**Análisis Estadístico de la Encuesta de Carga Financiera y Educación Financiera de los Hogares 2018**

Presentado por:

Jorge Andres Buitrago Cepeda

David Santiago Salazar Garcia

German Dario Martinez Robayo

Docente:

Fernando Ariel Bomba Bombo

Universidad Central

Facultad de Ciencias administrativas, económicas y contables

Maestría en Analítica de datos

Bogotá

2022

**Índice**

[1. Introducción 3](#_Toc117201046)

[2. Justificación 3](#_Toc117201047)

[3. Objetivos minería 3](#_Toc117201048)

[3.1. Objetivo General 3](#_Toc117201049)

[3.2. Objetivos específicos 4](#_Toc117201050)

[4. Objetivos comerciales 4](#_Toc117201051)

[4.1. Objetivo General 4](#_Toc117201052)

[4.2. Objetivos específicos 4](#_Toc117201053)

[5. Comprensión de los datos. 5](#_Toc117201054)

[5.1. Variables. 5](#_Toc117201055)

[6. Comprensión de los datos 6](#_Toc117201056)

[7. Preparación de los datos 11](#_Toc117201057)

[7.1. Limpieza. 11](#_Toc117201058)

[7.2. Codificación Variables Categoricas 12](#_Toc117201059)

[7.3. Datos Atipicos 14](#_Toc117201060)

[7.3. Datos Atipicos 19](#_Toc117201061)

# Introducción

La Encuesta de Carga Financiera y Educación Financiera de los Hogares (IEFIC) tiene el objetivo de “obtener información de calidad sobre la situación financiera de los hogares colombianos” (Departamento Administrativo Nacional de Estadística, 2019).

Esta encuesta busca construir indicadores asociados con los aspectos financieros de las familias como: nivel de endeudamiento, capacidad de pago, conocimiento de los hogares de variables monetarias y financieras y cómo esto afecta sus decisiones de consumo, ahorro e inversión.

Con esta encuesta se busca tener datos confiables y seguros, para identificar variables financieras de los hogares relacionadas con sus decisiones de consumo estudiando los principales gastos de acuerdo a los ingresos y cómo esto puede explicar aspectos como el tipo de vivienda y la posibilidad de tener un negocio propio.

# Justificación

Con la información La Encuesta de Carga Financiera y Educación Financiera de los Hogares (IEFIC), se pretende dar respuesta a algunas preguntas de interés y cumplimiento de objetivos tanto desde la minería datos como comerciales referente a la capacidad de los Hogares para la adquisición de vivienda propia

Los objetivos comerciales van primero que los de MD

# Objetivos minería

## 3.1. Objetivo General

Desarrollar soluciones basadas en Técnicas de minería de datos con el insumo de la encuesta de Carga Financiera y Educación Financiera de los Hogares (IEFIC) en el 2018 para las ciudades de Bogotá, Cali y Medellín, adaptadas a las necesidades en términos de educación financiera de los hogares que permitan a una empresa del sector financiero tomar decisiones soportadas en el conocimiento de los datos

## 3.2. Objetivos específicos

1. Caracterizar a los encuestados de acuerdo a las siguientes variables demográficas: sexo y nivel educativo.
2. Realizar un análisis detallado de los ingresos de los encuestados en relación a los gastos de: alimentación, vestuario, servicios públicos, servicios tecnológicos y salud.
3. Identificar las condiciones de propiedad de la vivienda en torno a la dependencia con otras variables como: sexo, nivel educativo, negocio propio y municipio.
4. Estimar a partir de una muestra la media y el total del ingreso de los encuestados.
5. Desarrollar un análisis comparativo sobre el nivel de ingresos de los encuestados de las ciudades de Medellín y Cali frente a Bogotá.

# Objetivos comerciales

## 4.1. Objetivo General

Analizar las características de los consumidores de los hogares Colombianos encuestados, por medio de sus variables demográficas para diseñar un producto financiero que se adapte a cada necesidad según los grupos demográficos encontrados.

## 4.2. Objetivos específicos

Realizar un análisis de los grupos demográficos encontrados para proponer un portafolio de productos que se adapte a cada grupo.

Proponer un portafolio de productos basado en los perfiles demográficos de los encuestados para evaluar su viabilidad.

