# 가설 수립

## **'서울시 생활정보 기반 대중교통 수요 분석'**

자치구 인구수

도착시간, 나이, HW

많은 top3 : 송파구 강서구 강남구

적은 top3 : 중구 종로구 용산구

1. 승객 대비 노선 개수

적은 top3 : 강동구 광진구 도봉구

2. 사람수 당 이동시간

많은 top3 : 금천구 도봉구 강북구

3.

이동인구(합)과 종사자들의 총합 강한 상관관계

이 지표가 강한 지역 중에 노선 개수가 부족한 지역이 있다면

노선을 늘리는 방향

직원들의 지표 찾기 ^.^

최종 가설)

전제1 : 우리는 직장인들의 자치구 별 이동시간, 이동지역 등을 바탕으로

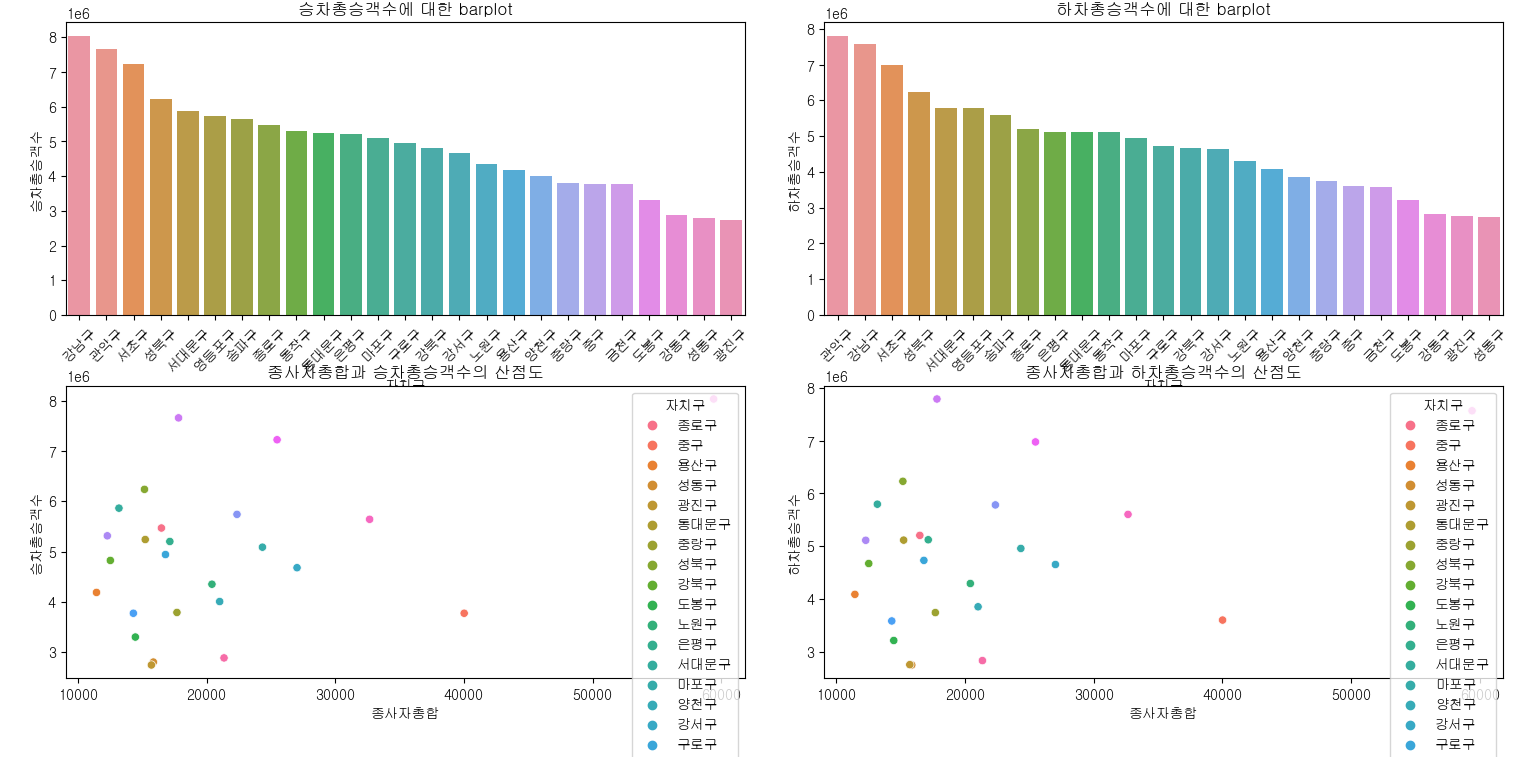
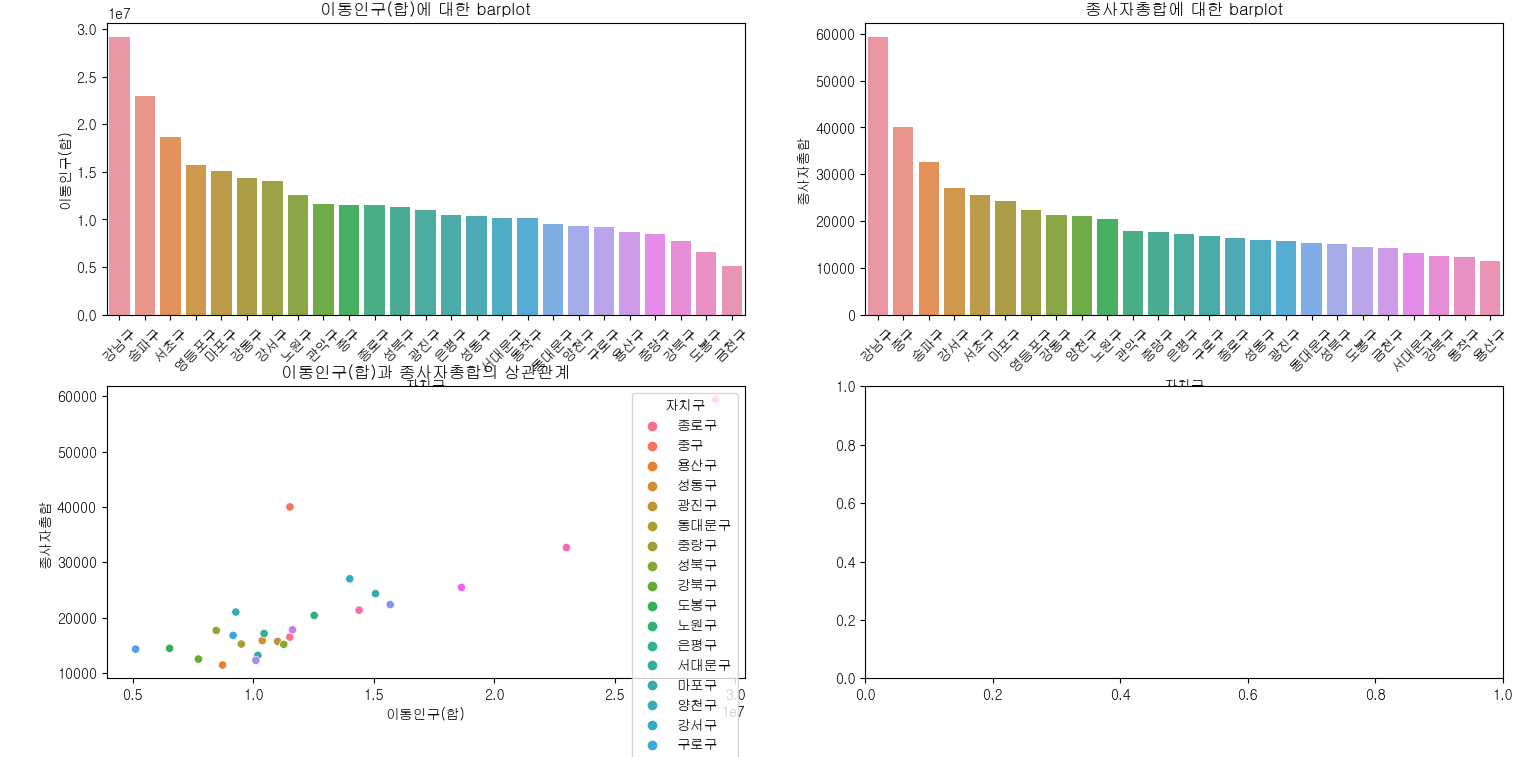
직장인들을 위한 증선이 가장 필요한 지역이 어디인지 알아보고자 한다.

