**DESARROLLO DE UN MODELO DE RECOMENDACIÓN DE LIBROS PERSONALIZADO UTILIZANDO INTELIGENCIA ARTIFICIAL PARA FOMENTAR LA LECTURA EN ADULTOS**

**Integrantes**

Angélica María Arboleda Salazar

María Paz Estrada Granada

Luis Alberto Ramírez Ospina

**Ejecutores**

Natalia Betancur Herrera

Frank Yesid Zapata Castaño

Margarita María Orozco

**Universidad de Antioquia, Universidad de Caldas, UBICUA**

**BOOTCAMP Inteligencia Artificial**

**Talento TECH 2.0**

**Colombia, Manizales**

**Marzo 2025**

**Agradecimientos**

Gracias al programa Talento Tech, a los profesores Frank Yesid Zapata y Natalia Betancur, a los monitores Andrés Sánchez y Natalia Suárez y a todas las personas que hicieron posible la segunda versión del bootcamp. Por el apoyo y su guía fundamental en este proyecto.

Gracias también a UBICUA, Universidad de Antioquia y Universidad de Caldas por sus instalaciones para la realización del trabajo.

### 1. INTRODUCCIÓN

La lectura ha sido históricamente un pilar fundamental en el desarrollo personal y cultural de las sociedades. Desde una perspectiva literaria, los libros no solo son vehículos de entretenimiento, sino también herramientas poderosas para la introspección y el crecimiento personal. La literatura ofrece a los individuos la oportunidad de explorar mundos diversos, empatizar con personajes de diferentes contextos y reflexionar sobre sus propias experiencias y valores. Este proceso de inmersión y reflexión contribuye al desarrollo de habilidades críticas, como el pensamiento analítico y la capacidad de resolución de problemas.

Desde el ámbito político, la promoción de la lectura es esencial para el fortalecimiento de una ciudadanía informada y participativa. La capacidad de leer y comprender textos complejos permite a los ciudadanos evaluar críticamente la información, participar en debates públicos y tomar decisiones informadas sobre asuntos que afectan sus vidas y comunidades. Además, la lectura fomenta la inclusión social al proporcionar acceso a una amplia gama de conocimientos y perspectivas, lo que es crucial en sociedades cada vez más diversas y globalizadas.

Culturalmente, la lectura actúa como un puente entre generaciones, preservando y transmitiendo el patrimonio cultural y las tradiciones. Al mismo tiempo, abre las puertas a nuevas ideas y movimientos culturales, enriqueciendo el tejido social con una diversidad de voces y experiencias. En un mundo donde la tecnología y los medios digitales dominan gran parte de la comunicación, la lectura sigue siendo una práctica vital que nutre la imaginación, la creatividad y el entendimiento intercultural.

Es posible afirmar entonces que el ejercicio de la lectura es una herramienta fundamental para el desarrollo cognitivo, emocional y social de las personas. Sin embargo, aunque el nivel de alfabetización se ha incrementado en algunos países a lo largo de los años, paralelamente, en la era digital, muchas personas han perdido el interés por la lectura, lo que limita su capacidad para desarrollar habilidades críticas como el pensamiento analítico y la creatividad. Este proyecto busca abordar este desafío mediante el desarrollo de un modelo predictivo basado en inteligencia artificial que recomiende libros personalizados, con el objetivo de fomentar el hábito de la lectura y mejorar la experiencia lectora en diferentes segmentos de la población.

En este caso, la herramienta de la inteligencia artificial permitirá analizar las preferencias, intereses y características individuales de cada persona para ofrecer recomendaciones de libros que se ajusten a sus gustos y necesidades. Este enfoque busca no solo incrementar la tasa de lectura, sino también generar un impacto social importante al facilitar el acceso a libros relevantes y atractivos para cada lector.

El proyecto se estructura en tres fases principales: caracterización de lectores, desarrollo del modelo y el posible impacto de esta estrategia para fomentar la lectura. Dicho enfoque integral busca no solo resolver el problema a nivel individual, sino también sentar las bases para una intervención comunitaria más amplia que impulse el acceso a la lectura en distintos contextos.

### 2. PLANTEAMIENTO DEL PROBLEMA

La baja tasa de lectura en la población adulta de México, donde solo el 42.5% de las personas de 18 años en adelante han leído al menos un libro en el último año representa un desafío significativo para el desarrollo cultural y educativo del país. Esta problemática se ve exacerbada por factores como la falta de tiempo, el acceso limitado a libros y la ausencia de recomendaciones personalizadas que puedan motivar a los individuos a leer más.

El problema que se plantea es: **¿Cómo puede un sistema de recomendación basado en inteligencia artificial mejorar significativamente los hábitos de lectura en la población adulta de México, al proporcionar recomendaciones de libros personalizadas que consideren las preferencias, intereses y características individuales de cada lector?**

Para abordar este problema, se plantea desarrollar un modelo predictivo que identifique patrones en las preferencias de lectura y, asimismo, sea capaz de adaptarse a las necesidades cambiantes de los lectores. Este modelo debe integrar datos demográficos, históricos de lectura, y posiblemente datos contextuales como el tiempo disponible para la lectura y el acceso a diferentes formatos de libros (físicos, digitales, audiolibros).

El objetivo final es crear un sistema que no solo genere un incremento en la cantidad de libros leídos por individuo, sino que también mejore la calidad de la experiencia lectora, fomentando un hábito sostenible y enriquecedor. Este enfoque tiene el potencial de transformar la relación de los individuos con la lectura, contribuyendo al desarrollo personal y cultural de la población.

Conforme la información consultada en la cual basaremos este proyecto se tiene que:

* Según el Módulo de Lectura (MOLEC) del INEGI, en México, solo el 42.5% de la población de 18 años y más leyó al menos un libro en los últimos 12 meses (datos más recientes disponibles).
* La falta de hábitos de lectura se asocia con factores como la falta de tiempo, acceso a libros y recomendaciones poco personalizadas.
* Un sistema de recomendación basado en IA puede ayudar a conectar a los lectores con libros que se ajusten a sus intereses y necesidades.

Este proyecto busca abordar esta problemática mediante el desarrollo de un modelo predictivo basado en inteligencia artificial que recomiende libros personalizados, teniendo en cuenta las preferencias, intereses y características individuales de cada lector. **La aspiración a largo plazo es fomentar el hábito de la lectura en la población y mejorar su experiencia lectora de manera significativa.**

### 3. OBJETIVOS

#### 3.1 Objetivo General

Desarrollar un sistema de recomendación de libros personalizado utilizando técnicas de inteligencia artificial (IA) para incentivar el hábito de la lectura en adultos y de esta manera enriquecer su desarrollo personal y cultural.

#### 3.2 Objetivos Específicos

* Realizar una revisión parcial de la literatura y bases de datos existentes sobre hábitos de lectura y preferencias literarias en diferentes segmentos de la población.
* Analizar datos sobre los hábitos de lectura y preferencias de los participantes mediante metodologías de recolección de datos y análisis de patrones.
* Desarrollar y validar un modelo basado en inteligencia artificial para recomendar libros personalizados.

