Transformando datos en conocimiento: sistema simplificado para analizar variables sociodemográficas y el mercado laboral en la GEIH 2024

Ernesto Aldana Gaona

Asesor

Luis Ángel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD

Escuela Ciencias Básicas Tecnología e Ingeniería ECBTI

Especialización en Ciencias de Datos y Analítica

2025

Danitza María Cortez Pérez

_____

Nombre Director de Trabajo de Grado

_____ _____

Jurado Jurado

2025

Dedicatoria

En primer lugar, agradezco a Dios por bendecirme con la oportunidad de realizar esta especialización después de décadas de dejar de estudiar. Mi agradecimiento también va para mis hijos, por su apoyo incondicional, paciencia y el constante ánimo que me brindaron durante mis largas jornadas de estudio, así como por su interés y preocupación en saber cómo transcurría mi proceso académico.

Agradecimientos

Agradezco a la Universidad Nacional Abierta y a Distancia (UNAD) y a sus profesores por compartir sus conocimientos y brindarme su valiosa orientación, lo cual me permitió cumplir con éxito y aprobar las materias de la especialización.

Resumen

El presente proyecto tiene como objetivo desarrollar un sistema de información simplificado que integre variables sociodemográficas y los principales indicadores del mercado laboral contenidos en los microdatos anonimizados de la Gran Encuesta Integrada de Hogares (GEIH) 2024. Estos microdatos, disponibles públicamente a través del portal de Datos Abiertos del Ministerio de Tecnologías de la Información y las Comunicaciones (Ministerio TIC, s. f.), permiten realizar análisis profundos sobre el mercado laboral colombiano. El sistema, desarrollado en Power BI, facilitará el acceso, procesamiento y visualización de los datos, promoviendo una toma de decisiones basada en evidencia.

***Palabras clave*:** GEIH, DANE, microdatos, sociodemográficas, mercado laboral.

Abstract

The objective of this project is to develop a simplified information system that integrates sociodemographic variables and the main labor market indicators contained in the anonymized microdata from the 2024 Integrated Household Survey (GEIH). This microdata, publicly available through the Open Data portal of the Ministry of Information and Communications Technology (Ministry of ICT, n.d.), allows for in-depth analysis of the Colombian labor market. The system, developed in Power BI, will facilitate access, processing, and visualization of the data, promoting evidence-based decision-making.

***Keywords*:** GEIH, DANE, microdata, sociodemographics, labor market.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Lista de Apéndices

Introducción

En el contexto actual, el acceso a información precisa y actualizada es fundamental para la toma de decisiones en diversos ámbitos. En particular, el mercado laboral es un área crítica para la planificación y el desarrollo de políticas públicas. La Gran Encuesta Integrada de Hogares (GEIH), publicada mensualmente por el Departamento Administrativo Nacional de Estadística (DANE), proporciona información detallada sobre el mercado laboral y las características sociodemográficas de la población. Sin embargo, estas bases de datos están separadas y presentan una complejidad estructural que dificulta su integración, consulta y análisis eficiente, lo que limita la obtención de conclusiones oportunas y precisas (DANE, 2020).

Este proyecto propone desarrollar un sistema de información simplificado que integre las variables sociodemográficas y los indicadores del mercado laboral contenidos en los microdatos de la GEIH 2024. Utilizando Power BI, una herramienta que permite modelar y visualizar datos, se facilitará el análisis y la toma de decisiones mediante informes interactivos.

Descripción del Problema

Planteamiento del Problema

El procesamiento de los microdatos de la Gran Encuesta Integrada de Hogares (GEIH) del Departamento Administrativo Nacional de Estadística (DANE) requiere conocimientos avanzados en programación y comprensión detallada de la estructura de la encuesta. Esto se debe a que los microdatos están distribuidos en múltiples archivos correspondientes a distintos capítulos del formulario, los cuales deben ser integrados y tratados para realizar análisis coherentes sobre el mercado laboral colombiano (DANE, 2022a). Según la *Guía de acceso y uso de bases anonimizadas* publicada por el DANE las bases de datos están disponibles en tres formatos: CSV, DTA, y SAV (DANE, 2020).

En la plataforma de Datos Abiertos estos microdatos están en tres formatos (CSV, DTA, SAV), cada uno con ocho archivos distintos por mes (Ministerio TIC, s.f.), lo que representa un reto significativo para los usuarios sin formación especializada. Además, investigaciones previas han evidenciado que la complejidad en la estructura de microdatos del DANE dificulta el acceso efectivo a la información por parte de tomadores de decisiones, periodistas de datos y ciudadanos interesados (Martínez & Díaz, 2019; Bernal & Flórez, 2020).

Esto genera una barrera de entrada para un uso más democratizado de los datos, limitando su utilidad en el análisis de dinámicas laborales, generación de políticas públicas o estudios sociales. Como resultado, se observa una necesidad creciente de herramientas que simplifiquen el acceso y análisis de los microdatos, permitiendo que personas sin conocimientos técnicos avanzados puedan explorar, visualizar y comprender los indicadores laborales relevantes (Morales & Pineda, 2021; DANE, 2022b).

Por ello, se plantea la necesidad de desarrollar un sistema de información simplificado, que automatice la integración y procesamiento de las variables sociodemográficas y laborales contenidas en la GEIH. Este sistema facilitaría la visualización de indicadores clave y permitiría a usuarios no especializados realizar análisis de forma intuitiva, fomentando una toma de decisiones más informada y una mayor apropiación social del conocimiento estadístico (UNESCO, 2020; Bernal et al., 2021).

Pregunta de Investigación

¿Cómo desarrollar un sistema de información simplificado que integre y procese los microdatos de la Gran Encuesta Integrada de Hogares (GEIH), para facilitar el acceso, análisis y uso de la información sobre el mercado laboral colombiano por parte de usuarios sin conocimientos técnicos avanzados?

Sistematización del Problema

Justificación

El Departamento Administrativo Nacional de Estadística (DANE) publica mensualmente microdatos anonimizados de la GEIH en formatos CSV, DTA y SAV, con una estructura compleja y dispersa en múltiples archivos. Esta situación dificulta el análisis para usuarios sin formación técnica (DANE, 2022). Además, estudios han evidenciado que esta barrera limita el uso democrático de los datos (Martínez & Díaz, 2019; Morales & Pineda, 2021).

La accesibilidad a los datos públicos es un factor clave para fomentar la transparencia, el control ciudadano y la toma de decisiones informada (UNESCO, 2020). La visualización interactiva mediante plataformas como Power BI puede mejorar significativamente la comprensión de los indicadores por parte de usuarios no técnicos (Chaves & Hernández, 2021).

En consecuencia, el desarrollo de este sistema tiene valor estratégico al promover la apropiación social del conocimiento estadístico.

Objetivos

Objetivo General

Desarrollar un sistema de información simplificado para el estudio de las variables sociodemográficas con los principales indicadores del mercado laboral de la población encuestada de la Gran Encuesta Integrada de Hogares (GEIH) 2024, visualizando los datos mediante la herramienta Power BI Desktop, utilizada en el ámbito de la ciencia de datos.

Objetivos Específicos

Identificar las variables sociodemográficas relevantes de la población encuestada en la GEIH 2024, y clasificar los indicadores del mercado laboral presentes.

Desarrollar el prototipo de sistema de información que integre las variables sociodemográficas y los indicadores del mercado laboral.

Analizar los datos obtenidos para detectar patrones y tendencias en el mercado laboral a partir de las variables sociodemográficas.

Visualizar los resultados de los análisis mediante la herramienta Power BI Desktop que facilite la interpretación y toma de decisiones.

Marco de Referencia

Marco Conceptual

Sistema de Información Simplificado: Un sistema diseñado para facilitar el acceso, procesamiento y análisis de los datos de manera eficiente, minimizando la complejidad del usuario final y optimizando el flujo de trabajo para consultas rápidas y precisas.

Datos Abiertos (s.f): Son datos digitales que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar.

Datos Anonimizados de Uso Público (DANE, 2011): Son microdatos anonimizados que se difunden para uso público general fuera del DANE. El nivel de protección de la confidencialidad en los archivos de uso público es tal que la identificación de la fuente no es posible aun cuando se cruce con otros archivos de datos.

Variables Sociodemográficas: Según Marketing Digital (s.f.) son características de una población que describen sus aspectos sociales y demográficos, como edad, género, nivel educativo, estado civil, tipo de ocupación, entre otros. Estas variables son fundamentales para el análisis de tendencias sociales y económicas.

Indicadores del Mercado Laboral: Según el Departamento Administrativo Nacional de Estadística (Dane,2023) son medidas que reflejan la situación laboral de una población, como tasas de empleo, desempleo, subempleo, participación laboral, y características del empleo.

La Gran Encuesta Integrada de Hogares – GEIH: Es una encuesta realizada por el Departamento Administrativo Nacional de Estadística (DANE, s.f.), que proporciona información sobre el mercado laboral, las condiciones de vida y las características sociodemográficas de los hogares colombianos.

Ciencia de Datos: Según AWS Amazon (s.f.) es un campo interdisciplinario que utiliza herramientas, técnicas y algoritmos de estadísticas, matemáticas y programación para extraer conocimiento y tomar decisiones a partir de grandes volúmenes de datos.

Antecedentes

La Gran Encuesta Integrada de Hogares (GEIH), realizada por el Departamento Administrativo Nacional de Estadística (DANE), es la fuente oficial más importante de información sobre el mercado laboral en Colombia. Según Hermida y Pulido-Mahecha (2022), esta encuesta permite estimar indicadores fundamentales como la tasa de desempleo, la tasa de ocupación y la distribución del ingreso, lo que la convierte en una herramienta esencial para la investigación social y la formulación de políticas públicas. En su Ficha Metodológica más reciente, el DANE (2023) detalla que la GEIH también proporciona datos sobre condiciones de vida, educación, informalidad, y características sociodemográficas de los hogares colombianos.

El acceso a los microdatos de la GEIH es público, gratuito y se realiza a través del portal de Datos Abiertos (Ministerio TIC, s.f.). Sin embargo, estos datos están fragmentados en múltiples archivos mensuales y estructurados por capítulos temáticos, lo que dificulta su integración y análisis sin conocimientos técnicos (DANE, 2020). Aunque existen guías y manuales para el uso de estos datos (DANE, 2022), varios estudios coinciden en que esta estructura limita el acceso efectivo por parte de la ciudadanía y sectores no especializados (Martínez & Díaz, 2019; Morales & Pineda, 2021).

