## **Artificial**

## **Intelligence and Machine Learning**

Project Report

Semester-IV (Batch-2022)

**IPL SPORTS PREDICTOR**



**Supervised By: Submitted By:**

Ms.Monica Dutta Mehak Garg - 2210990572

Mehak Goel - 2210990573

Mehak Kharbanda -2210990574

Mehak Sharma - 2210990575

**Department of Computer Science and Engineering**

**Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab**

**INDEX**

|  |  |  |
| --- | --- | --- |
| S.no | Topics | Page No. |
| 1.  2.  3.  4.  5.  6. | Introduction 1.1 Background 1.2 Motivation 1.3 Objectives 1.4 Significances 1.5 Features  Problem Definition and Requirements 2.1 Problem Statement 2.2 Software Requirements 2.3 Requirements 2.4 Datasets Hardware  Methodology 3.1 Problem Understanding and definition 3.2 Data Acquisition and exploration 3.3 Data Preprocessing and Feature Engineering 3.4 Model Selection and Requirement 3.5 Model Evaluation and Validation 3.6 User Interface and Development 3.7 Collaboration  Results  Conclusion  Reference | 1-5  5-7  7-9  9-22  23  24 |

1. **INTRODUCTION**

**1.1 BACKGROUND**

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India, which has not only refined the landscape of cricket but also become a cultural phenomenon since its inception in 2008. Organized by the Board of Control for Cricket in India (BCCI), the IPL features franchise teams representing different cities and regions of India, each comprising a mix of domestic and international cricketing talent.

The IPL has become synonymous with entertainment, attracting millions of viewers worldwide with its high-energy matches, celebrity endorsements and innovative marketing strategies. Its fast-paced format, coupled with the presence of renowned cricketers from around the globe, ensures that each game is a spectacle, filled with drama, suspense, and moments of brilliance.

The IPL Sports Predictor project aims to harness the power of data analytics and machine learning to develop robust predictive models that can accurately forecast match results and player performances in the IPL. By using advanced algorithms and techniques, such as regression analysis, classification models, and ensemble methods, the project seeks to provide cricket enthusiasts and stakeholders with valuable insights and predictions to enhance their IPL experience.

* 1. **MOTIVATION**

The motivation behind the IPL Score Predictor project lies in the intricate and dynamic nature of IPL matches, where a multitude of factors converge to determine the final outcome. These matches are not merely contests between bat and ball but are influenced by a complex interplay of variables, including player form, team strategies, pitch conditions, weather dynamics, and even intangible factors such as momentum and psychological factors.

For cricket enthusiasts, predicting the outcome of an IPL match adds an extra layer of excitement and engagement to the viewing experience. However, due to the inherent unpredictability of the sport and the multitude of factors at play, making accurate predictions manually can be challenging and often subjective.

This project seeks to address this challenge by harnessing the power of machine learning algorithms and advanced analytics to analyze historical match data and identify patterns and trends that can be used to predict match winner with greater accuracy.

With numerous factors influencing match outcomes, ranging from team composition and player to pitch conditions and weather dynamics, accurately predicting IPL results poses a significant challenge. By using machine learning algorithms and advanced analytics, this project endeavors to unravel the underlying patterns and trends in IPL data, enabling more informed predictions and strategic insights.

* 1. **OBJECTIVES**
     1. **Match Outcome Prediction:** Develops model to accurately predict the outcomes of IPL matches, including identifying the winning team.
     2. **Player Performance Analysis:** Analyze player statistics to predict individual performances, such as runs scored, wickets taken and other relevant metrics.
     3. **Data Integration:** Gather and integrate historical match data, team statistics, player metrics, and other relevant factors to enhance the predictive capabilities of the model.
     4. **User-Friendly Interface:** Design an intuitive and user-friendly interface to present predictions and insights, making them accessible to cricket enthusiasts and stakeholder.
     5. **Model Evaluation and Improvement:** Continuously evaluate model performance and refine algorithms to enhance accurate and reliable predictions throughout the IPL season.
     6. **Scalability and Performance:** Ensure scalability and performance of the predictive model to handle large volumes of data and deliver timely predictions during peak IPL periods.
     7. **Real-time updates:** Implement mechanisms to provide real-time updates on match predictions and player performances as match progress, ensuring that users have access to the latest insights and information.
     8. **Feature Importance Analysis:** Conduct feature importance analysis to identify the key factors that influence match outcomes and player performances, providing valuable insights into the dynamics of IPL matches.
     9. **Interpretability:** Enhance the interpretability of the predictive models by providing explanations or visualizations of the factors driving the predictions, enabling users to understand the rationale behind the predictions and build trust in the model.
     10. **Adaptive Learning:** Implement adaptive learning techniques to continuously update the predictive models based on new data and emerging trends, ensuring that the models remain relevant and effective throughout the IPL season.
     11. **Long-term Trend Analysis:** Conduct long-term trend analysis to identify recurring patterns and dynamics in IPL matches and player performances over multiple seasons, enabling stakeholders to anticipate and adapt to evolving trends and strategies.
     12. **Education and Outreach:** Develop educational resources and outreach programs to raise awareness about the role of data analytics and machine learning in sports prediction, empowering enthusiasts and professionals to harness data-driven insights for decision-making in cricket and beyond.
  2. **SIGNIFICANCES**
     1. **Educational Purposes:** The project serves as an educational tool for students and enthusiasts interested in learning about data science, machine learning and sports analytics
     2. **Enhanced Fan Experience:** By providing accurate predictions of match outcomes and player performances, the project enhances the overall viewing experience for cricket fans. It adds an extra layer of excitement and track player performances more closely.
     3. **Trend Analysis:** The project enables trend analysis and identification of patterns in team performance, player form, and match dynamics over time.
     4. **Global Reach:** The IPL sports predictor project transcends geographical boundaries and reaches cricket fans worldwide. It serves as a unifying platform for fans from diverse backgrounds to come together, share their passion for the sport and engage in discussions around match-predictions and outcomes.
     5. **Real-time Updates:** Implement mechanisms to incorporate real-time data updates during matches to enhance the timeliness and relevance of predictions.
     6. **Community and Social Interaction:** The project fosters community engagement and collaboration among cricket enthusiasts, data scientists and technologists.
     7. **Promotion of Data Literacy:** The project promotes data literacy by demonstrating the practical applications of data analysis and machine learning techniques in a context that resonates with a wide audience.
     8. **Technological Innovation:** The project drives innovation in the field of sports analytics by exploring new data sources, algorithms, and methodologies, pushing the boundaries of what is possible in predictive modeling and decision support systems.
     9. **Predictive Analytics Education:** The project serves as a practical case study for teaching predictive analytics concepts and techniques in educational settings, helping students develop critical thinking skills and real-world problem-solving abilities.

* 1. **FEATURES**
     1. **Match Outcome Prediction**: Predict the outcome of IPL matches, including the winning team.
     2. **Data Integration**: Gather and integrate diverse datasets, including historical match data, team statistics, player metrics and other relevant factors.
     3. **Machine Learning Models**: Develop and deploy machine learning algorithms, such as regression models, classification models, ensemble methods, and deep learning techniques, for predictive analysis.
     4. **Player Performance Analysis**: Analyze individual player performances, including runs scored and wickets taken.
     5. **User-Friendly Interface**: Design an intuitive and user-friendly interface for accessing predictions, exploring match insights, and interacting with the system effortlessly**.**
     6. **Historical** **Data Visualization**: Visualize historical IPL data using charts, graphs, heatmaps, and interactive visualizations to facilitate exploration and analysis by users.
     7. **Academic Research:** The project serves as a valuable resource for academic research in fields such as sports analytics, data science, and predictive modeling, contributing to advancements in knowledge and methodology.
     8. **Performance Evaluation:** The project facilitates the evaluation of team strategies and player performances, allowing teams to assess their strengths and weaknesses and make adjustments for future matches.
     9. **Technological Innovation:** The project drives innovation in the field of sports analytics by exploring new data sources, algorithms, and methodologies, pushing the boundaries of what is possible in predictive modeling and decision support systems.
     10. **Live Match Tracking:** Implement real-time tracking of IPL matches, providing users with updates on match progress, key events, and performance indicators as the action unfolds.
     11. **Match Momentum Analysis:** Develop algorithms to analyze the momentum shifts within matches based on key events, such as wickets, boundaries, and strategic decisions, providing insights into the ebb and flow of gameplay.

1. **PROBLEM DEFINITION AND REQUIREMENTS**
   1. **PROBLEM STATEMENT**

Predictive modelling, facilitated by advancements in computer technology, plays a

pivotal role in forecasting future events across various domains. In sports,

particularly cricket and football, prediction techniques are utilized to anticipate

scores, rankings, winners and other outcomes.

However, developing a reliable machine learning model to accurately predict IPL

match outcomes and analyse player performances poses a significant challenge. The

dynamic nature of cricket, coupled with the complexity of relevant factors, requires

a comprehensive approach that interconnected elements and advanced

computational techniques.

The IPL Sports Predictor Project aims to overcome this challenge by leveraging historical data, advanced predictive modeling techniques, and computerized algorithms to create a robust predictor. By accurately forecasting match outcomes and player performances, the project seeks to enhance user engagement and provide actionable insights for cricket enthusiasts and stakeholders. Through the integration of probability analysis, regression modeling, and other predictive techniques, the IPL Sports Predictor Project aims to revolutionize the way IPL matches are analyzed and experienced, offering users a valuable tool for strategic planning and informed decision-making in the realm of cricket.

* 1. **SOFTWARE REQUIREMENTS**
     1. **Programming Language:** Python
     2. **Data Analysis and Visualization:** Pandas, NumPy, Matplotlib, Seaborn
     3. **Machine Learning Libraries:** Scikit-learn
     4. **Development Environment:** Jupyter Notebook, Visual Studio Code
  2. **HARDWARE REQUIREMENTS**
     1. **Processor:** Multi-core processor
     2. **RAM:** 8GB or more
     3. **Storage:** 100GB or more
     4. **Operating System:** Windows, macOS, Linux
     5. **Internet Connection:** Required for downloading datasets, libraries and updates
  3. **DATASETS**

We have downloaded Datasets from the below link:

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 566 | 5 |

* + 1. **Players Dataset:** Player name, date of birth, country of origin, batting preference and bowling preference

**LINK:**

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=Players.xlsx>

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 179078 | 21 |

* + 1. **Delivery Dataset:** Delivery information

**LINK:**

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=deliveries.csv>

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 756 | 18 |

* + 1. **Matches Dataset:** Details of the match ties and matches where duckworth

lewis applied

**LINK:**

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=matches.csv>

#### **most\_runs\_average\_strikerate Dataset:** Batsmen Total runs average and

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 516 | 6 |

their strike rate

**LINK:**

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 15 | 1 |

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=most_runs_average_strikerate.csv>

#### **Team Dataset:**

**LINK:**

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=teams.csv>

#### **Teamwise\_home\_and\_away Dataset:** team home wins and away wins

|  |  |
| --- | --- |
| Total Rows | Total Columns |
| 15 | 7 |

**LINK:**

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set/data?select=teamwise_home_and_away.csv>

1. **METHODOLOGY**
   1. **PROBLEM UNDERSTANDING AND DEFINITION**
      1. Define the objectives of the project, including the specific prediction tasks (e.g., match outcome prediction, score estimation, player performance analysis) and the target audience for the predictor.The primary objective is to predict the outcome of IPL matches accurately. This includes forecasting the winning team.
      2. Conduct a thorough analysis of the problem domain, understanding the dynamics of IPL matches including the format (Twenty20),rules, team, compositions, player roles, strategies commonly employed by teams, relevant factors influencing outcomes, and existing approaches to sports prediction.
   2. **DATA ACQUISITION AND EXPLORATION**
      1. Gather historical IPL match data, team statistics, player performance metrics, pitch conditions, weather data, and other relevant features from reliable sources.

**HISTORICAL DATA:**

Collect data on past IPL matches, including match dates, teams playing venues, toss outcomes, innings details and match results (win/loss).

**TEAM STATISTICS:**

Obtain team-level statistics such as total runs scored, total wickets taken and win-loss records.

* + 1. Explore the collected data through descriptive statistics, data visualization, and exploratory data analysis (EDA) to gain insights into patterns, trends, and relationships within the dataset.

**DATA VISUALIZATION:**

Create visualizations such as histograms, box plots, scatter plots, and heatmaps to visualize the distribution of variables, relationships between variables, and trends over time.

**EXPLORATORY DATA ANALYSIS (EDA):**

Conduct exploratory data analysis techniques to identify patterns, trends, and relationships within the dataset.

* 1. **DATA PREPROCESSING AND FEATURE ENGINEERING**
     1. Clean the raw data to handle missing values, outliers, and inconsistencies, ensuring data quality and reliability.

**DATA CONSISTENCY CHECK:**

Perform checks for data inconsistencies, such as duplicate records, conflicting information or data format errors and resolve them to ensure data consistency and integrity.

