Q1: What are the products

1. What are the most frequently bought products?

Quantity	Description	Family
12 047	2.5 L Vanille	Scoop Ice
8730	Cornets d'Amour	Hand Ice
7661	L Vanille	Scoop Ice
4345	Maxi Vanille/27	Hand Ice
3212	Vanille 2.5L +0.5L gratis	Scoop Ice
3107	Cornets d'Amour Vanille/16	Hand Ice
3091	Expo met zachte wafel/12	Indiv dessert
2987	Dessert Dame Blanche/14	Cups
2659	Assortiment XXL	Assortments

All are "Verkoopsartikelen"
All other items can be found in
BestAmountSoldProducts dataframe

2. Which products render the most revenue?

Revenue	Description	Family
€ 89 147.80	2.5L **Vanille**	Scoop Ice
€ 82 935	Cornets d'Amour	Hand Ice
€ 39 619.10	Assortiment XXL	Assortments
€ 36 932.50	Maxi **Vanille**/24	Hand Ice
€ 29 971.50	Big Chocolate/20	Hand Ice
€ 29 516.50	Cornets d'Amour Vanille/16	Hand Ice
€ 27 579.60	1 L **Vanille**	Scoop Ice

€ 27 230.10	5 L **Vanille**	Scoop Ice
€ 26 584.30	Dessert Dame Blanche/14	Cups
€ 23 768.80	**Vanille** 2.5 L +0.5L gratis	Scoop Ice

All are “Verkoopsartikelen”
All other items can be found in **BestRevProducts** dataframe

3. Which products are bought the most in the region of Brussels, Antwerp, ...?

We are gonna give an overview of the 7 most bought products per depot, together with its quantity. The depot in Vilvoorde gives an indication about which products are bought most in the region of Brussels for example

However, before this overview we also give an indication about the quantity of products bought per depots:

Depot Location	**Quantity**
Deinze	34433
Aarschot	12898
Antwerpen	37305
Vilvoorde	61176

We start with Deinze:

Product	**Quantity**	**Family**
2,5 L.Vanille	3527	Scoop Ice
Cornets d'Amour Bres/16	2536	Hand Ice
1 L.Vanille	1915	Scoop Ice
Maxi Vanille/24	1231	Hand Ice
Cornets d'Amour Vanille/16	860	Hand Ice
Vanille 2,5L + 0,5L gratis	741	Scoop Ice
Dessert Dame	688	Cups

Blanche/14		

In the region of Aarschot:

Product	**Quantity**	**Family**
2,5 L.Vanille	1041	Scoop Ice
Cornets d'Amour Bres/16	785	Hand Ice
1 L.Vanille	637	Scoop Ice
Picco Praline/6	450	Individual Dessert
Maxi Vanille/24	436	Hand Ice
Dessert Dame Blanche/14	274	Cups
Assortiment XXL	240	Assortments

In the region of Antwerpen:

Product	**Quantity**	**Family**
2,5 L.Vanille	2785	Scoop Ice
Cornets d'Amour Bres/16	2489	Hand Ice
1 L.Vanille	2025	Scoop Ice
Maxi Vanille/24	1170	Hand Ice
Cornets d'Amour Vanille/16	1100	Hand Ice
Vanille 2,5L + 0,5L gratis	1059	Scoop Ice
Expo met zachte wafel/12	1034	Individual Dessert

In the region of Vilvoorde (Brussels):

Product	**Quantity**	**Family**
2,5 L.Vanille	4992	Scoop Ice

1 L.Vanille	3224	Scoop Ice
Cornets d'Amour Bres/16	3084	Hand Ice
Maxi Vanille/24	1587	Hand Ice
Dessert Dame Blanche/14	1456	Cups
Picco Praline/6	1328	Individual Dessert
Vanille 2,5L + 0,5L gratis	1305	Scoop Ice

4. Are product purchases correlated? Are some products often purchased together?

We will answer both questions separately. First, we look at what products are most often bought together:

Product A	Product B	Frequency
1 L.Vanille	1 L.Mokka	769
2,5 L.Vanille	Cornets d'Amour Bres/16	594
2,5 L.Vanille	2,5 L.Mokka	432
1 L.Vanille	1 L.Chocolade	310
2,5 L.Vanille	Maxi Vanille/24	303
1 L.Vanille	Cornets d'Amour Bres/16	299
1 L.Stracciatella	1 L.Vanille	289
2,5 L.Vanille	2,5 L.Stracciatella	280
Dessert Dame Blanche/14	Cornets d'Amour Bres/16	278
Vanille 2,5L + 0,5L gratis	Cornets d'Amour Bres/16	276

Further, we also look at the which product families are most often bought together:

Family A	Family B	Frequency
Hand Ice	Hand Ice	27944
Scoop Ice	Hand Ice	10759

Hand Ice	Cups	8778
Cups	Cups	4618
Ice Cakes	Ice Cakes	3226
Garniture	Scoop Ice	3104
Scoop Ice	Cups	3017
Individual Dessert	Hand Ice	2453
Individual Dessert	Scoop Ice	1926
Ice Cakes	Meals	1659

Here, we notice that there is one product family that is clearly bought most with other product families, namely Hand Ice. Thus, if we have no idea what to recommend, it is clear that we should recommend the Hand Ice products. Further, most product families are bought most together with a product from its own family. This gives the seller some basic indications of what to recommend if he has to think quickly.

