## Title: IPL Infographics: Data Analytics & Data Visualization

1. **Project Statement:**

The Indian Premier League (IPL) is a popular Twenty20 cricket tournament held annually in India. This project aims to perform exploratory data analysis and create data visualizations on an IPL dataset.

**Outcomes:**

* Gain insights into various aspects of the IPL, such as player performance, team trends, and venue analysis.
* Identify patterns and relationships within the data through visualizations.
* Create informative infographics that effectively communicate these insights to a wider audience.

1. **Modules to be Implemented:**
2. Data Collection
3. Data Exploration and Preprocessing
4. Model Building
5. Model Evaluation & Presentation
6. **Week-wise Module Implementation and High-Level Requirements with Outputs Screenshots**

**Milestone 1: Weeks 1 & 2**

**Week 1 (DC1, DC2, DC3):**

* + Understand the project goals and data analysis techniques.
  + Identify relevant IPL data sources (e.g., official websites, sports analytics platforms).
  + Acquire data from multiple sources, ensuring data quality and consistency.
  + Clean and prepare the data by merging datasets, handling missing values, and formatting data types.

**Deliverables:**

* Approved master dataset.

**Week 2 (DEP1, DEP2):**

* + Conduct exploratory data analysis to understand the overall statistics of the tournament and individual variables.
  + Perform univariate analysis to examine the distribution and characteristics of each variable.
  + Generate visualizations like histograms, box plots, and scatter plots to explore relationships between variables.

**Deliverables:**

* A detailed report on the variables used in the dataset.

**Week 3 (DEP3, DEP4, DEP5, DEP6):**

* + Address data type inconsistencies by converting variables to appropriate formats (e.g., numerical, categorical).
  + Analyze missing value patterns and employ suitable techniques like imputation or deletion to handle missing data.
  + Identify outliers using methods like z-scores or interquartile range (IQR) and determine appropriate treatment strategies.
  + Investigate trends, seasonality, and randomness components within the data. Perform time series decomposition if applicable.

**Deliverables:**

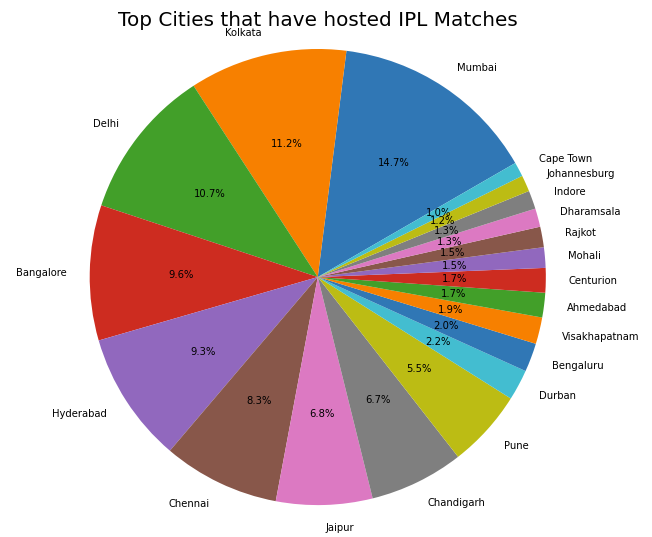
* Analysis of the venues for the tournament, potentially including visualizations like bar charts or maps.

**Week 4 (DEP4, DEP5):**

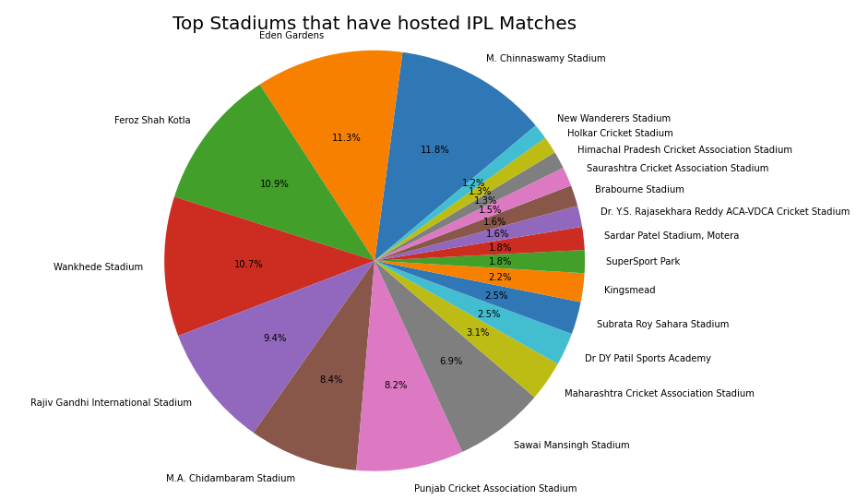
* + Employ various techniques like mean/median imputation, forward fill, or interpolation to address missing values based on data characteristics.
  + Implement different outlier detection methods and document observations on data quality and potential impacts.
  + Create visualizations showcasing the top cities and venues with the most IPL matches played.

