6. Volume adjustment

Volume being **excluded** from the comparable selection, we can have big gaps between a target and its comparables. That is why we need a specific volume adjustment step.

The objective is to adjust the prices of comparables to answer the following question : **"what would be the comparable cpc price if it was sold at the same volume as the target cpc ?"**

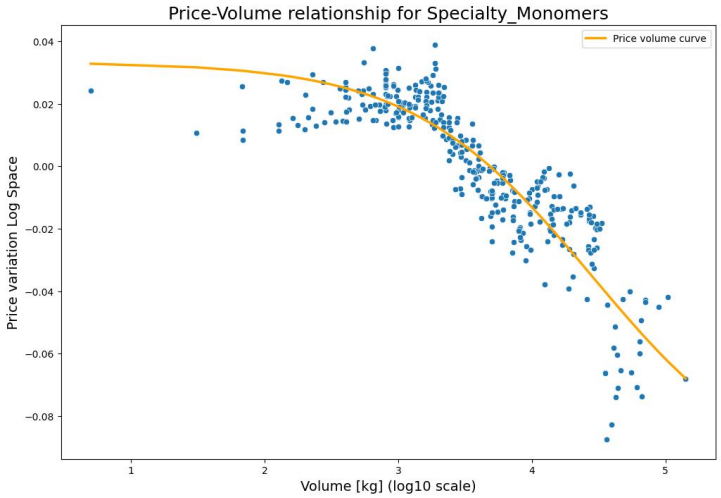
To do this, we apply the following two methods:

Volume adjustment function

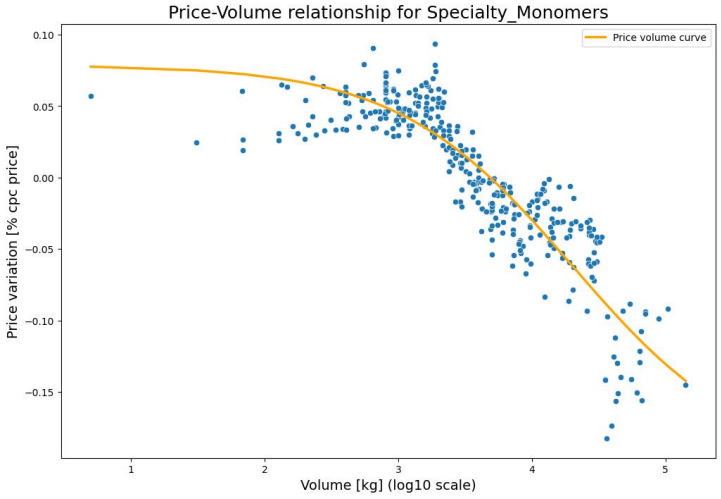
Example for Specialty Monomers (CS):

- We start from the SHAP value of the volume (please refer to section 4 above).
  Each dot here represents a CPC and the x-axis (SHAP value of cpc_volume_log) goes from -0.09 to +0.06

- We add the volume as a second dimension on the X axis to create a scatterplot.
  - On the Y-axis, we retrieve our SHAP values with the scale between -0.09 and +0.04.
  - Each dot still represents a CPC.
  - We can see that the SHAP values decreases as volume increase.



- We find the best fit function to the scatterplot and display it as a curve (in yellow above).
- Finally, we apply the log inverse function (10^x) to the y-axis to obtain a result that we can interpreted. The y-axis value is now a price variation due to volume, compared to the average prices of the CPC of the family.
  Note : the shape of the curve does not change, only the Y-axis is modified



