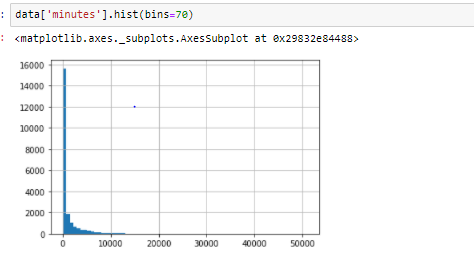
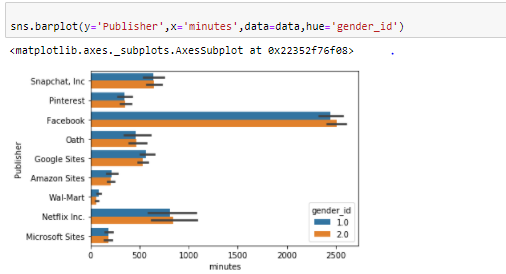
App Usage Analysis

* Initially text file is convert to dataframe and merging it with other files through inner join and dropped duplicates
* Minutes column is positively skewed.

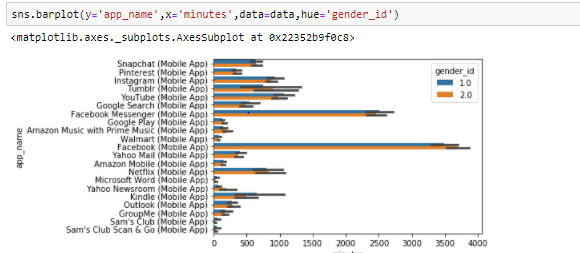


* We have applied barplot on minutes, publisher. From plot, we can clearly state that there is equal distribution of usage w.r.t gender\_id.



Above performed plot is made on mean of minutes.

* We have applied barplot on minutes, apps. From plot, we can clearly state that there is equal distribution of usage w.r.t gender\_id.



Above performed plot is made on mean of minutes.

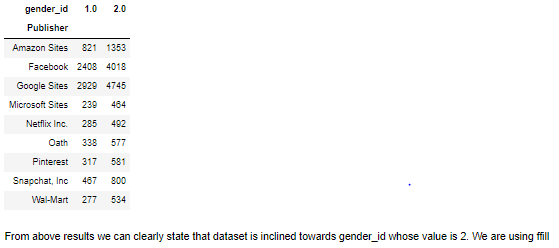
* From boxplot, we can see there are outliers whose duration of app usage in extremely high
* From below boxplot, we can see there are outliers whose duration of app usage from particular publisher varies extremely high

|  |  |
| --- | --- |
|  |  |

We can clearly see there are high number of outliers in Facebook, Google sites than remaining publishers.

These outliers will impact on mean so we cannot replace entire dataset mean. So we are replacing nan values of minutes with mean of individual publisher minutes.

* From below crosstab, we can clearly see in each of publisher we have dominance of users whose gender\_id is 2.0.



|  |  |  |
| --- | --- | --- |
| Rankings of individual Apps based on total users | Rankings of individual Apps based average minutes per user | Rankings of the publishers as per total devices across applications |
|  |  |  |

* From above table, we can see an interesting observation though google apps publisher has highest devices registered and google apps also have highest users for individual apps still facebook and facebook messenger is in top in terms of engaging users for longer duration.