



# **Análise da relação entre características demográficas e hábitos de consumo entre jovens**

**Gabrielly Castilho Guimarães – 805757**

**Luiza Neves Graça – 791328**

**Miguel Antonio de Oliveira – 772180**

**Rubens Neto – 759550**



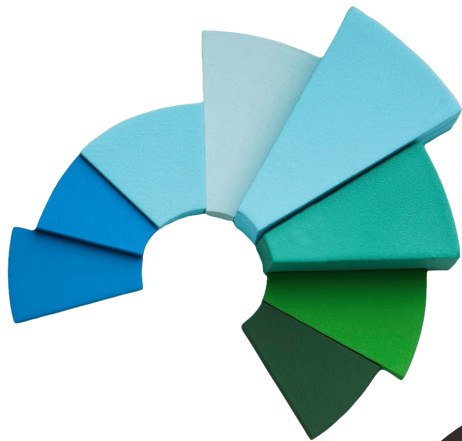
# Introdução e Obejtivo



- . Traçar uma relação hábitos de consumo e o seu perfil demográfico
- . Realizar um estudo de dados demograficos
- . Auxiliar análises de mercado e publico alvo
- . Agrupar individuos em segmentos
- . Conhecer melhor os habitos de consumo dos jovens

# A Base de dados

Baseados em um estudo de 2013 que entrevistou 1010 jovens de 15 a 30 anos e coletou ao todo dados sobre 150 variáveis: (i) 17 atributos referentes ao gosto musical dos respondentes; (ii) 12 atributos referentes às preferências de filmes; (iii) 32 atributos sobre hobbies e interesses; (iv) 10 atributos em relação a fobias; (v) 3 atributos sobre os hábitos de saúde; (vi) 57 atributos referentes à personalidade e opiniões; (vii) 7 atributos referentes aos hábitos de consumo; e (viii) 10 atributos referentes ao perfil demográfico dos participantes.



# Os atributos utilizados foram:



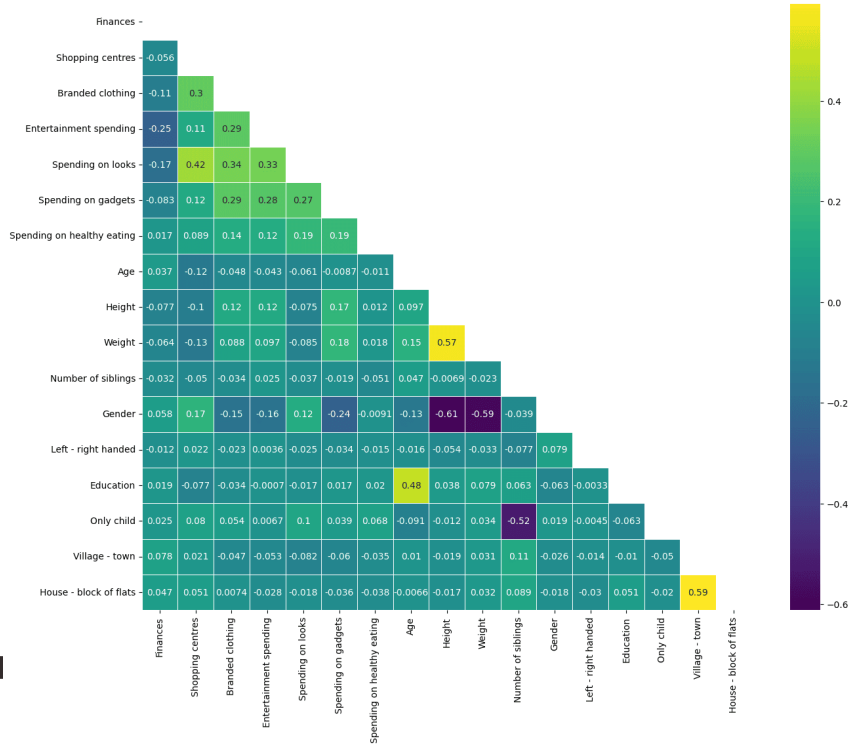
## Atributos referentes ao perfil demográfico

- Idade (assume valores inteiros de 15 a 30)
- Altura (assume valores inteiros em cm)
- Peso (assume valores inteiros)
- Quantidade de irmãos
- Gênero (binário)
- Mão dominante
- Se é um filho único (binário)
- Se cresceu em uma cidade ou vila
- Se cresceu em apartamento ou casa

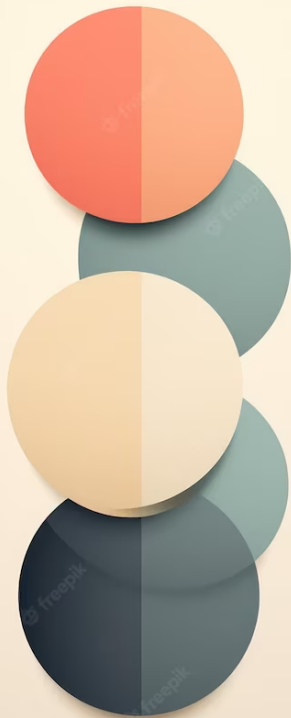
## Atributos referentes aos hábitos de consumo

