OFFICIAL STATISTICS

Measuring tourism using mobile network data

Belén González Olmos[1], María Velasco Gimeno[2]

[1] National Statistics Institute (Spain), belen.gonzalez.olmos@ine.es
[2] National Statistics Institute (Spain), maria.velasco.gimeno@ine.es

---

Abstract: IIn Spain, Tourism basic statistics are responsibility of the National Statistical Institute (INE) and traditionally are based on surveys. In recent years, due to the challenges associated with the collection of data from persons, especially during the COVID-19 pandemic, national statistics offices have explored the access to data generated by private sector using two different approaches: based on either a specific arrangement or taking advantage of a regulatory framework. In this article, the Spanish experience in the use of mobile phones positioning data is explained.

INE uses mobile phone positioning data as auxiliary information to tourism surveys with the objective of improving the geographical breakdown of tourism figures. The use of this source of information allows obtaining new products with a granularity in terms of origin/destination of tourists that would be impossible to achieve using traditional techniques, without increasing the cost of the statistics and the burden on the informant. Nonetheless, it poses some challenges in terms of quality assurance and sustainable access. The results are published as experimental statistics, but the ultimate aim is to integrate them with traditional tourism surveys.

---

1 Introduction

The production model of official statistics in the near future has to be adapted to the new situations of competition, data availability and user requirements. Improvements in the production model have a direct impact on improving the quality of statistical products of all kinds. Currently, the national statistical system is still mainly based on traditional surveys; however, the use of administrative registers and the use of Big Data in statistical production have been incorporated into the various operations for several years now.

The use of new available data sources (private databases, Big Data, ...) in the compilation of official statistics is a path that has already been undertaken. The use of these new sources of information opens up new possibilities to compile statistics more quickly, with a greater geographical and functional disaggregation and to address the study of emerging phenomena in a shorter period of time. Moreover, they represent a fundamental way to reduce the response burden on informants, as they are based on existing and available information.

Given the need to study in depth the quality, procedures, applicable statistical techniques, etc., of these sources, the strategy followed by the European Statistical System (ESS), which has already been joined by different national statistical offices such as the INE, as well as Eurostat, is to disseminate these operations under the name of experimental statistics. The idea is to be able to test and try out these new data sources without the restrictions and limitations imposed by official statistics. In this way, the organization is acquiring the necessary knowledge to make these statistics official in the near future.

Experimental statistics use new data sources and methods in an effort to better respond to the needs of our users in a timely manner. The contents they present are considered experimental because they have not yet reached sufficient maturity in terms of reliability, stability or data quality to be included in official statistics. Nevertheless, the available results are offered to users for their use and evaluation, due to the relevance they may have and as a means to improve the products themselves by gathering the opinion of the final users of the information.

Until an experimental statistic has reached sufficient stability and maturity, it will not be proposed for inclusion as official, and therefore will not be included in the National Statistical Plan.

In the case of mobile network data, the main problem countries are encountering is access to those data. They are strongly protected by national and international legal regulations, due to highly sensitive issues of confidentiality and personal privacy. In addition, there are economic interests to commercially exploit them by the phone companies, so access to the data is the first major problem to be addressed. The objective is, in the short term, to have access to specific datasets for use in research activities and, in the long term, to investigate the feasibility of sustained access under standard production conditions, as well as the required characteristics of such data (what information it needs to contain).

In relation to access, there is no guiding principle or golden rule for success, as the situation is markedly different per country (different legal regulations, different contents in the datasets offered) and per mobile network operator (different company structures and different commercial interests), and the role that national data protection agencies can play is important.

The generation of mobile network data has one characteristic that specifically affects processing: the data are not generated with a specific metadata structure for statistical purposes. Moreover, it is characterized by the fact that the data does not contain information from the data provider but from third parties (the customers) and that it is information that plays a central role in the provider's business.

These data (as with many other Big Data sources) have been generated for purposes very different from statistical production, even before their potential use for statistical purposes has been identified. Therefore, there is no proper metadata structure included in the generation process.

The raw data are extremely technical and some of them are only stored temporarily, which makes pre-processing necessary to generate exploitable data for statistical purposes (microdata).

