

Pontificia Universidad Javeriana

Proyecto Visualización de datos 2022-3

Análisis de la pobreza multidimensional a través de modelos gráficos en Power BI

Samuel Peña García

Paula Sofía Torres Rodríguez

Juan Sebastián Córdoba Valderrama

Índice

Introducción.....	1
Contexto	1
El contexto empresarial	1
El contexto de los datos	3
Objetivos.....	5
Objetivo general:	5
Objetivos específicos:.....	5
Indicadores.....	5
Herramientas.....	6
Power BI y Power Query	6
Análisis Exploratorio de Datos (EDA).....	7
Diseño	7
Análisis Exploratorio de los datos	7
Identificación de variables objetivo para cumplimiento del proyecto.....	10
Visualizaciones	11
Selección.....	11
Modelamiento	11
Evaluación	11
Elaboración de encuesta	11
Resultados de encuesta	12
Grafica Evaluada #1	12
Gráfica evaluada #2	13
Gráfica evaluada #3	14
Gráfica evaluada #4	15
Gráfica evaluada #5	16
Análisis de resultados y desarrollo de objetivos.....	17
Conclusiones.....	23
Lecciones aprendidas.....	25
Referencias	26

Introducción

Hoy en día, gracias a la evolución de la tecnología tenemos distintas formas y recursos para analizar diversos temas y/o contratiempos. Hoy fácilmente podemos hacer en uno o dos días un trabajo que hace décadas requería de meses para su realización, como lo es el caso de los análisis de datos. Como fue mencionado previamente, hay distintas formas y recursos para analizar hechos, y, en este reporte, se expone una de estas formas.

El problema del cual se hablará a lo largo del documento es la pobreza monetaria y desigualdad en Colombia en el año 2020. Un tema que durante este periodo de tiempo quiso ser analizado por el Departamento Administrativo Nacional de Estadística DANE con el propósito de suplir o dar a conocer de forma cuantificada la situación del país frente a variables correspondientes a condiciones económicas, familiares y laborales de distintos individuos. Por lo cual, se decidió acudir a la variedad de información que el DANE es capaz de proporcionar a través de sus metadatos con el fin de identificar variables que puedan ser útiles para la visualización y planteamiento de indicadores que correspondan con el análisis necesario para el cumplimiento de determinados objetivos planteados en el Plan Estratégico 2020-2022 del Ministerio de Trabajo colombiano.

De esta forma, se identificaron cuáles de los múltiples propósitos mencionados en el informe del ministerio podrían ser enriquecidos empleando las bases de datos encontradas. Asimismo, se determinaron herramientas como Microsoft Power BI, PowerQuery y demás softwares que permitirán mostrar los resultados a los entes involucrados por medio de un informe que incluya gráficas y, como producto, un dashboard dinámico para promover el entendimiento de los resultados encontrados.

Contexto

El contexto empresarial

El Ministerio de Trabajo es una entidad que, de acuerdo con la Ley 1444 de 2011 de la constitución política colombiana se encarga de la “formulación y adopción de las políticas, planes generales, programas y proyectos para el trabajo” (Departamento Administrativo de la Función Pública, 2011). En este sentido, el mismo tiene la responsabilidad de identificar y plantear junto con todos los organismos del Estado proyectos que correspondan con el Plan de Desarrollo general del gobierno. Para el año 2019, con el posicionamiento del presidente Iván Duque —quien dio a conocer una línea de gobierno que racionalizaba el gasto público, promovía el empleo formal y ofrecía trabajo a largo plazo—, esta entidad creó, conforme a las condiciones políticas, económicas y sociales del país, nuevos objetivos que contribuyeran a estas propuestas. Entre las cuales se pueden destacar:

- a. Promover la calidad de vida y bienestar de los colombianos por medio del acceso a trabajo digno y ampliación de esquemas de protección social.

- b. Promover el trabajo decente fortaleciendo políticas y estrategias orientadas a ingreso y empleo productivo.
- c. Garantizar derechos fundamentales en la ejecución del trabajo.
- d. Fortalecer el sistema de protección y seguridad social en materia de protección a la vejez, riesgos laborales y subsidio familiar.
- e. Mejorar la gestión institucional orientada a resultados.

Claramente, para cada uno de estos objetivos, existe iniciativas estratégicas, indicadores y acciones que guían a los funcionarios y entes gubernamentales para su cumplimiento. Sin embargo, nos enfocaremos especialmente en el mencionado en el punto b.

En la tabla 1 se mostrarán las dos iniciativas estratégicas, temáticas asociadas e indicadores que plantea el Ministerio de Trabajo para dar cumplimiento a este lineamiento:

Objetivos estratégicos	Iniciativas estratégicas	Temáticas asociadas a las iniciativas	Indicadores
Promover el trabajo decente a través de la formulación y fortalecimiento de políticas y estrategias orientadas a la generación de ingresos y de empleo productivo.	Formular, dirigir, coordinar, ejecutar, articular y evaluar políticas, programas, proyectos y lineamientos dirigidos a la promoción de las condiciones que permitan la generación de ingresos y empleo digno, la inclusión productiva y la formalización laboral.	Formalización Laboral <ul style="list-style-type: none"> • Inclusión Laboral • Migración Laboral • Economía Solidaria • Emprendimiento Aplicaciones y plataformas tecnológicas <ul style="list-style-type: none"> • Sistema de Información del Servicio Público de Empleo 	Personas capacitadas o sensibilizadas con estrategias de Formalización Laboral
			Porcentaje de jóvenes que no estudian y no tienen empleo
			Tasa de orientados colocados en la red pública del Servicio Público de Empleo
			Tasa de Desempleo Juvenil
	Formular, dirigir, coordinar, ejecutar, articular y evaluar políticas, programas, proyectos y lineamientos de la formación para el trabajo.	<ul style="list-style-type: none"> • Cierre de brechas de capital humano • Sistema Nacional de Cualificaciones • Formación en entornos laborales • Prácticas laborales 	Brecha de ingreso mensual promedio entre hombres y mujeres
			Jóvenes beneficiarios del Programa Estado Joven
			Tasa desempleo
			Tasa de formalidad laboral
			Porcentaje de avance en el fortalecimiento articulación y actualización de la herramienta de difusión de información del mercado laboral.

Indicadores del plan de acción institucional 2020 Objetivo 2

Dada esta información, se comenta en el documento que, para asegurar resultados de estos proyectos, se hará su correspondiente evaluación y seguimiento periódico. En este sentido, se tendrán en cuenta iniciativas estratégicas, temáticas y, especialmente los indicadores, para realizar un estudio de variables determinantes del conjunto de datos adquiridos para la ejecución, planteamiento y modelamiento de visualizaciones, actividades y proyectos que contribuyan al cumplimiento y realización de este objetivo.

El contexto de los datos

El Archivo Nacional del DANE es la rama del DANE que “tiene como objetivos garantizar la producción, disponibilidad y calidad de la información estadística estratégica, y dirigir, planear, ejecutar, coordinar, regular y evaluar la producción y difusión de información oficial básica.” (Departamento Administrativo Nacional de Estadística, 2004). Por lo cual, año tras año es uno de los entes de investigación y medición de eventos, problemáticas e impactos de diferente tipo a lo largo de la región colombiana. Entre ellas, se puede encontrar el proyecto de *Medición de Pobreza Monetaria y Desigualdad en el año 2020*, una investigación que derivó de la iniciativa Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad - MESEP. La cual fue creada en enero de 2009 mediante un convenio DANE-DNP con el objetivo de evaluar los factores que afectaron la comparabilidad de las cifras de mercado laboral y pobreza en los hogares.

