



# 지하철 편의 개선 제안

기업 프로젝트 기반 AI 빅데이터 융합 전문가 양성과정

현정환 | 송수린 | 이윤  
2023.03.03

## 01. 주제 및 목적

“지하철 이용이  
불편한 이유는  
무엇일까?”

연구 주제:

1. 지하철 혼잡도의 원인을 파악한다.
2. 지하철 혼잡도를 완화하기 위한 방법을 모색한다.
3. 해결 방안 적용 시 기대효과를 예측한다.

## 01. 주제 및 목적 | 연구배경

“2300만 인구가 모여 사는 수도권. 과밀은 일상이 됐다.

서울 용산구 이태원 참사 이후 다수의 인원이 좁은 공간에 밀집하는 상황을 개선해야 한다는 지적이 나온다.

...

압박감으로 인해 지하철에서 실신을 경험하거나 목격했다는 증언도 있었다.

김포골드라인을 타고 서울로 출퇴근하는 최아라(26·여)씨는 “출퇴근 시간, 산소가 부족하다고 느낄 정도로 심각할 때가 많다”며 “한 달 전쯤 실신할 뻔한 적도 있다.

그때 제 앞에 계시던 분도 쓰러졌다. 종점에서 내리면 헉헉대며 숨을 고르는 사람들도 종종 보인다”고 말했다.

김은지(34·여)씨도 “(압박받는 가운데) 히터까지 틀면 가슴이 심하게 답답할 때가 있다”며 “겨울철에는 열차에서 내린 후 숨쉬기가 답답한지 플랫폼 의자에 누워 쉬는 사람들도 있다”고 설명했다.

...

김포골드라인 정원은 172명이지만 출퇴근 시간대에는 300~400명의 시민이 탑승한다. 김포시에 따르면 1일 기준 출근 시간대 평균 혼잡률은 230%다.”

## 01. 주제 및 목적 | 연구배경

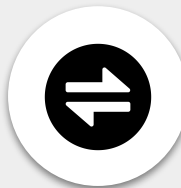
- 대중교통 수단 선택시 이용객의 주요 관심사는 “시간”과 “혼잡도”
- 지하철 평균 혼잡도는 희망 수준의 2.3배
- 출퇴근 시 지하철 출퇴근자 행복지수에 혼잡도가 큰 영향을 준다는 연구 결과
- 환승 혼잡도가 대중교통 정책 우선순위 고려사항이 되어야 한다는 제안

출처: <http://www.sijung.co.kr/news/articleView.html?idxno=96114>

## 02. 연구 방법 및 과정 | 연구 가설

가설 1: 지하철 **이용객 수**와 혼잡도가 양의 상관관계를 보일 것이다.

- 혼잡도와 이용객 수 사이의 상관관계 분석
- 환승 가능 여부와 이용객 수 사이의 상관관계 분석
- 총 이용객 수 상위 30개 역의 환승 가능 여부 비율 분석
- 총 이용객 수 상위 6개, 하위 6개 역의 시간대 별 이용객 수 분석



## 02. 연구 방법 및 과정 | 연구 가설

가설 2: 무임승차자 수와 열차의 혼잡도는 양의 상관관계를 보일 것이다.

- 시간대 별 혼잡도 분석
- 노인 승하차 수가 많은 역 5개의 시간대별 혼잡도 분석
- 주요 무임승차 대상인 노인의 승차자 수와 혼잡도간 상관관계 분석



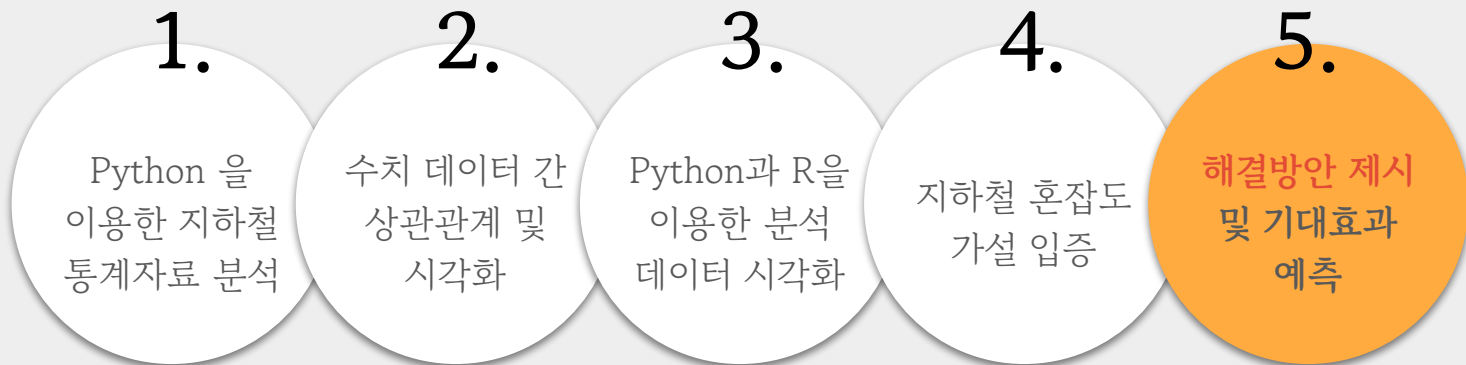
## 02. 연구 방법 및 과정 | 연구 가설

가설 3: **지하철 역의 구조적 특성**과 **혼잡도**는 상관관계를 보일 것이다.

- 역사의 혼잡도, 층수, 면적 간 상관계수 분석
- 출근시간대 혼잡도 상위 25%역의 층수, 면적, 혼잡도 간 상관계수 분석
- 퇴근시간대 혼잡도 상위 25%역의 층수, 면적, 혼잡도 간 상관계수 분석

