

Desafio Cartola

Contexto:

O Cartola é um jogo eletrônico amplamente reconhecido no Brasil, desenvolvido pelo Grupo Globo. Trata-se de uma plataforma que permite aos entusiastas do futebol montarem seus próprios times virtuais com jogadores da vida real, competindo em ligas de fantasia baseadas no Campeonato Brasileiro de futebol. Os participantes devem usar seus conhecimentos esportivos e estratégias para selecionar os melhores jogadores e criar escalasções vencedoras a cada rodada.

Nesse contexto, surge uma questão fundamental que desejamos explorar: a identificação dos fatores e padrões que influenciam a propensão dos jogadores a se tornarem **assinantes da versão pró** do Cartola. O nosso objetivo é compreender o que motiva um usuário a aderir à versão premium do jogo, com o intuito de ampliar o engajamento e a receita relacionados ao Cartola.

Para alcançar esse objetivo, é essencial conduzir uma análise exploratória de dados abrangente, permitindo-nos compreender os diversos comportamentos de consumo dos jogadores. Além disso, é fundamental construir um **modelo de propensão, como uma classificação binária**, que viabilize a avaliação da hipótese levantada, identificando grupos com maior probabilidade de conversão em assinantes. Os resultados dessa análise deverão ser consolidados e apresentados de forma convincente. O público-alvo da apresentação incluirá **designers, Product Owners (POs) e gerentes de produto**, aos quais você defenderá a implementação da sua solução.

Objetivos:

- Realizar uma análise exploratória dos dados com o objetivo de encontrar informações úteis para identificação de perfis de usuários propensos a assinarem o Cartola. **Utilize técnicas de estatística e visualização de dados para apoiar suas conclusões.**
- Ajustar um modelo preditivo simples para identificar os perfis de usuários propensos a assinarem o Cartola. **Apresente os métodos de limpeza e processamento de dados** para gerar a entrada do modelo preditivo. **Justifique a escolha das técnicas e modelo utilizados.**
- **Elaborar uma apresentação** com a análise, resultados e sugestões de melhorias para o projeto desenvolvido.
- **Disponibilizar o código fonte** utilizado no exercício. Sua solução deve ser organizada, legível e limpa, evitando soluções com over-engineering. **Tente ser sucinto, mas preze pela qualidade do código.**

Entregáveis

- código/notebooks/imagens/documentos gerados durante a análise;
 - Sugestão de ferramenta: Python, SQL, BigQuery, PostgreSQL
- material que será utilizado na apresentação dos resultados
 - Sugestão de ferramenta: ppt, keynote, google slides, pdf

Base de dados

O arquivo anexo, **base_cartola_desafio.csv**, contém informações agregadas de uma amostra de usuários do *GloboEsporte.com*, que acessaram editoriais de diversas modalidades esportivas, durante um período de Olimpíadas. A tabela a seguir descreve sobre os campos inseridos na base de dados:

Atributo	Descrição
user	código único do usuário
sexo	gênero do usuário
uf	estado da federação
idade	idade do usuário
dias	número de dias que veio no período
pviews	quantidade de páginas acessadas
visitas	quantidade de visitas
tempo_total	tempo total em segundos
device	usuário consumiu somente no pc (pc_only), no mobile (m_only) ou em ambos (pc_e_m)
futebol	consumo da modalidade esportiva (em segundos)
futebol_internacional	consumo da modalidade esportiva (em segundos)
futebol_olimpico	consumo da modalidade esportiva (em segundos)
blog_cartola	consumo da modalidade esportiva (em segundos)
atletismo	consumo da modalidade esportiva (em segundos)
ginastica	consumo da modalidade esportiva (em segundos)
judo	consumo da modalidade esportiva (em segundos)
natacao	consumo da modalidade esportiva (em segundos)
basquete	consumo da modalidade esportiva (em segundos)
handebol	consumo da modalidade esportiva (em segundos)
volei	consumo da modalidade esportiva (em segundos)
tenis	consumo da modalidade esportiva (em segundos)
canoagem	consumo da modalidade esportiva (em segundos)
saltos_ornamentais	consumo da modalidade esportiva (em segundos)
home	consumo da modalidade esportiva (em segundos)
home_olimpiadas	consumo da modalidade esportiva (em segundos)
cartola_status	usuário assina o Cartola (Cartola Pro), utiliza a versão gratuita (Cartola Free) ou não utiliza o Cartola (Não Cartola)

Figura 1: Tabela de atributos e descrições