**빅데이터 분석 결과 보고서**

“MZ세대를 위한 생활 안전 / 생활 편의 기반

거주지 추천 서비스”

**WhereHouse**

**한준원, 왕승철**

**목차**

1. **분석 기획 (Planning)  
   1.1 문제 정의  
   1.2 기획 의도**
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 분석 결과**
4. **활용방안**
5. **참고문헌**

**1. 분석 기획**

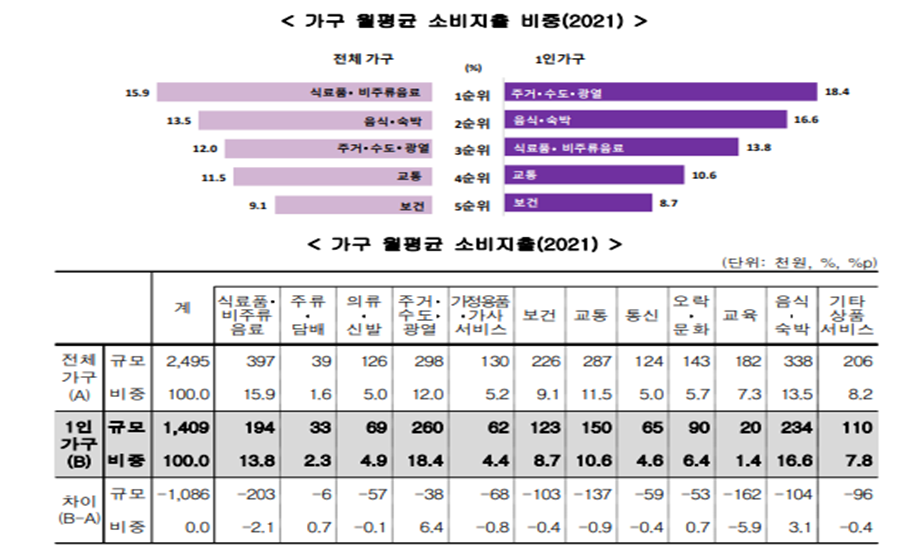
1. **문제 정의**
2. **기획 의도**

**2. 데이터 준비**

**2.1 필요 데이터 정의 및 수집**

**- 편의시설 데이터**

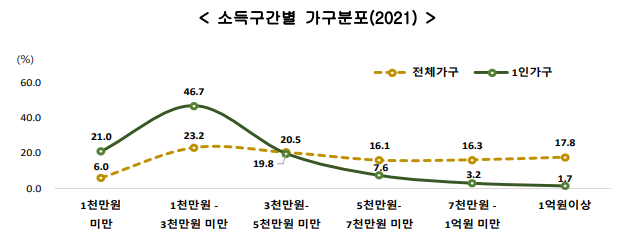
편의시설 데이터는 MZ세대 1인가구에게 필요하다고 생각되는   
편의점, 카페, 지하철, 음식점, 다이소, 올리브영, 백화점으로 선정하였다.   
선정 기준은 가구 월 평균 주요 소비 지출에서 볼 수 있듯, 음식, 숙박과 관련된 음식점, 식료품, 비주류 음료와 관련된 편의점, 카페   
그리고, 교통과 관련된 지하철, 마지막으로 의류, 신발과 기타와 관련된 다이소, 올리브영, 대규모점포(백화점, 대형마트) 등을 선정하였다.



* 음식점 : 서울시 일반음식점 인허가 정보
* 편의점, 카페 데이터 : 서울 휴게 음식점 인허가 정보
* 지하철 : 서울 교통 공사 자치구별 지하철역 정보
* 다이소, 올리브영 : 각 회사 홈페이지 구 별 점포 정보
* 대규모 점포 (백화점, 대형마트) : 서울시 대규모 점포 인허가 정보

