

인천광역시 일자리 미스 매치 해소를 위한

일자리 미스 매칭 현황 분석 결과 보고서



2020. 12. 28.

[핵심요약]

- 배경 및 목적

경인지방통계청 2020년 6월 인천광역시 고용동향에 따라 인천광역시의 실업률 상승을 비롯한 자료를 기반으로 본 분석에서는 기존 표준분석모델과 다르게 인천광역시 최신 워크넷 구직/구인 데이터와 통계청의 인구통계 데이터를 활용해 구직자의 현황 파악으로 인천 일자리 창출 정책 수립 등의 기초자료 활용을 목적으로 한다.

- 활용데이터

활용 데이터		
인천광역시 구직 데이터	인천광역시 성별 경제활동인구 데이터	인천광역시 교육정도별 경제활동인구 데이터

- 분석절차



- 분석결과

- 1) 취업률은 2016년부터 2020년까지 하락세.
- 2) 희망급여의 경우 학력이 올라갈수록 높아짐.
- 3) 성별, 학력에 따라 희망직종의 차이가 있음.
- 4) 초등학교 & 중학교 졸업자의 경우 경력은 길지만 희망하는 급여가 낮아서 취업이 잘 되는 편인 것으로 추정.
- 5) 무학의 경우 2019년의 구직자들의 취업률이 급감한 이유로 취업이 잘 되는 50, 60대의 희망이 줄고, 취업이 어려운 10, 20대의 희망이 늘어나서 전체적으로 무학의 취업률이 급감했을 것으로 추정.
- 6) 고등학교 졸업자의 경우 남녀에 따라 희망하는 직종의 차이가 큼.
- 7) 대학교 & 대학원 졸업자는 희망급여가 높고 연구직의 비중이 높아짐.
- 8) 취업유무를 분류하는데 연도, 희망급여가 중요해보임.

- 제언

- 1) 여성을 위한 다양한 직업 교육
- 2) 저학력자(무학, 초졸, 중졸)의 재취업을 위한 일자리 마련
- 3) 인천지역 2030 청년 유출 방지 정책

목 차

I. 분석 개요	4
1. 개요	5
2. 데이터 설명	7
3. 데이터 속성	7
II. 분석 과정	10
1. 분석 프로세스	11
2. 데이터 전처리	12
3. 분석방법 및 절차	18
III. 분석 결과	19
1. 워크넷 데이터 현황 분석	20
2. 모델링	57
IV. 요약 및 제언	64
1. 결과 요약	65
2. 한계점 및 제언	69
부록	72

분석 개요

- 1. 개요
- 2. 데이터 설명
- 3. 데이터 속성

1. 개요

가. 추진배경 및 목적

- 만성적인 일자리 부족, 지역기업의 구인난은 국정 및 자치단체의 핵심과제이다.



- 통계청이 발표한 '2020년 6월 고용동향'에 따르면 1999년 이후 역대 처음으로 10%대를 경신하며 청년 10명 중 1명은 일자리를 얻지 못한다. 청년층 취업만 힘든 것이 아닌 60대를 제외한 모든 연령 계층에서 1년 전과 비교하여 20대(-15만1000명), 30대(-19만5000명), 40대(-18만 명), 50대(-14만6000명) 등 취업자가 감소했다. 취업률이 증가한 연령대는 60대 이상이 유일하다. 1년 전과 비교하면 33만 8000명이 증가했다.
- 특히, 경인지방통계청에 따르면 인천광역시는 지난 5월 실업률 4.7%에서 6월 5%로 0.3%p 증가해 전국 평균 실업률(4.3%)보다 0.7%p 높았다.
- 6월 인천지역 취업자수는 156만 6천명으로 지난해 같은 기간 대비 3만 3천명 감소했다. 전년 동월 대비 2만 4천명이 줄어들었던 5월보다도 감소폭이 더욱 확대되어 취업자수가 줄고 있는 것으로 나타나 인천 지역 구직자들의 연령대, 최종학력, 희망직종 등 현황을 파악하여 인천 지역 일자리 창출 정책 발전에 기여할 것으로 기대한다.
- 또한, 일자리 현안문제 해결을 위한 빅데이터 분석 기반의 맞춤형 일자리 분석모델은 자치단체에 확산 보급 시 고용안정성 실현에 기여할 것으로도 기대한다.

나. 기존 일자리 미스 매칭 표준분석모델

구 분	내 용
미스매치 원인분석	지역산업구조와 고용 인력의 관계성 분석 일자리 미스매치의 각종 조건별 원인 분석 지역·성별·연령별 선호 일자리 분석
프로파일 구축	구직자/구인기업 직무별 선호특성을 반영한 지능형 프로파일링
일자리 추천모델	직장 맞춤형 일자리 추천 모델 구인기업 구인 성공예측조건 추천 모델

[표 1-1] 사업내용

<행정안전부 일자리 미스 매칭 표준분석모델 사업내용>

- 기존 일자리 미스 매칭 표준분석모델은 경상남도 김해시 기준 2006년 ~ 2015년 워크넷 구인/구직 데이터와 통계청의 인구통계자료 등을 활용한 분석 모델이다.
- 본 분석에서는 기존과 동일하게 표준분석모델의 추진배경과 목적으로 진행하며, 인천광역시 최신 워크넷 구직/구인 데이터(2016.01~2020.06)와 통계청의 인구통계 자료를 활용하여 구직자의 현황 파악으로 인천 일자리 창출 정책 수립 등의 기초자료 활용을 목적으로 한다.

다. 분석 주제 및 데이터

- 주 제 : 인천광역시 일자리 미스 매칭 분석
- 데 이 터 : ‘16. 1. 1. ~ 20. 6. 30.’ 4년 6개월 간의 인천광역시 워크넷 구직 데이터 및 인구 통계 데이터
- 데이터 수집
 - 워크넷 구직 데이터 : 고용노동부 워크넷 데이터 수령
 - 경제활동인구 통계 : 통계청 수집

2. 데이터 설명

가. 데이터 개요

구 분	데이터	데이터 내용
워크넷	인천광역시 구직 데이터	워크넷 등록 구직자의 성별, 나이, 주소 등 정보 데이터
경제활동인구	인천광역시 성별 경제활동인구 데이터	표본가구 내 상주하는 만 15세 이상 가구원의 성 별 인구 데이터
경제활동인구	인천광역시 교육정도별 경제활동인구 데이터	표본가구 내 상주하는 만 15세 이상 가구원의 교 육정도별 인구 데이터

3. 데이터 속성

○ 워크넷 구직 데이터 (1,160,510 건)

항목	데이터 형태	활용 유무	비고
행번호	int	x	분석에 무의미한 데이터
연도	int	o	
성별	chr	o	
나이	chr	o	
주소	chr	o	인천광역시 주소만 사용
구직등록일자	date	x	분석에 무의미한 데이터
해외취업희망여부	chr	x	데이터 편중이 심함
여성가장여부	chr	x	데이터 편중이 심함
도서(섬)지역거주여부	chr	x	데이터 편중이 심함
장애여부	chr	x	데이터 편중이 심함
워크넷공개여부	chr	x	분석에 무의미한 데이터
문자알림수신여부	chr	x	분석에 무의미한 데이터
최종학력	chr	o	
졸업여부	chr	x	최종학력과 유사
희망직종1	chr	o	
희망직종2	chr	x	결측치 많음(약 70%)
희망직종3	chr	x	결측치 많음(약 90%)
신입경력여부1	chr	o	
신입경력여부2	chr	x	결측치 많음(약 70%)

신입경력여부3	chr	x	결측치 많음(약 90%)
경력1(년수)	int	o	
경력1(개월수)	int	o	
경력2(년수)	int	x	결측치 많음(약 74%)
경력2(개월수)	int	x	결측치 많음(약 77%)
경력3(년수)	int	x	결측치 많음(약 84%)
경력3(개월수)	int	x	결측치 많음(약 85%)
희망직무내용1	chr	x	분석 시간 부족
희망직무내용2	chr	x	결측치 많음(약 70%)
희망직무내용3	chr	x	결측치 많음(약 90%)
희망근무지역1	chr	o	
희망근무지역2	chr	x	결측치 많음(약 51%)
희망근무지역3	chr	x	결측치 많음(약 81%)
희망급여형태	chr	x	연봉으로 통일하여 급여로 사용
희망급여	int	o	
고용형태(작업조정여부)	chr	x	분석에 무의미한 데이터
고용형태	chr	o	
고용형태(파견근로여부)	chr	x	분석에 무의미한 데이터
고용형태(대체인력 취업 가능여부)	chr	x	분석에 무의미한 데이터
병역특례희망	chr	x	분석에 무의미한 데이터
희망근무시간	chr	x	정제하기 힘든 데이터
희망근무시간 (2교대 가능여부)	chr	x	결측치 많음(약 88%)
희망근무시간 (3교대 가능여부)	chr	x	결측치 많음(약 89%)
그밖의희망사항	chr	x	정제하기 힘든 데이터
운전가능여부	chr	o	
차량소유여부	chr	o	
컴퓨터활용능력 -문서작성여부	chr	o	
컴퓨터활용능력 -스프레드시트여부	chr	x	데이터 편중이 심함
컴퓨터활용능력 -프리젠테이션여부	chr	x	데이터 편중이 심함
컴퓨터활용능력 -회계프로그램여부	chr	x	데이터 편중이 심함
컴퓨터활용능력 -기타프로그램여부	chr	x	데이터 편중이 심함
경력건수	int	o	
구직종료처리명	chr	x	분석에 무의미한 데이터
취업유무	chr	o	

○ 행정구역 성별 경제활동인구 데이터 (504 건)

항목	데이터 형태	활용 유무	비고
시도별	chr	o	인천광역시만 사용
성별	chr	o	
15세이상인구 (천명)	int	o	
경제활동인구 (천명)	int	o	
비경제활동인구 (천명)	int	o	
경제활동참가율 (%)	int	o	
실업률 (%)	int	o	
고용률 (%)	int	o	
15~64세 고용률 (%)	int	o	

○ 행정구역 교육 정도별 경제활동인구 데이터 (1,134 건)

항목	데이터 형태	활용 유무	비고
시도별	chr	o	인천광역시만 사용
교육정도별	chr	o	
15세이상인구 (천명)	int	o	
경제활동인구 (천명)	int	o	
비경제활동인구 (천명)	int	o	
경제활동참가율 (%)	int	o	

분석 과정

- 1. 분석 프로세스
- 2. 데이터 전처리
- 3. 분석방법 및 절차

1. 분석 프로세스

프로세스	구분	내용
데이터 수집	워크넷 데이터	워크넷 데이터 관리 담당자(고용노동부)에게 요청 및 수령
	통계 데이터	kosis 홈페이지를 통해 수령
데이터 전처리	이상치 제거	경력(년수), 희망 급여, 경력 건수의 이상치 제거
	결측치 제거	경력2(년수), 경력2(개월수), 경력3(년수) 등 결측치 제거
	변수 변환	희망근무지역, 전체경력기간 생성, 직종 분류 통합
데이터 분석	EDA	파이썬 시각화 라이브러리를 통해 EDA 실시
	로지스틱 회귀	희망 직종 변수를 사용하여 모델링 후 오즈비 확인
	그래디언트 부스팅	모델링 후 변수 중요도 확인
	랜덤 포레스트	모델링 후 변수 중요도 확인 및 shap를 통한 모델 해석 실시
결과 도출	EDA 결과 도출	EDA 결과를 통해 정책적 제언 작성
	모델링 결과 도출	모델링을 통한 변수 중요도 정리 및 shap를 통해 모델 해석 정리

2. 데이터 전처리

변수	결측치	이상치
경력1(개월수)	0으로 대체	x
경력1(년수)	아래참조	Q1 - 2*IQR 이하값, Q3 + 2*IQR 이상값 제거
경력건수	아래참조	Q1 - 2*IQR 이하값, Q3 + 2*IQR 이상값 제거
희망급여	x	연봉 : 하위 10,000,000이하 150,000,000원 이상 제거 월급 : 하위 1,000,000이하 10,000,000원 이상 제거

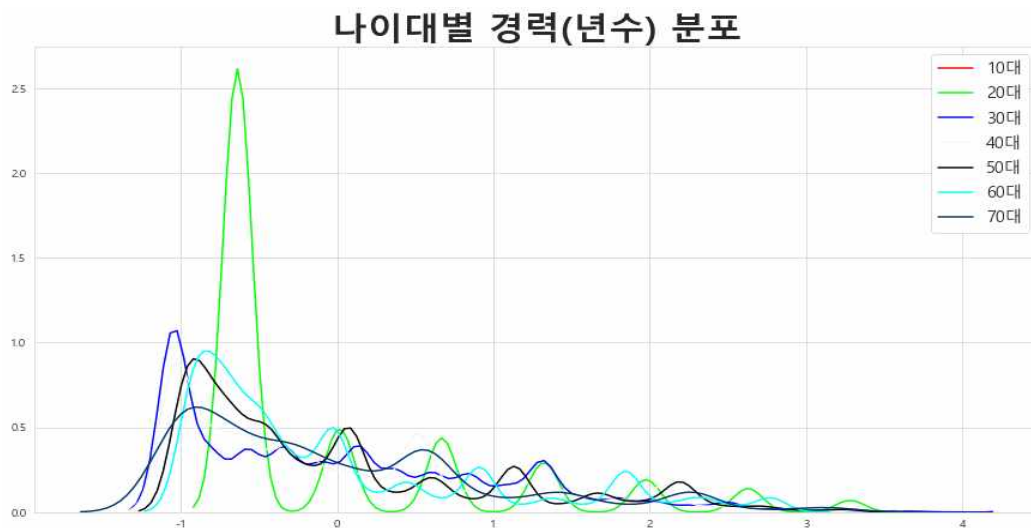
가. 경력1(년수), 경력1(개월수), 경력건수

- ‘경력2’, ‘경력3’은 결측치가 많아서 분석에서 제외하고, ‘경력1(년수)’, ‘경력1(개월수)’의 데이터 값을 결합하여 ‘전체경력기간’ 파생변수 생성.

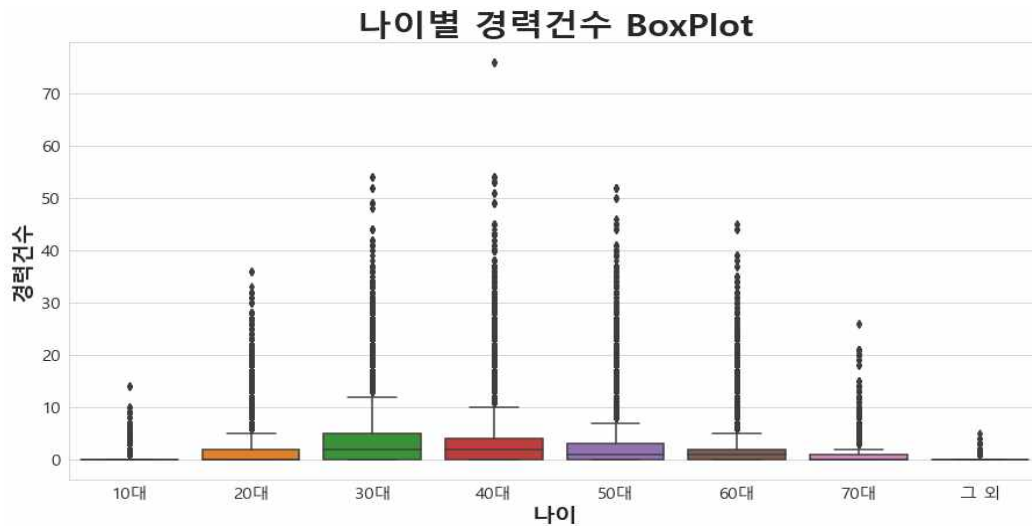
○ 이상치 처리

- 나이대로 나누어 이상치 제거(10대, 20대, 30대, 40대, 50대, 60대, 70대 기준으로 이상치 파악).

- 나이대별 경력(년수) 분포



- 나이별 경력건수 Boxplot



* box는 25%~75% 구간이고 가장 위쪽 경계선은 4분위선으로 그 위로 벗어나 있는 점들이 이상치임

변수	이상치 비율
10대 경력(년수)와 경력건수	13%
20대 경력(년수)와 경력건수	4%
30대 경력(년수)와 경력건수	1%
40대 경력(년수)와 경력건수	3%
50대 경력(년수)와 경력건수	3%
60대 경력(년수)와 경력건수	4%
70대 경력(년수)와 경력건수	6%

- 경력1(년수)가 이상치이거나 경력건수가 이상치인 경우 제거.
- 그 외로 분류된 나이는 결측치가 대부분이므로 모두 제거(약 90%).
- 데이터 개수 : 1,160,510건 > 1,118,197 건으로 변화.

○ 나이대별 경력과 경력건수 맞추기

- 나이에 비해 경력기간과 경력건수가 맞지 않는 경우가 있음.
- 경력기간과 경력건수가 맞지 않는 경우가 있음.
(ex. 경력은 10년인데 경력건수가 0)
- 경력1(년수)가 0이 아닌데 경력건수가 0인 경우 경력1(년수)에 따라 경력건수 최빈값으로 대체.
- 남은 데이터 중 경력1(년수)가 결측치의 경우 중앙값으로 대체.

○ 경력1(년수), 경력1(개월수) 통합

- 경력1(년수)와 경력1(개월수)/12를 더하여 전체경력기간 변수 생성.

나. 직종 분류

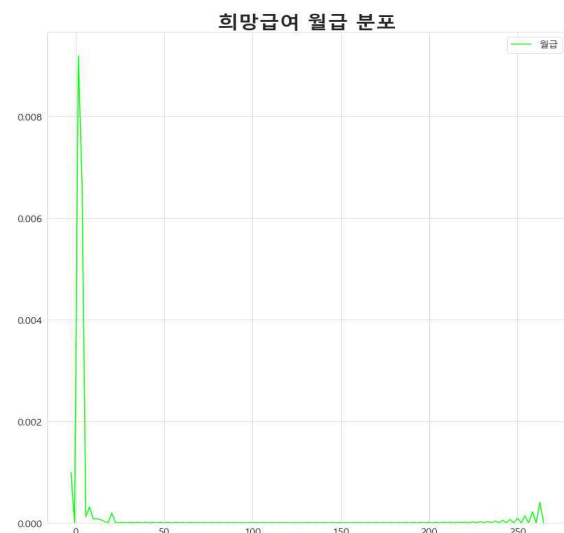
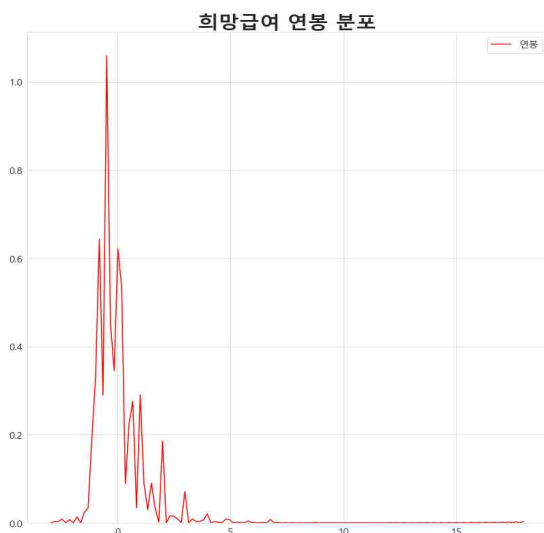
○ 직종분류 대분류로 통합

- 한국고용직업분류표(2018)을 통해 소분류 직종을 대분류로 통합. * [참고2] (대분류로 통합한 이유는 많은 범주로 인한 분석 시간을 줄이기 위함)
- 총 10개의 대분류('경영·사무·금융·보험직', '연구직 및 공학 기술직', '교육·법률·사회복지·경찰·소방직 및 군인', '보건·의료직', '예술·디자인·방송·스포츠 직', '미용·여행·숙박·음식·경비·청소직', '영업·판매·운전·운송직', '건설·채굴직', '설치·정비·생산직', '농림어업직') 생성.

* 한국고용직업분류 상의 직업의 정의 및 직업분류의 원칙은 ILO의 국제표준직업분류(ISCO-08) 및 한국표준직업분류(KSCO-2017)의 정의를 그대로 따르고 있으며, KSCO와KECO간 연계표 제공하고 있음.

다. 희망급여 & 희망급여형태

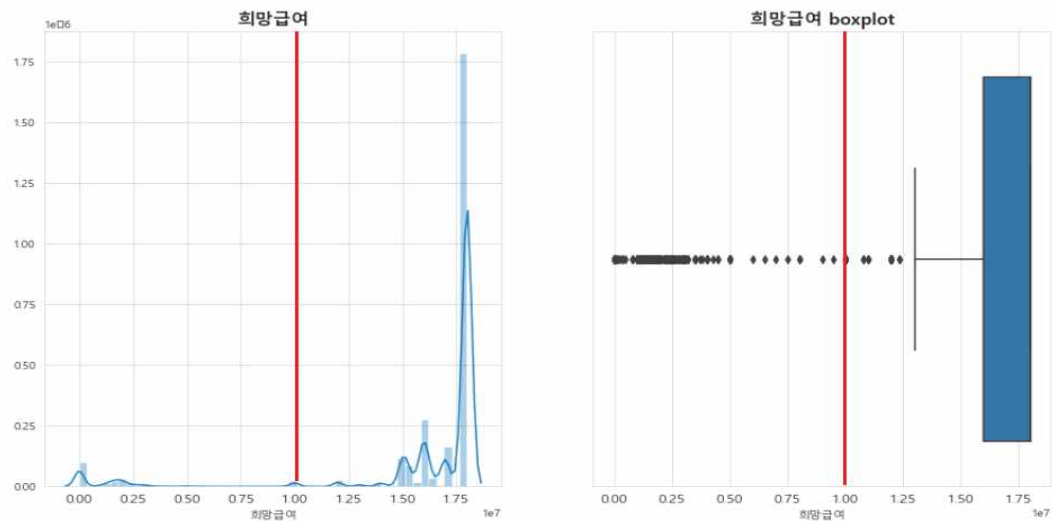
○ 이상치 파악



- 오른쪽 값의 영향이 매우 커보임. 특히 월급의 이상치가 심해보임.
- 시급의 경우 2016년 최저시급 6,030원 기준, 일급의 경우 2016년 일 8시간
* 시급/일급은 일용직, 단기직인 경우가 많을거라 판단하여 사용하지 않음.
- 연봉/월급 상위 5%, 연봉/월급 하위 5% 기준으로 이상치 임의로 판단.
* 연봉 상·하위 이상치를 모두 제거하지 않고, 분포 그래프로 적절히 판단하여 기준 정함.

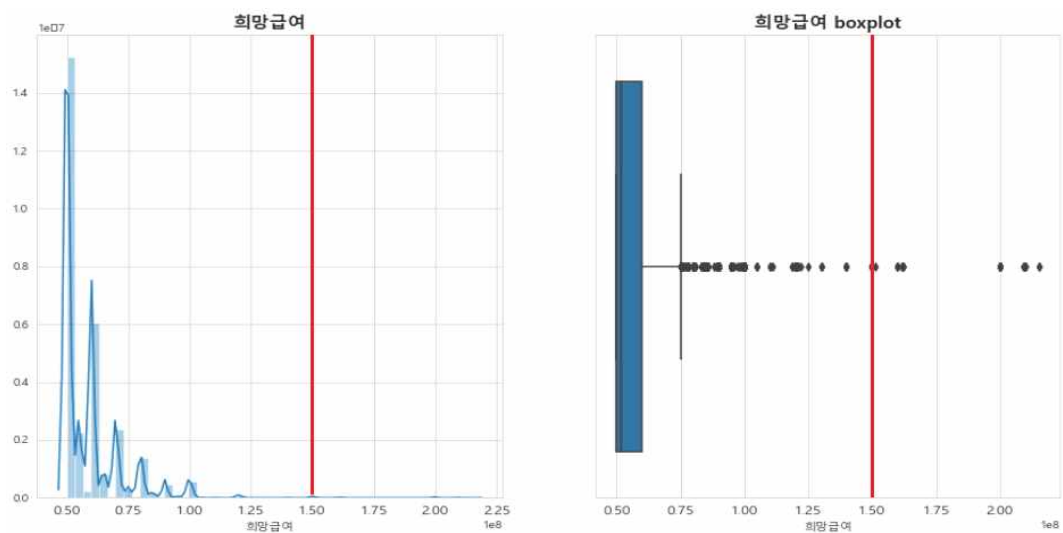
○ 연봉 상.하위 5% 이상치 판단 기준선

<연봉 하위 5%>



- 연봉 하위 이상치 기준으로 10,000,000원 이하 제거.

