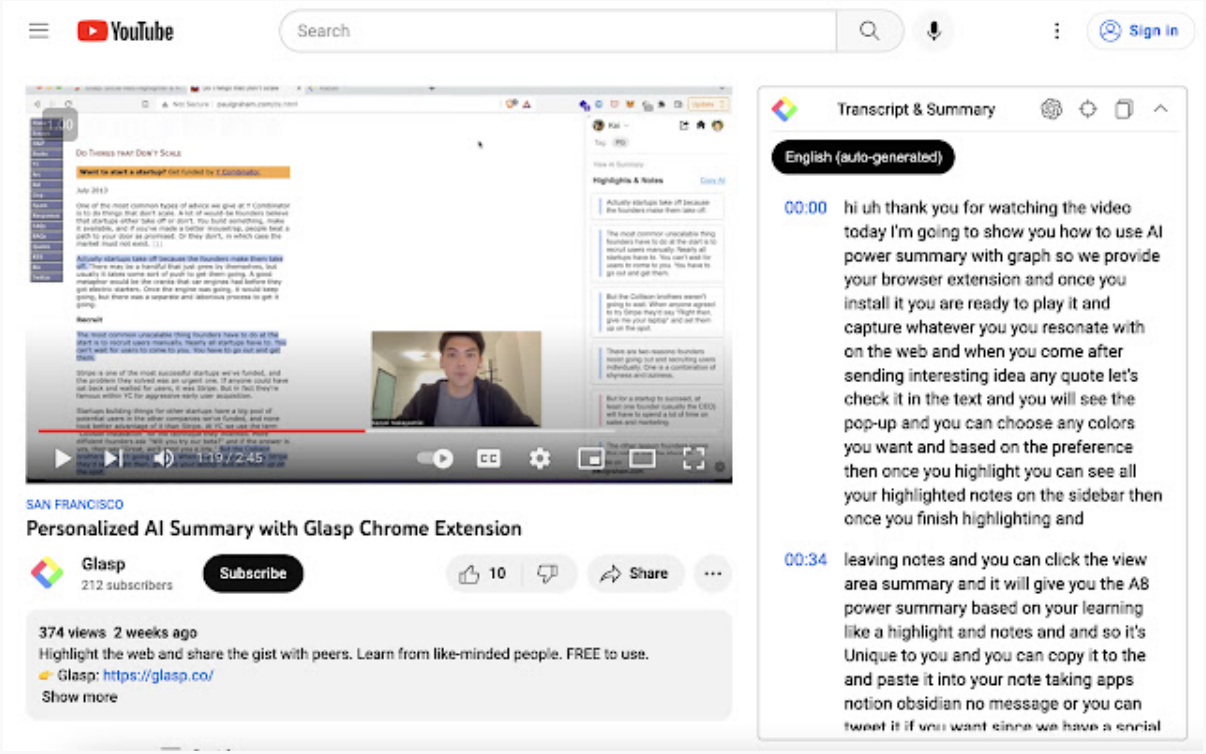
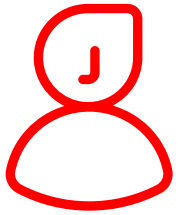


Youtube Video Summarizer Presentation

An introduction to a tool that can automatically summarize Youtube videos using transcript analysis.

