

## Data Preparation

Analysis of combined\_data.csv

### Sample Selection

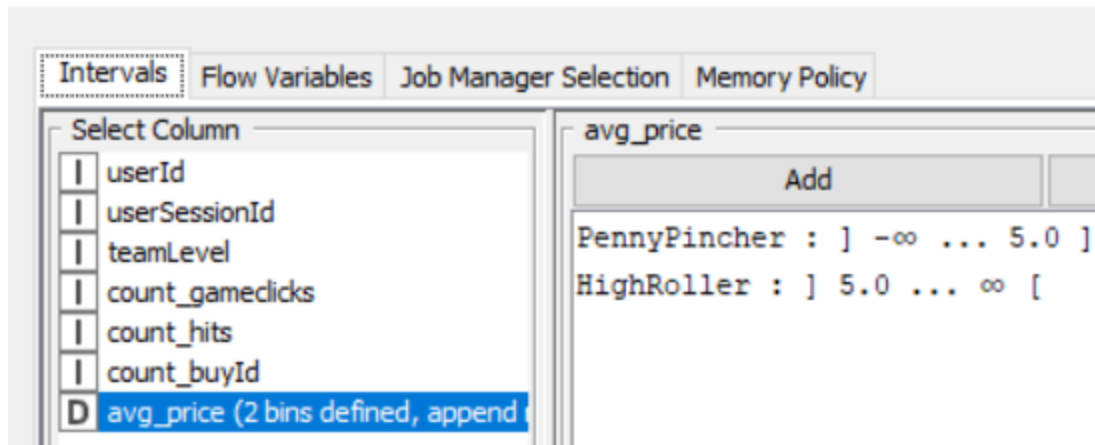
Item	Amount
# of Samples	4619
# of Samples with Purchases	1411

### 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:

▲ Dialog - 4:12 - Numeric Binner (Categorical Target)

File



A Categorical attribute named '*User\_Category*' is created with given conditions. PennyPincher are those with  $\text{avg\_price} \leq 5$ , and HighRoller are those with  $\text{avg\_price} > 5$ .

The creation of this new categorical attribute was necessary because solving a classification problem requires the target variable to be categorical but  $\text{avg\_price}$  is numerical.

### Attribute Selection

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

Attribute	Rationale for Filtering
avg_price	Target variable is derived from this attribute, so it can not become part of model features otherwise we'd see model overfitting
userId	Id field adds no information hence can be excluded from modelling
userSessionId	Id field adds no information hence can be excluded from modelling
<Optional Fill in>	<Optional Fill in 1-3 sentences>

## Data Partitioning and Modeling

The data was partitioned into train and test datasets.

The <training> data set was used to create the decision tree model.

The trained model was then applied to the <test> dataset.

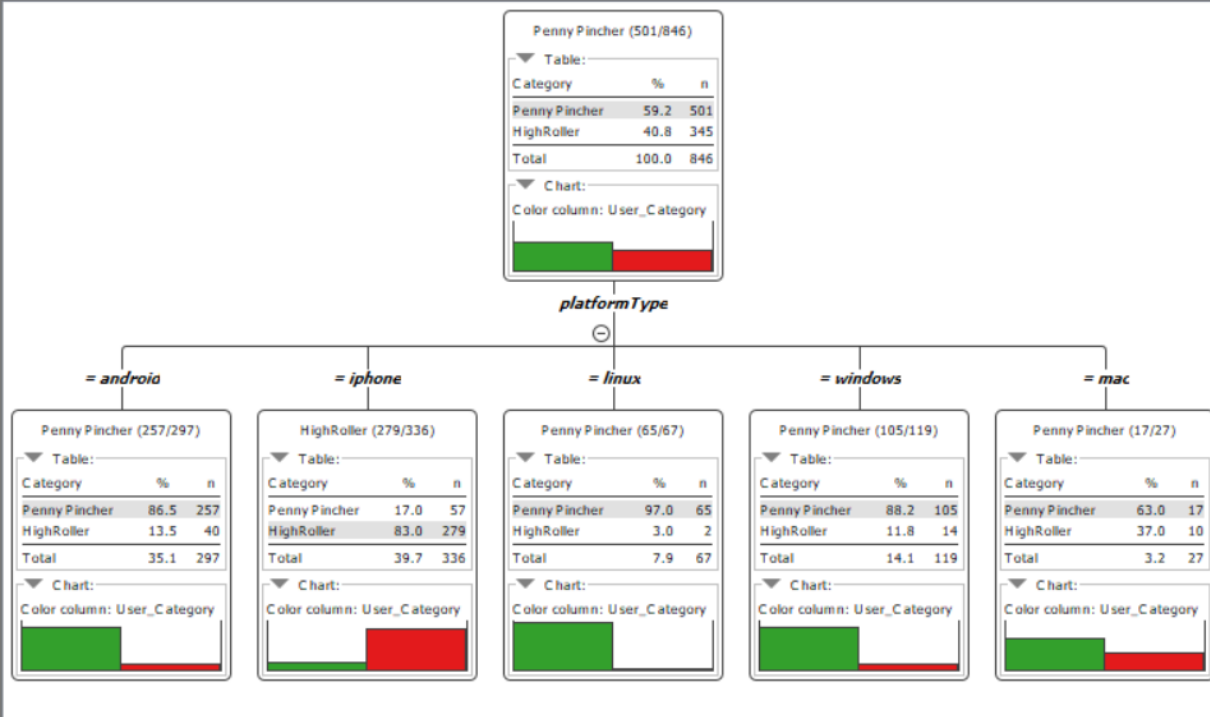
This is important because we need to evaluate the performance of our model on unseen dataset.

When partitioning the data using sampling, it is important to set the random seed because we want to be able to regenerate the same results while comparing with others.

A screenshot of the resulting decision tree can be seen below:

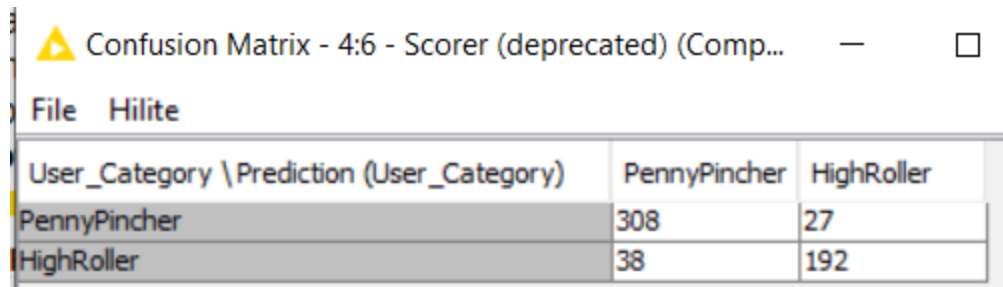
Decision Tree View - 4:10 - Decision Tree Learner (Train model)

File HiLite Tree



## Evaluation

A screenshot of the confusion matrix can be seen below:



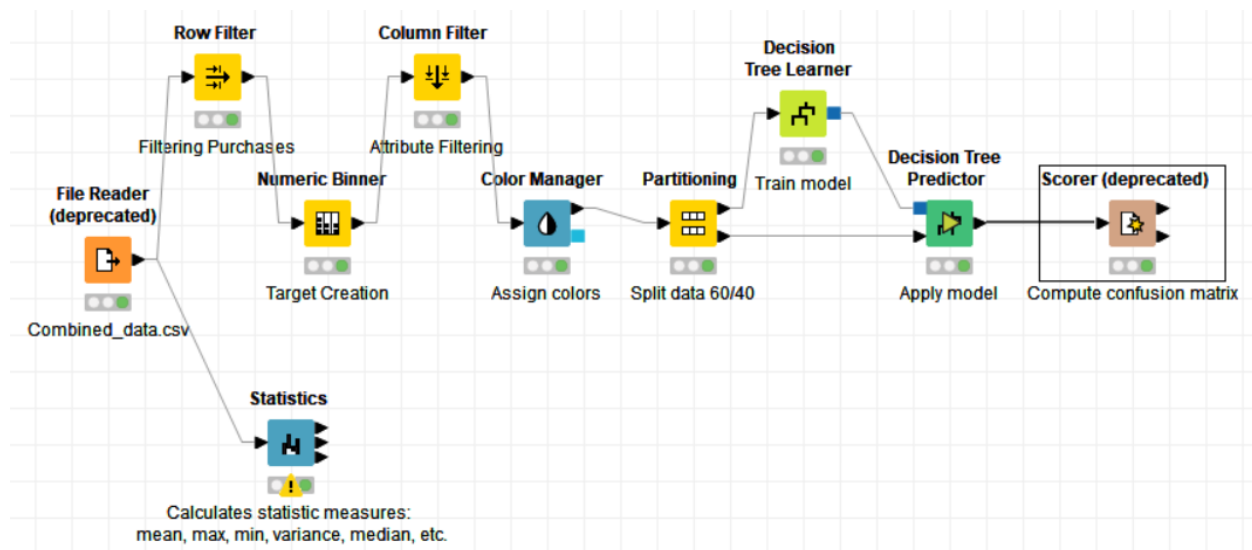
User_Category \ Prediction (User_Category)	PennyPincher	HighRoller
PennyPincher	308	27
HighRoller	38	192

As seen in the screenshot above, the overall accuracy of the model is <88.496%>

38 HighRollers were incorrectly predicted as PennyPincher by the Decision Tree model. Likewise, 27 PennyPinchers were incorrectly predicted as HighRoler. Decision Tree model correctly predicts 308 PennyPincher and 192 HighRoller. Overall, 500 accurate predictions with 65 incorrect predictions.

## Analysis Conclusions

The final KNIME workflow is shown below:



## What makes a HighRoller vs. a PennyPincher?

Users with iphone platformType are mostly HighRoller.

Users with Linux platformType are all (but two) PennyPinchers.

Mac users are more inclined to spend more than linux/windows/android.

Specific Recommendations to Increase Revenue
1. Iphone and Mac users may be targeted with expensive items to generate more revenue out of them
2. Other platform users can be targeted with a wider variety of cheaper options to generate

more revenue from them.