**Open Source Programming Group Project**

***Group 11***

**Stephanie Beyer (**stephanie.beyer@ieseg.fr**)**

**ShabeNoor Kamal (**shabenoor.kamal@ieseg.fr**)**

**Omar Abdelgelil (**[omar.abdelgelil@ieseg.fr](mailto:omar.abdelgelil@ieseg.fr)**)**

**Prof. Dr. Matthijs Meire**

**Business Case**

This case is prepared for an online gambling website, with international users, who were offered a wide array of gambling services. The company has collected a wealth of data from its databases but now aims to convert this into knowledge.

The company’s requirements for this project are to get some basic marketing descriptions for their customer base and to create a marketing data mart with aggregated information per customer.

To aid this process, we were provided with multiple data tables that contain detailed data on clients including demographics, user daily aggregation and poker chip conversion. The time range of the data is from February 1, 2005 up Until September 30, 2005. Our goals in assisting this institution were multi-fold:

1. Our initial aim was to clean, merge and aggregate the different data tables provided by the company in a way that it is easy to get meaningful data on each individual client in their database.
2. Second, to create some additional marketing summary statistics to allow deeper insight into each client.
3. Third, to create a data manual for marketing analysts, that contains the overall summary statistics.
4. Lastly, to create a Shiny application, which provides interactive features to investigate the marketing numbers.

**Variables and Datamart Creation**

Initially we had four separate tables. A list of the tables is mentioned below. All of the tables were joined on User ID:

* Demographics – each record describes characteristics of a client.
* User Daily Aggregation table – contains betting information related to each product
* Poker Chip Conversions – contains poker chip transaction information
* Analytics Dataset - contains actual betting behavior data for both fixed-odds and live-action sports book betting for each participant aggregated

Additional variables were created to get further insights from the data. Initial variables were created in the individual tables before merging them. Dates were converted and missing values were identified and excluded from charts.

Variables were created within the existing tables to aggregate them in a way that would allow us to get relevant information (for example number of bets, average bet amount etc) at a unique UserID level. These aggregated tables were then merged together to create a basetable with one row per unique user ID.

Additional aggregated datamarts were created, such as for country, product, seasonality (by month) and transaction type for building a Shiny application with different views.

**Marketing Summary Statistics and Metrics**

**Important metrics:**

* Win vs Stake ratio: This ratio is calculated as mean of the average wins divided by the average stakes for each product, across all users. What this tells us is, if users are winning as much money as they stake the ratio would be closer to one. Therefore, while a higher ratio would be better for gamblers (higher wins) a lower ratio would be better for the company (higher stakes, lesser wins). This information is displayed in the Seasonality tab.
* Total count of users for each gambling product: Calculated as the total number of users who participate in each gambling product.
* Total count of transactions for each gambling product: Calculated as the total number of transactions processed for each gambling product. Combined with the previous metric, it is a good way to assess how popular each product is. Both metrics are available in the Product tab.
* Average wins, average bets and average stakes: these metrics are calculated per product by taking an average from aggregated user data. They can be especially helpful when crossed by other factors such as seasonality, or other demographic information, such as age, country, etcetera, to get a wider understanding of each product offered by the company. The relation between all 3 can be seen on the tab of User Info.
* **Frequency** = how many times did a customer play, with any gambling product, during his active period. The average frequency is displayed on the User Info summary table.
* **AverageProfitPerUsr** = the total profits made by user divided by his frequency, displayed on Country tab.
  + The Frequency is already created
  + Then we need to create a variable for the total profits made by user:

Basetable$SumProfitPerUser <- Basetable$Profit\_Sports + Basetable$Profit\_Poker + Basetable$Profit\_Games + Basetable$Profit\_Casino + Basetable$Profit\_Supertoto

* + Then create our variable **AverageProfitPerUsr:**

Basetable$AverageProfitPerUsr <- Basetable$ SumProfitPerUser / Basetable$Frequency

* **RevenueMargin** = Stakes - Winnings /Stakes \*100 and then we can divide into three categories (High,Medium , Low margins), which is displayed in the User Info.
  + First we need to create a variable for Total Stakes:

Basetable$TotalStakes <- Basetable$SumofStakes\_Sports + Basetable$SumofStakes\_Games+Basetable$SumofStakes\_Casino + Basetable$SumofStakes\_Supertoto + Basetable$TotalAmount\_Bought (for poker)

* + Secondly we need to create a variable for Total Winnings :
  + Basetable$TotalWinnings <- Basetable$SumofWinnings\_Sports + Basetable$Sumofwinnings\_Games + Basetable$SumofWinnings\_Casino + Basetable$SumofWinnings\_Supertoto + Basetable$Total\_Amount\_Sold (for Poker)
  + Then create the required variable **RevenueMargin:**

Basetable$RevenueMargin <- ((Basetable$TotalStakes - Basetable$TotalWinnings) / Basetable$TotalStakes ) \* 100

* + Then based on our needs we can classify the customers into three categories (High Margin, Medium Margin, Low Margin)

Basetable$RevenueMarginGrouped <- ifelse(Basetable$RevenueMargin <= X, 'Low Margin',

ifelse((Basetable$Loyalty>Y &Basetable$Loyalty<=Z),Medium Margin,

ifelse(Basetable$Loyalty >Z,'High Margin,'NA’)))

**For poker gambling behavior, some important metrics that were calculated were:**

* Average buy amount: the average amount (value) of chips bought by users
* Average sell amount: the average amount (value) of chips sold by users
* Average times bought: the average number of times chips are bought
* Average times sold: the average number of times chips are sold

**Shiny Application**

The Shiny Application created consists of multiple tabs and interactive charts to allow an easy understanding of the data.

