

ANALISIS DE SALARIO ANUAL Y DIVISAS

Limpieza de datos y Dashboard

Autor: Daren Joan Rodríguez Cáceres

1. Introducción

La información recolectada corresponde al salario anual que reciben personas de todo el mundo. Este salario se encuentra en la moneda que recibe cada persona sus ingresos, así como información referente a características generales como: edad, género, nivel educativo, país, ciudad entre otras variables.

Mucha de esta información contiene errores y se requiere realizar la limpieza de la misma, así como ordenar la información que se requiera para darle sentido a los datos y generar nuevas variables en pesos colombianos, a continuación, se describe el procedimiento a seguir y las indicaciones.

2. Origen de la información

Archivo original:

<https://docs.google.com/spreadsheets/d/1IPS5dBSGtwYVbjsfbaMCYIWnOuRmJcbequohNxCyGVw/edit?resourcekey#gid=1625408792>

Archivo modificado: Ask A Manager Salary Survey 2021_modificado.xlsx

3. Variables y tipos de datos

Variables Originales	Descipción	Tipo de variable
Timestamp	hora y fecha de ingreso de la información	Fecha y hora
How old are you?	Rango de edad: menor de 18, 18 a 24, 25 a 34, 35 a 44, 45 a 54, 55 a 64, mayor de 65	Texto(Rango Numérico entero)
Industry	Industria o sector económica	Cadena de texto
Job title	Cargo laboral o nombre del empleo	Cadena de texto
Additional context on job title	Descripción adicional del empleo	Cadena de texto
Annual salary	Salario anual	Numero entero
Other monetary comp	Otras compensaciones monetarias	Numero entero
Currency	Moneda o divisa del salario anual, compensaciones y otros ingresos	Cadena de texto
Currency - other	Monedas o divisas no comunes	Cadena de texto
Additional context on income	Contexto adicional sobre los ingresos	Cadena de texto
Country	País	Cadena de texto
State	Estado o provincia	Cadena de texto
City	Ciudad	Cadena de texto

Overall years of professional experience	Rango en años de experiencia profesional general: menos de 1 año, 2 a 4, 5 a 7, 8 a 10, 11 a 20, 21 a 30, 31 a 40, 41 o mas.	Texto(Rango Numérico entero)	
Years of experience in field	Rango en años de experiencia en el campo específico: menos de 1 año, 2 a 4, 5 a 7, 8 a 10, 11 a 20, 21 a 30, 31 a 40, 41 o mas.	Texto(Rango Numérico entero)	
Highest level of education completed	Nivel educativo:College degree, Master's degree, High School,Some college,PhD	Cadena de texto	
Gender	Genero :Woman, Man, Non-binary, other or prefer not to answer	Cadena de texto	
Race	Raza: Black or African American, Hispanic, Latino, or Spanish origin, White,Native American or Alaska Native,Asian or Asian American, Middle Eastern or Northern African,Another option not listed here or prefer not to answer	Cadena de texto	
Variables Adicionales		Descipción	Tipo de variable
TRM_COP		Tasa de cambio en pesos	Numérica, decimal, moneda
Annual_salary_COP		Salario anual en pesos colombianos	Numérica, decimal, moneda
Compensatory_COP		Compensaciones monetarias en pesos colombianos	Numérica, decimal, moneda
Annual_salaryandcompensatory COP		Suma del salario anual y compensaciones en pesos colombianos	Numérica. decimal. moneda

4. Limpieza de datos

La base de datos se actualiza constantemente y para futuros análisis se debe tomar la mas reciente y seguir los siguientes pasos:

- La variable Annual Salary solo puede tener como menor valor el numero cero, el resto de números que sea menos de 1000 debe multiplicarse por mil, esto porque la mayoría de gente asume que es en miles el salario en algunos casos. Asi un salario de 18 realmente corresponde a 18000 ya que es un valor anual.
- La variable Currency corresponde a las divisas o moneda en la que se debe entender la cifra del salario anual y se presenta en siglas, en algunos casos esta sigla no aparece y se debe usar el dato de la siguiente variable currency-other o de acuerdo al contexto general como país u otra información dentro del formulario se puede conocer la moneda en cada caso. Por ejemplo el dólar Australiano tiene la sigla AUD y es muy probable que las personas de

ese país lo usen entonces al encontrar currency vacía puede corresponder a esta moneda de acuerdo al contexto.