Youtube's Global Command

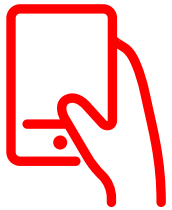


YouTube has over 2 billion monthly users

YouTube is the second most visited website on the internet

Users watch 1 billion hours of YouTube videos daily

This accounts for around 5 billion videos being watched per day



YouTube app has over 10 billion installs on Android

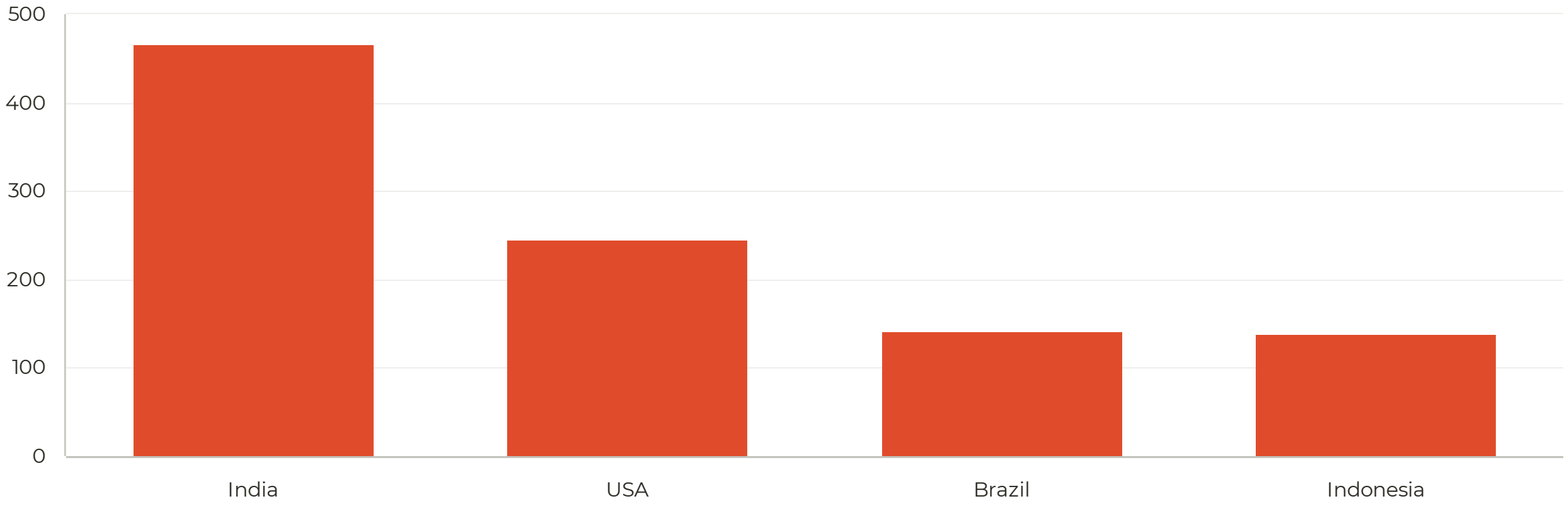
This is more installs than the entire world population

T-Series has over 205 million YouTube subscribers

The most subscribers of any channel on the platform

YouTube dominates online video streaming with billions of users and hours watched daily

