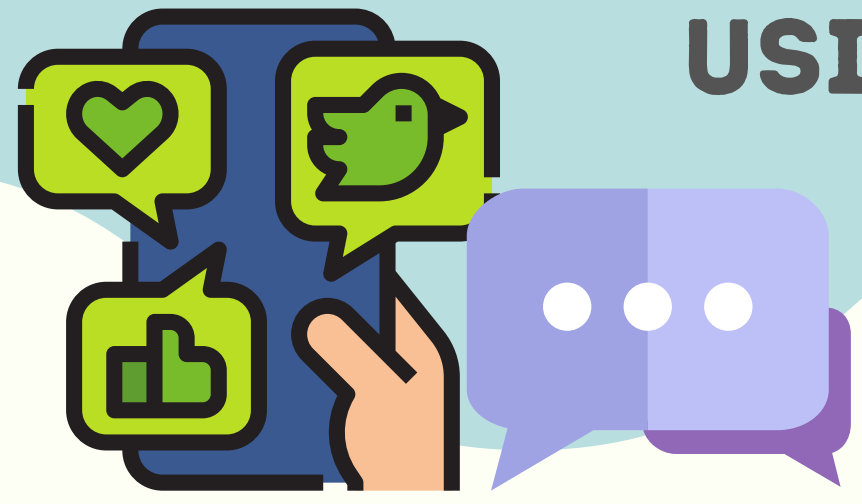


# OPINION ANALYSIS ON TWEETS ABOUT HOUSING

## USING NATURAL LANGUAGE PROCESSING TECHNIQUE



### INTRODUCTION

PLEASE FEEL FREE TO  
GIVE ADVICE FOR  
IMPROVEMENT

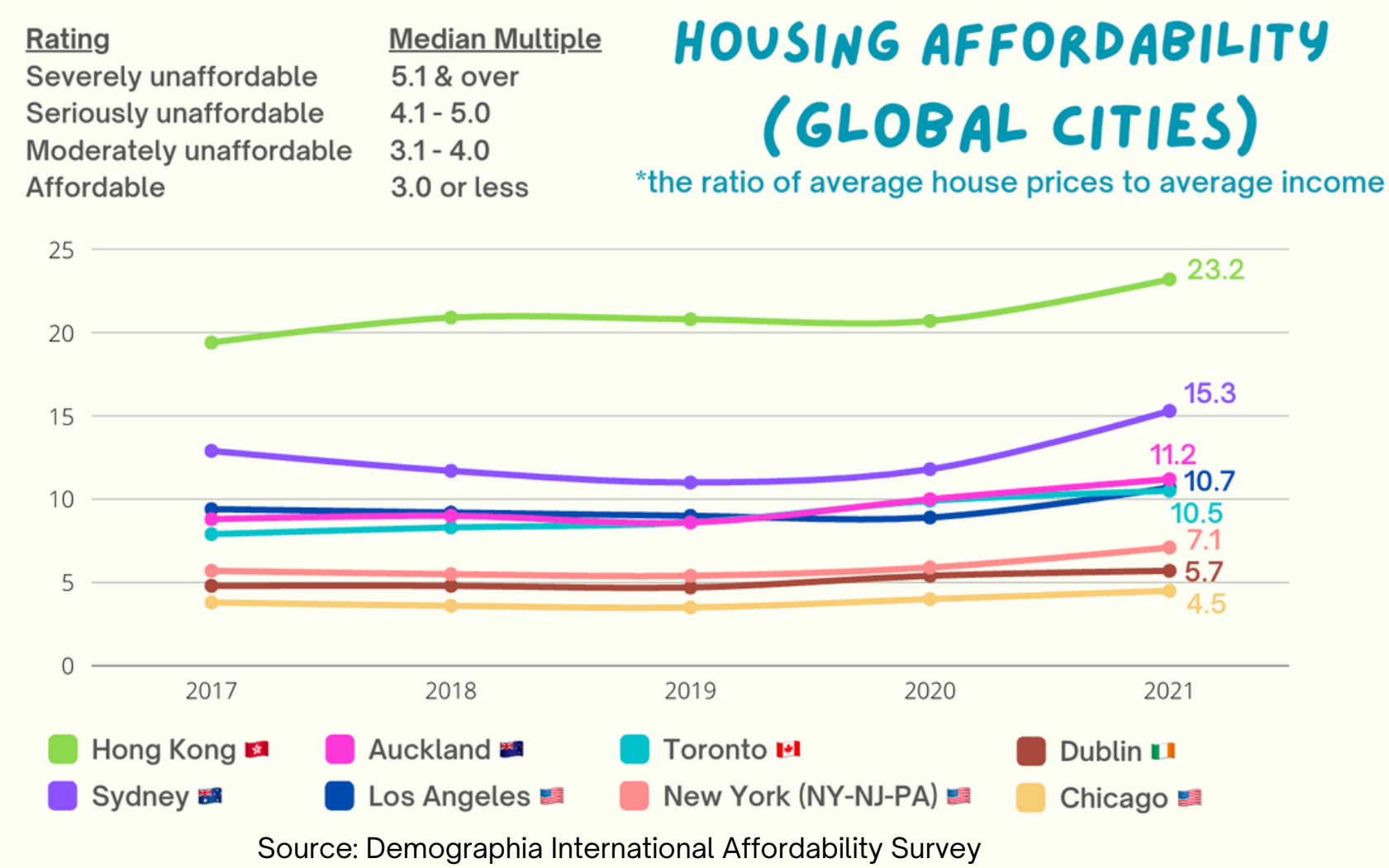
BY THAPANEE PUNNAVIRAT  
thapanee.punnavirat@gmail.com



