

## **INDEX**

- 1. ABSTRACT**
- 2. INTRODUCTION**
- 3. PROJECT AND DATAMART DESIGN**
- 4. OVERALL SUMMARIZATION**
- 5. METRICS DETAIL EXPLANATION**

## 1. ABSTRACT

Globalization and the rapid changes in technology have let new players to keep entering in markets that have had high barriers before. Companies now are more concerned on what they make their users feel instead of what they offer.

Because of this, they find themselves in a constant seek of understanding their customer, of knowing what motivates their purchasing decisions, what they feel with the product and why they decide to return. A lot of shifts in the managerial cultures are been made, this time companies are becoming more customer centric and thanks to technology, the marketing strategies and its metrics can be done in a more individual customer level.

Companies are in the pursuit of obtaining as much information as possible from their customers, such as how often they use their products (frequency), how much they invest or bet in the products (monetary), how much they win or lose or when they will stop using them (churn). The answers to these questions let companies have a more effective marketing strategy and a direct communication with their customers. Finding what metrics are relevant and which customers they should focus more their attention to.

Bwin.com is one of these companies, that are investing a lot of efforts in understanding their customers and surely fits in the description mentioned above.

**Key words: Metrics, Customer, Marketing, Bwin.**

## 2. INTRODUCTION

Bwin.com is an online gambling site, considered among the 10 largest companies in the world for offering a wide variety of products and having vast amount of transactions made by its users.

In order to get as much information as possible from its customers and secure a one to one relationship by understanding their behavior on the website, what products they prefer and how often they used them. We will answer questions such as what is the most attractive program, what is the most common nationality or language; are women as engaged as men and what do they like? Insights on gender or language are just a small example of how precise we could understand our market.

Thereby, it is possible to also make estimates on the value of each customer, knowing how much he or she provides to the big basket and even predict when this particular customer would churn. Being able to distinguish each customer, understanding that everyone is different and has different behaviors, knowing how much he or she is worth, and when will leave, allows Bwin to create more efficient loyalty programs and retention strategies by sending every time more personalized messages that would increase customer engagement to the site and reduce churn.

In this project, after having the user data from an eight-month period, we will try to come up with certain metrics that could help Bwin.com or any other online gambling website maintain a one to one relationship with its customers and understand how valuable each of them is. By analyzing and using these metrics, any company would be able to create less costly strategies to have long lasting relationships with their customers.

### 3. PROJECT AND DATAMART DESIGN

Our project was conducted to develop certain metrics for the Bwin Enterprise and its online gambling website, that could help us better understand their customers and get some insights from their behavior and their gambling preferences.

For this assignment, we were given three different sources (which will be explained later on) from which we could gather as much information as possible. The main purpose was to clean the data from each table and organize it in a fully comprehensive and easy to understand datamart.

The main primary files, mentioned above, were raw data from which we started were:

- **Raw Data Set for Demographics.** Contains information on demographic in a customer level such as the Country of Residence, Language, Gender, Registration Date and when they became active in each of the offered online betting products.
- **Raw Data Set for Daily Activity.** Contains the actual betting information associated with each product such as Sports Betting, Casino, Poker or Online Games for each participant for each calendar day with at least one transaction from February 1, 2005 through September 30, 2005
- **Raw Data Set for Poker Chip Transactions.** Contains the information of each poker chip transaction from February 1, 2005 through September 30, 2005 handed in by the Division of Addictions (DOA). These summaries were obtained for each poker play in terms of poker chip transactions to
- yand from the poker site for each poker session.

Each raw data set contained specific information on the users that could be compiled by the user ID as the primary key. In order to provide a clean datamart and develop metrics that would help obtain valuable insights for each user, we proceeded to do the step by step explained in the following chapter.

