AIRBNB SENTIMENT ANALYSIS

Using NLTK + VADER

Andrew Wong

ABOUT ME

Based in Melbourne, Australia. I have worked / lived in 14 countries in Europe and Asia.

A high-performing commercial technologist and solution delivery leader with a track record in working with data science, machine learning, mobile, digital and software engineering program budget up to AUD$50 million.

I have specialist data science domain in the followings

- *Agile Development: Rapid iteration of data science projects through small, incremental value delivery.*

- *Product Development: Developing minimum viable data science / machine learning product (MVP) through Design Thinking and Design Sprinting.*
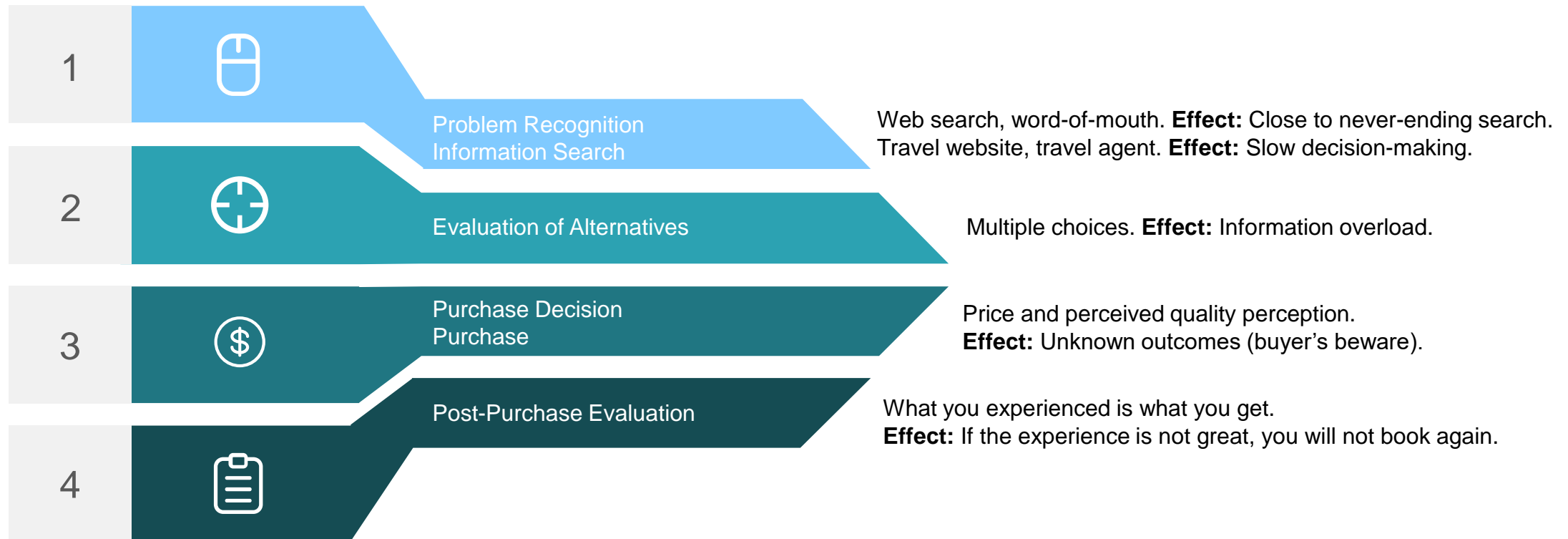
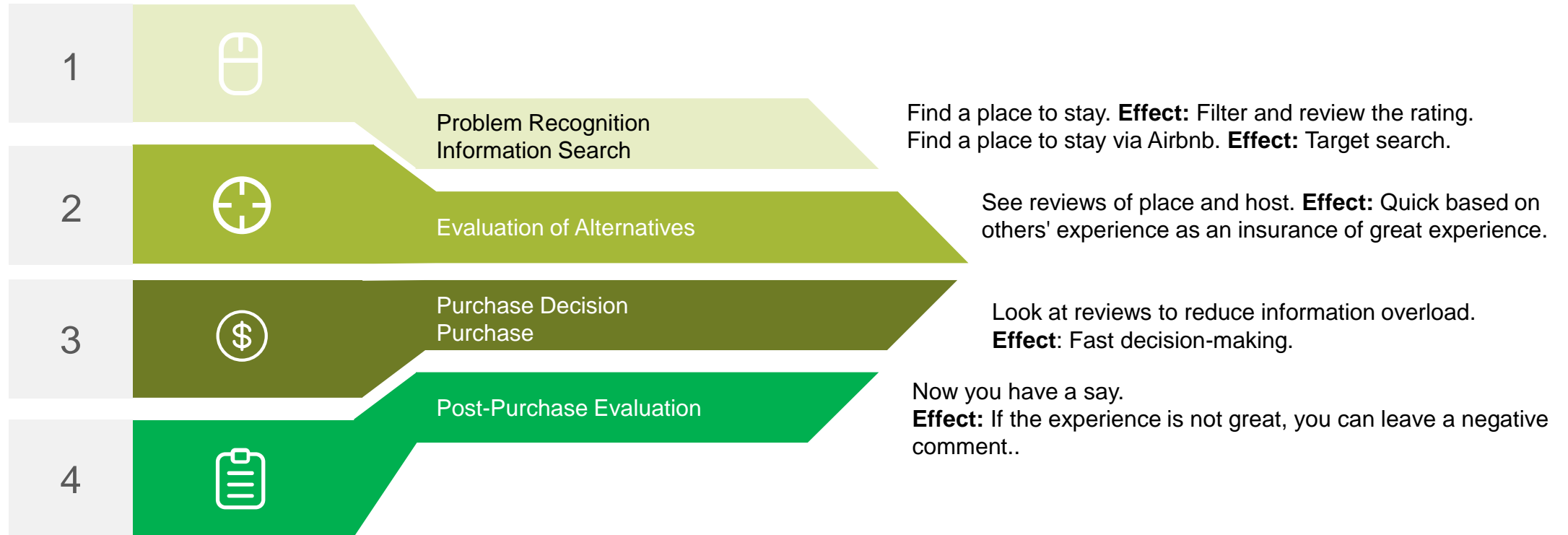- *Industry Verticals: Start-ups, MedTech, EdTech, FinTech*

