

▼ PROJECT DELIVERABLE OF  
SAMSUNG SW-AI ACADEMY FOR YOUTH

PERIOD / 2025.08.25-10.02  
CODE IDENT. / S13P21E203

# in/site

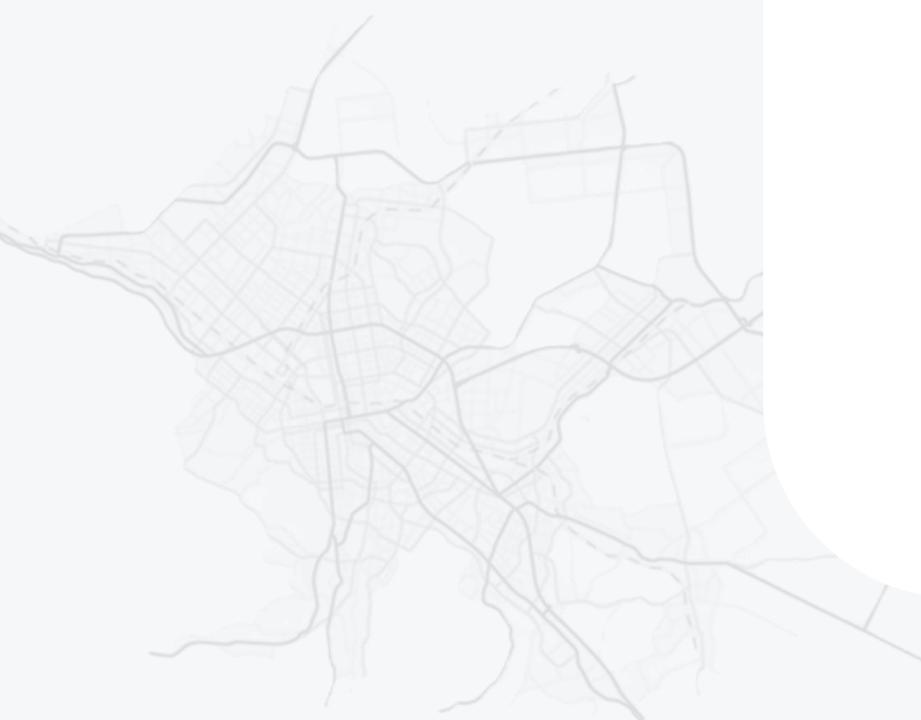
# Final Report

▼ TEAM INSITE / FROM INSIGHT TO SITE -  
POWERED BY DATA.

