



STUDIO SHODWE

Social Buzz Data Analysis

BY ERLANDO REGITA



About Company

Social Buzz was founded by two former engineers from a large social media conglomerate, one from London and the other from San Francisco . Over the past 5 years, Social Buzz has reached over 500 million active users each month. They have scaled quicker than anticipated and need the help of an advisory firm to oversee their scaling process effectively.



Clients Request

To start our engagement with Social Buzz, we are running a 3 month initial project in order to prove to them that we are the best firm to work with. They are expecting the following:

- An audit of their big data practice
- Recommendations for a successful IPO
- An analysis of their content categories that highlights the top 5 categories with the largest aggregate popularity



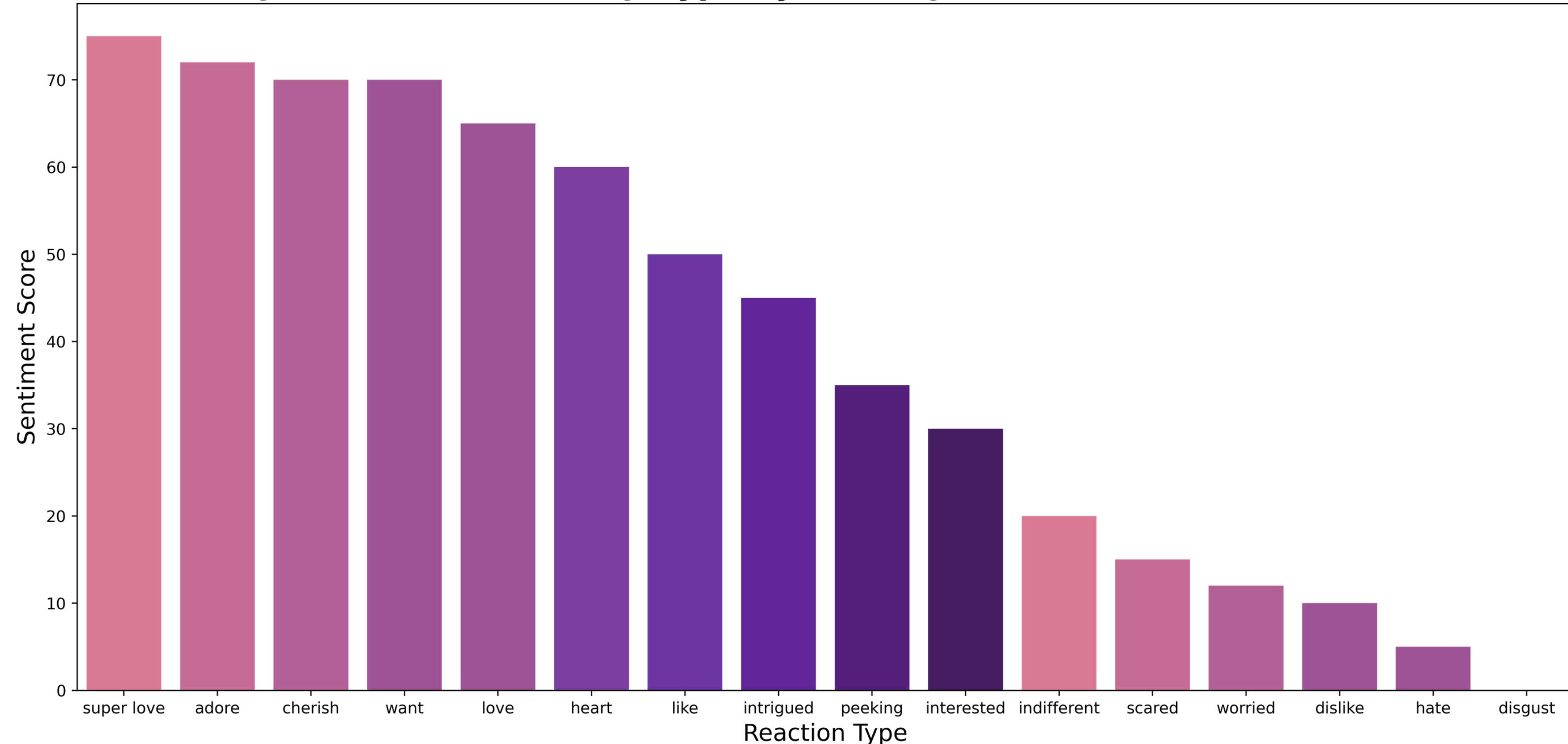
STUDIO SHODWE

Data Exploration



Reaction Type by per Sentiment Score

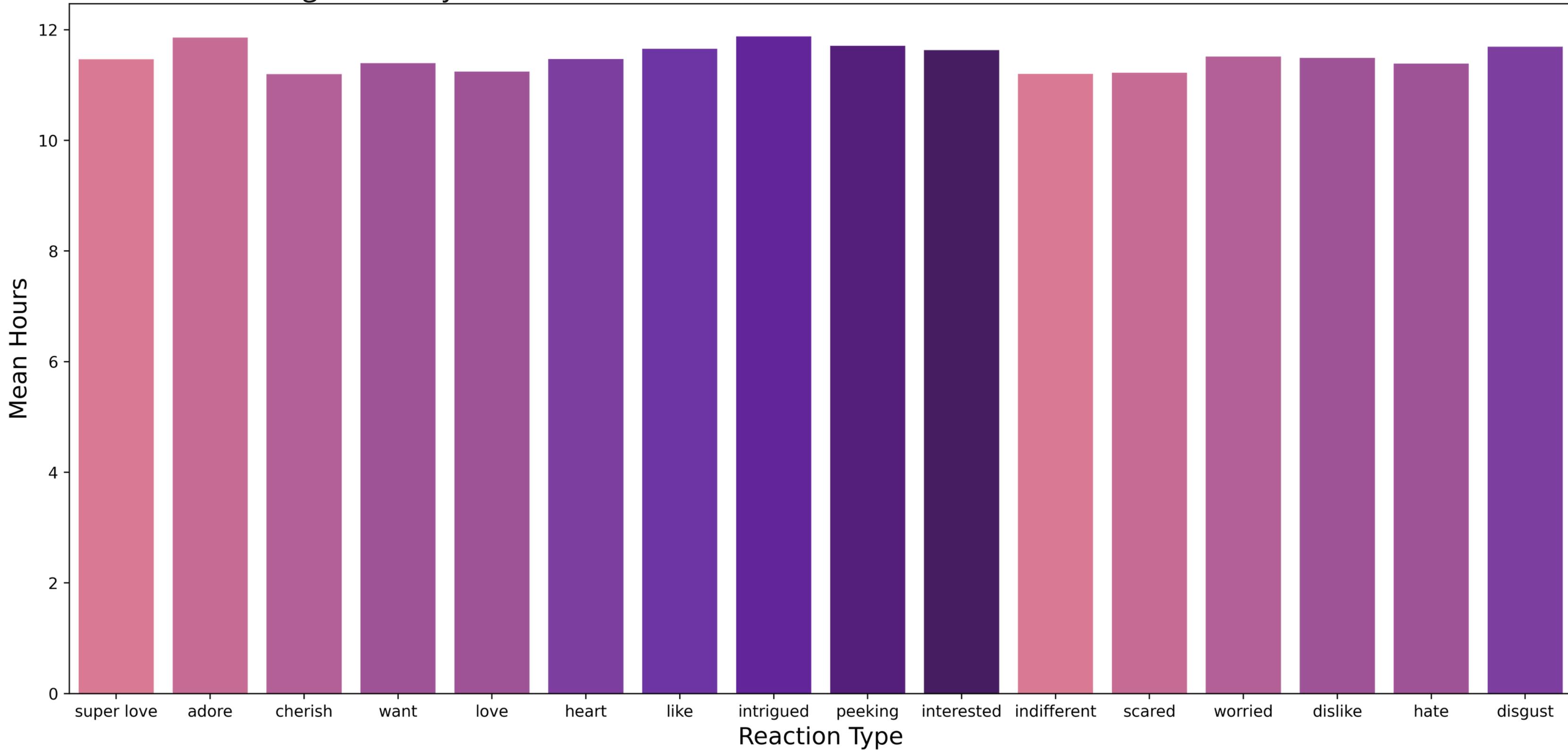
reaction type distribution looks normal if its measured by sentiment score. The left one is the positive sentiment (high sentiment score). The right bar chart is the negative sentiment





