**PROJETO ML 1 CRISP-DM**

**Fase 1 do projeto:**

Otimização de campanhas de marketing digital, com base em uma empresa fictícia.

Vamos chamar essa empresa de "TechGadgets", uma loja online que vende uma variedade de dispositivos eletrônicos e acessórios.

**Contexto da TechGadgets:**

TechGadgets tem experimentado um crescimento estável nos últimos anos.

No entanto, nos últimos trimestres, a empresa notou um platô nas vendas e um aumento na concorrência de outras lojas online.

A equipe de liderança quer revitalizar o crescimento, focando em campanhas de marketing digital mais eficazes.

Tarefas:

1. Definir Objetivos de Negócios:

Objetivo Principal: Aumentar as vendas online em 15% no próximo trimestre.

Objetivos Secundários:

Melhorar a taxa de conversão de visitantes do site em compradores em 5%.

Aumentar o engajamento com campanhas de e-mail marketing em 20%.

Identificar e focar nos segmentos de clientes mais lucrativos.

2. Situação Atual:

Desafio: Apesar de um aumento constante no tráfego do site, a taxa de conversão e o valor médio do pedido não aumentaram conforme esperado.

Contexto: Análises iniciais sugerem que as campanhas de marketing não estão tão personalizadas quanto poderiam estar, e os concorrentes estão oferecendo promoções mais atraentes.

3. Critérios de Sucesso:

Quantitativos:

Aumento de 15% nas vendas online em comparação ao trimestre anterior.

Aumento de 5% na taxa de conversão do site.

Aumento de 20% na taxa de abertura e clique em campanhas de e-mail.

Qualitativos:

Feedback positivo dos clientes sobre as campanhas de marketing.

Melhoria na percepção da marca, medido através de uma pesquisa de satisfação do cliente.

Perguntas para Reflexão:

Quais características dos clientes estão associadas a uma maior probabilidade de compra?

Esta pergunta ajudará a identificar os segmentos de clientes a serem focados.

Quais campanhas anteriores foram mais bem-sucedidas e por quê?

Compreender o que funcionou no passado pode iluminar estratégias para o futuro.

Stakeholders:

Equipe de Marketing: Diretamente envolvida na criação e execução de campanhas.

Equipe de Vendas: Beneficiada pelo aumento das vendas e pode oferecer insights sobre as preferências dos clientes.

Liderança Executiva: Interessada nos resultados gerais e no retorno sobre o investimento.

Clientes: Embora não diretamente envolvidos, seus feedbacks e comportamentos são cruciais para o sucesso do projeto.

**Conclusão da Fase 1:**

Com essa configuração, definimos claramente o que a TechGadgets deseja alcançar, entendemos o contexto atual e estabelecemos critérios claros para medir o sucesso.

Estas são as fundações sobre as quais construiremos nas próximas fases do projeto.

Próximos Passos: Agora que temos uma compreensão sólida do negócio e de seus objetivos, a próxima fase será mergulhar nos dados disponíveis para a TechGadgets, entender o que temos, o que precisamos e como podemos usá-los para atingir nossos objetivos.

**Fase 2: Entendimento dos Dados**

Seleção de Dados: Eu sugerirei fontes de dados públicas que podemos usar ou criarei um cenário com dados simulados que representam as vendas online, dados dos clientes e informações de campanha.

Acesso aos Dados: Fornecerei instruções sobre como acessar e carregar os dados em seu ambiente de análise (como Python ou R).

Exploração Inicial: Eu fornecerei códigos e orientações sobre como realizar uma análise exploratória básica para entender a estrutura, qualidade e características dos dados.

**Fase 3: Preparação dos Dados**

Limpeza e Transformação: Eu guiarei você através dos processos de limpeza de dados e transformação, incluindo tratamento de valores ausentes, codificação de variáveis categóricas e normalização.

Códigos de Implementação: Fornecerei códigos de exemplo para realizar essas tarefas usando linguagens de programação de análise de dados, como Python ou R.

**Fase 4: Modelagem**

Seleção de Modelos: Discutiremos quais modelos são apropriados para o nosso objetivo, seja para segmentação de clientes, previsão de vendas ou análise de campanhas.

