

Data Preparation

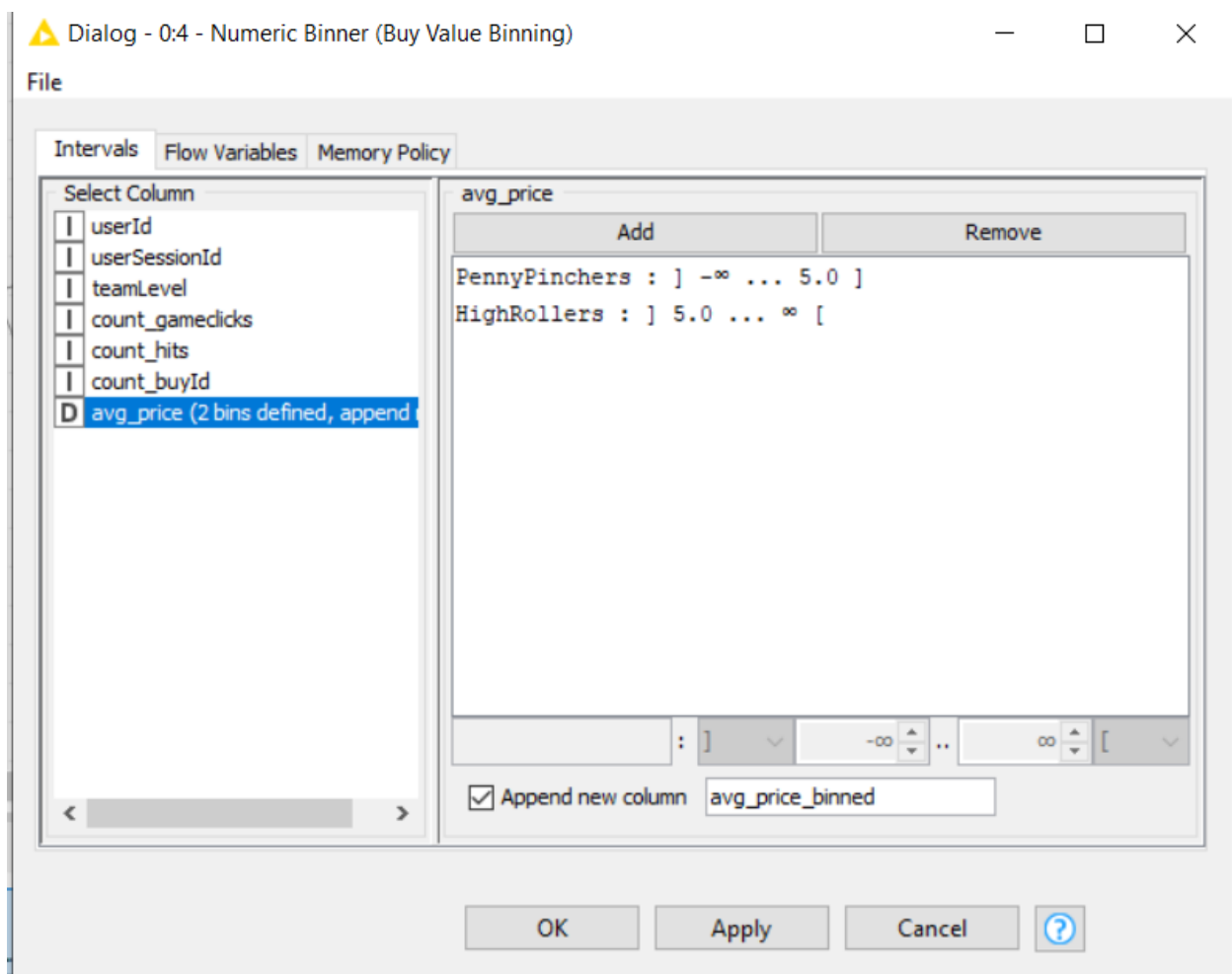
Analysis of combined_data.csv

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:



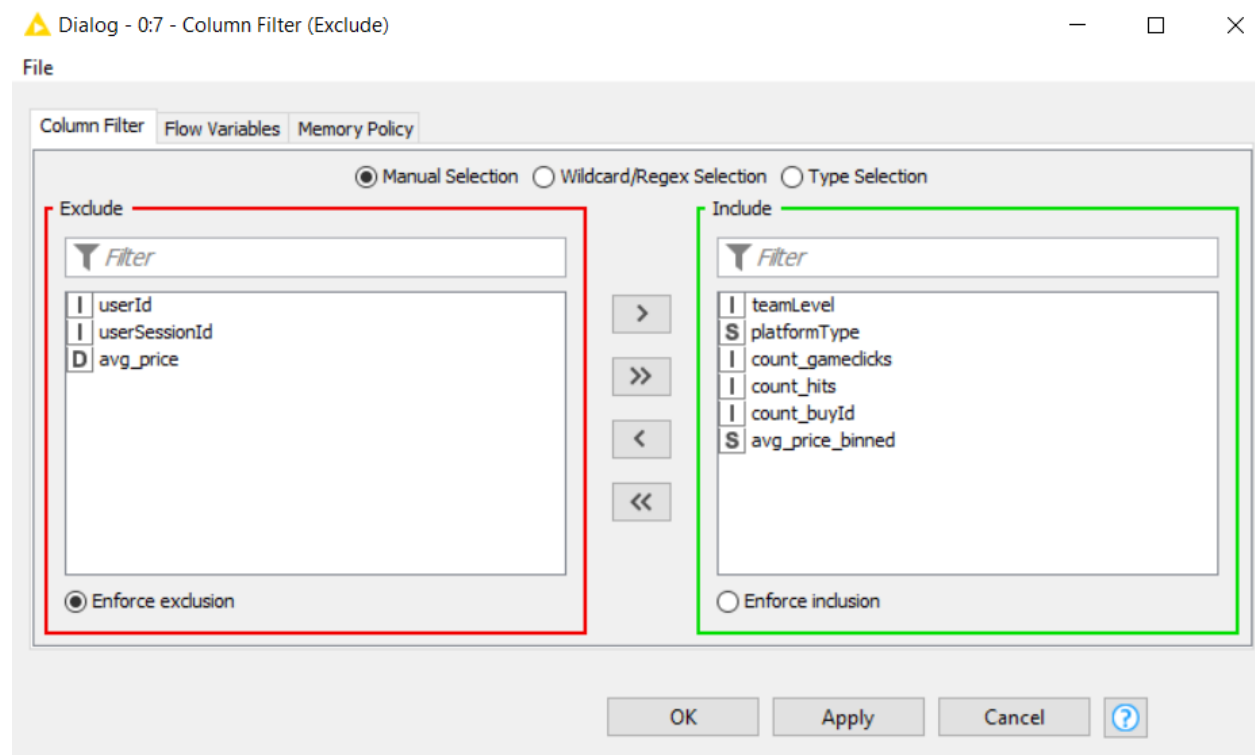
Created two classifications, One for PennyPinchers whose average spending is less than \$5.00 and one for HighRollers whose average spending is more than \$5.00. Note that PennyPinchers are also spenders, as we have filtered who did not spend at all!

The creation of this new categorical attribute was necessary because:

To separate PennyPinchers and Highrollers's buying power so that we can analyse them as two datasets.

Attribute Selection

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



Attribute	Rationale for Filtering
UserID	There is no reason to analyze a specific User. It is important that decisions we make are agnostic to a user.
UserSessionID	Same reason why Userid is filtered. On top of that reason, UserSessionID seems to have no effect on buying habits.
Avg_price	We have already have avg_binned attribute , hence we don't need this attribute.

