

PROGRAMA CODEAR FUTURA 2023

# INTELIGENCIA ARTIFICIAL EN EL EMPLEO:

Sus atributos e influencia en el  
mercado laboral global.

DOCUMENTACIÓN  
PROYECTO FINAL

---

SARAH V. PEÑA



# Índice

Explorando la Evolución Laboral en la Era de la Inteligencia Artificial.....	3
Tipo de análisis, objetivo y público objetivo.....	3
Descripción y ejecución del modelo ETL.....	4
Descripción del Modelo Entidad - Relación.....	5
Aplicando SQL al Dataset.....	7
Estrategia de Storytelling (Mockup del Tablero).....	10
Sobre el Dashboard.....	11
Disposiciones finales.....	12

# Explorando la Evolución Laboral en la Era de la Inteligencia Artificial

En un mundo donde la transformación digital se torna cada vez más una realidad innegable, el impacto de la inteligencia artificial en el panorama laboral ha desencadenado una revolución en la forma en que entendemos y concebimos el trabajo. Este viaje por la incursión de la inteligencia artificial en el ámbito laboral se traduce en una ventana para observar cómo la ciencia de datos y la emergente tecnología IA están moldeando los roles y las percepciones salariales a nivel global.

Desde una perspectiva descriptiva y exploratoria, el análisis presentado tiene como objetivo proporcionar una visión panorámica sobre el impacto de la inteligencia artificial en el ámbito laboral, ofreciendo datos actualizados sobre la adopción global de la IA y los ingresos asociados a roles vinculados con esta tecnología emergente.

Pensado para diversos actores del entorno laboral, este análisis está diseñado para sumergir a todos los interesados en el cambiante panorama, especialmente a profesionales que buscan adentrarse o evolucionar en el campo de la ciencia de datos. Este informe provee valiosa información acerca de la creciente demanda de roles relacionados con datos, necesaria para quienes buscan prepararse y navegar exitosamente en este contexto de cambio y evolución constante.

En definitiva, este análisis representa un viaje significativo que brinda una comprensión más amplia sobre cómo la inteligencia artificial está influenciando y transformando el entorno laboral a nivel global. Al proporcionar una visión general y detallada de los impactos, se convierte en un recurso valioso para todos aquellos que desean explorar las dinámicas cambiantes de la ciencia de datos y los ingresos asociados a esta revolución laboral en la era de la inteligencia artificial.

## Tipo de análisis, objetivo y público objetivo:

El enfoque de análisis contemplado es de carácter descriptivo y exploratorio.

La meta en este proyecto es proporcionar una visión general de cómo la introducción de la inteligencia artificial ha impactado en el ámbito laboral, presentando datos actualizados sobre su adopción a nivel global y los ingresos asociados a roles vinculados a esta tecnología emergente, desglosados por regiones.

Este análisis se destina a todos los involucrados en el entorno laboral, siendo su enfoque en:

- **Carrera en Ciencia de Datos:** Dirigido a profesionales que buscan cambiar de carrera o entrar en la ciencia de datos, el informe ofrece información sobre la creciente demanda de roles relacionados con datos y consejos para prepararse.

## Descripción y ejecución del modelo ETL

En este contexto este proceso consistiría en la obtención, preparación y carga de datos relacionados con la transformación del empleo y el impacto de la inteligencia artificial en los ingresos a nivel internacional.

### ❖ **Extracción (Extract):**

Obtención de datos de ingresos de fuentes como estadísticas gubernamentales y bases de datos de empleo, así como de datos sobre la influencia de la inteligencia artificial en estudios y empresas de tecnología.

Los datos utilizados para este proyecto se obtuvieron de los siguientes enlaces, ambos del archivo de datasets de Kaggle:

1. [Data Science Salaries 2023](#)
2. [AI Global Index](#)

Se descargaron los datasets disponibles en formato CSV. Se tuvo en cuenta que los mismos fueran recientes, ambos están actualizados al 2023.

