



# MIAD

Maestría  
en Inteligencia  
Analítica de Datos

# Dashboard 2: visualización y documentación para narrar a Un equipo de traba

Steven Vargas Riaño



## Descripción del tablero y hallazgos importantes

### Conclusiones Detalladas del Tablero de Visualización

El tablero titulado "¿Cuánto ganas en lo que haces?" presenta un análisis detallado de las respuestas de una encuesta sobre salarios, distribuidas por variables demográficas, geográficas, educativas y de industria. A continuación, se detallan las conclusiones derivadas de las visualizaciones mostradas:

---

#### 1. Datos Generales

- **Total de respuestas:** La base de datos contiene 28,339 encuestas, lo que representa un volumen significativo y suficiente para realizar análisis representativos.
- **Número de trabajos:** Se han registrado 12,141 trabajos diferentes, lo que refleja una amplia variedad de ocupaciones.
- **Cobertura geográfica:** La encuesta abarca 107 países, lo que proporciona una visión global de las condiciones salariales.

---

#### 2. Datos Salariales

- **Promedio salarial en COP:** El promedio salarial reportado es de **787 millones de pesos colombianos (COP)**, un valor que probablemente esté influido por trabajos de altos ingresos en la base de datos.
- **Promedio de compensación adicional:** Los encuestados reportan un promedio de **50 millones de COP** en compensaciones adicionales, que incluyen bonos, horas extras, entre otros beneficios.
- **Promedio de salario más compensaciones:** El ingreso total promedio, sumando salario base y compensaciones, asciende a **824 millones de COP**.

---

#### 3. Distribución por Edad y Género

- La mayor parte de los encuestados se encuentra en los rangos de edad **25–34** y **35–44 años**, lo que coincide con la población laboralmente activa más significativa.
- La mayoría de los participantes son mujeres, seguidos por hombres, con una pequeña proporción de personas que se identifican como no binarias u otras categorías.

---

#### 4. Distribución Geográfica

- El mapa de calor muestra una alta concentración de encuestados en **América del Norte**, **Europa** y algunos puntos de **Sudamérica**, mientras que regiones como **África** y partes de **Asia** tienen menor representación.
- Esto podría reflejar la ubicación de industrias de alto ingreso o una participación desigual en la encuesta.

---

#### 5. Nivel Educativo

- **Educación superior predominante:** El mayor porcentaje de encuestados tiene un **título universitario (13,651 personas)**, seguido de aquellos con **maestrías (8,934 personas)**.
  - **Otros niveles educativos:** Una proporción más baja tiene un título de **PhD (1,433 personas)**, **grados profesionales** como medicina o derecho (1,335 personas) o educación secundaria (**High School** con 649 personas).
  - Este patrón indica que el acceso a salarios más altos está correlacionado con niveles educativos avanzados.
-

## 6. Raza de los Participantes

- **Predominio de personas blancas:** El 87.48% de los encuestados se identifica como blancos, lo que sugiere una posible sobrerrepresentación de esta población.
- **Otras razas:** Una pequeña proporción incluye personas asiáticas, negras u otras razas, destacando la importancia de fomentar una mayor diversidad en futuras encuestas.

---

## 7. Relación entre Industria y Promedio Salarial

- **Industrias con mayores ingresos promedio:**
  - **Accounting Banking and finance:** Es la industria que sin duda registra los mejores salarios, por encima de **11 mil millones de COP.**
  - **Social networks:** Se observa que esta industria tiene los salarios más altos, con un promedio cercano a los **8 mil millones de COP.**
  - **Ipr:** También reporta un promedio alto, cercano a los **5 mil millones de COP.**
- **Industrias tradicionales con ingresos más bajos:**
  - **Accounting, Banking & Finance, Medical/Pharmaceutical y Petroleum** tienen salarios más bajos en comparación con las industrias tecnológicas.
- Este análisis muestra cómo las industrias tecnológicas y de redes sociales están liderando en términos de remuneración, en contraste con sectores más tradicionales.

