MARKETING DATAMART - Manual

Group 6 : NITESH SHANBHAG, JUSTINE DELEWAL & LIU YI

Contents

0.1 Approach

We started off with reading all the four SAS Input files(Demographics, UserDaily Aggregation, Poker Chip Conversions and Analytical Dataset), and cleaning them. All the date columns were converted into R `Date` objects to allow calculations on dates. From the analytic dataset, only the `Age` column was taken since this information is not present in Demographics dataset. The First Step was to clean individual datasets and Convert each Table to 1 row per user Id to get the complete information about each User in the Company. Here All the transactions before Feb1 2005 and after September 30 2005 were deleted and Only the users with registration date between 1 and 27 february, and transactions done after `FirstPay` date were retained. There were Total of 4 individual Tables created

1. Sports(Which consits of Live Action and Fixed Odds)
2. Poker(contains Poker Boss Media)
3. Games(Contains SuperToto, GameVS and GamesBwin)
4. Casino(contains Casino Boss Media and Casino Chartwell)

Basetable for 4 Individual Tables have also been created suppose if the Marketing teams want to focus individually on each Product with deeper insights on each Product. For the Final Basetable, to give an overall picture of each User Id, each Individual Products were merged with each other on User ID and Demographic Info of the User giving rise to a Final Basetable with One row Per User ID with Global Marketing Metrics created for each User. Some Important Product-wise metrics were also retained.

0.2 Structure

The Basetable consists of the following main columns :

- **UserID** - *[From Demographics table]* Unique identifier of the user assigned at the time of registration.
- **Gender** - *[From Demographics table]* User's Gender.
- **Country** - *[From Demographics table]* User's country of residence.
- **Continents** - _Calculated from Country of the User.
- **Language** - *[From Demographics table]* Language of the User
- **Application** - *[From Demographics table]* Application of the User as mentioned in Application table
- **Age** - *[Analytic Dataset]* User's Age
- **Age Group** - *[Calculated]* Age was made into 4 groups :'18-35 Years old','35-55 Years old','55-75 Years,'75+ Years old','Not Known'
- **GGR(Gross Gaming Revenue)** - *[Calculated]* Calculated as the Total Profits to the Company. User Stakes Minus the Total Winnings of the User from all Products
- **Ranked Segment** - *[Calculated]* Profitability rank of the segment to which the user belongs. See *Metrics* section for details.
- **Profit Margin** - *[Calculated]* Total profit margin from the user for bwin.
- **Lifetime Value (Indicative)** - *[Calculated]* Descriptive approximate lifetime value of the user for bwin over his active play period.
- **Total Bets Placed** - *[Calculated]* The Total Bets Placed by the User (excluding Poker)
- **First&Last_Active_Date** - *[Calculated]* The Very First & Last Transaction date of the User. Note that it is after First Payment Date and this implies to all Dates as Before Payment Dates Transactions have been deleted.
- **Length of Relationship(In Days)** - *[Calculated]* Last_Day - First_Day, which represents the total number of days the user has been with the company in that time frame.
- **Loyal** - *[Calculated]* Loyalty was calculated as the Last_Day(Sep30) minus the Last Transaction Date of the User. It would give an hint on whether the user is still with the company and loyal to the company. These were also Divided into High / Medium and Low Loyalty Class.
- **ARPU(Average Revenue Per User Per Login)** - *[Calculated]* Total Profit Per User / Frequency
- **Frequency** - *[Calculated]* Total number of Logins. It was also divided into High, Medium and Low Classes to indicate how frequent the Customer is. It is used as one of the measure to calculate the Market Segments as explained below.
- **Observed Stakes** - *[Calculated]* Based on the Total Stakes of the User, Each user has been distinguished into 3 Classes: High Stakes which indicates User's Monetary Value, Medium Stakes and Low Stakes. Stakes has been taken into consideration because Winnings depends on the User's game and does not give an indication about the User's monetary value.
- **Total Stakes/Winnings** - *[Calculated]* Total Stakes and Total Winnings from all 4 Products has been calculated
- **Individual Profits Each Product** - *[Calculated]* Indivdual Profits by each Product from the User.
- **No of Days Played for each Product** - *[Calculated]* Number of Days Played by each Product has been calculated
- **First & Last Transaction Date of Each Product** - *[Calculated]* First & Last Transaction Date of Each product has been calculated to give an idea on which Product the User started to bet.