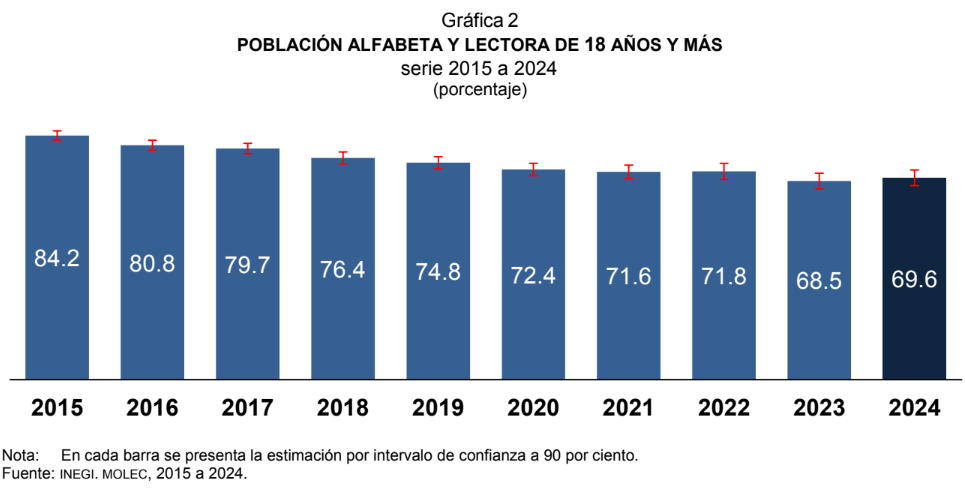
**4. JUSTIFICACIÓN**

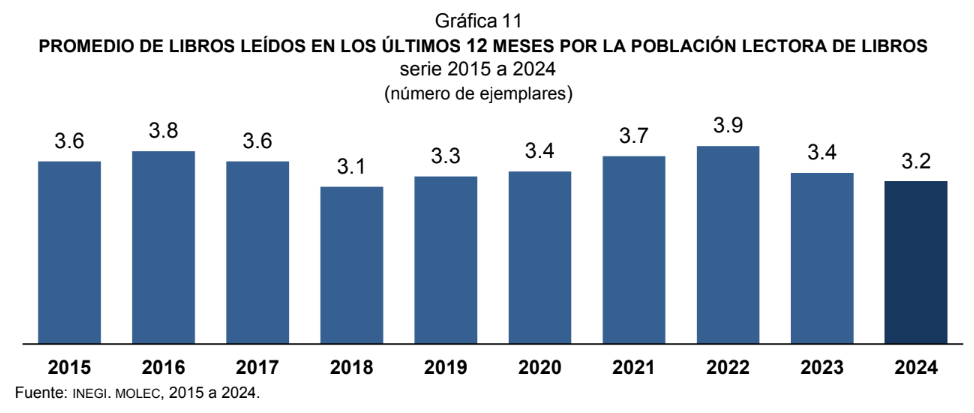
Según el DANE y estudios como ENLEC 2017, en Colombia el índice de lectura ha mejorado, especialmente en ciudades como Bogotá, sin embargo, aún hay desafíos para mantener el hábito en adultos jóvenes y mayores.

La pandemia generó un aumento en la lectura, como lo indica El País (2024), pero es necesario consolidar este hábito en la post pandemia.

La Cámara Colombiana del Libro reporta que los colombianos leen en promedio 3.7 libros al año (Infobae, 2024), lo que muestra un avance, pero aún está por debajo de países líderes en lectura como Argentina o España (Lectupedia).

En México, según SciELO, el porcentaje de lectura es bajo, lo que contrasta con países como Argentina, donde se leen más libros per cápita (Lectupedia).





En la era digital, la lectura se presenta como un refugio que contrasta con la inmediatez y superficialidad de la información online. Lejos de ser solo una herramienta para mejorar el rendimiento académico, la lectura es una experiencia enriquecedora que abre puertas a nuevos mundos, promueve la reflexión y fortalece las conexiones humanas.

Se evidencia un incremento en la alfabetización de la población urbanizada de México, sin embargo, paralelamente, es posible observar que la cantidad

“La alfabetización es mucho más que saber leer y escribir, es cómo nos comunicamos en la sociedad. Tiene que ver con las prácticas y relaciones sociales, con el conocimiento, el idioma y la cultura. La alfabetización se adentra en muchos aspectos de la vida. La alfabetización abre oportunidades de aprender, encontrar un mejor empleo y recibir remuneraciones más altas. La alfabetización también repercute en la condición social, la participación política, la expresión cultural, la supervivencia lingüística, el acceso a la atención de la salud y la prestación eficaz de servicios sociales, entre otras cosas.” (UNESCO, 2013).

Desde el enfoque del desarrollo humano, la lectura y la escritura son prácticas esenciales que facilitan el acceso a la información y la comunicación, promoviendo la autonomía y la toma de decisiones. Estas actividades están relacionadas con los avances sociales y la participación política, y contribuyen a ampliar la libertad humana, permitiendo a las personas vivir según sus propias elecciones (CERLALC, 2006; Sen, 2002).

Así pues, la evaluación y medición de los impactos de las políticas de lectura y escritura, y su relación con el desarrollo sociocultural y económico de los países, es de vital trascendencia.

Psicología del lector: Estudios como el de Santo Tomás (repositorio USTA) y la Maestría sobre Lectura y Desarrollo Humano (FLACSO) destacan la importancia de la personalización para fomentar la lectura.

Adicionalmente, existen muchos beneficios que se presentan al incrementar el nivel de lectura, de los cuales se mencionan algunos a continuación:

**Transformación personal y social**  
La lectura permite adentrarse en diferentes culturas, realidades y perspectivas, ofreciendo mucho más que conocimientos técnicos o académicos. A través de los libros, las personas pueden:

* **Explorar universos paralelos:** Cada obra es una invitación a descubrir historias, contextos y formas de pensar que enriquecen la imaginación y la creatividad.
* **Desarrollar empatía y comprensión:** Al conocer las vivencias y emociones de personajes diversos, se fomenta la capacidad de ponerse en el lugar del otro y comprender realidades distintas a la propia.
* **Estimular el pensamiento crítico:** La lectura invita a cuestionar, analizar y reflexionar sobre el entorno, ayudando a construir opiniones fundamentadas y a tomar decisiones informadas en la vida cotidiana.

**Beneficios más allá del aula**  
Aunque el impacto en el ámbito académico es innegable, los beneficios de la lectura van mucho más allá del rendimiento escolar. Entre ellos destacan:

* **Reducción del estrés:** Sumergirse en una buena historia puede ser una forma efectiva de relajarse y desconectar de las presiones diarias.
* **Fomento de la creatividad:** La exposición a ideas y narrativas variadas estimula la imaginación y abre la mente a nuevas formas de ver el mundo.
* **Fortalecimiento del sentido de comunidad:** Los clubes de lectura, foros y encuentros literarios crean espacios para el diálogo, el intercambio de ideas y la construcción de redes de apoyo social.

**Re-conceptualizando la lectura**  
Este proyecto propone cambiar la percepción de la lectura, alejándose de la idea de una actividad exclusivamente académica o tediosa. La intención es mostrar que leer es una experiencia dinámica y liberadora, que permite a cada individuo explorar nuevas dimensiones del conocimiento y del ser. Se busca incentivar la práctica lectora mediante:

* **Actividades interactivas y participativas:** Espacios donde los lectores puedan compartir sus experiencias, discutir sobre sus lecturas y descubrir recomendaciones que se adapten a sus intereses.
* **Acceso a materiales atractivos y relevantes:** La selección de libros que conecten con las realidades y pasiones de la comunidad, facilitando una identificación más profunda con las historias y los personajes.
* **Uso de herramientas digitales:** Integrar plataformas y recursos online que hagan de la lectura una experiencia accesible, moderna y alineada con los hábitos tecnológicos actuales.