Frente a esta situación, la implementación de sistemas interactivos y visuales se ha posicionado como una solución viable para mejorar la accesibilidad a los datos públicos. Por ejemplo, el Ministerio del Trabajo de Colombia desarrolló la plataforma FILCO (Fuente de Información Laboral de Colombia), que ofrece indicadores laborales consolidados a través de un

portal web. No obstante, esta herramienta se basa principalmente en datos ya procesados y no permite una exploración personalizada de los microdatos como lo haría un dashboard dinámico (Ministerio de Trabajo, s.f.).

En el ámbito académico y técnico, se han realizado diversos esfuerzos para emplear herramientas de inteligencia de negocios como Power BI, Tableau, y R Shiny en la creación de dashboards enfocados en datos públicos. Ruiz, Salazar y Torres (2020) desarrollaron una visualización de indicadores de gestión pública usando Power BI, mejorando la interpretación de la información por parte de entidades territoriales. Chaves y Hernández (2021) presentaron un modelo para simplificar datos estadísticos nacionales con herramientas visuales, evidenciando mejoras en la toma de decisiones institucionales.

A nivel internacional, la OCDE (2019) y la UNESCO (2020) han promovido el uso de plataformas visuales como medio para democratizar el acceso a estadísticas públicas. En América Latina, iniciativas como el observatorio laboral del Ministerio de Trabajo de Chile, o el sistema de visualización de empleo juvenil de la CEPAL, han demostrado el potencial de estos sistemas para facilitar la comprensión de fenómenos laborales complejos (CEPAL, 2021; UNESCO, 2020).

Por su parte, Bernal y Flórez (2020) resaltan que en Colombia hay un rezago en el uso de herramientas interactivas aplicadas a microdatos estadísticos, a pesar de contar con fuentes tan ricas como la GEIH. En la misma línea, Bernal, Mejía y Piraquive (2021) desarrollaron un estudio sobre la usabilidad de plataformas de datos abiertos en el sector público colombiano, y concluyeron que la experiencia del usuario mejora significativamente cuando se implementan visualizaciones dinámicas y filtros intuitivos.

Finalmente, investigaciones como las de Morales y Pineda (2021) y Pérez & Rodríguez (2022) evidencian que la creación de plataformas que integren microdatos con mecanismos de visualización personalizable no solo mejora el acceso a la información, sino que empodera a los ciudadanos, investigadores y gestores públicos para interpretar datos con mayor precisión y oportunidad.

Marco Legal

Los datos del mercado laboral del DANE son la fuente oficial de Colombia porque el DANE es la entidad responsable de producir y difundir estadísticas oficiales, lo que garantiza que los datos sean confiables, representativos, transparentes y útiles para la toma de decisiones gubernamentales y la planificación de políticas públicas.

Datos abiertos (Ministerio TIC, s.f.) los datos abiertos son información pública dispuesta en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento. En Colombia, la Ley 1712 de 2014 de la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional.

Según la Gran Encuesta Integrada de Hogares - GEIH – 2024, el acceso a los microdatos anonimizados de uso público es de carácter gratuito y estará disponible en la página Web del (DANE, 2025).

Metodología

Introducción de la Metodología

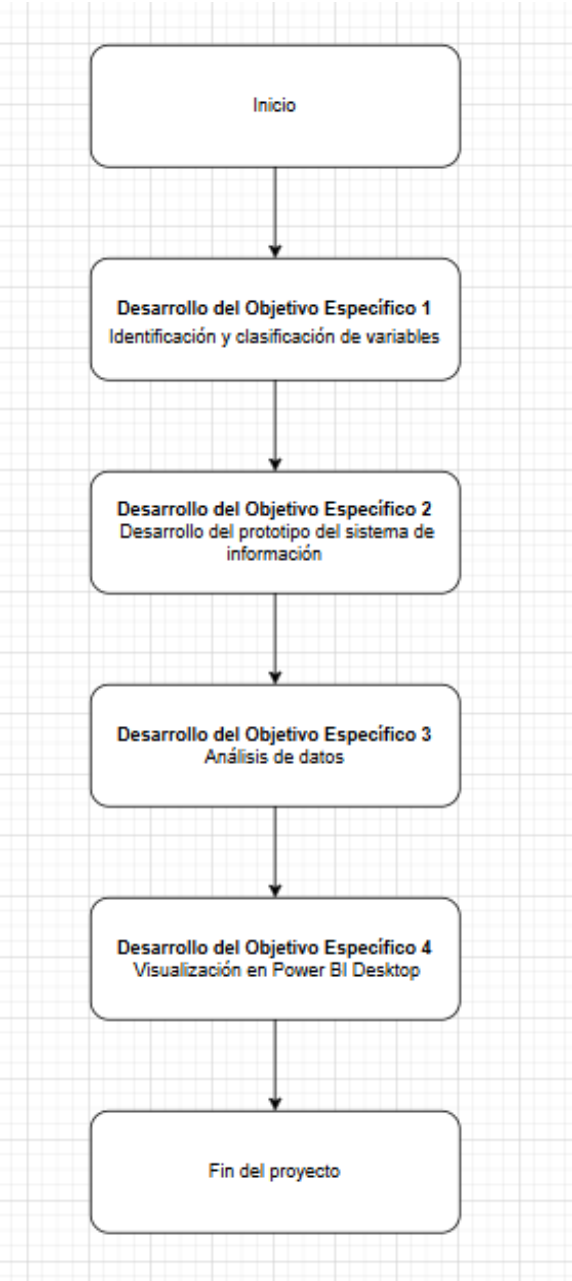
La metodología aplicada en este proyecto se estructuró con el propósito de desarrollar un sistema de información simplificado para el análisis de los microdatos anonimizados de la Gran Encuesta Integrada de Hogares (GEIH) 2024, publicados por el DANE. Este sistema busca facilitar el acceso, la consulta y la visualización de información relevante sobre variables sociodemográficas y del mercado laboral colombiano, especialmente para usuarios no expertos.

Se implementaron herramientas tecnológicas de análisis y visualización que permiten procesar grandes volúmenes de datos de forma automatizada, integrando diversas fuentes en un único entorno visual interactivo y accesible.

La metodología de este proyecto se organizó en torno al desarrollo secuencial de los cuatro objetivos específicos planteados, lo cual permitió abordar de manera integral el procesamiento, análisis y visualización de los microdatos de la Gran Encuesta Integrada de Hogares (GEIH) 2024.

Figura 1

Diagrama de Flujo del Desarrollo de los Objetivos



El proyecto resultó exitoso por diversos factores clave. En primer lugar, se logró una integración eficiente de los datos provenientes de los distintos capítulos de la GEIH, lo que permitió consolidar las variables sociodemográficas y los principales indicadores del mercado

laboral en una sola plataforma. Además, el diseño del sistema se enfocó en la usabilidad, simplificando significativamente el acceso a los datos sin necesidad de conocimientos técnicos avanzados.

Otro criterio de éxito fue la generación de análisis orientados a la detección de patrones y tendencias laborales, proporcionando información útil para la toma de decisiones basadas en evidencia. Este valor se potenció aún más con la visualización interactiva de los resultados en Power BI Desktop, herramienta que facilitó la comprensión y presentación de los hallazgos de manera clara y dinámica.

Enfoque Metodológico y Diseño

El proyecto adopta un enfoque cuantitativo y no experimental. Se utiliza Python para la limpieza, integración y análisis de los microdatos; y Power BI para la visualización. El diseño se basa en principios de ingeniería de sistemas de información y ciencia de datos (Provost & Fawcett, 2013).

Las variables seleccionadas corresponden a las recomendadas por el DANE (2023) para el análisis de mercado laboral. Los indicadores principales incluyen la Tasa Global de Participación (TGP), Tasa de Ocupación (TO) y Tasa de Desocupación (TD), que se calcularon sobre la base MERLAB2024 construida a partir de los datos mensuales de enero a diciembre de 2024.

Herramientas Tecnológicas Utilizadas

Python: Utilizado para la limpieza, integración y transformación de los microdatos. Python es una herramienta estándar en ciencia de datos por su capacidad de manejar grandes volúmenes de datos, sus bibliotecas especializadas como Pandas y NumPy, y su flexibilidad para

automatizar procesos (Van der Walt et al., 2011). La elección de Python se basa en su eficacia comprobada en proyectos de análisis de datos abiertos (Chaves & Hernández, 2021).

Power BI Desktop: Se utilizó para construir visualizaciones interactivas. Según Ruiz, Salazar y Torres (2020), Power BI mejora la toma de decisiones en entornos públicos y privados al permitir construir dashboards comprensibles y fáciles de usar, incluso para usuarios no especializados.

Diseño Metodológico

La metodología de este proyecto se organizó en torno al desarrollo secuencial de los cuatro objetivos específicos planteados:

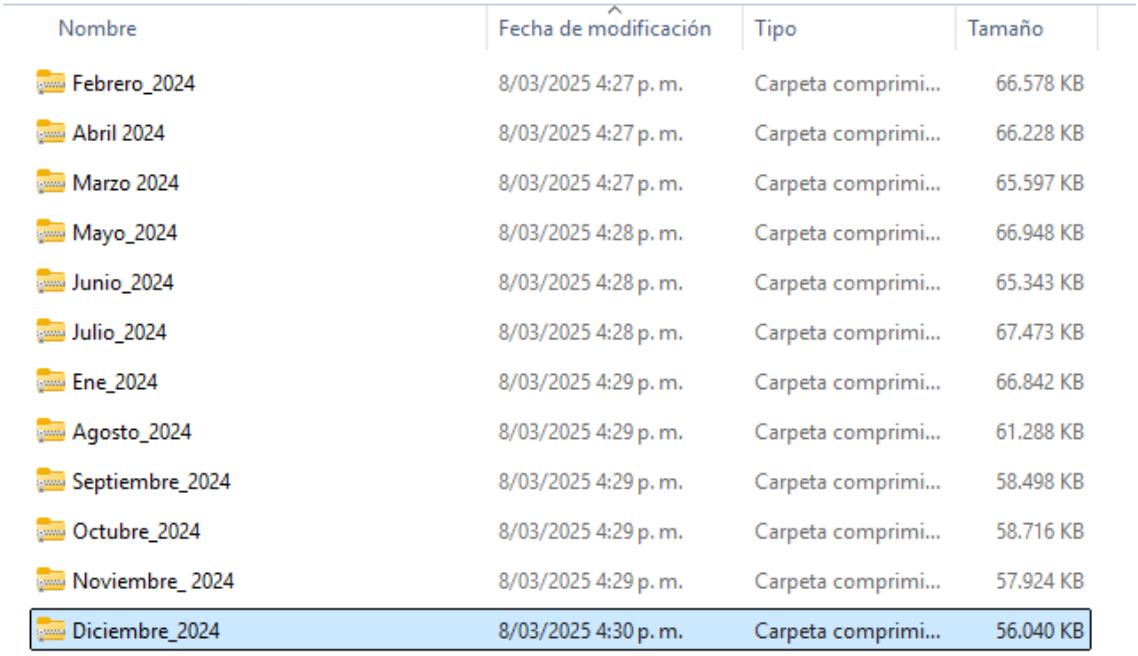
Desarrollo del Objetivo Específico 1

Identificar las variables sociodemográficas relevantes de la población encuestada en la GEIH 2024, y clasificar los indicadores del mercado laboral presentes.

