**Data Preparation**

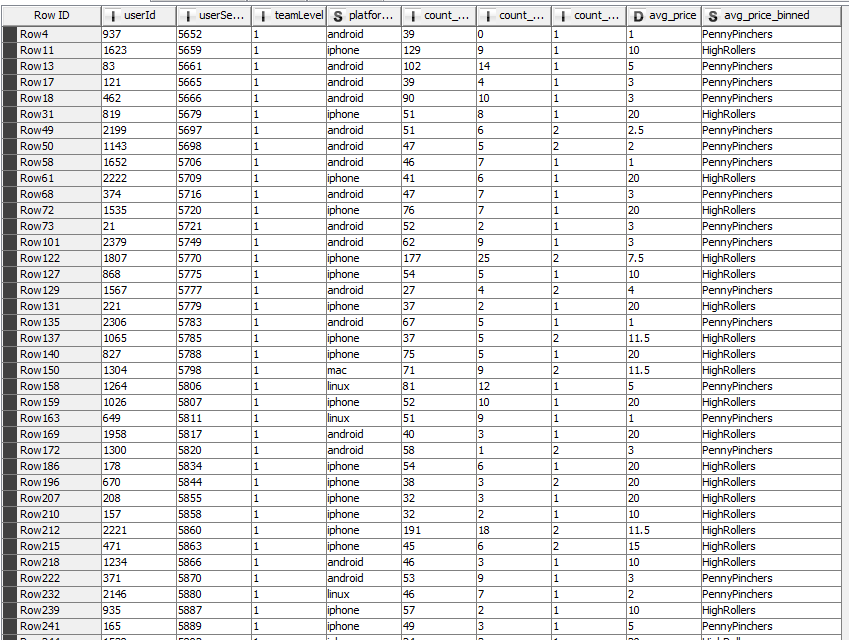
Analysis of combined\_data.csv

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 attribute is called avg\_price\_binned. The users who buy with average price above 5 are HighRollers and below or equal to 5 are PennyPinchers. It helps us to categorize users who buy more or buy expensive items from the less spending ones.

The creation of this new categorical attribute was necessary because **it helps determine which users have contributed most to the revenue of the company.**

.

Attribute Selection

The following attributes were filtered from the dataset for the following reasons: **It means I kept these attributes because rest didn’t make any special impact.**

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userId | It was filtered because it is the unique key for every single user and is thus always required to identify individual user. |
| platformType | Will help us in analyzing whether certain platform users make majority of “HighRollers” or “PennyPinchers” |
| avg\_price\_binned | This is the only dependent variable which is to be predicted and is thus required in the dataset so that predictions can be made later. |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |