**Mini project BI**

**Realized by:**

**Ben Afia Edriss (TD1 TP1)**

**Ben Nsib Nassim (TD1 TP1)**

**E-mails:**

[**bennsib.nassim@gmail.com**](mailto:bennsib.nassim@gmail.com)

[**edriss.benafia1@gmail.com**](mailto:edriss.benafia1@gmail.com)

**Professor:**

**Gzara Mariem**

**Level:LA3INFO**

**TD: TD1**

**TP: TP1**

**Year: 2020-2021**

**Data Description**

**Dimension of data:**

There are 98913 observations (98913 users)

There are 24 Variables

**Names of variables, data types and data description**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Type** | **Description** |
| **identifierHash** | **Num** | Anonymous unique id. |
| **type** | **Chr** | The type of entity. This file contains only "user" entities. |
| **country** | **Chr** | User's country (written in french). See also the column "Country Code" if you prefer an ISO identifier. |
| **language** | **Chr** | The user's preferred language (the language of their interface when using the site) |
| **socialNbFollowers** | **Int** | Number of users who follow this user's activity. New accounts are automatically followed by the store's official accounts |
| **socialNbFollows** | **Int** | Number of user account this user follows. New accounts are automatically assigned to follow the official partners. |
| **socialProductsLiked** | **Int** | Number of products this user liked. |
| **productsListed** | **Int** | Number of currently unsold products that this user has uploaded. |
| **productsSold** | **Int** | Number of products this user has sold. |
| **productsPassRate** | **Num** | % of products meeting the product description. (Sold products are reviewed by the store's team before being shipped to the buyer.) |
| **productsBought** | **Int** | Number of products this user bought |
| **gender** | **Char** | user's gender |
| **civilityGenderId** | **Int** | civility as integer |
| **civilityTitle** | **Chr** | Civility title |
| **hasAnyApp** | **Chr** | user has ever used any of the store's official app |
| **hasAndroidApp** | **Chr** | user has ever used the official Android app |
| **hasIosApp** | **Chr** | user has ever used the official iOS app |
| **hasProfilePicture** | **Chr** | user has a custom profile picture |
| **daysSinceLastLogin** | **Int** | Number of days since the last login. All user data were fetchedthe same day. See also "seniority". |
| **seniority senio** | **Int** | Number of days since the user registered |
| **seniorityAsMonths** | **Num** | see seniority. Here, expressed in months |
| **seniorityAsYears** | **Num** | see seniority. Here, expressed in years |
| **countryCode** | **Chr** | user's country (ISO-3166-1) |

**Number of distinct values ​​for each variable**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Number of distinct values** | **Values description** |
| **type** | **1** | User : The type of entity. This file contains only "user" entities |
| **country** | **200** | There are 200 name of country |
| **language** | **5** | En : English  Fr : French  De : Deutsche  Es : Spanish  It : Italian |
| **gender** | **2** | F : Female  M : Male |
| **civilityTitle** | **3** | Mrs : Madam  Mr : Mister  Miss :Miss |
| **hasAnyApp** | **2** | True and false |
| **hasAndroidApp** | **2** | True and false |
| **hasIosApp** | **2** | True and false |
| **hasProfilePicture** | **Chr** | True and false |
| **countryCode** | **200** | There are 200 ISO codes of country |

**Missing values**

No missing data

**Useless varaibles**

There are some useless variables (duplicated variables, variables have no statistics values...)

Identifierhash and type have not statistics values.

civilityGenderId and gender : have the same meaning

country and countryCode : have the same meaning

**🡺We should to delete the useless columns.**

**New dimension of data:**

There are 98913 observations (98913 users)

There are 20 Variables

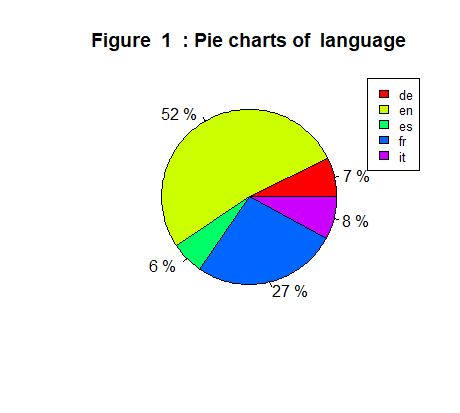
**Data Exploration**

**I) Data Exploration (Univariate exploration)**

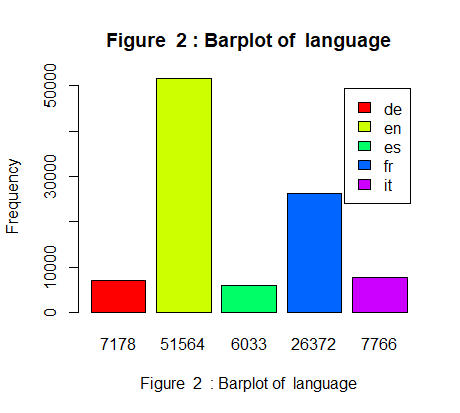
1. **The categorical variables**

**How do you explore categorical data? We use the pie charts and the barplot**

1. **Language**

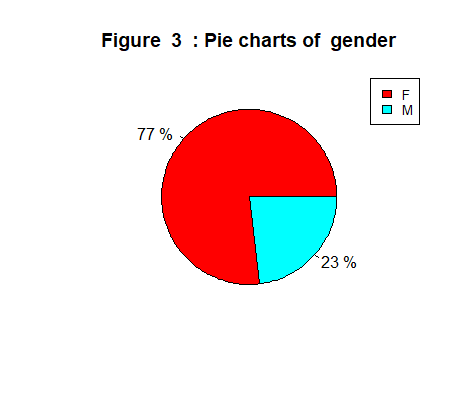


* **The majority of user preferred the English language 52 % of users preferred it and the second preferred language is French 27% of users preferred it**

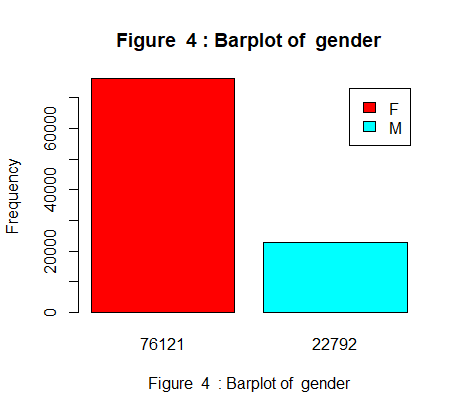


* **More than 50000 users preferred the English language**
* **About 30000 users preferred the English language**

1. **Gender**



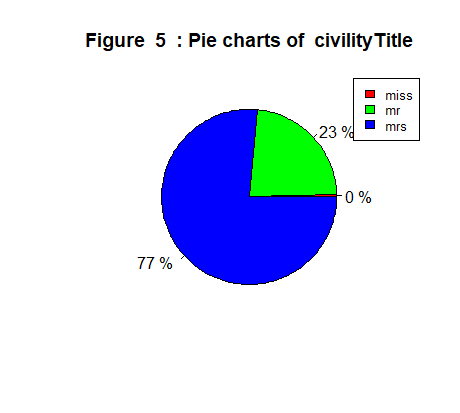
* **The majority of users are females (77%) and 23% are males**

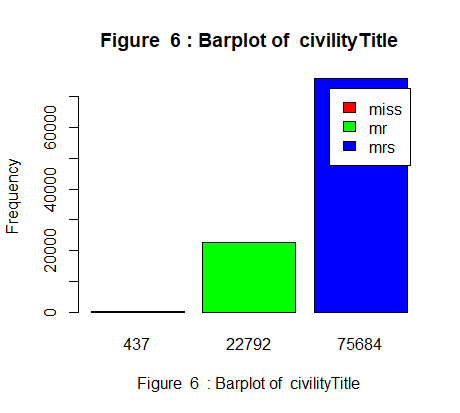


* **More than 76000 of users are females**
* **More than 22000 of users are males**

