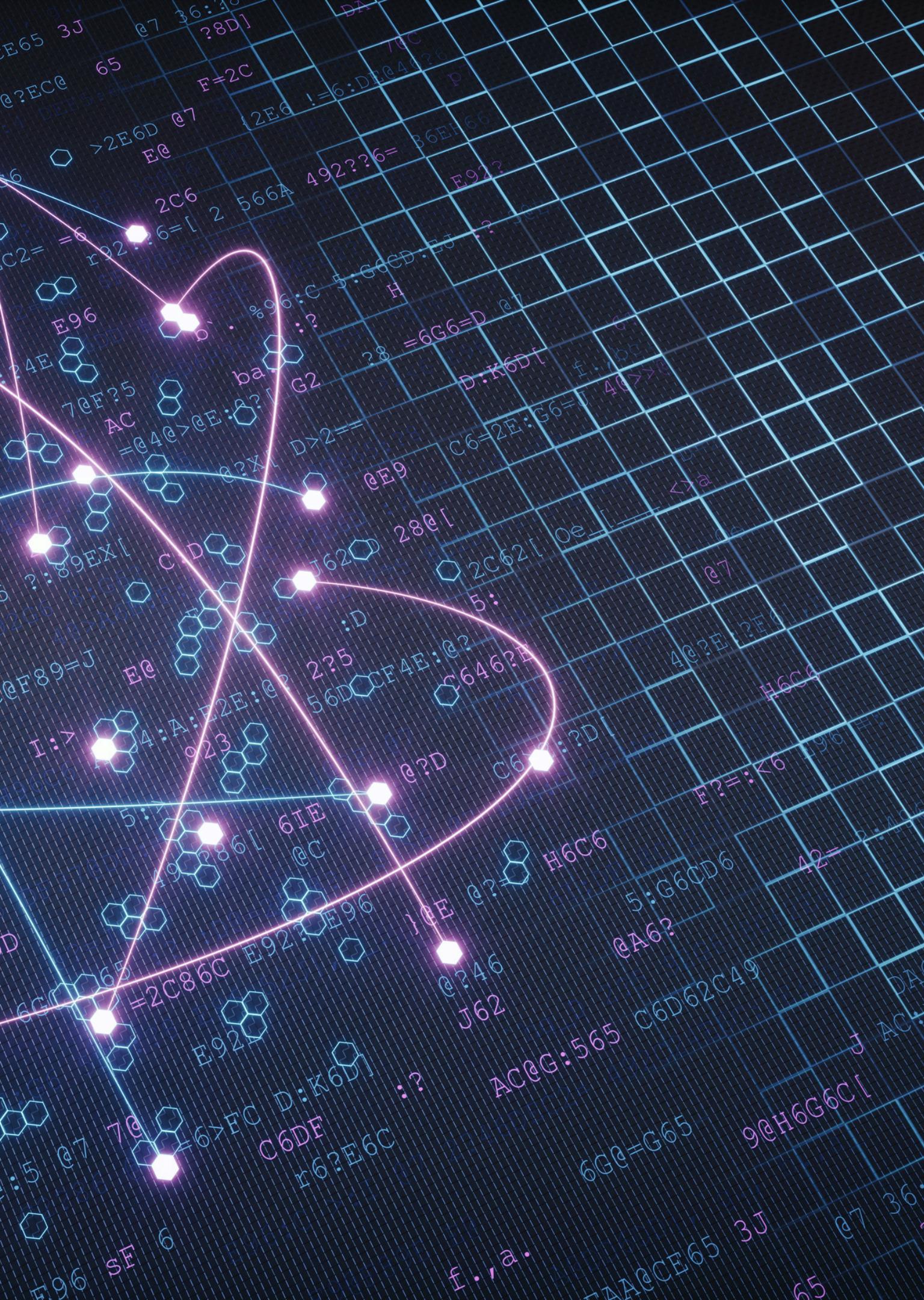


FAKE JOB POSTING

INTEGRANTES:
Jharold Alonso Mayorga
Villena



¿Qué problema identifican en el dataset?



El dataset tiene una desproporción considerable entre las ofertas laborales fraudulentas (1) y las no fraudulentas (0).



