

# ANÁLISE DE INFORMATIVO NUTRICIONAL DA REDE MCDONALD'S ATRAVÉS DE CLUSTERIZAÇÃO COM FOCO EM APRESENTAR UMA PROPAGANDA ABRANGENTE

Cientista de dados: **Arthur Lucena Silva**

Email: [arthur.als95@outlook.com](mailto:arthur.als95@outlook.com)

Data: 20/07/2021

## 1. INTRODUÇÃO

Uma das principais redes de *fastfood* do mundo, o McDonald's está presente em mais de 20 países e atende cerca de 46 milhões de clientes por dia. Seu principal produto é o Hambúrguer, que será o foco de nossa pesquisa (FIGUINHA, 2005). Atualmente o “Méqui”, como é popularmente conhecido, é considerado a maior franquia de comida rápida do mundo, contudo, precisa brigar pelo *market share* com outras empresas bem avaliadas no mercado. De acordo com o site SUAFRANQUIA.COM (2019) e também o site T4.AI (2021), os principais concorrentes seriam: KFC, Subway, e Domino's Pizza considerando o estudo mais recente e também selecionando especificamente o ramo gastronômico, o restaurante Burger King a título de curiosidade entra bastante enfraquecido na briga pela fatia de mercado com apenas 1,2% ao lado do Pizza Hut, conforme a Figura 1. Com isso em mente, é fundamental para essas companhias realizarem um ótimo serviço de marketing e adotarem estratégias eficazes na tomada de decisão de suas propagandas, que por sua vez não é algo barato, de acordo com a VEJA (2021) 30s de break no programa *Fantástico* custam cerca de 1.05 milhão de reais.

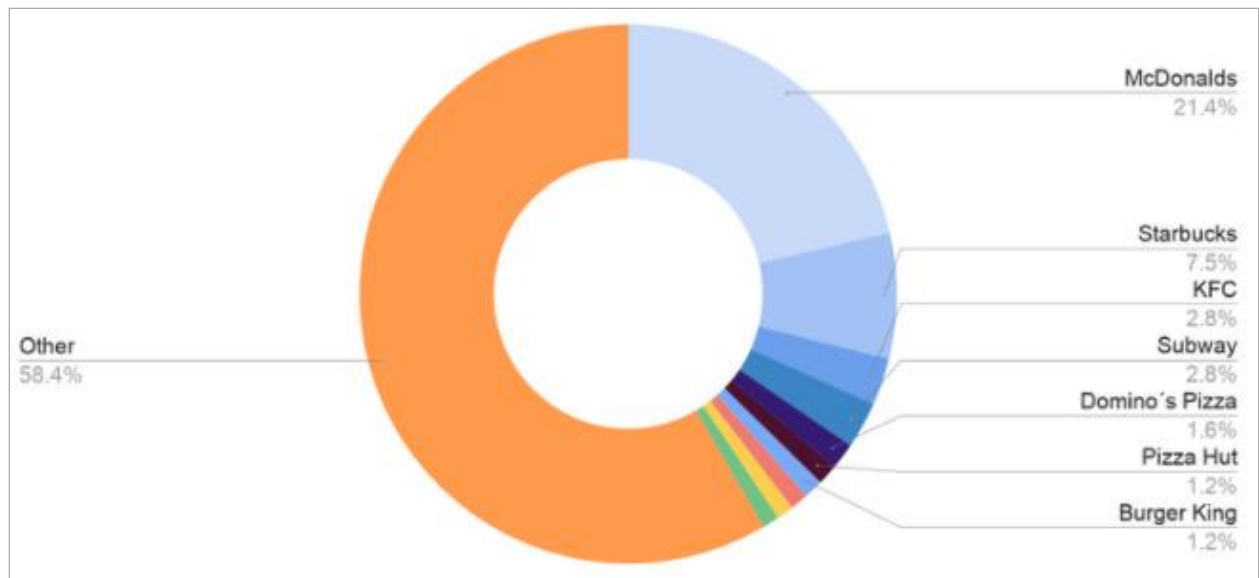


Figura 1 - Market share das principais franquias do ramo alimentício em 2021

## 2. OBJETIVO

Com base nos dados nutricionais dos lanches disponíveis no cardápio da rede McDonald's, foi solicitado um estudo para determinar – “Quais são os lanches da rede do McDonald's que devem ser transmitidos durante o intervalo de um programa televisivo de alta audiência, que demonstrem com eficácia a variedade de nosso cardápio”.