**3) civilityTitle**

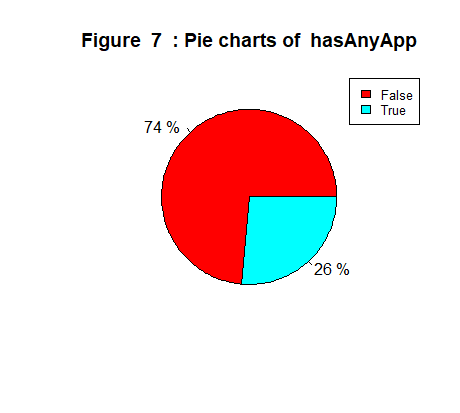
* **77% of users are Mrs.**
* **23% of users are Mrs.**
* **About 0% of users are Miss**



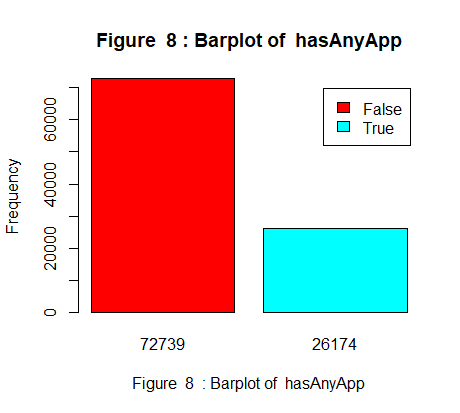


* **More than 75000 of users are Mrs.**
* **More than 22000 of users are Mrs.**
* **About 437 of users are Miss (the lower)**

**4) hasAnyApp**

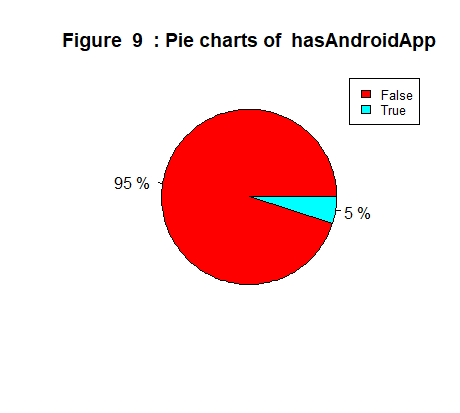


* **74% of users have not any of the store's official app**
* **26% of users have a store's official app**

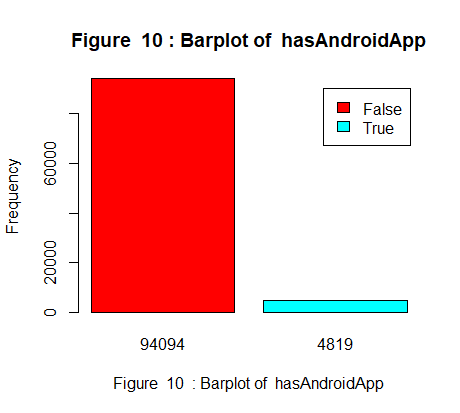


* **More than 72000 of users have not any of the store's official app**
* **More than 26000 of users have a store's official app**

**5) hasAndroidApp**

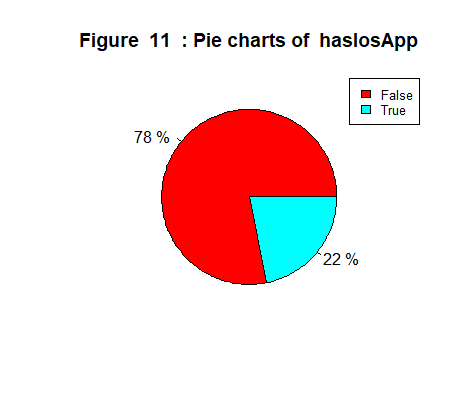


* **95% of users have not the android app**
* **26% of users have the android app**

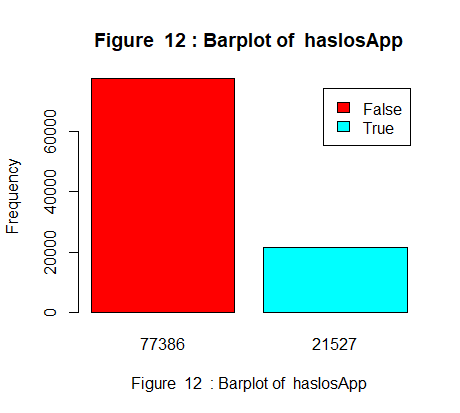


* **More than 94000 of users have not the android app**
* **More than 4800 of users have the android app**

**6) hasIosApp**

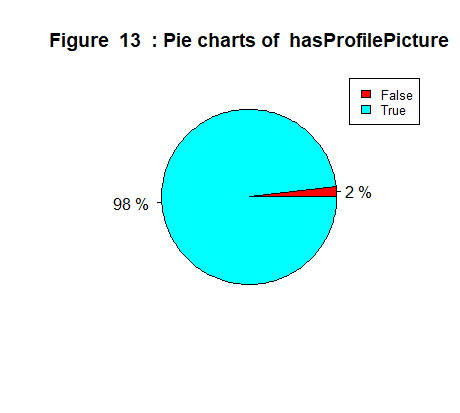


* **78% of users have not the iOS app**
* **26% of users have the iOS app**

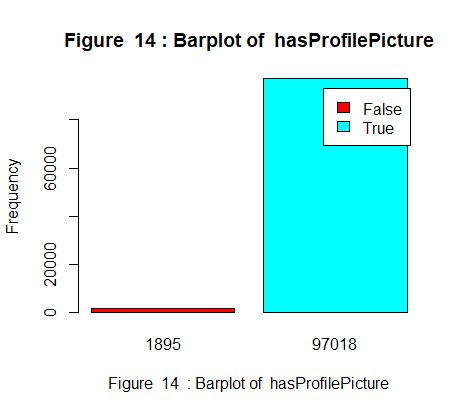


* **More than 77000 of users have not the iOS app**
* **More than 21000 of users have the iOS app**

**7) hasProfilePicture**

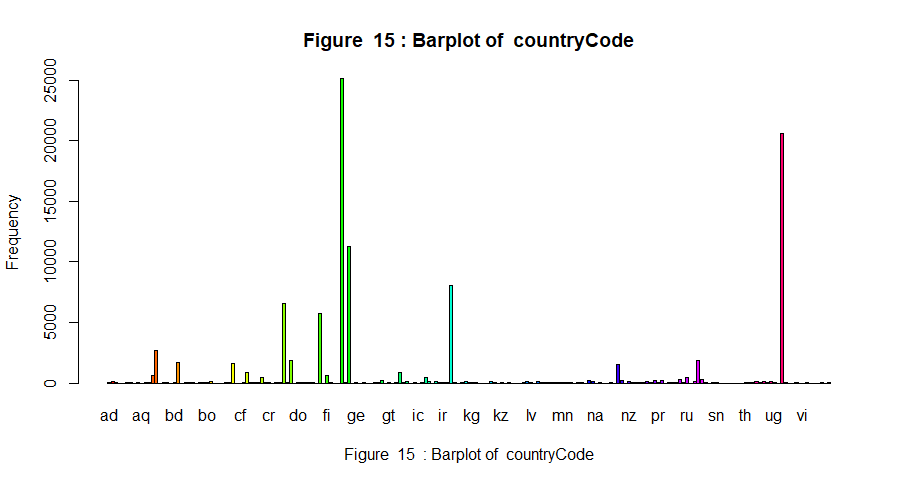


* **98% of users have not a profile picture**
* **2% of users have profile picture**



