**Introduction**

**Indian Premier League**

The **Indian Premier League** (**IPL**), officially **Vivo Indian Premier League** for sponsorship reasons, is a professional Twenty20 cricket league in India contested during April and May of every year by teams representing Indian cities and some states. The league was founded by the Board of Control for Cricket in India (BCCI) in 2008, and is regarded as the brainchild of Lalit Modi, the founder and former commissioner of the league.

[**DATA GATHERING:**](https://towardsdatascience.com/data-science-for-startups-tracking-data-4087b66952a1)

Discusses the motivation for capturing data from applications and web pages, proposes different methods for collecting tracking data, introduces concerns such as privacy and fraud.

[**BUSINESS INTELLIGENCE:**](https://towardsdatascience.com/data-science-for-startups-business-intelligence-f4a2ba728e75)

Identifies common practices for ETLs, automated reports/dashboards and calculating run-the-business metrics. Presents an example with Database Managements and Data Studio(graphs).

**PLATFORM USED IN DATA ANALYSIS (DATA MANAGEMENT):**

1. **Microsoft Excel 2016(version 2016)**

**VISUALIZE:**

a). Pivot Table

b)**.** Simple graphs and Pictorial representations.

**Scope of the Analysis**

The report segments the IPL data from season 2008 to 2017. There are three files. One file contains the information about each match of all the seasons. Other file contains the information about what happens on each ball of the match. Both the files are related to each other by Match Id.

Cricket is a game of bat and ball where 2 teams of 11 players each play. The data which I have is for 20-20 format. In this both teams bat (and ball) for 20 overs each and whoever scores more runs wins. 11 players can generally be classified in 4 different categories

1. Batsman: Objective is to score runs as quickly as possible (4-6 Players in each team)
2. Bowler: Objective is to take wickets and give as less as runs possible (3-5 in each team)
3. All Rounder: Does both batting and bowling (3-5 in each team)
4. Wicket Keeper + Batsman: Stands behind the stumps for fielding team. Objective is to collect the ball and help in dismissing the batsman. Every team needs to have 1 specialist keeper.

Along with that all these players field when a team is bowling where there objective is to stop the ball, take catches and runout

The scope of the data tells various aspects of Indian Premier League. In cricket every team and every player is analyzed. By the data which I have collected I have compare various teams across various seasons, compared players strike rate and average. Players which have higher strike rate can be sent in slog overs to get a good total for the team.

Every Year auction of IPL takes place. Each team owner opts for some players. Based on the credentiality of the player owner’s pay the price to the players. If the a player is a hard hitting batsman which can be discovered by seeing his strike rate then for sure he will get a good sum of money.

Thus every year a player is analysed on the basis of his performance. Moreover analysis of each stadium whether it is good for batting first or balling first is also analysed. Also toss results can be analysed which can give information regarding which team won when batting first and which team won when fielding first.

Economy rate of each player is also calculated. It can be calculated using the formula runs given by the bowler divided by the number of balls he bowled.

The method which is used to analyse the data is pivot table and different types of graph in Microsoft Excel. Many pivot tables are generated to make the graph and carryout different analysis. Many bar charts and pie charts are made to fetch the information about each player and each stadium.

**Existing System**

Existing system is the collection of the vast data without formatting or anything else. It’s just raw data and there is not much information that can be extracted from that. Though that raw data is important for drawing out the major results and for finding out the major factor. The current existing system needs to be processed to form information. By applying proper formatting and cleaning we can represent the data in an impressive way which would actually help in analysis.

Existing system is the older way to represent the data with the new technology and the new tools available we can create the model and represent the data in new form which is more interactive and seems to be more informative then the existing one. The data needs to be processed fast and needs to provide the detailed view which has been made possible with the help of the new system available to us.

