

Análisis sobre Salarios en Data Science 2022/23



- Proyecto Final | Segunda Entrega
 - Fabrizio Mateo Heck
 - Tutor: Leonel Bordon
 - Profesor: Dario Ceballos
 - Julio 2023

CODER HOUSE

Contenido

Descripción de la temática:.....	3
Destinatario	3
Hipótesis:	3
Diagrama ER.....	3
Listado de Tablas y Campos	4
Tabla Employee	4
Tabla Company.....	5
Tabla Country	5
Tabla Industry.....	5
Tabla Size	5
Tabla JobTitle	5
Transformaciones.....	6
Medidas Calculadas	6
Marcadores y Tooltips	7
Conclusiones:	7
Futuros Lineamientos:	7

Descripción de la temática:

En el presente proyecto se trabajará con datos extraídos de la página web de Kaggle (fuente citada al final), sobre personal en Data Science de 3755 ejemplares entre 2022 y 2023, con sus respectivos salarios y tamaño de empresas donde llevan a cabo sus tareas diarias.

Sumado a esto, se recopilaron datos de salarios mínimos en cada país del empleado, industria en la que trabajan y porcentaje del tiempo de trabajo remoto con respecto al de presencial.

Destinatario

La idea es poder mostrar a los ya introducidos en el mundo del Data Science un parámetro para saber en dónde se encuentran y entender los factores que pueden hacer variar su salario. Del mismo modo, pretende servir de guía para todo aquel que este en el proceso de elección o formación en el rubro.

Hipótesis:

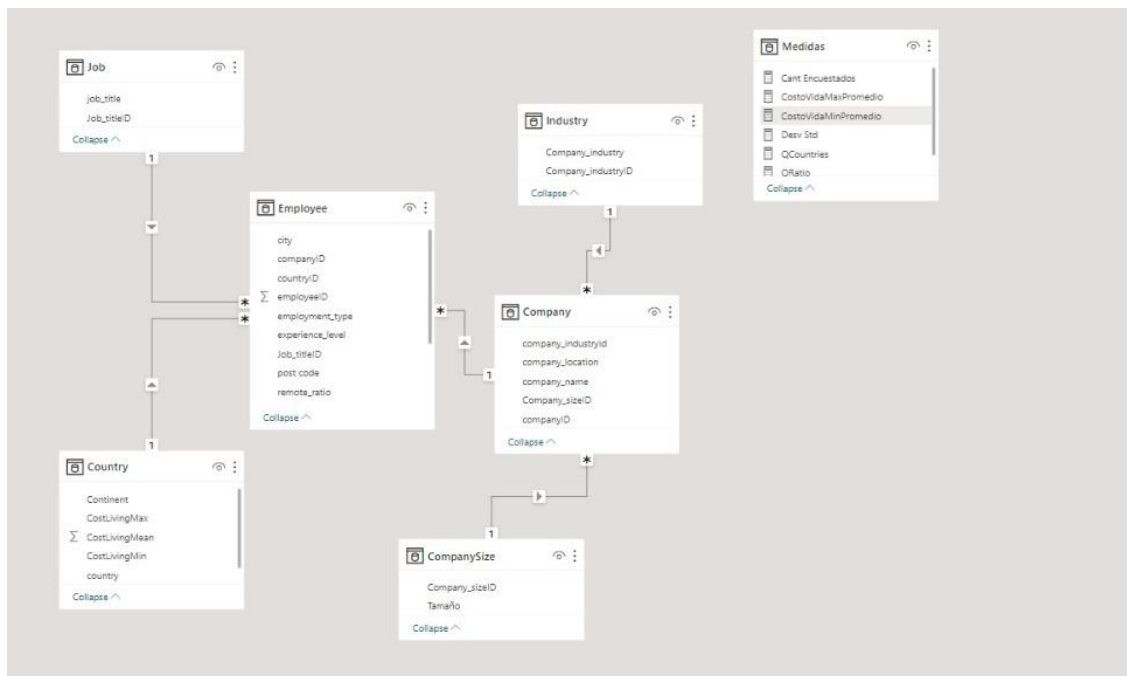
Se buscará comprobar si realmente las empresas de mayor tamaño abonan salarios superiores, como también el qué tan influyente es el nivel de experiencia tanto para lograr puestos en grandes empresas como para recibir mejores pagos anuales. Sumado a esto, interesa saber de dónde son la mayoría de las empresas, a qué países pertenecen las de mayor calibre y qué puesto tiene mayor salario en promedio.

Por último, se comparará con el costo de vida promedio de cada país de residencia del empleado para esbozar la capacidad de ahorro, suponiendo que los salarios en la industria son competitivos a nivel general y permiten tener, relativamente, buen estándar de vida en cualquier parte del mundo.

Fuente de base de datos (incompleta): <https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023?resource=download>

Diagrama ER

A continuación, se presenta el Diagrama Entidad-Relación creado por el mismo PowerBI, con la intención de organizar y comprender las relaciones entre las distintas tablas que conforman la base de datos a utilizar:



Listado de Tablas y Campos

Tabla Employee

Como tabla central, es la más extensa y repleta de datos con 13 campos y es la que describe cada característica de cada empleado, desde el año en que se le consultó su salario, hasta el ratio remoto-presencial en el que trabaja:

Employee		
Campo	Tipo de Campo	Tipo de Clave
EmployeeID	Int	PK – Index
WorkYear	Int	
EmployeeType	Varchar, 50	
ExperienceLevel	Varchar, 50	
JobTitleID	Varchar, 50	FK
Salary	SmallMoney	
SalaryCurrency	Int	
SalaryInUSD	Int	
CountryID	Varchar, 50	FK
Street	Varchar, 50	
PostCode	Varchar, 50	
City	Varchar, 50	
RemoteRatio	Int	
CompanyID	Int	FK

Tabla Company

Esta tabla muestra las características de la compañía para la cual trabajan los empleados, con los siguiente 5 campos:

Company		
Campo	Tipo de Campo	Tipo de Clave
CompanyID	Int	PK
Location	Varchar, 50	
Size	Varchar, 50	FK
Name	Varchar, 50	
IndustryID	Int	FK

Tabla Country

Esta tabla contiene un identificador con cada país de residencia y salario mínimo de cada trabajador.