Statistical microdata can be analysed in many different ways and the purpose for which they are used will condition the aggregation process required. The experience described above has been oriented towards a particular type of aggregated data: those that provide counts of individuals from

a given target population (general population, tourists, resident tourists, travelers, ...) by territorial cells and time interval.

Aggregate data must be somehow linked to the target population, i.e. an inference exercise must be carried out, given that NIS only access data from some telephone companies. The problem is that, because of the way the data have been generated and collected, there is a need for a new methodology because the traditional sample design methodology is no longer applicable.
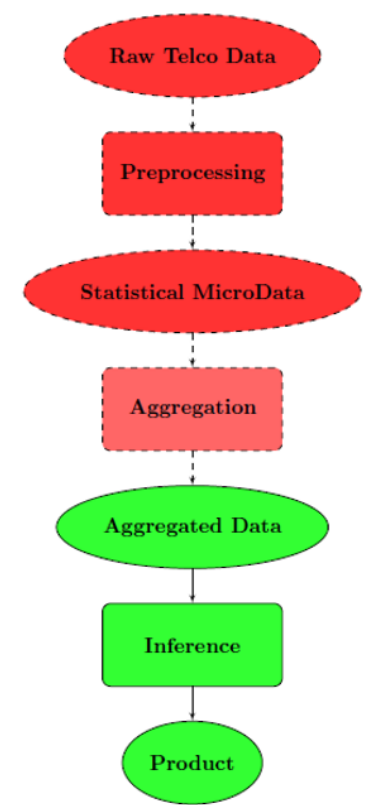

Figure 1: Mobile network data process in mobile phone companies.

The process can be represented as in the Figure 1 (the oval elements indicate data sets and the rectangular ones indicate steps in the process; in green the phases in which the National Statistical Institutes have access to the data -the three at the bottom- and in red those that are exclusively accessed by the source -the three at the top-, although the aggregation phase will depend on each specific case).

As can be seen, there is no access to the raw telecommunications data originating in the network, nor even to the pre-processing to produce statistical microdata or even to the next stage of aggregation.

The contents of this paper are the following. In Section 2 the experimental statistic to measure tourism flows with mobile phone positioning data is presented. Section 3 goes into more detail on methodological aspects such as the definitions and processing data. Section 4 presents how the results are disseminated in INE website with maps, and some examples of the usefulness of this information. Finally in Section 5 some conclusions are included.

2 Experimental statistics: Measurement of tourism using mobile network data

Since 2015, the Spanish National Statistics Institute (INE) has been responsible for the statistical operation Resident Tourism Survey (RTS) and the Border Tourism Movements Survey (FRONTUR). Both operations are based, in a large part of their elaboration, on surveys aimed at individuals and households. While the main objective of the former is to estimate the number of trips made by Spanish residents both within and outside Spain, that of the latter is to estimate trips made by non-residents in Spanish territory.

In both cases the cost of collecting questionnaires is high and the results obtained do not provide a good geographical granularity (Autonomous Community at most) because they do not have the necessary sample support. Regarding data availability, FRONTUR disseminates its results around one month after the end of the reference month. The ETR, on the other hand, publishes its results quarterly three months after the end of the reference quarter, which significantly reduces its timeliness. For example, data for the summer season are released around Christmas, three months after the end of the season.

Fortunately, technology is evolving, and many tools or devices have entered the daily lives of citizens. This fact, combined with the decreasing prices of using these devices and the continuous growth of the capacity to process and analyse the immense volumes of data (Big Data) is creating a whole new range of data sources that cannot be ignored by the official statistics system.

Due to this, INE, in contact with the three large mobile companies in Spain, has developed a project for the exploitation of aggregated mobile telephone data from which the movements of resident and foreign tourists can be known, breaking down the information by Autonomous Community, provinces and municipalities through which their trips take place. In the same way, the countries to which tourists resident in Spain travel to when they go abroad are also known.

The complexity of the data captured by an antenna requires specialised processing to transform them into a set of information valid for statistical processing. This process has been carried out by the three mobile telephone operators, through their algorithms. This work has required continuous interaction by the INE over two years with the three operators for the definition of the algorithms, the detailed analysis of the results, the detection of systematic differences and the drawing up of conclusions that have enabled the transformation of mobile data to be adapted to international definitions of tourism.

