New data source for calculating ADR
(average daily rate per occupied room)

Introduction

Mindful of the vast opportunities for obtaining information through the use of the various technologies that Big Data encompasses, EUSTAT is working on different projects to incorporate data obtained through these techniques in its estimation processes. Information collection procedures and their use in official statistics are being studied both as a single data source and in combination with data from sample and/or administrative sources.

In this context, a project was launched in 2017 to obtain the average daily rate per occupied room (ADR) for the Survey on Tourist Establishments, using Big Data methodology.

From October 2020, the data published for this variable (ADR) and the derived variable, revenue per available room (RevPAR), originate from the capture of information available on the internet via online booking platforms for hotel rooms.

Web scraping is done in accordance with the Eurostat's recommendations regarding legal safeguards, competence, data protection and statistical confidentiality.

Modelling the ADR

For the final estimate of the ADR of hotel establishments in the Basque Country, the following statistical model was defined, which combines information obtained from online booking platforms using web scraping techniques, information in the survey directory, and information obtained from surveys of the establishments themselves:

$$ADR = \beta_0 + \beta_1\,Median + \beta_2\,Max + \beta_3\,Min + \beta_4\,Category$$

$$+\beta_5\,Stratum + \beta_6\,OnlineBooking + \beta_7\,Occupation + \beta_8\,Size + \varepsilon$$

where the explanatory variables are given by:

- ***Median***, ***Max*** and ***Min*** are the median, maximum and minimum daily prices captured on the internet over the previous 4 months
- ***Category***: 5 levels are considered: 4 and 5 stars, 3 stars, 2 stars, 1 star and guesthouses.
- ***Stratum***: geographical area in which the establishment is located
- ***OnlineBooking:*** a binary variable indicating the presence or otherwise of the establishment on online room booking platforms.
- ***Occupation***: the occupation rate per room at an establishment during the month of reference
- ***Size***: size range according to the number of rooms at the establishment (<9, 9-15, 16-39, > 40 rooms).

Information is obtained from online booking platforms for each hotel establishment using web scraping once a day for each of the previous 120 days, so that there are 120 prices per booking platform for each establishment and for each day. Therefore, a total of 3,720 different prices would be obtained in a 31-day month for each establishment and platform.

The ADR obtained based on the survey information has been compared to the ADR estimate using different data capture models and procedures for a two-year period, and the most satisfactory is the model proposed above.

Both the information capture and the estimation procedure are incorporated into the processes for producing the Eustat Survey on Tourist Establishments.