



This is the **published version** of the bachelor thesis:

Marcos Martínez, Gabriel; Antens, Coen Jacobus, tut. Análisis de datos mediante Business Intelligence : Un estudio de casos prácticos. 2025. (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/317525>

under the terms of the  license

Anàlisi de dades mitjançant Business Intelligence: Un estudi de casos pràctics

Gabriel Marcos Martínez

Resum— Aquest projecte s'inicia amb l'estudi i anàlisi de tiquets de compra del supermercat Mercadona, amb l'objectiu d'obtenir coneixement sobre els hàbits dels consumidors, identificar patrons de compra recurrents, anomalies i segmentar clients segons les seves preferències. Per aconseguir-ho, s'han utilitzat eines com Power BI i Python, que permeten extreure, netejar, transformar i visualitzar dades de forma eficient. Aquesta fase introdueix tècniques bàsiques d'anàlisi sobre conjunts de dades estructurades i de mida reduïda. A mesura que el projecte evoluciona, l'atenció es desplaça cap al sector musical, on s'empren datasets públics de gran volum provinents de plataformes com Spotify i Kaggle. Mitjançant aquests conjunts de dades, es poden analitzar tendències musicals, característiques sonores i comportaments d'usuari en la creació de playlists. Aquesta segona part culmina amb la implementació d'un sistema recomanador híbrid, que integra tant el filtratge col·laboratiu com les metadades musicals. El fil conductor del projecte és l'exploració del potencial de l'anàlisi de dades per transformar informació en coneixement útil, aplicable a àmbits molt diversos.

Paraules clau— Tiquets de compra, Mercadona, Power BI, anàlisi de dades, Python, Spotify, Kaggle, música, datasets massius, tendències musicals, recomanador híbrid, filtratge col·laboratiu, metadades, visualització de dades, aprenentatge automàtic.

Abstract—This project begins with the analysis of purchase tickets from the supermarket Mercadona, aiming to uncover insights about consumer habits, detect purchase patterns, identify anomalies, and segment users by behavior. Tools such as Power BI and Python are employed to extract, clean, transform, and visualize data effectively. This initial phase focuses on small, structured datasets and basic data analytics techniques. As the project progresses, attention shifts to the music industry, using large-scale public datasets from platforms like Spotify and Kaggle. These datasets enable the study of musical trends, sound features, and user behavior in playlist creation. The second part culminates in the development of a hybrid recommender system that combines collaborative filtering with musical metadata. The project as a whole explores how data analysis and machine learning can generate actionable knowledge across diverse domains.

Index Terms— Purchase receipts, Mercadona, Power BI, data analysis, Python, Spotify, Kaggle, music, large datasets, music trends, hybrid recommender system, collaborative filtering, metadata, data visualization, machine learning.

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

Vivim en una era on les dades tenen un paper clau en la presa de decisions. Empreses com Netflix analitzen milions de visualitzacions per recomanar continguts personalitzats als seus usuaris. Glovo, per la seva banda, optimitza les rutes de repartiment i anticipa la demanda utilitzant dades en temps real. Supermercats com LIDL i Mercadona aprofiten l'anàlisi de tiquets de compra per detectar patrons de consum i ajustar estratègies de màrqueting i logística.

Aquestes organitzacions tenen un denominador comú: usen el **Business Intelligence (BI)** per transformar dades brutes en coneixement útil. El BI no es limita a la creació d'informes, sinó que inclou l'anàlisi predictiva, la visualització interactiva i l'ús de tècniques d'intel·ligència artificial per optimitzar processos i millorar el rendiment.

Aquest projecte es divideix en dues fases: una primera d'anàlisi comercial a partir de tiquets de compra, i una segona centrada en dades musicals per construir un sistema recomanador basat en Spotify i altres fonts públiques.

Segons Gartner, el Business Intelligence (BI) és un conjunt d'eines, infraestructures i pràctiques que permeten analitzar dades per millorar decisions i optimitzar el rendiment empresarial [1]. Més enllà de la tecnologia, el BI implica una estratègia que integra processos i persones.

Forbes i Power BI destaquen com aquestes eines ajuden sectors com el retail a entendre millor el comportament dels clients i adaptar les estratègies comercials [2]. A més, plataformes com Power BI i Tableau han contribuït a fer l'anàlisi de dades més accessible, democratitzant-ne l'ús entre perfils no tècnics [3].

- E-mail de contacte: gabriel.marcos@uab.cat
- Menció realitzada: Enginyeria del Software
- Treball tutoritzat per: Coen Antens (Àrea de Ciències de la Computació)
- Curs 2024/25

2 PROPOSTA

Amb aquesta proposta es pretén mostrar com l'aplicació de tècniques de Business Intelligence pot aportar valor en diferents àmbits mitjançant l'anàlisi de dades. A continuació es detallen els objectius, l'abast i les possibles aplicacions del projecte.

2.1 Objectiu

El projecte planteja dues línies d'anàlisi diferenciades però complementàries, amb l'objectiu d'extreure informació rellevant a partir de dades estructurades i no estructurades mitjançant tècniques de Business Intelligence (BI).

En primer lloc, l'estudi de tiquets de compra de supermercats, amb l'objectiu d'identificar patrons i comportaments de consum. Aquest procés inclou la digitalització automatitzada dels tiquets mitjançant tècniques de visió per computador i l'anàlisi de les dades amb eines com Power BI per obtenir informació útil per a la presa de decisions. Aquest anàlisi es realitza a petita escala a través de dades d'un únic usuari al llarg del temps.

En segon lloc, l'anàlisi de dades musicals obtingudes de Spotify, centrada en les característiques de les cançons com el tempo, l'energia, la valència o l'acústica. A partir d'aquestes dades, es pretén estudiar l'evolució de la música al llarg del temps i dissenyar un sistema recomanador de cançons similars basat en aquestes característiques, sense necessitat d'informació directa dels usuaris. Per a aquest cas l'anàlisi té un volum superior a l'anterior, obtenint dades de datasets públics.

2.2 Abast i objectius específics

- Digitalitzar i estructurar les dades dels tiquets de compra mitjançant **Visió per Computador**¹.
- Explorar les capacitats de diferents eines de BI com Power BI per crear dashboards i informes interactius.
- Dissenyar un sistema recomanador de cançons basat en similituds detectades a través de l'anàlisi de dades musicals.
- Analitzar la viabilitat d'aplicar aquests processos i metodologies a altres àmbits i sectors, demostrant la gran escalabilitat que permeten aquestes eines.

2.3 Exploració en altres sectors

Tot i que el focus inicial d'exploració estarà centrat en l'anàlisi de tiquets de compra de supermercats i dades musicals de Spotify, les eines i metodologies que utilitzarem per fer-ho són escalables i poden aplicar-se en altres sectors, per exemple:

- Streaming (Netflix, Spotify, YouTube): L'anàlisi de reproduccions ajuda a entendre els usuaris i millorar les recomanacions.
- E-commerce i màrqueting digital: Plataformes com Amazon o Shopify usen el BI per ajustar estratègies de

venda segons el comportament dels clients.

- Transport i logística: Empreses com Uber o Glovo analitzen rutes, temps i costos per ser més eficients.

3 METODOLOGIA

Per garantir un desenvolupament òptim i flexible del projecte, s'ha optat per usar Kanban com a metodologia de treball, implementada mitjançant l'eina Trello per al seguiment i gestió de tasques. Aquesta elecció es basa en la naturalesa oberta i escalable del projecte, que requereix una organització adaptable a mesura que es vagin obtenint resultats i conclusions.

3.1 Enfocament de treball

El projecte ha seguit un enfocament flexible i iteratiu, sense partir d'un objectiu tancat ni d'un conjunt de requisits estrictament definits. Aquesta metodologia ha permès adaptar les tècniques i eines d'anàlisi de dades en funció dels resultats obtinguts i les necessitats que han anat sorgint durant el desenvolupament.

En la primera fase, centrada en l'anàlisi de tiquets de compra de supermercats, es va començar amb un nombre reduït de documents (2 tiquets), per després ampliar la mostra a 15 i finalment a 45, permetent una anàlisi més robusta i completa.

La segona fase del projecte es va enfocar en dades extretes de Spotify. Tot i que el procés de neteja i normalització de les dades també va implicar un treball important, el pas cap a la seva visualització i anàlisi va resultar més àgil gràcies a l'experiència prèvia adquirida. Aquesta part ha permès estudiar l'evolució musical en base a característiques tècniques de les cançons i iniciar el desenvolupament d'un sistema recomanador que identifiqui similituds entre temes.

3.2 Eines per a la gestió de projectes

- **Trello**: és una eina basada en Kanban que permet organitzar tasques en llistes i targetes. És intuïtiva i facilita la col·laboració entre equips.
- **Microsoft Project**: ideal per a la gestió de projectes grans, permet la planificació de tasques, recursos i pressupostos amb eines avançades d'anàlisi.

3.3 Eines i tecnologies utilitzades

Les eines seleccionades per al desenvolupament del projecte són les següents:

- **Python** per al processament de dades, fent especial èmfasi en l'ús de la llibreria **pandas** per extreure informació dels tiquets en format digital.
- **Power BI** per l'anàlisi, visualització i creació d'informes interactius.
- **Trello** per a la gestió de tasques i organització del projecte i Microsoft Projects pels Diagrames de Gantt, veure Il·lustració 25 (Apèndix 3).

¹ La **Visió per Computador** és una branca de la intel·ligència artificial que permet a les màquines interpretar i extreure informació de dades visuals

com imatges o vídeos.

4 ESTAT DE L'ART

L'estat de l'art recull els coneixements i avenços actuals relacionats amb les tecnologies i metodologies clau del projecte. En aquest apartat es presenten els fonaments del Business Intelligence i el funcionament del sistema recomanador de Spotify.

4.1 Definició de Business Intelligence

El Business Intelligence (BI) es defineix com el conjunt de mètodes, processos i tecnologies que permeten a les organitzacions recopilar, analitzar i transformar dades en informació útil per a la presa de decisions. I aquest ha evolucionat al llarg dels anys:

- **Anys 60-80:** En aquesta etapa inicial, els sistemes de suport a la decisió (DSS) ajudaven en tasques bàsiques de processament de dades. Tot i les limitacions, aquests sistemes van establir les bases per a l'ús de les dades en la presa de decisions [4].
- **Anys 90:** S'introdueixen els Data Warehouses i les empreses comencen a emmagatzemar les dades en repositoris centralitzats. Això permet una millor integració i accés als grans volums d'informació, fent els anàlisis més fàcils de forma sistemàtica [5].
- **Anys 2000:** Apareixen els primers sistemes de BI com els coneixem en l'actualitat, oferint una millora en les visualitzacions i una experiència d'usuari intuïtiva.
- **Anys 2010-Actualitat:** El BI està integrat amb el Big Data, IA i Machine Learning, esdevenint molt més potent l'anàlisi i el processament de dades. Les plataformes actuals són capaces de predir comportaments futurs i automatitzar recomanacions mitjançant models entrenats amb dades de múltiples fonts [5].

