

DOCUMENTACIÓN DEL MODELO

Fuente: <https://www.askamanager.org/2021/04/how-much-money-do-you-make-4.html>

Es una fuente que nos habla de los sueldos que tienen profesionales en diferentes disciplinas (principalmente en Estados Unidos) y estos datos son recolectados por "Ask a Manager"

Tiene:

Número de registros (filas): 28.190

Número de variables (columnas): 18

Formato: csv

Plataforma: Looker studio

Presentado por: Juan Carlor Gutiérrez Camperos

1. VARIABLES EN BASE DE DATOS ORIGINAL.

Análisis de variables

1. Timestamp

- Tipo: Fecha/Hora
- Descripción: Fecha y hora en que el encuestado diligenció el formulario. Útil para análisis temporales y control de duplicados.

2. How old are you?

- Tipo: Categórica (ordinal)
- Descripción: Rango de edad del encuestado (p. ej., 25–34). Permite segmentar salarios por grupo etario.

3. What industry do you work in?

- Tipo: Categórica (nominal)
- Descripción: Industria en la que trabaja el encuestado. Clave para comparaciones salariales sectoriales.

4. Job title

- Tipo: Texto
- Descripción: Cargo o puesto de trabajo reportado. Puede requerir normalización para análisis agregados.

5. If your job title needs additional context, please clarify here

- Tipo: Texto
- Descripción: Aclaraciones adicionales del cargo. Campo descriptivo, no estructurado.

6. What is your annual salary?

- Tipo: Numérica (entero)
- Descripción: Salario anual reportado en la moneda indicada. Variable central del análisis.

7. How much additional monetary compensation do you get?

- Tipo: Numérica (decimal)
- Descripción: Ingresos adicionales (bonos, comisiones). Puede contener ceros o nulos.

8. Please indicate the currency

- Tipo: Categórica (nominal)
- Descripción: Moneda del salario (USD, EUR, etc.). Requiere normalización para comparaciones.

9. If "Other," please indicate the currency here

- Tipo: Texto
- Descripción: Especificación de moneda cuando se selecciona "Other". Complementa la variable anterior.

10. If your income needs additional context, please explain here

- Tipo: Texto
- Descripción: Comentarios sobre condiciones del ingreso (medio tiempo, contratos, etc.).

11. What country do you work in?

- Tipo: Categórica (nominal)
- Descripción: País donde trabaja el encuestado. Requiere limpieza por variaciones de escritura.

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

- Tipo: Categórica (nominal)
- Descripción: Estado de EE. UU. Aplica solo para registros de ese país.

13. What city do you work in?

- Tipo: Texto
- Descripción: Ciudad de trabajo. Útil para análisis geográficos detallados.

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

- Tipo: Categórica (ordinal)
- Descripción: Años totales de experiencia profesional.

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

- Tipo: Categórica (ordinal)
- Descripción: Años de experiencia específica en el campo actual.

16. What is your highest level of education completed?

- Tipo: Categórica (ordinal)
- Descripción: Máximo nivel educativo alcanzado. Relevante para análisis de brechas salariales.

17. What is your gender?

- Tipo: Categórica (nominal)
- Descripción: Género reportado por el encuestado. Uso frecuente en análisis de equidad.

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

- Tipo: Categórica múltiple / Texto
- Descripción: Raza/etnia reportada. Puede requerir descomposición en variables binarias.

2. VARIABLES LUEGO DE MODELADAS

1. Ciudad(campo calculado)

- Tipo: Ciudad (Información geográfica)
- Descripción: Variable derivada que normaliza los valores de la columna “What city do you work in?” a un conjunto controlado de ciudades estándar. La normalización se realiza mediante reglas explícitas que agrupan múltiples variantes textuales bajo un único nombre canónico de ciudad (por ejemplo, “New York”, “Los Angeles”, “San Francisco”).

