



Tipificación de clientes por tienda de OFFCORSS mediante Análisis de Datos

Johan Daniel Zuluaga Gómez

Informe de prácticas para optar por título de Ingeniero de sistemas

Asesor interno

Daniela Serna Buitrago

Asesor externo

Nathalia Velez Velasquez

Universidad de Antioquia

Facultad de ingeniería, UdeA

Ingeniería de sistemas

Medellín, Antioquia, Colombia

2024

Cita

Zuluaga Gómez [1]

Referencia

[1] J. D. Zuluaga Gómez, “Tipificación de clientes por tienda de OFFCORSS mediante Análisis de Datos”, Semestre de industria, Ingeniería de Sistemas, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2024.
Estilo IEEE (2020)



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio César Saldarriaga Molina.

Jefe departamento: Diego José Luis Botía Valderrama.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

TABLA DE CONTENIDO

RESUMEN	6
ABSTRACT	7
I. INTRODUCCIÓN	8
II. OBJETIVOS	9
III. MARCO TEÓRICO	10
IV. METODOLOGÍA	12
V. RESULTADOS	14
VI. ANÁLISIS	18
VII. CONCLUSIONES	19

LISTA DE FIGURAS

Fig. 1. Diagrama del proceso de minería de datos CRISP-DM	13
Fig. 2. Sección con indicadores de las tiendas de OFFCORSS	15
Fig. 3. Diapositiva con los resultados del análisis de indicadores para la tienda	16
Fig. 4. Matriz de entrada para el modelo NMF	17

SIGLAS, ACRÓNIMOS Y ABREVIATURAS

PCA	Principal Component Analysis
NMF	Non-negative Matrix Factorization
CRISP-DM	Cross-Industry Standard Process for Data Mining
UdeA	Universidad de Antioquia

RESUMEN

Este artículo presenta un análisis detallado de datos para generar perfiles de clientes en cada tienda de OFFCORSS, utilizando dos enfoques principales: análisis de indicadores y modelos de factorización de matrices no negativas (NMF). El estudio se basa en la metodología CRISP-DM y revela variaciones significativas en el desempeño de las tiendas, así como patrones de consumo distintivos entre diferentes segmentos de clientes y productos. Los resultados proporcionan información valiosa para mejorar la estrategia comercial de OFFCORSS, adaptándola a las necesidades y preferencias específicas de cada tienda.

Palabras clave — **Análisis de datos, Perfiles de clientes, Indicadores de rendimiento, Modelos de factorización de matrices no negativas (NMF), Comportamiento del consumidor, Optimización de inventario.**

ABSTRACT

This article presents a detailed data analysis to generate customer profiles in each OFFCORSS store, using two main approaches: indicator analysis and non-negative matrix factorization (NMF) models. The study is based on the CRISP-DM methodology and reveals significant variations in the stores' performance, as well as distinctive consumption patterns among different customer and product segments. The results provide valuable information to improve OFFCORSS's commercial strategy, adapting it to the specific needs and preferences of each store.

***Keywords* — Data analysis, Customer profiles, Performance indicators, Non-negative matrix factorization (NMF) models, Consumer behavior, Inventory optimization.**

I. INTRODUCCIÓN

En el competitivo mercado de la moda infantil, comprender las preferencias y comportamientos de los clientes es fundamental para mantenerse relevante y competitivo. En este contexto, el análisis de datos se ha convertido en una herramienta invaluable para las empresas que buscan entender mejor a su audiencia y adaptar sus estrategias comerciales en consecuencia. En este artículo, se presenta un análisis exhaustivo de datos aplicado a las tiendas de OFFCORSS, una reconocida marca de ropa infantil. El objetivo principal es generar perfiles detallados de clientes para cada tienda, utilizando dos enfoques principales: análisis de indicadores y modelos de factorización de matrices no negativas (NMF). Este estudio se basa en la metodología CRISP-DM y proporciona insights valiosos sobre el desempeño de las tiendas, así como sobre los patrones de consumo de los clientes. Los resultados obtenidos tienen el potencial de informar y mejorar significativamente la estrategia comercial de OFFCORSS, permitiendo una mayor personalización y adaptación a las necesidades específicas de cada mercado local.