---

## Conclusiones Clave

1. **Relación educación-ingresos:** Los salarios más altos están asociados con niveles educativos avanzados, como maestrías y doctorados.
2. **Desigualdad geográfica:** Las regiones con mayor desarrollo económico (América del Norte y Europa) concentran los ingresos más altos, mientras que otras regiones tienen una menor representación.
3. **Industria tecnológica dominante:** Los trabajos relacionados con redes sociales y tecnología generan los mayores ingresos, lo que resalta la transformación del mercado laboral hacia sectores digitales.
4. **Necesidad de diversidad:** La sobrerrepresentación de ciertas razas y géneros sugiere que futuras encuestas deben trabajar en ser más inclusivas para obtener una visión más equilibrada.

Este tablero brinda información clave para entender las dinámicas salariales globales y destaca la influencia de factores como la educación, la geografía y la industria en la remuneración laboral.

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:

nombre de la variable (Inglés)	Tipo de variable	Descripción (Español)
City_Refinada.Reemplazar	Texto	Ciudad refinada donde trabaja la persona.
City-What city do you work in?	Texto	Pregunta sobre la ciudad en la que trabaja la persona.
Country_Refinada.Reemplazar	Texto	País refinado donde trabaja la persona.
How many years of professional work experience do you have in your field?	Texto	Años de experiencia profesional en el campo de trabajo específico de la persona.
How many years of professional work experience do you have overall?	Texto	Años de experiencia profesional total en cualquier campo.
Suma de How much additional monetary compensation do you get?	Número	Suma de la compensación monetaria adicional anual (bonos, horas extras, etc.).
How old are you?	Texto	Rango de edad de la persona.
If you're in the U.S., what state do you work in?	Texto	Estado de los EE. UU. donde trabaja la persona, si aplica.
If "Other," please indicate the currency here:	Texto	Campo para especificar otra moneda si no es USD.
If your income needs additional context, please provide it here:	Texto	Campo para proporcionar contexto adicional sobre los ingresos.
If your job title needs additional context, please clarify here:	Texto	Campo para aclarar detalles adicionales sobre el puesto de trabajo.
Job title	Texto	Título del trabajo o puesto de la persona.
Please indicate the currency	Texto	Moneda en la que se encuentra expresado el salario.
Timestamp.1	Fecha/Hora	Fecha y hora del registro inicial de la encuesta.
Timestamp.2	Fecha/Hora	Fecha y hora adicional del registro de la encuesta.
What industry do you work in?	Texto	Industria en la que trabaja la persona (por ejemplo, educación, manufactura, etc.).
Suma de What is your annual salary?	Número	Salario anual equivalente, considerando 40 horas semanales y 52 semanas al año.
What is your gender?	Texto	Género de la persona.
What is your highest level of education completed?	Texto	Nivel educativo más alto completado por la persona (por ejemplo, título universitario, maestría, etc.).
What is your race? (Choose all that apply.)	Texto	Raza o etnia seleccionada por la persona.

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:

Nombre de la variable (Inglés)	Tipo de variable	Descripción (Español)
City_Refinada.Reemplazar	Texto	Ciudad refinada donde trabaja la persona.
Suma de compensaciones COP	Número	Suma total de compensaciones adicionales en pesos colombianos (COP).
Suma de Índice	Número	Índice que suma algún valor asociado a las respuestas.
Suma de PROMEDIO	Número	Valor promedio calculado (sin unidades específicas).
PROMEDIO SALARIOS	Número	Promedio de los salarios en la base de datos.
Suma de Salario + Compensacion	Número	Suma del salario anual más compensaciones monetarias adicionales.
Suma de salario_anual COP	Número	Salario anual total expresado en pesos colombianos (COP).
Timestamp.1	Fecha	Fecha del registro inicial de la encuesta.
Timestamp.2	Hora	Hora adicional del registro de la encuesta.
Suma de TRM	Número	Tasa Representativa del Mercado (TRM) utilizada en el cálculo del salario en COP.

