

## Data Preparation

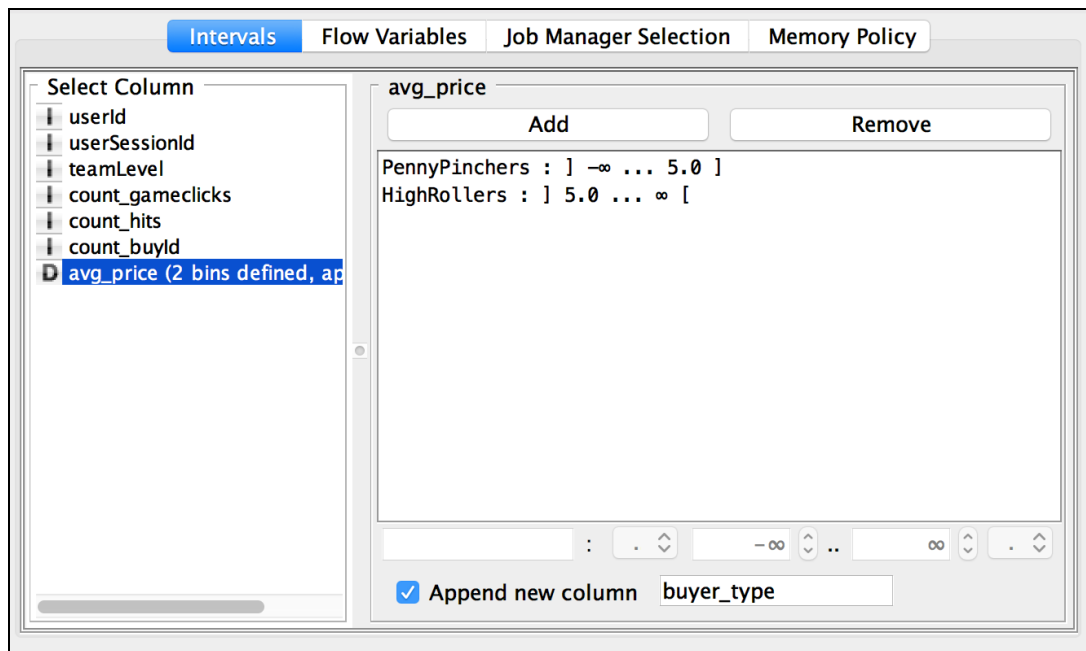
Analysis of combined\_data.csv

### 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:



It was created a new categorical variable named “buyer\_type”. It is derived directly from the avg\_price attribute and generates two bins (categories). PennyPinchers for those with purchases  $\leq 5.0$  and HighRollers for those with purchases  $> 5.0$ . This categorical variable is appended as a new column as shown in the creation dialog.

The creation of this new categorical attribute was necessary to categorize (to label) current rows considering the purchased amount, and map this value into a desired types of users (PennyPinchers and HighRollers) in our case study.

### Attribute Selection

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

Attribute	Rationale for Filtering
avg_price	The categorical variable buyer_type has been generated based in this attribute, so considering it in our classification process would be redundant.
userId	This attribute identify the user saying who is the user (It can be repeated), but it does not influence the category a user could be considered based on their purchases.
sessionId	This attribute identify the session. Sessions do not influence how the user performs and also their purchase behavior.