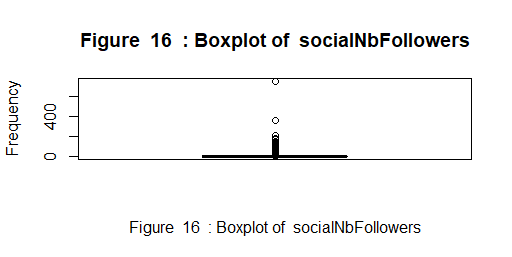
* **More than 97000 of users have profile picture**
* **1895 of users have profile picture**

**8) countryCode**

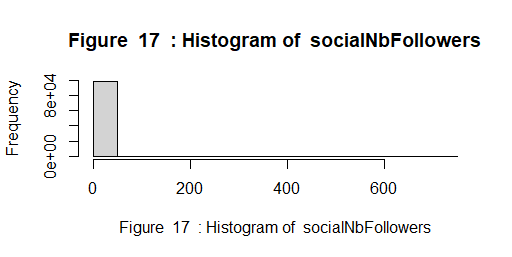


* **More than 25000 users live France**
* **More than 20000 of users live US**

1. **The numeric variables**
2. **socialnbFollowers**

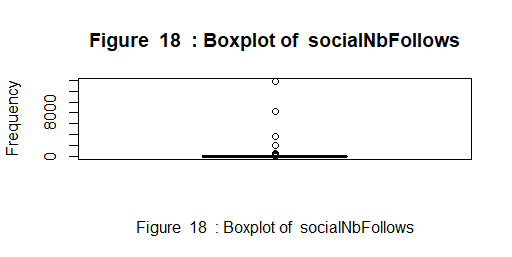


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

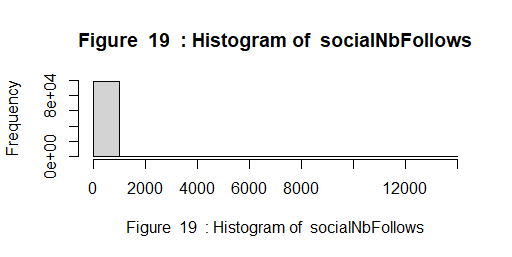


* Uni modal data distribution
* There is one peak of data

1. **socialnbFollows**

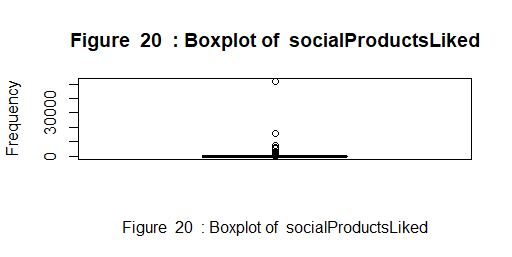


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

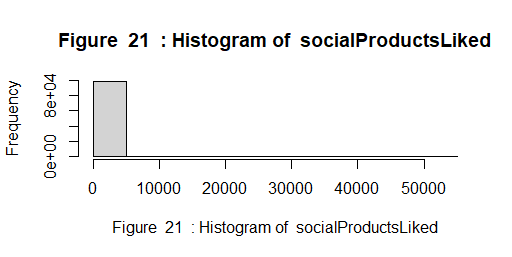


* Uni modal data distribution
* There is one peak of data

1. **socialProductsLiked**

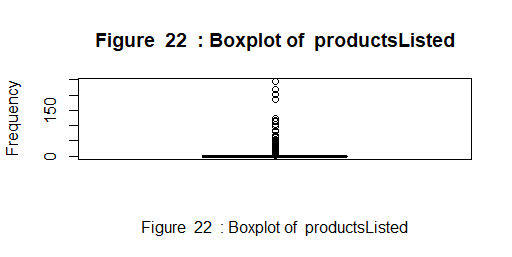


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**
* Uni modal data distribution
* There is one peak of data

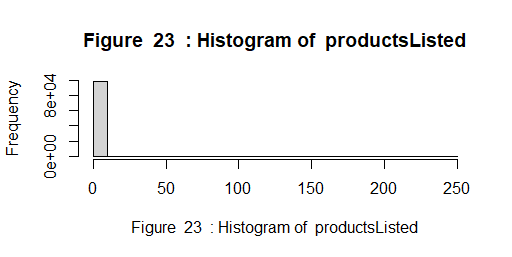


1. **ProductsListed**

* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

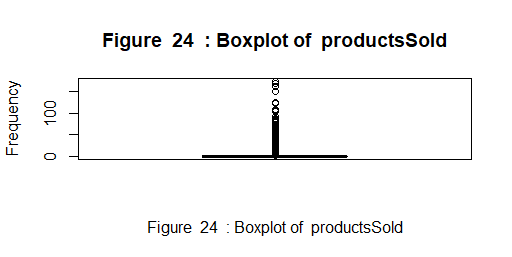


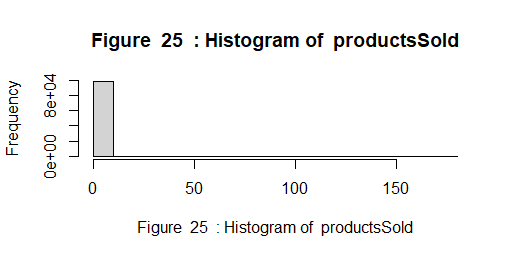
* Uni modal data distribution
* There is one peak of data



1. **ProductsSold**

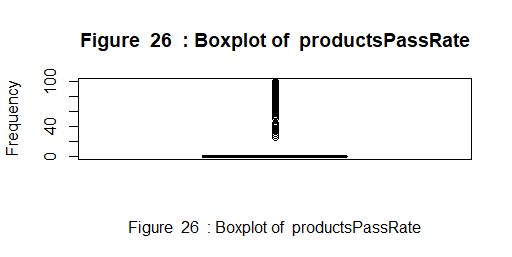
* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**



