International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

IPL DATASET ANALYSIS USING PANDAS

[1]N.Bhargavi,[2].M.Lasya Priya,[3]G.Sravan kumar,[4]J.Akshay

[1]Assistant Professor, [234]Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

bhargavi.cse@siddhartha.co.in, 23TQ1A6750@siddhartha.co.in, 23TQ1A6745@siddhartha.co.in, 23TQ1A6713@siddhartha.co.in

ABSTRACT

The Indian Premier League (IPL) is one of the most popular professional cricket leagues in the world, generating large volumes of match and player performance data every season. Analyzing this data helps identify trends in team strategies, player performance, and match outcomes. This project focuses on analyzing IPL datasets using the Python library pandas, which is widely used for data manipulation and exploratory data analysis. The dataset typically includes detailed information such as match results, teams, venues, players, runs scored, and wickets taken. Such datasets enable analysts to study patterns in cricket matches and extract meaningful insights from historical IPL seasons. The main objective of this project is to perform exploratory data analysis on IPL datasets to understand various aspects of the tournament. The dataset generally contains two major files: matches data, which includes information such as teams, toss decisions, match winners, and venues, and deliveries data, which provides ball-by-ball details like batsman runs, bowler performance, and wickets. By applying Pandas operations such as filtering, grouping, aggregation, and sorting, the project identifies key insights including the most successful teams, top-performing players, match trends across seasons, and the impact of toss decisions on match outcomes. The analysis also uses visualization tools to represent patterns and statistics through graphs and charts, making the results easier to interpret. Through this process, the project demonstrates how data analytics techniques can be applied to sports data to support performance evaluation and strategic decision-making. Overall, the project highlights the importance of data

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

analysis in modern sports and shows how Python and Pandas can be effectively used to explore large datasets and derive valuable insights from cricket match data.

I INTRODUCTION

The Indian Premier League (IPL) is one of the most popular professional Twenty20 cricket leagues in the world. Since its launch in 2008, the league has attracted millions of fans and generated large amounts of statistical data from every match, including information about teams, players, venues, scores, and match results. This vast amount of cricket data provides an excellent opportunity for data analysis and sports analytics. By analyzing such data, meaningful insights can be obtained about player performance, team strategies, and match outcomes. In recent years, data analytics has become an important tool in sports for evaluating team performance and making strategic decisions. The IPL dataset contains detailed records of matches and ball-by-ball deliveries, which allow analysts to study various aspects of the game. Typical IPL datasets include files such as matches.csv, which contains summary information about each match, and deliveries.csv, which provides ball-by-ball details including batsman, bowler, runs scored, and wickets taken. These datasets help analysts understand patterns such as the most successful teams, the best players, and trends across different IPL seasons. This project focuses on analyzing IPL datasets using the Python library pandas, which is widely used for data manipulation and analysis. Pandas allows efficient handling of large datasets through operations such as filtering, grouping, aggregation, and statistical analysis. By applying these techniques, the project aims to explore IPL match statistics, identify performance trends, and visualize insights about teams, players, and match results. The analysis demonstrates how data science techniques can be applied to sports datasets to generate meaningful conclusions and support better understanding of cricket performance data.

II LITERATURE SURVEY

Several studies have been conducted on analyzing cricket data, especially from the Indian Premier League, using data analytics and machine learning techniques. Researchers have shown that IPL generates large volumes of data including match results, player statistics, venue details, and ball-by-ball events. These datasets can be analyzed to identify trends in team performance, batting and bowling efficiency, and match outcomes. Previous research has applied statistical analysis and visualization methods to evaluate player consistency, scoring patterns, and team strategies. Such studies demonstrate that sports analytics can help coaches, analysts, and teams make better strategic decisions based on historical match data. Many research works have

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

also used programming tools such as Python and libraries like Pandas, NumPy, and Matplotlib to analyze IPL datasets. These tools help perform data preprocessing, filtering, grouping, and aggregation on large datasets to extract useful insights. Some studies have gone further by applying machine learning algorithms such as Support Vector Machines, Random Forest, and Logistic Regression to predict match outcomes and player performance based on historical IPL data. These approaches show that data-driven techniques can significantly improve understanding of cricket statistics and help build predictive models for sports analytics. Recent studies also emphasize the role of data analytics and programming tools in understanding cricket performance and match strategies. Researchers have used IPL datasets containing match summaries and ball-by-ball information to study factors such as batting strike rates, bowling efficiency, venue influence, and team winning patterns. Many studies use Python libraries like Pandas, NumPy, and Matplotlib to clean and process the data, perform statistical analysis, and visualize trends across multiple IPL seasons. These analytical approaches help identify key performance indicators for players and teams and reveal patterns that influence match outcomes. Such research highlights how data-driven analysis can improve decision-making in cricket analytics and demonstrates the practical application of data science techniques in sports performance evaluation.

