



Proyecto EDA

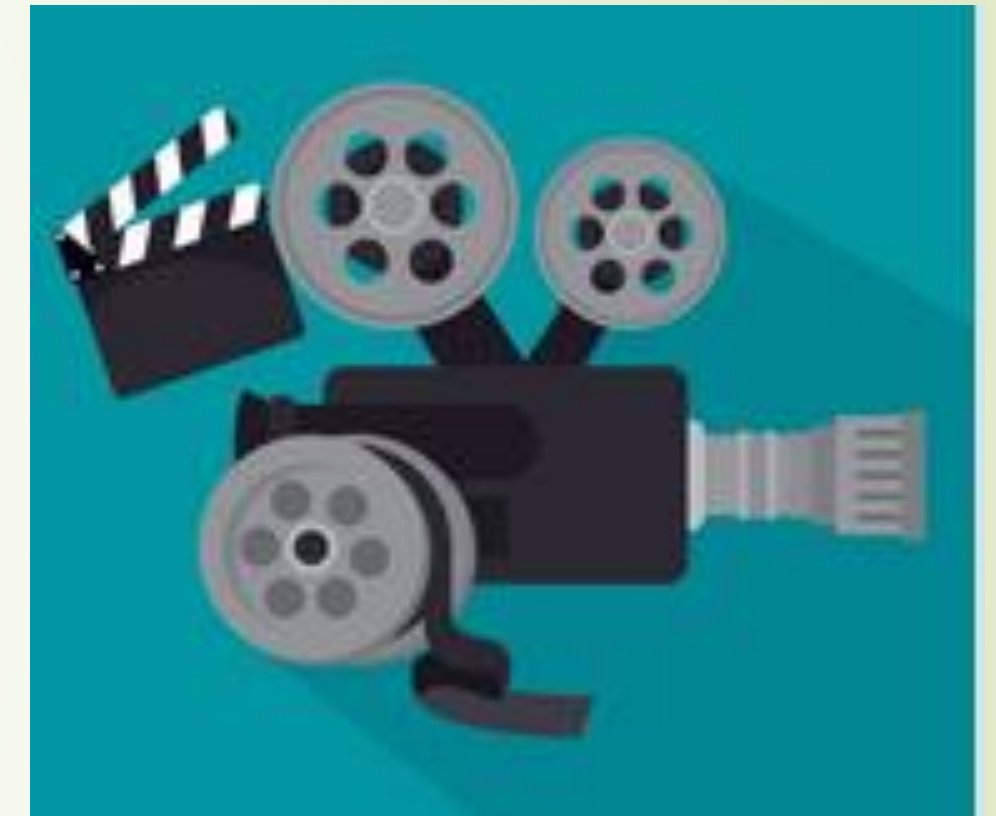
Bootcamps en Data Science 2023

José YAYA



1. INTRODUCCION

- El presente proyecto EDA se centra en el análisis del cine de género drama y terror, con un enfoque particular en el cine de terror en lengua española.
- Surge de la inquietud por comprender las razones detrás de la escasa producción de películas de terror en este idioma.



2. HIPÓTESIS:

¿La lengua en la que se redacta el guion de una película influye significativamente en su clasificación dentro del género cinematográfico, afectando la percepción y la recepción del género de terror en el cine hispanohablante?

Esta hipótesis se basa en la idea de que factores como la historia de la producción cinematográfica, estereotipos culturales, tabúes, la promoción y el interés del público, así como el impacto del idioma en la narrativa, pueden contribuir a la escasa producción de películas de terror en lengua española.



ABORDAJE

En esta fase, se abordaron temas tales como:

- Limpieza de datos
- Análisis y tratamiento de las películas
- Análisis descriptivo de las películas
- Visualización de datos para las películas

En una siguiente fase, se centra en realizar un estudio estadístico profundo. Se llevará a cabo un análisis descriptivo detallado y se aplicarán pruebas estadísticas, como la Chi-Cuadrado, para examinar relaciones y patrones en los datos. Este análisis incluirá variables adicionales como idioma, género, país, calificación y votos.



