Walkthrough

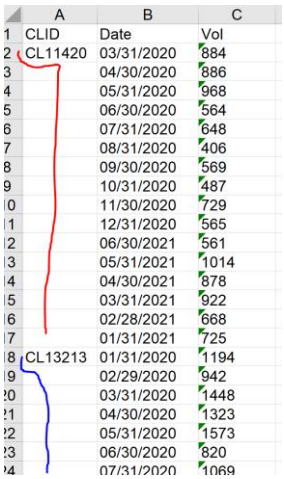
So, in this project, we need to figure out Q2 2021 volumes, for each region, and if everything "looks good". I assume that means, seeing if overall Q2 YoY, had AT LEAST positive growth, and ideally, is equal to or better than, Q1 YoY growth.

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.

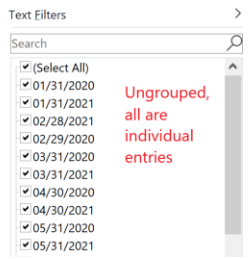
1. The volume data for each date, was grouped by CLID, like this:



	A	B	C
1	CLID	Date	Vol
2	CL11420	03/31/2020	884
3		04/30/2020	886
4		05/31/2020	968
5		06/30/2020	564
6		07/31/2020	648
7		08/31/2020	406
8		09/30/2020	569
9		10/31/2020	487
10		11/30/2020	729
11		12/31/2020	565
12		06/30/2021	561
13		05/31/2021	1014
14		04/30/2021	878
15		03/31/2021	922
16		02/28/2021	668
17		01/31/2021	725
18	CL13213	01/31/2020	1194
19		02/29/2020	942
20		03/31/2020	1448
21		04/30/2020	1323
22		05/31/2020	1573
23		06/30/2020	820
24		07/31/2020	1069

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

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



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

3. 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:

	A	B	C
1	CLID	Date	Vol
2	CL11420	31/03/2020	884
3	CL11420	30/04/2020	886
4	CL11420	31/05/2020	968
5	CL11420	30/06/2020	564
6	CL11420	31/07/2020	648
7	CL11420	31/08/2020	406
8	CL11420	30/09/2020	569
9	CL11420	31/10/2020	487
10	CL11420	30/11/2020	729

	A	B
1	CLID	GEOID
2	C-CL69323	GEO1001
3	C-CL97995	GEO1001
4	C-CL87299	GEO1003
5	C-CL38496	GEO1001
6	C-CL75562	GEO1003
7	C-CL31601	GEO1001
8	C-CL33189	GEO1001
9	C-CL22140	GEO1001
10	C-CL49960	GEO1002

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:

	A	B	C	D	E
1	CLID	Date	Volum	Len	Xlookup Re
2	CL13213	31/01/2020	1,194	7	GEO1001
3	CL13213	29/02/2020	942	7	GEO1001
4	CL13213	31/03/2020	1,448	7	GEO1001
5	CL13213	30/04/2020	1,323	7	GEO1001
6	CL13213	31/05/2020	1,573	7	GEO1001
7	CL13213	30/06/2020	820	7	GEO1001
8	CL13213	31/07/2020	1,069	7	GEO1001
9	CL13213	31/08/2020	571	7	GEO1001
10	CL13213	30/09/2020	947	7	GEO1001
11	CL13213	31/10/2020	694	7	GEO1001

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:

GEOID	Region	Total Volume
GEO1001	NAM	3,008,286
GEO1003	EMEA	880,760
GEO1002	APAC	562,005
GEO1004	LATAM	425,262

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:

	A	B	C	D	E	F
1	CLID	Date	Volum	Len	Region ID	Region
2	CL13213	31/01/2020	1,194	7	GEO1001	NAM
3	CL13213	29/02/2020	942	7	GEO1001	NAM
4	CL13213	31/03/2020	1,448	7	GEO1001	NAM
5	CL13213	30/04/2020	1,323	7	GEO1001	NAM
6	CL13213	31/05/2020	1,573	7	GEO1001	NAM
7	CL13213	30/06/2020	820	7	GEO1001	NAM
8	CL13213	31/07/2020	1,069	7	GEO1001	NAM
9	CL13213	31/08/2020	571	7	GEO1001	NAM
10	CL13213	30/09/2020	947	7	GEO1001	NAM

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:

G2 = "Q" & ROUNDUP(MONTH([@Date])/3, 0) & " " & YEAR([@Date])

	A	B	C	D	E	F	G	H
1	CLID	Date	Volum	Len	Region ID	Region	Quarter	
2	CL13213	31/01/2020	1,194	7	GEO1001	NAM	Q1 2020	
3	CL13213	29/02/2020	942	7	GEO1001	NAM	Q1 2020	
4	CL13213	31/03/2020	1,448	7	GEO1001	NAM	Q1 2020	
5	CL13213	30/04/2020	1,323	7	GEO1001	NAM	Q2 2020	
6	CL13213	31/05/2020	1,573	7	GEO1001	NAM	Q2 2020	
7	CL13213	30/06/2020	820	7	GEO1001	NAM	Q2 2020	
8	CL13213	31/07/2020	1,069	7	GEO1001	NAM	Q3 2020	

