Walkthrough

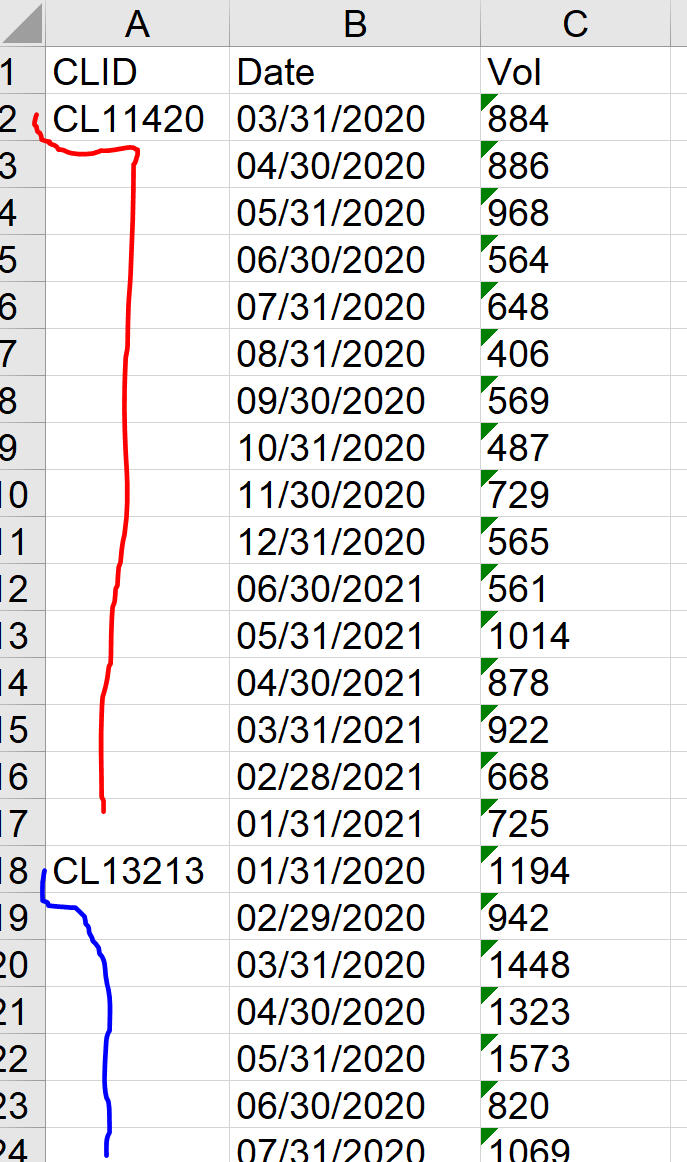
So, in this project, we need to figure out Q2 2021 volumes, for each region, and if everything “looks good”. I assume that means, seeing if overall Q2 YoY, had AT LEAST positive growth, and ideally, is equal to or better than, Q1 YoY growth.

Thus, we’ll need to figure out how much volume is in each region, not only for Q2 2021, but also of previous quarters, so we have something to compare against, to see how “good its going”. Let’s get started.

Data Cleaning

First, we need to clean the data, starting with the dates/volumes sheet.

1. The volume data for each date, was grouped by CLID, like this:



So, I inserted the appropriate CLID, for all blank cells.

1. Date stored as text (not dates), as the filter doesn’t group them, they’re all individual entries:

Table

Description automatically generated

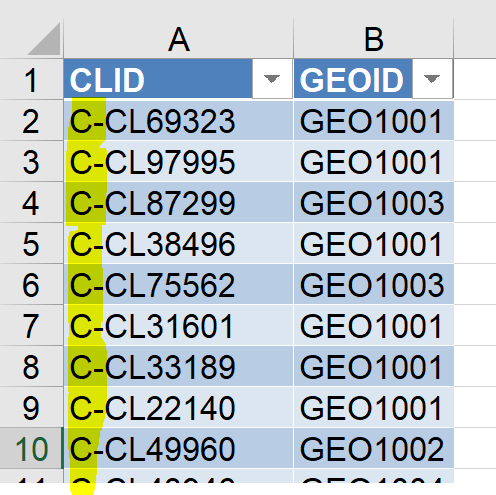
Thus, they were reformatted to proper dates, via Text to Columns wizard.

1. Vol values stored as text, not numbers, so they were reformatted to numbers, via Text to Columns wizard.

Next, on the “Sheet3” sheet, the following was done:

1. Need to figure out which GEOID’s, correspond to which regions (NAM, EMEA, APAC, LATAM). It’s stated that NAM is GEO1001, and EMEA is GEO1003, but we need to figure out the TOTAL volume for each region, to know what APAC and LATAM are (since LATAM is the lowest). So we’ll start with that.

