Data Preparation

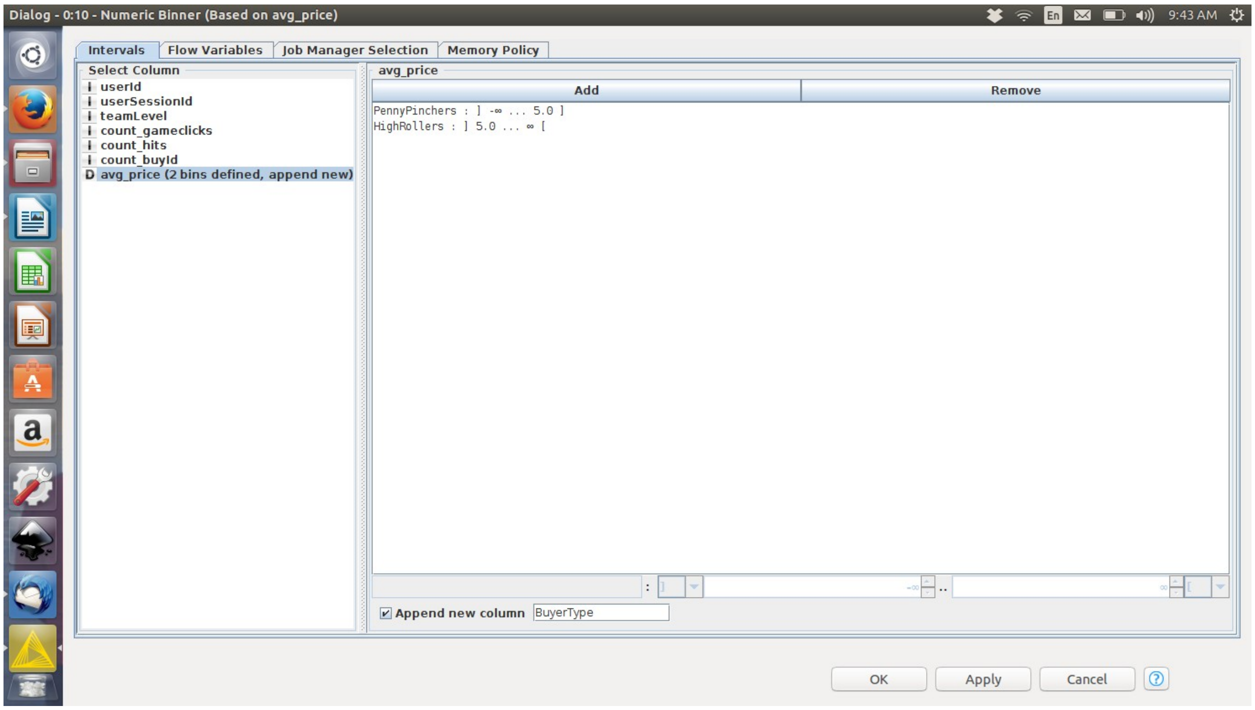
Analysis of combined_data.csv

Sample Selection

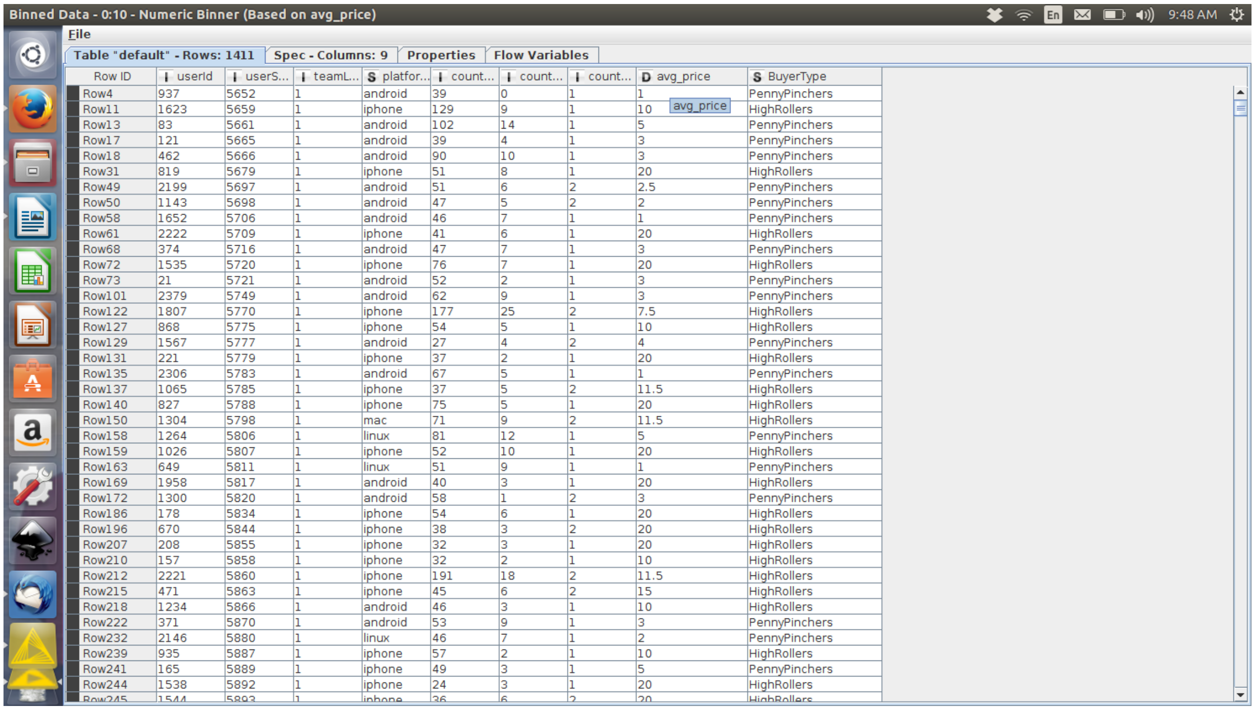
Item	Amount
# of Samples	4619
# of Samples with Purchases	1411

Attribute Creation

A new categorical attribute BuyerType was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers). A screenshot of the attribute follows:



Based on avg_price , another categorical vaiable named BuyerType is created . BuyerType is PennyPinchers if avg_price is less than or equal to 5. If avg_price is greater than 5 , its value is HighRollers.



Binned Data - 0:10 - Numeric Binner (Based on avg_price)

File

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

Row ID	userId	userS...	teamL...	platfor...	count...	count...	count...	avg_price	BuyerType
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

The creation of this new categorical attribute was necessary due to the following :

1. Here we had to predict the user that are likely to buy big-ticket items.
2. In the given data, average purchase price is given.This is a numerical variable.
3. Decision tree , requires the response variable to be categorical.
4. Hence created a categorical variable named BuyerType based on avg_price, which will be used as the response variable while building a model using decision tree.

Attribute Selection

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

Attribute	Rationale for Filtering
avg_price	As the response variable “buyerType” is derived from avg_price.Its variable has to be excluded from the selection.
user_id	userId is just an identifier for the user.It won't contribute anything the the model.
userSession_Id	userSession_id is also just an identifier.Hence filtering from the dataset.

