**Projeto da disciplina**

**Ciência de Dados e Inteligência Artificial**

**Fase 1**

|  |  |
| --- | --- |
| **Nome do estudante** | **Daiane Deponti Bolzan** |

|  |
| --- |
| SÍNTESE DO CONJUNTO DE DADOS COMO APRESENTADO NA SUA ORIGEM |
| *Apresente uma breve descrição, acerca do conjunto de dados escolhido, indicando as informações contidas nos dados selecionados. É importante, ao apresentar sua síntese, que você inclua elementos que permitam compreender o significado e a aplicação dos dados escolhidos.* |
| O conjunto de dados escolhido refere-se a taxas de câmbio diárias de diversas moedas em relação ao real, disponibilizado pelo Banco Central do Brasil. Ele inclui informações sobre valores de compra e venda para moedas como dólar americano, euro, iene, libra esterlina e dólar canadense. Os dados são apresentados em formato tabular, contendo datas e valores das cotações. |

|  |
| --- |
| DESCRIÇÃO DO SEU INTERESSE EM EXPLORAR ESSE CONJUNTO DE DADOS |
| *Em até 3 parágrafos, explique o interesse em explorá-lo.* |
| A exploração desse conjunto de dados permite compreender padrões e tendências no mercado cambial, essencial para análises econômicas e estratégias financeiras. Além disso, pode-se avaliar correlações entre diferentes moedas e o impacto de eventos econômicos globais.  Esse estudo também é relevante para prever movimentos de câmbio utilizando técnicas de aprendizado de máquina e estatística. Com essas previsões, é possível desenvolver estratégias de proteção contra variações cambiais.  Outro ponto importante é a aplicação desse conjunto de dados no setor empresarial, auxiliando em decisões sobre importação, exportação e investimentos em moedas estrangeiras. |

|  |
| --- |
| FINALIDADE DO CONJUNTO DE DADOS ESCOLHIDO |
| *Descreve para que serve esse conjunto de dados selecionado, ou seja, a razão pela qual eles foram coletados e estruturados.* |
| Os dados foram coletados para monitoramento das taxas de câmbio e auxiliar análises econômicas. Eles são utilizados por economistas, investidores e empresas para entender variações cambiais e prever tendências de mercado. São coletados e disponibilizados pelo Banco Central.  A ideia principal é criarmos um modelo que permita analisar se houve modificação nas taxas de câmbio durante o período da pandemia e quais as principais moedas que contribuíram para essa possível modificação se existir. Desta forma, utilizaremos uma variável binária como variável target para identificar essa possível modificação utilizando algoritmos. |

|  |
| --- |
| QUANTIDADE DE LINHAS E COLUNAS DO CONJUNTO DE DADOS |
| *Descreva apenas a quantidade de linhas e de colunas selecionadas.* |
| O conjunto de dados possui 6578 linhas e 14 colunas. |

|  |
| --- |
| QUAL O FORMATO QUE O CONJUNTO DE DADOS É DISPONIBILIZADO |
| *CSV, JSON, XLSX, etc.* |
| O conjunto de dados está disponível no formato CSV. |

