Flipkart

GRiD 4.0

2022 Campus Challenge

Extract Trends from social media data

Team Name: Vengeance

Institute Name:Anil Neerukonda Institute of Technology and Sciences (ANITS)

Team members details

Team Name	Vengeance		
Institute Name	Anil Neerukonda Institute of Technology and Sciences		
Team Members >	1 (Leader)	2	3
Name	Shaik Ahmed Razha Khan	Koralla Madhu Mita	Charitha Bhukaya
Batch	2019-2023	2019-2023	2019-2023

Deliverables/Expectations for Level 2 (Idea + Code Submission)

Deliverable 1:

Identification of trends from social media

1. Identify trends on social media based on category. Can restrict to Fashion as a category for the project. Ex: Polka dots dresses are trending on twitter.
2. Ranking/scoring logic for trends extracted.
3. Outcome format:
   a. Option1: List of trending keyword(s) along with list of sample images and respective links from which the trend is derived with most trending first:
      Example: Trends:[{Polka dot dresses, <list of links/images>,trending score}, {Bellbottom Jeans, <list of links/images>,trending score}..]
   b. Option 2: structured data according to flipkart category, sub category, vertical and product attributes
      Example: {category: Fashion, Sub-category: Women Western, vertical: Women dresses, trending attribute type: Pattern, trending attribute value: Polka Print, list of sample images and links from which the trend is derived}.
      Outcome with Option 2 format will be given bonus points.

Deliverable 2:

Mapping trends with Flipkart products:

1. Create mapping of extracted trending keyword(s) with Flipkart category, sub category, vertical and product attribute(s), search page links.
   Example:{category: Fashion, Sub-category: Women Western, vertical: Women dresses, trending attribute type: Pattern, trending attribute value: Polka Print}
   Note: Use category, Subcategory combination from the Flipkart Website
2. From a trending keyword, creating a corresponding searchable term on Flipkart which will lead to matching products.
   Example: Tropical Tops keywords will not give right results directly on Flipkart but we can construct search query for it using some intelligence.
3. Points will be given based on similarity between sample images for trends and product results on Flipkart.

Use-cases

User :

P0 - View trending keywords extracted from social media
P1 - View trending score assigned to each keyword
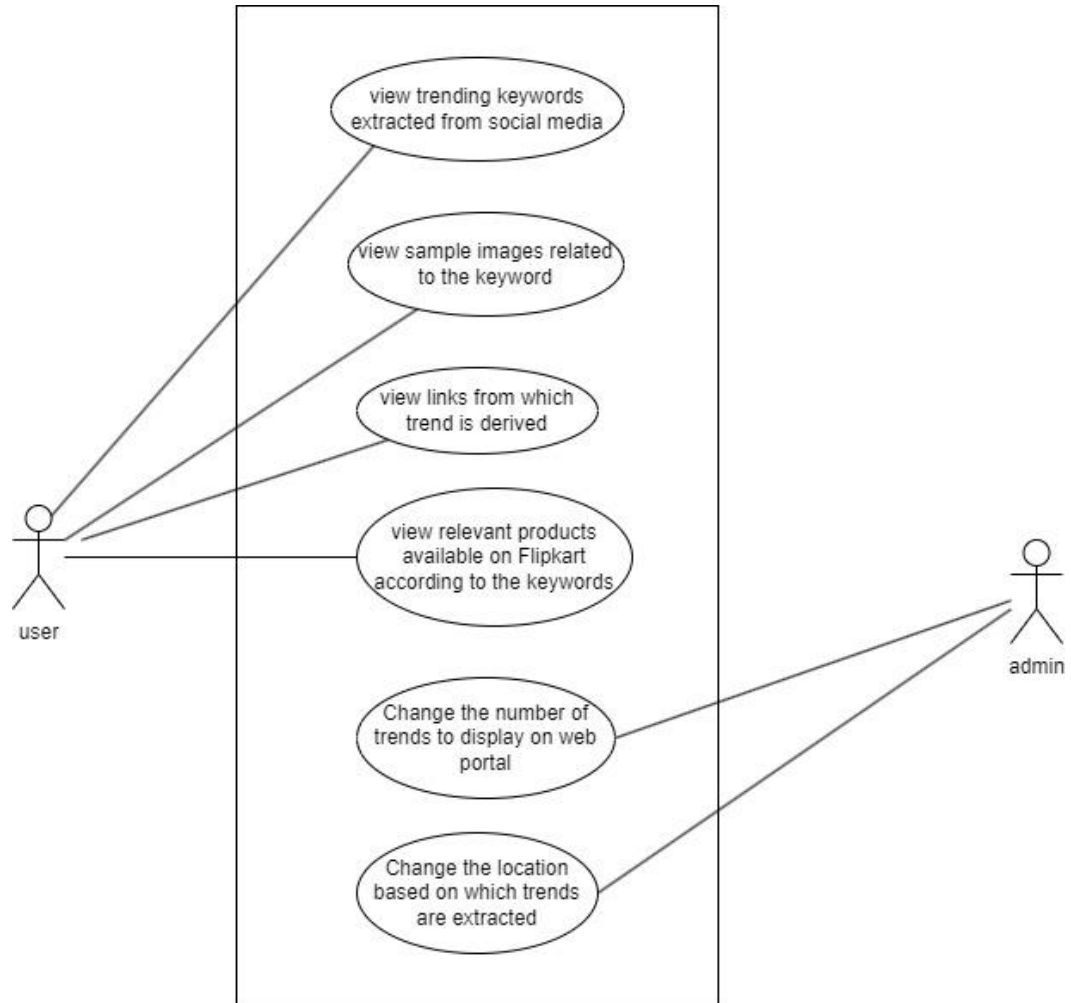P2 - View Sample images related to the keyword
P3 - View links from which the trend is derived
P4 - View relevant products available on Flipkart according to the keyword

Admin :
P5 - Change the number of trends to display on web portal
P6 - Change the location based on which trends are extracted

Solution statement/ Proposed approach

Overall Problem :

Identify trends on social media based on category. We are restricting to Fashion as a category for the project and displaying Ranking/scoring logic for trends extracted. This project is made in accordance with Option 1 in Deliverable 1.

