***MODELO PARA LA IDENTIFICACIÓN DE PATRONES CRÍTICOS EN SALUD MENTAL, QUE PROMUEVAN PROGRAMAS DE PROMOCIÓN Y DE PREVENCIÓN CON ENFOQUE DE GÉNERO EN EL DEPARTAMENTO DE CALDAS.***

**Integrantes**

**Cristina Quintero Escobar**

**Silvia Juliana Macias Parra**

**Gedny Libeth Hernández Montoya**

**Ejecutores**

**Natalia Betancur Herrera**

**Frank Yesid Zapata Castaño**

**Margarita Maria Orozco**

**Universidad de Antioquia, Universidad de Caldas**

**Talento TECH**

**BOOTCAMP Inteligencia Artificial**

**Marzo de 2025**

**TABLA DE CONTENIDO**

1. **INTRODUCCIÓN**3
2. **PLANTEAMIENTO DEL PROBLEMA**4
3. **OBJETIVOS**
   1. **OBJETIVO GENERAL** 5

**3.2 OBJETIVOS ESPECÍFICOS** 6

**4. JUSTIFICACIÓN**………………………………………............………………………….7

**5. ALCANCE** ……………………………………………............….…………………..……8

**6. FLUJOGRAMA**……………………………………………............………………………9

**7. PRESUPUESTO**…………..........…………………...........………….…………....……11

**8. METODOLOGÍA**……………………..……………………...........……………………..12

**8.1 DESCRIPCIÓN BASE DE DATOS**…….....…………………...........……………12

**8.1.2 RESUMEN GENERAL DEL DATASET**……………..…...….....................….12

**8.2 ESTADÍSTICA DESCRIPTIVA** ........................................................................12

**8.2.1 VARIABLES CATEGÓRICAS Y HALLAZGOS CLAVE** ..............................12

**8.2.2 VARIABLES NUMÉRICAS** ..........................................................................13

**8.3 CORRELACIÓN Y RELACIÓN ENTRE VARIABLES** ....................................14

**9. ENTENDIMIENTO DE DATOS** ..............................................................................19

**9.1 DESCRIPCIÓN DE LAS VARIABLES** ............................................................19

**9.2 VALIDACIÓN INICIAL DE LOS DATOS** .........................................................20

**9.2.1 ANÁLISIS EXPLORATORIO** ........................................................................20

**9.2.3 MANEJO DE VALORES NULOS Y CODIFICACIÓN**...................................21

**9.2.3 HERRAMIENTAS PARA EDA** .....................................................................21

**10. MODELADO** ........................................................................................................22

**10.1 PREPARACIÓN DEL MODELO**....................................................................23

**10.1.1 VARIABLE PREDICTORA (Y)**....................................................................23

**10.1.2 INTERPRETACIÓN DEL EDA GENERADAS EN COLAB**........................24

**10.2 CONCLUSIONES**...........................................................................................27

**11. INTERPRETACIÓN GLOBAL Y POSIBLES ACCIONES**....................................27

**12. CONCLUSIONES GENERALES Y RECOMENDACIONES**.................................28

**13. BIBLOGRAFÍA**......................................................................................................30

**1. INTRODUCCIÓN.**

El consumo de sustancias psicoactivas (como alcohol, tabaco, drogas ilegales o medicamentos sin prescripción) representa una amenaza creciente para la salud física, mental y social de las personas. Estas sustancias alteran el funcionamiento del cerebro, generando dependencia, daños orgánicos y riesgos conductuales que afectan no solo al consumidor, sino también a su entorno.

La normalización de su uso en contextos sociales, la presión de grupo o la búsqueda de evadir problemas emocionales son factores que llevan a muchas personas a experimentar con ellas, sin considerar las consecuencias a largo plazo. Sin embargo, el consumo no es inocuo: puede desencadenar adicción, deteriorar la memoria, aumentar la ansiedad o el riesgo de accidentes, y hasta provocar daños irreversibles en órganos vitales.

Además, las sustancias psicoactivas no son solución a problemas como el estrés, la soledad o la falta de motivación. Por el contrario, su uso frecuente agrava estos desafíos y limita el desarrollo personal, las relaciones interpersonales y la capacidad de tomar decisiones saludables.

Es fundamental romper el estigma asociado al consumo y fomentar un enfoque preventivo basado en la educación, el acceso a apoyo profesional y la promoción de hábitos saludables. Reconocer los riesgos y buscar ayuda es el primer paso hacia una vida más segura y plena.

La creación de políticas públicas efectivas para la equidad de género en el acceso a programas de atención, promoción y prevención, requiere de un análisis profundo de datos que evidencien las necesidades reales de la población femenina. Actualmente, los tomadores de decisión enfrentan dificultades para diseñar estrategias eficientes debido a la dispersión y falta de integración de la información sobre las mujeres en Caldas. Además, la toma de decisiones basada en intuiciones o aproximaciones limitadas puede generar programas con impacto reducido o ineficiente.

La evolución de la inteligencia artificial actualmente permite analizar grandes volúmenes de datos de manera automatizada, identificando patrones, tendencias y áreas prioritarias de intervención para la ejecución de políticas públicas. En el presente trabajo se propone un modelo no supervisado que establece patrones para la identificación de programas de promoción y prevención con enfoque de género para las mujeres de Caldas con alto riesgo por consumo de sustancias psicoactivas. Este enfoque reduce la subjetividad en la formulación de programas y contribuye a un uso más eficiente de los recursos destinados a la equidad de género.

Este trabajo no solo contribuirá al fortalecimiento de la equidad de género en el departamento, también servirá como modelo replicable en otras regiones del país, promoviendo la innovación social y el uso de la tecnología como herramienta clave para la transformación y el desarrollo humano con enfoque de género.

**2. PLANTEAMIENTO DEL PROBLEMA.**

Uno de los estudios más recientes sobre el uso y el abuso de sustancias psicoactivas fue realizado por la Secretaría Distrital de Salud y la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC), donde se evidencia que el consumo de sustancias psicoactivas lícitas e ilícitas en mujeres ha incrementado en todas las prevalencias[[1]](#footnote-0). Se identificó que este aumento está en consonancia con la tendencia internacional de cierre de la brecha de consumo por género, lo que ha merecido que las mujeres desarrollen por lo menos un trastorno por el uso de drogas lícitas o ilícitas con mayor rapidez que los hombres. Se calcula que las mujeres representan actualmente entre el 45% y el 49% de quienes consumen anfetaminas y de estimulantes farmacéuticos, opiáceos farmacéuticos, sedantes y tranquilizantes[[2]](#footnote-1).

La problemática se agudiza en tanto que la brecha en el acceso a tratamiento sigue siendo grande para las mujeres en todo el mundo. Aunque representan casi uno de cada dos consumidores de anfetaminas, sólo constituyen una de cada cinco personas en tratamiento por trastornos por su consumo[[3]](#footnote-2).

El consumo de sustancias psicoactivas se presenta como una solución a problemas como el estrés, la soledad o la falta de motivación para mantener un proyecto de vida, por el contrario, su uso frecuente agrava estos desafíos y limita el desarrollo individual, las relaciones interpersonales y la capacidad de tomar decisiones saludables.

Siendo las mujeres las depositarias de la asignación de roles de género que han sido y siguen siendo el resultado de interacciones complejas entre tradiciones coloniales, estructuras patriarcales y dinámicas culturales, las cuales han perpetuado desigualdades y estereotipos de género que vinculan lo femenino con la domesticidad, la maternidad y la subordinación[[4]](#footnote-3).

Este legado persiste especialmente en zonas donde el acceso a la educación y a programas de prevención y promoción es menor, y donde las expectativas culturales que idealizan a la mujer como "ángel del hogar" o "madre abnegada", han limitado su participación en ámbitos como la economía, la gobernanza y el acceso a tratamientos como aquellos que están ligados con el uso de sustancias psicoactivas lícitas y ilícitas.

En el presente documento, las autoras consideran que las creaciones simbólicas, culturalmente arraigadas y las creaciones sociales moldeadas por sistemas de poder y representaciones sociales que vinculan lo femenino con la domesticidad, la maternidad y la subordinación, tienen un doble efecto, ya que se convierten en un factor propulsor del consumo y a vez, una barrera para el acceso a atenciones especializadas, asociadas a trastornos por el uso de drogas lícitas o ilícitas, a la promoción de estilos de vida saludable y la prevención de uso.

