



Salarios para tipos de trabajos en el mundo de los datos (Informatico)

Institución

Coderhouse

Autor

Ing. Juan Ramirez

28-05-2024



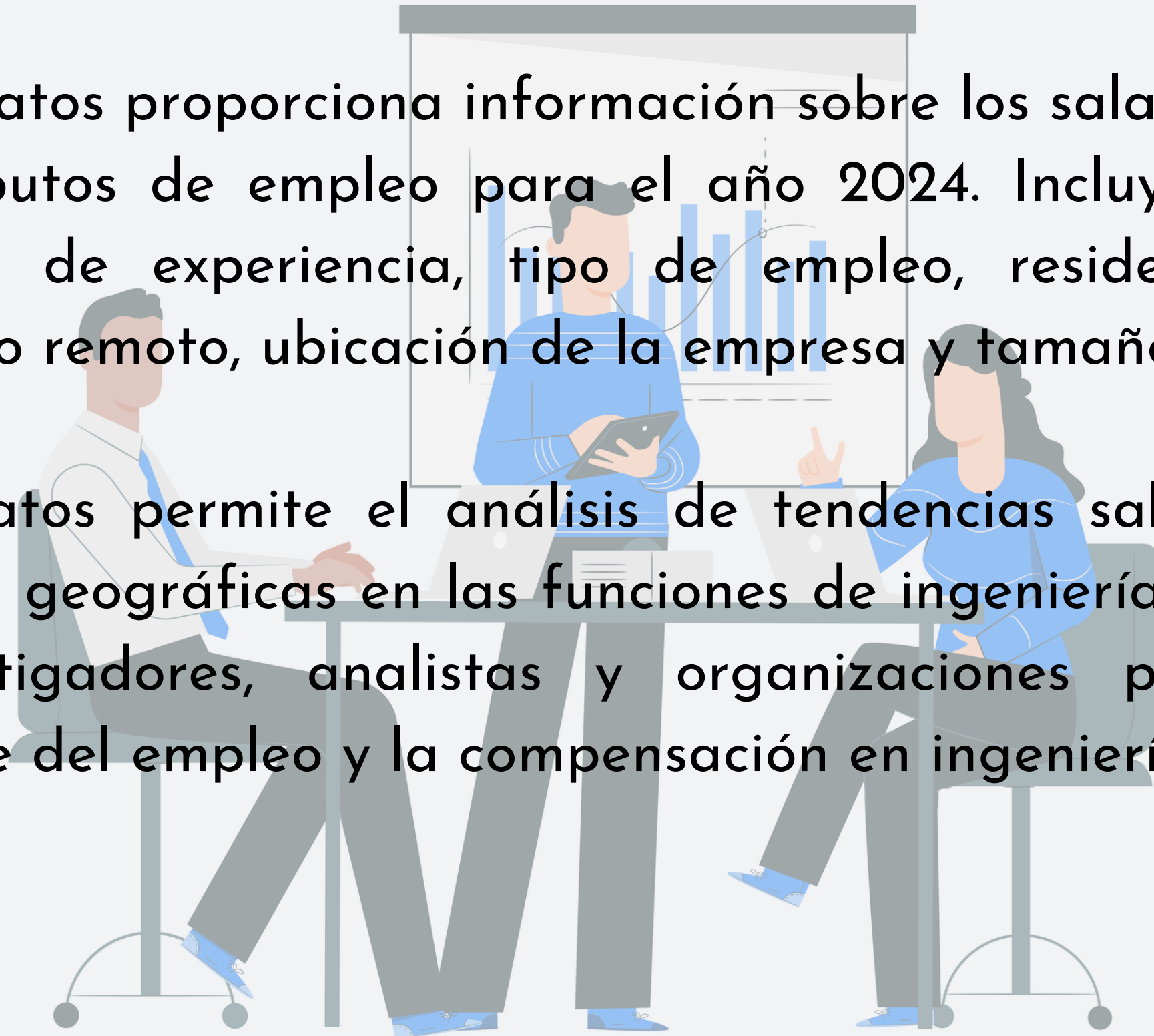
Indice

- 1. Introducción.....3
- 2. Contexto Comercial.....4
- 3. Objetivos y Alcance.....5
- 4. Hipótesis.....6
- 6. Variables del Data set.....8
- 7. Tipo de aprendizaje utilizado.....11
- 8. Modelo de resolución.....12
- 9. Resumen del Proceso de Análisis de Salarios y modelo de ML para Data Scientists.....13
- 10. Conclusión14
- 11. Resumen.....17