<연봉 상위 5%>

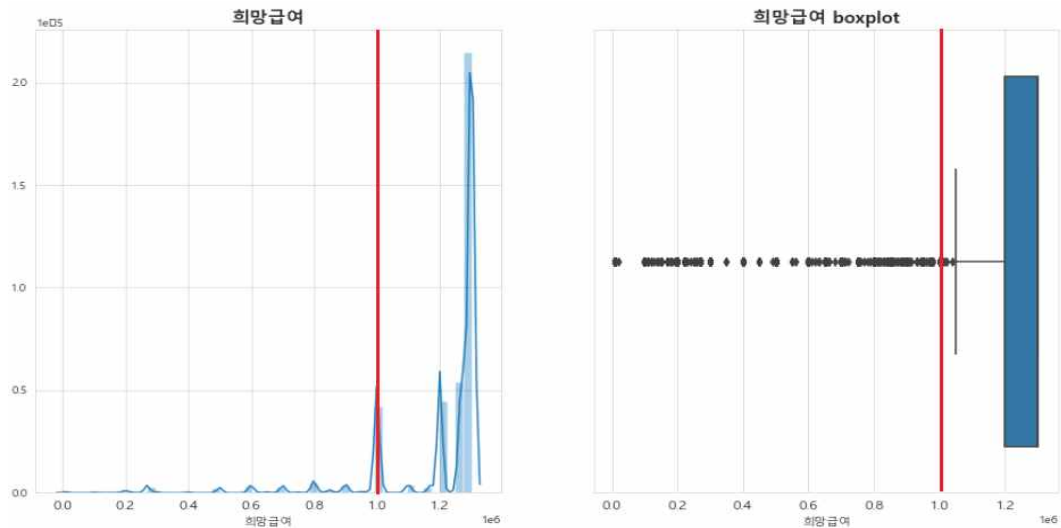


- 연봉 상위 이상치 기준으로 150,000,000원 이상 제거.

* 월급 상.하위 이상치를 모두 제거하지 않고, 분포 그래프를 보고 적절히 판단하여 기준을 정함.

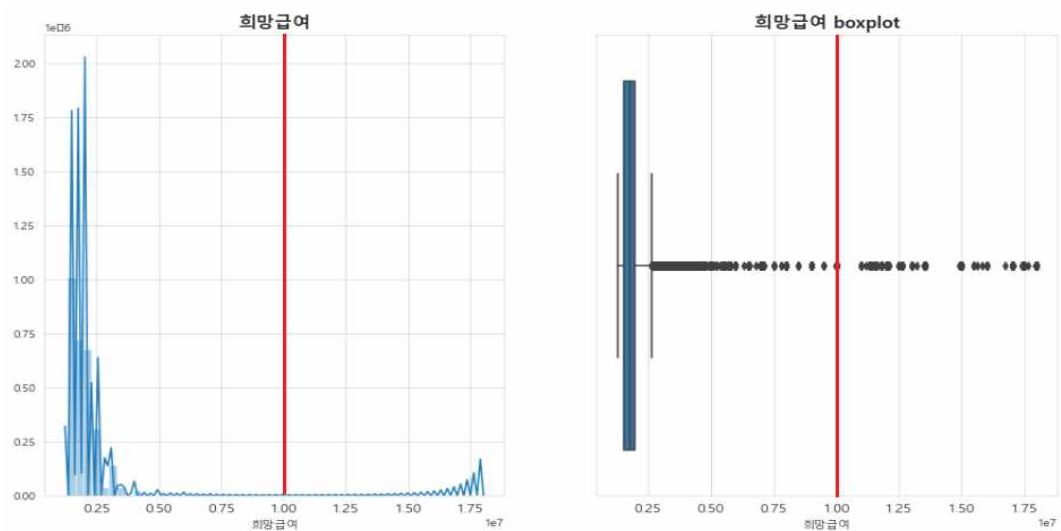
○ 연봉 상.하위 5% 이상치 판단 기준선

<월급 하위 5%>



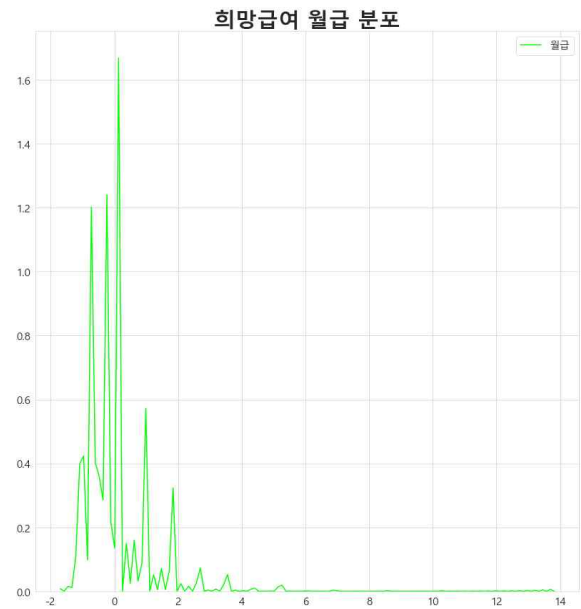
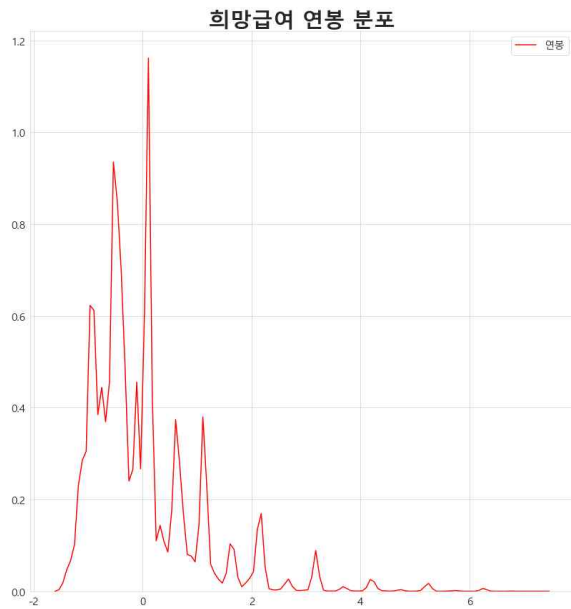
- 월급 하위 이상치 기준으로 1,000,000원 이하 제거.

<월급 상위 5%>



- 월급 상위 이상치 기준으로 10,000,000원 이상 제거.

* 월급 상.하위 이상치를 모두 제거하지 않고, 분포 그래프를 보고 적절히 판단하여 기준을 정함.



- 이전보다 상대적으로 퍼진 모양을 보임.
- 최종적으로 '희망급여'는 월급 데이터 값을 연봉으로 변환하여 급여를 연봉으로 통일함.

라. 희망근무지역

- 인천광역시시 기준 워크넷 구직자 데이터이므로 대부분 희망근무지역이 인천광역시였음.
- 인천(83%), 경기, 서울을 제외하고 나머지 지역의 비율은 5%임.
- 희망근무지역을 인천, 서울&경기, 비수도권으로 재분류.

3. 분석 방법 및 절차

가. 분석 방법

- 기술통계 : 빈도, 평균, 상관분석 등
- EDA (Exploratory Data Analysis 탐색적 데이터 분석) : 그래프나 통계적인 방법으로 데이터 분석을 하는 기법으로 구직 현황에 대한 파악.
- 로지스틱 회귀분석 : 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법으로 오즈비를 통해 어떤 희망직종이 취업확률을 높이는지 파악.
- 랜덤 포레스트 : 훈련 과정에서 구성된 다수의 결정 트리로부터 분류 또는 평균 예측치(회귀분석)를 출력하는 방법으로 워크넷에 등록된 인천광역시 구직자들의 취업 유무를 예측.
- shap : 해석이 불가능한 모델을 Shapley Values를 통해 해석 가능하게 해주는 패키지로 랜덤 포레스트를 해석하기 위해서 사용.
- 그래디언트 부스팅 : 여러 개의 약한 학습기를 순차적으로 학습-예측하면서 잘못 예측한 데이터에 가중치 부여(경사 하강법 사용)를 통해 오류를 개선해 나가면서 학습하는 방식으로 워크넷에 등록된 인천지역 구직자들의 취업 유무를 예측하고 어떤 변수가 취업 유무를 분류하는데 영향이 큰지 확인.
- 상관 계수 파악 : 두 변수 사이의 상관성의 정도를 나타내는 수치로 하나의 변수가 증가할 때 다른 변수도 증가하면 양의 상관을 하나의 변수가 감소할 때 다른 변수가 감소하면 음의 상관을 보인다고 함. 상관 계수는 $-1 \sim 1$ 의 값을 가질 수 있으며 1에 가까울수록 같이 증가하는 형태를 보이고 -1 에 가까울수록 같이 감소하는 형태를 보임. 두 변수의 상관성을 파악하기 위해 사용하였고, 인과관계를 뜻하지는 않음.

나. 분석 도구

- 언어 : python
- 라이브러리 : pandas, numpy, matplotlib, seaborn, scikit-learn

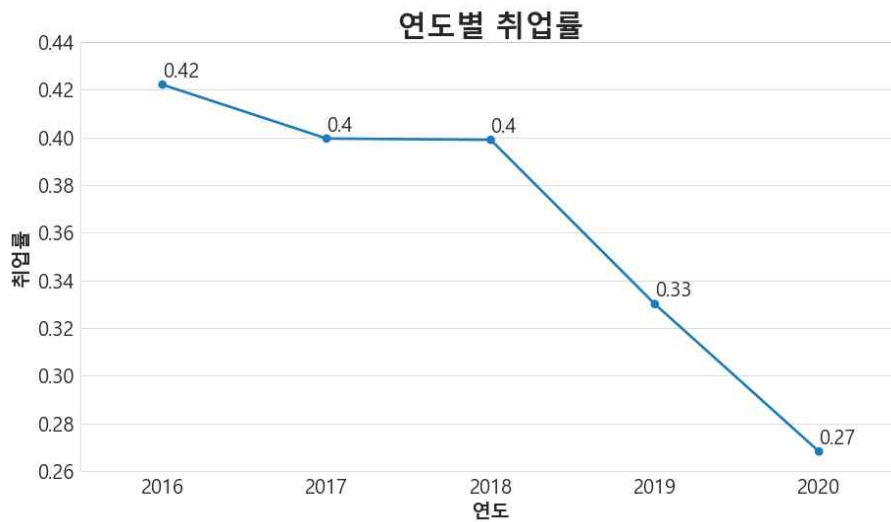
분석 결과

- 1. 워크넷 구직 데이터 현황 분석
- 2. 모델링

1. 워크넷 구직 데이터 현황 분석

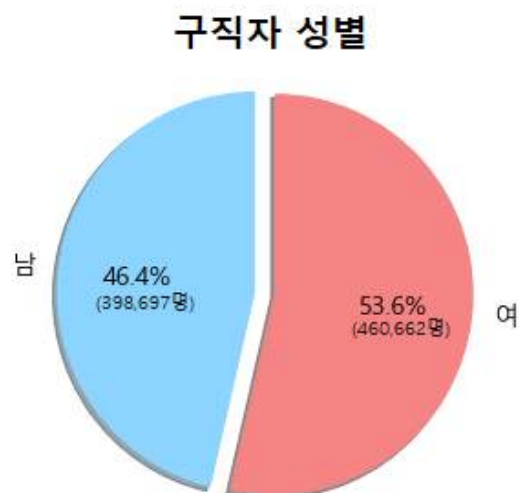
가. 구직자 현황

1) 연도별 취업률



○ 2016년부터 워크넷 구직 데이터 기준으로 취업률이 하락세를 보임.

2) 구직자 성별



○ 워크넷 등록 구직자는 여성(53.6%)이 남성(46.4%)보다 많음.

3) 구직자 연령대



○ 구직자 연령대는 20대(25%)가 가장 많고, 10대와 70대는 현저히 적음.

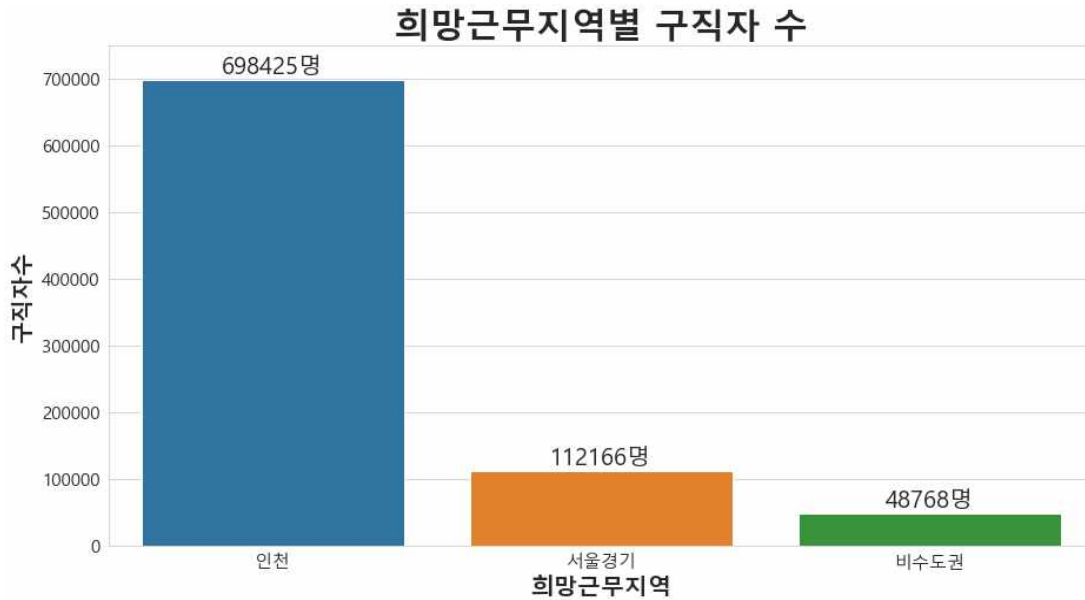
4) 행정구별 구직자 수



○ 인천광역시 10개 시·군구 중, 가장 구직자 등록 수가 많은 세 지역은 남동구(173,140명), 서구(165,399명), 부평구(160,641명) 순으로 나타남.

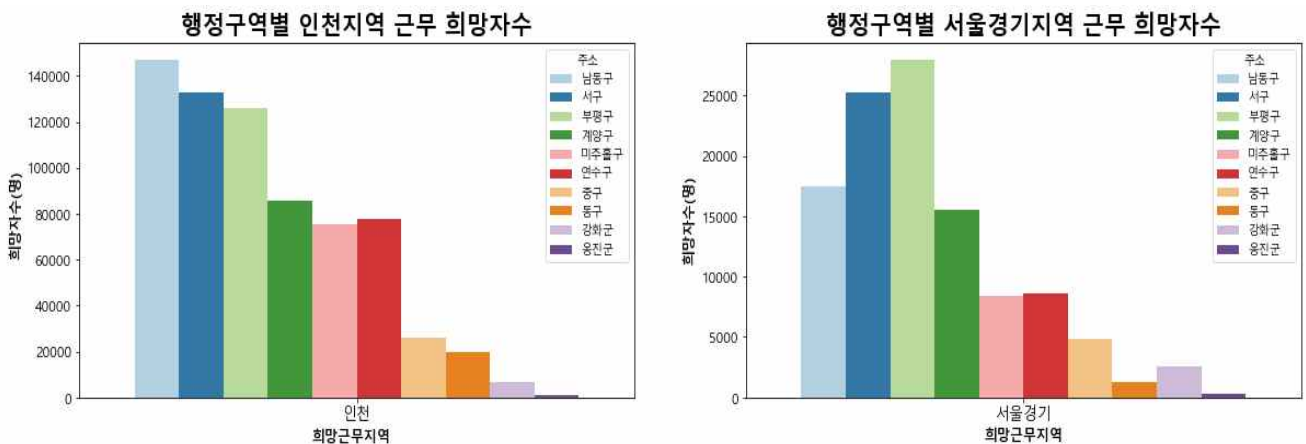
○ 구직자 등록 수가 가장 적은 시·군구는 용진군(1,452명).

5) 구직자 희망근무지역



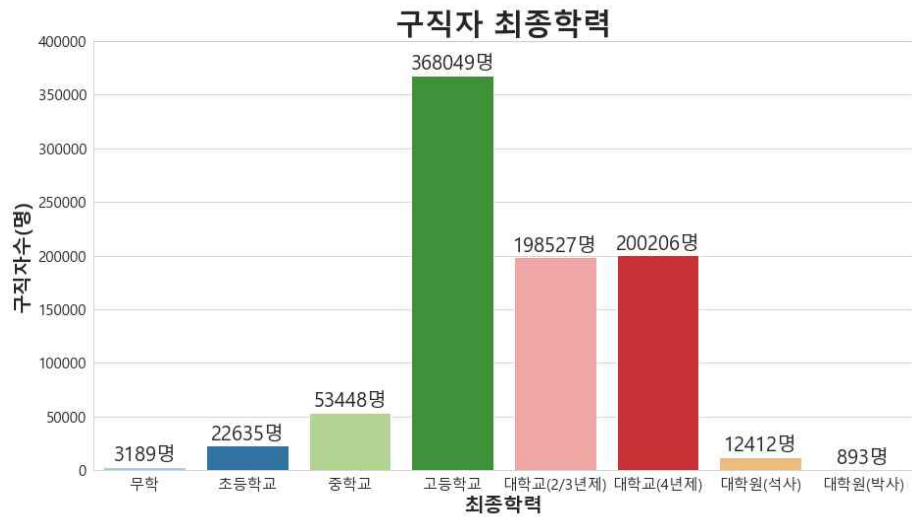
- 인천에 등록된 구직자들은 다른 지역보다 대부분 인천에서 근무하기를 원함.
- 그 다음으로는 서울·경기, 비수도권 순서임.

- 희망근무지역이 인천인 행정구역별 희망자수 와 서울경기인 희망자수



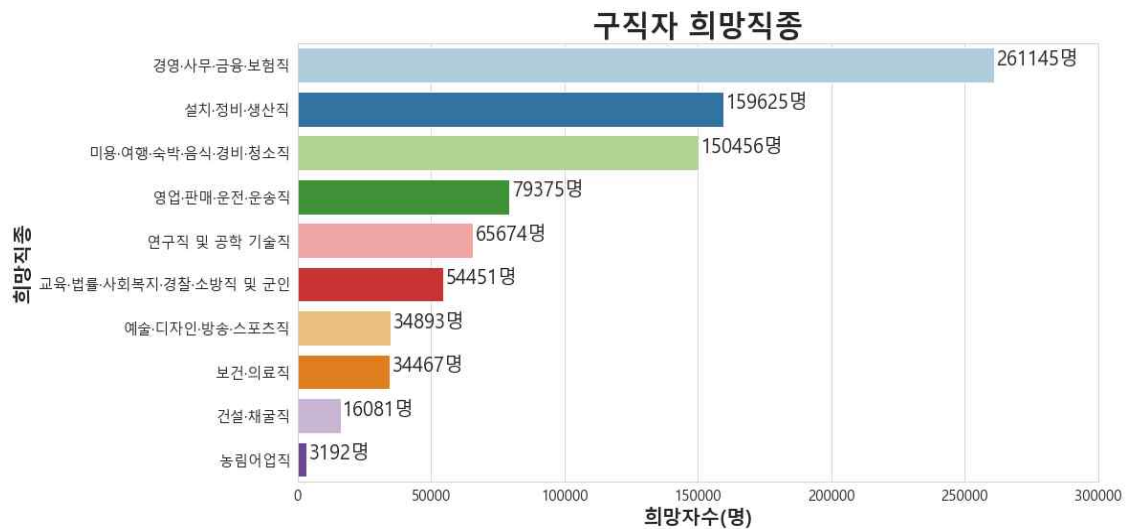
- 인천으로 근무를 희망하는 구직자의 행정구역별 분포는 ‘4) 군구별 구직자 수’의 분포와 비슷한 형태를 보임.
- 서울·경기지역으로 근무를 희망하는 구직자의 행정구역 분포는 인천지역을 희망하는 분포와 달랐으며 부평구와 서구에서 많이 희망함.

6) 구직자 최종학력



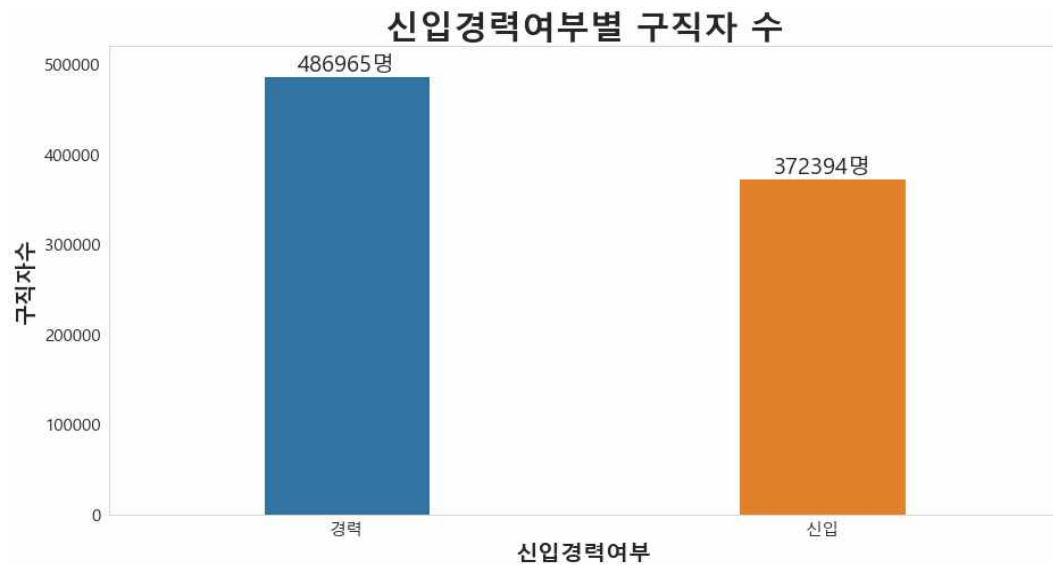
- 워크넷 등록 구직자들의 최종학력을 살펴본 결과, 고등학교 졸업(43%)의 최종학력이 압도적으로 많았고, 그 뒤로 대학교(4년제), 대학교(2/3년제)가 비슷한 수로 나타남.
- 무학과 대학원(박사)는 굉장히 적은 수로 나타남.

7) 구직자 희망직종



- 경영·사무·금융·보험직(30%)을 희망하는 구직자가 가장 많음.
- 농림어업직을 희망하는 구직자는 매우 적음.

8) 구직자 신입경력여부



- 워크넷 등록 구직자들은 신입보다 경력자가 더 많음.
- 경력직의 구직 등록이 많이 한 이유가 직업이나 직장의 만족도가 떨어져 이직률이 높아졌을 거라 생각함.

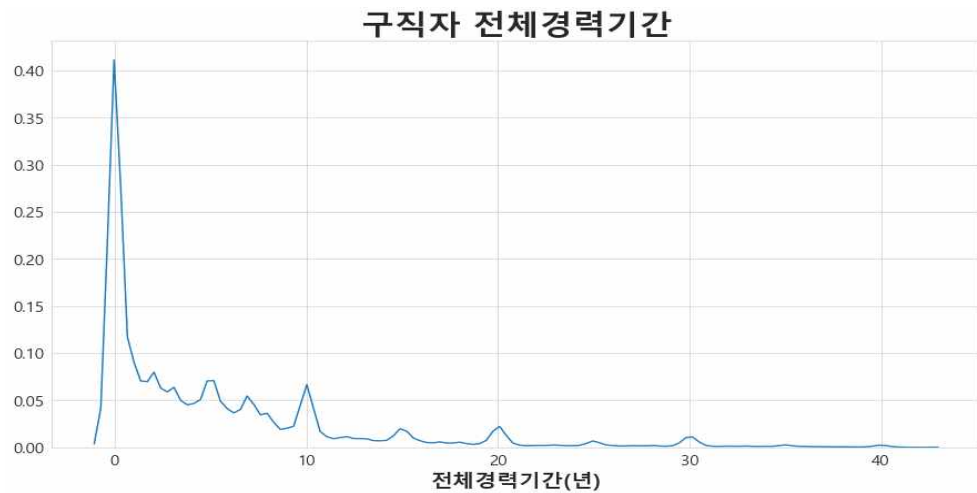
9) 구직자 경력건수



- 경력 건수가 대부분 0건이었고, 경력 건수가 많아질수록 구직자 수가 적어지는 모습임.

