

Case de Análise de Dados

Objetivo:

Avaliar a capacidade de realizar análises estatísticas, identificar correlações entre variáveis e propor implementações adicionais relevantes.

Descrição do Exercício:

Você receberá um conjunto de dados fictício contendo informações sobre a utilização de um plano de saúde. O conjunto de dados inclui as seguintes variáveis:

- EMPRESA: Nome da empresa
- DATA_REFERENCIA: Data de referência
- CODIGO_APOLICE: Código da apólice
- TITULARIDADE: Titularidade (Titular ou Dependente)
- GENERO: Gênero do beneficiário
- IDADE: Idade do beneficiário
- DATA_ATENDIMENTO: Data do atendimento
- FLAG_INTERNACAO: Indicador de internação (Sim ou Não)
- CODIGO_INTERNACAO: Código da internação
- DATA_INTERNACAO: Data da internação
- CODIGO_EVENTO: Código do evento
- VALOR_PAGO: Valor pago pelo evento
- DATA_PAGAMENTO_EVENTO: Data do pagamento do evento
- QUANTIDADE_SERVICOS: Quantidade de serviços prestados
- CIDADE: Cidade do atendimento
- REDE_REEMBOLSO: Rede de reembolso (Sim ou Não)
- PLANO_SAUDE: Tipo de plano de saúde
- ESPECIALIDADE: Especialidade médica do atendimento
- EVENTO_OPERADORA: Evento registrado pela operadora
- TIPO_EVENTO: Tipo de evento (Consulta, Exame, Cirurgia, etc.)
- CODIGO_SERVICO: Código do serviço prestado
- DESCRICAO_SERVICO: Descrição do serviço prestado

Tarefas:

Análise Estatística Descritiva:

Calcule as estatísticas descritivas (média, mediana, desvio padrão, etc.) para as variáveis IDADE e VALOR_PAGO.

Visualize a distribuição dos atendimentos por **ESPECIALIDADE**.

Análise de Correlação:

Calcule a correlação entre TIPO_EVENTO e VALOR_PAGO.

Visualize a correlação usando um gráfico de dispersão (scatter plot).

Análise de Tendências:

Analise a tendência de utilização do plano de saúde ao longo do tempo. Crie uma visualização que mostre a evolução dos atendimentos mensais.

Modelagem Preditiva:

Utilize os dados para treinar um modelo de machine learning que preveja o VALOR_PAGO nos próximos meses.

Divida os dados em conjuntos de treino e teste e avalie a acuracidade do modelo utilizando métricas apropriadas.

Visualize as previsões do modelo e compare com os valores reais.

Proposta de Implementação Adicional:

Proponha e implemente uma análise adicional que você considere relevante para entender melhor os dados. Pode ser uma análise de segmentação de beneficiários, análise de sazonalidade, ou qualquer outra que julgar pertinente.

Entrega:

Um relatório em formato Jupyter Notebook (.ipynb) contendo:

- Código utilizado para as análises.
- Visualizações geradas.
- Interpretação dos resultados obtidos.
- Descrição da análise adicional proposta e implementada.

Critérios de Avaliação:

- Clareza e organização do código.
- Qualidade das visualizações.
- Correção das análises estatísticas.
- Relevância e criatividade da análise adicional proposta.
- Clareza na interpretação dos resultados.