Daegeon Han ([1dae@g.skku.edu](mailto:1dae@g.skku.edu))

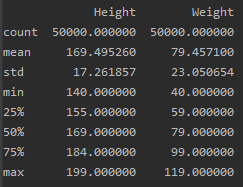
표준화데이터생성보고서

데이터 명 : 남녀 키, 몸무게 랜덤데이터

데이터 특징 :

남녀 키, 몸무게에 대한 랜덤한 데이터 생성 후 Discretization, Bivarization, Variable Transformation)를 제공합니다.

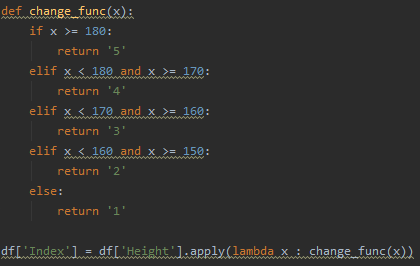
데이터 preprocessing :

1.  Discretization

* Classification 을 하기 위한 범위를 지정할 때

Height, Weight 연속적인 데이터에 5개의 카테고리로

동등하게 Mapping 할 수 있는 경계값을 구해보았습니다.

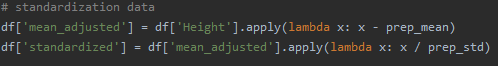
1. Variable Transformation (Simple Function)

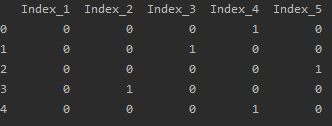
* 전체 데이터 5만개 중 Height 범위를

Col= Index에 1,2,3,4,5 데이터로 구분.

(Value를 10 단위로 구분)

* Height : Mean = 169.495260 , STD = 17.261857
* Weight : Mean = 79.457100 , STD = 23.050654

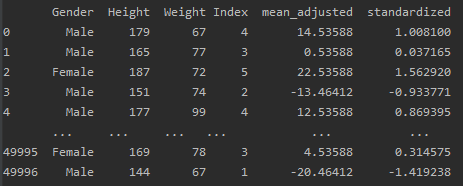


1. Bivarization

* 구분한 Index를 Bivarization

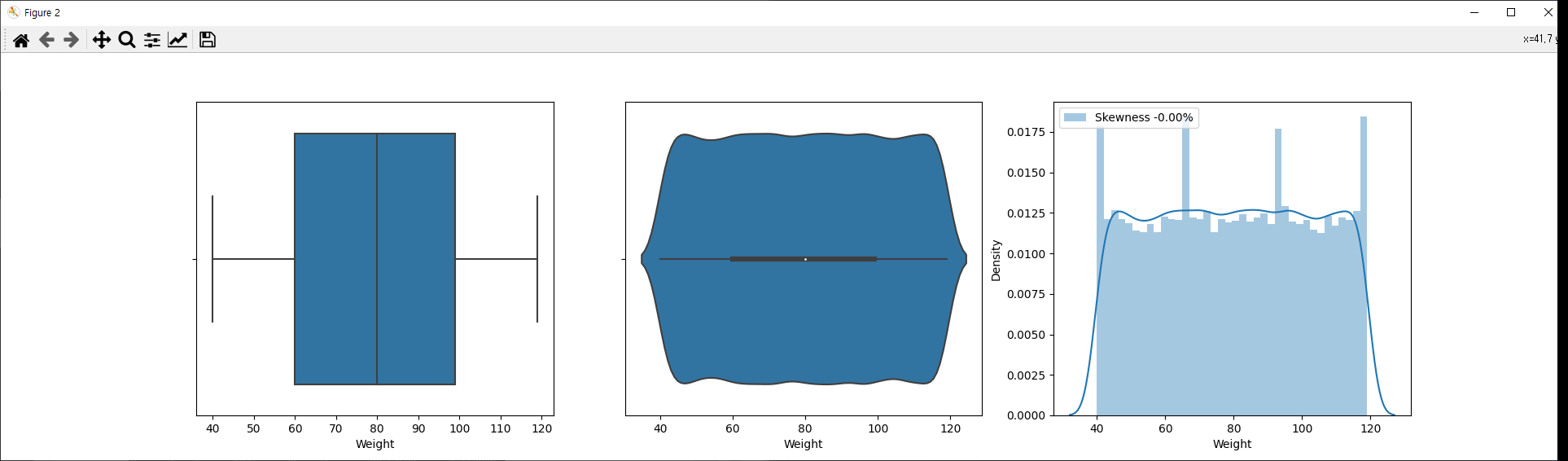
(one-hot encoding)