### 2.1. Entendendo a questão

A solução do caso exige uma resposta que contenha quais são os lanches que precisam ser passados no comercial, sabemos que estes lanches precisam ser distintos um do outro para que o comercial possa ter uma abrangência maior e consiga alcançar mais público do que o comum. Com base nessas considerações a questão engloba análise **descritiva**, e **exploratória**.

### 3. ANÁLISE EXPLORATÓRIA E DETALHAMENTO DOS DADOS

O estudo apresentado neste trabalho é baseado em um *dataset* do tipo “.csv” que contém 25 tipos de lanches comercializados pelo McDonald’s, e que contém no total, incluindo o tipo de lanche, 12 variáveis nutricionais do produto comercializado, onde a variável “Lanche” é do tipo *string*, e as demais (que são as nutricionais) são numéricas. Não foram identificados *missing values* na base de dados utilizada.

#### 3.1. Adequabilidade do *dataset* e revisão do objetivo

A base de dados possui variáveis numéricas que possibilitam a aplicação do método de “clusterização”, portanto não apresentou barreira para a elaboração do modelo. Os dados apresentaram uma confiabilidade de 90,51%, onde foi feito uma comparação com a descrição nutricional do lanche *Big Tasty* que está no *dataset*, com o que está no site do McDonald’s, conforme apresentado na Tabela 1. Esta confiabilidade foi proveniente das médias das razões das variáveis nutricionais expostas, é um valor estimado e possui a incerteza também da base utilizada, contudo possui um índice médio aceitável para o teste feito, representando veracidade nos dados apresentados e descartando a necessidade de nova base de dados para análise.

Fonte	Lanche	Valor Energetico	Carboidratos	Proteinas	Gorduras Totais	Gorduras Saturadas	Gorduras Trans	Colesterol	Fibra Alimentar	Sodio	Calcio	Ferro
Base utilizada	Big Tasty	843	45	41	55	24	1,7	104	5,1	1511	381	8,2
Site McDonalds	Big Tasty	944	45	43	66	26	1,6	121	4	1490	335	7,4
Confiabilidade		90,51%										

Tabela 1 - Comparativo nutricional do Big Tasty

## 4. METODOLOGIA E MODELAGEM APLICADA

Para o estudo feito foi utilizado uma técnica do tipo não supervisionada de *machine learning* denominada *cluster*, que visa o agrupamento de observações de tal forma que dentro de cada grupo as observações sejam semelhantes, e entre grupos sejam distintas, em outras palavras, dentro de cada grupo a variabilidade deve ser mínima enquanto que entre grupos deve ser máxima. Os esquemas de agrupamento podem ser classificados, basicamente, em dois tipos, conhecidos como hierárquicos e não hierárquicos. Enquanto que o tipo hierárquico privilegia uma estrutura como sua própria definição aponta, hierárquica, para a formação dos grupos, a variante não hierárquica de análise utiliza algoritmos para maximizar a homogeneidade entre os membros de cada grupo.

### 4.1. Padronização das variáveis

Ambos os métodos classificados que serão usados no estudo exigem que os dados correspondentes aos fatores sejam padronizados numericamente para que a distância euclidiana entre fatores possa ser medida e comparada entre grupos.

### 4.2. Método hierárquico

O agrupamento entre as amostras é realizado a partir das amostras mais próximas entre si (menor distância) para todas as técnicas. E as diferentes técnicas de agrupamento hierárquicos são aplicadas apenas para cálculo de um grupo para outro. Existem algumas variantes da técnica hierárquica, para o presente estudo foi adotada a variante *Average Linkage* que tem por essência: (1) Calcular a distância de todos os elementos observados contra todos, (2) Agrupar os elementos mais próximos (menor distâncias), (3) Define a distância do primeiro grupo contra todos os demais grupos baseados na **média** da distância entre cada integrantes do grupo com os demais, (4) Após chegar no último grupo, o dendrograma é criado baseado na distância calculada.