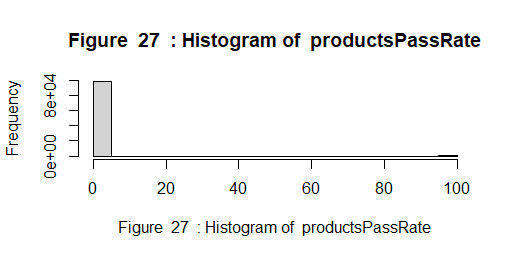


* Uni modal data distribution
* There is one peak of data

**6)ProductsPassRate**

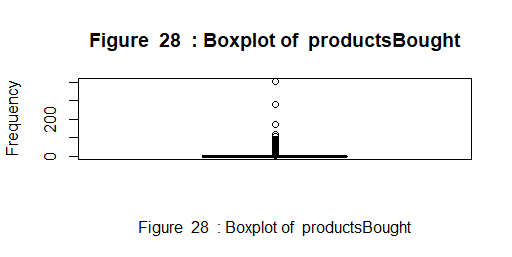


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

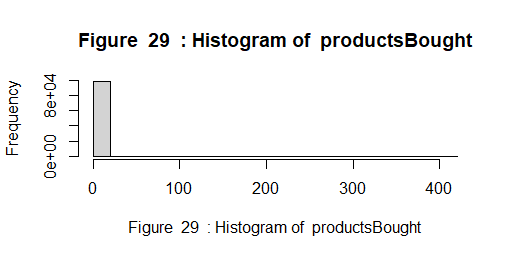


* Uni modal data distribution
* There is one peak of data

**7)ProductsBought**

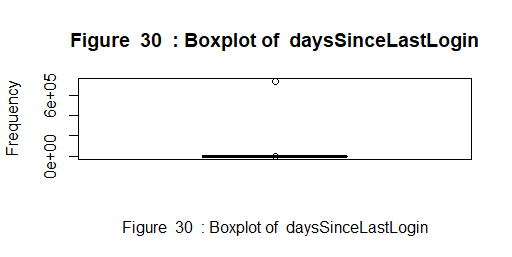


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

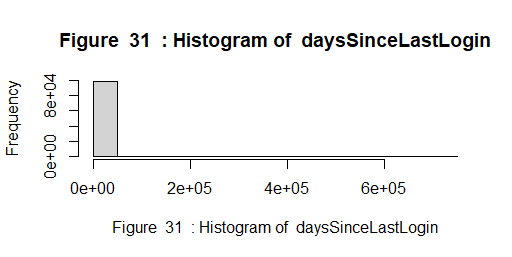


* Uni modal data distribution
* There is one peak of data

**8)daysSinceLastLogin**

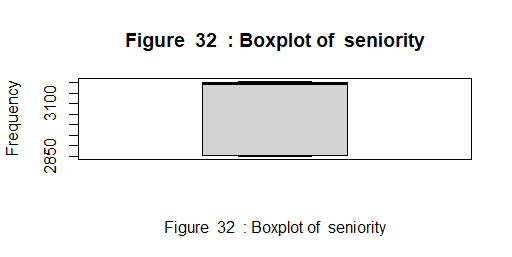


* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**

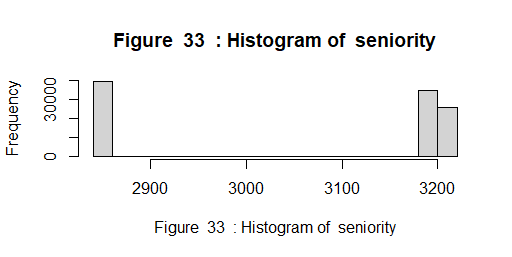


* Uni modal data distribution
* There is one peak of data

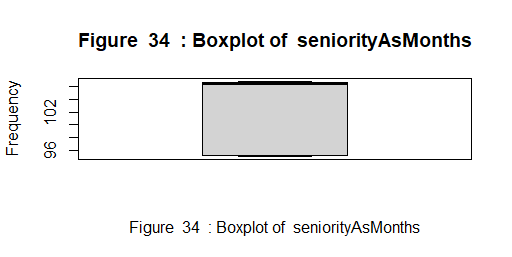
**9)seniority**

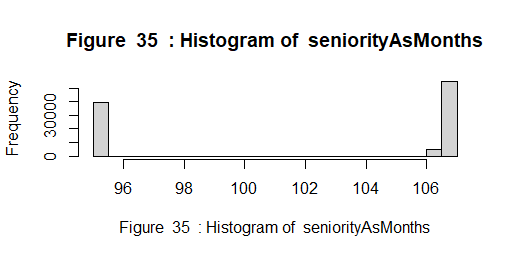


* **low density of data between Q1 and Mean**
* **Very higher of density in [min-Q1] and [median-max]**
* Bi modal data distribution
* There is two groups of data well separated



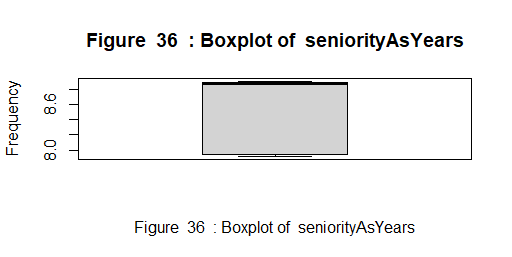
**10)seniorityAsMonths**



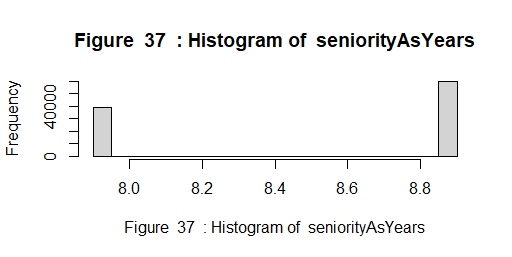


* Bi modal data distribution
* There is two groups od data well separated
* **low density of data between Q1 and Mean**
* **Very higher of density in [min-Q1] and [median-max]**

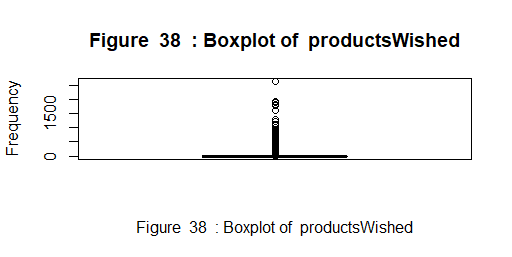
**11)seniorityAsYears**



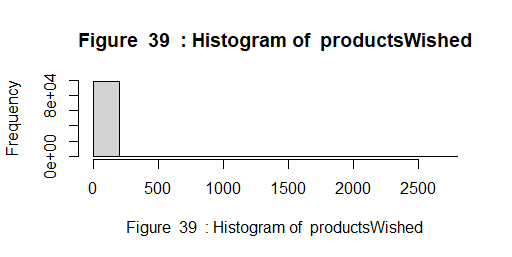
* **low density of data between Q1 and Mean**
* **Very higher of density in [min-Q1] and [median-max]**
* **Bi modal data distribution**
* **There is two groups of data well separated**
* **The peak is in the end of the histogram**



**12)productsWished**



* **The boxplot is so tight**
* **Presence of some outliers**
* **High density of data between Min and Max**



* **Uni modal data distribution**
* **There is one peak of data**

**II) Data Exploration (Bivariate exploration)**

**Correlation of numeric variables (with command cor(data))**

1. **socialNbFollowers**

* **Moderate positive correlation with socialProductsLiked, productsSold and socialNbFollows.**
* **Weak positive correlation with productsListed, productsPassRate and productsWished.**
* **There is independence with the other variables**

1. **socialNbFollows**

* **Strong positive correlation with socialProductsLiked.**
* **Moderate positive correlation with socialNbFollows.**
* **There is independence with the other variables**

1. **socialProductsLiked**

* **Strong positive correlation with socialNbFollows.**
* **Moderate positive correlation with socialNbFollowers.**
* **Weak positive correlation with productsWished .**
* **There is independence with the other variables.**

1. **productsListed**

* **Moderate positive correlation with productsSold**
* **Weak positive correlation with socialNbFollowers and productsPassRate .**
* **There is independence with the other variables.**

1. **productsSold**

* **Moderate positive correlation with socialNbFollowers and productsListed.**
* **Weak positive correlation with, productsPassRate.**
* **There is independence with the other variables.**

1. **productsPassRate**

* **Weak positive correlation with socialNbFollowers and productsSold.**
* **There is independence with the other variables.**

1. **productsWished**

* **Weak positive correlation with socialNbFollowers, socialProductsLiked and productsBought.**
* **There is independence with the other variables.**

1. **productsBought**

* **Weak positive correlation with productsWished.**
* **There is independence with the other variables.**

1. **daysSinceLastLogin**

* **There is independence with all variables.**

1. **seniority**

* **Very Strong positive correlation with** **seniorityAsMonths and seniorityAsYears.**
* **There is independence with the other variables.**

1. **seniorityAsMonths**

* **Very Strong positive correlation with** **seniority and seniorityAsYears.**

**There is independence with the other variables**

1. **seniorityAsYears**

* **Very Strong positive correlation with** **seniority and seniorityAsYears.**
* **There is independence with the other variables**

**Conclusion:**

**🡺We can now considerate seniority, seniorityAsMonths and seniorityAsYears as one column because they have a very strongly positively correlated and they having the same meaning.**

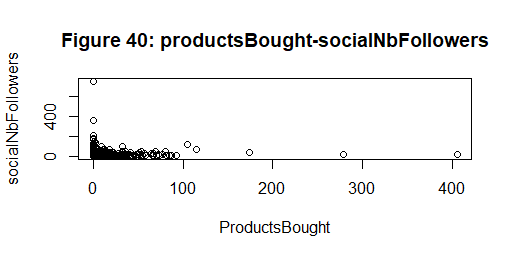
🡺**Now we have 18 variables**

Puisque on a pour but de déterminer les personnes qui vont rester dans le store et les personnes qui vont quitter le store on a consentré sur deux variables qui peut nous aider à atteindre notre but à savoir : productsBought & products Sold .

