

Instructivo para actualización de Dashboard Salarios de Profesionales

Elaborado por: Rebeca Gamboa Venegas

El siguiente documento está conformado por tres secciones:

1. Variables en base de datos original
2. Variables modeladas para la construcción del Dashboard
3. Pasos para la limpieza y construcción de variables

Estas secciones buscan ser una guía para entender la base de datos con la que se trabaja y replicar, en caso de que sea necesario, los procedimientos implementados para la limpieza inicial y la creación de nuevas variables.

Variables en base de datos original

Se describe el nombre original de la variable, el tipo de variable y una breve descripción de su significado.

Tabla 1. Descripción de variables originales

Nombre	Tipo	Descripción
Timestamp	Fecha	Fecha y hora en la que fue llenada la encuesta
How old are you?	Texto	Rango de edad de la persona
Industry	Texto	Industria en la que se desempeña la persona
Job title	Texto	Cargo que desempeña la persona
Additional context on job title	Texto	Contexto adicional de la profesión desempeñada. Este campo es opcional, únicamente se llena en caso de ser necesario
Annual salary	Número	Salario anual de la persona
Other monetary comp	Número	Otras compensaciones monetarias recibidas a lo largo del año por la persona
Currency	Texto	Moneda en la que se reporta el salario, puede elegir entre 10 monedas diferentes, sino elegir "Other" y escribir la moneda que aparecerá en "Currency-other"
Currency - other	Texto	Otra moneda diferente a las preseleccionadas
Additional context on income	Texto	Contexto adicional del salario reportado. Este campo es opcional, únicamente se llena en caso de ser necesario
Country	Texto	País de residencia de la persona

State	Texto	Estado de residencia - Aplica únicamente para los residentes en Estados Unidos
City	Texto	Ciudad de residencia de la persona
Overall years of professional experience	Texto	Rango de cantidad general de años de experiencia profesional de la persona
Years of experience in field	Texto	Rango con la cantidad de años de experiencia en la profesión
Highest level of education completed	Texto	Máximo nivel de educación completa alcanzado por la persona
Gender	Texto	Género de la persona
Race	Texto	Raza de la persona, existen 7 posibles respuestas de selección múltiple

Variables modeladas para la construcción del Dashboard

Se describe el nombre final de la variable, el tipo de variable y una breve descripción de su significado. Para este ejercicio se renombran las variables al español y además, se crean 3 nuevos campos importantes para lograr los objetivos del Dashboard, estos campos son “Compensaciones”, “Salario_anual”, y “Salario_total”

Tabla 2. Descripción de variables finales

Nombre	Tipo	Descripción
Fecha	Fecha	Fecha y hora en la que fue llenada la encuesta
Edad	Texto	Rango de edad de la persona
Industria	Texto	Industria en la que se desempeña la persona
Cargo	Texto	Cargo que desempeña la persona
Contexto profesión	Texto	Contexto adicional de la profesión desempeñada
Salario anual	Número	Salario anual
Otras compensaciones	Número	Otras compensaciones monetarias recibidas a lo largo del año
Moneda	Texto	Moneda en la que se reporta el salario
Moneda-otra	Texto	Otra moneda diferente a las preseleccionadas
Contexto ingreso	Texto	Contexto adicional del salario reportado
País	Geo	País de residencia de la persona, en formato georeferenciado
Estado	Geo	Estado de residencia, en formato georeferenciado - Aplica únicamente para los residentes en Estados Unidos
Ciudad	Geo	Ciudad de residencia de la persona, en formato georeferenciado

Años de experiencia profesional	Texto	Rango de cantidad de años de experiencia de la persona
Años de experiencia en el campo	Texto	Rango de cantidad de años de experiencia en el campo
Educación	Texto	Mayor nivel de educación completa alcanzado por la persona
Género	Texto	Género de la persona
Raza	Texto	Raza
Compensaciones	Número	Compensaciones en pesos colombianos, recibidas a lo largo del año
Salario_anual	Número	Salario anual en pesos colombianos
Salario_total	Número	Suma de las compensaciones + salario anual en pesos colombianos

Pasos para la limpieza y construcción de variables

Se describen los pasos necesarios para la limpieza y creación de las nuevas variables.

