Unsupervised Assignment

‘Customer segmentation for the telecommunications sector’

DECEMBER 2023

EXECUTIVE SUMMARY

Recent segmentation analysis has revealed fundamental findings that aim to facilitate the formulation of promotional strategies for our telecommunications company's marketing team. In the following report, results are presented that provide in-depth insight into the preferences and behaviors of our customers, identifying three key larger groups, that distinguish themselves by the type of payment each one use: "Group A", "Group B" and "Group C" and a fourth smaller but profitably relevant group, "Group D".

Among the outstanding results, the "Group A" segment stands out with an average card payment rate 73.61% higher than overall. This inclination presents valuable strategic opportunities, such as the introduction of a payment card linked to customers' credit cards for exclusive promotions. In addition, there is a clear preference for national long-distance calls, suggesting the possibility of targeting promotions to this type of service. The strategy for this group is crucial, as it represents 53.79% of the clients studied.

In contrast, "Group B" segment distinctly favors cash transactions, displaying an average payment rate 349% higher than the overall average. While their involvement in international calls is limited, their inclination towards discounts presents a strategic opportunity to capitalize on promotions in this domain, aiming to augment their engagement and establish them as more loyal and profitable customers.

The "Group C" segment, while not displaying a specific service preference, favors transactions through bank accounts as main characteristic. A recommended strategy involves reinforcing national long-distance and local call services, recognizing these as key areas where customers within this segment spend an extended duration.

The final group,"Group D", despite its small size, stands out as the most active segment in the use of the company's services. This group exhibits 383% higher average international call minutes and 144% higher average local call minutes compared to the data set average. Although this segment proves to be profitable, it faces a significantly higher churn rate. Therefore, it is imperative to implement specific promotional retention strategies focused on the long calls that characterize this group, with special attention to international calls, which have a remarkable relevance for them and have an average duration of 4 minutes.

In summary, a dual strategy is suggested for our promotional campaigns. For the first three groups, the focus should be on customer loyalty and increased service utilization, tailored to the distinctive characteristics of each segment. However, the group oriented to long call services demands a different strategy, focused on retention, with the objective of capitalizing on a segment that, although small in size, is profitable for the company in terms of the amount of services they consume.

APPROACH TO CLUSTER ANALYSIS AND PROFILING

In the initial stage, a thorough analysis of the various variables was carried out to understand them prior to their incorporation into the model. Given the dummy nature of many of these variables, the strategy of assigning a value of 0.7 to those that initially had a value of 1 was implemented in order to attenuate their influence on the overall significance of the model. Consequently, the dummy variables were adjusted to have values of 0 or 0.7.

The second stage focused on carrying out a correlation analysis by exploring the corresponding matrix. The objective was to determine the existence of significant relationships between the variables that could provide redundant information to the model. In general terms, low levels of correlation between the variables were observed. However, the correlation between "gender" and "churn" is notable, which was subsequently investigated without obtaining conclusive and relevant results. In addition, a clear negative correlation was identified between the different payment methods, which was accentuated when the groups were differentiated later in the analysis.

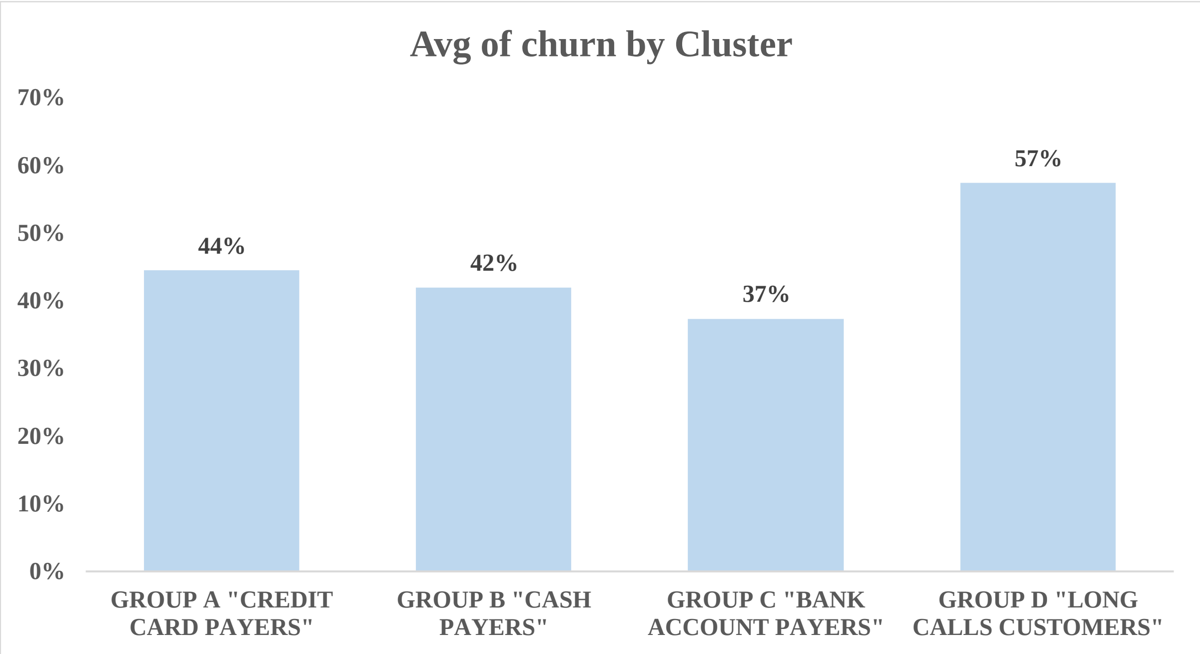
In the third phase, the clustering model was implemented using the K-Means method in Dataiku. Initially, all available variables were used, excluding the "churn" variable, since the interest is focused on observing the behavior of the clusters with respect to this variable. Several cluster options were explored, varying between 3, 4, 5, 6 and 7, along with an outlier cluster with a threshold of 5% and 150 observations.

