



UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE QUÍMICA

PROJETO DE QUIMIOMETRIA

ESTUDO DE ÁGUA MINERAL COMERCIALIZADAS EM

BRASÍLIA-DF ATRAVÉS DE UM MODELO DE PCA

CAMILA CALANDRINY ROCHA DA COSTA

Brasília, DF

2022

INTRODUÇÃO

Água é o elemento indispensável para a sobrevivência de todos os organismos vivos, além de ser essencial para a manutenção do clima e ecossistemas da Terra. Águas minerais são águas subterrâneas que infiltram através do solo. De acordo com a Resolução da ANIVSA – RDC 173 de 13/11/2006 - ‘Água mineral é a água obtida diretamente de fontes naturais ou por extração de águas subterrâneas. É caracterizada pelo conteúdo definido e constante de determinados sais minerais, oligoelementos e outros constituintes considerando as flutuações naturais’.

A água mineral natural não apresenta adição de sais e pode ter ação medicamentosa. A diferença da “água de mesa” (água potável) é que esta tem composição normal provenientes de fontes naturais ou artificialmente captadas, que preencham os requisitos de potabilidade para a região. A água potável deve seguir parâmetros microbiológicos, físicos, químicos e radioativos definidos por lei e que não ofereça riscos à saúde humana.

Água mineral ainda pode ser distinguida pela composição química e pelas características da fonte. Neste trabalho há águas classificadas em: alcalino terrosas (carbonato de cálcio $\geq 120\text{mg/L}$), fluoretada (quando contiver fluoreto $\geq 0,02\text{ mg/L}$) e vanádica (quando contiver vanádio $\geq 0,03\text{ mg/L}$), litinada (quando contiver lítio $\geq 0,01\text{ mg/L}$), fracamente radioativas (teor de radônio mínimo entre 5 e 10 unidades Mache por litro, a 200C e 760 mm de Hg de pressão) e radioativas (teor de radônio entre 10 e 50 unidades Mache por litro, a 200C e 760 mm de Hg de pressão). Quanto a temperatura, as classificações das águas são hipotermiais, quando sua temperatura estiver compreendida entre 25 e 330C.

Hoje no mercado brasileiro há diversas marcas de águas minerais, de diversas fontes naturais e, também, com os mais variados preços. O objetivo desse trabalho é, com as informações fornecidas nos próprios rótulos das marcas, tentar identificar padrões e ver como essas diferentes marcas se relacionam. Também é verificar quais marcas se relacionam e como as fontes das águas se relacionam umas com as outras. Verificar quais propriedades são mais importantes para a correlacionar as diferentes águas. Há também um dado de uma água potável, não relacionada com as águas minerais, e o objetivo com essa amostra é ver quão relacionada ela é com as outras amostras de água mineral.

MÉTODOS

As amostras foram obtidas de dois mercados de Brasília (Big Box na 408 da Asa Norte e Atacadão no Setor Terminal Norte) nos dias 12 e 17 de abril deste ano. Foram coletadas as informações dos rótulos de 14 marcas de água mineral, sendo que 2 delas foram da mesma marca e mesma fonte, mas diferentes lotes. 4 amostras são da mesma marca, mas de diferentes fontes. A amostra de água potável foi obtida através de um laudo técnico de uma amostra de água de uma ETA localizada na Bahia e que foi coletada no dia 07 de fevereiro deste ano.

Os dados das amostras foram inseridos em uma tabela. O primeiro passo foi tratar os dados. Dados faltantes foram substituídos pela média e os dados foram autescalados centrando-os na média.

Foi feito um modelo de PCA através da decomposição em valores singulares (SVD) de dois componentes. Também foi feito uma análise hierárquica de agrupamentos (HCA) para analisar a distância entre as amostras através de um dendrograma. A linguagem utilizada foi Python, usando como ambiente computacional o Jupyter Notebook. Algumas bibliotecas e pacotes com funções e métodos já definidos foram usadas para esse estudo.

RESULTADOS

Foram analisadas 15 amostras no total. As tabelas com os dados são mostradas a seguir.