# Comprensión de los datos.

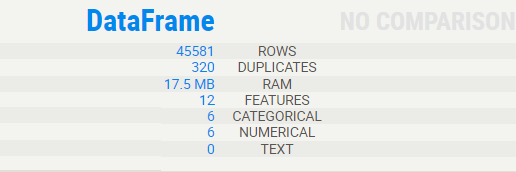
Desde el 2009 el Banco de la República y el Departamento Administrativo Nacional de Estadística han implementado la Encuesta de Carga Financiera y Educación Financiera de los Hogares (IEFIC) con el objetivo de “obtener información de calidad sobre la situación financiera de los hogares colombianos” (Departamento Administrativo Nacional de Estadística, 2019). Esta encuesta busca construir indicadores asociados con los aspectos financieros de las familias como: nivel de endeudamiento, capacidad de pago, conocimiento de los hogares, de variables monetarias y financieras y cómo esto afecta sus decisiones de consumo, ahorro e inversión.

## 5.1. Variables.

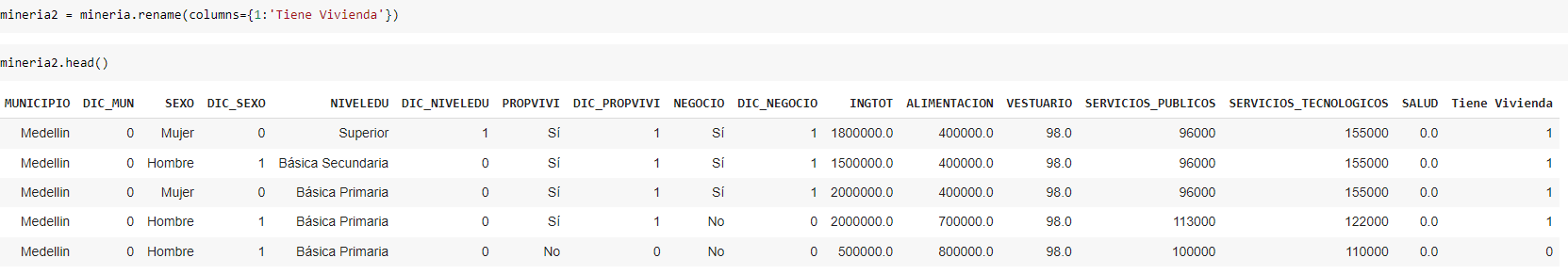
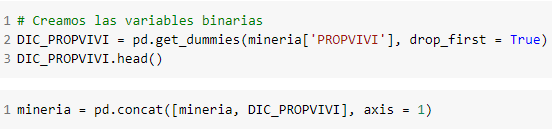
|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Tipo de Variable | Escala de medición | Descripción de la variable. |
| Municipio | Cualitativa | Nominal | Medellín  Bogotá  Cali |
| Sexo | Cualitativa | Nominal | Hombre  Mujer |
| Nivel educativo | Cualitativa | Ordinal | Ninguno  Preescolar  Básica primaria  Básica secundaria  Media  No sabe/No responde  Sin reporte |
| Vivienda propia | Cualitativa | Nominal | Si o No cuenta con vivienda propia |
| Negocio | Cualitativa | Nominal | Si o No cuenta con negocio propio |
| Ingreso total | Cuantitativa | Continua | Ingreso total por persona |
| Alimentación | Cuantitativa | Continua | Gasto en alimentación |
| Vestuario | Cuantitativa | Continua | Gasto en vestuario |
| Servicios públicos | Cuantitativa | Continua | Sumatoria de los gastos asociados a:  Luz  Agua  Gas |
| Servicios tecnológicos | Cuantitativa | Continua | Sumatoria de los gastos asociados a:  Celular  Internet  Telefonía |
| Salud | Cuantitativa | Continua | Gasto en servicios de salud |

