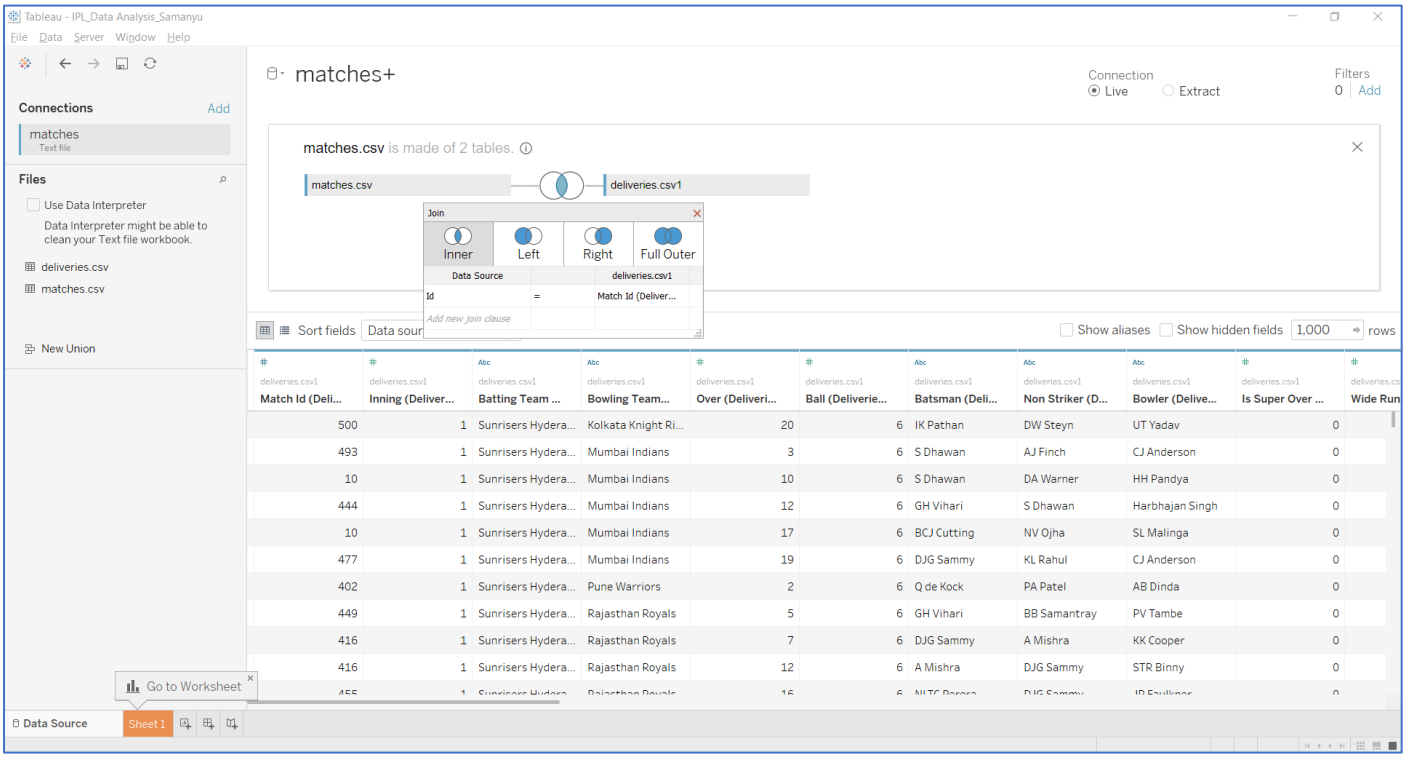
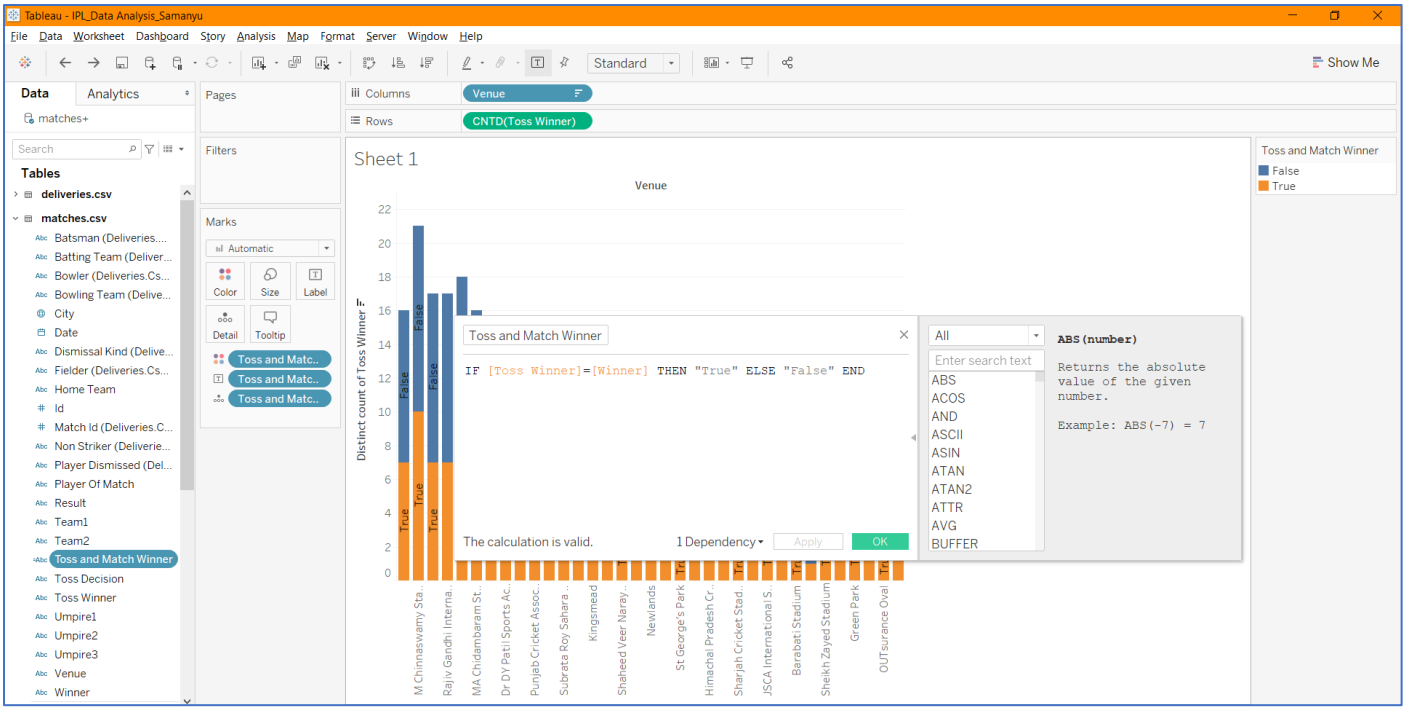
Tableau Assignment_IPL data set



1st we use **Inner Join** to join 2 tables (matches and deliveries) using common column **id(matches)** with match **id(deliveries)**

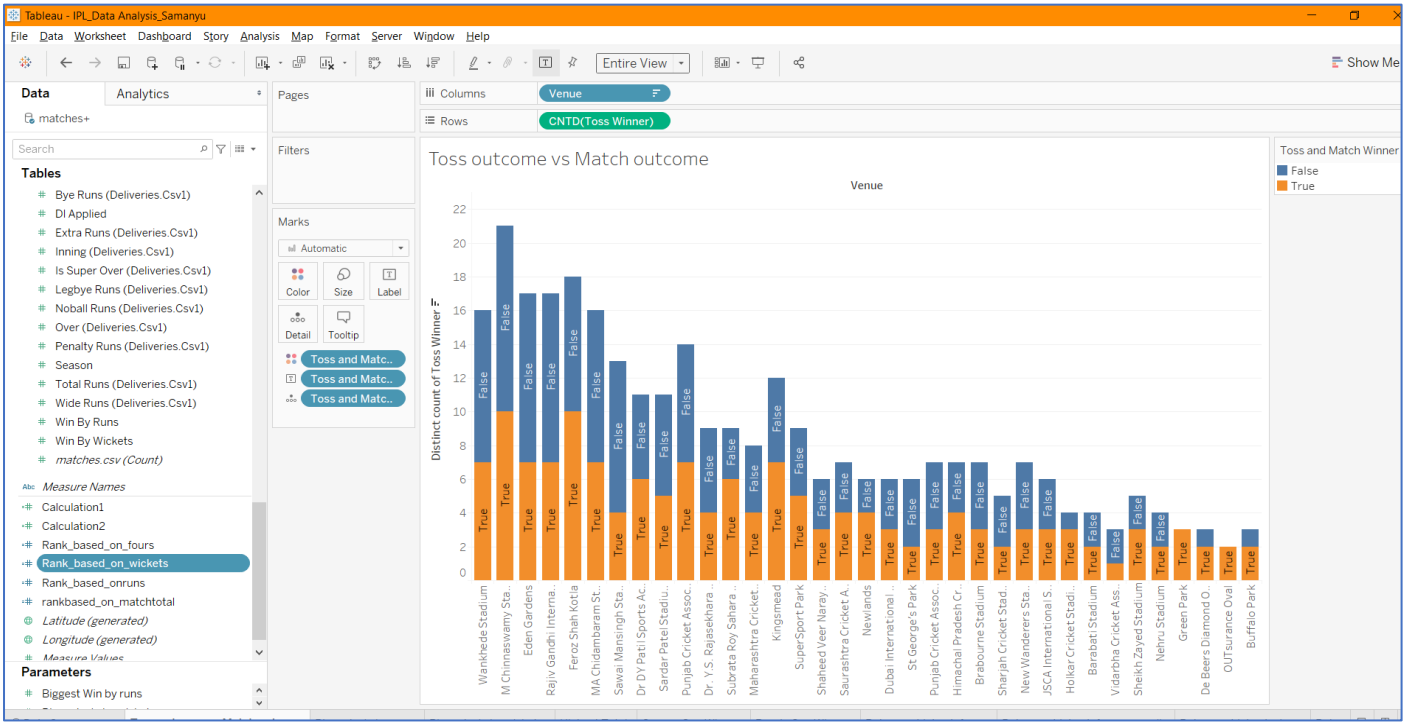
1. Toss_outcome vs Match_outcome(for each ground/venue)



I have created calculated field named as 'Toss and Match Winner' to find out if the toss_winner is indeed the match_winner. If it matches, I have considered it as True else False for each venue

