

Questionário

1. O principal objetivo do Exploratory Data Analysis é garantir que:

- a) Os dados não têm missing values
- b) Que o modelo Deep Learning funciona bem
- c) Que os dados têm o conteúdo e o formato adequado para a responder à pergunta em questão
- d) Existem correlações entre variáveis

2. Qual a diferença entre Dados e Informação?

- a) Informação são dados organizados de uma forma que permite fazer medições, análises e visualizações
- b) Não existe diferença, são apenas 2 nomes para o mesmo tipo de conteúdo
- c) Dados são conteúdos organizados de uma forma que permite fazer medições, análises e visualizações e informação são as ferramentas suporte à sabedoria
- d) Dados são conteúdos de informação organizados para modelação

3. Antes de considerar que os dados a que temos acesso são apropriados para responder à nossa pergunta devemos verificar se (marcar todas as validas):

- a) Os dados são relevantes
- b) Os dados estão organizados em tabelas indexadas
- c) Existem duplicados
- d) Há inconsistências

4. Que tipo de variáveis existem:

- a) Outliers e Missing Values
- b) Quantitativas (ordinais e continuas) e Qualitativas (nominais e discretas)
- c) Numéricas (ordinais e continuas) e Categóricas (nominais e discretas)
- d) Quantitativas (discretas e continuas) e Qualitativas (nominais e ordinais)

5. Qual o principal objetivo de uma tabela de frequências?

- a) Permite identificar problemas nos dados, como outliers
- b) Compreender a distribuição de uma variável por classe ou intervalo
- c) Identificar se os dados são adequados para o problema que queremos resolver
- d) Transformar os dados em sabedoria

6. Para organizar dados quantitativos em tabelas de frequência é necessário:

- a) Retirar as casas decimais
- b) Preencher missing values
- c) Criar intervalos de valores
- d) Retirar linhas repetidas

7. Os gráficos Pie Chart são os mais indicados para que tipo de variáveis?

- a) Quantitativas Contínuas
- b) Qualitativas Nominais
- c) Qualitativas Ordinais
- d) Quantitativas Ordinais

8. Para análise a relação entre 2 variáveis que tipo de gráfico é o mais adequado?

- a) Pie Chart
- b) Gráfico de Barras
- c) Scatter Plot
- d) Todos os anteriores podem ser usados

9. Imagina que recebes um novo conjunto de dados, quais os teus 3 primeiros passos e porquê?

1. Alinhamento com o problema de negócio: Analisar os dados para perceber se estão alinhados com o objetivo do estudo/análise/modelo, ou seja, se têm dimensão suficiente, se têm detalhe adequado, histórico suficiente, etc.
2. Formatos: Validar, e se necessário corrigir, formatos, escalas, tal como proceder à normalização de conteúdos.
3. Conteúdos: Explorar as variáveis por forma a conhecer melhor o seu comportamento e verificar se é necessário fazer algum tratamento ao mesmo, como preenchimento de *missing values*, *outliers*, etc.

Pontuação: Da pergunta 1 à pergunta 8 =10%, Pergunta 9 = 20%