**빅데이터 분석 결과 보고서**

(“MZ세대를 위한 거주지 추천 서비스”)

**WhereHouse**

**한준원, 왕승철**

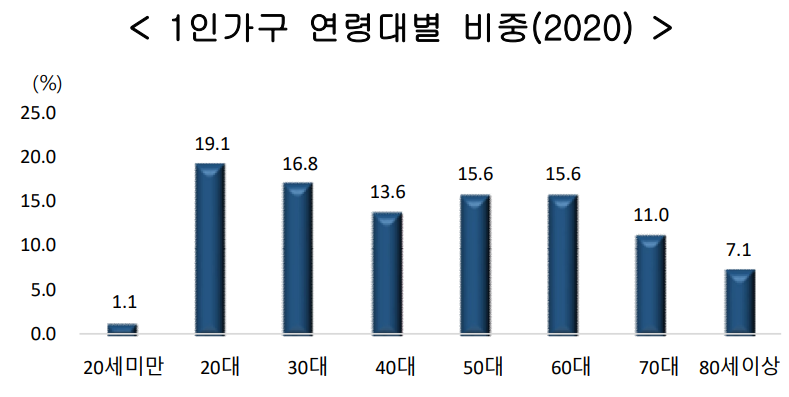
**목차**

1. **분석 기획 (Planning)  
   1.1 기획 의도  
   1.2 분석 목표**
2. **데이터 준비 (Preparing)  
   3.1 데이터 수집  
   3.2 데이터 정합성 검정  
   3.3 데이터 전처리**
3. **데이터 분석 (Analyzing)  
   3.1 분석 과정  
   3.2 상관 분석 결과  
   3.3 결과 해석**
4. **활용방안**
5. **참고문헌**

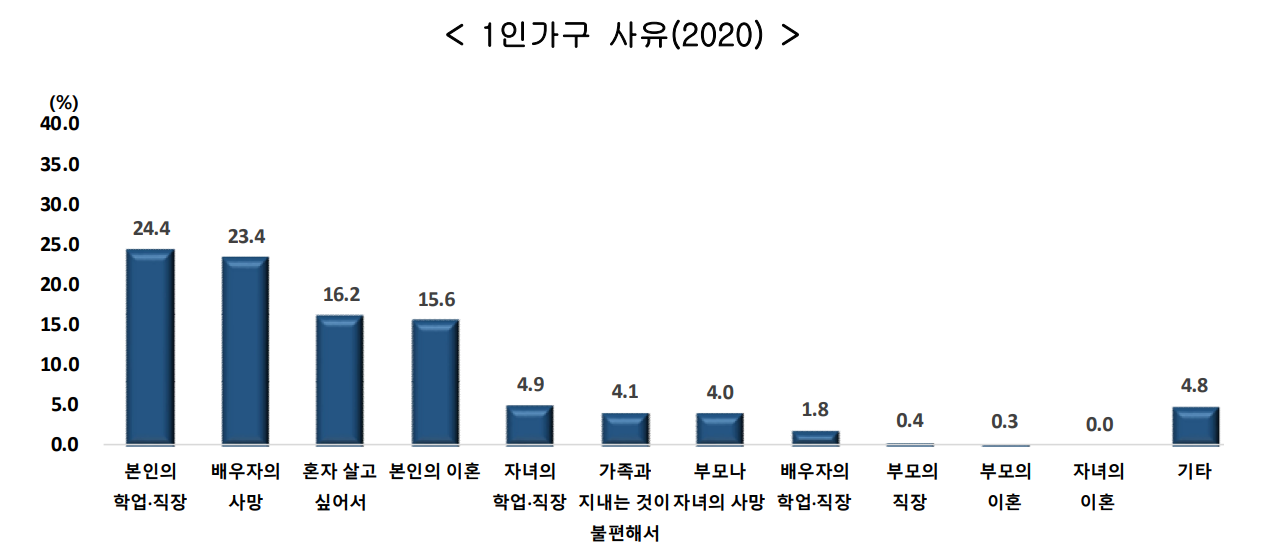
**1. 분석 기획**

**1.1 기획 의도**

최근 우리 사회는 전체 가구 유형 중 1인가구가 가장 큰 비중을 차지함과 동시에 MZ세대가 속한 20대의 1인 가구율이 가장 높은 수치를 기록하고 있다.