The improvements seen in this project compared to traditional surveys are:

- The pool of individuals available to the operators is much larger than the traditional survey samples. Considering that a large percentage of the population has a mobile phone and that the main operators have around 25% market share each, the sample of each of them can be around 3 orders of magnitude larger than that of the surveys.
- With a much larger number of individuals for whom travel information is available, the geographical breakdowns provided are much broader. Traditional surveys provide information for both resident tourism (ETR) and non-resident tourism (FRONTUR) at Autonomous Community level. With this new source, information is provided up to municipal level.
- In the case of resident tourism, the temporal availability improves considerably. The ETR publishes quarterly data three months after the end of the reference quarter, which means that the data for the first month of the published quarter is published with a time lag of 5 months. With this study, this information is published with only a one-month lag.

- With respect to outbound tourism, the ETR provides annual information for the four main countries receiving resident tourism, as well as for groupings of these countries. Through this study, monthly information is provided for all countries with a minimum number of trips.
- Regarding inbound tourism, FRONTUR provides monthly information for a few countries individually and for groupings of these. With this study, monthly information is given for practically all countries (if information is available for a minimum number of trips).
- Subjectivity and errors that may be introduced by informants (actively or passively) when providing information in the questionnaire are eliminated.

On the other hand, as mentioned above, this data source does not allow us to obtain qualitative variables such as type of accommodation used, reason for the trip, form of organization or mode of booking, so it is still necessary to maintain some kind of field operation to obtain them.

Due to the wide coverage of this project, which includes domestic, inbound and outbound tourism, 3 independent experimental statistics are carried out. The first publication was made in May 2022, including data from July 2019. Since then, monthly publications are made, around 35 days after the reference month.

- Inbound tourism (Spain, 2022b): The number of trips, overnight stays and the corresponding average stay of tourists coming to Spain from any country are published. They are provided by municipalities, provinces and Autonomous Communities. This is a great improvement in terms of the geographical breakdown provided by FRONTUR.
- Outbound tourism (Spain, 2022c): The number of trips made by residents in Spain to foreign countries is published, as well as the associated overnight stays and average duration. This is broken down by municipalities, provinces and Autonomous Communities of residence, and for all countries to which residents have made trips. It offers an improvement over the ETR in terms of both geographical breakdown and timeliness.
- Domestic tourism (Spain, 2022a): This is published for residents in Spain, the number of trips outside the province of residence, as well as their overnight stays and average duration. The geographical breakdown level, both by origin and destination, is Autonomous Community, province and municipality. As with outbound tourism, it offers an improvement over the ETR in terms of both geographical breakdown and timeliness.

The study period for these three operations is the month. In the first dissemination, the months from July 2019 to April 2022 were published. Since then, they have been updated in monthly basis.

In parallel to these publications, INE is working to integrate this information with the ETR and FRONTUR, which would be the final objective of these experimental statistics. In this way, it will be possible to provide leading indicators that will make it possible to provide travel information much earlier than in the current deadlines. Special improvement will be obtained in the data on domestic and foreign trips (ETR) as the information will be published in less than a month and without waiting the current 3-5 months (depending on the month) for the information to be available. On the other hand, this integration will bring granularity to the surveys, as well as a reduction in the size of the samples and questionnaires, which will reduce the burden on the respondent, as well as the cost of the operations.

3   Methodology

3.1   Adjustment of definitions

For the development of this project, the definitions and concepts of traditional surveys, which follow international methodologies and standards, have been adjusted and adapted to the information available to mobile companies. By way of example, the adaptation of the main concepts is shown below:

- Trip in the field of tourism statistics (UNWTO, 2008): Tourism trips are all trips with a main destination outside the person's place of usual residence, involving an overnight stay outside the place of usual residence and lasting less than one year, provided that the main purpose of the trip, including business, leisure or other personal reasons, is other than employment in a company established in the place visited. They are outward and return journeys and end when the person returns to his/her place of usual residence.
- Journey adapted to mobile network data: A tourism trip is considered to have taken place when a mobile phone has been detected for a longer period between 22:00 and 06:00 in a municipality or country other than that of usual residence and, in addition, has also been captured the following day (from 06:00 onwards) in that municipality. The journey ends when the mobile phone is again detected for a longer period between 22:00 and 06:00 in the municipality or country of residence.
- Overnight stay in the field of tourism statistics: Number of consecutive nights that a person spends in a municipality or country other than that of residence as part of a trip.
- Overnight stay adapted to mobile network data: Number of consecutive nights in which a mobile phone has been detected longer between 22:00 and 06:00 in a municipality or country other than the municipality or country of residence and has also been captured the following day (from 06:00 onwards) in that municipality or country.
- Main destination of the trip in the field of tourism statistics: This is the place where the respondent has spent the greatest number of nights.
- Main destination adapted to mobile network data: This is the place where a mobile phone has made the most overnight stays as part of a trip (according to the definition).

3.2   Data collection and integration

INE receives the tabulated and aggregated data prepared by each company. It does not have individual device information, only receives the aggregated information provided by the mobile telephone operators. It receives both raw data and data aggregated to the population.

Depending on the type of tourism (inbound, outbound or domestic), the operators' files are processed and integrated in different ways. To mention one of them as example, the procedure for outbound tourism is detailed below.

The first step consists of processing the operators' files. The files sent to INE monthly contain information on trips and overnight stays abroad, by Autonomous Community, province and municipality of origin, made by resident tourists. INE carries out a prior filtering of these files, in order to suitably adapt their format before processing the information they contain.

The second step is the estimation of the totals by country. For this purpose, the trips to the corresponding country provided by the three operators in the Autonomous Community files are added up. Since the three operators do not cover 100% of mobile telephony users in Spain, correction factors are estimated to bring this sum to the total population. These factors vary according to the

number of operators providing data for each country and are estimated by quarter using market share data from the CNMC (Spanish National Markets and Competition Commission). Similarly, total overnight stays are estimated for each country. Average durations are calculated as the quotient of trips and overnight stays for each country estimated independently.

Finally, these totals must be distributed by Autonomous Community, province and municipality of origin. This process is followed for the Autonomous Communities (ccaa, from now on):

- For each country, the percentage of trips (and overnight stays) from each autonomous community is determined for each of the operators.
- For each ccaa-country crossing, the average of the three percentages obtained in the previous point is calculated.
- The averages calculated are adjusted so that their total sum per country is 100%.
- These percentages are applied to the estimate of trips (and overnight stays) for each country, thus distributing them by ccaa.

The distribution of totals by province and municipality is done in a similar way.

More information on the different methodologies for each type of tourism (domestic, outbound and inbound) can be found in the technical projects on the website: https://www.ine.es/experimental/turismo_moviles/experimental_turismo_moviles.htm.

4   Dissemination of results

The variables published are the number of tourists and the overnight stays and average duration associated with their trips.

The geographical breakdown variables are countries, Autonomous Communities, provinces and municipalities. The time disaggregation variable is the month.

For the publication of the results, tables are used, where you can select the different variables that you want to consult, as well as infographics that allow you to select the countries, ccaa, provinces and municipalities on different maps, to evaluate the complete series or the data on trips in a specific month.

In the case of outbound tourism, the following maps are presented in the infographic:

- World map with trips in each month to each destination country (see Figure 2).
- Map and line graph for each continent (see Figure 3).
- Map and graph for Autonomous Communities, provinces and municipalities. When an area is selected, the 10 most visited countries are displayed in the graph (see Figure 4).

Similarly for inbound tourism, the following maps are available:

- World map with monthly trips from each country (see Figure 5):
- Map and line graph from each continent (see Figure 6).
- Map and graph for Autonomous Regions, provinces and municipalities. By selecting an area, the graph shows the 10 countries of origin with the highest number of tourists (see Figure 7).

