Group\_Project\_Markdown

# Project Introduction

Group Assignment Open Source Programming 2019  
Prof. Dr. Matthijs Meire  
Group : BORCHERT Philipp, CELI Fabian, HUTIN Claire  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Introduction:

The purpose of this project is to provide an online gamblind website some  
basic marketing descriptives for the customer base and create a marketing data mart per customer.  
In order to do so, data was cleaned and aggregated by User, to later explore some variables that made sense for us to achieve our task.

In this project, we were given several databases. We aggregated them in one final data mart in which we focused on the main marketing metrics  
for the gambling company.  
As marketing metrics are required, we focused on the customer profile of the gamblers, e.g. where they come from, how old they are, their loyalty.

We of course looked at where they gamble the most at, the bettings frequency, what they can expect to win based on the amount of bets they make  
and the amount of money they bet each time.

Our aim is to make this report and Shiny App as readable as possible for the betting company to find out some key patterns in amoungs their  
customers’ habits.

# Data Description:

The data sets contain information about an online gambling website, with information from February to September 2005.

In the code and DataMart the datasets were save as follow:

|  |  |
| --- | --- |
| File Name | Code Name |
| “RawDataIIIPokerChipConversions.sas7bdat” | conversion |
| “AnalyticDataInternetGambling.sas7bdat” | ADIG |
| “RawDataIDemographics.sas7bdat” | Demo |
| “RawDataIIUserDailyAggregation.sas7bdat” | DaylyAGG |

The data sets were cleaned and merged starting from Demographic data set because it contains the information per costumer.

In general, the majority of gamblers are males. The continent with more clients is Europe, particularly Germany. The most common age of clients is 25 years old.

The two main gambling products are Betted\_Casino\_chartwell and Betted\_Sports\_fixed\_odd. However, the average winnings are in Betted\_Casino\_chartwell and Betted\_Sports\_fixed\_live.

For further and interactive analysis please click on the following link:

<https://pborieseg.shinyapps.io/R_Assignment/>

# Variable description

The following table will present the variables in the final datamart and a description of each of these:

|  |  |
| --- | --- |
| Variable Name | Description |
| UserID | User ID was of each participant at time of registration assigned by bwin |
| total\_conv\_amount | The sum of poker chip transaction amount per User ID |
| avg\_conv\_amount | The average of poker chip transaction amount per User ID |
| total\_sell\_amount | The total amount sold per User ID |
| total\_buy\_amount | The total amount bought per User ID |
| avg\_timediff\_hours | The average time difference between conversions |
| nr\_conversions | The number of poker chip transactions per User ID |
| RegistrationDate | Date of registration of each participant |
| Age | Participant’s age |
| FOTotalStakes | Total stakes each participant betted on fixed-odds plays from February 1, 2005 through September 30, 2005 |
| FOTotalWinnings | Total winnings each participant received on fixed-odds plays from February 1, 2005 through September 30, 2005 |
| FOTotalBets | Total number of bets each participant betted on fixed-odds plays from February 1, 2005 through September 30, 2005 |
| FOFirstActiveDate | First date of fixed-odds play of each participant from February 1, 2005 through September 30, 2005 |
| FOLastActiveDate | Last date of fixed-odds play of each participant from February 1, 2005 through September 30, 2005 |
| FOTotalDaysActive | Total days of active fixed-odds plays of each participant from February 1, 2005 through September 30, 2005 |
| LATotalStakes | Total stakes each participant betted on live-action plays from February 1, 2005 through September 30, 2005 |
| LATotalWinnings | Total winnings each participant received on live-action plays from February 1, 2005 through September 30, 2005 |
| LATotalBets | Total number of bets each participant betted on live-action plays from February 1, 2005 through September 30, 2005 |
| LAFirstActiveDate | First date of live-action play of each participant from February 1, 2005 through September 30, 2005 |
| LALastActiveDate | Last date of live-action play of each participant from February 1, 2005 through September 30, 2005 |
| LATotalDaysActive | Total days of active live-action plays of each participant from February 1, 2005 through September 30, 2005 |
| FirstSportsActiveDate | First date of active sports book play for each participant from February 1, 2005 through September 30, 2005 |
| RegDate | Date of registration |
| FirstPay | Participant’s first betting money deposits date |
| FirstAct | Participant’s first active play date |
| First\_time\_sports | Participant’s first sports book play date |
| First\_time\_casino | Participant’s first casino play date |
| First\_time\_games | Participant’s first games play date |
| First\_time\_poker | Participant’s first poker play date |
| ApplicationID | Different route of access to bwin (ID) |
| avg\_bet | Average number of daily bets per participant |
| min\_bet | Minimum number of bets made by a participant |
| max\_bet | Maximum number of bets made by a participant |
| total\_number\_bets | Total number of bets the participant has made |
| avg\_winnings | Average betting winnings per participant |
| min\_winnings | Minimum betting winnings per participant |
| max\_winnings | Maximum betting winnings per participant |
| total\_money\_bets | Total betting stakes per participant, in € |
| min\_money\_bets | Minimum betting stakes per participant, in € |
| max\_money\_bets | Maximum betting stakes per participant ,in € |
| mean\_money\_bets | Average betting stakes per participant ,in € |
| Betted\_Casino\_boss | Number of bets the participant made for this product category |
| Betted\_Casino\_chartwell | Number of bets the participant made for this product category |
| Betted\_Games\_bwin | Number of bets the participant made for this product category |
| Betted\_Games\_VS | Number of bets the participant made for this product category |
| Betted\_Sports\_fixed\_live | Number of bets the participant made for this product category |
| Betted\_Sports\_fixed\_odd | Number of bets the participant made for this product category |
| Betted\_Supertoto | Number of bets the participant made for this product category |
| Gender | Participant’s Gender (1: Female; 0: Male) |
| Language | Participant’s language code |
| Country | Participant’s country of residence code |
| country\_name | Participant’s country of residence name |
| country | Participant’s country of residence abbreviation |
| latitude | Participant’s country of residence latitude |
| longitude | Participant’s country of residence longitude |
| Continent\_Name | Participant’s country of residence continent name |
| Continent\_Code | Participant’s country of residence continent code |
| application\_description | Name of the different routes of access to bwin |
| TotalDaysActive | Number of days each participant was active |
| gender\_description | Participant’s gender |
| LastActiveDate | Last date each participant was active |

# Summary statistics of the datamart

As the summary statistics show,on average users transact €36.30 per day, when the average total amount per user is of €7930. Users have made 116 different transactions on average during the given period. Users are on average 31 years old., and 92% of them are Male. Out of all users, they bet around 9 times during the time period, considering that each bet made is around 71€,  
when the winning are on average of €63? which means that betters usually lose money. They can, however, expect great earning,  
as the average maximum bet per user is of€481. Most betters are European, as the below bar chart shows.

library(readr)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
datamart <- read\_csv("C:/Users/hutin/Desktop/IESEG/MASTER BIG DATA/R/datamart.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegistrationDate = col\_date(format = ""),  
## FOFirstActiveDate = col\_date(format = ""),  
## FOLastActiveDate = col\_date(format = ""),  
## LAFirstActiveDate = col\_date(format = ""),  
## LALastActiveDate = col\_date(format = ""),  
## FirstSportsActiveDate = col\_date(format = ""),  
## RegDate = col\_date(format = ""),  
## FirstPay = col\_date(format = ""),  
## FirstAct = col\_date(format = ""),  
## First\_time\_sports = col\_date(format = ""),  
## First\_time\_casino = col\_date(format = ""),  
## First\_time\_games = col\_date(format = ""),  
## First\_time\_poker = col\_date(format = ""),  
## country\_name = col\_character(),  
## country = col\_character(),  
## Continent\_Name = col\_character(),  
## Continent\_Code = col\_character(),  
## application\_description = col\_character(),  
## gender\_description = col\_character(),  
## LastActiveDate = col\_date(format = "")  
## )

## See spec(...) for full column specifications.

summary(datamart, digits = 3)