II. OBJETIVOS

A. Objetivo general

Realizar un análisis exhaustivo de datos para generar perfiles detallados de clientes para cada tienda de OFFCORSS. Para esto se usarán dos enfoques principales.

B. Objetivos específicos

1. Analizar indicadores para todas las tiendas y
2. Aplicar modelos de factorización de matrices no negativas (NMF) para identificar patrones de consumo.

III. MARCO TEÓRICO

La generación de perfiles de clientes ha sido un tema fundamental en marketing. Se propone utilizar dos enfoques principales:

1. Análisis de Indicadores: Este método implica la generación y análisis de indicadores de consumo para cada tienda, segmento de clientes, y de productos, lo que permite comparar el desempeño de cada tienda con la empresa en general e identificar áreas de oportunidad. La generación de indicadores requiere de herramientas estadísticas como promedios, porcentajes y crecimientos.

2. Factorización de Matrices No Negativas (NMF): Es un método de álgebra lineal que descompone una matriz en múltiples matrices más simples, todas ellas con valores no negativos. Este enfoque resulta especialmente útil en el análisis de datos cuando se trabaja con matrices que representan cantidades no negativas, como es el caso de datos de consumo, calificaciones, o frecuencias. [1]

En el contexto de análisis de segmentación de clientes y productos, la NMF permite descomponer una matriz que contiene en sus filas los segmentos de clientes y en sus columnas los segmentos de productos, donde cada elemento de la matriz representa la cantidad de consumo de un cliente a un producto específico. Esta matriz se descompone en tres matrices: una que representa la relación entre los segmentos de clientes y los patrones de consumo, otra que representa la relación entre los patrones de consumo y los segmentos de productos, y una tercera matriz que contiene el error de la aproximación.

La interpretación de los resultados de la NMF es intuitiva y de gran utilidad en la práctica. La primera matriz resultante nos proporciona información sobre los patrones de consumo más relevantes para cada segmento de clientes, mientras que la segunda matriz nos indica cuáles son los productos más relevantes para cada patrón de consumo. Esta flexibilidad en la interpretación de los resultados permite identificar diferentes comportamientos de consumo para cada segmento de clientes, lo que facilita la toma de decisiones estratégicas personalizadas.

Además, se ha observado que la NMF tiende a ofrecer mejores resultados en comparación con otros métodos de análisis de datos, como el análisis de componentes principales (PCA) o el

clustering. Esto se debe a su capacidad para capturar patrones de consumo complejos y su capacidad para manejar datos no negativos de manera natural.

En resumen, la factorización de matrices no negativas es una técnica poderosa y versátil en el análisis de segmentación de clientes y productos, que proporciona una comprensión profunda de los patrones de consumo y facilita la toma de decisiones estratégicas en el ámbito empresarial.

IV. METODOLOGÍA

La metodología utilizada en este proyecto sigue el enfoque CRISP-DM (Cross-Industry Standard Process for Data Mining) [2], que es un marco ampliamente reconocido para la minería de datos. A continuación se detalla cada una de las etapas de esta metodología adaptada al contexto del proyecto:

Entendimiento del negocio: Esta etapa comienza con la comprensión profunda de los objetivos y necesidades del negocio. Se lleva a cabo mediante exposiciones y entrevistas con la asesora externa y la líder de inteligencia de negocios. El objetivo es identificar los desafíos y oportunidades clave que guiarán todo el proceso de minería de datos.

Entendimiento de los datos: En esta fase, en colaboración con la líder de inteligencia de negocios, se realiza una exploración exhaustiva de los sistemas de información y bases de datos disponibles. Esta exploración sirve como punto de partida para la comprensión de los datos relevantes para el proyecto, identificando su calidad, estructura y posibles limitaciones.

Preparación de los datos: El siguiente paso implica la extracción y transformación de los datos para satisfacer las necesidades específicas de los modelos que se implementarán en las etapas posteriores. Esto implica limpieza de datos, integración de fuentes, selección de características relevantes y cualquier otra preparación necesaria para garantizar la calidad y la idoneidad de los datos para el análisis.