This dissemination is complemented with specific infographics for the annual data where, for example, the countries that receive the highest percentage of Spanish tourists in summer or winter (outbound tourism) can be visualized (see Figure 8) or the municipalities with the most tourists by country of origin (inbound tourism, see Figure 9)
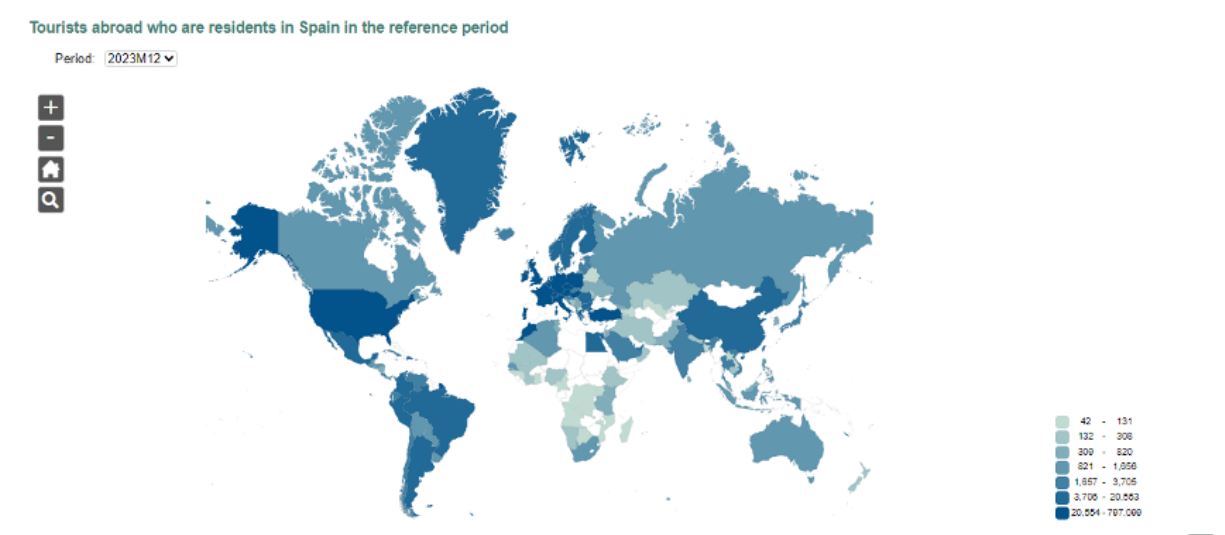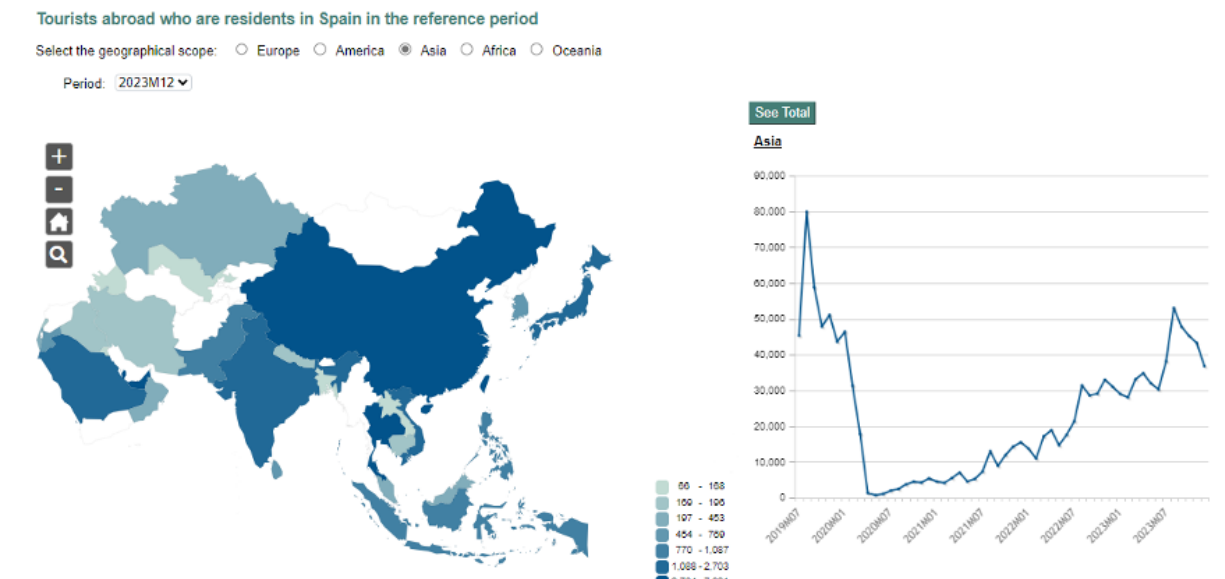
Figure 2: World map for outbound tourism data

Figure 3: Outbound tourism data by destination continents.

