

# Sumário

1. Contextualização
2. Introdução
3. Referencial Teórico
4. Metodologia
5. Resultados
6. Discussão e conclusões
7. Ameaças a validade

# Contextualizando

- O universo musical passou por diversas mudanças no decorrer dos anos.
- Um aumento em quantidade de empresas que buscam encontrar artistas para investir.
- Um meio, atual, para esses empresários observarem são as plataformas de streaming.
- A plataforma de streaming Youtube foi ganhando uma grande quantidade de usuários por sua facilidade em gerar conteúdo.



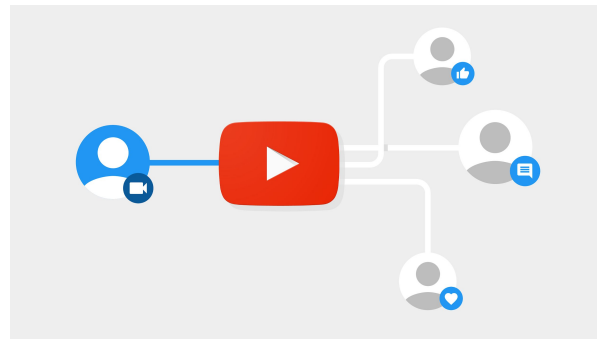
## Tecnologia em música

(<https://artelinguagemcomunicacao.blogspot.com>)

# Introdução

- Auxiliar à empresários encontrar e observar artistas na plataforma de streaming da Google (YouTube).
- Técnicas de mineração de dados (Data Mining), a fim de responder a questão de pesquisa ao lado.
- Realização de um pré-processamento e análise dos dados.
- Para determinar esses fatores, foi utilizado uma técnica chamada de **regressão linear** nos dados para obter informações relevantes, referente a audiência em um vídeo, logo, determinando os fatores para o canal de sucesso.

Q1: Quais os fatores que são determinantes na definição de um canal de sucesso?

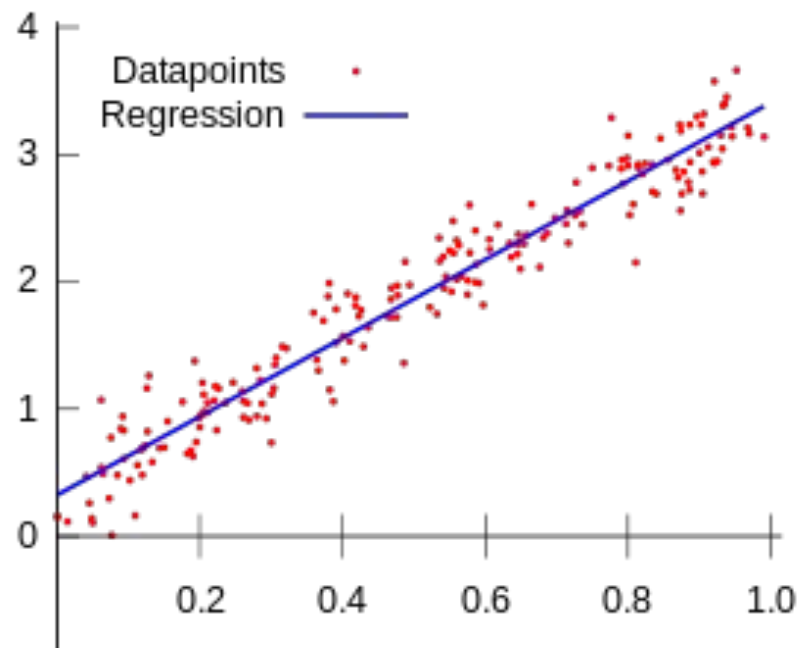


Youtube Data API

(<https://developers.google.com/youtube/v3>)

# Referencial Teórico

- Regressão linear, é realizada através de vários pontos, a fim de se obter uma regressão que possa prever o ideal entre labels pré estabelecidas.
- Data Analytics Youtube (google), uma ferramenta para conseguir extrair dados/informações da plataforma de streaming.
- O poder da mineração de dados (Data mining) para a realização de possíveis previsões no nosso cotidiano.



# Metodologia

- Plataforma escolhida: Youtube. Pois apresenta uma quantidade enorme de artistas no ramo musical.
- A API utilizada para captura de dados foi a da própria ferramenta (apresentada no slide a seguir).
- A API, mesmo sendo uma ferramenta com vários recursos para captura dos dados, ela apresenta um sistemas de “quota” diária para a captura dos dados.