**Deliverables:**

* A list of the Top 20 Cities where the most number of matches have been played.



* A list of the Top 20 venues where the most number of IPL matches have been played.

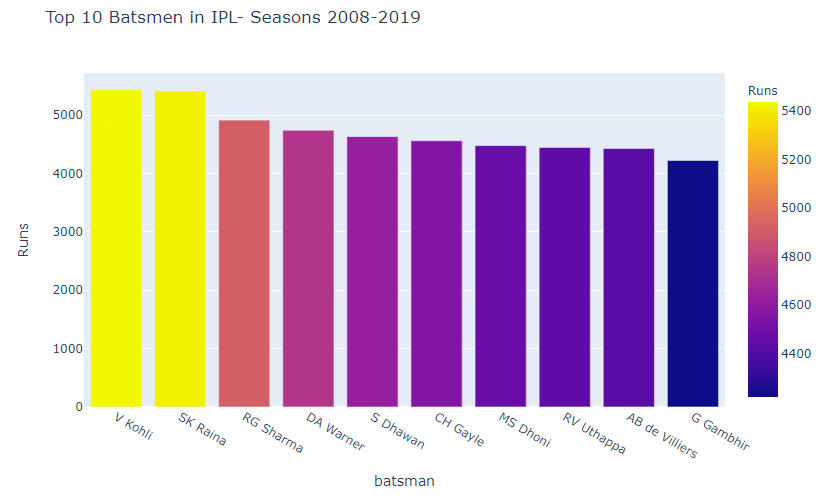


**Week 5 (MB1, MB2):**

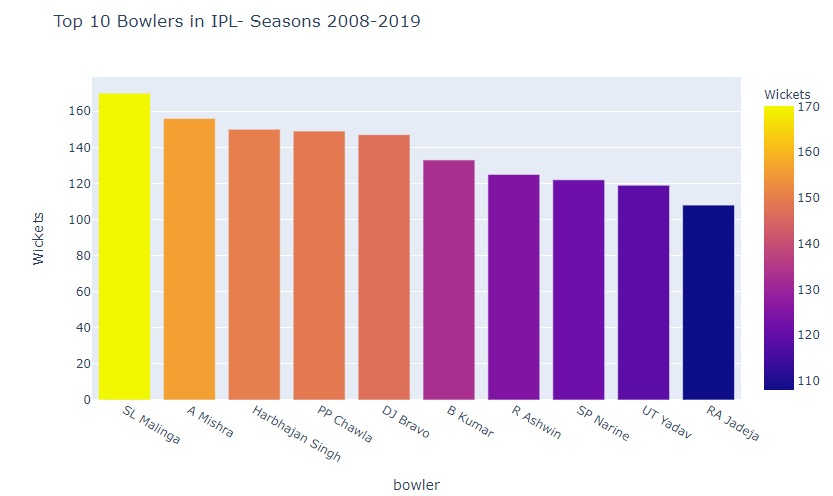
* This week focuses on time series forecasting, relevant if the project aims to predict future trends.
  + Identify suitable time series models (e.g., ARIMA, SARIMA) based on data characteristics.
  + Split the data into training, testing, and validation sets for model evaluation.
  + Build multiple models on the training data, potentially addressing class imbalance if present.

**Deliverables :**

* Player Level Analysis reports including:
  + Top 10 Scoring Batsman in the Tournament



* + Top 10 highest scorers in a match of IPL
  + Top 10 Bowlers with the highest number of wickets.