On a étudier les correlations entre productsBought et les autres variables numeriques de meme pour la variable productsSold

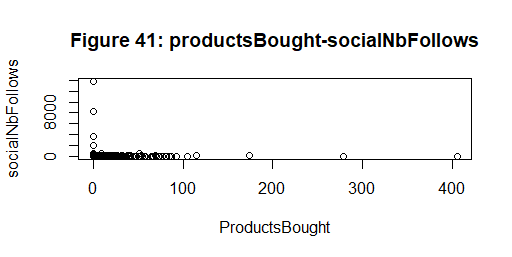
**Scatter plots productsBought:**

1. **productsBought – socialNbFollowers**



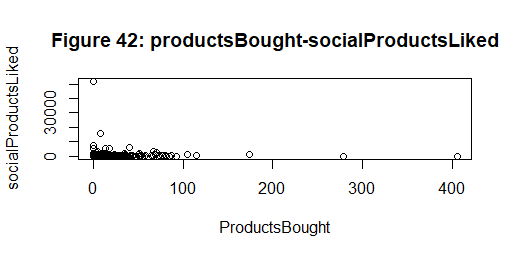
* **Presence of one group of data between 0 and 100**
* **High density of data**
* **Presence of some outliers**
* **Independent between the 2 variables**

**2)productsBought – socialNbFollows**



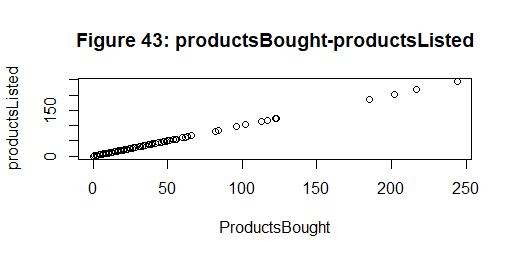
* **Presence of one group of data between 0 and 100**
* **High density of data**
* **Presence of some outliers**
* **Independent between the 2 variables**

**3)productsBought – socialProductsLiked**



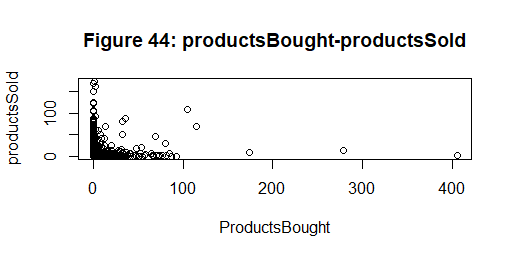
* **Presence of one group of data between 0 and 100**
* **High density of data**
* **Presence of some outliers**
* **Independent between the 2 variables**

**4)productsBought – productsListed**



* Perfect corroelation between the 2 variables
* Discontinious distribution of data
* Presence of some outiers

**5)productsBought – productsSold**



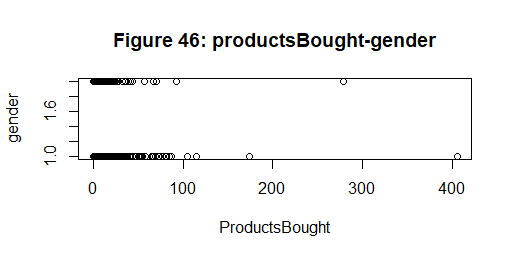
* **Presence of one group of data between 0 and 100**
* **High density of data**
* **Presence of some outliers**
* **Independent between the 2 variables**

**6)productsBought – productsPassRate**



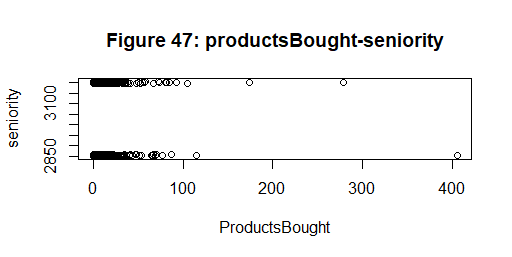
* **Presence of 2 groups of data between 0 and 100 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**7)productsBought – gender**



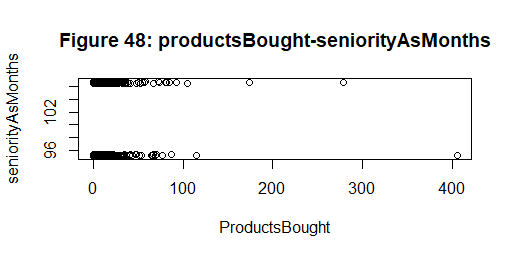
* **Presence of 2 groups of data between 0 and 100 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**8)productsBought – seniority**



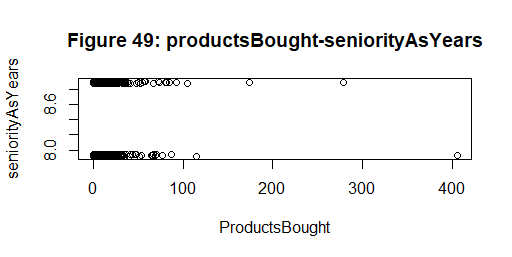
* **Presence of 2 groups of data between 0 and 100 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**9)productsBought – seniorityAsMonths**



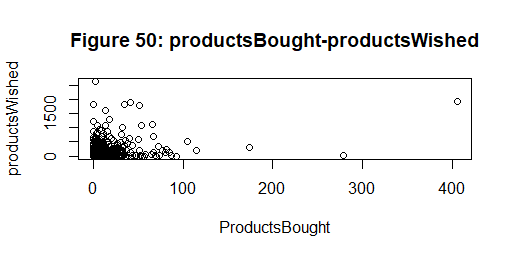
* **Presence of 2 groups of data between 0 and 100 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**10)productsBought - seniorityAsYears**



* **Presence of 2 groups of data between 0 and 100 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**11)productsBought - productsWished**



* **Presence of one group of data between 0 and 100**
* **High density of data**
* **Presence of some outliers**
* **Independent between the 2 variables**

**Scatter plots productsSold**

**1)productsSold– socialNbFollowers**



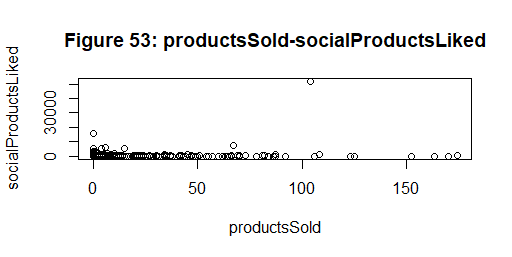
* **Weak correlation between the 2 variables**
* **Presence of some outliers**
* **High density between 0 and 50**

**2)productsSold – socialNbFollows**



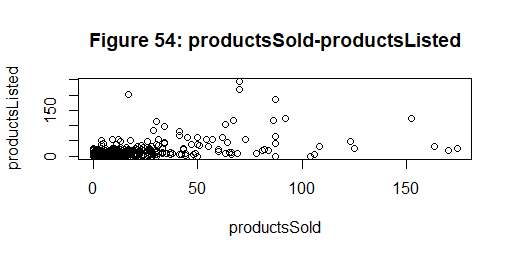
* **Independent correlation between the 2 variables**
* **Presence of some outliers**

**3)productsSold – socialProductsLiked**



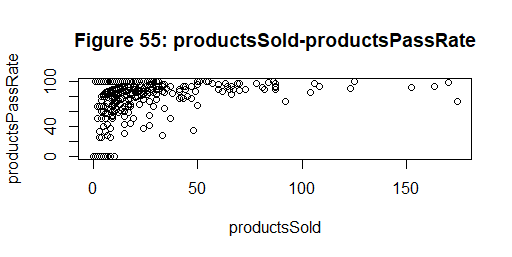
* **Independent correlation between the 2 variables**
* **Presence of some outliers**

