**Analyzing Location Vs Performance of the Users**

**AIM**:

To study the impact of the Location in the performance of Users in the Game of Kingdom of Camelot.

**ABSTRACT**:

Occupancy of the Users to build cities in the Location is ***sparser in the center of the Field***. Their settlements are ***concentrated on the outer core*** of the field. They are sporadic on the borders. They are sometimes thin and sometimes thick on the borders.

Win Count Vs Location: No concrete Relation.

Lose Count VS Location: No concrete Relation.

Draw cont Vs Location: No concrete Relation.

Attack Score Vs Location: No concrete Relation.

Defence Score Vs Location: No concrete Relation.

There is a ***NO strong/conclusive relationship*** between the Location and the User Performance Parameters. Correlation Coefficients are negligible almost showing that they are almost independent. The Kologomorov smiroff tests between the samples are inconclusive. The Mutual Information shows a weak positive information share between Draw count, Lose count and Defence score on the Location.

**PURPOSE**:

1. The Location in the Game of Kingdom of Camelot is chosen by the gamers to build their city
2. Location choice is strategically made. It needs to be in a suitable position favorable for both attack and defense and with the right set of neighbors.
3. We want to understand the Location choice of our gamers.
4. We want to study the correlation between the location choice with their user performance during the attack and the defense phases.
5. We can perhaps suggest a suitable location for future gamers via recommendation if there is strong correlation between the performance of user and the Location. It also gives us ideas on how to enrich their Gaming experience.

**PROCEDURE**:

The location of the attacks, user performance and user win count are computed from attacks2.txt

Parsing: Perl,Python Back End: Mysql.

We will study the relation between the Location and each of the defining criteria for user’s performance namely

1. The Win Count vs the Location of the cities.
2. The Lose Count vs the Location of the cities.
3. The Draw Count vs the Location of the cities.
4. The Defence Score vs the Location of the Cities.
5. The Attack Score vs the Location of the Cities.

**INPUT DATA**: Total Number of Locations: 171039 Total Number of Matches: 3052296.

**REGIONS IN THE MAP**:

**Inner Core/Center**: X = { (300 to 500) } and Y = { (300 to 500) }

**Outer Core/Wings**: X = { (100 to 200) , (500 to 700) } and Y = { (100 to 200) , (500,700) }

**Border**/**Crust**: X = { (0 to 100) , (700 to 800) } and Y = { (0 to 100) , (700,800) }

**GRAPH 1**: **Win Count Vs Locations**

We want to see if there is any relationship between location of the cities and the Win count of the Users in that location using a ***Bubble Chart***. A bubble chart is a type of chart that displays three dimensions of data. Each entity with its triplet of associated data is plotted as a *bubble* that expresses two of the vᵢ values through the *bubble*'s (X,Y) location and the third dimension*(The Winning Count)* through its size. Here the X and Y coordinates values corresponds to the actual X,Y coordinates of city.

