MZ세대 분석 서비스

빅데이터 분석 결과 보고서

2023.05.25 ~ 2023.06.25

팀명 : DreamCatcher

참여인원 : 윤희선, 조민정, 이준형

목차

1. 분석 프로젝트 개요
   1. 개요
   2. 분석 배경 및 기획의도
   3. 분석 목표
2. 분석 데이터
   1. 데이터 설명
3. 분석 과정
   1. 데이터 전처리
   2. 삶의 만족도 분석
4. 분석 결과
   1. 삶의 만족도 회귀 분석 결과
5. 결론
   1. 분석결과 요약
   2. 분석의 한계
   3. 활용방안
6. 참고문헌
7. 프로젝트 개요
   1. 개요
      * 프로젝트명 : MZ세대 삶의 질 분석 서비스
      * 프로젝트 기간 : 2023.05.25 ~ 2023.06.25
      * 데이터 분석 언어 : R
   2. 분석 배경 및 기획의도

MZ세대로 불리우는 2·30대 청년들이 살아가고 있는 지금의 시대는 유튜브, SNS등 다양한 플랫폼이 활성화되면서 자신과 타인의 삶을 비교하는 모습을 많이 볼 수 있다.

이에 따라 상대적인 박탈감에 의해 MZ세대의 행복지수는 낮아지고 있으며 삶에 대한 만족도 또한 낮아지고 있다.

이에 본 프로젝트는 다양한 삶을 살아가고 있는 사람들의 삶과 실제 느끼는 만족도의 데이터 분석을 통해 MZ세대의 삶의 만족도에 영향을 미치는 요인과 향상 시킬 수 있는 방법을 찾아 제안하고자 한다.

* 1. 분석 목표
* 빅데이터를 활용하여 20~30대 청년들이 느끼고 있는 삶의 만족도와 삶의 질을 파악
* 분석을 통해 도출한 MZ세대의 삶의 만족도에 영향을 미치는 요인을 토대로 개인의 조건을 구체적으로 비교·분석해 만족도를 높일 수 있는 개선 방향을 제시

1. 분석 데이터
   1. 데이터 설명

* 한국노동연구원에서 약 1만 3000여가구를 24년 간 추적 조사한 데이터 중 4개 년도의 데이터를 수집

|  |  |
| --- | --- |
| 데이터명 | 데이터 설명 |
| 한국노동패널조사  설문조사 데이터 | - 한국노동연구원에서 진행한 24회차 중 4개 년도의 설문 결과 데이터 (12,962건)  - 37개의 변수 중 분석에 필요한 13개만 추출   |  |  | | --- | --- | | 컬럼 ID | 컬럼명 | | year | 조사년도 | | p\_age | 나이 | | p\_edu | 교육수준 | | p\_firm\_size | 종업원 규모 | | p\_hours | 주당 평균근로시간 | | p\_jobfam2017 | 직종 | | p\_married | 가구원혼인상태 | | p\_region | 거주지역 | | p\_sex | 성별 | | p\_wage | 월평균 임금 | | p0201 | 취업/미취업 여부 | | p0317 | 정규직/비정규직 여부 | | p6508 | 삶의 만족도 | |

1. 분석 과정
   1. 데이터 전처리

|  |  |
| --- | --- |
| 구분 | 처리 내용 |
| 결측값 처리 | - 데이터 내의 월 임금을 제외한 다른 변수에서의 결측 값은 NA로 대체 |
| 미취업자의 임금 | - 미취업자의 경우 결측 값이었던 월 임금을 0으로 대체 |
| 특정 연령 추출 | - 전 연령으로 구성되어 있는 데이터 중 2·30대의 데이터만 추출 |
| 연령 그룹화 | - 2·30대의 연령은 유사한 환경에 속한 그룹으로 묶음  - 20대 초반(20~22), 20대 중반(23~26), 20대 후반(27~29),  30대 초반(30~32), 30대 중반(33~36), 30대 후반(37~39)  6개의 그룹으로 한정 |
| 삶의 만족도 수치 | - 만족도는 최소 1점에서 최대 5점의 범위를 가짐 |