Inconsistencias en Ubicaciones



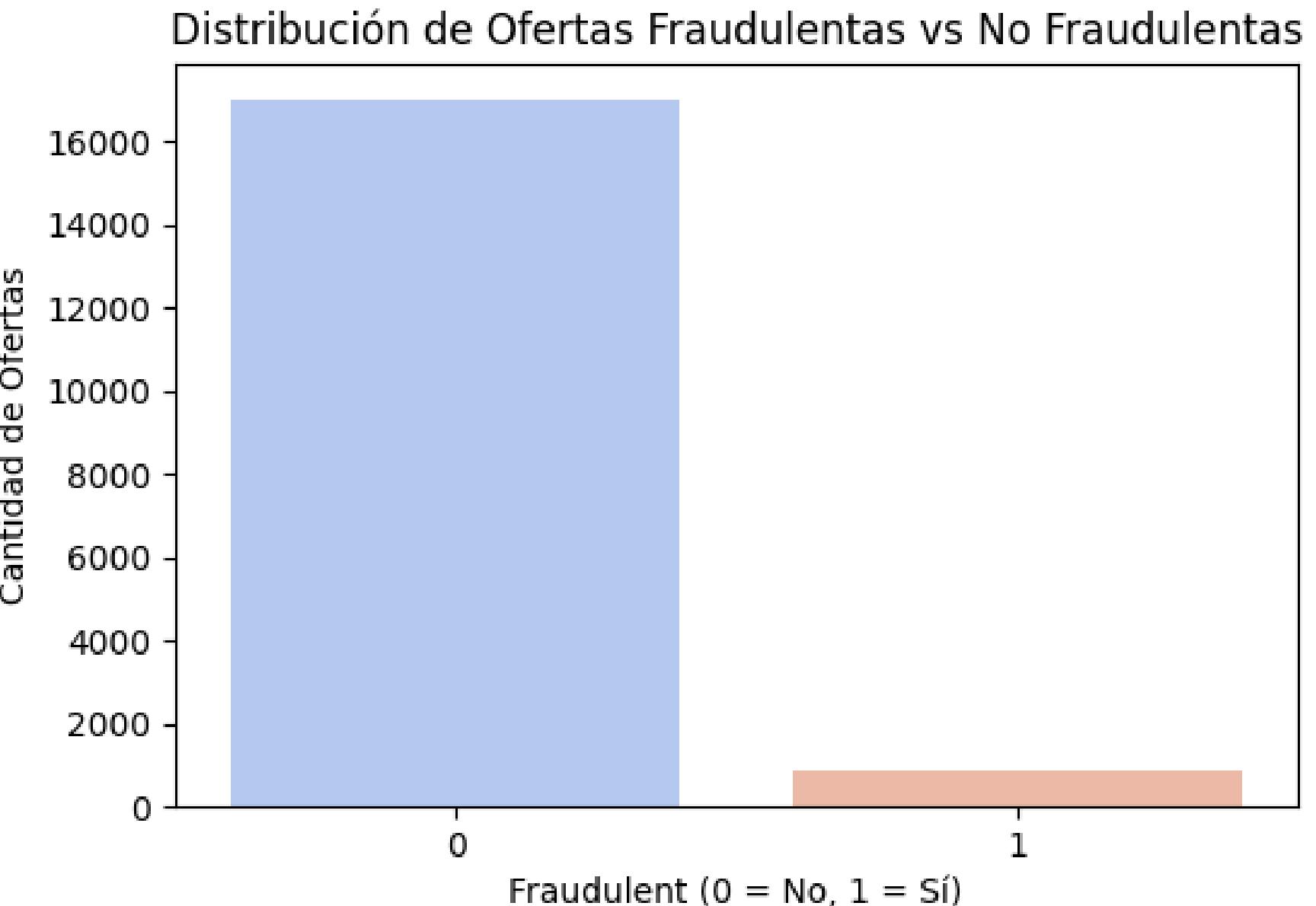
Falta de datos y datos inconsistentes dificultando una análisis profundo

1

Desproporción entre datos clave “fraudulentos”

El dataset tiene una desproporción considerable entre las ofertas laborales fraudulentas (1) y las no fraudulentas (0).

1. **Este desbalance de clases es un problema común en tareas de clasificación y puede requerir técnicas de balanceo de clases**
2. **Esto si no se corrige se puede llevar a un sesgo de la clase mayoritaria.**

