

Documentado del modelado

Alumno Erick Andrés Pérez

El presente documento, detalla el modelado de los datos y la forma en que puede ser replicada por parte de un tercero. El dashboard elaborado puede ser visualizado en el siguiente link:

<https://lookerstudio.google.com/reporting/5403f6ee-9b81-4921-a887-3a20a91b693a>

1. Variables en base de datos original.

En primer lugar, se van a exponer una a una las variables originales obtenidas de la encuesta “Ask a Manager”¹, señalando el tipo de variables y la descripción de cada variable, según el modelo brindado.

I. Timestamp (esta oculta)

- a. **Tipo de variable:** Texto/Fecha-hora
- b. **Descripción:** Esta variable registra la fecha y hora en que se completó la encuesta

II. How old are you?

- a. **Tipo de variable:** Texto (categórica ordinal)
- b. **Descripción:** Va por rango de edad del encuestado, expresado en grupos etarios (por ejemplo: "25-34", "35-44", "45-54")

III. What industry do you work in?

- a. **Tipo de variable:** Texto (categórica nominal)
- b. **Descripción:** Es una variable de industria o sector en el que trabaja el encuestado (por ejemplo: "Education (Higher Education)", "Computing or Tech", "Accounting, Banking & Finance", "Nonprofits", "Publishing")

IV. Job title

- a. **Tipo de variable:** Texto

¹ Disponible en: <https://www.askamanager.org/2021/04/how-much-money-do-you-make-4.html>

- b. **Descripción:** Se incluye el título o cargo laboral del encuestado (por ejemplo: "Research and Instruction Librarian", "Change & Internal Communications Manager", "Marketing Specialist")

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

- a. **Tipo de variable:** Texto (respuesta abierta)
- b. **Descripción:** Campo opcional para proporcionar contexto adicional o aclaraciones sobre el título del trabajo cuando el encuestado así lo considere

VI. What is your annual salary?

- a. **Tipo de variable:** Número (entero/decimal)
- b. **Descripción:** Salario anual del encuestado, esta pregunta se expresa en la moneda que maneja el encuestado.

VII. How much additional monetary compensation do you get, if any?

- a. **Tipo de variable:** Número (entero/decimal)
- b. **Descripción:** Indica compensaciones monetarias adicionales al promedio anual, como bonos, horas extras, etc. No incluiría el valor de beneficios no monetarios que hubiese percibido, y como la variable anterior se encuentra expresada en la moneda que maneja el encuestado.

VIII. Please indicate the currency

- a. **Tipo de variable:** Texto (categórica nominal)
- b. **Descripción:** Se refiere a la moneda en la que se expresa el salario y la compensación adicional (por ejemplo: "USD", "GBP", "JPY", "EUR"), esta respuesta es cerrada.

IX. If "Other," please indicate the currency here:

- a. **Tipo de variable:** Texto (respuesta abierta)
- b. **Descripción:** Campo para especificar la moneda si no se encuentra en las opciones predeterminadas.

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

- a. **Tipo de variable:** Texto (respuesta abierta)

- b. **Descripción:** Campo opcional para proporcionar contexto adicional sobre los ingresos reportados.

XI. What country do you work in?

- a. **Tipo de variable:** Texto (categórica nominal)
- b. **Descripción:** País donde trabaja el encuestado (por ejemplo: "United States", "United Kingdom").

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

- a. **Tipo de variable:** Texto (categórica nominal)
- b. **Descripción:** Estado de Estados Unidos donde trabaja el encuestado (aplica solo si trabaja en Estados Unidos).

XIII. What city do you work in?

- a. **Tipo de variable:** Texto
- b. **Descripción:** Ciudad donde trabaja el encuestado

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

- a. **Tipo de variable:** Texto (categórica ordinal)
- b. **Descripción:** Años totales de experiencia laboral profesional, expresados en rangos (por ejemplo: "2 - 4 years", "5-7 years", "8 - 10 years", "21 - 30 years")

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

- a. **Tipo de variable:** Texto (categórica ordinal)
- b. **Descripción:** Años de experiencia laboral profesional específicamente en su campo o industria actual, expresados en rangos (por ejemplo: "2 - 4 years", "5-7 years", "8 - 10 years", "21 - 30 years")

XVI. What is your highest level of education completed?

- a. **Tipo de variable:** Texto (categórica ordinal)
- b. **Descripción:** Nivel educativo más alto completado por el encuestado (por ejemplo: "College degree", "Master's degree")

XVII. What is your gender?

- a. **Tipo de variable:** Texto (categórica nominal)

- b. **Descripción:** Identidad de género del encuestado (por ejemplo: "Woman", "Man", "Non-binary")

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

- a. **Tipo de variable:** Texto (categórica nominal, respuesta múltiple)
- b. **Descripción:** Raza o etnia del encuestado. Permite seleccionar múltiples opciones (por ejemplo: "White", "Hispanic, Latino, or Spanish origin")

2. Variables luego de modeladas

En el modelado se eliminó la variable con el “*Timestamp*”, a su vez se eliminaron las variables “*Please indicate the currency*” y las variables “*If Other, please indicate the currency here:*”, toda vez que todas las variables se transformaron a signo pesos. Se cambió la variable “*What is your annual salary?*” a “*Anual salary*” y “*How much additional monetary compensation do you get*” se transformó a “*Compensation*”. Se incluyó una nueva variable producto de la suma de “*Anual salary*” y “*Compensation*”, de nombre “*Anual Salary + Compensation*”. El resto de las variables se mantuvieron en su forma original luego de modeladas.

