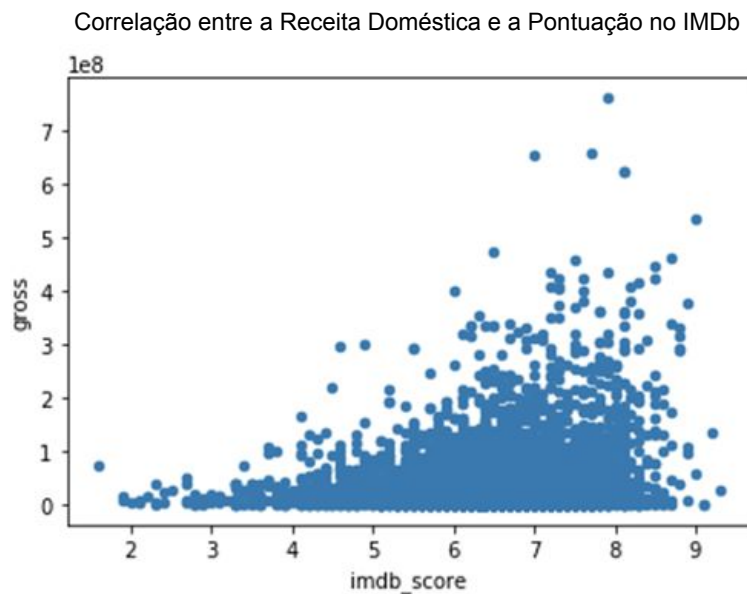
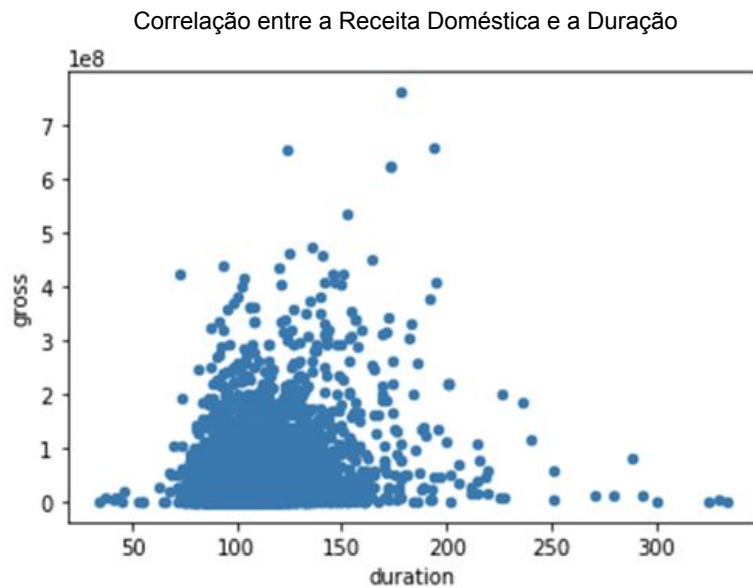


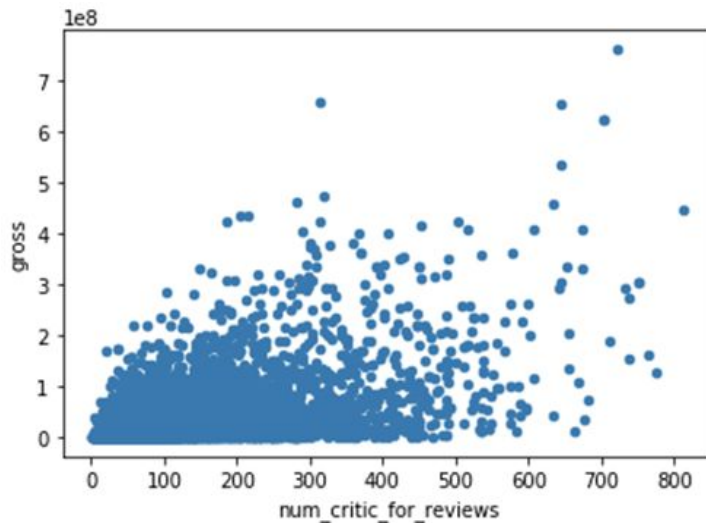
Lista 3 – Introdução ao Python

Autores: Lucas Prosdocimi, Nicolas Seiji, Marco Brum, Pedro Nader

- **Pergunta 1: Em geral, existe algum fator que dentre os seguintes: duração, avaliação do IMDb e número de críticas para revisão; impacta, ou seja, apresenta correlação com a receita doméstica dos filmes?**