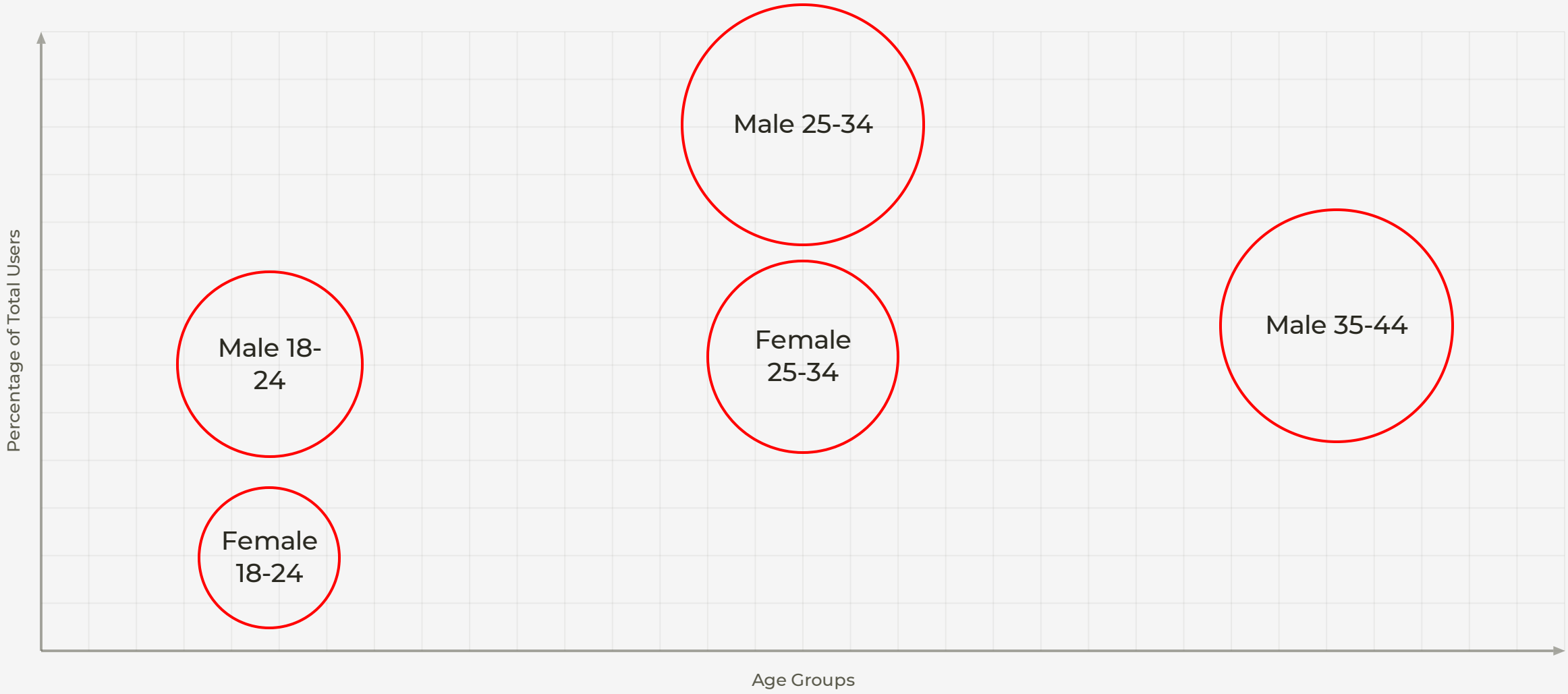
Active YouTube Users by Country



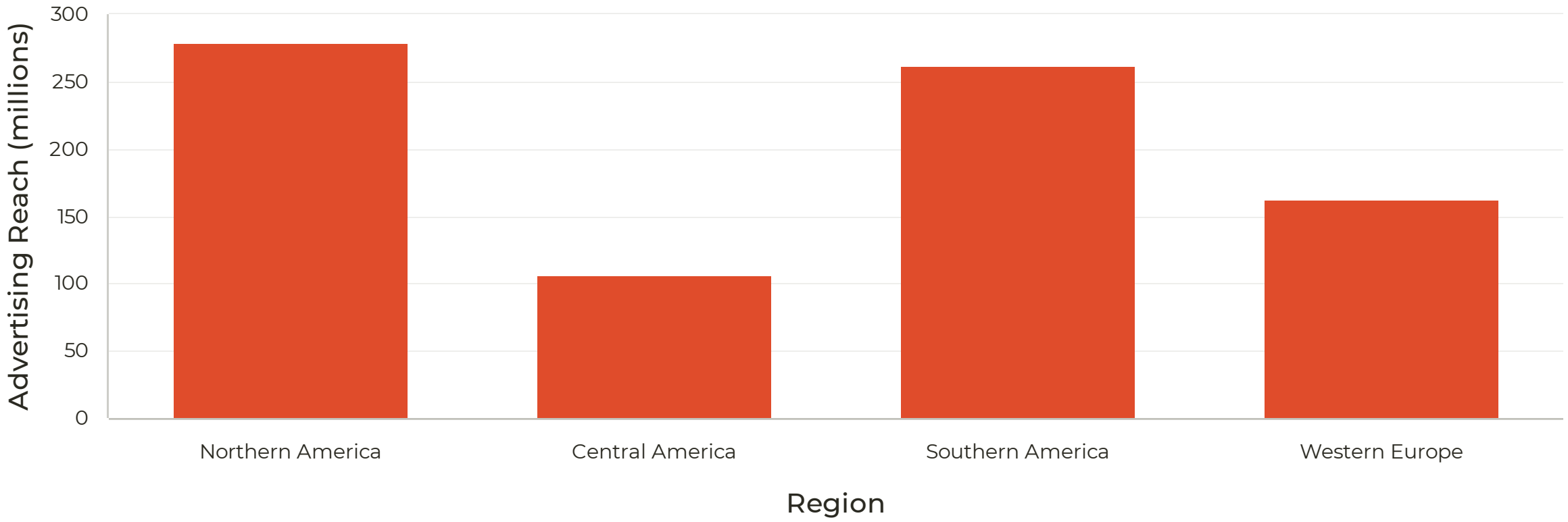
India has the most active YouTube users.

