MARKETING DATA MANUAL

Business Analytics Tools Open Source

Brigth Arias and Josephine Schneegans

2022-12-22

Introduction

The data mart called DataMartMarketing has been created in order to understand the behavior of gambling players, according to different characteristics and actions. To do so, we have used information from three datasets that gave us demographic, transactions and daily utilisation information but all poker insight.

Before starting the interpretation of the data, we had to clean and transform them in order to have useful and adequate datas. We were then able to create an organized datamart to optimize the creation of marketing insights like the Shiny App (vizualisation).

Datasets

We have started the work with three datasets, given in .SAS format read by 'read_sas' function.

`DataI_Demo`

```
##  [1] "ApplicationID" "Country"       "FirstAct"      "FirstCa"
##  [5] "FirstGa"       "FirstPay"      "FirstPo"       "FirstSp"
##  [9] "Gender"        "Language"      "RegDate"       "UserID"
```

`DataII_Uda`

```
## [1] "Bets"      "Date"      "ProductID" "Stakes"    "UserID"    "Winnings"
```

`DataIII_Pcc`

```
## [1] "TransAmount"   "TransDateTime" "TransType"     "UserID"
```

Data Processing

Raw Datasets Cleaning

`DataI_Demo`

- UserID was in numeric format so we changed it into character as we do not want UserID to be considerated as a number

- Dates were not in dates format, there were in character, numeric, date, integer… so we ha converting them into dates format

`DataII_Uda`

- UserID was in numeric format so we changed it into character as we do not want UserID to be considerated as a number
- The data was in charater so we change it as date format
- In order to have more practical numbers we have rouded Stakes and Winnings

`DataIII_Pcc`

- UserID was in numeric format so we changed it into character as we do not want UserID to be considerated as a number
- TransDateTime were not in dates format, there were in characterso we have converting them into dates format
- In order to have more practical numbers we have rouded TransAmount variable

Merging

We have created a DataMart by a full join merging between DataII_Uda and DataIII_Pcc on 'ProductID' and 'UserID' then the result has been left join with DataI_Demo on 'UserID'.

DataMart Cleaning

- We have drop NA value for FirstAct variable in order to take off clients who have missing value for the first active date
- We have excule record that from UserDailyAggregation that took place before the first pay-in date

Data Mart Marketing

Metric

`LOR`

DataMartMarketing$LOR < -as.integer(difftime(DataMartMarketing$Last_Active_Date, DataMartMarketing\$First_Active_Date, units = c("days")))

Here we did a segmentation about loyalty of clients. To do so, we use the RFM method (recency/frequency/monetary) and we segmented the client into 5 categories:

- Champions are your best customers, who bought most recently, most often, and are heavy spenders. Reward these customers. They can become early adopters for new products and will help promote your brand.
- Potential Loyalists are your recent customers with average frequency and who spent a good amount. Offer membership or loyalty programs or recommend related products to upsell them and help them become your Loyalists or Champions.

- New Customers are your customers who have a high overall RFM score but are not frequent shoppers. Start building relationships with these customers by providing onboarding support and special offers to increase their visits.
- At Risk Customers are your customers who purchased often and spent big amounts, but haven't purchased recently. Send them personalized reactivation campaigns to reconnect, and offer renewals and helpful products to encourage another purchase.
- Can't Lose Them are customers who used to visit and purchase quite often, but haven't been visiting recently. Bring them back with relevant promotions, and run surveys to find out what went wrong and avoid losing them to a competitor.

`RFM_score_betting`

RFM_score_betting = 100 *R_score +10* F_score + M_score

`Loyalty_segmentation_bettings`

Loyalty_segmentation_bettings= case_when(

between(RFM_score_betting,0,89)~"Can't Lose Them",

between(RFM_score_betting,90,178)~"At Risk Customers",

between(RFM_score_betting,179,266)~"New Customers",

between(RFM_score_betting,267,355)~"Potential Loyalists",

TRUE~"Champions"))

`RFM_score_pokerChip`

RFM_score_pokerChip = 100 *R_score +10* F_score + M_score

`Loyalty_segmentation_pokerChip`

Loyalty_segmentation_pokerChip= case_when(

between(RFM_score_pokerChip,0,89)~"Can't Lose Them",

between(RFM_score_pokerChip,90,178)~"At Risk Customers",

between(RFM_score_pokerChip,179,266)~"New Customers",

between(RFM_score_pokerChip,267,355)~"Potential Loyalists",

TRUE~"Champions"))

