


붙임 1

제5회 대구 빅데이터 분석 경진대회 참가신청서

접수번호	※작성하지 않음 ※ 해당되는 부분에 <input checked="" type="checkbox"/> 체크 바랍니다.			
지원형태	<input type="checkbox"/> 개인 <input checked="" type="checkbox"/> 팀			
신청(대표)자명	ChatKPT	생년월일	1999.01.20	
주 소	대구광역시 달서구 호산동로 35길 28-19			
연락처(휴대폰)	010-8858-0833	E-Mail	jun99120@gmail.com	
소 속	계명대학교			
공모 과제	■ 데이터 분석 및 활용 분야 일반 부문(<input checked="" type="checkbox"/>) 금융 부문(<input type="checkbox"/>) 소비자 부문(<input type="checkbox"/>) 에너지 부문(<input type="checkbox"/>)			
구 분	성 명	연락처(휴대폰)	소속/직위	E-Mail
대표자	윤태준	010-8858-0833	계명대학교	jun99120@gmail.com
팀 원	이도건	010-2460-2821	계명대학교	dlehrjs10@naver.com
팀 원	김민경	010-5038-3121	계명대학교	mrud0424@naver.com

준수사항 동 의	1. '제5회 대구 빅데이터 분석 경진대회'의 제반 규정 및 가이드라인을 준수하며, 이를 준수하지 않을 경우 불이익을 받을 수 있습니다. 2. 본 경진대회에 제출한 산출물이 타 공모전 및 경진대회의 수상작이거나 내용의 상당 부분에 유사성이 발견될 경우, 허위사실 기재, 제 3자의 지식재산권 및 정보 등의 무단사용을 하였을 경우, 입상 취소, 상금 환수 등의 제재 조치를 따를 것이며, 타인과의 법적 분쟁은 참가자 본인의 책임입니다. 3. 본 경진대회에 제출된 모든 서류(참가신청서, 분석결과서 등)는 일체 반환되지 않습니다. 4. 심사결과에 따라 적합한 수상작이 없을 경우 수상작을 선정하지 않거나 시상내역이 변동될 수 있으며, 참가자는 이에 대하여 이의를 제기하지 않습니다. 5. 수상작의 경우 대구광역시가 정책개선 및 공공 활용을 위해 D-datahub 및 대구빅데이터활용센터 홈페이지에 게재하여 활용 및 공개할 수 있습니다. 동의함 <input checked="" type="checkbox"/> 동의하지 않음 <input type="checkbox"/>
-------------	---

위와 같이 『제5회 대구 빅데이터 분석 경진대회』 참가를 신청합니다. 2023년 5 월 31 일 신청자(대표자) 윤태준 	
(재)대구디지털혁신진흥원장 귀하	
첨 부	(붙임2) 개인정보 수집·이용 동의서 1부, (붙임3) 과제분석계획서 1부.

【개인정보 수집·이용 동의 안내】

「개인정보 보호법」 제15조 및 제23조에 따라 개인정보를 수집·이용하는 경우 개인의 동의를 얻어야 합니다.

가. 개인정보 수집·이용 목적

- 「제5회 대구 빅데이터 분석 경진대회」를 위한 '수상작 선정평가 및 경진대회 운영·관리'를 목적으로 이용됩니다.

나. 개인정보 수집 항목

- 성명, 소속/직위, 휴대전화, 전자우편, 거주지역

다. 개인정보의 보유·이용기간

- 개인정보는 수집·이용에 관한 동의일로부터 경진대회 **최종 결과 발표일로부터 6개월간** 이용목적을 위하여 보유·이용되며, 경진대회 수상자는 최종 결과 발표일로부터 2년간 보유·이용됩니다.

라. 개인정보 수집·이용에 동의하지 않을 권리 및 동의하지 않을 경우의 불이익

- 정보주체는 「제5회 대구 빅데이터 분석 경진대회」에 개인정보 수집·이용의 동의를 거부할 권리가 있습니다.
- 개인정보 수집·이용에 동의하지 않을 경우, 본 대회에 참가신청이 불가합니다.

