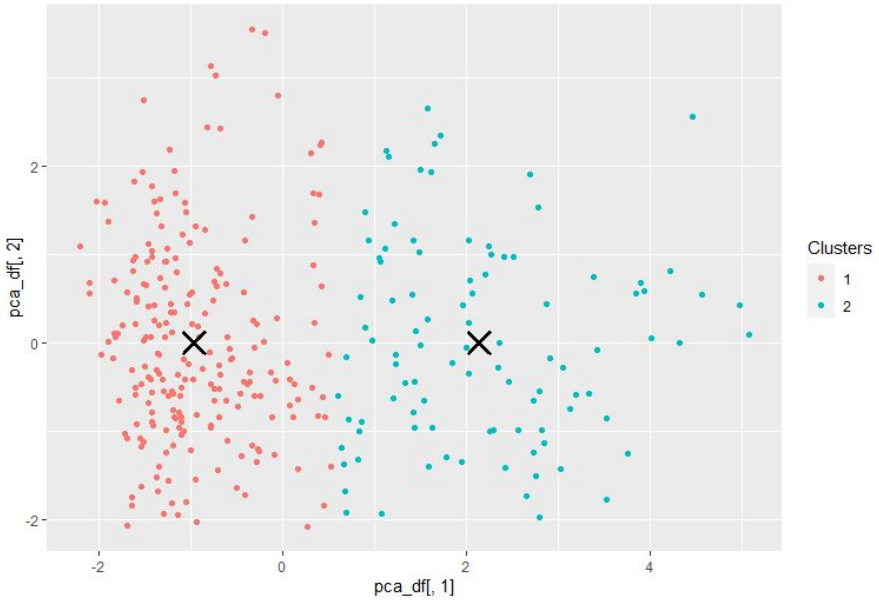


De acordo com o método da silhueta aplicado a um modelo k-means com 2 clusters, o número ótimo de clusters a ser utilizados na modelagem é 2.



O gráfico acima sugere que também é possível trabalhar com 5 clusters para essa base de dados, já que eles apresentam um valor significativo pela curva da silhueta. Vale ressaltar que o tamanho médio da silhueta para dois clusters é maior (0.37) do que para 5 clusters (0.29). Porém, ainda assim será feito também uma análise para 5 clusters, a fim de buscar obter mais informações por meio dos modelos de clusterização.

Gráfico dos 2 clusters obtidos pelo modelo k-means

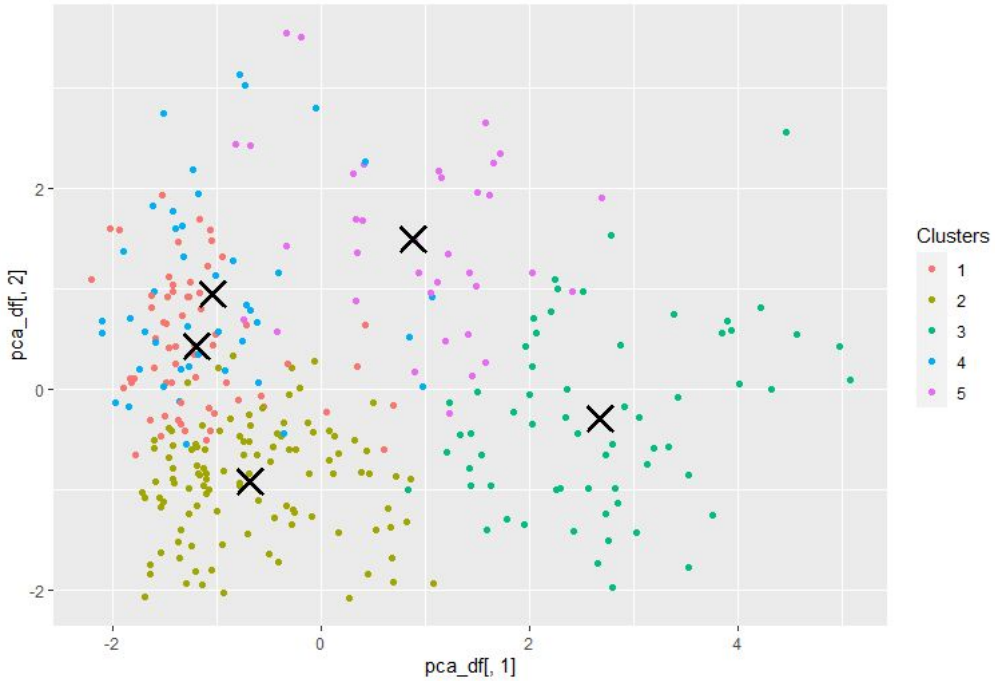


Este gráfico mostra os pontos que foram agrupados em seus respectivos clusters, onde os mais próximos ao centróide possuem um comportamento mais parecido com a média.

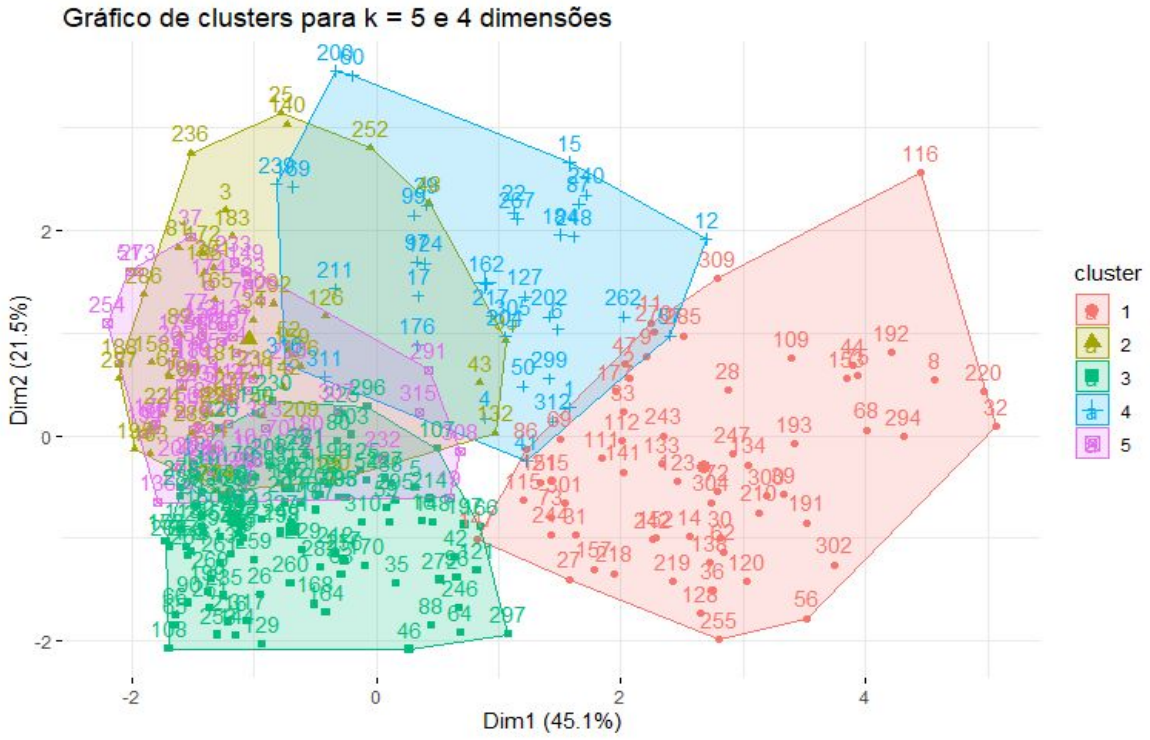
Ao comparar as observações de cada cluster com os labels originais da variável Channel foram obtidos 285 verdadeiros e 32 falsos, o que indica que apenas 11% das observações não foram caracterizadas iguais a variável Channel.

