**Documentación el modelado**

1. **Variables en base de datos original. Tengan en cuenta que, al describir esta base de datos, los nombres de las variables deben estar escritas en inglés, tal cual las encuentran en la base de datos original, mientras que las descripciones sí deben estar escritas en español. Esto permite a los otros miembros del grupo volver a la base original y reconocer las variables que están describiendo:** 
   1. **Tipo de variable (número, texto, etc.).**
   2. **Descripción de variable.**

**How old are you?**

***Edad***

Variable tipo texto que contiene rango de edades ej: 25-34, o 35-44, etc.

**What industry do you work in?**

***Profesión***

Variable tipo texto que describe la industria en la que trabaja el encuestado ejemplo: Computing or Tech, Nonprofits, etc.

**Job title**

***Titulo profesional***

Variable tipo texto que menciona el título del empleo/trabajo que desempeña el encuestado, ejemplo: Data developer, Manager, Program director, etc.

**If your job title needs additional context, please clarify here:**

***Observaciones del titulo***

Variable tipo texto que ofrece una descripción extra del titulo del trabajo de ser necesario, no se tienen todas las respuestas pues es opcional

**What is your annual salary? (You'll indicate the currency in a later question. If you are part-time or hourly, please enter an annualized equivalent -- what you would earn if you worked the job 40 hours a week, 52 weeks a year.)**

***Salario anual***

Variable tipo numérica que contiene el salario anual del encuestado, la pregunta también agrega que si el trabajo es a tiempo medio se debe responder con el equivalente anualizado de los ingresos si se trabajara 40 horas a la semana 52 semanas al año.

**How much additional monetary compensation do you get, if any (for example, bonuses or overtime in an average year)? Please only include monetary compensation here, not the value of benefits.**

***Compensación anual***

Variable tipo numérica que contiene la compensación adicional que el encuestado puede obtener ya sean bonos, o pagos por tiempo extra.

**Please indicate the currency**

***Tipo de moneda***

Variable tipo texto que contiene la divisa en la que se obtiene el salario anual y compensación adicional, ejemplo: USD, CAD, GBP, EUR, “other”, etc.

**If "Other," please indicate the currency here:**

***Otro tipo de moneda***

Variable tipo texto que contiene la divisa en la que se obtiene el salario anual y compensación adicional que pudo no estar en las opciones brindadas anteriormente y estaba en “other”, ejemplo: Peso argentino, INR (Indian Ruppee), etc.

**If your income needs additional context, please provide it here:**

***Observaciones sobre los ingresos***

Variable tipo texto que contiene contexto adicional acerca del ingreso del encuestado. ejemplo: tipo de bonos extra obtenidos, como funciona su salario, etc.

**What country do you work in?**

***País***

Variable tipo geográfica que contiene el país de trabajo del encuestado.

**If you're in the U.S., what state do you work in?**

***Estado de trabajo (EEUU)***

Variable tipo geográfica que pregunta al encuestado en caso de que el país de trabajo sea en EEUU en qué estado trabaja.

**What city do you work in?**

***Ciudad***

Variable tipo geográfica que contiene la ciudad en la que el encuestado trabaja.

**How many years of professional work experience do you have overall?**

***Experiencia laboral total***

Variable tipo texto que contiene los años de experiencia profesional total que tiene el encuestado, este valor está en rangos, ejemplo: 2-4 años, 5-7 años, 8-10 años, etc.

**How many years of professional work experience do you have in your field?**

***Experiencia laboral profesional total***

Variable tipo texto que contiene los años de experiencia profesional en su campo de trabajo que tiene el encuestado, este valor está en rangos, ejemplo: 2-4 years, 5-7 years, 8-10 years, etc.

**What is your highest level of education completed?**

***Nivel más alto de educación completada***

Variable tipo texto que contiene el máximo nivel de educación obtenido por el encuestado, ejemplo: Master’s Degree, PhD, Colege Degree, etc.

**What is your gender?**

Variable tipo texto que contiene el genero con el que el encuestado se identifica, ejemplo: Man, woman, non-binary, etc.

**What is your race? (Choose all that apply.)**

***Raza o grupo étnico***

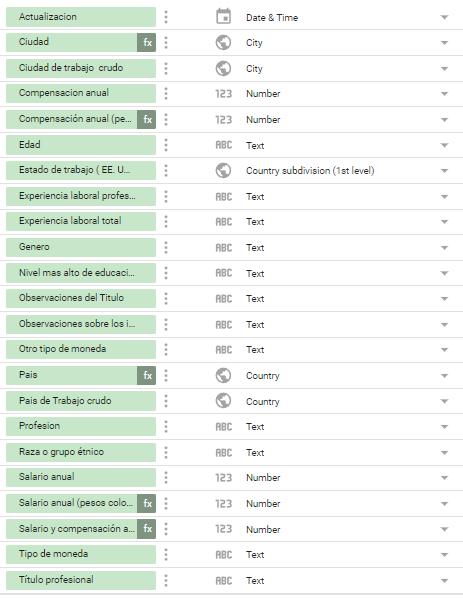
Variable tipo texto que contiene la raza con la que el encuestado se identifica, se pueden elegir varias que aplican, ejemplo: White, hispanic, latino, asian, etc.