- Para cada item abaixo, as pessoas responderam o quanto concordam com a frase em uma escala de 1 a 5, sendo 1 "discordo totalmente" e 5 "concordo totalmente"
- Eu economizo todo o dinheiro que posso
  - Eu gosto de frequentar Shoppings e grandes centros comerciais
  - Prefiro roupas de marca
  - Gasto muito dinheiro com festas e eventos sociais
  - Gasto muito dinheiro com a minha aparência
  - Gasto muito dinheiro com gadgets e aparatos tecnológicos
  - Eu não me incomodaria de pagar mais por uma comida boa, de qualidade ou saudável

# Pré processamento e agregação



## • Agrupamento K-Means

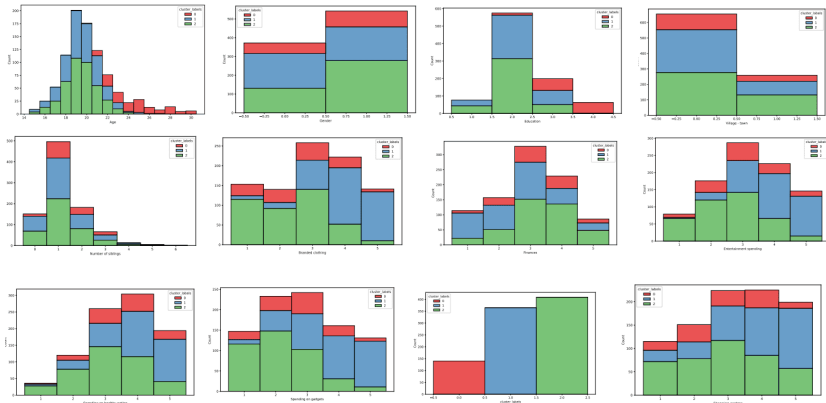
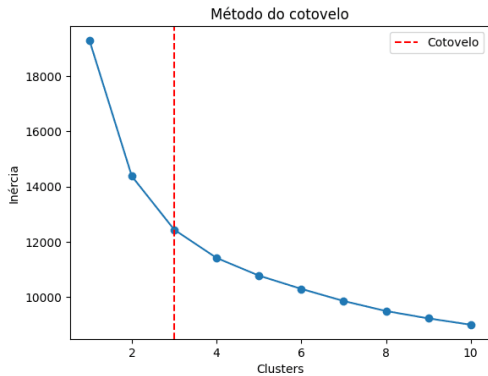


-O primeiro grupo, representado em vermelho, contém pessoas mais velhas e com uma escolaridade maior, frequentam shoppings e consomem comidas gourmets ou roupas de grife, e moderadamente economizam dinheiro em gadgets eletrônicos.

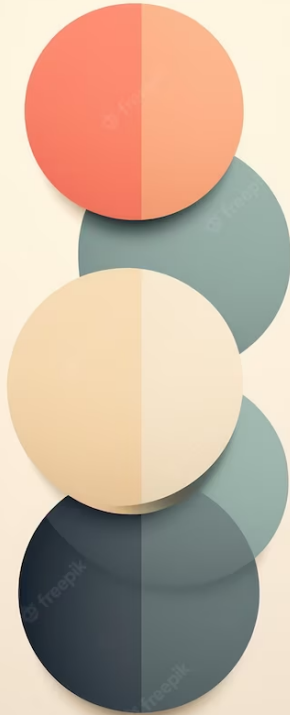
-O segundo grupo, em azul, é possuí pessoas mais jovens com um grau de escolaridade médio. Essas pessoas tendem a valorizar bem menos a economia de dinheiro, e tendem a gastar mais, com roupas de marca, festas, idas em Shoppings, comidas, e gadgets tecnológicos.

-O terceiro grupo, respresentado em verde. Ele é caracterizado por pessoas mais jovens que o primeiro grupo e possuem uma escolaridade menor. Seus hábitos financeiros são, estas são pessoas que valorizam muito mais a poupança de dinheiro e estão menos inclinadas a gastar com supérfluos como roupas de marca, festas e eventos, Shoppings, comidas gourmet, e aparelhos eletrônicos.