Reaction Type per Mean Hours

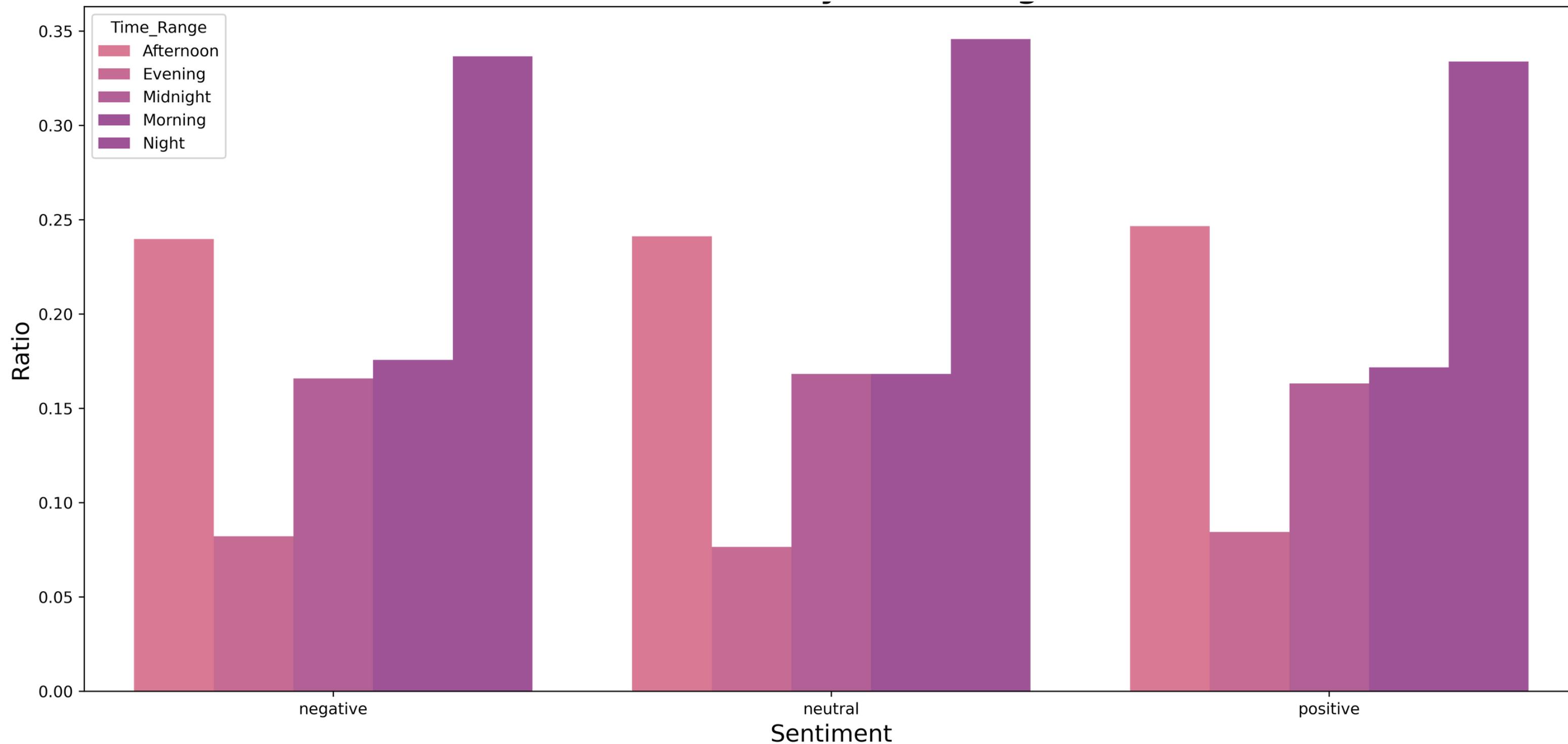
Reaction type by mean hours, this data seems computer generated data. It seems impossible because almost all reaction type have similar/almost the same mean hours. Our deduction this dataset is generated by computer, not a real dataset generated from user





