**Analyzing User Expenditure Vs Social Interaction and Performance**

**AIM**:

To study the social behaviors and Performances of high spending users in the Game of Kingdom of Camelot.

**ABSTRACT**:

***Social Interaction Vs User Expenditure***:

There is a ***positive, but a weak correlation*** between the User Expenditure and the social interaction by the user. Graph shows an increasing trend of interactions with the increase in the amount of cents spent by the user. *Correlation coefficient* between them is *0.289496*.***Distance Correlation*** between them is ***0.3758944*.** The ***mutual Information*** shared by them is **4.26 bits.** These statistics shows that the relationship between them is not strong

***User Performance Vs User Expenditure:***

There is a ***positive, but a very weak correlation*** between the User Expenditure and the Performance by the user. . The ***Logarithmic increase in the expenditure*** increases the User Performance **very weakly**. The User Performance is not affected by User’s expenditure. Perhaps our performance metric is too strict. *Correlation coefficient* between them is ***0.1235699***. ***Distance Correlation*** between them is ***0.1880827*.** The ***mutual Information*** shared by them is **5.203 bits.** These statistics shows that the relationship between them is very weak.

**PURPOSE**:

1. The Gamers who spend money on games are our VIP customers.
2. We need to study their performance via Performance metrics.
3. We need to know if they are struggling to have a feel-good experience while gaming or feeling bored being the best and are not challenged enough to be excited.
4. By analyzing the User Performances with their expenditure, we can understand if their money is well spent.

We need to know if we had made them feel the worth of the money they spent and they are encouraged to pay more.

1. We need to understand their impressions, ideas, likes and dislikes of our game when they play.
2. The interaction logs obtained from the alliance chat rooms give a peek at what the players are communicating to each other.
3. By analyzing the relationship between the User expenditure and the User social interaction gives us a notion of what they feel about our games. Do they want to sustain the relationship/business with us? Do they find their time and money are worth spent? Do they feel excited for our new games /levels/changes/. Are they encouraged to spend more and buy more games from us?
4. It also gives us ideas on how to enrich their Gaming experience.

**PROCEDURE**:

The total number of messages given by each user is computed from alliance\_chat.txt

The total amount spend the individual user is obtained from User\_data.txt

The User Performance is obtained from attacks2.txt by metrics mentioned before.

Parsing: Perl, Python Back End: Mysql.

**USER EXPENDITURE AND SOCIAL BEHAVIOUR OF THE USER**

**GRAPH**:

The total number of messages given by the users is obtained plotted with the total amount spent by the Users. We have the following graph.

**X AXIS**:

The ***Amount of money spent by User in Cents***. In LOG SCALE (base 10).

**Y AXIS**:

The ***Count of messages sent by User in alliance chat rooms***.

The X axis value is in Log scale. A logarithmic scale is a nonlinear scale used when there is a

Large range of quantities as compared to a linear scale present in Y axis.

**DATA**: Alliance\_chat.txt

Total Number of Users who participated in chat rooms by typing messages: 6160.

Total Number of Users who spent money: 506707.

We focused on the ones who has at least one message in the chat rooms.

**OBSERVATION**: We see a slightly increasing trend in the number of messages corresponding to the amount of money spent. This could be explained at the Users would want to get more feel of the game, becoming more serious about gaming and engage themselves showing an involvement given the amount of money they have spent. On the other hand, there is still a great number of people who does not engage socially while spending more money on games. They are probably more of single person.

**USER EXPENDITURE AND SOCIAL BEHAVIOUR OF THE USER**

**GRAPH**:

The Performance of the User (computed in another report) is plotted with the total amount spent by the Users.

We have the following graph.