BUSINESS PROBLEMS

Customers Reviews in This Attention Deficit World

THE SCIENCE OF REVIEWS IN THIS ATTENTION DEFICIT WORLD: PRE-REVIEW WORLD
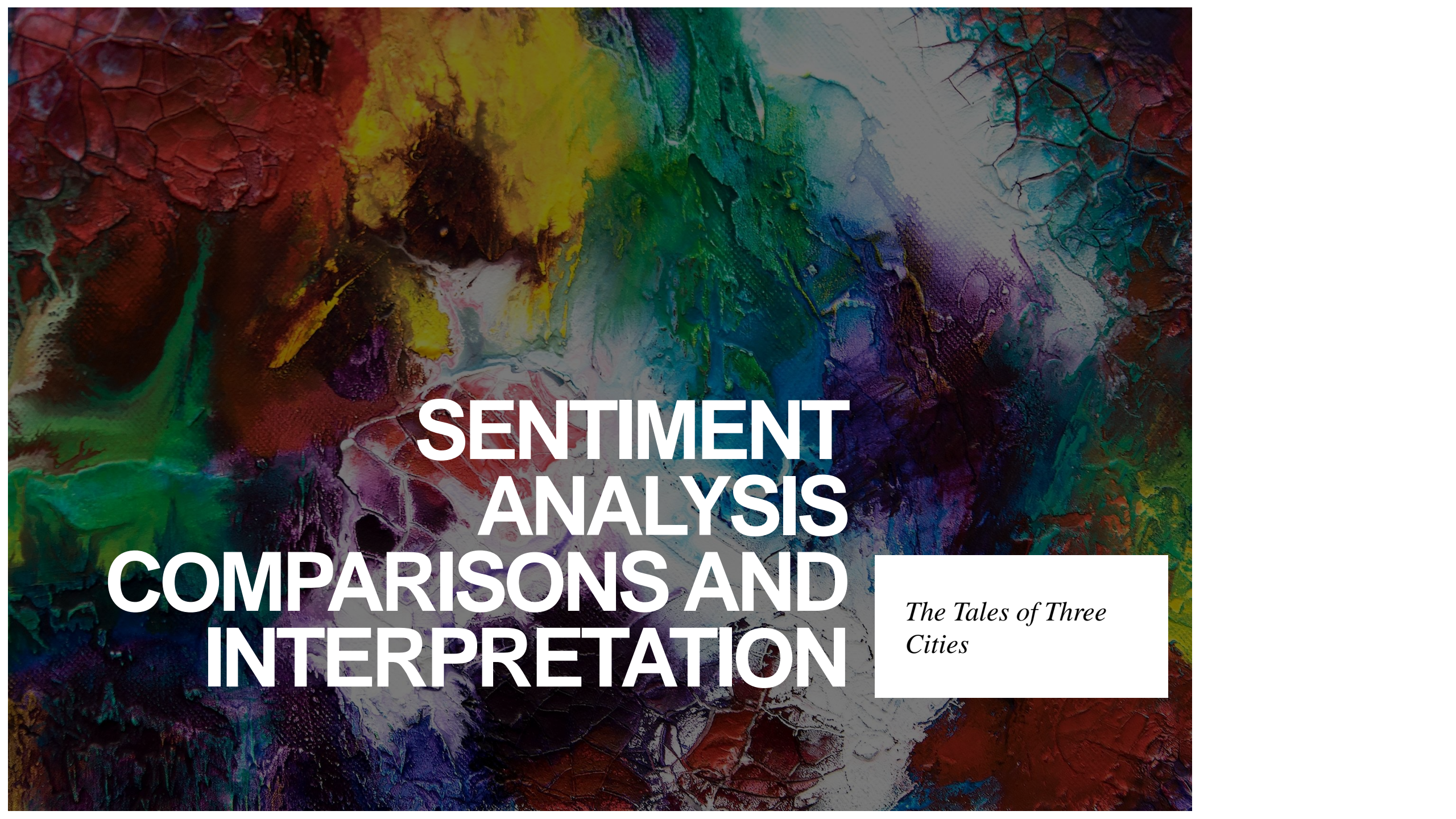
1

Problem Recognition
Information Search

Web search, word-of-mouth. **Effect:** Close to never-ending search.
Travel website, travel agent. **Effect:** Slow decision-making.

2

Evaluation of Alternatives

Multiple choices. **Effect:** Information overload.

3

Purchase Decision
Purchase

Price and perceived quality perception.
Effect: Unknown outcomes (buyer's beware).

4

Post-Purchase Evaluation

What you experienced is what you get.
Effect: If the experience is not great, you will not book again.

THE SCIENCE OF REVIEWS IN THIS ATTENTION DEFICIT WORLD: HYPER-REVIEW WORLD

1

Problem Recognition
Information Search

Find a place to stay. **Effect:** Filter and review the rating.
Find a place to stay via Airbnb. **Effect:** Target search.

2

Evaluation of Alternatives

See reviews of place and host. **Effect:** Quick based on others' experience as an insurance of great experience.

3

Purchase Decision
Purchase

Look at reviews to reduce information overload.
Effect: Fast decision-making.

Post-Purchase Evaluation

Now you have a say.
Effect: If the experience is not great, you can leave a negative comment..

4

BUSINESS OBJECTIVES

- To extract and explore Airbnb customer reviews.

- To compare and contrast Airbnb customer reviews based on three very unique UNESCO Cities of Literature - Dublin, Edinburgh, and Melbourne.

