DSCI 510 – Fall 2025 – University of Southern California

Instructor: Dr. Alexey Tregubov

Final Project Progress Report

Title: Which Topics Are Trending Across Online Learning Platforms?

Project Scope Update

I analyzed how Udemy and Coursera topics change over time and compared them with global interest using Google Trends through Pytrends. I built a full pipeline that cleans the Kaggle data, normalizes topics by year, and pulls yearly "machine learning'' interest. The code is modular, `fp_topics.py` cleans data, `fp_trends.py` handles the API, and main.py creates the plots and saves everything to the `results/` folder.

Data Source Table

- Udemy Online Courses (Kaggle): 3,678 courses with titles, subjects, and publish dates
- Coursera Courses & Skills Dataset 2025 (Kaggle): 3,404 courses with subject and skill data
- Google Trends (Pytrends API): public interest data for "machine learning"

Used API: I used the Pytrends library, I retrieved weekly data, Google search interest for "machine learning", aggregated it by year, and saved both a CSV (`trends_machine_learning_preview.json`), and JSON preview for verification.

Findings / Results So Far

I found that Udemy is trending toward Web Development, Graphic Design, and Business Finance, while Coursera focuses more on Data Science and Business. Global interest in "machine learning'' has risen steadily since 2010. My regression showed an R^2 of about 0.036, meaning only a small link between global interest and Udemy's topic growth.

Issues / Difficulties

Coursera's data only provides a 2025 snapshot, which limits how far I can compare trends over time. To handle Google Trends rate limits, I added retry delays in my code, and I manually aligned the two platforms' category systems so the topics match consistently across datasets.