PROBLEM STATEMENT

1. You are given with the transactional (Sales of various categories of SKUs- Stock Keeping Units) and economic parameters of the cities (CPI, City Type – City or Town, Area, Literacy Rate, Population, etc.) of various outlets or stores of a retail chain, Streets-Mart.
2. Please cluster these stores optimally based on the given data (transactional and economic parameters), identify parameters that affects the revenue of the stores to be different for different group for that specific retail chain, Streets-Mart.
3. Based on the above, prescribe the marketing strategy to enhance the revenue for various stores of different city types.
4. Also assume if GDP (or rather GSDP – Gross State Domestic Product) of India increases or decreases or if CPI changes then will there be any change in your above strategy?

Note 1: Apart from the given transactional data and economic parameters, you are free to take/consider various other data information (e.g. socio-economic, market data, demographic data, macroeconomic [GSDP – Gross State Domestic Product, Per capita Income of the cities, population density, etc.]) to fine tune your outcome.

Note 2: Any Statistical / Machine Learning Programming language /package/ platform may be used (R, Python, etc.)

Note 3: As the number of rows is not very high you can keep a small set of data for validation.

Note 4: The sources of economic data are various governments and other websites. While collecting additional data, you may not find all data in one place or there could be missing data; in those situations you can take judicious call for missing value treatment or take close by data rationally.

Sale – The Revenue of a category of SKU (Stock Keeping Unit) for 8 months of a Financial Year (2014-15)

CPI – Consumer Price Index (average CPI computed for those 8 months for a state)

City Name: As given

City Type – If it is major City/ Metro or Big Town

Literacy Rate – Literacy Rate of the City

Population – It is the Population of the city as per the last census 2011 (yes, it will be different for the said year). The population of the city that has been considered is the one without / excluding agglomeration to make more sense in the said context of target audience/ shoppers.