

# BUMBLE

GOOGLE PLAY STORE - APP REVIEWS

Sameera  
June 2022

# INTRODUCTION

---

- Bumble is a dating application.
- Most popular dating app, after tinder.
- Profiles of potential matches are displayed to users, who can **swipe left** to reject a candidate, or **swipe right** to indicate interest.
- Users can sign-up using their phone number or Face Book profile.
- Subscription fee (3 types) or free.
- The users have options for searching romantic partners, or friends in **BFF** mode.
- In heterosexual matches, only **female** users can make the first contact with matched male users.
- The company reports having over 55 million users in 150 countries as of 2019.

# GOAL

- To analyze the reviews –
  - User reviews give a chance to make the application better.
- Topic Modeling -
  - Divide the reviews into clusters of different topics
  - Know the percentage of each topic

# DATA

---

- Data Set
  - Initially there were - 110031 entries
  - 10 columns – reviews, ratings, date, ...
  - Feature of Interest – **content**, has reviews by users
  - Different languages with script in English and Non-English
  - Extracted only the reviews in English language using **langdetect**
  - 89472 rows – Only English language entries

# TOOLS

**Pandas** – Clean, Explore and Feature Engineering

**Scikit-Learn** – Build different Classification models and perform cross validation, variable selection and regularization

**Matplotlib/ Seaborn** – Visualizing data exploration, modeling and results

**Python 3.8.5** – to run all of the above

**nltk** - Natural language toolkit, to work with human language data.

# RESULTS

---

- Five different topics were discovered –
  - Bad Reviews For Paid Subscriptions -- 69%
  - Profile Match -- 22.7 %
  - Good Reviews About The App -- 3.4%
  - Good Reviews About Other Users -- 2.3%
  - Easy To Use -- 2.2%



