

## QUESTIONÁRIO CIENCIA DE DADOS

### 1. O que é a função `pd.read_csv()` no pandas?

- a) Carrega dados de um arquivo CSV em uma variável do tipo string
- b) Carrega dados de um arquivo CSV em um DataFrame
- c) Cria um gráfico a partir de um arquivo CSV
- d) Cria uma nova coluna em um DataFrame a partir de um arquivo CSV
- e) Converte um arquivo CSV em formato Excel

**Resposta correta: b) Carrega dados de um arquivo CSV em um DataFrame**

**Feedback:** A função `pd.read_csv()` é usada para carregar dados de arquivos CSV em um DataFrame no pandas, permitindo análise e manipulação de dados tabulares.

### 2. Como você monta a conexão com o Google Drive no Google Colab?

- a) `from google_drive_downloader import GoogleDriveDownloader as gdd`
- b) `from google.colab import drive`
- c) `drive.mount("/content/drive")`
- d) Ambos b e c
- e) Nenhuma das alternativas acima

**Resposta correta: d) Ambos b e c**

**Feedback:** Para acessar o Google Drive no Google Colab, utilizamos `from google.colab import drive` e `drive.mount("/content/drive")` para montar o Google Drive no Colab.

### 3. Qual é a função do comando `dados.head()` no pandas?

- a) Exibe as últimas 5 linhas de um DataFrame
- b) Exibe as primeiras 5 linhas de um DataFrame
- c) Exibe o tamanho do DataFrame
- d) Exibe a soma das colunas do DataFrame
- e) Exibe as colunas de um DataFrame

**Resposta correta: b) Exibe as primeiras 5 linhas de um DataFrame**

**Feedback:** O método `head()` retorna as primeiras 5 linhas de um DataFrame, permitindo uma visão rápida dos dados.

**4. O que a expressão `dados["quality"] == 5` faz?**

- a) Seleciona as linhas em que a coluna "quality" é igual a 5
- b) Exclui as linhas em que a coluna "quality" é igual a 5
- c) Ordena os dados pela coluna "quality" em ordem crescente
- d) Cria uma nova coluna chamada "quality" com valor 5
- e) Converte os valores da coluna "quality" para 5

**Resposta correta: a) Seleciona as linhas em que a coluna "quality" é igual a 5**

**Feedback:** A expressão cria uma máscara booleana, filtrando as linhas do DataFrame onde a coluna "quality" é igual a 5.

**5. O que é o comando `dados.iloc[0, 1]` no pandas?**

- a) Acessa o valor da primeira linha e segunda coluna
- b) Acessa a primeira coluna e a primeira linha
- c) Acessa o valor da última linha e da última coluna
- d) Acessa a segunda linha e a segunda coluna
- e) Exclui a primeira linha e a segunda coluna

**Resposta correta: a) Acessa o valor da primeira linha e segunda coluna**

**Feedback:** O método `iloc[]` permite acessar dados por posição de linha e coluna no DataFrame. A posição começa em 0.

**6. O que faz a máscara `(dados["pH"] >= 3.4) & (dados["free sulfur dioxide"] >= 10)` no pandas?**

- a) Filtra os dados onde a coluna "pH" é maior ou igual a 3.4 e "free sulfur dioxide" é maior ou igual a 10
- b) Substitui todos os valores da coluna "pH" por 3.4
- c) Ordena os dados pela coluna "pH"
- d) Substitui os valores da coluna "free sulfur dioxide" por 10
- e) Cria uma nova coluna com os resultados das condições

**Resposta correta: a) Filtra os dados onde a coluna "pH" é maior ou igual a 3.4 e "free sulfur dioxide" é maior ou igual a 10**

**Feedback:** A expressão cria uma condição de filtro, retornando apenas os dados que atendem a ambas as condições.

**7. O que acontece ao executar `dados["free sulfur dioxide"] >= 20` no pandas?**

- a) Exibe os valores da coluna "free sulfur dioxide" que são maiores ou iguais a 20
- b) Exclui as linhas com valores menores que 20 na coluna "free sulfur dioxide"
- c) Substitui os valores menores que 20 por 20 na coluna "free sulfur dioxide"
- d) Filtra as colunas da DataFrame, excluindo a coluna "free sulfur dioxide"
- e) Não faz nada, é uma expressão inválida

**Resposta correta: a) Exibe os valores da coluna "free sulfur dioxide" que são maiores ou iguais a 20**

**Feedback:** A expressão retorna uma máscara booleana que indica quais valores na coluna "free sulfur dioxide" são maiores ou iguais a 20.