# Comprensión de los datos

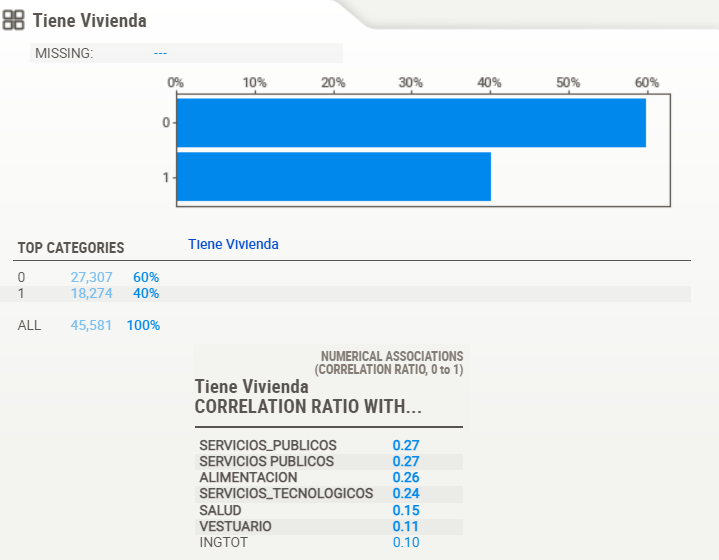
Se ingresa el archivo ajustado de la Encuesta de Carga Financiera y Educación Financiera de los Hogares 2018, donde se han seleccionado 11 variables, 6 cuantitativas y 5 cualitativas de las 331 preguntas que desarrolla la encuesta, para el caso tomaremos como variable Objetivo la variable “PROPVIVI” la cual hace referencia a si el hogar cuenta con vivienda propia o no.

****

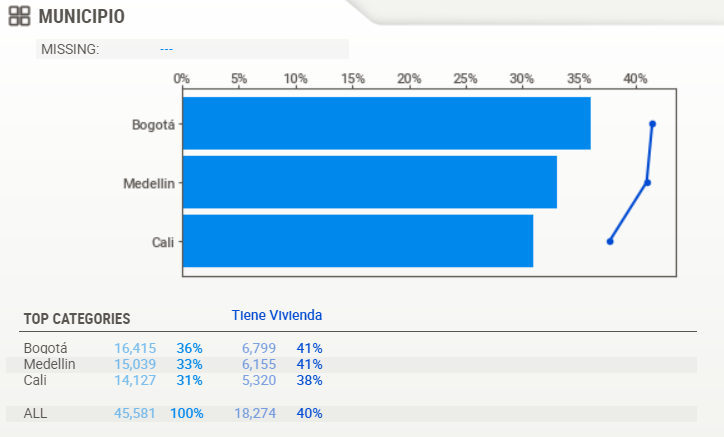
Inicialmente realizamos un proceso de transformación de de Variable “Porpviv” en una variable de tipo numérica 1 Tiene vivienda 0, No tiene vivienda Propia.



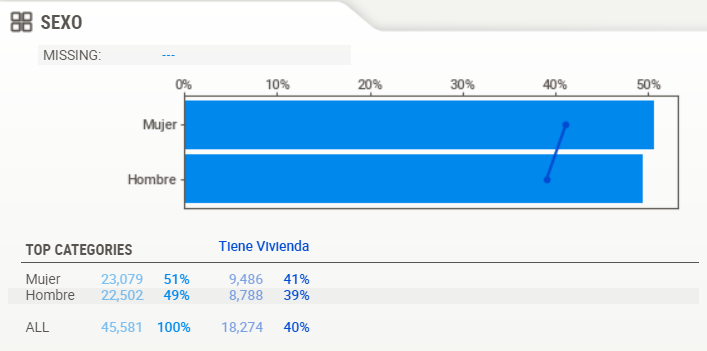
Encontramos como del total de encuestados solo el 40% cuenta con vivienda propia. La variable “Tiene Vivienda” no presenta correlación significativa con variables numéricas incluidas dentro del estudio.