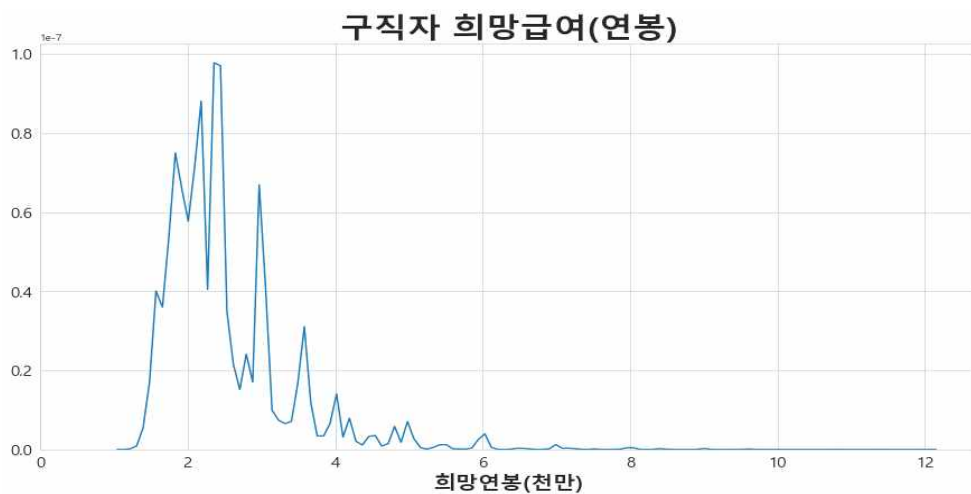
* 신입(372,394)과 경력 0건(312,336) 건수가 맞지 않는 것은 전처리 과정에서 발견하지 못한 차이임.

○ 10) 구직자 전체경력기간



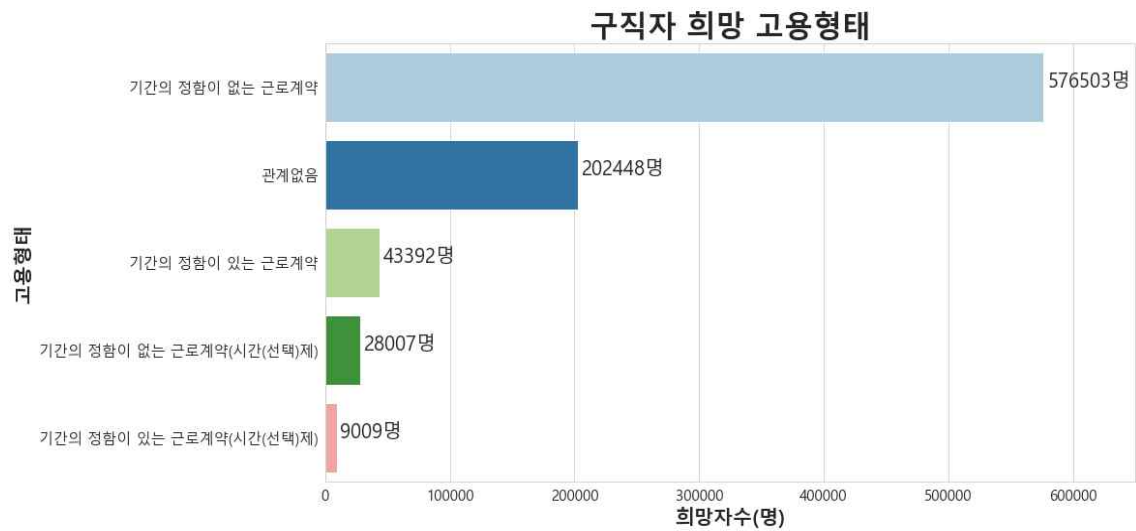
- 데이터가 좌측으로 몰려있는 것을 보면, 경력이 짧은 구직자가 전체적으로 많이 등록되어있다는 것을 알 수 있음.

11) 구직자 희망급여(연봉)



- 구직자의 희망급여(연봉) 값이 왼쪽으로 치우쳐있는 형태를 보이고 있음.
- 워크넷 구직자 데이터는 낮은 급여를 희망하는 구직자가 많고 소수의 고액 급여를 희망하는 구직자가 있음.

12) 구직자 희망 고용형태

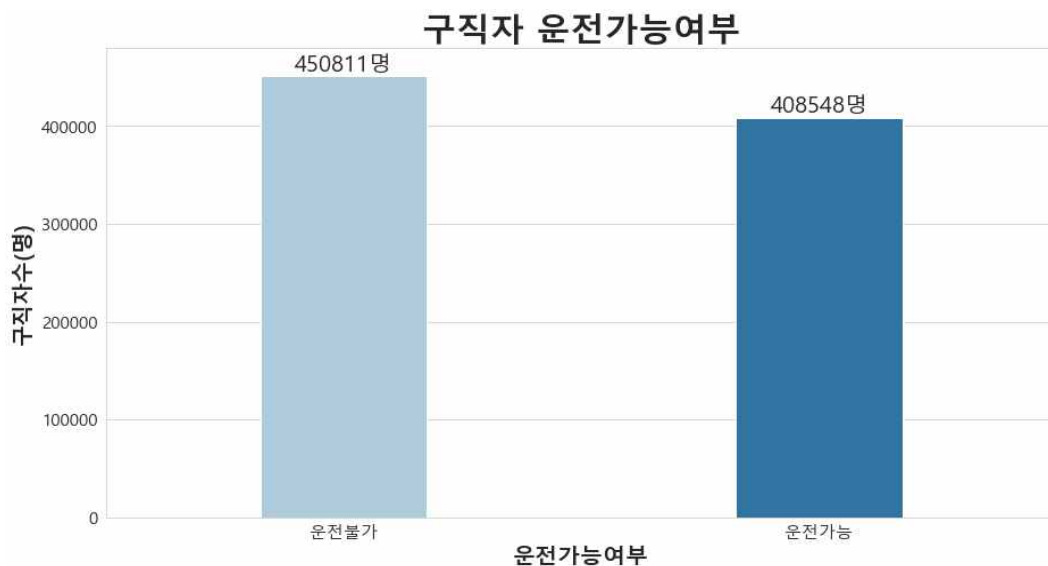


* 기간의 정함이 없는 근로계약 : 정규직 또는 무기계약직

* 기간의 정함이 있는 근로계약 : 계약직

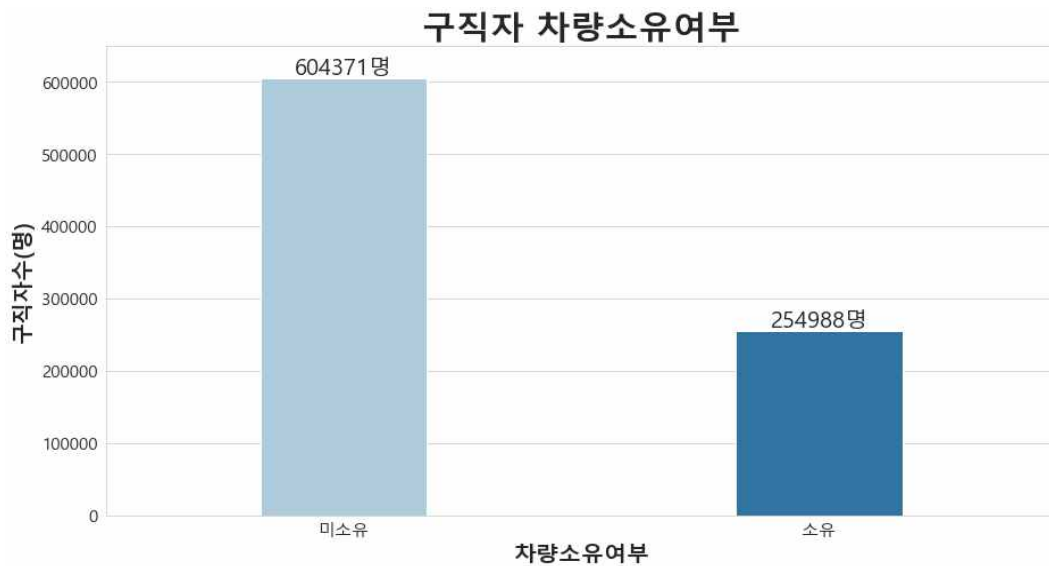
- 워크넷 등록 구직자들의 희망 고용형태는 기간의 정함이 없는 근로계약(576,503명)이 가장 많음.

13) 운전가능여부



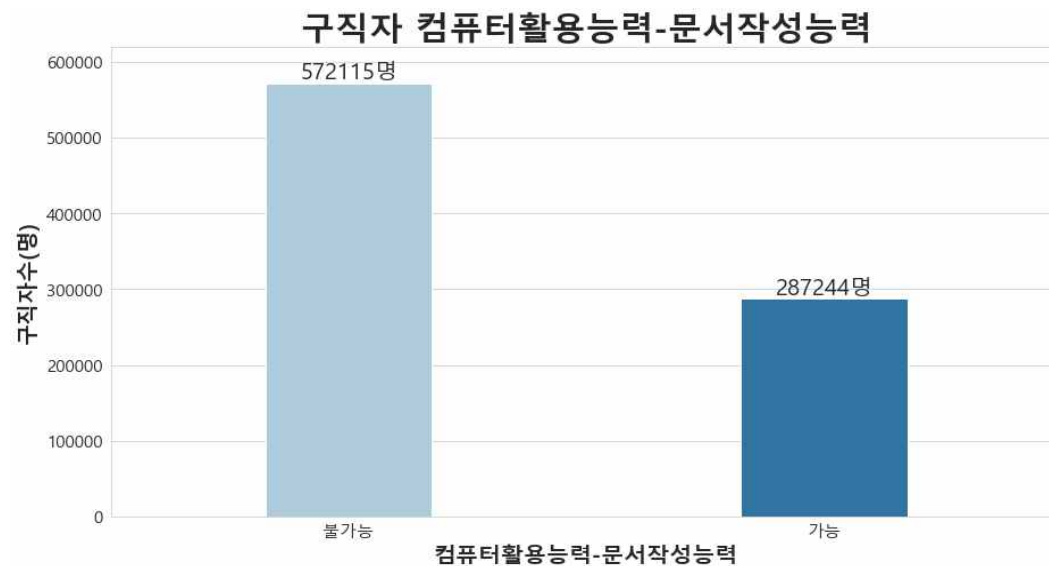
- 운전 가능한 구직자보다 운전이 불가능한 구직자가 더 많음.

14) 구직자 차량소유여부



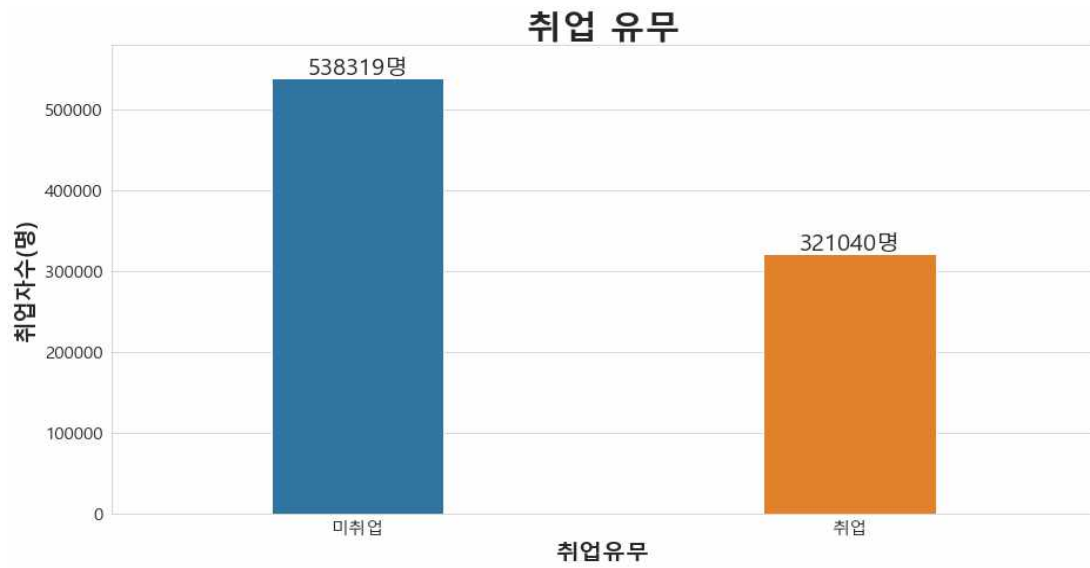
○ 차량을 소유하지 않은 구직자가 소유한 구직자보다 많음.

15) 컴퓨터활용능력-문서작성능력



○ 문서작성능력이 있는 구직자보다 불가능한 구직자가 더 많음.

16) 구직자 취업유무

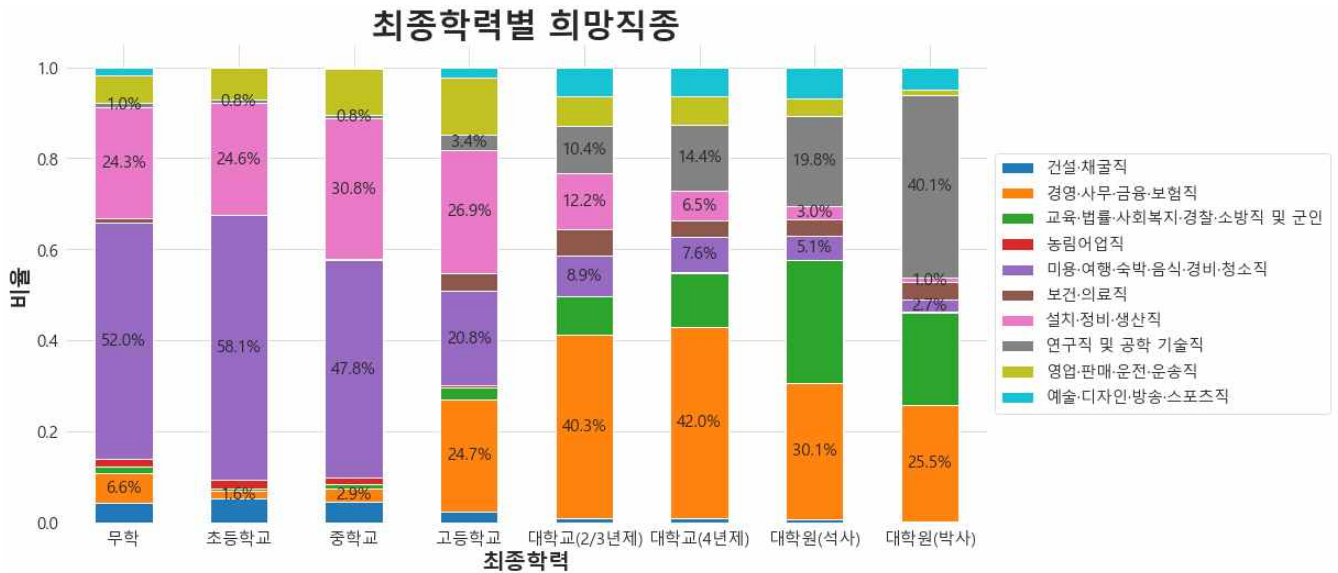


○ 현재 취업자보다 미취업자가 더 많음.

나. 희망직종

구직자 현황 분석에서 구직자의 희망직종과 최종학력 중심으로 쏠림이 있는 것을 확인하고 최종학력별로 희망하는 직종의 차이가 있을 것으로 판단하여 희망직종별로 구체적으로 확인하고자 함.

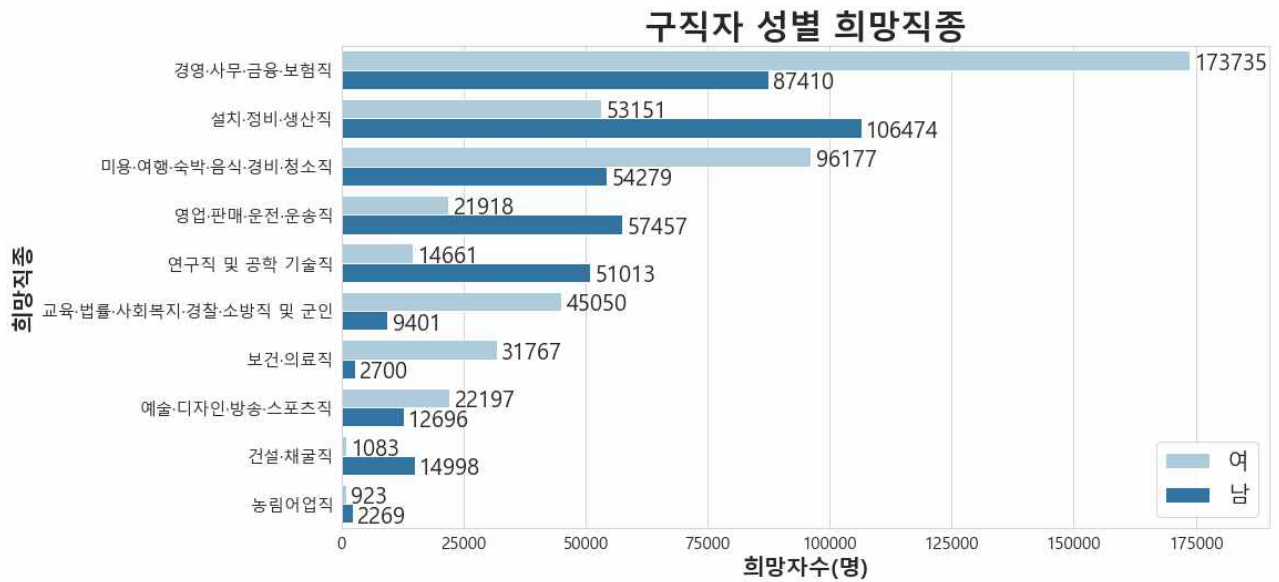
1) 최종학력별 희망직종



- 무학에서 대학교(4년제)까지 경영·사무·금융·보험직의 비율 늘어나다가 대학원(석사, 박사)로 갈수록 경영·사무·금융·보험직의 비율이 줄어듦.
- 고학력이 될수록 연구직 및 공학 기술직의 비율이 늘어남. (1.0% -> 40.1%)
- 저학력(무학, 초졸, 중졸, 고졸)의 경우, 미용·여행·숙박·음식·경비·청소직¹⁾, 설치·정비·생산직의 희망 비율이 높고, 경영·사무·금융·보험직, 연구직 및 공학 기술직의 희망 비율이 적음.
- 고학력으로 갈수록 경영·사무·금융·보험직과 연구직 및 공학 기술직의 희망자가 증가하는 모습을 보임.

1) 추후 제언 2의 근거

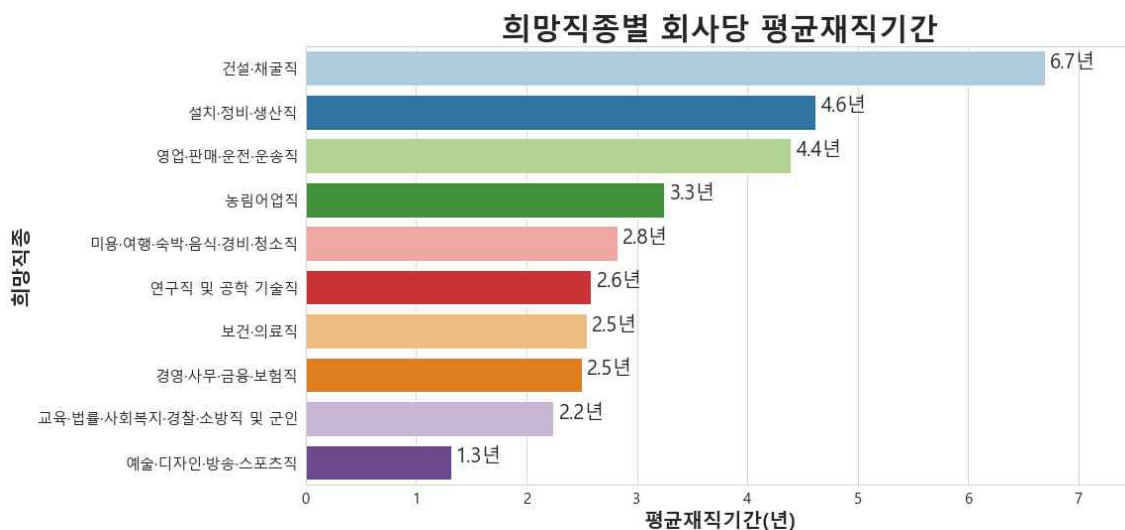
2) 구직자 성별 희망직종



○ 직종별 성별 분포를 보았을 때, 여성은 경영·사무·금융·보험직을 가장 많이 선호하고 남성은 설치·정비·생산직 직종을 가장 많이 선호함.

○ 특히 여성의 경우 경영·사무·금융·보험직(38%)에 몰려 있음²⁾

3) 희망직종별 회사당 평균재직기간³⁾



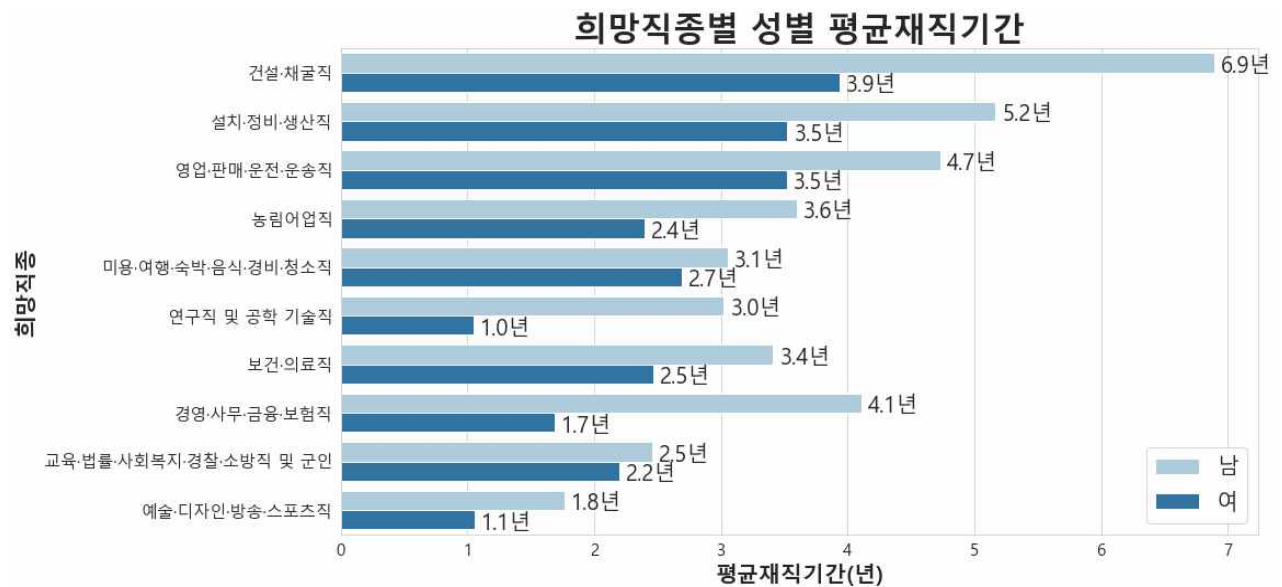
○ 평균재직기간 = $\frac{\text{직종별 전체경력기간}}{\text{경력건수}}$

2) 추후 제언 1의 근거

3) 추후 제언 2의 근거 (미용·여행·숙박·음식·경비·청소직의 평균재직기간이 길지 않다)

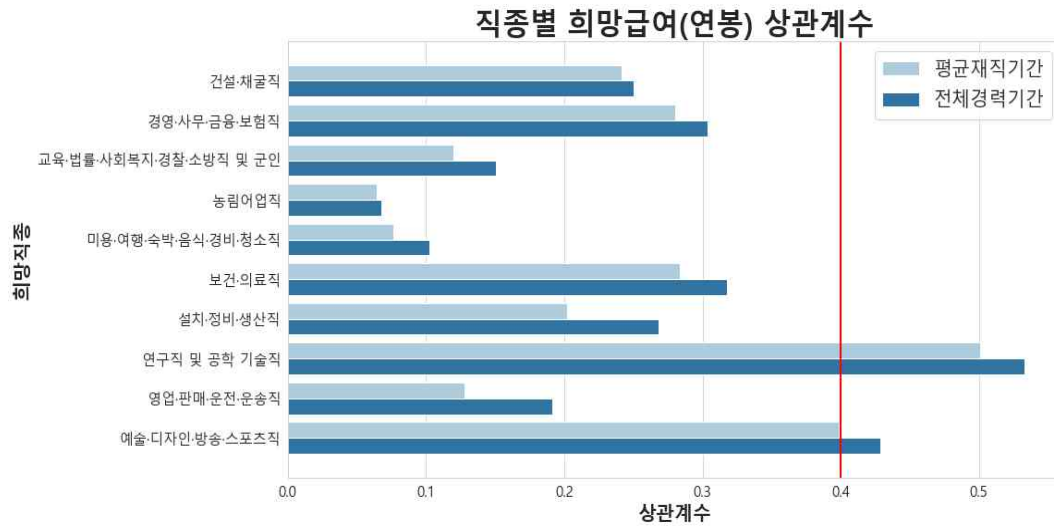
- 직종별 평균재직기간이 가장 길었던 직종은 건설·채굴직.
- 가장 짧은 직종은 예술·디자인·방송·스포츠직.
- 경영·사무·금융·보험직을 희망하는 희망자가 많음에도 평균재직기간은 짧은 편임.