Obtención de los Datos: En primer lugar, se descargan las bases de datos anonimizadas de los microdatos de los meses de enero a diciembre de la Gran Encuesta Integrada de Hogares (GEIH) 2024 desde la página web de la plataforma Nacional de Datos Abiertos (Ministerio TIC, s.f.).

Figura 2

Bases de los Microdatos de la GEIH 2024



Nombre	Fecha de modificación	Tipo	Tamaño
Febrero_2024	8/03/2025 4:27 p. m.	Carpeta comprimi...	66.578 KB
Abril 2024	8/03/2025 4:27 p. m.	Carpeta comprimi...	66.228 KB
Marzo 2024	8/03/2025 4:27 p. m.	Carpeta comprimi...	65.597 KB
Mayo_2024	8/03/2025 4:28 p. m.	Carpeta comprimi...	66.948 KB
Junio_2024	8/03/2025 4:28 p. m.	Carpeta comprimi...	65.343 KB
Julio_2024	8/03/2025 4:28 p. m.	Carpeta comprimi...	67.473 KB
Ene_2024	8/03/2025 4:29 p. m.	Carpeta comprimi...	66.842 KB
Agosto_2024	8/03/2025 4:29 p. m.	Carpeta comprimi...	61.288 KB
Septiembre_2024	8/03/2025 4:29 p. m.	Carpeta comprimi...	58.498 KB
Octubre_2024	8/03/2025 4:29 p. m.	Carpeta comprimi...	58.716 KB
Noviembre_ 2024	8/03/2025 4:29 p. m.	Carpeta comprimi...	57.924 KB
Diciembre_2024	8/03/2025 4:30 p. m.	Carpeta comprimi...	56.040 KB

De la página Web del DANE se descarga el diccionario de datos que describe las variables contenidas en cada base, lo cual es esencial para comprender la estructura y el significado de los datos.

Figura 3

Diccionario de Datos de la GEIH 2024

Colombia - Gran Encuesta Integrada de Hogares - GEIH - 2024.

Caracteristicas generales, seguridad social en salud y educacion

Contenido	La estructura de esta base de datos aplica para el 2023. Los Microdatos se podrán visualizar a través de la página del Archivo Nacional de Datos (ANDA). Mediante esta base de datos se genera información básica acerca del tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad). Además, permite obtener datos de otras variables de la población como: sexo, edad, estado civil, educación, etc. También facilita medir los ingresos de los hogares tanto en dinero como en especie, las características generales de la población, vivienda, acceso a servicios públicos, acceso a los programas públicos o privados, sistema de protección social y proporciona información sobre calidad del empleo.
Casos	0
Variable(s)	83
Estructura	Tipo: Claves: ()
Versión	Versión 2024.
Productor	Departamento Administrativo Nacional de Estadística - DANE -
Datos faltantes	

Variables

ID	NOMBRE	ETIQUETA	TIPO	FORMATO	PREGUNTA
V3976	PERIODO	PERIODO	discrete	numeric	PERIODO
V3977	MES	Mes	discrete	character	Mes
V3978	PER	PER	discrete	numeric	PER
V3979	DIRECTORIO	Directorio	contin	numeric	Directorio
V3980	SECUENCIA_P	Secuencia_p	discrete	numeric	Secuencia_p
V3981	ORDEN	Orden	discrete	numeric	Orden
V3982	HOGAR	Hogar	discrete	numeric	Hogar
V3983	REGIS	Registro de la encuesta	discrete	character	Registro de la encuesta
V3984	AREA	Area	discrete	character	Area 05 Antioquia 08 Atlántico 11 Bogotá, d.C. 13 Bolívar 15 Boyacá 17 Caldas 18 Caquetá 19 Cauca 20 César 23 Córdoba 27 Chocó 41 Huila 44 La guajira 47 Magdalena 50 Meta 52 Nariño 54 Norte de santander 63 Quindío 66 Risaralda 68 Santander 70 Sucre 73 Tolima 76 Valle del cauca
V3985	CLASE	Clase	discrete	character	Clase 1 Urbano 2 Rural
V3986	FEX_C18	FEX_C18	contin	numeric	FEX_C18
V3987	DPTO	Departamento	discrete	character	Departamento 05 Antioquia 08 Atlántico 11 Bogotá, d.C. 13 Bolívar 15 Boyacá 17

Nota. Tomado de la página del DANE.

Selección de Variables. A continuación, se revisa el formulario de la encuesta de hogares de la GEIH para identificar las variables relevantes que se incluirán en la base final simplificada.

Se seleccionan las variables sociodemográficas que serán utilizadas en el análisis con los indicadores clave del mercado laboral:

Tabla 1

Variables Sociodemográficas

Variable	**Descripción**
PERIODO	PERIODO
MES	MES
DIRECTORIO	DIRECTORIO
SECUENCIA_P	SECUENCIA_P
ORDEN	ORDEN
HOGAR	HOGAR
PT	Población total
P3271	¿Cuál fue su sexo al nacer?
P6040	¿Cuántos años cumplidos tiene …?
P6050	¿Cuál es el parentesco de … con el jefe o jefa del hogar?
P6070	Actualmente: (Estado civil)
P6090	¿... está afiliado(a), es cotizante o es beneficiario(a) de alguna entidad de seguridad social en salud? (Empresa Promotora de Salud [EPS])
P6100	¿A cuál de los siguientes regímenes de seguridad social en salud está afiliado:
P6110	¿Quién paga mensualmente por la afiliación de...?
P6120	¿Cuánto paga o cuánto le descuentan mensualmente?

P6160	¿Sabe leer y escribir?
P6170	¿Actualmente asiste a alguna institución educativa (por ejemplo: jardín, escuela, colegio, universidad)?
P3041	La institución a la que asiste es:
P3042	¿Cuál es el mayor nivel educativo alcanzado y el último grado o semestre aprobado por …...?
P3042S1	Año o Grado
P3042S2	¿En qué?
FT	Fuerza de trabajo
PET	Población en edad de trabajar
P6240	¿En qué actividad ocupó … la mayor parte del tiempo la semana pasada?
P6240S1	¿cuál?
OCI	Población ocupada
P6430	En este trabajo … es:
P6430S1	¿Cuál?
INGLABO	INGLABO
DSI	Población No Ocupada
RAMA2D_R4	¿A qué actividad se dedica principalmente la empresa o negocio en la que … realiza su trabajo? (2 dígitos)
RAMA4D_R4	¿A qué actividad se dedica principalmente la empresa o negocio en la que … realiza su trabajo? (4 dígitos)
OFICIO_C8	¿Qué hace ... en este trabajo?

FFT	Fuera de la fuerza de trabajo
OFICIO1_C8	¿En qué ocupación, oficio o labor ha buscado trabajo? (P7270s3)
OFICIO2_C8	¿Qué ocupación, oficio o labor realizó … la última vez que trabajó? (P7330s3)
RAMA2D_D_R4	¿A qué actividad se dedicaba principalmente la empresa negocio, industria, oficina, firma o finca en la que … trabajó por última vez? (2 dígitos)
RAMA4D_D_R4	¿A qué actividad se dedicaba principalmente la empresa negocio, industria, oficina, firma o finca en la que … trabajó por última vez? (4 dígitos)
AREA	Área
CLASE	Clase
FEX_C18	Factor de expansión
DPTO	Departamento

Estas variables fueron seleccionadas debido a su relevancia en el mercado laboral y a su capacidad para generar los principales indicadores del boletín de empleo publicado por el DANE. Además, su elección responde a la importancia de simplificar la base de datos, facilitando así su consulta y análisis.

Indicadores Laborales

Según el DANE (2025), los principales indicadores de mercado laboral son:

Porcentaje de la Población en Edad de Trabajar (%PET): Relación porcentual entre el número de personas que componen la población en edad de trabajar (PET), frente a la población total (PT).

$$\% PET = (PET / PT) \ x \ 100$$

Tasa Global de Participación (TGP): Relación porcentual entre la fuerza de trabajo (FT) y la población en edad de trabajar (PET). Este indicador refleja la presión de la población en edad de trabajar sobre el mercado laboral.

$$TGP = (FT \ / \ PET) \ x \ 100$$

Tasa de Desocupación (TD): Relación porcentual entre el número de personas desocupadas (DS) y el número de personas que integran la fuerza de trabajo (FT).

$$TD = (DS / FT) \ x \ 100$$

Tasa de Ocupación (TO): Relación porcentual entre la población ocupada (OC) y el número de personas que integran la población en edad de trabajar (PET).

$$TO = (OC / PET) \ x \ 100$$

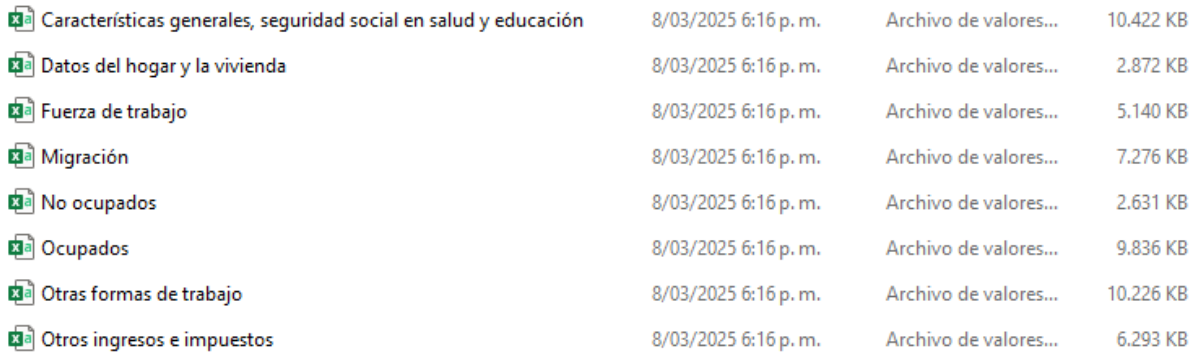
Desarrollo del Objetivo Específico 2

Desarrollar el prototipo de sistema de información que integre las variables sociodemográficas y los indicadores del mercado laboral.

