



Desafio Data science

Leroy Merlin

Descrição do desafio



As compras online fazem, mais do que nunca, parte do dia a dia das pessoas. Mais do que nunca, usar os dados para tomadas de decisões é a grande aposta das empresas no mundo.

Entender o perfil da carteira de clientes para gerar campanhas personalizadas é imperativo em um mundo competitivo como o de hoje.

Além disso, um bom planejamento das vendas é crucial para o setor de supply chain conseguir alimentar a rede e evitar problemas como *sold out* e *overstocking*.

Vamos nesse desafio trabalhar com uma base de compras via e-commerce que nos permite construir visões sobre clientes e projeções sobre vendas futuras.

Tema: E-commerce mundial

Histórico: 1-jan-2011 até 31-dez-2014

Colunas:

- Order ID
- Order Date
- Ship Date
- Ship Mode
- customer ID
- Customer Name
- Segment
- City
- State
- Country
- Postal Code
- Market
- Region
- Product ID
- Category
- Sub-Category
- Product Name
- Sales
- Quantity
- Discount
- Profit
- Shipping Cost
- Order Priority

Tamanho: 7.2 mb

Formato: xlsx

1. Data Preparation / Data investigation

- Data cleansing: limpeza da base de dados, outliers, missings (explicitar técnica)
- Levantar grandes números que descreva a base de customers e segmentos do e-commerce
- Pedidos específicos:
 - Qual é o tempo médio de entrega em todos os países?OK
 - Qual segmento de clientes é mais rentável em cada ano ?
 - Qual é o segmento com maior volume de vendas? (entender volume por número de vendas)OK
 - Como os clientes são distribuídos pelos países?
 - Crie um Perfil dos clientes com base em sua frequência de compra - calcular frequência de compra para cada cliente
 - Os clientes mais frequentes estão contribuindo com mais receita?
 - Eles também são rentáveis?

2. Customer segmentation

- Construção de segmentação de customers com foco em pirâmide de valores: Recência*, Frequência e Valor

3. Projeção de vendas por segmento

- Treinar um modelo de time series forecast para prever as vendas (coluna Sales) para o segmento com maior volume de vendas (em todo o histórico) - frequência diária
- Validação do modelo: mês de Dez 2014 (frequência diária)
- Usar métricas MAPE, SMAPE e RMSLE para calcular o resultado em Dez 2014.

Recência = tempo desde última compra do customer*

- **Notebook Jupyter:** comentado com os itens 1 a 3 pedidos no slide anterior e respostas do item 1. Explicitar técnicas de preparação de dados para o item 3.
- **Item 3:** Base CSV com o resultado do modelo para Dez de 2014 e com as colunas:
Data | Vendas reais | Forecast vendas | MAPE | SMAPE | RMSLE