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 attributes follows:

Binned Data - 0:11 - Numeric Binner(Categorize HighRollers)

File

Table "default" - Rows: 1411 | Spec - Columns: 9 | Properties | Flow Variables

Row ID	userId	userSe...	teamLevel	platfor...	count_...	count_...	count_...	avg_price	avg_pri...
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
Row169	1958	5817	1	android	40	3	1	20	HighRollers
Row172	1300	5820	1	android	58	1	2	3	PennyPinchers
Row186	178	5834	1	iphone	54	6	1	20	HighRollers
Row196	670	5844	1	iphone	38	3	2	20	HighRollers
Row207	208	5855	1	iphone	32	3	1	20	HighRollers
Row210	157	5858	1	iphone	32	2	1	10	HighRollers
Row212	2221	5860	1	iphone	191	18	2	11.5	HighRollers
Row215	471	5863	1	iphone	45	6	2	15	HighRollers
Row218	1234	5866	1	android	46	3	1	10	HighRollers
Row222	371	5870	1	android	53	9	1	3	PennyPinchers
Row232	2146	5880	1	linux	46	7	1	2	PennyPinchers
Row239	935	5887	1	iphone	57	2	1	10	HighRollers
Row241	165	5889	1	iphone	49	3	1	5	PennyPinchers
Row244	1538	5892	1	iphone	24	3	1	20	HighRollers
Row245	1544	5893	1	iphone	36	6	2	20	HighRollers
Row261	2218	5909	1	android	80	6	1	3	PennyPinchers
Row262	1162	5910	1	windows	192	16	1	2	PennyPinchers
Row266	1821	5914	1	windows	178	22	1	1	PennyPinchers
Row271	2133	5919	1	android	87	14	1	3	PennyPinchers
Row272	1027	5920	1	iphone	52	5	3	15	HighRollers
Row273	518	5921	1	linux	121	16	1	1	PennyPinchers
Row282	2029	5930	1	iphone	89	7	1	10	HighRollers
Row286	2384	5934	1	windows	41	5	1	1	PennyPinchers
Row290	1155	5938	1	iphone	71	16	1	20	HighRollers
Row292	564	5940	1	linux	34	3	1	1	PennyPinchers
Row293	97	5941	1	android	74	10	1	1	PennyPinchers
Row297	253	5945	1	android	66	12	1	3	PennyPinchers
Row302	934	5950	1	windows	43	4	1	10	HighRollers
Row307	2009	5955	1	iphone	27	4	1	5	PennyPinchers
Row324	1815	6088	1	iphone	272	24	1	3	PennyPinchers
Row325	864	6173	1	android	79	9	1	3	PennyPinchers

The avg_price_binned attribute was created by binning on the avg_price attribute and distinguishes between “HighRollers” and “PennyPinchers.” HighRollers are buyers of items that cost more than $5.00. PennyPinchers are buyers of items that cost $5.00 or less.

The creation of this new categorical attribute was necessary because we want to separate players into two categories based on their purchasing behavior.