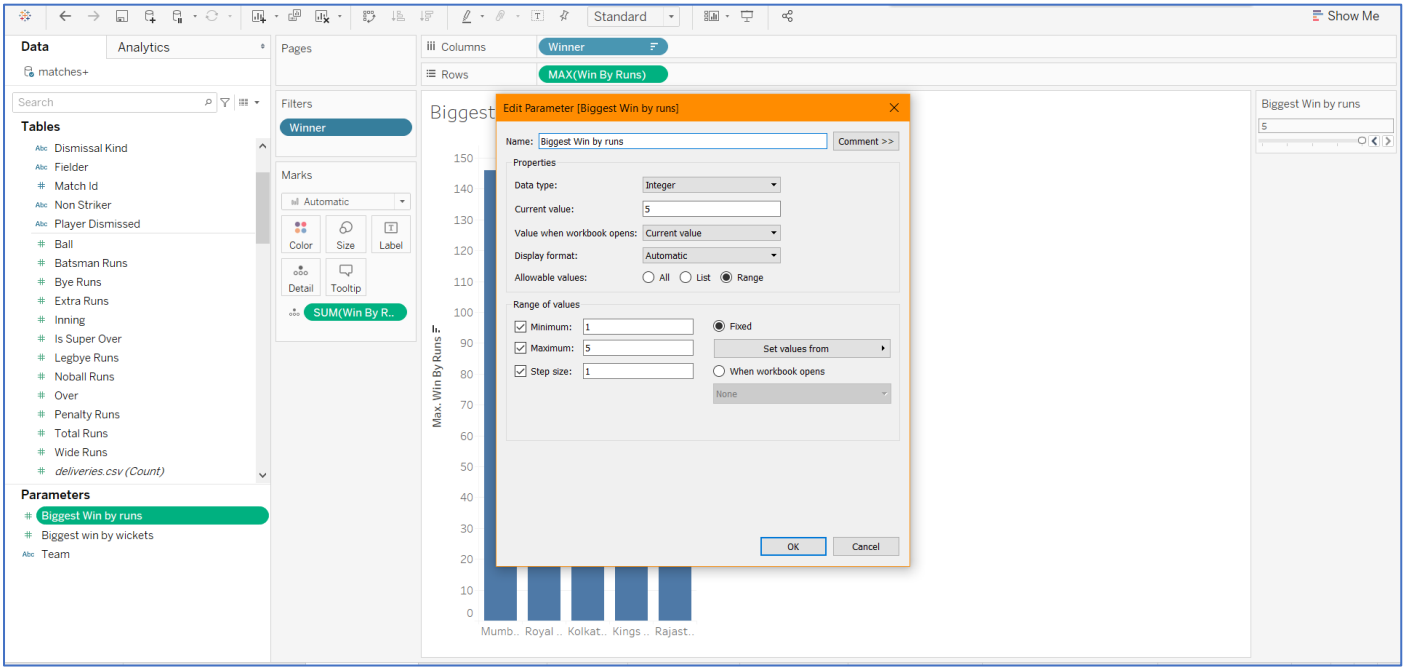
IF [Toss Winner]=[Winner] THEN "True" ELSE "False" END

I have found that almost in all the cases where the team who won the toss also won the match for each of the venues as given below

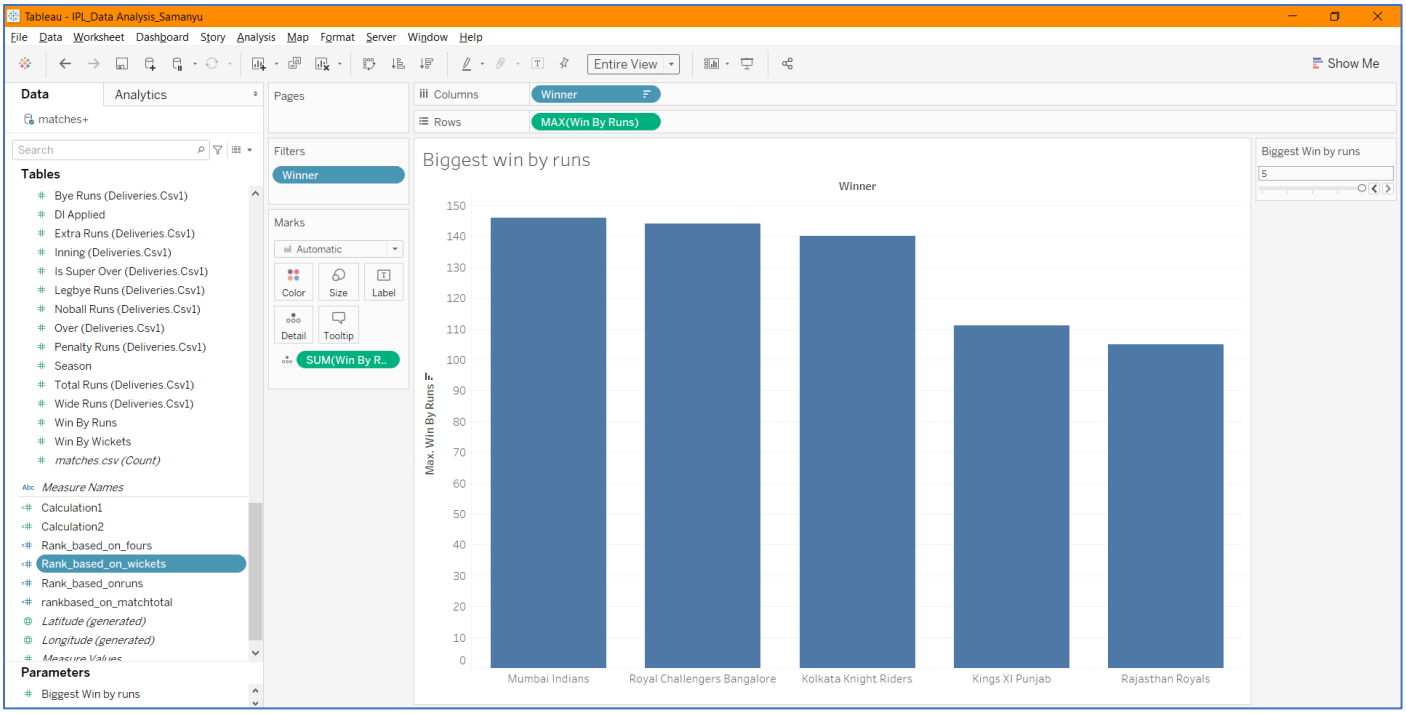


2. Biggest win by runs

First I have created filter by dropping 'Winner' to the column, max of 'Win by Runs' in rows, drop 'Win by run' in detail with measure as 'sum' and then created a parameter 'Biggest Win by runs' with data type as 'Integer', current value as 5 and min, max range as 1 and 5.

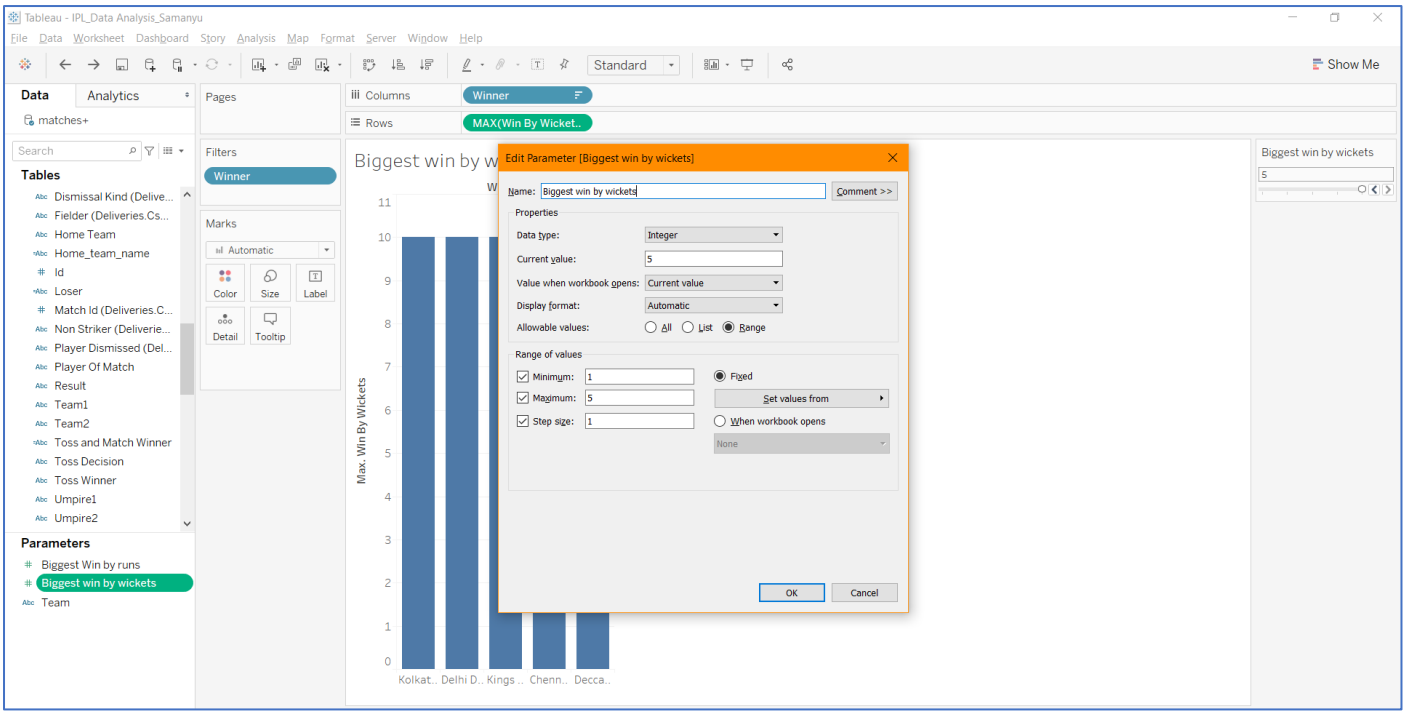


As I have limited scope to show all the winners by runs, so I have limit this to top 5 biggest win by runs and that can be toggle as per the user choice. The top 5 teams with biggest wins by runs are **Mumbai Indians, Royal Challenger Bangalore, Kolkata knight, Kings XI Punjab and Rajasthan Royals** as given below

