

# Customer Analytics and Customer Insights

## WS19/20 – Final Assignment

*Anna Franziska Bothe (576309)*

The key for successful targeting is the knowledge about the customers' behaviour as well as their preferences. Based on this knowledge, optimal pricing strategies can be determined and thus, the profit can be maximized. To understand and predict the customers' behaviour, a data analysis is necessary.

The underlying data set was collected within the seminar "Customer Analytics and Customer Insights" in 2019/20. The survey deals with the preferences regarding the purchase of 10 different jeans with a focus on Levi's, Tommy Hilfiger, Guess and a mystery product. After data cleaning, in total 780 respondents participated. For the analysis only a subset of 650 respondents is selected – depending on the matrikel number functioning as an unique ID.

### **Cluster Analysis**

Since it is very expensive to target each customer individually, a prerequisite for effective targeting is the detection of customer groups that behave similarly and have the same preferences. In order to find these groups a hierarchical cluster analysis with help of the ward (Ward.D2) algorithm is conducted. Based on the dendrogram and an elbow plot, three clusters are chosen. While trying to maximize the homogeneity within groups, three clusters seem to be a reasonable amount of target groups for developing marketing targeting campaigns.

The 650 observations are split into groups of 136, 180 and 334 respondents. The following table shows the mean part-worths and the importance per attribute of each cluster as well as the willingness-to-pay (WTP) for all non-price attributes levels in each cluster:

Attributes	Cluster 1 (136 respondents)			Cluster 2 (180 respondents)			Cluster 1 (334 respondents)		
	PW	IMP	WTP	PW	IMP	WTP	PW	IMP	WTP
Price		<b>55.59</b>			24.49			29.80	
Mystery	-0.16	7.18	-2.02	-0.31	8.76	-14.23	-0.46	<b>31.56</b>	-33.58
Hilfiger	0.23		3.15	-0.06		-10.59	0.19		-20.03
Guess	-0.06		-0.67	-0.32		-14.31	-0.88		<b>-42.36</b>
Levi's	-0.01		0	0.69		0	1.14		0
Straight	0.59	24.23	-0.37	0.66	<b>62.24</b>	-51.12	1.07	31.47	<b>12.40</b>
Boot Cut	-0.50		-14.73	-2.03		-89.45	-0.61		-22.77
Wide	-0.70		-17.44	-2.89		-101.67	-0.94		-29.84
Skinny	0.61		0	4.26		0	0.48		0
Black	0.41	13.00	<b>6.91</b>	0.16	4.52	<b>-0.25</b>	0.17	7.16	1.28
Light Blue	-0.30		-2.44	-0.34		-7.38	-0.29		-8.33
Blue	-0.11		0	0.18		0	0.11		0

Table 1: Results of cluster analysis (PW = part-Worth, IMP = importance, WTP = willingness-to-pay)

As observed, group 1 cares a lot about the price which is the most important feature. It is more than double as important as for the other groups. Brand is the least important attribute. However, they prefer – as the only group – Hilfiger over Levi's.

In contrast, the brand as well as the fit are almost equally important for group 3. The importance for the brand is about four to five times higher as for the other groups. They would pay about 43 Euro less for a Guess than for a Levi's jeans. Levi's jeans are strongly preferred over any other brand.

Group 2 does not differentiate a lot between the mystery product and the jeans brands. However, the brand is not that important for this group anyway. The fit is highly important though. The importance of the fit for group 2 is double as high as for any other group. They dislike wide and boot cut jeans a lot; they would pay about 102 Euro less for a wide than for a skinny jeans. This is by far the biggest range of the WTP between two attribute levels.

Group 1 does not differentiate a lot between straight and skinny fit. Group 3 even prefers straight legs over a skinny fitting jeans. Nonetheless, all groups dislike the wide or boot cut jeans.