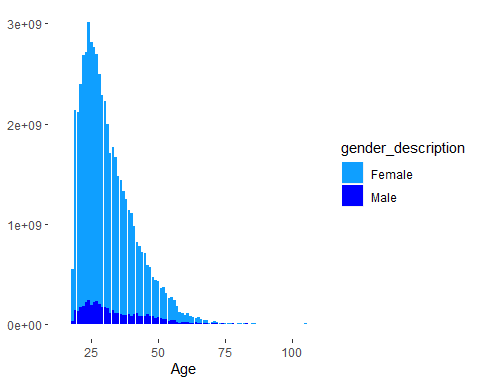
## UserID total\_conv\_amount avg\_conv\_amount total\_sell\_amount   
## Min. :1324354 Min. : 0 Min. : 0.01 Min. :0.00e+00   
## 1st Qu.:1344918 1st Qu.: 70 1st Qu.: 7.11 1st Qu.:4.88e+01   
## Median :1364861 Median : 419 Median : 13.99 Median :2.59e+02   
## Mean :1364846 Mean : 7930 Mean : 36.60 Mean :4.10e+03   
## 3rd Qu.:1384877 3rd Qu.: 2301 3rd Qu.: 29.37 3rd Qu.:1.33e+03   
## Max. :1405190 Max. :4506058 Max. :6964.54 Max. :2.25e+06   
## NA's :41908 NA's :41908 NA's :41908   
## total\_buy\_amount avg\_timediff\_hours nr\_conversions   
## Min. :0.00e+00 Min. : 0.00 Min. : 1   
## 1st Qu.:1.81e+01 1st Qu.: 3.13 1st Qu.: 5   
## Median :1.58e+02 Median : 12.10 Median : 29   
## Mean :3.83e+03 Mean : 42.59 Mean : 116   
## 3rd Qu.:9.92e+02 3rd Qu.: 35.89 3rd Qu.: 128   
## Max. :2.26e+06 Max. :3937.98 Max. :2968   
## NA's :41908 NA's :41928 NA's :41908   
## RegistrationDate Age FOTotalStakes FOTotalWinnings   
## Min. :2015-02-02 Min. : 14.0 Min. : 0 Min. : 0.0   
## 1st Qu.:2015-02-10 1st Qu.: 24.0 1st Qu.: 51 1st Qu.: 16.9   
## Median :2015-02-17 Median : 29.0 Median : 148 Median : 100.8   
## Mean :2015-02-16 Mean : 31.4 Mean : 729 Mean : 632.6   
## 3rd Qu.:2015-02-23 3rd Qu.: 37.0 3rd Qu.: 450 3rd Qu.: 368.8   
## Max. :2015-02-28 Max. :105.0 Max. :380094 Max. :420949.3   
## NA's :3796 NA's :3796 NA's :4576 NA's :4576   
## FOTotalBets FOFirstActiveDate FOLastActiveDate   
## Min. : 1 Min. :2015-02-02 Min. :2015-02-02   
## 1st Qu.: 11 1st Qu.:2015-02-13 1st Qu.:2015-03-21   
## Median : 36 Median :2015-02-22 Median :2015-07-04   
## Mean : 135 Mean :2015-02-27 Mean :2015-06-24   
## 3rd Qu.: 109 3rd Qu.:2015-02-28 3rd Qu.:2015-09-25   
## Max. :26345 Max. :2015-09-30 Max. :2015-10-01   
## NA's :4576 NA's :4576 NA's :4576   
## FOTotalDaysActive LATotalStakes LATotalWinnings LATotalBets   
## Min. : 1.0 Min. : 0.0 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 5.0 1st Qu.: 15.0 1st Qu.: 6.5 1st Qu.: 4.0   
## Median : 14.0 Median : 60.5 Median : 45.0 Median : 15.0   
## Mean : 24.3 Mean : 1318.6 Mean : 1233.6 Mean : 99.3   
## 3rd Qu.: 32.0 3rd Qu.: 281.3 3rd Qu.: 241.8 3rd Qu.: 60.0   
## Max. :310.0 Max. :384518.3 Max. :375006.0 Max. :21230.0   
## NA's :4576 NA's :19501 NA's :19501 NA's :19501   
## LAFirstActiveDate LALastActiveDate LATotalDaysActive  
## Min. :2015-02-02 Min. :2015-02-02 Min. : 1.0   
## 1st Qu.:2015-02-17 1st Qu.:2015-03-10 1st