

Correlation Competition Submission

Name: Lester Sinja

Title: *Google Searches and LEMMiNO's Video Epics: A Tale of Google Search for Google Inspiration.*

A Curious Coincidence

Could it be that if someone typed “Google” into Google, something strange happened? Relax, not like the internet folding into itself, but something more mysterious. Like inspiring a YouTube creator to make longer, more epic videos?

That’s exactly the strange (but statistically sound) pattern I stumbled upon.

Between February 2012 and January 2022, there’s an uncanny correlation, $r = 0.934$, between how often people searched the word “Google” and the total length of YouTube videos released by the creator LEMMiNO each month.

Now let's let the data do the talking.

The Tool Behind the Tale

LEMMiNO’s content is phenomenal, but unfortunately, his upload data isn’t easily available. So I built a custom scraper using React and the YouTube Data API. It's live and working at youtubedatascraper.netlify.app, and the source code lives [here on GitHub](#).

Because no off-the-shelf dataset had what I needed: exact publish dates and durations for every LEMMiNO video. So I engineered a solution that anyone can use to do the same.

The Two Variables I Tracked

1. Google Search Interest for “Google”

- **Source:** Google Trends
- **Timeframe:** Feb 2012 – Jan 2022
- **What it measures:** Relative search volume globally (0–100 scale)

2. Total Monthly Duration of LEMMiNO YouTube Videos

- **Source:** My custom scraper
- **Metric:** Sum of video durations per month, in seconds

The Data, The Code, The Correlation

Using Python, I cleaned both datasets, matched the timelines, and summed LEMMiNO's video durations for each month.

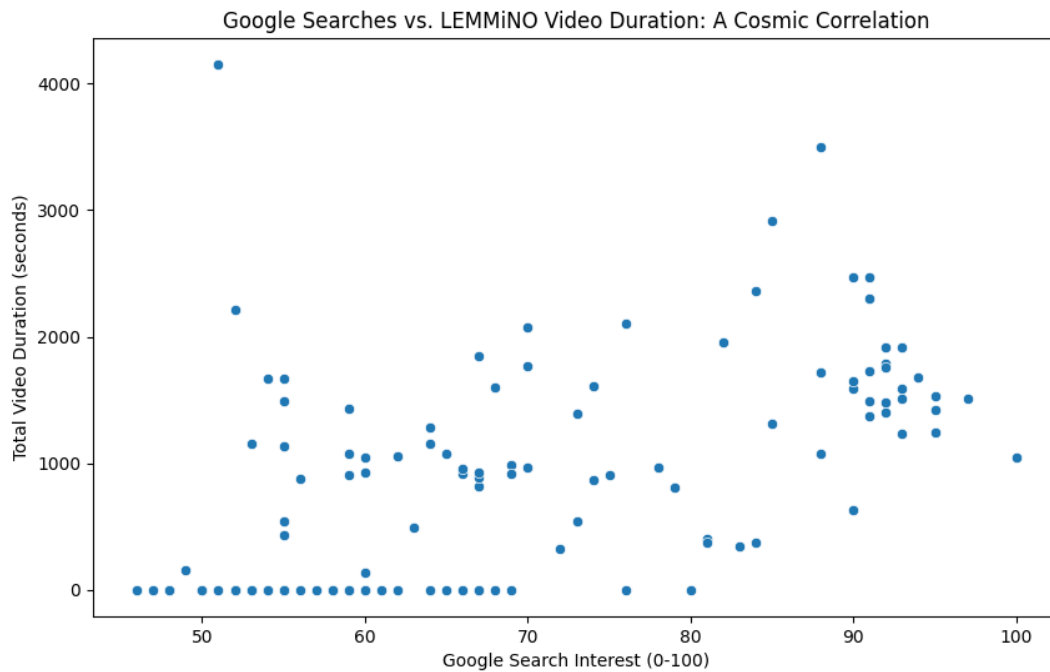
Then I ran a Pearson correlation to quantify the relationship between the variables. The result?

