

Introduction to Data Programming Mid Project

# 좌석 개선을 통한 대중교통 이용 편의성 증대 방안 연구

강성운, 정민우

# Agenda

- 1 연구 배경
- 2 문제 정의-데이터 분석을 활용하여
- 3 연구 주제 및 목표
- 4 향후 계획

# 연구 배경

사회 : 사회일반

## "압사하는거 아니냐"...의자 없애려는 지하철에 쏟아진 불만

중앙일보 | 입력 2023.11.02 17:04 업데이트 2023.11.02 22:40



## 출퇴근 지하철 일부 '좌석 다 빼고' 혼잡도 줄인다

파이낸셜뉴스 | 입력 2023.11.01 06:00 수정 2023.11.01 06:00

내년 1월중 지하철 4·7호선에서 시범 적용  
플랫폼 계단 근처 혼잡도 높은 객차 1~2칸 대상  
"효과 입증되면 추후 확대 시행"



## 서울 지하철 4·7호선, 혼잡 완화 위해 '좌석 없는 칸' 도입

내년 1월부터 시범 운영

## 내년 출퇴근 시간대 지하철 4·7호선 좌석 없는 열차 달린다

김표형 기자 | 구독 + | 입력 2023.11.01 15:03 수정 2023.11.01 15:13 | 11면

0 0

| 혼잡도 최대 40% 완화 기대

많이 본 뉴스

# 문제 정의

01

문제 정의 - 사용데이터

02

문제 정의 - 사용코드

03

문제 정의 - 분석결과

# 문제 정의 - 사용데이터

## 데이터 정보

공개일자	2017.12.26.	최신수정일자	2022.04.12.
갱신주기	정기(매년)	분류	교통
원본시스템		저작권자	서울특별시 디지털정책관
제공기관	서울특별시	제공부서	디지털정책관 빅데이터담당관
담당자	최성용 (02-2133-4365)		
원본형태	DB	제3저작권자	없음
라이선스	 저작권자표시(BY): 이용이나 변경 및 2차적 저작물의 작성을 포함한 자유이용을 허락합니다.		
관련태그	도보, 버스, 승용차, 오토바이, 자전거, 지하철, 택시, 통근, 통학, 교통, 서울서베이, 등교		

서울시 통근·통학시 이용하는 교통수단 통계 /자료갱신일 : 2023-09-01 / 수록기간 2005년 ~ 2022년

<http://data.seoul.go.kr/dataList/10283/S/2/datasetView.do>

# 문제 정의 - 사용코드

```
#install.packages("readr")
#install.packages("dplyr")

# csv 파일을 읽기
library(readr)
library(dplyr)

file_path <- 'C:/Users/WOON/Downloads/Transportation Data.csv'
df <- read_csv(file_path)

# 열 합쳐 '버스를 이용하는 이동'과 '이용하지 않는 이동' 데이터 작성
selected_columns_df <- df %>%
  select(버스, `지하철(철도)`, `버스 + 지하철`)

# 결측치 제거
filtered_df <- selected_columns_df %>%
  filter(
    between(버스, mean(버스, na.rm = TRUE) - sd(버스, na.rm = TRUE),
            mean(버스, na.rm = TRUE) + sd(버스, na.rm = TRUE)),
    between(`지하철(철도)`, mean(`지하철(철도)`, na.rm = TRUE) -
            sd(`지하철(철도)`, na.rm = TRUE),
            mean(`지하철(철도)`, na.rm = TRUE) + sd(`지하철(철도)`, na.rm = TRUE)),
    between(`버스 + 지하철`, mean(`버스 + 지하철`, na.rm = TRUE) - sd(`버스 + 지하철`, na.rm = TRUE),
            mean(`버스 + 지하철`, na.rm = TRUE) + sd(`버스 + 지하철`, na.rm = TRUE))
  )

# 'Using Bus'와 'Not Using Bus' 열을 만들고
# 첫 번째 열에 행 번호 추가
final_df <- filtered_df %>%
  mutate(ID = row_number(), # 행 번호 추가
         `Using Bus` = 버스 + `버스 + 지하철`,
         `Not Using Bus` = `지하철(철도)` ) %>%
  select(ID, `Using Bus`, `Not Using Bus`) # 열 순서 조정

# 각 열의 평균과 표준편차 계산
summary_stats <- final_df %>%
  summarise(
    Mean_Using_Bus = mean(`Using Bus`, na.rm = TRUE),
    SD_Using_Bus = sd(`Using Bus`, na.rm = TRUE),
    Mean_Not_Using_Bus = mean(`Not Using Bus`, na.rm = TRUE),
    SD_Not_Using_Bus = sd(`Not Using Bus`, na.rm = TRUE)
  )

# 마지막 행 평균, 표준편차 추가
final_df <- rbind(final_df,
                  c("Mean", summary_stats$Mean_Using_Bus,
                    summary_stats$Mean_Not_Using_Bus,
                    c("S.D", summary_stats$SD_Using_Bus,
                      summary_stats$SD_Not_Using_Bus)))

# 데이터 프레임 열 이름 설정
colnames(final_df) <- c("ID", "Using Bus", "Not Using Bus")

# 결과를 출력 및 CSV 파일 저장
print(final_df)
write_csv(final_df, 'C:/Users/WOON/Downloads/final_data.csv')
```



R 이용, csv 파일 데이터를 1시그마 이내의 데이터만으로 **전처리** 및 **평균, 표준편차** 계산하여 새로 저장하는 코드

## 문제 정의 - 분석결과

	A	B	C
1	ID	Using Bus	Not Using Bu
2	1	44.5	14.3
3	2	38.9	19
4	3	38.1	17.5
5	4	40.4	14.6
6	5	32.3	17.7
7	6	49.8	13.8
8	7	44	18.7
9	8	36.7	20.3

52	51	35.2	19.3
53	52	41.7	15
54	53	49.1	16
55	54	38.5	15.9
56	55	39.4	15.4
57	56	42.4	15.5
58	57	38.4	20.4
59	Mean	40.72456	16.56667
60	S.D	3.842956	2.412714

분석 결과 최근 3년간 서울시의 통근 및 통학 방법중 지하철 '만'을 이용하는 경우는 16.57%에 불과했다. 통근의 옵션에 버스를 이용하는 방법이 포함 되어있는 방법이 40.72%에 해당된다.

# 연구 주제 및 기대효과



좌석 개선을 통한 대중교통 이용  
편의성을 효과적으로 증대 하려면  
버스의 좌석의 개선도 필수적이다.

## 기대효과

1. 지하철 '만'을 개선하는 16% 정도 의 개선효과 을 벗어난 57%의 진짜 대중교통 문제 해결
2. 지하철 포화도 못지 않은 버스 포화도 문제 해결을 단순한 해결법으로 해결
3. 대중교통 수송능력 증대로 배차를 최적화 하여 교통비 상승 억제



# 추후 일정 (Gantt chart)

	7주차	8주차	9주차	10주차	11주차	12주차	13주차
팀 선정							
아이디어 브레인 스토밍							
대주제 선정							
데이터 수집							
주제 선정을 위한 데이터 분석							
문제 정의							
주제 구체화							
중간 발표 자료 준비							
중간 발표							
데이터 연구							
연구 결과 해석 및 정리							
최종 발표 자료 준비							
최종 발표							

Introduction to Data Programming Mid Project

**THANK YOU  
FOR WATCHING!**