**Proyecto Final de Ingeniería de Datos**

**[Título por definir]**

**William Sebastián Cabrera Buitrago**

**Paola Andrea Cortés Montes**

**Andrés Felipe Miranda Mendoza**

**Sara Sofía Torres Méndez**

**Simón Vélez Castillo**

**Profesor: Javier Casas Salgado**

**Universidad del Rosario**

**Ingeniería de Datos**

**2024 - I**

**Descripción**

**Problema**

Se dispone de un conjunto de datos que contiene información demográfica y socioeconómica de Argentina en el 2019, sujeta a ciertas restricciones y reglas. El objetivo es realizar un análisis detallado de estos datos para responder a una serie de preguntas relacionadas con los ingresos, la educación, el estado conyugal, la situación laboral y otros aspectos relevantes de la población. El enfoque principal está orientado a entender los distintos factores que influyen en la obtención de salarios elevados, identificando así las variables determinantes en el panorama socioeconómico del país.

**Reglas**

* Una persona puede tener solo un estado conyugal a la vez
* En cada hogar debe haber por lo menos un miembro.
* En cada una casa(id) puede haber un hogar o más.
* Cada hogar solo debe tener un jefe.
* Los años de escolaridad no pueden ser mayor a la edad de la persona.
* Todos los ingresos de las personas son no negativos.
* El sexo de las personas solo puede ser mujer o varón.
* El id de cada casa es único.
* Cada miembro debe tener un hogar.
* Cada hogar debe tener una casa(Id).

**Preguntas**

* ¿Cuáles son los ingresos de las personas las cuales han terminado sus estudios universitarios a comparación de aquellas personas que no la han completado?
* ¿Cómo compara la situación laboral de la gente que vive con sus padres versus quienes no?
* ¿En promedio cuantos ingresos de familias cuya cabeza de hogar sea mujer superan el salario mínimo en argentina durante el 2019?
* ¿Como se comparan los ingresos familiares y el salario entre universidad privada y pública?
* ¿Qué lugar de nacimiento (de los presentes en la encuesta) cuenta con los mayores salarios?
* ¿Las familias más grandes tienden a tener una mejor, igual o peor educación que las familias más pequeñas?
* ¿Las personas con ingresos familiares per cápita superiores a la media tienen salarios superiores a la media?

