# Project Task Client X– By Eva Santing

**Q1. Who are the 10 most valuable customers for the last 2 years? How would you define most valuable?**

A1. The most valuable customers can be defined by multiple factors. I chose Profit to be the main factor, as it is often the most relevant to a company. Sorting the most profitable customers, gives a total profit of $19.3M profit in 2020 & 2021 combined, which is 41.7% of total profit in this period. 57246571 is the most profitable customer as seen in the table “Total Profit per Customer” (codes are used because names are less clear in this specific dataset).

Additional graphs illustrate the contribution to sales, number of orders per Customer, average profit per order and last transaction date. These Graphs are based on the Top 10 most profitable Customers to further distinguish between their value. They show no strong outliers and appear to follow a more balanced value distribution among all graphs.

57246571 contributed to 20.5% of the total Sales. In “Number of orders per Customer”, it shows only 161 orders and “Avg Profit per Order per Customer” shows $25K, indicating that this Customer orders less frequent, but places larger orders on mid-priced products. Their last transaction was on 29-12-2021. This is a recent sale, indicating they are an active customer. Overall, this customer is highly valuable as it appears in the upper range of all graphs, showing loyalty, high budget and dependency on “X”. It will be interesting to dive deeper into their needs, such as new products or targeted discounts (e.g. volume-based) that could increase the relationship and retention chance with this customer.

Some interesting insights are that customer 877685933 generates the least profit ($0.8M) but has the largest amount of orders (559), with their latest order on 30-12-2021, indicating a stronger return rate, even though they do not contribute as much to profit. 877685933 shows that it is likely a smaller company with strong loyalty to “X”.

Meanwhile, Customer 804419273 has very few orders (29) with a larger profit per order ($59K) and overall Profit ($1.7M). This Customer could have a large budget who buys from more different places, or simply need fewer, more high-priced products. This could be more susceptible to change of distributor/supplier and should be investigated further.

A deep dive into the needs of 877685933 and 804419273 could be beneficial to the relationship and optimize their purchases. For 877685933 it is important to assess if their high order frequency is due to budget or product availability (are they often out of stock and thus need to be ordered more often?). They could benefit from potential bulk discount.

For 804419273 we can investigate if “X” is their sole supplier and if they can be positioned as such. Understanding the products they buy and field they are in, could give insight into what specific product needs they might have.

Finally, there appears to be a good strong basis for the top 10 customers, based on their sale frequency and profitability. Tracking their data in the years after (2022-2025) could show more insights into their long-term contribution as well as the relationship growth. Tracking their sales data in different years or seasons with retention strategies (such as discounts) could be useful to see their effectiveness.

**Q2. What were the sales of the highest selling distributor in July 2021?**

A2. In July 2021, 68.6% of all profit was generated by a distributor we will call OFFYoRm for short. OFFYoRm sold a total of 1668 units, with $5.3 average profit per item results in a total profit of $13K. They mainly sold product 234048 which accounted for $6.8K of the total revenue.

Furthermore, OFFYoRm sold mainly to 3 customers, 968191911 ($6.9K), 64463565 ($4.4K) and 122263023 ($1.3K), which accounts for around 97% of the total sales.

OFFYoRm mainly sells 5 products to 3 different customers with 234048 being their most profitable product ($6.8K). To be able to tell if this is low or high performance, the contribution of other distributors over several months should be compared. Another insight should be created to see if these are returning customers of OFFYoRm, by comparing their past and future purchases.

It will be interesting to see how OFFYoRm is performing in July as comparison to other months or years to see if the distributor is growing. Showing the contribution of OFFYoRm to total sales in comparison with other distributors in 2021 could show more clearly their value to “X”.

**Q3a. What was the impact of Covid on the business? How do you define Covid breakout period?**

A3a. According to an article in Statista (2022), Covid impact in Europe and the USA were mainly felt in 2020, showing significant revenue drops across companies. However, many companies managed to recover in 2021, often exceeding the 2019 sales. In census.gov (2021) it is stated that most restrictions were lifted in the third quarter of 2020, making 2021 open for increased revenue. Therefore, 2020 shall be used as main point for the Covid breakout period.

The impact in 2020 can clearly be seen among the graphs provided. The average unit price dropped by 8.6% in 2020, compared to the 3 years before. Profit in 2020 was $14.1M, 39.2% lower than in the years before. Units sold were also down by 11.9% with 5.4M units sold in 2020. Number of customers in 2020 were 67K, 12.9% lower than in the years before. Only the number of Distributers increased by 9.1% and appears unaffected by covid, with a continuous growth going into 2021 as well.

