

# SMU

AI+X 선도인재양성 프로그램

데이터분석과 인사이트 도출

생각할 것  
기반  
대중 교통 수요 분석



LEE  
HOJUN

이 호 준

(E-mail : lee.hojun@kt.com)

### 주요프로젝트

관제 데이터 활용, 이상징후 판단 및 동적 임계치 모델링  
관제 시스템 GENIO+ AI적용 및 고도화 프로젝트  
KT 로그인 관제 시스템 구축 프로젝트  
KT 로그 아키텍처 V-TF

### 강의 이력

AI (Machine Learning) 사내 멘토링 (KT IT부문 인프라본부, 2021)  
KT 신입사원, 인턴 AI 과제 코칭 및 멘토링(KT, 2020~2021)  
KT AIVLE SCHOOL 코치 (2021~)  
상명대 AI Jam 코칭 및 멘토링 (2022)



<https://www.sli.do/>

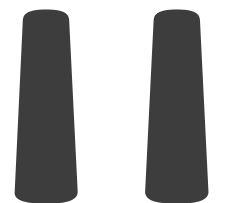
Enter code : #3649668

# 신입 개발자를 위한 PCM

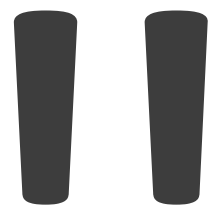
Programmer Competency Matrix

**Subject**  
오늘의 주제

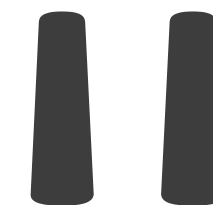
!! 오늘부터 **미니프로젝트**를 통해  
**서울시 생활정보 데이터로**  
**대중교통 수요분석**을  
**해보겠습니다!**



