**Mobile atlas (with price and released time with monthly accuracy) + Device clustering + Segmentation rules & Facebook audience creation:**

Versions: 20170219 RS: Initial version

1. Impala: Create a summarize table of mobile atlas data available from Tapad’s apollo.dtac\_vertical\_dataset table. Here, I choose to extract all the models appeared in Jan 2017 from event\_source = idsync, reach = distinct(tapad\_id) (0.TapadMobileAtlas.sql)



OUTPUT: devicelist\_idsync\_jan17.csv

1. R:

* 1.scrape\_gsmarena\_urls

Get the urls of phones in gsmarena.com for mobile atlas table creation

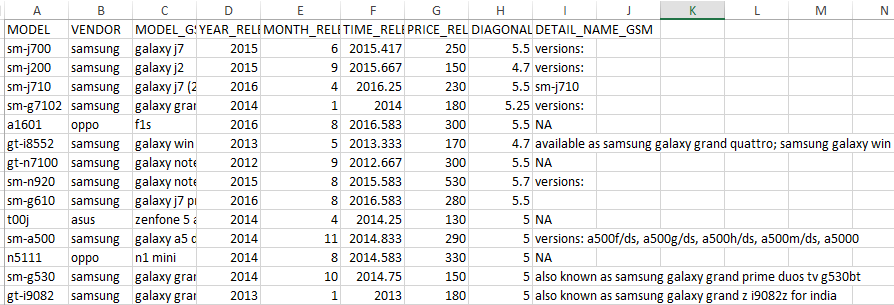
* 2.create\_mobileatlas\_table

Scrape the phone information on gsmarena.com to create a mobile atlas table. Here we scrape model, marketing name, vendor, released year, screen size, released month and released price. The release month and release price are very important variables for clustering which are not yet available in Tapad data (Petter is working out to have it ready)

* 3.match\_useragent\_mobileatlas

Match the mobile info scraped from 2 to model in devicelist\_idsync\_jan17.csv

OUTPUT: mobile\_atlas\_gsmarena.csv

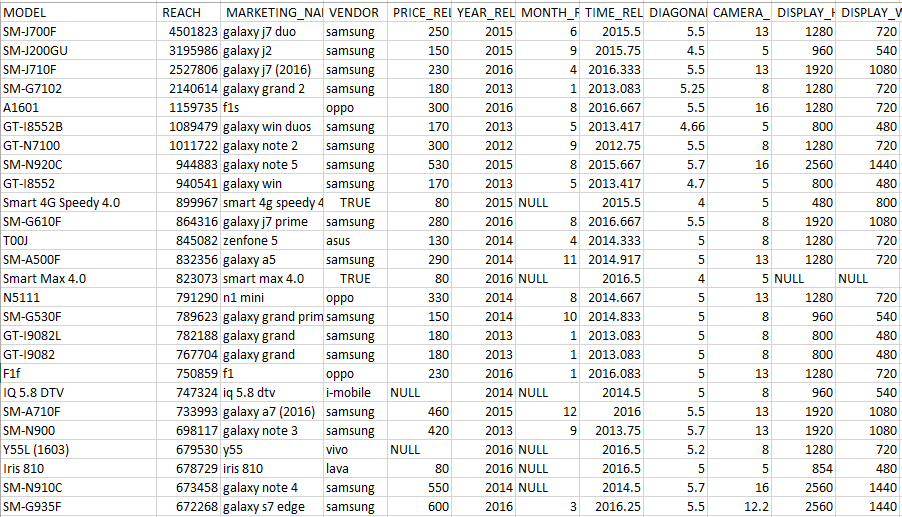


* The model names in mobile\_atlas\_gsmarena.csv for samsung are in the cut-version (not sm-j700f but sm-j700 for example).
* The mobile\_atlas\_gsmarena.csv table contain only info of models that are available on gsmarena.com.
* The model names are all in lcase. But for creating segmentation rules for DSP, it is still have to be a case sensitive now.
* Missing price for some phones (OEM, i-mobile, wiko, some samsung, etc.)

We need a complete table of all the models in Tapad + info from gsmarena (doing left join). I tried with R but it didn’t work so I do it in Impala instead.

1. Impala: Join mobile info from gsmarena to mobile atlas from Tapad & deal with missing price (merge\_model2gsm.sql)

OUTPUT: techname\_matched2atlas.csv This table can already be used for device based segmentation rules. (select by vendor, price, released time, screen, camera)



1. Python: Device Clustering (5.cluster\_dvc.py) (I used Spyder)

* K-means clustering, HAC, Dbscan
* Elbow method and silhouttte method for determining an appropriate cluster number
* Test clustering using (price, age, screen), (price, age, camera) and (price, age, screen, camera)

OUTPUT: DvcClusteringResults.csv (Augmented mobile atlas + clustering results)

Augmented = + released time with monthly precision + released price

1. Tableau: DVCClusteringViz
2. Impala: 6.DeviceBasedRuleCreation.sql

* Create segmentation rules for Android for Kaidee, Pantip, Facebook (not case sensitive) and Sanook (DSP) (case sensitive)
* Create Facebook audience using dtac\_vertical\_dataset + augmented mobileatlas with cluster