Big Basket:

Question 1:

Big basket is planning to start a recommendation engine based on the previous orders placed by the customer.

Question 2:

In Amazon there are several ways in which a recommendation is provided to the customer other than from previous orders. Few of the ways are below.

1. Recommendation from browser history of the user.
2. Based on the previous order
3. Personalised recommendation from the search history.
4. Last but not the least the best seller, promotional etc;

Question3 :

In Bigbasket people usually buy the household grocery items's so it would be suitable if the recommendation is given from the previous order of the user or by using the covariance matrix logic of Netflix recommendation engine. For example: if a person is buying protein then if any list of all the other customers buying protein and other item along with it should be recommendation engine.

Question4 :

IF a separate matrix is created for every user, where his list of orders is placed then all the system has to do id Runa recommendation engine, every time a user logs in and suggest from the freqold orders ,if a user is missing to any frequent purchased order.

Question5 :

Fist step is cleaning the data and convert the file to the binary file. It is done in the python notebook which I have attached.

Code:

**BB_member=subset(BB,BB$Member=='M04158')**

*In the above step it subset data based on the user in one dataframe. The below code is run in r , We have attached the script in the folder under the name "Association rules and graph.rmd"*

**rules = apriori(as.matrix(BB_member[,3:222]), parameter=list(support=0.05, confidence=0.5,minlen=1)) ## the first column in mydata**

*In the above step association rule is applied on the member dataset with supporrt=0.05 and confidence level of 0.5 and min Len of 1.*

*Below result is 8 rules are generated.*

Output:

```
    lhs                              rhs                  support    confidence lift     count
[1]  {Sugar}                       => {Root.Vegetables}   0.06060606 0.6666667  1.725490  8
[2]  {Glucose..Marie...Milk.Biscuits} => {Other.Vegetables}  0.05303030 0.5000000  1.375000  7
```

```
[3]  {Beans}                      => {Exotic.Vegetables} 0.05303030 0.5833333  2.851852  7
[4]  {Beans}                      => {Other.Vegetables}  0.05303030 0.5833333  1.604167  7
[5]  {Beans}                      => {Root.Vegetables}   0.06060606 0.6666667  1.725490  8
[6]  {Other.Dals}                 => {Root.Vegetables}   0.06818182 0.5294118  1.370242  9
[7]  {Brinjals}                   => {Other.Vegetables}  0.09848485 0.5000000  1.375000 13
[8]  {Exotic.Vegetables}          => {Root.Vegetables}   0.12121212 0.5925926  1.533769 16
[9]  {Other.Vegetables,Brinjals}  => {Root.Vegetables}   0.05303030 0.5384615  1.393665  7
[10] {Root.Vegetables,Brinjals}   => {Other.Vegetables}  0.05303030 0.5833333  1.604167  7 `
```

*Item frequently bought by the user is Sugar,root vegetables in a pair more frequently.*
*Marie and root vegetables together.*
*Biscuit anf vegetables together*
*Beans,other vegetabls together.*

*So when look at the output it looks like user buys Vegetable,Beans,Bringals and root vegetables.*
*So when the user is checking out we can suggest it at time of checking out.*

*Question 6 :*

*Taken a subset of two members from the main dataset.*

***BB_member=subset(BB,BB$Member %in% c('M04158' ,'M08075' ) )***
***BB_member***

*To generate the rules:*

***rules = apriori(as.matrix(BB_member[,3:222]), parameter=list(support=0.05, confidence=0.5,minlen=1)) ## the first column in mydata***

***The above query return 180 rules , which is mixtur of both users.***

***Few of the users are.***

```
[1]   {}                    => {Root.Vegetables}  0.50267380 0.5026738
1.000000 94
[2]   {Urad.Dal}            => {Root.Vegetables}  0.06417112 0.8571429
1.705167 12
[3]   {Cream.Biscuits}      => {Beans}            0.05882353 0.7333333
2.364368 11
[4]   {Cream.Biscuits}      => {Other.Vegetables} 0.07486631 0.9333333
1.917949 14
[5]   {Cream.Biscuits}      => {Root.Vegetables}  0.06417112 0.8000000
1.591489 12
[6]   {Sooji...Rava}        => {Gourd...Cucumber} 0.05347594 0.6250000
2.540761 10
[7]   {Sooji...Rava}        => {Beans}            0.06951872 0.8125000
2.619612 13
[8]   {Sooji...Rava}        => {Other.Vegetables} 0.06417112 0.7500000
1.541209 12
[9]   {Sooji...Rava}        => {Root.Vegetables}  0.05882353 0.6875000
1.367686 11
```