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.

## Fases del Proceso

### 1. Extracción

El archivo fuente es un CSV que se encuentra en una ubicación específica del sistema local. El archivo contiene 18 columnas y utiliza coma (,) como delimitador. Se especifica la codificación UTF-8 para garantizar la correcta interpretación de los caracteres especiales.

---

### 2. Promoción de Encabezados

Se toma la primera fila del archivo como encabezados para las columnas, estableciendo los nombres que describen los datos de forma más legible y comprensible.

---

### 3. Limpieza de Datos

- **Reemplazo de valores:** Se eliminan espacios en blanco adicionales en las columnas relevantes, como la que contiene información sobre los salarios, para estandarizar los datos.
- **Filtrado de valores no deseados:** Se eliminan filas que contienen valores irrelevantes o erróneos en las columnas principales, como Timestamp.

---

### 4. Transformaciones de Columnas

- **Asignación de tipos de datos:**
  - Las columnas con datos textuales (por ejemplo, "Ciudad", "País" y "Título del trabajo") se configuran como tipo texto.
  - Las columnas con valores numéricos (por ejemplo, "Salario anual" y "Compensación adicional") se configuran como tipo número.
- **División de columnas:**
  - La columna Timestamp, que contiene fecha y hora, se divide en dos columnas: una para la fecha (Timestamp.1) y otra para la hora (Timestamp.2).

---

### 5. Estandarización de Texto

- Los valores en la columna What country do you work in? se transforman para que cada palabra inicie con mayúscula, facilitando la lectura y el análisis.

---

### 6. Filtrado Avanzado

Se eliminan filas con valores específicos no deseados, como errores o inconsistencias

en columnas clave, asegurando que los datos restantes sean útiles para el análisis.

---

## 7. Uniones de Tablas

Se realizan dos uniones externas (tipo Left Outer Join) para enriquecer los datos:

### 1. Unión con la tabla de países (Country\_Refinada):

- Se agrega información refinada de los países basada en coincidencias en la columna Country-What country do you work in?.

### 2. Unión con la tabla de ciudades (City\_Refinada):

- Se agrega información refinada de las ciudades basada en coincidencias en la columna City-What city do you work in?.

---

## 8. Reorganización de Columnas

Las columnas se reordenan para seguir una estructura lógica, priorizando aquellas que son más relevantes para el análisis, como fechas, títulos de trabajo y compensaciones monetarias.

---

## 9. Creación de Índices

Se añade una columna de índice para numerar de manera única cada fila, facilitando referencias y análisis posteriores.

Anexos

# Tabla TRM:

## 1. Extracción

- **Fuente de datos:** Se utiliza una API JSON, accesible a través de una URL pública: <https://www.datos.gov.co/resource/32sa-8pi3.json>.
- **Método de extracción:**
  - Se realiza una solicitud HTTP para obtener los datos directamente desde la API.
  - Los datos extraídos se convierten a un formato interno de Power Query para su procesamiento.

---

## 2. Conversión de Datos

- Los datos extraídos del JSON se convierten en una tabla para facilitar las transformaciones posteriores. Cada elemento del JSON se desglosa en filas de una tabla.

---

## 3. Expansión de Columnas

- **Descripción:**
  - Se descomponen los registros anidados de la columna principal (Column1) en múltiples columnas:
    - valor: Representa el valor relevante del dato.
    - unidad: Indica la unidad de medida asociada al valor.
    - vigenciadesde: Marca el inicio de la vigencia del dato.