SEPT 29, 2025

# Index

데이터와 AI로,  
카페 상권 선택을 더 쉽게



- 1 기획 배경 & 서비스 소개
- 2 프로젝트 시연
- 3 주요 기능
- 4 사용 기술 및 모델
- 5 기대효과 & 마무리

# in/site / Our Team



천광민

팀장



남다현

FE개발



정필교

DB/API



김범주

모델개발



김재유

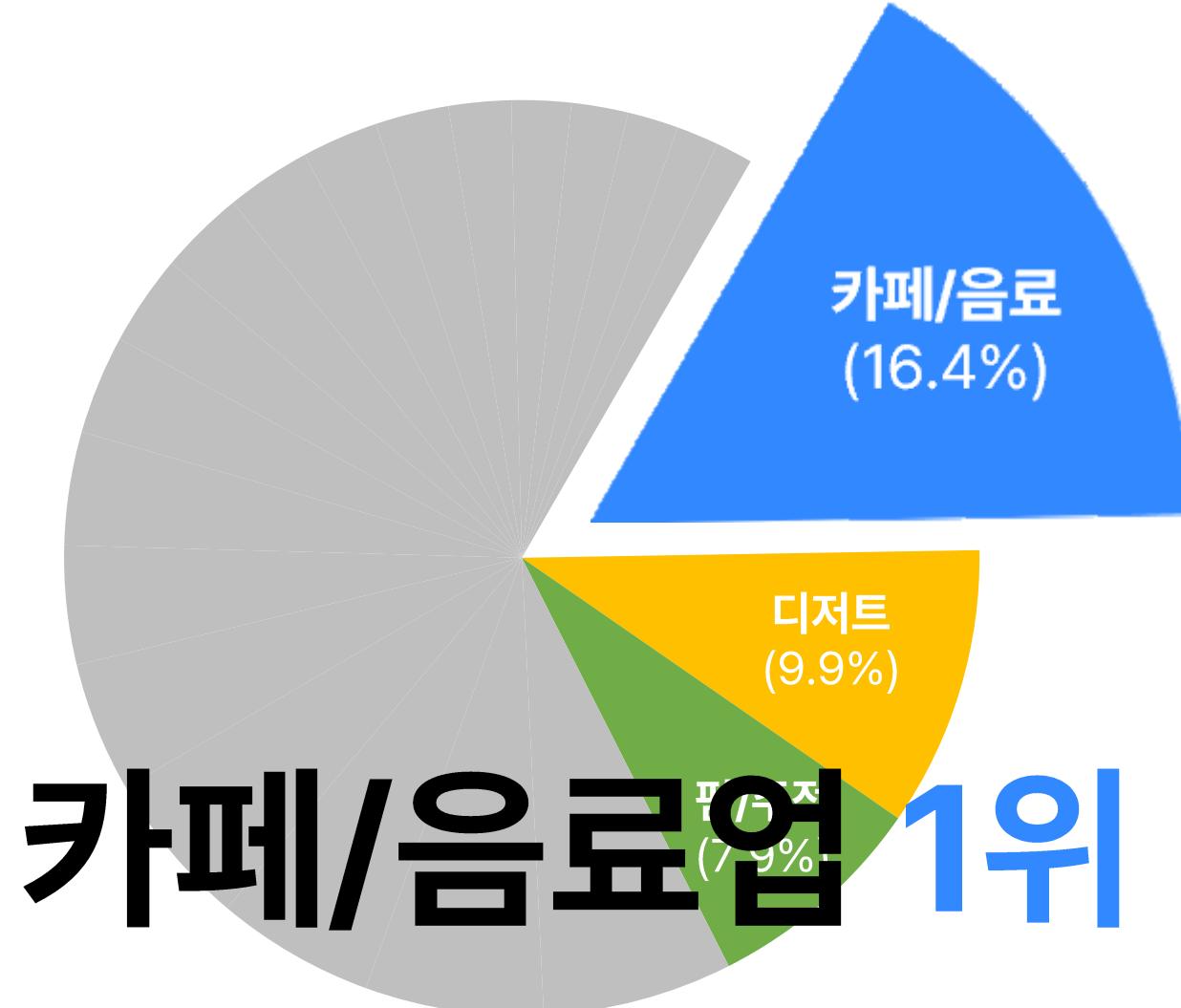
모델개발

2015

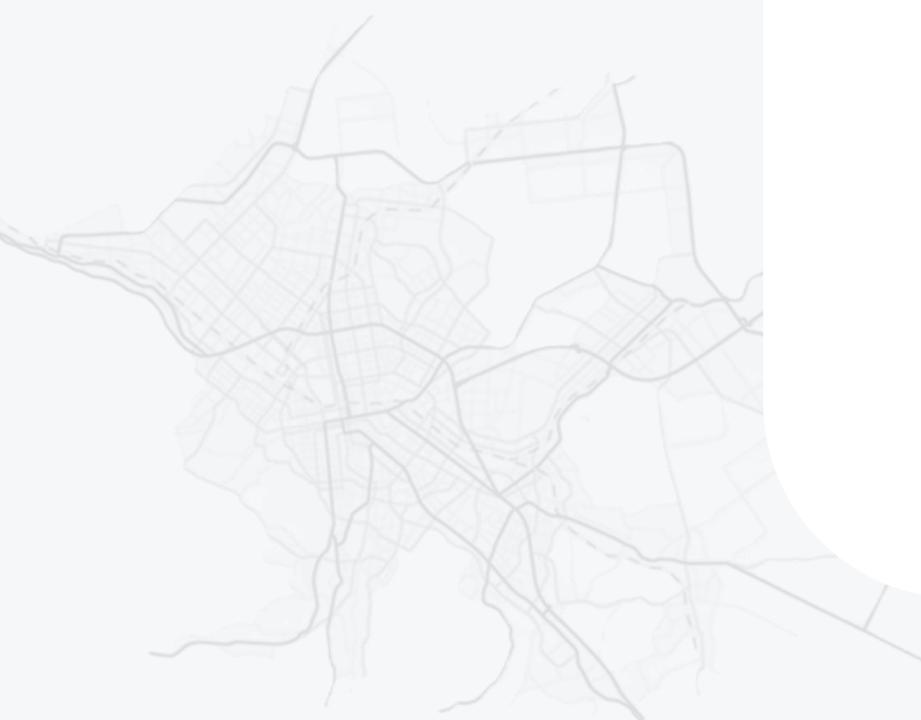
가수

# 카페 창업 선호도

예비창업자 희망 창업  
업종 설문조사  
(출처: 2022 우리동네씨이오)