본인은 「제5회 대구 빅데이터 분석 경진대회」에서 본인의 개인정보를 수집·이용하는 것에 동의합니다.

(동의함 ☒ / 동의하지 않음 ☐)

2023년 5 월 31 일

신청인 _____ 윤태준 (인)

【개인정보 수집·이용 동의 안내】

「개인정보 보호법」 제15조 및 제23조에 따라 개인정보를 수집·이용하는 경우 개인의 동의를 얻어야 합니다.

가. 개인정보 수집·이용 목적

- 「제5회 대구 빅데이터 분석 경진대회」를 위한 '수상작 선정평가 및 경진대회 운영·관리'를 목적으로 이용됩니다.

나. 개인정보 수집 항목

- 성명, 소속/직위, 휴대전화, 전자우편, 거주지역

다. 개인정보의 보유·이용기간

- 개인정보는 수집·이용에 관한 동의일로부터 경진대회 **최종 결과 발표일로부터 6개월간** 이용목적을 위하여 보유·이용되며, 경진대회 수상자는 최종 결과 발표일로부터 2년간 보유·이용됩니다.

라. 개인정보 수집·이용에 동의하지 않을 권리 및 동의하지 않을 경우의 불이익

- 정보주체는 「제5회 대구 빅데이터 분석 경진대회」에 개인정보 수집·이용의 동의를 거부할 권리가 있습니다.
- 개인정보 수집·이용에 동의하지 않을 경우, 본 대회에 참가신청이 불가합니다.

본인은 「제5회 대구 빅데이터 분석 경진대회」에서 본인의 개인정보를 수집·이용하는 것에 동의합니다.

(동의함 ☒ / 동의하지 않음 ☐)

2023년 5 월 31 일

신청인 _____ 이도건 이도건

【개인정보 수집·이용 동의 안내】

「개인정보 보호법」 제15조 및 제23조에 따라 개인정보를 수집·이용하는 경우 개인의 동의를 얻어야 합니다.

가. 개인정보 수집·이용 목적

- 「제5회 대구 빅데이터 분석 경진대회」를 위한 '수상작 선정평가 및 경진대회 운영·관리'를 목적으로 이용됩니다.

나. 개인정보 수집 항목

- 성명, 소속/직위, 휴대전화, 전자우편, 거주지역

다. 개인정보의 보유·이용기간

- 개인정보는 수집·이용에 관한 동의일로부터 경진대회 **최종 결과 발표일로부터 6개월간** 이용목적을 위하여 보유·이용되며, 경진대회 수상자는 최종 결과 발표일로부터 2년간 보유·이용됩니다.


라. 개인정보 수집·이용에 동의하지 않을 권리 및 동의하지 않을 경우의 불이익

- 정보주체는 「제5회 대구 빅데이터 분석 경진대회」에 개인정보 수집·이용의 동의를 거부할 권리가 있습니다.
- 개인정보 수집·이용에 동의하지 않을 경우, 본 대회에 참가신청이 불가합니다.

본인은 「제5회 대구 빅데이터 분석 경진대회」에서 본인의 개인정보를 수집·이용하는 것에 동의합니다.

(동의함 ☒ / 동의하지 않음 ☐)

2023년 5 월 31 일

신청인 _____ 김민경 

접수번호

※작성하지 않음

「제5회 대구 빅데이터 분석 경진대회」 과제 분석 계획서

신청자명(팀명) ChatKpt

분석주제명 대구 청년들의 장기근속요인 파악

1. 기획 배경 및 개요

통계청 발표에 따르면, 고용률은 증가하고 실업률은 감소하고 있다. 이는 한 시점의 결과이기 때문에 취업의 성공 요인으로 고용률만 고려하는 것은 무리가 있다. 취업을 한 것으로 끝이 아니라, 취업 후 장기적으로 근무를 하는 것도 중요한 취업 흐름의 요인으로 봐야 한다.

