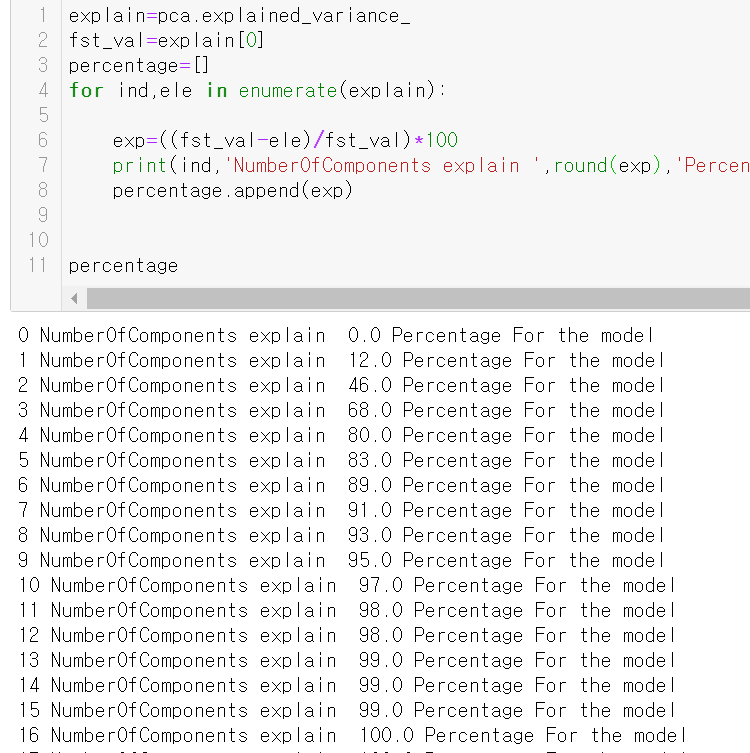


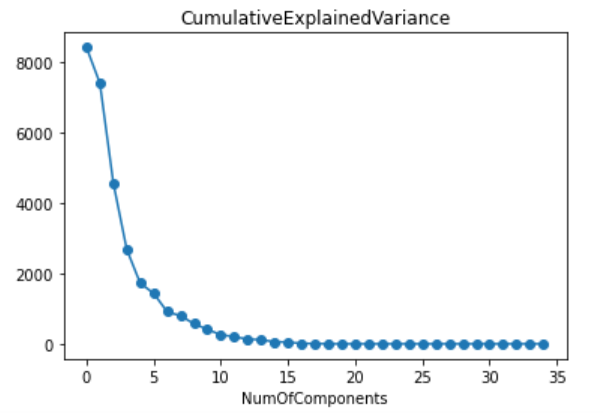
(a)

**PROCESS**

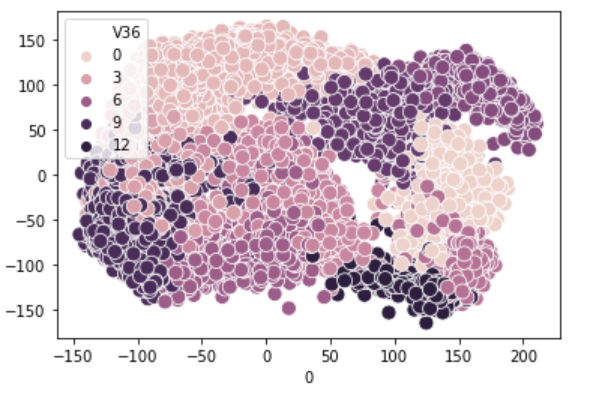
1. Split the data into test and train.
2. Fit PCA with n\_components:35
3. Get Cumulative explained varicance by the number of the components
4. Plot biplots with pc1, pc2 as well as three-dimenstional score plot with pc1, pc2 and pc3

