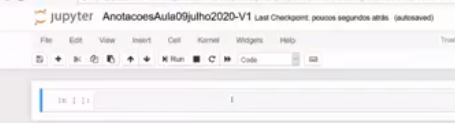
09\_07\_2020

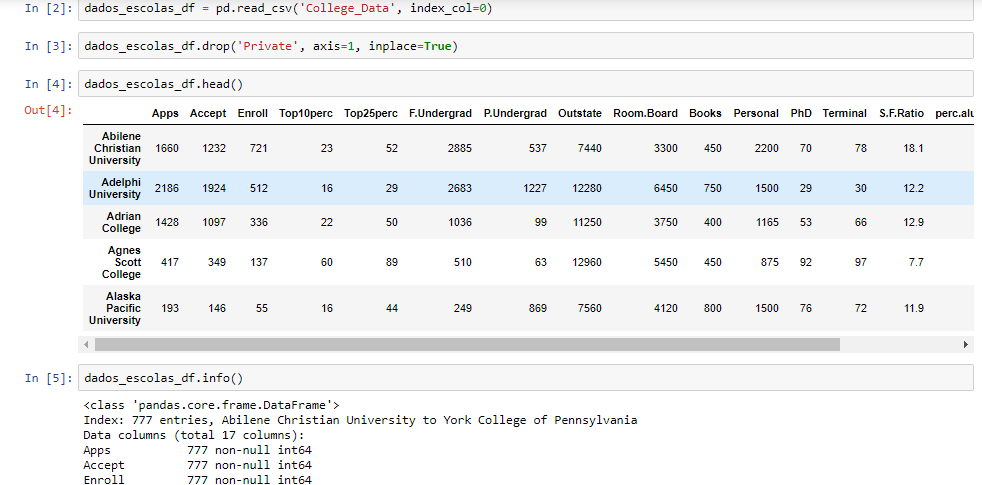
K-mean

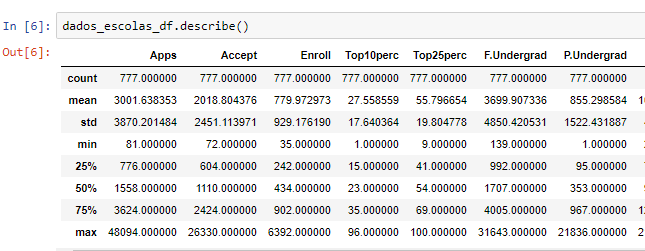


O algoritmo *k-* mean procura um número predeterminado de clusters dentro de um conjunto de dados multidimensionais não rotulados. Isso é feito usando uma concepção simples da aparência do cluster ideal:

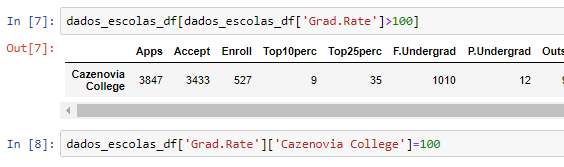
* O "centro do cluster" é a média aritmética de todos os pontos pertencentes ao cluster.
* Cada ponto está mais próximo de seu próprio centro de cluster do que de outros centros de cluster.