#### 4.2.1. Gráfico de Elbow

Esta técnica tem como objetivo auxiliar na escolha do número de *clusters*. Assim como seu princípio, a ideia é minimizar a variabilidade dentro de cada *cluster*, ou seja:

$$\text{Minimize}(\sum_{i=1}^n W(C_k))$$

Onde  $C_k$  é o  $k$  (*cluster*), e  $W(C_k)$  é a variação dentro do *cluster*. Então a soma dos quadrados total dentro de cada *cluster* (*wss*) mede a homogeneidade do *cluster* e queremos que seja a menor possível. Dessa forma foi gerado o gráfico da Figura 2, que nos permite identificar a criação de 4 grupos de *cluster*. Não é o nosso caso, porém se o tempo de propaganda, ou o capital de investimento na propaganda, fosse um fator limitante poderíamos diminuir o número de *clusters* para se adequar aos parâmetros exigidos da situação.

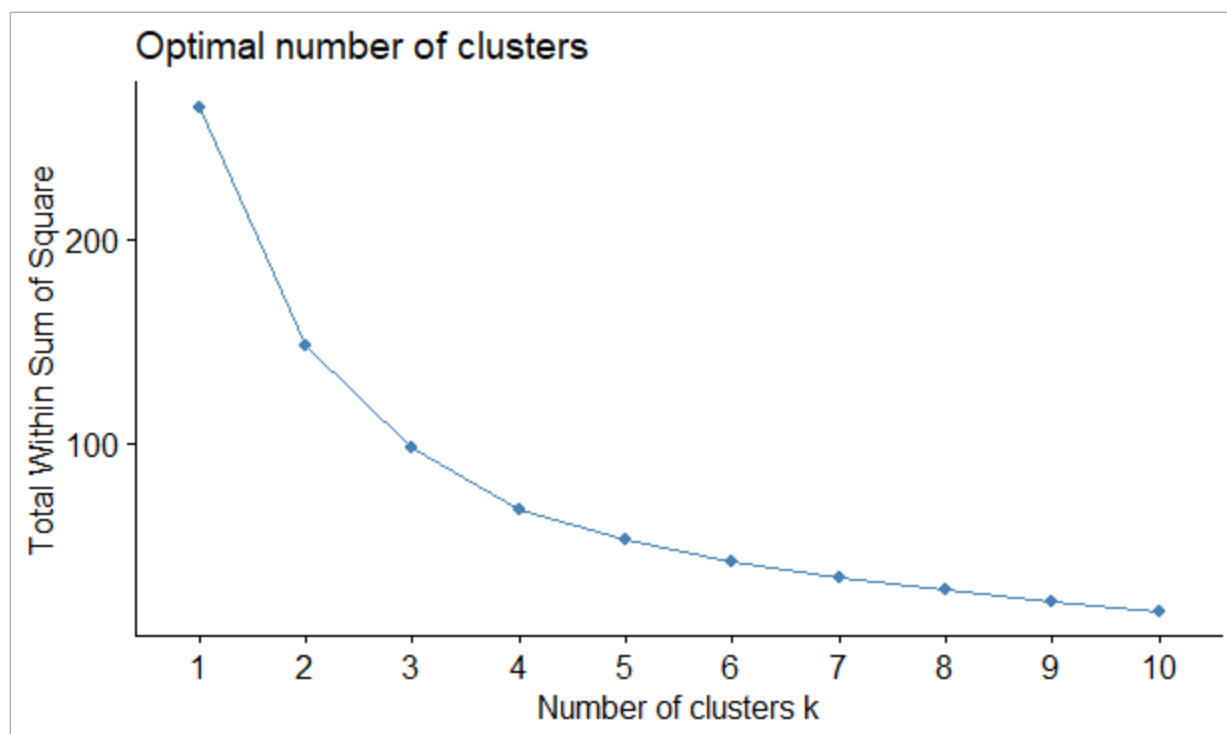


Figura 2 - Gráfico de Elbow

#### 4.2.2. Dendograma

A Figura 2 apresenta o dendrograma da análise pelo método *average*, em seguida a Figura 3 demonstra a criação de 4 grupos distintos entre si, porém semelhantes entre integrantes.