1. Introducción

Este conjunto de datos proporciona información sobre los salarios de los ingenieros de datos y los atributos de empleo para el año 2024. Incluye información como salario, cargo, nivel de experiencia, tipo de empleo, residencia del empleado, proporción de trabajo remoto, ubicación de la empresa y tamaño de la empresa.

El conjunto de datos permite el análisis de tendencias salariales, patrones de empleo y variaciones geográficas en las funciones de ingeniería de datos. Puede ser utilizado por investigadores, analistas y organizaciones para comprender el panorama cambiante del empleo y la compensación en ingeniería de datos.



2. Contexto Comercial

Una empresa está en proceso de crear su propio departamento de análisis de datos para reducir su dependencia de consultoras externas. Para lograr este objetivo, es crucial entender los salarios del mercado para diferentes roles de ingeniería de datos y niveles de experiencia. La problemática principal es determinar qué salarios ofrecer a los candidatos para atraer y retener talento calificado en este campo competitivo.

3. Objetivos y Alcances

General

Determinar qué salarios ofrecer a los candidatos para atraer y retener talento calificado en campo de trabajo de datos ("Data Scientist","Data Analyst","Machine Learning Engineer","Data Engineer").

Especifico

- **Establecer una estructura salarial competitiva:** Basada en datos del mercado, para los diferentes roles y niveles de experiencia en ingeniería de datos.
- **Generar un modelo de Machine learning:** desde el conjunto de datos, establecer un patron de aprendizaje para un modelo predictivo salarial.

4. Hipótesis

- **H1:** Los ingenieros de datos con mayor nivel de experiencia ('EX') reciben salarios significativamente más altos en USD que los ingenieros de nivel inicial ('EN'), nivel medio ('MI') y sénior ('SE').
- **H2:** La diferencia salarial entre los niveles 'SE' y 'MI' es significativa, con los ingenieros sénior ganando más.
- **H3:** Utilizando los datos historicos y bajo un modelo supervisado, podemos generar un modelo que prediga con precision los salarios que deden ser asignados a los distintos tipos de cargos



6. Variables del Data set

work_year: El año en que se recogieron los datos (2024).

experience_level: El nivel de experiencia del empleado, categorizado como EX (ingeniero experto), SE (ingeniero sénior), MI (ingeniero de nivel medio) o EL (ingeniero de nivel inicial).

employment_type: El tipo de empleo, como tiempo completo (FT), tiempo parcial (PT), contrato (C) o autónomo (F).

job_title: El título o función del empleado dentro de la empresa, por ejemplo, ingeniero de inteligencia artificial.

salary: El salario del empleado en la moneda local (por ejemplo, 202 730 USD).

salary_currency: La moneda en la que está denominado el salario (por ejemplo, USD).