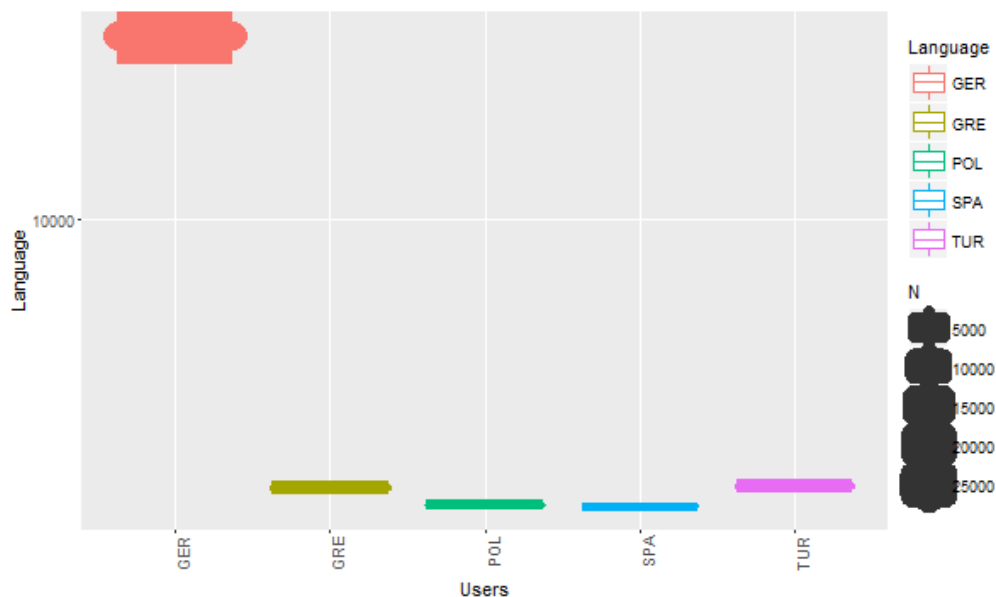
#### 4. OVERALL SUMMARISATION

After preparing a comprehensive and organized datamart, we decided to revise it on a customer level and furthermore on an aggregated level.

Some of the insights we visualized were the following:

##### 4.1. Languages most used by users

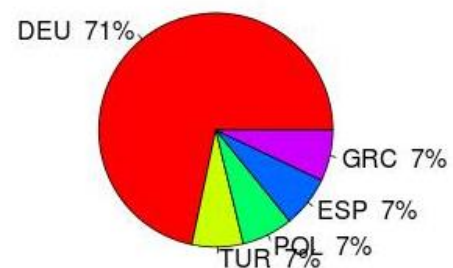
This variable will let Bwin.com overview from the offered languages, which are the ones the users prefer. Understanding this will let them improve their website according to language preferences and make sure their audience will get the message they are sending.



##### 4.2. Countries with most users

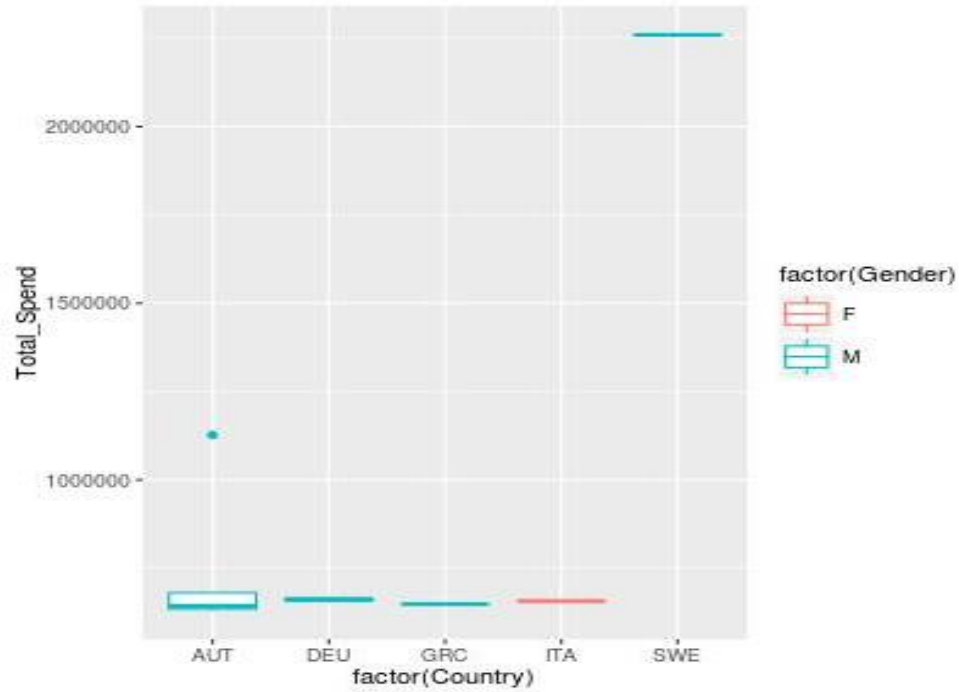
Knowing this will let Bwin.com focus their strategies within the countries that have most users. It also helps to see if it's possible to invest in the countries where their presence is less.