El creixement del Big Data i Machine Learning ha portat a fer més potents les eines ja existents de BI, fent-les capaces de predir comportaments futurs a través de models entrenats amb les dades, automatitzar recomanacions a través d'algoritmes i de gestionar grans volums de dades provinents de diverses fonts d'informació.

4.2 Sistema Recomanador de Spotify

El sistema recomanador de Spotify integra diverses tècniques per oferir suggeriments musicals segons els gustos dels usuaris, fonamentant-se en tres components.

En primer lloc, Spotify genera un mapa de proximitat on cada punt representa una cançó o artista, i la distància entre punts indica el grau de similitud. Aquest mapa es construeix a partir de metadades (nom, artista, gènere) i atributs musicals quantitatius, com el ritme, l'energia, el volum i la tonalitat. Per millorar la precisió i evitar agrupacions incorrectes de temes amb continguts similars, però estils diferents, s'aplica un filtratge basat en el contingut, que analitza en profunditat les característiques pròpies de cada peça [9].

² **Processament de Llenguatge Natural** (Natural Language Processing, NLP) és un conjunt de tècniques d'intel·ligència artificial que permeten a les màquines interpretar, analitzar i generar llenguatge humà.

En segon lloc, el sistema fa servir tècniques de processament de llenguatge natural² (NLP) per analitzar textos associats a les cançons i artistes, recollits de fonts com articles, blogs i ressenyes [9][10].

Finalment, s'incorpora un model d'anàlisi d'àudio directe que extreu característiques objectives del senyal sonor, com la durada, tempo i altres mètriques acústiques, facilitant la classificació i agrupació quan la informació textual és limitada [10].

Aquest conjunt de models s'integra en un sistema híbrid, que es refina constantment amb dades reals d'interacció dels usuaris (playlists, reproduccions, salts), per generar recomanacions personalitzades i de gran precisió [9][10].

5 DESENVOLUPAMENT – ESTUDI DE TIQUETS

Abans d'analitzar en detall el procés tècnic d'extracció i tractament de dades, és útil observar un exemple real d'un tiquet de Mercadona. La Il·lustració 1 permet identificar visualment els camps clau com la data, hora, descripció dels productes, imports i mètode de pagament, que són essencials per a les etapes posteriors del projecte.

5.1 Extracció de dades des de PDF



MERCADONA, S.A. A-46103834

C/ VERGE DE L' ASSUMPCIÓ 32

08210 BARBERA DEL VALLES

TELÉFON: 937191004

27/01/2024 13:36 OP:386433

FACTURA SIMPLIFICADA: 3228-020-674932



Descripció	P. Unit	Import
1 BROU CASOLÀ DE PEIX		1,15
1 BOSSA PLÀSTIC		0,15
1 FLOR. BROCOLI		1,52
1 CEBA TUB		1,85
1 ALL SEC 400 G		2,10
2 ALLIOLI TERRINA	1,35	2,70
1 FLAN QUESO		2,00
1 BEURE F.SILVESTRE 0%		1,45
TOTAL (€)		12,92
TARGETA BANCÀRIA		12,92
IVA	BASE IMPOSABLE (€)	QUOTA (€)
10%	6,64	0,66
21%	0,12	0,03
0%	5,47	0,00
TOTAL	12,23	0,69

Il·lustració 1: Exemple de tiquet

El procés d'extracció de dades va començar llegint tiquets en format PDF amb la llibreria **pdfplumber**³, que permet obtenir el text pàgina a pàgina amb alta precisió. Aquesta informació textual es va processar per generar un fitxer **CSV**⁴ estructurat, amb camps com data, hora, productes, imports, IVA, pàrquing i mètode de pagament, que va servir com a base per les fases següents del projecte.

³ **pdfplumber**: Biblioteca de codi obert per a Python que facilita l'extracció de text, taules i metadades de fitxers PDF [4].

⁴ **CSV**: Format de fitxer de text en pla que permet emmagatzemar dades tabulades separant cada valor per comes.

5.2 Data Curation - Problemes en l'extracció de dades des de PDF

Malgrat l'ús de la llibreria **pdfplumber**³, es van detectar diversos reptes derivats de la variabilitat dels formats dels tiquets.

1. Extracció de la informació de pàrquing:

El codi original delimitava la secció de productes entre les línies "Descripció P. Unit Import" i "ENTRADA". En tiquets sense informació de pàrquing, on "ENTRADA" no apareixia, la secció quedava buida i el tiquet no es desava correctament. Per solucionar-ho, es va ampliar el criteri de límit acceptant també "TOTAL" com a marcador de tancament. A més, es va afegir una comprovació per garantir la captura del contingut encara que no s'ajustés al patró inicial (veure II-Il·lustració 2).

```
for i, line in enumerate(lines):
    if "Descripció P. Unit Import" in line:
        prod_start = i + 1
    if prod_start is not None and (line.startswith("ENTRADA")
    or line.startswith("TOTAL")):
        prod_end = i
        break
prod_lines = lines[prod_start:prod_end] if prod_start is not None else []
```

Il·lustració 2: Codi acceptació "ENTRADA" i "TOTAL"

2. Extracció de l'adreça:

El codi inicial reconeixia només adreces que començaven per "C/", excloent altres formats comuns com "C.C." o "CTRA". Aquesta limitació causava pèrdua d'informació rellevant. La solució va consistir a ampliar la lògica per capturar qualsevol línia que contingui una adreça significativa, independentment del prefix.

5.3 Data Curation - neteja i estructuració de la informació

Després d'extreure les dades dels tiquets i guardar-les en un fitxer CSV, s'ha realitzat un procés de neteja per garantir la seva coherència i compatibilitat amb Power BI, a la II-Il·lustració 3 observem un exemple. Aquest procés ha estat essencial per assegurar la qualitat i homogeneïtat de les dades.

S'han normalitzat els noms dels productes, corregint variacions en majúscules, accents i abreviatures (com ara "bròcoli", "brocoli", "BROU CASOLÀ", "brou casolà", etc.), i s'han unificat els noms en castellà per facilitar l'anàlisi de dades de Barbastro. També s'han gestionat els valors nuls per evitar errors en les visualitzacions.

Descripció	P. Unit	Import
1 ROTLLE LLAR DOBLE		2,25
1 HIGIÈNIC DOBLE ROTLL		4,40
1 BEURE FRUITES BOSC		2,05
1 BEURE F.SILVESTRE 0%		1,45
2 PAPER HUMIT WC	1,55	3,10
1 BÍFIDUS NATURAL		1,35
1 BÍF FRUITES SIL		1,30

Il·lustració 3: Exemple noms diferents

Altres ajustos que s'han fet per facilitar la creació d'algunes gràfiques posteriorment han sigut:

- **Separació de Data i Hora:** Divisió en dues columnes per una millor manipulació temporal.
- **Concatenació d'Adreça i Ciutat:** Creació d'un camp combinat per facilitar l'anàlisi geogràfica.
- **Normalització de les hores de compra:** Agrupament en intervals de 30 minuts per identificar patrons horaris.
- **Assignació de noms setmanals:** Creació d'un camp per establir el nom del dia de la setmana per cada data.
- **Càlcul del temps d'estada al pàrquing:** Diferència entre hora d'entrada i sortida per mesurar el temps d'estada dels clients.

5.4 Enriquiment amb dades nutricionals

Per millorar les dades recollides, s'ha incorporat informació nutricional de manera separada per facilitar el seu anàlisi a Power BI.

El primer pas ha estat normalitzar els noms dels productes per garantir una identificació clara a la base de dades nutricional. Després de la normalització, s'han seleccionat dues columnes clau: el nom del producte i el preu unitari (o l'import total si no està disponible). Aquests camps són essencials per identificar cada producte i facilitar la cerca a la base de dades nutricional, en tenim un exemple del format a la II-Il·lustració 4.

```
BROU CASOLÀ DE PEIX,"1,15"
FLOR. BROCOLI,"1,52"
CEBA TUB,"1,85"
ALL SEC 400 G,"2,10"
ALLIOLI TERRINA,"1,35"
FLAN QUESO,"2,00"
BEURE F.SILVESTRE 0%,"1,45"
PROT. NORMAL,"1,85"
SOPA POLLASTRE FIDEU,"1,00"
ALBERGÍNIA,"1,99"
PEBROT VERMELL,"2,69"
BEGUDA AVENA,"0,95"
NABIU 500 GR,"6,49"
SALAMI PACK-4,"1,97"
BÍFIDUS CREMÓS COCO,"1,30"
BÍFIDUS FIBRES I CER,"1,30"
```

Il·lustració 4: Mostra Productes i Preus

Un cop disponibles les dades bàsiques (nom del producte i preu unitari), s'inicia la integració amb l'eina **mercaGPT**⁵, que obté automàticament la informació nutricional de cada producte, principalment a partir del nom i, en alguns casos, també considerant el preu unitari per reforçar la identificació. Per optimitzar el procés i evitar sobre càrregues, les dades s'envien en tandes, assegurant una execució més eficient. L'eina retorna la informació nutricional completa per a cada producte.

⁵ **mercaGPT**: Eina desenvolupada mitjançant intel·ligència artificial per ajudar a millorar els hàbits de consum mitjançant la consulta de dades

nutricionals dels productes de Mercadona[5].

Finalment, tota aquesta informació s'emmagatzema en un nou fitxer enriquit, que conté tant les dades originals com els valors nutricionals detallats. Els camps inclosos són: **Producto, Precio (€/unidad), Calorías (kcal), Grasas (g), Grasas Saturadas (g), Hidratos (g), Azúcares (g), Proteínas (g), Sal (g), Sin Gluten i Sin Lactosa**. El resultat ho veiem amb una mostra a la Il·lustració 5.

