ICT4SM LABORATORY 2 DESCRIPTION 2024/25

Nikhil Jha

E-mail: nikhil.jha@polito.it

Accessing the data

You will use 3 different datasets about Turin:

- Carsharing data – trips from **carshari**ng users. Stored in a MongoDB server running on the server `bidgatadb.polito.it:27017`. Two indexed collections, "ictts_PermanentBookings" for **Car2Go** and "ictts_enjoy_PermanentBookings" for **Enjoy**.
- IMQ data – survey on trips made by a sample of car users. In **dropbox** `Laboratories/ lab2/data/IMQ.zip`.
- UnipolTech data – sample of trips made by cars insured by the company Unipol. In **dropbox** `Laboratories/lab2/data/Unipoldata.zip`

Lab description

Each group has to work on the project assignment and submit a report.

- Max 6 pages which describes the findings.
- One header page with the group number, members, etc.
- Max one extra page for each optional task.
- Code, scripts, etc., must be added as separate files.

Tasks

Focus on Torino, apply the filters you used in Lab 1. Process it to further analyse it by producing the following plots and results:

1. For carsharing data, check if the booking records requires to be filtered to get valid rentals. Derive the OD matrix based on the zones of Torino. **Remember to normalize the matrix.** Plot the results using heatmaps.
   a. Compute OD matrices on different periods (weekdays vs weekends, daytime vs night-time). Is it possible to spot any difference?
   b. Consider the same period, compute OD matrices on different dates (e.g., OD matrices of weekdays in two different weeks). is it possible to spot any difference? How does the different compare with different periods?
   c. Consider the same period, compute OD matrices on same date for Car2GO and Enjoy. Is there any difference between different platforms? How does it compare with previous ones?
2. Derive metrics to compute the similarity between OD matrices. Test different metrics under different scenarios (e.g., the comparison in task 1, compare two random matrices, compare two completely dissimilar matrices…), report the numerical result. Choose the most appropriate metric and explain it. (1-b and 1-c with quantitative metrics)
3. Compute OD matrices using UnipolTech dataset and IMQ dataset and estimate the similarity using the metric you derived in task 2.
   a. Compare results from these 2 datasets. Do people with similar profiles (gender/ age/commercial for UnipolTech and age/purpose/… for IMQ) have similar mobility behaviours?
   b. Consider IMQ dataset, what gender of user has greater differences in behaviour across age groups?
   c. Consider IMQ dataset, what kind of users are more likely to use Car2GO and Enjoy?

      d. Compare the OD matrices from the 3 datasets overall.
4. (Optional) Consider carsharing and UnipolTech data, are there better solutions to divide the city area in terms of mobility analytics rather than using these zones?
5. (Optional) For UnipolTech and for CarSharing estimate continuous density for trip generation and trip distribution, then find a way to compare them.