Here we have calculations (min/max/avg/sum) of transaction amount (buy and sell)

`MaxPokerTransDate`

MaxPokerTransDate = max(TransDatePoker)

`MinPokerTransDate`

MinPokerTransDate = min(TransDatePoker)

`TotalPokerTranSell`

TotalPokerTranSell = sum(TransAmount[TransType==24])

`TotalPokerTranBuy`

TotalPokerTranBuy = sum(TransAmount[TransType==124])

`MaxPokerTranSell`

MaxPokerTranSell = max(TransAmount[TransType==24])

`MaxPokerTranBuy`

MaxPokerTranBuy = max(TransAmount[TransType==124])

`MinPokerTranSell`

MinPokerTranSell = min(TransAmount[TransType==24])

`MinPokerTranBuy`

MinPokerTranBuy = min(TransAmount[TransType==124])

`AvgPokerTranSell`

AvgPokerTranSell = mean(TransAmount[TransType==24])

`AvgPokerTranBuy`

AvgPokerTranBuy = mean(TransAmount[TransType==124])

`CountPokerTranSell`

CountPokerTranSell = sum(TransType==24)

`CountPokerTranBuy`

CountPokerTranBuy = sum(TransType==124)

`NumberOfProducts` *calculation of the total product per customer*

NumberOfProducts = length(unique(ProductID, na.rm = TRUE))

`Total_Stakes` *calculation of the total stakes per clients*

total_stakes <- data.frame(UserID = DataMart$UserID, Stakes = DataMart$Stakes)

head(total_stakes)

total_stakes <- total_stakes %>%

group_by(UserID) %>%

summarise(Total_Stakes = sum(Stakes, na.rm = TRUE))

`Total_Winnings` *calculation of the total winnings by clients*

total_winnings <- data.frame(UserID = DataMart$UserID, Winnings = DataMart$Winnings)

head(total_winnings)

total_winnings <- total_winnings %>%

group_by(UserID) %>%

summarise(Total_Winnings = sum(Winnings, na.rm = TRUE))

`Total_Bets` *calculation of the total bets per clients*

total_bets <- data.frame(UserID = DataMart$UserID, Bets = DataMart$Bets)

total_bets <- total_bets %>%

group_by(UserID) %>%

summarise(Total_Bets = sum(Bets, na.rm = TRUE))

`Final_balance`

final_balance = Total_Winnings - Total_Stakes

Here we have calculated some variables in order to have metrics per visit

`NumDaysOfVisited`

NumDaysOfVisited = n_distinct(Date)

`Avg_winnings_per_visit`

avg_winnings_per_visit = mean(Winnings)

`Avg_stakes_per_visit`

avg_stakes_per_visit = mean(Stakes)

`Avg_bets_per_visit`

avg_bets_per_visit = mean(Bets)