### 6. ALCANCE

Este proyecto busca transformar la experiencia de la lectura en una actividad más accesible y atractiva para todas las personas desde los 18 años en adelante, sin importar su contexto educativo. Para ello, se recopilaron bases de datos que permiten comprender los hábitos de lectura y las preferencias literarias de los participantes. Con base en esta información, se desarrollará un sistema de recomendaciones personalizadas utilizando inteligencia artificial y aprendizaje automático, asegurando que cada lector descubra libros alineados con sus intereses y nivel de lectura.

El proyecto se enfocará en adultos de 18 años en adelante, utilizando datos de hábitos de lectura en México y se integrarán bases de datos como MOLEC para entrenar el modelo de IA.

Además, se validará y ajustará el modelo predictivo para mejorar su precisión y aplicabilidad en distintos contextos. La implementación del sistema permitirá que cualquier persona encuentre lecturas más significativas y motivadoras, impulsando el hábito lector de manera natural.

El impacto del proyecto no se limitará a un grupo específico, sino que también contribuirá a la comunidad en general al generar información valiosa sobre estrategias efectivas de fomento a la lectura. Los resultados obtenidos servirán para sensibilizar sobre la importancia del acceso a lecturas personalizadas y para desarrollar iniciativas que fortalezcan el hábito lector en distintos entornos, ya sea en el hogar, espacios públicos o comunidades digitales.

La evaluación del Plan de Lectura realizada planteó como hipótesis central que la generación de una cultura de lectura y escritura contribuye a la consolidación de espacios de participación social, al desarrollo social y económico y al ejercicio responsable de los derechos políticos y sociales de los ciudadanos de un territorio. De esta forma, el posicionamiento y fortalecimiento de los planes de lectura dentro de las agendas gubernamentales es decisivo para el desarrollo de capacidades de la población participante.

Este enfoque busca trascender la visión tradicional de la lectura como una obligación académica, promoviendo una cultura en la que leer sea visto como una fuente de placer, conocimiento y conexión con diferentes realidades.

El proyecto también incluye la validación y ajuste del modelo para asegurar su precisión y aplicabilidad en diversos contextos. Adicionalmente, se planea implementar un sistema de recomendaciones personalizadas que permita a los estudiantes descubrir libros que se ajusten a sus intereses y niveles de lectura.

### 7. METODOLOGÍA

#### Diseño del proyecto:

#### 7.1 Investigación y descripción de la base de datos

Obtención de información sobre hábitos de lectura mediante encuestas y cuestionarios digitales (MOLEC).

La base de datos recopila información clave sobre los hábitos de lectura y las preferencias literarias y características demográficas de los participantes. Las variables consideradas incluirán:

* **Variables numéricas:** Edad, tiempo promedio dedicado a la lectura, número de libros leídos en el último año, frecuencia de lectura semanal.
* **Variables categóricas:** Género, género literario preferido (ficción, no ficción, fantasía, ciencia ficción, etc.), formato de lectura preferido (libro físico, digital, audiolibro), entre otros.

En este caso El MOLEC del INEGI proporciona datos detallados sobre los hábitos de lectura en México, incluyendo:

* Porcentaje de lectores por género, edad y nivel educativo.
* Preferencias de lectura (libros, revistas, periódicos, etc.).
* Razones por las que las personas no leen (falta de tiempo, interés o acceso).
* Estudios complementarios muestran que la falta de recomendaciones personalizadas es una barrera para fomentar la lectura.

#### 7.2. Recopilación de datos sobre lectura

Utilizar bases de datos como datos.gov.co, Kaggle, y ENLEC 2017 para obtener información sobre hábitos de lectura.

#### 7.3 Preprocesamiento y limpieza de datos

Normalización de datos, tratamiento de valores nulos y eliminación de posibles inconsistencias.

Limpieza y organización de datos con herramientas como Python (Pandas, NumPy).

1. **Desarrollo del modelo:** Implementación de técnicas de inteligencia artificial para realizar el código.
2. **Validación y ajuste del modelo:** Evaluación de la precisión y optimización del modelo para mejorar su desempeño.
3. **Pruebas con datos reales de usuarios de México.**

#### 7.3 Entendimiento de los Datos

Se llevará a cabo un análisis exploratorio de datos (EDA) para detectar patrones, tendencias y correlaciones entre variables. Esto incluirá la visualización de correlaciones y la detección de valores atípicos. El análisis permitirá una mejor comprensión de los factores que influyen en la elección de libros y ayudará a afinar el modelo predictivo.

#### 7.4 Preparación del Modelo

Para la creación del sistema de recomendación se utilizará el algoritmo RTDa, reconocido por su eficiencia en la clasificación de datos desbalanceados. Este algoritmo permitirá generar correlaciones precisas basadas en los intereses y hábitos de lectura de las personas.

#### 7.5 Evaluación del Modelo

El desempeño del modelo se evaluará utilizando una matriz de correlación y métricas como precisión, recall y F1-score.

**8. MODELADO**

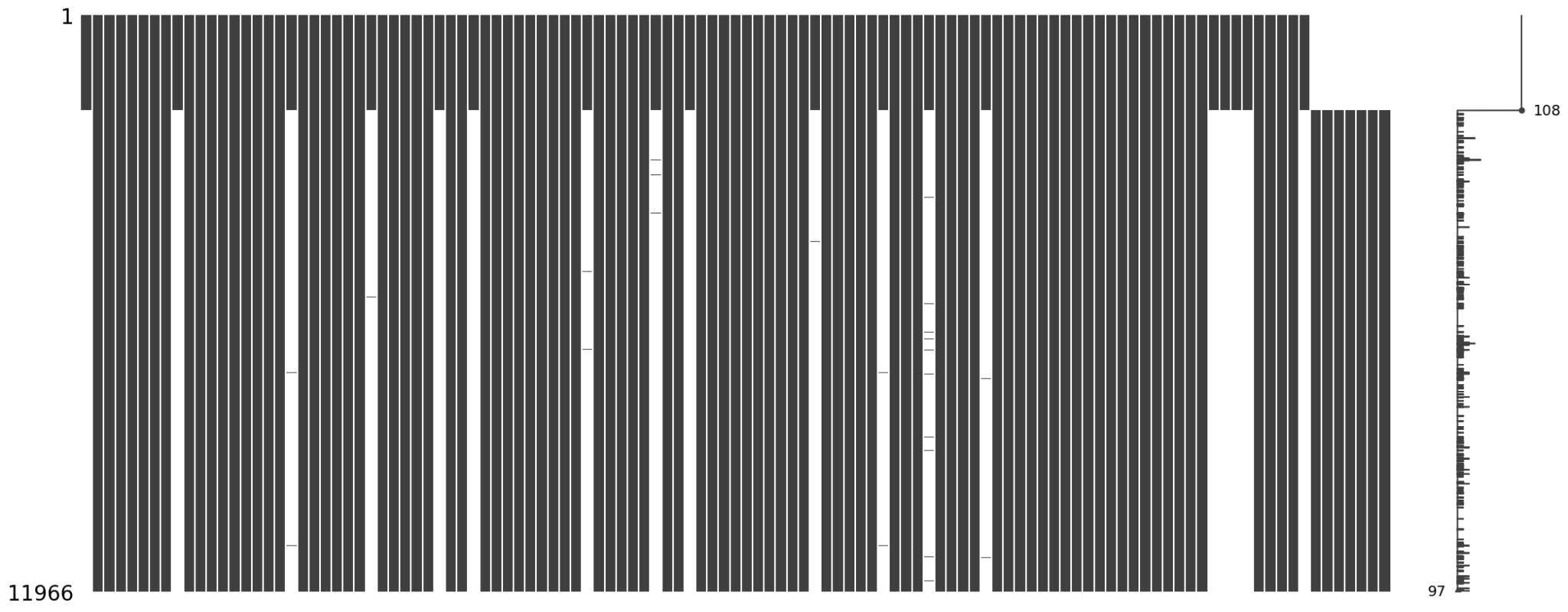
Con respecto a las variables, se presenta la traducción de su significado en el archivo “Resumen” de cada carpeta correspondiente. Sigue el Resumen del año 2024 como muestra.