* + 1. Engineer new features and transform existing features to capture relevant information and enhance the predictive power of the dataset.
  1. **MODEL SELECTION AND DEVELOPMENT**
     1. Select appropriate machine learning algorithms (e.g., regression, classification) based on the prediction tasks and characteristics of the dataset.
     2. Split the dataset into training, validation, and testing sets to train and evaluate the performance of the selected models.
     3. Train the models using the training data, optimizing hyperparameters and model parameters through techniques.

* 1. **MODEL EVALUATION AND VALIDATION**
     1. Validate the models on unseen data from the testing set to confirm their reliability and robustness in real-world scenarios.
     2. Evaluate the trained models on the validation set using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score) to assess their performance and generalization capabilities.

**APPROPRIATE EVALUATION METRICS:**

For match outcome prediction: Accuracy, Precision, F1-Score

**ADDITIONAL CONSIDERATIONS:**

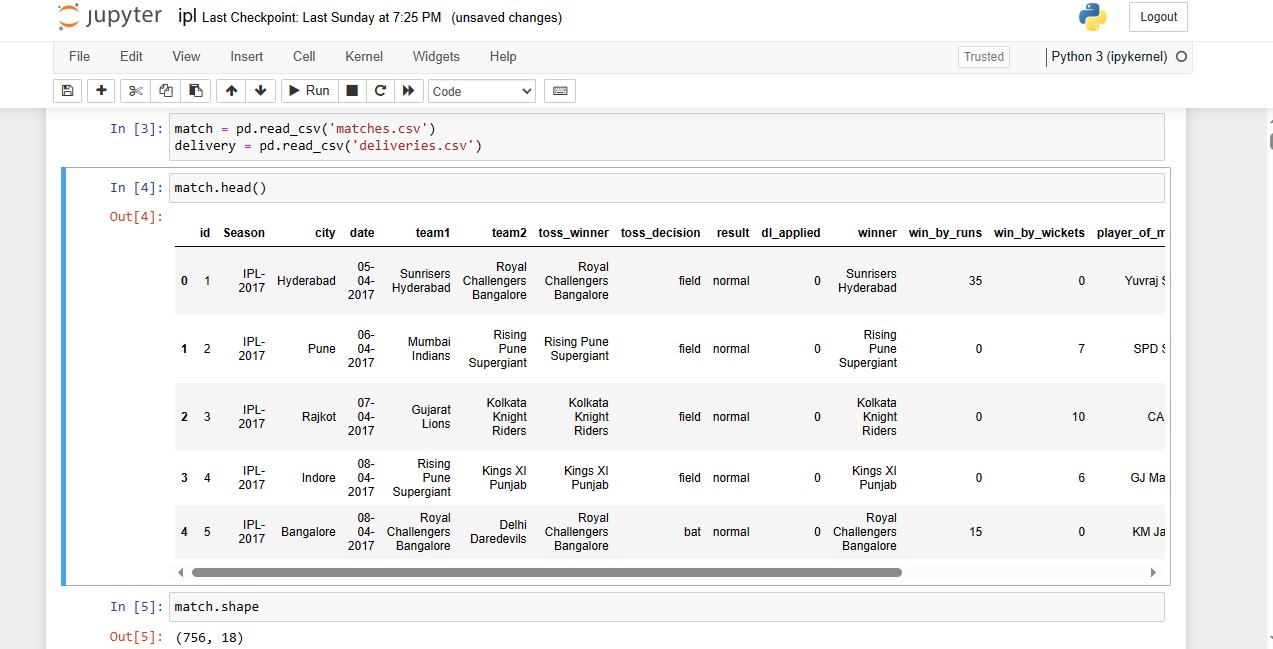
Overfitting: High performance (High Accuracy) on training data but poor performance (Poor Accuracy) on testing data.

Underfiiting: Poor performance (Poor Accuracy) on both testing and training data.

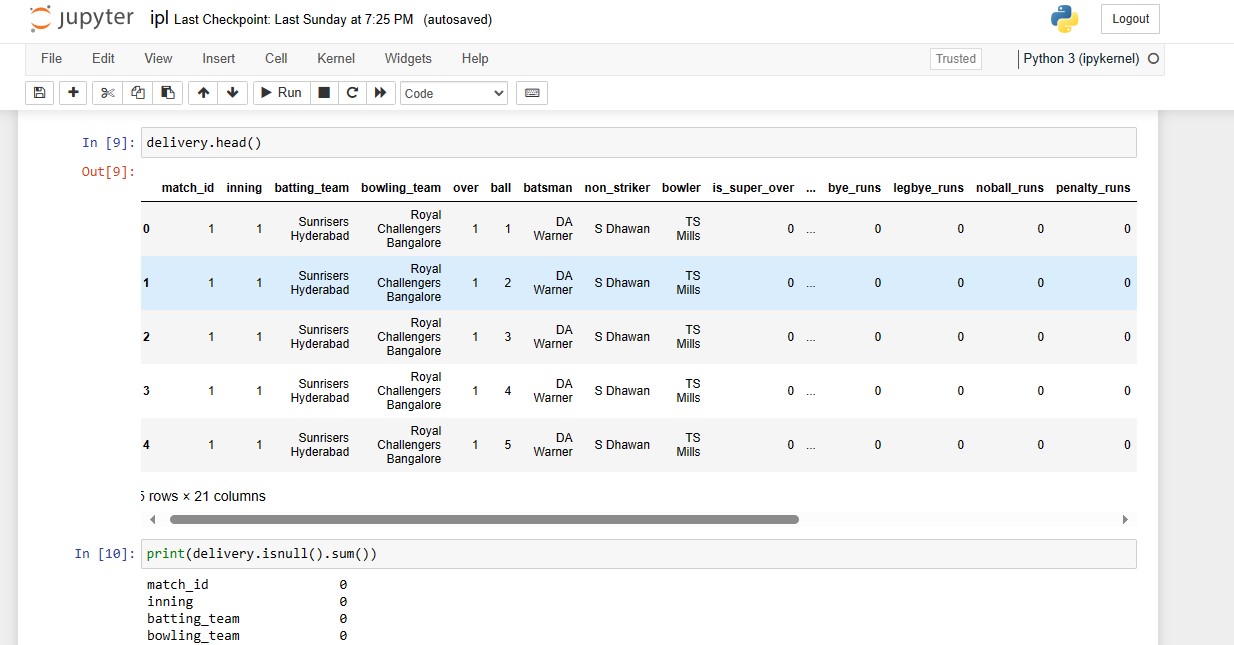
* 1. **USER INTERFACE AND DEVELOPMENT**
     1. Design interactive visualizations to present results and engage users effectively.
  2. **COLLABORATION**
     1. Foster collaboration among team members to leverage diverse perspectives and expertise in model development and feature engineering.

