Smots – 8/4/22

Who I am: I run a small automation software company selling into mid-sized US companies.

The task overall:

The task is to generate a script to recreate the following image (each with the 5 bar charts + legend, but without the blue boxes on the left and the [Geography 1] title on the top) for each of the Analytical Regions listed in the in the data set. As a result there will be 10 or so images, each with 5 bar charts on them.

The client will receive these images and can copy them onto a Powerpoint slide for their annotation, etc. The dimensions of each image should be 10” x 5.75”

Chart, bar chart

Description automatically generated

The output should be a single .py script that, when fed the inputs, generates a series of image outputs.

Geo mapping:

Before we go into the details for each chart, I’ll give a brief overview of the mapping. The external data provider for the very top chart provides CBSA codes. However, for internal data sets, there are “Club Names.” Both of these can be mapped into “Analytical Regions” per the below excel input.   
Graphical user interface, text, application

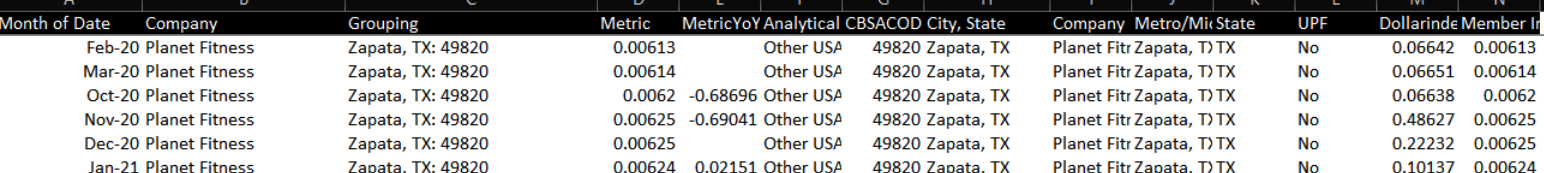
Description automatically generated

Each analytical region defined in this document should receive its own image. There are ten or so analytical regions and analytical regions contain multiple clubs.

Chart 1: Market Share Trends

The first chart is a look at relative market shares between the client and their competitors. This chart is the most complex, but should not be too tough.

The input file for this is “MSCience Raw Data\_1.xlsx” (use the tab labeled “MScience Raw Data MSA”) – it looks like:



The relevant columns are the date, the company, the CBSA code (used for geo mapping to the client’s analytical regions), and Member Index. The task is to recreate the bar chart by calculating the share of member indexes, by company, by date for each company to create a market share picture.

*Note – ignore the ‘analytical region’ column and the second ‘company’ column in this Excel as they are outdated, but still provided by the data provider.*

For detailed chart configuration, see “Market Share Bar Chart Configuration.xlsx” – this file has a definition for which competitors should be named on the bar chart – all other competitors can fall into the “other” bucket. You will also notice hex colors for the bar chart that you can pull from. Lastly, you’ll notice some competitors go by multiple names, in those situations we will define them in this Excel to combine into one.

Graphical user interface, text

Description automatically generated

For chart labeling, we want to label the most recent month and each prior December – and they only need to be displayed where the bar is sufficiently large to hold a label, as shown in the screenshot. If this proves challenging, please provide alternatives.

This chart will also require a legend along the top. It does need to be at the top, but it does not need to be in this precise formatting (i.e., can be along the top in one row rather than two, etc).

Chart 2: UPF Members

This chart should be far more straightforward. The key data file is “Active Members.xlsx”:

Table

Description automatically generated

You will also need the club name to analytical region mapping from “Geo Mapping.xlsx.” This is simply a matter of summing all clubs in each analytical region and charting. I expect this chart to be fairly straightforward, so will keep the directions short. Labeling and formatting can be consistent with the image above. Please display in rounded thousands as seen on the chart above.

Please feel free to ask questions as needed.

Chart 3: # UPF Clubs

This chart will require only “Geo Mapping.xlsx” (re-pasting screenshot):  
Graphical user interface, text, application

Description automatically generated

This chart requires counting clubs open since the open-date as listed above and summing by analytical region. You can assume any club displayed here is open and that no clubs have closed.

Similarly, I imagine this to be fairly straightforward, so will keep the directions short. Labeling and formatting can be consistent with the image above.

Chart 4: UPF Members per Club

This chart requires no new data – it is a division between the two prior chart (UPF Members divided by # UPF Clubs). I will similarly keep this instruction short, please label and format consistent with the image above. Please display as rounded numbers (not in thousands).

Chart 5: Est. Overall Gym Members (K)

This chart also requires no new data and can be derived by two prior charts. In the first chart, we establish Planet Fitness market share as a % of the overall market. In the second chart, we establish Planet Fitness members. This is a division of Planet Fitness members / Planet Fitness % market share. I will similarly keep this instruction short, please label and format consistent with the image above. Please display as rounded thousands.

Additional notes

1. While the chart at the very beginning of this document goes back to Jan 2017, please make the chart go back to the same month 5 years ago – i.e., if the “newest” month in the dataset (as defined by the newest month in the is “MSCience Raw Data\_1.xlsx” document) is May 2022, have the chart go back to May 2017
2. For the x-axis labels, please display them as “Mmm ‘YY” i.e., Jan ’19. This is intentionally different than the MM/D/YY label shown.
3. The chart has a gap from 3/1/20-9/1/20 – you can ignore the gap and produce the data as if that gap was not there. The recipient of the data can recreate that gap via powerpoint.
4. There are a couple small requirements regarding conforming to our platform:
   1. We use python 3.9
   2. Please use matplotlib
   3. Use [typer](https://typer.tiangolo.com/) for inputs – our scripts typically look like:  
        
      Text

      Description automatically generated

*[ guts of script ]*

Graphical user interface, text

Description automatically generated

While our inputs in this example were ‘--olr’ and ‘--key’, for this project they will be ‘--mscience’, ‘--chartconfiguration’, ‘--geomapping’, etc, etc for each of the input files.

1. Lastly, I am pythonically fluent and have written most of the past scripts myself – if you have any question about how to handle certain instances, don’t hesitate to ask (ajskhan@gmail.com)