4) 직종별 성별 평균재직기간



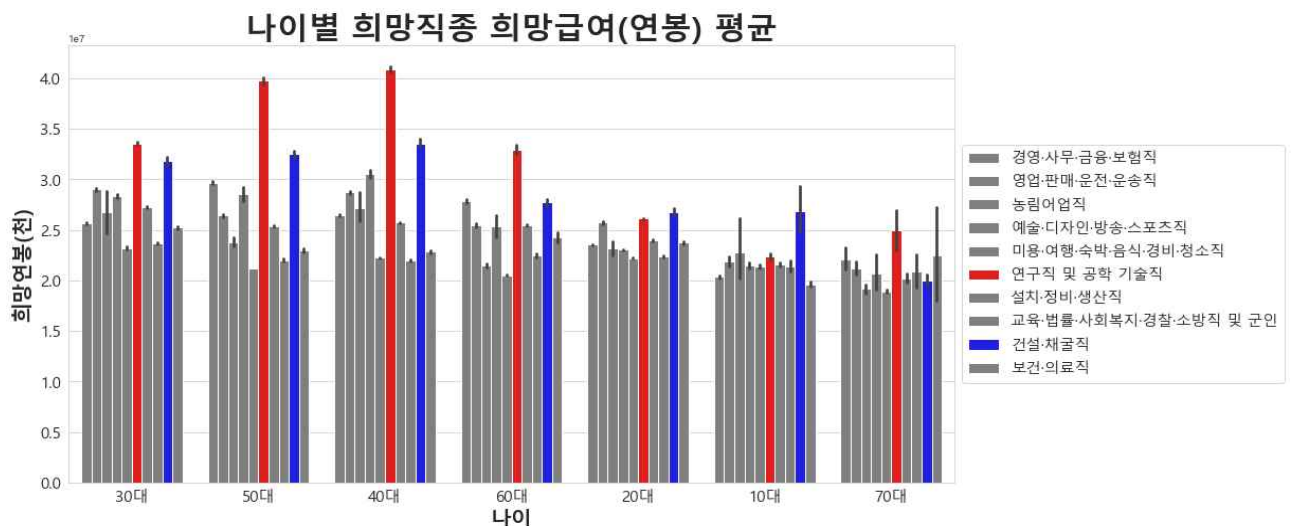
- 모든 직종에서 남성의 평균재직기간이 높음.
- 남성의 평균재직기간이 가장 긴 직종은 건설·채굴직(설치·정비·생산직, 영업·판매·운전·운송직 순서)
- 여성의 경우, 수치의 차이가 있지만 전체적인 경향은 변하지 않음.
- 여성 또한 평균재직기간이 가장 긴 직종은 건설·채굴직(설치·정비·생산직, 영업·판매·운전·운송직 순서)
- 예술·디자인·방송·스포츠직은 가장 짧은 평균재직기간을 보였고, 여성의 경우 연구직 및 공학 기술직 또한 비슷한 기간을 보여줌.

5) 직종별 희망급여(연봉) 상관계수



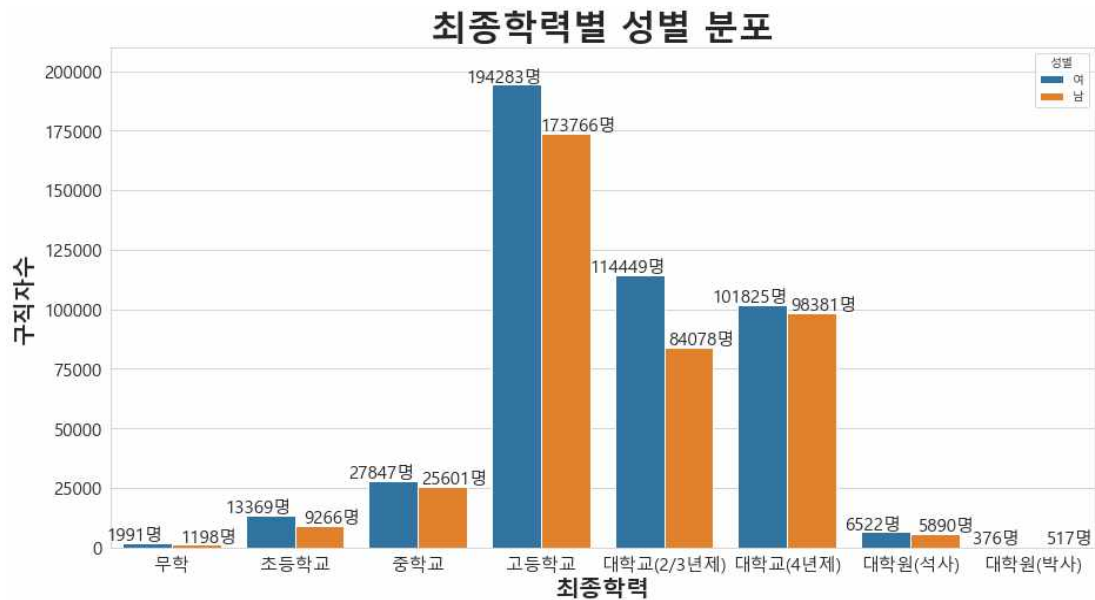
- 대부분의 희망직종은 희망급여(연봉)별 평균재직기간과 전체경력기간과의 상관관계가 없음.
- 예술·디자인·방송·스포츠직은 약간의 상관관계를 보였으며, 연구직 및 공학 기술직도 보통의 상관관계를 보여 평균재직기간(0.5)과 전체경력기간(0.53)이 길어질수록 희망급여(연봉)가 올라가는 경향을 보임.

6) 나이별 희망직종, 희망급여(연봉) 평균



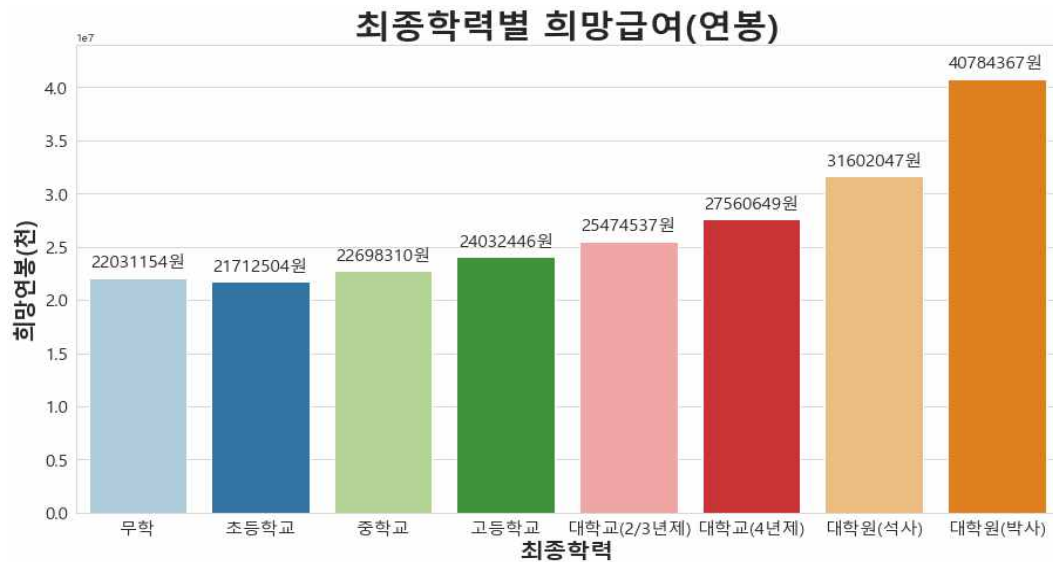
- 가장 눈에 띄는 희망직종으로 연구직 및 공학 기술직이고, 대부분의 나이대에서 높은 희망급여(연봉)을 보임.
- 그 다음으로는 건설·채굴직이 높은 희망급여(연봉)을 원하는 것으로 보임.

7) 최종학력 성별 분포



- 여성은 고등학교 >> 대학교(2/3년제) >> 대학교(4년제) 순으로 분포됨.
- 남성은 고등학교 >> 대학교(4년제) >> 대학교(2/3년제) 순으로 분포됨.

8) 최종학력별 희망급여(연봉)



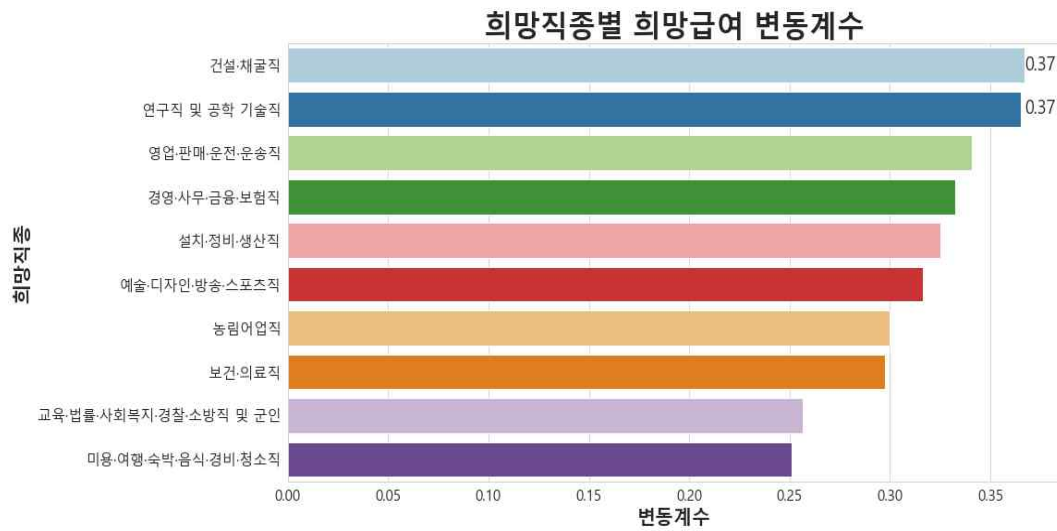
- 대학원(박사)가 희망급여가 가장 높음.
- 무학, 중학교, 초등학교는 희망급여 수준이 비슷함.

9) 최종학력별 희망직종 희망급여(연봉) 평균치

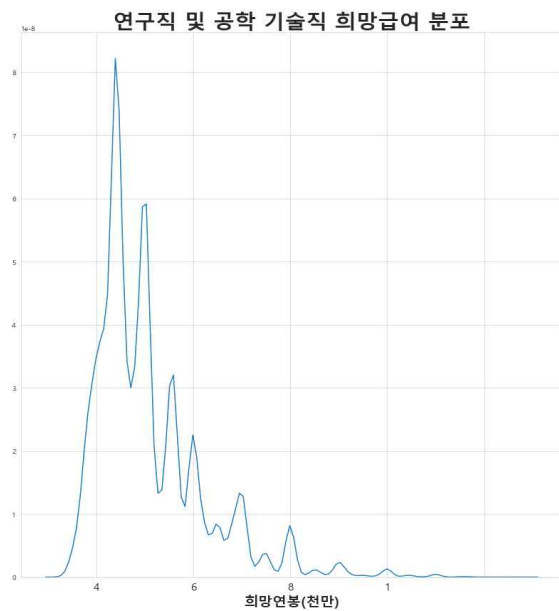
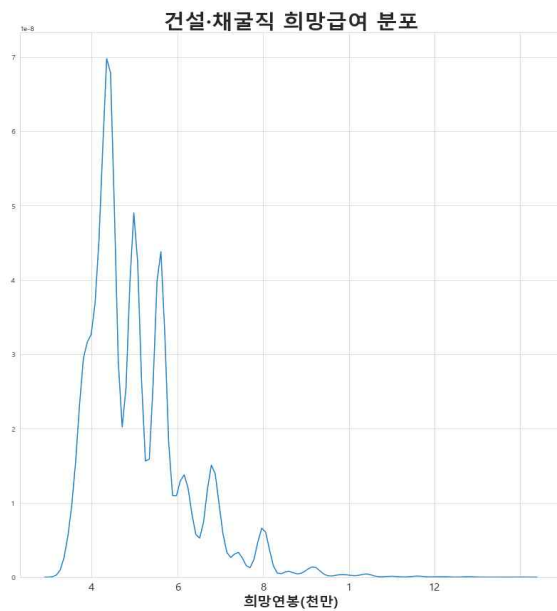


- 무학에서 대학교(4년제)까지 희망급여가 높은 두 직종은 연구직 및 공학 기술직과 건설·채굴직을 뽑을 수 있음.
- 대학원(석사), 대학원(박사)부터는 건설·채굴직의 희망급여가 줄어드는 것을 볼 수 있음.
- 또한 연구직은 꾸준히 높은 희망급여를 보이며, 비교적 희망급여가 낮은 편이었던 경영·사무·금융·보험직이 대학원(석사) 이상에서는 높은 희망급여를 보인다는 것을 알 수 있음.

10) 희망직종별 희망급여 변동계수



○ 희망급여가 가장 높았던 두 직종이 희망급여의 변동성이 다른 직종에 비해서 더 큼.

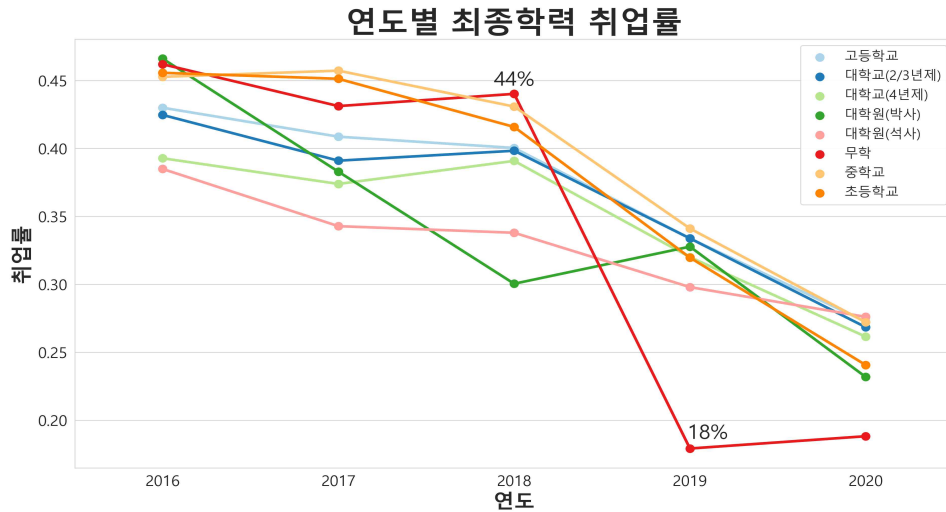


○ 분포도를 확인해보니 매우 높은 희망급여를 원하는 사람들이 있어보임.

○ 전체적으로 다양한 분포를 보임.

다. 최종학력

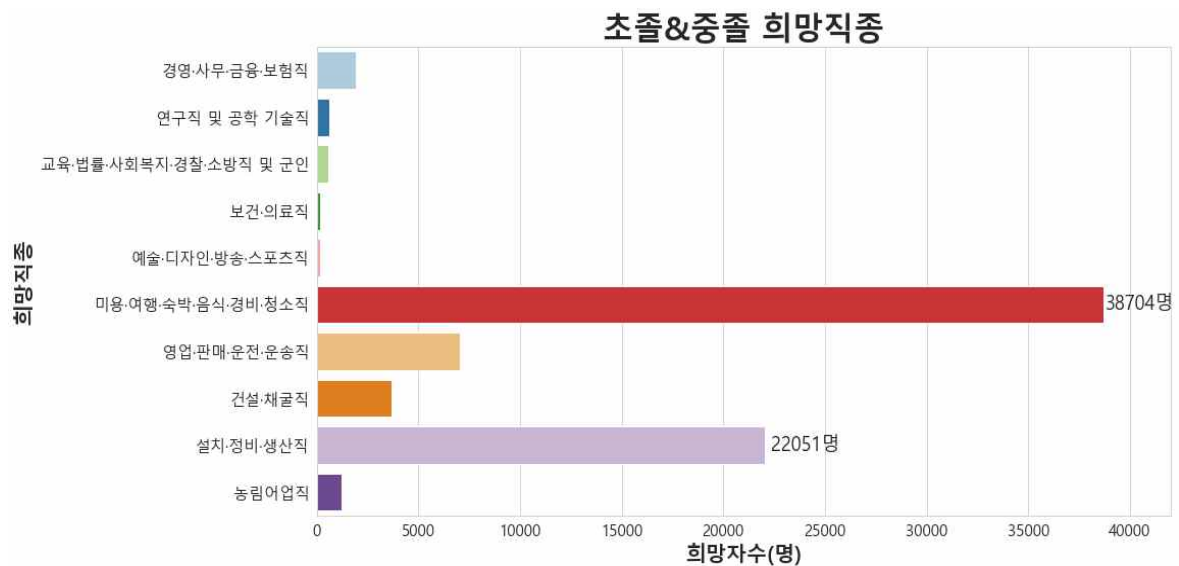
1) 연도별 최종학력 취업률



○ 모든 최종학력의 취업률이 감소하는 모습을 보임.

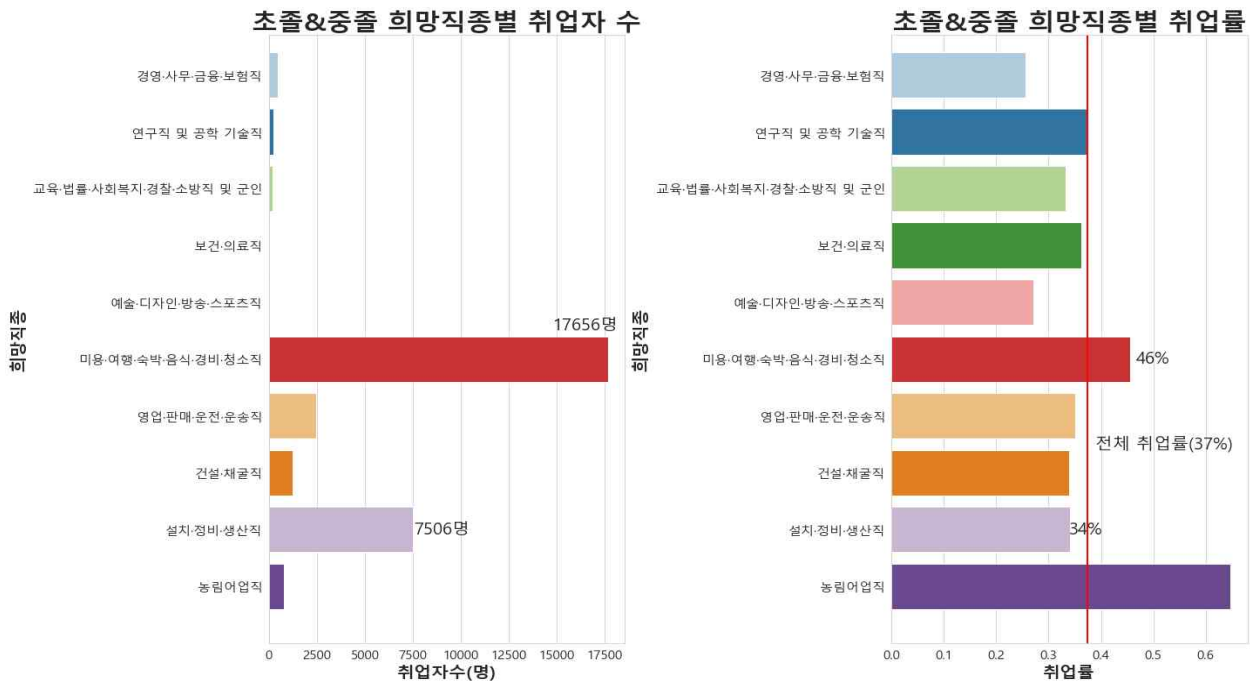
○ 그 중에서 무학의 2019년 취업률 급감이 눈에 띈.

2) 초등학교, 중학교 졸업 구직자의 희망직종



○ 초등학교, 중학교 졸업 구직자들은 미용·여행·숙박·음식·경비·청소직을 압도적으로 많이 희망하였고, 그 다음으로는 설치·정비·생산직임.

3) 초등학교,중학교 졸업 구직자의 희망직종별 취업자 수 & 희망직종별 취업률



○ 미용·여행·숙박·음식·경비·청소직은 희망자 수가 많은 만큼 취업자 수도 많았고, 취업률 또한 평균치를 상회함.

○ 농림어업직은 희망자 수가 현저히 적었지만, 취업률은 평균치를 꽤 상회함.

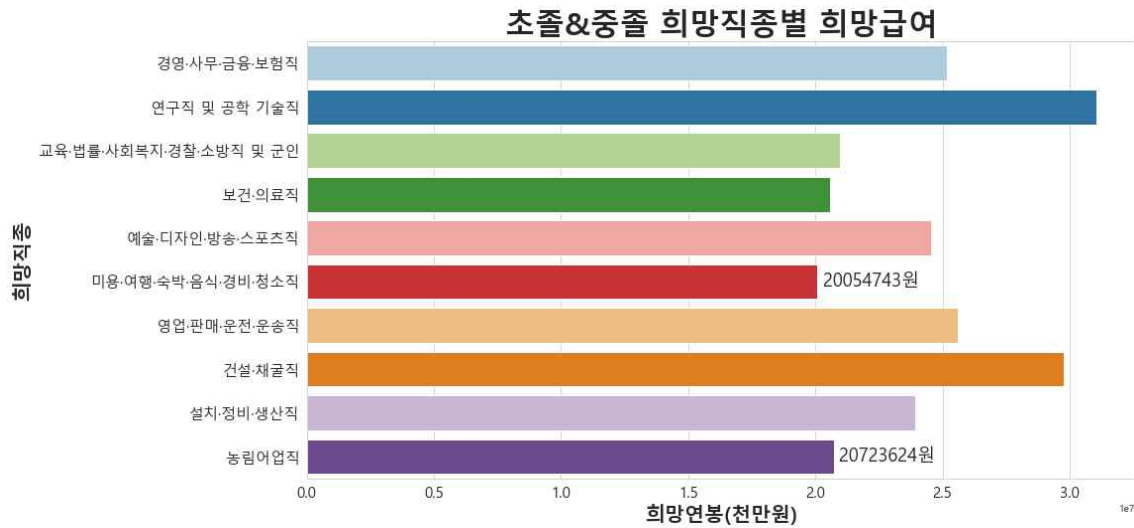
4) 초등학교, 중학교 졸업 구직자의 나이



○ 초등학교, 중학교 졸업의 구직자는 50,60대가 상당부분 차지함.⁴⁾

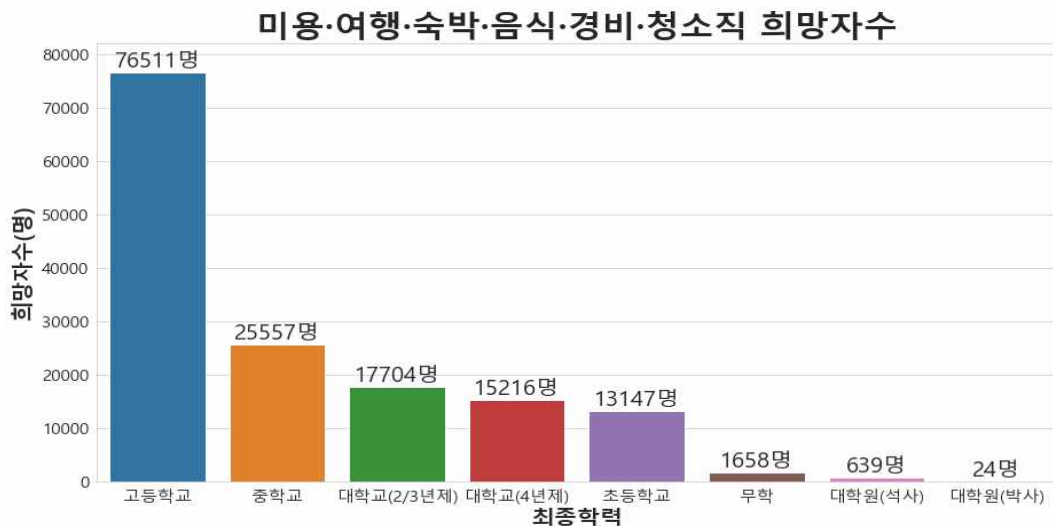
4) 추후 제언2의 근거

5) 초등학교, 중학교 졸업 구직자의 희망직종별 희망급여



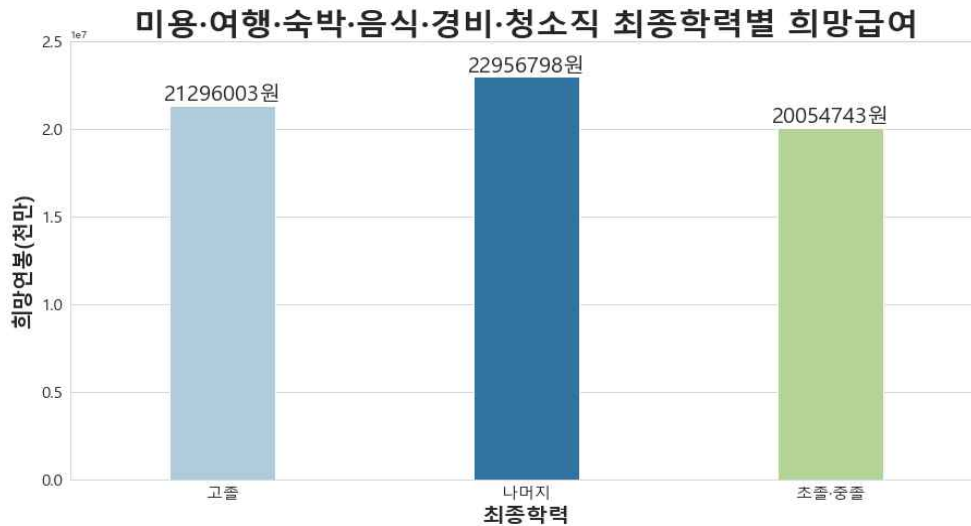
- 희망직종별 희망급여를 보면 희망자 수가 적었던 연구직 및 공학 기술직과 건설·채굴직이 가장 높음.
- 희망자 수가 가장 많고, 취업률도 높았던 미용·여행·숙박·음식·경비·청소직은 희망급여가 가장 낮은 것을 볼 수 있음.

