Data Analyst Report

**Dating Apps data in the India**

Data Observation and Visualization

horizontal line

# Placeholder image

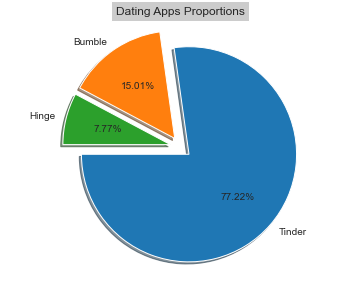
# Introduction

The data is collected from India by Sidharth Kirplani (<https://www.kaggle.com/datasets/sidharthkriplani/datingappreviews>) . The data is from apps like Tinder, Hinge and Bumble. Data like text reviews, ratings, name and more. In this report, I will be showing the general observation and some visualization report regarding the data. Lastly, I will discuss some inspiration I have got while analyzing the data.

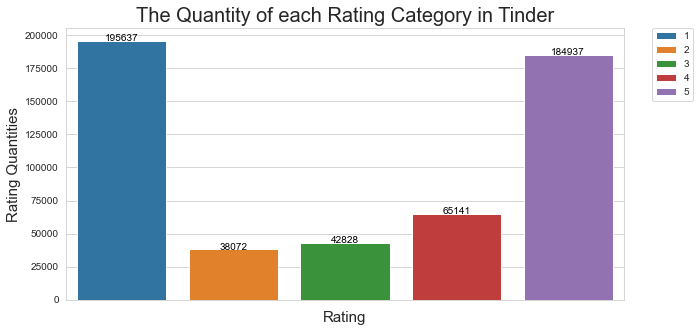
# Data Explanation

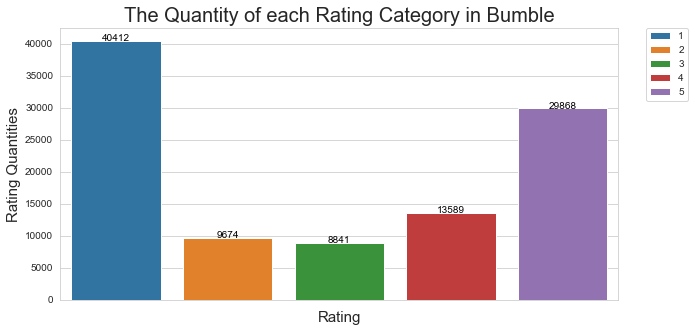
1. Name  
   The name of the user who reviewed on the google app store
2. Review  
   Th text review that the respective user input
3. Rating  
   The rating submission from the respective user
4. Thumbs Up  
   The ‘likes’ on the review, received by other user who agreed on the respective review
5. Data and Time  
   The date and time when the user input the data
6. App  
   The name of the app