전제2 : 모든 데이터는 나이 별로 조사될 것이다.

**가설1: 이동 인구의 합과 해당 지역 종사자 수와는 관련이 있을 것이다?**

가설1 계수

**가설2. 하차승객수와 해당 지역 종사자 수는 관련이 있을 것이다?**



가설2 계수



**가설3: 자치구별 직장인 (이동시간, 20-40세) 수와 노선수의 연관성,이동유형(hw,wh)**

— 직장인들의 버스 수요가 가장 높을 것

—- 가설1,가설2를 통해 자치구 별로 (평일 이동하는 지역 종사자 수와 노선수의 연관을 찾는다) 7-9 / 5-7)

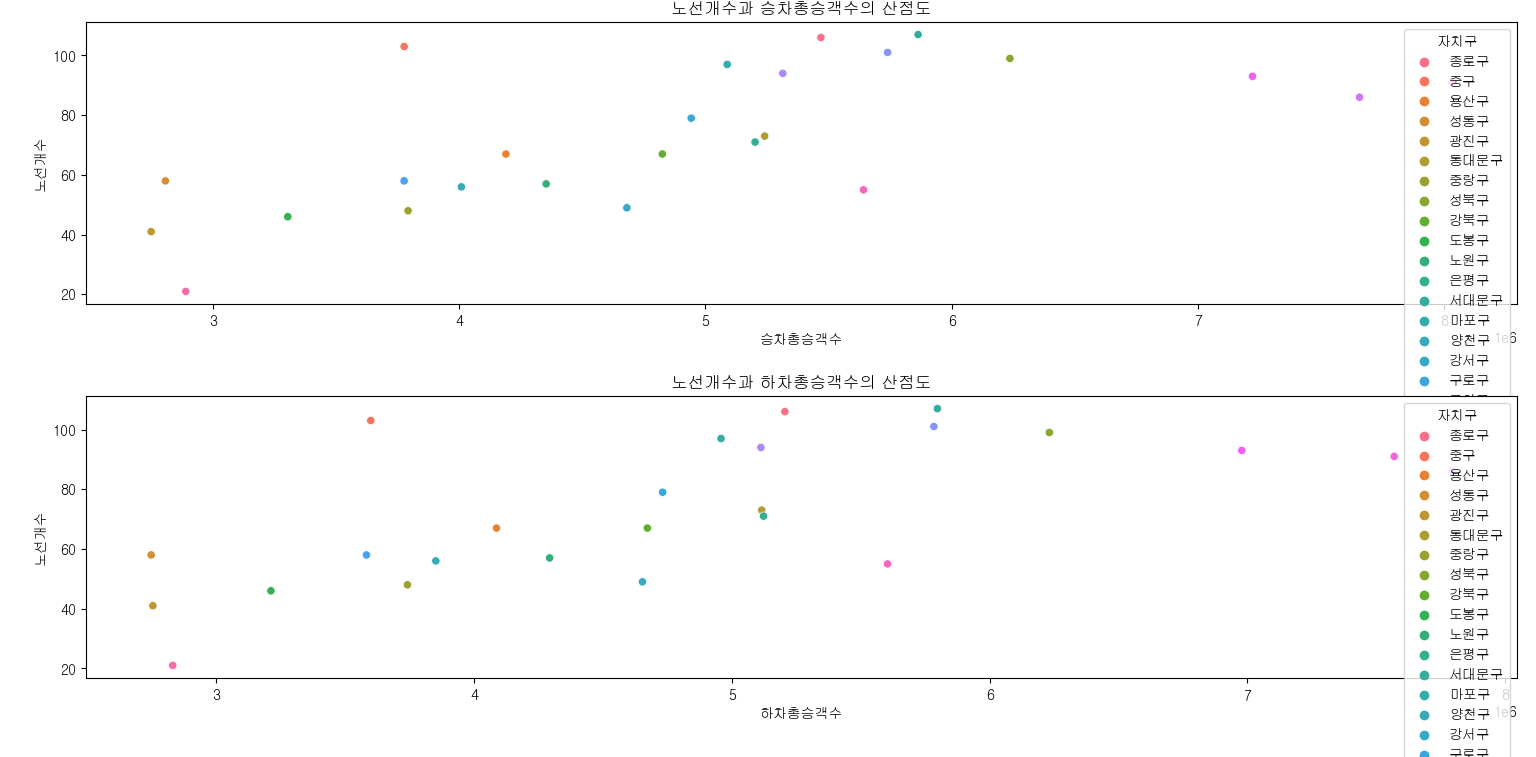
— 자치구별로 직장인의 숫자와 노선수를 비교하여 상관계수를 확인하고

