

# Consum de Tapes als Locals Barcelonins

*Entrega Teoria 1*



**Integrants del grup:**

Pau Queralt Muñoz  
Alberto Moreno Marinez  
Gerard Fideu Garcia

Abril 2025

# Índex

<b>1</b>	<b>Introducció</b>	<b>2</b>
<b>2</b>	<b>Objectius</b>	<b>2</b>
<b>3</b>	<b>Model Entitat-Relació</b>	<b>3</b>
3.1	Entitats . . . . .	3
3.2	Relacions i Cardinalitats . . . . .	4
3.3	Model E-R: . . . . .	5
<b>4</b>	<b>Recol·lecció i Neteja de Dades</b>	<b>7</b>
4.1	Fonts de Dades . . . . .	7
4.2	Procés de Neteja de Dades . . . . .	8
4.3	Futur: Plataforma Col·laborativa . . . . .	8
<b>5</b>	<b>Avanç de Consultes Mitjançant l'Àlgebra Relacional</b>	<b>9</b>
5.1	Consulta 1: Quin usuari ha consumit més en un bar? . . . . .	9
5.2	Consulta 2: Quin bar ha tingut més consumicions? . . . . .	9
5.3	Consulta 3: Quin és l'ingredient més consumit? . . . . .	9
<b>6</b>	<b>Conclusions</b>	<b>10</b>
6.1	Aportacions: . . . . .	10
<b>7</b>	<b>Anexos</b>	<b>11</b>
7.1	Repositori GitHub . . . . .	11

# 1 Introducció

Amb el creixent auge dels brunchs i dels cafès d'especialitat, observem com es dilueix progressivament la cultura dels bars tradicionals que ofereixen menjar casolà i típic. Les tapes, un dels pilars fonamentals de la gastronomia barcelonina, sovint queden relegades a un segon pla davant de propostes més modernes i instagrammables”.

Com a grup d'amics apassionats per aquesta cultura, tenim el costum de relaxar-nos amb una bona tapa i un refresc en qualsevol bar autèntic de la ciutat. De fet, fa temps que compartim recomanacions, registrem quins bars ens agraden i fins i tot valorem les tapes que tastem (“Les bombes de la Cova Fumada són les millors: 9.75/10”).

A més, un dels membres del grup és celíac, fet que ens ha portat a tenir en compte les restriccions alimentàries a l'hora d'escollir locals: saber si ofereixen opcions sense gluten o plats aptes per a vegetarians és fonamental.

Aquest projecte neix, doncs, de tres motivacions principals:

- **Preservar** i catalogar bars i tapes autèntiques de Barcelona abans que desapareguin.
- **Garantir** la seguretat alimentària de col·lectius amb restriccions (persones celíaques, vegetarianes, amb al·lèrgies, etc.).
- **Analitzar** patrons de consum: quines tapes són més populars? Hi ha estacionalitat? Quines combinacions o quantitats es repeteixen?

# 2 Objectius

L'elecció de les tapes com a eix temàtic per a la nostra base de dades respon a un conjunt d'objectius concrets que busquem assolir a través de la seva construcció i posterior explotació:

- **Dissenyar i estructurar una base de dades relacional robusta**, que reculli informació rellevant sobre tapes i establiments de restauració barcelonins: ingredients, preus, opcions per a dietes especials, valoracions, zones i moments de consum.
- **Permetre consultes complexes mitjançant lògica relacional**, que ajudin a filtrar tapes segons criteris com el tipus d'ingredient, disponibilitat per a celíacs o vegetarians, barri, horari o preu mitjà.
- **Analitzar patrons de consum** a través de consultes SQL orientades a detectar quines tapes són les més demanades, quines combinacions apareixen amb més freqüència i si existeixen preferències estacionals o per franges horàries.
- **Estudiar les quantitats de consum i la seva variabilitat**, identificant establiments amb altes ràtios de comanda per client, o tapes que tenen més recurrència entre consumidors.
- **Proporcionar una eina útil per a la comunitat gastronòmica**, oferint una base de dades oberta que permeti tant a professionals com a aficionats localitzar tapes tradicionals, comparar establiments i prendre decisions de consum informades.

