

Impacto de la identidad y los factores socioeconómicos en la prevalencia de enfermedades mentales en Bogotá

Esteban Rico Moreno

Tomas Calixto Bedoya

Nicolas Londoño Cuellar

Amalia Jaramillo Arango

Wendy Argüello Hernández

23 de mayo de 2024

Inteligencia de Negocios

0. Resumen

El presente proyecto es un trabajo multidisciplinar realizado por un grupo de estudiantes de medicina e ingeniería de sistemas de la Universidad de los Andes. El proyecto pretende utilizar los datos de encuestas multipropósito realizadas por el DANE en Bogotá y municipios aledaños para apoyar un trabajo analítico con enfoque médico.

En este documento, se presentan los objetivos del proyecto a nivel general y a nivel particular de cada carrera, se presenta el enfoque de investigación, se resumen los procesos de diseño del Data Mart desde la explicación de su modelo multidimensional, así como la explicación del proceso ETL para la carga, organización y utilización de los datos en un tablero de control utilizado para apoyar el proceso de análisis.

Finalmente, se discuten las conclusiones del proceso analítico y se discuten brevemente las limitaciones que se encontraron a lo largo del trabajo de este proyecto.

Contenido

0. Resumen.....	2
1. Introducción.....	4
1.1 Objetivos del Proyecto.....	4
1.2 Contexto y Justificación.....	4
2. Metodología	5
2.1 Identificación de Requerimientos Analíticos	5
2.2 Entendimiento de los datos	6
2.3 Arquitectura de la solución construida y Diseño de Data Marts	6
2.4 Diseño e Implementación de ETL	8
3. Análisis Descriptivo	10
3.1 Tableros de Control.....	10
3.2 Análisis Realizados e interpretación de los resultados.....	11
4. Discusión y Conclusiones	11
4.1 Significancia Estadística.....	11

4.2 Significancia Clínica	12
4.3 Limitaciones del Estudio	12
5. Video	12
6. Anexos	13
6.1 Documentación Técnica.....	13
6.2 Material Complementario.....	13
7. Referencias	13

1. Introducción

1.1 Objetivos del Proyecto

Como objetivo global, el proyecto buscó estudiar cómo la intersección entre diversos hábitos, factores socioeconómicos e identidad de género modula la prevalencia de enfermedades mentales en la población de Bogotá. Para dicho propósito, se manejaron desde ambas disciplinas los siguientes objetivos particulares:

Ingeniería de Sistemas:

- Proporcionar al equipo de análisis médico una herramienta que facilite la visualización e interpretación de datos relevantes para el objetivo de estudio, utilizando los datos provistos por las encuestas multipropósito del DANE

Medicina:

-

1.2 Contexto y Justificación

La salud mental es una preocupación creciente a nivel mundial, y las disparidades en la prevalencia de trastornos como la ansiedad y la depresión dentro de las comunidades marginalizadas son particularmente alarmantes. Investigaciones previas han demostrado que la pertenencia a la comunidad LGBTQ+ está asociada con un mayor riesgo de trastornos mentales como la ansiedad y la depresión (Moagi et al, 2021).

Como se estableció previamente, las personas que pertenecen a la comunidad LGBTQIA+ experimentan mayores tasas de ansiedad y depresión en comparación con la población heterosexual. Adicionalmente, según Gómez-Chica et al. (2024), también presentan mayor riesgo de suicidio y esta situación se ve agravada por diversos factores, como el estrés minoritario, la discriminación, la victimización y la falta de acceso a servicios de salud mental afirmativos. La pandemia de COVID-19 ha exacerbado aún más las desigualdades en salud mental para los jóvenes LGBT. Un informe de The Trevor Project (2021) encontró que los jóvenes LGBT experimentaron mayores niveles de ansiedad y depresión durante la pandemia, y que estos desafíos se vieron agravados por el aislamiento social y la falta de acceso a recursos de salud mental.

Por otro lado, la falta de atención adecuada y accesible a la salud mental también es un problema importante para los jóvenes LGBT. Hafeez et al. (2017) encontró que los jóvenes LGBT enfrentan barreras significativas para acceder a la atención de salud mental, incluyendo la falta de proveedores capacitados y la discriminación en el sistema de atención médica. Los profesionales de la salud mental pueden tener sesgos o falta de conocimiento sobre las necesidades específicas de la comunidad LGBT, lo que puede llevar a diagnósticos erróneos o subestimaciones de la gravedad de los síntomas.