**X AXIS**: The ***X coordinate of the City***.

**Y AXIS**: The ***Y coordinate of the city.***

**BUBBLE SIZE:** The ***Total Number of matches won in that location.***

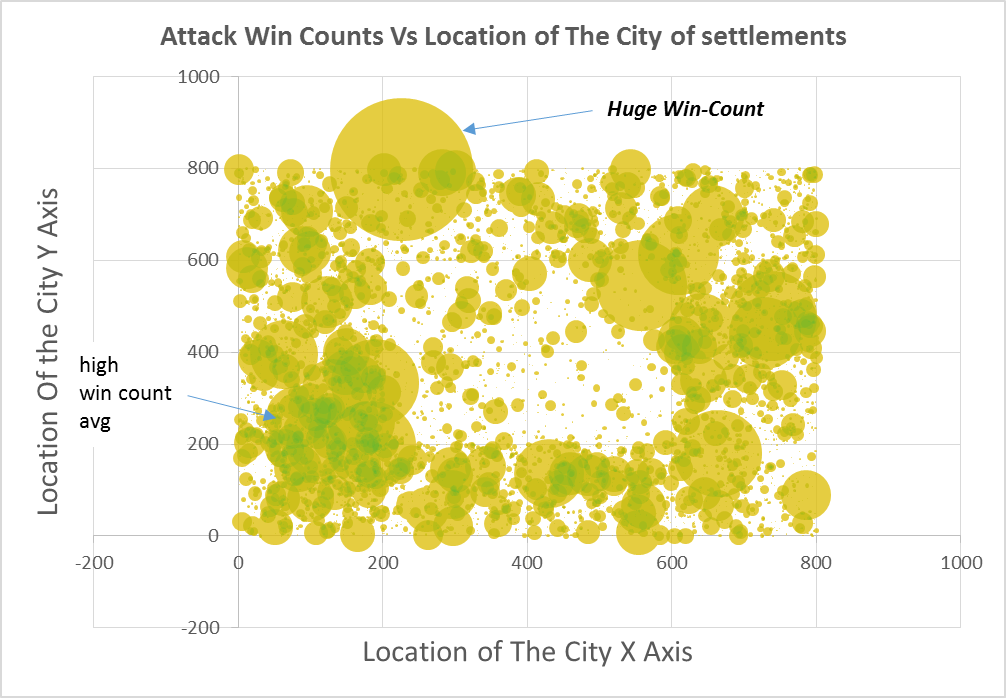
**DATA**: The win count is defined as the number of times the Defenders successfully defended their city plus the

Number of successful attacks made by the attackers in the location.

*Average Win Count Per Location: 7.77322 Standard Deviation: 14.44*

**OBSERVATION:**

1. Bubbles with high win count concentrated with around the region marked Red.
2. Bubble marked in **Grey** seems to have warriors with a potential which could become greater in future.
3. The Lower Win Count are concentrated on the border lines and in between the edges marked Blue.

****

**Low Win count**

**GRAPH 2**: **Lose Count Vs Locations**

We want to see if there is any relationship between location of the cities and the Win count of the Users in that location using a ***Bubble Chart***.

**X AXIS**: The ***X coordinate of the City***.

**Y AXIS**: The ***Y coordinate of the city.***

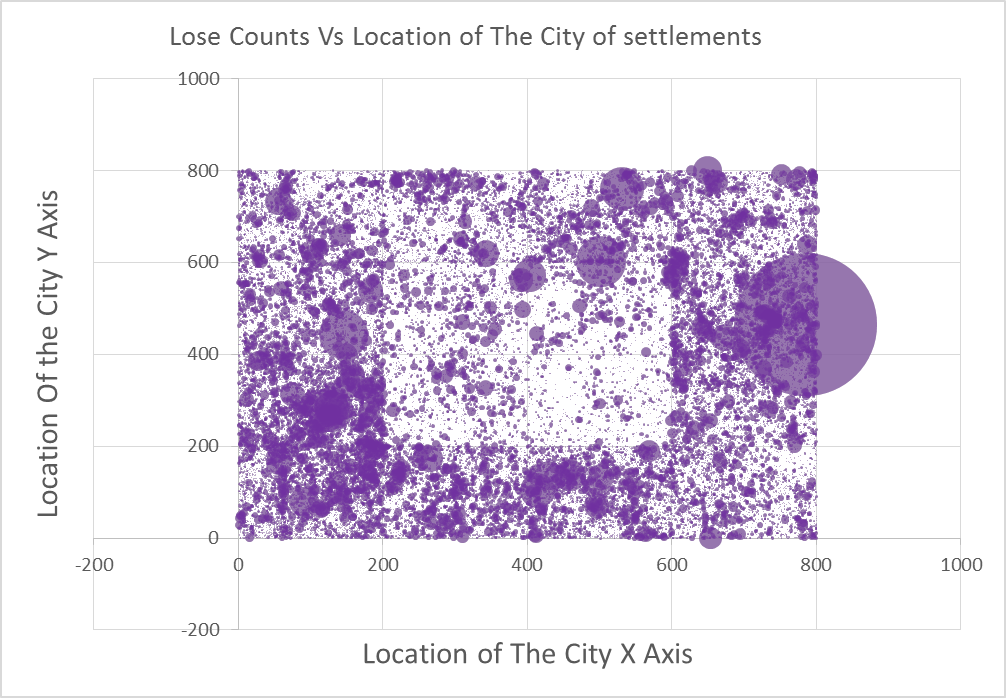
**BUBBLE SIZE:** The ***Total Number of matches Lost in that location.***

**DATA**: The Lose count is defined as the number of times the Defenders failed to defend their city summed with the Number of failed attacks made by the attackers in the location. Match Result=0.