Respecto a lo anterior, se puede decir que la recolección de los datos mencionados se llevó a cabo en todas las áreas metropolitanas tradicionales en Colombia, en donde se analizaron mediciones de alrededor de 168 variables que comprenden tópicos como el estrato económico, condiciones de vida, condiciones laborales, nivel de ingreso, nivel de estudios, papeles familiares, categorización de poblaciones vulneradas y el nivel de trabajo. Donde las mismas, además, fueron adquiridas haciendo uso de herramientas como lo son las Metodologías de Medición de Pobreza – propuestas tanto con por la Comisión Económica para América Latina y el Caribe (CEPAL como por la MESEP) en Colombia con el propósito de instaurar indicadores determinantes como lo son Índice de Necesidades Básicas Insatisfechas (NBI) y el nuevo Índice de Pobreza Multidimensional (IPM).

En este sentido, se hicieron uso de diversos elementos y métodos para la generación de datos procesos derivados del establecimiento de información frente al tópico de pobreza monetaria

Elementos requeridos	Metodologías de medición	Procesos y datos resultantes
<ul style="list-style-type: none"> • Valor de las líneas de pobreza: Índice que está sujeto a cambios que dependen un panorama preciso del costo de necesidades básicas como alimento, vestuario y vivienda. • Ingreso per-cápita disponible de los hogares: Se solicita información sobre las condiciones de empleo de las personas (si trabajan, en qué trabajan, cuánto ganan, si tienen seguridad social en salud o si están buscando empleo) y características generales de la población como sexo, edad, estado civil y nivel educativo. 	<ul style="list-style-type: none"> • Método del Consumo Calórico: Especificación del nivel de gasto/ingreso total que permitiría la consecución de los requerimientos nutricionales mínimos. • Método del Costo de las Necesidades Básicas: Define cumplimiento nutricional del componente alimentario para determinar líneas de pobreza Primaria -- ingresos no son suficientes para cubrir los requerimientos mínimos -- o secundaria -- ingresos habrían sido suficientes de no ser porque una porción de estos se destinó a otro tipo de consumo) • Metodología aplicada por la CEPAL: Combinación de los dos métodos descritos anteriormente. 	<ul style="list-style-type: none"> • Construcción del gasto corriente per cápita a nivel de la unidad de gasto. • Ordenamiento de los hogares por percentil de ingreso per cápita deflactado. • Selección de la población de referencia para instrucción de campañas. • Determinación de porcentajes de trabajo infantil e informal. • Definición de los dominios para la construcción de canastas y líneas de pobreza.

Elementos, metodologías y proceso derivados de la recolección de datos en el proyecto de *Medición de Pobreza Monetaria y Desigualdad en el año 2020*

De esta forma, como resultado de la investigación propuesta por estas entidades estatales, se generaron tres conjuntos de datos con distintos enfoques: hogares, personas y víctimas

Nombre Set de datos	Información	Número de variables
Hogares	Contiene la información necesaria sobre ingresos de los hogares y unidades de gasto y el valor de las líneas para el cálculo de la incidencia de pobreza e indigencia de 2020.	23
Personas	Contiene las variables a nivel de personas, necesarias para realizar los procedimientos de conformación, depuración, imputación y empalme del ingreso, para el cálculo de la pobreza monetaria, así como las variables construidas a través de dichos procedimientos.	137
Víctimas	Muestra si personas pertenecientes a un determinado grupo familiar fue víctima o no de algún tipo de conflicto u hecho violento.	8

Conjuntos de datos generados.

Cabe resaltar que, originalmente, el conjunto de datos que más información recaudaba después de su limpieza contenía alrededor de 1.240.000 filas, las cuales representan grupos familiares que, en su **mayoría**, estaban conformadas por entre 2 a 5 personas, residían en restos urbanos y, además, se encontraban entre estratos 1,2 y 3.

El **alcance** de este análisis se limitará a esta muestra de la población que fue seleccionada por los investigadores del DANE.

Objetivos

Teniendo en cuenta todo lo anterior, aunado al contexto colombiano actual, se identificaron y definieron diversos objetivos para tener en cuenta para focalizar el alcance y la utilización de los recursos disponibles, dando como resultado una investigación clara, concisa y aplicable para los grupos de interés del proyecto, tales como el ministerio de trabajo. El cual, en el año 2020 compartió un plan estratégico que propone distintos esquemas de mejoramiento frente a hechos como lo son acceso al trabajo decente; acceso a mercados dignos y acceso a la inclusión productiva.

Siendo así, los objetivos son los siguientes:

Objetivo general:

- Diseñar una interfaz gráfica que permita identificar y visualizar la relación existente entre variables relacionadas con pobreza monetaria y desigualdad en el territorio colombiano con el fin de dar cumplimiento a algunos proyectos planteados por el Ministerio de Trabajo en su plan estratégico 2020-2022.

Objetivos específicos:

- Identificar brechas de ingreso mensual promedio entre hombres y mujeres para la formulación de planes de estrategias de equidad laboral de género.
- Visualizar individuos en estado de indigencia y la ubicación en la que mayor concentración de este tipo de población se encuentra.
- Reconocer características socioeconómicas, educacionales y espaciales en las que más propenso es encontrar mayor mano de obra menor de edad (trabajo infantil).
- Distinguir sectores de acuerdo con estado socioeconómico y laboral en los que se presenta una mayor tasa de informalidad.

Indicadores

De acuerdo con los objetivos planteados anteriormente y la alineación de este proyecto con los objetivos propuestos en el plan estratégico del Ministerio de Trabajo, se considerarán esencialmente dos de los nueve indicadores que fueron preestablecidos por el mismo y se propondrá uno adicional que asegure la ejecución, evaluación y seguimiento del segundo proyecto instaurado, el cual propone promover el trabajo decente y el bienestar a través de la formulación, fortalecimiento de políticas y estrategias orientadas a la generación de ingresos y de empleo productivo. Estos son:

Indicador	Valor general de acuerdo con el total del conjunto de datos seleccionado	Valor general 2020
Brecha de ingreso mensual promedio entre hombres y mujeres.	13%	5.8%
Tasa de formalidad laboral.	51%	49%
Tasa de trabajo infantil	0,5%	4,8%

Valor de indicadores seleccionados según estadísticas del DANE para el año 2020 a nivel general en la población colombiana.

Como se puede evidenciar, cada uno de los indicadores está directamente relacionado con el análisis, desarrollo y cumplimiento de los objetivos específicos. Por lo cual, su consideración y representación visual son prioridad para la ejecución del presente trabajo.

Herramientas

Power BI y Power Query

Teniendo en cuenta el objetivo general de este proyecto, el cual es desarrollar una interfaz visual que nos permita resumir y comunicar el estado de la situación actual en Colombia respecto a brechas salariales de género, niveles de propensión a ser víctimas en poblaciones indigentes, tasa de trabajo infantil e informalidad laboral se necesitó de una herramienta que nos permitiera cargar, filtrar y plasmar gráficamente la información resultante del proyecto de Medición de Pobreza Monetaria y Desigualdad en el año 2020.

Dicho lo anterior, se seleccionó el software de Power BI —un aplicativo de la suite de Microsoft— el cual es catalogado como un “sistema predictivo, inteligente y de gran apoyo, capaz de traducir los datos (simples o complejos) en gráficas, paneles o informes”. El cual se caracteriza por dar cumplimiento de los requisitos normativos y de privacidad para cualquier sistema, además de proveer una interfaz de trabajo intuitiva que contribuye a la presentación eficaz y verídica de la información en ámbitos de inteligencia empresarial o de proyectos.

Cabe resaltar que esta plataforma tiene su propio motor de preparación de datos y transformación de datos: Power Query. Este, es capaz de dar solución a múltiples problemáticas que se ven involucradas en el proceso de conexión, búsqueda, modelamiento y reconocimiento apropiado de los datos puesto que presta funcionalidades como lo son la transformación de datos, creación de flujos e instauración de lenguaje de fórmulas.

