


# Guía para Actualización y Modelado de Datos de Salarios

Este documento describe el paso a paso para actualizar los datos y aplicar el modelado diseñado. La persona que lo use debe seguir estas instrucciones en orden, asumiendo que la estructura de la base de datos original se mantiene en versiones futuras.

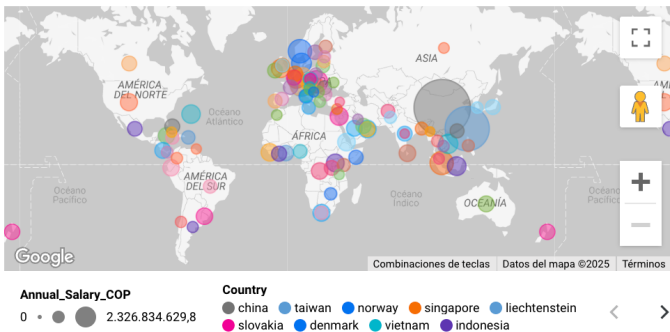
URL del tablero: <https://lookerstudio.google.com/reporting/ffe7abd1-a115-47e0-8f89-5c625b108442>



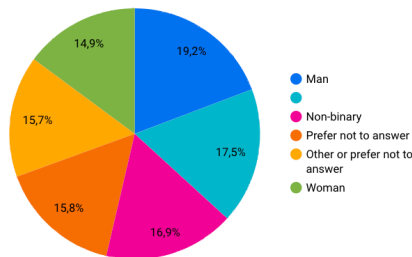
Analisis Ingresos - Encuesta salarial 2021

Total de respuestas  
28.112

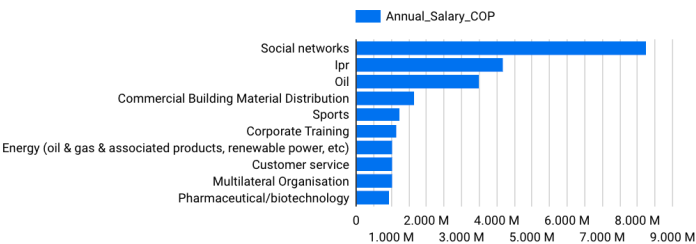
Salarios anuales promedio por pais - Moneda COP



Salarios anuales promedio por genero



TOP 10 Industrias con mayores salarios promedio al año



Se tiene un total de encuestados de 28.112 personas. Podemos observar de los diferentes graficos:

- Diferentes paises en los cuales viven las personas encuestadas asi como su salario promedio anual en pesos colombianos.
- Distribucion salarial promedio por genero. Valores en pesos colombianos.
- Top 10 de industrias con salarios promedio mas altos. Valores en pesos Colombianos.

Base de datos TRM: <https://www.xe.com/es/currencyconverter/convert/?from=USD&to=COP> Consultar valores TRM el día 07 Feb 2025

## Variables en base de datos original.

Variable	Tipo	Descripción
How old are you?	Texto	Rango de edad de la persona encuestada.
What industry do you work in?	Texto	Industria en la que trabaja la persona encuestada.
Job title	Texto	Título del puesto de trabajo de la persona encuestada.

<b>If your job title needs additional context, please clarify here:</b>	Texto	Espacio para aclaraciones adicionales sobre el título del puesto.
<b>What is your annual salary?</b>	Número	Salario anual de la persona encuestada, convertido a un equivalente a tiempo completo.
<b>How much additional monetary compensation do you get, if any?</b>	Número	Compensación monetaria adicional recibida en promedio al año (por ejemplo, bonos u horas extras). No incluye beneficios.
<b>Please indicate the currency</b>	Texto	Moneda en la que está expresado el salario y la compensación adicional.
<b>If "Other," please indicate the currency here:</b>	Texto	Si la moneda seleccionada no está en la lista, se especifica aquí.
<b>If your income needs additional context, please provide it here:</b>	Texto	Campo para proporcionar contexto adicional sobre los ingresos.
<b>What country do you work in?</b>	Texto	País donde trabaja la persona encuestada.
<b>If you're in the U.S., what state do you work in?</b>	Texto	Estado de EE.UU. en el que trabaja la persona encuestada (si aplica).
<b>What city do you work in?</b>	Texto	Ciudad donde trabaja la persona encuestada.
<b>How many years of professional work experience do you have overall?</b>	Texto	Rango de años de experiencia laboral profesional.
<b>How many years of professional work experience do you have in your field?</b>	Texto	Rango de años de experiencia laboral en el campo específico de la persona encuestada.
<b>What is your highest level of education completed?</b>	Texto	Nivel más alto de educación completado por la persona encuestada.
<b>What is your gender?</b>	Texto	Género con el que se identifica la persona encuestada.
<b>What is your race? (Choose all that apply.)</b>	Texto	Identificación racial de la persona encuestada, pudiendo elegir múltiples opciones.