Country		
Campo	Tipo de Campo	Tipo de Clave
CountryID	Varchar, 50	PK
Country	Varchar, 50	
Continent	Varchar, 50	
CostLivingMin	Varchar, 50	
CostLivingMax	Varchar, 50	

Tabla Industry

De aquí en adelante, las tablas tienen solo dos campos que describen el ID en cada uno de los casos, por lo cual se decide exponerlo de forma más sencilla. En este, hace referencia a qué industria pertenece la compañía para la cual se trabaja.

1. IndustryID (PK, Int);
2. Industry (Varchar, 50).

Tabla Size

Define en uno de los tres tamaños posibles de empresa

1. SizeID (PK, Varchar, 50);
2. Size (Varchar, 50).

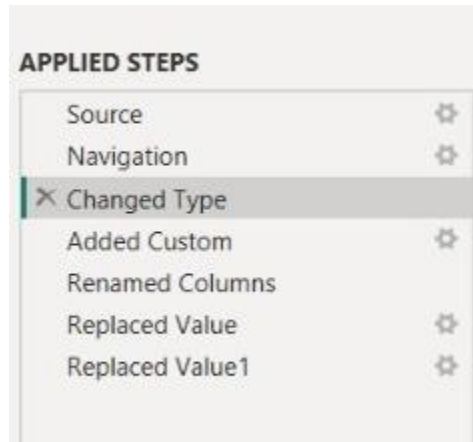
Tabla JobTitle

Cada uno de los 93 puestos de trabajo, relacionados al Data Science, están identificados aquí.

1. JobTitleID (PK, Varchar, 50);
2. JobTitle (Varchar, 50).

Transformaciones

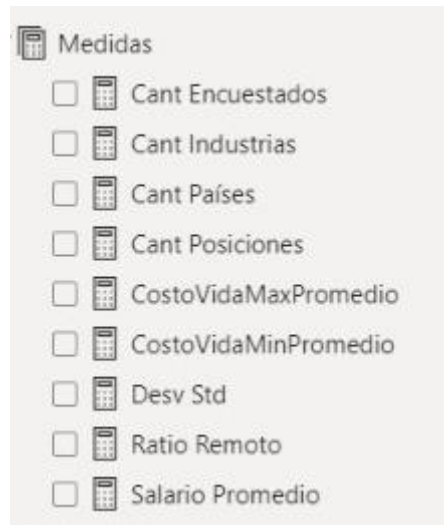
Siendo que era una base de datos bastante limpiada en Excel, solo restó hacer algunos detalles como cambiar el tipo de dato del campo countryID, agregar una nueva columna en la tabla Country, llamada CostLivingMean, que muestra el promedio entre CostLivingMin y CostLivingMax. Sumado a eso, se modificó el nombre del campo que quedó como countryID y se reemplazaron dos valores en country que pasaron de “States” y “New” a “United States” y “New Zeland” respectivamente.



Medidas Calculadas

Con respecto a las medidas calculadas generadas, se realizaron 7:

- Cant Encuestados = `COUNT(Employee[employeeID]);`
- Cant Industrias = `Count(Industry[Company_industryID]);`
- Cant Países = `COUNT(Country[countryID]);`
- Cant Posiciones = `Count(Job[Job_titleID]);`
- CostoVidaMaxPromedio = `AVERAGE(Country[CostLivingMax]);`
- CostoVidaMinPromedio = `AVERAGE(Country[CostLivingMin]);`
- Desv Std = `STDEV.S(Employee[salary_in_usd]);`
- RatioRemoto = `Count(Employee[remote_ratio]);`
- Salario Promedio = `AVERAGE(Employee[salary_in_usd]).`



Marcadores y Tooltips

En las solapas Industria y Top se crearon dos marcadores para cada una con la intención de mejorar la visualización de los gráficos allí expuestos. En ello se pueden ver, por un lado, “Salario Promedio” y “Empleados Industria” y, por el otro, “Top Países” y “Top Puestos” respectivamente.

Por otro lado, se creó un Tooltip aplicado en la solapa Tamaño, el cual muestra la cantidad de trabajadores por tamaño de empresa.

Conclusiones:

Podemos ver que los salarios en términos generales han aumentado vigorosamente en los cuatro años trabajados a medida que la desviación estándar se reduce, mientras que la cantidad de países en lo que residen los trabajadores, se ve en gran aumento.

En segundo lugar, se observa que son las medianas compañías las que ofrecen mejores salarios, y que, en promedio, la mitad de los trabajadores, asisten de forma presencial a sus trabajos. Sumado a esto, la consultoría es tanto la que mejores salarios otorga y la que mayor cantidad de trabajadores emplea.

En cuanto a las diferencias entre continentes, se muestra que los mayores costos de vida tienen, son también los que más pagan. Además, son Israel, Malasia y Puerto Rico los que mayores salarios pagan, siguiéndole Estados Unidos y Canadá, mientras que los puestos mejor pagos son los relacionados a posiciones de jefe como “Lead”, “Director” o “Principal”.

Futuros Lineamientos:

A medida que haya más cantidad de encuestados, tendremos información más precisa y certera. También se podría hacer un análisis más acabado de los puestos de trabajo para aglomerar los que implican tareas similares.