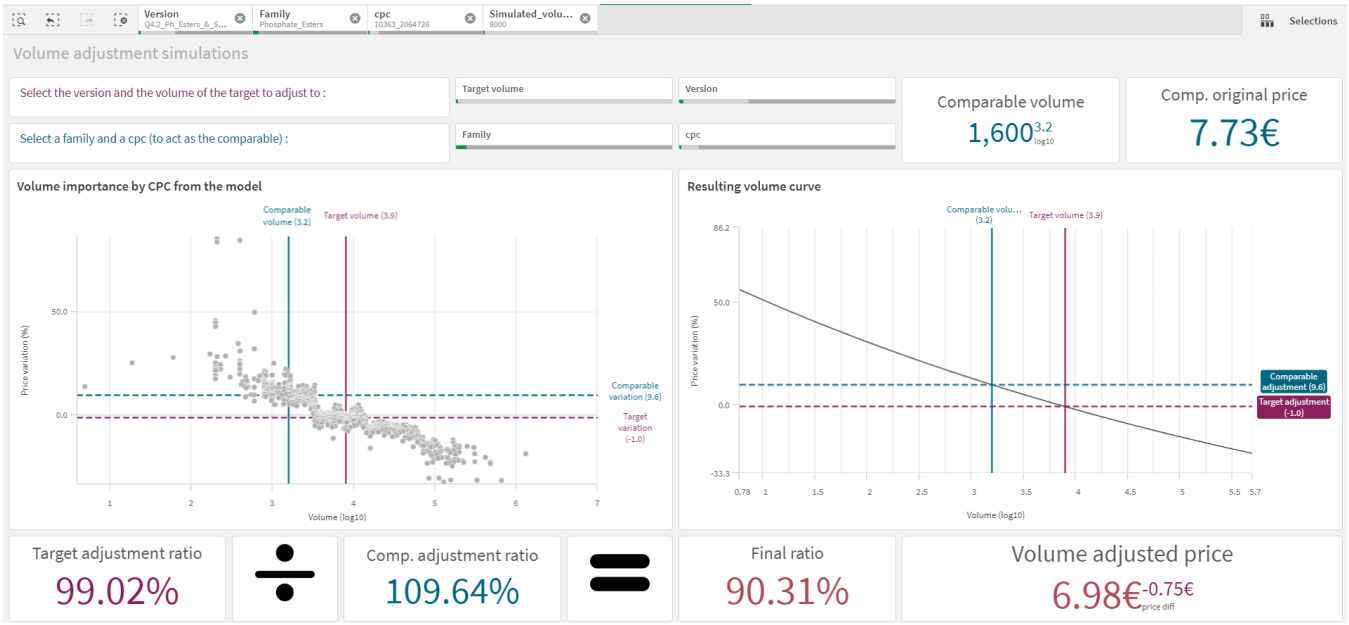
- Then, we adjust every comparable using the fit function.

Dashboard simulation tool

These steps are simulated in the **"Volume adjustment simulations"** tab of the Qlik dashboard for a better understanding based on real values of the chosen run (version).

To use the tab, you need to select a family and a CPC that will be used as a comparable. Then, you can enter the volume you want to adjust to. This simulates the volume of the target cpc.
By changing the volume, you can see how the volume adjusted price is affected.

Phosphate_Esters example (CS):



- The selected CPC has a volume of 1 600 and we are comparing to a 8 000 volume target.
  When adjusting, we expect to have an decrease in price since the target has a lower volume.

- The scatterplot on the left is a representation of all the CPCs of the family and their **price variation due to volume** (based on SHAP values from the first model).
  It is the **same scatterplot** as the one displayed above.

- The line chart on the right is the **fit function** for the family, originated from the scatterplot.

- **X-axis** (volumes) on both of the graphs are on a **log scale** to make it more readable and less prone to outliers.

- **Y-axis** (price variations) on both of the graphs are **out of the scale** and represent understandable price variations. That is why the scale is similar to the second scatterplot above, and not the first one.

- Our **comparable** CPC:
  - Has a volume of 1 600, which equals **3.2** in log10 scale.
  - Crosses the fit function at an adjustment value of **9.64%**.
    - This means that its volume increases its price by 9.64% compared to the family average.
    - This leads to a ratio of **109.64%** (= 100% + 9.64%).

- Our **target** CPC(not defined since it works for any target with this volume) :
  - Has a volume of 8 000, which equals 3.9 in log10 scale.
  - Crosses the fit function at an adjustment value of **-0.98%**.
    - This means that its volume lowers its price by 0.98% compared to the family average.
    - This leads to a ratio of **99.02%** (= 100% - 0.98%).

- To adjust our comparable to the target volume, we compute the **ratio** between the target value and the comparable value.
  - 99.02% / 109.64% = **90.31%**

- **In the end, this comparable will have its price decrease by 9.69% if we were to compare it with a target having a volume of 8 000.**

- **Its price goes from 7.73€ to 6.98€. The volume adjusted price decrease is 0.75€.**

This volume adjustment is applied on every comparable of every target CPC, no adjustment is applied on the target directly !

Group adjustment

We mentioned before the "cpc_volume_log" feature which is showing the volume of a CPC. We also have a feature named "group_volume_but_cpc_label" looking at the total volume of the customer group of the CPC, **excluding volume of the CPC itself.**

Indeed, the size of the entire group is supposed to have an impact on the CPC price. That is why this group adjustment step is needed.