Attribute Selection

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

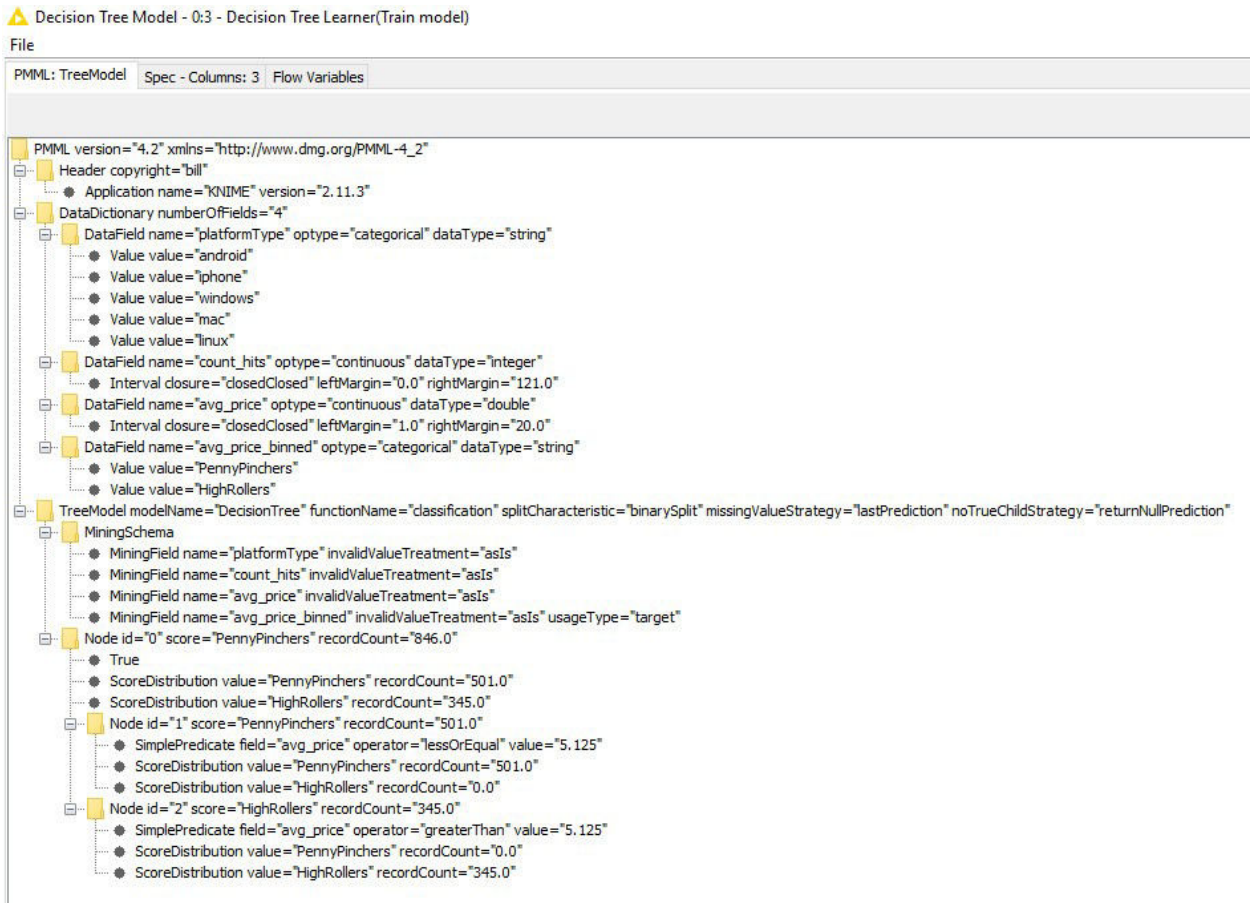
Attribute	**Rationale for Filtering**
platformType	To identify the platform most used by HighRollers and PennyPinchers
count_hits	To identify the number of game hits achieved by HighRollers and PennyPinchers
avg_price	To identify HighRollers and PennyPinchers

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 test our model on data that was not used to train (i.e., create) it.

When partitioning the data using sampling, it is important to set the random seed because we want to get reproducible results.

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



Evaluation

A screenshot of the confusion matrix can be seen below:

Confusion matrix - 0:6 - Scorer(Compute confusion matrix)

File

Table "spec_name" - Rows: 2 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	PennyPinchers	HighRollers
PennyPinchers	335	0
HighRollers	0	230

As seen in the screenshot above, the overall accuracy of the model is 100%

A screenshot of the accuracy statistics can be seen below:

Accuracy statistics - 0:6 - Scorer(Compute confusion matrix)