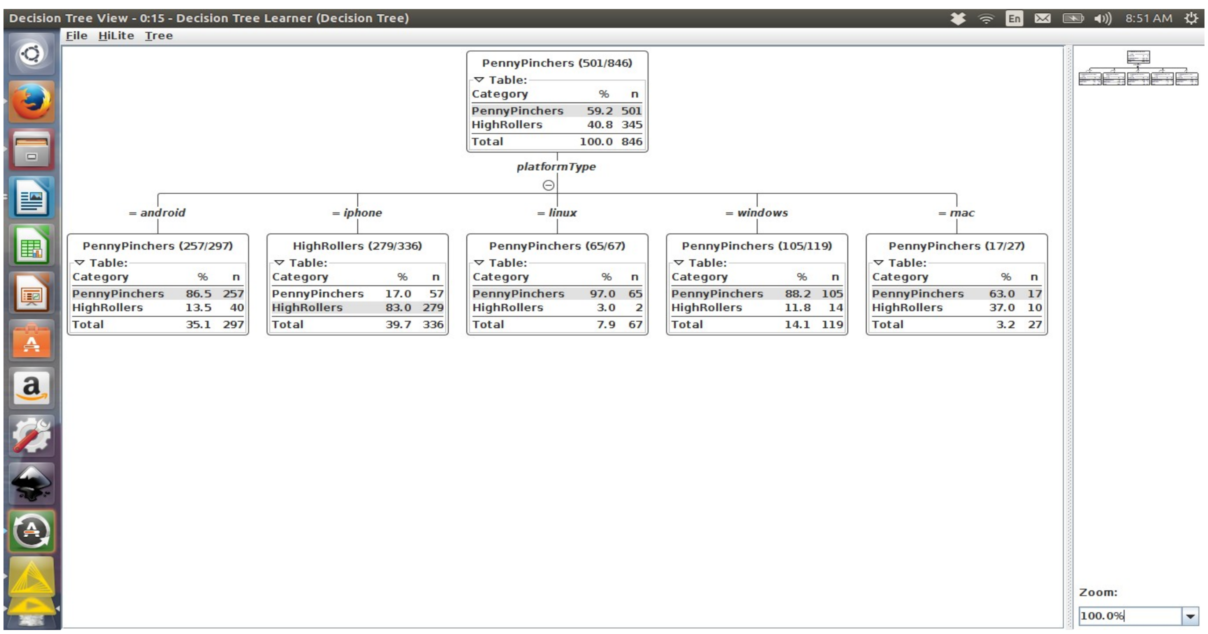
With the given dataset , following variables are being considered for modeling :

1. Response Variable : BuyerType
2. Exploratory Variables : teamLevel, platformType,count_gameclicks,count_hits,count_buyid.

Data Partitioning and Modeling

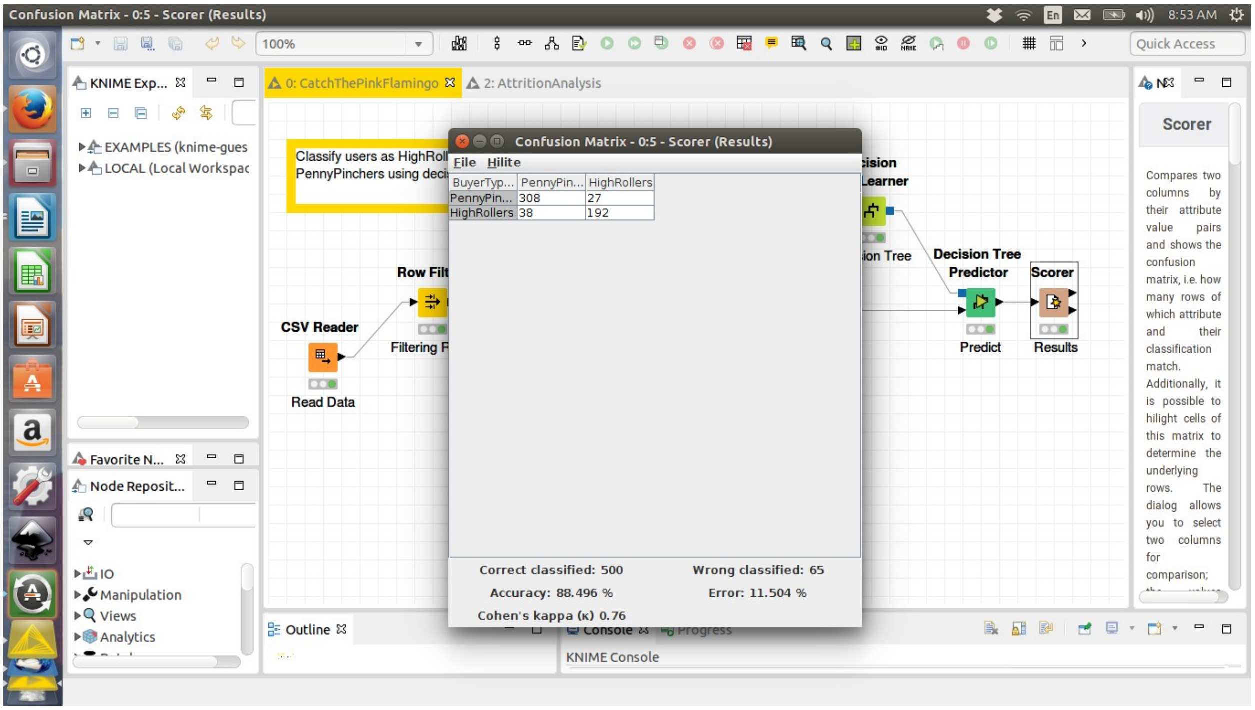
The data was partitioned into train and test datasets.
The **training** data set was used to create the decision tree model.
The trained model was then applied to the **testing** dataset.
This is important because **the model has to be tested on the data that was not used to train the model**.
When partitioning the data using sampling, it is important to set the random seed for the following reason :
To ensures that same partition are formed every time the partition node is executed.Hence get reproducible results.