**Percentage of Customers by language spoken**

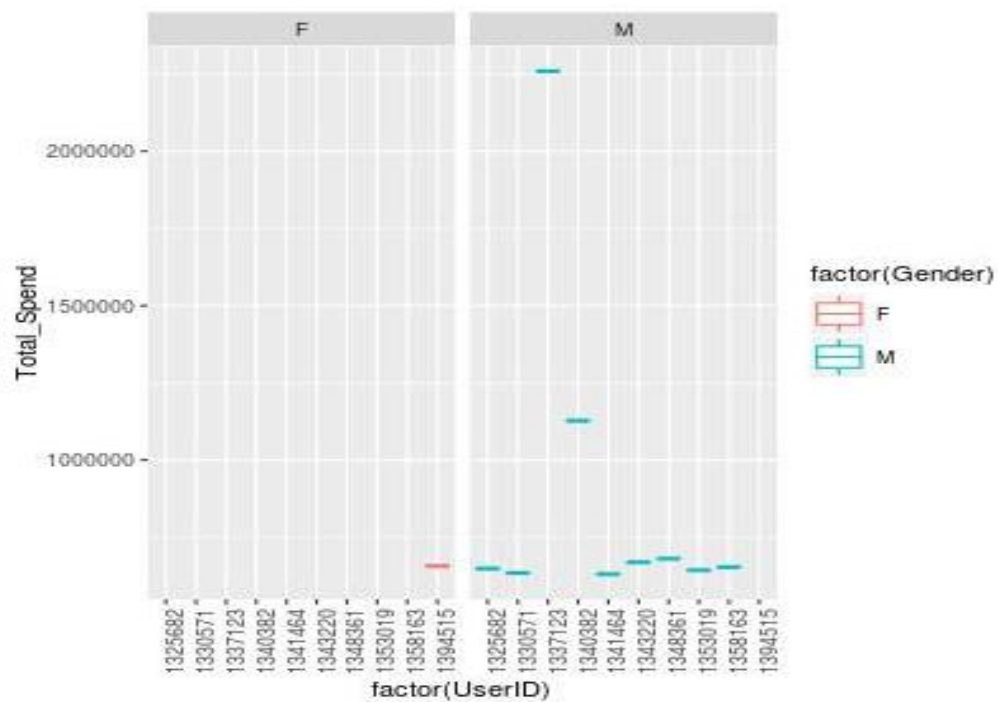


### 4.3. Top Countries according to gender

By this we can aim marketing strategies according to gender in the countries that have the most participation.



### 4.4. Users that bring most revenues to the company according to gender



## 5. METRICS DETAIL EXPLANATION

Variables on “mkt\_datamart.csv”:

- Period: 1<sup>st</sup> February 2005 – 30<sup>th</sup> September 2005
- Product catalogues:
  - **SP** - Sports book fixed-odd, Sports book live-action, Supertoto
  - **PO** - Poker BossMedia
  - **CA** - Casino BossMedia, Casino Chartwell
  - **GA** - Games VS, Games bwin
- Raw databases:
  - **Demographics** - RawDataIDemographics.sas7bdat
  - **UserDaily** - RawDataIIUserDailyAggregation.sas7bdat
  - **PokerChip** - RawDataIIIPokerChipConversions.sas7bdat

