**ANALYSE PREMIER LEAGUE PLAYERS**

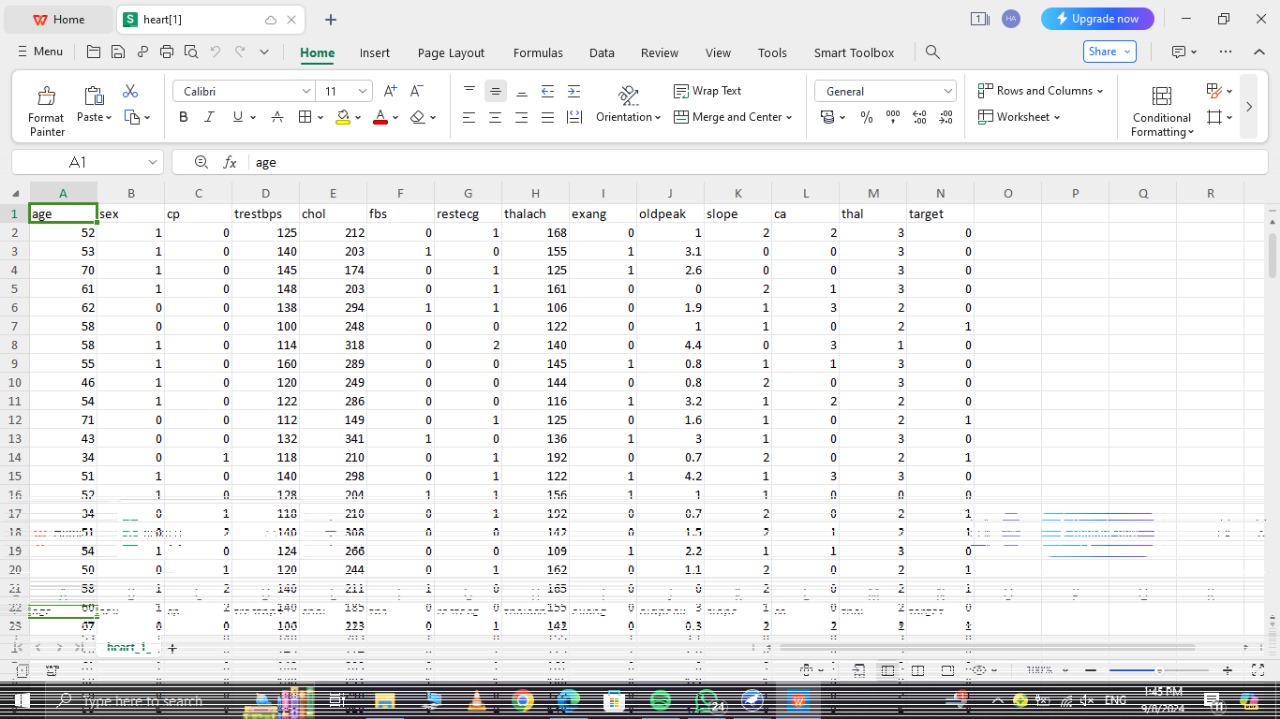
**SETTING THE RESEARCH GOAL:**

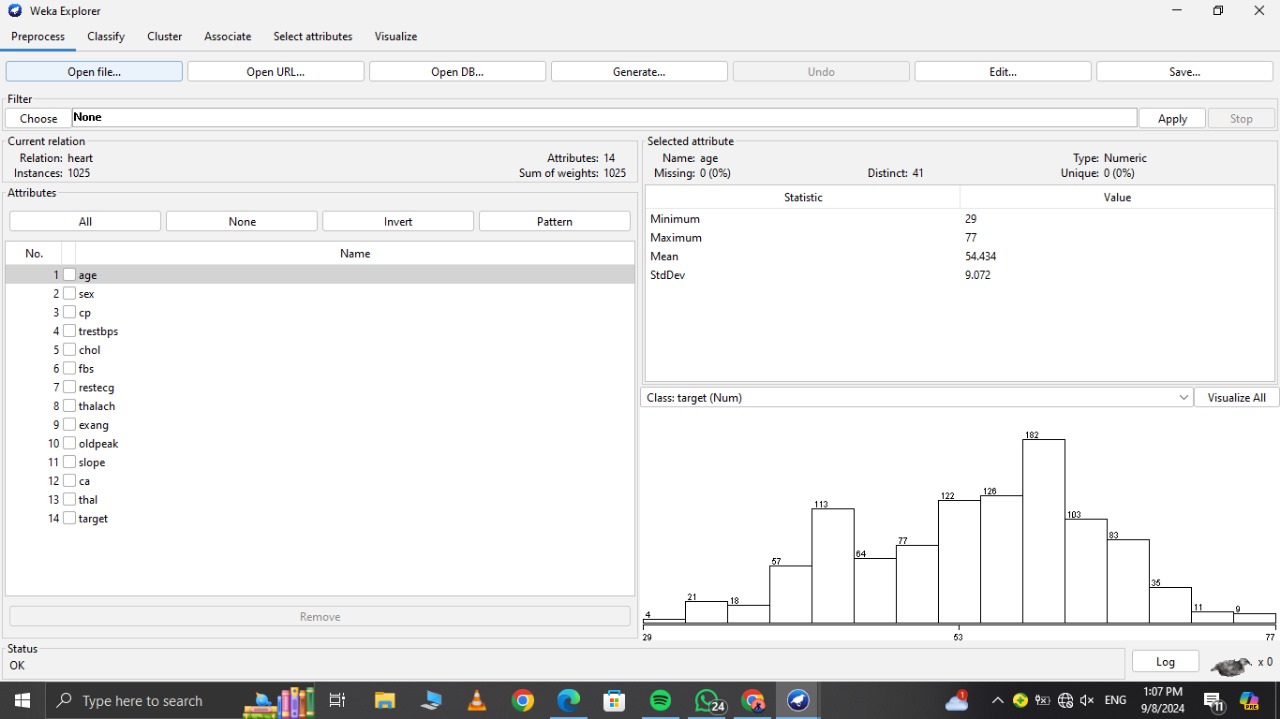
The research goal is to develop and assess a predictive model for heart disease using patient health metrics. Objectives include exploring and preprocessing the data, selecting significant features, training and evaluating various machine learning models, and interpreting results. The focus will be on identifying key predictors and achieving high model accuracy. The ultimate aim is to provide actionable insights for improving heart disease prediction and prevention.