**4)productsSold – productsListed**



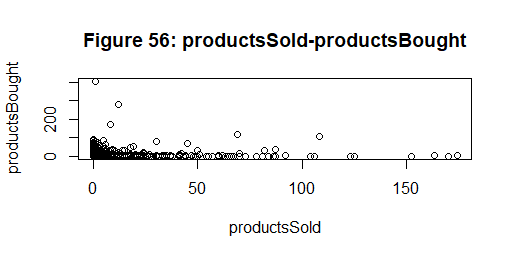
* **Moderate correlation between the 2 variables**
* **High density of data between 0 and 50**
* **Presence of some outliers**

**5)productsSold – productsPassRate**



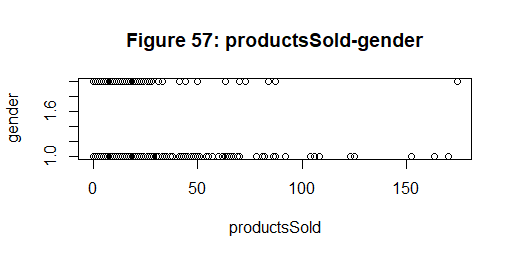
* **Presence of two groups of data**
* **Presence of some outliers**
* **Weak correlation between the two variables**

**6)productsSold – productsBought**



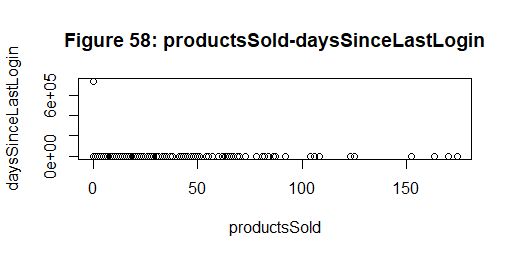
* **Independent**
* **High density of data**
* **Presence of some outliers**

**7)productsSold – gender**



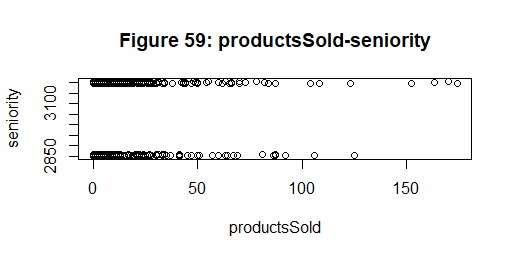
* **Presence of 2 groups of data between 0 and 70 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**8)products Sold – daysSinceLastLogin**



* **High density of data between 0 and 70**
* **Presence of some outliers**

**9)productsSold – seniority**



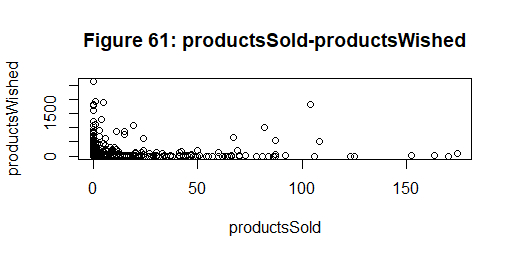
* **Presence of 2 groups of data between 0 and 50 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**10)productsSold - seniorityAsMonths**



* **Presence of 2 groups of data between 0 and 70 well separated**
* **High density of data in the 2 groups**
* **Presence of some outliers**

**11)productsSold – productsWished**



* **Independent**
* **High density of data**
* **Presence of some outliers**

**Data clustering and cluster validation**

**We have now 18 variables: 2 variables output (productsBought,** **productsSold) and 16 decisions variables.**

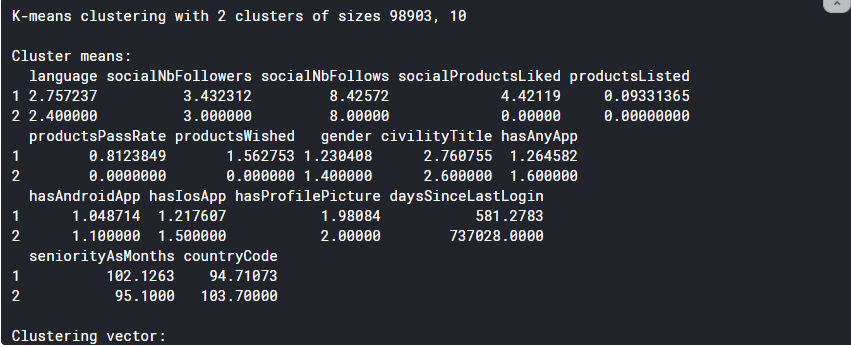
**So:**

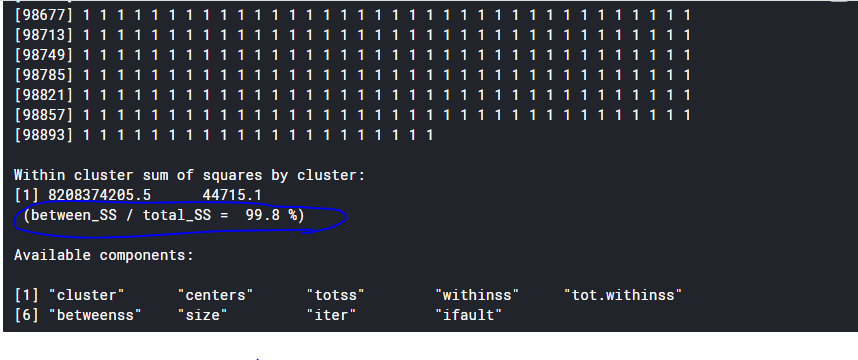
* **First: we will replace productsBought and** **productsSold with other column (**leaveStore) **:**

**All users bought or sold a product they take no and others they will take no**

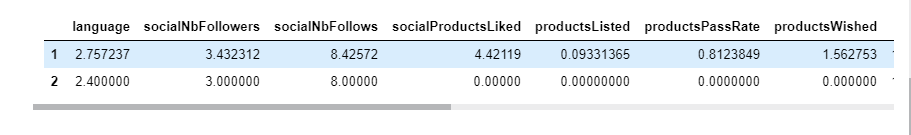
* **Second: we will convert all categorical variables to numeric variables because we will be using the k-mean in clustering and this algorithm will only work with numeric variables.**

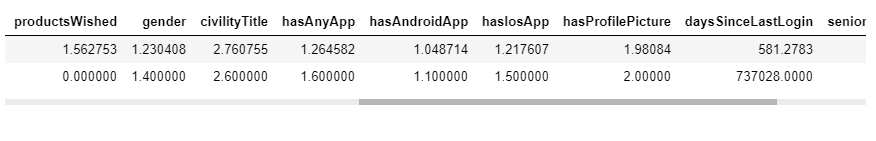
**1-Exploration the result of k-means**

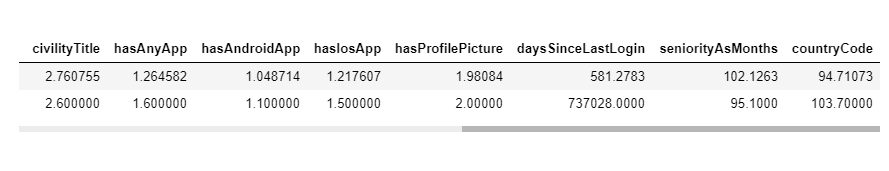




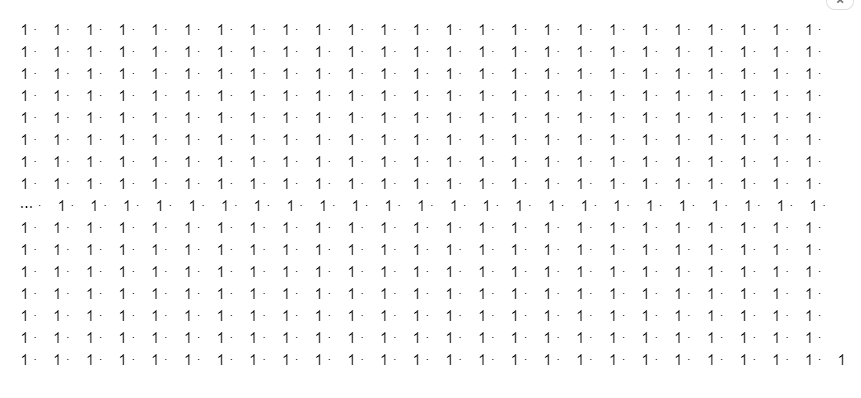
**Centers of clusters**





****

**Clusters**



**Size of clusters**

First cluster has 98903 users and the second has only 10 users.