**4.RESULTS**

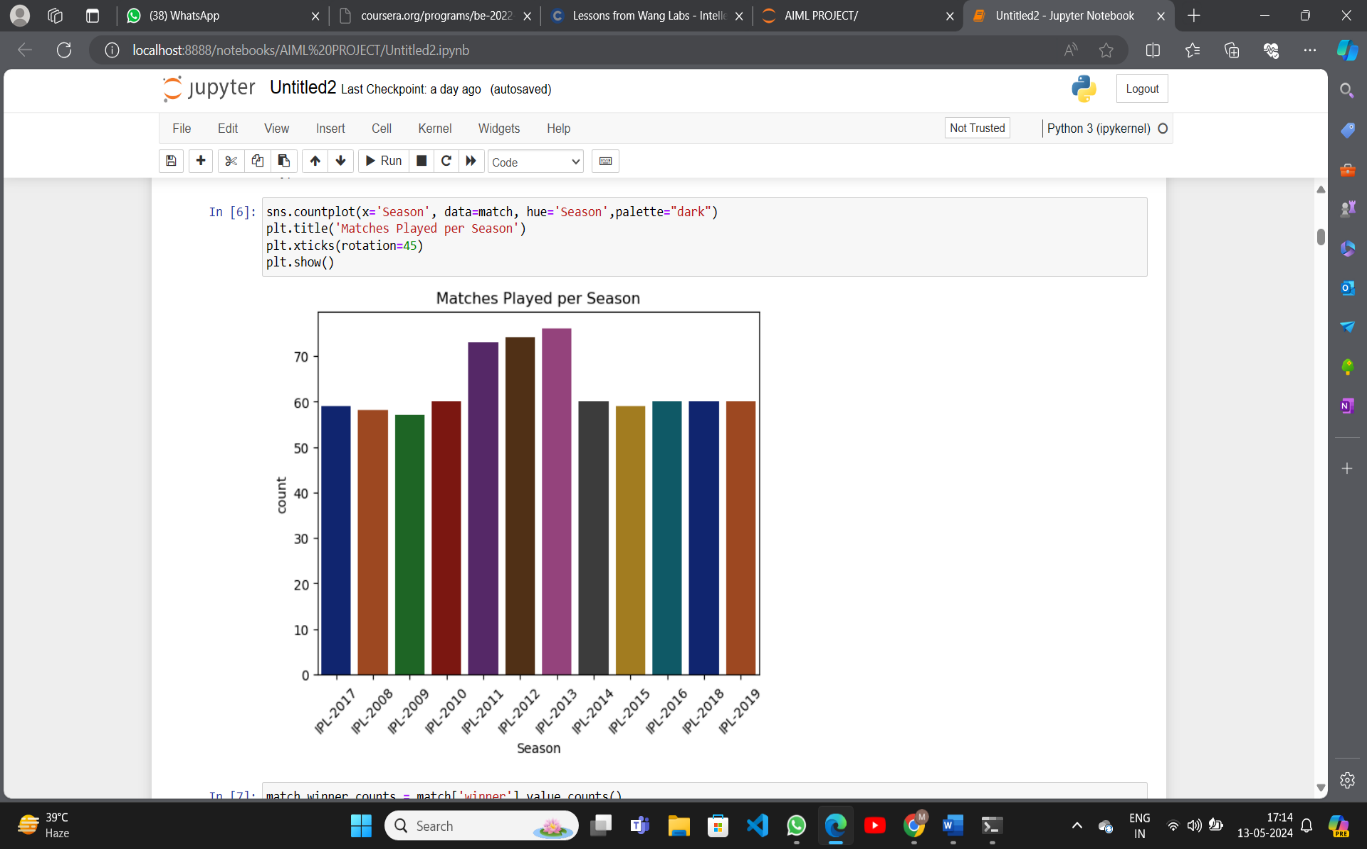
MATCH DATA



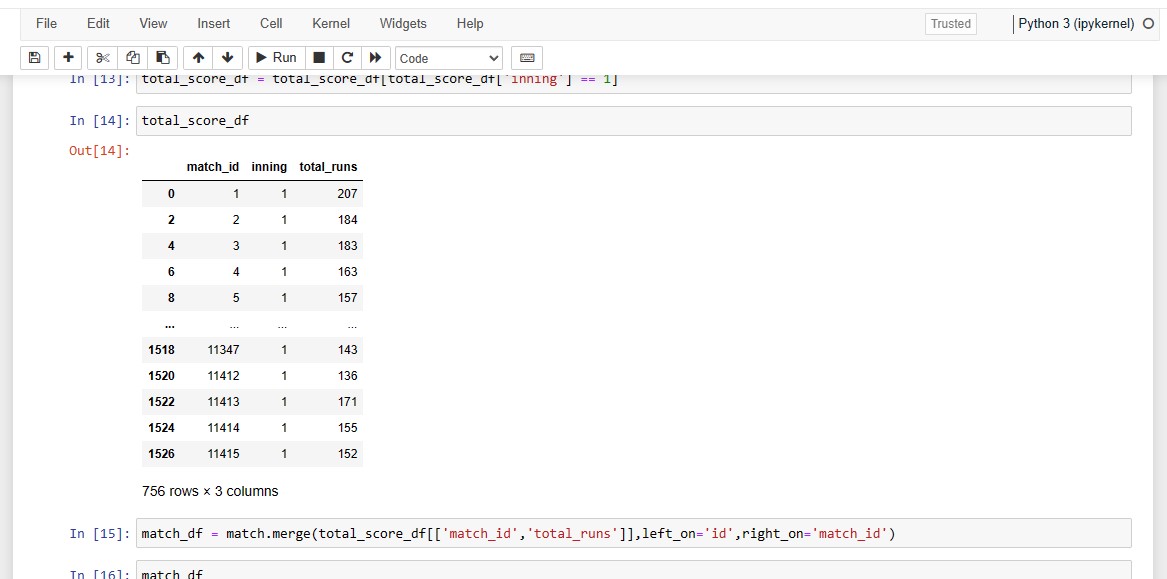
DELIVERY DATA



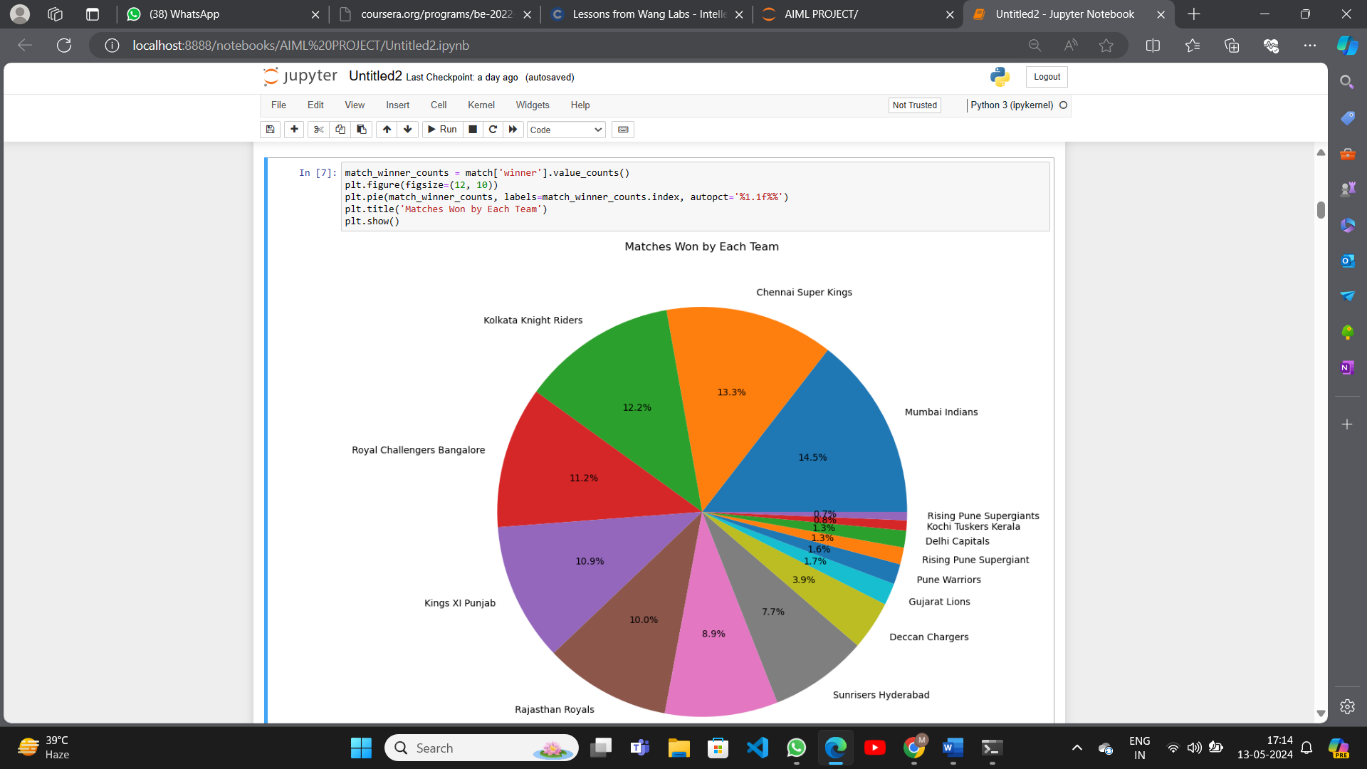
NUMBER OF MATCHES PLAYED PER SEASON



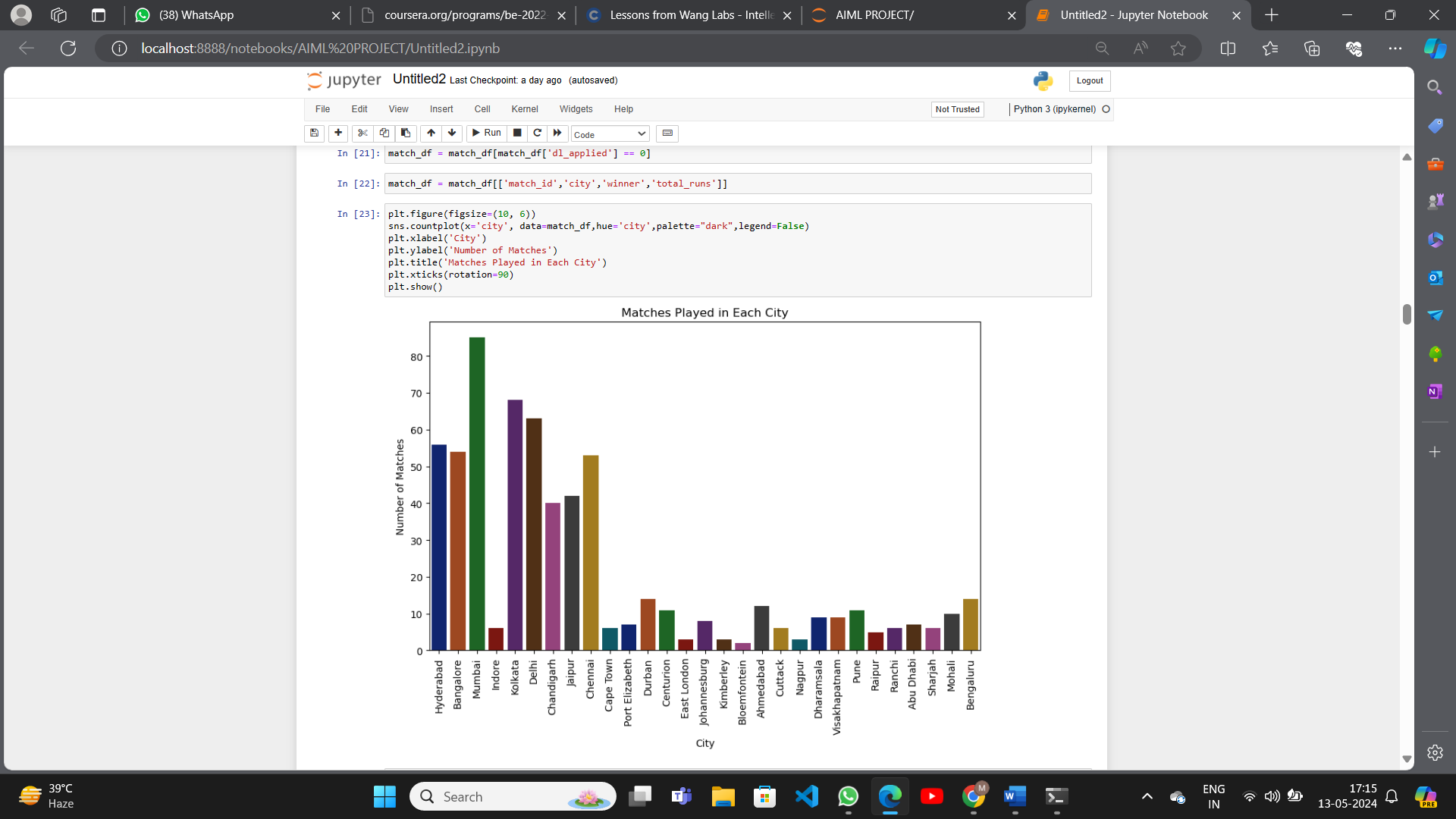
TOTAL RUNS OF FIRST INNINGS



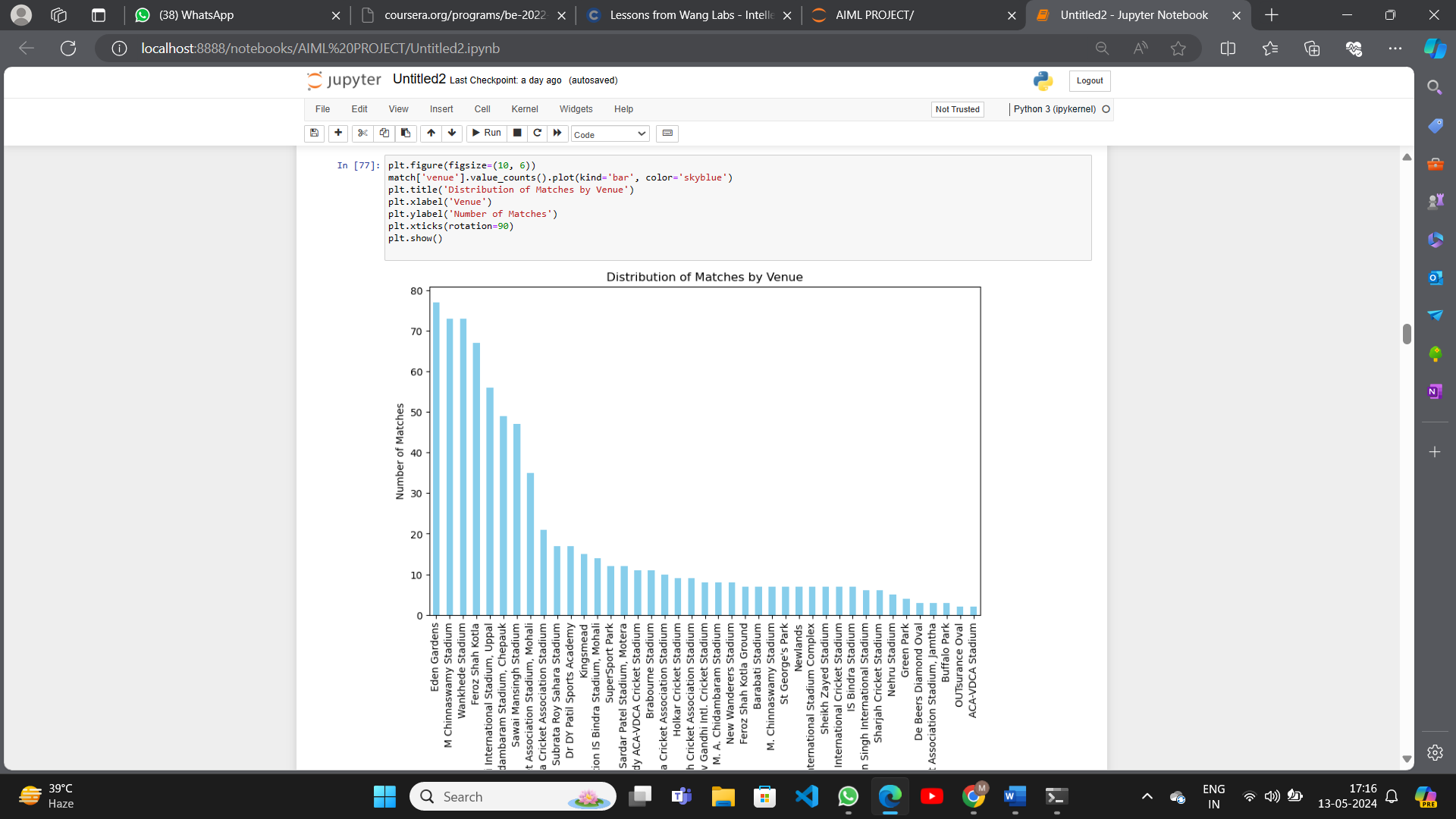
PERCENTAGE OF MATCHES WON BY EACH TEAM



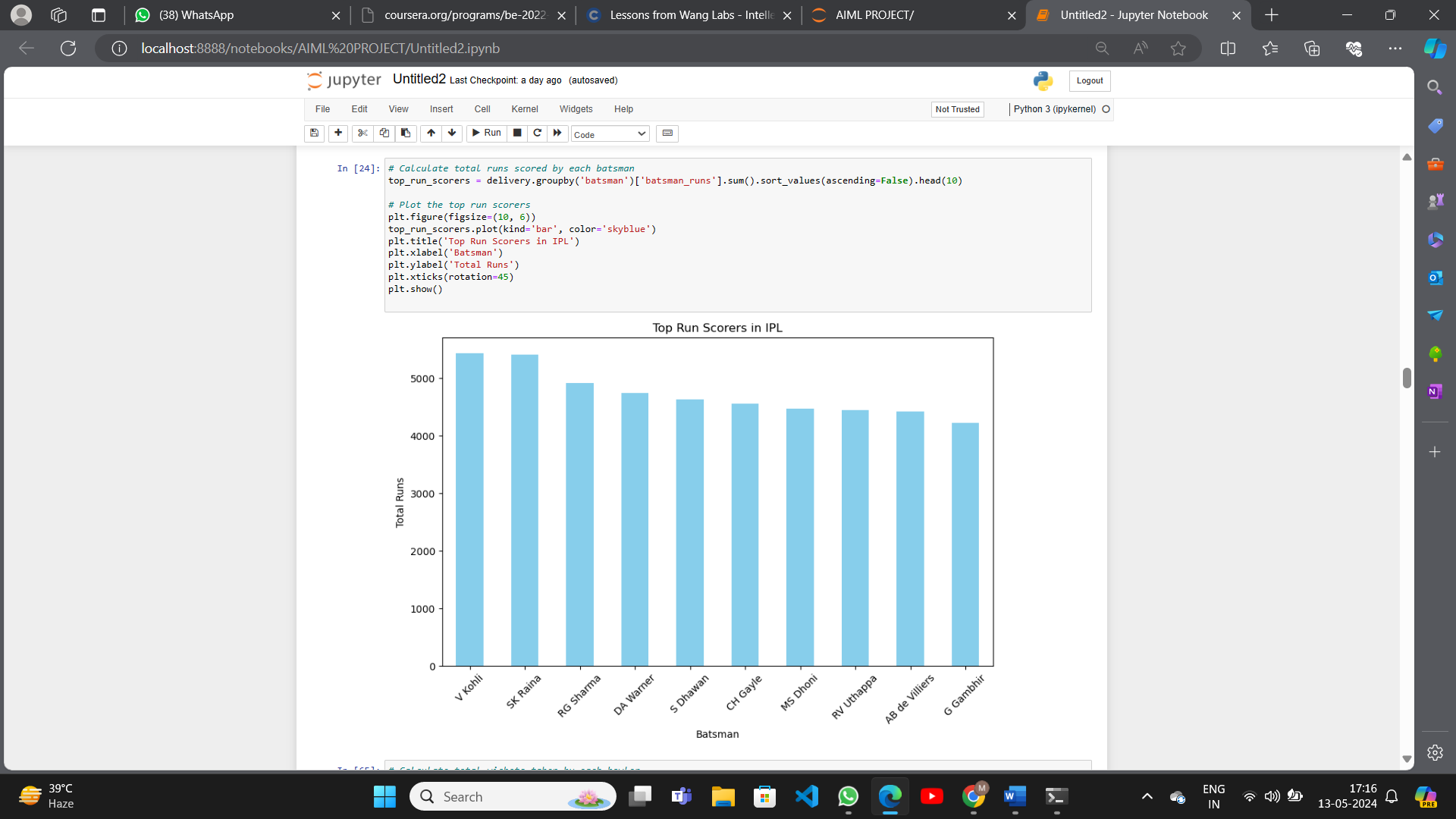
NUMBER OF MATCHES PLAYED IN EACH CITY



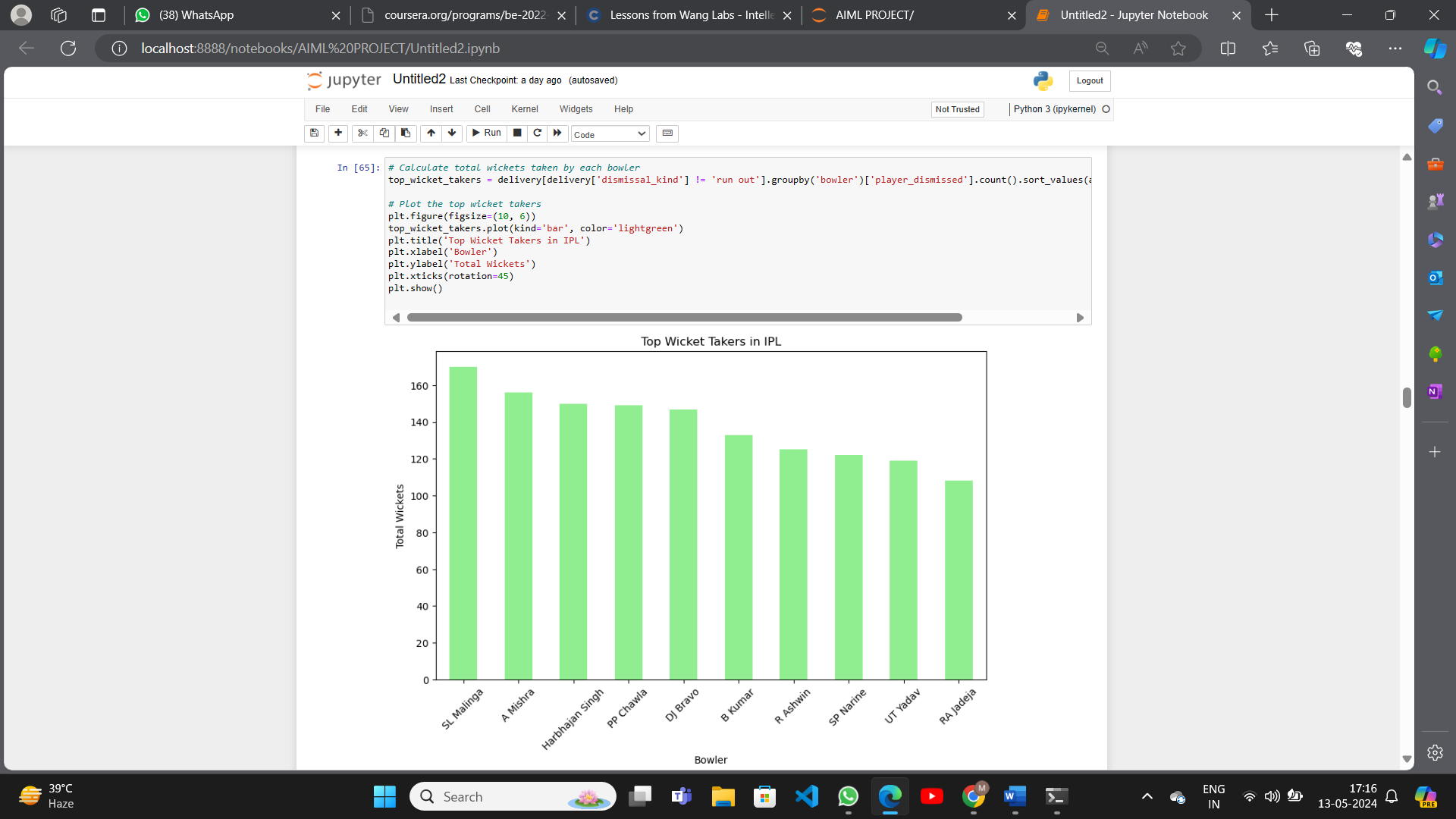
NUMBER OF MATCHES PLAYED IN EACH VENUE

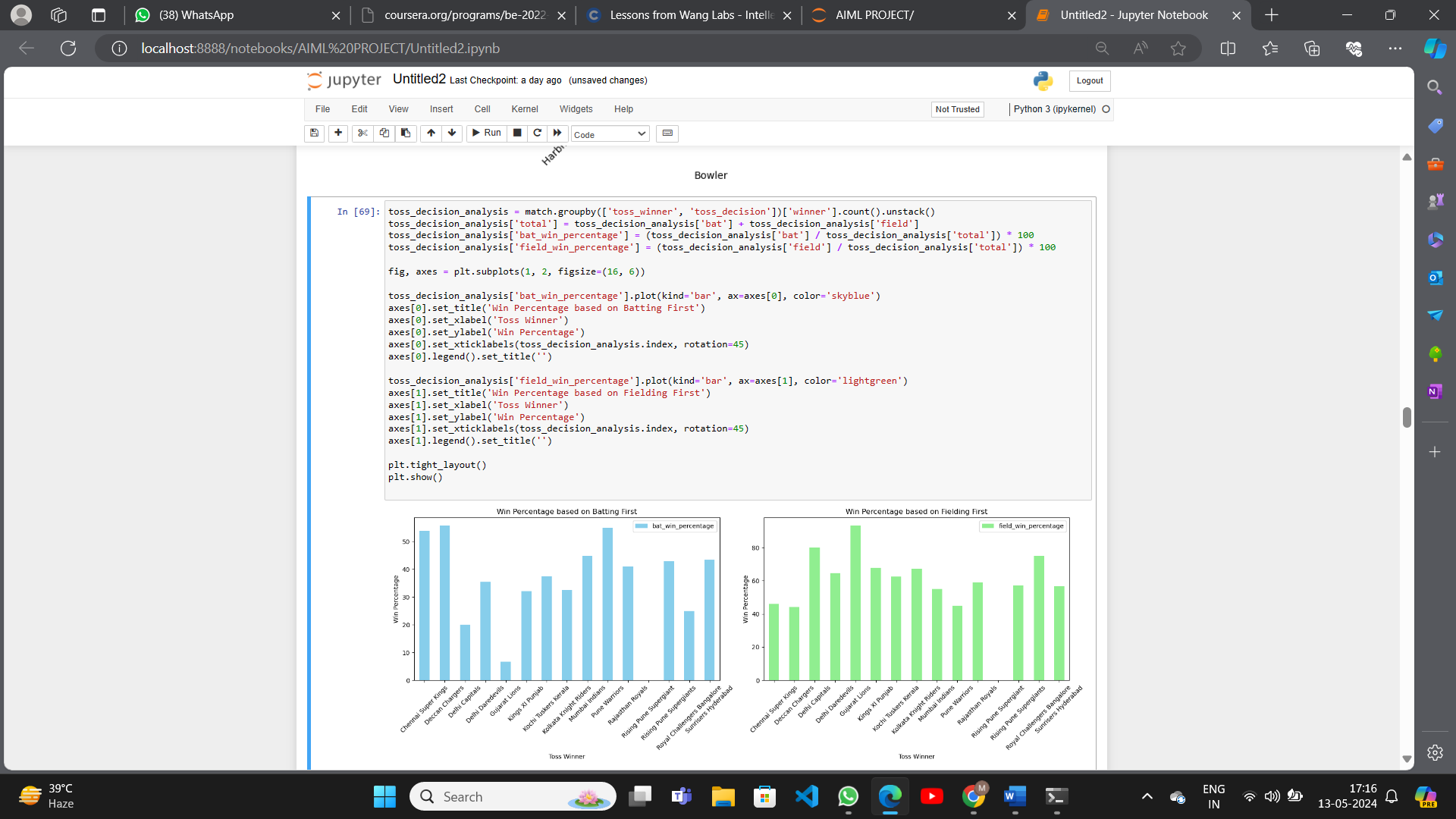