#### **8. O que é necessário para usar o GoogleDriveDownloader no Google Colab?**

- a) Instalar a biblioteca google-colab
- b) Criar uma conta no Google Cloud Platform
- c) Usar o comando `gdd.download_file_from_google_drive()`
- d) Configurar a API do Google Drive no Colab
- e) Não é necessário nada, é automático

**Resposta correta: c) Usar o comando `gdd.download_file_from_google_drive()`**

**Feedback:** O comando `gdd.download_file_from_google_drive()` é utilizado para baixar arquivos do Google Drive, mas requer que a biblioteca `google_drive_downloader` seja instalada previamente.

#### **9. Como você pode visualizar as últimas 5 linhas de um DataFrame no pandas?**

- a) `dados.tail()`
- b) `dados.first()`
- c) `dados.last()`
- d) `dados.tail(0)`
- e) `dados.end()`

**Resposta correta: a) `dados.tail()`**

**Feedback:** O método `tail()` retorna as últimas 5 linhas do DataFrame. Você pode passar um número como argumento para visualizar um número diferente de linhas.

#### **10. O que faz o comando `dados.shape` no pandas?**

- a) Exibe o número de linhas e colunas do DataFrame
- b) Exibe as colunas do DataFrame
- c) Exclui as colunas do DataFrame
- d) Exibe os valores das células no DataFrame
- e) Exibe a estrutura do arquivo CSV original

**Resposta correta: a) Exibe o número de linhas e colunas do DataFrame**

**Feedback:** O comando `shape` retorna uma tupla com o número de linhas e colunas do DataFrame, ajudando a entender o tamanho do conjunto de dados.

**11. Qual das seguintes afirmações é verdadeira em relação ao método `loc` no Pandas?**

- a) **Utilizado para selecionar linhas e colunas por rótulos (textos com os nomes das colunas).**
- b) Utilizado apenas para selecionar linhas com base em índices inteiros.
- c) Utilizado exclusivamente para seleção de colunas por rótulos.
- d) Utilizado para realizar operações matemáticas em um DataFrame.

**Resposta correta: a)**

**Feedback:** O método `loc` é utilizado para acessar um grupo de linhas e colunas por rótulos ou uma matriz booleana. Ele seleciona dados com base nos rótulos (nomes) das colunas ou índices.

**12. Qual é a principal diferença entre `loc` e `iloc`?**

- a) **`loc` é usado para seleção por rótulos (textos com os nomes das colunas), enquanto `iloc` é usado para seleção por índices inteiros.**
- b) `loc` é usado apenas para seleção de colunas, enquanto `iloc` é usado apenas para seleção de linhas.
- c) Ambos são idênticos e podem ser usados de forma intercambiável.
- d) `iloc` é uma versão mais antiga e depreciada de `loc`.

**Resposta correta: a)**

**Feedback:** `loc` seleciona com base em rótulos, enquanto `iloc` usa índices inteiros (posições) para seleção. São métodos diferentes para seleção de dados.

**13. O que o método `info()` faz em um DataFrame?**

- a) Retorna informações sobre as operações recentes realizadas no DataFrame.
- b) Fornece um resumo estatístico das colunas do DataFrame.
- c) **Exibe informações sobre o tipo de dados, valores ausentes e uso de memória no DataFrame.**
- d) Calcula a média das colunas numéricas em um DataFrame.

**Resposta correta: c)**

**Feedback:** O método `info()` fornece uma visão geral do DataFrame, incluindo o tipo de dados de cada coluna, a contagem de valores não nulos e o uso de memória.

