

**Universidade Federal de Ouro Preto**  
**Instituto de Ciências Exatas e Aplicadas - ICEA**  
**Campus João Monlevade - MG**

LINCOLN REBOUÇAS - 20.1.8068

## **Relatório de Coleta de Dados - Nuvem de Palavras**

João Monlevade  
2024

## Introdução

Este relatório apresenta uma análise de vídeos do YouTube relacionados ao tema “Programação e Dados”, com foco nas linguagens de programação **Java** e **Python**, além de **Banco de Dados**. O objetivo da pesquisa foi identificar os vídeos mais populares e relevantes entre 2020 e 2022. Foram utilizadas métricas como *visualizações*, *curtidas (likes)* e *comentários (replies)* para avaliar o desempenho dos vídeos.

## Palavras-chave e Consultas

### Palavras-chave utilizadas:

- Programação
- Java
- Python
- Dados

### Consultas realizadas:

- Melhores cursos de programação
- Java
- Python
- Banco de dados

## Intervalo de Tempo

O intervalo de tempo considerado para a coleta dos dados foi de **1 de janeiro de 2020 a 31 de dezembro de 2022**. Durante esse período, a pandemia de COVID-19 incentivou o crescimento da educação online, aumentando a busca por cursos de programação e dados.

## Coleta de Dados

A coleta e análise dos dados foram realizadas utilizando Jupyter Notebook, que permitiu uma exploração interativa e visual dos dados obtidos. Durante a coleta, que durou **15 minutos**, foram coletados **166 vídeos** e **46.665 comentários**. Esse processo foi fundamental para garantir a qualidade e a relevância dos dados analisados. Os dados coletados incluem o título do vídeo, data de publicação, categoria, visualizações, curtidas e comentários. A seguir, apresentamos as estatísticas descritivas das principais métricas. Os detalhes da implementação, bem como o código utilizado para a coleta dos dados, estão

disponíveis no repositório do GitHub. Você pode acessar o repositório através do seguinte link: Repositório de Coleta e Análise de Dados

## Métricas

### Visualizações

Os vídeos analisados tiveram entre 27 e 8.595.003 visualizações, com uma média de aproximadamente **495.949 visualizações**.

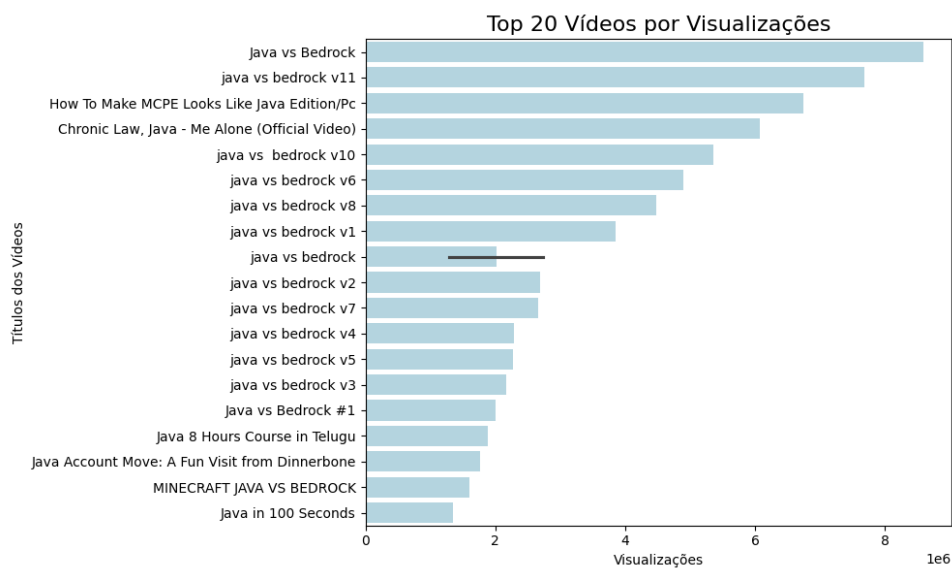


Figura 1 – Top 20 vídeos por visualizações.

