



ANALYZING SOCIAL MEDIA DATA ON THE TOILETS OF SINGAPORE

A project by Eugene Choy, Sim Li Jin, Li Xiaohang, Ho Winston & Ragini Verma

OUR CLIENT

The Restroom Association of Singapore (RAS) was set up in 1998 as a non-profit organization that advocates for clean public toilets. RAS works closely with the National Environmental Agency (NEA) and Ministry of Health (MOH) to create initiatives that promote clean toilets. In 2003, RAS launched the Happy Toilet Programme (HTP) as a star-grading initiative for public toilets. Toilet owners can apply for their toilets to be assessed by RAS and are then awarded a certificate with a rating of the toilet.

PROBLEM

- Part of the mission of Restroom Association Singapore is to 'Lobby the cleaning industry to meet the right standards'. With around 1150 toilets to look out for, it can be challenging to know which toilet to focus lobbying attention on
- Toilet owners registered under HTP schedule an audit by Restroom Association. The results of these audits are valid for a year afterwards. However, there is a likelihood that the toilets won't remain clean after the audit is completed.

Our solution analyzes social media mentions of toilets in Singapore, and integrates these sentiments and comments into the current Happy Toilet Programme. This will allow for a more dynamic review of the toilets, as well as for concrete comments to be reflected so that RAS can take action according to the feedback



TECHNOLOGIES

Python Programming Tools:
Jupyter
Pandas
NLTK
Textblob
Selenium
Scrapy
Gensim

APIs:
Google Maps API
One Map API
Reddit API
Twitter API

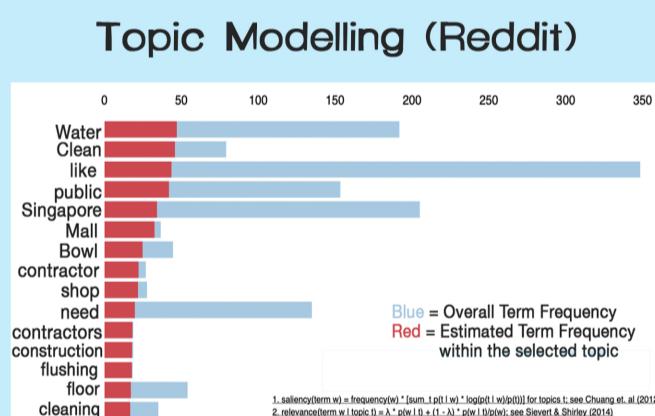
TYPES OF ANALYSIS



After scraping data from Reddit, TripAdvisor, Instagram & Twitter, we conducted 2 main types of analysis:

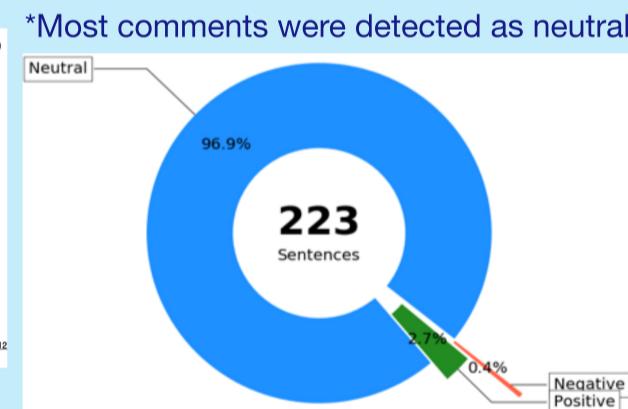
- Natural Language Processing (NLP) to identify locations, conduct Sentiment Analysis and identify Trending topics and words
- Geographical Visual Analysis to display data in a map form and therefore visualize the locations of dirty toilets

RESULTS & INSIGHTS



*Topics associated with opinions on 'toilets', found by conducting LDA topic modelling

Sentiment Analysis (TripAdvisor)



We also conducted:
Sentiment Analysis (Reddit)
Trending Words Analysis (Twitter)

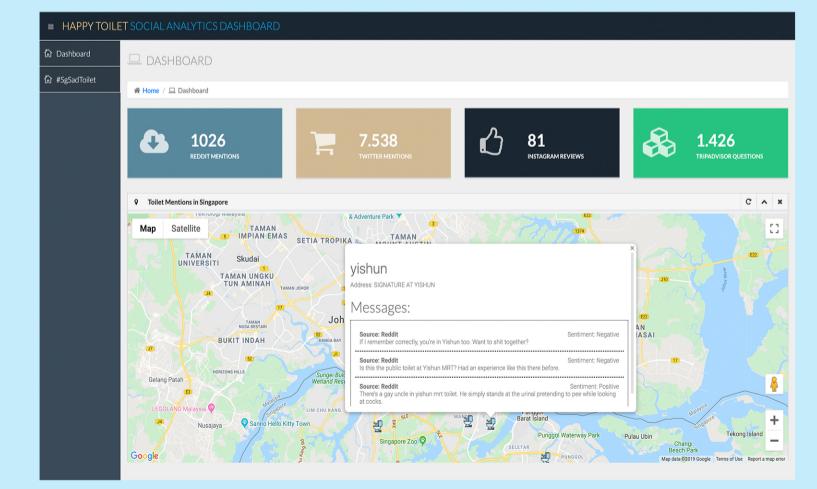
Trending Word Cloud (Instagram)



*Interesting words identified include 'ply' and 'butt spray'. This could imply that users may keep a lookout for the number of ply toilet paper or the availability of bidets in toilets

Geographical Visual Analysis

We developed a simple web application to integrate the various components of our analysis. We also mapped the locations of the posts to visualize



VALUABLE INSIGHTS

- Insight Generation:** Integrated Public Toilet Rating System: For RAS to dynamically identify locations of public toilets with good or bad ratings
- Possible Enhancements of Happy Toilet Audit Metrics:** Key phrases identified can explain what users look out for in a good toilet, these features can be added to the current rubric to differentiate the happier toilets better
- Lastly, we suggest a **social media campaign** to encourage users of public toilets to post their sentiments using specific hashtags (**#SGHappyToilet** & **#SGSadToilet**) to make it easier for our application to collect relevant data regarding sentiments on toilets