

## Training Data Set Creation

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

```
In [153]: users.toPandas().head(5)
```

Out[153]:

	userId	sum(price)	sum(count_gameclicks)	count(adId)
0	231	63.0	262	19
1	2032	20.0	638	39
2	233	28.0	250	37
3	433	0.0	34	11
4	34	95.0	665	34

Dimensions of the training data set (rows x columns) : 529\*3

# of clusters created: 4