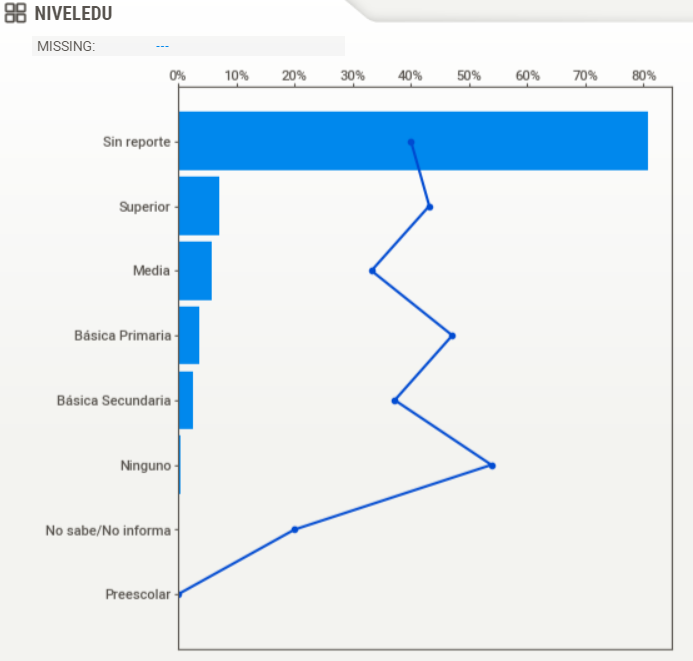
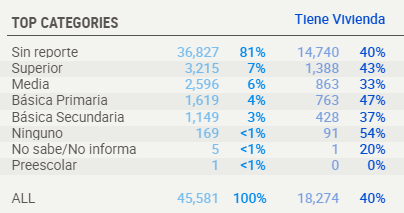
Los 45.581 encuestados se distribuyen dentro de las principales ciudades capital, Bogotá, Medellín, Cali.



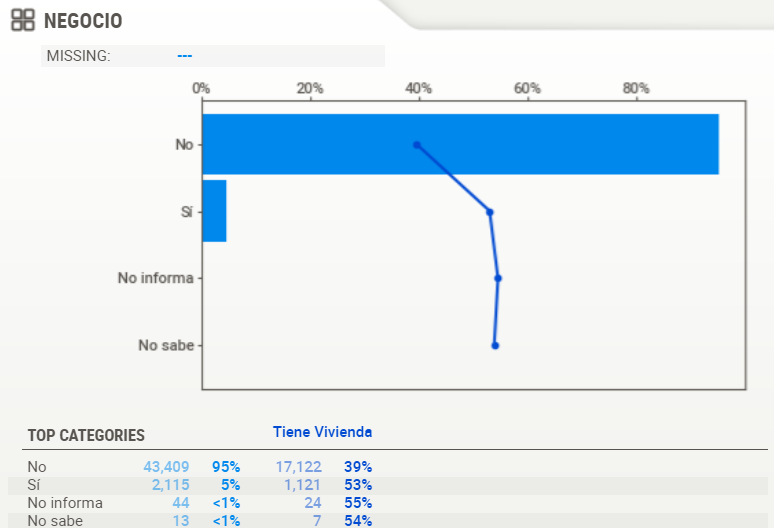
Encontramos la Particularidad que el porcentaje de Mujeres con vivienda propia se encuentra dos puntos porcentuales Por encima de los Hombres.



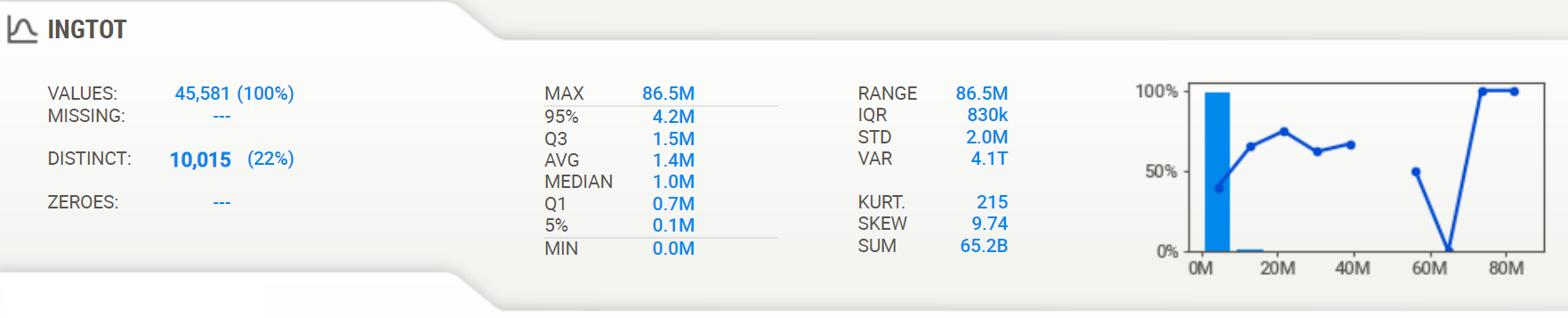
Encontramos un 81% de la población que no brinda reporte sobre el nivel de escolaridad. Como particularidad evidenciamos como el 54% de la población sin título académico cuenta con vivienda propia.



