DataMart_MarketingManual

Ying Xiao, Alice Wolfe, Bharanidharan Murugesan

20 December 2023

Background

Online gambling draws a large variety of people from around the world to play it's games. In this report and the complimentary app, your users, their use habits, and impacts on the business are explored. These explorations will help you better understand how to market to new and existing users for your business.

Please explore our app here: https://datamart-group2.shinyapps.io/datamart_app/

The DataMart

From the raw data your company provided on user demographics, poker statistics and daily interactions with products, deep insights can be drawn. Below is a description of the metrics put together to maximize your insights into your customers and their product use patterns.

Table of Marketing Metrics

Variable	Type	Description
UserID	double class	Unique IDs for each user
Language Description	Character	Language of each user
Country Name	Character	Country of each user
RegDate	Date	Registration date
FirstPay	Date	Date of first payment
FirstAct	Date	Date of first use of a product
FirstSp	Date	Date of first sports book use
FirstCa	Date	Date of first casino play
FirstGa	Date	Date of first games play
FirstPo	Date	Date of first poker play
Gender	Factor	Gender of user
FirstAggDate	Date	Date of first betting activity
LastAggDate	Date	Date of last betting activity
TotalAggDays	Double Class	Length of betting relationship in days
TotalWinnings	Double Class	Total winnings across all products per user
AvgWinnings	Double Class	Average winnings across all products per user
TotalStakes	Double Class	Total stakes across all products per user
AvgStakes	Double Class	Avg stakes made per play per user
TotalBets	Double Class	Total bets made per user
Earning	Double Class	Total winnings less losses per user
AllProductDescriptions	Character	List of all products a user interacts with

Variable	Type	Description
ProductComboCount	Integer	Count of the number of users with the same product combination
TransactionDays	Double Class	Length of poker relationship in days
TimesOfPlaying	Double Class	Number of times a user played poker
Buy_Amount	Double Class	Total buy amount per user
Sell_Amount	Double Class	Total sell amount per user
LoR_days	Double Class	Length of relationship from registration date to last action
TimeToFirstBet_days	Double Class	Number of days between registration and first bet made
TotalSpend	Double Class	Total spend across all products per user
SpendCategory	Character	Spending category per user: Low, Medium, High or Profit Risk (indicating they win more than they spend)
TotalAvgWins	Double Class	Average winnings per product combination
AvgTotalBets	Double Class	Average bet amount per product combination
AvgDaysToFirstBet	Double Class	Average days until first bet per product combination
Cas_BM_X	Double Class	Group of summary statistics by user for Pro Casino BossMedia product
Cas_CW_X	Double Class	Group of summary statistics by user for Pro Casino Chartwell product
G_Bwin_X	Double Class	Group of summary statistics by user for Pro Games bwin product
G_VS_X	Double Class	Group of summary statistics by user for Pro Games VS product
Pok_BM_X	Double Class	Group of summary statistics by user for Pro Poker BossMedia product
SB_FO_X	Double Class	Group of summary statistics by user for Pro Sports book fixed-odd product
SB_LA_X	Double Class	Group of summary statistics by user for Pro Sports book live-action product
TOTO_X	Double Class	Group of summary statistics by user for Pro Supertoto product

Creating the DataMart and Important Marketing Metrics

To create the DataMart, we started by cleaning and processing the raw data provided. This included converting data into appropriate data types, removing duplicate values, handling missing bet values by filling with 0, and other general cleaning activities.

Once each raw dataset was cleaned, we transformed the data through aggregation to create a final table that was on a per user level AKA one row of data for each user. Through this aggregation, the initial marketing metrics were created as follows:

AGGREGATION SUMMARY PER USERID We calculated the Total Aggregation Days (TotalAggDays) by selecting the first and last aggregation date for each user, and calculating the difference.

```
DailyAgg <- DailyAgg %>%
  group_by(UserID) %>%
  mutate(
```

```
    FirstAggDate = min(Date),
    LastAggDate = max(Date),
    TotalAggDays = as.numeric(difftime(max(Date) + 1, min(Date), units = "days"))
  ) %>%
  ungroup()
```

SUMMARY RELATED TO WINNING, STAKES, AND BETS We calculated aggregated summary statistics of users' winnings, stakes and bets then found each user's earnings by subtracting stakes from winnings.

```
DailyAgg <- DailyAgg %>%
  group_by(UserID) %>%
  arrange(UserID) %>%
  mutate(TotalWinnings = sum(Winnings),
         AvgWinnings = mean(Winnings),
         TotalStakes = sum(Stakes),
         AvgStakes = mean(Stakes),
         TotalBets = sum(Bets),
         Earning = sum(Winnings - Stakes))
```

