Attribute Selection

Attribute	Rationale for Selection
Average ad-click count	Similar to the in-app purchases, we record each ad-click by the user and using this we can further predict the tendency or chances that the user will click on the ads. With time some users will click ads more oftern than the others Thus it helps us to compare these. We will calculate the average ad-click count by counting all the ad-clicks for a user divide by the number of sessions played.
Average buy-click count	Whenever the user makes an in-app purchase it's recorded as a buy-click. So, as the game progress users make more purchases whilst making their way through the different levels in the game. We can calculate the avg-buy-click per level. From here we can see the trend if the users are making more or less purchases or anything else. Calculation : Count all the buy-clicks for a user dividing it by the number of sessions played.
Average Expenditure	We can study the average expenditure per level as there is a range of in-app game items for users to purchase. With this we can study the highrollers vs pennypinchers. Calculation : Count all the item-prices paid by a user then divide it by the number of sessions played.

Training Data Set Creation

The training data set used for this analysis is shown below (first 5 lines):

avgPurchaseCount	avgAdClicks	avgExpPerSession
0	0	0
0.714285714	5.142857143	8.571428571
0	0	0
0	0	0
0	0	0
1.142857143	4.857142857	3.714285714
2	6.5	31.5
1	5	6.5
1.5	5.666666667	3.333333333
0	0	0

Dimensions of the training data set (rows x columns) : 1091 rows x 3 columns

# of clusters created: 3

Cluster Centers

Cluster #	Cluster Center
1 FreeLoaders	Average buy-clicks :: 0.02 Average Expenditure :: 0.03 Average ad-clicks :: 0.23 *Cluster size :: 589, 54% of dataset*
2 PennyPicnhers	Average buy-clicks :: 0.94 Average Expenditure :: 4.30 Average ad-clicks :: 5.83 *Cluster size :: 390, 35.7% of dataset*
3 HighRollers	Average buy-clicks :: 1.58 Average Expenditure :: 20.19 Average ad-clicks :: 5.71 *Cluster size :: 112, 10.3% of dataset*

These clusters can be differentiated from each other as follows:

Cluster 1(freeloaders) is different from the others in that…
It has captured most of the users that essentially play the game without spending at all and typically clicks on 1 advertisement in every 4-5 games. They provide the basic behaviour of the users in the game

Cluster 2(pennypinchers) is different from the others in that…
It is characterised as users who makes near to 1 purchase every level. Their expenditure is around $4.303 within a level. They also display a tendency to click on nearly 6 advertisements per level. It is much higher than the freeloaders.

Cluster 3(highrollers) is different from the others in that…
1.58 purchases per level(68% more than the pennypinchers) and spending almost $20 per level (470% more than the pennypinchers). Also ad-click rate is just slightly lower by 2% than penny-pinchers, they also click close to 6 advertisements per level which is much higher than the freeloaders

Log file from cluster analysis :

Highlighted text indicates details of the model selected for evaluation.

[SAMPLE DATASET]

[array([0., 0., 0.]), array([0.71428571, 8.57142857, 5.14285714]), array([0., 0., 0.]), array([0., 0., 0.]), array([0., 0., 0.]),
array([1.14285714, 3.71428571, 4.85714286]), array([2. , 31.5, 6.5]), array([1. , 6.5, 5.]),
array([1.5 , 3.33333333, 5.66666667]), array([0., 0., 0.])]

[TRAIN DATASET SHAPE]

(1091, 3)

[K]=1

[CENTER(S)]

[array([0.50749203, 3.62997822, 2.79677448])]

[COST]=59925.05543342645

[CLUSTER SIZES]

dict_items([(0, 1091)])

[K]=2

[CENTER(S)]

[array([1.37703252, 17.03899811, 5.90320122]), array([0.35365747, 1.25772443, 2.24720167])]

[COST]=23212.28217540738

[CLUSTER SIZES]

dict_items([(0, 164), (1, 927)])

[K]=3

[CENTER(S)]

[array([0.93948718, 4.30493547, 5.8309768]), array([1.57476616, 20.18807802, 5.71264881]),
array([0.01850594, 0.03449349, 0.23324844])]