A seguir será plotado o gráfico para o 5 clusters obtidos com k-means.

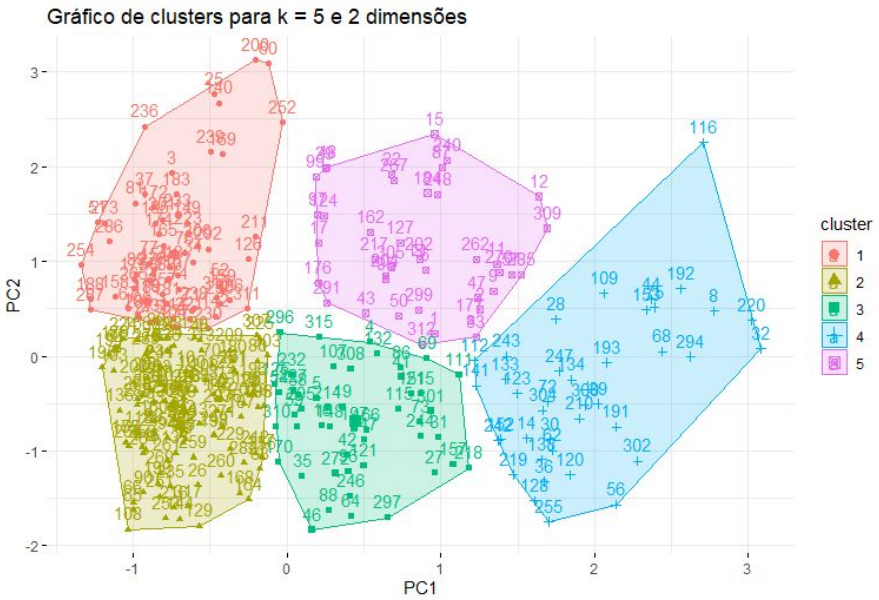
Gráfico dos 5 clusters obtidos pelo modelo k-means



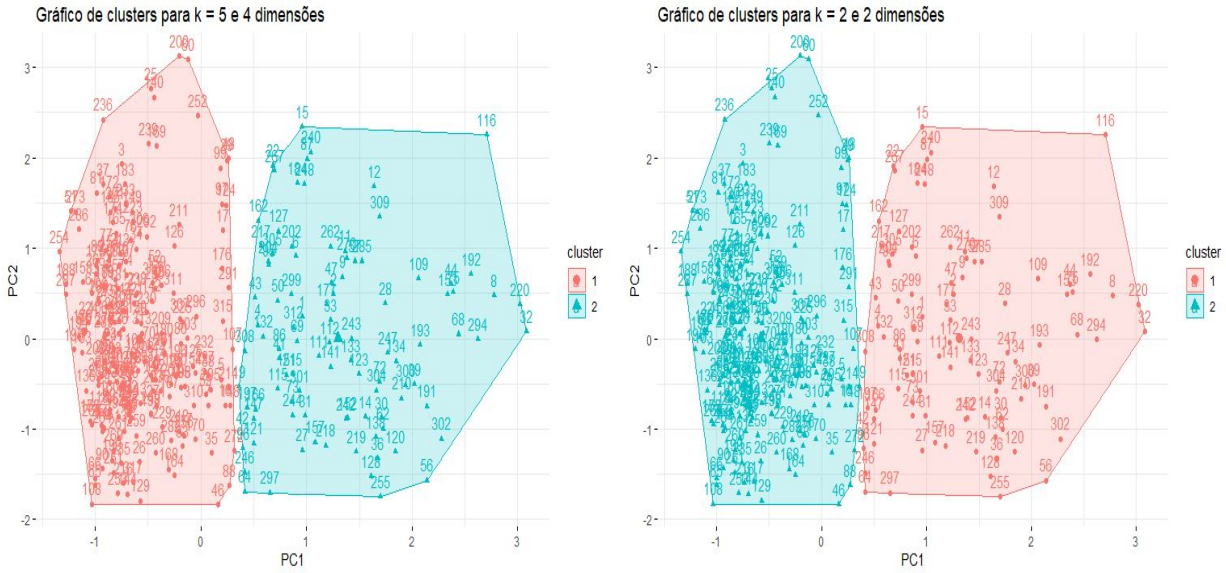
No gráfico acima podemos ver os centróides de cada um dos 5 clusters obtidos pelo modelo k-means.



Uma outra forma de visualizar os clusters é por este gráfico acima. Porém, como o PCA está utilizando 4 componentes principais a visualização dos clusters fica sobreposta. Para contornar isto foi realizado abaixo o plot com o PCA com 2 componentes principais para melhorar a visualização dos 5 clusters..



Ao utilizar 2 componentes principais a visualização dos 5 clusters fica mais fácil.



