

FORMACIÓN PROFESIONAL · INVESTIGACIÓN COMERCIAL · UNIDAD 4

Fuentes secundarias y suficiencia

Lunes 22 de febrero de 2027. Unai llega con un informe gratuito: «El 38 % de los hogares compraría kits de comida preparada». Nora hace la cuenta: con unos 6.000 hogares en el barrio, 2.280 compradores; «no hace falta encuesta». Pero la encuesta es de otra ciudad, de 2022, voluntaria y cuenta como «hogar» a cada persona que responde. ¿Qué sirve y para qué?

PARA EMPEZAR

En esta unidad aprenderás a…

- Planificar una búsqueda online y offline con buscadores, operadores avanzados y criterios de selección, y registrarla.
- Valorar por separado la fiabilidad, la cobertura, la actualidad y las condiciones de uso de una fuente.
- Normalizar datos con sus metadatos, leer una serie temporal y deflactar importes con un índice de precios.
- Decidir, objetivo por objetivo, si la información disponible basta y qué falta por preguntar.
- Archivar las fuentes con versión, responsable y fecha para que otro equipo pueda recuperarlas.

SITUACIÓN DE PARTIDA

«No hace falta encuesta»

Lunes 22 de febrero de 2027. Unai llega con un informe gratuito que ha encontrado en internet: «El 38 % de los hogares compraría kits de comida preparada». Nora hace la cuenta en voz alta: «Si en el barrio hay unos 6.000 hogares, son **2.280 hogares**. No hace falta encuesta».

Adrián lee la letra pequeña: la encuesta es **de otra ciudad, de 2022**, la contestó **quien quiso** y cuenta como «hogar» a **cada persona que responde**. Sobre la mesa hay tres fichas: **A**, el padrón; **B**, ese informe; **C**, las ventas de la prueba de kits. Nora te encarga el **inventario de fuentes y lagunas** del estudio de kits para hogares, para el viernes 26/02, antes de pagar ningún dato nuevo.

❓ ¿Qué quiere decidir Nora con el 38 % y qué tendría que comprobar antes?

❓ ¿Qué sabemos de verdad sobre los hogares del barrio y qué estamos suponiendo?

❓ ¿Qué dato del informe permitiría revisar la cuenta de Nora?

❓ ¿Por qué no se pueden multiplicar personas por un porcentaje de «hogares»?

MAPA DE LA UNIDAD

Lo que vamos a ver

4.1 Búsqueda online y offline

4.2 Fiabilidad y representatividad

4.3 Organización y metadatos

4.4 Suficiencia y lagunas

4.5 Archivo y recuperación

Muchos datos, ¿de quién?



Este informe tiene muchos datos.

¿Habla de nuestra población?





No exactamente; señalaré la diferencia.

CLAVE

Una fuente prestigiosa de otra población no se vuelve representativa del barrio por ser prestigiosa.

Primero el dato, después dónde buscar



Plan de búsqueda

Para cada objetivo: qué variable, de qué población y de qué periodo. De ahí salen los **términos** (y sus sinónimos: «comida preparada», «platos cocinados», «comida para llevar») y los **lugares**.



Fuentes online

Organismos estadísticos (el INE publica resultados agregados), portales de datos abiertos, ayuntamientos, cámaras de comercio, asociaciones sectoriales y de consumidores, publicaciones especializadas.



Fuentes offline

Archivo interno (tiques, hojas de pedidos, libreta de peticiones), publicaciones impresas, bibliotecas y documentación pedida a un organismo. También se fechan y se registran.



Motores de búsqueda

El buscador generalista ordena por relevancia, popularidad y anuncios: el primero es el mejor posicionado, no el mejor. Los buscadores internos y los catálogos de biblioteca filtran por tema, territorio y periodo.

CLAVE

Se busca con una pregunta delante: primero se decide qué dato hace falta y después dónde puede estar.

Operadores de búsqueda avanzada

OPERADOR	QUÉ HACE	EJEMPLO
"..."	Busca la frase exacta	"comida preparada" hogares
-	Excluye un término	kits comida -receta
OR	Acepta cualquiera de los dos	padrón OR población
site:	Limita a un dominio	padrón site:ine.es
filetype:	Limita al tipo de archivo	encuesta hogares filetype:pdf

Registro de búsqueda: «22/02/2027 · buscador general · "comida preparada" hogares filetype:pdf · 30 resultados revisados · elegido: informe de consumidores (ficha B) · límite: otra ciudad, 2022, «hogar» = persona».

CLAVE

Cada buscador tiene sus operadores: compruébalos en su ayuda. Cada búsqueda deja una línea: fecha, lugar, consulta exacta, resultados revisados, fuente elegida y motivo.

Elegir, ir al origen y no acumular

✅

Criterios de selección

Quién publica y con qué propósito; si explica el método; ámbito y periodo; fecha de actualización; condiciones de uso; coste. Sin autor ni método, no sirve como dato.