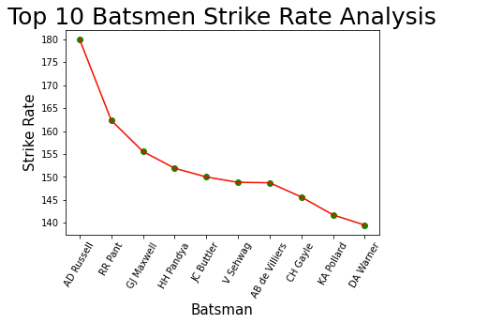


## Week 6 (MB3, MB4):

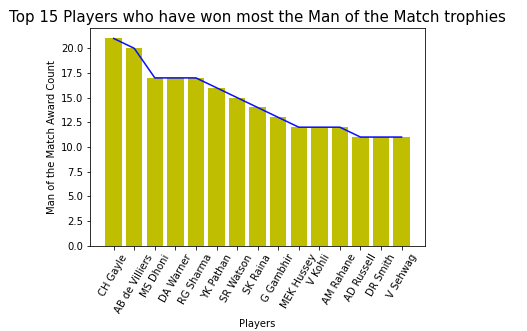
* This week focuses on finalizing the model selection and hyper parameter tuning.
  + Compare model performance using metrics like mean squared error (MSE) or R-squared to select the best model.
  + Select the best performing model based on validation set performance metrics.
  + Fine-tune the hyper parameters of the chosen model to optimize its accuracy and generalization capabilities.
  + If time series forecasting is not pursued, dedicate this week to further data exploration and analysis to identify interesting insights for visualization.

**Deliverables :**

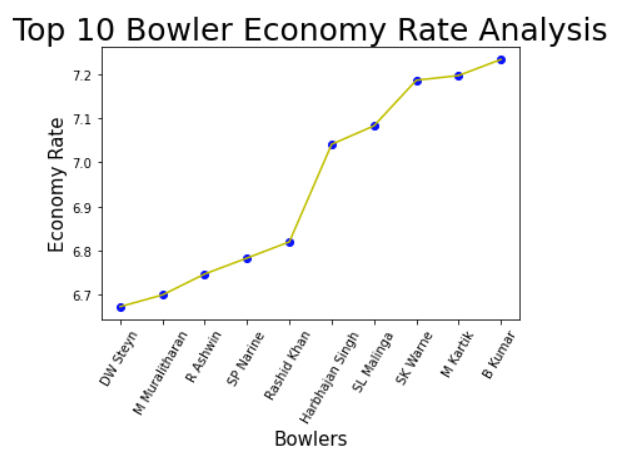
* Player Level Analysis reports including:
  + Strike Rate calculation for batsmen with a minimum target run threshold.



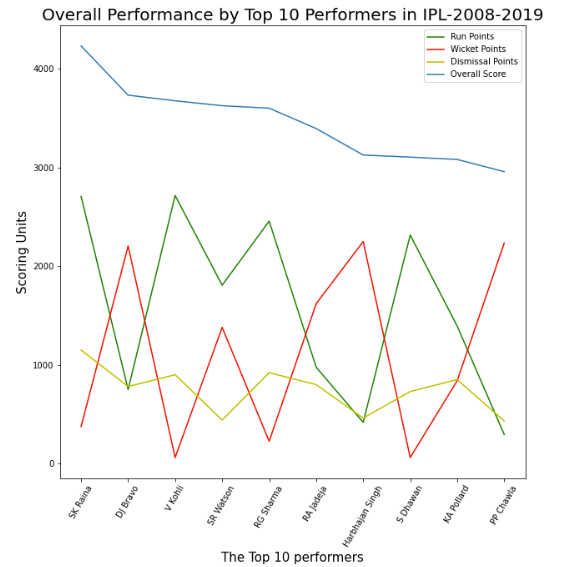
* + List of players with the highest number of 'Man of the Match' awards.



* + Economy rate calculation for bowlers exceeding a specific ball limit.



* + Best all-rounder performance considering batting, bowling, and fielding factors.