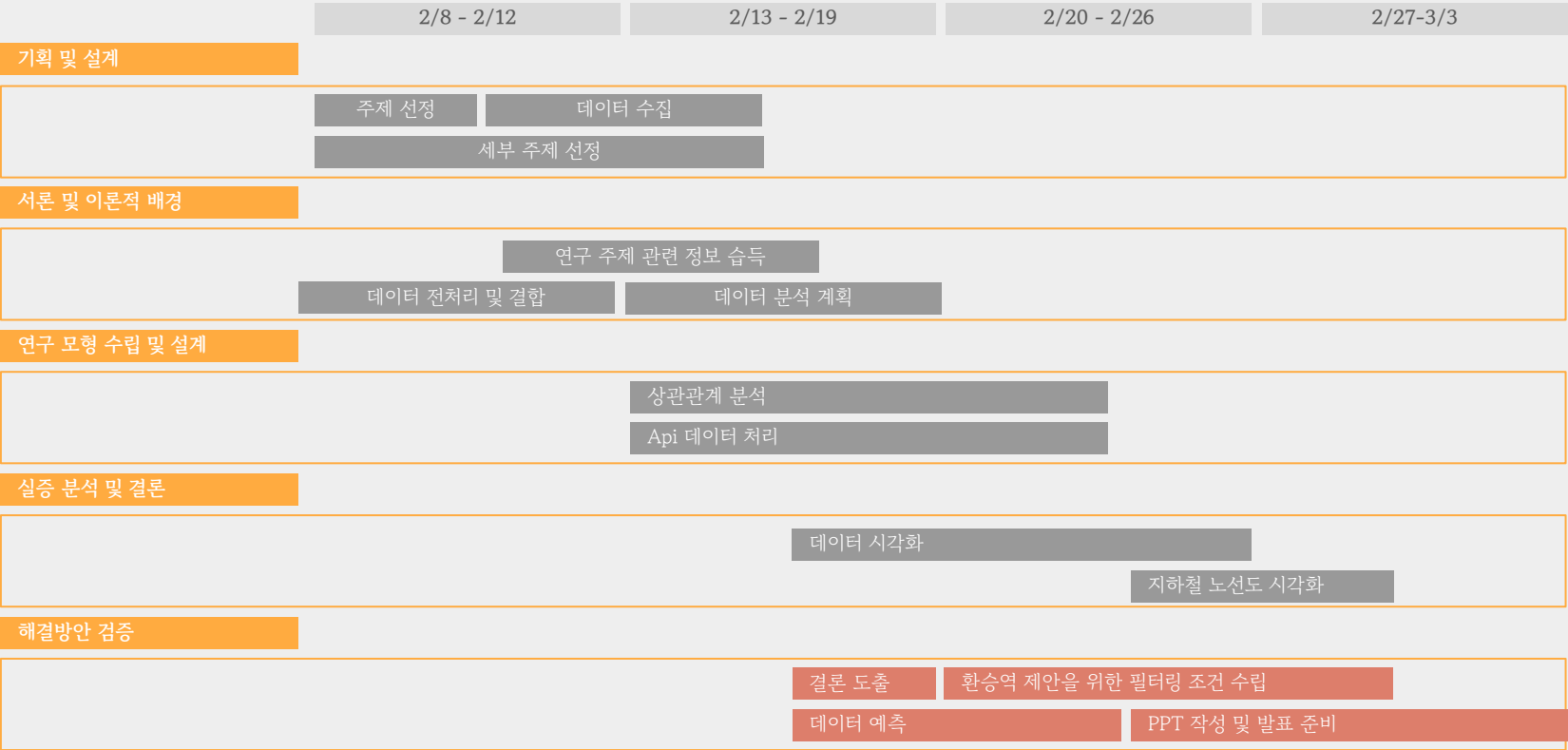


## 02. 연구 방법 및 과정 | 연구 과정





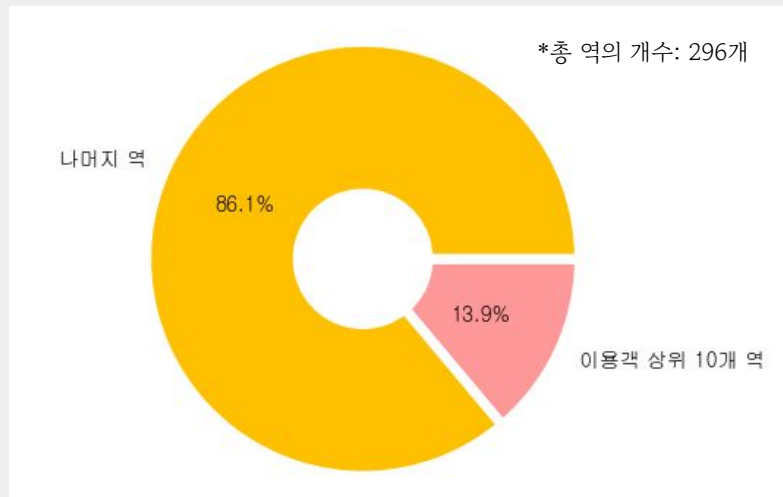
# 02. 연구 방법 및 과정 | 프로젝트 관리



### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### 배경 및 분석 대상

- 특정 역에 승하차인원 쏠림 현상
- 특정 호선에 승하차인원 쏠림 현상
- 이용객수가 유난히 많은 역을 살펴 혼잡도가 높은 원인을 규명

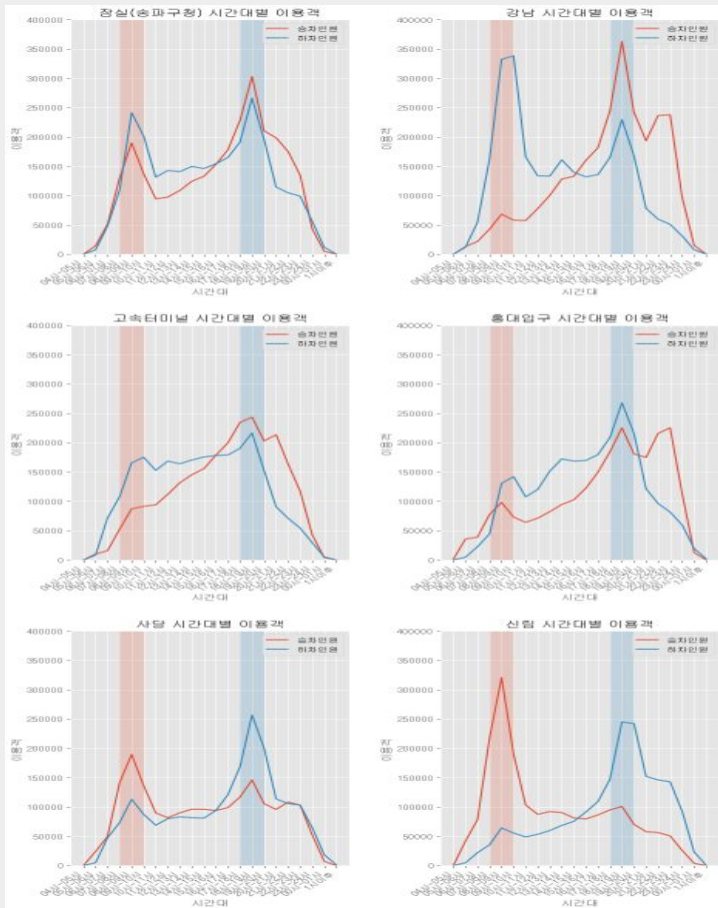


총 이용객 수가 많은 역 10개에 지하철 전체 이용객의 13.9%가 포함된다.

모든 역에 이용객이 골고루 분포되어 있다면 10개의 역에는 총 이용객 수의 약 3.37%가 포함되어야 한다.

따라서 현재 특정 역에 이용객이 몰려있다는 것을 알 수 있다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도 - 이용객 수 상위 6개 역 시간대별 승하차인원



총 이용객 수 상위 6개 역을 뽑아 본 결과 잠실, 강남, 고속터미널, 홍대입구, 사당, 신림역이 추출되었다.

출근시간대(7시~10시)는 빨간색 띠로, 퇴근시간대(17시~20시)는 파란색 띠로 표시하였다.

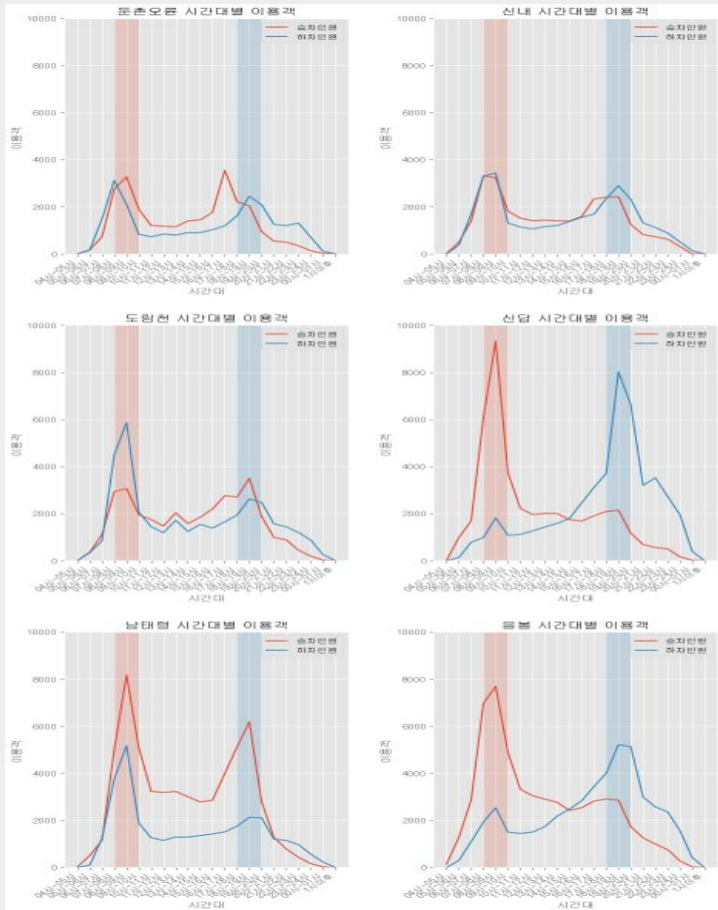
출퇴근 시간대에 최소 약 55,000명에서 최대 약 360,000명의 인원이 이용하였다.

고속터미널역과 홍대입구역 외에는 모두 출퇴근시간대 승하차인원이 눈에 띄게 많다. 이는 특정 시간대에 상대적으로 더 혼잡함을 알 수 있다.

한편 출근시간대에 승차가 많은 역은 거주지역으로, 퇴근시간대에 승차가 많은 역은 회사들이 몰려있는 지역으로 예상된다.