1. Ingrese al siguiente link:
<https://docs.google.com/spreadsheets/d/1IPS5dBSGtwYVbjsfbaMCYIWnOuRmJcbequo hNxCyGVw/edit?resourcekey#gid=1625408792>
2. Copie la base de datos en una hoja de cálculo de Google
3. Enlace la hoja al Google Data Studio, dando clic en “Añadir base de datos” y seleccione la opción “Hojas de cálculo de Google”, finalmente seleccione la base de datos copiada en el paso anterior.

Limpieza

Limpieza de columna “Country”

Opción 1:

1. Dentro de la hoja de cálculo diríjase a la columna “State”, como esta columna es única para los habitantes de Estados Unidos, coloque “United States” en el campo “Country”.
2. Como se está realizando la modificación desde la hoja de cálculo, es importante recalcar que esta no distingue entre mayúsculas y minúsculas

a la hora de filtrar, pero Google Data Studio si lo hace, por lo que es necesario unificar las casillas de texto. Para esto se puede hacer una nueva columna llamada "Country-clean" a la cual le aplicaremos dos funciones, una llamada "TRIM" que elimina los espacios al frente y al final del texto y la función "LOWER" la cual vuelve todos los caracteres a minúsculas. Estas funciones se colocan en la columna "Country-clean" y se le aplican a "Country" de la siguiente manera:

`=TRIM(LOWER(Casilla_Country))`

Finalmente, se corre la función en todas las casillas de la columna.

3. Utilice los filtros de la columna "Country" para observar relaciones y unificar los campos, por ejemplo, para los países de "United Kingdom" tome en consideración los campos de "Wales", "Scotland", "England", "North Ireland", "Britain", entre otros.

Opción 2:

1. Dentro de Google Data Studio seleccione la opción "Recurso", seguidamente "Gestionar las fuentes de datos añadidas", luego dé clic al botón "editar" relacionada con la fuente de datos.
2. Una vez ahí, de clic a "Añadir un campo" al que llamará "País-limpio" y escriba un código que busque en la columna "Country" semejanzas. Para esto usaremos igualmente la función "TRIM" Y "LOWER" las cuales tienen la misma función que en la hoja de cálculo, pero, además, es necesario usar la función "REGEXP_CONTAINS" que nos ayuda a identificar si en las filas de "Country" existen los valores descritos, en caso de ser así retorna el valor dicho después del "THEN".
3. Ingrese el código:

CASE

```
WHEN TRIM(LOWER(Country)) IN ( "united sates of america", "uniyed states","unites status","unites statea","us of a","uniter statez","untied states","us govt employee overseas, country withheld","usaa","usat","usab","united states","uss","united statesp","unyted statesp", "unted states","united stattes","united statees","united statws","uniyes states","united states of americas","us","unites statea","united sates","u.s.", "united states", "us", "america", "the united states", "the us", "u. s", "u. s.", "u.a.", "u.s", "u.s.", "u.s>", "u.sa", "ua", "u.s.a.", "usa", "u.s.a", "uxz", "united y", "united
```

```
statss", "uniteed states", "usa (company is based in a us territory, i work remote)", "usa--  
virgin islands", "usa, but for foreign gov't", "uniited states", "unite states", "united states",  
"united states of america", "unites states", "usa tomorrow", "unitef stated", "united states  
of american", "united stares", "united state", "united state of america", "united stated",  
"united stateds", "united states is america", "unitedstates" ) THEN "United States"  
WHEN TRIM(LOWER(Country)) IN ( "britain", "united  
kingdom", "uk", "scotland", "england", "scotland", "uk", "great britain", "england", "northern  
ireland", "england/uk", "england, uk.", "united kingdom (england)", "united  
kingdom.", "u.k.", "united kindom", "england, uk", "great britain", "u.k.", "uk (northern  
ireland)", "uk for u.s. company", "united kingdom", "wales (united kingdom)", "england,  
gb", "u.k. (northern england)", "u.k.", "england, united kingdom", "englang", "wales", "uk  
(england)", "uk, remote", "scotland, uk", "unites kingdom", "northern ireland", "wales,  
uk", "england, united kingdom", "wales (uk)", "northern ireland, united kingdom", "uk, but for  
globally fully remote company") THEN "United Kingdom"  
WHEN TRIM(LOWER(Country)) IN ("new zealand", "new zealand aotearoa", "nz", "new  
zealand", "from new zealand but on projects across apac", "aotearoa new zealand") THEN  
"New Zealand"  
WHEN TRIM(LOWER(Country)) IN ("netherlands", "the netherlands", "nederland") THEN  
"Netherlands"  
WHEN TRIM(LOWER(Country)) IN  
("canada", "can", "csnada", "canda", "canad", "canadw", "canadá") THEN "Canada"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'canada') THEN "Canada"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'd[a-z]nmark') THEN "Denmark"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'switzerland') THEN  
"Switzerland"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'china') THEN "China"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'afghanistan') THEN  
"Afghanistan"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'france') THEN "France"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'm[a-z]xico') THEN "Mexico"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'bra[a-z]il') THEN "Brazil"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'australi*') THEN "Australia"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'austria*') THEN "Austria"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'argentina*') THEN "Argentina"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'japan*') THEN "Japan"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'italy*') THEN "Italy"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'luxem*') THEN "Luxembourg"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'hong kon*') THEN "Hong Kong"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'phili*') THEN "Philipinnes"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'germany') THEN "Germany"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'spain') THEN "Spain"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'czech republic') THEN "Czech  
Republic"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'south africa') THEN "South  
Africa"  
WHEN REGEXP_CONTAINS(TRIM(LOWER(Country)), 'united sta*') THEN "United  
States"  
ELSE TRIM(Country)  
END
```

Limpeza de columna "City"

1. Para este caso, desde la hoja de cálculo de Google, coloque los valores de la columna "City" de manera ascendente.
2. Cree una hoja adicional, donde copiará a partir de la celda A1, los datos de "City"
3. Copie la celda A1 en la B1
4. Aplique la siguiente formula en la celda B2 y corrala en el resto de las celdas hacia abajo:

$$=IF(LEFT(A2,6) = LEFT(A1,6), B1 , A2)$$

5. Esta función permite comparar los primeros 6 caracteres de todas las casillas, en caso de existir una coincidencia, se elige la casilla de arriba, de lo contrario mantiene el nombre original del campo.
6. Una vez realizado el procedimiento, copie como valores la columna B de la nueva hoja, en la columna "City"

Editar columnas

Para editar las columnas en data Studio, siga las siguientes instrucciones:

1. Ingrese a "Recurso", dé clic a "Gestionar las fuentes de datos añadidas", luego dé clic al botón "editar" relacionada con la fuente de datos.
2. Cambie los nombres de las columnas a español, a continuación, se presenta una tabla con los nombres:

Nombre en inglés	Nombre en español
Timestamp	Fecha
How old are you?	Edad
Industry	Industria
Job title	Cargo
Additional context on job title	Contexto profesión
Annual salary	Salario anual
Other monetary comp	Otras compensaciones
Currency	Moneda
Currency - other	Moneda-otra

Additional context on income	Contexto ingreso
Country	País
State	Estado
City	Ciudad
Overall years of professional experience	Años de experiencia profesional
Years of experience in field	Años de experiencia en el campo
Highest level of education completed	Educación
Gender	Género
Race	Raza

Cambio de tipo de variables

1. Ingrese a “Recurso”, dé clic a “Gestionar las fuentes de datos añadidas”, luego dé clic al botón “editar” relacionada con la fuente de datos.
2. De clic a “ABC” (Tipo de variable) del campo “País” y cámbielo a “Información Geográfica” –“País”
3. Realice el mismo procedimiento para “Ciudad” y “Estado”, cambiando únicamente la opción “País” por “Ciudad” y “Subdivisión del país (primer nivel”, respectivamente.