3. DATOS

Hemos utilizado una muestra de datos obtenida de Kaggle (7 MB, con más de 9000 datos y 14 columnas)



3. DATOS

Como dificultades encontradas, el dataset contenía lo siguiente:

- ➡ Archivo pesado (7 MB)
- ➡ Datos nulos.
- ➡ Información innecesaria a no ser utilizada.
- ➡ Además del tipo de películas, incluía documentales, series de TV, cortos, entre otros.
- ➡ Información contenida entre corchetes y comillas simples

[illegible]

4. SOLUCIONES

Limpieza, análisis y tratamiento de datos:

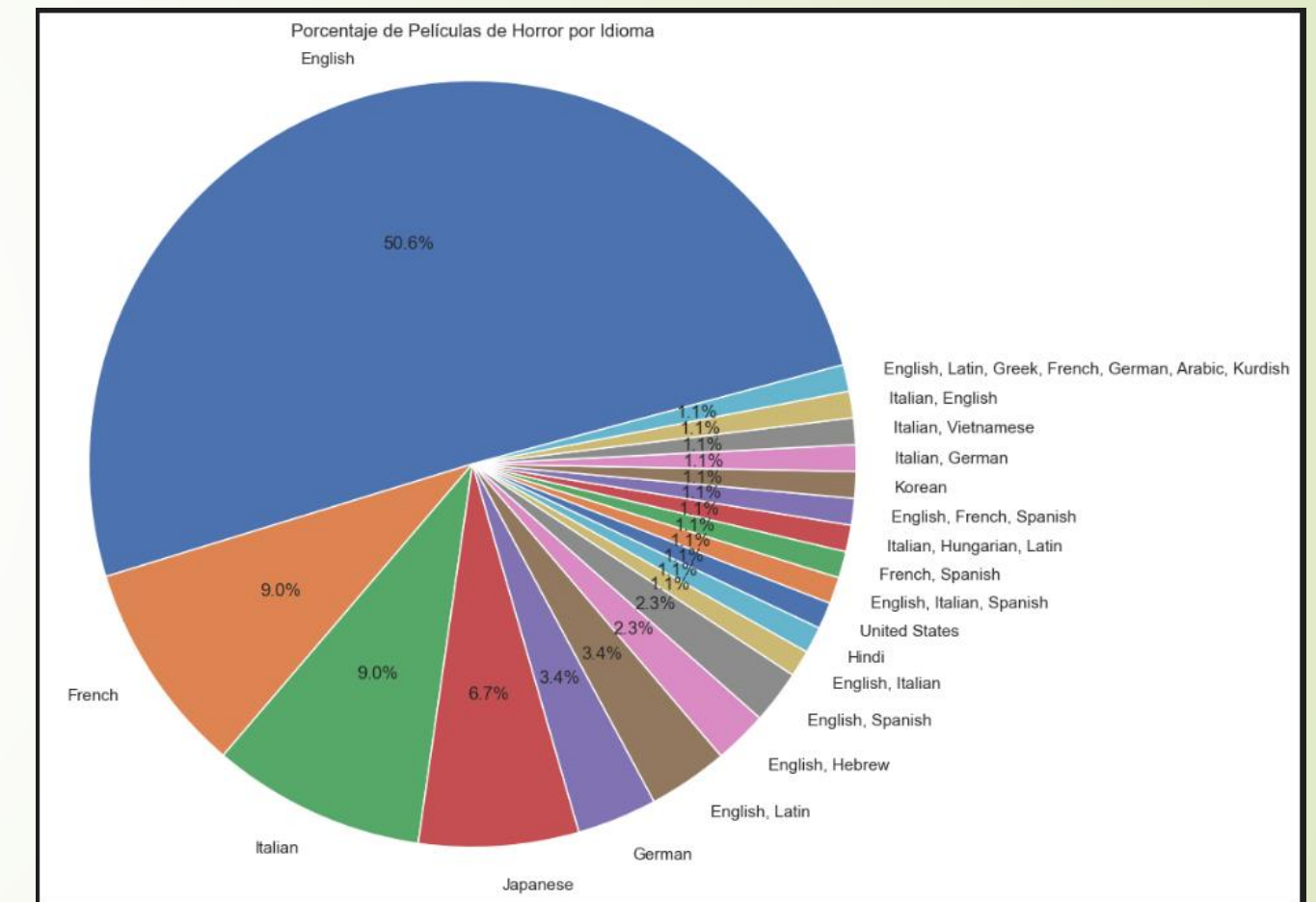
- Eliminación de caracteres no deseados: método `str.replace()`
- Eliminación de espacios en blanco: `str.strip()`
- Uso de expresiones regulares: `regex=True`
- Filtrado por tipo
- Filtrado por género
- Selección de columnas
- Análisis de valores nulos
- Reemplazo de nulos: Se reemplazan los valores nulos en la columna 'language' por el correspondiente valor de la columna 'country', y se verifica nuevamente la cantidad de nulos en el DataFrame.



