

## Data Preparation

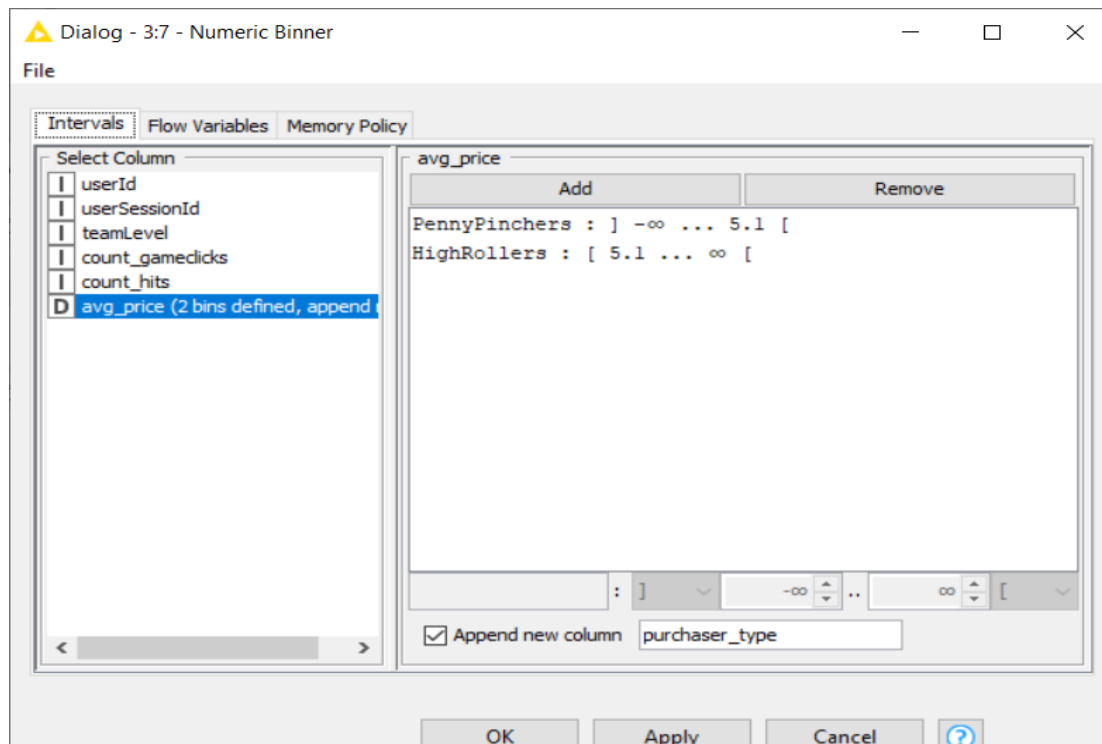
Analysis of combined\_data.csv

### 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 design description:

- High roller those who are spending more than 5.00\$.
- Penny pinchers those who spend 5.00\$ or less.
- The new column builds based on the average price that make us use it for classifying the users.

The creation of this new categorical attribute was necessary because:

We want to understand the attribute that we have and use it to know who make the large purchases in the game in order to make decision and this categorical variable are the foundation in our decision tree.

Attribute Selection

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

Attribute	Rationale for Filtering
userId	does not have relevant to the workflow.
userSessionId	does not have relevant to the workflow.
avg_price	This column was used to create the categorical column "purchaser_type" we will try to predict based in the new split data so that why we dont need this futaur in our workflow.