

EUSTAT. Basque Statistical Institute

Jorge Aramendi Rique j-aramendi@eustat.eus
Javier San Vicente Alfambra j-sanvicente@eustat.eus

Automatic hotel prices collection on the Internet for the Tourism Survey in the Basque Country

Keywords: Big Data, prices, hotel room prices, web scraping

EUSTAT, aware that Big Data is an interesting source of information for statistical offices, has carried out a pilot study for the study of the daily series of hotel room prices and its possible use in the Survey of Tourist Establishments Receivers (ETR).

The objective was to obtain the average daily prices (ADR – Average Daily Rate) of hotel establishments, and the unit of analysis is the standard double room with bathroom, without breakfast and without VAT.

We decided to select only one of the websites that offer room reservations, Booking.com, estimating that it would be enough to highlight the possible outcomes derived from both the computer tool and the method to calculate the average daily prices.

We used a web scraping tool to extract information from websites, *import.io*, a tool that allows creating scraping templates adapted to a particular website to capture the information.

A key element to determine the advantage of using Big Data to estimate hotel prices for the Tourist Establishments Receivers survey is the high coverage ratio, which is around 88% for bed places, with respect to the Eustat Tourism Register.

The survey questionnaire includes the corresponding question about the average daily rate applied to a double room (ADR), distinguishing nine different types of clients, including clients who have booked the room through the “Online Tour operator”. This category of customer, and its ADR, would, therefore, be comparable to the average price obtained on Booking.com.

Prices analyzed in Booking practically do not present differences according to the term of the consultation, 7 or 14 days in advance, and on average, prices of survey for the type “Online Tour operator” are lower than the result of Booking.com.