*Kepios

YouTube User Demographics



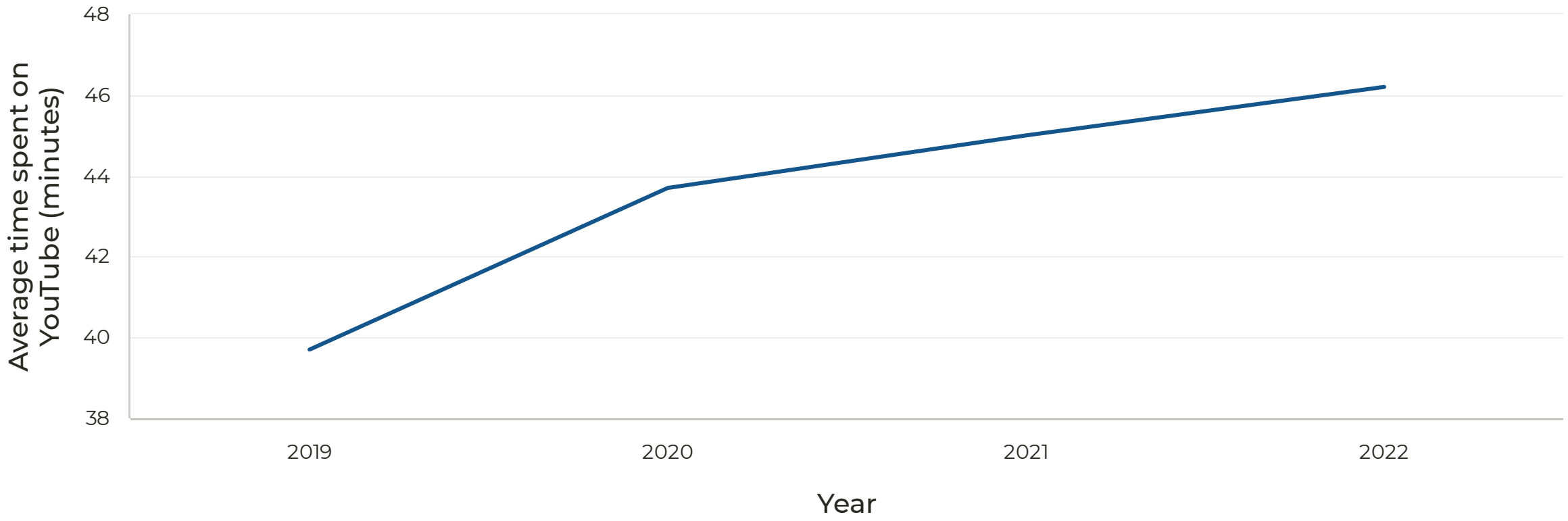
YouTube Advertising Reach by Region



YouTube has a significant advertising reach across all regions.