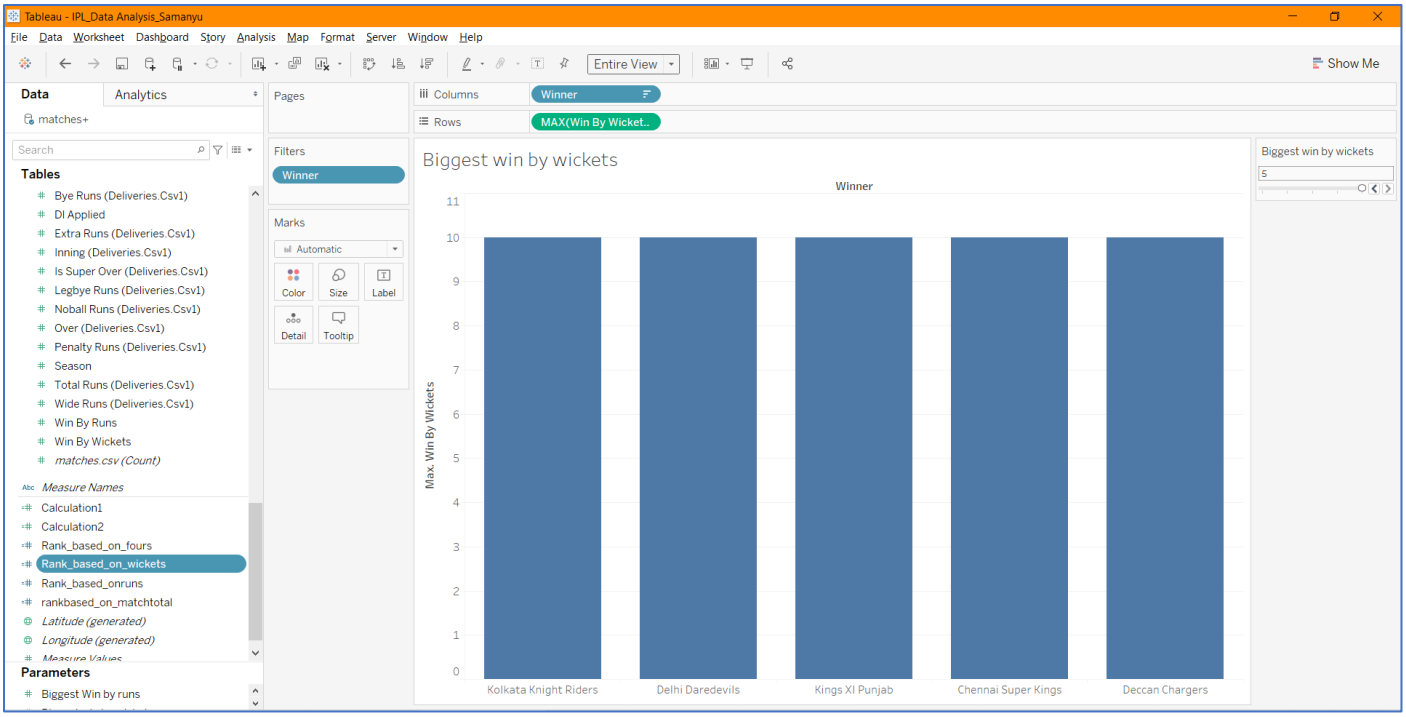


3. Biggest Win by Wickets

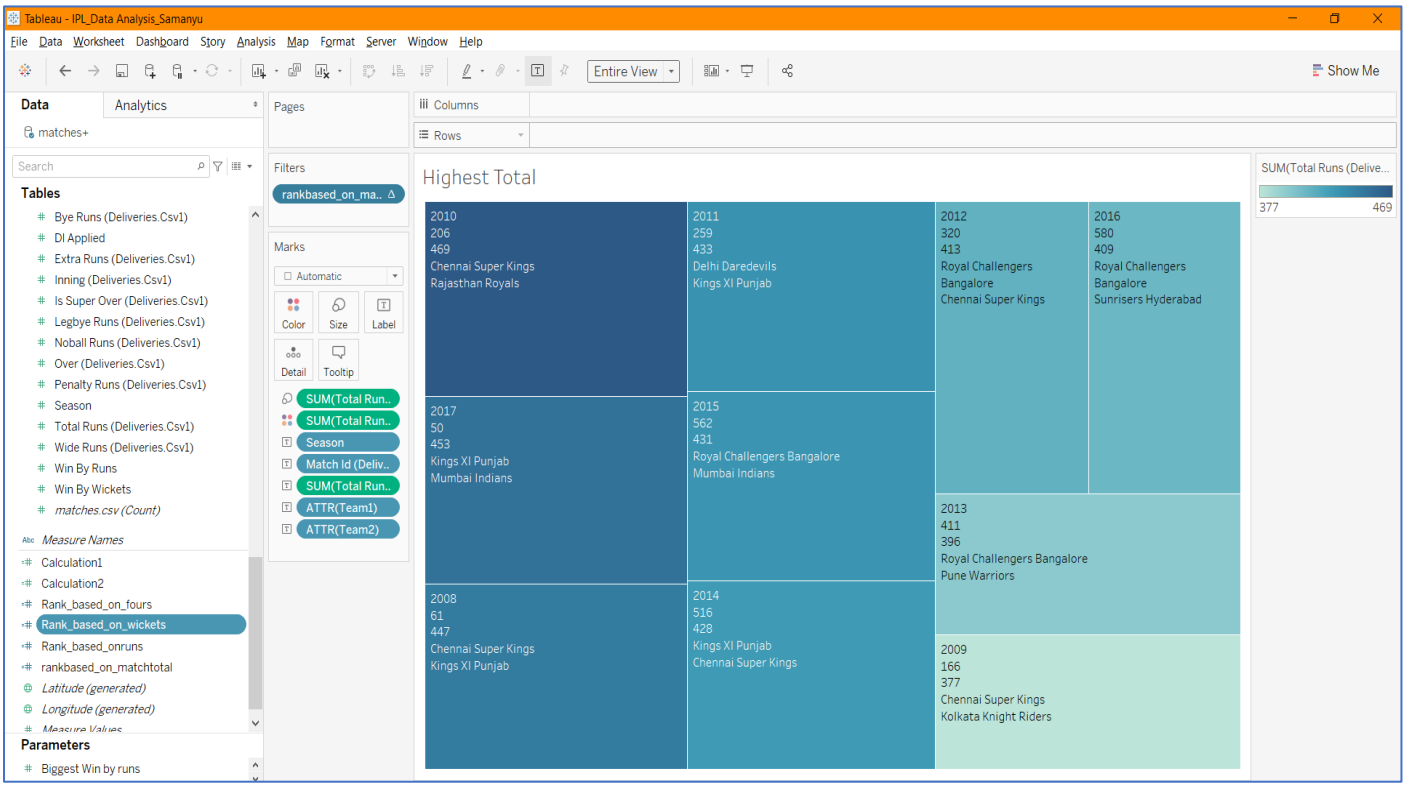
First I have created filter by dropping 'Winner' to the column, max of 'Win by Wickets' in rows, drop 'Win by Wickets' in detail with measure as 'sum' and then created a parameter 'Biggest win by wickets' with data type as 'Integer', current value as 5 and min ,max range as 1 and 5



As I have limited scope to show all the winners by runs, so I have limit this to top 5 biggest win by runs and that can be toggle as per the user choice. The top 5 teams with biggest wins by runs are **Kolkata knight, Delhi Daredevils, Kings XI Punjab, Chennai Super Kings and Deccan Chargers** as given below



4. Highest totals (across all the seasons)



For the above analysis I have dragged and dropped ‘season’ in rows and sum of ‘Total Runs’ in columns. Then I have selected the chart type to tree map. Again we dragged the ‘Match Id’ on Tooltip and created a calculated field named ‘rankbased_on_matchtotal’ and wrote the formula as follows:

RANK(sum([Total Runs (Deliveries.Csv1)]),'desc')

After that I have dragged the new field ‘rankbased_on_matchtotal’ to the filters. Right click on the same and select Edit Table calculation. Selected Specific Dimension and choose match Id. Dragged ‘Team1’ and ‘Team2’ into the tooltip. I also dragged sum of ‘Total Runs’ on colors.

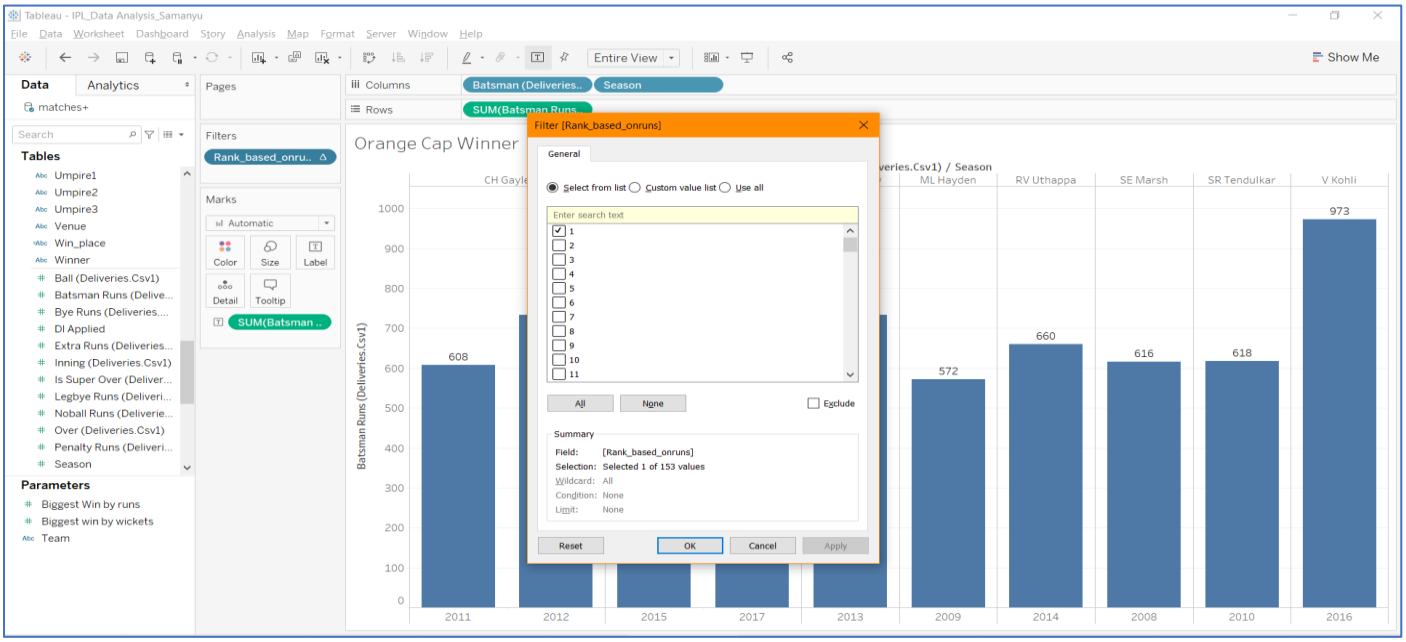
It is evident from the above tree map that it represents the highest totals made by the teams in a match across all seasons. I can see that in most of the cases Royal Challengers Bengaluru and Chennai Super Kings are involved in achieving such feat. Again in the year **2010** Highest total has been created and the match was played between **Chennai Super Kings and Rajasthan Royals with a highest total score of 469**

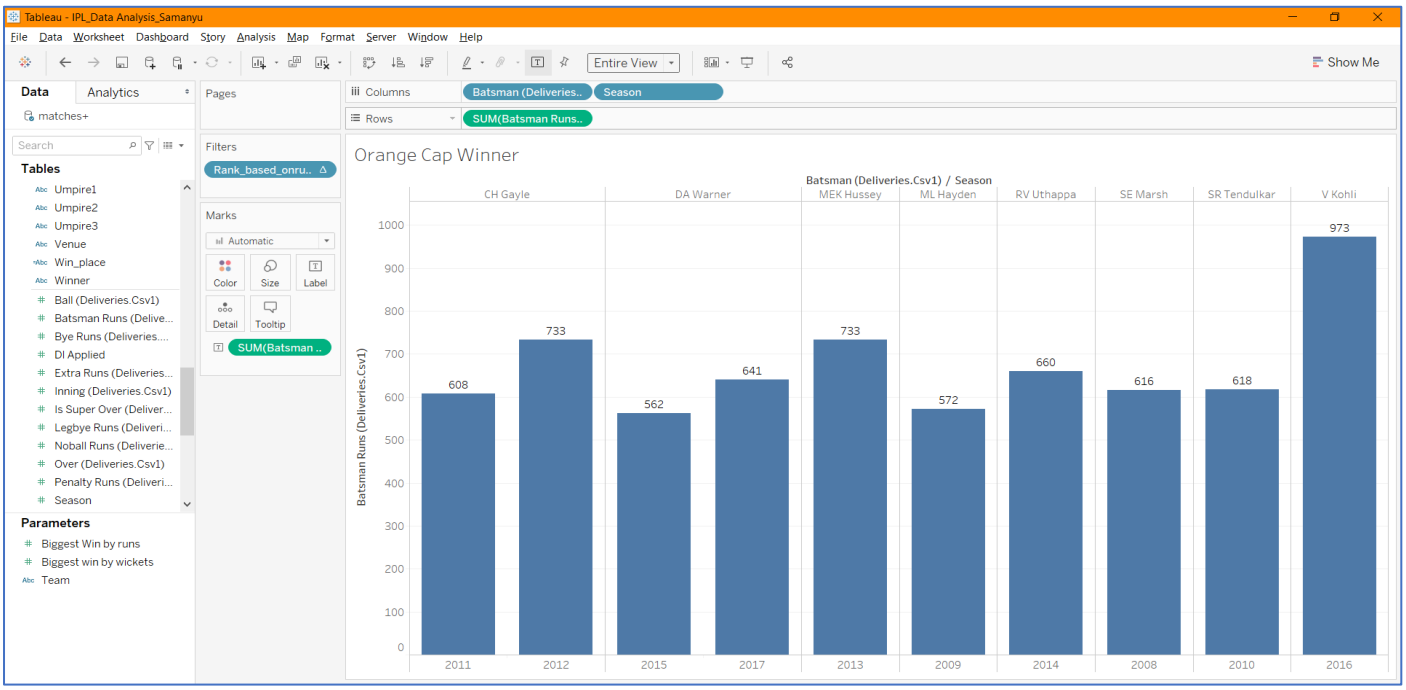
5. Orange Cap contenders (The batsmen who have scored the maximum number of runs in a particular season)

To find this I have dragged the Season on the columns and Batsman Runs on the rows .After that dragged Batsman in columns. Then we dragged Batsman on the tooltip and created a new calculated field named Rank_based_onruns. Here we wrote a formula :-

RANK(SUM([Batsman Runs (Deliveries.Csv1)])),'desc')

Then I dragged the calculated field 'Rank_based_onruns' to the filter , click on the filter field, select 'Edit Filter', select check box which is denoting 1 as I have to take out the Batsman with highest score.





