

(a) Key Themes

- luxe hair care
- unite
- happy mother's day
- island adventure
- toons box ship
- travel essential
- welcome tauerville
- myths legends
- leather suede
- home holidays
- modern man
- cracking case
- deadpool home
- moodboard year
- rush relax
- this mean war
- hydrate
- magical mystery
- operation mystery box
- the irish cottage
- galactic
- women power
- mad scientist
- hamilton
- pool party
- good morning sunshine
- darkness
- robots
- natural beauty
- grateful thankful
- japan's iconic snacks
- wonderland
- sorting hat
- banned books
- watching
- happy happy halloween
- christmas
- rogues rebels
- hello sanrio
- serenity
- legends
- yule ball
- happy new year
- the world traveler wonder
- book musical

(b) Analysis Plan

- Preprocess data set :-
 - From the given DOM webpages, we create the BeautifulSoup object by converting it into lxml format
 - We extract all the reviews, urls, keywords, tags and other metadata from the retrieved lxml model
 - And store them in the csv file processed.csv
- Text processing :-
 - change case to lower
 - tokenize text and remove punctuation
 - remove words that contain numbers
 - remove stop words