Variable / Metric	Data Type / Summary	Calculation	Meaning
<b>Group 1 – Gamblers Information &amp; Behaviors</b>			
<b>UserID</b> (User ID)	Numeric	Only selected UserIDs which have First Active Date (First_Act) different from #NA and in the chosen period.	The unique ID for each gambler registered at bwin.com.
<b>Country</b> (Country)	Character	Converted from country code to UN-3 Letter Abbreviation Code.	Country of gambler. e.g. FRA (France), DEU (Germany), BEL (Belgium), etc.
<b>Language</b> (Language)	Character	Converted from language code to ISO 639-2 Codes.	Language used by gambler. e.g. GER (Germany), FRE (French), ENG (English), etc.
<b>Gender</b> (Gender)	Character (M, F)	Converted from gender code (0 and 1) to character.	Gender of gambler. i.e. 0 = F (Female) and 1 = M (Male).
<b>ApplicationID</b> (Application ID)	Numeric	Import from Demographics raw data.	The code of websites (sources) that redirected gamblers to bwin.com to make the registration.
<b>Reg_Date</b> (Registration Date)	Date Min: 2005-02-01 Max: 2005-02-27	Import from Demographics raw data.	Gambler registration date.
<b>First_Pay</b> (First Payment Date)	Date Min: 2005-02-01 Max: 2005-10-02	Import and standardize from Demographics raw data.	Gambler first betting money deposits date.
<b>First_Act</b> (First Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Only selected First Active Date (First_Act) different from #NA and in the chosen period.	Gambler first active (play) date.

<b>First_SP</b> (First Sport Book Active Date)	Date Min: 2005-02-01 Max: 2005-10-02	Import and standardize from Demographics raw data.	Gambler first play date in Sport book product catalogue. Some UserIDs have no recorded values (NA).
<b>First_CA</b> (First Casino Active Date)	Date Min: 2005-02-01 Max: 2005-10-02	Import and standardize from Demographics raw data.	Gambler first play date in Casino product catalogue. Some UserIDs have no recorded values (NA).
<b>First_GA</b> (First Game Active Date)	Date Min: 2005-02-01 Max: 2005-10-02	Import and standardize from Demographics raw data.	Gambler first play date in Game product catalogue. Some UserIDs have no recorded values (NA).
<b>First_PO</b> (First Poker Active Date)	Date Min: 2005-02-01 Max: 2005-10-02	Import and standardize from Demographics raw data.	Gambler first play date in Poker product catalogue. Some UserIDs have no recorded values (NA).
<b>Last_Act</b> (Last Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Calculate from the maximum of Last Active (play) Date of the 4 product catalogues, i.e. SP, CA, GA and PO.	Gambler last active (play) date.
<b>Last_SP</b> (Last Sport Book Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Take the last activity of SP product catalogues from UserDaily raw data.	Gambler last play date in Sport book product catalogue. Some UserIDs have no recorded values (NA).
<b>Last_CA</b> (Last Casino Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Take the last activity of CA product catalogues from UserDaily raw data.	Gambler last play date in Casino product catalogue. Some UserIDs have no recorded values (NA).
<b>Last_GA</b> (Last Game Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Take the last activity of GA product catalogues from UserDaily raw data.	Gambler last play date in Game product catalogue. Some UserIDs have no recorded values (NA).
<b>Last_PO</b> (Last Poker Active Date)	Date Min: 2005-02-01 Max: 2005-09-30	Take the last activity of PO product catalogues from PokerChip raw data.	Gambler last play date in Poker product catalogue. Some UserIDs have no recorded values (NA).
<b>Gap_Reg_Act</b> (Gap of Registration and Active Date)	Numeric	Calculate by First Active Date (First_Act) minus Registration Date (Reg_Date).	The number of delaying days between registration date and first active date.
<b>Gap_Reg_Pay</b> (Gap of Registration and Payment Date)	Numeric	Calculate by First Payment Date (First_Pay) minus Registration Date (Reg_Date).	The number of delaying days between registration date and first payment date.