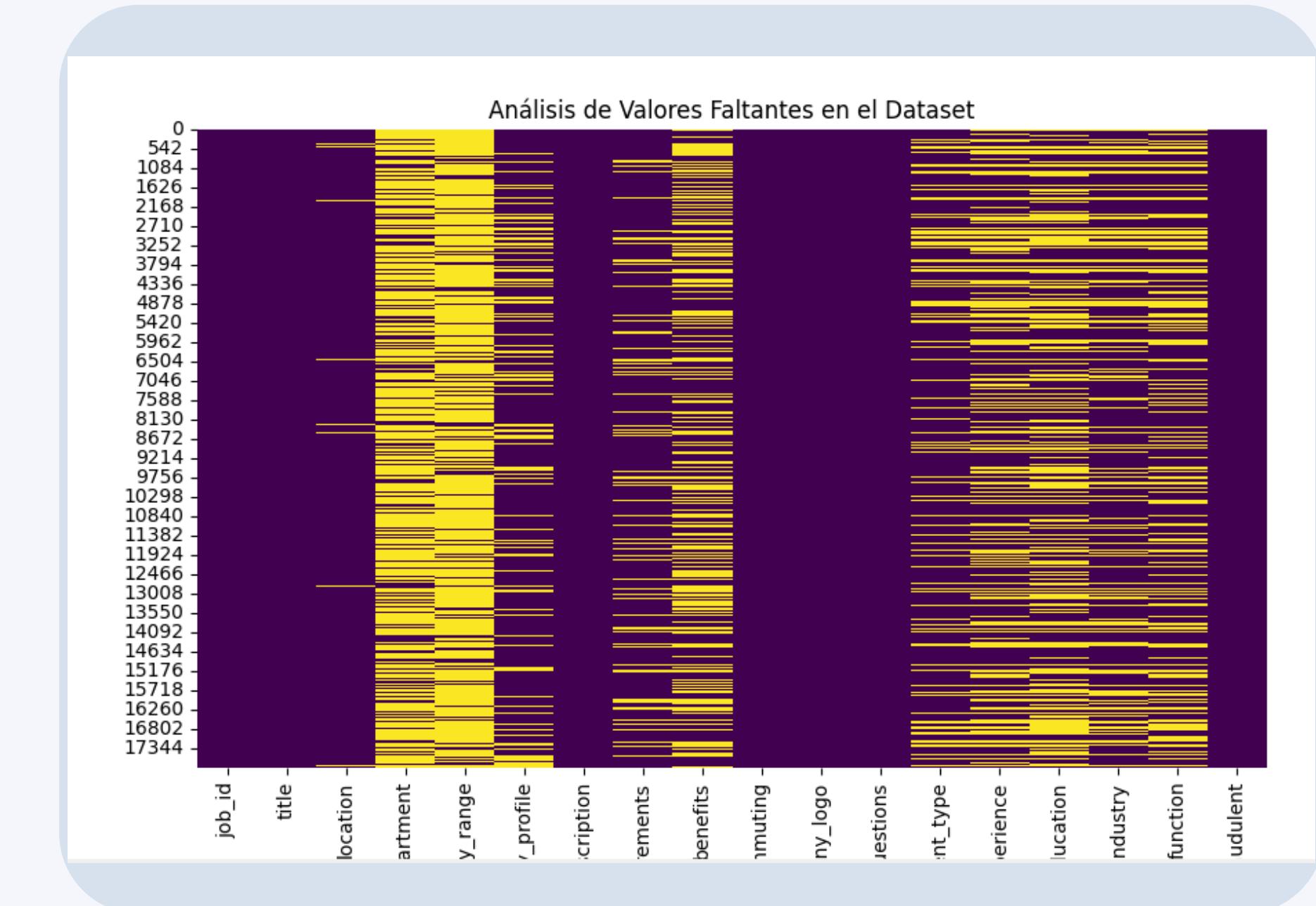


2

Falta de datos

El dataset tiene muchas columnas con valores faltantes, especialmente en las columnas salary_range, department, required_experience, required_education, industry.

- 1. La importancia de manejar estos valores faltantes.**
- 2. La falta de equilibrio de datos en el dataset.**
- 3. Es necesario abordar estos valores faltantes, por ejemplo, mediante imputación (llenar con valores promedio o mediana) o eliminando las filas con datos incompletos si es necesario.**