Análisis Exploratorio de Datos (EDA)

El Análisis Exploratorio de Datos es una técnica empleada para analizar e investigar conjuntos de datos y resumir sus principales características, esto con el propósito de estudiar detenidamente la información presentada antes de hacer cualquier tipo de supuesto (IBM Cloud Education, 2020). Esta, a menudo, se lleva a cabo empleando métodos de visualización de datos, ya que facilita la identificación de datos atípicos; identificación de la distribución y conocimiento de posibles clasificaciones de la información. Sin embargo, actualmente, se habla de cuatro tipos principales de EDA:

- **No gráfico univariante.** Datos que se analizan consisten en una sola variable. El objetivo principal del análisis univariante es describir los datos e identificar los patrones que existen en ellos.
- **Gráfico univariante.**
 - Diagramas de tallo y hojas, que muestran todos los valores de datos y la forma de la distribución.
 - Histogramas, diagramas de barras en los que cada barra representa la frecuencia (recuento) o la proporción (recuento/recuento total) de casos para un rango de valores.
 - Diagramas de caja, que representan gráficamente el resumen de cinco números de mínimo, primer cuartil, mediana, tercer cuartil y máximo.
- **No gráfico multivariante:** Muestran la relación entre dos o más variables de los datos a través de la tabulación cruzada o las estadísticas.
- **Gráfico multivariante:** El gráfico más utilizado es un diagrama de barras agrupadas o un gráfico de barras donde cada grupo representa un nivel y cada barra dentro de un grupo representa los niveles de la otra variable.

Diseño

Análisis Exploratorio de los datos

En primer lugar, se quiso identificar por medio de Power Query la información a la cual se tenía y no se tenía acceso. Puesto que, aunque panorámicamente parecía que se conocían gran cantidad de atributos por cada uno de los entes registrados en los data sets, se encontró que existían múltiples columnas que carecían de hasta el 99% de la información que se suponía debían reflejar. Por lo cual, se evaluó —de acuerdo con los porcentajes de ocupación— a las variables que podrían ser eliminadas por la alta carencia de información, como las mostradas a continuación:

ABC 123 Cotiza pensión?	A ^B C Trabajo o negocio adicional	A ^B C Ingresos trabajo	1 ² 3 p5100
● Válido 56 %	● Válido 56 %	● Válido 27 %	● Válido 3 %
● Error 0 %	● Error 0 %	● Error 0 %	● Error 0 %
● Vacío 44 %	● Vacío 44 %	● Vacío 73 %	● Vacío 97 %

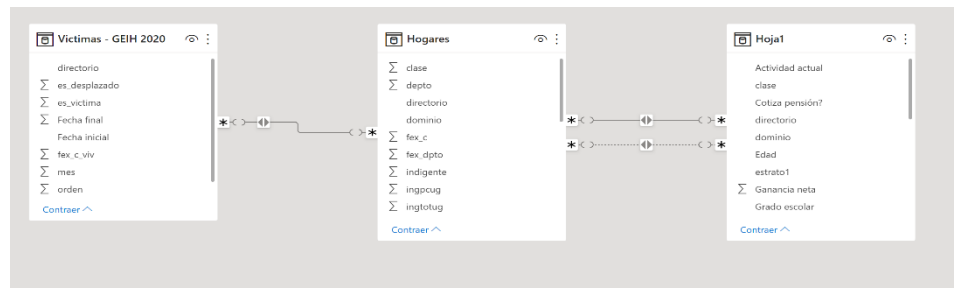
Columnas con porcentajes mínimos de ocupación

Adicionalmente, durante este proceso, se encontraron múltiples datos que no fueron formateados de forma correcta —teniendo en cuenta la información consignada en el diccionario de datos con el cual viene adjunto junto con los CSV— bien sea porque el software no fue capaz de identificarlo correctamente o porque existía algún tipo de error de digitación. Por lo cual, se reemplazaron múltiples valores por su correspondiente etiqueta.

directorio	secuencia_p	mes
5105379	1	13
5105380	1	13
5105381	1	13
5105382	1	13
5105383	1	13
5105384	1	13
5105385	1	13
5105386	1	13

Ejemplos datos mal formateados o digitados

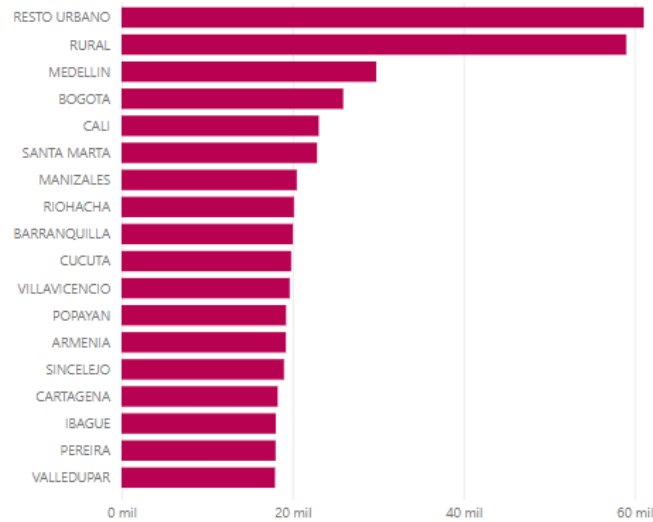
Además, indagando un poco más acerca de las características que se podían encontrar de cada objeto, se evidenció que existían atributos como lo eran el *dominio* (cualidad categórica que indicaba el área metropolitana a la que pertenecía una persona) y *directorio* (variable cualitativa que asignaba una especie de “id” a los grupos registrados). Lo que posteriormente permitió establecer relaciones entre los conjuntos de datos, como se consigna en la siguiente imagen:



Relaciones entre conjuntos de datos

Habiendo hecho lo anterior, se empezó a trabajar con la interfaz gráfica, en donde —sin tener en cuenta atributos como lo son el diseño y la estética— se empezó a evaluar la información, distribución y cantidad de datos que se tenían. Entre estos, se destacaron:

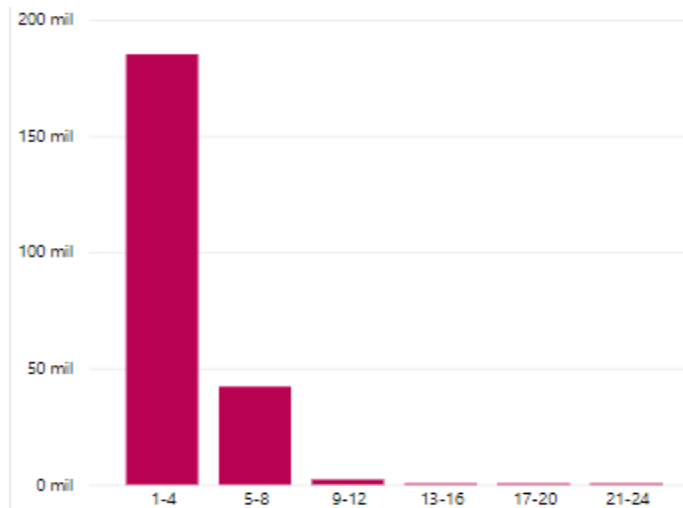
- Distribución del **total general de los hogares** según el **área metropolitana**.