- **Length Play of Each Product** - *[Calculated]* Length of Play of Each Product has been calculated
- **Most Played Month of each Product** - *[Calculated]* Most Played Month of Each User per Product has been calculated to see which month was prefered on overall by the Users to target the month of coming marketing campaigns
- **Most Played Day of the Week for each Product** - *[Calculated]* Most Played Day of the week for each User was also calculated to give a better picture on which Day of the Week to target the upcoming Marketing campaigns.
- **Amount Lost Last Day for each Product** - *[Calculated]* Amount Lost by the User on the last day of his/her play has also been calculated per Product to indicate the Marketers whether he / she may churn if the amount lost is huge but this was not included in Final basetable but available in individual Product Basetables

**Note - If a user has not played a particular product, it means there is no data for that user on that particular product and all the corresponding columns will have 0/NA in such cases.*

0.3 Important Metrics Description

Two types of marketing metrics have been provided in the Datamart : *Global* (per user across all products) and *Product* (per user per product). This section describes their meaning and how they have been calculated in the order they appear in the Datamart :

1. **GGR** - GGR stands for Gross Gaming Revenue in the gaming and online casino world. It is calculated as the difference between the net profit (the amount players wager minus the amount they win) and the tax an establishment pays in regards to gaming supplies.

Here tax was not taken into Consideration hence (GGR = Total Stakes - Total Winnings)

2. **Ranked segment** - User segments have been created by a combination of Frequency and Total Stakes and assigned to each user. The segments were ranked as per logic explained below and the ranks were directly put in the Basetable. Grouping by profitable segments would allow the marketing to target each group of users in a customized way. The Stakes were used as an important metrics to calculate the Segment because we Can judge the Players Monetary Worth based on his Stakes and not by his Profit.This is why we wanted to segment the Players based on their Monetary Value and Frequency of visit to the Gambling site. The user segment matrix created is as follows :

Frequency	Stakes(euros)	Rank
>= 150	>= 10000	1
>= 75 & < 150	>= 10000	2
<75	>= 10000	3
>= 150	>= 1000 & < 10000	4
>= 75 & < 150	>= 1000 & < 10000	5
<75	>= 1000 & < 10000	6
Irrespective	< 1000	7

Which can also be translated as below:

Frequency	Stakes(euros)	Rank
High	High	1
Medium	High	2
Low	High	3
High	Medium	4
Medium	Medium	5
Low	Medium	6
Irrespective	Low	7

3. **Overall profit margin** - This is the most direct indicator of how profitable a user has been so far for bwin. Postive margin indicates bwin is making money from the user whereas negative means bwin is losing money to the user. Lower margin indicates the cost to serve that user is high. The marketing can focus only on the most profitable users, and in conjunction with the other metrics, try to encourage users with low margin to play more. This has been calculated as:

```
[(Overall Stakes - Overall Winnings) / Overall Stakes] * 100
```

*_Note - The Promotional Money has not been considered to calculate any value and is a direct reflection to User's own money.

4. **Lifetime value (Indicative)** - This is a descriptive metric calculated using historical data on the user's activity on bwin and indicates the approximate cash flow bwin can expect from the user over each period of similar duration for which the user was active on bwin if he/she were to continue playing. Marketing can project this cash flow into future, diminishing it over time based on some parameters, and

apply the weighted average cost of capital and the retention probability (out of scope of the Datamart) to get the NPV of the user over the desired projected period. Only the users with postive lifetime are valuable and need to be focused on. Bwin can also try to devise strategy to extract additional value from lower valued user, based on the analysis on whether they are more loyal or cost less to serve. The metric has been calculated as follows:

```
[(Overall Stakes - Overall Winnings) / TotalActDays_Overall]
*
[LastActDate_Overall - FirstActDate_Overall + 1]
```

5. **Total active days overall** - This gives a general idea about the duration for which different users have been active on bwin over the period of study. This has been calculated as the count of unique dates in the Poker transaction and User aggregation tables combined.

6. **Overall playing frequency** - This metric indicates how much the user has actually been playing and winning or losing money. Higher the ratio, the more addicted the user is to online gambling and has better prospect of being a loyal customer. This has been calculated as follows:

   ```
   TotalActDays_overall / (LastActDate_Overall - FirstActDate_Overall + 1)
   ```

7. **Loyal** - This representsthe Loyalty of the User to Bwin. If he has been active until the end of the Last day of the Analysis then he / she is considered to be loyal.

If No_of_Days <= 61 then High Loyalty Between 61 and 153 then Medium Loyalty If > 153 Then Low Loyalty

Please note :This has not taken into consideration his start Date / Frequency as we wanted his loyalty as of the end of the Sep 30th.

8. **ARPU** - This represent Average Revenue Per User Per Login : This is to indicate how much does each user contribute to the company's Profits per login.

