

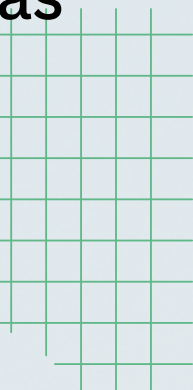


ANÁLISIS COMPLETO DE PELÍCULAS DE DISNEY: EXPLORACIÓN, TENDENCIAS Y PREDICCIONES

ESTUDIO DETALLADO CON PYTHON



Erik Bañez Rojas





Contexto

Disney, una de las compañías de entretenimiento más icónicas del mundo, ha producido un extenso catálogo de películas desde 1937. Con el paso de los años, estas producciones han variado en género, recaudación y relevancia cultural. Este proyecto tiene como objetivo explorar las tendencias y el rendimiento financiero de las películas de Disney para identificar patrones clave y factores de éxito.

Objetivos del Proyecto

- Analizar tendencias en la producción cinematográfica:
- Estudiar cómo ha evolucionado la frecuencia y diversidad de géneros en las películas de Disney a lo largo de las décadas.
- Explorar el rendimiento financiero:
- Investigar los ingresos brutos de las películas y ajustarlos por inflación para evaluar la relevancia de las producciones clásicas y modernas.
- Comprender la influencia de franquicias:
- Evaluar cómo las franquicias (como Marvel y Star Wars) han impactado los ingresos generales de Disney.
- Identificar patrones de éxito:
- Determinar cuáles factores, como género, clasificación MPAA y época, están asociados a mayores ingresos.
- Aplicar técnicas predictivas:
- Usar modelos de machine learning para predecir los ingresos ajustados por inflación de futuras producciones.

Limpieza de Datos

Preparación y Transformación de los Datos

Descripción de las variables

- 📌 movie_title: Título de la película.
- 📌 release_date: Fecha de estreno.
- 📌 genre: género de la película
- 📌 mpaa_rating: Clasificación según la Motion Picture Association.
- 📌 total_gross: Ingresos brutos totales (USD).
- 📌 inflation_adjusted_gross: Ingresos ajustados por inflación (USD).

Revisión de datos faltantes

- Se encontraron valores nulos en las columnas genre y mpaa_rating.
- Acción: Los valores nulos se reemplazaron con "NaN" para poder identificar estas categorías más adelante.

```
df = df.fillna("NaN")
```

Conversión de formatos

- release_date se convirtió al formato datetime para permitir análisis temporales y la extracción de variables como décadas y años.

```
df['release_date'] = pd.to_datetime(df['release_date'])
```

Creación de una nueva columna: decade

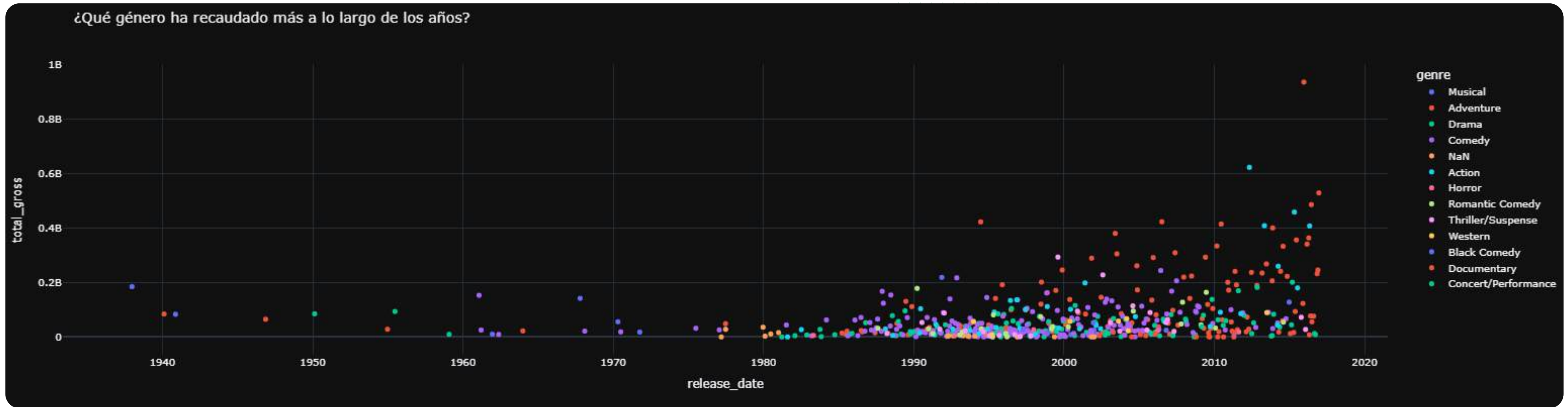
- Se agruparon las películas en décadas para analizar tendencias de producción y recaudación a lo largo del tiempo.

```
lis=[]
for i in range(579):
    if df['release_date'][i].year > 2010:
        lis.append('2010-2020')
    elif df['release_date'][i].year <= 2010 and df['release_date'][i].year > 2000:
        lis.append('2000-2010')
    elif df['release_date'][i].year <= 2000 and df['release_date'][i].year > 1990:
        lis.append('1990-2000')
    elif df['release_date'][i].year <= 1990 and df['release_date'][i].year > 1980:
        lis.append('1980-1990')
    elif df['release_date'][i].year <= 1980 and df['release_date'][i].year > 1970:
        lis.append('1970-1980')
    elif df['release_date'][i].year <= 1970 and df['release_date'][i].year > 1960:
        lis.append('1960-1970')
    elif df['release_date'][i].year <= 1960 and df['release_date'][i].year > 1950:
        lis.append('1950-1960')
    else:
        lis.append('<1950')
df['decade'] = lis
```

579 rows x 7 columns

Análisis Exploratorio

Frecuencia de géneros por décadas.



Visualización de tendencias utilizando la variable release_date

```
fig = px.scatter(df, x='release_date', y="total_gross", template='plotly_dark', title='¿Qué género ha recaudado más a lo largo de los años?', color='genre')  
fig.data[0].line.color = 'limegreen'  
fig.show()
```



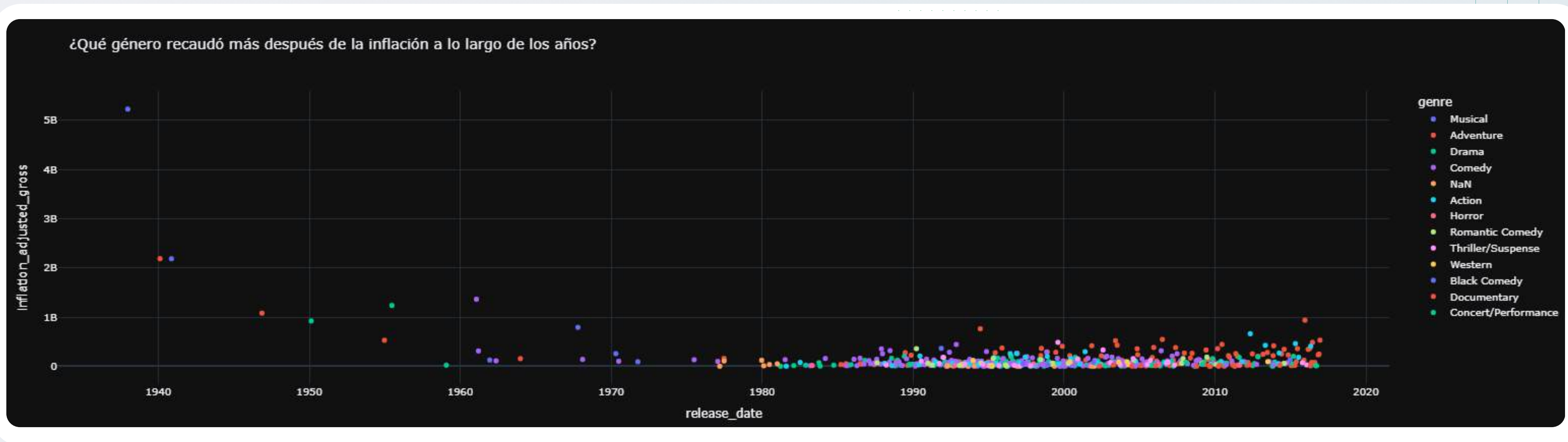

OBSERVACIONES



- 📌 La primera película producida por Disney fue un musical, el clásico 'Blancanieves y los siete enanitos'.
- 📌 La frecuencia de las películas aumenta considerablemente a lo largo de las décadas.
- 📌 Los ingresos brutos de las películas muestran un crecimiento exponencial, alcanzando sus picos más altos después del 2000.
- 📌 La popularidad de los géneros cambió con el tiempo: musicales dominaron antes de 1960, mientras que aventura y acción lideran desde 1990.
- 📌 Desde 1990, se observa una mayor diversidad de géneros, en contraste con las décadas anteriores, que estaban dominadas por unos pocos géneros.
- 📌 Las producciones de alto presupuesto después del 2000 han generado los mayores ingresos brutos, destacando el impacto de franquicias y éxitos recientes.
- 📌 La variable de la década permite un análisis más profundo sobre cómo cambiaron las tendencias de géneros y recaudación a lo largo del tiempo.

Análisis Exploratorio

Relación entre Géneros, Años de Lanzamiento e Ingresos Ajustados por Inflación.



Exploración de la recaudación ajustada por inflación en diferentes géneros a lo largo del tiempo.

```
fig = px.scatter(df, x='release_date', y="inflation_adjusted_gross", template='plotly_dark', title='¿Qué género recaudó más después de la inflación a lo largo de los años?', color='genre')
fig.data[0].line.color = 'limegreen'
fig.show()
```



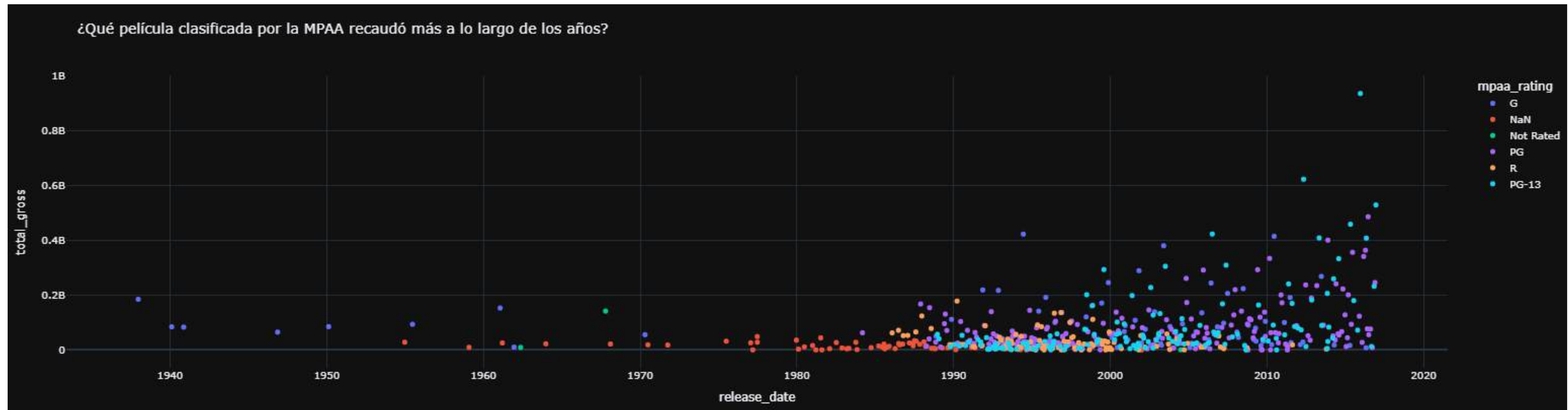

OBSERVACIONES



- 📌 Con el ajuste por inflación, los ingresos de las películas clásicas se incrementaron notablemente, posicionando a 'Blancanieves y los siete enanitos' como la producción más valiosa de Disney en términos de recaudación, superando incluso a Avengers: Endgame.
- 📌 Las películas anteriores a 1960, como los musicales, dominan los ingresos ajustados por inflación, destacándose por ser clásicos duraderos.
- 📌 Desde 1990, los géneros de acción y aventura comenzaron a liderar los ingresos brutos ajustados por inflación, mostrando un cambio en las preferencias del público.
- 📌 Aunque el número de películas producidas creció con los años, los ingresos ajustados por inflación de las películas más recientes son más consistentes pero no alcanzan los niveles de clásicos como Blancanieves.
- 📌 Los géneros musicales y de aventuras generaron los picos más altos en ingresos ajustados, destacándose en diferentes períodos de tiempo.
- 📌 En las décadas recientes, aunque se observa mayor diversidad de géneros, ninguno logra los niveles de ingresos ajustados alcanzados por los grandes clásicos de Disney.
- 📌 Este análisis resalta cómo las películas históricas siguen siendo económicamente relevantes debido al ajuste por inflación, subrayando el impacto de producciones de la época dorada de Disney.

