

Etapa 3

Perfil Cientista de Dados



Compartilhamos um cenário (hipotético) sobre aquisição de um conteúdo. O objetivo principal é responder quanto aproximadamente vale um determinado conteúdo (custo de licença), considerando o catálogo apresentado na base anexa, servindo de bússola para a decisão de se seguir com a compra.

Você possui em mãos uma base com ~200 títulos (conteúdos) já disponíveis na plataforma de streaming Globoplay. Estamos no momento de planejar a aquisição de novos conteúdos para 2023, realizando uma árdua pesquisa de mercado e definindo os limites orçamentários para o ano.

Com os dados desta base, apresente um modelo estatístico que seja capaz de determinar intervalos de valores possíveis para o custo de licença de cada um dos conteúdos abaixo:

Title	Genre	Release Year	Duration Minutes	Country	\$ Prod
Admirer Of The Solstice	Romance	2011	95	Índia	97,2 MM
Electricity Can Sing	Humor	2018	116	Suíça	63,5 MM
Annual Women	Nonfiction	2017	129	Alemanha	30,4 MM
Pigs Of Silver	Children	2011	123	Estados Unidos	83,3 MM
Mystery In The Dark	Horror	2017	89	Reino Unido	110,4 MM
Doctors Without Shame	Adventure	2020	105	Cingapura	14,7 MM
Symbols Of Space	Sci-Fi	2015	88	Países Baixos	97,3 MM

Após a exploração:

1. apresente a modelagem;
2. explique a adoção do modelo (porque escolheu esse modelo);
3. apresente o resultado do modelo;
4. aponte as limitações do modelo adotado e modelagens alternativas;
5. responda a pergunta de negócio de forma amigável.

Todos os dados apresentados são fictícios e devem ser utilizados da maneira como foram apresentados. Caso identifique algum possível problema de data quality, identifique-o e apresente solução de contorno, seguindo o teste.