

Outcome of the Extra Delivery in Cricket

Data warehousing and Data mining approach

PPG Dinesh Asanka
Pearson Lanka (Pvt) Ltd.
Sri Lanka.

Abstract— There is a common belief in cricket that extra delivery in an over will cost more to the bowling team than to the batting team. This research paper is to verify the above statement. Many parameters were identified and this research is to verify what the leading factors which will affect the outcome of the extra delivery. Apart from the standard static analysis data warehousing and data mining approach will be used. After analysis data, it was discovered that more than individual factors, team cricket status will more matter to the outcome of the extra delivery.

Index Terms — Cricket, T20, Extra Delivery, Association Rule, Clustering Data ware House, Data Mining

I. INTRODUCTION

In cricket, one over consist of six deliveries and in case there is an illegal delivery, either wide or no ball, bowler has to deliver an extra delivery. There is a common belief mostly among cricket experts that this extra delivery cost more to the bowling team than to batting team. This research paper is to find out, whether that statement is true and in what conditions this statement is true. To verify this hypothesis, data ware house and data mining techniques are used as those techniques will provide more accurate results than standard statistics methods. Another outcome of this research is to find out the reasons for the result of extra delivery.

II. DATA COLLECTIONS

First, data collected for all three formats of cricket, i.e. Test cricket, One day International (ODI) and Twenty-Twenty International (T20I). However, Test and ODI data had to be ignored and Twenty-Twenty International was selected as the data format due to following reasons.

A. Test Matches

There are lots of variations in test cricket.

- When end of each session (Lunch / Tea / End of days play) is approaching players will not take risks hence they don't care about last delivery.
- If any player is approaching a land mark (Century or any other carrier land marks) players try to play without taking any risks.
- In long running innings, there can be multiple balls will be used. Depending on the condition of the ball players will play safely.
- If one team wants to play to avoid defeat, they won't take risks

- If the tail-enders (players who don't have skills with batting) will want proper batsmen to keep the strike in the next over hence they will play safe despite the extra delivery. .
- Weather conditions will impact test matches since match will last for five days,

B. One Day Internationals (ODI)

- Main reasons for not to consider ODI is rules of ODI were changed very frequently.
- Super sub player was introduced and abandoned.
- Two new balls were used from 2012. Due to this rules swinging condition of the ball got changed.
- Field restrictions were changed time to time. With new rules in 2012, only four fielders can be stationed during non-power play overs.
- Free hit for following delivery after a front foot no-ball was introduced in 2012.
- Power play rules changed very frequently.
- Batting team has the option of taking the batting power play when they need.
- Since power play overs can be taken between 11th to 39th overs, power play will have different types of flow in the match.
- In multi country matches, the winning team gets a bonus point, if they maintain run rate of 1.25 times the other teams. So in the event that a team looking for the bonus point match will become 40 overs match.

C. Twenty-Twenty International (T20I)

T20I started recently around 5 years and rules are stable. After the T20I was introduced only rule changed was introduction of super over in which for tie matches there will be additional one over like extra time in football. . This rule does not affect this research as this research considers only 40 overs. Also fact that T20I will last around twenty overs it can be considered that it is consistent during the entire match.

D. Selected Matches

Initially, it was decided to select T20I matches between two world cups, but then there won't be enough data to be analyzed. However, matches of the World cup 2012 in Colombo, Sri Lanka was considered, those matches were more skewed to Sri Lanka. Then it was decided to collect more data. However, most of the matches weather conditions were changed and some matches were played between test matches and ODI. Hence less importance were given to the

T20I. Therefore, it was decided to collect data from one tournament and 2014 T20I world cup in Bangladesh [11] was selected as the data set for the research as every country give high priority for this event.

E. Rules

Few rules were placed when data collections to maintain the consistency of the collected data.

- Half way interrupted matches/innings were ignored. Also, results obtained from Dougworth Lewis method [9] was ignored as well. However, during the 2014 world cup, there were no such matches.
- Multiple extra delivery overs were ignored. This means that in a space of one over, there are more than one extra delivery was sent. When there are multiple extra deliveries it is difficult to identify what the extra delivery is. There are 29 such incidents in the entire tournament. During the entire tournament, there 224 instances of extra deliveries. Therefore, 13% of those incidents were ignored.
- Half-way completed overs also were ignored since half-way completed overs does not have an extra delivery. However, during this world cup there were no such incidents.

After studying the commentators' feedback on different matches following factors were identified as the possible reasons for the extra delivery.