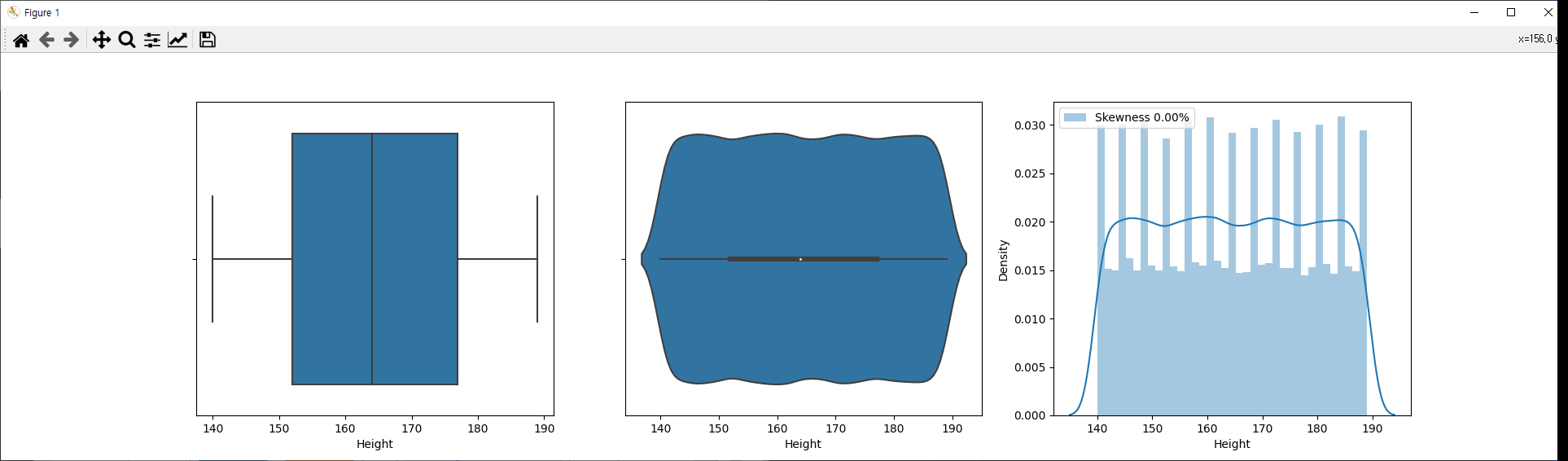
결과의 함의 :



