

La trampa de los 'rankings' incompletos o sesgados

JOSU MEZO

Cuando se hace un estudio sistemático de cualquier realidad social o económica en muchos lugares (varias ciudades o varios países), es muy habitual que en el informe de resultados se incluyan clasificaciones o *rankings*: listas de lugares ordenados según el valor que presenten en el asunto estudiado. En un artículo anterior de esta sección¹, expliqué ya algunos de los problemas que se pueden dar al informar sobre estas clasificaciones. El más importante es lo que se puede llamar la falacia del medallero: poner el foco en la posición precisa que ocupa en la clasificación nuestra ciudad o país, sus subidas o bajadas, sin reparar en que esas variaciones pueden deberse a cambios insignificantes o diferencias mínimas con territorios similares al nuestro. El reverso de ese error suele ser prestar escasa atención a historias interesantísimas de transformaciones enormes que se producen en países tal vez más lejanos y distintos al nuestro, pero que

por su población y peso económico tienen gran trascendencia.

En esta ocasión, quiero llamar la atención sobre otros dos problemas potenciales de los *rankings*: que sean incompletos (no están todos los lugares que deberían estar) o que sean sesgados (el método de elaborarlos favorece la presencia de informaciones de ciertos tipos de lugares más que de otros). Si sucede cualquiera de esas dos cosas, la posición en la tabla clasificatoria de nuestra ciudad o nuestro país no será simplemente poco relevante, sino directamente engañosa.

Por eso es necesario, como siempre, buscar la fuente de la que procede la información. Como pasa también con otro tipo de noticias, nos podemos llevar la sorpresa de que no aparezca ninguna. Todo periodista ha experimentado alguna vez la frustración de buscar y rebuscar la fuente original

Josu Mezo es profesor de la Universidad de Castilla-La Mancha y editor del blog *Malaprensa*

¹ "¿Cómo va lo nuestro?", *Cuadernos de Periodistas*, n.º 22, mayo de 2011, pp. 127-132

de una afirmación que circula prácticamente como si fuera saber común, encontrando solamente ejemplos de noticias, declaraciones, artículos, cada vez más antiguos, que repiten la misma frase o idea sin citar fuentes o haciendo una atribución muy genérica del tipo “según la ONU”. A veces acaba resultando que no hay ningún informe ni estudio que sustente la afirmación. Esto le sucedió al periodista Manuel Ansede² cuando trató de localizar el origen de la aserción, prácticamente parte de la sabiduría popular, según la cual España es el segundo país más ruidoso del mundo, después de Japón, reproducida por múltiples medios e incluso dada por buena por alguna autoridad pública. Tras encontrar y descartar atribuciones a la Organización Mundial de la Salud o la National Academy of Sciences (de Estados Unidos), concluyó que el mito se originó en un congreso en Zaragoza, en el que alguien declaró que supuestamente procedía de un informe de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el cual nadie ha sido capaz de encontrar. El congreso se celebró nada menos que en 1987; y, desde entonces, ese *ranking* imaginario del ruido, en el que somos medalla de plata, ha ido rodando de noticia en noticia sin que nadie lo compruebe.

Lo más habitual, sin embargo, será que la clasificación de la que tenemos noticias no sea un puro invento, y proceda de algún estudio al que, con más o menos esfuerzo, podamos acceder. En ese caso no es raro que descubramos que las clasificaciones que se pretenden exhaustivas no lo sean tanto, y resulte que no han incluido datos de todos los casos posibles, sino solo de unos pocos. Así, a menudo, informes que se difunden diciendo cosas del tipo “España es el país europeo donde más pasa A” o “donde menos sucede B” abarcan solamente datos de cinco, siete o diez países europeos.

En la misma línea podemos encontrar que lo que se presenta como “el país más [pongan aquí su adjetivo] del mundo” acaba siendo solo el país más [...] de diez, 20 o 30 países estudiados. O ni siquiera eso. Un ejemplo famoso, ya de hace unos años, fue la afirmación, difundida por una asociación del sector y repetida sin pestañear por casi todos los medios, de que España era el décimo país del mundo donde había más piratería de música (aún se hablaba más de piratería a través del top manta que de las descargas *online*). Lo del puesto décimo fue una total invención, o tal vez –pensemos benévolamente– un malentendido, producido

2 Ansede publicó sus hallazgos en un divertido hilo de Twitter disponible aquí: <http://tiny.cc/ansede-ruido>

por la explicación oral de otro contenido del informe en la rueda de prensa.