*Average Lose Count Per Location: 7.77322 Standard Deviation: 14.44*

**OBSERVATION**

1. The Lose counts are more uniform throughout the regions unlike the Winning counts which concentrated on the few. This is explained the fact that the great warriors from the few regions conquer from a huge area around them and not just from few focused regions to plunder wealth.
2. Bubbles with high Lose count concentrated with around the region marked **Orange,** ***Incidentally we saw high Win Count in that region in the previous graph*. *This region is filled with great warriors and great failures.***

******

**High Lose count**

**GRAPH 3**: **Draw Count Vs Locations**

We want to see if there is any relationship between location of the cities and the Draw count of the Users in that location using a ***Bubble Chart***. The Draw matches are the matches with result 2. We add the number of timer the region drew to its location.

**X AXIS**: The ***X coordinate of the City***.

**Y AXIS**: The ***Y coordinate of the city.***

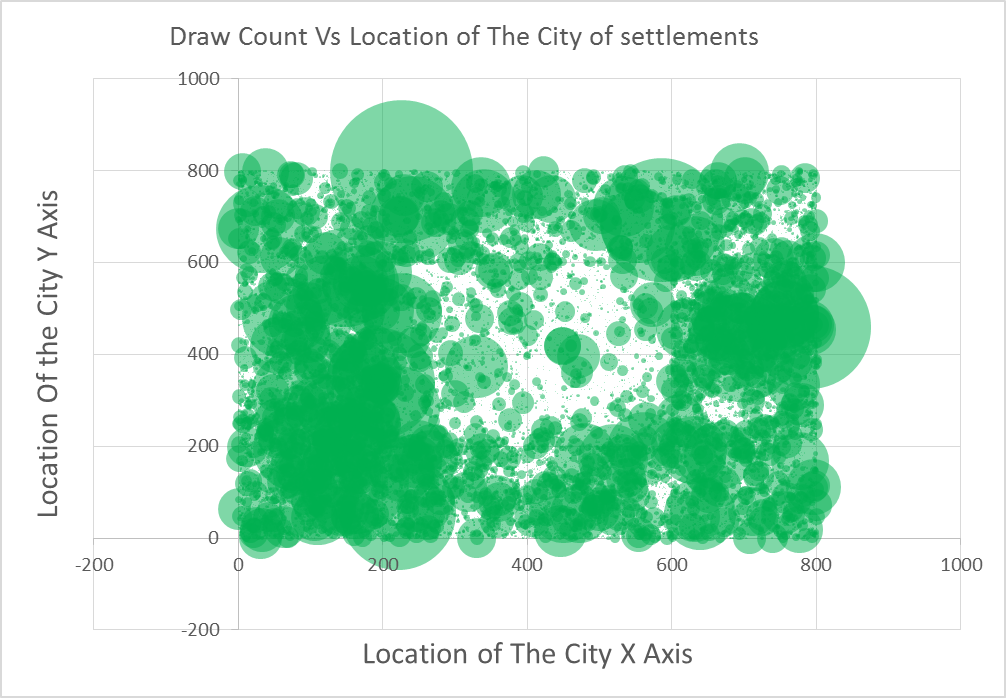
**BUBBLE SIZE:** The ***Total Number of matches Drawn in that location.***

**DATA**: The Draw count is defined as the number of times the neither the defenders nor the attackers won the match. Match Result=2.But there was plunder of wealth from the Defenders by the attackers (other than Gold).

*Average Draw Count Per Location: 19.92 Standard Deviation: 92.924*

**OBSERVATION**

1. The size of the Draw counts are bigger than lose counts and win counts for the reason that the wealth accumulation seems like the main reason for the drawn matches. And literally almost everybody would want to accumulate as much wealth as possible to survive the battles/build the cities.
2. We see heavy concentration of huge bubbles almost uniformly throughout the regions. Bubbles with high Draw count concentrated with around the region marked **Orange,** ***Consequentially we saw high Win Count/Lose in that region in the previous graphs*. *This region is filled warriors who are easily good at accumulating wealth.*** There are also regions of high drawn matches marked by **Dark Blue.** *These new cities might be building up to be the future great warriors.*



**High Avg**

**Draw counts**

**High Avg**

**Draw counts**

**High Draw count**

**GRAPH 4**: **Attack Score Vs Location**

We want to see the relationship between location of the cities and the Attack Scores of the Users in that location using a ***Bubble Chart***. The Attack Score is computed from the procedure mentioned in the User Performance.

