

“AÑO DE LA RECUPERACIÓN Y CONSOLIDACIÓN  
DE LA ECONOMÍA PERUANA”.



ESCUELA PROFESIONAL DE CIENCIA DE LA  
COMPUTACIÓN  
TOPICOS EN CIENCIA DE DATOS

---

# Análisis Exploratorio de Datos - Data Wrangling

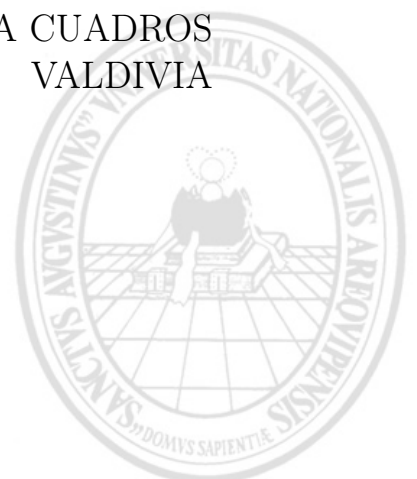
---

*Estudiantes:*

Jharold Alonso Mayorga Villena

*Docente :*

ANA MARIA CUADROS  
VALDIVIA



# 1. Análisis del Comportamiento de los Datos

Imagina que estás buscando empleo. Hoy en día, muchas personas utilizan Internet para encontrar trabajos en sitios web de empleo, aplicaciones y redes sociales. Al buscar una oferta de trabajo, ¿cómo sabes si la oferta es real o falsa? Algunas ofertas de trabajo pueden ser legítimas, mientras que otras pueden ser estafas que intentan aprovecharse de quienes buscan empleo. ¿Cómo puedes diferenciar una oferta real de una falsa?

En este proyecto, utilizamos un dataset llamado **Real / Fake Job Posting Prediction**, que contiene información sobre miles de ofertas de trabajo publicadas en línea. Este dataset tiene como objetivo enseñarnos a detectar patrones en las ofertas de trabajo para poder identificar cuáles son fraudulentas y cuáles son legítimas.

## 1.1. Descripción de los Datos

Cada columna del dataset tiene una función específica que nos da información sobre las ofertas de trabajo. Mientras que cada fila del dataset representa una oferta de trabajo que segun las características que contiene sobre cada columna se puede determinar su nivel de credibilidad. A continuación, se describen las columnas clave del dataset, con ejemplos de los valores que podrían contener.

Columna/Variables	Descripción
job_id	Identificador único para cada oferta de trabajo.
title	El título del trabajo ofrecido. Ejemplo: "Desarrollador de Software".
location	Ubicación geográfica de la oferta de trabajo. Ejemplo: "US, NY, New York".
department	El departamento que ofrece el trabajo. Ejemplo: "Marketing".
salary_range	El rango salarial, si está disponible. Ejemplo: "50,000 - 60,000 USD".
company_profile	Descripción breve sobre la empresa.
description	Detalles sobre las responsabilidades del trabajo.
requirements	Requisitos para aplicar al trabajo.
benefits	Beneficios adicionales que ofrece la empresa.
telecommuting	Indica si el trabajo es remoto (1 = sí, 0 = no).
has_company_logo	Si la oferta tiene el logo de la empresa (1 = sí, 0 = no).
has_questions	Si la oferta incluye preguntas adicionales (1 = sí, 0 = no).
employment_type	Tipo de empleo (por ejemplo, "Full-time", "Internship").
required_experience	La experiencia mínima requerida para el puesto.
required_education	El nivel de educación requerido.
industry	La industria a la que pertenece el trabajo (por ejemplo, "Marketing", "IT").
function	La función dentro de la empresa (por ejemplo, "Marketing", "Customer Service").
fraudulent	Indica si la oferta es fraudulenta o no (1 = fraudulenta, 0 = legítima).

Cuadro 1: Ejemplo de un registro del dataset.

## 1.2. Tipos de Datos y Formato

Es esencial analizar los tipos de datos y el formato de cada columna para asegurar que estén listos para el análisis y modelado posterior. A continuación se detallan los tipos de datos de cada columna y su clasificación.

Columna	Tipo de Dato	Clasificación	Formato
job_id	Cuantitativo	Discreto	Numérico
title	Cualitativo	Nominal	Texto (Categoría)
location	Cualitativo	Nominal	Texto (Categoría)
department	Cualitativo	Nominal	Texto (Categoría)
salary_range	Cuantitativo	Continuo	Numérico (Rango)
company_profile	Cualitativo	Nominal	Texto
description	Cualitativo	Nominal	Texto
requirements	Cualitativo	Nominal	Texto
benefits	Cualitativo	Nominal	Texto
telecommuting	Cuantitativo	Discreto	Binario (0, 1)
has_company_logo	Cuantitativo	Discreto	Binario (0, 1)
has_questions	Cuantitativo	Discreto	Binario (0, 1)
employment_type	Cualitativo	Nominal	Texto (Categoría)
required_experience	Cuantitativo	Ordinal	Numérico
required_education	Cualitativo	Ordinal	Texto (Categoría)
industry	Cualitativo	Nominal	Texto (Categoría)
function	Cualitativo	Nominal	Texto (Categoría)
fraudulent	Cuantitativo	Discreto	Binario (0, 1)

Cuadro 2: Tipos de datos y formato de las columnas del dataset.

En la tabla anterior, se especifica el tipo de cada columna:

- **Cuantitativo:** Datos numéricos que pueden ser medidos y analizados en términos de cantidades, como el salario o los años de experiencia.
- **Cualitativo:** Datos no numéricos, como el título del trabajo o la ubicación, que representan categorías.
- **Nominal:** Variables que no tienen un orden inherente, como el título del trabajo o la ubicación.
- **Ordinal:** Variables con un orden específico, como los años de experiencia, donde 3 años es más que 2, pero menos que 5.
- **Continuo:** Variables numéricas que pueden tomar cualquier valor dentro de un rango, como el salario.
- **Discreto:** Variables numéricas que solo toman valores específicos, como la cantidad de trabajo remoto (0 o 1).

## 2. Análisis de la Columna title

En este análisis, exploramos la columna "title" del dataset, que contiene los títulos de las ofertas de trabajo. Dado que la columna "title" es de tipo categórico, no realizamos análisis numéricos tradicionales como en las columnas de salario o experiencia. Sin embargo, realizamos un análisis detallado para entender cómo se distribuyen los títulos y si existe alguna relación interesante con otras variables, como el salario.

### 2.1. Clasificación de los Títulos y Conteo de Frecuencia

Para comenzar, observamos cómo están clasificados los títulos de trabajo en el dataset, es decir, cuántos títulos únicos existen y cuántas veces se repite cada uno. Este análisis es importante para ver si hay una concentración de títulos similares o si existe una gran diversidad en los tipos de trabajos ofrecidos.

#### Gráfico 1: Distribución de los Títulos de Trabajo más Frecuentes.

El gráfico a continuación muestra los 20 títulos de trabajo más frecuentes. Esto nos da una idea de los trabajos más comunes en el dataset y nos permite identificar posibles tendencias o patrones en las ofertas de trabajo.

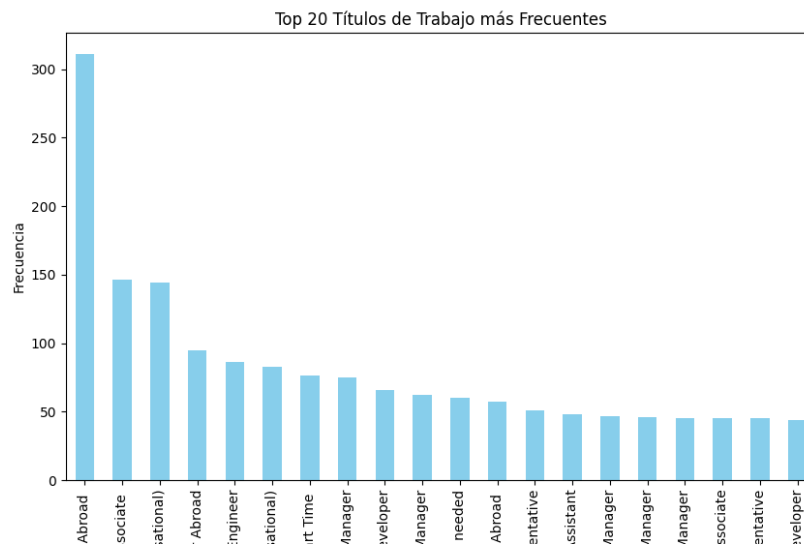


Figura 1: Top 20 Títulos de Trabajo más Frecuentes

**Explicación del gráfico:** El gráfico de barras presenta los 20 títulos más comunes de ofertas de trabajo. El eje  $x$  representa los títulos de trabajo, y el eje  $y$  muestra la frecuencia con la que cada título aparece en el dataset.

## 2.2. Correlación entre los Títulos y el Salario

Aunque los títulos de trabajo son categóricos, podemos investigar si existe alguna relación entre los títulos y otras variables numéricas, como el salario. Para ello, utilizamos un gráfico de `boxplot` que muestra cómo varían los salarios para cada título de trabajo.

### Gráfico 2: Distribución de Salarios por Título de Trabajo.

El siguiente gráfico de `boxplot` nos muestra la distribución del salario para cada título de trabajo. Este gráfico nos ayuda a entender si hay títulos asociados con salarios más altos o más bajos, y si existen valores atípicos (*outliers*).

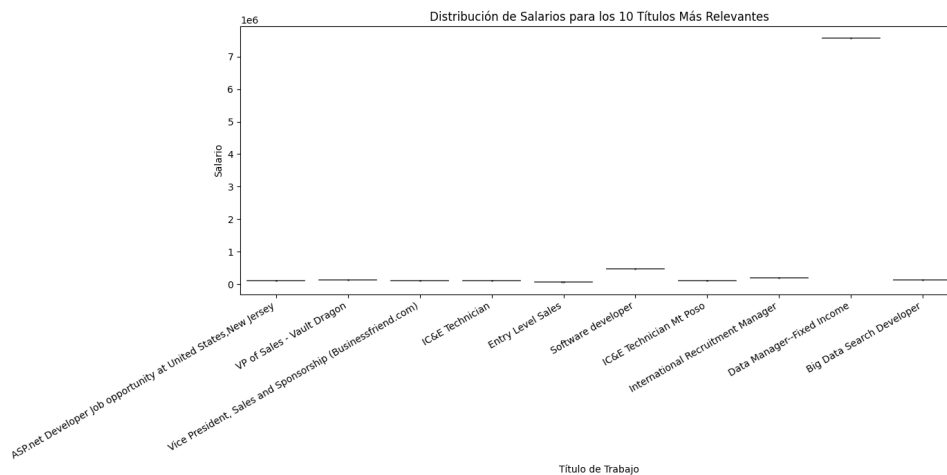


Figura 2: Distribución de Salarios por Título de Trabajo

**Explicación del gráfico:** El `boxplot` muestra la distribución de los salarios para cada título de trabajo. La línea dentro de la caja indica la mediana, los bordes de la caja son los cuartiles (Q1 y Q3), y los puntos fuera de los "bigotes" son los valores atípicos (*outliers*).

## 2.3. Outliers en los Títulos de Trabajo

En un análisis de datos, es importante identificar los posibles valores atípicos o *outliers*. En este caso, los *outliers* en los títulos de trabajo son aquellos que ocurren con una frecuencia muy baja, lo que podría indicar errores de entrada de datos o categorías raras que no representan la mayoría de las ofertas.

### Gráfico 3: Títulos de Trabajo con Baja Frecuencia.

A continuación, mostramos un gráfico de barras que visualiza los títulos de trabajo que aparecen con menos frecuencia en el dataset. Estos títulos podrían ser errores o valores atípicos que merecen una revisión más profunda.

**Explicación del gráfico:** El gráfico de barras muestra los títulos de trabajo que aparecen con poca frecuencia. Si algún título aparece solo una o dos veces, podría ser un error en los datos o una categoría irrelevante para el análisis.

## 2.4. Conclusión del Análisis de la Columna title

El análisis de la columna "title" nos ha permitido identificar varios aspectos interesantes sobre las ofertas de trabajo en el dataset:

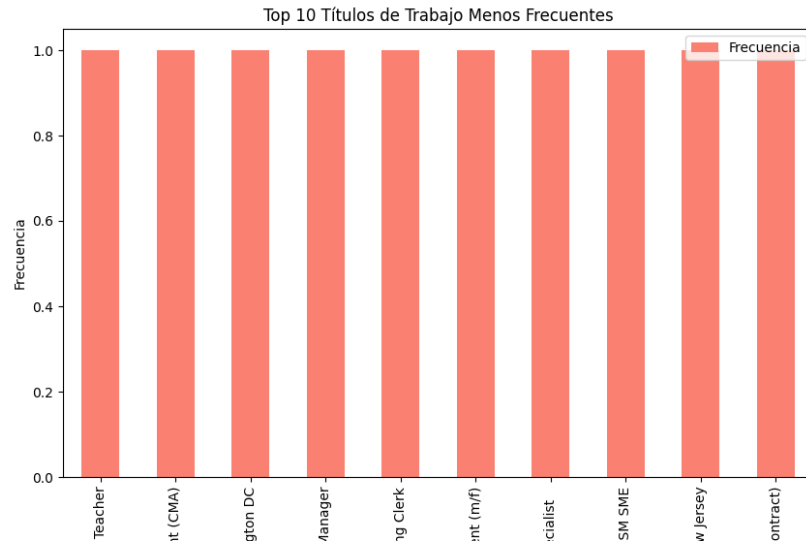


Figura 3: Títulos de Trabajo con Baja Frecuencia

- La mayoría de los trabajos están concentrados en ciertos títulos, lo que refleja las áreas más comunes de empleo en el dataset.
- Hemos observado una variabilidad significativa en los salarios para diferentes títulos de trabajo. Algunos títulos están asociados con salarios más altos, mientras que otros están relacionados con salarios más bajos.
- Los títulos de trabajo con baja frecuencia podrían ser considerados como outliers, y algunos de ellos pueden ser errores o categorías no relevantes.

