KABAM DATA SCIENCE A CASE STUDY

YOGESH PADMANABAN

WHY DATA SCIENCE?

Data Science is the art and science of knowledge extraction from data.

Given the Data,

1. What does it mean/imply to have these observations?
2. What are the Patterns and anomalies hidden in the data.
3. What stories does it tell ? The ecosystem behind the data.
4. What should we expect in the future from the data.

DATA SET

1. Attacks Data.
2. Alliance Chat.
3. User Data.
4. Date: 2013/10/25 to 2013/12/18

AGENDA

1. Measuring User Performance
2. Classification of Users as Spenders and Non Spenders.
3. Relationships between User Expenditure Vs Performance & Social Interaction.
4. Text Processing of Alliance Chats.
5. User Choices of Alliances, Knights and Geo Politics.
6. Location Vs User Performance.
7. IDEAS

I.MEASURE USER PERFORMANCE

PURPOSE:

- Gauge User proficiency and Alliance proficiency.
- Help the poor performers. (Boosts, Training, Easy Mode)
- Challenge the Strong ones. (Medals, Challenging Rounds, Tough Modes).
- Establish the Leaderboard for Users/Alliances/Knights.

I.MEASURE USER PERFORMANCE

PROCEDURE: User Performance is computed for each match.

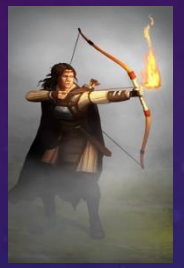
It is then averaged across matches. It is defined by the following scores.

- *Defense Score*
- *Attack Score*
- *Knight Attack Score*
- *Knight Defense Score*
- *Winning Count*
- *Losing Count*
- *Draw Count*

$$User\ Performance = Wincount - LostCount + 0.5 \times DrawCount + DefenceScore + AttackScore + KnightAttackScore + KnightDefenceScore$$

I. MEASURE USER PERFORMANCE





MIGHT SCORE:

$$Defender\ S0Might = \sum^{Army\ Size} Might\ of\ the\ Defense\ Troop \times \#\ of\ members$$

$$Defender\ S0MightLost = \sum^{Army\ Size} Might\ of\ the\ Defense\ Troop\ died \times \#\ of\ members\ died$$

LOOT SCORE:

$$LOOTSCORE = LOOTGOLD \times 10 + LOOTFOOD \times 5 + LOOTWOOD \times 3 + LOOTORE \times 2 + LOOTSTONE$$





I.MEASURE USER PERFORMANCE

DEFENSE SCORE: It is a comprehensive score measuring User's defending capabilities. It is defined by the following:

- Ratio Of *Wall strength to Rounds fought (+).*
- *S0Might* of Defending User (+).
- *S0MightLost* by Defending User (-15%).
- *S1MightLost* of Attacking Enemy (+).
- *S1Boosts* taken by Attacking Enemy (+10%).
- *S0Boost* taken by Defending User (-1%).
- Enemy's Combat level (+ 10%).
- Defending User's Combat level (+).
- Enemy's Loot Score(-15%).

$$DefenseScore = log\left\{\left(\frac{Wall}{Rounds}\right) + S0Might - (S0Boost) \times 0.01 - 0.15 \times (S0MightLost + LootScore) \right.$$
$$\left. +0.1 \times (S1Boost + \ S1CombatLevel\ ) + \ S1MightLost \ + S0CombatLevel \right\}$$

I.MEASURE USER PERFORMANCE

ATTACK SCORE: It is a comprehensive score measuring User's attacking capabilities. It is defined by the following:

- Ratio Of *Wall strength to Rounds fought (-).*
- *S1Might* of Attacking User (+).
- *S1MightLost* by Attacking User (-15%).
- *S0MightLost* of Defending Enemy (+).
- *S0Boosts* taken by Defending Enemy (+10%).
- *S1Boost* taken by Attacking User (-1%).
- Enemy's Combat level (+ 10%).
- Defending User's Combat level (+).
- Loot Score(+).
- Experience (+).

$$AttackScore = log\left\{\begin{matrix} S1Might - \left(\frac{Wall}{Rounds}\right) + LootScore - (S1Boost) \times 0.01 - 0.15 \times (S1MightLost) \\ +0.1 \times (S0Boost + \ S0CombatLevel\ ) + \ S0MightLost \ + S1CombatLevel + XP \end{matrix}\right\}$$

I.MEASURE USER PERFORMANCE

1. Best Performing Users

Userid	Defense Score	Attack Score	Win Count	Lost Count	Draw Count	User Performance
100218	326.8772	4955.199	17688	151	150	23271.08
6344024	1278.306	12549.81	7518	75	155	21576.12
3765281	3674.813	4366.809	12027	257	1026	21351.62

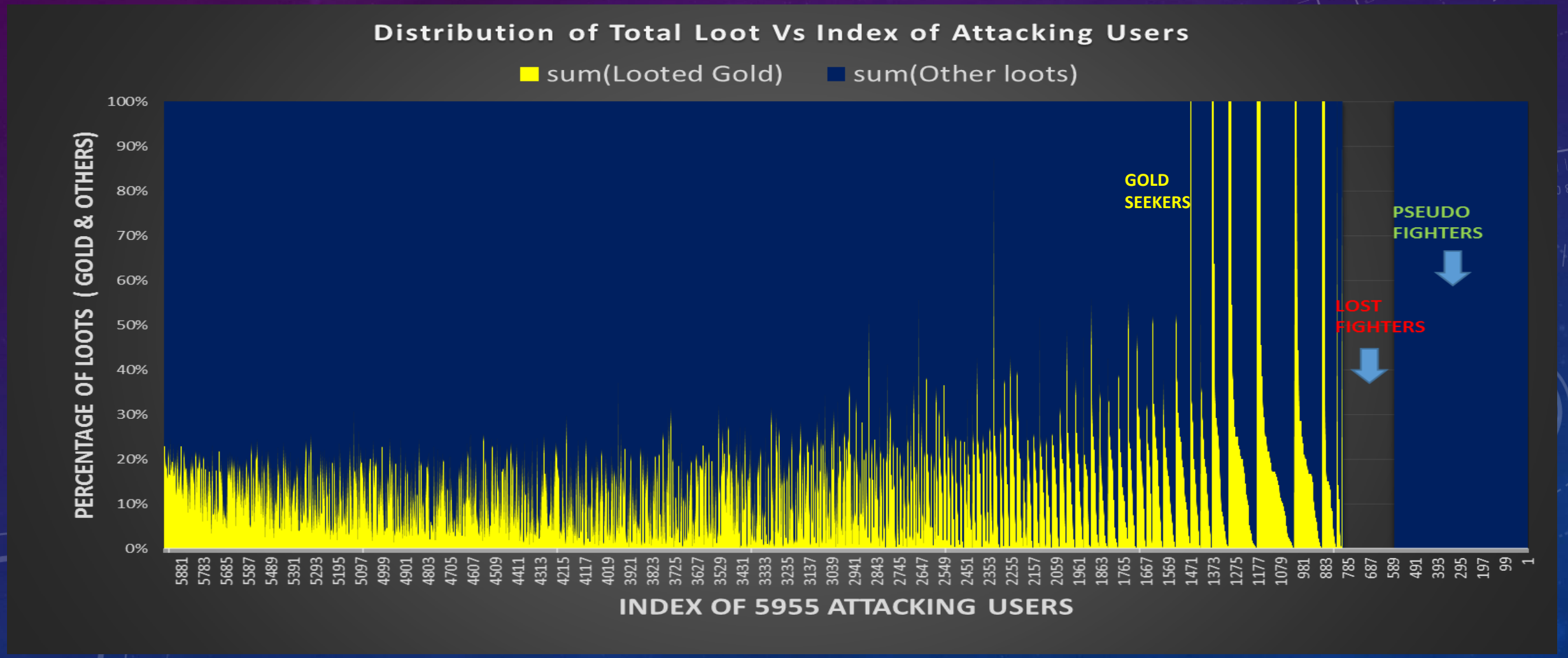
2. Best Defending Users

Userid	Defense Score	Attack Score	WinCount	LostCount	DrawCount	User Performance
1492912	15060.61	-4192.5	1391	92	1896	14247.11
25609	10325.44	1959.227	1046	335	102	13767.67
9114900	8671.873	0	0	0	**4226**	12897.87

Stationary Defenders?

I.MEASURE USER PERFORMANCE

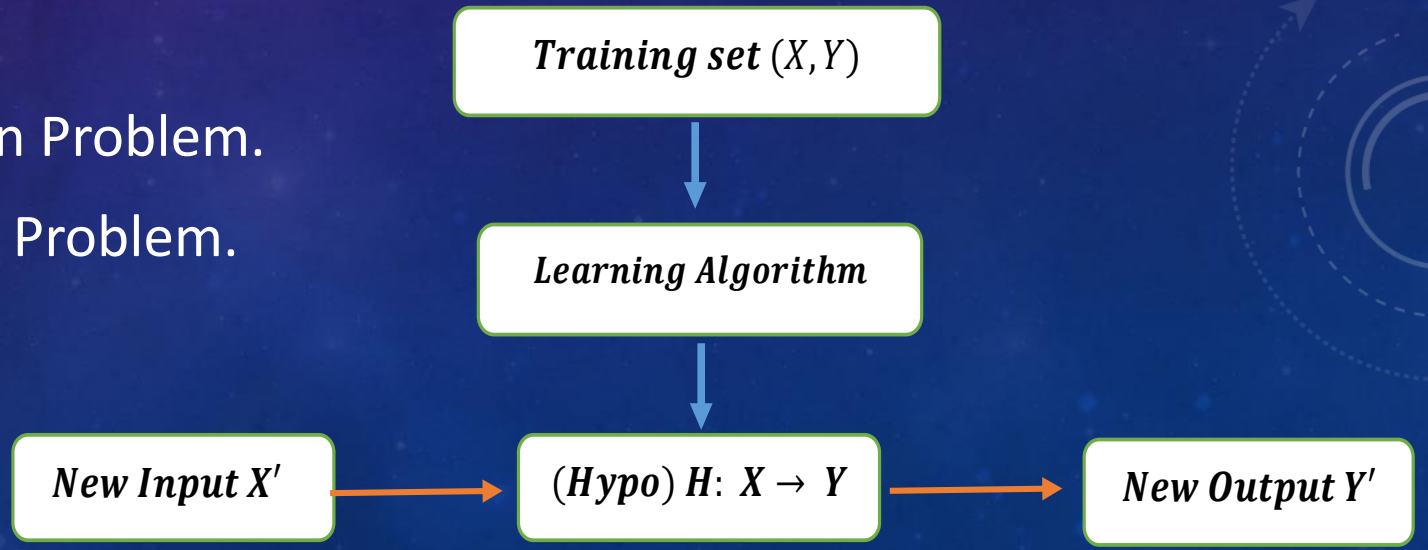
PSEUDO WARRIORS They wage wars with an intent of only collecting resources.



II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

SUPERVISED LEARNING:

- *Training Set*: Set of *m* input data, X (set of Independent variables) and its Output data, Y (Dependent variable).
- *Hypothesis/Classifier H:* A mapping function between the Input and the Output.
- *Learning Algorithm:* The Algorithm by which we obtain the hypothesis H to map the input to Output.
- Continuous Output: Regression Problem.
- Discrete Output: Classification Problem.



II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

LINEAR REGRESSION

- The hypothesis is a linear function of input X.
- *Training set* consists of m pairs of X and Y.
- X , a set of independent variables. $X = \{x_1, x_2, \dots, x_n\}$
- Y , a continuous dependent variable. $Y = \{y\}$
- $Hypothesis, \quad h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \Theta^T X$
- θ_i 's : weights updated using training set till convergence.
- Cost : squared error between Y and $h_\theta(x)$.