TOTAL PRODUCTION DESCRIPTION We created the Total product description for each UserID & Earning Per Customerby aggregating by userID and concatenating all unique product descriptions associated with that user.

```
DailyAgg <- DailyAgg %>%
  group_by(UserID) %>%
  arrange(ProductID) %>%
  mutate(AllProductDescriptions = paste(unique(‘Product Description‘), collapse = ", "))
```

COUNTS OF PRODUCTION DESCRIPTION OCCURENCES We calculated how many users had each specific combination of products in order to find the most popular product combinations.

```
DailyAgg <- DailyAgg %>%
  group_by(UserID, AllProductDescriptions) %>%
  mutate(ProductComboCount = n()) %>%
  ungroup()
```

SUMMARY FOR EACH PRODUCT We calculated aggregated summary statistics by individual user for each product. The statistics we gathered were sum and average of winnings, sum and average of stakes, sum of bets and the earnings calculated in the same manner as previously described. Here is an example code for product Casino BossMedia.

```
PRODUCT <- DailyAgg %>%
  group_by(UserID,ProductID)%>%
  summarise(Cas_BM_TWin = ifelse( ‘Pro Casino BossMedia‘  == 1, sum(Winnings),0),
            Cas_BM_AWin = ifelse(‘Pro Casino BossMedia‘ == 1, mean(Winnings),0),
            Cas_BM_TStakes = ifelse(‘Pro Casino BossMedia‘ == 1, sum(Stakes),0),
            Cas_BM_AStakes = ifelse(‘Pro Casino BossMedia‘== 1, mean(Stakes),0),
            Cas_BM_TBets = ifelse(‘Pro Casino BossMedia‘ == 1, sum(Bets),0),
            Cas_BM_Earn = ifelse(‘Pro Casino BossMedia‘ == 1, sum(Winnings-Stakes),0))
```

TOTAL DATE OF PLAYING We created a variable to capture the length of time in days a user has played for.

```
PokerChip <- PokerChip %>%
  group_by(UserID) %>%
  mutate(TransactionDays = as.numeric(difftime(max(TransDateTime) + 1, min(TransDateTime), units = "days
```

TIMES OF PLAYING We considered each sell action as one time of playing. Here we count how many times each User played.

```
PokerChip <- PokerChip %>%
  filter(tolower(TransType) == "sell") %>%
  group_by(UserID) %>%
  mutate(TimesOfPlaying = n())
```

TRANSACTION SUMMARY We calculated the total buy and sell amount for each user.

```
# Summarize buy and sell amount for each customer
PokerChip <- PokerChip %>%
  group_by(UserID) %>%
  mutate(Buy_Amount = sum(TransAmount[TransType == "Buy"]),
         Sell_Amount = sum(TransAmount[TransType == "Sell"]))
```

LENGTH OF RELATIONSHIP We calculated the Length of relationship as the time gap between registration date and last betting activity date.

```
DataMart <- DataMart%>%
  mutate(LoR = LastAggDate - RegDate)
#convert Length of Relationship to numeric for future calculations
DataMart$LoR_days <- as.numeric(DataMart$LoR, units='days')
DataMart <- subset(DataMart, select = -c(LoR))
```

TIME TO FIRST BET We calculated Time to first bet as the time gap between registration date and first betting activity date.

```
DataMart <- DataMart%>%
  mutate(TimeToFirstBet = FirstAggDate - RegDate)
#convert Length of Relationship to numeric for future calculations
DataMart$TimeToFirstBet_days <- as.numeric(DataMart$TimeToFirstBet, units='days')
DataMart <- subset(DataMart, select = -c(TimeToFirstBet))
```

LOYALTY Loyalty shows the relationship between user's total bets and the length of relationship.

```
DataMart <- DataMart%>%
  mutate(Loyalty = TotalBets/LoR_days)
```

SPEND CATEGORIES We calculated the total spending of each user.

```
DataMart <- DataMart%>%
  mutate(TotalSpend = Buy_Amount - Sell_Amount - Earning )
```

Using these total amounts, we calculated summary statistics on total spend to determine spend categories

```
summary(DataMart$TotalSpend)
```