***[10] {Glucose..Marie...Milk.Biscuits} => {Beans} 0.07486631 0.5600000 1.805517 14***
***[11] {Glucose..Marie...Milk.Biscuits} => {Other.Vegetables} 0.08021390 0.6000000 1.232967 15***
***[12] {Glucose..Marie...Milk.Biscuits} => {Root.Vegetables} 0.08021390 0.6000000 1.193617 15***

*****************************************************************************************************
*7. Bigbasket is interested in introducing a "Smart Basket" feature that will identify a list of items a customer is more likely to buy. Discuss how this feature can be created?*

***A.*** *Smart Basket basically mean the items which customers are more likely to buy. These days more than discount people think about saving time, i.e why the online shopping is a boom, because people avoid going to super market and buy Daya to day usage item.Now people will prefer saving time while online shopping if the items which a user frequently buy is present in smart basket then the user doesnt have to search, he or she can just reorder the basket. Because mostly the household item which a user buys weekly is same.To get this information following steps should be followed.Need to identify the items which the customer has frequently bought in the past orders.We follow the below steps:*

*1. When the user logs in take the subset of all the previous order the user has done.*
*2. Create a network graph of the member and his orders.*
*3. Check he items with maximum degree the user will tend to buy these items more frequently.*

*****************************************************************************************************
***8. Pick a customer of your choice and create a "smart basket" for that customer. Please explain***

***A.*** *Steps to follow to create a smart basket for a customer.*
*he below code is run in r , We have attached the script in the folder under the name "Association rules and graph.rmd"*

*1. Create a subset for any of the customer.*
*2. Calculate indegree,outdegree and betweenness for the previous order after creating a graph*
*3. Run a K means on the basket,*
*4. After running the K means cluster 1 contains only member id*
*cluster 2 contains the frequently bought product*
*cluster3 contains the rest products.*

*So we could use the output of cluster 2 in frequently bought product.*

***1.*** *Creating the graph*

```
x<-read.csv("POS DATA-Table 1.csv")
x_member=subset(x,x$Member %in% c('M04158' ) )
AirlineNW <- graph.edgelist(as.matrix(x_member[,c(1,5)]), directed=TRUE)
```

*2. Calculate indegree,outdegree and betweenness.*

```
outdegree <- degree(AirlineNW,mode="out")
indegree <- degree(AirlineNW,mode="in")

btwn <- betweenness(AirlineNW,normalized = TRUE)
```

*3. Run K means cluster.*

*centralities <- cbind(indegree, outdegree)*
*colnames(centralities) <- c("inDegree","outDegree")*
*Cluster_Variability <- matrix(nrow=20, ncol=1)*
*for (i in 1:5) Cluster_Variability[i] <- kmeans(centralities,centers=i, nstart=3)$tot.withinss*
*plot(1:20, Cluster_Variability, type="b", xlab="Number of clusters", ylab="Within groups sum of squares")*

*4. Create the cluster:*
*X <- kmeans(centralities,centers=3)*

*kmean_clust<- cbind(centralities, X$cluster)*
*5. Subset of cluster 2:*

| | inDegree | outDegree | |
|---|---|---|---|
| Bread | 28 | 0 | 1 |
| Banana | 26 | 0 | 1 |
| Root Vegetables | 59 | 0 | 1 |
| Other Dals | 24 | 0 | 1 |
| Exotic Vegetables | 32 | 0 | 1 |
| Other Vegetables | 61 | 0 | 1 |
| Brinjals | 26 | 0 | 1 |

*In the above manner we can create a smart basket.*

************************************************************************************************************************

***9. What sort of data challenges do you anticipate while building a recommendation engine for Bigbasket?***

**A.** *One of the major challenges that we face is when a new customer logs in , we will have no data about the customer to suggest him anything in smart basket.*
*In the above scenario we could request the user to fill a form with basic information like address, bachelor or family, veg or non veg. Based on this new data we could use the Netflix recommendation engine (basically using the correlation matrix) logic and provide the user smart basket option*

************************************************************************************************************************

***10. What are the implementation and deployment challenges of a recommendation engine for Bigbasket?***

**A.** *Implemention and deployment challenges:*
*1. We can't store the data in the bulk format, each customer information should be in separate database. Because it would take a lot of time to create a subset at runtime.*
*2. It would be advisable to store the data in different graphs for each customer in the graph database to make the process quick.*
*3. Other major issues is Lack of data.*
*4. And the data which we have is dynamic data, it changes over time.*
*5. A separate logic to be implemented for a new product being launched in market ,if we have to promote it.*

************************************************************************************************************************