|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | This survey data is recorded by the G+I Nuremberg, a subsidiary company of the Gesellschaft für Konsumforschung (Society for Consumer Research) within the framework of its household-panel. Description of the variables is given separately (Desc.docx). The data are based on the coffee bargains of 2752 households, which belonged to the panel between 1.1.1988 and 31.12.1990. All of these households were located in former West-Germany, and a total of 130986 bargains was registered.   |  | | --- | | **Description of the covariates** | | |  |  | | --- | --- | | Column | Number of packets bought | | A | 1 : one packet 2 : two packets 3 : three or more packets | | B | Selling-price per packet of coffee | | 1 : less than 6,50 DM (Currency Deutsche Mark) 2 : 6,50 DM to 8,50 DM 3 : more than 8,50 DM | | C | Identification number of household | | D | Coffee Brand | | 1 : Jacobs Krönung 2 : Jacobs other 3 : Aldi 4 : Aldi other 5 : Eduscho Gala 6 : Eduscho other 7 : Tchibo Feine Milde 8 : Tchibo other 9 : Andere Kaffeemarken | | E | Period between two coffee purchases in days | | F | Age of householder | | 1 : less than 24 years 2 : 25 to 39 years 3 : 40 to 49 years 4 : 50 to 59 years 5 : 60 years or more | | G | Socioeconomic level of householder | | 1 : upper class 2 : upper middle class 3 : middle class 4 : lower middle class 5 : lower class | | H | Monthly income of household | | 1 : less than 1499 DM 2 : 1500 to 2499 DM 3 : 2500 to 3499 DM 4 : 3500 DM or more | | I | Size of household | | J | Price-consciousness | | 1 : not at all 2 : a little 3 : price-conscious 4 : distinctly price-conscious | | K | Education of householder | | 1 : nine-year elementary school 2 : intermediate high school 3 : high-school / university | | L | Brand loyalty | | 1 : loyal 2 : not loyal | | | |

The file (Coffee.xlsx) contains survey data with detailed description given in the attached doc (Desc.docx). The survey aims to understand coffee consumption pattern of households. Specifically, a few of the questions the survey may attempt to answer are

* Which brands of coffee are more popular? Given a brand, are all variants equally preferred?
* What are the prices of different brands of coffee?
* How frequently does a household buy coffee? How many packets of coffee are bought at a time?
* What are the factors that have an impact on a household’s coffee purchase pattern? Does brand preference depend on household size? Does purchase depend on a person’s income or education level?

There may be many other related and relevant questions that a company may be interested in. But before we get to do that, we need to ‘prepare’ the data and do some simple checks. The list is endless. However, a starting point could be the following:

* Are there any variables/columns you think will be eventually irrelevant for the study? Why?
* Can you possibly identify some outliers?
* For better analysis do you think a few brand variants needs to be clubbed together?
* Can you do some probability estimates- Say what is the probability that a randomly chosen household belongs to the top two social levels? What is the probability that households belonging to the top two social levels buy more expensive coffee?

You are expected to discuss on the above-mentioned points. You are also expected to identify a few (between 2 and 4) relevant business questions, aligning data against each question or pointing out if any relevant data is not available. Request you to participate and share your insights.