가설3 계수



(근데 이 지표의 계수는 큰 의미를 둔다기 보다 인구 대비 노선 수가 많으면 좋은 거고, 부족하면 늘려야 한다는 이번 데이터 분석의 취지와 연결이 되기 때문)

**참고. 노선 개수와 승객수의 상관관계**

****

* **승차 상관관계**
* **상관계수: 0.67, p-value: 0.00028**
* **하차 상관관계**
* **상관계수: 0.65, p-value: 0.00041**

**결론. 승하차승객수와 노선수간의 연관이 있을 것이다?**

X -> Y

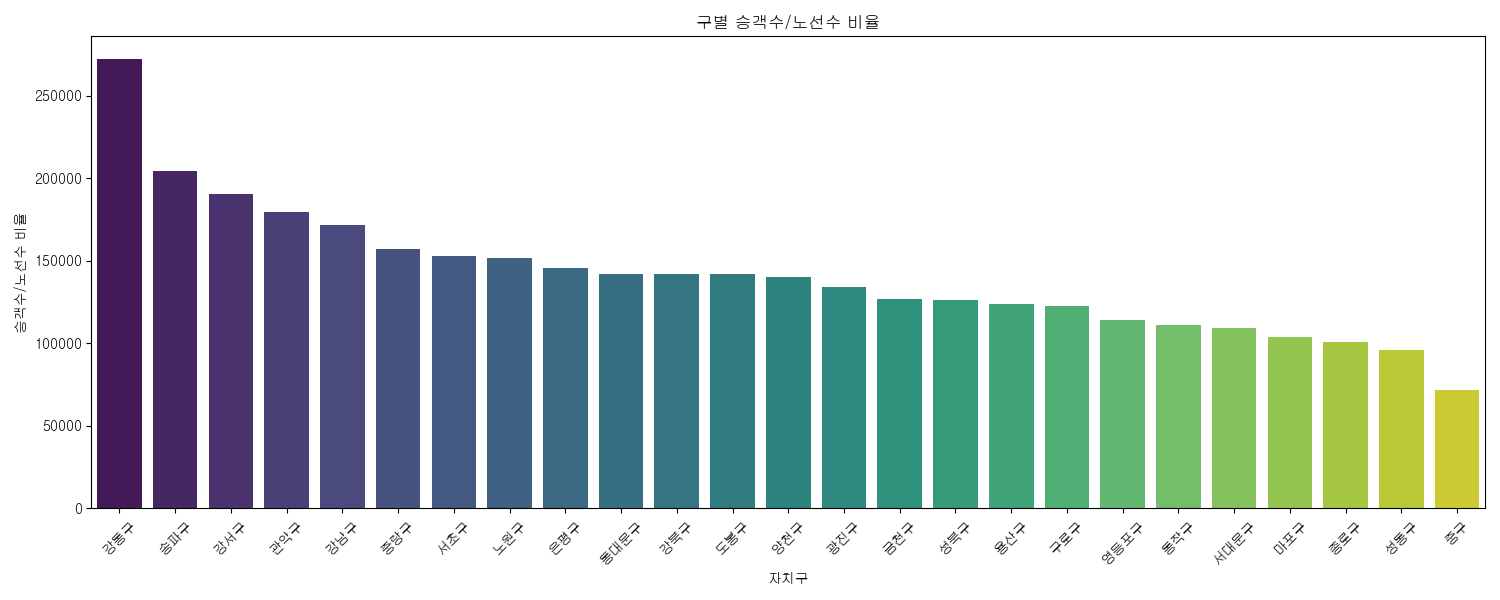
승하차가 많은데 노선수가 적으면 좀 그러니까….