The CLID is inconsistent between the two tables, as shown here:

Table

Description automatically generated

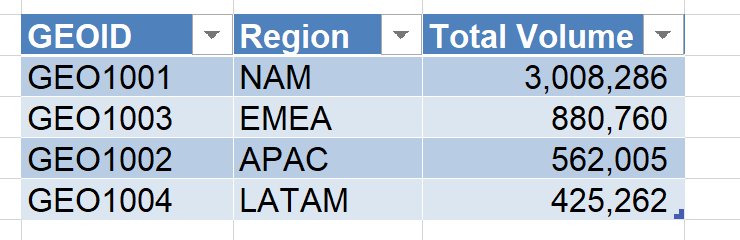
From calculating the length of CLID in both tables, and using a filter, we can see that there is always an extra “C-“. So, we created a new column, with it cut off, using the MID function.

Now that we have a matching column in both tables, we can use XLOOKUP, to extract the GEOID’s:

Table

Description automatically generated

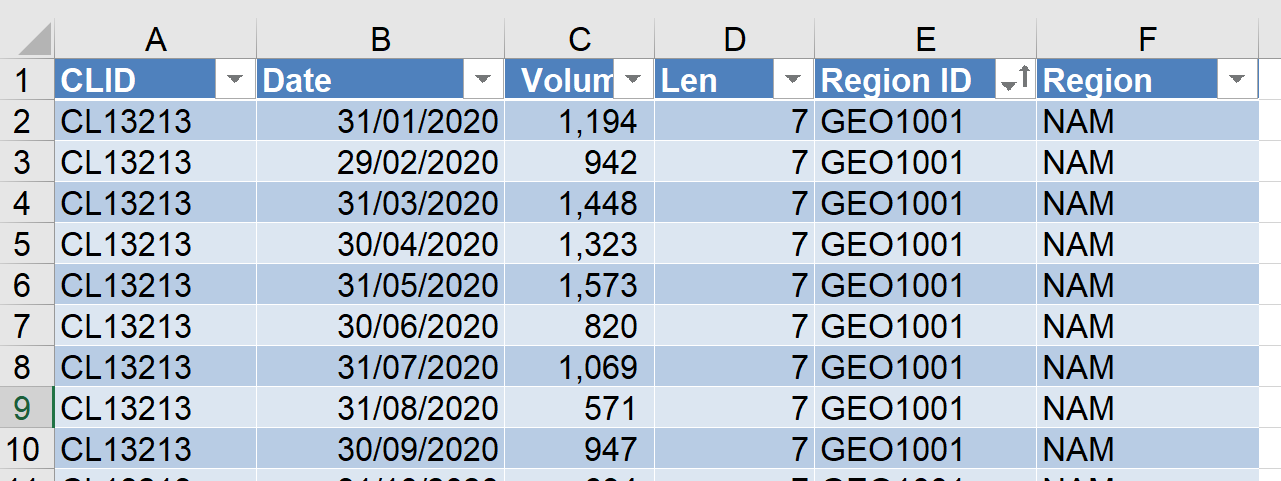
We now know the volume for each client, on each date, and what region they were in. Thus, we can figure out the total volume for each region, via a SUMIF:



The lowest volume belonged to GEO1004, thus LATAM is GEO1004, and APAC is GEO1002.

Excellent, we now know all 4 regions, and how much volume is in each OVERALL, but we still need to figure out how much is in each, PER QUARTER.

Okay, first, we’ll just add a region name column to the table containing the dates, via a VLOOKUP, to make things a bit clearer:



Okay, now we need to figure out which quarter, each row falls into. We can do this, by building a column just for the quarter, and working out what it is, via the date column, and a custom formula:

Table

Description automatically generated

Now, we can just sum together all volumes, by Quarter, AND by Region (or Region ID), using a SUMIFS:

Table

Description automatically generated

And voila! We’ve managed to clean/transform the data, to a point where we can begin to answer the original request, thus concluding the data cleaning stage. Now, we move onto the exploring/analysis stage:

Data Exploring & Analysis

First, after looking at line charts, of the volume by region and quarters:

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

It appears that this company has seasonality, as its volume ramps up across ALL regions, from mid Q3, to the start of Q2 next year, then has a large slowdown.

So, it would be misleading/pointless to compare Q2 2021, to Q1 2021, as the business is in its natural rising period of its seasonality, so of course it would “look good” in that comparison.

