Facultad de Ciencias Naturales y Matemáticas

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

ESTADÍSTICA

Proyecto

ANÁLISIS EXPLORATORIO DE LOS SALARIOS DE PROFESIONALES EN EL CAMPO DEL DATA SCIENCE

Integrantes:

1. Rivas Abad Braulio de Jesús
2. Romero Rodríguez Luis Fernando
3. Sornoza Vera Alejandro Francisco
4. Tomalá Tumbaco Karelys Milena
5. Zurita Guerrero Angelo Saul

Tabla de Contenidos

Índice de Tablas

Índice de Gráficas

Índice de Figuras

1 Identificación del Problema

1.1 Motivación

El conjunto de datos seleccionado corresponde a los salarios de los profesionales en el campo de la ciencia de datos, disponible en la plataforma data.world. La elección de este conjunto de datos se basa en varias razones clave. En primer lugar, la ciencia de datos ha experimentado un crecimiento sustancial en los últimos años, consolidándose como una disciplina central en diversas industrias debido a su capacidad para transformar datos en decisiones estratégicas. La relevancia de este campo ha aumentado exponencialmente, impulsada por la creciente necesidad de analizar grandes volúmenes de información en un entorno digital cada vez más complejo. Por lo tanto, analizar los salarios de los profesionales en esta área no solo proporciona una visión actual del mercado laboral, sino que también refleja las tendencias emergentes y la evolución de un campo en constante expansión. Estos factores hacen que el análisis de los salarios de los data scientists sea particularmente pertinente y valioso en el contexto del presente estudio.

1.2 Objetivos

1.2.1 Objetivo General

Realizar un análisis exploratorio de los salarios de profesionales relacionados al campo de los datos entre los años 2020 y 2023 para determinar la profesión mejor pagada.

1.2.2 Objetivos Específicos

a. Comparar los salarios promedio entre científicos de datos y otros profesionales en áreas de datos para obtener aquel que tiene mejor salario.
b. Examinar la relación entre el tamaño de la empresa y el nivel de experiencia de los empleados.
c. Explorar la relación que existe entre la experiencia laboral y el salario en USD de los profesionales para extraer un modelo de regresión lineal que muestre la correlación entre ambas variables.

2 Marco Teórico

2.1 Demanda y Oferta Laboral

La ciencia de datos se ha convertido en una de las profesiones más demandadas debido al aumento exponencial de datos generados diariamente. Empresas de todos los sectores buscan científicos de datos para analizar y extraer información valiosa que les permita tomar decisiones informadas (Coursera, 2023). Conocer los salarios ayuda a entender la oferta y demanda en el mercado laboral, permitiendo a las empresas ajustar sus estrategias de contratación y retención de talento.

2.2 Competitividad Salarial

Los salarios de los científicos de datos varían significativamente según la región, la industria y el nivel de experiencia (Pykes, 2024). Un análisis estadístico permite a los profesionales y a las empresas comparar salarios y asegurarse de que están ofreciendo o recibiendo una compensación competitiva. Esto es esencial para atraer y retener a los mejores talentos en un mercado laboral competitivo.

2.3 Evolución del Campo

El campo de la ciencia de datos está en constante evolución, con nuevas herramientas y técnicas emergiendo regularmente. Los salarios pueden reflejar estas tendencias, mostrando cómo la demanda de habilidades específicas (como el aprendizaje automático o la inteligencia artificial) impacta en la compensación. Un análisis detallado puede revelar qué habilidades son más valoradas y cómo se traduce esto en términos salariales (MisApuntes, 2023).

2.4 Equidad Salarial

Investigar los salarios también es importante para abordar cuestiones de equidad salarial. La ciencia de datos, como muchos otros campos tecnológicos, ha enfrentado desafíos en términos de diversidad e inclusión. Un análisis estadístico puede ayudar a identificar disparidades salariales basadas en género, raza u otros factores, y proporcionar datos para impulsar políticas de equidad salarial (Coursera, 2023).

2.5 Planificación de Carrera

Para los profesionales de la ciencia de datos, conocer las tendencias salariales es vital para la planificación de su carrera. Les permite tomar decisiones informadas sobre su desarrollo profesional, como qué habilidades adquirir o qué industrias explorar. Además, proporciona una base para negociar salarios y beneficios de manera efectiva (Pykes, 2024).

2.6 Impacto Económico

Finalmente, la ciencia de datos tiene un impacto significativo en la economía global. Las empresas que utilizan datos de manera efectiva pueden mejorar su eficiencia, innovar y crecer más rápidamente. Conocer los salarios de los científicos de datos ayuda a cuantificar el valor económico de esta profesión y a justificar inversiones en formación y desarrollo de talento en este campo (MisApuntes, 2023).

Investigar los salarios de los científicos de datos no solo proporciona una visión clara del mercado laboral, sino que también impulsa la competitividad, la equidad y el desarrollo profesional en un campo que es crucial para el futuro de muchas industrias.

3 Marco Analítico

Para llevar a cabo el análisis exploratorio de los salarios de profesionales en el campo de la ciencia de datos, se ha utilizado un conjunto de datos que contiene información detallada sobre las empresas y sus empleados en Estados Unidos. Este dataset es fundamental para alcanzar los objetivos establecidos en la investigación.

