

Análisis sobre Salarios en Data Science 2022/23



- Proyecto Final | Segunda Entrega
 - Fabrizio Mateo Heck
 - Tutor: Loenel
 - 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 Industry.....	5
Tabla Size	5
Tabla JobTitle	5
Tabla Country	5

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:

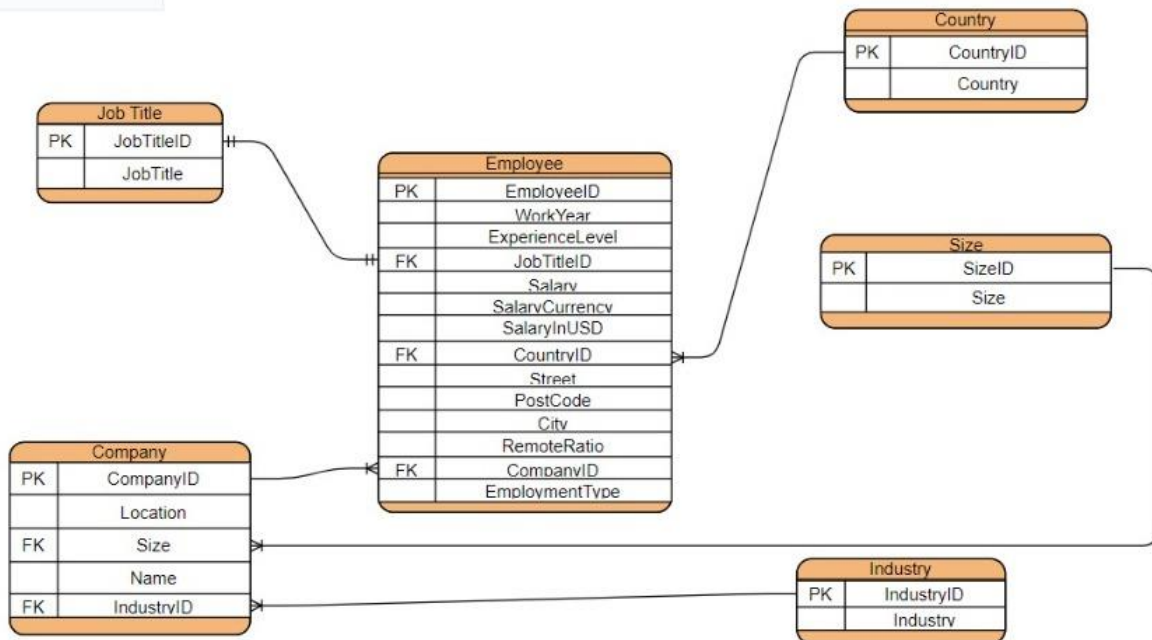
Se buscará comprobar si realmente las empresas de mayor tamaño abonan salarios superiores, como también qué tan influyente es el nivel de experiencia tanto para lograr puestos en grandes empresas como para recibir mejores pagos anuales. Sumado a esto, interesa saber de dónde son la mayoría de las empresas, 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 en Visual Paradigm 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	
ExperienceLevel	Varchar	
JobTitleID	Varchar	FK
Salary	SmallMoney	
SalaryCurrency	Int	
SalaryInUSD	Int	
CountryID	Int	FK
Street	Varchar	
PostCode	Varchar	
City	Varchar	
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	
Size	Varchar	FK
Name	Varchar	
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	PK
Country	Varchar	
MinimumWage	SmallMoney	

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).

Tabla Size

Define en uno de los tres tamaños posibles de empresa

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

Tabla JobTitle

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

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