#### Variables luego de modeladas.

Variable	Tipo	Descripción
<b>Age</b>	Texto	Rango de edad de la persona encuestada.
<b>Industry</b>	Texto	Industria en la que trabaja la persona encuestada.
<b>Job title</b>	Texto	Título del puesto de trabajo de la persona encuestada.
<b>If your job title needs additional context, please clarify here:</b>	Texto	Espacio para aclaraciones adicionales sobre el título del puesto.
<b>Annual_Salary</b>	Número	Salario anual de la persona encuestada, convertido a un equivalente a tiempo completo.

<b>Additional_Compensation</b>	Número	Compensación monetaria adicional recibida en promedio al año (por ejemplo, bonos u horas extras). No incluye beneficios.
<b>Currency</b>	Texto	Moneda en la que está expresado el salario y la compensación adicional.
If "Other," please indicate the currency here:	Texto	Si la moneda seleccionada no está en la lista, se especifica aquí.
If your income needs additional context, please provide it here:	Texto	Campo para proporcionar contexto adicional sobre los ingresos.
<b>Country</b>	Texto	País donde trabaja la persona encuestada.
<b>State_US</b>	Texto	Estado de EE.UU. en el que trabaja la persona encuestada (si aplica).
<b>City</b>	Texto	Ciudad donde trabaja la persona encuestada.
<b>Years_Professional_Experience</b>	Texto	Rango de años de experiencia laboral profesional.
<b>Professional_Expr_inField</b>	Texto	Rango de años de experiencia laboral en el campo específico de la persona encuestada.
<b>Highest_Education</b>	Texto	Nivel más alto de educación completado por la persona encuestada.
<b>Gender</b>	Texto	Género con el que se identifica la persona encuestada.
<b>Race</b>	Texto	Identificación racial de la persona encuestada, pudiendo elegir múltiples opciones.
<b>Annual_Salary_COP</b>	Número	Salario anual en pesos colombianos de la persona encuestada, convertido a un equivalente a tiempo completo.
<b>Compensation_COP</b>	Número	Compensación monetaria en pesos colombianos adicional recibida promedio al año (por ejemplo, bonos u horas extras). No incluye beneficios.
<b>Total_Annual_Salary_COP</b>	Número	Total salario en pesos colombianos incluyendo compensaciones. Suma de Annual_Salary_COP y Compensation_COP

## 1. Carga y Normalización de Datos

### 1.1 Cargar el archivo CSV

El archivo de datos se encuentra en formato CSV y debe ser cargado en un DataFrame de pandas.

```
import pandas as pd

# Cargar el archivo CSV
file_path = "Ask A Manager Salary Survey 2021 (Responses) - Form Responses 1.csv"
df = pd.read_csv(file_path)
```

## 2. Estandarización de Países

### 2.1 Normalización de nombres de países

Los nombres de los países presentan errores de escritura y deben ser estandarizados. Para esto, se usa un diccionario con posibles variaciones mal escritas.

#### Procedimiento

1. Convertir los nombres a minúsculas.
2. Eliminar espacios en blanco al inicio y final.
3. Comparar con el diccionario y asignar el nombre correcto.

```
# Mapeo de países
country_map = {
    "usa": ["united states", "united sates", "america", "usa", "us", "u.s."],
    "canadá": ["can", "canad", "canada"],
    "mexico": ["mexico", "mexico"],
    "not defined": ["n/a", "unknown", "global", "international"]
}

# Función de normalización
def normalize_country(value):
    if pd.isna(value):
        return None
    value = value.strip().lower()
    for country, aliases in country_map.items():
        if value in aliases:
            return country
    return value

# Aplicar normalización
df["What country do you work in?"] = df["What country do you work in?"].apply(normalize_country)

# Guardar resultado
df.to_csv("normalized_survey.csv", index=False)
print("Limpieza de países completada.")
```

#### ¿Cómo agregar otro país?

Si necesitas estandarizar otro país, agrégalo en el country\_map de la siguiente manera:

```
"france": ["france", "france "]
```

