

**Proyecto: Tendencias Del Mercado Laboral Global De IA y Perspectivas  
Salariales Para 2025**

**Alumno: Dana Domínguez Arce**

**Profesor: Jaime Alejandro Romero Sierra**

**Materia: Introducción A La Ciencia De Datos**

## **Introducción (Objetivo, justificación y datos):**

Este proyecto busca analizar Tendencias del mercado laboral global de IA y perspectivas salariales para 2025, con el objetivo principal de ofrecer información estratégica que ayude a empresas, reclutadores y profesionales de recursos humanos a mejorar sus procesos de contratación, retención y desarrollo de talento especializado. Para ello se estudiarán factores como el salario, las habilidades más demandadas en el sector, los niveles de experiencia, la industria, la ubicación geográfica y las modalidades de trabajo. Uno de los objetivos específicos es construir un modelo de referencia que permita ajustar políticas salariales y estrategias de contratación, con la meta de reducir en un 15 % el tiempo necesario para cubrir vacantes en roles relacionados con la IA. Al mismo tiempo, esta información brindará a los solicitantes y profesionales una visión clara de los requisitos y tendencias actuales del mercado.

El análisis permitirá identificar los títulos profesionales con mayor demanda a nivel mundial, así como examinar la relación entre salario, experiencia, tipo de empresa y sector industrial. También se evaluarán las habilidades técnicas más solicitadas y su impacto en la remuneración, además de observar las tendencias de trabajo remoto e híbrido dependiendo de la región y el tipo de puesto. Este estudio es especialmente relevante porque, en los últimos años, la rápida expansión de la IA y la creciente necesidad de análisis de datos han generado una fuerte competencia global por talento especializado. Sin embargo, muchas organizaciones carecen de información confiable y actualizada sobre cuáles habilidades priorizar, cuánto pagar o qué modalidad de trabajo ofrecer, lo que ocasiona vacantes difíciles de cubrir, alta rotación laboral y desajustes entre las capacidades disponibles y las necesidades reales del mercado.

Además, existen importantes desigualdades regionales en salarios, nivel educativo y acceso al trabajo remoto, lo que complica la creación de estrategias de contratación equitativas y efectivas. Para abordar estas problemáticas, el proyecto se basa en una base de datos de más de 15,000 ofertas de empleo recopiladas entre octubre de 2024 y julio de 2025, lo que garantiza un análisis actualizado y representativo. En este contexto, la IA se ha convertido en uno de los motores principales de transformación digital, impulsando la automatización, la toma de decisiones y la innovación tecnológica. Pero esta misma

evolución ha generado una brecha significativa entre la oferta y la demanda de talento capacitado, lo que dificulta un equilibrio en el mercado laboral.

Tanto empresas como gobiernos y profesionales necesitan información sistematizada y confiable que les permita tomar decisiones adecuadas. Las organizaciones requieren datos precisos para diseñar políticas salariales competitivas, definir habilidades prioritarias y mejorar sus procesos de reclutamiento, mientras que los profesionales necesitan conocer qué habilidades desarrollar y dónde se encuentran las mejores oportunidades laborales. También las instituciones educativas pueden beneficiarse de este análisis para adaptar sus programas a las exigencias actuales del mercado.

En conjunto, este proyecto proporciona una visión basada en datos reales sobre cómo se comportan los salarios, las habilidades más valoradas, los niveles de experiencia y las modalidades laborales en el terreno de la inteligencia artificial. Su aporte permite no solo comprender la situación actual, sino anticipar las tendencias que definirán el futuro inmediato del sector. Con ello, contribuye a cerrar la brecha de habilidades, mejorar la empleabilidad y fortalecer la competitividad global de las organizaciones, promoviendo un mercado laboral más equilibrado y alineado con las necesidades emergentes de la economía digital.

- **Datos:** Breve descripción de las fuentes de datos disponibles, por cada columna que significa.

Nombre de la columna	Descripción	Valores de ejemplo
id_del_trabajo	Identificador único para cada oferta de empleo	AI00001, AI00002
título profesional	Título/función anunciada	Científico investigador de IA, analista de datos
salario_usd	Salario anual convertido a USD	90376, 124355
moneda_salaria	Moneda en la que se ofrece el salario	USD, EUR, GBP

nivel de experiencia	Nivel de experiencia requerido (EN=Entrante, MI=Medio, SE=Senior, EX=Ejecutivo)	SE, EN, MI, EX
tipo_de_empleo	Tipo de contrato (CT=Contrato, FT=Tiempo completo, PT=Tiempo parcial, FL=Freelance)	Tiempo completo, tiempo parcial, tiempo ...
ubicación de la empresa	País donde tiene su sede la empresa	China, Canadá, Alemania
tamaño de la empresa	Tamaño de la empresa (S=Pequeña, M=Medianas, L=Grande)	Tallas pequeñas, medianas y grandes
residencia del empleado	País de residencia del empleado	China, Irlanda, Singapur
relación remota	% de trabajo remoto permitido (0=Presencial, 50=Híbrido, 100=Remoto)	0, 50, 100
habilidades requeridas	Habilidades técnicas clave requeridas (separadas por comas)	Python, SQL, Tableau, Docker
educación_requerida	Nivel mínimo de educación requerido	Licenciatura, Maestría, Doctorado, Asociado
años_de_experiencia	Se requieren años mínimos de experiencia	0, 4, 9, 15
industria	Sector industrial del puesto de trabajo	Automoción, medios de comunicación, salud
fecha de publicación	Fecha en que se publicó el trabajo (AAAA-MM-DD)	18 de octubre de 2024
fecha límite de solicitud	Última fecha para postular al empleo (AAAA-MM-DD)	7 de noviembre de 2024
longitud de la descripción del trabajo	Longitud de la descripción del puesto en caracteres	1076, 2340
puntuación de beneficios	Puntuación numérica (0-10) que refleja la calidad de los beneficios	5.9, 9.4

nombre de empresa	Nombre de la empresa contratante	Análisis inteligente, TechCorp Inc
-------------------	----------------------------------	---------------------------------------

## Metodología

### Proceso de limpieza de datos:

Para el proceso de limpieza se utilizó la base de datos alterada, generada a partir de la base principal original de 15,000 registros. Tras el proceso de modificación dirigido por el profesor, el conjunto de datos resultante contuvo 15,633 registros, tras haber sido ensuciada por un programa dejando más NaN y la palabra basura Auto% de manera aleatoria por el dataset.

Se comenzó con la carga y exploración básica inicial, con el fin de conocer un poco más nuestro dataset.

Para una mayor comodidad el código estará marcado por el color azul.

#### 1. Carga y exploración

- El notebook comienza cargando el fichero sucio con `pd.read_csv('df_sucio_eseya.csv')`.
- Luego se realiza una exploración rápida con `df.describe()` y `df.isnull().sum()` para conocer dimensiones, estadísticas básicas por columna y la cantidad de valores nulos por columna. Esto permitió identificar las columnas numéricas y categóricas con datos faltantes, revisar rangos y desviaciones.

#### 2. Detección y eliminación de duplicados

- Se verificó la presencia de filas duplicadas usando `df2.duplicated()` y `df2.duplicated().sum()`.
- Las filas duplicadas fueron eliminadas con `df3 = df2.drop_duplicates()`. Con esto se aseguró que cada oferta/registro sea único antes de proceder al resto de limpiezas.

#### 3. Limpieza de errores textuales y normalización

- Se aplicaron reemplazos puntuales con `replace(..., regex=True)` para corregir patrones erróneos en columnas (por ejemplo `df3.replace(to_replace='Auto%#', value=0,`

`regex=True, inplace=True)).` Esto indica limpieza de tokens o caracteres extraños que impedían interpretar correctamente valores numéricos o categorías.

- También hay mapeos de valores categóricos para unificar nomenclaturas, por ejemplo,  
`df4['titulo_trabajador'] = df4['titulo_trabajador'].replace(traducion_titulos)`, donde se reemplazan variantes de títulos por una versión estandarizada.

#### 4. Manejo de valores ausentes (missing values)

- **Numéricas:** se llenaron con la media (o promedio) redondeada de la columna:  
`df3[col] = df3[col].fillna(promedio_redondeado).` Esto evita perder registros por dropna y mantiene la columna numérica utilizable en análisis posteriores.
- **Categóricas/texto:** se llenaron con una etiqueta fija (por ejemplo 'sin dato') para marcar explícitamente la ausencia de información: `df3[k] = df3[k].fillna('sin dato').`
  - a. El uso de `isnull().sum()` antes y después permite comprobar que los rellenos eliminaron (o redujeron) los nulos detectados.

#### 5. Transformaciones de columnas complejas

- Para columnas con listas o cadenas separadas por comas (por ejemplo habilidades), se transformaron a listas con `.apply(lambda x: x.split(',') if isinstance(x, str) else [])`. Esto facilita análisis por habilidad.
- Se revisaron y ajustaron tipos de datos cuando fue necesario (por ejemplo conversiones implícitas con astype en varias partes del notebook, cuando se requiere que una columna sea numérica o fecha).

#### 6. Guardado del conjunto limpio

- El dataset final limpio se exportó a CSV con `df4.to_csv ("Base_limpia_IA.csv", index=False)`. Así se preserva la versión procesada para análisis posteriores (modelado, visualizaciones, etc.).

## Análisis Exploratorio de Datos (EDA):

- **Descripción general de los datos**
- **Visión General**

El data set determinado como df4 cuenta con 15468 registros y 19 variables, es decir 19 columnas.