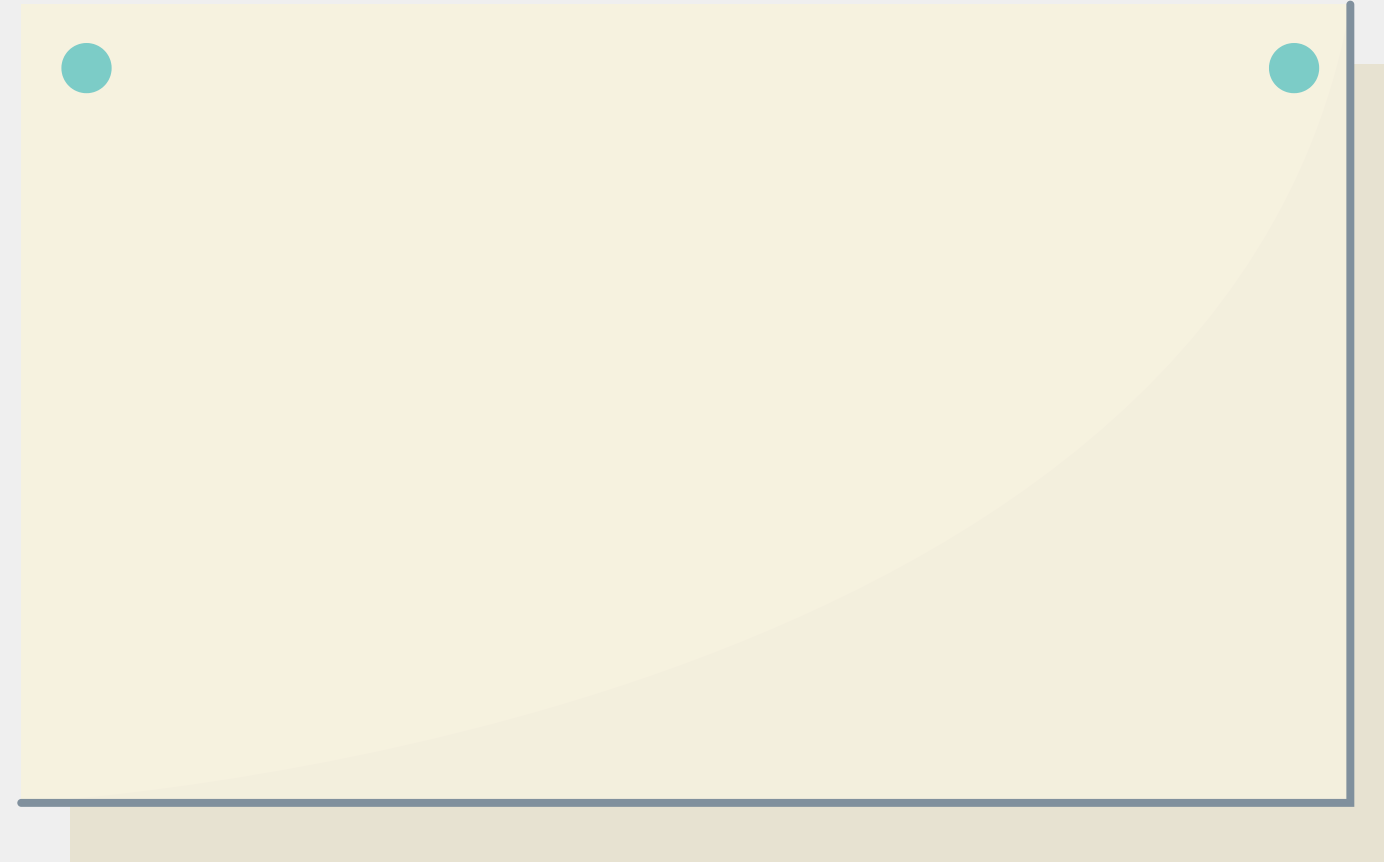
salary_in_usd: El salario convertido a dólares estadounidenses para fines de estandarización.

employee_residence: El país de residencia del empleado.

remote_ratio: La proporción que indica el alcance del trabajo remoto permitido en el puesto (0 para ningún trabajo remoto, 1 para completamente remoto).

company_location: La ubicación de la empresa donde trabaja el empleado.

company_size: El tamaño de la empresa, a menudo categorizado por el número de empleados (S para pequeña, M para mediana, L para grande).