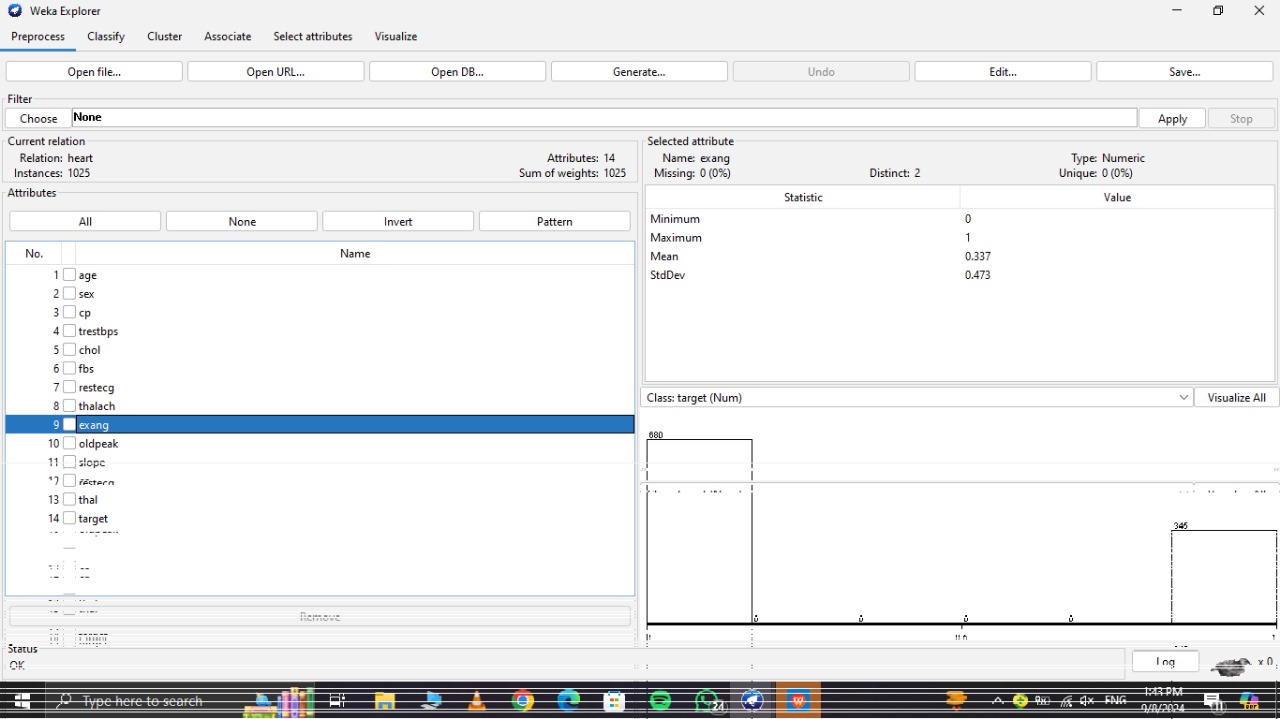
**RETRIEVING THE DATA:**

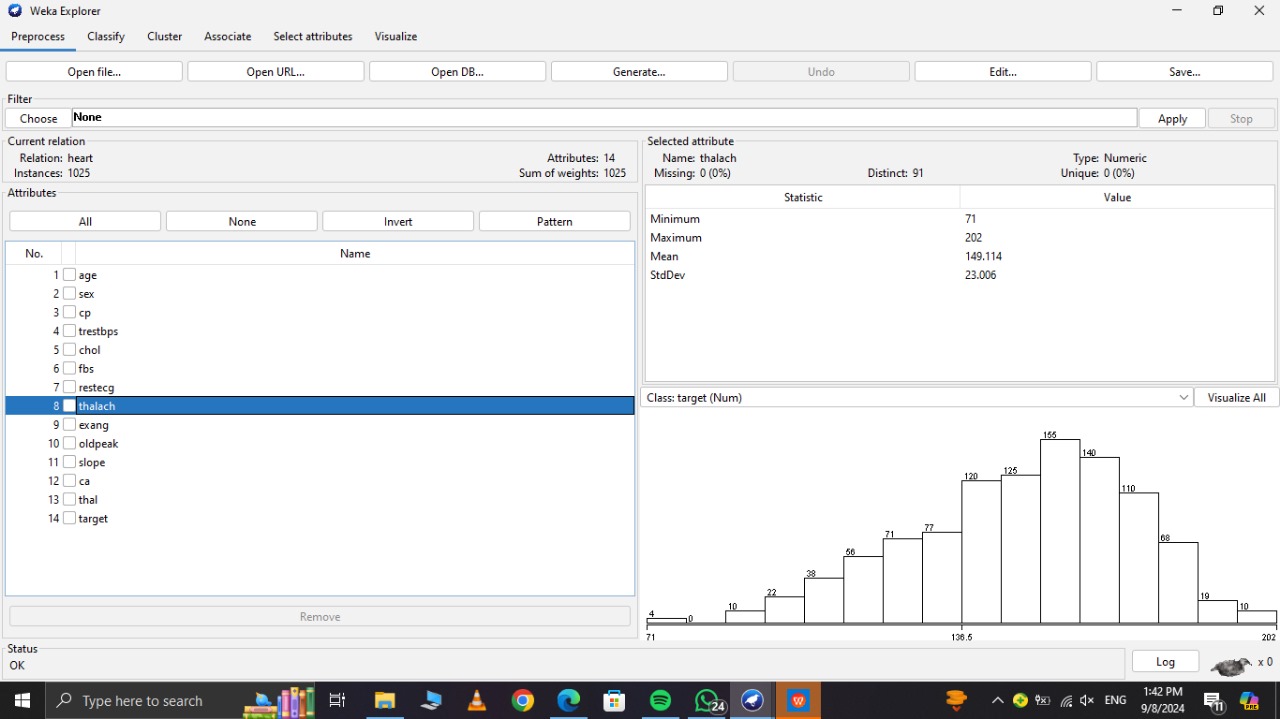
To retrieve data for heart disease analysis, start by sourcing a dataset from repositories like UCI Machine Learning Repository or Kaggle. Download the dataset in CSV format. Load the data into your working environment using tools such as Python with Pandas. Verify the data for any missing values or inconsistencies. Clean and preprocess the data by handling missing values, encoding categorical variables, and normalizing features. Once prepared, the dataset will be ready for analysis and model development.

