Coffee over Statistical Methods for Decision Making

~Shikhar Parashar

September 27, 2017

MENTOR: Dr. Bappaditya Mukhopadhyay

PGP - BA-BI Great Lakes Institute Of Management (Class of '17-'18)

Reading the dataset into R and checking the dimensions of the dataset

## [1] 130986 12

## [1] "Coffee dataset provides 130986 records with 12 columns."

Identifing Variables of the coffee dataset and updating behaviour / nature of variables as per the Desc.doc

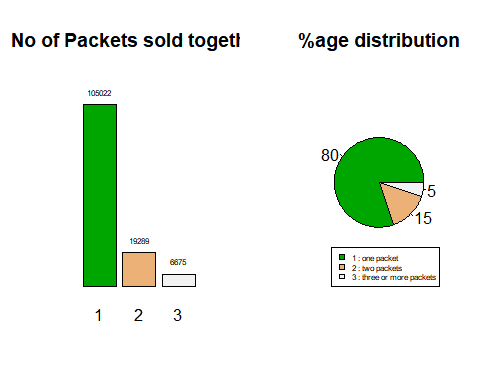
## 'data.frame': 130986 obs. of 12 variables:  
## $ ï..No\_of\_Packet : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Price\_per\_Packet : Factor w/ 3 levels "1","2","3": 1 1 1 1 2 1 1 1 1 2 ...  
## $ IDNo : Factor w/ 2111 levels "1","8","14","17",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Brand : Factor w/ 9 levels "1","2","3","4",..: 4 4 4 4 9 4 9 9 4 9 ...  
## $ Days\_between\_Purchase: int 11 24 22 24 11 10 12 13 24 4 ...  
## $ Age : Factor w/ 5 levels "1","2","3","4",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ SEC : Factor w/ 5 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ Income : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Household\_Sz : Factor w/ 5 levels "1","2","3","4",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Price\_Conscious : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Education : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...  
## $ Loyalty : Factor w/ 2 levels "1","2": 1 1 1 1 2 2 2 1 2 2 ...

### Exploratory Data Analysis

Now we further explore the nature of each individual variable in detail.

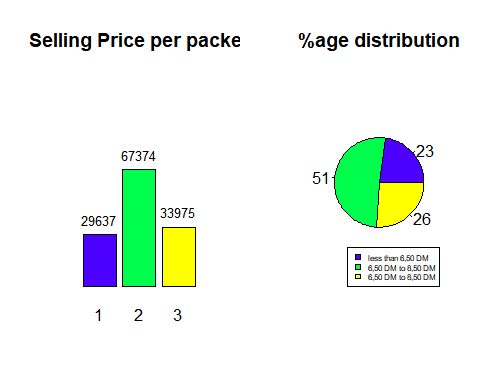
## ï..No\_of\_Packet Price\_per\_Packet IDNo Brand   
## 1:105022 1:29637 560 : 251 9 :35548   
## 2: 19289 2:67374 4627 : 245 8 :20055   
## 3: 6675 3:33975 1268 : 243 1 :15662   
## 985 : 242 3 :13832   
## 3669 : 239 6 :13274   
## 2898 : 228 4 : 9370   
## (Other):129538 (Other):23245   
## Days\_between\_Purchase Age SEC Income Household\_Sz  
## Min. : 1.00 1: 2675 1:24203 1:33480 1:21587   
## 1st Qu.: 6.00 2:16944 2:22302 2:34543 2:48624   
## Median : 9.00 3:26749 3:49052 3:40557 3:27910   
## Mean : 15.14 4:46601 4:25830 4:22406 4:20200   
## 3rd Qu.: 17.00 5:38017 5: 9599 5:12665   
## Max. :741.00   
##   
## Price\_Conscious Education Loyalty   
## 1:75051 1:37108 1:75057   
## 2:25694 2:38641 2:55929   
## 3: 5068 3:55237   
## 4:25173   
##   
##   
##

### No\_of\_Packet



Conclusion: As one can see, the Sales of Individual Packets outclass the sales of 2 and 3+ packs at a time by 80%.