**Subject**  
오늘의 주제



# 서울시 생활정보데이터(인구, 상권) 기반 대중교통(버스) 수요 분석



*Example*

실제 사례 소개

# 빅데이터를 활용한 서울시 심야버스 노선 선정 성공 사례

- 기간 : 2013 ~ 현재
- 세계 최초 빅데이터 활용 교통 정책 반영 케이스
- 정책 참여 그룹 : 서울시, KT



## Example

# 실제 사례 소개 – 노선 결정 AS IS – TO BE

## AS - IS



- 교통 전문가의 직관에 의존
- 심야 교통 데이터의 부재

## TO - BE



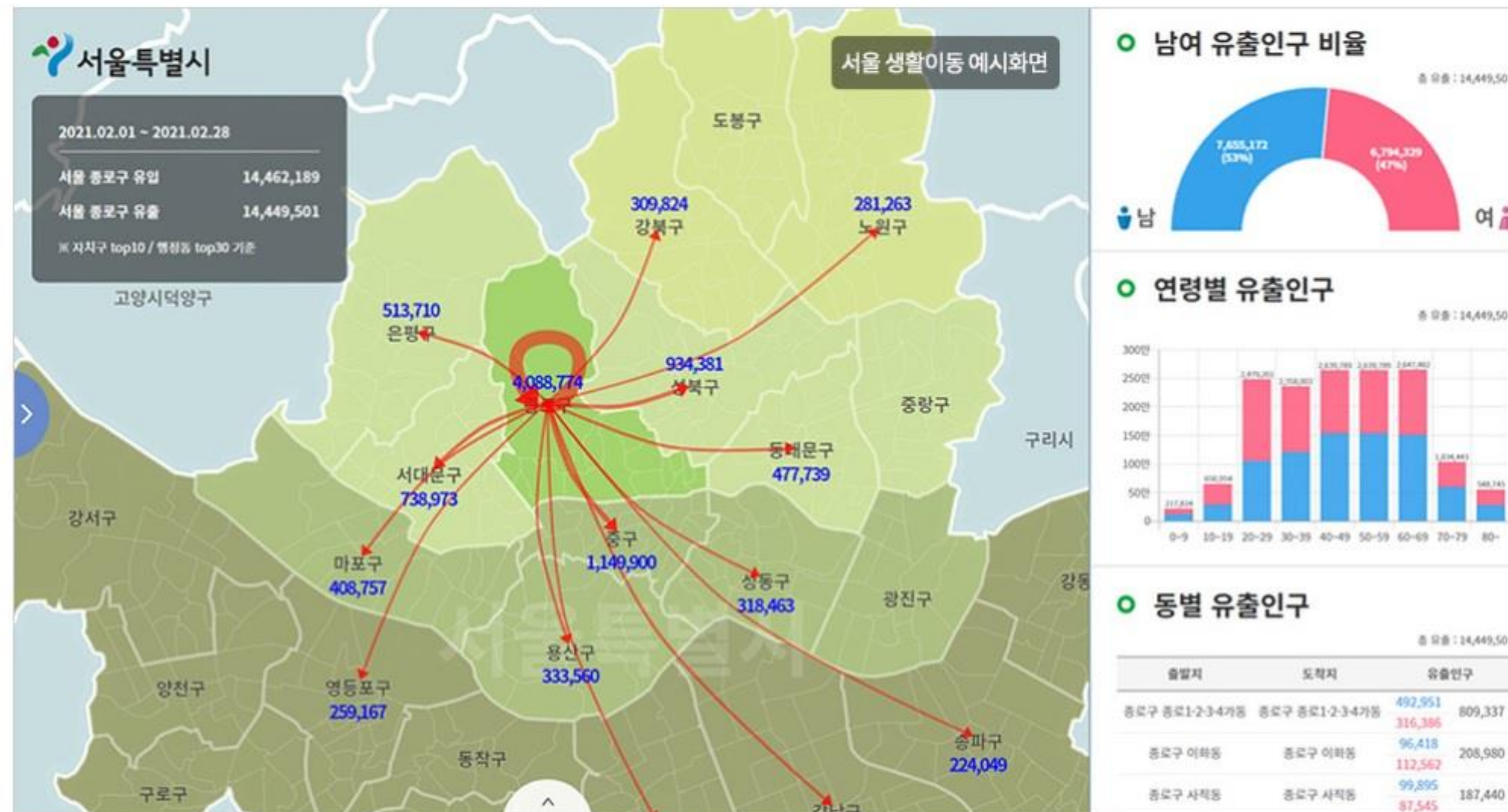


## Example

### 실제 사례 소개 – 서울 생활 이동



서울 생활이동 사이트 바로가기



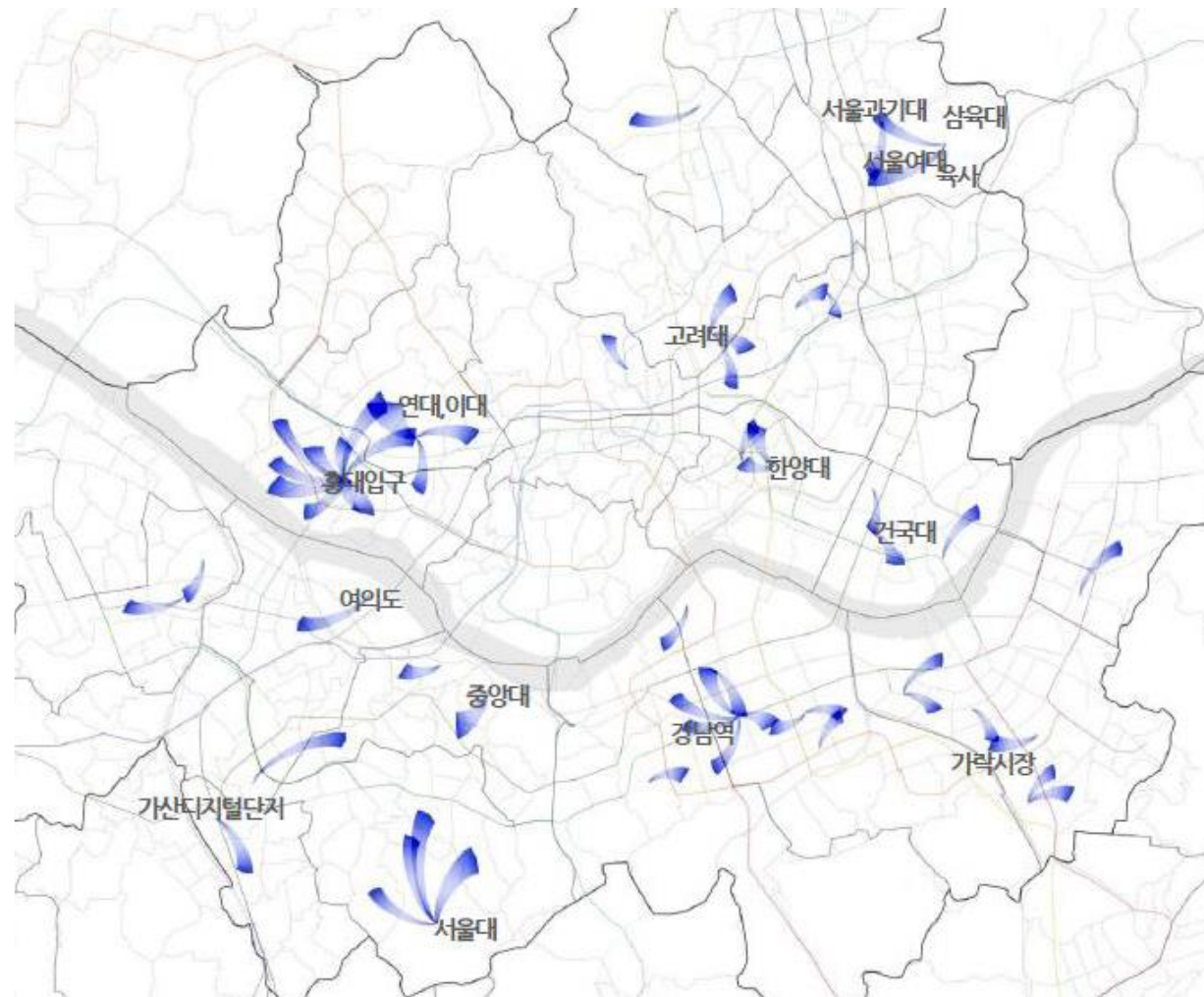
<https://data.seoul.go.kr/dataVisual/seoul/seoulLivingMigration.do>