TABLE I. IDENTIFIED REASONS

Factor	Description	Reason
Bowler	Name of the player who delivered the extra delivery.	There are varies types of bowlers. Therefore, bowler needs to be captured.
Batsmen	Name of the player who faced the extra delivery.	There are varies types of batsmen. There batmen and few other parameters is also needs to be captured.
Runs scored for batsman	When batsman facing the extra delivery how many runs he has scored.	Depending on much batsman has scored will depend on how much he will score.
Number of deliveries batsman has faced	When batsman facing the extra delivery how many deliveries he has faced prior.	Normally when a batsman starts, he needs few time to gets his focus. There can be dependency to the extra delivery outcome from the number of deliveries he has faced.
Innings	Whether it is batting first or second innings of the match.	It is assumed that while chasing there can be a different approach.
Type of the extra delivery	Whether it is no-ball or a wide ball.	Bowlers' rhythm might be different from no-ball to a wide ball.
Ball number	Ball number in which the wide / no-ball was sent by the bowler.	This to identify for which delivers bowler has to focus. It is said that bowlers tend to send wide at the start of over and tend to send no-ball at the end of overs.
Runs in the previous over	How many runs were scored during the previous over?	It is essential to identify whether there is a pattern of runs scoring.

Runs in the after over	How many runs were scored during the after over of the extra delivery over?	It is essential to identify whether there is a pattern of runs scoring.
Runs of the over	How many runs were scored in the over, excluding the extra delivery	It is essential to identify whether in the given over there is a pattern of scoring. If it is high scoring over whether it is an extra ball or not, last delivery will have a high runs.
Score in extra ball	How many runs were scored for the last delivery?	This is the final outcome of the research.
Power play	Whether the delivery was sent during the power play. Power play is first 6 overs in T20I.	During power play only two fielders can be stationed outside 30 yards.
Partnership	What is the partnership?	Normally in front and middle order more capable batmen will plan the extra delivery.
Partnership Runs	How many runs partnership has yielded before the extra delivery was sent in.	
Partnership deliveries	How many deliveries were used by this partnership before the extra delivery was sent in.	
Wicket in the same over	If the wicket was fallen during the same over.	If a wicket is fallen batsmen tends to play safely as bowlers are also putting their extra effort.
FreeHit	Whether the last delivery is delivery is sent as a free hit.	If it is free hit batsmen can be get out only by run-out therefore, batsmen tends to play more freely.
Over Number	Over number	Over Number indicates the phase of the match.

Batsmen's and bowlers' parameters were also identified. There are two types of parameters.

1. Fixed parameters like batting hand (right and left), bowling hand (right or left), country etc. These values are captured one time only.
2. Match dependent parameters. Parameters like number of runs or number of wickets are varying from match to match. Therefore those needs to be taken before every match.

Fixed parameters identified for bowlers and batsmen are shown in Table II and Table III.

TABLE II. FIXED PARAMETERS FOR BOWLERS

Factor	Description
Country	Which country bowler represents.
Bowling Hand	Whether the bowler is right or left hand bowler.
Bowling Style	Leg Spinner, off spinner, fast medium, fast, medium
All Rounder	Whether the player has the capabilities in both batting and bowling.

TABLE III. FIXED PARAMETERS FOR BOWLERS

Factor	Description
Country	Which country batsman represents.
Batting Hand	Whether the bowler is right or left hand bowler.
Wicket Keeper	Whether the batsmen is the wicket keeper
All Rounder	Whether the player has the capabilities in both batting and bowling.

There are instances where some players played for different country. For example, ED Joyce [5] started to play for Ireland, then he moved to play England then back again to Ireland. Also, Luke Ronchi [4] started to play for Australia and now currently playing for New Zealand. However, this is not needed to consider as in the world cup there is no possibility for players to move between countries.

Match dependent parameters for bowlers and batsmen are shown in Table IV and Table V.

TABLE IV. IDENTIFIED PARAMETERS FOR BOWLERS

Factor	Description	Reason
Number of Matches	How many matches has the player played?	This gives an idea about the experience of the player.
Wickets	No of wickets player has taken.	
Overs	Number of overs player has sent	
Economy Rate	How many runs were scored against the player per over.	This indicates whether it is difficult to score runs against the bowler.
Batting Ranking	ICC LG Player Ranking was used.	

TABLE V. IDENTIFIED PARAMETERS FOR BOWLERS

Factor	Description	Reason
Number of Matches	How many matches has the player played?	This gives an idea about the experience of the player.
Runs	How many runs were scored by the batsmen	
Strike Rate	Number of runs scored by the batsmen per 100 deliveries.	This will indicate whether the batsman is a striking batsmen.
Average	Number of run scored per innings. Number of innings is calculated by counting number of inning batsmen got dismissed.	
Bowling Ranking	ICC LG Player Ranking was used.	