Sentiment Type per Time Range

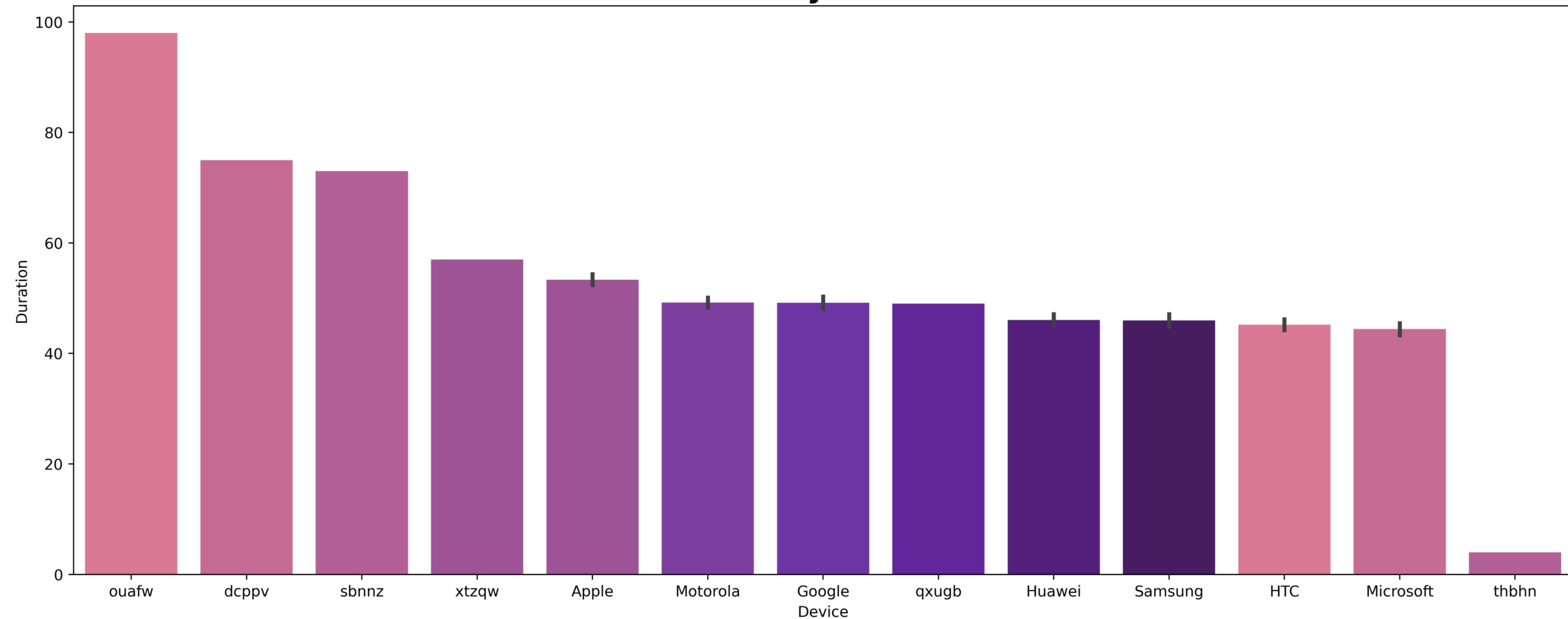
Again, almost all time range between sentiment type have the similar ratio. This is strengthen our argument in previous slide that this dataset generated from computer, not user based dataset. But in the next exploration, we will provide some insights based on this dataset (assumption this dataset is not validated as computer generated data)





