Aula 02

**Nomenclaturas Importantes:**

**DataSet** - Significa Conjunto de Dados

**Dicionário de Dados** - " legendas "

**Outlier** - Número discrepante, "Diferentão"

* Ele pode ou não ser válido para a análise de dados
* Ele é um ponto de atenção. Ele pode ou não ser um problema. Tudo dependerá da média estabelecida!

**Média**(Indicador enviesado)

Cálculo da média:

* Soma de registros do atributo dividido pelas quantidades de linhas do atributo

**Mediana(MD) -**

Para solucionar o problema do indicador enviesado(média) utilizamos a mediana.

**Cálculo da posição da mediana:**

* É necessário que os registros estejam em ordem crescente

Posição da mediana = (número de registros + 1) / 2

***Pno = (n + 1) / 2***

**Mediana é igual ao registro da posição da mediana**

**obs: quanto maior for a diferença da mediana maior a possibilidade de encontrar um OUTLIER**

**Desvio Padrão - se difere da mediana por não estar em raiz quadrada**