BROU CASOLÀ DE PEIX,1.15,15,0.1,0.0,0.5,0.3,2.1,0.7,Si,Si
 FLOR. BRÒCOLI,1.52,34,0.4,0.1,7.2,1.7,2.8,0.02,Si,Si
 CEBA TUB,1.85,40,0.1,0.0,9.3,4.7,1.2,0.01,Si,Si
 ALL SEC 400 G,2.1,149,0.5,0.1,33.0,1.0,6.4,0.02,Si,Si
 ALLIOLI TERRINA,1.35,700,75.0,11.0,2.5,1.0,2.0,1.2,No,No
 FLAN QUESO,2.0,120,3.5,2.1,18.5,16.2,3.2,0.1,No,No
 BEURE F.SILVESTRE 0%,1.45,46,0.1,0.0,11.0,10.0,0.2,0.0,Si,Si
 PROT. NORMAL,1.85,80,2.3,0.9,9.2,8.4,5.1,0.1,No,No
 SOPA POLLASTRE FIDEU,1.0,48,1.2,0.4,7.8,1.5,2.3,0.8,No,No
 ALBERGÍNIA,1.99,25,0.2,0.0,5.7,3.4,1.2,0.01,Si,Si
 PEBROT VERMELL,2.69,31,0.3,0.0,6.0,4.2,1.0,0.02,Si,Si
 BEGUDA AVENA,0.95,42,1.4,0.2,6.6,4.0,0.3,0.1,Si,Si
 NABIU 500 GR,6.49,57,0.3,0.0,14.0,10.0,0.7,0.01,Si,Si
 SALAMI PACK-4,1.97,410,37.0,15.0,2.5,0.9,17.0,3.2,No,Si
 BÍFIDUS CREMÓS COCO,1.3,92,4.5,3.2,8.5,6.2,4.1,0.1,No,No
 BÍFIDUS FIBRES I CER,1.3,78,3.2,1.8,10.2,7.5,3.6,0.1,No,No
 BÍFIDUS NATURAL,1.4,62,3.0,1.7,6.8,5.1,3.5,0.1,No,No
 PATATES BRAVES ALLIO,2.5,260,14.0,2.5,32.1,2.3,3.8,1.5,No,No
 CRUNCHY CHICKEN,2.15,320,18.0,4.8,22.5,1.2,21.0,1.8,No,No

Il·lustració 5: Mostra de dades nutricionals

5.5 Anàlisi i visualització amb Power BI

Després d'obtenir les dades de compres i les nutricionals, es preparen per a l'anàlisi amb Power BI. Es treballa amb dos fitxers separats: un amb la informació de compres i un altre amb les dades nutricionals.

Despesa Total:

La despesa total es calcula sumant els imports de tots els tiquets recollits, obtenint un total de 6.060 € durant el període analitzat.

Mitjana de despesa per compra:

Per conèixer la despesa mitjana, s'analitzen tots els imports dels tiquets. La Il·lustració 20 (Apèndix 1) mostra algunes compres amb imports alts, útils per detectar excepcions. La despesa mitjana per compra és de 50,46 €, valor que permet entendre els hàbits de consum i fer comparacions entre períodes.

Productes més adquirits:

L'anàlisi permet identificar els productes més adquirits. A la Il·lustració 22 (Apèndix 2) podem observar els tres productes més comprats: Beure F.Silvestre 0%, Beguda avena i Meló pell de gripau.

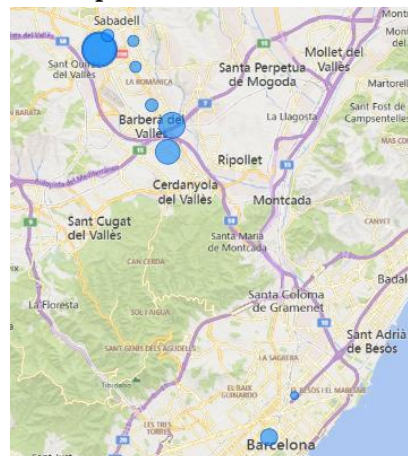
Despesa segons ubicació:

Per analitzar si la ubicació del supermercat influeix en la despesa, es va calcular la mitjana de despesa per compra segons l'adreça.

Els resultats mostren que el Carrer Santa Anna (Cerdanyola) és on es gasta més per compra, seguit del Carrer Viladomat i el Centre Comercial Vía Sabadell (tots dos a Sabadell). Aquesta informació es pot veure a la Il·lustració 21

(Apèndix 1). Aquesta dada ajuda a identificar diferències de preu o hàbits de consum segons la ubicació. A més a més, al mapa de la Il·lustració 6, es visualitzen les zones on s'han fet compres. La mida dels cercles indica el volum de compres en cada establiment: cercles més grans indiquen més compres.

Patrons de compra:

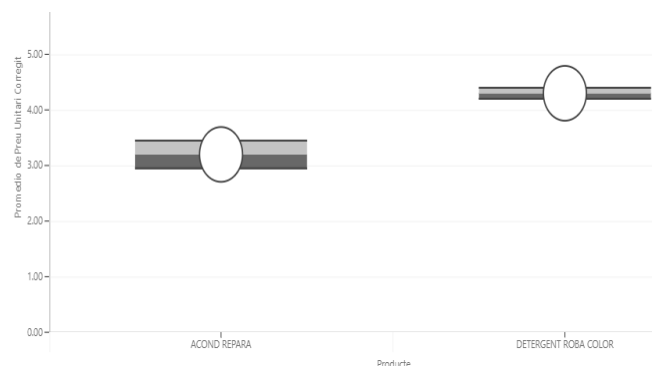


Il·lustració 6: Mapa de compres

Per identificar patrons setmanals de compra, s'han analitzat tres variables: l'hora de la compra (en franges de 30 minuts), el dia de la setmana i l'adreça de l'establiment. A la Il·lustració 23 (Apèndix 2) es pot observar la concentració de compres els divendres, especialment en dues ubicacions: Carrer Santa Anna (Cerdanyola del Vallès) i Carretera Nacional 150 (Barberà del Vallès). Aquest patró podria indicar un hàbit de compra setmanal al final de la setmana.

Evolució dels preus dels productes:

Per estudiar l'evolució dels preus, hem utilitzat un **box plot**⁶ que mostra les variacions al llarg del temps. A la Il·lustració 7, es poden veure els exemples dels productes ACOND REPARA i DETERGENT ROBA COLOR, que han variat de preu en un any. Per exemple, el preu d'ACOND REPARA va passar de 3,45 € a 2,95 €, indicant una lleugera disminució. Tot i que la majoria dels productes tenen preus estables, alguns poden presentar variacions puntuals.



Il·lustració 7: Boxplot ACOND REPARA i DETERGENT ROBA COLOR

⁶ **Boxplot:** Gràfic que mostra la distribució d'un conjunt de dades

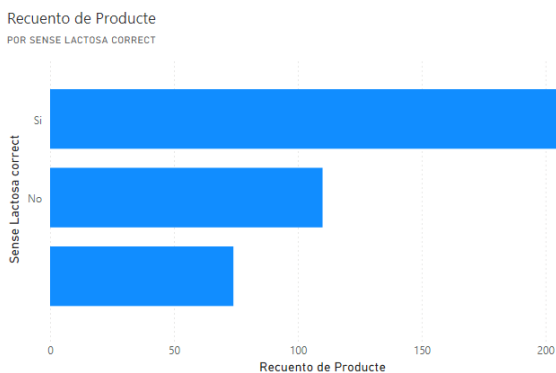
destacant la mediana, els quartils i els valors atípics.

Valors nutricionals dels productes:

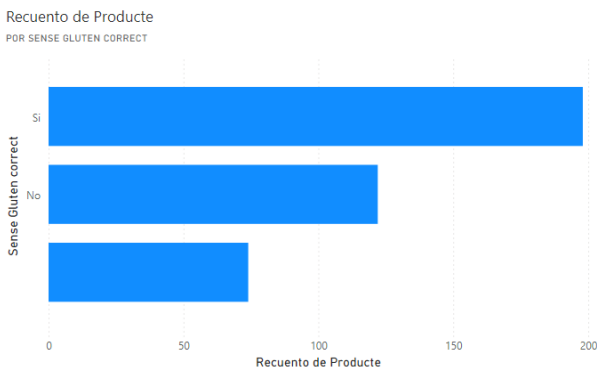
L'anàlisi nutricional s'ha centrat en observacions generals sobre nutrients com sucres, greixos, proteïnes, greixos saturats i calories. Per exemple, a la Il·lustració 24 (Apèndix 3), observem els productes amb més sucre, destacant els sucres, caramels, melmelades, cremes i sucre.

Intoleràncies:

S'ha analitzat la presència de gluten i lactosa als productes. La Il·lustració 8 mostra que 210 productes no contenen lactosa, 110 si en tenen i 74 no estan classificats. A la Il·lustració 9 es veu que 198 productes no contenen gluten, 122 si en tenen i 74 no disposen d'informació.



Il·lustració 8: Productes sense lactosa



Il·lustració 9: Productes sense gluten

6 DESENVOLUPAMENT – ESTUDI DE SPOTIFY

Després d'analitzar dades de tiquets de compra amb Power BI, la fase següent del projecte es centra en el sector musical. L'objectiu és desenvolupar un sistema recomanador basat en la similitud entre cançons, utilitzant característiques musicals rellevants com el tempo o l'energia, extrems de conjunts de dades disponibles.

6.1 Recol·lecció de dades

Inicialment, es va plantejar utilitzar l'API⁷ pública de Spotify per obtenir dades musicals detallades, com característiques acústiques de les cançons i informació relacionada amb artistes i àlbums. No obstant això, les limitacions actuals de l'API impedeixen accedir a certa informació que abans estava disponible. Per això, es va optar per treballar

amb datasets públics, com els que es troben a **Kaggle**⁸ que inclouen dades musicals històriques i atributs acústics necessaris per al nostre model.

- **Dataset 1:** Music Recommendation System using Spotify Dataset [11].
- **Dataset 2:** Spotify Tracks Dataset [12].

6.2 Fusió de dades

Per integrar la informació dels dos datasets seleccionats, es va desenvolupar un codi que els fusiona mitjançant l'identificador de Spotify, permetent combinar els camps disponibles i identificar les entrades amb dades incompletes. Un dels problemes principals va ser la manca de l'any de publicació de moltes cançons. Aquesta dada, essencial per analitzar l'evolució musical, es va decidir recuperar-la utilitzant l'API de Spotify.

