**NLP Proposals**

(Arbnor, Lukas and Marion)

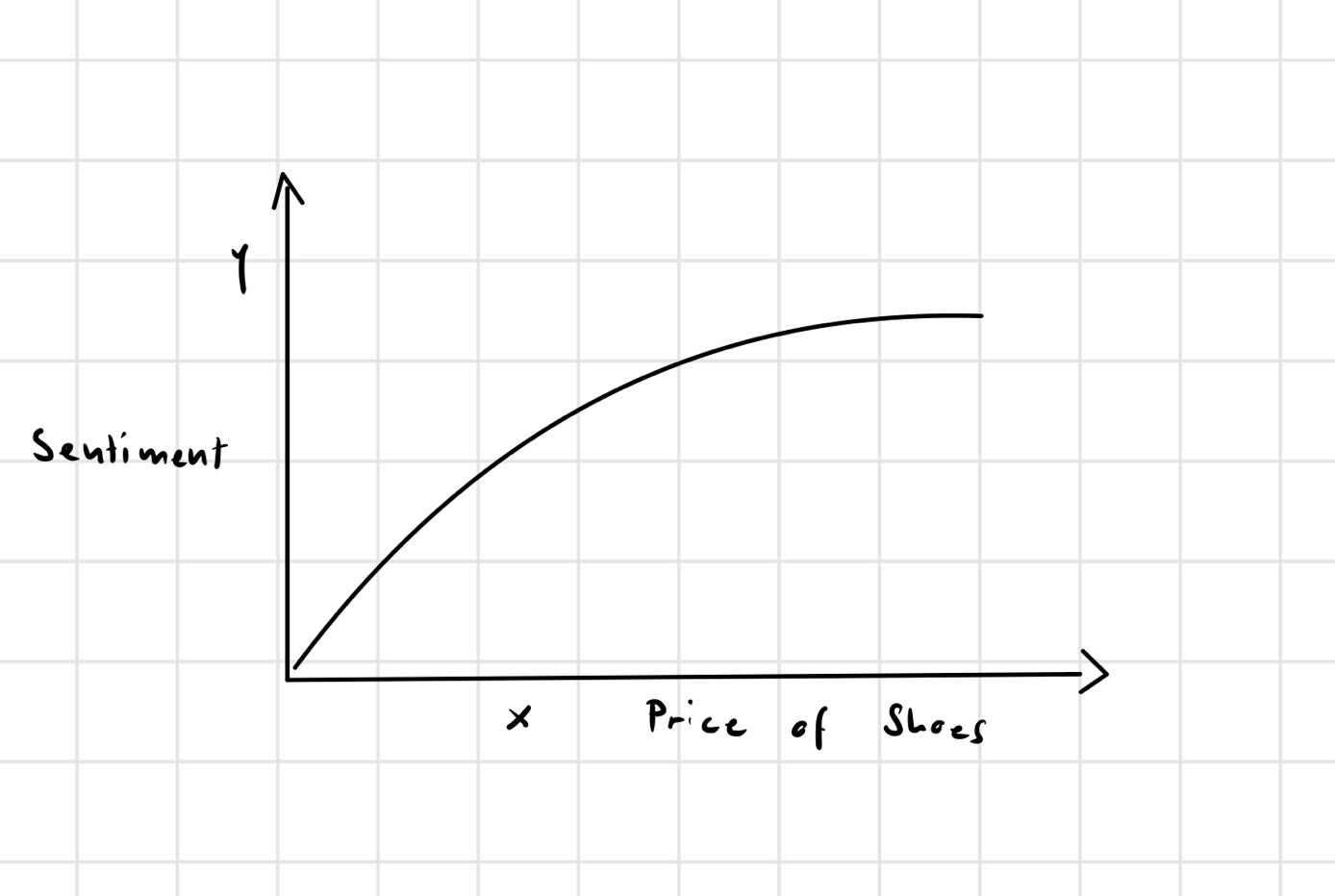
**Proposal 1  
Shoe reviews on Amazon**

RQ: Do customers who buy more expensive shoes leave more positive reviews than customers who buy less expensive shoes.

Motivation: we love shoes!

Based on utility maximization, we assume that more expensive shoes result in a higher utility and therefore in more positive sentiments. We analyze the dataset of amazon <https://www.kaggle.com/datasets/daishinkan002/men-women-shoes-reviews>

and perform a sentiment analysis on the reviews. Then we plot it against shoe prices (see example below).



**Proposal 2**

**CNN-Articles Review**

RQ: Did newspapers express more negatively during the pandemic than before and after and did the sentiments reflect the different stages of the pandemic?

Motivation: we would like to know if the media have properly reflected the sentiments of the population.

CNN Dataset of Articles between 2011 – 2022. We would concentrate on news articles between 2019 and 2022 regarding the pandemic.

We assume that the sentiments in the articles are more positive before and after the pandemic, and that the sentiments will also vary - but less - during the pandemic, because of the nature of the virus and the restrictions of the government.