또한 6개의 역 중 4개의 역이 퇴근시간대에 승차가 많은 것으로 보아 직장인 몰려있는 지역이 지하철 이용객이 많은 것으로 추측된다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도 - 이용객 수 하위 6개 역 시간대별 승하차인원



총 이용객 수 하위 6개 역을 뽑아 본 결과 둔촌오륜, 신내, 도림천, 신답, 남태령, 응봉역이 추출되었다.

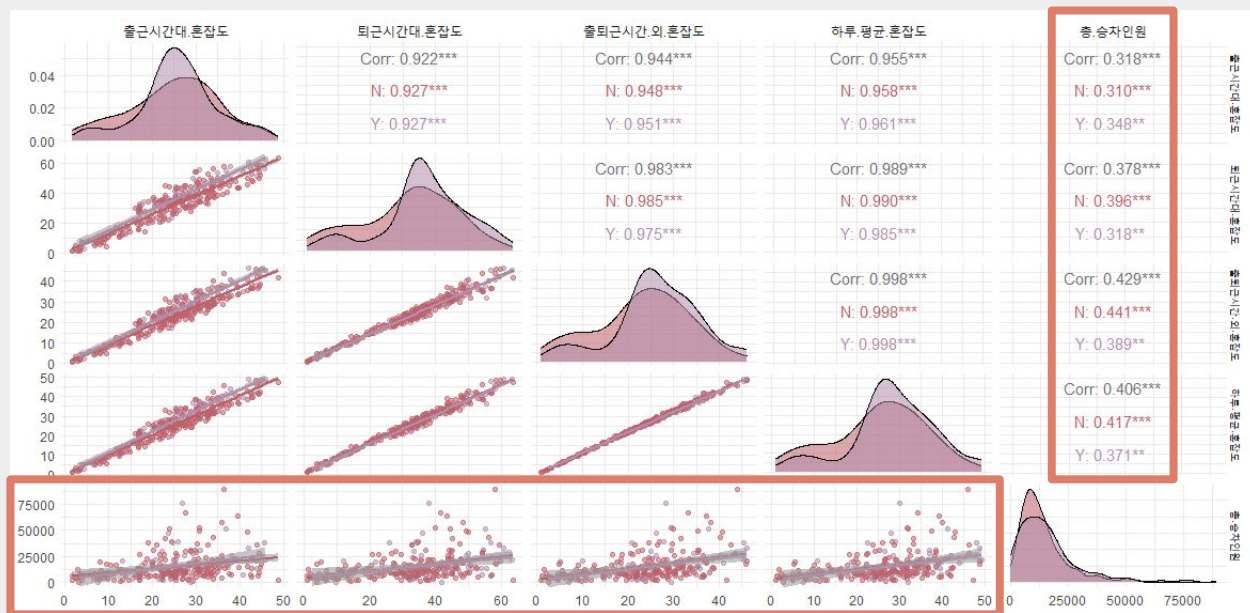
출근시간대(7시~10시)는 빨간색 띠로, 퇴근시간대(17시~20시)는 파란색 띠로 표시하였다.

출퇴근 시간대에 최소 약 2,000명에서 최대 약 9,500명의 인원이 이용하였다.

6개의 역 모두 앞서 본 상위 6개 역과 같이 출퇴근시간대에 승하차인원이 눈에 띄게 많은 것을 볼 수 있다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### - 총 승차인원과 혼잡도 관계



총 승차인원과 혼잡도 사이의 상관계수를 구하였다.

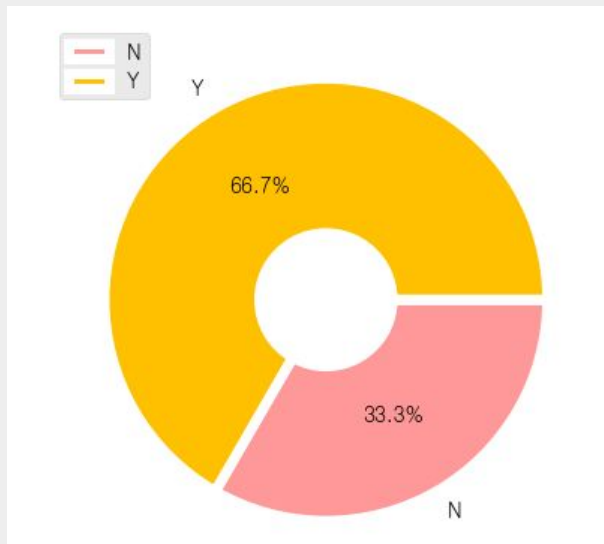
출퇴근시간대 혼잡도, 그 외 시간대 혼잡도, 전체 운영 시간대 혼잡도 모두 총 승차인원과 유의미한 상관계수가 도출되었다.

아래 산점도 또한 선형적인 모양을 가지고 분포되어있는 것을 볼 수 있다.

이를 통해 총 승차인원과 혼잡도는 약한 양의 상관관계를 가짐을 알 수 있다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### - 총 이용객 수 상위 30개 역의 환승 가능 여부 비율



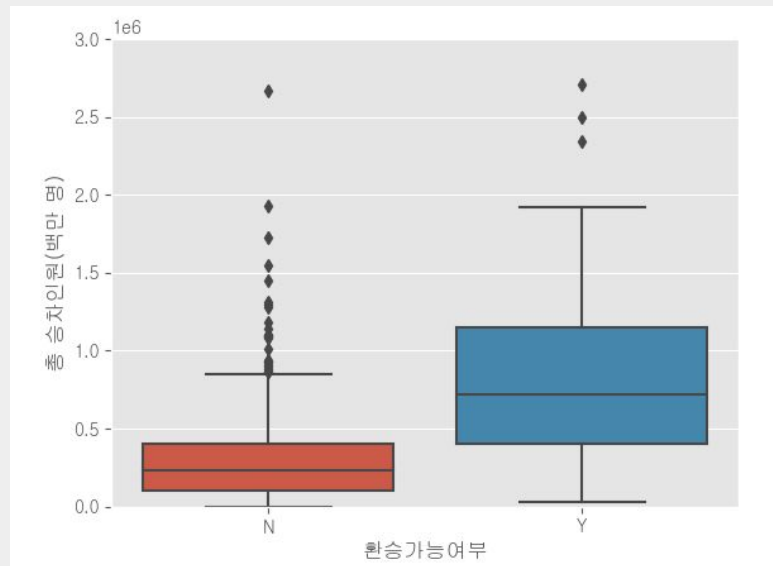
이용객 수가 많은 역의 특징을 알아보기 위해 총 이용객 수를 기준으로 정렬해보았다.

이용객 수가 많은 역들은 대체적으로 환승역인 것을 알 수 있었다. 정확한 수치를 얻기 위해 총 이용객 수 상위 30개의 역을 환승 가능 여부에 따라 구성비를 구해보았다. 그 결과 환승이 가능한 역이 30개 역 중 66.7%를 차지한다는 것을 확인할 수 있었다.

이를 통해 환승 가능 여부와 총 이용객 수 사이에 관련성이 있을 것이라고 추측하고 이에 대해 알아보고자 한다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### - 환승 가능 여부와 총 승차인원 관계



보다 구체적인 환승 가능 여부에 따른 이용객 수를 보기 위해  
역별 총 승차인원을 환승 가능 여부에 따라 박스 플롯을  
그려보았다.

박스플롯을 통해 환승 가능한 역의 총 승차인원 중앙 값이  
환승 불가능한 역보다 확실히 더 큰 것을 알 수 있고, 상자의  
위치 또한 환승 가능한 역의 상자가 높게 형성 됨을 볼 수  
있다.

이는 환승 가능 여부와 총 이용객 수 사이에 관련성이 있음을  
시각적으로 보여준다.

### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### - 환승 가능 여부와 총 승차인원 관계

Independent Samples T-Test

		Statistic	df	p
총 승차인원	Student's t	-12.1 *	520	< .001

Note.  $H_0: \mu_N = \mu_Y$

\* Levene's test is significant ( $p < .05$ ), suggesting a violation of the assumption of equal variances

‘환승 가능 여부에 따라 총 승차인원이 같다’ 라는 귀무가설을 통계적으로 확인하기 위해 독립 t검정을 해보았다.