De hecho, el informe no hacía un *ranking* individualizado de los países, sino que los agrupaba en cuatro niveles de piratería, según el porcentaje que supusiera la música pirata sobre el total de música comprada (pirata más legal)³. El primer nivel, en el que más del 50% de la música sería pirata, incluía a 31 países; en el segundo nivel, con ventas piratas entre el 25 y el 50% del mercado, estaban 25 países más; España era uno de los once países en el tercer nivel, con ventas piratas entre el 10 y el 24% del total, y finalmente otros 16 países estaban en el cuarto nivel, con menos del 10% de la ventas no autorizadas. En una supuesta clasificación individual, por lo tanto, habría 56 países con más piratería que España, que estaría en algún lugar entre los puestos 57 y 67. Y un imaginario titular –poco probable, por falta de gancho– tendría que decir algo así como que “España es el 57.º país con más piratería musical”, tal vez aclarando en el subtítulo que empataba en esa posición con otros diez países.

Pero aquí viene la clave: aun suponiendo que todos estos países estuvieran

ordenados individualmente, y España ocupara precisamente la posición 57, tampoco ese titular sería correcto, puesto que el estudio solo cubría 83 países. ¿Quizá podríamos arreglarlo diciendo algo como “España, en la posición 57 de 83 países en piratería musical”? Seguiría siendo un titular engañoso, porque los países incluidos no venían en igual proporción de todos los continentes y niveles de riqueza. Estaban sobrerrepresentados los países desarrollados, de los que apenas faltaba alguno, y no aparecían la mayoría de los países en desarrollo, en particular africanos (había cinco) o asiáticos (se incluían doce). Asimismo, la mayor parte de los países ricos y desarrollados estaban en los grupos tercero y cuarto (con menos piratería), y la mayor parte de los países en desarrollo estaban en los grupos primero y segundo (con más piratería). De manera que se podía razonablemente estimar que una buena parte de los más de 100 países ausentes en la tabla habrían estado, de ser incluidos, en los grupos primero y segundo, con niveles de piratería superiores a los de España. Y entonces, la posición “57 entre 83”, que dejaría a España más o menos detrás de dos tercios de los países con más piratería y delante de un tercio con menos pira-

3 No entro aquí en la discusión sobre si el término “pirata”, cargado de connotaciones, es el más apropiado para hablar de la música vendida o distribuida sin autorización del propietario de los derechos de autor. Lo acepto por comodidad y popularidad del término

tería, sería una distorsión de la realidad que habríamos encontrado si hubiéramos podido incluir a todos los países del mundo.

En definitiva: el estudio no era exhaustivo, porque no incluía a todos los países; lo cual, en sí mismo, no tiene por qué ser criticado. Solo las grandes organizaciones internacionales, como las vinculadas a las Naciones Unidas, tienen recursos para generar o compilar informaciones estadísticas de todos los países. Un estudio limitado solo a unos cuantos países puede ser perfectamente legítimo, informativo y científico. Pero lo importante es que al resumirlo y difundirlo no se tergiverse su significado. La posición relativa de cualquier país en un estudio como este (tanto da que fuera una de las primeras como una de las últimas) no debería resumirse nunca como “el país con más piratería del mundo” o “el país con menos piratería”, ni ninguna otra posición. Si decía más arriba que en la mayoría de los *rankings* sociales o económicos es mucho más importante el valor conseguido que la posición ocupada, y que debemos evitar la falacia del medallero, esto se hace aún más imperativo cuando directamente dicho medallero es incompleto porque faltan competidores.

Una variante especialmente grave de esta situación se da si, en realidad, el informe o texto que es la fuente de la información no había ni siquiera inten-

tado hacer un mínimo *ranking*, sino que simplemente citaba unos pocos casos como ejemplos. Esto es lo que pasó a comienzos de 2019, cuando nada menos que la delegada del Gobierno para la Violencia de Género, Pilar Llop, declaró que España, “según datos de la ONU, se ha erigido en el tercer país del mundo en demanda de prostitución”, aseveración llamativa que fue recogida en titulares por varios periódicos nacionales. Tras unas cuantas búsquedas frustradas, y con ayuda de mis lectores en Twitter, llegué a la conclusión de que el origen de esa afirmación que nos deja en tan mal lugar sería el informe *The Globalization of Crime*, de la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC), publicado en 2010, que en su página 49 contenía el siguiente párrafo [traducción propia]:

“Los datos de encuestas nacionales sugieren que el porcentaje de hombres que han comprado servicios sexuales alguna vez en su vida varía considerablemente entre países y en el tiempo. Según las encuestas de Kinsey en la década de 1940 [en Estados Unidos], el 70% de los hombres adultos declararon haber pagado por sexo al menos una vez en sus vidas, pero en esa época el sexo extramarital no pagado era mucho menos común que hoy. Encuestas más recientes sugieren que la cifra actual es más cercana al 19%. Encuestas recientes en otros países sugieren una cifra

similar en Suecia (13%), los Países Bajos (14%), Australia (15%) y Suiza (19%). España (39%) es un valor atípico en Europa, como lo es Puerto Rico (61%) en América del Norte. La cifra comparable es aún mayor en Tailandia (73%)”.

Claramente, el dato español es llamativamente más alto que el de otros países desarrollados mencionados, aunque se podría discutir mucho sobre su fiabilidad, ya que es antiguo y es más alto que los publicados en otros estudios, incluidos algunos del Instituto Nacional de Estadística (INE) y el Centro de Investigaciones Sociológicas (CIS). Pero lo importante en este caso es que de un párrafo que menciona ocho casos sueltos no se puede de ninguna manera deducir cuál es “el tercer país del mundo en demanda de prostitución”, ni tampoco –claro– el primero ni el último. Es una tergiversación total de la información.

Un último escollo con la interpretación de estas clasificaciones, quizá el más difícil de esquivar, se da cuando un estudio no ha tomado datos solamente de algunos países, ha utilizado un método de investigación sesgado, que en la práctica ha favorecido a algunos lugares sobre otros. Por ejemplo, hace ya unos cuantos años un periódico

español informaba de que la administración electrónica española estaba por detrás de la de Armenia, Bolivia, Togo o Corea del Norte. La información procedía de un informe, aparentemente muy serio, elaborado por investigadores de una universidad norteamericana, examinando páginas web de la administración de 198 países, así que no nos enfrentábamos esta vez al problema del *ranking* incompleto. Sin embargo, entre líneas, era posible darse cuenta de que el estudio tenía severos problemas: había examinado solo unas pocas webs de cada país, daba mucha importancia a la información orientada hacia el exterior, para turistas o inversores extranjeros, y tenía un sesgo lingüístico favorable a las páginas traducidas al inglés. Por todo ello, a pesar de ser un *ranking* aparentemente completo, estaba sesgado y tenía muy poco valor⁴.

En otras ocasiones ha sucedido que estudios supuestamente globales, pero realizados por empresas o instituciones españolas, tenían un sesgo lingüístico o tecnológico (o ambos) que les llevaba a encontrar más casos en nuestro país, o en otros de habla hispana, dando lugar a afirmaciones muy chocantes, como que España es el país del mundo con más ordenadores infectados por bots o que es el segundo país del mundo con mayor posesión o tráfico de imágenes

4 Más detalles en: <http://tiny.cc/mp-adm-elect>

de pornografía infantil (en número absoluto de archivos). La primera afirmación se basaba en el estudio de una empresa española de *software* antivirus (lógicamente, con muchos más clientes en España) y la segunda partía de un estudio lanzado desde España, el cual había explorado las redes P2P más populares en nuestro país, y probablemente también había usado palabras en español para rastrear los archivos⁵.

En definitiva, ya sabíamos que es una mala idea centrarse en la posición de nuestro país o nuestra ciudad en un *ranking* completo y bien elaborado.

No obstante, hay que estar aún más alerta con los *rankings* incompletos o sesgados. En el primer caso, nuestra posición en el medallero es engañosa, y debemos evitar hablar de ella. Si el estudio es bueno, aún puede tener interés comentar los valores que tiene cada territorio (el nuestro y otros similares) en cada uno de los indicadores. En el segundo caso, realmente, es muy poco lo que podemos extraer de interés. No vale la posición, no valen los datos. A veces el informe sirve para poco más que dar a conocer a quien lo ha hecho. Pensemos bien si merece que le hagamos ese favor.

5 Más información sobre este último caso en: <http://tiny.cc/mp-Alia2>