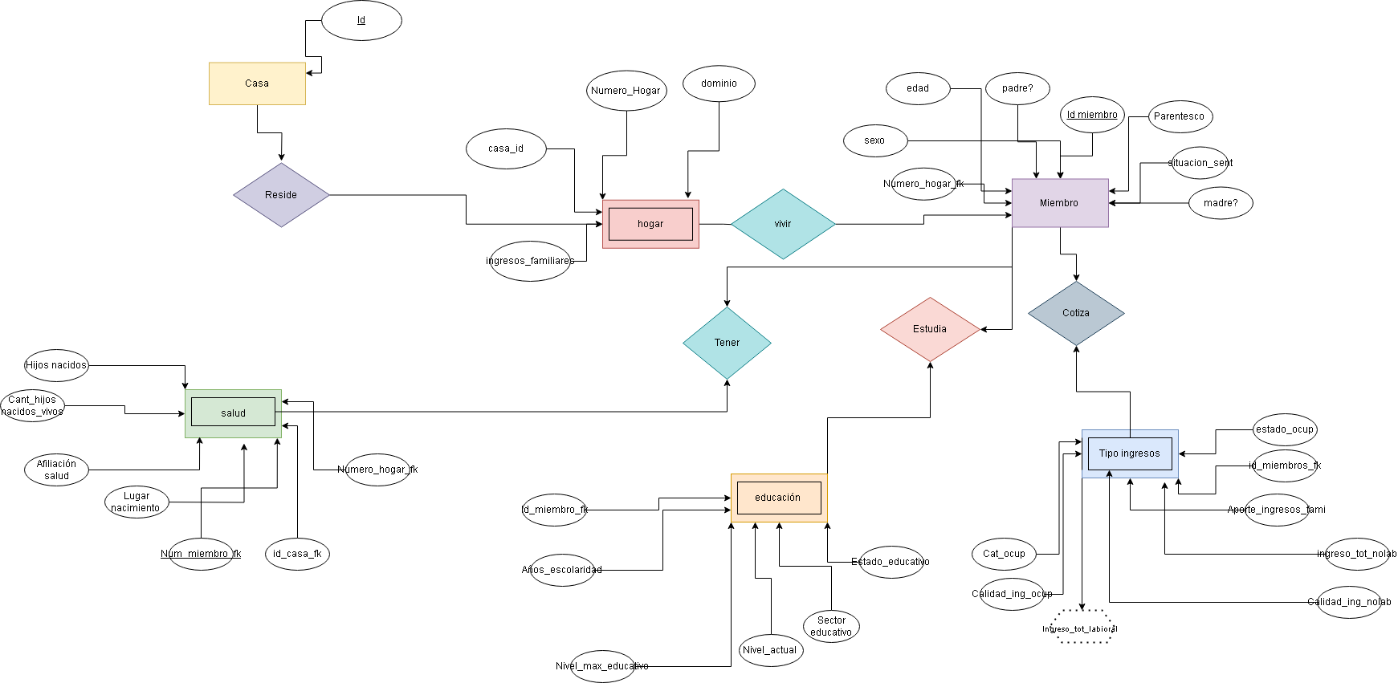
**Análisis.**

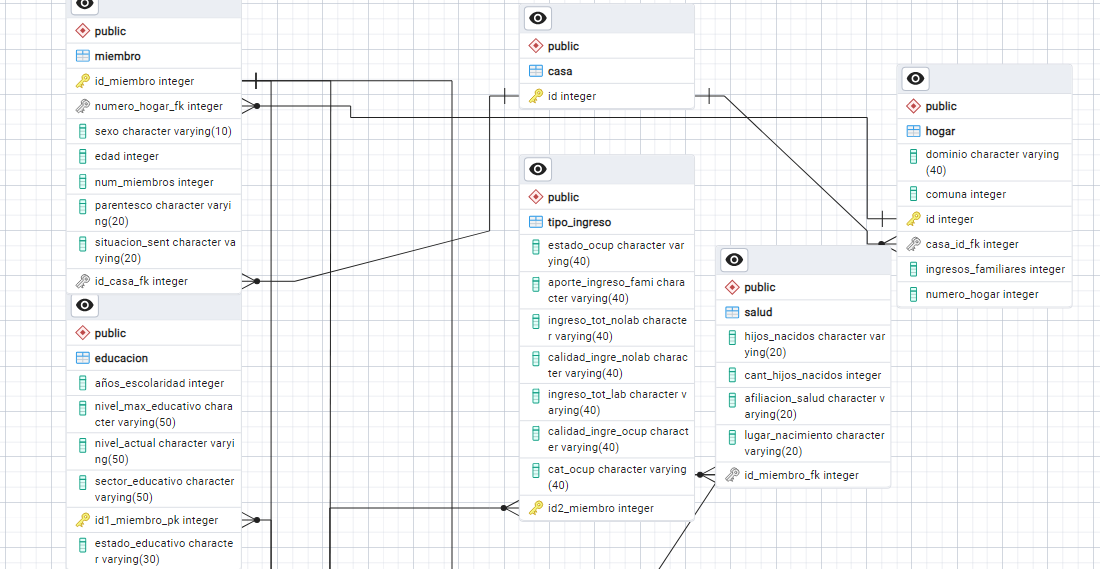
**.**

* Entidades:
  + Casa (Fuerte): se refiere al sitio físico donde pueden vivir uno o más nucleos familiares, sus atributos son ID.
  + Hogar (Débil): se refiere al núcleo familiar al que se realizó la encuesta; sus atributos son dominio, comuna, id, casa\_id\_fk, ingresos\_familiares, numero\_hogar,calidad\_ingresos familiares.
  + Miembro (Fuerte): se refiere a la persona que vive dentro del hogar;sus atributos son id\_miembro, numero\_hogar\_fk, sexo,edad,num\_miembros,parentesco,situacion\_sent,id\_casa\_fk
  + Tipo ingresos (Débil): da detalles de los ingresos familiares de un hogar; sus atributos son estado\_ocup, cat\_ocup, aporte\_ingreso\_fami, ingreso\_tot\_nolab, ingreso\_tot\_lab,calidad\_ingre\_ocup, calidad\_tot\_nolab, calidad\_ingresos\_totales,ingresos\_totales, ingreso\_per\_capita\_familiar, id2\_miemmbro\_pk
  + Educación (Débil): da información del nivel académico alcanzado por la persona, sus atributos son años\_Escolaridad, nivel\_max\_educativo,nivel\_actual, sector\_educativo, id1\_miembro\_pk, estado\_educativo
  + Salud (Débil):da información médica de la persona sus atributos son hijos\_nacidos,cant\_hijos\_nacidos, afiliacion\_salud, ligar\_nacimiento, id\_miembro\_fk.
* Atributos:
  + **Id:**describe el id de cada casa y es un entero. Pertenece a la entidad Casa
  + **nhogar**:este describe el número de hogar de cada casa correspondiente y es un entero. Pertenece a Hogar
  + **id\_miembro**:es el id de cada miembro y es un entero.
  + **miembro**: describe el número de miembro correspondiente a cada hogar.
  + **comuna**:describe el dominio de cada casa (id)y este puede contener los datos de “Villas de la emergencia” o “Resto de la ciudad” y es un string.
  + **sexo**:desribe el sexo binario de cada miembro y solo recibo los datos de “Mujer” y “Varón” y es un string.
  + **edad**: describe la edad de un miembro en años y este siempre es un número no negativo, es un entero.
  + **situacion\_conyugal**: describe el estado civil de cada miembro y este puede contener los datos de “Soltero/a”, “Viudo/a”, “Unido/a”, “No corresponde”, “Divorciado/a”; este es un string.
  + **Dominio**:el lugar donde se encuentra cada hogar y es un stsring que contiene “Resto de la Ciudad” o “Villas de emergencia”
  + **parentesco**: describe el rol de la persona en el hogar y es un string.
  + **num\_miembro\_padre**: describe el número del miembro del padre si vive en el hogar, si no vive en el hogar le otorgamos un -1 y si no corresponde -2, el tipo de dato es string.
  + **num\_miembro\_madre**: describe el número de miembro de la madre si vive en el hogar, si no vive en el hogar le otorgamos un -1 y si no corresponde -2, el tipo de dato es string.
  + **estado\_ocupacional**: contiene los datos de “Inactivo”, “Ocupado” y “Desocupado” y es un string.
