**IPL Dataset**

Problem statement to build an interactive Tableau dashboard for a newsletter on IPL, and it includes infographics and derived statistics from the dashboard. This is groundwork for understanding the two pivotal datasets that form the backbone of our analysis:

**matches.csv:** This dataset provides detailed match-level data across IPL seasons from 2008 to 2017, offering a historical view of the games.

**deliveries.csv:** This dataset contains ball-by-ball information on each match from 2008 to 2017.

We have to perform tasks, which are categorised into three main areas of focus:

* Match stats

Toss vs match outcome

Biggest win

Highest total

* Player stats

Orange cap contender

Purple cap contender

Batsmen with the highest number of fours and sixes

* Team stats

Season-wise team performance

Winning percentage

First we will load our critical ‘matches.csv’ and ‘deliveries.csv’ datasets into Tableau. It is here that we will see how the raw data starts shaping itself into a powerful visual tool. We will navigate the Tableau interface, understand where and how to import files, and learn the significance of joining these data sets to create a comprehensive analytical platform.

Here’s a structured breakdown of what we learnt:

**Loading data sets in Tableau:** We started by initiating a new workbook in Tableau. To load the data sets, we navigated to the application menu and selected ‘Text File’ since ‘matches.csv’ and ‘deliveries.csv’ are in CSV format. We then walked through how to locate and open the ‘matches.csv’ file, setting the stage for our data analysis.

**Joining data sets for a unified view:** Next, we focused on merging our datasets to enrich our analysis. By dragging ‘deliveries.csv’ into the dialogue box already containing ‘matches.csv’, we initiated a join. The colour indicators play a crucial role here; orange signifies a successful join, while red indicates an issue that needs addressing. This step is crucial as it combines match-level data with detailed delivery information, providing a more comprehensive data set for analysis.

**Building informative dashboards:** With our data loaded and joined, we turned our attention to creating dashboards. We differentiated between dimensions (indicated by ‘Abc’), metrics (indicated by ‘#’) and dates (marked by the calendar icon). The process of creating three distinct dashboards focused on match, player and team statistics, aligning with the project’s scope.

We’ll concentrate on crafting detailed visualisations for match, player and team statistics. Get ready to see how numbers and information come to life, offering insights into the vibrant world of IPL. Let’s dive in!

We focused on two critical aspects of the IPL match statistics dashboard in Tableau: ‘Toss vs Match Outcome’ and ‘Biggest Wins by Runs’. Here are the key takeaways:

**Toss vs Match Outcome**

**Dimension creation:** We explored how to create a new dimension, ‘Toss vs Outcome’, to analyse the influence of winning the toss on the match result. By using a calculated field, we compared the ‘Toss Winner’ with the ‘Match Winner’ to categorise each match as either ‘Toss Winner Won’ or ‘Toss Winner Lost’.

**Visualising the impact:** We then visualised this relationship by creating a chart that compares toss outcomes to match outcomes for each venue. This visualisation helps in understanding the strategic advantage of winning the toss in different conditions and venues.

**Biggest Wins by Runs (Sum of Win by Runs)**

**Highlighting dominant victories:** We delved into identifying the most substantial victories in IPL history, focusing on the biggest wins by runs. We learnt how to filter and sort to focus on the top 10 most significant victories by the run margin.

**Building an interactive chart**: Techniques were shared for enhancing interactivity, such as colour-coding by the winning team and adding detailed tooltips. We also discussed how to aggregate the data effectively to focus on the most impactful wins, ensuring that the visualisation is both informative and engaging.

Now, let’s delve deep into the nuances of cricket analytics. We will now shift our focus to understanding the distribution of wins by wickets, a key indicator of match dominance. You will learn how to create a distribution plot that reveals patterns and trends of how teams clinch their victories. But that’s not all; we’ll also examine the highest totals across all seasons, identifying standout innings that have left a long-lasting mark on the IPL history.

let’s consolidate our understanding of a few significant aspects of IPL match analysis:

**Wins by wickets:** We navigated through creating a distribution plot in Tableau, strategically placing the ‘Wins by Wickets’ dimension and ‘match ID’ to visualise the frequency of victories based on wickets taken. This technique gave us a clear view of how often teams win with specific wicket margins, providing a deeper insight into match outcomes. We refined our analysis by focusing on non-zero wicket outcomes, which indicate a batting team win, ensuring that our visualisation reflects only the instances where wickets directly influenced the win. In our ‘wins-by-wickets’ plot, we applied thoughtful filtering and exclusion, particularly removing instances of zero-wicket wins to concentrate on more telling data. We also implemented a date filter to span all IPL seasons, offering a comprehensive look at trends over time.

**Highest totals:** Moving on, we constructed a visualisation to display the total runs scored in IPL matches. By aggregating the total runs and pairing them with match IDs and batting teams, and then filtering the scores by innings, we created a comprehensive view that highlights the scoring capability of each team in individual matches. This exploration not only highlighted standout innings but also showed us the trends and exceptions in scoring across IPL seasons. Moreover, you observed that the ‘Royal Challengers Bangalore’ notably secured the top three spots, indicating their dominant batting performance.

**Dashboard construction:** We then brought all four insights together in a cleanly presented 2x2 tiled dashboard. This layout choice, favouring simplicity and clarity, allowed us to integrate the necessary filters and legends without overwhelming the viewer. The approach ensured that even those new to Tableau could follow along and appreciate the strategic placement of each element.

Having completed our exploration of match statistics, let’s now delve into the realm of player statistics.

You have now developed a comprehensive understanding of player statistics in IPL, brought to life through vibrant ‘tree plots’ and ‘bubble charts’. You saw how the ‘Orange Cap’ contenders light up the field with their scoring, represented through interactive treemaps that reveal their proportional impact. Similarly, the ‘Purple Cap’ contenders’ wicket-taking abilities were vividly illustrated, giving you a clear view of the top bowlers. Moreover, the bubble charts for the most sixes and fours scored provided a dynamic and colourful representation of the players’ boundary-hitting abilities.

By now, you should feel more confident in your ability to use Tableau to create impactful and informative visualisations that tell the stories behind the statistics.

Now, let’s evaluate the season-wise performance of IPL teams from 2008 to 2017.

We embarked on a journey to help you understand how teams have fared across different IPL seasons, focusing on wins, losses and the margins of those outcomes. We created a ‘Teams’ parameter and various calculated fields such as ‘Win Flag’, ‘Outcome’, ‘Against’ and ‘Margin by’ to transform raw data into meaningful metrics. We built a graph displaying aggregated wins and losses over the years, colour-coded for clarity and enriched with tooltips for detailed insights. Further, we introduced the ‘Win%’ metric and the ‘Home vs Away’ dimension to add depth to our analysis, showcasing how teams perform at home games as compared to away games. It is important to mention that the stacked bar chart might not be the optimal visualisation here. Alternatives like highlight tables or horizontal bars, especially when filtering out years with null values, can provide clearer insights into win percentages. The dashboard will demonstrate how these alternatives can enhance the clarity of the data.

**Please be aware that:**

The ‘NA’ notation within the ‘Home\_team‘ column indicates that the match was held at a neutral location.

Predominantly, the entries for 2009 and 2014 are marked as ‘NA’, reflecting that those matches took place on neutral grounds.

For the Chennai Super Kings, the absence of data, primarily during 2016 and 2017, is due to the team’s disqualification.

You can also filter the ‘null’ values in the above visualisations to get a better graph.

Finally, we combined all these elements into a ‘Team Statistics’ dashboard, offering a multifaceted view of team performances across seasons.