**X AXIS**: The ***X coordinate of the City***.

**Y AXIS**: The ***Y coordinate of the city.***

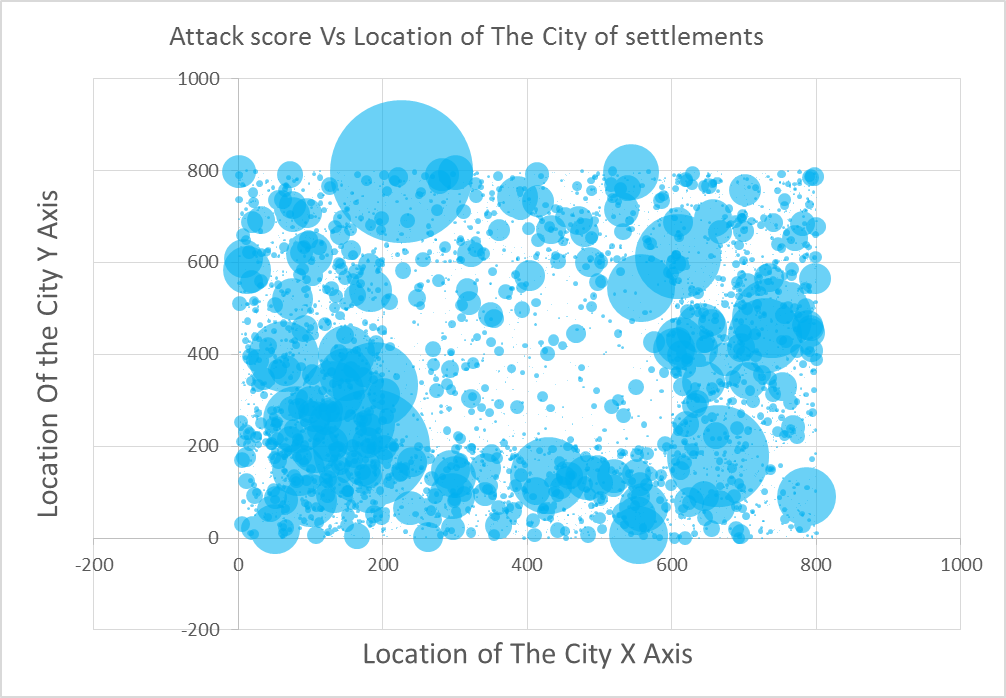
**BUBBLE SIZE:** The ***Sum of Attack Score in that location.***

**DATA**: Attack Score It is computed based on various criteria such as winning count; losing count; drawn matches; Might gathered; Might loss of the opponent; Might lose of the User; Combat Levels of the User and the Opponent; Being the Knight of the attack (Winning/Losing/drawn); Loot of Gold, Food, Wood, Ore, Stone; Combat Levels of the Knights; the strength of the wall and the rounds of fighting; we introduce logarithms to certain values in order to dampen the numerical overflow.

*Average Lose Count Per Location: 7.859 Standard Deviation: 84.02*

**OBSERVATION**

1. As expected the region marked by **Orange**, we see a great score for Attacking as we circled a bunch of warriors in that region.
2. We also see a next generation of warriors as marked by **Dark Blue** in the Previous showing a good attack strength.
3. There is one huge kingdom on (200,800) who had managed great Win count, Draw count and Attack Score.



**High Avg**

**Attack Score**

**High Attack Score**

**GRAPH 4**: **Defense Score Vs Location**

We want to see the relationship between location of the cities and the Defense Scores of the Users in that location using a ***Bubble Chart***. The Defense Score is computed from the procedure mentioned in the User Performance.

