

YouTube Trends

Link de referência: <https://www.kaggle.com/datasnaek/youtube-new>

UPDATE: código-fonte usado para coletar esses dados [divulgados aqui](#)

O YouTube (o site de compartilhamento de vídeos mundialmente famoso) mantém uma lista dos [principais vídeos de tendências](#) na plataforma. De acordo com a revista *Variety*, “para determinar os vídeos mais populares do ano, o YouTube usa uma combinação de fatores, incluindo a medição das interações dos usuários (número de visualizações, compartilhamentos, comentários e curtidas). Observe que eles não são os vídeos mais vistos em geral no ano civil”. Os principais artistas da lista de tendências do YouTube são vídeos de música (como o famoso "Gangnam Style"), performances de celebridades e / ou reality shows e vídeos virais de cara com câmera que o YouTube é bem conhecido.

Este conjunto de dados é um registro diário dos principais vídeos populares do YouTube.

Observe que esse conjunto de dados é uma versão estruturalmente aprimorada [desse conjunto de dados](#).

Inspiração

Os usos possíveis para esse conjunto de dados podem incluir:

- Análise de sentimentos de várias formas
- Categorizando vídeos do YouTube com base em seus comentários e estatísticas.
- Treinar algoritmos de ML como RNNs para gerar seus próprios comentários no YouTube.
- Analisando quais fatores afetam a popularidade de um vídeo do YouTube.
- Análise estatística ao longo do tempo. (modelos regressivos, ARMA, previsões)

Para mais inspiração, consulte os kernels neste conjunto de dados!

Conteúdo

Esse conjunto de dados inclui vários meses (e contando) de dados em vídeos diários de tendências do YouTube. Os dados são incluídos para as regiões dos EUA, GB, DE, CA e FR (EUA, Grã-Bretanha, Alemanha, Canadá e França, respectivamente), com até 200 vídeos de tendências listados por dia.

EDIT: agora inclui dados das regiões RU, MX, KR, JP e IN (Rússia, México, Coreia do Sul, Japão e Índia, respectivamente) no mesmo período.

Os dados de cada região estão em um arquivo separado. Os dados incluem o título do vídeo, o título do canal, o tempo de publicação, tags, visualizações, curtidas e não curtidas, descrição e contagem de comentários.

Os dados também incluem um `category_id` campo, que varia entre as regiões. Para recuperar as categorias de um vídeo específico, localize-o no associado JSON. Um desses arquivos é incluído para cada uma das cinco regiões no conjunto de dados.

Para obter mais informações sobre colunas específicas no conjunto de dados, consulte os [metadados](#) da [coluna](#) .