

Miniproject 3  
BMI 555 IEE 520  
Fall 2018  
Due 11/20/2018

**Please submit through the digital drop box. After you upload you must also submit.**

Use the dataset in python

```
from sklearn import datasets  
mydata=datasets.load_diabetes()
```

**Exclude the target attribute**

You can look at the following site for examples (there are many other sites also)

<https://stackabuse.com/hierarchical-clustering-with-python-and-scikit-learn/>

1. Execute  $K$ -means clustering for 2, 3, 4, 5, 10 clusters  
Compute a measure of clustering quality and plot the measure against  $K$   
Use your plot to select a reasonable number of clusters  
What measure might you use to validate your clustering against the target attribute that was excluded?
2. Execute hierarchical clustering with single, complete, average, and Wards' distance  
Graph the dendrogram for each linkage.

Which linkage do you prefer for your results? Why?

What is a reasonable number of clusters for the selected linkage?