**X AXIS**: The ***Amount of money spent by User in Cents***. In LOG SCALE

**Y AXIS**: The ***Performance of the User in alliance chat rooms***. In LOG SCALE

The Both X and Y axes value is in Log scale. A logarithmic scale is a nonlinear scale used when there is a

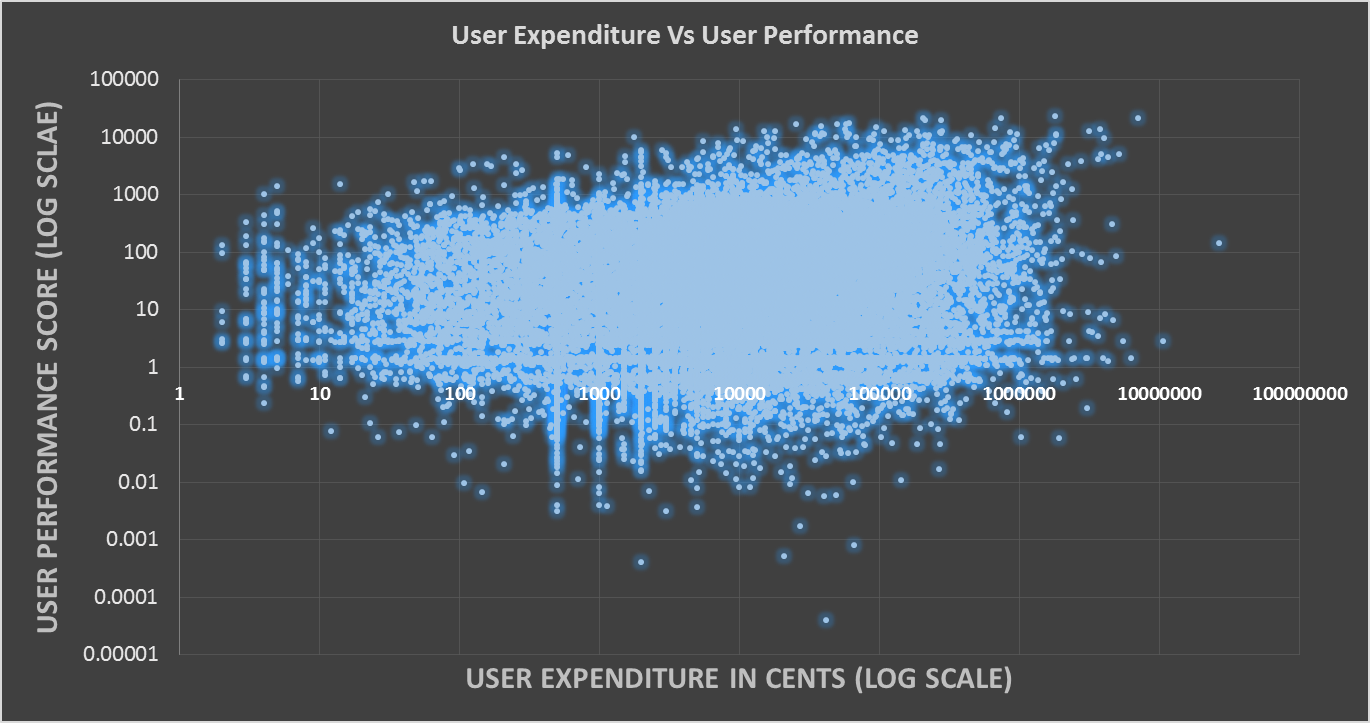
Large range of quantities.

**DATA**: Alliance\_chat.txt

Total Number of Users who spent money: 506707.

Total Number of User who .

**OBSERVATION**: We see an almost random distribution of the Performance of the User with the Expenditure with a ***very weakly increasing*** trend shown here. The ***Logarithmic increase in the expenditure*** increases the User Performance **very weakly**. This is not a great encouragement for a User. If the User has to spend logarithmic amount of money to show a slight increase in the User Performance, there is not much value provided. Perhaps the Metric to measure User Performance is too straight.

****

Weakly increasing trend.

**STATISTICS:**  (using R)

**Correlation Coefficient(Pearson’s r)**: It is a measure of the strength and direction of the linear relationship between two Random variables(X,Y).  A ***Random variable*** is a real-valued function defined on a set of possible outcomes. Here, the number of chat messages per User and the money spent by the User are two random variables we want to study. Computing the Correlation-Coefficient helps us find the linear relationship between them. The values lies between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation.