🡺Clust1 >> Clust2

**Totss**

Totss = 5431197845194.74 : very big value

**withinss of clusters**

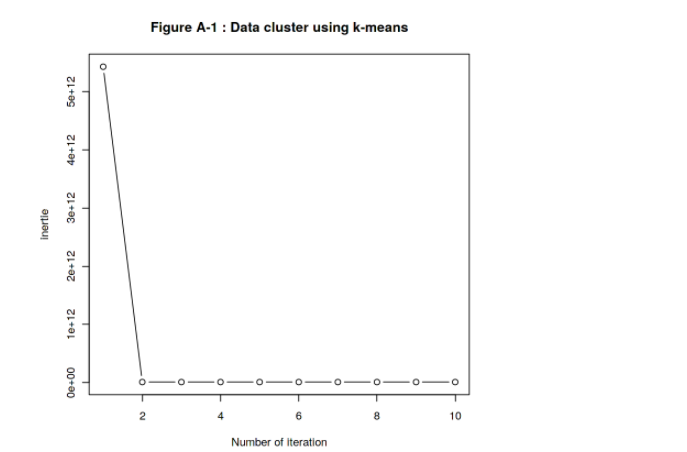
First clust : 8208374205.53964

Second clust : 44715.1

**tot.withinss**

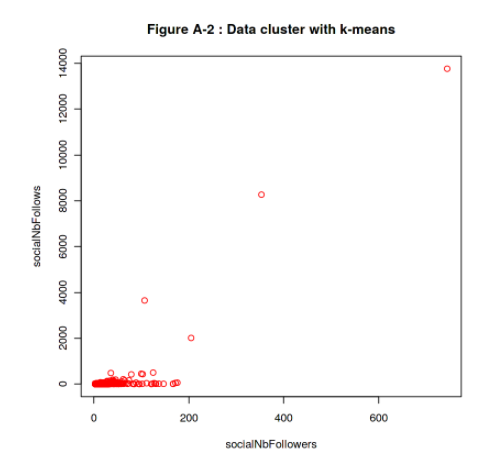
Tot.withinss: 8208418920.63964

**2-Evolution of inertie**

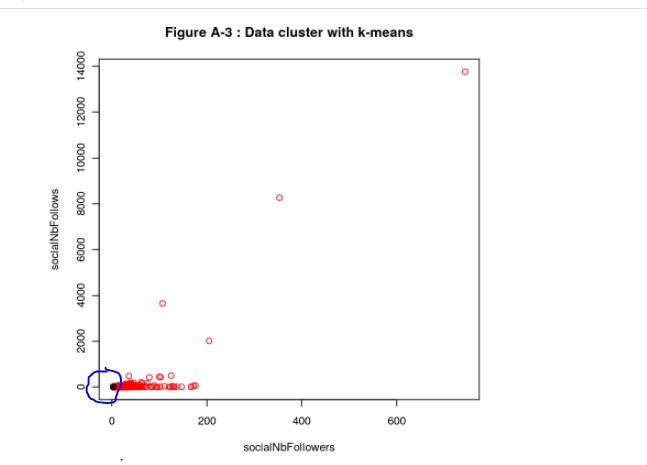


**🡺In the second time, the value of inertia is reduced and remains constant**

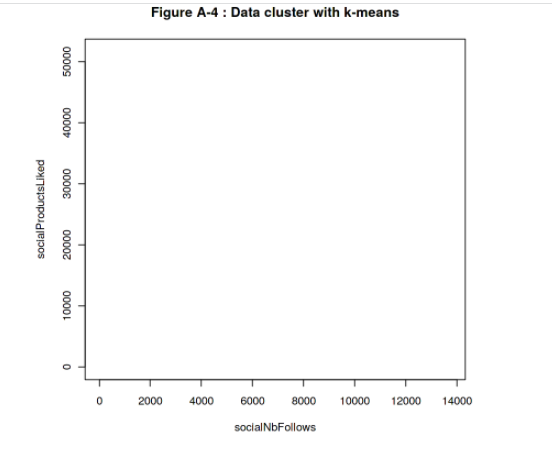
**3-Plot of clust**



* **There are some outliers**
* **Higher density in [0,200]**
* **We show only one clust because the other clust has only 10 items very smaller than the first clust**



* **The black group is very small**



* **No groups**

We thing the clustring of users is not powerful to create 2 cluster of users (leave or not leave the store)

**Data classification and validation**

**We have 18 variables and we replace two columns productsBought and** **productsSold with leaveStore because all users they buy or sold a product they stay in the store.**

**leaveStore has 2 class (“no” and “yes”)**

**🡺15 decision variables: inputs**

**🡺1 column (leaveStore): output**

**1) Data classification**

**Our tree structure:**

**Conditional inference tree with 54 terminal nodes**

**Response: leaveStore**

**Inputs: language, socialNbFollowers, socialNbFollows, socialProductsLiked, productsListed, productsPassRate, productsWished, gender, civilityTitle, hasAnyApp, hasAndroidApp, hasIosApp, hasProfilePicture, daysSinceLastLogin, seniorityAsMonths, countryCode**