Este análisis nos proporciona una base para entender cómo los títulos de trabajo están distribuidos en el dataset y cómo pueden estar relacionados con otras variables importantes como el salario.

### 3. Análisis de la Columna location

En esta sección, realizamos un análisis detallado de la columna "location" del dataset, que contiene las ubicaciones geográficas de las ofertas de trabajo. La columna "location" es categórica, lo que significa que contiene descripciones de texto que indican la ciudad, país o si el trabajo es remoto. A pesar de ser categórica, podemos obtener información clave sobre la distribución geográfica de las ofertas y si existen relaciones interesantes con otras variables, como el salario.

#### 3.1. Clasificación de Ubicaciones y Conteo de Frecuencia

Primero, analizamos cómo están clasificadas las ubicaciones en el dataset, es decir, cuántas veces se repite cada ubicación. Este análisis nos ayuda a ver si las ofertas de trabajo están concentradas en ciertas ubicaciones o si existen ofertas distribuidas en varias regiones geográficas.

##### Gráfico 1: Distribución de las Ubicaciones más Frecuentes.

A continuación, se muestra un gráfico de barras que visualiza las ubicaciones más frecuentes en el dataset. Esto nos permitirá observar si hay ciudades o países con una alta concentración de ofertas o si las ofertas están distribuidas de manera más homogénea.

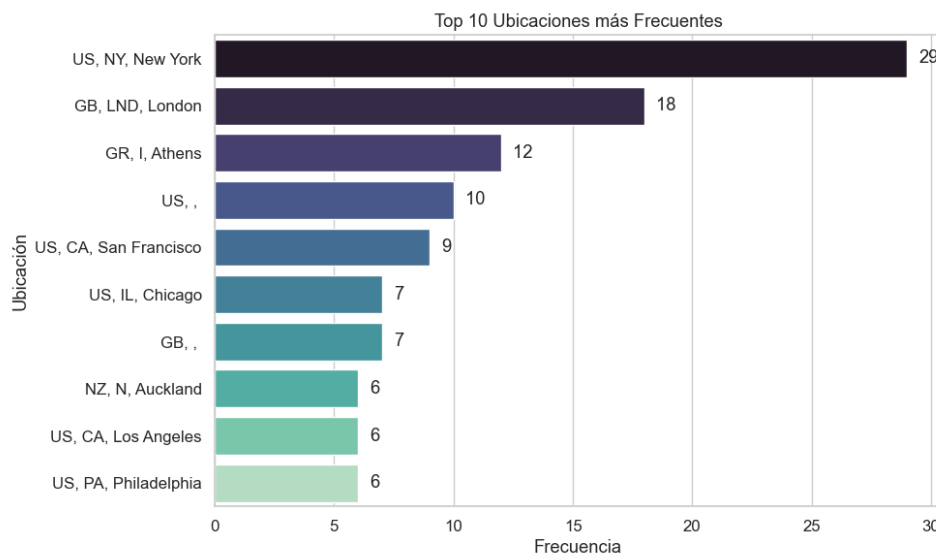


Figura 4: Top 20 Ubicaciones más Frecuentes

**Explicación del gráfico:** El gráfico de barras muestra las 20 ubicaciones más comunes de las ofertas de trabajo. El eje  $x$  representa las ubicaciones, y el eje  $y$  muestra la frecuencia con la que aparece cada ubicación en el dataset. Esto nos ayuda a identificar si hay una alta concentración de ofertas en ciertos lugares, como en grandes ciudades o países específicos.

#### 3.2. Identificación de Outliers en las Ubicaciones

Aunque las columnas categóricas no tienen outliers en el mismo sentido que las columnas numéricas, podemos considerar como outliers aquellas ubicaciones que aparecen

con muy baja frecuencia. Estas ubicaciones podrían ser errores de entrada o categorías que no aportan valor al análisis.

### Gráfico 2: Ubicaciones con Baja Frecuencia.

El gráfico siguiente muestra las ubicaciones que ocurren con una baja frecuencia en el dataset. Esto nos ayudará a identificar posibles errores en la entrada de datos o ubicaciones irrelevantes para el análisis.

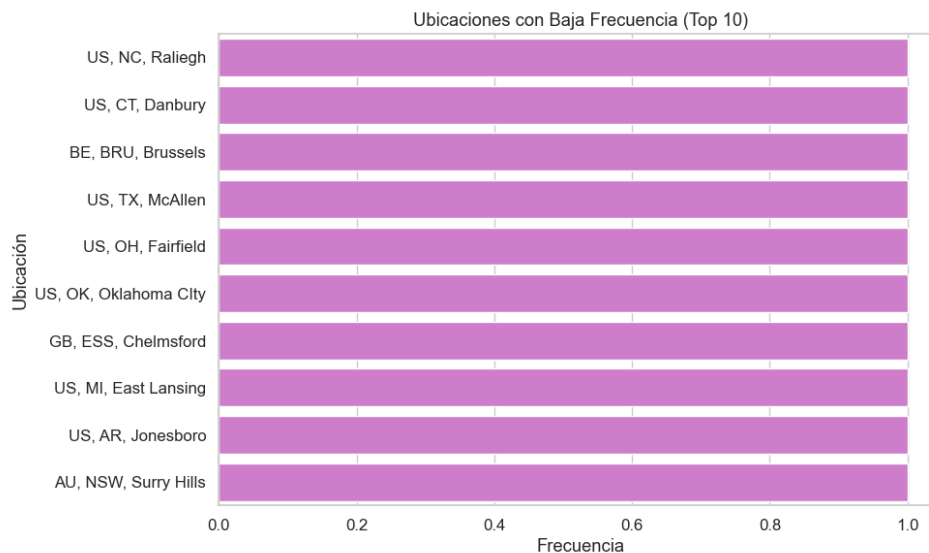


Figura 5: Ubicaciones con Baja Frecuencia

**Explicación del gráfico:** Este gráfico de barras muestra las ubicaciones que ocurren muy pocas veces en el dataset. Las ubicaciones con una frecuencia baja (menos de dos veces) son consideradas outliers y podrían requerir revisión. Estas categorías raras podrían ser errores o simplemente ubicaciones no relevantes.

### 3.3. Correlación entre Ubicación y Salario

Aunque "location" es una columna categórica, podemos explorar si existe alguna relación entre las ubicaciones y el salario. Utilizamos un boxplot para visualizar cómo varían los salarios en función de la ubicación de cada oferta de trabajo.

#### Gráfico 3: Distribución de Salarios por Ubicación.

A continuación, se presenta un boxplot que muestra cómo se distribuyen los salarios para cada ubicación. Este gráfico nos permite observar si hay algunas ubicaciones asociadas con salarios más altos o más bajos, y si existen valores atípicos.

**Explicación del gráfico:** El boxplot muestra la distribución del salario para cada ubicación. La línea dentro de cada caja indica la mediana del salario, mientras que los bordes de la caja representan el primer y tercer cuartil. Los puntos fuera de los "bigotes" son los outliers, que indican valores extremos en el salario para ciertas ubicaciones. Este gráfico nos ayuda a entender si hay ubicaciones asociadas con salarios inusuales.



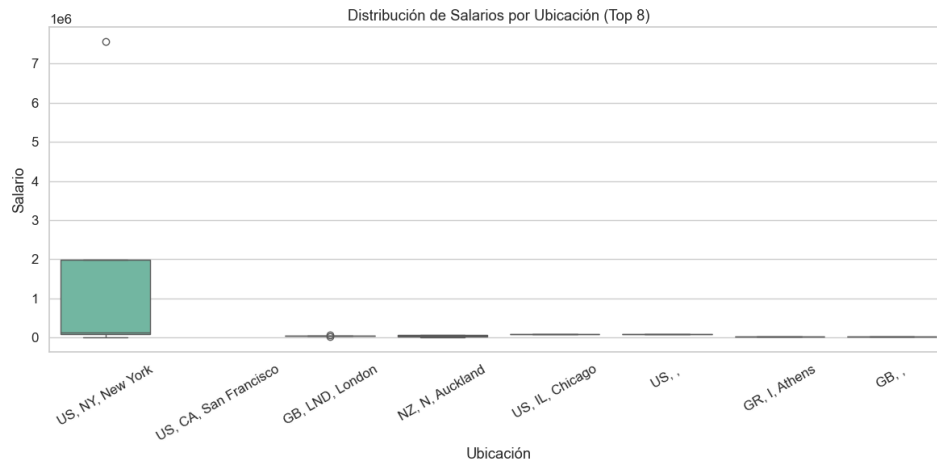


Figura 6: Distribución de Salarios por Ubicación

### 3.4. Clasificación de Ubicaciones: Remoto vs. Local

Además de analizar las ubicaciones geográficas, también podemos clasificar las ofertas de trabajo como remotas o locales. En este caso, clasificamos las ubicaciones en dos categorías: trabajos remotos (si la ubicación contiene la palabra "remote") y trabajos locales (si especifican una ciudad o país).

#### Gráfico 4: Distribución de Ubicaciones: Remoto vs. Local.

A continuación, se presenta un gráfico que muestra cuántas ofertas de trabajo son remotas y cuántas son locales. Este gráfico nos permitirá ver cómo se distribuyen las ofertas de trabajo en función de su tipo de ubicación.

**Explicación del gráfico:** Este gráfico de barras muestra la distribución de ofertas de trabajo remotas frente a ofertas locales. Las barras indican la frecuencia de cada tipo de ubicación, lo que nos permite observar qué proporción de ofertas de trabajo permite el trabajo desde cualquier lugar (remoto) y cuáles requieren estar ubicados físicamente en un lugar específico (local).

### 3.5. Conclusión del Análisis de la Columna location

El análisis de la columna "location" nos ha permitido obtener una visión más clara de cómo se distribuyen las ofertas de trabajo geográficamente:

- Las ubicaciones están distribuidas de manera desigual, con algunas ciudades o países concentrando más ofertas de trabajo que otros.
- Hemos identificado outliers en las ubicaciones, que son valores que ocurren con poca frecuencia y podrían ser errores de entrada de datos.
- La relación entre la ubicación y el salario muestra variabilidad, lo que nos permite observar si ciertos lugares están asociados con salarios más altos o bajos.
- Finalmente, hemos clasificado las ubicaciones en remotas y locales, proporcionando una visión de las tendencias actuales en cuanto a trabajos remotos.

Este análisis nos proporciona una base sólida para entender cómo la ubicación puede influir en el salario y las características de las ofertas de trabajo.

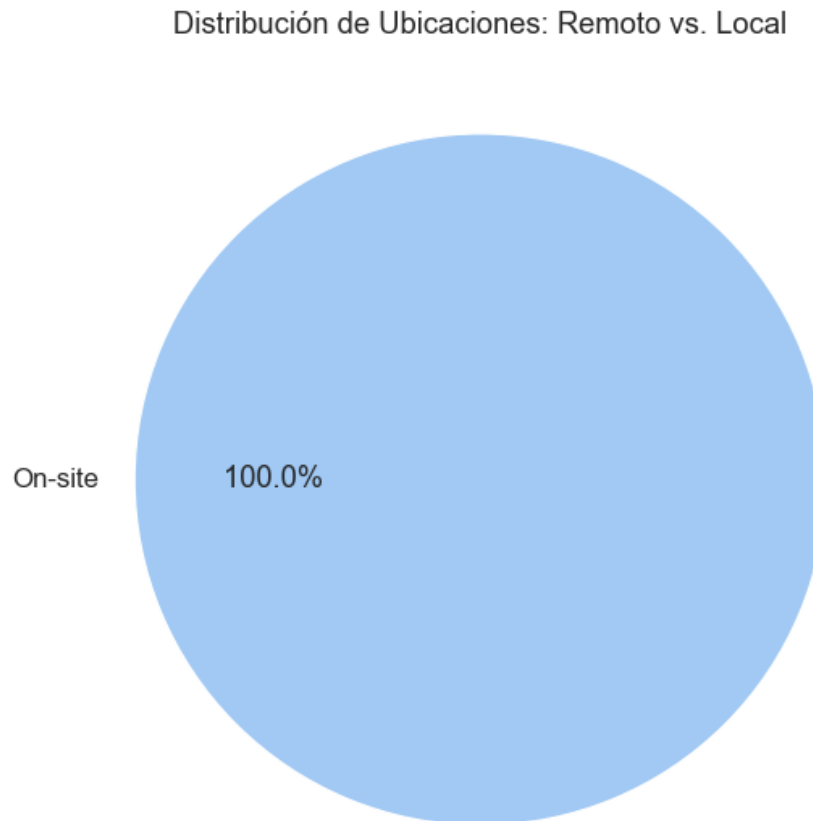


Figura 7: Distribución de Ubicaciones: Remoto vs. Local

## 4. Análisis de la Columna department

En esta sección, realizamos un análisis detallado de la columna `department` del dataset, que contiene los departamentos a los cuales pertenecen las ofertas de trabajo. La columna `department` es de tipo categórico, lo que significa que contiene valores de texto que representan diferentes departamentos, como Marketing, IT, Finanzas, etc. Aunque no es una columna numérica, podemos analizar cómo se distribuyen las ofertas de trabajo entre los departamentos y cómo varía el salario o la frecuencia de las ofertas.