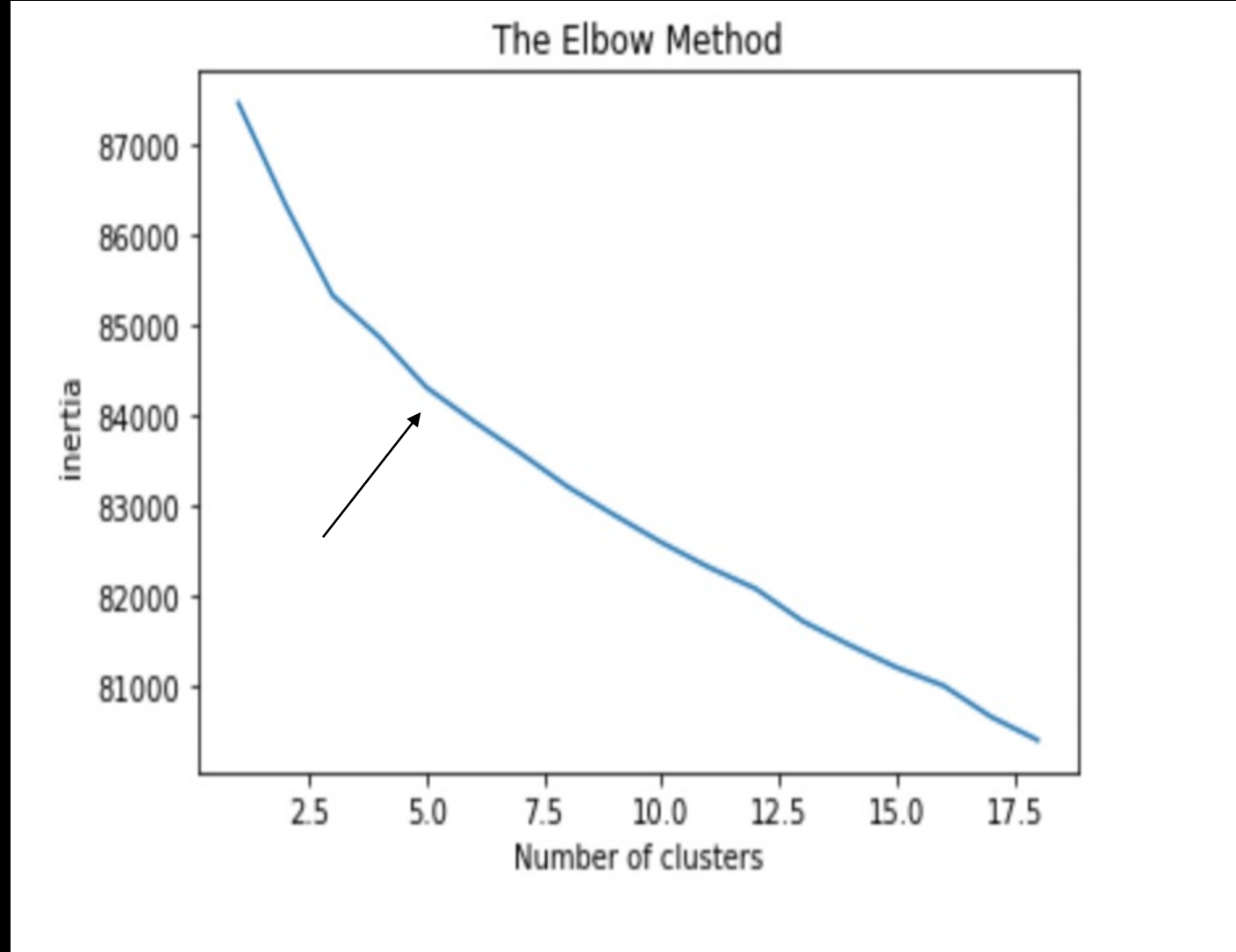
# METHODOLOGY

---

- WordNetLemmatizer was used.
- Custom stop words were added to the stopwords set.
- Transform the text data into a matrix, which represents the weights of each word in the corpus.
- Algorithms Used -
  - K-Means Algorithm
  - NMF – Non-Negative Matrix Factorization
  - Naïve Bayes
- Use the elbow method to decide the number of clusters.