자치구 직장인 나이를 얻은 상태

강남구 - 노선수는 비교적 적음

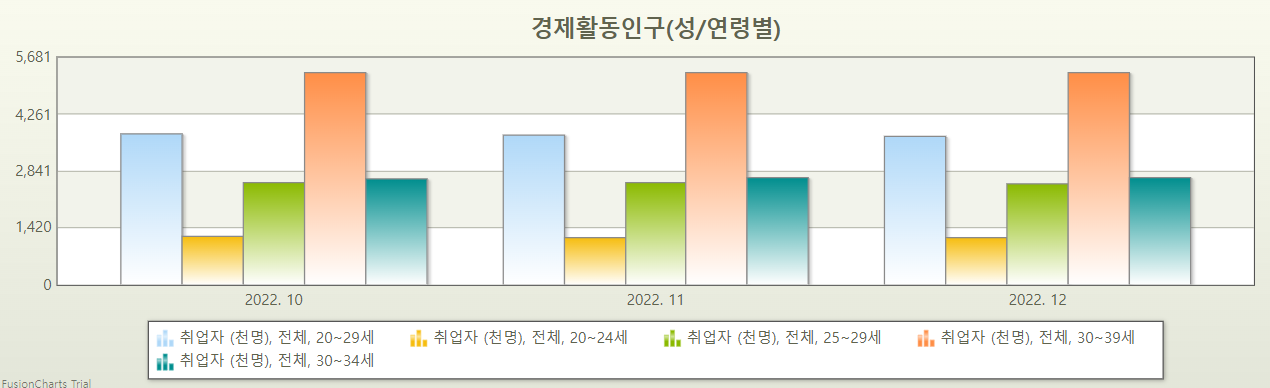
귀무가설 :

자치구별 이동 인구(합)->해당 지역 종사자 수???