6) 미용·여행·숙박·음식·경비·청소직 최종학력별 희망자 수



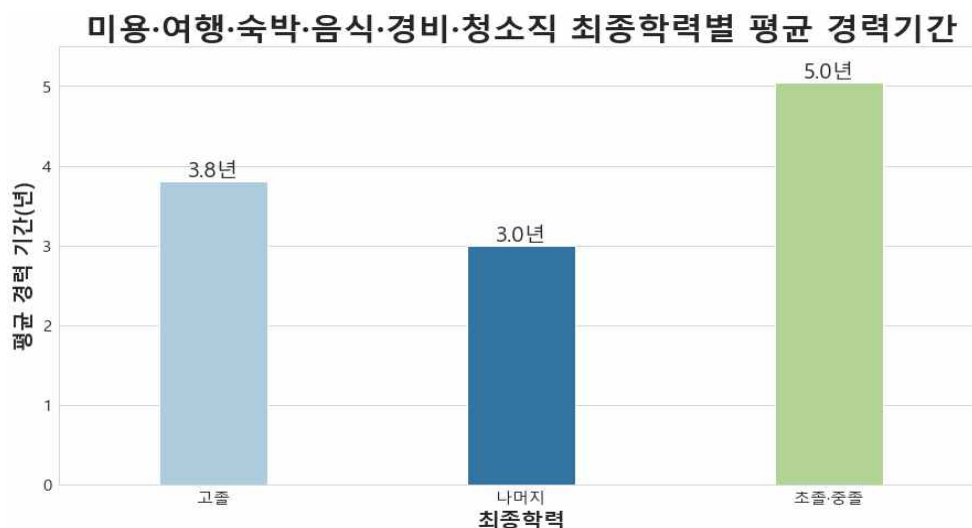
- 초등학교와 중학교 졸업의 구직자가 가장 선호했던 직종이며, 현 직종 안에서 희망자 수는 고등학교 졸업인 구직자가 가장 많은 것을 볼 수 있음.

7) 미용·여행·숙박·음식·경비·청소직 최종학력별 희망급여



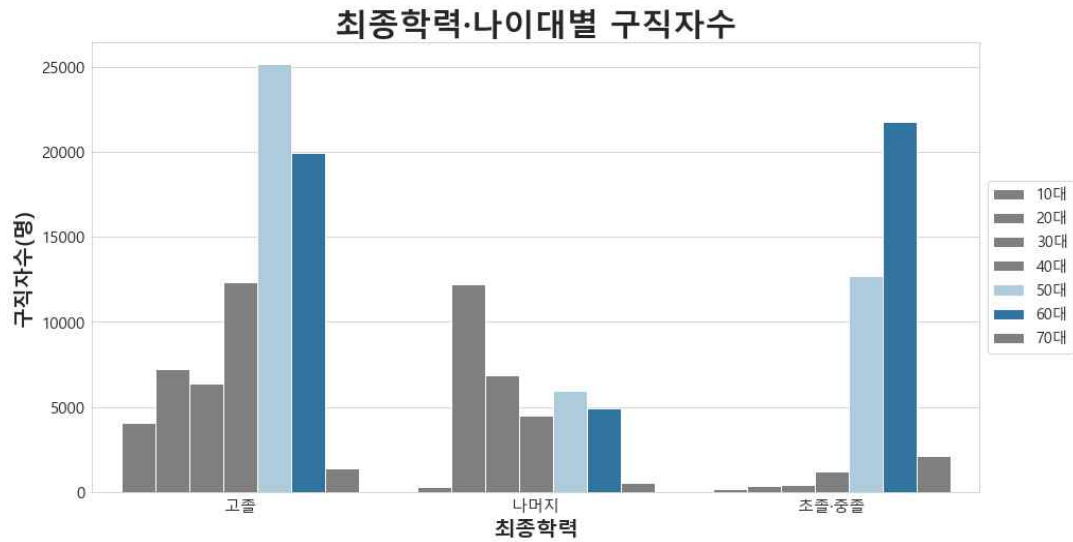
- 미용·여행·숙박·음식·경비·청소직 직종은 초등학교와 중학교 졸업인 구직자들이 가장 희망하는 직종이지만, 다른 최종학력의 구직자에 비해 희망급여가 낮음.
- 최종학력이 높을수록 희망급여가 높아지는 모습을 볼 수 있음.

8) 미용·여행·숙박·음식·경비·청소직 최종학력별 경력기간



- 평균 경력기간으로 보면 최종학력이 낮아질수록 경력기간이 길어지는 것을 알 수 있음.
- 초등학교와 중학교 졸업인 구직자의 경우, 평균 경력기간이 다른 최종학력의 구직자에 비해 높음에도 희망급여가 낮은 모습임.

9) 최종학력과 나이별 구직자 수

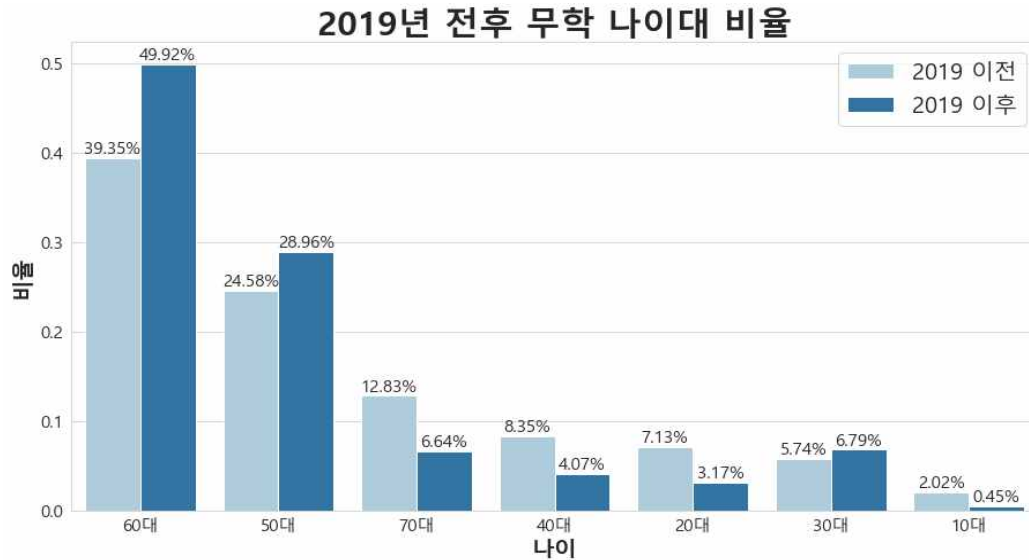


- 초등학교와 중학교 졸업의 구직자들은 50-60대의 비중이 높음. 경력기간이 긴 이유가 50-60대의 비율이 높기 때문인 것으로 보임.
- 50-60대의 비율이 높은 초등학교와 중학교 졸업인 구직자들이 경력도 높고, 희망급여도 낮기 때문에 상대적으로 취업률이 높은 것으로 예상.
- 고등학교 졸업도 50-60대의 구직자 수가 많지만, 초등학교와 중학교 졸업의 나이의 분포보다 다양한 나이의 분포를 보여줌.
- 대학교 졸업 이상의 구직자는 20대가 가장 많음.

10) 무학

무학의 경우 2019년에 취업률이 급감했으므로 한번 살펴볼 필요가 있다고 판단.

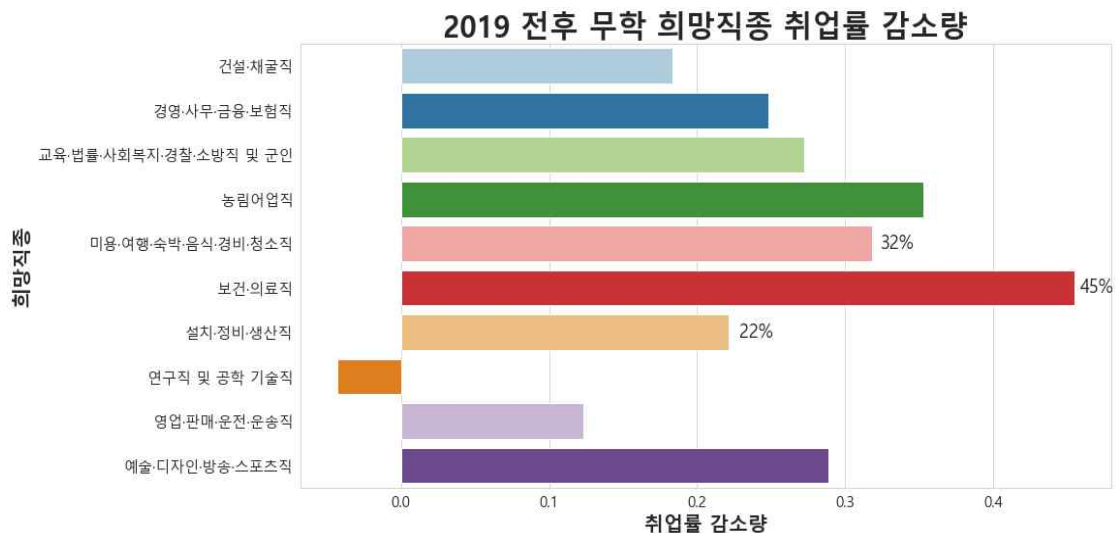
10-가) 2019년 전후로 무학인 구직자들의 나이 비율 비교



○ 2019년 전과 후, 무학인 구직자들의 나이를 파악해보니 2019년 이전에도 50-60대의 비율이 높았지만 2019년 이후로 50-60대와 30대의 구직 등록이 더욱 늘었음.

○ 다른 나이에서는 2019년 이후로 오히려 구직 등록이 줄었음.

10-나) 2019년 전후로 무학인 구직자들의 희망직종별 취업률 감소량



그래프 42. 2019년 전후 무학의 희망직종 취업률 감소량

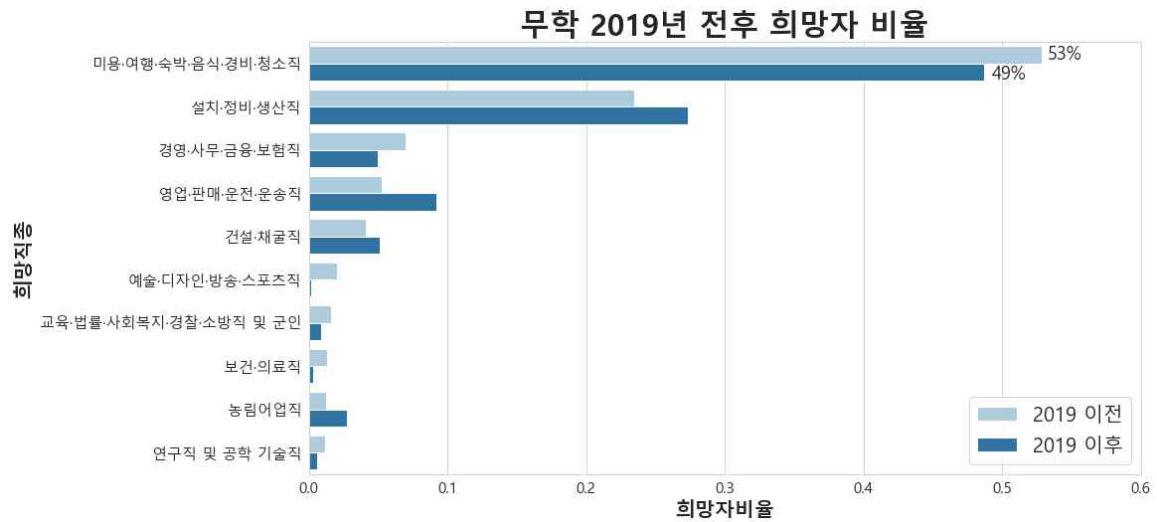
- 2019년 이후, 무학이면서 보건·의료직에 희망했던 구직자들의 취업률이 가장 크게 감소.
- 그 다음으로 농림어업직, 미용·여행·숙박·음식·경비·청소직 순으로 취업률 감소가 큼.

10-다) 무학 구직자들의 희망직종별 취업자 수



- 무학인 구직자 중, 미용·여행·숙박·음식·경비·청소직 희망한 구직자들의 취업률이 가장 높았고, 그 다음으로 설치·정비·생산을 희망한 구직자들의 취업률이 높음.
- 이를 제외한 다른 희망직종에서는 적은 수의 취업자를 보여줌.
- 2019년 이후 취업률 감소가 많았던 보건·의료직과 농림어업직을 희망하던 취업자 수가 위 그래프와 같이 현저히 적었기 때문에 2019년 이후 전체 취업률이 급감한 이유는 아닌 것 같다고 판단함.

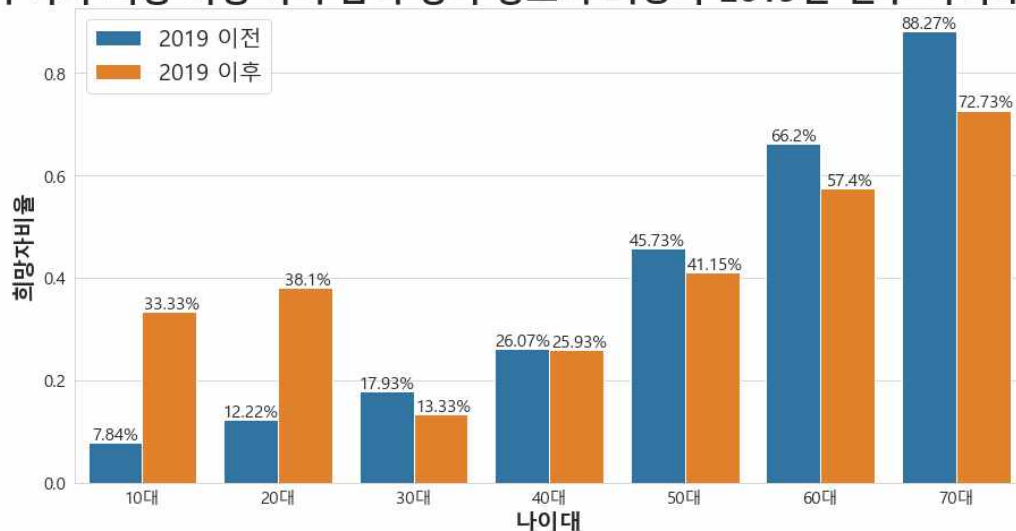
다음으로 무학의 희망직종 중, 취업률 감소량이 많았던 미용·여행·숙박·음식·경비·청소직을 살펴보겠다.



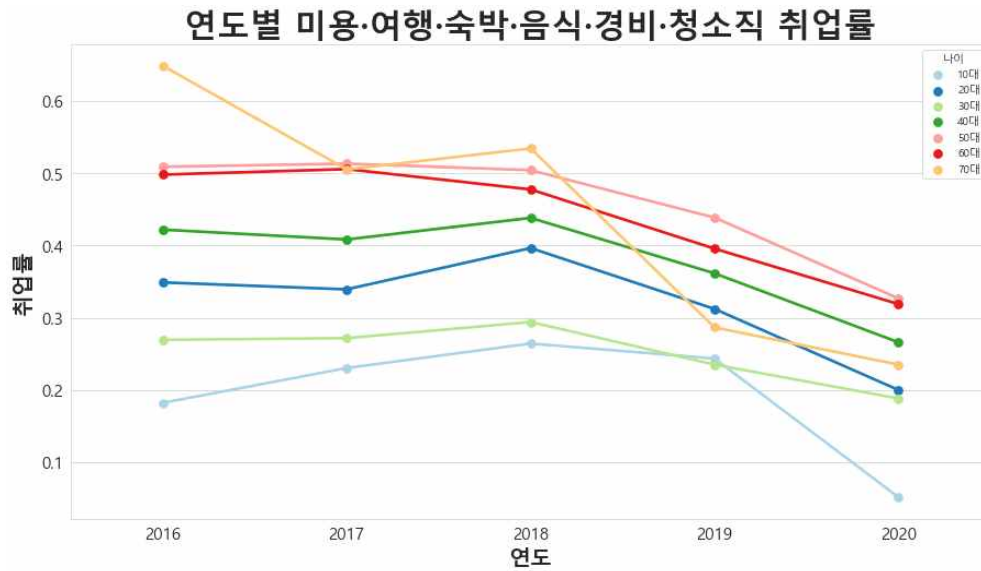
그래프 87. 무학 2019년 전후 희망자 비율

- 2019년 이후로 미용·여행·숙박·음식·경비·청소직을 희망하는 비율은 조금 줄었지만, 큰 차이는 안보임.
- 하지만, 그래프 42. 보면 취업률은 크게 감소.
- 반대로 설치·정비·생산직과 영업·판매·운전·운송직 그리고 농림어업직의 희망자 비율은 늘었음.

무학자 미용·여행·숙박·음식·경비·청소직 희망자 2019년 전후 나이대 비율



- 2019년 이후, 10-20대 구직자들의 해당 직종 희망 비율이 높아지고, 30대 이상부터는 희망 비율이 감소.



그래프 89. 연도별 전체 구직자의 미용·여행·숙박·음식·경비·청소직 나이별 취업률

- 무학인 구직자 중에서 미용·여행·숙박·음식·경비·청소직을 희망하는 나이대의 비율 변화가 있었지만 전체 최종학력의 미용·여행·숙박·음식·경비·청소직 나이대의 취업률 순위는 거의 변화가 없음.
- 무학의 경우 나이대의 비율 변화가 있었지만 나이대별 취업률 순위 변화가 없었기 때문에 취업률이 감소한 것으로 예상됨.

11) 고등학교

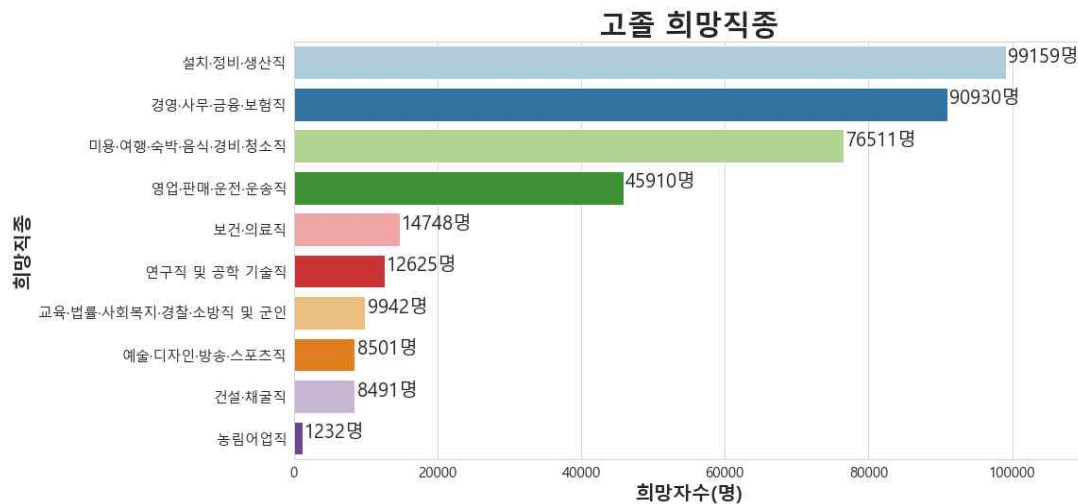
11-가) 고등학교 졸업 구직자의 나이



○ 고등학교 졸업 구직자의 경우, 전체 구직자의 나이 분포 그래프 7. 와 다른 모습임.

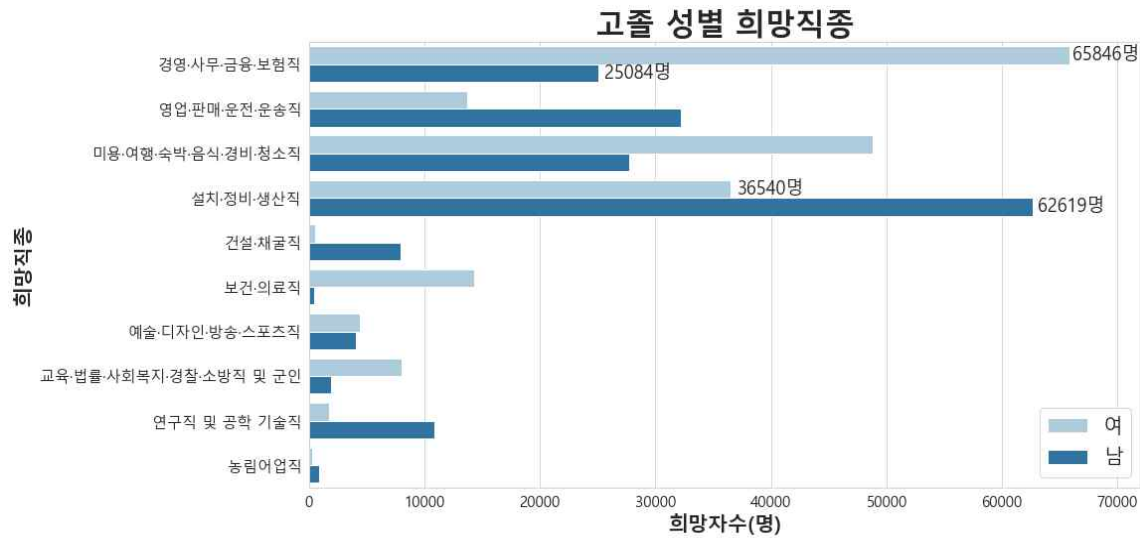
○ 40-50대가 가장 많고, 30대 >> 20대 >> 60대 >> 10대 >> 70대 순서임.

11-나) 고등학교 졸업 구직자의 희망직종



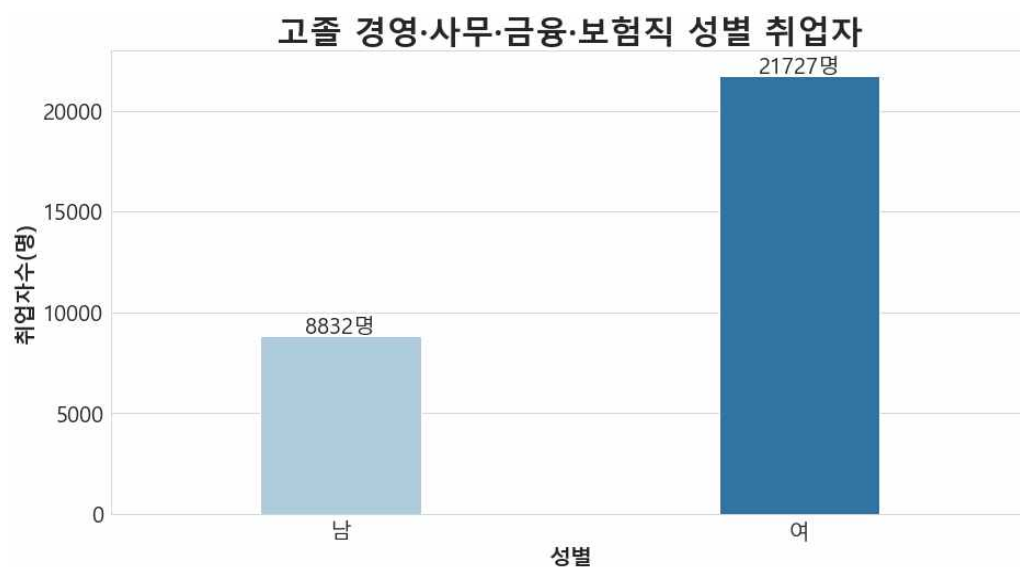
○ 고등학교 졸업인 구직자의 경우, 가장 희망하는 직종으로 설치·정비·생산직이 가장 많았고, 그 다음으로 경영·사무·금융·보험직이 많음.