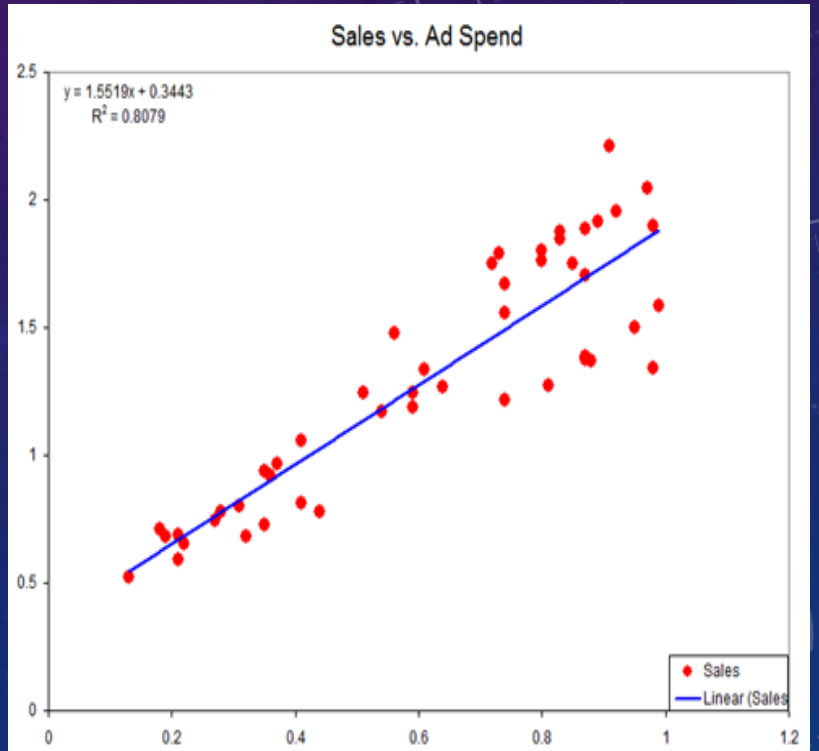
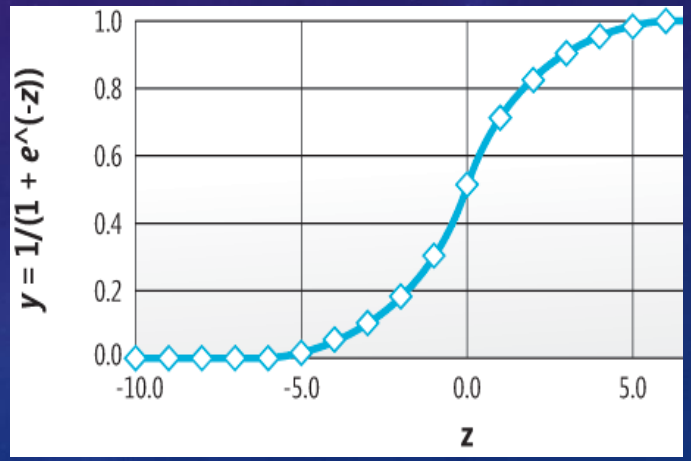


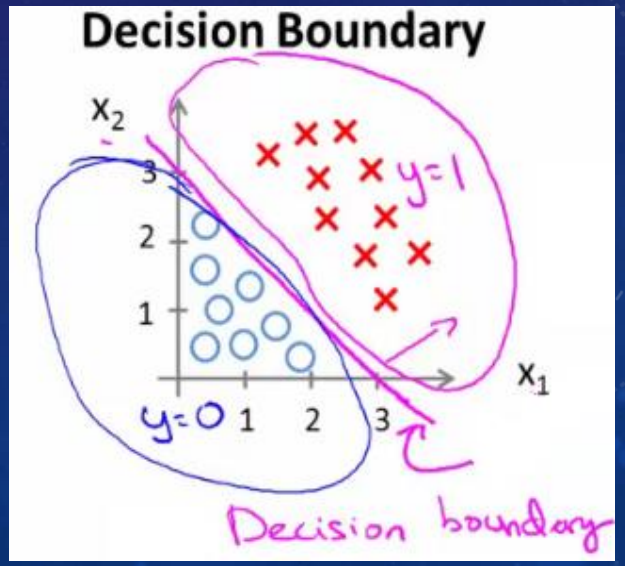
Fig 1. Linear Regression Model. We can make an intuitive assessment that increase in *Ad spend* also increases *Sales*. Using the straight line, we may also be able to predict.

II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

LOGISTIC REGRESSION

- For a binary output variable, $Y = \{0,1\}$ we use sigmoid function $g(z) = \frac{1}{1+e^{-z}}$ to smoothly transition between 0 and 1.
- The *log likelihood of the outcome* Y is modeled as a *linear combination* of the input X.
- $Logistic\ Hypothesis\ , h_\theta(x) = \frac{1}{1+e^{-\Theta^T X}}$
- $p(y = 1|X;\theta) = h_\theta(x)\ ;$
- $p(y = 0|\mathrm{X};\theta) = 1 - h_\theta(x)$
- $p(y|x;\theta) = \left(h_\theta(x)\right)^y * \left(1 - h_\theta(x)\right)^{1-y}$
- $Likelihood = P(Y|X;\theta) = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta)$
- $Cost:\ Log\ (Likelihood)$

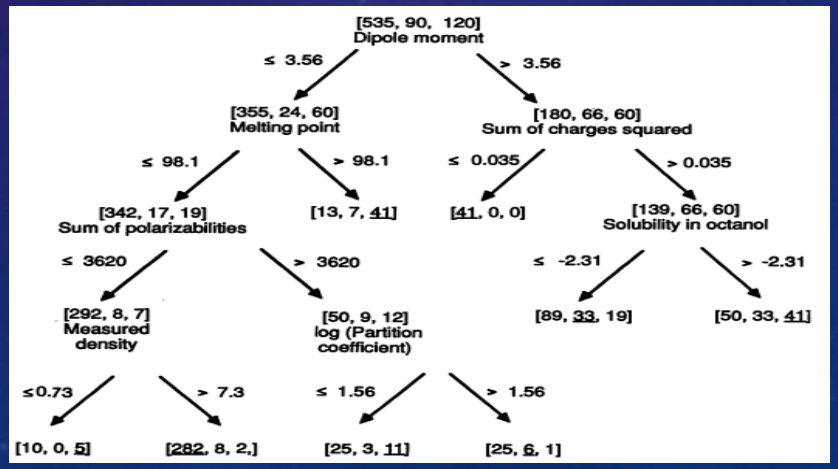


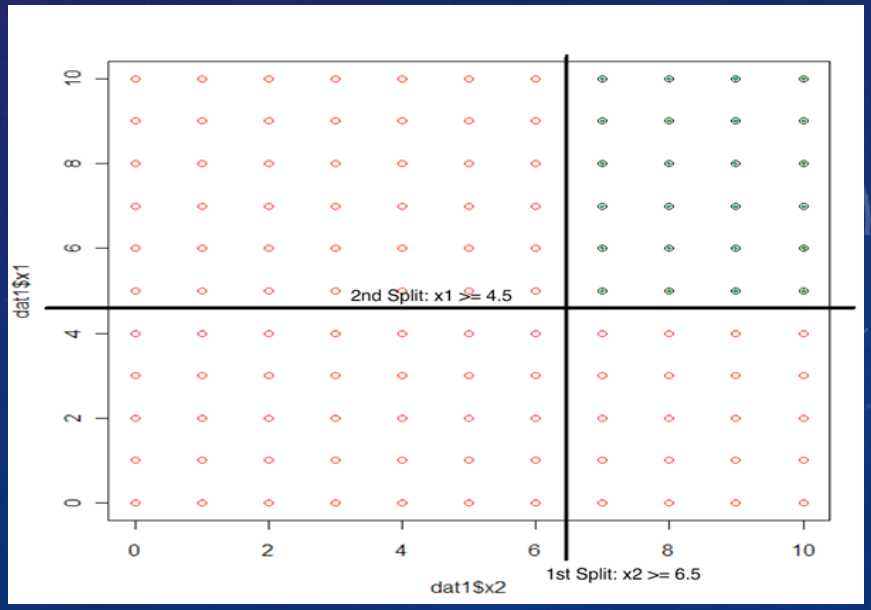


II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

DECISION TREES

- Classification tree models recursively partition the data space one variable at a time.
- Choosing the next variable to split the data is based on Information measures entropy for Recursive Partition ($rpart$) Trees in R.
- They Fit a simple prediction within each of the rectangular partitions.
- *Assumption:* decision boundaries are parallel to axes.