Disadvantage of Current Existing System

* We need to create a relationship between the data available
* Data needs to be processed further to make it more presentable
* The data lacks visualization and data is huge and tedious making it difficult to study.
* There can’t be any results drawn by just looking at the data.
* Data needs conditional splitting in order to determine the teams performance

**Source of Dataset**

* **Kaggle:-** Kaggle is an online community of data scientists and machine learners, owned by Google, Inc. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle got its start by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and short form AI education.
* Wikipedia

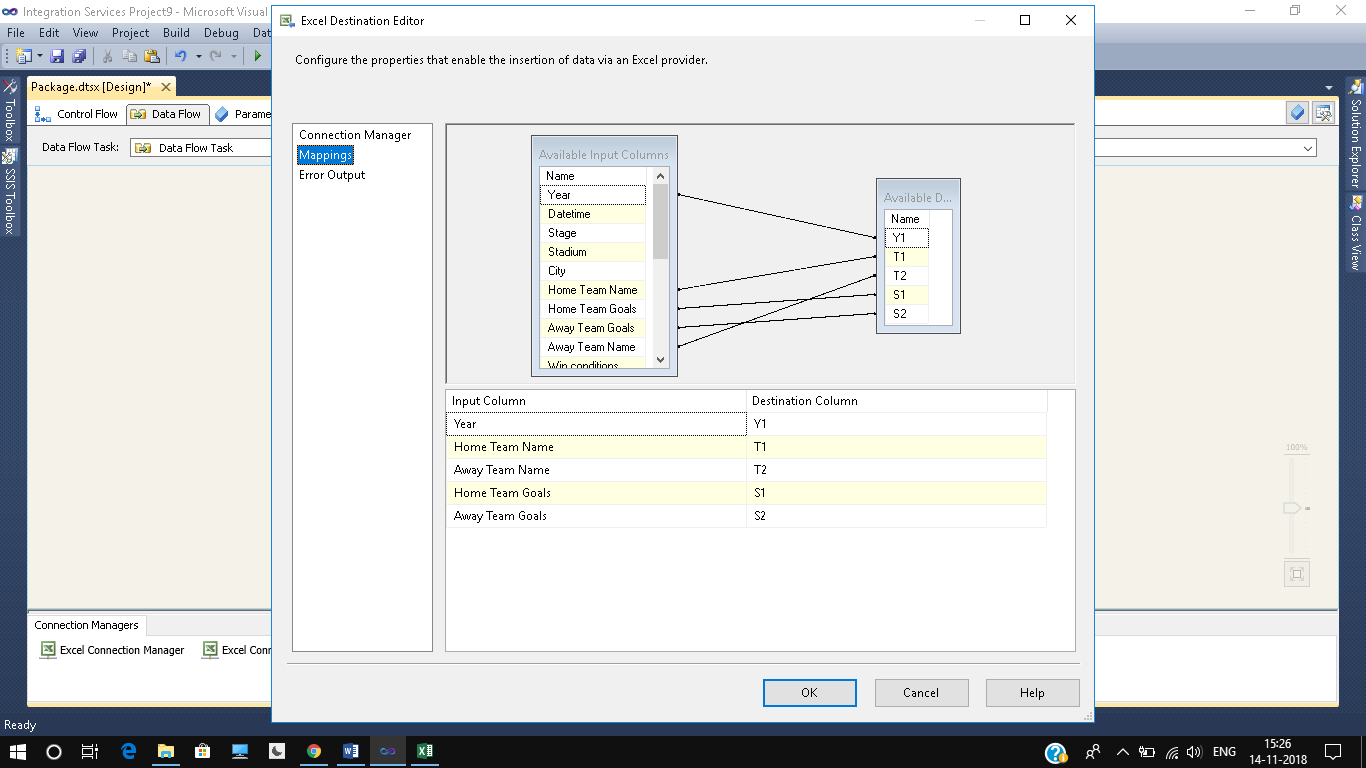
**Extraction Transform and Load**

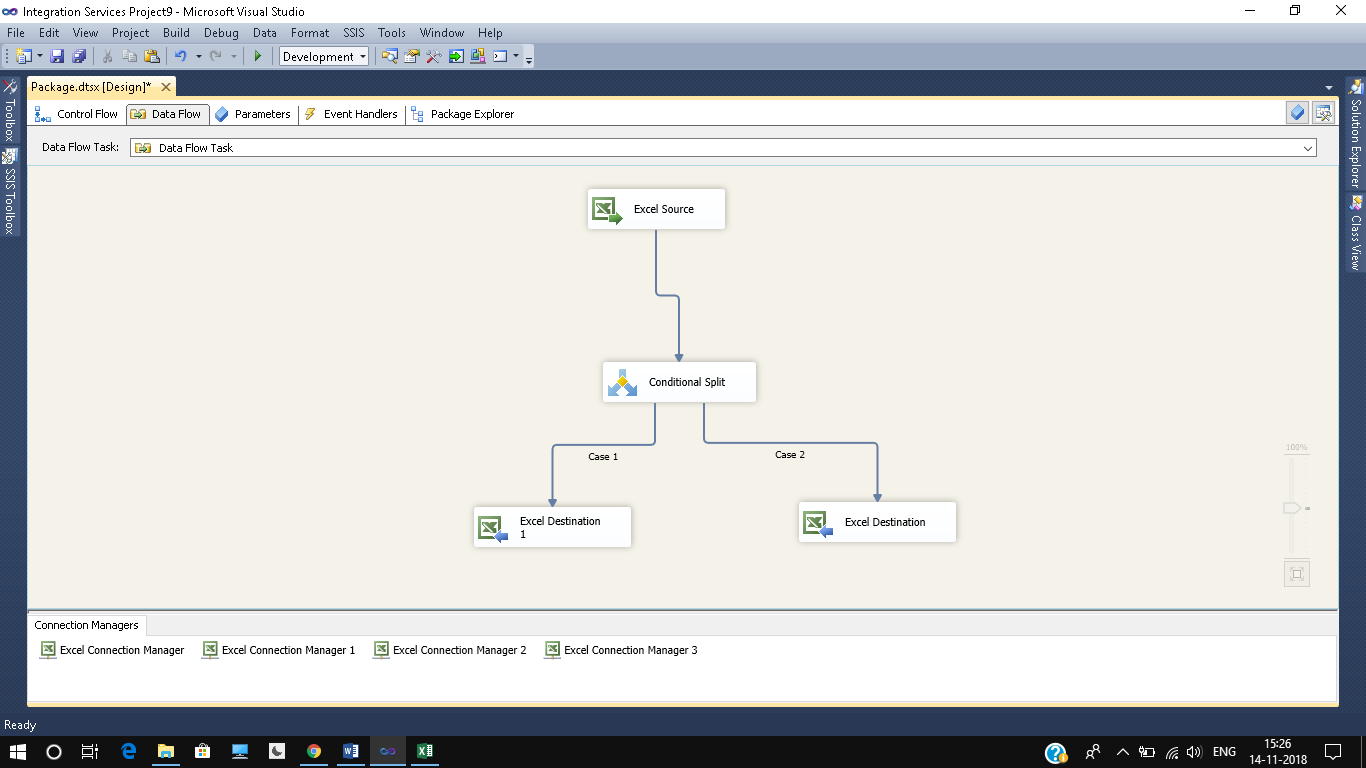
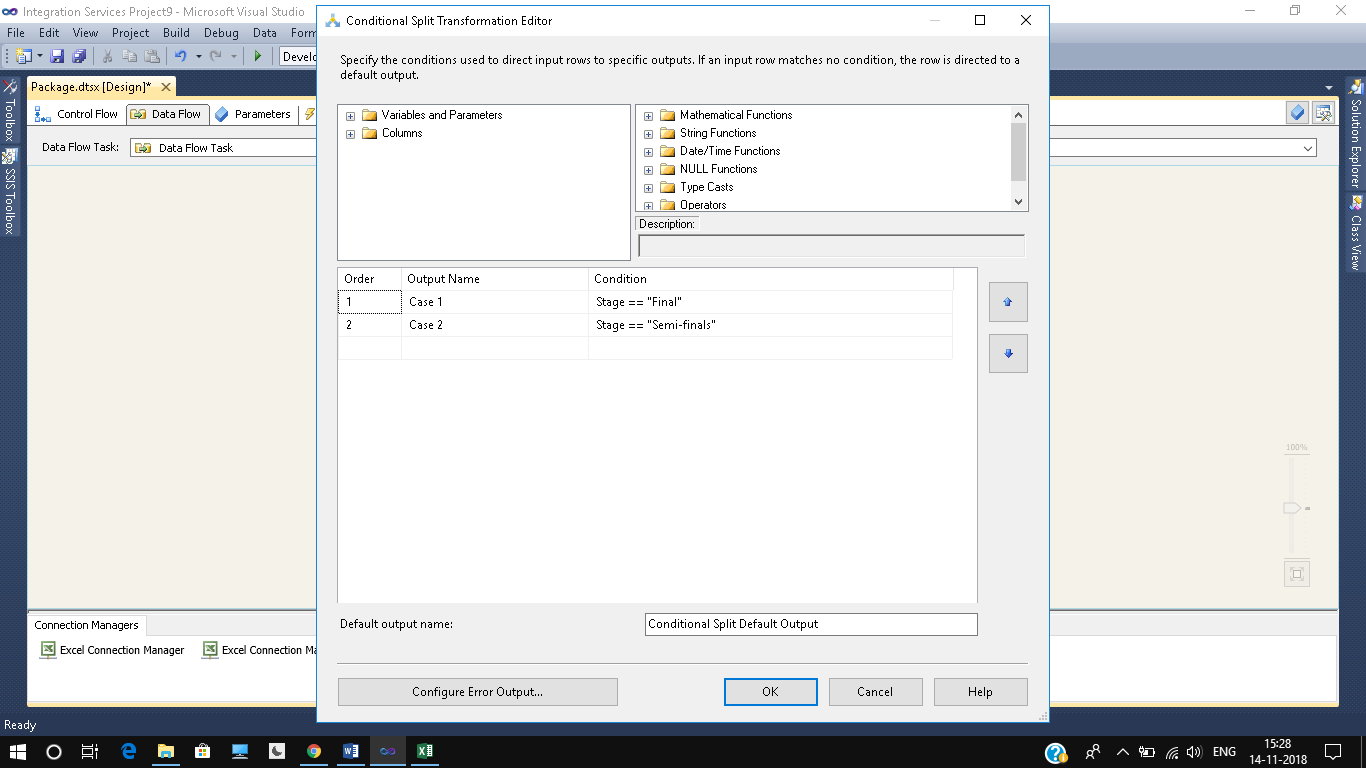
ETL process stands for the Extract, Transform and Load. ETL plays the major role in this analysis also because at first the data obtained from the excel sheet was taken cleaned that is what we call the extraction and transformation. The transformation contains the cleaning of data, formation of tables, formation of pivot tables accordingly and then loading it on to excel again. So the process goes from the excel to excel in this, However the main purpose of ETL in this was to make the data available for the analysis and making it clean.