- La variable Country presenta la lista de países de las personas del formulario, sin embargo, mucha información está mal escrita, incompleta o confusa. Por ejemplo, para referirse a Estados Unidos se puede encontrar EU, United States, USA entre otras. La idea es modificar y unificar una sola palabra o sigla por país que lo haga fácil de reconocer, por ejemplo, se cambian todas a USA. Si no se tienen información se revisa la variable City para dar un mejor contexto y poder definir el país. Para esto se pueden utilizar herramientas de filtrado en Excel y luego se reemplazan o herramientas de código, sin embargo, países como Canadá, USA o UK presentan mucha variación y algunas no son fácilmente identificables sin ver contexto. Por ejemplo, algunas personas pueden colocar IS o nombres mal escritos que con código no son fácilmente identificables y un filtro con la ciudad facilitan cual puede ser el país de origen.
- La variable City presenta la lista de ciudades y se maneja de manera similar a Country, con la diferencia que se tiende a priorizar la información que no es clara o no es fácil de interpretar como N/A. Por ejemplo, algunas personas no ponen su ciudad y ponen trabajo remoto, WFH o simplemente no quieren poner su ciudad por seguridad, todos estos casos se dejan como N/A. En el caso de caso de ver información de una ciudad repetida, pero con caracteres adicionales o información extra la idea es unificar. Por ejemplo al encontrar Los Angeles, LA, Los angeles ,CA se unifica en Los Angeles. Se usan las herramientas de filtrado y reemplazando por el valor unificado. No siempre se puede hacer esto por tiempo y se recomienda empezar por los países que tienen más información de ciudades como Canadá, USA y Uk.
- Para los dos casos anteriores es útil generar una columna temporal que puede servir para hacer algunos cambios en sus nueva información integrada, por ejemplo para el caso de USA pueden haber varias siglas, EU, US, EUA, United States y en Excel se pueden usar formulas para reemplazar estos valores SI(valor de celda= "EU"; "USA"; valor de celda), la formula anterior tb es útil cuando varias siglas o nombres de país inician por la misma letra, por ejemplo añadiendo la formula izquierda SI(IZQUIERDA(Valor de celda)="E";"USA";Valor de celda)
- En todos los casos al filtrar país por país se facilita el llenado de datos erróneos o vacíos.

5. Generación de variables adicionales

Se requiere ver las nuevas variables en pesos colombianos COP y para esto se requiere el cambio de la divisa o TRM en el día de realizarse el informe, generando una tabla como la siguiente:

País/ Continente	Moneda	Sigla	TRM_COP
Estados Unidos	Dollar Americano	USD	3936
Canadá	Dollar Canadiense	CAD	3090
Gran Bretaña	Libra Esterlina	GBP	5337
Europa	Euro	EUR	4565
Australia	Dollar Australiano	AUD	2807
Nueva Zelandia	Dollar zelandes	NZD	2615
Suiza	Franco suizo	CHF	4250
Suecia	Corona Sueca	SEK	421.64

Japon	Yen	JPY	34026
Sudafrica	Rand	ZAR	258.73
Hong Kong	Dollar	HKD	504.94
Noruega	Corona Noruega	NOK	444.43
Brasil	Real	BRL	749.82
Korea del Sur	Won	KRW	3.28
Singapore	Dollar	SGD	2923
Israel	Shequel	ILS	126.74
Arabia	Saudi Riyal	SAR	1050
Tailandia	Baht	THB	120.36
Sri Lanka	Rupia	LKR	19.4
India	Rupia	INR	52.26
Dinamarca	Corona	DKK	600.83
Argentina	Peso Argentino	ARS	36.94
Malasia	Ringgit malayo	MYR	936.27
Trinidad	Dólar de Trinidad	TTD	576.1
Mexico	Peso Mexicano	MXN	191.13
Suecia	Corona Sueca	CZK	181.76
Bangladesh	Taka de Bangladesh	BDT	45.55
Filipinas	Peso Filipino	PHP	76.43
Polonia	Zloty polaco	PLN	975.37
Turquia	Lira Turca	TRY	290.57
China	Yuan Chino	CNY	617.48
Taiwan	Nuevo dólar Taiwandes	TWD	140.71
Indonesia	Rupia Indonesia	IDR	0.27
Croacia	kuna Croata	HRK	591.57