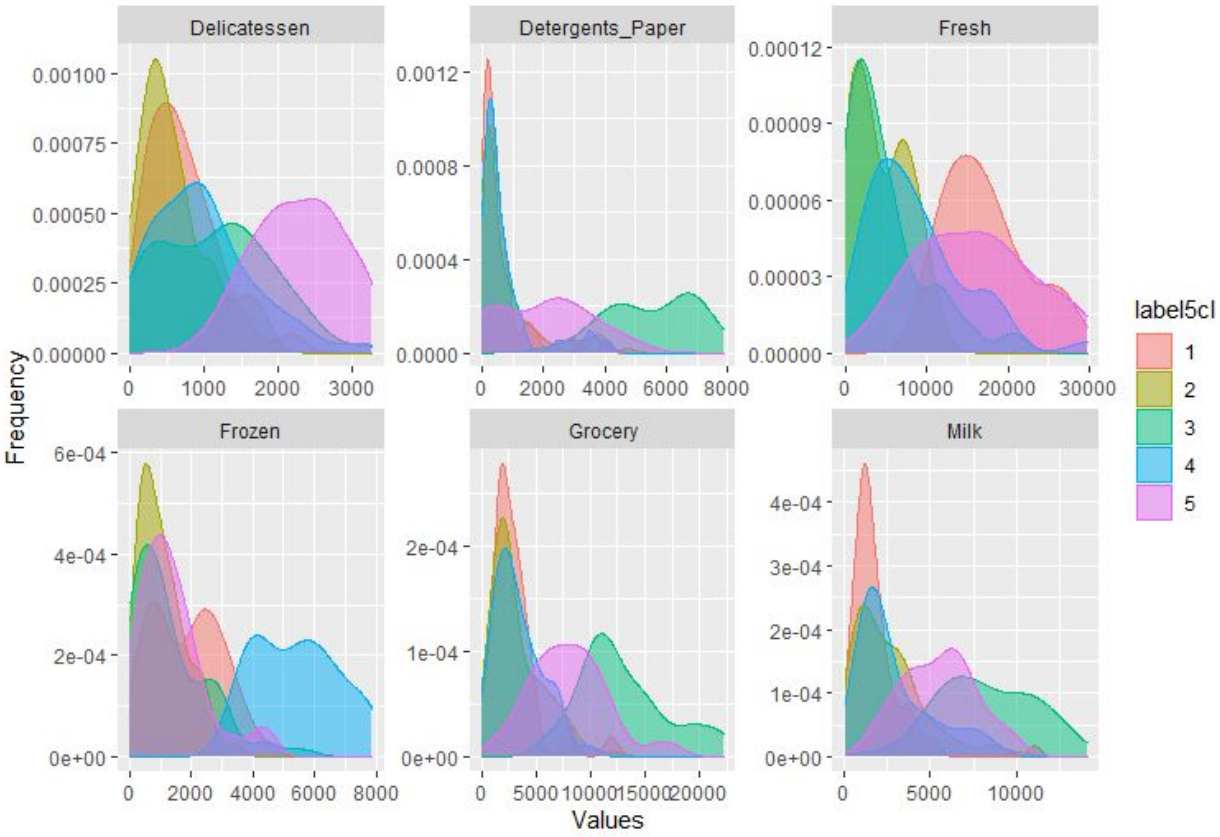
O mesmo procedimento foi aplicado para o modelo k-means com 2 clusters.

Compra média anual para k-means com 5 clusters e 4 dimensões no PCA

Cluster	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
1	17269.	1842.	2996.	1831.	571.	692.
2	4586.	2647.	3274.	1065.	850.	615.
3	5014.	8261.	13010.	1307.	5546.	1152.
4	8749.	2895.	3367.	5421.	778.	974.
5	15877.	5492.	7983.	1385.	2064.	2277.

De acordo com a tabela acima o cluster 1 possui maiores quantidades de Fresh, Grocery, Milk e Frozen empatados. Para o cluster 2, Fresh, Grocery e Milk. Para o cluster 3, Grocery, Detergents_paper e Milk. O cluster 4, Fresh, Frozen e Grocery. Por último, o cluster 5, Fresh, Milk e Grocery. Para traçar os perfis de compra de cada cluster vou utilizar abaixo um gráfico da densidade da distribuição de cada variável para cada cluster.

Gráfico de densidade da distribuição de cada variável para os clusters de 1 a 5



Analisando a tabela do padrão de compra médio e o gráfico de densidade acima, o perfil de cada um dos 5 clusters em ordem decrescente ficou assim:

(1) Fresh, Delicatessen, Milk, Frozen, Grocery e Detergents_paper

(2) Fresh, Grocery, Milk, Delicatessen e Frozen. Detergents_paper.

(3) Grocery, Milk, Detergents_paper, Fresh, Frozen e Delicatessen

(4) Fresh, Frozen, Delicatessen, Grocery, Milk e Detergents_paper

(5) Fresh, Grocery, Milk, Delicatessen, Detergents_paper e Frozen

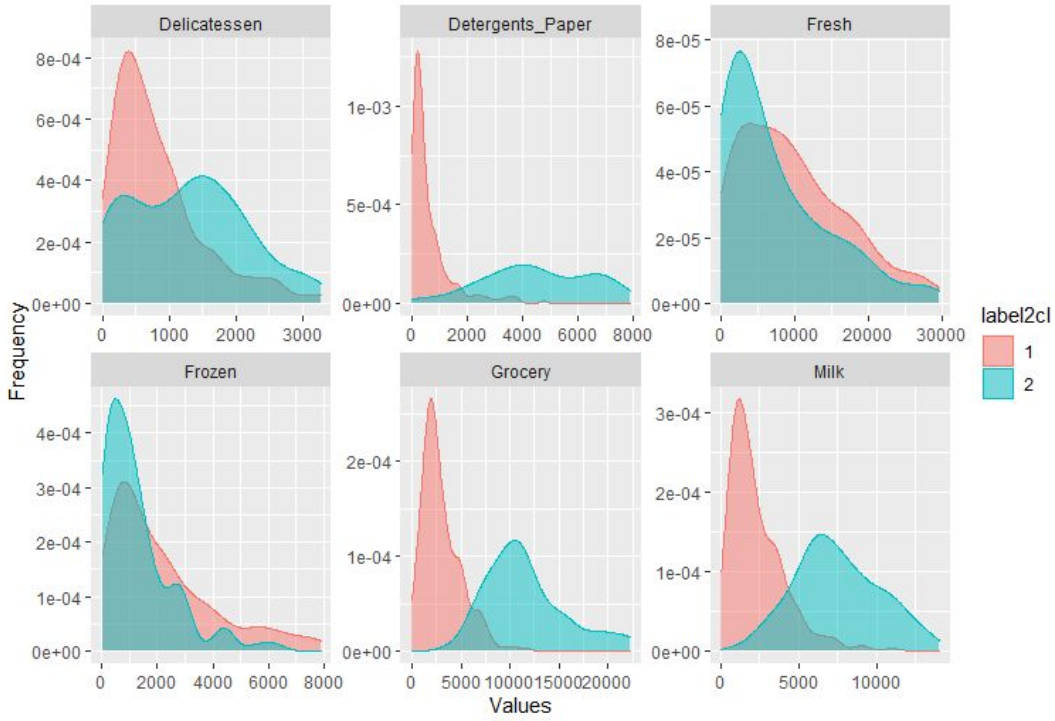
A análise do perfil de cada cliente foi realizada com base nos gráficos de densidade e na compra média de cada cluster do modelo k-means. A ordem de importância levou em conta a área da curva e o valor da compra média para cada categoria. Assim foi possível escolher as variáveis com uma distribuição mais próxima da normal ou menos assimétrica, e que possuem maiores níveis de compra média para cada cluster.

