



MSIS-2621- BUSINESS INTELLIGENCE AND DATA WAREHOUSING

FINAL PROJECT REPORT

INDIAN PREMIER LEAGUE



SUBMITTED BY :
Megha Sharma
Prachi Tamhankar
Sarthak Biyani
Yugal Karamchandani

TABLE OF CONTENTS

Sr. No	TOPIC	PAGE NUMBER
1.	Introduction and business scenario	3-4
2.	Swim Lane Diagrams	5-7
3.	Data Models	8-10
4.	Architecture Diagram	11
5.	ETL Mappings	12-14
6.	Dashboards	15-20
7.	Key Learnings	21
8.	References	22

What is Indian Premier League?

Indian Premier League is a Twenty20 cricket tournament been played in India founded by Board of Cricket Control in India (BCCI) where various franchise-based teams compete in a league tournament against each other to emerge as the winner of the cricket tournament. The tournament is held in cricket stadiums viewed by a large audience of people in stadium and over the television broadcast both.

Various teams play in the league stages and the top four teams qualify for the playoffs, out of the playoffs the top two teams play the final of the tournament.

In this twenty over tournament, first toss takes place to decide which team will bat first and which team will bat second. There are two teams with eleven players in the side and these eleven players are a mix of Batsmen, Bowlers, Wicket Keeper and All-rounder. Each team bats for the twenty overs in an innings, every over there are six legal balls that are bowled by the bowling team and the batting team tries to score the maximum runs in the innings and sets a target for the other team. Once the batting team finishes its innings the team that bowled starts to bat for twenty over and tries to chase its target set by the team batting first, if the team chases the target within the twenty overs it emerges as the winner but if the team is not able to chase the set target, the team that batted first wins the match.

Business Scenario- We have analyzed the scenario in perspective of an IPL team owner considering the Business Processes mentioned below. The owners face challenges in identifying the players performances, Price of the Match Tickets of each stadiums that generate revenue for the franchise, which players to retain and which not to and Analysis of Toss Results of each stadium.

We have four different Business Processes for the Owners of Team of Indian Premier League: -

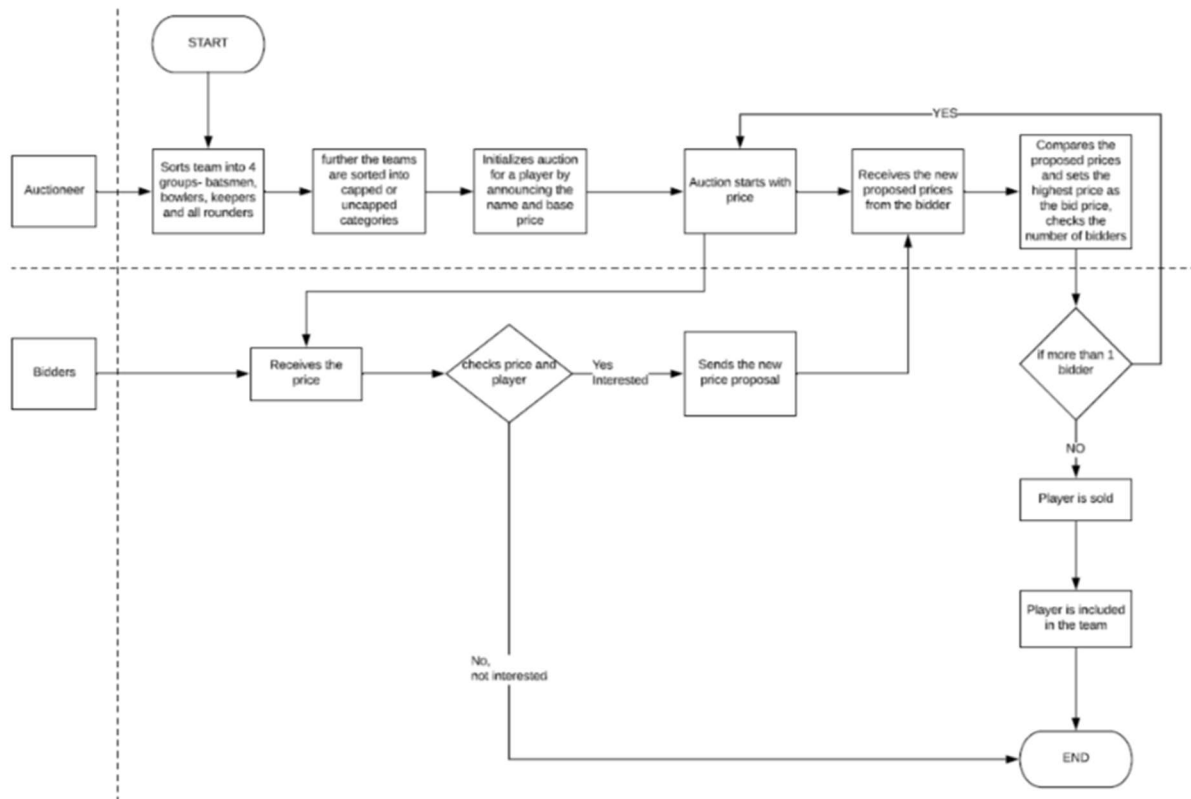
1. **Descriptive Analysis of Player Performance-** For more than 500 players who have played or are currently playing in the league we have performed analysis of their performances. For the batsman, we have calculated the total number of runs scored and their strike rate over the seasons and for the particular seasons. For the bowlers, we have calculated the total number of wickets taken and their average over the seasons and for the particular season's user want to view. This would make easier for the team owners to choose players from the entire list they want to keep in their teams.
2. **Predictive Analysis of Price for Auction of Players-** Different franchises bid for different players to form a strong and desirable team that can win the tournament for them. The starting auction amount of each player is decided by the player itself and then based on the highest amount spent by franchise, that franchise gets the player into its team. The same auction amount is paid by the franchise to the players at the end of tournament. We have predicted in the dashboard whether a player's price will go up or down based

on their total performance and performance over the season. Dashboard for such scenarios help team owner if they want to increase their bet on the same player and try to go for some other player in the auction.