Por las razones mencionadas anteriormente, es necesario realizar más investigaciones para comprender mejor las complejidades de la salud mental en la comunidad LGBT y desarrollar estrategias efectivas para la prevención, el diagnóstico y el tratamiento de la ansiedad y la depresión en esta población. Por último, también es necesario implementar acciones para reducir el estigma y la discriminación hacia las personas LGBT, aumentar el acceso a servicios de salud mental y capacitar a los profesionales de la salud mental para brindar una atención médica adecuada a esta comunidad.

2. Metodología

2.1 Identificación de Requerimientos Analíticos

Dado que el tema analítico es el impacto de factores socioeconómicos, con énfasis en identidad de género, en el desarrollo de enfermedades mentales, los análisis requeridos son:

- Evaluación de elementos demográficos como la edad, la orientación sexual, el género, y la educación.
- Evaluación de condiciones de vivienda, como la vivienda en sí misma y los servicios a los que tiene acceso el individuo.
- Los hábitos del individuo.

2.2 Entendimiento de los datos

Los datos provienen de las encuestas multipropósito realizadas por el DANE en Bogotá y municipios aledaños. En ella, se registra información acerca de diferentes características de los individuos, sus hogares, y sus condiciones de vida. La encuesta está enfocada al hogar, por lo que cada registro de los datos tiene la identificación de un hogar particular, pero un registro particular corresponde a un individuo. No obstante, no es posible determinar si un hogar o individuo del 2017 es homólogo a uno del 2021, por lo que los casos de ambos años son tomados como singulares a pesar de que, en realidad, pueda haber personas que aparecen 2 veces a través de ambas encuestas. Adicionalmente, dadas ciertas preguntas, es posible concluir que las entrevistas fueron respondidas por mayores de 18 años, y todo menor de 18 debió ser representado por un mayor de edad, por lo que la información referente a menores de edad se encuentra en el registro de un adulto responsable.

Dada la amplitud de los datos de la encuesta, para el entendimiento de los datos nos centramos en aquellos más relevantes para nuestro enfoque de investigación: factores socioeconómicos, hábitos, y componentes de la identidad del individuo. Las preguntas realizadas por el DANE son, por lo general, de 2 tipos: el primer tipo son opciones múltiples, preguntas como el material de los pisos de la casa, el estrato, el sexo y el estado civil; el segundo tipo son de sí o no, preguntas como si un servicio es cubierto en un hogar, si ha realizado una actividad particular en los últimos 30 días, o si la persona entrevistada es cabeza de hogar.

Dicho lo anterior, consideramos importante mencionar que, si bien la mayoría de las preguntas que utilizamos fueron realizadas iguales tanto en el 2017 como en el 2021, hubo una excepción. Las preguntas de género cambian entre 2017 y 2021; mientras que en el 2017 se tiene la pregunta de opción múltiple “¿Con qué género se identifica?” (Femenino, Masculino y Transgenero son las respuestas posibles), en el 2021 esta pregunta se parte en 2: La primera es la pregunta de sí o no “¿Usted se siente identificado/a con el sexo que le fue asignado al nacer?”, y la segunda es la pregunta de opción múltiple “Usted se reconoce como:” con hombre trans, mujer trans y otro como respuestas posibles.

Como último aspecto pertinente, pudimos ver que el único dato sobre cuando se hicieron las encuestas es el año, pero no hay mayor especificación de fechas para cada individuo particular.