# 창업 고려 사항



상권분석

입지·공간

자금·비용

운영·인력·관리

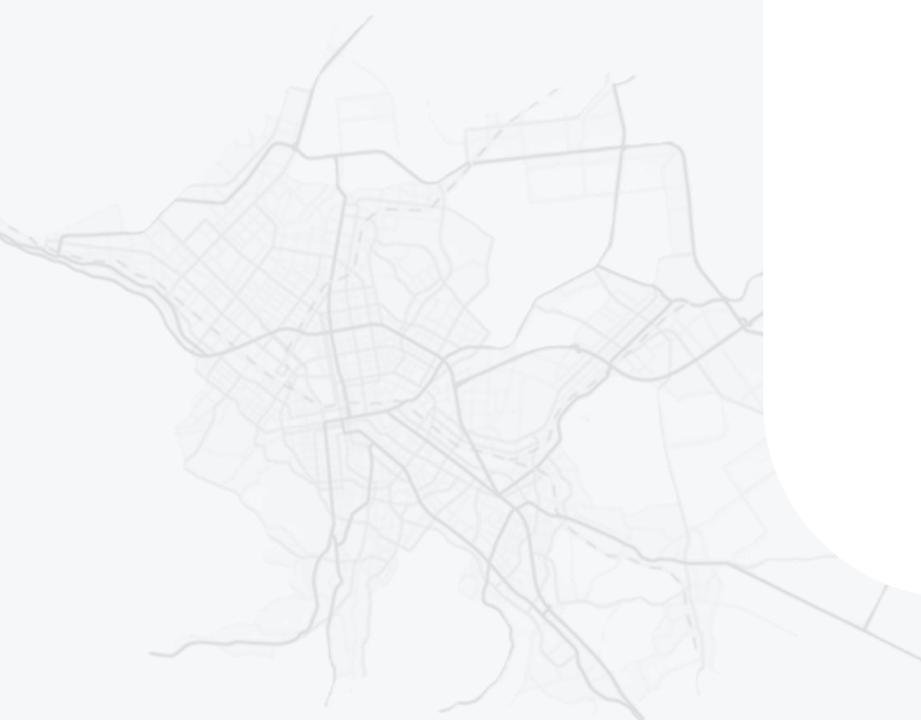
메뉴·분위기

데이터 기반 운영

마케팅 전략

리스크 관리

# 창업 준비 기간

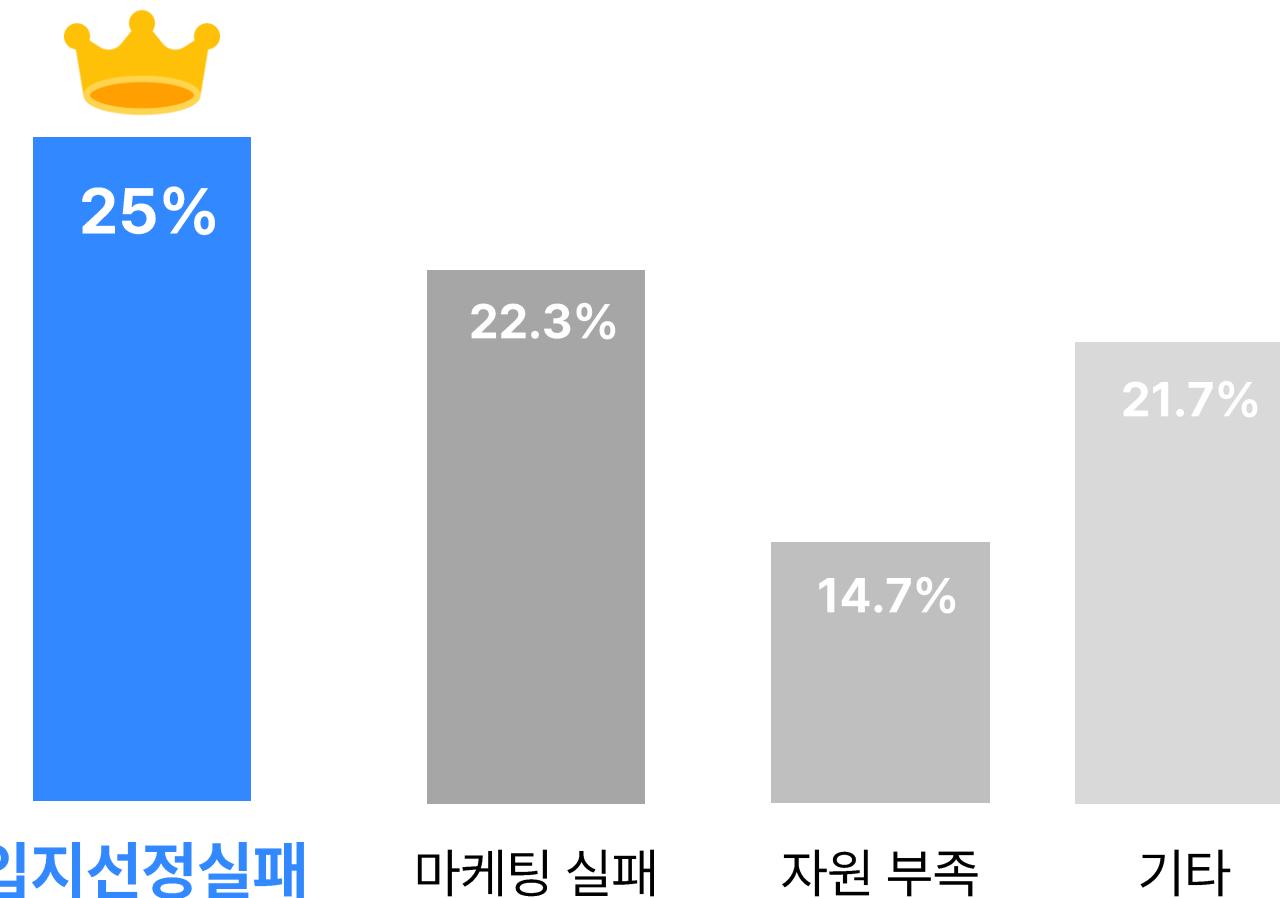
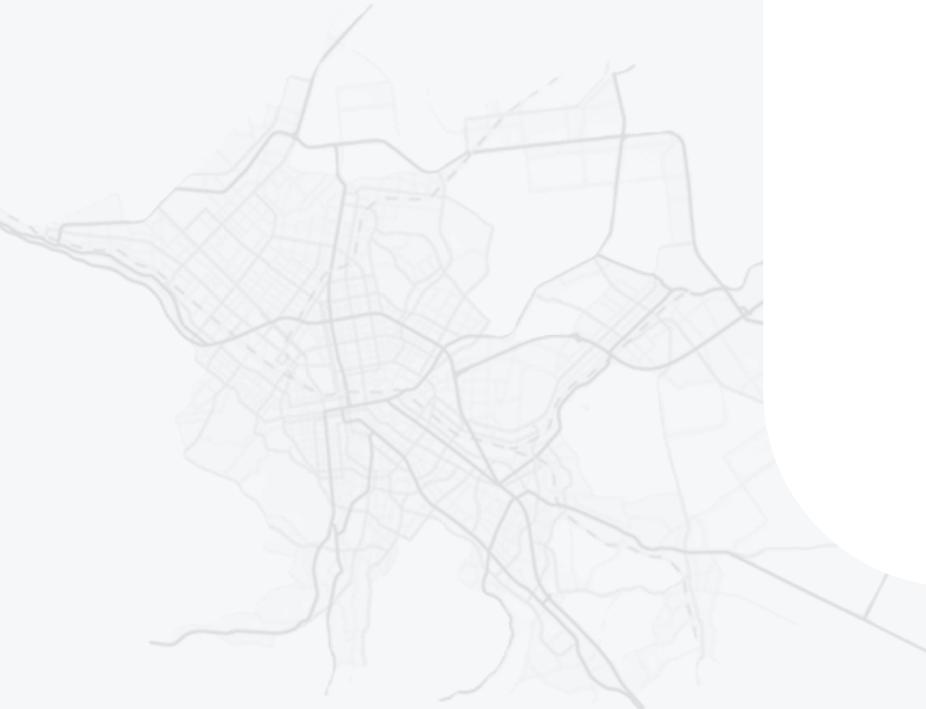


창업 준비 기간(3~6개월)

상권 분석(2주~1개월)

# 폐업 사유

가게 운영 경험 설문  
(출처: 2024 핀다포레스트)





in/site



시연

Development Tools



## 주요 기능

Key Functions

# 추천

Trade Area  
Recommendation

# 비교

Trade Area  
Comparison



**상권 추천**

Trade Area Recommendation

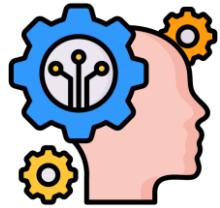
# 👍 상권 추천 | 특장점

Trade Area Recommendation



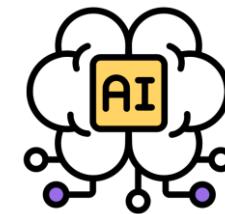
## 맞춤형 추천

사용자가 직접 선택한 자치구  
와 상권 유형(발달/골목)을  
반영하여, 단순히 인기 있는  
곳이 아니라 개인 상황에 맞  
는 최적 상권을 추천