### Curtidas

A média de curtidas por vídeo foi de **11.595**, variando entre 3 e 403.250. O gráfico abaixo mostra os 20 vídeos com mais curtidas.

### Comentários

A média de comentários por vídeo foi de **281**, com um máximo de 4.684 comentários.

### Nuvem de Palavras

A partir dos 46.665 comentários coletados, foi gerada uma nuvem de palavras, destacando os termos mais frequentes nos vídeos sobre programação e dados. As principais palavras encontradas foram: **bedrock**, **java**, **python**, **minecraft**, **better**, **thank**, e **player**.

Esses termos indicam uma forte relação entre a programação em **Java** e **Python** com o universo do **Minecraft**, um dos jogos mais populares e amplamente modificado por

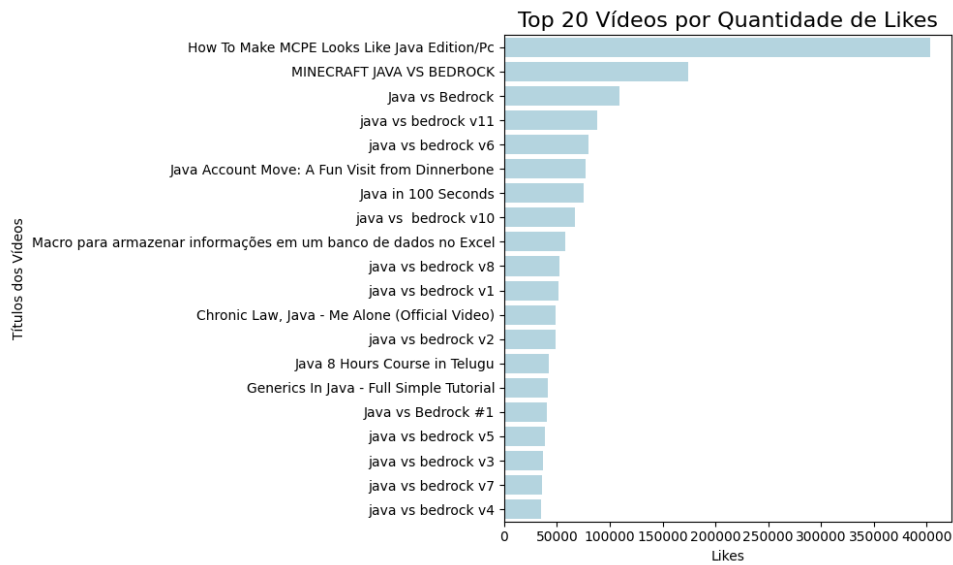


Figura 2 – Top 20 vídeos por likes.



Figura 3 – Top 20 vídeos por comentários.

usuários, explicando a presença de termos como **bedrock** (referente a uma versão do Minecraft) e **player** (jogador).

A palavra **better** está provavelmente associada às discussões sobre melhorias no código ou comparações entre as versões **Java** e **Bedrock** do Minecraft, enquanto **thank** reflete interações positivas entre criadores de conteúdo e a comunidade.

## Conclusão

O período analisado, entre 2020 e 2022, foi marcado por um aumento significativo no interesse por conteúdos educacionais em programação e dados, particularmente em



Figura 4 – Nuvem de palavras gerada a partir dos comentários dos vídeos.

**Java, Python e Banco de Dados.** Os vídeos mais visualizados e curtidos ofereciam cursos completos e tutoriais práticos, sendo altamente valorizados pelos usuários. A análise das métricas de *visualizações*, *likes* e *comentários* revelou um alto engajamento da comunidade nesses temas.

Este relatório fornece *insights* valiosos sobre os vídeos mais relevantes para quem deseja aprender programação e trabalhar com dados, com destaque para as linguagens **Python** e **Java**, bem como o uso de **bancos de dados**.

## Referências

COMMUNITY, S. O. Python: wordcloud, repetitive words. Available online: <https://stackoverflow.com/questions/43954114/python-wordcloud-repetitive-words>: :text=wordcloud

DEVELOPERS, G. Youtube data api v3. Available online: <https://developers.google.com/youtube/v3>. 2022.

(DEVELOPERS, 2022) (COMMUNITY, 2017)