Análisis Exploratorio

Recaudación Total por Películas Clasificadas por MPAA a lo Largo del Tiempo

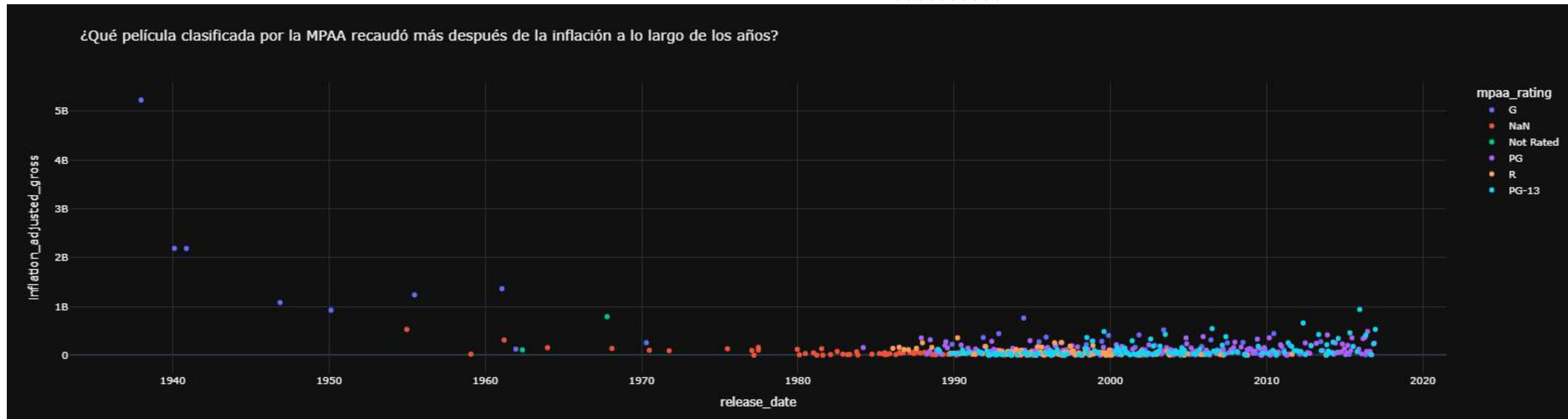


Evaluación de la recaudación total de películas en relación con su clasificación MPAA

```
fig = px.scatter(df, x='release_date', y="total_gross", template='plotly_dark', title='¿Qué película clasificada por la MPAA recaudó más a lo largo de los años?', color='mpaa_rating')
fig.data[0].line.color = 'limegreen'
fig.show()
```


Análisis Exploratorio

Recaudación Ajustada por Inflación para Películas Clasificadas por MPAA



Impacto de la inflación en las recaudaciones de películas según su clasificación

```
fig = px.scatter(df, x='release_date', y="inflation_adjusted_gross", template='plotly_dark', title='¿Qué película clasificada por la MPAA recaudó más después de la inflación a lo largo de los años?', color='mpaa_rating')
fig.data[0].line.color = 'limegreen'
fig.show()
```




OBSERVACIONES



✚ Podemos ver tendencias similares en la clasificación de la MPAA: las primeras películas de Disney solían tener clasificación G, pero las más recientes tienen en su mayoría clasificación PG y PG-13.

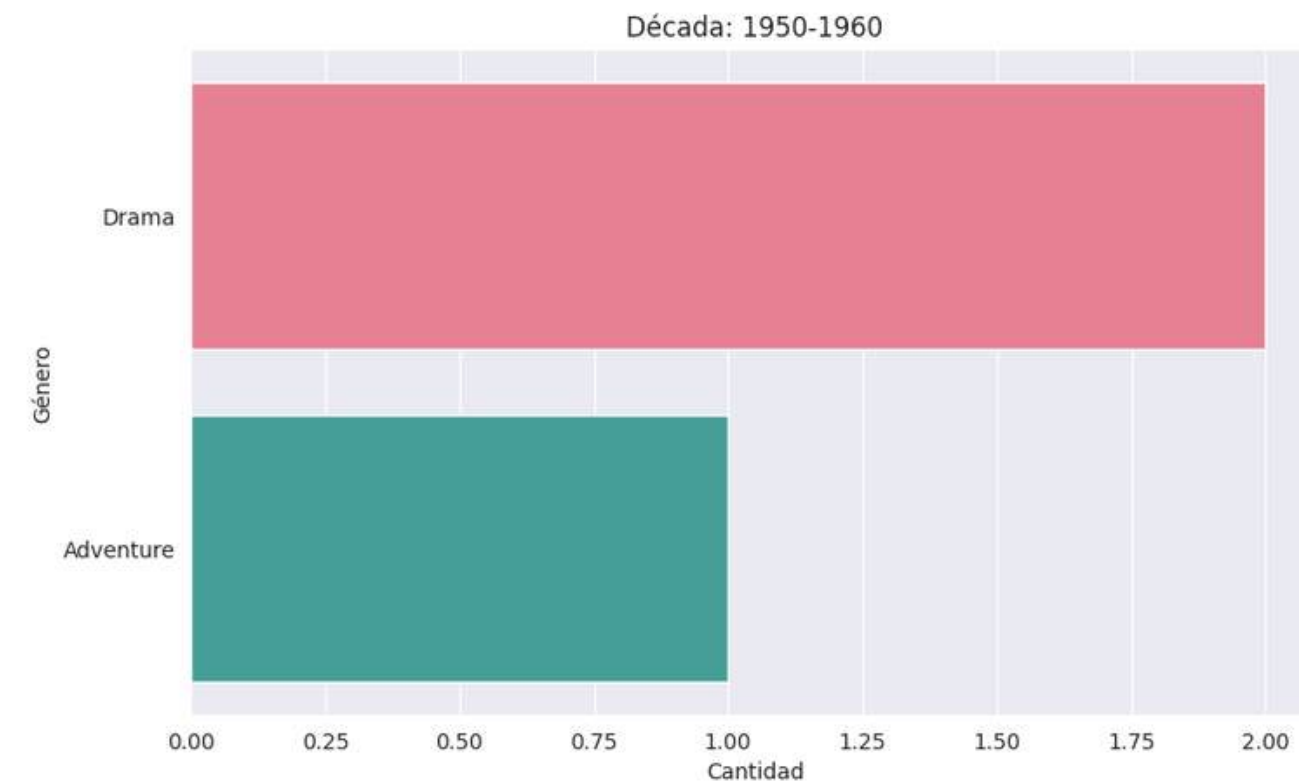
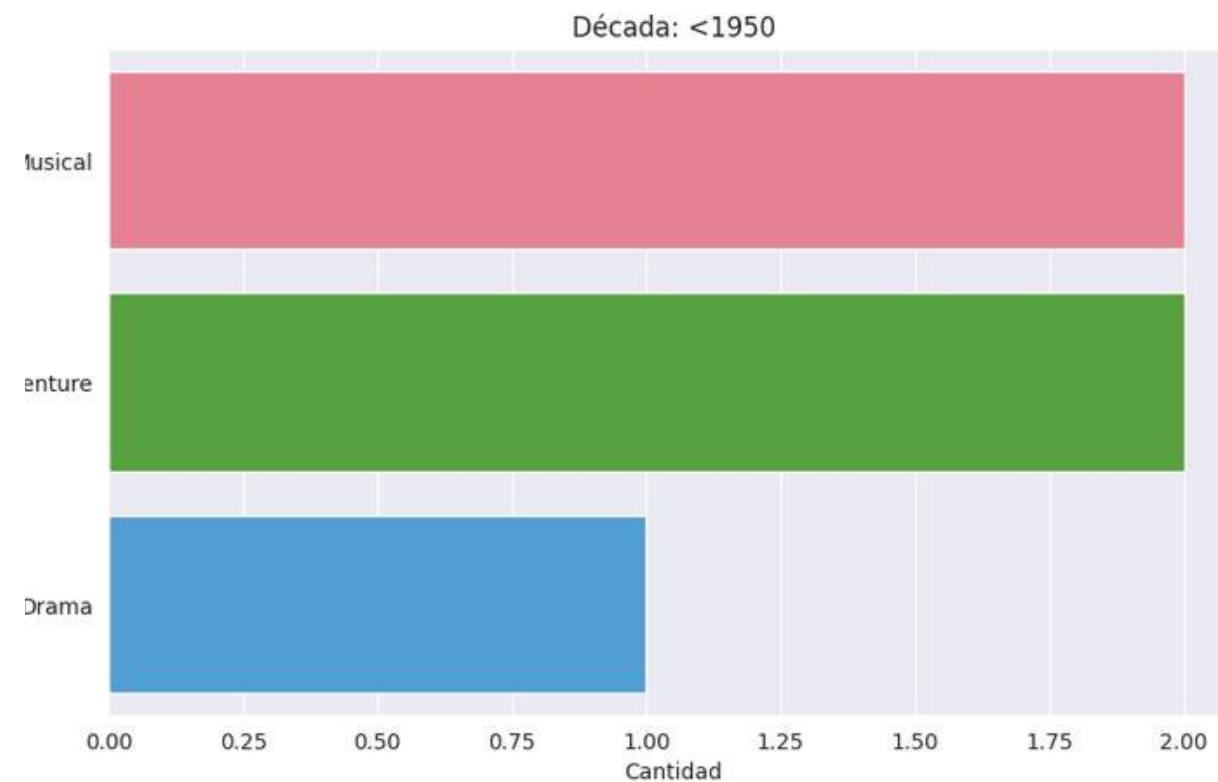
✚ Debido a la inflación en los ingresos brutos de las películas, las primeras películas que tenían clasificación G recaudaron más que las películas actuales con clasificación PG o PG-13.