**14. Quais tipos de informações são comumente obtidas usando o método `info()`?**

- a) Quantidade de linhas e colunas no DataFrame.
- b) Informações sobre o tipo de dados, valores ausentes e uso de memória.
- c) Nomes das colunas e rótulos das linhas.
- d) **Todas as opções anteriores estão corretas.**

**Resposta correta: d)**

**Feedback:** O método `info()` fornece uma visão geral das colunas, tipos de dados, valores ausentes e o uso de memória, além de exibir a quantidade de linhas e colunas no DataFrame.

**15. Qual método seria mais apropriado para extrair um resumo contendo as informações da quantidade de valores não nulos por coluna, nome das colunas e tipo de dados em cada coluna de um DataFrame?**

- a) `describe()`
- b) `head()`
- c) **`info()`**
- d) `tail()`

**Resposta correta: c)**

**Feedback:** O método `info()` é o mais adequado para exibir essas informações

básicas sobre um DataFrame, como valores não nulos, tipos de dados e a quantidade de colunas e linhas.

**16. Qual comando usado em aula para importar a biblioteca Pandas em Python?**

- a) import pd
- b) **import pandas as pd**
- c) import pandas
- d) from pandas import pd

**Resposta correta: b)**

**Feedback:** A importação correta do Pandas em Python é import pandas as pd, permitindo o uso da convenção pd para facilitar o código.

**17. Qual comando é usado para ler um arquivo CSV em um DataFrame Pandas?**

- a) read\_data()
- b) **read\_csv()**
- c) load\_csv()
- d) import\_csv()

**Resposta correta: b)**

**Feedback:** O método read\_csv() é o comando usado para carregar um arquivo CSV em um DataFrame no Pandas.

**18. Qual comando é usado para selecionar uma única coluna em um DataFrame Pandas?**

- a) select\_column()
- b) get\_column()
- c) select()
- d) **[]**

**Resposta correta: d)**

**Feedback:** Para selecionar uma única coluna em um DataFrame Pandas, usamos o comando `[]` com o nome da coluna como chave, por exemplo: `df['coluna']`.

**19. Qual método é usado para agrupar dados em um DataFrame Pandas?**

- a) `group_by()`
- b) `group()`
- c) **`groupby()`**
- d) `aggregate()`

**Resposta correta: c)**

**Feedback:** O método `groupby()` é usado para agrupar dados em um DataFrame com base em uma ou mais colunas.

**20. Qual método é usado para ordenar um DataFrame Pandas com base em uma ou mais colunas?**

- a) `sort_by()`
- b) `order()`
- c) **`sort_values()`**
- d) `arrange()`

**Resposta correta: c)**

**Feedback:** O método `sort_values()` é usado para ordenar um DataFrame com base em uma ou mais colunas.

**21. Qual comando é usado para calcular a média de uma coluna em um DataFrame Pandas?**

- a) **`mean()`**
- b) `average()`
- c) `median()`
- d) `calc_mean()`

**Resposta correta: a)**

**Feedback:** O método `mean()` é usado para calcular a média de uma coluna numérica em um `DataFrame`.

**22. Qual comando é usado para criar um gráfico de caixa (boxplot) em Seaborn?**

- a) `scatterplot()`
- b) **`boxplot()`**
- c) `box_plot()`
- d) `scatterplot()`

**Resposta correta: b)**

**Feedback:** O método `boxplot()` em Seaborn é utilizado para criar gráficos de caixa (boxplots), que são úteis para mostrar a distribuição dos dados e os valores discrepantes.

**23. Qual comando é usado para criar um gráfico de histograma em Seaborn?**

- a) `box_plot()`
- b) `histogram()`
- c) **`histplot()`**
- d) `scatterplot()`

**Resposta correta: c)**

**Feedback:** O comando `histplot()` em Seaborn é utilizado para criar gráficos de histograma, que representam a distribuição de frequências dos dados.

**24. Qual parâmetro do método permite definir o número de barras do histograma?**

- a) **`bins`**
- b) `hue`
- c) `palette`
- d) `bar_color`



**Resposta correta: a)**

**Feedback:** O parâmetro bins define o número de barras (intervalos) em um histograma.

**25. Qual parâmetro do método scatterplot permite colorir os pontos com base em uma variável categórica?**

- a) **hue**
- b) color
- c) palette
- d) marker

**Resposta correta: a)**

**Feedback:** O parâmetro hue permite que os pontos em um gráfico de dispersão sejam coloridos com base em uma variável categórica.