# ELBOW METHOD –

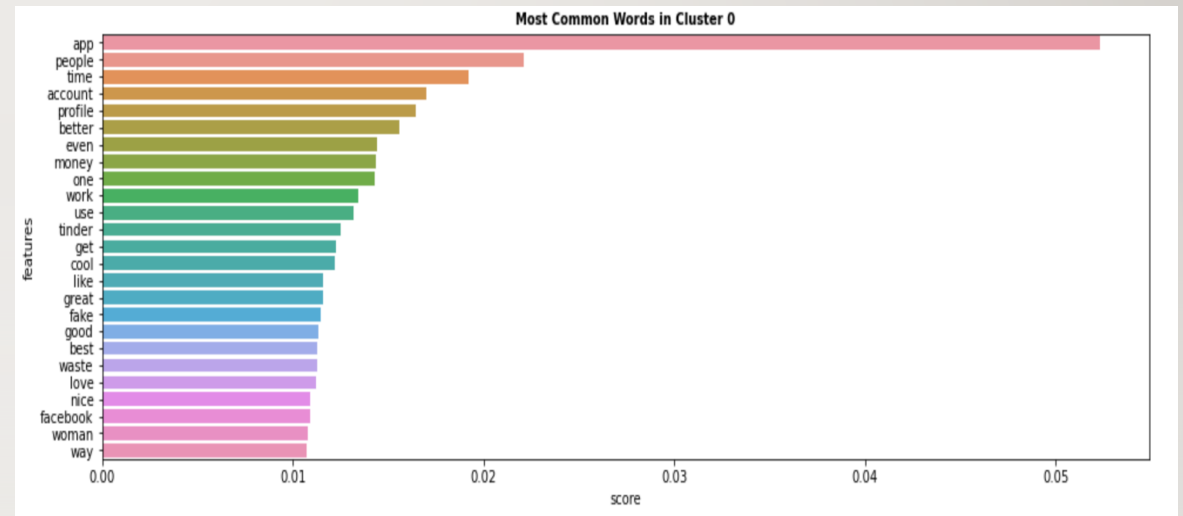
NUMBER OF CLUSTERS VS INERTIA



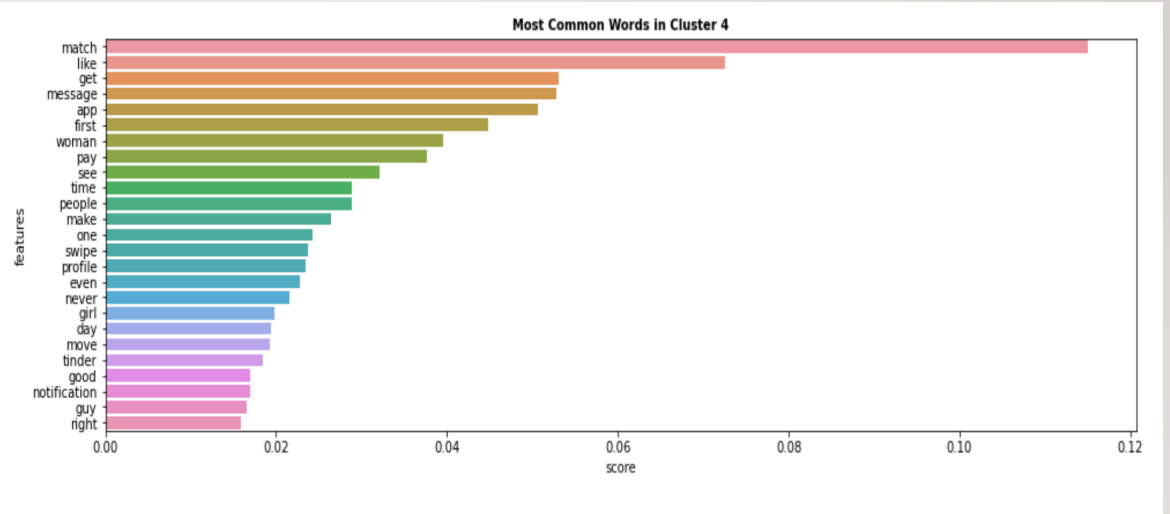


# VISUALIZATIONS

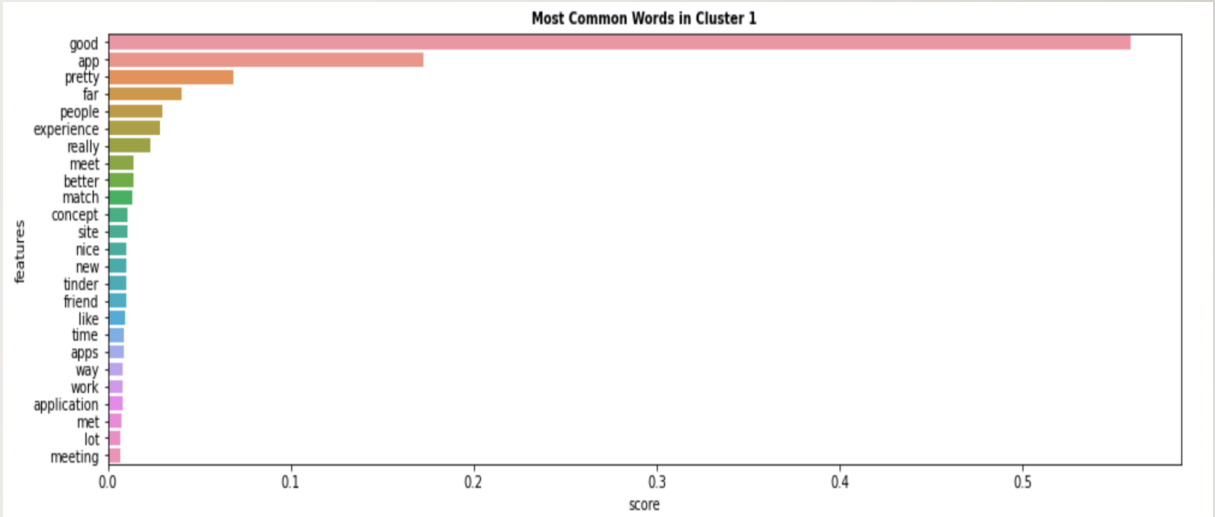