### 4.1. Clasificación de Departamentos y Conteo de Frecuencia

Para comenzar, analizamos cómo están clasificados los departamentos en el dataset, es decir, cuántas veces se repite cada departamento. Este análisis nos ayudará a comprender si las ofertas de trabajo están concentradas en algunos departamentos o si están distribuidas de manera más equitativa entre diferentes áreas.

#### Gráfico 1: Distribución de las Ofertas de Trabajo por Departamento.

El siguiente gráfico circular (pie chart) muestra las 20 ubicaciones más frecuentes y su distribución en el dataset. Este gráfico es ideal para entender cómo se reparten las ofertas de trabajo entre los departamentos y destacar las áreas más comunes.

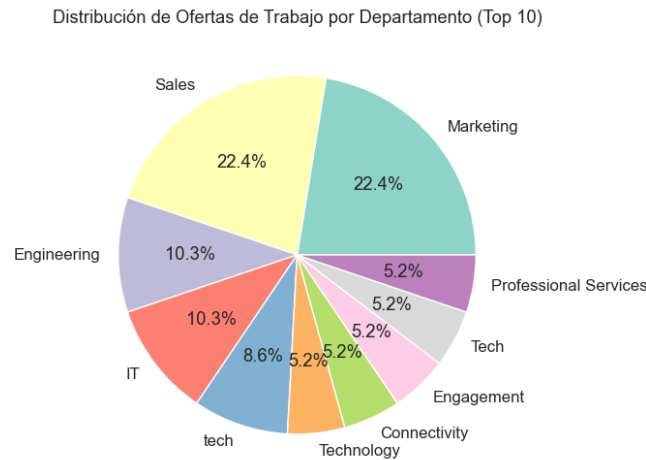


Figura 8: Distribución de las Ofertas de Trabajo por Departamento

**Explicación del gráfico:** El gráfico de pastel muestra la distribución de las ofertas de trabajo entre los diferentes departamentos. Cada segmento del gráfico representa un departamento y su tamaño es proporcional al número de ofertas de trabajo en ese departamento. Este gráfico es útil para ver rápidamente qué departamentos tienen más ofertas y cuáles son menos representados.

## 4.2. Relación entre el Departamento y el Salario

Aunque la columna `department` es categórica, podemos explorar cómo se distribuyen los salarios entre los diferentes departamentos. Para ello, utilizamos un gráfico de violín que muestra cómo varían los salarios dentro de cada departamento.

### Gráfico 2: Distribución de Salarios por Departamento.

El siguiente gráfico de violín compara la distribución de los salarios en función del departamento. Este gráfico permite observar cómo se dispersan los salarios dentro de cada departamento, lo que nos ayuda a detectar si existen departamentos asociados con salarios más altos o más bajos.

**Explicación del gráfico:** El gráfico de violín combina un boxplot y una distribución de densidad, lo que permite observar la mediana de los salarios, su rango intercuartílico y la presencia de valores atípicos (outliers) en cada departamento. El gráfico nos ayuda a ver cómo se comparan los salarios entre departamentos y si hay una gran variabilidad dentro de cada uno.

## 4.3. Identificación de Outliers en los Departamentos

En el análisis de categorías, podemos considerar como outliers aquellos departamentos que tienen una frecuencia extremadamente baja. Los departamentos con solo una o dos ofertas podrían ser considerados outliers, ya que representan una pequeña proporción de las ofertas de trabajo.

### Gráfico 3: Departamentos con Baja Frecuencia.

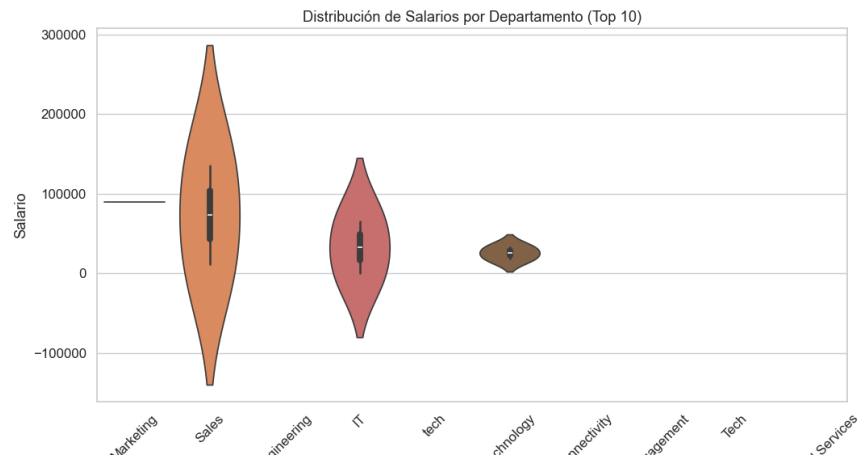


Figura 9: Distribución de Salarios por Departamento

Este gráfico de barras muestra los departamentos con una frecuencia baja, es decir, aquellos departamentos que tienen solo unas pocas ofertas de trabajo. Estos departamentos son considerados outliers y podrían ser errores o categorías irrelevantes.

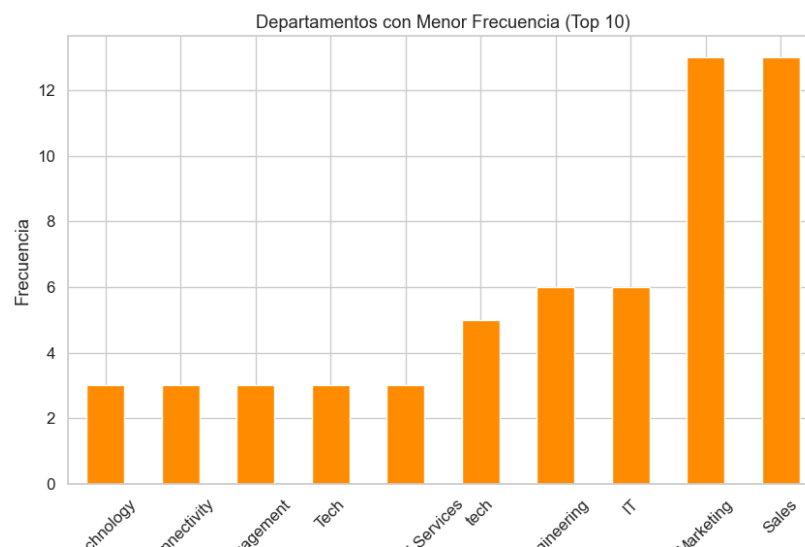


Figura 10: Departamentos con Baja Frecuencia

**Explicación del gráfico:** Este gráfico de barras visualiza los departamentos que tienen una baja frecuencia de ofertas (menos de dos). Los departamentos con poca representación en el dataset podrían ser outliers y requieren revisión para verificar si son datos relevantes o si hay algún error en su clasificación.

#### 4.4. Clasificación de Departamentos por Frecuencia

Finalmente, para entender mejor cómo se distribuyen las ofertas de trabajo en cada departamento, usamos un gráfico de barras horizontal. Este gráfico muestra cuántas ofertas de trabajo existen para cada departamento de manera clara y visual.

#### Gráfico 4: Frecuencia de Ofertas de Trabajo por Departamento.

A continuación, se presenta un gráfico de barras horizontal que muestra cuántas ofertas de trabajo existen para cada departamento. Este gráfico nos ayuda a comparar rápidamente la cantidad de ofertas entre departamentos.

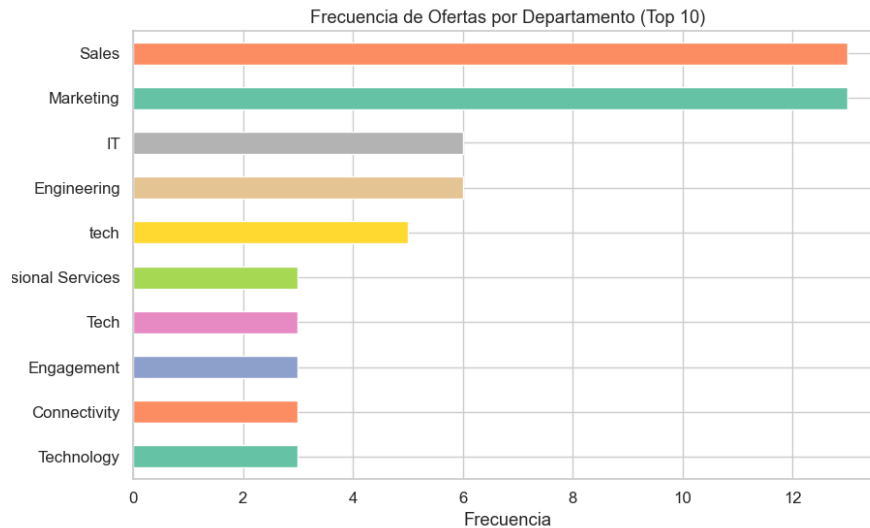


Figura 11: Frecuencia de Ofertas de Trabajo por Departamento

**Explicación del gráfico:** El gráfico de barras horizontal muestra la cantidad de ofertas de trabajo para cada departamento. Los departamentos con más ofertas están en la parte superior y los menos representados en la parte inferior. Este gráfico es útil para identificar rápidamente qué departamentos tienen la mayor cantidad de trabajos disponibles.

#### 4.5. Conclusión del Análisis de la Columna department

El análisis de la columna department nos ha proporcionado una visión más clara de cómo se distribuyen las ofertas de trabajo en el dataset:

- La distribución de las ofertas de trabajo entre los diferentes departamentos muestra qué áreas son más comunes, con un gráfico circular que resalta los departamentos con mayor representación.
- La relación entre el departamento y el salario se visualiza de manera efectiva mediante un gráfico de violín, que muestra la variabilidad de los salarios en cada departamento.
- Los outliers en la frecuencia de departamentos han sido identificados con un gráfico de barras, permitiéndonos detectar departamentos con baja frecuencia que podrían ser errores o categorías irrelevantes.
- Finalmente, la frecuencia de ofertas por departamento ha sido comparada utilizando un gráfico de barras horizontal, lo que nos permite ver de manera clara cuáles departamentos tienen más ofertas.

Este análisis proporciona una base sólida para comprender cómo las ofertas de trabajo están distribuidas en el dataset y cómo se relacionan con otras variables, como el salario.

## 5. Análisis de la Columna salary range

En esta sección, realizamos un análisis detallado de la columna `salary range`, que contiene los rangos salariales de las ofertas de trabajo. Dado que esta columna es numérica, realizaremos el análisis de las principales estadísticas descriptivas y visualizaremos los resultados con gráficos interactivos que incluyen valores clave como la media, mediana, mínimo, máximo, y los percentiles 25 % y 75 %.

### 5.1. Estadísticas Descriptivas para salary range

Primero, calculamos las estadísticas descriptivas más importantes de la columna `salary range`, incluyendo el mínimo, máximo, media, mediana, y los percentiles 25 % y 75 %, que nos ayudarán a entender la distribución de los salarios en el dataset.

#### Gráfico 1: Distribución de Salarios y Estadísticos

El siguiente gráfico muestra la distribución de los salarios junto con las líneas verticales que representan los valores de mínimo, máximo, media, mediana y los percentiles 25 % y 75 %.

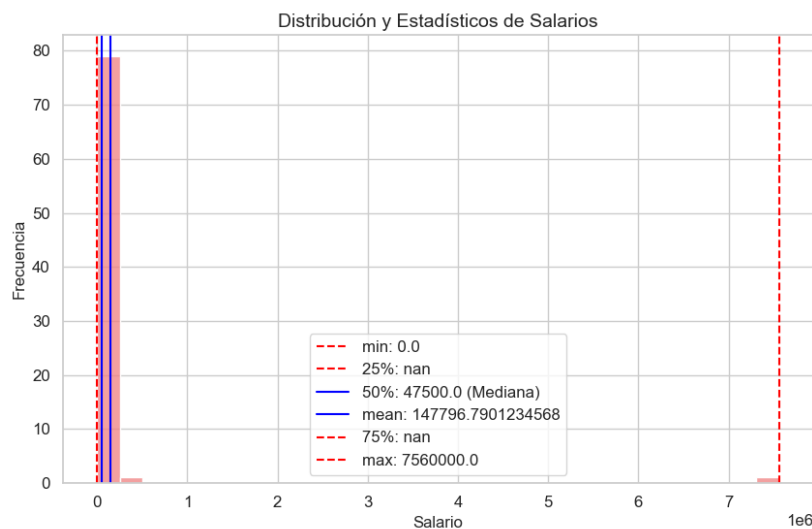


Figura 12: Distribución de Salarios con Estadísticos (mínimo, 25 %, 50 %, media, 75 %, máximo)

**Explicación del gráfico:** El gráfico de histograma muestra cómo se distribuyen los salarios en el dataset. Las líneas verticales representan los valores calculados:

- *Líneas rojas:* Muestran los percentiles 25 % y 75 %, que nos indican en qué rangos caen el 50 % central de los salarios.
- *Líneas azules:* Representan la media y la mediana de los salarios.
- *Líneas negras:* El mínimo y máximo de los salarios, que nos dan los rangos de los valores extremos.