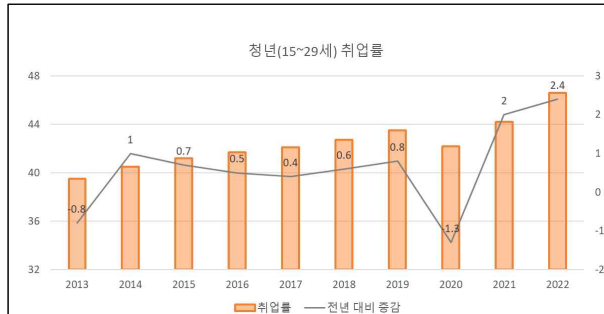


그림 2 청년(15~29세)의 취업률

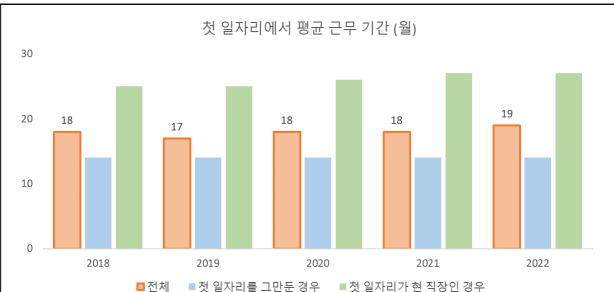


그림 3 청년(15~29세)의 첫 일자리에 평균 근무 기간

근속을 하지 않을 시 기업과 구직자 모두에게 손실이 발생한다. 구직자의 경우 업무와 관련된 경력을 쌓기 어려우며, 이는 곧 전문성을 높이는 개인의 발전에 걸림돌이 될 수 있다. 기업의 경우 시간과 비용을 들여 교육하지만, 근무 기간이 짧다면 직원을 다시 채용 후 교육해야 한다.

'3명 중 1명 조기퇴사'...커리어 고민 큰 MZ세대 '퇴준생' 신세

등록 2022.03.24 07:50:00 수정 2022.03.24 07:46:43

지난해 신규입사자 28% 5개월 내 퇴사
2019년 퇴사율에 비해 10% 이상 늘어
이름값보다 성장 가능성 중시하는 MZ
"이직이 곧 성장·해결책이 돼버린 듯"
"월급만으로 내집 마련 불가능 점 영향"

그림 4 조기퇴사 관련, 출처: 뉴시스

육재는 시의원 "MZ세대 공무원 퇴사율 급증...사기진작을 위한 대책 필요"

신성대 기자 | 승인 2023.04.05 17:40 | 댓글 0

MZ세대 공무원 의원면직률 19년 4.7%에서 22년 8.5%로 2배 증가
육재는 시의원, "대면 서비스 질 향상 및 불심해 일하는 분위기 조성 위한 공무원 사기진작 대책 필요"



출처기사
후한진 2022.03.05. 광명동자이 호텔청지 기아본...
"185가 공무원들의 재발구입...가" 광명동자이...
선공위 자내자를 의해...출처...10명 이상...
100여 명을...최신지...100여 명이...
선생는 우리 내장...해...이...대...
출처: 파이낸스 투데이

그림 5 공무원 퇴사관련, 출처: 파이낸스 투데이

현재 사회적으로 사기업, 공공기관 구분 없이 MZ라고 칭하는 세대의 장기 근속률이 낮아지고 있으며, 특히 청년(~29세)까지의 인구의 근속률이 낮다.

이번 프로젝트를 통해 대학 졸업자의 첫 일자리에 근무한 사람과 그렇지 못한 사람의 요인을 알아보고, 이와 관련된 개인의 특성, 나아가 첫 일자리에 근무한 사람이 다닌

기업의 요인도 알아보고자 한다. 이 연구 결과들을 토대로 근속 기간을 높일 수 있는 방향성을 제시하고자 한다. 분석 결과를 통해 기업 입장에서 채용 공고를 게재할 때, 어떤 정보가 있으면 구직자들이 관심을 가지는지 알 수 있을 것이다. 그리고 기업의 정책을 파악 및 보완할 수 있을 것이다. 그리고 구직자의 역량과 기업의 직무를 정확히 파악하는 것이 어떤 영향을 미치는지 파악할 수 있을 것이다.