3. Paso a paso lo que se debe hacer para actualizar los datos y aplicar el modelado que se ha diseñado.

A continuación, se describe el paso a paso, de las actividades a adelantar para actualizar los datos y aplicar el modelado que se ha diseñado a los datos originales.

Paso 1: Abrir el archivo de datos original en el software de su preferencia (Excel, Python, R, etc.).

Paso 2: Eliminar las siguientes columnas del dataset:

- Columna "Timestamp"
- Columna "Please indicate the currency"
- Columna "If 'Other' please indicate the currency here:"

Paso 3: Renombrar las siguientes columnas:

- Cambiar el nombre de "What is your annual salary?" por "Annual Salary"
- Cambiar el nombre de "How much additional monetary compensation do you get" por "Compensation"

Paso 4: En la columna de país (identificar el nombre exacto de esta columna en la base de datos), aplicar las siguientes sustituciones:

Para Estados Unidos:

- Reemplazar "US" por "United States"
- Reemplazar "U.S." por "United States"
- Reemplazar "U.S" por "United States"
- Reemplazar "USA" por "United States"
- Reemplazar "U.S.A" por "United States"
- Reemplazar "U.S.A." por "United States"

Revisar errores ortográficos comunes (ej: "Unites States", "United State") y reemplazar por "United States"

Para Reino Unido:

- Reemplazar "UK" por "United Kingdom"
- Reemplazar "U.K." por "United Kingdom"
- Reemplazar "U.K" por "United Kingdom"

Revisar errores ortográficos comunes y reemplazar por "United Kingdom"

Para otros países:

- Reemplazar "Can" o "CAN" por "Canada"
- Reemplazar "Aus" o "AUS" por "Australia"

Revisar cada país de forma individual para identificar abreviaturas, errores ortográficos o variaciones en la escritura

Estandarizar usando el nombre oficial completo del país en inglés

Paso 5: Aplicar la conversión monetaria en la columna "Annual Salary":

Para registros en USD: multiplicar el valor por el TRM del día en que se hace la conversión, en nuestro caso fue \$3.760

Para registros en otras monedas: primero convertir a USD y luego multiplicar por el TRM, tener cuidado con la columna otras las monedas más comunes son renmibi,

rupias indias o indonesas, coronas danesas suecas o noruegas, baht tailandés y francos suizos.

Para registros ya en pesos colombianos se mantiene el valor original.

Fórmula: Nuevo valor = Valor original en USD × TRM

Paso 6: Aplicar la misma conversión en la columna "Compensation":

- Seguir el mismo procedimiento del Paso 5
- Si el valor es nulo o está en blanco, mantenerlo así

Paso 7: Calcular los valores de "**Annual Salary + Compensation**".

Si "Compensation" está vacía o es nula, considerar su valor como 0

Fórmula: Annual Salary + Compensation = Annual Salary + Compensation

Si "Compensation" está vacía o es nula, considerar su valor como 0

Ejemplo

Si Annual Salary = 100,000,000 COP y Compensation = 20,000,000 COP

Entonces Annual Salary + Compensation = 120,000,000 COP

Paso 8: Procesar la columna "*Annual Salary + Compensation*", eliminando outliers, en especial los valores muy bajos o altos que no correspondan a la muestra.

Paso 9: Verificar que todas las transformaciones se aplicaron correctamente:

- Confirmar que las columnas eliminadas ya no existen
- Confirmar que los nombres de las columnas renombradas son correctos
- Revisar una muestra aleatoria de países para verificar la estandarización
- Revisar una muestra aleatoria de conversiones monetarias

Paso 10: Verificar que no hay valores negativos o anómalos en las columnas monetarias.

Paso 11: Verificar que la nueva columna "*Annual Salary + Compensation*" se calculó correctamente para todos los registros y se eliminaron los outliers.

Paso 12: Guardar el archivo transformado con un nombre diferente al original (ej: "*datos_modelados.xlsx*" o "*datos_limpios.csv*").

Paso 13: Documentar la fecha de procesamiento y la TRM utilizada (3,760 COP/USD) para referencia futura.

Paso 14: El archivo resultante servirá para el análisis en looker studio y actualizar las tablas del tablero elaborado.