Ranking are taken from the International Cricket Rankings (ICC) [6], LG player rankings [7] which is the standard ratings used by ICC. However, ranking were not updated for each and every match during the 2014 T20I world cup. Therefore, existing rankings were obtained before the matches. 196 instances of extra delivery incidents were collected. Following are the countries which have sent extra deliveries.

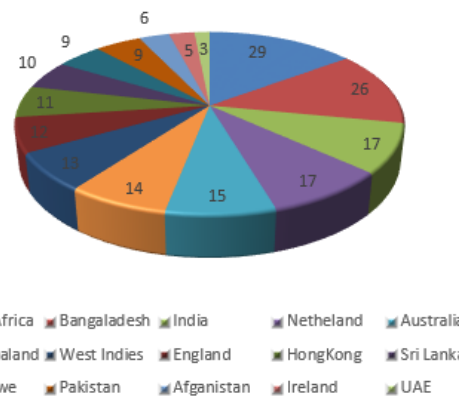


Fig. 1. Country Wise Extra Delivery Sent

South Africa and Banladesh have sent many extra deliveries. Fig 2 shows countries which were benifited from the extra deliveries.

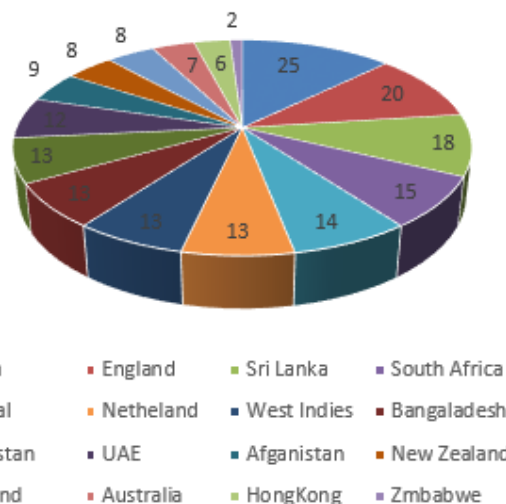


Fig. 2. Country Wise Extra Delivery Sent

India, England and Sri Lanka are the countires benefited from the extra deliveries. Fig 3 shows types of extra deliveries consumed by the bowlers.

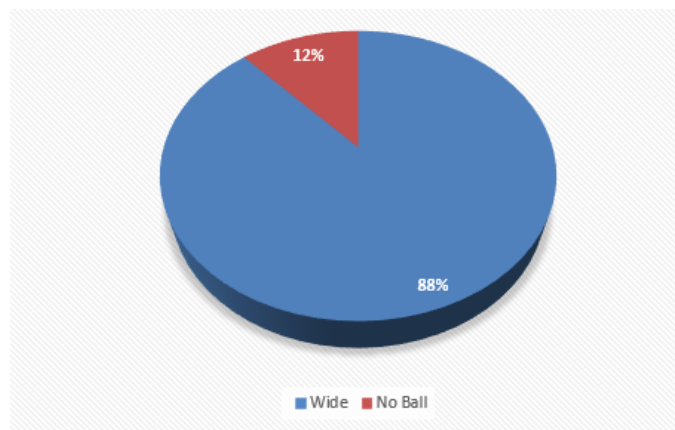


Fig. 3. Type of Extra Deliveries

Out of all the extra deliveries, 88% of them are wide and only 12% are no ball deliveries. After introduction of free hit into T20I, bowlers have improved by not sending more no balls.

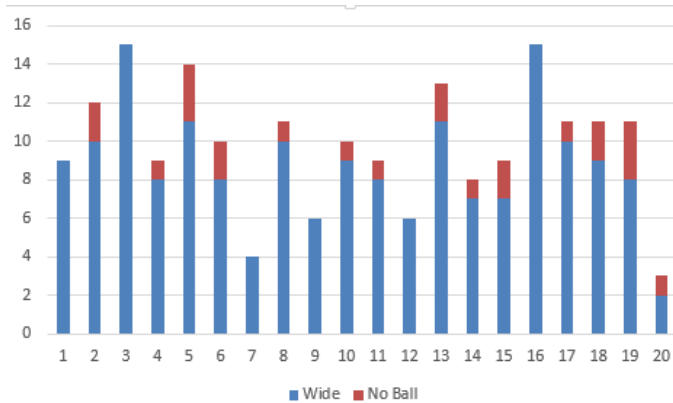


Fig. 4. Extra Delivery Sent Overs

Fig 4 shows distribution of extra deliveries with respect to different overs. 3rd and 16th overs are the overs which were seen high number of extra deliveries. However, both overs did not see any no ball deliveries.

Fig. 5 shows delivery number which extra delivery was sent.