- **Tipos de Variables:**

Se identificaron los tipos de variables, encontrando únicamente categóricas y numéricas, remplazando el tipo “object” por categórica y “float64” por numéricas.

	<b>Columna</b>	<b>Tipo de Dato</b>
0	id_trabajo	Categoricas
1	titulo_trabajador	Categoricas
2	salary_usd	Numericas
3	Moneda_salario	Categoricas
4	Nivel_experiencia	Categóricas
5	tipo_empleo	Categóricas
6	Ubicacion_empresa	Categóricas
7	tamaño_empresa	Categóricas
8	residencia_empleado	Categóricas
9	remote_ratio	numéricas
10	habilidades_empleados	Categóricas
11	educacion_requerida	Categóricas
12	years_experience	numéricas

13	industria	Categóricas
14	fecha_publicacion	Categóricas
15	fecha_límite	Categóricas
16	job_description_length	Numéricas
17	benefits_score	Numéricas
18	nombre_empresa	Categóricas

- **Resumen estadístico**

El resumen estadístico obtenido mediante `df.describe()` permitió observar medias, medianas y rangos de valores, brindando una primera comprensión de la distribución interna de cada variable.

Estadísticas descriptivas para variables numéricas:

Variable	Media	Mediana	Desv. Estándar	Mínimo	Máximo
salario_usd	112960. 55	100483. 50	60670. 87	0.0	39909 5.0
relación remota	49.40	50.00	39,76	0.0	100.0
años_de_experiencia	6.12	5.00	5.47	0.0	19.0
longitud de la descripción del trabajo	1501.47	1501.47	567.16	500.0	2499.0

<b>puntuación de beneficios</b>	7.50	7.50	1.43	5.0	10.0
---------------------------------	------	------	------	-----	------

Estadísticas descriptivas para variables categóricas:

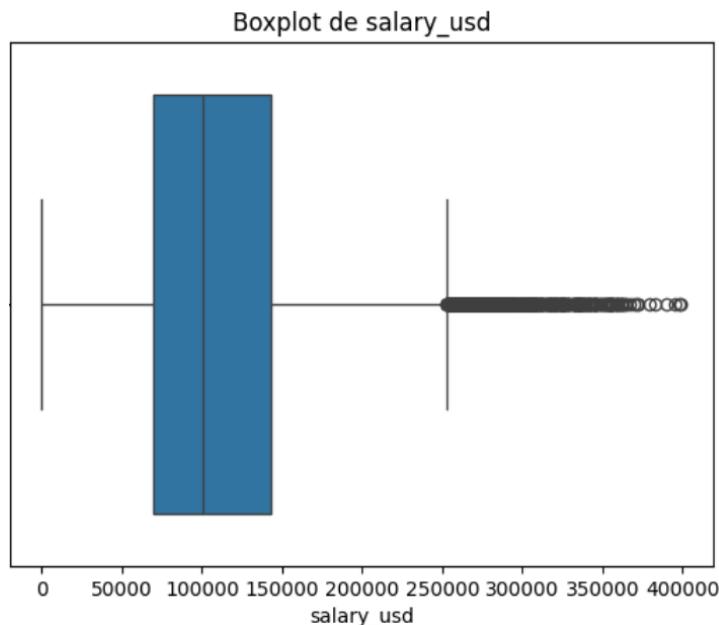
Título del Trabajador	Frecuencia
<b>Investigador en Aprendizaje Automático</b>	786
<b>Ingeniero de Software en IA</b>	767
<b>Ingeniero de Sistemas Autónomos</b>	764
<b>Ingeniero de Aprendizaje Automático</b>	761
<b>Arquitecto de IA</b>	759
<b>Ingeniero en Robótica</b>	757
<b>Jefe de IA</b>	752

## 2. Visualización y Distribución de Variables Individuales

### \* Variables Numéricas

Se analizaron como variables principales: salary\_usd, years\_experiencey benefits\_score. utilizando **histogramas** (para ver la forma de la distribución) y **boxplot** (para identificar asimetrías y valores atípicos).

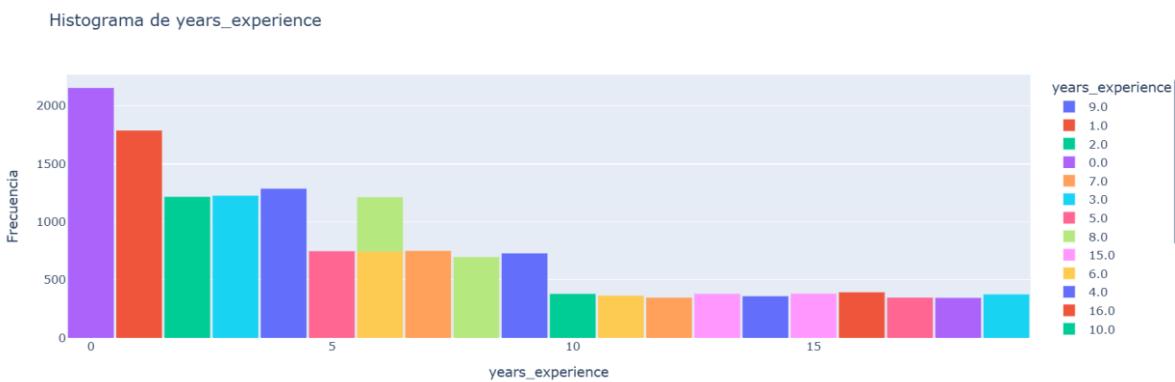
## 1. Salario\_USD (Salario en USD)

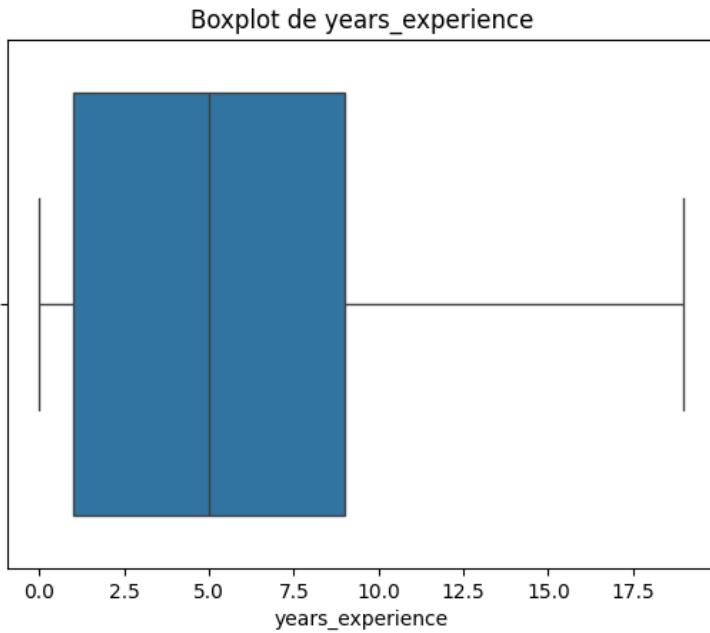


La distribución del salario está sesgada hacia la derecha indicando así un sesgo positivo. La mayor concentración de salarios se encuentra entre \$50,000\$y\$150,000 Dólares estadounidenses. La frecuencia máxima se observa en la clase salarial alrededor de \$100,000USD, mientras que en su boxplot se observa que la mediana esta entre \$100,000 y \$150,000 USD. Además se destaca la presencia de outliers.

