|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Segmentation data - Legend** | | |  |  |  |  |  |  |  |  |
|  | The dataset consists of information about the purchasing behavior of 2,000 individuals from a given  area when entering a physical ‘FMCG’ store. All data has been collected through the loyalty cards they use at checkout.  The data has been preprocessed and there are no missing values. In addition, the volume of the dataset has been  restricted and anonymised to protect the privacy of the customers. | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Variable** | **Data type** | **Range** | **Description** |  |  |  |  |  |  |  |
|  | **ID** | numerical | Integer | Shows a unique identificator of a customer. | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Sex** | categorical | {0,1} | Biological sex (gender) of a customer. In this dataset there are only 2 different options. | | |  |  |  |  |  |
|  |  |  | 0 | male |  |  |  |  |  |  |  |
|  |  |  | 1 | female |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Marital status** | categorical | {0,1} | Marital status of a customer. | |  |  |  |  |  |  |
|  |  |  | 0 | single |  |  |  |  |  |  |  |
|  |  |  | 1 | non-single (divorced / separated / married / widowed) | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Age** | numerical | Integer | The age of the customer in years, calculated as  current year minus the year of birth of the customer at the time of creation of the dataset | | | | | | |  |
|  |  |  | 18 | Min value (the lowest age observed in the dataset) | |  |  |  |  |  |  |
|  |  |  | 76 | Max value (the highest age observed in the dataset) | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Education** | categorical | {0,1,2,3} | Level of education of the customer | |  |  |  |  |  |  |
|  |  |  | 0 | other / unknown |  |  |  |  |  |  |  |
|  |  |  | 1 | high school |  |  |  |  |  |  |  |
|  |  |  | 2 | university |  |  |  |  |  |  |  |
|  |  |  | 3 | graduate school |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Income** | numerical | Real | Self-reported annual income in US dollars of the customer. | |  |  |  |  |  |  |
|  |  |  | 35832 | Min value (the lowest income observed in the dataset) | |  |  |  |  |  |  |
|  |  |  | 309364 | Max value (the highest income observed in the dataset) | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Occupation** | categorical | {0,1,2} | Category of occupation of the customer. | |  |  |  |  |  |  |
|  |  |  | 0 | unemployed / unskilled | |  |  |  |  |  |  |
|  |  |  | 1 | skilled employee / official | |  |  |  |  |  |  |
|  |  |  | 2 | management / self-employed / highly qualified employee / officer | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | **Settlement size** | categorical | {0,1,2} | The size of the city that the customer lives in. | |  |  |  |  |  |  |
|  |  |  | 0 | small city |  |  |  |  |  |  |  |
|  |  |  | 1 | mid-sized city |  |  |  |  |  |  |  |
|  |  |  | 2 | big city |  |  |  |  |  |  |  |