After breaking the overall problem into sub problems, we get :

1. Get a list of trending keywords from the internet.
2. Select only those keywords that are creating a lot of buzz on social media.
3. Calculate trending scores of these selected keywords.
4. Get sample links from which the trend is derived.
5. Get sample images related to the keyword.
6. Get flipkart products related to the keyword.

Solution to Sub Problems

1.Get a list of trending keywords from the internet.
We are using google trends to tackle this sub problem. Using google trends, we can get the trending keywords present in queries searched on the Google website. We can access Google Trends data in python using Pytrends API. We are filtering results by Shopping category, Fashion sub-category and Location as India.

2. Select only those keywords that are creating a lot of buzz on social media.
Twitter has a lot of textual content, so we're relying on it to check whether these selected keywords are creating buzz on twitter or not. We are using Tweepy API to do this.

3. Calculate trending scores of these selected keywords.
We can calculate the trending score by storing number of tweets posted in the last one week, containing our trending keyword. We have sorted the keywords based on this count and have returned the result as a dictionary.

4. Get sample links from which the trend is derived.
We can search the recent tweets containing our keyword using Tweepy and get the id of these tweets, using this id we have constructed the URL of the tweet.

5. Get sample images related to the keyword.
To get the sample images, we're taking the help of icrawler, an API that's used to scrap Google Images. We're searching for related images on google images site and downloading these images into seperate folders - trend 1, trend 2....trend n. We'll use these images on our web portal.

6. Get flipkart products related to the keyword.
We're using the Flipkart scraper API to scrape search results and product details from Flipkart. We'll use our found keywords, search them on Flipkart using the API and get the product name and url. We'll display these details on our web portal and users can directly go to Flipkart from these links.

Result
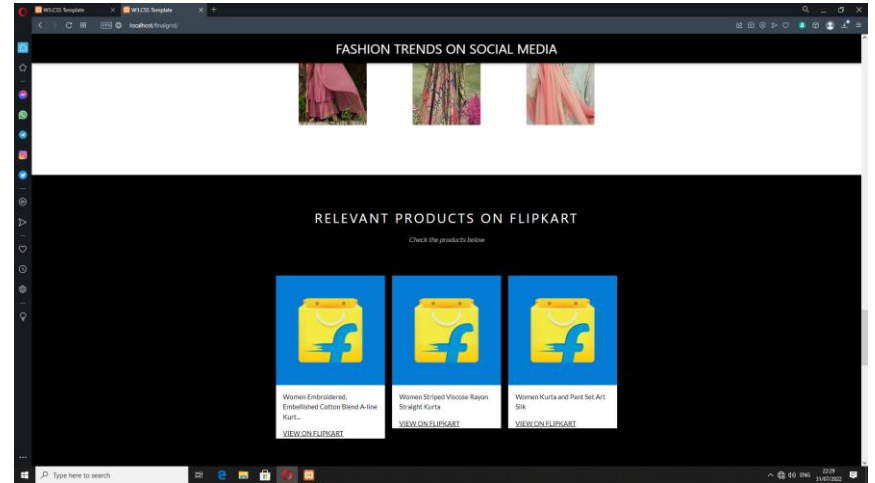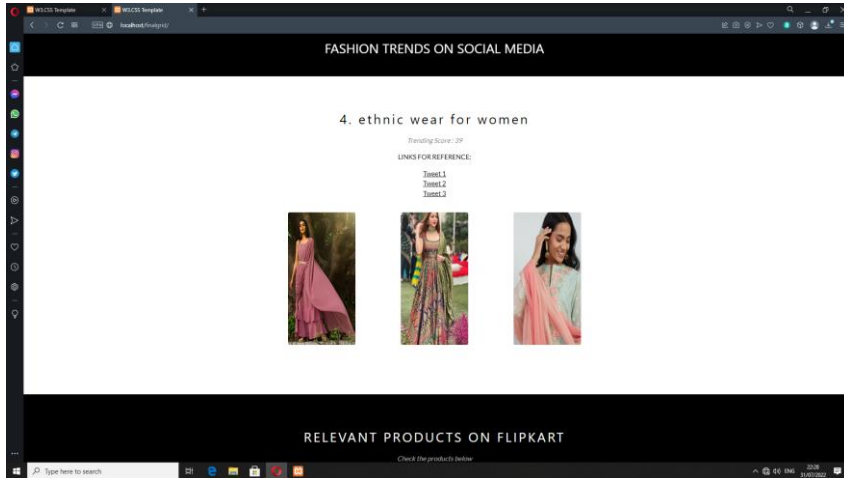
Limitations

1. Our Project depends on some APIs, if any of these APIs become unstable, the project won't work correctly.

2. The trends and tweets extracted are only based on Indian trends, the admin has to change the geocode attribute in the source code to set it to any other region.

3. Only top 5 trends are displayed, slight modifications in the code are required to get more trends.

3. The web portal needs an active internet connection to load the CSS fonts from the web.

Future Scope

1. We would like to scrape and add the Flipkart thumbnail for the Flipkart products being displayed on the web portal.

2. Besides fashion, we would also like to add many more categories or perform mapping with Flipkart categories and sub categories.

3. We would like to use Facebook, Instagram and other social media sites to calculate trending scores and improve accuracy.

4. We would also like a feature that allows the user to select the number of trends to display (As of now, it is restricted to 5).