Implementação do Modelo: Fornecerei códigos de exemplo e explicarei como treinar e validar os modelos escolhidos.

**Fase 5: Avaliação**

Avaliação de Modelos: Mostrarei como usar métricas para avaliar o desempenho dos modelos e como interpretar esses resultados no contexto dos objetivos de negócios.

Interpretação dos Resultados: Discutiremos como os resultados podem ser usados para tomar decisões de negócios informadas.

**Fase 6: Implantação**

Plano de Ação: Discutiremos estratégias para implementar insights derivados da análise, como ajustar campanhas de marketing.

Monitoramento: Falaremos sobre como configurar um sistema para monitorar o desempenho das campanhas e o impacto das mudanças implementadas.

Suporte Contínuo:

Perguntas e Esclarecimentos: Em cada etapa, estarei aqui para responder a quaisquer perguntas, fornecer esclarecimentos e ajudar a superar quaisquer desafios.

Adaptação e Flexibilidade: O projeto será adaptado ao seu ritmo e às suas necessidades de aprendizado. Podemos ajustar e expandir as etapas conforme necessário.

Estou aqui para ajudar você a desenvolver suas habilidades de análise de dados e aplicá-las em um contexto real de negócios!

—---------------------------------------

Aqui está uma visão geral das bibliotecas e ferramentas Python que provavelmente utilizaremos:

Bibliotecas Comuns em Projetos de Análise de Dados:

NumPy: Para operações matemáticas e trabalhar com arrays.

Pandas: Para manipulação e análise de dados tabulares.

Matplotlib e Seaborn: Para visualização de dados.

Scikit-learn: Para algoritmos de machine learning e modelagem estatística.

SciPy: Para tarefas científicas e técnicas mais avançadas.

Ambiente de Desenvolvimento:

Jupyter Notebook ou JupyterLab: Ambientes interativos que facilitam a escrita e a execução de código, bem como a visualização de dados.

Estrutura do Projeto:

Configuração do Ambiente: Certificar-se de que seu ambiente Python está pronto e todas as bibliotecas necessárias estão instaladas.

Importação e Configuração de Dados: Aprender como carregar e configurar seus dados para análise.

Análise Exploratória de Dados (EDA): Explorar os dados usando estatísticas descritivas e visualizações para entender melhor o que você está trabalhando.

Preparação dos Dados: Limpar e transformar os dados para modelagem.

Modelagem: Implementar e treinar modelos de machine learning.

Avaliação e Interpretação: Avaliar os modelos e interpretar os resultados.

Implantação e Monitoramento: Discutir estratégias para colocar seus insights em prática.

—------------------------------------------

Próximos Passos:

Quando você estiver pronto, podemos começar com a Fase 2:

Entendimento dos Dados.

Eu fornecerei um conjunto de dados simulado ou sugerirei um conjunto de dados público para trabalharmos, juntamente com instruções passo a passo sobre como carregá-lo e começar a explorá-lo usando Python.

Em cada fase, fornecerei códigos de exemplo e explicarei como você pode aplicá-los ao seu projeto. E claro, estarei disponível para responder a quaisquer perguntas e ajudar com quaisquer desafios que você encontrar.

—------------------------------------------

Vamos abordar o projeto de forma estruturada e progressiva, começando com os conceitos e técnicas básicas e avançando gradualmente para aspectos mais complexos e avançados.

A ideia é construir uma base sólida de conhecimento e habilidades, garantindo que você entenda cada etapa antes de passar para a próxima.

Aqui está como podemos estruturar nossa jornada de aprendizado:

**Etapa 1: Fundamentos de Python e Análise de Dados**

Python Básico: Revisão de conceitos fundamentais de Python, como variáveis, tipos de dados, estruturas de dados, loops e funções.

Manipulação Básica de Dados com Pandas: Aprender a carregar, explorar e manipular dados usando o Pandas.

Visualização Básica de Dados: Introdução à visualização de dados com Matplotlib e Seaborn, aprendendo a criar gráficos como gráficos de linha, barras e dispersão.

**Etapa 2: Análise Exploratória de Dados (EDA)**

Estatísticas Descritivas: Compreender medidas de tendência central e dispersão.

Limpeza de Dados: Tratar dados ausentes, remover duplicatas e lidar com erros.

