



Desafio: Construindo um modelo de Regressão para marketing

☰ Escola	Dados
☷ Habilidades	Regressão
# N°	5



Desafio



Arquivo do Desafio:

- **Python Graph Gallery** → Repositório com o passo a passo de como gerar gráficos utilizando as principais bibliotecas de Python
- **SciKit Learn** → Documentação com os principais modelos utilizados para clusterização
- **Base de dados:**

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/6acf00f3-d13e-45bf-8b24-ff63a6c85bf4/MKT.csv>

Prepare seu dataset para modelagem de dados

Aplique os conhecimentos de regressão para estruturar uma análise em um dataset de campanha de marketing. O modelo de machine learning mais utilizado no mercado são os de regressão, com isso é importante que você consiga desenvolver esses modelos para se destacar.

Contexto - Introdução

Uma empresa está investindo mensalmente em plataformas de publicidade online, como Youtube, Facebook e newspaper, para a prospecção de leads (pessoas interessadas em seus produtos). A fim de acompanhar o desempenho desses investimentos, a empresa registra todos os gastos com publicidade e todos os retornos de vendas gerados a partir desses investimentos.

Para **entender** melhor **a relação entre as variáveis** presentes nesses registros e **identificar os fatores que mais impactam** na geração de leads, a empresa solicitou a análise de um especialista em dados. **Além disso, a empresa busca criar um modelo de predição** de valores para estimar o retorno de vendas que pode ser gerado a partir de um determinado investimento em publicidade.

Sobre os dados

A tabela contém informações dos investimentos feitos pelo youtube, facebook, newspaper e também a quantidade de cada.

Coluna	Descrição
youtube	Investimento youtube
facebook	Investimento facebook
newspaper	Investimento newspaper
sales	Valor das vendas

Contexto - Como começar?

Comece baixando o arquivo csv. disponibilizado no início desta página. Feito isto, busque tratar e interpretar os dados que receberão, eles serão essenciais para a construção das próximas etapas.

1. Importe o dataset para o colab, utilizando uma biblioteca de manipulação de dados
2. Entenda os dados que está visualizando! Para isso, explore as informações das variáveis presentes no dataset para identificar possíveis inconsistências, como valores ausentes, dados duplicados ou outliers.
3. Faça uma análise descritiva das variáveis
4. Entenda os dados com a análise exploratória. Nesta etapa do desafio, você pode utilizar gráficos e visualizações para identificar padrões, relações e tendências entre as variáveis e investigar possíveis correlações.
5. Realize a montagem do modelo de regressão para construir um modelo que seja capaz de estimar o retorno de vendas a partir de um determinado investimento.

Utilize também dos materiais complementares para te guiar neste desafio!

- **Python Graph Gallery** → Repositório com o passo a passo de como gerar gráficos utilizando as principais bibliotecas de Python
- **SciKit Learn** → Documentação com os principais modelos utilizados para clusterização

Etapas de Desenvolvimento

Para te ajudar nesse processo, detalhar o processo nas etapas a seguir:

Etapa 01) Análise Descritiva

Esta etapa consiste em explorar os dados do dataset para **compreender melhor as variáveis e identificar problemas**. Para isso, é recomendado utilizar a biblioteca **Pandas** para importar e manipular os dados e realizar cálculos estatísticos, além das bibliotecas de visualização.

É importante investigar o tipo de dado em cada variável, os valores e a distribuição dos dados. Ao final, espera-se ter uma interpretação sólida dos dados para avançar para a próxima etapa.



Dica: Utilize a biblioteca Pandas e bibliotecas de visualização de dados para entender como seu dataset está.

Etapa 02) Análise Exploratória

Nesta etapa iremos explorar mais a fundo os dados, **identificando relações entre as variáveis e descobrindo padrões relevantes**. Para isso, utilize técnicas de visualização de dados e análises estatísticas, buscando possíveis correlações e identificando possíveis outliers ou desvios da normalidade.



Dica: Análise a correlação dos dados e a distribuição dos mesmos

Etapa 03) Modelagem

Para esta etapa, deve-se **construir um modelo** simples de **regressão** que permita a previsão solicitada pela empresa, com base nos dados disponíveis. Para isto, importe as bibliotecas necessárias e carregue os conjuntos de dados para iniciar a sua construção!



Dica: Utilize as técnicas ensinadas para a criação de um modelo de regressão simples utilizando a biblioteca sklearn.

Etapa 04) Calculando predição

Para concluirmos a demanda solicitada pela empresa, iremos **aplicar o modelo de regressão construído** nas etapas anteriores **para realizar as previsões** de retorno de vendas que pode ser gerado a partir de um determinado investimento em publicidade e assim, poderemos apresentá-lo a empresa.

Através dessas previsões, poderemos avaliar o impacto dos diferentes níveis de investimento em marketing nas vendas, auxiliando na tomada de decisões e na definição de estratégias de negócio.



Dica: Utilize os conhecimentos aprendidos para calcular o valor de vendas com base nos investimentos em marketing. Utilize técnicas e ferramentas de análise de dados com Python.



Critérios de Avaliação

Os critérios de avaliação mostram como você será avaliado em relação ao seu desafio.

Atendeu às Especificações	Atendeu as expectativas	Pontos
Análise Descritiva	Durante a etapa de Análise Descritiva, é esperado o uso da função <code>describe</code> para analisar como está distribuído o dataset.	25
Análise Exploratória	Durante a etapa de Análise Exploratória, é esperado uma análise das variáveis para identificar possíveis correlações relevantes entre elas, o que será fundamental para a modelagem posterior.	25
Modelagem	Durante a etapa de Modelagem, é esperado que o modelo de regressão seja testado e treinado de forma a identificar possíveis ajustes necessários para garantir que o modelo esteja adequado para a realização de previsões precisas.	25

Calculando predição	Espera-se que o cálculo do coeficiente de determinação (r^2) seja realizado e que esse resultado seja utilizado para fazer a predição das vendas de acordo com o investimento.	25
---------------------	--	----



Entrega



Você deverá submeter o link compartilhável do colab!



Dica: Se preocupe em detalhar cada etapa do seu código