Correlação entre a Receita Doméstica e o Número de Críticas para revisão

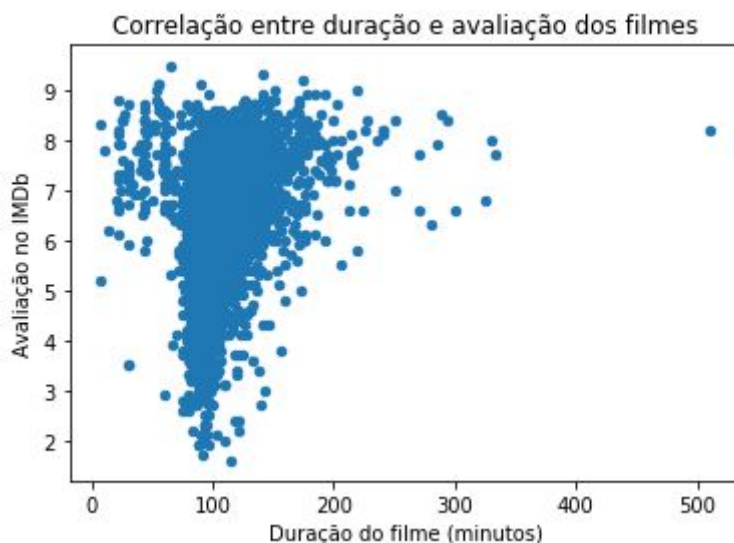


Considerando-se os 3 gráficos acima, os quais apresentação a correlação da receita doméstica com a duração, a pontuação no IMDb e o número de críticas para revisão, pode concluir que aqueles filmes que apresentam as maiores receitas domésticas, também possuem maior pontuação no IMDb.

Já em relação aos outros fatores, não é possível observar uma correlação direta, de maneira a caracterizar um impacto na receita doméstica.

- **Pergunta 2: Existe alguma correlação da duração dos filmes com sua avaliação pelo público? Hipótese: quanto maior o filme, mais "chato ou não o público acharia"**

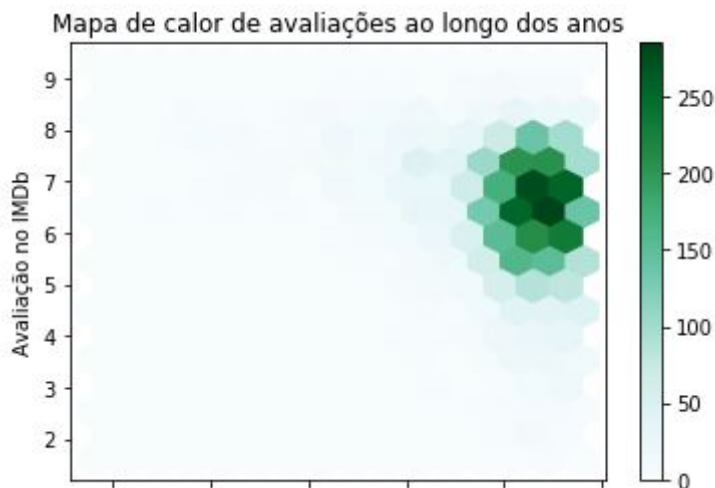
```
df.plot.scatter(x='duration',y='imdb_score')  
plt.title('Correlação entre duração e avaliação dos filmes')  
plt.xlabel('Duração do filme (minutos)')  
plt.ylabel('Avaliação no IMDb')
```



A partir do gráfico gerado, não é possível afirmar que há uma correlação clara entre a duração de um filme e sua avaliação no IMDb, uma vez que grande parte dos filmes encontram-se na faixa de 100 minutos e tem avaliações variando da faixa de 2 até 9 aproximadamente.