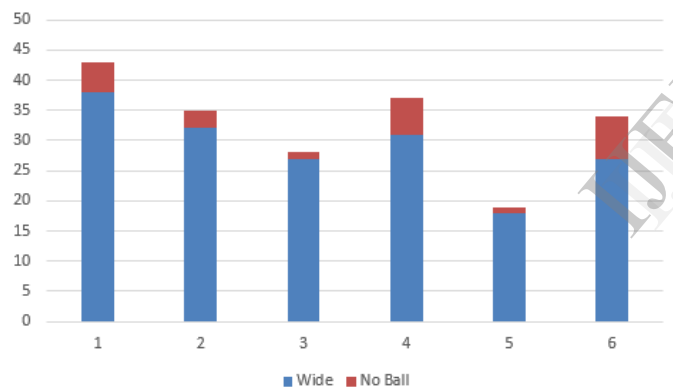


Fig. 5. Delivery Number wise Extra Delivery

From the above Fig 5, it is evident that first delivery has chance of being called extra delivery hence bowlers needs to be more focus when they start a new over.

Fig 6 shows runs scored in the extra delivery.

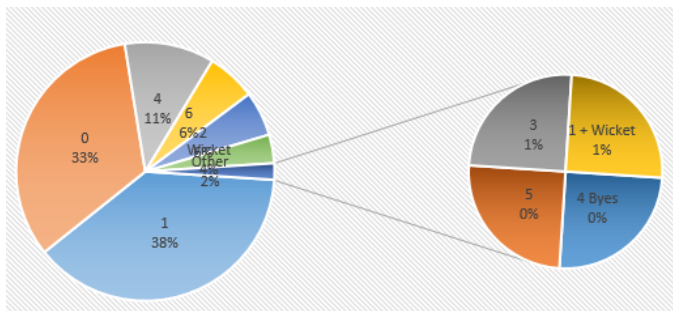


Fig. 6. Runs Scored for the Extra Delivery

It is observed that more than 60% of extra deliveries has cost only no runs or only one run as it can be observed from Fig 6. This indicates that, in general extra delivery does not cost much for the batting team.

III. TECHNOLOGIES USED

Data ware house is a system used for reporting and data analysis [1] [2]. Data ware house is used in most of sectors such as Retail, e-commerce, Procurement, Customer Relationship Management, Financial Services, Education, Health care etc. [3]. Data mining is used to predict the data. So in this research both of these techniques are used along with conventional statistics techniques.

Fact and Dimension tables are identified for the collected data set. Also, range dimensions were identified to improve the end user analysis.

Nature of the data set does not allowed to use star schema hence snow flex schema was used.

Microsoft SQL Server 2014 [11] used as the storage for these data set while for analysis purposes SQL Server Analysis Service 2014 was used. For the better presentation and the simplicity of usage Microsoft Excel 2014 was used.

Clustering and association techniques are used in this research. Clustering technique is used to identify the natural grouping so that it will be identify the same data sets which falls to the same group.

IV. DESIGN OF DATA STORE

Data ware house was desinged to accomadate these data so that it can be analysed using data ware house techniques. DimCountry, DimPlayer and DimMatch are the main dimensions while FactExtraDelivery, FactBatsmen, FactBowler and FactTeamRankings are the Fact tables. Proposed data ware house desinged is shown in Fig 7.

For all the measures columns in the fact tables, range value is introduced to improve the analysis.

Categories in FactExtraDelivery

Two mechanics were used to identify the ranges for measures columns. They are depending on the domain and simple statistical method.

Following are the range dimension identified by the domain.

Partnership: Depending on the partnership stage with respect to the innings categorization was done. Partnerships between 1 and 3 categorized to Front Order, Partnerships between 4 and 7 categorized Middle Order and partnerships between 8 and 10 categorized as Late Order.

Over Number: There is already a categorization for power play overs which are sent between 1 and 6 and same categorization is used. Overs between 7 and 10 are categorized to After Power Play and Overs between 11 and 16 are categorized as Middle Overs and overs between 17 and 20 categorized as Final Overs.

Score in Extra Ball: Since there are several scores were made to the extra delivery that is also categorized depending on the impact to the batting team. If it is 0 or 1 run it is categorized as No Impact while 2 or 3 is categorized as average impact. Meanwhile 4, 5 or 6 is categorized as Large Impact. If the wicket has fallen in the extra delivery it is categorized as Adverse Impact.

Using equal frequency distribution method [8] ranges were identified for following measures.

Measure Column	Low	Medium	High
PartnershipDeliveries	0 - 7	8 - 19	20 +
PartnershipRuns	0 - 7	8 - 21	22 +
RunsScoredforBatsmen	0 - 6	7 - 25	26 +
BatsmenFacedDeliveries	0 - 6	7 - 17	18 +
RunsInthePreviousOver	0 - 5	6 - 9	10 +
RunsInAfterOver	0 - 5	6 - 9	10 +
RunsOftheovers	0 - 5	6 - 9	10 +

Categories in FactBatsmen

Rankings were categorized by allocating high ranking for 1 – 10, medium ranking for 11 – 50 and low ranking for 51 – 100.