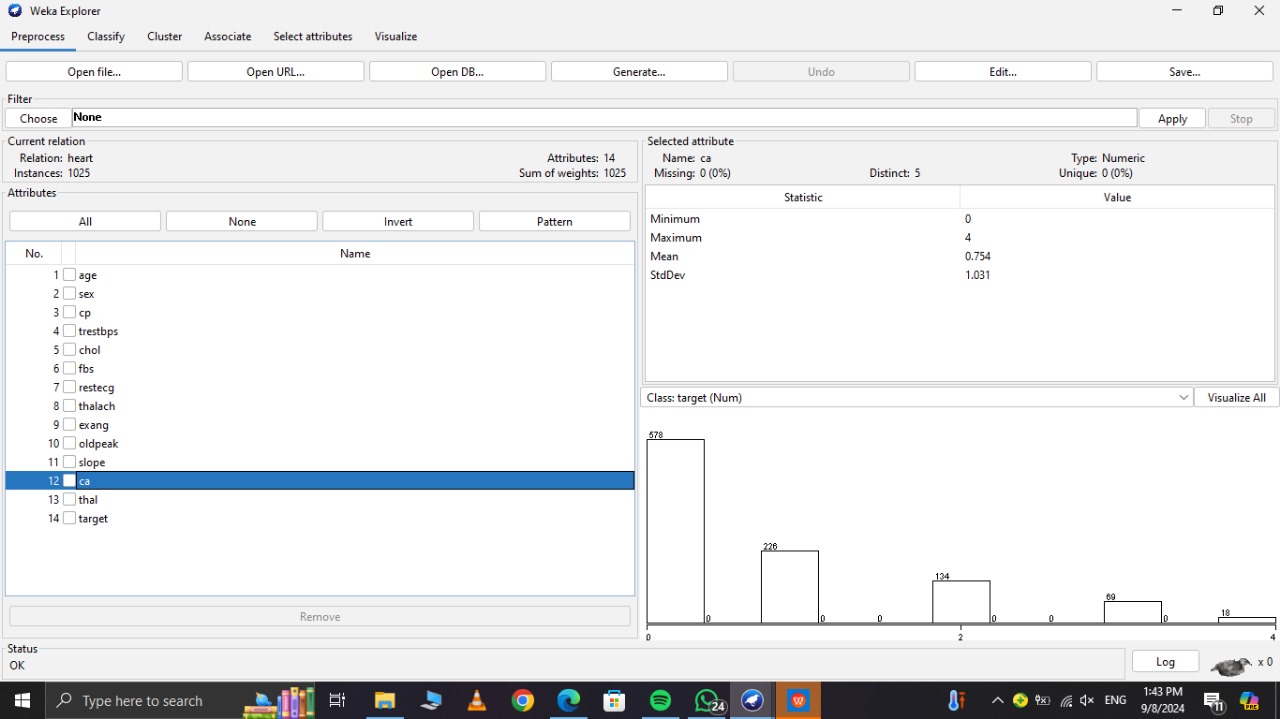
**DATA PREPARATION**:

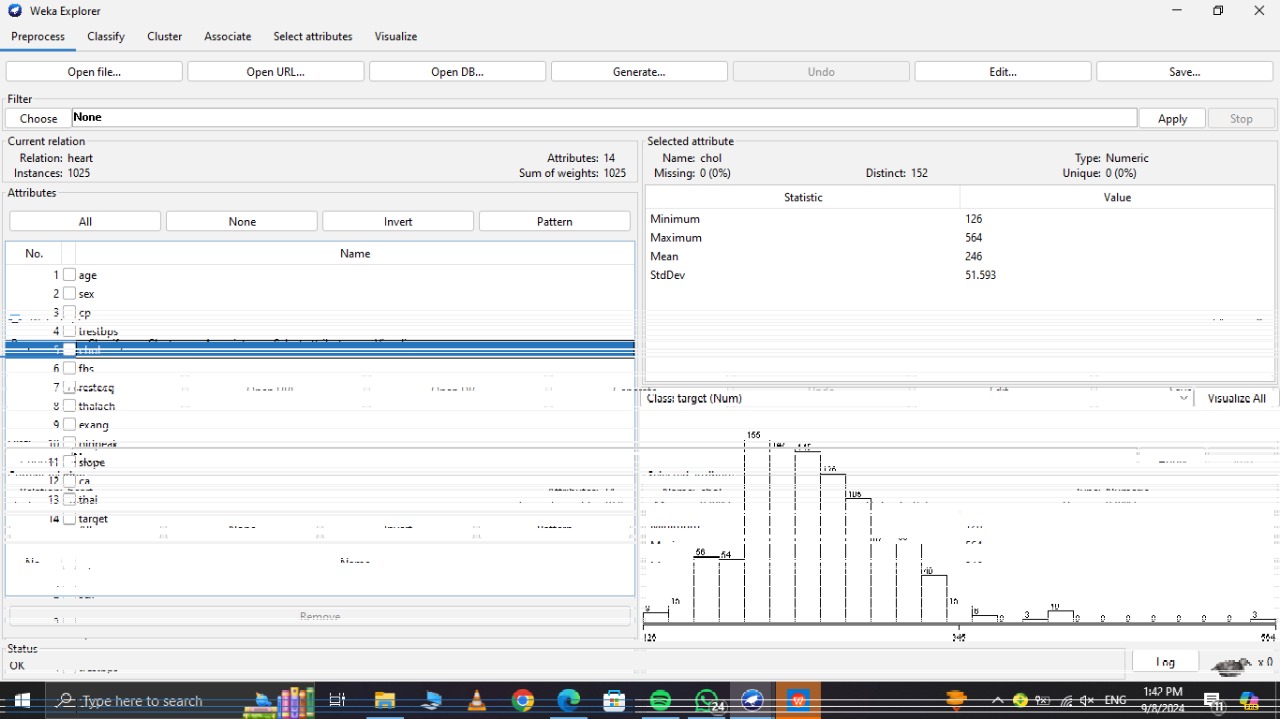
DATA CLEANING:





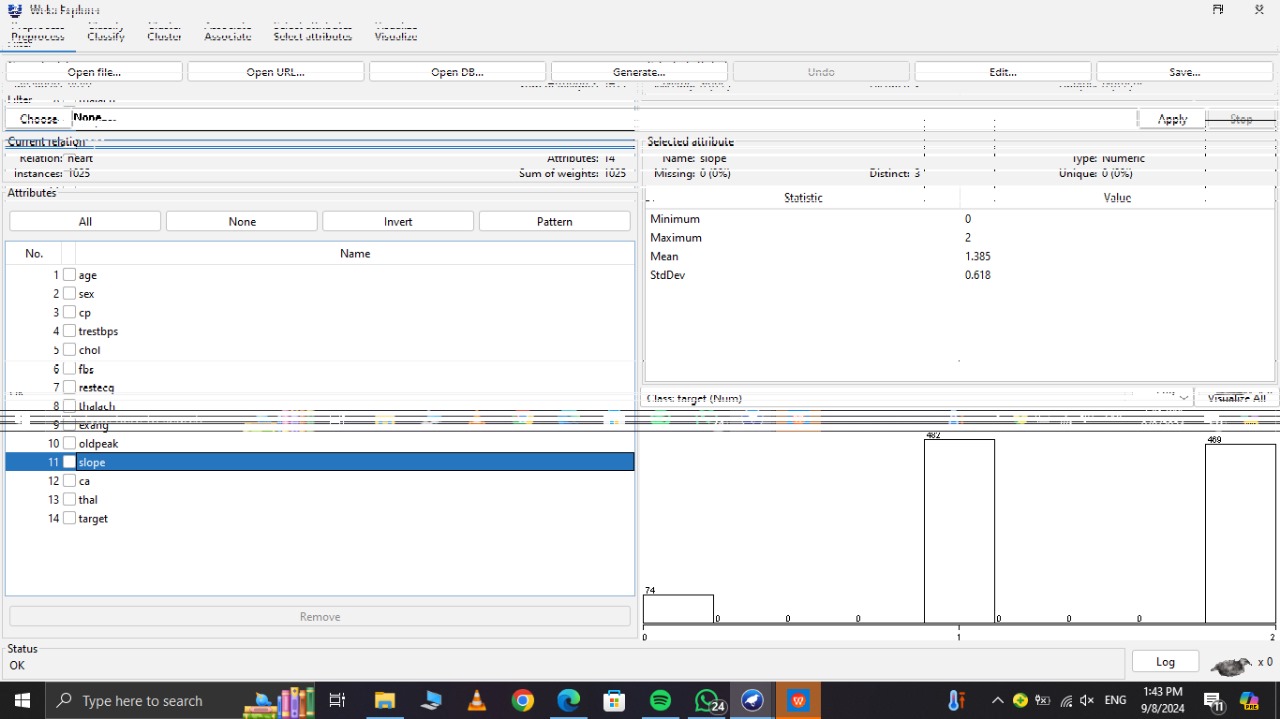






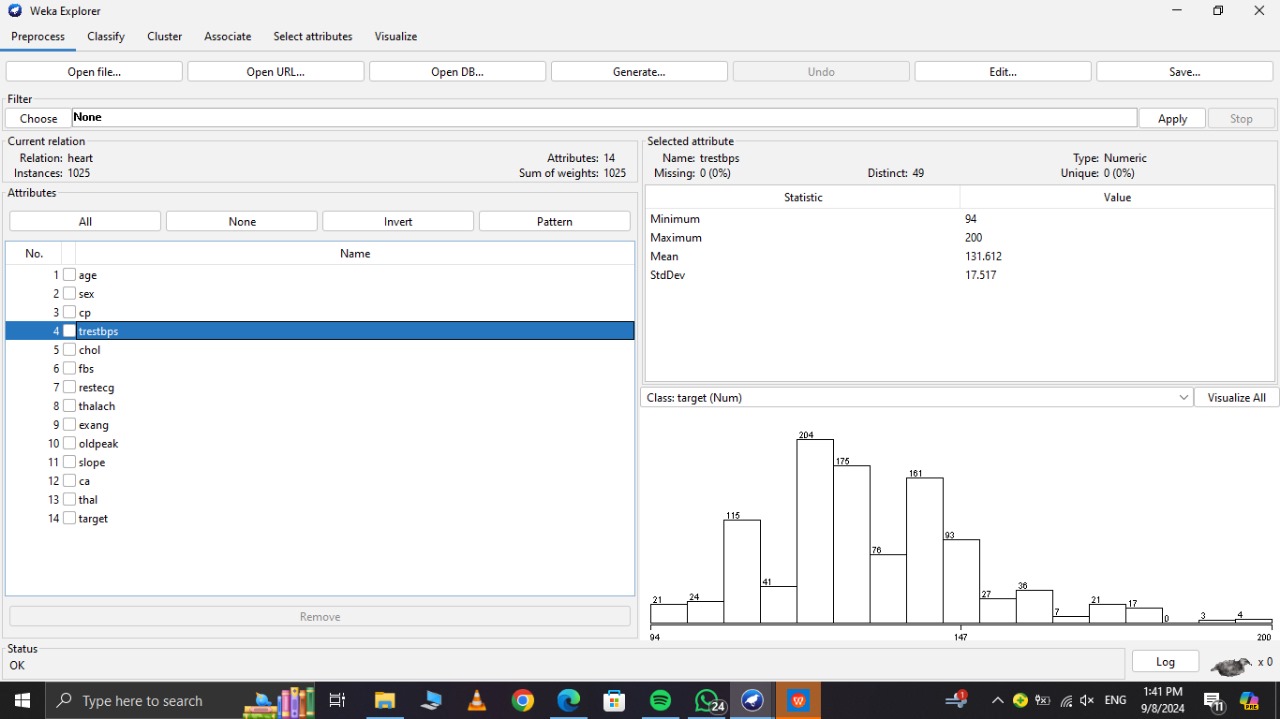
**DATA EXPLORATION**:

Data exploration for heart disease analysis involves examining the dataset’s structure, including column names and data types. It includes generating summary statistics to understand feature distributions and detecting anomalies. Checking for missing values is crucial for data integrity. Exploring relationships between features and the target variable helps identify patterns. Analyzing the balance of classes in the target variable reveals distribution. Visualization techniques, like plots and pair plots, aid in understanding feature relationships and overall data trends.



**EXPLORATORY DATA ANALYSIS**:

To explore data analysis for heart disease, start by loading and previewing the dataset to understand its structure. Generate summary statistics to identify distributions and outliers. Check for missing values and perform data cleaning as needed. Analyze correlations between features and the target variable to identify key predictors. Examine the distribution of the target variable to understand class balance. Use visualizations to explore relationships and feature distributions. Evaluate feature importance to determine the most significant predictors for heart disease.



**Presentation**:

Start with an introduction to the research goal and dataset. Display key statistics and visualizations to highlight data characteristics and relationships. Summarize model development, performance metrics, and key findings. Conclude with significant predictors and implications for players statistics

**Automation:**

Develop scripts to automate data processing, exploration, and analysis. Create a pipeline for pre processing, model training, and evaluation. Use libraries to generate and save visualizations automatically. Automate report generation for consistent and efficient presentations.