Explorar os dados

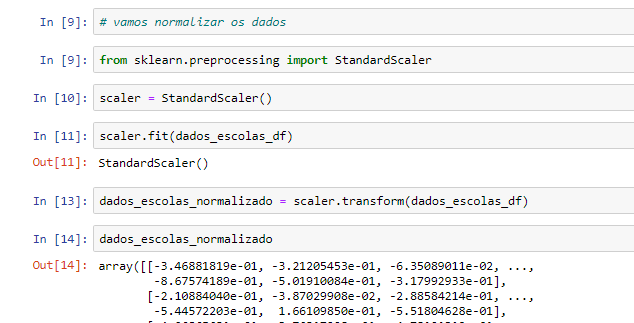




Retirar dados output



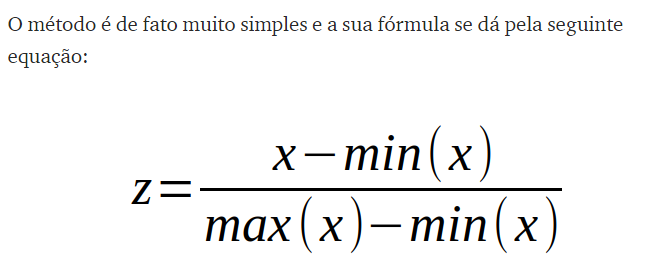
Normalizar os dados



Normalização de Dados

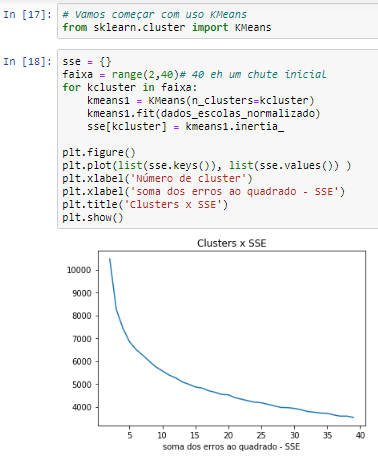
A normalização visa colocar os dados dentro de um padrão que facilite a sua análise.

A normalização de dados Min-Max é um processo de transformar valores discrepantes usando uma escala que vai de 0,0 (para o menor) e 1,0 para o maior valor. Isso facilita a comparação de valores que foram medidos usando diferentes escalas, por exemplo, Celsius e Fahrenheit ou de diferentes unidades de medida como velocidade e distância.

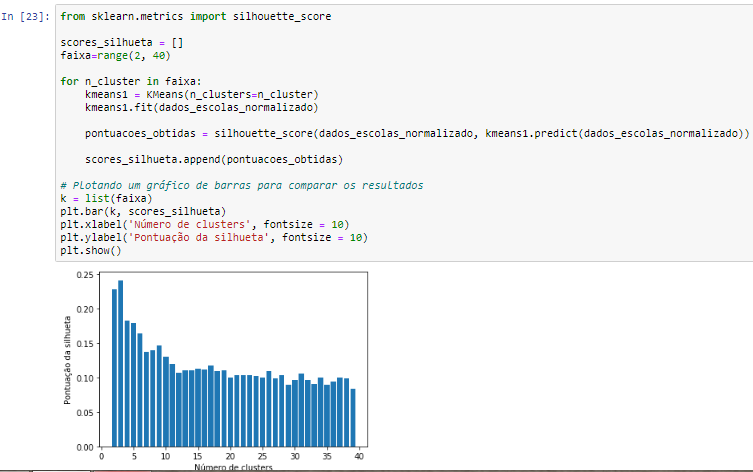


Começar a usar Kmeans

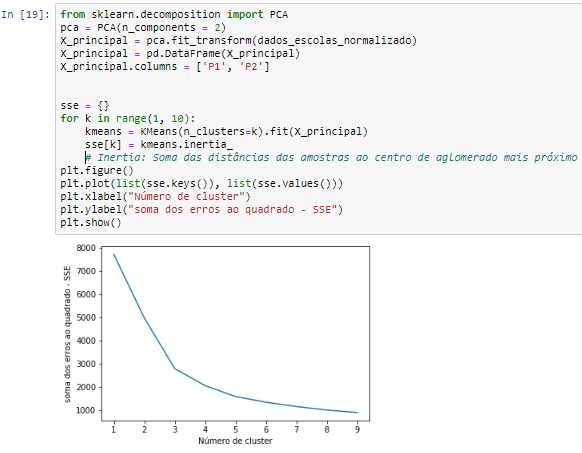
Descobrir o (K)