GRAPHS AND WORDCLOUD

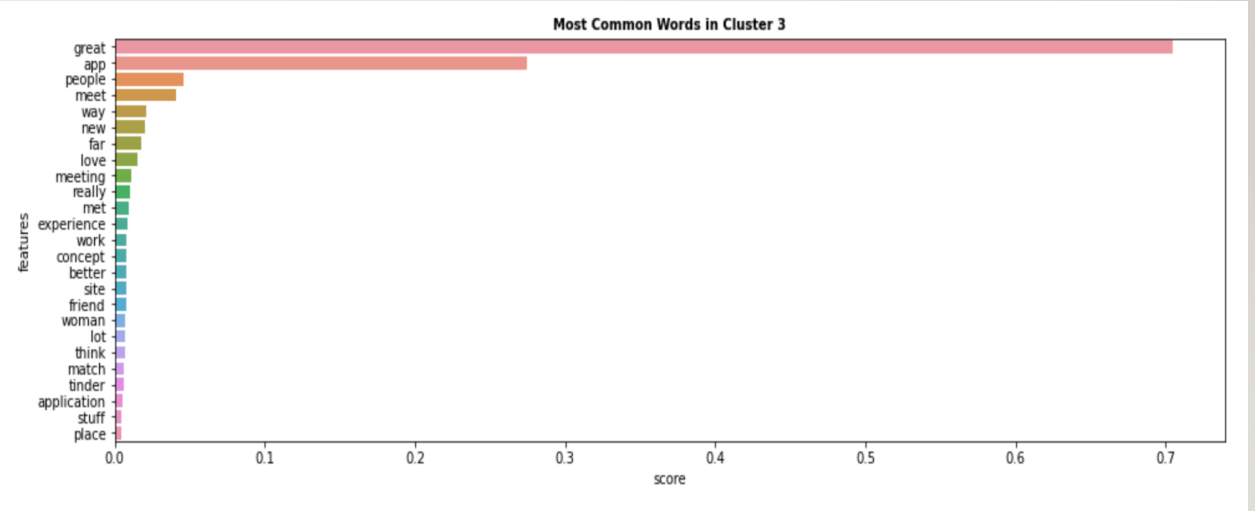
[illegible]

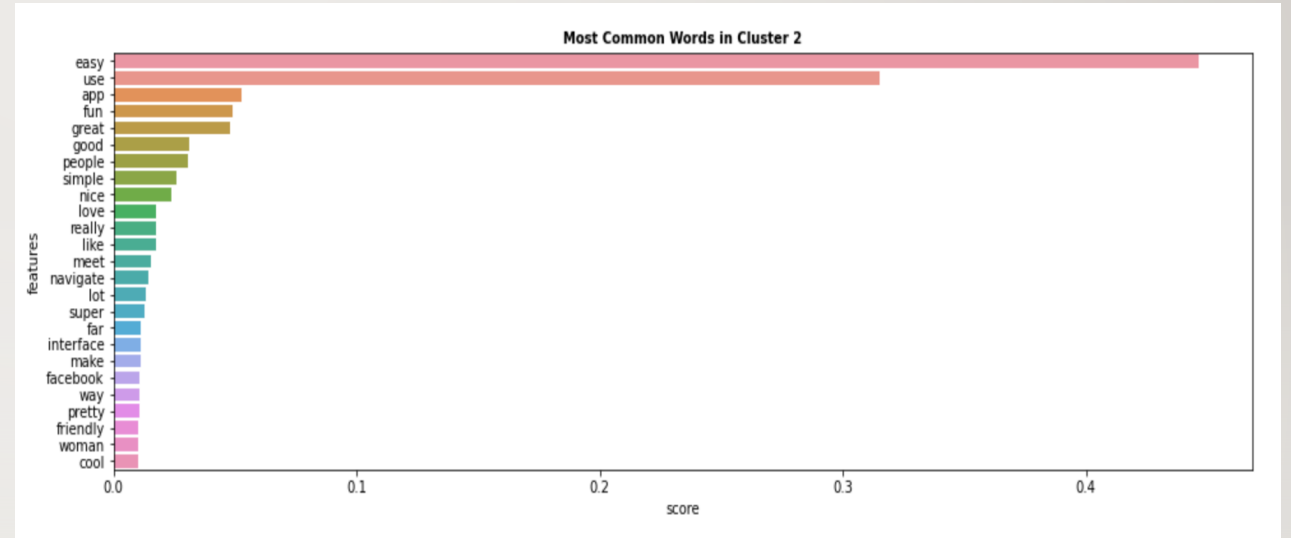
## CLUSTER 0 – BAD REVIEWS FOR SUBSCRIPTIONS

[illegible]



