

Documentación de Modelado y Visualización

1. Descripción general del proyecto

Este documento describe el proceso de limpieza, modelado y visualización aplicado a la base de datos **Ask A Manager Salary Survey 2021**, con el objetivo de analizar patrones salariales globales y normalizar los ingresos a Pesos Colombianos (COP). El resultado es un conjunto de datos listo para análisis y un dashboard en Looker Studio que sirve como insumo interno para equipos de analytics.

2. Variables de la base de datos original

A continuación se describen las principales variables de la base original. Los nombres se mantienen **en inglés**, tal como aparecen en la fuente, mientras que las descripciones están en español.

Variables originales

- **Timestamp**

Tipo: Fecha

Descripción: Momento en el que la persona respondió el formulario.

- **How old are you?**

Tipo: Texto

Descripción: Rango de edad declarada por la persona encuestada.

- **What industry do you work in?**

Tipo: Texto

Descripción: Industria o sector económico en el que trabaja la persona.

- **Job title**

Tipo: Texto

Descripción: Cargo o título del puesto de trabajo.

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

Tipo: Texto

Descripción: Aclaraciones adicionales sobre el cargo.

- **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.)**

Tipo: Numérico

Descripción: Salario anual reportado por la persona, en su moneda local.

- **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.)**

Tipo: Numérico

Descripción: Compensaciones monetarias adicionales como bonos u horas extra.

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

Tipo: Texto

Descripción: Moneda en la que se reporta el salario diferente a las tradicionales.

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

Tipo: Texto

Descripción: Contexto adicional sobre los ingresos.

- **Please indicate the currency**

Tipo: Texto

Descripción: Moneda en la que se reporta el salario.

- **What country do you work in?**

Tipo: Texto

Descripción: País donde trabaja la persona.

- **What city do you work in?**

Tipo: Texto

Descripción: Ciudad donde trabaja la persona.

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

Tipo: Texto

Descripción: Rango de años totales de experiencia profesional.

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

Tipo: Texto

Descripción: Rango de años de experiencia en el sector específico.

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

Tipo: Texto

Descripción: Nivel educativo más alto alcanzado.

- **What is your gender?**

Tipo: Texto

Descripción: Género con el que se identifica la persona.

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

Tipo: Texto

Descripción: Identidad racial reportada por la persona.

3. Variables luego del modelado

Las variables modeladas mantienen nombres **en inglés** para consistencia con la base original. Todas las descripciones están en español.

- **annual_salary**

Tipo: Numérico

Descripción: Salario anual convertido a formato numérico, tras eliminar separadores de miles.

- **country_normalized**

Tipo: Texto

Descripción: País normalizado a una categoría estándar (por ejemplo, múltiples variantes de "United States" unificadas como "USA").

- **cop_rate**

Tipo: Numérico

Descripción: Tasa de cambio utilizada para convertir el salario a pesos colombianos.

- **annual_salary_cop**

Tipo: Numérico

Descripción: Salario anual expresado en pesos colombianos (COP).

- **additional_compensation_raw_cop**

Tipo: Numérico

Descripción: Compensaciones adicionales convertidas a COP.

- **total_salary_cop**

Tipo: Numérico

Descripción: Suma del salario anual y las compensaciones adicionales, ambos en COP.

- **fx_error**

Tipo: Booleano

Descripción: Indicador de error en la conversión por moneda no soportada, es decir monedas diferentes a USD, EUR, GBP, CHF, CAD y AUD/NZD.

- **years_experience_field**

Tipo: Texto

Descripción: Rango de años de experiencia específicamente en el campo, industria o rol actual reportados por el encuestado. Representa experiencia especializada, no trayectoria laboral total.

- **years_experience_field**

Tipo: Texto

Descripción: Rango de años experiencia laboral total acumulada, independientemente del campo o industria. Incluye experiencia previa en otros sectores.

4. Descripción del modelado de datos

El proceso de modelado se diseñó para minimizar supuestos y maximizar la reproducibilidad:

1. **Renombrado de columnas** para simplificar los nombres y facilitar el trabajo analítico.
 2. **Normalización de países y ciudades**, unificando múltiples variantes escritas libremente.
 3. **Limpieza de salarios:**
 4. Eliminación de separadores de miles.
 5. Conversión a tipo numérico.
 6. Identificación de valores menores a 100 en monedas principales (USD, EUR, GBP, CHF, CAD, AUD/NZD) y multiplicación por 1.000 cuando se interpreta como salario expresado en miles.
 7. **Conversión de moneda:**
 8. Solo se convierten monedas: USD, GBP, EUR, CHF, CAD y AUD/NZD.
 9. Se utiliza la TRM obtenida de Google Finance.
 10. **Tratamiento de outliers:**
 11. En visualizaciones se usan percentiles P10 y P90 en lugar de mínimos y máximos.
 12. Se excluyen países e industrias con menos de 10 observaciones para evitar anomalías.
-

5. Paso a paso para actualizar los datos

Este procedimiento describe de forma detallada cómo replicar el proceso de limpieza y modelado sobre versiones actualizadas de la base de datos, asumiendo que la estructura del archivo original no cambia.