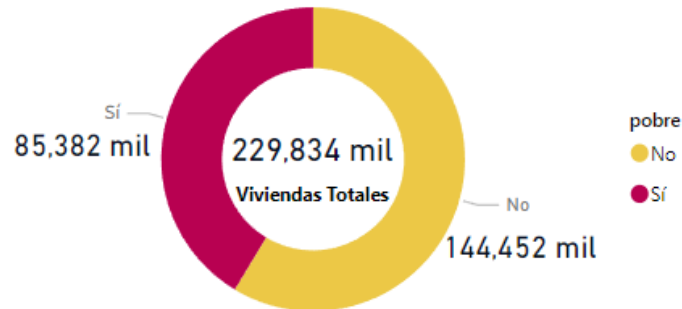
- Distribución del **total general de los hogares** según el **estrato**.



- Distribución del **total general de los hogares** según la **cantidad de individuos que lo conforman**.



- Distribución del **total general de los hogares** según su **clasificación de pobreza**.



Finalmente, tras la limpieza y entendimiento de los datos, se establecieron y definieron de forma más objetiva los posibles alcances que tendrían los análisis y acciones que derivan de la formulación del presente proyecto.

Identificación de variables objetivo para cumplimiento del proyecto

Por lo mencionado en la sección de indicadores, encontramos distintas variables que son específicas para el trabajo de cada uno de los objetivos formulados. En este sentido —tras realizar el análisis exploratorio de los datos— se identificó que los valores disponibles son:

Objetivo	Valor	Tipo
Identificar brechas de ingreso mensual promedio entre hombres y mujeres para la formulación de planes de estrategias de equidad laboral de género.	Ingreso	Número decimal
	Sexo	Categorico
	Puesto de trabajo	Categorico
	Salario	Número decimal
Precisar zonas que requieran de un plan de rehabilitación y protección para individuos en estado de indigencia que son propensos a ser víctimas de cualquier tipo.	Dominio	Categorico
	Indigente	Booleano
	Es_victima	Booleano
Reconocer características socioeconómicas, educacionales y espaciales en las que más propenso es encontrar mayor mano de obra menor de edad (trabajo infantil).	Edad	Número entero
	Nivel_educativo	Categorico ordinal
	Es_victima	Booleano
	Es_Desplazado	Booleano
	Puesto de trabajo	Categorico
Distinguir sectores de acuerdo con estado socioeconómico y laboral en los que se presenta una mayor tasa de informalidad.	Estrato	Categorico ordinal
	Puesto de trabajo	Categorico
	Régimen de seguridad social	Categorico

	Dominio	Categorico
	Cotiza Pensión	Booleano
	Seguridad Social	Booleano

Variables objetivo para el cumplimiento de objetivos.

Visualizaciones

Selección

Se puede reconocer que, de acuerdo con las diferentes variables existentes, los volúmenes de datos obtenidos y el propósito de la visualización se deben seleccionar distintos tipos de gráficos que representen forma verídica y correcta la información a presentar.

Tediando en cuenta que los volúmenes de datos son medios y que se tienen diferentes cuando se trata de relacionar variables, los gráficos seleccionados fueron (véase tabla):

Propósito	Variables involucradas	Gráfico
Comparación	Variables categóricas y numéricas	Gráfico de barras (horizontales y verticales)
Composición	Cambio temporal • Meses	Gráfico de línea
	Estadístico • Datos de tipo booleano	Gráfico de tartas
Distribución	Variables categóricas y numéricas	Histograma de columna

Gráficos seleccionados según propósito

Modelamiento

Evaluación

Elaboración de encuesta

Después de modelar gráficos de acuerdo a parámetros preestablecidos para la elaboración de los mismos, se decidió realizar una encuesta en la que los participantes debían visualizar gráficos (los cuales, se consideraba, ya cumplían con los parámetros para ser catalogados como buenos), para que fuesen valorados de acuerdo con las siguientes afirmaciones del 1-5 haciendo uso de la escala de Likert (donde 1: Totalmente en desacuerdo, 2: En desacuerdo, 3: Ni de acuerdo ni en desacuerdo, 4: De acuerdo, 5: Totalmente de acuerdo).

Número	Afirmación
1	La relación entre los colores y la temática son consideradas como adecuadas
2	Los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia
3	Los gráficos seleccionados son los pertinentes para la visualización de los tipos de datos que se están representando
4	Los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico
5	Haría uso de estas visualizaciones para incluirlas en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados

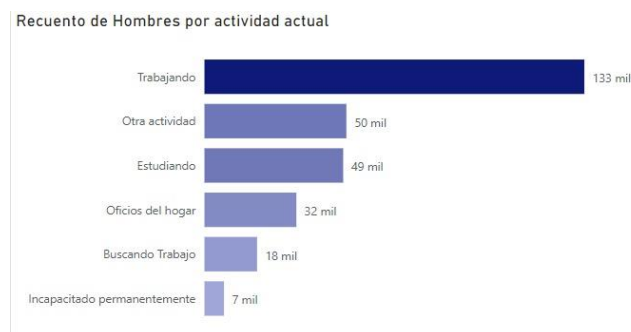
Afirmaciones incluidas en la encuesta

Resultados de encuesta

Se plantea y despliega una encuesta en el formato “Google Forms”, donde se invita a diferentes personas de manera voluntaria a responder 25 preguntas de tipo cuestionario en formato Likert donde 1 es totalmente desacuerdo y 5 totalmente de acuerdo, se logró recolectar la información de 38 sujetos.

Las preguntas están divididas homogéneamente en 5 visualizaciones distintas, es decir, cada grafico es sometido a los mismos 5 interrogantes que abordan los distintos tópicos de este mismo (relación de colores y temática, simplificación del contexto, adecuación de su uso, comprensión y relevancia).

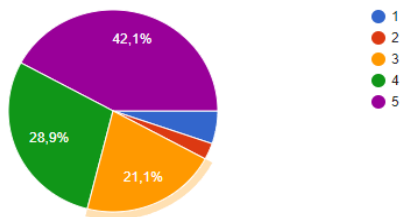
Grafica Evaluada #1



Resultados:

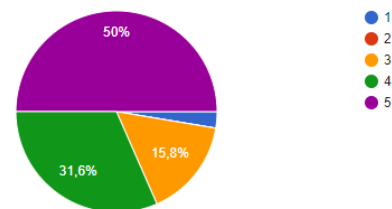
Teniendo en cuenta este gráfico, ¿la relación entre los colores y la temática son consideradas como adecuadas?

38 respuestas



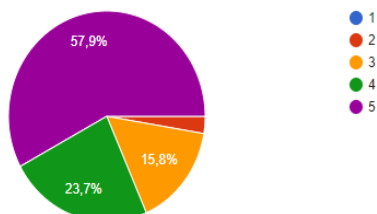
Teniendo en cuenta este gráfico, ¿los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia?

38 respuestas



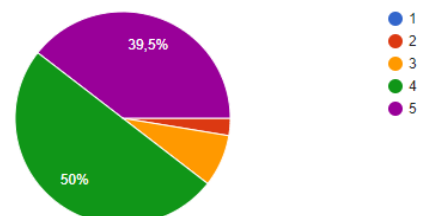
Teniendo en cuenta este gráfico, ¿los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico?

38 respuestas



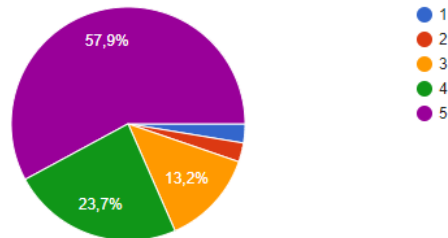
Teniendo en cuenta este gráfico, ¿el gráfico seleccionado es pertinente para la visualización del tipo de datos que se están representando?

38 respuestas



Teniendo en cuenta este gráfico, ¿haría uso de este gráfico para incluirlo en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados?