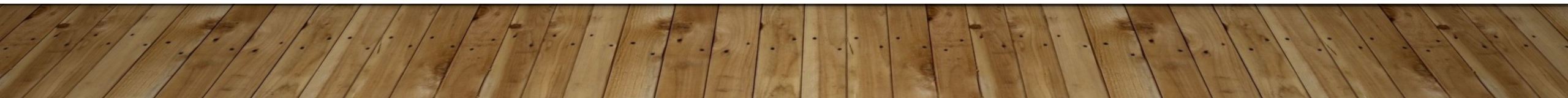
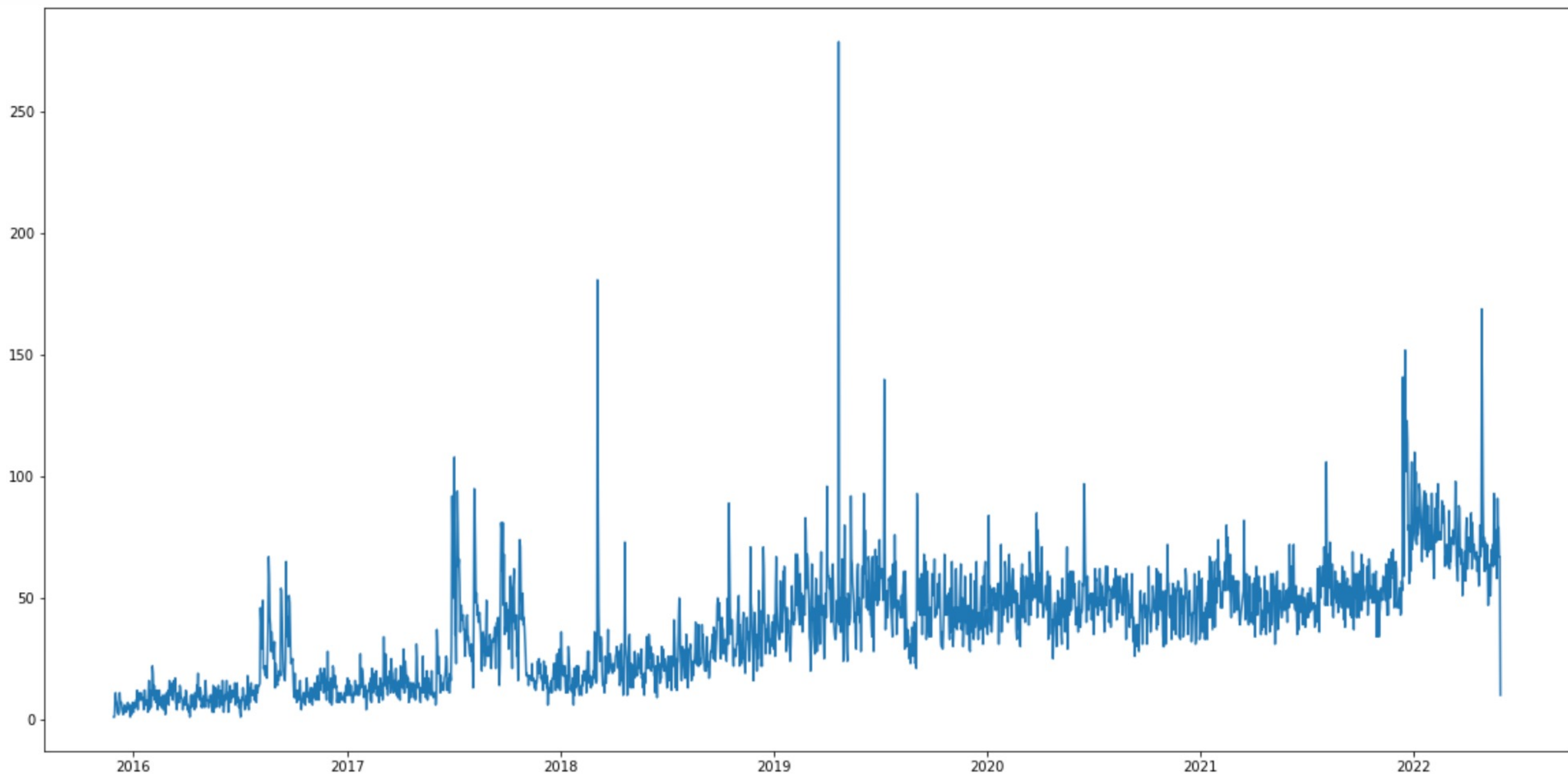
[illegible]





## CLUSTER 2 – EASY TO USE





# NAÏVE BAYES MODELS

---

- Gaussian NB Score – 0.503
- Multinomial NB Score – 0.855

# VADER SENTIMENT

---

- Using SentimentIntensityAnalyzer method
- Best Review –
  - Score - 0.9062
  - Review - good app thanks dear woman you are beautiful
- Worst Review –
  - Score - - 0.7997
  - Review - after being a premium user i m not able to login to my id pathetic experience no customer support is provided

# FUTURE WORK

- Translate the remaining 20 K non-English entries to English language
- Find the best Coherence Score for NMF model
- Retrain the models
- Deploy the model