**Data Preparation**

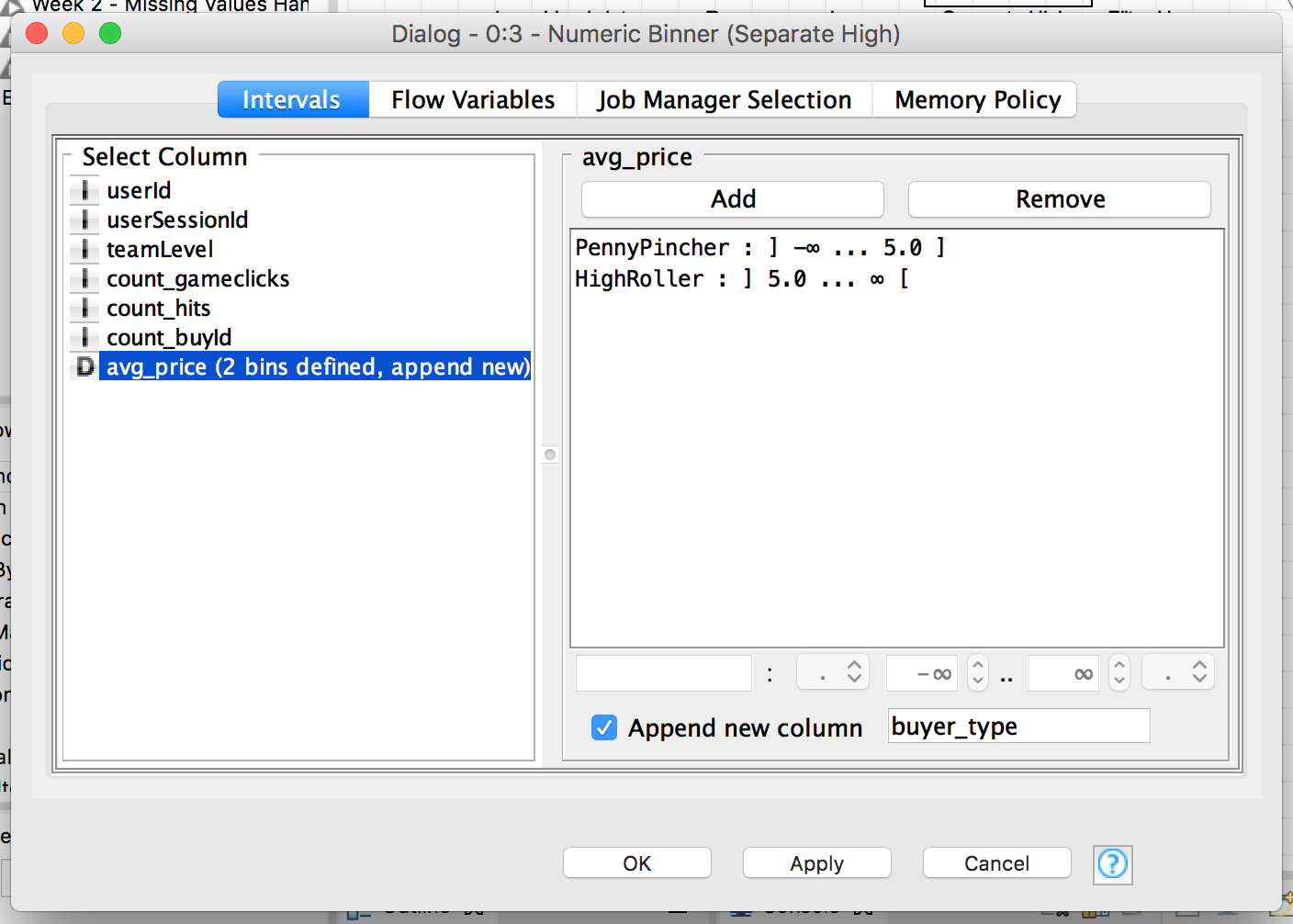
Analysis of combined\_data.csv

Sample Selection

|  |  |
| --- | --- |
| **Item** | **Amount** |
| # of Samples | 4619 |
| # of Samples with Purchases | 1411 |

Attribute Creation

A new categorical attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers). A screenshot of the attribute follows:



High rollers are defined as those who purchase items over $5.00. Defining a new column based on the avg\_price allows us to classify users accordingly.

The creation of this new categorical attribute was necessary because:

Our goal is to understand the attributes of who makes large purchases. This categorical variable is what we are going to base our decision tree upon.

Attribute Selection

The following attributes were filtered from the dataset for the following reasons:

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userId | Not relevant in determining the user’s behavior. |
| userSessionId | Not relevant in determining the user’s behavior. |
| avg\_price | This variable was used to create the categorical variable “buyer\_type”, the variable we are trying to predict based on other data elements. Therefore, we do not want to include this as a criteria in our decision tree model. |