SENTIMENT ANALYSIS COMPARISONS AND INTERPRETATION

The Tales of Three Cities

UNESCO CITIES OF LITERATURES

DUBLIN
IRELAND

EDINBURGH
UNITED KINGDOM

MELBOURNE
AUSTRALIA

SENTIMENT ANALYSIS THROUGH NLTK WORD CLOUD

Dublin

English Comments

SENTIMENT ANALYSIS THROUGH NLTK WORD CLOUD
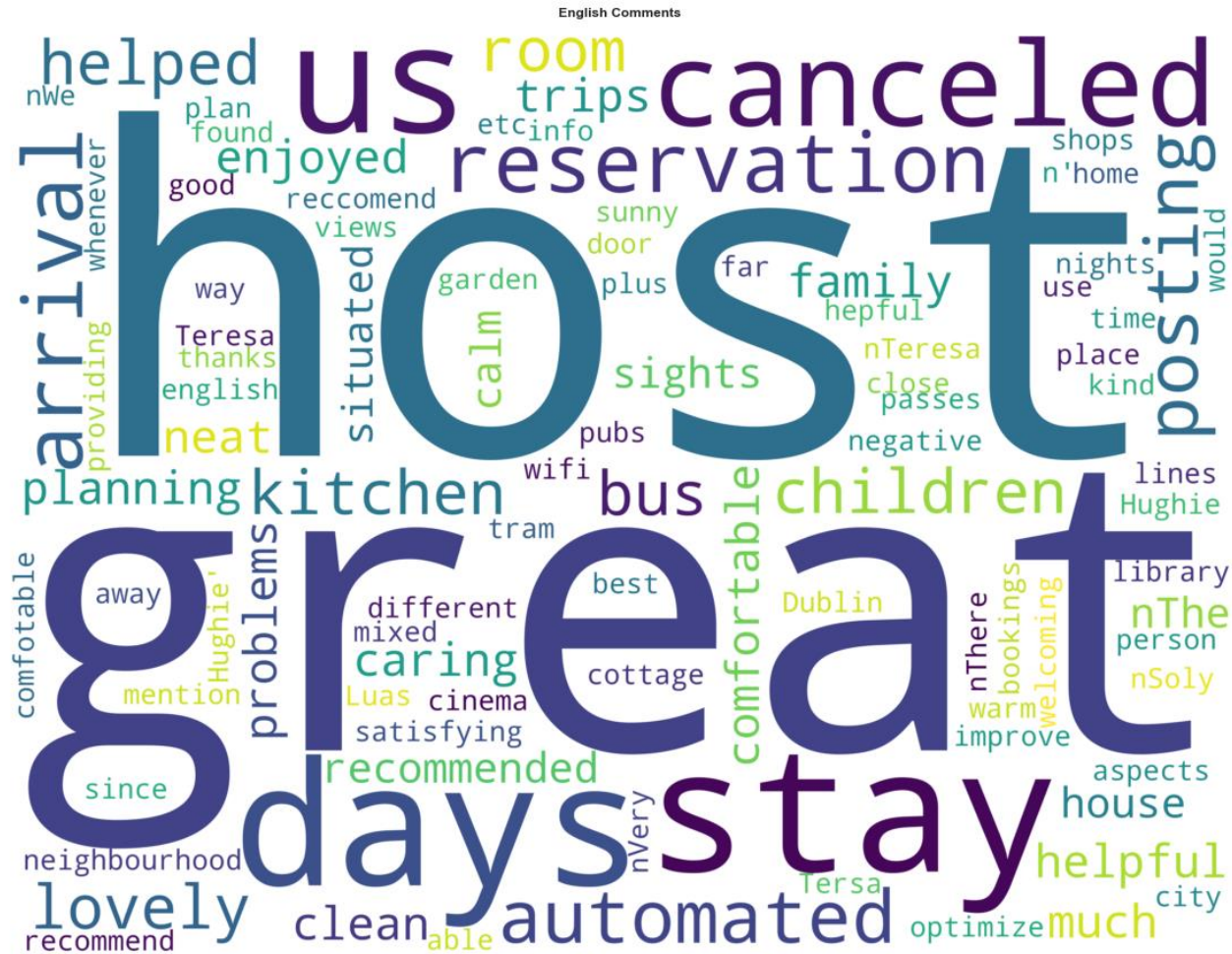
Edinburgh
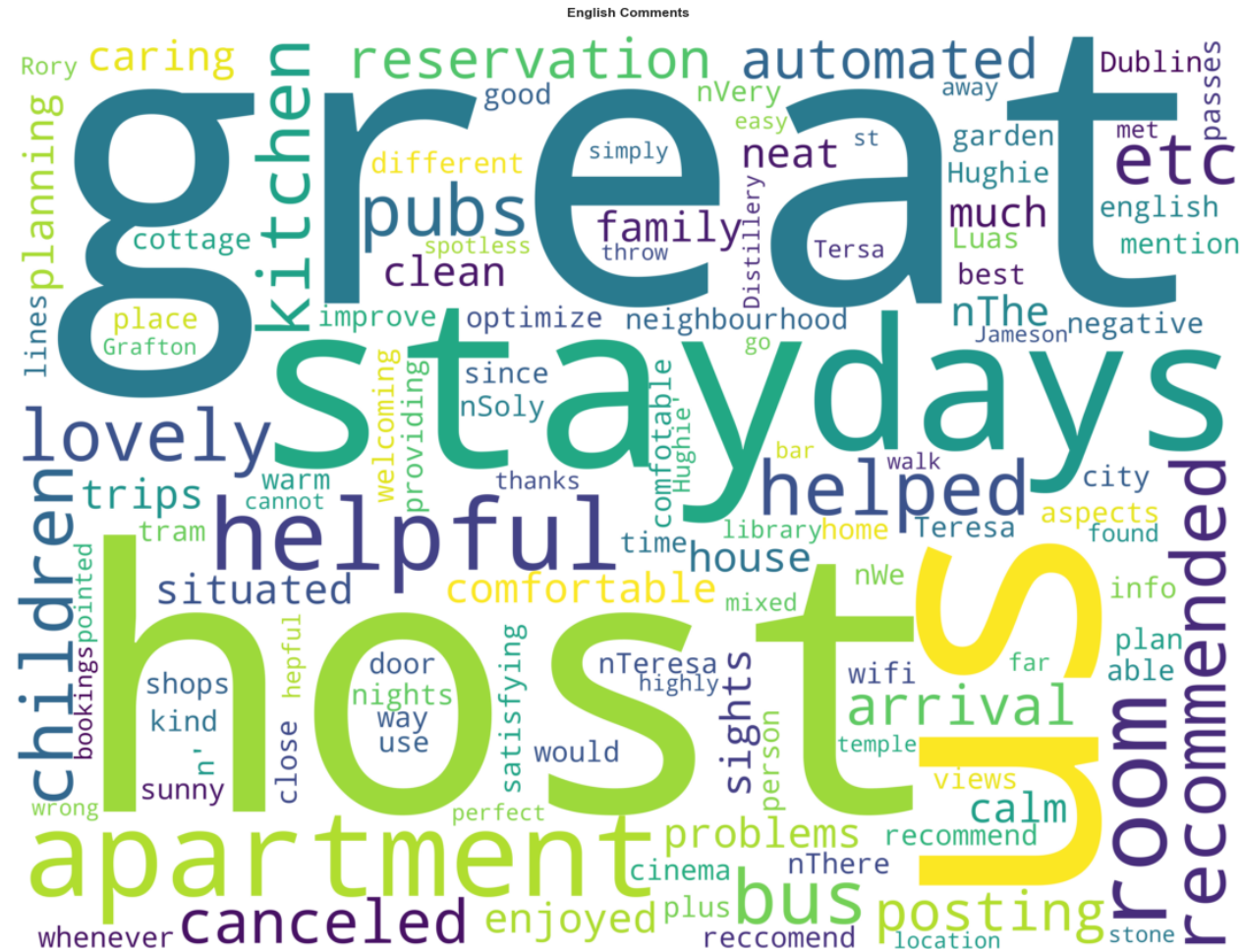
SENTIMENT ANALYSIS THROUGH NLTK WORD CLOUD
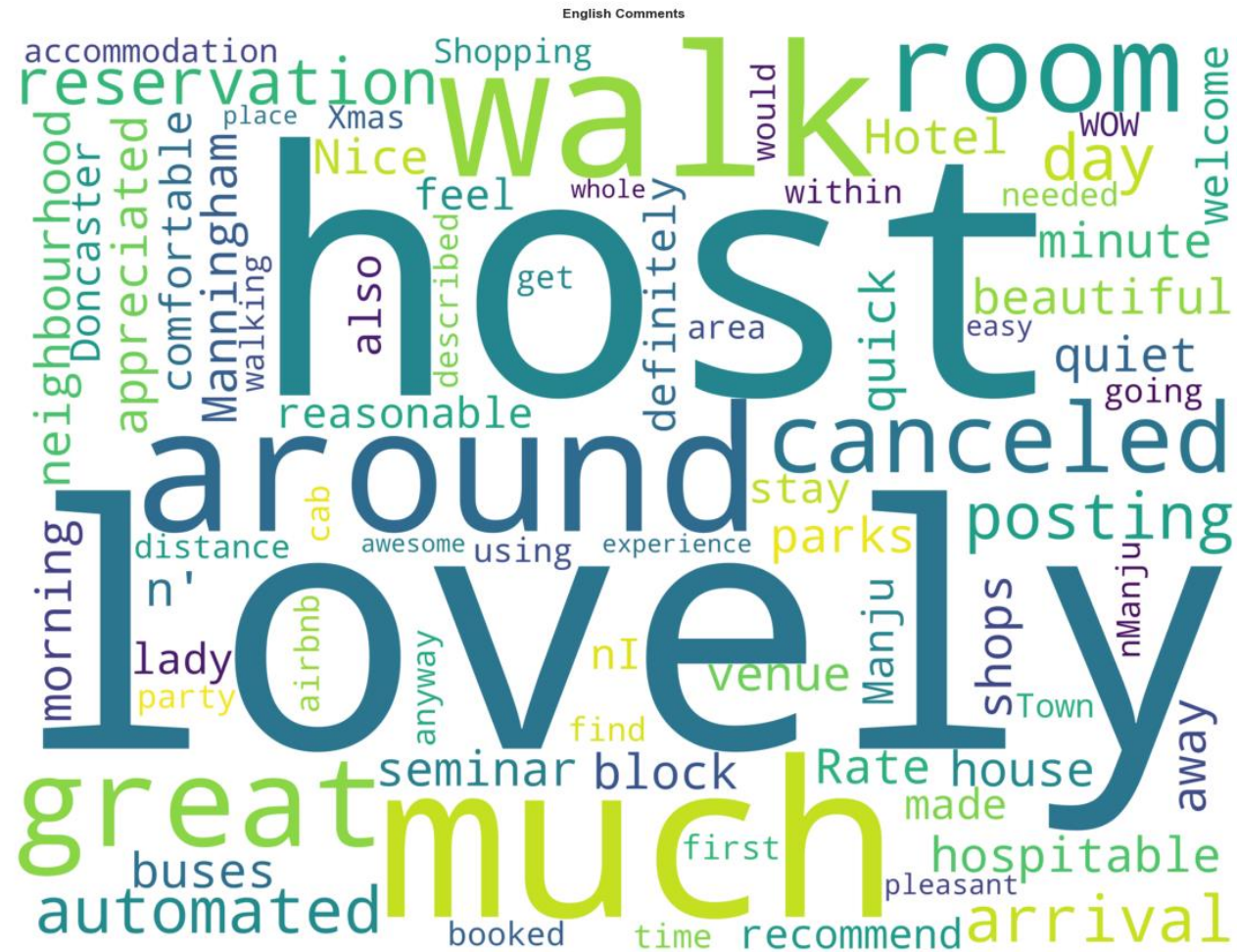
Melbourne