**Extraction**

Extraction is the process of extracting the data from somewhere but in my case the data was taken from the kaggle in the csv format. It was already in xlsx format so I did not have to do any kind of extraction. The whole process of extracting the data from the web to excel is the process of the extraction.

**Transform**

Transform allows the user to modify, clean and make the data appropriate so that it can be used for the purpose of the analysis and that is why this is considered as the most important part of the data analysis. In my case also the data was not that clean, I had to clean the data where I had to look for the rows which were missing values or were having N/A values. The values were required to be filled with the appropriate values to get the correct result and at the same there were rows which were required to drop to get results. I also have to convert the data from the string to the integer format and have to look for combining the data into one.The data cleaning is not just about cleaning the values and there is more to it in which we have to look after the format of the data. We must classify rows into the strings, numbers, dates and currencies and so on. In order to maintain the proper data format.



**Loading**

Loading means once the data is transformed then it is ready to be processed for the task of analysis. In the excel once the data is cleaned then the data is loaded on to form the pivot table, tables, and power pivot in order to get the desired outputs. Once these results are done then the data is loaded and the next form which are visualization and the other things start taking place. Once the data is cleaned and loaded one can proceed on to with the process of creating visualizing and creating formulas to do analysis and justify it and this also contains the part of the correlation for my project.

**Analysis On Dataset**

Let us look at the objectives which we will be doing in this analysis and then we would jump into the detailed view of it.

Objectives

1. **ANALYSIS OF UMPIRES AND VENUES->** We will analyse the most used venues and most number of times standby Umpire.
2. **ANALYZING BATSMAN PERFORMANCE ->** In this analysis we are going to look at the top 10 run scorers in the IPL.
3. **ANALYZING NUMBER OF TIMES TEAM WON THE TOSS** -> In this we will analyse the number of times team won the toss across all seasons of the IPL..
4. **ANALYZING TOSS BASED DECISIONS ->** In this we will see how many times team won the toss opt for bat or bowl.
5. **SUMMARY OF ALL IPL->** This analysis tells about the Number of matches played in each season of the IPL, total runs made in each season of the IPl.

**Analysis of Umpires and Venues**

**Introduction**

In each match there are two onfield umpires who takes decision about each delivery. Some umpires who are good are used in more matches than the other. Similarily different matches are played in different venues. In this we analyse the number of times a venue is used and umpire work.

**General Description**

The analysis is divided into three different part on the basis of the number of umpires umpired in each seasons. Firstly the runs scored by each player is calculated and then it is sorted for each season. Then a table is created having players with highest runs in each season. A pivot table is made and a bar graph is made for visualisation.

**Specific Requirements, Function and Formulas**

The analysis requires Microsoft excel 2010 or above, requires pivot table and different type of graphs. The function and formula of sum is used to calculate the number of runs scored by each player and then sorting is applied. It requires curiosity and lots of interest in analysis.

**Analysis Result**

The analysis results in the table having players with most runs in each season.

**Visualisation**

Below is visualisation of different players of the analysis.

**Analysi Batsman performance**

**Introduction**

In every IPL many player play and perform to the best of their ability. Some players score many runs and receive many man of the match awards. In this analysis we will see the top 10 run scorer of IPL and top 5 players with most man of the match award.

**General Description**

Players score many runs and receive many man of the match awards. In this analysis we will see the top 10 run scorer of IPL and top 5 players with most man of the match award.

**Specific Requirements, Function and Formulas**

The analysis requires Microsoft excel 2010 or above, requires pivot table and different type of graphs. The function and formula of sum and countis used for calculation of number of wins. It requires curiosity and lots of interest in analysis.

**Analysis Result**

Analysis Gives top 10 players with most runs in the ipl and top 5 players with most man of the match award.

* + 1. **Visualisation**

Below is visualisation of the analysis.

**Analysis Of Win Percentage Based on Batting First or Fielding**

**Introduction**

A team can either opt for batting or fielding on winning the toss. In this section we will analyse the percentage of times a team won batting first and fielding first.

**General Description**

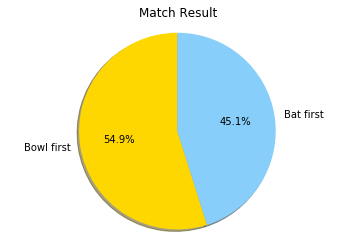
In this we will first calculate the number of times a team won batting first and fielding first and then we divide the result by 100 to get the percentage.