Modelado: En esta etapa, se desarrollan y evalúan diferentes modelos de minería de datos para abordar los objetivos del proyecto. Se lleva a cabo una evaluación técnica de los modelos, así como una evaluación en términos del negocio para garantizar que los modelos seleccionados sean adecuados para resolver los desafíos identificados en la etapa de entendimiento del negocio. Este proceso puede implicar ajustes iterativos en el enfoque de modelado para garantizar que se alcancen los objetivos del negocio dentro del plazo establecido.

Evaluación: Una vez que se han desarrollado los modelos, se evalúan para determinar su eficacia y su capacidad para abordar los problemas de negocio identificados. Se realiza una

revisión exhaustiva de los resultados y se comparan con los criterios de éxito definidos en la etapa de entendimiento del negocio.

Despliegue: Finalmente, los modelos seleccionados se entregan al entorno operativo del negocio para su uso. Se establecen procesos para mantener los modelos, asegurando que sigan siendo relevantes y efectivos a lo largo del tiempo.

Esta metodología CRISP-DM proporciona un marco estructurado e iterativo para el proceso de minería de datos, asegurando que se aborden de manera efectiva los desafíos del negocio y se generen insights valiosos a partir de los datos disponibles.

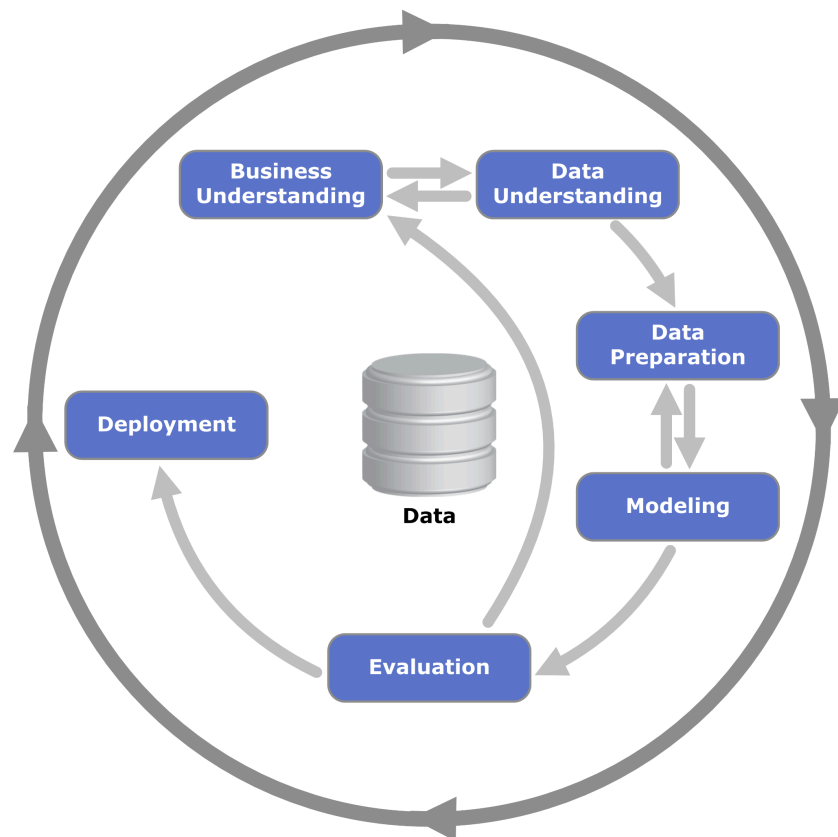


Fig. 1. Diagrama del proceso de minería de datos CRISP-DM

Nota. fuente https://upload.wikimedia.org/wikipedia/commons/b/b9/CRISP-DM_Process_Diagram.png

V. RESULTADOS

El estudio se inició conforme a la metodología establecida, comenzando con la comprensión del contexto empresarial y los datos disponibles, con el objetivo de identificar la necesidad previamente señalada y los recursos de información pertinentes para abordarla. La información disponible del cliente se centra en su historial de compras. Para cada establecimiento, se analizaron diversas métricas correspondientes al año 2023, incluyendo el número de clientes, número de clientes del año anterior, el total de facturas, los ingresos por ventas, y las unidades vendidas. A partir de estos datos, se calcularon una serie de indicadores clave, como la frecuencia de compra, el valor promedio de las facturas, la cantidad media de unidades por cliente y el crecimiento de la base de clientes.