# Youtube API

Overview

- Activities
- Captions
- ChannelBanners
- Channels
- ChannelSections
- Comments
- CommentThreads
- GuideCategories
- PlaylistItems
- Playlists
- Search
- Subscriptions
- Thumbnails
- VideoAbuseReportReasons
- VideoCategories
- Videos
- Watermarks
- Parâmetros de consulta padrão
- Erros

## Parâmetros

A tabela a seguir lista os parâmetros que esta consulta suporta. Todos os parâmetros listados são os parâmetros de consulta.

### Parâmetros

#### Parâmetros obrigatórios

**part**

**string**

O parâmetro **part** especifica uma lista separada por vírgulas de uma ou mais propriedades de recurso **search** que serão incluídas pela resposta da API. Os nomes **part** que podem ser incluídos no valor do parâmetro são **id** e **snippet**.

Se o parâmetro identificar uma propriedade que contenha propriedades criança, estas serão incluídas na resposta. Por exemplo, em um resultado **search**, a propriedade **snippet** contém outras propriedades que identificam o título, a descrição, entre outros, do resultado. Se você definir **part=snippet**, a resposta da API também conterá todas essas propriedades aninhadas.

**Filtros** (especifique exatamente um dos seguintes parâmetros)

# Metodologia - Coleta dos dados

- Seleção de canais:
- Captura de dados desses canais durante duas semanas
- Através de parâmetros, escolhemos as informações úteis.

```
    },  
    "statistics": {  
      "viewCount": "2322054",  
      "likeCount": "43272",  
      "dislikeCount": "485",  
      "favoriteCount": "0",  
      "commentCount": "1374"  
    }  
  }  
]  
}
```

- Boyce Avenue
- Mariana Nolasco
- Ana Gabriela
- Sofia Karlberg
- Cimorelli the band
- Gabi Luthai
- Daniela Sings
- Joana Castanheira
- Carina Mennitto

# Metodologia - Tratamento dos dados

## Canais:

- Id canal
- Quantidade de inscritos
- Quantidade vídeos publicados
- Quantidade total de visualizações

## Vídeos:

- Id do vídeo
- Quantidade de visualizações
- Quantidade de like
- Quantidade de não like
- Quantidade de comentários
- Id do canal



# Metodologia - Armazenamento

- Dados foram cadastrados em tabelas (utilizando banco de dados MySQL).
- Capturados através de um algoritmo criado de requisições.
- Hospedado em um servidor em nuvem, serviços da Digital Ocean.
- Para análise os dados foram capturados e analisados em computadores locais.

# Metodologia - Regras e padrões

## Regressão

Atributos	Descrição
Id do vídeo	Representa qual vídeo tem as informações abaixo
Média visualizações	Média de visualizações
Desvio padrão visualizações	Desvio padrão de visualizações
Média like	Média de like
Desvio padrão like	Desvio padrão de like
Média não like	Média de não like
Desvio padrão não like	Desvio padrão de não like
Média comentários	Média de comentários
Desvio padrão comentários	Desvio padrão de comentários
Inscritos no canal	Quantidade de inscritos no canal (na última captura)

# Metodologia - Regras e padrões

## Canais:

- Quantidade de inscritos (da última captura dos dados sobre aquele canal)

## Regressão:

- De acordo com o conjunto de valores passado para a regressão, determinar a audiência pela quantidade de inscritos no canal.

## Vídeos:

- Para cada dia, foi realizado uma média dos valores para resultar em um vídeo por id em um determinado dia.
- Realização de uma linearização dos valores para trabalharmos com a grandeza dos números, e não valores brutos.
- Realização de uma normalização dos dados, para um melhor desempenho e resultado da regressão utilizada