****

그들의 1인가구 사유는 [학업,직장]이 24.4%로 가장 많은 걸 알 수 있고, 1인 가구 지역별 분포도를 보면 서울의 비중이 가장 높기 때문에 서울에 있는 지역구를 거주지로 추천해주는 서비스를 MZ세대들에게 제공한다. 각 지역 별 가격 데이터와 편의 시설 데이터들을 수집하여 분석하고, 평균 가격과 편의성의 정도를 종합하여 각자 입맛에 맞게 거주지를 선택할 수 있도록 도움을 제공한다.



**1.2 분석 목표**

편의성이란 지표는 정성적인 데이터이기 때문에 정량적인 데이터가 존재하지 않아   
편의성과 연관도가 높은 지표를 찾아 임의적으로 수식을 선정해야 한다.

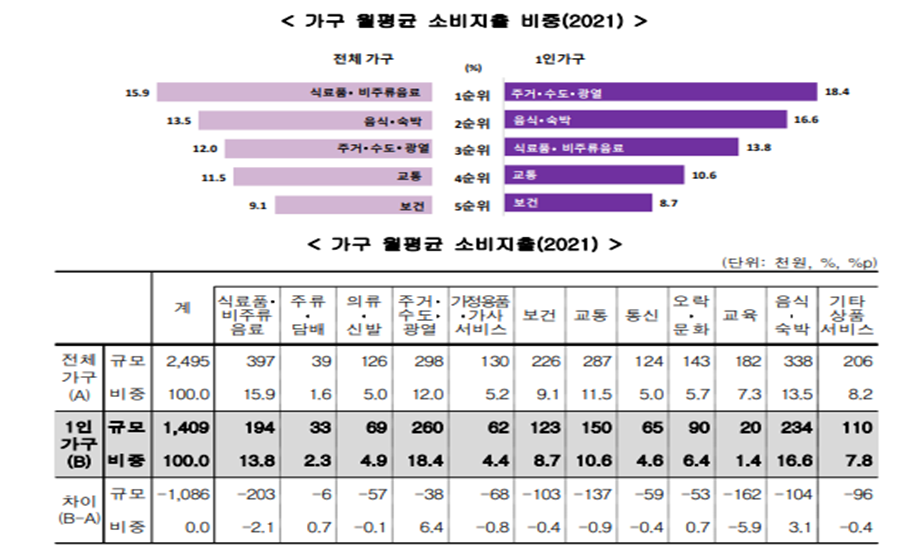
먼저 1인가구의 월평균 소비지출 통계를 보면 1위가 주거, 수도, 2위가 음식, 숙박, 3위가 식료품, 비주류 음료, 4위가 교통, 5위가 안전인 것을 알 수가 있고, 1인 가구의 어려움 통계표를 보면, 균형 잡힌 식사가 어려움이 42.4%로 가장 높은 것을 확인할 수가 있다. 따라서, 위 통계를 기반으로 1인가구 월 평균 소비지출 통계를 인용하여 선정한 수식과,음식과 관련된 데이터들인 음식점 데이터, 편의점, 카페 데이터, 교통과 관련된 데이터인 지역구 별 지하철 데이터 등을 분석하여 이들간의 상관 관계를 분석해 MZ세대들에게 제공 해줄 수 있는 유의미한 결과를 도출하려고 한다.

또한, 편의 시설 데이터와 MZ세대 1인 가구 비율 데이터를 통해 편의 시설이 MZ세대들의 거주지 선택에 있어 어느 정도의 영향을 미치는지 분석한 후, 이를 통해 종합 점수 선정에 적합성을 판단하도록 한다.