## 5.2. Detección de Outliers en Salarios

Para detectar los outliers o valores atípicos, utilizamos el Rango Intercuartílico (IQR). Cualquier salario fuera del rango de  $Q1 - 1.5 * IQR$  a  $Q3 + 1.5 * IQR$  se considera un outlier.

### Gráfico 2: Outliers en los Salarios

A continuación, se muestra un boxplot que visualiza los outliers en la distribución de los salarios. Los valores fuera de los bigotes del boxplot representan los salarios atípicos.

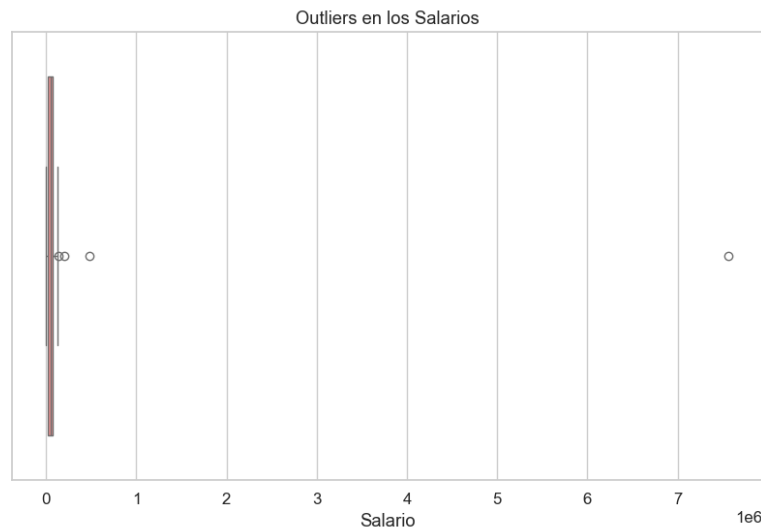


Figura 13: Outliers en los Salarios

**Explicación del gráfico:** El boxplot muestra la distribución de los salarios. Los puntos fuera de los bigotes son considerados outliers. Estos salarios podrían ser indicativos de errores de entrada de datos o de ofertas fraudulentas.

## 5.3. Correlación entre Salario y Experiencia Requerida

Finalmente, exploramos si existe alguna relación entre el salario y la experiencia requerida. Un gráfico de dispersión nos ayuda a visualizar esta posible correlación.

### Gráfico 3: Correlación entre Salario y Experiencia Requerida

A continuación, se muestra un gráfico de dispersión que visualiza la relación entre los años de experiencia y el salario. Este gráfico nos permitirá observar si existe alguna tendencia en la que los salarios aumenten con la experiencia.

**Explicación del gráfico:** El gráfico de dispersión nos permite ver cómo se distribuyen los salarios en función de los años de experiencia. Si existe una correlación positiva, esperaríamos que los salarios aumenten a medida que la experiencia requerida también aumenta.

## 5.4. Conclusión del Análisis de la Columna salary range

El análisis de la columna `salary range` nos ha proporcionado una visión clara sobre la distribución de los salarios en el dataset:



Figura 14: Correlación entre Salario y Experiencia Requerida

- La distribución de salarios muestra que hay un rango significativo de salarios, con algunos valores extremos que podrían ser outliers.
- Los outliers fueron identificados utilizando el IQR y visualizados mediante un box-plot, lo que nos permite detectar posibles datos erróneos o fraudulentos.
- La correlación entre el salario y la experiencia requerida muestra si los salarios aumentan conforme se requiere más experiencia.

Este análisis proporciona una base sólida para entender cómo se distribuyen los salarios y cómo están relacionados con otras variables, lo que nos ayudará a identificar patrones relevantes y posibles ofertas fraudulentas.



## 6. Análisis de la Columna telecommuting

En esta sección, realizamos un análisis detallado de la columna `telecommuting`, que contiene información sobre si las ofertas de trabajo son remotas (`telecommuting`) o no remotas. Esta columna es categórica y binaria, con dos posibles valores: 0 (no remoto) y 1 (remoto). Analizaremos cómo se distribuyen estas ofertas y cómo se relacionan con otras variables, como el salario y el tipo de empleo.

### 6.1. Clasificación de Ofertas de Trabajo: Remoto vs. No Remoto

Primero, observamos la distribución de las ofertas de trabajo entre los trabajos remotos (valor 1) y los no remotos (valor 0). Para esto, utilizamos un gráfico circular que nos muestra la proporción de trabajos remotos frente a los no remotos.

#### Gráfico 1: Distribución de Trabajos Remotos vs No Remotos.

A continuación, se muestra un gráfico circular que visualiza la distribución de trabajos remotos y no remotos en el dataset.

Distribución de Trabajos Remotos vs No Remotos

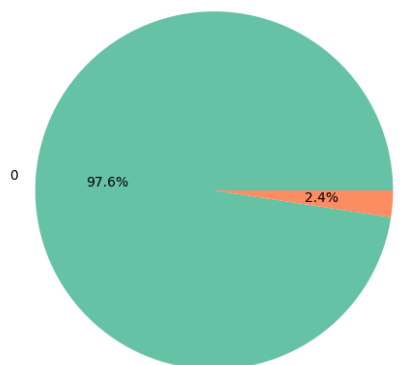


Figura 15: Distribución de Trabajos Remotos vs No Remotos

**Explicación del gráfico:** El gráfico de pie chart muestra la proporción de ofertas de trabajo remotas (valor 1) y no remotas (valor 0). Este gráfico es útil para ver la distribución de trabajos remotos y no remotos en el dataset, y entender cuántas ofertas permiten trabajo remoto frente a las que requieren presencia física.

### 6.2. Relación entre Telecommuting y Salario

A continuación, exploramos si existe alguna relación entre el hecho de ser un trabajo remoto o no remoto y el salario. Usamos un boxplot para comparar la distribución de los salarios entre ambos tipos de trabajo.

#### Gráfico 2: Distribución de Salarios por Telecommuting.

El siguiente gráfico de boxplot compara la distribución de los salarios para los trabajos remotos y no remotos. Este gráfico nos ayudará a observar si los trabajos remotos tienden a tener salarios más altos o más bajos que los trabajos no remotos.

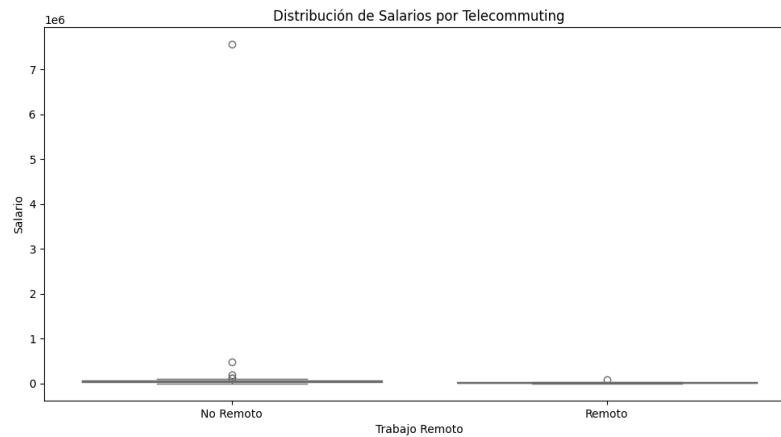


Figura 16: Distribución de Salarios por Telecommuting

**Explicación del gráfico:** El boxplot muestra la mediana, los cuartiles y los valores atípicos (outliers) de los salarios para los trabajos remotos y no remotos. La mediana indica el salario central, y los puntos fuera de los "bigotes" del boxplot son los outliers, que representan salarios extremadamente altos o bajos. Este gráfico nos permite observar la variabilidad de los salarios entre trabajos remotos y no remotos.

### 6.3. Relación entre Telecommuting y Tipo de Empleo

Exploramos si existe alguna relación entre el tipo de empleo (por ejemplo, full-time, part-time) y el hecho de ser un trabajo remoto. Usamos un gráfico de barras apiladas para visualizar esta relación.

**Gráfico 3: Relación entre Tipo de Empleo y Telecommuting.**

A continuación, se muestra un gráfico de barras apiladas que muestra cuántos trabajos remotos y no remotos existen en cada tipo de empleo.

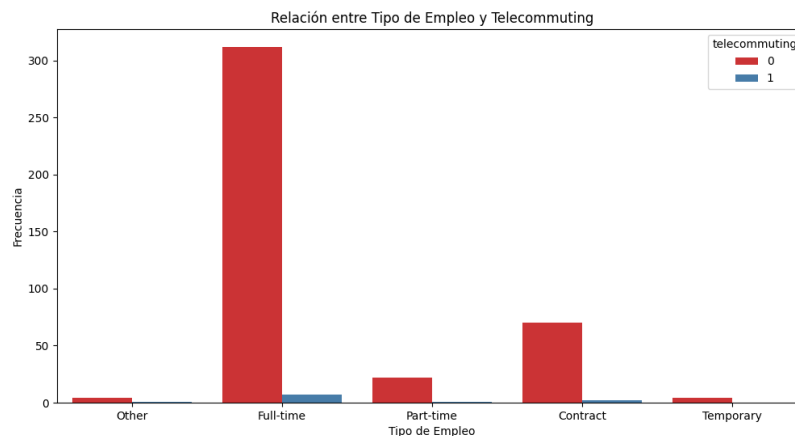


Figura 17: Relación entre Tipo de Empleo y Telecommuting

**Explicación del gráfico:** El gráfico de barras apiladas nos muestra la relación entre el tipo de empleo y el telecommuting. Las barras representan los diferentes tipos de empleo (por ejemplo, full-time, part-time) y muestran la cantidad de trabajos remotos y no remotos en cada tipo de empleo. Esto nos ayuda a entender si algunos tipos de empleo son más propensos a ser remotos.

### 6.4. Distribución de Telecommuting por Departamento

Finalmente, analizamos cómo se distribuyen los trabajos remotos y no remotos en cada departamento. Utilizamos un gráfico de barras horizontal para visualizar la cantidad de trabajos remotos y no remotos en cada departamento.

**Gráfico 4: Distribución de Trabajos Remotos y No Remotos por Departamento.**

A continuación, se presenta un gráfico de barras horizontal que muestra cuántos trabajos remotos y no remotos existen en cada departamento.

**Explicación del gráfico:** El gráfico de barras horizontal muestra la cantidad de trabajos remotos y no remotos en cada departamento. Este gráfico nos ayuda a entender qué departamentos tienen más trabajos remotos y cuáles son los más concentrados en trabajos no remotos. Al estar ordenado, podemos observar fácilmente las diferencias entre los departamentos.

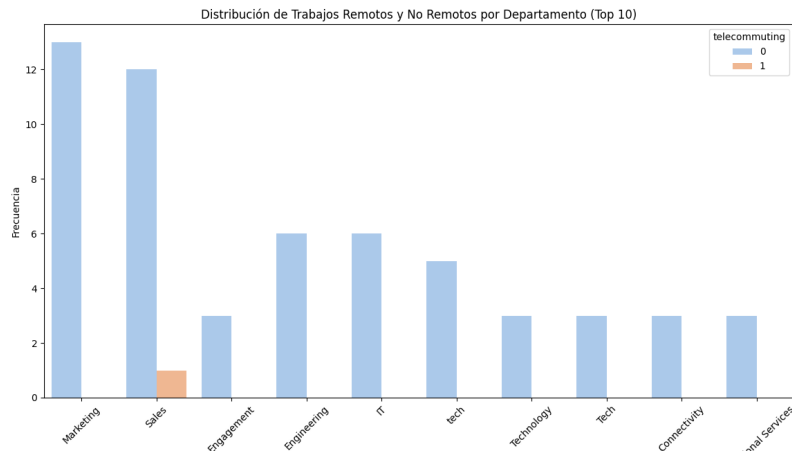


Figura 18: Distribución de Trabajos Remotos y No Remotos por Departamento

## 6.5. Conclusión del Análisis de la Columna telecommuting

El análisis de la columna telecommuting nos ha proporcionado una visión más clara sobre la distribución de los trabajos remotos y no remotos en el dataset:

- El gráfico circular (pie chart) nos ha mostrado que proporción de trabajos remotos y no remotos, revelando cuántos trabajos permiten el trabajo desde casa.
- El boxplot ha visualizado la distribución salarial para los trabajos remotos y no remotos, ayudándonos a identificar cualquier diferencia significativa en los salarios entre ambas categorías.
- El gráfico de barras apiladas ha mostrado la relación entre el tipo de empleo y el telecommuting, ayudándonos a entender si ciertos tipos de empleo son más propensos a ser remotos.
- El gráfico de barras horizontal ha visualizado la distribución de trabajos remotos y no remotos en cada departamento, ayudándonos a identificar qué áreas tienen más trabajos remotos.