HUBWAY_STATIONS.CSV

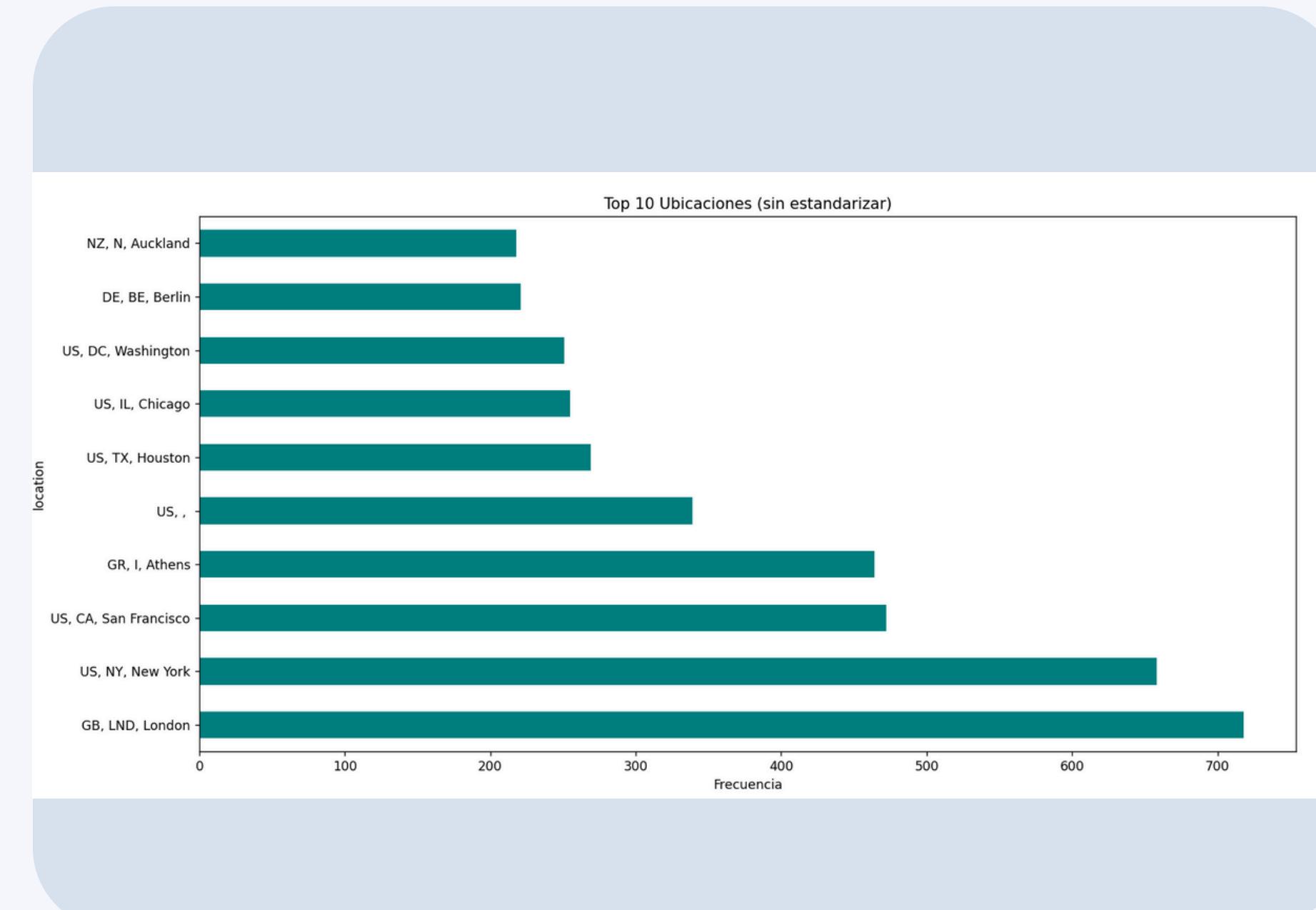
HUBWAY_TRIPS.CSV

3

Inconsistencias en Ubicaciones

La columna location contiene formatos no estandarizados (ej. "New York, NY" vs "NYC").

Existen múltiples formatos para la misma ubicación (ej. 'Remote', 'REMOTE', 'Work From Home'). Se requiere estandarización con expresiones regulares.