Device Type per Duration Used

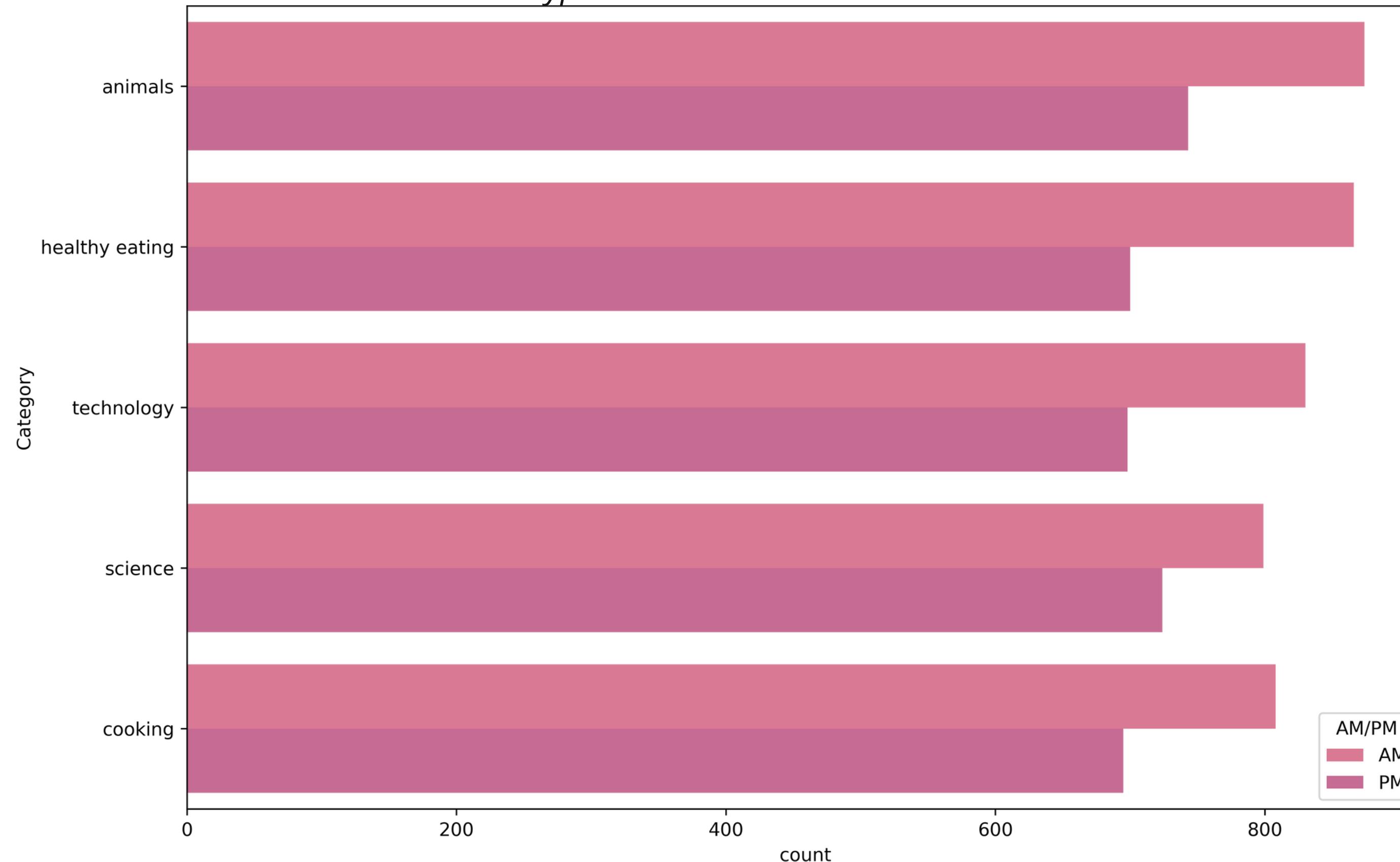
There's some other unfamiliar device type have high duration used, instead devide from lets say apple, motorola, and google device have average duration used.