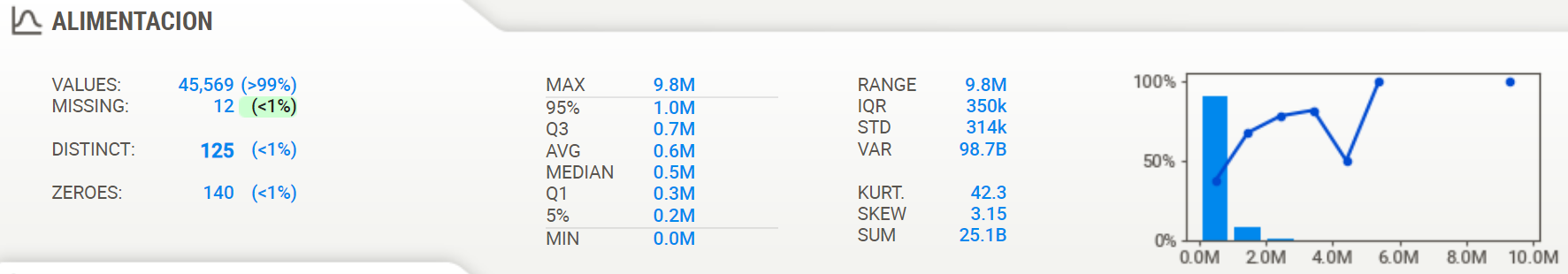
Solo el 5% de la población encuestada cuenta con negocio propio, donde el 53% de esta población tiene vivienda propia.



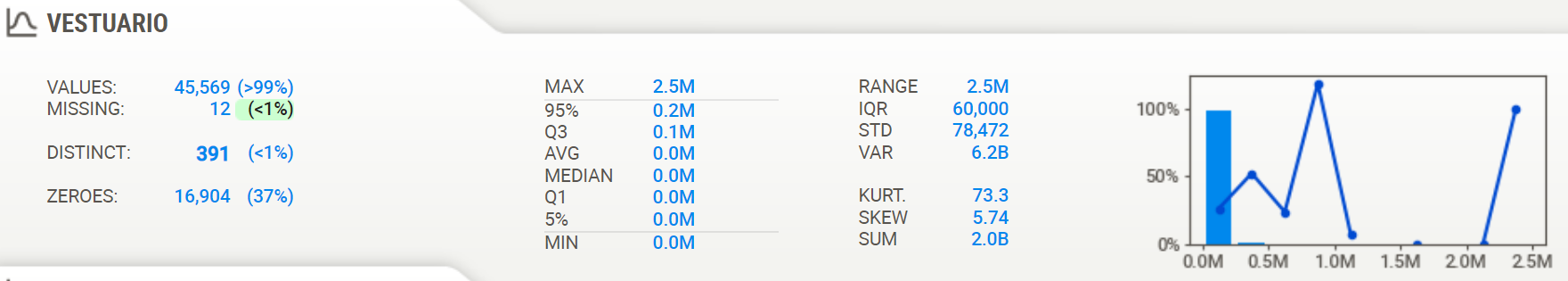
El ingreso promedio de los hogares encuestados se encuentra en 1.4 Millones, el 95% de la población encuestada se encuentra por debajo de los 4.5 millones.