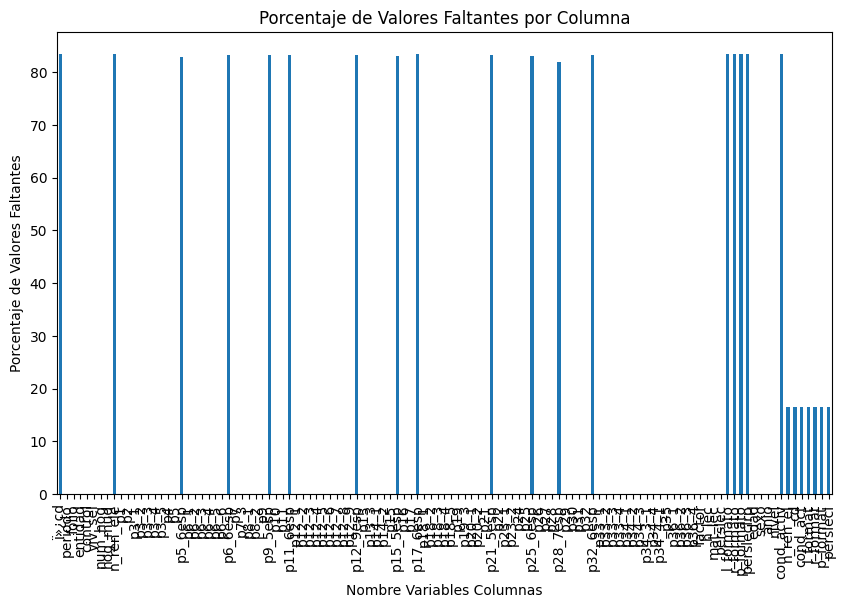
**Ruta:**(IA) Inteligencia Artificial>datos >conjunto\_de\_datos\_molec\_2024\_>catalogos>Resumen