- vigenciahasta: Representa el fin de la vigencia del dato.
  - **Resultado:**
    - Cada columna ahora contiene un atributo único del registro original.
- 

#### 4. Cambio de Tipos de Datos

- **Descripción:**
    - Se ajustan los tipos de datos de las columnas:
      - unidad: Tipo texto (type text).
      - vigenciadesde y vigenciahasta: Tipo fecha y hora (type datetime).
  - **Motivo:** Asegurar que los datos sean interpretados correctamente para análisis posteriores.
- 

#### 5. Eliminación de Columnas Irrelevantes

- **Descripción:**
    - Se elimina la columna vigenciahasta porque no es relevante para el análisis.
  - **Resultado:**
    - La tabla queda simplificada con las columnas necesarias: valor, unidad y vigenciadesde.
- 

#### 6. Filtrado de Filas

- **Descripción:**
    - Se filtran los datos para conservar únicamente los registros cuya columna vigenciadesde coincide con la fecha 2025-02-07.
  - **Motivo:**
    - Esto permite trabajar exclusivamente con los datos correspondientes a una fecha específica, ajustada a los requisitos del análisis.
- 

#### 7. Limpieza de Valores

- **Descripción:**
    - Se reemplazan los puntos (.) en los valores numéricos de la columna valor por comas (,), cumpliendo con el formato numérico estándar utilizado en muchos países hispanohablantes.
  - **Herramienta:**
    - Utiliza la función de reemplazo de texto para realizar esta operación.
- 

#### 8. Cambio de Tipo de Datos (Final)

- **Descripción:**
  - La columna valor se convierte a tipo numérico (type number), asegurando que pueda ser utilizada para cálculos y análisis cuantitativos.
- **Resultado:**
  - Los datos numéricos están listos para operaciones matemáticas o estadísticas.

## 9. Resultados Finales

El proceso final produce una tabla estructurada con las siguientes características:

- **Columnas:**
  - valor: Representa el valor numérico relevante para análisis.
  - unidad: La unidad de medida asociada al valor.
  - vigenciadesde: Fecha de inicio de la vigencia del dato.
- **Filtrado:** Solo contiene datos correspondientes a la fecha 2025-02-07.
- **Formato numérico:** Los valores están en un formato estándar compatible con operaciones matemáticas.

✕

✓

fx

= Table.TransformColumnTypes(#"Valor reemplazado",{{"valor", typ

	1.2 valor	AB <sub>C</sub> unidad	vigenciadesde
1	4150,99	COP	7/02/2025 12:00:00 a. m.

## TABLA MONEDAS

### 1. Extracción

- **Fuente de datos:** Los datos se cargan desde un objeto JSON comprimido que contiene una tabla con información de monedas. Este objeto se descomprime y convierte en una tabla para su procesamiento.
- **Columnas iniciales:**
  - Moneda: Nombre completo de la moneda (por ejemplo, "Dólar estadounidense").
  - Abreviatura: Código internacional de la moneda (por ejemplo, "USD").
  - Equivalente en USD: Valor de la moneda en dólares estadounidenses.
  - Equivalente en COP: Valor de la moneda en pesos colombianos.

### 2. Eliminación de Columnas

- **Descripción:**
  - Se elimina la columna Equivalente en COP, ya que no es relevante para los cálculos en esta etapa.
- **Motivo:**
  - Reducir la tabla a las columnas necesarias para el análisis basado en dólares estadounidenses (USD).

### 3. Limpieza de Valores

- **Descripción:**
  - En la columna Equivalente en USD, se reemplazan los puntos (.) por comas (,).
- **Motivo:**
  - Ajustar los valores al formato numérico estándar utilizado en algunos

países de habla hispana, donde las comas representan decimales.

#### 4. Cambio de Tipos de Datos

- **Descripción:**
  - La columna Equivalente en USD se transforma al tipo numérico (type number).
- **Motivo:**
  - Asegurar que los valores de la columna puedan ser utilizados en operaciones matemáticas o cálculos posteriores.

