# **Algorithm and Data Analysis**

The algorithm that we are going to use are not yet included on the 30% system requirements, so the table below provides the algorithm that we may or may not use (still not decided) on the 100% working system.

The algorithm below are based on the types of clustering algorithm:

|  |  |  |
| --- | --- | --- |
| **CLUSTERING ALGORITHM** | **DEFINITION** | **USAGE** |
| Connectivity-Based Clustering | A method of unsupervised machine learning clustering where it begins with a pre-defined top to bottom hierarchy of clusters. It then proceeds to perform a decomposition of the data objects based on this hierarchy, hence obtaining the clusters. | Since we are trying to find the similarities between many objects (like: Users behavior, business similarities, etc...), a connectivity-based cluster with a pre-defined hierarchy of clusters will be a good start. |
| Centroid-Based Clustering | Considered as one of the simplest clustering algorithms, yet the most effective way of creating clusters and assigning data points to it. The intuition behind centroid based clustering is that a cluster is characterized and represented by a central vector and data points that are in close proximity to these vectors are assigned to the respective clusters. | It is based on distance, so something like if one point is similar to the other that could be a cluster, so a centroid-based cluster is on the list of a priority when it comes to the usage of algorithm on our system. |

**Team Pythoneers**

Binuya, John Alex

Calinawan, Jimmy

Delos Reyes, John Lloyd

Hasta, Reynan