Next, we also looked at the product purchase correlations, as we can gain some insight to what products are often bought together, independent on the volume.

Product A	**Product B**	**Correlation**
2x2,5L Caramelo Crunch (Caddy)	2x2,5L Malaga (caddy)	0.666641
5 L.Aardbei Extra (Caddy)	2x 2,5LRio (caddy)	0.629788
2x2,5L Caramelo Crunch (Caddy)	2x2,5LSorbet Framboos (Caddy)	0.577336
2x2,5LSorbet Framboos (Caddy)	2x2,5L Malaga (caddy)	0.577336
2x2,5L Caramelo Crunch (Caddy)	2x2,5L Cappucino(caddy)	0.577314
2x2,5L Speculoos (Caddy)	2x2,5L Stracciatella(caddy)	0.577259
2x2,5L Malaga	2x 2,5LRio (caddy)	0.561905

(caddy)		
2x2,5L Cappucino(caddy)	2x2,5L Stracciatella(caddy)	0.553336
2x2,5L Speculoos (Caddy)	2x2,5L Cappucino(caddy)	0.547664
2x2,5LCitroen (caddy)	5 L.Aardbei Extra (Caddy)	0.539204

5. What are the total sales generated for each product family?

Family	Revenue
Hand Ice	399 042.20
Scoop Ice	279 235.80
Cups	110 634.70
Ice Cakes	62 858.50
Assortements	61 107.90
Individual desserts	54 172.20
Meals	39 075.60
Coffee	31 738.70
Garniture	13203.10
BIG	10 240.00
Coupon	2433.00
Pos inside	708.10

This is the complete **BestRevFamily** dataframe

6. Does the weather / seasonal changes have an effect on the total revenue

If we have a look at the share for the whole revenue, for each season

Autumn	23.12%
Spring	29.36%
Summer	29.24%
Winter	18.28%

We can clearly see that the most is sold in Spring and

	Summer. This doesn't have to be surprising since we are a dessert company overall. People tend to buy more desserts/ice creams etc in hotter temperatures. We have a big drop in revenue: about 60% of the revenue are from Spring and Summer, but only 18% of the revenue is made in the Winter. We clearly have a seasonal and weather effect on revenue. Higher temperatures = higher revenues. When looking deeper into the family sales between seasons, we see that the same family products stay on top. Each season has the same patterns but there is a change in Autumn! In Autumn we have a high rise in IceCakes and Meals also risealot **Conclusion: There is a seasonal effect: Spring and Summer have more sales. Very little sales in the Winter. Higher Temp = more sales** **Downscale in winter. No patterns found between seasons: Hand ice stays the most popular along with scoop ice. But in Autumn Ice cakes and meals rise alot higher.**

Q2: Who are the customers?

1. What do the customers buy?	We decided to skip this question: We already answered which are the most sold products in Q1. This is already an answer on what customers buy. Further exploratory insights on the customers themselves and associating this with buying patterns with certain products can be found in further questions from Q2 onwards.
2. Which customers left the company?	Customers that have left the company are defined as a churner. In our analysis we see a churner as a customer that has not made any purchases for one year (365 days) or longer. The reason for this timeframe is to subdue seasonal effects, since we clearly have seen this effect on revenue earlier. Given the fact that our database is made on a two year span, we can not go any higher for our churning span as this would be larger than 50% of our database. Based on the last time a customer made a purchase, and that time being more or less than 365 days ago, we were able to define our churners. Some customers are neither in the churned or non churned database, these customers have not bought anything in the past two