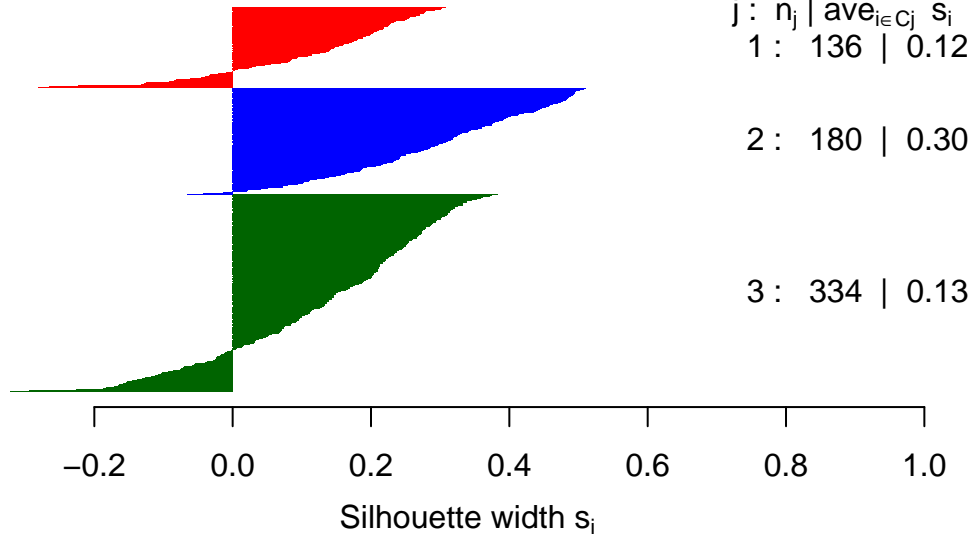
Group 1 and 3 prefer black jeans slightly over blue jeans. However, Group 2 prefers

blue jeans even though they also like black jeans. The color is the least important feature for group 2 and 3.

The silhouette width gives an indication of how close a data point is to the neighboring clusters data points. The distance can be from -1 (data point might be assigned to the wrong cluster) to 1 (perfectly separated data point). The value 0 indicates that a data point is very close to the decision boundary whether it belongs one or another cluster.

### Silhouette Plot

n = 650



Average silhouette width : 0.17

Figure 1: The silhouette plot indicates how well the observations are clustered

As observed in the silhouette plot, cluster 1 and 3 contain observations that probably belong to another cluster. Cluster 2 has the highest average silhouette width with 0.3. The negative values of cluster 1 and 3 obviously have an impact on their average silhouette width of 0.12 and 0.13 which is quite low.

The average distance between clusters is 4.70 and the average distance within clusters 3.54 which also shows that the clusters are very similar to each other.

An explanation for the unfavorable results might be the homogeneity among respondents. More than 88% of the respondents are 30 or younger. At the same time, the education level is very high with about 77% having a Bachelor, Master or PhD. About 88% come either from an Asian or western culture. In general, all are optimistic

people. Characteristics are highly correlated with preferences and behaviour. Hence, it is not surprising that the clusters are very homogeneous.

However, it is still possible to characterize the cluster:

Cluster	No.	Characteristics	Synonyms
1	136	Very young and some older respondents, in average half a year older than cluster 2, price sensitive, low brand awareness, high purchase probability, like Hilfiger, least optimistic	Bargain Hunters
2	180	In average youngest group with about 62% females and 74% students, interested in fashion, most optimistic about the future	Fashionistas
3	334	Oldest group, only group with slightly more males, know exactly what they want (Levi's and Lee), highest income on average, most optimistic about the past	Originals

Table 2: Characteristics of the clusters 1 to 3 formed by the conducted cluster analysis

The respondents of the first cluster consider Hilfiger as slightly more credible than Levi's which it is in line with the results of the higher WTP for a Hilfiger than for a Levi's jeans. The awareness of jeans brands is compared to other group almost always the lowest and almost 50% of the respondents know none or only 1-2 of the top brands which are the three most known brands (Levi's, Tommy Hilfiger, Diesel, Esprit). Interestingly, at the same time they have the highest purchasing probability of 8 of 9 brands (except for Levi's cluster 3 has the highest probability). Even though they seem not to have a lot of knowledge about jeans brands, they are willing to buy them. Also, the cluster is very price sensitive and prefers low prices regardless of the brand. Therefore, they are considered to be the Bargain Hunters. The Bargain Hunters are in comparison to the other group less optimistic. Only 35% think the Mystery product might be their favourite jeans: Tommy Hilfiger. In comparison: about 62% and 52% of the other groups think that the mystery product is their favourite jeans brand.

The Fashionistas build the second cluster. They are very optimistic about their future which is in line with the highest score of 62% that the mystery product is their favourite jeans brand Levi's. They are really into fashion. Out of 9 brands they are the most aware of 7. About 70% of the respondents know at least 3 of the 4 top brands and approx. 62% are girls. Levi's has the highest purchase probability which indicates that it is the favourite brand of the Fashionistas. This indication is supported by the results of table 1.