**3. OBJETIVOS.**

**3.1 OBJETIVO GENERAL.**

Desarrollar un modelo no supervisado, con inteligencia artificial, que le permita a los actores sociales y territoriales del departamento de Caldas, identificar patrones críticos en salud mental para la formulación y puesta en marcha de programas de promoción y prevención en salud, con enfoque de género.

**3.2 OBJETIVOS ESPECÍFICOS.**

* Identificar los patrones críticos en salud mental que permitan tener un panorama de las condiciones de salud de las mujeres en Caldas.
* Proponer estrategias a mediano y largo plazo que mejoren las condiciones de vida de las mujeres que responda a la realidad encontrada.

**4. JUSTIFICACIÓN.**

La tecnología tiene la capacidad de permear aspectos importantes en nuestra vida cotidiana, transformando la manera en que nos relacionamos, comunicamos, aprendemos y socializamos, potenciando el desarrollo humano desde sus diversas áreas, lo que facilita el alcance a nivel local y global de los objetivos del desarrollo sostenible, especialmente el ODS 3 el cual promueve la salud y bienestar[[5]](#footnote-4).

De manera particular, las metas 3.4 y 3.5 de los ODS se enfocan en los siguientes aspectos:

* + Reducir la mortalidad por enfermedades no transmisibles (como diabetes o cáncer) y promover la salud mental y el bienestar.
  + Incluye acciones para prevenir trastornos mentales y mejorar el acceso a servicios de salud mental.
  + Prevenir y tratar el abuso de sustancias adictivas (alcohol, drogas), vinculado a problemas de salud mental como la depresión o ansiedad.

Otro avance que proporciona la tecnología y la inteligencia artificial es el acceso, uso, visualización e interpretación de datos. A través de gráficos y representaciones visuales, se pueden identificar patrones, tendencias y anomalías en los datos, que no podrían ser evidentes de otra manera. Además, la aplicación de herramientas tecnológicas en la planificación de políticas públicas es un avance crucial hacia la modernización del Estado y la mejora de la gobernanza.

La inteligencia artificial permitirá generar escenarios prospectivos, anticipar problemáticas y recomendar acciones específicas que respondan a la realidad de las mujeres en Caldas. Con ello, se fortalecerá la implementación del sistema de cuidado, facilitando el acceso a servicios de salud, educación, empleo y protección para las mujeres, en especial aquellas en situación de vulnerabilidad.

**5. ALCANCE.**

El alcance del modelo permitirá la identificación de patrones críticos que incidirán en la toma de decisiones para la implementación de programas de atención, promoción y prevención en salud, con enfoque de género y enfoque territorial.

En Colombia, el Ministerio de Salud y Protección Social promueve la Política Nacional de Salud Mental, a partir de la resolución 4886 del 2018 que debe ser aplicada en todo el territorio colombiano. Dicha política parte de la formulación de un perfil epidemiológico y una encuesta nacional de salud mental. El análisis de datos permitió identificar que una de las principales problemáticas en salud mental es el consumo de sustancias psicoactivas y la falta de acceso a programas que atiendan y promuevan la salud mental. Además, dicha Política insta a los prestadores de servicios de salud el desarrollo de programas de respuesta de primer y segundo nivel de atención basados en los siguientes enfoques: Desarrollo; Curso de Vida; Género; Territorial y Psicosocial[[6]](#footnote-5).

Por lo anteriormente expuesto, *el modelo no supervisado para la identificación de patrones críticos en salud mental,* fue creado con base en inteligencia artificial y aprendizaje automático, utilizando el lenguaje de programación Python y herramientas avanzadas de análisis de datos, que permitieron identificar patrones, tendencias y necesidades prioritarias dentro de la población femenina del departamento. El análisis cuantitativo se basó en 19.530 Registros Individuales de Prestación de Servicios de Salud (RIPS); éstos corresponden a un conjunto de datos mínimos y básicos que el Sistema General de Seguridad Social en Salud requiere para los procesos de dirección, regulación y control.

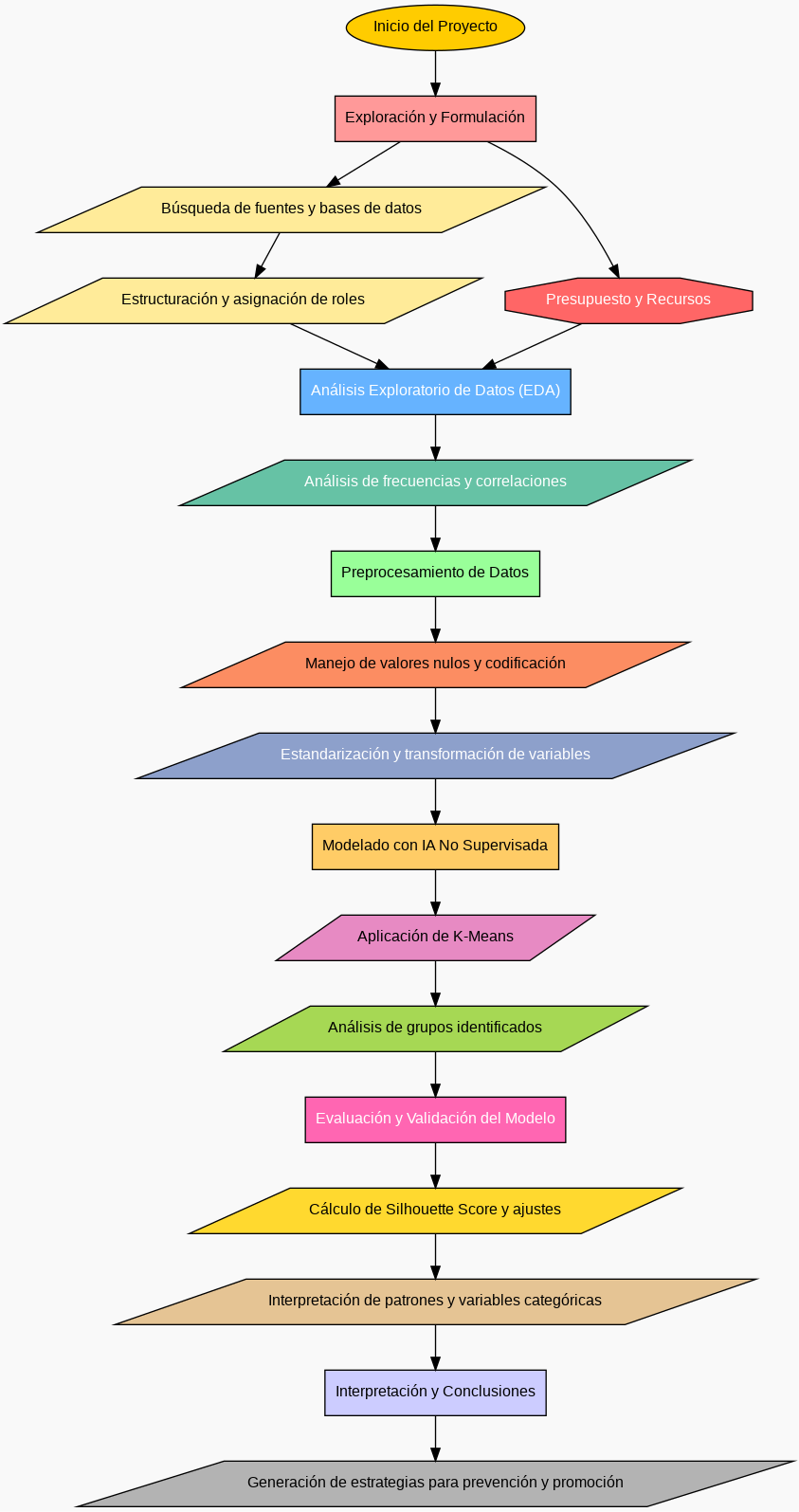
Con la información recolectada en los RIPS, el Ministerio de Salud y otras entidades pueden diseñar estrategias, asignar recursos y fortalecer programas de prevención y tratamiento. Además, sirven para evaluar la efectividad de las intervenciones en salud mental, como la atención psicosocial en hospitales, el acceso a servicios y la garantía de calidad y mejora de la atención.

