

Reporte: los salarios de los científicos de datos

Antonio David Gutiérrez Páez, Arnold Torres Maldonado

2023-05-03

Resumen

Resumen tipo TT

Introducción

La ciencia de datos es una disciplina que ha estado a la alza recientemente con los avances tecnológicos y con la cantidad de datos que se producen todos los días, siendo actualmente una disciplina indispensable para el análisis y tratamiento de datos en masa. Debido a esto, la relación entre el salario percibido para esta disciplina contra el año en el que una persona se dedicaba a esto va en subida, siendo el último año uno de los mejores pagados para esta profesión, esto también debido a la alta demanda de procesamiento de datos y la poca disponibilidad de personas profesionales que se dediquen a esta disciplina.

ai-jobs.net es el sitio web de donde proviene

Materiales y métodos

Para el flujo de trabajo del proyecto se hizo uso del modelo CRISP-DM (Cross Industry Standard Process for Data Mining) el cual plantea seis fases que tiene que pasar un proyecto de minería de datos: comprensión de problema, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación del modelo generado. Para observar la comparación entre los años y sus salarios percibidos, Utilizamos el método de regresión lineal por mínimos cuadrados así como el modelo de predicción para conocer como es que se comportarán los salarios para esta profesión a lo largo de los años 2020 - 2023.

***explicar que es la reg lineal, describir el dataset, explicar el crispedm**

Experimentación y resultados

***desglosar el modelo crisp, aplicar todas las secciones del crispedm**

Conclusiones

Referencias bibliográficas

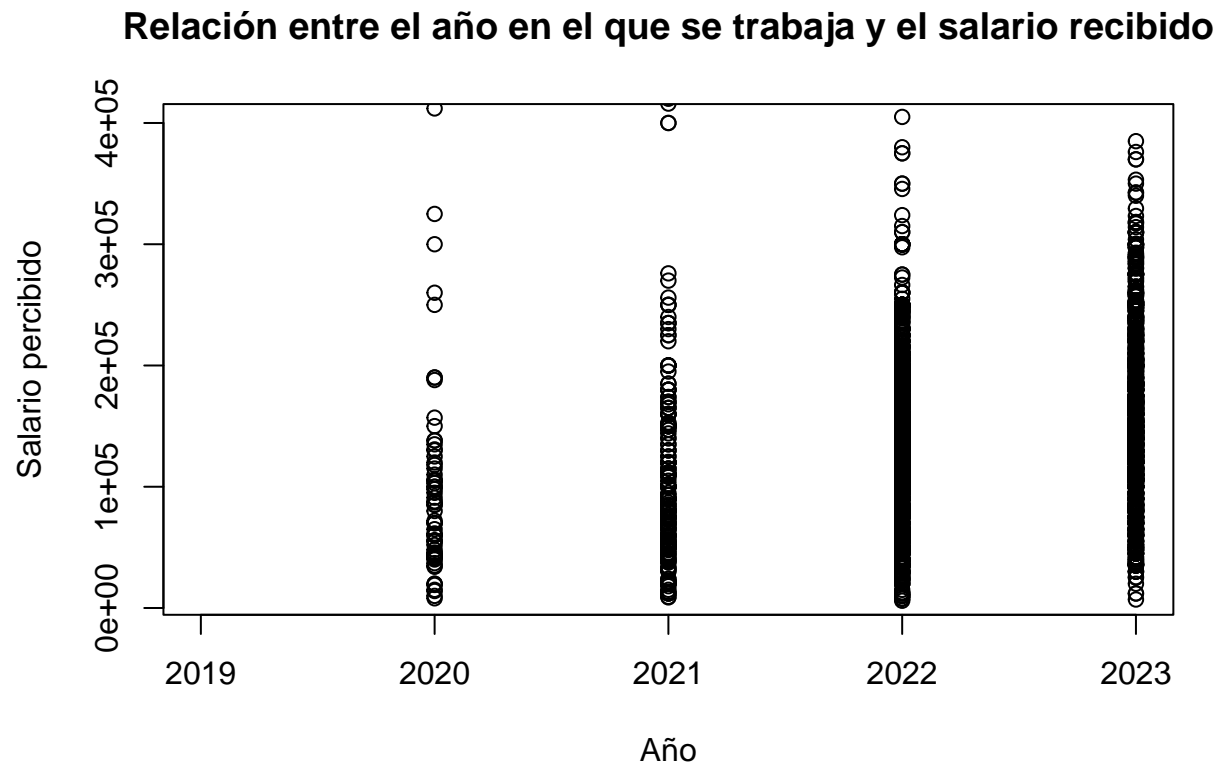
Chaki, A. (2023). Data Science Salaries 2023. [Conjunto de datos]. Kaggle. <https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023>

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.