### ❖ **Transformación (Transform):**

Este es el paso que implica limpiar datos, estandarizar formatos, integrar información de diferentes fuentes, realizar cálculos y segmentar datos por criterios como país e industria. Además, se agregan datos para resumirlos a niveles superiores, como ingresos promedio por país.

Luego de descargar los datasets, se realizaron las limpiezas básicas todas realizadas en Excel:

- Búsqueda y eliminación de nulos y duplicados
- Reorganización de columnas

Seguido a esto, se exploró la data para buscar maneras de relacionar las dos tablas a través de una columna que tuviera información con formato común. Las columnas que más se acercaban a este requerimiento en la tabla Data Science Salaries eran: `company_location` y `employee_residence`, que contienen el código país según el ISO 3166-1 alpha-2 (combinación única de 2 letras mayúsculas por país); y de la tabla AI Global Index la columna `country_code`.

Sin embargo, las informaciones disponibles en la tabla de salarios tienen países diferentes a los contenidos en el listado de se ejecutó un proceso de normalización de los datos, lo que creó la necesidad de normalizar los datos mediante una tercera tabla que contenga todos los países, convirtiéndose en la tabla intermediaria entre las dos.

A partir de esto, se ejecutó un proceso de normalización de los datasets, tomando como referencia la explicación de las formas normales [aquí](#). Resultando un dataset más robusto con

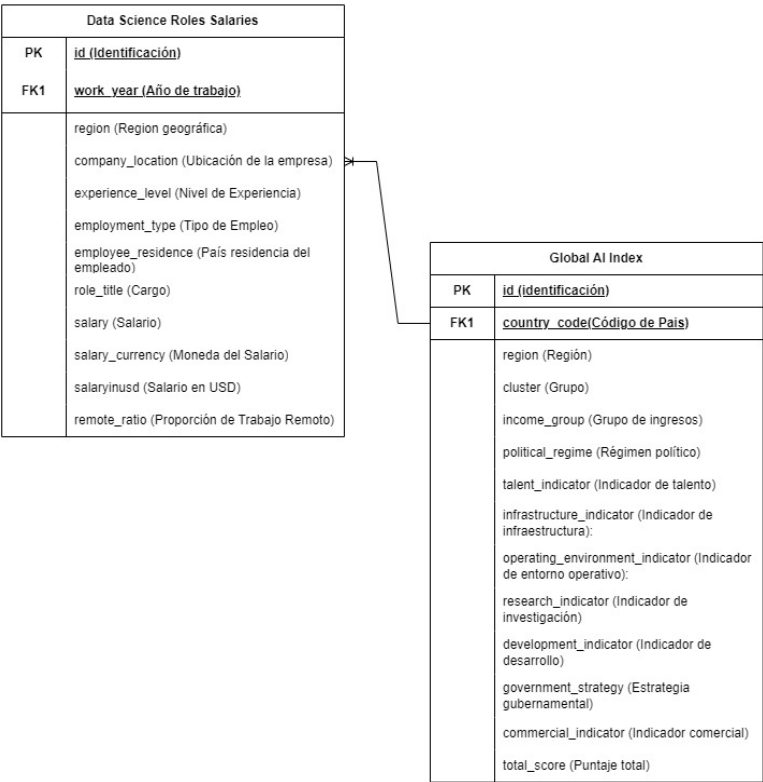
las tablas principales interconectadas por la tabla de los países, regiones y continentes y cada tabla principal tiene sus tablas anexas para identificar de manera única los valores distintos de cada atributo. Luego de hacer este proceso, se pudo añadir las columnas “id” a ambas tablas principales con la idea de que ya cada fila es un registro único de información. El dataset normalizado está disponible [aquí](#).

❖ **Carga (Load):**

Una vez procesados, los datos se almacenan y están listos para su análisis, permitiendo a los usuarios acceder a información valiosa mediante filtros y visualizaciones. El archivo de excel ya normalizado fue cargado a SQL Server para realizar queries e ir generando insights y a PowerBI para realizar las visualizaciones.