PERCENTILE NEGATIVE, POSITIVE, AND VERY POSITIVE COMMENTS

Dublin

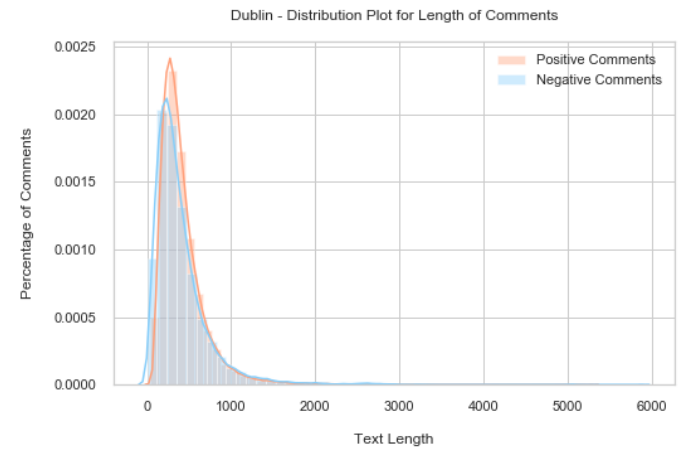- Positive Comments — 22.43%
- Negative Comments — -0.81%
- Very Positive Comments — 78.38%

Edinburgh

- Positive Comments — 22.48%
- Negative Comments — -1.01%
- Very Positive Comments — 78.54%

