**Descrição:**

***Agrupamento\_Kmeans:*** esse código escrito em python usa algorítmos de classificação não supervisionada para agrupar produtos recebendo os valores de uma dataset, fornecido, no formato csv.

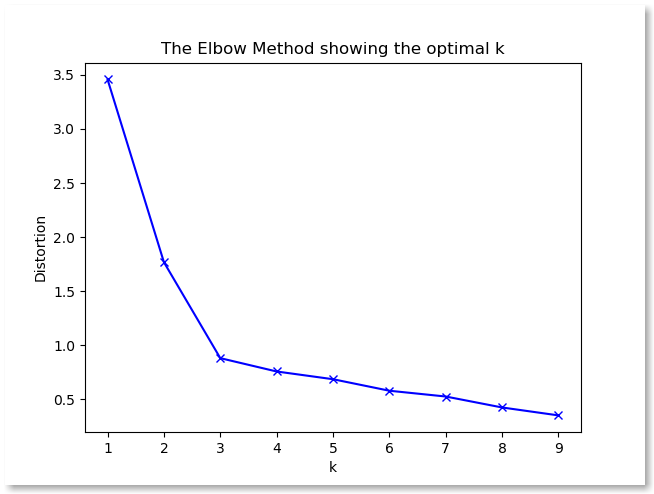
***Classificação\_Knn:*** uma vez identificados os grupos, outliers e etc, esse algorítimo faz a classificação dos produtos e projeção de vendas.

**Detalhes :**

**Kmeans :** também conhecido como algorítmo de Lloyd, é uma técnica de agrupamento que maximiniza dados semelhantes usando a distância do cluster. Uma característica do kmeans, diferente do Meanshift, é que você precisa definir inicialmente o número de cluster antes de iniciar a classificação. Para isso eu usei a técnica do cotovelo ( Elbow method ) .

***Obs.:*** se você não definir o número de cluster, a biblioteca kmeans no scikit learning usado, assume um valor default de 8 clusters.

**Elbow method :** *na prática você utiliza um plano cartesiano ( x,y ), o número k de cluster e a proporção da dispersão ( variância ) entre o modelo e o centro do cluster. Para um k inicial, a dispersão decaí rapidamente a medida em que os valores de k aumentam, o que fica evidenciado pela curva.*



*No ponto em que a curva se acentua, k = 3 na figura, formando a imagem de um cotovelo, temos um valor de k em que a partir daí valores maiores de k não significam uma melhora significativa nos resultados do agrupamento.*

**Knn:** técnica de aprendizado supervisionado. O princípio por trás do Knn é localizar um número pré-definido de amostras ( previamente treinadas ) em distância para um ponto novo ( a que se quer fazer uma previsão ) e prever o objeto.

**Conclusão:** Usando essas técnicas foi possível agrupar os produtos com característiscas semelhantes, e dispor aqueles com características peculiares. Posteriormento, classificar e fazer uma projeção.

***Obs:*** Os códigos foi criado usando, dentro do anaconda : ipython(5.3.0), scikit-learning ( 0.18.1 ), numpy ( 1.12.1 ), pandas(0.20.1) e matplotlib( 2.0.2). Dentro do código podemos observar a utilização de 2 algorítmos de classificação não supervisionada: Kmeans e MeanShift, ficando a critério escolher.