BIG DATA IN STATISTICS ON PASSENGERS TRANSPORT – A CASE STUDY ON LISBON METROPOLITAN AREA

Name(s) of author(s):

Rute Cruz Calheiros, Porfírio Leitão

Statistics Portugal (PORTUGAL)

rute.cruz@ine.pt, porfirio.leitao@ine.pt

Organization: Statistics Portugal

Abstract:

The Lisbon Metropolitan Area corresponds to 3.3% of the national territory surface and comprises 18 municipalities, which are distributed in equal numbers between the north and south sides of the Tagus River.

The resident population in this region (NUTS II) amounts to a total of 2.8 million people, corresponding to 27% of the population in Portugal.

Given the high population density and intensive economic activity in the region, passengers transport systems play a crucial role to day-to-day citizen's life and businesses. Thus, the Lisbon M.A. transport system is a responsibility of an administrative organization that goes by the same name: the Lisbon Metropolitan Area, bringing together the common interests of the municipalities around the capital of Portugal.

The most important public transportation operators in the region are organized in the form of a consortium, the OTLIS, with a view, among other things, on the relationship between companies in the management of the ticketing system and passes from multi companies.

The ticketing system works with contactless cards, which can be charged with single tickets, sets of tickets, one company passes or multi companies passes.

Each time a card is used in the system, either at the entrance or exit, a record in the database is generated, which is composed by card ID, transaction type (in/out), day/hour, transport operator and local.

The transport system databases compile an enormous amount of relevant information for the knowledge of commuting, allowing an accurate perspective not only of the utilization level of each mode of transport but also the variability throughout the day and especially the construction of origin/destination matrices.

The follow-up of each card ID also allows the construction of daily trips by single stages concatenation, showing how users choose to combine the various modes of transport at their disposal.

The challenges faced in the treatment of this huge set of information relate not only to the need to design models to build complete trips, but also to the management of extraordinary volumes of information, which continuously feed the system and configure the profile of Big Data.

Statistics Portugal has carried out a case study concerning a part of these databases with the aim to assess the relevance of their use for the production of passengers transport official statistics, applied to the geographical area concerned, the Lisbon Metropolitan Area.