Melbourne

- Positive Comments — 25.61%
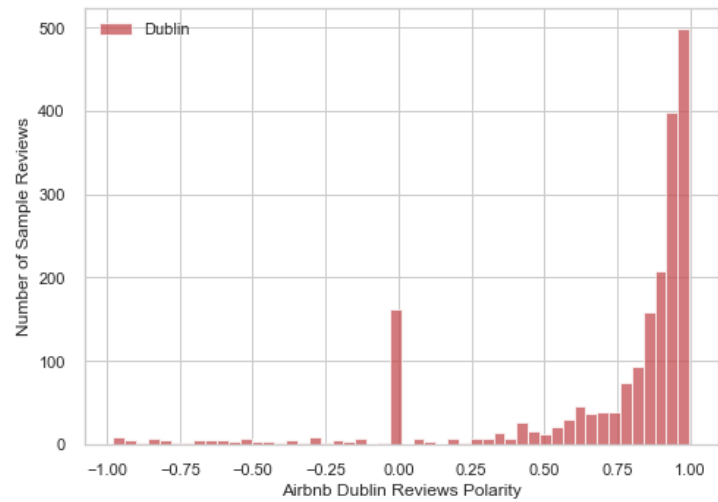- Negative Comments — 0.00%
- Very Positive Comments — 74.39%
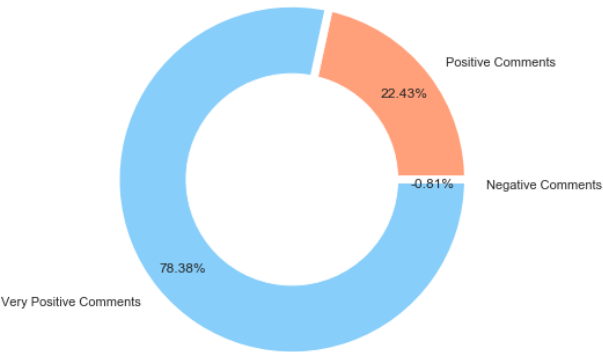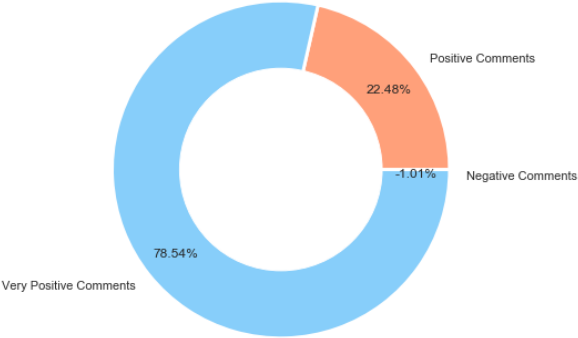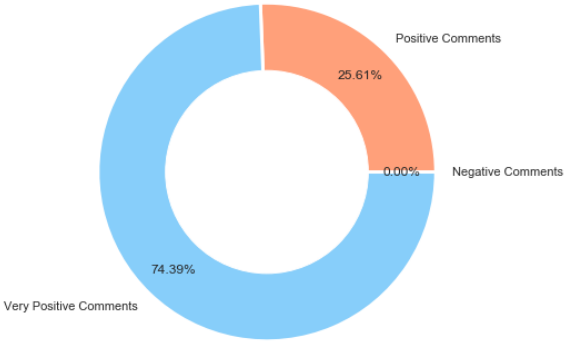
POSITIVES AND NEGATIVES COMMENTS THROUGH DISTRIBUTION SEABORN PLOT