Además, se segmentan los clientes según su ciclo de vida en la marca, lo que incluyó categorías como NUEVOS (clientes que compraron por primera vez en el último año), CONSERVADOS (clientes que compraron en el último año y en el año pasado), RECUPERADOS (clientes que compraron en el último año, que no compraron en el año pasado, pero sí antes), INACTIVOS (clientes que no compraron en el último año, pero compraron el año pasado) y PERDIDOS (clientes que llevan 2 años sin comprar en la marca). Estas segmentaciones se complementaron con la clasificación de los clientes según su nivel de inversión en la marca, dividiéndolos en categorías como NEW TALENT (han comprado menos de 300.000), STAR (han comprado entre 300.000 y menos de 1.000.00), COOLSTAR (han comprado entre 1.000.000 y 2.000.000) y ROCKSTARS (han comprado una cantidad mayor o igual a 2.000.000).

Para cada una de estas segmentaciones, se calcularon las respectivas proporciones dentro de la clientela. Posteriormente, se examinaron los datos de compra de los clientes en relación con los productos, considerando aspectos como las unidades vendidas según categorías de productos y colecciones. Esto incluyó análisis de los productos por "mundos" (Newborn, Bebé, Niño cada una con su variante femenina), tipos de artículo (accesorios, calzado, etc.) y colecciones específicas (ABC, Básicos, Playa, Special).

Para el análisis y manipulación de estos indicadores, se empleó Microsoft Excel, facilitando comparaciones entre tiendas y con la empresa en su conjunto. Los datos fueron

extraídos de la base de datos y organizados en una hoja de cálculo, donde cada fila representaba una tienda y cada columna un indicador específico como se observa en la figura 2.

	A	B	C	D	E	F	G	H
1	INDICADORES							
2	TIENDA	CLIENTES	CLIENTES AÑO PASADO	FACTURAS	VENTAS	UNIDADES	FRECUENCIA	TICKET PROMEDIO
3	Compañía	549,937	581,709	818,288	118,455,890,265	1,629,532	1.5	144,761
4	Offcorss Sogamoso	2,945	3,202	4,513	706,259,828	9,279	1.5	156,495
5	T. Gran Plaza Soacha	5,774	6,057	7,978	1,354,763,199	16,983	1.4	169,812
6	Tda Aguachica	2,400	2,455	4,167	592,619,253	8,170	1.7	142,217
7	Tda Alcaravan	6,640	5,462	10,465	1,374,720,269	19,259	1.6	131,364
8	Tda Andino	7,817	8,785	10,319	1,737,534,597	22,563	1.3	168,382
9	Tda Arauca	4,359	4,400	8,891	1,153,918,071	16,669	2.0	129,785
10	Tda Arkadia	7,854	7,964	10,841	1,606,914,879	22,558	1.4	148,226
11	Tda B/vista Monteri	5,439	5,086	9,939	1,268,086,034	19,204	1.8	127,587
12	Tda B/vista Sta M.	7,234	8,103	10,161	1,464,919,125	21,340	1.4	144,171
13	Tda Bolivia	2,022	2,622	3,509	123,042,823	6,915	1.7	35,065
14	Tda Buenavista Bqll	8,702	9,846	11,893	1,673,754,490	23,661	1.4	140,734
15	Tda C Internacional	64	419	75	10,628,576	177	1.2	141,714
16	Tda Cacique	9,213	9,671	12,600	2,100,311,866	28,941	1.4	166,691
17	Tda Calima Bogota.	5,644	5,912	7,367	1,164,988,632	15,020	1.3	158,136
18	Tda Caribe Plaza	12,304	13,193	19,773	2,469,421,578	35,456	1.6	124,889
19	Tda Centro Mayor	9,987	11,337	13,143	2,233,471,215	27,806	1.3	169,936
20	Tda Chipichape 2	10,176	12,178	14,973	2,331,891,680	33,917	1.5	155,740
21	Tda Cosmocentro Ca	6,460	6,388	9,097	1,193,284,178	17,706	1.4	131,173
22	Tda Diverplaza	4,430	4,616	6,419	901,391,995	11,798	1.4	140,426
23	Tda Dmoda	403	935	460	49,268,796	719	1.1	107,106
24	Tda Ecoplaza	6,111	6,484	8,867	1,422,782,933	17,452	1.5	160,458
25	Tda El Tesoro	12,596	13,492	20,386	2,729,089,461	38,889	1.6	133,871
26	Tda Estación Ibague	8,718	9,241	13,084	2,027,997,462	26,854	1.5	154,998
27	Tda Fontanar	7,337	7,437	10,485	1,670,624,383	20,423	1.4	159,335
28	Tda Fusagasuga	3,418	1,672	5,235	777,505,473	10,185	1.5	148,521
29	Tda Gran Estación	8,017	9,645	10,161	1,635,034,008	21,280	1.3	160,913
30	Tda Gran Plaza	6,125	5,590	10,553	1,671,479,668	22,971	1.7	158,389
31	Tda Guacari	6,554	6,351	11,234	1,345,676,562	20,123	1.7	119,786
32	Tda Guatapuri	6,807	6,653	9,680	1,516,539,700	22,543	1.4	156,667
33	Tda Hayuelos bogt.	3,859	4,543	5,144	751,109,177	9,270	1.3	146,017
34	Tda Jardin Plaza	12,663	13,440	18,525	2,644,706,305	38,226	1.5	142,764

