Data Preparation

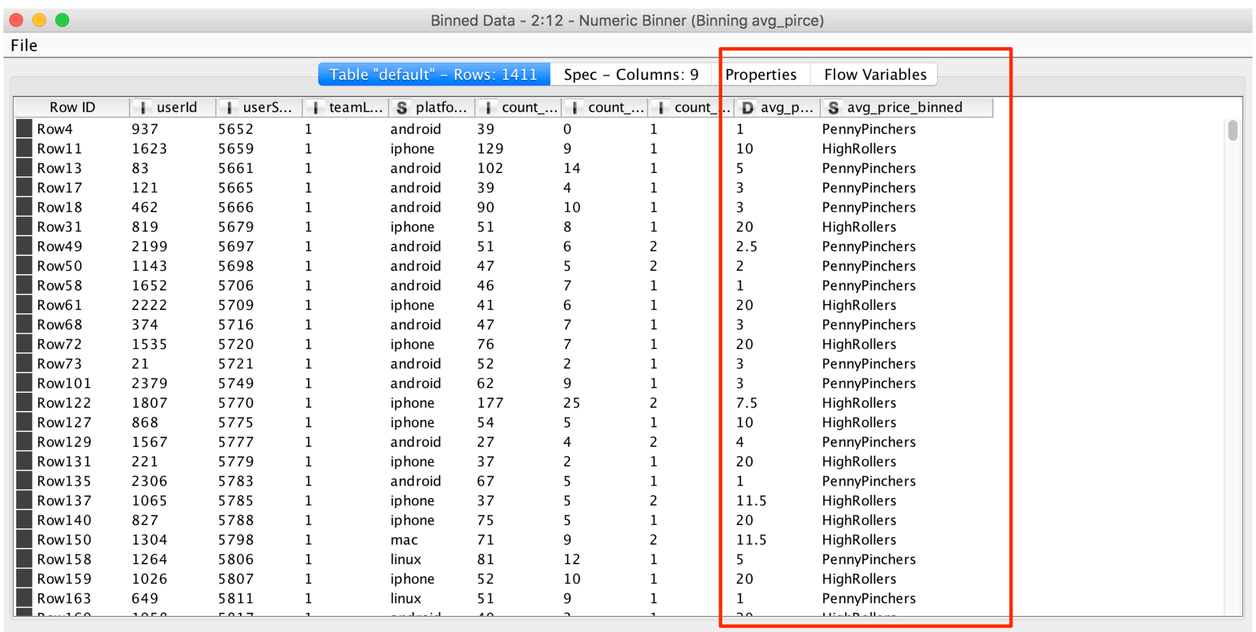
Analysis of combined_data.csv

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:



Row ID	userId	userS...	teamL...	platfo...	count_...	count_...	count_...	avg_p...	avg_price_binned
Row4	937	5652	1	android	39	0	1	1	PennyPinchers
Row11	1623	5659	1	iphone	129	9	1	10	HighRollers
Row13	83	5661	1	android	102	14	1	5	PennyPinchers
Row17	121	5665	1	android	39	4	1	3	PennyPinchers
Row18	462	5666	1	android	90	10	1	3	PennyPinchers
Row31	819	5679	1	iphone	51	8	1	20	HighRollers
Row49	2199	5697	1	android	51	6	2	2.5	PennyPinchers
Row50	1143	5698	1	android	47	5	2	2	PennyPinchers
Row58	1652	5706	1	android	46	7	1	1	PennyPinchers
Row61	2222	5709	1	iphone	41	6	1	20	HighRollers
Row68	374	5716	1	android	47	7	1	3	PennyPinchers
Row72	1535	5720	1	iphone	76	7	1	20	HighRollers
Row73	21	5721	1	android	52	2	1	3	PennyPinchers
Row101	2379	5749	1	android	62	9	1	3	PennyPinchers
Row122	1807	5770	1	iphone	177	25	2	7.5	HighRollers
Row127	868	5775	1	iphone	54	5	1	10	HighRollers
Row129	1567	5777	1	android	27	4	2	4	PennyPinchers
Row131	221	5779	1	iphone	37	2	1	20	HighRollers
Row135	2306	5783	1	android	67	5	1	1	PennyPinchers
Row137	1065	5785	1	iphone	37	5	2	11.5	HighRollers
Row140	827	5788	1	iphone	75	5	1	20	HighRollers
Row150	1304	5798	1	mac	71	9	2	11.5	HighRollers
Row158	1264	5806	1	linux	81	12	1	5	PennyPinchers
Row159	1026	5807	1	iphone	52	10	1	20	HighRollers
Row163	649	5811	1	linux	51	9	1	1	PennyPinchers