De esta manera, el modelo no supervisado facilitará y permitirá a todos los actores involucrados en el sistema de salud, enfocar sus acciones, programas y recursos de forma prioritaria en aquellos municipios donde la problemática de salud mental requiera mejorar la toma de decisiones en la formulación e implementación de políticas públicas con perspectiva de género y por lo tanto la disminución de la brecha de atención y el estigma relacionado con los estereotipos de género. Además, los datos de entrada que se requieren para este tipo de análisis, pertenecen a datos que ya están incorporados en el sistema de prestación de servicios de salud, por lo que no se requiere implementar softwares o programas adicionales para alimentar el modelo.

Finalmente, los resultados y hallazgos del modelo no supervisado serán socializados con instituciones públicas, organizaciones sociales y académicas, fomentando la sensibilización y el desarrollo de políticas basadas en evidencia. El proyecto busca ser un modelo replicable y escalable, con posibilidad de aplicarse en otros departamentos y contextos nacionales para promover la equidad de género y el desarrollo sostenible.

**6. FLUJOGRAMA.**

El flujograma representa el flujo de trabajo del proyecto, desde la exploración y análisis de datos hasta la interpretación de resultados. Muestra de manera estructurada cómo procesamos el proyecto y las variables categóricas, se aplicó K-Means para identificar patrones y se integraron los recursos del presupuesto en la toma de decisiones.

****

**7. PRESUPUESTO.**

El modelo no supervisado fue creado por tres mujeres, las cuales se unieron en este proyecto con la intención de generar propuestas de apoyo, cambio y procesos de empoderamiento de las mujeres que habitan en los diferentes municipios del departamento de Caldas. A continuación, se detalla el presupuesto.

| **PRESUPUESTO** | | | |
| --- | --- | --- | --- |
| **RECURSOS HUMANOS** | **JUSTIFICACIÓN** | **VALOR UNITARIO /3** | **VALOR TOTAL** |
| **Profesional Psicóloga Magister**  **Profesional Psicóloga Especialista**  **Profesional en Ingeniería Comercial** | Etapa exploratoria, realización de lluvia de ideas en equipo, busqueda de fuentes de información y bases de datos | 1.000.000 | 3.000.000 |
|
|
| Etapa de formulación de proyecto, estructuración y construcción de ruta y asignación de roles | 1.500.000 | 4.500.000 |
| Etapa de programación, conceptualización y presentación del mismo |  |  |
| **Ingeniero de sistemas magister en programación e IA** | Experto apoyo en programación, para la construcción del modelo | 2.500.000 | 2.500.000 |
|  |  | **VALOR PARCIAL** | **10.000.000** |
| **RECURSOS FÍSICOS** | **JUSTIFICACIÓN** | **VALOR UNITARIO** | **VALOR TOTAL** |
| **ALQUILER DE EQUIPOS** | para socializar el proyecto en varias entidades | 500.000 | 500.000 |
| **MATERIAL IMPRESO** | Socialización del proyecto y memoria del mismo | 100.000 | 100.000 |
|  |  |  |  |
|  |  | **VALOR PARCIAL** | **600.000** |
|  |  | **VALOR TOTAL PROYECTO** | **10.600.000** |

**8. METODOLOGÍA.**

**8.1 Descripción base de datos.**

## 8.1.2 Resumen General del Dataset.

El análisis exploratorio de datos (EDA, por sus siglas en inglés) proporciona un análisis descriptivo de las variables más relevantes en la base de datos. Los aspectos principales incluyen:

* Número total de registros: Se muestra el tamaño de la base de datos, lo que permite evaluar su representatividad.
* Número de variables: Se han seleccionado las características más relevantes para el estudio.
* Distribución de valores faltantes: Se analizan las columnas con datos ausentes y su impacto en el análisis.
* Tipos de datos: Se identifican variables categóricas y numéricas, cruciales para aplicar modelos no supervisados.

**8.2 Estadísticas Descriptivas.**

**8.2.1 Variables Categóricas y hallazgos clave.**

Son un tipo de variables que toman un número limitado y discreto de valores, que representan categorías o grupos. Para efectos del presente trabajo, las variables categóricas utilizadas y caracterizadas son: Trastornos por consumo de sustancias psicoactivas; trastornos del humor.

El análisis explora la frecuencia y distribución de las variables:

* Género
* Ubicación (municipios de Caldas)
* Año de reporte
* Categoría Principal (trastornos relacionados con el consumo de sustancias, trastornos del estado de ánimo, etc.)
* Código de diagnóstico (según el CIE-10)

Frente a estas variables, se puede dar cuenta de los siguientes hallazgos:

* Hay más registros de mujeres que de hombres en la base de datos.
* Ciertos municipios tienen mayor concentración de casos de alto riesgo.
* Algunos trastornos específicos tienen mayor presencia en mujeres en comparación con hombres.

**8.2.2** **Variables Numéricas.**

Estas variables realizan los respectivos análisis cuantitativos que, para el caso del modelo, arrojaron las tendencias sobre los principales diagnósticos consignados en los Registros Individuales de Prestación de Servicios de Salud (RIPS). Dichos diagnósticos corresponden a un conjunto de datos mínimos y básicos que el Sistema General de Seguridad Social en Salud requiere para los procesos de dirección, regulación, seguimiento y control.

De igual forma se consultó la información estadística y de acceso público que presenta la Dirección Territorial de Salud en el Departamento. Dicha información permite identificar datos relacionados con la situación en salud en Caldas; el informe en mención se conoce como el documento ASIS y, para efectos del presente proyecto, la información fue consultada para cuantificar y cualificar el estado de la salud de las mujeres en Caldas. Se identifica que la información básica que contine el informe ASIS es alimentada en gran parte por los RIPS que se le realiza a cada paciente que vivita su respectiva IPS. Para el caso de la identificación de las características de la salud mental, se tuvieron en cuenta los ASIS de los años 2018, 2019 y 2020.

Se examinaron medidas como:

* Máximos y mínimos de valores en cada categoría.
* Media y mediana de casos por ubicación y categoría.
* Desviación estándar para entender la variabilidad en los datos.

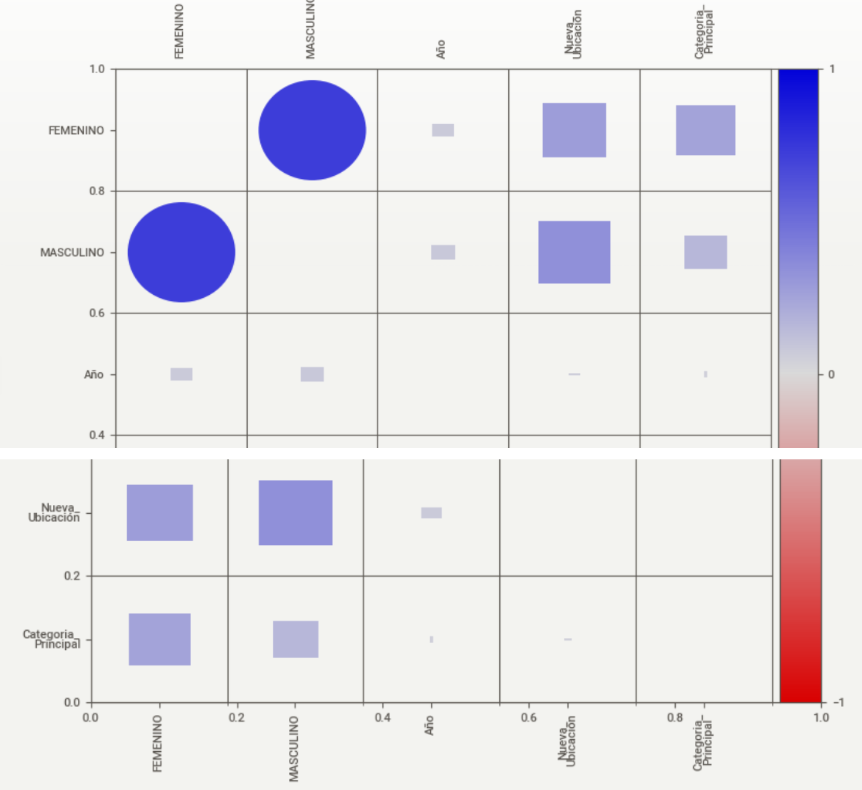
Hallazgos Clave:

* Hay una fuerte desigualdad en la distribución de casos por género y ubicación.
* Se identifican valores atípicos en ciertos diagnósticos, lo que sugiere posibles patrones inusuales en la data.