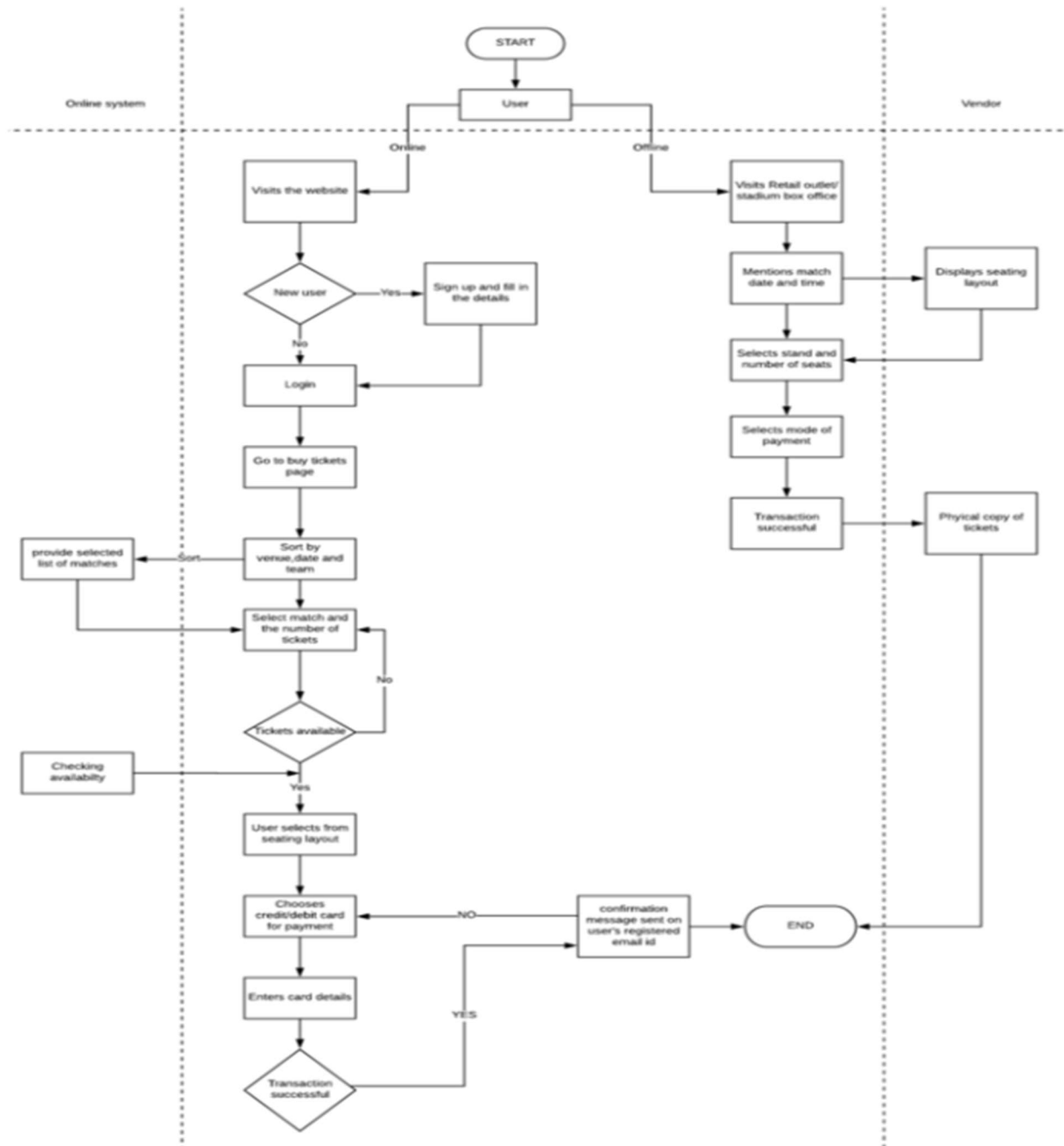
3. **Prescriptive Analysis for Retention of Players-** Every franchise can retain up to 3 players based on the performance so the franchise owners can set up the criteria such as number of runs and wickets taken to find which players to retain and which not to retain.
4. **Analysis of Toss Results-** This marks as the commencement of the match and the two team captains along with the match referee gather in the center of the stadium for the toss where it is decided that which team will bat first and which will bat second. The captain that wins the toss is given the choice to choose either to bat first or bat second. Toss plays crucial role in the results of the match as the captains can choose based on the pitch conditions and players in the team batting first is their stronger aspect or batting second is aspect. Old toss results will provide an insightful analysis about the stadium on which stadium to bat first and which to bowl first.