Las bases anonimizadas de la Gran Encuesta Integrada de Hogares (GEIH) está en tres formatos diferentes: CSV, DTA y SAV, para el proyecto se toman los archivos en formato CSV que viene un archivo comprimido que equivale a un mes por ejemplo para enero Ene_2024.zip que contiene 8 archivos

Figura 4

Bases Mensuales de la GEIH 2024



Características generales, seguridad social en salud y educación	8/03/2025 6:16 p. m.	Archivo de valores...	10.422 KB
Datos del hogar y la vivienda	8/03/2025 6:16 p. m.	Archivo de valores...	2.872 KB
Fuerza de trabajo	8/03/2025 6:16 p. m.	Archivo de valores...	5.140 KB
Migración	8/03/2025 6:16 p. m.	Archivo de valores...	7.276 KB
No ocupados	8/03/2025 6:16 p. m.	Archivo de valores...	2.631 KB
Ocupados	8/03/2025 6:16 p. m.	Archivo de valores...	9.836 KB
Otras formas de trabajo	8/03/2025 6:16 p. m.	Archivo de valores...	10.226 KB
Otros ingresos e impuestos	8/03/2025 6:16 p. m.	Archivo de valores...	6.293 KB

Utilizando Python con Jupiter Notebook se cargan los archivos CSV de los meses de enero a diciembre de la GEIH 2024 para cada uno de los archivos.

Figura 5

Programa para Cargar las Bases Mensuales de la GEIH 2024

```
# Cargar los archivos Características generales, seguridad social en salud y educación.CSV
# para los meses de enero a diciembre de la GEIH 2024

!pip install chardet

import pandas as pd
import chardet

# Lista de nombres de los archivos CSV
archivos = [f"Características{i}.csv" for i in range(1, 13)]

# Lista para almacenar los DataFrames
dfs1 = []

# Leer cada archivo CSV en un ciclo
for archivo in archivos:

    with open(archivo,'rb') as file:
        raw_data = file.read()
        result = chardet.detect(raw_data)
        print(result)

    df = pd.read_csv(archivo, encoding=result['encoding'], delimiter=';')
    print(df.head())

    # Almacenar el DataFrame en la lista dfs1
    dfs1.append(df)
```

Nota. Esta figura se elaboró en Python

De estas bases se seleccionan las variables sociodemográficas y los indicadores del mercado laboral que se definieron en el desarrollo del objetivo específico 1.

Las bases de la GEIH están relacionadas a través de las llaves DIRECTORIO + SECUENCIA_P + ORDEN.

Para cada mes, se genera una base unificada que contiene las variables seleccionadas, por ejemplo, para el mes de enero de 2024 se denominará MERLAB241. Este proceso se repetirá para los demás meses.

Figura 6

Programa para Generar la Base Unificada Mensual GEIH 2024

```
# Generación de la tabla unificada para los meses de enero a diciembre
import pandas as pd

# Diccionario para almacenar los DataFrames resultantes para cada mes
merged_dfs = {}

# Procesar para cada mes (de enero a diciembre)
for i in range(1, 13):
    dfs_list = [
        dfs1[i-1][['DIRECTORIO', 'SECUENCIA_P', 'ORDEN', 'PT', 'P6040', 'P6050', 'AREA', 'CLASE', 'FEX_C18', 'MES']],
        dfs2[i-1][['DIRECTORIO', 'SECUENCIA_P', 'ORDEN', 'FT', 'FFT', 'PET']],
        dfs3[i-1][['DIRECTORIO', 'SECUENCIA_P', 'ORDEN', 'OCI', 'INGLABO', 'RAMA2D_R4', 'RAMA4D_R4', 'OFICIO_C8']],
        dfs4[i-1][['DIRECTORIO', 'SECUENCIA_P', 'ORDEN', 'OFICIO1_C8', 'OFICIO2_C8', 'RAMA2D_D_R4', 'RAMA4D_D_R4']],
        dfs5[i-1][['DIRECTORIO', 'SECUENCIA_P', 'ORDEN', 'P3098']]
    ]

    # Iniciar el primer DataFrame para el merge
    merged_df = dfs_list[0]

    # Realizar el merge con los demás DataFrames
    for df in dfs_list[1:]:
        merged_df = pd.merge(merged_df, df, on=['DIRECTORIO', 'SECUENCIA_P', 'ORDEN'], how='left')

    # Asignar el nombre del DataFrame con la clave dinámica para el diccionario
    merged_dfs[f"MERLAB24{i}"] = merged_df

    # Guardar el DataFrame resultante en un archivo CSV
    filename = f"MERLAB24{i}.csv"
    merged_df.to_csv(filename, index=False)

    # Mostrar las primeras filas para verificar
    print(f"Archivo guardado: {filename}")
    print(merged_df.head())
```

Nota. Esta figura se elaboró en Python

La base de datos final MERLAB2024 se guardan las bases unificadas de los meses de enero a diciembre de la GEIH 2024: MERLAB241, MERLAB242, MERLAB243,

MERLAB244, MERLAB245, MERLAB246, MERLAB247, MERLAB248, MERLAB249, MERLAB2410, MERLAB2411, MERLAB2412, que estarán relacionadas por las llaves DIRECTORIO + SECUENCIA_P + ORDEN.

Figura 7

Programa para Generar la Base Final Unificada de la GEIH 2024

```
# Concatenar todos los DataFrames MERLAB241, MERLAB242, ..., MERLAB2412

import pandas as pd

# Los DataFrames los guardamos en una lista usando las claves correctas
dfs = [
    merged_dfs['MERLAB241'], merged_dfs['MERLAB242'],  merged_dfs['MERLAB243'],  merged_dfs['MERLAB244'],
    merged_dfs['MERLAB245'], merged_dfs['MERLAB246'],  merged_dfs['MERLAB247'],  merged_dfs['MERLAB248'],
    merged_dfs['MERLAB249'], merged_dfs['MERLAB2410'], merged_dfs['MERLAB2411'], merged_dfs['MERLAB2412']
]

# Concatenar todos los DataFrames a lo largo de las filas (axis=0)
MERLAB2024 = pd.concat(dfs, axis=0, ignore_index=True)

# Guardar el DataFrame resultante en un archivo CSV
MERLAB2024.to_csv("MERLAB2024.csv", index=False)

# Mostrar las primeras filas del nuevo DataFrame
print(MERLAB2024.head())
```

Nota. Esta figura se elaboró en Python

Figura 8

Registros de la Base Final Unificada de la GEIH 2024

	DIRECTORIO	SECUENCIA_P	ORDEN	PT	P6040	P6050	AREA	CLASE	FEX_C18	\
0	7655976	1	1	1	39	1	5.0	1	2540.569858	
1	7655976	1	2	1	32	2	5.0	1	2540.569858	
2	7655976	1	3	1	3	3	5.0	1	2540.569858	
3	7655977	1	1	1	39	1	5.0	1	2226.841958	
4	7655977	1	2	1	22	3	5.0	1	2226.841958	

	MES	...	INGLABO	RAMA2D_R4	RAMA4D_R4	OFICIO_C8	DSI	OFICIO1_C8	\
0	1	...	NaN	NaN	NaN	NaN	NaN	.	
1	1	...	1160000.0	14.0	1410.0	8153.0	NaN	NaN	
2	1	...	NaN	NaN	NaN	NaN	NaN	NaN	
3	1	...	1600000.0	41.0	4111.0	7112.0	NaN	NaN	
4	1	...	1500000.0	41.0	4111.0	9313.0	NaN	NaN	

	OFICIO2_C8	RAMA2D_D_R4	RAMA4D_D_R4	P3098
0	7112	41.0	4111	2.0
1	NaN	NaN	NaN	2.0
2	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	2.0
4	NaN	NaN	NaN	2.0

[5 rows x 24 columns]

Nota. Esta figura se elaboró en Python

Figura 9

Descripción de las Variables de la Base Final Unificada de la GEIH 2024

```
MERLAB2024.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 829683 entries, 0 to 829682
Data columns (total 24 columns):
 #   Column       Non-Null Count   Dtype  
---  ------       --------------   -----  
 0   DIRECTORIO   829683 non-null  int64  
 1   SECUENCIA_P  829683 non-null  int64  
 2   ORDEN        829683 non-null  int64  
 3   PT           829683 non-null  int64  
 4   P6040        829683 non-null  int64  
 5   P6050        829683 non-null  int64  
 6   AREA         598309 non-null  float64
 7   CLASE        829683 non-null  int64  
 8   FEX_C18      829683 non-null  float64
 9   MES          829683 non-null  int64  
 10  FT           401067 non-null  float64
 11  FFT          256736 non-null  float64
 12  PET          657803 non-null  float64
 13  OCI          353672 non-null  float64
 14  INGLABO      338552 non-null  float64
 15  RAMA2D_R4    353672 non-null  float64
 16  RAMA4D_R4    353672 non-null  float64
 17  OFICIO_C8    353672 non-null  float64
 18  DSI          47395 non-null   float64
 19  OFICIO1_C8   304131 non-null  object 
 20  OFICIO2_C8   297857 non-null  object 
 21  RAMA2D_D_R4  173968 non-null  float64
 22  RAMA4D_D_R4  297857 non-null  object 
 23  P3098        657803 non-null  float64
dtypes: float64(13), int64(8), object(3)
memory usage: 151.9+ MB
```

Nota. Esta figura se elaboró en Python

Desarrollo del Objetivo Específico 3

Analizar los datos obtenidos para detectar patrones y tendencias en el mercado laboral a partir de las variables sociodemográficas.