장기 근속에 대한 기준은 현재 정부에서 청년과 기업의 상생을 위한 제도로 2년 만기형인 청년내일채움공제를 운영하는 것으로 보아 최소 2년 이상 근무를 한다면 근속하는 것으로 간주했다. 따라서 종속변수로 졸업 후 첫 일자리에서 일자리를 그만둔 시점까지 근무 기간이 24개월 이상이면 근속을 한 것으로 보고, 졸업 후 첫 일자리에서 조사 시점까지 근무 기간이 24개월 이상인 경우도 최소기준을 넘겼기 때문에 근속을 한 것으로 포함했다.

2. 분석 내용 요약

1. 대졸자직업이동경로조사(GOMS) 데이터를 이용하여, 대구 청년들의 장기 근속 여부 파악
2. 장기 근속 여부 파악 이후, 영향을 미치는 요인 탐색

3. 분석 방법 및 계획

☞ 사용데이터

- 대졸자직업이동경로 조사(GOMS: Graduate Occupational Mobility Survey)
2017년 8월 ~ 2019년 2월 졸업자를 대상으로, 2019년과 2020년에 설문조사를 진행한 GOMS18, GOMS19사용

☞ 분석 방법

종속변수 “2년 이상 근속 여부”로 이분형 자료로 설정

- 연속형 변수의 경우 이 표본 t 검정(Independent two sample t-test) 수행
- 범주형 자료의 경우 카이제곱 검정(Chi-square test) 수행

가. 로지스틱 회귀분석 (Logistic Regression)

로지스틱 회귀분석(Logistic Regression) 은 하나 이상의 독립 변수와 종속 변수 간의 관계를 분석하는 데 사용되는 통계 방법이다. 로지스틱 회귀분석은 입력 변수를 기반으로 특정 사건이 발생할 확률을 계산한다.

$$\log\left[\frac{P(Y_i = 1|x_i)}{1 - P(Y_i = 1|x_i)}\right] = \beta_0 + \beta_1 X_{(1i)} + \beta_2 X_{(2i)} + \dots + \beta_k X_{(ki)} \dots (1)$$

$$[\exp\{\hat{\beta}_j - z_{(\alpha/2)} \text{s.e}(\hat{\beta}_j)\}, \exp\{\hat{\beta}_j + z_{(\alpha/2)} \text{s.e}(\hat{\beta}_j)\}] \dots (2)$$

수식(1)에서 $\log(\frac{p}{1-p})$ 는 취업 후 2년 이상 근속 여부에 대한 log 값을 나타내고, x_i 는 대졸자 특징 모델 변수 (인구 특성, 대학 생활 요인, 삶의 만족도) 및 기업 특징 모델 변수(기업 복지, 근무 강도, 직장 만족도)를 나타낸다. 마지막으로 ϵ_i 는 오차항을 의미한다. 로지스틱 회

귀분석의 결과 해석은 구해진 β 의 지수값인 오즈비(odds ratio)를 사용하고 로지스틱 회귀분석의 분포가 정규성을 띤다고 보기 어렵기 때문에 식(2)와 같이 오즈비의 95% 신뢰구간(95% confidence interval)이 1을 포함하지 않으면 통계적으로 유의하다고 판단한다.