Optimització de les consultes a l'API:

Un dels principals reptes va ser la limitació del nombre de consultes per minut a l'API de Spotify. Inicialment, es pensava fer una consulta per cada track_id, però això suposava fins a 100.000 peticions, cosa inviable.

Per optimitzar, es va decidir buscar per nom d'àlbum en lloc de per cançó, obtenint així l'any de publicació per totes les pistes associades a l'àlbum. Aquesta estratègia va reduir les consultes a unes 40.000, corresponents als àlbums únics del dataset. A la Il·lustració 10: Codi Python per obtenir any de publicació d'àlbum, es mostra el fragment de codi que implementa aquesta optimització.

```
def get_album_year(album_name):
    try:
        results = sp.search(q=album_name, type='album', limit=1)
        if results['albums']['items']:
            album = results['albums']['items'][0]
            release_date = album['release_date']
            year = int(release_date.split('-')[0])
            return year
    except Exception as e:
        safe_print(f"Error en la cerca de l'àlbum {album_name}: {e}")
    return None
```

Il·lustració 10: Codi Python per obtenir any de publicació d'àlbum

Per assegurar la continuïtat i evitar repeticions en cas d'interrupcions, les dades es guarden cada 10 àlbums en un fitxer CSV, permetent reprendre el procés des de l'últim punt guardat sense duplicar consultes.

6.3 Anàlisi de les dades

El conjunt de dades integra característiques musicals objectives (energia, ballabilitat, valence) i metadades descriptives (nom, artista, àlbum, any de publicació), així com informació sobre popularitat, gènere i contingut explícit. Per explorar i visualitzar aquestes dades de forma eficient, s'ha utilitzat Power BI, posant especial atenció a l'evolució temporal de certs atributs i a les diferències segons gènere i popularitat.

⁷ API: conjunt de regles que permeten que diferents programes es comuniquin entre si.

⁸ Kaggle: plataforma en línia per a competicions i intercanvi de dades i models d'intel·ligència artificial.

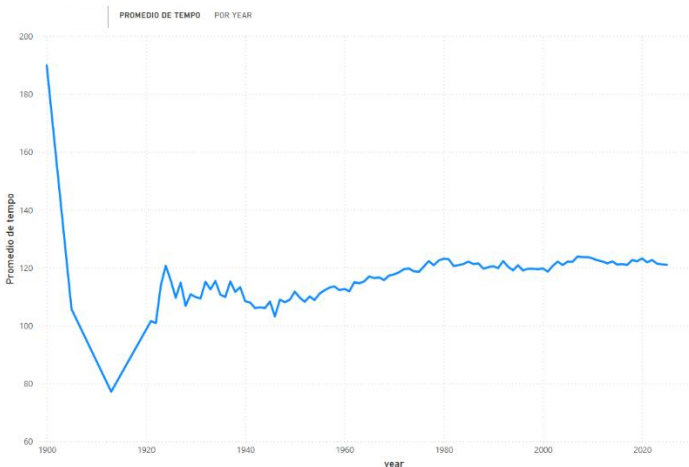
Evolució de característiques per any:

Per analitzar l'evolució temporal de la música, s'han seleccionat quatre variables: danceability (ballabilitat), valence (emoció positiva), energy (intensitat sonora) i tempo (velocitat en BPM).

- **Danceability:** Indica com de ballable és una cançó (0 a 1).
- **Valence:** Mesura la positivitat emocional de la música (0 = trista, 1 = alegre).
- **Energy:** Reflecteix la intensitat i activitat sonora (0 = suau, 1 = molt energètica).
- **Tempo:** Indica la velocitat rítmica d'una cançó, mesurada en beats per minut (BPM).

Fins als anys 60, l'energia era baixa (<0,4) i ha anat pujant fins a 0,6-0,7, mentre la valence es manté estable amb una lleugera baixada. La danceability va ser alta entre 1920 i 1940, va baixar després, va pujar als 80 i ha anat disminuint fins a 0,45 actualment. A la Il·lustració 26 (Apèndix 4), es pot observar la tendència d'aquestes variables al llarg dels anys.

El tempo s'ha mantingut estable entre 100 i 120 BPM, a la Il·lustració 11, s'observa aquesta estabilitat, tot i que a les darreres dècades es nota una tendència a superar els 120 BPM.



Il·lustració 11: Evolució del tempo per dècades

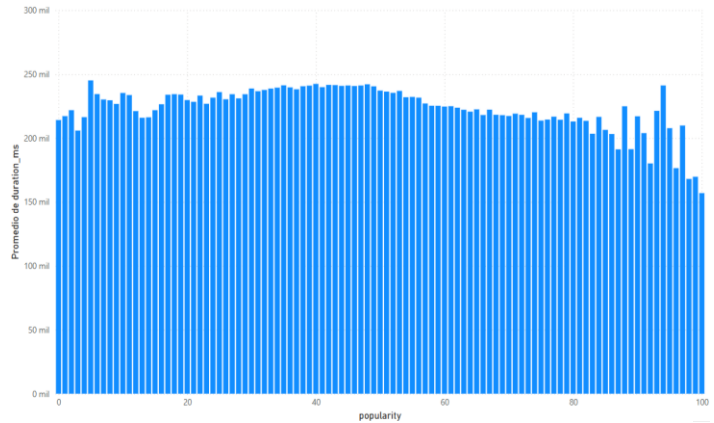
Popularitat per Gènere:

Aquest apartat analitza la freqüència de gèneres musicals al dataset, comptant quantes cançons hi ha per cada gènere, sense utilitzar la popularitat individual de les cançons. A la Il·lustració 27 (Apèndix 4), es mostren com els gèneres més comuns són la música clàssica, òpera, comèdia i study. Cal tenir en compte que, segons l'API de Spotify, els gèneres s'associen als artistes, no a les cançons.

Distribució de la popularitat en funció de la durada:

S'ha analitzat la relació entre la durada de les cançons (en mil·lisegons) i la seva popularitat. A la Il·lustració 12 es mostra que les cançons amb durada al voltant de 240.000 ms tenen una popularitat mitjana d'uns 40. A mesura que la popularitat augmenta, la durada tendeix a reduir-se fins

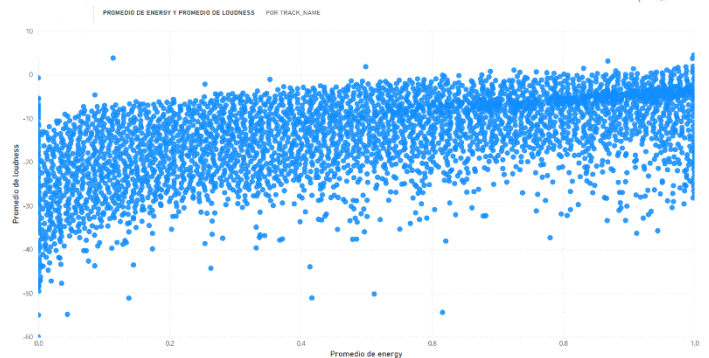
a uns 220.000 ms. Tot i això, quan la popularitat supera 90, la durada és molt variable, indicant que no existeix una durada òptima que garanteixi major popularitat.



Il·lustració 12: Distribució popularitat segons durada

Relació entre Energy i Loudness:

A la Il·lustració 13 observem com les cançons amb un nivell d'energia més elevat, proper a 1.0 (Eix X), tendeixen a presentar també un augment en el valor de loudness (Eix Y), traduït com a major volum sonor, que es mesura en decibels (dB). Cal destacar que el loudness habitualment es manté per sota de 0 dB per evitar distorsions audibles.

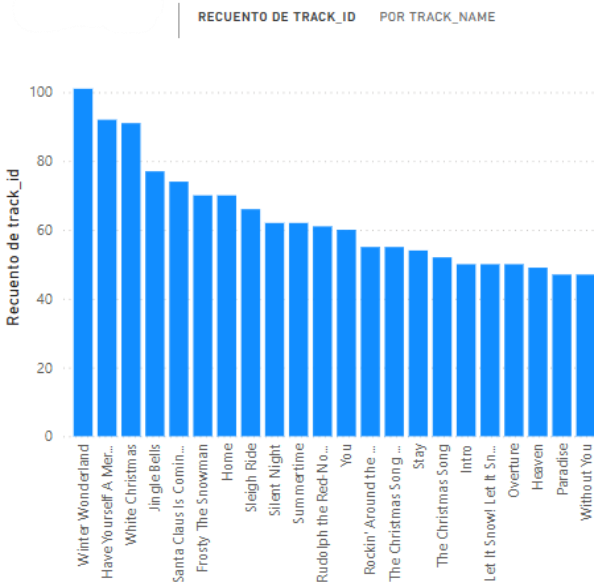


Il·lustració 13: Relació entre Energy i Loudness

Noms de cançons que més es repeteixen:

Alguns títols de cançons es repeteixen moltes vegades dins del nostre conjunt de dades, ja que diversos artistes poden publicar temes amb el mateix nom. Per exemple, cançons com *Winter Wonderland*, *White Christmas* o *Home* apareixen fins a un centenar de vegades. En aquest anàlisi s'han comptabilitzat totes les cançons que contenen aquest text al títol, encara que el nom complet sigui més extens.

Aquest fenomen permet identificar patrons en la tria dels noms, sovint influïts per temàtiques comunes. A la Il·lustració 14, es pot veure com títols com *Jingle Bells* o *Santa Claus Is Coming To Town* són molt freqüents i clarament associats a la temàtica nadalenca. D'altra banda, també emergeixen paraules com *You*, *Home* o *Stay*, que es repeteixen sovint i podrien considerar-se paraules clau, probablement per la seva simplicitat i connexió amb temàtiques amoroses o emotives.

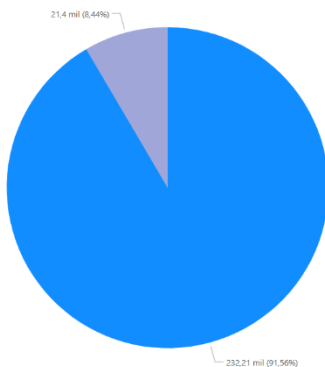


Il·lustració 14: Repetició de noms per a cançons

Contingut explícit:

Segons la classificació utilitzada per plataformes com Spotify, el contingut explícit fa referència a aquelles cançons que contenen paraules malsonants, llenguatge ofensiu o fan referència a temes considerats sensibles o inapropiats per a certs públics.