38 respuestas



Gráfica evaluada #2

Promedio de Salario en el puesto seleccionado por Sexo por Sexo



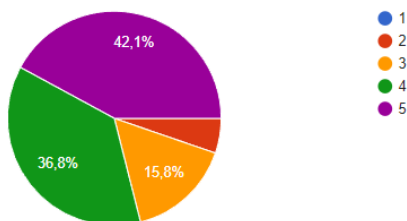
Puesto de trabajo

Empleado doméstico	Obrero o empleado de empresa particular
Jornalero	Obrero o empleado del gobierno

Resultados

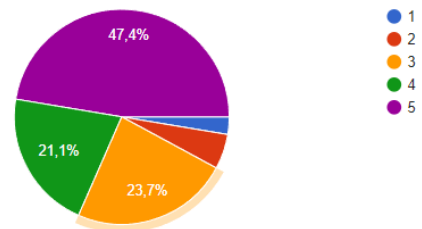
Teniendo en cuenta este gráfico, ¿la relación entre los colores y la temática son consideradas como adecuadas?

38 respuestas



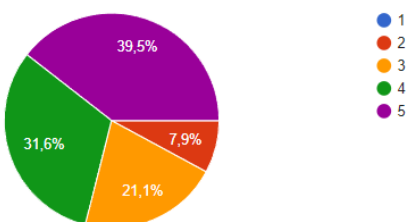
Teniendo en cuenta este gráfico, ¿los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia?

38 respuestas



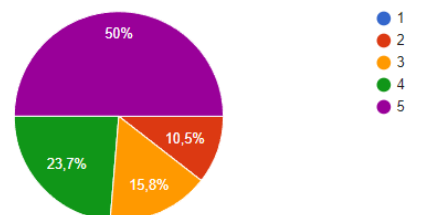
Teniendo en cuenta este gráfico, ¿el gráfico seleccionado es pertinente para la visualización del tipo de datos que se están representando?

38 respuestas



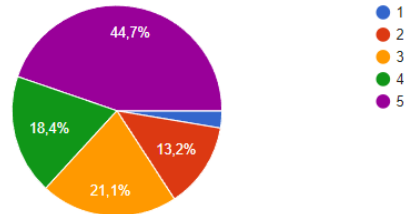
Teniendo en cuenta este gráfico, ¿los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico?

38 respuestas



Teniendo en cuenta este gráfico, ¿haría uso de este gráfico para incluirlo en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados?

38 respuestas



Gráfica evaluada #3

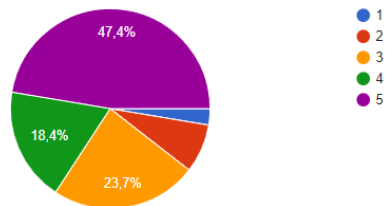
Grado escolar por estrato

Grado escolar	1	2	3	4	5	6	Total
0	42.37%	27.12%	10.17%	6.78%	10.17%	3.39%	100.00%
1	58.33%	20.83%	4.17%	4.17%	12.50%		100.00%
2	50.00%	35.71%	7.14%	7.14%			100.00%
3	63.41%	21.95%	9.76%	4.88%			100.00%
4	62.50%	17.50%	20.00%				100.00%
5	47.18%	23.94%	16.20%	9.15%	3.52%		100.00%
6	52.83%	25.79%	10.06%	6.92%	2.52%	1.89%	100.00%
7	41.77%	23.42%	13.29%	12.03%	6.96%	2.53%	100.00%
8	36.92%	20.77%	26.15%	9.23%	6.15%	0.77%	100.00%
9	38.06%	24.63%	18.66%	8.21%	9.70%	0.75%	100.00%
10	30.77%	24.36%	16.67%	14.10%	12.82%	1.28%	100.00%
11	22.14%	26.72%	25.95%	13.74%	9.16%	2.29%	100.00%
Total	42.08%	24.29%	16.64%	9.25%	6.41%	1.33%	100.00%

Resultados

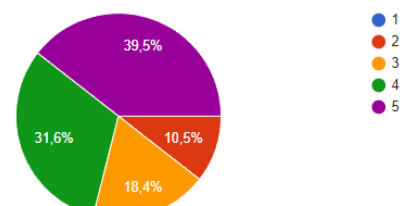
Teniendo en cuenta este gráfico, ¿la relación entre los colores y la temática son consideradas como adecuadas?

38 respuestas



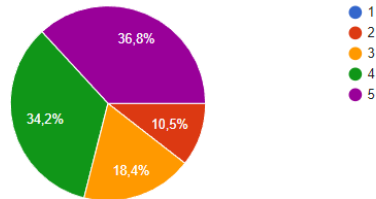
Teniendo en cuenta este gráfico, ¿los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia?

38 respuestas



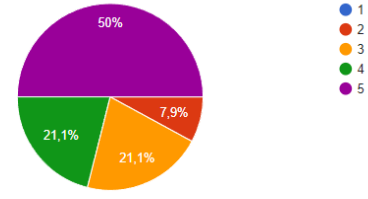
Teniendo en cuenta este gráfico, ¿el gráfico seleccionado es pertinente para la visualización del tipo de datos que se están representando?

38 respuestas



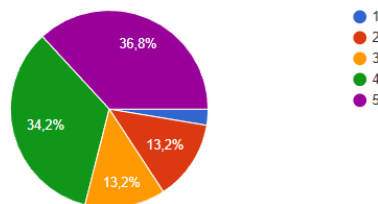
Teniendo en cuenta este gráfico, ¿los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico?

38 respuestas



Teniendo en cuenta este gráfico, ¿haría uso de este gráfico para incluirlo en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados?

38 respuestas



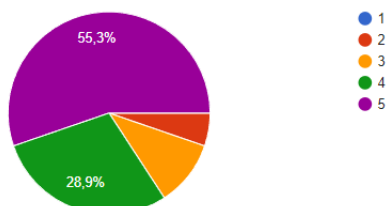
Gráfica evaluada #4



Resultados:

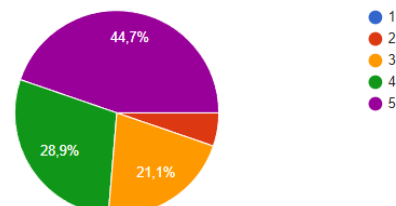
Teniendo en cuenta este gráfico, ¿la relación entre los colores y la temática son consideradas como adecuadas?

38 respuestas



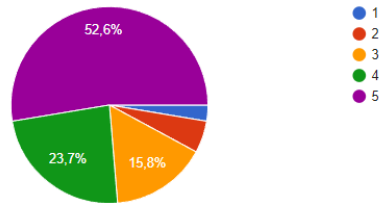
Teniendo en cuenta este gráfico, ¿los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia?

38 respuestas



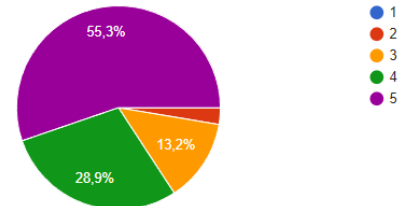
Teniendo en cuenta este gráfico, ¿el gráfico seleccionado es pertinente para la visualización del tipo de datos que se están representando?

38 respuestas



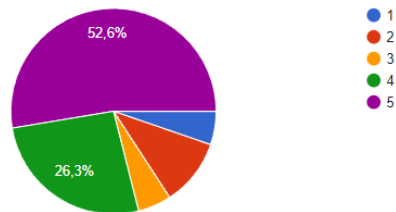
Teniendo en cuenta este gráfico, ¿los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico?

