

Respostas

1. Faça uma análise exploratória dos dados (EDA), demonstrando as principais características entre as variáveis e apresentando algumas hipóteses relacionadas. Seja criativo! ([analise_exploratoria_de_dados.ipynb](#))
2. Responda também às seguintes perguntas:
 - a. Qual filme você recomendaria para uma pessoa que você não conhece? **Para as pessoas que eu não conheço, facilmente recomendaria filmes de diretores que estão no quadrante superior direito, ou seja alto faturamento e uma boa crítica do público.**
 - b. Quais são os principais fatores que estão relacionados com alta expectativa de faturamento de um filme? **Gênero, Diretor, Crítica são os principais responsáveis pelo faturamento, pois são eles que levam até o público a vontade de ir ao cinema cada um da sua forma, por exemplo eu iria ao cinema ver qualquer filme que o Tim Burton for lançar simplesmente por ser fã do trabalho dele.**
 - c. Quais insights podem ser tirados com a coluna *Overview*? É possível inferir o gênero do filme a partir dessa coluna? **Sim, é possível mesmo não me aprofundando em provar esse ponto é possível ver que existem palavras chaves, vou dar um exemplo para o gênero de ação a palavra "War" se destaca em várias Overview mostrando que está ligada ao gênero.**
3. Explique como você faria a previsão da **nota do imdb** a partir dos dados. Quais variáveis e/ou suas transformações você utilizou e por quê? Qual tipo de problema estamos resolvendo (regressão, classificação)? Qual modelo melhor se aproxima dos dados e quais seus prós e contras? Qual medida de performance do modelo foi escolhida e por quê? **É um modelo de regressão pois temos que identificar um valor e não classificar em um grupo. Foi utilizado as variáveis Genre, Runtime, Meta_score, Released_Year que foi transformado na idade do filme, No_of_Votes, Gross. O modelo escolhido foi o LightGBM se aproxima dos dados deste problema. Tem como pró a velocidade e como contra em datasets pequenos ele corre o risco de acabar**

decorando e não aprendendo. R^2 (R-squared - Coeficiente de Determinação) onde seu valor vai mostrar a porcentagem da influência da nota pelas variáveis que utilizamos.

4. Supondo um filme com as seguintes características:

```
{ 'Series_Title': 'The Shawshank Redemption',  
  'Released_Year': '1994',  
  'Certificate': 'A',  
  'Runtime': '142 min',  
  'Genre': 'Drama',  
  'Overview': 'Two imprisoned men bond over a number of years,  
finding solace and eventual redemption through acts of common  
decency.',  
  'Meta_score': 80.0,  
  'Director': 'Frank Darabont',  
  'Star1': 'Tim Robbins',  
  'Star2': 'Morgan Freeman',  
  'Star3': 'Bob Gunton',  
  'Star4': 'William Sadler',  
  'No_of_Votes': 2343110,  
  'Gross': '28,341,469' }
```

Qual seria a nota do IMDB? O IMDB Rating previsto pelo modelo para The Shawshank Redemption é: 8.69