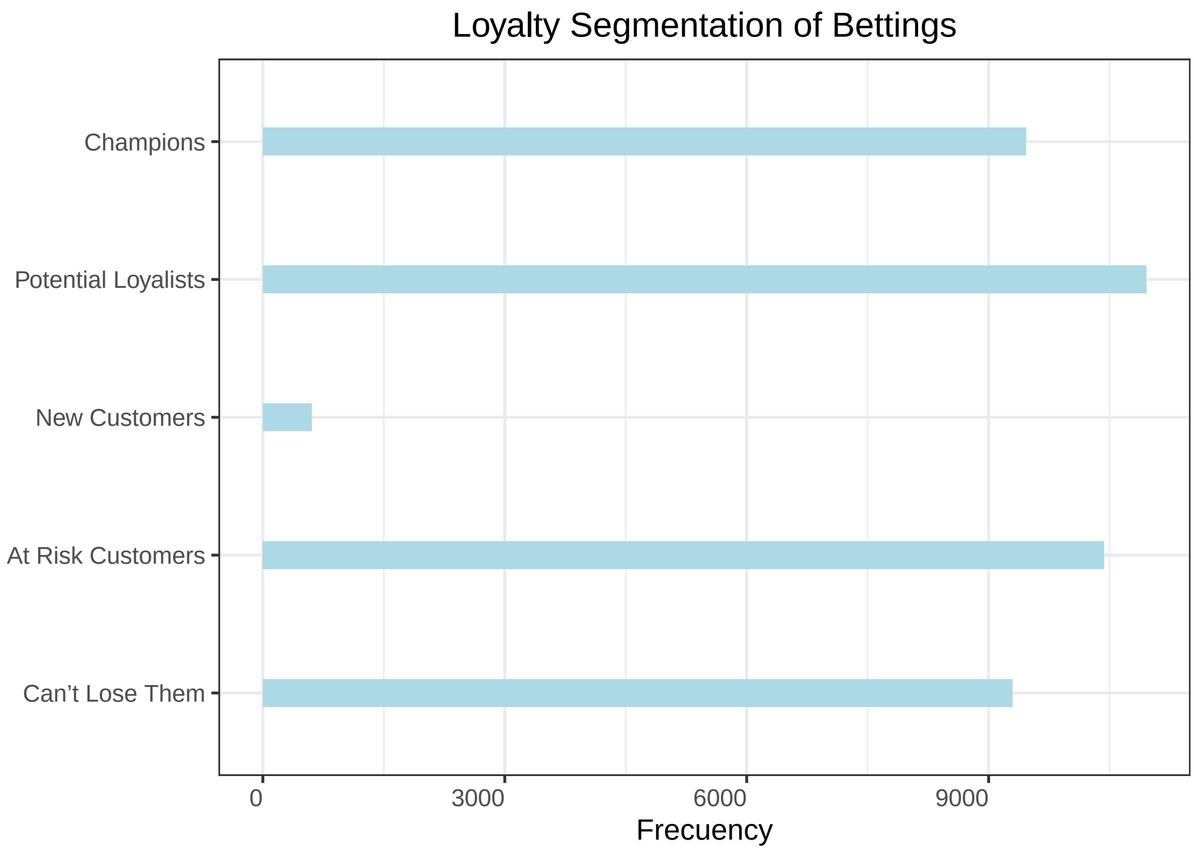
`Ratio_win_stake`

ratio_win_stake = avg_winnings_per_visit/avg_stakes_per_visit

Marketing Insight

Bettings and loyalty

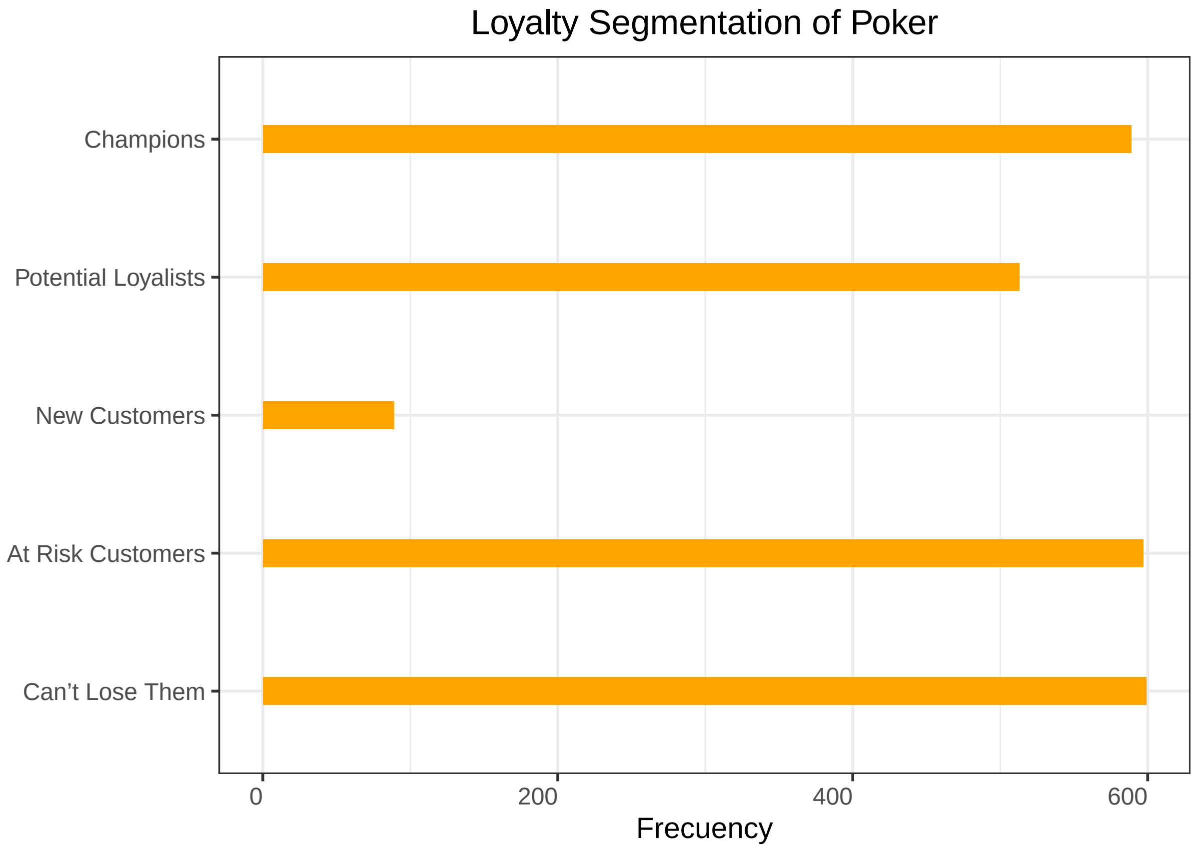
The following barchart is representing the number of client of each Bettings Segmentation:



Here we can see the number of bettings per segment. The most represented segment is the Potential Loyalist, with more than 10500 bettings. They are the recent customers with average frequency and good amount spending and the less represented segment is the New Customers with less than 750 bettings. They are customers that have a high RFM but they are not frequent clients.

Pokerchip and Loyalty

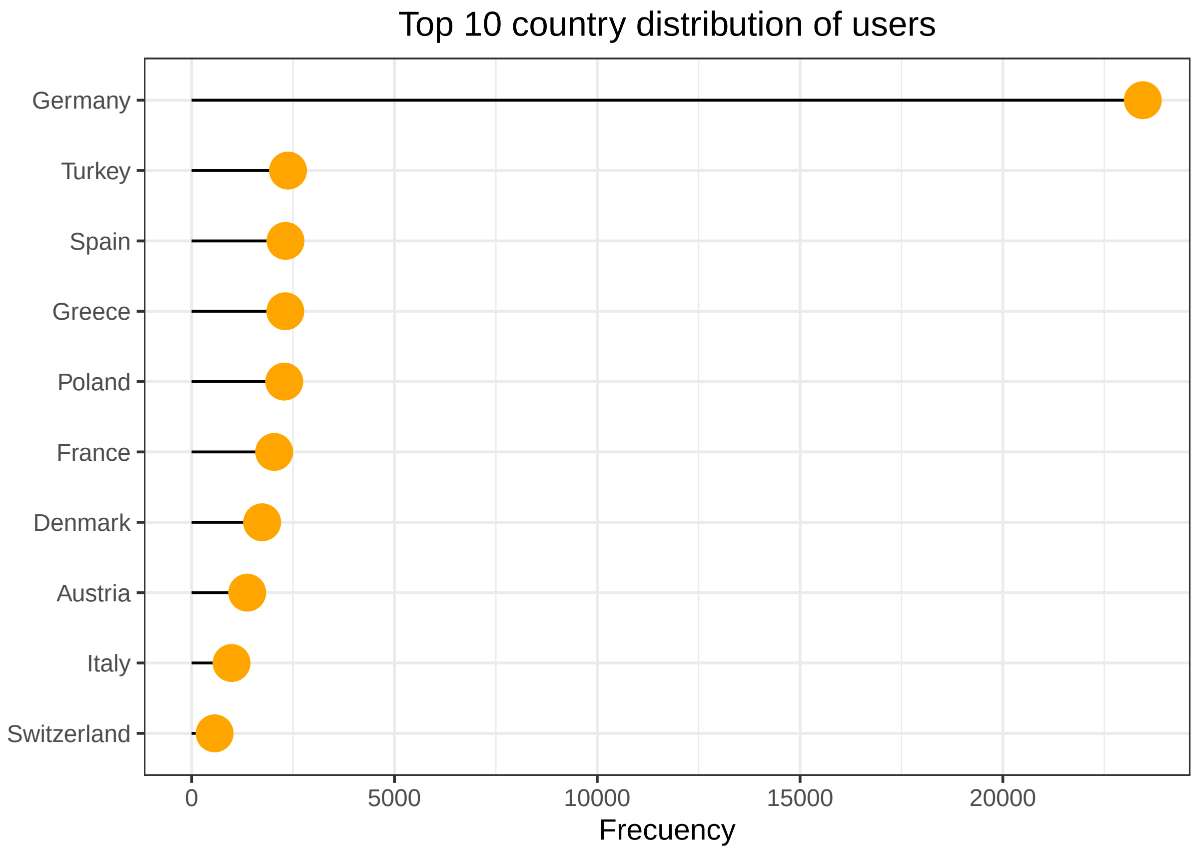
The following barchart is representing the number of client of each PockerChip Segmentation:



Here we can see the number of bettings per segment. The most represented segment are the At Risk Customers and the Can't Lose Them, with 600 PockerChip purchase1. At Risk Customers are the one that are often purchasing with big amount but not recently and Can't Lose Them are the clients who used to purchased and visit often but not recently. The less represented segment is the New Customers with less than 100 PockerChip purshase. They are customers that have a high RFM but they are not frequent clients.

Client and Country

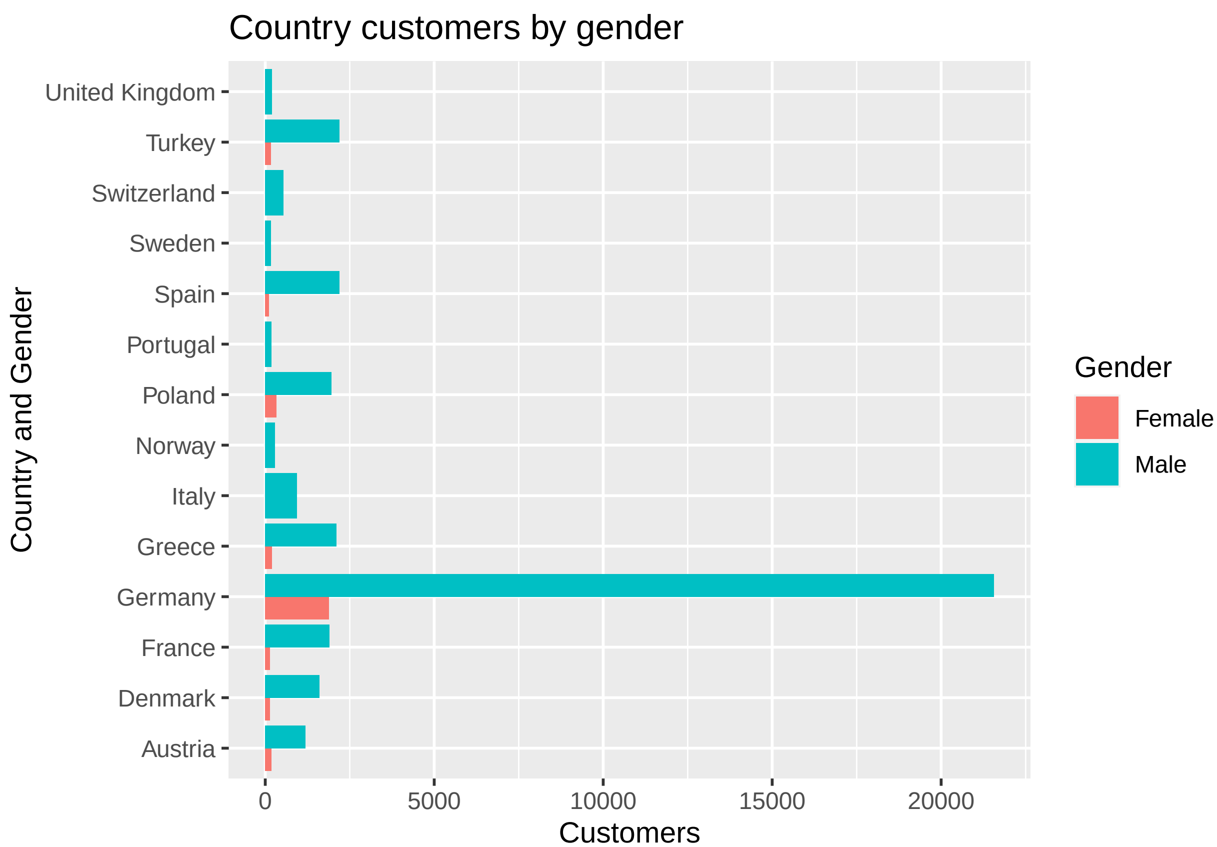
The following barchart is representing the number of client of each different country:



The most represented Country is by far Germany with more than 22500 clients. Then we have a group of four Country, Turkey, Spain, Greece and Poland that all have approximately 2500 clients and then we have France, Danemark, Autria, Italy and Switzerland that have between 1300 and 600 clients.