File

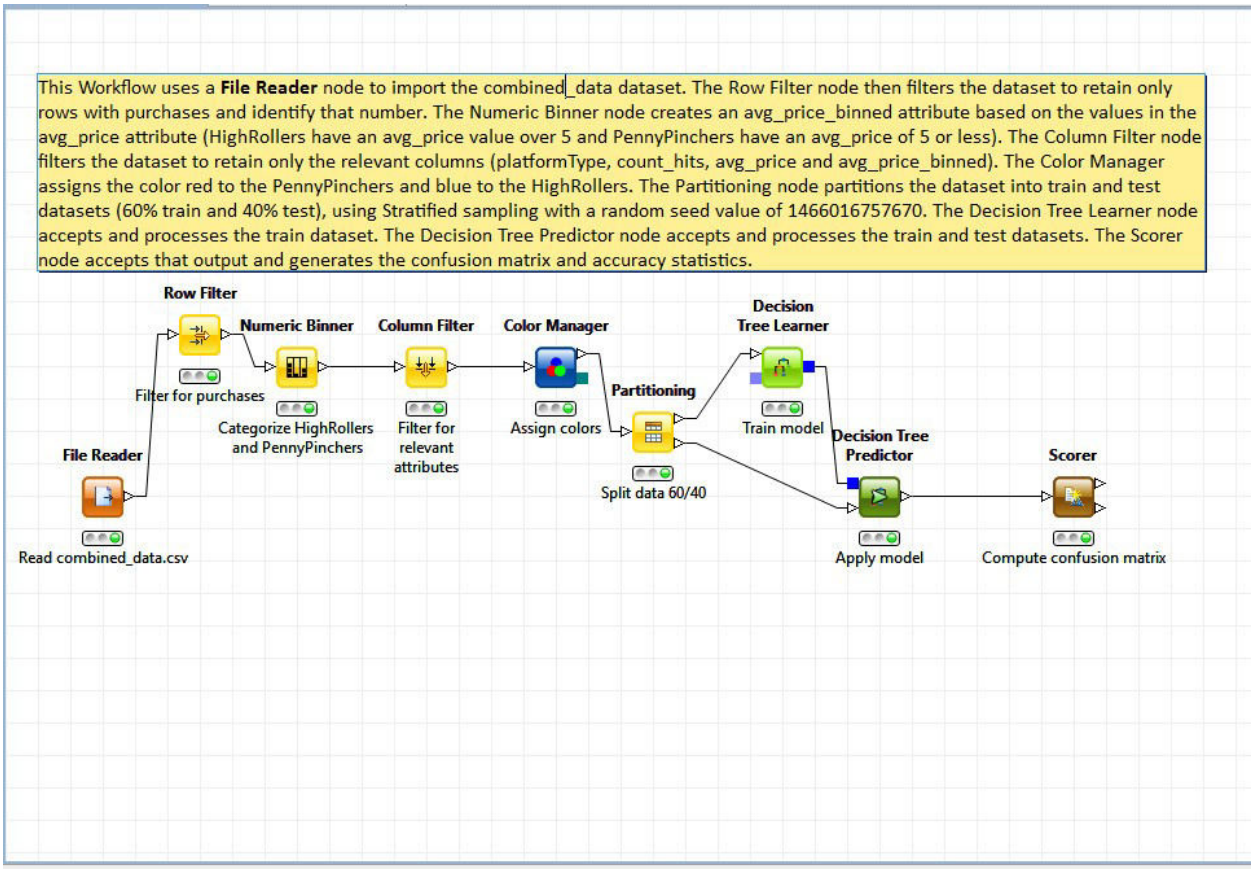
Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specifity	F-measure	Accuracy	Cohen's kappa
PennyPinchers	335	0	230	0	1	1	1	1	1	?	?
HighRollers	230	0	335	0	1	1	1	1	1	?	?
Overall	?	?	?	?	?	?	?	?	?	1	1

There are 335 PennyPinchers and 0 HighRollers for the PennyPinchers category. There are 230 HighRollers and 0 PennyPinchers for the HighRollers category.

Analysis Conclusions

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

We defined a HighRoller as a user who buys items that cost more than $5.00 and a PennyPincher as a user who buys items that cost $5.00 or less.

Building on our observations from the Week1Technical Appendix, where we found the top 3 spenders use the iPhone platform, we can now confirm that the majority of HighRollers use the iPhone platform according to the Classified Data found in the Decision Tree Predictor. We did not find any reasonable correlation between count_hits and HighRollers or PennyPinchers in this analysis.

A screenshot of the Classified Data can be seen below.

