**Data Preparation**

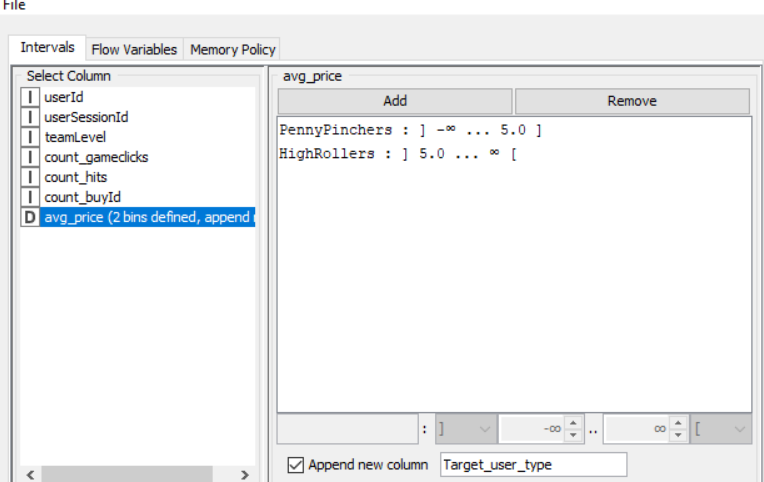
Analysis of combined\_data.csv

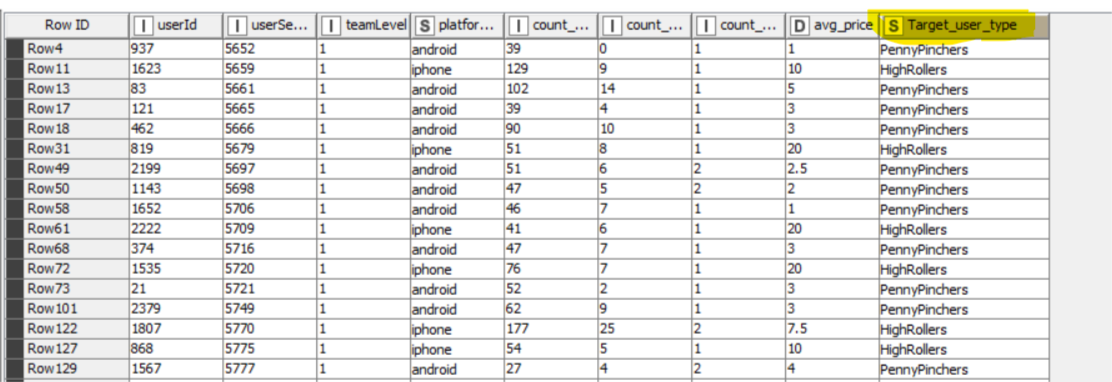
Sample Selection

|  |  |
| --- | --- |
| **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:





**Categorical attribute name:** *Target\_user\_type*

**Description:** It isderived by binning avg\_price attribute between HighRollers (buyers of items that cost more than $5.00) and PennyPinchers (buyers of items that cost $5.00 or less).

The new attribute will act as our target variable for training our classifier and segmenting the users between HighRollers and PennyPinchers. This classier is then used for predicting the type of new user (unseen data) in future.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| UserId | Id field |
| userSessionId | Id field |
| Avg\_price | Numerical field from which target field is created |