Qu.: 2.0   
## Median :2015-02-27 Median :2015-05-13 Median : 5.0   
## Mean :2015-03-16 Mean :2015-06-02 Mean : 13.5   
## 3rd Qu.:2015-03-17 3rd Qu.:2015-09-07 3rd Qu.: 14.0   
## Max. :2015-10-01 Max. :2015-10-01 Max. :234.0   
## NA's :19501 NA's :19501 NA's :19502   
## FirstSportsActiveDate RegDate FirstPay   
## Min. :2015-02-02 Min. :2005-02-01 Min. :2005-02-01   
## 1st Qu.:2015-02-13 1st Qu.:2005-02-09 1st Qu.:2005-02-11   
## Median :2015-02-22 Median :2005-02-16 Median :2005-02-20   
## Mean :2015-02-27 Mean :2005-02-15 Mean :2005-02-28   
## 3rd Qu.:2015-02-28 3rd Qu.:2005-02-22 3rd Qu.:2005-02-26   
## Max. :2015-09-30 Max. :2005-02-27 Max. :2005-10-02   
## NA's :3796 NA's :1646 NA's :1646   
## FirstAct First\_time\_sports First\_time\_casino   
## Min. :2005-02-01 Min. :2005-02-01 Min. :2005-02-01   
## 1st Qu.:2005-02-12 1st Qu.:2005-02-12 1st Qu.:2005-02-16   
## Median :2005-02-21 Median :2005-02-21 Median :2005-02-26   
## Mean :2005-03-01 Mean :2005-03-02 Mean :2005-03-25   
## 3rd Qu.:2005-02-27 3rd Qu.:2005-02-27 3rd Qu.:2005-04-07   
## Max. :2005-10-02 Max. :2005-10-02 Max. :2005-10-02   
## NA's :1648 NA's :2882 NA's :38407   
## First\_time\_games First\_time\_poker ApplicationID   
## Min. :2005-02-01 Min. :2005-02-01 Min. : 1.00   
## 1st Qu.:2005-02-19 1st Qu.:2005-02-12 1st Qu.: 1.00   
## Median :2005-03-04 Median :2005-02-24 Median : 3.00   
## Mean :2005-04-05 Mean :2005-03-31 Mean : 3.91   
## 3rd Qu.:2005-05-04 3rd Qu.:2005-04-26 3rd Qu.: 3.00   
## Max. :2005-10-02 Max. :2005-10-02 Max. :42.00   
## NA's :41022 NA's :42153 NA's :1646   
## avg\_bet min\_bet max\_bet total\_number\_bets  
## Min. : 0.273 Min. : -1.00 Min. : 1.0 Min. : 1   
## 1st Qu.: 1.358 1st Qu.: 0.00 1st Qu.: 4.0 1st Qu.: 16   
## Median : 2.306 Median : 0.00 Median : 9.0 Median : 61   
## Mean : 8.815 Mean : 1.16 Mean : 54.2 Mean : 362   
## 3rd Qu.: 4.631 3rd Qu.: 0.00 3rd Qu.: 26.0 3rd Qu.: 200   
## Max. :1986.604 Max. :1343.00 Max. :6223.0 Max. :193442   
## NA's :444 NA's :444 NA's :444 NA's :444   
## avg\_winnings min\_winnings max\_winnings   
## Min. : 0.00 Min. : -40.00 Min. : 0.0   
## 1st Qu.: 1.93 1st Qu.: 0.00 1st Qu.: 15.1   
## Median : 6.34 Median : 0.00 Median : 48.7   
## Mean : 62.20 Mean : 9.37 Mean : 498.2   
## 3rd Qu.: 20.62 3rd Qu.: 0.00 3rd Qu.: 159.6   
## Max. :63100.00 Max. :22790.00 Max. :270222.0   
## NA's :444 NA's :444 NA's :444   
## total\_money\_bets min\_money\_bets