2.3 Arquitectura de la solución construida y Diseño de Data Marts

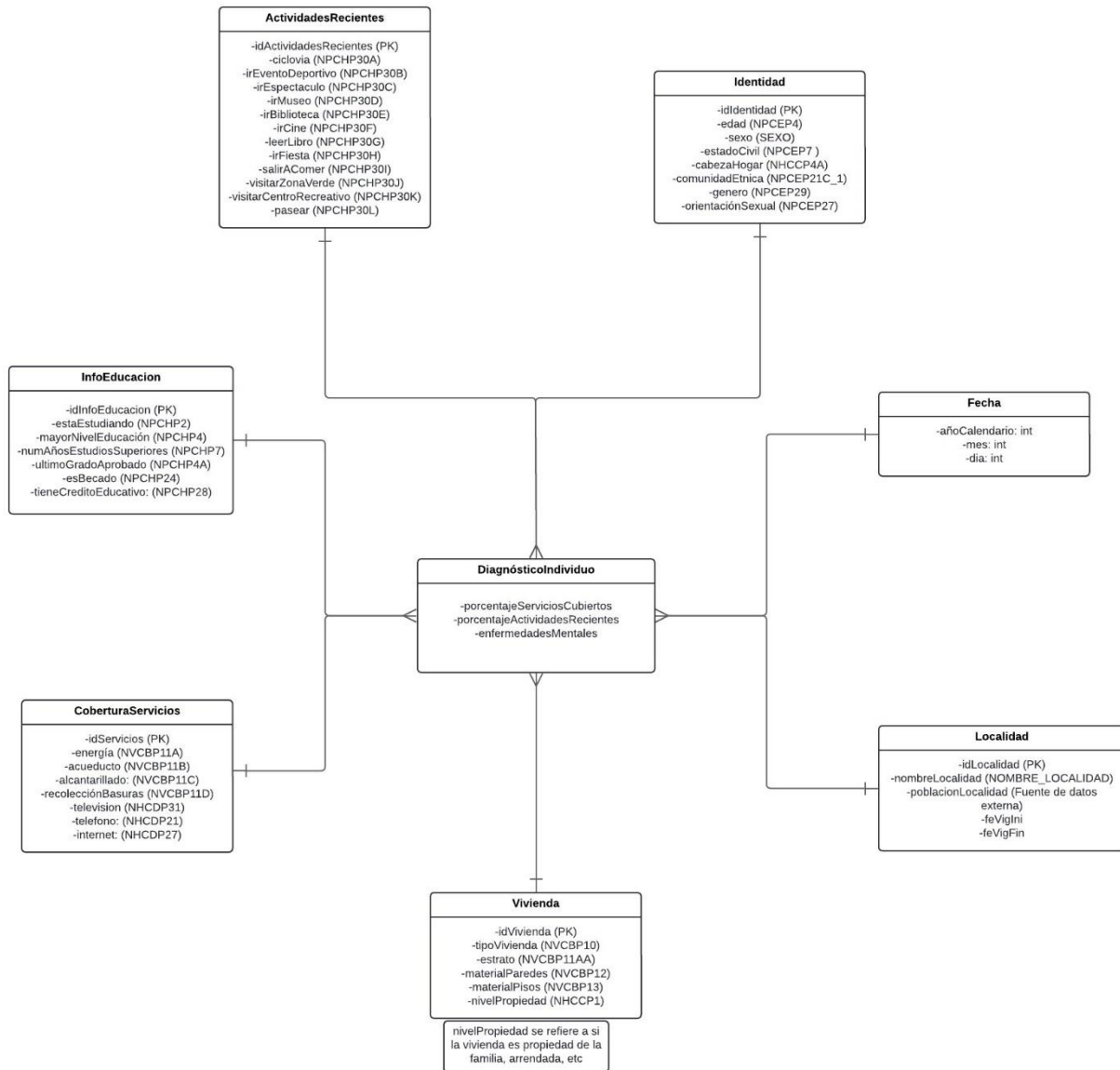


Figura 1: Modelo Multidimensional del Data Mart

El modelo multidimensional toma datos de la encuesta multipropósito considerados relevantes para el proyecto. Tiene una tabla de hechos, que corresponde a un diagnóstico individual, y los datos considerados relevantes se agruparon en dimensiones que representan la vivienda donde vive la persona, su localidad, la fecha en la que se realizó el diagnóstico, información sobre el historial educativo del individuo, la cobertura de servicios en el hogar del individuo, las actividades que ha realizado en los últimos 3 días, y finalmente, información acerca de su identidad y sexualidad. La tabla de hechos tiene una granularidad a nivel de persona puesto que se desea facilitar el análisis de factores que inciden en el desarrollo de enfermedades mentales a nivel individual.

Además del diagnóstico con respecto a si el individuo tiene o no una enfermedad mental, la tabla de hecho incluye dos medidas por su posible interés analítico, siendo

En el modelo incluimos de qué columnas de la encuesta multipropósito se extrae la información. Nótese que para el atributo género en la dimensión de Identidad, la columna incluye un asterisco. Esto se debe a lo mencionado en el entendimiento de datos, siendo esto para el 2021 esta columna se parte en 3. Más adelante, en el proceso de ETL, explicaremos qué hicimos para solucionar estas discrepancias.