- **Pergunta 3: como se comporta as avaliações do IMDb ao longo do tempo?**

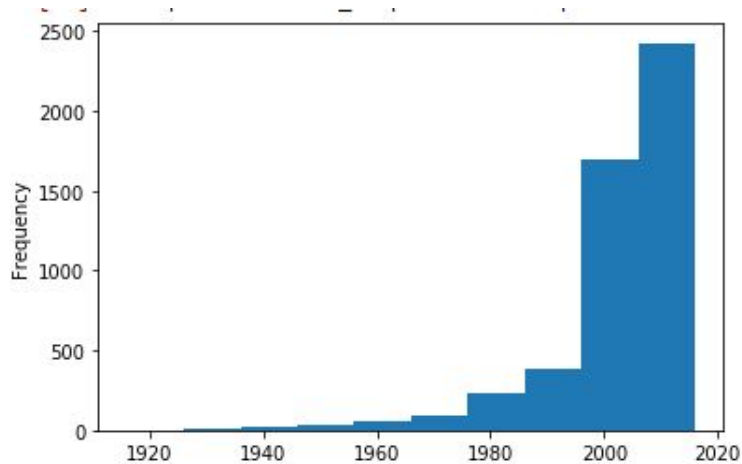
```
df.plot.hexbin(x='title_year', y = 'imdb_score', gridsize=15)
plt.title('Mapa de calor de avaliações ao longo dos anos')
plt.xlabel('Anos')
plt.ylabel('Avaliação no IMDb')
```



Percebe-se que as avaliações são mais frequentes para os anos mais recentes, seja por uma popularização da plataforma, seja pela criação da plataforma ser de anos mais recentes, o que faz com que os títulos mais antigos sejam menos avaliados.

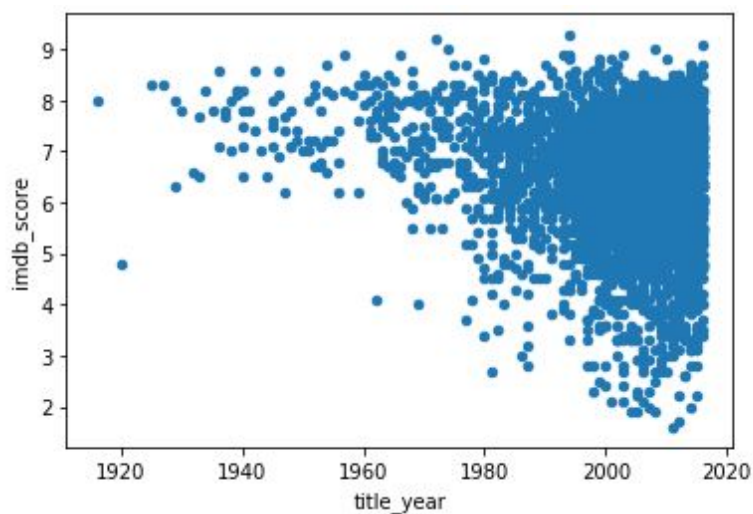
- **Pergunta 4: Os filmes têm tido pontuações no IMDb melhores com o passar dos anos?**

```
df['title_year'].plot.hist()
```



Primeiro, é possível perceber que a base contém muito mais filmes recentes do que antigos.

