# Decoding Gambling Behavior

Problem Statement

Bustabit, an online gambling platform, has been giving random multipliers of bonuses to players after each winning bet. In order to better cater to the players and increase house’s income and leverage increase in games played by the players, Bustabit would like to know different types of players based on their betting behaviour and risk-taking patterns.

Online gambling is the act of placing wagers on risk-reward games for a chance of winning money. Each person places bets based on their own individual intuition and financial availability. Understanding the similarities and differences in the customers’ betting behaviour in this context can help the gambling platform make informed decisions that can help increase revenue of the platform.

There are a few basic rules for playing a game of Bustabit:

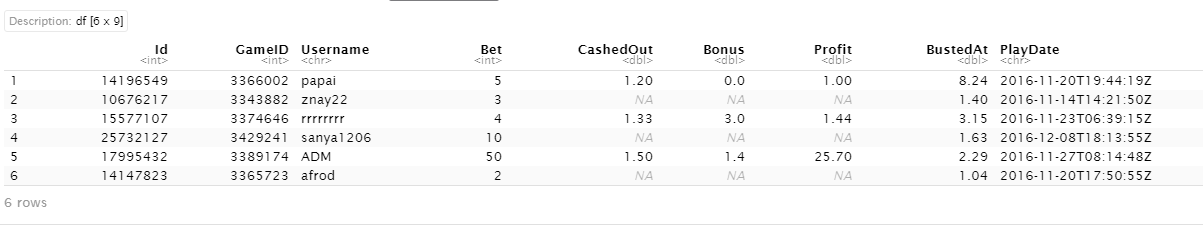
1. You bet a certain amount of money (in Bits, which is 1 / 1,000,000th of a Bitcoin) and you win if you cash out before the game **busts**.
2. Your win is calculated by the multiplier value at the moment you cashed out. For example, if you bet 100 and the value was 2.50x at the time you cashed out, you win 250. In addition, a percentage Bonus per game is multiplied with your bet and summed to give your final Profit in a winning game. Assuming a Bonus of 1%, your Profit for this round would be (100 x 2.5) + (100 x .01) - 100 = 151
3. The multiplier increases as time goes on, but if you wait too long to cash out, you may bust and lose your money.

Data Set

For the purpose of this project, the following data has been collected from Bustabit. The dataset consists of information on 50000 games of Bustabit played by 4150 different players on the platform. The data includes the following variables:

1. **Id** - Unique identifier for a particular row (game result for one player).
2. **GameID** - Unique identifier for a particular game.
3. **Username** - Unique identifier for a particular player.
4. **Bet** - The number of Bits (1 / 1,000,000th of a Bitcoin) bet by the player in this game.
5. **CashedOut** - The multiplier at which this particular player cashed out.
6. **Bonus** - The bonus award (in percent) awarded to this player for the game.
7. **Profit** - The amount this player won in the game, calculated as (Bet \* CashedOut) + (Bet \* Bonus) – Bet.
8. **BustedAt** - The multiplier value at which this game busted.
9. **PlayDate** - The date and time at which this game took place.

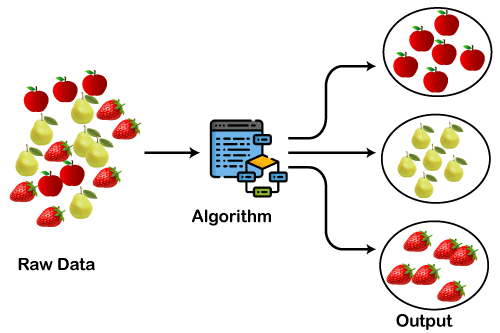
The data set consisting of all the above variables looks as follows.

****

Dataset as collected from the source.

Approach

Giving higher bonus to cautious players and lower bonus to risk takers is not a productive way to leverage players into betting again. Hence, in order to determine appropriate amount of bonus to be given to the players, we apply K-Means clustering algorithm on the players’ data to divide them into different clusters based on their betting behaviour.