SWIM LANE DIAGRAMS

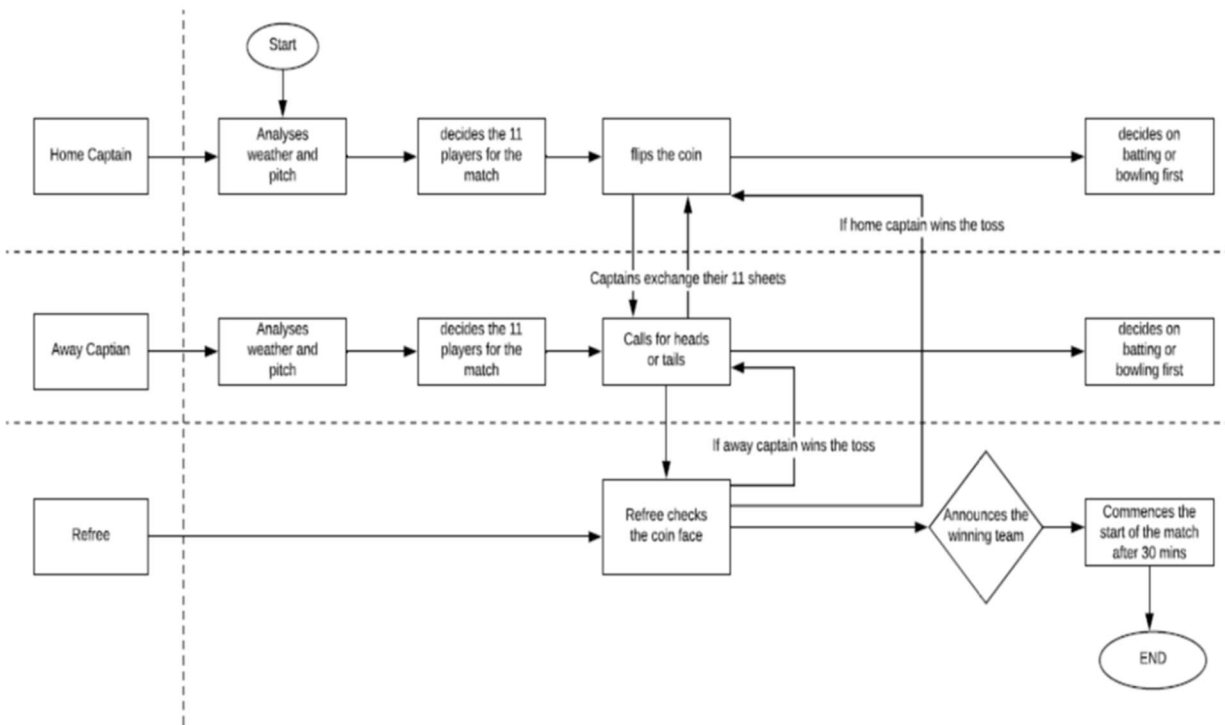
1. Player Auction



2. Match Tickets

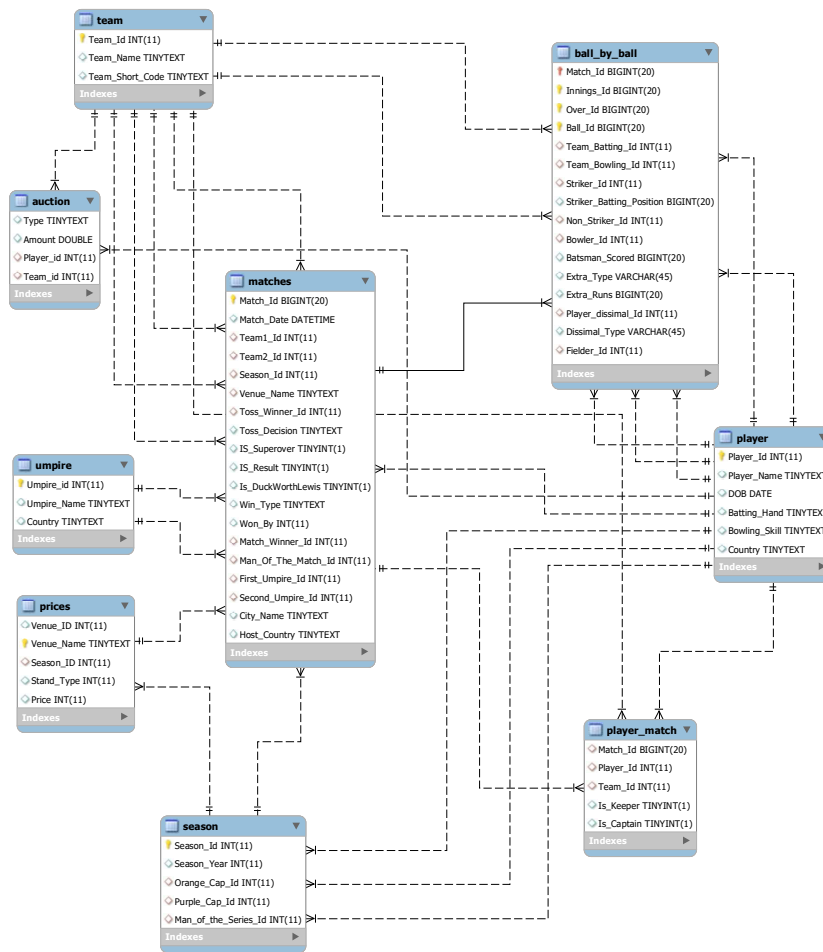


3. Toss of Match