| NOMBRE\_CAMPO | NEMÓNICO | CATÁLOGO |
| --- | --- | --- |
| Folio | folio |  |
| Entidad | entidad | entidad |
| Control | control |  |
| Vivienda seleccionada | viv\_sel | viv\_sel |
| Número de hogar | num\_hog | num\_hog |
| hogar mudado | hog\_mud | hog\_mud |
| Número de renglón del informante elegido | n\_ren\_el | n\_ren\_el |
| Ciudad autorepresentada | cd | cd |
| Periodo | periodo | periodo |
| Sexo | sexo | sexo |
| Edad | edad | edad |
| Nivel de instrucción | nivel | nivel |
| ¿Hasta qué año o grado aprobó en la escuela? | anio | anio |
| Condición de actividad ? La semana pasada… | cond\_act | cond\_activ |
| 1. ¿Usted sabe leer y escribir un recado? | p1 | c\_si\_no |
| 2. ¿Usted acostumbra leer? | p2 | c\_si\_no |
| 3a. en los últimos doce meses, leyó LIBROS? | p3\_1 | c\_si\_no |
| 3b. en los últimos tres meses, leyó REVISTAS? | p3\_2 | c\_si\_no |
| 3c. la semana pasada, leyó PERIÓDICOS? | p3\_3 | c\_si\_no |
| 3d. lee HISTORIETAS como Archie, Memin Pinguin o Marvel? | p3\_4 | c\_si\_no |
| 3e. sin considerar redes sociales como Facebook y Twitter, lee PÁGINAS DE INTERNET, FOROS O BLOGS? | p3\_5 | c\_si\_no |
| 4. ¿Cuántos libros leyó en los últimos doce meses? | p4 | cp\_cantidad |
| 5. ¿Cuál fue el motivo principal por el que usted leyó el (los) libro(s)? | p5 | cp\_motivo |
| 5. ¿Cuál fue el motivo principal por el que usted leyó el (los) libro(s)? Especifique | P5\_6esp |  |
| 6. ¿El (Los) libro(s) que leyó fue(ron) manual(es), guía(s) o recetario(s)? | p6\_1 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) específico(s) de alguna materia o profesión, libro(s) de texto o de uso universitario (ciencia, tecnología, matemáticas, medicina, comercio, derecho, etcétera)? | p6\_2 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) de autoayuda, superación personal o religioso(s)? | p6\_3 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) de literatura (novela, cuento, ciencia ficción, poesía, teatro, etcétera)? | p6\_4 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) de cultura general (acontecimientos, biografías, temas sociales o de interés, etcétera)? | p6\_5 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) Otro | p6\_6 | c\_si\_no |
| 6. ¿El (Los) libro(s) que leyó fue(ron) Otro. Especifique | P6\_6esp |  |
| 7. ¿El libro (La mayoría de los libros) fue(ron) descargado(s) gratuitamente, regalado(s) o prestado(s) o comprado(s) (impreso o digital)? | P7 | cp\_adquisición |
| 7. ¿El libro (La mayoría de los libros) fue(ron) fue(ron)comprado(s) (impreso o digital). Aproximadamente, ?cuánto gastó? | P7\_3 | cp\_gasto |
| 8. ¿El (Los) libro(s) que leyó se encontraba(n) en formato digital? | p8\_1 | c\_si\_no |
| 8. ¿El (Los) libro(s) que leyó se encontraba(n) impreso(s)? | p8\_2 | c\_si\_no |
| 9. ¿En qué lugar leyó el (los) libro(s) principalmente? | p9 | cp\_lugar |
| 9. ¿En qué lugar leyó el (los) libro(s) principalmente? Especifique | p9\_5esp |  |
| 10. ¿Cuántas revistas leyó en los últimos tres meses? | p10 | cp\_cantidad |
| 11. ¿Cuál fue el motivo principal por el que usted leyó la(s) revista(s)? | p11 | cp\_motivo |
| 11. ¿Cuál fue el motivo principal por el que usted leyó la(s) revista(s)? Especifique | p11\_6esp |  |
| 12. ¿La(s) revista(s) que leyó fue(ron) de deportes o automovilismo (Sports Illustrated, Automóvil panamericano, Motociclismo, etcétera)? | p12\_1 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) de política(s), económica(s) o financiera(s) (Proceso, Expansión, Newsweek, Alto nivel, Poder y negocios, etcétera)? | p12\_2 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) didáctica(s), sobre manualidades, cocina o decoración (Carpintería, Dos agujas, Cocina Fácil, etcétera)? | p12\_3 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) de bienestar o salud (Men’s health, Women’s health, Cuerpo mente, Salud vital, etcétera)? | p12\_4 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) religiosas (Atalaya, Vida nueva, Despertares, etcétera)? | p12\_5 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) especializada(s), técnica(s) o científica(s) (National Geographic, Conozca más, Muy interesante, Quo, etcétera)? | p12\_6 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) de cultura general o temas de interés (Selecciones, Contenido, Algarabía, etcétera)? | p12\_7 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) de entretenimiento (TV Notas, Cosmopolitan, Vogue, GQ, Premiere, Rolling Stone, etcétera)? | p12\_8 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron)? Otro | p12\_9 | c\_si\_no |
| 12. ¿La(s) revista(s) que leyó fue(ron) otro.Especifique | p12\_9esp |  |
| 13. ¿La revista (La mayoría de las revistas) fue(ron) descargada(s) gratuitamente, regalada(s) o prestada(s) o comprada(s) (impresa o digital)? | p13 | cp\_adquisición |
| 13. ¿La revista (La mayoría de las revistas) fue(ron) comprada(s) (impresa o digital). Aproximadamente, ? cuánto gastó? | P13\_3 | cp\_gasto |
| 14. ¿La(s) revista(s) que leyó se encontraba(n) en formato digital? | p14\_1 | c\_si\_no |
| 14. ¿La(s) revista(s) que leyó se encontraba(n) impresa(s) | p14\_2 | c\_si\_no |
| 15. ¿En qué lugar leyó la(s) revista(s) principalmente? | p15 | cp\_lugar |
| 15. ¿En qué lugar leyó la(s) revista(s) principalmente? Otro. Especifique | p15\_5esp |  |
| 16. ¿Cuántos periódicos leyó la semana pasada? | p16 | cp\_cantidad |
| 17. ¿Cuál es el motivo principal por el que usted leyó el (los) periódico(s)? | p17 | cp\_motivo |
| 17. ¿Cuál es el motivo principal por el que usted leyó el (los) periódico(s)? Otro. Especifique | p17\_6esp |  |
| 18. Del (De los) periódico(s), leyó la sección de política? | p18\_1 | c\_si\_no |
| 18. Del (De los) periódico(s), leyó la sección de deportes? | p18\_2 | c\_si\_no |
| 18. Del (De los) periódico(s), leyó la sección de cultura o espectáculos? | p18\_3 | c\_si\_no |
| 18. Del (De los) periódico(s), leyó la sección de sociales? | p18\_4 | c\_si\_no |
| 18. Del (De los) periódico(s), leyó la sección policiaca? | p18\_5 | c\_si\_no |
| 19. ¿El periódico (La mayoría de los periódicos) fue(ron) descargado(s) gratuitamente, regalado(s), prestado(s) o comprado(s) (impreso o digital)? | p19 | cp\_adquisición |
| 19. ¿El periódico (La mayoría de los periódicos) fue(ron) comprado(s) (impreso o digital). Aproximadamente, ?cuánto gastó? | p19\_3 | cp\_gasto |
| 20. ¿El (los) periódico(s) que leyó se encontraba(n) en formato digital? | p20\_1 | c\_si\_no |
| 20. ¿El (los) periódico(s) que leyó se encontraba(n) impresas | p20\_2 | c\_si\_no |
| 21. ¿En qué lugar leyó el(los) periódico(s) principalmente? | p21 | cp\_lugar |
| 21. ¿En qué lugar leyó el(los) periódico(s) principalmente? Otro. Especifique | p21\_5esp |  |
| 22. ¿Usted acostumbra leer historietas como Archie, Memín Pinguín o Marvel… | p22 | cp\_frecuencia |
| 23. ¿La mayor parte de las historietas que lee se encuentran en formato digital? | p23\_1 | c\_si\_no |
| 23. ¿La mayor parte de las historietas que lee se encuentran impresas | p23\_2 | c\_si\_no |
| 24. Sin considerar redes sociales como Facebook y Twitter, ¿las páginas de internet, foros o blogs acostumbra leerlos… | p24 | cp\_frecuencia |
| 25. ¿Cuál es el motivo principal por el que usted leyó páginas de internet, foros o blogs? | p25 | cp\_motivo |
| 25. ¿Cuál es el motivo principal por el que usted leyó páginas de internet, foros o blogs? Otro. Especifique | p25\_6esp |  |
| 26. Aproximadamente, ¿cuántos minutos continuos lee algún material de lectura? | p26 | cp26 |
| 27. Cuando lee, ¿realiza otra actividad al mismo tiempo? | p27 | c\_si\_no |
| 28. Principalmente, ¿qué otra actividad realiza mientras lee? | p28 | cp28 |
| 28. Principalmente, ¿qué otra actividad realiza mientras lee? Otra Especifique | p28\_7esp |  |
| 29. Cuando lee algún Alfanumérico , ¿cree que la velocidad de su lectura es… | p29 | cp29 |
| 30. Cuando usted lee algún Alfanumérico , ¿la comprensión de su lectura es… | p30 | cp30 |
| 31. ¿Acostumbra consultar otros materiales (diccionario, enciclopedia, internet, etcétera) para buscar más información sobre su lectura? | p31 | c\_si\_no |
| 32. ¿Cuál es el motivo principal por el que usted no lee los materiales de lectura antes mencionados? | p32 | cp32 |
| 32. ¿Cuál es el motivo principal por el que usted no lee los materiales de lectura antes mencionados? Otro Especifique | p32\_6esp |  |
| 33. En los últimos tres meses, ¿usted asistió a la sección de libros y revistas de una tienda departamental? | p33\_1 | c\_si\_no |
| 33. En los últimos tres meses, ¿usted asistió a un puesto de libros o revistas usados? | p33\_2 | c\_si\_no |
| 33. En los últimos tres meses, ¿usted asistió a una biblioteca? | p33\_3 | c\_si\_no |
| 33. En los últimos tres meses, ¿usted asistió a una librería? | p33\_4 | c\_si\_no |
| 34. ¿Cuándo usted era niño(a) sus padres o tutores lo(a) llevaban a bibliotecas o librerías? | p34\_1 | c\_si\_nr |
| 34. ¿Cuándo usted era niño(a) veía a sus padres o tutores leer? | p34\_2 | c\_si\_nr |
| 34. ¿Cuándo usted era niño(a) sus padres o tutores le leían? | p34\_3 | c\_si\_nr |
| 34. ¿Cuándo usted era niño(a) ? ¿Quién le leía principalmente? | p34\_3\_1 | cp34c1 |
| 34. ¿Cuándo usted era niño(a) sin contar los libros de texto, había libros en su casa? | p34\_4 | c\_si\_nr |
| 34. ¿Cuándo usted era niño(a)?Aproximadamente había… | p34\_4\_1 | cp34d1 |
| 35. ¿En su infancia usted asistió a la escuela primaria? | p35 | c\_si\_no |
| 36. ¿Cuando usted era niño(a), en su escuela primaria sus maestros le pedían leer otros libros además de los de texto? | P36\_1 | c\_si\_nr |
| 36. ¿Cuando usted era niño(a), en su escuela primaria sus maestros después de realizar alguna lectura, le pedían comentar o exponer lo leído? | P36\_2 | c\_si\_nr |
| 36. ¿Cuando usted era niño(a), en su escuela primaria sus maestros lo(a) motivaban para que leyera otros libros? | P36\_3 | c\_si\_nr |
| 36. ¿Cuando usted era niño(a), en su escuela primaria sus maestros lo(a) animaban para visitar bibliotecas? | p36\_4 | c\_si\_nr |
| Factor | factor |  |
| Hábito de lectura | h\_lec | cph\_lec |
| Materiales de lectura | mat\_lec | cpmat\_lec |
| Status de las personas en condición de lectura de materiales considerados por MOLEC | perslec | cpperslec |
| Formato de libros | l\_format | cp\_formato |
| Formato de revistas | r\_format | cp\_formato |
| Formato de periódicos | p\_format | cp\_formato |
| Personas lectoras de libros, revistas o periódicos | perslecl | c\_si\_no |

