QBUS2820 Assignment 1

This version was compiled on September 29, 2020

Introduction

Data processing and exploratory data analysis

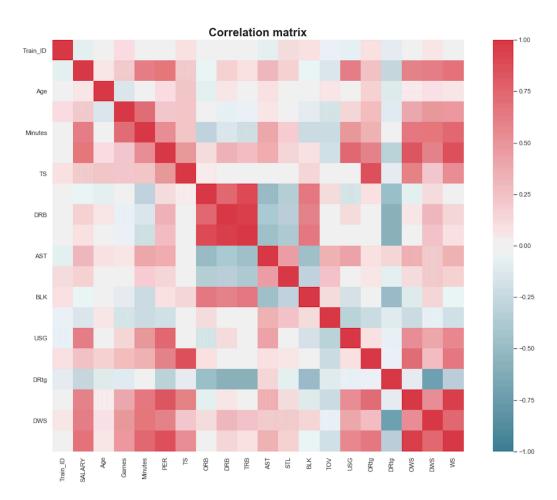


Fig. 1. Caption

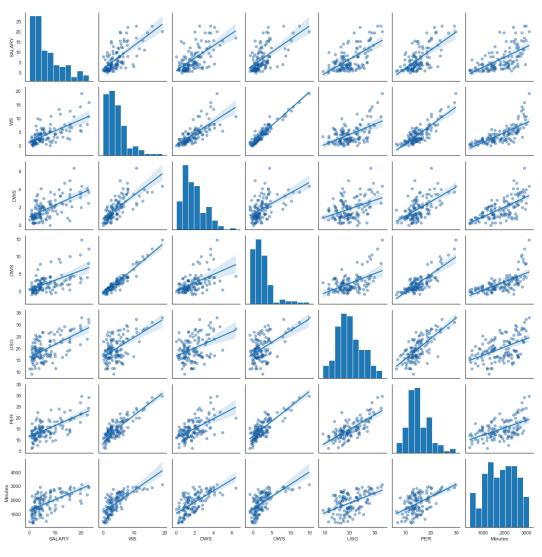


Fig. 2. Caption

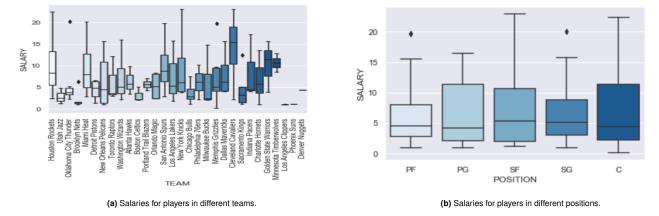


Fig. 3. Boxplots of salaries for players in different teams and positions.

As shown in Figuredsgi jsafoisjdf oifj salkadfj oij lfjsadfisoijfosfijsafdsnglkjgsdifjoasdj dsfjladkfeknwafka

Feature engineering

Methodology of the linear regression and kNN regression models

| Variable | Number of neignbours | Validation error | Polynomial degree | Validation error | |
|---|----------------------|------------------|--|---|--|
| DWS | 20 | 4.1311 | [2, 3] | 4.340583 | |
| WS | 8 | 4.1856 | [2, 3, 4] | 4.350843 | |
| Minutes | 9 | 4.2967 | [2] | 4.361685 | |
| ows | 27 | 4.5046 | [2, 3, 4, 5] | 8.342030 | |
| PER | 18 | 4.7217 | [2, 3, 4, 5, 6] | 8.407928 | |
| (a) Top 5 K nearest neighbbour models with the highest validation errors. | | | (b) Top 5 polynomial linear regression m | (b) Top 5 polynomial linear regression models with the highest validation err | |

Fig. 4. Validation errors of 10 models developed.

Methodology of the model that is not covered in this unit Test set performance