3.1 Descripción de las variables

Variables cualitativas	**Tipo**	**Soporte**
experience_level	Ordinal	[Entry-level EN, Mid-level MI, Senior-level SE, Experienced EX]
job_tittle	Nominal	[Data Engineer, Data Scientist, Data Analyst, Machine Learning Engineer, Analytics Engineer, Data Architect, Research Scientist, Data Science Manager, Applied Scientist, Research Engineer, ML Engineer, Data Manager, Machine Learning Scientist, Data Science Consultant, Data Analytics Manager, Computer Vision Engineer, AI Scientist, Other]

company_size	Ordinal	[Small, Medium, Large]
employment_type	Nominal	[Full-time FT, Part-Time PT, Contract CT, Freelance FL]
salary_currency	Nominal	[United States Dollar (USD), Euro (EUR), British Pound Sterling (GBP), Indian Rupee (INR), Canadian Dollar (CAD), Australian Dollar (AUD), Singapore Dollar (SGD), Brazilian Real (BRL), Polish Zloty (PLN), Swiss Franc (CHF), Hungarian Forint (HUF), Danish Krone (DKK), Japanese Yen (JPY), Turkish Lira (TRY), Thai Baht (THB), Israeli Shekel (ILS), Hong Kong Dollar (HKD), Other]
employee_residence	Nominal	[United States (US), United Kingdom (GB), Canada (CA), Spain (ES), India (IN), Germany (DE), France (FR), Portugal (PT), Brazil (BR), Greece (GR), Netherlands (NL), Australia (AU), Mexico

		(MX), Italy (IT), Pakistan (PK), Japan (JP), Ireland (IE), Other]
company_location	Nominal	[United States (US), United Kingdom (GB), Canada (CA), Spain (ES), India (IN), Germany (DE), France (FR), Brazil (BR), Australia (AU), Greece (GR), Netherlands (NL), Mexico (MX), Portugal (PT), Ireland (IE), Singapore (SG), Austria (AT), Japan (JP), Other]

Tabla 1. Descripción de variables cualitativas

Variables cuantitativas	**Tipo**	**Soporte**
work_year	Discreta	[2020-2023]
salary_in_usd	Continua	[5k – 450k]
remote_ratio	Discreta	[0, 50, 100]
salary	Continua	[6k – 30.4M]

Tabla 2. Descripción de variables cuantitativas

3.2 Análisis Exploratorio de Datos

Se realizó el análisis exploratorio de datos para comprender mejor la estructura y características del conjunto de datos, con el fin de identificar patrones y obtener una visión general de la distribución de los datos antes de aplicar métodos de análisis más complejos. Como resultado, se obtuvieron las siguientes gráficas, que permitieron encontrar hallazgos importantes.

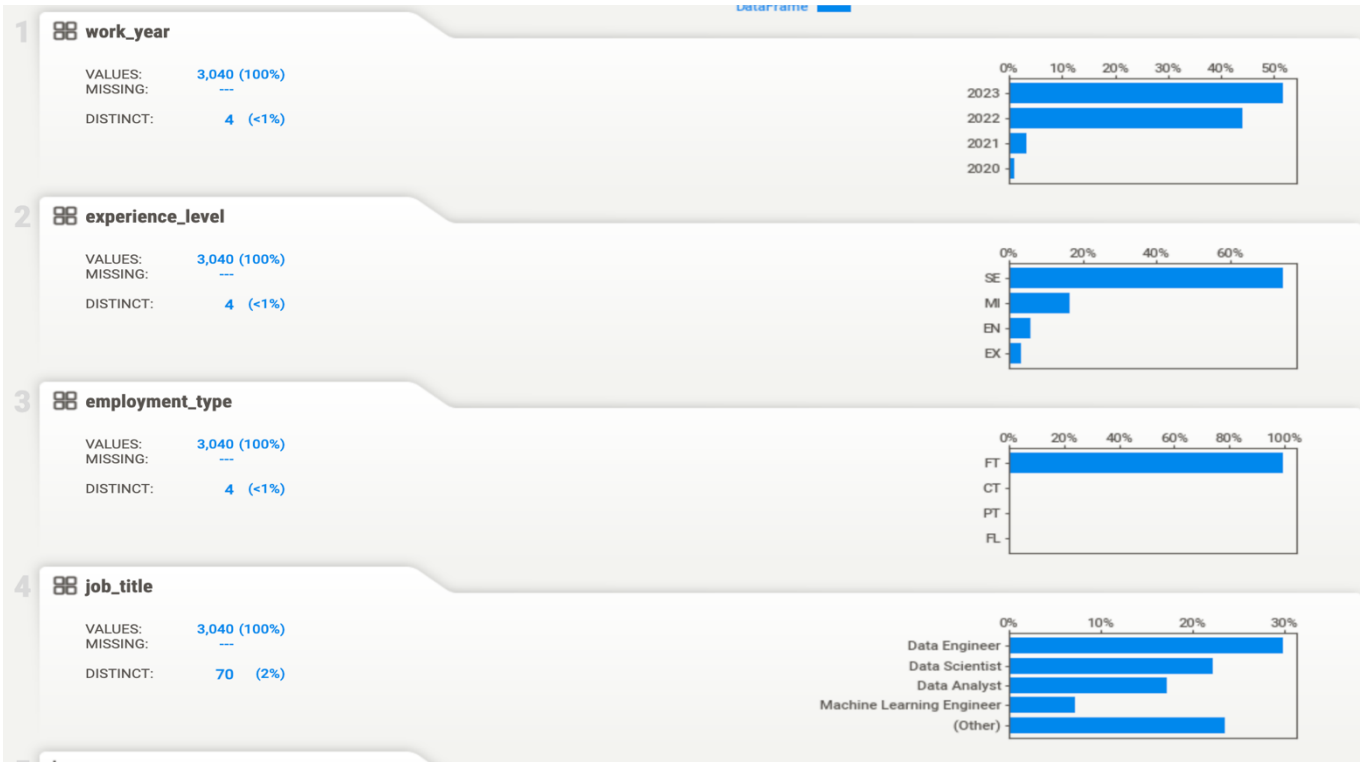

Figura 1. Análisis exploratorio de datos de salarios de profesionales en ciencia de datos parte 1

