Title: Facebook Social Media Analysis
By Yi-Hsuan Hsu, Cindy Tai, Navado Wray, and Judy Zuo

Overview/Summary:

Social media platforms are widely used in order to spread information and used for entertainment. One of the world's largest platforms, Facebook, has billions of accounts with a diverse user base. Statista.com showed that in February 2019, Facebook dominated over 70 percent of online users in each age group in the US except seniors over 65 which is around 40 percent. The number of monthly active users has reached 2500 millions by the end of 2019. In 2016, the percentage of people who had over 200 Facebook "friends" was close to 60 percent (statista.com). However, this information only concludes that this platform is used by many people without much consideration for what makes the users different from each other.

This dataset describes user accounts on Facebook through the number of likes (both given and received through different forms of access), date of birth, gender, number of friends, and account age in days. We hope to discover insights within the dataset that will lead to a better understanding of our social media platform and how it is being utilized. Specifically, we wish to examine if specific user characteristics such as age or gender correlate to individual account activity. Based on these discoveries, we can determine areas of focus on which the platform can be improved, creating a better browsing experience for all current and future users.

Problem Definition:

1. Does the ratio of friendships initiated to total friends differ based on user gender?

Gender differences is what this question is focusing on. The group wants to know if there's a correlation between what gender a user is and the likelihood of them initiating a friendship with another user. By looking at the difference between a male and a female's ratio of friendships initiated to total friends, a conclusion can be drawn about their usage habits. Those who have a higher ratio may use Facebook for forming new friendships while those who have a lower ratio may prefer to use Facebook to keep in touch with their current friends. The group is hypothesizing that males will have a higher ratio of friendship initiation than females. Testing this hypothesis can lead to further insights regarding each gender and the way they form relationships/friendships with others.

2. Does age correlate to whether users prefer mobile vs web experience?

This question allowed for a better understanding of how age affects social media usage, particularly the platform on which users browse. People who give out more "likes" on mobile (versus the web) most likely prefer using the app over the desktop version. The same can be said for those who give out more likes on the web, that they most likely prefer the desktop version over the app. Answering this question can lead to optimization ideas for the app as well as web development. If a specific age group prefers one method of accessing social media, then that method can be reworked to better suit the needs of that age group. For example, if the older generation prefers desktop, then the web version can have larger text to account for possible poor eyesight, a problem shared by many older users.

3. Does account age correlate to *"success"* on the social media platform?

First off, the group would like to define *"success"* as *"the ratio of likes and friendships received versus the ratio of likes and friendships given."* The group would like to determine whether or not experience with the platform over time correlates to an increased *"success"* metric. As in, do more experienced users actively achieve more *"success"* compared to newer users? Answering this question may help with spotting, for example, voting manipulation through bot networks; if a new account is gaining more than the average likes for its account age, it is possible that the account is using bots to artificially inflate their like count. This can also be used to isolate "power users"; if an old account is getting a disproportionately large amount of likes, they could be abusing the algorithms of the platform to get their content to as many people as possible, which can lead to questions about misinformation, fake news, and so on.

Period of Analysis:

2020-02-10 to 2020-04-24

Vendor Contact:

Dataset received from:
Professor Connor Watson
cww5@njit.edu

Dataset comes from:
Facebook
There is currently no way to get in contact with support outside of messaging them on their page:
https://www.facebook.com/facebookapp

Facebook HQ
(650) 960-1300
1601 Willow Rd, Menlo Park, California 94025

Vendor Documentation:

None provided. We will be creating the documentation for this data set as part of the fulfillments of this analysis.

Sample Data:

Link to dataset:
https://drive.google.com/open?id=1PPrGOfUFO3zOboAu55F_MEO1rLDBBN5T