38 respuestas



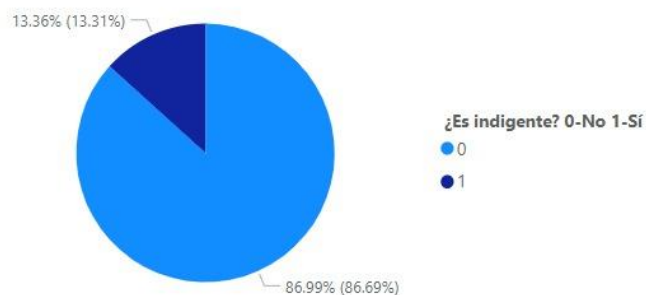
Teniendo en cuenta este gráfico, ¿haría uso de este gráfico para incluirlo en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados?

38 respuestas



Gráfica evaluada #5

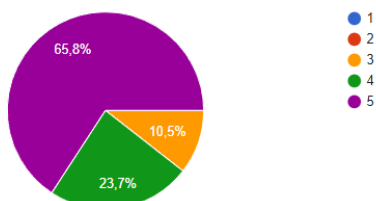
Porcentaje de indigencia respecto al total de encuestados



Resultados

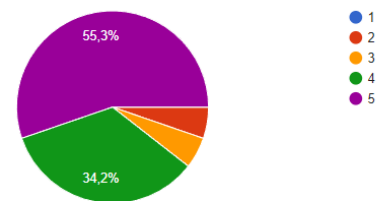
Teniendo en cuenta este gráfico, ¿la relación entre los colores y la temática son consideradas como adecuadas?

38 respuestas



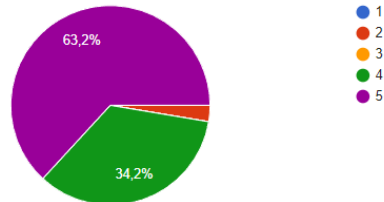
Teniendo en cuenta este gráfico, ¿los datos que se logran visualizar en el plano dan una idea del contexto al que está haciendo referencia?

38 respuestas



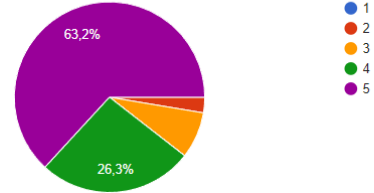
Teniendo en cuenta este gráfico, ¿el gráfico seleccionado es pertinente para la visualización del tipo de datos que se están representando?

38 respuestas



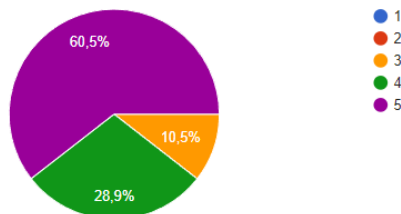
Teniendo en cuenta este gráfico, ¿los títulos y leyendas seleccionadas contribuyen a la comprensión del gráfico?

38 respuestas



Teniendo en cuenta este gráfico, ¿haría uso de este gráfico para incluirlo en algún trabajo, reporte o presentación que tenga relación con los tópicos presentados?

38 respuestas



Tras revisar las respuestas conjuntamente, se puede afirmar correctamente que los gráficos generados son útiles para la descripción de los datos, ya que a nivel general se percibe un porcentaje de 52,4% en “*Totalmente de acuerdo*”, 29,4% “*De acuerdo*” y los votos por “*Totalmente en desacuerdo*” o “*Desacuerdo*” no son representativos en los porcentajes generales, por lo tanto, se puede afirmar que los dashboards, serán modificados ligeramente para cumplir con las expectativas de visualización. Sin embargo, estas no serán reformuladas por completo.

Análisis de resultados y desarrollo de objetivos

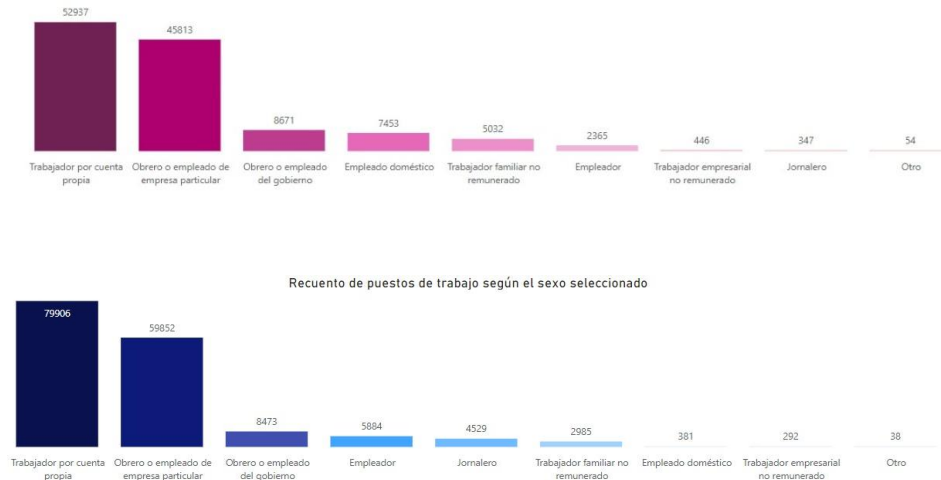
A continuación, se enumerarán uno a uno los objetivos y se mostrarán los distintos gráficos que resultaron del análisis empleado. Cada uno de ellos empleando al menos dos de las variables objetivo que fueron **especificadas** en la sección de **diseño**. Cabe resaltar, que cada una de las visualizaciones descritas a continuación podrán ser visualizadas más nítidamente en los dashboards adjuntos junto con este reporte.

1. Identificar brechas de ingreso mensual promedio entre hombres y mujeres para la formulación de planes de estrategias de equidad laboral de género.

Desde la exploración de estos datos se identificó que efectivamente el data set seleccionado reflejaba la realidad colombiana: existe una brecha salarial de género. El índice calculado al nivel de data set alcanza el 13%. Dónde el salario promedio recibido por las mujeres es de

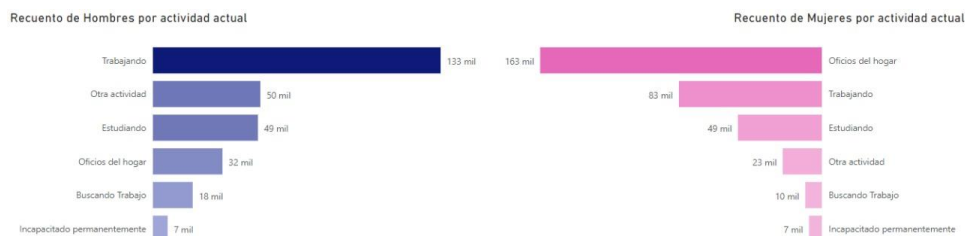
un millón trescientos mil pesos, mientras que el de los hombres es de un millón trescientos cincuenta mil pesos. Respecto a esto, se evaluaron otros atributos cómo:

- Se puede ver, que las labores están distribuidas de una forma bastante parecida. Por lo que no se detectó ninguna discriminación especial por parte de algún sector.



- Aquí el gráfico da a conocer algo que, a pesar de que se considera erróneamente que los índices de desigualdad son obsoletos, en el grupo analizado se evidencia que la mayoría de las mujeres realiza tareas del hogar. Mientras que, para el caso de los hombres, gran parte —casi en la misma proporción que las mujeres que trabajan en tareas de casa— se encuentran trabajando.