TOTAL RUNS SCORED BY BATSMAN

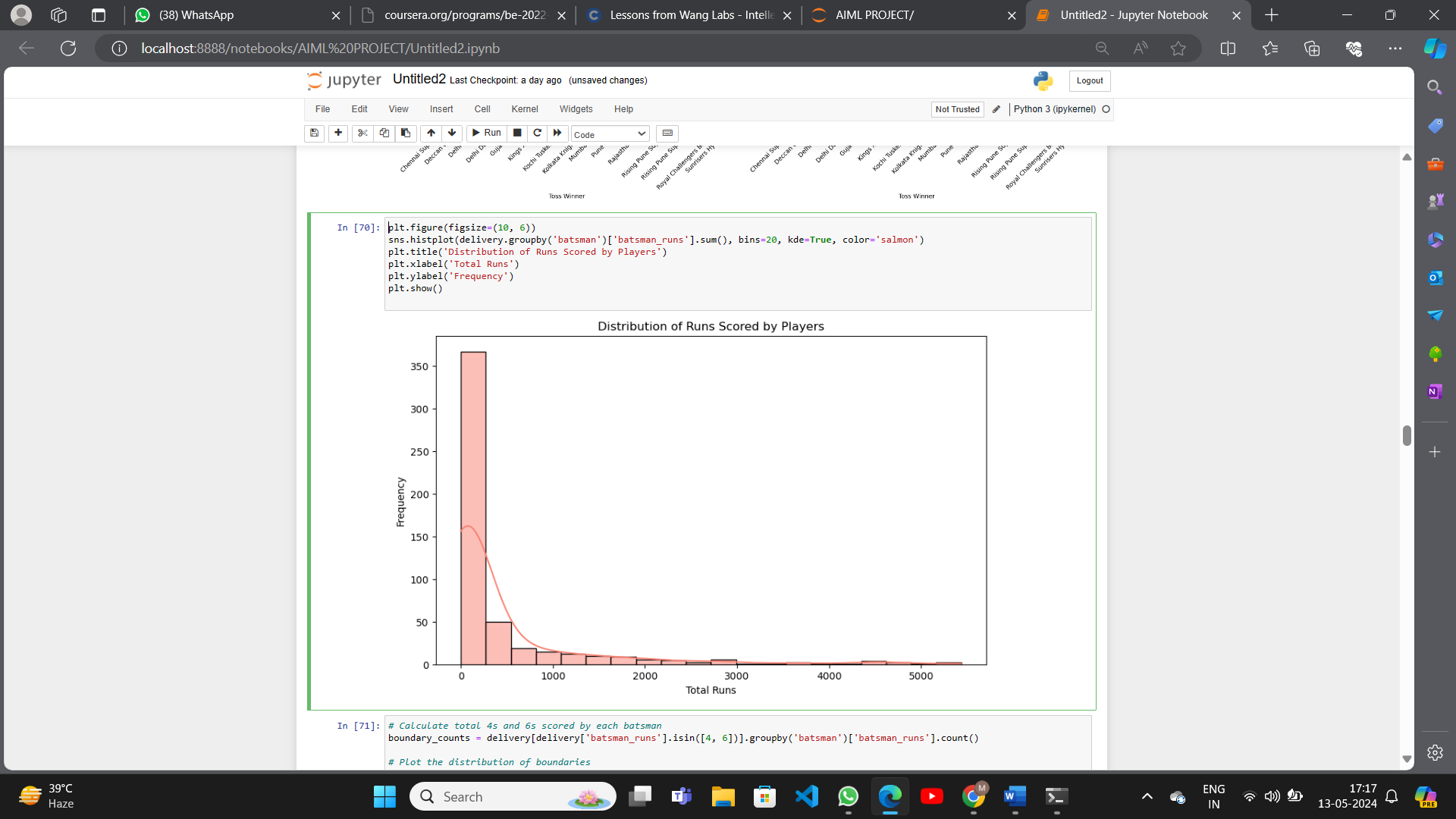


TOTAL WICKETS TAKEN BY A BOWLER

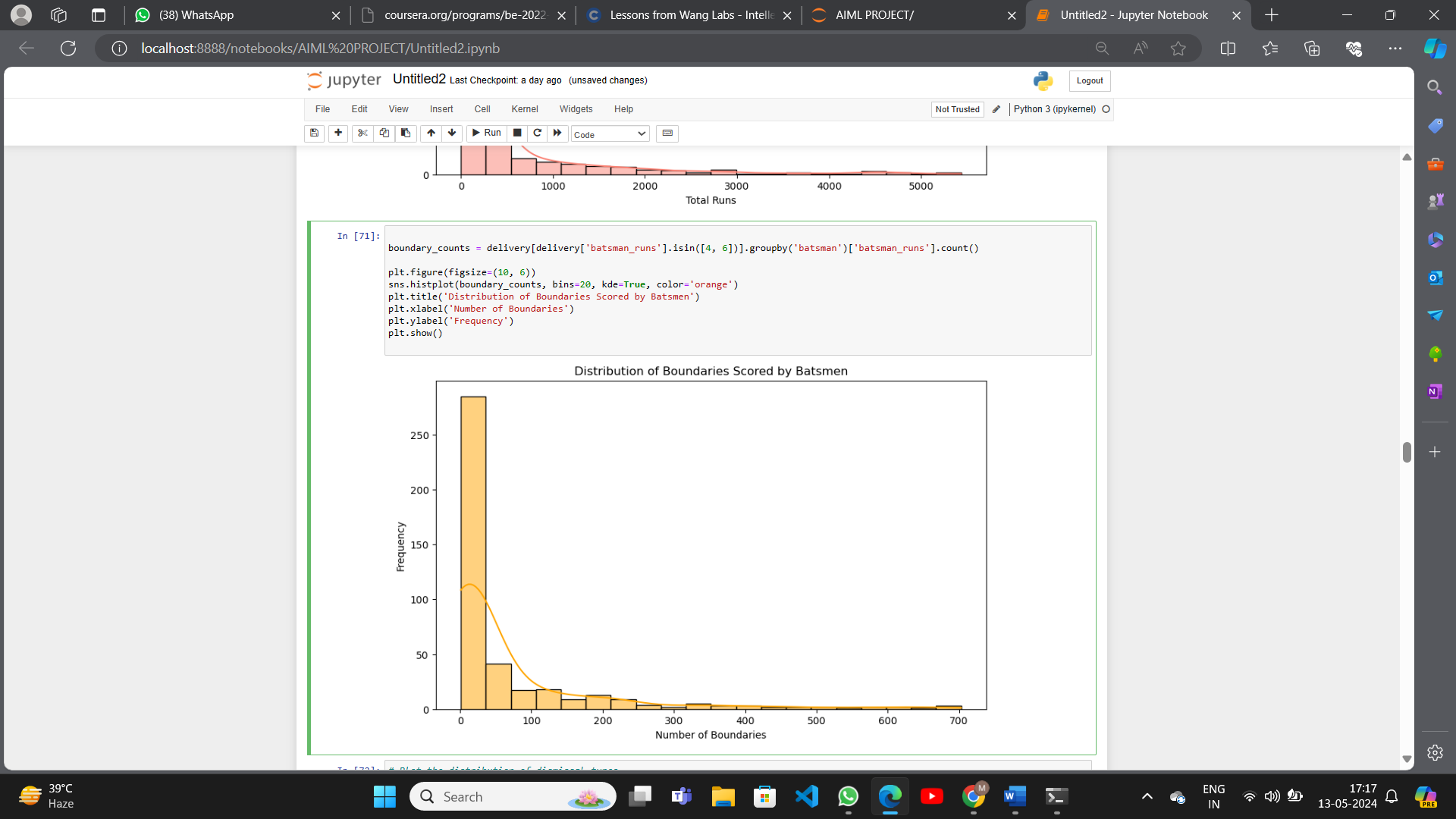


WIN PERCENTAGE OF BATTING AND FIELDING

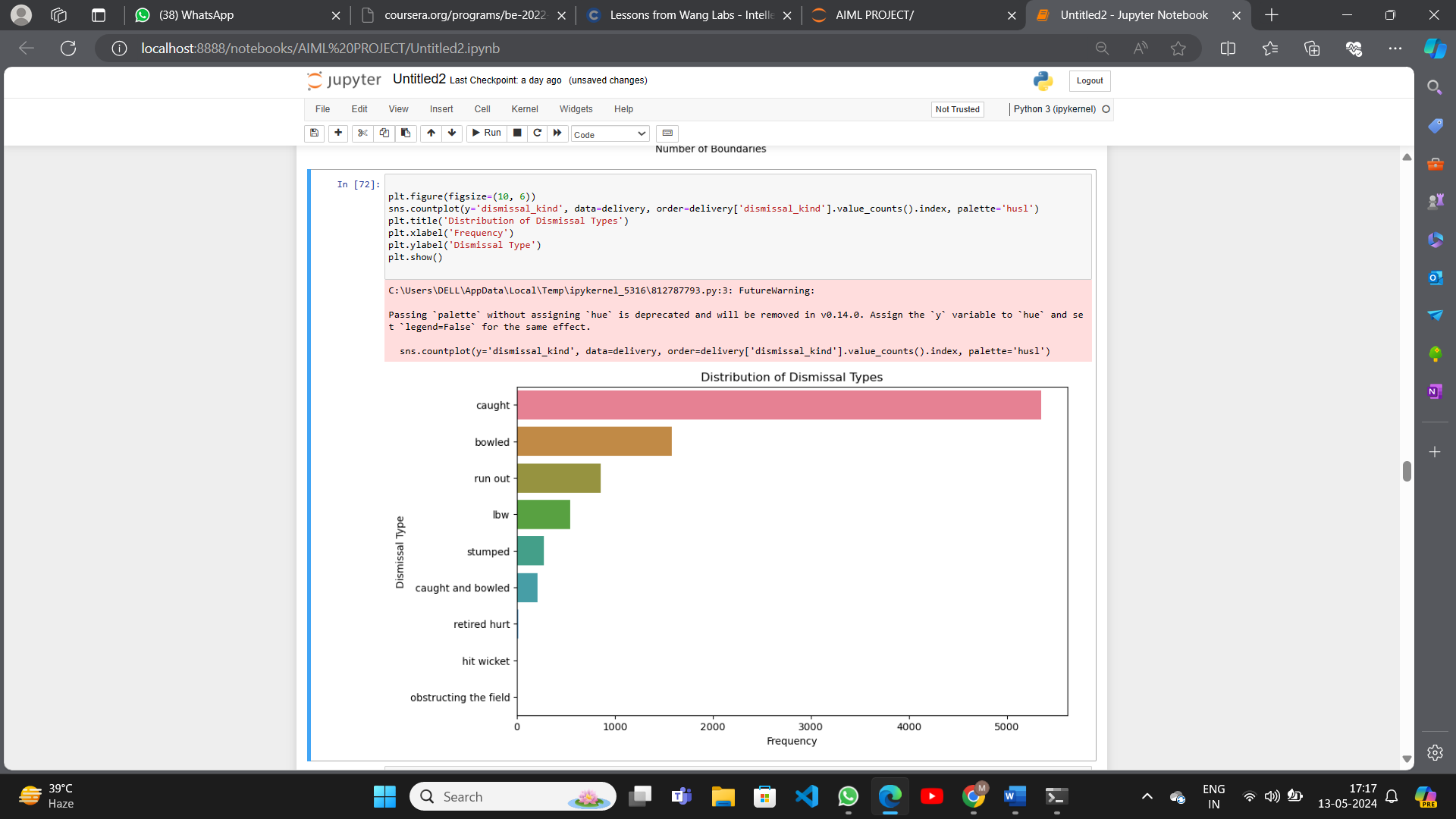
DISTRIBUTION OF RUNS



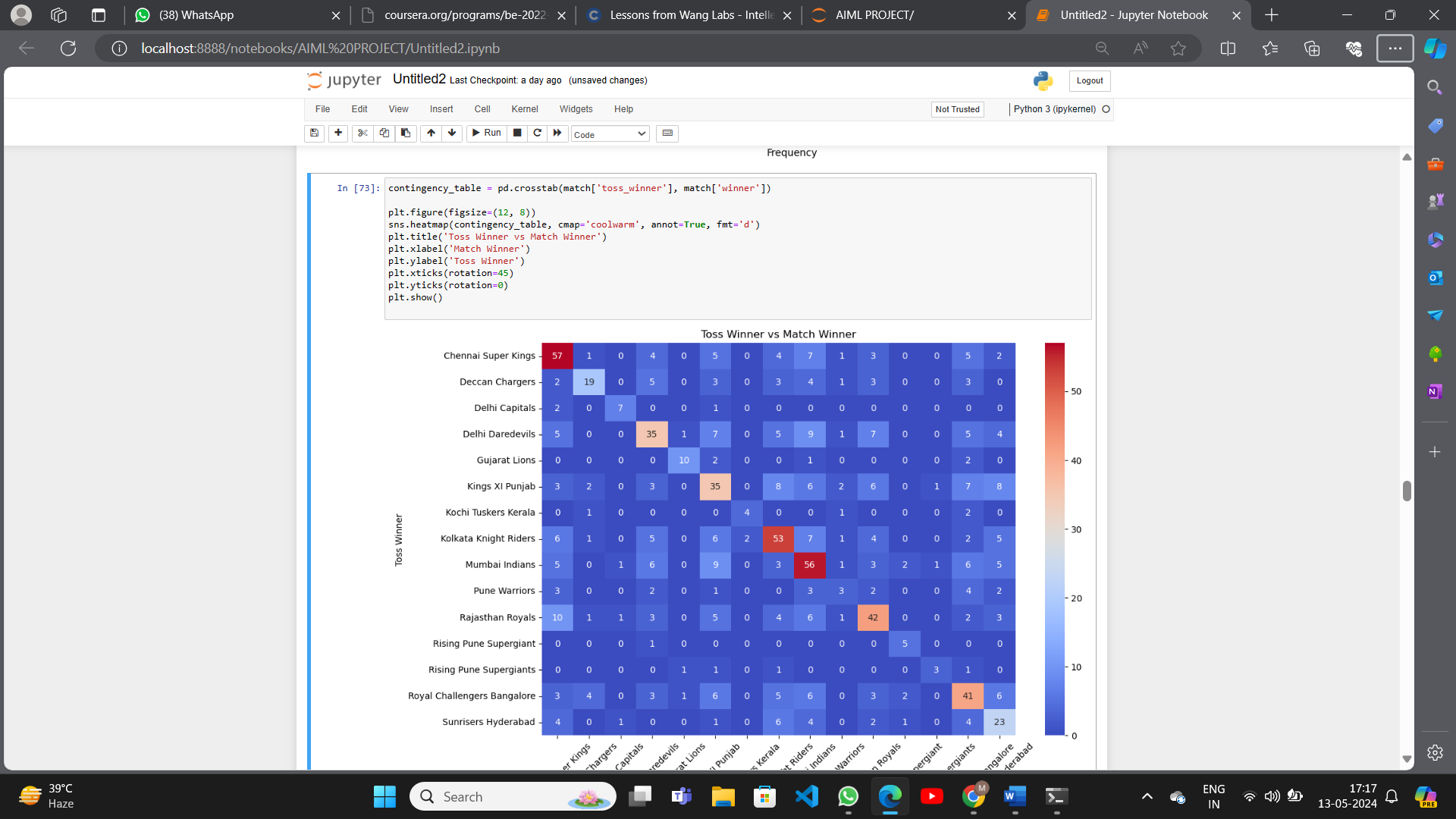
DISTRIBUTION OF BOUNDARIES



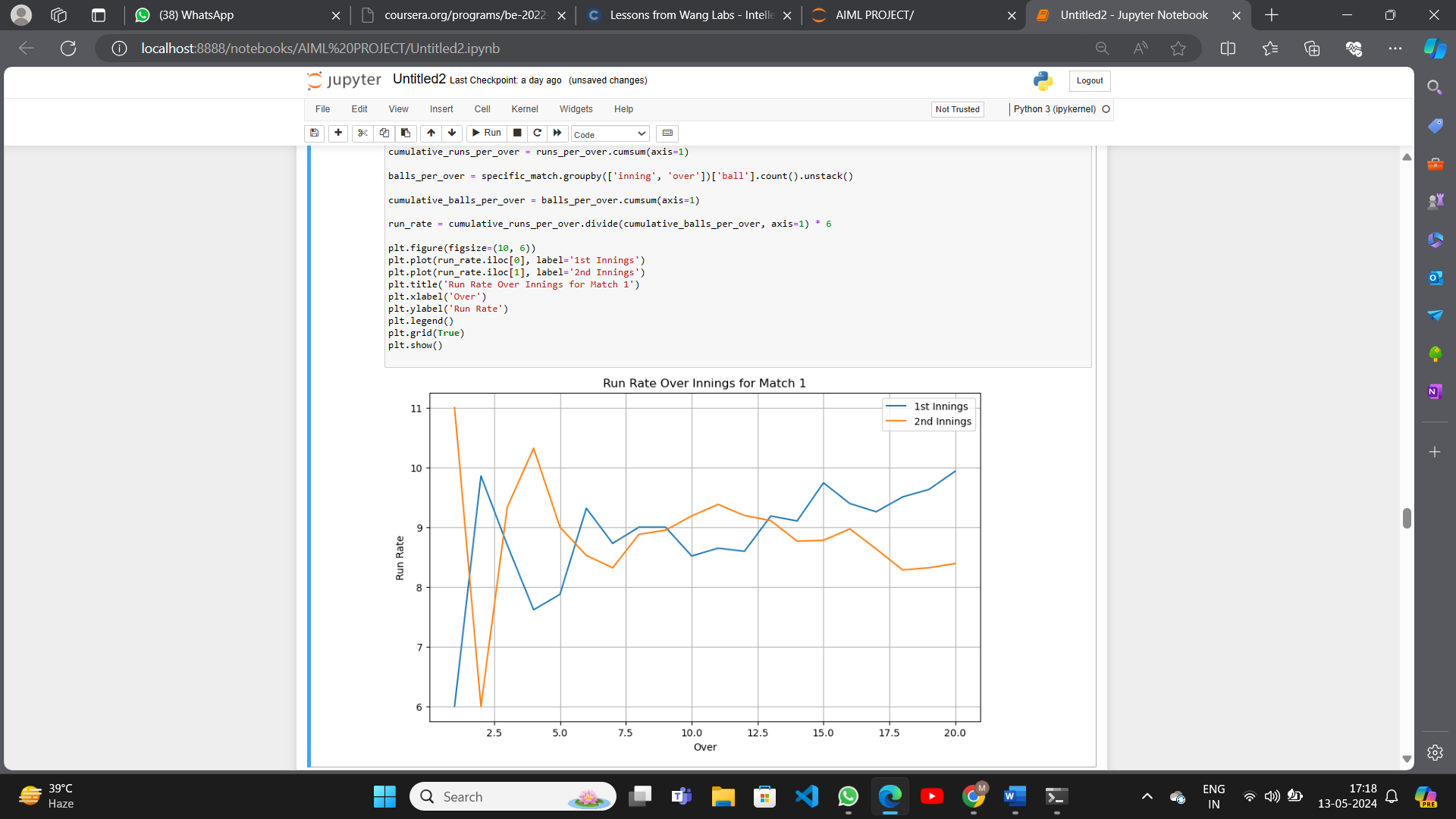
DISTRIBUTION OF DISMISSAL

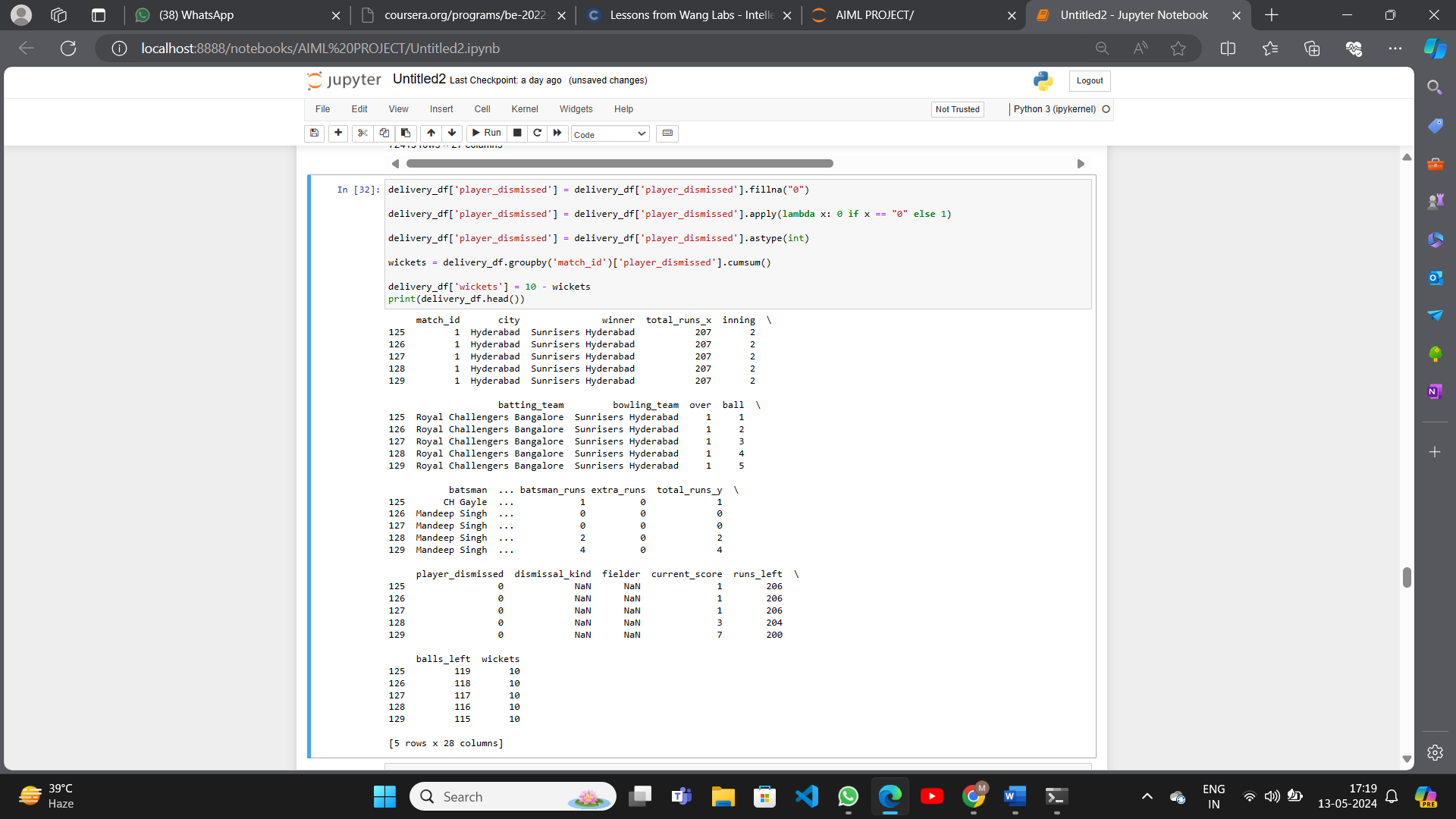


TOSS WINNER VS MATCH WINNER



RUN RATE

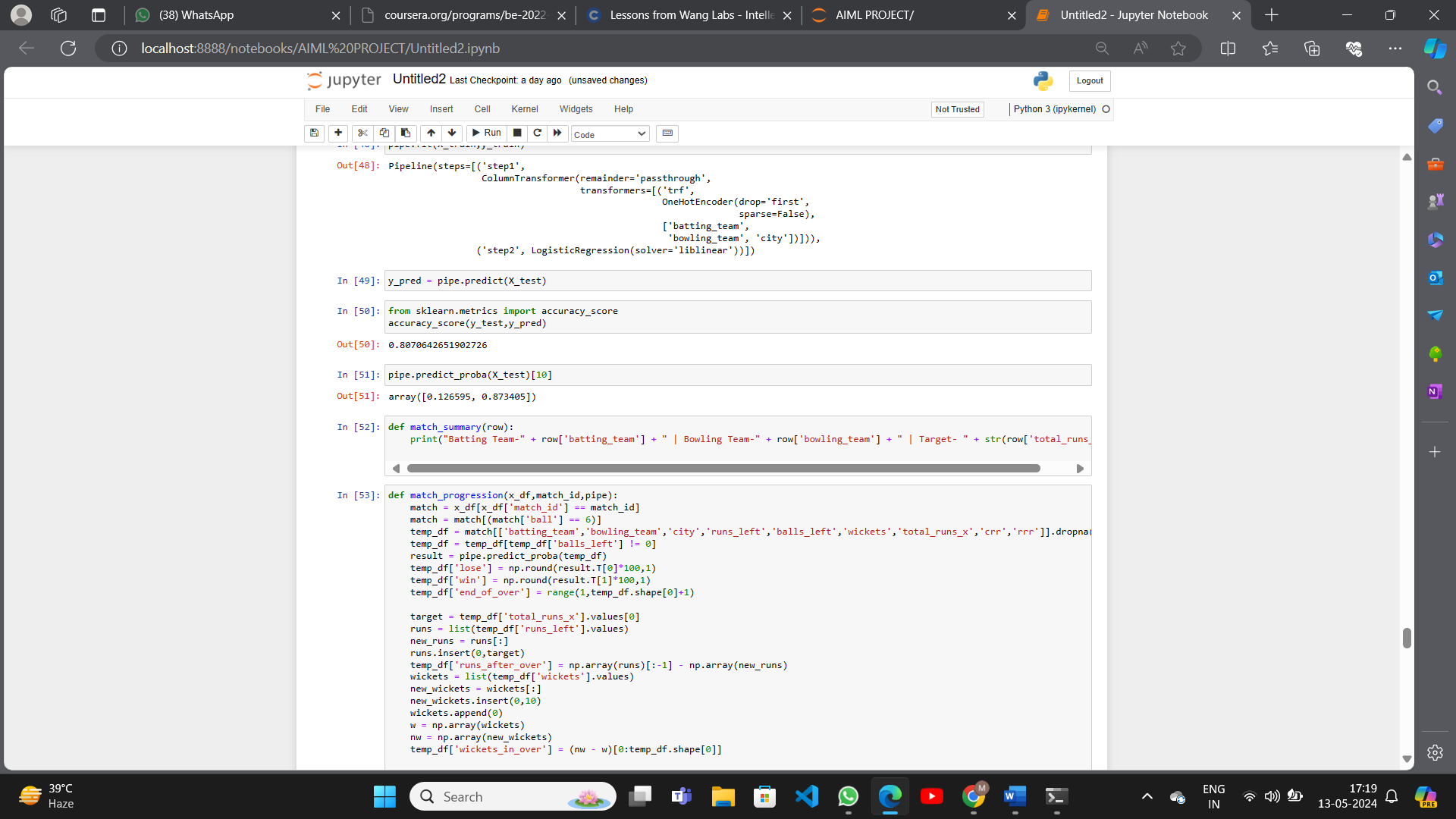


MERGED DATA