P-value 값이 유의수준 0.05보다 작아 귀무가설을 기각하게 되면서 환승 가능 여부에 따라 총 승차인원이 다르다는 결론에 도달하였다.

```
> cor(df$환승, df$총.승차인원, method = 'spearman', use = 'complete.obs')  
[1] 0.4025975
```

추가적으로 환승 가능 여부와 총 승차인원 간의 스피어만 상관계수를 구해보았는데, 약 0.4로 양의 상관관계가 있다고 할 수 있다.



### 03. 분석 및 시각화 | 이용객 수 및 혼잡도

#### - 환승역 개설 전 후 총 이용객 수 변화

고속버스터미널 역

연도	3	7	9	총합	
2007년	122,107명	29,496명		151,603명	환승역 개설 전
2008년	116,699명	32,386명		149,085명	
2009년	115,646명	33,342명	15,394명	164,382명	
2010년	115,888명	34,055명	19,564명	169,507명	환승역 개설 후
2011년	113,515명	35,362명	20,700명	169,577명	

종로3가 역

연도	1	3	5	총합	
1997년	93,417명	16,217명	자료없음	109,634명	환승역 개설 전
1998년	93,086명	15,078명		108,164명	
1999년	자료없음	자료없음		-	
2000년	99,957명	13,773명	24,117명	137,847명	환승역 개설 후
2001년	102,596명	13,742명	25,299명	141,637명	
2002년	99,379명	13,857명	25,051명	138,287명	

약수 역

연도	3	6	총합	
1998년	23,508명			환승역 개설 전
1999년	자료없음			
2000년	26,222명			
2001년	26,617명	4,872명	31,489명	환승역 개설 후
2002년	26,836명	5,587명	32,423명	
2003년	26,406명	6,320명	32,726명	

이수 역

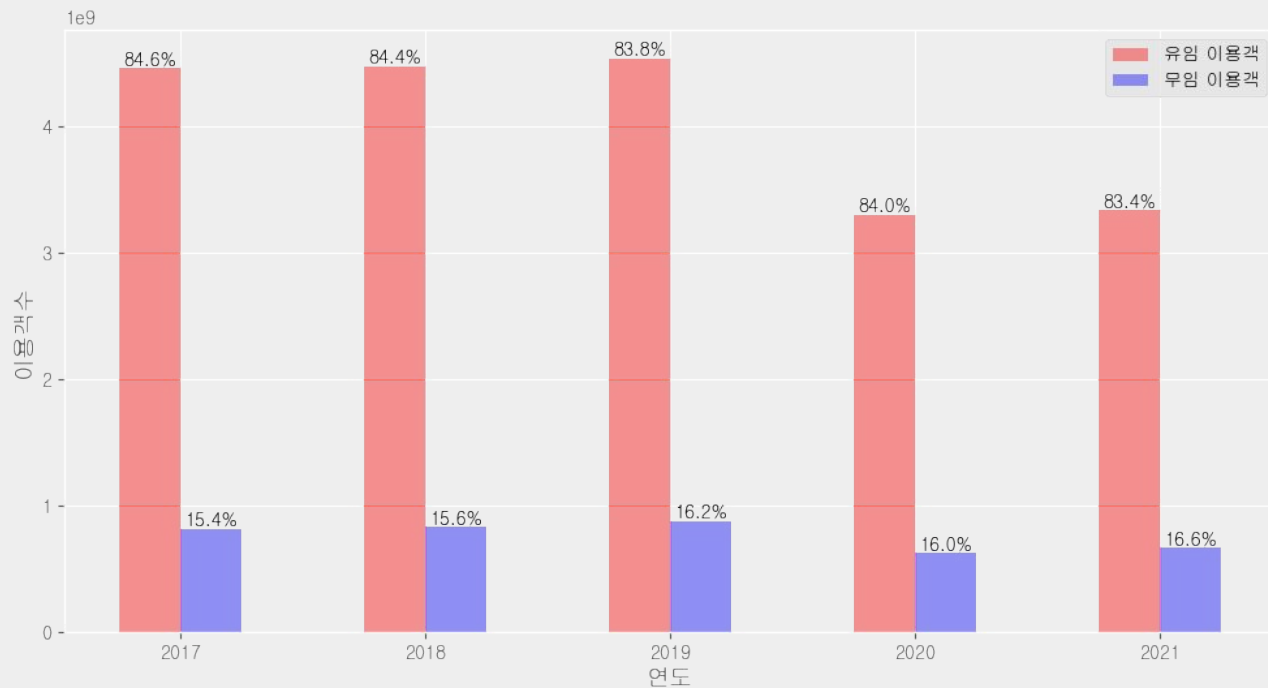
연도	4	7	총합	
1997년	63,897명			환승역 개설 전
1998년	62,701명			
1999년	자료없음			
2000년	56,173명	19,104명	75,277명	환승역 개설 후
2001년	51,704명	23,718명	75,422명	
2002년	51,382명	25,863명	77,245명	

실제로 환승역 개설 후 총 이용객 수가 어떻게 변화하였는지 살펴보기 위해 현재 환승역인 역들 중 몇 개를 가져와 보았다.

대체로 환승역 개설 전보다 환승역을 개설 한 후에 이용객 수가 확연히 증가한 것을 볼 수 있다.

### 03. 분석 및 시각화 | 무임승차 및 혼잡도

#### 배경

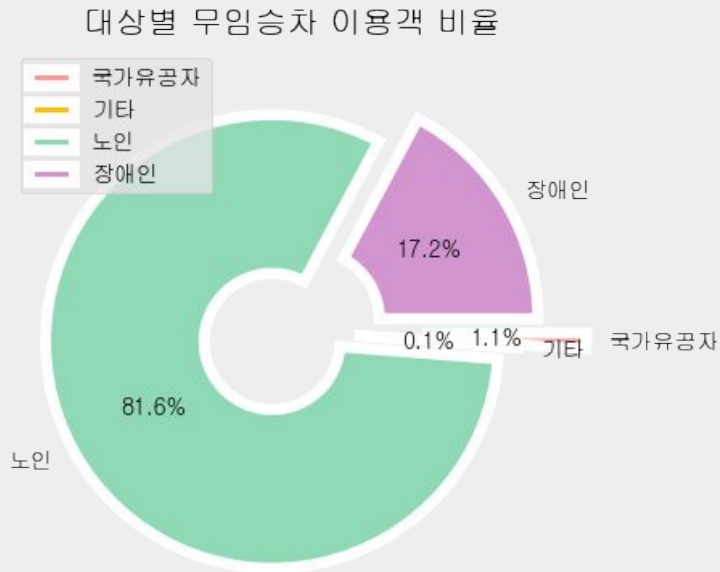


지하철 이용객 중 무임 승차자의 비율이 16%로 유임 승차자의 약 1/6을 차지한다.

### 03. 분석 및 시각화 | 무임승차 및 혼잡도

#### 배경

- 지하철 무임승차에 노인이 큰 비중을 차지



5년(2017~2021)간의 무임승차자 구성을 보았을 때, 무임승차자 중에서 노인이 80% 이상을 차지하고 있다.

따라서 노인 무임승차자에 초점을 맞춰 분석을 진행하도록 하였다.