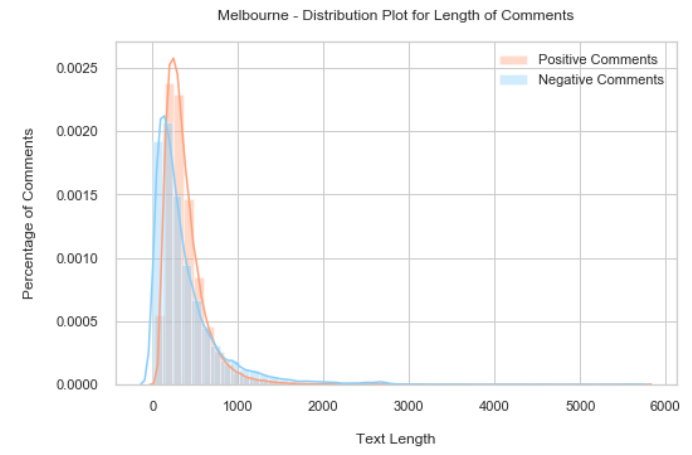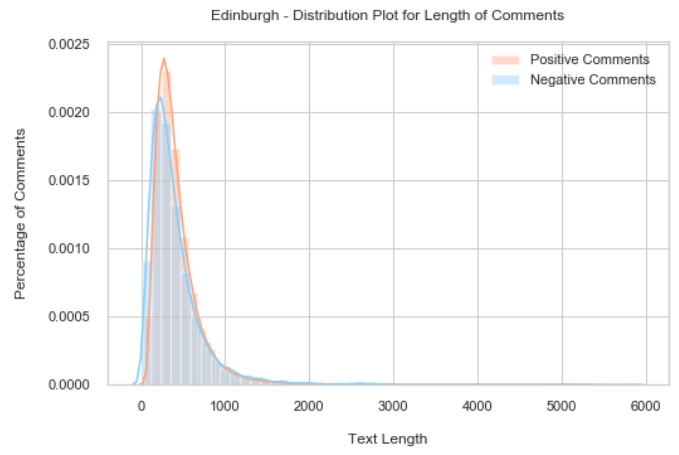
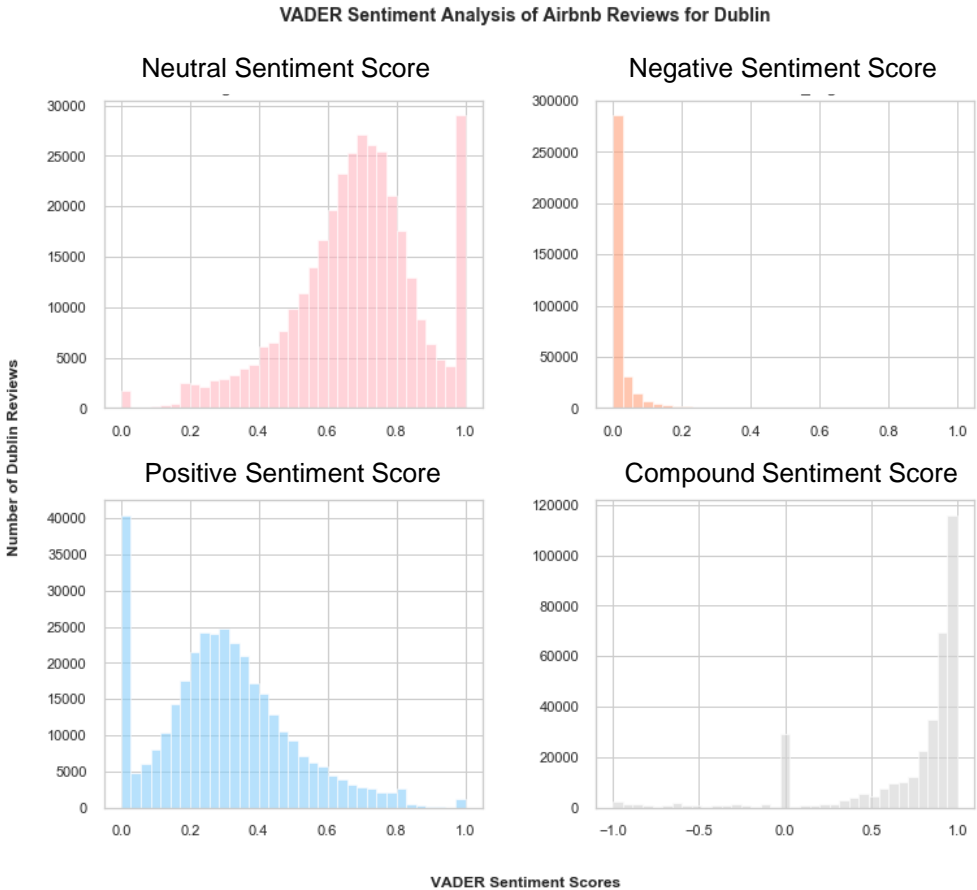# RULE-BASED SENTIMENT ANALYSIS TOOL CALLED VADER
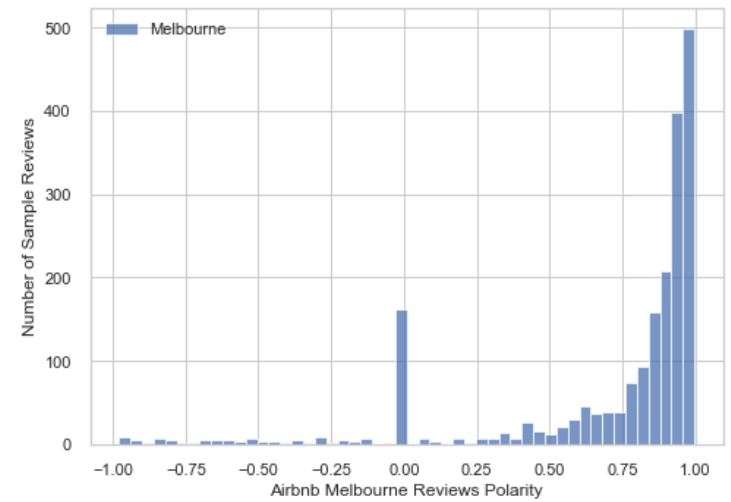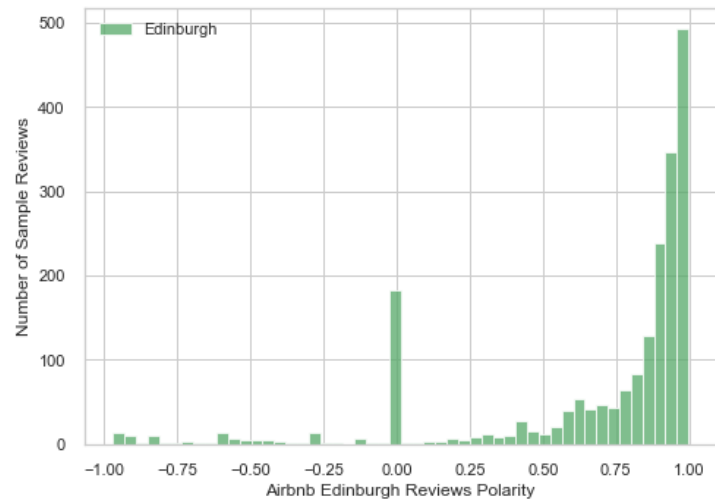
VADER SENTIMENT ANALYSIS
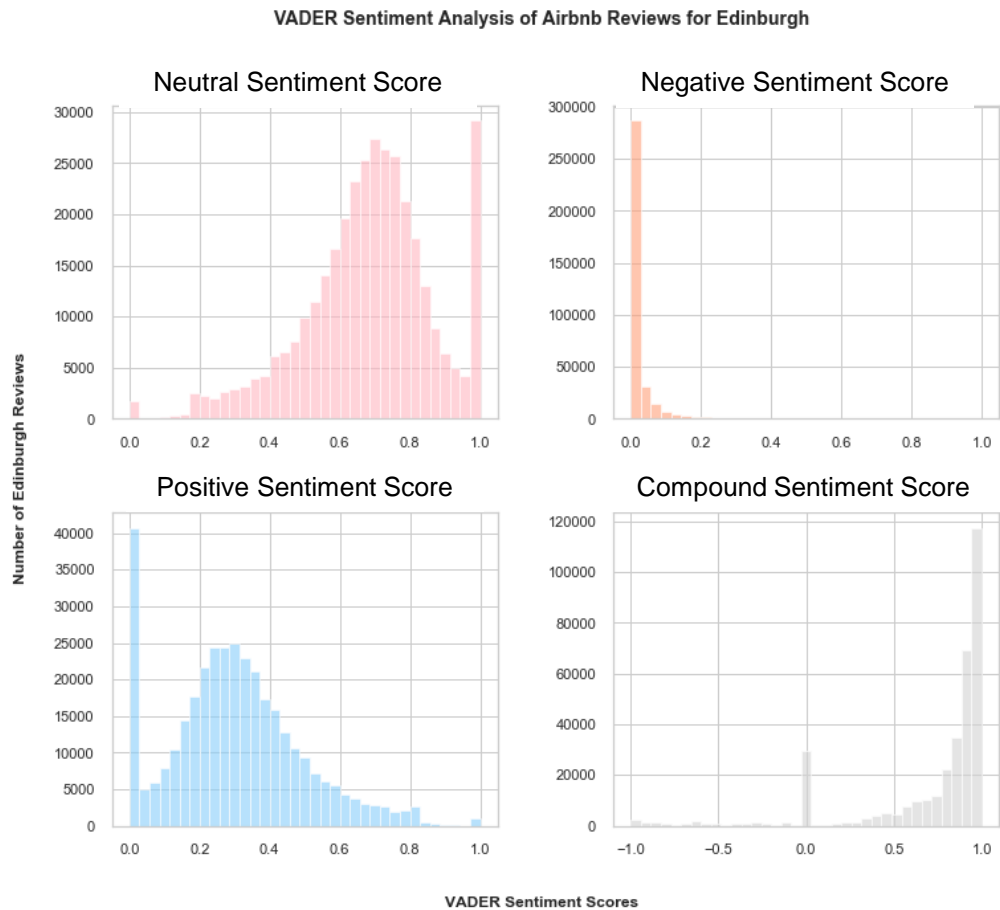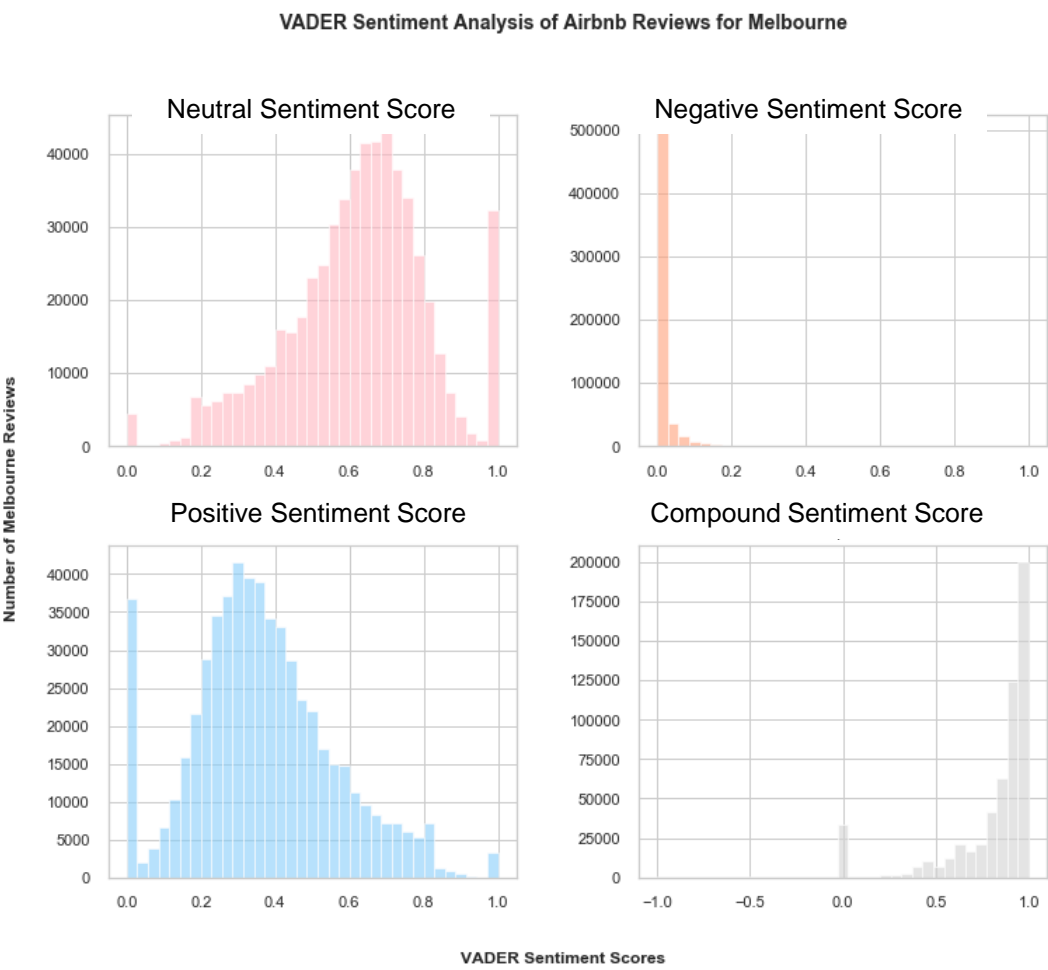
Dublin