## 8.3 Correlaciones y relaciones entre Variables

Las variables en la preparación del modelo son la base de la predicción. Definir correctamente qué variables se usarán y cómo se procesarán (limpieza, escalado, codificación) influye directamente en la precisión del modelo.

* Se presentan matrices de correlación para identificar asociaciones entre características.
* Se utilizan mapas de calor y gráficos de dispersió**n** para visualizar cómo interactúan ciertas variables.



Este gráfico de correlación visualiza las asociaciones entre diferentes variables. Se representa con un mapa de calor con círculos de diferente tamaño y color, donde:

* Círculos más grandes y oscuros indican correlaciones más fuertes.
* Círculos más pequeños y claros representan correlaciones más débiles.

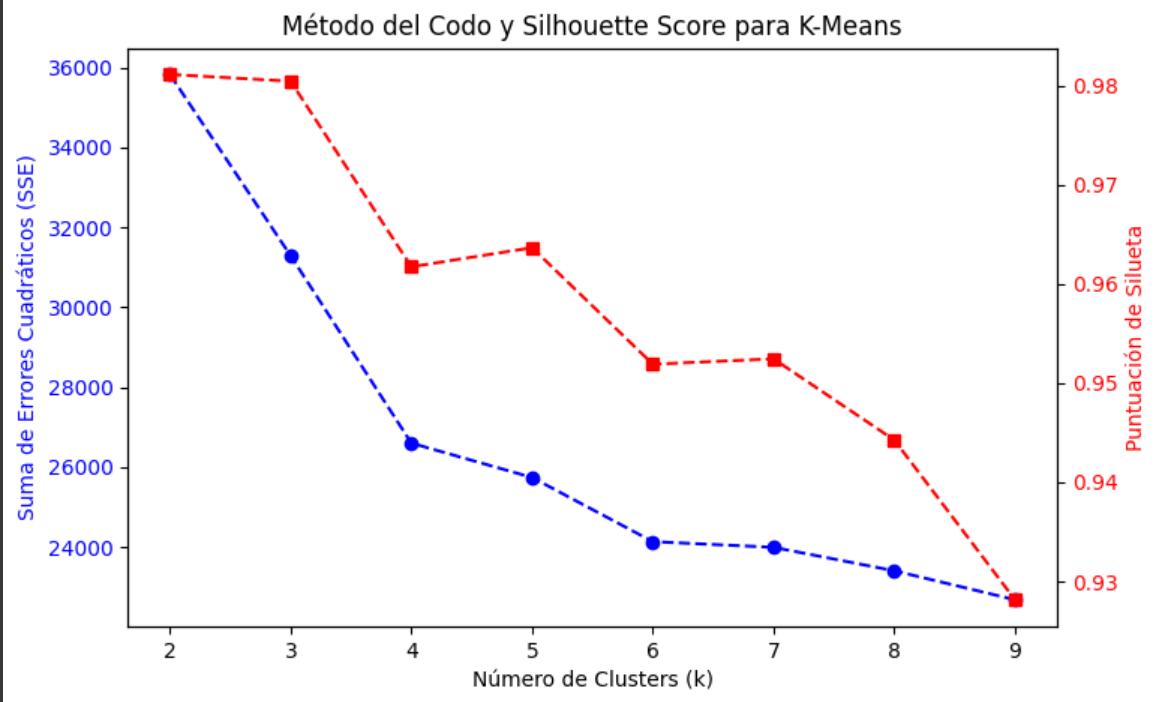
### Observaciones clave:

La categoría "FEMENINO" y "MASCULINO" tienen una correlación fuerte consigo mismas, lo cual es esperado.

Hay asociaciones más débiles entre el género y otras variables como "Año", "Nueva Ubicación" y "Categoría Principal".

Frente a las interacciones con la categoría "Nueva Ubicación" y "Categoría Principal", se observa que algunas ubicaciones y categorías principales tienen cierta asociación con el género, lo que indica que el consumo puede estar influenciado por la localización y el tipo de diagnóstico.

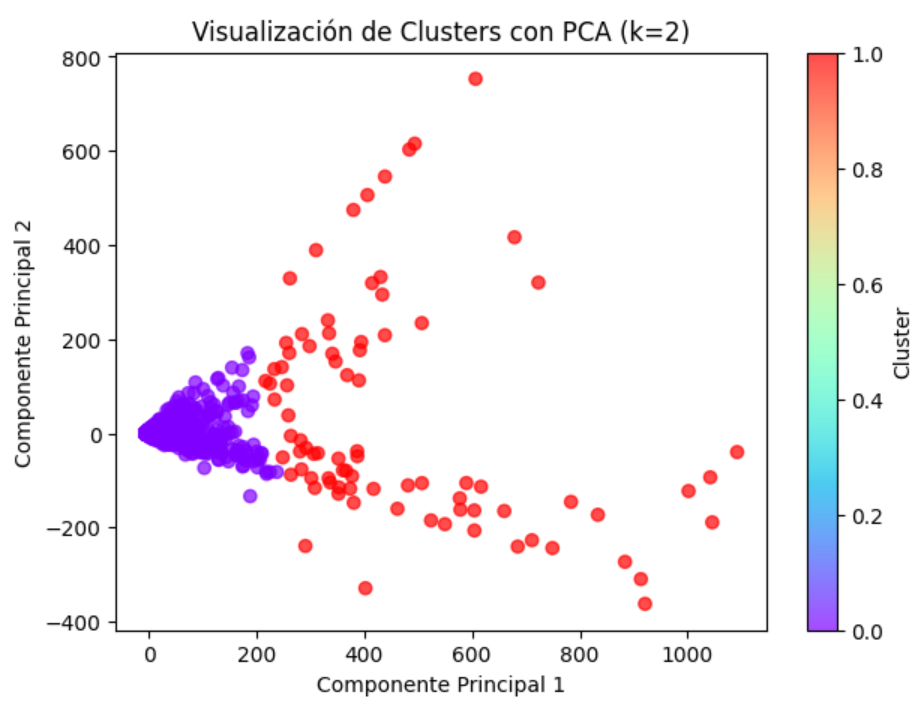
En conclusión, la correlación más fuerte está en la diferencia de género y su asociación con ciertas categorías diagnósticas y ubicaciones, lo cual podría indicar que ciertos lugares o trastornos tienen mayor prevalencia en mujeres o en hombres. Existen asociaciones fuertes entre ciertas categorías diagnósticas y el género. La ubicación geográfica influye en la cantidad de casos reportados.



Este gráfico combina dos métricas clave para determinar el número óptimo de clusters (k) en el modelo de K-Means.

* Línea azul (puntos circulares): Representa la Suma de Errores Cuadráticos (SSE) en el eje izquierdo (azul).
  + A medida que aumentamos el número de clusters (k), el SSE disminuye, lo que significa que los datos se ajustan mejor dentro de sus respectivos clusters.
  + Sin embargo, después de cierto punto, la reducción es menos significativa, lo que indica el punto óptimo de k.
* Línea roja (cuadrados): Representa el Silhouette Score en el eje derecho (rojo).
  + Esta métrica mide la calidad de la separación entre clusters (valores más altos indican mejores agrupaciones).
  + Se observa que al aumentar k, el Silhouette Score tiende a disminuir después de cierto punto.
* Punto de inflexión en SSE
  + El punto donde la curva azul deja de descender abruptamente marca el número óptimo de clusters.
  + En este gráfico, parece estar alrededor de k = 4 o k = 5, ya que después la reducción de SSE es mínima.
* Máximo del Silhouette Score
  + Indica cuán bien agrupados están los datos.
  + Aquí, la puntuación es más alta para valores entre k = 4 y k = 5, lo que sugiere que es un buen número de clusters para nuestro modelo.

En conclusión, con k = 4 o 5, se obtiene un equilibrio entre precisión y separación de los clusters. Más clusters no siempre es mejor, ya que el Silhouette Score baja, lo que indica que los grupos se vuelven menos diferenciables. Esta elección nos permite segmentar a las mujeres en alto riesgo de consumo de sustancias psicoactivas en grupos más representativos para diseñar estrategias de prevención personalizadas.



Este gráfico muestra la segmentación de datos en dos clusters utilizando el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad.