**2. 데이터 준비**

**2.1 필요 데이터 정의 및 수집**

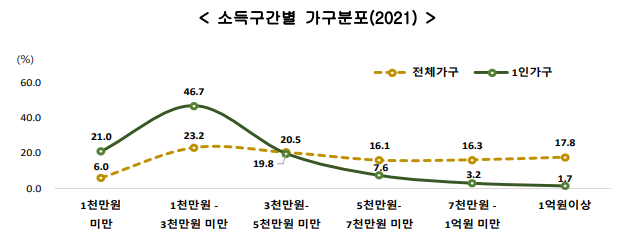
**편의시설 데이터**

편의시설 데이터는 MZ세대 1인가구에게 필요하다고 생각되는   
편의점, 카페, 지하철, 음식점, 다이소, 올리브영, 백화점으로 선정하였다.   
선정 기준은 가구 월 평균 주요 소비 지출에서 볼 수 있듯, 음식, 숙박과 관련된 음식점, 식료품, 비주류 음료와 관련된 편의점, 카페   
그리고, 교통과 관련된 지하철, 마지막으로 의류, 신발과 기타와 관련된 다이소, 올리브영, 대규모점포(백화점, 대형마트) 등을 선정하였다.

* 음식점 : 서울시 일반음식점 인허가 정보.csv(서울 열린 데이터 광장)
* 편의점, 카페 : 서울시 휴게 음식점 인허가 정보.csv(서울 열린 데이터 광장)
* 지하철 : 서울 교통 공사 지하철역 주소 및 전화번호 정보.csv(서울 열린 데이터 광장)
* 다이소, 올리브영 : 각 회사 홈페이지 구 별 점포 정보(엑셀 수기 작성)
* 대규모 점포 (백화점, 대형마트) : 서울시 대규모 점포 인허가 정보.csv(서울 열린데이터 광장)

**월세, 전세 데이터**

주머니 사정이 넉넉치 않은 MZ세대 1인 가구의 경우 거주지 선택 시  
편의성 뿐 아니라 월세(보증금, 월세), 전세 가격도 매우 중요한 선택의 지표 중 하나이다. 따라서 MZ세대 거주지 추천을 위해서 월세, 전세 각각의 평균 가격과   
실제 거주하고 있는 MZ세대 1인 가구 간의 상관관계를 분석하기 위해   
월세, 전세 실제 거래가 데이터를 사용하고자 한다.



* 서울시 부동산 전월세가 정보.csv(서울 열린 데이터 광장)

**1인가구 인구, 인구 밀집도 데이터**

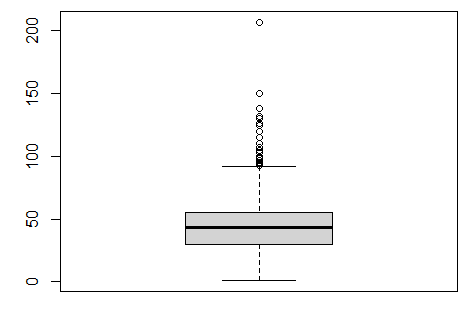
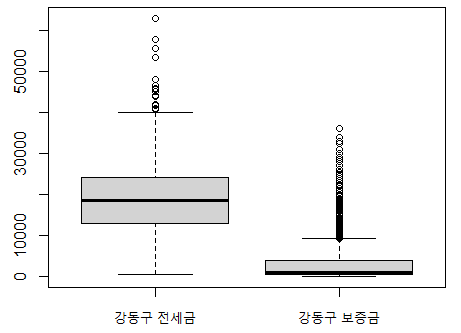
마지막으로 MZ세대 1인 가구를 대상으로 한 서비스를 위해 MZ 세대 1인 가구가 가장 많이 거주하고 있는 지역과 위에서 열거한 데이터들의 상관관계 분석을 위해 1인가구 인구에서의 MZ세대 1인 가구 거주 비율을 사용하여 이를 실현하고자 한다.

