

Data Preparation

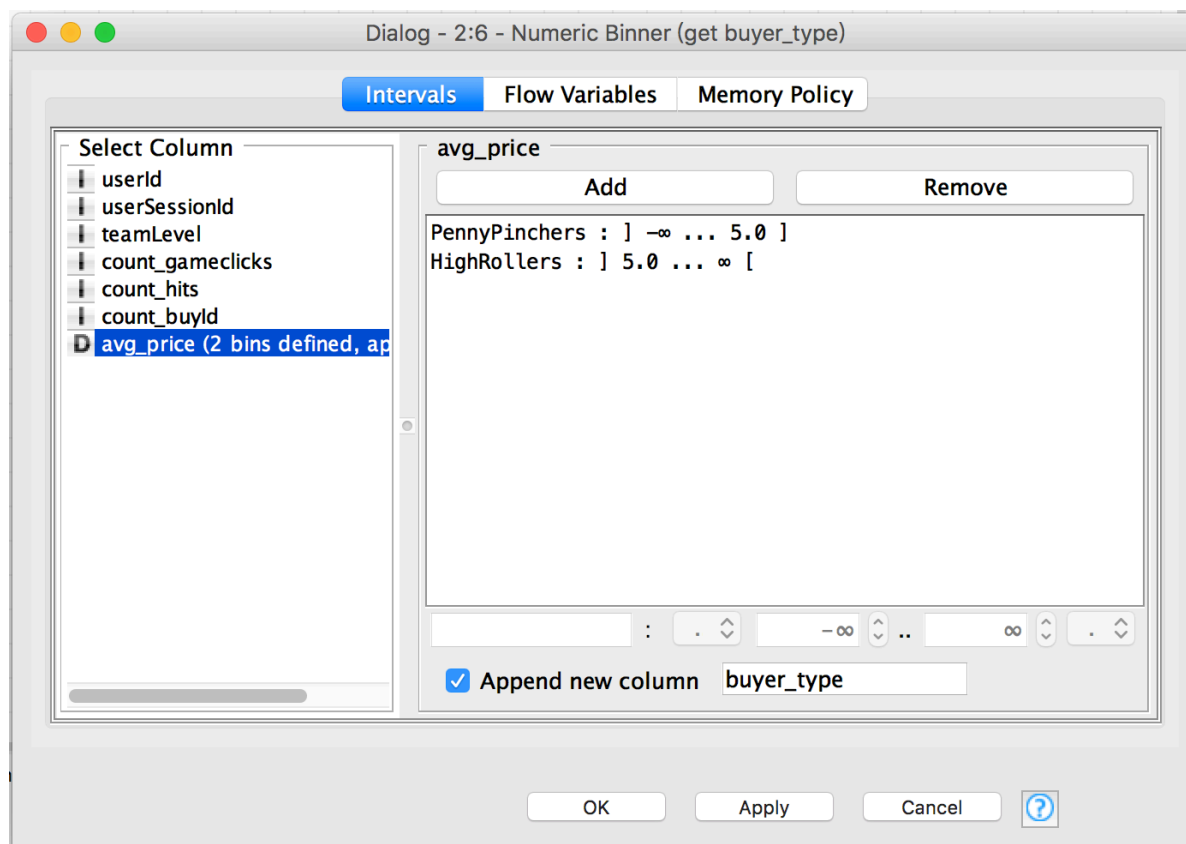
Analysis of combined_data.csv

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 column of avg_price represent average per orders of a user. If avg_price greater than 5, then the user is HighRollers. If avg_price less than or equal 5, then the user is PennyPinchers.

The creation of this new categorical attribute was necessary because.

The task is classifying users as HighRollers or PennyPinchers, we need categorical variate, instead of continuous variate.

Attribute Selection

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

Attribute	Rationale for Filtering
userID	It is uniquely user id, has no sense to classify.
userSession	It is uniquely user session id, has no sense to classify.
avg-price	The target label derived from it