* Cada punto es un registro en la base de datos (una mujer con características específicas).
* Los colores representan dos clusters identificados por K-Means (k=2):
  + Cluster 0 (morado/azul): Grupo con características similares en una región más compacta.
  + Cluster 1 (rojo): Grupo más disperso con características más variables.
* Los ejes Componente Principal 1 y 2 representan la transformación de los datos a un espacio bidimensional con PCA, que conserva la mayor cantidad de información posible.
* El Cluster 0 (morado) tiene una alta concentración de puntos en un área específica, lo que sugiere que este grupo comparte patrones similares de riesgo o comportamiento.
* El Cluster 1 (rojo) está más disperso, lo que indica mayor variabilidad en los datos, posiblemente con características más diversas o menos definidas.
* PCA reduce la dimensionalidad de los datos manteniendo la mayor cantidad de información posible.
* Permite visualizar mejor la separación de clusters en un espacio de 2 dimensiones, ya que los datos originales pueden tener muchas variables.

En conclusión, se lograron dos grupos distintos, lo que indica que las mujeres en riesgo pueden segmentarse en al menos dos perfiles diferenciados. El cluster 0 representa un grupo homogéneo, mientras que Cluster 1 muestra mayor dispersión. Este análisis ayuda a diseñar estrategias personalizadas de prevención y promoción de salud mental.

**9. Entendimiento de los Datos.**

La fase de entendimiento de los datos se enfoca en la recopilación, exploración y validación de la información necesaria para alimentar el modelo no supervisado. Con base en los Registros Individuales de Prestación de Servicios de Salud (RIPS) en el departamento de Caldas, a continuación, se describe el contenido y propósito de cada grupo de variables:

**9.1 Descripción de las Variables.**

| **VARIABLE** | **DESCRIPCIÓN DE LA VARIABLE** | **IDENTIFICADOR** |
| --- | --- | --- |
| **ATENCIÓN** | Aquí se encuentran descritos los diagnósticos diferenciales con base en la codificación del CIE-10 | 0 |
| **FEMENINO** | Corresponde a la identidad de género que se establece en la atención | 1 |
| **MASCULINO** | Corresponde a la identidad de género que se establece en la atención | 1 |
| **AÑO** | Describe el año de ocurrencia del evento | 2 |
| **NUEVA UBICACIÓN** | Corresponde al municipio en donde se generó la atención. | 1 |
| **CATEGORÍA PRINCIPAL** | Corresponde a la frecuencia de presentación del evento, bajo el diagnóstico diferencial del CIE-10 | 1 |
| **SUBCATEGORÍA** | Es una definición detallada del diagnóstico y el trastorno. | 1 |

La categorización se hizo tomando en cuenta lo siguiente:

| 0 | No se tiene en cuenta |
| --- | --- |
| 1 | No ordinal |
| 2 | Ordinal |

**9.2 Validación Inicial de los Datos.**

**9.2.1 Análisis exploratorio.**

La visualización de correlaciones entre variables mediante gráficos, para el desarrollo del modelo, se hizo mediante un **Análisis Exploratorio de Datos (EDA),** el cual es un proceso inicial que se realiza al trabajar con un conjunto de datos, donde se explora su estructura, se identifican patrones, anomalías, relaciones entre variables y otros aspectos clave. El objetivo del EDA es entender mejor los datos antes de aplicar modelos o realizar análisis más complejos, asegurando así una base sólida para los siguientes pasos en el análisis.

Este análisis sirve para: conocer la estructura de los datos (revisando el tamaño, el tipo de variables (numéricas, categóricas), y las características generales); identificar patrones y relaciones (explorando correlaciones entre variables, tendencias o comportamientos específicos en los datos); detectar valores atípicos o errores (localizar valores que no tienen sentido o errores de captura de datos que puedan afectar los resultados); evaluar la distribución de las variables (observando si los datos están sesgados, presentan valores extremos, o siguen una distribución normal) y hacer la preparación para la limpieza de datos (podemos identificar valores nulos, datos duplicados o inconsistencias, para luego limpiarlos y obtener un conjunto de datos más manejable).

**9.2.2 Manejo de valores nulos y codificación**

En la limpieza de la base de datos se decidió no contar con los valores nulos ya que estos pueden generar resultados inconsistentes y afectar el proceso de normalización, coherencia y eficiencia en análisis.

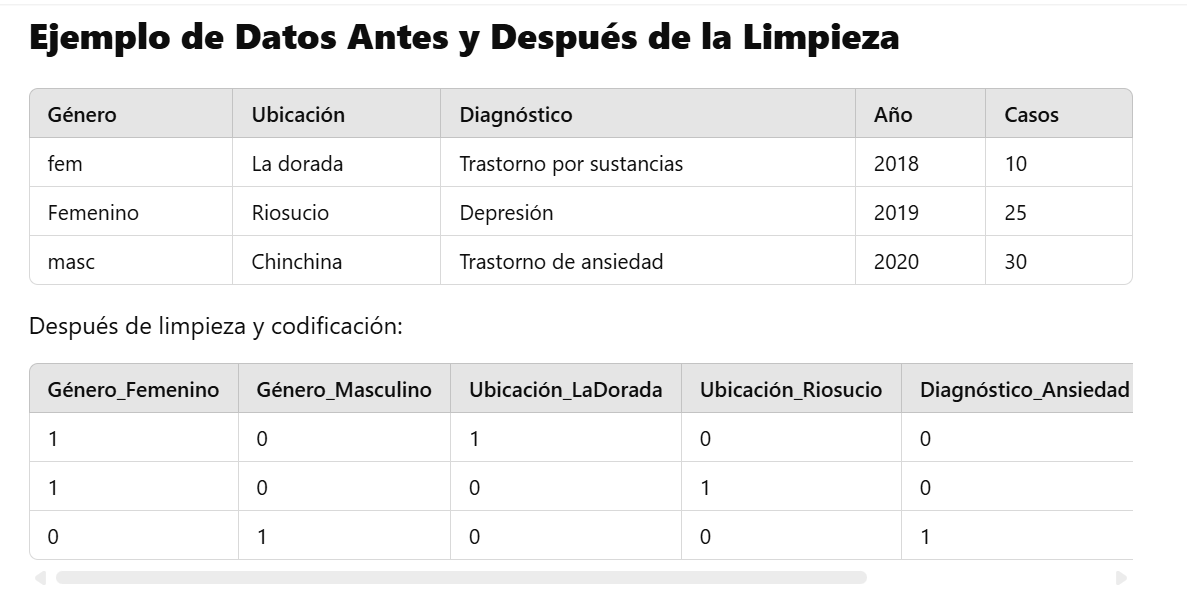
**9.2.3 Herramientas para EDA**

En programación, librerías como **Pandas, Matplotlib y Seaborn** en Python son ampliamente utilizadas para realizar un EDA de manera rápida y visual. También existen librerías específicas, como Sweetviz, que generan informes automáticos de EDA para conjuntos de datos. Se utilizó **nump**y para realizar cálculos numéricos y análisis de datos, además se utilizó **missingno** que permite ver de una forma visual, donde se encuentran los valores nulos en un conjunto de datos.

El análisis de cada una de las gráficas obtenidas y su correlacion se puede observar en el apartado de: Interpretación del EDA Generadas en Colab

* **Limpieza de datos**:

Se realizó al inicio, mediante la organización de la información y acorde a la disponibilidad de la misma, teniendo en cuenta que algunas columnas tenían información incompleta, lo cual no genera una métrica. Por tanto, se tomó la decisión de eliminar el 20% de la información inicial, la cual estaba ubicada en dos columnas de la base de datos inicial, para garantizar la calidad del análisis o el modelado.



**10. MODELADO.**

Para este proyecto se seleccionó el modelo **K-means,** éstees un algoritmo no supervisado de clustering. Se utiliza cuando existe una gran cantidad de datos sin etiquetar y tiene como objetivo dividir un segmento de datos entre un número de grupos (llamados clusters) basados en su similitud.

El Análisis de Componentes Principales (PCA, por *Principal Component Analysis*) es una técnica de reducción de dimensionalidad no supervisada. El PCA es muy eficaz para visualizar y explorar conjuntos de datos de alta dimensión, o datos con muchas características, ya que puede identificar fácilmente tendencias, patrones o valores atípicos. El PCA se utiliza comúnmente para el preprocesamiento de datos para su uso con algoritmos de machine learning.