**X AXIS**: The ***X coordinate of the City***.

**Y AXIS**: The ***Y coordinate of the city.***

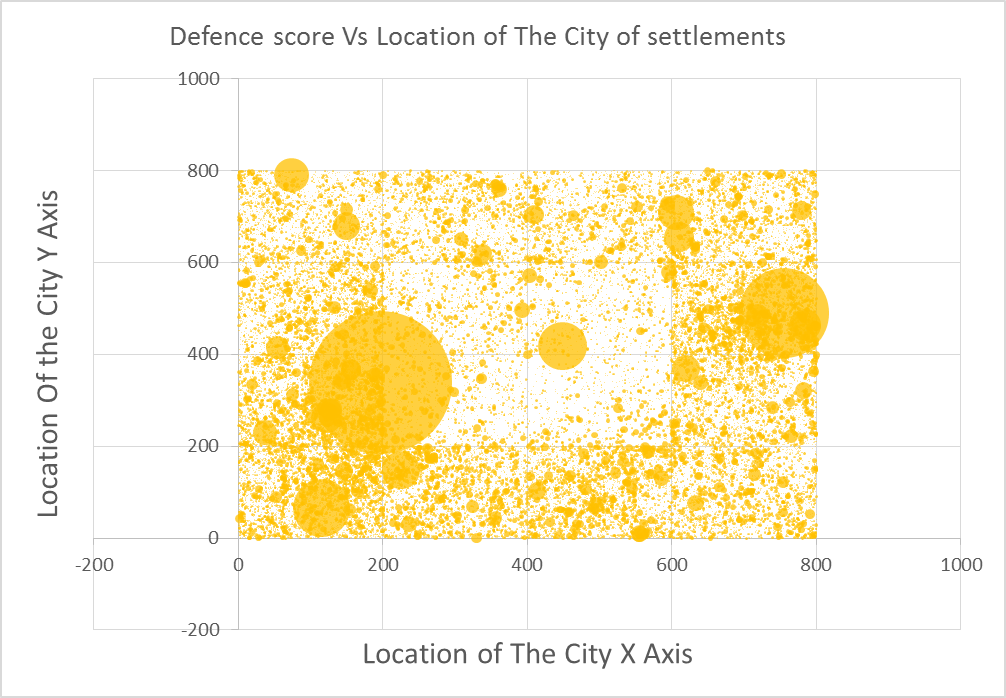
**BUBBLE SIZE:** The ***Sum of Defense Score in that location.***

**DATA**: Defense Score is computed based on various criteria such as winning count; losing count; drawn matches; Might gathered; Might loss of the opponent; Might lose of the User; Combat Levels of the User and the Opponent; Being the Knight of the attack (Winning/Losing/drawn); Loot of Gold, Food, Wood, Ore, Stone; Combat Levels of the Knights; the strength of the wall and the rounds of fighting; we introduce logarithms to certain values in order to dampen the numerical overflow.

*Average Lose Count Per Location: 20.859 Standard Deviation: 42.02*

**OBSERVATION**

1. Defense scores are little more uniform than the attack score with lesser variation. There is an overall medium level defense from whole of the gamers.
2. The Great Warriors who had won before marked in **Red** are reasonably fine at defending. While we do not see a stark differences like before, there are few stalwarts with great scores. This is also seen at **Dark Blue** region.

****

**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. 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. [3]

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}}

**Mutual Information (MI)**:

We can now look into the information they share amongst each other independent of any relationships. How knowing one of the Random variables does affects the knowledge of the other random variable? In other words. How does ***knowing the location of the city is this place or fact that the there is a change in the User’s Location on games affects our knowledge on user Performance.*** Mutual information provides the amount of information one variable reveals about another. 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. [2]

**2-Sample Kolgomorov Smirnoff Test**:

The two sample Kolmogorov-Smirnov test is a nonparametric test that compares the cumulative distributions of two data sets to check if the samples are from same distribution. This is a test performed to check if values given by two data sets comes from a similar or same distribution. It is a measure of similarity between two data sets. [1]

For Any test of Hypothesis: we need null hypothesis, The alternate Hypothesis, The test statistic, the P value and the threshold of acceptance.

***The null hypothesis*** is that both groups were sampled from populations with identical distributions. It tests for any violation of that null hypothesis -- different medians, different variances, or different distributions.

***Alternate Hypothesis***: The samples are from different distributions.

***Test Statistic***: The empirical distribution function  for  iid observations  is defined as

F_n(x)={1 \over n}\sum_{i=1}^n I_{X_i\leq x}

where  I_{X_i\leq x}  is the indicator function, equal to 1 if  and equal to 0 otherwise.

