**Data Science**

**Curso I – Primeiros Passos:**

Link da documentação das bibliotecas que utilizaremos nesse curso:

- Pandas: <https://pandas.pydata.org/pandas-docs/stable/>;

- Seaborn: <http://seaborn.pydata.org/introduction.html>.

Vou colocar todo arquivo escrito na pasta da aula, mas aqui está o link do drive, de qualquer forma: <https://drive.google.com/drive/folders/1YV0fPKe-_f5beekL2cBUVSOsZ_2G6uaS>.

1. **Aula 1 – Data Science – Dados e Visualizações:**
   1. Podemos sim alterar o nome das colunas passando uma lista com os novos nomes a partir do atributo colunas do nosso df:

notas.columns = ['usuarioId', 'filmeId', 'nota', 'momento']

* 1. O pandas possui visualização e construção de dados a partir do método plot().
     1. O plot recebe o parâmetro kind=’’, onde podemos escolher o tipo de gráfico, como um hist de histograma, por exemplo. Se colocar somente plot() ele cria um plot normal onde não vemos nada:

notas.nota.plot(kind='hist')



notas.nota.plot()



* + - 1. O hist é um tipo de gráfico que conta uma história, mostrando no eixo x as notas que um filme recebeu e no y a quantidade de notas. Basicamente o inverso do plot.
    1. S.median(): retorna a mediana de uma series, mostrando exatamente qual o valor que dividi um conjunto de dados no meio:

notas.nota.median()



* + 1. Importamos a seaborn e fizemos um boxplot para visualizar melhor os gráficos:

import seaborn as sns

sns.boxplot(notas.nota)



* + - 1. A mediana é o traço no meio do box. Quartil 25 é o final esquerdo do box e 75 o final direito. As linhas com traço na ponta dos lados esquerdo e direito são, respectivamente os limites inferior e superior, ou seja, da mediana para as laterais do box temos 25% cada lado, os li e ls são outros 25%, somando 50% no box e 50% fora.
      2. Com esse gráfico podemos concluir que metade das pessoas, ou seja, o box inteiro, deram notas entre 3 e 4.