	years so no decision could be made. In total we have 5079 Customers who have bought items. Of these customers by our churner definition we have 152 churned customers and 5460 nonchurner 5079+152 != 5460 (Because not having bought = not included) customers **Conclusion: df_Churnedcustomers and df_nonChurnedcustomers show us which customer churned or didn't.**
3. Which customers have the highest CLV	We defined CLV as the amount spent on products over the complete lifetime of a customer. To calculate CLV we have to sum all amounts spent over all visits of a single customer. We created a separate dataframe containing all customers with their corresponding CLV. The top 5 customer CLV's are: 14 211,50 / 9938,90 / 8899,80 / 8503,20 / 7041,70 In later questions we go deeper into customer CLV and use this dataframe for further analysis. **Conclusion: df_CLV contains all this data**
4. What is the relationship between leaving the company and buying patterns?	Firstly we just have look at the differences in normal spending amounts compared to factors. When we compare based on seasonality we get a nice insight: 72.62% of the revenue from Churners are generated in Spring compared to 31 % from nonChurners. Also Churners barely buy in the summer: 17% share of total revenue over all seasons. Non churners have 31% so this is quite a big decrease. Surprisingly churners only buy 0.87% during Autumn while nonChurners have a share of 21% during Autumn. So we notice big changes. Also Churners buy half as much in the winter When it comes to weekend/day: Churners buy slightly more during weekends, but not significant enough. When it comes to the Family: Churners buy double the amount of Coffee as non churners and spend more on Cups (17% vs 11%) and Ice cakes other notable changes are that they buy less meals Further separations are not deemed significant **Halfway Conclusion: Churners buy close to exclusively during Spring (63% and Summer 25%), but close to nothing in Autumn and a whole lot less during the winter.**

Churners buy a lot more Ice cakes, but less Hand Ice and Scoop Ice. They also buy more cups.
We go deeper:

Several buying patterns or customer properties that we calculated are:

- average number of purchases per customer
- average amount spent per purchase per customer
- average spending per month per customer
- average time between purchases per customer
- reviews of customer

When we state a variable as an ‘average’ we calculate the properties of this customer divided by it’s active time. This active time is defined as the time between the last and first purchase in our database. So a customer that bought something in January and December is 12 months active, a customer that bought something in January and February is only active for one month. This is done to normalize the data.

By plotting this information we can see that

- Customers who **purchase less than once a month or once a month** on average are more likely to churn
- Customer who don’t buy something for a very long time OR customers that buy something on a **30 to 60 day average between purchases** are more likely to churn
- Customers that spent **less than 200** on average **per purchase**
- Customer that spend **less than 350** on average **per month**

- Customers that have **no opinion, are not completely satisfied or are very unsatisfied lead to 94.24% of the churners**. (67 no op, 62 not comp satisfied, 2 very unsatisfied out of 139 churners).
(Need to upload data!)

<table>
<tr>
<td>5. Are there clusters of customers? How would you describe these different clusters?</td>
<td>Geographically speaking we can define three or four clusters centered around Ghent, Antwerp and
- Kortenberg if we look at three clusters
- Brussels and Leuven/Holsbeek if we look at four clusters

Based on the type of customer we can’t really identify a cluster, however we do see the proportion of catering to horeca in the region leuven is higher than anywhere else in Belgium.

Based on reviews we can see a cluster of (very) satisfied customers in the region of Ghent, Brussels and even Leuven, with more of lower scoring review in the region between Antwerp, Brussels and Leuven. Overall Ghent and Brussels score very well A strong management question could be to really try and get an opinion as the customers opinion is very valuable.</td>
</tr>
<tr>
<td>6. Do customers have different buying patterns during the weekend?</td>
<td>Total sales during 5 weekdays: 693 761.90
during sat and sun 370 687.90

So you sell way more on a normal weekendday than a weekday: (693 761.90/5) < (370 687.90/2)

We then explored if there were different buying patterns in buying patterns for products, families etc. We looked at relative comparisons. No significant changes were found.

Conclusion: Weekend = more sales than during the workweek. No other sign changes in buying behaviour</td>
</tr>
<tr>
<td>7. Do customer sales differ across different cities? Is there a relationship between customer sales and average income per inhabitant (and other factors)?</td>
<td>Firstly we looked at different regions CLV and found the following:
<table>
<tr><td>B02</td><td>290 525.55</td></tr>
<tr><td>B25</td><td>318 953.00</td></tr>
<tr><td>C04</td><td>226 554.05</td></tr>
<tr><td>C17</td><td>228 417.20</td></tr>
</table>
Using this we can see that the cities located in region B25 and B02 have the highest Rev and CLV. There is a big difference in the C region having a CLV drop by nearly 80 000, which is a huge difference. We can conclude there is a big difference between cities. This might be related to the depots as well.</td>
</tr>
</table>

If we have a look at CLV by Depot we find the following:

Aarschot	92 116.50
Antwerpen	286 587.15
Deinze	259 064.85
Vilvoorde	449 843.50

Here we also notice big differences. Vilvoorde is by far the best performer, with a shared second place to Antwerp and Deinze. Aarschot drastically underperforms however. Clearly there are differences between cities, regions and depots!

We then had a look at a new dataset showing us the average and median income for cities and tried to link this to our analysis. By performing linear regression however no significant effects were found for average as median income so we omit this hypothesis that these would influence customer behaviour.