## 03. 분석 및 시각화 | 무임승차 및 혼잡도

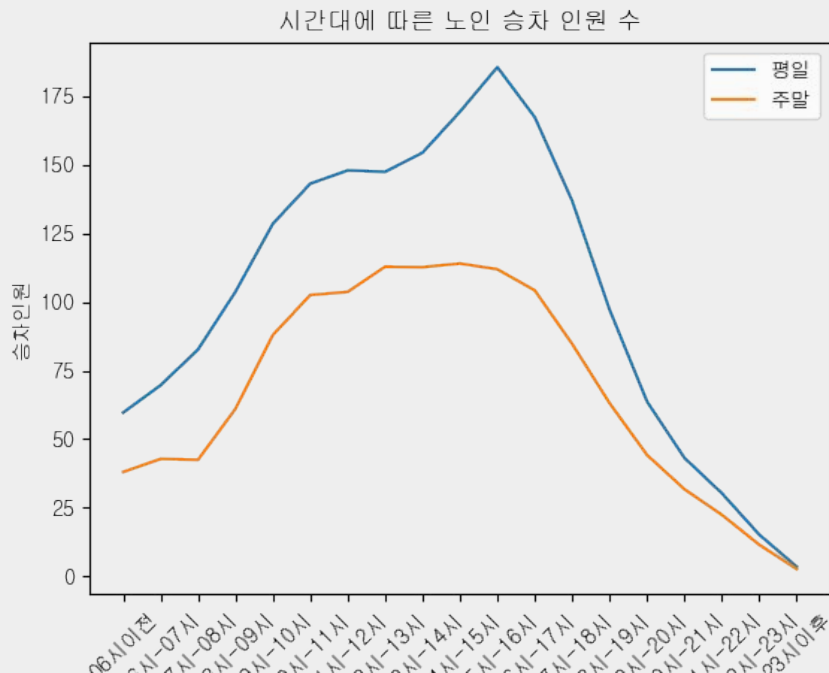
### 분석 대상

- 시간대별 노인 승하차 인원 (2019-2022)
- 운영시간대 혼잡도와 노인승차 인원의 상관관계
- 출퇴근 시간대 혼잡도와 노인승차 인원의 상관관계

### 03. 분석 및 시각화 | 무임승차 및 혼잡도

#### 분석 내용

- 평일과 주말의 노인승차 인원 비교
- 주말보다 평일에 노인 승차자 수가 많음
- 따라서 다른 시간대보다 출퇴근시간대에 많은 지 확인해 볼 필요 있음



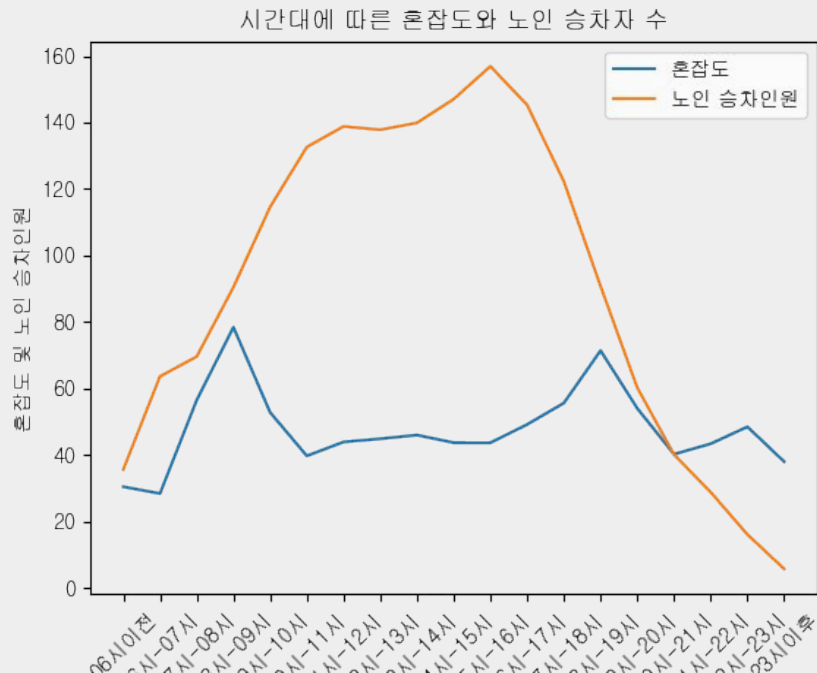
### 03. 분석 및 시각화 | 무임승차 및 혼잡도

#### 분석 내용

- 운영시간대 (6시 이전-23시 이후) 지하철역 혼잡도 및 노인승차 인원 비교

- 혼잡도는 출퇴근 시간에 집중적으로 높은 반면,  
노인 승차인원 수는 출근시간부터 꾸준히 증가하여  
오후 3~4시에 가장 많음

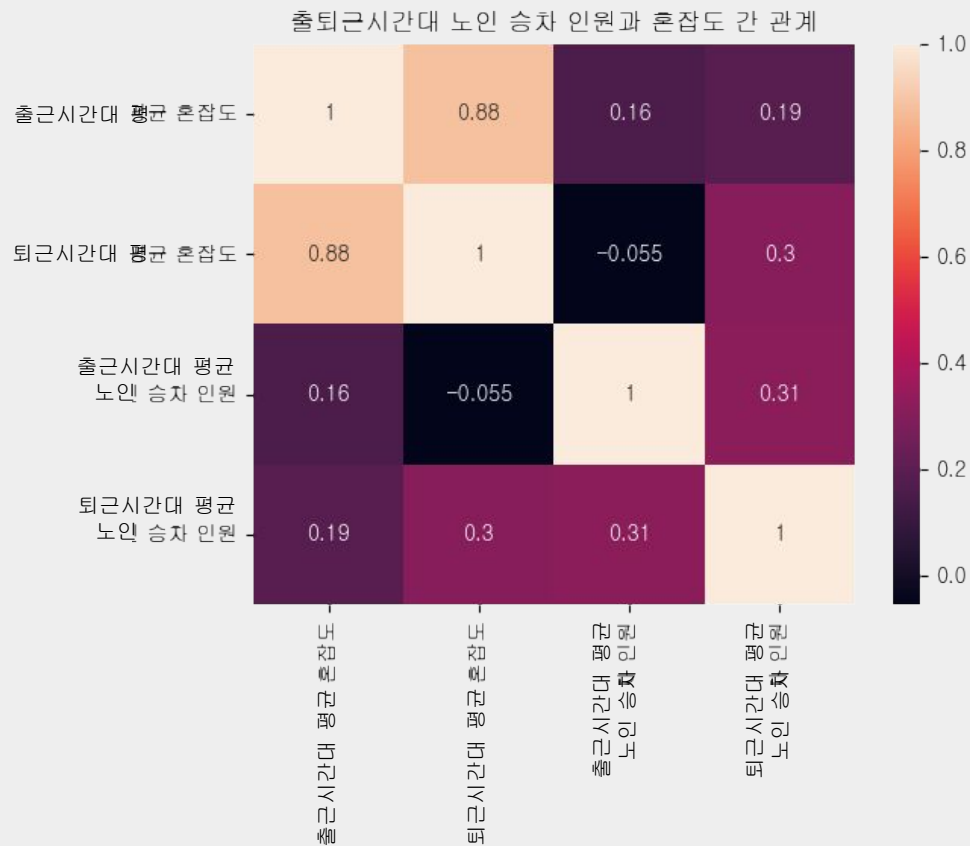
- 노인 승차인원 수가 출퇴근 시간대 혼잡도에 크게  
영향을 주지 않을 것이라고 예상됨



### 03. 분석 및 시각화 | 무임승차 및 혼잡도

#### 분석 내용

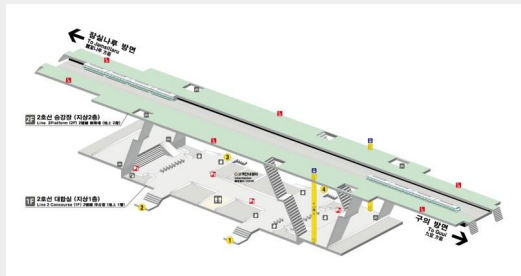
- 출근시간대에는 0.16으로 유의미한 상관관계가 있다고 보기 어려움
- 퇴근시간대에도 0.3으로 유의미한 상관관계가 있다고 보기 어려움



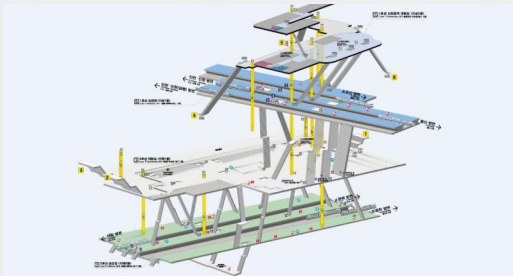
### 03. 분석 및 시각화 | 역 구조 및 혼잡도

#### 배경 및 분석 대상

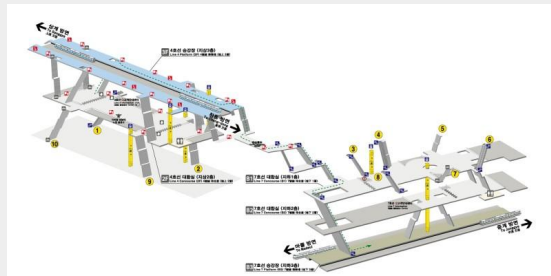
- 지하철 이용자의 역 내 동선을 수평 이동 (면적)과 수직 이동 (층수)을 통해 파악한다.
- 특정 역 구조가 혼잡도 해소에 도움을 줄 수 있는지 분석한다.



