

## Data Preparation

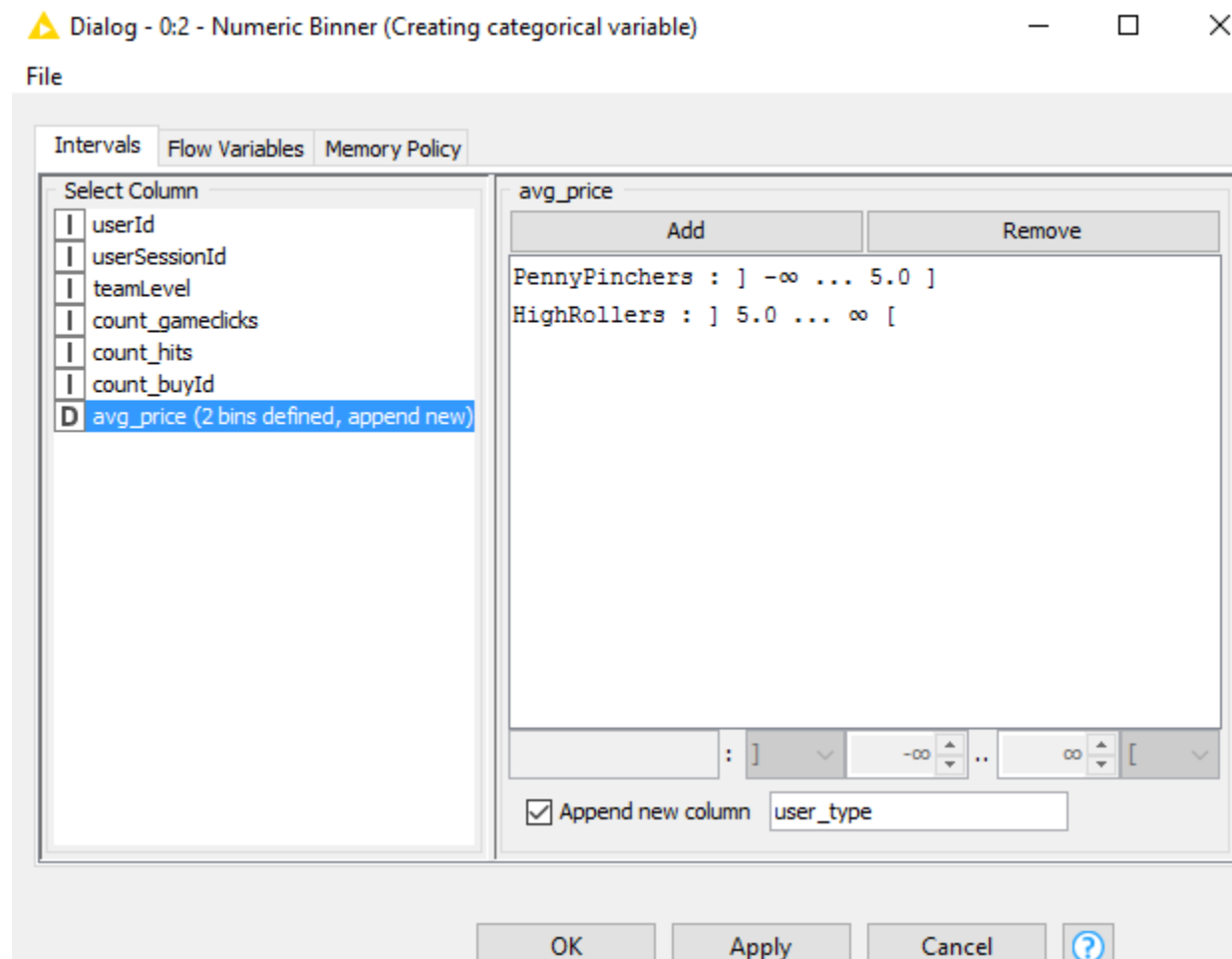
Analysis of combined\_data.csv

### 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 creation of this new categorical attribute was necessary because it helps to categorize HighRollers(more than \$5.00) and PennyPinchers(less than or equal to \$5.00)

### Attribute Selection

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

Attribute	Rationale for Filtering
userId	This attribute can be removed because our target is predicting which user is likely to purchase big-ticket items and this attribute is unnecessary
userSessionId	The userSession id attribute has no use in our analysis so it can be removed because our target is predicting which user is likely to purchase big-ticket items
avg_price	The creation of new attribute 'user_type' makes the attribute 'avg_price' unnecessary so we can remove it.