11-다) 고등학교 졸업 구직자의 성별 희망직종



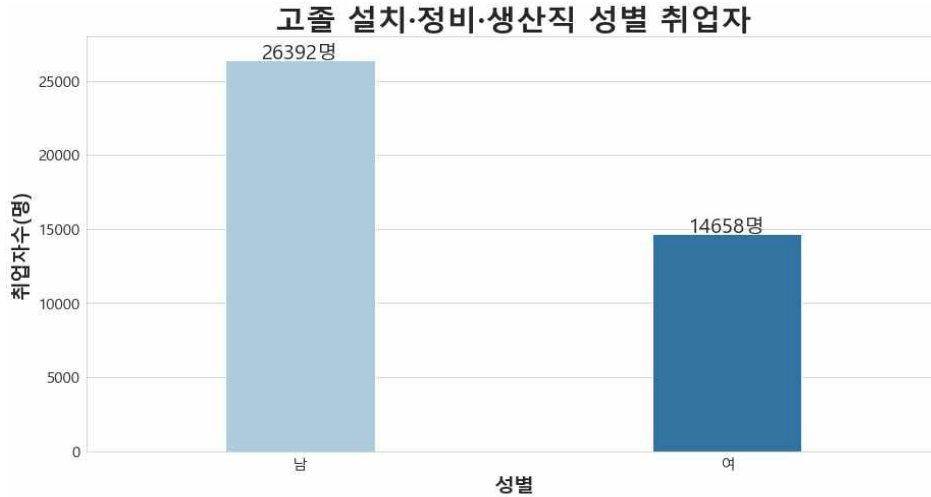
- 고등학교 졸업인 구직자가 가장 많이 희망했던 경영·사무·금융·보험직과 설치·정비·생산직에서 남녀 비율의 차이가 보임.
- 경영·사무·금융·보험직의 경우 여성이 많은 비율을 보였고, 설치·정비·생산직의 경우 남성의 비율이 더 많음.

11-라) 고등학교 졸업 구직자의 경영·사무·금융·보험직 성별 취업자 수



- 여성의 희망 비율이 더 많았던 경영·사무·금융·보험직 직종에서 취업자의 비율도 여성이 많았음.

11-마) 고등학교 졸업 구직자의 설치·정비·생산직 성별 취업자 수



그래프 49. 고등학교 졸업 설치·정비·생산직 희망자 성별 취업자수

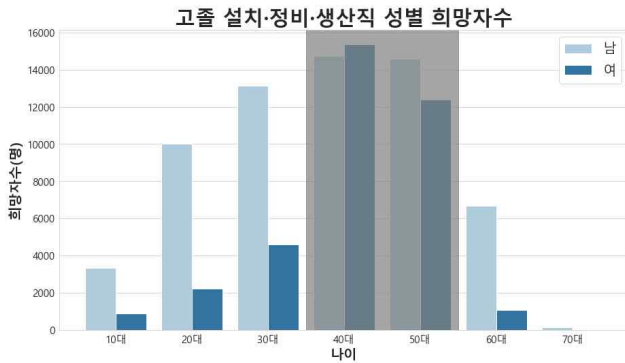
- 남성의 희망 비율이 더 많았던 설치·정비·생산직 직종에서 취업자의 비율도 남성이 더 많았음.
- 해당 희망직종 내에서 성별에 따라 희망하는 수가 많은 만큼 취업자 수도 차이가 있음.

11-바) 고등학교 졸업 희망직종 TOP 2 나이

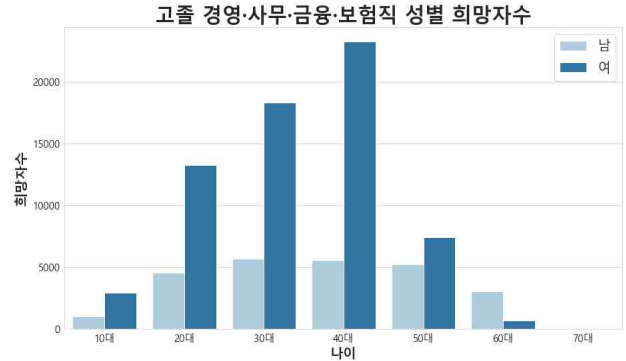


- 남녀 성비가 크게 나뉘었던 두 희망직종의 나이대를 확인해보면,
경영·사무·금융·보험직은 40대가 가장 많고, 다음은 30대>>20대 >>50대 순.
- 설치·정비·생산직도 40대가 가장 많고, 그 다음은 50대>>30대>>20대 순임.
- 두 희망직종 간 희망하는 구직자의 나이대 차이가 보임.

11-사) 고등학교 졸업 구직자의 설치·정비·생산직 성별 희망자 수와
경영·사무·금융·보험직 성별 희망자 수 비교



그래프 51. 고등학교 졸업 구직자 중 설치·정비·생산직 성별 나이별 희망자 수

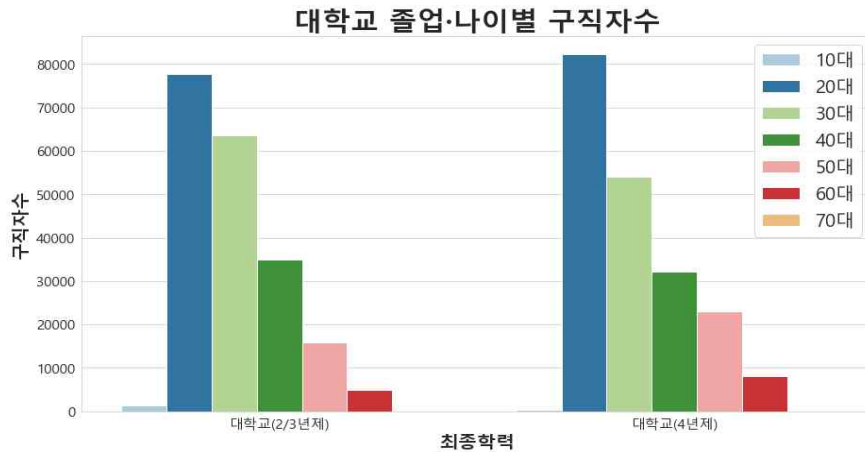


그래프 52. 고등학교 졸업 구직자 중, 경영·사무·금융·보험직 성별, 나이별 희망자 수

- 그래프 51. 와 그래프 52.을 비교 하였을 때,
설치·정비·생산직의 경우 40대를 제외하고는 다양한 나이에 성별의 비중이 많았고, 경영·사무·금융·보험직의 경우 10-50대의 여성 비중이 많음.
- 그래프 49.에서 보듯 설치·정비·생산직에서 남녀 희망자 수의 차이가 뚜렷하게 보였지만, 나이대별로 확인했을 때 40-50대에서는 성별 차이없이 많은 희망자를 보여주며 설치·정비·생산직의 전체 연령대 중 40-50대의 비중을 높인 것으로 보임.
- 경영·사무·금융·보험직의 경우 희망자의 성별 차이가 꾸준히 보이고, 특히 20-40대에서 성별 차이가 심함. 다만 50대에서는 여성의 희망 수가 줄었기 때문에 경영·사무·금융·보험직은 설치·정비·생산직에 비해 20-30대의 희망 비율이 높은 것으로 보임.

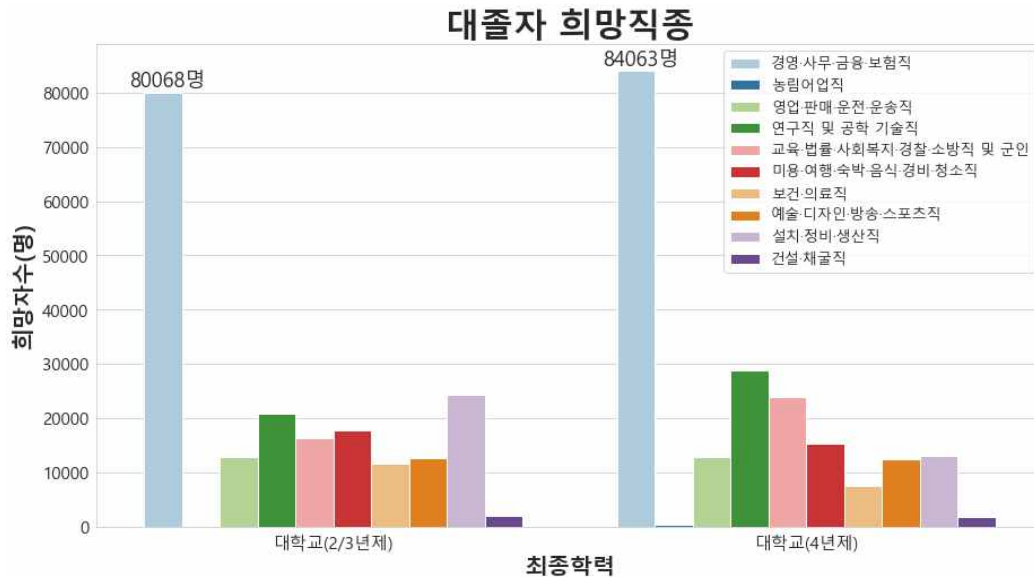
12) 대학교

12-가) 대학교 졸업 구직자의 나이



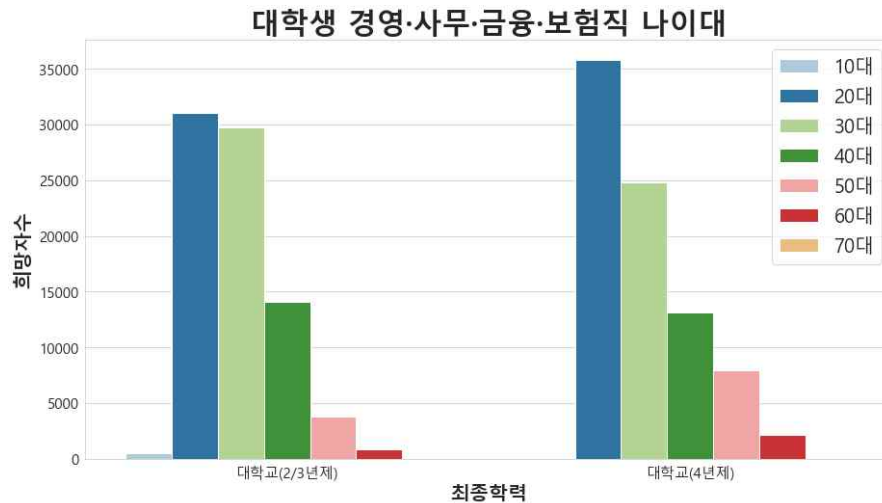
- 구직자의 최종학력이 대학교(2/3년제)와 대학교(4년제) 졸업 모두 나이대별 비슷한 비율을 보임.

12-나) 대학교 졸업 구직자의 희망직종



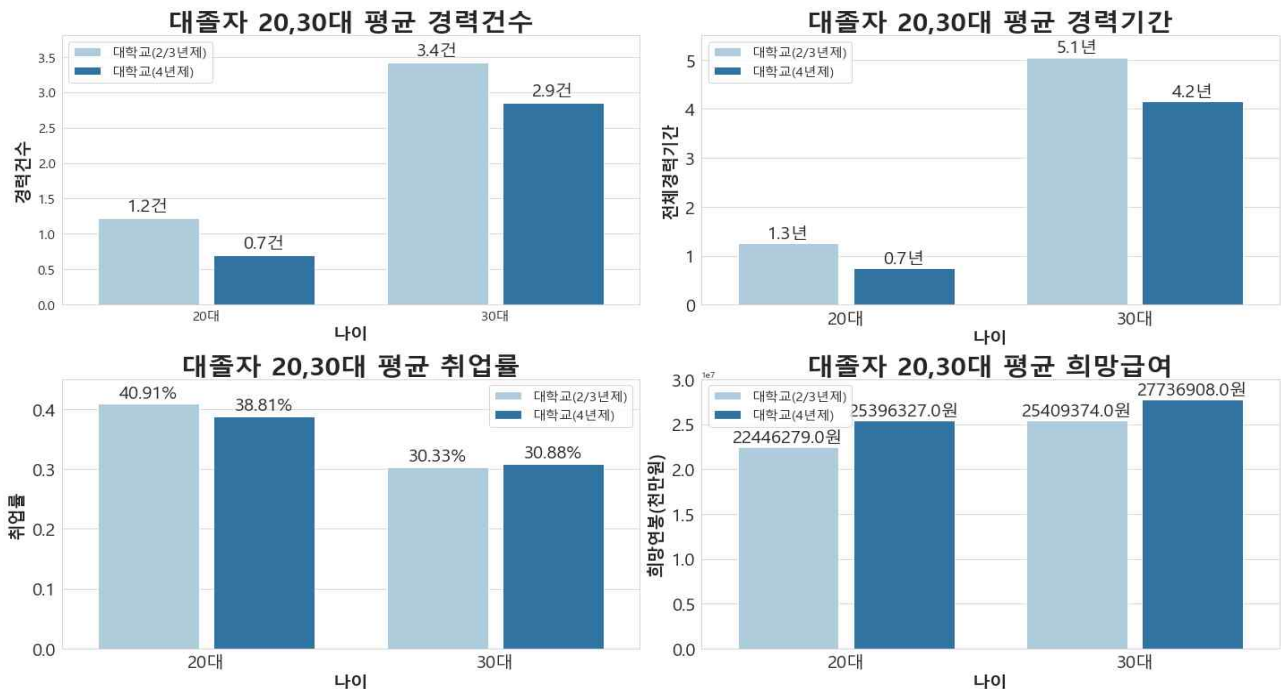
- 최종학력이 대학교(2/3년제)와 대학교(4년제)의 희망직종은 약간 다른 모습을 보임.
- 두 최종학력 모두 가장 희망하는 직종으로 경영·사무·금융·보험직을 선택.
- 다만, 대학교(2/3년제)는 설치·정비·생산직이 두 번째로 많은 희망자 수를 보이고, 대학교(4년제)는 연구직 및 공학 기술직이 희망자 수가 두 번째로 많은 것을 볼 수 있음.

12-다) 대학교 졸업 구직자의 경영·사무·금융·보험직 나이



- 대학교(2/3년제)와 대학교(4년제) 졸업의 구직자들이 가장 많이 희망했던 경영·사무·금융·보험직의 나이대를 살펴보면, 20대가 가장 많고 그 다음으로 30대가 많은 비율을 차지함.

12-라) 대학교 졸업 20-30대 평균 경력건수&경력기간&취업률&희망급여



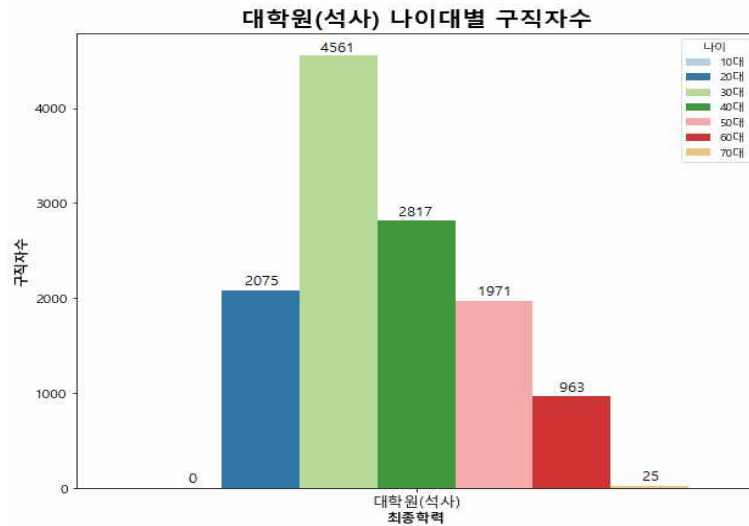
- 대학교 졸업인 구직자의 희망직종으로 가장 높았던 경영·사무·금융·보험직의 20-30대의 경력건수와 경력기간을 살펴보았을 때, 모두 대학교(2/3년제)가

높음. 대학교(4년제) 졸업인 구직자의 비해 비교적 빠른 졸업으로 인한 결과인 것으로 봄.

- 평균 취업률은 대학교(2/3년제) 졸업 구직자들이 20대에서는 조금 더 높았지만, 30대에는 대학교(4년제)가 약간 더 높음.
- 평균 희망급여는 20-30대에서 모두 대학교(4년제)가 높음.

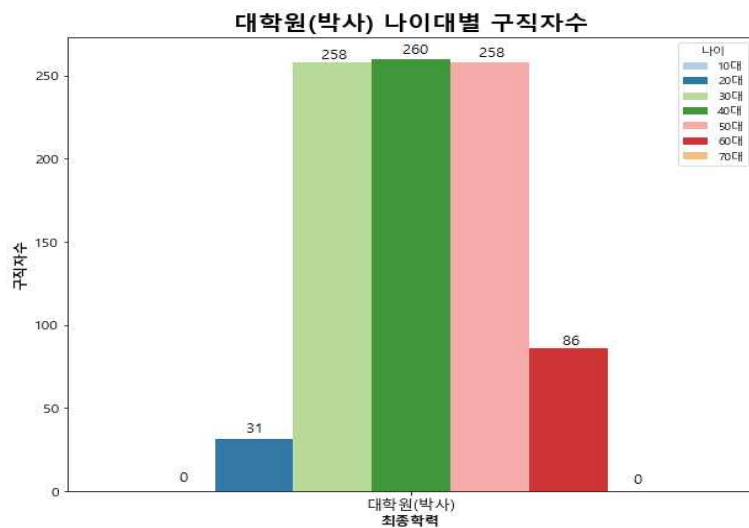
13) 대학원

13-가) 대학원(석사) 졸업 구직자 나이



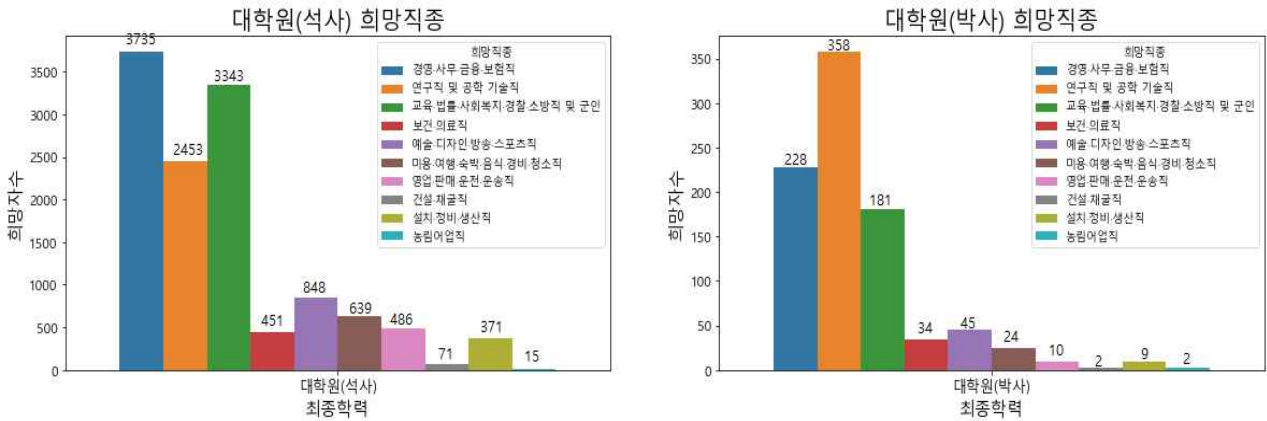
○ 대학원(석사) 졸업 구직자들은 상당 부분 30대였고, 그 다음으로 40대 >> 20대 >> 50대 >> 60대 순임.

13-나) 대학원(박사) 졸업 구직자 나이

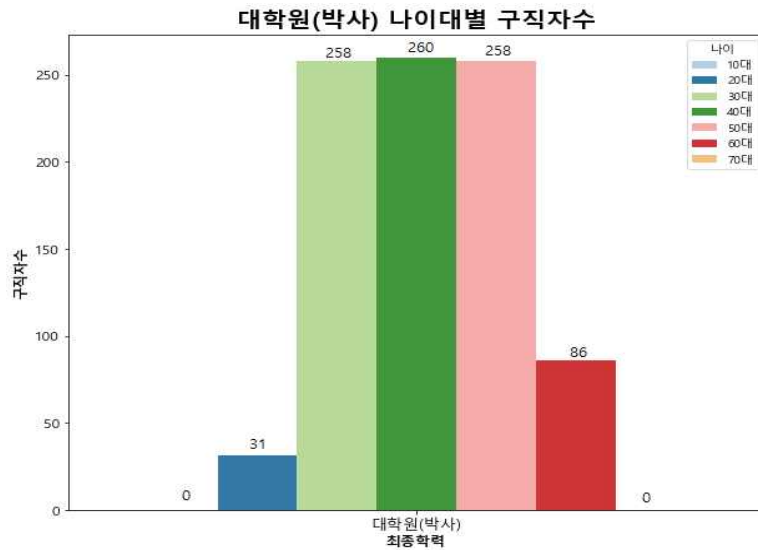


○ 대학원(박사) 졸업 구직자 수는 30-50대가 고르게 퍼져있고, 그 다음으로 60 >> 20대 순임.

13-다) 대학원 졸업 구직자의 희망직종

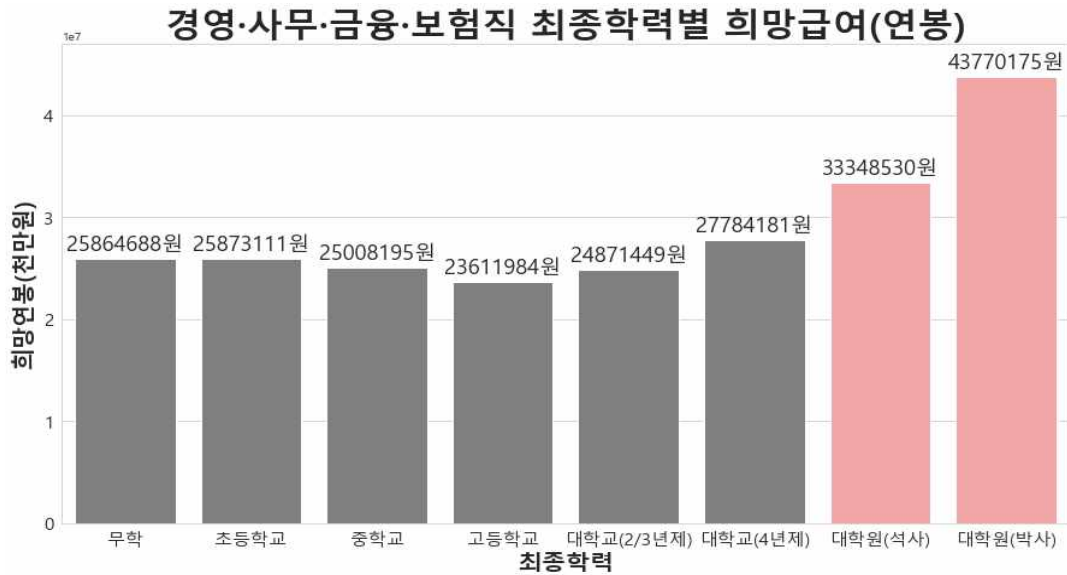


○ 대학원(석사)와 대학원(박사) 졸업 구직자의 희망직종 중 가장 희망자 수가 많은 직종 3가지 내에서 다르게 나타남.