max\_money\_bets mean\_money\_bets   
## Min. :0.00e+00 Min. : -28.3 Min. : 0 Min. : 0.00   
## 1st Qu.:7.77e+01 1st Qu.: 0.0 1st Qu.: 20 1st Qu.: 5.23   
## Median :2.34e+02 Median : 0.0 Median : 44 Median : 11.41   
## Mean :2.77e+03 Mean : 12.6 Mean : 481 Mean : 71.01   
## 3rd Qu.:8.19e+02 3rd Qu.: 0.0 3rd Qu.: 129 3rd Qu.: 30.00   
## Max. :1.13e+06 Max. :23790.0 Max. :272416 Max. :64600.00   
## NA's :444 NA's :444 NA's :444 NA's :444   
## Betted\_Casino\_boss Betted\_Casino\_chartwell Betted\_Games\_bwin  
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 0.000 Median : 0.000   
## Mean : 0.387 Mean : 0.757 Mean : 0.175   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.000   
## Max. :170.000 Max. :154.000 Max. :171.000   
## NA's :444 NA's :444 NA's :444   
## Betted\_Games\_VS Betted\_Sports\_fixed\_live Betted\_Sports\_fixed\_odd  
## Min. : 0.000 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 6.0   
## Median : 0.000 Median : 1.00 Median : 17.0   
## Mean : 0.276 Mean : 8.13 Mean : 29.9   
## 3rd Qu.: 0.000 3rd Qu.: 7.00 3rd Qu.: 41.0   
## Max. :111.000 Max. :240.00 Max. :342.0   
## NA's :444 NA's :444 NA's :444   
## Betted\_Supertoto Gender Language Country   
## Min. : 0.000 Min. :0.000 Min. : 1.00 Min. : 8   
## 1st Qu.: 0.000 1st Qu.:1.000 1st Qu.: 2.00 1st Qu.:276   
## Median : 0.000 Median :1.000 Median : 2.00 Median :276   
## Mean : 0.106 Mean :0.916 Mean : 3.82 Mean :359   
## 3rd Qu.: 0.000 3rd Qu.:1.000 3rd Qu.: 6.00 3rd Qu.:300   
## Max. :78.000 Max. :1.000 Max. :17.00 Max. :895   
## NA's :444 NA's :1647 NA's :1646 NA's :1646   
## country\_name country latitude longitude   
## Length:44295 Length:44295 Min. :-40.9 Min. :-106.3   
## Class :character Class :character 1st Qu.: 46.2 1st Qu.: 10.5   
## Mode :character Mode :character Median : 51.2 Median : 10.5   
## Mean : 48.7 Mean : 12.0   
## 3rd Qu.: 51.2 3rd Qu.: 10.5   
## Max. : 71.7 Max. : 174.9   
## NA's :1707 NA's :1707   
## Continent\_Name Continent\_Code application\_description  
## Length:44295 Length:44295 Length:44295   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## TotalDaysActive gender\_description LastActiveDate   
## Min. : 2.0 Length:44295 Min. :1970-01-01   
## 1st Qu.: 13.0 Class :character 1st Qu.:2015-03-08   
## Median : 28.0 Mode :character Median :2015-06-09   
## Mean : 44.8 Mean :2011-08-01   
## 3rd Qu.: 57.0 3rd Qu.:2015-09-23   
## Max. :464.0 Max. :2015-10-01   
## NA's :20281