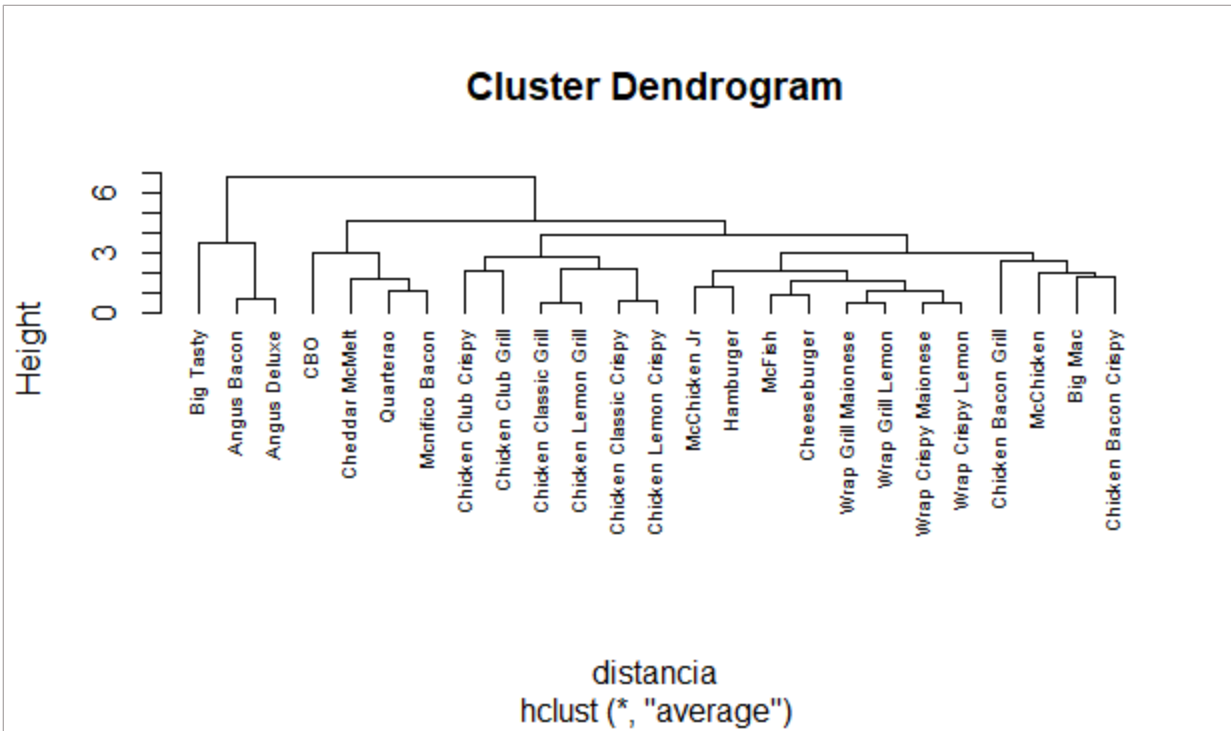


Figura 3 - Dendograma cluster variante average

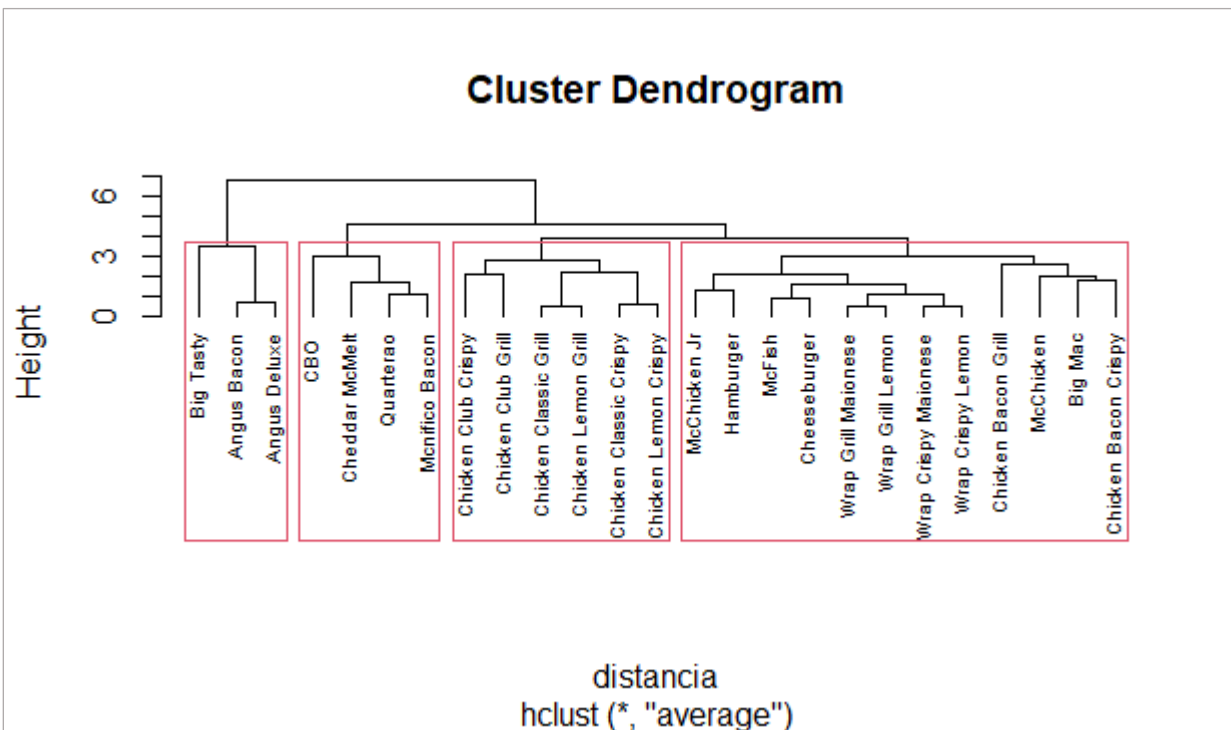


Figura 4 – Divisão de grupos clusterizados

### **4.3. Método não hierárquico**

O método não hierárquico é recomendado para ser aplicado quando se tem um grande volume de dados, o que acaba dificultando a definição de *clusters*. Não é o nosso caso, porém é uma prática altamente recomendada usar mais de um modelo nas análises quando possível com o intuito de alinhar as expectativas do objetivo e checar o padrão de resposta de cada modelo. Para o estudo feito foi adotada a variante *k-means* que opera essencialmente da seguinte maneira: (1) Seleciona, ou especifica, os centroides iniciais de cada aglomeração, (2) Analisa as informações observadas e atualiza a distância dos centroides, onde esse processo é repetido até a convergência ser alcançada, (3) Após chegar a convergência é feita uma leitura final dos dados onde cada observação é assinalada ao centroide mais próximo.

#### **4.3.1. Gráfico de Elbow**

Para a elaboração dos grupos, vamos seguir com a mesma metodologia utilizada no item 4.2.1. A partir da análise da Figura 5, definimos a criação de 4 grupos, mesma quantidade utilizada no método anterior.



