

NBA Linear Regression

Kevin Joo - DA Flex Student

Introduction

Objective: Analyze NBA Team stat data to predict win percentage

Methodology

Data:

- 20 years of data from the ESPN website**

Tools:

- NumPy and Pandas for data manipulation**
- Matplotlib and Seaborn for data visualizations**
- BeautifulSoup for web-scraping**
- Sklearn for linear modeling**

Methodology

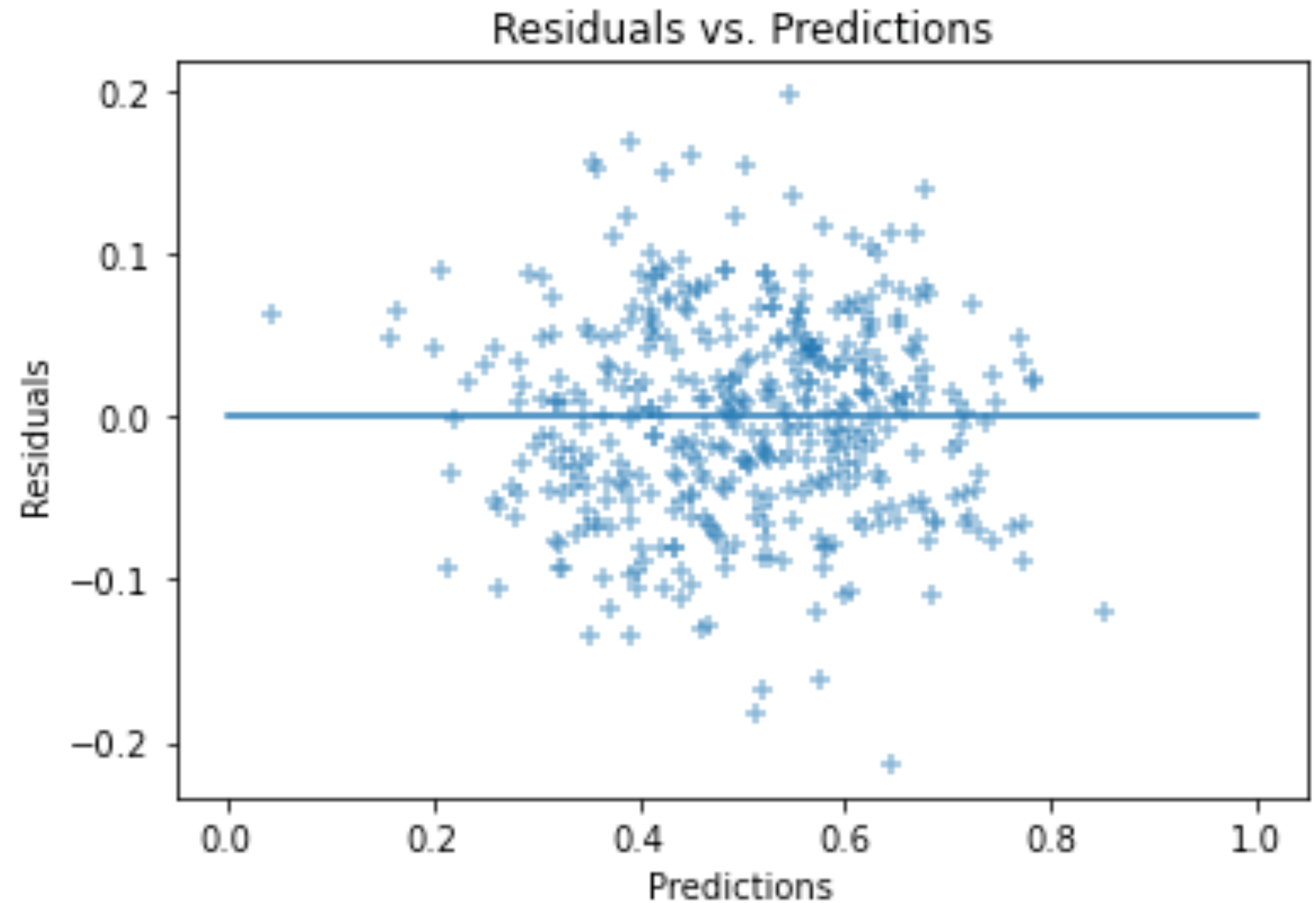
Predictor Values: "PTS", "FGM", "FGA", "FG%", "3PM", "3PA", "3P%", "FTM", "FTA", "FT%", "OR", "DR", "REB", "AST", "STL", "BLK", "TO", "PF"