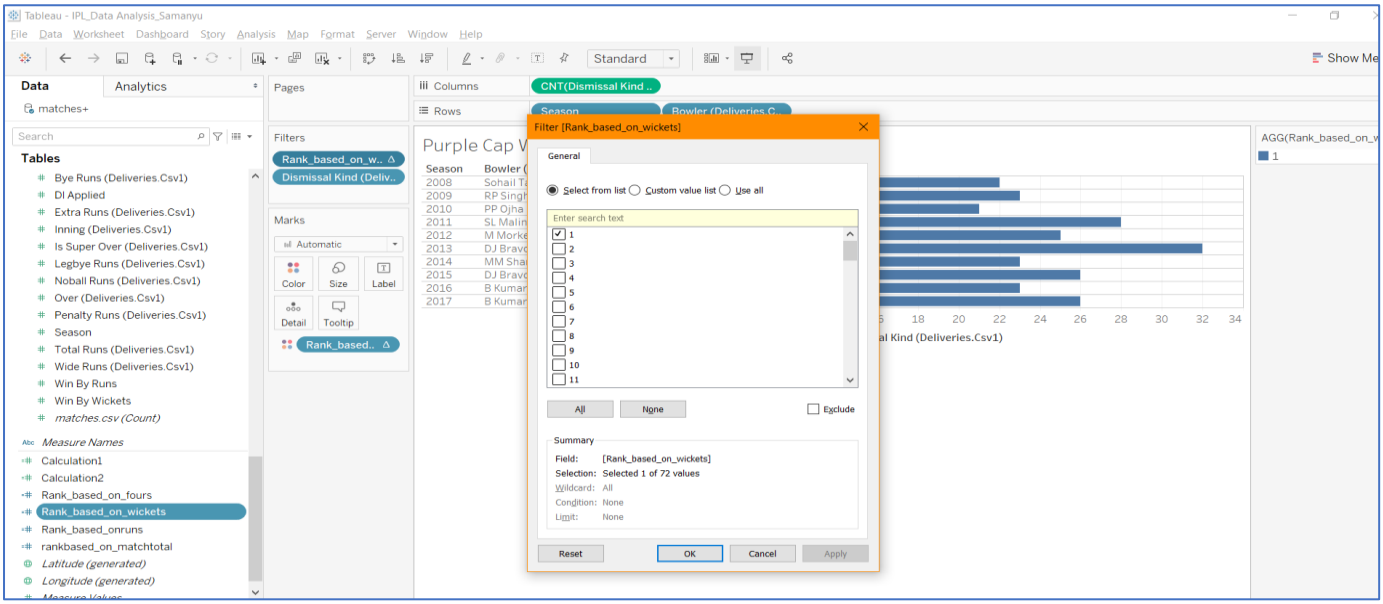
From this Bar chart we may infer that Most of the time overseas players have won Orange cap. Out of these players **Virat Kohli** has scored the highest runs in a single year i.e; in the year **2016** with highest runs of **973**

6. Purple Cap contenders (The bowlers who have taken the maximum number of wickets in a particular season)

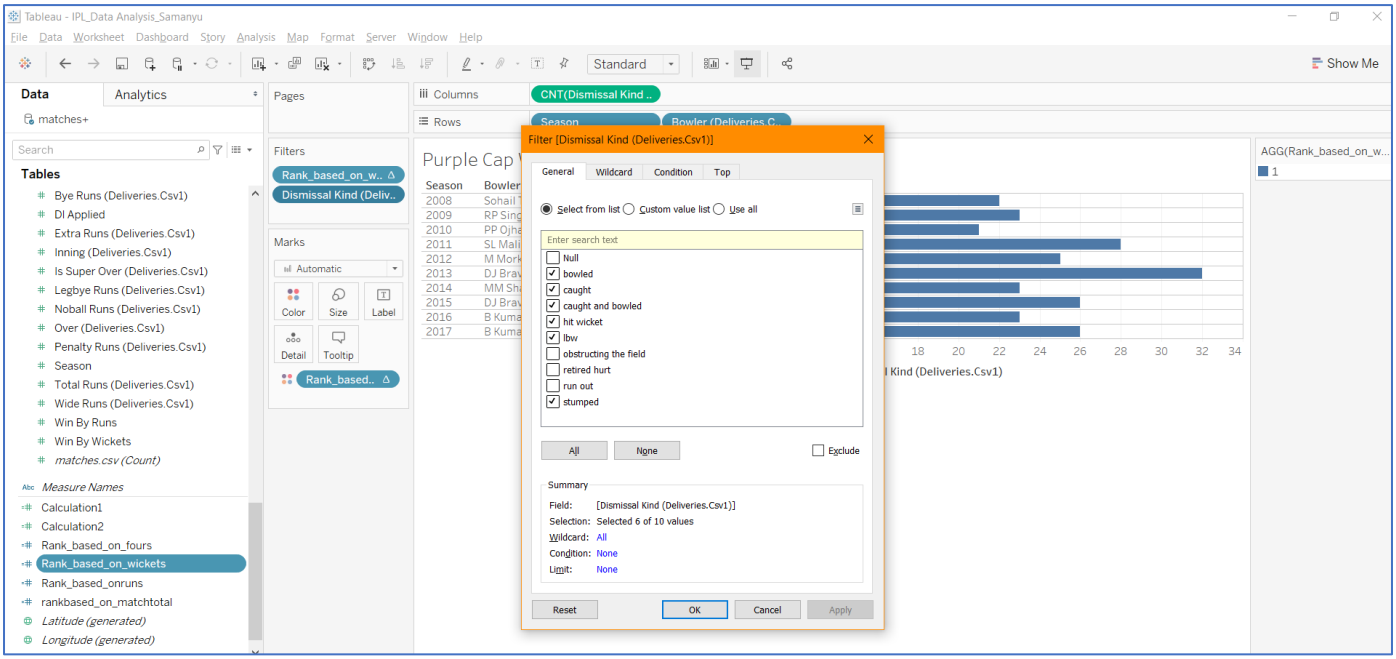
To find this we have dragged the 'Season', 'Bowler' on the rows and Count of 'Dismissal kind' on the columns. Then we have created calculated field 'Rank_based_on_wicket' where we have used the following formula to rank the bowlers.

RANK(count([Dismissal Kind (Deliveries.Csv1)]),'desc')

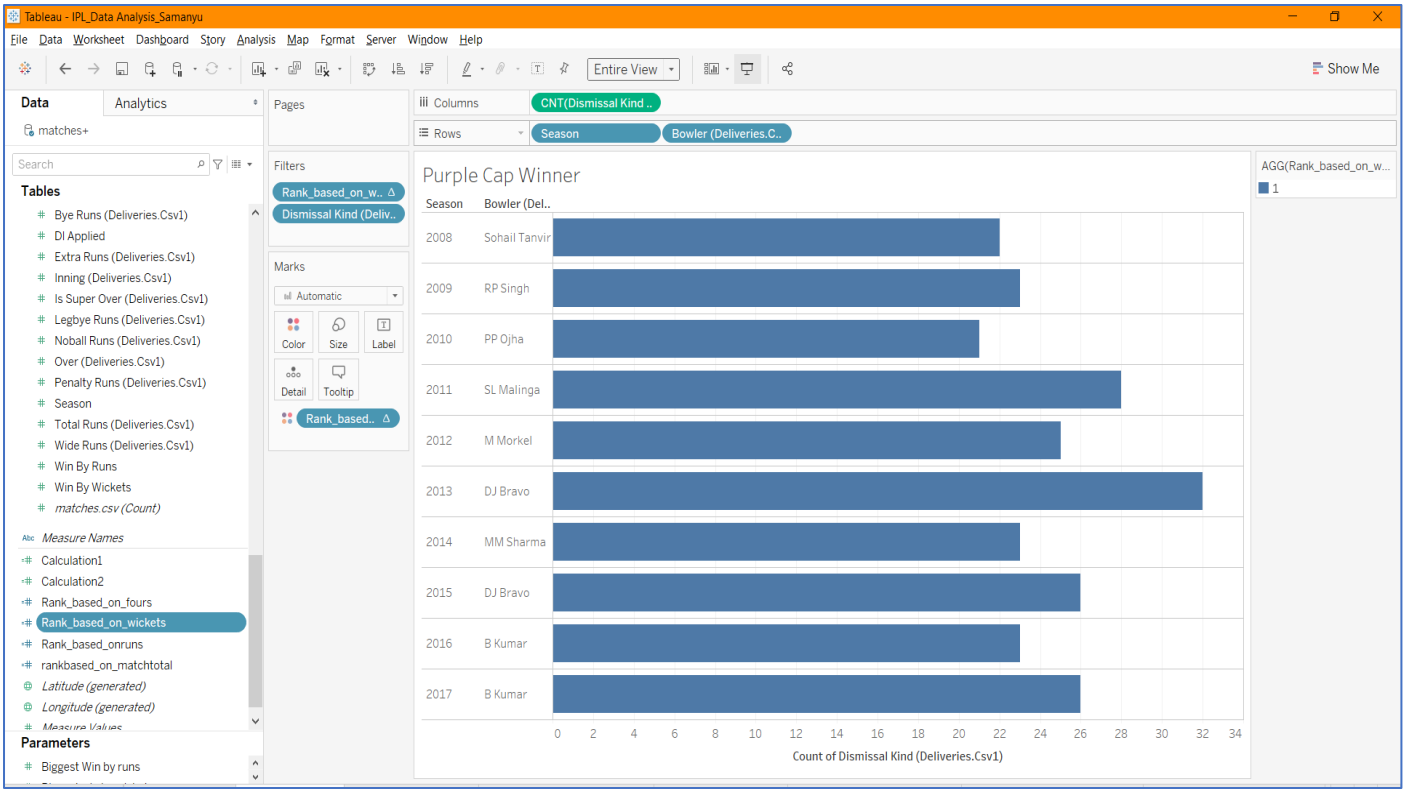
Now I have dragged the calculated field to filter and tick marked 1 only as we want the highest wicket taker in each season that means the Rank should be 1 only