**26. Qual parâmetro do método scatterplot é usado para especificar o estilo dos pontos?**

- a) **style**
- b) point\_style
- c) marker\_style
- d) markers

**Resposta correta: a)**

**Feedback:** O parâmetro style é usado para especificar o estilo dos pontos no gráfico de dispersão, como diferentes formas ou marcas.

**27. Qual é a principal função de um gráfico de caixa (boxplot)?**

- a) **Mostrar a distribuição dos dados de forma resumida, incluindo medidas de tendência central e dispersão, incluindo valores discrepantes (outliers).**
- b) Destacar exclusivamente os valores discrepantes (outliers) nos dados.
- c) Visualizar a relação entre duas variáveis.

- d) Representar a frequência de ocorrência de diferentes categorias de dados.

**Resposta correta: a)**

**Feedback:** O gráfico de caixa (boxplot) resume a distribuição dos dados, destacando a mediana, os quartis e os valores discrepantes (outliers).

**28. Qual é a principal função de um histograma?**

- a) Visualizar a relação entre duas variáveis.
- b) Identificar valores discrepantes (outliers) nos dados.
- c) **Representar a distribuição de frequência de uma variável quantitativa.**
- d) Mostrar a média e o desvio padrão dos dados.

**Resposta correta: c)**

**Feedback:** O histograma é utilizado para representar a distribuição de frequência de uma variável quantitativa, mostrando como os dados estão distribuídos em intervalos.

**29. Qual é a principal função de um gráfico de dispersão?**

- a) Mostrar a distribuição de frequência de uma variável quantitativa.
- b) Identificar valores discrepantes (outliers) nos dados.
- c) **Visualizar a relação entre duas variáveis quantitativas.**
- d) Representar a distribuição de frequência de uma variável categórica.

**Resposta correta: c)**

**Feedback:** O gráfico de dispersão é utilizado para visualizar a relação entre duas variáveis quantitativas, mostrando como elas se correlacionam.

**30. Qual das opções a seguir descreve uma distribuição normal?**

- A) Uma distribuição assimétrica em que a concentração de dados está mais à esquerda.
- B) Uma distribuição simétrica com dados concentrados ao redor da média.
- C) Uma distribuição com picos pronunciados nas extremidades.
- D) Uma distribuição em que todos os dados são iguais.
- E) Uma distribuição onde os dados são distribuídos uniformemente.

**Resposta correta: B**

**Feedback:** A distribuição normal é simétrica e possui a maior concentração de dados ao redor da média. Ela é frequentemente representada por um gráfico em forma de sino.

**31. O que é a distribuição normal padrão?**

- A) Uma distribuição com média igual a 0 e desvio padrão igual a 1.
- B) Uma distribuição com média igual a 1 e desvio padrão igual a 0.
- C) Uma distribuição com valores negativos de média.
- D) Uma distribuição com desvio padrão constante.
- E) Uma distribuição com média e desvio padrão variáveis.

**Resposta correta: A**

**Feedback:** A distribuição normal padrão é uma distribuição normal que foi transformada para ter uma média igual a 0 e um desvio padrão igual a 1. Ela é usada como referência para calcular probabilidades de outras distribuições.

**32. Qual é o modelo da distribuição normal em termos de suas variáveis?**

- A) Média  $\mu = 0$ , desvio padrão  $\sigma = 2$ .
- B) Média  $\mu = 1$ , desvio padrão  $\sigma = 1$ .
- C) Média  $\mu = 0$ , desvio padrão  $\sigma = 1$ .
- D) Média  $\mu = 0$ , desvio padrão  $\sigma = 3$ .
- E) Média  $\mu = 2$ , desvio padrão  $\sigma = 0$ .

**Resposta correta: C**

**Feedback:** A distribuição normal padrão tem média igual a 0 e desvio padrão igual a 1. Isso facilita o uso de tabelas para calcular probabilidades.