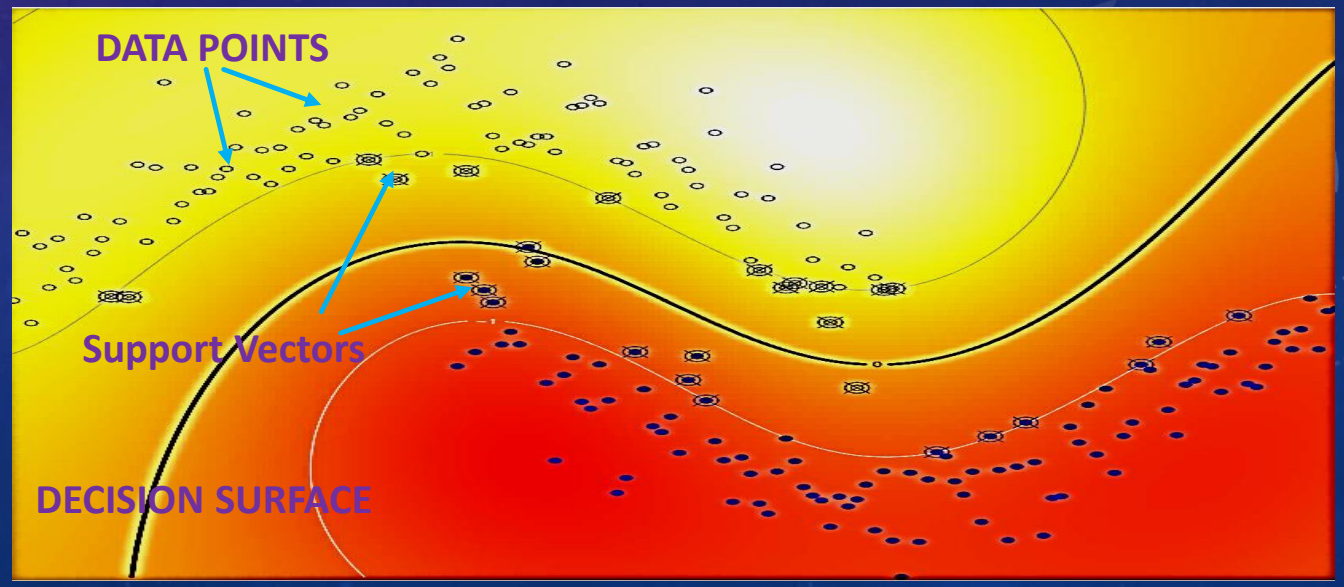




II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

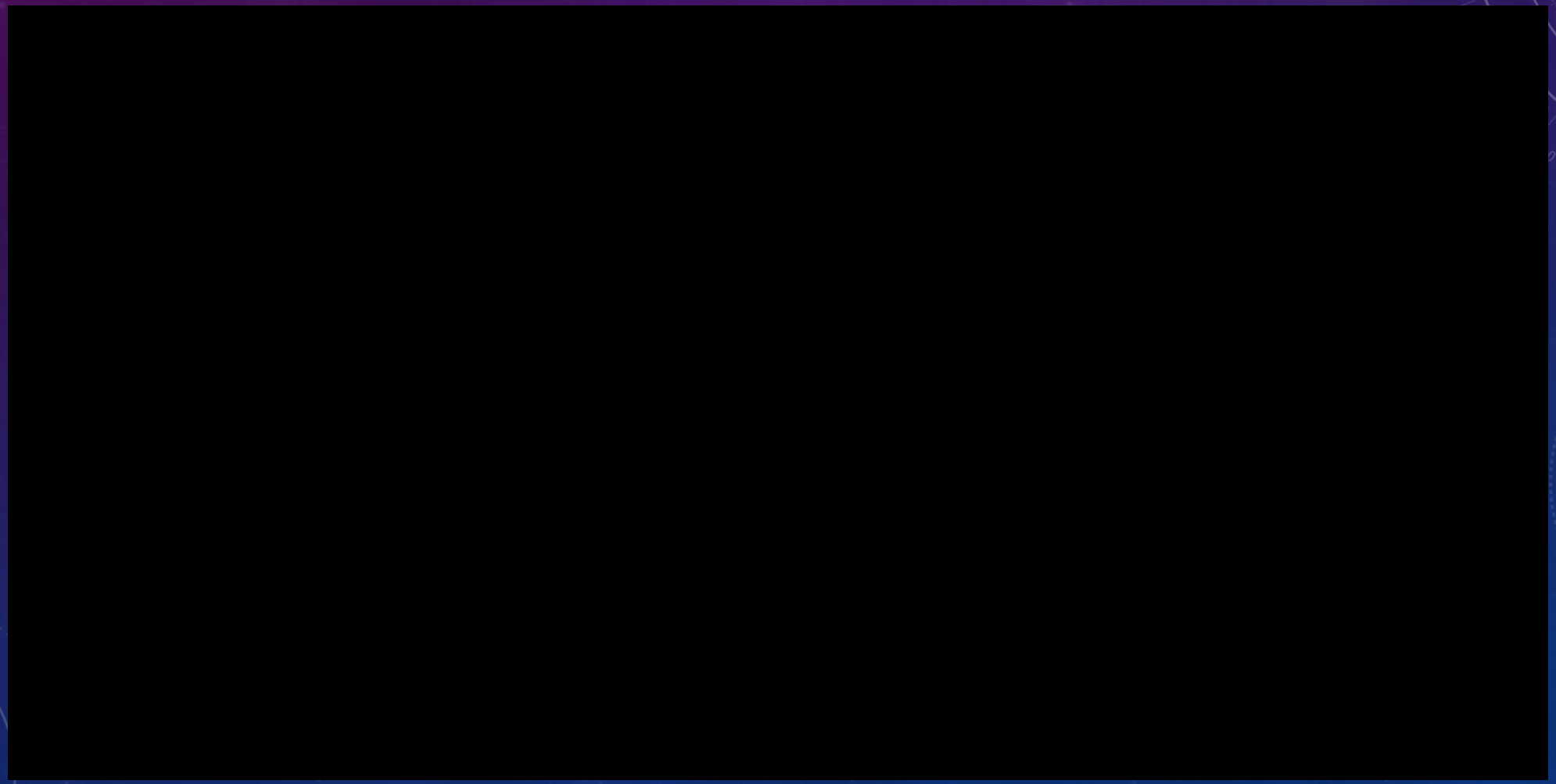
Support Vector Machines

- They separate the data using a (Hyperplane) with maximum *Margin*.
- Margin is the closest distance between data points & decision surface.
- The decision surface is specified by a subset of the data, the *Support Vectors*.
- They use Kernels to map the data into higher dimension to make it linearly separable.



II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

Support Vector Machines *kernels*



II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

FEATURES SET

- *User Performance* : User scores, Match Results, Alliance association, match frequencies.
- *Alliance Performance* : Alliance Scores, Match results.
- *Knight Performance* : Knight Scores, Match results, match frequencies.
- *User Social* : User Sentiments, User subjectivity, chat counts.
- *User Delight* : Loots, Boosts, Mights, Wall Strengths & XPs.

TRAINING DATA: Feature Set of 4000 Users from total of 5387 User data.

TEST DATA: Feature Set of 1387 User Data.

Number of Features: ~44

II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

FEATURES SET

- *User Performance* : User scores, Match Results, Alliance association, match frequencies.
- *Alliance Performance* : Alliance Scores, Match results.
- *Knight Performance* : Knight Scores, Match results, match frequencies.
- *User Social* : User Sentiments, User subjectivity, chat counts.
- *User Delight* : Loots, Boosts, Mights, Wall Strengths & XPs.

TRAINING DATA: Feature Set of 4000 Users from total of 5387 User data.

TEST DATA: Feature Set of 1387 User Data.

Number of Features: ~44

II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

Metrics Of Evaluation:

- *Precision*: Ratio of correctly classified positives to number of items classified as “positive”.
- *Recall or True Positive rate*: Ratio of correctly classified positives to total number of positives.
- *Accuracy*: Ratio of correctly classified items(+ve and –ve) to the total number of items.
- *Specificity*: Ratio of correctly classified negatives to total number of negatives.
- *Fall Out or False Positive rate or :* Ratio of correctly classified negatives to number of items classified as “negative”.
- *Mutual Info:* Amount of information shared between the test truth & test predictions.
- *F1 Score:* Weighted average of Precision and Recall.
- *Area Under the ROC Curve:* Area of region under the curve between Recall and Fall Out.

II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

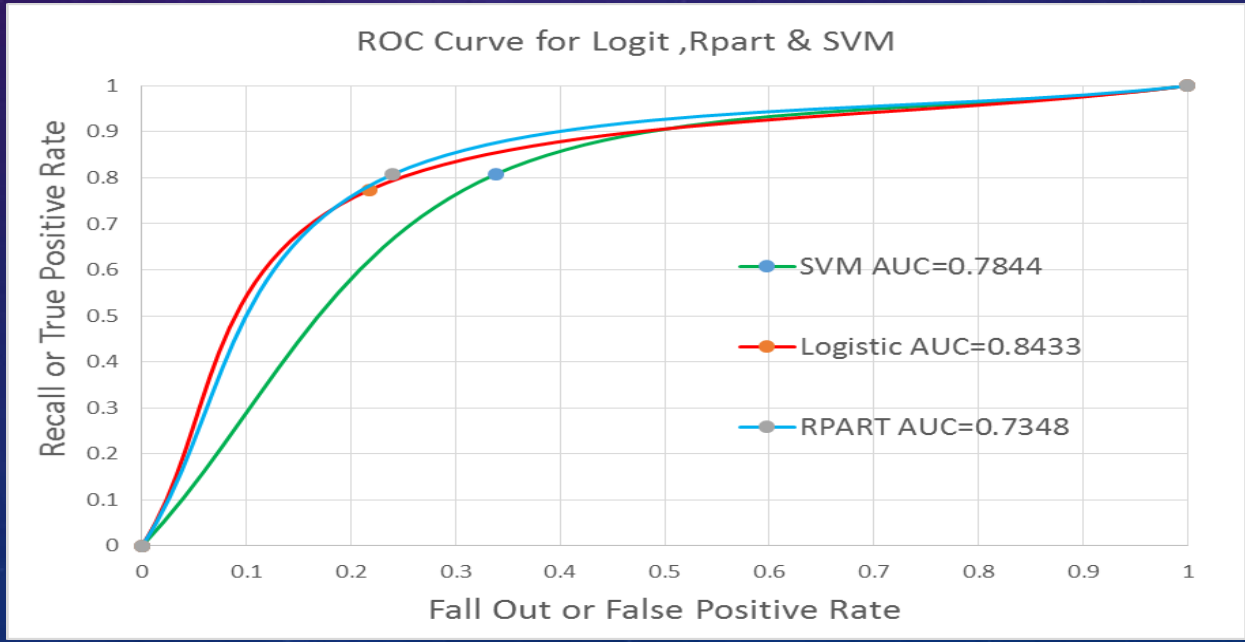
RESULTS

Metric	Formula	Logistic Regression	Recursive Partition	Support Vector Machines
Precision	TP/(TP+FP)	85.94%	***85.33%***	81.45%
True Positive Rate or Recall	TP/(TP+FN)	77.44%	***80.86%***	89.06%
Accuracy	$TP + TN/(TP + TN + FP + FN)$	77.72%	***79.09%***	80.24%
F1 Score	$\frac{2 * Precison * Recall}{Precision + Recall}$	81.015%	***83.034%***	85.085%
Specificity	$TN/(FP + TN)$	71.19%	***76.03%***	65.03%
False Positive Rate or Fall-Out	$FP/(FP + TN)$	28.80%	***23.96%***	34.97%

II. CLASSIFICATION OF USERS (SPENDERS OR NOT)

RESULTS

Metric	Formula	Logistic Regression	Recursive Partition	Support Vector Machines
Mutual Information (Test Prediction , Test truth)	$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right),$	0.1528	***0.1619***	0.1113
Area Under the Curve	Area under the ROC curve	0.8433	***0.7844***	0.7347



III. EXPENDITURE VS PERFORMANCE & SOCIAL INTERACTION

AIM

To study the Performances and Social Interaction of high spending users in the Game of Kingdom of Camelot.

PURPOSE

- The Gamers who spend money on games are our VIP customers.
- Are they struggling to have a feel-good experience while gaming ?
- Do they need more training? more boosts?
- Are they feeling bored being the best and are not challenged enough to be excited?
- By Analyzing the alliance chats, We can get their impressions, ideas, likes and dislikes of our game.
- 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?

Where, X is the mean of X, Y is the mean of Y and n the number of items

III. EXPENDITURE VS PERFORMANCE & SOCIAL INTERACTION

Entropy

Entropy is a measure of uncertainty of a random variable. It is a measure of the amount of information required on the average to describe the random variable. $p(x)$ is the probability mass function of X.

$$H(X) = -\sum_{x \in X} p(x) \log p(x)$$

Relative Entropy

The relative entropy is a non symmetric measure of the distance between two probability distributions. The relative entropy $D(p||q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p.

$$D(p||q) = \sum_{x \in X} p(x) \log \left(\frac{p(x)}{q(x)}\right)$$

Mutual Information

Consider two random variables X and Y with a joint probability distribution $p(x, y)$ and the marginal probability $p(x)$ and $p(y)$. The Mutual Information $I(X, Y)$ is the relative entropy between the joint probability distribution and the product distribution $p(x)p(y)$.

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

Where, X is the mean of X_i, Y is the mean of Y_i and n the number of items

III. EXPENDITURE VS PERFORMANCE & SOCIAL INTERACTION

Correlation Coefficient (Pearson's *r*)

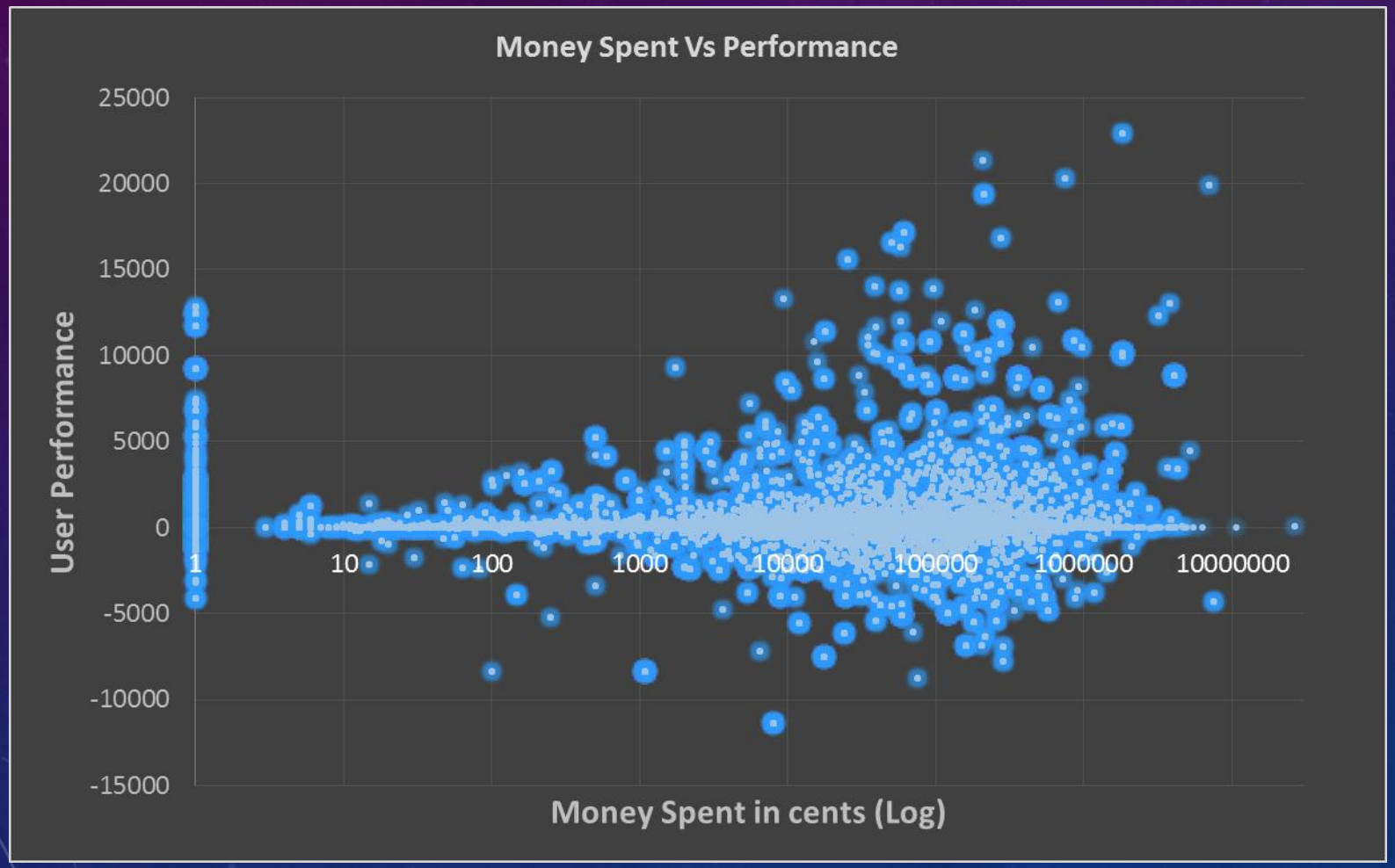
It is a measure of the strength and direction of the linear relationship between Random variables(X *and* Y). The values lies between +1 (positive correlation) and −1 (negative correlation). Value 0 is no linear correlation. $\bar{X}$ is the mean of X_i. $\bar{Y}$ is the mean of Y_i *and* n, the number of items.