* 서울시 1인가구(연령별) 통계.xlsx

**2.2 데이터 정합성 검정**

분석을 위해 데이터 정합성 검정을 수행하였다.

전세, 보증금, 월세 데이터에서 이상치 값을 boxplot을 통해 다수 확인할 수 있었다.

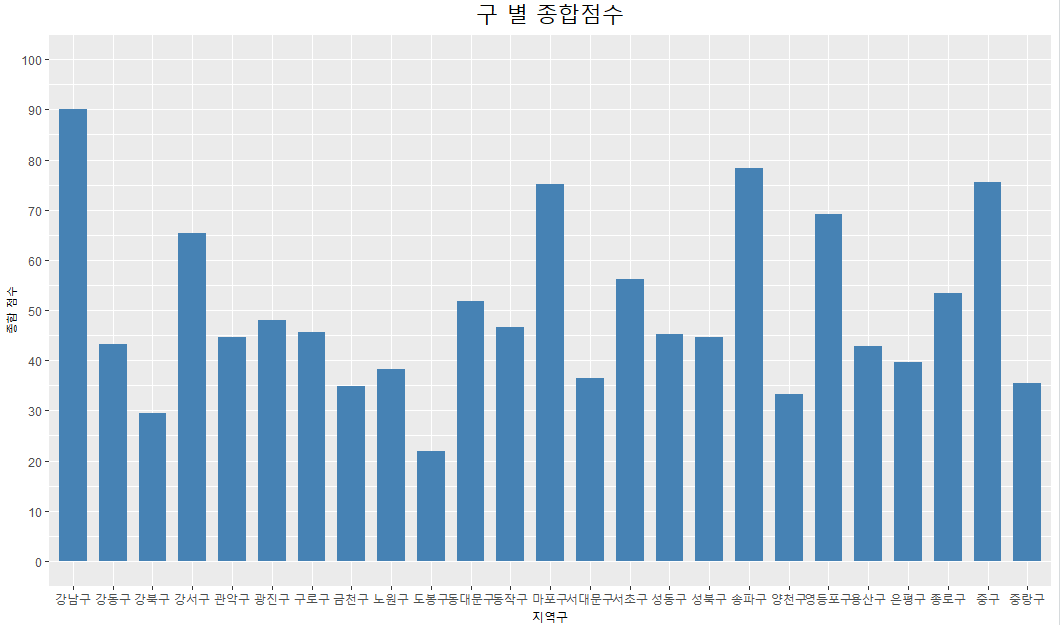


이를 제외한 데이터를 통해 전세금, 보증금, 월세 데이터를 이용하여 분석하도록 하며,  
또한, 구 별 지하철 역 개수가 실제 개수와 데이터 상의 개수가 상이한 것을 확인할 수 있었다. 따라서, 이를 올바른 데이터로 수정한 후 분석에 이용하도록 한다.

**2.3 데이터 전처리  
편의점, 카페, 음식점 데이터**  
해당 데이터의 경우 폐업된 가게가 데이터에 포함되어 있는 것을 확인하여,  
정상 영업중인 점포만 사용하도록 정제하였으며,  
구 별 점포 개수를 카운트하여 data.frame형태로 만들어 사용한다.  
  
**MZ세대 1인가구 비율 데이터**해당 데이터의 경우 20~69세 이상까지 분류되어 있는 데이터이기 때문에  
우리의 타겟인 MZ세대를 20~35세까지로 책정하고 전체 1인가구 인구 대비  
MZ세대 인구 비율을 새롭게 만들어 사용한다.

**3. 데이터 분석**

**3.1 분석 과정**  
1인가구의 월평균 소비지출 통계를 참고하여 선정한 수식을 통해 각 구의 편의시설 (편의점, 카페, 음식점, 지하철, 다이소, 올리브영, 쇼핑몰) 개수를 종합 점수 100점이 되도록 환산하여 이를 사용하도록 하였으며, 수식은 아래와 같다.