* 1. 삶의 만족도 분석
  + 여러 개의 독립변수들이 종속변수에 미치는 영향을 분석하기 위해 다중회귀분석을 실시하였다. 그 결과 예측에 대한 정확도로 48%가 산출되어 다소 낮은 수치에 회귀가 어렵다고 판단되어 독립변수가 2개 이상일 경우 실행하는 다중 로지스틱 회귀 분석을 진행하였다.
  + Z-test로 변수 별로 p-value를 산출했을 때 여러 독립변수 중 유의미한 결과를 보였던 변수들을 채택했으며 연령은 p-value값이 0.5를 넘기는 값이었지만 다른 변수 들과의 상호작용이 존재하여 영향이 없지 않았으므로 포함
* 변수 설정

|  |  |
| --- | --- |
| 종속변수 | 삶의 만족도 (sat\_lifeTotal) : 1점 ~ 5점  만족도를 점수별로 나누어 5개의 범주로 진행 |
| 독립변수 | - 연령(age)  1 = 20대 초반 / 2 = 20대 중반 / 3 = 20대 후반 / 4 = 30대 초반 / 5 = 30대 중반 /6 = 30대 후반  - 성별(sex)  1 = 남자 / 2 = 여자  - 학력(edu)  1 = 무학 / 2 = 고등학교 졸업 미만 / 3 = 고등학교 졸업 / 4 = 대학 재학, 중퇴 / 5 = 2·3년제 대학 졸업 / 6 = 4년제 대학 졸업 이상  - 취업여부(hire)  1 = 취업 / 2 = 미취업  - 혼인상태(marriage)  1 = 미혼 / 2 = 유배우자 |

* 다중 로지스틱 회귀 분석(Multinomial Logistic Regression) 진행

R에서 multinom() 함수를 사용하여 로지스틱 회귀 분석을 실시

총 12,962개의 데이터 중 75%인 9,712개를 학습 데이터로 나누어 로지스틱 모델을 학습시켜 각 분류 별로 회귀 모델 도출한 결과

Call:

multinom(formula = sat\_lifeTotal ~ age + hire + marriage + sex +

edu, data = train.set)

Coefficients:

(Intercept) age hire marriage sex edu

2 2.9328753 0.08915888 -0.6188779 -1.402977 0.13586685 0.9434152

3 3.7799499 0.10452629 -1.6718984 -2.353177 -0.04319590 1.3975842

4 0.3499232 0.17639996 -2.3634291 -2.059885 -0.06258468 2.0507193

5 -1.9850527 0.16977247 -2.2859893 -2.300455 -0.14621132 2.3454761

Std. Errors:

(Intercept) age hire marriage sex edu

2 0.7706066 0.08277204 0.2762476 0.3416481 0.2718459 0.1196034

3 0.7768380 0.08411287 0.2789100 0.3460566 0.2739208 0.1211181

4 0.7917008 0.08581562 0.2823217 0.3492683 0.2770621 0.1237764

5 0.8531236 0.09022255 0.2920795 0.3589628 0.2852425 0.1339159

Residual Deviance: 21272.01

AIC: 21320.01

위와 같은 다중 로지스틱 회귀분석의 계수를 구할 수 있다.

계수를 통해 만족도 분류 별로 어떤 변수가 영향을 많이 미치는지 알 수 있다.

또한 회귀 모델에 사용자가 입력한 조건을 변수로 대입하여 predict() 함수를

이용하여 사용자의 삶의 만족도 예측 값을 산출해낼 수 있다.

1. 분석결과

* 삶의 만족도 분석 결과

삶의 만족도에 대하여 분석한 결과 모든 그룹에서 취업여부, 혼인상태, 학력이 가장 많은 영향을 미치는 것을 알 수 있었다.