De manera general, el modelo K-Means comienza seleccionando aleatoriamente puntos como centroides iniciales. Estos centroides son los centros de los clusters. El algoritmo asigna cada punto de datos al cluster cuyo centroide es más cercano (usualmente basado en la distancia euclidiana). Después de asignar todos los puntos a un cluster, se recalcula la posición de los centroides como el promedio de todos los puntos en ese cluster.El algoritmo repite los pasos de asignación y recálculo de centroides hasta que los centroides ya no cambian significativamente o se alcanza un número máximo de iteraciones. Al final se puede visualizar los clusters y los centroides para entender cómo se han agrupado los datos. Para concluir, K-Means es sensible a la escala de los datos, por lo que es recomendable escalar las características antes de aplicar el algoritmo.

**10.1 Preparación del Modelo**

En Machine Learning (ML), las variables juegan un papel fundamental en la preparación del modelo, ya que representan los datos de entrada y salida que alimentarán el algoritmo.

Tipos de variables en la preparación del modelo

1. Variables de entrada (features o características) Representan los atributos que el modelo usará para hacer predicciones.
2. Variables auxiliaresNo se usan directamente en el modelo, pero ayudan en la limpieza y transformación de datos. Variables de preprocesamiento, para convertir variables categóricas a numéricas.

**10.1.1 Variable Predictora (Y)**

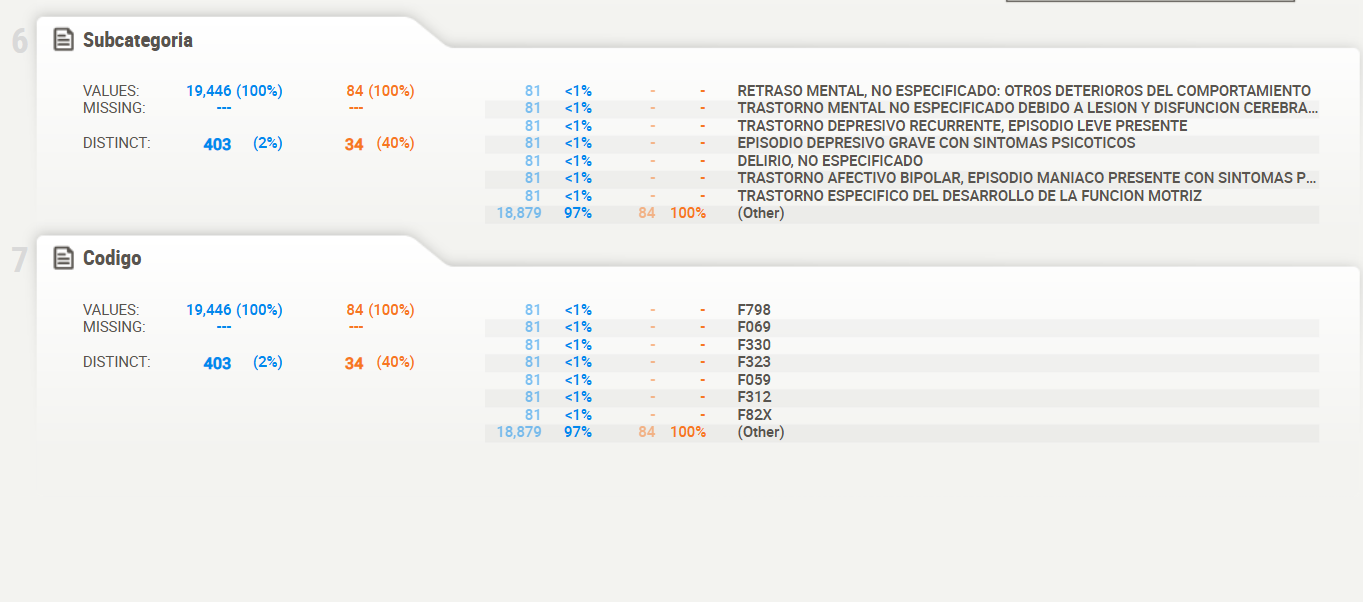
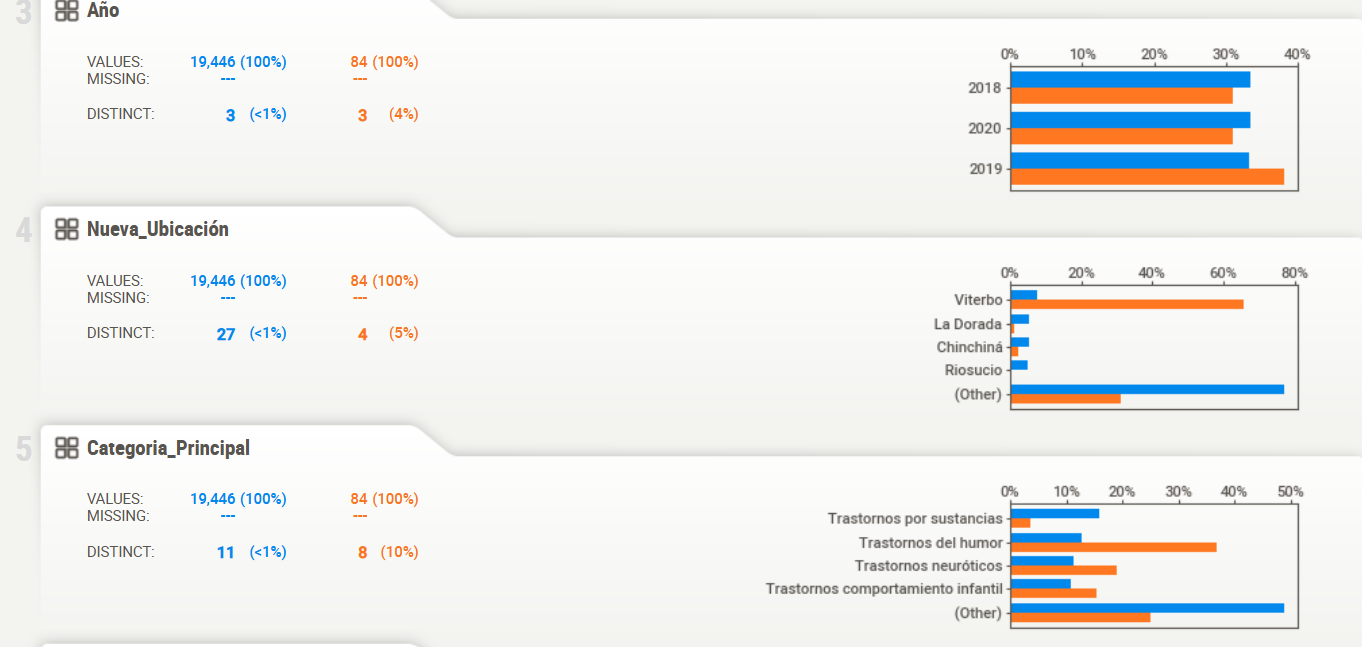
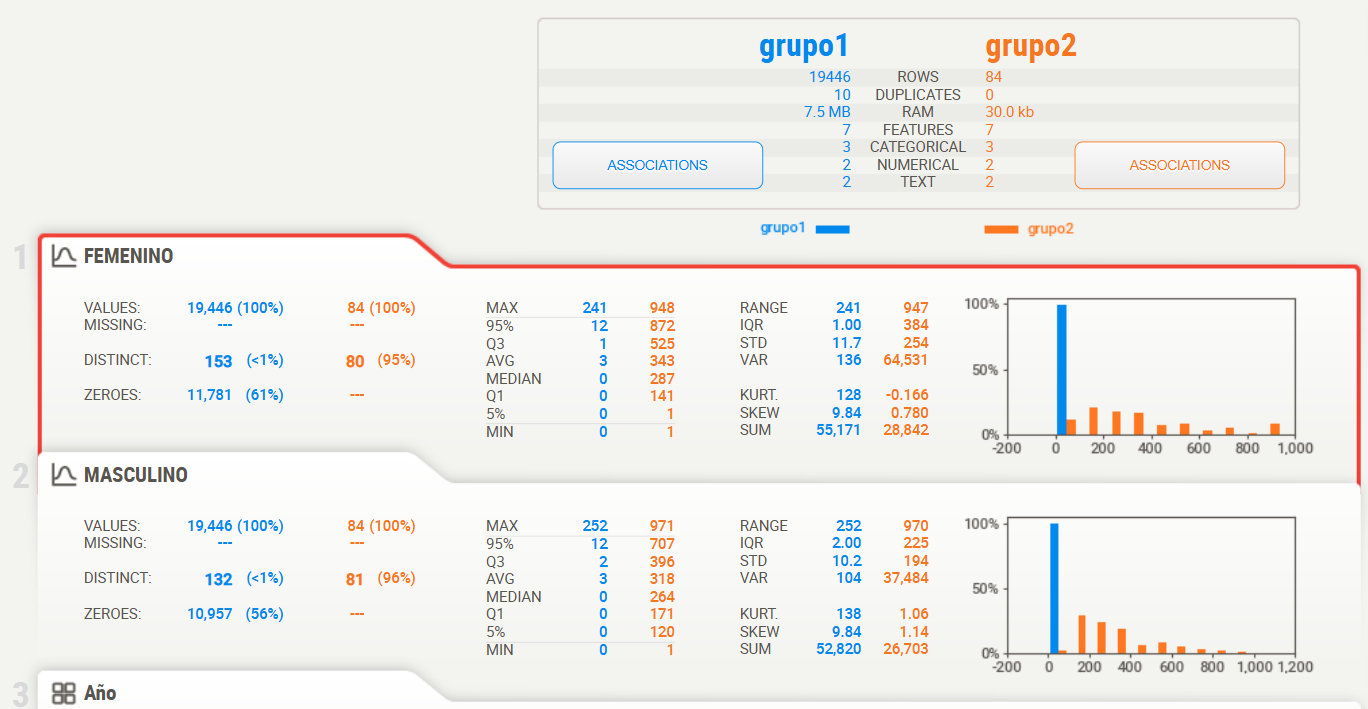
En el modelo K-Means, las variables predictoras son los atributos o características de los datos que el algoritmo usa para agrupar observaciones en clusters. Como K-Means es un método de aprendizaje no supervisado, estas variables no tienen etiquetas predefinidas, sino que el algoritmo busca patrones en ellas para formar grupos homogéneos.