🧬

Ir al origen

Se sigue la cadena hasta el **origen primario** y se anotan su definición y su método. Si se corta: «origen no localizado», solo como pista.

🔁

La cifra copiada

Una cifra redonda que aparece en muchas webs sin fecha ni autor suele ser la misma cifra copiada, no varias confirmaciones.

⚠️

Errores habituales

Quedarse con la primera página; confundir fecha de publicación y de los datos; tomar un resumen de prensa por el informe; buscar solo lo que confirma la idea de partida.

CLAVE

Buscar no es acumular: se guarda lo que responde a una variable, una población y un periodo del plan, con su origen, su definición y su fecha.

Cuatro preguntas por separado



Fiabilidad

¿Quién la elabora, con qué método documentado y con qué control de errores? ¿Se puede comprobar?



Cobertura

¿La población, el territorio y la definición coinciden con los del objetivo? (Representatividad.)



Actualidad

¿De cuándo son los datos, no la publicación? ¿Ha cambiado algo desde entonces?



Condiciones de uso y seguridad

¿Se puede reutilizar y citar? ¿Contiene datos personales? ¿El archivo es íntegro?

CLAVE

No se pone una nota única: el padrón es muy fiable y, aun así, no habla de compras.

412 voluntarios o 100 al azar



Ficha B: 412 respuestas

Encuesta voluntaria en línea: responde quien quiere, sobre todo los más interesados. Ni 400 respuestas ni 4.000 lo corrigen. Y «hogar» es cada persona que responde: si en una casa contestan dos, cuenta dos veces.

VS



100 elegidas al azar del barrio

Una muestra representa si el marco cubre la población, se elige con un diseño apropiado (mejor, al azar) y se trata la no respuesta: poder participar no basta.

CLAVE

El tamaño reduce el azar, **no el sesgo**. Por eso la ficha B es menos representativa que 100 respuestas elegidas al azar del barrio.

🙋 PREGUNTA PARA LA CLASE

Nora calcula: «0,38 × 6.000 hogares = 2.280 hogares compradores». ¿Qué falla?

A Nada: es una regla de tres

B El 38 % es de otra ciudad, de 2022, voluntario y con otra definición de hogar, y los 6.000 hogares no tienen fuente

C Que habría que usar los 14.520 empadronados

✔ undefined

El padrón: un registro de vecinos



«El Padrón municipal es el registro administrativo donde constan los vecinos de un municipio».

Ley 7/1985, de 2 de abril, reguladora de las Bases del Régimen Local, art. 16.1 · Consulta la versión vigente

SUJETO OBLIGADO

Cada ayuntamiento, que lo gestiona y remite sus datos al INE (art. 17.3) para coordinar los padrones de todos los municipios.

CLAVE

La ficha A es un registro administrativo, no una encuesta: cuenta **personas** empadronadas, no hogares, y no dice nada de lo que compran. Sirve para el contexto de población.

Aceptar, limitar o descartar

✔️

Aceptar

Responde al objetivo con su límite escrito.

🟡

Limitar

Solo como contexto o hipótesis.

✖️

Descartar

Llevaría a una conclusión equivocada.

CLAVE

La decisión se toma **por objetivo** y se anota junto a la fuente, con su criterio. Las reseñas y opiniones de una red social, elegidas por quien escribe y sin saber si compró, son indicio, nunca medida (Ley 3/1991, art. 27.7 y 27.8: reseñas falsas o no comprobadas).

4.3 · TABLA 1

Fichas del inventario, normalizadas

FICHA	VARIABLE Y VALOR	UNIDAD	PERIODO	POBLACIÓN	LIMITACIÓN
A	Empadronados: 14.520	Personas	01/01/2026	Secciones del barrio	No da hogares ni compra
B	Compraría kits: 38 %	% de quienes responden (412)	2022	Otra ciudad, voluntarios	«Hogar» = persona que responde
C	Kits vendidos: 648	Kits	11/01-20/02/2027	Compradores de K01 y K02	213 kits sin cliente identificado

CLAVE

Los datos se normalizan sin borrar el original: unidad, periodo y definición viajan con cada cifra; lo que no cabe en una columna, a la limitación.

Los datos que describen el dato

Metadatos

Definición, unidad, fuente con su enlace o ubicación, fecha de los datos y de la consulta, versión, transformaciones y responsable. Sin ellos, nadie sabrá si «648» eran kits, tiques o euros.

No mezclar unidades

Personas y hogares no se suman ni se dividen sin decirlo: 14.520 ÷ 6.000 = **2,42 personas por hogar**, un dato derivado.

Con sus fuentes

Ese 2,42 se anota con fuentes y fechas: padrón, 01/01/2026; hogares, sin fuente ni fecha.