Also I dragged 'Dismissal Kind' to filter and filter only bowled, caught, caught and bowled, hit wicket, lbw, stumped as all these are considered as wicket taken by the bowler



And at last, I dragged ‘Rank_based_on_wicket’ onto color to get the final chart



So from the above chart, it is evident that **D J Bravo** is the highest wicket taker with **32 wickets** among all the season and he took it in the year **2013**

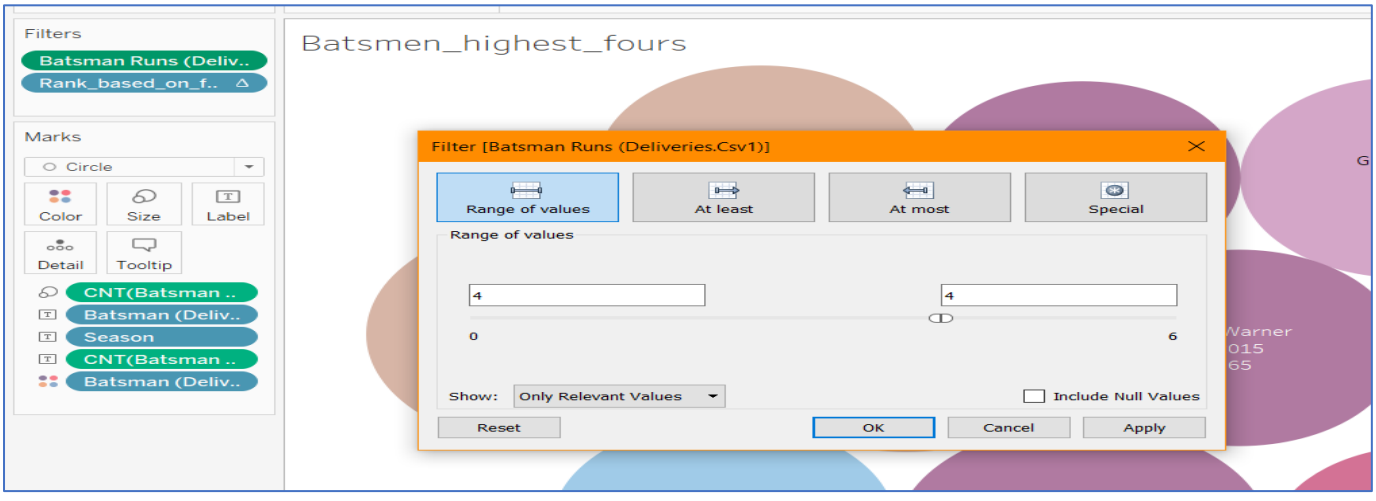
7. Batsmen who have hit the most number of fours and sixes (per season and overall)

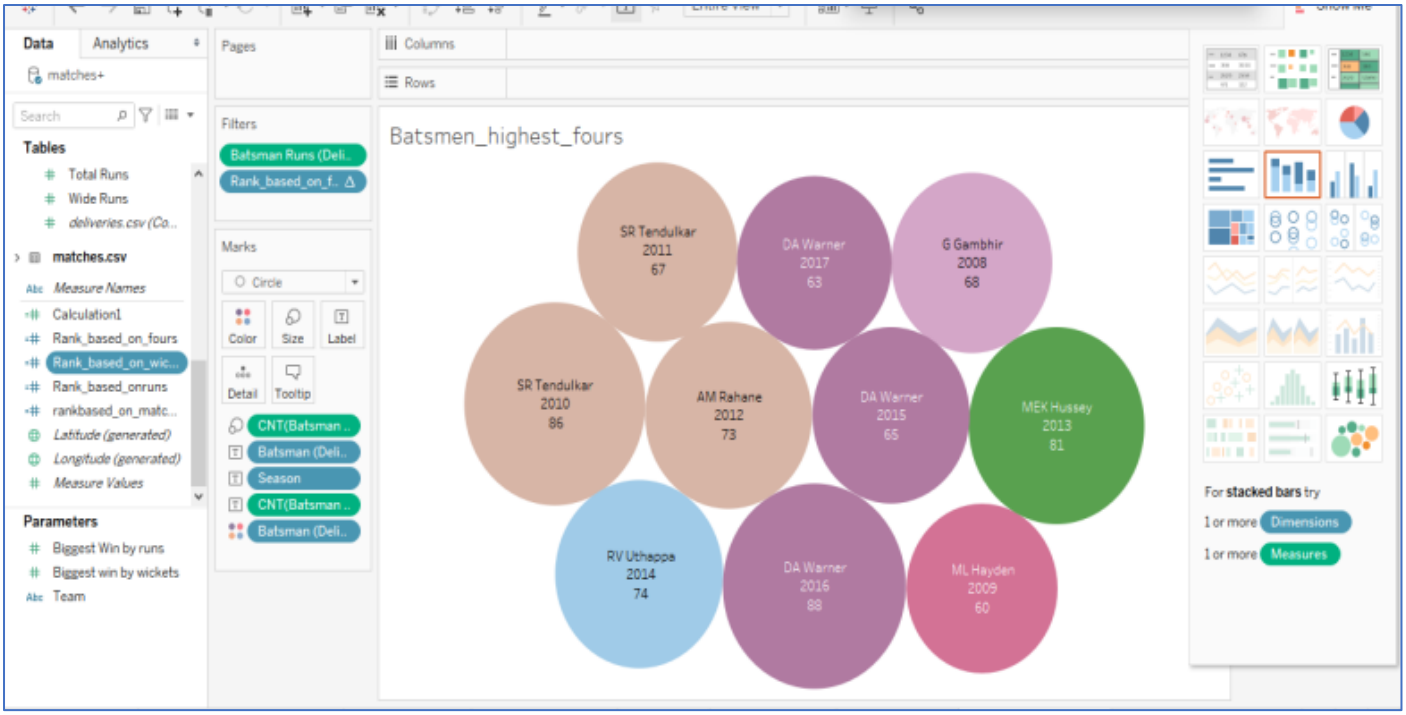
i) Batsmen with most number of Fours :-

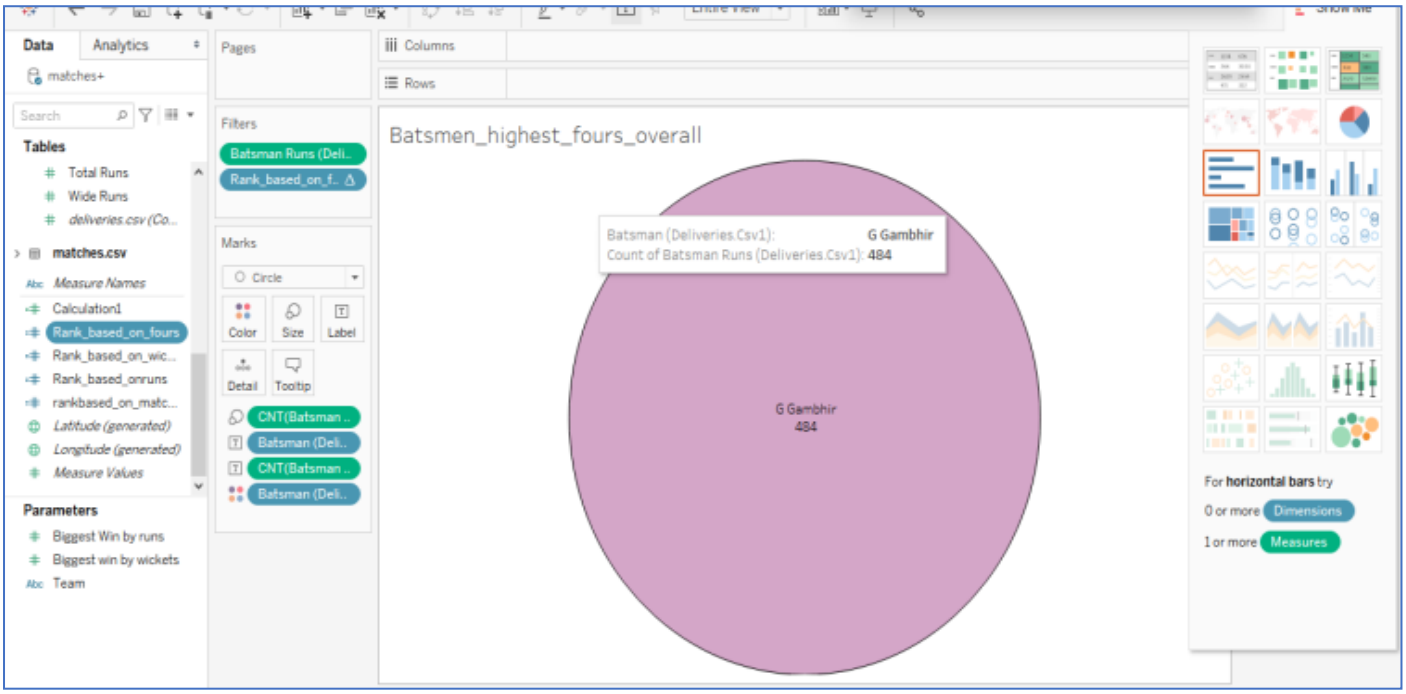
To find highest number of fours, I have dragged the ‘season’ on the rows and Count of ‘Batsman Runs’ on the columns .Then I dragged ‘Batsman’ on the tooltip and created a new calculated field named ‘Rank_based on_fours’ .Here I wrote a formula :-

RANK(COUNT([Batsman Runs (Deliveries.Csv1)])),'desc')

Then right click on the same and select edit Table Calculation. I have opted for specific dimensions and selected ‘Batsman’. I again dragged ‘Batsman’ on colors and count of ‘Batsman Runs’ on Size. I have dragged ‘Batsman Runs’ in the filters column and click on Edit values and gave Range of values as 4 to 4 and dragged ‘season’ on label







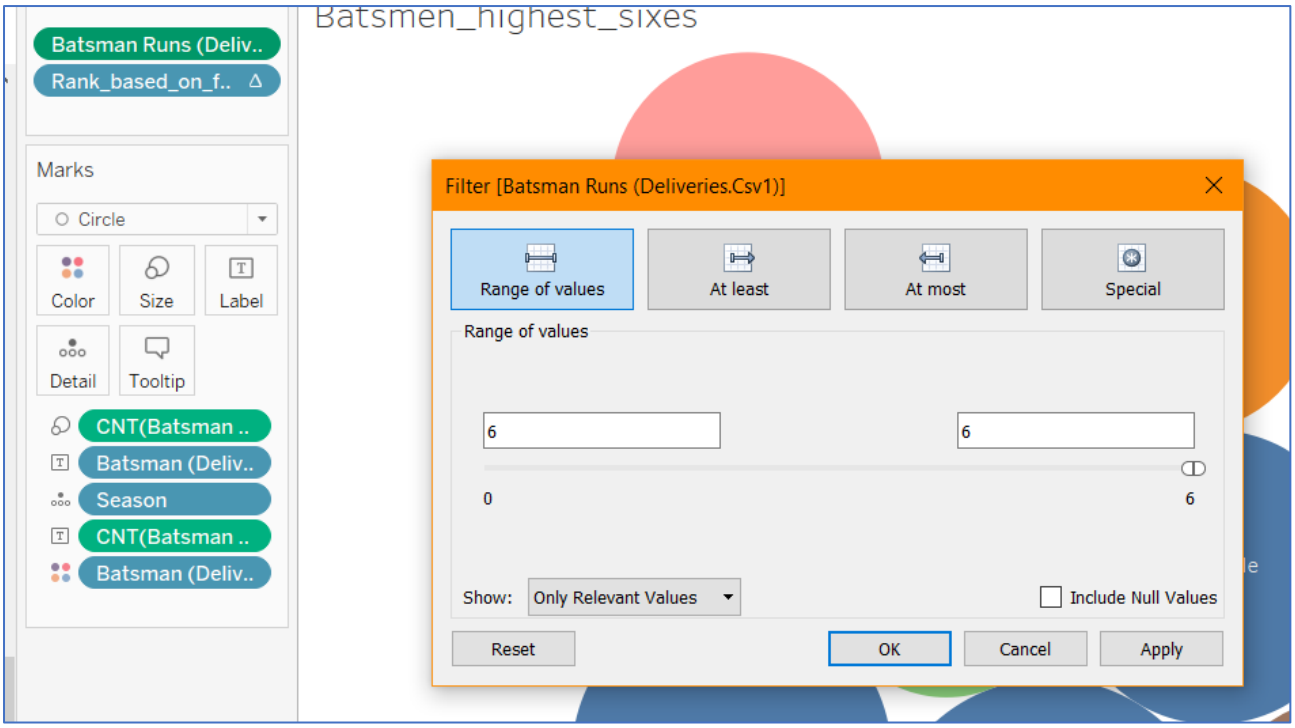
From this Bubble chart it may be inferred that Out of these players DA Warner has hit the highest no of fours in a single year i.e; in the year 2016.SR Tendulkar is on the second place with 86 fours in the year 2010.Moreover in the second chart given above it is clear that **Gautam Gambhir** has hit the highest no of fours i.e; **484** across all seasons.