Tabela 1: Propriedades físico-químicas

Amostra	Marca	Fonte	Lote	Classificação	pH 25°C	Temperatura (C°)	Condutividade a 25°C (µS/cm)	Resíduo Evaporação (mg/L)	Radioatividade 760 de mmHg
1	Indaiá	Primavera	1	Alcalino terrosa fluoretada	7,27	20,8	327	198,62	NaN
2	Minalba	Água Santa	1	Fluoretada Fracamente Radioativa	7,76	17,9	143,5	109,95	9,79
3	Ambev	São José	1	Fluoretada Fracamente Radioativa	6,56	23	206	127,06	8,84
4	Crystal	Gênesis 1	1	Fluoretada e Hipotermal	7,38	25,2	147,7	82,71	NaN
5	Mariza	Mariza	1	Fluoretada Fracamente Radioativa	7,97	26,8	116,1	92,83	6,29
6	La Priori	Topázio	1	Fluoretada Litinada e Fracamente Radioativa	6,21	23	36,7	25,93	5,86
7	Nestlé	Esmeralda	1	Hipotermal	4,72	27,1	50,6	42,4	NaN
8	Hay Up	Quedas do Iguaçu	1	Fluoretada e Vanádica	9,75	21	370	246,09	NaN
9	Indaiá	Santa Rita 1	1	Hipotermal	4,52	28,2	79,3	47,45	NaN
10	Indaiá	Marisa	1	Fluoretada Fracamente Radioativa	7,57	18,4	150,2	117,65	7,68
11	Prata	Leve	1	Fluoretada Radioativa na fonte	6,49	22,3	76,7	73,47	23,08
12	Crystal	Gênesis 1	2	Fluoretada e Hipotermal	7,38	25,2	147,7	82,71	NaN
13	Indaiá	Esmeralda	1	Hipotermal	4,72	27,1	50,6	42,4	NaN
14	São Lourenço	Oriente	1	Alcalino terrosa fluoretada	5,63	22,3	590	338,69	NaN
15	Saída ETA	Embasa	1	Potável	6,83	5,8	307	199,6	NaN

Tabela 2: Composição química

Amostra	Bicarbonato (mg/L)	Ca (mg/L)	Mg (mg/L)	Cl ⁻ (mg/L)	NO ₃ ⁻ (mg/L)	Na (mg/L)	K (mg/L)	SO ₄ ²⁻ (mg/L)	F ⁻ (mg/L)	Sr (mg/L)	Ba (mg/L)	Br ⁻ (mg/L)	PO ₄ ³⁻ (mg/L)	Li (mg/L)	Zn (mg/L)	V (mg/L)
1	169,14	31,000	16,500	10,850	5,150	4,180	3,990	1,110	0,11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	110,78	16,800	11,100	0,11	0,77	1,210	1,240	0,13	0,05	0,02	0,018	NaN	NaN	NaN	NaN	NaN
3	75,05	13,900	5,420	7,100	9,720	10,000	2,480	2,260	0,07	0,086	0,073	0,02	0,14	NaN	NaN	NaN
4	78,51	21,303	2,201	0,05	0,11	1,009	0,855	0,9	0,06	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	70,02	12,613	1,820	0,04	NaN	8,372	1,643	1,660	0,15	0,086	NaN	NaN	NaN	NaN	NaN	NaN
6	17,64	0,465	3,523	0,04	0,01	0,135	0,223	1,810	0,02	NaN	NaN	NaN	NaN	0,011	NaN	NaN
7	6,92	1,080	0,937	7,310	1,040	5,036	0,755	1,870	NaN	NaN	NaN	0,03	NaN	NaN	NaN	NaN
8	104,57	0,825	0,072	0,2	NaN	86,145	0,524	2,140	0,32	0,01	NaN	NaN	NaN	0,005	0,01	0,407
9	1,76	0,264	0,635	17,300	3,540	11,641	0,796	0,95	NaN	NaN	0,022	0,03	NaN	NaN	NaN	NaN
10	118,09	17,800	11,900	NaN	0,24	1,110	1,190	0,11	0,09	0,015	0,015	NaN	0,04	NaN	NaN	NaN
11	50,71	8,800	4,070	0,85	1,16	2,390	2,760	2,95	0,22	0,109	0,051	NaN	0,37	NaN	NaN	NaN
12	78,51	21,303	2,201	0,05	0,11	1,009	0,855	0,9	0,06	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	6,92	1,080	1	7,31	1,04	5,036	0,755	1,87	NaN	NaN	NaN	0,03	NaN	NaN	NaN	NaN
14	361,79	36,934	16,657	0,11	1,68	38,839	41,934	1,93	0,05	0,052	0,486	NaN	NaN	0,041	NaN	NaN
15	34,9	34,900	34,900	40,400	0,1	NaN	NaN	6,100	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Os valores definidos como NaN foram valores que não estavam nos rótulos ou não informados. Como se pode ver na tabela, muitas marcas não têm muitos dados e por isso foram escolhidos apenas algumas variáveis de estudo: pH, temperatura, condutividade, resíduo de evaporação, bicarbonato, cálcio, magnésio, cloreto, nitrato, sódio, potássio e sulfato.