Gender repartition by country

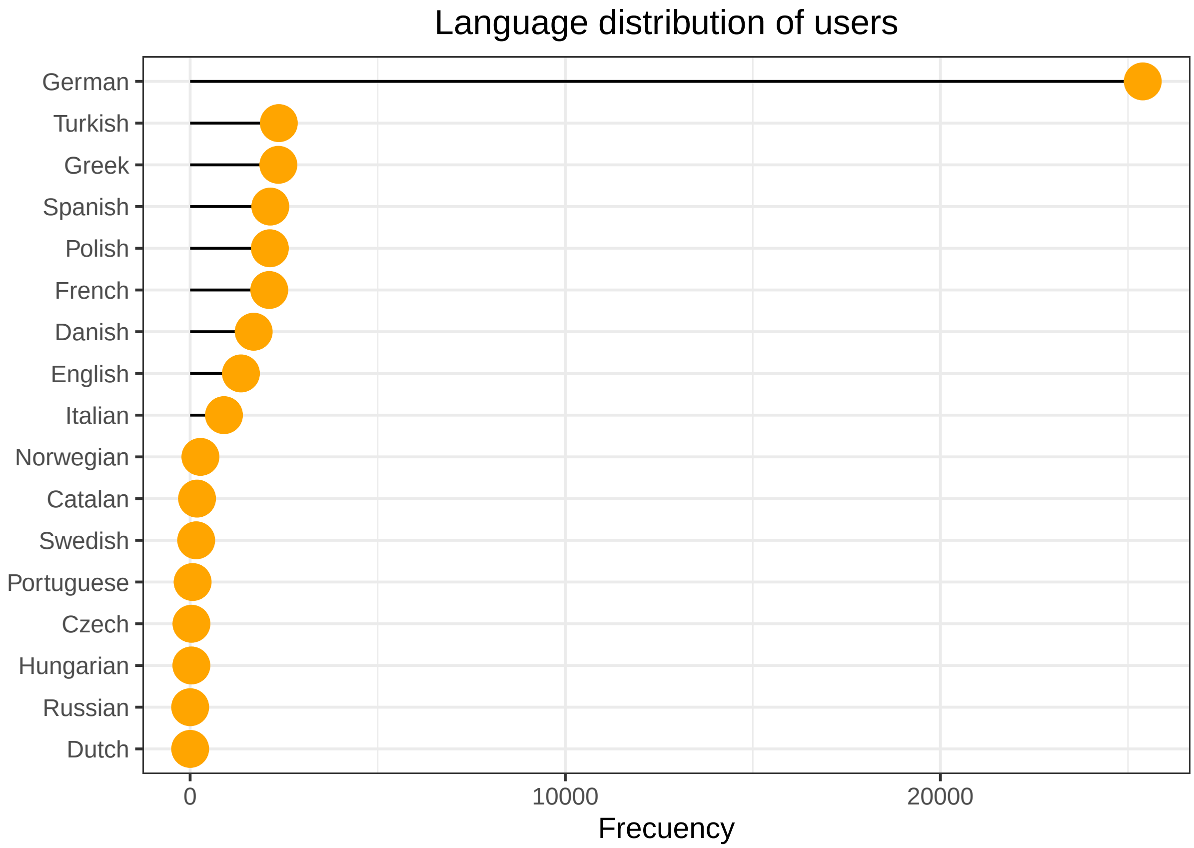
The following graphic is showing the client's repartition of female and male by country:



Here we have the customer group by gender and country. The most represented clients are male from Germany. We can see that in every country, female are really less represented, so we can assume that in general, clients are mostly males.