Visualização Avançada: Construir visualizações mais complexas para entender as relações entre variáveis.

**Etapa 3: Preparação de Dados para Modelagem**

Transformação de Dados: Aprender técnicas para normalização, padronização e codificação de variáveis categóricas.

Criação de Recursos (Feature Engineering): Criar e selecionar recursos úteis para modelagem.

**Etapa 4: Introdução à Modelagem Estatística e Machine Learning**

Modelos Lineares: Implementar modelos lineares simples para entender a relação entre variáveis.

Validação Cruzada: Aprender a técnica para avaliar a performance de modelos em dados não vistos.

Métricas de Avaliação: Compreender diferentes métricas para avaliar modelos de regressão e classificação.

**Etapa 5: Modelos de Machine Learning Avançados**

Modelos de Classificação: Aprofundar em modelos como Regressão Logística, Árvores de Decisão, e Florestas Aleatórias.

Modelos de Agrupamento: Explorar técnicas de clusterização como K-means.

Ajuste de Hiperparâmetros: Aprender a otimizar modelos para melhorar o desempenho.

**Etapa 6: Interpretação de Modelos e Comunicação de Resultados**

Interpretação de Resultados: Aprender a interpretar os resultados dos modelos e o que eles significam para o negócio.

Visualizações Avançadas: Criar visualizações complexas para apresentar os resultados dos modelos.

**Etapa 7: Implantação e Monitoramento de Modelos**

Implantação de Modelos: Discutir estratégias para implementar modelos em produção.

Monitoramento e Manutenção: Aprender a monitorar o desempenho do modelo ao longo do tempo e atualizar quando necessário.

**Suporte Contínuo e Avanço Progressivo:**

Resolução de Dúvidas: Em cada etapa, estarei disponível para responder suas perguntas e fornecer explicações detalhadas.

Exercícios Práticos: Propor exercícios e projetos práticos para reforçar o aprendizado.

Revisão e Feedback: Oferecer revisões e feedbacks sobre os exercícios e projetos que você realizar.

**Adaptando o Ritmo:**

Pausas para Consolidação: Teremos pausas para garantir que você está confortável com o material antes de avançar.

Revisões Regulares: Faremos revisões regulares para reforçar conceitos anteriores e garantir uma compreensão sólida.

Essa estrutura é flexível e será adaptada conforme suas necessidades e progresso.

O objetivo é garantir que você não apenas entenda cada conceito e técnica, mas também saiba como aplicá-los em situações reais para fornecer serviços de alta qualidade na área de análise de dados.

—-----------------------------------------------------------------

**START DO PROJETO**

Fase 1 - Entendimento do Negócio, focando na etapa 1, que é a introdução aos fundamentos de Python necessários para análise de dados.

Isso será feito no contexto das necessidades da empresa fictícia "TechGadgets", onde o objetivo é entender e definir claramente o problema de negócios, os objetivos e os critérios de sucesso.

**Fase 1: Entendimento do Negócio**

Etapa 1: Fundamentos de Python para Análise de Dados

Antes de mergulharmos nas necessidades específicas da TechGadgets, é essencial que você esteja confortável com os fundamentos do Python. Vamos cobrir:

1 Ambiente de Desenvolvimento:

Certifique-se de ter o Python e as ferramentas necessárias instaladas (como Anaconda, que inclui Python, Jupyter Notebook e as principais bibliotecas).

Abra um novo Jupyter Notebook, que será sua área de trabalho interativa.

2 Python Básico:

Variáveis e Tipos de Dados: Aprenda a trabalhar com números, strings e booleanos.

Estruturas de Dados: Familiarize-se com listas, tuplas, conjuntos e dicionários.

Controles de Fluxo: Pratique com loops (for e while) e condicionais (if, elif, else).

3 Introdução ao Pandas:

Séries e DataFrames: Entenda as estruturas de dados principais do Pandas para manipulação de dados.

Leitura de Dados: Aprenda a carregar dados de um arquivo CSV em um DataFrame.

Exploração Básica de Dados: Use métodos como .head(), .describe(), .info() para explorar os dados carregados.