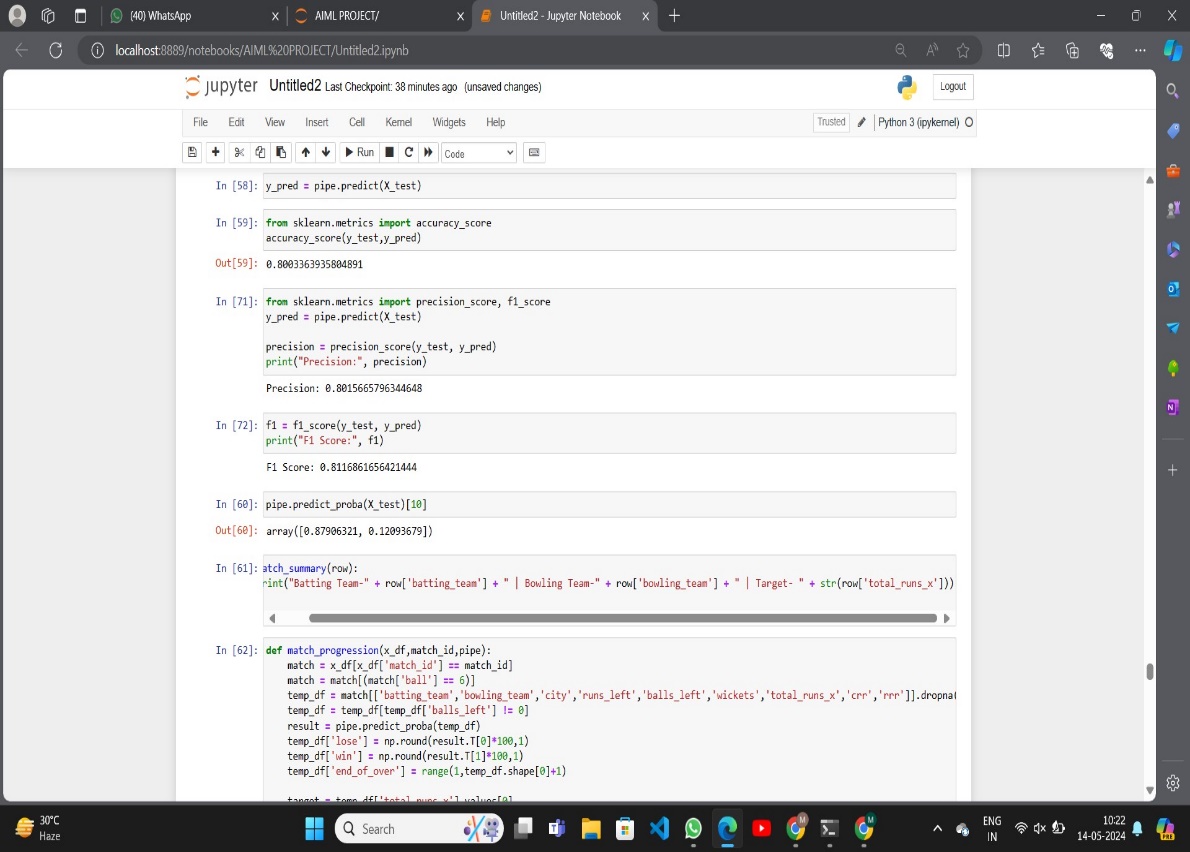
TRAINING THE DATA



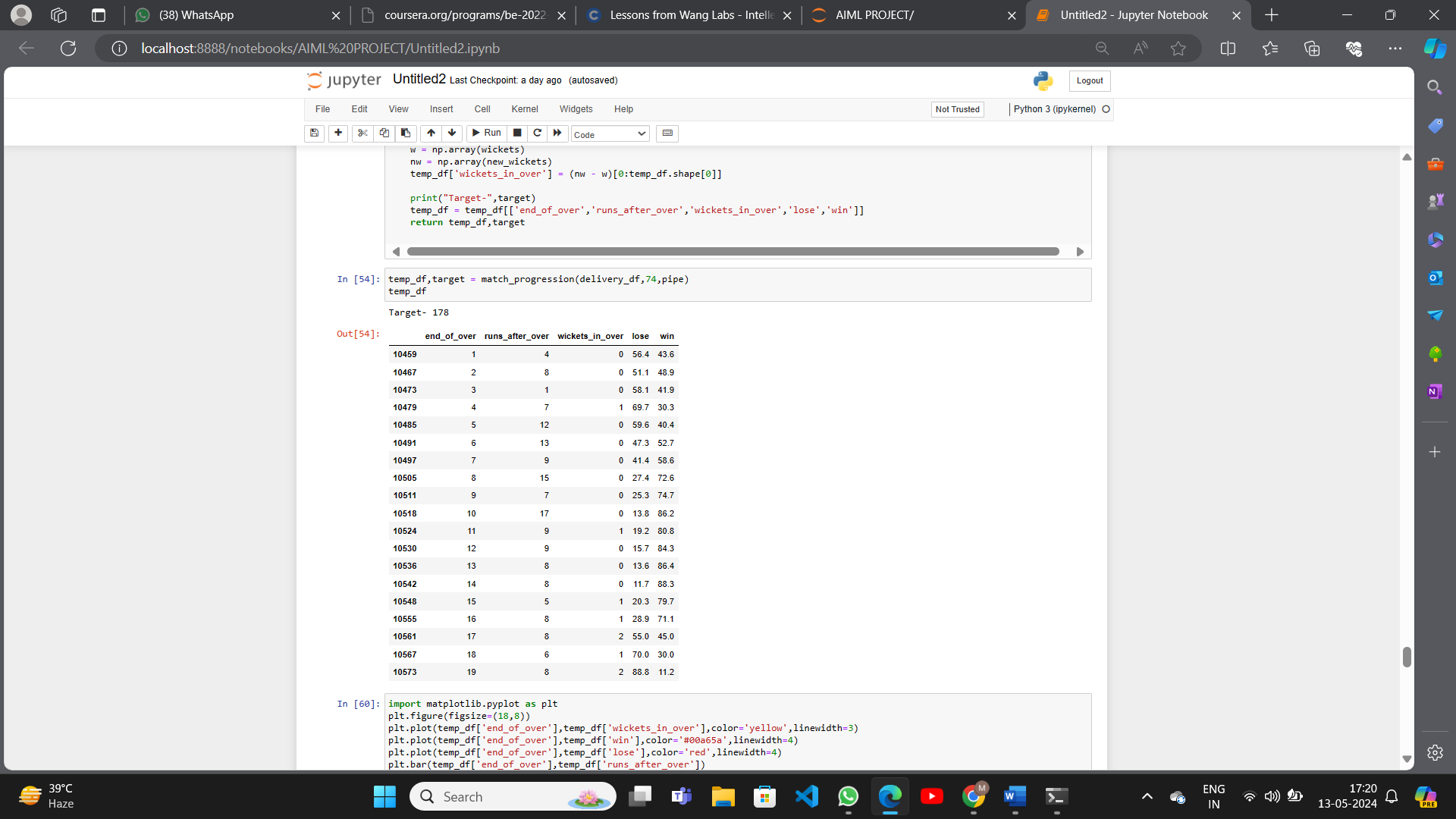
ACCURACY

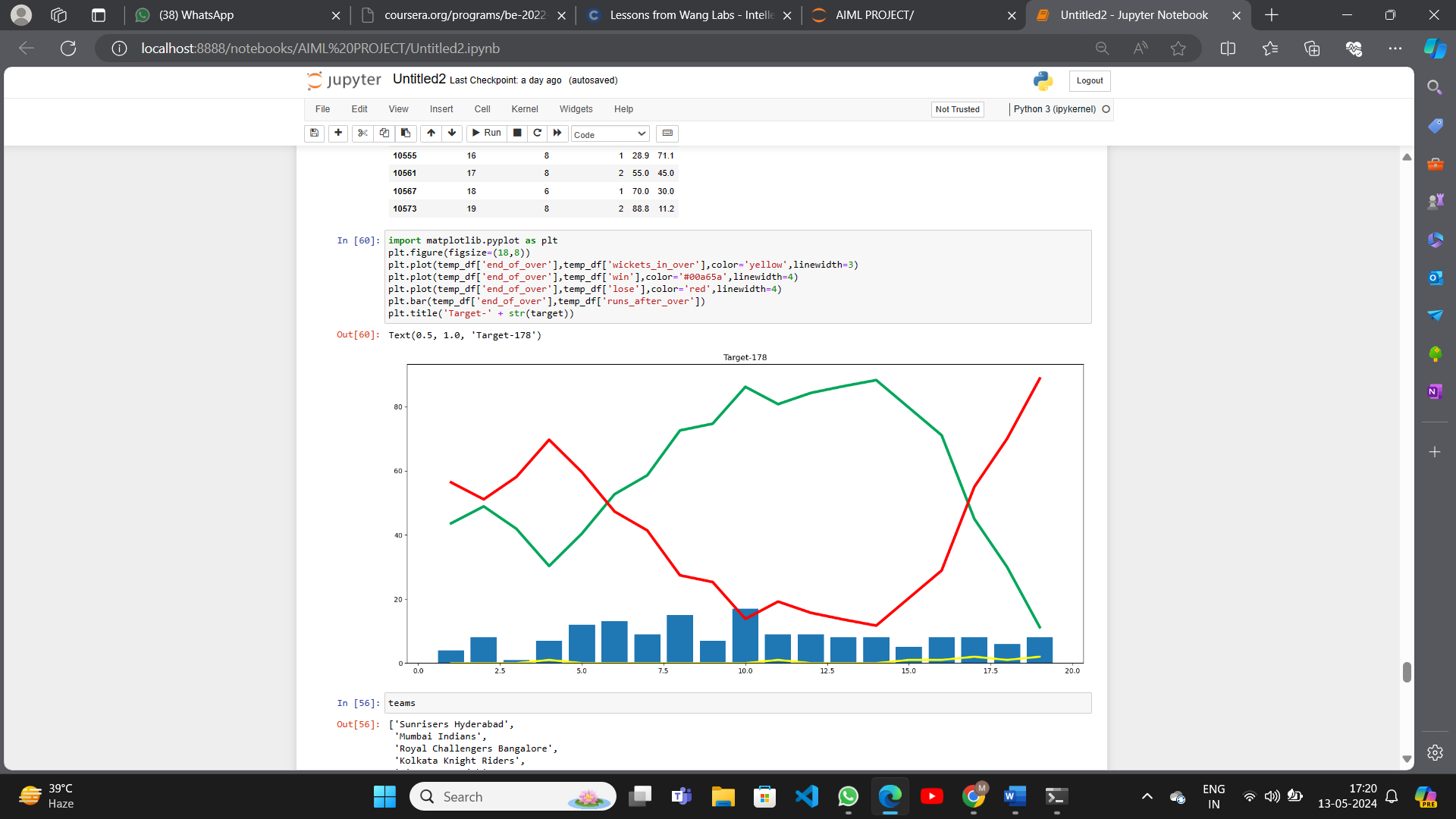


PRECISION AND F1-SCORE



DATA REQUIRED FOR PREDICTING



GRAPH PREDICTING WHETHER TEAM IS LOSING OR WINNING

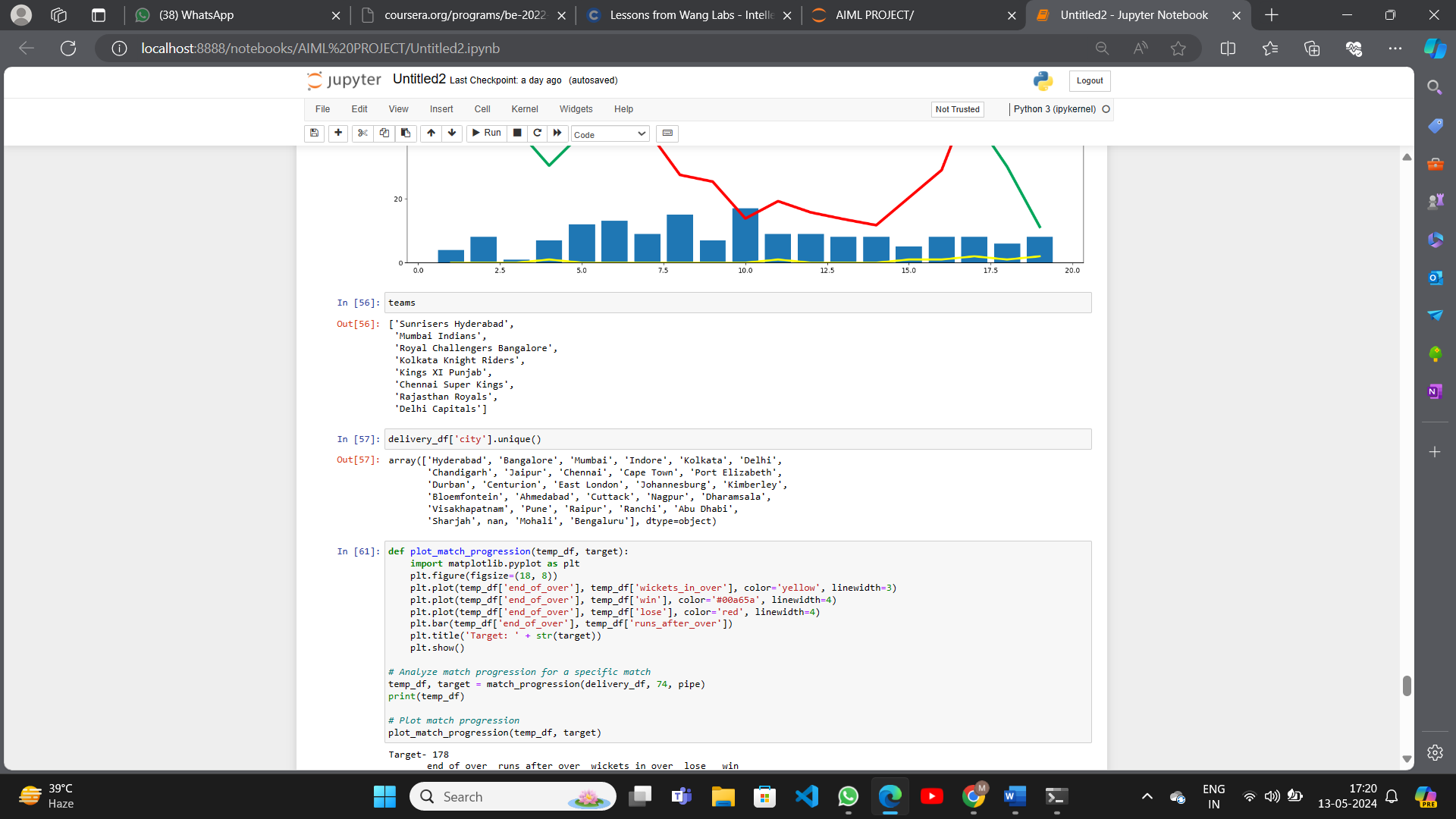
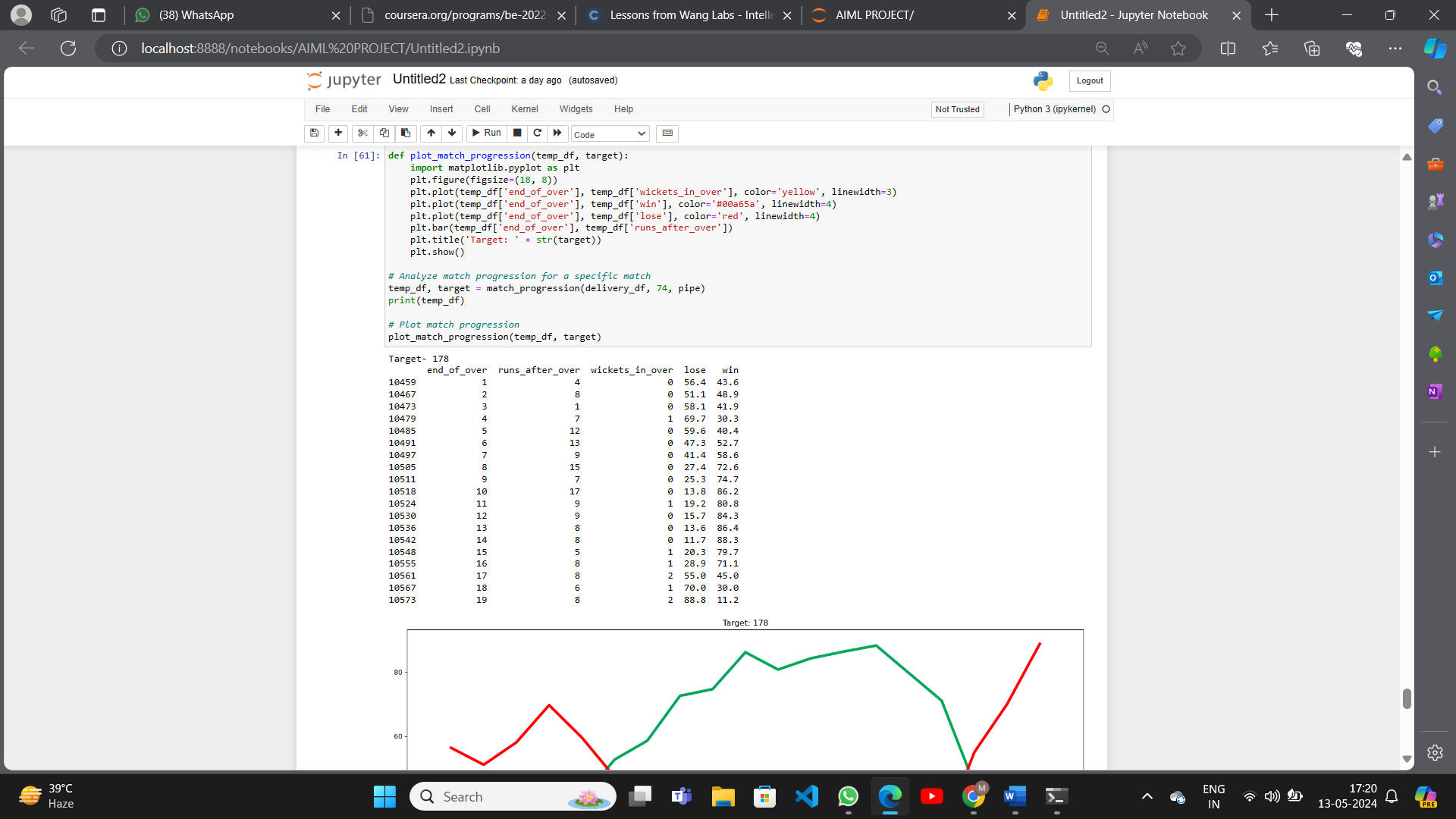
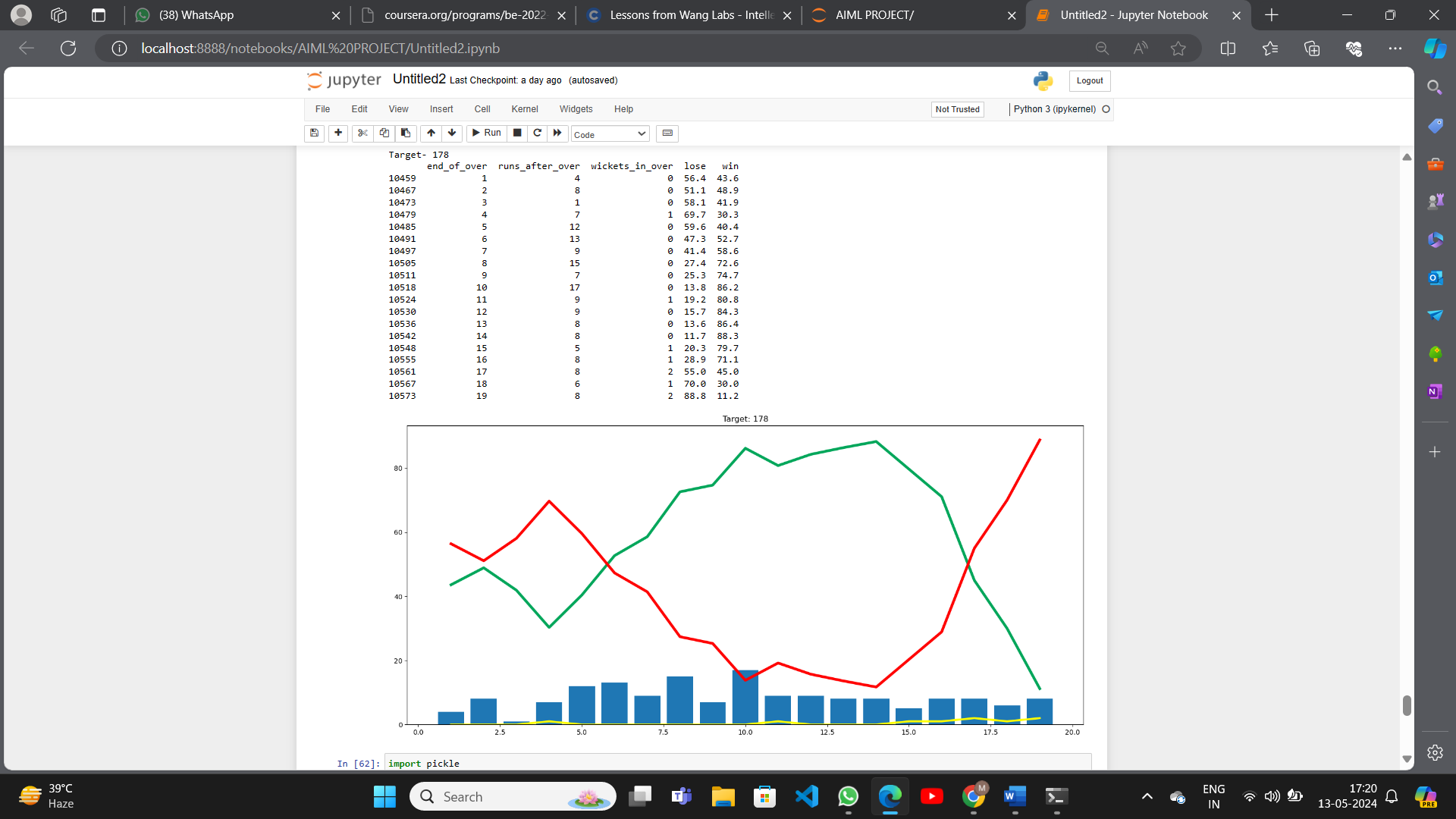
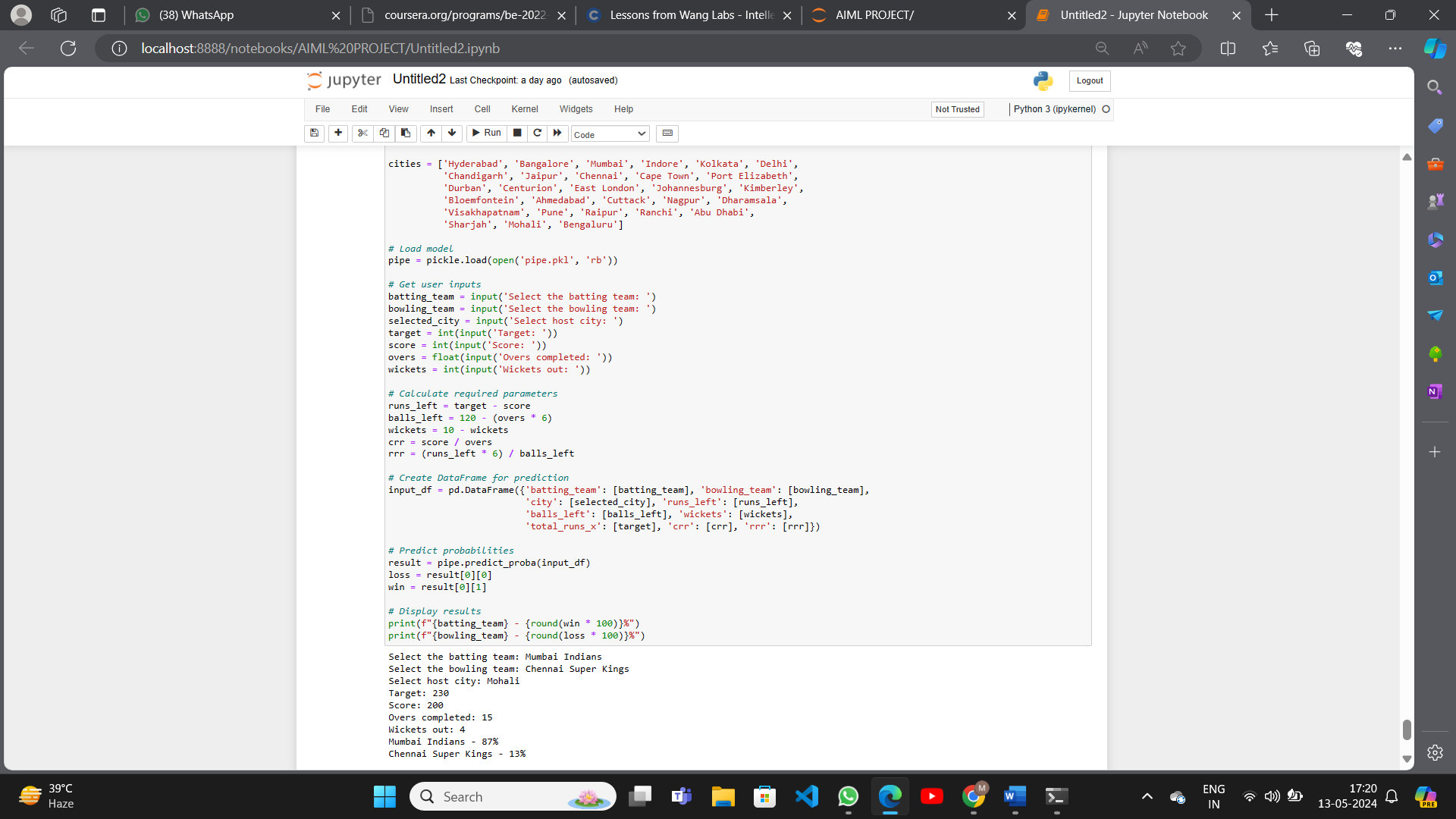
DISPLAYING TEAM NAMES AND CITIES WHERE PLAYED

TABLE DEPICTING MATCH PROGRESS



GRAPH DEPICTING MATCH PROGRESS

PREDICTION MADE