*Hootsuite

Average Time Spent on YouTube

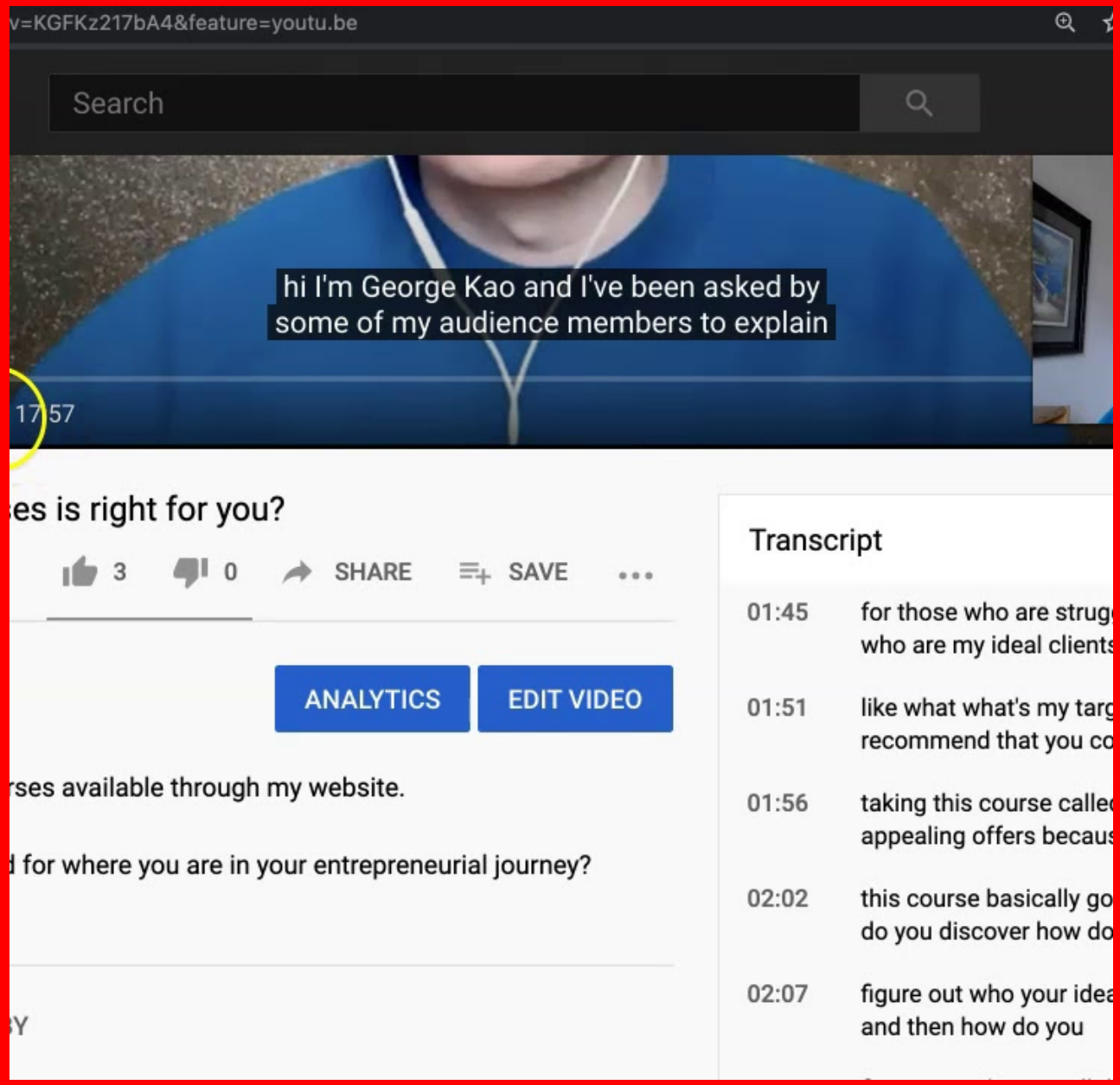


The average time spent on YouTube has steadily increased from 2019 to 2022.

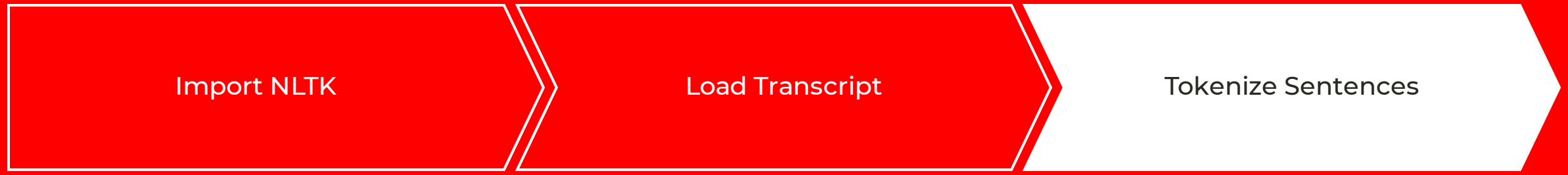
*Insider Intelligence

Fetching Video Transcripts

The YouTube API allows us to seamlessly retrieve video transcripts based on a provided URL or video ID. This automation eliminates the need to manually find and copy transcripts.



Sentence Tokenization



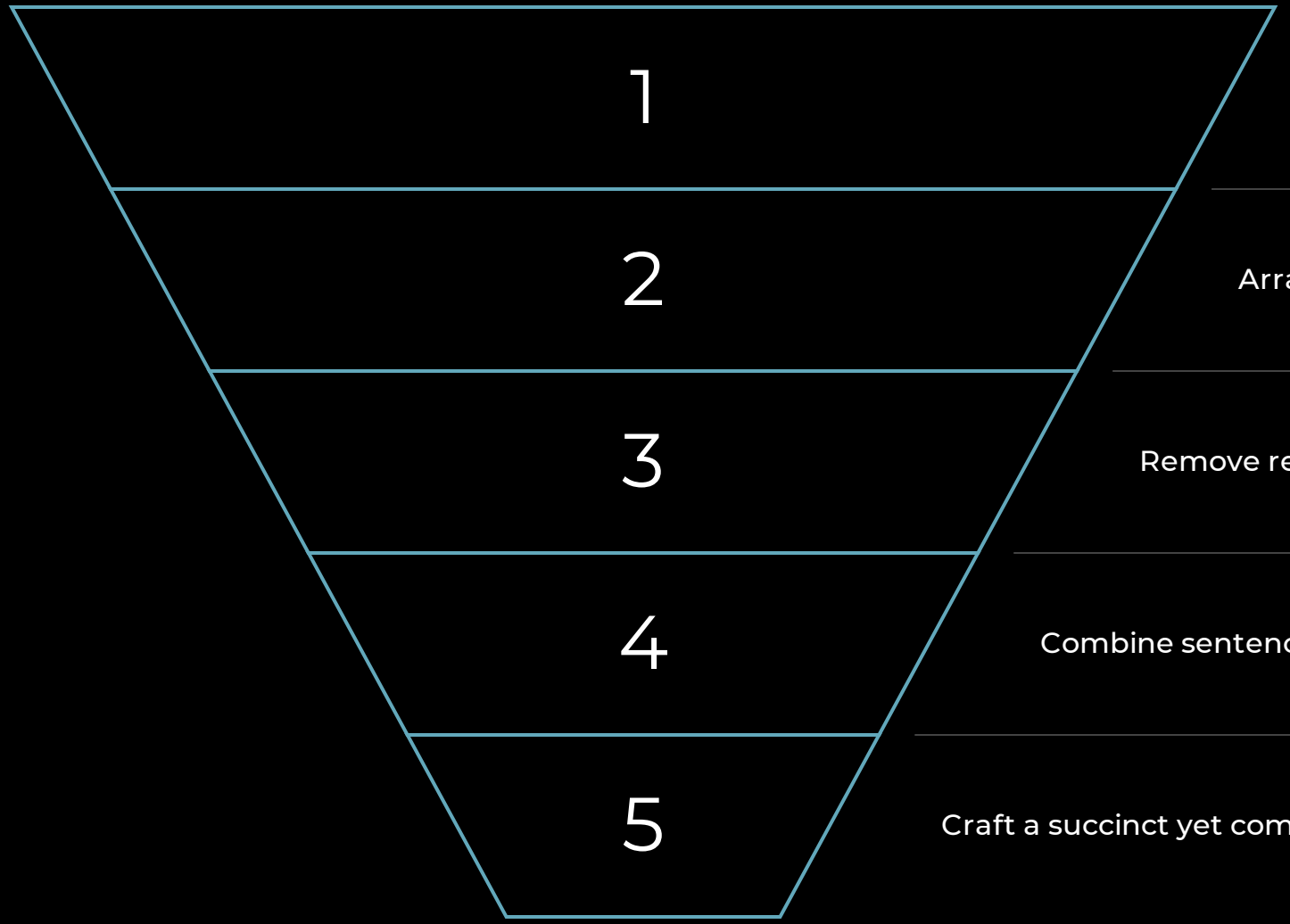
Import the Natural Language Toolkit library for Python.

Load the video transcript text into a string variable.

Use NLTK's sent_tokenize() function to split the transcript into a list of sentences.

