

## Análise de Dados para Business Intelligence

2024/2025

### Avaliação Teórico-Prática 2

#### Enquadramento

O *SaborUrbano* é um restaurante inovador que utiliza análise de dados e machine learning para aprimorar a aquisição e retenção de clientes. Para identificar padrões no comportamento dos consumidores e otimizar campanhas de marketing, lançou um desafio exclusivo. A gerência do restaurante pediu o nosso apoio para realizar um estudo aprofundado sobre o comportamento dos seus clientes.

Para tal, disponibilizou-nos uma base de dados com uma amostra representativa do seu universo de clientes. Nestes dados, podemos encontrar:

Variável	Descrição
CustomerID	Identificação única do cliente
Name	Nome do cliente
Age	Idade do cliente
Education	Nível de escolaridade do cliente
Marital_Status	Estado civil do cliente
Gender	Género do cliente
Income	Rendimento anual familiar do cliente
Kid_Younger6	Número de crianças menores de 6 anos no agregado familiar
Children_6to18	Número de crianças entre 6 e 18 anos no agregado familiar
Time_Adherence	Número de meses desde a adesão do cliente ao cartão do restaurante
Recency	Número de dias desde a última compra do cliente
MntMeat&Fish	Montante gasto em pratos de carne e peixe
MntEntries	Montante gasto em entradas
MntVegan&Vegetarian	Montante gasto em pratos veganos e vegetarianos
MntDrinks	Montante gasto em bebidas
MntDesserts	Montante gasto em sobremesas
MntAdditionalRequests	Montante gasto em pedidos adicionais
NumOfferPurchases	Número de compras feitas com ofertas promocionais
NumAppPurchases	Número de compras feitas através de apps de entrega de alimentos
NumTakeAwayPurchases	Número de compras para take-away
NumStorePurchases	Número de compras realizadas na loja
NumAppVisitsMonth	Número médio de acessos ao restaurante em apps de entrega de alimentos

Complain	Indicador de que o cliente fez uma reclamação
Response_Cmp1	Indicador de que o cliente aceitou a oferta na campanha 1
Response_Cmp2	Indicador de que o cliente aceitou a oferta na campanha 2
Response_Cmp3	Indicador de que o cliente aceitou a oferta na campanha 3
Response_Cmp4	Indicador de que o cliente aceitou a oferta na campanha 4
Response_Cmp5	Indicador de que o cliente aceitou a oferta na campanha 5

## Notas sobre a resolução

A resolução deverá ser **entregue num documento em PDF**, que deverá incluir:

- Nomes dos elementos do grupo e respetivos números de aluno.
- A questão a responder seguida da respetiva resposta.
- Todos os outputs, gráficos ou tabelas utilizados para responder às questões devem estar incluídos na resposta.
- **O código utilizado para a resolução do problema não deve ser incluído na resposta.**

## I. Análise Preditiva de Dados - Regressão Linear:

1. A gerência do restaurante deseja prever os gastos dos clientes em pratos veganos e vegetarianos (MntVegan&Vegetarian) com base em outras variáveis do dataset.
  - a. Crie um modelo de Regressão Linear Múltipla para prever **MntVegan&Vegetarian**, com 4 variáveis explicativas numéricas à sua escolha. Justifique a sua escolha e apresente o modelo teórico.
  - b. Qual é a qualidade do ajustamento? Justifique e interprete.
  - c. O model é válido? Justifique.
  - d. Interprete os coeficientes de cada uma das variáveis incluídas no modelo.
  - e. Investigue sobre a significância das variáveis incluídas no modelo e justifique. Dessas variáveis, existe alguma que não é estatisticamente significativa? Se sim, o que deve ser feito e porquê?
  - f. Verifique os pressupostos do modelo. Apresente os resultados.
  - g. Verifique a presença de multicolinearidade no modelo. Justifique.
  - h. Inclua no modelo as variáveis **Gender** e **Response\_Cmp4**. Como interpreta os coeficientes associados no contexto do problema?

## II. Análise Descritiva de Dados Multivariados – Análise em Componentes Principais:

O restaurante deseja entender quais variáveis têm maior impacto no comportamento geral dos clientes, reduzindo a complexidade do dataset.

1. Aplique PCA para reduzir a dimensionalidade do dataset, utilizando as seguintes variáveis: MntMeatFish, MntEntries, MntVeganVegetarian, MntDrinks, MntDesserts, MntAdditionalRequests.
  - a. Apresente as medidas de adequabilidade dos dados a esta técnica exploratória. O que se pode concluir?
  - b. Quantas componentes são extraídas pelos vários métodos de extração de componentes principais? Justifique.
2. Utilize 3 componentes principais.
  - a. Qual a variância total explicada por esta solução? Justifique.
  - b. Aplique a rotação de Varimax. Interprete a solução e atribua nomes sugestivos a cada uma das componentes extraídas.
  - c. Construa os índices associados às componentes principais, utilizando a soma das variáveis. Podemos utilizar estes índices em substituição das componentes principais? Justifique.

## III. Análise Descritiva de Dados Multivariados – Análise de Clusters:

A equipa de marketing deseja criar campanhas personalizadas baseadas no comportamento e nas características dos clientes. Para isso, é essencial identificar grupos de clientes com comportamentos semelhantes. Esses grupos serão usados para definir ações específicas, como promoções, ofertas ou recomendações personalizadas.

1. Utilize os índices calculados na ACP, **Income** e **Age** como variáveis para o clustering.
  - a. Construa um dendograma utilizando a métrica de distâncias Euclidiana e o método de Ward. Identifique o número de clusters que achar mais adequado com base no dendograma. Justifique.
  - b. Aplique o método de K-means para formar 4 grupos.
  - c. Calcule os centros finais dos clusters, apresente os resultados em gráficos que achar adequados.

- d. Que características distinguem os grupos formados? Justifique.
- e. Os 4 clusters são formados por quantos clientes? Era espectável termos discrepância no número de casos em cada cluster? Justifique.
- f. Sugira uma campanha de marketing direcionada para um dos clusters. Explique como os dados analisados justificam essa escolha.