HUBWAY_STATIONS.CSV

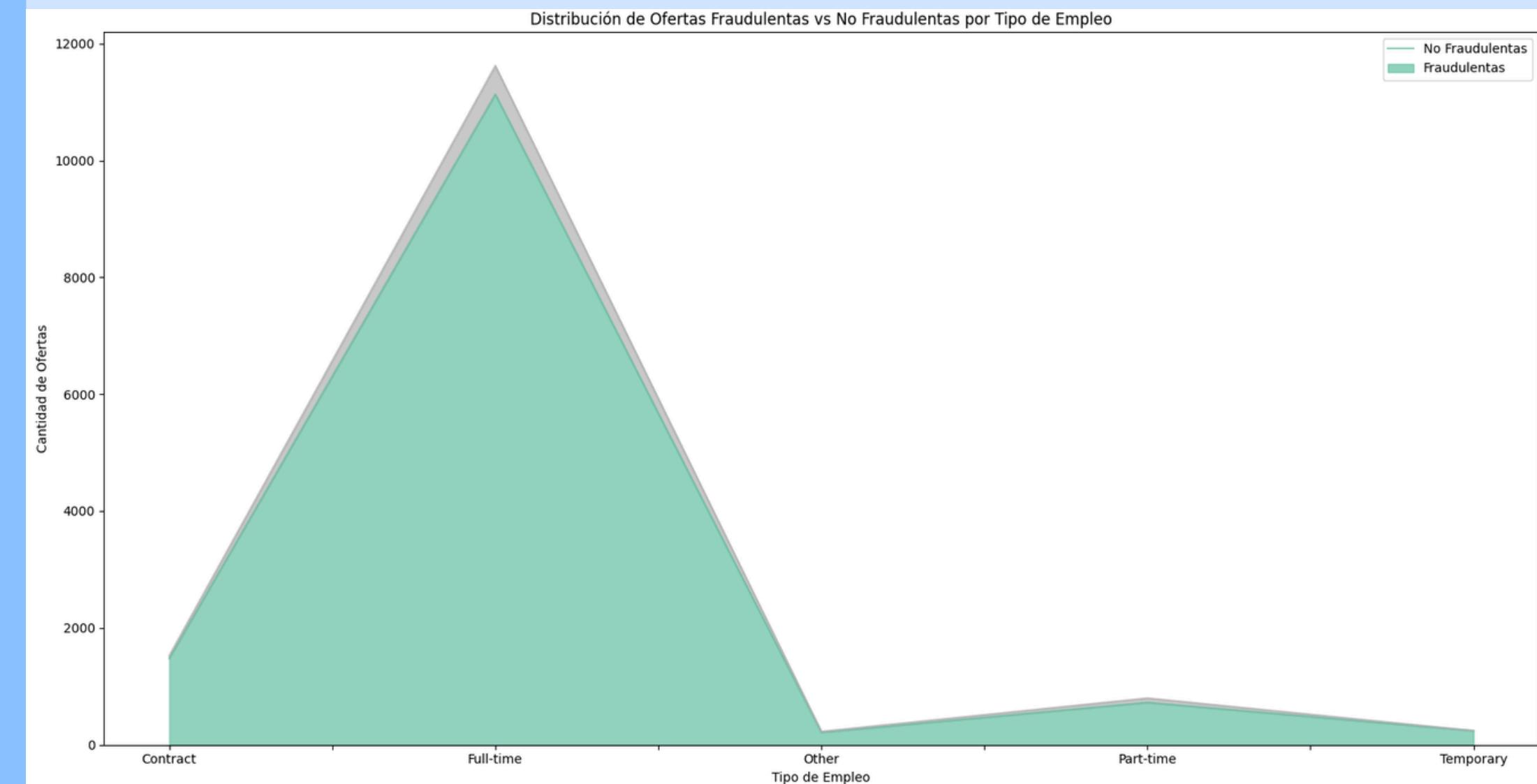
HUBWAY_TRIPS.CSV

¿Qué descubrieron al analizar los datos?

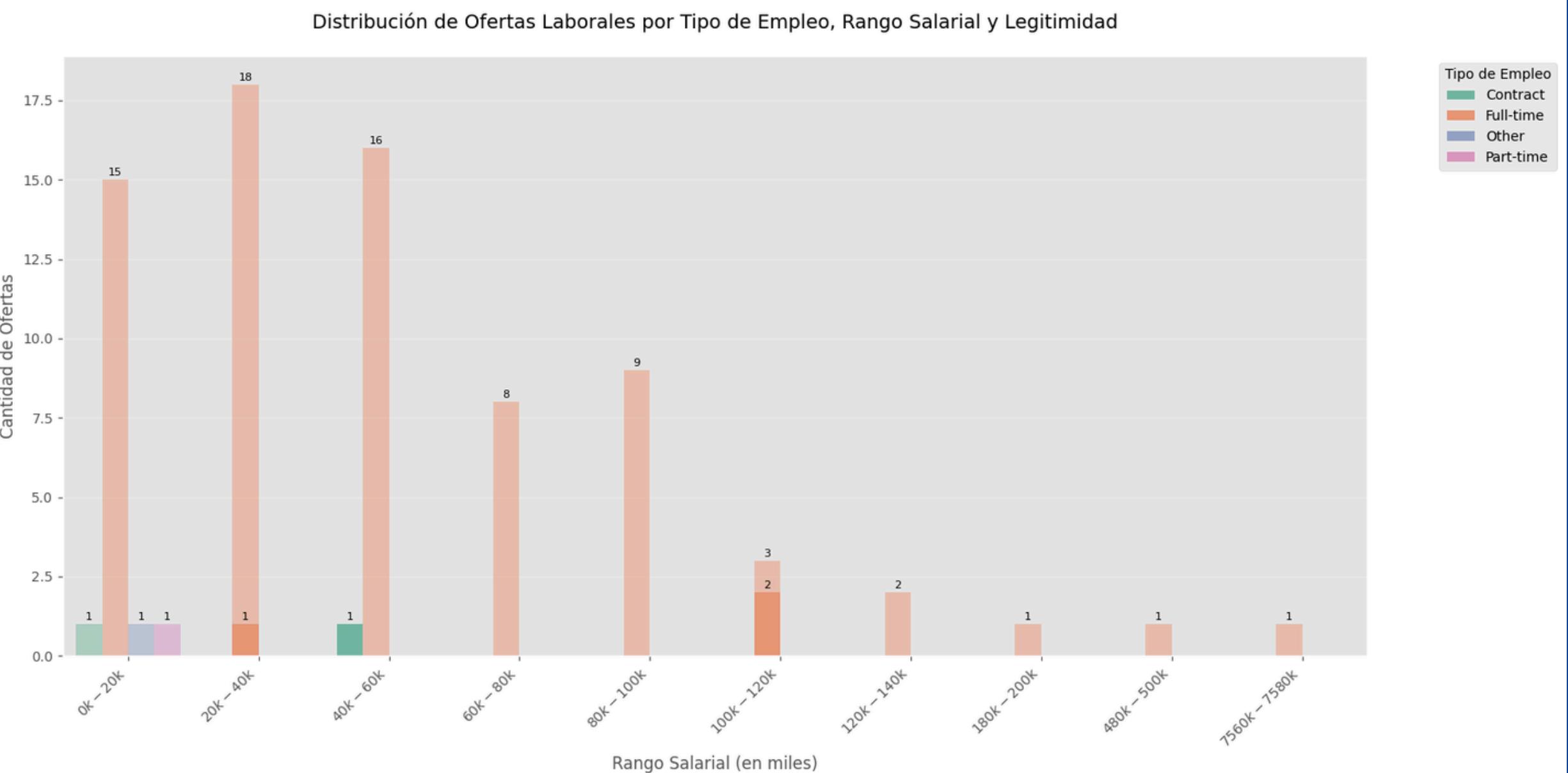
Este patrón sugiere que las ofertas de tipo pasantía o freelance podrían ser más propensas a ser fraudulentas. Esto podría deberse a la naturaleza de estos trabajos, que a menudo son menos formales, tienen menos requisitos de experiencia y pueden atraer a personas que buscan oportunidades laborales a corto plazo sin mucha investigación. Las ofertas fraudulentas pueden aprovechar esta falta de formalidad y la urgencia de los solicitantes para engañarlos.

