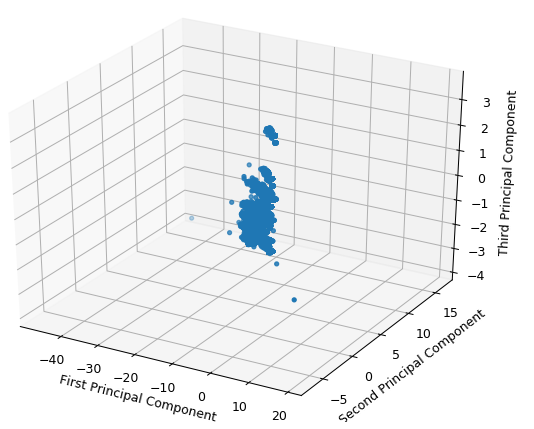
**DNSC 6279 Spring 2019 Assignment 4**

## Group 6 -- Yixuan Huang, Ada Li, Xinyu Wang, Jingsi Wu

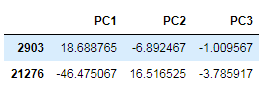
1.Your plot of the first 3 principal components with outliers (1 pt.)



**Figure 1: The first three principal components of the training data with outliers.**

Two outliers

* 'PC1' <-40
* 'PC1' > 10



2.Your plot of the first 3 principal components without outliers (1 pt.)

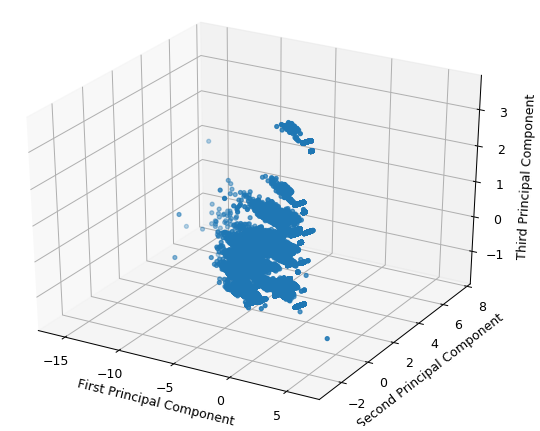


Figure 2: The first three principal components of the training data without outliers.

3.Your plot of k vs. total in-cluster error (2 pts.)

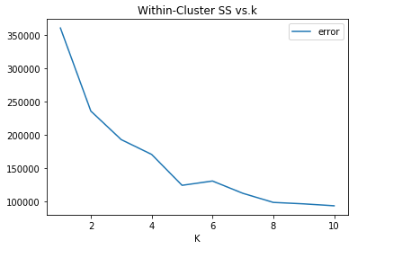
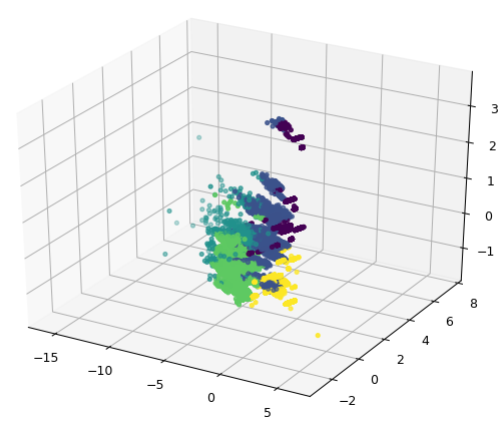


Figure 3: The number of clusters, k, vs. the total within-cluster error

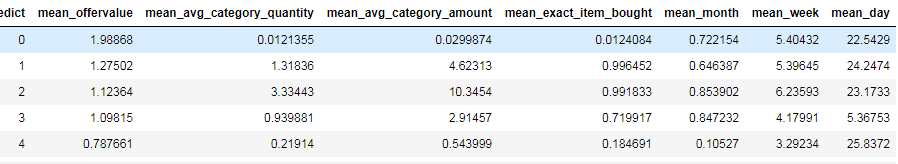
K=5

4.Your plot of the first 3 principal components without outliers colored by cluster label (2 pts.)



**Figure 4: The first three principal components of the training data without outliers colored by cluster label.**

5.Your cluster centroids and with brief profiles (2 pts.)



**Figure 5: Centroids for k-means clusters in the training data.**

#### centroids 0: Relatively have higher offer value, have lower probablility to purchase，purchased happened on medium in month week, day（latter on that month），have lower value in avg\_category\_quantity，avg\_category\_amount

#### centroids 1: Relatively have medium offer value, more likely to do purchase， with medium category\_quantity, category\_amount, purchased during medium month and week, on higher value of day（latter on that month）

#### centroids 2: Relatively have medium in offer value, have much higher avg\_category\_quantity and avg\_category\_amount，more likely to purchase during higher value of month and week on higher value of day（latter on that month）

#### centroids 3: Relatively have medium in offer value, have medium avg\_category\_quantity and avg\_category\_amount，likely to purchase，purchase happened during higher value of month and medium value of week on lower value of day（earlier on that month）

#### centroids 4: Relatively have lower in offer value, have lower avg\_category\_quantity and avg\_category\_amount，less likely to purchase，purchase happened in during lower value of month（earlier） and medium value of week on higher value of day（latter on that month）

#### Bonus: We might see that the expected effect is that it not only shows the degree the 2 outliers will increase the within-cluster error of the model，but also shows that that the difference of the within-cluster error will decrease as the centroid number (K) increase.[¶](http://localhost:8888/notebooks/Downloads/DM_hyx/Ass04_Group6_final.ipynb#We-might-see-that-the-expected-effect-is-that-it-not-only-shows-the-degree-the-2-outliers-will-increase-the-within-cluster-error-of-the-model%EF%BC%8Cbut-also-shows-that-that-the-difference-of-the-within-cluster-error-will-decrease-as-the-centroid-number-)