1. **CONCLUSION**

The IPL Sports Predictor Project has successfully demonstrated the potential of data analytics and machine learning in enhancing the understanding and engagement of cricket enthusiasts during IPL matches. Through a systematic approach to data acquisition, preprocessing, model development, and validation, the project has achieved notable results in predicting match outcomes and analyzing player performances.

The project's methodology, which prioritized domain expertise integration, model interpretability, and has ensured the reliability, transparency, and relevance of the predictive models and user interface. By leveraging historical match data, team statistics, player metrics, and other relevant factors, the IPL sports predictor has provided valuable insights and actionable predictions for cricket enthusiasts, fantasy cricket players, and sports analysts.

Moving forward, the IPL Sports Predictor Project holds immense potential for further innovation and impact in the field of cricket analytics. Future enhancements may include refining prediction models, integrating real-time data streams, expanding user engagement features, and exploring new prediction tasks and performance metrics. By continuing to iterate, collaborate, and adapt to evolving user needs and technological advancements, the project aims to remain at the forefront of cricket prediction and analysis, contributing to the excitement and enjoyment of IPL matches for years to come.

1. **REFERENCES**

For Datasets:

<https://www.kaggle.com/datasets/ramjidoolla/ipl-data-set>

For Project:

<https://www.codespeedy.com/sports-predictor-using-python-in-machine-learning/>