

# Social media market segmentation

Dhwanit

4/19/2020

## Problem statement

Twitter tweets from the followers of a consumer brand were analyzed. Every tweet has been classified as belonging to a category like sports, cooking, beauty, fitness etc. by annotators.

These tweets can help in market segmentation i.e., to classify social market into segments to help understand the customer base better in terms of their interests, preferences, social media activity etc. This can help in targeted advertisement campaigns, better reach to customers, tweak promotional offers according to customers etc., in effect helping the business to grow.

Our task here to is to identify these segments based on intuition and sound data analysis and interpret them in a coherent manner.

## Steps

The broad outline of the approach we took to carry out this task as follows:

### 1. Exploratory data analysis

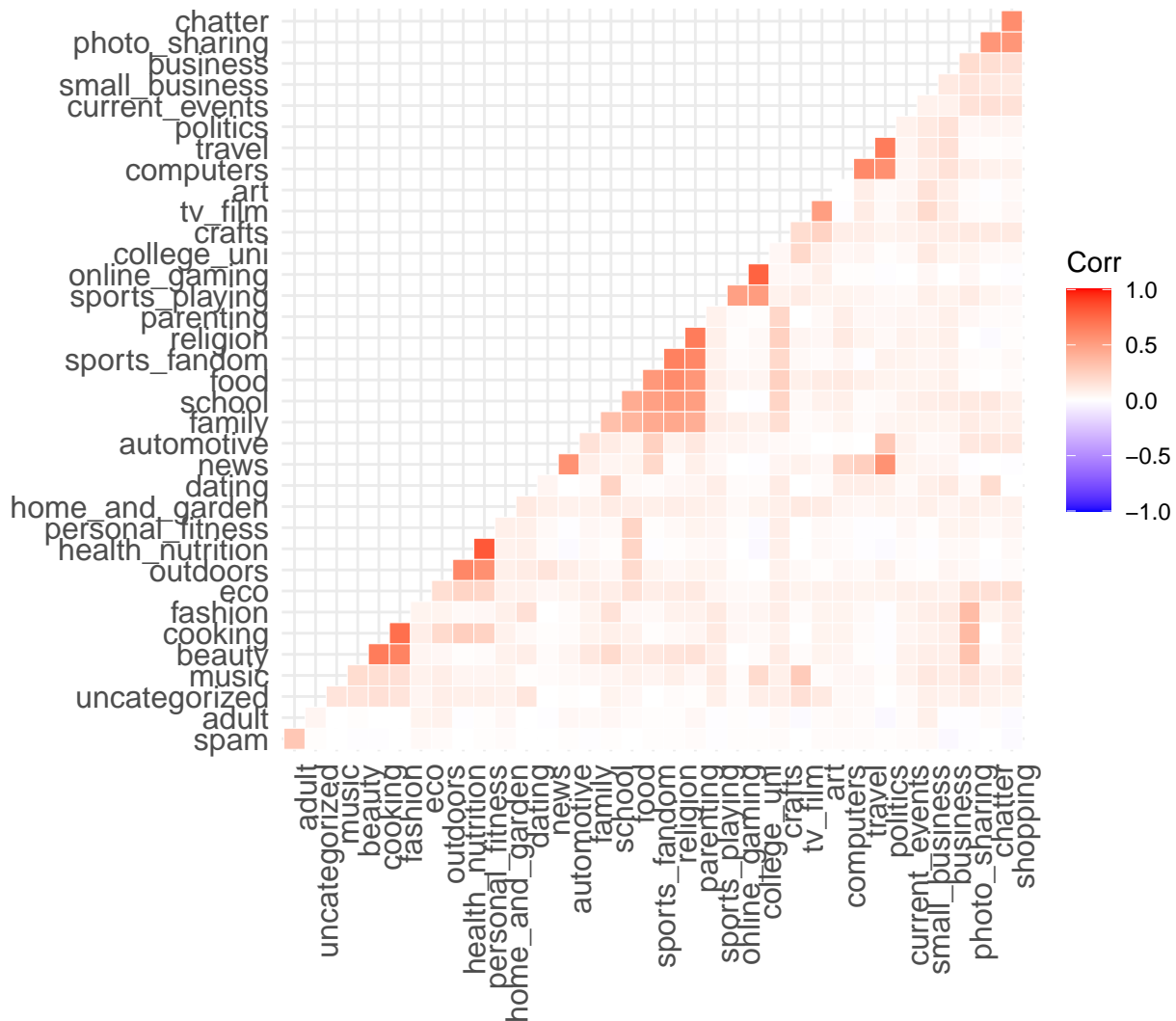
We find out that the data has 36 features available for each user. These 36 features specify the number of tweets in each category (like online\_gaming, politics, cooking etc.) by that user. It is intuitive to first check for correlation between these features as that can help us in finding out correlated features. Since correlation is not sensitive to scaling, we just run plot correlation for the raw data.

```
sm_data <- read.csv("social_marketing.csv")
#summary(sm_data)
#head(sm_data, 10)

sm_feat_raw <- sm_data[,2:length(sm_data)]

# Creating a correlation plot

cormat <- round(cor(sm_feat_raw), 2)
#head(cormat[, 1:6])
ggcorrplot(cormat, hc.order = TRUE,
            type = "lower",
            outline.color = "white") + theme(axis.text.x = element_text(angle=90, hjust=1))
```



