Proposal: Time Series Analysis of Sports Data

1. Introduction:

For this project, I will research NBA data in order to understand player performance trends, impact of injuries and fatigue, team dynamics, game strategies, career trajectories, and player development. In this project, I will conduct time series analysis on sports data to uncover insights and make predictions about various aspects in the NBA.

2. Data Selection:

Dataset: I will analyze the NBA player stats dataset, which contains detailed statistics for NBA players over multiple seasons. The dataset includes information such as points scored, rebounds, assists, steals, blocks, and turnovers. This dataset provides a collection of player-level statistics over time, making it suitable for conducting time series analysis on player performance.

Location: The dataset can be accessed on Kaggle's website with this link: https://www.kaggle.com/datasets/vivovinco/2023-2024-nba-player-stats/data

3. Model Selection and Equation:

Model: I will employ a simple linear regression model to analyze the trends and patterns in player performance metrics over time. In the context of sports data analysis, linear regression can help to identify trends, predict future performance of a player, and assess the impact of various factors on player statistics over time. The equation for the linear regression model is given by:

$y=mx+c$

“y” represents the dependent variable (for example: points scored, assists),

“x” represents the independent variable (time),

“m” represents the slope of the regression line, and

“c” represents the intercept.

4. Test Data Generation:

I will generate synthetic test data to validate the performance of our linear regression model. This test data will include random fluctuations in player statistics such as points scored, rebounds, and assists, while still maintaining realistic patterns observed in the actual dataset. In addition, I will utilize the NBA Player Stats dataset mentioned earlier as our primary dataset for analysis. This dataset will serve as the basis for training and testing our linear regression model.

5. Data Filtering and Preprocessing:

A quality filter will be applied to the dataset to remove outliers, missing values, and irrelevant variables that could potentially skew the analysis. This ensures that our model is trained on high-quality, relevant data. The removal of outliers and irrelevant variables can help us focus on the most meaningful aspects of player performance and avoid potential biases in our model.

6. Model Fitting and Error Analysis:

I will fit the linear regression model to the filtered dataset using regression techniques.

Error analysis will be performed to evaluate the accuracy of our model's predictions and identify any discrepancies between the predicted and actual values of player performance metrics.

7. Explanation of Model Fit:

I will provide a detailed explanation of how the linear regression model fits the data, including insights gained from the analysis of player performance trends over time.

The interpretation of regression coefficients, significance tests, and goodness of fit measures will be explained in order to illustrate the effectiveness of the model in capturing patterns in the data.