### Pasos para el Análisis de Datos

1. **Descarga y Exploración de Datos:**
   * Descarga de los datos desde el enlace proporcionado.
   * Carga de los datos en un entorno de análisis, como Python (usando pandas) o R.
   * Realización de una exploración inicial para entender la estructura de los datos: tipos de variables, valores faltantes, etc.
2. **Preprocesamiento de Datos:**
   * Limpieza de los datos eliminando o imputando valores faltantes.
   * Normalización o estandarización de las variables si es necesario.
   * Realización de transformaciones de datos, como codificación de variables categóricas.
3. **Análisis Descriptivo:**
   * Cálculo de estadísticas descriptivas para entender la distribución de las variables.
   * Visualizaciones (histogramas, gráficos de barras, etc.) para identificar patrones o tendencias (este apartado se omite por practicidad).
4. **Modelado Predictivo:**
   * Selección de un modelo de inteligencia artificial adecuado, como regresión logística, árboles de decisión, o redes neuronales, dependiendo de la naturaleza de los datos y el objetivo.
   * División de los datos en conjuntos de entrenamiento y prueba.
   * Entrenamiento del modelo con el conjunto de entrenamiento.
5. **Evaluación del Modelo:**
   * Evaluación del rendimiento del modelo usando métricas como precisión, recall, F1-score, o AUC-ROC, dependiendo del tipo de problema (clasificación o regresión).
   * Validación cruzada para asegurar la robustez del modelo.
6. **Interpretación:**
   * Análisis de los resultados para identificar factores clave que influyen en la lectura.

**matriz de valores faltantes**

****

### 8. ANÁLISIS DE RESULTADOS Y DISCUSIÓN

El uso de RTDA en la recomendación de libros permitirá un enfoque robusto y eficiente para la finalidad del proyecto, referente a la experiencia de lectura, incentivando el interés y la participación de más personas en la lectura voluntaria.

Al combinar datos demográficos y hábitos lectores, se logrará un mejor entendimiento de los factores que influyen en la construcción del hábito lector, facilitando el diseño de estrategias innovadoras para fomentar la lectura de manera efectiva y atractiva.

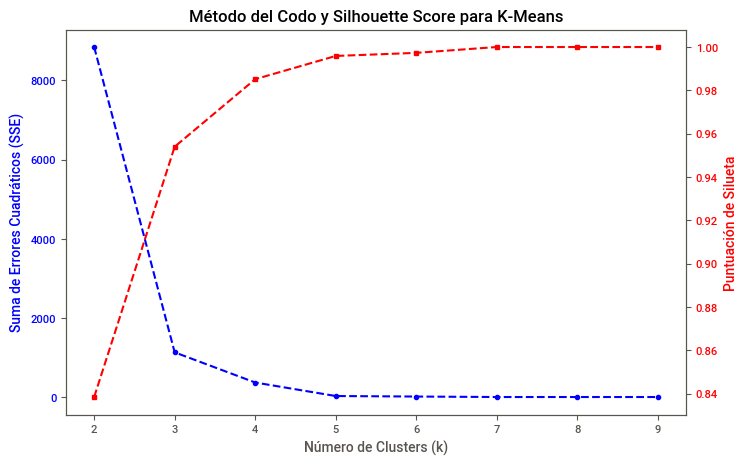
A pesar de los beneficios potenciales de los modelos de recomendación basados en inteligencia artificial para fomentar la lectura en adultos, es importante reconocer ciertas limitaciones inherentes a su implementación.

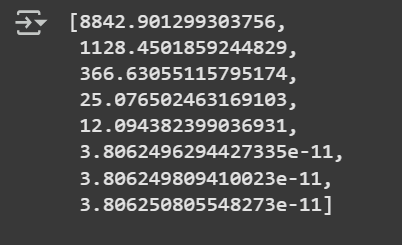
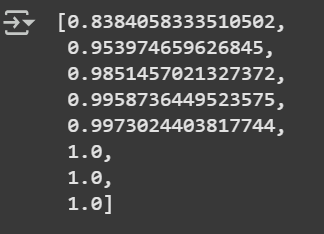
Dependencia de la calidad de los datos: La efectividad de los modelos de recomendación está intrínsecamente ligada a la calidad y precisión de los datos utilizados para entrenarlos. Datos incompletos, desactualizados o incorrectos pueden llevar a recomendaciones inexactas, lo que podría disminuir la confianza de los usuarios en el sistema y, en última instancia, afectar negativamente su experiencia de lectura.

Posibles sesgos en las recomendaciones: Los modelos de IA pueden heredar sesgos presentes en los datos de entrenamiento, lo que podría resultar en recomendaciones que no reflejan adecuadamente la diversidad de intereses y preferencias de los usuarios. Este sesgo puede limitar la exposición de los usuarios a una variedad más amplia de géneros y autores, restringiendo su experiencia de lectura y potencialmente perpetuando estereotipos o desigualdades existentes.

Abordar estas limitaciones requiere un enfoque cuidadoso en la recopilación y gestión de datos, así como en el diseño de algoritmos que sean transparentes y equitativos. Además, es crucial implementar mecanismos de retroalimentación que permitan a los usuarios influir en las recomendaciones, asegurando que el sistema evolucione y mejore continuamente en respuesta a sus necesidades y preferencias cambiantes.

Ahora se presentan los métodos o métricas consideradas para este modelo.





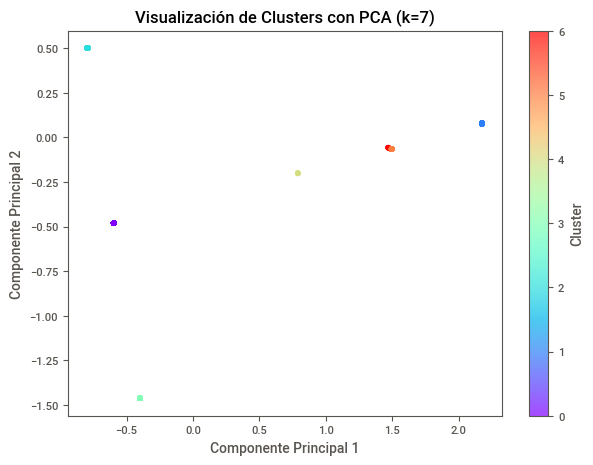
Esta gráfica muestra la evaluación del número óptimo de clusters (kkk) para el modelo **K-Means**, utilizando dos métricas clave:

* **Método del Codo (SSE - Suma de Errores Cuadráticos, en azul)**
  + La curva azul representa la SSE, que mide la compactación de los clusters.
  + A medida que kkk aumenta, la SSE disminuye porque los clusters son más pequeños y mejor definidos.
  + El punto donde la disminución se vuelve menos pronunciada es el **"codo"**, que sugiere un valor óptimo de kkk.
* **Coeficiente de Silhouette (en rojo)**
  + Esta métrica evalúa la separación entre clusters, con valores más altos indicando mejores agrupaciones.
  + Generalmente, buscamos el kkk donde esta puntuación es alta.

### Análisis para el modelo no supervisado:

* **El codo parece estar alrededor de k=3k = 3k=3 o k=4k = 4k=4**, lo que indica que dividir los datos en 3 o 4 grupos es una buena opción.
* **El coeficiente de Silhouette es más alto en k=3k = 3k=3 o k=4k = 4k=4,** lo que refuerza la idea de que estos valores de kkk generan clusters bien separados.
* Si se busca un equilibrio entre compacidad y separación, se **podría elegir k=3k = 3k=3 o k=4k = 4k=4 como el número óptimo de clusters** para el sistema de recomendación de libros.