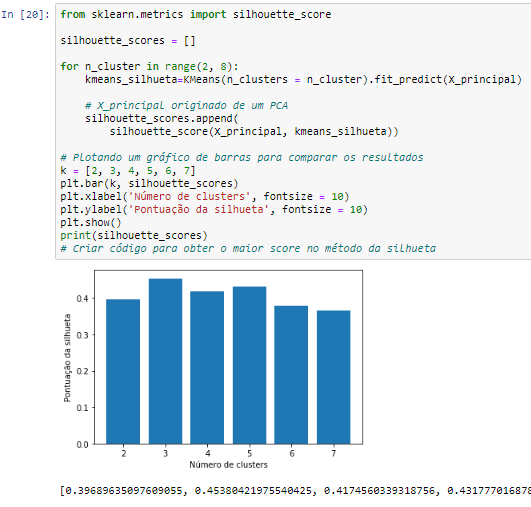
Usar a silhoueta



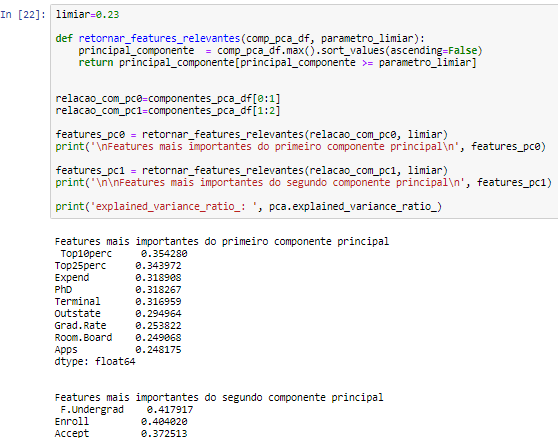
Usando o PCA



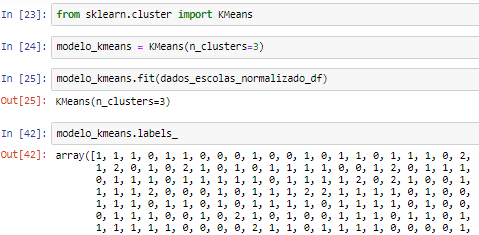
Usando a silhoueta



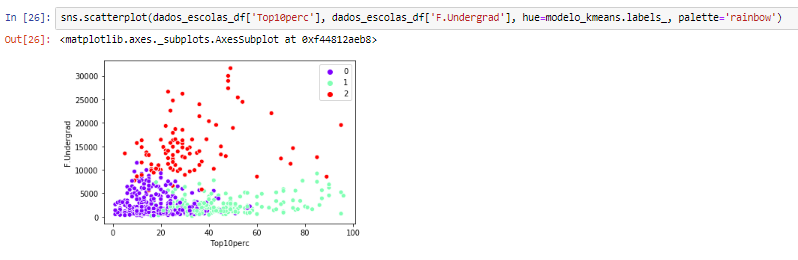
Verificando as features mais importantes



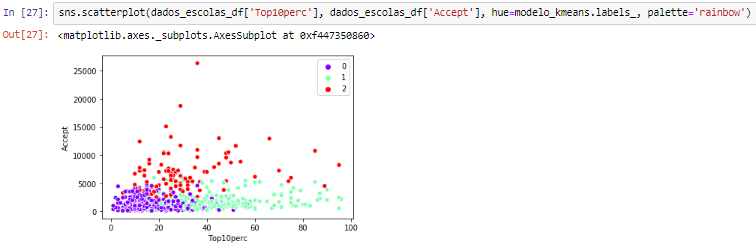
Fazendo Kmeans do clusters = 3



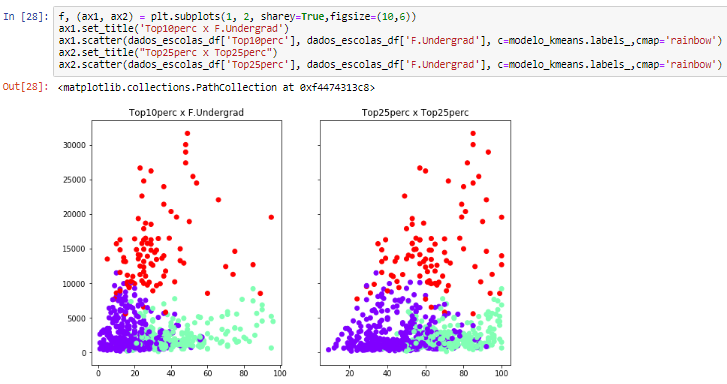
Usando scatterplot - 2 campos ( Top10perc ) e ( F.Undergrad )



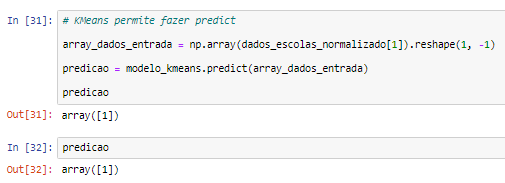
Usando scatterplot - 2 campos ( Top10perc ) e ( Accpet )



Subplot ('Top10perc x F.Undergrad') e ("Top25perc x Top25perc")



Predicao



SVM