AB Moneda	AB Abreviatura	1.2 Equivalente en USD
Dólar estadounidense	USD	1
Libra esterlina	GBP	1,31
Dólar canadiense	CAD	0,75
Euro	EUR	1,1
Dólar australiano	AUD	0,7
Dólar neozelandés	NZD	0,65
Rupia india	INR	0,01
Peso argentino	ARS	0,01
Franco suizo	CHF	1,1
Ringgit malayo	MYR	0,24
Rand sudafricano	ZAR	0,07
Corona sueca	SEK	0,11
Dólar de Hong Kong	HKD	0,13
Corona noruega	NOK	0,12
Real brasileño	BRL	0,2
Corona danesa	DKK	0,15
Dólar de Trinidad y Tobago	TTD	0,15
Peso mexicano	MXN	0,05
Corona checa	CZK	0,05
Taka bangladesí	BDT	0,01
Peso filipino	PHP	0,02
Zloty polaco	PLN	0,26
Lira turca	TRY	0,07
Yuan chino	CNY	0,16
Shekel israelí	ILS	0,29

### Tabla Ciudad (City)

#### 1. Extracción

- **Fuente de datos:** El archivo CSV se encuentra en una ubicación local especificada.
- **Características del archivo:**
  - Delimitador: El archivo utiliza punto y coma (;) como separador entre columnas.
  - Número de columnas: Tres columnas.

- Codificación: UTF-8 para garantizar la correcta interpretación de caracteres especiales.

## 2. Transformaciones

### 2.1. Cambio de Tipos de Datos Inicial

- Se asignan tipos de datos preliminares a las columnas:
  - Column1: Tipo texto (type text).
  - Column2: Tipo texto (type text).
  - Column3: Tipo número entero (Int64.Type).

### 2.2. Promoción de Encabezados

- La primera fila del archivo se promueve como encabezados, lo que convierte los nombres de las columnas en:
  - City- What city do you work in?: Nombre de la ciudad en la que trabaja la persona.
  - Reemplazar: Texto para realizar reemplazos o modificaciones específicas.
  - 1: Valor numérico asociado, posiblemente un índice o dato cuantitativo.

### 2.3. Cambio de Tipos de Datos Final

- Se aseguran los tipos de datos correctos para las columnas después de la promoción de encabezados:
  - City- What city do you work in?: Tipo texto (type text).
  - Reemplazar: Tipo texto (type text).
  - 1: Tipo número entero (Int64.Type).

## 3. Resultados Finales

### 3.1. Columnas Resultantes

- **City- What city do you work in?:**
  - Tipo: Texto.
  - Descripción: Ciudad en la que trabaja la persona, utilizada para análisis geográficos o categorizaciones.
- **Reemplazar:**
  - Tipo: Texto.
  - Descripción: Campo de texto que podría ser utilizado para realizar modificaciones, reemplazos u otras transformaciones en los datos.
- **1:**
  - Tipo: Número entero.
  - Descripción: Representa un valor cuantitativo, posiblemente un índice o una métrica asociada.

### 3.2. Tabla Procesada

La tabla final está estructurada, limpia y lista para integrarse en herramientas analíticas. Se ha asegurado la consistencia en los tipos de datos y el formato de las columnas.

	A <sub>C</sub> City- What city do you work in?	A <sub>C</sub> Reemplazar	t <sub>3</sub> 1
1		No Aplica	2
2	-	No Aplica	3
3	---	No Aplica	4
4	----	No Aplica	5
5	-----	No Aplica	6
6	(Anonymous)	No Aplica	7
7	(Province@of@Alberta)	No Aplica	8
8	(Rural)	No Aplica	9
9	**Too@small@a@sample@size,@can t@share	No Aplica	10
10	.	No Aplica	11
11	..@	No Aplica	12
12	/	No Aplica	13
13	@	No Aplica	14
14	@@	No Aplica	15
15	@@@	No Aplica	16
16	@Berkley@Heights@	Berkley	17
17	@Boston@	Boston	18
18	@Byron@Bay	Byron	19
19	@Charlottesville	Charlott	20
20	@Chicago	Chicago	21
21	@Chicago@	Chicago	22
22	@Los@Angeles	Los Angeles	23
23	@N/A@	No Aplica	24
24	@Rural@Florida@Panhandle	No Aplica	25
25	@Ventura@	No Aplica	26