Hierarchical Clustering N.C. = 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Observation** | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **comments** |
| **Size** | 2206 | 997 | 938 | 858 |  |
| **Gender** | Mostly female (87.08%) | Mostly male (97.69%) | Mostly female (62.90%) | Mostly female (80.77%) | The gender is unbalanced. 3227 female and 1772 male. |
| **Relative % of gender** | 59.53 fem.  16.08 male | 0.71 fem.  54.97 male | 18.28 fem.  19.64 male | 21.48 fem.  9.31 male |  |
| **Car** | No (76.11%) | Yes (61.28%) | No (52.13%) | No (80.54%) |  |
| **Child (mean)** | 0.4914 | 0.4473 | 0.5789 | 0.0303 |  |
| **Income (mean)** | 152103 | 185769 | 210535 | 133146 |  |
| **Credit (mean)** | 559203 | 596484 | 747231 | 541100 |  |
| **Loan (mean)** | 25322 | 28460 | 32868 | 23629 |  |
| **Price (mean)** | 499619 | 529871 | 679467 | 483970 |  |
| **Job status**  Commercial  Pensioner  State servant  working | 568 | 245 | 313 | 0 | Cluster 4 are only pensioner. Hence, the lowest income and num. of child. |
| 1 | 0 | 0 | 858 |
| 107 | 28 | 232 | 0 |
| 1530 | 724 | 393 | 0 |
| **Study** | Sec. edu (84.45%) | Sec. edu (86.46%) | High Edu (82.19%) | Sec. edu (81.23%) |  |
| **Family stats.** | Single (14%) | Single (25%) | Married (85%) | Widow (16%) | For all the clusters, the main stats. Is married |
| **House** | No provided info | | | | |
| **Age (mean)** | 40.59 | 38.81 | 39.17 | 59.25 |  |
| **Job duration** | 6.79 | 5.50 | 6.83 | 10.89 |  |
| **Job occupation** | Sales (19.72%) | Laborers (41.02%) | Core staff (22.7%) | Unknow (100%) (pensioner) |  |
| **N enquiries** | 0.26 | 0.28 | 0.35 | 0.21 |  |

Now, we can try to extract some basic features description for each cluster.

**Cluster 1** and **Cluster 2** are from similar social class, but distinguished with male and female. The cluster 1 are mostly composed by women (87%) while cluster 2, by man (97.7%). Their education level is similar, which the vast majority have only secondary education (84% and 86%). Even though their income has some discrepancy (152k vs. 185k vs), this could be result of gender inequality on their income.

And in the target level, the cluster 1 tends to overdue the payment (12,78%) slightly more than the cluster 2 (3%).

In the **Cluster 3**, the gender is more homogeneously distributed. Even though there is 62.9% of female (consequence of unbalanced gender distribution), the relative frequency is that 18.28% of female and 19.64% belongs to this cluster.

We believe that this cluster represents the grouping of people who belongs to an higher social level, with higher income (210.535) and education (82.19%). This also can be observed by their job type, where core staff represents 22.7% of all the occupations. On the other hand, they have a more stable familiar situation (85% married, with higher nº child mean). In the target level, the percentage of overdue is 6.5%.

In the **Cluster 4**, there is a grouping of retired people, whose income is much lower (133k), their jobs status are completely pensioners. The average age of this cluster is 59 years old, and the percentage of widow is significantly higher than normal another cluster. In the target level, the percentage of overdue is 6.3%.

 