Training Data Set Creation

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

```
In [2]: cloudantdata = flightPredict.loadDataSet("flightstats_training_data_for_flight_delay_predictive_app_training_set",\
                                                  "training")
        cloudantdata.printSchema()
        cloudantdata.count()
        Successfully registered SQL table training
        root
        |-- _id: string (nullable = true)
        |-- _rev: string (nullable = true)
        |-- actualRunwayArrival: struct (nullable = true)
        |    |-- dateLocal: string (nullable = true)
        |    |-- dateUtc: string (nullable = true)
```

Dimensions of the training data set (rows x columns) : 5

# of clusters created: 3

Attribute Selection

Attribute	Rationale for Selection
User_ID	The user ID is the key for tying in game purchases and ad-click data back to the actual buyers.
Ad-clicks	The ad clicks detail the types of ads the users clicked on.
Buy-clicks	The buy clicks detail what the users bought.

Cluster Centers

Cluster #	Cluster Center
User_ID & Buy_Clicks	Clusters users with purchases
User_ID & Ad_Clicks	Clusters users with ads clicked on

These clusters can be differentiated from each other as follows:

Cluster 1 is different from the others in that it looks at what the users bought.

Cluster 2 is different from the others in that it looks at the ads that were clicked on.

Recommended Actions

Action Recommended	Rationale for the action
Increase advertising on computers and games.	Computers and games are the advertisements that are clicked on the most.

Stop spending money on automotive advertising.	Automotive ads are clicked on 20-25% less than the other ads clicked on at a higher rate.
Develop higher end computer products that the users can buy.	Users spend 4x more on computers than anything else. This suggests the users would be willing to spend more money on higher priced products.
Sports items have the lowest spend with the fourth highest advertisement click rate/	Tie a sports product to a computer product in order to increase the users spending on sports related items.