- **Años\_Experiencia (Años de Experiencia)**

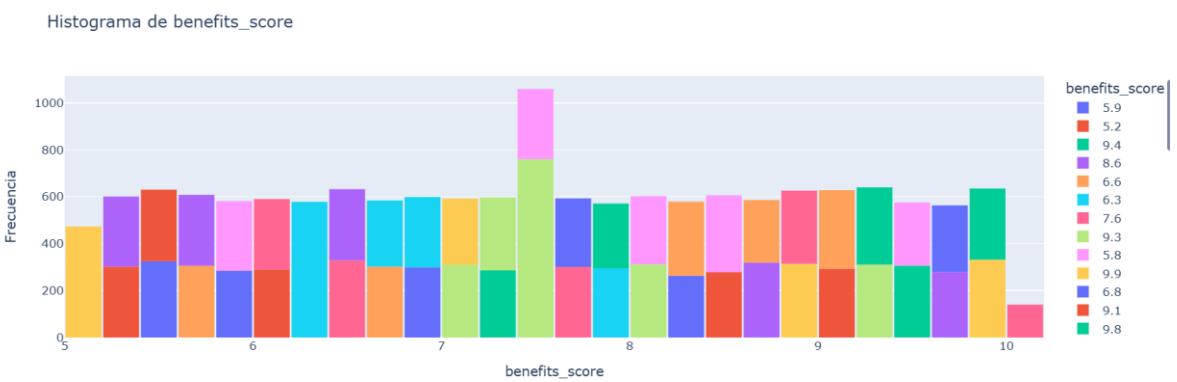
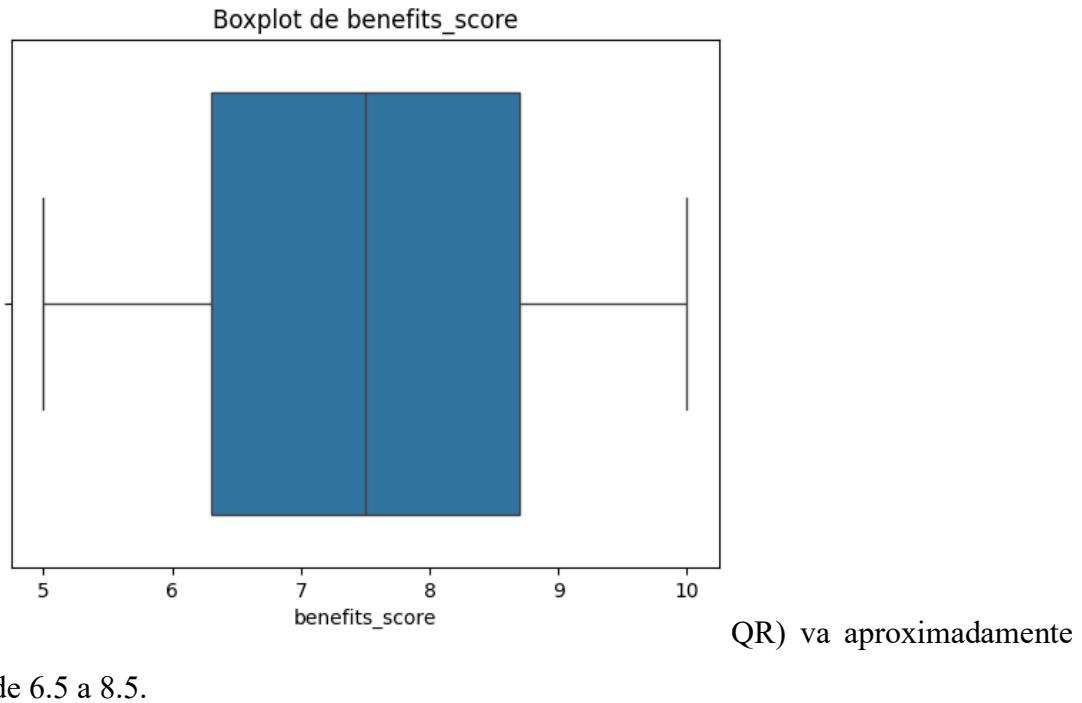
La mayor frecuencia se encuentra en el rango de 0 a 5 años de experiencia. El valor más frecuente parece ser 0 años con una frecuencia de más de 2000, seguida por 1 año.





- Se observan **valores atípicos** más allá de 17 años.
- **Benefits\_Score (Puntuación de Beneficios)**

La distribución es más uniforme que las anteriores, mostrando frecuencias relativamente similares para la mayoría de las evaluaciones entre 5 y 10. Hay un pico notable alrededor de 7.5 superando los 1,000 en frecuencia, en el boxplot se muestra la mediana ligeramente debajo del 7.5. El rango intercuartílico (I



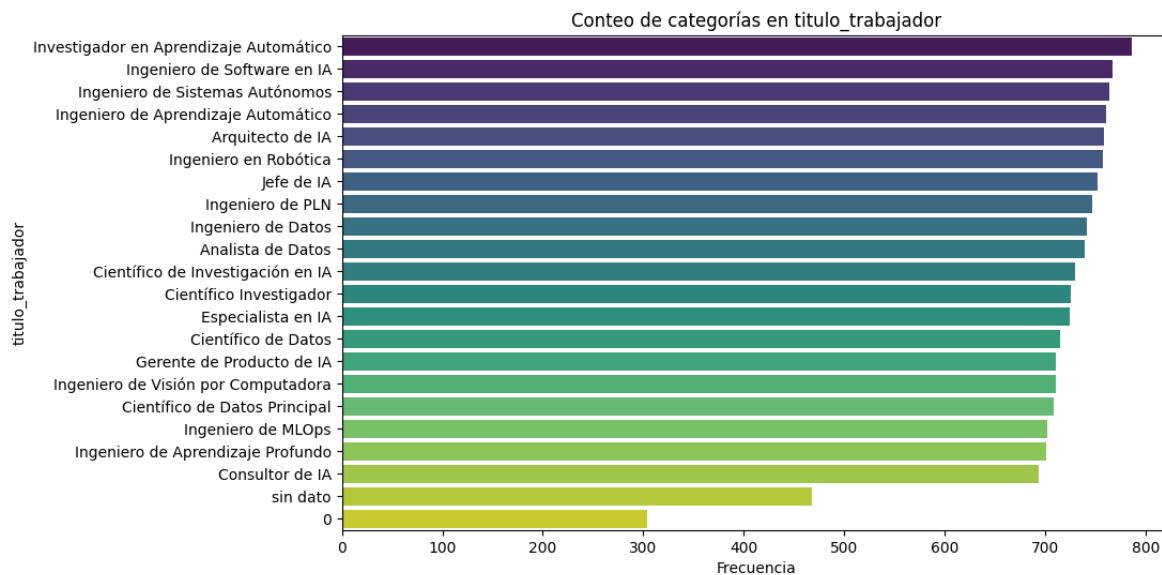
- **Remote\_Ratio (Tasa de Trabajo Remoto)**

Esta variable es peculiar porque tiene valores muy concentrados. La distribución es **bimodal/trimodal**, mostrando picos muy altos en(Trabajo Presencial),(Híbrido) y(Remoto).

## Variables Categóricas

- **Título del trabajador**

La categoría más frecuente es **Investigador en Aprendizaje Automático**, con una frecuencia cercana a 800. Le siguen de cerca **Ingeniero de Software en IA**, **Ingeniero de Sistemas Autónomos** e **Ingeniero de Aprendizaje Automático**.

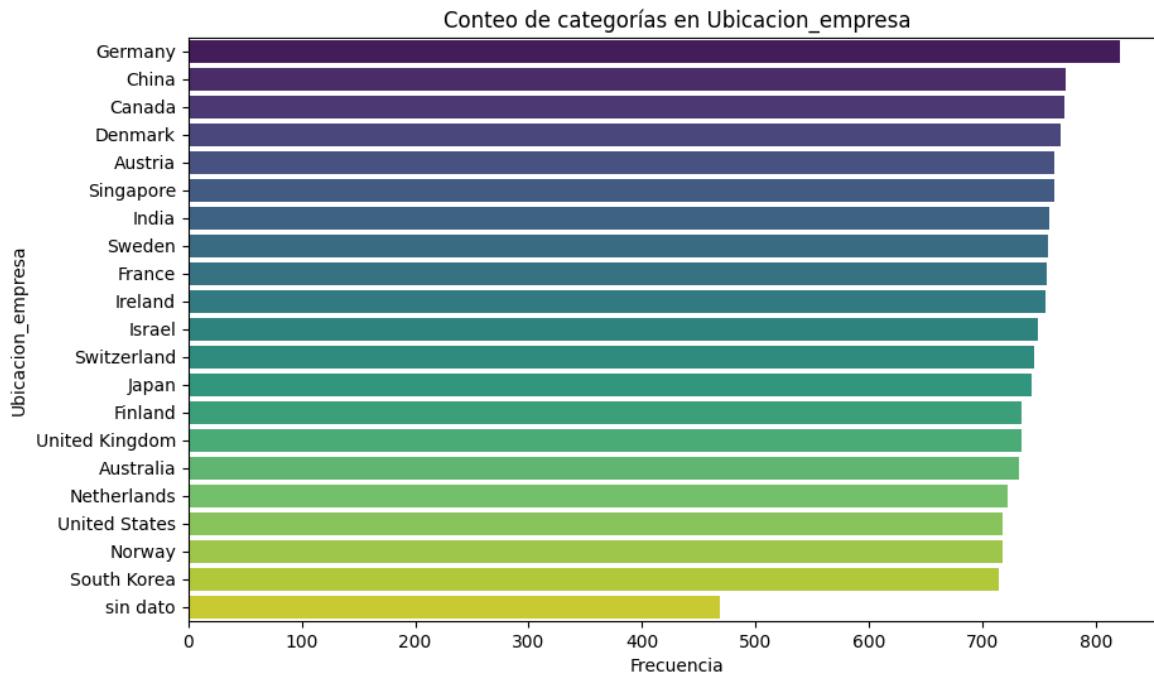


### - Moneda del Salario

La moneda más común con gran diferencia es el **USD**, con una frecuencia cercana a 12 000. La segunda más frecuente es el **EUR**, con una frecuencia de poco más de 2000.

### - Ubicación de la Empresa (*Ubicacion\_empresa*)

La frecuencia de la ubicación de la empresa es alta para muchos países, siendo **Alemania**, **China**, **Canadá** y **Dinamarca** los que tienen las frecuencias más altas (alrededor de 750-800).



- **Residencia del Empleado ( residencia\_empleado)**

Los países con la mayor frecuencia de residencia son **Suecia, Dinamarca, India y Austria**.