III SYSTEM ANALYSIS

The system analysis for IPL dataset analysis using Indian Premier League focuses on understanding how large-scale cricket data is processed, analyzed, and transformed into meaningful insights using the Python library Pandas. The system is designed to handle structured datasets containing match details such as teams, players, scores, venues, and outcomes. It begins with data collection from reliable sources (CSV files, APIs, or databases), followed by preprocessing steps like handling missing values, removing duplicates, and formatting data for consistency. The core processing layer uses Pandas for operations such as filtering matches, grouping team performances, calculating player statistics (like strike rate, average), and identifying trends across seasons. The system also supports exploratory data analysis (EDA), enabling visualization of insights like top-performing teams, win percentages, and venue-based advantages. The output is presented in the form of tables, graphs, or dashboards, helping users easily interpret cricket analytics. Overall, this system improves decision-making, supports performance evaluation, and demonstrates efficient handling of sports data using data science techniques.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Existing system

The existing system for analyzing data from the Indian Premier League is mostly based on traditional and manual methods. Earlier, cricket data analysis was done using spreadsheets like Excel or basic database queries without advanced tools such as Pandas. Analysts had to manually collect match data, clean it, and perform calculations, which was time-consuming and prone to errors. Data was often scattered across different files, making it difficult to manage and analyze large datasets efficiently.

Additionally, these systems lacked automation and advanced analytical capabilities. Complex operations like grouping, filtering, and trend analysis required significant manual effort. Visualization options were also limited, making it harder to derive meaningful insights about team performance, player statistics, or match outcomes. As the IPL data grew over multiple seasons, handling large volumes of data became inefficient and slow in the existing system.

Overall, the existing system was less efficient, lacked scalability, had higher chances of human error, and did not provide deep insights compared to modern data analysis approaches.

DisAdavantages of Existing system

- Manual data handling for Indian Premier League increases chances of human errors
- Time-consuming process for data cleaning and analysis without automation
- Difficult to manage large datasets across multiple IPL seasons
- Limited analytical capabilities compared to tools like Pandas
- Lack of efficient data filtering, grouping, and aggregation features

Proposed system

The proposed system for analyzing data from the Indian Premier League is an automated and efficient data analysis solution built using Pandas. In this system, IPL datasets are collected and stored in a structured format such as CSV files, which are then processed using Pandas for data cleaning, transformation, and analysis. The system eliminates manual effort by automating tasks like handling missing values, removing duplicates, and organizing data for consistency. It enables advanced operations such as filtering matches, grouping team and player statistics, and performing comparative analysis across seasons. Additionally, the system supports

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

data visualization using libraries like Matplotlib or Seaborn to present insights in the form of graphs and charts. This helps in identifying patterns such as top-performing teams, player performance trends, and match outcomes effectively. Overall, the proposed system is faster, more accurate, scalable, and capable of providing deep insights, making it highly suitable for modern sports data analysis.

Advantages of Proposed System

- Faster data processing compared to manual methods for **Indian Premier League**
- Automated data cleaning (handles missing values and duplicates efficiently)
- High accuracy with reduced human errors
- Efficient handling of large datasets using **Pandas**
- Easy data filtering, grouping, and aggregation operations

IV METHODOLOGY

The methodology of the IPL Dataset Analysis using Pandas project describes the systematic process used to collect, process, and analyze IPL cricket data in order to extract meaningful insights. The project follows several stages including data collection, data preprocessing, exploratory data analysis, and visualization. Each stage plays an important role in converting raw cricket data into useful information about teams, players, and match outcomes. The methodology ensures that the dataset is properly organized and analyzed using Python tools, mainly the Pandas library, which is widely used for data manipulation and analysis.

1. Data Collection The first step in the methodology is collecting the IPL dataset from reliable sources such as Kaggle or other open cricket databases. The dataset usually consists of multiple CSV files that contain historical match information. Two common files used in IPL analysis are matches.csv, which contains match details such as teams, venues, toss results.
2. Data Preprocessing and Cleaning After collecting the dataset, the next step is data preprocessing and cleaning. In real datasets, there may be missing values, duplicate records, or inconsistent data formats that must be handled before performing analysis.
3. Exploratory Data Analysis (EDA) The third step is Exploratory Data Analysis (EDA), which involves examining the dataset to identify patterns and

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

relationships. In this stage, various Pandas operations such as filtering, grouping, sorting, and aggregation are applied to analyze the dataset.

