



REPORTE FASE I

Descubrimiento del proyecto

“Análisis de los salarios globales para científicos de datos”

Introducción a la ciencia de datos

Ricardo Xochitiotzi Flores

20/10/2024

M.C. Jaime Alejandro Romero Sierra



Objetivo.

Comprender las variables que impactan en los sueldos y las posibilidades de ser contratado, esto en base a la predicción del salario según el puesto de trabajo, la experiencia, entre otros factores clave.

Descripción del problema.

El mercado de trabajo para científicos de datos es diverso y constantemente en la actualidad seguirá creciendo significativamente, y los salarios varían ampliamente según la ubicación geográfica, la experiencia, el tipo de trabajo y el sector. El problema radica en identificar las variables más influyentes para hacer predicciones salariales realistas y evaluar la probabilidad de empleo para recién egresados en diferentes roles de ciencia de datos.

Recursos digitales.

Se utilizarán herramientas como Python y “Pandas” para la manipulación de datos, además de “Matplotlib” para visualización inicial y ocuparemos “Seaborn” para una visualización mas avanzada de los datos.

Nuestras fuentes de datos disponibles se basan en:

- Año en el que se pagó el salario.
- **Experiencia del científico de datos.**
- Tipo de empleo que ocupa.
- **El puesto desempeñado (título).**
- **El monto pagado.**
- **Datos demográficos de la empresa.**
- Tamaño de la empresa.



Hipótesis iniciales.

- Los científicos de datos con mayor experiencia recibirán salarios más altos independientemente de la ubicación.
- Los trabajos de ciencia de datos en empresas grandes (L) tienden a pagar más que en empresas pequeñas (S) o medianas (M).
- El trabajo remoto (50% o 100%) influirá positivamente en los salarios debido a la flexibilidad y demanda global.

Stakeholders principales

- **Recién graduados:** Interesados en conocer las oportunidades laborales y expectativas salariales.
- **Científicos de datos empleados:** Buscar prepararse mejor para un mejor empleo.
- **Empresas tecnológicas:** Comprender las tendencias globales actuales en empleo y salarios en diversas profesiones.
- **Instituciones educativas:** Interesadas en conocer qué carreras conducen a empleos mejor pagados.

Preguntas clave

- ¿Qué factores son los más determinantes en la variación salarial?
- ¿Cuál es la probabilidad de que un recién egresado sea contratado en un rol de ciencia de datos?
- ¿Cómo varían los salarios según el nivel de experiencia y la ubicación geográfica?
- ¿Influye el tamaño de la empresa en los salarios y en las probabilidades de contratación?



Fuentes de datos

- Datos salariales de las empresas.
- Datos de encuestas sobre la experiencia que tienen los empleados.
- Datos demográficos de las empresas y empleados.

Justificación del proyecto

El proyecto de análisis de salarios globales en el ámbito de la ciencia de datos es crucial en un momento donde la demanda de científicos de datos está en constante crecimiento y transformación. La ciencia de datos se ha convertido en una de las profesiones más demandadas a nivel global, impulsada por la necesidad de las empresas de adaptarse al mundo digital, extraer valor de los datos y tomar decisiones más inteligentes y basadas en evidencia, más recientemente con ayuda de modelos de predicción con ayuda de machine learning. Sin embargo, este crecimiento exponencial plantea importantes preguntas sobre las dinámicas salariales y las oportunidades de empleo, especialmente para recién graduados o profesionales que están considerando ingresar en este campo.

Este proyecto tiene el potencial de llenar un vacío informativo crítico al proporcionar un análisis exhaustivo que permita comprender cómo las variables como la experiencia, el país de residencia, el tipo de empleo, el tamaño de la empresa y la proporción de trabajo remoto afectan los salarios en este sector. No sólo se busca responder preguntas fundamentales como “¿Qué puede esperar ganar un recién egresado que entre al campo de la ciencia de datos?”, sino también proporcionar a las empresas datos valiosos sobre cómo estructurar sus políticas de compensación para ser más competitivas en un mercado globalizado.

En un contexto económico cada vez más competitivo y global, tanto los profesionales de la ciencia de datos como las empresas tecnológicas están interesados en tener acceso a información clara y confiable que les permita tomar decisiones informadas. Para los empleados actuales o potenciales, este



análisis podría significar una mejor comprensión de dónde posicionarse geográficamente y qué tipo de experiencia o habilidades necesitan desarrollar para mejorar sus ingresos. Para las empresas, significa acceder a datos que les permitan ajustar su estrategia de reclutamiento, sabiendo cuáles son las condiciones de mercado que atraerán o retendrán el mejor talento.

El impacto de este proyecto también se extiende más allá del análisis salarial. Las universidades y otras instituciones educativas pueden beneficiarse de los resultados para diseñar sus programas académicos en torno a las habilidades más demandadas y los roles mejor remunerados. Al conocer las tendencias de contratación y los requisitos salariales, estas instituciones pueden guiar mejor a sus estudiantes en términos de formación y expectativas profesionales, lo que mejoraría la empleabilidad de los egresados.

Por otro lado, el enfoque en los recién egresados añade un componente social y económico relevante. En un mundo post-pandemia, donde el trabajo remoto ha reconfigurado la dinámica laboral, es fundamental para los nuevos profesionales entender si pueden competir por los mismos empleos que aquellos que viven en grandes centros tecnológicos, pero desde cualquier parte del mundo. Este análisis también puede proporcionar una visión más justa e inclusiva de cómo la ubicación y el trabajo remoto están democratizando el acceso a trabajos altamente remunerados.

Este proyecto no solo tiene relevancia para quienes buscan obtener o proporcionar empleo en ciencia de datos, sino que también ofrece un marco de análisis que contribuye a entender y cerrar brechas salariales, mejorar las expectativas de empleo y ayudar a crear un entorno laboral más informado y equilibrado. Los resultados serán una fuente esencial para la toma de decisiones estratégicas en este sector en constante evolución.



¿Cuántos datos y que tipo son?

En total hay 47,579 datos repartidos en 11 columnas y se clasifican principalmente en dos:

- **Numéricos:** Salarios y proporción de trabajo remoto.
- **Categoricos:** Título del trabajo, tipo de empleo, nivel de experiencia, entre otros.