## 3. Estandarización de Ciudades

La columna de ciudades también contiene errores tipográficos. Se usa un proceso similar al de los países.

```
# Mapeo de ciudades
city_map = {
    "new york": ["ny", "new york city", "nyc"],
    "los angeles": ["la", "los angeles", "losangeles"],
    "london": ["london", "londres"]
}

# Función de normalización de ciudades
def normalize_city(value):
    if pd.isna(value):
        return None
    value = value.strip().lower()
    for city, aliases in city_map.items():
        if value in aliases:
            return city
    return value

# Aplicar normalización
df["What city do you work in?"] = df["What city do you work in?"].apply(normalize_city)

# Guardar resultado
df.to_csv("normalized_survey.csv", index=False)
print("Limpieza de ciudades completada.")
```

#### ¿Cómo agregar otra ciudad?

Si una nueva ciudad debe ser estandarizada, agrégala en city\_map con sus variaciones mal escritas.

## 4. Renombrar Columnas

Algunas columnas tienen nombres largos o poco claros. Se renombrarán para facilitar el análisis.

```
# Diccionario de mapas
column_map = {
    "What industry do you work in?": "Industry",
    "What is your annual salary?": "Annual_Salary",
    "Please indicate the currency": "Currency",
    "What country do you work in?": "Country",
    "What city do you work in?": "City",
    "What is your highest level of education completed?": "Education",
    "What is your gender?": "Gender"
}

# Renombrar columnas
df.rename(columns=column_map, inplace=True)

# Verificar cambios
df.head()
```

## 5. Conversión de Monedas a COP

Los salarios se reportan en diferentes monedas y deben convertirse a pesos colombianos (COP).

### Procedimiento

1. Definir un diccionario de códigos de moneda.
2. Obtener tasas de cambio desde una API o archivo estático.
3. Calcular los salarios en COP.

```
import requests

# Diccionario de monedas
currency_map = {
    "USD": "USD", "EUR": "EUR", "GBP": "GBP", "JPY": "JPY", "AUD": "AUD", "BRL": "BRL",
    "COP": "COP", "CHF": "CHF", "CAD": "CAD"
}

# Obtener tasas de cambio desde una API
url = "https://api.exchangerate-api.com/v4/latest/USD"
response = requests.get(url)
exchange_rates = response.json()["rates"]

# Función de conversión a COP
def convert_to_cop(row):
    currency = row["Currency"]
    salary = row["Annual_Salary"]
    if pd.isna(salary) or currency not in currency_map:
        return None
    rate = exchange_rates.get(currency_map[currency], 1)
    return salary * rate

# Aplicar conversión
df["Annual_Salary_COP"] = df.apply(convert_to_cop, axis=1)

# Guardar resultado
df.to_csv("converted_survey.csv", index=False)
print("Conversión de salarios a COP completada.")
```

### ¿Cómo agregar otra moneda?

Si aparece una moneda nueva, agrégala al diccionario `currency_map`, por ejemplo:

```
"INR": "INR"
```

## 6. Calcular la columna Total\_Annual\_Salary\_COP

En este paso, vamos a crear una nueva columna llamada `Total_Annual_Salary_COP`, que representa la suma de las siguientes columnas:

- **Annual\_Salary\_COP**: Salario anual convertido a pesos colombianos.
- **Compensation\_COP**: Otros tipos de compensación (bonos, comisiones, etc.) convertidos a pesos colombianos.

Dado que algunas filas pueden tener valores nulos (NaN) en estas columnas, usaremos `.fillna(0)` para reemplazar los valores faltantes por 0 antes de realizar la suma.

```
df["Total_Annual_Salary_COP"] = df["Annual_Salary_COP"].fillna(0) + df["Compensation_COP"].fillna(0)
```

## 7. Guardado Final

Después de completar los pasos anteriores, guardar el archivo final.

```
df.to_csv("final_survey_data.csv", index=False)  
print("Archivo final guardado.")
```

## Resumen de Pasos

1. Cargar los datos desde el archivo CSV.
2. Estandarizar nombres de países con `normalize_country()`.
3. Estandarizar nombres de ciudades con `normalize_city()`.
4. Renombrar columnas para mejorar la legibilidad.
5. Convertir salarios a COP usando tasas de cambio actuales.
6. Calcular salario total en pesos Colombianos.
7. Guardar archivo final.

El código completo se puede descargar directamente del siguiente repositorio:

<https://github.com/ocamilot/visualizacin-y-storytelling.git>