We found the minimum Total Spent by a user was $-37,037.76, Median was $49.60, Average was $201.75 and the maximum was $76,165.57

Then we created a function to categorize the TotalSpend column into categories: Low Spend, Medium Spend, High Spend and Profit Risk (AKA users who win more than they spend) through the use of an lapply function.

```
category <- function(x){
  if (is.na(x)){
    return('N/A')
  }else if (x<0){
    return('PROFIT RISK')
  }else if ((x>=0)&(x<=50)){
    return('low')
  }else if ((x>50)&(x<=1000)){
    return('medium')
  }else {
    return('high')
  }
}

DataMart$SpendCategory <- unlist(lapply(DataMart$TotalSpend, category))
```

AVGWINNINGS BY PRODUCT COMBINATION We then looked at average winnings across all product combinations that users had in order to see trends across popular product combinations.

```
Avgwin <- DataMart %>%
  group_by(AllProductDescriptions) %>%
  summarise(TotalAvgWins = mean(AvgWinnings))
```

TOTAL BETS BY PRODUCT COMBINATION This process was repeated with total bets data.

```
AvgTotBet <- DataMart %>%
  group_by(AllProductDescriptions) %>%
  summarise(AvgTotalBets = mean(TotalBets))
```

TIMETOFIRSTBET_DAYS BY PRODUCT COMBINATION Finally, we looked at the average number of days until a users first bet aggregated by their specific product combination.

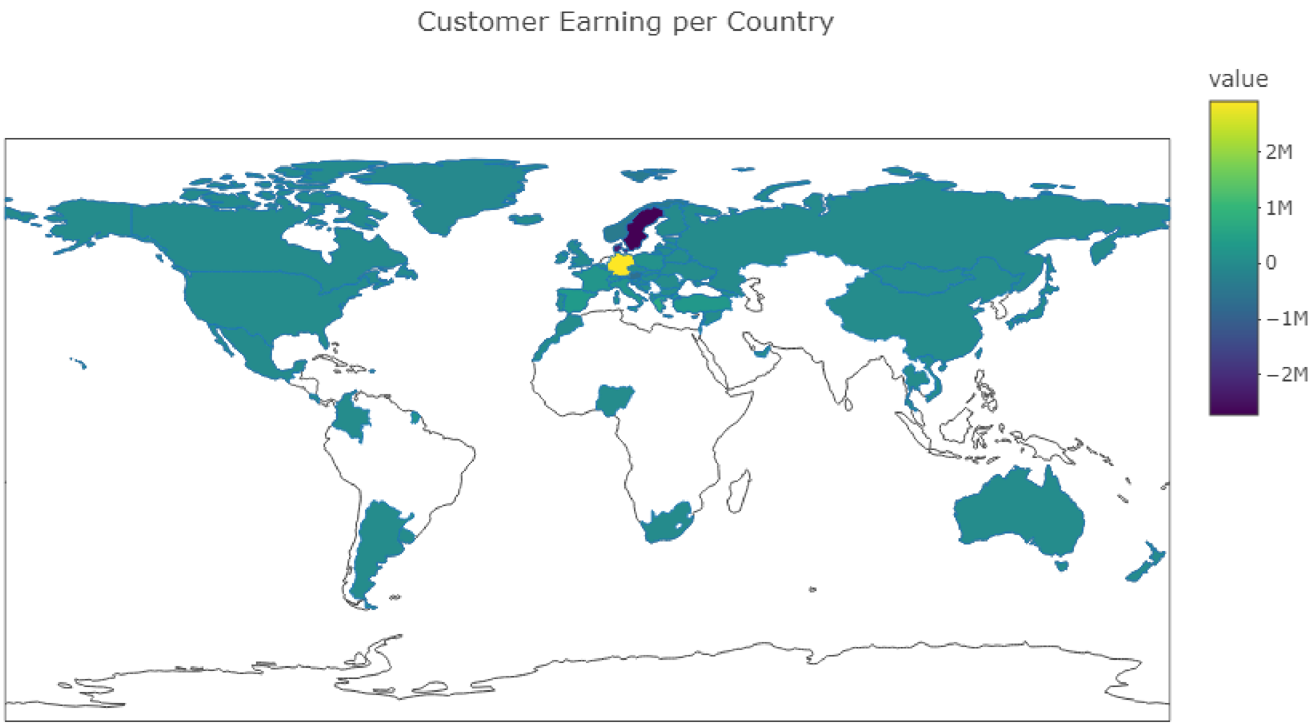
```
AvgDaysToFirstBet <- DataMart %>%
  group_by(AllProductDescriptions) %>%
  summarise(AvgDaysToFirstBet = mean(TimeToFirstBet_days, na.rm = TRUE))
```

Summary Statistics

These marketing metrics can be best explored on our convenient Shiny App, however for ease, we've highlighted key summary statics for each app section.