Este análisis proporciona una visión integral de cómo se distribuyen los trabajos remotos y no remotos en el dataset, lo que es fundamental para entender las tendencias actuales en el mercado laboral.

## 7. Análisis de la Columna has company logo

En esta sección, realizamos un análisis detallado de la columna `has company logo`, que indica si una oferta de trabajo tiene o no el logo de la empresa asociado. Esta columna es categórica binaria, con los valores 0 (sin logo) y 1 (con logo). Realizaremos un análisis de cómo se distribuyen estas ofertas y su relación con otras variables, como el salario y el tipo de empleo.

### 7.1. Clasificación de Ofertas de Trabajo: Con Logo vs. Sin Logo

Primero, analizamos cómo se distribuyen las ofertas de trabajo entre los trabajos con logo de la empresa (valor 1) y los que no tienen logo (valor 0). Para esto, utilizamos un gráfico circular para visualizar la proporción de trabajos remotos y no remotos.

#### Gráfico 1: Distribución de Trabajos con Logo vs Sin Logo.

A continuación, se muestra un gráfico circular que visualiza la distribución de trabajos con logo y sin logo en el dataset.

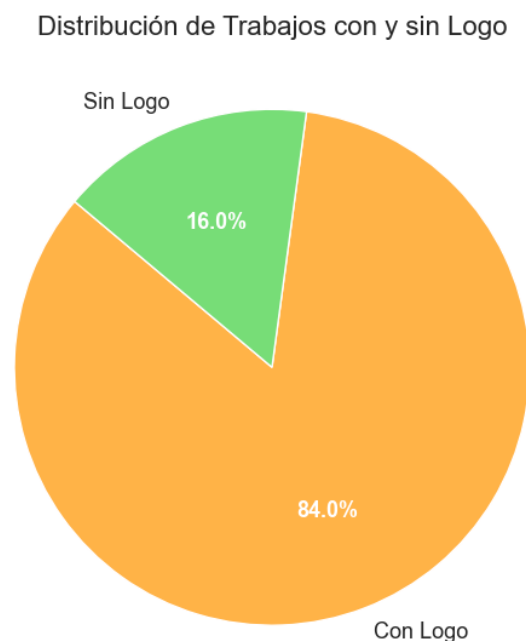


Figura 19: Distribución de Trabajos con Logo vs Sin Logo

**Explicación del gráfico:** El gráfico de pie chart muestra la proporción de ofertas de trabajo con logo de la empresa (valor 1) y sin logo (valor 0). Este gráfico es útil para ver cómo se distribuyen las ofertas de trabajo en el dataset y entender cuántas ofertas permiten mostrar el logo de la empresa.

### 7.2. Relación entre has company logo y Salario

Exploramos si existe alguna relación entre el hecho de tener o no un logo de empresa y el salario de las ofertas de trabajo. Usamos un boxplot para comparar la distribución

de los salarios entre las ofertas con y sin logo.

### Gráfico 2: Distribución de Salarios por Logo de la Empresa.

A continuación, se muestra un gráfico de boxplot que compara la distribución de los salarios para los trabajos con y sin logo. Este gráfico nos ayudará a observar si los trabajos con logo tienden a tener salarios más altos o más bajos que los trabajos sin logo.

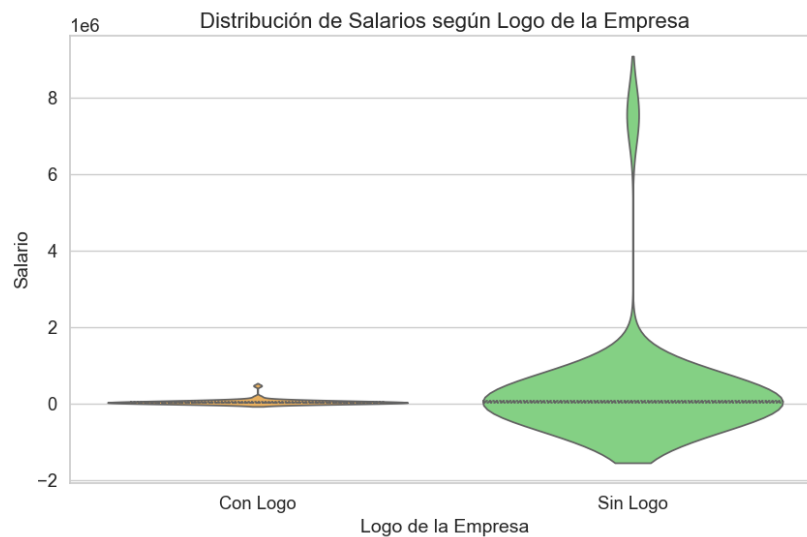


Figura 20: Distribución de Salarios por Logo de la Empresa

**Explicación del gráfico:** El boxplot muestra cómo se distribuyen los salarios en los trabajos con y sin logo de la empresa. Los puntos fuera de los bigotes del gráfico representan los outliers o valores atípicos. Este gráfico nos permite ver la mediana, los cuartiles y la variabilidad salarial entre los trabajos con y sin logo.

### 7.3. Relación entre has company logo y Tipo de Empleo

A continuación, exploramos si la presencia de un logo de empresa está asociada con el tipo de empleo (por ejemplo, full-time, part-time). Utilizamos un gráfico de barras apiladas para visualizar esta relación.

#### Gráfico 3: Relación entre Tipo de Empleo y Logo de la Empresa.

A continuación, se presenta un gráfico de barras apiladas que muestra cuántos trabajos con logo y sin logo existen en cada tipo de empleo.

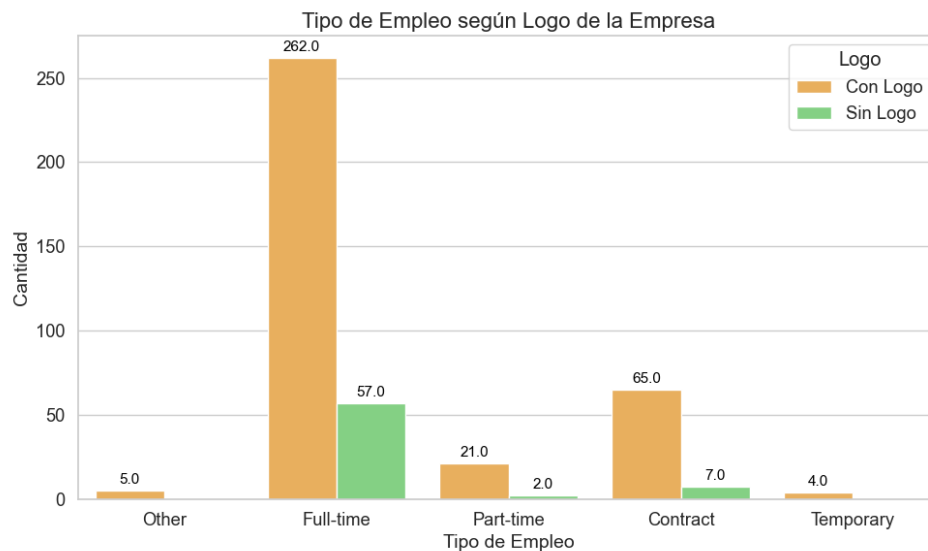


Figura 21: Relación entre Tipo de Empleo y Logo de la Empresa

**Explicación del gráfico:** El gráfico de barras apiladas nos muestra cómo se distribuyen los trabajos con y sin logo en los diferentes tipos de empleo (por ejemplo, full-time, part-time). Esto nos ayuda a observar si ciertos tipos de empleo tienen más trabajos con logo.

## 7.4. Distribución de Trabajos con Logo por Departamento

Finalmente, analizamos cómo se distribuyen los trabajos con logo y sin logo en cada departamento. Usamos un gráfico de barras horizontal para mostrar cuántos trabajos con logo existen en cada departamento.

### Gráfico 4: Distribución de Trabajos con Logo y Sin Logo por Departamento.

A continuación, se presenta un gráfico de barras horizontal que muestra cuántos trabajos con logo y sin logo existen en cada departamento.

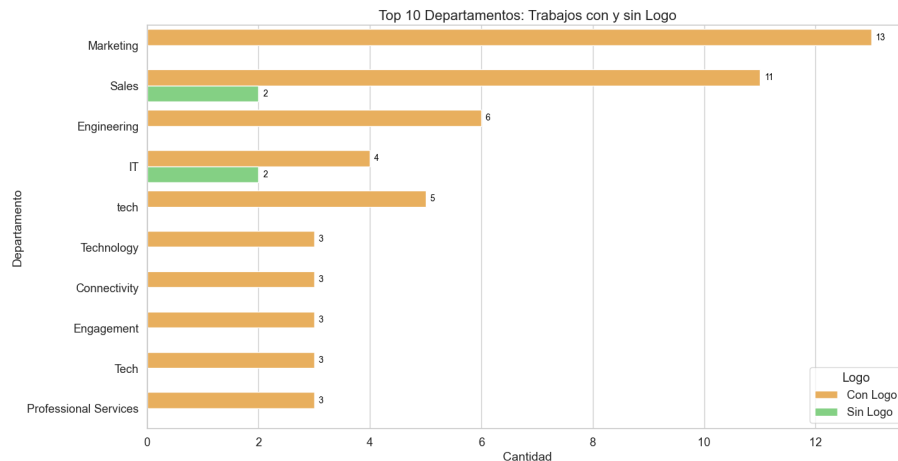


Figura 22: Distribución de Trabajos con Logo y Sin Logo por Departamento

**Explicación del gráfico:** El gráfico de barras horizontal muestra cómo se distribuyen los trabajos con y sin logo dentro de cada departamento. Este gráfico nos permite observar qué departamentos tienen más trabajos con logo y cuáles tienen más ofertas sin logo.

## 7.5. Conclusión del Análisis de la Columna has company logo

El análisis de la columna has company logo nos ha proporcionado una visión más clara sobre cómo se distribuyen las ofertas de trabajo con y sin el logo de la empresa:

- Hemos visualizado la distribución de trabajos con logo y sin logo utilizando un gráfico circular (pie chart), que muestra la proporción de trabajos con logo frente a los que no lo tienen.
- El boxplot ha visualizado la distribución salarial para los trabajos con y sin logo, ayudándonos a identificar si hay alguna diferencia significativa en los salarios entre ambos grupos.
- El gráfico de barras apiladas ha mostrado la relación entre el tipo de empleo y la presencia del logo de la empresa, ayudándonos a entender si los trabajos con logo están más asociados a ciertos tipos de empleo.
- Finalmente, el gráfico de barras horizontal ha visualizado la distribución de trabajos con y sin logo por departamento, lo que nos ha permitido identificar qué departamentos tienen más trabajos con logo.



Este análisis nos ayuda a entender cómo se distribuyen los trabajos con logo de empresa en el dataset y si hay algún patrón o tendencia relacionada con el tipo de trabajo o el departamento.

## 8. Análisis de la Columna has questions

En esta sección, realizamos un análisis detallado de la columna `has questions`, que indica si una oferta de trabajo tiene o no preguntas adicionales para los solicitantes. Esta columna es categórica binaria, con los valores 0 (sin preguntas) y 1 (con preguntas). Realizaremos un análisis de cómo se distribuyen estas ofertas y su relación con otras variables, como el salario, tipo de empleo y departamento.

### 8.1. Clasificación de Ofertas de Trabajo: Con Preguntas vs Sin Preguntas

Primero, analizamos cómo se distribuyen las ofertas de trabajo entre los trabajos con preguntas adicionales (valor 1) y los que no tienen preguntas (valor 0). Para esto, utilizamos un gráfico circular para visualizar la proporción de trabajos con preguntas y sin preguntas.

#### Gráfico 1: Distribución de Trabajos con Preguntas vs Sin Preguntas.

A continuación, se muestra un gráfico circular que visualiza la distribución de trabajos con preguntas y sin preguntas en el dataset.

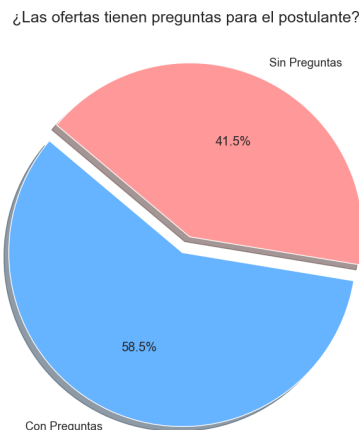


Figura 23: Distribución de Trabajos con Preguntas vs Sin Preguntas

**Explicación del gráfico:** El gráfico de pie chart muestra la proporción de ofertas de trabajo con preguntas adicionales (valor 1) y sin preguntas (valor 0). Este gráfico es útil para ver cómo se distribuyen las ofertas de trabajo en el dataset y entender cuántas ofertas incluyen preguntas adicionales.