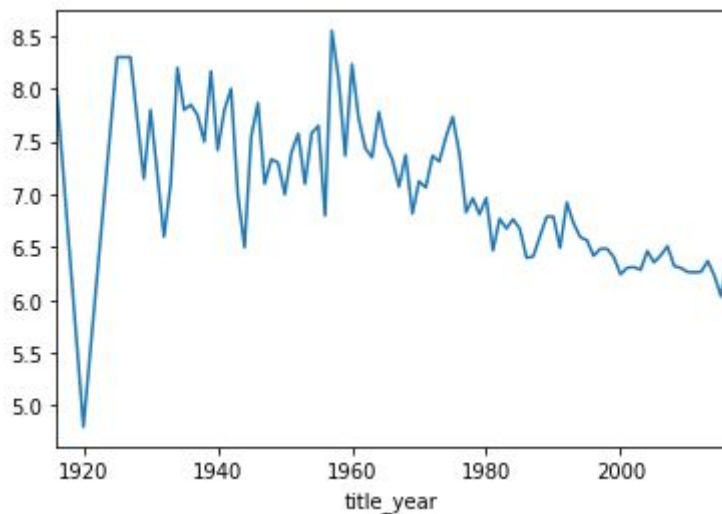
```
df.plot.scatter(x='title_year',y='imdb_score')
df.corr()['title_year'].sort_values()
```



```
imdb_score          -0.209167
duration            -0.135038
director_facebook_likes -0.063820
num_user_for_reviews -0.003147
num_voted_users      0.007397
gross               0.030886
budget              0.045726
facenumber_in_poster 0.061504
actor_1_facebook_likes 0.086873
actor_3_facebook_likes 0.096137
actor_2_facebook_likes 0.101890
cast_total_facebook_likes 0.109971
aspect_ratio        0.159973
movie_facebook_likes 0.218678
num_critic_for_reviews 0.275707
title_year           1.000000
Name: title_year, dtype: float64
```

Com esse gráfico e a análise de correlação, não aparenta possuir nenhuma correlação expressiva, mas é possível perceber um comportamento de expansão da pontuação do IMDb conforme passam os anos. Portanto, vamos verificar a média da pontuação do IMDb ao longo dos anos para verificar se há uma tendência de comportamento

```
agrupado=df.groupby(['title_year']).mean()  
agrupado['imdb_score'].plot()
```

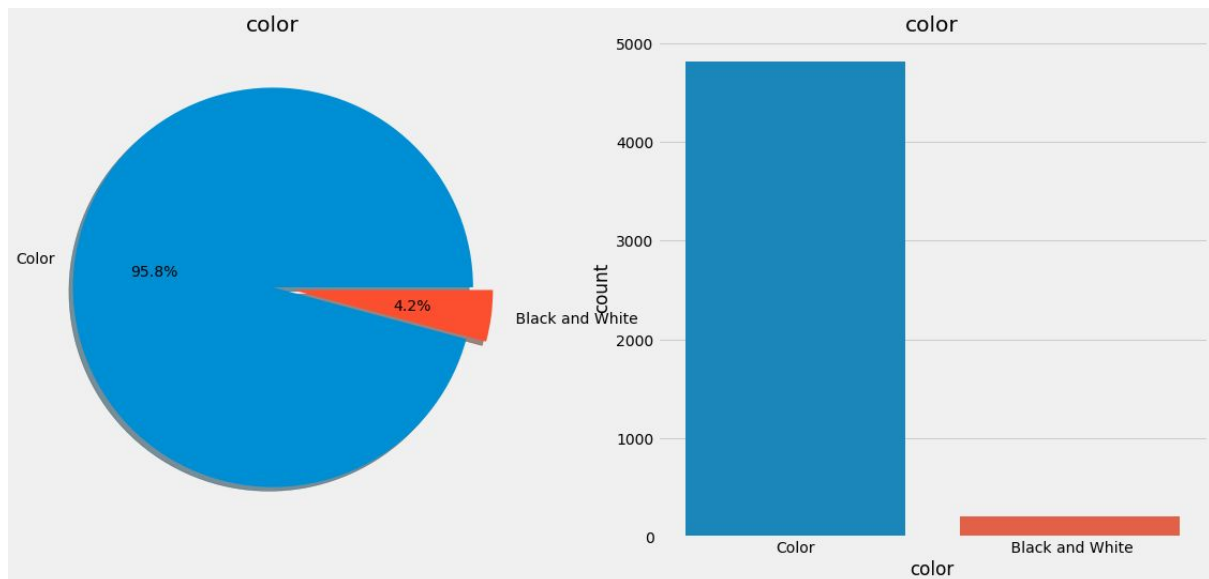


Quando se agrupa a partir do ano de lançamento, a média da pontuação no IMDb abaixa consideravelmente. Portanto, embora tenha variações, há constância de decréscimo da média ao longo dos anos, o que responde a pergunta acima de forma negativa. Vale ressaltar que isso se dá a partir da base de dados que possuímos, que provavelmente selecionou apenas bons filmes antigos. Isso não prova que filmes mais antigos eram melhores, por exemplo.