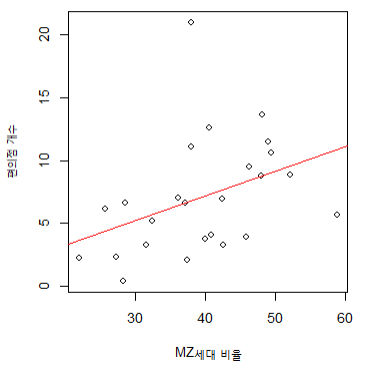


구 별 종합점수를 막대그래프로 시각화 하면 위와 같다.

**3.2 상관 분석 결과**

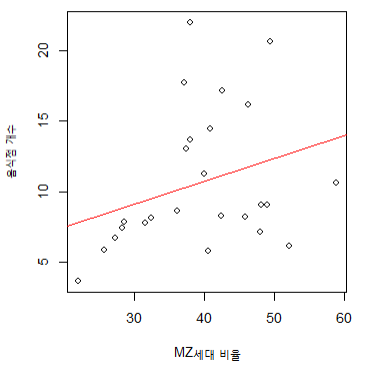
**1) MZ세대 비율과 편의점**

|  |  |
| --- | --- |
| 상관계수 | 0.3876116 |

MZ세대 1인가구 비율과 편의점 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 편의점 개수는  
MZ세대 거주 비율에 0.3876116만큼   
약한 양의 상관관계에 있는 것을 확인할 수 있다.

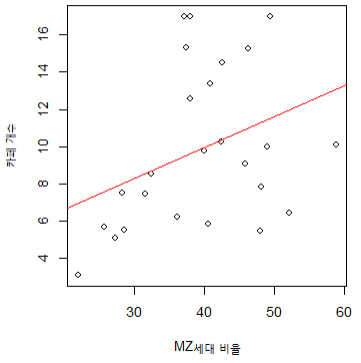
**2) MZ세대 비율과 음식점**

|  |  |
| --- | --- |
| 상관계수 | 0.302745 |

MZ세대 1인가구 비율과 카페 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 음식점 개수는  
MZ세대 거주 비율에 0.302745만큼   
약한 양의 상관관계에 있는 것을 확인할 수 있다.

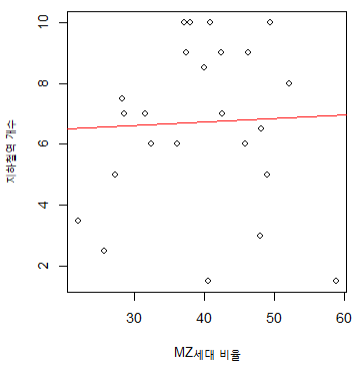
**3) MZ세대 비율과 카페**

|  |  |
| --- | --- |
| 상관계수 | 0.3559358 |

****MZ세대 1인가구 비율과 카페 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 카페 개수는  
MZ세대 거주 비율에 0.3559358만큼   
약한 양의 상관관계에 있는 것을 확인할 수 있다.

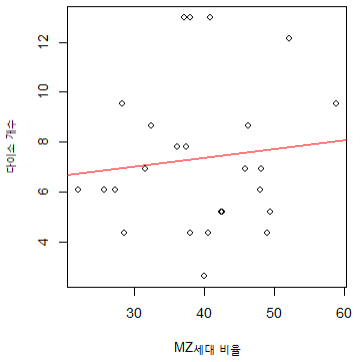
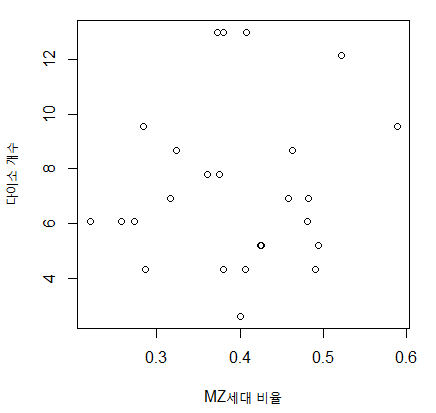
**4) MZ세대 비율과 지하철**

|  |  |
| --- | --- |
| 상관계수 | 0.03813673 |

****MZ세대 1인가구 비율과 카페 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 지하철역 개수는  
MZ세대 거주 비율에 0.03813673으로  
거의 상관없는 것을 확인할 수 있다.