Análisis Exploratorio

Utilizamos la variable Década para realizar un análisis de tendencias más detallado.

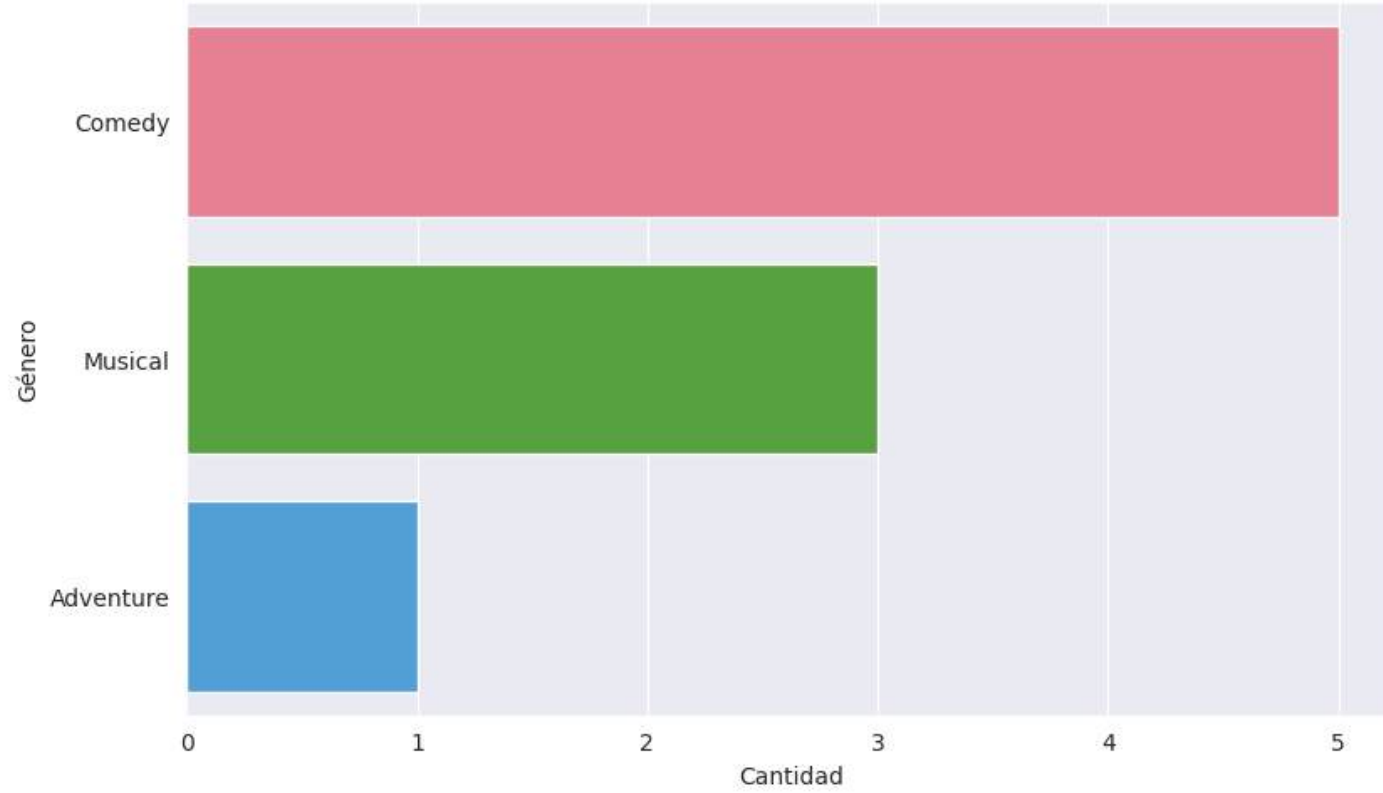
```
fig = px.scatter(df, x='release_date', y="inflation_adjusted_gross", template='plotly_dark', title='¿Qué película clasificada por la MPAA recaudó más después de la inflación a lo largo de los años?', color='mpaa_rating')
fig.data[0].line.color = 'limegreen'
fig.show()
```

Análisis detallado de la evolución de los géneros cinematográficos de Disney a lo largo de las décadas.