  + **Cat\_ocupacional:** contiene la información de si la persona es o no asalariada, as un string
  + **calidad\_ingreso\_lab**: Indica la calidad de los ingresos de un individuo, lo que podría incluir aspectos como la estabilidad del empleo, la regularidad de los ingresos, entre otros. Es un string.
  + **calidad\_ingreso\_familiares**: Indica la calidad de los ingresos de un hogar, lo que podría incluir aspectos como la estabilidad del empleo, la regularidad de los ingresos, entre otros. Es un string.
  + **ingreso\_total:** Representa el ingreso total percibido por un individuo en un período determinado. Es un entero
  + **calidad\_ingresos\_nolab:** Similar a calidad\_ingreso, \_lab pero específicamente relacionado con ingresos no laborales, como rentas, intereses, etc. Es string.
  + **ngreso\_total\_nolab:** El ingreso total derivado de fuentes no laborales. Es entero.
  + **ingreso\_total\_lab:** El ingreso total derivado de fuentes laborales. Es entero.
  + **calidad\_ingresosTotales:** Es la calidad de los ingresos laborales y no laborales. Es un string.
  + **ingresos\_familiares:** Representa el total de ingresos percibidos por una familia en un período determinado. Es un entero.
  + **ingresos\_per\_capita\_familiar**: Los ingresos totales de la familia divididos por el número de miembros de la familia. Es un entero.
  + **años\_escolaridad**: describe la cantidad de años que ha estudiado la persona y es un entero.
  + **estado\_educativo**: Indica el estado del nivel educativo actual de un individuo, como asiste o no asiste etc. Es categórica.
  + **sector\_educativo**: Se refiere al tipo de institución educativa en la que está inscrito un individuo, como pública, privada, etc. Es un string.
  + **nivel\_actual**: El nivel educativo actual de un individuo, que podría ser grado, curso, año, etc. Es un string.
  + **nivel\_max\_educativo**: El nivel educativo máximo alcanzado por un individuo hasta el momento. Es categórica.
  + **cantidad\_hijos\_nac\_vivos:** El número de hijos que nacieron con vida de cada miembro y es un entero.
  + **hijos\_nacidos\_vivos**: Describe si tiene hijos o no y es un string.
  + **afiliacion\_salud**: Describe el tipo de afiliación que tiene cada miembro y es un string.
  + **Lugar\_nacimiento**: describe la locación de donde nació un miembro y es un string.
* Relaciones:
  + Reside: esta relación nos indica cuantos hogares hay por casa
  + Vivir: Describe cuantas personas hay por hogar
  + Cotiza: indica los ingresos de cada miembro
  + Estudia: indica el nivel académico de cada miembro
  + Tener: indica si cada miembro tiene plan de salud

Model**o Entidad Relación**

**Modelo Relacional**





Enlace base de datos

<https://docs.google.com/spreadsheets/d/1yvxhY7NS5FK6kLnpIqBCFv5bZn3UvFZzG7IOxGMUJy0/edit?usp=sharing>