4 Visualização Básica com Matplotlib/Seaborn:

Gráficos de Linha e Barras: Crie gráficos simples para entender as tendências dos dados.

Histogramas: Visualize a distribuição dos dados.

Tarefas:

1 Configurar o Ambiente:

Instale o Python e o Jupyter Notebook via Anaconda (se ainda não o fez).

Abra um novo Jupyter Notebook.

2 Revisar Python Básico:

Crie algumas variáveis, listas, e dicionários.

Escreva um loop simples e uma declaração condicional.

3 Prática com Pandas:

Importe a biblioteca Pandas.

Carregue um arquivo CSV de exemplo em um DataFrame.

Explore os dados usando métodos básicos do Pandas.

4 Exploração Visual dos Dados:

Importe as bibliotecas Matplotlib e Seaborn.

Crie um gráfico simples para visualizar os dados.

Aplicando ao Contexto da TechGadgets:

Enquanto você pratica essas habilidades básicas, pense em como elas podem se aplicar ao contexto da TechGadgets.

Por exemplo, ao carregar um conjunto de dados, imagine que são os dados de vendas da empresa ou informações sobre as campanhas de marketing.

Próximo Passo: Depois de concluir estas tarefas introdutórias e se sentir confortável com o básico, você estará pronto para começar a explorar e entender os dados específicos da TechGadgets, o que faremos na Fase 2.

—---------------------------

**Fase 1: Entendimento do Negócio - Detalhamento**

1. Contexto e Desafio da TechGadgets:

A TechGadgets notou uma desaceleração nas vendas apesar de um aumento no tráfego do site.

A concorrência está aumentando, e a empresa precisa de uma estratégia de marketing mais eficaz para se destacar.

2. Definição de Objetivos de Negócios:

Objetivo Principal: Aumentar as vendas online em 15% no próximo trimestre.

Objetivos Secundários:

Melhorar a taxa de conversão de visitantes do site em compradores em 5%.

Aumentar o engajamento com as campanhas de e-mail marketing em 20%.

3. Critérios de Sucesso:

Quantitativos: Especificar metas numéricas claras (como as mencionadas acima).

Qualitativos: Melhorar a satisfação do cliente e a percepção da marca.

4. Identificação de Perguntas Chave:

Quais segmentos de clientes são mais propensos a comprar?

Quais produtos têm o maior potencial de crescimento de vendas?

Quais foram os atributos das campanhas mais bem-sucedidas no passado?

5. Planejamento de Projeto:

Cronograma: Estabelecer um cronograma realista para as fases do projeto.

Recursos Necessários: Determinar quais dados e ferramentas serão necessários.

Riscos e Suposições: Identificar quaisquer potenciais riscos para o projeto e suposições feitas.

Tarefas:

Documentar o Contexto e Desafio: Escreva um breve resumo sobre a situação atual da TechGadgets e os desafios que ela enfrenta.

Definir Objetivos Específicos: Refine os objetivos de negócios listados acima, tornando-os tão específicos e mensuráveis quanto possível.

Estabelecer Critérios de Sucesso: Liste os critérios específicos que serão usados para avaliar o sucesso do projeto.

Formular Perguntas Chave: Com base nos objetivos do negócio, elabore uma lista de perguntas que a análise de dados precisa responder.

Esboçar um Plano de Projeto: Crie um esboço simples do plano do projeto, incluindo cronograma, recursos necessários e qualquer suposição ou risco relevante.

Aplicando ao Contexto da TechGadgets:

À medida que você trabalha nessas tarefas, mantenha o foco nas necessidades específicas da TechGadgets. Cada decisão e plano deve ser feito com os objetivos finais da empresa em mente.

Próximos Passos: Depois de completar estas tarefas, você terá uma compreensão clara e um plano bem definido para abordar o problema da TechGadgets. Então, estaremos prontos para avançar para a Fase 2: Entendimento dos Dados, onde começaremos a explorar e analisar os dados que ajudarão a informar e orientar suas estratégias de marketing.

Por favor, prossiga com as tarefas listadas e me avise quando estiver pronto para avançar ou se tiver quaisquer dúvidas ou precisar de esclarecimentos em qualquer uma das etapas!

