COMSCORE DATA ANALYSIS

In our first look, we observe some information missing in the 'minutes' column and 'gender_id' column. Let's see how we deal with the missing values.

Missing-value Treatment

There are 3% null values in minutes column for few apps for a device_id. This pattern gives an impression that either user might have downloaded and uninstalled the app which created the row in the database but usage was not found or user never used the app in that entire month. In the worst case, there could have been a problem with the data collected from the app. In the worst-case scenario, there are many ways we can impute missing values with some meaningful statistic. Since there are only 3% missing values and the time considered is for an entire month for comparatively well-known apps, I do not want to create any bias by assuming worst case scenario and imputing values in place of missing information.

Also, in demographic information around 3% of gender information is missing for some device_id's. device_id is an identifier column and we cannot assume anything upon identifier column. Hence, I'm removing missing value rows in this table.

Outlier Treatment

We observed there are some outliers in the minute's column. I noted down these outliers and perform further analysis accordingly.

Insights

Which App is most installed?

Of all the apps, Google and Facebook apps are majorly used applications.

Which app is most used or has more customer engagement?

Facebook products still stand at the top. As of now, it looks like most of the user's installed Facebook apps as well as spend most of their time on Facebook products. However, since we observed a few outliers above I am introducing another metric to evaluate the most used app.

Metric Design

Average App usage per device: (Total Minutes per app) / (Number of devices per app)

On average users spend the most time on Facebook apps followed by YouTube, Instagram, Tumblr & Netflix. An interesting observation is one of the most used apps Tumblr did not show up in previous plots.

And the most installed app 'Google Play' is not the most used on an average.

Gender-wise analysis

- Our data set has 62% of female users and 38% of male users.
- So, now let's analyze how the app usage varies with gender.
- Both the genders mostly use Facebook apps & YouTube.
- Besides that, we can see that male users prefer using Instagram, Tumbler & Snapchat.
- Female users prefer using Netflix, Instagram, Tumblr, & Kindle.
- Depending on the target audience the above plots helps us to drive the content through different platforms for better reach.

Analyzing Additional-data

- In additional data, users considered seems to exhibit different patterns compared to the initial set of users.
- Firstly, the average-usage is slightly lower. Additional data users seem to be using Tumblr more than YouTube & Instagram which is opposite with initial set users.

Publisher-wise Analysis

- **Google Sites:** Though YouTube seems to be slightly less installed app among all the Google applications, YouTube has the highest average-usage rate.
- **Facebook:** All the three Facebook apps stands in the top 4 highest average-user apps with the Facebook app being first followed by Facebook Messenger & Instagram.
- **Amazon Sites:** Kindle has better Average-user rate among all the apps from Amazon.
- **Oath:** Tumblr stands top with the highest Average-user rate among all the apps from Oath.
- **Walmart:** All the Walmart apps stands in the top 4 least Average-user ratings.
- **Microsoft Sites:** Outlook has the highest average-user rating as well as the most installed app of all the Microsoft apps followed by GroupMe and Word.