Last but not least, many academic respondents are grouped – with almost 80% of the group members who have either a Bachelor, Master or PhD degree. Moreover, they are into old jeans brands like Lee and Levi’s. They have the highest awareness for Lee and the highest purchase probability for both brands. Those brands are the only once founded in the 19th century and are known as the original brands. Hence, the group is named the “Originals”. About 29% of the Originals are older than 27. The Originals form 70% of the top income class with an income of >100k Euro per year. In contrast to all other groups, more men (60%) than women are in this group. In comparison to the Fashionistas, the Originals are a little less optimistic.

### Market Share

In the given market simulation all products are skinny, blue jeans. Only the brands and prices vary.

Brand	Price	Total MS	Bargain Hunters MS	Fashionistas MS	Originals MS
Mystery	70	0.23	0.28	0.29	0.18
Hilfiger	79	0.23	0.25	0.20	0.23
Guess	69	0.23	0.33	0.32	0.13
Levi’s	99	0.18	0.06	0.16	0.25
None	0	0.14	0.08	0.04	0.22

Table 3: The market shares (MS) of all brands in the total market and in each cluster

Since the Fashionistas are very optimistic about the mystery product, the market share is very high. It is also high for the Bargain Hunters . This is due to their price sensitivity. As shown in table 1, the price is the most important feature for the Bargain Hunters and the mystery product is the purchasable good with a lower price than all other brands except Guess. In fact, the Bargain Hunters mystery market share decreases with an increasing price. In contrast, the Originals know what they want: Levi’s. Even though the price is way higher than for the other jeans, Levi’s has the biggest market share. The high value of none derives from the fact that only skinny, blue jeans are offered. But the Originals clearly prefer a straight fit.

The effect of a changing price from 0 to 200 on the market share of the mystery with no competitive reaction can be observed in figure 2.

## Market Structure for Different Prices of the Mystery Product

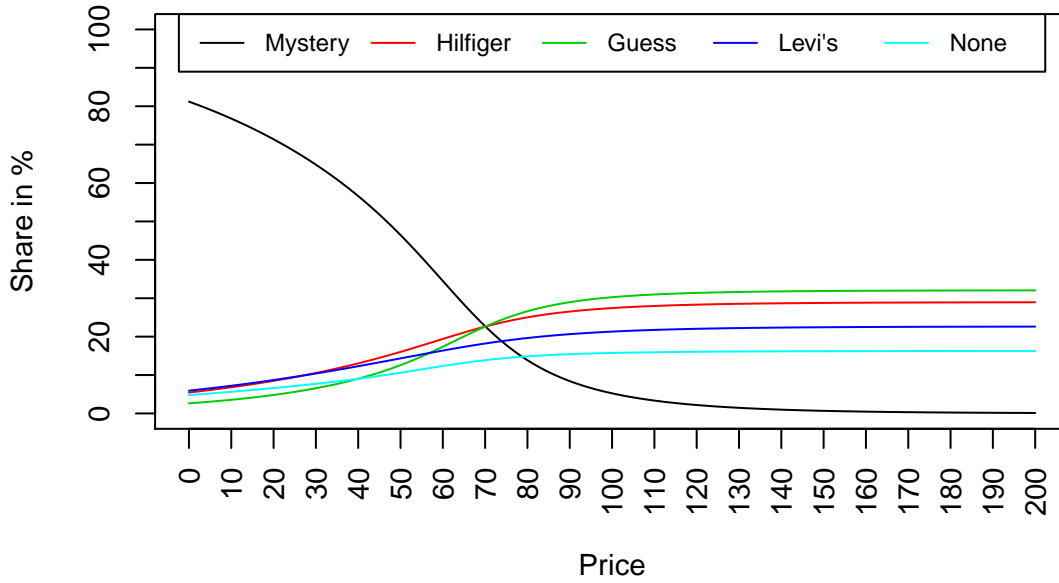


Figure 2: The market share of the mystery product with no competitive reaction

The mystery product only reaches a maximum market share of about 81%. With a price around 100 it has only 5% market share, starting at 140 Euro the market share is below 1% and starts to approach 0%. Even if the price of the mystery product would be 0, some respondents would still prefer to buy the other brands. The market share of Levi's is with almost 6% the highest as long as the mystery product price is very low. This indicates that especially the Originals stick with their favourite brand.

Within the framework of the survey, only 46% of respondents said that they would consider buying a mystery product. Anyhow, customers that are very price sensitive, such as the Bargain Hunters, will prefer the mystery product if the price is low and the fit as well as the color appeals to their preferences.