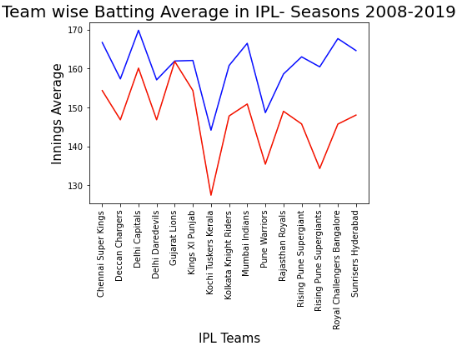
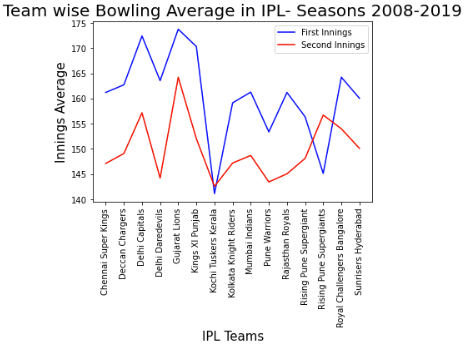


**Week 7 (MEP1, MEP2, MEP3):**

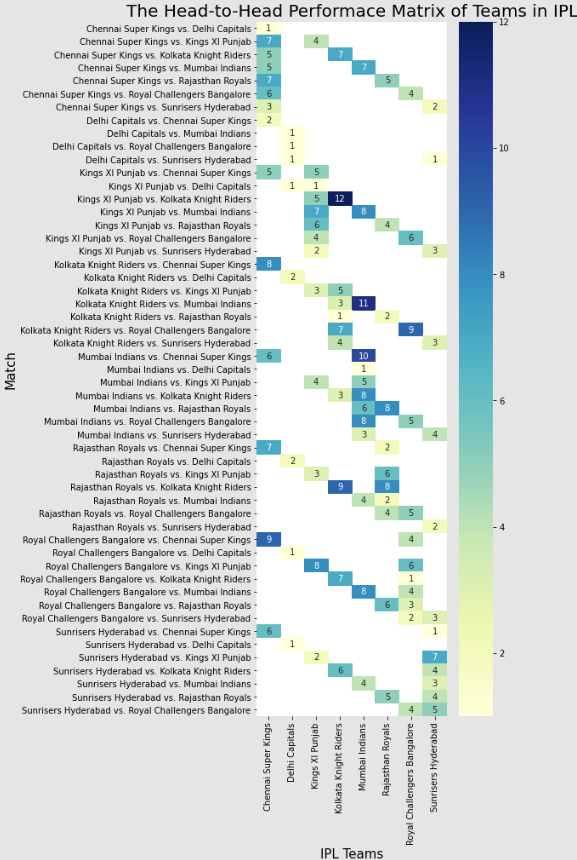
* + Finalize the chosen model (if applicable) and ensure its robustness.
  + Calculate performance metrics on the validation data to assess the model's generalizability.
  + If model performance is unsatisfactory, consider revisiting previous steps or exploring alternative approaches.
  + Prepare a presentation or project document summarizing the data analysis process, key findings, and visualizations.

**Deliverables:**

* Team-wise Analysis reports including:
  + Innings-wise batting and bowling averages for each team.



* + Win/loss analysis by runs or wickets for each team.
  + Head-to-head match analysis between IPL teams.