$$Pearson\ r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

2 sample Kolgomorov Smirnoff Test

- It is a nonparametric test that compares the cumulative distributions of two data sets to check if the samples are from same distribution. (Null Hypo: = ,Alternate Hypo: ≠)
- $Test\ Statistic\ D = \max_{1<i<N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right)$
- *P value:* **p-value** is the probability of obtaining a test statistic result at least as extreme or as close to the one that was actually observed, assuming that the null hypothesis is true

III. EXPENDITURE VS PERFORMANCE & SOCIAL INTERACTION



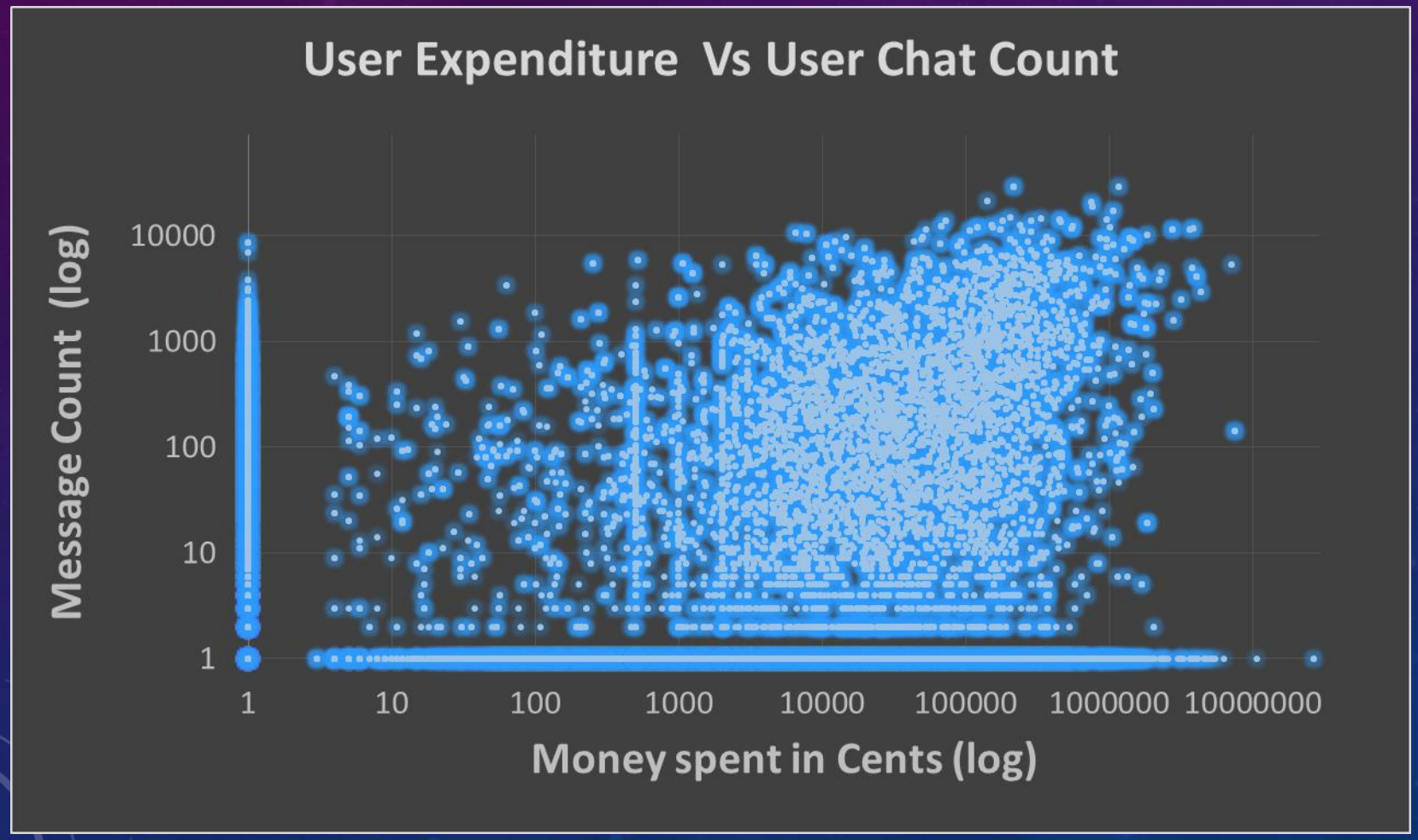
Correlation Coefficient
0.11859842

Kolgomorov-Smirnoff Test
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
3.72616

RESULT
Weak relationship between Expenditure and Performance of the User

III. EXPENDITURE VS PERFORMANCE & SOCIAL INTERACTION



Correlation Coefficient
0.23367334

Kolgomorov-Smirnoff Test
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
0.88868

RESULT
Weak relationship between the Expenditure and the Social interaction of the User.

IV. TEXT PROCESSING OF ALLIANCE CHATS

AIM: Text processing of Alliance chats.

PURPOSE:

- Understanding User's messages gives essence of their game play strategies, idea, goals.
- It gives us their impressions on the game, on the alliances, on each other.
- We can see the sentiments of the User, Alliances and understand who likes and is happy about the games and who would is not happy.
- We can classify the Users geographically based on the language of communication.
- We can correlate the actions taken by the User to the communication they had.
- We can understand the Organization of the Alliances and their activities.
- We can recommend happy teams/alliances for an unhappy Users.
- We can understand the subjectivity of their language. This is useful to understand their personality. People who are more objective vs people who are more subjective.

IV. TEXT PROCESSING OF ALLIANCE CHATS

LINGO

- Users tend to use short form English words called as Lingo.
- "***peopl clse 2 my hart alw get D 1st msg 4 D dy.. gud mornin*** " which means "People close to my heart always get the first message for the day. Good Morning"
- Words from http://transl8it.com/largest-best-top-text-message-list/

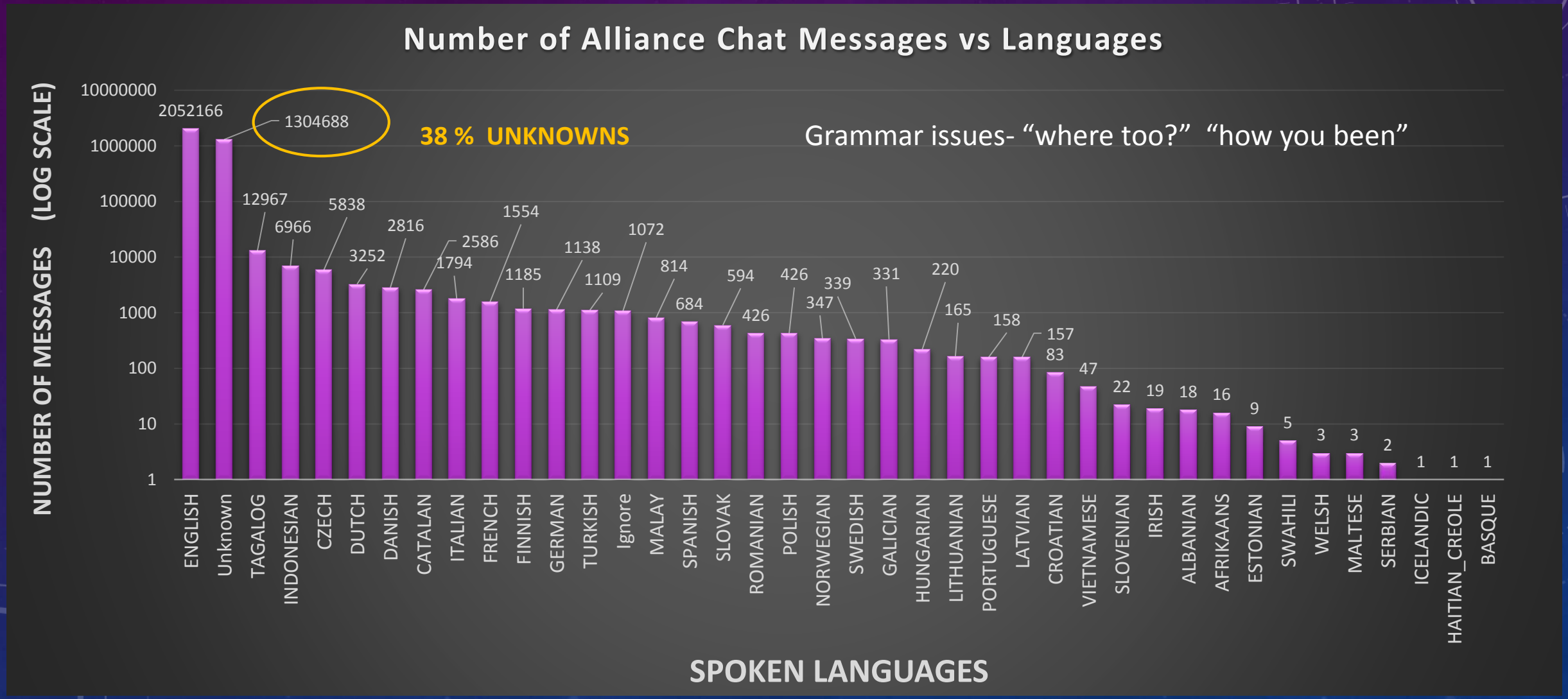
CAPTURING EMOTICONS

- Shorter symbols express more meaning and context.
- The messages are selectively parsed for the emoticons such as **: -)**
- The Challenge: Lingo and emoticons evolve over time. Symbol " **~:** " . This means an **Elvis**