K-Means clustering is one of commonly used clustering algorithms where K is the numbers of clusters the players are divided into. The clusters are formed by comparing the similarities and differences (in variables) between observations (game info) and each cluster has multiple observations which are similar to each other and each cluster is separated from the other based on the differences among observations from one cluster to another.

Representation of working of K-Means Clustering Algorithm.

The similarities between two observations can be calculated in terms of distance in clustering methods. More similarities decrease the distance among the observations while less similarities increase the distance among them, thus forming the clusters and their centroids. A new observation added to the cluster changes the shape of the cluster and hence the centroid observation, until all the input observations have been segmented into appropriate clusters.

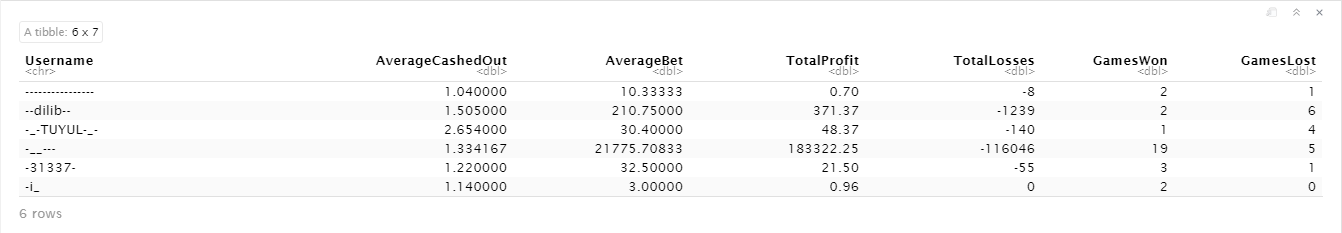
Data Preparation and Application of Algorithm

In order to perform the analysis, let us first take a look at how the data looks like;

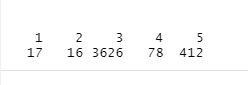
Dataset sorted in descending order of BustedAt value.

The above table consists of the basic information provided by Bustabit with game level statistics, that was sorted in descending order of the value at which a game was busted.

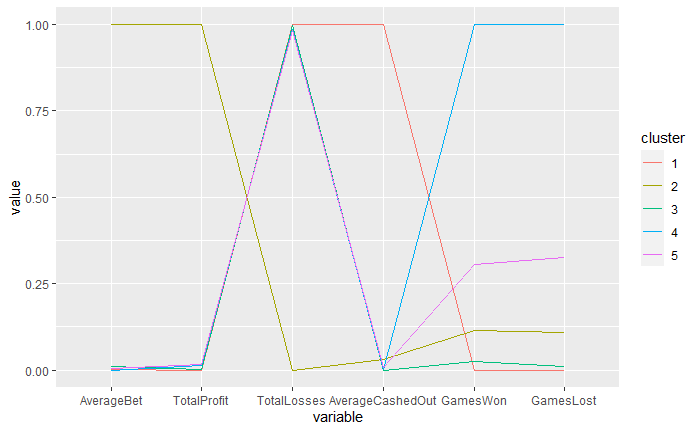
In order to cluster the players, we need to prepare player level statistics from the given game-wise data. Also, to better quantify the player behaviour, we will need some more variables like losses incurred when a game was busted and an indicator variable that quantifies whether the game was a win or loss. To create per-player statistics, we find average values of all the variables of the new data for each player.

Upon applying all the above data cleaning and preparation steps, the following would be the final input data with per-player stats.

Representation of Input data with Average Stats per Player

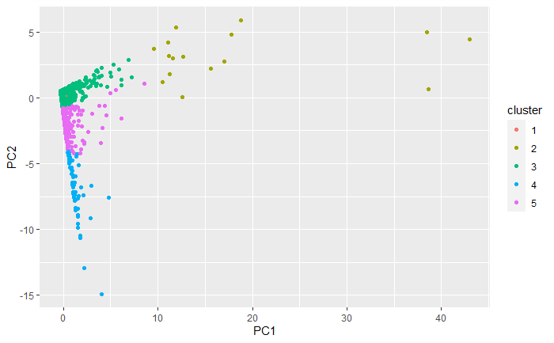
With the standardized data of per-players features, we now apply the K-Means clustering algorithm in order to cluster the players based on their gambling behaviour. Upon applying the algorithm, the following were the number of observations or players sorted into 5 different clusters.

