**Data Preparation**

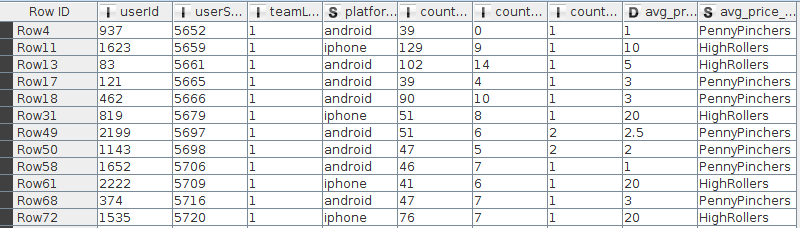
Analysis of combined\_data.csv

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 new attribute avg\_price\_binned uses the avg\_price attribute to classify the instances. When the value of avg\_prive is less than 5 it classify the instance as “PennyPinchers”. And when it is grater than 5 as “HighRollers”

The creation of this new categorical attribute was necessary because it will be the target attribute that we are going to use in the next steps to train the decision tree.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userId | Its excluded because its not a significance value, its only the Id to identify the session |
| userSesionId | Its excluded because its not a significance value, its only the Id to identify the session |
| avg\_price | Its excluded because the target attribute its created from this attribute |