

Make your IPL team win!

- MANOJ GHUHAN ARIVAZHAGAN, NIKHIL YADALA

Introduction:

Indian Premier League is a T20 cricket league, contested by 8 teams based out of 8 different Indian cities. In a single T20 game, 2 teams contest against each other. Each team gets to bat an innings of a maximum of 20 overs(120 balls) and a maximum of 10 wickets. The team that wins the toss gets to decide if they want to bat/bowl first. The 1st 6 overs of each inning are called powerplay overs, during which the bowling team can place only 2 players outside the 30-yard circle.

Goals:

In this assignment, we try to analyze and understand different factors that can influence/enhance the winning probability of a team.

More specifically, we try to address the following questions:

1. Does dominating the powerplay overs(1-6) boost the chances of winning?
2. Does retaining wickets and dominating the death overs(16-20) boost the chances of winning?
3. Does the choice of batting/bowling first affect the chances of winning when conditioned on the venue?

We provide the user with an option to choose the IPL team they want to analyze. The users can also choose the timeframe of the matches to be analyzed. This will help the users in studying the change in trend of metrics over time.

Context:

The playing grounds at different venues vary significantly in their characteristics. For example, the ground in Bangalore is small and is friendly for batting whereas the ground in Chennai is big and assists bowling better. Since most of the games are day-night matches, the dew factor can also influence the ease of batting/bowling. Therefore we hypothesize that the type/nature of the venue can influence the winning probabilities. In order to test the hypothesis, we plot the percentages of wins by the team batting first.

Observations and Design Decisions:

We observe that there exists a definite bias in the wins when conditioned on the venue. Further, we also analyzed the percentage of teams that win the toss and decide to bat. We were surprised to find that in some of the venues like Kolkata, these 2 values did not match with each other. The IPL teams should take into account, these venue statistics to make their decision

which might boost their winning probabilities. Since we are comparing 2 percentages/probabilities, we decided to plot a horizontal stacked bar plot to represent the 2 values. These plots are directly plotted on the geographical map for better understanding and visualization of the environmental factors(humidity, dew, temperature, etc.)

Context:

In order to answer questions 1 and 2, we decided to analyze 2 different types of metrics, runs scored per over and wickets lost per over for both the team of interest and their opponents. We perform 2 separate analyses for the matches won and matches lost separately to understand the trend better.

Observation and Design Decisions:

We observe that in the case of winning matches, the winning team generally dominates the powerplay by scoring more runs, losing fewer wickets, and scoring more runs in the death overs. The exact opposite trend is observed in the case of losing matches. For the visualizations of these metrics, we used line plots as the metrics are essentially frequencies across different overs. An alternate plot, we considered was the bar plot.

We also plotted a scatter plot of runs scored vs wickets lost in a match, to see if a pattern emerges. We noticed that saving more wickets at the end of an innings didn't have a major impact on the chances of winning. This is primarily because teams having more wickets tend to take more risk to score runs in the death overs, thereby ending up losing wickets.

Dataset:

We used the [IPL Complete Dataset \(2008-2020\)](#) available on the Kaggle for our study. This dataset has Ball-by-Ball statistics of all the matches that happened between the years 2008 and 2020. We performed exploratory analysis on the dataset after cleaning the data and ended up removing data from the few matches that were hosted outside India and also removed teams that participated in less than 3 seasons.

Overview of the Development process:

Both the teammates worked together and made equal contributions to all the following processes:

1. Filtering and choosing the dataset - 4 hours
2. Data Cleaning and Exploratory Data Analysis - 4 hours
3. Deciding the right set of questions to answer - 1 hour
4. Getting familiar with streamlit - 2 hours
5. Coding our application - 5 hours
 - a. Our page layout was inspired by a couple of examples([proj1](#),[proj2](#)) provided on the streamlit website. Part of code snippets for the layout was resued from these projects.

6. Hosting the Demo - 30 mins
7. Writing the report - 3 hours