- **Preservar i donar visibilitat a la cultura gastronòmica local**, posant en valor les tapes clàssiques davant l'impacte creixent de tendències gastronòmiques globals i efímeres.

### 3 Model Entitat-Relació

A continuació, es mostren les entitats clau que conformen el nostre model entitat-relació, orientat a gestionar i analitzar informació sobre tapes i consum en bars de Barcelona.

#### 3.1 Entitats

##### Bar

- **id\_bar (PK)**: Identificador únic del bar.
- **nom**: Nom comercial del bar.
- **direccio**: Adreça del local.
- **barri**: Zona o districte on es troba ubicat.
- **horari**: Horari d'obertura i tancament.

##### Tapa

- **id\_tapa (PK)**: Identificador únic de la tapa.
- **id\_bar (FK)**: Bar que serveix la tapa.
- **id\_consumicio (FK)**: Referència a la consumició on es demana.
- **nombre**: Nom de la tapa.
- **descripcio**: Breu descripció del plat.
- **preu\_base**: Preu base sense complements.

##### Ingredient

- **id\_ingredient (PK)**: Identificador únic de l'ingredient.
- **id\_tapa (FK)**: Tapa on apareix l'ingredient.
- **nom**: Nom de l'ingredient.
- **es\_alergen**: Booleà que indica si pot causar al·lèrgies.

##### Alergen

- **id\_alergen**: Identificador de l'al·lèrgen.
- **nom**: Nom de l'al·lèrgen (ex: gluten, ou).
- **efectes**: Breu explicació dels seus efectes.

## Client

- **id\_cliente (PK)**: Identificador del client.
- **nom**: Nom del client.
- **mail**: Correu electrònic.
- **genere**: Gènere del client (si es registra).
- **es\_alergen**: Booleà per indicar si té alguna al·lèrgia.

## Consumicio

- **id\_consumicio (PK)**: Identificador de la comanda o consumició.
- **id\_bar (FK)**: Bar on es fa la consumició.
- **id\_cliente (FK)**: Client que realitza la comanda.
- **preu\_total**: Import total de la consumició.
- **data\_hora**: Data i hora de la comanda.
- **nota**: Comentaris o puntuació.

## InfoConsumicio

- **id\_consumicio (PK)**: Referència a la consumició.
- **id\_tapa (PK)**: Tapa que forma part de la comanda.
- **quantitat**: Quantitat d'unitats d'aquesta tapa en la consumició.

## 3.2 Relacions i Cardinalitats

- **Bar  $\longleftrightarrow$  Tapa (1:N)**  
*Explicació:* Un bar pot oferir moltes tapes diferents, però cada tapa està associada a un únic bar concret. Relació de "un a molts".
- **Bar  $\longleftrightarrow$  Consumicio (1:N)**  
*Explicació:* Un bar pot registrar moltes consumicions fetes en el seu establiment. Cada consumició està vinculada a un únic bar. Relació de "un a molts".
- **Tapa  $\longleftrightarrow$  Ingredient (1:N)**  
*Explicació:* Cada tapa pot contenir diversos ingredients, però un ingredient concret només pertany a una tapa en aquest model. Relació de "un a molts".
- **Ingredient  $\longleftrightarrow$  Alergen (N:M)**  
*Explicació:* Un ingredient pot contenir diversos al·lèrgens (ex: ou i gluten), i un al·lèrgens pot aparèixer en múltiples ingredients. Relació de "molts a molts".
- **Client  $\longleftrightarrow$  Alergen (N:M)**  
*Explicació:* Un client pot ser al·lèrgic a diversos al·lèrgens, i un al·lèrgens pot afectar múltiples clients. Relació de "molts a molts".

- **Client**  $\longleftrightarrow$  **Consumicio** (1:N)

*Explicació:* Un client pot fer diverses consumicions al llarg del temps, però cada consumició està associada a un sol client. Relació de "un a molts".

- **Consumicio**  $\longleftrightarrow$  **InfoConsumicio** (1:N)

*Explicació:* Una consumició pot incloure diverses tapes (amb quantitats diferents), però cada entrada d'InfoConsumicio està vinculada a una sola consumició. Relació de "un a molts".

