HOW TO INCREASE REVENUE IN THE 'CATCH THE FLAMINGO' GAME

Prepared by

Ken Wood

December 10, 2020

Data Preparation

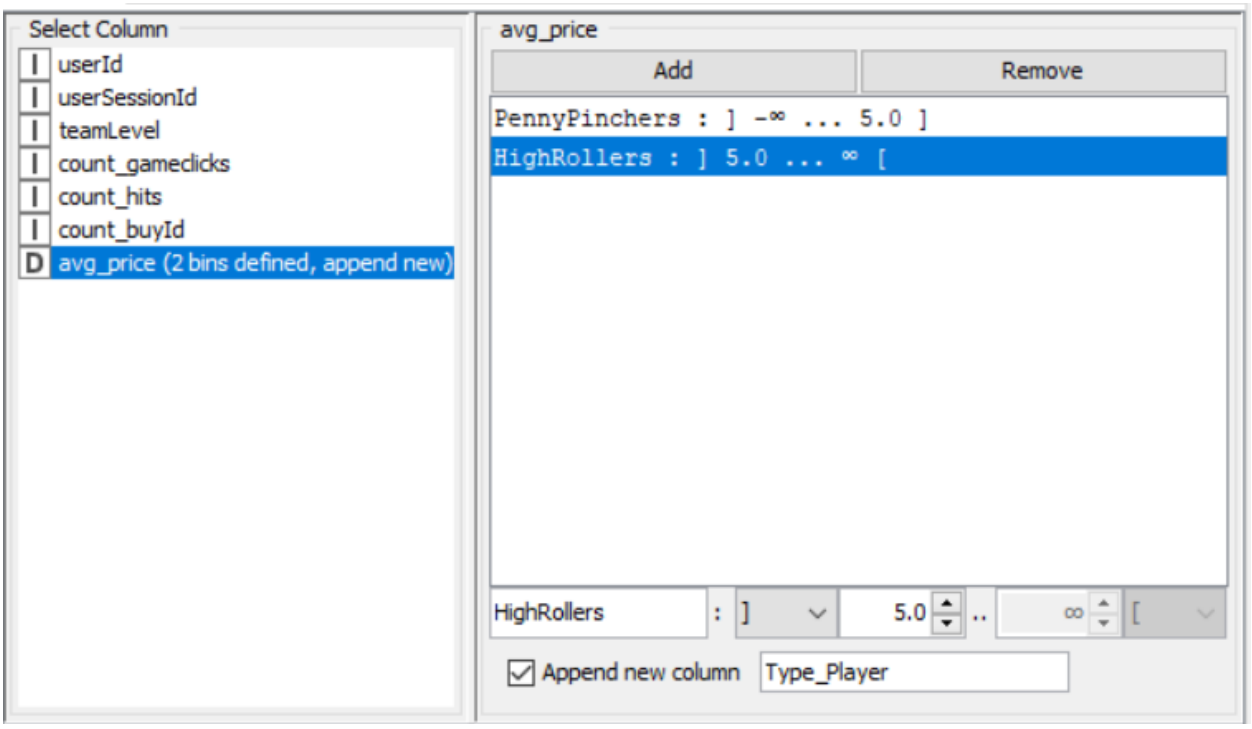
Analysis of combined_data.csv

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:



New column named “avg_price_binned” is the new attribute where avg_price > 5 belongs to “HighRollers” because the prices of them are over $5, while avg_price <= 5 belongs to “PennyPinchers” because the prices of those are not over $5.

Attribute Deletion

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

Attribute	Rationale for Filtering
avg_price	We don't need the average price anymore since we are coding it into a categorical variable
userId	Don't need this since it is just a computer-generated number
userSessionId	Don't need this since it is just a computer-generated number

Data Partitioning and Modeling

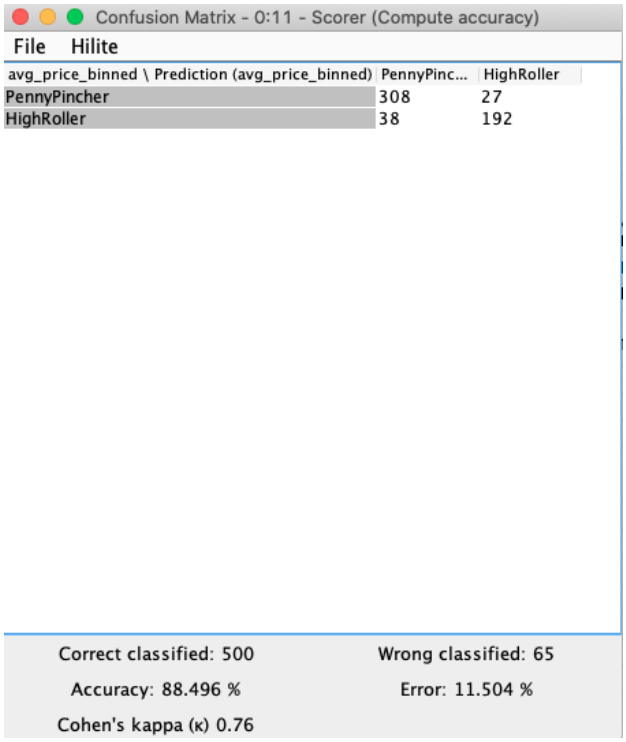
The data was partitioned into train and test datasets.
The train data set was used to create the decision tree model.
The trained model was then applied to the test dataset.
This is important because we want to ensure that we are not overfitting our Decision Tree Model on the train dataset. Running the model on the test data set allows us to determine whether our model generalizes well when introduced with new data.

When partitioning the data using sampling, it is important to set the random seed because we want to be able to obtain repeatable results.

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

Evaluation

A screenshot of the confusion matrix can be seen below:



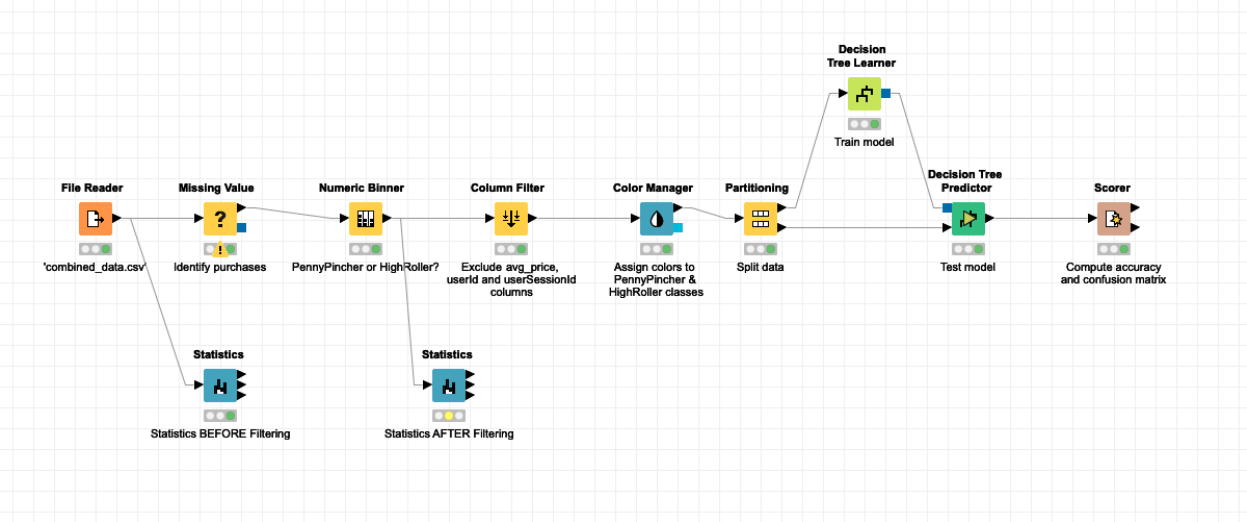
	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Therefore: TN = 308, FP = 27, FN = 38, TP = 192, total = 565

Accuracy = (TP + TN)/total = (192+308)/565 = 0.885
Error = (FP + FN)/total = (27+38)/565 = 0.115
Precision = TP/(TP+FP) = 192/(192+27) = 0.877
Recall = TP/(TP+FN) = 192/(192+38) = 0.835
F1 Score = 2(Precision*Recall)/(Precision+Recall) = 2*0.887*0.835/(0.877+0.835) = 0.865

Analysis Conclusions

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

Looking at the decision tree created by the model, HighRollers mostly use the iPhone as their platform to play the game, followed (distantly) by Android. PennyPinchers are more prevalent on desktops or laptops.

Overall, we can see that more valuable purchases are made on mobile devices rather than desktops or laptops.

Specific Recommendations to Increase Revenue
1. Create a marketing initiative that entices PennyPinchers to play the game on a mobile device.
2. Build out more game features on mobile platforms to maintain a continued and growing interest in the game.