Machine learning mini project for Master Smart Mobility

- Download the Hubway data from http://hubwaydatachallenge.org/
- The idea is to extract temporal patterns from hubway bike sharing data using different machine learning approaches:

1) Unsupervised Learning

- Aggregate origin-destinations data on hourly base, so you will have all OD flows in hourly basis.
- Perform a clustering algorithm without considering the days feature, to see if you can cluster the O/D flows in two clusters: weekend and week days.
- Then run another clustering algorithm to cluster all the stations who share similar temporal patterns (ex. Early morning flow, etc)
- Try different numbers of clusters and investigate the optimal number of clusters by justifying your approach and interpreting the results.
- Try to cluster different profiles of users based on their information (age, flow locations, etc) and show how the clustering result present different types of profiles.

2) Supervised Learning

- The idea is to predict bike flow based on history of rideships.
- Split your date into Train and Test set. You can learn on 1 year data and test on 1 month. (considering the yearly growth of users)
- Feel free to add features from other sources (eg. types of days, school holidays, weather information, sociodemographic data, etc)
- Try a supervised learning model (ANN, XGBoost, SVM, etc) on your choice.
- Predict the flow for your test data and compare your results with the real data.
- Calculate the errors and explain how you can improve your results.

Presentation:

- Deadline is march 5th before midnight. Send all the reproducible codes & reports in a zip file to fereshteh.asgari@gmail.com & ons.jelassi@telecom-paristech.fr
- Prepare some slides to present your results on Wednesday March 6th (each presentation between 15-20 minutes)