- **InfoConsumicio**  $\longleftrightarrow$  **Tapa** (N:1)

*Explicació:* Una entrada d'InfoConsumicio fa referència a una única tapa, però una mateixa tapa pot aparèixer en moltes consumicions. Relació de "molts a un".

### 3.3 Model E-R:

A continuació es mostren els diagrames dels nostres models Entitat-Relació, un primer prototip que vam fer a paper, i un model polit en PlantUML que permet la modificació i ampliació senzilla del model.

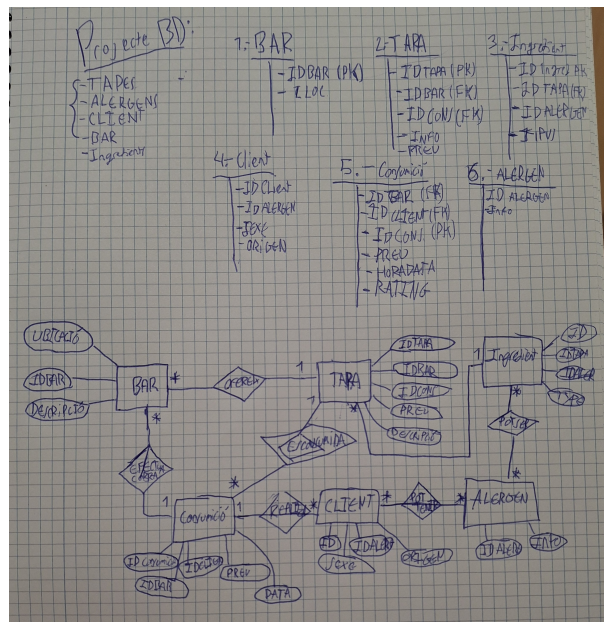


Figura 1: Prototip-ER-1 a paper

El següent diagrama mostra el model Entitat-Relació generat amb PlantUML, on es detallen les entitats clau del nostre sistema i les seves relacions. Aquest model és l'evolució del prototip inicial i està dissenyat per facilitar futures modificacions.