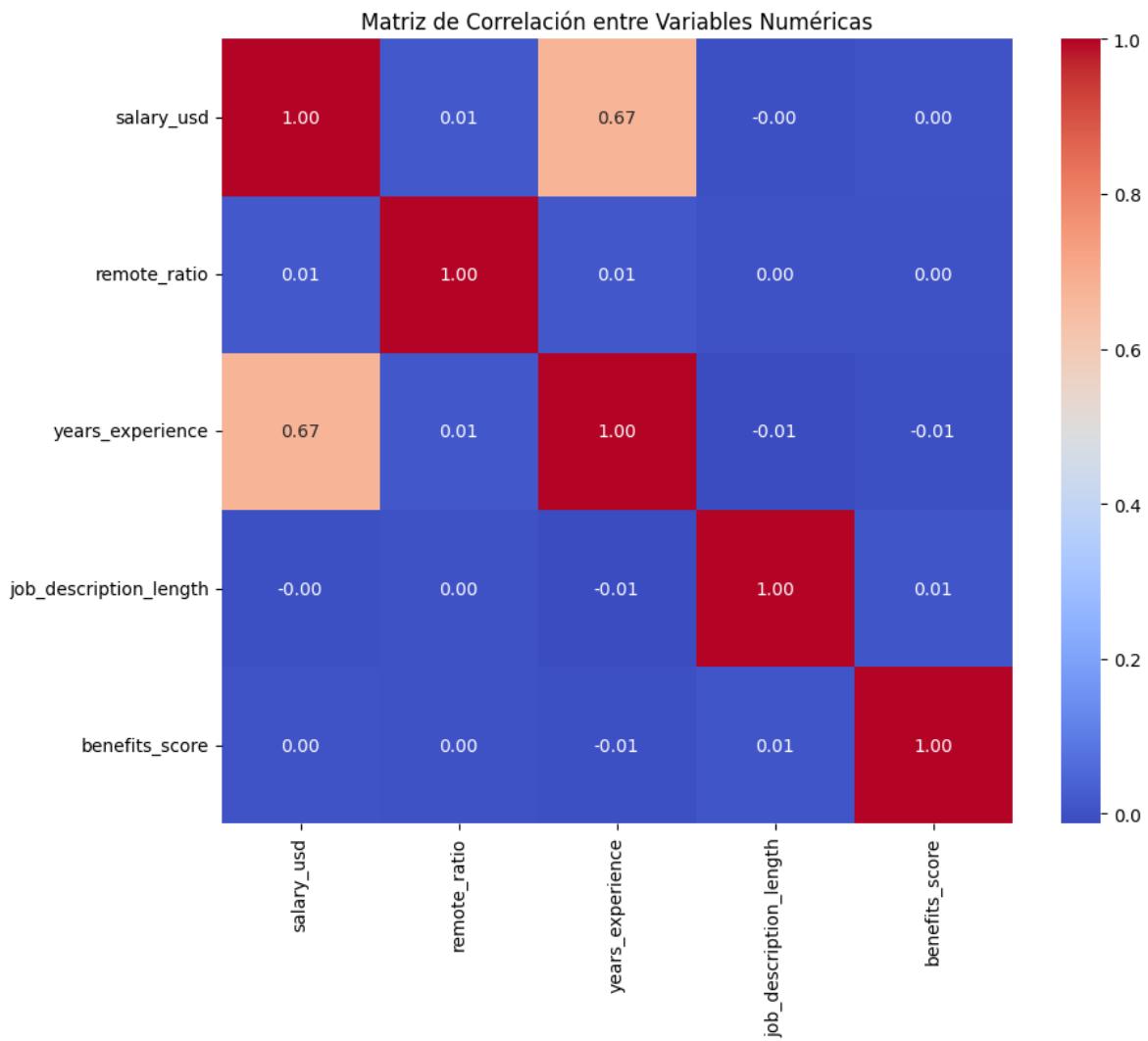
- **Tamaño de la Empresa ( tamaño\_empresa)**

Las categorías de tamaño de empresa **L (Grande) , M (Mediana) y S (Pequeña)** tienen frecuencias muy altas y muy similares, todas alrededor de 5000.

Aquí tienes \*\*todo tu análisis convertido en prosa formal\*\*, tal como debe aparecer en la \*\*descripción de la metodología de tu EDA\*\*, redactado de manera clara, académica y coherente con los lineamientos del PDF.

### 3. Correlación entre Variables

Dentro del EDA se elaboró una matriz de correlación con el objetivo de identificar relaciones lineales entre las variables numéricas del dataset. Esta herramienta pudo permitir observar de manera visual la intensidad de las relaciones mediante un heatmap, donde los valores mas cerca de 1.00 indican una correlación positiva mayor o fuerte, mientras que aquellos cercanos a -1.00 representan una correlación negativa significativa. Por otro lado, los valores próximos a 0 nos revelan la ausencia de relación lineal entre las variables.



Desde lo general podemos observar que las variables numéricas presentan correlaciones muy débiles o prácticamente nulas. No obstante, la variable salary\_usd mantiene una correlación positiva moderada de 0.67 con años de experiencia. Esta relación indica que, conforme aumentan los años de experiencia laboral solicitados, también tiende a aumentar el salario ofrecido en las vacantes. Este resultado es consistente con el comportamiento habitual del mercado laboral en áreas tecnológicas.

Por otro lado, variables como remote\_ratio, job\_description\_length y benefits\_score mostraron correlaciones cercanas a cero tanto entre ellas como con el salario y los años de experiencia. Esto sugiere que no existe una relación lineal entre ellos.

#### 4. . Análisis de Valores Atípicos (Outlier)

Variable	Número de valores atípicos detectados
salario_usd	<b>551</b>
relación remota	0
años_de_experiencia	0
longitud de la descripción del trabajo	0
puntuación de beneficios	0
nivel_exp_num	0
número de empresa de tamaño	0

El análisis por el proceso IQR ayudo a detectar 551 valores atípicos en la variable objetivo salary\_usd. Estos outliers representan salarios excepcionalmente altos. Se decidió mantener estos registros en el conjunto de datos para no perder información valiosa sobre el potencial de ingresos en el mercado laboral al meterlos al modelo de machine learning.

#### 5 . Análisis de Valores faltantes

El análisis de valores nulos o faltantes en la base de datos revela que la mayoría de las variables están **completas**, con un 0% de datos faltantes. Sin embargo, hay dos variables que si presentan muchos valores faltantes significativos.

	Valores Faltantes	Porcentaje (%)
nivel_exp_num	773	4.997414
tamano_empresa_num	468	3.025601

La estrategia de imputación utilizada fue el remplazo de valores faltantes por el promedio, dando esto como resultado:

- Para la variable salary\_usd los valores nulos son remplazados por 112960.55148694079
- Para la variable remote\_ratio los valores nulos son remplazados por 49.397944142746326
- Para la variable years\_experience los valores nulos son remplazados por 6.117995862425653
- Para la variable job\_description\_length los valores nulos son remplazados por 1501.4730727954488
- Para la variable benefits\_score los valores nulos son remplazados por 7.498151021463666
- Para la variable nivel\_exp\_num los valores nulos son reemplazados por 2.506362708404219
- Para la variable tamano\_empresa\_num los valores nulos son reemplazados por 2.997133333333333

## **6. Relación entre Variables Categóricas y Numéricas**

Se utilizaron boxplots para poder analizar pares de variables que aporten para seguir con las hipótesis iniciales que se generaron al inicio del proyecto, las cuales fueron:

### **Hipótesis 1:**

Aquellos profesionales con un nivel de experiencia mayor como Senior (SE) y Ejecutivo (EX) presentan salarios significativamente más altos que los de nivel medio (MI) o de entrada (EN, Principiante), independientemente de la ubicación geográfica.

### **Hipótesis 2:**

Los roles de inteligencia artificial que requieren habilidades avanzadas en programación (como Python, TensorFlow o PyTorch) y gestión de datos (SQL, AWS, MLOps) presentan salarios promedio superiores a los puestos con habilidades con menos técnicas.

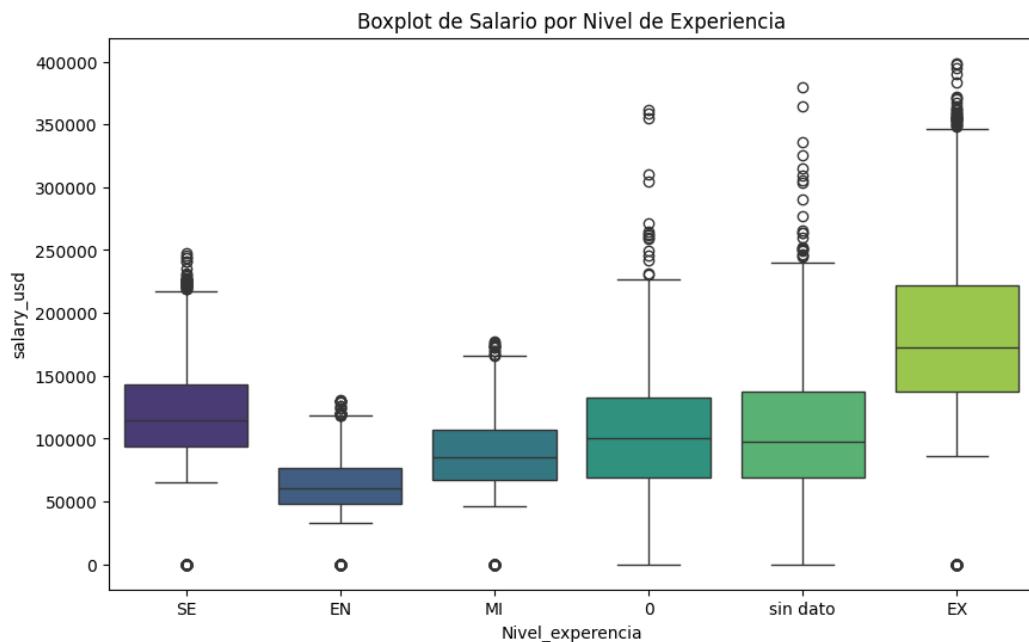
### **Hipótesis 3:**

Las ofertas laborales de empresas que permiten una mayor proporción de trabajo remoto son más susceptibles a ofrecer salarios ligeramente más bajos, sin embargo, atraen un más candidato.