Generar nuevas variables

Para generar las nuevas columnas en data Studio, siga las siguientes instrucciones:

1. Ingrese a “Recurso”, dé clic a “Gestionar las fuentes de datos añadidas”, luego dé clic al botón “editar” relacionada con la fuente de datos.
2. Una vez ahí, de clic a “Añadir un campo” al que llamará “Compensaciones”
3. Paralelamente ingrese a la página: [Xe Live Exchange Rates](#)
4. Ingrese el siguiente código para crear “Compensaciones”

```
CASE  
WHEN Moneda = "USD" THEN Otras compensaciones * 3939.48
```

```
WHEN Moneda = "CAD" THEN Otras compensaciones * 3092.60
WHEN Moneda = "GBP" THEN Otras compensaciones * 5339.22
WHEN Moneda = "EUR" THEN Otras compensaciones * 4466.91
WHEN Moneda = "AUD/NZD" THEN Otras compensaciones * 2809.65
WHEN Moneda = "CHF" THEN Otras compensaciones * 4254.86
WHEN Moneda = "SEK" THEN Otras compensaciones * 421.90
WHEN Moneda = "JPY" THEN Otras compensaciones * 34.13
WHEN Moneda = "ZAR" THEN Otras compensaciones * 258.94
WHEN Moneda = "HKD" THEN Otras compensaciones * 505.08
WHEN Moneda = "Other" AND Moneda-otra = "USD" THEN Otras compensaciones *
3939.48
ELSE Otras compensaciones
END
```

6. De clic en guardar y regrese a la pantalla donde se muestran todos los campos

IMPORTANTE: Modifique los datos de Tipo de Cambio encontrados en la página del punto 3.

7. En esta misma página agregue un campo nuevo, para ello de clic a “Añadir un campo” y nómbrelo como “Salario_anual”
8. Ingrese el siguiente código:

```
CASE
WHEN Moneda = "USD" THEN Salario anual * 3939.48
WHEN Moneda = "CAD" THEN Salario anual * 3092.60
WHEN Moneda = "GBP" THEN Salario anual * 5339.22
WHEN Moneda = "EUR" THEN Salario anual * 4466.91
WHEN Moneda = "AUD/NZD" THEN Salario anual * 2809.65
WHEN Moneda = "CHF" THEN Salario anual * 4254.86
WHEN Moneda = "SEK" THEN Salario anual * 421.90
WHEN Moneda = "JPY" THEN Salario anual * 34.13
WHEN Moneda = "ZAR" THEN Salario anual * 258.94
WHEN Moneda = "HKD" THEN Salario anual * 505.08
WHEN Moneda = "Other" AND Moneda-otra = "USD" THEN Salario anual * 3939.48
ELSE Salario anual
END
```

9. De clic en guardar y regrese a la pantalla donde se muestran todos los campos

IMPORTANTE: Al igual en la columna “Compensaciones”, modifique los datos de Tipo de Cambio encontrados en la página del punto 3.

10. En esta misma página agregue un campo nuevo, para ello de clic a “Añadir un campo” y nómbrelo como “Salario_total”

11. Ingrese el siguiente código:

=Salario_anual+Compensaciones

12. De clic en guardar y regrese a la pantalla donde se muestran todos los campos.

Una vez realizada la limpieza y la creación de nuevas variables puede actualizar el Dashboard.