Reddit Live Dashboard

A. Preparing Reddit Live data:

For analysis the following cleaning is done on post comments:

1). **Fixing contractions**: Words containing "'" are first expanded, such as "I'm" is expanded to "I am"
2). **Removing special characters**: Special characters are removed from comments, so that there is no breakage in comments in to rows while processing the comment texts
3). **Lower case**: All characters in comment are converted to lower case
4). **Removing stop words**: A few words are very common in comments and do not carry any sentiment, words like 'begin' and 'news'
5). **Removing words with length less than 3**: Further words less than 3 characters are filtered out and replaced by empty strings
6). **Slang removal**: Reddit is known for posts which have a lot of slangs in the post comments. These slangs are removed from analysis
7). **Removing additional stop words**: Consists of using stop_word data frame from tidy text
8). **Blank removal**: In some cases, the comment rows are empty after the above steps, these rows are filtered out before sentiment analysis.

B. Steps to run Reddit Live Dashboard:

1. Fill Google sheet

The google sheet is setup to input the keyword that needs to be searched on Reddit. This google sheet also acts as an input to the R executor workflow. Below is the google sheet with sample entry:

A	B	C	D	E	F
Reddit_keyword	Name	Email	Date(YYYY-MM-DD)	Insight_Flag	Relevancy_(1- 10)
Avon Cosmetics	[redacted]		2018-07-23	No	7
BMW			2018-06-21	No	6
Weight Watchers			2018-06-21	No	6
Coventry University			2018-06-21	No	6
ASDA			2018-06-21	No	6
Travis Perkins			2018-06-21	No	6
DOMU			2018-06-21	No	6
Flybe			2018-06-21	No	6

Google sheet for entering reddit keywords

Only the Reddit_ keyword, date, and insight flag, relevance is used in the R script. The remaining columns are just used to keep tab of the people using the reddit insight automation.

2. Fetching Reddit Data

The keyword is extracted from the google sheet from the first row that has 'Yes' as the insight flag of the google sheet.

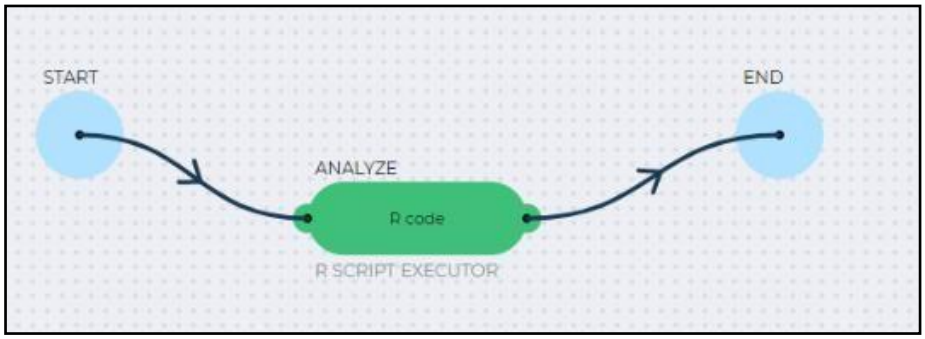
```
# Variables to input data
#Reading Reddit Keyword and Start Date
data_main$Reddit_keyword[rownum[1]-1] <- paste0(data_main$Reddit_keyword[rownum[1]-1])
data_main$`Date(YYYY-MM-DD)`<- paste0(data_main$`Date(YYYY-MM-DD)`[rownum[1]-1])

reddit_links <- reddit_urls(
  search_terms   = data_main$Reddit_keyword[rownum[1]-1],
  page_threshold = 500
)

reddit_thread <- reddit_content(reddit_links$URL)
```

Fetching reddit threads with keyword from the google sheet

We get id, structure, post_date, comm_date, num_comments, subreddit, upvote_prop, post_score, author, user, comment_score, controversiality, comment, title, and domain.

3. Run R Script Executor workflow

An R Script Executor workflow is setup to run the above script on need basis. The user must make the entry in google sheet and then execute the workflow. Data is processed and stored in Redshift table.



R Script Executor workflow to run Reddit R code

4. Visualization on Reddit Live Dashboard

Tableau Dashboard tool has been used to visualize the Reddit Insights. It is constructed in a way to have 2 slides. Slide one, to showcase the overview insights, which would contain total comments, top performing domains, average comments per post, temporal trends for last 1 year and filter by sentiment of posts. Second slide contains the contextual analysis

which contains the top keywords and their frequency.

C. Summary