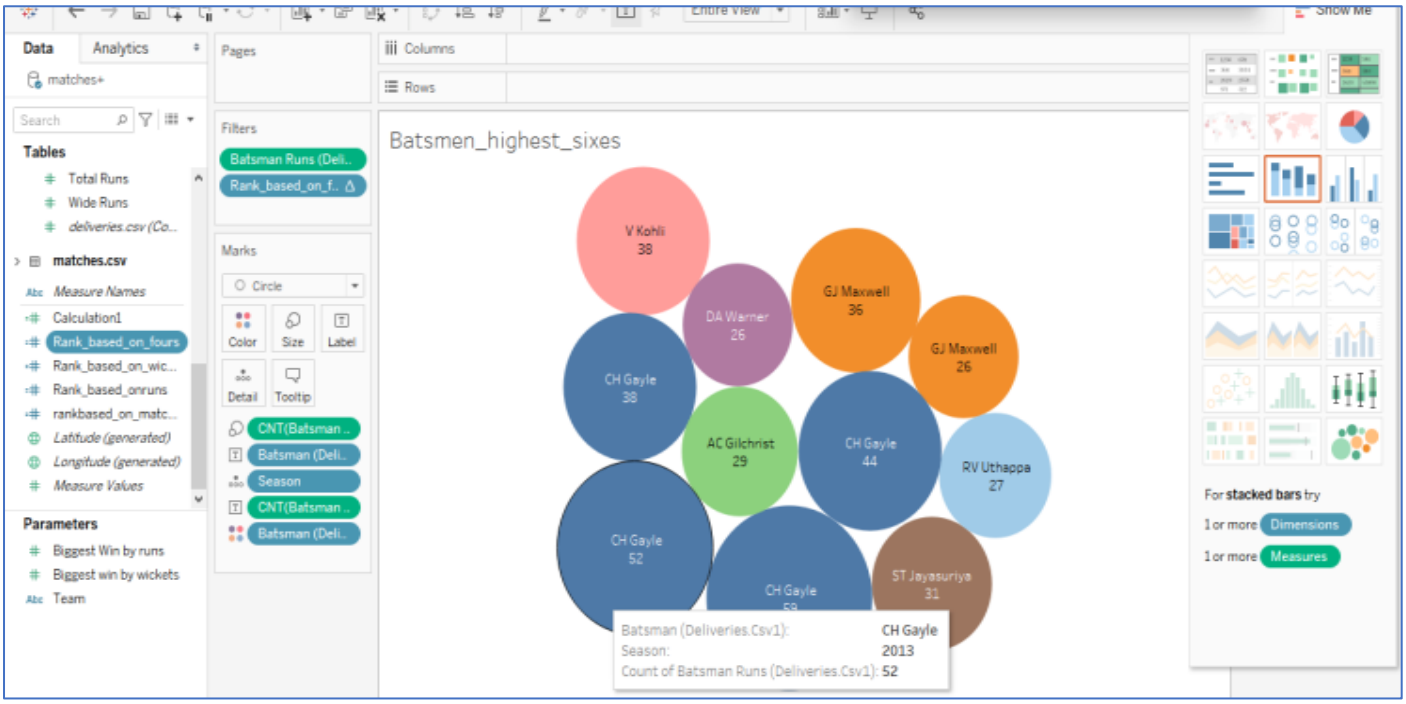
ii) Batsmen with most number of Sixes :-

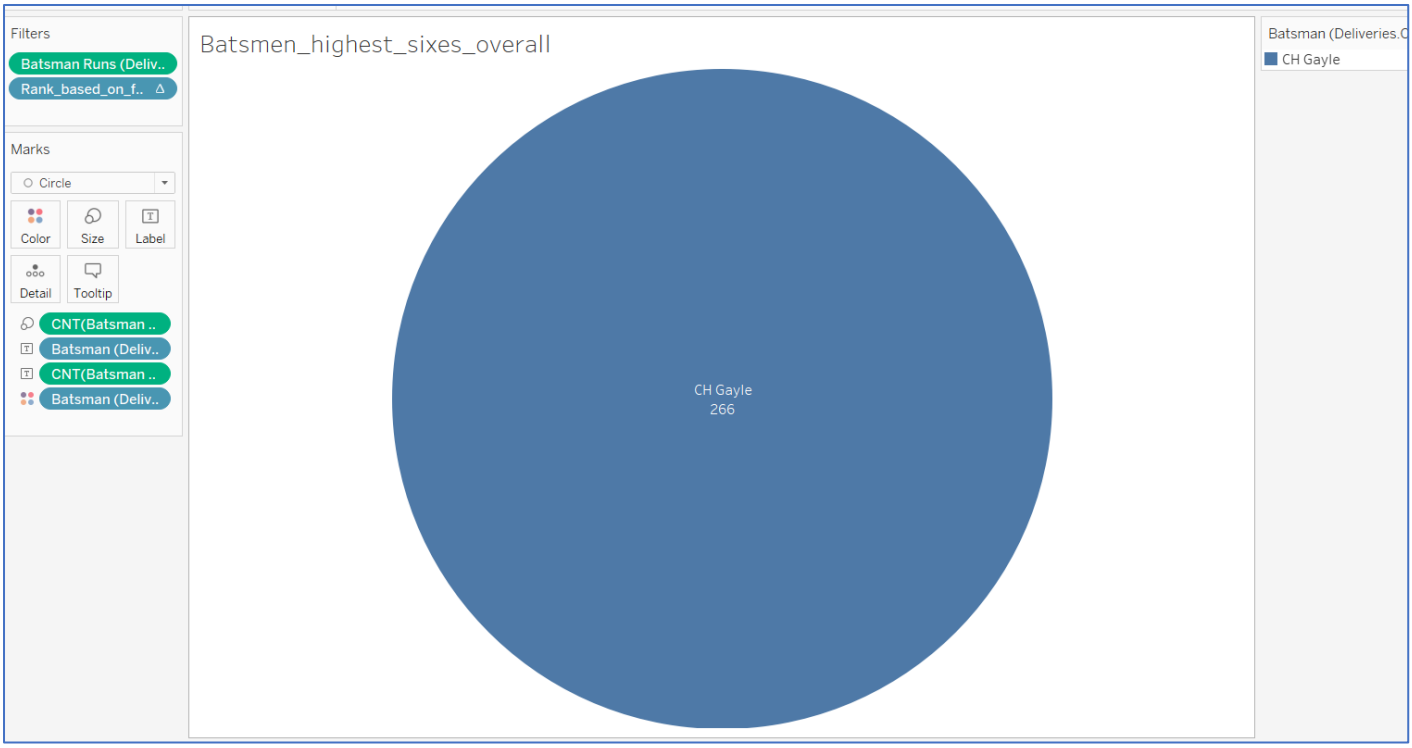
To find highest number of sixes, I have dragged the Count of 'Batsman Runs' on the rows and 'Batsman' on the columns .Then I have created a new calculated field named 'Rank_based on_Sixes' .Here I wrote a formula :-

RANK(COUNT([Batsman Runs (Deliveries.Csv1)])),'desc')

Then right click on the same and select edit Table Calculation.I opted for specific dimensions and selected Batsman. Again dragged 'Batsman' on colors and count of 'Batsman Runs' on Size.I have dragged 'Batsman Runs' too in the filters column and after right click on the same we opted Edit Filter and selected the range of values from 6 to 6.Then I dragged season on Details.







From this Bubble chart it may be inferred that Out of these players **Chris Gayle** has hit the highest no of sixes in a single year i.e; in the year 2012.We can see that most of the players are overseas players. Moreover in the second chart given above it is clear that **Chris Gayle** has hit the highest no of sixes i.; 266 across all seasons.

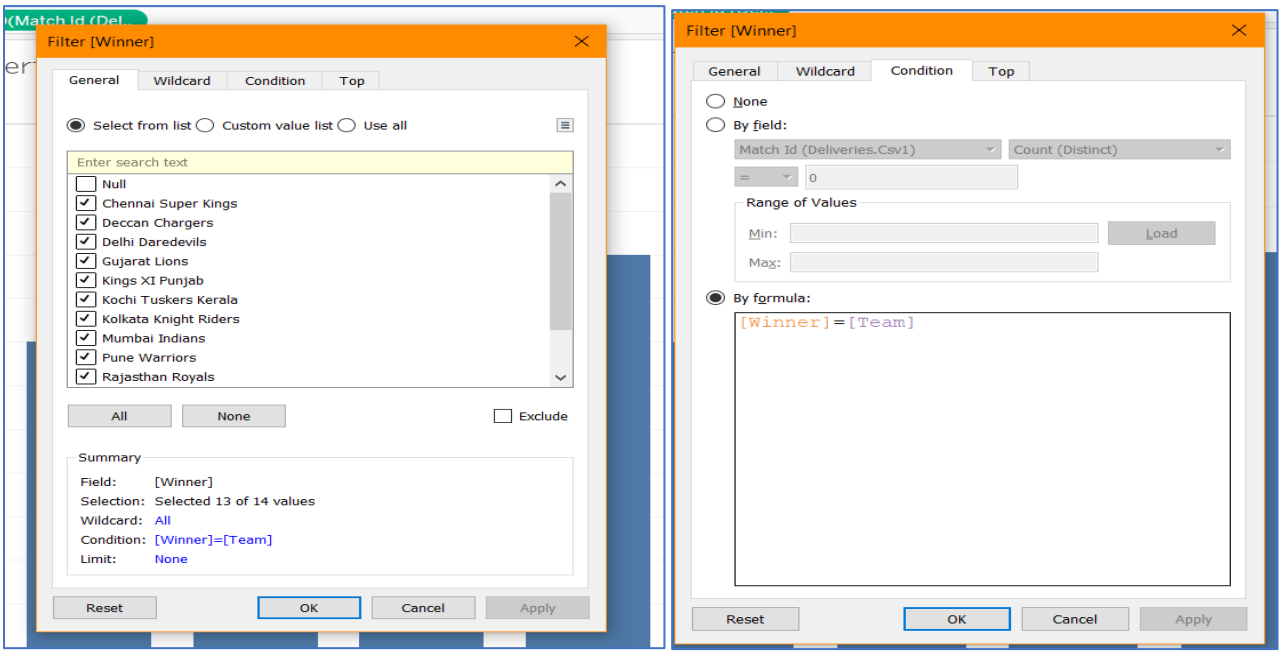
8. Season-wise team performance (wins vs losses)