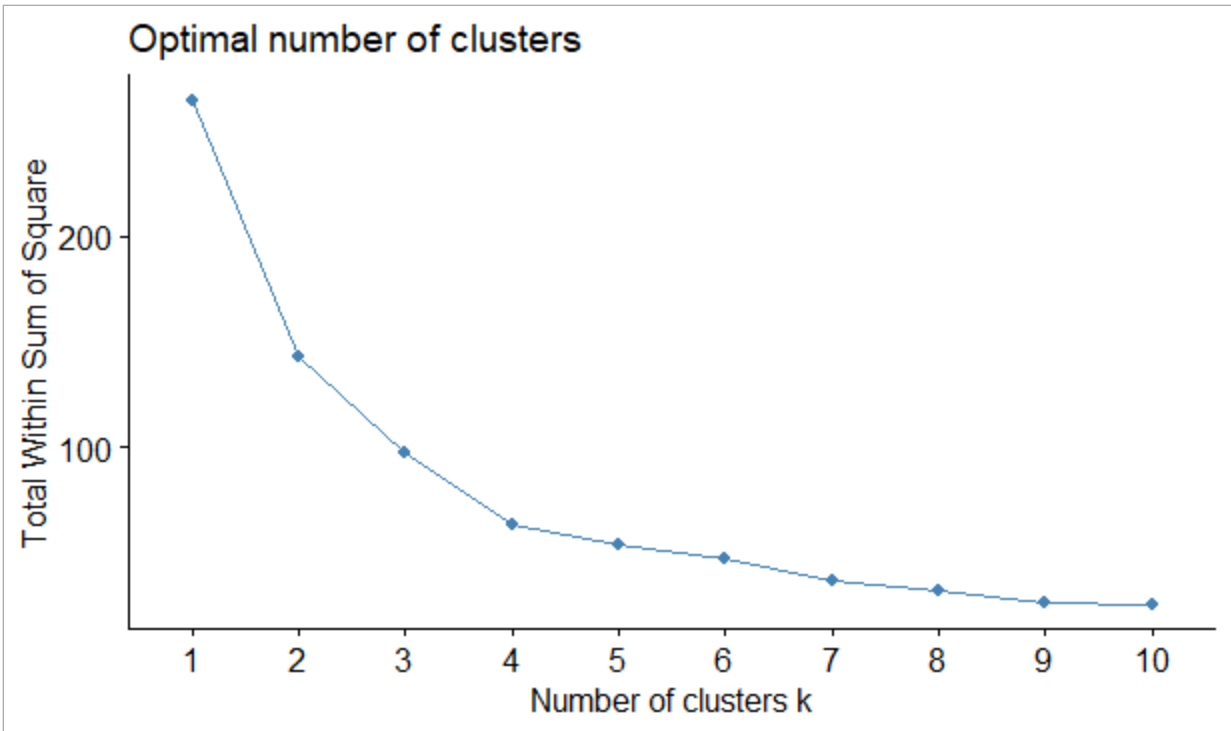


Figura 5 - Gráfico de Elbow para método k-means

#### 4.3.2. Centroides especificados

Para a elaboração dos grupos, vamos seguir com o valor escolhido no item 4.2.1, ou seja, 4 grupos. A Figura 4 apresenta a especificação dos centroides de cada grupo criado pelo algoritmo.