Paso 1. Descarga de la fuente de datos

1. Acceder al Google Sheets oficial de *Ask a Manager Salary Survey 2021* mediante el enlace público.
2. Descargar manualmente el archivo en formato **CSV** desde Google Sheets.
3. Guardar el archivo localmente, manteniendo el nombre original o uno identificable por fecha.

Paso 2. Carga de datos en Python

1. Abrir el notebook de Python utilizado para el modelado.
2. Cargar el archivo CSV utilizando la librería `pandas`, verificando que el número de registros coincida con el archivo descargado.

Paso 3. Renombrado de columnas

1. Renombrar las columnas originales utilizando un diccionario definido directamente en el notebook de Python, donde cada nombre original en inglés se mapea a un nombre simplificado y estandarizado (por ejemplo, `annual_salary_raw`, `country_raw`, `city_raw`, etc.).
2. En este paso no se eliminan columnas; únicamente se renombran para facilitar el análisis posterior.

Paso 4. Normalización de país

1. Limpiar previamente el campo de país (`country_raw`) para estandarizar mayúsculas y eliminar espacios innecesarios.
2. Normalizar el país mediante un diccionario manual (lookup) definido en el código:
3. Múltiples variantes textuales se agrupan bajo categorías estándar como `USA` y `UK`.
4. Si el país no está contemplado en los diccionarios definidos, se conserva el valor original.
5. Crear el campo `country_normalized` como resultado de esta normalización.

Paso 5. Normalización de ciudad

1. Aplicar reglas de normalización al campo de ciudad (`city_raw`) mediante un diccionario manual y reglas de prioridad:
2. Primero se identifica si el registro corresponde a trabajo remoto y se asigna el valor `Remote`.
3. Luego se normalizan ciudades específicas como `New York` y `Washington DC`.
4. Las ciudades no contempladas se conservan, aplicando capitalización estándar (`title()`).
5. Crear el campo `city_normalized`.

Paso 6. Limpieza de variables salariales

1. Limpiar los campos salariales eliminando separadores de miles (comas).
2. Convertir los campos `annual_salary_raw` y `additional_compensation_raw` a tipo numérico.
3. Para las monedas **USD, EUR, GBP, CHF, CAD y AUD/NZD**, identificar valores menores a 100 e interpretarlos como salarios expresados en miles, multiplicándolos por 1.000.
4. Este ajuste no se aplica a monedas fuera de este conjunto.

Paso 7. Conversión de moneda (TRM)

1. Obtener manualmente la Tasa Representativa del Mercado (TRM) del día desde **Google Finance** para cada moneda soportada.
2. Registrar explícitamente la fecha de la TRM utilizada.
3. Convertir el salario anual y las compensaciones adicionales a Pesos Colombianos (COP).
4. Si la moneda del registro no se encuentra dentro del conjunto soportado, marcar el campo `fx_error = TRUE` y excluir estos registros del análisis posterior.

Paso 8. Creación de variables finales

1. Crear el campo `annual_salary_cop`.
2. Crear el campo `additional_compensation_raw_cop`.
3. Calcular el campo `total_salary_cop` como la suma del salario anual y la compensación adicional, ambos en COP.

Paso 9. Exportación del dataset

1. Guardar el dataframe final en formato **CSV** de manera local.
2. Utilizar este archivo como fuente de datos en Looker Studio, reemplazando el dataset anterior.

Paso 10. Validación en Looker Studio

1. Verificar que el total de registros sea consistente con el dataset exportado.
 2. Revisar visualmente los siguientes módulos del dashboard:
 3. **Módulo 1:** Contador total de respuestas.
 4. **Módulo 2:** Mapa por países (o ciudades, según configuración).
 5. **Módulo 3:** Relación entre industria y salario promedio en COP.
 6. **Módulo 4:** Relación entre experiencia profesional y salario en COP.
 7. Confirmar que filtros, rangos salariales y distribuciones no presenten valores anómalos.
-

6. Créditos

- **Fuente de datos:** Ask a Manager Salary Survey 2021
- **TRM:** Google Finance
- **Fecha de extracción TRM y Fuente de datos:** 07/02/2026

Este documento sirve como referencia interna para replicar, auditar y extender el análisis salarial presentado en el dashboard.