Screenshot of the resulting decision tree is as shown below:



Evaluation

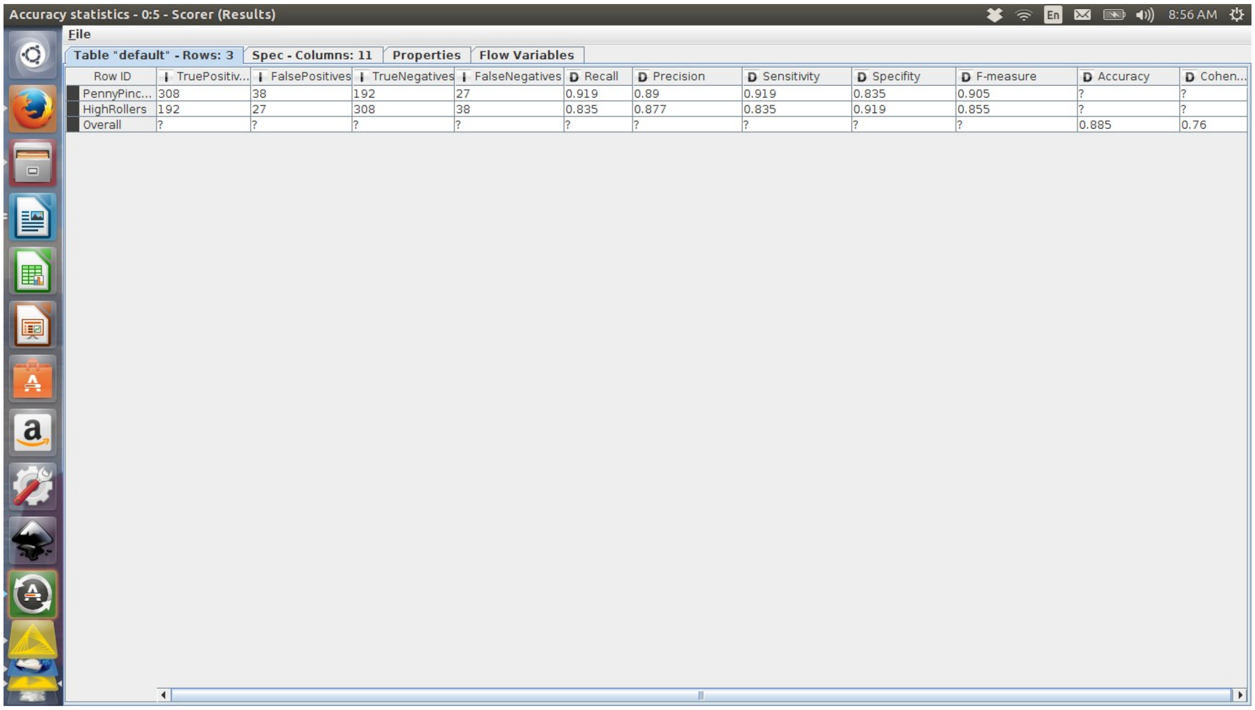
Screenshot of the confusion matrix can be seen below:



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

500 are correctly classified , where as 65 are wrongly classified.89% of times users were correctly classified as PennyPinchers. Where as 87% times they were correclty identified as HighRollers.