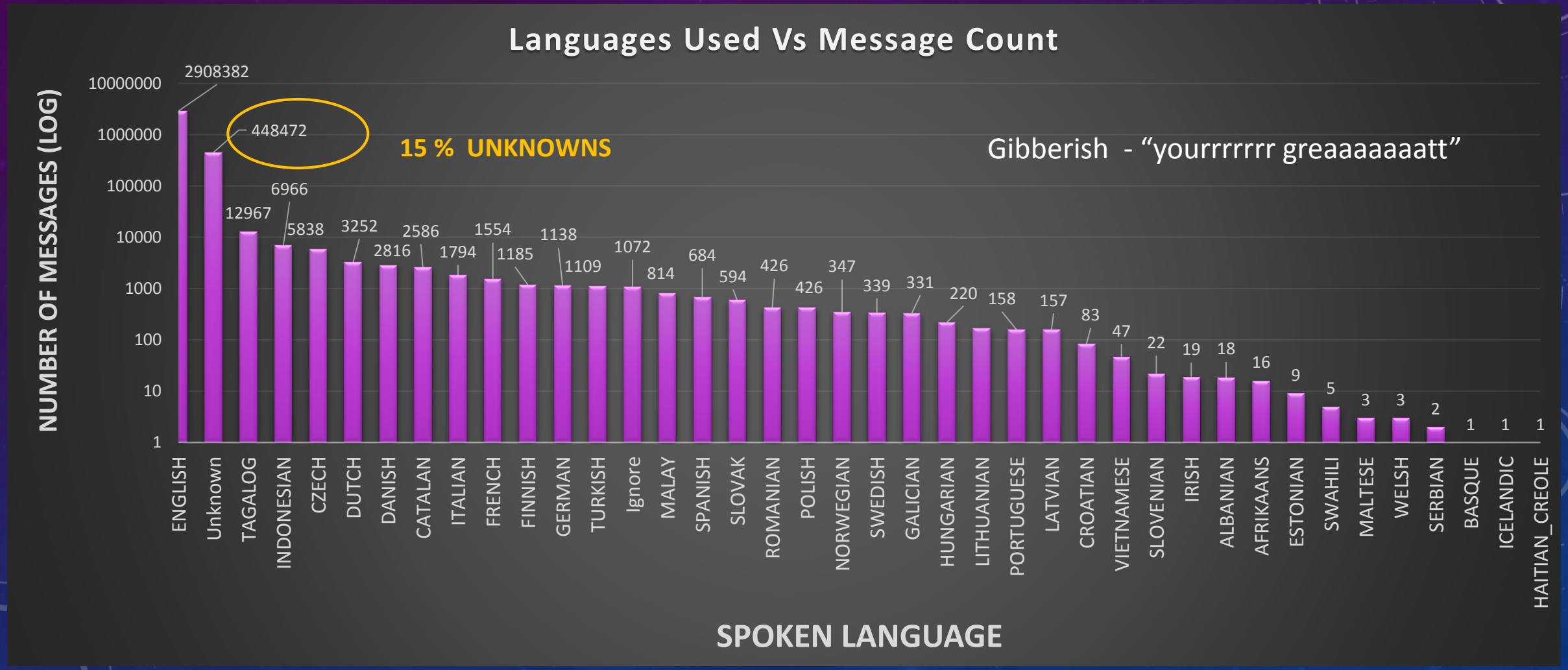
DETECT LANGUAGE

- Use ***Chromium compact language detector*** from Google package for python.
- Gramar issues- "where too?" "stupid keyboard" "how you been"
- Gibberish - "yourrrrrrr greaaaaaaatt"
- Parts of Speech Tagging.
- Using **TextBlob** Package in python.

IV. TEXT PROCESSING OF ALLIANCE CHATS



IV. TEXT PROCESSING OF ALLIANCE CHATS



IV. TEXT PROCESSING OF ALLIANCE CHATS

SENTIMENT ANALYSIS

- A basic task in ***sentiment analysis*** is classifying the ***polarity*** of a given text at the document, sentence, or feature/aspect level.
- Polarity is a real value between +1(positive) ,0(neutral) & -1(negative).
- **TextBlob** package in Python.
- Uses a Native Bayes Classifier for Sentiments.

SUBJECTIVITY ANALYSIS

- The **linguistic** expression of somebody's **opinions**, **sentiments**, emotions, evaluations, beliefs, speculations *(private states which are personal to themselves).*
- The following *"**I love my food**" has a positive subjectivity while. "**sunny weather**" has* zero subjectivity.
- ***TextBlob*** package in python.

IV.TEXT PROCESSING ALLIANCE CHATS

$$User\ delight\ =\ abslog(sum\ of\ User\ Sentiment\ *\ sum\ of\ subjectivity * count\ of\ comments)$$

Abslog finds –log ||num|| for -ve numbers; Log (num+1) for others.

1. Happiest Users

UserID	Sentiment	subjectivity	count	Delight
1493831	2508.977	10743.93	29136	11.895
578124	2148.754	7827.686	21083	11.549
3765281	2309.957	6930.6	20577	11.517

2. Happiest Alliances

AllianceID	sum(SI.sentiment)	sum(SI.subjectivity)	count	Delight
21114|2	9831.675	58812.22	169784	13.991
22587|21	10858.63	39995.48	136743	13.773
8647|2	9369.567	39571.57	144412	13.728

IV.TEXT PROCESSING ALLIANCE CHATS

$$Feel\ Good\ = Win - Lost + abslog(\ Sentiment + \ sum\ of\ subjectivity + \ count\ of\ comments)$$

1. Happiest Users (Feel Good)

User id	Subjectivity	Sentiment	Count	Win Count	Lost Count	FEEL GOOD
100218	202.3972	46.14497	755	17688	151	**17540.69**
3765281	5627.079	2071.066	21567	12027	257	**11775.16**
1888783	266.7217	117.4101	981	8069	202	**7870.94**

2. Disappointed Users (Feel Good)

User id	Subjectivity	Sentiment	Count	Win Count	Lost Count	FEEL GOOD
8648066	144.0611	42.18857	1211	19	574	**-551.16**
2773733	27.03306	8.68584	194	378	760	**-378.94**
1098040	401.2151	88.81813	1495	40	324	**-280.00**

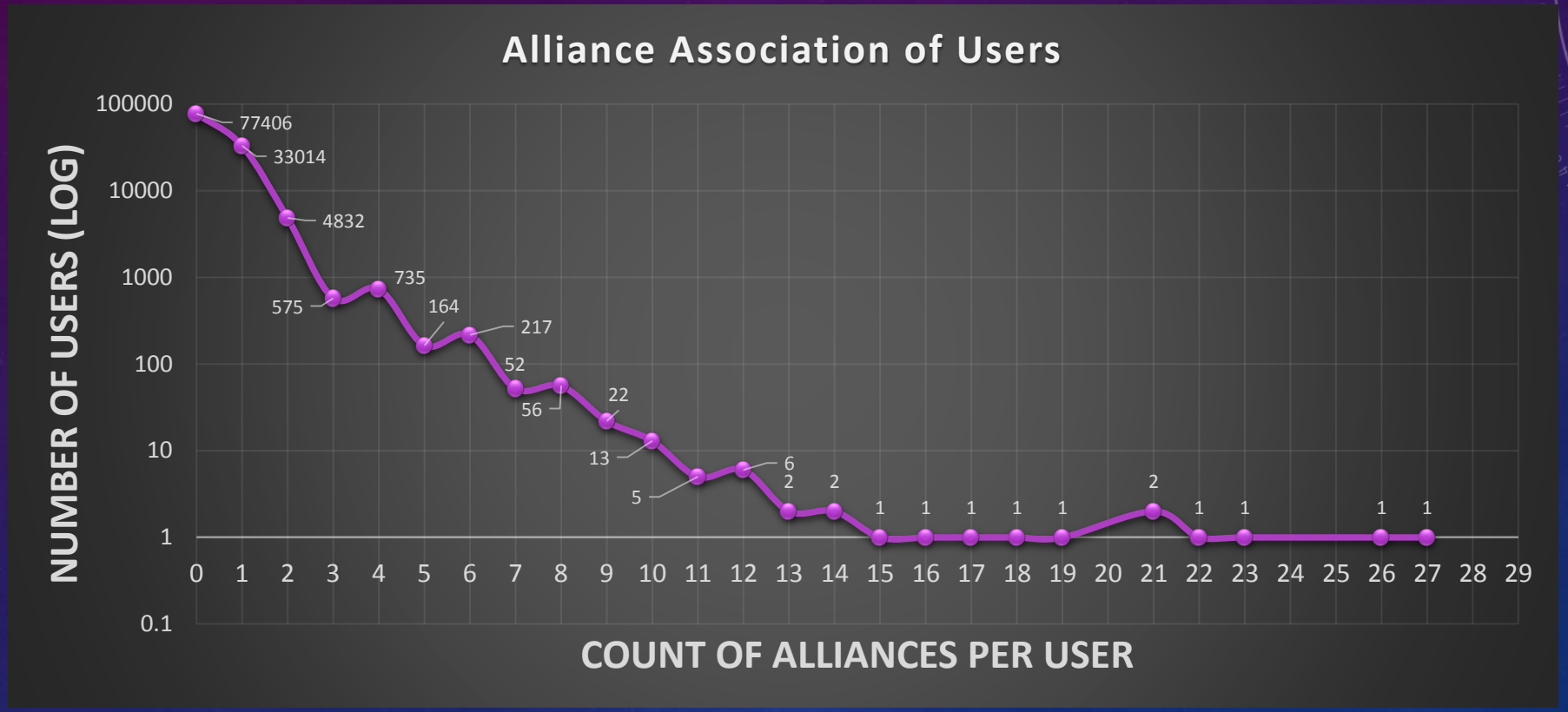
V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS

- **AIM:** We want to understand the How users make their choices in the Alliances, Knights and Organized politics .

PURPOSE:

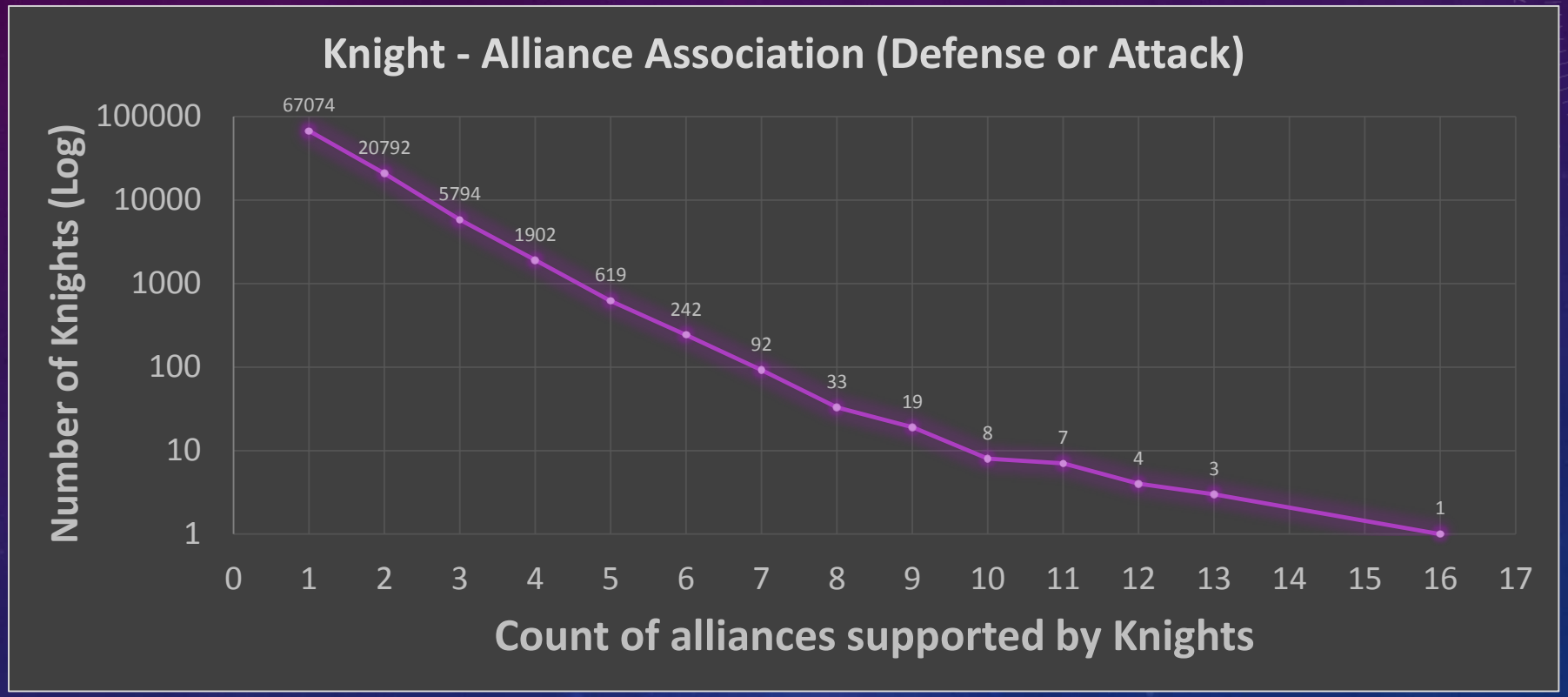
- Understanding User's choices on Alliances, gives us their user behavior on what they prefer to do.
- It helps understand their strategies and customization to prepare for the war.
- We can get to know them better and helps us understand their decision criteria.
- We can take advantage using their behavior and apply predictive analytics.
- Observing the level of partnerships between the Alliances helps us to create partner alert system which could alert trouble /victory for the partner alliances.
- We can recommend Knights for the Troops based on the Knights past performance.
- We can advise the Alliance to be more cohesive rather than adhesive or vice versa.
- We could recommend new Friends/partners to a common enemy.

