Pivotal™

PIVOTAL DATA SCIENCE CANDIDATE EXERCISE

DESCRIPTION

The annual "family-level" data files included here (available on the National Bureau of Economic Research (NBER) web site, www.nber.org) are the result of linking the four quarterly interviews for each Consumer Expenditure Survey (CE) respondent family. The processed data also collapses the hundreds of spending, income, and wealth categories into a consistent set of categories across all the years.

Our customer is a manufacturer of consumer goods and we want to find a way to use this data to help our customer.

TASK

Propose one way of using this data employing **one** of the following methods: regression, classification or clustering. Execute your proposal and discuss your methodology, justify your algorithm/ feature selection and share insights from the model. We're interested in understanding your approach as well as your ability to communicate any insights derived from your model. The output should include a Microsoft Word, Power Point or PDF document.

GUIDELINES

- You have 24 hours to complete this exercise though we expect you to spend 2-3 hours
- Please send your completed work to the Pivotal employee who sent you this exercise within this time frame.
- We cannot provide answers to any questions about the data. All information we can provide is contained in this package.
- You are free to use any tool you prefer (SQL, R, Python, Matlab, Weka, RapidMiner, Mahout, etc.)

DATA FEATURES

- Family characteristics (head of household's age, income, marital status, employment, number of kids, ages of kids, spouse's age, spouse's employment status, etc.)
- Income sources for the family (wages, business, pension, rents etc.)
- Expenditures for many categories (housing, food, clothes, utilities, entertainment, etc.)
- The quarter the interview took place
- Income and Expenditure for the previous period (lagged variables)

A complete data dictionary describing all of the features is included in the package (Data Dictionary.pdf)

DATASET

- Consumer Expenditure Survey for 1996-2000 (12k rows, 220 columns)
  Data_consumer_expenditure_survey.csv
- Prices for goods and services by quarter 1990-2000 (43 rows, 40 columns)
  Data_Supplementary_Prices.csv