1. GP\_통합(3years)\_ver1.csv

=> 모든 변수 포함된 통합본

1. GP\_통합(3years)\_ver2.csv

=> 종사장 지위 컬럼 삭제 (취업여부와 동일한 값)

=> 석/박사 재학 여부 확인하여 미취업인데 석/박사 재학중인 데이터 삭제

(54,443 -> 51,690 (2,753개 삭제)

1. GP\_통합(3years)\_ver2\_취업예측용.csv

=> ver2에서 취업예측용 독립변수만 선정하여 추출

(제외한 컬럼 :

‘산업분류',

'직업분류',

'직원수',

'사업체위치(시도)',

'정규직여부',

'근로소득(현재\_월\_만원)' ,

'근로소득(초임\_월\_만원)' ,

'신입여부'

1. GP\_통합(3years)\_ver2\_타분석용.csv

=> 위에서 제외한 컬럼만 모아놓은 csv 파일

1. GP\_통합(3years)\_ver3\_원핫\_취업예측용.csv

범주형 변수들에 대해서 원핫인코딩 진행

1. GP\_통합(3years)\_ver4.csv

=> GP\_통합(3years)\_ver1.csv 에서 종속변수 2개 추가

1. GP\_통합(3years)\_ver4\_취업예측용.csv

ver4에서 컬럼 제외

(제외한 컬럼 :

‘산업분류',

'직업분류',

'직원수',

'사업체위치(시도)',

'정규직여부',

'근로소득(현재\_월\_만원)' ,

'근로소득(초임\_월\_만원)' ,

'신입여부'

1. GP\_통합(3years)\_ver5\_원핫.csv