**Tab 1: Demographics of users by country**

This tab consists of an interactive map of world, for countries where the users of this website are present. The colors represent the Revenue Margin of the website in each country. Additionally a filter allows the user to filter the chart by continent. A hover tool option is available so that the information for each country can be easily identified. Finally, a table is included with a summary of the top 10 countries, by average profit for the continent or continents selected.

**Findings:**

Some interesting findings from this chart include:

* The lowest average age of users is mostly found in South American countries (ie in Mexico the average age is 23, in Argentina, it is 24 and in Colombia, it is 20. Additionally, Taiwan has an average user age of 22. These countries may need to be targeted differently since their demographic is much younger.
* The oldest average user age (above 40 years old) is in countries including Japan, Singapore, South Africa, Puerto Rico and Mauritius. This should be taken into consideration when targeting these users with offers as due to their age, they may have more money to spend on gambling.

**Tab 2: User Information**

This shows the density distribution for gender by age and a scatter plot for Average Wins by Average Stakes, showing the Average Bets in size and the Revenue Margin segment in color. Both charts can be filtered by Continent and Age Range, with a summary table showing other metrics for the selected features.

**Findings:**

Some interesting findings from this chart include:

* In most countries, the majority of users are male; some exceptions to this are Thailand and Vietnam where the majority of users are female. Because of this, we decided to display Gender density grouped by continent. Additionally, Gender distribution varies by age, with the age range between 20 and 35 having more female participation.
* The segments show users are well informed, as most are concentrated in Higher and Medium revenue segments. There is no significant difference in age averages, but Frequency (or number of plays) decreases for higher segments.
* There is a positive relation between Average Stakes and Average Wins, with Medium and High revenue segments being concentrated in lower Average Stakes and Average Wins. This, coupled with the previous finding, suggests that users are risk averse.

**Tab 3: Product Information**

The third tab created includes some charts related to product information.

The first chart gives the win to stake ratio for all the gambling products offered by the company. This allows marketing analysts to compare between products and identify which ones would be most profitable for the company, and direct efforts and promotions on those products.

The second chart gives information on the count of users that are participating in each gambling product, along with the count of the total count of transactions for each product.

Viewers can choose to look at all the products, or they may select certain products to compare.

**Findings:**

* From the first chart we can see that casino (bossmedia and chartwell) along with the supertoto product are more profitable for users, having a win to stake ratio of greater than 0.8, which means users would typically win back 80% or more of the money they stake. However, for the gambling website, games (bwin and VS) along with sports book fixed odd are more profitable products with a ratio between 0.67-0.69 which means users would win back only 67-69% of the money they stake.
* The second chart shows us that the highest number of users are for the sports book fixed odd product, and it also has the highest betting transactions. However, for this specific product each user is likely to make many more transactions compared to other products. This, along with the low win to stake ratio, would make sports book fixed odd the most profitable product for the website, which can incentivize marketers to invest more resources in promoting it.
* The second highest transactions and users are for the sports book live action product. Even though the win to stake ratio is not as high, it is important information for marketing analysts that sports book betting (fixed odd and live action together) makes up the largest proportion of transactions and users.
* While game products also have a lower win to stake ratio, making them more profitable for the company, the lower number of users and transactions may deter a lot of investment in promoting them. Amongst these two, games bwin fares better.

**Tab 4: Seasonality**

This tab contains two charts where the data is shown by the month of transaction. The first chart shows the average wins and average stakes by month. Again, a viewer may select all products to look at the general trend and impact of seasonality, or they can focus on a product(s) to look at how seasonality has an impact on that specific product.

The second chart shows the average number of bets by month, this can help identify if users are more likely to gamble in certain months compared to others.

**Findings:**

* Overall, considering all products, the highest average wins and stakes are in the months of March and August while the lowest are in April and June
* For sports book related products the trends are different – in live action the wins and stakes pick up over the initial months, reaching a peak in July, and then falling again. This is matched by the second chart, as the highest number of bets are also placed in July. For fixed odd these metrics are almost constant across the year, but fall after July. This means that users may start to spend less on sports book products after July. However, the number of bets placed does not match this pattern as it reaches a low in June and then rises again in the following months.
* For games products (bwin and chartwell) July and August are the most popular months, with the highest average wins, stakes and transactions.

**Tab 5: Transaction Type**

The last tab looks into some metrics regarding poker chips, such as the average buy and sell amount, average number of times bought and sold (number of transactions) and number of users carrying out these transactions.

The data is displayed throughout time, with filters available to zoom into each month and select each metric.

**Findings:**

* Average buying amount: If you look at this by the month where users made their last buying transaction month, this metric peaks for people who made their last transaction in September and is the lowest for those who made it in May. However if this is viewed from their initial buying transaction month, there is a downward trend, ie people who made their first transaction in February are more likely to have a higher average buying amount vs those who made it in October.
* Average selling amount: The highest average selling amount is for users who make their last selling transaction in September, followed by June. This may be due to the summer or Thanksgiving holidays but requires further exploration by country. This also applies to users who make their initial selling transaction in June, followed by February.
* Average times bought and average times sold: The results from these graphs make sense as users who made their initial buying transaction earlier in the year have bought and sold a higher average number of times, following a downward trend as the months increase. While users who made their last buying transaction later in the year also have higher average times bought and sold, following an upward trend as months increase. However, it is unclear if this is due to seasonality or marketing strategies.