Predicting Value: "WP" (Winning Percentage)

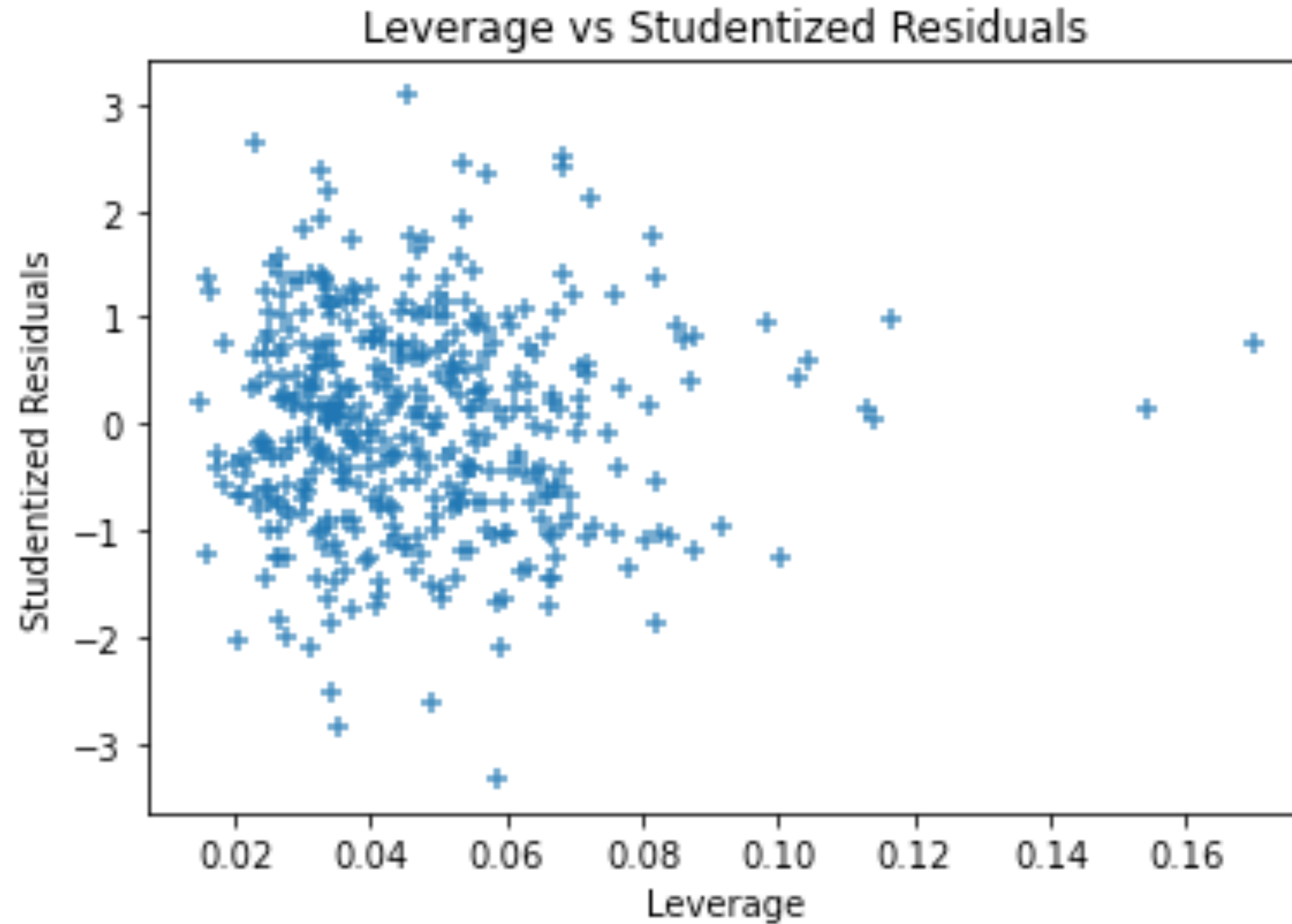
Results

Insight: (Homoscedasticity)

Table to the right do not follow a pattern, therefore there are no biased and/or skewed test results