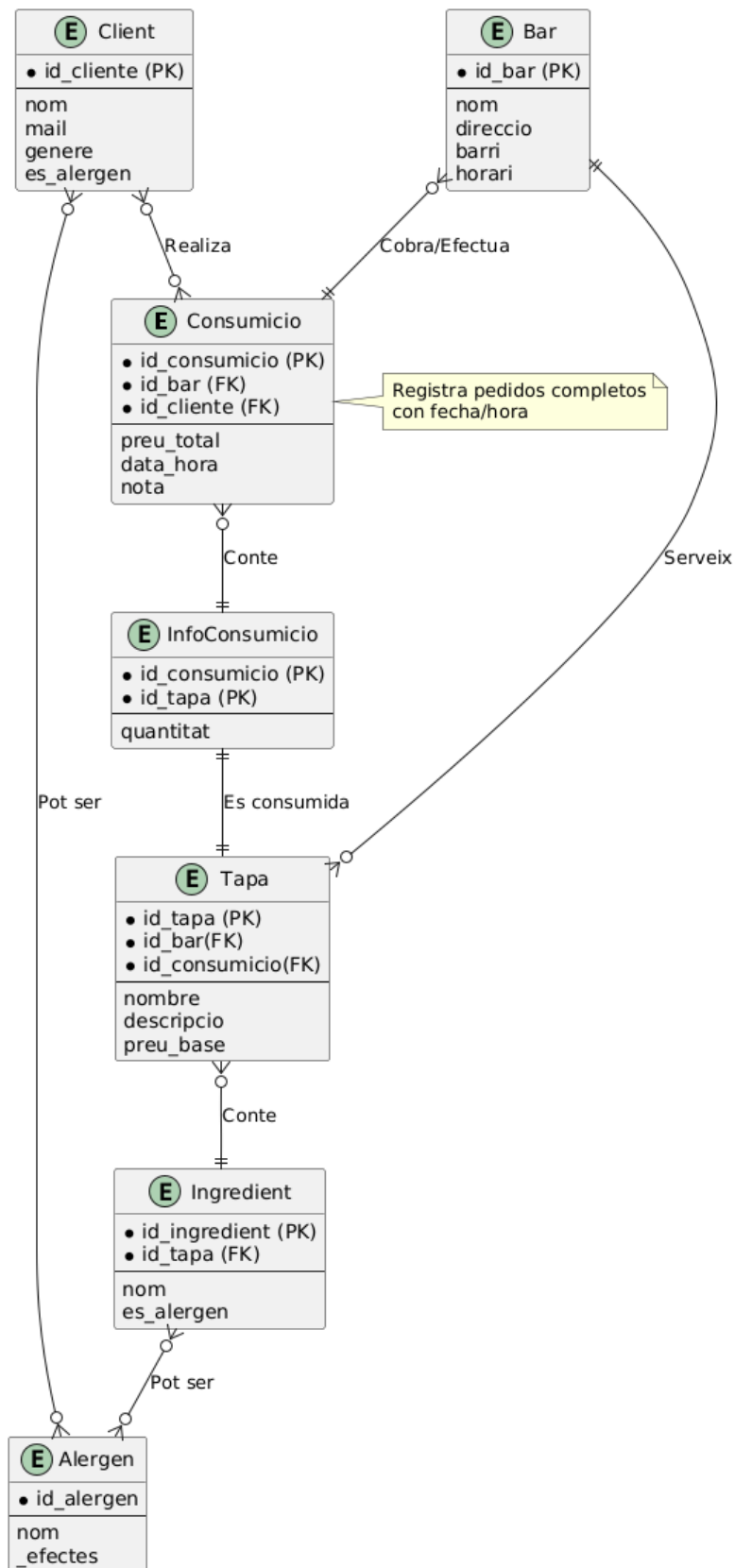


Figura 2: Diagrama E-R generat amb PlantUML

## 4 Recol·lecció i Neteja de Dades

Per a alimentar la nostra base de dades de bars i tapes de Barcelona, hem seguit una metodologia estructurada que inclou diverses fonts de dades i un procés rigorós de neteja. Això ens ha permès garantir la qualitat i coherència de la informació recopilada.

### 4.1 Fonts de Dades

Les fonts de dades utilitzades en aquest projecte inclouen:

- **Plataformes de Reviews (Scraping Inicial):**
  - **TripAdvisor, TheFork, i Google Maps:** Es va extreure informació sobre els noms de bars, direccions, puntuacions i comentaris sobre tapes específiques (per exemple, *L-Les bombes d'aquest bar són espectaculars*).
  - **Eines:** Python, amb llibreries com BeautifulSoup o Scrapy, per realitzar el web scraping.
  - **Desafiaments:** Identificar i estructurar dades no homogènies (exemple: formats diferents de puntuacions) i filtrar ressenyes falses mitjançant anàlisi de sentiment.
- **Dades de Consum Propi (Manual):**
  - **Registres personals del grup:** Arxius amb dades sobre consumicions pròpies, com les ressenyes detallades de tapes com les *bombes* (*"Bar X: Bombes 9/10 – crosta cruixent, farcit cremós"*).
  - **Format:** Excel/CSV amb camps estructurats (nom\_bar, tapa, puntuació, data\_consumicio).
  - **Enquestes a amics:** Recollida de dades mitjançant enquestes sobre hàbits de consum d'amics propers (exemple: *"Quants cops has demanat braves aquest any?"*).
- **Pàgines Web de Bars (Scraping):**
  - **Menús online i webs de bars comuns:** Extracció d'informació sobre llistats de tapes, preus i ingredients (per exemple, *Can Paixano, Quimet & Quimet*).
  - **Eines:** Selenium per a pàgines dinàmiques amb JavaScript.
- **Xarxes Socials (Hashtags i Geolocalització):**
  - **Instagram i TikTok:** Cercada de posts amb hashtags com #TapasBarcelona, #BombesBCN, o geolocalitzacions de bars.
  - **Filtrat:** Anàlisi d'imatges i descripcions que mencionin tapes específiques.