Como variable principal el salario en USD (`salary\_usd`) varía según diferentes categorías, permitiendo observar distribuciones, medianas, dispersión y valores atípicos.

### Nivel de Experiencia vs. Salario

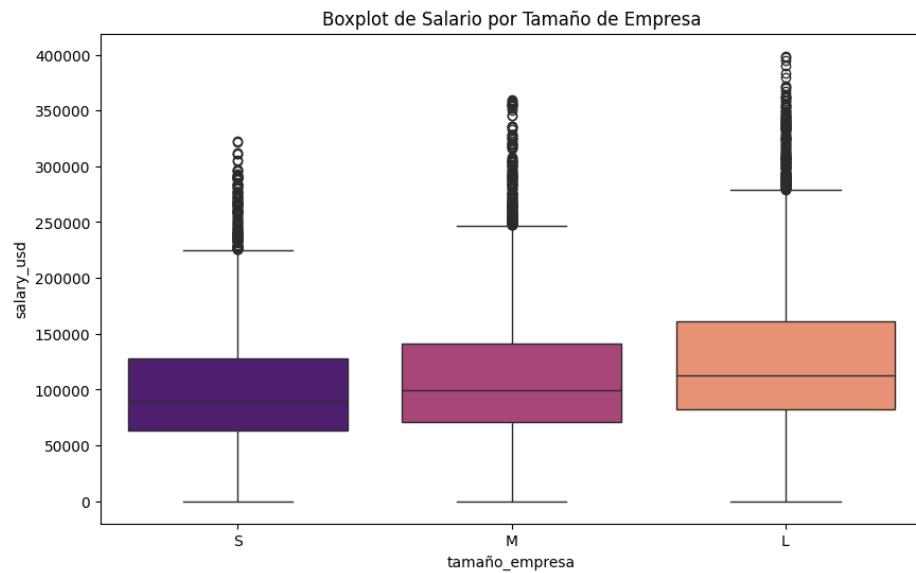
Se identifica una relación directamente proporcional entre experiencia y salario. Los roles ejecutivos (EX) presentan la mediana salarial más elevada, aproximadamente \$175,000 USD, junto con la mayor dispersión salarial, lo que refleja una amplia variabilidad en los ingresos de alta jerarquía. Los niveles senior (SE) muestran medianas superiores a los intermedios (MI), mientras que los principiantes (EN) registran los valores más bajos. Esto confirma que la experiencia es uno de los factores más determinantes en la estructura salarial.



### Tamaño de Empresa vs. Salario

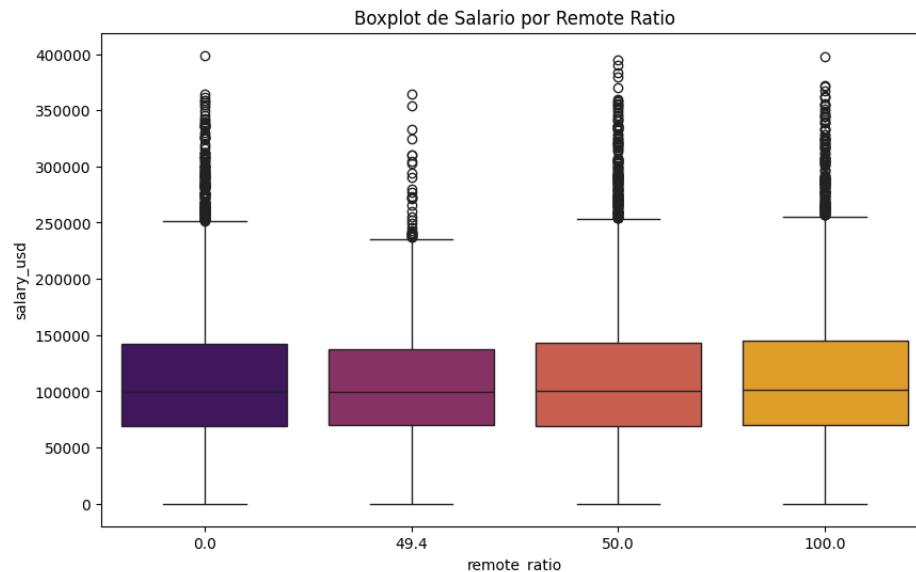
Las empresas grandes (L) ofrecen una mediana salarial ligeramente superior en comparación con las medianas (M) y pequeñas (S), aunque las diferencias entre categorías no son marcadas. Las tres presentan un rango intercuartílico similar, lo que sugiere que, si bien las organizaciones más grandes tienden a pagar mejor, el tamaño

empresarial no define de forma contundente el nivel salarial en este dataset.



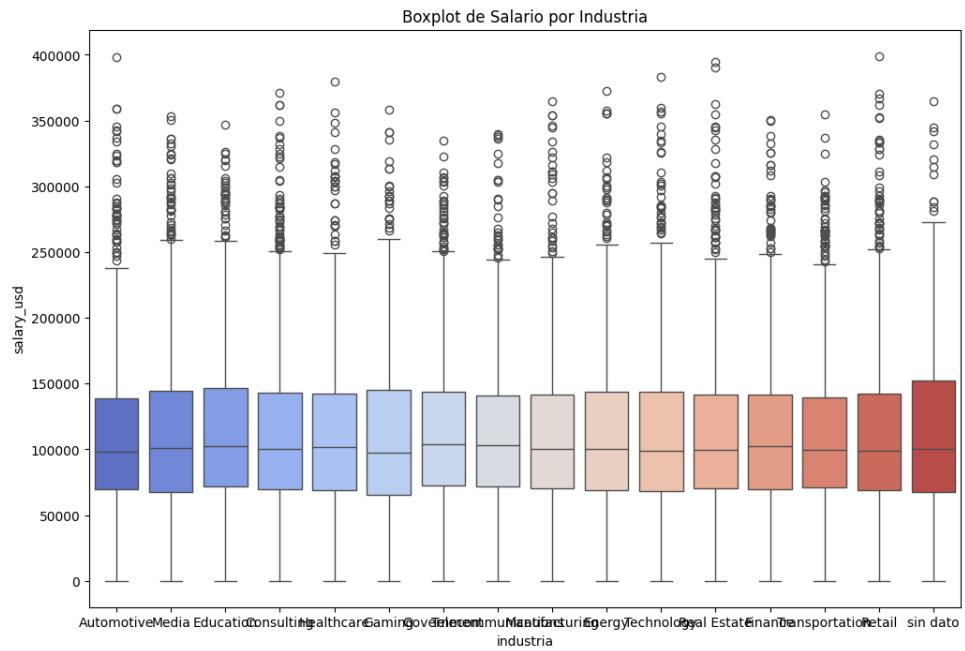
### Modalidad de Trabajo vs. Salario

No se observan diferencias significativas en los salarios entre modalidades presenciales, híbridas o remotas. Las medianas para los tres tipos son prácticamente idénticas, ubicándose ligeramente por encima de los \$100,000 USD, con dispersiones y cantidades de valores atípicos similares. Esto indica que, en este contexto, la modalidad de trabajo no influye de manera relevante en el nivel salarial.



### Industria vs. Salario

La mayoría de las industrias analizadas presentan medianas y rangos salariales muy similares, sin que ninguna se destaque notablemente por salarios significativamente altos o bajos. Esto sugiere que, dentro de este conjunto de datos, el sector industrial no es un factor diferenciador clave en la determinación del salario.



## Habilidades vs. Salario

Esta grafica nos sugiere que las habilidades más fundamentales que se relaciona con el nivel del salario son **Python** y **SQL**. Sin embargo **GCP** está asociado con el valor más bajo. Esta jerarquía de valor es una variable importante a considerar para el modelado de machine learning.