Results



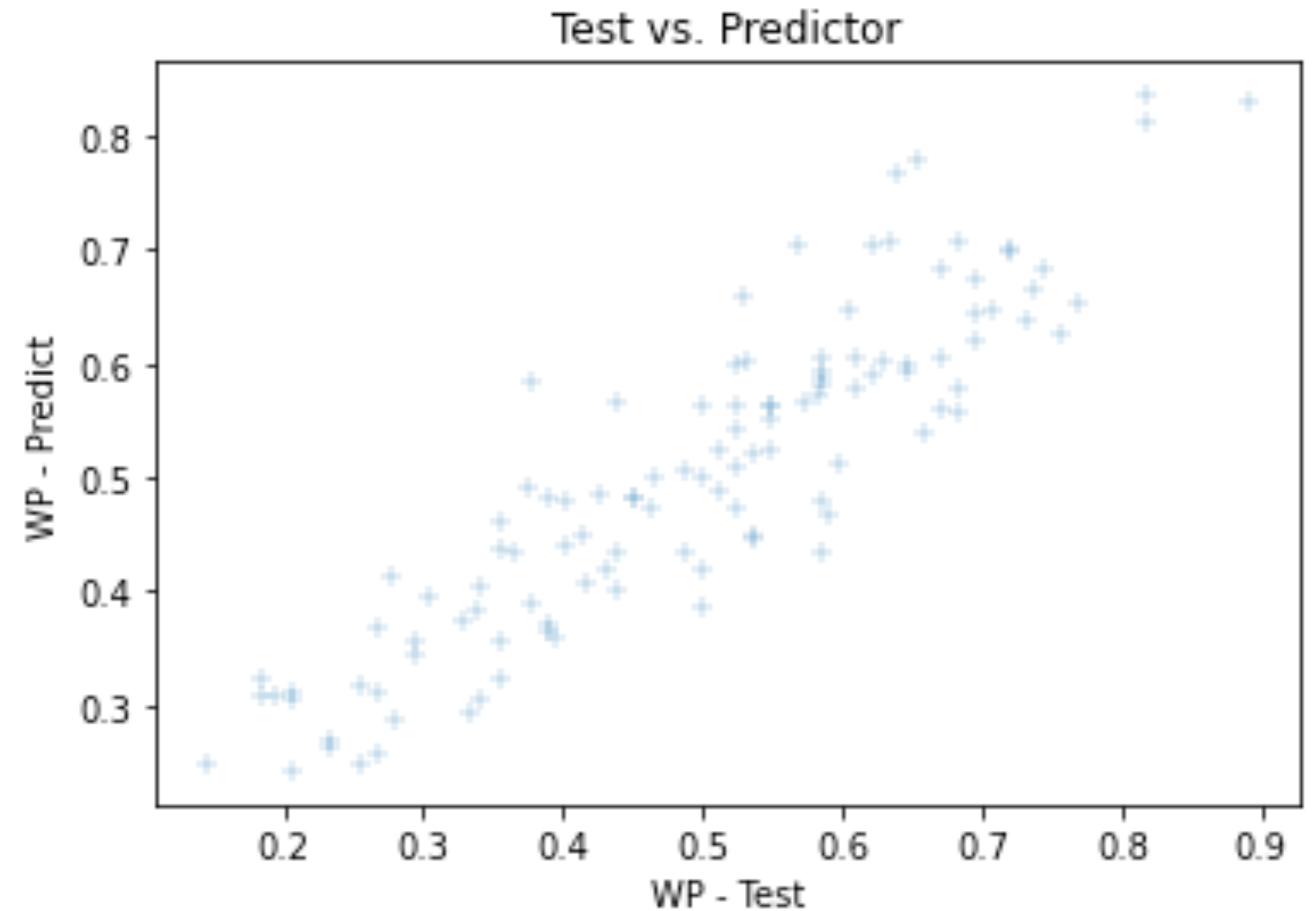
Results

Insight:

- Tested and Trained on different model-methods:
 - a. Used a 80/20 split for testing
 - b. Used a 75/25 split for training and validation
- Polynomial:
 - a. Training set score: 0.84
 - b. Validation set score: 0.45
- Linear Regression:
 - a. Training set score: 0.79
 - b. Validation set score: 0.83

Results

Simple Linear Regression:
Mean Squared Error (MSE): 0.005
Coefficient of Determination (R^2): 0.81



Future Work

If I had more time:

-Look at player stats and integrate a time-series into the model to see how player's development impacts the over-all team stats and the WP

End