

# Cafe Bazaar Report

Mohammad Mazraeh

May 14, 2016

## Synopsis

In this research we want to analyse user payment data and cluster users in order to plan for promotions or solutions to make more money (Increase users long term value)!

## Data Processing

### Generating Data

In this part we would simulate some data as input for our analysis.

$n$  is number of data to generate.

*numUsers* is the number of users.

*typeFactor* is a factor including software types. in this analysis it's App or Game.

We assume the data table would have one ID and 3 features:

**type**: the payment was for an Application or a Game **amount**: the amount of payment **OrderDate**: payment date It is assumed that installations with no payment are recorderd as  $amount = 0$ . Also any payment (including in-app purchases) is recorderd in the table.

```
n <- 3000
numUsers <- 500
appCount <- 100
typeFactor <- as.factor(c('App','Game'))
```

Let's generate the data: for each record *CustID* is sampled from user IDs. *type* is set to App or Game with eual probability. for *amount* variable 0.9 of records will have zero value (Installations with no payement) and others are assumed to be from a normal distribution with arbitrary mean and std. at last *orderDate* is sampled randomly between "2013/01/01" and "2016/01/01".

```
orders <-
  data.table(CustID = sample(1:numUsers,size = n,replace = TRUE),
    type = sample(typeFactor,n,replace = TRUE, prob = c(0.5,0.5)),
    appID = sample(1:appCount,replace = TRUE),
    amount = round(sample(c(abs(rnorm(n = 0.1*n,mean = 3000,sd = 2000))),rep(0,0.9*n))),-2)
    OrderDate = sample(seq(
      as.Date('2013/01/01'),
      as.Date('2016/01/01'),
      by="day"), n,replace = TRUE))
```

## Convert Data to RFM Model

One of the most useful techniques to analyse user purchase history is to use RFM Models. R,F and M stand for Recency,Frequency and Monetary respectively. In this part we want to convert our input data into RFM Features:

***FirstPurchaseDate***: Roughly shows how long the user is with us.

***LastPurchaseDate***: basis of Recency.

***NumberofOrders***: basis of Frequency.

***TypeCount***: basis of breadth

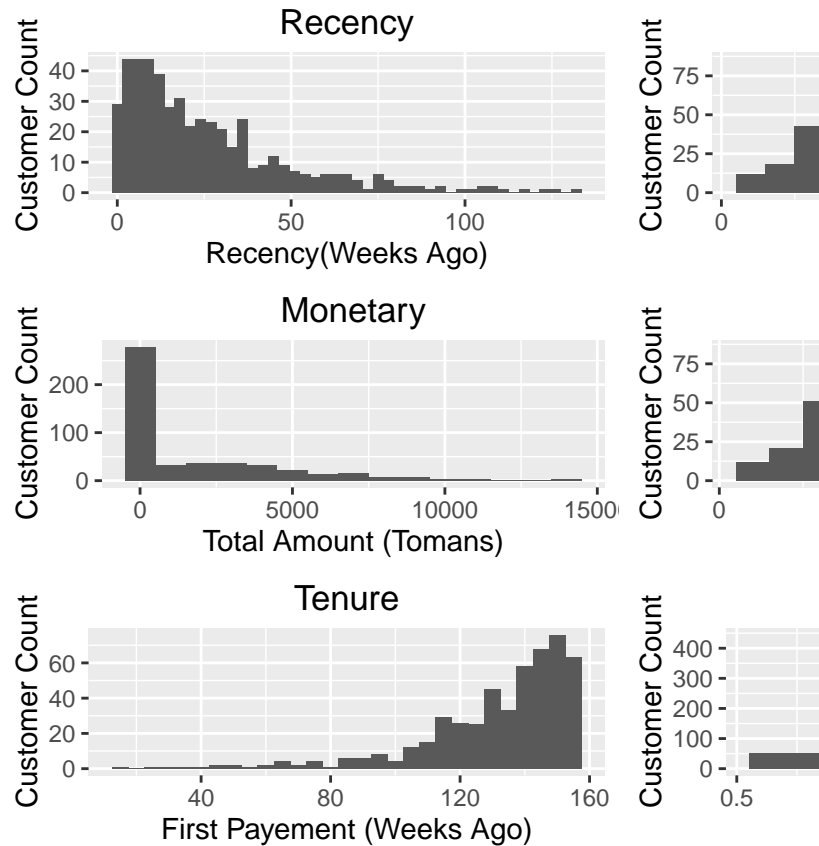
***TotalAmount***: basis of Monetary.

```
# Building RFM Features
RFM <- orders %>%
  group_by(CustID) %>%
  summarise(FirstPurchaseDate = min(OrderDate),
            LastPurchaseDate = max(OrderDate),
            NumberOrders = n(),
            TypeCount = length(unique(type)),
            Breadth = length(unique(appID)),
            TotalAmount = sum(amount))

LastDate <- max(RFM$LastPurchaseDate)
RFM <- RFM %>% mutate(R = as.numeric(LastDate-LastPurchaseDate))
```