# Demographics charts

**Users Age** As we can see from the chart below, most of the company’s customers are around 30 years old, with an clear presence of Male over Female.

x <- datamart %>%  
 filter(is.na(gender\_description)==FALSE)  
  
ggplot(x, aes(x = Age, y = UserID, fill=gender\_description)) + geom\_bar(stat = "identity") +  
 ylab(NULL)+  
 theme(panel.background = element\_rect(fill="white"))+  
 scale\_fill\_manual(values=c("#0f9fff","blue"))

## Warning: Removed 2150 rows containing missing values (position\_stack).



**Gender chart**

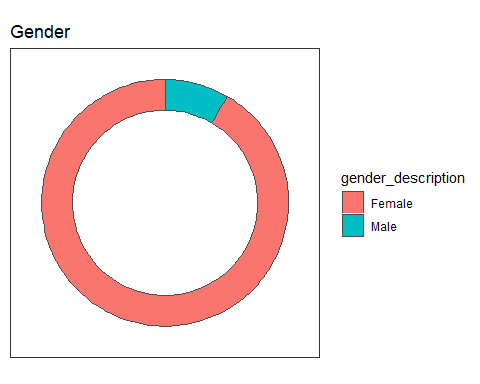
As the following graph shows, there are 39060 Male participants against 3588 Female participant, which are about 92% Male being active on Internet sports Betting.

# Gender Donut Chart  
library(readr)  
library(dplyr)  
library(ggplot2)  
datamart <- read\_csv("C:/Users/hutin/Desktop/IESEG/MASTER BIG DATA/R/datamart.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## RegistrationDate = col\_date(format = ""),  
## FOFirstActiveDate = col\_date(format = ""),  
## FOLastActiveDate = col\_date(format = ""),  
## LAFirstActiveDate = col\_date(format = ""),  
## LALastActiveDate = col\_date(format = ""),  
## FirstSportsActiveDate = col\_date(format = ""),  
## RegDate = col\_date(format = ""),  
## FirstPay = col\_date(format = ""),  
## FirstAct = col\_date(format = ""),  
## First\_time\_sports = col\_date(format = ""),  
## First\_time\_casino = col\_date(format = ""),  
## First\_time\_games = col\_date(format = ""),  
## First\_time\_poker = col\_date(format = ""),  
## country\_name = col\_character(),  
## country = col\_character(),  
## Continent\_Name = col\_character(),  
## Continent\_Code = col\_character(),  
## application\_description = col\_character(),  
## gender\_description = col\_character(),  
## LastActiveDate = col\_date(format = "")  
## )

## See spec(...) for full column specifications.

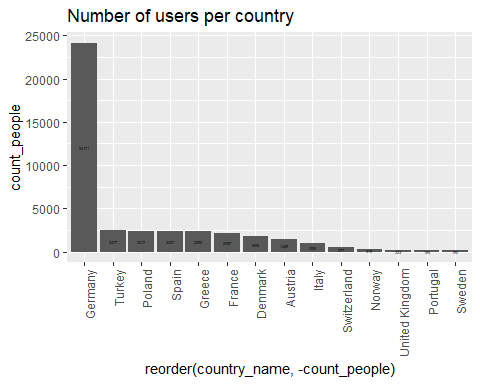
gender1 <- datamart%>%   
 filter(is.na(gender\_description)==FALSE) %>%  
 group\_by(gender\_description) %>%   
 summarise(count\_gender=n())  
  
# Add addition columns, needed for drawing with geom\_rect.  
  
gender1$fraction = gender1$count\_gender / sum(gender1$count\_gender)  
gender1 = gender1[order(gender1$fraction), ]  
gender1$ymax = cumsum(gender1$fraction)  
gender1$ymin = c(0, head(gender1$ymax, n=-1))  
  
  
ggplot(gender1, aes(fill=gender\_description, ymax=ymax, ymin=ymin, xmax=4, xmin=3)) +  
 geom\_rect(colour="grey30") +  
 coord\_polar(theta="y") +  
 xlim(c(0, 4)) +  
 theme\_bw() +  
 theme(panel.grid=element\_blank()) +  
 theme(axis.text=element\_blank()) +  
 theme(axis.ticks=element\_blank()) +  
 labs(title="Gender")



**Number of users per country**

The following graph represents the top 15 countries in terms of number of participants. Germany is by far the most represented country with 24000 users, followed by Turkey, Poland, Spain, Greece, France and Denmark.

