**Data Preparation**

Analysis of combined\_data.csv

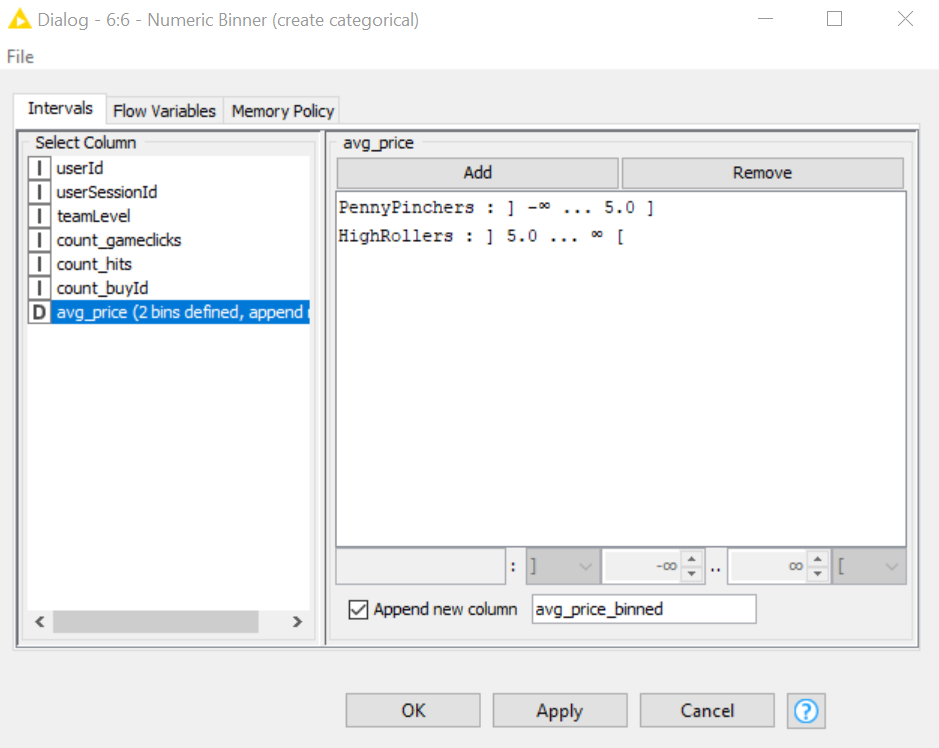
Sample Selection

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

<Fill In Screenshot>



<Fill In: Describe the design of your attribute in 1-3 sentences.>

**The new column is called avg\_price\_binned , and made of 2 categorical values : PennyPinchers paying less than $ 5 in average, and HighRollers paying more $ 5 in average.**

**We “Append new column” to use this new column in the flow.**

The creation of this new categorical attribute was necessary because <Fill in 1-2 sentences>.

**The goal of the present analytics exercise is to predict which users will fall HighRollers category , or in PennyPinchers category.**

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| **avg\_price** | **Correlated with the new column avg\_price\_binned** |
| **User\_id** | **No need, it’s a number generated by the computer** |
| **User\_session** | **No need, same reason like user\_id** |
| ~~<Optional Fill in>~~ | ~~<Optional Fill in 1-3 sentences>~~ |