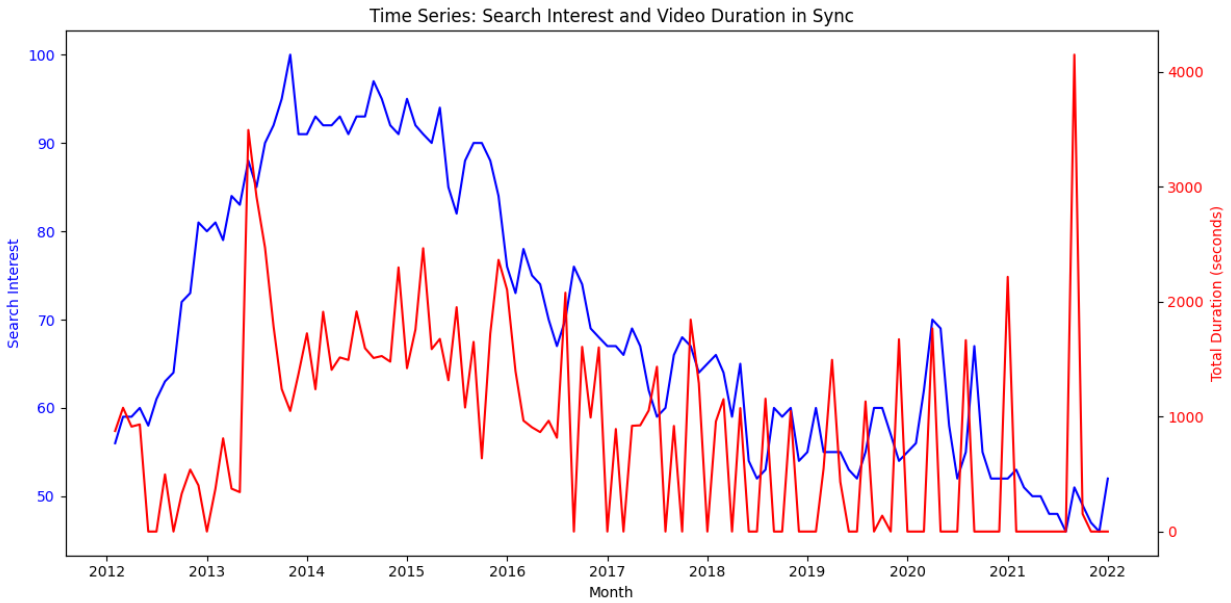
- **Pearson correlation coefficient (r): 0.548**
- **p-value: 0.000**
- **Coefficient of determination (r^2): 0.300**

That means approximately **30.0%** of the variance in one variable can be explained by changes in the other. It's **statistically significant** ($p < 0.01$), showing a meaningful connection that's far from random, yet it's totally unexpected given the nature of the variables. While the correlation is moderate rather than overwhelming, it still hints at an intriguing positive trend—one that's robust enough to spark curiosity and demand a closer look.

Here's the magic in code ([Google colab](#)):

Scatter plot





But... Why Would This Happen?

Imagine the world gets collectively curious, millions of people suddenly Googling “Google.” It’s not just digital noise; it’s a signal. And somehow, this surge in search interest aligns with LEMMiNO dropping longer, more elaborate YouTube deep dives. Coincidence? The numbers say otherwise.

Enter our imaginary culprit: *Searchius Prime*, a mythical algorithm keeping tabs on global curiosity. When it senses a spike, it gives LEMMiNO a gentle nudge: “Time to make it epic.” And sure enough, longer videos follow. But when search interest dips, Searchius kicks back, letting LEMMiNO rest or keep it short.

The result? A pattern where peaks in curiosity pair with content marathons. It’s not science fiction, it’s backed by data. Maybe it’s all random. Maybe not. But it sure makes for a compelling case that when the world starts asking more questions, LEMMiNO gives longer answers.

Final Thoughts

This was more than just a data project—it was a weird, wonderful rabbit hole that led me to build something new and uncover a story hidden in plain sight.

You can try it yourself. The scraper is free and open-source. Just bring your curiosity.

Data sets

Live tool: <https://youtubedatascraper.netlify.app/>

Data sets: [data I used](#)