Conclusion: There are differences regionally. B02 and B25 highly outperform C04 and C17. For the Depots Vilvoorde has enormous CLV. Antwerp and Deinze ok about the same, but Aarschot has really low CLV and should be evaluated.

The relationship between income and sales is not easy to determine due to the fact that we don't have a clear dataset about income based on region. However we were able to plot our customers based on the CLV where we split the CLV up in several stages:

Extremely Low CLV	CLV<100
Very Low CLV	100<CLV<200
Low CLV	200<CLV<300
Medium CLV	300<CLV<400
Good CLV	400<CLV<500
High CLV	CLV>500

By inspecting the elements we can firstly conclude that there is a very large overlap of all the layers. On first sight however we can see more blue spots in brussel than green spots, unlike Ghent where green spots are in

<table>
<tr><td></td><td>the majority. This gives us a first indication that higher income locations such as brussel increase the CLV.

If we go further after skimming the lower layers and transitioning to the higher layers we can start to see a pattern emerge. Most of the higher CLV’s have a tendency to be located closer to the city centre of a (larger) city. For each level where we increase the CLV we can see most of the values receding towards a city(centre).

A good example would be Ghent and Leuven for example. All of the customers are spread out however the biggest CLV’s are located a lot closer to the centre of Ghent and Leuven than the lower CLV’s.

Most of the highest CLV’s are located on top of a named city, this gives us a strong indication that locations with higher income (such as cities.</td></tr>
</table>

Q3: Who are the employees

For more insights about the employees, we would like to refer to our code and presentation. Because we have also found some insights about where the employees live and about the languages that they speak.

<table>
<tr><td>1. What are the routes of the employees?</td><td>We created a dataframe that has in one column the Route_ID’s and in the other one the Employee_ID’s. In this way you can see which employees are connected to which routes and the other way around.

<table>
<tr><th>Employee_ID</th><th>RouteTemplate_ID</th></tr>
<tr><td>18503648417252</td><td>289661606.0</td></tr>
<tr><td>18821748667448</td><td>219019324.0</td></tr>
<tr><td>18934048107254</td><td>289413943.0</td></tr>
<tr><td>18934048107254</td><td>289657598.0</td></tr>
<tr><td>18934048107254</td><td>289414945.0</td></tr>
<tr><td>18934048107254</td><td>289416949.0</td></tr>
<tr><td>18934048107254</td><td>289412941.0</td></tr>
<tr><td>18934048107254</td><td>289418953.0</td></tr>
<tr><td>18934048107254</td><td>289419955.0</td></tr>
<tr><td>18934048107254</td><td>289411939.0</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>...</td></tr>
<tr><td>2. What is the turnover for each employee?</td><td>For all the years:

Name	Turnover
VAN HECKE jan	254484.50
PEETERS michel	238142.45
BUFFEL Sandy	168118.15
VERRESEN Dirk	84362.55
VERSTRAETE Erik	48427.65
MATTIJS stijn	43707.65
DE PAUW Kim	41163.30
KERCKAERT Pieter	18551.55
CLAUS David	14980.15
BALLINGS Hendrik	8552.20
TACK Hann	5974.15
VAN HUFFEL Jan	4781.00

…
For 2020:

Name	Turnover
PEETERS michel	130312.90
VAN HECKE jan	121714.80
BUFFEL Sandy	106949.10
VERSTRAETE Erik	48427.65
DE PAUW Kim	29904.30
CLAUS David	14980.15
TACK Hann	5974.15
VAN HUFFEL Jan	4781.00
PAUWELS Jeffrey	3588.85
KERCKAERT Pieter	2174.85
DE BRUYNE Chris	484.40
DE MILT Tomas	120.60
JACOBS Michel	24.20
STEEMAN Evy	22.75

We also calculated the sales rate. The</td></tr>
</table>

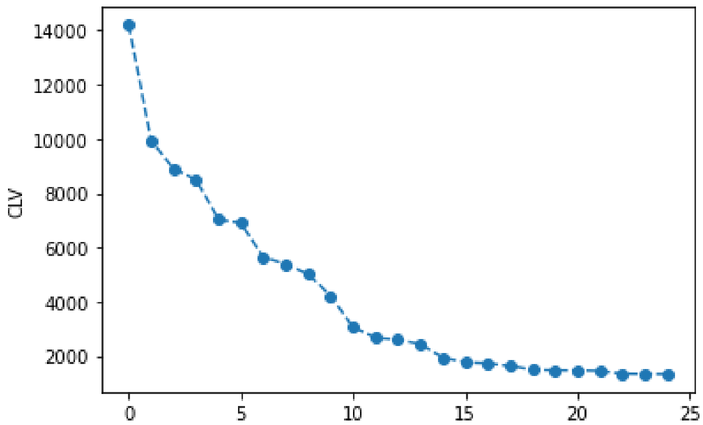
	sales rate is the number of visits that have an outcome of 2 (sale happened) divided by all the visits. However it is also important to keep in mind the turnover in terms of money too. Since the best performing employees in terms of turnover, only have a rate betwen 30 and 40 %. To conclude you should encourage your employees to do as many visits as possible in a good way without panicking about the sales rate, since this is very normal that it is not much higher than 40%.