Content Type by Time AM/PM

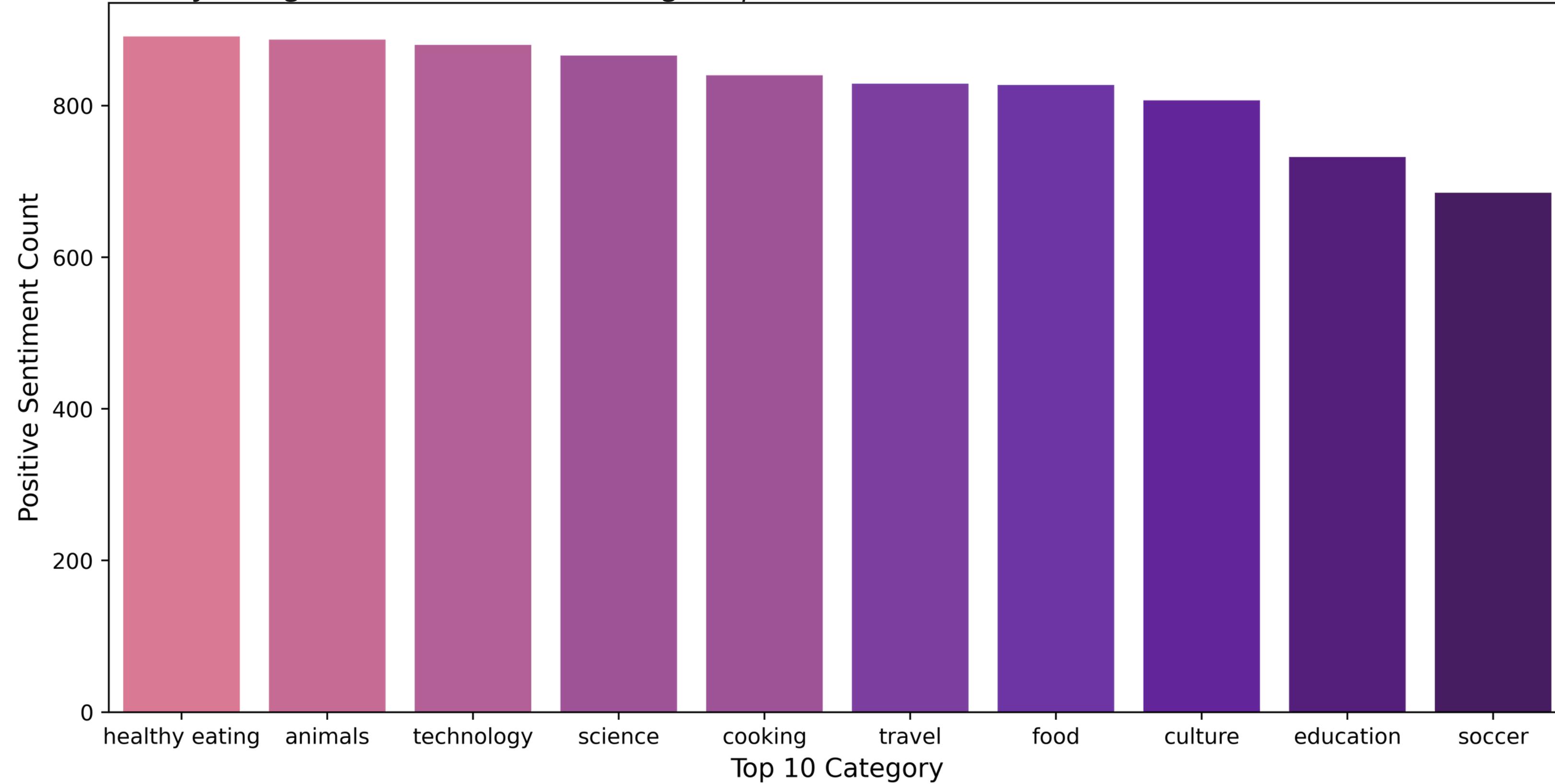
Most of content is watched by user in the morning (AM). Animals and healthy eating is most wathed content type





