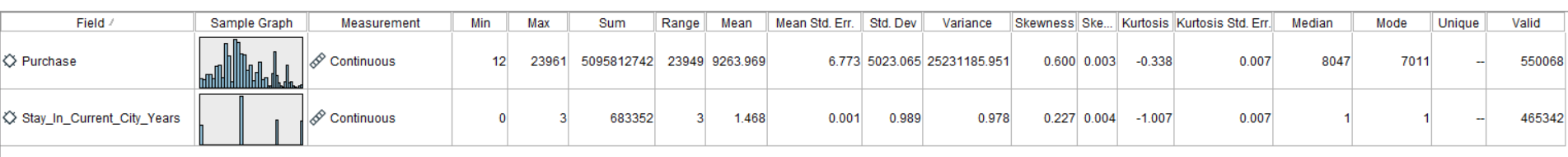
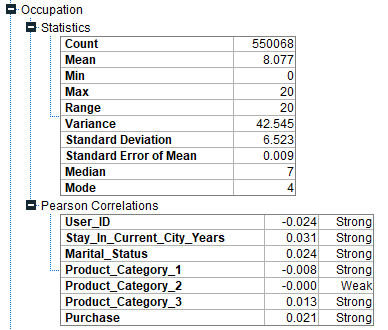
**Phase II: Analysis Outline**

* Submit a copy of your final dataset (if no NDA) and a Word (or PowerPoint file) containing the outline of your data analysis and the analyses that you have done (or plan to do).
* You do not need to have completed all the analysis at this point, but you should at least list what techniques you plan to use and a description of the way you plan to present your results visually.
* Your outline should contain at least the following sections. You can submit this as a draft of your report in phase III in Word.
  + 1. Introduction
  + 2. Problem Description and Research Questions (or conceptual view of the study)
  + 3. Data
  + 4. Analysis (methods, the relevance of methods, evaluations, etc.)
  + 5. Conclusions

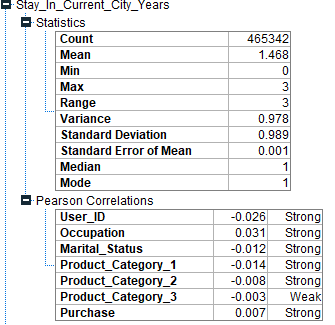
Data Audit for 2 continuous variables:

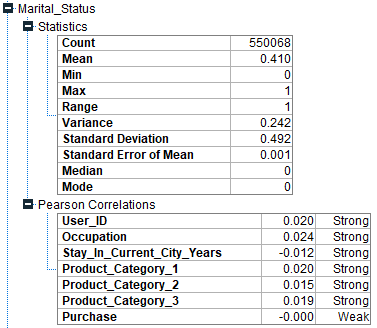


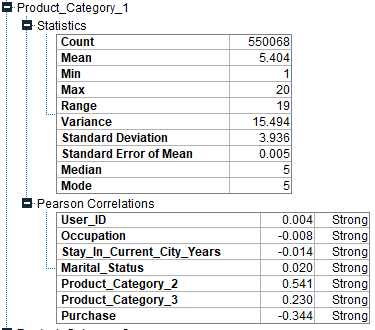
We have performed Data Audit for only 2 continuous variables because most of the variables in our dataset are categorical and ordinal.



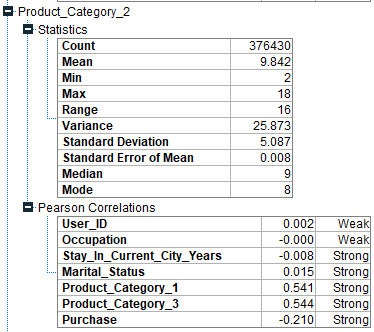
For the occupation variable, the mean is 8.077, the max and range is 20. Occupation has no strong correlation with any of the other variables.



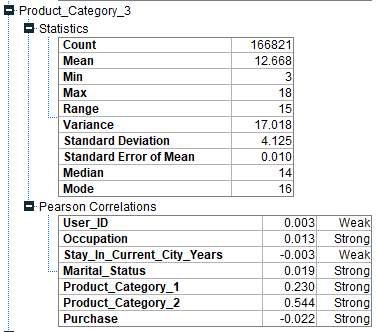


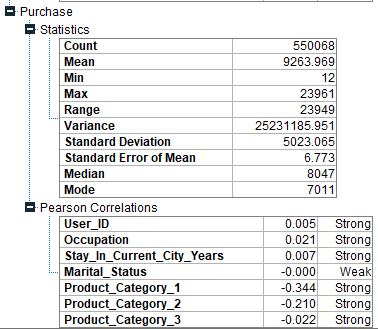


Product Category 1 has a strong correlation with Product Category 2

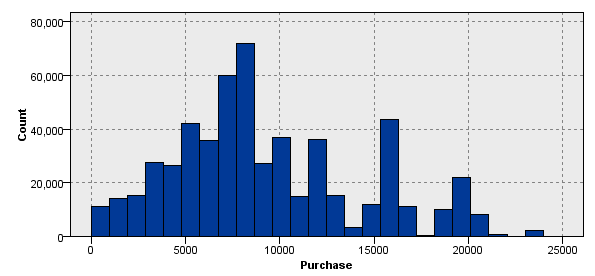


Product Category 2 has a strong correlation with Product Category 3

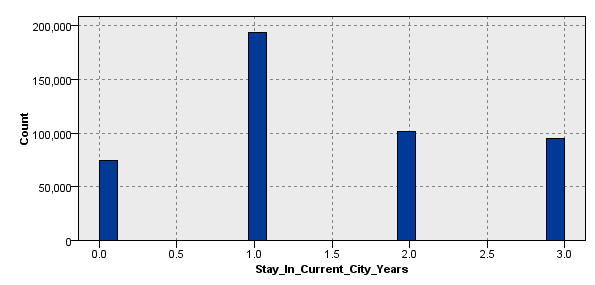


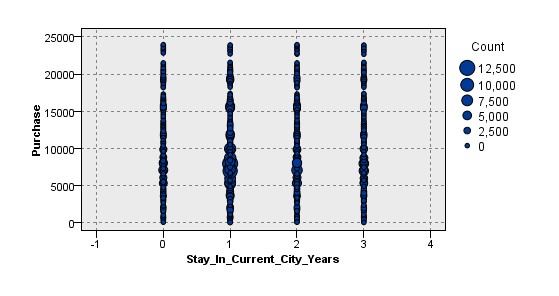


Histogram of Purchase



Histogram of Stay in Current years as they are categorical variables





For the occupation variable, the mean is 8.077, the max and range is 20. Occupation has no strong correlation with any of the other variables. Product Category 1 has a strong correlation with Product Category 2. Product Category 2 has a strong correlation with Product Category 3. As most of the variables are categorical, we have not produced scatterplots. The purchase histogram is left skewed which means that the mean is less than mode and this also shows that the data is negatively skewed.