- **Pergunta 5: Existe algum país com uma cultura mais forte de realizar filmes em preto e branco?**

color	country	
Black and White	Australia	1
	Canada	6
	China	3
	Denmark	1
	France	8
	Germany	7
	Hong Kong	2
	India	1
	Ireland	1
	Italy	1
	Japan	2
	Kenya	1
	Libya	1
	Poland	1
	Russia	1
	South Africa	1
	Soviet Union	1
	Switzerland	1
	UK	20
	USA	149

Conforme a tabela acima, é possível observar que, em geral, os países realizam poucos (esporádicos) filmes em preto e branco. Uma exceção à regra é o UK (Reino Unido) que apresenta quase 15% dos filmes realizados sem coloração.



Conforme pode ser observado em gráfico acima, apenas 4,2% dos filmes catalogados no IMDB possuem a coloração P&B.

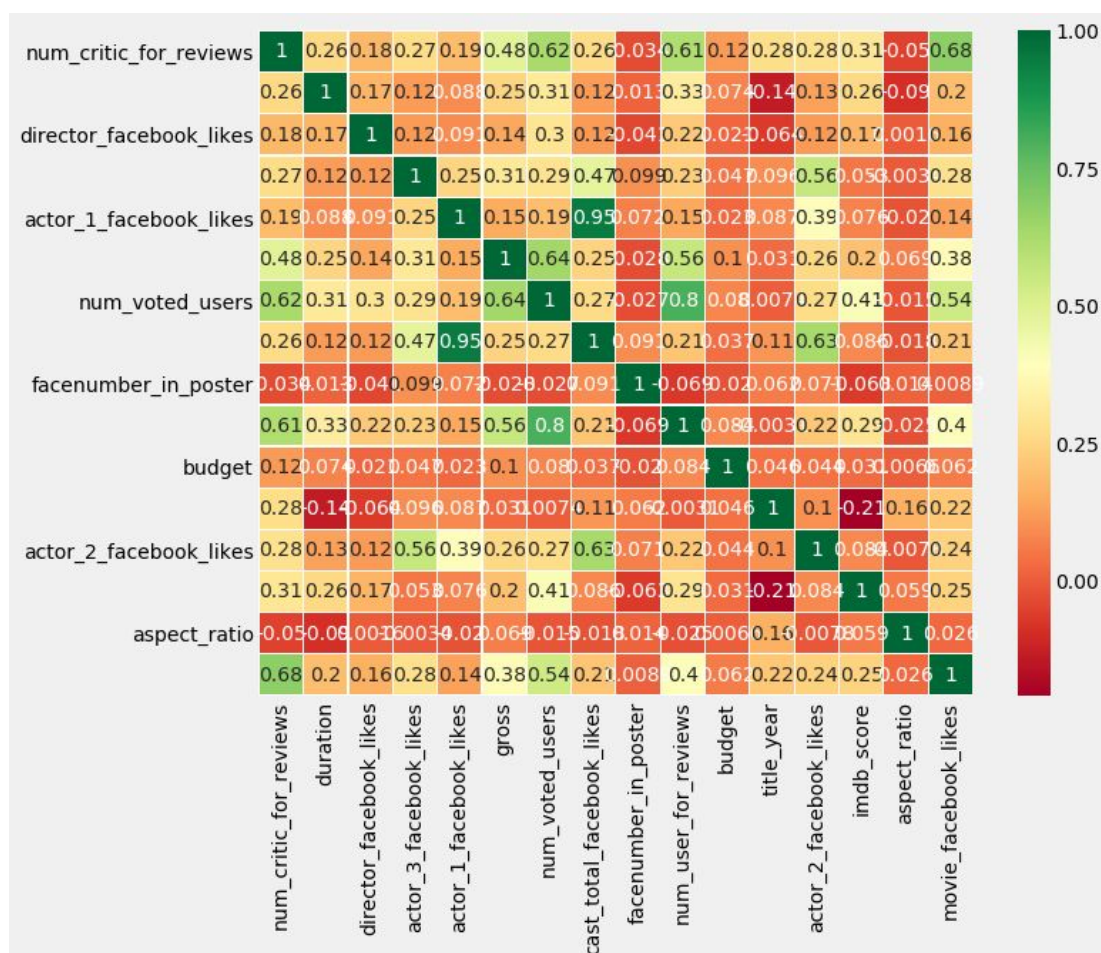
- **Pergunta 6: Em geral, como se comporta a distribuição do tempo duração dos filmes?**

```
print('Filme com maior duração foi:',data['duration'].max(), 'minutos')
print('Filme com menor duração foi de:',data['duration'].min(), 'minutos')
print('Média de duração dos filmes é de',data['duration'].mean(), 'minutos')
```

```
Filme com maior duração foi: 511.0 minutos
Filme com menor duração foi de: 7.0 minutos
Média de duração dos filmes é de 107.2010739856802 minutos
```