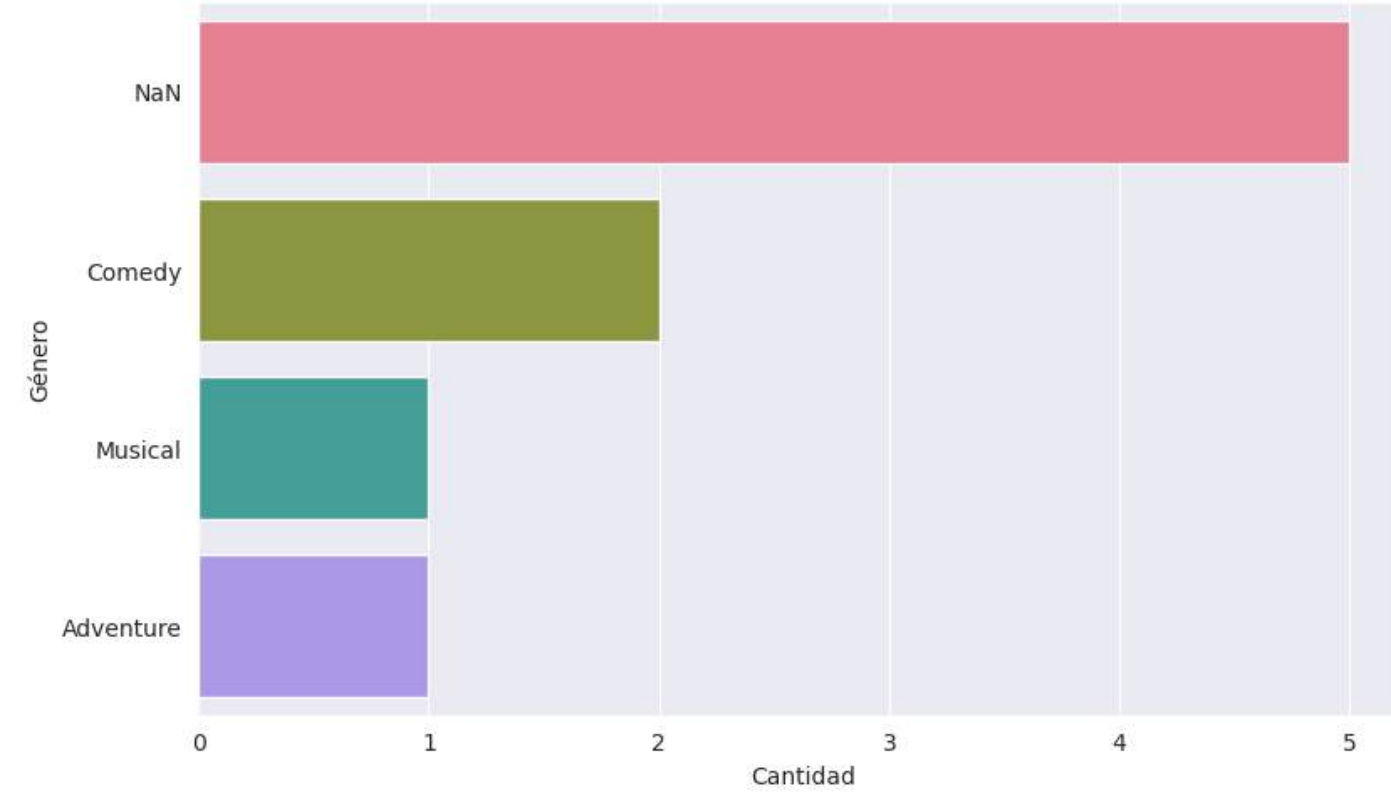


RESULTADOS

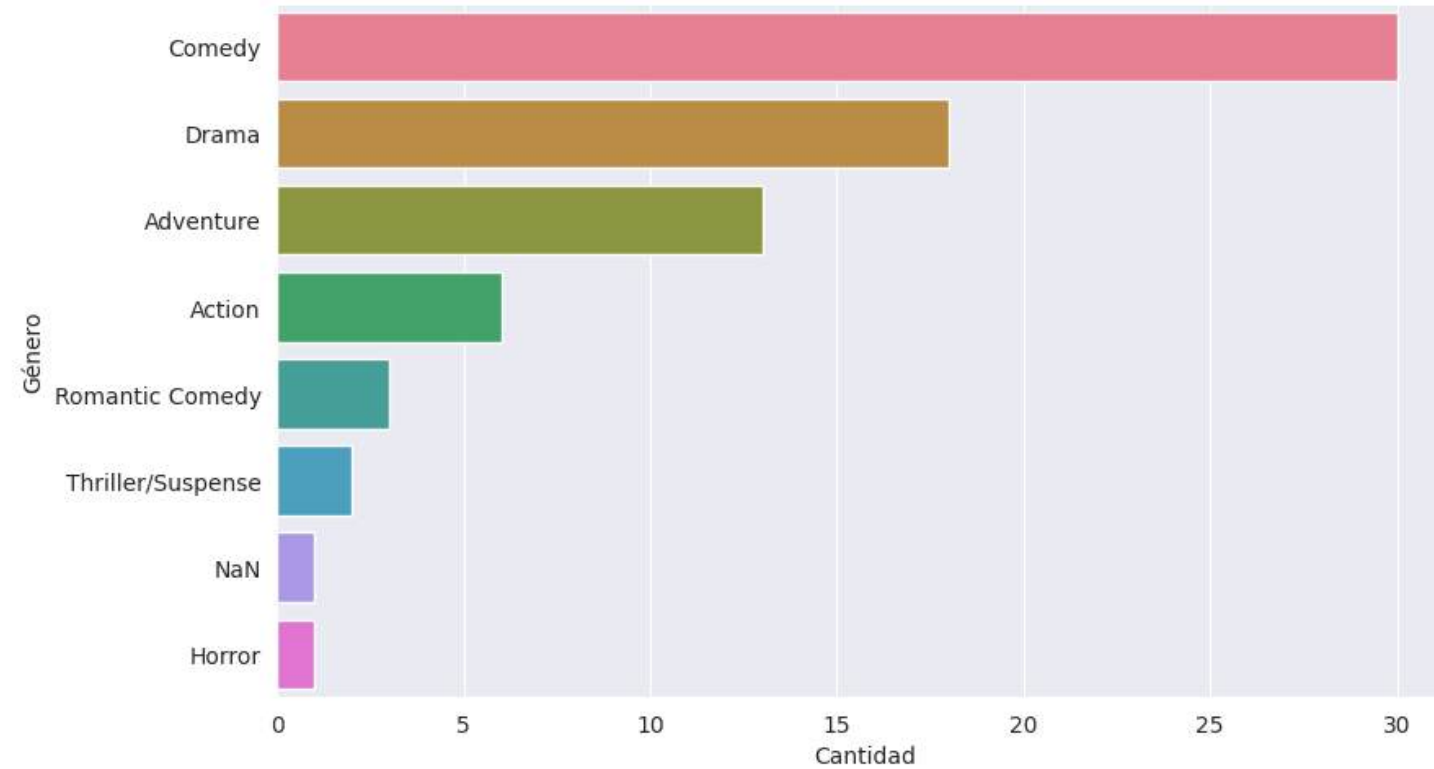
Década: 1960-1970



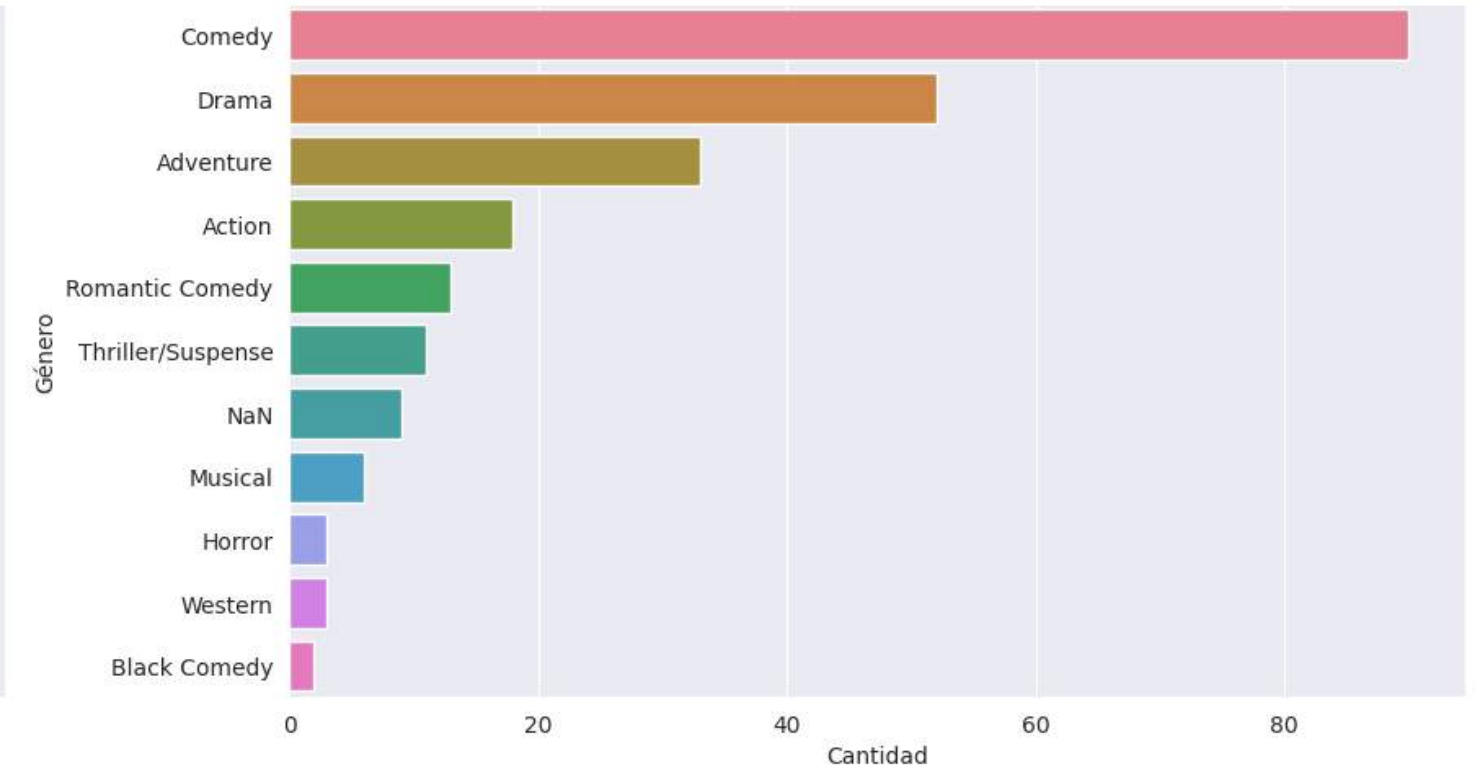
Década: 1970-1980



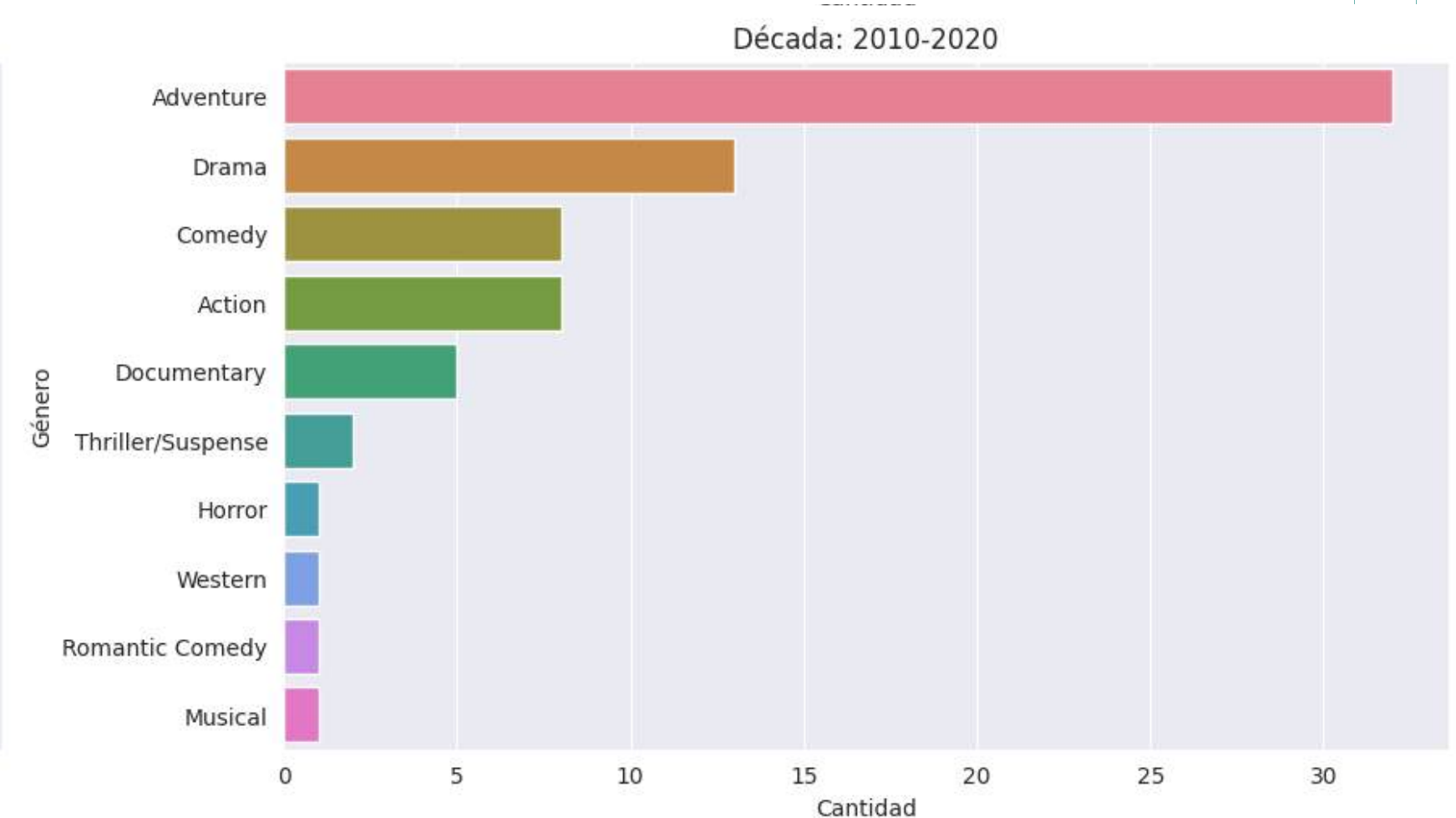
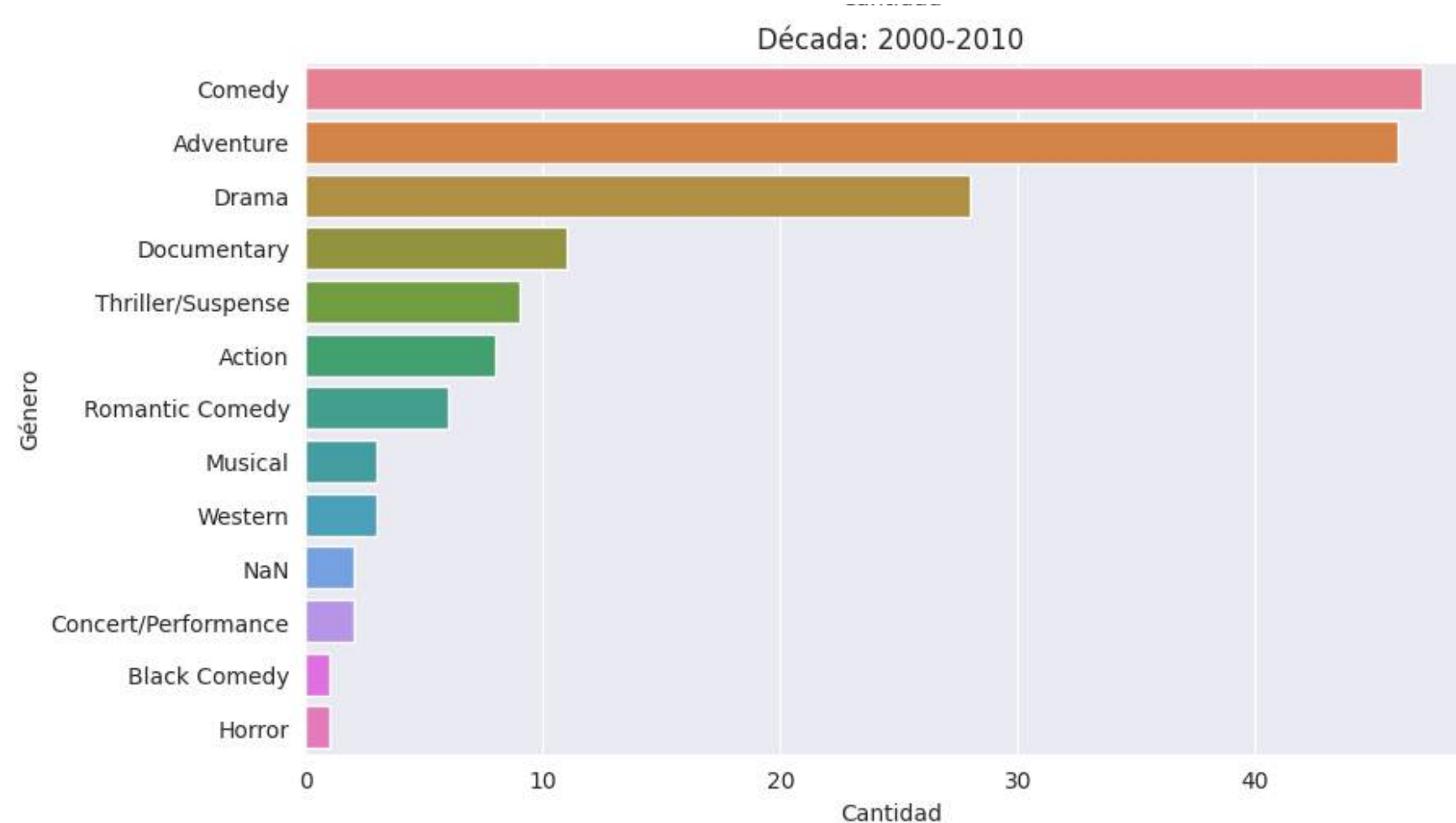
Década: 1980-1990



Década: 1990-2000



RESULTADOS



OBSERVACIONES

✚ Podemos ver que en diferentes décadas Disney se centró en hacer películas de diferentes géneros.