## ML 기반 점수

단일 지표가 아닌 인구·매출·  
경쟁·변화 추세 등 다차원 데  
이터를 종합 분석한 점수를  
활용 → 객관적이고 신뢰도  
높은 추천 가능



## AI 추천 이유

추천 결과에는 단순 점수뿐  
아니라 상권 분위기와 추천  
이유가 함께 제시되어, 사용  
자가 결과를 이해·납득하고  
의사결정에 활용할 수 있음



**상권 비교**

Trade Area Recommendation

# vs 상권 비교 | 특장점

Trade Area Recommendation



## 직관적 비교

인구·점포·매출·상권변화지표  
·종합추천점수를 한 화면에서  
나란히 비교할 수 있어, 조건  
별 차이를 한눈에 확인 가능



## 데이터 시각화

복잡한 수치를 표로만 보여주는 것이 아니라, 차트·그래프로 시각화하여 초보 창업자도 직관적으로 상권의 장단점을 파악할 수 있음



## 의사결정 도움

단순 나열이 아닌 종합추천 점수와 요소별 점수(지속성, 수익성, 접근성, 위험도, 경쟁강도)까지 함께 제공하여, 비교 결과가 곧 창업 의사결정 지표로 활용 가능

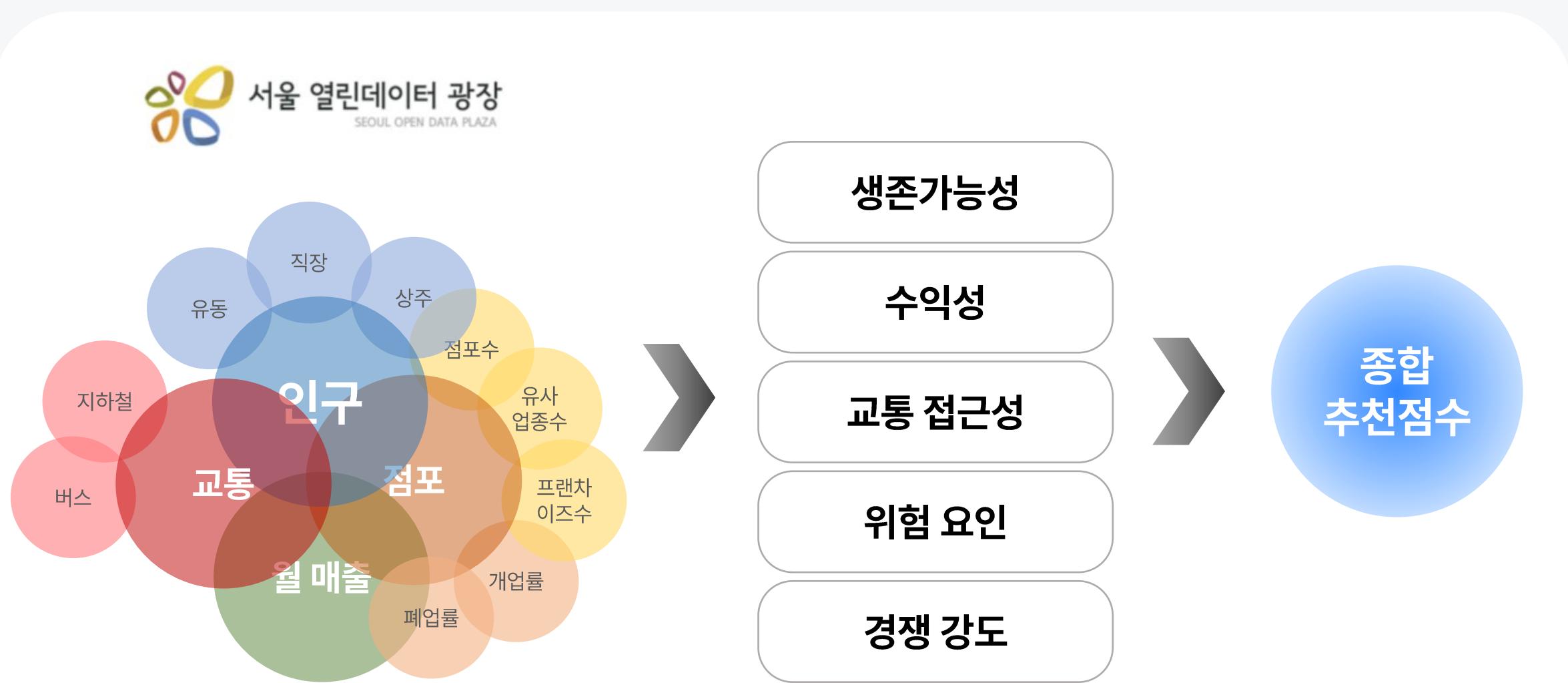


추천 모델

Recommendation Model

# 추천 모델

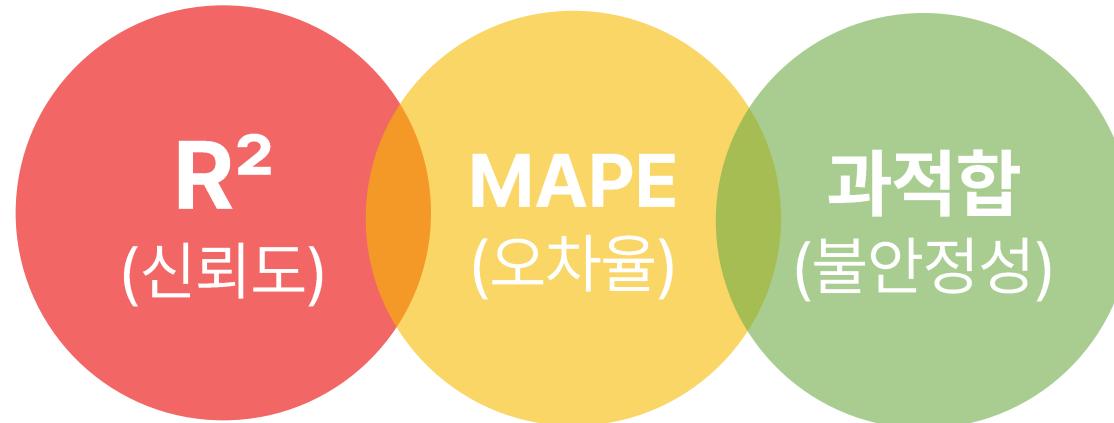
Recommendation Model





# 추천 모델

Recommendation Model



- **R<sup>2</sup>(신뢰도)** 모델이 실제 데이터를 얼마나 잘 예측할 수 있는지를 나타내는 지표입니다.
- **MAPE(오차율)** 예측 값과 실제 값의 차이를 백분율로 나타내어 오차 정도를 보여줍니다.
- **과적합(불안정성)** 학습 데이터만 맞추어 실제 상황에서는 성능이 떨어지는 현상을 방지하는 지표입니다.