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS



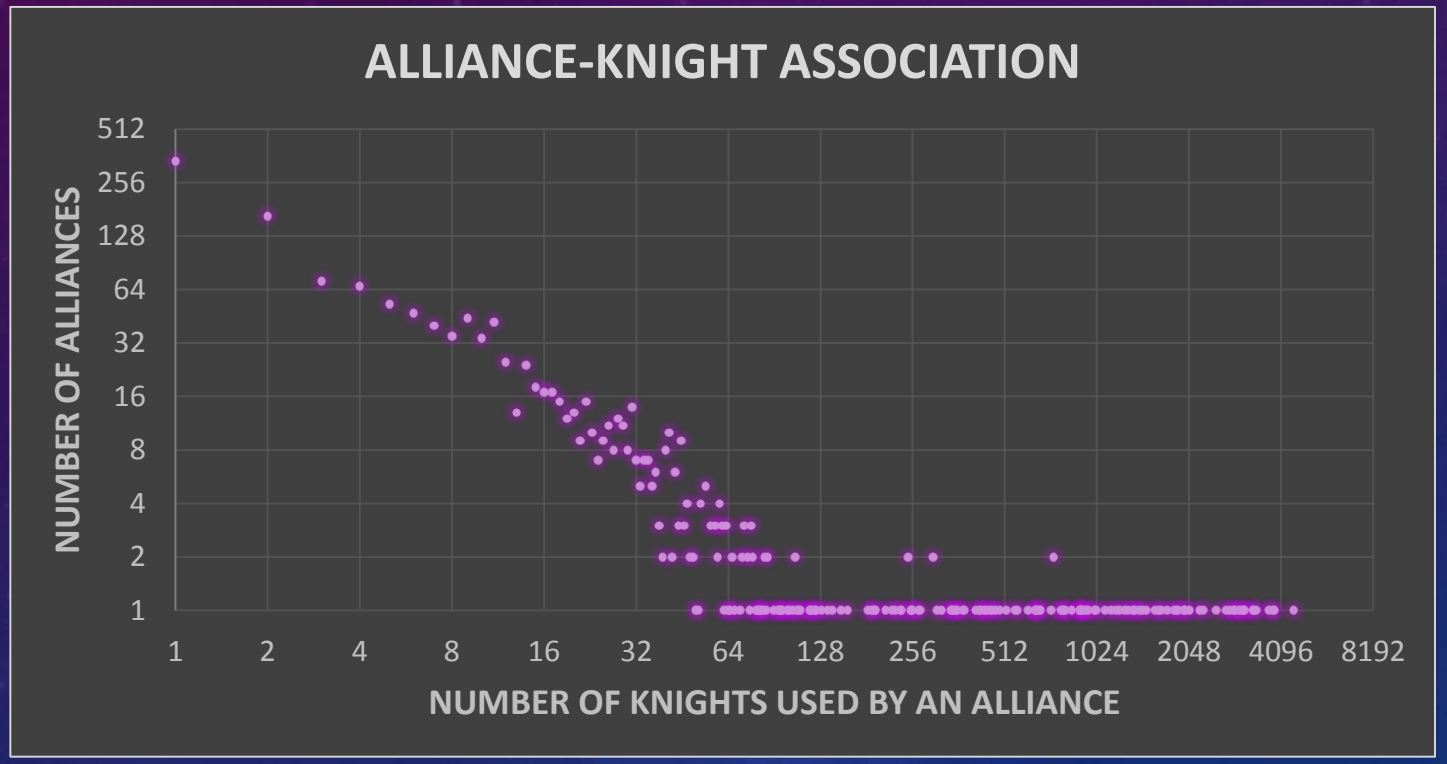
Number of Alliance associated	None	One	Two	>2
Percentage of Users	66%	28%	4%	0.5%

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS



Number of Alliances supported	One	Two	Three	> Three
Percentage of Knights	69.44%	21%	6%	3.56%

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS



Number of Knights Equipped	One	Two	Three	> Three
Percentage of Alliances	69.44%	21%	6%	3.56%

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS

INSIGHTS ON RIVALRY:

- *No fights amongst the members of the same alliance*.
- Users and Alliances have their rival and archenemies with whom they keep fighting again and again.

No	S0UserID	S1UserID	Match Count
1	25112	2486380	19256
2	21536	6745138	2691
3	9114900	1508200	2114

No	S0AllianceID	S1AllianceID	Match count
1	336|1	20118|1	20230
2	8647|2	21114|2	5103
3	336|1	516|1	4558

- ***Match Marathons:*** Relentless fighting on a single day till their moms shout at them.

No	S0Userid	S1UserID	Date	Match Count
1	25112	2486380	10/26/2013	19152
2	7452322	1458478	11/1/2013	785
3	5573073	7384945	11/15/2013	658

No	S0Alliance	S1Alliance	Date	Match Count
1	336|1	20118|1	10/26/2013	19160
2	19527|21	15805|21	12/15/2013	752
3	22211|21	23227|21	11/15/2013	702

- We should celebrate the valor of these fighters with Rival leaderboards and trophies.

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS

OBSERVATIONS ON KNIGHTS:

Presence or Absence of Knights:

- Is a knight necessary?

Categories	Match Side	Matches with No Knights	Matches with Knights	% of Matches with No Knights	% of Matches with Knights
Total Matches	Defending	2929282	123014	96%	4%
	Attacking	21	3052275	99.99%	0.00%

Internal or External Knights:

- how the users prefer a Knight from the same alliance to a Knight from a different alliance.

Categories	Match Side	Matches with Same Alliance Knight	Matches with External Alliance Knight	% of Matches with Internal Knight	% of Matches with External Knight
Total Matches	Defending	119341	3673	97.01%	2.98%
	Attacking	3047421	4854	99.84%	0.15%

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS

GEO POLITICS ON ALLIANCES:

Sharing of Knights: Friendship. The more the knights shared the lesser the distance between them.

For alliance A and alliance B,

$$Distance(A,B)= \frac{1}{1+Number\ of\ knights\ shared\ between\ A\ \&\ B}$$

Categories	Match Side	Clustering Method	Number of clusters
Alliance	Defending	K means	12
	Attacking	K means	9

V. LOCATION VS PERFORMANCE

AIM:

To study the impact of the Location in the performance of Users.

PURPOSE:

- The Location in the Game of Kingdom of Camelot is chosen by the gamers to build their city
- Location choice is strategically made. It needs to be in a suitable position favorable for both attack and defense.
- It should have the right set of neighbors friends or enemies.
- We can perhaps suggest a suitable location for future gamers via recommendation.

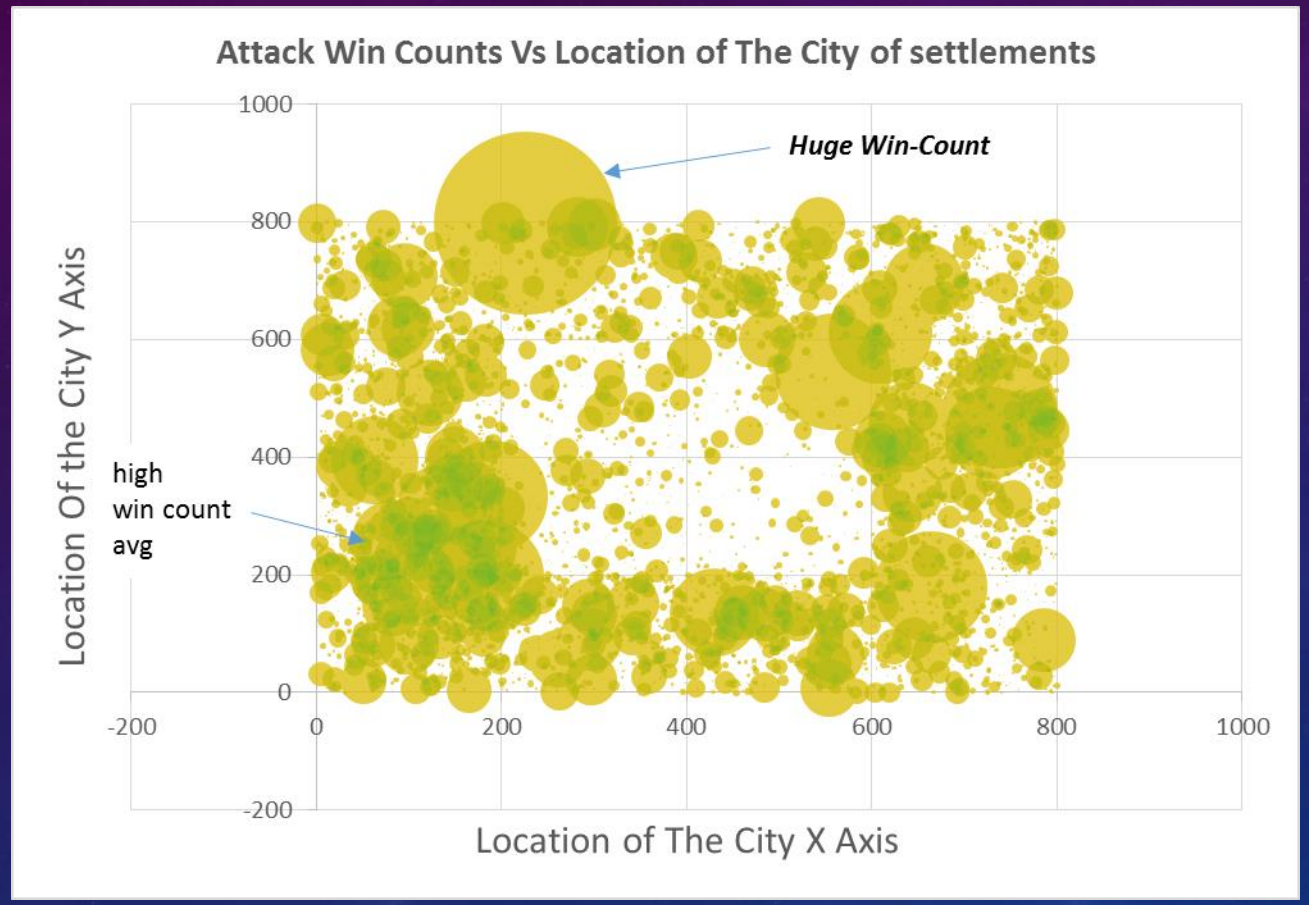
Metrics Used Location & Performance Scores

1.Correlation Coefficient .

2.Mutual Information.

3.Kolmogorov Test of similarity in the distributions.