Language repartition

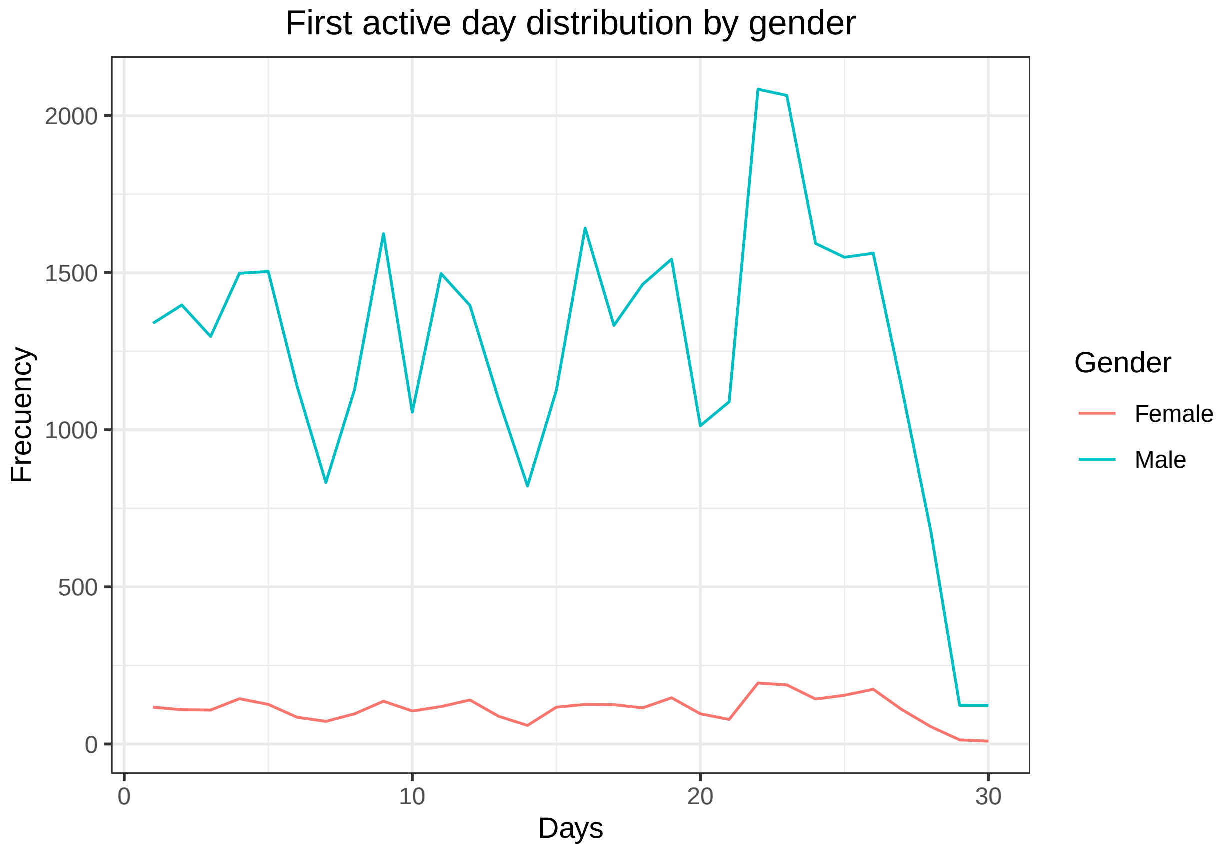
The following graphic is showing the client language's repartition:



As we see previously, Germany is the country with the most clients. We can see on this graph that German is the most spoken language with more than 25000 clients, and then we have a group of five languages: Turkish, Greek, Spanish, Polish and French with a little bit less than 2500 clients, then we have Danish, English, Italian ad Norwegian speakers that have less than 4000 clients.