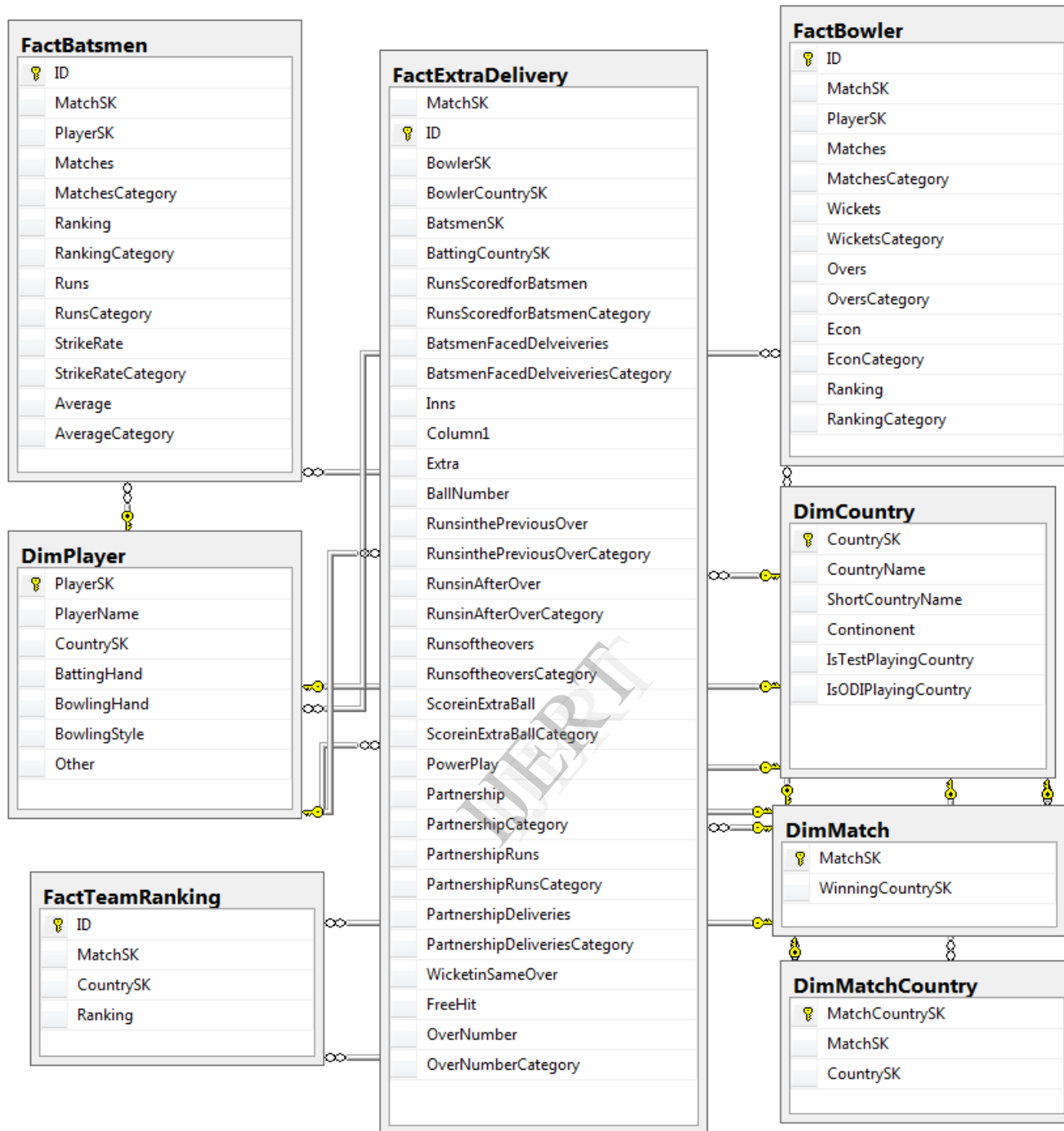
	Low	Medium	High
Ranking	51 - 100	11 - 50	1 - 10
Matches	0 - 14	15 - 32	33 +
Runs	0 - 230	231 - 580	481 +
StrikeRate	0 - 117	117.01 - 130	130.01 +
Average	0 - 21	21.01 - 29.00	29.01 +

Categories in FactBowler

Rankings were categorized by allocating high ranking for 1 – 10, medium ranking for 11 – 50 and low ranking for 51 – 100.

