

## Unsupervised Training Data Set

Descriptive modeling is designed to identify the underlying patterns in the data without the existence of an outcome or response variable; it is also called *unsupervised training*. That is, observations might be formed into groups without prior knowledge of the data or unknown links and associations might be discovered between certain variables in the underlying data. Cluster analysis, SOM/Kohonen maps, and principal components are considered descriptive modeling techniques. In Enterprise Miner, all the variables in the analysis are set to a variable role of **input** when performing the various unsupervised training techniques.

The following is the data set that is used in the various Enterprise Miner nodes based on unsupervised training for explaining both the configuration settings and the corresponding results. The SAS data set is based on major league baseball hitters during the 2004 season. The data set consists of all baseball hitters in the major leagues that had at least 150 at bats, which resulted in 371 observations of hitters in the major leagues. The data consisted of twenty variables in the data mining data set with fifteen numeric variables and five categorical variables. The following table displays the variable names, model role, measurement level, and variable description of the variables.

Name	Model Role	Measurement Level	Description
TEAM	Input	Nominal	Major League Team
FNAME	Rejected	Nominal	First Name
LNAME	Rejected	Nominal	Last Name
POS1	Input	Nominal	Position
G	Input	Interval	Number of Games Played
OPS	Input	Interval	On-base percentage plus slugging pct.
SLG	Input	Interval	Slugging Percentage <i>Total Bases Divided by the Number of At Bats</i>
AVG	Input	Interval	Batting Average <i>Hits Divided by the Number of At Bats</i>
OBP	Input	Interval	On-Base Percentage $(H + BB + HBP) / (AB + BB + HBP + SF)$
AB	Input	Interval	At Bats
R	Input	Interval	Runs Scored
H	Input	Interval	Hits
H2	Input	Interval	Doubles
H3	Input	Interval	Triples
HR	Input	Interval	Home Runs
RBI	Input	Interval	Runs Batted In
BB	Input	Interval	Walks
SO	Input	Interval	Strikeouts
SB	Input	Interval	Stolen Bases
LEAGUE	Input	Binary	American League or National League