Nome: Leonardo Padilha de Lima

**Questão 1 (10 Pontos)**

Liste o maior número possível de casos de uso para o conjunto de dados.

1. ***Sistema para Previsão de vendas por Vendedor, Canal, Produto, Período, Grupo de clientes e Segmento de clientes.***
2. ***Sistema para Clusterização***
3. ***Sistema para Estimativa de Churn (Perda de clientes)***
4. ***Ranqueamento de vendas por Vendedor, Canal, Produtos, Períodos, Grupos de Clientes e Segmento de clientes.***
5. ***Sistema de recomendação de produtos para os clientes de acordo com seus padrões de consumo.***

**Questão 2 (10 Pontos)**

Escolha um dos casos de uso listados na pergunta 1 e descreva como a criação de um modelo estatístico com base no conjunto de dados poderia ser usado para melhorar os negócios dos quais esses dados vêm.

***R: Sistema para Estimativa de Churn (Perda de clientes)***

**Questão 3 (20 Pontos)**

Implemente o modelo que você descreveu na pergunta 2, de preferência em Python. O código deve recuperar os dados, treinar e testar um modelo estatístico e relatar critérios de desempenho relevantes. Idealmente, poderemos replicar sua análise a partir do código-fonte enviado; portanto, explique as versões das ferramentas e pacotes que você está usando.

***R: A respostas para esta questão no jupyter do github.***

**Questão 4 (60 Pontos)**

1. Explique todas as opções do projeto e use os notebooks jupyter. (por exemplo, pré-processamento, seleção de modelo, hiper parâmetros, critérios de avaliação). Compare e contraste suas escolhas com metodologias alternativas
2. Descreva como você melhoraria o modelo na Questão 3 se tivesse mais tempo

***R: As respostas para os itens da questão 4 estão no jupyter do github. Fiz todo o comentário passo a passo.***