Fig. 2. Sección de hoja de cálculo de Microsoft Excel con indicadores de las tiendas de OFFCORSS

La fila 3 corresponde a la compañía que consiste en la data agregada de las tiendas, esta fila nos permite hacer una comparación adecuada entre tiendas y la compañía, posteriormente entre tiendas. Con estos indicadores se procede a sacar conclusiones para cada tienda, tomando los atributos con las diferencias más grandes, siendo estos los que le dan la identidad a la tienda misma.

Para una presentación efectiva de los resultados, se automatizó un proceso utilizando Python y las bibliotecas pandas y pptx, generando diapositivas visuales para la alta dirección de mercadeo. Estas diapositivas proporcionaron una representación clara y amigable de los datos, facilitando la identificación de tendencias y patrones relevantes como se muestra en la figura 3.

Para la implementación del modelo de factorización de matrices no negativas, se llevaron a cabo varias etapas clave. En primer lugar, se definieron grupos de clientes en función de su comportamiento de gasto y frecuencia de compra, así como grupos de productos basados en diversas características, como la presencia de descuentos, la colección a la que pertenecen, el

segmento de mercado al que están dirigidos y el tipo de artículo que representan. Estos datos se extrajeron de la base de datos y se organizaron en una estructura de matriz, cruzando los grupos de clientes con los grupos de productos. Este proceso se realizó utilizando Python junto con la biblioteca pandas, como se ilustra en la figura 4.