En un entorno como Google Colab con Python, normalmente se trabajó con bibliotecas como scikit-learn para implementar K-Means. Aquí continuación sepresenta el flujo básico para entender el papel de las variables predictoras en este modelo:

1. Carga de datos: Se obtuvo un conjunto de datos con varias características (variables predictoras).
2. Preprocesamiento: Se estandarizaron las variables, dado que K-Means es sensible a las escalas.
3. Selección de variables predictoras: Se eligieron las columnas que se usarían para el clustering.
4. Aplicación de K-Means: Se ejecutó el algoritmo para encontrar patrones en los datos.
5. Análisis de los resultados: Se visualizaron los clusters y se interpretó el papel de cada variable en la agrupación.

### 10.1.2 Interpretación del EDA Generadas en Colab

Análisis de gráficos y tablas que se han generado para identificar patrones clave en tu base de datos y cómo estos pueden contribuir a la identificación de programas de promoción y prevención con enfoque de género en mujeres en alto riesgo de consumo de sustancias psicoactivas.



En estas imágenes se presentan estadísticas detalladas de los datos en dos grupos:

* **Grupo 1** (19,446 registros)
* **Grupo 2** (84 registros)

### Hallazgos importantes por variables clave:

### Género (Femenino vs Masculino)

* **Valores perdidos:** Ninguno.
* **Proporción de valores cero:**
  + 61% en femenino.
  + 56% en masculino.  
     Esto sugiere que más de la mitad de los registros tienen valores en cero en la variable analizada, lo que puede indicar que una gran proporción no presenta ciertos diagnósticos o factores de riesgo.
* **Máximos y Promedios:**
  + Se observa una diferencia significativa en los valores máximos y promedio entre los grupos, lo que puede reflejar variaciones en la cantidad de casos reportados por género.

### Año

* Solo hay **tres años distintos** en los datos (2018, 2019, 2020).
* Se nota una distribución pareja, aunque el año 2019 parece tener más registros en el grupo de alto riesgo.

### Nueva Ubicación

* Hay **27 ubicaciones distintas** en el grupo general y **4 en el grupo de alto riesgo**.
* Esto sugiere que el grupo de alto riesgo se concentra en menos ubicaciones, con Viterbo destacándose como una zona crítica.

### Categoría Principal

* Hay 11 categorías en el grupo general y 8 en el grupo de alto riesgo.
* Entre las más comunes en el grupo de alto riesgo se encuentran:
  + Trastornos por sustancias
  + Trastornos del humor
  + Trastornos neuróticos
  + Trastornos de comportamiento infantil

### Subcategoría y Código

* En el grupo de alto riesgo, hay 34 subcategorías y códigos específicos frente a 403 en el grupo general.
* Entre los más prevalentes en el grupo de alto riesgo están:
  + F798 (Retraso mental no especificado)
  + F330 (Trastorno depresivo recurrente, episodio leve presente)
  + F312 (Trastorno afectivo bipolar, episodio maníaco con síntomas psicóticos)

**10.2 Conclusiones**

Las mujeres en alto riesgo parecen estar más concentradas en ciertas ubicaciones y trastornos, especialmente los relacionados con el consumo de sustancias, el humor y el comportamiento.

Este análisis muestra que hay patrones claros en términos de género, ubicación y diagnóstico, lo cual es fundamental para diseñar estrategias de prevención y promoción más efectivas.

**11. Interpretación Global y Posibles Acciones**

Con base en estos hallazgos, algunas acciones clave para programas de prevención y promoción con enfoque de género podrían ser:

1. **Focalización geográfica:**
   * Concentrar esfuerzos en localidades específicas como **Viterbo** y **Riosucio**, donde hay una mayor prevalencia de casos en mujeres de alto riesgo.
2. **Enfoque en trastornos específicos:**
   * Dirigir programas hacia **trastornos por sustancias, trastornos del humor y trastornos neuróticos**, que tienen una mayor incidencia en la población femenina en alto riesgo.
3. **Aprovechar la segmentación temporal:**
   * Dado que los datos se concentran en 2019, se podrían analizar qué eventos ocurrieron ese año para entender el impacto en el consumo de sustancias.
4. **Atención a los códigos diagnósticos recurrentes:**
   * Desarrollar estrategias de intervención basadas en diagnósticos específicos como **F330 (depresión recurrente leve)** o **F312 (bipolaridad con síntomas psicóticos)**.

### 12. Conclusiones generales y recomendaciones.

Cuando se piensa en una problemática social como la situación de las mujeres en Caldas, la información institucional (bases de datos de los RIPS) no es suficiente, dado que esta información está en función de la salúd pública, pero no del desarrollo integral de las mujeres, de esta manera se sugiere para próximos proyectos formular otro tipo de variables que abarquen más elementos y más realidades, así como una base datos de entidades generadoras de datos específicos y así tener un diagnóstico más efectivo de la situación de las mujeres y poder presentar alternativas con enfoque diferencial a todo nivel.

Los RIPS revisados para efectos de este proyecto, fueron los RIPS de salud mental del departamento de Caldas, que permiten a las autoridades sanitarias identificar tendencias y patrones en los problemas de salud mental, como el aumento de casos de depresión, ansiedad, consumo de sustancias o trastornos psicóticos. Esto facilita la toma de decisiones basadas en evidencia.

Dentro de las bases de datos consultadas se evidenció que el Trastorno mental y del comportamiento debido al uso de sustancias psicoactivas (F10–F19) es el diagnóstico más recurrente entre los diagnóstico de salud mental en el departamento de Caldas. El código (F-10 y hasta el F19) son códigos que corresponden a los trastornos por uso de sustancias. y estos se encuentran enmarcados en el CIE 10, este es un sistema de clasificación Internacional de Enfermedades, 10ª Revisión) es un sistema Organización Mundial de la Salud (OMS). y su importancia se enmarca en el Diagnóstico y tratamiento, Epidemiología y Salud Pública, Gestión del Sistema de Salud, Investigación Médica y Psicológica. En Colombia , el uso del CIE-10 es obligatorio en los sistemas de salud, lo que permite una clasificación adecuada para la atención de los pacientes con trastornos por consumo de sustancias psicoactivas y el diseño de políticas públicas en salud mental.