|  |
| --- |
| Escolha pelo menos 10 colunas totalmente preenchidas (as mais importantes) e, para cada coluna (inclusive para a coluna alvo):   * Qual o nome e o que representa? * Qual o tipo de dados? Nominal/Ordinal/Numérico/Data e/ou hora? * Quais são os valores considerados válidos? * Quantos valores distintos aparecem na coluna? * Qual o menor e o maior valor, e qual a moda? * Os valores da coluna são numéricos? Qual a média e qual o desvio-padrão? Qual a mediana? |
| *Você pode apresentar essas informações da maneira como entender mais conveniente, como por exemplo em uma tabela, parágrafos, etc. É importante que sejam inseridas as imagens e as evidências do passo a passo realizado na atividade.* |
| |  |  |  | | --- | --- | --- | | Nome da Coluna | Descrição | Tipo de Dados | | Pandemia | Variável Binária para identificar o período da pandemia | Categórico | | Data | Data da cotação | Data | | Dólar (Venda) | Taxa de venda do dólar americano | Numérico | | Dólar (Compra) | Taxa de compra do dólar americano | Numérico | | Euro (Venda) | Taxa de venda do euro | Numérico | | Euro (Compra) | Taxa de compra do euro | Numérico | | Iene (Venda) | Taxa de venda do iene | Numérico | | Iene (Compra) | Taxa de compra do iene | Numérico | | Libra (Venda) | Taxa de venda da libra esterlina | Numérico | | Libra (Compra) | Taxa de compra da libra esterlina | Numérico | | Dólar Australiano (Venda) | Taxa de venda do dólar australiano | Numérico | | Dólar Australiano (Compra) | Taxa de compra do dólar australiano | Numérico | | Dólar Canadense (Venda) | Taxa de venda do dólar canadense | Numérico | | Dólar Canadense (Compra) | Taxa de compra do dólar canadense | Numérico |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Nome da Coluna | Média | Moda | Mediana | Dispersão | Mínimo | Máximo | Dados  faltantes | Mediana |  |  |  |  |  |  |  |  |  |  | | - | 0 | 0 | - | 0,367 | 0 | 1 | 0 |  |  |  |  |  |  |  |  |  |  |  | | Data | 07/02/2012 | 31/12/1998 | 05/02/2012 | ~27 anos | 31/12/1998 | 12/03/2025 | 1 | - |  |  |  |  |  |  |  |  |  |  | | Dólar (Venda) | 3,05806 | 2,0266 | 2,6368 | 0,418092 | 1,2078 | 6,2086 | 1 | 3.20 |  |  |  |  |  |  |  |  |  |  | | Dólar (Compra) | 3,05736 | 1,6712 | 2,636 | 0,418208 | 1,207 | 6,208 | 1 | 3.18 |  |  |  |  |  |  |  |  |  |  | | Euro (Venda) | 3,5413 | 3,7528 | 3,2089 | 0,378434 | 1,39445 | 6,9422 | 1 | 3.80 |  |  |  |  |  |  |  |  |  |  | | Euro (Compra) | 3,53854 | 2,4967 | 3,2069 | 0,378798 | 1,39093 | 6,9392 | 1 | 3.75 |  |  |  |  |  |  |  |  |  |  | | Iene (Venda) | 0,0270885 | 0,02034 | 0,02502 | ,0347231 | 0,010715 | 0,05545 | 2 | 0.030 |  |  |  |  |  |  |  |  |  |  | | Iene (Compra) | 0,0270673 | 0,02267 | 0,024996 | 0,347611 | 0,010687 | 0,05543 | 2 | 0.029 |  |  |  |  |  |  |  |  |  |  | | Libra (Venda) | 4,48062 | 3,2398 | 4,1257 | 0,318171 | 1,97597 | 8,1143 | 1 | 4.50 |  |  |  |  |  |  |  |  |  |  | | Libra (Compra) | 4,47712 | 2,7417 | 4,128 | 0,318392 | 1,97148 | 8,1111 | 1 | 4.48 |  |  |  |  |  |  |  |  |  |  | | Dólar Australiano(venda) | 2,24597 | 20769 | 2,08405 | 0,370275 | 0,740699 | 4,4954 | 2 |  |  |  |  |  |  |  |  |  |  |  | | Dólar Australiano(Compra) | 2,24383 | 2,1614 | 2,08255 | 0,370827 | ,0738261 | 4,4944 | 2 |  |  |  |  |  |  |  |  |  |  |  | | Dólar Canadense (Venda) | 2,40977 | 1,7424 | 2,14518 | 0,370566 | ,0783634 | 4,6193 | 2 | 3.10 |  |  |  |  |  |  |  |  |  |  | | Dólar Canadense (Compra) | 2,40793 | 16347 | 2,14185 | 0,370958 | 0,781828 | 4,6173 | 2 | 3.08 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  | |

|  |
| --- |
| Qual a oportunidade para um projeto de ciência de dados foi identificada a partir dessa análise? Justifique sua resposta! |
| *Descreva as potencialidades que o conjunto de dados selecionado apresenta do ponto de vista da Ciência de Dados e Inteligência Artificial.* |
| A partir da análise inicial, observou-se uma grande oportunidade em explorar esse conjunto de dados com técnicas de **Ciência de Dados e Inteligência Artificial**. A criação da **variável binária “Pandemia”**, que classifica os dados entre os períodos de pandemia (de 11/03/2020 a 05/05/2023, conforme a OMS) e os demais, permite conduzir análises comparativas.  Essa categorização possibilita o desenvolvimento de um **modelo de classificação supervisionado**, tendo como objetivo identificar **se houve alteração significativa nas taxas de câmbio durante a pandemia**. Além disso, permite verificar **quais moedas apresentaram maiores variações** e se determinados padrões foram influenciados diretamente por este evento global.  Dessa forma, a aplicação de algoritmos de machine learning — como **Random forest, regressão logística ou Naive Bayes** — pode oferecer insights preditivos relevantes, tanto do ponto de vista econômico quanto estratégico, auxiliando na **tomada de decisões em ambientes de alta volatilidade**. |

|  |
| --- |
| LINK PARA ACESSO AOS DADOS SELECIONADOS |
| *Inclua aqui o link para o conjunto de dados escolhido. No caso de ter sido feito um recorte/amostragem do conjunto original, documentar quais foram os passos feitos para produzir o recorte.* |
| Todos os dados e arquivos utilizados estão disponíveis em: <https://github.com/Daaaiii/cienciadados>  O conjunto de dados original pode ser acessado através do Banco Central do Brasil: [Banco Central - Séries Temporais.](mailto:https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries) Na opção de seleção por tema, selecionei as seguintes opções: |

Cliquei em consultar valores e depois ordenei conforme a data de início e selecionei o período inicial.

Tela de computador com texto preto sobre fundo branco

O conteúdo gerado por IA pode estar incorreto.

Ao clicar em visualizar valores a seguinte página abrirá com a possibilidade de exportar o arquivo csv;

Interface gráfica do usuário, Aplicativo, Tabela

O conteúdo gerado por IA pode estar incorreto.

Como o site permite exportar 10 variáveis por vez, precisei criar 2 arquivos no Orange, transformar os valores de texto para número com o “Edit domain” e fazer o merge das tabelas com o Merge Data.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo, Email

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Aplicativo

O conteúdo gerado por IA pode estar incorreto.