El original no se toca

Se trabaja sobre una copia y cada transformación (cambio de unidad, media móvil, deflactado) se anota con su fórmula.

CLAVE

Así se puede volver al dato bruto, repetir el cálculo y comprobar que sale lo mismo.

4.3 · TABLA 2

Kits por semana y media móvil de tres

Una **serie temporal** tiene **tendencia**, **estacionalidad** (las meriendas del sábado), **ciclo** (varios años) e **irregular** (un pedido grande, una semana de frío).

SEMANA	11-16/01	18-23/01	25-30/01	01-06/02	08-13/02	15-20/02
Kits	92	126	98	124	100	108
Media móvil (3)	—	105,33	116,00	107,33	110,67	—

(92 + 126 + 98) ÷ 3 = 105,33. La serie suavizada oscila alrededor de 108 kits (648 ÷ 6); las semanas de 126 y 124 tienen explicación propia (la prueba del K02, del 18 al 23/01, y un pedido de formación más el anuncio, sin separar, del 01 al 06/02).

CLAVE

La media móvil centrada suaviza lo irregular y deja ver la tendencia. Seis semanas no bastan para ver una tendencia a largo plazo ni estacionalidad: haría falta un año de datos, mejor dos.

4.3 · TABLA 3

Deflactar: euros de 2023

Valor real = valor nominal ÷ índice × 100. Índice de los kits ficticio; en un estudio real, el IPC del INE (general, base 2025 = 100: 94,8 en 2023; consulta del 29/10/2026).

AÑO	2023	2024	2025	2026
Tique medio nominal (€)	12,80	13,30	13,70	14,00
Índice de kits (2023 = 100)	100,0	102,8	105,6	108,4
Tique medio real (€ de 2023)	12,80	12,94	12,97	12,92

2026: 14,00 ÷ 108,4 × 100 = 12,92 €. Variación nominal (14,00 − 12,80) ÷ 12,80 = 9,4 %; real (12,92 − 12,80) ÷ 12,80 = 0,9 %.

CLAVE

Casi toda la subida es de precios. Se anota qué índice, con base y fecha; y nunca se deflacta una cantidad física (kits, tiques): solo importes en euros.

🙋 PREGUNTA PARA LA CLASE

Actividad 6. Los vecinos de diario gastaban 8,40 € por visita en 2023 y 9,00 € en 2026 (índice 108,4). En euros de 2023…

A gastan 9,76 €: un 16,2 % más

B gastan 8,30 €: un 1,2 % menos

C gastan lo mismo

✔ undefined

4.4 · TABLA 4

Matriz objetivo-evidencia-suficiencia (22/02/2027)

OBJETIVO	EVIDENCIA DISPONIBLE	VEREDICTO
O1. Hogares del barrio	A: 14.520 personas; unos 6.000 hogares sin fuente documentada	Parcial
O2. Proporción de hogares que compraría kits	B: otra ciudad, 2022, voluntaria, otra definición	Sin evidencia
O3. Perfil y frecuencia de los compradores actuales	C: 435 kits de 107 titulares (4,07 por titular); 213 sin identificar	Parcial
O4. Servicio de pedido preferido	Libreta: peticiones sueltas de reparto por la tarde	Parcial (solo indicio)

Cubierto si: mide la variable con la misma definición; misma población y periodo vigente; precisión que permite elegir; equivocarse cuesta menos que pagar datos nuevos. Si falla una, parcial; si fallan las dos primeras, sin evidencia; en otros casos, parcial si mide la variable y la población, y sin evidencia si no.

CLAVE

Las **lagunas** justifican pagar información primaria: O2, encuesta a una muestra de hogares del barrio; O4, la misma encuesta o entrevistas; O1, preguntar al ayuntamiento si publica hogares por sección.

Un dato que otro puede encontrar



Estructura

Una carpeta por proyecto: **brutos** (sin tocar), **trabajo** (copias normalizadas) y **resultados**. Nombres `AAAA-MM-DD_tema_estado_vN`: 2027-02-22_padron_bruto_v1.



Ficha de fuente

Título, organismo, enlace o ubicación, fecha de los datos y de consulta, condiciones de uso, versión, responsable y uso en el estudio.



Versiones

El padrón de 2027 no sobrescribe al de 2026: nueva versión con fecha, se conserva la anterior y se anota en el **registro de cambios** qué, quién, cuándo y por qué.



Acceso y seguridad

Brutos en solo lectura; copia de seguridad en otro soporte con restauración probada; datos personales cifrados, copias seudonimizadas y no más tiempo del necesario (RGPD, art. 5.1 e).

CLAVE

La prueba de un buen archivo: otro equipo, sin preguntar, localiza el dato, lee su definición y reconstruye cada transformación.