1. *Positive correlation* between X and Y is when X increases Y also increases or when X decreases Y also decreases.
2. *Negative correlation* between X and Y is when X increases Y decreases and vices versa.
3. *Zero correlation* between X and Y is when increase or decrease in X does not affect Y linearly.

r = \frac{\sum ^n _{i=1}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum ^n _{i=1}(X_i - \bar{X})^2} \sqrt{\sum ^n _{i=1}(Y_i - \bar{Y})^2}}

1. ***Correlation-coefficient between the amount spent and the number of chat messages sent by the user*** is ***= 0.289496.***
2. ***Correlation-Coefficient between the amount spent and the User Performance is=0.1234699***

These indicates a weak positive linear correlation between them.

**Distance Correlation(Dcor)**:

The correlation Coefficient doesn’t help us with Non-Linear correlation between the Random variables X and Y. That’s where Distance correlation helps us [2]. Zero correlation in Correlation-Coefficient means that the Random variables are not related linearly. But this could mean they are related in a non-linear fashion.

Distance correlation is a measure of statistical dependence between two random variables. It uses pair-wise distances between the observations as part of the calculation. Distance correlation satisfies 0 ≤ Dcor ≤ 1, and d = 0 only if X and Y are independent.

1. ***Distance Correlation between the amount spent and the number of chat messages sent by the user is*** ***=0. 3758944***
2. ***Distance Correlation between the amount spent and the User Performance is 0.1880827***

Thus the Random variables Money spent by the User and the chat messages sent by the User are not Independent. But are not strongly correlated. (The formula/calculations for distance correlation are complex [1]. Avoided for clarity)

**Mutual Information (MI)**:

Now that we have computed the linear and non-linear correlations between the two random variables. We can now look into the information they share amongst each other independent of any relationships between them. How knowing one of the Random variables does affects the knowledge of the other random variable? In other words. How does ***knowing the fact that the there is an increase/decrease in the User’s spending on games affects our knowledge on the amount of interactions the user is bound to have in alliance chats or the Performance of the User .*** Mutual information provides the amount of information one variable reveals about another [3]. It ranges from 0 to Infinity, does not depend on the functional form underlying the relationship. It is measure in bits. The higher the Mutual Information the higher the information share.

 I(X;Y) = \sum_{y \in Y} \sum_{x \in X} 
                 p(x,y) \log{ \left(\frac{p(x,y)}{p(x)\,p(y)}
                              \right) }, \,\!


where  is the joint probability distribution function of *X* and *Y*.

are the marginal probability distribution functions of *X* and *Y* respectively.

1. ***Mutual Information between the number of chat messages and the money spent by User = 4.269688 bits***
2. ***Mutual Information between the User Performance and the money spent by User= 5.205188***

This indicates that there is a positive information share between the amount of money spent by the Users to the number of chat messages by the User as well as the Performance. There is little more information share by the User Performance to the money spent than the number of chats.

**CONCLUSION:**

***Social Interaction Vs User Expenditure***:

There is a ***positive, but a weak correlation*** between the User Expenditure and the social interaction by the user. Graph shows an increasing trend of interactions with the increase in the amount of cents spent by the user. *Correlation coefficient* between them is *0.289496*.***Distance Correlation*** between them is ***0.3758944*.** The ***mutual Information*** shared by them is **4.26 bits.** These statistics shows that the relationship between them is not strong

***User Performance Vs User Expenditure:***

There is a ***positive, but a very weak correlation*** between the User Expenditure and the Performance by the user. . The ***Logarithmic increase in the expenditure*** increases the User Performance **very weakly**. The User Performance is not affected by User’s expenditure. Perhaps our performance metric is too strict. *Correlation coefficient* between them is ***0.1235699***. ***Distance Correlation*** between them is ***0.1880827*.** The ***mutual Information*** shared by them is **5.203 bits.** These statistics shows that the relationship between them is very weak.

**REFERENCES:**

1. http://en.wikipedia.org/wiki/Distance\_correlation.
2. <http://www3.nd.edu/~mclark19/learn/CorrelationComparison.pdf>
3. http://www.scholarpedia.org/article/Mutual\_information.