

## Assignment 3. Due date: 24/09

<http://dataanalyticsedge.com/2019/11/29/diamond-price-prediction-using-lasso-and-ridge-regression-in-python/>

이 사이트처럼 주제는 다이아몬드의 가격을 예측하는 것입니다. 많은 모형들이 있지만 우리는 회귀모형 4가지 (linear regression, ridge, lasso, elastic net)을 배우고 있으니 이것만 이용하겠습니다.

머신 러닝을 위해서는 우리가 무엇을 하고자 하는지 목적이 필요합니다. 그 순서대로 잘 적힌 사이트라서 참고해서 작성해보면 좋을 것 같습니다. 보고서는 모드 파이썬으로 작성하시고 zip파일로 제출하지 말아주시기 바랍니다.

자료를 받으면 먼저

1. 탐색
2. 데이터 정제 (missing value 등의 처리)
3. 목적에 맞는 분석

의 단계를 거치게 됩니다. Regularization이 필요한 이유는 먼저 회귀분석에서 문제가 발생했기 때문입니다. 그럼 이 문제가 발생하는지 먼저 check를 해보아야 하겠지요? 그 과정이 바로 correlation matrix를 살펴보거나 각각의 그림을 그려서 관계를 유추해보는 것입니다. 참고하도록 드린 코드를 참고하셔서 원하는 분석 순서와 원하는 표 및 그림의 모양대로 만들어내어 제출해주시기 바랍니다. 목적은 다이아몬드의 가격 예측입니다. 반드시 보내 드린 코드로만 분석하지 않아도 됩니다. 목적에 맞도록 논리적으로 작성해보시면 되겠습니다. 분석을 할 때는 이미 짜여진 코드를 자신의 목적에 맞게 수정해가면서 배워나가다 보면 더 깊이 있는 분석코드를 작성할 수 있게 됩니다.

분석 참고 Kaggle

<https://www.kaggle.com/shrutimechlearn/types-of-regression-and-stats-in-depth>

<https://www.kaggle.com/danielmarrama/diamond-exploratory-analysis-and-visualizations>

<https://www.kaggle.com/hakeemtfrank/predict-diamond-prices-elasticnet-random-forest>

<https://www.kaggle.com/fuzzywizard/diamonds-in-depth-analysis>