**참고. 승객수 / 노선수** 

학생, 군인, 주부 등 비경제활동인구

자치구별 이동 인구의 합 - 단변량

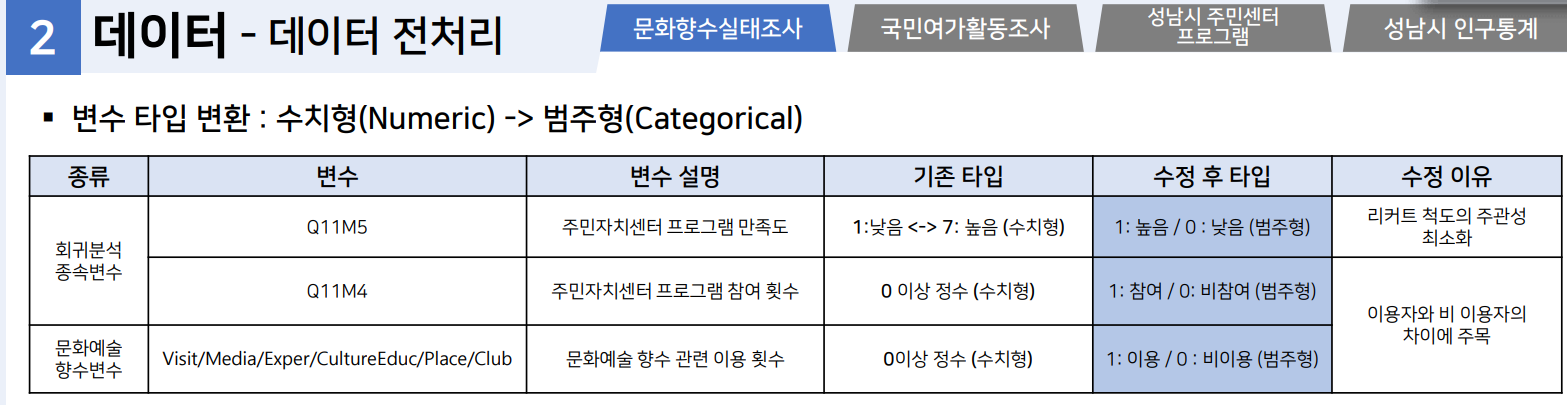


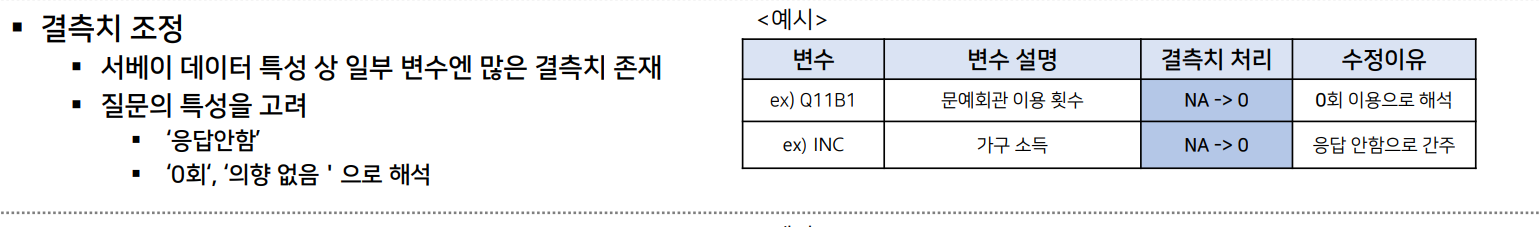
* 가설이 사실임을 증명할 가능성이 있어야 합니다.
* 가설이 거짓임을 증명할 가능성이 있어야 합니다.
* 가설의 결과는 재현 가능해야 합니다.