4. Data Visualization The final step in the methodology is data visualization, where the results of the analysis are presented in graphical form. Visualization tools such as Matplotlib or Seaborn are used to create charts and graphs including bar charts, pie charts, and line graphs

System Architecture

The system architecture of the IPL Dataset Analysis using Pandas project describes the structure and workflow used to process and analyze IPL match data. The architecture follows a layered approach where the dataset flows through different stages such as data collection, preprocessing, analysis, and visualization. Each layer performs a specific task that transforms raw IPL data into meaningful insights. This architecture ensures efficient data handling, easy analysis, and clear presentation of results. The IPL dataset typically consists of structured CSV files such as matches.csv and deliveries.csv, which store detailed information about match results, teams, players, runs, wickets, venues, and other match statistics. These datasets are loaded into Pandas DataFrames, which allow efficient manipulation and analysis of large datasets using Python.

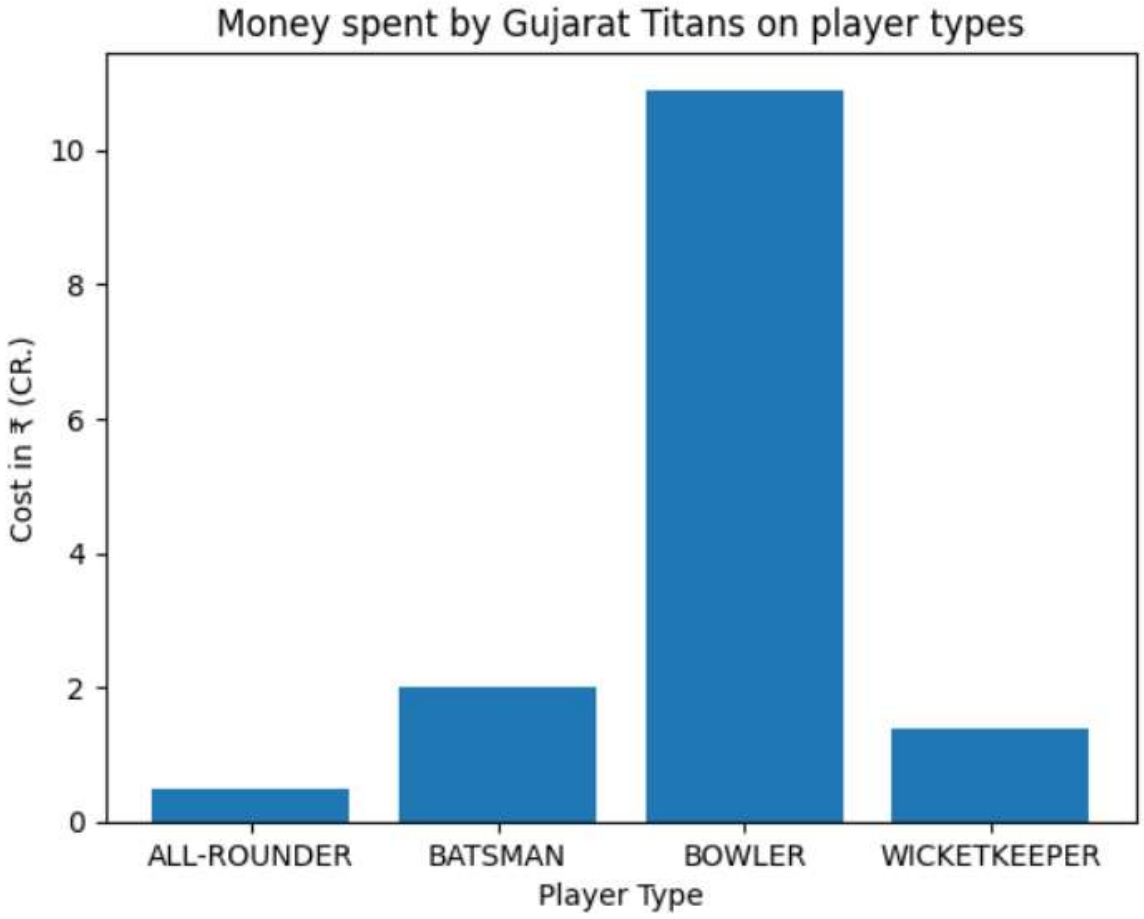
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