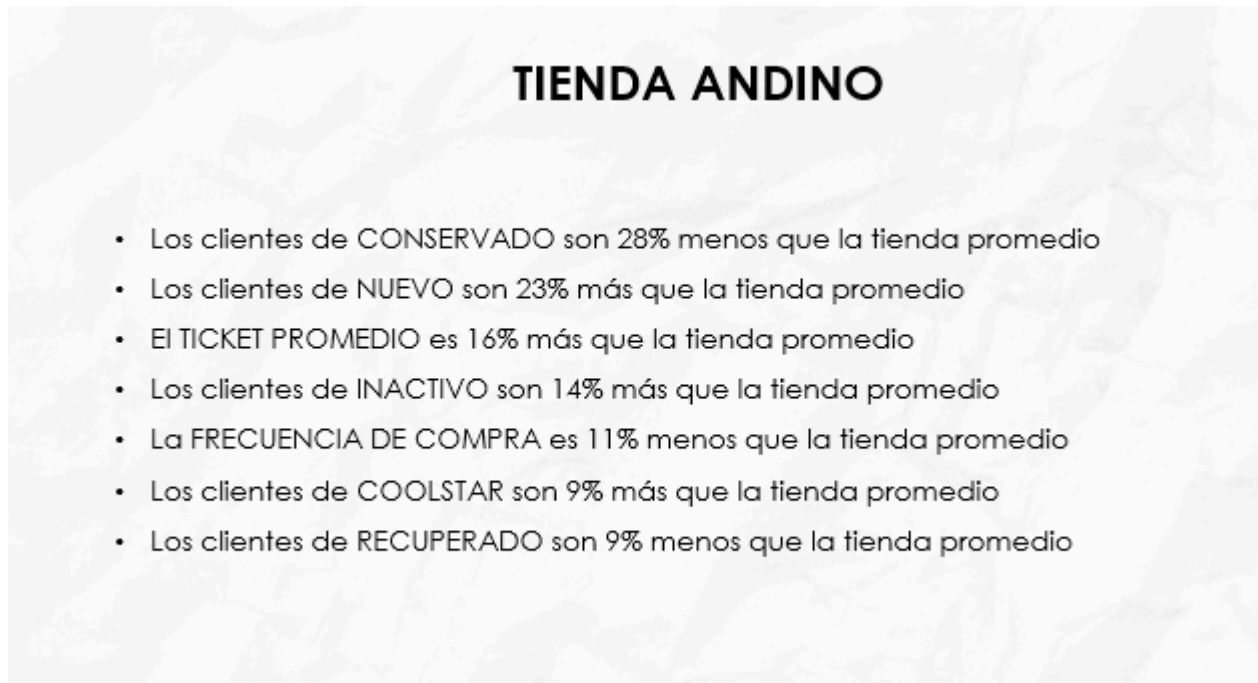


Fig. 3. Diapositiva con los resultados del análisis de indicadores para la tienda Andino.

Para garantizar la precisión del análisis, se excluyeron los clientes clasificados como inactivos o perdidos en términos de su frecuencia de compra, ya que su inclusión podría distorsionar los resultados al no representar los patrones de compra relevantes, el interés está en los clientes activos. Asimismo, se aplicó un filtro para reducir la complejidad del análisis, seleccionando únicamente las 340 columnas más significativas de un total de 1899, que representaban el 90% de las ventas totales. Esto permitió simplificar el cálculo y concentrarse en los datos más relevantes. El resultado fue una matriz con 11 filas, que representan los distintos grupos de clientes, y 340 columnas, que representan los grupos de productos.

Una vez que se construyó esta matriz, se procedió a ajustar el modelo de factorización correspondiente. La implementación del modelo se realizó utilizando la biblioteca scikit-learn [3]. A partir de la primera matriz, se identificaron los patrones que mejor representaban a cada grupo de clientes. Estos patrones se utilizaron luego para filtrar la segunda matriz, donde se inspeccionaron los resultados finales que ya podían ser interpretados. Este proceso permitió identificar tendencias y relaciones significativas entre los distintos grupos de clientes y

productos, proporcionando información valiosa para la toma de decisiones estratégicas en términos de gestión de inventario, segmentación de mercado y estrategias de marketing.

product	0, BASICOS, BEBÉ NIÑA, CAMISETA	0, BASICOS, NIÑO, CAMISETA	0, BASICOS, BEBÉ NIÑO, CAMISETA	0, PLAYA, BEBÉ NIÑO, ROPA DE BAÑO	0, CUIDADO PERSONAL, CATEGORÍAS, CUIDADO PERSONAL	0, INTERIOR, NIÑO, SET	0, BASICOS, NEWBORN BOY, CONJUNTO	0, BASICOS, NIÑA, CAMISETA	0, PIJAMAS, NIÑO, PIJAMA	0, PIJAMAS, NIÑA, PIJAMA	...
group											
COOLSTAR, CONSERVADO	0.0	0.0	17.0	0.0	17.0	0.0	0.0	18.0	0.0	17.0	...
COOLSTAR, NUEVO	13.0	0.0	0.0	0.0	13.0	0.0	0.0	0.0	0.0	0.0	...
COOLSTAR, RECUPERADO	0.0	0.0	0.0	0.0	52.0	0.0	0.0	26.0	0.0	0.0	...
NEW TALENT, CONSERVADO	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	...
NEW TALENT, NUEVO	2.0	5.0	3.0	0.0	0.0	7.0	1.0	3.0	0.0	1.0	...
NEW TALENT, RECUPERADO	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	...
ROCKSTAR, CONSERVADO	15998.0	15998.0	15236.0	15156.0	14314.0	12630.0	11788.0	10946.0	10946.0	10104.0	...
ROCKSTAR, NUEVO	333.0	222.0	357.0	222.0	333.0	0.0	135.0	0.0	0.0	111.0	...
STAR, CONSERVADO	5.0	25.0	12.0	34.0	12.0	0.0	40.0	0.0	0.0	0.0	...
STAR, NUEVO	16.0	30.0	7.0	8.0	0.0	0.0	0.0	28.0	12.0	15.0	...
STAR, RECUPERADO	0.0	6.0	5.0	0.0	0.0	6.0	14.0	0.0	0.0	6.0	...