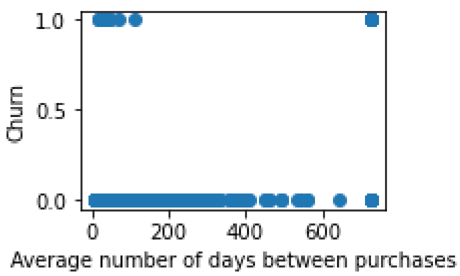
Q4: What are the routes

1. How are the customers divided into regions?	Each customer has its own unique complementary variables such as customer id, route id, postcode and region. By looking at every single customer ID and trying to find a pattern we can see that every single region corresponds to a select number of postal codes. Indeed after plotting the postal codes we can differentiate four distinct regions which correspond to the four regions. **Conclusion: regions are based on postal code.**
2. Which customers are assigned to which routes?	Every customer has its own unique route id, so this means that a customer is only visited by one route and not multiple routes. After comparing databases we were able to pinpoint which exact route each customer is referenced to. One step further is plotting on our map of Belgium of every customer, however we give each group of customers a different colour for visual reference. **Conclusion: the database df_CRV shows each route corresponding with a customer, furthermore a map colored with different routes visualizes this data.**
3. Which routes are assigned to which depot	There are four possible depots present; Deinze, Aarschot, Antwerpen and Vilvoorde. Each route is assigned to one specific depot. This means that the route starts and ends at this depot when the employee does his route. A comparison of databases shows us every unique route and which depot it is linked to. One step further is plotting all the customer locations

	color coded based on which depot their delivery comes from. **Conclusion: the database df_customer_Route_Depot shows us an overview of all routes and their assigned depots, furthermore a map colored with the depot property visualizes this data.**

Q5: How can the company improve its service

1. Which customers should be rewarded?	You should reward the customers with high CLV. Customer retention is very important. These customers buy a lot frequently leading to high CLV, so losing these customers could lead to a huge change in sales. In total we have 5709 customers. We firstly check if there is a lot of difference between customer CLV  Looking at the graph we can see that the best CLV customers outshine all others by a big margin. We notice an elbow after the top 10 best customers. Here CLV stabilizes/ converges. These top customers all are Dutch and are Horeca and Catering customer types. If we consider the total CLV 1 064 449.8. The best performing customer holds a 1.335% share in this. having the top 10: they share 7.12 % of the total CLV. So it is clear these customers are very valuable and important. Using other numbers we find that these 10 customers represented by 0.18 % of all your customers already have this huge CLV share. Not even a half percentage of the customers gives a 7.12% share of CLV. This is quite insane. So extra attention should go to these customers. The first private customer is ranked at place 15 for

	CLV. This is quite a high rank for a non caterer/horeca customer. Rewarding private customers is a lot less expensive,but a good practice. Not being as big, it is normal private customers do not have a CLV as high. Our recommendation would be to reward the top 10/15 Horeca/Catering businesses since these proportionally bring in the most CLV shown by the elbow on the curve Private customers should be rewarded as well. As it is not as expensive for reward programs, we recommend rewarding the top 50 best CLV private customers. After the top 50 best CLV private customers CLV quickly diverges, so rewarding these people is not as important. these can companies and private people to be rewarded can be found in the upgraded df_CustomerType and df_RewardFramePrivate respectively Also important of note. If we group this data we find that combining the top 10 horeca/catering customers and the top 40 private customers compromise about 1 percent of your total customers but have a share in the CLV of a staggering 19%! You could also argue that pre-churners should be rewarded. Customers that show behaviour specific to churning.  This plot shows churners buy quickly and then burn-out. It might be a good idea to define an underbound and reward customers that frequently and quickly buy many products to prevent churning and reward them to keep them in the company. **Conclusion:** **Reward the top 15 companies** **Reward the top 50 private customers since they have a big share in CLV** **Reward quick buying patterns to prevent churning**
2. Which employees should be	For this you could take in the first place into account