First active day by gender

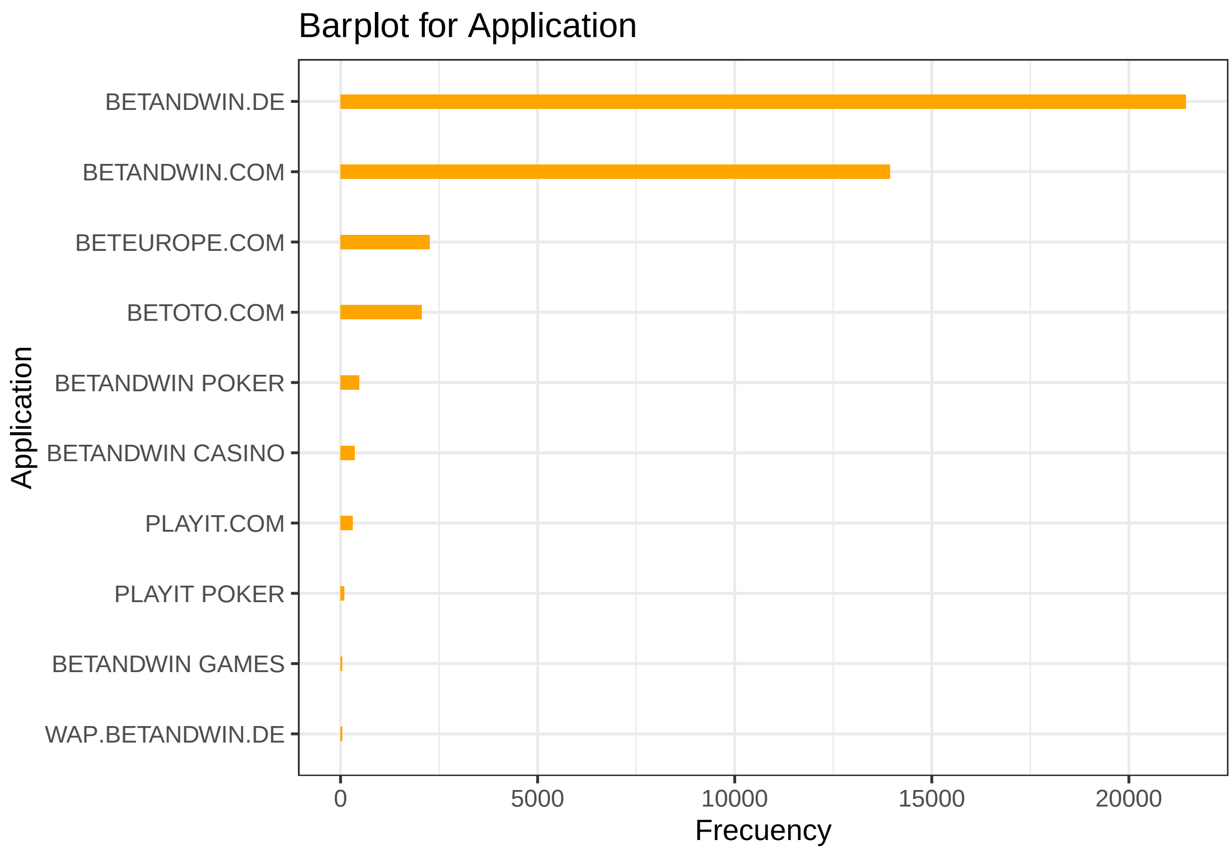
The following plot shows the frecuency of first active day of the month of the users:



In this graph we can see two important things: First the male has more active days frecuency of first activation day than female, and there are more first activation during the last days of the month.

Application repartition

The following barchart is representing the number of client for each different application:



As we can see, the top 5 routes more used to access to bwin are Betandwin.de, Betandwin.com, Beteurope.com and Betoto.com.

Link of the application

The synthesis of the analysis has been created by Shiny app, which can be viewed at the following link: https://brigthygaby.shinyapps.io/OnlineGambling/

References

https://medium.com/analytics-vidhya/customer-segmentation-using-rfm-analysis-in-r-cd8ba4e6891

https://www.programmingr.com/rfm-analysis/

http://rstudio.github.io/shinydashboard/

https://github.com/fondaa/bigdata/blob/main/Gambling%20Company%20Market%20Analysis%20with%20R.R

https://github.com/Tutoman/Betwin-Datamart/blob/main/Datamart_Code.ipynb

https://clevertap.com/blog/rfm-analysis/#:~:text=RFM%20Score%20Finally%2C%20we%20can%20rank%20these%20custo

https://rstudio.github.io/shinydashboard/

https://stackoverflow.com/questions/24027605/determine-the-number-of-na-values-in-a-column

https://r-graph-gallery.com/267-reorder-a-variable-in-ggplot2.html

https://9to5answer.com/finding-the-max-of-a-r-dataframe-column-ignoring-inf-and-na

https://github.com/rstudio/shiny-examples/tree/main/087-crandash