Online Jobs Advertisements (OJAs) Experimental Statistics: Bulgaria

Aims

Bulgarian National Statistical Institute (BNSI) is interested in producing experimental statistical information about the OJAs as new statistics derived from the big data sources and as complementary statistics to the regular JVS as well.

In order to be methodologically correct, the following facts and assumptions must be taken into account when producing the experimental OJAs statistics:

- The traditional JVS takes account of job vacancies declared by employers at the end of each quarter (according to a European Regulation);
- On-line job vacancy (OJVs) is a different concept from OJAs. Although an OJA may contain 0, 1 or more OJVs, here it is assumed an OJA contains exactly one OJV. In this case, the two concepts could be used as synonyms;
- The distribution by characteristics covers only those OJAs, where the relevant characteristics were found during the scraping.

Data Sources

BNSI used the extracted OJAs from the two biggest Internet job-portals in Bulgaria: www.jobs.bg and www.zaplata.bg to produce experimental statistical.

These two jobs portals publish daily almost 95% of the total number of OJAs from all job portals in Bulgaria and this makes them sufficiently representative to produce OJAs experimental statistics. Therefore, the BNSI decided not to scrape the other small jobs portals to save effort and resources.

For the purposes of analyzing the COVID-19 situation, the administrative daily data on registered unemployment has been received from Employment Agency.

Methodology

The monthly OJAs indicator is based on a count of OJAs newly daily published on www.jobs.bg and www.zaplata.bg during the month. Duplicate advertisements are removed before the OJAs are coded based on the NUTS 3 and NACE Rev.2 classifications.

The OJAs indicators do not reflect the total number of job advertisements on the Bulgarian labor market as it does not include jobs advertised through other channels – enterprises websites, social media, newspapers or through informal communications such as word of mouth. In addition, it is possible that some vacant job positions are not advertised at all.

The experimental OJAs monthly and quarterly statistics have been produced regularly for the period July 2019 – March 2020 according to different characteristics, e.g. NUTS 3 level, NACE categories, educational level and etc.

During the uncertain COVID-19 situation, the detailed weekly OJAs statistics have been produced from the beginning of March 2020 – till 19th April 2020 to reflect changes and lockdown effect on the

numbers of the published OJAs in Bulgaria. This is a clear example how OJAs data could be used efficiently as a complementary and timely information for labour insights.

The used software is Python.

BNSI is doing OJAs experimental statistics in four main steps, which are performed in daily, weekly monthly or quarterly basis. The steps are executed automatically with scheduled Python scripts.

Step 1: Scraping

BNSI is doing scraping of OJAs on daily basis. During the scraping, we are saving OJAs that are published only yesterday, i.e. the day before the real scraping. We are using Python's Scrapy library with custom written spiders. They are using configuration files in JSON format with instructions what to be scraped and from where. A log file is written during the scraping. The spiders are saving the scraped data in CSV files by days. The CSV files have 26 columns of semi-structured information.

Step 2: Validation

BNSI is doing validation step daily on the scraped CSV file with yesterday OJAs data by Python script. The validation step is checking whether the obtained OJAs records are well structured, i.e. dose the column contains expected data or not. In addition, the script is transforming date column in the same pattern. Then the script is saving the OJA records without errors in CSV files by days.

Step 3: Encoding

BNSI is doing encoding step daily on the all validated CSV files with OJAs data by Python script. First, the script de duplicates all records by all columns and keeps only the first record. Then the script de duplicates all records by date, OJA publisher (employer), populated place, country and OJA title and keeps only the first record. Then the script encodes the OJA by the following characteristics:

- Educational levels;
- Permanent or temporary work;
- Full or part time work;
- Kind of payment;
- Currency;
- Country codes;
- NUTS 3 levels – from populated places classification and OJA populated place;
- Enterprise's ID code (Statistical Business Register code) – full string comparison of OJA publisher and SBR enterprise name;
- NACE rev.2 codes – from enterprise's identification code and SBR NACE information.

The script is saving the encoded OJAs data in two CSV files:

- All OJAs records;
- OJAs records placed in Bulgaria.

Step 4: Experimental statistics and statistical indicators

BNSI is doing experimental statistics on weekly, monthly and quarterly basis with OJA records placed in Bulgaria by Python script. These are the indicators:

- Number of OJAs and change by weeks;

- Number of OJAs by educational level of and change by weeks;
- Number of OJAs by full and part time work and change by weeks;
- Number of OJAs by NACE Level 1 and change by weeks;
- Number of OJAs by NUTS 3 and change by weeks;
- Number of OJAs and change by months;
- Number of OJAs by educational level and change by months;
- Number of OJAs by full and part time work and change by months;
- Number of OJAs by NACE Level 1 and change by months;
- Number of OJAs by NUTS 3 and change by months;
- Ratio of OJAs to registered unemployed by weeks

Results

The experimental data for OJAs are available on the WPB wiki page:

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WPB_Experimental_statistics