Table "default" - Rows: 565 | Spec - Columns: 5 | Properties | Flow Variables

Row ID	platfor...	count_...	avg_price	avg_pri...	Predicti...
Row17	android	4	3	PennyPinchers	PennyPinchers
Row58	android	7	1	PennyPinchers	PennyPinchers
Row61	iphone	6	20	HighRollers	HighRollers
Row73	android	2	3	PennyPinchers	PennyPinchers
Row101	android	9	3	PennyPinchers	PennyPinchers
Row127	iphone	5	10	HighRollers	HighRollers
Row129	android	4	4	PennyPinchers	PennyPinchers
Row135	android	5	1	PennyPinchers	PennyPinchers
Row150	mac	9	11.5	HighRollers	HighRollers
Row158	linux	12	5	PennyPinchers	PennyPinchers
Row159	iphone	10	20	HighRollers	HighRollers
Row172	android	1	3	PennyPinchers	PennyPinchers
Row207	iphone	3	20	HighRollers	HighRollers
Row212	iphone	18	11.5	HighRollers	HighRollers
Row232	linux	7	2	PennyPinchers	PennyPinchers
Row241	iphone	3	5	PennyPinchers	PennyPinchers
Row245	iphone	6	20	HighRollers	HighRollers
Row262	windows	16	2	PennyPinchers	PennyPinchers
Row282	iphone	7	10	HighRollers	HighRollers
Row290	iphone	16	20	HighRollers	HighRollers
Row293	android	10	1	PennyPinchers	PennyPinchers
Row325	android	9	3	PennyPinchers	PennyPinchers
Row328	windows	4	1	PennyPinchers	PennyPinchers
Row344	windows	1	3	PennyPinchers	PennyPinchers
Row349	iphone	1	10	HighRollers	HighRollers
Row358	iphone	16	20	HighRollers	HighRollers
Row368	android	23	2	PennyPinchers	PennyPinchers
Row369	windows	5	5	PennyPinchers	PennyPinchers
Row371	android	13	3	PennyPinchers	PennyPinchers
Row380	iphone	5	6	HighRollers	HighRollers
Row386	iphone	11	7.5	HighRollers	HighRollers
Row387	iphone	6	3	PennyPinchers	PennyPinchers
Row391	linux	7	1	PennyPinchers	PennyPinchers
Row392	iphone	2	12.5	HighRollers	HighRollers
Row395	iphone	22	15	HighRollers	HighRollers
Row408	android	3	20	HighRollers	HighRollers
Row440	android	5	11.5	HighRollers	HighRollers
Row441	windows	5	2	PennyPinchers	PennyPinchers
Row450	linux	8	2	PennyPinchers	PennyPinchers
Row455	iphone	7	5	PennyPinchers	PennyPinchers
Row466	iphone	3	20	HighRollers	HighRollers
Row482	android	11	3	PennyPinchers	PennyPinchers
Row530	android	9	6.5	HighRollers	HighRollers
Row533	android	4	3	PennyPinchers	PennyPinchers
Row541	windows	10	1	PennyPinchers	PennyPinchers
Row553	iphone	8	20	HighRollers	HighRollers
Row561	android	11	3	PennyPinchers	PennyPinchers
Row582	iphone	2	20	HighRollers	HighRollers
Row589	windows	28	10	HighRollers	HighRollers
Row605	linux	18	1	PennyPinchers	PennyPinchers
Row616	windows	12	1	PennyPinchers	PennyPinchers
Row636	windows	30	1	PennyPinchers	PennyPinchers
Row641	linux	11	1	PennyPinchers	PennyPinchers
Row646	mac	4	10	HighRollers	HighRollers
Row650	android	14	3	PennyPinchers	PennyPinchers
Row653	windows	7	1	PennyPinchers	PennyPinchers

Specific Recommendations to Increase Revenue
1. Given that the HighRollers typically use iPhones, it would be beneficial to offer more ads for iPhone-related products in general.
2. Assuming it is possible to generate targeted ads, offering ads for in-game purchases (e.g., binoculars) to players with low count_hits should increase revenue.