Compra média anual para k-means com 2 clusters e 4 dimensões no PCA

Cluster	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
1	9831.	2332.	3015.	2165.	632.	832.
2	7393.	7530.	11602.	1290.	4540.	1287.

O perfil de compra médio para o k-means com 2 clusters ficou com o Cluster 1 possuindo maiores quantidades de produtos das categorias Fresh, Milk, Grocery e Frozen e para o cluster 2 Fresh, Milk e Grocery e Detergents_paper. O perfil do cluster 1 se assemelha ao de HRC(hotel, restaurantes e cafés), porque possui maiores quantidades de produtos frescos e congelados para preparo de comidas. Enquanto que o cluster 2 se assemelha aos Retail (varejista), onde possui altas quantidade de Milk e Grocery comumente comprados em lojas do tipo mercearia, além de uma alta quantidade de produtos frescos que podem ser verduras frutas e hortaliças compradas para casa. Para auxiliar na análise do perfil de compra dos clusters foi plotado o gráfico de densidade abaixo.

Gráfico de densidade da distribuição de cada variável para os clusters de 1 e 2

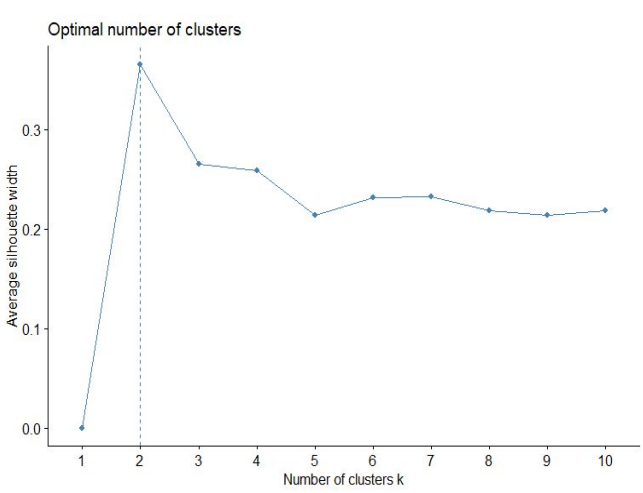


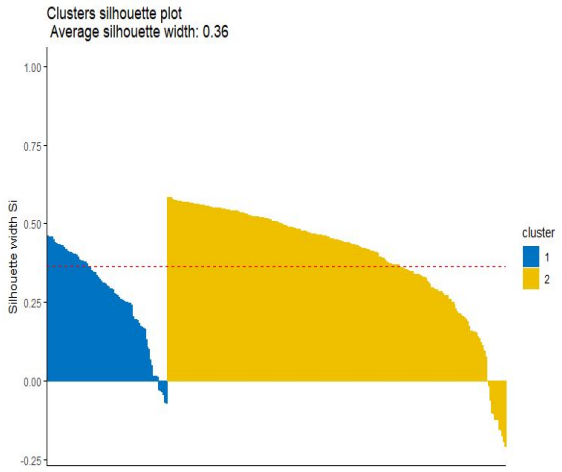
Como visto acima, o cluster 1 tem maior predominância das variáveis Fresh, Frozen, Delicatessen, pois apresentam maior área na curva de densidade. O cluster 2 tem maior predominância das variáveis Detergents_paper, Grocery e Milk, pois apresentam maior área na curva de densidade. Além disso, é válido ressaltar que as variáveis Milk, Fresh e Frozen possuem participação semelhante em ambos os clusters.

É de se esperar que compras feitas por hotéis e restaurantes (cluster 1) possuam mais quantidade de valores altos e menos de valores baixos, enquanto que as compras feitas por varejista possuem menor quantidade de valores altos e mais de valores baixos. Pois hoteis e restaurantes compram produtos da categoria Fresh e Frozen em maiores quantidades. A variável Delicatessen também apresenta este padrão. Já as outras três (Detergents_paper, Grocery e Milk) apresentam o inverso deste padrão, a quantidade de valores baixos é maior e de valores maiores é menor para o cluster 1. E a quantidade de valores baixos é menor e de valores maiores é maior para o cluster 2. Como é de se esperar de um cliente varejista, já que ele irá comprar maiores quantidade de Detergents_paper, Grocery e Milk do que um hotel e restaurante, para vender para seus clientes, enquanto que um hotel ou restaurante usaria apenas para os seus banheiros o Detergents_paper e Grocery e Milk são menos consumidos por clientes de hotéis e restaurantes normalmente.