2.4 Diseño e Implementación de ETL



8

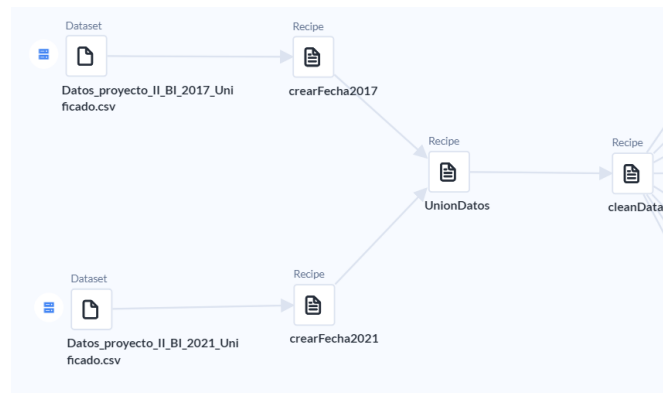


Imagen 2. ETL primera parte en dataprep

En lo que respecta a la primera parte, esta consiste en un conjunto de pasos previos para unificar los datasets de 2021 y 2017, y una vez unificados realizar una sección de limpieza de datos. En lo que respecta a los pasos previos se asegura de que para cada dataset se cree una fecha que contenga el año 2021 y 2017, con esto se puede realizar la validación del dataset para un futuro. Luego se unifica en base al nombre, para finalmente realizar un proceso de limpieza, en este proceso de limpieza primero se cambia la columna que relaciona los géneros, pues la toma de datos no fue la misma en los años, por lo que se realizó una modificación para que tomar el valor de 1 o 2, siendo cisgénero y transgénero las opciones (ya que son de extrema relevancia para el requerimiento analítico). Segundo se realizó una limpieza exhaustiva cambiando valores inválidos, o valores como true o false a 0 y 1 para mantener el formato de entero en todos los datos, por último, se realizó el cambio de valores nulos a 0. Al final el dataset que genera este recipe limpio es el que se utiliza para todos los pasos.

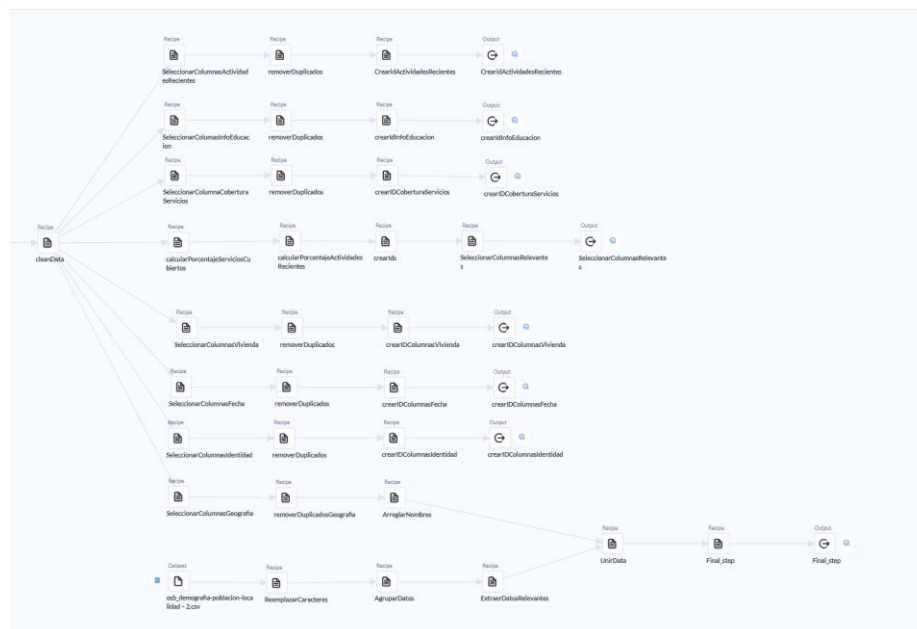


Imagen 3. ETL segunda parte en dataprep

Para este punto es bastante relevante mencionar que todo se deriva de la parte de clean data, además la gran mayoría de dimensiones siguen un patrón bastante similar, a excepción de 2: la de geografía, y la de diagnostico de individuos.