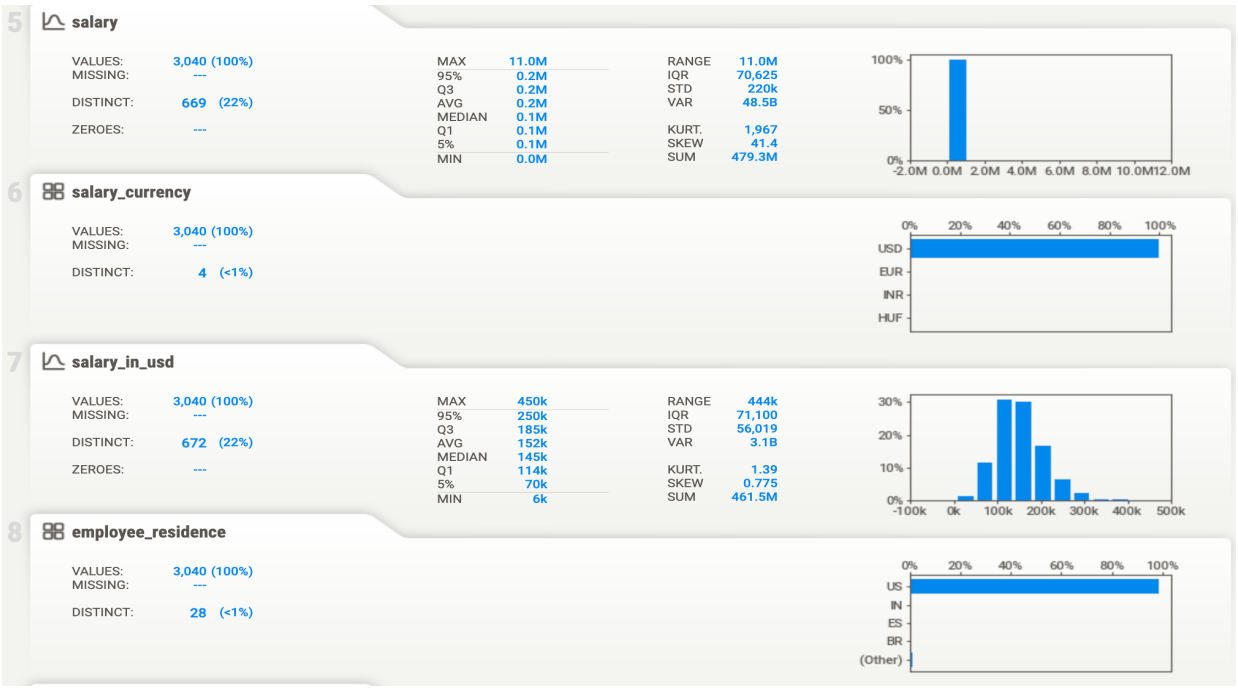

Figura 2. Análisis exploratorio de datos de salarios de profesionales en ciencia de datos parte 2

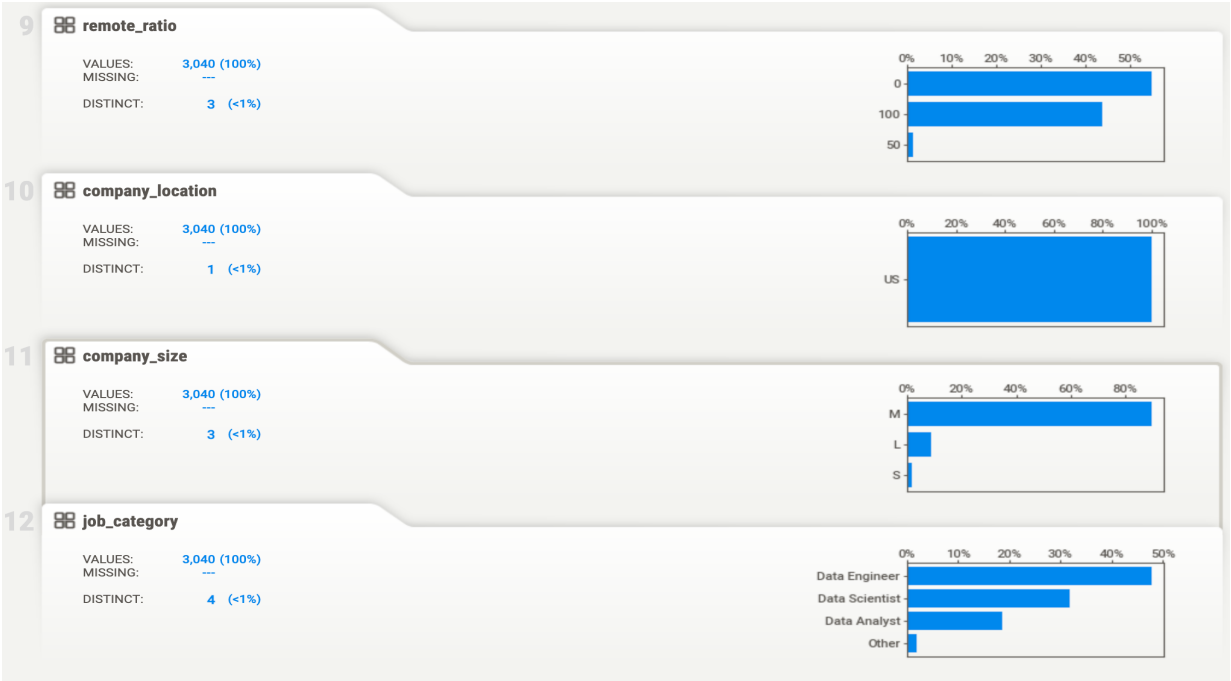

Figura 3. Análisis exploratorio de datos de salarios de profesionales en ciencia de datos parte 3

3.3 Hallazgos

3.3.1 Figura 1:

- Observamos que el nivel de experiencia más común es Senior SE, lo que indica que la mayoría de los trabajadores son experimentados en el área.