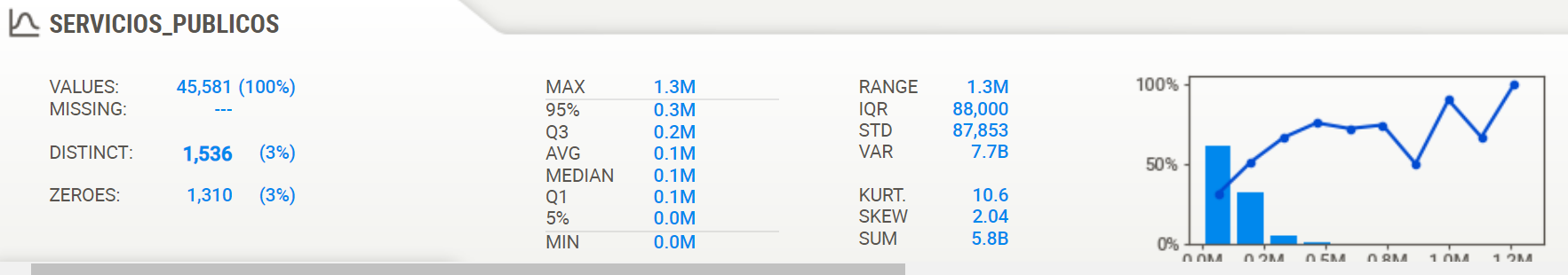
En promedio los hogares destinan 600.000$ en gastos de alimentación correspondientes al 42% del ingreso de los hogares. 95% de la población encuestada destina 1.000.000$ o menos.



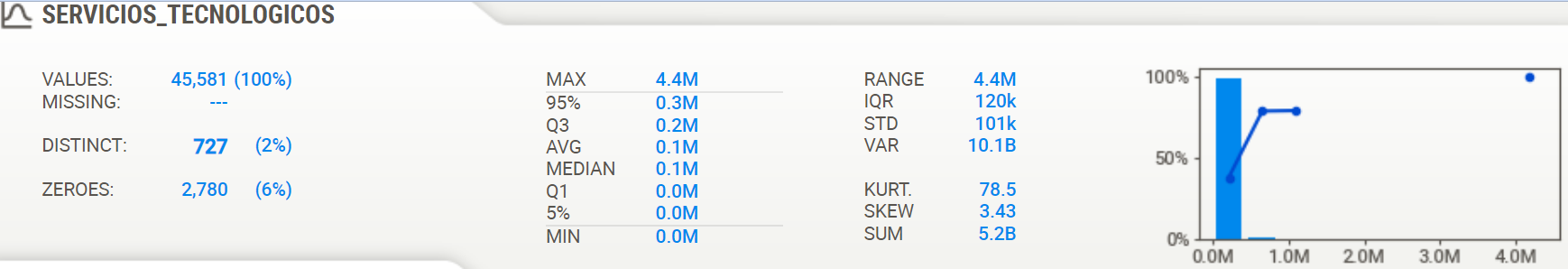
El gasto de los hogares en Vestuario se encuentra en promedio por debajo de los 100.000. El 95% de los hogares destinan menos de 200.000$ en vestuario al mes



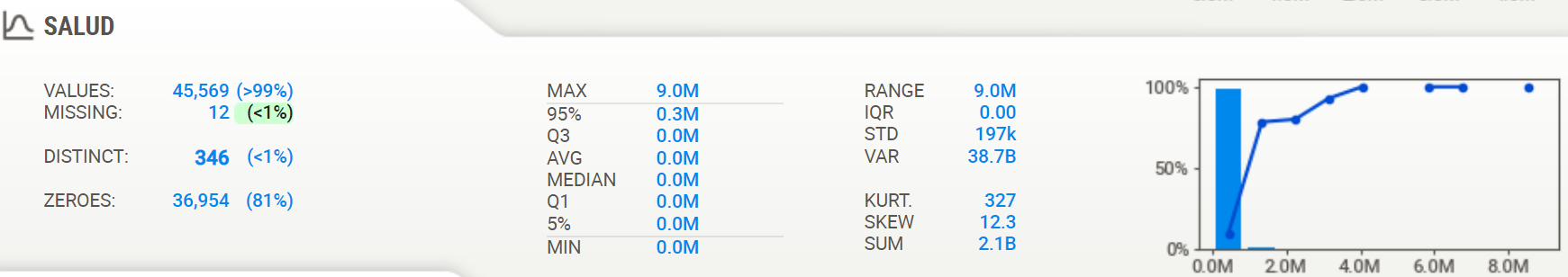
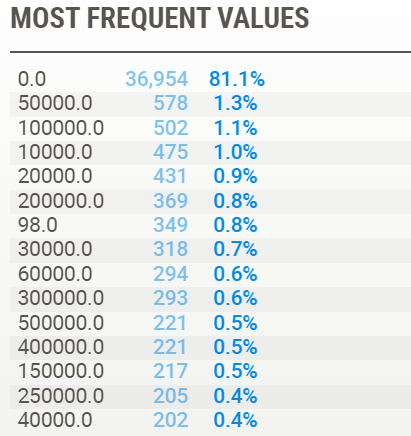
En promedio los hogares destinan 100.000$ en gastos de servicios Públicos 7% a 10% del Ingreso de los Hogares. 95% de los hogares destinan 300.000$ o menos a gastos en servicios públicos.