Antes de começar a tratar os dados, alguns valores de cloreto, nitrato, sódio e potássio estão faltando em algumas amostras. Decidiu-se substituir esses valores pelos valores de média da coluna.

Em seguida, os dados foram autoescalados e centrados na média.

Com os valores tratados, fez-se um dendrograma para verificar graficamente a similaridade entre as amostras. A distância entre as amostras foi a distância euclidiana.

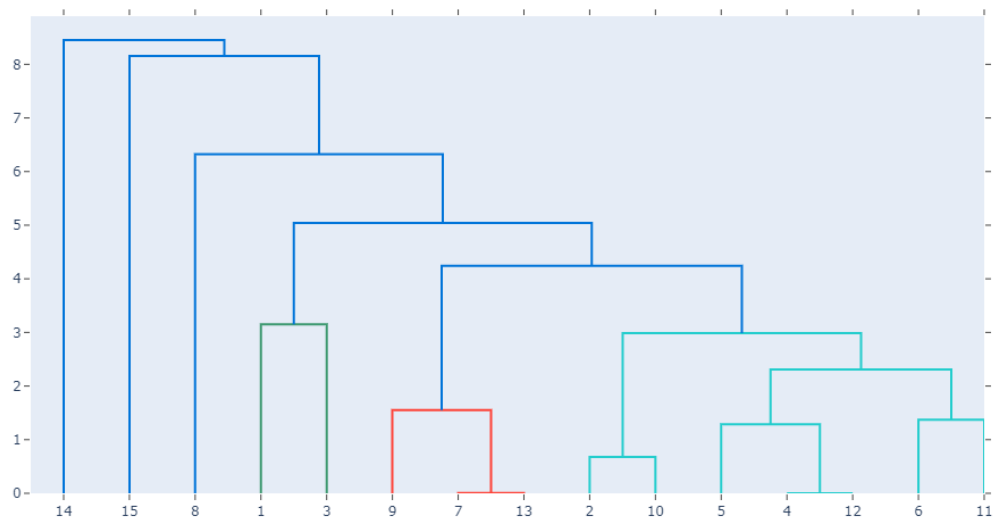


Figura 1: Dendrograma de amostras de água usando dados autoescalados.

Algumas informações já podem ser retiradas do dendrograma. Já se pode ver que as amostras 4 e 12 são iguais, o que já era esperado uma vez que são da mesma marca (Crystal) e fonte (Gênesis 1). As amostras 7 e 13 também são iguais o que era esperado uma vez que a fonte é a mesma e os mesmos dados foram fornecidos, a distinção fica apenas na marca. As amostras 14 e 15 são as mais distintas entre todas. A amostra 15 é a amostra de água potável e mostra que realmente a água potável é diferente de águas minerais.

Em seguida, um mapa de calor com as correlações entre as variáveis foi obtido.