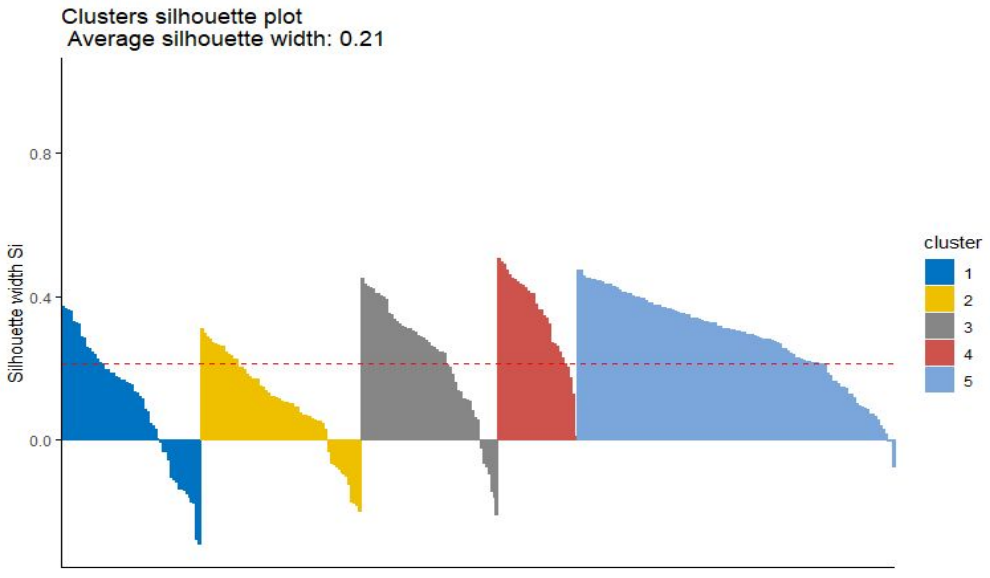
Clusterização Hierárquica

A seguir foi analisado a clusterização desta base de dados utilizando a clusterização hierárquica. Mas antes foi analisado o número ótimo de clusters para o método de clusterização hierárquica.

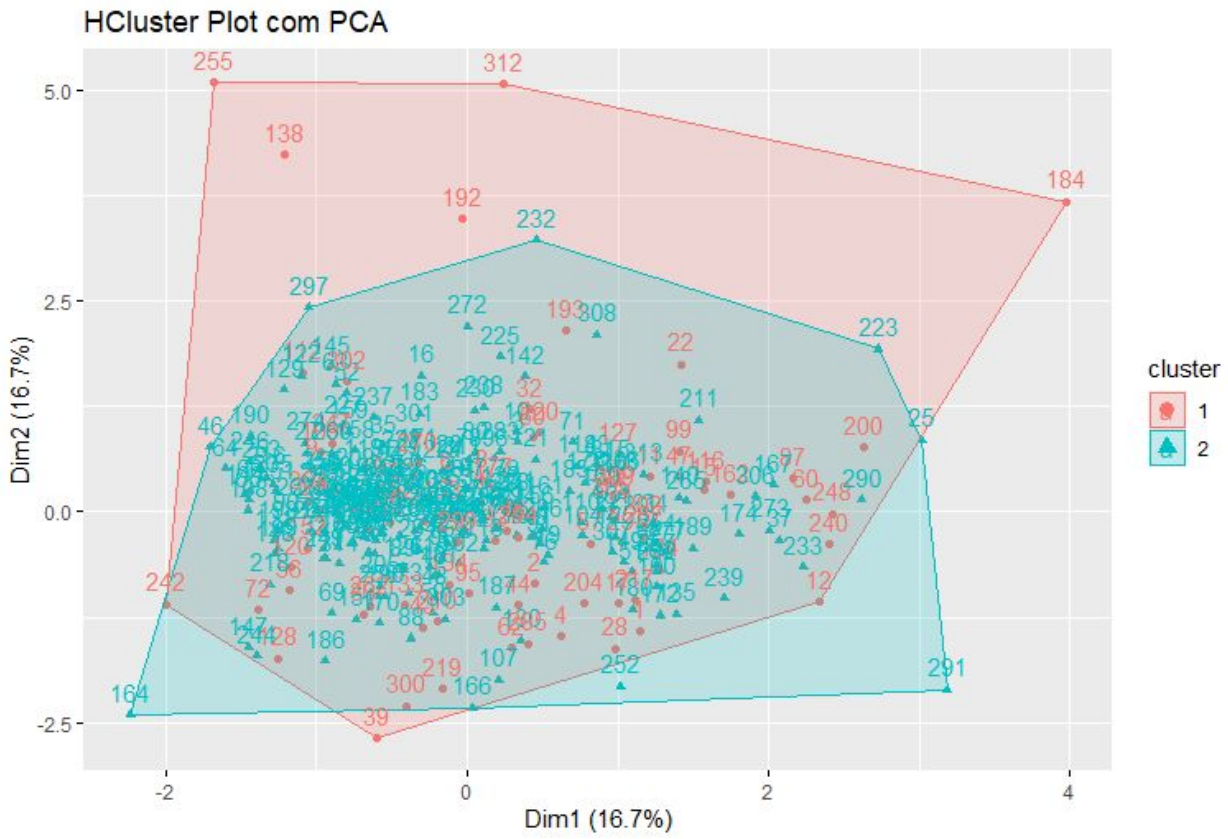




De acordo com o gráfico acima, pode-se concluir que o número ótimo de clusters é 2, o mesmo resultado obtido com o outro algoritmo (k-means). Porém, ao aplicar o modelo hierárquico com k=5 a curva da silhueta também mostrou-se ser significante para os cinco clusters.