V. LOCATION VS PERFORMANCE –WIN COUNTS



Average	Stddev	Max	Min
7.773	72.788	6747	0

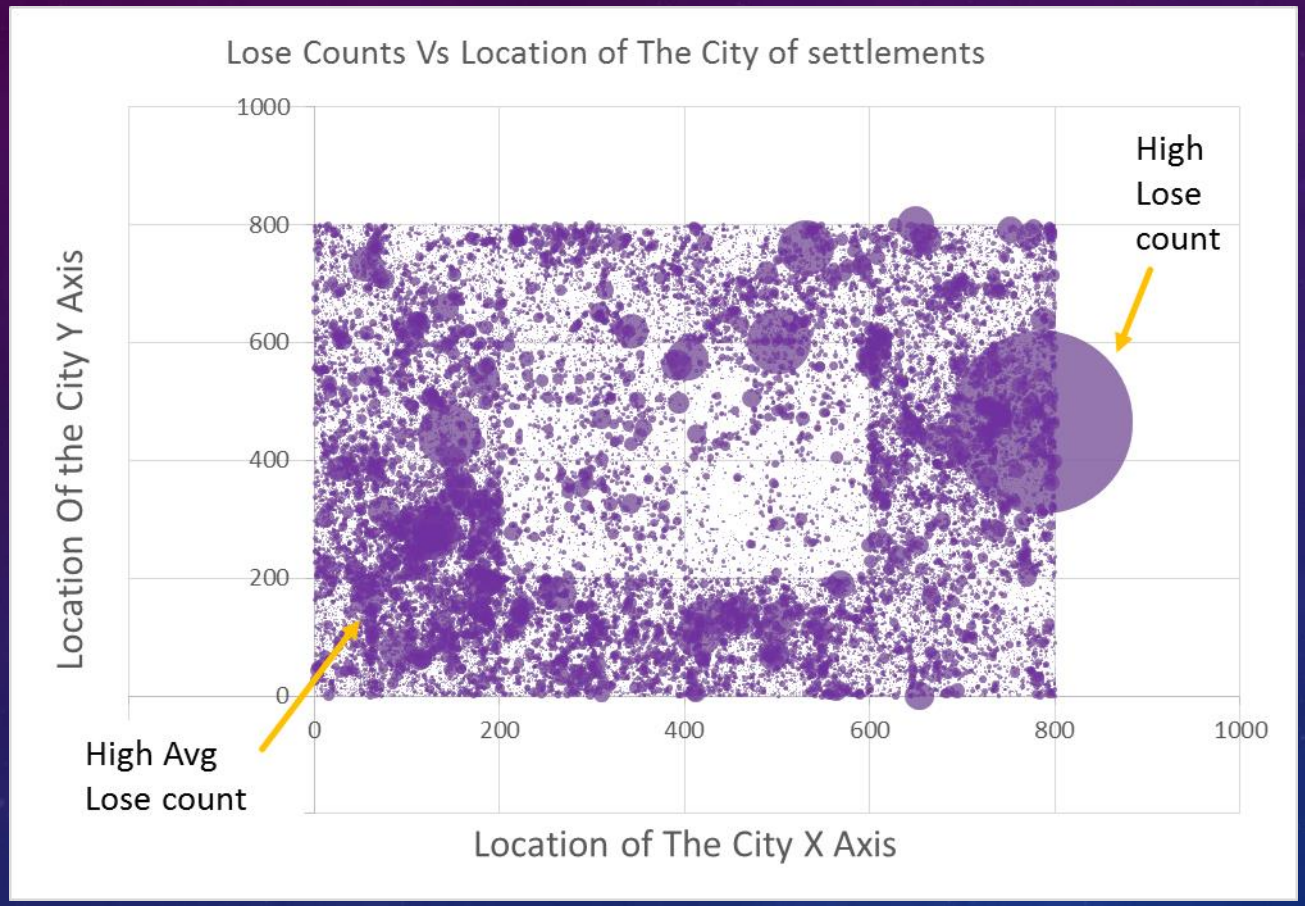
Correlation Coefficient
-0.00053

Kolgomorov-Smirnoff Test
D=0.9985
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
0.959954

RESULT
Weak relationship between Win count and Location of the User

V. LOCATION VS PERFORMANCE-LOSE COUNTS



Average	Stddev	Max	Min
7.773	14.449	1588	0

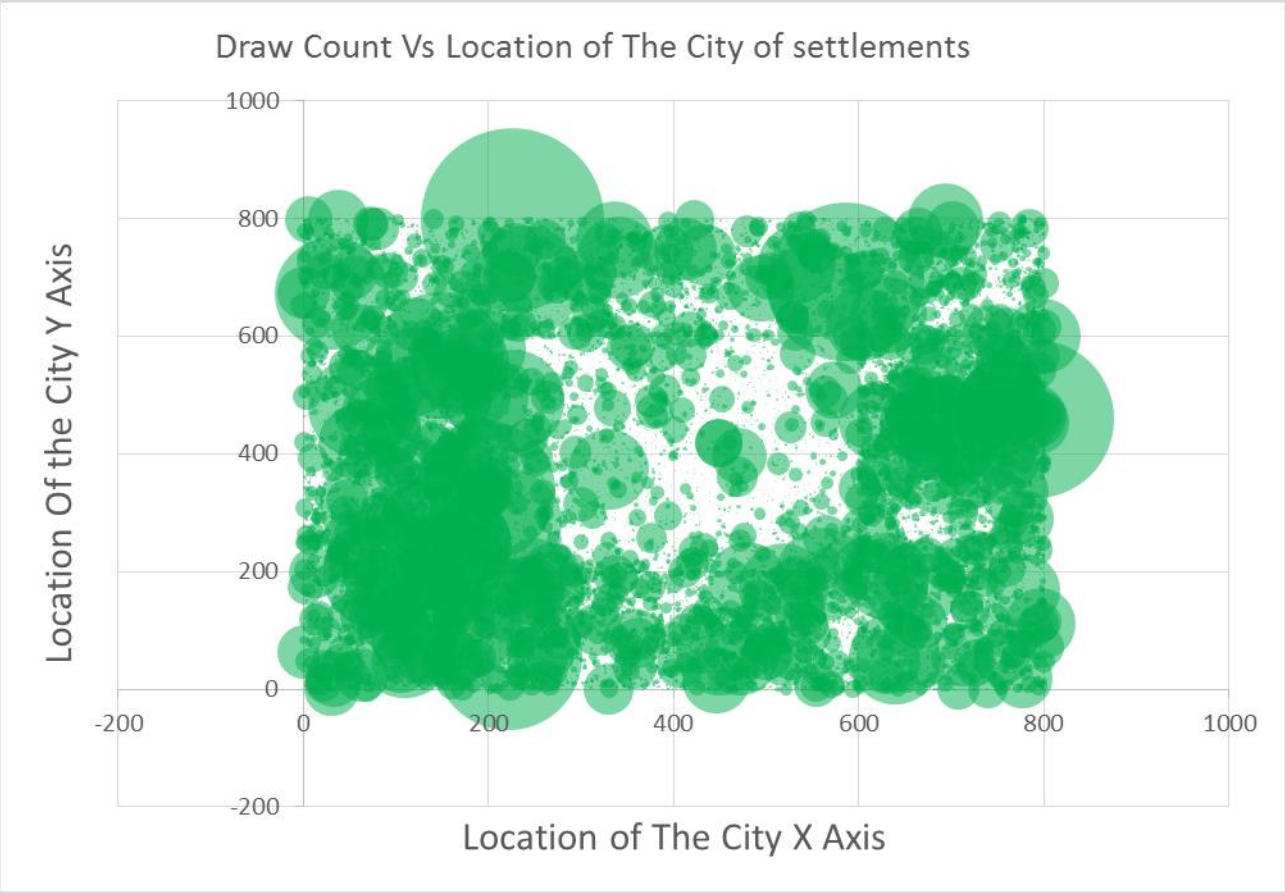
Correlation Coefficient
0.00884

Kolgomorov-Smirnoff Test
D=1
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
2.934839

RESULT
Weak relationship between Lose count and Location of the User

V. LOCATION VS PERFORMANCE



Average	Stddev	Max	Min
19.922	92.928	5537	0

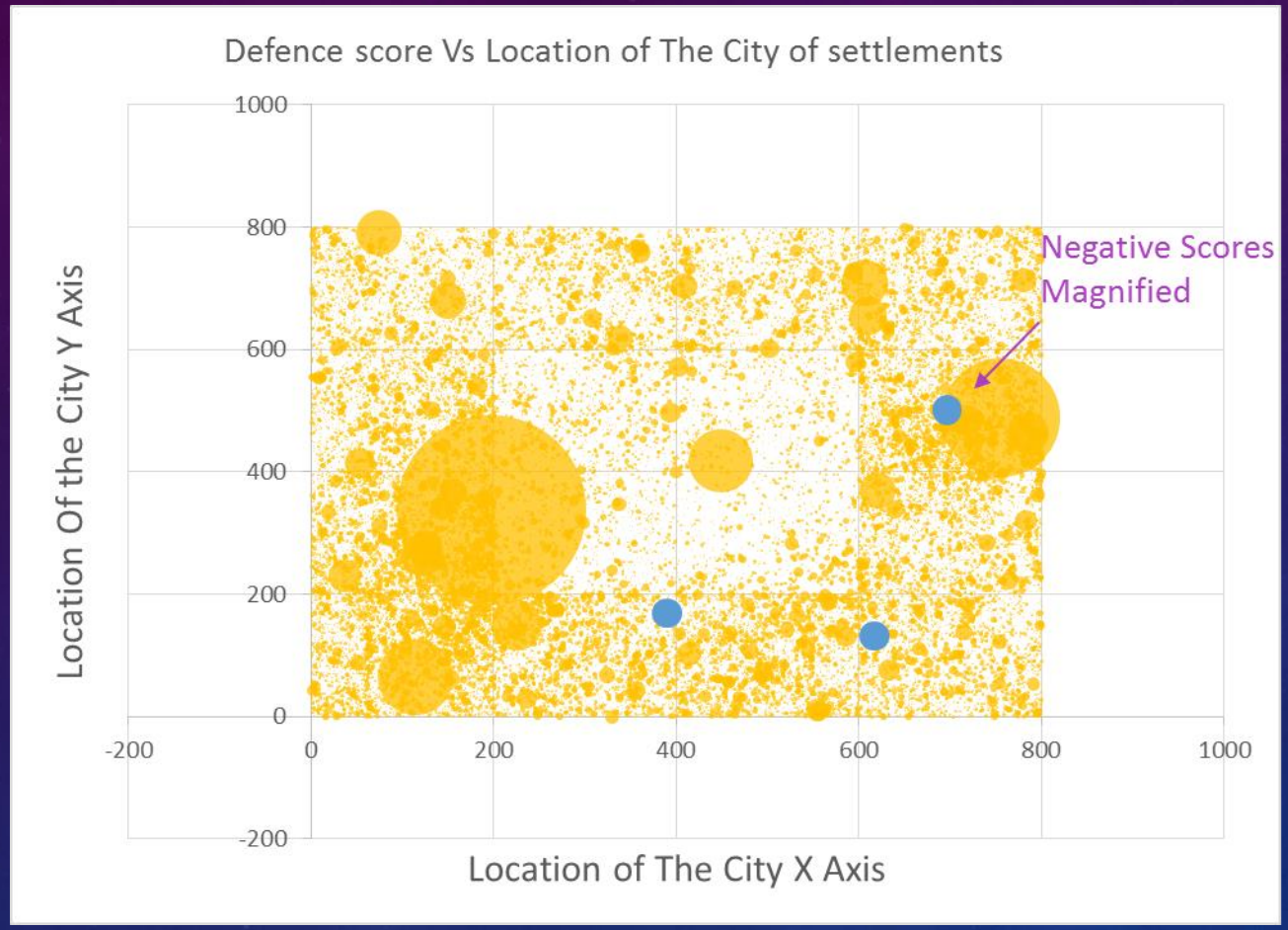
Correlation Coefficient
0. 01043

Kolgomorov-Smirnoff Test
D=0.9977
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
3.16612