- Se observa que la mayoría de los datos corresponden a los años 2022 y 2023, además existe una alta concentración en el 2023.

- Los roles más comunes son Data Engineer, Data Scientist y Data Analytics, lo que sugiere una alta demanda de estos perfiles en los últimos tres años.

3.3.2 Figura 2:

- La distribución salarial muestra una amplia variabilidad, con algunos salarios significativamente más altos que otros. Se puede ver que los salarios se encuentran en un promedio.

- Se observa que menos del 50% de las posiciones son completamente remotas, lo que indica una tendencia media-baja hacia el trabajo remoto.

- La mayor parte de las empresas están ubicadas en Estados Unidos, lo que nos lleva a centrarnos exclusivamente en el mercado estadounidense. Esto también explica la predominancia de la moneda local (dólar).

3.3.3 Figura 3:

- La mayoría de las posiciones provienen de empresas medianas y grandes, lo que indica que son las más propensas a ofrecer estos tipos de trabajos.

4 Comparación de salarios promedio entre científicos de datos y otros profesionales en áreas de datos.

Para realizar la comparación de los salarios promedio entre Científicos de Datos (Data Scientists) y otros profesionales en el ámbito de los datos, específicamente Ingenieros de Datos (Data Engineers) y Analistas de Datos (Data Analysts), se utilizó el conjunto de datos ds_salaries_US.csv, el cual proporciona información detallada sobre los salarios en diferentes categorías profesionales dentro del campo de la ciencia de datos. Se enfocó en las variables job_category (categoría de trabajo) y salary_in_usd (salario en dólares estadounidenses) para realizar el análisis.

```
[1] " Media Data Scientist "
[1] 163644.7
[1] " Media Data Engineer "
[1] 158215.4
[1] " Media Data Engineer "
[1] 115438
```

Figura 4. Salarios promedio por categoría de profesionales en áreas de datos.

Primero, se filtraron los datos para cada categoría profesional y se calcularon los salarios promedio para cada grupo. Los resultados obtenidos muestran que el salario promedio para los Científicos de Datos(mean_DS USD) es $163644.7, para los Ingenieros de Datos(mean_DE USD) es $158215.4 , y para los Analistas de Datos(mean_DA USD) es $115438 .

Con base a estos se realizaron distintas pruebas de hipótesis, considerando que los Científicos de Datos son los que más ganan respecto a los de áreas de datos.

```r
# Ho: median_DS >= median_DE
# Ha : median_DS < median_DE
test_result_DS_DE <- t.test(salaries_DS, salaries_DE, alternative = "less",
                      conf.level = 0.95)
```

Figura 5. Código R de Prueba de Hipótesis entre Científicos de Datos y Ingenieros de Datos

```
        Welch Two Sample t-test

data:  salaries_DS and salaries_DE
t = 2.3504, df = 2022.3, p-value = 0.9906
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 9230.485
sample estimates:
mean of x mean of y
 163644.7  158215.4
```

Figura 6. Resultado de Prueba de Hipótesis entre Científicos de Datos y Ingenieros de Datos

Para evaluar si las diferencias observadas en los salarios promedio, se realizaron pruebas t. La comparación entre Científicos de Datos e Ingenieros de Datos reveló un valor p de 0.9906. Dado que este valor p es mayor que el nivel de significancia de 0.05, no se rechaza la hipótesis nula o lo que nosotros supusimos, por lo tanto, se puede concluir que los Científicos de Datos ganan más que los Ingenieros de Datos en promedio.

```r
# Ho: median_DS >= median_DA
# Ha : median_DS < median_DA

test_result_DS_DA <- t.test(salaries_DS, salaries_DA, alternative = "less",
                            conf.level = 0.95)
```

Figura 7. Código R de Prueba de Hipótesis entre Científicos de Datos y Analistas de Datos

```
        Welch Two Sample t-test

data:  salaries_DS and salaries_DA
t = 19.302, df = 1459.9, p-value = 1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 52317.34
sample estimates:
mean of x mean of y
 163644.7  115438.0
```

Figura 8. Resultado de Prueba de Hipótesis entre Científicos de Datos y Analistas de Datos

Así mismo, se realizó la comparación entre Científicos de Datos e Analistas de Datos donde se obtuvo un valor p aproximado de 1. Dado que este valor p vuelve a ser mayor que el nivel de significancia de 0.05, no se rechaza la hipótesis nula o lo que nosotros supusimos, por lo tanto, se puede concluir que los Científicos de Datos ganan más que los Analistas de Datos en promedio.

En resumen, los resultados de las pruebas de hipótesis realizadas revelan que, aunque no se detectaron diferencias estadísticamente significativas en los salarios promedio entre Científicos de Datos e Ingenieros de Datos, ni entre Científicos de Datos y Analistas de Datos, los Científicos de Datos presentan los salarios promedio más altos en comparación con los otros dos roles. No obstante, es relevante observar que los Ingenieros de Datos están representados en un 50.1% más que los Científicos de Datos en el conjunto de datos, con 1,453 casos frente a 968. Esto subraya que, aunque los Científicos de Datos reciben una compensación más alta, su número de profesionales es menor en comparación con los Ingenieros de Datos. La mayor representación de Ingenieros de Datos podría reflejar una creciente demanda de este rol, pero también pone de relieve el valor y la competitividad asociados con la profesión de Científico de Datos.