Apesar da média de duração dos filmes ser de 107 minutos (1h 47min) existem filmes de curta metragem com duração de 7 minutos e outros de extrema longa metragem com duração de mais de 8h.

- **Pergunta 7: Quais dados apresentados na tabela possuem maior correlação entre si? Que conclusões podem ser tiradas a respeito?**

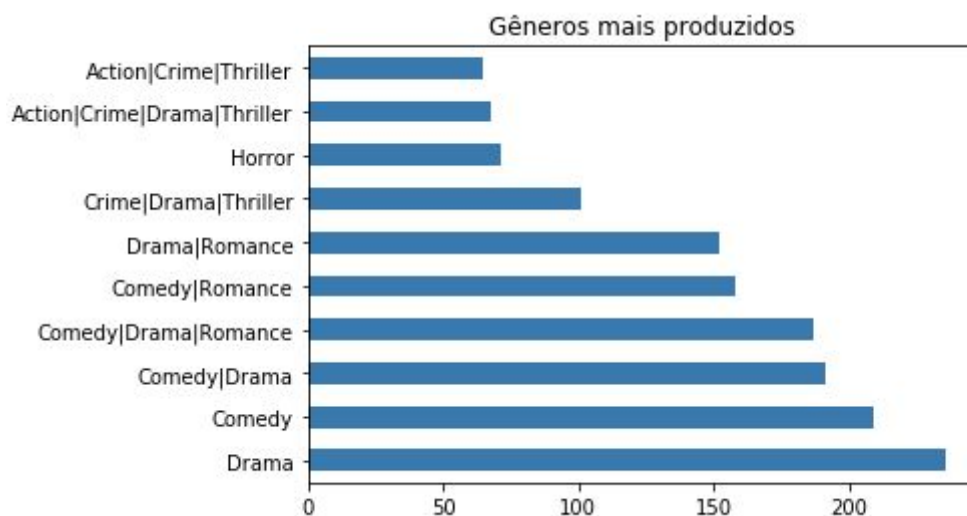


A partir do gráfico acima é possível tirar diversas conclusões. Uma delas é a alta correlação entre a receita do filme e o número de usuários que votaram na avaliação. Essa correlação faz sentido pois um filme com maior receita consequentemente teve mais pessoas o assistindo. Logo, o número de pessoas que fazem a avaliação do filmes tende a ser maior.

Outra correlação interessante é entre o número de votos de avaliação e o número de likes do filmes no Facebook. A conclusão que pode ser tirada é a de que, pessoas que votam na avaliação do filmes são mais presentes e ativas na internet e nas redes sociais. Logo, não só avaliam o filme no IMDB como interagem com ele em outras redes sociais, como o Facebook. Documen

- **Pergunta 8: Pode-se dizer que o gênero de filmes "Documentário" é o mais produzido ao longo da história?**

```
df['genres'].value_counts()  
(df['genres'].value_counts())[:10].plot('barh');  
plt.title('Gêneros mais produzidos');
```



Segundo o gráfico acima, não pode-se dizer que o gênero "Documentário" foi o mais produzido, já que este nem sequer aparece no Top 10 Gêneros mais feitos, o qual inclusive apresenta em algumas posições, uma mesclagem de gêneros. Além disso, percebe-se também, que os gêneros "Drama" e "Comédia" lideram no aspecto abordado.