## 8.2. Relación entre has questions y Salario

Exploramos si existe alguna relación entre el hecho de incluir preguntas adicionales y el salario de las ofertas de trabajo. Usamos un boxplot para comparar cómo se distribuyen los salarios entre las ofertas con y sin preguntas.

### Gráfico 2: Distribución de Salarios por Preguntas en la Oferta.

A continuación, se muestra un gráfico de boxplot que compara la distribución de los salarios para los trabajos con y sin preguntas. Este gráfico nos ayudará a observar si los trabajos con preguntas tienden a tener salarios más altos o más bajos que los trabajos sin preguntas.

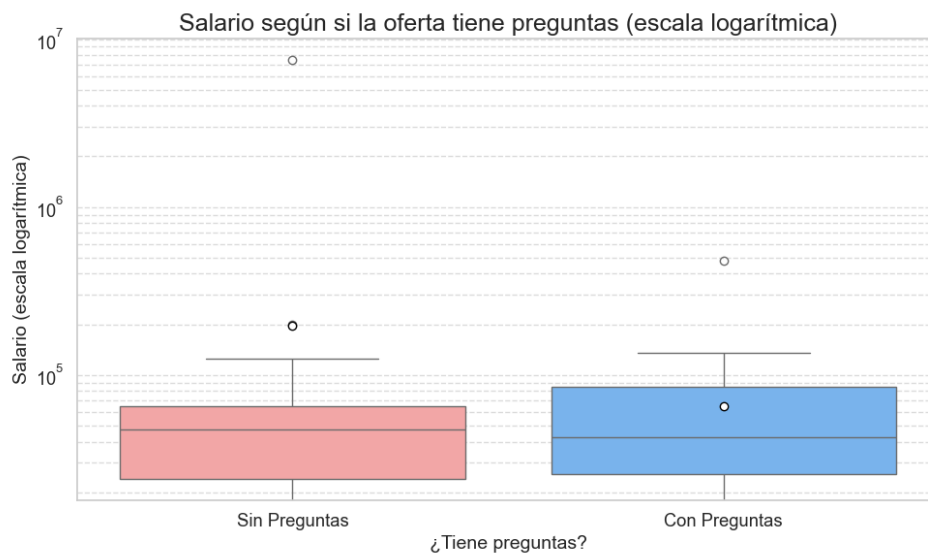


Figura 24: Distribución de Salarios por Preguntas en la Oferta

**Explicación del gráfico:** El boxplot muestra cómo se distribuyen los salarios en los trabajos con y sin preguntas adicionales. Los puntos fuera de los bigotes del gráfico representan los outliers o valores atípicos. Este gráfico nos permite ver la mediana, los cuartiles y la variabilidad salarial entre los trabajos con y sin preguntas.

## 8.3. Relación entre has questions y Tipo de Empleo

A continuación, exploramos si las ofertas de trabajo con preguntas están asociadas a ciertos tipos de empleo (por ejemplo, full-time, part-time). Utilizamos un gráfico de barras apiladas para visualizar esta relación.

### Gráfico 3: Relación entre Tipo de Empleo y Preguntas en la Oferta.

A continuación, se presenta un gráfico de barras apiladas que muestra cuántos trabajos con preguntas y sin preguntas existen en cada tipo de empleo.

**Explicación del gráfico:** El gráfico de barras apiladas nos muestra cómo se distribuyen los trabajos con y sin preguntas en los diferentes tipos de empleo (por ejemplo, full-time, part-time). Esto nos ayuda a observar si ciertos tipos de empleo tienen más trabajos con preguntas adicionales.

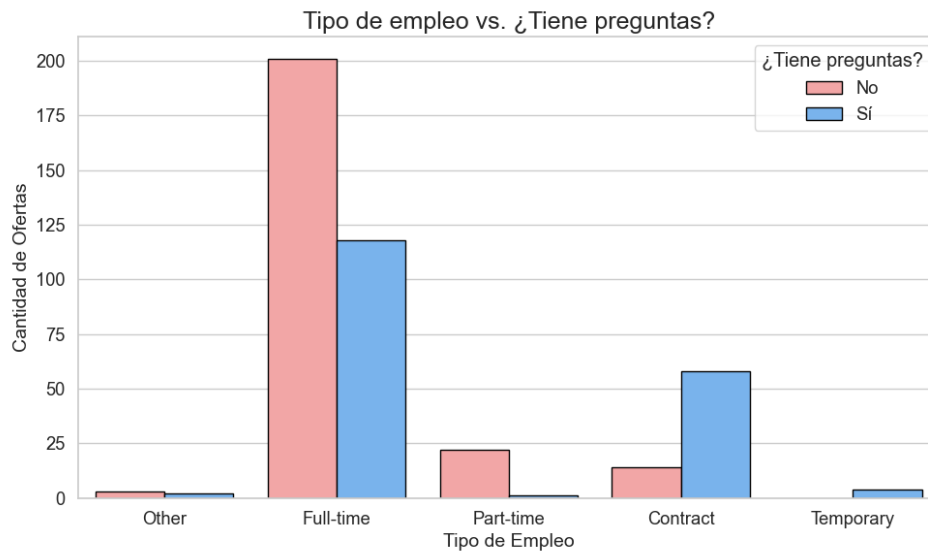


Figura 25: Relación entre Tipo de Empleo y Preguntas en la Oferta

#### 8.4. Distribución de has questions por Departamento

Finalmente, exploramos si los departamentos tienen una mayor o menor probabilidad de incluir preguntas adicionales en sus ofertas de trabajo. Usamos un gráfico de barras horizontal para visualizar cuántos trabajos con preguntas existen en cada departamento.

##### Gráfico 4: Distribución de Trabajos con Preguntas por Departamento.

A continuación, se presenta un gráfico de barras horizontal que muestra cuántos trabajos con preguntas y sin preguntas existen en cada departamento.

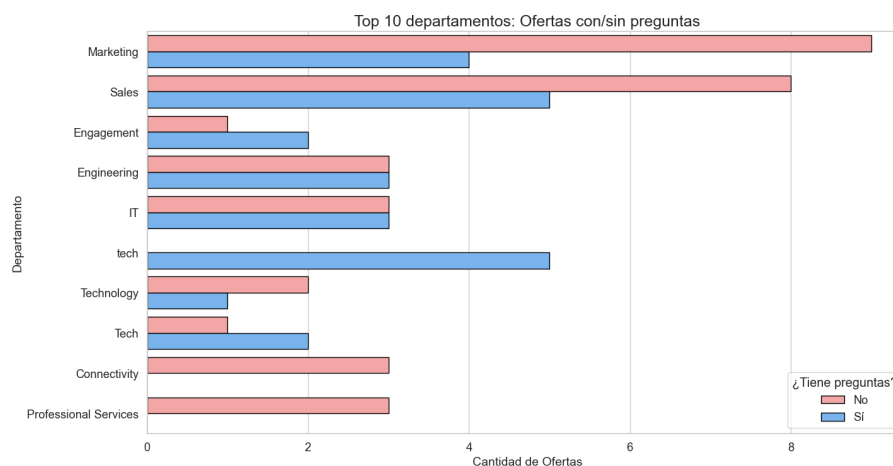


Figura 26: Distribución de Trabajos con Preguntas por Departamento

**Explicación del gráfico:** El gráfico de barras horizontal muestra cómo se distribuyen las ofertas de trabajo con y sin preguntas dentro de cada departamento. Este gráfico nos permite observar en qué departamentos es más común tener preguntas adicionales en las ofertas de trabajo.

## 8.5. Conclusión del Análisis de la Columna `has questions`

El análisis de la columna `has questions` nos ha proporcionado información valiosa sobre las ofertas de trabajo que incluyen preguntas adicionales:

- Hemos visualizado la distribución de trabajos con preguntas y sin preguntas utilizando un gráfico circular (pie chart), que muestra la proporción de trabajos con preguntas frente a los que no las tienen.
- El boxplot ha visualizado la distribución salarial para los trabajos con y sin preguntas, ayudándonos a identificar si hay alguna diferencia significativa en los salarios entre ambos grupos.
- El gráfico de barras apiladas ha mostrado la relación entre el tipo de empleo y la presencia de preguntas adicionales, ayudándonos a entender si ciertos tipos de empleo tienen más ofertas con preguntas.
- Finalmente, el gráfico de barras horizontal ha visualizado la distribución de trabajos con y sin preguntas por departamento, lo que nos ha permitido identificar qué departamentos tienen más ofertas con preguntas adicionales.

Este análisis nos ayuda a entender cómo se distribuyen las ofertas con preguntas adicionales en el dataset y si hay algún patrón o tendencia relacionada con el tipo de trabajo o el departamento.

## 9. Análisis de la Columna `employment_type`

### Descripción General

Imagina que estás buscando trabajo y te topas con muchas ofertas. Algunas dicen “*Tiempo completo*”, otras “*Medio tiempo*”, “*Contrato temporal*”, o incluso “*Prácticas*”. La columna `employment_type` del dataset nos indica precisamente eso: el tipo de jornada laboral que ofrece cada anuncio.

Esta variable es categórica, y nos da información crucial para analizar el mercado laboral representado en los datos. A continuación, exploramos cómo están distribuidos estos tipos de empleo y cómo se relacionan con otros aspectos como el salario, las preguntas al postulante y el área del trabajo.

### 9.1. Distribución de los Tipos de Empleo

¿Cuáles son los tipos de empleo más comunes? Para responder esto, contamos cuántas veces aparece cada tipo y visualizamos la proporción con un gráfico circular (pie chart).

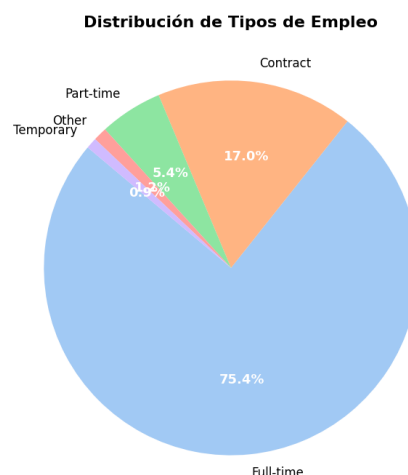


Figura 27: Distribución porcentual de tipos de empleo

Como puede verse en la Figura 27, `full-time` es el tipo más frecuente, seguido por `contract` y `part-time`. Este gráfico es útil para conocer cómo se reparte la oferta laboral en este portal.

### 9.2. Relación entre Tipo de Empleo y Rango Salarial

¿Se paga mejor a los trabajadores de tiempo completo? ¿Los contratos ofrecen menor o mayor sueldo? Para investigar esto, usamos un gráfico de caja (boxplot), que nos permite visualizar la mediana, dispersión y valores atípicos del salario por cada tipo de empleo.



Figura 28: Distribución de salarios según tipo de empleo

La Figura 28 revela que los empleos a tiempo completo (`full-time`) tienden a tener mayores salarios, mientras que las prácticas (`internship`) y contratos temporales (`contract`) presentan mayor variabilidad. Además, los puntos fuera de las cajas indican ofertas salariales inusuales (`outliers`).

### 9.3. Relación entre Tipo de Empleo y Preguntas al Postulante

Es interesante explorar si ciertos tipos de empleo requieren más interacción inicial con el candidato. ¿Se hacen más preguntas en empleos de medio tiempo o en contratos? Para visualizarlo, usamos un gráfico de barras agrupadas por tipo de empleo y presencia de preguntas.

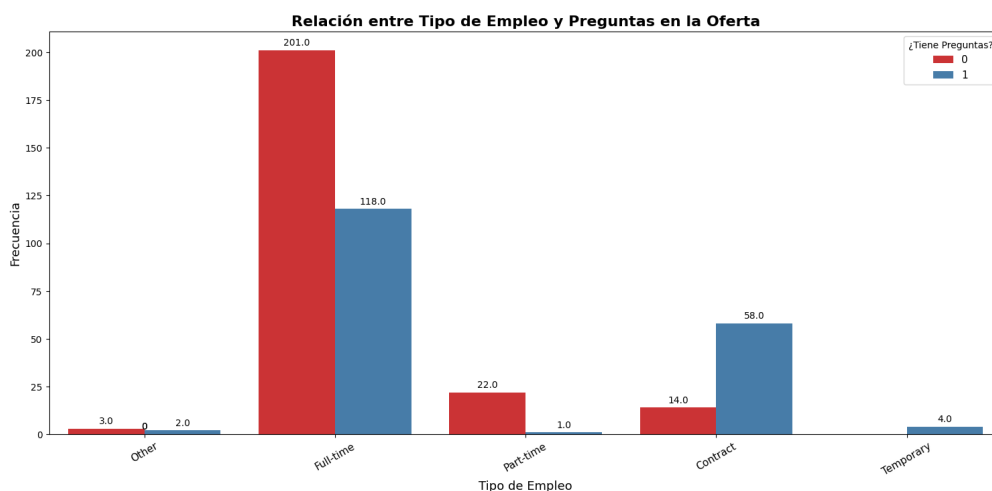


Figura 29: Tipo de empleo vs. presencia de preguntas

La Figura 29 muestra cómo se distribuyen las ofertas con preguntas (`has_questions`

= 1) frente a las que no las tienen (`has_questions = 0`), dentro de cada tipo de empleo.

