

Business Case 7 - Limpando e explorando dados

Nesta última parte do business case você se aprofundou mais em ciência de dados, principalmente com manipulação de dados. Aprendeu Python e pandas que são extremamente importantes para manipular dados além de outros “tricks” do Python e R para tarefas desse tipo. E, além disso, você aprendeu os fundamentos da base de ciência de dados, que é a estatística.

Agora vamos colocar em prática o que você vem estudando nas aulas.

Visão geral

Um cliente pequeno contratou a Luby para um trabalho de ciência de dados. Você, junto com o time de ciência de dados serão encarregados de analisar esses dados.

O cliente é do ramo de marketing e quer ser direcionado em como investir melhor com propagandas. São dados de vários canais de comunicação diferentes, até o momento foram disponibilizados os dados do youtube que eles têm na base de dados deles. Em resumo, quer-se saber os vídeos que são mais tendência para poder investir em tais segmentos.

Para isso há vários passos que um cientista de dados deve fazer para conseguir chegar ao resultado esperado e este será seu trabalho aqui.

Objetivos

Você deve fazer uma análise exploratória dos dados. Esta análise deve obedecer alguns requisitos e responder algumas perguntas. Seu objetivo é:

- Limpar os dados de entrada
- Fazer uma análise exploratória respondendo às seguintes perguntas:
 1. Os usuários gostam mais de vídeos de qual categoria?
 2. Os usuários gostam mais de vídeos de qual categoria?
 3. Top 5 vídeos que são tendência no país à qual a base de dados foi coletada?
 4. Esses vídeos têm algo em comum?

5. O vídeo mais curtido também é o vídeo mais popular?
6. Número máximo de dias até o status de tendência de um vídeo?
7. Os usuários comentam em qual categoria mais?

Requisitos

Alguns requisitos devem ser seguidos para este trabalho:

- Os códigos devem ser escritos estritamente em R ou Python
- O código deve ser bem comentado indicando o os passos seguidos
- Deve ser feito um arquivo markdown com as respostas das perguntas feitas
- Devem ser indicados também as transformações feitas nos dados na fase de limpeza, por exemplo: mudar as colunas para os tipos corretos.

Dados

Os dados estão em um arquivo csv e correspondem a vários vídeos do youtube e suas respectivas informações. Os mesmos podem ser acessados no link:

<https://drive.google.com/file/d/12wHjH-fAy9tBHooJeVcR5l6QzKaVmUKn/view?usp=sharing>

Entregáveis

- Código em Python ou R contendo a resolução do case (Sugestão: Um jupyter notebook também é bem vindo)
- Arquivo markdown contendo as respostas e qualquer outro comentário relevante (caso seja entregue o Jupyter notebook, os comentários podem ser feitos lá e anulam a necessidade de um arquivo markdown)
- Todos os arquivos devem estar em um link público ou privado no github/gitlab/etc.