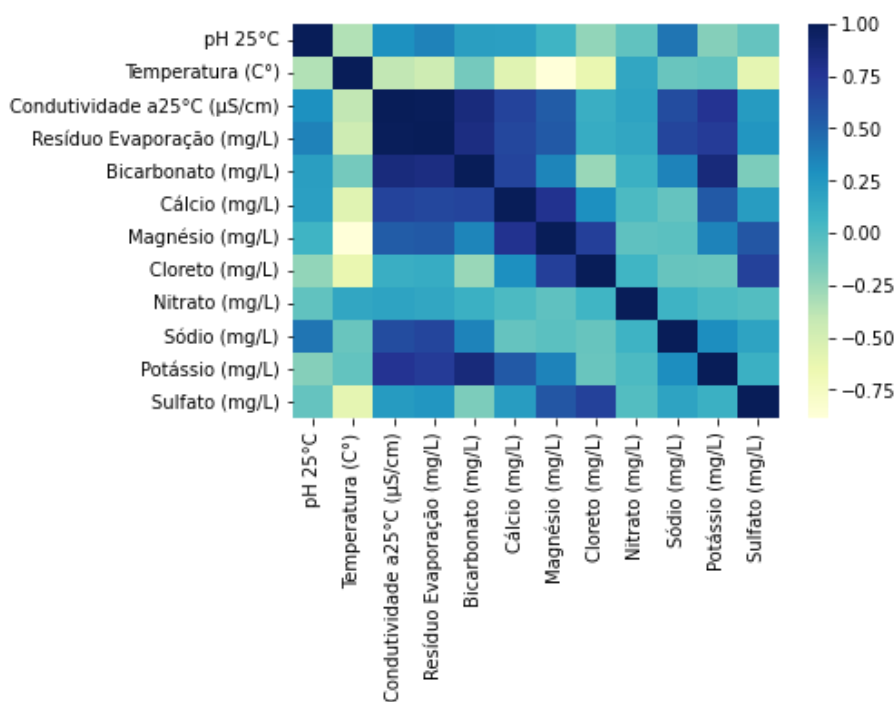


Figura 2: Mapa de calor correlacionando as variáveis.

Com o mapa de calor, pode-se ver que o resíduo de evaporação tem uma forte correlação com a condutividade. O cloreto também tem uma boa correlação entre o sulfato e a temperatura não tem quase correlação com o magnésio. Assim, com o mapa de calor, pode-se ver fatores que são dependentes ou independentes entre si dentro da análise de água.

Partindo agora para a análise com PCA, a tabela a seguir mostra a variância explicada para componente e a variância acumulada.

Tabela 3: Componentes principais e variâncias

Componente Principal	Variância Explicada (%)	Variância Acumulada (%)
1	42,70	42,70
2	22,91	65,61
3	12,40	78,01
4	10,06	88,07
5	7,72	95,79
6	2,41	98,21
7	1,41	99,61
8	0,19	99,80
9	0,14	99,94
10	0,04	99,98
11	0,01	99,99
12	0	100

Para dois componentes principais, é explicada 65,61% da variância. As duas primeiras componentes são as que mais explicam a variabilidade dos dados. 5 componentes principais explicam mais de 95% e é o suficiente para dar informações importantes. Porém, para esse estudo, o foco será as duas primeiras componentes principais.

Os gráficos seguintes são os gráficos de escores e pesos para duas componentes principais.