# Resultados

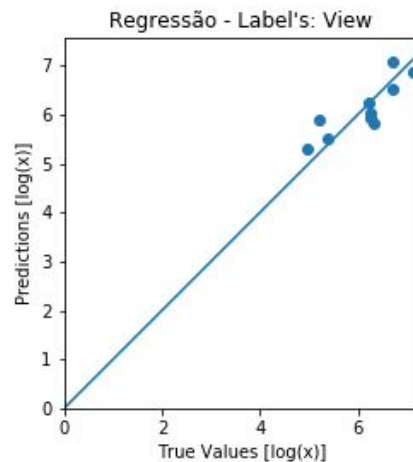
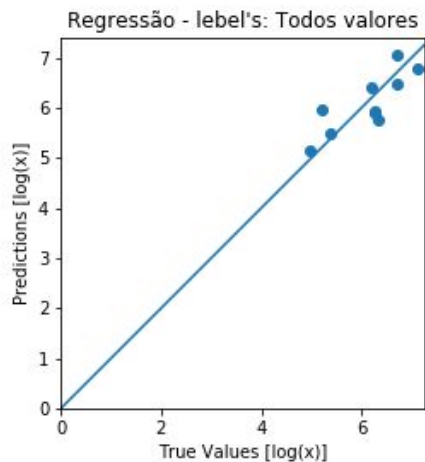
- Os dados, citados anteriormente, que foi separado para instância de teste e de treino, sendo separado 20% para testes e 80% para treinamento.
- Utilizando scikit-learn, foi realizada a regressão nessa base e como resultado foi obtido os parâmetros mais importantes para a regressão.

Position	Parametro	Coefficiente para regressão
2	view_mean	0.3732200360586011
0	likes_mean	0.1678438286517744
7	dislike_std	0.15496454397202333
6	dislike_mean	0.05756773420383723
1	likes_std	0.027823830633219138

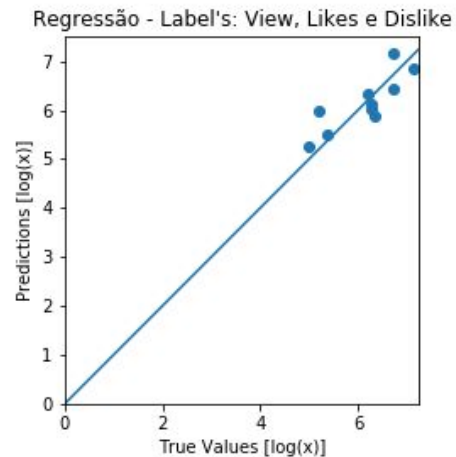
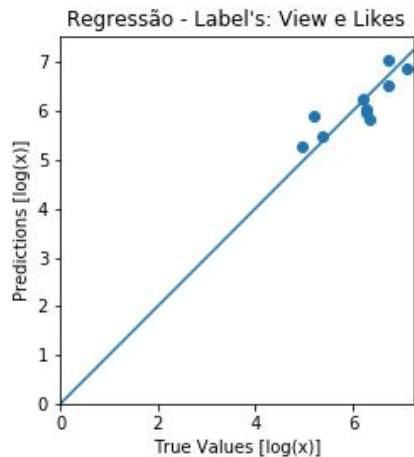
# Resultados

- Foi selecionado os 3 primeiros coeficientes de média e foi realizado novamente a regressão para esses valores para conseguir determinar os principais parâmetros encontrados.
- Com eles conseguimos estipular o quanto aquele parâmetro ajuda na previsão para a audiência daquele canal

# Resultados



# Resultados



# Resultados

- Para as tabelas mostradas anteriormente, podemos relacionar os valores obtidos com as métricas do erros absoluto (MAE) e a média dos erros ao quadrados (MSE).
- A apresentação a seguir do MSE foi com base em sua raiz (para mostrar os dados obtidos reais)

Base utilizada	MAE	MSE
Geral	0.3507371548651297	0.3977645810891414
Views	0.3099424225394416	0.3607402872847538
Views e Likes	0.29458746570548594	0.34720609048117634
Vies, likes e dislikes	0.31518495805187347	0.3707967124521461



# Conclusão

- Observando os valores de erros, concluímos que os principais fatores para determinar um canal de sucesso, tendo como base os inscritos no canal, são os valores de visualizações(Views) e curtidas(Likes).
- Pois a partir da realização de uma regressão com outros valores, a regressão começa a ter um aumento na quantidade de erros existentes.

# Ameaças a validade

- Após analisar melhor e visualizar os valores capturados, somente de canais de possíveis sucessos, foi concluído que obter uma quantidade maior de resultados, uma vez que a regressão utiliza os parâmetros existentes de curtidas, em média, para concluir uma quantidade de inscritos (sua real audiência).
- Uma captura de dados durante um período maior.

Obrigado!