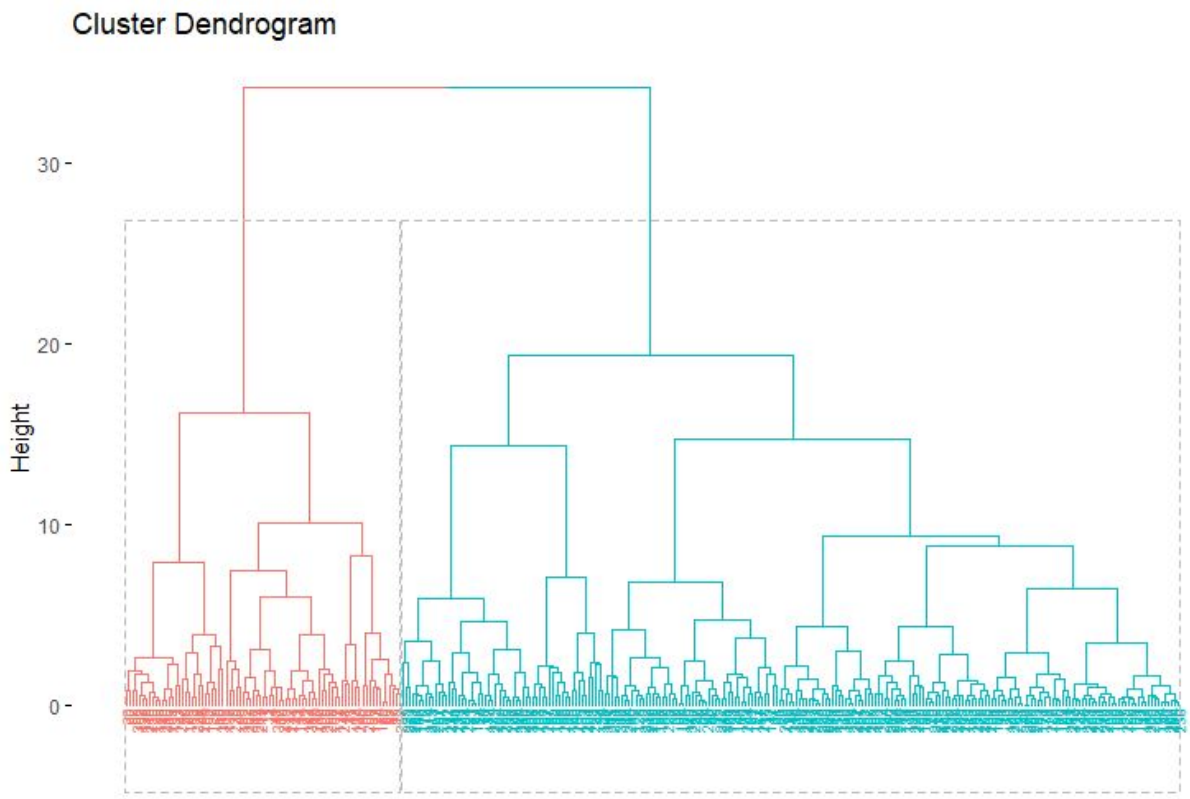


Em seguida foi plotado os clusters obtidos com o método hierárquico.

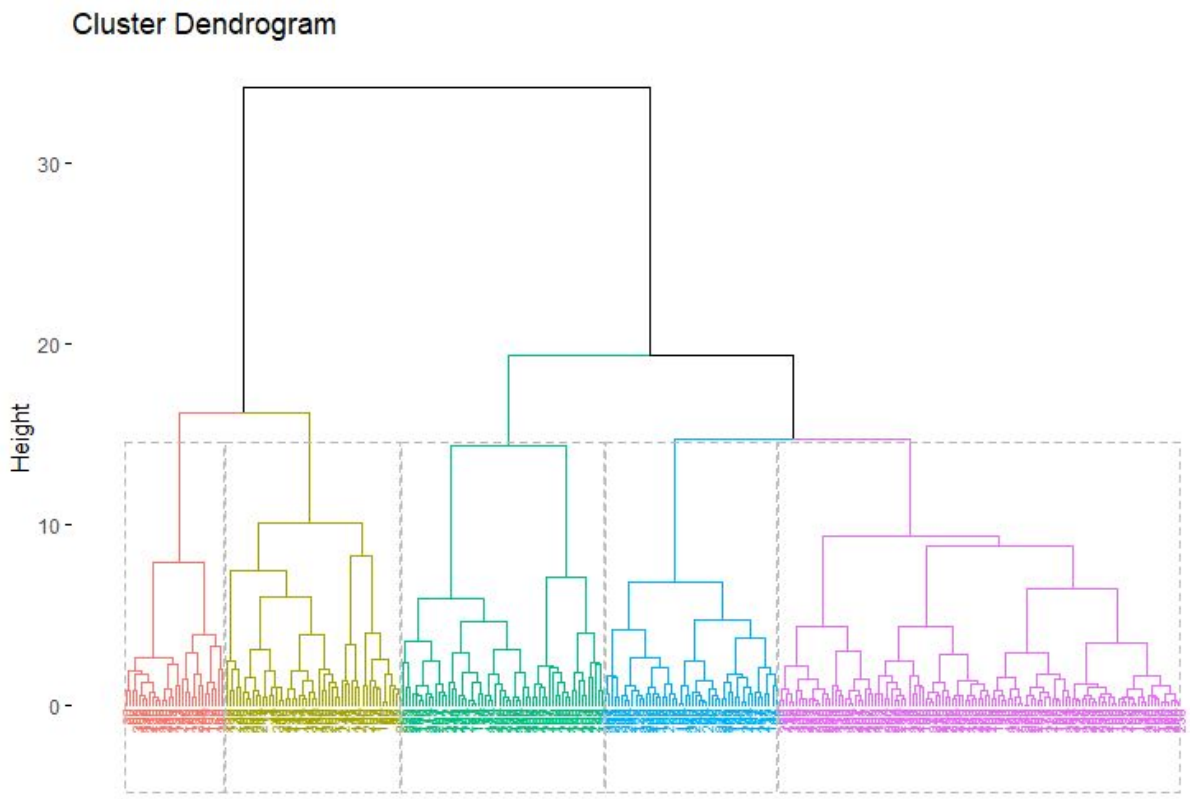


O gráfico acima revela o padrão de agrupamento obtido com a clusterização hierárquica dos clusters 1 e 2 quando aplicada na base de dados escalonada e centralizada e com PCA. O resultado obtido foi semelhante ao obtido pela clusterização com k-means, apesar de ocorrer uma sobreposição de parte dos dados sobre ambos os clusters. Abaixo foi plotado o mesmo algoritmo para a base de dados escalonado, centralizada e com PCA.

O próximo gráfico é um dendograma que representa como o algoritmo funcionou para chegar aos dois clusters obtidos.