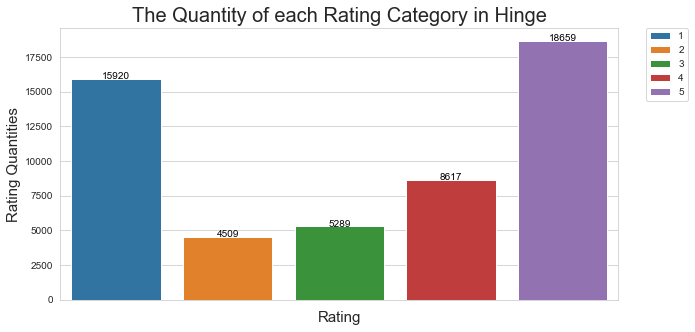
# Data Observation



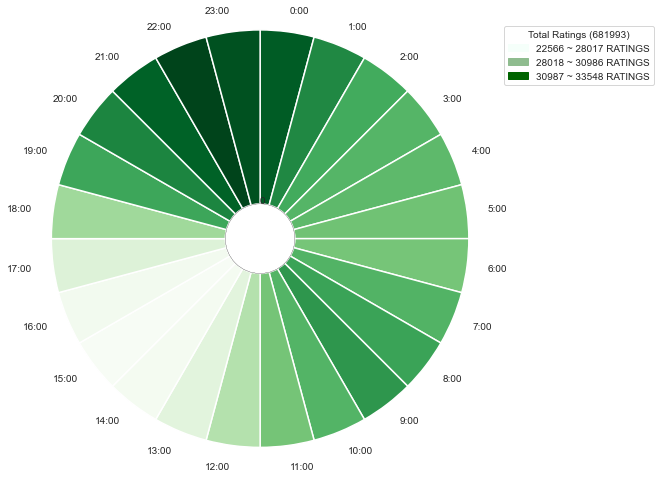
On 70,000 sets of data, these are the proportions on each app (Tinder, Hinge and Bumble). As we can see, Tinder has the most data than the other two, containing 77% among the others. While Hinge has the smallest proportions which is around 8% among the others. As we can have an assumption that Tinder is a popular option for finding dating partners online in India. Unfortunately, we do not have the data on the proportions of gender in order to have a better observation.





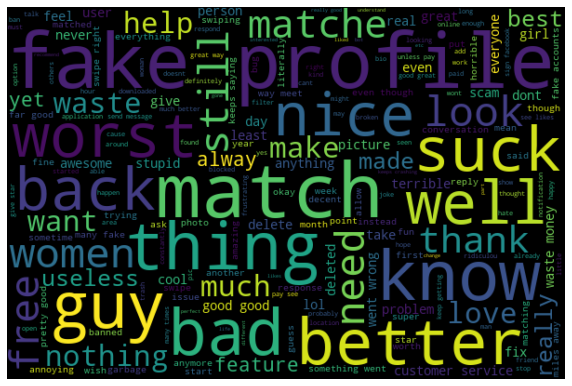


Based on the ratings proportion on these 3 applications (Tinder, Bumble and Hinge). We can see a pattern that vast users tend to rate 1 star or 5 star as their review on the respective application.

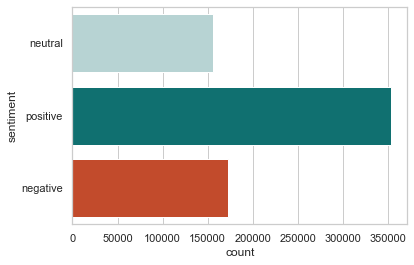


The graph above will show the rating time (on the whole day in international time format basis) on all of the users. The darker the color is, the heavier traffic will be in the review application section. The color tends to become darker between 21:00 (9pm) to 00:00 (12am), meaning that most of the users will comment and give review about the respective application between 21:00 (9pm) to 00:00 (12am).

# Word Analyst (python NLTK)



These are the words that often appear in all of the reviews from the 3 applications (Tinder, Bumble and Hinge). I have removed most of the neutral words (like ‘men’, ‘lady’, ‘use’, ‘by’ etc). This is not the perfect word analyst from the other reports that you will see, in order to improve the quality and insight requires more time and effort. (which I don’t want to use most of my time improving the analyst, I am just showing the basic concept of what word analyst is.)



The graph above will be the sentiment of all of the reviews. Have 3 categories neutral, positive and negative. For accuracy we can focus on positive and negative (because some of the neutral words I have removed). As we can see, positive words tend to appear more than negative words in all of the reviews from India.

# Summary

1. Tinder is a better well-known dating application in India, among Hinge and Bumble
2. Reviews tend to be very extreme, majority of the reviews either is rated as 1 or 5, while on the middles are the minority of the reviewers
3. It is common that people will use their electronic gadgets for leisure after work hours, most of the reviews were reviewed during the night time
4. The reviews are kind of extreme too, most of them are using very vulgar language complaining the flaws of the applications, or very thankful of the application on what it provided to them
5. There are more positive words than negative words on all of the reviews collected

# What could I do more on this topic

1. Honestly speaking, I lost interest in this topic so just decided to do a general analyst on what I observe in this data, but there are more meets the eye
2. The reviews can be counted and analyze the words counts and the rating given, to see the correlation between them
3. The NLTK can be enhanced by removing more neutral words and focusing on the positive and negative words
4. We can collect data similar to this topic from other countries and make a comparison on this
5. We can focus on words like ‘match’ and ‘no matches’ on the reviews, then make a general prediction (regarding your personal status) on the matching rate on each dating application (but there are a better dataset to make this model prediction, so no point going deep in this dataset)