—-----------------------------------------------------------

Aplicando ao Contexto da TechGadgets:

Vamos imaginar que a TechGadgets tem um conjunto de dados de vendas que inclui informações como data da venda, categoria do produto, quantidade e receita. Você usará o Pandas para carregar esses dados e o Matplotlib/Seaborn para criar visualizações que ajudarão a entender as tendências de vendas e o comportamento do cliente.

Tarefas:

1 Instalar o Python e o Jupyter Notebook (se ainda não estiver instalado).

2 Praticar Python Básico: Experimente criar variáveis, listas, loops e condicionais.

3 Explorar Dados com Pandas: Carregue um conjunto de dados de exemplo e pratique a visualização das primeiras linhas e estatísticas.

4 Visualizar Dados: Crie gráficos simples para entender as tendências nos dados.

Próximos Passos:

Após concluir essas tarefas, você terá uma boa compreensão das ferramentas básicas necessárias para análise de dados. Então, estaremos prontos para aprofundar na análise específica para a TechGadgets, incluindo a definição de objetivos, a formulação de perguntas e o planejamento do projeto.

—-------------------------------------------------------------------------------

**HANDS ON**

DATASETS ON LINE RENTAIL

https://archive.ics.uci.edu/dataset/352/online+retail

<https://archive.ics.uci.edu/datasets?search=Online+Retail>

**ANÁLISE INICIAL**

você forneceu um arquivo CSV chamado "Online Retail - Online Retail.csv", que será a base para nossa análise de dados.

Vou começar carregando e examinando os dados para entender sua estrutura e conteúdo.

Vamos verificar as primeiras linhas, a quantidade de dados e as informações gerais sobre o conjunto de dados para planejar nossa análise.

O conjunto de dados "Online Retail" contém 541,909 entradas e 8 colunas. Aqui estão os detalhes das colunas:

1 InvoiceNo: Número da fatura de compra.

2 StockCode: Código do item em estoque.

3 Description: Descrição do produto.

4 Quantity: Quantidade de cada produto por transação.

5 InvoiceDate: Data e hora da compra.

6 UnitPrice: Preço unitário do produto.

7 CustomerID: ID numérico do cliente.

8 Country: País de origem do cliente.

Das primeiras linhas, podemos ver exemplos de produtos vendidos, como "WHITE HANGING HEART T-LIGHT HOLDER" e "WHITE METAL LANTERN", com detalhes sobre a quantidade, preço e o cliente que fez a compra.

Observações Iniciais:

A coluna Description tem alguns valores faltantes, o que pode ser relevante dependendo do tipo de análise.

CustomerID também tem valores faltantes. Isso é crítico para qualquer análise que dependa da identificação de clientes únicos.

InvoiceDate está no formato de string; para análises temporais, precisaremos convertê-la para um formato de data.

**PRÓXIMOS PASSOS DA ANÁLISE**

Próximos Passos na Análise:

1 Limpeza dos Dados:

Lidar com valores faltantes.

Converter tipos de dados onde necessário (por exemplo, InvoiceDate para datetime).

2 Exploração e Análise de Dados:

Análise descritiva (médias, mediana, modas).

Distribuição de quantidades e preços.

Análise de clientes e países.

3 Análise Temporal:

Tendências ao longo do tempo.

Padrões sazonais.

4 Segmentação de Clientes:

Identificação de segmentos de clientes.

Análise de comportamento de compra.

5 Análise de Produtos:

Produtos mais vendidos.

Relação entre preço e quantidade vendida.

6 Insights de Negócios:

Identificar oportunidades para campanhas direcionadas.

Avaliar a eficácia das estratégias de preços.

MAIS DETALHES DA ANÁLISE + UMA ANÁLISE MAIS PROFUNDA

Vamos mergulhar em uma análise profunda dos dados do "Online Retail".

Vou seguir uma estrutura lógica para explorar e analisar o conjunto de dados, fornecendo insights relevantes para negócios e marketing.

Aqui está o plano detalhado de nossa análise:

1. Limpeza e Preparação dos Dados:

Tratamento de Valores Faltantes: Decidir como lidar com os dados faltantes nas colunas Description e CustomerID.

Conversão de Tipos de Dados: Transformar InvoiceDate em datetime e verificar outros tipos de dados.