Nombre del campo	ID de campo
por ejemplo, Nuevo campo calculado Ciudad	ID de campo calc_pxbrhkeo0d
Fórmula	
<pre> 1 CASE 2 WHEN LOWER(TRIM(What city do you work in?)) IN ("new york","nyc","new york city") THEN "New York" 3 WHEN LOWER(TRIM(What city do you work in?)) IN ("los angeles","la","l.a.") THEN "Los Angeles" 4 WHEN LOWER(TRIM(What city do you work in?)) IN ("san francisco","sf","san fran","bay area") THEN "San Francisco" 5 WHEN LOWER(TRIM(What city do you work in?)) IN ("chicago","chi") THEN "Chicago" 6 WHEN LOWER(TRIM(What city do you work in?)) IN ("boston") THEN "Boston" 7 WHEN LOWER(TRIM(What city do you work in?)) IN ("seattle") THEN "Seattle" 8 WHEN LOWER(TRIM(What city do you work in?)) IN ("austin") THEN "Austin" 9 WHEN LOWER(TRIM(What city do you work in?)) IN ("miami") THEN "Miami" 10 WHEN LOWER(TRIM(What city do you work in?)) IN ("atlanta") THEN "Atlanta" 11 WHEN LOWER(TRIM(What city do you work in?)) IN ("denver") THEN "Denver" 12 </pre>	

2. País(campo calculado)

- Tipo: País (Información geográfica)
- Descripción: La variable País se construye a partir de la variable original “What country do you work in?”, el cual presenta alta variabilidad debido a que es de texto libre. Para su modelado, se aplican transformaciones de limpieza (eliminación de espacios y conversión a minúsculas) y reglas de normalización que agrupan diferentes formas de escritura de un mismo país bajo un nombre estándar. Este proceso reduce el ruido en los datos y permite realizar análisis agregados confiables por país.

Nombre del campo	ID de campo
por ejemplo, Nuevo campo calculado País	ID de campo calc_6u9oejeo0d
Fórmula	
<pre> 1 CASE 2 WHEN LOWER(TRIM(What country do you work in?)) IN ("us","usa","u.s.","u.s.a","united states","united states of america","america") THEN "United States" 3 WHEN LOWER(TRIM(What country do you work in?)) IN ("canada","ca") THEN "Canada" 4 WHEN LOWER(TRIM(What country do you work in?)) IN ("mexico","mexico","mx") THEN "Mexico" 5 6 WHEN LOWER(TRIM(What country do you work in?)) IN ("argentina","ar") THEN "Argentina" 7 WHEN LOWER(TRIM(What country do you work in?)) IN ("bolivia","bo") THEN "Bolivia" 8 WHEN LOWER(TRIM(What country do you work in?)) IN ("brazil","brasil","br") THEN "Brazil" 9 WHEN LOWER(TRIM(What country do you work in?)) IN ("chile","cl") THEN "Chile" 10 WHEN LOWER(TRIM(What country do you work in?)) IN ("colombia","co") THEN "Colombia" 11 WHEN LOWER(TRIM(What country do you work in?)) IN ("ecuador","ec") THEN "Ecuador" 12 WHEN LOWER(TRIM(What country do you work in?)) IN ("peru","peru","pe") THEN "Peru" </pre>	

3. Salario_anual(campo calculado)

- Tipo: Moneda(COP_pesos Colombianos)
- Descripción: La variable salario_anual se construye a partir del ingreso anual reportado por el encuestado y su respectiva moneda. Para su cálculo, se aplican factores de conversión que transforman los valores originales a una moneda base común, permitiendo comparaciones homogéneas entre registros. En los casos en que la moneda es reportada como "Other", se utiliza la moneda especificada manualmente por el usuario para determinar la tasa de conversión correspondiente. Este proceso mejora la consistencia y comparabilidad del dato salarial.

Nombre del campo	ID de campo
por ejemplo, Nuevo campo calculado : salario_anual	ID de campo calc_bu99fneo0d
Fórmula	
<pre>1 CASE 2 WHEN Currency = "USD" THEN Annual_salary * 3956.54 3 WHEN Currency = "CAD" THEN Annual_salary * 3181.64 4 WHEN Currency = "GBP" THEN Annual_salary * 5356.28 5 WHEN Currency = "EUR" THEN Annual_salary * 4529.98 6 WHEN Currency = "AUD/NZD" THEN Annual_salary * 2797.49 7 WHEN Currency = "CHF" THEN Annual_salary * 4759.21 8 WHEN Currency = "SEK" THEN Annual_salary * 488.69 9 WHEN Currency = "JPY" THEN Annual_salary * 23.59 10 WHEN Currency = "ZAR" THEN Annual_salary * 227.23 11 WHEN Currency = "HKD" THEN Annual_salary * 473.55 12 WHEN Currency = "Other" AND If 'Other', please indicate the currency ... = "USD" 13 THEN Annual_salary * 3956.54 14 ELSE Annual_salary 15 END</pre>	