	Low	Medium	High
Ranking	51 - 100	11 - 50	1 - 10
Matches	0 - 9	10 - 26	27 +
Wickets	0 - 8	9 - 22	23 +
Overs	0 - 25	25.1 - 70	70.1 +
Economy	0 - 6.85	6.86 - 7.7	7.71 +

Figure 7 is the proposed design for above data ware house.



FiFig 7. Proposed Data Ware House Designed

V. ANALYSIS

Analysis was done using category of runs scored for extra delivery which has described in Section III. Since there are lot of different measures, category was used to identify influence of other parameters.

When the data was analyzed, it was observed that 72% of incidents does not have major impact whereas only 24% of incidents are either has large or average impact.

Outcome was analyzed depending on the non-test playing and test playing teams. It was observed that there is no much of a difference whether the batting team or bowling team is test playing or a non-test paying country as shown in figure 8.

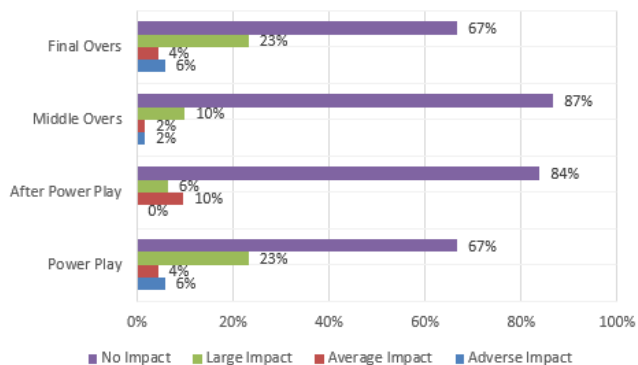


Fig 10 : Outcome Depending on the Over Number

As shown in the figure 10, just after power play overs and in the middle over there is a no impact from the extra delivery than it compared to power play and final overs. Also, during power play overs and in the the final over impact is high. Typically in final overs and late overs batmen look to score runs in every oppertunity.

Extra delivery was analysed with respect to the batting partnership. Since T20I are played in for 120 deliveries, most of the time it all the wickets are not needed and mostly front line and middle order batsmen are playing. From the collected data set, 57 % of extra deliveries were faced by front line batsmen and another 39 % by the middle order where as mere 4 % for late order.

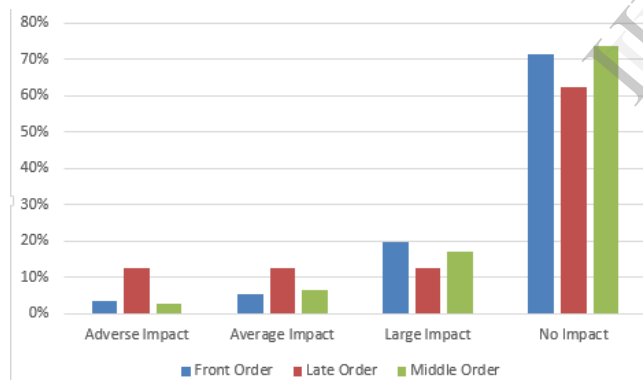


Fig 11 : Outcome Depending on the Batsmsen Partnership

Figure 11 values trivial since it shows as front line batsmen will have higher impact for the extra delivery.

When the partnership is well established, there is a tendancy for batsmen to score freely. Establishment of partnership is idenfied from the number of runs of the partnership as well as the number of deliveries of the partnership.

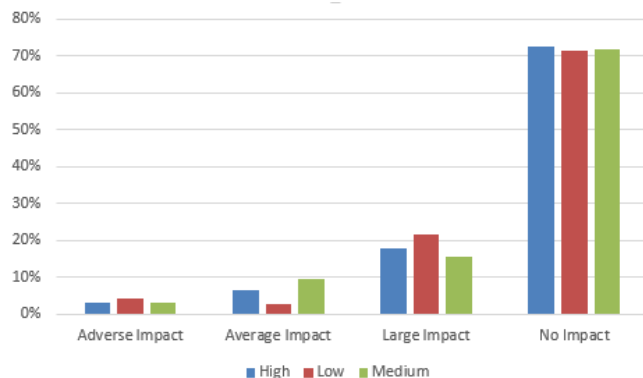


Fig 12 : Outcome Depending on the Batsmsen Partnership Deliveries

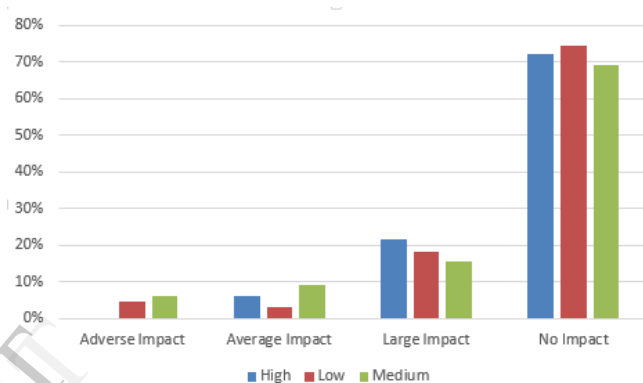


Fig 13 : Outcome Depending on the Batsmsen Partnership Runs

Above figures 12 and 13 indicate the same behaviour which was observed in the other findings.