<b>Num_Product</b> (Number of Playing Products)	Numeric Min: 1 Max: 4 Mean: 1.2	Calculate by the availability of First Active Day of 4 product catalogues, i.e. First_SP, First_CA, First_GA and First_PO.	Number of products was playing by a gambler.
<b>Max_Bet</b> (Maximum Betting)	Numeric Min: 0 Max: 27350.33 Mean: 53.27	Calculate by taking the maximum amount of Stakes per UserID from UserDaily raw data.	Maximum betting amount of a gambler to 3 product catalogues SP, CA and GA.
<b>Min_Bet</b> (Minimum Betting)	Numeric Min: 0 Max: 3959.000 Mean: 6.220	Calculate by taking the minimum amount of Stakes per UserID from UserDaily raw data.	Minimum betting amount of a gambler to 3 product catalogues SP, CA and GA.
<b>Avg_Bet</b> (Average Betting)	Numeric Min: 0 Max: 3959.000 Mean: 15.884	Calculate by taking the average amount of Stakes per UserID from UserDaily raw data.	Average betting amount of a gambler to 3 product catalogues SP, CA and GA.
<b>Group 2 – Gamblers RFM value in general</b>			
<b>Total_Earning</b> (Total Earning)	Numeric Min: -20021.2 Max: 1093403.0	Total_Earning_SP + Total_Earning_CA + Total_Earning_GA + Total_Earning_PO	Total earning or losing (negative value) of a gambler on 4 product catalogues.
<b>Total_Spend</b> (Total Spending)	Numeric Min: 0 Max: 2259812.0	Stakes_SP + Stakes_CA + Stakes_GA + Buy_PO	Total spending of a gambler on 4 product catalogues.
<b>Total_Active</b> (Total Active Days)	Numeric Min: 1.00 Max: 235.00	Merge UserDaily and PokerChip by UserID and Date. Put value 1 per activity then aggregate by UserID.	Total active days of gambler on 4 product catalogues.
<b>Winning_Rate</b> (Winning Rate)	Numeric Min: 0 Max: 100.00	$100 * (\text{Win\_SP} + \text{Win\_CA} + \text{Win\_GA} + \text{Win\_PO}) / (\text{Win\_SP} + \text{Win\_CA} + \text{Win\_GA} + \text{Win\_PO} + \text{Loss\_SP} + \text{Loss\_CA} + \text{Loss\_GA} + \text{Loss\_PO})$	Overall winning rate of a gambler on 4 product catalogues.
<b>Avg_Earning_Day</b> (Average Earning Per Day)	Numeric Min: -1000.00 Max: 63100.00	Total_Earning / $(\text{Win\_SP} + \text{Win\_CA} + \text{Win\_GA} + \text{Win\_PO} + \text{Loss\_SP} + \text{Loss\_CA} + \text{Loss\_GA} + \text{Loss\_PO})$	Average earning or losing (negative value) per day of a gambler on 4 product catalogues.
<b>Recency</b> (Customer Recency)	Numeric Min: 3.000 Max: 10.000	10 – (September – Last Active Month)  * September: the last month in the period	The overall recency scores of a gambler. From 0 (least recently visited) to 10 (most recently visited) on 4 product categories.