**Housing crisis** becomes a topic creating concerns for many people around the world mostly due to a lack of housing that is affordable to purchase or rent. Referring to the World Bank (2021)'s statement, 1.6 billion people tend to be impacted by the global housing shortage by 2025. In many countries, house price increases dramatically without corresponding to income growth which accelerates at a slower rate. According to the data of Demographia International Housing Affordability updated on March 2022 by Urban Reform Institute and Frontier Centre for Public Policy, the cities such as Hong Kong, Sydney, and Dublin have ratings on unaffordability at a severe level (median multiple at 5.1 or above).

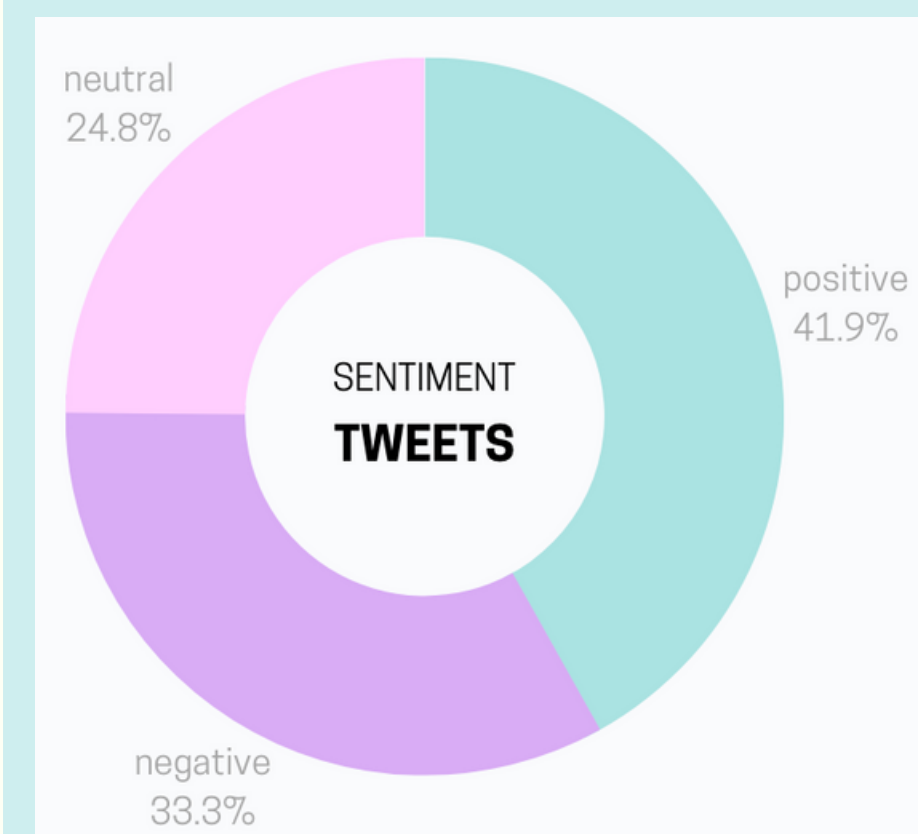
The housing crisis has been caused by many factors. One of the factors is the impact of low-interest rates and the shift to remote work in many countries during the COVID-19 pandemic, which accelerates the demand for houses. However, many aspects of housing problems are still implicit and not yet mentioned widely.