VADER Sentiment Analysis of Airbnb Reviews for Dublin

VADER SENTIMENT ANALYSIS

Edinburgh

VADER Sentiment Analysis of Airbnb Reviews for Edinburgh

VADER SENTIMENT ANALYSIS

Melbourne

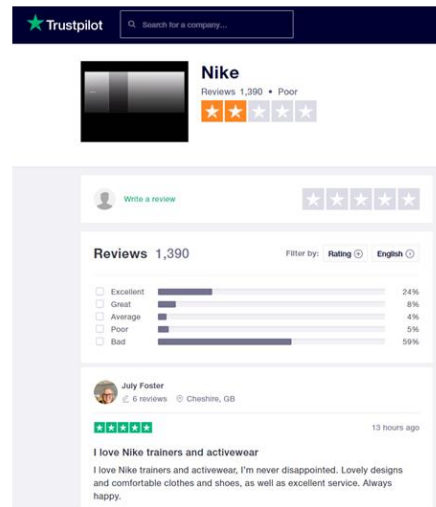VADER Sentiment Analysis of Airbnb Reviews for Melbourne

BUSINESS RESEARCH OUTCOMES

Data Science +

RECOMMENDATIONS FOR BUSINESS DECISION-MAKERS (AIRBNB EXECUTIVE)

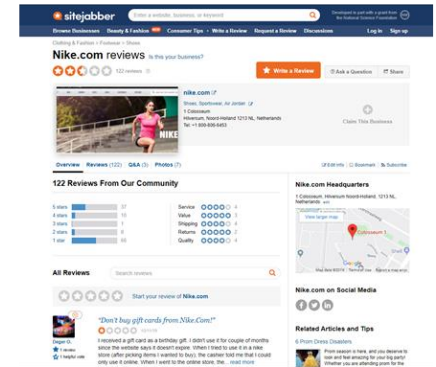*Build a better customer reviews site by 'stealing with pride' from the best of the best*

Trustpilot about giving voice to every review is an experience that matters
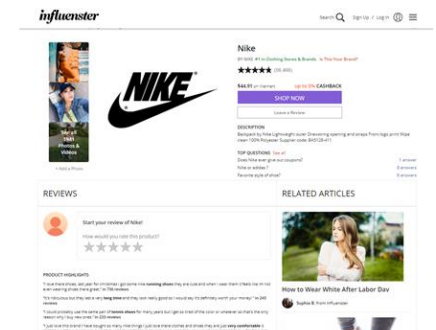

ProductReview is a community of consumers helping each other make better purchasing decisions.
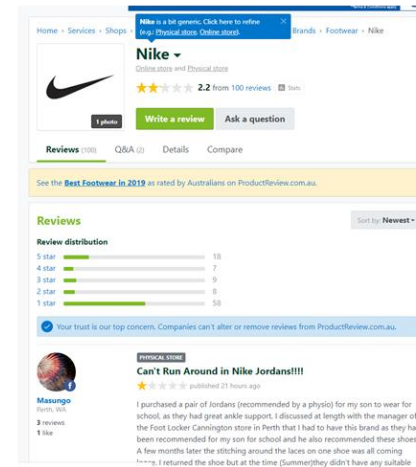
Sitejabber is helping smart buyers choose the best businesses

Reviews.io is about growing business and build trust with reviews

Influenster is a product discovery platform
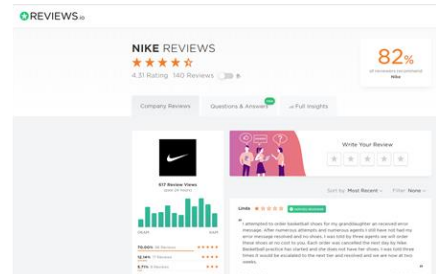
RECOMMENDATIONS FOR DATA SCIENTISTS

Learn the way of design thinking, design sprinting, and agile development

LIMITATIONS AND FUTURE SCOPE

Next Chapter

LIMITATIONS

The Deep Neural Network (DNN) not working as expected; need to find root cause. *What can do to debug?*

Predictive Modeling – limiting to Random Forrest, SVM, XG Boost, and Logistic Regression. *Can I do more?*
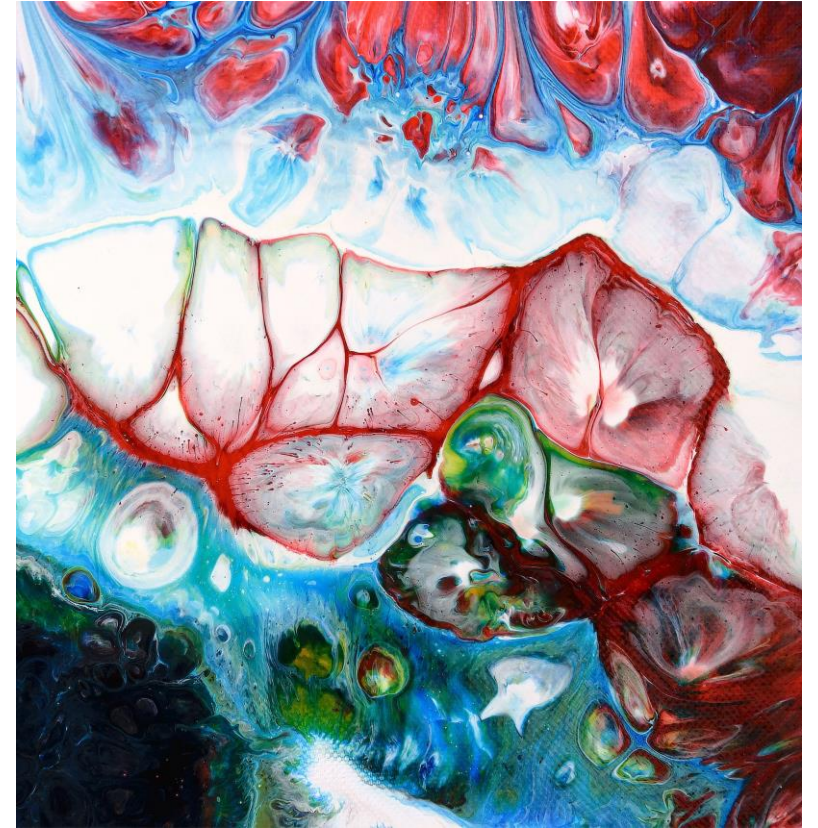
FUTURE SCOPE

Experimenting with *Phrase Net*. Connecting key words in a text using lines to show linkages.

Experimenting with *Word Tree*. Displaying of the words in qualitative dataset, where frequently used words are connected by branches to the other words that appear nearby in the data.