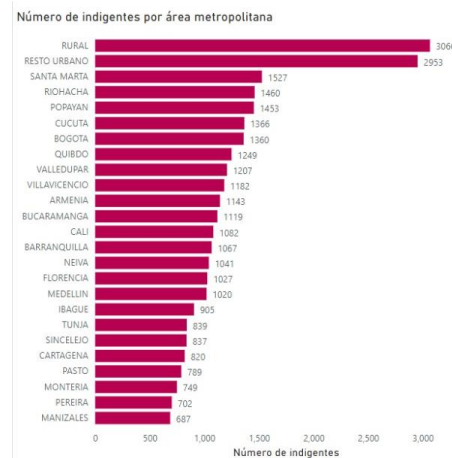
Sin embargo, se evidencia que los dos siguientes grupos más grandes de las mujeres involucran a aquellas que se encuentran trabajando y estudiando, de forma parecida a los hombres, qué comparten el mismo porcentaje de la población considerada como estudiante. Por lo cual, aunque se reconoce que sigue existiendo una diferencia considerable en lo que parecen ser oportunidades de trabajo o estigmatización de género, existen puntos en los que, parcialmente hay equidad.



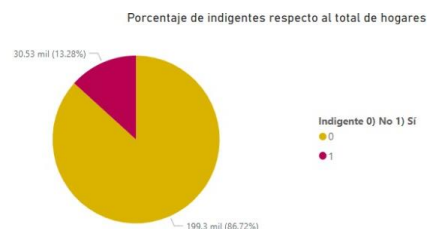
2. Visualizar individuos en estado de indigencia y la ubicación en la que mayor concentración de este tipo de población se encuentra.

- En este gráfico obtenemos dos zonas críticas en las que la mayoría de los individuos considerados en la investigación de pobreza monetaria se encuentran en estado de indigencia. Estas son las de carácter rural y resto urbano. Locaciones que, aunque son muy distintas entre sí, representan la mayor parte del territorio colombiano.

Por otra parte – considerando zonas aún más específicas -- se ve como áreas metropolitanas como lo son Santa Marta, Riohacha, Popayán, Cúcuta y Bogotá recopilan la mayoría de la población indigente en Colombia.

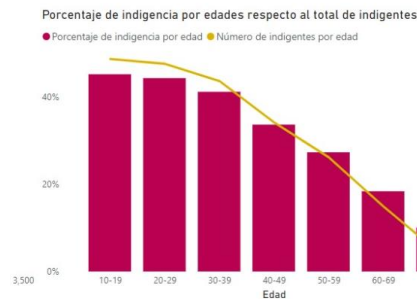


- Para denotar un porcentaje se estableció frente al total general de edades un gráfico que permitiera distinguir entre habitantes de calle y personas que no lo eran. Lo cual, nos permitió conocer que aproximadamente el 13% de la muestra analizada pertenecía a la población indigente. Lo cual, a consideración individual, representa un porcentaje bastante alto.



- Buscando evidenciar la tendencia en el porcentaje de indigencia se establecieron distintos rangos de edad, para demostrar esto correctamente se empleó un diagrama

de barras con el que se puede concluir que entre menor sean los rangos de edad se presenta la mayor cantidad de indigencia; paralelo al grafico se usa una línea de tendencia, la cual busca representar con mayor claridad la disminución del índice de indigencia a medida del aumento en la edad.



3. Reconocer características socioeconómicas, educacionales y espaciales en las que más propenso es encontrar mayor mano de obra menor de edad (trabajo infantil).

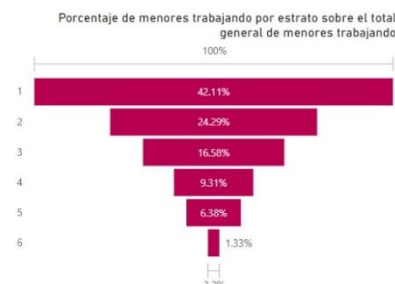
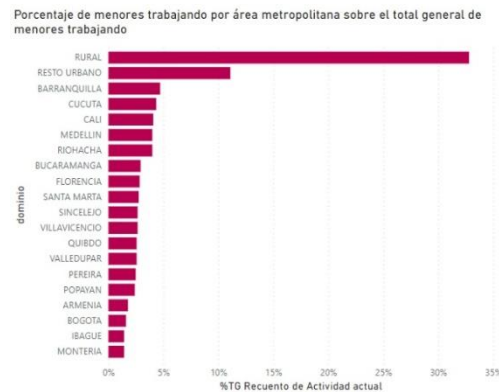
- Para reconocer de forma más amena esta temática, quisimos estudiar un factor primordial en la infancia: La educación. Se consideró, de esta forma, la distribución del nivel educativo de la población menor de 18 años que ya trabaja con el propósito de conocer a que edad o en que etapa de la vida académica de un menor de edad este iniciaba a trabajar. Por medio de una matriz y de un gráfico de barras, se ve que hay un pico de este índice de trabajo desde el 5 grado del colegio.



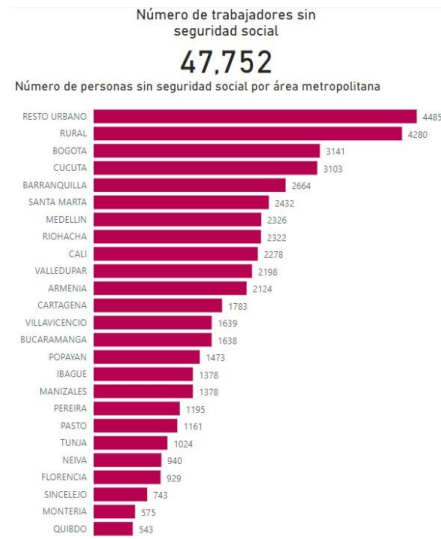
Grado escolar por estrato

Grado escolar	1	2	3	4	5	6	Total
0	42.37%	27.12%	10.17%	6.78%	10.17%	3.39%	100.00%
1	58.33%	20.83%	4.17%	4.17%	12.50%		100.00%
2	50.00%	35.71%	7.14%	7.14%			100.00%
3	63.41%	21.95%	9.76%	4.88%			100.00%
4	62.50%	17.50%	20.00%				100.00%
5	47.18%	23.94%	16.20%	9.15%	3.52%		100.00%
6	58.87%	25.79%	10.06%	6.92%	2.52%	1.89%	100.00%
7	41.77%	23.42%	13.29%	12.03%	6.96%	2.53%	100.00%
8	36.92%	20.77%	26.15%	9.23%	6.15%	0.77%	100.00%
9	38.09%	24.83%	18.68%	8.21%	9.70%	0.75%	100.00%
10	30.77%	24.36%	16.67%	14.10%	12.82%	1.28%	100.00%
11	22.14%	26.72%	25.95%	13.74%	9.16%	2.29%	100.00%
Total	42.08%	24.29%	16.64%	9.25%	6.41%	1.33%	100.00%

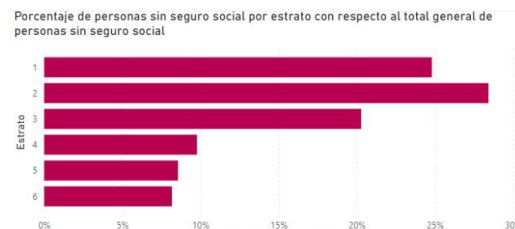
- Se quiso identificar, además, si eran zonas rurales o urbanas las que más tendían a reclutar mano de obra joven para realizar actividades de tipo laboral. De manera casi determinante, se ve como las zonas rurales contratan mayor mano de obra que tiene menos de 18 años respecto al total general de personas que se encuentran trabajando.



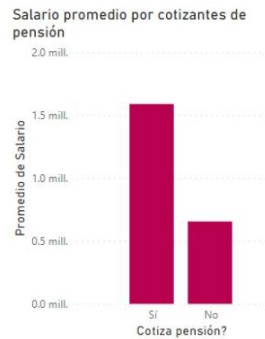
- Distinguir sectores de acuerdo con estado socioeconómico y laboral en los que mayor tasa de informalidad.
 - En este grafico se compara las zonas metropolitanas en general con el número de personas sin seguridad social, paralelo al grafico de la indigencia poseen características muy similares, nuevamente las zonas de carácter Rural y el resto urbano, son las más representativas. Así también las demás localizaciones más específicas se presentan nuevamente.