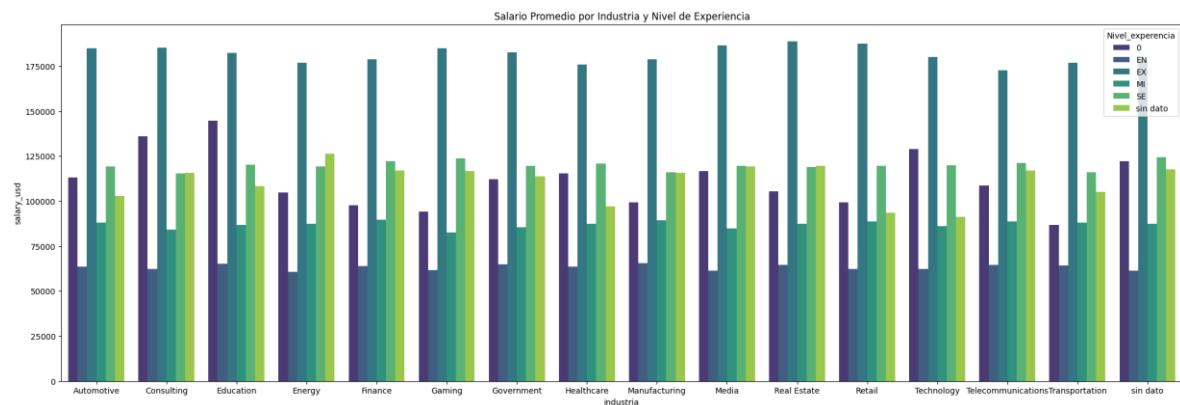


## 7. Observaciones y Hallazgos Importantes

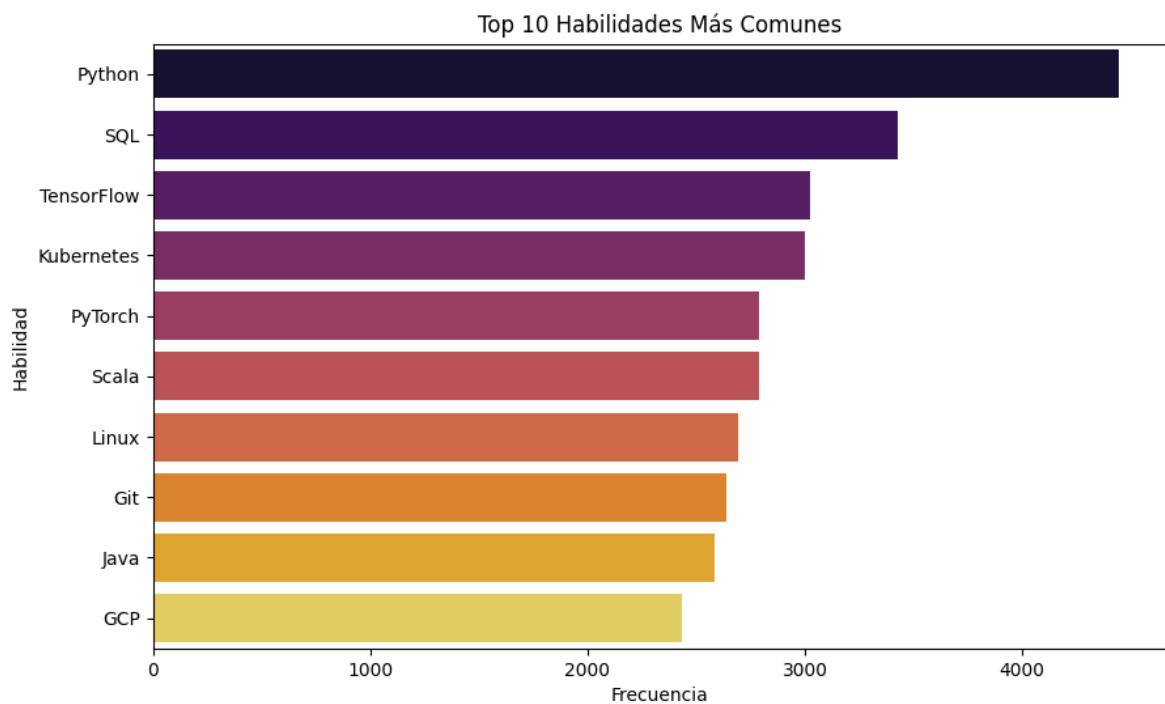
**Podemos decir que:**

El nivel de experiencia emerge como el factor más influyente en el salario, con una progresión clara y esperada según la jerarquía del puesto. Por el contrario, el tamaño de la empresa, la industria y la modalidad de trabajo muestran un impacto limitado, con distribuciones salariales notablemente uniformes entre categorías. Estos hallazgos permiten categorizar como más importante a las variables relevantes para futuros modelos predictivos mas desarrollados y estrategias de compensación.

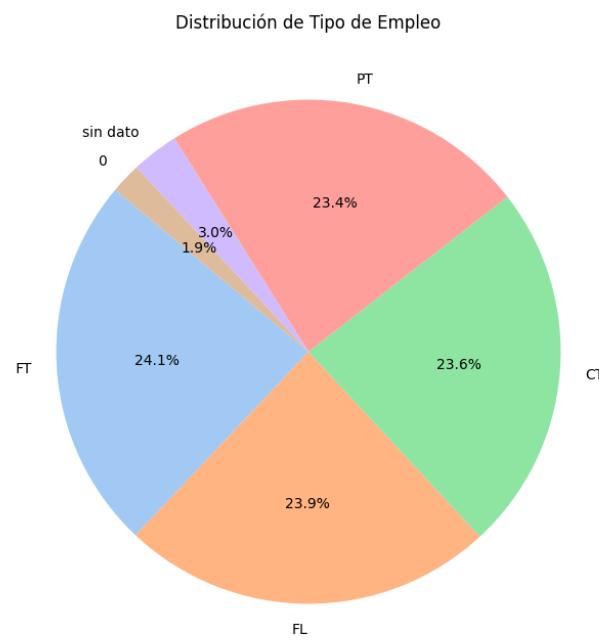
**Relación Salario-Experiencia:** Se encontró con la existencia de un patrón escalonado y robusto que relaciona el nivel de experiencia con el salario promedio, lo que es lógico, pero esencial para la validación de políticas internas.



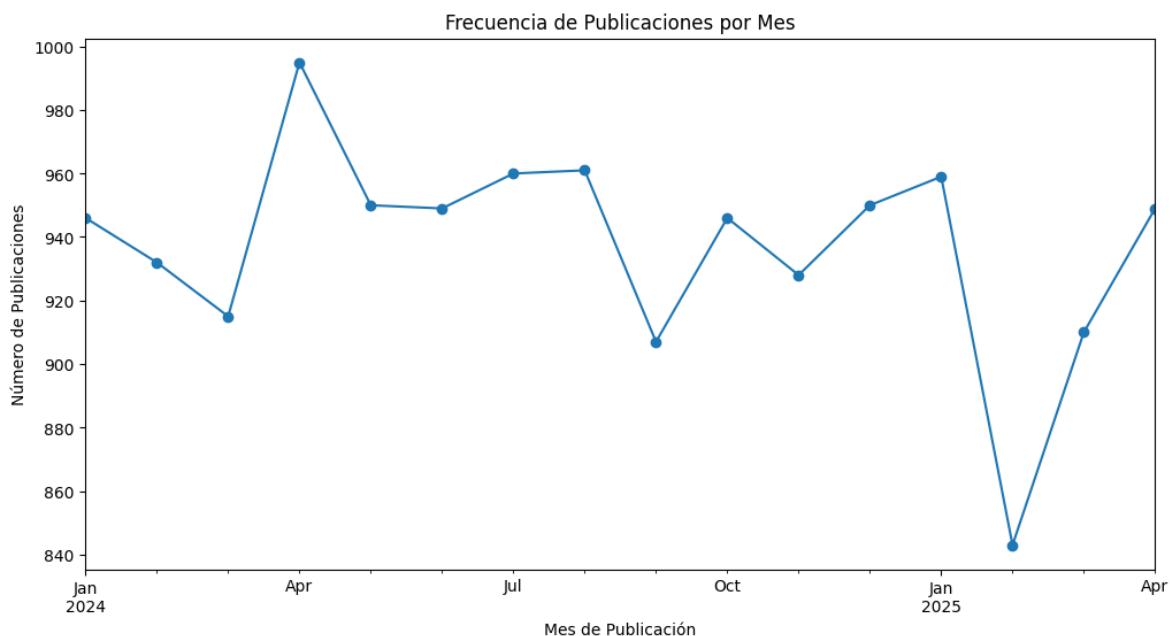
**Demandado por Habilidades:** Las habilidades más solicitadas se concentran fuertemente en Python y SQL , seguidas de frameworks de Deep Learning como TensorFlow. Esto podría ser de gran utilidad ya que las empresas lo podrían ocupar para enfocar programas de capacitación .



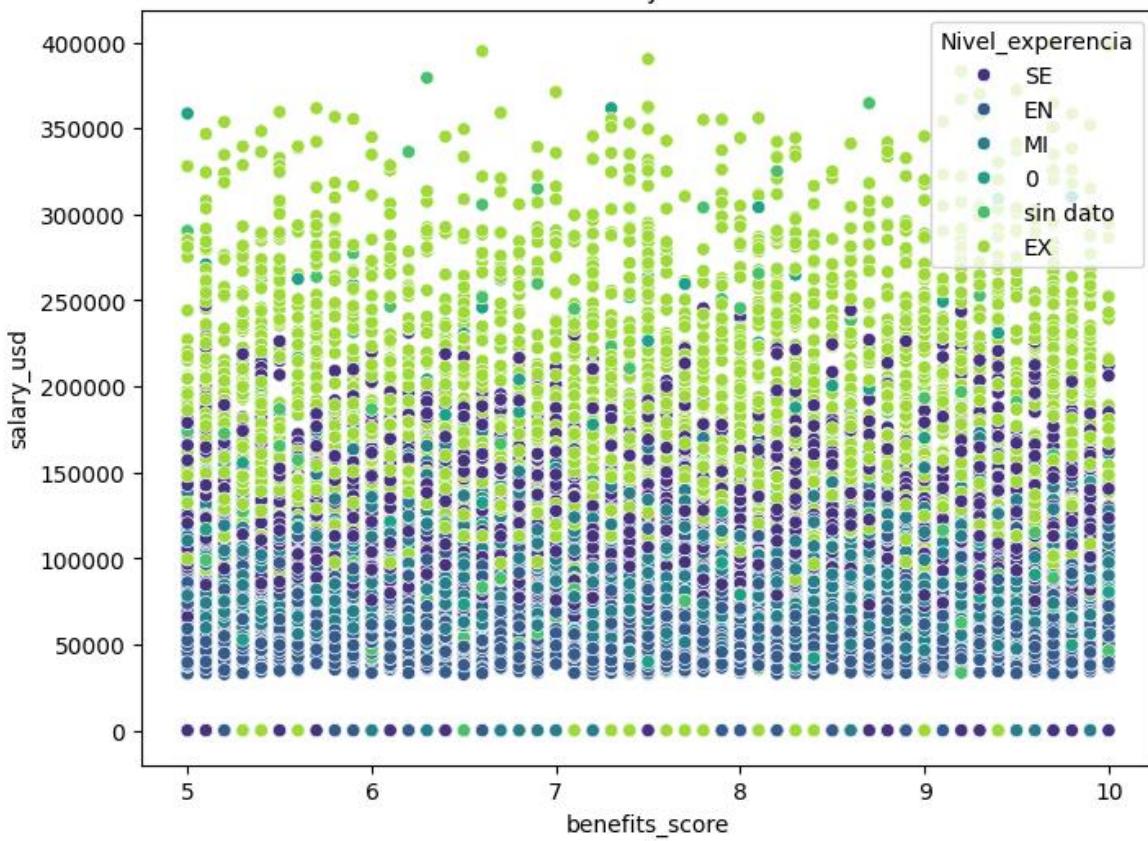
**Tipo de Empleo Dominante:** La gran mayoría de las ofertas corresponden a la modalidad de **Tiempo Completo (FT)**, lo que indica que, a pesar de la tendencia remota, la estructura laboral sigue priorizando la vinculación a tiempo completo.