Al analizar el dataset, un descubrimiento clave es que las ofertas de empleo categorizadas como:

- "Internship" (Pasantía) y "Freelance" tienen una probabilidad significativamente más alta de ser fraudulentas en comparación con otras categorías de empleo como "Full-time" (Tiempo Completo).

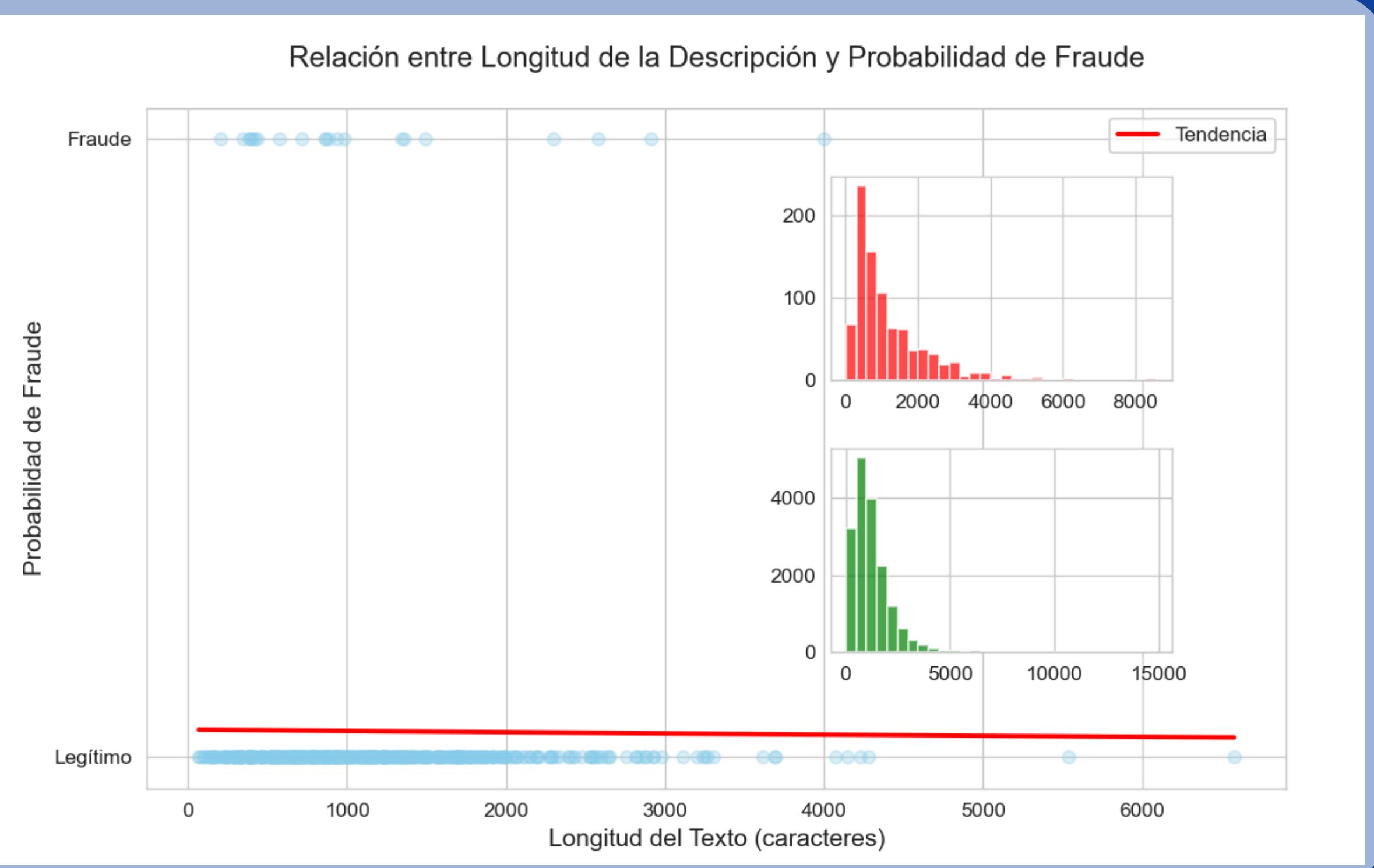


¿Qué reflejan los patrones de tendencia?