Instead, to effectively evaluate performance, we should compare Q2 2021, to the same quarter last year (a “Year-over-Year” approach), as now we’re cutting out the seasonality aspect (it’s the same season in both cases). We’ll also do a YoY for Q1 2021, as we have enough data to cover it:

Table

Description automatically generated

**Now we have a BASIC answer to the original request, Q2 2021 has experienced a 2.67% YoY growth in volume. However…**

We don’t know why it’s 2.67% (Q1 experienced 4.04% after all), and these results aren’t exactly pleasant to look at. So, we’ll continue on, to see if we can figure out why Q2 2021’s YoY growth was only 2.67%, compared to Q1 2021 (4.04%), and we’ll also make a visualisation at the end, so our findings are more presentable.

First, we’ll look at how each regions Q2 YoY, differs from its Q1 YoY:

Table

Description automatically generated

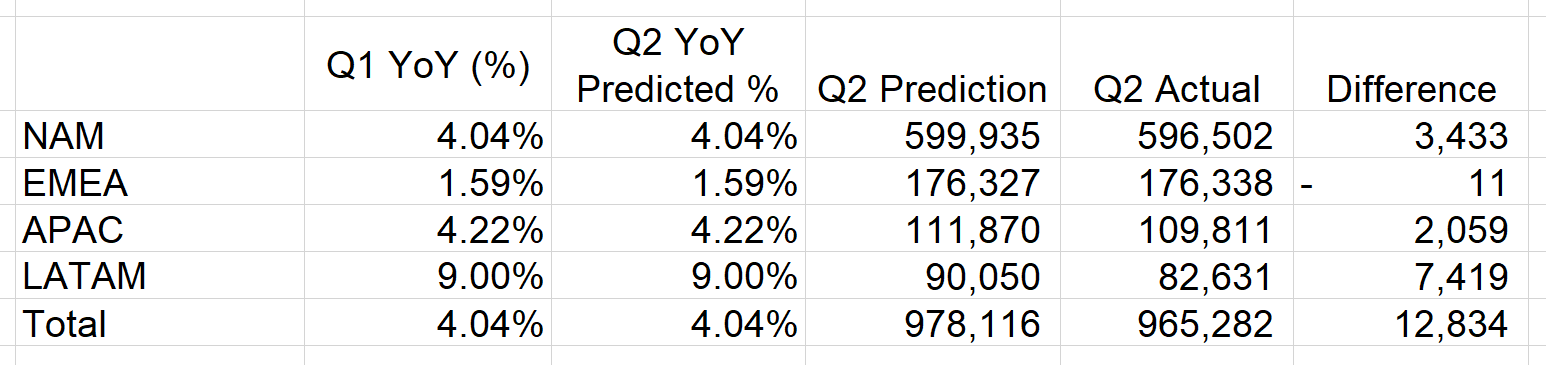
In the above, we can see that for every region, except EMEA, Q2 YoY is noticeably different, to Q1 YoY (it’s smaller).

Specifically:

* NAM’s Q2 YoY, is 85.28% the size of the Q1 YoY.
* APAC’s Q2 YoY, is 54.57% the size of the’ Q1 YoY.
* LATAM’s Q2 YoY, is only 0.17% the size of the Q1 YoY.
* EMEA, the exception, has a Q2 YoY, that is actually 0.4% larger than the Q1 YoY (basically the same).

So, we can see that for every region except EMEA, the quarterly YoY is lower, ESPECIALLY in LATAM’s case, where there was almost no growth at all.

We’ll try predict Q2 YoY, based off Q1 YoY (so we’ll assume Q2 YoY, would have the same growth as Q1, for each region):



Now, interestingly, when we look at our Q2 2021 deviation, which is +12,834, we can see that most of it came from LATAM, which makes sense. But is it just because LATAM makes up the majority of volume? No, see next page:

A picture containing calendar

Description automatically generated

Despite LATAM making up 58% of the deviation, it only accounts for 9% of the actual prediction.

So, let’s see if we can explain this deviation, (why Q2 YoY is lower than Q1 YoY). We’ll start with LATAM, as it seems quite unusual compared to the rest, we may find something uniquely interesting.

If we look through the customer transactions for LATAM, we find this:

Graphical user interface, table, Excel

Description automatically generated

CL22675, emerged in Q3 2020, and had the third largest volume of Q1 2021, then vanished in Q2 2021. Additionally, CL11420had placed a much smaller order than usual, in Q1 2020, compared to Q1 2021.

So the sole cause of 58% of the deviation (that comes from LATAM), is due to CL22675 vanishing in Q2 2021, and to a lesser extent, CL11420 placing a much larger order in Q1 2021 compared to Q1 2020.

There are other customers, like CL43946, who decrease Q2 growth by 655 units compared to Q1 growth, but clients like CL49900, raise it by 723, so overall it cancels out. It really is just CL22675 being responsible for 80% of the 58% deviation, and CL11420 for the remaining 20% of the 58% deviation.