5. CONCLUSIONES

Tendencia del idioma en el género horror:

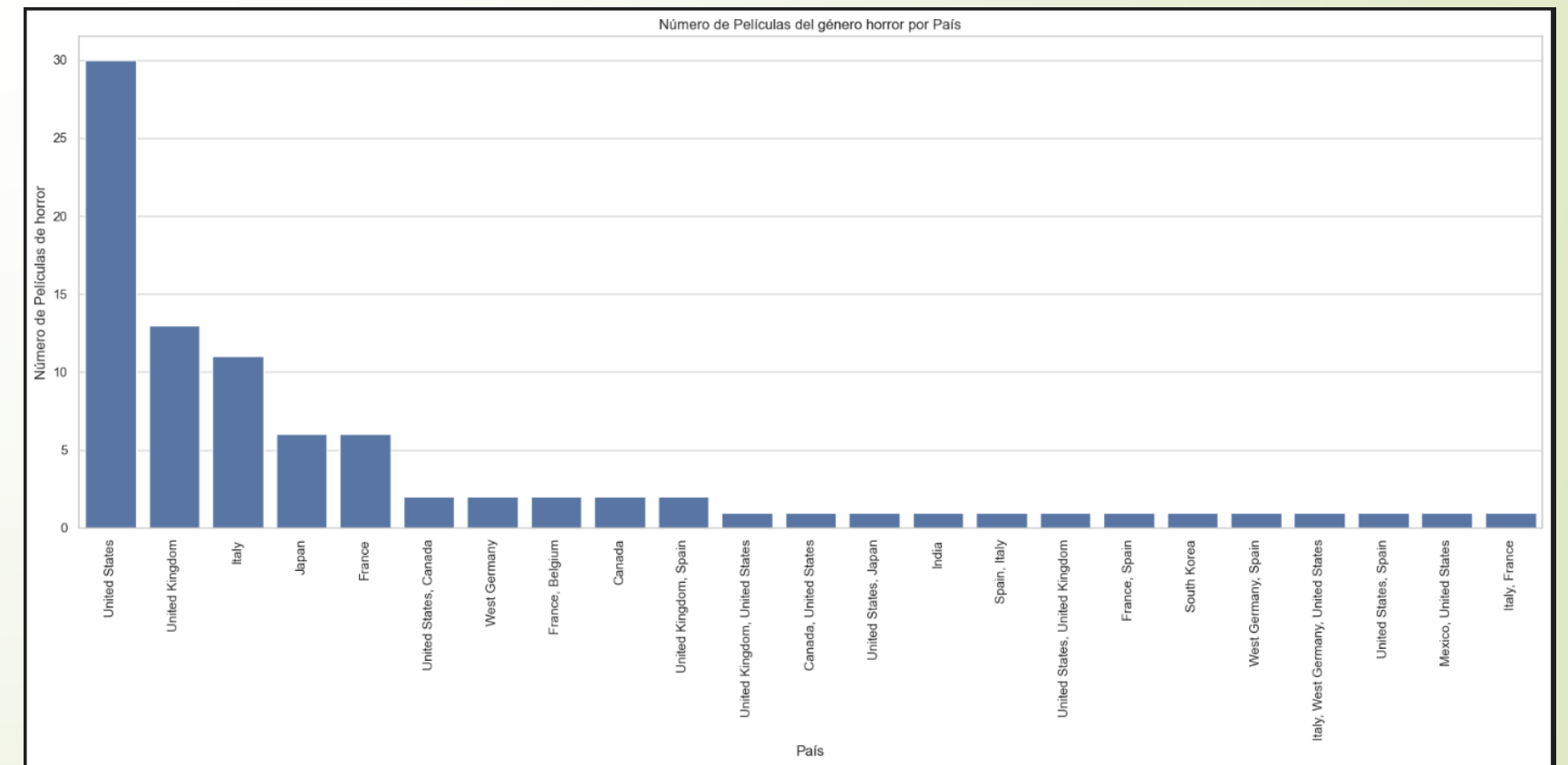
- El análisis revela que el inglés es el idioma más utilizado en películas de horror, con un 50.6%. Este idioma parece contener vocabulario que evoca miedo y tensión.
- Otros idiomas prevalentes en este género son el francés y el italiano (9% cada uno), seguidos por el japonés (6.7%).
- Pensamos que el estilo narrativo y las diferencias culturales, reflejadas en el idioma, pueden influir en cómo se percibe el miedo en las historias de horror.