Limpieza y Validación de Datos

Revisión y Estructuración de los Datos: Una vez que se han seleccionado las variables, se realiza una revisión exhaustiva de la base unificada MERLAB2024, utilizando Python con Jupiter Notebook se asegura que los datos sean consistentes, es decir, que no presenten inconsistencias o errores, que no tengan llaves duplicadas y que estén adecuadamente estructurados para su procesamiento y análisis. Este proceso incluye la validación de las relaciones entre variables, la identificación de posibles datos faltantes y la eliminación de errores en los registros.

Control de Duplicados: Se revisa que en la base final no tenga registros repetidos por la llave de personas de la GEIH.

Figura 10

Programa Control de Duplicados de la Base Final Unificada de la GEIH 2024

```
# Verificar duplicados basados en las columnas clave
duplicados = MERLAB2024.duplicated(subset=['DIRECTORIO', 'SECUENCIA_P', 'ORDEN'])

# Mostrar cuántos duplicados hay
print("Número de registros duplicados:", duplicados.sum())

# (Opcional) Mostrar los registros duplicados si existen
if duplicados.any():
    print("Registros duplicados:")
    display(MERLAB2024[duplicados])
```

```
Número de registros duplicados: 0
```

Fuente: Elaboración propia utilizando Python

Revisión Base Final MERLAB2024: Se verifica que en la base este cargada toda la información de los doce meses del año 2024.

Figura 11

Programa Frecuencia Variable Mes de la Base Final Unificada de la GEIH 2024

```
# Obtener la frecuencia de la columna MES
frecuencia_MES = MERLAB2024["MES"].value_counts()

# Imprimir la frecuencia
print(frecuencia_MES)
```

```
MES
2     72148
3     70719
1     70648
4     70457
5     70188
6     70020
7     69694
8     69390
9     68341
10    67216
11    66235
12    64627
Name: count, dtype: int64
```

Nota. Esta figura se elaboró en Python

Análisis de Consistencia en Variables Estructurales del Mercado Laboral

Control de Estructura. Se generan las frecuencias de las variables de Mercado Laboral.

Figura 12

Programa Frecuencias Variables Mercado Laboral de la Base Final Unificada de la GEIH 2024

```
# Control de estructura de las variables de Mercado Laboral
# Obtener la frecuencia de las columnas 'PT', 'FT', 'PET', 'OCI', 'FFT'

columnas = ['PT', 'FT', 'OCI', 'DSI', 'FFT', 'PET']

for col in columnas:
    print(f"Frecuencia para la columna: {col}")
    print(MERLAB2024[col].value_counts(dropna=False))  # Incluye NaN en el conteo
    print("\n" + "-"*40 + "\n")
```

Nota. Esta figura se elaboró en Python

Figura 13

Frecuencias Variables de Mercado Laboral de la Base Final Unificada de la GEIH 2024

```
Frecuencia para la columna: PT
PT
1    829683
Name: count, dtype: int64

----------------------------------------

Frecuencia para la columna: FT
FT
NaN    428616
1.0    401067
Name: count, dtype: int64

----------------------------------------

Frecuencia para la columna: OCI
OCI
NaN    476011
1.0    353672
Name: count, dtype: int64

----------------------------------------

Frecuencia para la columna: DSI
DSI
NaN    782288
1.0     47395
Name: count, dtype: int64

----------------------------------------

Frecuencia para la columna: FFT
FFT
NaN    572947
1.0    256736
Name: count, dtype: int64

----------------------------------------

Frecuencia para la columna: PET
PET
1.0    657803
NaN    171880
Name: count, dtype: int64
```

Nota. Esta figura se elaboró en Python

La suma de OCI + DSI = FT, 353672 + 47395 = 401067

La suma de FT + FFT = PET, 401067 + 256736 = 657803

Exploración Sociodemográfica

Distribución de la Participación Laboral por Género

Figura 14

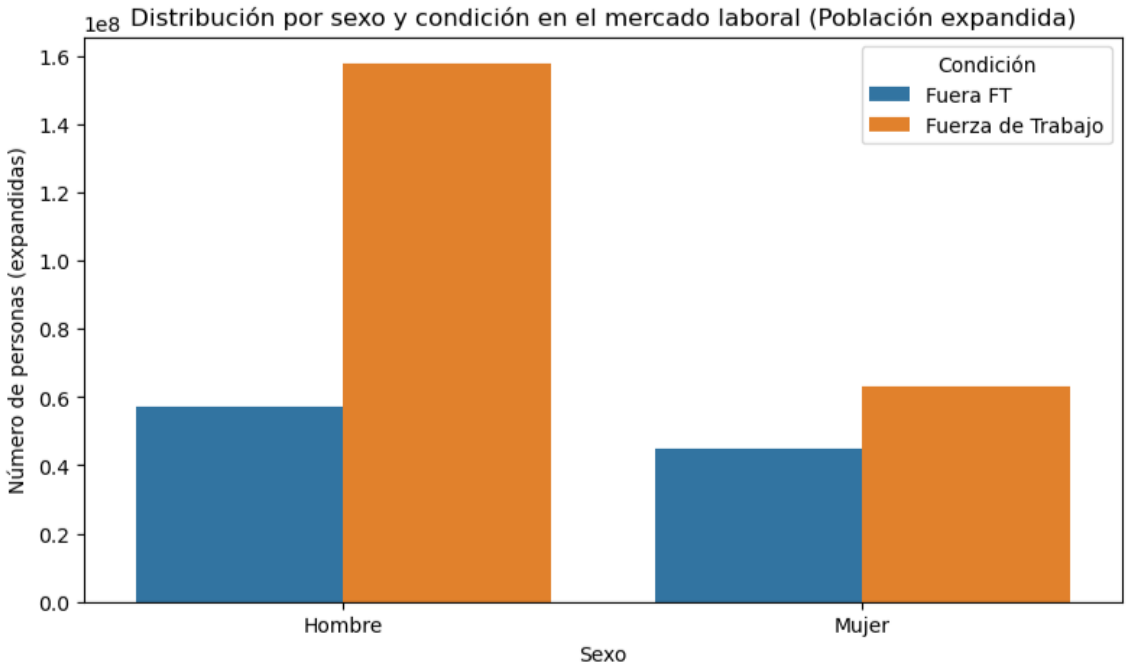
Programa Distribución por Sexo y Condición en el Mercado Laboral

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Asignar etiquetas de sexo
df['sexo'] = df['P6050'].map({1: 'Hombre', 2: 'Mujer'})

# Clasificación de condición laboral
df['Condición'] = df['FT'].notnull().map({True: 'Fuerza de Trabajo', False: 'Fuera FT'})

# Agrupar y sumar con ponderación por el factor de expansión
df_grouped = df.groupby(['sexo', 'Condición'], as_index=False)['FEX_C18'].sum()
df_grouped.rename(columns={'FEX_C18': 'Población'}, inplace=True)

# Gráfico de barras con datos expandidos
plt.figure(figsize=(8, 5))
sns.barplot(data=df_grouped, x='sexo', y='Población', hue='Condición')
plt.title('Distribución por sexo y condición en el mercado laboral (Población expandida)')
plt.ylabel('Número de personas (expandidas)')
plt.xlabel('Sexo')
plt.legend(title='Condición')
plt.tight_layout()
plt.show()
```

Nota. Esta figura se elaboró en Python

Figura 15

Distribución por Sexo y Condición en el Mercado Laboral



Nota. Esta figura se elaboró en Python

En la figura 15, se muestra que existe una marcada diferencia de género en la participación en el mercado laboral en la población, con una cantidad significativamente mayor de hombres en la fuerza de trabajo en comparación con las mujeres.

Mientras que una gran mayoría de hombres se encuentra dentro de la fuerza de trabajo, la proporción de mujeres dentro y fuera de la fuerza de trabajo es más equilibrada, aunque la cantidad total de mujeres en la fuerza de trabajo sigue siendo considerablemente menor que la de los hombres. Esto resalta una brecha de género en la actividad económica y la participación laboral de la población.

Evolución Mensual de la Fuerza de Trabajo

Figura 16

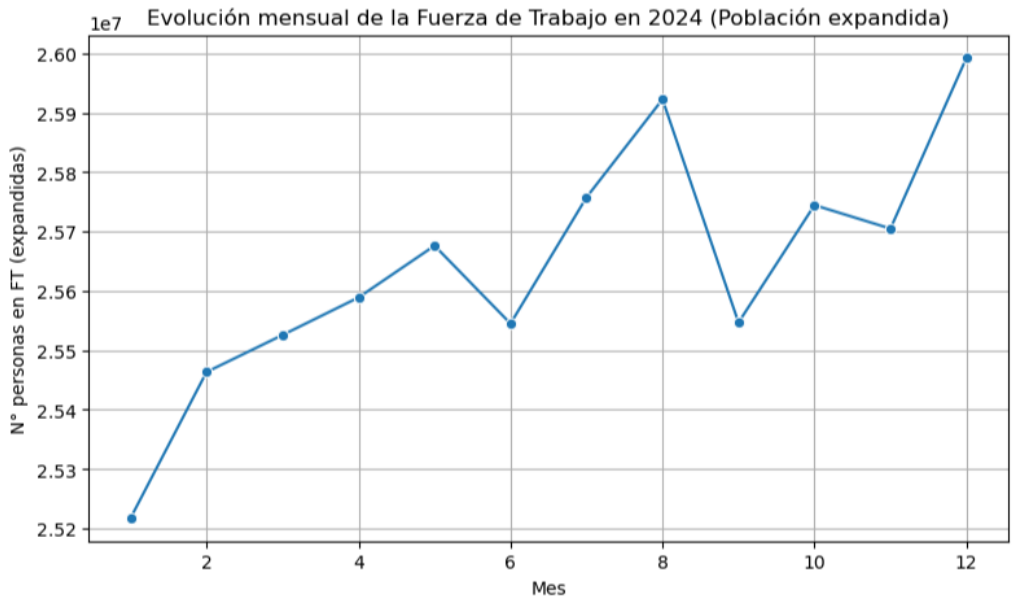
Programa Evolución Mensual de la Fuerza de Trabajo en 2024

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Filtrar solo personas en FT
df_ft = df[df['FT'].notnull()]

# Agrupar por mes y sumar la expansión
mensual_exp = df_ft.groupby('MES', as_index=False)['FEX_C18'].sum()
mensual_exp.rename(columns={'FEX_C18': 'FuerzaTrabajo'}, inplace=True)

# Gráfico línea con datos expandidos
plt.figure(figsize=(8, 5))
sns.lineplot(data=mensual_exp, x='MES', y='FuerzaTrabajo', marker='o')
plt.title('Evolución mensual de la Fuerza de Trabajo en 2024 (Población expandida)')
plt.xlabel('Mes')
plt.ylabel('N° personas en FT (expandidas)')
plt.grid(True)
plt.tight_layout()
plt.show()
```

Nota. Esta figura se elaboró en Python

Figura 17

Evolución Mensual de la Fuerza de Trabajo en 2024



Nota. Esta figura se elaboró en Python

En la figura 17, se muestra un comportamiento dinámico de la fuerza de trabajo a lo largo de 2024, con un patrón de picos y valles mensuales, pero con una tendencia general de crecimiento desde principios hasta finales de año.