i) Win

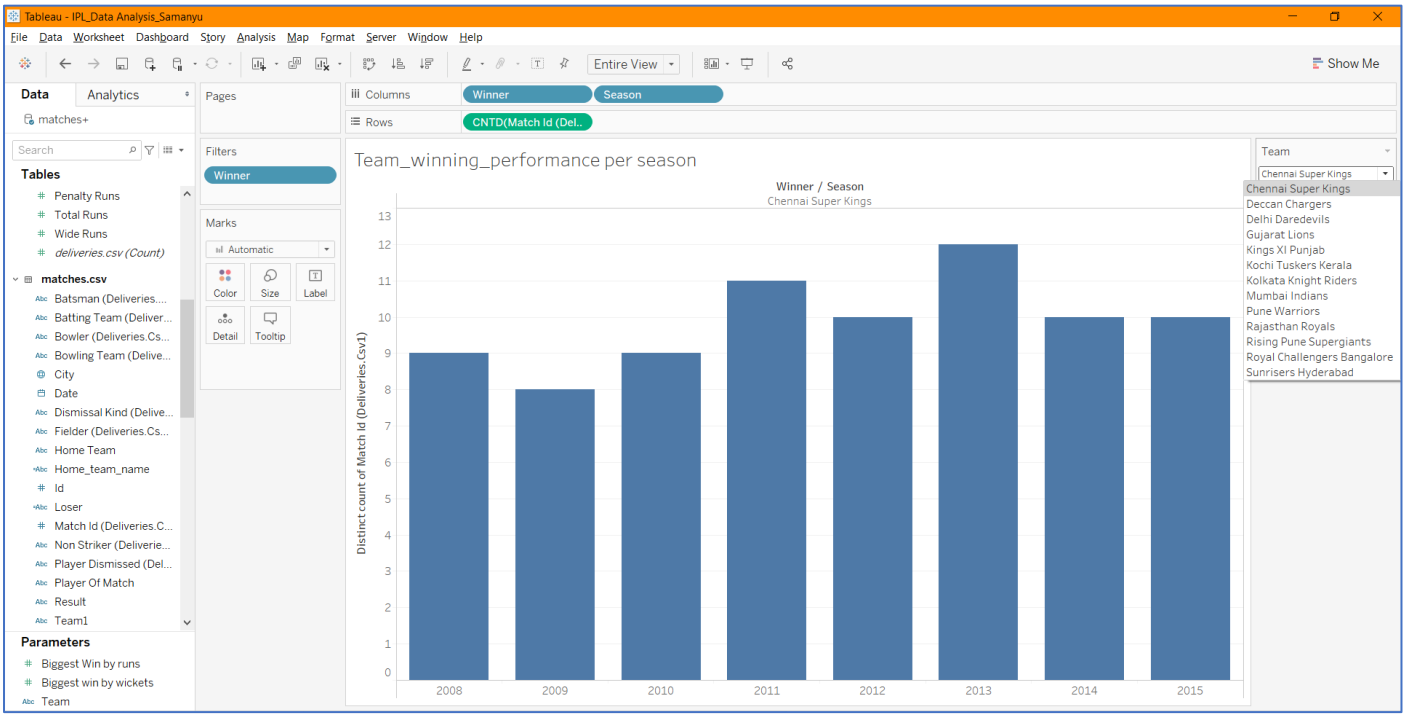
Here I have dragged 'winner' and 'season' on the columns and count of 'Match Id' on the Rows. Then we dragged 'Winner' on the Filters and right click on the same. I opted of Edit Filters in general tab I unticked null and on condition tab we selected by formula and wrote the below mentioned formula: -

[Winner]=[Team]

Now I have kept 'Team' as a show filter status and as per the choice of user, one can draw the winning performance for each team



`

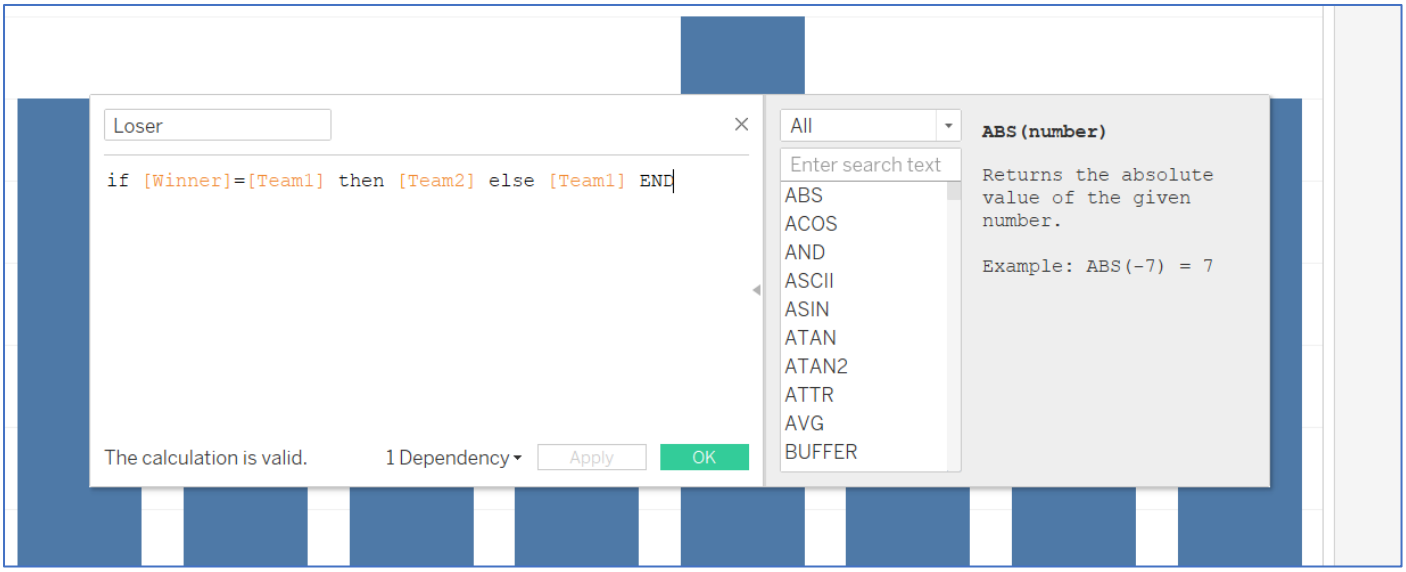


So if I select 'Chennai Super kings', I can see that it has won maximum number of matches in the year 2013 with 12 wins

ii) Loss

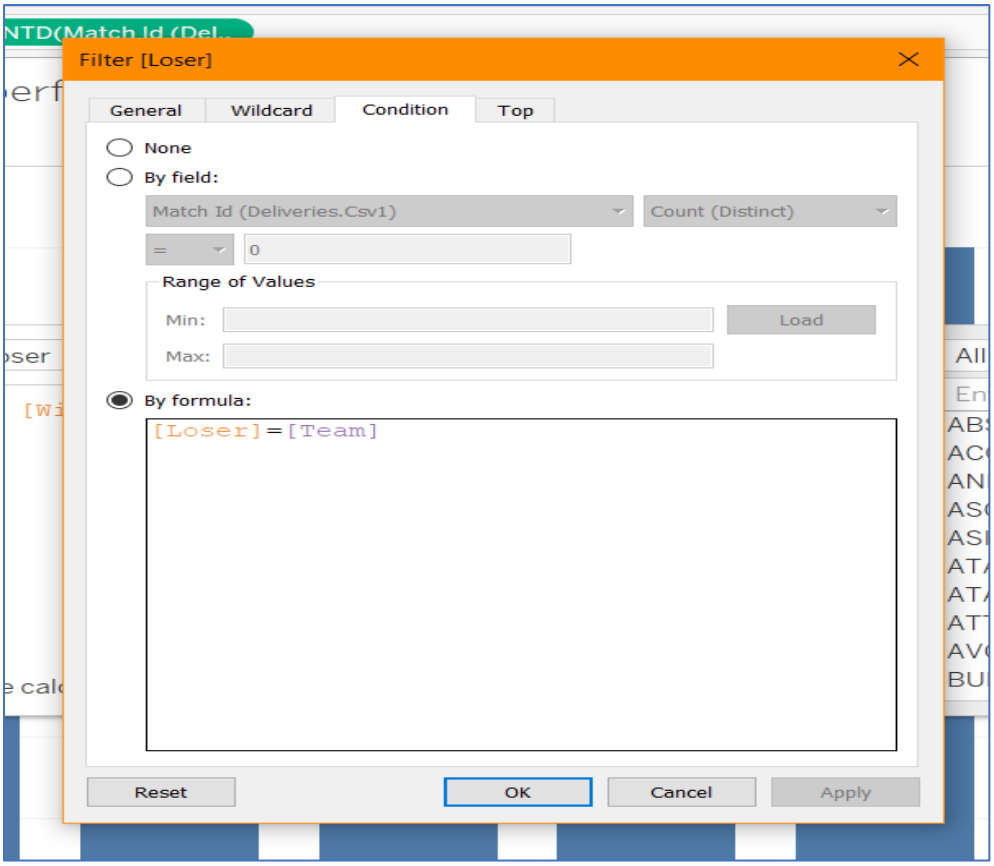
For this, I have creates a calculated field by the name 'Loser' and used the following formula to create loser