#Country Chart  
clients <- datamart%>%   
 group\_by(country\_name) %>%   
 summarise(count\_people=n())%>%  
 arrange(desc(count\_people))%>%  
 slice(1:15)  
clients<-na.omit(clients)  
  
  
  
ggplot(clients, aes(x=reorder(country\_name, -count\_people), y=count\_people, label=count\_people)) +   
 geom\_bar(stat = "identity")+  
 scale\_fill\_manual("legend", values = "Blue")+#c("A" = "black", "B" = "orange", "C" = "blue"))+  
 geom\_text(size=1, position=position\_stack(vjust=0.5))+  
 theme(axis.text.x=element\_text(angle=90, hjust=1))+  
 labs(title= "Number of users per country")

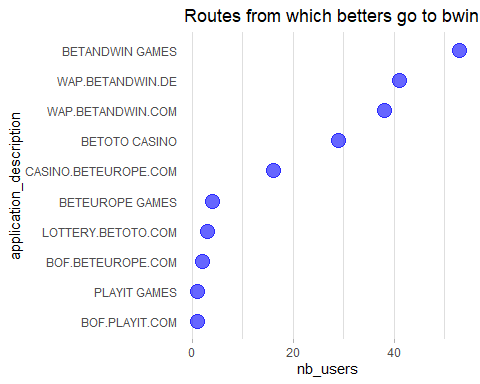


# Other key findings

**Routes from which users are coming from on the website**

The following graph shows that users mainly come from BETANDWIN GAMES to access to the bwin website.  
However, the highest average bets per customer are made from the CASINO.BETEUROPE.COM website.  
This means that the company should not only focus on its top 1 route access, but rather on the top 5 route accesses in order  
to gain and retain as many customer as possible.

applications<- datamart%>%   
 group\_by(ApplicationID) %>%   
 summarise( avg\_bets=sum(mean\_money\_bets),nb\_users=n())%>%  
 arrange(desc(nb\_users))  
applications<-na.omit(applications)  
  
  
applications <-mutate(applications,   
 application\_description=case\_when(ApplicationID==1 ~ 'BETANDWIN.COM',  
 ApplicationID==2 ~ 'TRIPLE-A-CASINO.COM',  
 ApplicationID==3 ~ 'BETANDWIN.DE',  
 ApplicationID==4 ~ 'WAP.BETANDWIN.COM',  
 ApplicationID==5 ~ 'SMS MOBILE BETTING APPLICATION',  
 ApplicationID==6 ~ 'CHARITY',  
 ApplicationID==7 ~ 'DOLCEVITACASINO.COM',  
 ApplicationID==8 ~ 'BALLS OF FIRE',  
 ApplicationID==9 ~ 'BETEUROPE.COM',  
 ApplicationID==10 ~ 'BAHSEGIR.COM',  
 ApplicationID==11 ~ 'CASINO.BETEUROPE.COM',  
 ApplicationID==12 ~ 'WWW.CASINOTURK.COM',  
 ApplicationID==13 ~ 'WWW.SANALCASINO.COM',  
 ApplicationID==14 ~ 'BETOTO.COM',  
 ApplicationID==15 ~ 'PLAYIT.COM',  
 ApplicationID==16 ~ 'CASINO.PLAYIT.COM',  
 ApplicationID==17 ~ 'THECROUPIER.COM',  
 ApplicationID==18 ~ 'SMS.BETANDWIN.COM',  
 ApplicationID==19 ~ 'WAP.BETANDWIN.DE',  
 ApplicationID==21 ~ 'BOF.PLAYIT.COM',  
 ApplicationID==22 ~ 'BOF.BETEUROPE.COM',  
 ApplicationID==23 ~ 'BETANDWIN POKER',  
 ApplicationID==24 ~ 'BETANDWIN CASINO',  
 ApplicationID==26 ~ 'SMS.BETANDWIN.DE',  
 ApplicationID==27 ~ 'LOTTERY.BETOTO.COM',  
 ApplicationID==28 ~ 'BETWORK.COM',  
 ApplicationID==29 ~ 'WAP.PLAYIT.COM',  
 ApplicationID==30 ~ 'PLAYIT POKER',  
 ApplicationID==31 ~ 'BETEUROPE POKER',  
 ApplicationID==32 ~ 'BETOTO POKER',  
 ApplicationID==33 ~ 'BETANDWIN GAMES',  
 ApplicationID==36 ~ 'BETOTO CASINO',  
 ApplicationID==38 ~ 'BETEUROPE GAMES',  
 ApplicationID==42 ~ 'PLAYIT GAMES'))  
  