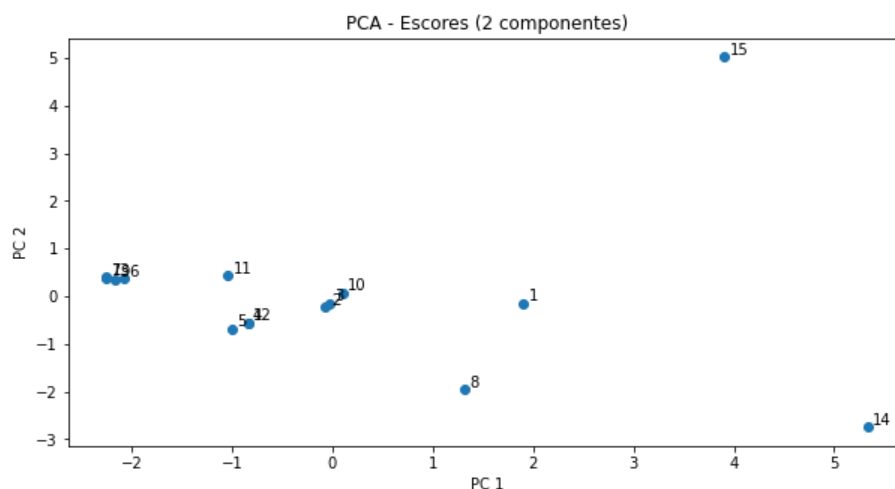


Figura 3: Gráfico de escores para 2 componentes

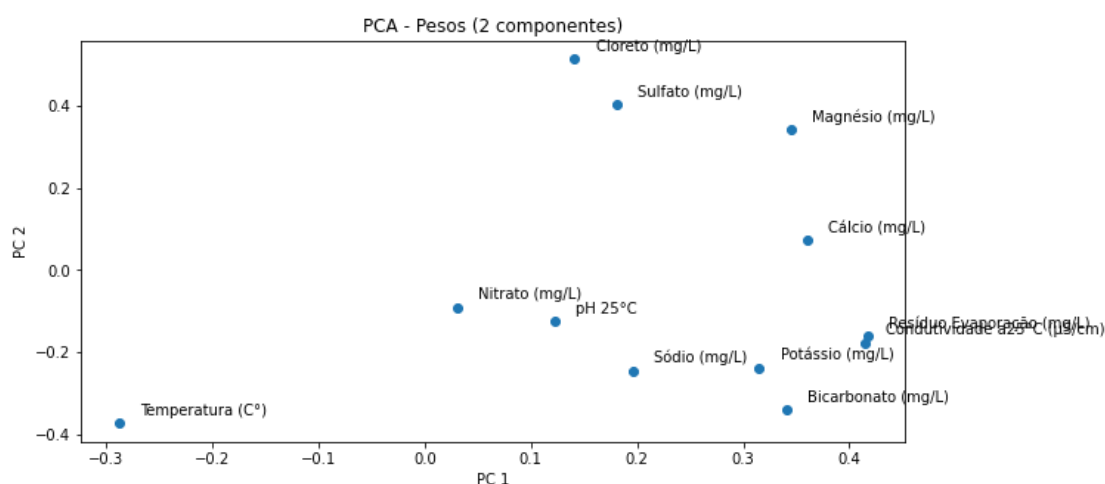


Figura 4: Gráfico de pesos para 2 componentes

No gráfico de pesos, é possível ver a formação de grupos definidos de amostras que se relacionam. As amostras 2, 3 e 10 estão bem próximas, mostrando que as fontes Água Santa, São José e Marisa se relacionam. As marcas Minalba, Ambev e Indaiá (Marisa) acabam se relacionando por consequência. As amostras 4, 5 e 12 também são bem próximas, assim, as fontes Mariza e Genesis 1 se relacionam e, consequentemente as marcas Crystal e Mariza, lembrando que as amostras 4 e 12 diferem somente do lote. Há ainda um *cluster* mais a esquerda no gráfico onde se juntam as amostras 6, 7, 9 e 13. As amostras 7 e 13 são da mesma fonte, mas marcas diferentes. Dessa forma, as fontes Topázio, Esmeralda e Santa Rita 1 se relacionam, então as marcas La Priori, Nestlé e Indaiá são águas bem próximas entre si.

As amostras 14 e 15 são as que mais diferem entre todas as amostras e as que mais estão afastadas. A amostra 15 era esperado, uma vez que ela é uma amostra de uma água potável. A 14 também tem propriedades bem diferenciadas das outras

amostras de água mineral., a fonte dela tem uma composição muito distinta de todas as outras.

Algumas proximidades já haviam sido notadas no dendrograma, como a relação entre as amostras 7,9 e 13, as amostras 5,4 e 12 e a 2 e 10. Já o dendrograma se difere nas amostras 1 e 3 que coloca as junto, enquanto no gráfico essas amostras têm uma distância considerável.

Quanto ao gráfico de pesos, a condutividade e o resíduo de evaporação estão bem próximos e tem uma forte relação, como já se havia notado no mapa de calor. A temperatura é a que menos tem influência pois é a que mais se afasta, seu peso é baixo. Ao fazer a tabela com os pesos dos componentes principais, dá para saber quem tem os maiores pesos e, assim, quais variáveis são importantes para fornecer as maiores informações.

Tabela 4: Pesos e componentes principais

Componente Principal	Variável	Peso	Tipo
PC1	Resíduo de Evaporação	0,417655	Bom
PC2	Cloreto	0,513858	Bom
PC3	pH 25°C	0,656209	Bom
PC4	Nitrato	0,588739	Bom
PC5	Nitrato	0,756585	Bom

Como já se tinha visto antes, até a quinta variável, se explica 95% das informações e as variáveis que explicam são as apresentadas na tabela, sendo o resíduo de evaporação e o cloreto as duas primeiras componentes principais.