Aprofitant que els datasets contenen aquesta informació, volem analitzar quin és el número de cançons catalogades com a contingut explícit. La Il·lustració 15 mostra la proporció de cançons amb contingut explícit dins del nostre conjunt de dades. Es pot observar que aproximadament 21.400 cançons inclouen contingut sensible, mentre que 232.210 no en contenen, podem concloure que una petita part de les cançons realment contenen contingut explícit.



Il·lustració 15: Contingut Explícit

6.4 Sistema Recomanador Híbrid

En aquesta secció es presenta el desenvolupament d'un sistema de recomanació musical capaç de proposar cançons similars a partir de les seves característiques sonores i del seu ús agregat en playlists. El sistema combina tècniques de filtratge basat en el contingut (Content-Based Filtering) i filtratge col·laboratiu (Collaborative Filtering) en una arquitectura híbrida.

L'objectiu és oferir recomanacions més robustes i precises, tenint en compte tant l'essència musical de cada cançó com els patrons de coaparició observats en col·leccions públiques de playlists, sense disposar d'informació directa sobre usuaris individuals.

1. Obtenció de les dades

Per construir aquest sistema recomanador híbrid, ha estat necessari combinar dues fonts de dades diferenciades però complementàries:

- Dades musicals (content-based):

Provenen d'un conjunt de dades estructurat amb informació tàcnica i sonora de cada cançó. Aquestes dades inclouen variables com valence, energy, danceability, acousticness, loudness, tempo, entre d'altres, que permeten descriure l'essència musical de cada peça. Aquest tipus d'informació és fonamental per a sistemes de recomanació basats en el contingut, ja que es poden comparar les cançons segons les seves propietats sonores. Són les dades que hem utilitzat anteriorment per a l'anàlisi mitjançant Power BI.

- Dades col·laboratives (collaborative filtering):

S'han extret del Spotify Million Playlist Dataset [14], un recurs públic que conté un milió de playlists distribuïdes en 1000 fitxers JSON. Tot i que aquest conjunt no inclou informació directa d'usuaris, permet analitzar patrons d'ús agregat observant quines cançons apareixen conjuntament dins les playlists. A la Il·lustració 16 veiem un petit exemple de l'estructuració de la informació dins d'aquests fitxers JSON.

L'anàlisi s'ha centrat en identificar les cançons del nostre dataset inicial que també apareixen en aquest dataset comparant tots dos a través de l'atribut track ID, per tal d'obtenir informació rellevant sobre la seva presència, posició i exposició dins les playlists.

```

"playlists": [
  {
    "name": "Throwbacks",
    "collaborative": "false",
    "pid": 0,
    "modified_at": 1493424000,
    "num_tracks": 52,
    "num_albums": 47,
    "num_followers": 1,
    "tracks": [
      {
        "pos": 0,
        "artist_name": "Missy Elliott",
        "track_uri": "spotify:track:0UaMYEVWZi0ZqiD0oHU3YI",
        "artist_uri": "spotify:artist:2wIVse2owCl7go1WT98tk",
        "track_name": "Lose Control (feat. Ciara & Fat Man Scoop)",
        "album_uri": "spotify:album:6vV5UrXcfyQD1wu4Qo219K",
        "duration_ms": 226863,
        "album_name": "The Cookbook"
      },
      {
        "pos": 1,
        "artist_name": "Britney Spears",
        "track_uri": "spotify:track:619VzXrHx09rA9A5euc8Ak",
        "artist_uri": "spotify:artist:26dSoYclwsYLMaK03tp0r4",
        "track_name": "Toxic",
        "album_uri": "spotify:album:0z7pVBG0D7HCIB7S8eLkLI",
        "duration_ms": 198800,
        "album_name": "In The Zone"
      }
    ]
  }
]

```

Il·lustració 16: Estructura d'un fitxer JSON del Spotify Million Playlist Dataset

2. Característiques col·laboratives extretes

A partir de l'anàlisi del Spotify Million Playlist Dataset, s'han derivat un conjunt de variables que reflecteixen el comportament agregat dels usuaris en l'ús de cançons dins playlists. Aquestes característiques no descriuen la cançó a nivell sonor, sinó que informen sobre com, on i amb quina freqüència apareix en col·leccions públiques, capturant així la seva popularitat contextual i recurrència social. Les variables calculades són les següents:

- **avg_pos**: Posició mitjana de la cançó dins les playlists on apareix.
- **min_pos**: Posició més alta (més a prop del principi) registrada.
- **max_pos**: Posició més baixa (més a prop del final) registrada.
- **total_followers**: Suma de seguidors de totes les playlists que contenen la cançó.
- **avg_followers**: Mitjana de seguidors de les playlists on apareix.
- **num_playlists**: Nombre total de playlists on es troba la cançó.
- **avg_num_tracks**: Nombre mitjà de cançons per playlist.

3. Fusió del dataset Híbrid

Un cop extretes les característiques musicals (content-based) i col·laboratives (playlist-based), es duu a terme la fusió de tots dos conjunts en un únic dataset híbrid. Aquesta integració es fa utilitzant el camp `track_id` com a clau comuna, ja que és l'identificador únic que permet enllaçar cada cançó amb les seves metadades sonores i les seves estadístiques de comportament en playlists.

L'objectiu d'aquesta fusió és obtenir una representació completa de cada cançó, que combini informació intrínseca (com l'energia, valence, tempo...) amb informació contextual i social (com la posició mitjana o el nombre de playlists on apareix). Aquesta doble visió permet capturar tant l'essència sonora com el seu ús dins del consum musical agregat.

4. Preprocessament de les dades

Es carrega el conjunt complet de cançons amb totes les característiques musicals i col·laboratives. Es descarten atributs no numèrics, com el nom de la cançó o l'artista, i es normalitzen totes les variables numèriques amb la tècnica **StandardScaler**⁹ de la llibreria **scikit-learn**¹⁰ per garantir que totes tinguin el mateix per en el càlcul de similitud. Per evitar penalitzar les cançons que no apareixen en cap playlist —i que per tant no tenen dades col·laboratives— s'imputen valors neutres (com la mitjana o zero) en aquestes variables. Així, aquestes peces continuen tenint oportunitat de ser recomanades a partir de les seves característiques sonores.

5. Càlcul de la similitud

Per trobar les cançons més semblants a una donada, s'utilitza la similitud cosinus, que compara l'angle entre

vectors i s'adapta bé a dades normalitzades. Es carrega el dataset ja preprocessat i s'identifica la cançó de referència. A partir del seu vector de característiques musicals (content-based) i col·laboratives (playlist-based), es calcula la similitud cosinus amb la resta del conjunt.

Aquesta operació es realitza per separat sobre cada tipus de característiques i després es combinen en una puntuació híbrida mitjançant una ponderació controlada pel paràmetre α . Aquest paràmetre defineix el pes relatiu de la similitud basada en contingut (content-based) i la basada en col·laboració (collaborative filtering), permetent ajustar la influència que cada tipus de característica té en el resultat final.

Un valor d' $\alpha = 0.75$ fa que el pes de les característiques musicals sigui major que el de les dades col·laboratives, ja que les playlists reflecteixen preferències més personals i específiques dels usuaris que les creen. A la Il·lustració 17 es mostra una representació gràfica dels pesos assignats.

$$\text{hybrid_scores} = \alpha * \text{sim_cb} + (1 - \alpha) * \text{sim_cf}$$

Il·lustració 17: Fórmula del híbrid_score

Aquest procés és possible gràcies a la similitud cosinus, una tècnica habitual en sistemes de recomanació basada en vectors. La mesura compara l'angle entre dos vectors en un espai multidimensional. A la Il·lustració 18 observem la fórmula corresponent, identificant A i B com els vectors que representen les característiques musicals de les dues cançons. El resultat estarà entre el -1 i 1, indicant un valor proper a 1 una gran similitud i un proper a 0 una baixa similitud.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Il·lustració 18: Fórmula Cosinus

6. Resultats:

Un cop executat, el programa mostra un missatge de benvinguda i segueix una seqüència d'interacció amb l'usuari com s'observa a l'exemple de la Il·lustració 19:

- **Entrada del nom de la cançó**: L'usuari introdueix el nom d'una cançó.
- **Cerca de coincidències**: El sistema retorna cançons que continguin el text introduït.
- **Selecció**: L'usuari escull el número corresponent a la cançó desitjada.
- **Recomanacions**: Es calculen les similituds i es mostren les 10 cançons més similars, incloent el seu nom, artista i la puntuació (hybrid_score).
- **Iteració**: L'usuari pot repetir el procés tantes vegades com vulgui o escriure sortir per acabar.

⁹ **StandardScaler**: mètode de normalització que transforma les dades perquè tinguin mitjana 0 i desviació estàndard 1.

¹⁰ **scikit-learn**: llibreria de Python per a aprenentatge automàtic i preprocessament de dades.

Benvingut al recomanador de cançons!

Introdueix el nom (o part) de la cançó (o 'sortir' per acabar): Cornfield Chase

Possibles cançons trobades:

22472: 'Cornfield Chase - Piano-Cello Version' - Michael Forster

224140: 'Cornfield Chase' - Hans Zimmer

Introdueix el número de la cançó que vols recomanar: 224140

Recomanacions per 'Cornfield Chase' - Hans Zimmer (alpha=0.75):

Pos	Track Name	Artist	Score
1	Watchman's Ease	Jeremy Soule	0.9720
2	From Shores To Mountains	The 2 Inversions	0.9708
3	Gamma Freq Pads	Masters of Binaurality	0.9683
4	Futari no kimochi	Kaoru Wada	0.9673
5	no, the moon	Teen Suicide	0.9587
6	Just for You	Giovanni	0.9579
7	Chalkboard	Jóhann Jóhannsson	0.9577
8	Anakin's Dark Deeds	John Williams,London S...	0.9566
9	A New Beginning	Tenno	0.9550
10	Chevaliers De Sangreal	Hans Zimmer	0.9547

Introdueix el nom (o part) de la cançó (o 'sortir' per acabar): sortir
Fins aviat!

Il·lustració 19: Exemple Recomanador

7 CONCLUSIONS

Aquest projecte ha evidenciat com l'anàlisi de dades pot tenir aplicacions molt diverses, des del tractament d'informació personal, com els tiquets de compra d'un únic usuari, fins a grans volums de dades musicals provinents de fonts obertes com l'API de Spotify. Aquesta evolució ha permès aplicar tècniques de Business Intelligence adaptades a cada situació, mostrant com la qualitat, la neteja i l'estructuració de les dades són fonamentals per extreure'n valor. Les visualitzacions generades han contribuït a detectar patrons, tendències i comportaments, consolidant la BI com una eina potent d'exploració i coneixement.