## Example

# 실제 사례 소개 – 서울 생활 이동 분석 케이스

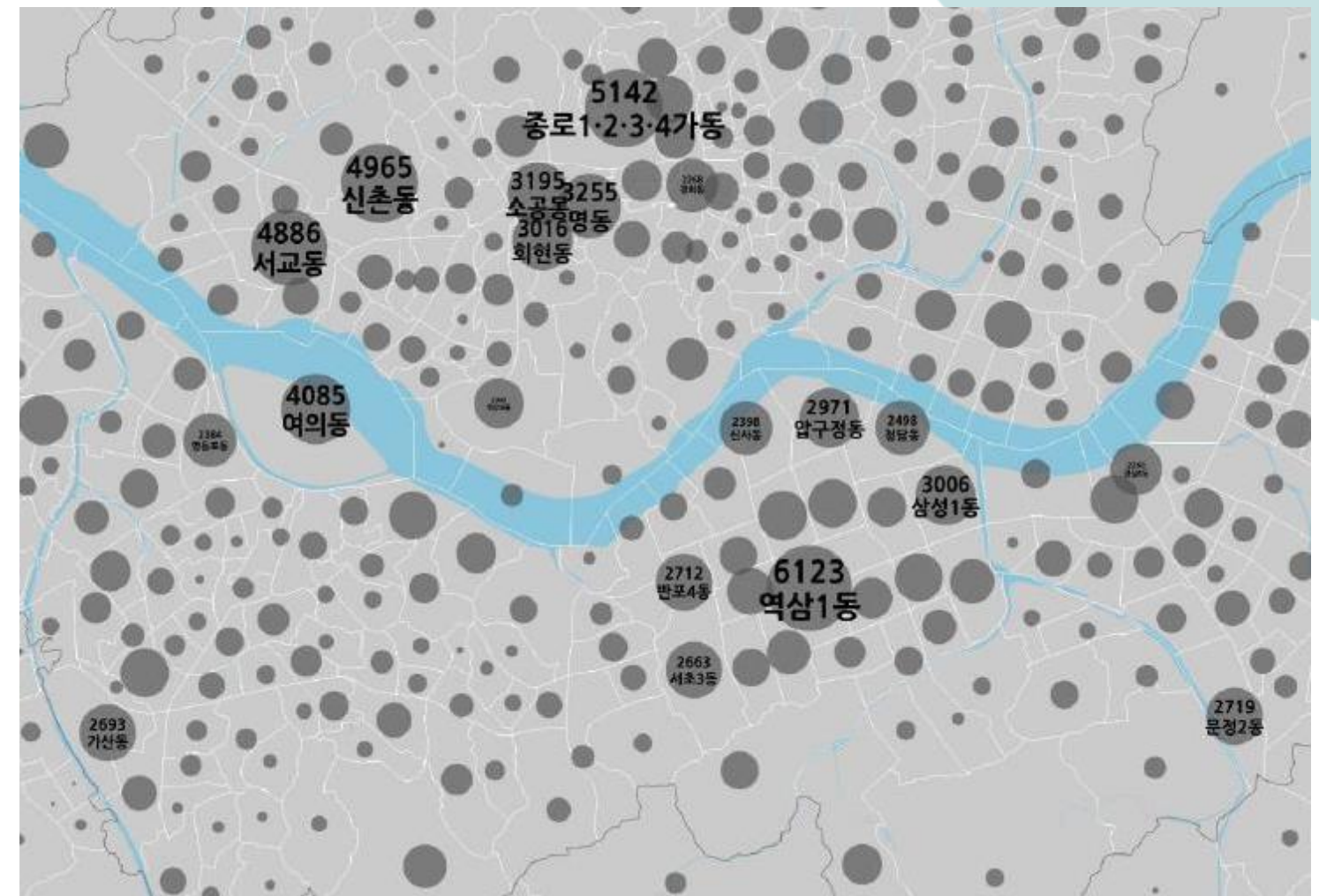
## Case 1



평일 심야 근거리 이동은 주로 대학가

-> 근거리 교통망 확충 및 가로등 설치 등  
안전 정책과도 연계

## Case 2



주말 출근 장소 순위 분석

-> 강남, 명동, 홍대 등  
사람이 많이 모이는 곳의 서비스업 종사자

## Example

### 실제 사례 소개 – 서울 생활 이동

## 프로젝트 활용 Data 예시

	대상연월	요일	도착시간	출발 시군구 코드	도착 시군구 코드	성별	나이	이동유형	평균 이동 시간(분)	이동인구(합)
0	202201	일	0	11010	11010	F	10	EH	10	23.44
1	202201	일	0	11010	11010	F	10	HE	20	5.87
2	202201	일	0	11010	11010	F	10	HW	10	5.82
3	202201	일	0	11010	11010	F	10	HH	10	5.97
4	202201	일	0	11010	11010	F	15	HE	10	4.08
...	...	...	...	...	...	...	...	...	...	...

“2022년 1월 일요일에 11010(종로구)에서 11010(종로구)로 이동하여  
0시~1시 사이에 도착했고 집에서 기타장소로 이동한 10~15세 여성은 5.87명  
이고 이들의 평균 이동시간은 20분이다”



**도메인 이해**

**우리가 먼저 알아두어야  
할 것은 무엇일까요?**



- 분야 관련 Issue
- 관련 지식

## *Domain Knowledge*

### 교통 분야 관련 issue

- **이용자의 수요 반영이 중요**
- **이용 패턴 정보 반영, 수요자 중심 정책 설계 가능**
- **공공 데이터 활용, 버스 노선 관련 인사이트 도출 및 정책 수립**



## *Domain Knowledge*

### 관련 지식 – 서울시 관련

1. 서울은 몇 개의 구로 이루어져 있을까요?
2. 인구가 가장 많은 구는? 가장 적은 구는?
3. 면적이 가장 큰 구는? 면적이 가장 작은 구는?
4. 도로 길이 합이 제일 긴 구는?(고속도로 제외)
5. 도로 면적 합이 제일 큰 구는?

## *Domain Knowledge*

### 관련 지식 – 버스 노선

1. 서울 버스 노선의 수는? (지선 + 간선 만)
2. 야간 노선의 운행 시작 시간은?
3. 서울내 버스 노선의 길이는? (지선 + 간선 만)

Outline

프로젝트 개요

서울시 생활정보 기반

대중교통 수요 분석

## — • 목표

서울시 제공 공공데이터를 활용,  
버스 노선 수요를 분석하고  
인사이트를 도출하여 개선안을 제시한다



Outline

문제정의

# 우리가 풀어야 할 문제는 무엇일까요?

---

서울 유동인구 등의 인구 정보와 버스 승하차 인원 및 운행 노선 분석을 통하여

**버스 시설 추가 필요 대상 지역**을 선정한다.



*Dataset*

## 데이터셋 소개

- **서울 버스 승하차 이용 데이터**
- **서울 구별 유동 인구 데이터**
- 서울 구별 주민 등록 인구 데이터
- 서울 구별 업종 등록 데이터