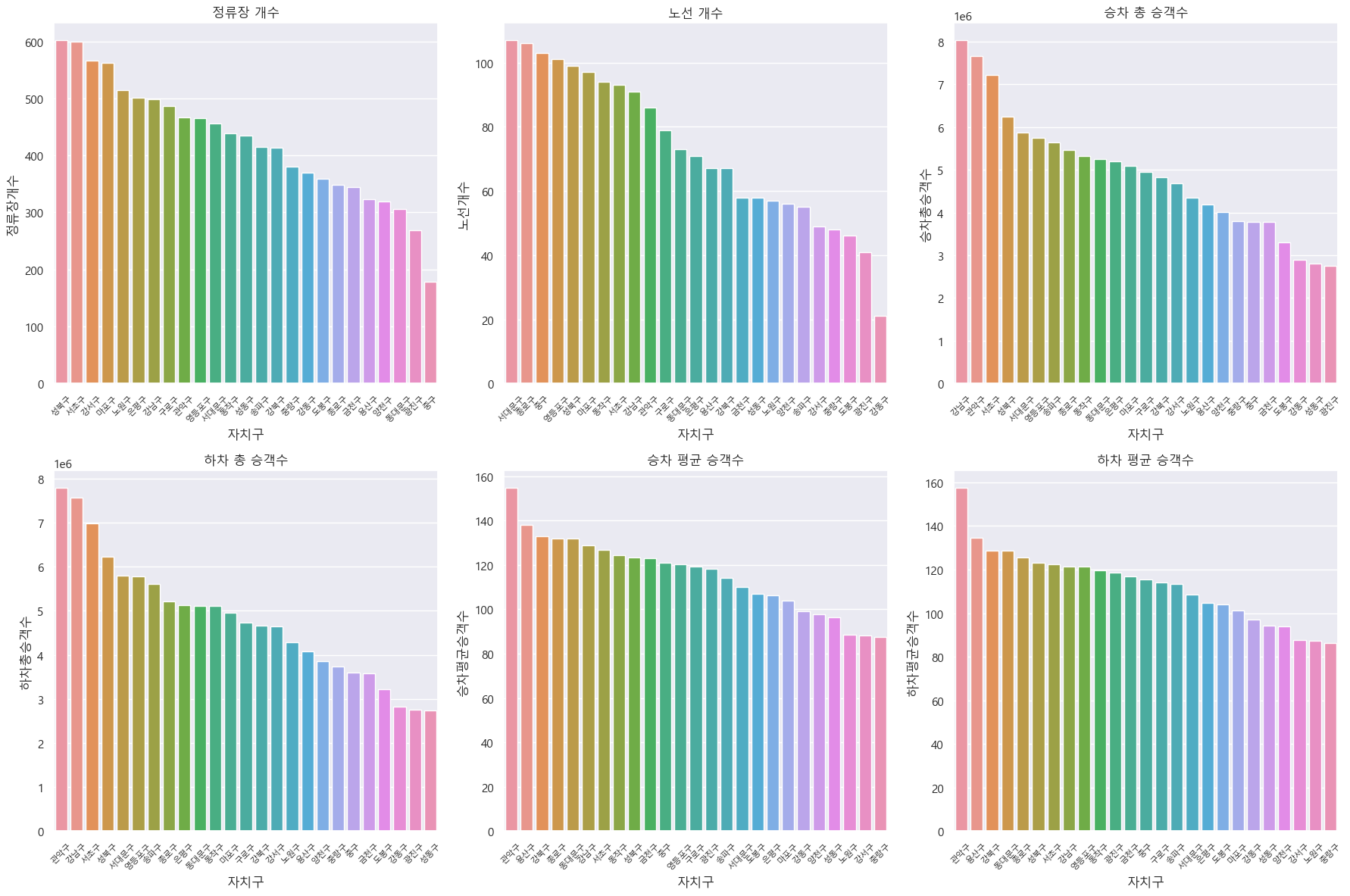
귀무가설\_(H0) : 보편적으로 옳다고 믿어지는 가설

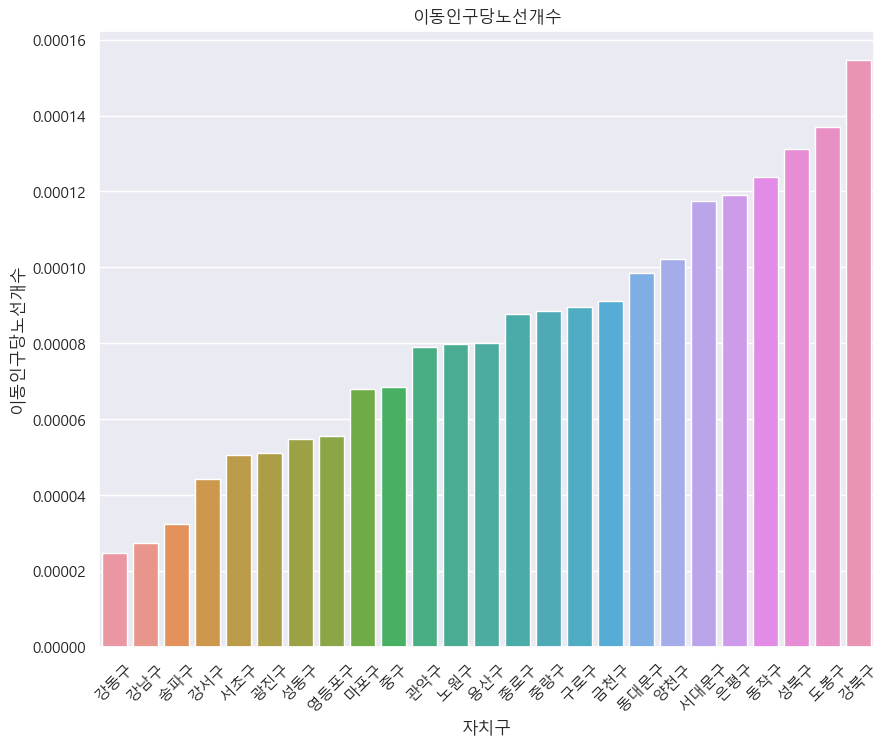
- 대립가설\_(H1) : 기존 주장에 문제점을 제기하는 새로운 가설

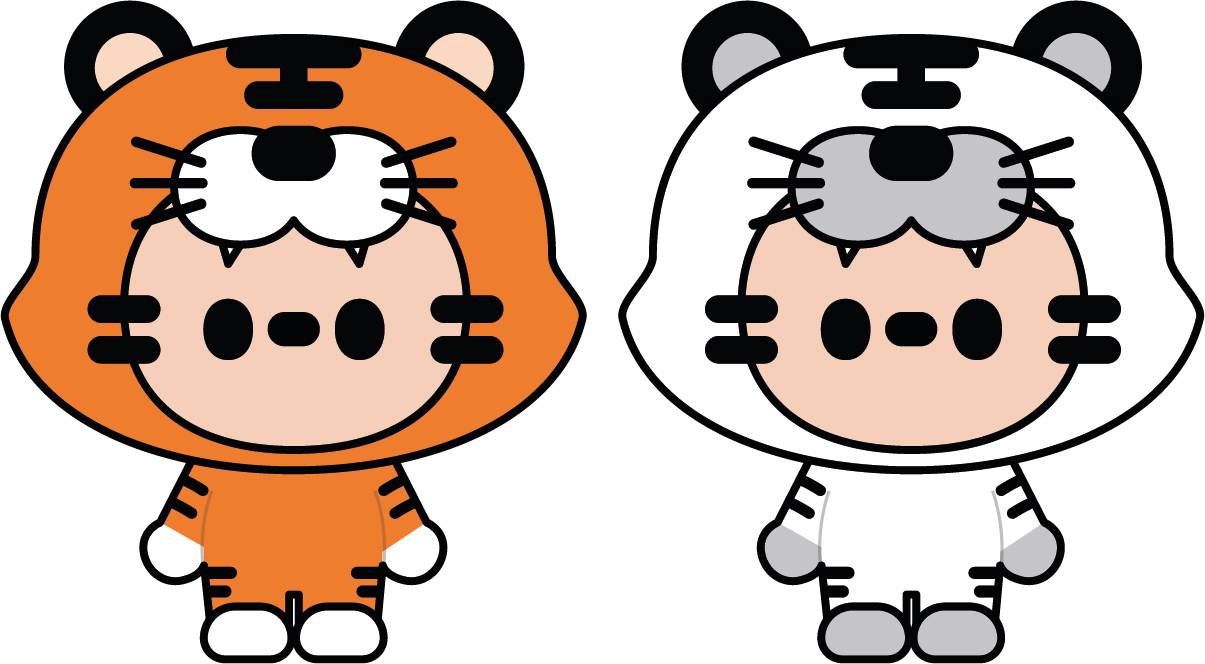