**OUTPUT**

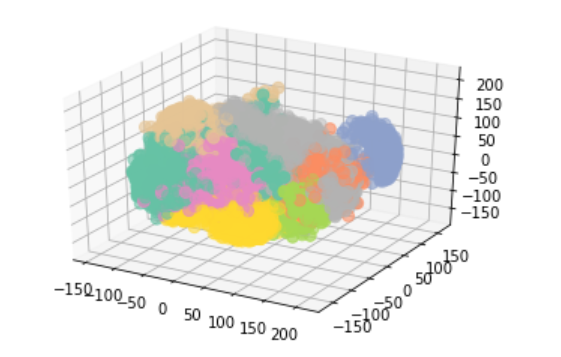




**PC1, PC2**



**PC1, PC2, PC3**

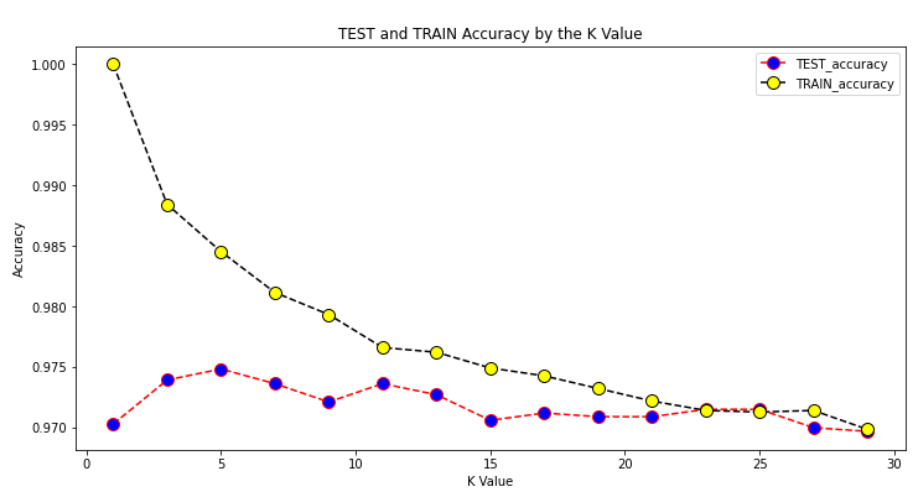


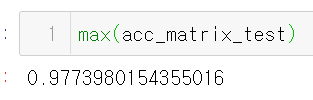
(b)

**PROCESS:**

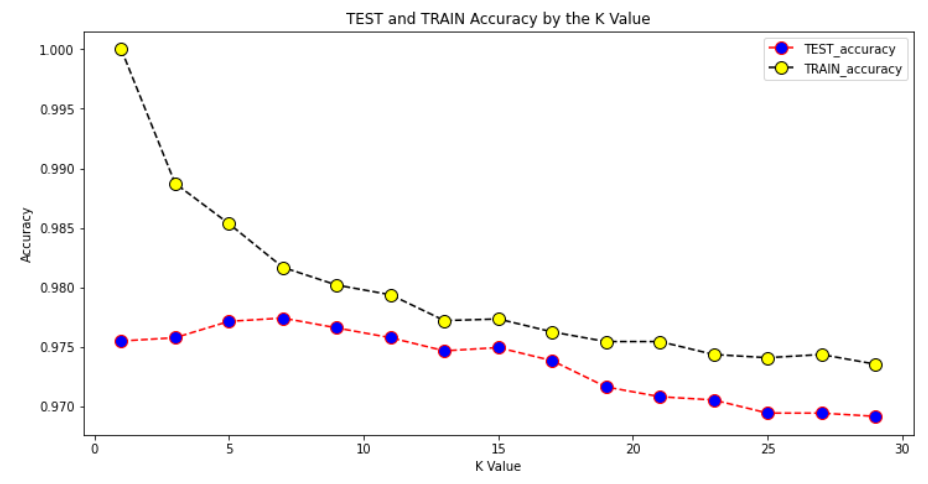
1. Split the data into test and train.
2. Fit and plot KNN **acc** with the data
3. Fit PCA with n\_components:7(90%)
4. Fit and plot KNN **acc** with the PCA data(7)

**KNN WITH ROW DATA**

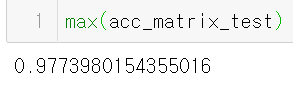




**KNN WITH PCA(n\_components = 7)**

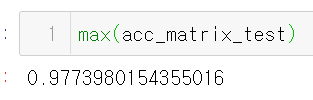


**Max test acc(k=7)**

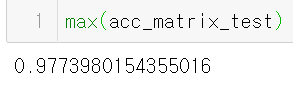


**DISSCUSSION**

ORIGINAL DATA TEST **ACC**(k:7)



PC DATA TEST **ACC**(k:7)



**Both accuracies are similar.**

**Only with 7PCs can get same performance with the model fitted by all the features. This implies that we can avoid ‘curse of dimensionality’.**