**Number of observations: 65942**

**1) hasProfilePicture == {False}; criterion = 1, statistic = 8071.497**

**2) daysSinceLastLogin <= 233; criterion = 1, statistic = 322.369**

**3) productsPassRate <= 0; criterion = 1, statistic = 122.698**

**4) daysSinceLastLogin <= 17; criterion = 1, statistic = 44.809**

**5)\* weights = 183**

**4) daysSinceLastLogin > 17**

**6) hasAndroidApp == {True}; criterion = 0.979, statistic = 33.732**

**7)\* weights = 33**

**6) hasAndroidApp == {False}**

**8) socialProductsLiked <= 87; criterion = 0.996, statistic = 37.455**

**9) productsListed <= 0; criterion = 0.978, statistic = 32.947**

**10)\* weights = 196**

**9) productsListed > 0**

**11)\* weights = 52**

**8) socialProductsLiked > 87**

**12)\* weights = 23**

**3) productsPassRate > 0**

**13)\* weights = 229**

**2) daysSinceLastLogin > 233**

**14) productsPassRate <= 0; criterion = 1, statistic = 66.314**

**15) daysSinceLastLogin <= 644; criterion = 1, statistic = 40.509**

**16) socialNbFollowers <= 3; criterion = 0.994, statistic = 31.66**

**17) socialProductsLiked <= 2; criterion = 0.968, statistic = 9.5**

**18)\* weights = 72**

**17) socialProductsLiked > 2**

**19)\* weights = 7**

**16) socialNbFollowers > 3**

**20)\* weights = 222**

**15) daysSinceLastLogin > 644**

**21) socialNbFollows <= 8; criterion = 0.998, statistic = 14.723**

**22)\* weights = 196**

**21) socialNbFollows > 8**

**23)\* weights = 15**

**14) productsPassRate > 0**

**24)\* weights = 24**

**1) hasProfilePicture == {True}**

**25) socialNbFollowers <= 5; criterion = 1, statistic = 6058.467**

**26) socialNbFollowers <= 3; criterion = 1, statistic = 3652.25**

**27) socialProductsLiked <= 3; criterion = 1, statistic = 1414.29**

**28) productsWished <= 0; criterion = 1, statistic = 689.095**

**29)\* weights = 50861**

**28) productsWished > 0**

**30) daysSinceLastLogin <= 374; criterion = 1, statistic = 140.349**

**31) civilityTitle == {miss, mr}; criterion = 1, statistic = 52.053**

**32) gender == {F}; criterion = 0.956, statistic = 24.611**

**33)\* weights = 9**

**32) gender == {M}**

**34)\* weights = 190**

**31) civilityTitle == {mrs}**

**35) productsListed <= 0; criterion = 1, statistic = 43.994**

**36)\* weights = 723**

**35) productsListed > 0**

**37)\* weights = 15**

**30) daysSinceLastLogin > 374**

**38) productsWished <= 18; criterion = 1, statistic = 31.18**

**39) socialNbFollows <= 8; criterion = 0.985, statistic = 31.328**

**40)\* weights = 1563**

**39) socialNbFollows > 8**

**41)\* weights = 99**

**38) productsWished > 18**

**42)\* weights = 40**

**27) socialProductsLiked > 3**

**43) daysSinceLastLogin <= 146; criterion = 1, statistic = 265.529**

**44) productsWished <= 0; criterion = 0.999, statistic = 56.525**

**45)\* weights = 481**

**44) productsWished > 0**

**46) daysSinceLastLogin <= 11; criterion = 0.983, statistic = 39.194**

**47)\* weights = 69**

**46) daysSinceLastLogin > 11**

**48)\* weights = 591**

**43) daysSinceLastLogin > 146**

**49) socialNbFollows <= 8; criterion = 1, statistic = 63.29**

**50) daysSinceLastLogin <= 677; criterion = 1, statistic = 47.926**

**51)\* weights = 1125**

**50) daysSinceLastLogin > 677**

**52)\* weights = 442**

**49) socialNbFollows > 8**

**53) productsWished <= 19; criterion = 1, statistic = 17.649**

**54) hasAndroidApp == {True}; criterion = 0.959, statistic = 15.23**

**55)\* weights = 19**

**54) hasAndroidApp == {False}**

**56)\* weights = 111**

**53) productsWished > 19**

**57)\* weights = 8**

**26) socialNbFollowers > 3**

**58) daysSinceLastLogin <= 384; criterion = 1, statistic = 1277.87**

**59) daysSinceLastLogin <= 70; criterion = 1, statistic = 84.642**

**60) productsPassRate <= 0; criterion = 1, statistic = 49.211**

**61) productsWished <= 1; criterion = 0.999, statistic = 51.017**

**62)\* weights = 361**

**61) productsWished > 1**

**63)\* weights = 309**

**60) productsPassRate > 0**

**64)\* weights = 43**

**59) daysSinceLastLogin > 70**

**65) productsPassRate <= 0; criterion = 1, statistic = 45.162**

**66) gender == {M}; criterion = 1, statistic = 47.111**

**67)\* weights = 205**

**66) gender == {F}**

**68) productsWished <= 3; criterion = 0.985, statistic = 41.361**

**69)\* weights = 640**

**68) productsWished > 3**

**70)\* weights = 125**

**65) productsPassRate > 0**

**71)\* weights = 20**

**58) daysSinceLastLogin > 384**

**72) daysSinceLastLogin <= 690; criterion = 1, statistic = 188.943**

**73) productsWished <= 0; criterion = 1, statistic = 53.011**

**74) socialNbFollowers <= 4; criterion = 1, statistic = 52.513**

**75) gender == {M}; criterion = 0.999, statistic = 55.689**

**76)\* weights = 284**

**75) gender == {F}**

**77)\* weights = 838**

**74) socialNbFollowers > 4**

**78)\* weights = 386**

**73) productsWished > 0**

**79)\* weights = 255**

**72) daysSinceLastLogin > 690**

**80) seniorityAsMonths <= 95.23; criterion = 1, statistic = 45.201**

**81)\* weights = 1584**

**80) seniorityAsMonths > 95.23**

**82) daysSinceLastLogin <= 695; criterion = 1, statistic = 50.33**

**83)\* weights = 58**

**82) daysSinceLastLogin > 695**

**84) socialNbFollowers <= 4; criterion = 1, statistic = 46.454**

**85) productsWished <= 0; criterion = 1, statistic = 35.155**

**86) gender == {M}; criterion = 0.975, statistic = 35.006**

**87)\* weights = 279**

**86) gender == {F}**

**88)\* weights = 1094**

**85) productsWished > 0**

**89)\* weights = 52**

**84) socialNbFollowers > 4**

**90) daysSinceLastLogin <= 701; criterion = 1, statistic = 37.049**

**91)\* weights = 42**

**90) daysSinceLastLogin > 701**

**92)\* weights = 383**

**25) socialNbFollowers > 5**

**93) daysSinceLastLogin <= 377; criterion = 1, statistic = 265.065**

**94) productsPassRate <= 0; criterion = 1, statistic = 95.284**

**95) daysSinceLastLogin <= 16; criterion = 0.995, statistic = 56.338**

**96)\* weights = 156**

**95) daysSinceLastLogin > 16**

**97) civilityTitle == {miss, mr}; criterion = 0.966, statistic = 54.219**

**98)\* weights = 100**

**97) civilityTitle == {mrs}**

**99)\* weights = 373**

**94) productsPassRate > 0**

**100)\* weights = 251**

**93) daysSinceLastLogin > 377**

**101) productsPassRate <= 0; criterion = 1, statistic = 63.674**

**102) civilityTitle == {miss, mr}; criterion = 1, statistic = 66.87**

**103)\* weights = 126**

**102) civilityTitle == {mrs}**

**104) countryCode == {ch, cn, dk, fi, fr, gb, gu, hr, hu, it, la, lb, mc, nl, nz, pr, ro, sa, se, tw, ua, us}; criterion = 0.992, statistic = 67.857**

**105)\* weights = 268**

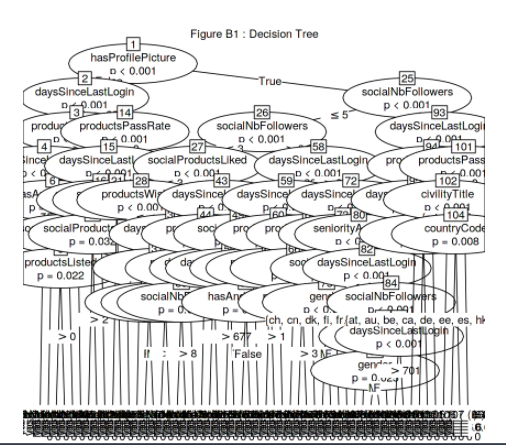
**104) countryCode == {at, au, be, ca, de, ee, es, hk, ie, jp, lt, sg, sk}**

**106)\* weights = 93**

**101) productsPassRate > 0**

**107)\* weights = 19**

**plot(tree)**



**There are many nodes we can't read the graphic.**

**Many decision variable 🡺 many nodes**

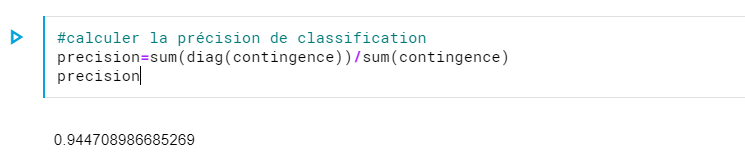
**Prédiction**



**Prédiction table**



**Calculer la précision**



**Calculer l’erreur**

**Erreur = 1-precision =**0.0552910133147311

**==>Donc on obtient un modèle très fortes puisque la précision plus que 94%**