Top 10 Category with Highest Average Positive Sentiment Score

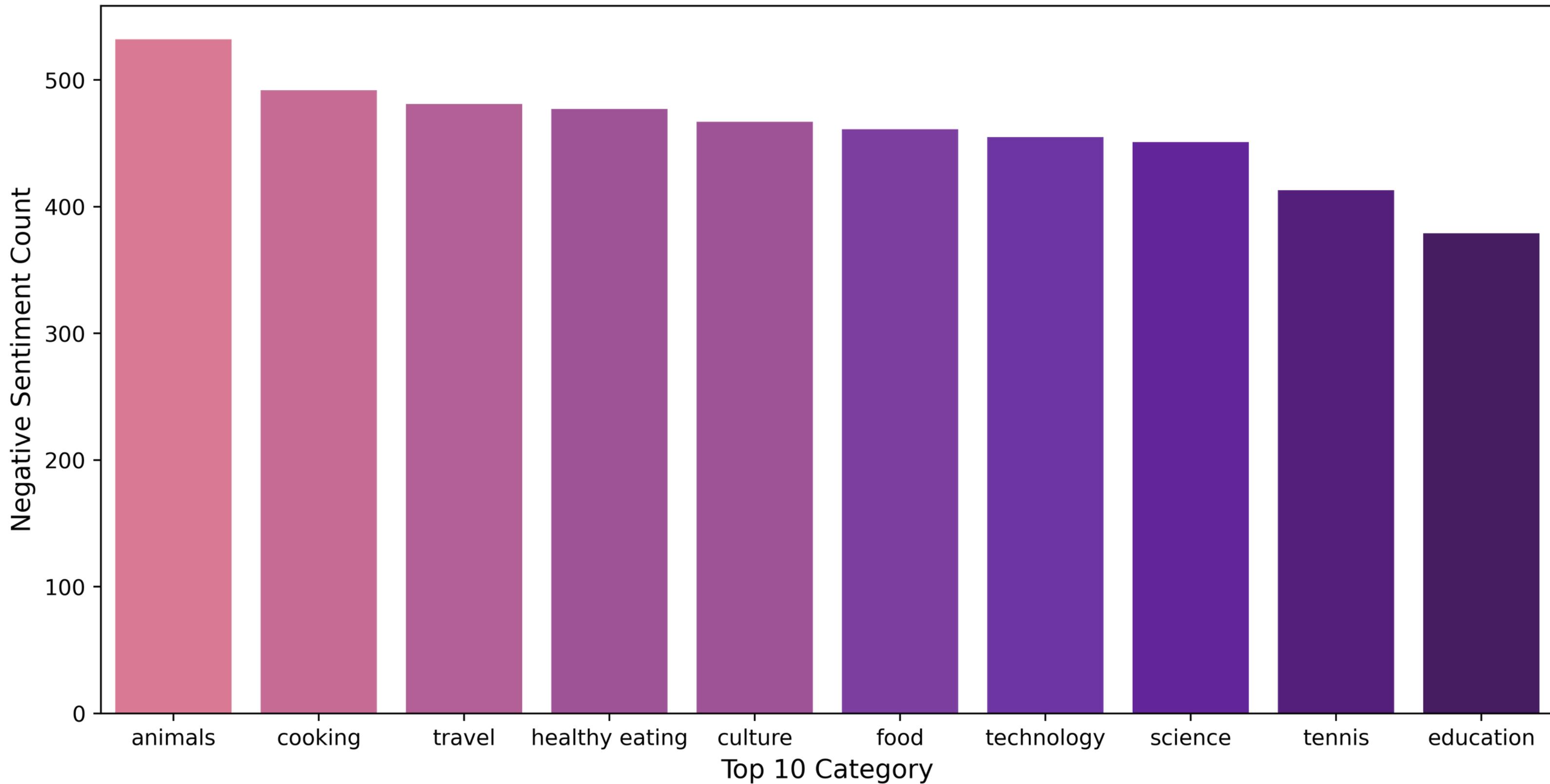
Healthy eating and animals are most 2 highest positive sentiment score





Top 10 Category with Highest Average Negative Sentiment Score

animals and cooking are most 2 highest negative sentiment score





Insight Summary



Computer Generated Data

There are some columns suspected as computer generated data. Because some of columns are constant and don't represent a pattern

Most Watched Content Type

We just can provide insights about most watched content type, highest positive and negative sentiment score, etc



Recommendation

We need additional data, during analysis we realized that there's something wrong on the data that client's provided. Client need to evaluate the data team about how they generate the data. Our recommendation for this case is data team and data environment evaluation.





STUDIO SHODWE

Contact Us



0821-1000-4094



erlando_regita@sbm-itb.ac.id



www.brian-insights.site



linkedin.com/in/erlandoregita/



github.com/erlndofebri





STUDIO SHODWE

Thank You

SLIDE PRESENTATIONS DESIGN