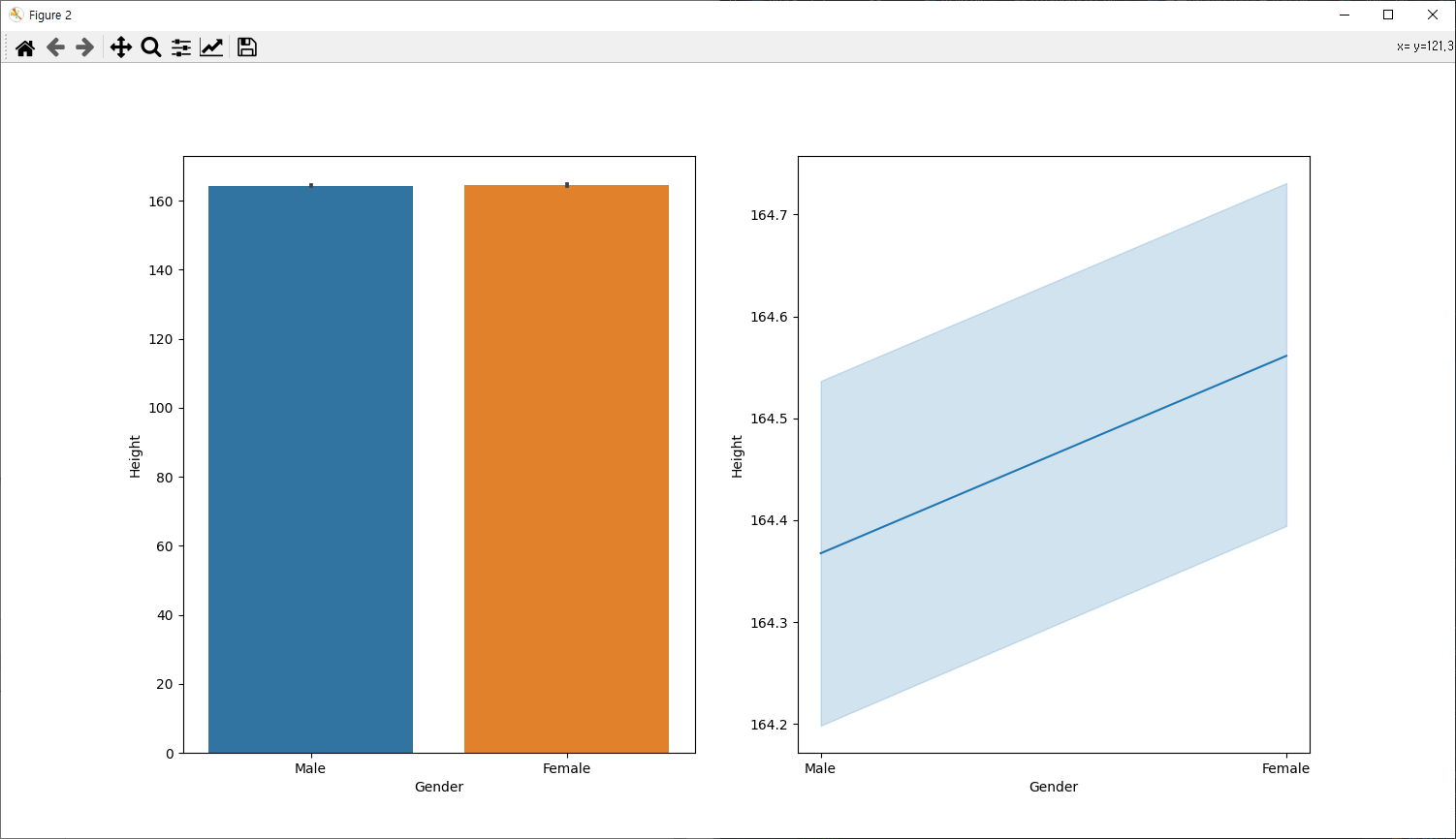
Data Preprocessing 을 통해 분석해본 결과

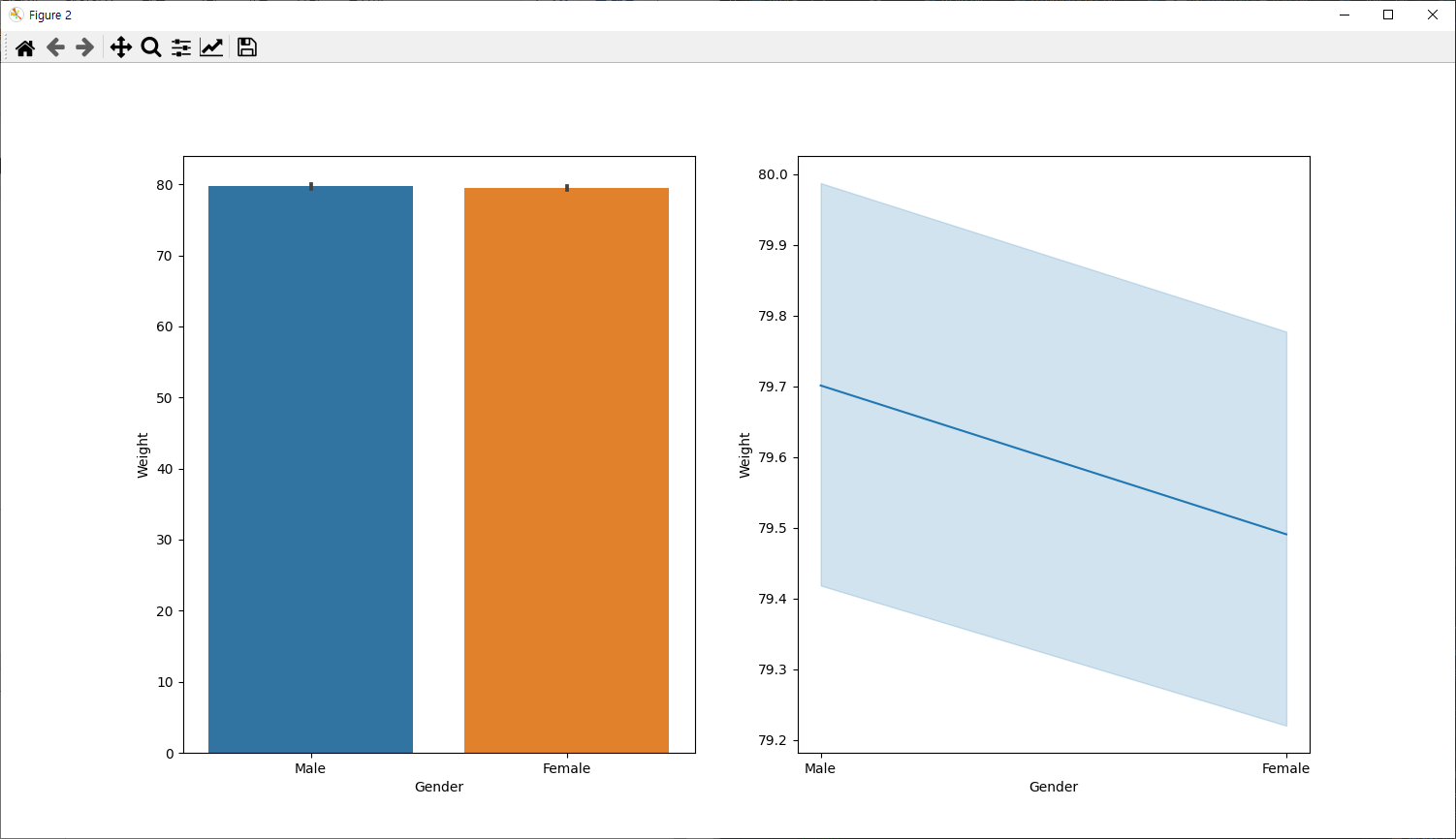
# UNIVARIATE ANALYSIS



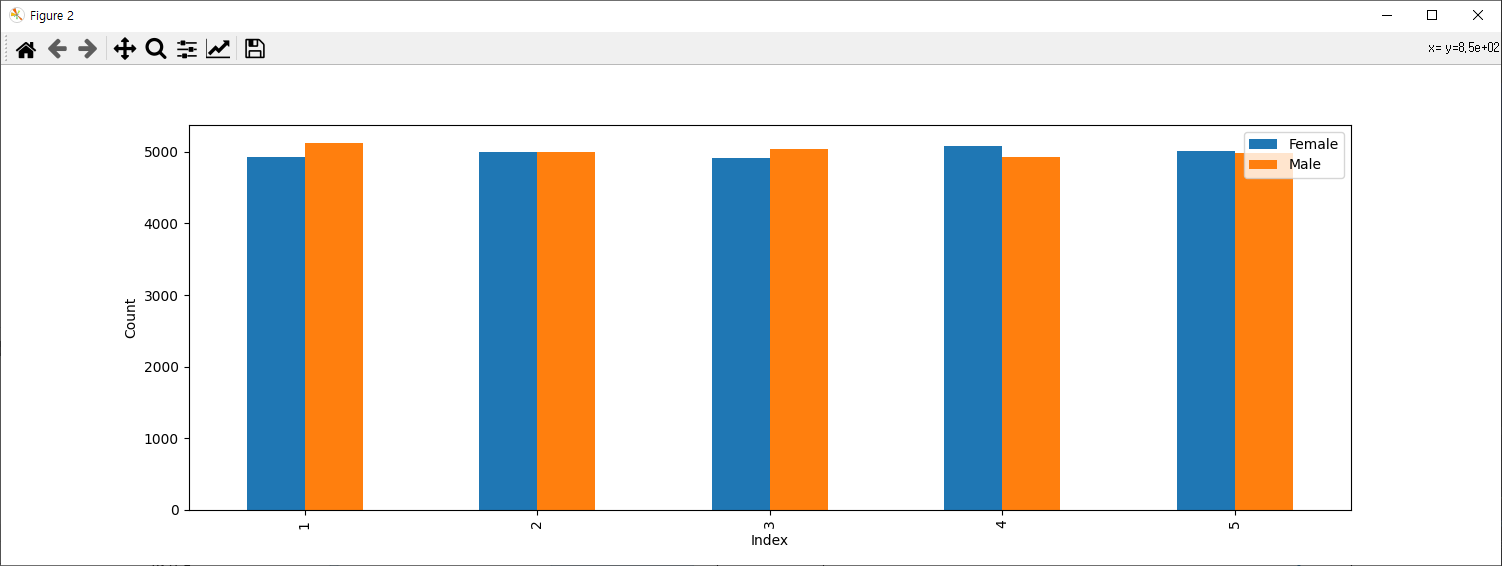