- Es común que las ofertas fraudulentas estén más asociadas a trabajos que no requieren mucha experiencia o que son poco definidos (por ejemplo, prácticas o empleos a medio tiempo sin detalles claros).
- Aquí estamos observando cómo se distribuyen las ofertas laborales según los rangos salariales. Las ofertas de trabajos legítimos suelen tener rangos salariales más definidos, mientras que las fraudulentas pueden intentar atraer a los candidatos ofreciendo salarios extremadamente altos o bajos sin detalles claros.

Tendencia: Longitud del Texto vs. Fraude



Patrón:

- Los anuncios fraudulentos tienen descripciones más cortas (promedio: ~500 caracteres) vs. los legítimos (~1000 caracteres).

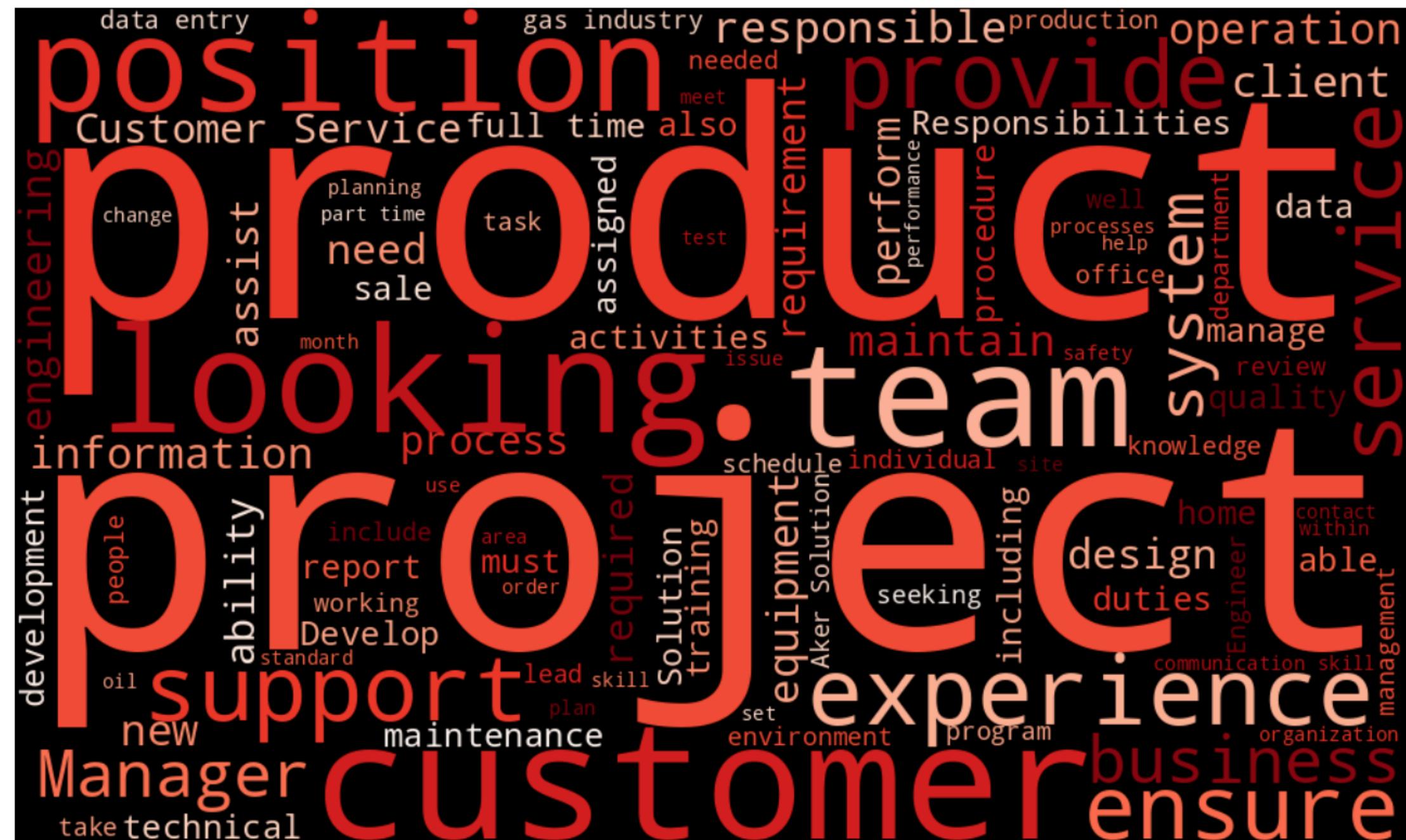
- La probabilidad de fraude disminuye exponencialmente al aumentar la longitud del texto.

¿Por qué ocurre?

- Los fraudes usan lenguaje genérico para atraer a más víctimas sin dar detalles específicos.