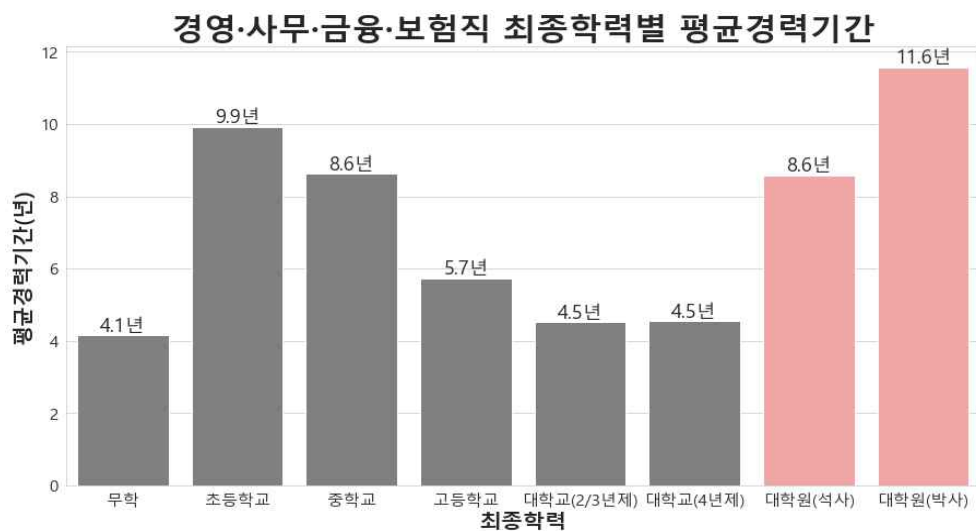


○ 대학원(석사)는 경영·사무·금융·보험직 >> 보건·의료직 >> 연구직 및 공학 기술직 순이고, 대학원(박사)는 연구직 및 공학 기술직 >> 경영·사무·금융·보험직 >> 보건·의료직 순임.

13-라) 경영·사무·금융·보험직 최종학력별 경향

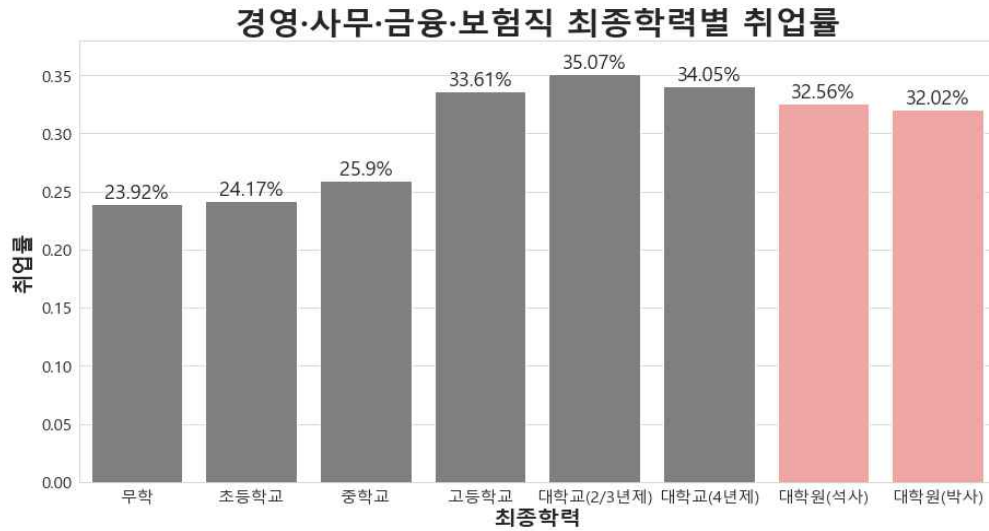


- 대학원(석사) 졸업인 구직자가 경영·사무·금융·보험직에 가장 많이 희망했으나 희망급여는 대학원(박사)가 더 높음.
- 대학원 졸업인 구직자를 제외한 다른 최종학력에서는 희망급여가 큰 차이가 없음.
- 무학부터 고등학교 졸업인 구직자들은 대학교 졸업 구직자들에 비해 경력이 길고, 대학교 졸업 구직자들은 희망급여가 더 높기 때문에 비슷한 희망급여를 보이는 것 같음.



그래프 107. 경영·사무·금융·보험직 최종학력별 평균경력기간

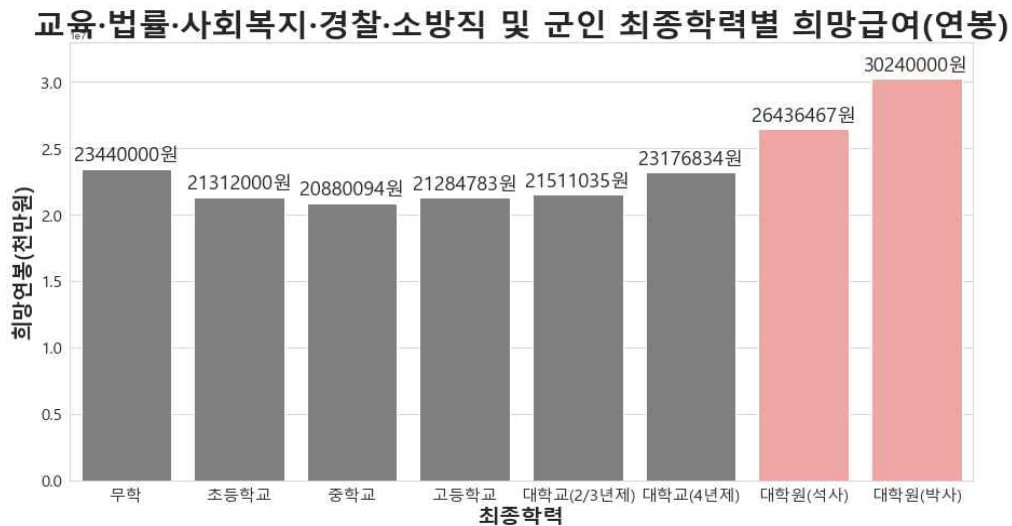
- 대학원(박사)가 경력기간이 제일 길고, 그 다음으로는 초등학교 졸업임.



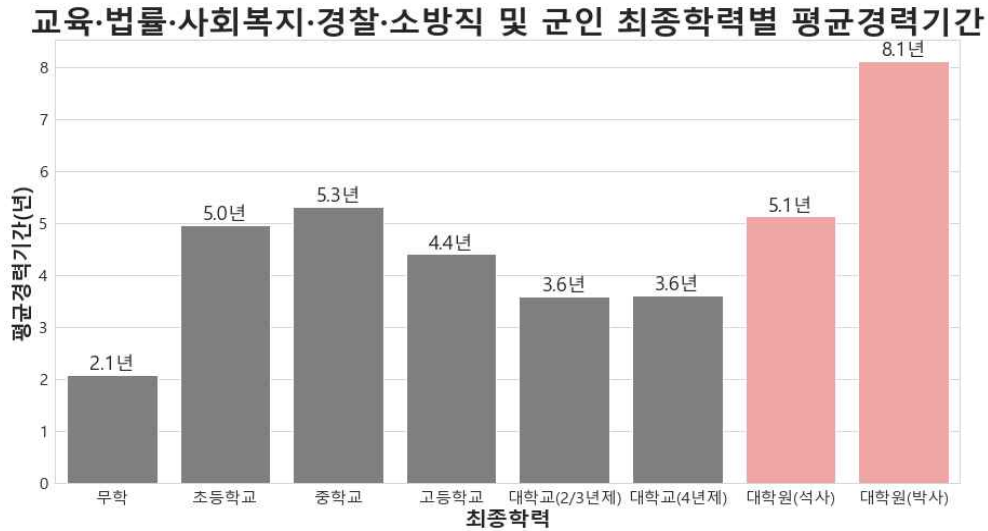
그래프 108. 경영·사무·금융·보험직 최종학력별 취업률

- 초등학교부터 중학교 졸업인 구직자의 경력기간이 길고, 고등학교부터 대학교 졸업 구직자와 희망연봉도 비슷했지만 실제 취업률은 가장 낮은 편에 속함.

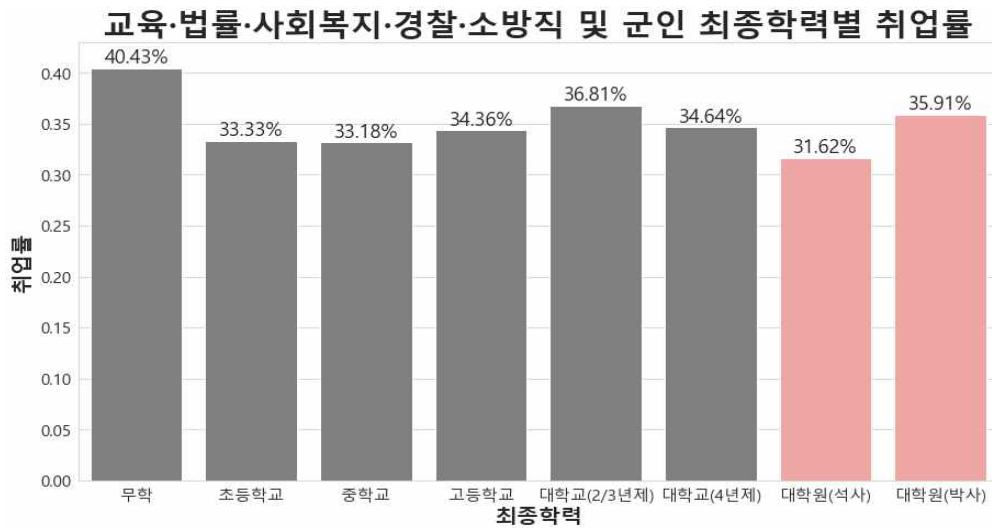
13-마) 교육·법률·사회복지·경찰·소방직 및 군인 최종학력별 경향



- 대학원(박사) 졸업 구직자들은 다른 최종학력 구직자들에 비하여 희망급여가 가장 높았고, 그 다음으로는 대학원(석사), 대학교(4년제) 졸업 아래로는 비슷한 양상을 보임.



- 교육·법률·사회복지·경찰·소방직 및 군인의 평균경력기간은 그래프 63. 경영·사무·금융·보험직과 비슷한 분포를 보이지만 평균경력기간은 더 짧아 보임.



- 무학의 취업률 경우에는 경영·사무·금융·보험직보다 조금 더 높은 것으로 보이고, 대학원(석사)의 경우는 거의 같은 수준을 보임.
- 전체적으로 경영·사무·금융·보험직과 비슷한 모양이지만 희망 급여, 평균 경력은 더 낮은 모습을 보였고 취업률의 경우 대학원(박사)는 조금 늘었지만 대학원(석사)의 경우 같은 수준을 보임.

2. 모델링

가. 로지스틱 회귀분석

- 희망 직종에 따라 취업유무에 어떤 영향을 미치는지 확인하는 것이 목적

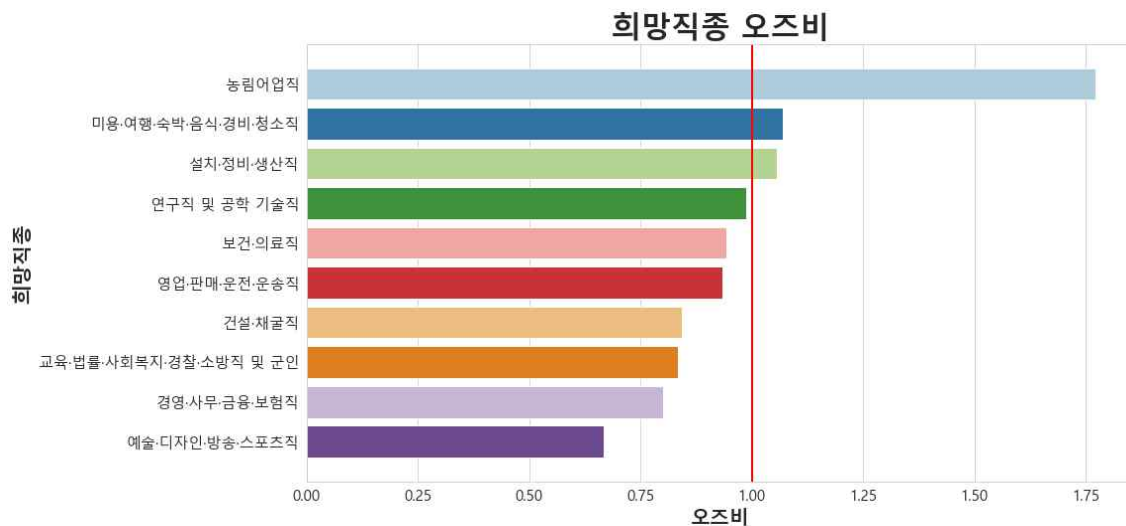
1) 데이터 사용

- 훈련 데이터 : 70%, 테스트 데이터 : 30% 사용
- 훈련 데이터와 테스트 데이터는 층화추출법(취업유무 기준)을 통해 추출
- 사용 변수 : 취업유무(종속변수), 희망직종(독립변수)

2) 결과

- 최적의 하이퍼 파라미터 : {C : 0.1, penalty : l2}
- 정확도 : 63%

3) 해석



- 다른 변수들이 동일하다는 조건하의 해석임.
- 농림어업직을 희망하면 농림어업직을 희망하지 않을 때 보다 약 1.77배 취업 확률이 올라감.
- 미용·여행·숙박·음식·경비·청소직의 경우 1.07배, 설치·정비·생산직의 경우 1.05배 정도 취업 확률이 올라감.
- 나머지는 오히려 해당직종을 희망할 경우 취업 확률이 내려가는 모습임.

나. 그래디언트 부스팅

구직자들의 취업유무에 있어 독립변수들 중 중요도가 높은 변수들을 파악하는 것이 목적

1) 데이터 사용

○ 훈련 데이터 : 70%, 테스트 데이터 : 30% 사용

○ 사용 변수

- 종속변수 : 취업유무
- 독립변수 : 연도, 성별, 나이, 주소, 희망직종, 신입경력여부, 희망급여(연봉), 고용형태, 운전가능여부, 차량소유여부, 컴퓨터활용능력, 문서작성여부, 경력건수, 희망근무지역, 전체경력기간

2) 결과

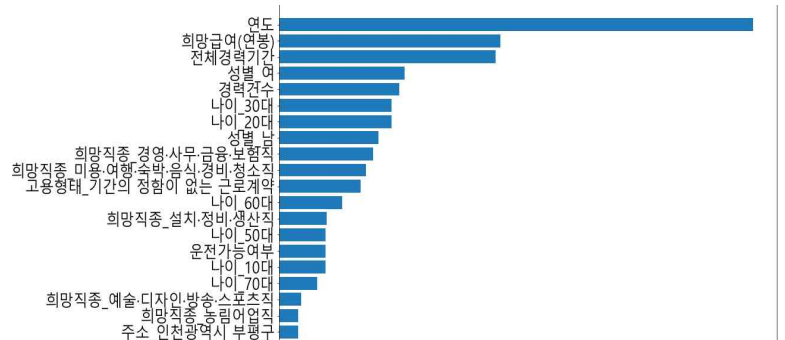
○ 최적의 하이퍼 파라미터: { min_samples_leaf : 40, max_depth : 4, n_estimators : 100, learning_rate : 0.2 }

○ 정확도 : training set 64.3% / test set 64.3%

○ confusion matrix : $\begin{bmatrix} 148486 & 12944 \\ 79102 & 17276 \end{bmatrix}$

○ 변수 중요도

	Feature	Importance
0	연도	0.206
1	희망급여(연봉)	0.096
6	전체경력기간	0.094
8	성별_여	0.054
5	경력건수	0.052
11	나이_30대	0.049
10	나이_20대	0.049
7	성별_남	0.043
44	희망직종_경영·사무·금융·보험직	0.041
47	희망직종_미용·여행·숙박·음식·경비·청소직	0.038
56	고용형태_기간의 정함이 없는 근로계약	0.036
14	나이_60대	0.027
49	희망직종_설치·정비·생산직	0.021
13	나이_50대	0.020
2	운전가능여부	0.020
9	나이_10대	0.020
15	나이_70대	0.016
52	희망직종_예술·디자인·방송·스포츠직	0.009
46	희망직종_농림어업직	0.008
26	주소_인천광역시 부평구	0.008



3) 해석

- 중요 변수로는 ‘연도(20.6%)’, ‘희망급여(9.6%)’, ‘전체경력기간(9.4%)’, ‘성별_여(9.4%)’, ‘경력건수(9.2%)’, ‘나이_30,20대(4.9%)’ 순.

* 중요 변수는 해당 모델이 종속 변수를 분류하는데 많은 영향을 준 변수

- 변수 중요도가 10% 이상인 변수는 ‘연도’뿐 임.
- ‘주소’ 변수는 영향도가 가장 낮은 편이라고 볼 수 있음.

다. 랜덤 포레스트

다수의 결정 트리로부터 취업유무를 분류하여 중요한 변수들을 확인하는 목적

1) 데이터 사용

- 훈련 데이터 : 70%, 테스트 데이터 : 30% 사용
- 훈련 데이터와 테스트 데이터는 층화추출법(연도, 취업유무 기준)을 통해 추출
- 사용 변수 :
 - 종속변수 : 취업유무
 - 독립변수 : 연도, 성별, 신입경력여부, 희망급여(연봉), 운전가능여부, 경력건수, 전체경력기간, 운전가능여부, 차량소유여부, 컴퓨터활용능력-문서작성여부, 나이, 최종학력, 희망직종, 고용형태, 취업유무(종속변수)

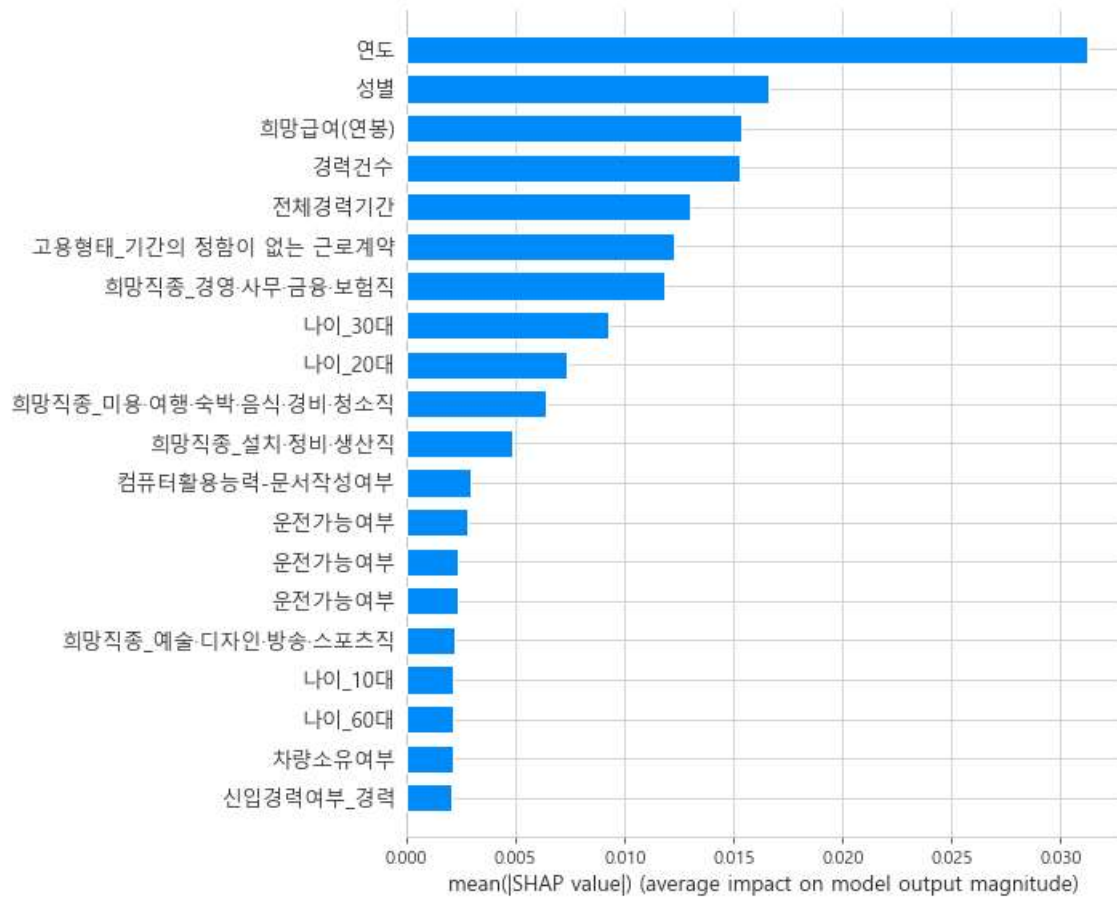
2) 결과

- 최적의 하이퍼 파라미터 : {max_depth : 12, min_samples_leaf : 8, min_samples : 8, n_estimators : 100}
- 정확도 : 63%

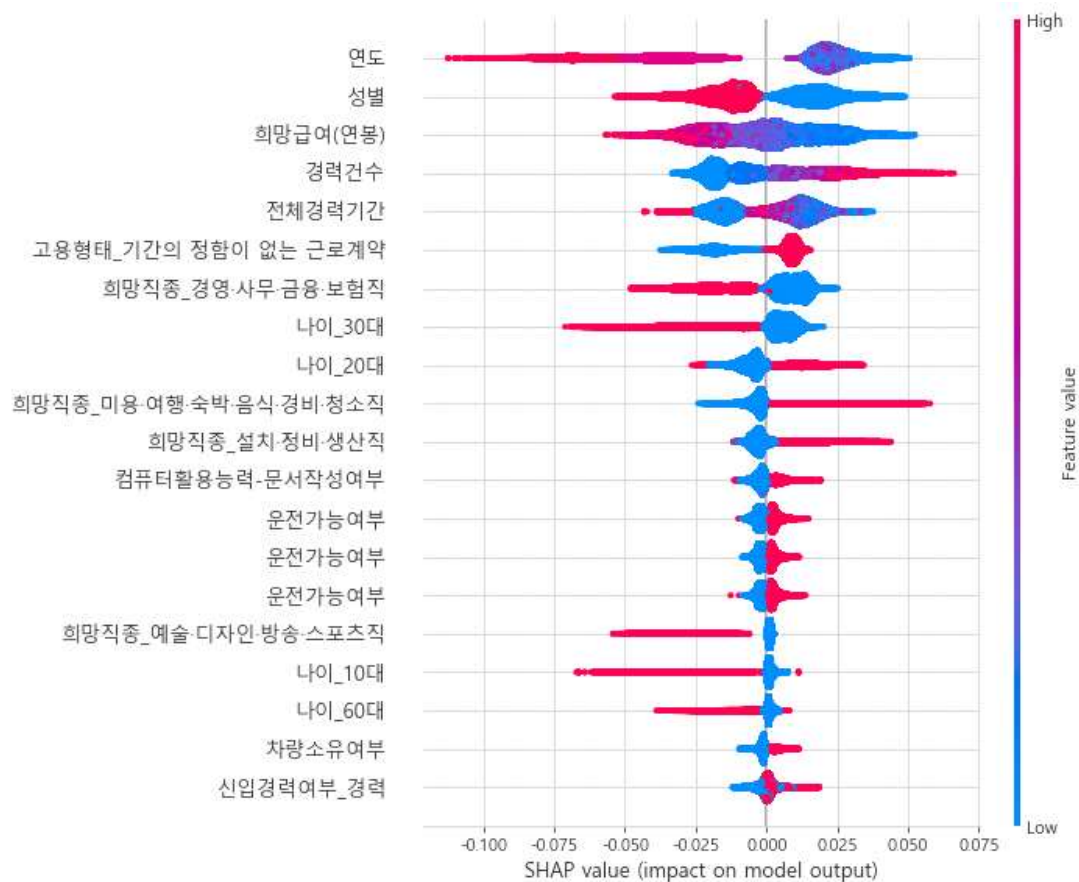
3) 해석

- 랜덤 포레스트는 모델 자체로 해석이 불가능하므로 해석을 위해 shap 패키지 사용

- summary_plot(변수 중요도)



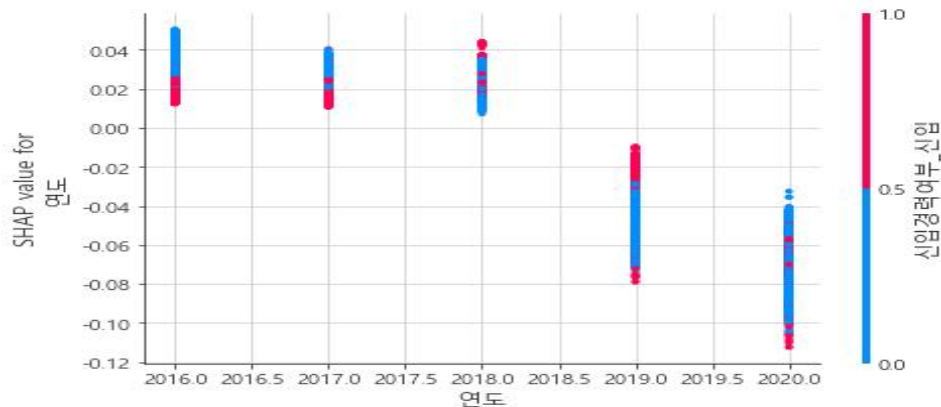
- 랜덤 포레스트가 취업으로 분류하는데 가장 큰 영향을 준 변수는 '연도' 변수임.
- * 중요 변수는 해당 모델이 종속 변수를 분류하는데 많은 영향을 준 변수
- 그 다음으로는 '성별', '희망급여(연봉)', '경력건수', '전체경력기간' 순서임.
- '고용형태'는 기간이 정함이 없는 근로계약이 가장 큰 영향을 미쳤고, '희망직종'에서는 경영·사무·금융·보험직이 가장 큰 영향을 미침.
- '나이'로는 30대 변수가 가장 큰 영향을 미친 것으로 보임.