APPENDIX
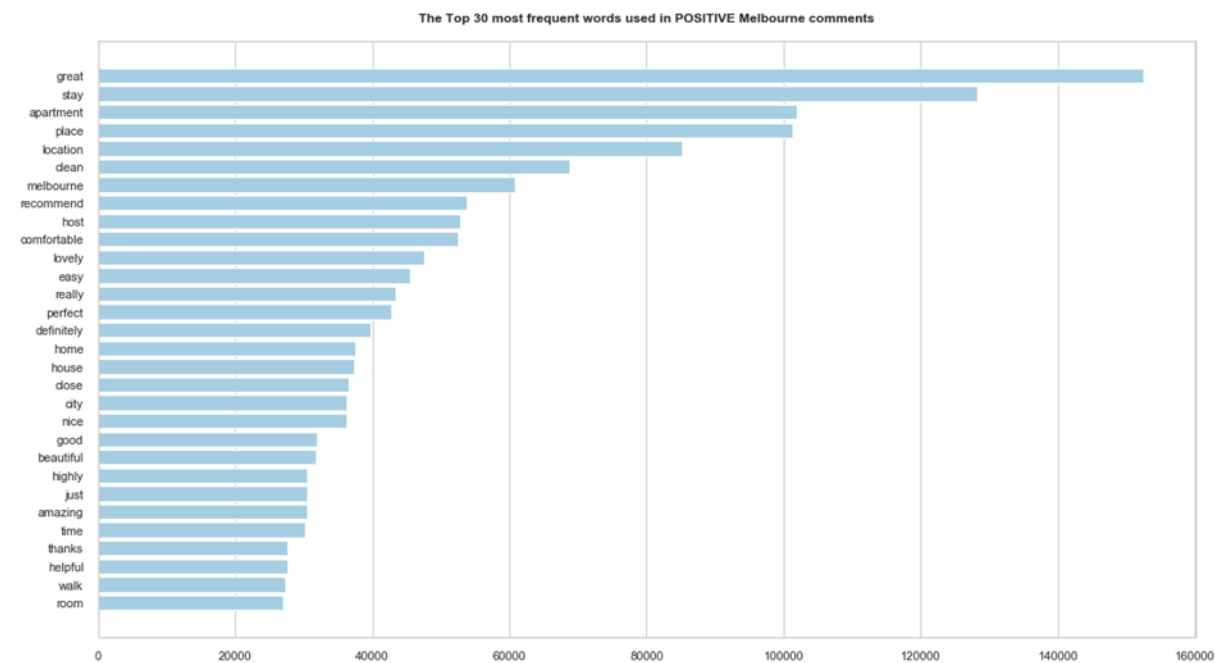
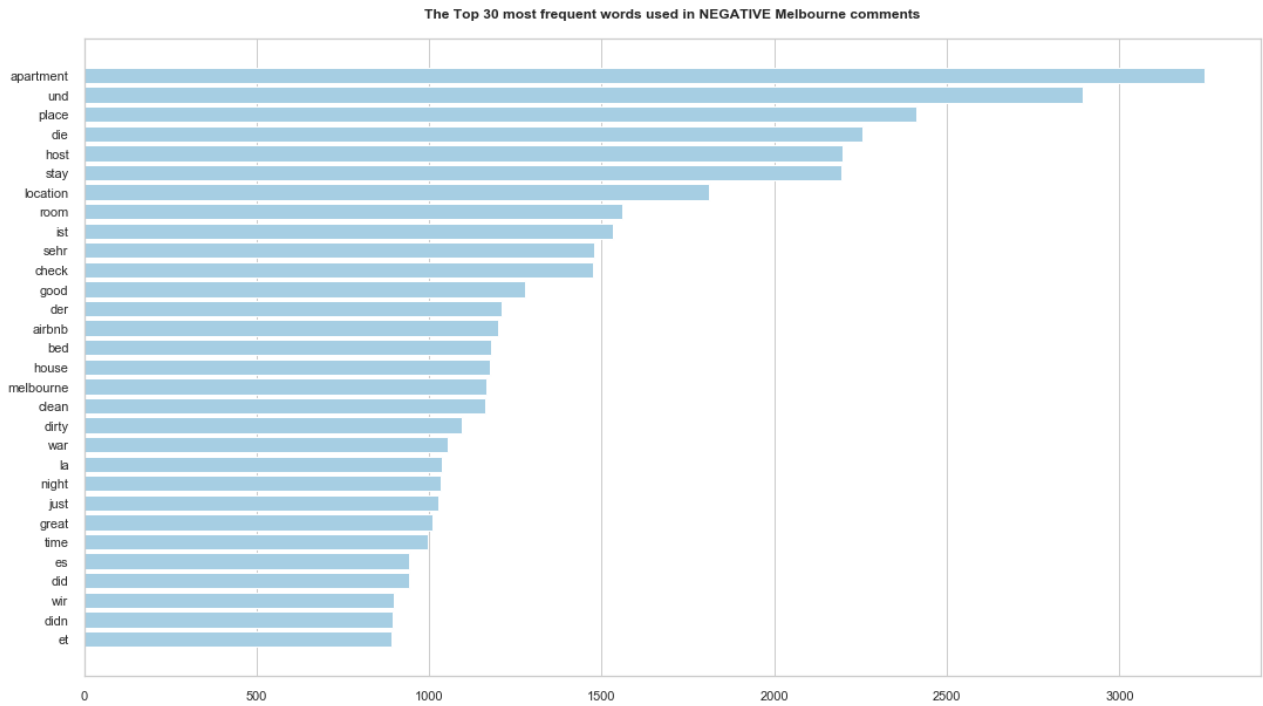References, additional analysis and figures

POSITIVE SENTIMENT ANALYSIS THROUGH NLTK WORD CLOUD

Melbourne

Melbourne - Positively Tuned Comments

NEGATIVE SENTIMENT ANALYSIS THROUGH NLTK WORD CLOUD

Melbourne

Melbourne - Negatively Tuned Comments

NEGATIVE SENTIMENT ANALYSIS THROUGH NLTK TERM VECTORIZATION

Melbourne

The Top 30 most frequent words used in POSITIVE Melbourne comments

NEGATIVE SENTIMENT ANALYSIS THROUGH NLTK TERMS VECTORIZATION

Melbourne

The Top 30 most frequent words used in NEGATIVE Melbourne comments

THANK YOU
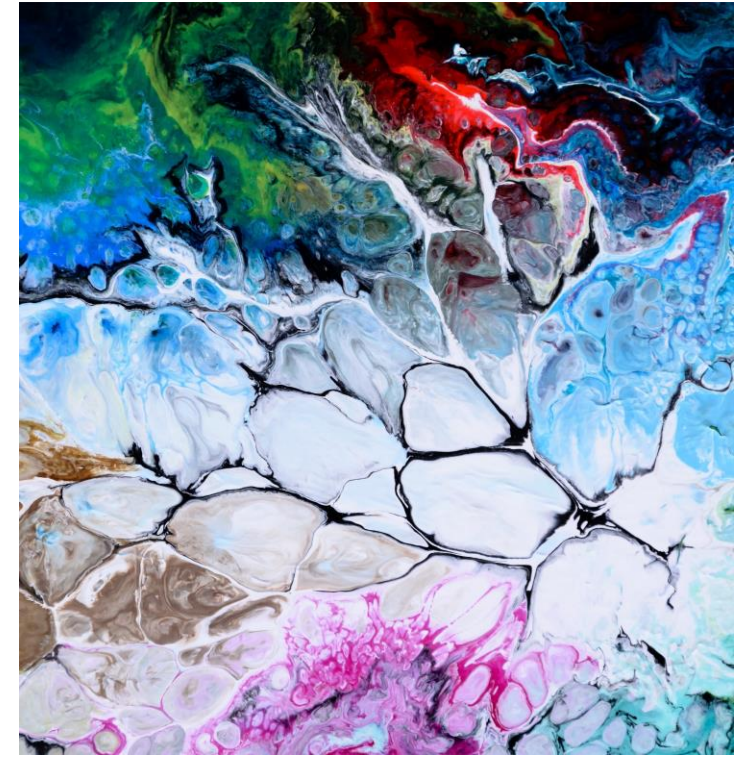
Andrew Wong

📱 *+61 416527928*

✉ *andrewwongls@outlook.com*

🔗 *LinkedIn: linkedin.com/in/andrewongls*

GitHub: github.com/andrewongls

Blog: medium.com/human-science-ai