**- 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

**- 1인가구 인구, 인구 밀집도 데이터**

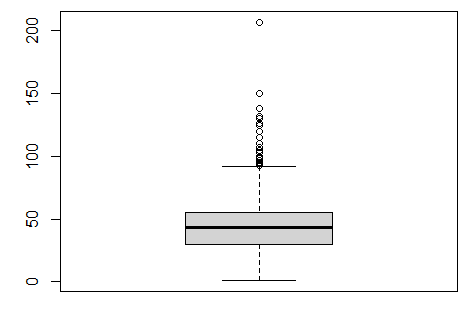
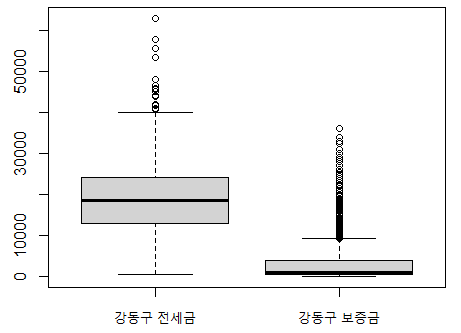
마지막으로 MZ세대 1인 가구를 대상으로 한 서비스를 위해 MZ 세대 1인 가구가 가장 많이 거주하고 있는 지역과 위에서 열거한 데이터들의 상관관계 분석을 위해 1인가구 인구에서의 MZ세대 1인 가구 거주 비율을 사용하여 이를 실현하고자 한다.

* 1인가구(연령별) 데이터

**2.2 데이터 정합성 검정**

분석을 위해 데이터 정합성 검정을 수행하였다.

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



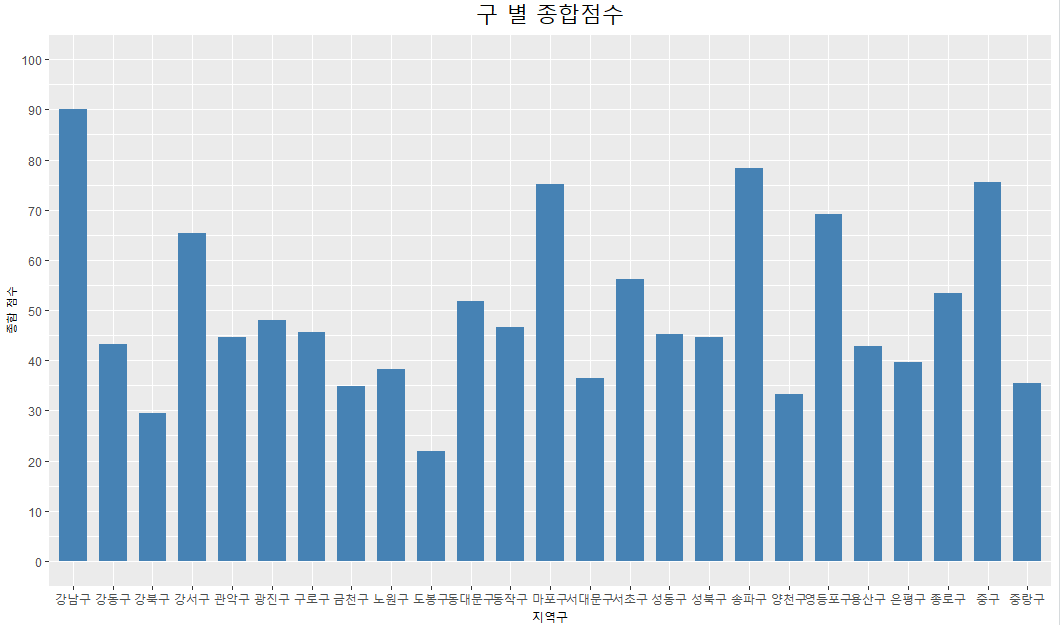
이를 제외한 데이터를 통해 전세금, 보증금, 월세 데이터를 이용하여 분석하도록 하며,  
또한, 구 별 지하철 역 개수가 실제 개수와 데이터 상의 개수가 상이한 것을 확인할 수 있었다. 따라서, 이를 올바른 데이터로 수정한 후 분석에 이용하도록 한다.

**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**

먼저 편의성이란 정량적인 데이터가 존재하지 않으며, 정성적인 데이터이므로,  
1인가구의 월평균 소비지출 통계를 통해 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.  
종합 점수 :

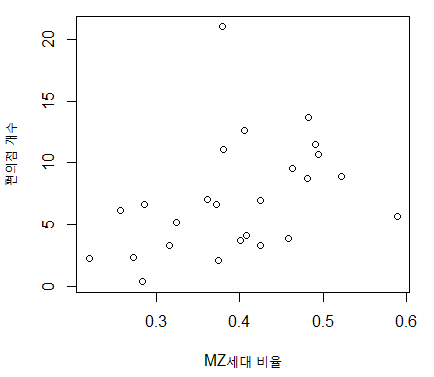


구 별 종합점수를 막대그래프로 시각화 하면 위와 같다.