		** Last Active Month: calculate from Last_Act metrics	
<b>Frequency</b> (Customer Frequency)	Numeric Min: 0.04255 Max: 10.00000	10 * Total_Active / MaxActive  * Total_Active: calculated above  ** MaxActive: take the maximum active days value, give it 10 point	The overall frequency scores of gambler. From 0 (least frequently visited) to 10 (most frequently visited) on 4 product categories.
<b>Monetary</b> (Customer Monetary Spending)	Numeric Min: 0 Max: 10.000000	10 * Total_Spend / MaxSpend  * Total_Spend: calculated above  ** MaxSpend: take the maximum spending value, give it 10 point	The overall monetary scores of a gambler. From 0 (least spending) to 10 (most spending) on 4 product categories.
<b>RFM</b> (Customer RFM)	Numeric Min: 3.043 Max: 27.191	Recency + Frequency + Monetary	The overall RFM (Recency- Frequency-Monetary) scores of a gambler. From 0 (least valuable) to 30 (most valuable) on 4 product categories.
<b>Rank</b> (Customer RFM Rank)	Numeric Min: 1 Max: 42616	Sort data by RFM by descending then put value from 1 to .N to Rank column.	The rank of gambler by his RFM score. From 1 (highest ranking) to 42616 (lowest ranking).
<b>Group 3, 4, 5 – Gamblers RFM per betting product catalogue (XX = SP or CA or GA)</b>			
<b>Stakes_XX</b> (Stakes XX)	Numeric	Aggregate daily Stakes per product catalogue from UserDaily raw data.	Total Stakes spending (in EUR) by gambler on product catalogue XX.
<b>Winnings_XX</b> (Winning XX)	Numeric	Aggregate daily Winnings per product catalogue from UserDaily raw data.	Total Winnings amount (in EUR) of a gambler on product catalogue XX.
<b>Bets_XX</b> (Betting Time XX)	Numeric	Aggregate daily Bets per product catalogue from UserDaily raw data.	Total betting times of a gambler on product catalogue XX.
<b>Win_XX</b> (Winning Times XX)	Numeric	If daily Winnings > 0, then Win = 1, Loss = 0. Then aggregate per product catalogue.	Total winning times of a gambler on product catalogue XX.

<b>Loss_XX</b> (Losing Times XX)	Numeric	If daily Winnings = 0, then Win = 0, Loss = 1. Then aggregate per product catalogue.	Total losing times of a gambler on product catalogue XX.
<b>Total_Earning_XX</b> (Total Earning XX)	Numeric	Calculate by daily Winnings minus daily Stakes. Then aggregate per product catalogue.	Total earning or losing (negative value) of a gambler on product catalogue XX.
<b>Total_Active_XX</b> (Total Active Days XX)	Numeric	Put value 1 for each activity in UserDaily raw data. Then transpose and aggregate per product catalogue.	Total active days of gambler on product catalogue XX.
<b>Winning_Rate_XX</b> (Total Winning Rate XX)	Numeric	$\text{Win\_XX} / (\text{Win\_XX} + \text{Loss\_XX}) * 100$	Winning rate of a gambler on product catalogue XX.
<b>Avg_Earning_XX</b> (Average Earning XX)	Numeric	$\text{Total\_Earning\_XX} / (\text{Win\_XX} + \text{Loss\_XX})$	Average earning or losing (negative value) per day of a gambler on product catalogue XX.
<b>Recency_XX</b> (Customer Recency XX)	Numeric (Score: 0-10)	10 – (September – Last Active Month)  * September: the last month in the period  ** Last Active Month: calculate from Last_XX metrics	The recency scores of a gambler. From 0 (least recently visited) to 10 (most recently visited) on product catalogue XX.
<b>Frequency_XX</b> (Customer Frequency XX)	Numeric (Score: 0-10)	$10 * \text{Total\_Active\_XX} / \text{MaxActive\_XX}$  * Total_Active_XX: calculated above  ** MaxActive_XX: take the maximum active days value, give it 10 point	The frequency scores of gambler. From 0 (least frequently visited) to 10 (most frequently visited) on product catalogue XX.
<b>Monetary_XX</b> (Customer Monetary XX)	Numeric (Score: 0-10)	$10 * \text{Stakes\_XX} / \text{MaxSpend\_XX}$  * Stakes_XX: calculated above	The monetary scores of a gambler. From 0 (least spending) to 10 (most spending) on product catalogue XX.