O gráfico seguinte mostra uma elipse com suas bordas representando o limite de 95% de variância. É uma elipse de T^2 Hotelling que expressa a distância que uma amostra se encontra do centro do modelo de PCA. Valores superiores do limite de T^2 significam uma maior distância do centro

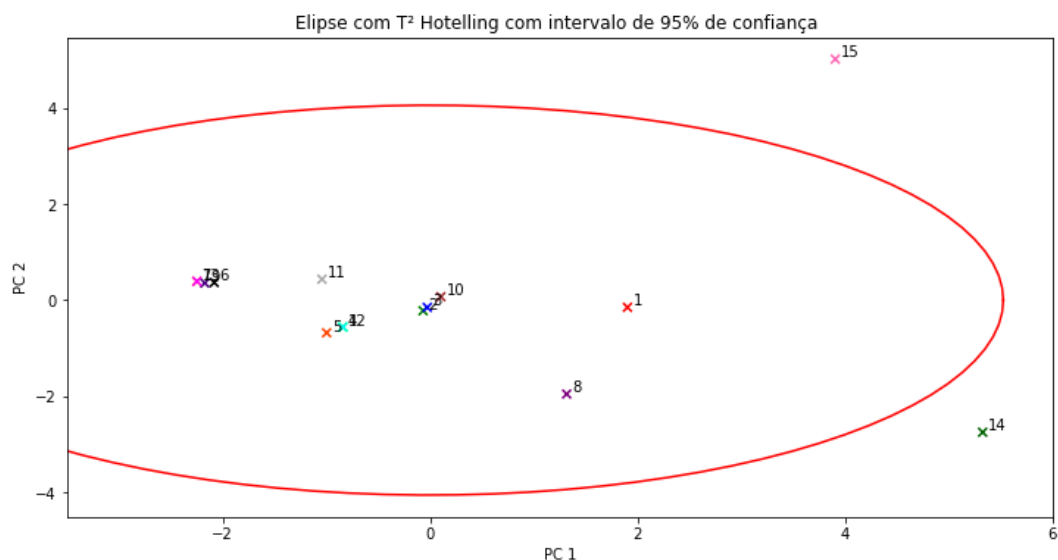


Figura 5: Elipse T² Hotelling 95% de confiança

As amostras 14 e 15 estão fora dos limites de confiança de 95%. Algo já esperado para a amostra 15. A amostra 14 realmente é bem diferente de todas as outras águas minerais.

O gráfico seguinte mostra a relação dos pesos com os escores e os limites de T² Hotelling.