  
ggplot(applications, aes(x=application\_description, y=nb\_users, label=round(nb\_users,1))) +  
 geom\_segment( aes(x=reorder(application\_description, nb\_users), xend=application\_description, y=nb\_users, yend=nb\_users), color="skyblue") +  
 geom\_point( color="blue", size=5, alpha=0.6) +  
 theme\_light() +  
 coord\_flip() +  
 theme(  
 panel.grid.major.y = element\_blank(),  
 panel.border = element\_blank(),  
 axis.ticks.y = element\_blank()  
 )+  
 labs(title= "Routes from which betters go to bwin")



**Users loyalty based on the 2 months prior to the end of the study**

The following bar chart tackles customers being active or not. As we had the information about their last connection both on  
fixed odds and live actions plays, we wanted to see who were the customers that did not play within the last two months.  
For this, we picked the last active date based on the two types of play and checked whether or not this was within the two months time frame.  
Results show that almost half of customers are considered as active players. But since we know that in the gambling world it is best to have  
users that are active weekly, or at least monthly, the company should look for a way of making sure users will play at least monthly.  
This can be an interesting path to follow for an upcoming strategy.

#Packages  
#install.packages("sas7bdat")  
library(sas7bdat)  
library(stringr)

## Warning: package 'stringr' was built under R version 3.6.2

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

library(tidyr)

## Warning: package 'tidyr' was built under R version 3.6.2

library(dplyr)  
library(jsonlite)  
library(rvest)

## Warning: package 'rvest' was built under R version 3.6.2

## Loading required package: xml2

##   
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':  
##   
## guess\_encoding

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday,  
## week, yday, year

## The following object is masked from 'package:base':  
##   
## date

library(readxl)

## Warning: package 'readxl' was built under R version 3.6.2

library(stringr)  
library(haven)  
library(ggplot2)  
  
  
# Read Dataset   
ADIG1 <- read.sas7bdat("C:/Users/hutin/Documents/GitHub/R\_Group\_Assignment/AnalyticDataInternetGambling.sas7bdat")  
#AnalyticDataInternetGambling.sas7bdat"))  
  
ADIG1$RegistrationDate<-as\_date(as.numeric(ADIG1$RegistrationDate))  
ADIG1$FOFirstActiveDate<-as\_date(as.numeric(ADIG1$FOFirstActiveDate))  
ADIG1$FOLastActiveDate<-as\_date(as.numeric(ADIG1$FOLastActiveDate))  
ADIG1$LAFirstActiveDate<-as\_date(as.numeric(ADIG1$LAFirstActiveDate))  
ADIG1$LALastActiveDate<-as\_date(as.numeric(ADIG1$LALastActiveDate))  
ADIG1$FirstSportsActiveDate<-as\_date(as.numeric(ADIG1$FirstSportsActiveDate))  
ADIG1$last\_active\_date <- pmax(ADIG1$FOLastActiveDate, ADIG1$LALastActiveDate, na.rm=TRUE)  
  
ADIG1$Active\_customer<- ifelse(ADIG1$last\_active\_date > as.Date("2015-08-01") ,"active","not active")  
loyalty<- ADIG1%>%  
 group\_by(Active\_customer)%>%  
 summarise(nb\_users=n())  
  
ggplot(loyalty, aes(x=Active\_customer, y=nb\_users, label=nb\_users, fill=Active\_customer)) +   
 geom\_bar(stat = "identity")+  
 geom\_text(size=3, position=position\_stack(vjust=0.5))+  
 labs(title= "Nb of active VS. Non active customers based on last active date")

