

TESTE ESTÁGIO DATA SCIENCE / BUSINESS INTELLIGENCE

Temos preferência que a questão 1 seja resolvida em Python ou R. Porém, o candidato pode resolver também com a linguagem que se sentir mais confiante. Só não será permitido a solução destas questões em Excel ou ferramentas similares. Se houver qualquer indício de plágio, o candidato será desclassificado do processo.

Caso o candidato opte por uma linguagem diferente de Python ou R, é muito importante que o candidato gere um arquivo com instruções para executar os códigos (uma espécie de “*READ ME*”). Além disso, é interessante que o código (para questão 1) tenha comentários para facilitar a compreensão do raciocínio utilizado.

Para questão 2, a solução pode ser em qualquer linguagem baseada em SQL.

Já a questão 3 exige capacidade analítica e um certo “faro investigativo”. O candidato deverá apresentar um texto conciso que justifique de forma clara e concludente o comportamento identificado nas figuras localizadas no final deste teste.

Boa sorte!

Questão 1

- a) Ler o arquivo **house_infos.csv** na forma de um *dataframe*.
- b) Neste *dataframe*, qual a quantidade de casas que possuem “**LotArea**” > 5000 e “**CentralAir**” = N?
- c) Filtrar *dataframe* para conter apenas as linhas que tenham “**GrLivArea**” > 2000.
- d) A partir do *dataframe* filtrado no item anterior, crie uma coluna “**SalePrice**”, os valores desta coluna deverão ser o resultado da seguinte expressão: **(LotFrontage + log(LotArea) + TotalBsmtSF + GrLivArea) * OverallQual**.
- e) Crie a coluna “**AvgOverallQual**”, a qual deve conter a média aritmética de **OverallQual** para cada **YearBuilt**.
- f) Tratar os casos em que há NaN nos resultados do cálculo dos dados da coluna “**SalePrice**”, preenchendo-os com 0 (zero).
- g) Ler o arquivo **year_condition.csv** na forma de um *dataframe*.

- h) Fazer um *merge* entre o *dataframe* final do arquivo **house_infos.csv** e do arquivo **year_condition.csv** com base na coluna **YearBuilt**. Desta forma, o *dataframe* resultante deverá conter uma nova coluna **YearCondition**.
- i) Agrupar por **YearBuilt**, **AvgOverallQual** e **YearCondition** e obter a média. Ordenar o *dataframe* resultante por **YearBuilt** de forma crescente.
- j) Os valores da coluna **SalePrice** devem estar com apenas duas casas decimais.
- k) Com o *dataframe* resultante, faça um gráfico de barras para **SalePrice** x **YearBuilt**.
- l) Utilizando um eixo y secundário no mesmo gráfico, acrescente um gráfico de linha para **AvgOverallQual** x **YearBuilt**.
- m) Salvar o gráfico em formato png com o nome **house_analysis.png**.
- n) Exportar o *dataframe* final para um arquivo pickle com nome **processed_data.p**.

Questão 2

Imagine uma base de dados de uma universidade que contenha as seguintes tabelas e seus respectivos campos organizados da seguinte forma:

Alunos (RA, nome, endereço, cidade)

Disciplinas (COD_DISC, nome_disc, carga_hor)

Professores (COD_PROF, nome, endereço, cidade)

Turma (COD_DISC, COD_TURMA, COD_PROF, ANO, horário)

Histórico (RA, COD_DISC, COD_TURMA, COD_PROF, ANO, frequencia, nota)

- a) Monte uma consulta que informe quantos alunos cursaram a disciplina de Banco de Dados em 2009 e 2010.
- b) Monte uma consulta que exiba os nomes dos alunos que não foram reprovados em nenhuma disciplina (considere nota mínima igual a 6 para aprovação).

Questão 3

Você já ouviu falar de funil de vendas? O funil de vendas é um modelo estratégico de consumo que ilustra o percurso teórico que o cliente deve seguir até efetuar a compra de um produto ou serviço (Figura 1). Este percurso é denominado como jornada de

compra. Assim como no funil tradicional, o funil de vendas começa largo e vai diminuindo até chegar aos reais consumidores. As etapas do funil de vendas estão diretamente ligadas às fases da jornada de compra.

E de *checkout*, você já ouviu falar? No mundo das lojas virtuais, especificamente, significa encerrar o processo de compra. O usuário seleciona um produto que deseja comprar na sua loja (ou vários), coloca no carrinho e então se dirige para fechar a compra. Esse é o momento do checkout. De forma resumida, é todo o procedimento que vai desde o cadastro do usuário (se ele não for cliente), passando pela seleção do endereço de entrega a cálculo de frete até o pagamento.

Tendo em mente os conceitos de funil de vendas e de checkout, analise as duas figuras a seguir e explique detalhadamente quais seriam as razões pelas quais o site do Guichê Virtual apresentou, em um dia qualquer, uma queda acentuada em sua taxa de conversão (Figura 2). Extraia o máximo de informações possíveis das **DUAS** figuras apresentadas.

