**Snapshot Report Tool: Rebuild (Notes Tracker)**

**2/24/22**

**Jeff’s follow up notes on Phase 1 approach** *with Hong Comments in red italics*

* Ideally the baseline group can be defined as part of the setup/input. This helps if we wanted to compare say prospects or customers in a defined market. Let’s say it’s a regional bank with a 10-state footprint. Here we don’t want all USP.
  + *[1. we will include filters (or parameters) in the script when we build baseline data. 2. As an optional if apply those filters to other data (such as customer file). 3. The filters have to come from attributes in appended 3rd party data, say Epsilon as our first data we will work on]*
* Can we create an index value (against baseline) for all the input segments? This is difficult to visualize on a bar/line chart, but it would be helpful to have in the underlying data. We could also use a heatgrid for the index values
  + *[we can have all indexes against baseline data in excel report. As Phase I, we may only include one index on a bar/line chart and we set up optional if include more indexes on a bar/line chart in Phase II]*
* Anywhere we have an index value (looking at pgs 6-7) let’s make sure 100 is always shown as the midpoint – eg. 100 is white, higher is green, lower is red, etc. Ideally it’s the midpoint on a line chart too, but understand that could be tricky to configure
  + *[it will be tricky on this one because most of attributes are already binned in 3rd party data. Let’s think about this in Phase II]*
* Possible to include a penetration rate (segment count / baseline count) by state and perhaps top 25 DMAs?
  + *[State will be definitely included, also we can include top DMAs or other common-used pre-defined geo attributes]*

**2/15/22**

**Jeff Shafer Thoughts**

* It would be helpful to include more than two input groups to get distributions – for example base population and then customer group A, B, C… You can make the index value for A, B, C relative to base
* Yes to editable charts. It would be great if we could define the color scheme or style once and replicate – eg. may want to use client’s theme for colors
* Can we pull in the descriptions and recodings from the data dictionary. For instance, if we pick up income tiers 1,2,3… can we label these with the corresponding values from the dictionary, eg. $20-30k.. Also possible to get the variable description on the page as it is in the dictionary
* Possible to generate a summary page with the input quantities?
* Can we get a ranking of some measure of variable importance? For example, these are the top 20 attributes with the greatest index value difference, limited by some minimum sample size… something like that. Basically answering the question ‘what is most important’

The more we can make this point and click to presentation ready the better. This is a super valuable tool for clients and future sales!

**2/14/22**

**Hong – Initial Input**

Here is high level information about current snapshot tool (attached are examples of outputs)

* We have R script to generate first excel report (one variable per sheet). If you have 100 variables in your input file, the report will include 100 sheets.
* We use second excel macro to generate charts (with indexes) and generate ppt with charts. Again one slide per variable.

Here are several my thoughts about enhancement of this tool:

* Pre-format for all (or selected important) Epsilon variables. For example, in current report
  + If Gender with two values 1 (Male) and 2 (Female), current report will only report distribution on 1 and 2.
  + For continuous variables (such as home values), it will be binned into 10 groups. If the data are different, the binned will be also different. We need pre-format all needed continuous Epsilon variables so all reports are comparable.
* The charts in PPT are un-editable. We need make changes in Excel macro (or VBA) to make them editable. Ideally we can group variables into different categories (such as demo, finance, lifestyle, …) and output charts in PPT within different categories.
* Currently excel report includes one variable per sheet (the sheet name is just number from 1 to N). We need change the number to variable name so it will be easy to find. Or put all report in one sheet.
* Any other thoughts?