(2022년 04월 서울시 구별 데이터로 한정)

출처 : 서울시 공공데이터 포털





*Pre-Check*  
사전 체크



활용 라이브러리



유의 사항



## Library

### 라이브러리 지식 체크

- ✓ **Pandas : 데이터 핸들링과 분석에 유용한 라이브러리**
- ✓ **Numpy : 다차원 배열 및 수학적 기능을 지원하는 라이브러리**
- ✓ **Seaborn : 통계 데이터 시각화 라이브러리**
- ✓ **Matplotlib : 일반적인 파이썬 시각화 라이브러리**
- ✓ **Scipy : 수학, 과학, 등의 분야에서 많이 사용하는 계산 라이브러리**

*Important*

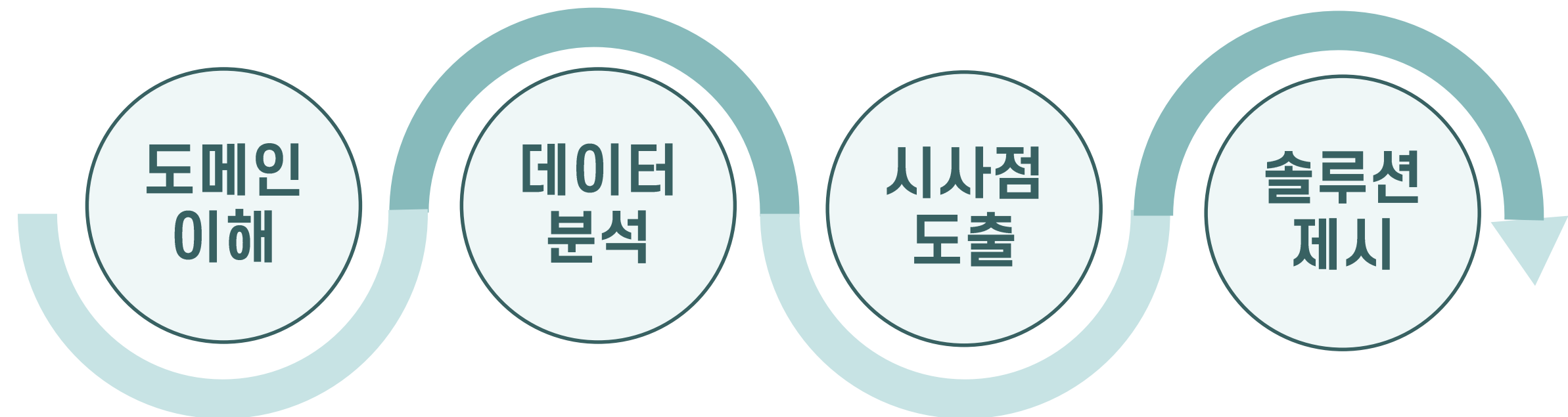
유의사항

본 프로젝트에서는 데이터 분석을 통해  
인사이트 도출을 시도하는 연습을 합니다.

**입체적인 접근**이 매우 중요합니다.

*Process*  
실습 프로세스

# 전체 과정 실습 EDA를 활용한 문제 해결의 'A~Z' 경험



- 실제 공공데이터를 기반으로 데이터 분석 노하우를 공유
- 문제 인식부터 데이터 분석을 활용한 인사이트 도출까지 단계별 실습 진행

*Practice*

## 미니프로젝트 시작



**미니 프로젝트  
시작합니다!**

Jupyter notebook을 실행해주세요

모델링을 위한 전과정 프로젝트 수행, 이제 시작합니다!



*Practice*

미니프로젝트 완료

# 미니프로젝트 완료!

## 데이터 분석 및 인사이트 도출

미니프로젝트를 통해 배우는 데이터 분석

자기주도형 실습으로 진행된 프로젝트,  
어떤걸 느끼셨나요?