- Para generar un porcentaje de la cantidad de personas sin seguridad social se remitió al estrato socioeconómico, buscando una relación de a mayor estrato, mayor porcentaje de personas que poseen seguro social. Después de realizar la visualización se evidencia mayor cantidad de personas que no poseen seguro social se sitúa entre los estratos 1-3.



- Pretendiendo una relación entre la tasa de informalidad y la cotización pensional, se encontró que el promedio de un salario de un sujeto que cotiza pensión en relación con el de una persona que no cotiza pensión, es desproporcional llegando a ser prácticamente 3 veces mayor, lo cual se evidencia de manera clara en el gráfico planteado.



Conclusiones

Respecto a cada uno de los objetivos se pudieron generar distintas conclusiones que son de gran utilidad para la elaboración de, por ejemplo, políticas públicas, campañas y eventos que mejoren la calidad de vida, salario y estado económico de muchos colombianos alrededor del territorio. Teniendo en cuenta que se trabajaron distintas problemáticas y distintos sectores sociales, se evaluaron las siguientes propuestas basados en los resultados y la información generada por los gráficos.

1. Brecha salarial

- **Información recopilada:**

- Para el grupo analizado existe una diferencia entre promedios de salarios de 50.000. Lo cual, equivale para el nivel de ingreso manipulado, a una brecha del 13%.
- Las mujeres, antes que ingresar a un trabajo formal, tienden a ejercer labores relacionadas con el hogar.
- No existe discriminación por parte de un sector específico.

- **Propuestas de acción:**

- Proponer al ministerio de trabajo realizar campañas que promuevan la igualdad de salarios y oportunidades en empresas como lo son las PIMES. Teniendo presente que la fuerza laboral se concentra más que todo en el área de trabajo “por cuenta propia”. Por lo cual identificar pequeños negocios, vendedores o prestadores de servicios, puede estimular la igualdad salarial.
- Proponer en nombre del ministerio de trabajo, al DANE, realizar estudios sectorizados que permitan determinar con más precisión si ambos sexos están

expuestos a igualdad de oportunidades o sí, por el contrario, siguen existiendo modelos mentales que estimulan la discriminación.

- Promover por medio de publicidad, comerciales y campañas escolares que la fuerza laboral que está siendo formada (estudiantes) reconozca la igualdad de género.

2. Población indigente

• Información recopilada:

- Aproximadamente el 13% de la muestra analizada pertenecía a la población indigente.
- La mayoría de esta población se encuentra ubicada en zonas rurales y determinados sectores urbanos.

• Propuestas de acción:

- Realizar un estudio que evalúe en un modelo de causa-raíz las condiciones que hicieron que estos individuos terminaran en un estado de indigencia.
- Proponer al gobierno y al ministerio de salud una colaboración en la que se genere algún tipo de programas de rescate para personas de mayores de 18 en estado de indigencia que puedan ser rehabilitadas y capacitadas para entrar en puestos laborales operativos.

3. Trabajo infantil

• Información recopilada:

- Existe un pico que demuestra que la mayoría de los niños que ejecutan algún tipo de actividad laboral lo hacen estando en quinto grado del colegio.
- La mayoría de los niños que trabajan pertenecen al estrato uno colombiano.

• Propuestas de acción:

- Analizar si las actividades laborales ejecutadas con mayor frecuencia en zonas rurales – como las agropecuarias— son realizadas o cedidas por adultos mayores a niños menores de edad.
- Determinar en compañía del SNIES y el SPADIES si efectivamente la mayoría personas que acceden a una institución de educación superior (y se encuentran en una edad inferior a los 18 años) no tienen antecedentes laborales.
- Estipular junto con el gobierno zonas de estrato unos en las que se deberían instaurar subsidios y colegios para desincentivar el trabajo infantil.

4. Informalidad laboral

• Información recopilada:

- **Sector urbano catalogado como uno de los más propensos a tener tasas de informalidad laboral.**
- Mayor cantidad de personas que no poseen seguro social se sitúa entre los estratos 1-3.
- **Propuestas de acción:**
 - Reconocer causas principales que hacen que el sector urbano sea uno de los más propensos a evadir el proceso de formalización laboral.
 - Enfocar la elaboración de campañas en estratos 1,2 y 3 – por medio de incentivos del gobierno – que permitan disminuir considerable el índice de informalidad —teniendo en cuenta que este está alcanzando casi el 50% a nivel general de Colombia.—

Lecciones aprendidas

A lo largo del proceso de elaboración de todo este plan existieron varios percances. De hecho, este fue una segunda opción después de haber considerado un proyecto que pretendía estudiar las causas de deserción académica en la Universidad Católica de Colombia; se establecieron objetivos, alcances e incluso se explicó con detalle el contexto al que se había enfrentado la institución durante los años a estudiar. Sin embargo, a pesar de planificar todo con tanto detalle se olvidó evaluar a profundidad una de las herramientas fundamentales para cumplir los objetivos declarados: Los datos. A pesar de que se supuso que el SNIES y el SPADIES tenían información completa, estas entidades solo suplían información dada en porcentajes a modo de resumen. Es decir, no nos permitía analizar diversas condiciones de individuos. Esto, generó que todo fuera obsoleto.

Dado lo anterior, se empezó a indagar mucho más sobre las fuentes de información, la calidad y completitud de los datos. Por lo cual, mientras se hacía la búsqueda de nuevos sets, se buscó que esto permitieran formular posibles estrategias que contribuyeran a bienestar social y visualizar óptimamente información que describiera la situación actual frente a diferentes situaciones en Colombia.

Por otro lado, con forme se explora más a detalle la información, los objetivos dejan de ser tan ambiciosos y se escalan a propuestas realistas que se adecuen al alcance de los datos. Así pues, se crearon, visualizaciones simples, pero objetivas, que permiten analizar y caracterizar el grupo objetivo. Estas, se realizaron teniendo en cuenta las distintas fases revisadas en clase: Entender el contexto (Ya que se evaluaron las fuentes de información y se reconoció la población evaluar), escoger la visualización (donde se aprendió -- de forma mucho más autónoma – a manipular el software de visualización), evaluar el diseño (por medio de la evaluación de encuestas que nos permitieron reevaluar nuestros diseños) y, finalmente, seleccionar nuestros mejores gráficos.

En resumen, no solo conocimos de qué forma se debe de emplear una aplicación como los Power BI, sino que, además, se supo cómo formular, evaluar y gestiona un proyecto teniendo en cuenta los recursos, alcances y limitaciones del entorno.

Referencias

El DANE informó que diciembre de 2020 cerró con una tasa de informalidad de 49%. (s. f.). Recuperado 8 de noviembre de 2022, de <https://www.larepublica.co/economia/el-dane-informo-que-diciembre-de-2020-cerro-con-una-tasa-de-informalidad-de-49-3124018>

Entre octubre y diciembre de 2021 se reportaron más de 508.000 menores en trabajo infantil, según DANE - Infobae. (s. f.). Recuperado 8 de noviembre de 2022, de <https://www.infobae.com/america/colombia/2022/04/09/entre-octubre-y-diciembre-de-2021-se-reportaron-mas-de-508000-menores-en-trabajo-infantil-segun-dane/>

¿Qué es el análisis exploratorio de datos? (2021, agosto 4). <https://www.ibm.com/cloud/learn/exploratory-data-analysis>

¿Qué es Power BI? | Deloitte España. (s. f.). Deloitte Spain. Recuperado 8 de noviembre de 2022, de <https://www2.deloitte.com/es/es/pages/technology/articles/que-es-power-bi.html>