Tanmateix, aquesta experiència també ha posat en relleu la importància de tenir una mirada crítica i responsable davant l'ús creixent de dades en tots els àmbits de la vida quotidiana. A mesura que les tecnologies avancen, les dades deixen de ser simples registres per convertir-se en el fonament de moltes decisions personals, comercials i institucionals. És per això que cal actuar amb responsabilitat i sensibilitat, evitant usos que puguin derivar en discriminacions o pèrdues de privacitat, tot garantint una anàlisi justa i respectuosa amb les persones i els contextos als quals pertanyen les dades.

Aquest treball, doncs, ha anat més enllà de l'àmbit purament tècnic, i ha estat també una aproximació a la complexitat social de l'entorn digital actual. L'anàlisi de dades ha de ser útil i eficient, però també transparent, comprensible i respectuosa. Aquest equilibri entre capacitat analítica i responsabilitat ha de guiar qualsevol procés de Business Intelligence, especialment en un món cada cop més interconnectat i dependent de la informació.

Bibliografia

[1] Gartner, "Business Intelligence," 2023. [En línia]. Disponible: <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>. [Consulta: 15-02-2025].

[2] Forbes, "How Business Intelligence is Transforming Retail," 2021. [En línia]. Disponible: <https://www.forbes.com>. [Consulta: 15-02-2025].

[3] Microsoft, "What is Power BI?," 2023. [En línia]. Disponible: <https://powerbi.microsoft.com>. [Consulta: 15-02-2025].

[4] pdfplumber, "pdfplumber: A Step-by-Step Guide to Parsing PDFs using the pdfplumber Library In Python," 2023. [En línia]. Disponible: <https://azhar-sayyad.medium.com/a-step-by-step-guide-to-parsing-pdfs-using-the-pdfplumber-library-in-python-c12d94ae9f07>. [Consulta: 07-03-2025].

[5] Mercagpt, "Mercagpt: Plataforma de Datos," 2025. [En línia]. Disponible: <https://www.mercagpt.com/>. [Consulta: 15-03-2025].

[6] TechTarget, "Data Curation," 2025. [En línia]. Disponible: <https://www.techtarget.com/searchbusinessanalytics/definition/data-curation>. [Consulta: 20-03-2025].

[7] Microsoft, "Tipos de visualizaciones de Power BI para informes y preguntas y respuestas," 2025. [En línia]. Disponible: <https://learn.microsoft.com/es-es/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a>. [Consulta: 16-03-2025].

[8] YouTube, "How Power BI Works, Boxplot" 2025. [En línia]. Disponible: <https://www.youtube.com/watch?v=iPbVdod9e-U>. [Consulta: 26-03-2025].

[9] Business Insider España, "Así funciona el algoritmo de Spotify: en detalle, qué sabe y cómo sabe lo que quieres escuchar," 2023. [En línia]. Disponible: <https://www.businessinsider.es/tecnologia/algoritmo-spotify-detalle-sabe-spotify-quieres-escuchar-1231984>. [Consulta: 03-05-2025].

[10] Nodd3r, "Cómo funciona el algoritmo y el sistema de recomendación de Spotify," s. f. [En línia]. Disponible: <https://nodd3r.com/blog/como-funciona-el-algoritmo-y-el-sistema-de-recomendacion-de-spotify>. [Consulta: 05-05-2025].

[11] Kaggle, "Music Recommendation System using Spotify Dataset," s. f. [En línia]. Disponible: <https://www.kaggle.com/code/vatsalmanvani/music-recommendation-system-using-spotify-dataset/input>. [Consulta: 09-05-2025].

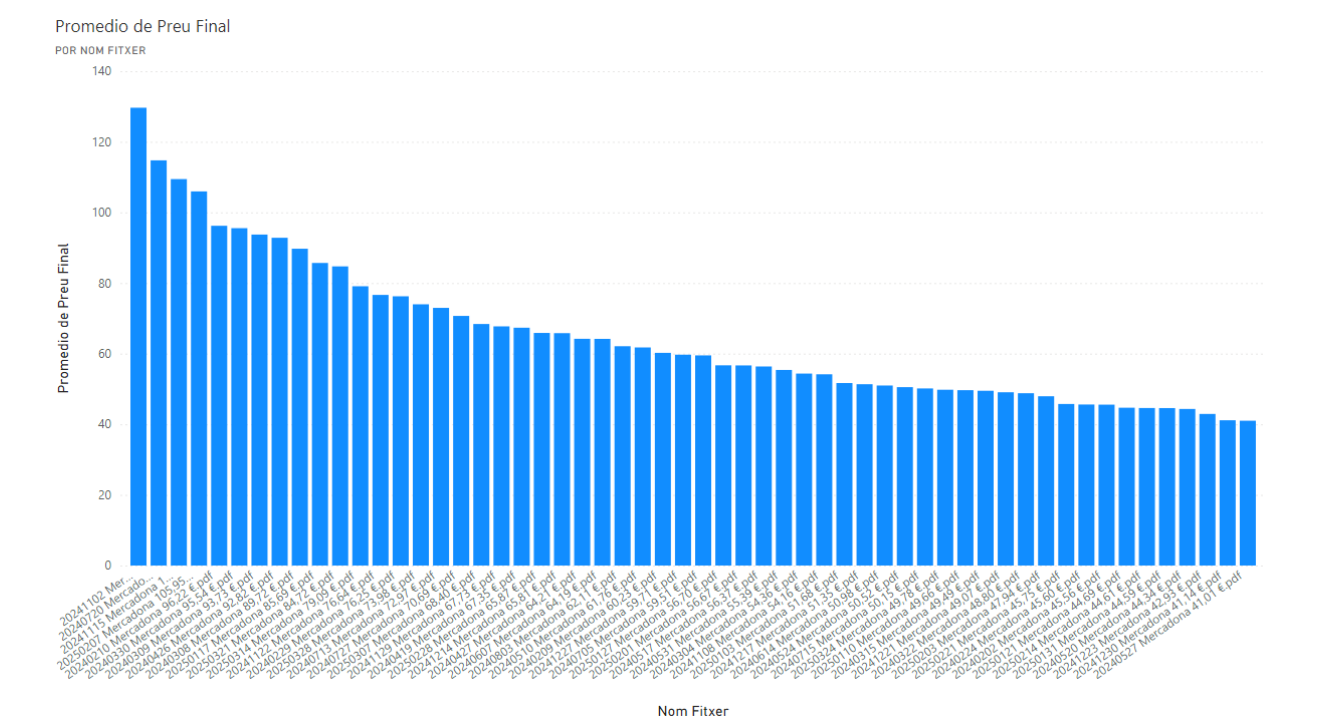
[12] Kaggle, "Spotify Tracks Dataset," s. f. [En línia]. Disponible: <https://www.kaggle.com/datasets/maharshipandya/spotify-tracks-dataset>. [Consulta: 11-05-2025].

[13] UNIR, "Distributed processing using cosine similarity for mapping Big Data in Hadoop," s. f. [En línia]. Disponible: <http://research.unir.net/unesco-congreso/wp-content/uploads/sites/76/2016/06/6-Distributed-processing-using-cosine-similarity-for-mapping-Big-Data-in-Hadoop1.pdf>. [Consulta: 16-05-2025]. [Imatge utilitzada a la Il·lustració 18: Fórmula Cosinus].

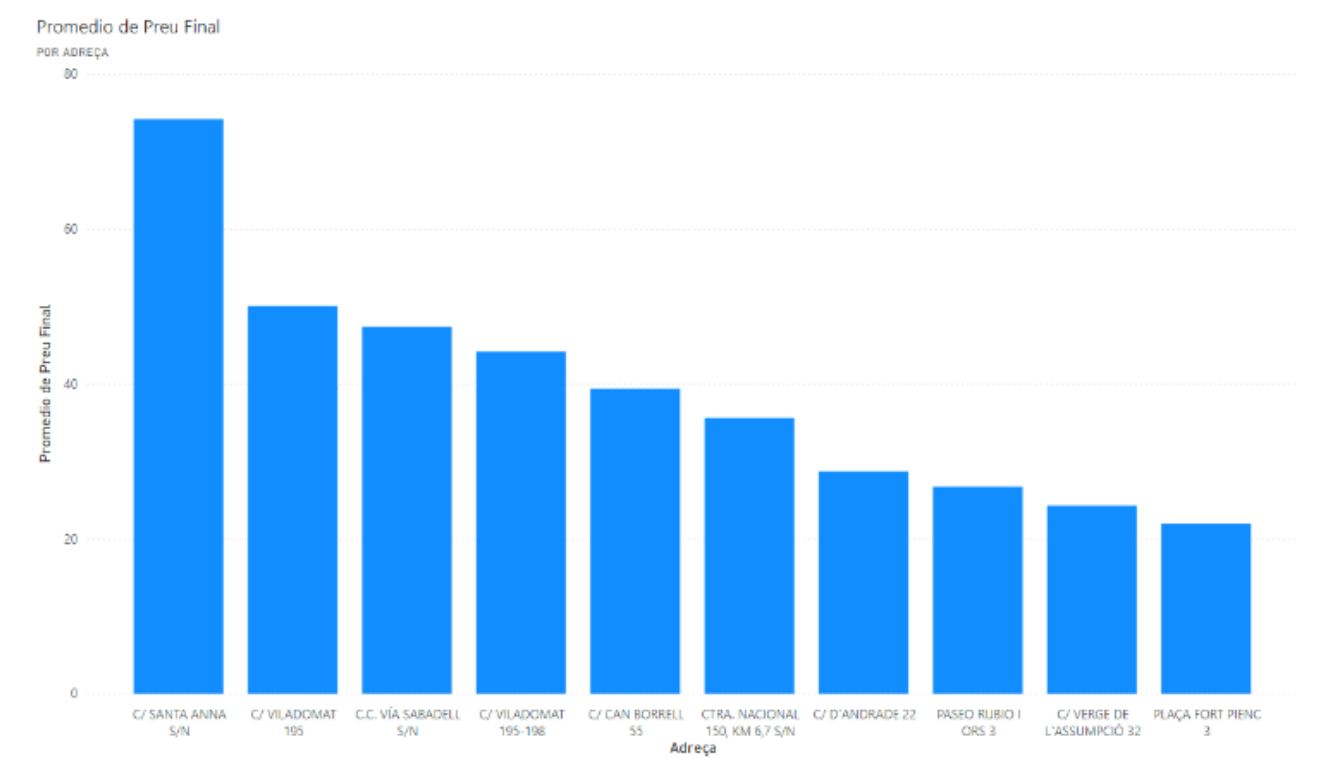
[14] Aicrowd, "Spotify Million Playlist Dataset Challenge," s. f. [En línia]. Disponible: <https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge>. [Consulta: 20-05-2025].