5. CONCLUSIONES

Idiomas en el género drama

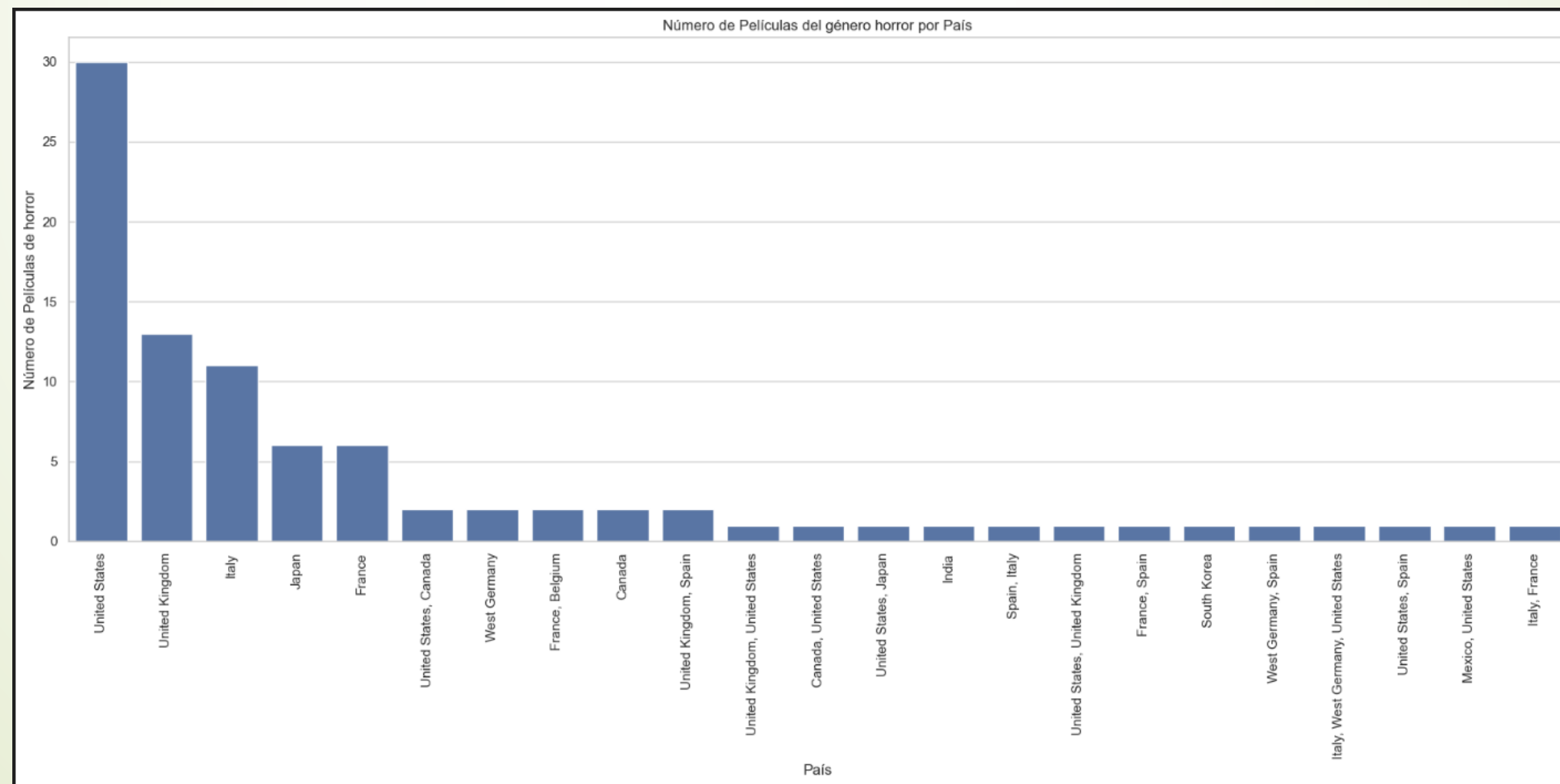
- En el análisis del género dramático, el inglés también es el idioma predominante, con un 46.1%. Le siguen el francés (5.9%), el japonés (3.1%) y el español (2.6%).
- Se observa que los países de habla inglesa dominan la producción cinematográfica, con Estados Unidos liderando (30%), seguido de Inglaterra (6.1%) y Canadá (5.4%).
- La producción de dramas en español es bastante limitada, con pocos países alcanzando cifras significativas