Another parameter would be the batmen match form. This is measured with the number of runs and number of deliveries he has faced before the extra delivery in the match.

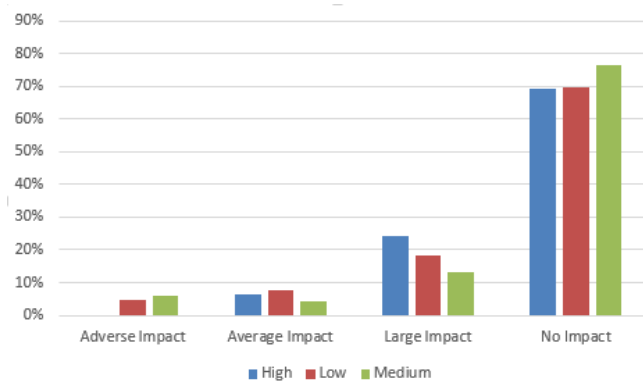


Fig 14 : Outcome Depending on the Runs scored by the Batsmsen

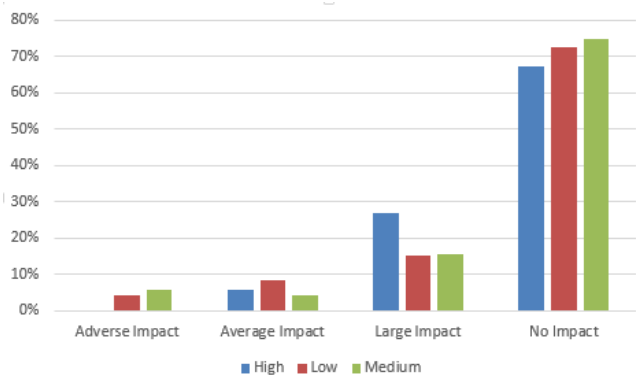


Fig 15 : Outcome Depending on the Number of Deliveries Faced by the Batsmen

According to figure 14 and figure 15, when the batsmen has scored high runs and faced high number of deliveries will make higher impact to the extra delivery.

Key Influencing Factors

Excel was used to analyse key influencing factors for outcome of the extra delivery. There was no clear indicator what the key influential factors are. However, It was identified that the playing country's ICC status is making a impact to the outcome of the extra delivery. For example, if the batsmen is not representing a one day international playing country, outcome of the last delivery is adversely effect to the batting team. Adversly means that wicket has fallen on the extra delivery.

Discrimination between factors leading to 'Large Impact' and 'Adverse Impact'			
Column	Value	Favors Large Impact	Favors Adverse Impact
Batsmen_IsODIPlayingCountry	Yes	Blue bar	
Batsmen_IsODIPlayingCountry	No		Orange bar
Bowler_IsTestPlayingCountry	No		Orange bar
Bowler_IsTestPlayingCountry	Yes	Blue bar	

Discrimination between factors leading to 'NO Impact' and 'Adverse Impact'			
Column	Value	Favors NO Impact	Favors Adverse Impact
Batsmen_IsODIPlayingCountry	Yes	Blue bar	
Batsmen_IsODIPlayingCountry	No		Orange bar
Batsmen_IsTestPlayingCountry	No		Orange bar
Batsmen_IsTestPlayingCountry	Yes	Blue bar	

Fig 16 : Comparison of Outcome of Extra Delivery with respect to the ICC Statuses of playing countries

Figure 16 indicates that when the batsmen is part of experience team, impact of the extra delivery is large.

Clustering

When clustering analysis was done, four major clusters were identified as shown in figure 17.

Characteristics for Cluster 1		
Variables	Values	Probability
Bowler_IsTestPlayingCountry	Yes	Blue bar
Batsmen_IsODIPlayingCountry	Yes	Blue bar
Batsmen_IsTestPlayingCountry	Yes	Blue bar
WicketinSameOver	NULL	Blue bar
PowerPlay	NULL	Blue bar
Bowler_BowlingHand	RH	Blue bar
Batsmen_BattingHand	RH	Blue bar
PartnershipCategory	Front Order	Blue bar
ScoreinExtraBallCategory	NO Impact	Blue bar