**33. O que significa o valor de Z em uma distribuição normal padrão?**

- A) A probabilidade de um valor ser maior que a média.
- B) A transformação de um valor para uma distribuição normal padrão com média 0 e desvio padrão 1.
- C) A medição de dispersão de uma distribuição.
- D) O desvio da média de um valor em unidades de desvio padrão.
- E) A medição de frequência dos dados em uma amostra.

**Resposta correta: B**

**Feedback:** O valor de Z é a transformação de um valor da distribuição original para a distribuição normal padrão, com média 0 e desvio padrão 1.

**34. Qual das afirmativas sobre o gráfico de uma distribuição normal é verdadeira?**

- A) Ele tem uma forma em "V" com dados concentrados nas extremidades.
- B) O gráfico é simétrico, com dados distribuídos de forma mais concentrada no centro.
- C) O gráfico é assimétrico, com dados concentrados à direita.
- D) O gráfico não possui pico; os dados são distribuídos uniformemente.
- E) O gráfico é assimétrico, com dados concentrados à esquerda.

**Resposta correta: B**

**Feedback:** A distribuição normal é simétrica em torno da média, com os dados mais concentrados no centro e se espalhando de forma gradualmente para as extremidades.

**35. Como pode-se verificar se um conjunto de dados segue uma distribuição normal?**

- A) Calculando a média e verificando a simetria dos valores.
- B) Observando o histograma e o gráfico de barras.
- C) Usando um gráfico qqplot para verificar a aderência à linha reta.
- D) Verificando se os dados têm variância nula.
- E) Calculando a mediana e comparando com a média.

**Resposta correta: C**

**Feedback:** O gráfico qqplot (quantile-quantile plot) é uma forma eficiente de verificar se os dados seguem uma distribuição normal. Ele mostra se os pontos se ajustam a uma linha reta, o que indica normalidade.

**36. O que representa a área sob a curva de uma distribuição normal?**

- A) A variância dos dados.
- B) O número total de dados no conjunto.
- C) A probabilidade total de ocorrência da variável de estudo, que é igual a 1.
- D) O desvio padrão dos dados.
- E) A mediana dos dados.

**Resposta correta: C**

**Feedback:** A área sob a curva de uma distribuição normal representa a probabilidade total de todos os resultados possíveis. Em uma distribuição normal, essa área é igual a 1, ou 100%.

**37. Qual é o método utilizado para transformar uma distribuição normal em uma distribuição normal padrão?**

- A) Multiplicando os valores pelos desvios padrão.
- B) Subtraindo a média e dividindo pelo desvio padrão.
- C) Adicionando a média ao desvio padrão.
- D) Calculando a variância dos dados.
- E) Subtraindo os valores das extremidades.

**Resposta correta: B**

**Feedback:** Para transformar uma distribuição normal em uma normal padrão, subtrai-se a média de cada valor e divide-se pelo desvio padrão. Esse processo gera uma distribuição com média 0 e desvio padrão 1.

**38. Em uma distribuição normal, qual é a probabilidade de um valor estar entre a média e um desvio padrão acima da média?**

- A) 34%.
- B) 50%.
- C) 68%.
- D) 95%.
- E) 99%.

**Resposta correta: C**

**Feedback:** Em uma distribuição normal, aproximadamente 68% dos dados estão dentro de um desvio padrão da média. Isso é conhecido como a regra empírica ou regra 68-95-99.

**39. O que a tabela Z fornece ao trabalhar com distribuições normais?**

- A) O valor da média para a distribuição.
- B) A probabilidade acumulada até determinado valor Z.
- C) O valor do desvio padrão para qualquer distribuição.
- D) A soma de todos os valores de Z.
- E) A variância de uma distribuição normal.

**Resposta correta: B**

**Feedback:** A tabela Z fornece a probabilidade acumulada até um determinado valor de Z em uma distribuição normal padrão. Isso ajuda a calcular a probabilidade de ocorrências dentro de certos intervalos.

40. Qual das alternativas abaixo melhor define a média de um conjunto de dados?

- A) A média é o valor que ocorre com maior frequência em um conjunto de dados.
- B) A média é a diferença entre o maior e o menor valor em um conjunto de dados.
- C) A média é a soma de todos os valores dividida pelo número de observações.
- D) A média é o valor que divide um conjunto de dados ordenado em duas metades.
- E) A média é o valor central de um conjunto de dados.