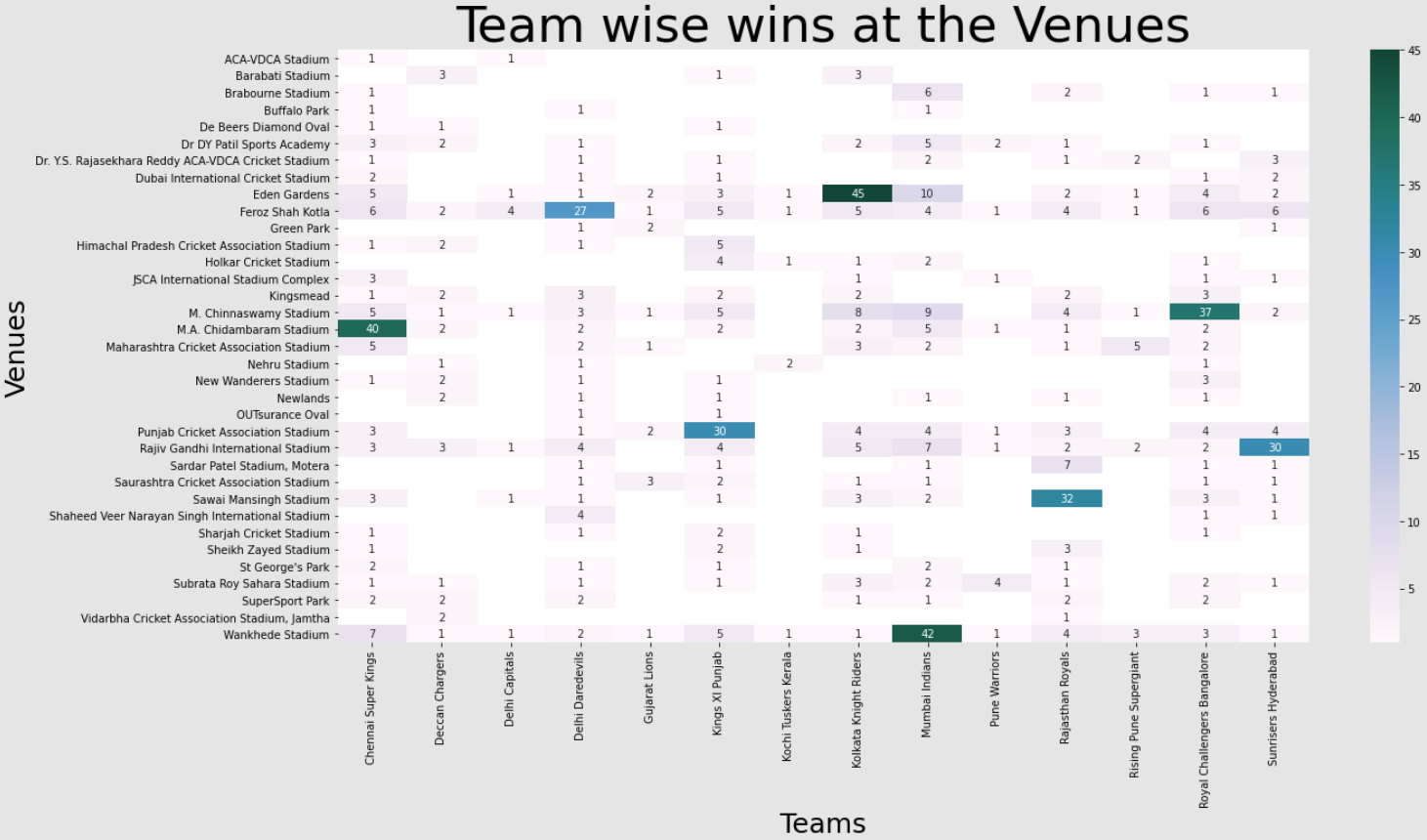


**Week 8 (MEP4, MEP5):**

* + Conduct final testing to ensure the project's functionality and presentation clarity.
  + Optimize and modularize code for better readability and maintainability.
  + Document the project comprehensively, including code comments, explanations, and user guides.
  + Develop a remediation plan outlining potential issues and mitigation strategies for future use.

**Deliverables:**

* Team-wise Analysis reports including:
  + Team winning performance at different venues.



* + Venue-wise best performers.
  + Heatmap of toss decisions taken by venue and their impact on win/loss.



* Project Documentation including:
  + Final presentation or report.
  + Final code with documentation.
  + Remediation plan.

1. **Evaluation Criteria:**

**Milestone 1 Evaluation (Week 2):**

* **Focus:** Data Collection (DC)
* **Deliverables:**
  + Completion of data acquisition from chosen sources.
  + Creation of a master dataset for analysis.
  + Univariate analysis report identifying relevant independent variables.

**Milestone 2 Evaluation (Week 4):**

* **Focus:** Data Exploration and Preprocessing (DEP)
* **Deliverables:**
  + Documentation of data cleaning and preprocessing techniques applied.
  + Report outlining chosen methods for handling missing values and outliers.

**Milestone 3 Evaluation (Week 6) - Optional**

* **Focus:** Model Building (MB)
* **Deliverables:**
  + Performance metrics reports for all models built during time series forecasting.

**Milestone 4 Evaluation (Week 8):**

* **Focus:** Model Evaluation & Presentation (MEP)
* **Deliverables:**
  + Finalized and approved model (if applicable).
  + Final project documentation or presentation summarizing the entire process, key findings, and visualizations.
  + Final code with proper documentation (if applicable).
  + Remediation plan outlining potential issues and mitigation strategies.
  + Submission of Action Tracking (AT) report (if applicable).
  + Code review completion in the designated Github repository.