# Analysis

## Exploratory Data Analysis



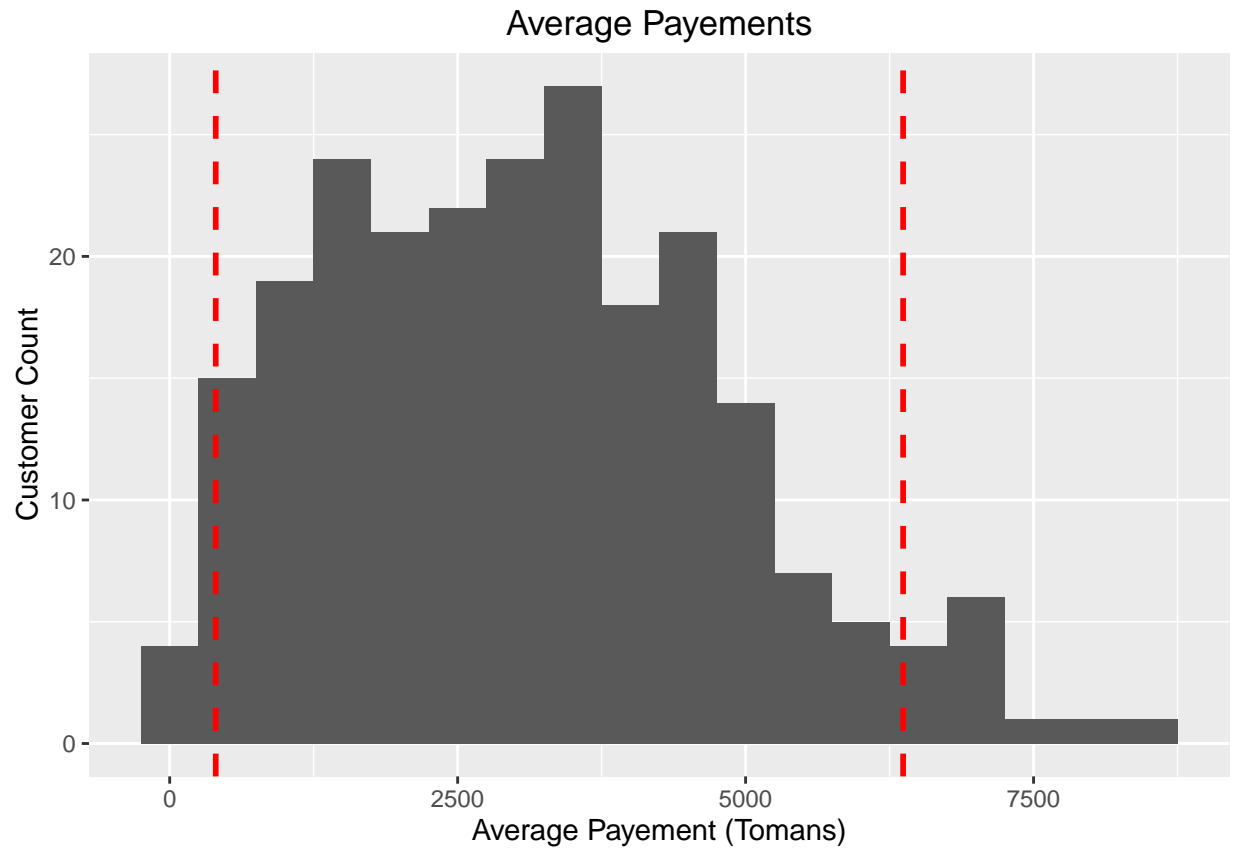
In this section we draw some EDA plots for RFM data.

By looking at recency plot we see that last payments in most recent weeks are more than some weeks ago, which seems to be normal. Frequency plot gives us a clue that most users install between 3-8 apps from cafe bazaar. In Monetary plot we see histogram for users total payments.

```
quantile(RFM$TotalAmount,seq(0.5,1,by = 0.1))
```

```
##    50%    60%    70%    80%    90%   100%
##      0   1260   2500   3980   5900  14400
```

Getting quantiles from total payments shows that 50% of customers won't pay for apps at all and 90% have paid less than 5900 tomans in their membership period. It's good to see how users pay for apps in average.



This shows that 90% of customers would pay an amount between 400 and 6367.5 Rials.

Customer Segmentation