## Summary

### 미니프로젝트 완료

서울시 구별 상권 등록 정보

서울시 인구 정보

서울시 버스 노선 관련 정보

**도메인 이해**

제공 데이터 분석

제공 데이터 시각화

데이터 이변량 분석

**데이터 분석**

**시사점 도출**

**솔루션 제시**

토론 방법론 학습 및 실습

가설 설정 및 분석

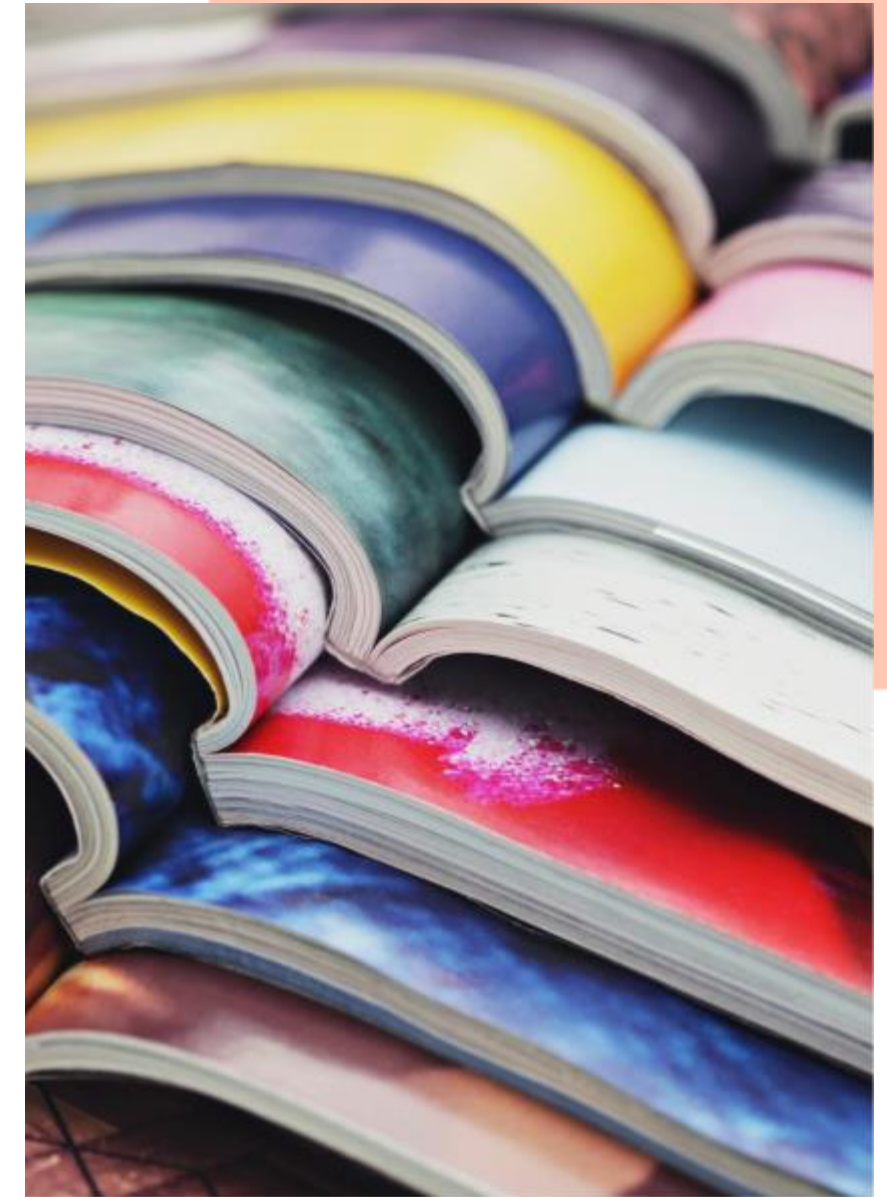
가설 검증 결과 기반 시사점 도의

“어느 구의 버스 노선/정류장 추가가 제일 필요한가요?”



**문제를 해결했다는 성취감**

**데이터 분석에 대한 자신감**



**협업을 통한 성장**

Summary  
마무리

!!

프로젝트를 통한 성취감과  
데이터 분석의 자신감을 얻은 오늘!

즐겁고 행복한

21년 시간 되시길 바랍니다!

!!

AI프로젝트 따라하기

***Thank you !***

**감사합니다.**