Frecuencia de publicaciones por mes:



Salario vs Puntaje de Beneficios



## **8. Modelo Machine Learning**

Dado que se identifican **relaciones no lineales** claras entre las variables categóricas (como el nivel de experiencia y la industria) y la variable objetivo ( salario) se decidió utilizar un modelo robusto capaz de manejar estas complejidades. El preprocesamiento se enfocó en la codificación de las variables categóricas para que el modelo pudiera interpretarlas adecuadamente. Para cumplir con el objetivo de establecer una referencia salarial , se implementó un modelo de regresión.

### **1. Descripción del Modelo**

Se implementó un Random Forest Regressor el cual es un modelo de aprendizaje supervisado diseñado para resolver problemas de regresión (predicción de valores numéricos continuos, como el salario). Este modelo se construye a partir de múltiples árboles de decisión que trabajan en conjunto para mejorar la precisión y evitar el sobreajuste.

### **2. Justificación**

La elección de Random Forest Regresor se basó en tres puntos principales:

1. Tipo de Variable Objetivo: Necesitábamos predecir el salario (numérico) , por lo que la regresión era obligatoria.
2. Robustez y Rendimiento: El Random Forest es conocido por ofrecer un mejor rendimiento y mayor resistencia al ruido y valores atípicos en comparación con modelos más simples como la Regresión Lineal, que tiene dificultades con las complejas interacciones de mercado.
3. Tamaño del conjunto de datos: Con más de 15.000 registros, el Random Forest es eficiente y maneja bien los conjuntos de datos grandes.

### **3. Implementación y Entrenamiento**

El conjunto de datos se dividió siguiendo el estándar el 80% de los datos para el entrenamiento y el 20% para la evaluación.

### **4. Resultados y Evaluación**

Se utilizaron métricas estándar de regresión para evaluar el desempeño del modelo sobre el conjunto de prueba, obteniendo los siguientes resultados:

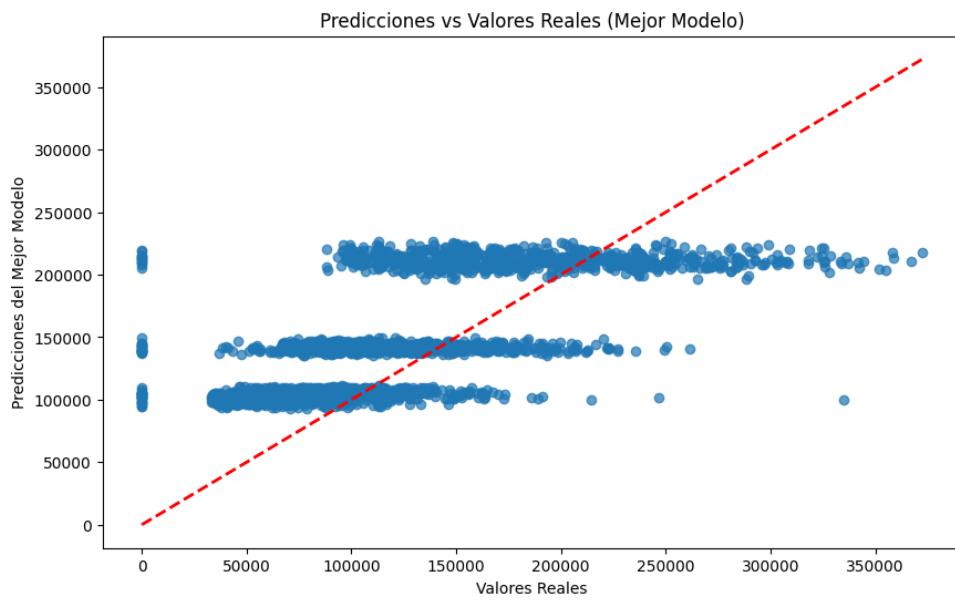
- **Error absoluto medio (EAM): 41493.14401010826**
- **Error cuadrático medio (ECM): 2623313963.8841515**
- **R<sup>2</sup>: 0.2954180755388003**

### Interpretación:

El modelo arrojó un R<sup>2</sup> de 0.2954 , lo que indica que las variables utilizadas solo logran explicar cerca del 29.5% de la variabilidad de los salarios en el mercado. El Error Absoluto Medio (EAM) de \$41,493 USD refuerza esta conclusión, sugiriendo que la predicción promedio se desvía considerablemente del valor real. Esto implica que el mercado salarial de IA está influenciado por factores no incluidos en el conjunto de datos (como la ciudad específica, el desempeño individual o el nombre exacto de la empresa), limitando la precisión del modelo en la predicción de un punto exacto.

## 5. Visualizaciones de resultados

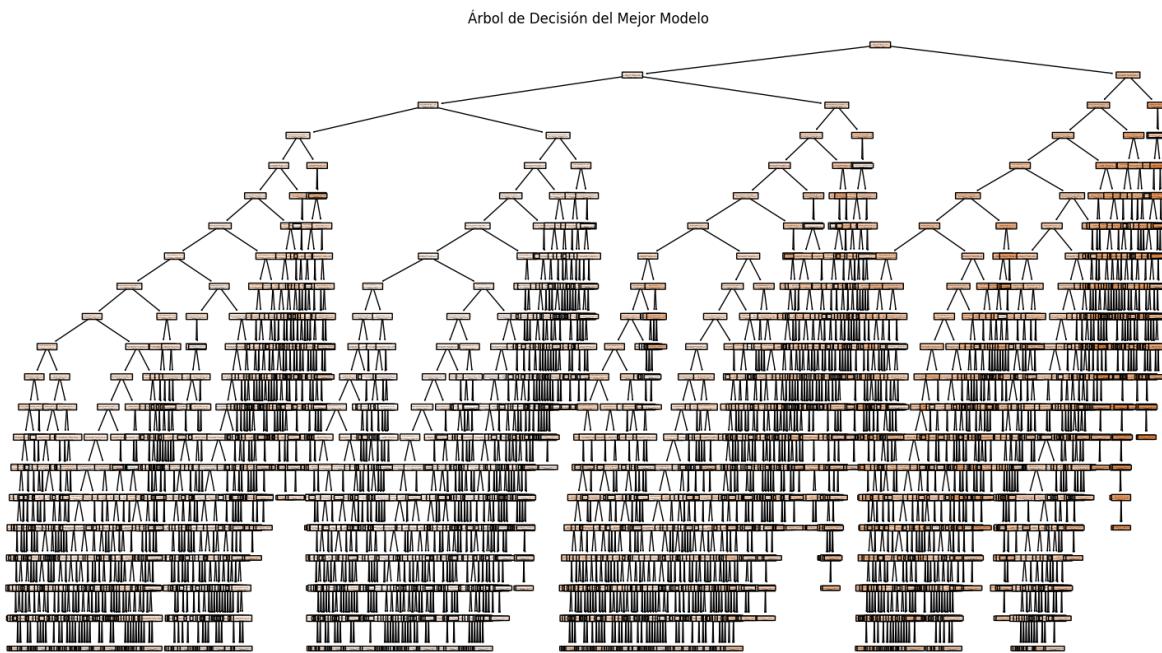
La visualización de las predicciones frente a los valores reales es clave para entender el



desempeño.

La dispersión muestra que, aunque hay una **tendencia lineal positiva** (el modelo predice salarios más altos para salarios reales más altos), los puntos están muy dispersos alrededor

de la línea ideal. Esto visualiza el **bajo R** y demuestra que el modelo es mejor para establecer rangos salariales amplios que para dar un valor exacto.