Number of Players sorted into 5 distinct clusters.

In order to clearly distinguish the clusters from one another, we compute the average values of the attributes in each cluster and compare them. The following is the final result of the clusters and their average values in each attribute which helps us determine the distinct features of each cluster based on their gambling behaviour.

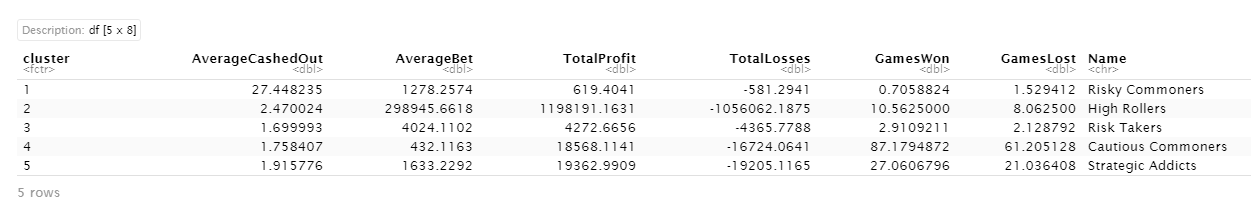
The above graph plots the values of variables of each cluster in a scale of 0 to 1, which differentiates one cluster from another with each cluster having different values across the variables. A value of 1 for a particular variable indicates that cluster has the highest value compared to all other clusters, and a value of 0 indicates that it has the lowest. This can help make relative comparisons between the clusters clearer.

The clusters can also be represented in a scatter plot of players plotted against two of the principal components to be able to interpret the plot in 2 dimensions. The below graph shows the scatter plot of the players divided into clusters with colour code. Each cluster can be differentiated looking at the placement of the observations of each cluster in the graph.



Scatter plot of players with principal components.

Output Analysis

****The application of K-Means clustering algorithm to the per-player input data gave us the following output based on the gambling behaviour of players.

Output of K-Means Clustering along with Labels

The above clusters can be interpreted as below based on the distinct values of each variable among the clusters. Each cluster is named appropriately based on the attributes of the clusters.

**Cautious Commoners:**

This is the largest of the five clusters, and might be described as the more casual Bustabit players. They've played the fewest number of games overall, and tend to make more conservative bets in general.

**Strategic Addicts:**

These users play a lot of games on Bustabit, but tend to keep their bets under control. As a result, they've made on average net positive earnings from the site, in spite of having the most games played. They seem to maintain a strategy (or an automated script/bot) that works to earn them money.

**Risky Commoners:**

These users seem to be a step above the Cautious Commoners in their Bustabit gambling habits, making larger average bets, and playing a larger number of games on the site. As a result, though they have about the same number of average games won as the Risk Takers, they have a significantly higher number of games lost.

**Risk Takers:**

These users have played only a couple games on average, but their average cashed out value is significantly higher than the other clusters, indicating that they tend to wait for the multiplier to increase to large values before cashing out.

**High Rollers:**

High bets are the name of the game for this group. They bet large sums of money in each game, although they tend to cash out at lower multipliers and thus play the game more conservatively, particularly compared to the Risk Takers. Interestingly, these users have also on average earned net positive earnings from their games played.

Conclusion

From the above analysis of Bustabit’s data and clustering based on gambling behaviour of players, we can conclude that the online gambling platform will now be able to make informed decision on the amount of bonus to be given to different clusters of players in order to leverage increased revenue with increase in number of games played by the customers. For example, a player that comes under the cluster ‘Strategic Addicts’ would be more inclined to play more games and increase his bet if he observes a pattern of bonus getting doubled every 3rd game that he plays.

Hence, K-Means clustering is very helpful for the gambling platform in making decisions on strategically advancing on their marketing techniques in order to improve their revenue and gains.