11 rows x 340 columns

Fig. 4. Matriz de entrada para el modelo NMF.

VI. ANÁLISIS

El análisis de los indicadores revela variaciones significativas en el desempeño de cada tienda de OFFCORSS, tanto en términos de ventas como en el comportamiento del cliente. Esta evaluación comparativa permite identificar áreas de oportunidad y fortalezas específicas en cada establecimiento. Por ejemplo, se observa que la tienda Andino tiende a atraer nuevos clientes que realizan compras sustanciales en su primera visita, pero muestra una baja tasa de retención. Este patrón de comportamiento sugiere la necesidad de estrategias enfocadas en la fidelización del cliente para mejorar la rentabilidad a largo plazo de la tienda.

Por otro lado, la aplicación del modelo de factorización de matrices no negativas (NMF) revela patrones de consumo distintivos entre diferentes segmentos de clientes y productos. Estos patrones proporcionan información valiosa sobre las preferencias de los clientes y las tendencias de compra, lo que permite adaptar estrategias de surtido de productos y marketing de manera más precisa. Por ejemplo, para los clientes de la tienda Andino en la categoría COOLSTAR y CONSERVADO, se identificó un patrón de compra específico que incluye productos relacionados con las categorías de SIN DESCUENTO, PLAYA, NIÑA, ROPA DE BAÑO, DIGITAL NATURE, BEBÉ NIÑA, CAMISETA, SPECIAL OCCASIONS, BEBÉ NIÑA, VESTIDO.

VII. CONCLUSIONES

En conclusión, el análisis de los indicadores y la aplicación del modelo de NMF proporcionan información valiosa para mejorar la estrategia comercial de OFFCORSS. Los hallazgos obtenidos permiten adaptar las estrategias comerciales a las necesidades y preferencias específicas de cada tienda, mejorando así la retención de clientes y la rentabilidad a largo plazo. Además, la capacidad de la NMF para identificar patrones de consumo distintivos entre diferentes segmentos de clientes y productos facilita la personalización de la oferta de productos y el diseño de estrategias de marketing más efectivas. En última instancia, el uso efectivo de estos análisis puede ayudar a OFFCORSS a mantener su posición como líder en el mercado de ropa infantil, ofreciendo una experiencia de compra única y relevante para cada cliente.

REFERENCIAS

- [1] J. An, H. Kwak, S. Jung, J. Salminen, M. Admad, and B. Jansen, "Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data," ACM Transactions on the Web, vol. 12, no. 5, November 2018, doi: 10.1145/3265986. Disponible en: https://www.researchgate.net/publication/328683848_Imaginary_People_Representing_Real_Numbers_Generating_Personas_from_Online_Social_Media_Data
- [2] Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 5, 13-22.
- [3] Scikit-learn developers (2023). Sklearn Decomposition NMF. Scikit-learn documentation. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html>

ANEXOS

1. Hoja de cálculo con el análisis de los indicadores de las tiendas de OFFCORSS.

Disponible en:

<https://docs.google.com/spreadsheets/d/1ua1w-VVMSHEUcFGNt3zCd6oYKV8OY2MI/edit?usp=sharing&oid=102234840403543070982&rtpof=true&sd=true>

2. Cuaderno de python con el modelado de NMF. Disponible en:

<https://drive.google.com/file/d/1FpW-SsBqjrK5SVoJmlR6Tt5677QBcmkz/view?usp=sharing>