Sin embargo, al hacer la evaluación, **se decidió usar k=7K = 7k=7 como se explica a continuación.**



Esta gráfica representa la **visualización de los clusters obtenidos con K-Means** utilizando **Análisis de Componentes Principales (PCA)** para reducir la dimensionalidad de los datos.

### Puntos clave a analizar:

1. **Número de Clusters (k=7)**
   * Se ha seleccionado k=7k=7k=7, lo que indica que los datos han sido divididos en 7 grupos.
   * Cada color representa un cluster diferente, como indica la barra de colores a la derecha.
2. **Distribución de los Clusters**
   * Los puntos están bastante separados, lo que puede sugerir que los clusters tienen buena diferenciación.
   * Sin embargo, algunos clusters parecen tener solo un punto, lo que podría indicar **sobre segmentación** (demasiados clusters).
3. **Evaluación de la Elección de kkk**
   * En la gráfica anterior (Método del Codo y Silhouette Score), k=7k=7k=7 no parecía el mejor valor, sin embargo, se ajusta a lo deseado.
   * Aunque, la dispersión y poca densidad de algunos clusters refuerza la idea de que **quizás un menor kkk, como 3 o 4, podría ser más adecuado**.

### Conclusión y próximos pasos:

* Si hay clusters con muy pocos elementos, se podría reducir kkk para mejorar la segmentación.
* Verificar si los grupos tienen sentido con respecto a las preferencias de los usuarios en la recomendación de libros.
* Si los clusters siguen estando dispersos, considerar probar otro algoritmo como DBSCAN o ajustar la métrica de distancia en K-Means.

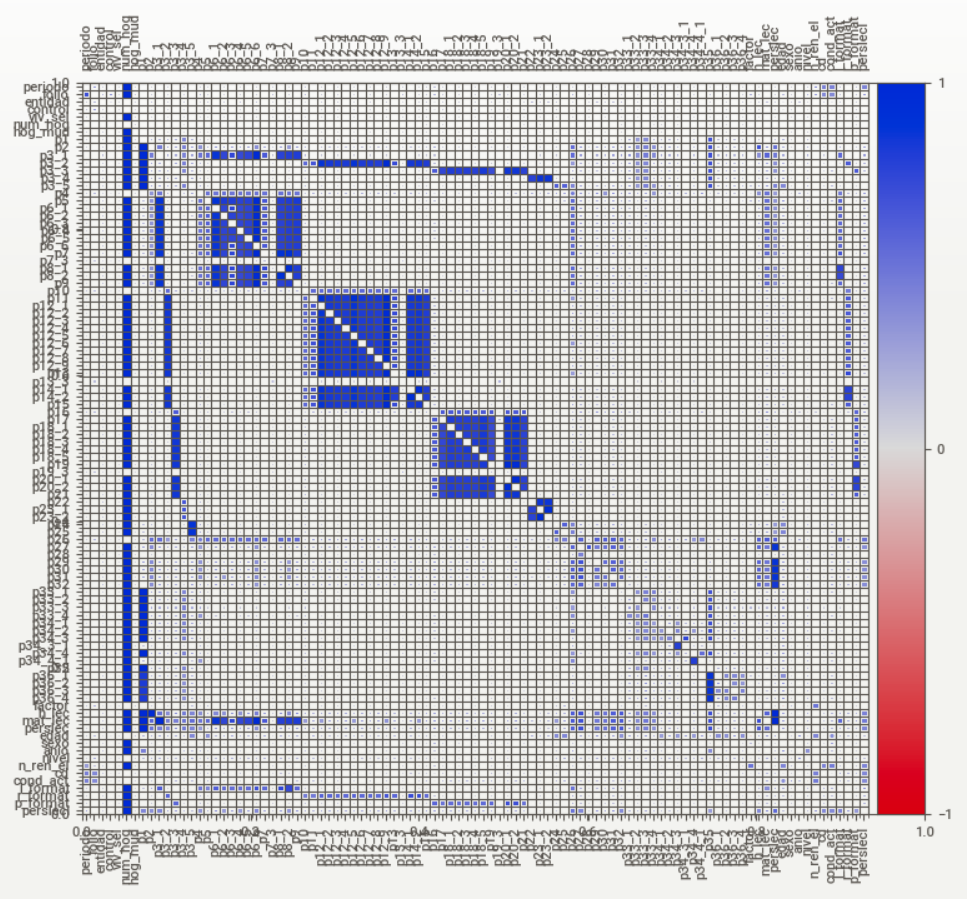
Ahora, con respecto al EDA depurado, se sabe que este informe detalla cómo interpretar el gráfico de asociaciones, proporcionando una guía clara para entender las relaciones entre las variables del conjunto de datos.

Este gráfico es útil para identificar relaciones significativas dentro del conjunto de datos.

El gráfico proporciona una representación visual clara de las asociaciones entre variables categóricas y numéricas en el conjunto de datos. Al interpretar estos elementos, se pueden identificar patrones y relaciones significativas que pueden informar decisiones y análisis posteriores.

En este caso, veremos que hay varias agrupaciones de variables con fuertes asociaciones, lo que podría indicar factores comunes o redundancia de información.

A continuación, la imagen de la matriz de correlaciones obtenida:



**Análisis de Asociaciones en el Conjunto de Datos**

En el presente informe, se analizan las asociaciones entre diferentes variables dentro de un conjunto de datos, utilizando un gráfico que representa tanto asociaciones categóricas como numéricas. A continuación, se detalla la interpretación de los elementos visuales del gráfico:

**Asociaciones Categóricas (Cuadrados):**

* **Descripción:** Los cuadrados en el gráfico representan asociaciones categóricas, evaluadas mediante el coeficiente de incertidumbre y la razón de correlación. Estos valores oscilan entre 0 y 1.
* **Coeficiente de Incertidumbre:** Este coeficiente es una medida asimétrica que indica la cantidad de información que los valores de las etiquetas de fila proporcionan a cada etiqueta en la parte superior del gráfico. Un valor más cercano a 1 sugiere una fuerte asociación, donde el conocimiento de una variable proporciona información significativa sobre la otra.
* **Interpretación:** Un cuadrado más oscuro o con un valor más alto indica una asociación más fuerte entre las variables categóricas correspondientes.

**Correlaciones Numéricas (Círculos):**

* **Descripción:** Los círculos representan correlaciones numéricas simétricas, calculadas mediante el coeficiente de correlación de Pearson. Estos valores varían de -1 a 1.
* **Coeficiente de Correlación de Pearson:** Este coeficiente mide la fuerza y la dirección de una relación lineal entre dos variables numéricas. Un valor de 1 indica una correlación positiva perfecta, -1 una correlación negativa perfecta, y 0 indica que no hay correlación lineal.
* **Interpretación:** Un círculo más grande o con un valor más extremo (cercano a -1 o 1) indica una correlación más fuerte. La dirección de la correlación (positiva o negativa) se debe considerar al interpretar estos valores.

**Diagonal Trivial:**

**Descripción:** La diagonal del gráfico se deja intencionalmente en blanco para mejorar la claridad visual. Esta diagonal representaría la correlación de cada variable consigo misma, que siempre es 1 en el caso de correlaciones numéricas, y no aporta información adicional relevante para el análisis de asociaciones entre diferentes variables.

### 9. IMPACTOS Y LÍNEAS FUTURAS

#### 9.1 Impactos

* **Impacto educativo:** Fomento de la lectura como un hábito natural, que a su vez potenciará habilidades como la creatividad, el pensamiento analítico y la comunicación efectiva.
* **Impacto social:** Fortalecimiento de la lectura como una herramienta para la construcción de identidad, el desarrollo emocional y social, teniendo como consecuencia la mejora en la calidad de vida de las personas.
* **Impacto cultural:** Promoción de la lectura como un pilar fundamental del crecimiento intelectual y la transmisión de valores.