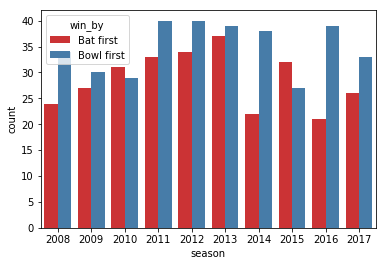
**Specific Requirements, Functions and Formulas**

The analysis requires Microsoft excel 2010 or above, requires pivot table and different type of graphs. The function and formula of sum and product is used for calculation of percentage. It requires curiosity and lots of interest in analysis

**Analysis Result**

Percentage of times a team won based on its decision of batting first or fielding first. A team can bat first or bowl first. Based on its decision the percentage of times a team won while batting first and bowling first.

**Visualization**

****

**Analysis Of Toss result**

**Introduction**

A team can either opt for batting or fielding on winning the toss. In this section we will analyse the number of times a team opt to bat or ball on winning the toss.

**General Description**

In this we will first calculate the number of times a team won batting first and fielding first and then we divide the result by 100 to get the percentage.

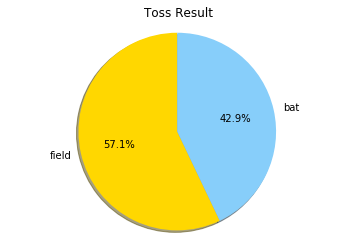
**Specific Requirements, Functions and Formulas**

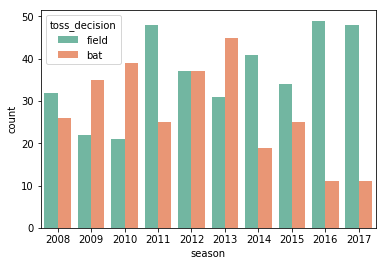
The analysis requires Microsoft excel 2010 or above, requires pivot table and different type of graphs. The function and formula of sum and product is used for calculation of percentage. It requires curiosity and lots of interest in analysis

**Analysis Result**

Teams decision on winning the toss. Before every match there is toss. On winning the toss a team can either opt for bat or field. This analysis shows the percentage of times a team opt for bat and bowl.

**Visualization**

****



**Analysis Of Summary of IPL**

**Introduction**

In this we will see the number of matches played in each IPL and the number of runs scored in each IPL.

**General Description**

In every Ipl there are number of matches played and runs scored. Here we will analyse the number of matched played and runs scored.

**Specific Requirements, Functions and Formulas**

The analysis requires Microsoft excel 2010 or above, requires pivot table and different type of graphs. The function and formula of sum and product is used for calculation of percentage. It requires curiosity and lots of interest in analysis

**Analysis Result**

Number of Fours and Sixes hit by a batsman.

**Visualization**

**List of Analysis with Result**

From all the analysis we have done so far, we have seen certain results which are quite evident from the visualizations that we have created. The first and foremost part was to make attractive pivots charts. That made the data representation quite easy to figure out. I made five analysis .Now I have detailed results which tells about the trend.

1. Chris Gayle and David Warner has made highest runs two times in the IPL.

2. Mumbai Indians has the most number of wins.

3. Season 2013 has the most number of matches.

4. 54.9 % of times bowling first team won the match and 45.1% batting first team won.

5. 57.1% of times team opt for fielding after winning the toss and 42.9% times team opt for batting after winning the toss.

6. Chris Gayle has the most number of sixes.

**Future Scope**

After analysing the various aspects of the IPL where the project can easily portrait the cricket data , it will be very easy for the future visualization.

* By adding the more data and making the the charts more “DYNAMIC” , it will become easy for us to visualize the better performance of the teams.
* Teams performance can easily be compared with the upcoming matches and it will be easy for predicting the winners for the matches.
* Teams performance with the styles and strategies will prove beneficial for prediction and analysing the teams future performances.
* Linear regression model can be deployed from the available information and many new relationship can be generated using power-pivot , if one need .

**References**  
• [www.google.com/sportkeeda.com](http://www.google.com/sportkeeda.com)

• [www.youtube.com/datasetipl](http://www.youtube.com/datasetipl)