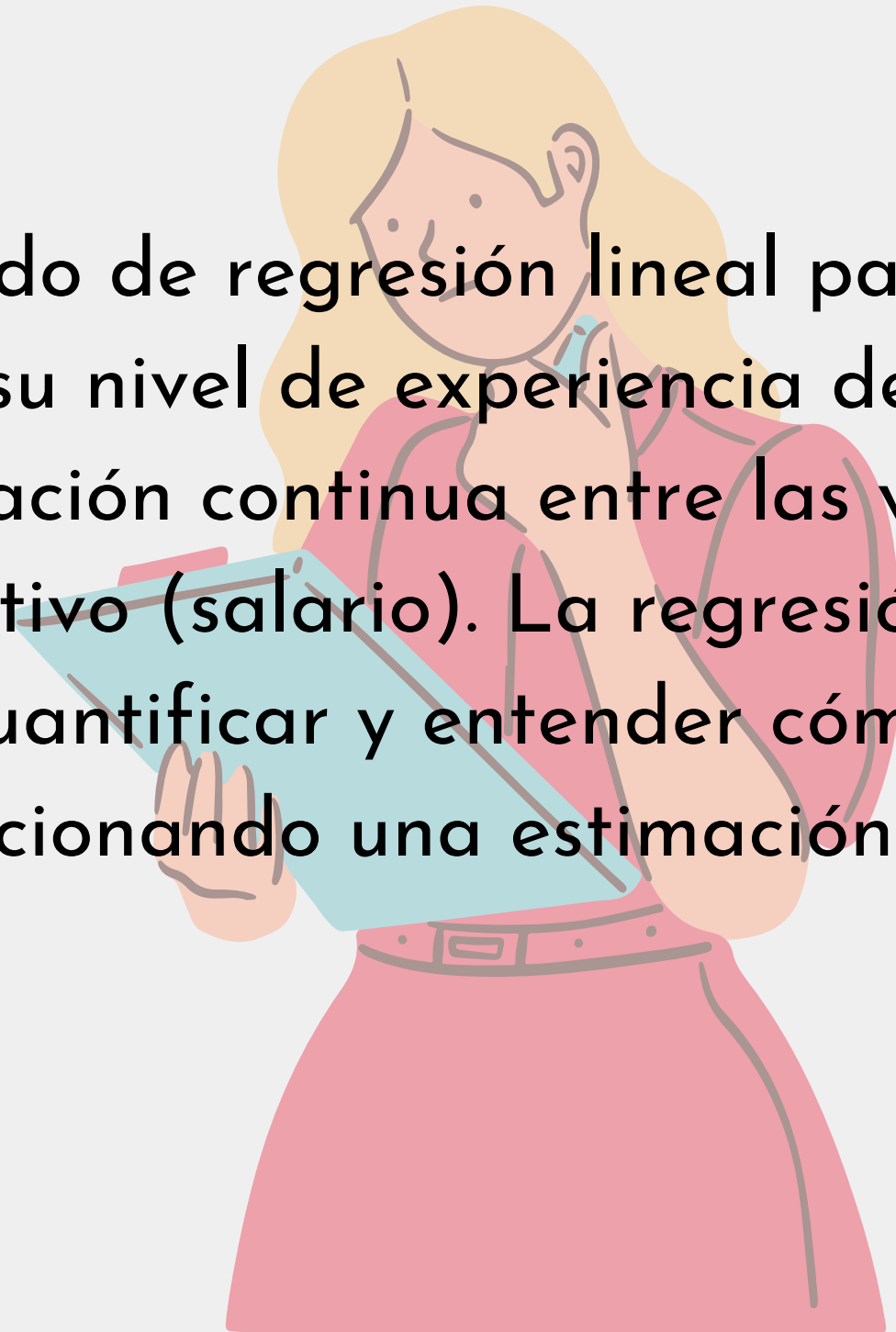
7. Tipo de aprendizaje utilizado

Se utilizó un modelo supervisados ya que son eficaces cuando se cuenta con un conjunto de datos etiquetado, permitiendo al modelo aprender de las observaciones conocidas para hacer predicciones precisas sobre nuevos datos. Esto es crucial en contextos salariales, donde se busca entender y prever tendencias económicas y de compensación basadas en la experiencia laboral.



8. Modelo de resolución

Se utilizó un modelo supervisado de regresión lineal para predecir los salarios de los Data Scientists en función de su nivel de experiencia debido a la naturaleza del problema, que implica una relación continua entre las variables predictoras (nivel de experiencia) y la variable objetivo (salario). La regresión lineal es adecuada para este análisis porque permite cuantificar y entender cómo varía el salario con respecto al nivel de experiencia, proporcionando una estimación directa de esta relación.



