REVIEW DATA CLEANSING

★★★★★

Implementing strategies to manage comment flood on establishment reviews in the Bar de Respeito map, ensuring platform trustworthiness post-traffic surge
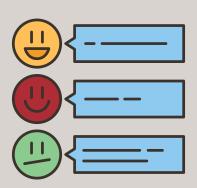
OBJECTIVE

Develop strategies for cleaning establishment reviews and structure queries to perform database deletions.
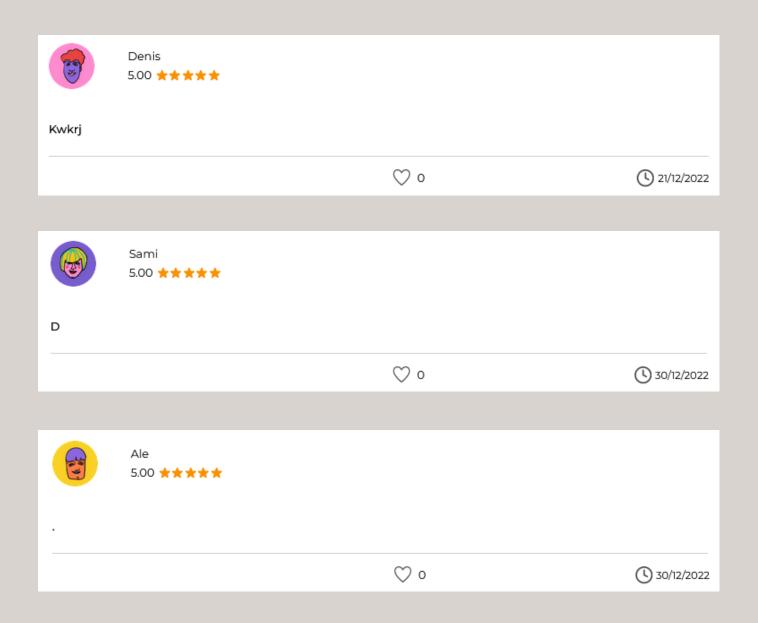
TOOLS

• PostgreSQL •

PROBLEM

- "What types of comments are pertinent in reviews?"
- "What is the pattern for non-pertinent comments?"
- "What characterizes a comment originating from a flood?"
- "Which period will be considered for deletion?"

EXAMPLES

Denis
5.00 ★★★★★

Kwkrj

♡ 0   🕐 21/12/2022

Sami
5.00 ★★★★★

D

♡ 0   🕐 30/12/2022

Ale
5.00 ★★★★★

.

♡ 0   🕐 30/12/2022

PROCEDURES

✓ Exploratory analysis of comments using SELECT
✓ Structuring of <u>queries</u>
✓ Measurement of results and future impact on map reviews
✓ Application in the test environment
✓ Comparison with the real environment
✓ Verification of improvements and optimizations of queries
✓ Application in the real environment

DEFINED STRATEGIES

- Users with comments starting with three consonants
- Users who posted comments with three or fewer characters
- Users who selected "not informed/other" in two registration steps
- Users who posted comments containing only consonants
- Users whose "name" data starts with three consecutive consonants

FINAL CONCLUSIONS

- More than 30,000 reviews were removed after filtering
- Discuss ways to prevent uncontrolled account creation by adding new verification steps in registration
- Create limitations on review submissions, such as a minimum character count