International Forum on Migration Statistics
IFMS 2023
24-26 January • Santiago, Chile

Exploring the usefulness of web scraping news and social media data for supplementing migration statistics in Mexico

Alejandra Figueroa, Víctor Silva, Olinca Páez & Elio Villaseñor.*
Instituto Nacional de Estadística y Geografía (INEGI)

Methodology

We developed an intensity indicator for international migrant presence in the Mexican territory based on the compilation of news (2020-2022) through web scraping techniques and data mining procedures to identify the municipalities mentioned in the text.

We also used data from Facebook´s Marketing API to detect people in transit in municipalities where the Population Census had identified lodgments for migrants.

Data analysis

The counting of news was mapped to visualize the territorial and temporal pattern of the migratory flow and to be compared to a similar indicator from a traditional source: the records of the authority on foreigners in irregular status.

Findings and conclusions

1) The news-based intensity indicator suggests a greater territorial dispersion of the migrant transit, in contrast to the pattern shown by the records of the authority, which points to a greater concentration of the migratory flow in the eastern part of the country.

2) Our indicator reflects the public perception of the migratory intensity and its evolution over time, and it is independent of the spatial distribution of infrastructure related to migration control or assistance.

3) Data from Facebook´s Marketing API based on the municipalities with migrant shelters does not correlate with the information in the registries, nor with the intensity indicator based on web scraping of news.

Insights of mobility through Mexico from two sources.

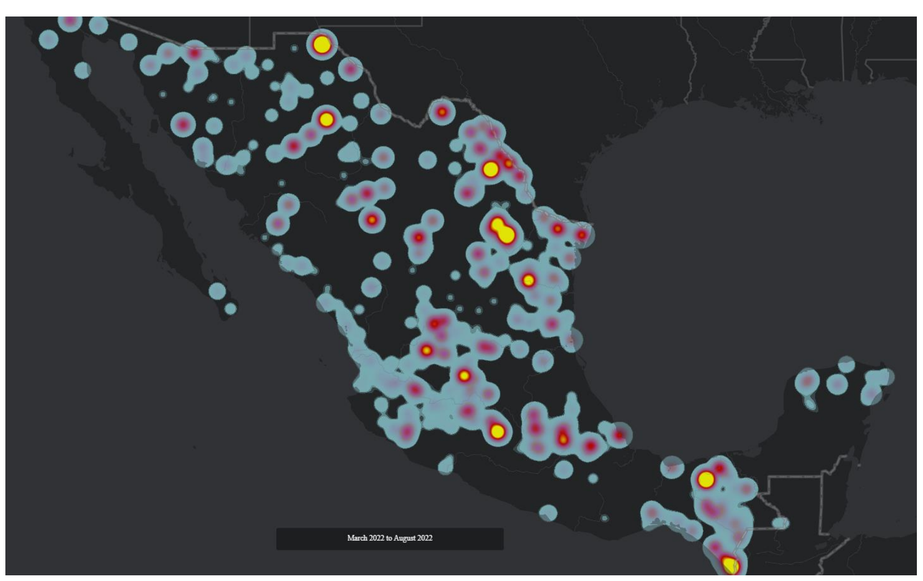

Fig. 1 Counting of news with the word “migrant” geolocated after the mention of a municipality (March-August 2022).

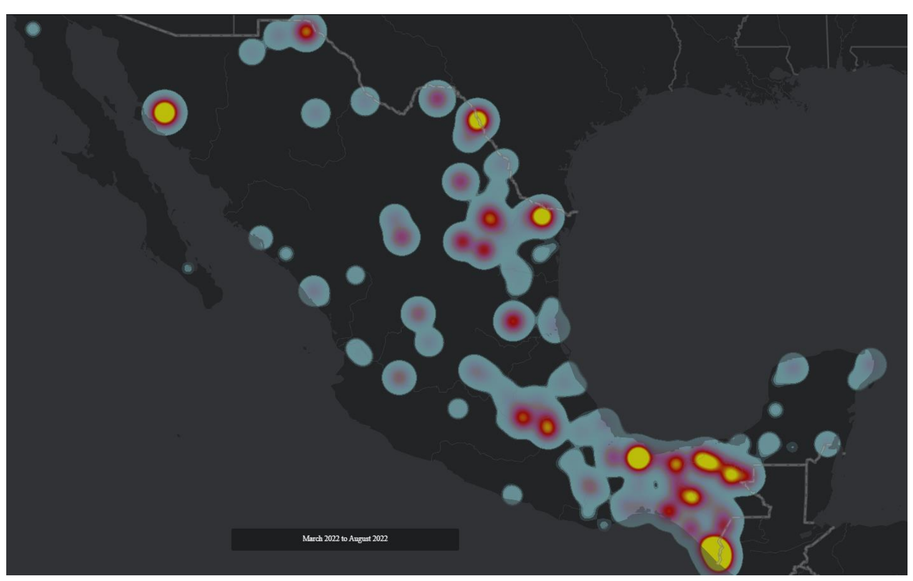

Fig. 2 Foreigners in irregular status (Events March-August 2022).

Contact information: olinca.paez@inegi.org.mx