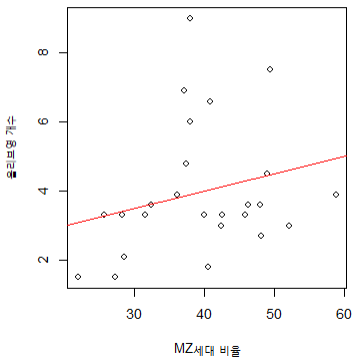
**5) MZ세대 비율과 다이소**

|  |  |
| --- | --- |
| 상관계수 | 0.1105288 |

****MZ세대 1인가구 비율과 다이소 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 올리브영 개수는  
MZ세대 거주 비율에 0.1105288으로  
지하철역 개수와 마찬가지로  
거의 상관없는 것을 확인할 수 있다.

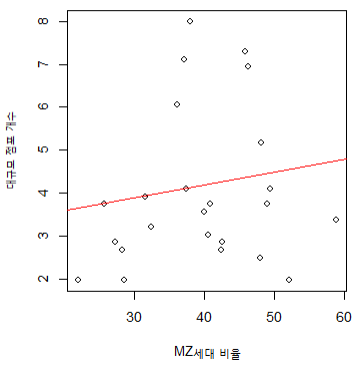
**6) MZ세대 비율과 올리브영**

|  |  |
| --- | --- |
| 상관계수 | 0.2446565 |

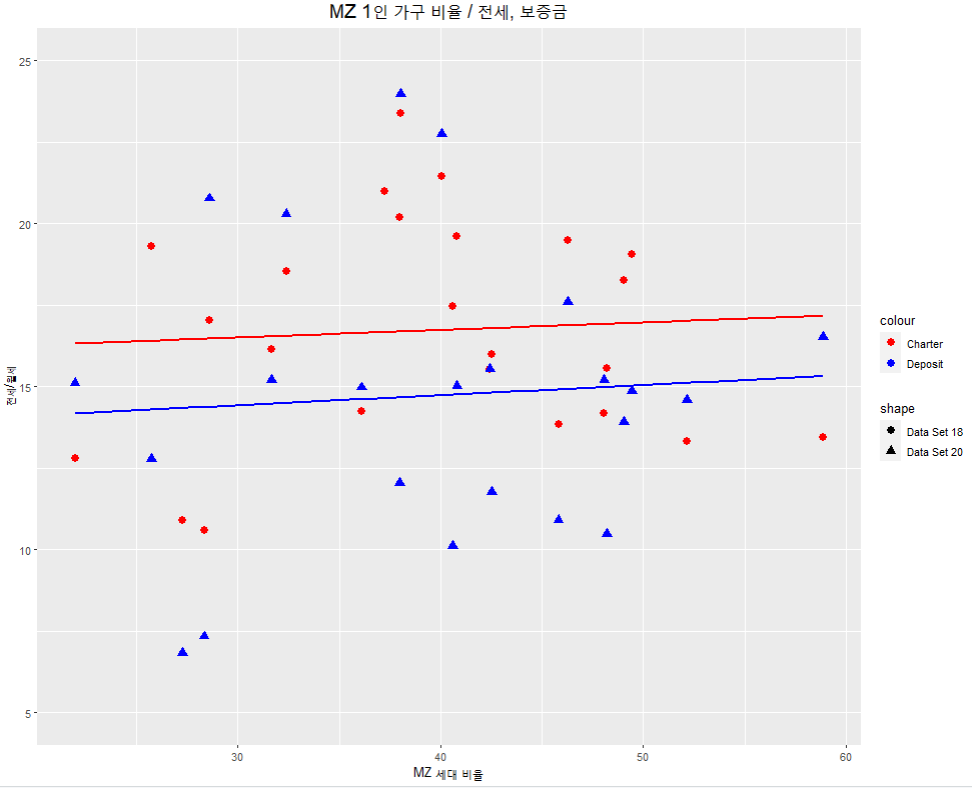
****MZ세대 1인가구 비율과 올리브영 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 올리브영 개수는  
MZ세대 거주 비율에 0.2446565만큼  
약한 양의 상관관계에 있는 것을 확인할 수 있다.