La información de la tabla anterior fue tomada el día 13 de feb. de 22 de la página web <https://www.xe.com/currencyconverter/convert/?Amount=1&From=CHF&To=COP> y de ser requerida para la moneda de otros países solo se debe buscar el país con la sigla requerida de su moneda al día del cambio.

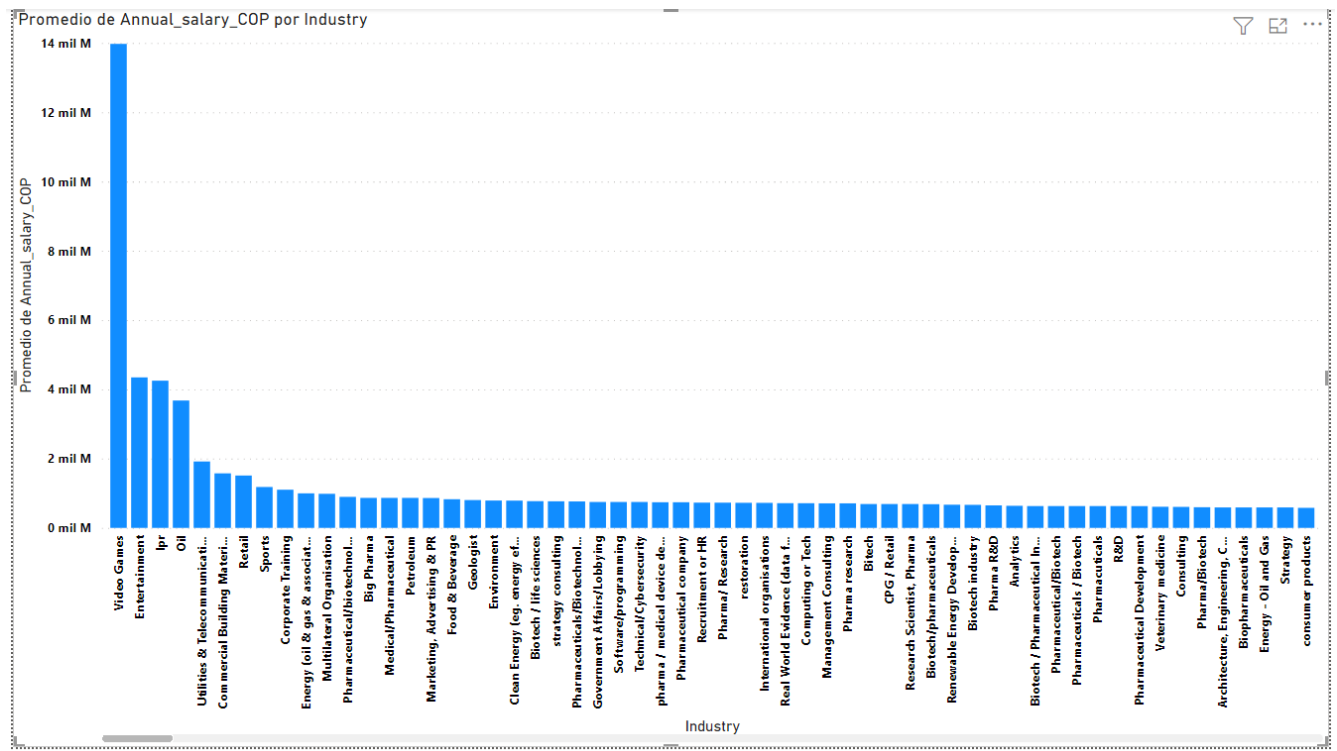
Para generar la variable TRM_COP se genera un buscar v en el archivo de Excel, usando la tabla anterior actualizada y usando las siglas de la variable Currency como la sigla de búsqueda.