\*예시: 수평 이동을 주요 동선으로 하는 강변역



\*예시: 수직 이동을 주요 동선으로 하는 신도림



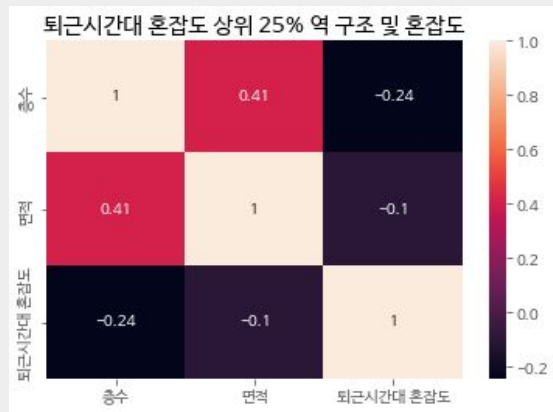
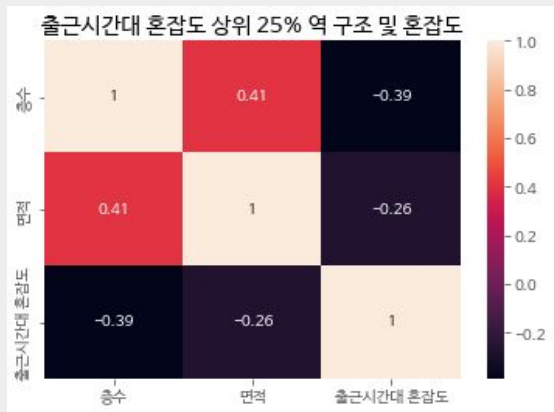
\*예시: 수직/수평 이동을 주요 동선으로 하는 청담



### 03. 분석 및 시각화 | 역 구조 및 혼잡도

#### 분석 내용

- 운영시간내 혼잡도와 층수, 면적 간 상관계수는 각 -0.36, -0.14
- 출근시간대(07-10시) 혼잡도와 층수, 면적 간 상관계수는 각 -0.39, -0.26
- 퇴근시간대(17-20시) 혼잡도와 층수, 면적 간 상관계수는 각 -0.24, -0.1
- 혼잡도와 층수는 강한 음의 상관관계, 혼잡도와 면적은 약한 음의 상관 관계



## 04. 결론 | 무임승차 및 혼잡도

- 노인 무임승차는 운영시간내 지하철 혼잡도에 영향을 미치고 있다고 보기 어렵다.
- 노인 무임승차는 출근시간 혼잡도에 영향을 준다고 보기 어렵다.
- 노인 무임승차는 퇴근시간 혼잡도에 영향을 준다고 보기 어렵다.

## 04. 결론 | 역 구조 및 혼잡도

- 역의 높은 층수는 혼잡도 해소에 도움을 주는 것으로 보인다.
- 역의 크기는 혼잡도 해소에 다소간 도움을 주는 것으로 보인다.
- 새로운 지하철 환승역 제안 시 많은 층수와 넓은 면적을 조건의 우선순위 부여

## 04. 결론 | 이용객 수 및 혼잡도

- 승하차 인원이 많은 역은 배차 간격 조정을 통해 혼잡도 완화를 위한 노력이 이루어진 것으로 보인다.
- 이용객 수와 혼잡도는 유의미한 양의 상관관계를 보인다.
- 승하차 인원과 혼잡도가 여전히 양의 상관관계와 유의미성을 보이는 것으로 보아 추가적 조정이 필요한 것으로 해석 된다.
- 환승 가능 여부와 이용객 수 사이에 양의 상관관계를 갖는다.

## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 환승역 제안

### 조건 적용 순서

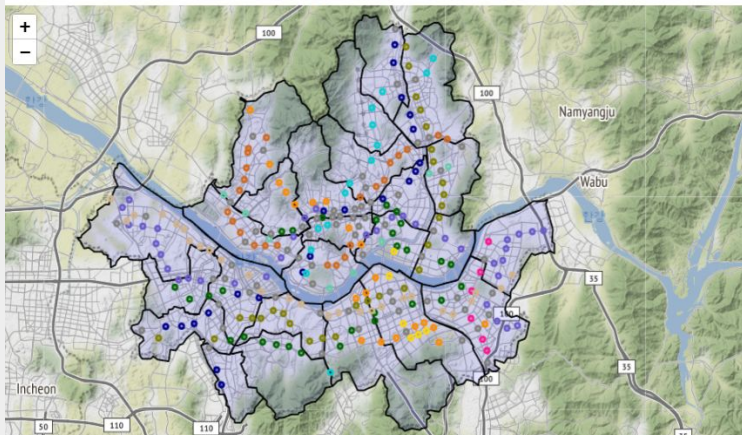
- 1-1. 현재 환승역인 역들의 층수의 1분위수인 3층을 기준으로 3층보다 많은 층수를 가진 역 채택
- 1-2. 현재 환승역인 역들의 면적의 1분위수인 10209.725m<sup>2</sup>를 기준으로 이보다 넓은 면적을 가진 역 채택
- 2. 이미 환승역인 역은 제외
- 3. 환승 후보역 반경 2.3km 내 후보역과 다른 호선인 역만 채택
- 4. 환승 후보역 반경 5.75km 내 해당 후보역과 3번에서 채택된 역간의 환승역이 존재하지 않는 경우 연결

환승 후보역

결합할 호선

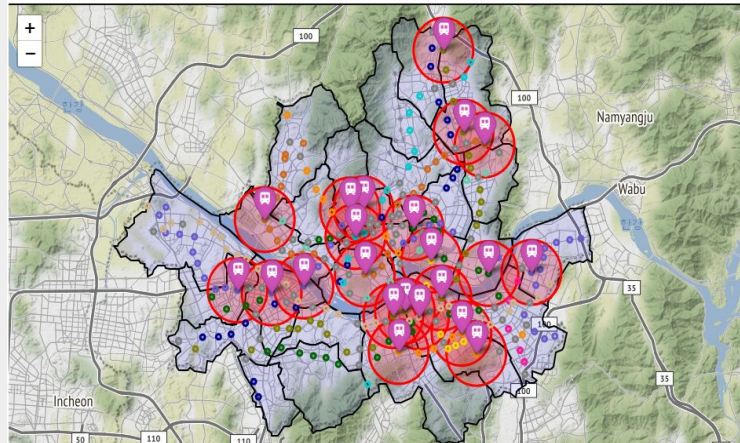
## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 환승역 제안

STEP 1: 현재 서울 내 운행 중인 지하철 표시



서울 내 지하철 역 노선별로 표시

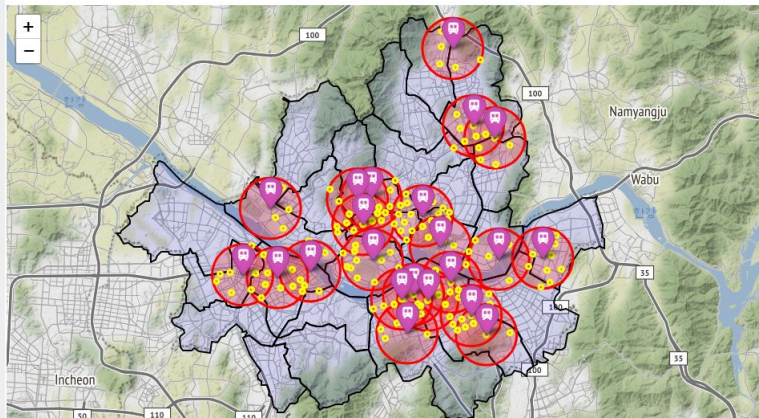
STEP 2: 환승 후보역 반경 2.3km 표시



각 환승 후보역마다 두 정거장의 거리인 2.3km를 반경으로  
인근역을 제한  
(한 정거장의 평균 거리 : 1.15km)