4.1   Cases of use

Two simple cases of us that can be carried out with the published data are shown below.

Objective: To analyse the evolution of the Asian countries most visited by residents in Spain in summer.

After downloading[1] the information on the number of monthly trips to the different countries on the Asian continent in the summer months (July, August and September) for the available years (2019-2023, see Figure 10).

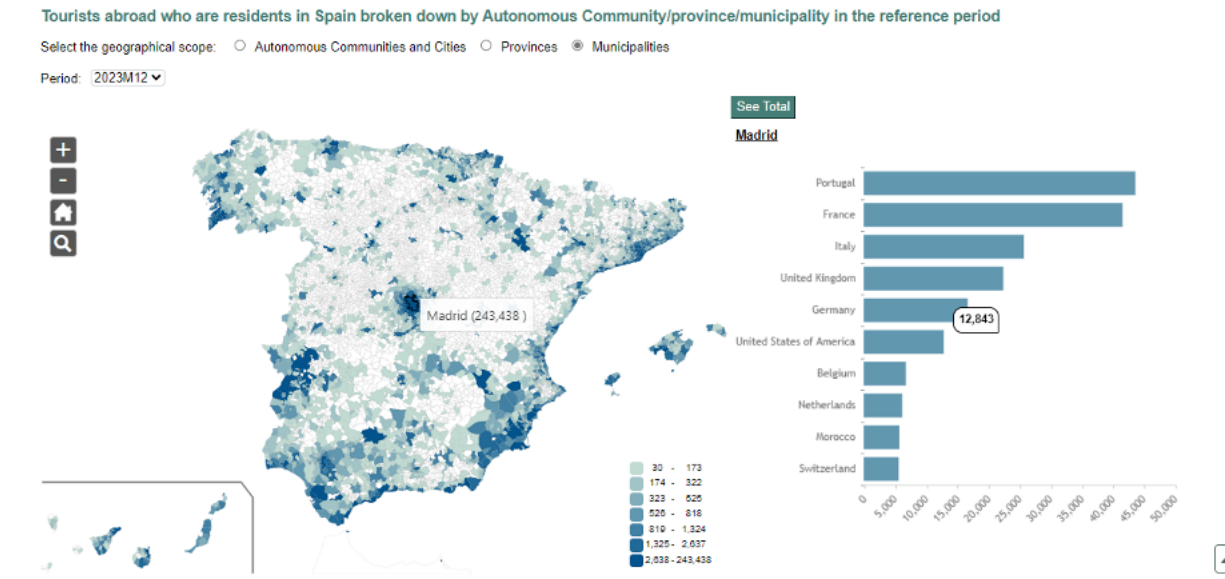[1] Available data: https://www.ine.es/dynt3/inebase/es/index.htm?padre=8578&capsel=8580

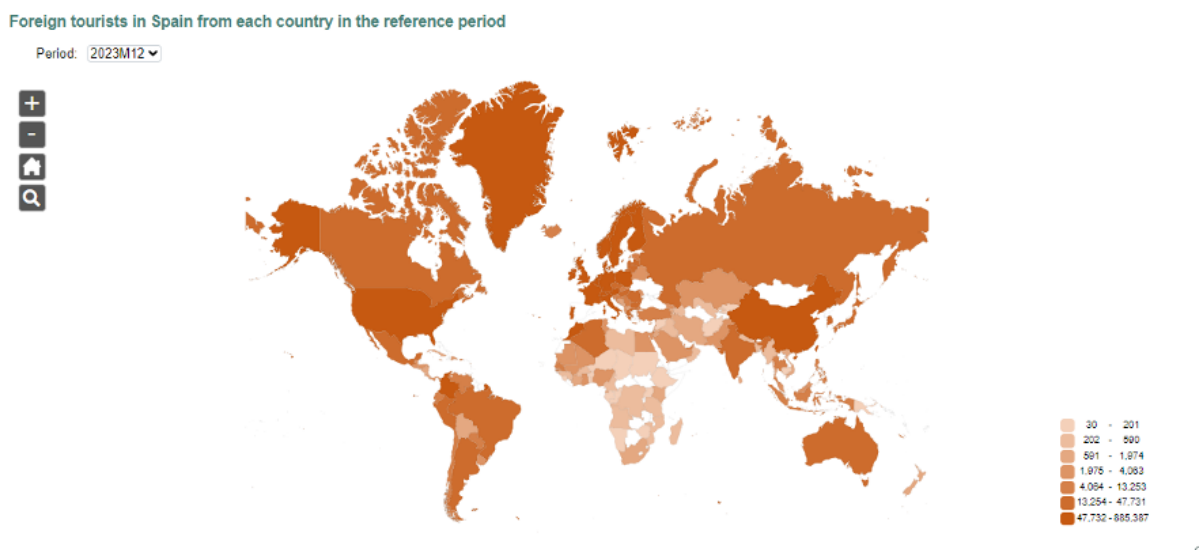Figure 4: Outbound tourism data by origin of trips.