### Price\_per\_Packet



Conclusion: We can see that 51% of all coffee packets are priced between 6,50 DM to 8,50 DM and hence is the dominating price category amongst the 3

### IDNo

Top 6 Houses with least number of Packets Purchased

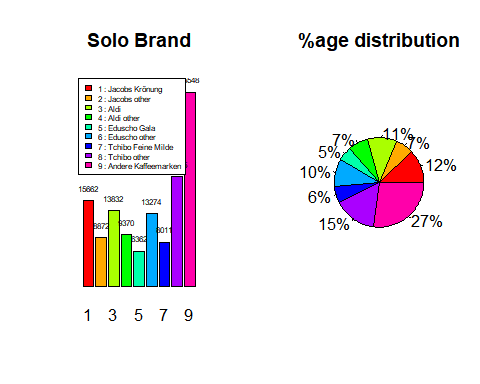
## 663 1665 1680 2005 4081 5773   
## 1 1 1 1 1 1

Top 6 Houses with most number of Packets Purchased

## 2898 3669 985 1268 4627 560   
## 228 239 242 243 245 251

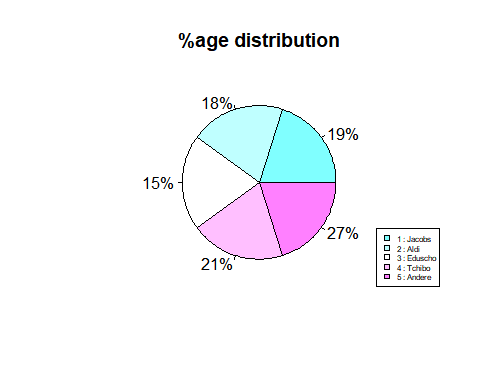
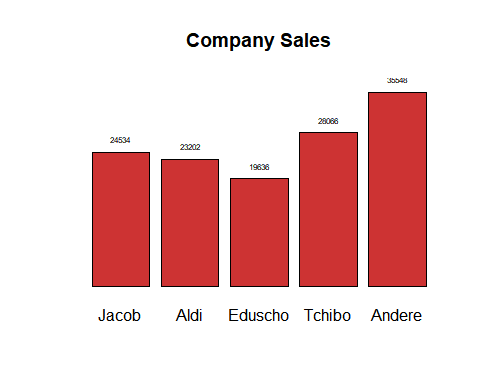
Conclusion: We can see that House ID: 560 is the largest purchaser of Coffee Packets whereas House ID: 663, 1665, 1680, 2005, 4081 and 5773 have purchased only 1 packet of Coffee

### Brand



Conclusion: When restricting ourselves to all individual Brands from different competetiors, we notice that the Brand : Andere Kaffeemarken has about 27% market share

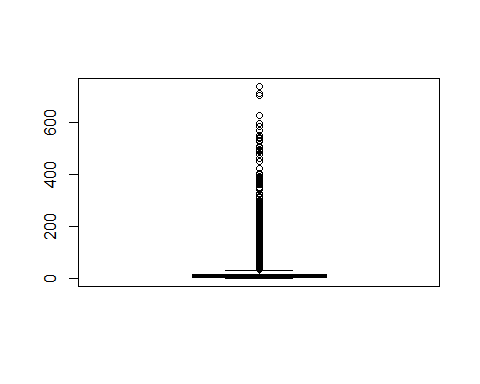
### Company



Conclusion: When looking for the Market Share for each company, it is evident that Andere leads the race with 27% market share, followed by Tchibo, Jacobs, Aldi and Eduscho respectively.