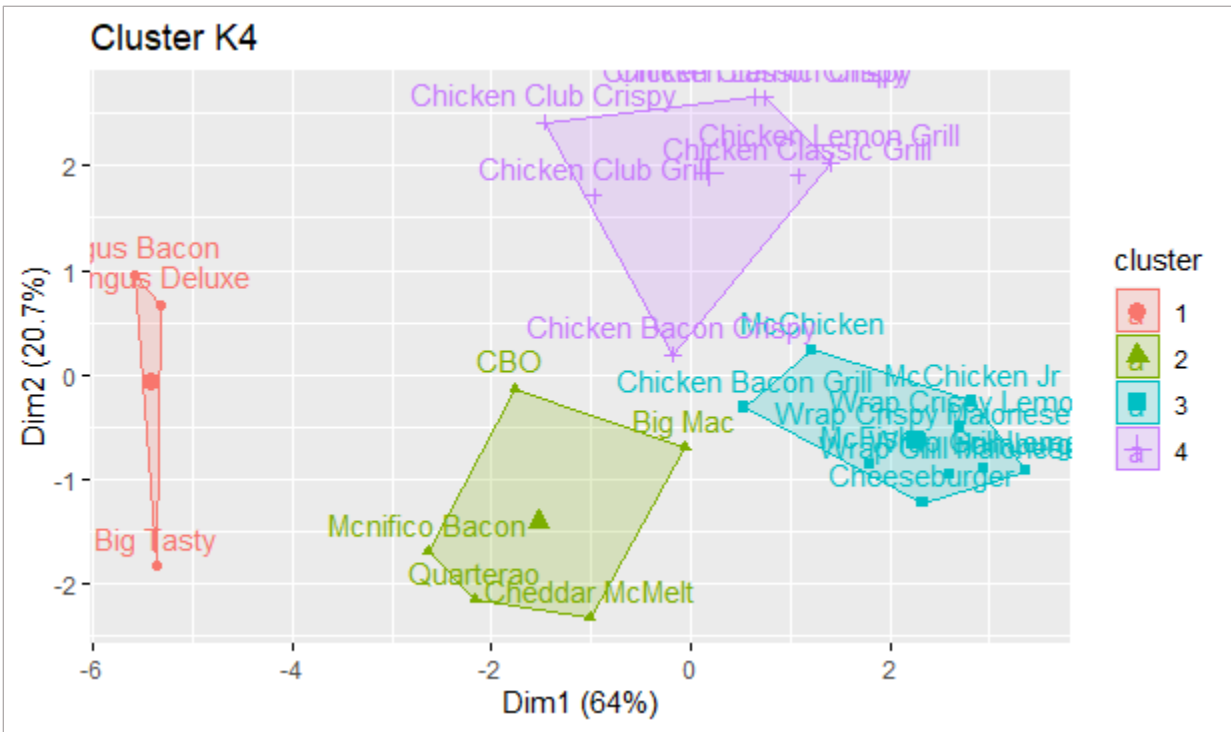


Figura 6 - Grupos criados e centroides definidos

## 5. RESULTADOS E DISCUSSÕES

Com base na análise hierárquica foi possível extrair o seguinte informativo estatístico dos grupos evidenciado na Tabela 2.

grupo_lanches4	n	Valor.Energetico	Carboidratos	Proteinas	Gorduras.Totais	Gorduras.Saturadas	Gorduras.Trans	Colesterol	Fibra.Alimentar	Sodio	Calcio	Ferro
1	12	377.5833	34.75000	19.50000	17.77500	5.650000	0.3000000	38.83333	2.316667	926.500	140.91667	3.107500
2	3	855.6667	52.66667	48.66667	50.00000	22.000000	1.7000000	129.33333	5.033333	1714.667	257.00000	5.333333
3	4	583.2500	40.75000	30.25000	33.25000	14.250000	1.1250000	83.25000	3.600000	1127.500	229.75000	9.475000
4	6	485.5000	49.33333	31.33333	17.96667	5.083333	0.2833333	57.16667	6.100000	1349.500	66.33333	2.683333

Tabela 2 - Resultado estatístico dos grupos

Como fica evidenciado, podemos determinar que: **Grupo(1)** Além de ser o com maior número de participantes, apresenta as opções mais *lights* do cardápio, nele se encontram os lanches com menor índice de “Sódio”, “Fibra alimentar”, “Carboidratos”, “Proteínas” dentre outros; O **Grupo(2)** representa as “bombas de carboidrato”, oposto ao Grupo (1) seria os lanches menos saudáveis e mais gordurosos; **Grupo(3)** Apresenta opções mais intermediárias, estaria classificado entre o Grupo(1) e o Grupo(2); **Grupo (4)** Realizou uma classificação interessante contendo os lanches montados com carne de frango, possuindo uma similaridade alta entre si.

### 5.1. Comparando os modelos estatísticos

Com base no site oficial do McDonald's os 3 principais produtos são os lanches: (1) *Big Tasty*, (2) *McNífico Bacon* e (3) *Big Mac*. O (1) encontra-se no grupo 2 e será o escolhido deste cluster; O (2) está no grupo 3 e será o representante deste agrupamento; Já o (3) encontra-se no grupo 1 pelo método hierárquico, porém quando analisado pelo não hierárquico ele fica localizado no Grupo(3) dos lanches intermediários, portanto nossa opção para representar os lanches mais *lights* será o *Cheeseburger*. Por último, nossa escolha do grupo 4 será o (4) *Chicken Classic Grill* seguindo o critério de popularidade dos lanches.

Quando comparamos os diagramas de Elbow para ambas técnicas, notamos uma variação muito sutil da variância entre os clusters a partir do terceiro grupo, como mostrado na Figura 7. Notou-se que um dos lanches escolhidos o (2) *McNífico Bacon* quando analisado pelo método não hierárquico, encontra-se no mesmo grupo do (3) *Big Mac* o que acaba levantando uma dúvida se estes são de fato diferentes entre si quando comparados com a maioria dos lanches.