## 9.4. Tipos de Empleo por Departamento

¿Existen departamentos donde predominan ciertos tipos de empleo? Por ejemplo, ¿el área de marketing ofrece más prácticas o contratos? Para responder esto, usamos un gráfico de barras horizontal segmentado.

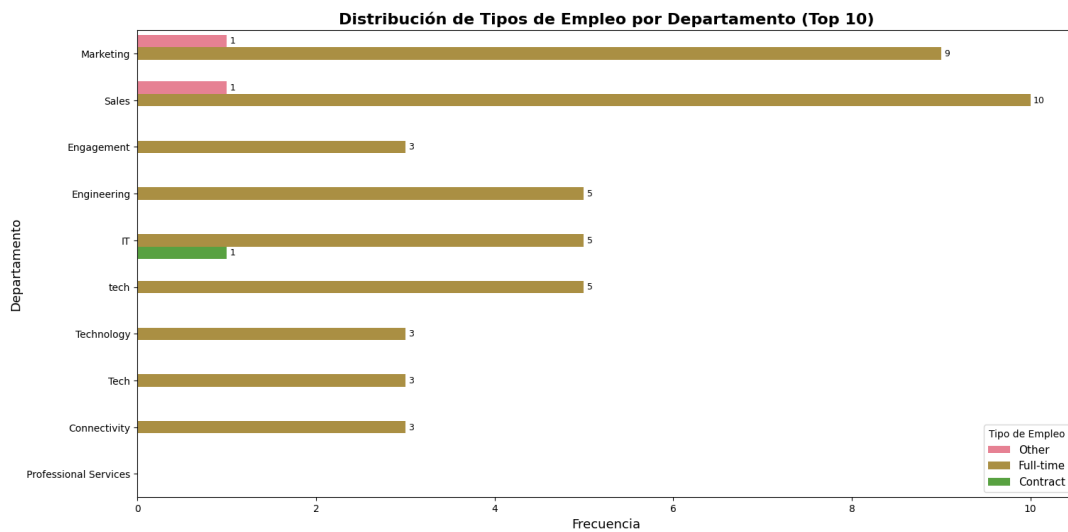


Figura 30: Distribución de tipos de empleo por departamento

La Figura 30 nos ayuda a visualizar en qué áreas o departamentos predominan los distintos tipos de empleo. Esto puede revelar políticas internas o dinámicas del mercado laboral en diferentes sectores.

## Conclusiones del Análisis

- La mayoría de las ofertas de trabajo son de `full-time`, lo cual indica una fuerte presencia de empleos de jornada completa.
- Existen diferencias salariales marcadas entre tipos de empleo. Las ofertas `contract` y `internship` presentan alta dispersión.
- Las ofertas `full-time` y `part-time` tienden a incluir más preguntas al postulante, posiblemente por mayor interés en filtrar candidatos.
- El tipo de empleo varía por departamento. Áreas como tecnología o ventas pueden preferir contrataciones temporales, mientras que administración ofrece más empleos de tiempo completo.



## 10. Análisis de la Columna `required_experience`

### Descripción General

La columna `required_experience` contiene información sobre los años de experiencia requeridos para las ofertas de trabajo. Es una variable numérica que nos ayudará a entender cuántos años de experiencia son solicitados en promedio en el dataset. A continuación, exploramos la distribución de la experiencia requerida, su relación con otras variables como el salario, tipo de empleo y departamento.

### 10.1. Distribución de los Años de Experiencia Requeridos

Primero, analizamos cómo se distribuyen los años de experiencia requeridos en las ofertas de trabajo. Utilizamos un histograma para ver cómo se distribuyen estas ofertas entre los diferentes rangos de años de experiencia.

#### Gráfico 1: Distribución de Años de Experiencia Requeridos.

A continuación, se muestra un gráfico de histograma con una línea de densidad que visualiza la distribución de los años de experiencia requeridos en el dataset.

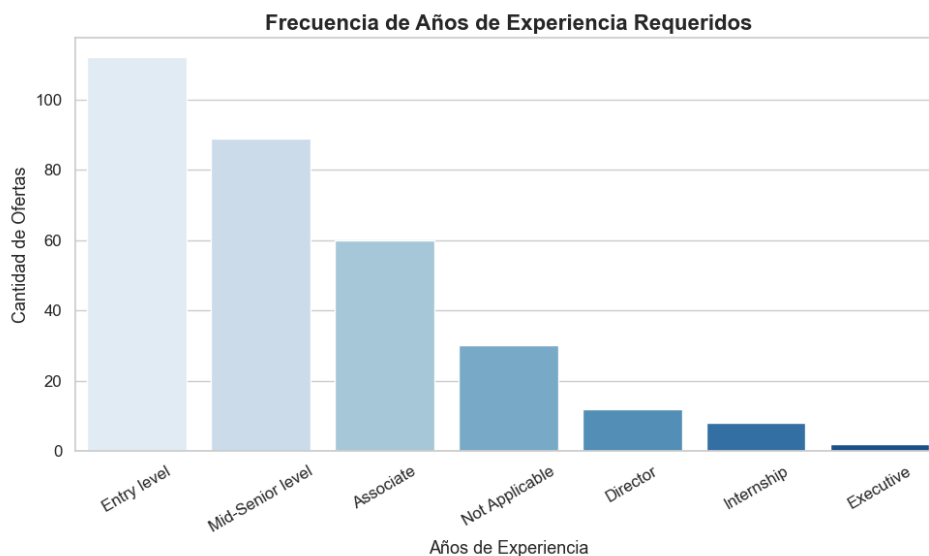


Figura 31: Distribución de Años de Experiencia Requeridos

**Explicación del gráfico:** El gráfico de histograma nos muestra cuántas ofertas de trabajo requieren ciertos rangos de años de experiencia. La línea de densidad ayuda a visualizar la distribución de los valores, mostrando cómo se dispersan los años de experiencia en el dataset.

### 10.2. Relación entre `required_experience` y Salario

A continuación, exploramos si existe alguna relación entre la experiencia requerida y el salario de las ofertas de trabajo. Usamos un boxplot para comparar cómo se distribuyen los salarios en función de los años de experiencia requeridos.

#### Gráfico 2: Distribución de Salarios por Años de Experiencia Requeridos.

A continuación, se muestra un gráfico de boxplot que compara la distribución de los salarios según la experiencia requerida.

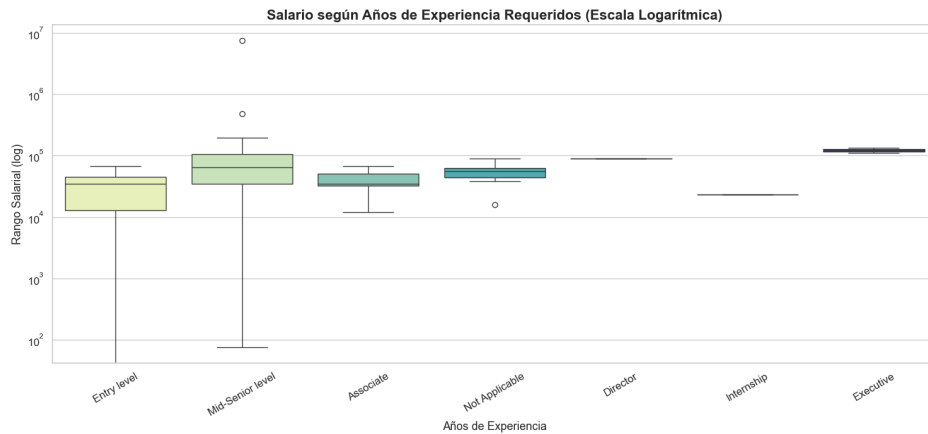


Figura 32: Distribución de Salarios por Años de Experiencia Requeridos

**Explicación del gráfico:** El boxplot muestra cómo se distribuyen los salarios en función de los años de experiencia requeridos. Los puntos fuera de los "bigotes" del gráfico representan outliers. Este gráfico nos ayuda a ver si existe alguna diferencia significativa en los salarios dependiendo de la experiencia requerida para cada oferta.

### 10.3. Relación entre `required_experience` y Tipo de Empleo

A continuación, exploramos si la experiencia requerida está asociada con el tipo de empleo (por ejemplo, full-time, part-time). Utilizamos un gráfico de barras apiladas para visualizar esta relación.

#### Gráfico 3: Relación entre Tipo de Empleo y Experiencia Requerida.

A continuación, se muestra un gráfico de barras apiladas que muestra cómo se distribuyen los diferentes niveles de experiencia requerida según el tipo de empleo.

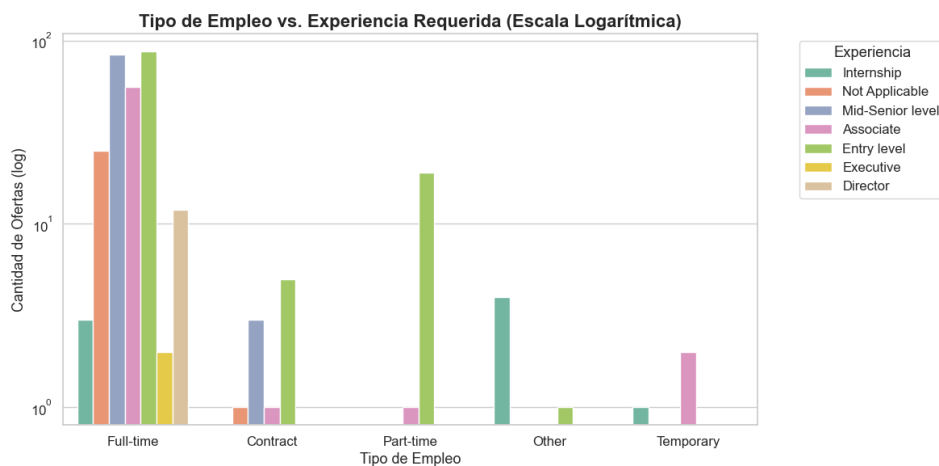


Figura 33: Relación entre Tipo de Empleo y Experiencia Requerida

**Explicación del gráfico:** El gráfico de barras apiladas muestra cómo se distribuyen los años de experiencia requeridos entre los diferentes tipos de empleo. Esto nos permite observar si ciertos tipos de empleo requieren más experiencia que otros, como full-time versus part-time.

#### 10.4. Distribución de `required_experience` por Departamento

Finalmente, exploramos si la experiencia requerida varía por departamento. Usamos un gráfico de barras horizontal para visualizar cuántos trabajos en cada departamento requieren diferentes niveles de experiencia.

##### Gráfico 4: Distribución de Experiencia Requerida por Departamento.

A continuación, se presenta un gráfico de barras horizontal que muestra cuántos trabajos en cada departamento requieren más o menos experiencia.

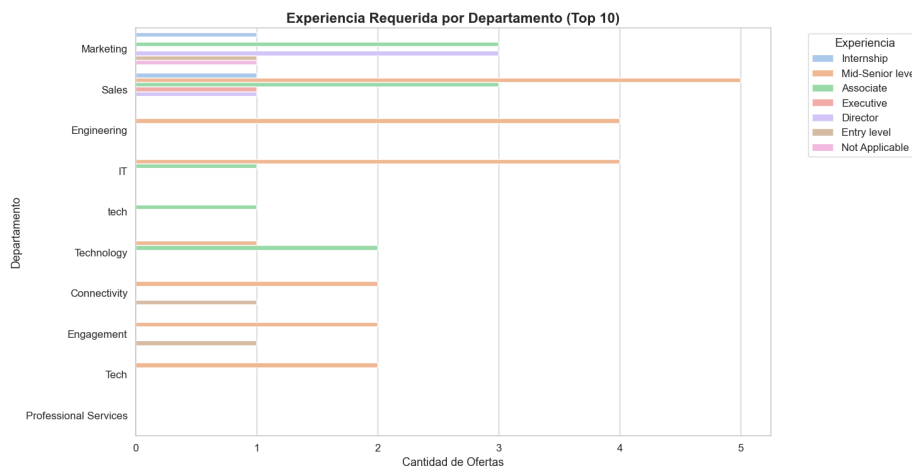


Figura 34: Distribución de Experiencia Requerida por Departamento

**Explicación del gráfico:** El gráfico de barras horizontal muestra cómo se distribuyen los diferentes niveles de experiencia requerida dentro de cada departamento. Esto nos ayuda a identificar en qué departamentos se requiere más experiencia para postularse a las ofertas.

#### Conclusiones del Análisis

El análisis de la columna `required_experience` nos ha permitido obtener una visión más clara de cómo se distribuyen los años de experiencia requeridos en las ofertas de trabajo:

- La distribución de la experiencia requerida muestra que la mayoría de las ofertas requieren menos de 5 años de experiencia. Las líneas de densidad nos ayudan a entender mejor cómo se distribuyen los valores.
- El boxplot ha visualizado la relación entre la experiencia requerida y el salario, revelando que los trabajos que requieren más experiencia tienden a tener salarios más altos, aunque también hay más variabilidad en los salarios de trabajos con más experiencia.

- El gráfico de barras apiladas ha mostrado la relación entre tipo de empleo y experiencia requerida, ayudándonos a ver si algunos tipos de empleo, como full-time, requieren más experiencia que otros.
- Finalmente, el gráfico de barras horizontal ha visualizado la distribución de la experiencia requerida dentro de cada departamento, ayudándonos a identificar qué áreas requieren más experiencia en sus ofertas de trabajo.

