**Data Preparation**

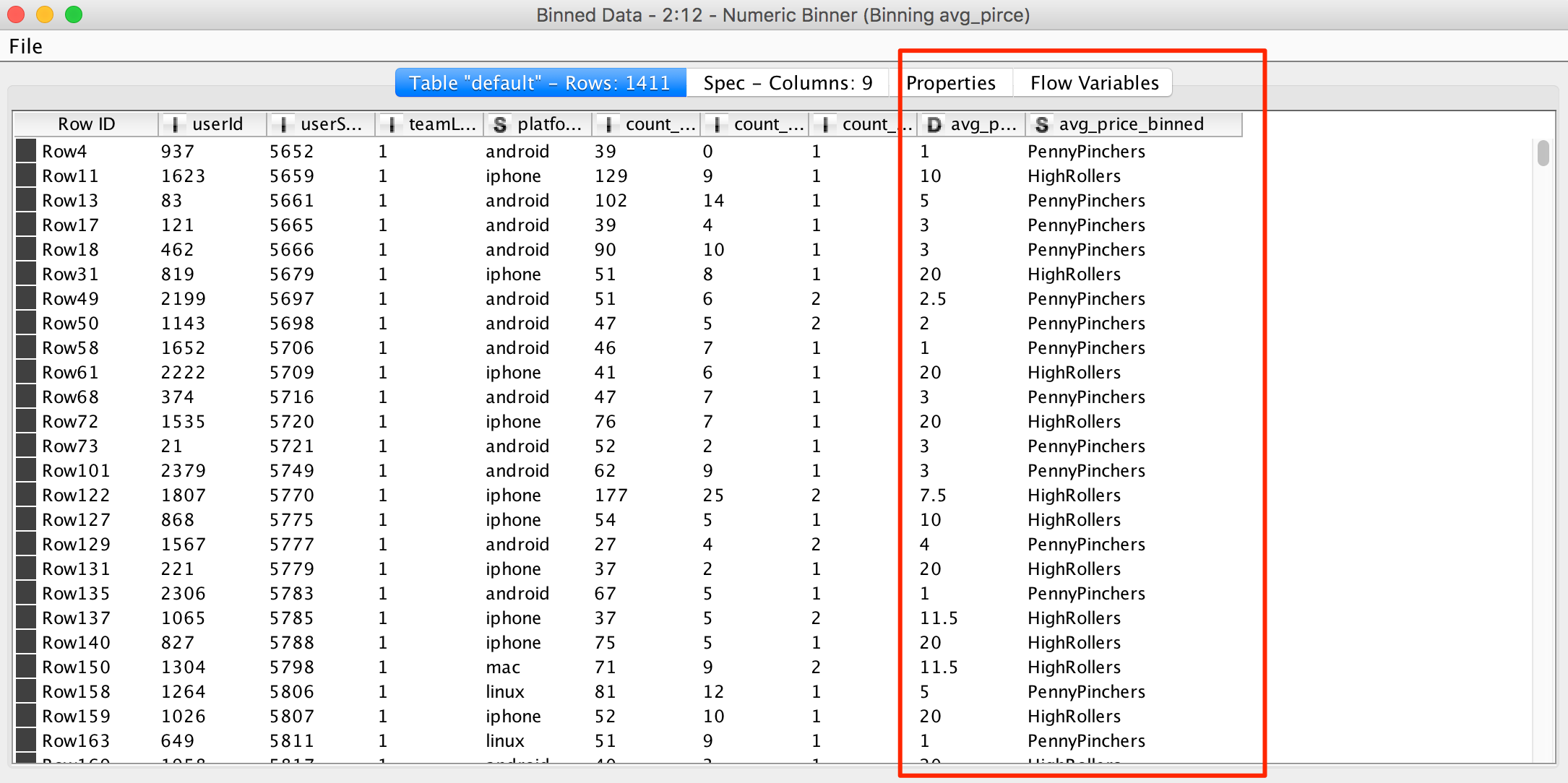
Analysis of combined\_data.csv

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 rows with average price more than 5$ are assigned the value of “HighRollers”, while ones with or less than 5$ are assigned the value of “PennyPinchers”.

The creation of this new categorical attribute was necessary because we are having a classification problem. And the label average price is of continuous value type. It’s necessary to transform it into categorical one.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userId | The index is useless as the attributes. |
| userSessionId | The index is useless as the attributes. |
| avg\_price | It is redundant now, since the binned average price has been used as the label. |
| <Optional Fill in> | <Optional Fill in 1-3 sentences> |