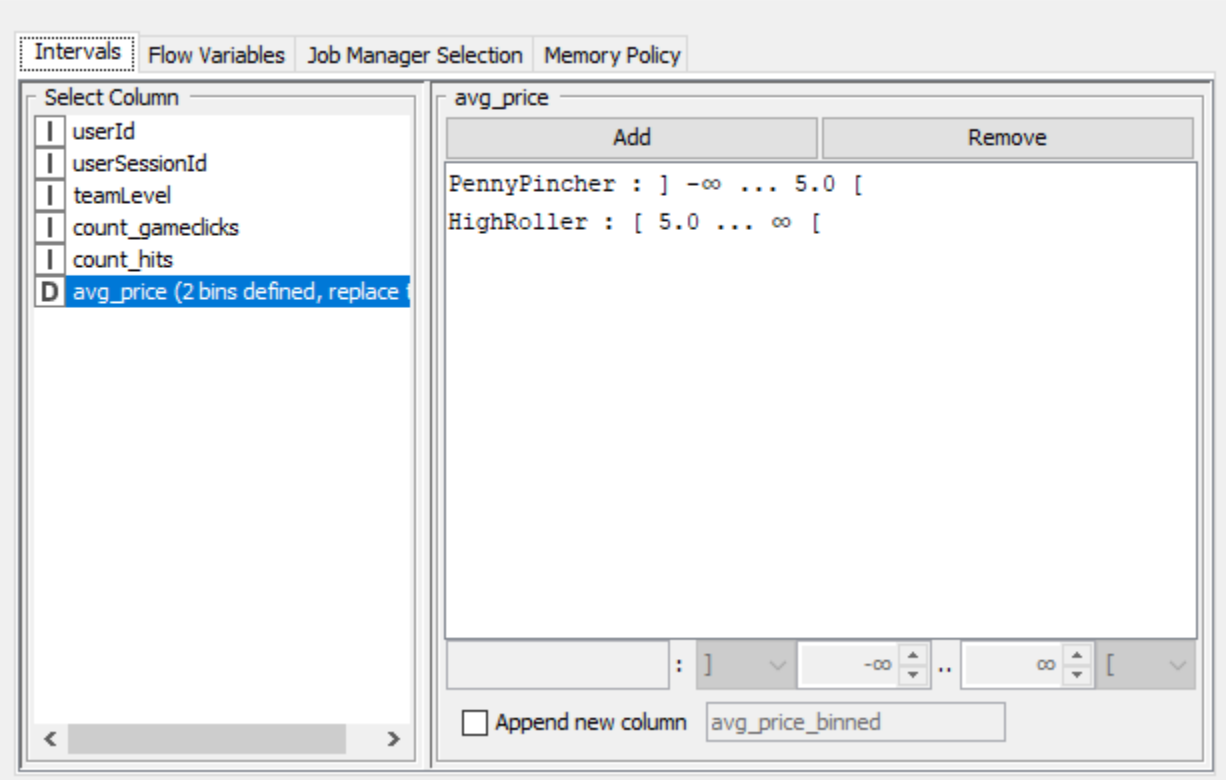
# Dara Preparation

### Sample Selection

| Item | Amount |
|---|---|
| **Samples** | 4619 |
| **Samples with purchases** | 1411 |

### Attribute Creation

A new categorical attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPitchers):



# Dara Partitioning and Modeling

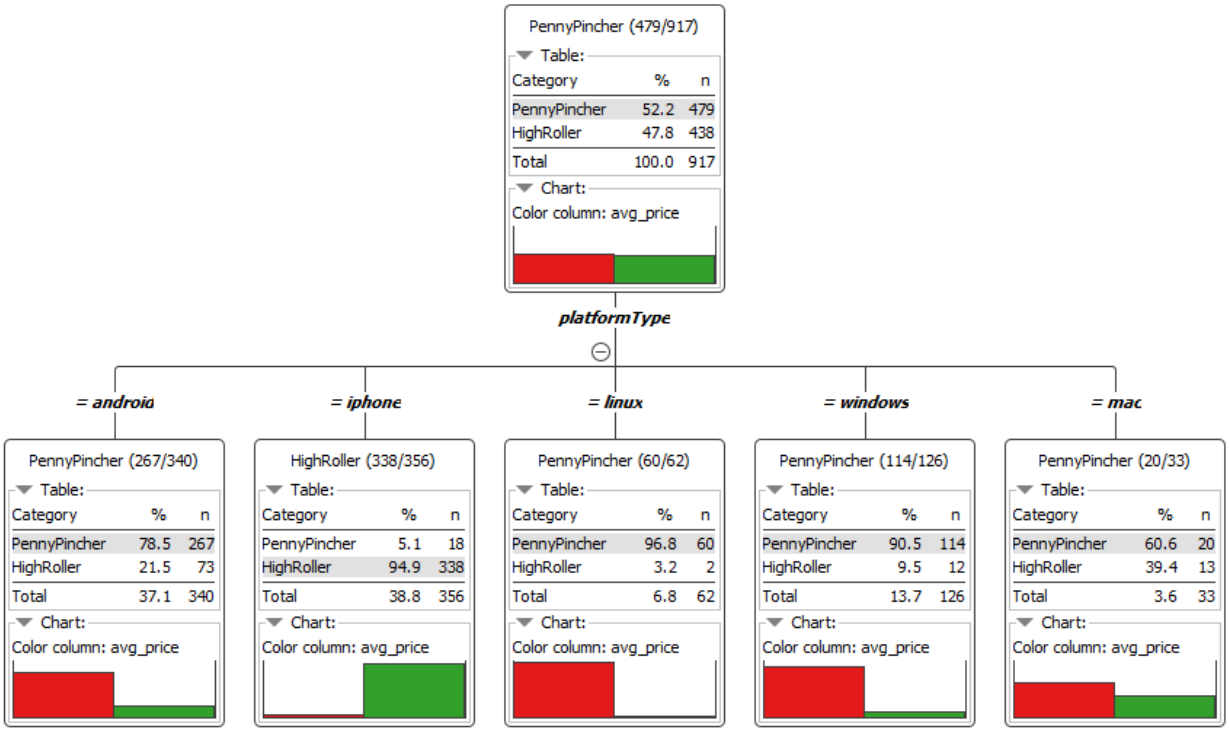
The data was partitioned into train and test datasets.