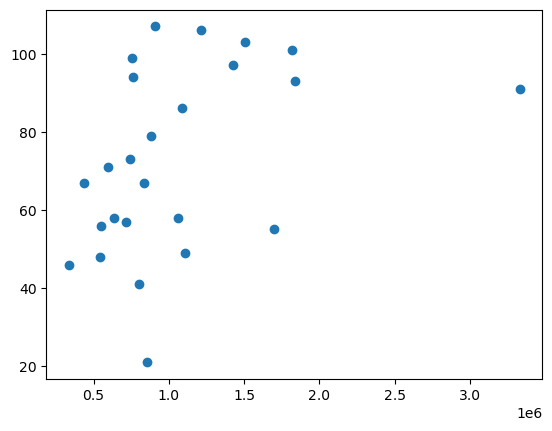


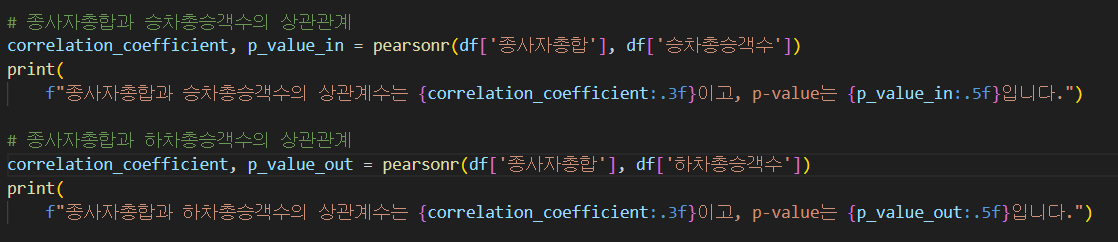




하하/호호

버스 일러스트





종사자총합과 승차총승객수의 상관계수는 0.380이고, p-value는 0.06085입니다.

종사자총합과 하차총승객수의 상관계수는 0.346이고, p-value는 0.09044입니다.



0.67 0.00028 0.65 0.00041