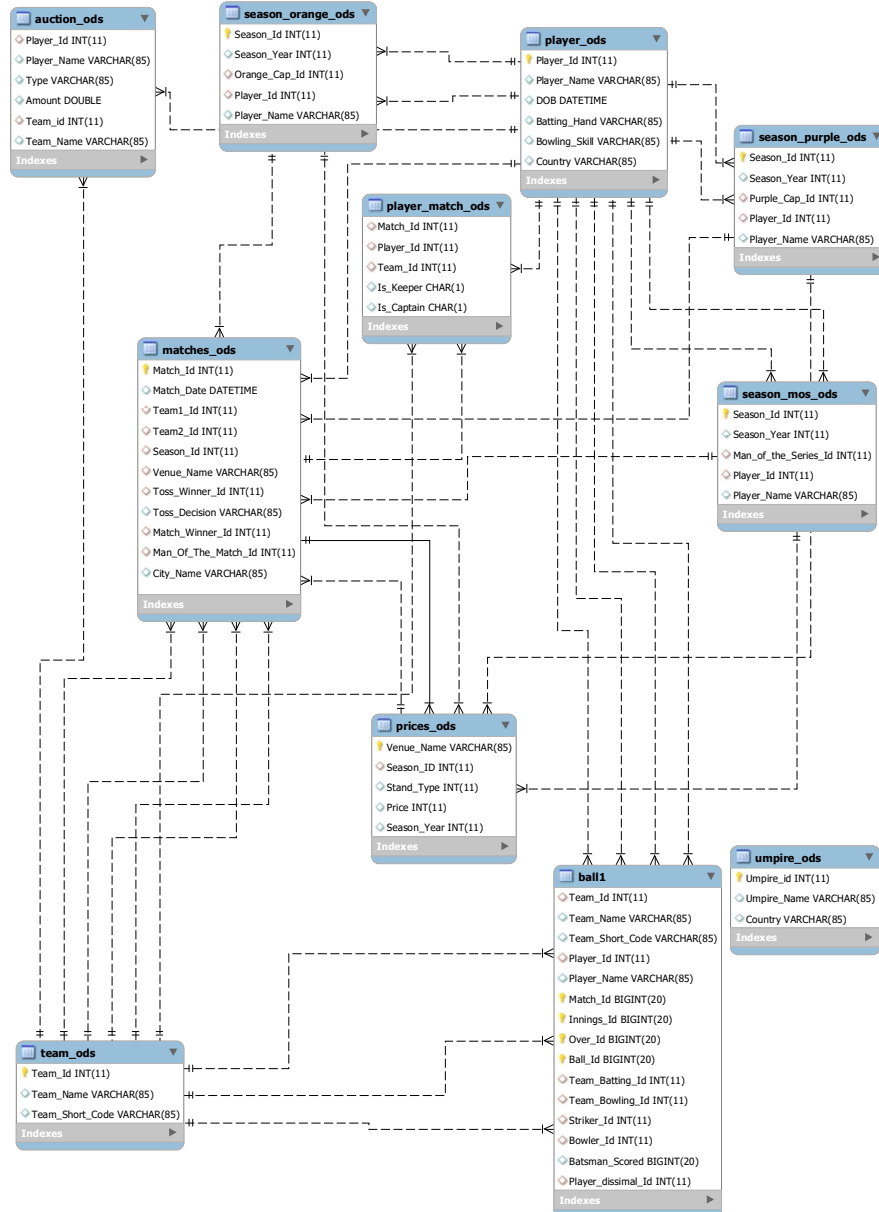


Data Model Diagrams-

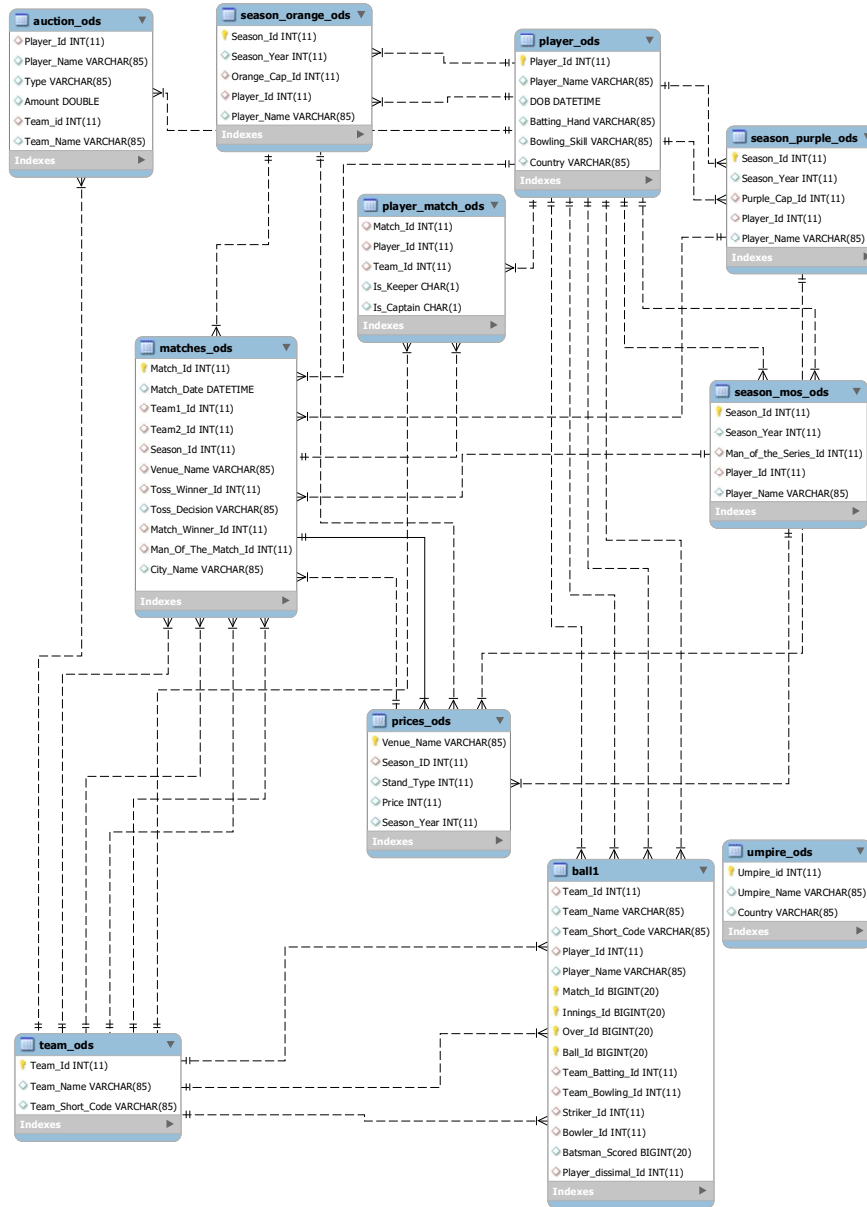
1. Online Transactional Processing Database (OLTP)-