✚ De 1980 a 2010 se centraron principalmente en películas de comedia, aunque las películas de aventuras tuvieron un rendimiento mucho mejor a partir de 2010, observaron este punto y comenzaron a centrarse más en el género de aventuras.

Análisis Exploratorio

Usamos movie_title para clasificar películas según su franquicia

```
movies_franchise_list = []
lis = df['movie_title']
for i in lis:
    a = re.search(r'Cars|Star Wars|Pirate|Prince of|Avengers|Captain America|Thor|Doctor Strange|Muppets|Toy Story|Ant Man|Guardian|Iron Man',i)
    if a:
        movies_franchise_list.append(i)
movies_name=['Cars','Star Wars','Pirates of the Caribbean','Prince of Persia','Toy Story','Muppets']
marvel = ['Avengers','Captain America','Thor','Doctor Strange','Iron Man','Guardians']
df['movie_franchise'] = 'None'
```

```
def movies_franchise(movies_name_list,replace_with_name_in_list=True,name=''):
    for i in movies_franchise_list:
        for k in range(len(df)):
            if df['movie_title'][k] ==i:
                for j in movies_name_list:
                    a= re.search(j,i)
                    if a:
                        if replace_with_name_in_list:
                            df['movie_franchise'][k] = j
                        else:
                            df['movie_franchise'][k] = name
movies_franchise(movies_name)
movies_franchise(marvel,False,'Marvel')
df_movie_franchise = df[df['movie_franchise'] != 'None']
```




RESULTADOS

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross	decade	movie_franchise
0	Snow White and the Seven Dwarfs	1937-12-21	Musical	G	184925485	5228953251	<1950	None
1	Pinocchio	1940-02-09	Adventure	G	84300000	2188229052	<1950	None
2	Fantasia	1940-11-13	Musical	G	83320000	2187090808	<1950	None
3	Song of the South	1946-11-12	Adventure	G	65000000	1078510579	<1950	None
4	Cinderella	1950-02-15	Drama	G	85000000	920608730	<1950	None
...
574	The Light Between Oceans	2016-09-02	Drama	PG-13	12545979	12545979	2010-2020	None
575	Queen of Katwe	2016-09-23	Drama	PG	8874389	8874389	2010-2020	None
576	Doctor Strange	2016-11-04	Adventure	PG-13	232532923	232532923	2010-2020	Marvel
577	Moana	2016-11-23	Adventure	PG	246082029	246082029	2010-2020	None
578	Rogue One: A Star Wars Story	2016-12-16	Adventure	PG-13	529483936	529483936	2010-2020	Star Wars
579 rows x 8 columns								

Nota: Por el momento, solo he considerado las siguientes franquicias de películas:

- Toy Story
- Piratas del Caribe
- Cars
- Príncipe de Persia
- Marvel
- Star Wars

Esta lista no es exhaustiva de todas las franquicias de Disney. En futuras versiones, podría incluir más franquicias, pero para el momento actual, mi análisis se basa únicamente en estas.

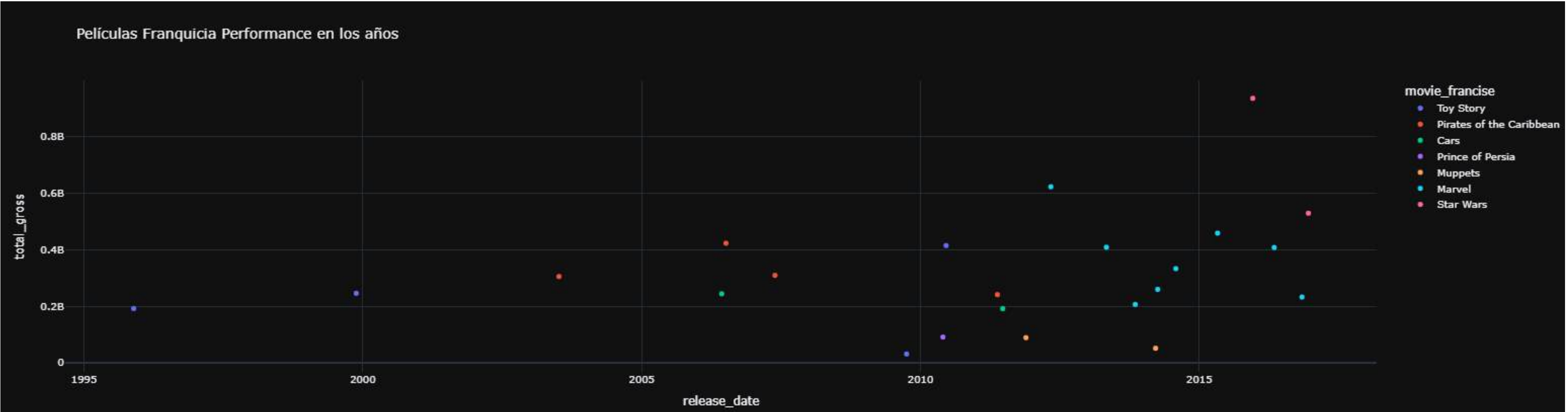


Análisis de Desempeño de Franquicias Cinematográficas

Evolución de las principales franquicias de Disney en términos de recaudación total.

```
fig = px.scatter(df_movie_franchise, x='release_date', y="total_gross",template='plotly_dark',title='Películas Franquicia Performance en los años',color='movie_franchise')
fig.data[0].line.color = 'limegreen'
fig.show()
```

Resultado

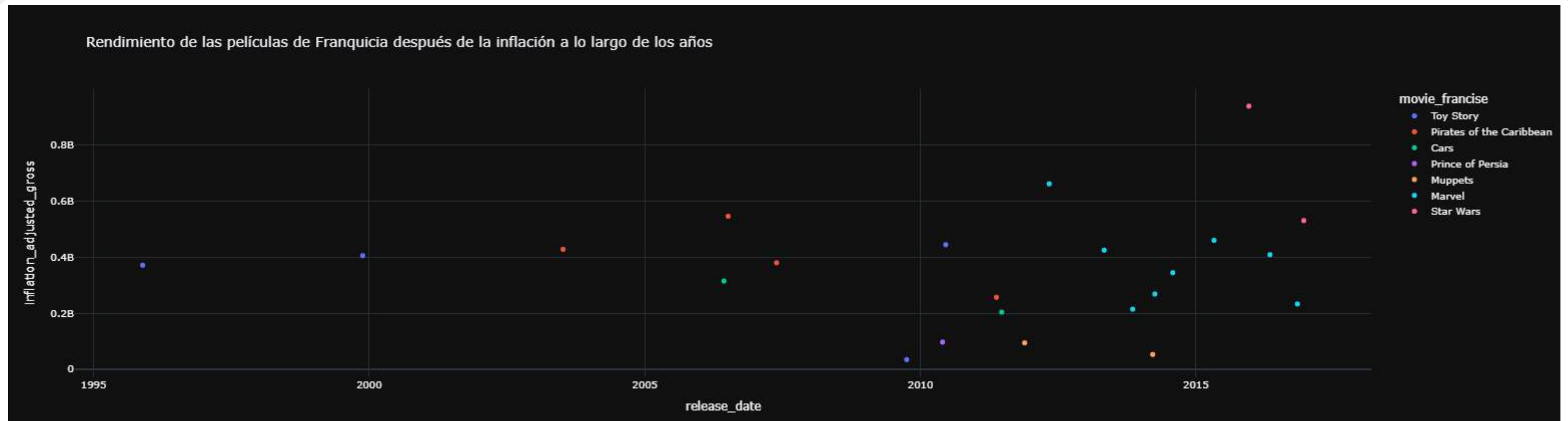


Análisis de Desempeño de Franquicias Cinematográficas Ajustado por Inflación

Análisis del rendimiento financiero de franquicias populares de Disney considerando la inflación.

```
fig = px.scatter(df_movie_franchise, x='release_date', y="inflation_adjusted_gross", template='plotly_dark', title='Rendimiento de las películas de Franquicia después de la inflación a lo largo de los años', color='movie_franchise')
fig.data[0].line.color = 'limegreen'
fig.show()
```

Resultado





OBSERVACIONES



📌 Una de las películas de Star Wars (Star Wars: El despertar de la fuerza) recaudó más al menos hasta 2016.

📌 Toy Story y Piratas del Caribe tuvieron un desempeño ligeramente mejor después de la inflación El conjunto de datos solo contiene películas que originalmente fueron producidas por Disney, por lo que se excluyen todas las películas que Disney compró más tarde; por ejemplo: solo se incluyen las películas de Star Wars realizadas después de 2012.