**Feedback:**

- A)** Errado. A média não se refere à frequência de ocorrência dos valores.
- B)** Errado. Isso descreve o conceito de **amplitude** (range).
- C)** Correto! A média é calculada somando todos os valores e dividindo pelo número de observações.
- D)** Errado. Isso descreve a **mediana**.
- E)** Errado. A definição é para a **mediana**, não a média.

41. A mediana de um conjunto de dados é:

- A) O valor que ocorre com maior frequência no conjunto de dados.
- B) O valor central quando os dados são organizados em ordem crescente ou decrescente.
- C) A média de todos os dados.
- D) A diferença entre o maior e o menor valor dos dados.
- E) A medida de dispersão dos dados em relação à média.

**Feedback:**

- A)** Errado. A descrição corresponde à **moda**, não à mediana.
- B)** Correto! A mediana é o valor que divide o conjunto de dados ao meio quando ordenado.
- C)** Errado. Isso descreve a **média**, não a mediana.
- D)** Errado. Isso é a definição de **amplitude**.
- E)** Errado. Isso descreve o **desvio padrão** ou a **variância**, não a mediana.

42. O que a moda de um conjunto de dados indica?

- A) O valor que ocorre com maior frequência.
- B) A média ponderada dos dados.
- C) O valor que divide o conjunto de dados em duas partes iguais.
- D) A soma dos quadrados dos desvios em relação à média.
- E) O valor central de um conjunto de dados ordenados.

**Feedback:**

- A)** Correto! A moda é o valor que aparece com maior frequência em um conjunto de dados.
- B)** Errado. A média ponderada não é a moda.
- C)** Errado. A descrição corresponde à **mediana**.
- D)** Errado. Isso descreve o cálculo da **variância**.
- E)** Errado. A definição corresponde à **mediana**, não à moda.

**43. Qual é a fórmula da variância em um conjunto de dados amostral?**

- A) A soma dos quadrados dos desvios em relação à média dividida pelo número de dados.
- B) A soma dos desvios em relação à média dividida pelo número de dados.
- C) A soma dos valores dos dados dividida pelo número de dados.
- D) A soma dos quadrados dos desvios em relação à média dividida pelo número de dados menos um.
- E) A média dos desvios em relação à média dos dados.

**Feedback:**

- A)** Errado. Esta descrição corresponde ao cálculo da variância populacional.
- B)** Errado. Essa descrição não é a fórmula correta.
- C)** Errado. Isso é a fórmula para a média, não para a variância.
- D)** Correto! A variância amostral é a soma dos quadrados dos desvios em relação à média dividida pelo número de dados menos um.
- E)** Errado. Isso descreve outra operação estatística, mas não é a variância.

**44. O desvio padrão de um conjunto de dados representa:**

- A) A média dos desvios em relação à média.
- B) A soma dos quadrados dos desvios em relação à média.
- C) A dispersão dos dados em torno da média, em unidades originais.
- D) A diferença entre o maior e o menor valor.
- E) O valor central do conjunto de dados.

**Feedback:**

- **A)** Errado. O desvio padrão é a raiz quadrada da variância, e não a média dos desvios.
- **B)** Errado. Isso descreve a variância.
- **C)** Correto! O desvio padrão mede a dispersão dos dados, mas na mesma unidade que os dados originais.
- **D)** Errado. Isso descreve a **amplitude**.
- **E)** Errado. O valor central é a **mediana**.

45. Qual é a principal finalidade do processo de limpeza de dados em Ciência de Dados?

A) Melhorar a visualização dos dados.

**B) Eliminar dados irrelevantes e corrigir inconsistências para garantir que os dados sejam de alta qualidade.**

C) Criar modelos preditivos a partir dos dados.

D) Organizar os dados em um formato específico de exibição.

E) Aumentar a quantidade de dados disponíveis para análise.