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>

Data Partitioning and Modeling

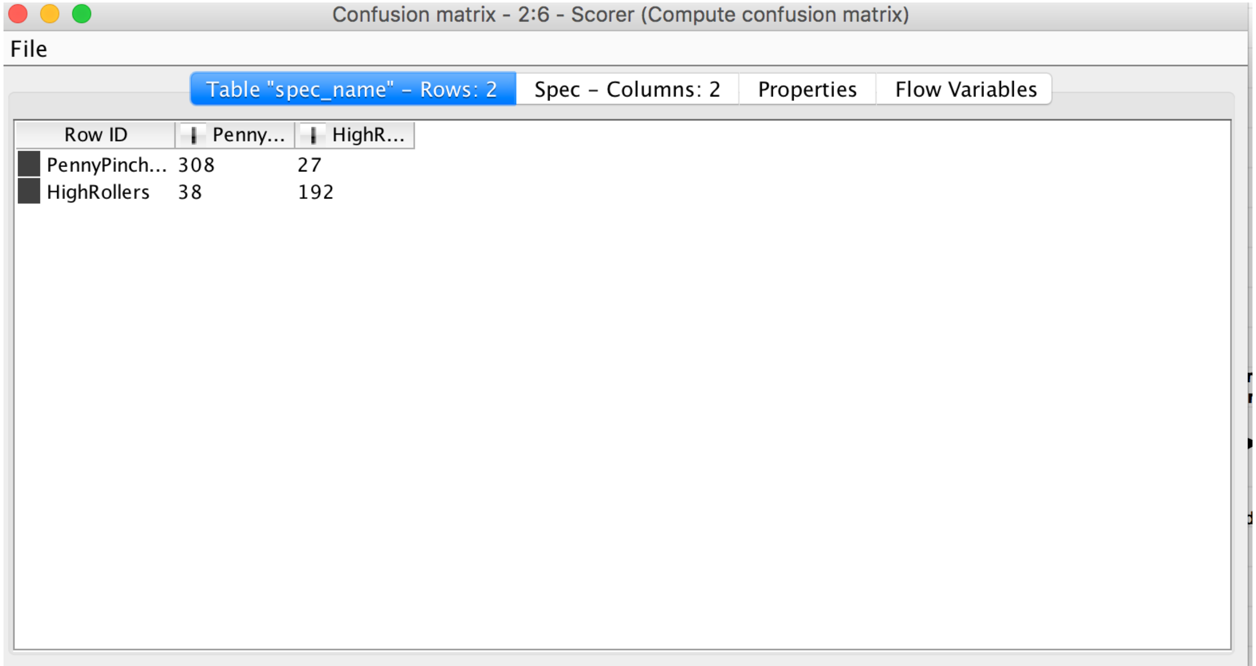
The data was partitioned into train and test datasets.
The training data set was used to create the decision tree model.
The trained model was then applied to the test dataset.
This is important because it could avoid overfitting. If we train all data to create the model and test it using the same data. The model is to memorize the data and unable to handle the unknown situation, which leads to overfitting.

When partitioning the data using sampling, it is important to set the random seed because we need the modeling results to be reproducible so that our conclusion could be persuasive.

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

Evaluation

A screenshot of the confusion matrix can be seen below:



Row ID	Penny...	HighR...
PennyPinch...	308	27
HighRollers	38	192

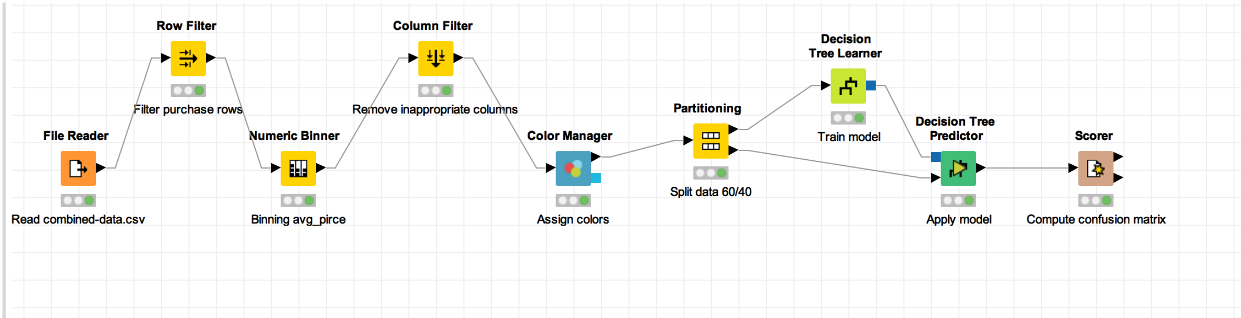
As seen in the screenshot above, the overall accuracy of the model is 88.5%

<Fill In: Write one sentence for each of the values of the confusion matrix indicating what has been correctly or incorrectly predicted.>

308: 308 true PennyPincher are correctly predicted as PennyPinchers
27: 27 true PennyPincher are incorrectly predicted as HighRollers
38: 38 true Highroller are incorrectly predicted as PennyPinchers
192: 192 true Highrollers are correctly predicted as HighRollers

Analysis Conclusions

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?
<Fill In 2-3 sentences answering this question based on insights from your analysis.>

1. The platformType makes the type of buyer type, and mobile platforms contributes more than PC platforms.
2. In mobile platform, iphone players are more likely to be HighRoller(83%), while android players tend to be PennyPinchers(86.5%)
3. In PC platform, players are generally PennyPinchers, but part of mac users are HighRollers(37%).

Specific Recommendations to Increase Revenue
1. Android and Windows are two big user group to develop more HighRollers.
2. Considering the potential HighRoller group in mac platform, it's worth investing to attract more players from mac platform.