APP TAB 1 |GEOGRAPHICAL IMPACT

For the Geographical impact, we want to check the number of customer and cutomer earning per country.

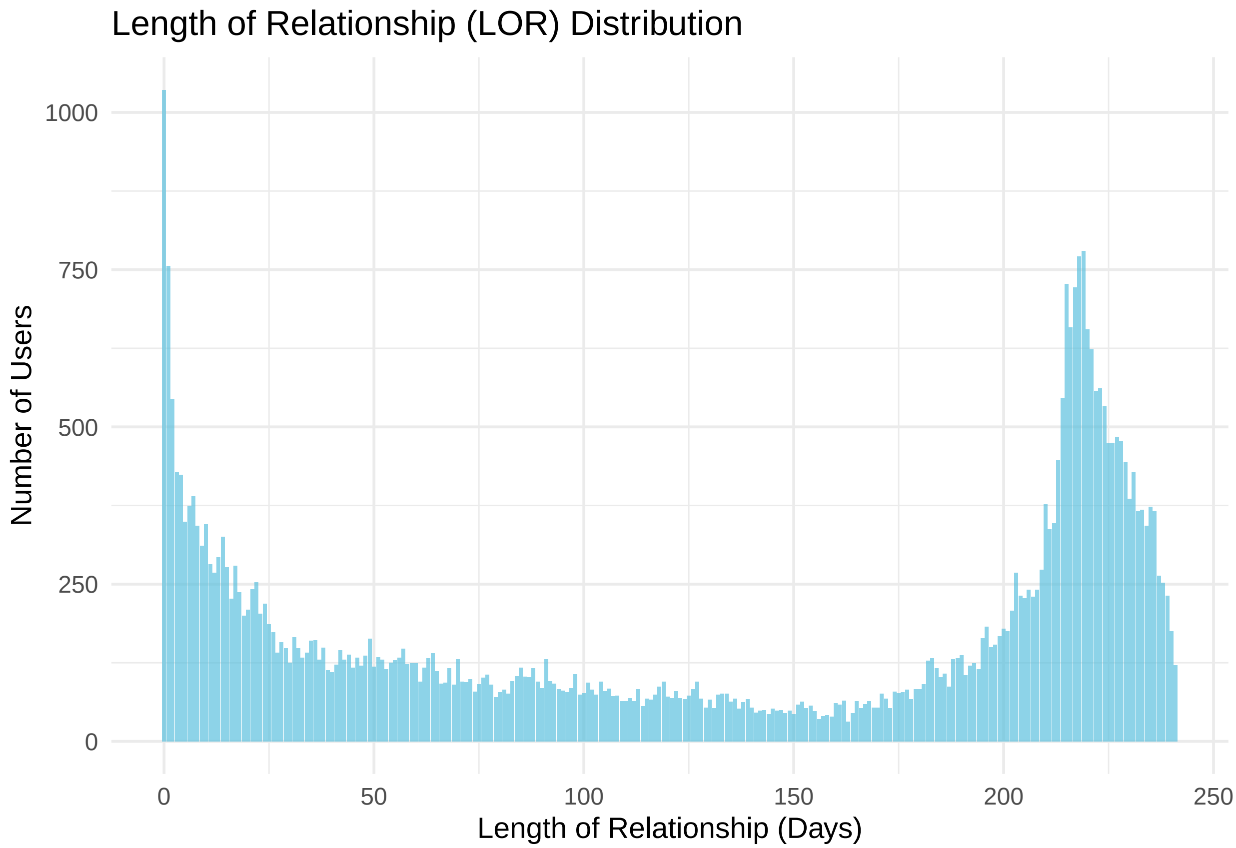


We found that most of our user (24k) are German and they earn in total 2.9 million Euro. However, we only have 190 users in Sweden, their earning is negative 2.7 million in total.