La información del CIE 10 se consigna en los RIPS En Colombia , el uso del CIE-10 es obligatorio en los sistemas de salud, lo que permite una clasificación adecuada para la atención de los pacientes con trastornos por consumo de sustancias psicoactivas y el diseño de políticas públicas en salud mental. no solo son registros administrativos, sino que cumplen un papel clave en la gestión, planificación y vigilancia de la salud mental.

De acuerdo a los patrones evidenciados en la base de datos de salud mental, los trastornos por uso de sustancias, involucran a un patrón de comportamientos en los que los pacientes continúan utilizando una sustancia a pesar de experimentar problemas significativos relacionados con su uso.

De acuerdo con los resultados arrojados por el modelo de Inteligencia artificial es necesario reflexionar sobre los factores predisponentes asociados a la salud pública, entendiendo que los factores predisponentes son condiciones estructurales, sociales y ambientales que incrementan la vulnerabilidad de una población ante problemas de salud. Estos factores no solo afectan la aparición de enfermedades, sino que profundizan las problemáticas al dificultar su prevención y manejo, de esta manera identificar estos determinantes pueden incidir generando una visión más amplia de las problemáticas que están desencadenando en problemas de salud pública. dentro de los factores predisponentes, se encuentran:

**1. Determinantes Sociales de la Salud:** estos representan las condiciones en las que las personas nacen, crecen, viven, trabajan y envejecen. Su impacto en la salud es profundo y desigual. Entre los determinantes se encuentran, pobreza y desigualdad económica**,** educación limitada**,** desempleo y trabajo informal**.**

**2. Falta de Acceso a Servicios de Salud:** Ésta se ve reflejada en Infraestructura deficiente, baja cobertura de salud pública, discriminación y barreras culturales**,**

**3. Entorno Físico y Ambiental:** Este entorno se dimensiona en la falta de acceso a agua potable y saneamiento**,** Contaminación del aire y del agua, condiciones de vivienda precariasy exposición a hacinamiento, plagas y temperaturas extremas.

**4. Factores Psicosociales y Culturales,** Este se identifica a partir de la normalización de hábitos de riesgo**,** estigma y violencia de género**,** falta de participación comunitaria**,** que debilita el tejido social y la capacidad de respuesta ante la crisis.

**5. Políticas Públicas y Gobernanza Débil,** falta de inversión en salud preventiva**,** corrupción y mala gestión, políticas inadecuadas**,** e impacto en la profundización de problemáticas

Estos factores generan **ciclos de desigualdad** donde los problemas de salud no solo persisten, sino que empeoran, afectando sobre todo a poblaciones más vulnerables. La respuesta debe ser integral, abordando tanto los síntomas como las causas estructurales de las problemáticas en salud pública

Para diseñar programas de prevención y promoción en salud mental, es clave identificar patrones críticos que afectan a la población. Desde un enfoque de **psicología social y comunitaria** , podemos considerar tres niveles de análisis: individual, comunitario y estructural. Aquí hay algunos patrones relevantes:

**1. Nivel individual**

* **Altos niveles de estrés y ansiedad** : relacionados con precarización laboral, inseguridad o carga excesiva de cuidado.
* **Depresión y desesperanza aprendida** : factores como la pobreza, la violencia o la falta de oportunidades pueden generar sentimientos de impotencia.
* **Baja percepción de autoeficacia y autoestima** : Común en mujeres con roles de cuidado que tienen barreras para la participación política.
* **Aislamiento social** : Personas con redes de apoyo débiles o comunidades fragmentadas tienen mayor riesgo de problemas de salud mental.

**2. Nivel comunitario**

* **Normalización de la violencia** : Tanto en el hogar como en espacios públicos, lo que impacta la salud mental de las mujeres y otros grupos vulnerables.
* **Débil tejido social** : Falta de confianza entre vecinos y poca organización comunitaria, lo que limita el apoyo mutuo y la acción colectiva.
* **Escasez de espacios seguros para el diálogo y la expresión emocional** : Sin lugares donde compartir experiencias y emociones, las personas pueden sentirse solas en sus problemas.

**3. Nivel Estructural**

* **Falta de acceso a servicios de salud mental** : Pocas políticas públicas orientadas a la prevención y atención temprana.
* **Carga desigual de roles de cuidado** : Mujeres asumidas más tareas sin reconocimiento ni apoyo institucional.
* **Brechas de participación política y liderazgo** : La exclusión de las mujeres en espacios de toma de decisiones refuerza la desigualdad y puede generar frustración y desgaste emocional.

**Acciones para un programa de prevención y promoción**

A partir de estos patrones, se pueden diseñar estrategias como:  
 ✅ **Redes de apoyo comunitario** : Grupos de autocuidado y liderazgo para mujeres.  
 ✅ **Estrategias de educación emocional** : Talleres para fortalecer la resiliencia y el manejo del estrés.  
 ✅ **Espacios de participación segura** : Promoción de encuentros comunitarios para incidencia política.  
 ✅ **Atención psicológica accesible** : Desde modelos comunitarios hasta telepsicología.

**13. BIBLIOGRAFÍA**

Alcaldía Mayor de Bogotá, Secretaría Distrital de Salud de Bogotá & Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC). Estudio de consumo de sustancias psicoactivas en Bogotá D.C., 2022. Tomado de: <https://www.unodc.org/documents/colombia/2023/septiembre-9/ESTUDIO_DE_CONSUMO_DE_SUSTANCIAS_PSICOACTIVAS_BOGOTA_2022.pdf>

Comunicado de prensa del 27 de junio de 2022, viena: “El Infome Mundial sobre las Drogas 2022 de la UNODC destaca las tendencias del cannabis posteriores a la legalización, el impacto ambiental de las drogas ilícitas y el consumo de drogas entre las mujeres y las personas jóvenes. Tomado de:

<https://www.unodc.org/unodc/es/press/releases/2022/June/unodc-world-drug-report-2022-highlights-trends-on-cannabis-post-legalization--environmental-impacts-of-illicit-drugs--and-drug-use-among-women-and-youth.html>

Marchionni (2018). Brechas de género en América Latina. Un estado de situación. Banco de desarrollo de América Latina. tomado de: <https://scioteca.caf.com/bitstream/handle/123456789/1401/Brechas%20de%20genero%20en%20America%20Latina.%20Un%20estado%20de%20situacion.pdf?sequence=5&isAllowed=y>

Ministerio de Salud y Protección Social. (2018). Resolución 4886 de 2018. Política Nacional de Salud Mental.

Oficina de las Naciones Unidad. Objetivos de desarrollos sostenible. Tomado de: <https://ods.dnp.gov.co/es/objetivos/salud-y-bienestar>

1. Alcaldía Mayor de Bogotá, Secretaría Distrital de Salud de Bogotá & Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC). Estudio de consumo de sustancias psicoactivas en Bogotá D.C., 2022. Tomado de: <https://www.unodc.org/documents/colombia/2023/septiembre-9/ESTUDIO_DE_CONSUMO_DE_SUSTANCIAS_PSICOACTIVAS_BOGOTA_2022.pdf> [↑](#footnote-ref-0)
2. Comunicado de prensa del 27 de junio de 2022, Viena: “El Informe Mundial sobre las Drogas 2022 de la UNODC destaca las tendencias del cannabis posteriores a la legalización, el impacto ambiental de las drogas ilícitas y el consumo de drogas entre las mujeres y las personas jóvenes. Tomado de:

   <https://www.unodc.org/unodc/es/press/releases/2022/June/unodc-world-drug-report-2022-highlights-trends-on-cannabis-post-legalization--environmental-impacts-of-illicit-drugs--and-drug-use-among-women-and-youth.html> [↑](#footnote-ref-1)
3. Ibidem. [↑](#footnote-ref-2)
4. Marchionni (2018). Brechas de género en América Latina. Un estado de situación. Banco de desarrollo de América Latina. tomado de: <https://scioteca.caf.com/bitstream/handle/123456789/1401/Brechas%20de%20genero%20en%20America%20Latina.%20Un%20estado%20de%20situacion.pdf?sequence=5&isAllowed=y> [↑](#footnote-ref-3)
5. <https://ods.dnp.gov.co/es/objetivos/salud-y-bienestar> [↑](#footnote-ref-4)
6. Ministerio de Salud y Protección Social. (2018). *Resolución 4886 de 2018. Política Nacional de Salud Mental*. [↑](#footnote-ref-5)