Key Sentence Extraction

Concise Summary Creation



Extract key sentences that convey the main ideas

Arrange extracted sentences logically

Remove redundant and extraneous sentences

Combine sentences smoothly into coherent paragraphs

Craft a succinct yet comprehensive summary

Effortless Video Transcript Extraction

Leverages YouTube API

Uses YouTube API to retrieve transcripts for any video

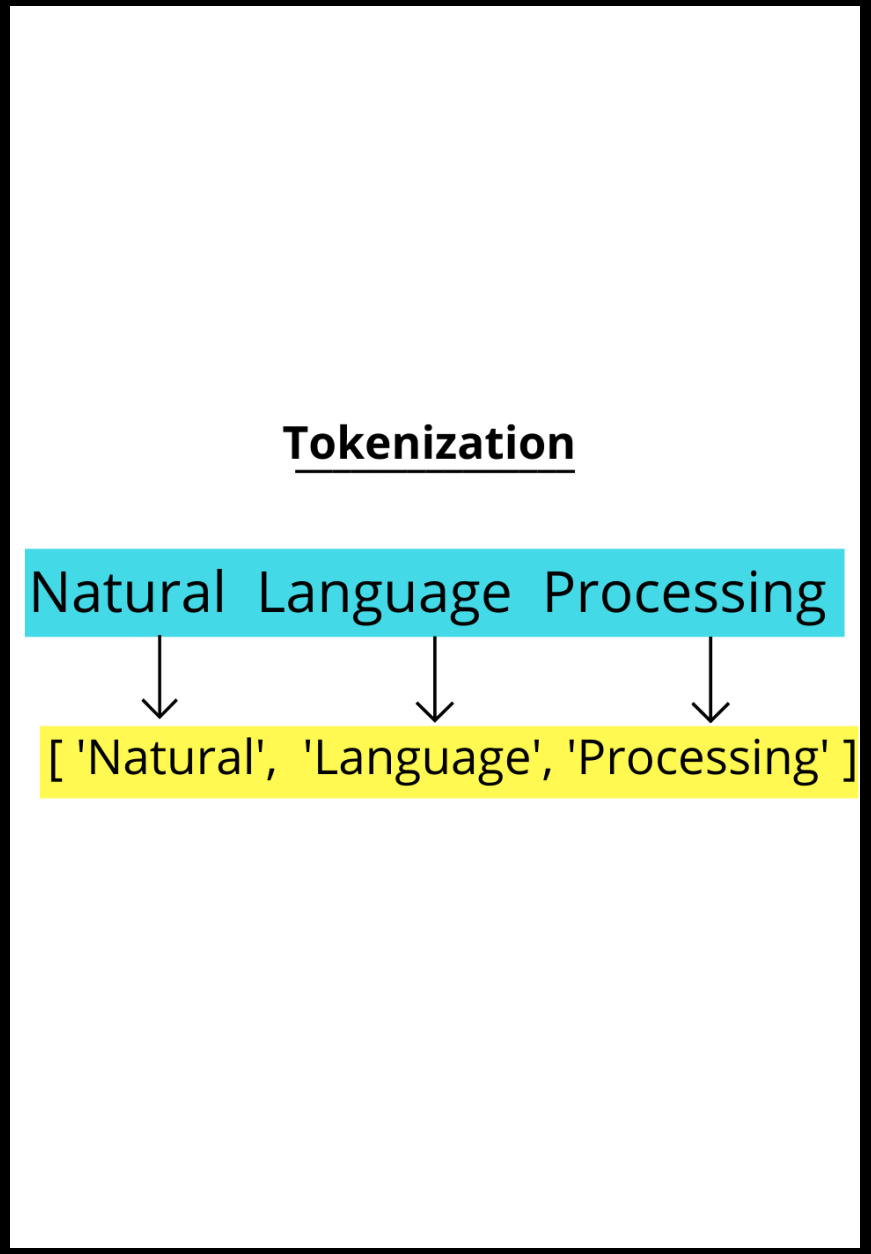
Seamless integration

Transcripts fetched directly from YouTube with minimal effort

Effortlessly extracts transcripts by integrating with YouTube API

NLTK Sentence Tokenization

The Natural Language Toolkit (NLTK) is a powerful Python library for natural language processing. We use NLTK's sentence tokenizer to break down the YouTube video transcripts into individual sentences. This prepares the transcripts for further analysis and summarization.



Summarization Process

- **Sentence Tokenization**

  Break down transcripts into individual sentences using NLTK's powerful tokenization capabilities

- **Key Sentence Extraction**

  Identify the most significant sentences in the transcript using NLTK's extractive summarization techniques

- **Summary Generation**

  Carefully combine extracted key sentences to create a concise yet informative summary

In a Nutshell

Word Frequency Counting

Iterates through each word, ignoring common words, and counts word frequency

Normalization

Divides frequencies by max value to balance word weights

Sentence Scoring

Iterates through sentences, adding word frequencies to get sentence scores

Summary Creation

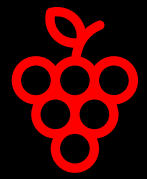
Extracts top scoring sentences to form 30% length summary

Key steps are counting word frequency, normalizing weights, scoring sentences, and extracting top sentences for the summary.

Purpose

Unlocking knowledge efficiently

YouTube summarizer allows viewers to quickly understand key information in videos



Promoting inclusivity and engagement

The summarizer accommodates different learning styles and attention spans, increasing video engagement

Shaping content consumption

Summaries break language barriers and make videos more accessible globally

Summaries unlock key points quickly, catering to viewers with limited time or those seeking specific information within lengthy videos.