<table>
<tr>
<td>rewarded?</td>
<td>the turnover of the employees, more specifically of the last year. (see question 3) But since that the reviews are a very important indicator of churners and customer satisfaction, we also take into account the number of good/bad reviews the employees have had. It is important to reward also those customers with a high percentage of good/excellent reviews in relation to all the reviews and to train the employees with a high percentage of bad/terrible reviews in relation to all the reviews. It could also be interesting to dig deeper into why these employees have these bad reviews. One could for example go talk to those customers who have had bad experiences with those employees and ask what exactly made them feel that way, what could be done better, etc. In this way you don’t necessarily just punish your employees, but you actually improve their performances!
For the dataframe containing these percentages and counts of reviews, you can see a screenshot below:
<table>
<tr><th>Name</th><th>Count_no_opinions</th><th>Count_good</th><th>Count_bad</th><th>Count_excellent</th><th>Count_terrible</th><th>Percentage_bad&terrible_over_all</th><th>Percentage_excellent&good_over_all</th></tr>
<tr><td>VERSTRAETE Erik</td><td>5701</td><td>8618.0</td><td>611.0</td><td>2248.0</td><td>50.0</td><td>3.836777339215231</td><td>63.07174367309032</td></tr>
<tr><td>DE PAUW Kim</td><td>4930</td><td>7557.0</td><td>536.0</td><td>1776.0</td><td>41.0</td><td>3.888140161725068</td><td>62.89083557951483</td></tr>
<tr><td>VAN HUFFEL Jan</td><td>287</td><td>432.0</td><td>34.0</td><td>113.0</td><td>3.0</td><td>4.25776754890679</td><td>62.715765247410815</td></tr>
<tr><td>VERRESEN Dirk</td><td>7034</td><td>10272.0</td><td>871.0</td><td>2391.0</td><td>54.0</td><td>4.485500921346135</td><td>61.40529531568228</td></tr>
<tr><td>BALLINGS Hendrik</td><td>523</td><td>729.0</td><td>40.0</td><td>126.0</td><td>3.0</td><td>3.0260380014074597</td><td>60.1688951442646</td></tr>
<tr><td>TACK Hann</td><td>544</td><td>760.0</td><td>81.0</td><td>155.0</td><td>8.0</td><td>5.749354005167959</td><td>59.10852713178295</td></tr>
<tr><td>MATTIJS stijn</td><td>3196</td><td>4111.0</td><td>289.0</td><td>512.0</td><td>21.0</td><td>3.813507196457129</td><td>56.870463771681635</td></tr>
<tr><td>BUFFEL Sandy</td><td>15146</td><td>18939.0</td><td>1905.0</td><td>3537.0</td><td>177.0</td><td>5.243804150715293</td><td>56.60890590368729</td></tr>
<tr><td>PAUWELS Jeffrey</td><td>398</td><td>531.0</td><td>70.0</td><td>83.0</td><td>9.0</td><td>7.241063244729606</td><td>56.278643446379476</td></tr>
<tr><td>KERCKAERT Pieter</td><td>2088</td><td>2627.0</td><td>396.0</td><td>532.0</td><td>48.0</td><td>7.801792303637322</td><td>55.5086979441223</td></tr>
<tr><td>DE BRUYNE Charlotte</td><td>222</td><td>202.0</td><td>31.0</td><td>37.0</td><td>3.0</td><td>6.8686868686868685</td><td>48.282828282828284</td></tr>
<tr><td>CLAUS David</td><td>1823</td><td>1345.0</td><td>632.0</td><td>223.0</td><td>80.0</td><td>17.353156227150865</td><td>38.21593955642213</td></tr>
<tr><td>JANNSENS Eddy</td><td>334</td><td>221.0</td><td>133.0</td><td>33.0</td><td>22.0</td><td>20.861372812920592</td><td>34.185733512786</td></tr>
</table>
</td>
</tr>
<tr>
<td>3. To which customers should the company send coupons in order to win them back?</td>
<td>As stated in question Q2.b we have defined our churners to be customers that are no longer active for one year or longer. These people can be potential customers to try and win back.

As stated in Q2.d customers that have not yet churned (so last purchase was < 365 days ago) but are likely to churn could be sent a coupon as well. These are customers who on average only make a purchase once or less a month, have around 30-60 days between purchases, spend less than 375 per purchase, spend on average less than 400 a month or have left no or a bad review.
As found earlier Churners also tend to spend a lot during the Spring but not during other seasons. It might be a good idea to put up promotions during the Spring, to get churners to get convinced to get dragged over to other seasons and boost sales as a whole.</td>
</tr>
<tr>
<td>4. Are there factors that the company can change in order to decrease the churning rate?</td>
<td>Q2.4 clearly states the following facts about churners:
1) Halfway Conclusion: Churners buy close to exclusively during Spring (63% and Summer 25%), but close to nothing in Autumn and a whole lot less during the winter.
2) Churners buy a lot more Ice cakes, but less Hand Ice and Scoop Ice. They also buy more cups.</td>
</tr>
</table>