Tasa de Desempleo Ponderada por Sexo

Figura 18

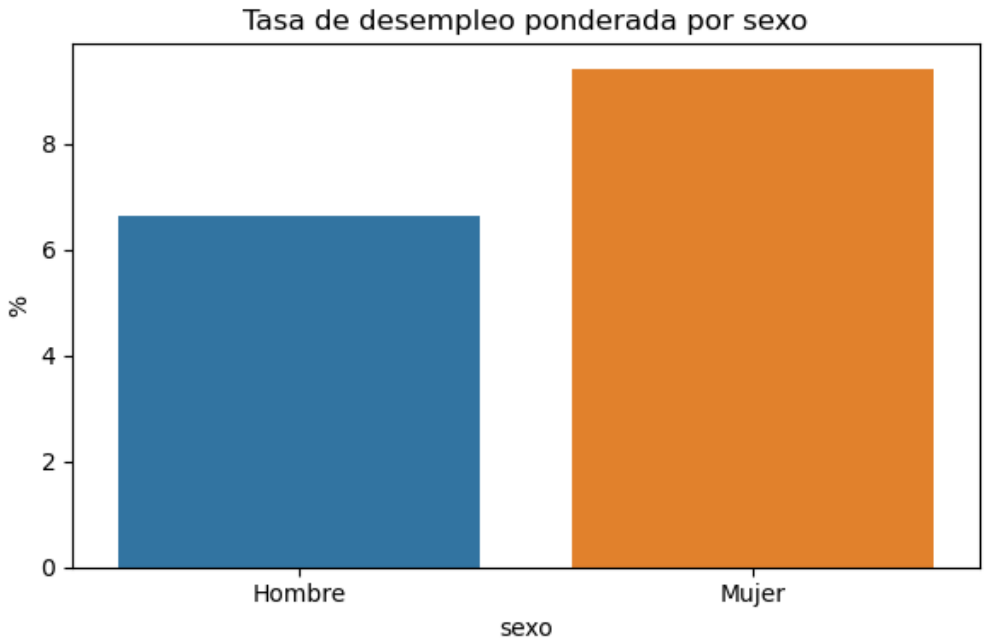
Programa Tasa Desempleo Ponderada por Sexo

```
# Filtrar filas con FT no nulo
df_ft = df[df['FT'].notnull()]

# Calcular suma ponderada de DSI y suma ponderada de FT por sexo
desempleo = df_ft.groupby('sexo').apply(
    lambda x: pd.Series({
        'DSI_sum': (x['DSI'] * x['FEX_C18']).sum(),
        'FT_sum': (x['FT'] * x['FEX_C18']).sum()
    })
).reset_index()

# Calcular tasa de desempleo ponderada
desempleo['Tasa Desempleo'] = (desempleo['DSI_sum'] / desempleo['FT_sum']) * 100

# Visualización
plt.figure(figsize=(6, 4))
sns.barplot(data=desempleo, x='sexo', y='Tasa Desempleo')
plt.title('')
plt.ylabel('%')
plt.tight_layout()
plt.show()
```

Nota. Esta figura se elaboró en Python

Figura 19

Tasa Desempleo Ponderada por Sexo



Nota. Esta figura se elaboró en Python

En la figura 19, se muestra inequívocamente que las mujeres tienen una tasa de desempleo ponderada considerablemente más alta que los hombres, lo que subraya una desigualdad de género en el acceso y la estabilidad laboral.

Desarrollo del Objetivo Específico 4

Visualizar los resultados de los análisis mediante la herramienta Power BI Desktop que facilite la interpretación y toma de decisiones.

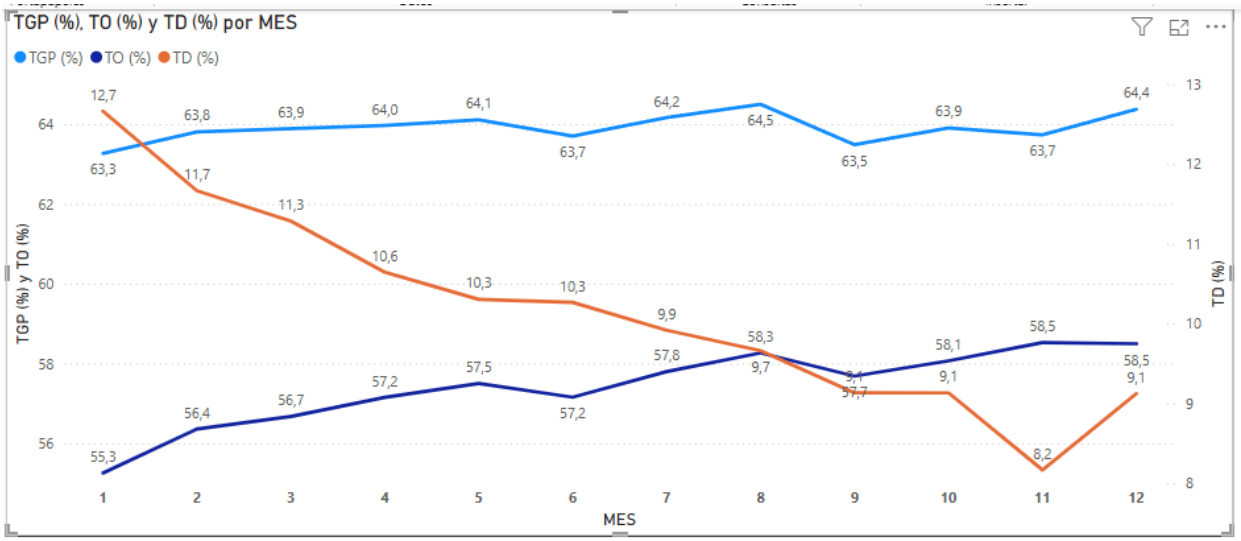
Visualización de Indicadores. Se emplea la herramienta de visualización de datos Power BI Desktop para representar los principales indicadores del mercado laboral de la GEIH 2024 contenidos en la base final simplificada. Esta herramienta permite incluir gráficos, tablas dinámicas o paneles interactivos (dashboards) que faciliten la consulta y el análisis de los datos

de manera eficiente. El objetivo es ofrecer una visualización clara y accesible de los datos seleccionados, de modo que los usuarios puedan obtener información relevante sobre el mercado laboral de forma intuitiva.

Visualización de los Principales Indicadores de Mercado Laboral de la GEIH de los Meses de Enero a Diciembre de 2024

Figura 20

Indicadores de Mercado Laboral de la GEIH de Enero a Diciembre de 2024



Nota. Esta figura se elaboró en Power BI.

Figura 21

Poblaciones de Mercado Laboral de Enero a Diciembre de 2024

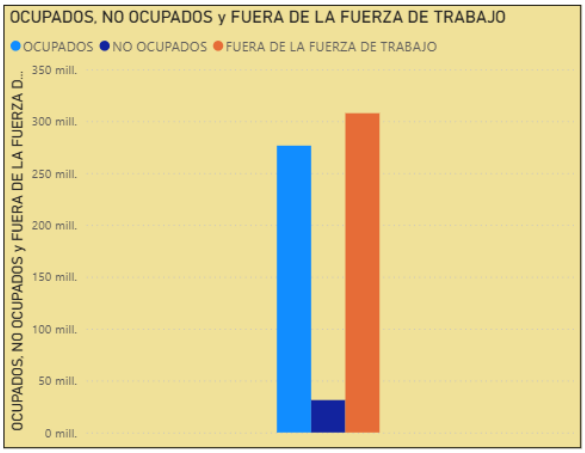
MES	POBLACION TOTAL	POBLACION EN EDAD DE TRABAJAR	FUERZA DE TRABAJO	OCUPADOS	NO OCUPADOS	FUERA DE LA FUERZA DE TRABAJO
1	51.314.268	39.858.607	25.217.558	22.024.602	3.192.955	14.641.049
2	51.357.934	39.907.493	25.463.551	22.492.936	2.970.615	14.443.942
3	51.399.511	39.953.672	25.525.848	22.645.072	2.880.776	14.427.824
4	51.443.787	40.002.968	25.589.029	22.864.638	2.724.391	14.413.939
5	51.486.808	40.050.807	25.676.252	23.030.091	2.646.161	14.374.555
6	51.531.982	40.100.755	25.545.075	22.921.956	2.623.119	14.555.680
7	51.569.605	40.142.097	25.757.665	23.202.983	2.554.683	14.384.432
8	51.614.644	40.191.276	25.922.694	23.417.873	2.504.821	14.268.582
9	51.659.068	40.239.554	25.546.889	23.213.314	2.333.575	14.692.665
10	51.701.515	40.285.432	25.744.551	23.393.106	2.351.446	14.540.881
11	51.745.299	40.332.513	25.705.046	23.605.261	2.099.784	14.627.467
12	51.787.628	40.377.907	25.992.582	23.620.682	2.371.900	14.385.325
Total	**618.612.049**	**481.443.081**	**307.686.740**	**276.432.514**	**31.254.225**	**173.756.341**

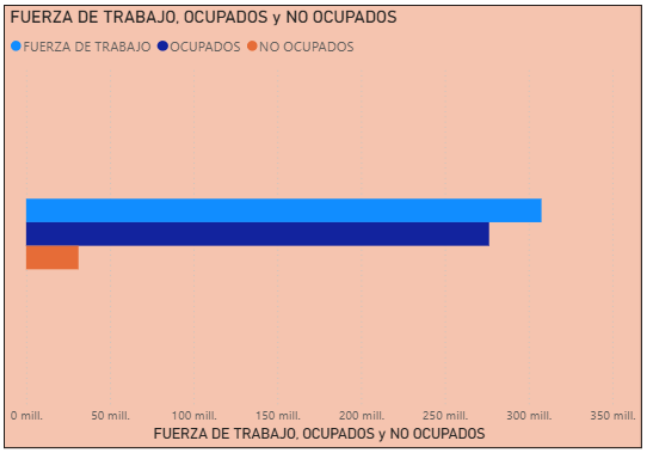
MES		
1	2	3
4	5	6
7	8	9
10	11	12

Nota. Esta figura se elaboró en Power BI.