Palabras Clave en Anuncios Fraudulentos (WordCloud + Frecuencias)

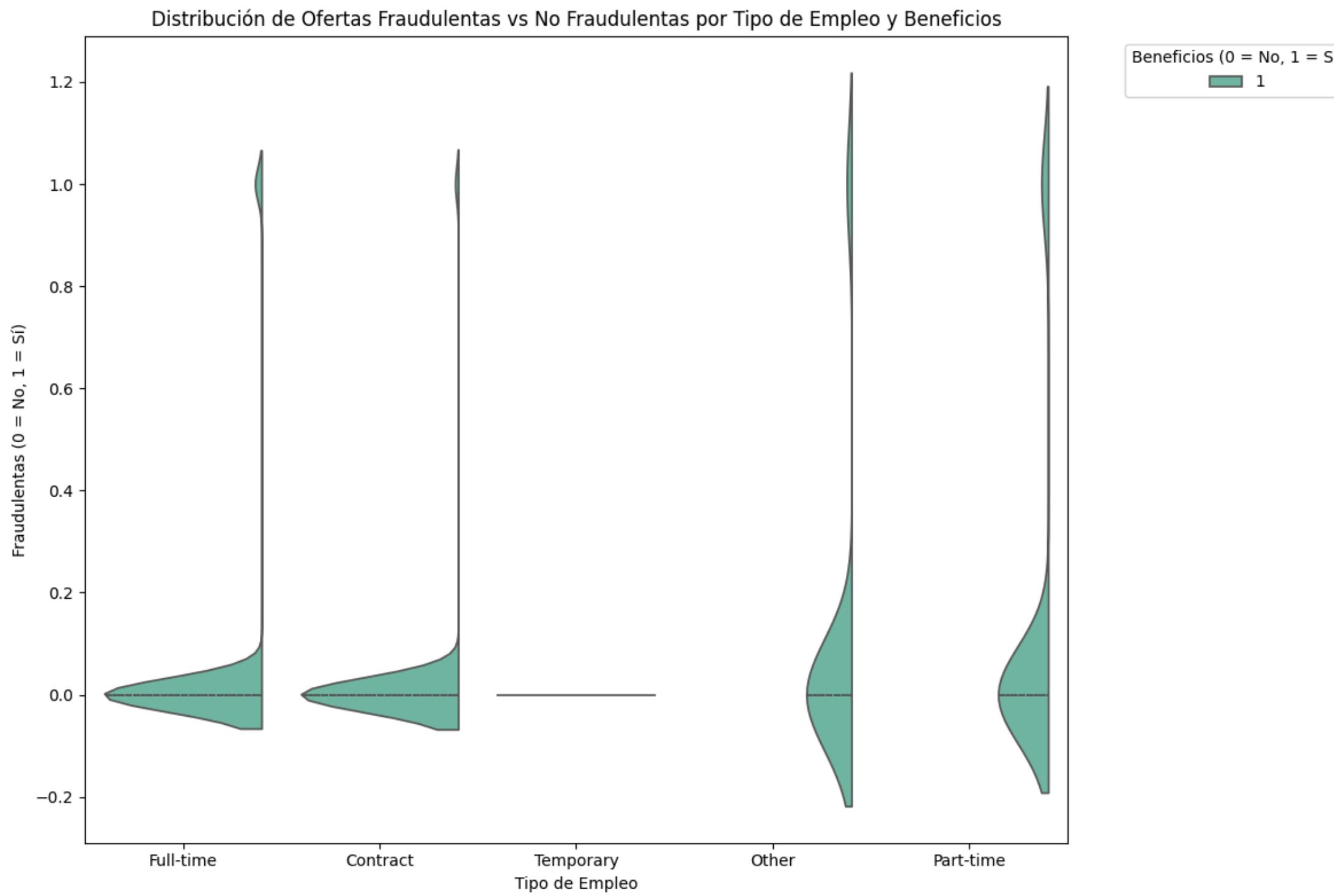
Palabras más Frecuentes en Anuncios Fraudulentos



Los anuncios fraudulentos tienen un 45% menos de caracteres en promedio que los legítimos.
Insight clave: Descriptions con menos de 1,000 caracteres son 3x más probables de ser fraudes.

- Usar **colores contrastantes** (rojo para fraudes, verde para legítimos).
 - Destacar la **línea de tendencia roja** con sombra para intervalo de confianza.
 - Incluir un **pie de foto** con el insight clave.

¿Cómo influye el tipo de empleo y los beneficios ofrecidos en la probabilidad de que una persona perciba una oferta de trabajo como fraudulenta?"



- La pregunta está enfocada en explorar si las personas perciben más las ofertas fraudulentas cuando ciertos beneficios no están presentes o si ciertos tipos de empleo (como pasantías o freelance) generan una mayor probabilidad de fraude.
- Las ofertas con beneficios ausentes podrían ser percibidas como más sospechosas.
- Las ofertas de empleo temporal o freelance pueden tener una mayor tendencia a ser vistas como fraudulentas debido a la naturaleza menos formal de esos trabajos.

GRACIAS