APÈNDIX

A1. PREU PER COMPRA

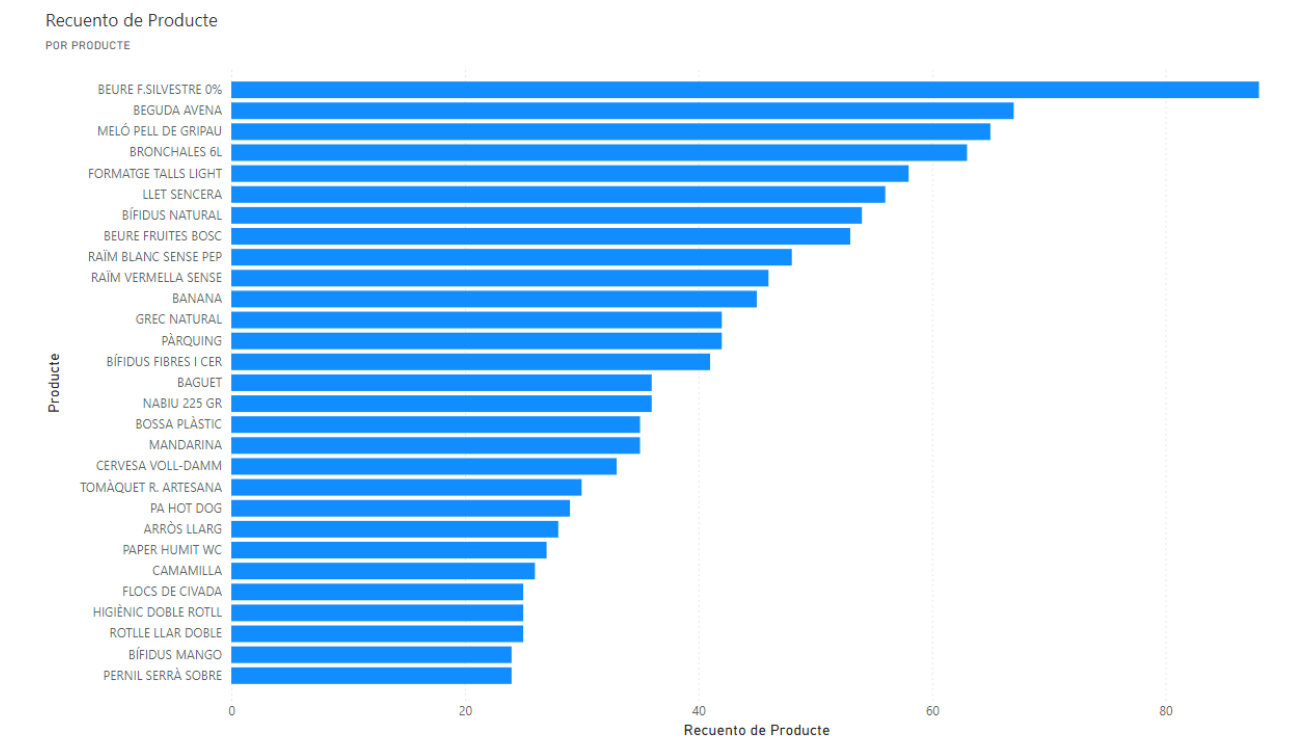


Il·lustració 20: Preu per compra

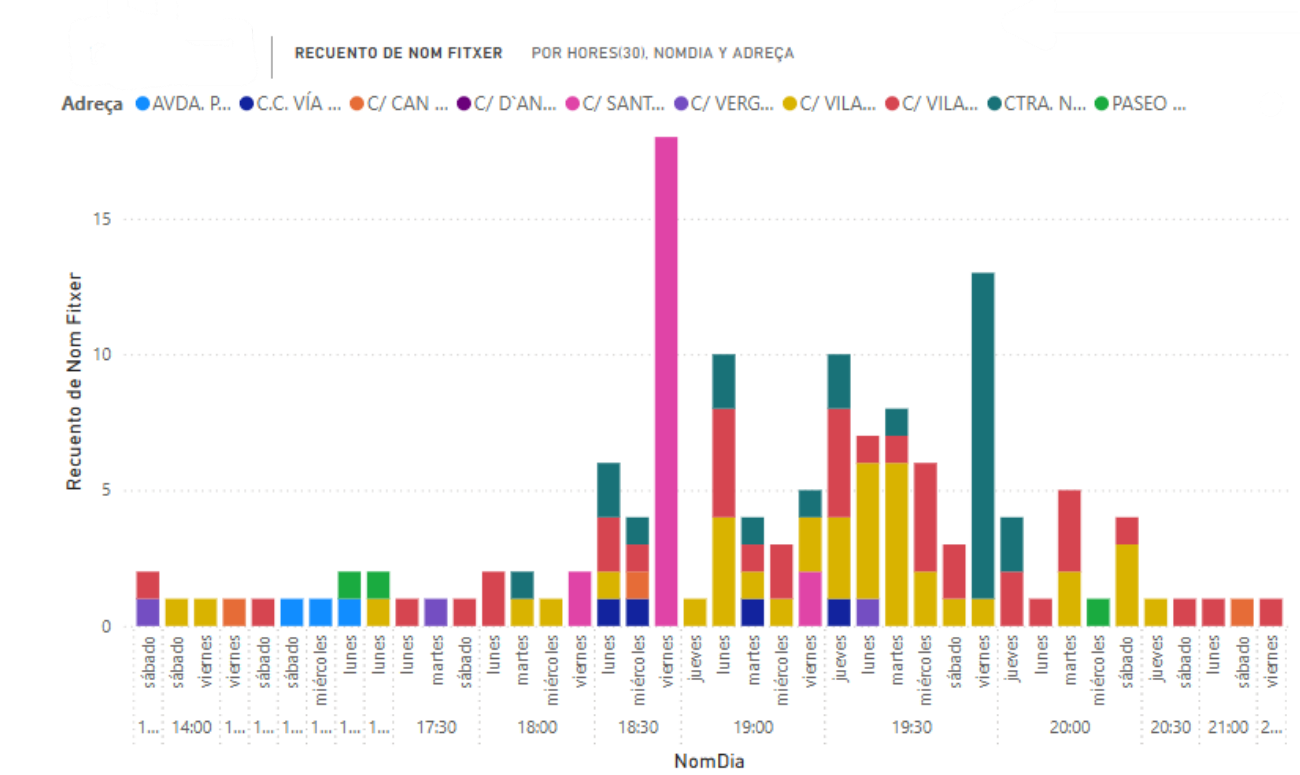


Il·lustració 21: Despesa promig segons establiment

A2. REPETICIONS DE PRODUCTE I PATRONS DE COMPRA

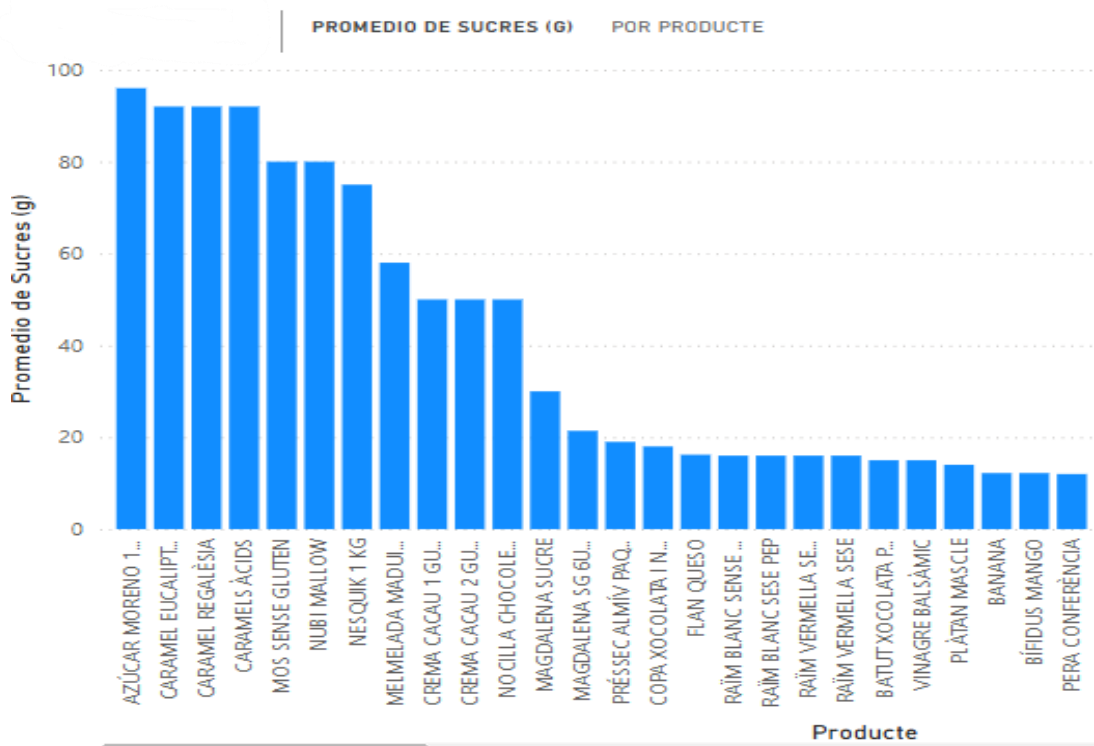


Il·lustració 22: Recompte de productes adquirits

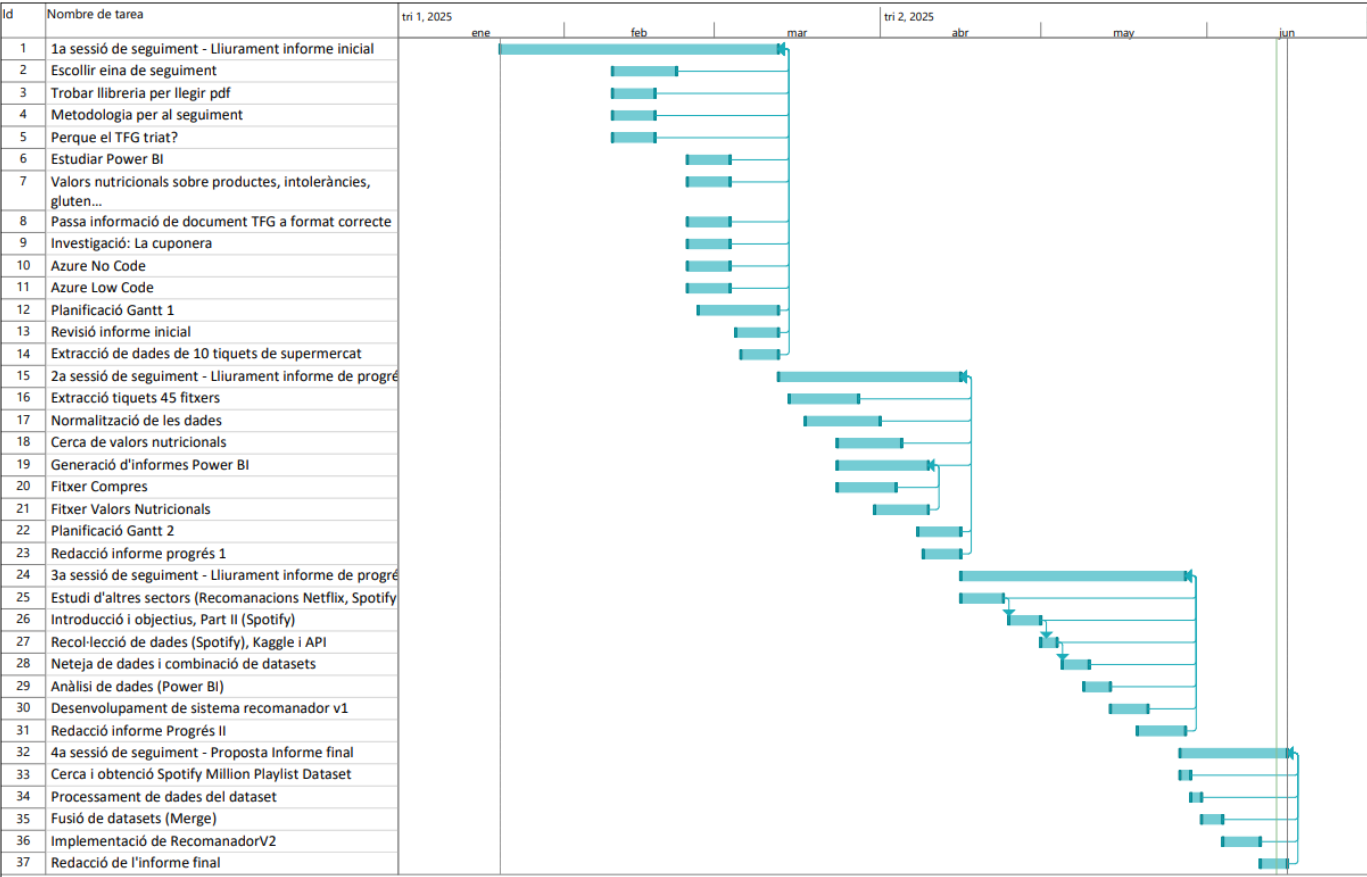