Descripción del Modelo Entidad - Relación

Se creó un esquema de base de datos que consta de dos tablas principales: una para los salarios en Ciencia de Datos y otra para el Índice Global de IA. La relación entre las dos tablas principales se establece mediante las columnas "company\_location" en la tabla de salarios y "country\_code" en la tabla del Índice Global de IA, mediadas por tres tablas intermedias: 'Continents', 'Regions' y 'Countries'. Este enfoque garantiza la concordancia entre la información de países en ambas tablas principales. El modelo entidad-relación completo con todas las tablas principales y secundarias está disponible [aquí](#). A continuación un resumen del modelo ER mostrando las 2 tablas principales:



El modelo de datos utilizado fue de constelación, o 'starflake', un híbrido entre modelo estrella y copo de nieve. En este caso, las tablas secundarias se relacionan directamente con las tablas principales, existiendo más de una tabla de hechos. Referencia conceptual [aquí](#).

#### *Tabla principal 1: Data\_Science\_Salaries*

1. **id [Clave Primaria]**: No. de identificación de la fila, numérica y autoincrementable
2. **work\_year (Año de trabajo) [Clave Foránea 2]**: El año en que se pagó el salario.
3. **experience\_level (Nivel de experiencia)**: El nivel de experiencia en el trabajo durante el año.
4. **employment\_type (Tipo de empleo)**: El tipo de empleo para el rol.
5. **job\_title (Cargo)**: El cargo en el que trabajó durante el año.
6. **salary (Salario)**: La cantidad total de salario bruto pagado.
7. **salary\_currency (Moneda del salario)**: La moneda del salario pagado
8. **salaryinusd (Salario en USD)**: El salario en dólares estadounidenses.
9. **employee\_residence (País de residencia del empleado)**: El país de residencia principal del empleado en el año de trabajo..
10. **remote\_ratio (Proporción de trabajo remoto)**: La cantidad general de trabajo realizado de forma remota.
11. **company\_location (Ubicación de la empresa) [Clave Foránea 1]**: El país de la oficina principal o sucursal contratante del empleador.
12. **company\_size (Tamaño de la empresa)**: El número mediano de personas que trabajaron para la empresa durante el año.

#### *Tabla principal 2: Global\_AI\_Index*

1. **id [Clave Primaria]**: No. de identificación de la fila, numérica y autoincrementable.
2. **country (País) [Clave Foránea]**: El país al que se refiere el índice.
3. **region (Región)**: La región geográfica a la que pertenece el país.
4. **cluster (Grupo)**: El grupo al que pertenece el país en términos de IA.
5. **income\_group (Grupo de ingresos)**: El grupo de ingresos económicos al que pertenece el país.
6. **political\_regime (Régimen político)**: El régimen político del país.
7. **talent\_indicator (Indicador de talento)**: Indicador que se centra en la disponibilidad de profesionales calificados para la prestación de soluciones de inteligencia artificial.
8. **infrastructure\_indicator (Indicador de infraestructura)**: Indicador que se centra en la confiabilidad y escala de la infraestructura de acceso, desde la electricidad hasta la capacidad de supercomputación.
9. **operating\_environment\_indicator (Indicador de entorno operativo)**: Indicador que se centra en el contexto regulatorio y la opinión pública sobre la inteligencia artificial.
10. **research\_indicator (Indicador de investigación)**: Indicador que se centra en la cantidad de investigación y de investigadores especializados en inteligencia artificial, investigando la cantidad de publicaciones y citas en revistas académicas creíbles.

11. development\_indicator (Indicador de desarrollo): Indicador que se centra en el desarrollo de plataformas y algoritmos fundamentales en los que se basan proyectos innovadores de inteligencia artificial.
12. government\_strategy (Estrategia gubernamental): Indicador de la profundidad del compromiso del gobierno nacional con la inteligencia artificial, investigando los compromisos de gasto y las estrategias nacionales.
13. commercial\_indicator (Indicador comercial): Indicador que se centra en el nivel de actividad de startups, inversión e iniciativas comerciales basadas en inteligencia artificial.