5 Relación entre Tamaño de la Empresa y Nivel de Experiencia de los Empleados

Para determinar si existe una asociación significativa entre estas dos variables, se creó una tabla de contingencia para representar la distribución de niveles de experiencia (junior o entry level, mid level o intermedio, senior, experto) en función del tamaño de la empresa (grande, mediana, pequeña). La tabla de frecuencia absoluta permitió visualizar la frecuencia de cada combinación de nivel de experiencia y tamaño de empresa, también se calculó el valor de la frecuencia esperada para cada una de las celdas.

	L	M	S	Sum
EN	41	123	19	183
EX	4	90	4	98
MI	60	423	14	497
SE	158	2087	17	2262
Sum	263	2723	54	3040

Figura 9. Tabla de frecuencia absoluta y conteos marginales

	L	M	S
EN	15.831908	163.91743	3.250658
EX	8.478289	87.78092	1.740789
MI	42.997039	445.17467	8.828289
SE	195.692763	2026.12697	40.180263

Figura 10. Tabla de frecuencia esperada

Para determinar si existe una asociación significativa entre el tamaño de la empresa y el nivel de experiencia, se realizó una prueba de chi-cuadrado de independencia, la cual compara las dos tablas de frecuencias. Esta prueba evaluó si las distribuciones de los niveles de experiencia son independientes del tamaño de la empresa.

```
        Pearson's Chi-squared test

data:  freqab
X-squared = 165.2, df = 6, p-value < 2.2e-16

> cat("El valor p es:", chisq_test$p.value, "\n")
El valor p es: 4.680104e-33
```

Figura 11. Captura en R de los valores la prueba Chi-cuadrado

Con un 5% de significancia la prueba de chi-cuadrado reveló una relación entre el tamaño de la empresa y el nivel de experiencia de los empleados, debido a que el valor p es extremadamente pequeño y menor al nivel de significancia. Esto sugiere que el tamaño de la empresa influye en la distribución del nivel de experiencia de sus empleados.

Debido a la relación entre las dos variables, se procedió a comparar las proporciones de empleados con un nivel de experiencia senior, en empresas grandes frente a empresas medianas. Se formularon las siguientes hipótesis:

- Hipótesis Nula (H0): La proporción de empleados con un nivel de experiencia senior en empresas grandes es menor o igual que la proporción en empresas medianas *(p1 <= p2)*.
- Hipótesis Alternativa (Ha): La proporción de empleados con un nivel de experiencia en empresas grandes es mayor que la proporción en empresas medianas *(p1 > p2)*.

Se utilizó una prueba de hipótesis de diferencia de proporciones, para evaluar si había diferencias significativas entre las proporciones del nivel de experiencia senior en las dos categorías de tamaño de empresa. La prueba Z para las proporciones mostró un estadístico Z de -5.94 y un valor p extremadamente alto (cercano a 1).

```
        2-sample test for equality of proportions with
        continuity correction

data:  c(p1, p2) out of c(n1, n2)
X-squared = 34.401, df = 1, p-value = 1
alternative hypothesis: greater
95 percent confidence interval:
 -0.2191902  1.0000000
sample estimates:
   prop 1    prop 2
0.6007605 0.7664341
```

Figura 12. Resultados de R de la prueba de proporciones

Se realizaron los cálculos en R y con base en estos resultados, no se rechazó la hipótesis nula, el valor p es mayor al nivel de significancia del 5%. Esto indica que no hay evidencia suficiente para afirmar que la proporción de empleados con un nivel de experiencia senior en empresas grandes es mayor que en empresas medianas.