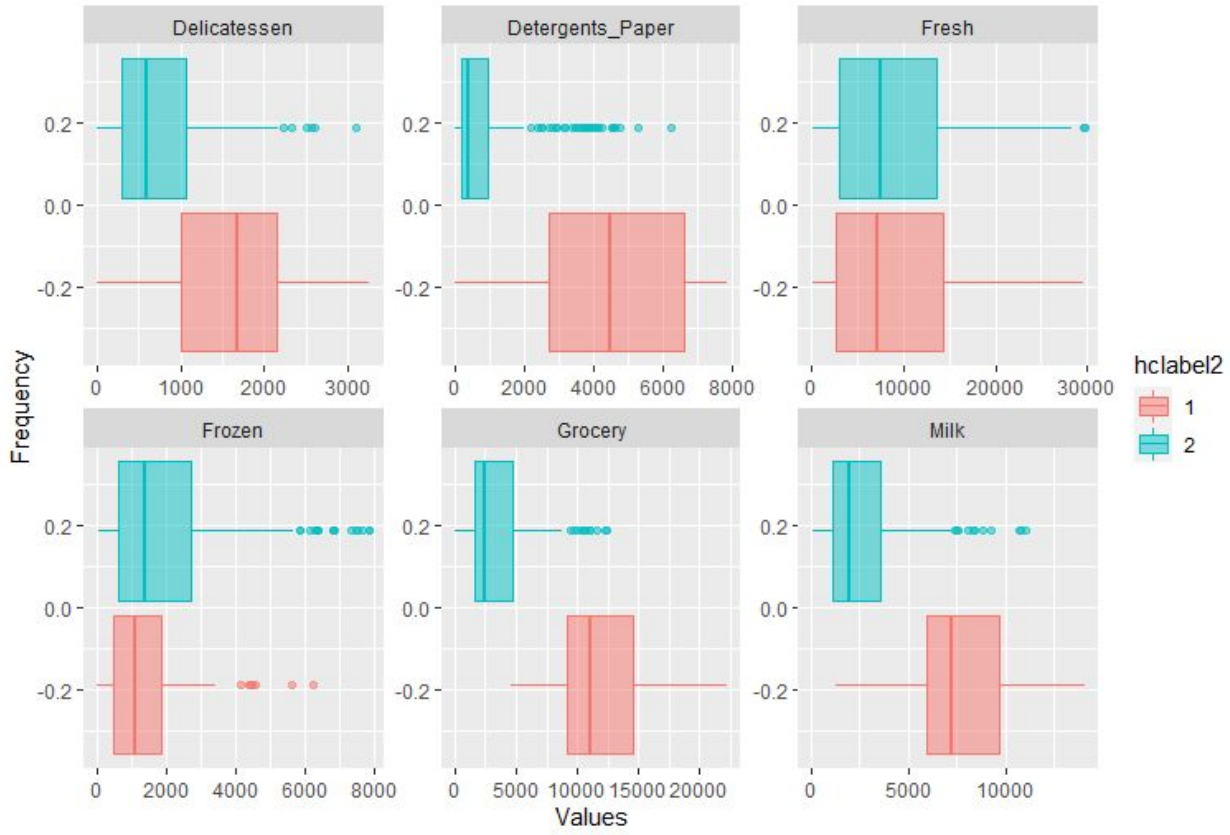


Em relação ao dendograma acima, pode-se notar que o agrupamentos foram feitos sucessivamente até que se chegasse a um único cluster. Ao realizar o recorte dos últimos dois grupos obtidos, se obtém os dois clusters encontrados pelo algoritmo hierárquico com k =2. A seguir foi plotado o dendograma para o modelo hierárquico com k =5.



Ao aplicar o modelo hierárquico com 5 clusters se obteve o dendograma acima como resultado dos sucessivos agrupamentos realizados até o corte final dos últimos 5 grupos.

Gráfico boxplot da importância de cada variável para cada clusters k = 2



Cluster	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
1	9178	7655	11877	1471	4490	1629.
2	9041	2621	3467	2044	900	740.

De acordo com a média e com a densidade da distribuição de cada variável:

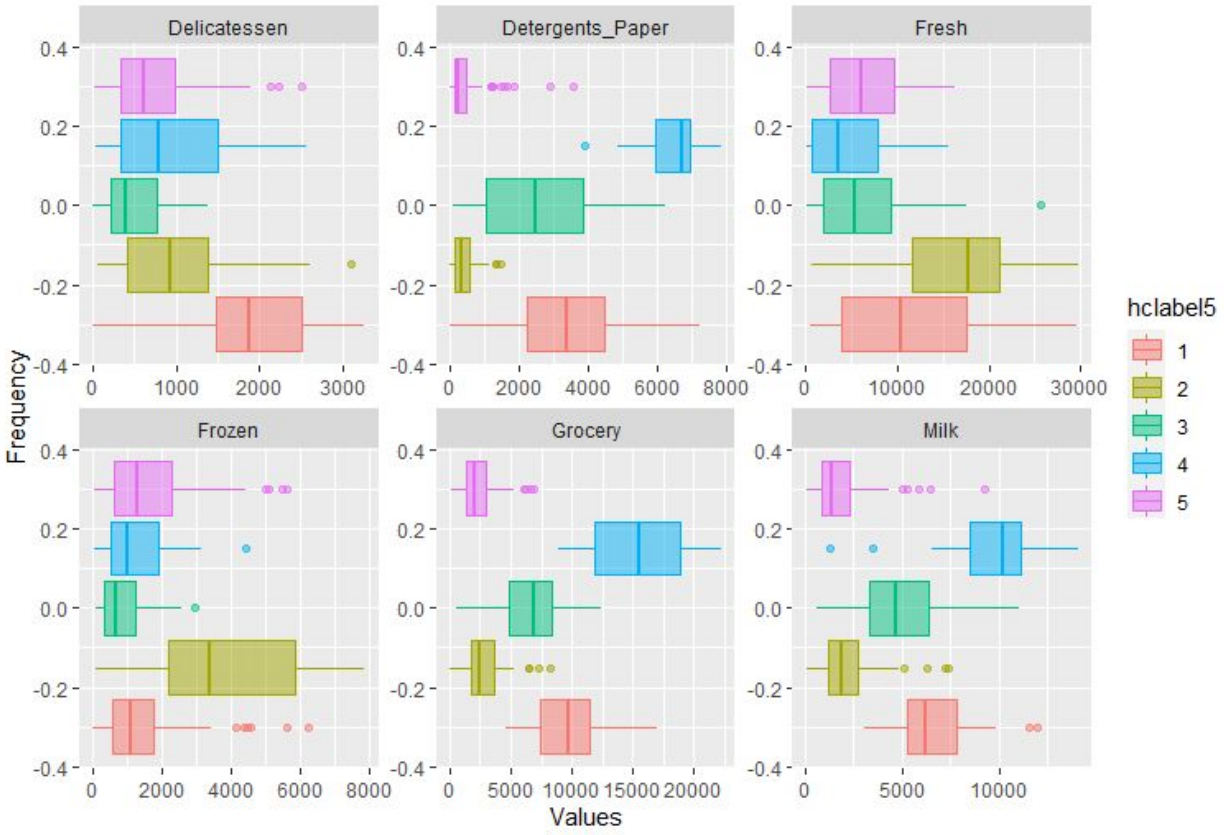
1 - Grocery, Detergents_paper e Milk são os mais relevantes.

