criteria:
RSE
adj r2
TSS & RSS

Rgression Ridge
Lasso
Cross validation then compare it to original LR

Covariance matrix

Plan:
Plot:
Debut on a fait un plot pour savoir comment les données varient en fonction de Time
D'apres la description des donnes on sait qu il va y avoir des correlations (interactions ?) entre facteur
Split Data:

RL:
    pour avoir une 1ere idée :
    on ne se fie pas du p value (cf names part 1)
    cook distance + leverage score ==> no problem in the data
    mean squared is HUGE
    QQplot ==> non linearity = heterosedacticity
R2 : 20% data sont expliques par LR —> PAS OUF!

KNN:
we also tried Knn but the mean squared is not any better

Subset selection:
    since we only have 32 variables we can afford having an exhaustive method
    best_bic :
        mean squared
    best R2:
        best r2 =0.29

but we still have same problem in QQ plot

Ridge model:

Lasso:

try to find something (alpha between 0 and 1) that works better

Principle component regression
    since we are sure there are interactions
    we find 4 components (we dont know which)
    for now it gives us the best mean squared error = 960