The resulting model exhibited a low silhouette score (0.278) and failed to provide clear identification of the distinguishing characteristics of each cluster. However, it is critical to note that a marked differentiation according to payment methods within the groups was evident from the outset. Many of the variables initially considered lacked relevance in the model. Variables such as age, gender, car, number of children and family income were progressively discarded, since they did not offer conclusive information for the differentiation of the groups as they presented practically identical means and, therefore, lacked relative importance in the model.

The most appropriate choice for the model included the variables: international discount, international call minutes, local call minutes, domestic long-distance call minutes, bank account payment, credit card payment and cash payment. When applying the K-means method, the optimal model was the one with k=3 and a group of outliers. The silhouette score of this model was 0.46, but the most relevant aspect is that it has allowed us to efficiently profile the groups, thus facilitating the identification of promotions that are more appropriate for each one.

In the next phase of the analysis, we proceeded to describe these groups in more detail, using some additional variables, specifically those that we had previously excluded from the analysis. We set out to describe these 4 clusters in terms of gender, age, marital status and number of children. Using graphs with these variables as responses, we again confirmed that there is no significant differentiation between the clusters in terms of the aforementioned variables. However, we were able to establish that these groups have an average age between 57 and 59 years, with an average of one child, are married and are of both genders. Therefore, we can add to our profile that most of the clients belong to families.

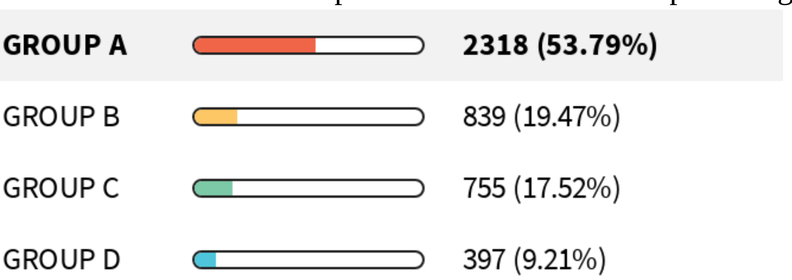
In the final phase, a comparison of our groups was carried out with the churn variable to determine which of them (if any) have a higher risk of dropping out. The results indicated that the first three groups have a mean churn rate ranging from 37% to 44%, suggesting that, although most clients tend to stay, some also drop out in these groups. However, the most vulnerable group is the fourth group, which has a mean churn rate of 57%. In other words, this segment is more prone to churn. In the following graph, the distribution of avg churn by cluster can be seen.



The general conclusions about the goups can be seen in the following table:

Segment	Profile
GROUP A	Pay with credit card, make more local calls, interested in national long-distance calls but the avg minutes are low.
GROUP B	Pay with cash, are interested in international discounts but avg min in international are low.
GROUP C	Pay with bank account, are not predefined for a specific service.
GROUP D	Make long international and local calls, are interested in international discounts, high profitable customers.

Distribution of customers per cluster in number and percentage.