## 4.2 Procés de Neteja de Dades

Un cop recollides les dades, hem seguit els següents passos per garantir la seva qualitat i coherència:

- **Estandarització de Formats:**

- Unificació de formats per a camps com preus (€4.50 → 4.50), dates (15/03/2025 → 2025-03-15), i barris (exemple: *Gracia* → *Gràcia*).
- **Eines:** Python amb Pandas i expressions regulars (Regex).

- **Identificació de Duplicats:**

- Eliminació d'entrades duplicades de bars o tapes, com quan el mateix bar es repeteix en diferents fonts (exemple: TripAdvisor i Google Maps).
- **Tècnica:** Comparació de *nom\_bar* i *direccio* mitjançant fuzzy matching per a detectar errors tipogràfics.

- **Gestió de Valors Faltants:**

- En cas de valors mancants (per exemple, el preu d'una tapa), es realitza una imputació mitjançant la mitjana dels preus d'altres tapes de la mateixa categoria (exemple: tapes *braves* = €4-6).
- Si un bar no té horari, es marcarà com *No disponible* per evitar inconsistències.

- **Validació d'Al·lèrgens:**

- Creació d'un diccionari d'ingredients i al·lèrgens (per exemple, *patata* → sense gluten) per a tapes amb etiquetatge poc clar.
- **Font Secundària:** Bases de dades públiques com FENIL per a validar ingredients comuns.

- **Geocodificació:**

- Conversió d'adreces de bars a coordenades GPS (latitud, longitud) mitjançant l'API de Google Maps o OpenStreetMap.

## 4.3 Futur: Plataforma Col·laborativa

En una fase posterior del projecte, es preveu/ ens agradaria implementar una plataforma on els usuaris puguin:

- Afegir nous bars i tapes mitjançant formularis estandaritzats.
- Puntuar consumicions i verificar al·lèrgens.
- Corregir errors mitjançant un sistema de revisió comunitària.

Aquesta plataforma permetrà crear una base de dades rica, precisa i útil, tant per a amants de les tapes com per a persones amb restriccions alimentàries.

## 5 Avanç de Consultes Mitjançant l'Àlgebra Relacional

Gràcies a la base de dades dissenyada, podem realitzar diverses consultes per obtenir informació rellevant sobre els consumos realitzats als bars de Barcelona. A continuació es presenten alguns exemples de consultes que es podrien fer, utilitzant l'àlgebra relacional per expressar-les de manera formal.

### 5.1 Consulta 1: Quin usuari ha consumit més en un bar?

Per a realitzar aquesta consulta, necessitem identificar quin client (usuari) ha consumit més quantitat en un determinat bar. La consulta es basarà en la taula 'Consumicio' per obtenir les quantitats consumides i l'usuari associat.

**Formula en àlgebra relacional:** Suposant que volem saber quin client ha consumit més en el bar amb l'id\_bar = 1:

$$\text{Resultat} = \pi_{\text{id\_cliente}, \text{quantitat}} (\sigma_{\text{id\_bar}=1}(\text{InfoConsumicio}) \bowtie \text{Consumicio})$$

Aquesta consulta filtra les consumicions per un determinat bar i després projecta els resultats en funció de l'id\_cliente i la quantitat consumida. Per trobar qui ha consumit més, caldria realitzar una agregació posterior.

### 5.2 Consulta 2: Quin bar ha tingut més consumicions?

Per a saber quin bar ha tingut més consumicions, podem realitzar una agregació de les comandes fetes a cada bar. Aquesta consulta farà ús de les taules 'Consumicio' i 'InfoConsumicio' per obtenir la informació relacionada amb els bars.

**Formula en àlgebra relacional:** Per obtenir el nombre de consumicions per a cada bar, fem:

$$\text{Resultat} = \gamma_{\text{id\_bar}, \text{count}(\text{id\_consumicio})}(\text{Consumicio})$$

Aquesta consulta retorna el nombre de consumicions per cada bar. Per a identificar el bar amb més consumicions.