Impacto Financiero de las Franquicias en Películas de Disney

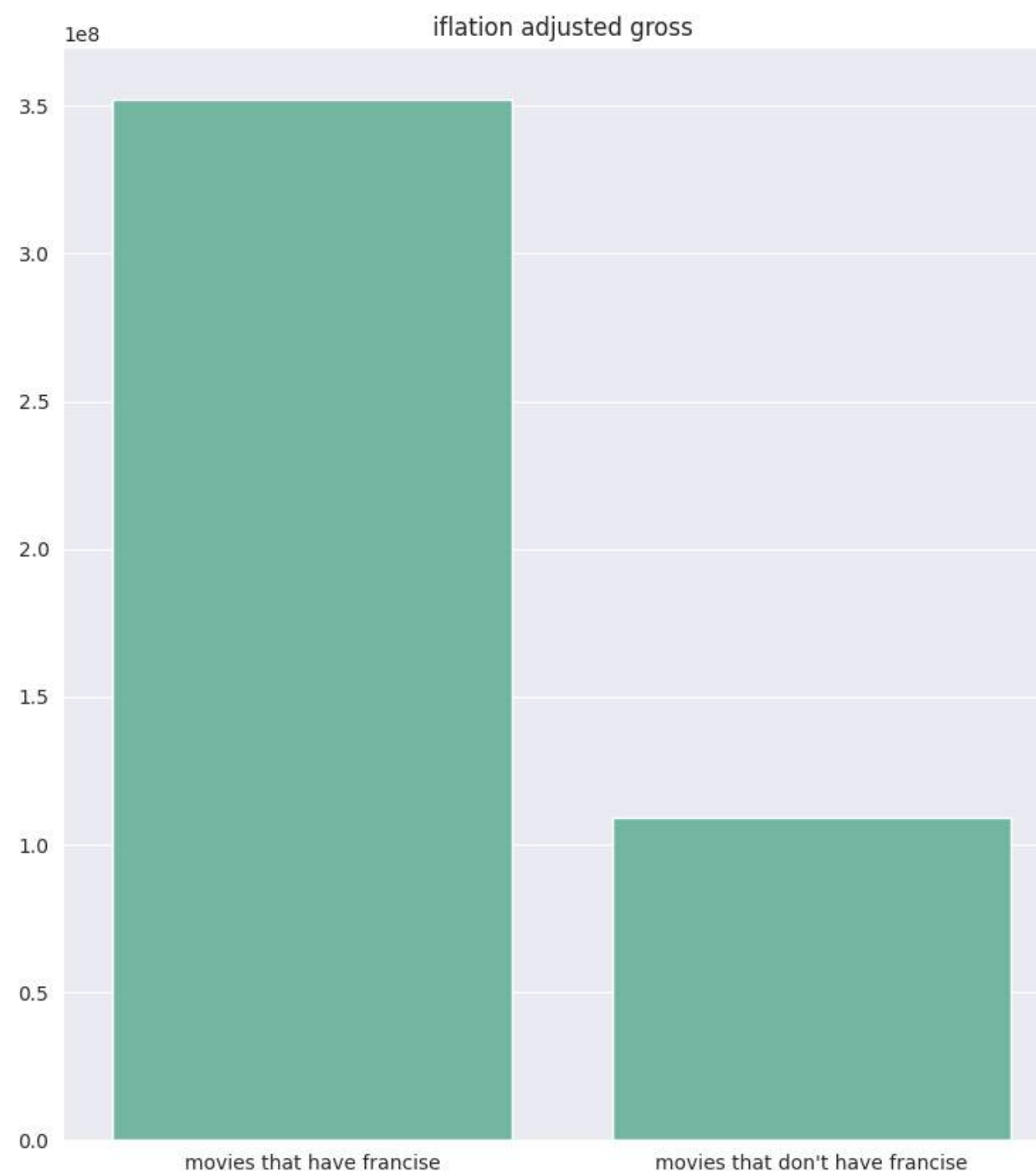
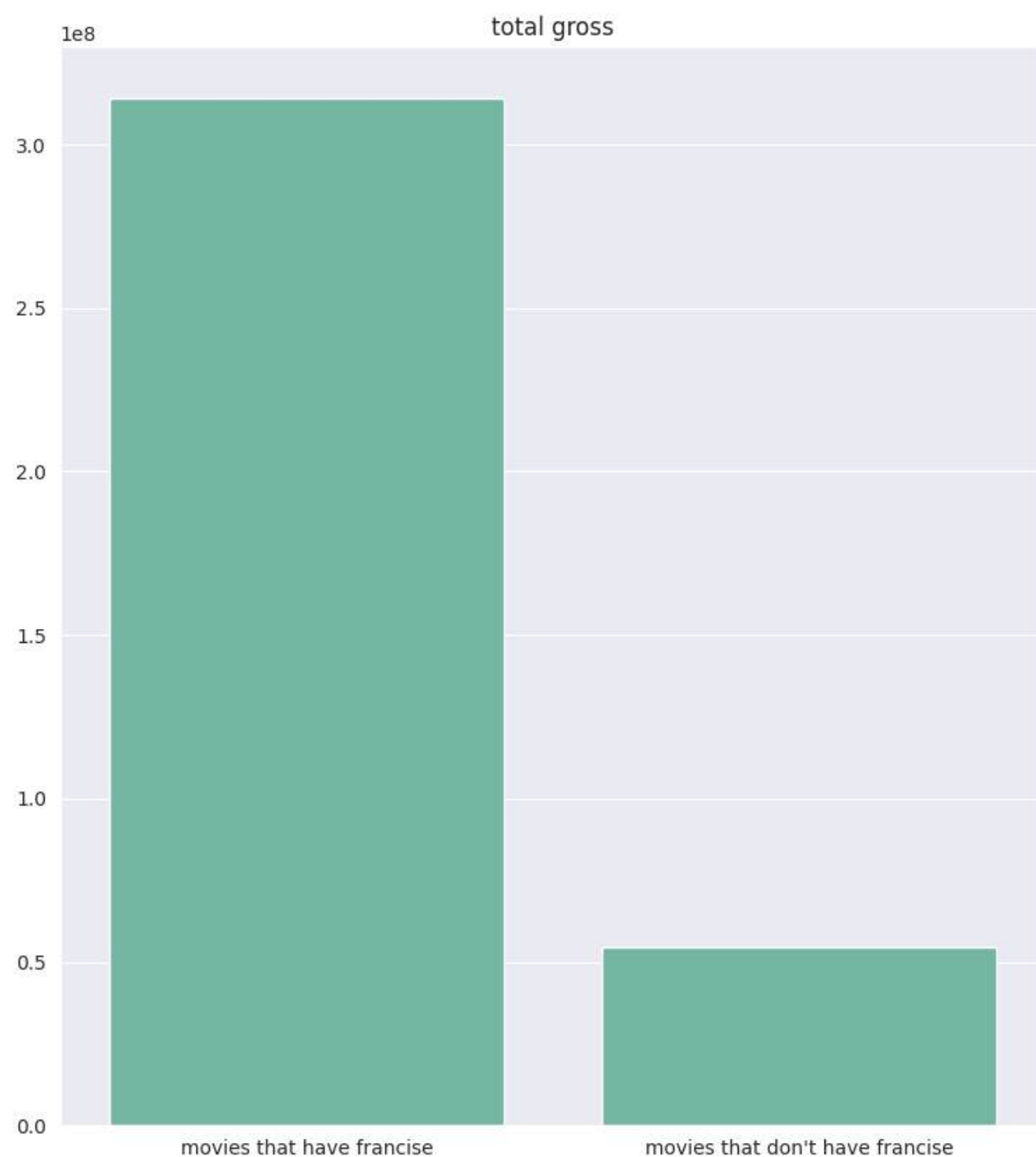
Comparación entre películas con franquicias y películas independientes.

Esta visualización analiza la diferencia de ingresos generados entre películas que forman parte de una franquicia y aquellas que no. Se presentan los datos tanto en términos absolutos (total gross) como ajustados por inflación (inflation adjusted gross), revelando el impacto sustancial de las franquicias en el éxito financiero de Disney.

```
fig, ax= plt.subplots(1,2,figsize=(20,10))
y=[]
y.append(np.sum(df[df['movie_franchise'] != 'None']['total_gross'])/len(df[df['movie_franchise'] != 'None']['total_gross']))
y.append(np.sum(df[df['movie_franchise'] == 'None']['total_gross'])/len(df[df['movie_franchise'] == 'None']['total_gross']))
x=["movies that have franchise","movies that don't have franchise"]
fig1 = sns.barplot(x=x,y=y,ax=ax[0])
fig1.set_title('total gross')
y=[]
y.append(np.sum(df[df['movie_franchise'] != 'None']['inflation_adjusted_gross'])/len(df[df['movie_franchise'] != 'None']['inflation_adjusted_gross']))
y.append(np.sum(df[df['movie_franchise'] == 'None']['inflation_adjusted_gross'])/len(df[df['movie_franchise'] == 'None']['inflation_adjusted_gross']))
x=["movies that have franchise","movies that don't have franchise"]
fig2 = sns.barplot(x=x,y=y,ax=ax[1])
fig2.set_title('inflation adjusted gross')
plt.show()
```




RESULTADO



Las películas de franquicias generan significativamente más ingresos, incluso ajustados por inflación, consolidándose como el principal motor financiero de Disney.

Distribución de Ingresos Ajustados por Género

Comparación de los géneros más lucrativos basados en ingresos ajustados por inflación.