**Feedback:**

- A) Errado. A limpeza de dados não está diretamente relacionada à melhoria da visualização, mas sim à qualidade dos dados.
- **B) Correto! A limpeza de dados envolve a correção de inconsistências, a remoção de valores faltantes e outros ajustes para garantir que os dados sejam de alta qualidade.**
- C) Errado. A criação de modelos preditivos faz parte da modelagem de dados, não da limpeza.
- D) Errado. A organização para exibição é parte da análise e não da limpeza de dados.
- E) Errado. Limpeza de dados não é focada em aumentar a quantidade de dados, mas sim em melhorar a qualidade dos dados já existentes.

46. O que são outliers em um conjunto de dados e qual seu impacto na análise?

A) Outliers são os valores mais comuns em um conjunto de dados e não afetam a análise.

B) Outliers são valores que aparecem com mais frequência que outros e são usados para melhorar a precisão do modelo.

**C) Outliers são valores que estão significativamente distantes da maioria dos**



**dados e podem distorcer os resultados da análise.**

D) Outliers são dados que não são relevantes para o modelo e devem ser sempre removidos antes da análise.

E) Outliers são valores que estão sempre dentro da faixa esperada e não têm impacto na análise.

**Feedback:**

- **A) Errado.** Outliers não são valores comuns e podem impactar a análise negativamente.
- **B) Errado.** Embora possam ser raros, outliers não são usados para melhorar a precisão do modelo e, frequentemente, são problemáticos.
- **C) Correto!** Outliers são valores que estão significativamente distantes dos dados principais e podem distorcer as conclusões da análise.
- **D) Errado.** Embora outliers possam ser removidos em alguns casos, não é uma regra geral, pois em certos contextos eles podem ser valiosos para a análise.
- **E) Errado.** Outliers são, por definição, valores fora da faixa esperada e têm impacto na análise.

47. O que é classificação em aprendizado de máquina?

**A) Um processo que agrupa dados em categorias.**

B) Um processo que prevê números contínuos.

C) Um processo que encontra padrões em dados sem rótulos.

D) Um processo que remove dados irrelevantes.

E) Um processo que organiza os dados em ordem crescente.

**Feedback:**

- **A) Correto!** A classificação é usada para atribuir dados a categorias ou classes distintas.
- **B) Errado.** Prever números contínuos é o objetivo da **regressão**, não da classificação.
- **C) Errado.** A classificação é uma técnica supervisionada, ou seja, requer dados rotulados.
- **D) Errado.** A remoção de dados irrelevantes é parte da **limpeza de dados**, não da classificação.
- **E) Errado.** Organizar dados em ordem crescente é uma tarefa simples de ordenação, não de classificação.

48. Qual comando é utilizado para exibir algo na tela em Python?

- A) print()
- B) display()
- C) show()
- D) echo()
- E) output()

**Feedback:**

- **A)** Correto! O comando **print()** é utilizado para exibir informações na tela.
- **B)** Errado. **display()** não é um comando padrão em Python.
- **C)** Errado. **show()** não é um comando padrão em Python.
- **D)** Errado. **echo()** é usado em outras linguagens, como PHP, mas não em Python.
- **E)** Errado. **output()** não é um comando válido em Python.

49. Como você cria uma lista em Python?

- A) lista = {}
- B) lista = []
- C) lista = ()
- D) lista = <>
- E) lista = //

**Feedback:**

- **A)** Errado. As chaves **{}** são usadas para criar um **dicionário** em Python, não uma lista.
- **B)** Correto! As listas são criadas usando colchetes **[]**.
- **C)** Errado. Os parênteses **()** são usados para criar **tuplas**, não listas.
- **D)** Errado. **<>** não é usado para criar listas em Python.
- **E)** Errado. **//** é um operador de divisão inteira, não para criar listas.

50. Qual é o resultado da expressão  $2 + 3 * 4$  em Python?

- A) 20
- B) 14
- C) 12

D) 17

E) 5

**Feedback:**

- **A)** Errado. O cálculo correto não resulta em 20.
- **B)** Correto! Em Python, a multiplicação tem precedência sobre a soma, então o cálculo é  $2 + (3 * 4)$ , resultando em 14.
- **C)** Errado. A expressão não resulta em 12.
- **D)** Errado. 17 não é o valor correto da expressão.
- **E)** Errado. O valor da expressão não é 5.