

Introdução à análise de dados

Felipe Carvalho *National Institute for Space Research (INPE), Brazil*

Felipe Menino *National Institute for Space Research (INPE), Brazil*

A aplicação dos conceitos vistos no minicurso de introdução à análise de dados é extremamente importante para a fixação dos conceitos. Desta forma, esta lista de exercícios busca criar situações que ajudem o aluno na aplicação e fixação de tais conceitos.

1. Contexto

Atualmente na **Google Play Store** existem milhares de aplicativos, com as mais variadas características. Um conjunto de dados disponibilizados no [Kaggle](#) traz informações gerais sobre 10 mil aplicativos disponibilizados na loja.

O interessante deste conjunto é tentar fazer verificações dos tipos mais populares de aplicativos, para quem sabe, ao começar um novo aplicativo, você já saber qual tendência seguir.

Para esta análise, um conjunto amostral de 1000 registros será utilizado, porém, encorajamos você a depois que finalizar a atividade, acessar a [página dos dados](#) e fazer a análise com todos os dados.

2. Atividade

Como apresentado ao longo das atividades do curso, a análise de dados é basicamente uma atividade utilizada para responder perguntas através de diferentes métodos de tratamento e manipulação dos dados.

Com base nisto, faça a análise do conjunto de dados **MIAD003**, que está no repositório do minicurso, dentro do diretório **atividades**.

Com este conjunto de dados, responda as seguintes perguntas:

1. Qual a categoria de aplicativo mais comum na **Google Play Store** ?
2. Qual a categoria dos aplicativos com maior média de avaliações ?
3. A maior quantidade de *reviews* pertence a qual aplicativo ?
4. Há mais aplicativos pagos ou grátis neste conjunto de dados ?

Dicas

Abaixo uma tabela com a relação de funções que podem ser úteis para a solução dos exercícios propostos

Descrição	R	Python
Leitura de dados	<code>readr::read_csv()</code>	<code>pd.read_csv()</code>
Agrupamento	<code>dplyr::group_by()</code>	<code>pd.DataFrame.groupby()</code>
Filtro	<code>dplyr::filter()</code>	<code>pd.DataFrame[Condição]</code>
Desvio padrão	<code>stats::sd()</code>	<code>pd.DataFrame.std()</code>
Média	<code>mean()</code>	<code>pd.DataFrame.mean()</code>
Mínimo	<code>min()</code>	<code>pd.DataFrame.min()</code>
Máximo	<code>max()</code>	<code>pd.DataFrame.max()</code>