5. CONCLUSIONES

Comparativa entre géneros:

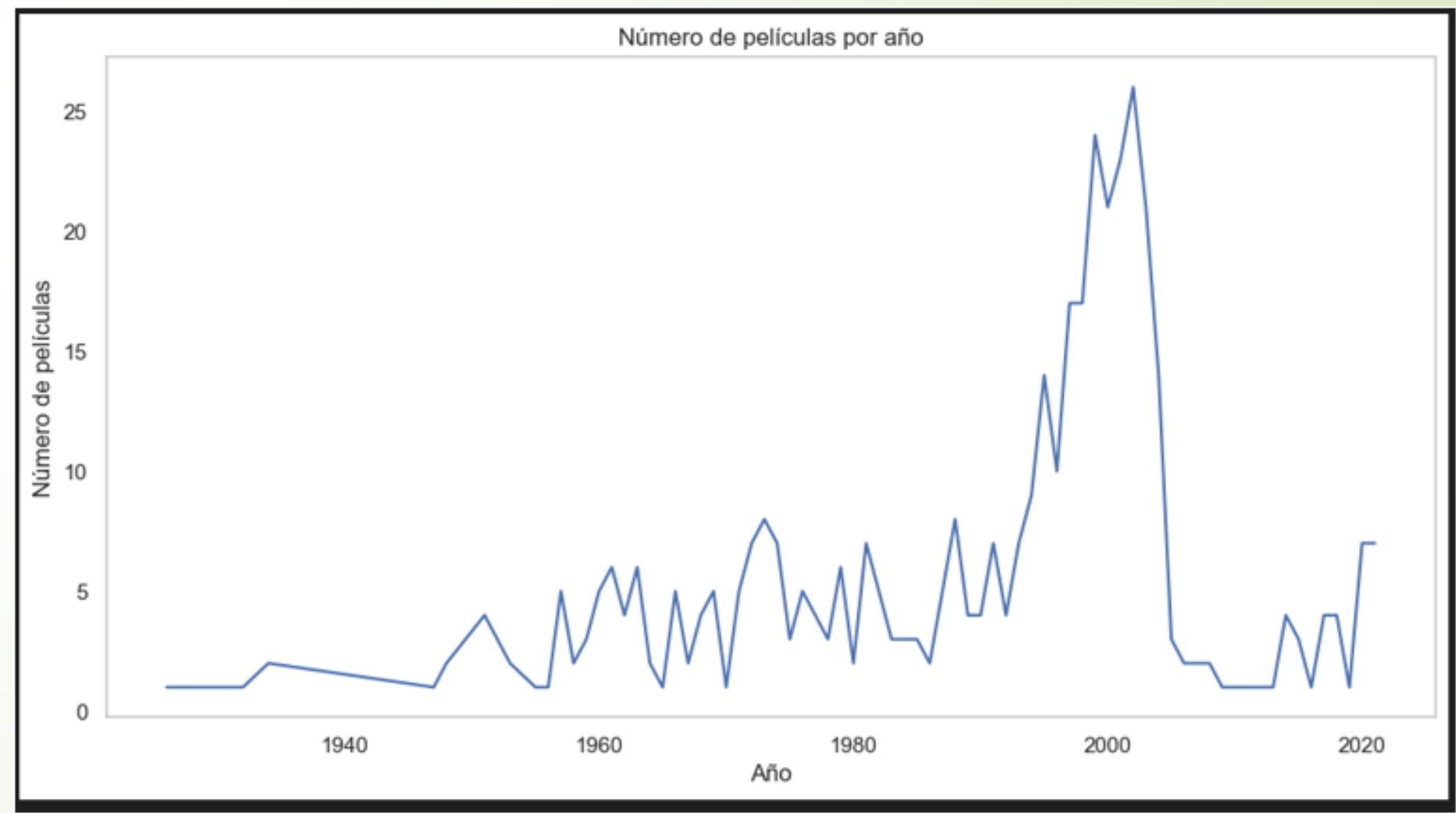
- Tanto en el género horror como en el drama, el inglés es el idioma dominante, aunque el horror tiene una mayor concentración de producciones en este idioma.
- La producción de cine en otros idiomas, incluidos el español, es baja en ambos géneros, destacando una tendencia similar en la predominancia del inglés.