Now, the remaining 42% of the deviation, comes from APAC and NAM, and they don’t seem like they’ll contain anomalies, as their share of the deviation is much more in line with how big they are to begin with, so its unlikely we’ll find something that really stands out, it’s likely more of a trend thing.

So, to detect these sorts of trends, we’ll first look at average volume per quarter, for each region.

Table

Description automatically generated

Interestingly, average volume actually suggests that Q1 YoY growth (4.04%), should be LOWER than Q2 YoY growth (2.67%), as Q1 had a change of -3.81%, and Q2 had a change of 2.67%. So this is actually evidence directly AGAINST what we’re investigating… interesting…

We’ll take a look at if any new customers are coming in (or leaving, which just means they had 0 volume):

A picture containing chart

Description automatically generated

Now this some more sense, Q1 total customers went up by 4, that’s an 8.16% increase, whereas Q2 had no change, so this supports Q1’s YoY growth being stronger than Q2’s. However, not all customers are equal, when it comes to volume, and this is pretty weak evidence, so we’re going to have to just look at each customer, for APAC, and NAM. (EMEA contributed virtually nothing to the deviation, its Q1 and Q2 YoY are basically the same).

Graphical user interface, text, application, table

Description automatically generated

For APAC customers (sorted largest volume to smallest), the main reason for Q2’s YoY being smaller than Q1’s, is due to APAC’s largest customer, CL49960, ordering only 484 more units (0.92%), for Q2 YoY, compared to 1,948 more units (4.25%), for Q1 YoY. That means Q2 YoY is 1,464 units behind Q1 YoY, which makes up almost the entire total difference of -1,569, and puts Q2’s YoY at only 2.3%, compared to Q1’ YoY at 4.22%.

Finally, for NAM customers, Q2 YoY here was 3.45%, compared to Q1 YoY’s 4.04%, so Q2 is only behind by 0.59%, meaning it may be tricky to notice the deviation as much.

Table

Description automatically generated

The only thing that stands out amongst NAM customers, is customer CL64939, whom placed a considerably smaller order in Q2 2021, compared to Q2 2020, the change in volume was -2,602. That by far had the biggest impact on Q2 YoY’s growth dropping, from the expected 4.04% as Q1 had, to a smaller, 3.45%.

With the remaining 42% deviation reasonably explained, we can move onto the summary, on the next page.

Summary of Findings

Overall, Q2 YoY growth, slowed to 2.67%, down from Q1 YoY growth of 4.04%.

NAM, EMEA, APAC, and LATAM, had Q1 YoY’s of 4.04%, 1.59%, 4.22%, and 9% respectively. Their Q2 YoY’s was 3.45%, 1.6%, 2.3%, and 0.02%, respectively.

Thus NAM is down 0.6%, EMEA is stable, APAC is down 1.9%, and LATAM is down 9%.

* 58% of this overall slowdown, is due to LATAM. Specifically, mainly two customers, "CL22675", and "CL11420". CL22675 placed a huge order (4,809 volume), in Q1 2021, then vanished in Q2 2021. CL11420, a smaller sized customer, had strong Q1 YoY growth (161%), as they were just starting up in Q1 2020, but had very weak Q2 YoY growth (1%).

Overall loss of ≈6.2k volume.

* 28% of this overall slowdown, is due to NAM. Specifically, mainly customer “CL64939”, whose Q2 YoY strongly declined (-61.72%), as their Q2 2021 volume, was far smaller than their Q2 2020 volume. (It was smaller, by 2,602 units of volume). Loss of ≈2.7k in actual volume.
* 16% of this overall slowdown, is due to APAC. Specifically, the largest customer in the region, “CL49960”. They had strong Q1 YoY growth (4.25%), but very weak Q2 YoY growth (0.92%). Loss of ≈1.5k in actual volume.
* The EMEA regions Q2 YoY growth (1.60%), was in line with Q1 YoY growth (1.59%), thus no investigation performed.

Improvements for the future

Unfortunately, activity in Q2, is generally on a larger scale, than activity in Q1, by about 14%. This means, if we compare the units of volume (not percentage, just actual units), between Q1 and Q2, then the comparison isn’t truly fair, as due to the nature of the business (based on the limited dataset), Q2 volumes are always going to be a bit bigger (on average, 14% bigger), than Q1 volumes.

Thus, if more time was permitted, and a more thorough investigation was desired, then early on, during data cleaning, we could scale down all of Q2’s volumes, by about 14%, so that it’s quantities for volumes, are now more fairly comparable, to Q1’s quantities for volume.

This means we could then compare their absolute differences, work out what the total absolute difference is, then scale the total difference back up by 13%, to get a more precise idea, of how much of the missing 12.8k expected volume, we’ve accounted/explained for (quick math shows 11.4k/12.8k explained, which is 89%, instead of the current 10.5k/12.8k, which is 82%)