#### 9.2 Líneas futuras

* **Recolección continua de datos:** Permite la actualización y mejora constante del modelo predictivo en su precisión.
* **Validación periódica del modelo:** Ajuste y optimización del sistema para adaptarse a los cambios en los hábitos y preferencias literarias.
* **Investigación adicional:** Incorporación de variables psicológicas como la motivación por la lectura, el entorno familiar, el nivel socioeconómico y la influencia de nuevas tecnologías en la lectura para una comprensión más profunda de los factores que afectan el hábito lector.

Este proyecto busca no solo aumentar el número de lectores, sino transformar la percepción de la lectura en un elemento esencial para el crecimiento personal y social, posicionándose como una experiencia enriquecedora y accesible para todas las personas.

Otros desarrollos futuros que se plantean son:

* **Generación de recomendaciones personalizadas basadas en intereses, nivel de lectura y comportamiento. (Ver código “piloto”)**
* Mejorar el modelo con más datos y técnicas avanzadas de IA.
* Expandir la plataforma a otros grupos de edad o idiomas.
* Realizar el modelo con repositorios de audiolibros.

**Plataforma de Implementación:**

* Aplicación web o móvil donde los usuarios puedan recibir y explorar sus recomendaciones.
* Posible integración con bibliotecas digitales o librerías para facilitar el acceso a los libros sugeridos.

**Evaluación y Mejora Continua:**

* Monitoreo del desempeño del sistema mediante métricas como tasa de interacción y satisfacción del usuario.
* Optimización constante del modelo con nuevos datos y retroalimentación.

**Impacto Esperado:**

* Aumento del interés por la lectura en adultos.
* Acceso equitativo a recomendaciones literarias personalizadas.
* Potencial colaboración con instituciones educativas y culturales.

**10. CONCLUSIONES**

Fomentar la lectura no es solo incrementar el tiempo que las personas dedican a leer, sino transformar su forma de interactuar con el conocimiento y el mundo. Este proyecto aspira a crear una cultura lectora que, lejos de ser una obligación académica, se convierta en una pasión y en una herramienta clave para el crecimiento personal, la innovación y la cohesión social. La lectura, al ser un medio para ampliar horizontes, se erige como un pilar fundamental para construir una sociedad más crítica, creativa y empática.

El hábito de lectura tiene un impacto profundo y multifacético en el PIB de un país. Desde el desarrollo del capital humano y la innovación hasta la reducción de costos en salud y el fortalecimiento de la industria editorial, la lectura es un motor clave para el crecimiento económico. Fomentar la lectura no solo beneficia a los individuos, sino que también contribuye al bienestar y la prosperidad de la sociedad en su conjunto.

La reducción de dimensionalidad con PCA mejoró la eficiencia del modelo al minimizar la redundancia de los datos y reducir la colinealidad entre variables, permitiendo que el algoritmo de clustering K-means identificara patrones más claros en los hábitos de lectura. Esto contribuyó a una segmentación más precisa de los usuarios y, por ende, a recomendaciones de libros más acertadas.

El presente proyecto ha demostrado que los modelos de recomendación basados en inteligencia artificial (IA) pueden constituir una herramienta altamente efectiva para fomentar la lectura en la población adulta. La capacidad de personalización inherente a estos sistemas es fundamental para captar y mantener el interés de los usuarios, adaptándose a sus preferencias individuales y hábitos de consumo de contenido.

El análisis de las razones para la falta de lectura revela que el 47.9% de los no lectores citan la falta de tiempo como el principal obstáculo, mientras que el 25.5% indica una falta de interés. Estos hallazgos subrayan la importancia de desarrollar soluciones que no solo optimicen la gestión del tiempo de los usuarios, sino que también despierten su interés a través de recomendaciones personalizadas que se alineen con sus gustos y necesidades específicas.

En cuanto a las preferencias de lectura, se ha identificado que los géneros más populares entre los lectores adultos son la literatura, incluyendo novelas y cuentos, así como los libros de texto o académicos. Este conocimiento permite a los modelos de recomendación priorizar estos géneros al sugerir nuevas lecturas, incrementando así la probabilidad de que los usuarios se comprometan con el material sugerido.

En conclusión, la implementación de modelos de recomendación basados en IA, con un enfoque en la personalización, tiene el potencial de transformar los hábitos de lectura en adultos, superando barreras como la falta de tiempo y el desinterés, y promoviendo una cultura de lectura más robusta y sostenida.

**11. REFERENCIAS**

Lectura en cada país  
<https://lectupedia.com/es/cantidad-de-libros-leidos-por-pais/>

DANE  
<https://www.dane.gov.co/index.php/estadisticas-por-tema/cultura/consumo-cultural>

Base de datos MOLEC México  
<https://www.inegi.org.mx/programas/molec/#datos_abiertos>

ENLEC 2017  
<https://sitios.dane.gov.co/enlec_dashboard/#!/>

Hábitos de lectura en Bogotá  
<https://www.canalcapital.gov.co/eventos-culturales/filbo-lectura-bogota-cifras>

Post Pandemia  
<https://elpais.com/america-colombia/2024-03-01/la-vida-pospandemia-los-colombianos-ahora-leen-mas-libros.html>

Libros más vendidos 2024-01  
<https://www.portafolio.co/tendencias/sociales/top-los-libros-mas-vendidos-en-colombia-en-el-primer-semestre-de-2024-610714>

Hábitos de lectura en Colombia  
chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://biblioteca.dane.gov.co/media/libros/LD_10753_EJ_5.PDF>

Para leer un país  
<https://www.mineducacion.gov.co/1621/article-122248.html>

Colombia dió un paso adelante en lectura impactando positivamente la calidad educativa  
<https://www.mineducacion.gov.co/portal/salaprensa/Comunicados/368450:Colombia-dio-un-paso-adelante-en-lectura-impactando-positivamente-la-calidad-educativa>

Santo Tomás   
<https://repository.usta.edu.co/items/55afd6ad-2db2-41d4-b30b-a7bc50d08863>

Datos.gov.co  
<https://www.datos.gov.co/Cultura/Vista-Bibloestaciones-de-Transmilenio-Bogot-D-C-A-/pzbk-7m7r/about_data>  
<https://datos.gov.co/browse?sortBy=newest&page=1&pageSize=20>

<https://datos.gov.co/>

Statista.com  
<https://www.statista.com/search/?q=lectura+en+colombia&p=1>

Infobae  
<https://www.infobae.com/colombia/2024/02/29/los-colombianos-estan-leyendo-mas-en-promedio-37-libros-al-ano/>

Portafolio  
<https://www.portafolio.co/tendencias/entretenimiento/aumento-el-indice-de-lectura-en-colombia-segun-camara-colombiana-del-libro-599499>

El espectador  
<https://www.elespectador.com/el-magazin-cultural/mejoraron-los-habitos-de-lectura-camara-colombiana-del-libro-noticias-hoy/>

Recopilaciones  
<https://www.kaggle.com/datasets/kevinespues/educacin-global-abandono-y-alfabetizacin>

Porcentaje de lectura en México  
<http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2145-94442016000100010>

Maestría sobre lectura y desarrollo humano  
<https://repositorio.flacsoandes.edu.ec/bitstream/10469/16169/1/TFLACSO-2019SFG.pdf>

**12. ANEXOS**

**Información adicional en las carpetas.**

****