6. ANÁLISIS EXTRA - ESTADÍSTICAS

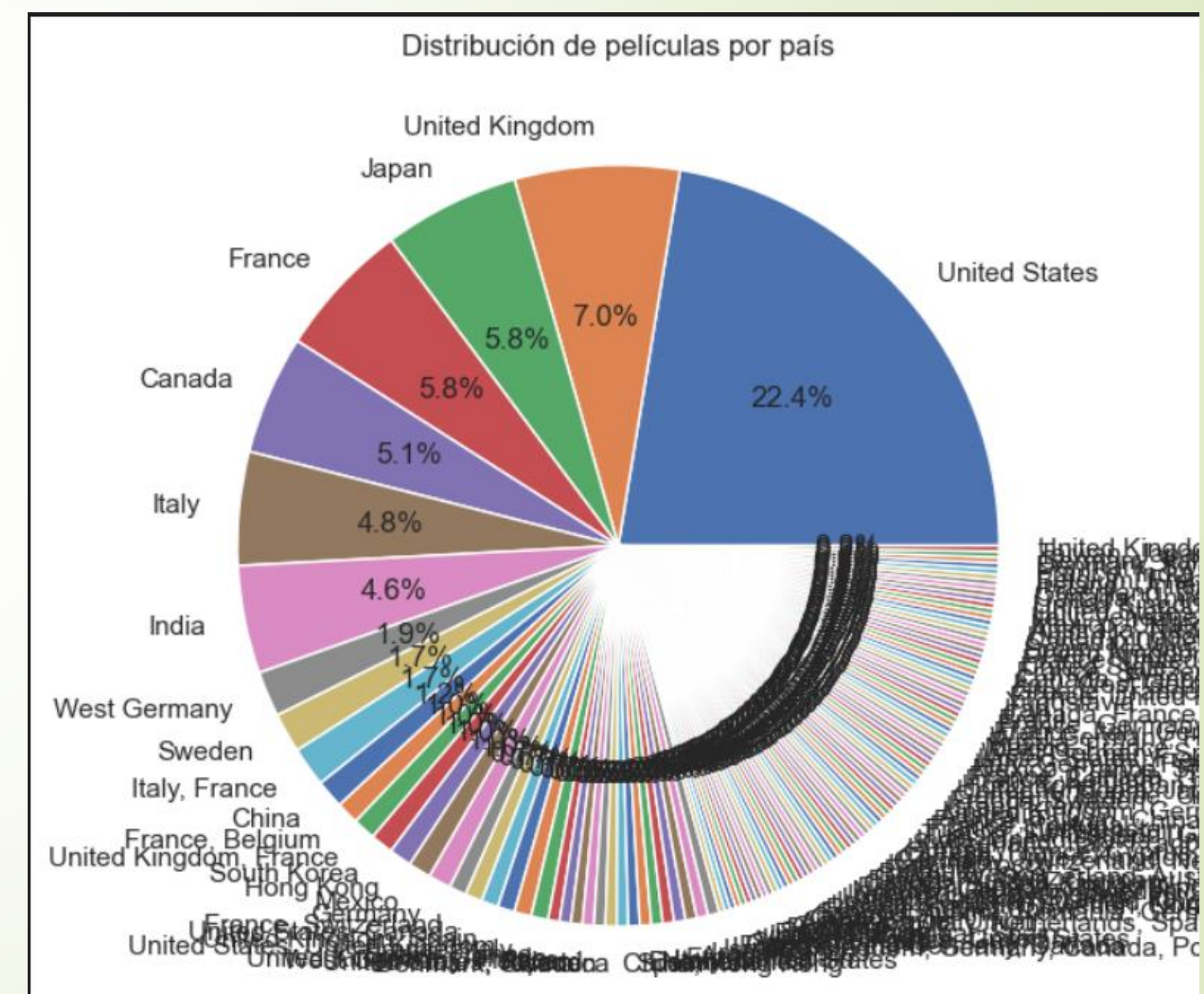
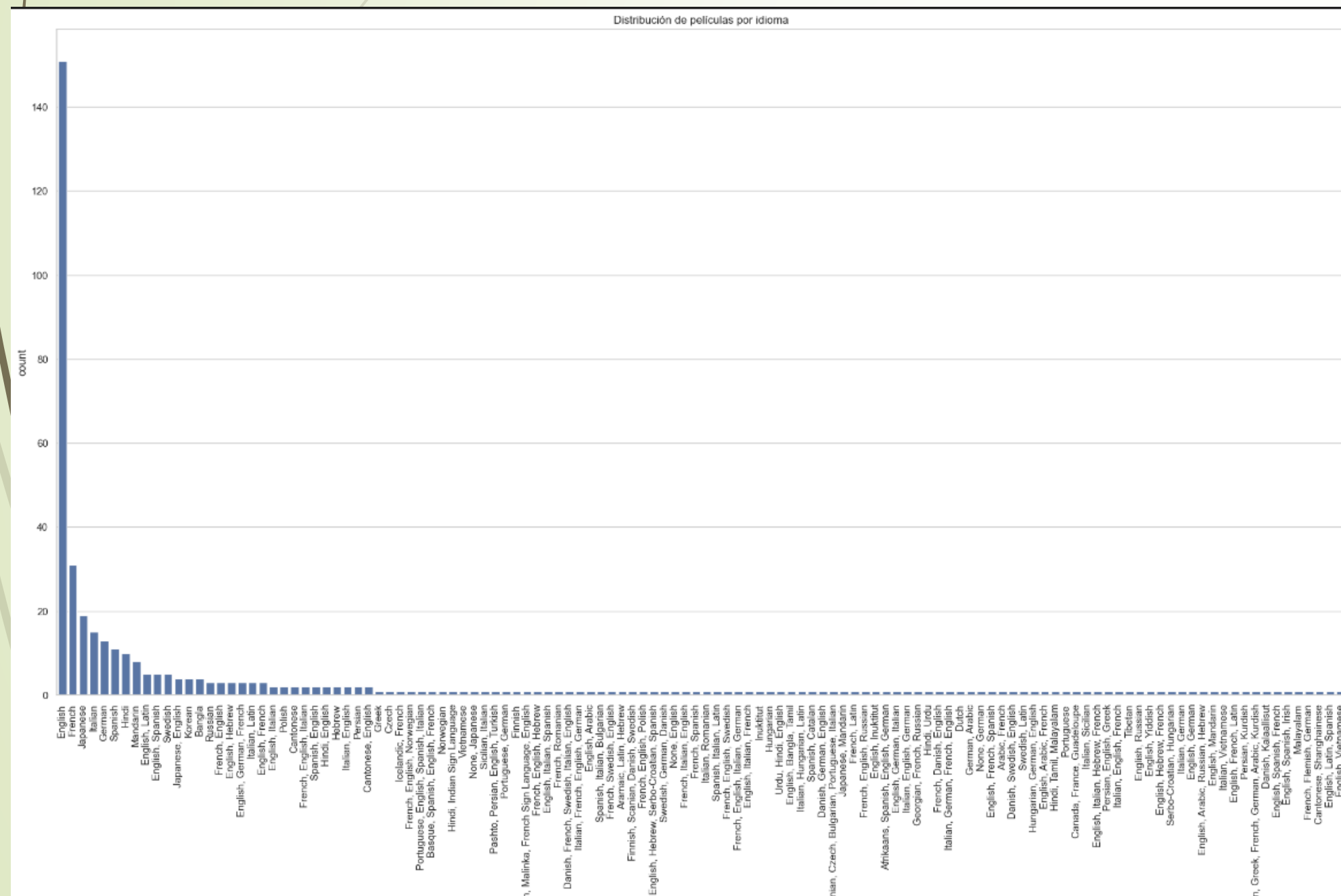
Se llevará a cabo un análisis descriptivo detallado y se aplicarán pruebas estadísticas para examinar relaciones y patrones en los datos. Este análisis incluirá variables adicionales como idioma, género, país, rating y votos.