Reutilizar información pública



La reutilización puede estar sometida, entre otras, a estas condiciones: «a) Que el contenido de la información, incluyendo sus metadatos, no sea alterado», «c) Que se cite la fuente» y «d) Que se mencione la fecha de la última actualización».

Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, art. 8 · Consulta la versión vigente

SUJETO OBLIGADO

Quien reutiliza información pública, como el equipo que trabaja con los datos del padrón.

CLAVE

Por eso el bruto se guarda sin tocar, con su fuente y su fecha, y cada transformación se hace en una copia.

CASO RESUELTO · PASO A PASO

El inventario de fuentes del estudio de kits



Tu equipo, para Nora

Lunes 22/02/2027: ¿basta con lo que existe para dimensionar la línea de kits para hogares o hay que pagar una encuesta? Caso ficticio.

1. **Decisión, datos y alcance.** Encargar o no una encuesta a hogares antes de ampliar la producción. A: padrón, 14.520 personas (01/01/2026). B: otra ciudad, 2022, 412 respuestas voluntarias, 38 %. C: 648 kits (11/01-20/02/2027). Alcance: hogares del barrio en 2027.

2. **Procedimiento.** Cada ficha se decide según su uso (O2, contexto o clientela actual). A: alta fiabilidad, personas y no hogares → limitar a contexto. B: fiabilidad baja (método sin control); otra ciudad, voluntaria, otra definición, 2022 → limitar a hipótesis. C: registro propio, solo compradores actuales → aceptar para la clientela actual.

3. **Comprobación.** 435 kits de 107 titulares (4,07 por titular) + 213 sin identificar = 648 (108 por semana). La cuenta de Nora (0,38 × 6.000 = 2.280) no es un dato.

4. **Interpretación.** O1 parcial, O2 sin evidencia, O3 parcial, O4 parcial. Permite concluir cuántas personas están empadronadas y cómo compran los clientes con tarjeta; no cuántos hogares comprarían kits.

✔ A contextualiza la población, no la demanda; B limita la extrapolación; C describe la clientela actual. Antes de recoger datos se registran los huecos: **O2 es la laguna que justifica la encuesta**.

I04 Inventario y lagunas

Para la investigación de tu plan I03:

1

Busca

Al menos tres fuentes secundarias (una online oficial y una offline), con su registro de búsqueda.

→ 2

Valora y normaliza

Las cuatro preguntas para cada una y una tabla normalizada con metadatos.

→ 3

Decide

La matriz objetivo-evidencia-suficiencia y las preguntas pendientes con su técnica primaria.

CLAVE

Comprueba tu dosier: cada fuente tiene ficha, versión y fecha de consulta; los brutos están sin tocar; incluyes archivos y registro de cambios, la decisión, el responsable y la fecha de revisión.

APRENDE CON LA IA

La respuesta que recibió Unai

Tres frases del asistente de IA sobre el padrón y una corrección. Pulsa cada tarjeta.

«El padrón es oficial y fiable: úsalo sin más comprobaciones»

BULO

Fiable no es pertinente: cuenta 14.520 personas, no hogares.

«Una fuente oficial responde a cualquier pregunta comercial»

BULO

La autoridad no sustituye la pertinencia: el padrón no mide compras.

«Aplica a los empadronados el porcentaje de compra que encuentres»

BULO

Mezcla personas con hogares y otra ciudad con el barrio.

«El padrón sirve como contexto de población del barrio»

REALIDAD

Limitar a contexto; la demanda (O2) pide una encuesta propia.

REPASO

Ideas clave

- Se busca con un plan: variable, población y periodo de cada objetivo; términos con sinónimos, operadores y un registro de cada búsqueda.
- Se sigue la cifra hasta su origen primario; una cifra redonda repetida sin autor suele ser la misma copiada.
- Fiabilidad, cobertura, actualidad y condiciones de uso se valoran por separado: el padrón es muy fiable y no habla de compras.
- El tamaño reduce el azar, no el sesgo: 412 respuestas voluntarias representan menos que 100 elegidas al azar.
- Cada fuente se acepta, se limita o se descarta para cada objetivo, con el criterio escrito.
- Tabla normalizada con metadatos: no se mezclan personas y hogares (14.520 ÷ 6.000 = 2,42, dato derivado).
- Media móvil de tres semanas: suaviza lo irregular (105,33; 116,00; 107,33; 110,67).
- Deflactar: nominal ÷ índice × 100. El tique sube un 9,4 % nominal y solo un 0,9 % real.
- La suficiencia se decide objetivo por objetivo (cubierto, parcial, sin evidencia), y el archivo guarda brutos, versiones y registro de cambios.



¡Ahora te toca a ti!

Repasa, practica y comprueba lo que has aprendido.

📝 **Resumen**

✅ **Test**

👆 **Interactivos**

Libro, presentaciones, resúmenes y tests gratis en **josesande.com**