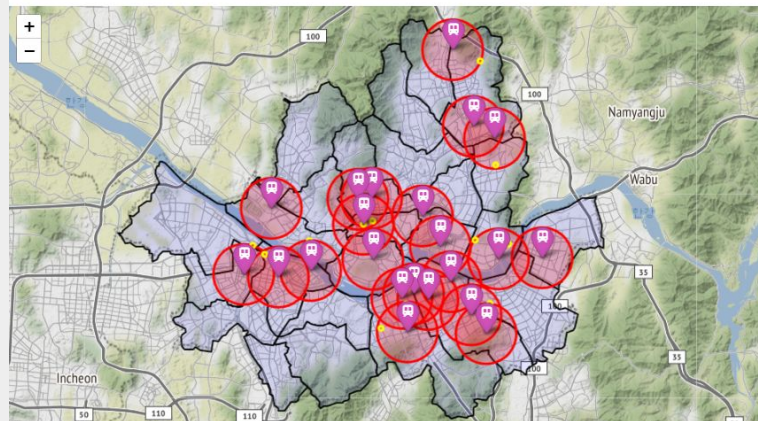
## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 환승역 제안

STEP 3: 환승 후보역 반경 2.3km 외 역 삭제



반경에 들어오지 못한 나머지 역을 지도에서 지우기

STEP 4: 2.3km 이내에 후보역과 다른 호선이면서 환승되지 않는 역 추출

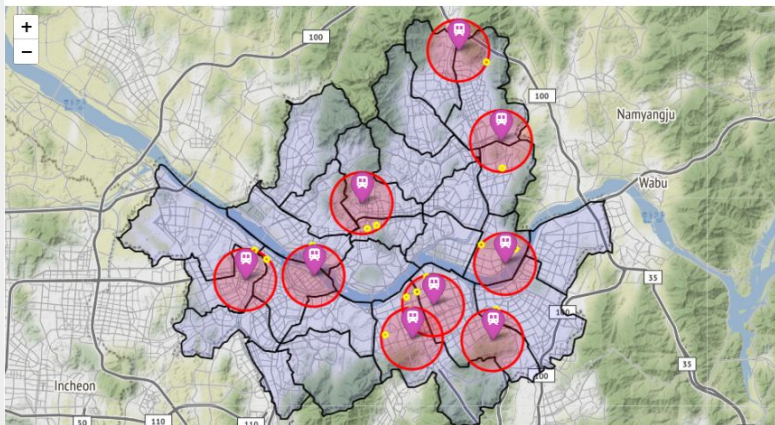


환승 후보역과 다른 호선이면서 환승역이 아닌 역 선택



## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 환승역 제안

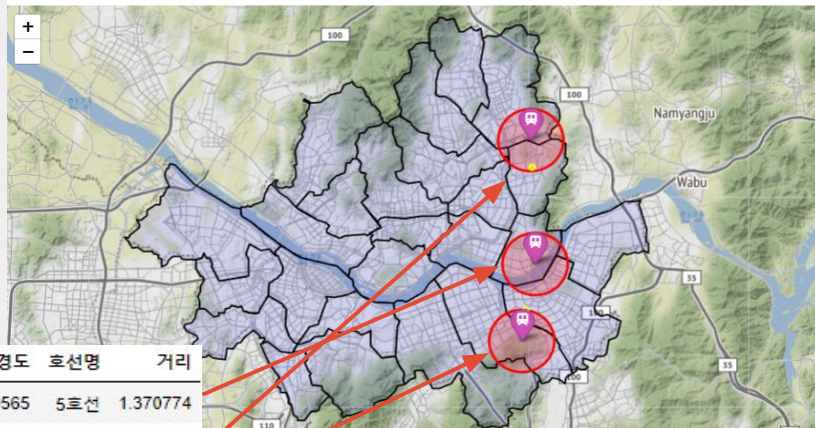
STEP 5: 예비 후보 역 추출



환승 후보역 중 반경 안에 역이 남아 있는 후보역만 선택

반경 5.75km(약 5정거장 거리)내에 후보역의 호선과 연결할 수 있는 역의 호선간 환승역이 존재하는 경우 후보역에서 제외

STEP 6: 최종 환승 후보 역



	후보역	지하철역	주소	위도	경도	호선명	거리
0	강변(동서울터미널)	광나루(장신대)	서울 광진구 광장동 237	37.545299	127.089565	5호선	1.370774
1	봉화산(서울의료원)	망우	서울 중랑구 상봉동 172	37.599296	127.092367	중앙선	2.005859
2	일원	삼전	서울 송파구 잠실동 347	37.504549	127.087364	9호선	2.299789



#### 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 기대효과

-기존 강변역 -> 광나루역 최소 소요 시간



환승역을 제안한 역 중 강변역을 5호선과 환승 할 수 있게 바꾼다고 가정해보았다.

기존에 강변에서 광나루를 가기 위해 두번을 환승하여 21분을  
돌아가야했었지만 환승역이 개설되면 지하철이 2.3분만에  
강변역에서 광나루역을 갈 수 있게 된다.

여기서 2.3분은 지하철이 한 정거장을 이동할 때 평균 속도 35km 정도로 달린다고 가정하고 계산한 시간이다.

-환승역 추가 후 강변역 -> 광나루역 예상 소요 시간 (시속 35km 기준)

	후보역	지하철역	거리	기존 소요시간	기대 소요시간
0	강변(동서울터미널)	광나루(장신대)	1.370774	21분	2.3

# 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 기대효과

-기존 봉화산역 -> 망우역 최소 소요 시간



이번엔 봉화산역에 중앙선을 연결한다고 가정해보았다.

기존에 봉화산역에서 망우역을 가기 위해 한번을 환승하여 22분을 돌아가야했었지만 환승역이 개설되면 지하철이 3.4분만에 봉화산역에서 망우역을 갈 수 있게 된다.

여기서도 마찬가지로 3.4분은 지하철이 한 정거장을 이동할 때 평균 속력 35km 정도로 달린다고 가정하고 계산한 시간이다.

-환승역 추가 후 봉화산역 -> 망우역 예상 소요 시간 (시속 35km 기준)

후보역	지하철역	거리	기존 소요시간	기대 소요시간
1 봉화산(서울의료원)	망우	2.005859	22분	3.4

# 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 기대효과

-기존 일원역 -> 삼전역 최소 소요 시간



마지막으로 일원역에 9호선을 추가한다고 가정해보았다.

기존에 일원역에서 삼전역을 가기 위해 두번을 환승하여 22분을 돌아가야했었지만 환승역이 개설되면 지하철이 3.9분만에 일원역에서 삼전역을 갈 수 있게 된다.

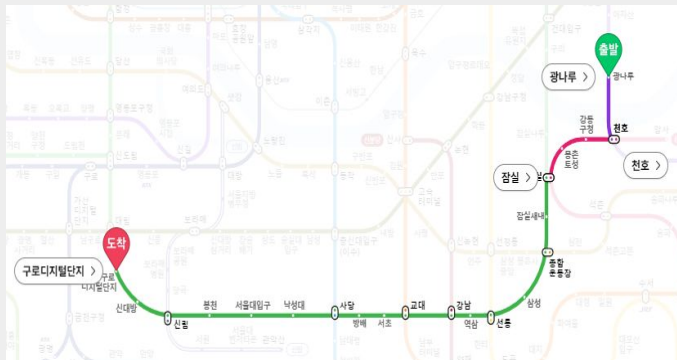
여기서 3.9분 또한 지하철이 한 정거장을 이동할 때 평균 속도 35km 정도로 달린다고 가정하고 계산한 시간이다.

-환승역 추가 후 일원역 -> 삼전역 예상 소요 시간 (시속 35km 기준)

	후보역	지하철역	거리	기존 소요시간	기대 소요시간
2	일원	삼전	2.299789	22분	3.9

## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 기대효과

- 기존 광나루 -> 구로디지털단지 최소 소요시간



앞서 본 것처럼 단거리가 아닌 장거리를 가는 경우를 가정해보았다.

광나루역에서 출근을 하기 위해 구로디지털단지역을 간다고 했을 때, 기존에는 두번을 환승하여 51분이 걸렸다.

그러나 강변역에 5호선 환승역을 개설하게 되면 총 소요시간이 41.3분으로 예상되어 이전보다 약 10분의 시간 절약을 기대해 볼 수 있다.