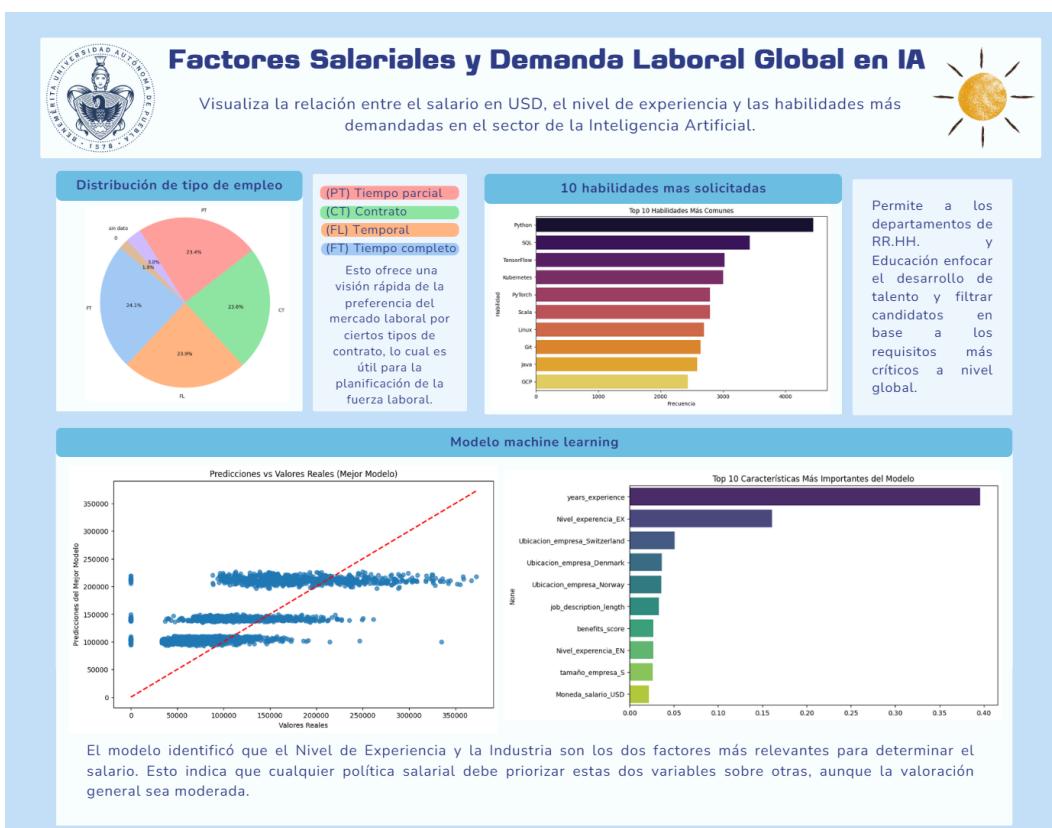
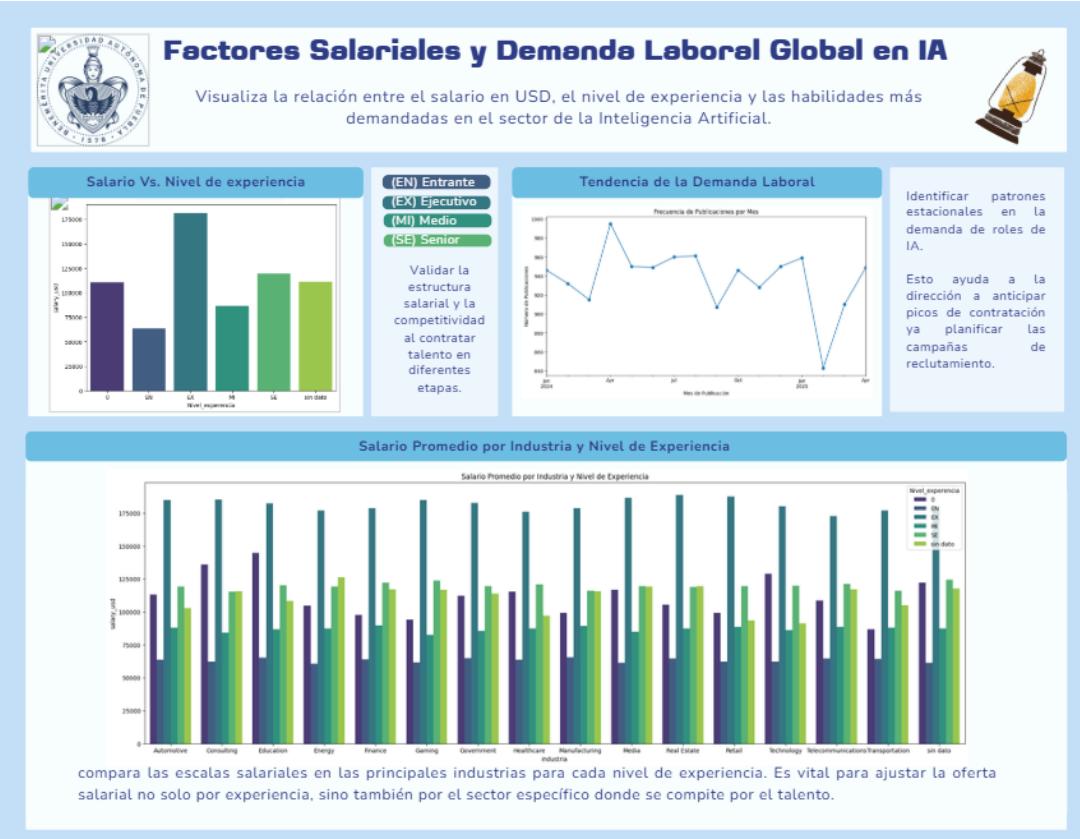
## 6. Conclusión del Modelo

El modelo **Random Forest Regressor** confirma la gran complejidad del mercado salarial de IA. Si bien la precisión no fue alta ( $R^2 = 0.2954$ ), las variables más influyentes, como el **Nivel de Experiencia** y la **Puntuación de Beneficios**, fueron debidamente identificadas.

Posibles mejoras para el futuro podrían ser

1. **Añadir variables geográficas** (país o ciudad) y de **competitividad de la empresa**.
2. **Optimizar los hiperparámetros** para reducir el margen de error
3. O Probar otros modelos más complejos para regresión

## 9.Capturas y Explicación del Dashboard



## **2. Uso y Beneficios del Panel de Control y conclusiones**

Este Dashboard de Factores Salariales y Demanda Laboral Global en IA está estratégicamente diseñado para ser la herramienta principal de Gerentes de Recursos Humanos, Directivos de Tecnología y Jefes de Adquisición de Talento que operan en el competitivo sector de la Inteligencia Artificial. Su objetivo no es solo presentar datos, sino transformar la información en una base para la toma de decisiones.

### **¿Qué decisiones pueden apoyar al usuario gracias al tablero?**

El tablero permite a los usuarios dejar de adivinar y comenzar a actuar con datos . La capacidad de visualización apoya decisiones críticas en tres áreas:

1. Ajuste de Política Salarial: La gráfica de Salario vs. Nivel de Experiencia y el cruce de Salario por Industria facilitan a los gerentes de RR.HH. validar si sus estructuras salariales son competitivas respecto al promedio global. Si el salario ofrecido para un perfil *Senior* está por debajo de la tendencia mostrada, el directivo puede decidir aprobar un ajuste presupuestario para garantizar la retención y atracción de ese talento clave.
2. Estrategia de Desarrollo de Talento: Al identificar el Top 10 de Habilidades Requeridas , el área de capacitación puede decidir enfocar la inversión en *bootcamps* internos en habilidades de alta demanda como Python o SQL, asegurando que la fuerza laboral interna se mantenga relevante y reduciendo la necesidad de buscar talento externo.
3. Planificación de la Fuerza Laboral: La Tendencia de la Demanda Laboral permite a los líderes anticipar períodos de alto o bajo volumen de contratación , ayudando a planificar los recursos del equipo de abastecimiento con meses de antelación.

### **¿Qué insights se pueden obtener con solo mirar las gráficas?**

- Valor de la Experiencia: Se visualiza al instante la prima que el mercado global impone por la experiencia, por ejemplo, cuánto se dispara el salario promedio al pasar del nivel Medio al Senior

- Apetito por Flexibilidad: El gráfico de Distribución por Tipo de Empleo revela rápidamente si la mayoría de las ofertas son de Tiempo Completo (FT) , indicando que el mercado aún se inclina por la vinculación estable.
- Factores No Salariales: La gráfica que relaciona el salario con la Puntuación de Beneficios puede indicar que las empresas están compensando con paquetes de valor agregado donde el salario base puede no ser el más alto.

### **¿Cómo se simplifica la interpretación del modelo o los resultados?**

La sección del panel dedicada a los resultados del Machine Learning traduce la complejidad matemática en términos de negocio:

- Identificación de los Conductores: En lugar de revisar los coeficientes del modelo, las gráficas de Importancia de Características señalan directamente que el Nivel de Experiencia y la Industria son los principales motores del salario . Esto simplifica la discusión ejecutiva, enfocándola en lo relevante.
- Manejo del Error: Al mostrar el bajo  $R^2(0.2954)$  y el alto EAM (\$41,493 USD), el tablero establece expectativas realistas . Permite al usuario comprender que las predicciones deben usarse para establecer bandas salariales amplias (rangos) y no para ofrecer un valor exacto.

El tablero facilita al gerente de reclutamiento identificar que la demanda de roles en crecimiento, lo que puede ayudar a adelantar la contratación de perfiles especializados antes de que la competencia sature mercado. Además, el Error Absoluto Medio del modelo fomenta la decisión de trabajar con rangos salariales conservadores hasta integrar más datos geográficos.

Este proyecto ayuda a poder proporcionar una referencia analítica para ajustar políticas salariales, aunque con cautela en el valor exacto.

**Apoyo a RR.HH.:** El tablero final es una herramienta visual que simplifica la complejidad de los datos , permitiendo a los gerentes tomar decisiones informadas sobre la competitividad de sus ofertas.

Posibles mejoras y direcciones para investigaciones futuras

## REFERENCIAS:

Puri. P. (2025). <https://www.kaggle.com/datasets/pratyushpuri/global-ai-job-market-trend-2025?resource=download>