Figure 5: World map for inbound tourism data.

The cases where a "." appears are because there are less than 30 trips and are hidden for statistical secrecy. For simplicity, it is assumed that there have been no trips in these cases.

Aggregating the data for the three summer months of each year and sorting by number of trips will result in the most visited Asian countries in the summer of each year (see Figure 11).

Some conclusions or conjectures could be drawn from these results:

- Tourist countries like Japan or Thailand disappeared from the Top-5 during 2020 and 2021 (Japan also in 2022), because of travel restrictions due to the pandemic.

Figure 6: Inbound tourism data by continents.

Figure 7: Inbound tourism data by destination.

- Maldives appears for the first time in the ranking in 2021. This may indicate that it has become fashionable among Spanish people or that perhaps they have put direct and cheaper flights to the country or had fewer travel restrictions.
- The year 2023 can be considered the year of normality in terms of international travel by Spaniards, as the ranking of the top 10 Asian countries is the same as in 2019 (except for changes in positions).

Objective: To identify favorite destination municipalities according to country of origin.
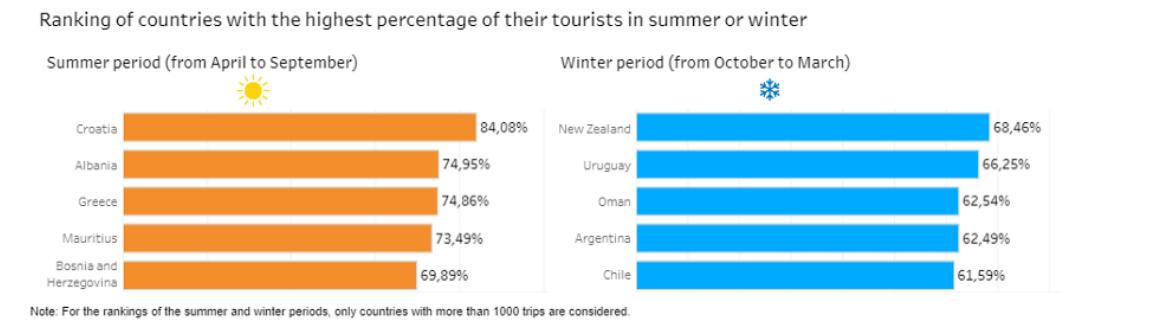
Figure 8: Ranking of countries with the highest percentage of tourist in summer or winter season.

Figure 9: Municipalities most visited by country.

Using monthly tourist data per destination municipality, broken down by continent and country of residence (for this example, data from August 2023[2] was used), it is possible to reveal the different preferences according to nationality. Another possible exercise along these lines would be to look at the different preferences of tourists from the same country throughout the year (see Figure 12).