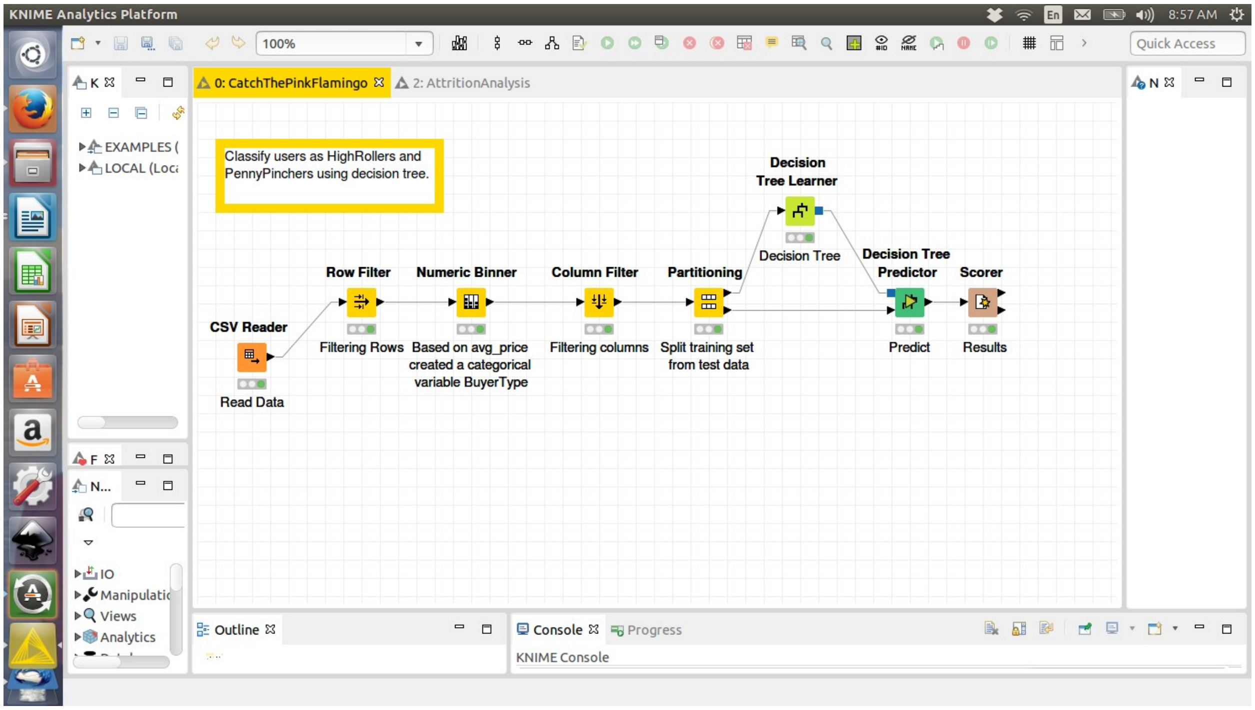
Other accuracy statistics are as below :



Row ID	TruePositiv...	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specifity	F-measure	Accuracy	Cohen...
PennyPinc...	308	38	192	27	0.919	0.89	0.919	0.835	0.905	?	?
HighRollers	192	27	308	38	0.835	0.877	0.835	0.919	0.855	?	?
Overall	?	?	?	?	?	?	?	?	?	0.885	0.76

Analysis Conclusions

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

Iphone users are HighRoller where as all other platform (i.e. android, linux, windows and mac) users are PennyPincher. Iphone forms 39.7% of user population, Android users are around 35.1 % , windows contribute 14.1% of total users,users using linux or mac are 7.9 and 3.2% respectively.

Specific Recommendations to Increase Revenue
1. Promote in-app purchases on android platform as it makes 35.1% of users.This will lead to increase in revenue proportionally. It is also recommended to investigate the reason for android users being PennyPincher
2. Promote in-app purchases to users who are classified as PennyPincher and are on iphone platform.As our findings suggest that iphone users are majorly HighRoller , there is a higher probability of them puchasing.Investigate further as to if there is any other reason for PennyPincher iphone users for not making in-app purchases.