### Days\_between\_Purchase

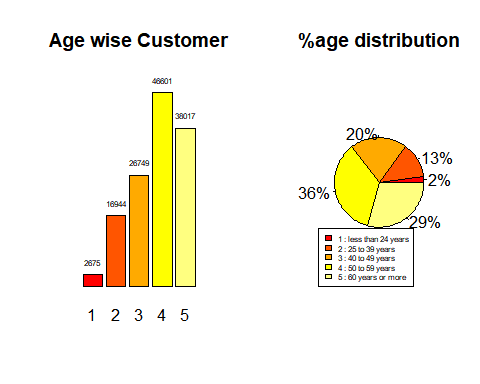
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 6.00 9.00 15.14 17.00 741.00



## [1] "53.49%" "46.51%"

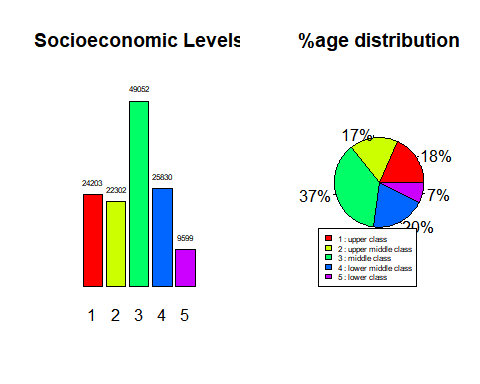
Conclusion: As evident from the summary and the resulting box plot: The number of days after which most households make their next purchase of coffee packets is 9 daysThe average number of days to make the purchase of next coffee packet is 15.14 days.46% Households buy another coffee packet within 6 days to 17 days of their previous packets purchase

### Age wise Customer Distribution



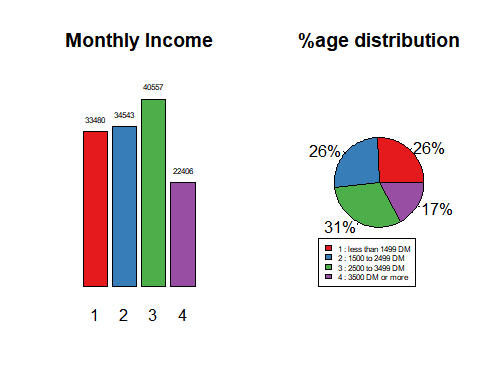
Conclusion: From the above, its crystal clear that, of all the customers researched, people within the Age Bracket of 50 to 59 years have the highest caffinne intakes whereas people within the age group of less than 24 years have the least caffinne intake.

### Socioeconomic level of householder



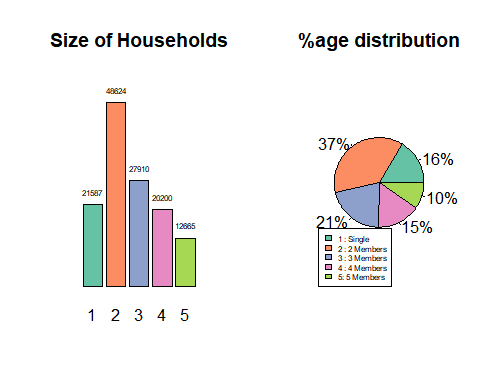
Conclusion: Middle Class is the Socio Economic class most likely to buy coffee whereas the Lower Class is the least likely of all the Socio Economic Classes to buy a packet of Coffee.

### Monthly Household Income



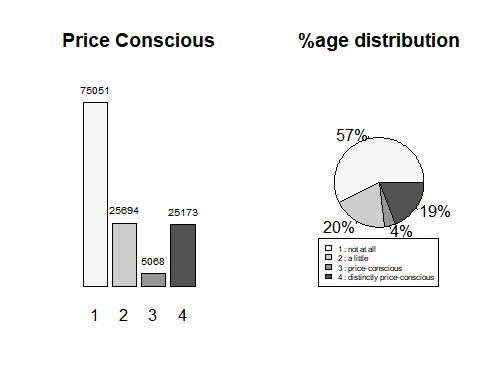
Conclusion: Household's earning 3500 DM or more happen to be smallest of consumers of Coffee whereas Households earning between 2500 DM to 3499 DM are the largest consumers of coffee.

