# Data Collection Methodology – Twitter API & Sentiment Analysis

This document outlines the methodology used to collect and process Twitter data for the research project titled:  
"The Impact of Social Media Sentiment Analysis on Consumer Behavior: A Case Study of the Fashion Industry."

## 1. Objective

To analyze fashion-related consumer sentiment using real-time Twitter data, providing insights into public perception and trends in the fashion industry.

## 2. Data Source

- Platform: Twitter (https://twitter.com)

- Access Method: Twitter API v2 (Search Recent Tweets endpoint)

- API Documentation: https://developer.twitter.com/en/docs/twitter-api

- Compliance: Data collection adhered to Twitter's Developer Agreement and Policy.

## 3. Tools & Libraries Used

- Programming Language: Python

- Libraries:  
 - Tweepy (for accessing Twitter API)  
 - Pandas (for data handling)  
 - TextBlob (for sentiment analysis)

## 4. Dataset Description

- Sample Size: 1,200 tweets

- Search Query: Tweets containing hashtags like #fashion, #style, #fashiontrends

- Date Range: Recent 30 days (collected in real-time)

- Language Filter: English only

- Exclusions: Retweets were excluded for originality

Variables Included:

|  |  |
| --- | --- |
| Column Name | Description |
| Timestamp | Time when the tweet was posted |
| Author\_ID | Unique (anonymized) Twitter user ID |
| Tweet | Full text of the tweet |
| Likes | Number of likes on the tweet |
| Retweets | Number of retweets |
| Sentiment | Sentiment label (Positive, Negative, Neutral) |

## 5. Code Summary

Below is a simplified code snippet used for collecting and analyzing the tweets:

import tweepy  
import pandas as pd  
from textblob import TextBlob  
  
client = tweepy.Client(bearer\_token='YOUR\_BEARER\_TOKEN')  
query = "#fashion OR #style OR #fashiontrends lang:en -is:retweet"  
tweets = client.search\_recent\_tweets(query=query, tweet\_fields=["created\_at", "text", "author\_id", "public\_metrics"], max\_results=100)  
  
data = []  
for tweet in tweets.data:  
 polarity = TextBlob(tweet.text).sentiment.polarity  
 sentiment = 'Positive' if polarity > 0 else 'Negative' if polarity < 0 else 'Neutral'  
 data.append([  
 tweet.created\_at, tweet.author\_id, tweet.text,  
 tweet.public\_metrics['like\_count'],  
 tweet.public\_metrics['retweet\_count'],  
 sentiment  
 ])  
  
columns = ['Timestamp', 'Author\_ID', 'Tweet', 'Likes', 'Retweets', 'Sentiment']  
df = pd.DataFrame(data, columns=columns)  
df.to\_csv("fashion\_sentiment\_dataset.csv", index=False)

## 6. Ethical Considerations

- Only public data was collected.  
- No personal identifiers (usernames, emails, location) were stored.  
- The dataset is used solely for academic purposes.

## 7. Conclusion

This method provides a verifiable, ethically sourced dataset with both qualitative (text content) and quantitative (likes, retweets, sentiment score) features, suitable for sentiment analysis in consumer behavior research within the fashion industry.

Prepared by: Anonymous

All the documents are solely for academic purposes. And any one can use the code, dataset present in this repo for their research.