5   Conclusions

After the experience gained in these years of work in this experimental statistic that we have just described, using cell phone positioning data, and in general, with the use of new sources of information, we can conclude that:

- It is important to use these sources of information, with rigor, within the framework of official statistics. The advantages they offer are many, among which stand out the more than significant increase in the granularity of the results obtained (difficult to obtain with traditional methods), the reduction of the burden on the informant or the improvement in the timeliness of the final results.
- In order to work accurately with these new sources of information it is very important to collaborate with the owners of these databases, especially if they have to make specific treatments to respond to statistical needs (adaptation of definitions, adjustment of fields of study, etc.). They must be aware of the relevance of the use of the information and there must be maximum collaboration between the statistical offices and the database owners.
- In general, the new data sources do not give a total answer or do not provide the complete information that in terms of variables, for example, of the traditional surveys, so the way to

---

[2]https://ine.es/experimental/turismo_moviles/exp_tmov_receptor_mun_2023.xlsx

	Tourists		
	2023M09	2023M08	2023M07
National Total			
Total	2.494.136	2.942.482	2.653.079
Asia	47.805	53.043	38.149
Afghanistan	.	.	.
Saudi Arabia	2.050	1.671	1.400
Azerbaijan	123	130	94
Bahrain	70	55	63
Bangladesh	113	96	31
Bhutan	.	67	
Brunei Darussalam	.	.	.
Cambodia	235	200	114
China	5.269	4.546	5.196
Korea, Republic of (South Korea)	1.092	1.221	768
North Korea	.	.	.
United Arab Emirates	5.728	5.282	4.477
Philippines	777	1.317	864
India	2.857	3.021	2.047
Indonesia	2.869	3.722	1.938
Iran, Islamic Republic of	305	226	159
Iraq	109	46	91
Marshall Islands	.	.	.
Israel	2.894	3.065	3.294
Japan	4.952	6.516	4.145

Figure 10: Trips to Asian countries

2023 (jul+aug+sep)	2022 (jul+aug+sep)	2021 (jul+aug+sep)	2020 (jul+aug+sep)	2019 (jul+aug+sep)
Japan	United Arab Emrates	Maldives	United Arab Emrates	China
United Arab Emrates	Israel	United Arab Emrates	Qatar	United Arab Emrates
China	Thailand	Qatar	China	Japan
Thailand	Qatar	Japan	Pakistan	Thailand
Israel	Jordan	Jordan	Lebanon	Israel
Qatar	Indonesia	Saudi Arabia	Japan	Indonesia
Indonesia	India	Pakistan	Corea	Vietnam
India	Saudi Arabia	Lebanon	Saudi Arabia	Qatar
Jordan	Vietnam	Israel	Israel	India
Vietnam	Maldives	China	India	Jordan

Figure 11: Ranking of most visited Asian countries by Spanish populations

follow integration of different sources. An example will be in the inbound tourism survey where the information provided by mobile telephony will serve to estimate the number of tourists visiting Spain, knowing the countries of origin and their places of destination, and with a survey it will be possible to know the characteristics of their trips such as the purpose of trips, type of accommodation or the tourism expenditure.

Germany	United Kingdom	Italy	USA	Japan
Palma	Calvià	Barcelona	Barcelona	Barcelona
Calvià	Barcelona	Madrid	Madrid	Madrid
Capdepera	Adeje	València	Palma	Donostia/San Sebastián
Barcelona	Arona	Sant Josep de sa Talaia	Sant Josep de sa Talaia	Granada
Madrid	Benidorm	Palma	València	Prat de Llobregat, El
Pájara	Sant Josep de sa Talaia	Formentera	Donostia/San Sebastián	Palma
Sant Llorenç des Cardassar	Palma	Calvià	Sevilla	Sant Josep de sa Talaia
San Bartolomé de Tirajana	Yaiza	Eivissa	Marbella	Málaga
Muro	Alcúdia	Lloret de Mar	Málaga	Bilbao
Llucmajor	Tías	Málaga	Eivissa	Palmas de Gran Canaria, Las

Figure 12: Ranking of most visited Spanish municipalities by tourist country of residence

References

Spain, NSI (2022a). Measurement of domestic tourism using mobile phone positioning. Technical report, Spain: National Statistics Institute.

Spain, NSI (2022b). Measurement of inbound tourism using mobile phone positioning. Technical report, Spain: National Statistics Institute.

Spain, NSI (2022c). Measurement of outbound tourism using mobile phone positioning. Technical report, Spain: National Statistics Institute.

UNWTO (2008). International recommendations for tourism statistics 2008. Technical report, Madrid, Spain: United Nations World Tourism Organization.