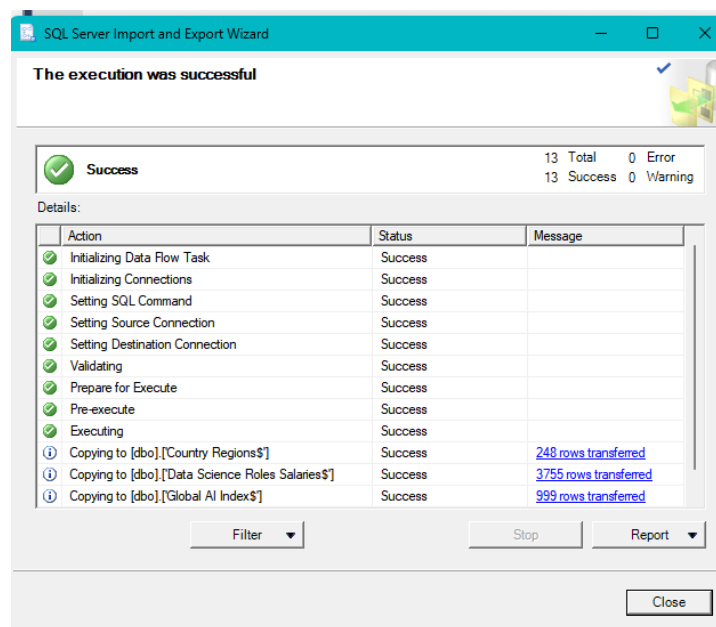
## Aplicando SQL al Dataset

El dataset base limpio utilizando excel optimizado para SQL Server se encuentra [aquí](#).

En primer lugar, se creó la base de datos para así importar el dataset al SQL Server y realizar los queries:

**CREATE DATABASE** Proyecto\_AT\_Sarah

Luego se cargó el Excel a SQL Server resultando el proceso así:



A continuación, los queries utilizados para las diferentes preguntas de investigación de este proyecto.

## Preguntas y Queries de SQL

1. Crear lista de los 10 países líderes en uso y desarrollo de la Inteligencia Artificial en base a su puntaje final.

```
SELECT TOP 10 GI.country_code, C.country_name, GI.total_score
FROM ['Global AI Index$'] AS GI
JOIN [Countries$] AS C ON GI.country_code = C.country_code
ORDER BY GI.total_score DESC
```

	country_code	country_name	total_score
1	US	United States of America	100
2	CN	China	62.92
3	GB	United Kingdom of Great Britain and Northern Ire...	40.93
4	CA	Canada	40.19
5	IL	Israel	39.89
6	SG	Singapore	38.67
7	KR	Republic of Korea	38.6
8	NL	Netherlands	36.35
9	DE	Germany	36.04
10	FR	France	34.42

2. Especificar las 3 regiones con mayor presencia en el desarrollo de la IA.

```
SELECT TOP 3 R.region_name, AVG(GI.total_score) AS 'Region Score'
FROM ['Global AI Index$'] AS GI
JOIN [Regions$] AS R ON GI.region_id = R.region_id
GROUP BY R.region_id, R.region_name
ORDER BY 'Region Score' DESC
```

	region_name	Region Score
1	Northern America	70.095
2	Eastern Asia	37.39
3	Western Europe	31.4571428571429

3. ¿Cuáles roles pagan más (Top 10)?

```
SELECT TOP 10 R.role_title, AVG(D.salary_in_usd) AS salary_average
FROM ['DataScience Salaries$'] AS D
JOIN [Roles$] AS R ON D.role_id = R.role_id
GROUP BY D.role_id, R.role_title
```



ORDER BY salary\_average DESC, D.role\_id

	role_title	salary_average
1	Data Science Tech Lead	375000
2	Cloud Data Architect	250000
3	Data Lead	212500
4	Data Analytics Lead	211254.5
5	Principal Data Scientist	198171.125
6	Director of Data Science	195140.727272727
7	Principal Data Engineer	192500
8	Machine Learning Software Engineer	192420
9	Data Science Manager	191278.775862069
10	Applied Scientist	190264.482758621