나. Kaplan-Meier & Log Rank test

대졸자의 졸업 후 첫 직장을 그만둘 때까지의 기간을 월(month)로 계산하여 생존자료의 분석 방법인 카플란-마이어(Kaplan-Meier)의 누적한계추정량 방법으로 중간 생존 시간(median survival time)을 확인했다. 그리고 근무 기간에 따른 퇴직자의 비율을 시각적으로 보기 위해 카플란-마이어 곡선(Kaplan-Meier Curve)을 제시했다. 각 시점에서 생존 확률을 추정하는 함수이며, 서로 다른 그룹 간의 생존 확률을 비교하거나 모집단의 중간 생존 시간을 식별하는데 사용할 수 있다. 그리고 그룹 간의 차이가 있는지 검정을 위해 로그-순위 검정(log-rank test)을 수행했다. 로그-순위 검정은 둘 이상의 그룹의 생존 분포를 비교하는 데 사용되는 통계적 가설 검정이다. 두 개 이상의 그룹 간의 중간 생존 시간에 유의미한 차이가 있는지 확인하기 위해 사용된다. 중앙값은 생존 시간 분포에서 50%의 대상의 관심 사건이 발생하는 시간을 의미한다. 따라서 각 그룹의 생존 시간 분포에서 표본의 근속 비율이 중앙값(survival medain time, 50%)이 되는 지점의 차이가 있는지 알아보고자 사용했다.

$$\hat{S}(t) = \begin{cases} 1 & t < t_1 \\ \prod_{j=1}^{t_i \leq t_i} (1 - \frac{d_i}{Y_j}) & t \geq t_1 \end{cases} \quad \dots(3)$$

$$\chi_{LR}^2 = \frac{[\sum_{i=1}^D (d_{1i} - Y_{1i} \frac{d_i}{Y_i})]^2}{\sum_{i=1}^D \frac{Y_{1i} Y_{2i}}{Y_i - 1} \frac{d_i}{Y_i} (1 - \frac{d_i}{Y_i})} \quad \dots(4)$$

χ_{LR}^2 = 로그-순위카이제곱검정 (log-rank chi-square test)으로, 각 그룹에서 발생한 사건 수와 그룹 내 위험률의 기댓값과 실제값의 차이

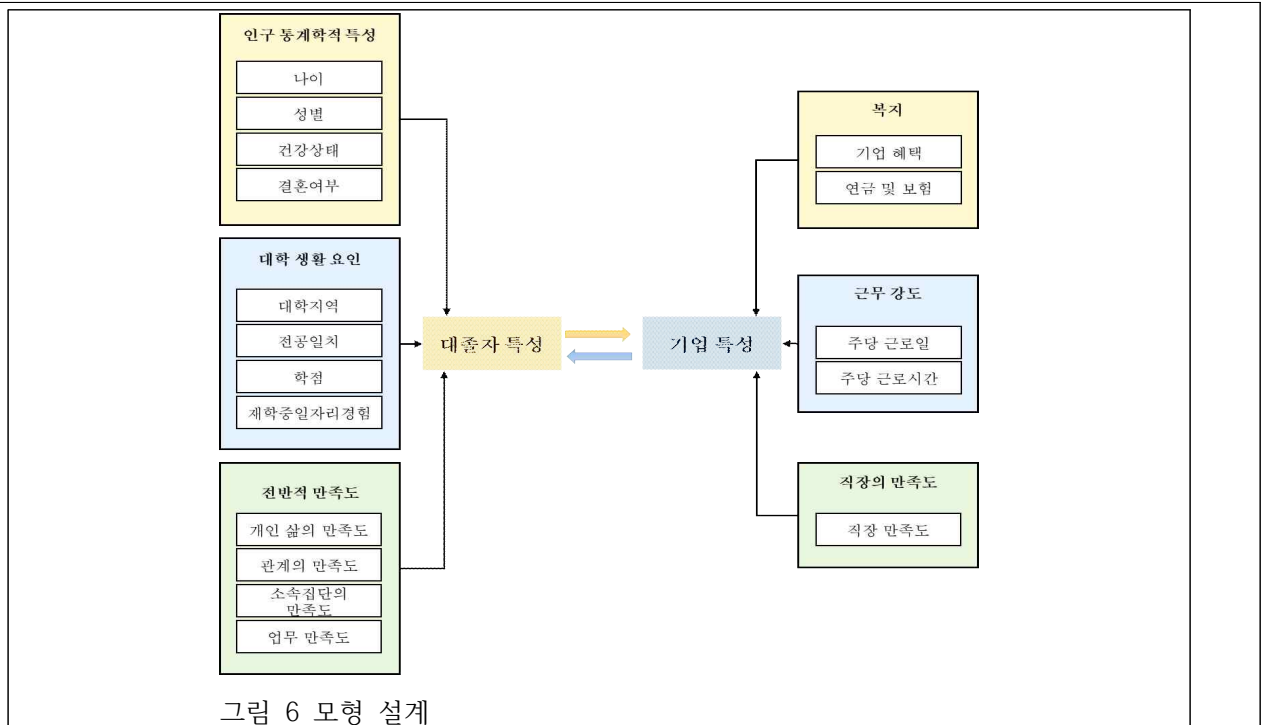
$d_i = \sum_{g=1}^K d_{gi}$ 는 t_i 시점에서 관측된 사건발생수