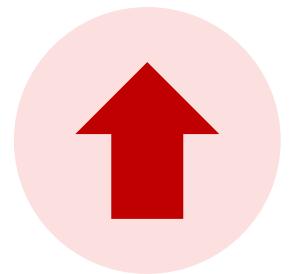
# 추천 모델

Recommendation Model

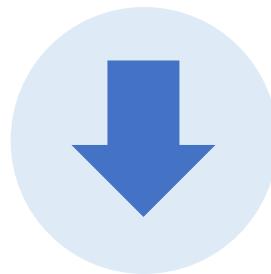
상권 유형	발달상권			골목상권		
	모델명	$R^2$ (신뢰도)	MAPE(오차율)	과적합(불안정성)	$R^2$ (신뢰도)	MAPE(오차율)
<b>HistGradientBoosting</b>	<b>0.8</b>	<b>23.7%</b>	<b>0.004</b>	<b>0.517</b>	<b>47.8%</b>	<b>0.032</b>
RandomForest	0.674	28.8%	0.001	0.401	52.3%	0.036
Ridge	0.712	30.2%	-0.024	0.474	50.1%	0.017
Lasso	0.742	30.8%	-0.018	0.455	50.6%	0.022
ElasticNet	0.633	32.4%	-0.032	0.384	51.9%	0.026
VotingEnsemble	0.719	28.5%	-0.021	0.455	50.4%	0.028
StackingEnsemble	0.730	29.1%	-0.010	0.481	49.9%	0.024

# 추천 모델

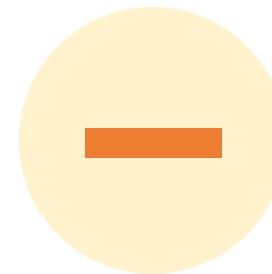
Recommendation Model



높은  
 $R^2$  (신뢰도)



낮은  
**MAPE** (오차율)



0에 수렴  
**과적합** (불안정성)

상권 유형	발달상권			골목상권			
	모델명	$R^2$ (신뢰도)	<b>MAPE</b> (오차율)	과적합 (불안정성)	$R^2$ (신뢰도)	<b>MAPE</b> (오차율)	과적합 (불안정성)
HistGradientBoosting		0.8	23.7%	0.004	0.517	47.8%	0.032

- 과적합(불안정성)이 0보다 큰 경우, OverFitting / 0보다 낮은 경우, UnderFitting

## “HistGradientBoosting” 모델 선정



# 추천 모델

Recommendation Model

# HistGradientBoosting



모델명	$R^2$ (설명도)	MAPE(오차율)	과적합(불안정성)	$R^2$ (설명도)	MAPE(오차율)	과적합(불안정성)
신뢰도는 높고 HistGradientBoosting	0.8	25.7%	낮고	0.5	47.8%	0.32

• 과적합(불안정성)이 0보다 큰 경우, OverFitting / 0보다 낮은 경우, UnderFitting

# 추천 시스템

Recommendation System



데이터 수집 및 로딩

데이터 정제

강화된 극값 처리

고급 모델 학습

추천 시스템

상권 데이터 통합  
업종 필터링  
발달/골목상권 추출  
시간별 데이터 병합

극단적 이상치 제거  
누락값 처리  
로그 변환  
새로운 특성 생성  
(인구/점포 밀도 등)