4. ¿Cuántos roles distintos de Ciencias de Datos existen?

```
SELECT COUNT(DISTINCT role_id) AS 'Role Count'
FROM ['DataScience Salaries$']
```

	Role Count
1	93

5. ¿Cuáles países pagan más en promedio (Top 10)?

```
SELECT TOP 10 DS.company_location, C.country_name,
AVG(DS.salary_in_usd) AS average_salary
FROM ['DataScience Salaries$'] AS DS
JOIN [Countries$] AS C ON DS.company_location = C.country_code
GROUP BY DS.company_location, C.country_name
ORDER BY average_salary DESC, DS.company_location
```

	company_location	country_name	average_salary
1	IL	Israel	271446.5
2	PR	Puerto Rico	167500
3	US	United States of America	151822.009539474
4	RU	Russian Federation	140333.333333333
5	CA	Canada	131917.689655172
6	NZ	New Zealand	125000
7	BA	Bosnia and Herzegovina	120000
8	IE	Ireland	114943.428571429
9	JP	Japan	114127.333333333
10	SE	Sweden	105000

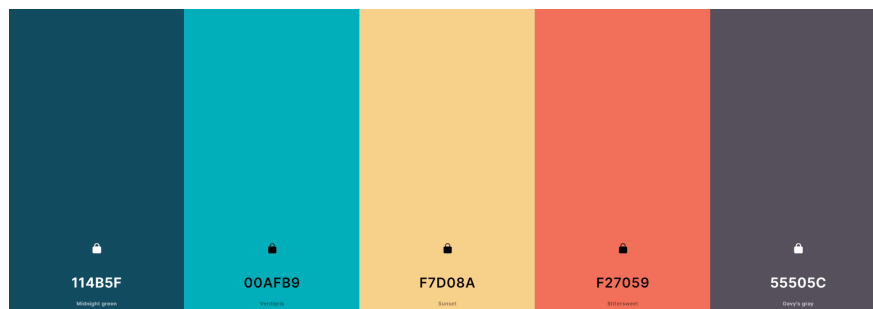
El archivo resultante con los queries para referencias está disponible [aquí](#).

## Estrategia de Storytelling (Mockup del Tablero)

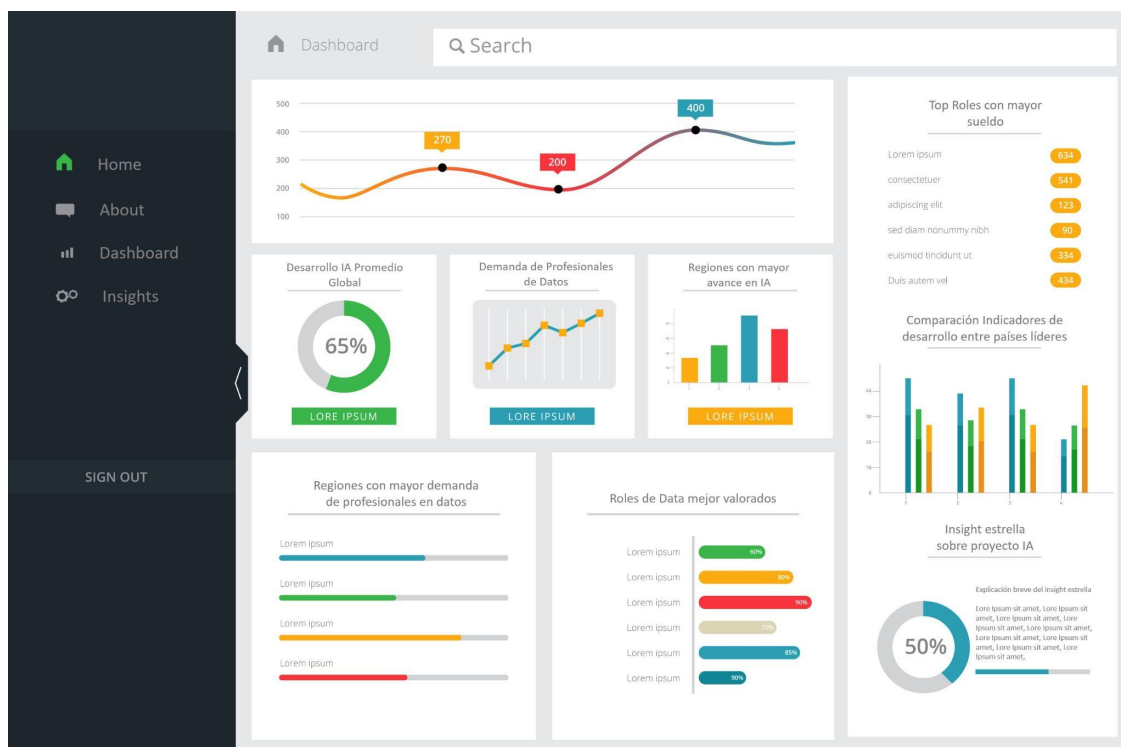
Los logotipos de Analista Tech para los fines de práctica de identidad y 'branding' son los siguientes:



La paleta de colores a utilizar es la siguiente:



El mockup de referencia del tablero es el siguiente (más detalles disponibles [aquí](#)):



## Sobre el Dashboard

### Cálculos en DAX realizados

Varias de las medidas creadas en DAX para PowerBI, considerando la estructura de la base de datos, son las siguientes:

#### **1. Promedio de Proporción de Trabajo Remoto:**

**Fórmula:** `Average Remote Ratio = AVERAGE('DataScience Salaries'[remote_ratio])`

**Explicación:** Calcula el promedio de la proporción de trabajo remoto en la tabla de salarios de ciencia de datos.

#### **2. Salario Promedio por Rol:**

**Fórmula:** `Avg Salary by Role = AVERAGEX(RELATEDTABLE('Roles'), [Avg Salary USD])`

**Explicación:** Utiliza la relación con la tabla de roles para calcular el salario promedio por cada rol en la tabla de salarios de ciencia de datos.

#### **3. Salario Promedio en USD:**

**Fórmula:** `Avg Salary USD = AVERAGE('DataScience Salaries'[salary_in_usd])`

**Explicación:** Calcula el salario promedio en dólares estadounidenses en la tabla de salarios de ciencia de datos.

#### **4. Países con Empresas Relacionadas con IA:**

**Fórmula:** `Countries with AI Related Companies = DISTINCTCOUNT('DataScience Salaries'[company_location])`

**Explicación:** Cuenta el número de países presentes en la tabla de salarios de ciencia de datos, lo que se interpreta como países con empresas relacionadas con inteligencia artificial.

#### **5. Top 5 Roles por Salario:**

**Fórmula:**

`Top5RolesBySalary =`

`VAR TopRoles =`

`TOPN(5, VALUES('Roles'[role_title]), CALCULATE(AVERAGE('DataScience Salaries'[salary_in_usd])),  
DESC)`

`RETURN`

`CALCULATE(  
AVERAGE('DataScience Salaries'[salary_in_usd]),  
'Roles'[role_title] IN TopRoles)`

**Explicación:** Identifica y presenta los cinco roles principales basados en el salario promedio en dólares en la tabla de salarios de ciencia de datos.

Estas medidas proporcionan insights clave sobre la proporción de trabajo remoto, salarios promedio por rol, salarios promedio en USD, la cantidad de países con empresas relacionadas con IA, y los roles más destacados por salario.

## Transformaciones de datos - Lenguaje M

En PowerBI, se llevaron a cabo transformaciones esenciales para garantizar la integridad y la coherencia de los datos. Esto incluyó la eliminación de registros con valores nulos, la corrección de tipos de datos incorrectos para asegurar la uniformidad, y la creación de una tabla calendario, la cual se integró al modelo y se relacionó con la tabla 'Work Years'. Estas acciones no solo contribuyeron a la limpieza de los datos, sino que también facilitaron un análisis más preciso y robusto, permitiendo una comprensión más profunda de la información temporal en el contexto de la evolución laboral y su relación con otros elementos del conjunto de datos.