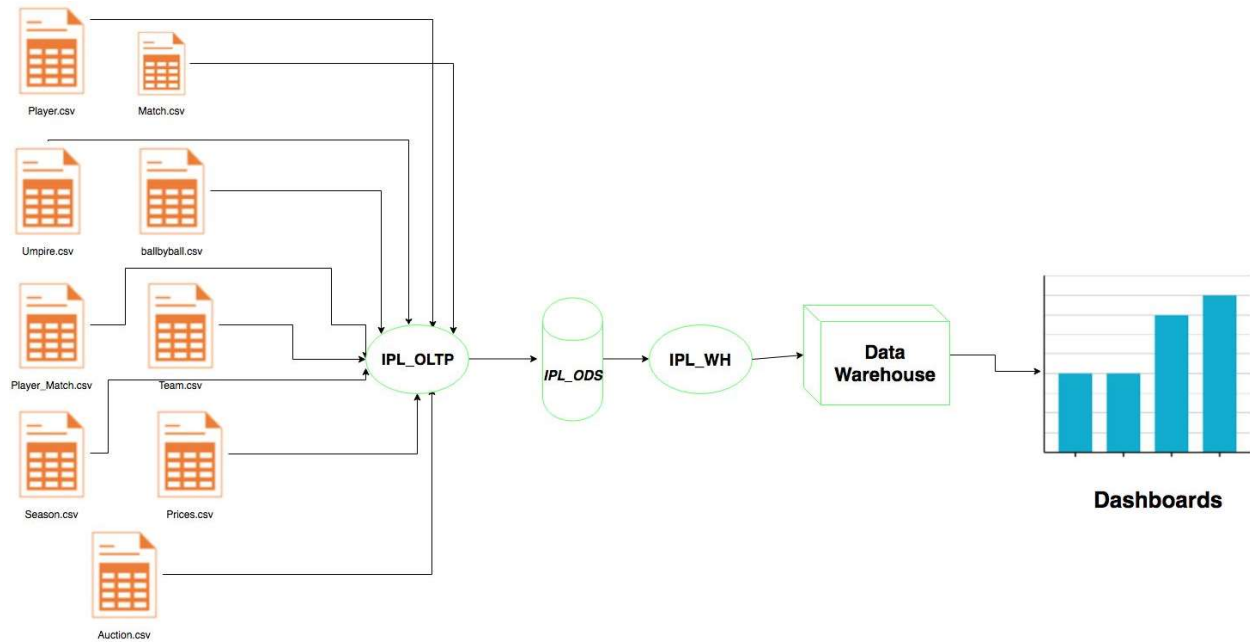
2. Operational Data Source



3. Warehouse Database



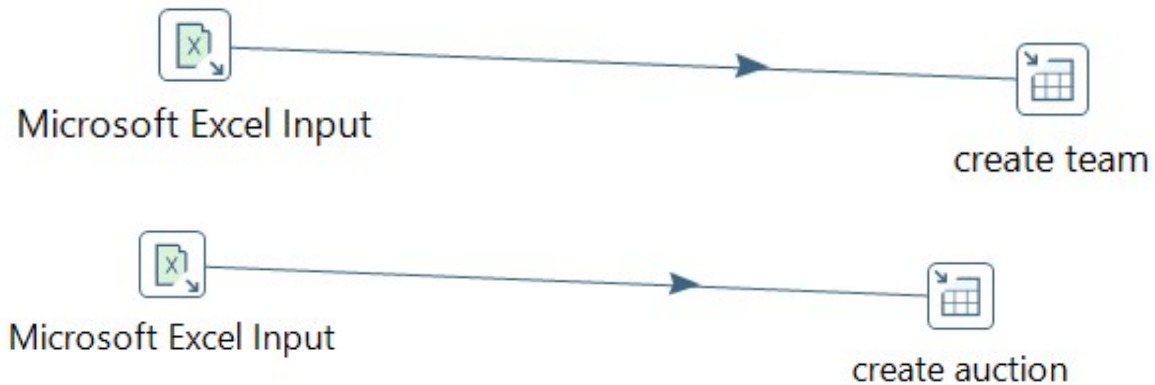
ARCHITECTURE DIAGRAMS



1. Initially we Took input files and loaded Data from CSV and Excel files into our database.
2. Designed an ODS database by joining multiple tables and keeping only the essential attributes which were required to create the dimension table and fact table for the Analysis.
3. Designed a Data Warehouse where we loaded the dimension tables from the ODS database keeping in mind different dimensions and facts for the purpose of analysis.
4. Created a fact table inside the Warehouse where only facts which were to be shown in Tableau were stored.

ETL Mappings

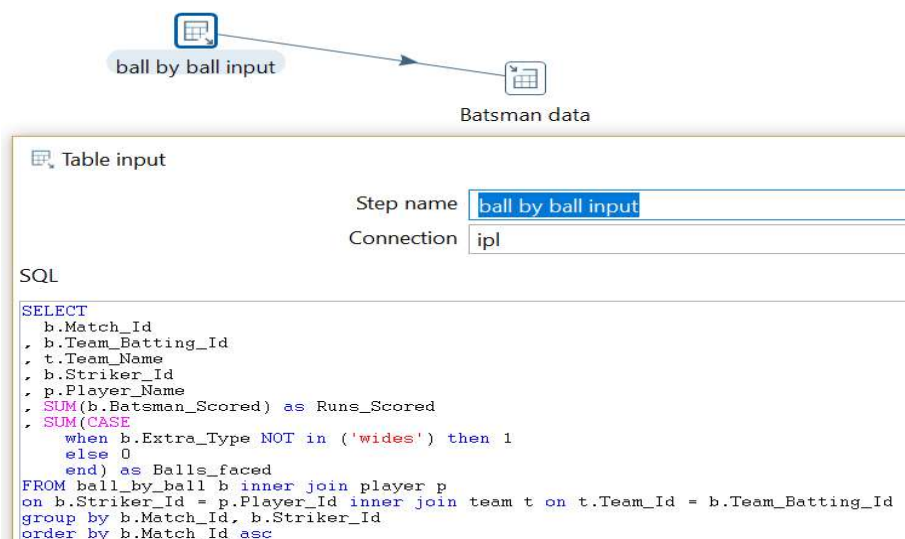
1. Creating the Online Transactional Processing Database by loading the input csv and excel files into it.



Here we can see how we uploaded the raw data from the csv files to create tables for the OLTP database. We performed the same steps for all the 9 CSV files.

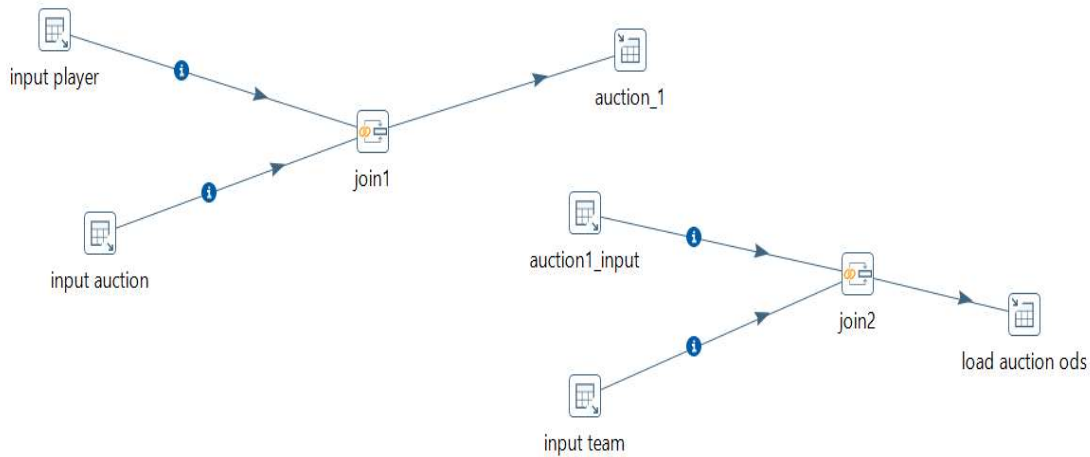
2. Generated the ODS tables from OLTP database where we have used various SQL queries for different conditions.

Example 1



We used SQL queries to calculate the runs scored by every batsman in every match for every season. We similarly calculated the number of wickets taken by bowlers for all the matches and generated these tables for ipl_ods.