APP TAB 2 | USER ANALYSIS

In this part, we mainly visualize the length of relationship and time to first bet.

```
## Warning: Removed 2092 rows containing non-finite values (‘stat_count()‘).
```



We observed that the majority of users can be categorized into two groups:

quick churn users (LOR less than 25 days)

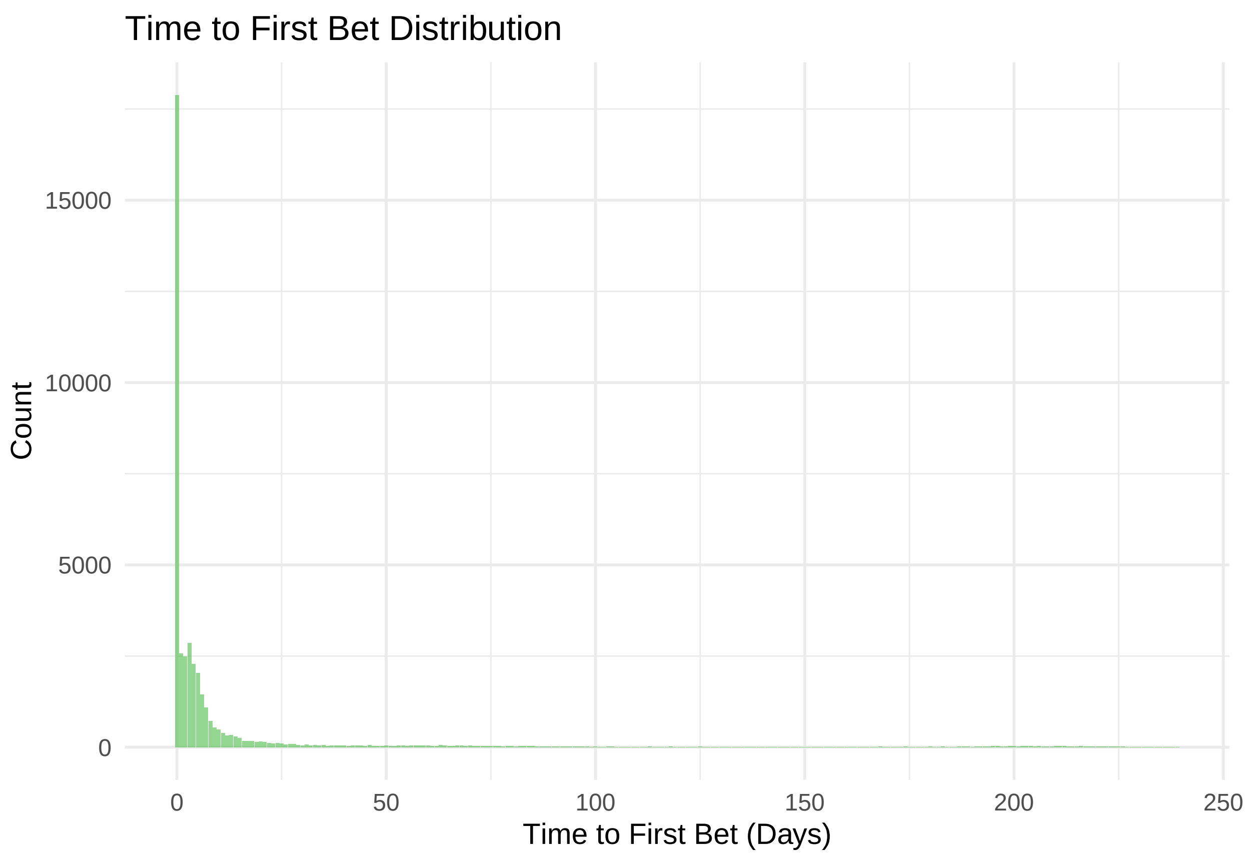
loyal users (LOP greater than 200 days)

There are over 1000 individuals who only attempt for one day.

Overall, the number of loyal users is significantly high.

```
## Warning: There were 2092 warnings in ‘summarise()‘.
## The first warning was:
## i In argument: ‘TimeToFirstBet_days = max(TimeToFirstBet_days, na.rm = TRUE)‘.
## i In group 11: ‘UserID = 1324371‘.
## Caused by warning in ‘max()‘:
## ! no non-missing arguments to max; returning -Inf
## i Run ‘dplyr::last_dplyr_warnings()‘ to see the 2091 remaining warnings.

## Warning: Removed 2092 rows containing non-finite values (‘stat_count()‘).
```