4. Compensaciones(campo calculado)

- Tipo: Moneda(COP_pesos Colombianos)
- Descripción: La variable compensaciones se obtiene a partir del valor original de compensación reportado por el encuestado y su respectiva moneda. Para su modelado, se aplican factores de conversión que transforman los valores a una moneda base común, permitiendo realizar comparaciones homogéneas entre registros. Este proceso reduce las distorsiones generadas por el uso de distintas monedas y mejora la consistencia del análisis de compensaciones.

Nombre del campo	ID de campo
por ejemplo, Nuevo campo calculado : compensaciones	ID de campo calc_7bf641eo0d
Fórmula	
<pre>1 CASE 2 WHEN Currency = "USD" THEN comp * 3956.54 3 WHEN Currency = "CAD" THEN comp * 3181.64 4 WHEN Currency = "GBP" THEN comp * 5356.28 5 WHEN Currency = "EUR" THEN comp * 4529.98 6 WHEN Currency = "AUD/NZD" THEN comp * 2797.49 7 WHEN Currency = "CHF" THEN comp * 4759.21 8 WHEN Currency = "SEK" THEN comp * 488.69 9 WHEN Currency = "JPY" THEN comp * 23.59 10 WHEN Currency = "ZAR" THEN comp * 227.23 11 WHEN Currency = "HKD" THEN comp * 473.55 12 ELSE Annual_salary 13 END</pre>	

5. Ingresol_total(campo calculado)

- Tipo: Moneda(COP_pesos Colombianos)
- Descripción: Variable agregada que representa el ingreso total anual del encuestado, calculado como la suma del salario_anual normalizado y las compensaciones. Ambos componentes han sido previamente convertidos a una moneda base común, lo que permite una suma coherente y comparaciones válidas entre registros.

Nombre del campo por ejemplo, Nuevo campo calculado	ID de campo ID de campo calc_dx4a04eo0d
Fórmula 	

3. PASO A PASO PARA ACTUALIZAR DATOS Y APLICAR EL MODELADO

Paso 1. Obtener la nueva versión de los datos

- Descargar la nueva versión del dataset original (CSV, Excel o Google Sheets).
- Verificar que la estructura no cambie y que existan las siguientes columnas:
 - What city do you work in?
 - What country do you work in?
 - Please indicate the currency
 - If "Other," please indicate the currency here:
 - What is your annual salary?
 - How much additional monetary compensation do you get?

Paso 2. Normalización del país

- En Looker Studio, crear el campo calculado País.
- Aplicar limpieza básica al texto:
- Eliminación de espacios iniciales y finales (TRIM)
- Conversión a mayúsculas o minúsculas.
- Aplicar un diccionario de mapeo mediante CASE para agrupar variantes:
- USA, US, U.S. → United States
- England, Scotland, UK → United Kingdom

Paso 3. Normalización de la ciudad

- Crear el campo calculado Ciudad en Looker Studio.
- Limpiar el texto (TRIM + mayúsculas/minúsculas).
- Aplicar reglas de normalización:
- new york, nyc, new york city → New York

Paso 4. Conversión del salario anual a COP

- Crear el campo calculado Salario_anual.
- Implementar una estructura CASE que:

-Evalúe la variable Currency

-Multiplique annual_salary por la tasa de conversión correspondiente.

- Para el caso Currency = "Other", utilizar la variable
- If "Other," please indicate the currency here: para identificar la moneda real.
- Expresar el resultado final en pesos colombianos (COP).

Paso 5. Conversión de compensaciones a COP

- Crear el campo calculado Compensaciones.
- Repetir el procedimiento del Paso 4, pero usando la variable comp.

Paso 6. Cálculo del ingreso total

- Crear el campo calculado Ingreso_total.
- Calcularlo como la suma: Salario_anual + Compensaciones

Paso 7. Actualizar el informe

- Verificar que todas las visualizaciones usen las variables modeladas.
- Validar resultados con registros de prueba.
- Publicar o compartir el informe actualizado