- 환승역 개설 후 광나루 -> 구로디지털단지 예상 소요시간

후보역	지하철역	거리	기존 소요시간	기대 소요시간
0 강변(등서출터미널)	광나루(장신대)	1.370774	21분	2.3

**39분** 최소시간

18개역 환승 0회 | 카드 1,550원

- 18:11 2 강변역 (214)
- 18:50 2 구로디지털단지역 (232)

**41.3분**

\*약 10분 시간 절약 예상

## 04. 결론 | 이용객 수 및 혼잡도 해결방안 (환승역 추가) - 기대효과

- 기존 구리 -> 월곡 최소 소요시간



이번에는 구리역에서 등교를 하기 위해 월곡역을 간다고 해보았다.

기존에는 두번을 환승하여 30분이 걸렸다.

그러나 봉화산역에 중앙선 환승역을 개설하게 되면 총 소요시간이 19.4분으로 예상되어 이전보다 약 10분의 시간 절약을 기대해 볼 수 있다.

- 환승역 개설 후 구리 -> 월곡 예상 소요시간

후보역	지하철역	거리	기존 소요시간	기대 소요시간
1 봉화산(서울의료원)	망우	2.005859	22분	3.4

**19.4분**

6분 최소시간

2개역 환승 0회 | 카드 1,250원

18:23	경의중앙	구리역 (1313)
18:29	경의중앙	망우역 (1315)

10분 최소시간

6개역 환승 0회 | 카드 1,250원

18:19	6	봉화산역 (647)
18:29	6	월곡역 (641)

\*약 10분 시간 절약 예상

## 05. 한계점

### 1.

데이터 수집의  
한계

- 수집할 당시의 형태 그대로인 로그 데이터가 아니라 1차 가공 된 통계 데이터를 토대로 분석이 진행 되었다.
- 데이터 형태가 상이해 지하철 상선과 하선 데이터에 의한 차이가 반영 되지 못했다.

## 05. 한계점

### 2.

변수 고려의  
한계

- 버스, 택시, 도보 이용 등 지하철 상하차 및 혼잡도에 영향을 줄만 한 기타 변수에 대한 고려가 이루어지지 못했다.
- 환승역 지정과 이용객 수의 일정한 상관관계를 파악하였음에도, 둘 사이의 인과관계를 파악하기 어렵기에, 환승역 지정으로 이용객 수 및 혼잡도를 효과적으로 낮췄다고 보기 어렵다.

## 05. 한계점

### 3.

검증 및 구현의  
한계

- 전체 역 이용시에 대한 검증이 아니라 특정 역 간 변화에 대한 결과만이 검증에 사용되었다.
- 기타 변수를 제한 하기 위해 여러가지 가정을 사용하여 해당 내용이 현실을 모두 반영한다고 보기 어렵다.
- 기타 변수 추가 및 시뮬레이션을 통해 더 정확한 제안 검증이 이루어져야 한다.



## 06. 느낀점

- 예상했던 바와 다르게 결과가 도출되었을 때 당황스럽기도 했지만 새로운 사실을 알 수 있는 좋은 기회였다고 생각한다.
- 지금까지 배웠던 분석 기법들을 직접 다시 사용해보고 문제가 생긴 부분들은 스스로 해결하면서 프로그램에 한층 가까워진 것 같아 뿌듯하다.
- 가설 검증을 위한 방법론에 대해 더 자세히 공부하면 유의미한 분석에 도움이 될 것이라는 생각이 든다.
- 공부해 왔던 파이썬과 R을 비롯해 api, folium, selenium을 적극적으로 사용한 점이 긍정적이었다. 또한 사회이슈와 관련해서 분석 및 해결방안을 제시했다는 점도 좋은 방향이었다.
- 중간에 주제가 몇 번 바뀌어 보다 깊이 있는 분석을 할 수 있었던 시간들을 허비하게 만든 점이 아쉬웠지만, 다음에는 데이터 분석을 미리 계획하고 구조적으로 데이터에 접근해 해결할 수 있을 것 같다.
- 초반에 팀원 한명이 개인적인 이유로 이탈하게 된 것이 역할 분담하는데 있어 어렵게 다가왔지만 개개인이 잘하는 부분을 빠르게 파악하고 분담한게 프로젝트 완성으로 이어질 수 있었던 것 같다.

## 07. 참조 | 업무 분장

### 현정환

- 지하철 이용현황, 유무임 승하차 현황  
전처리 및 호선별 시간대 이용객 수 분석
- 환승 가능 여부와 이용객수의 관계분석
  - 역 구조와 이용현황을 결합 및  
환승 후보역 추출, 제안 과정 구현

### 송수린

- 노인 무임 승하차, 혼잡도 데이터  
전처리 및 혼잡도 분석
- 지하철 노선과 환승역 제안 과정  
지도에 시각화
- 시간대 별 승하차인원과 혼잡도,  
환승 가능 여부 및 가능 노선 간의  
상관관계 분석 및 시각화

### 이윤

- 노인 무임 승하차 및 혼잡도  
데이터 분석 및 시각화
- 노인 무임 승하차 연령 변경 시  
데이터 예측
- 역 구조 및 혼잡도 데이터 분석 및  
시각화
- ppt 제작

## 07. 참조 | 개발 환경



Windows 10 pro

Python: Pandas,  
Numpy,  
Matplotlib,  
Folium, JSON,  
requests,  
seaborn,  
selenium

R: GGally, nord,  
tidyverse

## 07. 참조 | 활용 데이터

### 무임승차 및 혼잡도

-서울교통공사 지하철 혼잡도 정보 ([https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer\\_data\\_information](https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer_data_information))

-서울특별시 지하철 호선별 역별 유/무임 승하차 인원 정보 (<http://data.seoul.go.kr/dataList/OA-12251/S/1/datasetView.do>)

- 무임승차 대상별 현황 정보 ([https://kosis.kr/statHtml/statHtml.do?tblId=DT\\_357001\\_A027&orgId=357&language=kor&conn\\_path=&vw\\_cd=&list\\_id](https://kosis.kr/statHtml/statHtml.do?tblId=DT_357001_A027&orgId=357&language=kor&conn_path=&vw_cd=&list_id))

-서울교통공사 역별 월별 노인 승하차인원 정보 (<https://www.data.go.kr/data/15062070/fileData.do?recommendDataYn=Y>)

### 역 구조 및 혼잡도

-서울교통공사 지하철 혼잡도 정보 ([https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer\\_data\\_information](https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer_data_information))

-서울교통공사 연도별 일별 시간대별 역별 승하차 인원 (<https://data.seoul.go.kr/dataList/OA-12921/F/1/datasetView.do>)

-서울교통공사 역사 건축 정보 ([https://data.kric.go.kr/rips/M\\_01\\_01/intro.do](https://data.kric.go.kr/rips/M_01_01/intro.do))

-서울교통공사 지하철역사 건축 현황 ([http://data.seoul.go.kr/dataList/OA-11572/A/1/datasetView.do;jsessionid=395A41B820F7CAF76FA6A2A4BA7DF090.new\\_portal-svr-11](http://data.seoul.go.kr/dataList/OA-11572/A/1/datasetView.do;jsessionid=395A41B820F7CAF76FA6A2A4BA7DF090.new_portal-svr-11))

### 환승역 부족 및 혼잡도

-서울교통공사 지하철 혼잡도 정보 ([https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer\\_data\\_information](https://www.data.go.kr/data/15071311/fileData.do?recommendDataYn=Y#layer_data_information))

-서울교통공사 역사건축정보 ([https://data.kric.go.kr/rips/M\\_01\\_01/intro.do](https://data.kric.go.kr/rips/M_01_01/intro.do))

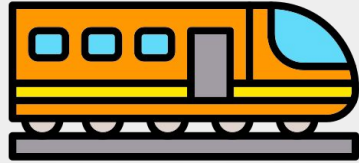
### API 및 크롤링

-서울시 역의 위치 데이터(카카오 API)

-네이버 지하철 (<https://map.naver.com/v5/subway/1000/-/-/?c=15,0,0,0,dh>)

### 연령 계층별 인구

-주요 연령계층별 추계인구(생산연령인구, 고령인구 등) / 전국 ([https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1BPA003](https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1BPA003))



THANK YOU FOR LISTENING