[COST]=12671.3123894749

[CLUSTER SIZES]

dict_items([(0, 390), (1, 112), (2, 589)])

[K]=4

[CENTER(S)]

[array([1.16061224, 11.48923622, 5.95703231]), array([0.01357597, 0.02816291, 0.15800116]),
array([1.84887218, 25.41132331, 5.46215539]), array([0.87687397, 2.79847867, 5.72487983])]

[COST]=7645.50590694292

[CLUSTER SIZES]

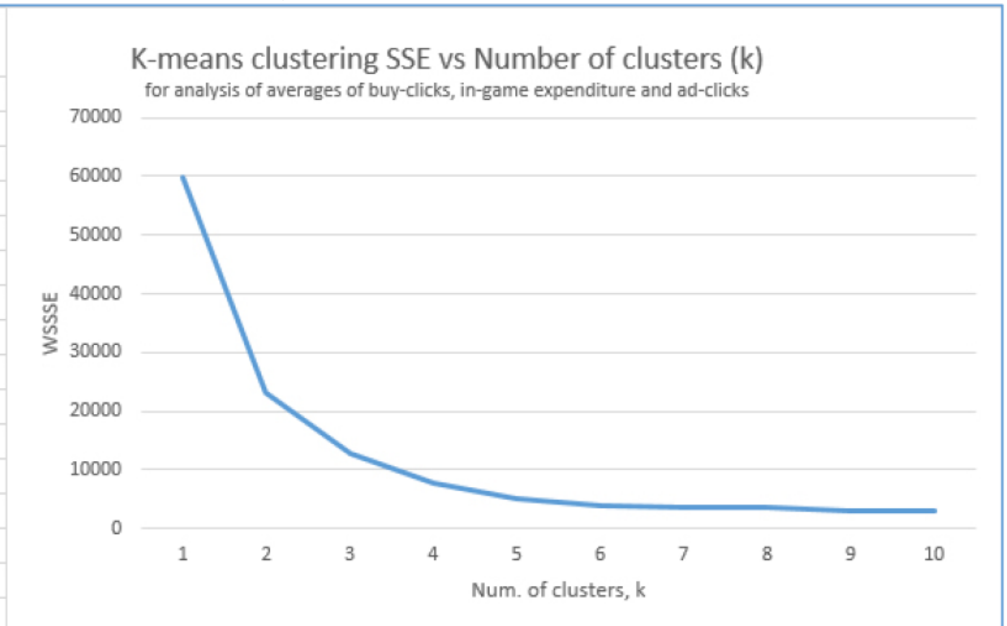
dict_items([(0, 140), (1, 577), (2, 57), (3, 317)])

[K]=5
[CENTER(S)]
[array([0.83032237, 2.41210121, 5.65188594]), array([1.14658163, 9.47171122, 6.06686224]),
array([0.01146922, 0.02439024, 0.14227642]), array([1.54826531, 19.3221017 , 5.38472789]),
array([2.28541667, 35.36238095, 6.1639881])]
[COST]=4983.973681127357
[CLUSTER SIZES]
dict_items([(0, 291), (1, 141), (2, 574), (3, 69), (4, 16)])

<Log truncated for brevity. k=6 onwards not included here>

Cost evaluation of cluster analysis

Num. of clusters, k	WSSSE
1	59925
2	23212
3	12671
4	7646
5	4984
6	3982
7	3610
8	3555
9	3166
10	3158



K=3 picked for evaluation of cluster analysis

Recommended Actions

Action Recommended	Rationale for the action
Provide multiple in-game purchases at time varying discounts	*Higher-rollers* make close to 1.5 purchases per level tend to spend much more than the *pennypinchers* (470%, i.e. almost 4 to 5 times more in the long run), thus a slight discount on bulk purchases can entice this group to continue purchasing the items. in the long term (perhaps maybe spend even more), or encourage the penny-pinchers to spend more. This would also lead to greater revenue generated from in-app purchases.
Assign high value ads to users who make at least 1 purchase per game.	The cluster analysis reveals that the users who make at least 1 ingame purchase per level have a tendency to click up close to 6 ads per level. This could optimise the revenue generated from 'high-rate' ads. The 'common-rate' ads can be assigned to the freeloaders.