

Appendix 1: Data preparation change log

Variable	Comment	Action
ZIP_3_CD	Modeling type should be nominal.	Open column information window and change it.
PROD_DESC, SEG_DESC, SUB_CAT_DESC	Description variables' modeling type should be set as unstructured text to find key words.	Change its modeling type as unstructured text.
SLS_DTE_NBR	Date variable.	First change its data type to character. Then, change it to numeric, choose ordinal as modeling type, and choose appropriate format.
EXT_SLS_AMT	Monetary variable.	Choose currency (US dollar) as the data format.
SLS_QTY, POS_TRANS	A zero value in sales quantity with a zero value in sales amount could show few inferences.	Hide and exclude these 25 rows by using row selection tool.
EXT_SLS_AMT, UNIT_PRICE	Create a new variable "UNIT_PRICE" to compute the unit price of each product transacted and find out outliers of the variable "EXT_SLS_AMT".	Set the upper limit of price, like 3 times the 90th percentile, to recognize outliers and the price greater than the limit should be outliers. Hide and exclude them.