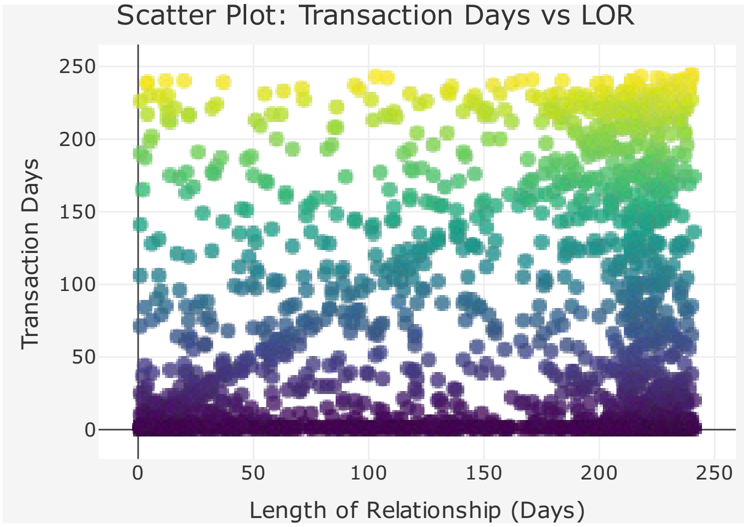


More than 95% of the Users play at the first day they registered. 99% of the users play within 25 days.

APP TAB 3 | LENGTH OF RELATIONSHIP ANALYSIS

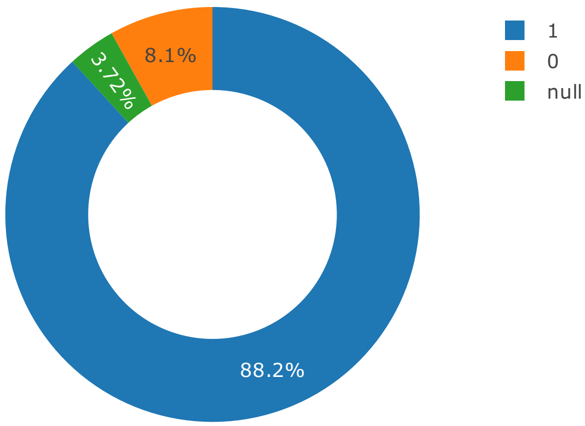
Users with longer relationships typically exhibit a sustained pattern of transactions over time, indicating their continuous engagement with the platform. However, there is also a subset of these users with relatively shorter transaction days. Overall, there is a clear positive correlation between transaction days and the length of the relationship.

```
## No trace type specified:
##   Based on info supplied, a 'scatter' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter
```



APP TAB 4 | GENDER INSIGHTS

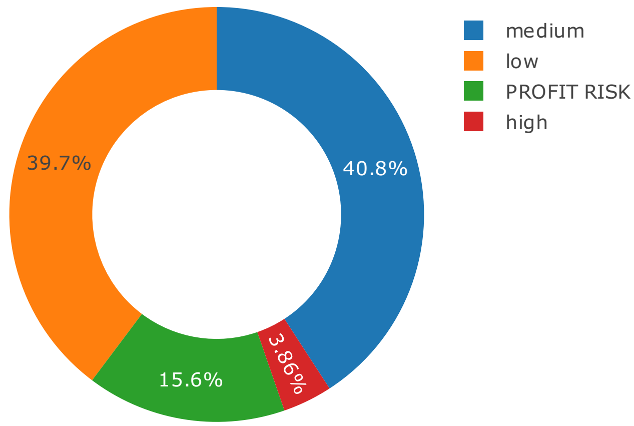
When inspecting the gender of all users, we find the vast majority of users are men. This should be accounted for in all marketing endeavors.



APP TAB 5 | SPENDING CATEGORIZATION INSIGHTS

When inspecting the categorization of user spending, we see that most users fall into the Low or Medium categories, spending between $0 - $1,000 on products. High spenders spending more than $1,000 on products are the smallest group although the maximum spend is $76,166.

Most interestingly, 15% of users are considered a Profit Risk meaning they win more money than they spend on products. These users should be flagged for special consideration in future marketing activities to either increase their spending, play with new products or discourage use of the products all together.



APP TAB 6 | PRODUCT COMBINATION INSIGHTS

Looking into the specific product combinations that users frequently use and how they impact their earnings and winnings can provide very valuable marketing insight. It can guide product bundling opportunities, assist in user segmentation and targeting and provide insights on how users play across products. Below please find the top 10 most popular product combinations.