Learning opinions relating to housing from society in a most recent and constant manner will lead to new insights into housing issues. For these reasons, this project has an objective to study the top-mentioned keywords, the sentiments, and the summaries of the tweets about housing by using the Natural Language Processing technique, which some challenges encountered from the sentiment analysis will also be introduced.

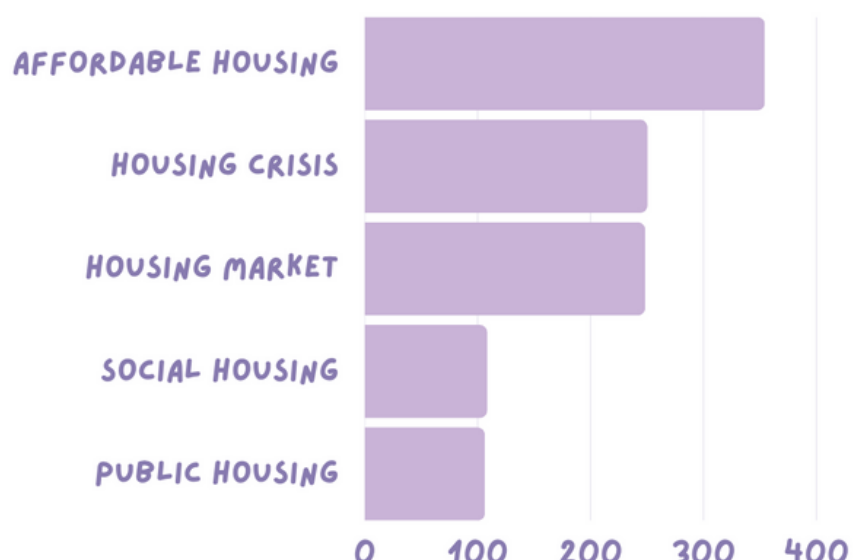


### RESULTS

After extracting words by using the Bi-gram technique, "affordable housing", "housing market", "housing crisis", "social housing", and "public housing" are ranked the top 5 mentioned keywords in tweets about housing respectively. When analyzing the sentiment of tweets based on these top keywords, "affordable housing" and "social housing" reflect positive sentiment, while the tweets with the "housing market", "housing crisis", and "public housing" keywords are more negative side. However, the overall outcome of the sentiment analysis, which also includes other keywords other than these top five keywords, shows that the majority of the tweets have positive sentiment (41.9%), followed by negative sentiment at 33.3% and neutral sentiment at 24.8%.



