

## **CSE6331-Spring 2017**

### **Microsoft Azure**

#### **Clustering (Machine Learning)**

1. On the cloud provider you will use a K-Means clustering tool to find clusters of data.
2. You will obtain the dataset(s)
3. K-means has been implemented in Python modules, Weka, and R (statistics), Java, and many others.
4. Your program will allow a user to specify number of clusters (and optionally datasets, and attributes)
5. You will display centroids and number of points in clusters, and other metrics (distances between clusters, etc.)