## Fuentes de datos

En la creación del tablero de PowerBI, se emplearon datos provenientes de una hoja de cálculo en Excel. El archivo fuente puede ser accedido [aquí](#). Esta fuente de datos proporciona la base para el análisis y la visualización en el tablero, permitiendo explorar de manera efectiva la dinámica del empleo en la era de la inteligencia artificial. La utilización de Excel como fuente de datos facilita la manipulación y preparación de la información antes de su importación a PowerBI, asegurando la calidad y coherencia de los datos que respaldan las diversas métricas y visualizaciones en el tablero final.

## Disposiciones finales

Resumen de Insights tomados de los tableros:

### **1. Tablero Índices Globales de la IA:**

- La democracia liberal encabezó la lista con el promedio más alto de total score, situándose en 31.95.
- La autocracia cerrada, la democracia electoral y la autocracia electoral siguieron en orden descendente en términos de total score.
- Existe una correlación positiva entre Talento Profesional Disponible y Demanda del Mercado.
- Este análisis ofrece una visión comparativa de los distintos regímenes políticos, resaltando la fortaleza relativa de la democracia liberal en los indicadores evaluados.
- A juzgar por los indicadores de menor demanda del mercado y mayor talento disponible en otros países, se puede deducir que por la posibilidad de trabajos remotos, la demanda de mercados mayoritarios está acaparando el talento disponible de otros países, cuya una de las razones probables es el menor costo.

## 2. Tablero Roles de Ciencia de Datos:

- El año 2023 destacó con el máximo total de salarios pagados en dólares, alcanzando la cifra más alta de 266,046,291 dólares.
- Hubo un notable aumento del 3,692.53% en los salarios respecto al año 2020, que registró la cifra más baja de 7,015,000 dólares.
- La distribución de salarios totales pagados en dólares siguió un patrón descendente de mayor a menor entre los años 2023, 2022, 2021 y 2020.
- El año 2023 representó significativamente el 51.50% del total de salarios pagados en dólares, indicando su papel preponderante en el panorama económico.
- En conjunto, el rango de salarios totales pagados en dólares abarcó desde 7,015,000 hasta 266,046,291, demostrando una variabilidad considerable a lo largo de los cuatro años analizados.
- En la revisión de roles en el ámbito de la ciencia de datos, se destaca que la categoría "Senior" ocupa la posición más alta con un total de 75 roles, seguida por "Mid-Intermediate", "Entry Level" y "Executive". Asimismo, se observa que la categoría "Senior" con 75 roles supera en un 275.00% a "Executive", que registra el total más bajo con 20 roles.

En conclusión, la exploración detallada de la evolución laboral en la era de la inteligencia artificial revela insights cruciales sobre la transformación del panorama laboral global. La combinación de datos actualizados y análisis profundo proporciona una visión panorámica sobre el impacto de la inteligencia artificial, desde el aumento significativo en los salarios en 2023 hasta la destacada influencia de la democracia liberal en los índices globales de IA.

Este enfoque descriptivo y exploratorio, respaldado por un modelo de datos robusto, ofrece una herramienta valiosa para profesionales en ciencia de datos y todos los interesados en comprender y adaptarse al cambiante entorno laboral impulsado por la inteligencia artificial.

La estrategia de storytelling, desde la extracción y transformación de datos hasta la creación del dashboard en PowerBI, subraya la importancia de la calidad y coherencia de los datos en la generación de insights significativos.

En última instancia, este trabajo no solo proporciona una radiografía detallada de la situación actual, sino que también sirve como un recurso esencial para quienes buscan comprender las dinámicas cambiantes y anticipar las demandas del futuro laboral en la era de la inteligencia artificial.