엄격한 극값 제거  
상위 0.5% 제거  
강건한 특성 생성  
고급 특성 엔지니어링

룰링 백테스트  
다양한 모델 비교  
특성 엔지니어링  
파이프라인  
매출 데이터 정규화  
교차검증

**HistGradientBoosting**  
지속성 +  
수익성 +  
접근성 +  
위험도 +  
경쟁강도  
= 종합추천점수

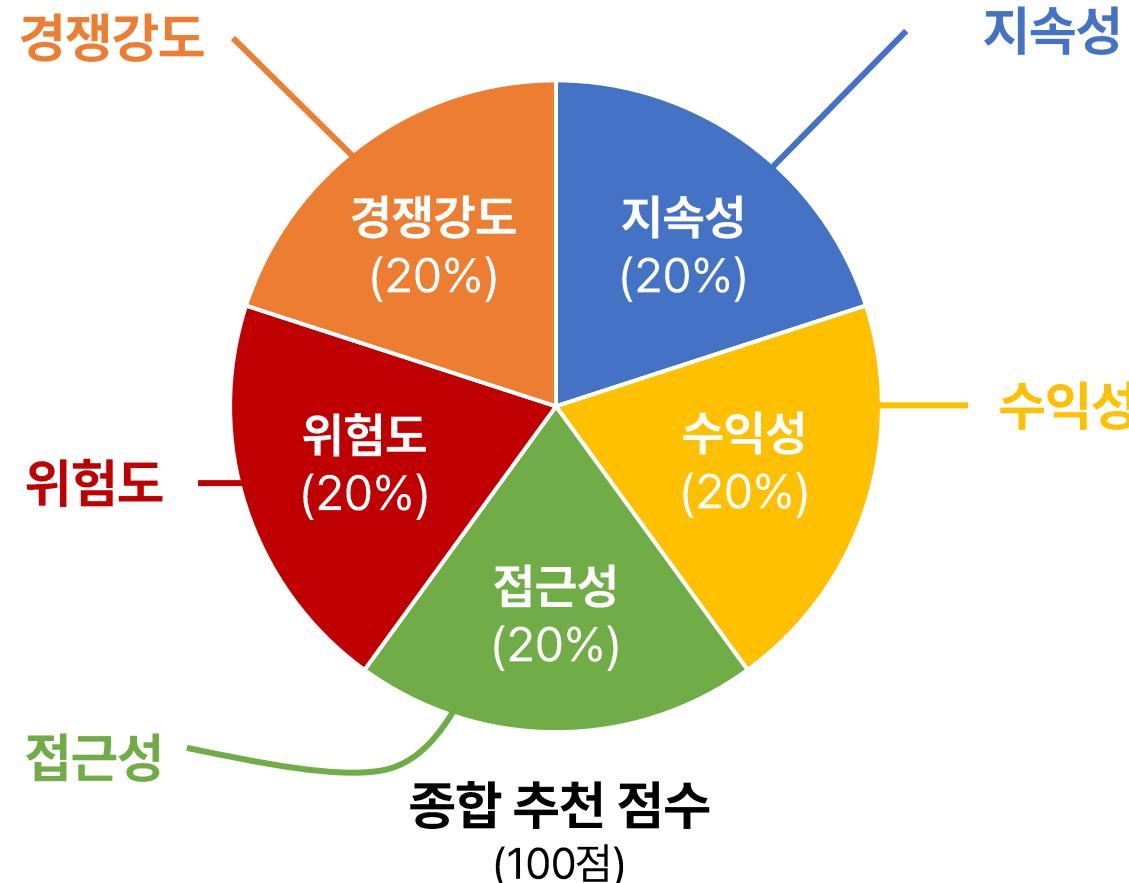
# 추천 시스템

Recommendation System

- 점포 수(50점)
- 유동 대비 점포(10점)
- 점포 밀도(40점)
- 폐업률(5점)

- 지하철역 거리(70점)
- 버스정류장 거리(30점)

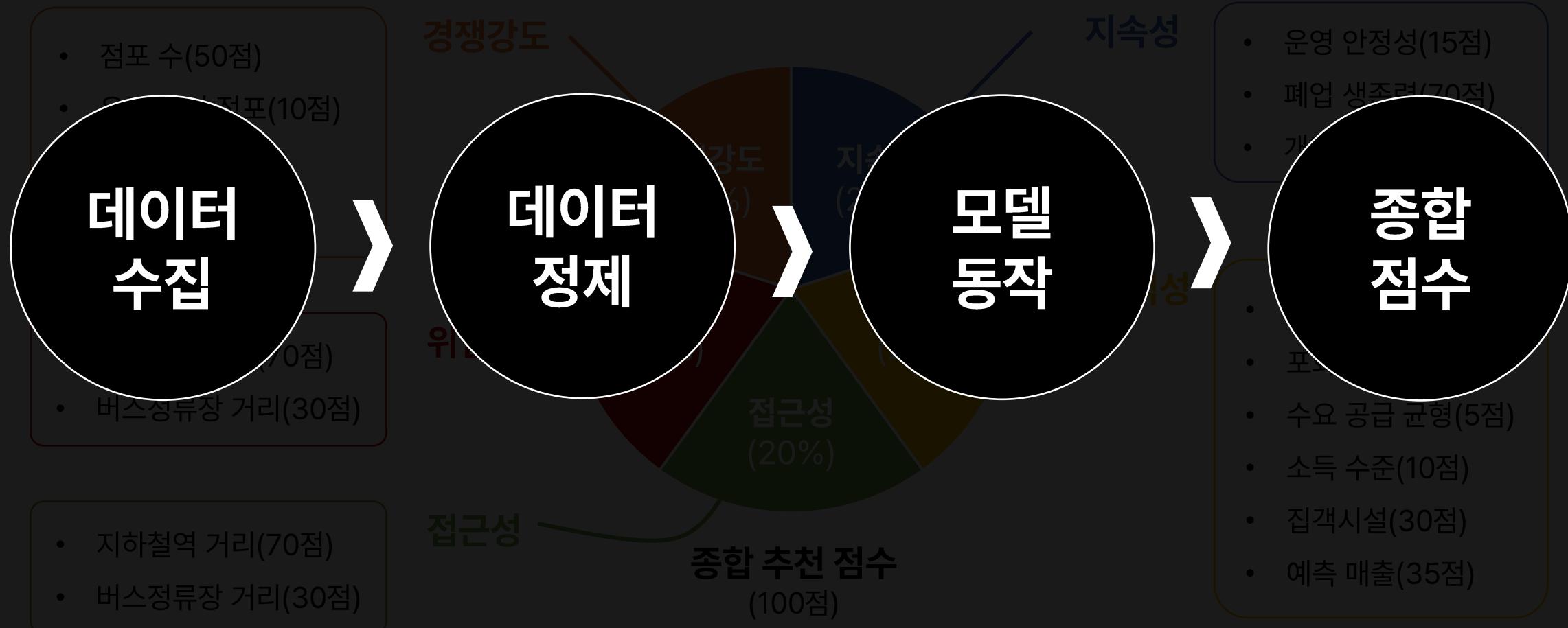
- 지하철역 거리(70점)
- 버스정류장 거리(30점)



- 운영 안정성(15점)
- 폐업 생존력(70점)
- 개업 성장성(15점)

- 시장 규모(10점)
- 포화도(10점)
- 수요 공급 균형(5점)
- 소득 수준(10점)
- 집객시설(30점)
- 예측 매출(35점)