The **Kolmogorov–Smirnov statistic** for a given cumulative distribution function *F*(*x*) is

D_{n,n'}=\sup_x |F_{1,n}(x)-F_{2,n'}(x)|,

where F_{1,n} and F_{2,n'} are the empirical distribution functions of the first and the second sample respectively.

***P value***: This is the answer to this question:

*If the two samples were randomly sampled from identical populations, what is the probability that the two cumulative frequency distributions would be as far apart as observed? More precisely, what is the chance that the value of the Komogorov-Smirnov D statistic would be as large or larger than observed?*

If the P value is small, conclude that the two groups were sampled from populations with different distributions. The populations may differ in median, variability or the shape of the distribution.

***Threshold of Acceptance***: The value is usually 0.01 or 0.05. If the P-value obtained is less than the threshold, we conclude that the Null –hypothesis is true and the samples are from different distribution. Otherwise, the sample are from similar distribution and the cumulative distance is lesser.

***STATISTICS RESULTS ON RELATIONSHIPS BETWEEN LOCATION AND USER PERFORMANCE.***

We did statistics test on three Location values X coordinates, Y coordinate, Location (800\*X+Y)

{ making it linear coordinates from an 2D- matrix }.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| User Performance | STATS TEST | X Coord | Y Coord | X\*800+Y | P Value of K-S Test |
| **WIN COUNT** | Correlation Coefficient | -0.00053 | -0.00154 | -0.00053 | NA |
| Mutual Info | 0.19065 | 0.191273 | 0.959954 | NA |
| Kolgomorov  Smirnov Test | D=0.9254 | D=0.9527 | D=0.9985 | p-value<2.2e-16 |
| **LOSE COUNT** | Correlation Coefficient | -0.00881 | -0.01936 | -0.00884 | NA |
| Mutual Info | 0.174633 | 0.173783 | **2.934839** | NA |
| Kolgomorov Smirnov Test | D=0.9007 | D=0.9044 | D=1 | p-value<2.2e-16 |
| **DRAW COUNT** | Correlation Coefficient | 0.010443 | -0.09889 | 0.01043 | NA |
| Mutual Info | 0.336754 | 0.33732 | **3.16612** | NA |
| Kolgomorov Smirnov Test | D=0.8328 | D=0.8407 | D=0.9977 | p-value<2.2e-16 |
| **Defence Score** | Correlation Coefficient | 0.012562 | -0.02347 | 0.012533 | NA |
| Mutual Info | 0.346854 | 0.346221 | **3.686872** | NA |
| Kolgomorov Smirnov Test | D=0.7991 | D=0.8106 | D=0.9999 | p-value<2.2e-16 |
| **Attack Score** | Correlation Coefficient | -0.00131 | -0.00225 | -0.00131 | NA |
| Mutual Info | 0.240538 | 0.2418 | 0.937399 | NA |
| Kolgomorov Smirnov Test | D=0.9261 | D=0.9258 | D=0.9982 | p-value<2.2e-16 |

There is a small positive mutual information between the location(X\*800+Y) and the Lose Count, Draw Count and the Defence Score but that doesn’t give us a strong relation between the variables. The correlation coefficient, Mutual Information and the Statistic test results shows that there is no relationship between the Location of the city to the User Performance.

There is ***no conclusive relationship*** between the Location with any of the following User Performance indicators:

1. Win Counts.
2. Draw Counts.
3. Lose Counts.
4. Defence Scores.
5. Attack Scores.

**CONCLUSION:**

There is a ***NO strong/conclusive relationship*** between the Location and the User Performance Parameters. Correlation Coefficients are negligible almost showing that they are almost independent. The ***Kologomorov smirnoff*** tests between the samples are ***inconclusive***. The Mutual Information shows a *weak positive information share* between ***Draw count, Lose count and Defense score on the Location.***

**REFERENCES:**

1. Kirkman, T.W. (1996) [Statistics to Use: Kolmogorov-Smirnov test](http://www.physics.csbsju.edu/stats/KS-test.html). (Accessed 10 Feb 2010)
2. <http://www.scholarpedia.org/article/Mutual_information>
3. http://stattrek.com/statistics/correlation.aspx