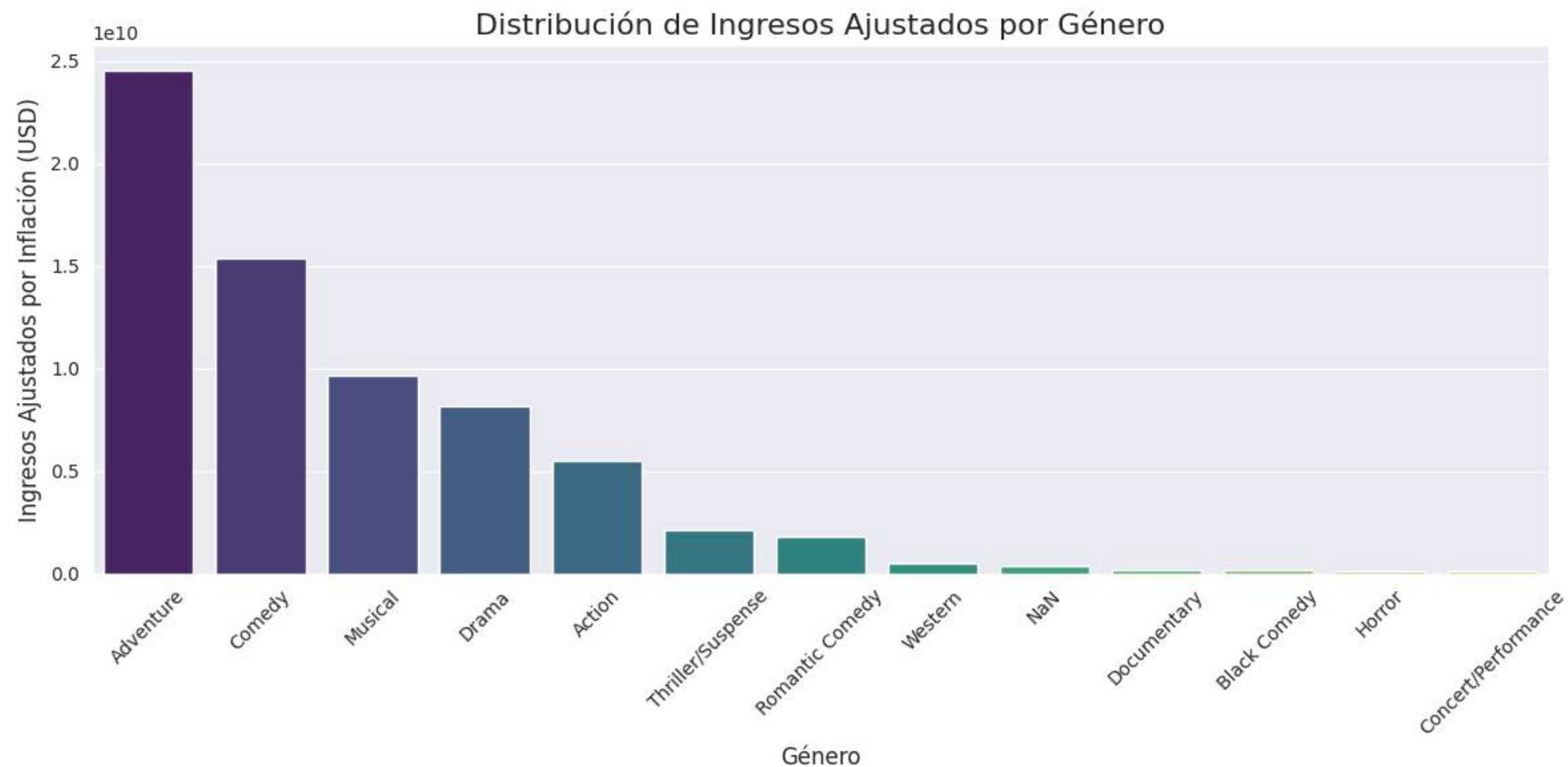
En esta gráfica se presenta el rendimiento financiero acumulado por género en ingresos ajustados por inflación. Este análisis destaca cuáles son los géneros más lucrativos en la historia de Disney, permitiendo entender las preferencias del público y las estrategias exitosas de la empresa.

```
genre_gross = df.groupby('genre')['inflation_adjusted_gross'].sum().sort_values(ascending=False)
plt.figure(figsize=(12, 6))
sns.barplot(x=genre_gross.index, y=genre_gross.values, palette='viridis')
plt.title('Distribución de Ingresos Ajustados por Género', fontsize=16)
plt.xlabel('Género', fontsize=12)
plt.ylabel('Ingresos Ajustados por Inflación (USD)', fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




RESULTADO

📌 Aventura lidera ampliamente:
El género de aventura es el que
más ingresos ha generado
ajustados por inflación,
mostrando su gran popularidad a
lo largo del tiempo.



Frecuencia de Películas por Género

Análisis de la cantidad de películas producidas por género en la historia de Disney.

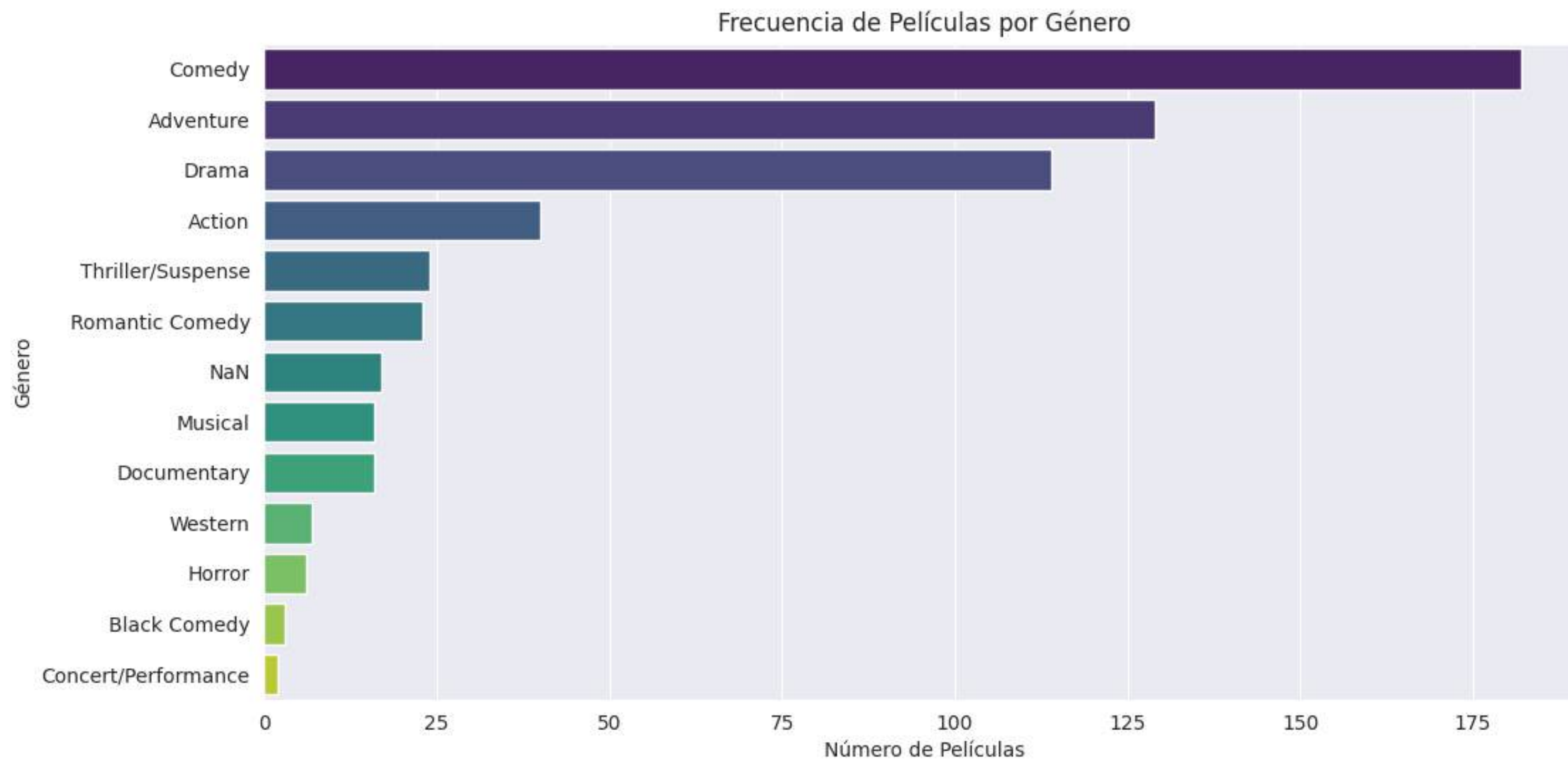
Esta gráfica presenta el número de películas producidas por Disney en cada género, permitiendo identificar cuáles son los más representados en su portafolio cinematográfico. Este análisis ayuda a entender la prioridad de la compañía en cuanto a géneros y su alineación con las preferencias del público.

```
plt.figure(figsize=(12, 6))
sns.countplot(y='genre', data=df, order=df['genre'].value_counts().index, palette='viridis')
plt.title('Frecuencia de Películas por Género')
plt.xlabel('Número de Películas')
plt.ylabel('Género')
plt.show()
```




RESULTADO

Comedia lidera en frecuencia: El género "Comedy" es el más producido, superando a otros géneros con más de 175 películas.



Recaudación Ajustada por Inflación a lo Largo de las Décadas

Análisis del desempeño económico de las películas de Disney en términos de ingresos ajustados por inflación por década.

Esta gráfica presenta los ingresos totales ajustados por inflación generados por las películas de Disney en cada década. Este análisis ayuda a evaluar el impacto financiero de sus producciones y cómo la compañía ha evolucionado a lo largo del tiempo.

```
sns.set_palette("husl")
sns.set_style("darkgrid")
plt.figure(figsize=(14, 7))
df.groupby('decade')['inflation_adjusted_gross'].sum().sort_index().plot(kind='bar', color=sns.color_palette("husl"))
plt.title("Recaudación Ajustada por Inflación a lo Largo de las Décadas", fontsize=16)
plt.xlabel("Década", fontsize=12)
plt.ylabel("Ingresos Ajustados por Inflación (USD)", fontsize=12)
plt.xticks(rotation=45)
plt.show()
```




RESULTADO

Incremento significativo a partir de los años 1990-2000: La recaudación ajustada por inflación muestra un aumento significativo a partir de esta década, destacándose como la más lucrativa.



Cantidad de Películas Producidas por Década

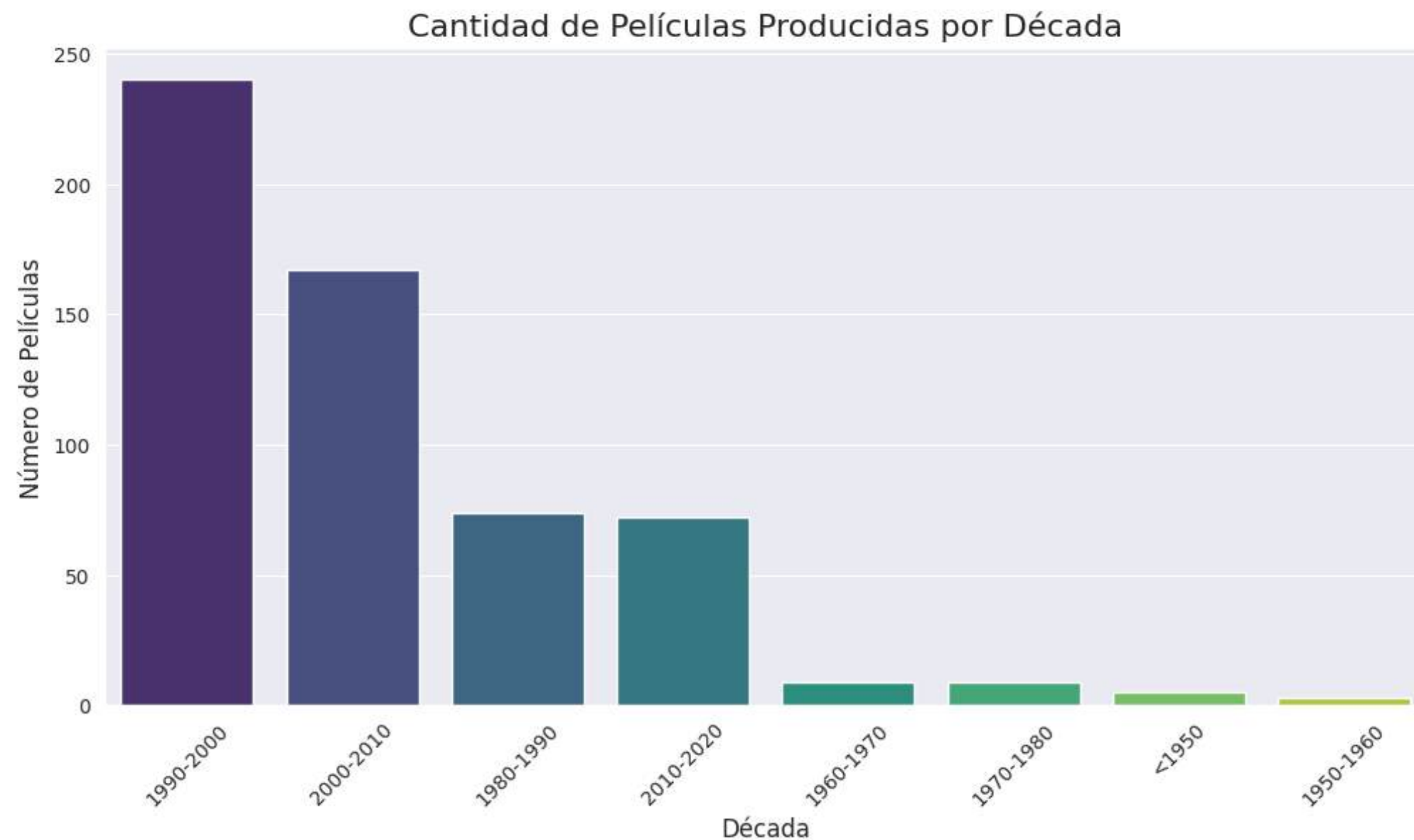
Evolución del volumen de producción cinematográfica de Disney a través de las décadas.

Este gráfico ilustra la cantidad de películas producidas por Disney en cada década, destacando las tendencias en la producción cinematográfica y la expansión de la empresa en diferentes períodos.

