Additional Information on Data Cleaning Decisions

Prospect Outcome Re-Classification:
- Decided to "place myself in 2020" for the analysis, meaning any customers who were still considering or needed more time to decide technically weren't lost yet. All prospect statuses relating to an ongoing decision process were changed to **'Pending'**. This decision greatly affects the analysis and it made more sense for me to approach it this way.

Band Reclassification:
- CONSIDERED re-doing band classification for all clients based on the following inferred employee counts per band:

employees (ranges inclusive)	Band
1	A / B (could be an error for those in B)
2-4	B
5-15	C1/C2
16-30	D
31-50	E
51 - 250 (roughly)	F
251+	G

- Decided to keep bands as they are and segment those in Band C into two segments (5-9 = C1, 10-15 = C2) which greatly affects the subsequent analysis
- Did this because I didn't have a definitive guideline for employee ranges per band, and there could be reasoning behind Band pricing of certain customers that seem to be outside of the typical employee ranges that I am not aware of

Renewing customers duplicate:
"QR2344", "2020-03-10", 6, "Safecontractor Band C",   "Renewed", "Won", 1,
"300.00", "Band C1"
Removed because:
- Since SafeContractor operates on annual subscriptions, I decided to keep the row with the most recent renewal date as two renewals in a year for a client could signify a corrected entry, mistake, but it could also be the actual case that a client renewed twice in a year. Logic could be flawed here.

Data Model Commentary:

My Data Model in Power BI has:
- Renewing Customers as the central table, Connected to a clean (deduplicated) Supplemental data table (Renewing → Supplemental: 1:1) to ensure ability to aggregate based on categories such as Company Type and Industry Sector without inflated results
- Connected Renewing Customers to Current Live Customers (Only 'Won' customers present in Current Live Customers), giving ability to look at more information relating to Contractors who renewed
- Current Live Customers was deduplicated so this was a 1:1 relationship as well (though deduplication of Current Live Customers may have been the wrong move…)
- Connected Price List to Renewing Customers (1:Many)
- Price list Connection not very necessary besides for filling in values in other tables. If database were more normalized, could just have a dimension price list table