2 - Fresh, Frozen e Delicatessen são os mais relevantes

Posteriormente foi realizada a avaliação interna de três métodos de clusterização diferentes (k-means, hierárquico e pam) com base na Conectividade (quanto menor, melhor conectados os clusters estão), Dunn (quanto maior, melhor o agrupamento em termos de separação e compactabilidade) e Silhouette (quanto maior,

mais similar cada objeto é para as observações do seu próprio cluster (similaridade) em comparação com as observações de outros clusters (dissimilaridade)).

Gráfico boxplot da importância de cada variável para cada clusters k = 5



Cluster	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
1	9626	6720	9888	889	3646	1522.
2	8078	2674	3046	5910	678	1126.
3	4774	2840	3743	1036	1085	563.
4	4921	9301	14629	1883	6096	1042.
5	15953	2219	3273	2186	578	1125.

De acordo com a média e com a densidade da distribuição de cada variável as categorias mais relevantes para cada cluster são (em ordem descrescente):

1 - Delicatessen, Fresh, Detergentes_paper, Grocery, Milk e Frozen

2 - Frozen, Fresh, Delicatessen, Milk, Grocery e Detergentes_paper

3 - Detergents_paper, Milk, Fresh, Grocery, Delicatessen e Frozen

4 - Grocery, Delicatessen, Milk, Fresh, Frozen e Detergentes_paper

5 - Fresh, Frozen, Delicatessen, Milk, Grocery e Detergents_paper

Em comparação com o modelo k-means, os perfis obtidos pelo modelo hierárquico se diferem um pouco. Porém, uma boa parte dos produtos mais significativos para cada cluster se repete em ambos os modelos, o que pode reforçar a importância daquela variável para um cluster específico. Isto foi visto na variável Fresh no cluster 1, 2, 3 e 4, Milk no cluster 2 e Detergents_paper no cluster 3. Nos outros casos as variáveis foram um pouco diferente. Como não se possui rótulos para cada um dos 5 cluster, é indicado o uso do modelo hierárquico neste caso, já que ele funciona de baixo para cima, realizando agrupamentos sucessivos que possibilitam encontrar os cluster de uma forma menos aleatória que o k-means.

Tabela de avaliação interna das técnicas de clusterização (k-means, hierárquico e pam)

hierarchical					
Connectivity	33.917	43.615	49.8262	54.2579	68.0940
Dunn	0.829	0.933	0.933	0.933	0.1031
Silhouette	0.3539	0.3220	0.2729	0.2471	0.2509
kmeans					
Connectivity	33.956	80.5143	79.6210	111.8778	121.512
Dunn	0.0783	0.660	0.657	0.600	0.648
Silhouette	0.3752	0.3062	0.2916	0.2854	0.2741
pam					
Connectivity	37.444	82.4258	129.5742	133.1302	155.1016
Dunn	0.610	0.576	0.452	0.452	0.523
Silhouette	0.3511	0.3279	0.2746	0.2886	0.2208

	Score	**Method**	**Clusters**
Connectivity	33.3917	hierarchical	2
Dunn	0.1031	hierarchical	6
Silhouette	0.3752	kmeans	2

As tabelas acima mostram que o melhor algoritmo em relação aos critérios de conectividade, Silhueta é o de clusterização hierárquica com 2 clusters. O critério Dunn indica que o melhor seria o método hierárquico com 6 clusters.

Conclusão

Então podemos concluir a partir da análise de clusters com o método k-means e hierárquico que esta base de dados pode ser analisada a partir de dois ou cinco grupos com padrões de compras anuais diferentes. A análise com dois clusters indicou que o primeiro cluster possui quantidades significativas de produtos da categoria Fresh, Frozen e Delicatessen, que são produtos consumidos normalmente por Hotéis, restaurantes e cafés para preparar comidas e vender produtos de delicatessen para os hóspedes/clientes. Já o segundo cluster possui quantidades mais significativas de produtos das categorias Milk, Detergents_paper e Grocery, o que reflete o padrão de compra de um cliente médio de um comércio varejista, pois são produtos comuns de serem comprados em mercearias ou comércios locais. Além disso, a base contava com a variável Channel que indicava se o cliente era HRC ou varejista.

Como o modelo de clusterização hierárquica é preferível quando não se tem os rótulos dos clusters, este foi o escolhido para esta análise. A análise com cinco clusters com o método hierárquico obteve os seguintes perfis como resultado:

1 - Delicatessen, Fresh, Detergentes_paper, Grocery, Milk e Frozen

2 - Frozen, Fresh, Delicatessen, Milk, Grocery e Detergentes_paper

3 - Detergents_paper, Milk, Fresh, Grocery, Delicatessen e Frozen

4 - Grocery, Delicatessen, Milk, Fresh, Frozen e Detergentes_paper

5 - Fresh, Frozen, Delicatessen, Milk, Grocery e Detergents_paper

O cluster 1 pode ser formado por pequenos restaurantes, lanchonetes ou cafés, já que vende mais produtos básicos e de delicatessen. O cluster 2 já aparenta ser formado por restaurantes e hotéis (HRC) por ter mais produtos como Frozen e Fresh. O cluster 3 aparente ser de clientes como mercearias, por vender mais Detergents_paper, Milk e Fresh, produtos geralmente vendidos por mercearias. O cluster 4 também tem um padrão de consumo semelhante a mercearias, porém uma mercearia que vende mais produtos de delicatessen e de grocery do que produtos frescos. O cluster 5

aparenta ser um hotel ou restaurante (HRC) grande, já que a maior parte dos produtos vendidos são Fresh e Frozen, indicando que produzem mais comida e vendem menos produtos de delicatessen que o cluster 1, que também aparenta ser HRC.

A análise de cluster com k =2 pelo modelo k-means obteve um maior acerto em relação ao modelo hierárquico. O k-means acertou 286 e errou 32 enquanto que o hierárquico com o método ward.D2 acertou apenas 40. O modelo hierárquico chegou a acertar 233 com o método centroid e 232 com o método single, porém a visualização dos dados clusterizados ficou comprometida, por isso foi utilizado o método Ward.D2. Como o acerto foi bem baixo para a clusterização hierárquica, neste caso o indicado é que seja utilizado o algoritmo k-means quando k = 2 para que seja mais fácil a visualização e para que acerte mais também.

Portanto, a existência de dois e cinco clusters nessa base de dados faz sentido de acordo com o que foi visto neste trabalho. Além disso, é possível saber o padrão de compra de um cliente médio para cada cluster e para cada modelo. O que possibilita obter melhores informações para guiar ações como compra de estoques, segmentação de campanhas publicitárias e promoções dos itens mais comprados por cada cluster.