Obs: conversão é um termo que se refere à venda.

Funil de vendas 17:00 - 18:00 (Situação Hipotética)

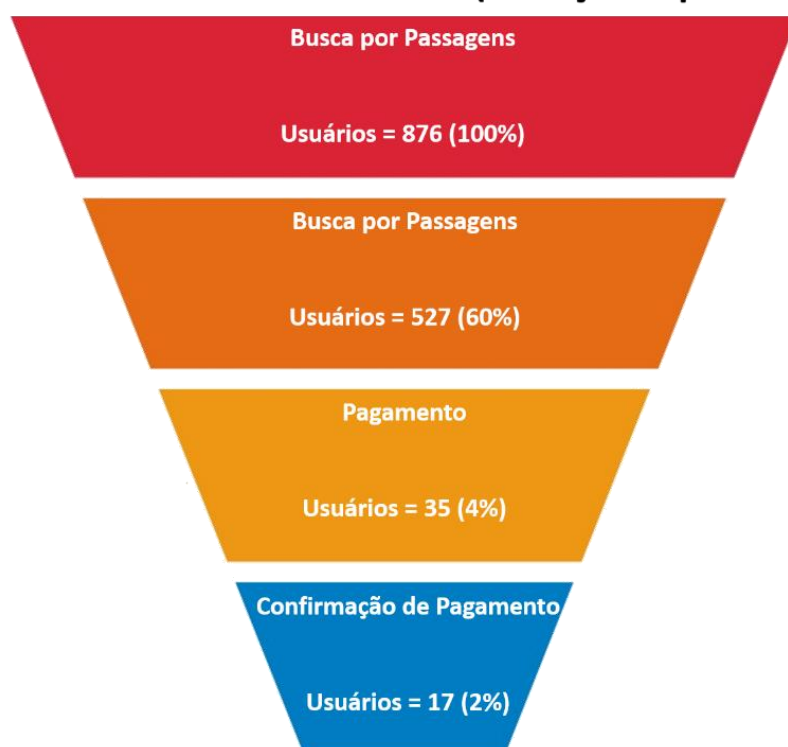


Figura 1 - Funil de vendas

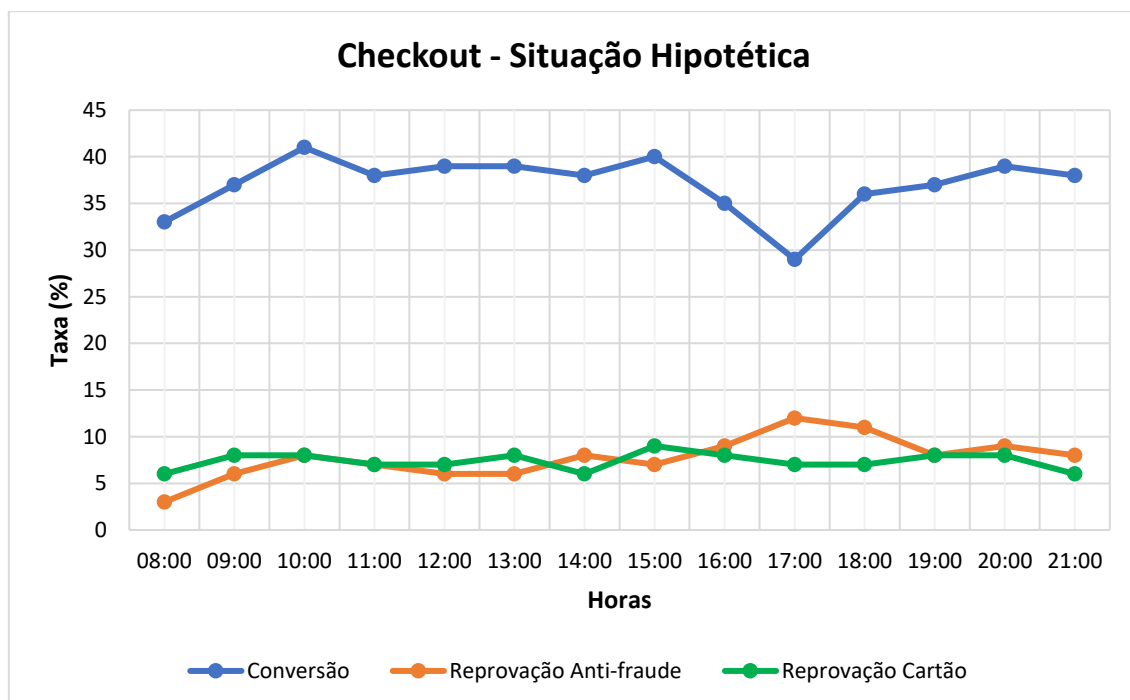


Figura 2 - Checkout por hora em um dia qualquer