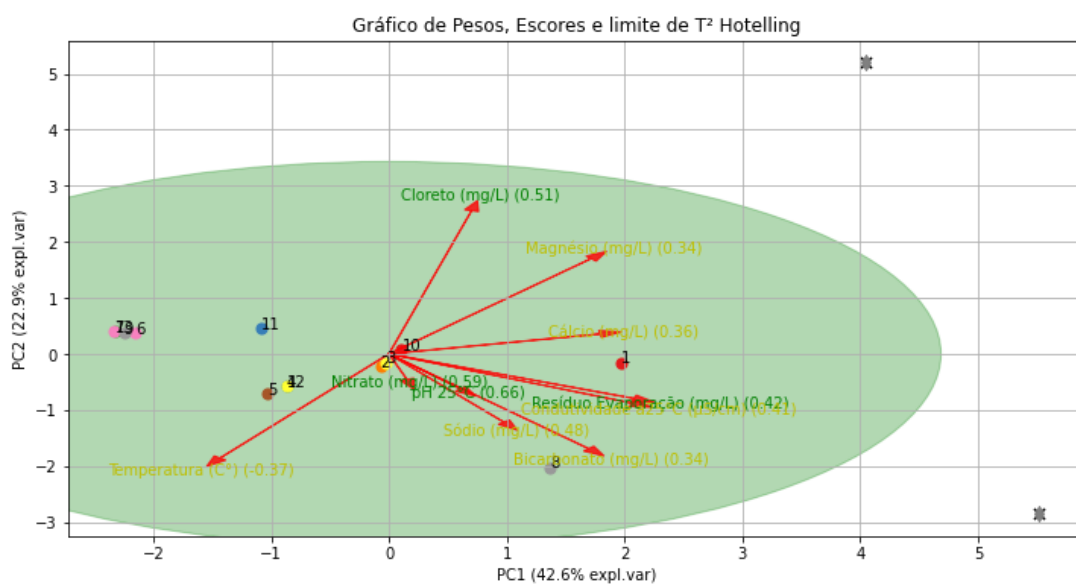


Figura 6: Gráfico de pesos, escores e limite de T² hotelling

As setas vermelhas são os pesos. Os pesos em verde são aqueles que explicam bem a variabilidade dos dados, enquanto os pesos em amarelo não são tão importantes assim para explicar a variância. A elipse verde representa o limite de T² hotelling e os pontos coloridos são as amostras. Vê-se que os pesos nitrato e pH estão bem próximos e que as amostras 2,3 e 10 estão bem relacionadas com nitrato, pH e cloreto, além de

estarem bem no centro do modelo. As amostras 5 e 4/12 estão também próximas do nitrato e temperatura (apesar de não ter um peso expressivo). A amostra 1 está mais próxima do cálcio, do resíduo de evaporação e da condutividade. As amostras 14 e 15 estão fora do limite.

CONCLUSÃO

O estudo de diferentes amostras de água mineral comercializadas em Brasília através de um modelo de PCA retornou bons resultados. Duas componentes principais conseguem explicar mais de 65% das informações sobre as amostras. Também foram identificadas as componentes principais com as maiores relevâncias. As amostras formaram grupos de correlações que conseqüentemente trouxeram correlações de marcas e fontes de água mineral.

Uma possível aplicação para esse estudo é mostrar ao consumidor quais marcas são mais próximas entre si e assim, de acordo com sua preferência, poder escolher entre essas marcas de acordo com preço e disponibilidade no mercado.

Idealmente, seria necessário analisar mais mostras de lotes diferentes e acompanhar a mudança de alguma composição físico-química com o passar do tempo para verificar se as componentes principais e as correlações permanecem as mesmas. Apesar de ter duas amostras de lotes diferentes, os valores das variáveis eram os mesmos, o que pode se dizer que são amostras iguais. Também, para duas amostras de mesma fonte, os valores eram os mesmos, sendo amostras iguais. Isso pode também chamar a atenção para uma melhor fiscalização para realmente verificar se as empresas estão fornecendo os valores corretos nos rótulos.

Com mais algumas amostras de água mineral, especialmente da mesma marca da amostra 14, talvez seja possível treinar um modelo que faça a classificação de água mineral. Uma visão mais otimista também seria poder treinar um modelo em que, com uma amostra desconhecida, pudesse identificar essa amostra e ainda determinar a qual fonte e marca ela pertence.

REFERÊNCIAS

ÁGUA MINERAL: O QUE É, DE ONDE VEM E COMO SE CLASSIFICA. **Serviço Geológico do Brasil – CPRM**. Disponível em: https://aguamineral.cprm.gov.br/agua_mineral.html. Acesso em 14 de abril de 2022.

BRASIL. **Decreto nº 5.440, de maio de 2005**. Estabelece definições e procedimentos sobre o controle de qualidade da água de sistemas de abastecimento e institui mecanismos e instrumentos para divulgação de informação ao consumidor sobre a qualidade da água para consumo humano. Brasília, 4 de maio de 2005, 184º da Independência e 117º da República. Disponível em: <https://caesb.df.gov.br/legislacao1/decretos/218-decreto-5-440-05-estabelece-definicoes-e-procedimentos-sobre-o-controle-de-qualidade-da-agua-de-sistemas-de-abastecimento-e-institui-mecanismos-e-instrumentos-para-divulgacao-de-informacao-ao-consumidor-sobre-a-qualidade-da-agua-para-consumo-humano.html>. Acesso em: 13 de abril de 2022.

ERDOGAN TASKESSEN. **PCA: python package**. Disponível em: <https://github.com/erdogant/pca>. Acesso em 12 de abril de 2022.

LAUDO DE ANÁLISES DE ÁGUA. **CÓD – AG2022146**. Bahia Analítica. Disponível em: https://www.embasa.ba.gov.br/images/Institucional/nossosservicos/analiselaboratorial/2022/canavieiras_fevereiro%202022_laudo%20qualidade%20da%20%C3%A1gua_coleta%20em%202007%20de%20fevereiro.pdf. Acesso em 14 de abril de 2022.