9. Resumen del Proceso de Análisis de Salarios y modelo de ML para Data Scientists

1. Importación y Preparación de Datos:

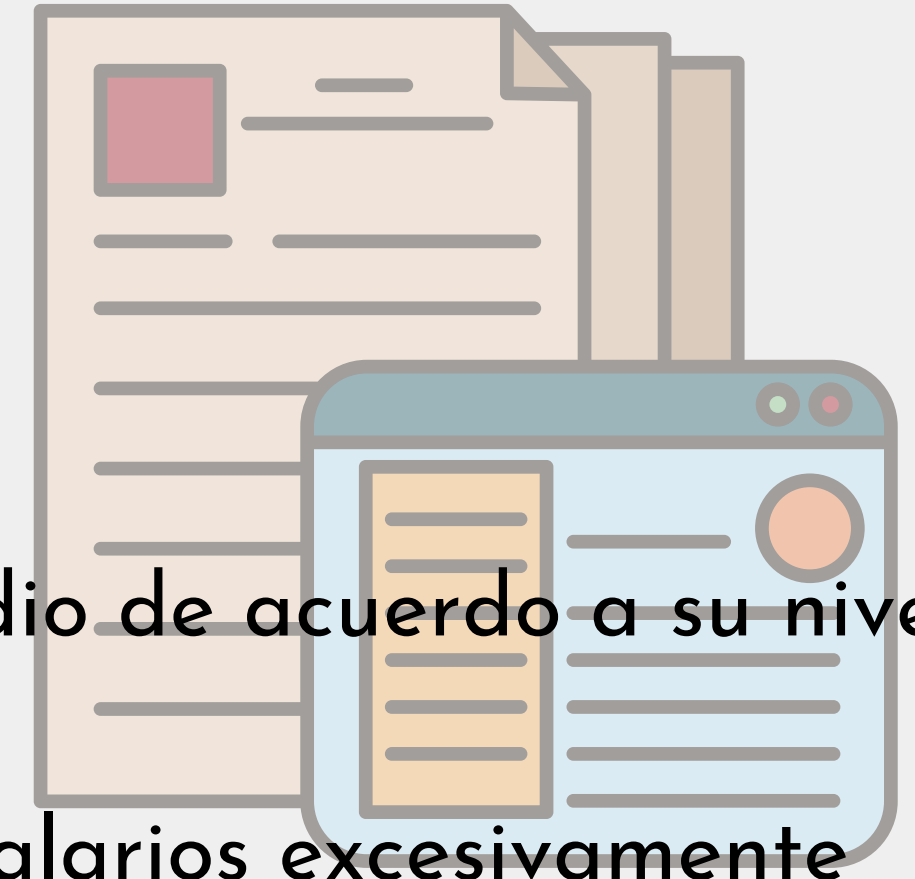
- Se importaron los datos desde un archivo CSV que incluía información sobre nivel de experiencia, tipo de empleo, título del trabajo, salario, residencia del empleado, proporción de trabajo remoto, ubicación de la empresa y tamaño de la empresa.
- Se verificaron existencia de datos nulos o perjudiciales para el set de datos a utilizar
- Se convirtieron las categorías del nivel de experiencia ('EN', 'EX', 'MI', 'SE') a valores numéricos ('1', '4', '2', '3') para facilitar el análisis.
- Se agruparon algunos títulos de trabajo similares bajo una misma categoría, por ejemplo, 'Data Science' y 'Data Scientist'

2. Análisis Exploratorio de Datos:

- Se visualizaron los tipos de trabajos y sus salarios promedio de acuerdo a su nivel de experiencia.
- Se identificaron y eliminaron outliers, como por ejemplo salarios excesivamente altos para 'Data Analyst'.

3. Correlación:

- Se calculó el coeficiente de correlación entre el nivel de experiencia y el salario para los Data Scientists, resultando en un valor de 0.4173, indicando una relación positiva moderada.



10. Conclusión

En el análisis realizado, se encontró que el uso de un modelo supervisado de regresión lineal fue viable para predecir los salarios de los Data Scientists en función de su nivel de experiencia. El modelo mostró una relación positiva moderada entre el nivel de experiencia y el salario, reflejada en un coeficiente de correlación de 0.4173.