Il·lustració 23: Patrons de compra

A3. VALORS NUTRICIONALS I DIAGRAMA DE GANTT

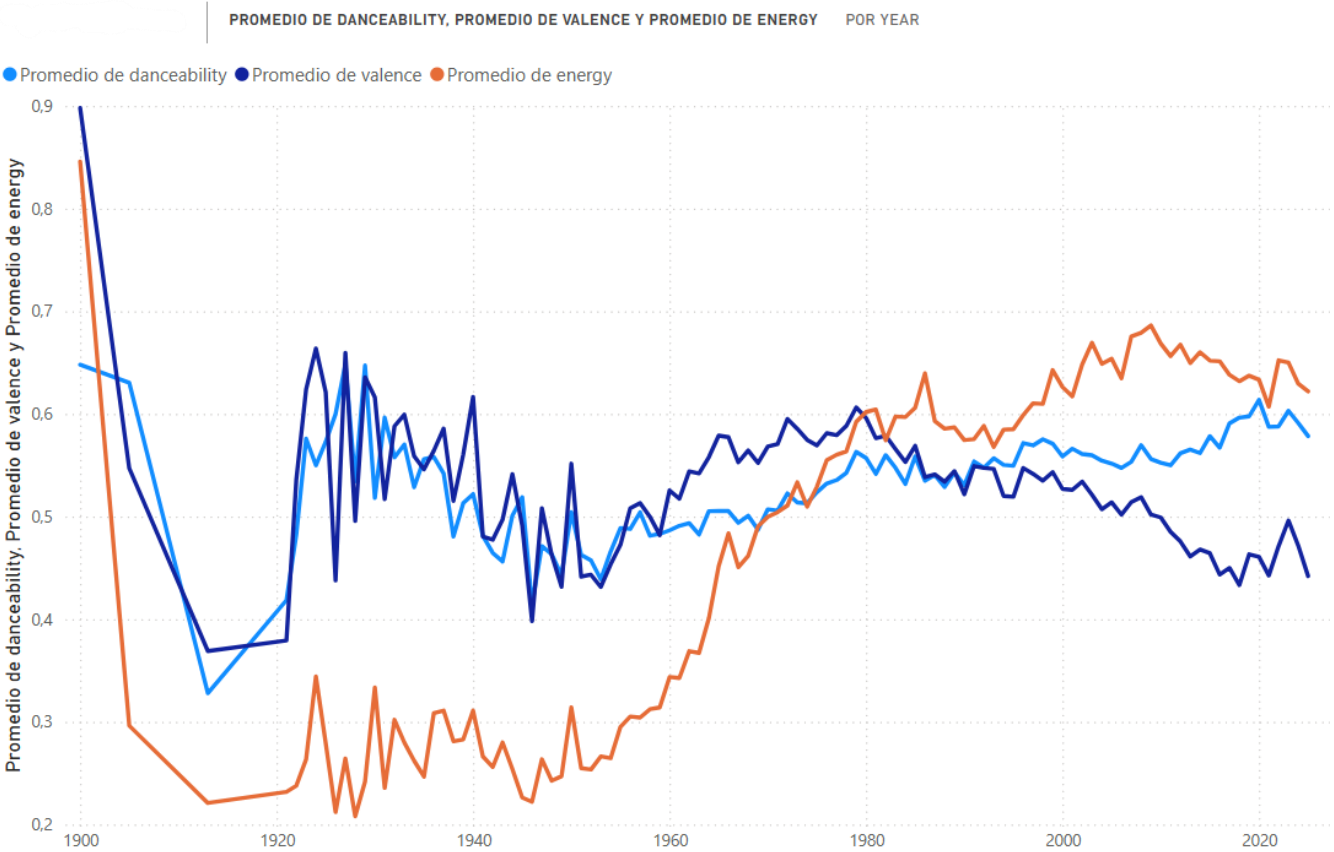


Il·lustració 24: Gràfic sucres

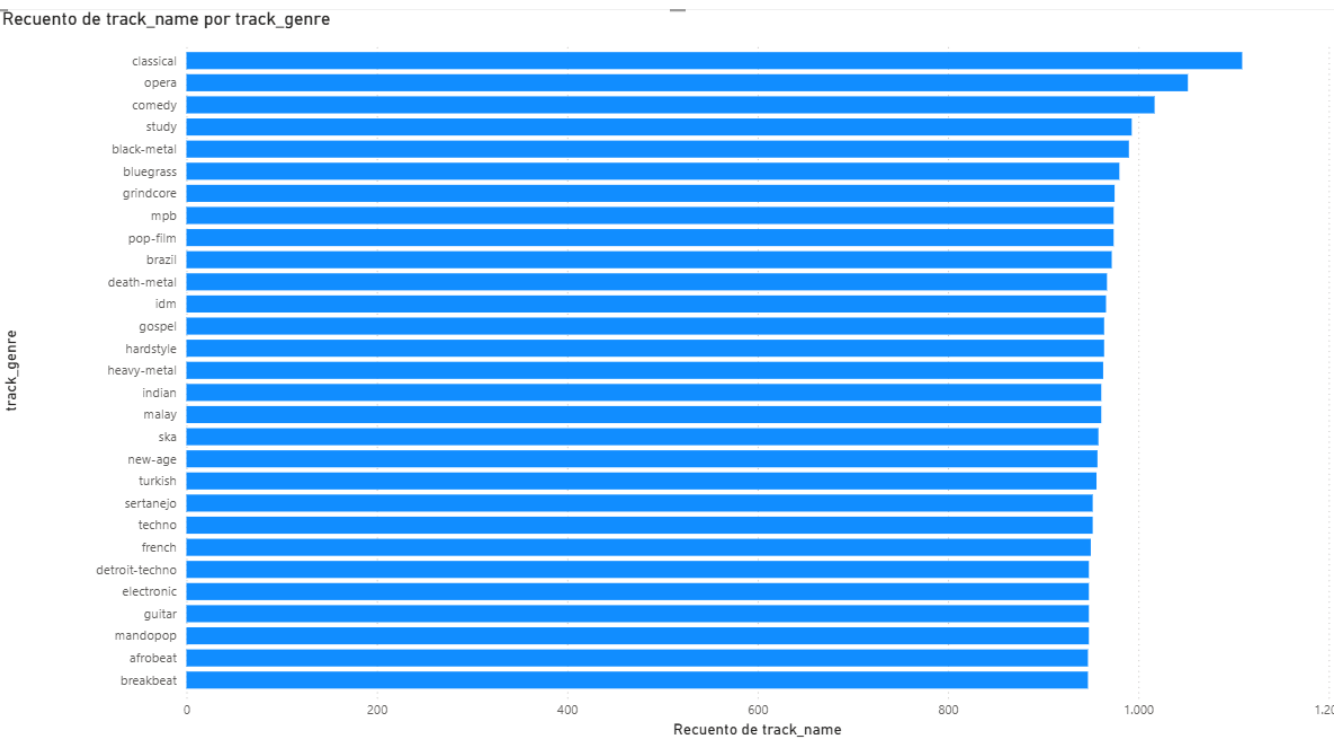


Il·lustració 25: Diagrama de Gantt

A4. ANÀLISI DE SPOTIFY



Il·lustració 26: Evolució de característiques per any



Il·lustració 27: Distribució de generes musicals