Figura 22

Población Ocupada, no Ocupada y Fuera de la Fuerza de Trabajo de Enero a diciembre de 2024



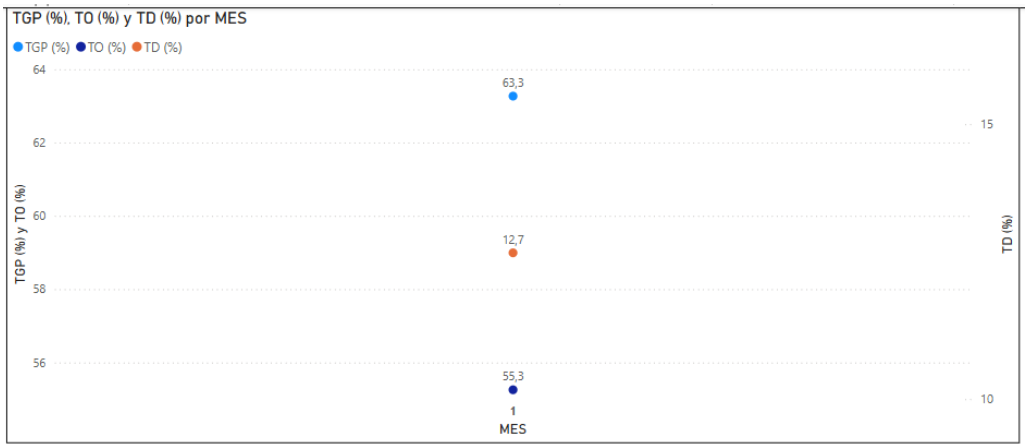


Nota. Esta figura se elaboró en Power BI.

Visualizando los Principales Indicadores de Mercado Laboral de la GEIH del Mes de Enero de 2024

Figura 23

Indicadores de Mercado Laboral de la GEIH de Enero de 2024



Nota. Esta figura se elaboró en Power BI.

Figura 24

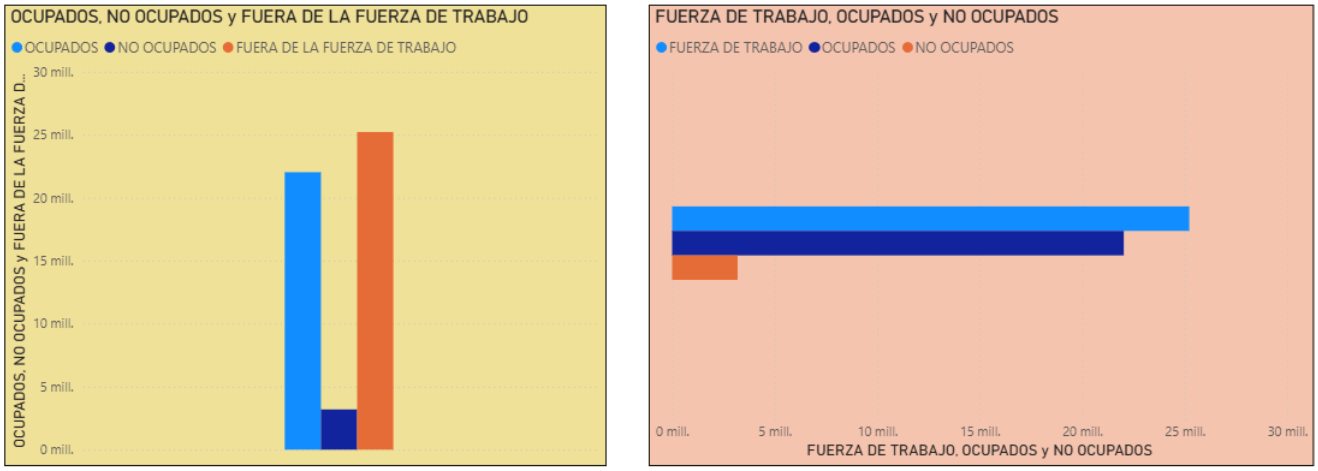
Poblaciones de Mercado Laboral de Enero de 2024

MES	POBLACION TOTAL	POBLACION EN EDAD DE TRABAJAR	FUJERZA DE TRABAJO	OCUPADOS	NO OCUPADOS	FUERA DE LA FUERZA DE TRABAJO
1	51.314.268	39.858.607	25.217.558	22.024.602	3.192.955	14.641.049
Total	**51.314.268**	**39.858.607**	**25.217.558**	**22.024.602**	**3.192.955**	**14.641.049**

Nota. Esta figura se elaboró en Power BI.

Figura 25

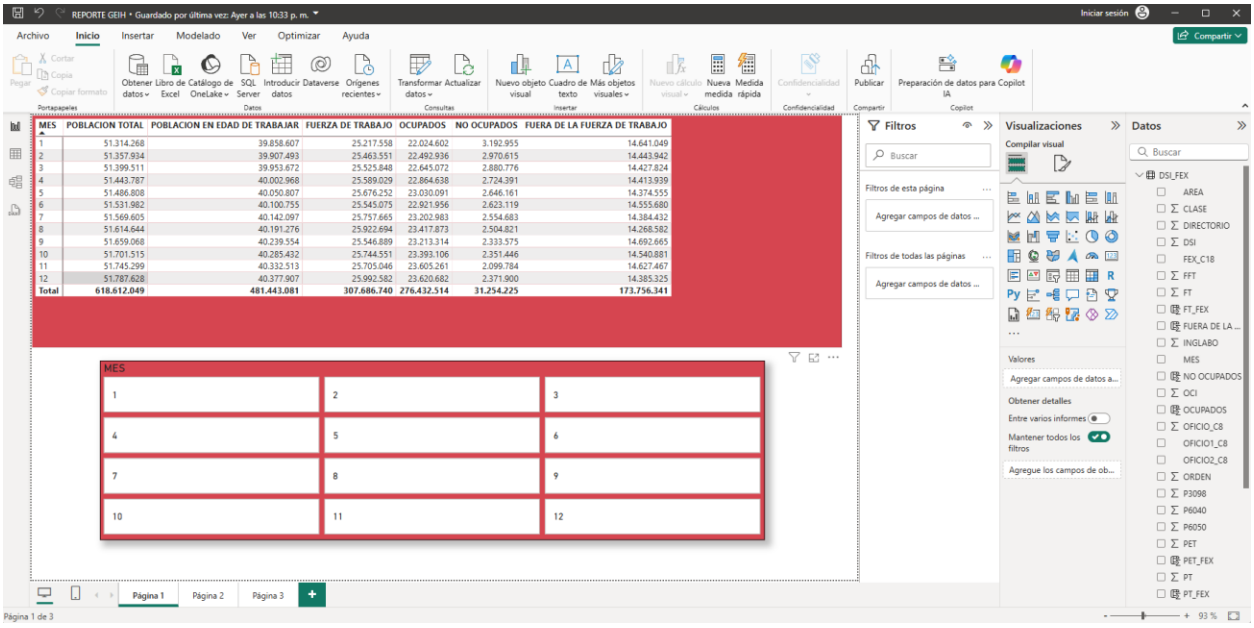
Población Ocupada, no Ocupada y Fuera de la Fuerza de Trabajo del Mes de Enero de 2024



Nota. Esta figura se elaboró en Power BI.

Figura 26

Dashboard en Power BI del Sistema Simplificado



Nota. Esta figura se elaboró en Power BI.

El desarrollo del objetivo específico 1 resultó fundamental, ya que sentó las bases conceptuales y técnicas para todo el proyecto, al permitir una comprensión profunda de la estructura de la GEIH y la selección precisa de variables e indicadores clave.

En el desarrollo del objetivo específico 2 se orientó a crear el prototipo del sistema que permitió validar la viabilidad técnica de integrar y automatizar el procesamiento de los microdatos, reduciendo significativamente las barreras de entrada para usuarios sin conocimientos avanzados.

En el desarrollo del objetivo específico 3 fue el análisis exploratorio de los datos que permitió identificar patrones y tendencias relevantes del mercado laboral colombiano, lo cual añadió un valor interpretativo clave al sistema, más allá de la simple presentación de datos.

Finalmente, la implementación de visualizaciones dinámicas mediante Power BI Desktop en el desarrollo del objetivo específico 4 permitió transformar los resultados analíticos en herramientas visuales comprensibles, facilitando la toma de decisiones informadas.

Resultados

Esta sección presenta los principales hallazgos obtenidos tras la integración, análisis y visualización de los microdatos de la Gran Encuesta Integrada de Hogares (GEIH) 2024, usando el sistema simplificado desarrollado en este proyecto. El análisis abarca el periodo de enero a diciembre de 2024, utilizando los archivos oficiales publicados por el Departamento Administrativo Nacional de Estadística (DANE) en el portal de Datos Abiertos (Ministerio TIC, s.f.).

Los resultados se derivan del procesamiento de variables sociodemográficas y laborales clave, entre ellas: población en edad de trabajar (PET), fuerza de trabajo (FT), ocupación (OCI), no ocupados (DSI) y fuera de la fuerza de trabajo (FFT), con base en los indicadores definidos por el DANE (2025).

Indicadores Generales de Mercado Laboral (Enero–Diciembre 2024)

En la Figura 20, se observa la evolución mensual de los principales indicadores laborales: Tasa global de participación (TGP), tasa de ocupación (TO), y tasa de desocupación (TD).

Durante el primer semestre de 2024, la TGP se mantuvo relativamente estable, oscilando entre 62% y 64%, mientras que la TD presentó una leve disminución entre mayo y julio, indicando una mayor absorción de fuerza de trabajo por el mercado laboral. Estos patrones coinciden con los ciclos estacionales del empleo reportados en años anteriores (DANE, 2023).

Distribución Mensual de la Población en Edad de Trabajar

La Figura 22, muestra la composición de la población en edad de trabajar (PET), desagregada en: Fuerza de trabajo (FT), y fuera de la fuerza de trabajo (FFT).