이 세가지 중에서도 각 그룹별로 가중치의 순위는 다르게 나타났다.

|  |  |
| --- | --- |
| 삶의 만족도 | 중요도 |
| 1점 | reference |
| 2점 | ① 학력 / ② 혼인상태 / ③ 취업여부 |
| 3점 | ① 혼인상태 / ② 취업여부 / ③ 학력 |
| 4점 | ① 취업여부 / ② 학력 / ③ 혼인상태 |
| 5점 | ① 학력 / ② 취업여부 / ③ 혼인상태 |

또한 변수의 값에 따라 삶의 만족도가 달라지는 것을 확인할 수 있었다.

* 연령: 연령이 증가할수록 삶의 만족도 증가
* 성별: 여자에 비해 남자가 삶의 만족도가 높음
* 학력: 고학력자일수록 삶의 만족도가 높음
* 혼인상태: 배우자가 있는 사람이 삶의 만족도가 높음
* 취업여부: 미취업자에 비해 취업자가 삶의 만족도가 높음

1. 결론
   1. 분석 결과 요약

데이터 분석의 목적은 MZ세대 즉 2·30대의 삶의 만족도에 영향을 미치는 요인을 파악하여 만족도가 낮은 사람에게 개선할 수 있는 방법을 제안해주고 결과적으로 삶의 만족을 높이고자 하는데 있다. 각 요인들이 삶의 만족도의 미치는 영향의 정도를 분석한 결과를 살펴보면 다음과 같다.

연령 또는 성별과 같이 바꿀 수 없는 요인은 삶의 만족도에 많은 영향을 미치지 못했다. 반대로 학력, 취업 여부와 같은 경제활동에 직결되는 요인들이 많은 영향을 미쳤다. 하지만 월 임금, 종사하고 있는 직종, 근무시간과 같은 직접적인 직무 요인은 삶의 만족도에 대한 영향력이 낮은 것으로 나타났다. 또한 배우자가 있는 사람들이 비교적으로 높고, 고연령일수록 높은 것으로 보아 안정적인 삶을 살아가는 사람들이 만족도가 높다는 것을 알 수 있었다.

* 1. 분석의 한계

한국노동패널데이터를 활용하여 분석하는 과정에서 다음과 같은 한계점이 나타났다.

첫째, 삶의 만족을 중점으로 조사된 데이터가 아니기 때문에 영향을 미치는 변수가 다양하고 사람마다 느끼는 정도가 주관적인 삶의 만족도를 정확하게 예측하는 것은 한계가 존재했다.

둘째, 설문조사로 취합된 데이터이기 때문에 중간 수치의 값으로 선택된 데이터가 많았다. 특히 삶의 만족도에 대한 데이터는 보통으로 선택 된 데이터가 대부분이어서 높고 낮음의 차이를 명확하게 분석하는 데에 한계가 존재하였다.

* 1. 활용방안

분석한 결과를 바탕으로 개인의 조건에 따라 만족도 향상을 위한 방법을 제시하는 서비스를 개발하고자 한다.

향후에는 현재 분석 대상인 MZ세대뿐만 아니라 세대를 넓혀 X세대, 은퇴를 바로 앞둔 베이비붐 세대를 위한 삶의 만족도 분석 서비스를 개발하고자 한다.

추가적으로 분석하는 데이터의 범위를 넓혀 취미 관련 빅데이터 분석을 통해 MZ세대의 개인 환경에 따라 선호도가 높을 것으로 예상되는 취미를 추천해주는 서비스를 개발하고자 하는 계획이 있다.

1. 참고문헌

선우지예(2022), 공공부문 종사자의 세대별 직무만족 비교분석 - 직무만족도 비교 및 직무만족 결정요인 분석, 서울대학교 행정대학원 행정학 석사 학위논문