		** MaxSpend_XX: take the maximum Stakes value, give it 10 point	
<b>RFM_XX</b> (Customer RFM XX)	Numeric (Score: 0-30)	Recency_XX + Frequency_XX + Monetary_XX	The RFM (Recency-Frequency-Monetary) scores of a gambler. From 0 (least valuable) to 30 (most valuable) on product catalogue XX.
<b>Rank_XX</b> (Customer Rank XX)	Numeric	Sort data by RFM_XX by descending then put value from 1 to .N to Rank_XX column.	The rank of gambler by his RFM score on product catalogue XX. From 1 (highest ranking) to 42616 (lowest ranking).
<b>Group 6 – Gamblers RFM of PO product catalogue</b>			
<b>Buy_PO</b> (Buy Poker Chip)	Numeric Min: 0 Max: 2259812.0 Mean: 213.1	Transpose and aggregate daily buying poker chip from PokerChip raw data.	Total amount of buying poker chip (in EUR) by gambler.
<b>Sell_PO</b> (Sell Poker Chip)	Numeric Min: 0 Max: 2246245.7 Mean: 228.6	Transpose and aggregate daily selling poker chip from PokerChip raw data.	Total amount of selling poker chip (in EUR) by gambler.
<b>Win_PO</b> (Winning Times Poker)	Numeric Min: 0 Max: 137.0000 Mean: 0.8267	If Sell_PO > 0, then Win_PO = 1 and Loss_PO = 0	Total winning times of a gambler on Poker.
<b>Loss_PO</b> (Losing Times Poker)	Numeric Min: 0 Max: 109.0000 Mean: 0.4952	If Sell_PO ≤ 0, then Win_PO = 0 and Loss_PO = 1	Total losing times of a gambler on Poker.
<b>Total_Earning_PO</b> (Total Earning Poker)	Numeric Min: -33306.77 Max: 39100.18 Mean: 15.45	Sell_PO – Buy_PO	Total earning or losing (negative value) of a gambler on Poker.
<b>Total_Active_PO</b> (Total Active Days Poker)	Numeric Min: 0 Max: 215.000 Mean: 1.322	Transpose and aggregate PokerChip raw data to daily activity. Then put value 1 for each daily activity and aggregate once again.	Total active days of gambler on Poker.
<b>Winning_Rate_PO</b> (Total Winning Rate Poker)	Numeric Min: 0 Max: 100.000 Mean: 3.594	$\text{Win\_PO} / (\text{Win\_PO} + \text{Loss\_PO}) * 100$	Winning rate of a gambler on Poker.

<b>Avg_Earning_PO</b> (Average Earning Poker)	Numeric Min: -2174.050 Max: 878.176 Mean: 1.061	Total_Earning_PO / (Win_PO + Loss_PO)	Average earning or losing (negative value) per day of a gambler on Poker.
<b>Recency_PO</b> (Customer Recency PO)	Numeric Min: 0 Max: 10.0000 Mean: 0.3868	10 – (September – Last Active Month)  * September: the last month in the period  ** Last Active Month: calculate from Last_PO metrics	The recency scores of a gambler. From 0 (least recently visited) to 10 (most recently visited) on Poker product.
<b>Frequency_PO</b> (Customer Frequency Poker)	Numeric Min: 0 Max: 10.00000 Mean: 0.06148	10 * Total_Active_PO / MaxActive_PO  * Total_Active_PO: calculated above  ** MaxActive_PO: take the maximum active days value, give it 10 point	The frequency scores of gambler. From 0 (least frequently visited) to 10 (most frequently visited) on Poker product.
<b>Monetary_PO</b> (Customer Monetary Poker)	Numeric Min: 0 Max: 1.00e+01 Mean: 9.43e-04	10 * Buy_PO / MaxSpend_PO  * Buy_PO: calculated above  ** MaxSpend_PO: take the maximum Buy_PO value, give it 10 point	The monetary scores of a gambler. From 0 (least spending) to 10 (most spending) on Poker product.
<b>RFM_PO</b> (Customer RFM Poker)	Numeric Min: 0 Max: 27.8605 Mean: 0.4492	Recency_PO + Frequency_PO + Monetary_PO	The RFM (Recency- Frequency-Monetary) scores of a gambler. From 0 (least valuable) to 30 (most valuable) on Poker product.
<b>Rank_PO</b> (Customer Rank Poker)	Numeric Min: 1 Max: 2369.00	Sort data by RFM_PO by descending then put value from 1 to .N to Rank_PO column.	The rank of gambler by his RFM score on Poker product. From 1 (highest ranking) to 2369 (lowest ranking).