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

The trained model was then applied to the test dataset

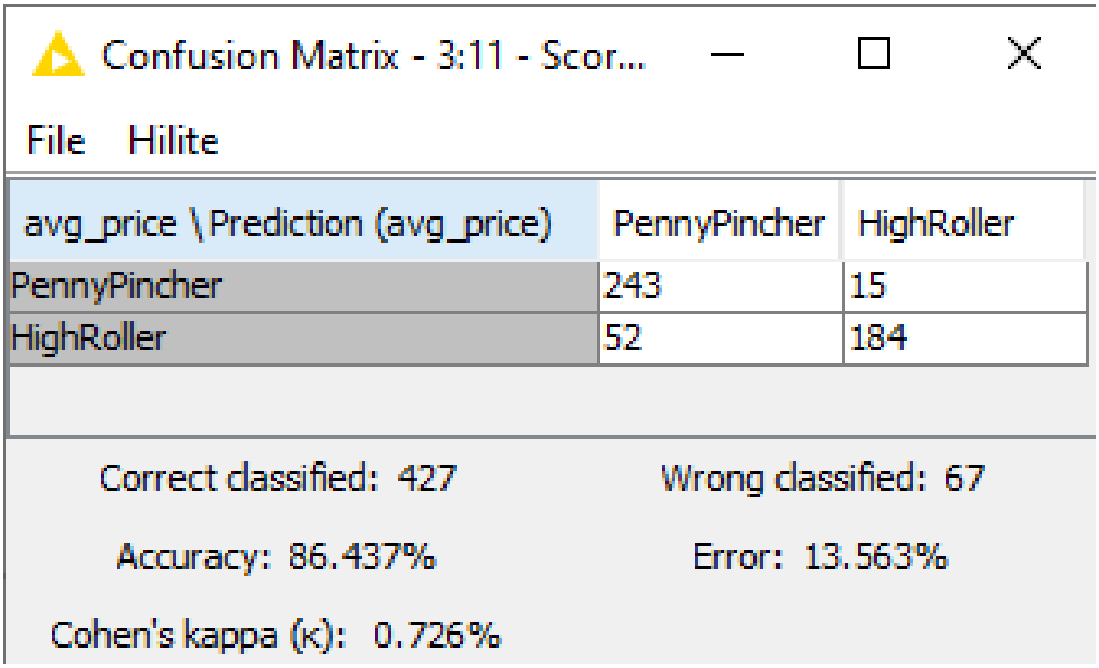
This is important because partitioning the data set into training and test data allows us to verify the accuracy of the trained model

When partitioning the data using sampling, it is important to set the random seed because it allows you to obtain reproducible results each time you run the partition



# Evaluation

Confusion matrix



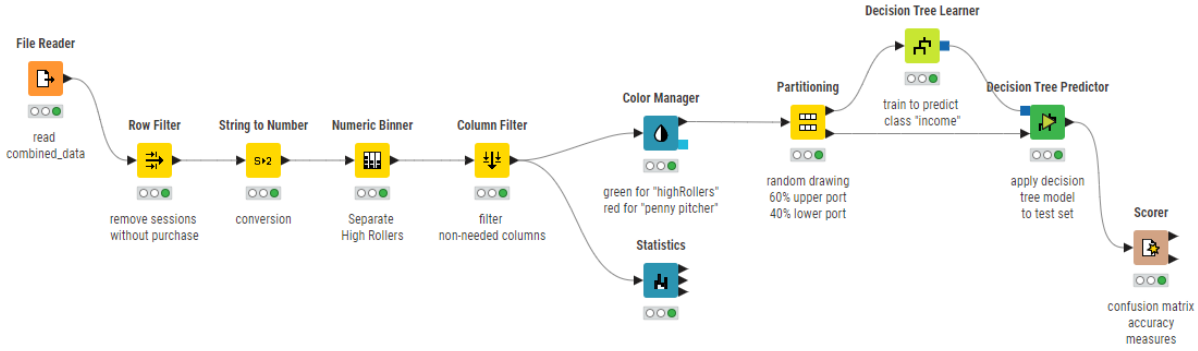

The accuracy of the model is 86.43%

When the buyer type is PennyPincher, the model classified correctly 243 times and incorrectly 15 times. When the byer type is HighRoller, the model classified correctly 192 times and incorrectly 52 times.

Coursera Big Data Specialization Capstone Project, Week 2
**Peer Graded Assignment: Classifying in KNIME to identify big Spenders in Catch the Pink Flamingo**

# Analysis Conclusions

The final KNIME workflow:




What makes a HighRoller vs PennyPincher?

The OS used by user. Users who are highRoller use IOS, PennyPincher use android, Mac, Windows, Linux

Specific Recommendations to increase Revenue

1. Target promotions to IOS users
2. Target future product develop towards IOS.

Michael Ventura