RECOMMENDATIONS FOR THE MARKETING TEAM

Given the purpose of the study, which seeks to reduce the service abandonment rate among the company's clients, it is suggested that a differentiated strategy be implemented for the first three groups and the fourth, considering the sensitivity of each group to abandonment. For the first groups, the focus should be on maintaining customer loyalty and encouraging increased use of services. This will not only strengthen loyalty, but also make these customers more profitable for the company. On the other hand, the strategy for the last group should be oriented towards retention, since, due to the number of services they use, especially in comparison with the other groups, they are valuable customers but tend to abandon the services.

Specific strategies for loyalty promotions and increased service usage:

Group A:
Promotions can be implemented that leverage credit card usage, such as creating a purchasing card directly associated with their credit card. Customers could obtain benefits, such as discounts on local calls, when using this card. Given the interest shown in national long-distance calls, even if their use is low, a strategy could include discount packages for these calls or even offer free minutes on national long-distance calls for the regular use of minutes on local calls.

Group B:
Considering the particular interest of this segment in making cash payments and obtaining discounts on international calls, it is suggested to focus promotions on these preferences. Despite a moderate use of local calls, with an average of approximately 45 minutes, a promotion could be devised to establish a connection between this use and international calls. This would serve as a stimulus to increase participation and consumption of services, taking advantage of the group's propensity for discounts on international calls and thus generating greater commitment to the services offered by the company.

Group C:
Since this group does not have a predetermination for a specific service, this opens up the opportunity to design customized promotions tailored to their individual usage patterns. The implementation of exclusive offers, tailored to their specific needs and behaviors, aims to encourage greater participation in the services offered by the company. This strategy will not only capture the attention of this particular group, but also strengthen their connection with the company by offering solutions that are precisely aligned with their preferences and consumption habits. By personalizing promotions, we seek to generate a more significant impact and increase the loyalty of this customer segment.

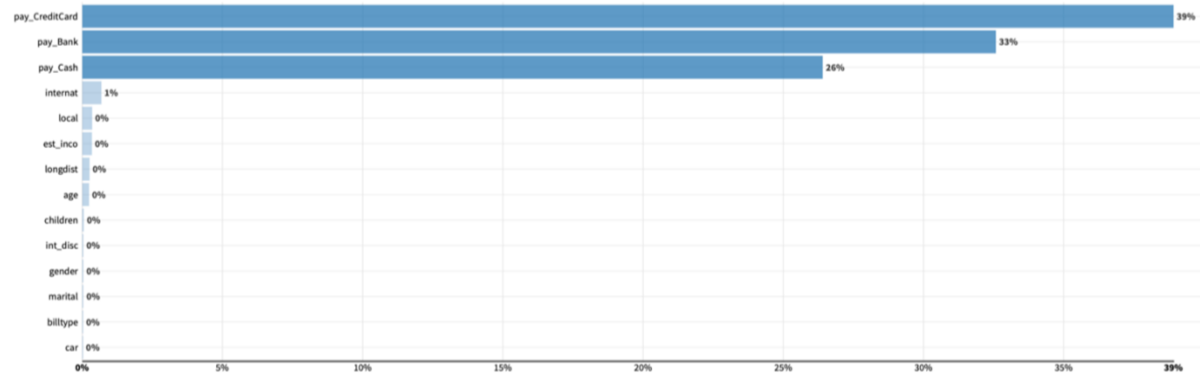
Specific strategies for retention promotions:

Group D:
The marketing team should devise specific promotional strategies tailored to this group's extended call patterns. Implementing personalized retention campaigns, exclusive international call promotions, and loyalty incentives can mitigate churn and maximize the profitability of this valuable segment. Additionally, proactive communication channels can be leveraged to keep this group informed about tailored offers and benefits, fostering a stronger connection with the brand and increasing the likelihood of retention.

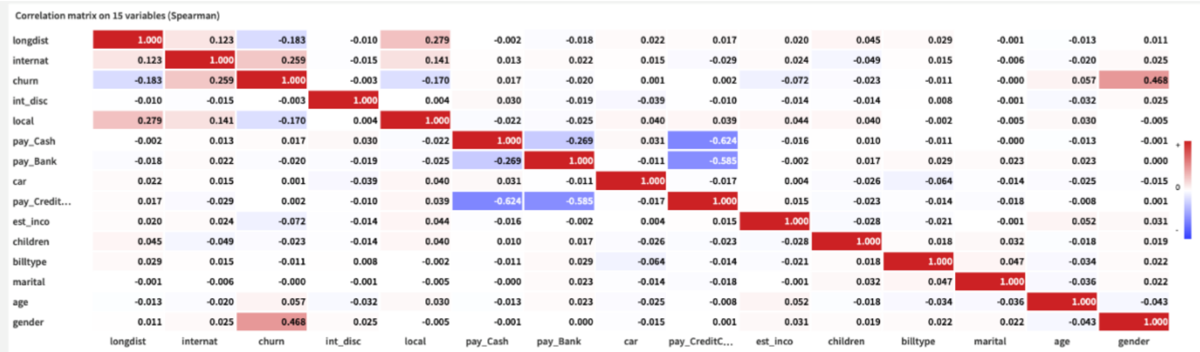
TECHNICAL ANNEX

Results of feature importance after the first analysis maintaining all the variables except churn:

Feature Importance

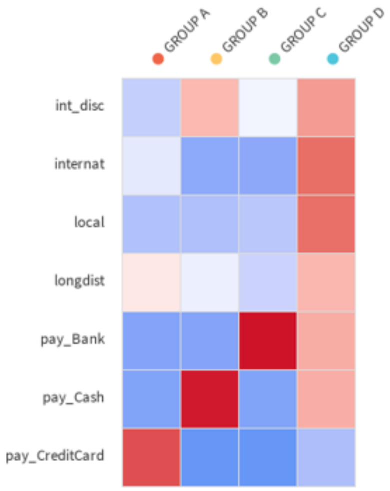


Correlation matrix considering the 15 variables of the dataset. Since gender is a highly correlated variable, it will not be included in the case study. We can also look at a negative correlation for the different payment methods, which makes total sense, implying that one customer must select one method.


Correlation matrix on 15 variables (Spearman)

	longdist	internat	churn	int_disc	local	pay_Cash	pay_Bank	car	pay_CreditC...	est_inco	children	billtype	marital	age	gender
longdist	1.000	0.123	-0.183	-0.010	0.279	-0.002	-0.018	0.022	0.017	0.020	0.045	0.029	-0.001	-0.013	0.011
internat	0.123	1.000	0.259	-0.015	0.141	0.013	0.022	0.015	-0.029	0.024	-0.049	0.015	-0.006	-0.020	0.025
churn	-0.183	0.259	1.000	-0.003	-0.170	0.017	-0.020	0.001	0.002	-0.072	-0.023	-0.011	-0.000	0.057	0.468
int_disc	-0.010	-0.015	-0.003	1.000	0.004	0.030	-0.019	-0.039	-0.010	-0.014	-0.014	0.008	-0.001	-0.032	0.025
local	0.279	0.141	-0.170	0.004	1.000	-0.022	-0.025	0.040	0.039	0.044	0.040	-0.002	-0.005	0.030	-0.005
pay_Cash	-0.002	0.013	0.017	0.030	-0.022	1.000	-0.269	0.031	-0.624	-0.016	0.010	-0.011	-0.000	-0.013	-0.001
pay_Bank	-0.018	0.022	-0.020	-0.019	-0.025	-0.269	1.000	-0.011	-0.585	-0.002	0.017	0.029	0.023	0.023	0.000
car	0.022	0.015	0.001	-0.039	0.040	0.031	-0.011	1.000	-0.017	0.004	-0.026	-0.064	-0.014	-0.025	-0.015
pay_Credit...	0.017	-0.029	0.002	-0.010	0.039	-0.624	-0.585	-0.017	1.000	0.015	-0.023	-0.014	-0.018	-0.008	0.001
est_inco	0.020	0.024	-0.072	-0.014	0.044	-0.016	-0.002	0.004	0.015	1.000	-0.028	-0.021	-0.001	0.052	0.031
children	0.045	-0.049	-0.023	-0.014	0.040	0.010	0.017	-0.026	-0.023	-0.028	1.000	0.018	0.032	-0.018	0.019
billtype	0.029	0.015	-0.011	0.008	-0.002	-0.011	0.029	-0.064	-0.014	-0.021	0.018	1.000	0.047	-0.034	0.022
marital	-0.001	-0.006	-0.000	-0.001	-0.005	-0.000	0.023	-0.014	-0.018	-0.001	0.032	0.047	1.000	-0.036	0.022
age	-0.013	-0.020	0.057	-0.032	0.030	-0.013	0.023	-0.025	-0.008	0.052	-0.018	-0.034	-0.036	1.000	-0.043
gender	0.011	0.025	0.468	0.025	-0.005	-0.001	0.000	-0.015	0.001	0.031	0.019	0.022	0.022	-0.043	1.000

The heatmap below shows how important each feature is to characterize a cluster. This heatmap shows that per each group of the first three, there is a high clustering based on the payment method preferences. For group A, the long distance calls is also an important feature for clustering, while for group B, International Discount is an important feature.



Legend: **Low** **High**