**7) MZ세대 비율과 대규모 점포(백화점, 대형마트)**

|  |  |
| --- | --- |
| 상관계수 | 0.142311 |

****MZ세대 1인가구 비율과 대규모 점포 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 대규모 점포 개수는  
MZ세대 거주 비율에 0.142311으로  
거의 상관없는 것을 확인할 수 있다

**8) MZ세대 비율과 전세, 보증금**

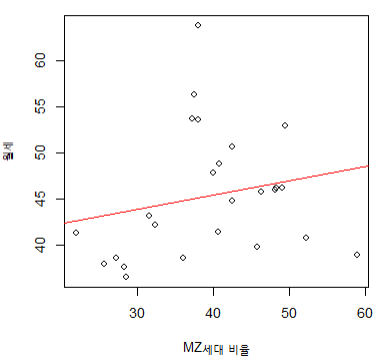
****MZ세대 1인가구 비율과 전세, 보증금(월세)을 통해 아래와 같은 산점도와 상관계수를   
얻을 수 있었다. 상관 분석을 통해 전세, 보증금(월세)는 MZ세대 거주 비율에 거의 상관없는 것을 확인할 수 있다,

|  |  |
| --- | --- |
| 전세 상관계수 | 0.03522923 |
| 보증금 상관계수 | 0.01032688 |

오른쪽 산점도는  
전세(빨간색 동그라미)는 1000만원 단위, 월세(파란색 세모)는 100만원 단위로   
단위를 맞춰서 표시하였다.

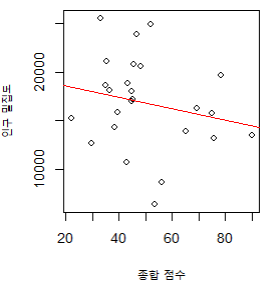
**9) MZ세대 비율과 월세**

|  |  |
| --- | --- |
| 상관계수 | 0.2099976 |

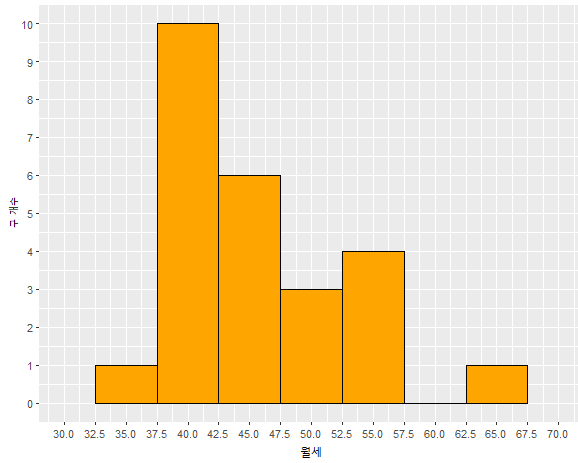
MZ세대 1인가구 비율과 월세를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 월세 가격은  
MZ세대 거주 비율에 0.2099976으로  
앞서 본 전세, 보증금보다는   
더 상관이 있는 것을 확인할 수 있다.

