# 1.4 Terceiro desafio: Knowledge Discovery in Databases (KDD)

1. Em relação aos dados disponibilizados, existem dados missing? Descreva o que foi encontrado. Em situações como essa, o que é necessário ser feito?

Texto

Descrição gerada automaticamente

Para esse dataset em questão não temos missing values, conforme mostra o resultado do nosso código, porém quando eles existem, deve-se analisar o contexto e utilizar uma das seguintes soluções:

Exclusão de Linhas ou Colunas: Se a quantidade de valores faltantes for pequena em relação ao tamanho do conjunto de dados, pode-se optar por remover as linhas (dropna() no Pandas) ou colunas que contenham valores faltantes. Isso pode ser aceitável se as informações perdidas não forem significativas para a análise.

Ex: df.dropna(axis=0, inplace=True) # Remove linhas com valores nulos

Preenchimento de Valores: Pode-se preencher os valores faltantes com estatísticas resumidas, como a média, mediana ou moda dos valores existentes na mesma coluna. Isso é útil quando os dados faltantes são poucos e podem ser estimados de maneira razoável.

Ex: df['coluna'].fillna(df['coluna'].mean(), inplace=True) # Preenche com a média da coluna

Interpolação: Para dados sequenciais ou temporais, a interpolação pode ser usada para preencher valores ausentes com base nos valores existentes.

Ex: df['coluna'].interpolate(method='linear', inplace=True) # Preenchimento por interpolação linear

Modelagem Preditiva: Em alguns casos, é possível usar modelos preditivos para prever valores faltantes com base em outras variáveis do conjunto de dados.

1. Analise os dados na perspectiva da coluna idade. Existem Outliers nos dados disponibilizados? É possível identificar algo em relação ao ticket médio de vendas relacionadas a esses Outliers? Justifique sua resposta.

Texto

Descrição gerada automaticamente

Identificamos que o Ticket Médio desses outliers de idade, são todos negativos ou zerados, sendo assim podemos removê-los de nossa análise de ticket médio visto que podem se tratar de um erro de sistema, uma vez que para toda compra não existe ticket médio negativo e muito menos idades como podemos ver nesses outliers.

1. Em relação à consistência do dado valor ticket médio, o que é possível refletir sobre seus conteúdos? Existem dados inconsistentes? Justifique como é possível corrigi-los e realize essa importante atividade, deixando esses dados prontos para análise.

Como observamos no tópico anterior, existem valores negativos no Ticket médio, que vem justamente dos clientes com as idades como 200, 900 etc. Sendo assim para limparmos esses casos, selecionamos os tickets médios somente quando forem >= 0, assim conseguimos ter uma análise real da base, como por exemplo nas duas análises abaixo, de média e de faixas do ticket médio.

Texto

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

1. A área comercial da Melhores Compras criou um conjunto de faixa etária para tentar compreender melhor o perfil do cliente, mas não conseguiu até o momento chegar a lugar algum.

Texto

Descrição gerada automaticamente

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Realizando a limpeza e iniciando as análises solicitadas:

1. Remoção de Outliers e descrição da coluna Ticket Médio

Texto

Descrição gerada automaticamente

1. Ticket Médio por Faixa Etária

Texto

Descrição gerada automaticamente

1. Idade média dos Clientes

Texto

Descrição gerada automaticamente

1. Variância da Idade

Texto

Descrição gerada automaticamente

1. Desvio Padrão da Idade

Texto

Descrição gerada automaticamente

1. Comparação Média x Mediana na Análise de Ticket Médio por Faixa etária

Texto

Descrição gerada automaticamente

1. Ranking de Vendas por Faixa Etária

Texto

Descrição gerada automaticamente

Depois de realizar o tratamento de dados, chegamos a alguns resultados interessantes. São eles:

### Análise Detalhada das Faixas Etárias

1. **Segmentação e Perfil de Consumo:**
   * Utilizar ferramentas de análise de dados avançadas para segmentar o público-alvo, especialmente a faixa etária de 45-54 anos, e identificar padrões de consumo. Por exemplo, análise de dados transacionais para entender quais categorias de produtos são mais populares nesse grupo e quais têm maior potencial de crescimento. É importante destacar que em quantidade esse grupo não é o principal, mas sim em valor consumido, o que é mais importante para nossa estratégia, por exemplo no grupo de 25 a 34 anos temos uma maior quantidade, porém com um valor consumido consideravelmente menor.
   * Exemplo: Com base em dados históricos, descobrimos que clientes na faixa etária de 45-54 anos tendem a preferir produtos de alta qualidade e exclusivos, priorizando segurança e praticidade. Isso se reflete em vendas consistentemente altas em categorias como produtos de tecnologia e saúde.
2. **Identificação de Necessidades e Preferências:**
   * Realizar pesquisas de mercado qualitativas para entender as necessidades específicas dessa faixa etária. Isso pode incluir grupos focais, entrevistas aprofundadas e análise de feedback nas redes sociais para capturar insights sobre o que realmente importa para esses consumidores.
   * Exemplo: Descobrimos que a segurança na transação online e a conveniência na entrega são aspectos críticos para consumidores de 45-54 anos, influenciando diretamente suas decisões de compra.

### Estratégia de Marketing e Comercialização

1. **Desenvolvimento de Campanhas Personalizadas:**
   * Criar campanhas de marketing direcionadas com mensagens que ressoem com a faixa etária de 45-54 anos. Isso pode incluir storytelling que destaca a qualidade, exclusividade e segurança dos produtos, utilizando canais como redes sociais, e-mail marketing e anúncios segmentados.
2. **Otimização da Experiência do Usuário:**
   * Investir na otimização da jornada do usuário, desde a navegação no site até o processo de checkout. Isso inclui melhorar a usabilidade do site, simplificar o processo de compra e personalizar recomendações de produtos com base no histórico de compras.
   * Exemplo: Após ajustes na interface do usuário para melhorar a simplicidade e clareza, observamos um aumento de 15% na taxa de conversão entre os clientes de 45-54 anos.

### Estratégias de Fidelização e Retenção

1. **Programas de Fidelização e Recompensas:**
   * Implementar programas de fidelidade robustos, como pontos de recompensa por compras, sistema de cashback e benefícios exclusivos para clientes frequentes. Isso incentiva compras repetidas e aumenta a lealdade à marca.
   * Exemplo: O programa de pontos "Valorize sua Fidelidade" não só aumentou as vendas recorrentes entre os clientes de 45-54 anos, mas também gerou um aumento médio de 30% no valor do carrinho de compras.
2. **Monitoramento e Análise Contínua:**
   * Utilizar análises preditivas para monitorar continuamente o comportamento dos clientes e ajustar as estratégias de marketing e vendas conforme necessários. Isso envolve a utilização de ferramentas de Business Intelligence (BI) para identificar tendências emergentes e oportunidades de mercado.
   * Exemplo: A análise preditiva indicou uma mudança nas preferências de produto dentro da faixa etária de 45-54 anos, permitindo ajustes antecipados na oferta de produtos e campanhas promocionais.

### Conclusão

Ao implementar uma abordagem baseada em dados e evidências para entender e atender melhor a faixa etária de 45-54 anos, podemos não apenas aumentar as vendas com esse segmento, mas também garantir uma base de clientes fiel e crescente. Essa estratégia não só se baseia em insights quantitativos e qualitativos sólidos, mas também utiliza exemplos concretos de sucesso e práticas comprovadas no mercado para maximizar o retorno sobre o investimento (ROI) em marketing e vendas.