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

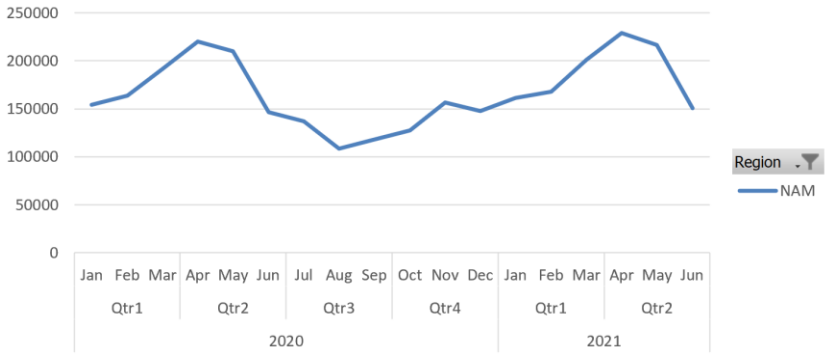
=SUMIFS(VolumeByClient[[Volume]:[Volume]], VolumeByClient[[Region]:[Region]], [@Performance], VolumeByClient[[Quarter]:[Quarter]], Performance[[#Headers],[Q1 2020]])

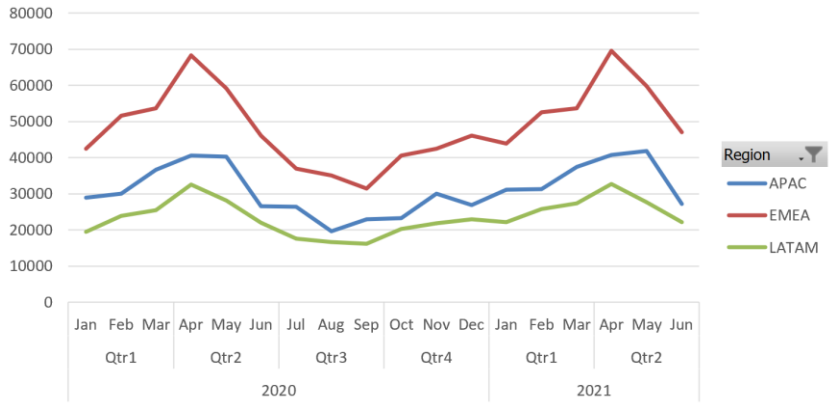
	L	M	N	O	P	Q	R	S	T
	Performance	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021	Q2 2021		
	NAM	509,419	576,618	363,694	432,034	530,019	596,502		
	EMEA	147,852	173,566	103,536	129,264	150,204	176,338		
	APAC	95,736	107,338	69,198	80,144	99,778	109,811		
	LATAM	69,053	82,618	50,574	65,121	75,265	82,631		

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:





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:

Variance Analysis 1

	Q1 YoY (%)	Q1 YoY (#)	Q2 YoY (%)	Q2 YoY (#)
NAM	4.04%	20,600	3.45%	19,884
EMEA	1.59%	2,352	1.60%	2,772
APAC	4.22%	4,042	2.30%	2,473
LATAM	9.00%	6,212	0.02%	13
Total	4.04%	33,206	2.67%	25,142

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:

Variance Analysis 1					
	Q1 YoY (%)	Q1 YoY (#)	Q2 YoY (%)	Q2 YoY (#)	Q2 vs Q1
NAM	4.04%	20,600	3.45%	19,884	85.28%
EMEA	1.59%	2,352	1.60%	2,772	100.40%
APAC	4.22%	4,042	2.30%	2,473	54.57%
LATAM	9.00%	6,212	0.02%	13	0.17%
Total	4.04%	33,206	2.67%	25,142	66.21%

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):

	Q1 YoY (%)	Q2 YoY Predicted %	Q2 Prediction	Q2 Actual	Difference
NAM	4.04%	4.04%	599,935	596,502	3,433
EMEA	1.59%	1.59%	176,327	176,338	- 11
APAC	4.22%	4.22%	111,870	109,811	2,059
LATAM	9.00%	9.00%	90,050	82,631	7,419
Total	4.04%	4.04%	978,116	965,282	12,834

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:

	Q1 YoY (%)	Q2 YoY Predicted %	Q2 Prediction	Prediction Proportion	Difference	Difference Proportion
NAM	4.04%	4.04%	599,935	61%	3,433	27%
EMEA	1.59%	1.59%	176,327	18%	- 11	0%
APAC	4.22%	4.22%	111,870	11%	2,059	16%
LATAM	9.00%	9.00%	90,050	9%	7,419	58%
Total	4.04%	4.04%	978,116	100%	12,834	100%

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:

	A	B	C	D	E	F	G	H	I	J	K	L
33												
34	**Sum of Volume**	**Column Labels**										
35		**2020**				**2021**						
36	**Row Labels**	**Qtr1**	**Qtr2**	**Qtr3**	**Qtr4**	**Qtr1**	**Qtr2**					
82	**LATAM**									ΔQ1	ΔQ2	DIFF
83	CL43946	41,282	49,071	28,827	36,607	41,985	50,429			703	1,358	- 655
84	CL85641	11,480	13,176	8,078	9,778	11,595	13,523			115	347	- 232
85	CL36191	4,139	4,910	2,891	3,665	4,268	4,961			129	51	78
86	CL81431	4,076	4,680	2,879	3,476	4,222	4,678			146	- 2	148
87	CL23634	2,665	3,174	1,864	2,376	2,667	3,248			2	74	- 72
88	CL11420	884	2,418	1,623	1,781	2,315	2,453			1,431	35	1,396
89	CL22675			1,249	3,569	4,809				4,809	-	4,809
90	CL37879	1,324	1,619	913	1,211	1,336	1,636			12	17	- 5
91	CL49900	1,639	1,879	1,153	1,402	483				- 1,156	- 1,879	723
92	CL67438	982	1,079	691	806	996	1,088			14	9	5
93	CL28683	582	612	406	450	589	615			7	3	4
94										6,205	10	6,195

CL22675, emerged in Q3 2020, and had the third largest volume of Q1 2021, then vanished in Q2 2021. Additionally, CL11420 had 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.

Average Volume								
	2020 Qtr1	2020 Qtr2	2020 Qtr3	2020 Qtr4	2021 Qtr1	2021 Qtr2	Q1 Δ Avg Volume	Q2 Δ Avg Volume
NAM	28,301	30,348	19,142	21,602	26,501	29,825	-6.36%	-1.72%
EMEA	18,482	21,696	12,942	16,158	18,776	22,042	1.59%	1.60%
APAC	7,364	8,257	4,943	5,725	7,127	8,447	-3.22%	2.30%
LATAM	6,905	8,262	4,598	5,920	6,842	9,181	-0.91%	11.13%
Total	16,777	18,803	11,289	13,331	16,137	19,306	-3.81%	2.67%

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):

Customer Flow								
	2020 Qtr1	2020 Qtr2	2020 Qtr3	2020 Qtr4	2021 Qtr1	2021 Qtr2	Q1 Δ Customer	Q2 Δ Customer
NAM	18	19	19	20	20	20	2	1
EMEA	8	8	8	8	8	8	0	0
APAC	13	13	14	14	14	13	1	0
LATAM	10	10	11	11	11	9	1	-1
Total	49	50	52	53	53	50	4	0

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).

L	M	N	O	P	Q	R	S	T
	Customer Volume Changes (%)					Customer Volume Changes (#)		
	Q1 %	Q2 %	DIFF %			ΔQ1	ΔQ2	DIFF
APAC	4.25%	0.92%	-3.33%			1,948	484	- 1,464
APAC	-0.04%	2.21%	2.25%			- 11	647	658
APAC	2.04%	3.47%	1.43%			87	176	89
APAC						53	73	20
APAC	1.02%	2.90%	1.88%			31	98	67
APAC	3.09%	0.84%	-2.26%			90	27	- 63
APAC	44.17%	2.56%	-41.61%			762	72	- 690
APAC	3.48%	2.70%	-0.77%			79	73	- 6
APAC	0.93%	0.27%	-0.66%			11	4	- 7
APAC	1.52%	1.54%	0.03%			13	14	1
APAC						1,132	1,254	122
APAC	1.56%	-0.38%	-1.94%			11	- 3	- 14
APAC	-39.27%	-100.00%	-60.73%			- 172	- 460	- 288
APAC	5.06%	9.03%	3.97%			8	14	6
APAC Total	4.22%	2.30%	-1.92%			4,042	2,473	- 1,569

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.

	Customer Volume Changes (%)			Customer Volume Changes (#)		
	Q1 %	Q2 %	DIFF %	ΔQ1	ΔQ2	DIFF
NAM	3.43%	3.20%	-0.22%	3,259	3,267	8
NAM	2.32%	1.53%	-0.79%	2,022	1,430	- 592
NAM	1.36%	3.57%	2.20%	939	2,820	1,881
NAM	2.42%	2.09%	-0.32%	1,387	1,379	- 8
NAM	1.40%	3.78%	2.37%	723	2,396	1,673
NAM	3.17%	1.97%	-1.20%	1,516	1,161	- 355
NAM	2.77%	0.48%	-2.29%	1,060	211	- 849
NAM	0.63%	0.96%	0.33%	128	201	73
NAM	2.70%	0.41%	-2.29%	218	34	- 184
NAM	0.22%	2.18%	1.97%	11	126	115
NAM	0.96%	2.88%	1.92%	46	164	118
NAM	2.13%	2.86%	0.73%	105	151	46
NAM	1.99%	3.77%	1.78%	90	203	113
NAM	-0.89%	1.40%	2.29%	- 34	61	95
NAM				5,531	4,475	- 1,056
NAM	0.81%	0.73%	-0.08%	29	27	- 2
NAM	2.34%	-61.72%	-64.06%	86	- 2,602	- 2,688
NAM	0.71%	1.36%	0.65%	19	39	20
NAM	0.00%	0.91%	0.91%	-	20	20
NAM				3,465	4,321	856
NAM Total	4.04%	3.45%	-0.60%	20,600	19,884	- 716

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%)