If [Winner]=[Team1] then [Team2] else [Team1] END

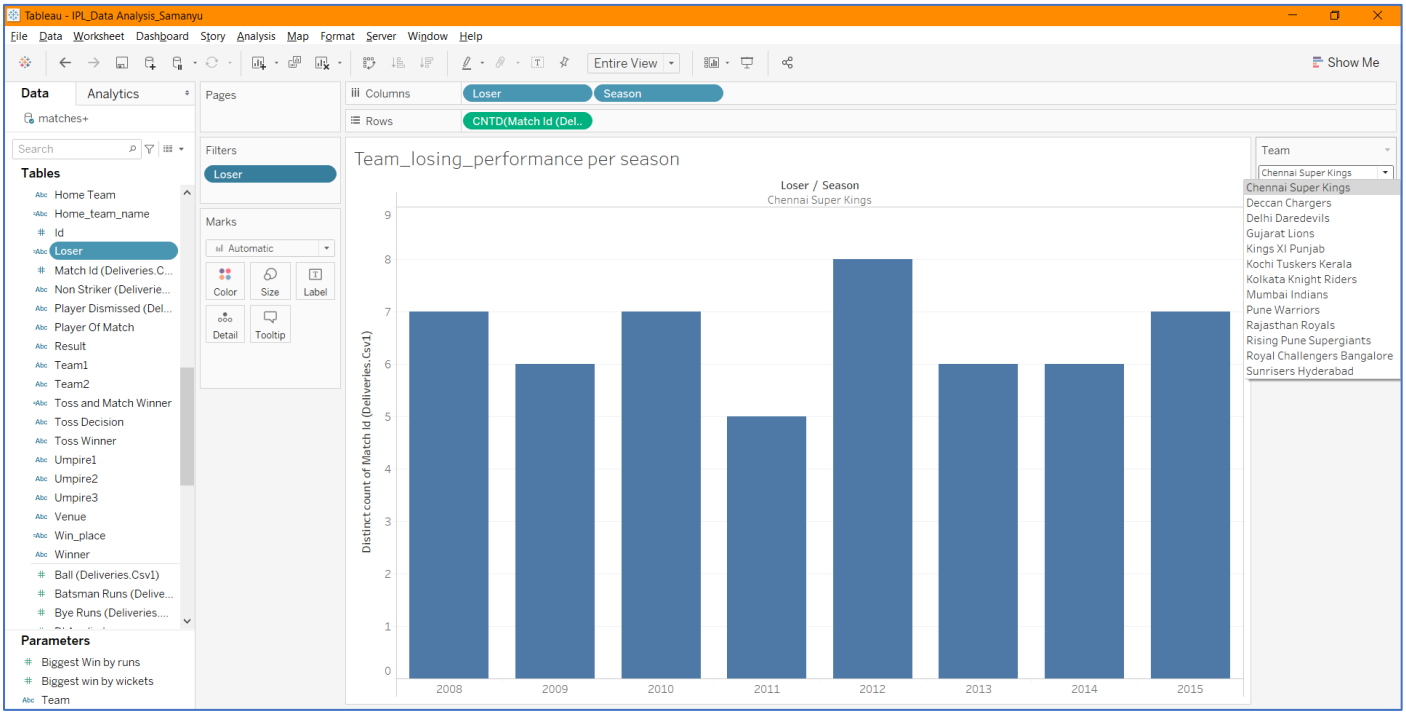


I have dragged 'Loser' and 'season' on the columns and count of 'Match Id' on the Rows. Then we dragged 'Loser' on the Filters and right click on the same. I opted of Edit Filters in general tab and on condition tab, I selected By formula and wrote the below mentioned formula: -

[Loser]=[Team]



Now I have kept 'Team' as a show filter status and as per the choice of user, one can draw the losing performance for each team

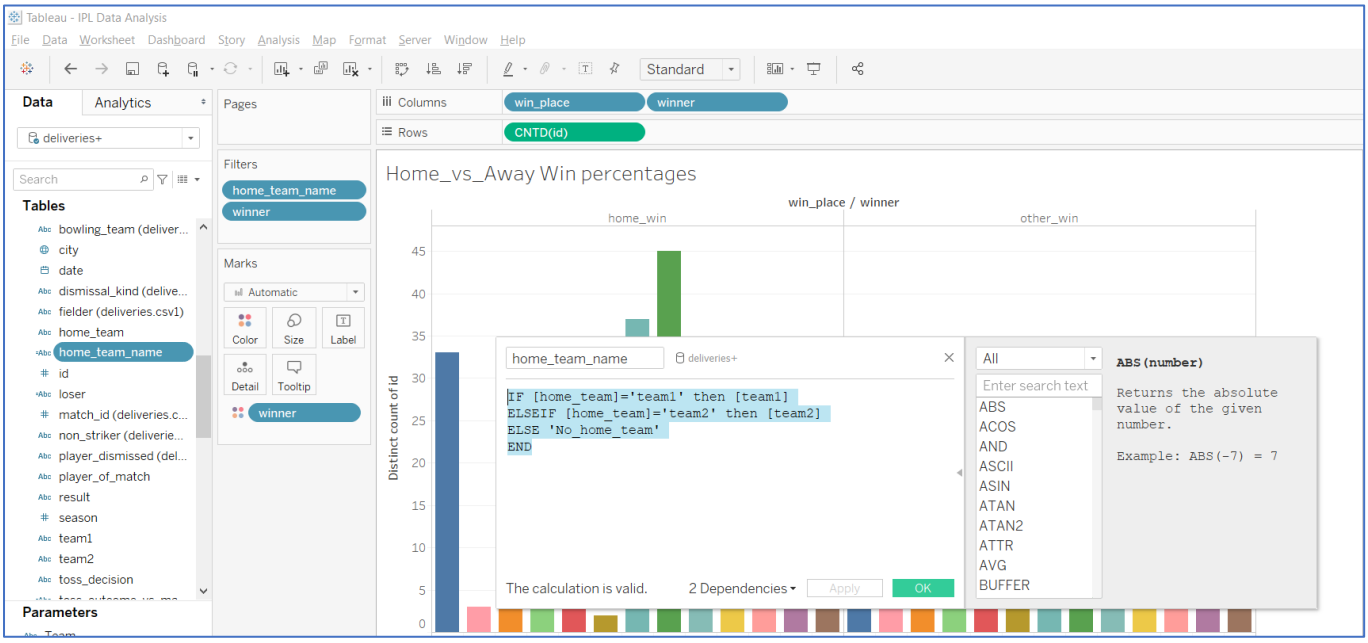


So if I select 'Chennai Super kings', I can see that it has lost maximum number of matches in the year 2012 with 8 losses

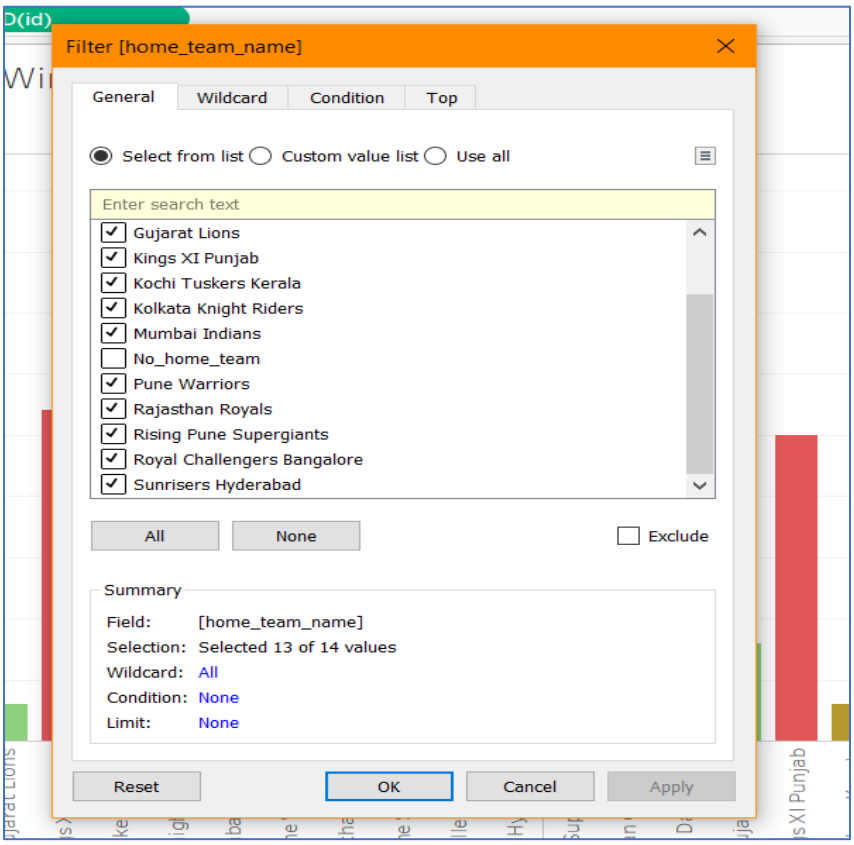
9. Win %age (Home vs Away)

I have created a calculated field by the name ‘home_team_name’ where I have used if else formula as given below

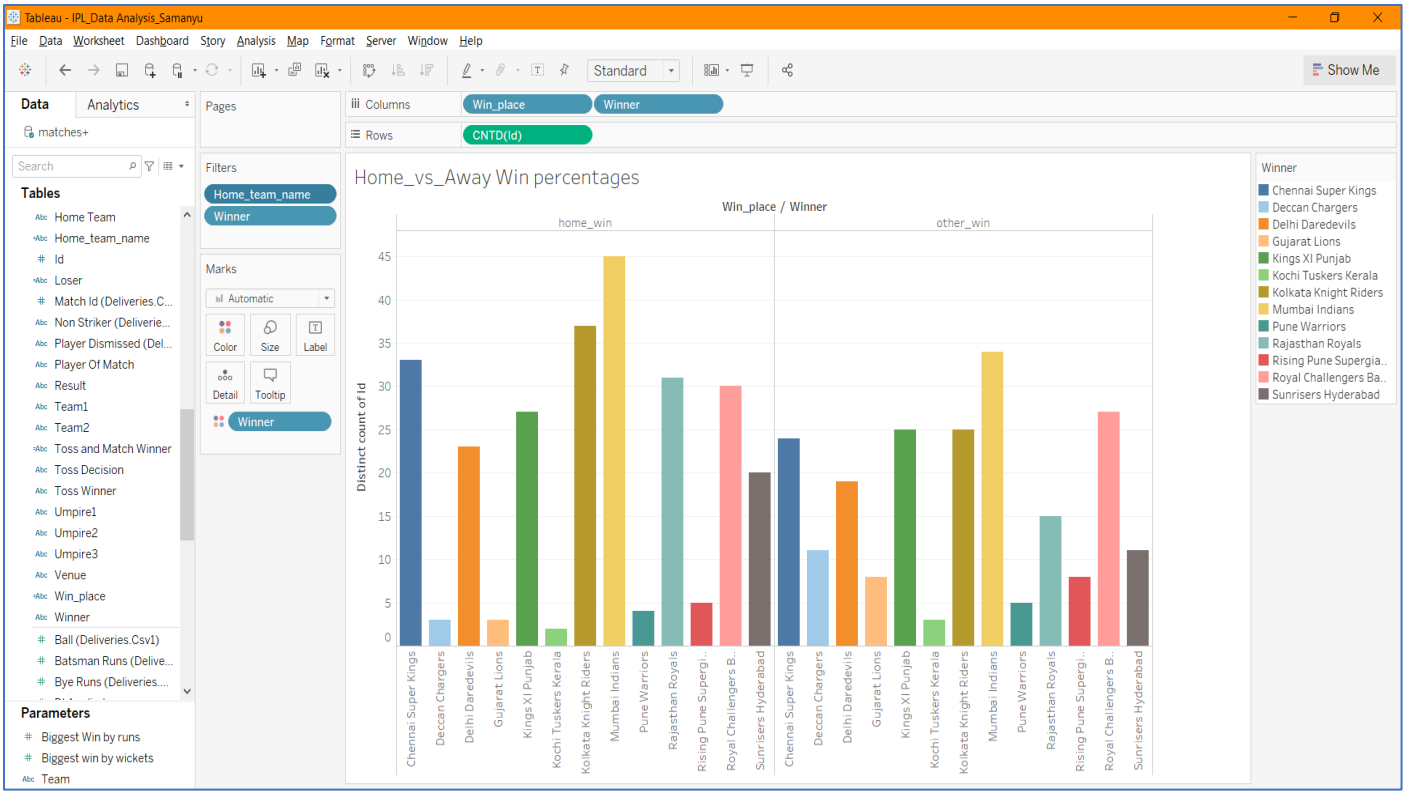
IF [home_team]='team1' then [team1] ELSEIF [home_team]='team2' then [team2] ELSE 'No_home_team' END



After that I dragged the calculated field ‘home_team_name’ into filter and unticked ‘No_home_team’



This is the final chart I have got as given below



From the above chart, it is evident that **Mumbai Indians** have got highest **Home_win** followed by 'Kolkata Knight Riders' whereas in case of Other-win which is away win, here also **Mumbai Indians** has the highest number of **away win** followed by 'Royal Challenger Bangalore'