1. **Variables luego de modeladas. Durante el modelado, ustedes pueden cambiar el nombre de las variables, las pueden dejar en inglés o traducirlas al español. En cualquier caso, por buena práctica, decidan qué idioma van a usar y sean consistentes con esa decisión en todas las variables:** 
   1. **Tipo de variable.**
   2. **Descripción corta, máximo 1 párrafo.**

Para todas las variables se decidió usar el español para el nombre de todas las variables, sin embargo, el valor de cada variable se mantuvo en su idioma para evitar más problemas de transformación, de esta manera las variables se llaman como se puede ver en el punto 1, pero sus valores si tienen valores en inglés, por ejemplo, para edad se menciona la palabra “years”, el género está en inglés (Man, Woman, etc.), lo mismo la raza, profesión y el título profesional.

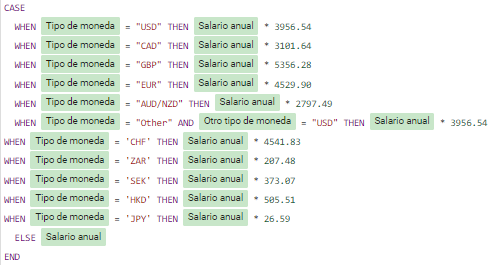
1. **Describir paso a paso lo que una persona debería hacer para actualizar los datos y aplicar el modelado que han diseñado. Piensen esta sección como una ayuda a alguien que los va a reemplazar en el trabajo. Esta persona debe entender con precisión cómo replicar lo que diseñaron. Este paso a paso asume que la estructura de la base de datos original no cambia en nuevas versiones. Puede ser tan detallada como ustedes quieran, pero debe ser muy claro cómo crear los nuevos campos y procedimientos para limpiar.**

La limpieza de estos datos fue hecha por medio de Looker Studio, por lo que todos Para realizar la limpieza primero fue necesario importar los datos, una vez hecho esto en Resource en la cinta superior de Looker Studio, darle Edit al “data source” cargado. Aquí dentro fue donde se hizo la limpieza. Primero se cambio el tipo de variable para todas las variables y el nombre.

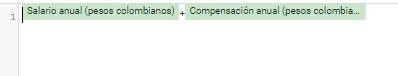


Primero se cambió los nombres a los mencionados en el punto 1 en español para que fuera mas comprensible y se cambio el tipo a las variables que fueran numéricas como salario, compensación, sus conversiones respectivas a pesos colombianos y la suma de ambas, también se cambió a variables tipo geográficas la ciudad, país y estado de EEUU. Como se puede ver en la foto se tiene “ciudad de trabajo crudo” y “país de trabajo crudo”, así como también “salario en pesos colombianos”, “compensación en pesos colombianos” y “salario y compensación en pesos colombianos”, estas variables existen porque se hizo uso de funciones calculadas usando “Add Field” dentro de este edit.

La creación de las variables que incluyen “… en pesos colombianos” se hizo con el mismo procedimiento, este fue tomar el valor de la variable original del salario y multiplicarlo por la divisa del tipo de moneda que es proporcionada por los datos, de esta manera se convierte el salario y la compensación en pesos colombianos como se ve a continuación:



Finalmente se sumó el salario anual y la compensación anual ambas en pesos colombianos para tener los ingresos totales anuales en pesos colombianos como se ve a continuación:



En el caso de la corrección de los países fue más complicado, para hacerlo se eliminaron espacios al principio y al final y se uniformaron todos los nombres en minúsculas. Luego, se revisó uno por uno para identificar aquellos que presentaban nombres diferentes, y se estandarizó a uno único, corrigiendo las diversas respuestas proporcionadas por los encuestados. Respecto a las ciudades, dado el gran número de estas, se llevó a cabo un análisis más detallado. En colaboración con el equipo, se decidió revisar los 600 registros más frecuentes, basándonos en las primeras cinco palabras de los registros de ciudad. Esta medida facilitó el agrupamiento y la toma de decisiones. Finalmente, se ajustaron los nombres, siguiendo el mismo procedimiento utilizado para los países.