- Se lleva a cabo una limpieza de datos.
- Análisis descriptivo de las columnas rating y vote
- Comparaciones entre las cantidades de películas del género horror y drama por idioma y país
- Análisis de tendencias en el tiempo



6. ANÁLISIS EXTRA - ESTADÍSTICAS

Se aborda la visualización de la distribución de películas según idioma y país utilizando gráficos.



6. ANÁLISIS EXTRA - ESTADÍSTICAS

- Se realizó un análisis estadístico utilizando la prueba Chi-Cuadrado para evaluar la influencia del idioma en la clasificación de películas de terror en el contexto hispanohablante, considerando variables como idioma, género, país, rating y votos.

1. Filtrado de Datos

2. Análisis Chi-Cuadrado:

- **Idioma y País:** Chi-Cuadrado, obteniendo un valor de **896.46** con un p-valor de **4.19e-44**, lo que indica una relación significativa entre el idioma y el país.
- **Idioma y Rating:** Se categorizó el rating en grupos, resultando en un Chi-Cuadrado de **47.20** y un p-valor de **0.10**, sugiriendo que no hay una relación significativa.
- **Idioma y Votos:** Se categorizó el número de votos, obteniendo un Chi-Cuadrado de **38.90** y un p-valor de **0.82**, lo que también indica una falta de relación significativa.

Resultados finales Chi-Cuadrado:

- **Idioma y País:** Relación significativa.
- **Idioma y Rating:** No significativo.
- **Idioma y Votos:** No significativo.

6. ANÁLISIS EXTRA - ESTADÍSTICAS

- Hemos compilado estos hallazgos en reportes que incluyen gráficos y análisis realizados.
- Con los resultados, hemos realizado un análisis más profundo, con pruebas estadísticas para validar la hipótesis.
- En las pruebas estadísticas se abordó la hipótesis planteada sobre la influencia del idioma en la clasificación dentro del género cinematográfico, especialmente en el cine de terror en el contexto hispanohablante, en este análisis estadístico se incluyeron las variables relevantes: idioma, género, país, rating y votos.
- Chi-cuadrado:
 - - Idioma y País: Existe una fuerte asociación.
 - - Idioma y Rating: No hay suficiente evidencia para afirmar que existe una asociación significativa.
 - - Idioma y Votos: No hay suficiente evidencia para afirmar que existe una asociación significativa.
- En conclusión, la hipótesis inicial se ve respaldada en cuanto a la influencia del idioma en la producción de películas de horror, especialmente en relación con el país de origen.

Atrévete a cruzar el portal de lo desconocido. Ahí donde el miedo termina, comienza la transformación.