```
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='decade', palette='viridis', order=df['decade'].value_counts().index)
plt.title('Cantidad de Películas Producidas por Década', fontsize=16)
plt.xlabel('Década', fontsize=12)
plt.ylabel('Número de Películas', fontsize=12)
plt.xticks(rotation=45)
plt.show()
```




RESULTADO



- ♦ La década de 1990-2000 destaca como la más prolífica en términos de producción, con más de 200 películas lanzadas.
- ♦ La década de 2000-2010 también muestra una alta actividad, ocupando el segundo lugar en cantidad de películas producidas.

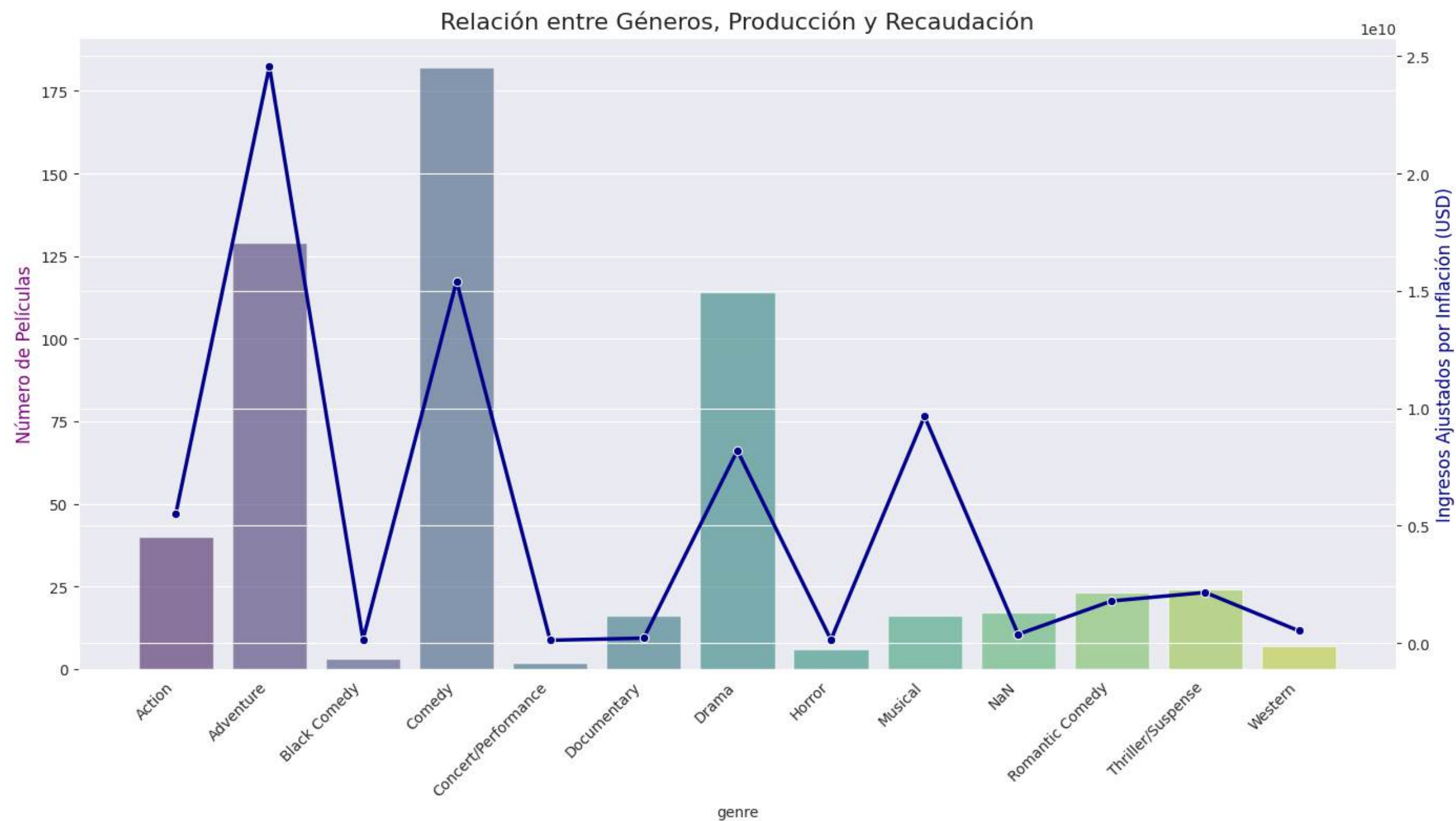
Relación entre Géneros, Producción y Recaudación

Análisis conjunto de la cantidad de películas producidas por género y su impacto en los ingresos ajustados por inflación.

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
genre_data = df.groupby('genre').agg({
    'inflation_adjusted_gross': 'sum',
    'movie_title': 'count'
}).reset_index()
genre_data = genre_data.rename(columns={'inflation_adjusted_gross': 'Total Ingresos Ajustados', 'movie_title': 'Número de Películas'})
fig, ax1 = plt.subplots(figsize=(14, 8))
sns.barplot(
    x='genre',
    y='Número de Películas',
    data=genre_data,
    palette='viridis',
    alpha=0.6,
    ax=ax1
)
ax1.set_ylabel('Número de Películas', fontsize=12, color='purple')
ax1.set_title('Relación entre Géneros, Producción y Recaudación', fontsize=16)
ax1.set_xticklabels(ax1.get_xticklabels(), rotation=45, ha='right')
ax2 = ax1.twinx()
sns.lineplot(
    x='genre',
    y='Total Ingresos Ajustados',
    data=genre_data,
    sort=False,
    marker='o',
    color='darkblue',
    linewidth=2.5,
    ax=ax2
)
ax2.set_ylabel('Ingresos Ajustados por Inflación (USD)', fontsize=12, color='darkblue')
plt.tight_layout()
plt.show()
```




RESULTADO



♦ Comedia es el género con mayor frecuencia de películas producidas, pero no lidera en ingresos ajustados por inflación.

♦ Aventura sobresale como el género con los mayores ingresos ajustados, destacándose a pesar de no ser el más producido.

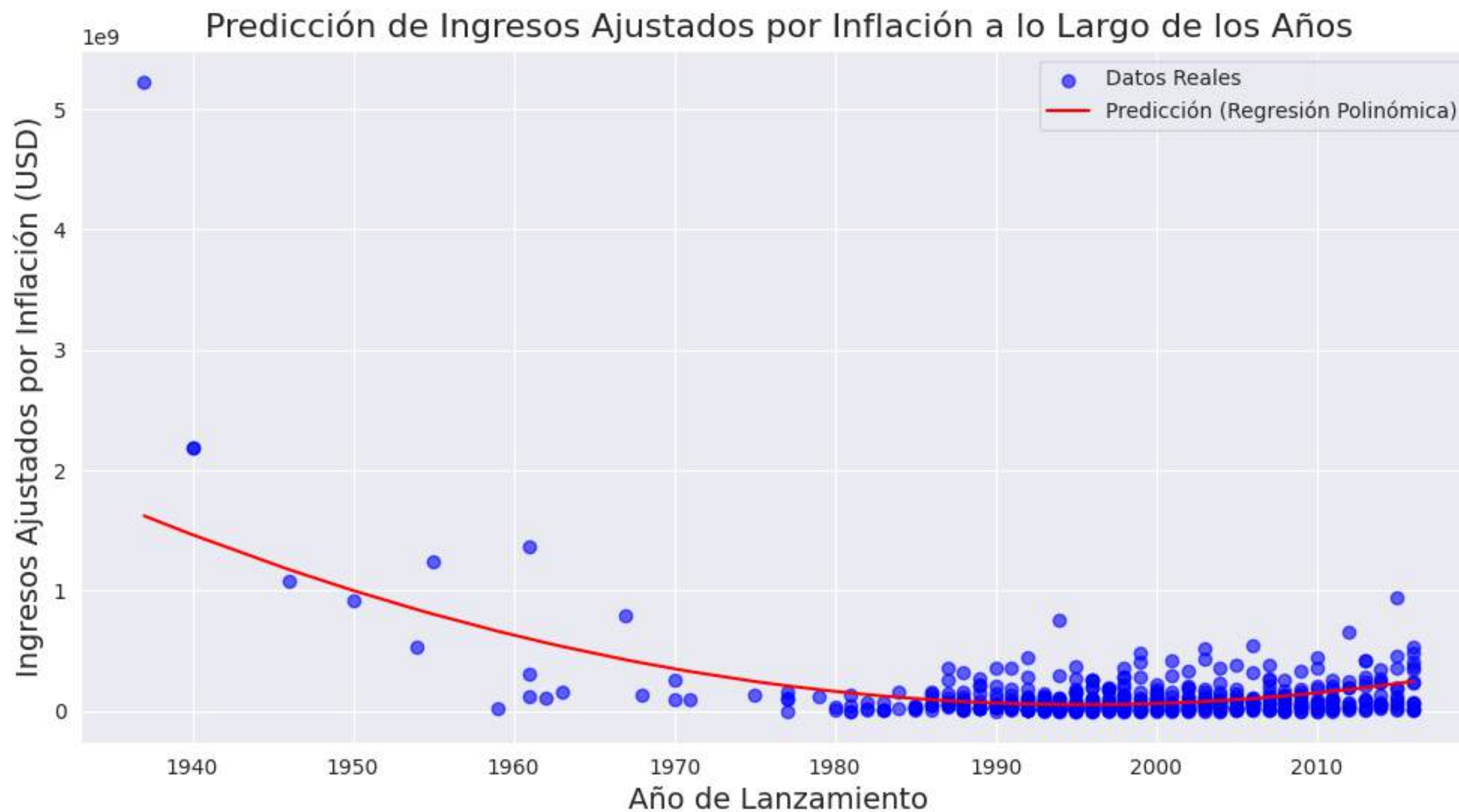
Predicción de Ingresos Ajustados por Inflación a lo Largo de los Años

En esta gráfica se muestra un modelo de regresión polinómica de grado 2 que predice la tendencia de los ingresos ajustados por inflación en función del año de lanzamiento de las películas de Disney.

```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
df_filtered = df.dropna(subset=['inflation_adjusted_gross'])
df_filtered['year'] = df_filtered['release_date'].dt.year
X = df_filtered[['year']].values
y = df_filtered['inflation_adjusted_gross'].values
X_train, X_test, y_train, y_test = train_test_split(X, y,
poly = PolynomialFeatures(degree=2)
X_train_poly = poly.fit_transform(X_train)
X_test_poly = poly.transform(X_test)
model = LinearRegression()
model.fit(X_train_poly, y_train)
y_pred = model.predict(X_test_poly)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error (MSE): {mse}")
print(f"R2 Score: {r2}")
plt.figure(figsize=(12, 6))
plt.scatter(X, y, color='blue', alpha=0.6, label="Datos Reales")
plt.plot(np.sort(X, axis=0), model.predict(poly.transform(np.sort(X, axis=0))), color='red', label="Predicción (Regresión Polinómica)")
plt.title("Predicción de Ingresos Ajustados por Inflación a lo Largo de los Años", fontsize=16)
plt.xlabel("Año de Lanzamiento", fontsize=14)
plt.ylabel("Ingresos Ajustados por Inflación (USD)", fontsize=14)
plt.legend()
plt.grid(True)
plt.show()
```





RESULTADO





OBSERVACIONES



 Tendencia decreciente en los ingresos ajustados por inflación:

Los ingresos ajustados por inflación tienden a disminuir para las películas más antiguas a medida que avanzan las décadas, especialmente antes de 1980.

 Incremento en la dispersión desde 1990:

A partir de los años 1990, aunque el modelo predice ingresos más bajos en promedio, hay una mayor variación, con algunas películas obteniendo ingresos significativamente altos.

 Relación no lineal capturada:

El modelo de regresión polinómica captura una relación no lineal que explica cómo las películas más recientes se desvían del patrón decreciente.

 Impacto de películas antiguas icónicas:

Las películas de antes de 1950, como "Snow White", destacan con ingresos excepcionalmente altos incluso después de ajustar por inflación, elevando la curva inicial.

 Limitación del modelo:

Aunque el modelo captura bien la tendencia general, la dispersión en los datos recientes sugiere que factores adicionales como género, presencia de franquicias y estrategias de marketing influyen significativamente en los ingresos.



MUCHAS GRACIAS



Erik Bañez Rojas