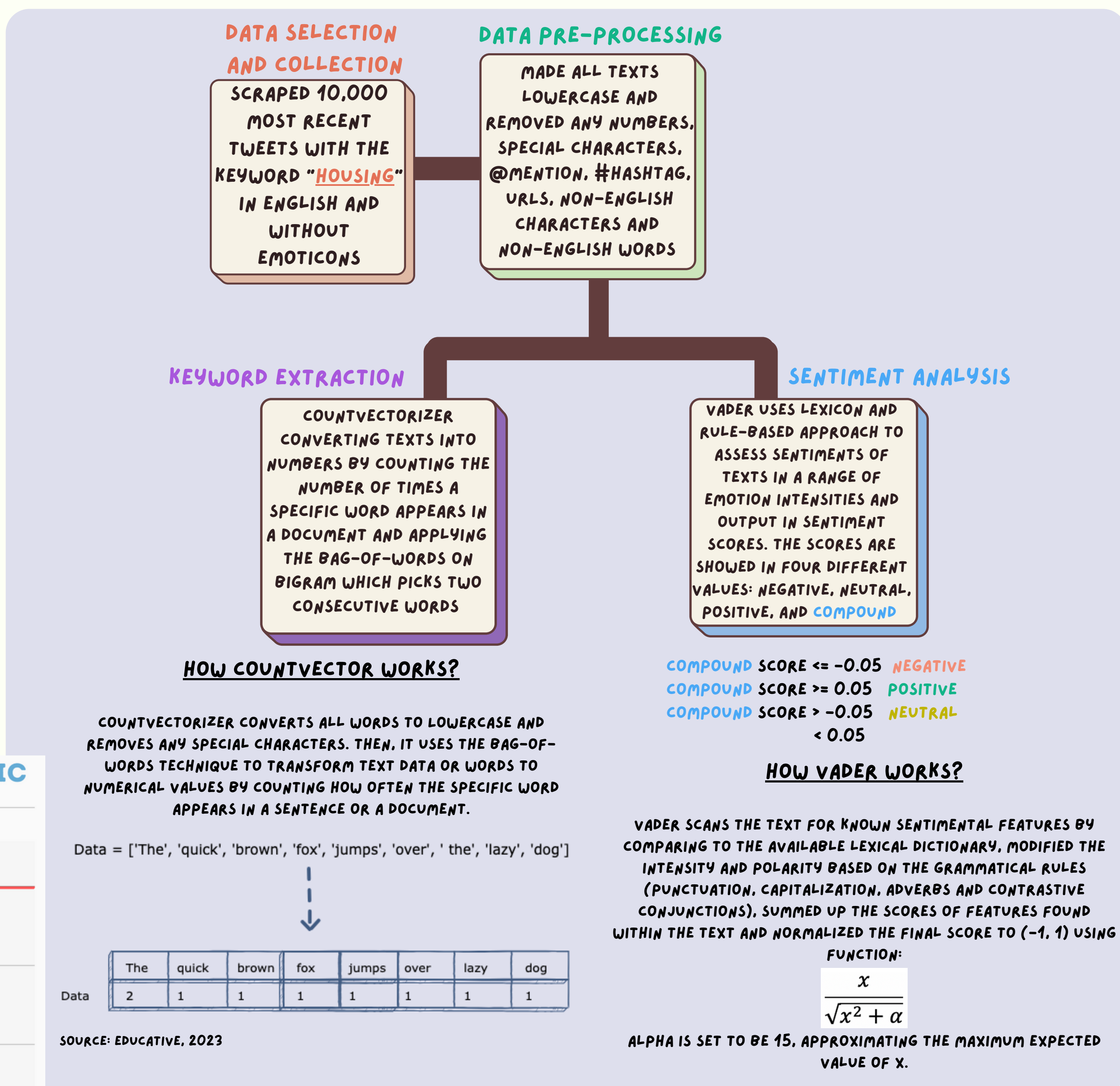
#### THE 5 MOST FREQUENT BI-GRAMS IN THE DATASET



#### REFERENCES:

Calderon, P. (2017). VADER Sentiment Analysis Explained.  
Delteil, C. (n.d.). Unsupervised Sentiment Analysis With Real-World Data: 500,000 Tweets on Elon Musk.  
Educative. (2023). CountVectorizer in Python.  
Jain, P. (2021). Basics of CountVectorizer.  
The Housing Agency. (2022). House Price to Income Ratio.  
World Bank. (2021). Can new technology solve the global housing crisis?

### WORKFLOW



### CHALLENGES

#### DOMAIN-SPECIFIC DICTIONARY OR CORPUS

The polarity of opinion words may vary for each specific domain. A positive sentence in one domain may be a negative sentence in another domain. For instance, the word "huge". "This chair is huge" may sound positive, but "This card is too huge to put in my bag." may sound negative.

#### SARCASM AND IRONIC PHRASES

Sometimes, the model cannot detect that the sentences containing positive words are actually sarcastic. For example, the model can predict a positive sentiment for the sentence "Brilliant! I am fired today." because it contains the word "Brilliant!"

#### NEOLOGISM

Languages change all the time. New words are created by groups of people from time to time. The lexical dictionary or corpus needs to be updated constantly to be able to predict sentiment analysis accurately.