IV RESULTS &OUTPUT

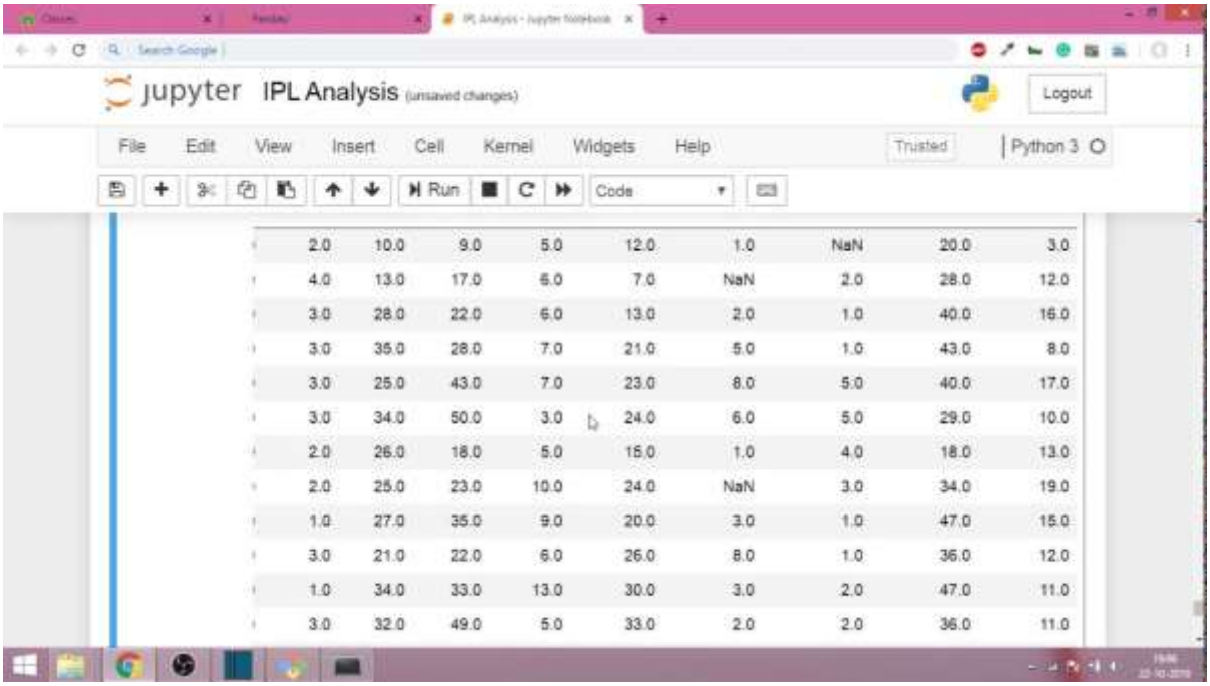


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



2.0	10.0	9.0	5.0	12.0	1.0	NaN	20.0	3.0
4.0	13.0	17.0	6.0	7.0	NaN	2.0	28.0	12.0
3.0	28.0	22.0	6.0	13.0	2.0	1.0	40.0	16.0
3.0	35.0	28.0	7.0	21.0	5.0	1.0	43.0	8.0
3.0	25.0	43.0	7.0	23.0	8.0	5.0	40.0	17.0
3.0	34.0	50.0	3.0	24.0	6.0	5.0	29.0	10.0
2.0	26.0	18.0	5.0	15.0	1.0	4.0	18.0	13.0
2.0	25.0	23.0	10.0	24.0	NaN	3.0	34.0	19.0
1.0	27.0	35.0	9.0	20.0	3.0	1.0	47.0	15.0
3.0	21.0	22.0	6.0	26.0	8.0	1.0	36.0	12.0
1.0	34.0	33.0	13.0	30.0	3.0	2.0	47.0	11.0
3.0	32.0	49.0	5.0	33.0	2.0	2.0	36.0	11.0

VI CONCLUSION

The IPL dataset analysis project using Pandas proves to be a highly effective solution for handling and analyzing large-scale cricket data from the Indian Premier League. The system not only simplifies complex data operations but also enables deeper insights into match performance, player efficiency, and team strategies through structured and well-organized analysis.

Furthermore, the project highlights the importance of data preprocessing and cleaning in achieving accurate results. By eliminating inconsistencies and handling missing data, the system ensures reliable outputs. The integration of visualization techniques also enhances the understanding of patterns and trends, making the analysis more interactive and user-friendly.

Additionally, the proposed system is scalable and can be extended in the future by integrating machine learning algorithms for prediction tasks such as match winners or player performance forecasting. It can also be connected to real-time data sources for live analytics. Overall, this project serves as a strong foundation for sports analytics and demonstrates the practical application of data science in real-world scenarios.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.