RESULT
Weak relationship between Draw Count and Location of the User

V. LOCATION VS PERFORMANCE



Average	Stddev	Max	Min
20.658	42.603	5912.72	-2.137

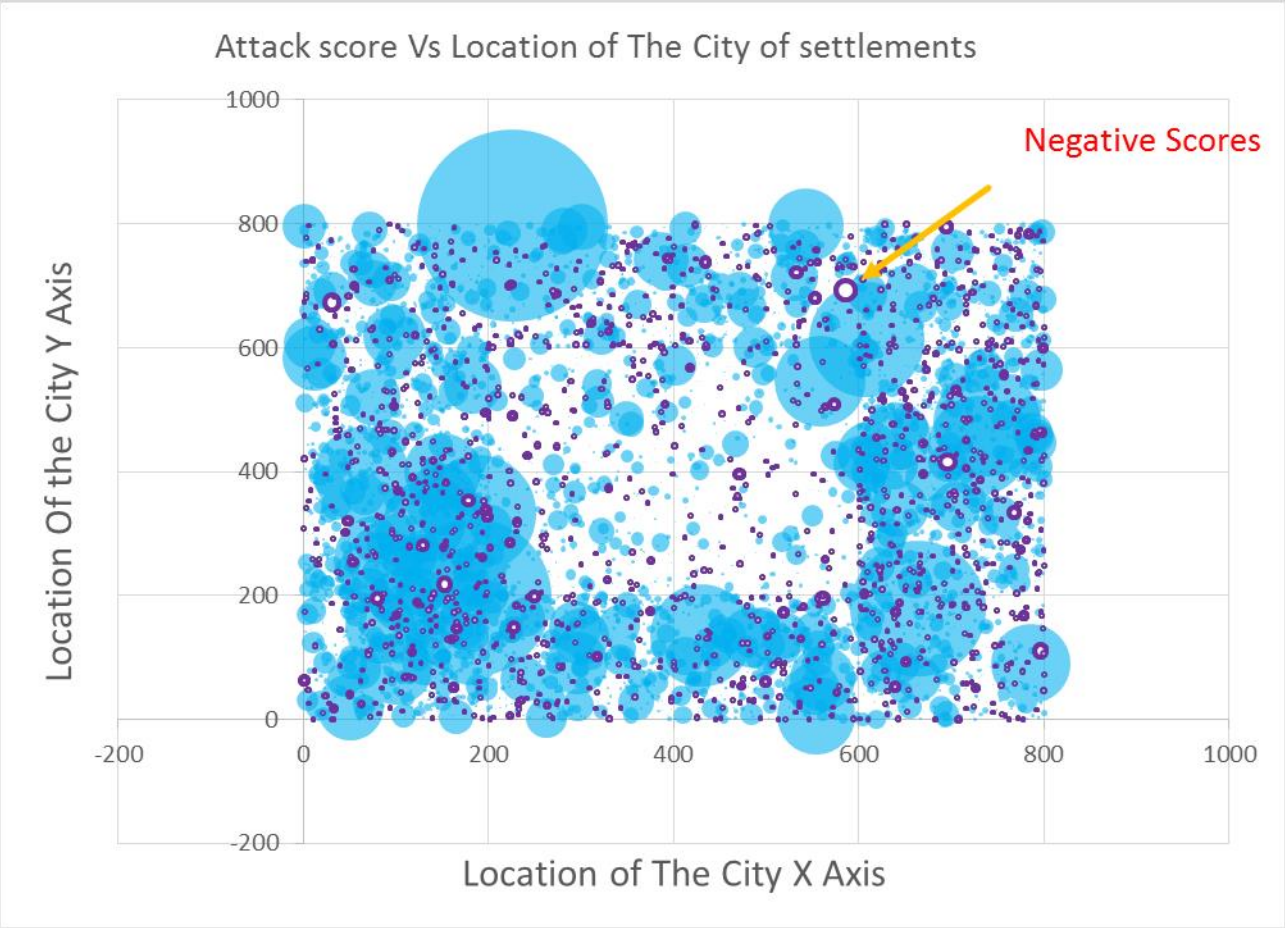
Correlation Coefficient
0.012533

Kolgomorov-Smirnoff Test
D=0.9999
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
3. 686872

RESULT
Weak relationship between Defense Score and Location of the User

V. LOCATION VS PERFORMANCE



Average	Stddev	Max	Min
7.859	84.032	8133.341	-804.015

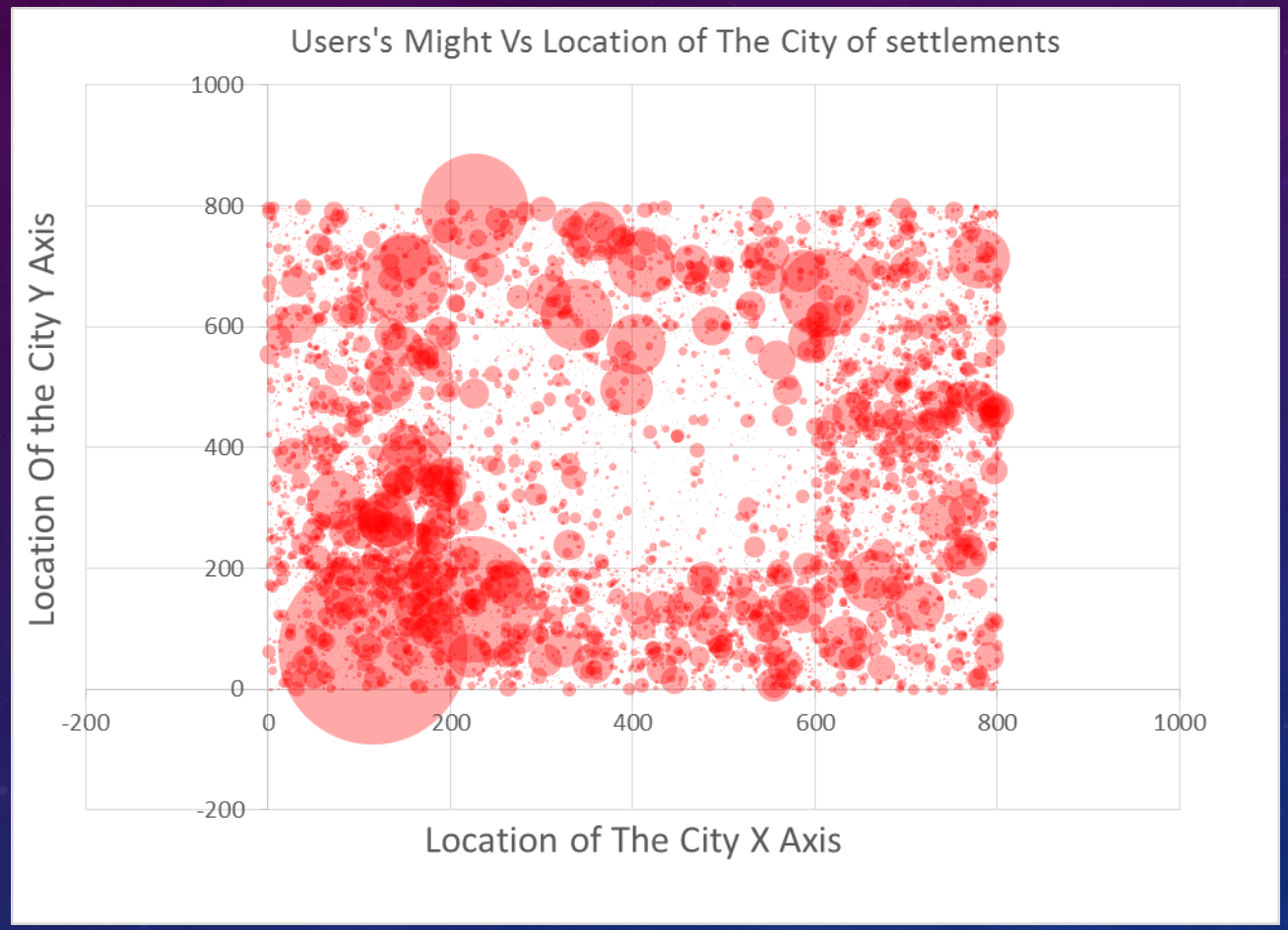
Correlation Coefficient
0.00131

Kolgomorov-Smirnoff Test
D=0.9982
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
0.937399

RESULT
Weak relationship between Losecount and Location of the User

V. LOCATION VS PERFORMANCE



Average	Stddev	Max	Min
30.066	139.732	15145.08	0.693

Correlation Coefficient
0.00884

Kolgomorov-Smirnoff Test
D=1
P value<2.2e-16
Null hypo: both are from same distribution. Alternate hypo cant be rejected.

Mutual Information
2.934839

RESULT
Weak relationship between Losecount and Location of the User

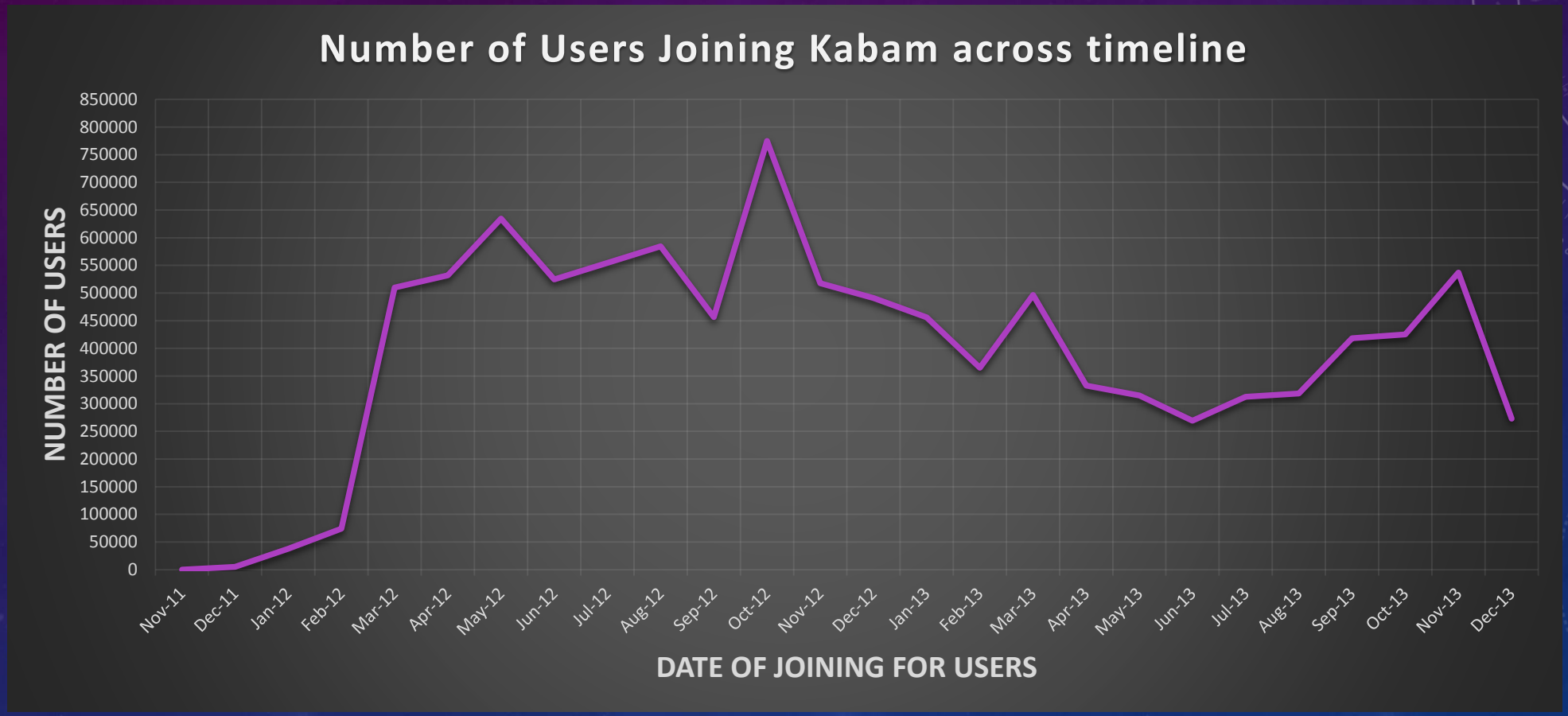
VII. FUTURE IDEAS

1. The Pathway of the warrior. How does the gamer starts from the beginning to become a leader board winner. What are the strength of the opponents he faces, the might he gathers, the loot he collects, the troops he gather, win-lost curve.
2. Cluster the Users based on their pathway to becoming a Winner vs Otherwise.
3. Establish Winning Secret.
4. Relationship between mood and Performance. When Good/Bad warriors fight with Good /Bad Mood.
5. Which is best way to spend your might? How to form the best combination of troop for the fight.
6. Is might the only winner of war?
7. Relationship between Language spoken by the User Vs Performance.
8. Predict might results.
9. Predict loss of might /resources in a fight even before fighting??
10. Predict change of alliances for a User.

VII. FUTURE IDEAS

1. Summarize the communications in the messages to understand the crux. Build word cloud to see the top most discussed words/features of the game/qualities of the game/impression (both positive and negative) of the User on the Game.
2. Positive impact of external competitions/new special rounds/challenge rounds on the User's Performance.
3. Establish Performance improvement methodologies. Evaluate their Effectiveness.
4. How does the happiness/Delight of the Users varies with expenditure of money, Improvement of Scores; Winning matches.
5. Introduce more rivalries between string fighters .. Like You have a new challenge!

V.USERS CHOICES ON ALLIANCE,KNIGHTS AND GEO POLITICS



III. USER EXPENDITURE VS PERFORMANCE & INTERACTION

AIM

WHAT NEXT?

1. Does expenditure improve performance?
2. what are potential matches for today?
3. More Info on Score computation.?
4. Add distribution of Happiness.