Correlation plot reveals some interesting and intuitive correlations between features. Some very strong +ve correlations can be seen, like between health nutrition and personal fitness, cooking and fashion, online\_gaming and college\_uni. These are intuitive like people who are interested in health\_nutrition also are likely to tweet about personal fitness. A rather peculiar one is between sports\_fandom and religion which doesn't seem so intuitive. (Maybe if you are a fan of God, you are likely to be a fan of sports team :) ).

## 2. Segmentation using clustering approach

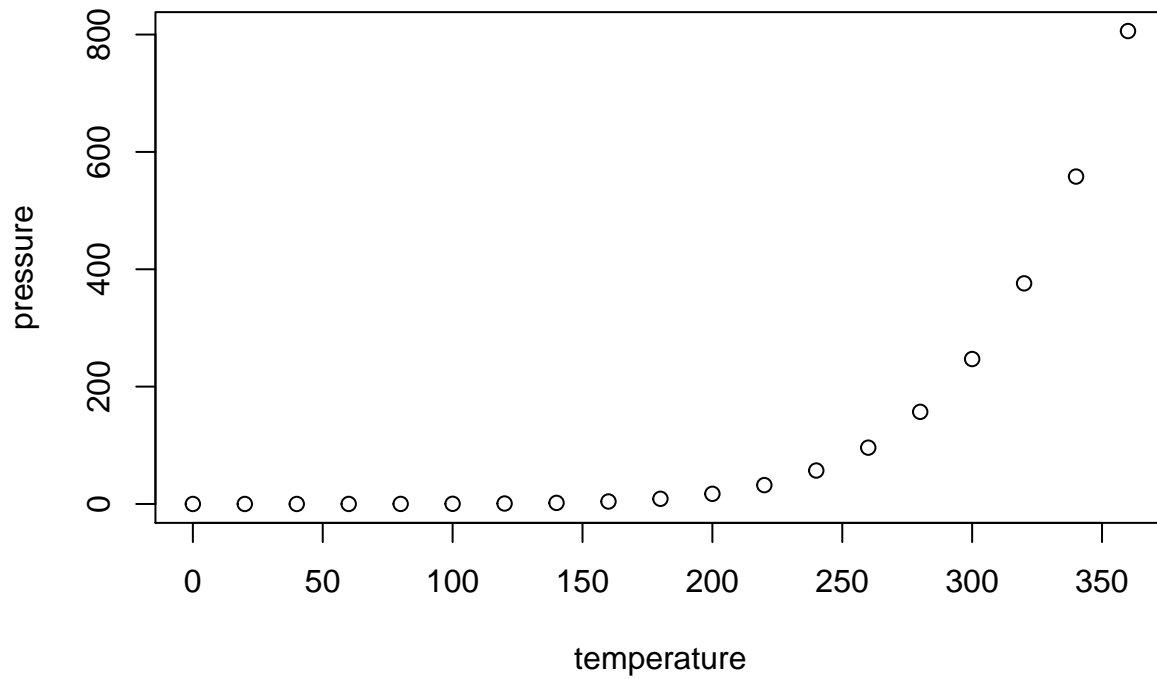
In order to see if these features and correlation can help to identify some interesting groups of users with shared features, we decide to try the clustering approach to help identify some number of clusters. We decide to go for the KMeans++ clustering approach. The first step in this clustering is to scale the data to zero mean and unit variance of the columns. This will even out the intrinsic variations in the features data and help us in clustering.

```
sm_user <- sm_data[, 1]

# Scaling and centering data
sm_sca = scale(sm_feat_raw, center=TRUE, scale=TRUE)
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.