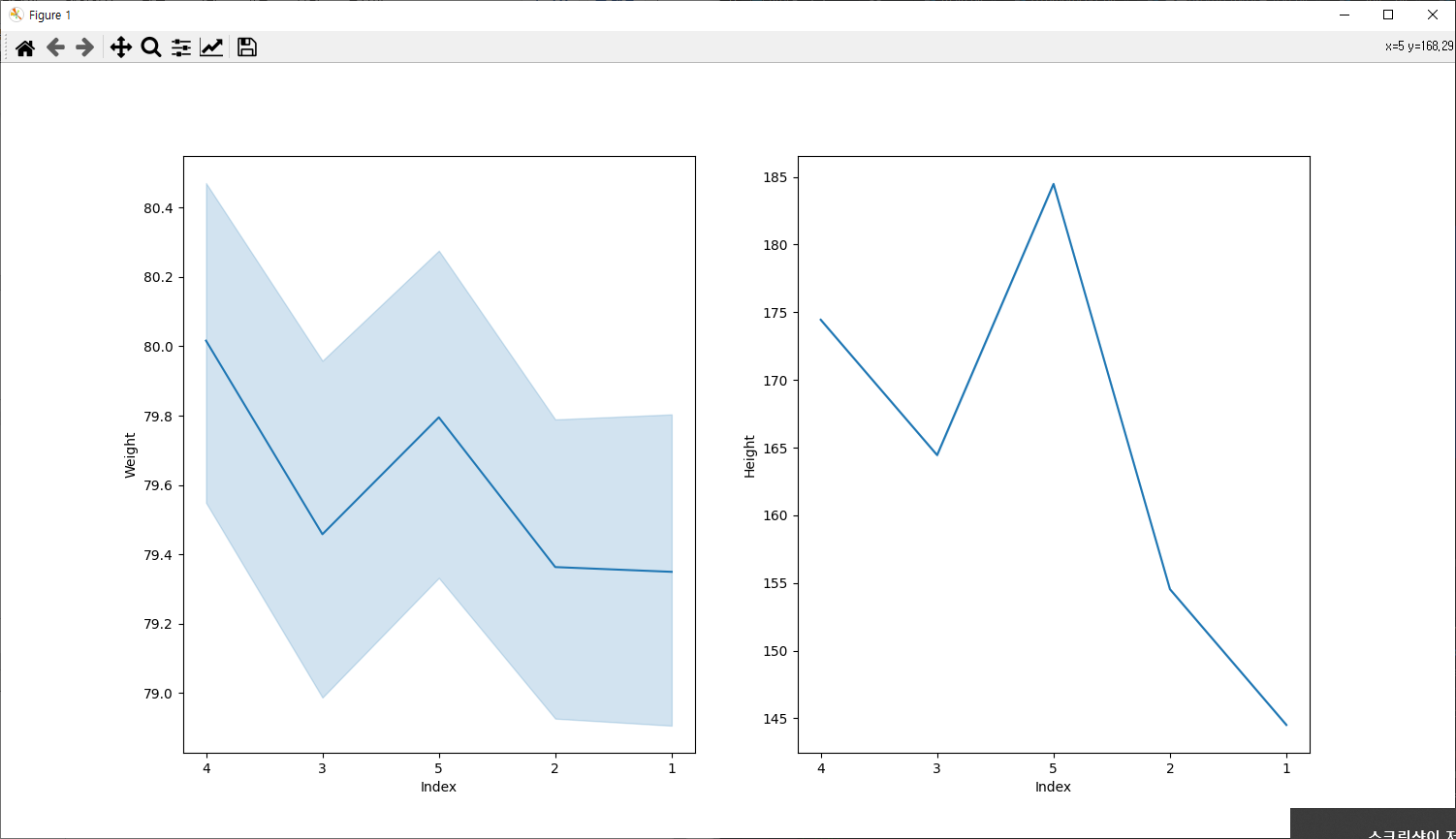
# BIVARIATE ANALYSIS





# The mean weight of Males is slightly more than that of females





# Females are more Extremely Weak than males  
# Males are more Weak than females  
# Females are more Normal than males  
# Females are more Overweight than males  
# Females are more Obese than males  
# Males are more Extreme Obese than females