$Y_i = \sum_{g=1}^K d_{gi}$ 는 t_i 시점에서 통합한 데이터에서 위험개체수

$\frac{d_i}{Y_i}$ 는 t_i 시점에서 관측된 위험률

$\frac{d_{gi}}{Y_{gi}}$ 는 t_i 시점에서 그룹 g 에서 관측된 위험률 $g = 1, 2, \dots, K$

수식(3)과 (4)를 통해 성별에서 남과 여, 전공에서 인문사회어문계열, 이공계열 및 예체능계열, 종사자 수 1,000명 미만과 1,000명 이상 기업의 차이를 파악한다.



👉 분석 Tool

- Python, R 탐색적 데이터 분석(시각화 포함) 및 통계 모델링에 이용하며, SAS를 사용하여 데이터 정리 및 전처리에 이용하여 분석 결과를 도출한다.

4. 분석 결과 활용 및 시사점

분석 결과를 통해 구직자와 기업에 장기 근속에 영향을 미치는 요인에 대한 정보를 제공하며, 이를 이용하여 현재 사회적인 문제인 일자리 미스매칭(Job mismatching)을 해소한다.

5. 참고문헌

1. 성웅현. (2021). 『응용 로지스틱 회귀분석 -이론, 방법론, SAS활용-』. 탐진. ISBN: 978-89-5540-363-3 93310.
2. 이승구, 오성욱. (2021). 「대학생의 직업가치에 따라 선택한 일자리가 직무만족과 삶의 만족에 미치는 영향 : 이직의도를 조절변수로」. 2021 고용패널조사 학술대회
3. 이영일, 황선주 and 정성광. (2020). 「직원 채용과정에서 NCS 교육과 기업 요구 역량에 기반한 잡 미스매칭 연구」. 전문경영인연구, 23(1), 373-395.
4. 중소기업청. (2012). 「알기 쉽게 풀어 쓴 중소기업 범위해설」
5. 중소벤처기업진흥공단. (2016). 「청년내일채움공제」
6. 통계청.(2023). 「2023년 3월 고용동향」. 통계청 사회통계국 고용통계과
7. 홍우형, 김우철 and 김상봉. (2019). 「기대임금 증가가 근로자 이직확률에 미치는 효과: 「내일채움공제」 사례를 중심으로」. 재정학연구, 12(3), 1-32.
8. 황광훈 (Hwang Kwanghoon). 「“청년취업자의 주관적 미스매치 결정요인 및 임금효과 분석.”」 산업노동연구 24.2 (2018): 181-214.
9. 황수빈, 박상순. (2021). 「코로나19 이후 노동시장 미스매치 상황 평가」. 한국은행 논고
10. 국가통계포털. (2022). 「경제활동 인구조사: 첫직장 근속기간」. 경제활동인구조사 청년층 부가조사