**3.2 상관 분석**

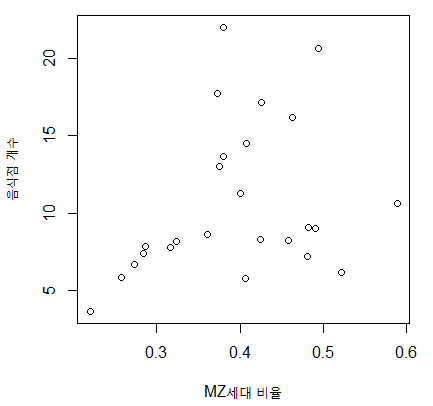
**1) MZ세대 비율과 편의점**

|  |  |
| --- | --- |
| 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.

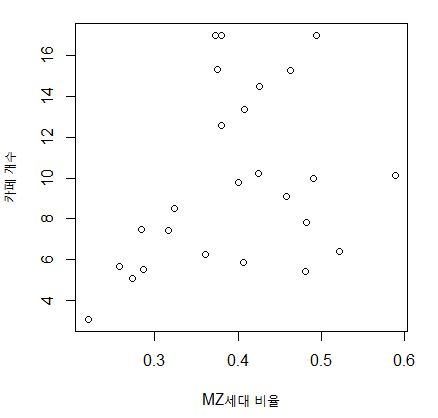
**2) MZ세대 비율과 음식점**

|  |  |
| --- | --- |
| 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.

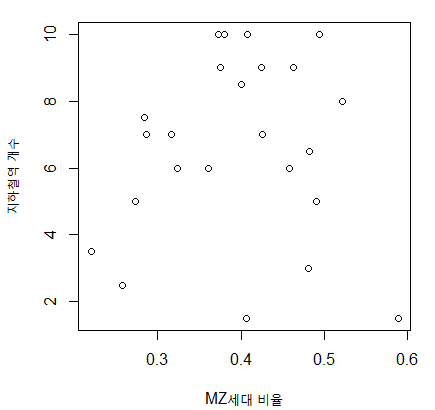
**3) MZ세대 비율과 카페**

|  |  |
| --- | --- |
| 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.

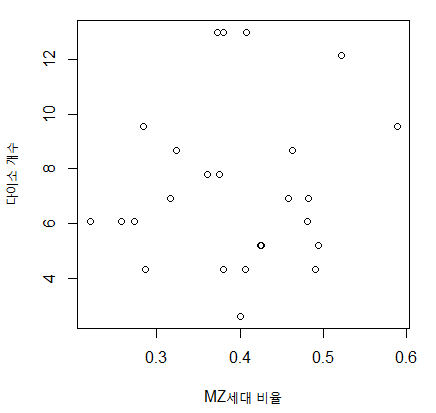
**4) MZ세대 비율과 지하철**

|  |  |
| --- | --- |
| 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.

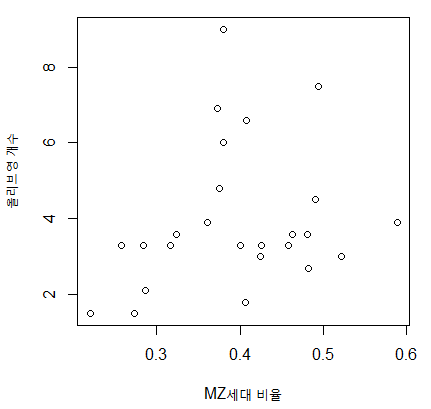
**5) MZ세대 비율과 다이소**

|  |  |
| --- | --- |
| 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.

