Project Overview

In this major capstone project, you will develop machine learning models to predict two key outcomes for IPL 2025:

1. **Toss Winner Prediction**: Predict which team will win the toss before each match.
2. **Match Winner Prediction**: Predict the winner of each match based on historical data.

This project will require students to analyze historical IPL data, extract meaningful insights, and build predictive models. The dataset provided consists of two key files:

- **matches.csv**: Contains match-level details, including teams, venues, toss results, match results, etc.
    - Key columns include:
    - id: Match ID
    - season: Year of the IPL season
    - city: City where the match was played
    - team1 & team2: Teams playing the match
    - toss_winner: Team that won the toss
    - toss_decision: Decision taken (batting or fielding)
    - winner: Match-winning team
    - player_of_match: Best performer of the match

- **deliveries.csv**: Contains ball-by-ball data for each match, including runs scored, wickets taken, and other key in-game statistics.
    - Key columns include:
    - match_id: Corresponds to id in matches.csv
    - inning: Inning number (1 or 2)
    - batting_team: Team currently batting
    - bowling_team: Team currently bowling
    - batsman, bowler, non_striker: Players involved in the ball
    - runs, wickets, extras: Ball outcome statistics

Additionally, students will use the **IPL 2025 schedule** to make predictions for the upcoming season.

Datasets:

- **[IPL Historical Data](#)**
- **[IPL 2025 Match Schedule](#) - You have to Make predictions from Match 1 to Match 70**

Data Preprocessing and Feature Engineering

- **Handling Missing Values**: Identify and handle missing data appropriately.
- **Feature Engineering**:
    - Calculate team performance metrics (win percentage, toss win percentage, average runs, etc.).
    - Generate features based on toss-winning patterns.
    - Venue-based and team-based performance analysis.
- **Data Splitting**: Separate data into **train (past seasons)** and **test (IPL 2025 schedule)**.

EDA

- Visualize trends in toss wins and match wins.
- Analyze home advantage, impact of toss decisions, and batting/bowling first preferences.
- Identify strong and weak teams over the years.
- Find key players and their performance impact.

Model Building

Model 1: Toss Winner Prediction

- Train a **classification model** using features such as:
    - Historical toss wins per team
    - Toss win percentage at different venues
    - Captain-based toss decisions
    - Match conditions (home/away, day/night)

Model 2: Match Winner Prediction

- Train another **classification model** using features like:
    - Team performance history
    - Toss results (impact of toss on win probability)
    - Past head-to-head records
    - Venue performance trends
    - Player form (batting & bowling averages)

Model Evaluation & Validation

- Use performance metrics such as:
    - **Accuracy, Precision, Recall, F1-score** for classification models.
    - **Confusion Matrix** to analyze correct and incorrect predictions.
    - **ROC-AUC Curve** to measure classification performance.
- Perform **cross-validation** to ensure model generalization.

Final Prediction for IPL 2025

- Use the trained models to predict:
    a. Toss winners for each match in the **IPL 2025 schedule**.
    b. Match winners based on model probabilities.
- Summarize and interpret the results.

General Instructions for Submission

The submission should follow a **data storytelling approach** where students present their findings match-by-match for IPL 2025. The submission format should be a **Markdown (.md) file uploaded on GitHub**.

report.md *(Main Submission File)*
- This file should follow a structured storytelling approach.
- For each IPL 2025 match, include:
  - **Match details** (teams, venue, date)
  - **Insights from historical data** (team form, head-to-head record, toss trends)
  - **Model prediction results** (who wins the toss, who wins the match)
  - **Code snippet of final decision from the model**
- This format ensures students practice **data storytelling**, a crucial skill in real-world data science projects.

Example of report.md file

```
# IPL 2025 Toss & Match Winner Predictions

## Introduction
This report presents match-by-match predictions for IPL 2025 using machine learning models.

---

## Match 1: Chennai Super Kings 🆚 Mumbai Indians
**Date:** March 22, 2025
**Venue:** Wankhede Stadium, Mumbai

### 🔍 Insights from Historical Data
- Mumbai Indians have won **70% of tosses** at Wankhede since 2010.
- Chennai Super Kings have won **8 of the last 10 matches** against MI.
- Teams batting second have a **65% win rate** at this venue.

### 📊 Model Predictions
✅ **Toss Winner:** Mumbai Indians
✅ **Match Winner:** Chennai Super Kings

### 👨‍💻 Code Snippet (Final Decision)
```