	This clearly shows the seasonal behaviour of churners. We recommend that churners should be made more aware of products for other seasons. Besides, some extra marketing during Autumn and the Winter might be necessary to entice these customers in buying our products more regularly. The fact that the core products are not often bought by churners gives an opportunity to upsell these products to the churners. Next, Q2.4 also states the following: 1) Customers who **purchase less than once a month or once a month** on average are more likely to churn 2) Customers who don't buy something for a very long time OR customers that buy something on a **30 to 60 day average between purchases** are more likely to churn. This indicates that churners should be enticed to buy the products more regularly. This could be achieved by flyers or customized ads on the internet. Finally, we also found the following result: 1) Customers that have **no opinion, are not completely satisfied or are very unsatisfied lead to 94.24% of the churners**. (67 no op, 62 not comp satisfied, 2 very unsatisfied out of 139 churners). We would recommend the company to gain more feedback from these customers and address their issues. This would not only gain the customer back, but improve the services for future customers.
5. Would it be valuable to recommend (upsell / cross sell) products to a customer?	It would be valuable to recommend certain products to a customer. In order to answer this question, we constructed a collaborative filtering analysis. Through this analysis, we are going to be able to give customer based product recommendations based on the history of the customer and the purchase history of other (similar) customers. Since these customers have bought similar items in the past, we will assume that the target customer has a high probability of buying the items he or she has not bought, but similar clients have. Now, what does this mean for the management of the company. We look at the effect of these product

	recommendations for the company. Suppose, we only look at the sales of 2020: 1) So they were 40133 sales made in 2020, good for a revenue of 524.402,3 euros 2) The mean amount for such a sale was on average 13,06 euros 3) Suppose we can increase the amount of the sales on average by 10 percent 4) This is a very conservative assumption as the average price of the products is 7,00 euros. This means that only two extra products have to be bought per 10 sales. This seems likely as the collaborative filtering analysis proposes products that have a high probability of being liked by the targeted customer. **If it would be possible to fulfill these assumptions, this strategy would yield an extra revenue of 40 133 * 1,306 euros = 52.413,70 euros euros or an increase of revenue by 10% each year.**
6. Which employees should be assigned to different routes?	We can divide this question into 2 parts. Since every employee first has to pass by a depot before going to the customer, we should optimize 2 things: 1) the distance of the employee to the depot & 2) the distance of the route to the depot. 1) will be solved in this while keeping in mind that the second part is *already optimized*. In the next question you can see how exactly this is done. To conclude, once you have the optimal depots for the employees (in terms of distance and equal number of km's they have to travel with regard to their colleagues), they will automatically be assigned to the optimal routes, since the routes' distances to that depot also are optimized. First we computed the distances of all the employees to all the different depots, and then we also added a column about how many times that employee already passed that depot in the past. As you can see, you can easily define for every customer, which depot is shorter or how you can better shift between the depots. Once a depot is chosen, an employee can choose from a number of optimal routes connected to that depot, which will be discussed in next question. A little screenshot from the dataframe you can see below:

Employee_ID	DEPOT	Count	Distance
18503648417252	Vilvoorde	21.0	126.72410838032812
18503648417252	Aarschot		83.37346557766134
18503648417252	Antwerpen		134.29204725783273
18503648417252	Deinze		227.4844205555618
18821748667448	Deinze	98.0	154.91125718000356
18821748667448	Aarschot		29.69513592409581
18821748667448	Antwerpen		55.29557890303353
18821748667448	Vilvoorde		62.20789474047806
18934048107254	Aarschot	61.0	50.556378956533216
18934048107254	Deinze	5833.0	194.1301398509361
18934048107254	Vilvoorde	48498.0	94.52184030000973
18934048107254	Antwerpen		98.340873759604
19902260358040	Vilvoorde	5.0	81.0279751146561
19902260358040	Aarschot		39.20469386147122
19902260358040	Deinze		177.978385901039
19902260358040	Antwerpen		79.70141959732516

7. Which routes should be reassigned to different depots?

The location of the depots are unfortunately unknown, however we can assume the coördinates of their location's city to be an accurate estimate. We looked at two different approaches for determining which routes should be assigned differently.

First the distance between every customer and its unique depot was calculated, if this distance exceeded a threshold (e.g. it was too far) then this route should be reassigned.
Next for an even more accurate result we plotted all our routes in different colors per depot, this allowed us to very easily see which routes were generally far away in from their depots or which routes were not clustered very well (e.g. they were spread out)

<u>Depot Deinze:</u>
Route 289414945 and route 289757472 are located very far away. Route 289414945 should switch towards Vilvoorde and

Route 289757472 should switch towards Antwerp.
70 km savings leads to **684.68 euro savings**

Route 289414945 should switch to Vilvoorde. 50 km savings leads to **489.06 euro savings**