On the positive side, 2021 shows recovery in most fields such as Profit, showing significant higher values than 2019 ($32.1M vs $22.6M). Unit price increased, surpassing the period of 2017-2019, but quantity sold barely increased compared to 2020, and still showing significantly lower value than 2019. This indicates that the sales are still in recovery, but as the average unit price is higher, profit is not harshly affected by this. More insights could be gained by seeing which products are sold more. It is possible that highly profitable items are sold more, while items that require higher quantities, are sold less.

Customer numbers are still declining, even after 2020, indicating a deeper issue in the customer base that hasn’t recovered from Covid yet. Reasons could be a potential switch in industry or demand in products. The increase in unit price may have driven away smaller customers, resulting in more relevance in larger customers.

Finally, Distributors are the only category showing a steady increase. A reason could be that businesses shifted more toward distributors, which could be caused by their need for reliability or simply availability of direct suppliers.

It will be useful to look into the recovery period after 2021, to get more insights into the duration of the recovery period from Covid. Looking into which products, customers and distributors are driving the recovery will gather more insights into their value and future retention and growth strategies.

https://www.statista.com/statistics/1337028/revenue-recovery-covid-19-industry/ https://www.census.gov/library/stories/2021/06/not-all-industries-experienced-declines-during-pandemic.html?

**Q3b. Was there any specific application level 2 (product category) that was impacted more than the others?**

A3b. The tree map shows the contribution of all categories between 2019-2021, showing Gasketing as the largest contributor to profit with $13M. Looking at the overall impact in the categories (Top 10 most impacted Categories in 2020), shows Structural reinforcement as largest negatively impacted category with -736% in 2020. However, this is a very small category with only a contribution of $1.4K. Therefore, I provided a graph with the 10 most profitable categories and their impact between 2019-2021 for more relevant information.

It shows Flexible bonding and sealing as most impacted category with -54.8%. It managed to bounce back in 2021 with a 70.6% increase with its total contribution being $6.5M. None of the most profitable categories have strong negative outliers and all showed increase in profit in 2021. Potting, Casting and Encapsulating as well as Gasketing increased with 583.3% and 307% respectively in 2021, showing a big rebound.

It will be interesting to dive into why Flexible bonding and sealing was impacted the most. Investigating if the drop is related to the industry in which it is used, can be insightful. For example, it could be related to construction, which was negatively impacted by Covid.

Finally, focusing on the heavily positive impacted categories can give insights as to why this happened and how to take advantage of it by using targeted advertisement, discounts or potential product expansion in these categories.

**Q4. Can you discuss over the data quality and completeness?**

A4. The dataset shows several issues in data quality and completeness

**Column Types:** Multiple columns were assigned incorrect data types, which were corrected  
**Lack of Documentation:** No proper documentation for most columns, making it difficult to understand the data structure.  
**Duplicate and inconsistencies:** in SapMaterialId had duplicate values. I created a SapMaterialID and SapMaterialIdhDescription combined column to make unique string values. Also one duplicate product code was removed from the product data (1.81812E+11).  
**Distributor:** More Distributor IDs than Distributor Names were present, meaning that distributors had multiple IDs assigned.  
**Data mismatch:** Many older sales records (mainly from 2017) referenced products that no longer or never existed in the product file. They were shown as blanks and had to be excluded from the datasets (in product category).  
**Formatting:** Many values were stored in E+11 instead of numerical format, they’ve been converted to strings.  
 Values used . instead of ,, not following proper US punctuation, which was changed.  
 NAICS numbers were stored as a decimal within a string  
**Data integrity:** PricedifferencePercentage column was mainly empty, didn’t have documentation and did not appear to be related to other datapoints, making it irrelevant for processing data.  
 Many products show a negative profit on sales, making some distributors lose money on sales  
**Readability:** Randomly generated names were difficult to read and appeared very similar. Using more simple and distinguishable names should be used for clarity.

Overall, the dataset appeared incomplete and required several corrections and data improvement techniques. However, no additional information was needed to fulfil the requested tasks. Overall, I would score this data 6/10. Doing a few of the above listed implementations when storing and importing data, will already greatly improve the data quality in the future.