

Webscraping Tripadvisor Hotel and restaurant toxicity

Lou Shangzhi

Cherel Bastien

Outlines of the presentation

- 1. Problematic
- 2. Methods
- 3. Problems & solutions
- 4. Results & conclusion



Problematic

Are comments about hotels more toxic than those about restaurants?

Methods

APIFY API

- Fetch hotels urls, rating
- Fetch restaurants urls, rating
- Source : https://apify.com



Extract the comments

Decide whether the comment is toxic or not

• Source: https://pypi.org/project/scrapegraphai/1.33.3/

Data analysis

- Proportional analysis between hotels and restaurants
- Draw conclusion about the initial problematic

Problems & solutions

Scraping trip advisor was impossible with Beautifulsoup



We chose to use APIS e.g. APIFY/GPT



Cette photo par Auteur inconnu est soumise à la licence CC BY-SA

Results & conclusion

Are comments about hotels more toxic than those about restaurants?

- More toxic comments for restaurants
- Raise awareness: the results might be biased because of unbalanced reviews

Toxicity Analysis: Category: Restaurants Total Comments: 37 Toxic Comments: 15 (40.54%) Non-Toxic Comments: 22 (59.46%) Category: Hotels Total Comments: 8 Toxic Comments: 2 (25.00%) Non-Toxic Comments: 6 (75.00%)

Results