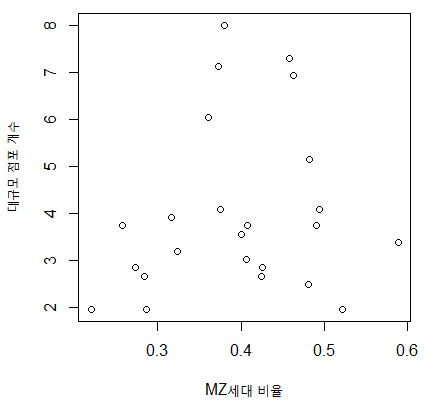
**6) MZ세대 비율과 올리브영**

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

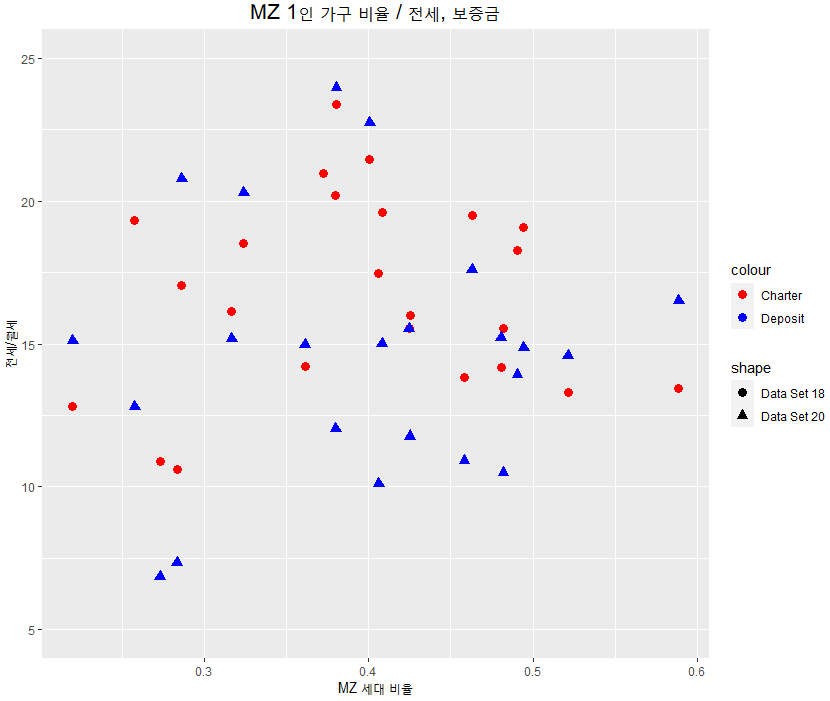
****MZ세대 1인가구 비율과 올리브영 개수를 통해   
오른쪽과 같은 산점도를 얻을 수 있었다.  
상관 분석을 통해 올리브영 개수는  
MZ세대 거주 비율에 0.2446565만큼  
약한 양의 상관관계에 있는 것을 확인할 수 있다.

**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**

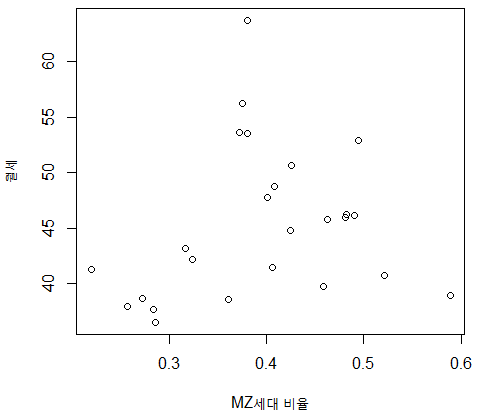
****MZ세대 1인가구 비율과 전세, 보증금(월세)를 통해 아래와 같은 산점도와 상관계수를   
얻을 수 있었다. 상관 분석을 통해 전세, 보증금(월세)는 MZ세대 거주 비율에 거의 상관없는 것을 확인할 수 있다,

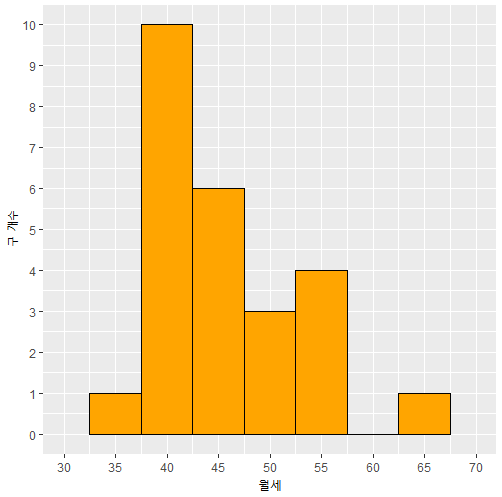
|  |  |
| --- | --- |
| 전세 상관계수 | 0.03522923 |
| 보증금 상관계수 | 0.01032688 |

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

**7) 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.

****

**4. 활용방안**

**6. 참고문헌**