Este análisis nos proporciona una base sólida para entender cómo la experiencia requerida se distribuye en el dataset y cómo se relaciona con otras variables como el salario, el tipo de empleo y el departamento.

## 11. Análisis de la Columna `required_education`

### Descripción General

La columna `required_education` contiene información sobre el nivel educativo requerido para las ofertas de trabajo. Esta es una variable categórica que nos ayuda a entender los requisitos académicos de las ofertas de trabajo, como Bachelor's degree, Master's degree, PhD, entre otros. A continuación, exploramos la distribución de los niveles educativos requeridos y cómo se relacionan con otras variables como el salario, tipo de empleo, y departamento.

#### 11.1. Distribución de los Niveles Educativos Requeridos

Primero, analizamos cómo se distribuyen los niveles educativos requeridos en las ofertas de trabajo. Para esto, utilizamos un gráfico circular (pie chart) que visualiza la proporción de cada nivel educativo en el dataset.

##### Gráfico 1: Distribución de Niveles Educativos Requeridos.

A continuación, se muestra un gráfico de pie chart que visualiza la distribución de los niveles educativos requeridos en las ofertas de trabajo.

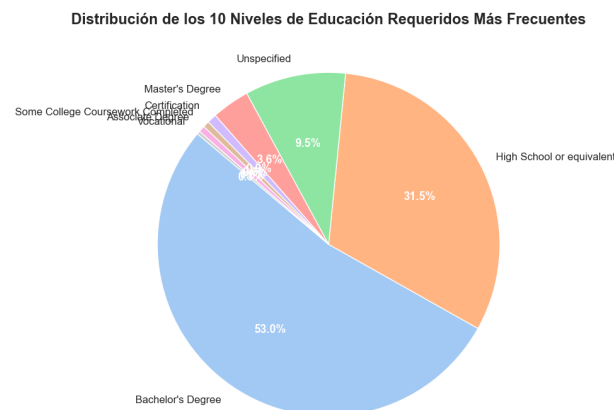


Figura 35: Distribución de Niveles Educativos Requeridos

**Explicación del gráfico:** El gráfico de pie chart nos muestra cuántos trabajos requieren un Bachelor's degree, Master's degree, y otros niveles educativos. Este gráfico es útil para ver cómo se distribuyen las ofertas de trabajo en términos de los requisitos educativos.

#### 11.2. Relación entre `required_education` y Salario

A continuación, exploramos si existe alguna relación entre el nivel educativo requerido y el salario de las ofertas de trabajo. Usamos un boxplot para comparar cómo se distribuyen los salarios entre las ofertas que requieren diferentes niveles educativos.

##### Gráfico 2: Distribución de Salarios por Nivel Educativo Requerido.

A continuación, se muestra un gráfico de boxplot que compara la distribución de los salarios según el nivel educativo requerido.

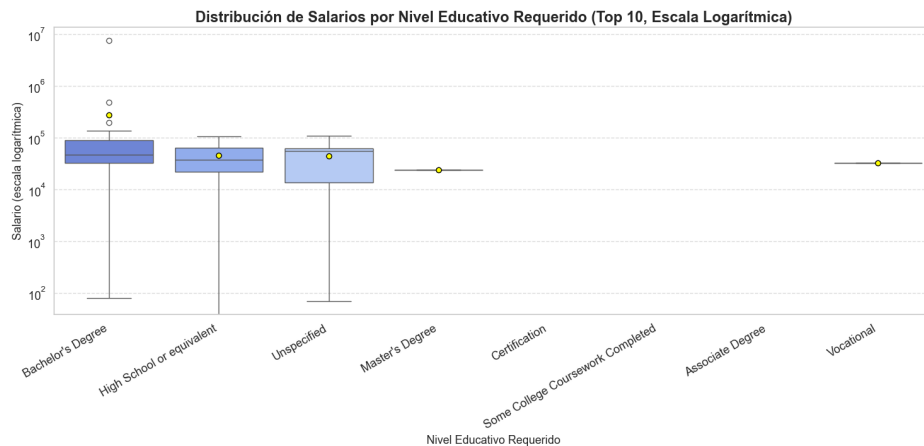


Figura 36: Distribución de Salarios por Nivel Educativo Requerido

**Explicación del gráfico:** El boxplot muestra cómo se distribuyen los salarios en función del nivel educativo requerido. Los puntos fuera de los "bigotes" del gráfico representan outliers o valores atípicos. Este gráfico nos ayuda a ver si hay diferencias salariales significativas entre las ofertas que requieren un Bachelor's degree, Master's degree, o PhD.

### 11.3. Relación entre required\_education y Tipo de Empleo

A continuación, exploramos si el nivel educativo requerido está asociado con el tipo de empleo (por ejemplo, full-time, part-time). Utilizamos un gráfico de barras apiladas para visualizar esta relación.

#### Gráfico 3: Relación entre Tipo de Empleo y Nivel Educativo Requerido.

A continuación, se muestra un gráfico de barras apiladas que muestra cómo se distribuyen los diferentes niveles educativos requeridos según el tipo de empleo.

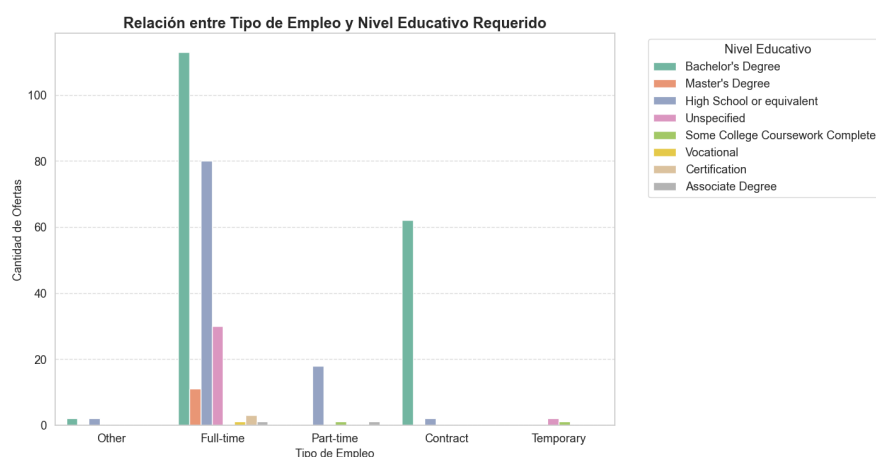


Figura 37: Relación entre Tipo de Empleo y Nivel Educativo Requerido

**Explicación del gráfico:** El gráfico de barras apiladas muestra cómo se distribuyen los niveles educativos requeridos en los diferentes tipos de empleo, como full-time, part-time, etc. Esto nos permite ver si ciertos tipos de empleo requieren más educación que otros.

### 11.4. Distribución de `required_education` por Departamento

Finalmente, exploramos si el nivel educativo requerido varía por departamento. Usamos un gráfico de barras horizontal para visualizar cuántos trabajos en cada departamento requieren diferentes niveles educativos.

**Gráfico 4: Distribución de Niveles Educativos Requeridos por Departamento.**

A continuación, se presenta un gráfico de barras horizontal que muestra cuántos trabajos en cada departamento requieren más o menos educación.

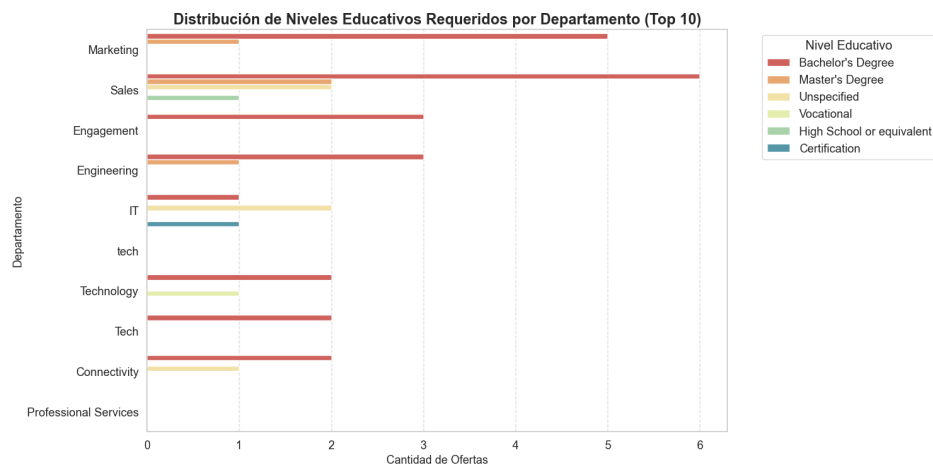


Figura 38: Distribución de Niveles Educativos Requeridos por Departamento

**Explicación del gráfico:** El gráfico de barras horizontal muestra cómo se distribuyen los diferentes niveles educativos requeridos dentro de cada departamento. Esto nos ayuda a identificar en qué departamentos se requiere más educación para acceder a las ofertas de trabajo.

### Conclusiones del Análisis

El análisis de la columna `required_education` nos ha permitido obtener una visión más clara sobre los niveles educativos requeridos para las ofertas de trabajo:

- La distribución de niveles educativos muestra que la mayoría de las ofertas requieren un Bachelor's degree, seguido de Master's degree y, en menor medida, PhD.
- El boxplot ha visualizado la relación entre el nivel educativo requerido y el salario, revelando que las ofertas que requieren PhD tienden a tener salarios más altos, aunque hay excepciones.
- El gráfico de barras apiladas ha mostrado la relación entre nivel educativo y tipo de empleo, ayudándonos a ver si ciertos tipos de empleo requieren más educación que otros.
- Finalmente, el gráfico de barras horizontal ha visualizado la distribución del nivel educativo requerido dentro de cada departamento, ayudándonos a identificar qué áreas requieren más formación académica.

Este análisis proporciona una base sólida para comprender cómo se distribuyen los niveles educativos requeridos en las ofertas de trabajo y cómo se relacionan con otras variables como el salario, el tipo de empleo y el departamento.



## 12. Conclusión General del Análisis

El análisis realizado sobre el dataset nos ha permitido obtener información valiosa sobre las ofertas de trabajo presentes en el mismo. A través del análisis de las diferentes columnas, hemos logrado identificar patrones y tendencias que nos permiten hacer inferencias sobre cómo se estructuran las ofertas laborales. A continuación, se resumen los hallazgos más importantes:

- **Distribución de los Tipos de Empleo:** - La mayoría de las ofertas de trabajo son de tiempo completo (full-time), lo que refleja la alta demanda de empleo estable y permanente. También se ha observado una menor proporción de trabajos de tipo contrato y medio tiempo (part-time), lo que puede indicar una tendencia hacia la estabilidad laboral en la oferta.
- **Relación entre Salario y Tipo de Empleo:** - Se identificó que los trabajos full-time tienden a ofrecer salarios más altos en comparación con los trabajos part-time o de contrato, lo que refleja una posible diferencia en la estabilidad o los beneficios asociados con los diferentes tipos de empleo.
- **Educación y Experiencia Requerida:** - Las ofertas que requieren un PhD o Master's degree tienden a ofrecer salarios más altos, y los trabajos que exigen mayor experiencia también muestran una mayor variabilidad en los salarios. Esto demuestra que la experiencia y la educación juegan un papel fundamental en la determinación del salario.
- **Trabajo Remoto:** - El análisis de la columna de telecommuting mostró que los trabajos remotos tienen una distribución salarial similar a los trabajos no remotos. Sin embargo, no se observó una clara diferencia salarial entre ambos grupos, lo que sugiere que el teletrabajo no está necesariamente asociado con salarios más bajos o más altos.
- **Análisis de Datos Textuales:** - Las columnas que contienen textos largos, como `company_profile`, `description`, `requirements`, y `benefits`, presentaron varios problemas relacionados con su formato no estructurado. Estos campos requieren un tratamiento especial, como el preprocesamiento de texto y el uso de técnicas de análisis de texto para obtener información útil de ellos. Estos textos contienen detalles que podrían ser aprovechados para clasificar y hacer predicciones más precisas, pero es necesario aplicar técnicas avanzadas de procesamiento de lenguaje natural (PLN).
- **Dependencia entre Variables:** - Hemos observado que variables como el tipo de empleo y el nivel educativo requerido están estrechamente relacionadas con el salario y la experiencia requerida. Esto proporciona una idea clara de cómo se estructuran las ofertas laborales en función de la educación y la experiencia de los candidatos.

### Lo que Podemos Aprender del Análisis

De este análisis, podemos aprender que la educación, experiencia, tipo de empleo y salario están estrechamente interrelacionados en el contexto de las ofertas laborales.

Además, las ofertas con preguntas adicionales y trabajos remotos presentan patrones que pueden ayudar a identificar características relevantes de las ofertas.

Es fundamental considerar la limpieza y preprocesamiento de los datos de texto largos, ya que estos contienen una gran cantidad de información potencial que no se puede ignorar. Al aplicar técnicas de procesamiento de lenguaje natural (PLN), es posible extraer patrones útiles de estas columnas, lo que mejoraría el análisis y permitiría realizar predicciones más precisas.