Analysis of INDIAN PREMIER LEAGUE dataset

The Indian Premier League is a professional men's Twenty20 cricket league, contested by ten teams based out of Ten Indian Cities. The league was founded by the Board of Control for Cricket in India in 2007.

Dataset

The dataset used for this project was found on Kaggle. The basic idea of analyzing the IPL dataset is to get a fair idea about 1) how many IPL matches were played in total? 2) How many seasons are we analyzing? 3) Which team scored the most runs?

These kind of analysis can be done using the data, by studying the factors such as
What is the probability of winning a match if the toss was won?
Which is the most successful IPL team with all the data at hand?

Tools & Libraries

• Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks

Data Description

- ID
- Season
- City
- Date
- Team1
- Team2
- Toss-Winner
- Toss-Decision
- Result
- dl-applied
- winner
- win by runs
- win by wickets
- Player of the match

- Venue
- Result-margin
- Eliminator
- Method
- Umpire1
- Umpire2
- umpire3

Data Cleaning

1) To find of NaN values
2) To Drop NaN Values

EDA

I looked at the different data and below is a few highlights of theanalysis.

- Import Libraries
- To read file using
- Size of columns and rows
- Top 5-Observation
- Bottom 5-Observation
- Information of Data
- Name of columns
- To find out missing values
- Heatmap for NaN values
- Drop NaN Values
- Heatmap after drop NaN Values
- Some statistical Inference
- No. of Season Played
- No. of matches Played

- The team won by Maximum Runs
- The team won by Maximum wickets
- The team won by Minimum wickets
- The team won by Minimum wickets
- The Bar plot for Season,result,toss_decision,winner
- Heatmap using Correlation
- Pairplot
- Probability of win if the toss was won
- Highest wins by teams per season
- In which city were the number of matches played