### Household Size



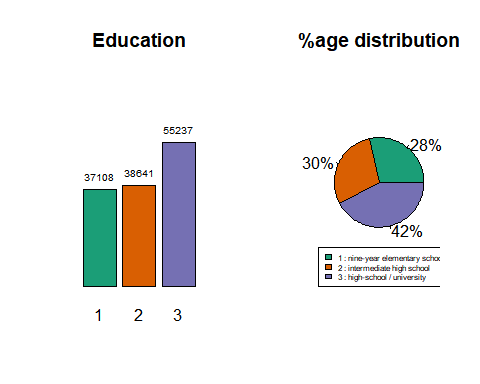
Conclusion: It is evident from the above that Small Families have a larger share of the Coffee sales. Households with 3 members or less account for 75% share of all coffee packets sold.

### Price Conciousness



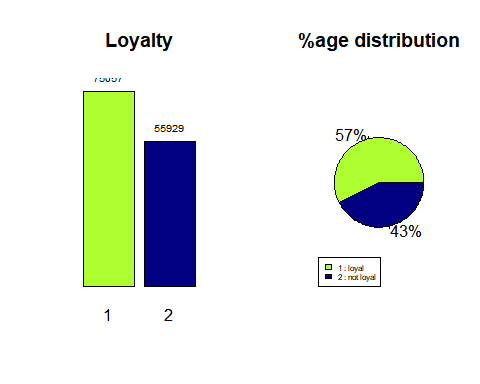
Conclusion: With more than 75% of all consumers either being 'not at all price concisous' or 'a little price concisous', the segment offers to be a Marketer's paradise for setting the Product Pricing.

### Education of householder



Conclusion: Based on Education we can infer that most coffee is consumed by people with high-school/university education.

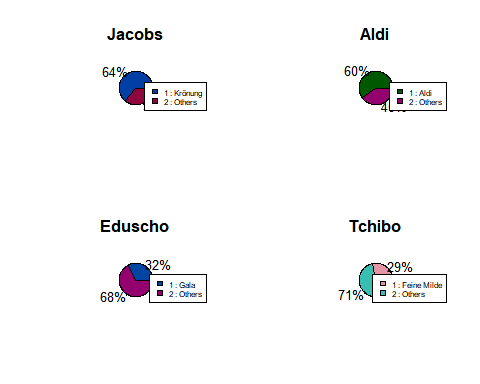
### Brand loyalty



Conclusion: Its evident from the above charts that Loyalty is not a big factor when making decisions on coffee. 43% non-loyal customers against 57% loyal customers is too close a call based on Loyalty.

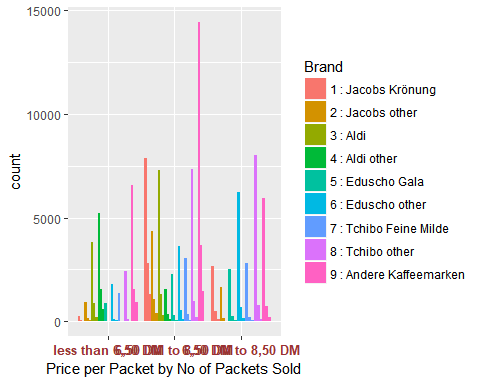
## Solutions to questions asked.

1. Which brands of coffee are more popular? Given a brand, are all variants equally preferred?



1. What are the prices of different brands of coffee?

## Warning: Ignoring unknown aesthetics: by



1. How frequently does a household buy coffee? How many packets of coffee are bought at a time?
2. What are the factors that have an impact on a household's coffee purchase pattern? Does brand preference depend on household size? Does purchase depend on a person's income or education level?