

## Data Preparation

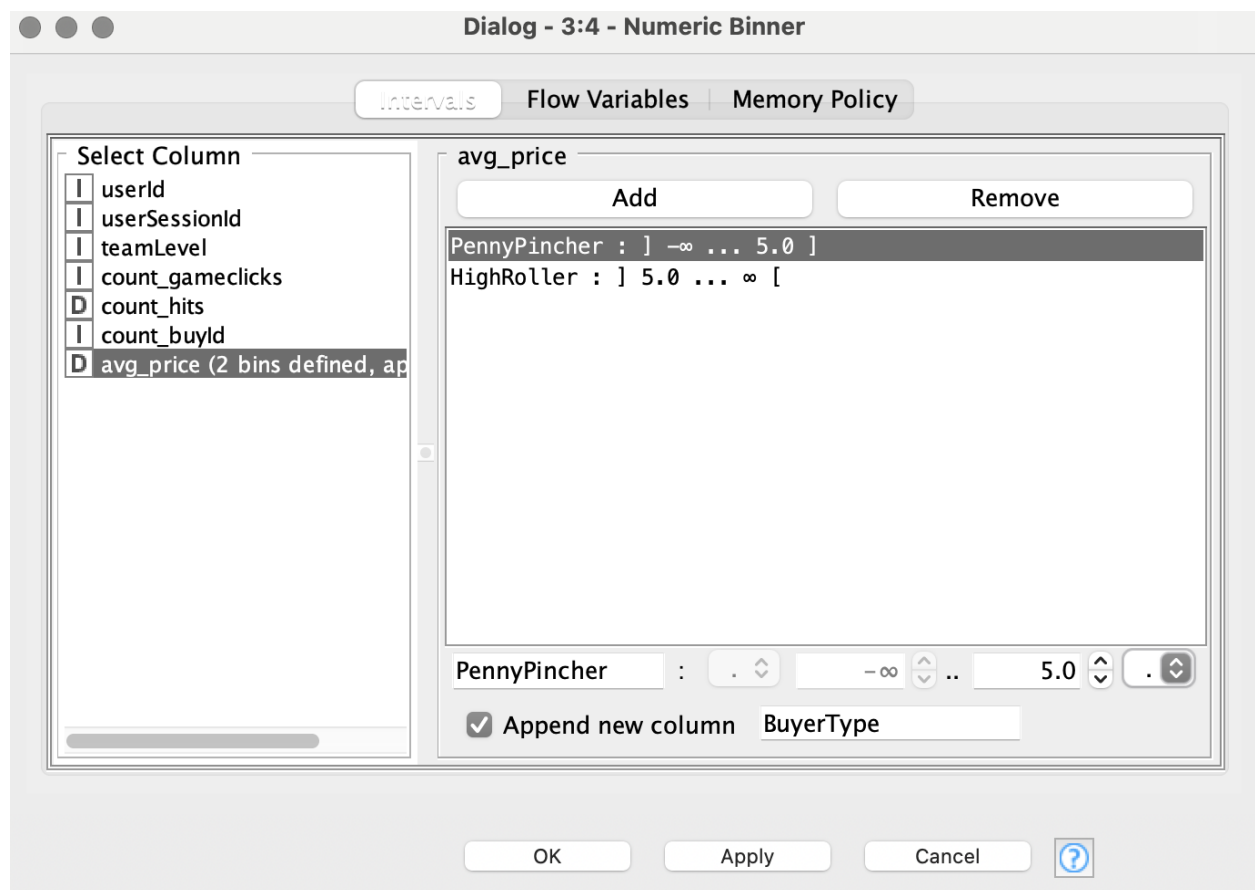
Analysis of combined\_data.csv

### 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 attribute consists of two classes. PennyPinchers spend less or equal than \$5 on average and HighRollers spend more than \$5 on average.

The creation of this new categorical attribute was necessary because it will give us insight into the users' buying behaviors. This will be the category our model will try to predict.

### Attribute Selection

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

Attribute	Rationale for Filtering
avg_price	This column was used to build the target column. Keeping would mean a data leakage
userId	Not influential in the model
sessionId	Not influential in the model