Characteristics for Cluster 2		
Variables	Values	Probability
FactBatsmen_RankingCategory	Low	Blue bar
FactBatsmen_RunsCategory	Low	Blue bar
FactBatsmen_MatchesCategory	Low	Blue bar
WicketinSameOver	NULL	Blue bar
Bowler_BowlingHand	RH	Blue bar
Batsmen_IsTestPlayingCountry	No	Blue bar
FactBowler_RankingCategory	Low	Blue bar
FactBatsmen_AverageCategory	Low	Blue bar
FactBatsmen_StrikeRateCategory	Low	Blue bar
ScoreinExtraBallCategory	NO Impact	Blue bar

Characteristics for Cluster 3		
Variables	Values	Probability
Batsmen_IsODIPlayingCountry	Yes	Blue bar
Batsmen_IsTestPlayingCountry	Yes	Blue bar
Bowler_IsTestPlayingCountry	Yes	Blue bar
ScoreinExtraBallCategory	NO Impact	Blue bar
WicketinSameOver	NULL	Blue bar
FactBatsmen_StrikeRateCategory	Medium	Blue bar
Batsmen_BattingHand	RH	Blue bar
PartnershipCategory	Front Order	Blue bar
FactBatsmen_RunsCategory	High	Blue bar
FactBatsmen_RankingCategory	Medium	Blue bar

Characteristics for Cluster 4		
Variables	Values	Probability
Batsmen_IsODIPlayingCountry	Yes	Blue bar
Batsmen_IsTestPlayingCountry	Yes	Blue bar
Bowler_IsTestPlayingCountry	Yes	Blue bar
WicketinSameOver	NULL	Blue bar
Batsmen_BattingHand	RH	Blue bar
ScoreinExtraBallCategory	NO Impact	Blue bar
Bowler_Continent	Asia	Blue bar
FactBowler_RankingCategory	Low	Blue bar
FactBowler_OversCategory	Low	Blue bar
FactBowler_MatchesCategory	Low	Blue bar

Fig 17 : Attributed of Major Clustering Groups

However, nothing could be gathered from the clustered obtained from the above figure.

VI. FUTURE WORK AND CONCLUSION

In most of the cases, it was reveled that impact of the extra delivery is not as costly as many experts believe. However, there few highlights found in the data analysis.

- There can be a large impact from the extra delivery with batsmen with medium strike rate and bowlers with high economy rates.
- Bowlers with high number of wickets have a little tendency to give away extra runs in the extra delivery.
- During power play overs and in the the final over impact is high.

- When the batsmen has scored high runs and faced high number of deliveries will make higher impact to the extra delivery.

After this research was carried there was a minor modification to the playing condition to the T20I. Therefore these results can be compared with next world cup. During the next world which is scheduled to be held in 2016 in India, same research will be done and will compare whether there is any other new trends. Also, 2016 women T20I data will be collected to analysis to verify whether there are any differences between men and women teams.

VII. REFERENCES

- [1] Bilal Ali Yaseen Alnassar, "Challenges in the Successful Implementation of Data Warehouse", Journal of Management Research, ISSN 1941-899X, 2014, Vol. 6, No. 3
- [2] Sanu Kumar, "Aspect Of Data Mining And Data Warehousing", International Journal of Technology Enhancements and Emerging Engineering Research, Vol 2, Issue 6 48. ISSN 2347-4289
- [3] Ralph Kimball , Margy Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, ISBN-13: 978-0471200246 ISBN-10: 0471200247, 2002, Wiley, 2 edition, pg. 5-10.
- [4] Luke Ronchi, CricInfo Players, <http://www.espnricinfo.com/new-zealand-v-south-africa-2014-15/content/player/7502.html>, Accessed on 2014-10-20.
- [5] Ed Joyce, CricInfo Players, <http://www.espnricinfo.com/ci/content/player/24249.html>, Accessed on 2014-10-20.
- [6] ICC LG Official Team Rankings, CricInfo, <http://www.espnricinfo.com/rankings/content/page/211271.html>, Accessed on 2014-10-20.
- [7] ICC LG Official Players Rankings, CricInfo, <http://www.espnricinfo.com/rankings/content/page/211270.html>, Accessed on 2014-10-20.
- [8] Statistics: Grouped Frequency Distributions, Jones, James <https://people.richland.edu/james/lecture/m170/ch02-grp.html>, Accessed on 2014-10-20.
- [9] The Duckworth-Lewis Method, CrinInfo, http://static.espnricinfo.com/db/ABOUT_CRICKET/RAIN_RULE_S/DUCKWORTH_LEWIS.html, Accessed on 2014-10-20.
- [10] ICC World Twenty20, <http://www.icc-cricket.com/world-t20>, Accessed on 2014-10-20.
- [11] Product Specifications for SQL Server 2014, Microsoft Developer Network, <http://msdn.microsoft.com/en-us/library/ms143287.aspx>, Accessed on 2014-10-20.