**10) 인구밀집도와 종합 점수**

|  |  |
| --- | --- |
| 상관계수 | -0.213787 |

****마지막으로 구 별 인구밀집도와 종합 점수를 통해  
오른쪽과 같은 산점도를 얻을 수 있었다.  
이는 음의 상관관계로 인구밀집도가 높으면  
종합 점수는 떨어지는 반비례 관계를 의미한다.

**11) 구 평균 월세 가격 히스토그램**

****주머니 사정이 좋지 않은 MZ세대 대다수를 위해서 1인가구 최소 주거 면적으로 지정된 (약 4.2평) 에서부터 1인당 주거면적 평균인 (약 10평) 까지 구 별 월세 가격 평균을 히스토그램으로 나타낸 것이다.

|  |  |
| --- | --- |
| **월세 범위 (만원)** | **구 개수** |
| 32.5~37.5 | 1개 |
| 37.5~42.5 | 10개 |
| 42.5~47.5 | 6개 |
| 47.5~52.5 | 3개 |
| 52.5~57.5 | 4개 |
| 57.5~62.5 | 0개 |
| 62.5~67.5 | 1개 |

**3.3 결과 해석**

먼저, MZ세대 1인가구 거주 비율과 편의시설들 간의 상관 분석을 통해 알 수 있듯  
거주지 선택에 엄청난 영향을 미치지는 않지만, 아예 상관이 없다고는 보기 힘든 것을 확인해볼 수 있으며 특히, 편의점, 카페, 음식점 3개와는 0.3 이상으로 해당 시설들이   
MZ세대 1인가구원들의 생활만족도에 어느정도 영향을 끼치는 것을 알 수 있다.

또한, 전세, 보증금, 월세의 경우 전세와 보증금은 거의 상관이 없지만,  
월세는 약한 양의 상관 관계를 가지는 것을 확인할 수 있었다.  
이를 통해 유추해볼 수 있는 것은 전세, 보증금의 경우는 돌려받을 수 있는 비용이지만,  
월세의 경우는 주기적으로 지출하는 금액으로 앞서 보았던 1인가구 소비 지출 1순위인 주거인 것을 고려한다면, 월세는 거주지 선택에 어느정도 영향을 미친다고 볼 수 있다.

마지막으로, 인구밀집도와 종합 점수 간의 상관 분석을 통해  
인구 밀집도가 높으면, 종합 점수가 떨어지는 것을 확인할 수 있었다.  
종합 점수는 구 별 편의 시설 개수를 통해 산정된 결과이므로, 개수에 비해 인구밀집도가 높다면, 이용하는데 불편함이 있기 때문에 이러한 결과가 도출된 것이라고 추측해볼 수 있을 것이다.

**4. 활용방안**

지역구 별 평균 전세/월세 가격 정보를 제공하기 때문에 MZ세대들이 실제 거래 시 해당 지역구의 평균 가격을 인지하고 좀 더 효율적인 거래를 할 수 있다.

분석한 데이터들을 기반으로 거주지 선택에 유용한 각 지역별 정보들을 제공하기 때문에MZ세대들이 거주지 선택을 함에 있어 시간과 비용을 절약 할 수 있다.

1인가구, MZ세대들의 소비 동향을 보면 전체 가구 대비 상대적으로 음식,숙박에 대한 지출 비중이 높고, 음식, 식료품이 삶의 만족도에서 큰 부분을 차지한다.  
이러한 MZ세대들의 소비 패턴을 고려하여 편의시설 별 점수 가중치를 다르게 두고 산정했기 때문에 보다 더 효율적인 소비를 할 수 있게끔 도움을 줌과 동시에 음식점, 편의점 등 식 관련 데이터를 그래프로 시각화하여 보여줌으로써 거주하고 싶은 지역구를 선택하는데 도움을 주게끔 활용할 수 있다.

**5. 참고문헌**

통계청, “2022년 통계로 보는 1인 가구”, 2021

통계청, “가계 동향 조사”

통계청, “사회 조사”

통계청, “인구 주택 총 조사”

여성 가족부, “가족 실태 조사”

국토 교통부, “주거 실태 조사”

문화 체육 관광부, “국민 여가 활동 조사”

서울특별시 디지털 정책관, “서울시 생활환경 만족도 통계”, 2022

정순둘, 성민현, “연령 집단 별 사회적 자본과 삶의 만족도 관계 비교”, 2012, 249-272