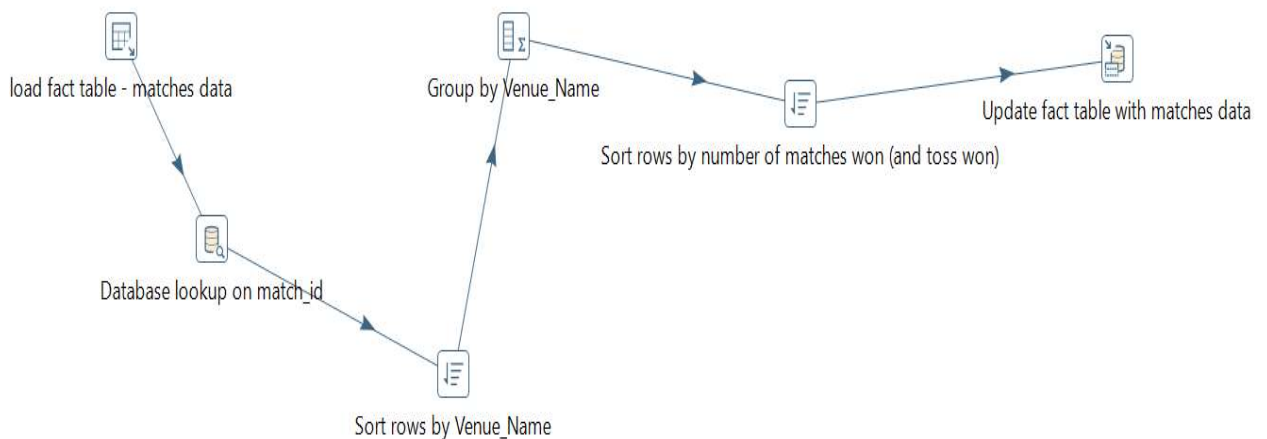
Example 2



We used 2 separate joins, first to load the auction data from OLTP Database and then created a join on auction data table and team table to create auction_ods in the ODS database.

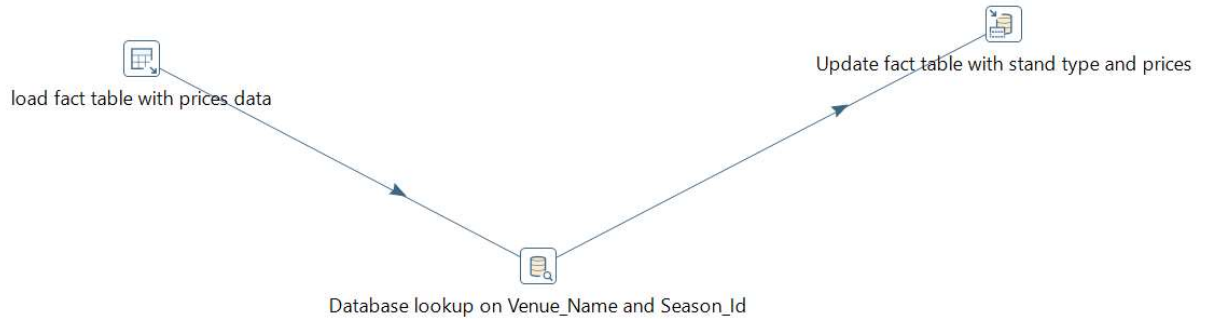
3. Loaded the Data Warehouse using the multiple ODS tables keeping only the dimensions needed for final analysis and their analysis factors in fact table.

Example 1



We used database lookup on table matches data, sorted the rows and grouped them by venue name to further sort them by number of matches won by the team that won the toss to create our fact table in ipl_wh for the same.

Example 2

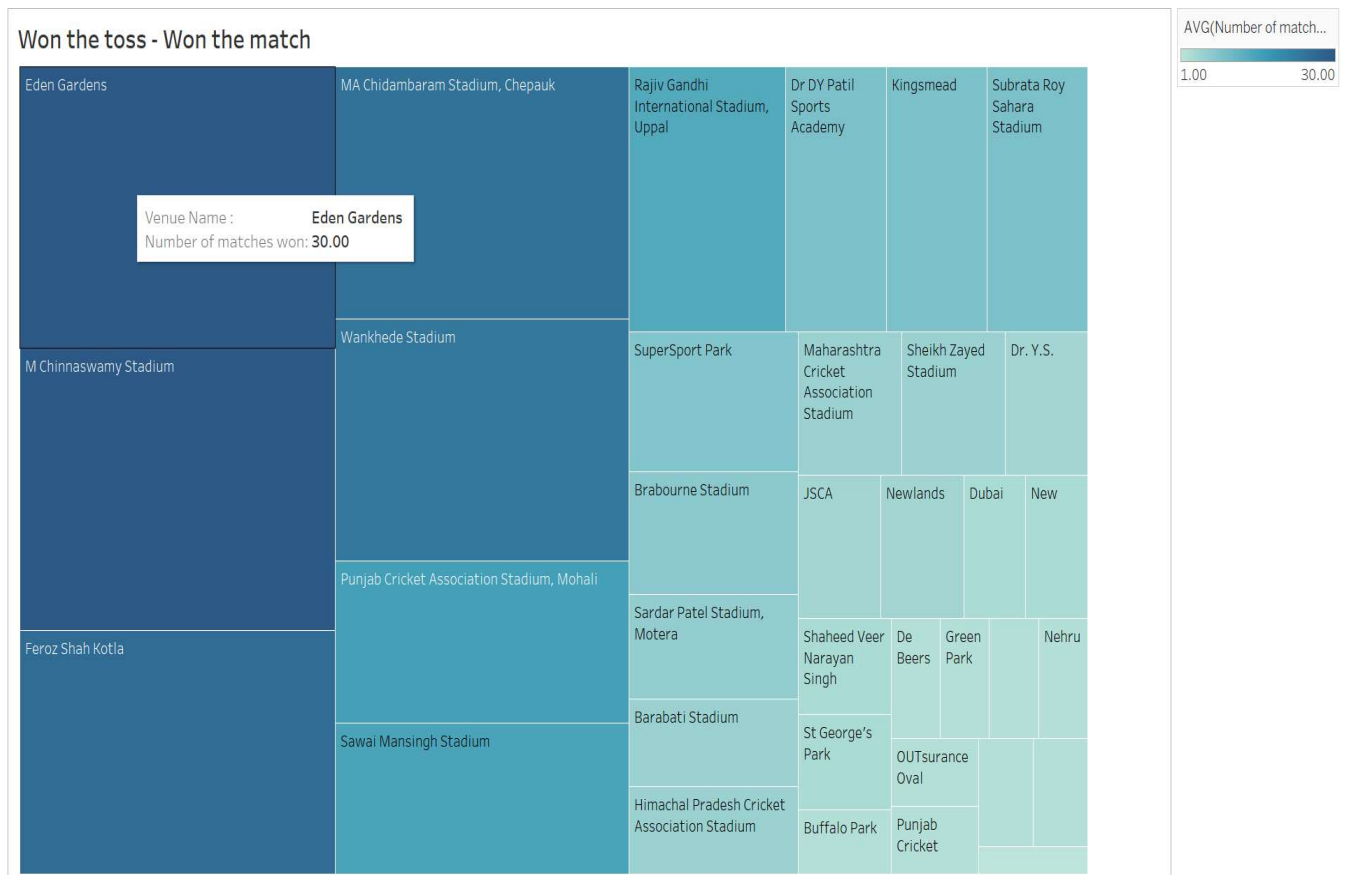


In this transformation, we created a lookup on venue name and season id for the price data fact table to update the table in ipl_wh where it would show the price per stand type for the matches from all seasons.