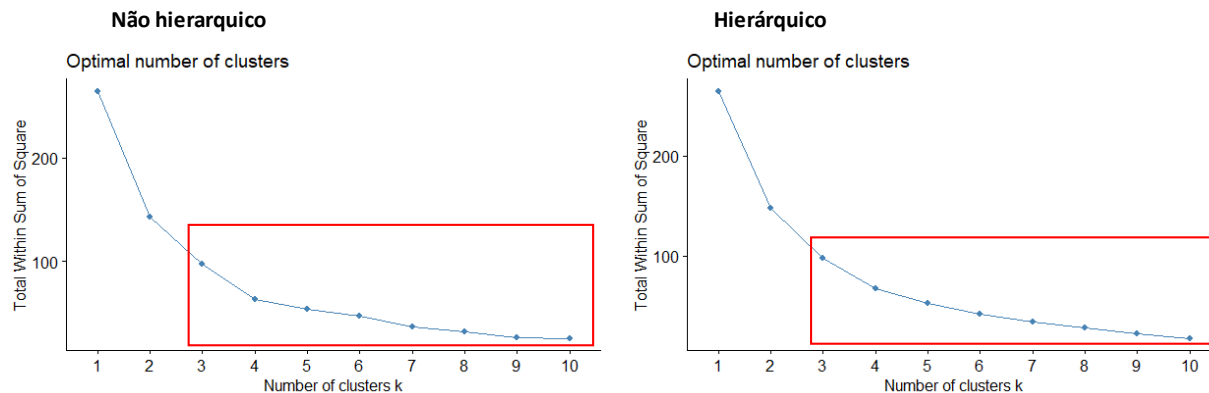


Figura 3 - Comparação da variância entre métodos

## 5.2. Incertezas

A análise fica sujeita a variação dos dados apresentados pelo McDonald's em sua base quando comparada com as informações apresentadas por outras instituições. Embora outros veículos de informação possam informar valores um pouco distintos, este estudo possui confiabilidade suficiente para a definição dos lanches para o comercial.

## 6. CONCLUSÃO

Por fim ficam evidentes neste trabalho os lanches escolhidos para serem transmitidos na propaganda com o foco de alcançar a máxima abrangência do público.

Grupo 1 – Light:	<i><b>Cheeseburger</b></i>
Grupo 2 – Ricos em carboidratos:	<i><b>Big Tasty</b></i>
Grupo 3 – Intermediários:	<i><b>McNífico Bacon</b></i>
Grupo 4 – Carne Branca:	<i><b>Chicken Classic Grill</b></i>

## 7. BIBLIOGRAFIA

PENG, Roger D.; MATSUI, Elizabeth. **The Art of Data Science**: a guide for anyone who works with data. Baltimore: LLC, 2016.

MCKINNEY, Wes. **Python para Análise de Dados**: tratamento de dados com pandas, numpy e ipython. 2. ed. São Paulo: Novatec, 2018.

GUERRA, Saulo; OLIVEIRA, Paulo Felipe de; MCDONNELL, Robert. **Ciência de dados com R**: introdução. São Paulo: Ibpad, 2018.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados**: um guia sobre visualização de dados para profissionais de negócios. Rio de Janeiro: Alta Books, 2017.

WHEELAN, Charles. **Estatística**: o que é para que serve como funciona. Rio de Janeiro: Zahar, 2016.

FIGUINHA, Renato. **Tudo Sobre o McDonald's**. 2005. Disponível em: <https://administradores.com.br/artigos/tudo-sobre-o-mc-donalds>. Acesso em: 18 jul. 2021.

T4.AI. **Fast Food Market Share**: about the fast food industry. About the Fast Food Industry. 2021. Disponível em: <https://www.t4.ai/industry/fast-food-market-share>. Acesso em: 18 jul. 2021.

VEJA (São Paulo). **Mais caro da Globo: Quanto custa um comercial no Jornal Nacional?** 2021. Disponível em: <https://vejasp.abril.com.br/cultura-lazer/quanto-custa-comercial-globo/>. Acesso em: 18 jul. 2021.

MCDONALD'S. **Cardápio**. 2021. Foi utilizado o seguimento de "Nossos Produtos" & "Clássicos". Disponível em: <https://www.mcdonalds.com.br/cardapio>. Acesso em: 18 jul. 2021.