Para generar la variable Annual_Salary_COP se requiere multiplicar la variable Annual_Salary por la variable TRM_COP. De igual manera con la variable Compensatory_COP se genera al multiplicar la variable Other_Monetary_Comp por la variable TRM_COP y por ultimo se suman las dos variables nuevas generadas para formar una variable llamada Annual_salaryandcompensatory_COP.

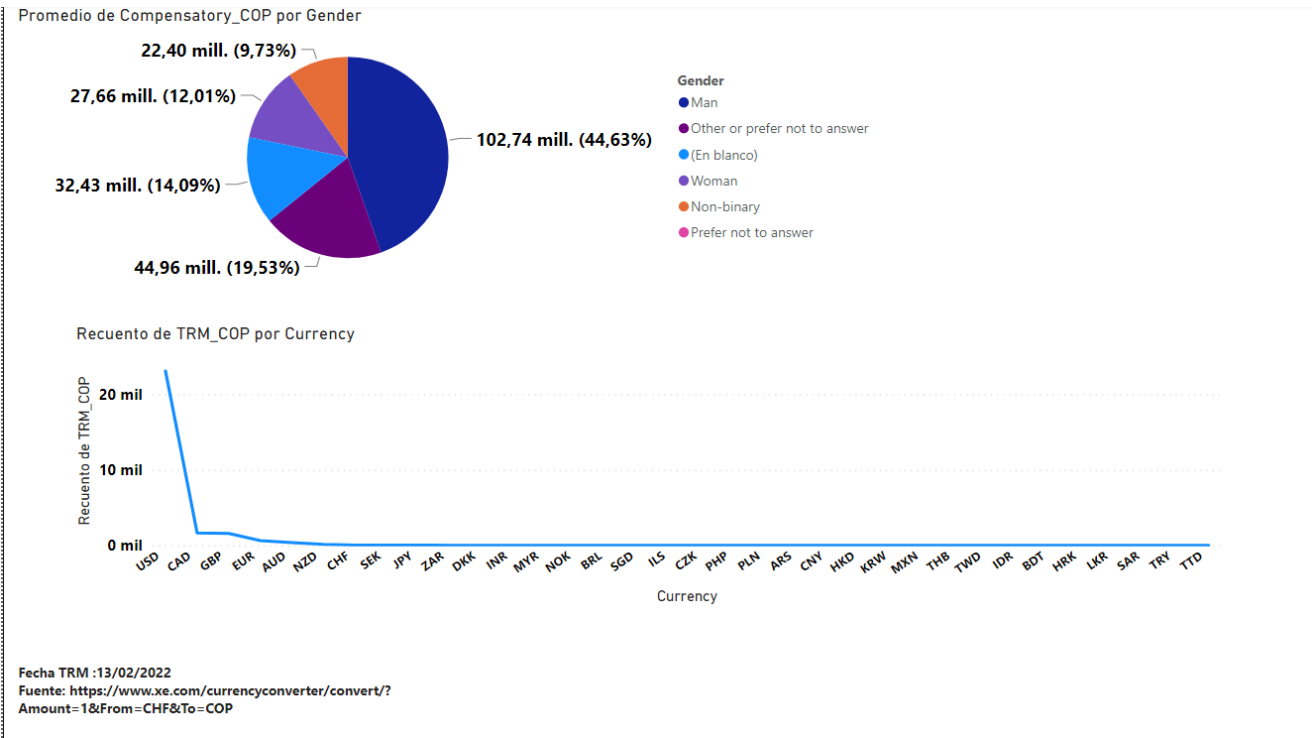
6. Generación del Dashboard en Power BI



Recuento de los datos y visualización de los países registrados.



Promedio anual en COP para cada sector económico



Promedio de ingresos compensatorios en COP por genero y recuento por diferentes divisas.