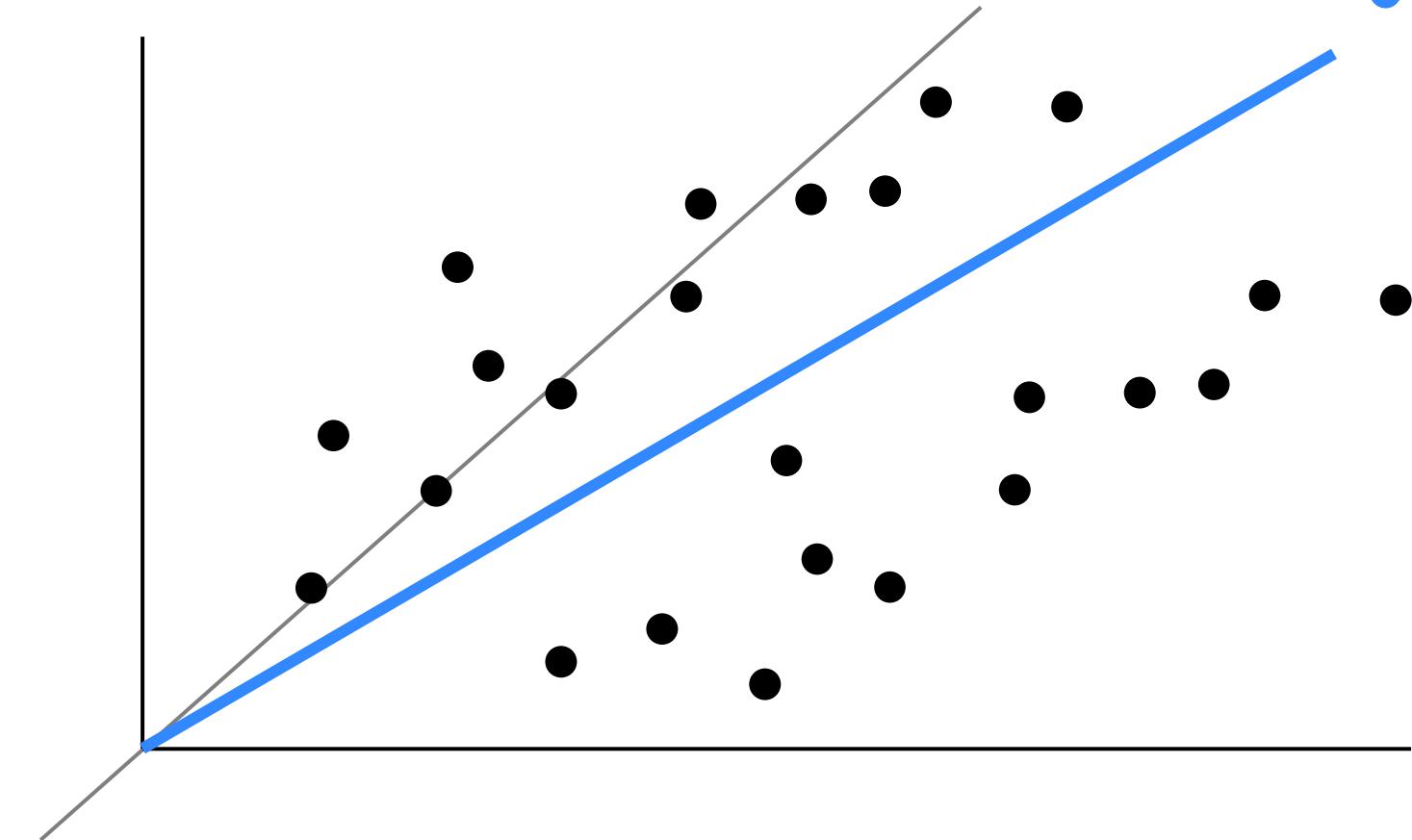
## 추천시스템 요약



# 추천 모델

Recommendation Model

예측치





추천 모델

Recommendation Model

모델 검증은?

분기별  
롤링  
백테스트

Repeated  
K  
Fold

홀드아웃



# 문제 해결

Recommendation Model

2초 캐시

정확하지 않은  
예측 모델



- 이상치 제거
- 코로나 초반 데이터 제외

- 치우쳐진 매출 데이터 분포를 고르게 만들어 평균적인 패턴 생성

- 변수들 간의 곱으로 숨은 관계 찾기
- 필요한 축만 남겨 불필요한 잡음 제거

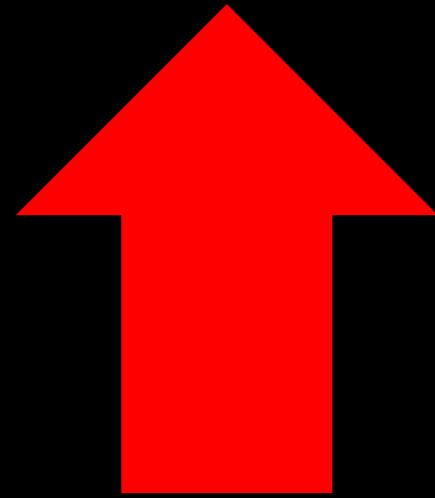
- 평균 대신 중앙값으로 기준을 맞추기

☞ 클린 데이터로 혼란도 감소

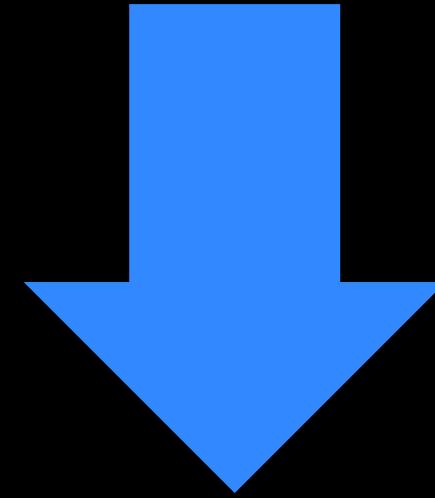
☞ 평균적 패턴 학습

☞ 깊고 정확한 패턴 학습

☞ 예측 오차 줄임



$R^2$   
(신뢰도)



MAPE  
(오차율)