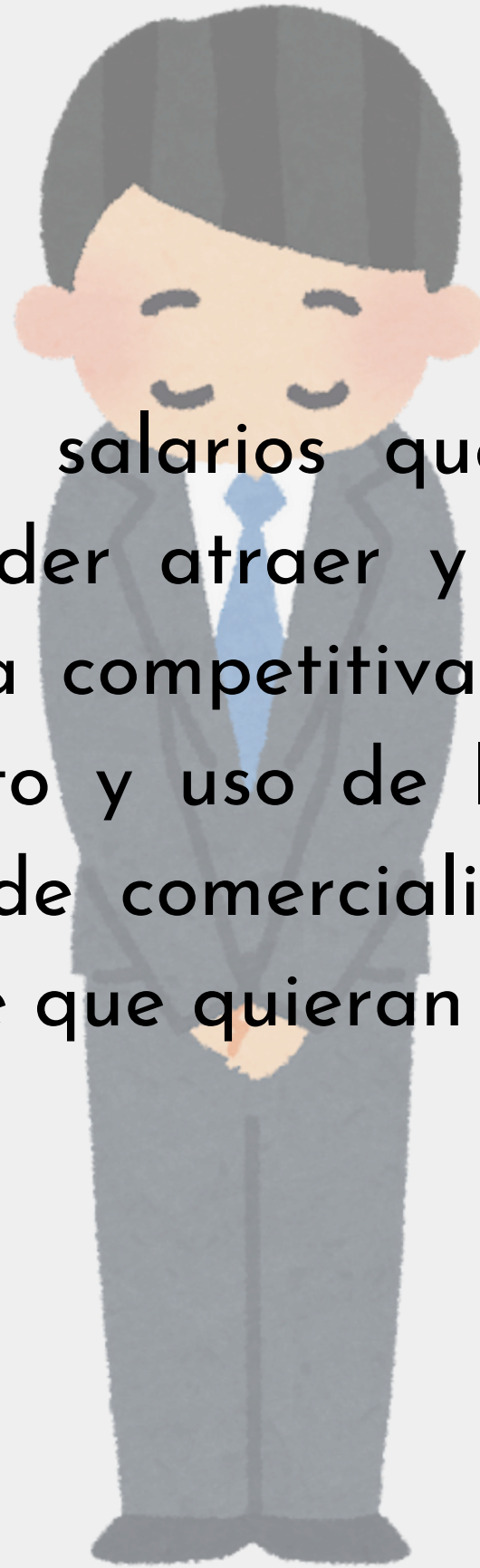
Esta relación indica que, en general, los salarios tienden a aumentar con el nivel de experiencia, lo cual es un resultado esperado y coherente con las prácticas laborales, resaltando así la afirmación de nuestra **Segunda Hipotesis**.

Sin embargo, se observó que en el los 'Data Analysts', los salarios no están alineados con los niveles de experiencia de manera consistente.



Específicamente, se encontró que los 'Data Analysts' con nivel de experiencia “senior o SE” ganan un poco más en promedio que aquellos con nivel de experiencia “experto o EX” (el cual es en terminos de experiencia mayor), lo cual es una anomalía en comparación con otras ocupaciones. Este analisis difiere desde el punto de vista de 'Data Analysts' con la “**Primera Hipotesis**” el cual indicaba que los ingenieros de datos con mayor nivel de experiencia ('EX') reciben salarios significativamente más altos en USD que los ingenieros de nivel

En contraste, los salarios definidos para niveles de experiencia en los distintos tipos de trabajos analizados (como Data Engineer, Machine Learning Engineer, y Data Scientist) son muy similares entre sí y siguen una tendencia esperada de aumento con la experiencia. Este hallazgo resalta una consistencia salarial en la mayoría de los roles, exceptuando el caso de los 'Data Analysts'.



En resumen, podemos determinar salarios que estan basados en la estructura competitiva del mercado para poder atraer y retener candidatos calificados. los cuales conformarán la nueva área competitiva de la empresa, esto con el fin de poder generar mejor procesamiento y uso de la informacion interna, ademas, de disponer del mismo como medio de comercializacion de servicios para analisis y disponibilización de datos de cliente que quieran mejorar su negocio.