This "**group_volume_but_cpc_label**" variable has 5 modalities :

- 0_one_cpc : The volume of the group for this family excluding the current CPC is equal to **zero**.
  This means that the group of the customer **only buys this CPC** and the adjustment needed is entirely covered by the step presented in section 5.
- Other CPCs whose groups have sales in the family (excluding the CPC ones) are split in 4 bins representing **quartiles** :
  - 1_small : 25% of the CPC having the smallest group volumes (first quartile)
  - 2_medium : 25% of the CPC representing the second quartile
  - 3_big : 25% of the CPC representing the third quartile
  - 4_top: 25% of the CPC having the biggest group volumes (fourth quartile)

Once every CPC is placed in its modality, we compute the median SHAP value of "group_volume_but_cpc_label" for every modality. This gives us the estimated impact of the group volume on price.

group_volume_but_cpc_label	median_shap
0_one_cpc	0.013
1_small	0.014
2_medium	0.009
3_big	0.002
4_top	-0.036

Example of data

As a reminder, these values come from the SHAP values and therefore are on the **log scale**. After transformation to get out of the log scale, they will be applied as a factor to previously computed volume adjusted price for the families having the group adjustment applied.

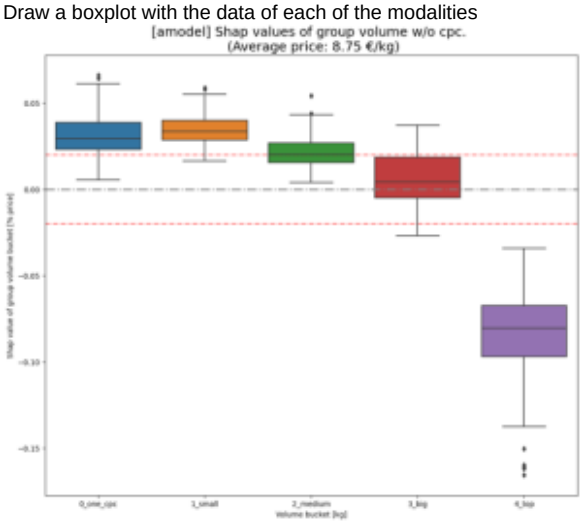
How to define which family has a group adjustment applied ?

To decide if we will apply a group volume adjustment for a family, we :

- Apply the transformation on SHAP values to get out of the log scale and obtain a price variation percentage.

group_volume_but_cpc_label	median_shap
0_one_cpc	0.029
1_small	0.033
2_medium	0.020
3_big	0.004
4_top	-0.080

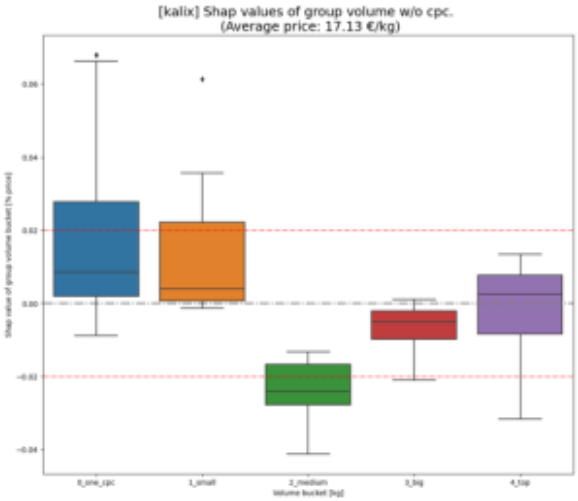
- Draw a boxplot with the data of each of the modalities



This boxplot shows us that customers belonging to small groups are buying at an average of 3% higher than the rest and customers belonging to the biggest groups 8% lower.

Red lines represent the -2% and +2% price variation due to group volume threshold. **We only apply a group volume adjustment when the medians of the different modalities are outside these bounds and the behavior of decreased price is coherent between groups.**

Another example :



- The median of one of the groups only is outside the bounds.
- The price variation is not coherent : prices increase for the biggest groups when we expect a decrease.
- **Conclusion : we do not apply a group volume adjustment for this family.**

Once we have the price variation due to group adjustment for our comparable CPC, we check the difference with our target CPC:

- **If they are both in the same modality, therefore having the same group size considered, we do not apply any adjustment.**
- **If they have different modalities, the adjustment will be the difference between the target bucket and the comparable bucket.**