7/29/2025
Database:
https://www.kaggle.com/datasets/diishasiing/cricket-analysis-dataset?resource=download
Description of Database:
Database uses statistics from 2022 IPL

(Description taken straight from Kaggle)
About Dataset
This dataset is ideal for cricket enthusiasts, data scientists, and analysts looking to derive insights into player performance trends, match outcomes, and predictive modeling in the sport of cricket.

This dataset has been meticulously scraped from ESPN's cricket match data. It includes comprehensive match data, player performance statistics, and game outcomes, which are essential for conducting in-depth analyses. The dataset has been cleaned and pre-processed to ensure accuracy, making it ready for data exploration, visualization, analytics and machine learning applications.

Files: (File description taken straight from Kaggle)
df_batting.csv : Contains detailed batting summary data, including player stats like runs, balls, and strike rate.
df_bowling.csv : Includes bowling stats such as overs, runs conceded, wickets, and economy rate.
df_match.csv : Captures match details including results, teams, winner, margin, venue, date, and match ID.
df_players.csv : Contains player profiles with team, batting/bowling style, role, and a brief description.

Deliverables:
Understand Dataset
Answer Analytical Questions
Create Dashboard - Player Stats Explorer
Optional: Create Model - Logistic Regression or XGBoost

- The model should predict what batsman and bowler would be a good matchup depending on the ground.

Project Goals:
Who are the most consistent players?

What factors contribute most to winning a match?

Is chasing better than batting first in IPL?

Tools:

Jupyter, Tableau

Libraries:
Pandas, NumPy, Matplotlib, and Seaborn

Project Summary:

Analyzed the databases of batting, bowling, match, and players. Did a base analysis using JupyterLab. Created a dashboard in Tableau that shows player stats and other stats related to the matches in IPL.

Key Findings:

- Highest runs: Virat Kohli
- Highest Wickets: Wanindu Hasaranga de Silva
- England had the highest number of wins.
- Most games are won by a 5 wicket margin.
- Highest Strike Rate player, Suryakumar Yadav with 1096 strike rate, is only behind Virat Kohli in runs by 57, who has a strike rate of 779.

Other key questions relate to AI and will be answered in another project that will have an AI setup to answer questions.

Future of project:

Analysis can be used to determine different winning attributes for IPL matches. The ability to see individual player stats allows viewers to compare and contrast players. The project is made to be a simple and fun overview of players during the 2022 Cricket IPL.

Documentation:
Use “python -m jupyterlab” to access jupyter lab through cmd

7/29/2025: Accomplished Analysis Deliverable
Used JupyterLab to analyze datasets with Pandas, NumPy, Matplotlib, and Seaborn libraries.

- Read data in
- Checked data structures
- Checked null values for datasets
- Made bar graph for top run scorers
- Made bar graph for top wicket takers
- Made bar graph for match wins per team
- Made bar graph for total runs per ground
- Checked top bowling averages

7/30/2025: Accomplished Dashboard with Player Stats Explorer
Used Tableau to create worksheets.

- Made bar graph for number of match wins per team
- Made bar graph for win run margin distribution
- Made bar graph for win wicket margin distribution
- Made bar graph for top 10 run scorers
- Made scatter plot for strike rate vs runs
- Made scatter plot for economy rate vs wickets
- Made player stats worksheet
- Made bar graph for player runs per ground
- Made bar graph for player runs per match

Used Tableau to create dashboard