Remoção de Dados Anômalos: Identificar e tratar possíveis outliers ou erros nos dados (por exemplo, quantidades negativas).

2. Análise Exploratória:

Estatísticas Descritivas: Resumo estatístico das colunas numéricas.

Distribuição de Quantidades e Preços: Analisar a distribuição e tendências de quantidades e preços.

Análise de Transações: Quantas transações por país, tendências ao longo do tempo, etc.

3. Análise Temporal:

Tendências ao Longo do Tempo: Como as vendas, quantidades e preços variam ao longo do tempo?

Padrões Sazonais: Existem padrões específicos por dia da semana, mês ou períodos do ano?

4. Análise de Clientes:

Segmentação de Clientes: Identificar diferentes tipos de clientes (por exemplo, com base no valor de compra, frequência, etc.).

Comportamento de Compra: Analisar como os clientes compram, quais produtos preferem e suas tendências de gasto.

5. Análise de Produtos:

Produtos Mais Vendidos: Quais são os produtos mais populares?

Relação entre Preço e Quantidade: Como o preço afeta a quantidade vendida?

6. Insights de Negócios:

Oportunidades para Campanhas: Identificar produtos ou segmentos de clientes para campanhas de marketing direcionadas.

Estratégias de Preços: Avaliar como o preço afeta a demanda e identificar pontos de preço ótimos.

Vou começar com a limpeza e preparação dos dados.

Isso inclui o tratamento de valores faltantes, conversão de tipos de dados e a remoção de possíveis dados anômalos.

LIMPEZA E PREPARAÇÃO DOS DADOS

Limpeza e Preparação dos Dados Concluída

1 Valores Faltantes:

Antes da Limpeza:

Description: 1,454 valores faltantes.

CustomerID: 135,080 valores faltantes.

Após Remover Transações com Quantidades Negativas:

Description: 592 valores faltantes.

CustomerID: 133,361 valores faltantes.

Conversão de Tipos de Dados:

InvoiceDate foi convertida para o formato datetime.

Remoção de Dados Anômalos:

Removemos transações com quantidades negativas, reduzindo as entradas de 541,909 para 531,285.

Observações:

CustomerID: Uma grande quantidade de dados faltantes permanece.

Para algumas análises, como a segmentação de clientes, precisaremos considerar apenas os registros com CustomerID.

Description: Poucos valores faltantes, que podem ser descartados ou mantidos, dependendo da análise.

PRÓXIMOS PASSOS

Agora que os dados estão limpos e preparados, podemos prosseguir com a análise exploratória.

Isso incluirá estatísticas descritivas, análise da distribuição de quantidades e preços, e uma visão geral das transações.

—--------------------------------------------------------------------------------

**1 PRÓXIMA ETAPA: LIMPEZA E PREPARAÇÃO DOS DADOS**

1. Importação das Bibliotecas e Carregamento dos Dados

Primeiro, importamos a biblioteca necessária e carregamos os dados do arquivo CSV.

2. Verificação de Valores Faltantes

Verificamos a quantidade de valores faltantes em cada coluna para entender onde precisamos focar nossos esforços de limpeza.

3. Conversão de Tipos de Dados

A coluna InvoiceDate estava no formato string. Para permitir análises temporais, convertemos para o tipo datetime.

4. Remoção de Dados Anômalos

Identificamos que algumas entradas tinham quantidades negativas, o que poderia indicar devoluções ou erros.

Decidimos remover essas entradas para focar apenas nas transações de compra regulares.

Nessa etapa é realizada uma limpeza básica e prepara os dados para análises mais profundas.

A quantidade de valores faltantes e o tipo de dados foram ajustados, tornando o conjunto de dados mais consistente e confiável para análises futuras.

—--------------------------------------------------------------------------------------------

**2 PRÓXIMA ETAPA: ANÁLISE EXPLORATÓRIA DOS DADOS**

Esta fase é crucial para entender as tendências gerais, padrões e características do nosso conjunto de dados

1. Estatísticas Descritivas

Primeiro, vamos gerar estatísticas descritivas para ter uma visão geral das colunas numéricas: Quantity e UnitPrice. Isso nos dará uma ideia da média, mediana, valores máximos e mínimos, entre outros.