Amazon Reviews

COM5508 Media Data Analytics

Lab 2

In this exercise, we use plain python (i.e., no numpy/pandas) to analyse Amazon reviews. You are free to select any of the amazon categories on `http://jmcauley.ucsd.edu/data/amazon/links.html`. If you want to use product information, i.e. the product names or "bought together" info, you need to download both the review files (please use the smaller 5-core subsets from the first table) and the metadata (from the "Per category files" in the second table) This data can be used to answer lots of different questions, some examples:

- What kind of products get reviewed the most?

- What is the overall distribution of scores?

- Are reviewers that review a lot of products very negative or very positive?

- Do products that contain a certain word in the title get many or high reviews?

- What kind of products are bought together with a specific product?

In this lab, you will forumate your own research question and answer it using the provided data.

Analysing Amazon Reviews

- Formulate a fun or interesting data research question. This can be as simple or complicated as you want (but of course you won't get an A+ grade for a trivial question)

- Load the reviews and/or metadata for the relevant category and read it into a suitable python data structure

- Find the needed data to answer the question using plain python (no numpy/pandas).

- Answer the question by providing and interpreting some key numbers or tables.

- Write the report in ipython-notebook, aim for about 2 pages.

- Submit in canvas as html or pdf