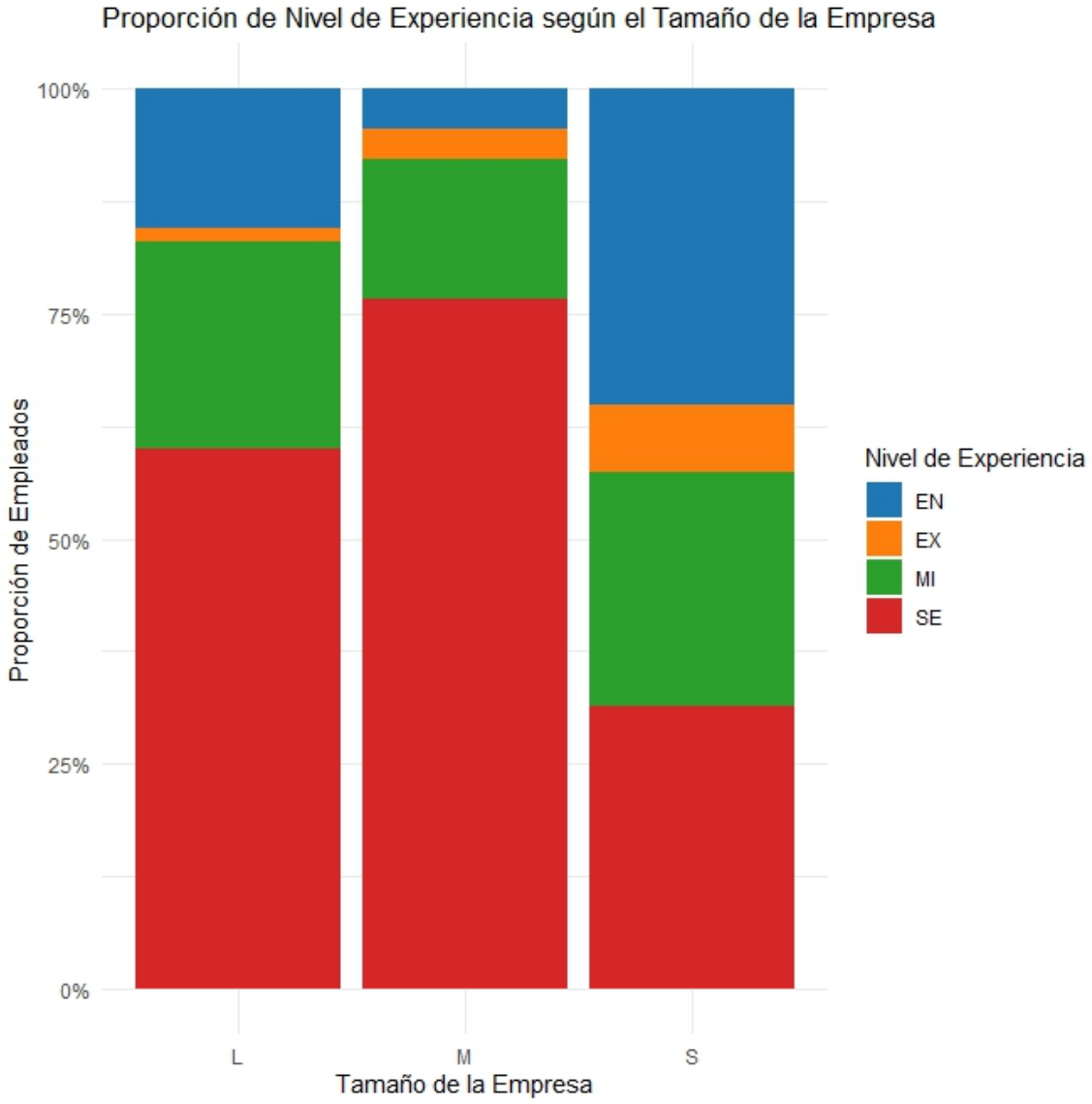

Gráfico1. Proporción del nivel de experiencia de los trabajadores según el tamaño de la empresa

Al hacer un gráfico de proporciones del nivel de experiencia de los trabajadores según el tamaño de la empresa, se tiene una visión general de su distribución, lo que corrobora la prueba de hipótesis anterior, y la proporción de trabajadores nivel senior es mayor en las empresas medianas frente a las grandes.

En el estudio se encontró una asociación significativa entre el tamaño de la empresa y el nivel de experiencia de los empleados, lo que indica que el tamaño de la empresa tiene un impacto en la distribución de los niveles de experiencia. Sin embargo, no se encontraron diferencias significativas en las proporciones del nivel de experiencia senior entre empresas grandes y medianas según la prueba de proporciones realizada.

6 Modelo de regresión lineal

Ahora se realizará el análisis entre las variables experience_level y salary_in_usd (nivel de experiencia y salario en USD) para poder determinar cuál de los niveles de experiencia tiene mayor sueldo y extraer un modelo de regresión lineal para poder correlacionar ambas variables. Para ello, primero se procederá a realizar un diagrama de cajas entre las variables.

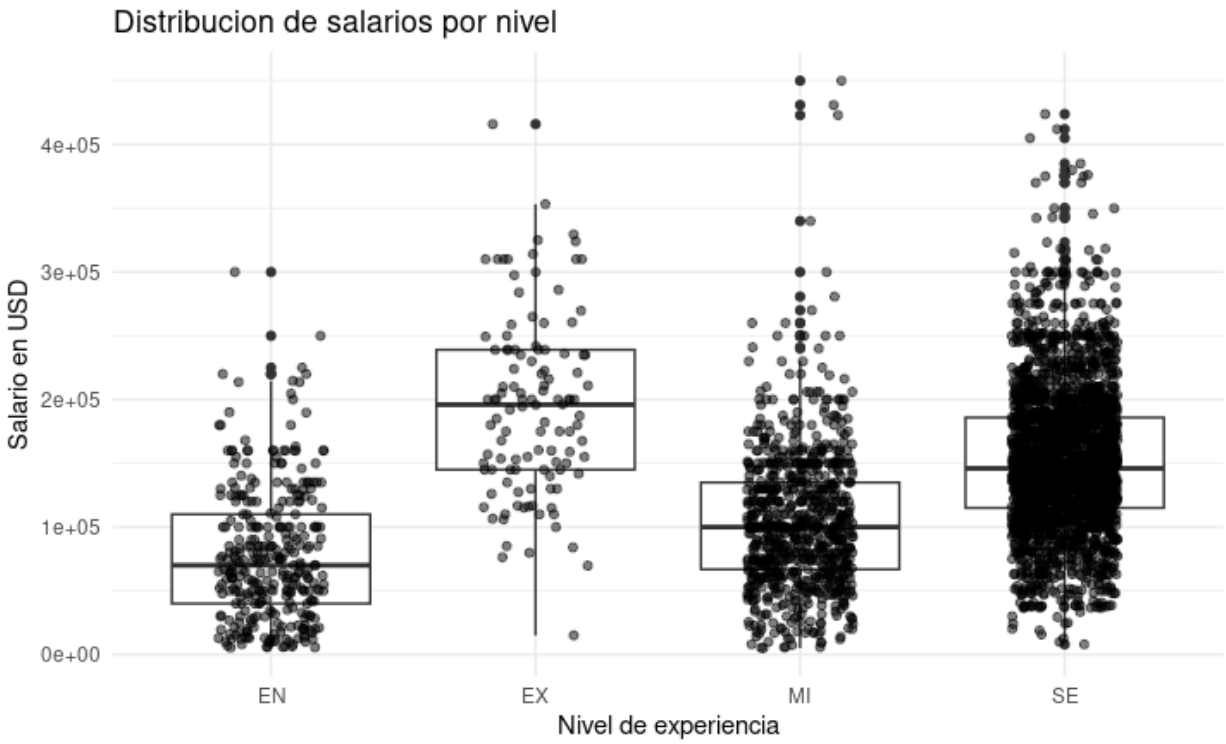

Gráfico2. Diagrama de cajas entre el nivel de experiencia y salario en usd

Nos podemos dar cuenta que parece existir cierta relación entre el nivel de experiencia y el salario percibido en dólares. El orden aparente de menor a mayor empieza por EN (entry level), MI (mid level), SE (senior engineer) y termina en EX (experienced) el cual tiene el mayor sueldo promedio. Ahora, se procederá a transformar los datos de nivel de experiencia, los cuales son ordinales, en

números, es decir, aplicar un proceso llamado encoding para poder obtener un modelo de regresión lineal entre dos variables que sean numéricas. Se obtienen los siguientes resultados:

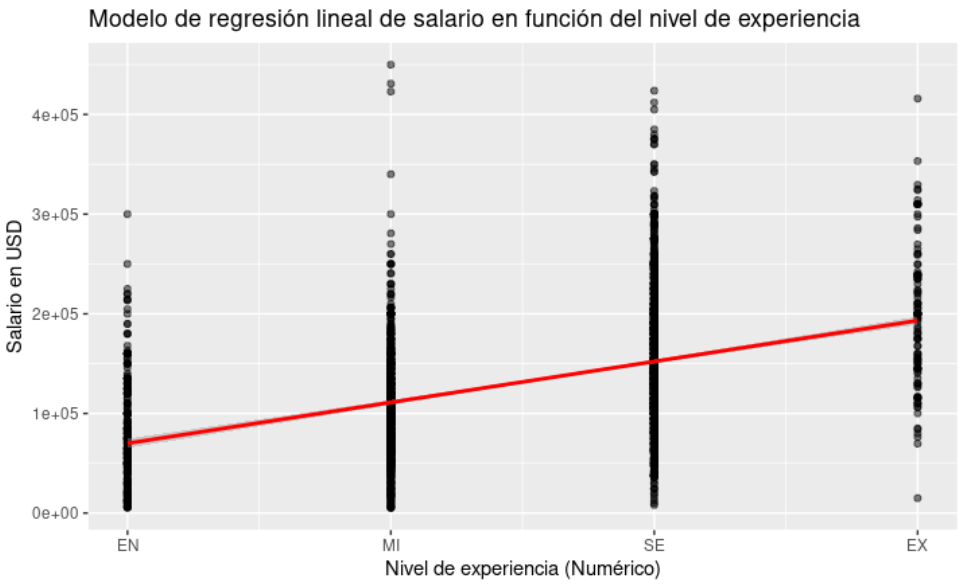

Gráfico3. Scatterplot entre el nivel de experiencia y salario en USD además del modelo de regresión lineal

Se puede observar de mejor manera que a medida que aumente el nivel de experiencia, también aumenta en general el salario en USD. La relación es positiva, es decir, que a medida que el nivel de experiencia aumenta, también lo hace el sueldo. Los parámetros extraídos del modelo son los siguientes.

```
Call:
lm(formula = salary_in_usd ~ experience_level_num, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-178181  -38152   -7124   33776  338934

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)            28951       3718   7.786 8.86e-15 ***
experience_level_num   41058       1361  30.158  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 56580 on 3753 degrees of freedom
Multiple R-squared:  0.1951,    Adjusted R-squared:  0.1949
F-statistic: 909.5 on 1 and 3753 DF,  p-value: < 2.2e-16
```

Figura 13. Captura de R studio sobre los parámetros del modelo de regresión lineal

Se observa que el intercepto es $\beta_0 = \$28951$ y la pendiente $\beta_1 = \$41058$. Los coeficientes del modelo (Intercept y experience_level_num) tienen valores significativos (***), lo que indica que ambos tienen un impacto significativo en la variable dependiente salary_in_usd. Esto se basa en los valores de p (Pr(>|t|)) que son muy pequeños (menores que 0.001). Multiple R-squared: 0.1951. Esto indica que aproximadamente el 19.51% de la variación en los salarios puede explicarse por el nivel de experiencia. Este valor es relativamente bajo, sugiriendo que hay otros factores importantes no incluidos en el modelo que podrían explicar la variación en los salarios. Adjusted R-squared: 0.1949. Es muy cercano al Multiple R-squared, lo cual es bueno, ya que indica que el modelo no está sobreajustado. Residual Standard Error: 56580. Esto representa la desviación estándar de los residuales. Un error estándar más bajo indica que las predicciones están más cerca de los valores observados, pero dado el contexto del rango de los salarios, podría ser considerado alto. El valor del F-statistic es 909.5 con un valor de p muy pequeño (< 2.2e-16), lo cual indica que el modelo en su conjunto es significativo. Esto significa que el nivel de experiencia está

relacionado con la variable dependiente. El modelo muestra que el nivel de experiencia tiene un impacto estadísticamente significativo en el salario, pero el R-cuadrado sugiere que el modelo no explica mucha de la variación en los salarios (solo alrededor del 19.51%). Esto indica que el modelo podría mejorarse añadiendo más variables predictoras relevantes para capturar mejor la variabilidad en los salarios. Sin embargo, parece ser un modelo suficiente hasta cierto punto dado que la variable experiencia inicialmente no era numérica y solo usamos una variable.

7 Conclusiones

Se realizó un análisis comparativo de los salarios promedio entre distintos roles profesionales dentro del campo de la ciencia de datos, centrándose en los años 2020 a 2023. Utilizando diversas herramientas, se investigó específicamente si los Data Scientists ganan más que otros profesionales como Data Engineers y Data Analysts. Los resultados mostraron que, aunque las diferencias no fueron estadísticamente significativas, los Data Scientists tienen el salario promedio más alto, lo que sugiere que este rol es de los más valorados y mejor remunerados en el sector.