### Optimal Price

Lastly yet importantly, an (individual) optimal price strategy within a non-reacting competitive market is determined. It aims to maximize the profit while considering the willingness-to-pay (WTP) of the customers.

The profit can be calculated as:

$$Profit = (OptimalPrice - Costs) * MarketShare * MarketSize \quad (1)$$

For simplicity the assumed market size is 100k market participants.

	Total Market	Bargain Hunters	Fashionistas	Originals
Optimal Price	67	66	66	69
Profit (in k €)	599	743	763	456

Table 4: The optimal price (OP) and profit of the given mystery product in the total market and in each cluster with an assumed market size of 100k

The optimal price for the Originals is higher than for the other group which is due to the higher income and higher willingness to spend much money on a jeans. The lower profit indicates a lower market share which is in line with the observed behaviour of the Originals. Mostly, they prefer to spend more money on a Levi's jeans for instance. The market share of the Bargain Hunters is a little smaller than for the Fashionistas even though the price is the same. Table 3 gives an indication that the Bargain Hunters prefer Guess a little more than the mystery product if the price is very similar. In fact, when having a look at the shares for mystery and Guess, the Bargain Hunters have a slightly little lower share percentage on the mystery product but therefore, a slightly higher share percentage of Guess.

Besides an uniform or a clustered pricing strategy, also individual optimal prices can be considered. The minima and maxima of the computed prices and profits are presented and discussed below:

	Price	Profit (in k €)	Cluster Membership
Highest OP	98	472	Originals
Lowest OP	55	411/428	Bargain Hunters
Highest Profit	76	1388	Bargain Hunters
Lowest Profit	87	7.65	Originals
Range	43	1380	

Table 5: Comparison of the highest and lowest individual optimal price (OP) and profit as a results of individual targeting

Table 5 shows that customers' willingness-to-pay for a jeans varies strongly. The difference between the maximum and the minimum optimal is price 43 Euro and the differences of the profits in the market with 100k participants is 1388k Euro.

The highest profit is with a lower price and the lowest profit with a higher price which indicates a negative relationship between price and profit. The computation of the correlation confirms this indication. The correlation is -0.42.

### Relationship between Individual Price and Profit

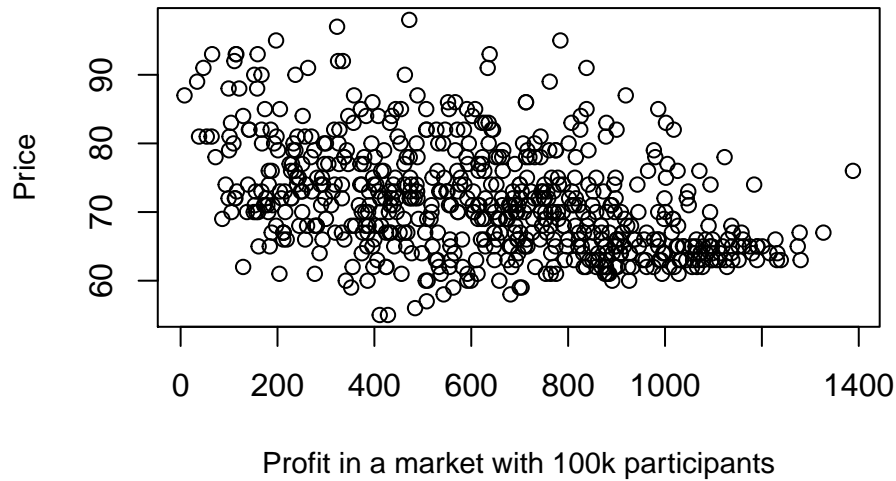


Figure 3: The scatterplot shows a negative relationship between profit and price

The correlation is in line with the assumptions of the formed clusters. Bargain Hunters like cheap prices and until a certain price, they are willing to buy the mystery product. Contrary, the Originals are willing to pay more for a product but they strongly prefer the jeans brand Levi's. Hence, the market shares for buying a mystery product is way lower as for the Bargain Hunters and this directly effects the aggregated profit.

The average profit of targeted pricing per customer is 62.8 Euro. For an uniform pricing strategy it is only 59.9 Euro per customer. Hence, the profit advantages of targeted pricing versus uniform pricing is on average 2.9 Euro per customer. In the assumed market of 100k participants that would be a total profit advantage of 290k Euro in comparison to uniform pricing. However, in a real life setting, targeted pricing is more costly than an uniform pricing strategy. The higher costs would reduce this profit advantage a bit.