Calculated as Total Profit Per User / Frequency

9. **Overall stakes** - This represents the total revenue from each user during his/her entire duration of activity, and is calculated as:

   ```
   Sum(Total Stakes on all P(i))
   ```

10. **Overall winnings** - This repesents the total cost for bwin of maintaining each user, and is calculated as:

    ```
    Sum(Total Winnings on all P(i))
    ```

**Note - The overall cost of a user also includes administrative and logistics cost but that data is not available.*

0.4 Basetable :

Below is the View of Basetable with few important Metrics

UserID	GGR	Profit_Margin	LifetimeValue	Total_BetsPlaced	Total_Length_Relationship
1329937	3155.64	9.7015384	15417.55445	12380	170
1399049	70.00	3.5897436	207.30769	292	76
1387163	120.00	27.2294078	1314.00000	35	218
1377820	-2.60	-0.2670776	-4.57600	57	43
1384050	24.47	13.7697517	86.03968	134	217
1386419	28.00	32.8253224	234.76923	33	217

0.5 Explanation for Graphical Representation in Shiny:

Please note that For all the below Graphs Filters have been kept with Applications, Continent and Gender for Management team to perform targeted marketin approach.

1. **Geographical Presence of Customomer** : This is a Geographical Representation of Presence of all the Bwin Customers on the World Map. This would help the Marketing Managers to target the locations for their next Marketing campaigns and would also give a view on starting their Marketing campaigns targeting newer countries / Areas which would have profitable business in the days ahead.
2. **Gambling Profiles** : This gives a graphical representation of 4 Different Profiles of Customers.

- Gamblers by Age - Histogram of Number of Gamblers by Age is plotted so as to provide a fair idea about which is the Age group that should be targeted more.(The Age Group 25-30 has the highest number of Gamblers)
- Gamblers by Product Category : This is a Plot to show number of players in each Product Category. This would give an idea about the popularity of each individual Business. We can see that Sports is played by most players followed by Casino and Games with Poker being played the least.
- Gamblers per Loyalty Class : Number of Players per Loyalty Class has been plotted. We can see that there are more people who are Highly Loyal followed by Low and Medium.
- Gamblers Per Language : Number of Gamblers per Language is plotted and we can see that we have most German Speakers. And also interesting thing to note here is if we exclude Europe we can find most spoken language to be Turkish. So Marketing Managers can use Language as a tool for their upcoming campaigns.

3. **Segmentation of Customers** : Here the number of Customers per Segment and Profit from each Segment is Plotted. It's an interesting result because there are way too less number of Customers from Segments 1,2,3. To be Precise (408 Customers in 1st Segment, 732 in 2nd Segment, 831 in 3rd and 32907 in 7 th Segment). But the Profits from 1st Segment is 1376012.73 Euros and similar figures in 2 and 3 with 1438677.6 Euros in 7th segment. It considerably helps Marketers to individually target each Segments.
4. **Sports / Poker / Casino / Games** : For each Individual Products the most Preferred Day of the Week and also the Most Preferred Month of the Year.

- Sports : So we can see Saturday has been preferred by 42% of the total Sports Players and also February and March being the preferred month.
- Poker : There was not an individual Day Preferred but Feb was the Preferred month.
- Casino : 26% of the Total Casino Players Preferred Friday and again February was the Preferred month.
- Games : 22% of the Total Games Players Preferred Friday and 20% of the Total Games Players Preferred Monday. Here February was the Preferred month, followed by March.

5. **Profits** : For each Individual Products the Profits to the company was displayed.Also Profits from each Loyalty Class was Displayed.

- We can see Fixed Odds has the highest Profits Followed by Live Action and Casino. Interesting thing to note is that All the Products contributed to Profits except Poker and there were huge loss from Poker even though the Number of Players Playing Poker were minimal.
- We can see that High Loyal Customers accounted to the Highest Profits followed by Low and Medium.

6. **Summary Statistics** :

The Statistics for some important Global Metrics have been calculated as displayed below.

Descriptive Statistic of few Global Metrics

Sum_stats	GGR	LifetimeValue	Total_BetsPlaced	ARPU	TotalStakes	TotalWinnings
Mean	172.80	590.51	359.31	8.03	2951.11	2778.25
Stdev	1124.95	3463.00	1981.75	43.86	22003.41	21460.80
Median	43.99	140.33	58.00	2.73	237.20	182.00
Minimum	-38504.47	-315250.00	0.00	-3250.00	0.00	0.00
Maximum	76165.51	146441.55	193442.00	1581.50	2259812.03	2246245.71
nobs	42564.00	42564.00	42564.00	42564.00	42564.00	42564.00

**Note - We also tried plotting relationships between various metrics that we have in the basetable but we could not find a valid relationship between any of these variables, which would mean our variables are not correlated to each other [except for individuals as discussed above]. For example, frequency and profits were not showing relationship, which could be infered as irrespective of the number of times a user plays a game, the profits is still random. Same applies for other metrics such as loyalty or higher stakes.*