DASHBOARD SCREENSHOTS

1. Dashboard for Toss Scenario

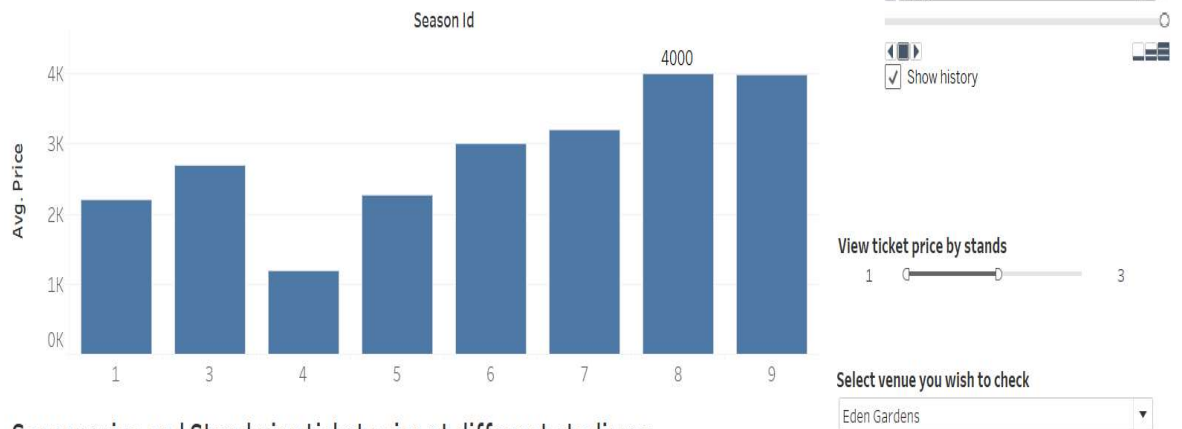
Here we have shown the heat map for the stadiums of IPL which indicates on which stadium the teams that have won the toss, have also won the match, this would help the teams to understand what things should be done on which stadium, whether they should bat first or bowl first.



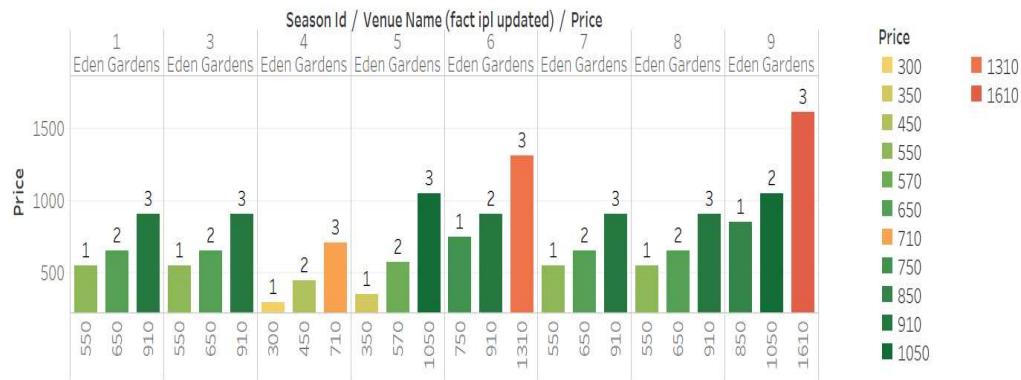
2. Dashboard for prices of match tickets

The bar chart in the dashboard denotes the average price of a ticket in an IPL season every year. Moreover, this dashboard provides full control to the user to check the prices of different stands at any venue for all the seasons.

Season wise average ticket price

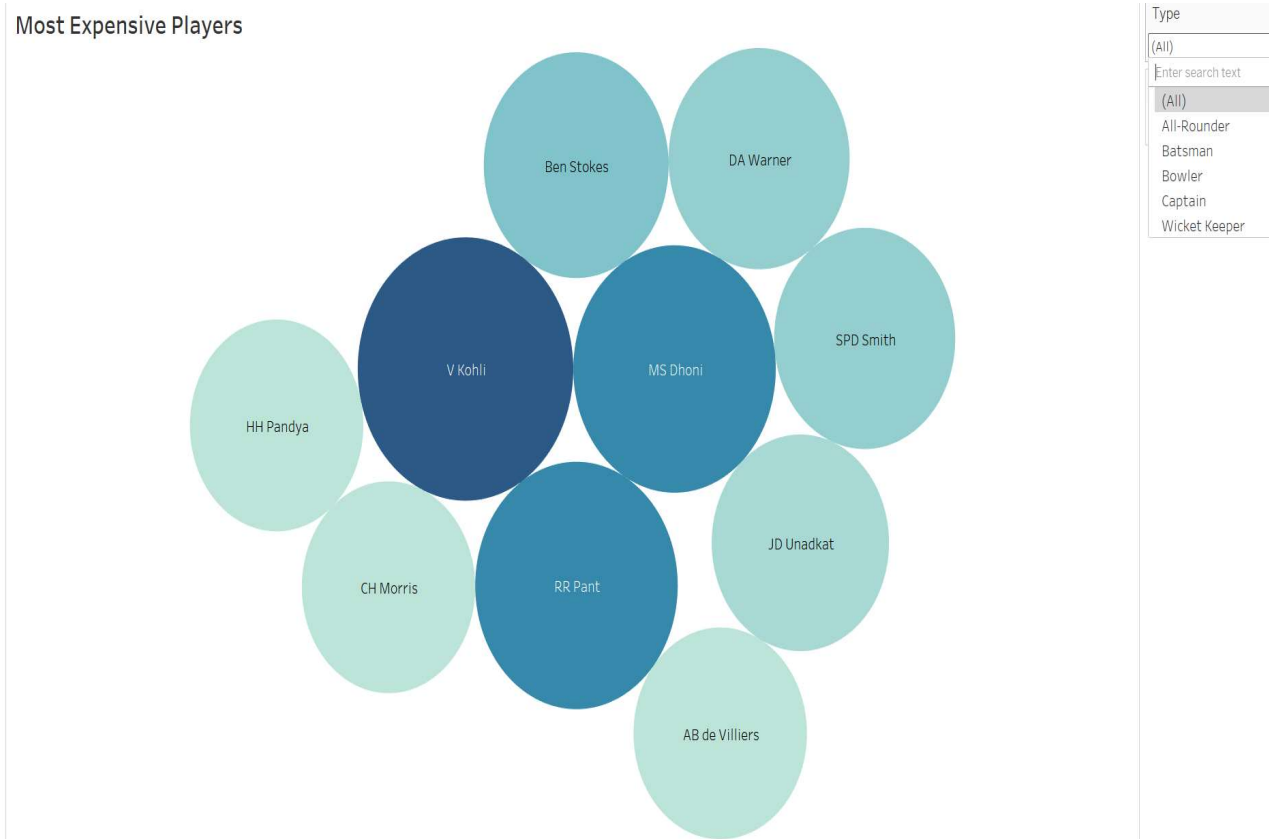


Season wise and Standwise ticket price at different stadiums



3. Dashboard for Most Expensive Player

The below bubble chart shows the most expensive buys of IPL auction 2018. It also gives user control to check the most expensive players according to the type of players. For example, user can view top 10 expensive batsman, bowlers, all-rounders and so on.



