Assignment 8 Functions

16 Dezember 2019

## Dataset description

Take the Facebook dataset. Column names are well defined so that everyone can interpret easily. userid age dob\_day dob\_year dob\_month gender tenure friend\_count friendships\_initiated likes likes\_received mobile\_likes mobile\_likes\_received www\_likes www\_likes\_received

## Task

Inspect the data set. Follow the recommended approach discussed in class. When examining the dataset, briefly comment on the steps/code/output as if you were an analyst.  
1. How would you describe the structure of the dataset?  
2. How many missing values are in the data set?  
3. Which variables are the missing values concentrated in? (Hint: apply or colSums function may be helpful)  
Make a decision on how you treat missing values. Reason your choice.  
4. Examine basic descriptives of the dataset using **summary()** or **describe()** function.

5. How can you characterize the sample in terms of gender?  
a. Explore the distribution of the friend count for both female and male users. For what gender, the average number of likes is higher?   
b. Which gender has likes by using web platform the most? Base your inference on the average number of likes.   
c. Which gender has likes by using the mobile application the most? Base your inference on the average number of likes.   
d. On average, who initiated more friendships in our sample: men or women?

1. Check if the Facebook mobile app is really beneficial for the company. Compare the type of device used for sending and receiving likes. What share of users receives/sends likes on a mobile device? What share of users interacts via the web?
2. Explore the number of friends by gender and age. What inferences can you make?
3. What users have higher friend count: people in older join (tenure) or those who joined later? Interpret the variable “tenure” as the number of days a user is registered on Facebook. Assume that those who use Facebook at least 500 days belong to an “older join” category and those who use Facebook less than 500 days belong to the “later join” category.

Please work in RMarkdown and submit the .pdf report as a result of your work. You can use any functions and packages for your analysis. Consider visualization tools as well to support you r analysis.