En promedio los hogares destinan 100.000$ en gastos de servicios Tecnológicos 7% a 10% del Ingreso de los Hogares. 95% de los hogares destinan 300.000$ o menos a gastos en servicios públicos.



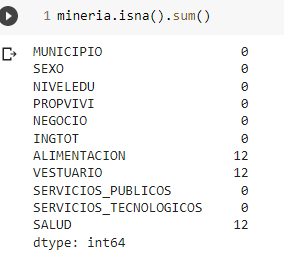
El 81% de los hogares no destinan un presupuesto mensual para gastos de salud.



# Preparación de los datos

# 7.1. Limpieza.

Al tratarse de un de una Base de datos proveniente de una encuesta DANE, los datos cuentan con un tratamiento previo por lo cual no encontramos datos faltantes N/A que resulten significativos.



# 7.2. Codificación Variables Categoricas

Contamos con un Total de 11 Variables(12 incluyendo codificación inicial de variable Propviv)Se Realizara la transformación de variables categóricas en Numéricas para la realización de Modelado

**Variables ordinales:** sus valores pueden ser ordenados jerárquicamente, como, por ejemplo, el nivel de educación de una persona (sin estudios, primaria, secundaria o superiores). Dentro de nuestra base de datos encontramos la variable “NIVELEDU” de tipo ordinal, donde podemos mediante técnicas de transformación de datos establecer un orden jerárquico donde el algoritmo establecerá 0 al nivel mínimo y una unidad adicional hasta el nivel máximo.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente



**Variables nominales:** no se puede establecer un orden en sus categorías. Algunos ejemplos son el sexo de una persona (hombre o mujer), el color del pelo (castaño, rubio, negro o rojo) o el país de nacimiento (España, Italia, Francia, etc.). Para las variables categóricas que no tienen un orden jerárquico utilizaremos la técnica One -hot. La anterior consiste en crear una nueva variable binaria (también llamada dummy) por cada categoría existente en la variable a codificar.  Así, estas nuevas variables contendrán 1 en aquellas observaciones que pertenezcan a esa categoría y 0 en el resto.



Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamenteInterfaz de usuario gráfica, Aplicación, Sitio web

Descripción generada automáticamente



**Variables numéricas:** Si bien tenemos completitud en la información, dentro de las variables categóricas encontramos valores atípicos que procederemos a ajustar en el proceso de modelado, estos valores se identifican dentro de la etapa de compresión de datos para cada una de las variables.



# 7.3. Datos Atipicos

**Datos atípicos:** Un inadecuado tratamiento de los datos atípicos durante el proceso tratamiento de los datos nos conduce a sesgos, pérdida de precisión y consecuentemente una incorrecta interpretación de los resultados. Posterior Se realizan eliminación de datos atípicos

Para el caso de la variable INGTOT identificamos gráficamente la existencia de Datos atípicos

Gráfico

Descripción generada automáticamente

Procedemos mediante el uso del IQR(rango intercuartílico) identificar los valores atípicos u outliers para proceder a su eliminación.

Texto

Descripción generada automáticamente

Encontramos un total de 4.634 outliers (10%)

El conjunto de datos resultante de la depuración pasa de 45.581 a 40.947

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Aplicamos los mismos criterios para eliminación de valores atípicos en variables Alimentación, vestuario, servicios públicos y servicios tecnológicos

Alimentación

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Servicios Públicos

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

Servicios Tecnológicos

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

# 7.4. Normalización de Variables

Para poder hacer uso de algoritmo de modelos de Machine Learning con funciones de distan cia para clasificación y segmentación, procedemos a normalizar los datos En este caso utilizaremos la función MinMaxScaler() que normaliza todos los datos entre [0, 1].



¿la fase 3 de la metodología?

¿Cuáles variables va a entra al modelo?

Hallazgo y concusiones