4. Dashboard for Player Analysis

This dashboard provides the player statistics of all seasons. The user can view highest run scorers and wicket takers till now. Also, the user can check the same stats for each season to know the Orange Cap winner (highest runs in a season) and Purple Cap winner. Moreover, the user can view stats for players for each team for all the seasons or any one season. It also displays the average strike rate and average bowling rate for each batsman and bowler respectively.

IPL Player Analysis

Teams

Team Short Code	
CSK	Chennai Super Kings
DC	Deccan Chargers
DD	Delhi Daredevils
GL	Gujarat Lions
KKR	Kolkata Knight Riders
KTK	Kochi Tuskers Kerala
KXIP	Kings XI Punjab
MI	Mumbai Indians
PW	Pune Warriors
RCB	Royal Challengers Bangalore
RPS	Rising Pune Supergiants
RR	Rajasthan Royals
SRH	Sunrisers Hyderabad

Season Id

(All) ▼

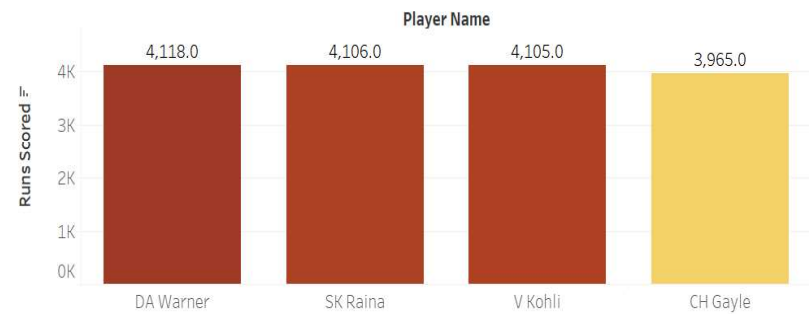
Select number of batsman

1 4
00

Select number of bowlers

1 8
00

Highest Run Scorers



Highest Wicket Takers



5. Batsman/Bowler Retention

This dashboard is designed for owners who wish to check stats of their players of all the seasons. Before a new season starts, the owners can put some runs criteria he/she wishes to check, and then it shows a list of players with recommendations whether a player should be retained or not based on the filter value entered by the owner/user.

Batsman Retention		
Player Name	Batsman Retention Prediction	Runs Scored
V Kohli	Batsman can be retained	4,105
MS Dhoni	Should not be retained	3,270
RR Pant	Should not be retained	198
Ben Stokes	Should not be retained	
DA Warner	Batsman can be retained	4,118
SPD Smith	Should not be retained	1,241
JD Unadkat	Should not be retained	3
AB de Villie..	Should not be retained	3,259
CH Morris	Should not be retained	285
HH Pandya	Should not be retained	156
KL Rahul	Should not be retained	725
MK Pandey	Should not be retained	1,827

AGG(Batsman Retentio...
☒ (All)
☒ Batsman can be reta..
☒ Should not be retain..
Runs Criteria for Batsm...
4,000

Bowler Retention		
Player Name	Bowler Retention Prediction	Wickets taken
V Kohli	Should not be retained	4.0
MS Dhoni	Should not be retained	
RR Pant	Should not be retained	
Ben Stokes	Should not be retained	
DA Warner	Should not be retained	
SPD Smith	Should not be retained	0.0
JD Unadkat	Bowler can be retained	28.0
AB de Villie..	Should not be retained	
CH Morris	Bowler can be retained	33.0
HH Pandya	Should not be retained	4.0

AGG(Bowler Retention ...
☒ (All)
☒ Bowler can be retai...
☒ Should not be retain...
Wickets Criteria for Bo...
20

6. Predictive analysis for upcoming auction

We created one calculation categorized on price of players in previous seasons. According to their price, the stats criteria is defined, and this dashboard predicts whether a player's price should increase or remain same in the upcoming auction.

Next season's price p
X

```

IF AVG([Amount])>100000000 AND (SUM([Runs Scored])>[Auction price predictor criteria f
ELSEIF AVG([Amount])>60000000 AND AVG([Amount])<100000000 AND (SUM([Runs Scored])>[Auc
ELSEIF AVG([Amount])<60000000 AND (SUM([Runs Scored])>[Auction price predictor criteri
ELSE "Same"
END

```

Score Well, Earn Higher

Player Name	Next season's price p..	Avg. Amount	Runs Scored	Avg. Batting Strike Rate	Wickets taken	Avg. Bowling Average
V Kohli	Increase	170,000,000	4,105	116	4	10
MS Dhoni	Same	150,000,000	3,270	129		
RR Pant	Same	150,000,000	198	105		
Ben Stokes	Same	125,000,000				
DA Warner	Increase	120,000,000	4,118	117		
SPD Smith	Increase	120,000,000	1,241	132	0	
JD Unadkat	Increase	115,000,000	3	55	28	14
AB de Villie..	Same	110,000,000	3,259	132		
CH Morris	Increase	110,000,000	285	116	33	18
HH Pandya	Increase	110,000,000	156	105	4	11

AGG(Next season's pric...

(All)

Auction price predictor ...

5,000

Auction price predictor ...

0

Key Learnings from the Project

1. Learnt how to handle high volume and variety of data altogether.
2. Learnt how to use SQL queries along with the cases inside ETL tool
3. We found that visualized data provided more insights than data represented in numeric and string form.
4. Swim Lane diagrams made it easier to understand the process as compared to directly generating the data model diagram for the same.

REFERENCES

<https://www.iplt20.com>

https://en.wikipedia.org/wiki/Indian_Premier_League

www.data.world.com