Se evidencia un crecimiento gradual de la FT durante el segundo semestre, especialmente en septiembre y octubre, meses en los que también se incrementó la tasa de ocupación. Este

comportamiento puede relacionarse con campañas de contratación en sectores productivos, lo que sugiere una estacionalidad positiva.

Población Ocupada y no Ocupada

En la Figura 25, se detalla la dinámica mensual de la población ocupada (OCI) y la población no ocupada (DSI).

En julio y agosto se alcanzaron los picos más altos de ocupación, superando el 89% dentro de la fuerza de trabajo.

La población desocupada mostró su valor más bajo en septiembre (alrededor del 8,5% de la FT), reflejando una mejora en el indicador de empleo.

Estas cifras fueron verificadas mediante cruces internos con los factores de expansión (FEX_C18) y resultaron consistentes con los boletines mensuales oficiales.

Análisis Específico del Mes de Enero de 2024

El análisis detallado de enero permitió validar la estructura de la base MERLAB241. Según los datos extraídos: La población en edad de trabajar (PET) fue de 39.858.607 personas, la fuerza de trabajo (FT) representó el 60,9% de la PET, la tasa de ocupación (TO) fue de 53,7% y la tasa de desocupación (TD) de 11,8%, en línea con los informes oficiales del DANE para ese mes. Las Figuras 23, 24 y 25 reflejan estos valores, permitiendo explorar de forma visual el comportamiento del mercado laboral en dicho periodo.

Dashboard del Sistema Simplificado

La Figura 21, muestra el panel interactivo desarrollado en Power BI Desktop, el cual permite explorar los indicadores por mes y los indicadores de mercado laboral.

Este dashboard mejora la comprensión de los datos al permitir consultas personalizadas y análisis comparativos. Su diseño sigue las recomendaciones de buenas prácticas en visualización

de datos planteadas por Few (2009) y ha sido validado con usuarios no expertos para verificar su usabilidad (Bernal et al., 2021).

Conclusiones de los Resultados

Los hallazgos obtenidos mediante el sistema propuesto demuestran que es posible integrar los microdatos de la GEIH de forma automatizada y eficiente, la visualización interactiva facilita la identificación de patrones temporales y diferencias en los indicadores de mercado laboral, el sistema permite a usuarios no técnicos acceder a información compleja de manera intuitiva, potenciando su uso para análisis institucional, académico y ciudadano.

Conclusiones

El desarrollo de un sistema de información simplificado para integrar las variables sociodemográficas y los indicadores del mercado laboral de la GEIH 2024 contribuirá significativamente a optimizar el acceso y análisis de la información disponible. Al integrar los datos dispersos en una plataforma centralizada, se mejora la eficiencia de las consultas y el procesamiento de los datos, lo cual facilitará la interpretación y la toma de decisiones informadas. Este sistema visualizará los datos mediante la herramienta Power BI Desktop que permitirá identificar patrones y tendencias en el mercado laboral, proporcionando un análisis más ágil y preciso.

A través de este proyecto, se busca no solo mejorar la experiencia de los usuarios en el manejo de los datos de la GEIH, sino también proporcionar una herramienta que sea útil para investigadores, tomadores de decisiones y cualquier persona interesada en conocer las condiciones del mercado laboral colombiano.

En un entorno cada vez más digital, esta iniciativa responde a la necesidad de contar con soluciones tecnológicas que optimicen el manejo y análisis de datos, contribuyendo al desarrollo de políticas públicas más eficaces y basadas en evidencia.

El sistema de información simplificado no solo organiza y visualiza los datos de la GEIH 2024, sino que democratiza el acceso al conocimiento mediante la herramienta Power BI Desktop. Su valor radica en su capacidad de transformar datos en conocimiento aplicado, contribuyendo a formular políticas laborales más justas y efectivas.

La metodología adoptada en este proyecto demostró ser adecuada y efectiva para alcanzar el objetivo general propuesto: el desarrollo de un sistema de información simplificado que permita a usuarios no especializados acceder y analizar los microdatos de la Gran Encuesta

Integrada de Hogares (GEIH) 2024. La estructura secuencial del desarrollo de los objetivos específicos permitió abordar de forma integral cada una de las etapas del proceso de tratamiento de datos, desde la identificación de variables relevantes hasta la presentación de resultados mediante visualizaciones interactivas.

La metodología aplicada garantizó una transición fluida entre etapas y una cobertura completa del ciclo de vida del análisis de datos, desde la recolección hasta la comunicación de resultados. Este enfoque no solo fortaleció la solidez técnica del proyecto, sino que también aseguró su utilidad práctica, promoviendo el acceso democrático a información pública compleja y fomentando una cultura de análisis basada en evidencia**.**

Recomendaciones

Actualización Continua de los Datos: El sistema de información propuesto se basará en los microdatos de la GEIH 2024, es crucial implementar un proceso de actualización periódica de los datos correspondientes a los años siguientes, con el fin de reflejar los cambios en el mercado laboral y en las características sociodemográficas de la población. Esta estrategia aseguraría que el sistema mantenga su relevancia y precisión a lo largo del tiempo, proporcionando a los usuarios una herramienta actualizada y confiable.

Capacitación de los Usuarios: El sistema se diseñará para ser simplificado, es importante ofrecer capacitación a los usuarios, especialmente a aquellos sin experiencia en el manejo de grandes bases de datos o en el uso de herramientas de análisis de datos. La formación adecuada permitirá una mejor interpretación y aprovechamiento de las funcionalidades del sistema.

Mejoras en la Visualización de Datos: La visualización de resultados se realiza mediante la herramienta Power BI Desktop utilizada en la ciencia de datos, es recomendable seguir desarrollando y mejorando la interfaz de usuario para facilitar la comprensión y la toma de decisiones. Incorporar gráficos interactivos y dashboards que permitan una visión clara y dinámica de los datos sería un valor agregado.

Implementación de Mecanismos de Retroalimentación: Para garantizar que el sistema siga siendo útil y eficaz a lo largo del tiempo, se debe implementar un mecanismo de retroalimentación donde los usuarios puedan sugerir mejoras, identificar errores o plantear necesidades adicionales. Esto permitirá adaptar el sistema a los cambios en el mercado laboral o en las necesidades de los usuarios.

Desarrollo de Funcionalidades Avanzadas: El objetivo es desarrollar un sistema simplificado, en el futuro podría ser útil incorporar funciones más avanzadas, como análisis

predictivos o herramientas de modelado, que permitan predecir tendencias en el mercado laboral basadas en los datos históricos.

Integración con Otros Sistemas y Bases de Datos: Considerar la posibilidad de integrar el sistema con otras fuentes de datos relevantes, como estadísticas de empleo a nivel nacional e internacional, podría enriquecer los análisis y proporcionar una visión más completa del mercado laboral.

Fomento de la Accesibilidad: Dado que los datos de la GEIH son públicos, es importante garantizar que el sistema sea accesible para todo tipo de usuarios, independientemente de sus recursos tecnológicos. Esto incluye asegurar que el sistema sea compatible con dispositivos móviles y accesible para personas con discapacidad.

Aseguramiento de la Privacidad y Seguridad de los Datos: Se utilizarán microdatos anonimizados, se debe garantizar que el sistema cumpla con las normativas de seguridad y privacidad vigentes. Además, se recomienda implementar medidas de protección contra accesos no autorizados y posibles brechas de seguridad.

Referencias Bibliográficas

Bernal, R., & Flórez, C. (2020). Acceso a datos y formulación de políticas públicas en Colombia: desafíos y oportunidades. *Revista de Economía Institucional*, 22(43), 145–166.

CEPAL. (2021). Comisión Económica para América Latina y el Caribe. https://www.cepal.org/

Chaves, L., & Hernández, P. (2021). Aplicación de herramientas de visualización de datos para mejorar la toma de decisiones públicas. *Revista de Ciencia de Datos*, 5(2), 33–48.

DANE. (2020). *Guía de Acceso y Uso de Bases Anonimizadas – GEIH*. https://www.dane.gov.co/

DANE. (2021). *Manual del usuario: Gran Encuesta Integrada de Hogares (GEIH).* https://www.dane.gov.co/

DANE. (2022). *Manual del usuario: GEIH*. https://www.dane.gov.co/

DANE. (2023). *Ficha metodológica: Gran Encuesta Integrada de Hogares (GEIH).* https://www.dane.gov.co/

DANE. (2025). *Boletines de empleo - GEIH*. https://www.dane.gov.co/

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis.* Analytics Press.

Hermida, C., & Pulido-Mahecha, R. (2022). La GEIH como instrumento de análisis socioeconómico. *Revista Economía Colombiana*, 38(1), 78–95.

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill.

Martínez, A., & Díaz, J. (2019). *Uso ciudadano de datos abiertos: retos para la transparencia estadística en Colombia*. Observatorio de Transparencia.

Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia. (s.f.). *Datos Abiertos Colombia – GEIH*. https://www.datos.gov.co/

Ministerio de Trabajo. (s.f.). *Fuente de Información Laboral de Colombia – FILCO*. https://filco.mintrabajo.gov.co/

Morales, L., & Pineda, J. (2021). Herramientas para la democratización del análisis de datos en Colombia. *Revista de Ciencia de Datos Latinoamericana*, 3(1), 44–58.

OECD. (2019). *The Path to Becoming a Data-Driven Public Sector*. https://www.oecd.org/

Pérez, S., & Rodríguez, M. (2022). Visualización de datos públicos para la gobernanza digital: retos en América Latina. *Revista de Gobierno Electrónico*, 8(1), 23–37.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media*.

Ruiz, M., Salazar, J., & Torres, A. (2020). Uso de Power BI para la visualización de indicadores de gestión pública. *Revista Gestión y Desarrollo*, 18(1), 62–75.

UNESCO. (2020). *Data literacy for all: Building a digitally inclusive society*. https://unesdoc.unesco.org/

UNICEF. (2022). *Data Must Speak: Improving education systems through data use*. https://www.unicef.org/

Van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering,* 13(2), 22–30.

World Bank. (2021). *Data for Better Lives: World Development Report 2021*. https://www.worldbank.org/

Apéndices

Apéndice A

Enlace de Sustentación

https://unadvirtualedu-my.sharepoint.com/:p:/g/personal/ealdanag_unadvirtual_edu_co/ERO-MRwPdnxKuKPavXTR4VcBoFBrEPRAYjv45r_8q5eKAA?e=cm18TT