Data visualization final project

The final project divides into 4 notebooks :

First notebook : we had to return to the classification project from the previous semester and make changes over it using PCA method and other models we learn about during the year.

Second notebook : Fashion Mnist data-set. This notebook is about clothing classification. At first I did all the usual preprocessing as check for null, duplicates, show the data as DF. Then, I did PCA ,and checked the models score before and after implementation of PCA.

Third notebook : cats VS. dogs data-set. On this data set we received pictures which we needed to be transform to DF and make preprocessing as usual. Here we did not had a test group so I split the data twice so I can have validation for the final test. I used PCA on all models.

Forth notebook : hand-positioning . on this data set we received many csv of many subjects. We had to combine them all together and prepare the data for the models. At first I uploaded all subjects individually, cleaned the data, merged same frames number, dropped duplicates and NaN, and combined all for 3 DF : train, validation and final test (which is one of the subjects).

In all 4 notebooks I used the same models from sklearn :

- KNN
- Random Forest
- XGBoost
- Bagging classifier
- Voting classifier
- AdaBoost

PCA – principal component analysis :

We use PCA for making a compact model, with the highest accuracy possible. Usually for compact model the score will slightly decrease. The idea of PCA is using dimensionality reduction which allows us to get great scores with a more compact models.

Throughout the project we tried to see were and when PCA will benefit us and when the we loose to much of the score for less dimensions.