<u>Depot Antwerpen:</u>

	Route 219018687 is located very far away, this route should switch towards Deinze. 10 km leads to **97.81 euro savings** Route 289756470 is located closer to Vilvoorde and can potentially switch as well. 10 km leads to **97.81 euro savings** Route 289755468 is located closer to Aarschot and should switch as well. 15 km leads to **146.72 euro savings** Other routes are a bit closer in euclidean distance but there are no major highways connecting them so getting there would be slower. **<u>Depot Vilvoorde:</u>** Most of the routes are located very close to Vilvoorde, however some of them are located practically next to the depot in Aarschot, routes 289656596, 289659602, 289761480 and 289661606 should relocate to Aarschot. 15 km per route leads to 60 km savings or **586.87 euro** Routes 289664612 and routes 289660604 are also almost as close or closer in euclidean distance but not necessarily faster. **<u>Depot Aarschot:</u>** Routes 289663610, 289662608, 289658600, 289657598 are all located very close to Aarschot, one can argue that it would be more efficient to have a more centralized depot if necessary in Vilvoorde but still these routes are assigned to the right depot. **2102.9 euro of fuel savings a year** https://www.webfleet.com/nl_nl/webfleet/blog/hoeveel-diesel-verbruikt-een-vrachtwagen-per-kilometer/ https://carbu.com/belgie/index.php/diesel
8. Which customers should be reassigned to different routes?	Customers that should be assigned a different route are customers that are either: - located very far away from the depot that their route is linked to - customers that are closer to the center of a cluster of a different route than to the center of their own cluster For possibility one we have: - All the customers belonging to route 289414945 (**originally Deinze**), they should be

	assigned a route from Vilvoorde, possibly split according to route 289413943, 289411393 and 289412941 since these routes are located very close by. - All the customers belonging to the route 289757472 (**originally Deinze**) should be assigned a route from Antwerp; no existing routes give a great match. - All the customers belonging to route 219018687 (**originally Antwerp**), should be assigned a route from Deinze, possibly split according to route 219019324 and 219024456. - All the customers belonging to route 289756470 (**originally Antwerp**), no existing routes give a great match. - All the customers belonging to route 289755468 (**originally Antwerp**), no existing routes give a great match. - All the customers from routes 289656596, 289659602, 289761480 and 289661606 (originally Vilvoorde) should be assigned a route from Aarschot, no existing routes are a good match. For possibility two we have: - Customer **from depot Vilvoorde** with route number 289656596 located near Heverlee, reassign to a more local route (such as 289418953) - Customers **from depot Vilvoorde** with route number 289419955 located at the south-west of Brussels, reassign to a more local route (such as 289411939 and 289416949) - Customer **from depot Deinze** with route number 219023811 located below Moerbeke, reassign to a more local route (such as 219019324) - Customer **from depot Deinze** with route number 21902445 located to the right of the Pinte, reassign to a more local route (such as 219021241) - Several customers **from depot Antwerp** with route number 289762482 located in the north of Antwerp and south-east can be reassigned to routes 28975468 and 289764486.
9. Which depots should be removed? Where should the company create new depots?	Remove/move: - The depot in Deinze is rather skewed to the south-west, a more central location could relatively shorten the travel time/distance for at least three of the routes. Savings of around 20km per route 219019962, 219019324 or

	219024456 could be motivating, however the closer to Ghent the more expensive. - Same argument goes for Antwerp, the depot is located to the left of the cluster, shifting towards the right can increase fuel efficiency, however a move like that can be very pricey. - A new depot is not really required, especially not if we reassign the right routes to the right depots. Suppose we do want to decrease the pressure on existing depots: a new depot could potentially be located in the centre between Antwerp - Vilvoorde - Aarschot OR a new depot could be opened in the south of Belgium with eyes on expanding. Number of customers per depot location: (see table below) Number of products per depot location: (see table below)
10. Which products should be added / removed from depots?	**Artur**
11. Does customer satisfaction relate to different factors? Can the company respond to these factors?	On average our company scores 6.27/10 in rescaled reviews with as std dev of 2.034. The first factor we looked at was the region. All

Number of customers per depot

Depot location	Customers (nr of)
Deinze	1481
Aarschot	525
Antwerpen	1328
Vilvoorde	1961

location:

Number of products per depot location:

Depot location	Products (nr of)
Deinze	290
Aarschot	261
Antwerpen	293
Vilvoorde	284

	regions were quite close except for region B25. B25 scored a 5.23/10. This is an odd find since B25 is also the region with the highest noted CLV. So clearly B25 has some issues since the variance of the region reviews are also extremely high. It needs further exploration, but might have to do with the location of the depot and **distance for delivery.** Next we looked at the spoken language. But our subset was too small as only 5 reviews from French speakers were found Dividing the reviews into customer types also did not lead to much insight. They show the same characteristics. Also catering and horeca are rather small with only 24 and 74 reviews respectively. So making an accurate comparison might be ill advised. **Conclusion: region B25 is very mixed and variating**