Para el caso de las dimensiones que comparten similitud en su patrón, estas, en general, consisten en hacer una selección de las columnas relevantes (dado el dataset y su código), renombrarlas, remover duplicados para asegurar unicidad, y luego generar un id, este id se realiza mediante una concatenación de cada valor junto con un '_', esto asegura que se pueda replicar luego en la tabla de hechos.

En el caso de geografía, este se divide en 2 flujos, el primero que proviene de clean data, en este solo se remueven los duplicados y se arreglan los nombres ya que se notó que por ejemplo Teusaquillo estaba como 2quillo. Mientras que en el otro flujo se toman los datos de otro dataset denominado demografía_población_localidad, para este solo se reemplazan los caracteres con el fin de eliminar las "ñ" y las tildes. Luego se agrupan los datos para poder conseguir la suma de la población por la localidad, para posteriormente extraer los datos relevantes como son los de la fecha 2017 y 2021 únicamente. Una vez obtenida esta información, se realiza la unión para como final_step eliminar columnas irrelevantes para el problema.

En el caso del diagnostico de individuo que es la tabla de hechos, primero se usa toda la información de clean_data para conseguir los porcentajes de servicios cubiertos y de actividades recientes por id, luego de esto se crean los ids en base a la información ya obtenida, una vez logrado se seleccionan las columnas relevantes y se genera el output de la dimensión.

Para el manejo de la diferencia entre el género del 2017 y el del 2021, lo que se hizo fue agrupar los valores en 2 categorías: Cisgenero y Transgénero. En el 2017, si el individuo respondía que se identificaba como transgénero, se clasificaba como tal, de lo contrario se clasificaba como Cisgenero. En cambio, en el 2021, si el individuo contestaba que se sentía identificado con el género asignado en su nacimiento, se le clasificaba como cisgénero, de lo contrario, se le clasificaba como transgénero.

Toda la información a detalle se encuentra en el Excel que modela el ETL adjunto en el repositorio en el que se encuentra el archivo.

3. Análisis Descriptivo

3.1 Tableros de Control

En el siguiente link están los tableros de control que se explicarán a continuación.

https://uniandes-my.sharepoint.com/personal/n_londonoc_uniandes_edu_co/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fn_londonoc_uniandes_edu_co%2FDocuments%2FBI%2FProyecto%202%20BI%2Epbix&parent=%2Fpersonal%2Fn_londonoc_uniandes_edu_co%2FDocuments%2FBI&ga=1

El archivo tiene tres tableros que funcionan como días una descripción analítica para responder varias preguntas y toparse con puntos interesantes sobre los datos. Para el primer tablero se cuenta con un mapa de calor de prevalencia de enfermedades mentales en las localidades de Bogotá, este puede actuar como filtro. Justo a este también se cuenta con una tabla de las localidades con la cantidad de gente encuestada y el porcentaje de afectados. Adicionalmente, se cuenta con graficas las cuales segregan por estrato. Es de notar que la mayoría de datos son de estratos 2 y 3.

En la siguiente página de los tableros tenemos factores relacionados a la identidad y sexualidad de los encuestados, en este encontramos varios filtros para segregar las gráficas por sexo, orientación sexual, localidad y estrato. Las gráficas incluyen las edades y la prevalencia de enfermedades mentales en la población.

La última página del tablero muestra una vista de los factores socioeconómicos y su relevancia frente a enfermedades mentales, entre estos factores se encuentran el estrato, máximo nivel educativo alcanzado, localidad, nivel de vivienda y estado civil. Estas gráficas nos dan una vista en que tan relevantes son estos factores a la hora de entender las enfermedades mentales y como afectan diferencialmente a la población.

3.2 Análisis Realizados e interpretación de los resultados

[Descripción general de los análisis realizados utilizando los tableros de control. Interpretación de los resultados (Hallazgos). Reflexione sobre cómo los resultados podrían contribuir a la **comprensión** del fenómeno estudiado.]