- ‘연도’의 값이 작으면 취업 확률이 높아지는 경향이 있음.
- ‘성별’의 경우 남성이면 취업 확률이 높아지는 경향이 있음.
- ‘희망급여(연봉)’이 낮을수록 취업 확률이 높아지는 경향이 있음.
- ‘경력건수’가 많을수록 취업 확률이 높아지는 경향이 있음.
- ‘전체경력기간’이 작을수록 취업 확률이 높아지는 경향이 있음.
- ‘고용형태’가 기간이 정함이 없는 근로계약일 경우 취업 확률이 높아지는 경향이 있음.
- ‘경영·사무·금융·보험직’을 희망하지 않는다면 취업 확률이 높아지는 경향이 있음.
- ‘나이’가 30대가 아닐 경우 취업확률이 높아지는 경향이 있음.

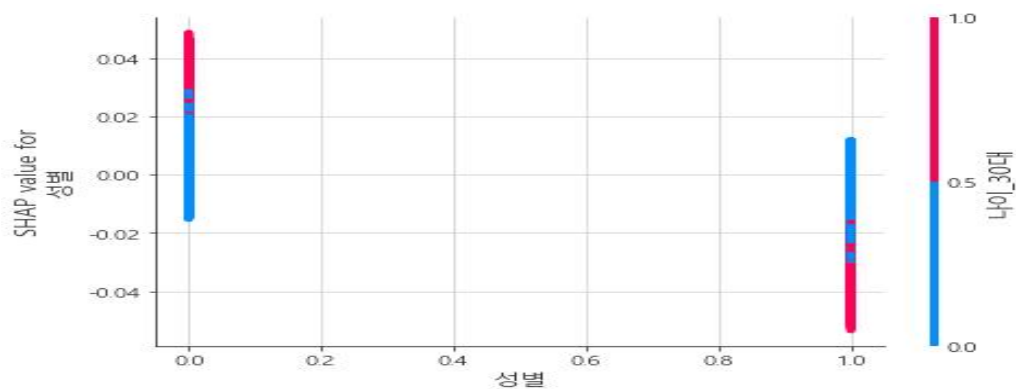
- dependence_plot

1) 연도



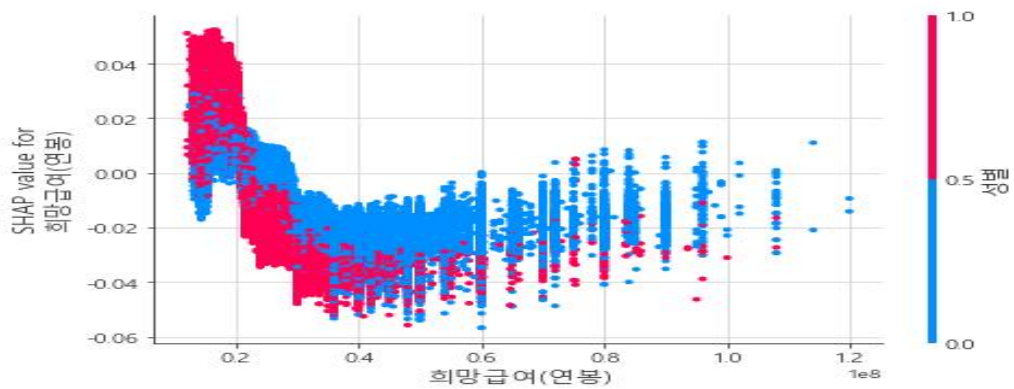
- ‘연도’ 변수와 상호작용 지수가 가장 큰 변수는 ‘신입경력여부_신입’ 변수임.
- 2016 ~2018년까지 shap value가 0을 넘겨 취업 확률을 높임.
- 2016, 2017년의 경우 신입일 경우 shape value가 떨어져 취업 확률을 상대적으로 경력에 비해 떨어트리는 모습임.
- 2016, 2017년에는 경력의 경우 취업의 확률을 높임.

2) 성별



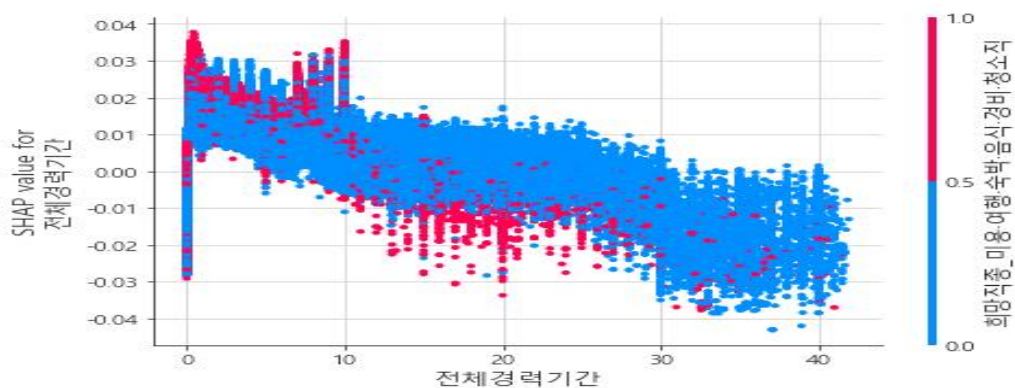
- ‘성별’ 변수와 상호작용 지수가 큰 변수는 ‘나이_30대’ 변수임.
- ‘성별’이 남자인 경우 여성보다 취업의 확률을 높임.
- 남성의 경우 30대가 아닌경우보다 30대인 경우 취업 확률을 높임.
- 여성의 경우 30대가 아닌경우보다 30대의 경우 취업 확률을 낮힘.

3) 희망급여(연봉)



- ‘희망급여(연봉)’ 변수와 상호작용 지수가 큰 변수는 ‘성별’ 변수임.
- ‘희망급여(연봉)’의 경우 2,000만원이 넘어가면 취업에 좋은 영향을 미치지 않음.
- ‘희망급여(연봉)’가 2,000만원 미만의 경우 ‘성별’이 여자인 경우가 취업 확률을 높였지만 그 이후로는 반대의 모습을 보이고 있음.⁵⁾

4) 전체경력기간



- ‘전체경력기간’ 변수와 상호작용 지수가 가장 컸던 변수는 ‘미용·여행·숙박·음식·경비·청소직’ 변수임.
- ‘전체경력기간’이 높아질수록 취업 확률에 좋지 않는 영향을 미치는 것으로 보임.
- ‘미용·여행·숙박·음식·경비·청소직’을 희망하지 않는 것이 취업에 더 좋은 영향을 미치는 것으로 보임.

5) 추후 제언1의 근거

요약 및 제언

- 1. 결과 요약
- 2. 한계점 및 제언

1. 결과 요약

가. 워크넷 구직 데이터 탐색적 분석 결과

1) 구직자

변수	설명
취업률	2016년부터 2020년까지 하락세(42% -> 27%)
성비	여자(54%) > 남자(46%)
나이대	20대(25%)로 가장 많음.
최종학력	고등학교 졸업(43%) > 대학교 졸업(2,3년제, 4년제) > 나머지
희망직종	경영(30%) > 생산 > 미용직
경력	경력(57%) > 신입
희망급여	전체적으로 낮고 비교적 높은 값들이 다수 포진
경력건수	대부분 경력이 0건
취업유무	미취업자(63%) > 취업자
희망근무지역	인천이 대다수(81%)
운전가능여부&차량소유여부	운전이 불가능한 구직자와 차량 미소유자가 많았다.

2) 희망직종

변수	설명
경영·사무·금융·보험직	- 여성이 가장 많이 희망하는 직종(38%) - 한 직장에서 재직기간이 짧은 편(평균 2.5년) - 고학력(대학교 ~ 대학원)이 많이 희망(64%)
설치·정비·생산직	- 남성이 가장 많이 희망하는 직종(27%) - 저학력(무학 ~ 고등학교)이 많이 희망(75%)
건설·채굴직	- 재직기간이 가장 긴 직종(남녀 모두 가장 김) - 급여가 높은 편(학력에 영향이 없어 보임) - 희망급여의 변동성이 가장 큼(연구직과 동일)
예술·디자인·방송·스포츠직	- 한 직장에서 재직기간이 가장 짧은 직종(평균 1.3년)
연구직 및 공학 기술직	- 희망급여의 변동성이 가장 큼(건설직과 동일, 변동계수 : 0.37) - 고학력(대학교 ~ 대학원)이 많이 희망(81%) - 평균재직기간, 전체재직기간과 상관 관계가 있어보임(각각 0.5, 0.53)
미용·여행·숙박·음식·경비·청소직	- 저학력(무학 ~ 고등학교)가 많이 희망(78%)
희망급여	- 학력이 올라갈수록 높아짐 - 저학력(무학 ~ 고등학교)은 직종별로 비슷
재직기간	- 남성 > 여성(모든 직종에 해당)

3) 최종학력

3-1) 초등학교, 중학교

변수	설명
나이	- 50,60대 비율이 높음(86%)
선호 직종	- 미용·여행·숙박·음식·경비·청소직(51%) > 설치·정비·생산직 순서
취업 직종	- 미용·여행·숙박·음식·경비·청소직 > 설치·정비·생산직 순서 - 미용·여행·숙박·음식·경비·청소직의 희망급여는 낮고 평균 경력기간은 높음 - 설치·정비·생산직은 취업률은 낮아보임.(34%)
경력	- 경력 기간은 다른 최종학력에 비해 긴 편(50,60대의 비중이 높기 때문으로 예상)(평균 5년)
- 경력은 길지만 희망하는 급여가 낮아서 취업이 잘 되는 편인 것으로 추정	

3-2) 무학

변수	설명
나이	- 50,60대 비율이 높음(67%)
취업률	- 2019년에 급감(44% -> 18%)
구직자	- 2019년 이후 50,60대 비율 증가 - 다른 젊은 나이대는 줄어드는 모습
취업률 감소	- 보건, 농림, 미용직(2019년 전후 기준) - 보건과 농림직의 경우 구직자 수, 취업자 수 모두 적어 미용직이 취업 감소량이 가장 커보임(32%)
미용직	- 10,20대 희망하는 비율 증가 - 50,60대 희망하는 비율 감소 - 나이대 별 취업률 순위는 큰 변화 없음
- 2019년 최종학력이 무학인 구직자들의 취업률이 줄어든 이유가 취업이 잘 되는 50, 60대의 희망이 줄고, 취업이 어려운 10,20대의 희망이 늘어나서 전체적으로 무학의 취업률이 급감했을거라 추정	

3-3) 고등학교

변수	설명
나이	- 40, 50대의 비율이 높음(51%)
남성	- 설치직의 비율이 높음(희망자 36%, 취업자 모두 높음) - 설치직의 나이대는 40,50,30대 순
여성	- 경영직의 비율이 높음(희망자 34%, 취업자 모두 높음) - 경영직의 나이대는 40,30,20대 순
<ul style="list-style-type: none"> - 두 직종(설치직, 경영직)에서 남녀 비율의 차이가 커보임 - 다만, 설치직 40,50대는 여성의 비율이 높은 편이고 경영직은 대부분 나이대에서 여성이 많은 것으로 보임 - 설치직의 경우 남성의 구직이 많은 직종이지만 40,50대의 여성이 많이 희망하면서 설치직을 희망하는 40,50대 비중이 높아진 것으로 보임 	

3-4) 대학교

변수	설명
대학교	
희망직종	- 경영직을 가장 희망(41%)
나이대	- 20, 30대의 비율이 높다(69%)
경력 기간	- 4년제 > 2년제
희망급여	- 4년제 > 2년제
대학원	
나이대	- 석사 : 30,40대(60%) - 박사 : 30,40,50대 고르게 분포(87%)
희망직종	- 연구직의 비중이 높아짐 - 석사 : 경영직(30%) > 교육직 > 연구직 - 박사 : 연구직(40%) > 경영직 > 경영직
희망급여	- 경영직 : 박사 > 석사 - 교육직 : 석사, 박사 모두 비슷(평균 3천)
경력기간	- 경영직 : 박사 > 석사

나. 워크넷 구직 데이터 모델링 결과

랜덤포레스트 (SHAP)	연도	성별	희망급여(연봉)	전체경력기간
상호작용 지수 큰 변수	신입경력여부_신입	나이_30대	성별	희망직종_미용·여행·숙박·음식·경비·청소직
해석	2016 ~ 2018년은 취업확률이 높아졌으며, 2017년까지는 신입의 경우 취업확률이 상대적으로 경력에 비해 떨어지는 모습.	남성의 경우, 30대가 아닌 경우 취업 확률이 높아지는 모습을 보임. 여성의 경우, 30대가 아닌 경우 취업 확률이 낮아지는 것으로 보임.	희망급여(연봉)이 2,000만원이 넘어가면 취업에 좋은 영향을 미치지 않는 것으로 보이며, 2,000만원 미만의 경우 성별이 여자인 경우 취업 확률을 높이고 그 이후로는 반대의 모습을 보이는 것으로 보임.	미용·여행·숙박·음식·경비·청소직을 희망하지 않는 것이 취업에 더 좋은 영향을 미치는 것으로 보임.

- 공통적으로 연도, 희망급여(연봉), 전체경력기간과 같이 변수 중요도가 상위 에 겹치는 변수도 있었지만 성별, 희망직종 등 다른 경우도 있음.

* 변수 중요도 : 독립 변수가 상대적으로 해당 모델에서 얼마만큼 종속변수에 영향을 주는지 확인

○ 랜덤 포레스트 모델 결과

- 랜덤 포레스트는 모델 자체로 해석이 불가능한 모델이므로 SHAP를 통해 모델 해석

	로지스틱 회귀	그라디언트 부스팅	랜덤포레스트	선택
accuracy	63%	64.3%	63%	그라디언트 부스팅 (64.3%)
유의한 변수 및 중요 변수 (top3)	농림어업직, 미용·여행·숙박·음 식·경비·청소직, 설치·정비·생산직	연도, 희망급여(연봉), 전체경력기간	연도, 성별, 희망급여(연봉)	공통 중요 변수 (연도, 희망급여)

2. 한계점 및 제언

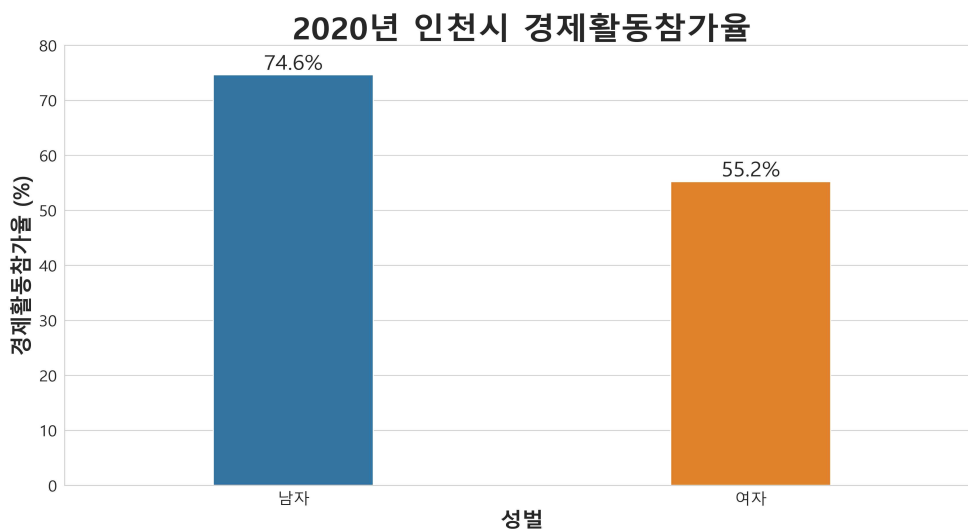
가. 한계점

- 워크넷 데이터 부실 오류많음.
 - 결측치 다수
 - 표본이 인천시 구직자를 대표하지 않는 것으로 추정
 - 컨텐츠나 마케팅 부족으로 인한 다른 구직자 사이트(사람인, 잡코리아 등) 대비 사용자가 적음 (*논문 인터뷰 내용 참조)
- 프로젝트 분석 기간이 짧아 일자리 미스매칭이 아닌 현황 분석에 그침.
- 취업유무 변수의 경우 어떤 방식으로 입력되었는지 확인이 불가하여 정확성이 떨어질 가능성이 있음.

나. 정책적 제언

1) 여성을 위한 다양한 직업 교육

- 여성은 현재 급여가 낮은 특정 직종(경영·사무·금융·보험직)에 몰려있어 전체적으로 취업률이 떨어지는 모습과 낮은 급여를 모습을 보여 다양한 직종을 희망할 수 있도록 기회 제공.



* 통계청 2020년 인천 성별 경제활동인구

또한 여성의 경우 경제활동참가율이 남성에 비해 많이 낮은 편이기 때문에 여성이 좀 더 경제활동에 참여할 수 있는 기회를 제공. (*아래쪽 참고자료 첨부)

2) 저학력자(무학, 초졸, 중졸)의 재취업을 위한 일자리 마련

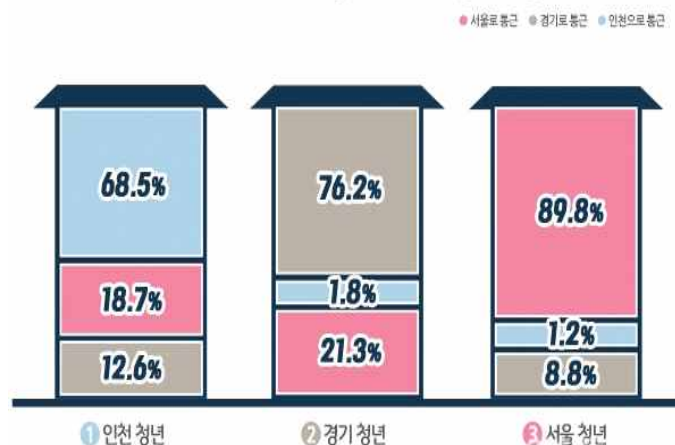
- 저학력자의 경우 코로나의 영향으로 실업자가 더 많이 늘어났다. 저학력자가 많이 희망한 미용·여행·숙박·음식·경비·청소직의 경우 평균 재직기간이 길지 않고, 50·60대의 비중이 높은 직업이기 때문에 재취업이 힘들 것으로 예상되며 특히 여행 업계는 피해가 막대한 상황이다. (*아래쪽 참고자료 첨부)

3) 인천지역 2030 청년 유출 방지 정책



- 통계청에 따르면 인천광역시 최종학력별 15세이상인구수(선그래프)는 대졸보다 고졸이 더 많다고 보여지지만, 워크넷 데이터(막대그래프)는 고졸보다 대졸로 등록된 구직자 수가 더 많다.

2018 수도권 통근자 이동 현황



출처: 통계청, 지역별고용조사(2018)

- 실제 서울과 경기도로 유출되는 인천 거주 청년들이 많다. 교육정보별 통계 그래프는 고졸 즉 20대의 사회초년생들이 인천지역에서 구직 활동을 하기보단 다른 지역으로 구직 활동을 하여 20대 청년들이 다른 지역으로 유출이 되고 있다는 2018년도 통계청 자료를 뒷받침하고 있다.

2018년 인천 지역 직업군 비중			
경영 및 회계 관련 사무직	16.5%	경영 금융 전문가 및 관련직	2.1%
장치기계 조작 및 조립 종사자	14.3%	정보 통신 전문가 및 기술직	1.6%
단순노무 종사자	14.0%	상담 안내 통계 및 기타 사무직	1.5%
판매 종사자	12.7%	금융 사무직	1.1%
기능원 및 관련기능 종사자	9.9%	경찰 소방 및 보안 관련 서비스직	0.9%
조리 및 음식 서비스직	5.1%	운송 및 여가 서비스직	0.8%
보건·사회복지 및 종교 관련직	4.9%	관리자	0.6%
교육 전문가 및 관련직	4.3%	농림어업 숙련 종사자	0.4%
돌봄·보건 및 개인 생활 서비스직	3.3%	법률 및 감사 사무직	0.2%
공학 전문가 및 기술직	3.1%	법률 및 행정 전문직	0.1%
문화·예술·스포츠 전문가 및 관련직	2.5%	과학 전문가 및 관련직	0.1%

출처: 통계청, 지역별고용조사(2018)

- 왼쪽은 인천지역 직종 비율

- 아래는 2030 청년들이 희망하는 직종 비율이다.

▲정보통신 전문가·기술직-서울 3.4%, 인천 1.6%, 경기 2.8% ▲경영·금융 전문가와 관련 직-서울 4.0%, 인천 2.1%, 경기 3.0% ▲문화·예술·스포츠 전문가와 관련직-서울 4.8%, 인천 2.5%, 경기 2.9% ▲경영·회계 관련 사무직-서울 17.2%, 인천 16.5%, 경기 15.6%

출처 :인천투데이

따라서, 2030 청년들이 희망하는 직종이 인천지역 내에 부족한 것으로 판단된다. 2030 청년들의 유출을 막기 위해서 새로운 산업군을 발굴하여 일자리를 창출하고 특히 4차산업 일자리 창출과 예술문화 일자리 창출이 필요해 보인다.

부록

[참고1]

- 행정안전부 일자리 미스 매칭 표준분석모델.pdf
- 2020년 6월 고용동향 (통계청)
- 2020년 6월 인천광역시 고용동향 (경인지방통계청)
- 한국고용직업분류표(2018).pdf
: <http://www.busanhrd.or.kr/kr/Business/data.php?mode=view&idx=279>
- 여성을 위한 다양한 직업 교육 (동아일보)
: <https://www.donga.com/news/Economy/article/all/20150817/73089901/1>
- 현재 서울시는 여성 취업 전문직업교육 28개 과정 무료 운영중 (서울일보)
: <https://news.seoul.go.kr/welfare/archives/506733>
- 저학력자(무학, 초졸, 중졸)의 재취업을 위한 일자리 마련 (조선일보)
: https://www.chosun.com/site/data/html_dir/2020/07/28/2020072801170.html
- 저학력 코로나 (조선일보)
: https://biz.chosun.com/site/data/html_dir/2020/05/13/2020051301218.html
- 인천지역 2030 청년 유출 방지 (인천투데이)
: <http://www.incheontoday.com/news/articleView.html?idxno=121810>
- 인천지역 2030 청년 유출 방지 (인천투데이)
: <http://www.incheontoday.com/news/articleView.html?idxno=121621>

[참고2]

취업알선직업분류 2018 분류단계별 항목 수

대분류	중분류	소분류	세분류	세세분류
10	35	136	450	933
0 경영·사무·금융·보험직	01 관리직(임원·부서장)	6	24	40
	02 경영·행정·사무직	9	32	103
	03 금융·보험직	3	14	20
1 연구직 및 공학 기술직	11 인문·사회과학 연구직	1	2	2
	12 자연·생명과학 연구직	2	5	7
	13 정보통신 연구개발직 및 공학기술직	6	13	31
	14 건설·채굴 연구개발직 및 공학기술직	1	7	18
	15 제조 연구개발직 및 공학기술직	9	25	62
2 교육·법률·사회복지·경찰·소방직 및 군인	21 교육직	5	16	48
	22 법률직	2	6	8
	23 사회복지·종교직	3	9	17
	24 경찰·소방·교도직	1	3	3
	25 군인	1	4	4
3 보건·의료직	30 보건·의료직	7	23	31
4 예술·디자인·방송·스포츠직	41 예술·디자인·방송직	7	30	60
	42 스포츠·레크리에이션직	1	5	12
5 미용·여행·숙박·음식·경비·청소직	51 미용·예식 서비스직	2	11	18
	52 여행·숙박·오락 서비스직	4	7	10
	53 음식 서비스직	2	13	23
	54 경호·경비직	2	5	13
	55 돌봄 서비스직(간병·육아)	1	2	6
	56 청소 및 기타 개인서비스직	2	11	20
6 영업·판매·운전·운송직	61 영업·판매직	7	20	62
	62 운전·운송직	4	15	30
7 건설·채굴직	70 건설·채굴직	6	24	43
8 설치·정비·생산직	81 기계 설치·정비·생산직	7	21	56
	82 금속·재료 설치·정비·생산직(판금·단조·주조·용접·도장 등)	6	19	38
	83 전기·전자 설치·정비·생산직	6	13	31
	84 정보통신 설치·정비직	2	5	7
	85 화학·환경 설치·정비·생산직	3	9	22
	86 섬유·의복 생산직	4	16	24
	87 식품 가공·생산직	3	12	17
	88 인쇄·목재·공예 및 기타 설치·정비·생산직	5	15	24
	89 제조단순직	1	1	7
9 농림어업직	91 농림어업직	5	13	16