### 5.3 Consulta 3: Quin és l'ingredient més consumit?

Per a saber quin és l'ingredient més consumit, necessitem fer una agregació de les quantitats de cada ingredient present a les consumicions. Aquesta consulta implicarà la taula 'Ingredient' i 'InfoConsumicio', i caldria comptabilitzar quins ingredients apareixen més sovint.

**Formula en àlgebra relacional:** Per a trobar l'ingredient més consumit, podríem utilitzar la següent expressió:

$$\text{Resultat} = \gamma_{\text{id\_ingredient}, \text{count}(\text{id\_tapa})}(\text{Ingredient} \bowtie \text{InfoConsumicio})$$

Aquesta consulta fa una unió entre les taules 'Ingredient' i 'InfoConsumicio', i després comptabilitza el nombre d'aparicions de cada ingredient. La resposta serà l'ingredient més freqüent.

## 6 Conclusions

En finalitzar aquesta primera fase del nostre projecte sobre el consum de tapes als locals barcelonins, hem establert una base sòlida per a la gestió i anàlisi de dades relacionades amb la gastronomia tradicional de la ciutat. Mitjançant el disseny d'un model Entitat-Relació detallat, hem definit les estructures necessàries per catalogar bars, tapes, ingredients i consumidors, així com les seves relacions.

Aquesta etapa inicial ens ha permès identificar les fonts de dades més rellevants —des de plataformes de reviews fins a registres personals— i establir un procés rigorós de neteja i estandardització per garantir la qualitat de la informació. A més, hem avançat en la definició d'àlgebra relacional per a consultes futures, com identificar els clients més freqüents, els bars amb més consum o els ingredients més populars.

El proper pas serà la implementació de la base de dades, que ens permetrà explotar aquesta informació per a:

- Preservar la cultura gastronòmica local, donant visibilitat als bars autèntics i les seves tapes.
- Oferir eines útils per a persones amb restriccions alimentàries, com celíacs o vegetarians.
- Analitzar patrons de consum (preferències per barris, horaris o estacionalitat) i tendències emergents.

El nostre objectiu final és que aquesta base de dades no només serveixi com a arxiu de la tradició culinària barcelonina, sinó també com a recurs per a consumidors i professionals del sector, fomentant decisions informades i contribuint a la sostenibilitat dels bars tradicionals en un context de canvis gastronòmics globals.

Amb aquests fonaments, estem preparats per avançar cap a la construcció d'un sistema funcional i escalable, que reflecteixi la riquesa de les tapes com a element identitari de Barcelona.

### 6.1 Aportacions:

El desenvolupament d'aquest projecte sobre el consum de tapes a Barcelona ha estat possible gràcies a les contribucions específiques de cada membre de l'equip:

- **Pau Queralt Muñoz:** S'ha encarregat de la redacció principal del document, participant activament en el disseny del model Entitat-Relació i en l'estructuració global del projecte.
- **Alberto Moreno Martinez:** Ha liderat la recollida inicial de dades, a més de contribuir significativament en la redacció dels continguts i en la ideació dels objectius del projecte.
- **Gerard Fideu Garcia:** Com a redactor secundari, ha participat en l'elaboració dels continguts i ha estat responsable del disseny i programació del diagrama UML que representa el model conceptual.

Aquest treball col·laboratiu ha permès crear una base sòlida per a l'anàlisi dels hàbits de consum de tapes a Barcelona, amb un model ben definit que permetrà futures ampliacions. Els pròxims passos se centraran en la implementació pràctica de la base de dades i en l'anàlisi i neteja de les dades recollides.

## 7 Anexos

### 7.1 Repositori GitHub

Tots els arxius del projecte, incloent-hi el codi font del model Entitat-Relació, les dades recopilades i altres elements relacionats, es poden trobar al nostre repositori de GitHub.

A continuació, proporcionem l'enllaç per accedir-hi:

<https://github.com/Pantanosso/BDProjectBombas>

Aquest repositori conté:

- El codi complet del model Entitat-Relació generat.
- Dades recopilades durant la fase inicial del projecte.
- Scripts i eines utilitzades per a la neteja i anàlisi de les dades.
- Documentació relacionada amb la implementació de la base de dades.