# K-means



# Agrupamento



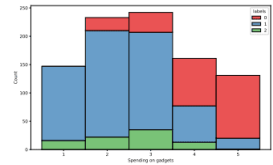
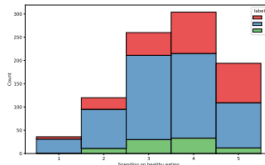
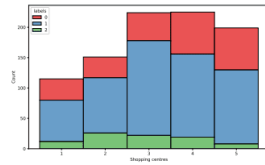
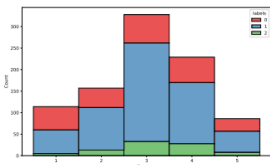
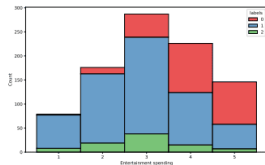
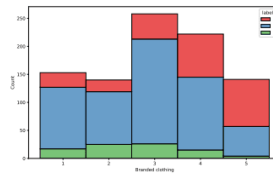
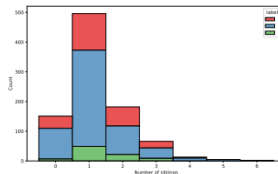
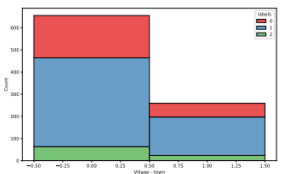
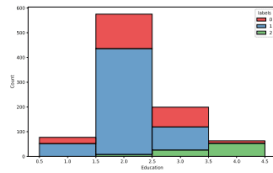
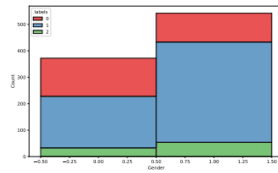
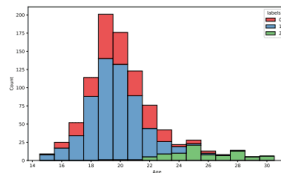
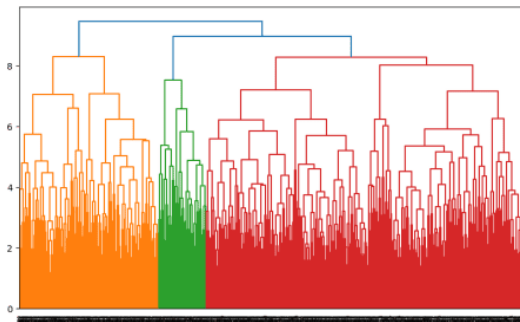
-O primeiro cluster, vermelho, é caracterizado por 20 anos em média e maioria homens sem ensino superior. Os seus hábitos financeiros, o grupo se importa menos com despesas financeiras e, de todos os grupos, é o que menos se importa com pagar mais caro para roupas de marca, festas, comidas gourmet ou saudáveis e, sobretudo, tendem a gastar mais com aparelhos eletrônicos

-O segundo cluster, azul, similar ao anterior sendo composto majoritariamente por mulheres. Esse grupo valoriza moderadamente a poupança de dinheiro e se opõe a pagar mais caro por roupas de marca, festas e comidas gourmet ou saudáveis e, acima de tudo, preferem não pagar muito por aparelhos eletrônicos.

-O terceiro grupo, verde, é o grupo em menor volume de observações. Eles tendem a representar pessoas mais velhas, em média com 25 anos, todos com grau de escolaridade superior aos outros grupos. O grupo tende a se importar mais com poupar dinheiro e tendem a se opor a pagar mais caro por roupas de marca, festas, idas em Shoppings e aparelhos eletrônicos. No entanto, não se importam tanto e pagar mais por comidas saudáveis ou gourmet.



# Agrupamento Hierárquico

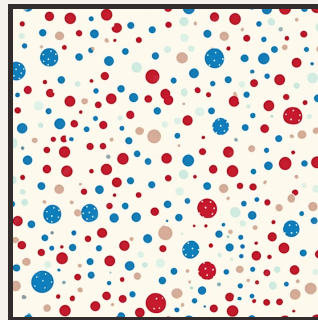


# Validação e comparação de modelos



# Modelos

De início, selecionamos as métricas relativas Silhouette e Davies-Bouldin, cujos resultados podem ser observados abaixo



Critérios Relativos		
	K- Means	Agrupamento Hierárquico
Silhouette	0.1481	0.094
Davies-Bouldin	2.09	2.50

**Em um primeiro momento notamos os baixos valores para o índice Silhouette, o que é esperado, uma vez que ele apenas capta separações lineares nos dados, resultando em baixos valores quando temos outros tipos de separação ou muitas dimensões.**



Em um primeiro momento notamos os baixos valores para o índice Silhouette, o que é esperado, uma vez que ele apenas capta separações lineares nos dados, resultando em baixos valores quando temos outros tipos de separação ou muitas dimensões.

Em seguida, para auxiliar na comparação dos dois modelos, também calculamos o índice Rand utilizando os grupos de um modelo para avaliar o outro.

Validação Externa	
Métrica	Valor
Rand	0.094

## Conclusão

- Desenvolvemos e avaliamos dois modelos de agrupamento para uma população jovem com base em dados demográficos e financeiros.
- No K-Means, testamos vários valores de K e escolhemos o melhor com base na inércia do método do cotovelo.
- No Agrupamento Hierárquico, utilizamos o dendograma para identificar partições potenciais.
- Validamos e comparamos os modelos, observando resultados semelhantes nas métricas internas, mas diferenças na métrica externa Rand.
- Concluímos que ambos os modelos são igualmente adequados para agrupar indivíduos por perfil socioeconômico.
- As métricas modestas sugerem influência de outros fatores nos resultados e a possível insignificância das diferenças nos clusters identificados pelos modelos.





# Obrigado!

Alguma duvida?  
O que acharam da aprensetação?

