**Data Preparation**

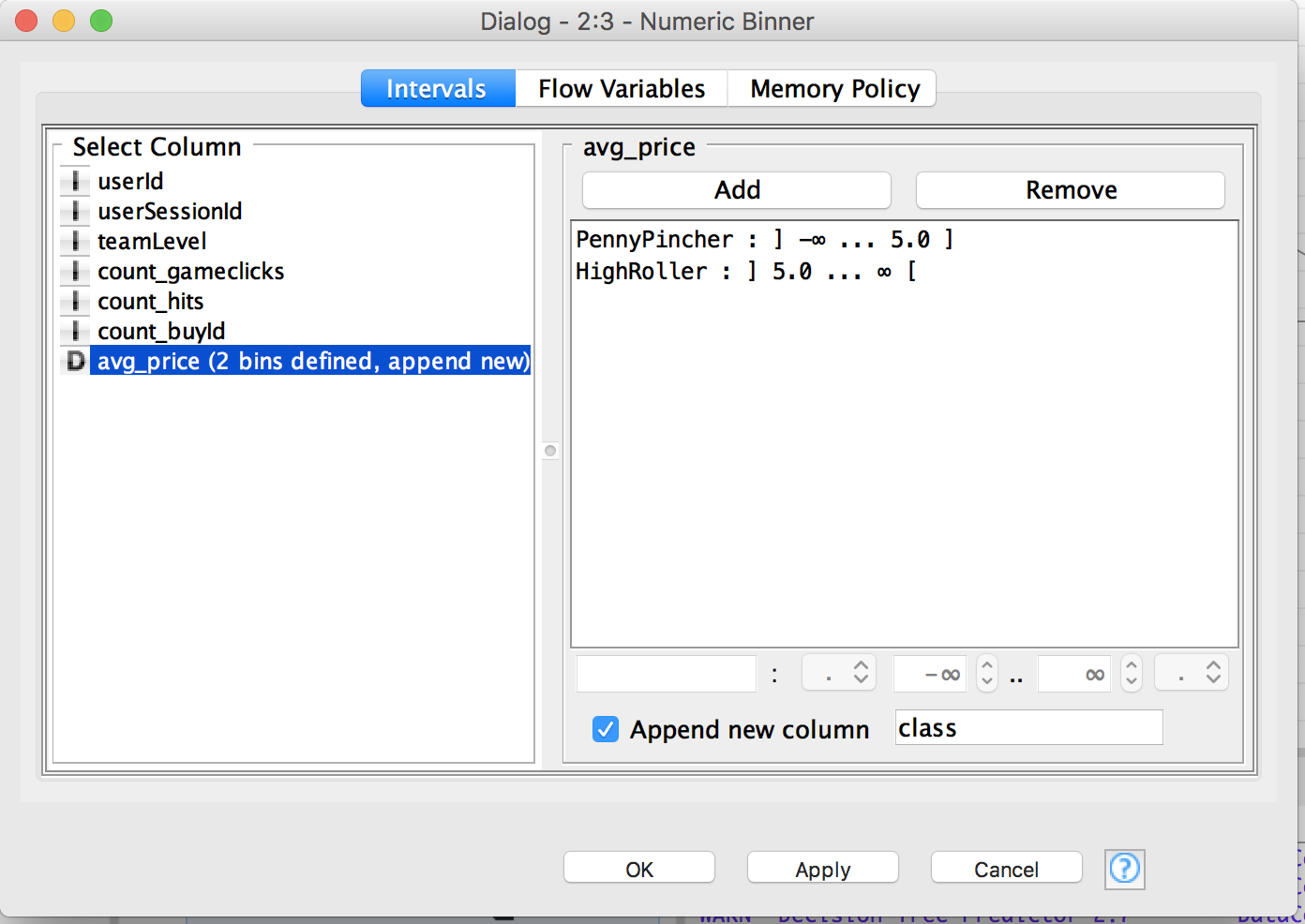
Analysis of combined\_data.csv

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:



The column class is the new categorical attribute. The value is either PennyPincher or HighRoller. Its value depends on avg\_price.

The creation of this new categorical attribute was necessary because the decision tree classifier needs a class label for each data record.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userId | UserId doesn’t represent any actual properties about the user. It is just an identifier. |
| userSessionId | Similar to userId. It is just an identifier. |
| avg\_price | We have already converted avg\_price into the class label. So we do not need it any more. |