Para validar estas observaciones, se aplicaron pruebas de hipótesis que compararon los salarios promedio de Data Scientists con los de Data Engineers y Data Analysts. A pesar de que los valores p obtenidos indicaron que no había diferencias significativas, los Data Scientists siguen liderando en términos de salario promedio. Este hallazgo es clave para entender la estructura salarial en el campo de la ciencia de datos y subraya la importancia del rol de Data Scientist en comparación con otros roles dentro del mismo ámbito.

Además, se logró identificar que los Data Scientists tienden a ser mejor remunerados en comparación con otros profesionales, aunque las diferencias no sean estadísticamente significativas. Este resultado proporciona una visión valiosa sobre las tendencias salariales en este campo, destacando el valor del rol de Data Scientist y su posicionamiento como uno de los más competitivos en el mercado laboral de la ciencia de datos.

Se exploró la relación entre el tamaño de la empresa y el nivel de experiencia de los empleados en ciencia de datos, para determinar si existe una asociación significativa entre estas dos variables.

Utilizando una tabla de contingencia, se analizaron las proporciones de empleados con diferentes niveles de experiencia en empresas de distintos tamaños, comparando específicamente las categorías de empresas grandes y medianas. Este análisis permitió visualizar la distribución de la experiencia laboral en función del tamaño de la empresa.

Para evaluar la asociación entre el tamaño de la empresa y el nivel de experiencia, se aplicó una prueba de chi-cuadrado de independencia. Los resultados de esta prueba revelaron que existe una asociación significativa, lo que indica que el tamaño de la empresa influye en la distribución de los niveles de experiencia de los empleados. Sin embargo, al realizar una prueba Z para comparar las proporciones de empleados con distintos niveles de experiencia entre empresas grandes y medianas, no se encontraron diferencias significativas, lo que sugiere que, aunque hay una asociación general, las diferencias en las proporciones no son suficientemente marcadas.

Encontramos que el tamaño de la empresa tiene un impacto en la distribución de los niveles de experiencia de los empleados en el campo de la ciencia de datos. Las pruebas indicaron que no hay evidencia suficiente para afirmar que las empresas más pequeñas tengan más empleados con un nivel de experiencia que las grandes. Estos hallazgos destacan la complejidad de la relación entre el tamaño de la empresa y la experiencia laboral, sugiriendo que otros factores también podrían influir en esta dinámica.

Exploramos la relación entre la experiencia laboral y el salario en USD de los profesionales en ciencia de datos, para determinar cómo se correlacionan estas dos variables. Para lograr esto, se construyó un modelo de regresión lineal que permitió analizar la relación entre el nivel de experiencia y el salario. Los resultados mostraron una correlación positiva significativa, indicando

que a medida que aumenta la experiencia del profesional, también lo hace su salario. Este hallazgo es consistente con las expectativas en el mercado laboral, donde la experiencia suele estar asociada con una mayor compensación.

El análisis reveló que, aunque la experiencia es un factor significativo en la determinación del salario, el modelo de regresión lineal utilizado explicó aproximadamente el 19.51% de la variabilidad en los salarios. Esto sugiere que, aunque la experiencia laboral es importante, existen otros factores que también influyen en la determinación de la compensación salarial y que no fueron incluidos en este modelo. Aun así, la relación positiva encontrada refuerza la importancia de la experiencia como un componente clave en la estructura salarial dentro del campo de la ciencia de datos.

Se confirmó que la experiencia laboral es un predictor crucial del salario en el campo de la ciencia de datos, con profesionales más experimentados recibiendo salarios más altos en promedio. Aunque el modelo de regresión lineal mostró que la experiencia no explica toda la variabilidad en los salarios, los resultados obtenidos proporcionan una base sólida para entender la importancia de la experiencia en la remuneración de los profesionales en este campo, destacando la necesidad de considerar también otros factores que podrían afectar la estructura salarial.

Concluimos que los Data Scientists son los profesionales mejor remunerados dentro del campo de la ciencia de datos en el período 2020-2023, destacando su relevancia y valor en el mercado laboral con sueldos que ***no dependen exclusivamente*** del nivel de experiencia o tamaño de la empresa. Este proyecto ha proporcionado una visión integral y clara de las dinámicas salariales en este campo, validando su importancia como herramienta para la toma de decisiones informadas.

8 Referencias

Coursera. (2023, 15 junio). *¿Qué es un data scientist? Salario, habilidades y cómo llegar a serlo*. Coursera. https://www.coursera.org/mx/articles/what-is-a-data-scientist

MisApuntes. (2023, 7 octubre). Salario científico de datos en España: ¿Cuánto se gana? - MisApuntes. *MisApuntes*. https://misapuntesdedatascience.es/cuanto-gana-un-cientifico-de-datos-en-espana/

Pykes, K. (2024, 29 Febrero). *La importancia de los datos: 5 razones principales*. https://www.datacamp.com/es/blog/importance-of-data-5-top-reasons

9 Anexos

Statistics/Obj1.R at d43d2aebbbb3bc0406e86fd976f40423138433e7 · aszurita/Statistics. (s.f.). GitHub. https://github.com/aszurita/Statistics/blob/d43d2aebbbb3bc0406e86fd976f40423138433e7/Obj1.R

Statistics/Obj2.r at d43d2aebbbb3bc0406e86fd976f40423138433e7 · aszurita/Statistics. (s.f.). GitHub. https://github.com/aszurita/Statistics/blob/d43d2aebbbb3bc0406e86fd976f40423138433e7/Obj2.r

Statistics/Obj3.R at d43d2aebbbb3bc0406e86fd976f40423138433e7 · aszurita/Statistics. (s.f.). GitHub. https://github.com/aszurita/Statistics/blob/d43d2aebbbb3bc0406e86fd976f40423138433e7/Obj3.R

AnalisisExploratorioDatos : https://statistics-70sf.onrender.com/