Para empezar, se tiene que solo el 1.5% de la población padece de alguna enfermedad mental. Este número luce sospechosamente bajo, considerando el estimado de la OMS de 1 de cada 8. Según las gráficas de edad, el grueso de la población encuestada se encuentra entre los 50 y 80 años. Enfocándose ya más en la pregunta de investigación, el segundo tablero nos indica de una situación interesante con los datos e identidad de género. En el total encuestado solo existen 365 personas que se denominan como pertenecientes a la comunidad trans. Aunque las gráficas indiquen que existe una correlación entre mayor tasa de enfermedad mental en poblaciones trans, consideramos que son bastante pocos datos en comparación como para sacar un resultado contundente. Estudios externos indican a que la proporción de población trans oscila entre el 0.4 y 2%. Sin embargo, la proporción que tenemos actualmente es 2 ordenes de magnitud más pequeño que lo esperado. Esto puede ser indicativo de subrepresentación en las encuestas. Similar alcanza a ocurrir con respecto a la orientación sexual, pero a menor escala. No obstante, asumiendo el valor de los datos y su veracidad, se encuentran patrones correlacionados entre tasa de enfermedad mental y poblaciones LGBT.

Finalmente, el último tablero nos indica resultados interesantes al segregar por nivel de vivienda. Un hallazgo interesante es evidente al seleccionar a toda la población con casa propia. Al hacer esto, se puede evidenciar como esta población se divide en dos franjas distintivas de edad. Adolescencia y adultez joven y tercera edad. Estos datos coinciden de igual manera con la población que padece de enfermedades mentales, indicando una posible correlación.

4. Discusión y Conclusiones

4.1 Significancia Estadística

[En esta sección, se analizan los resultados del proyecto desde la perspectiva de estadística inferencial, por medio de pruebas de hipótesis y/o intervalos de confianza.]

4.2 Significancia Clínica

La importancia clínica radica en entender las dinámicas específicas de la salud mental, así que se podrían desarrollar programas de apoyo psicosocial que aborden los síntomas y las causas subyacentes relacionadas con la comunidad LGBT y los factores socioeconómicos, lo que podría mejorar significativamente la calidad de vida de las personas afectadas. Así mismo, comprender cómo los factores socioeconómicos y el pertenecer a la comunidad LGBT se interceptan para influir en la prevalencia de enfermedades mentales lo cual permite identificar grupos de población que están en mayor riesgo siendo esto crucial para desarrollar intervenciones dirigidas y eficaces.

4.3 Limitaciones del Estudio

Uno de los objetivos iniciales de este estudio era realizar un análisis con un enfoque de género. Sin embargo, encontramos una limitación significativa en los datos recopilados mediante las encuestas. Dada la diferencia entre el modo de catalogar género entre el 2017 y el 2021, para el manejo de datos fue necesario dividir la categoría de género únicamente en 2, Cisgenero y Transgenero, limitando así la posible diversidad analítica.

Para futuros estudios, sería recomendable ajustar las herramientas de recolección de datos para incluir preguntas que permitan una identificación más detallada del género, a fin de proporcionar un análisis más exhaustivo y preciso en términos de este.

Así mismo, otra de las limitaciones de nuestro estudio radica en la ausencia de una pregunta específica que permita a los participantes identificarse explícitamente como miembros de la comunidad LGBT. Esta omisión implica que no hemos podido capturar completamente la diversidad de orientaciones sexuales presentes entre los participantes. Esta falta de datos precisos sobre la identificación LGBT podría haber influido en los resultados y en las conclusiones derivadas del estudio, sub representando o invisibilizando a algunos subgrupos dentro de la comunidad.

Para futuras investigaciones, recomendamos la inclusión de preguntas detalladas que permitan a los participantes especificar su orientación sexual y su identidad de género de manera más precisa. Esto no solo mejorará la representatividad de los datos, sino que también permitirá un análisis más inclusivo y matizado de las diversas experiencias dentro de la comunidad LGBT.

En general, las estadísticas de género y orientación sexual fueron algo problemáticas de trabajar. La población trans, por ejemplo, cuenta con muy pocos encuestados, demostrando un vacío en las estadísticas que el DANE tiene sobre miembros importantes de la comunidad LGBTI. Inevitablemente, esto compromete en cierto grado la confiabilidad del análisis.

5. Video

En el siguiente link se encuentra adjuntado el vídeo: <https://youtu.be/iUAQWVIPU0A>

6. Anexos

6.1 Documentación Técnica

Para la documentación técnica pueden consultarse los archivos de este repositorio. Encontrarán el documento de planeación del ETL, el documento de ejecución del ETL, y una imagen más grande del modelo multidimensional.

6.2 Material Complementario

7. Referencias

<https://www.who.int/news-room/fact-sheets/detail/mental-disorders>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5227946/>