신뢰도

22% 향상

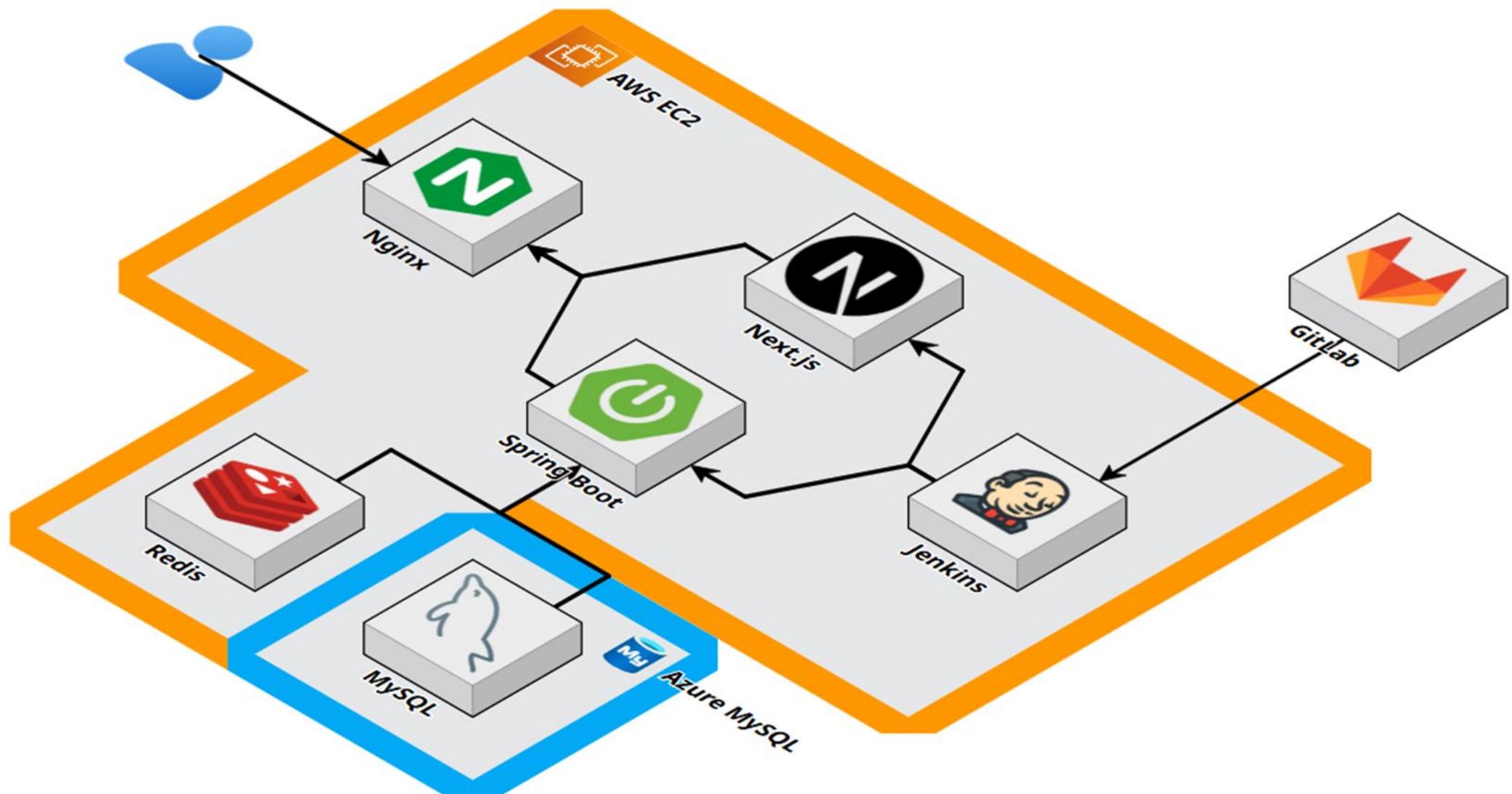
(발달 상권 기준)



# 시스템 구조도

System Architecture

# in/site Architecture

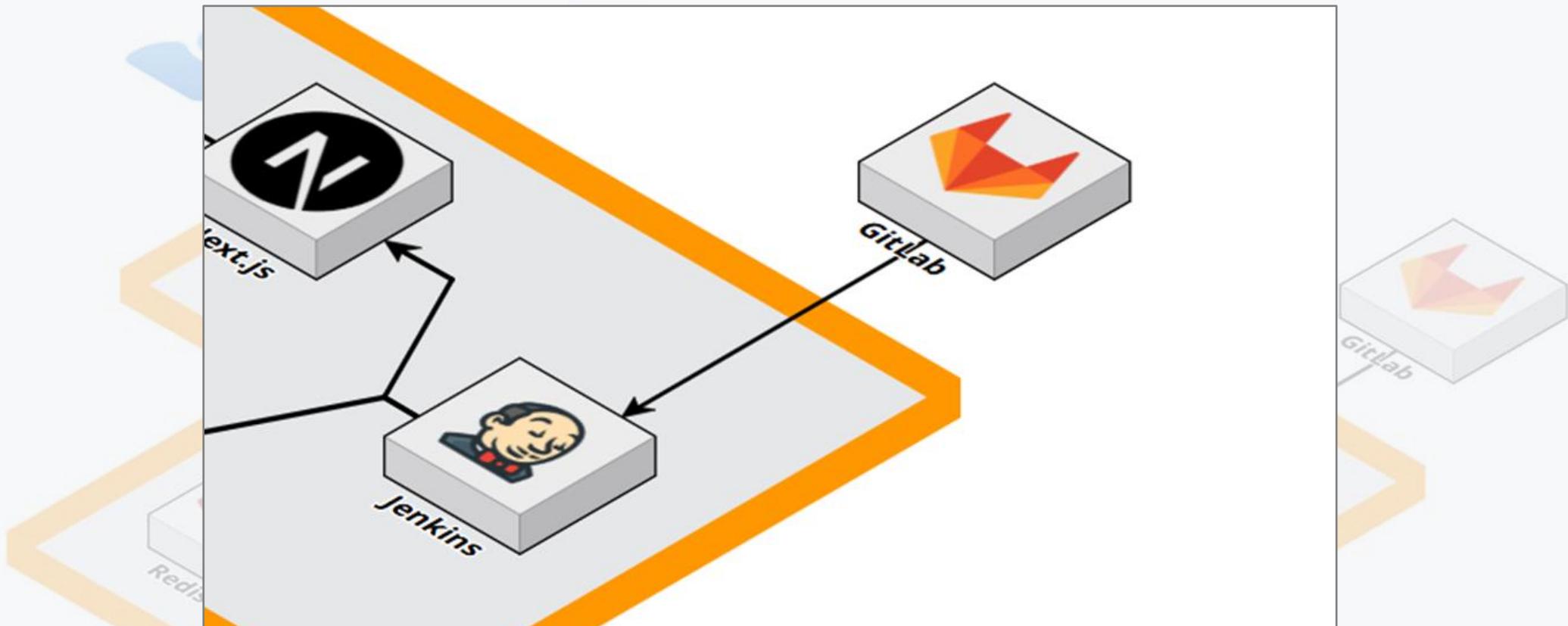




## 1. 클라이언트 요청 및 트래픽 분배

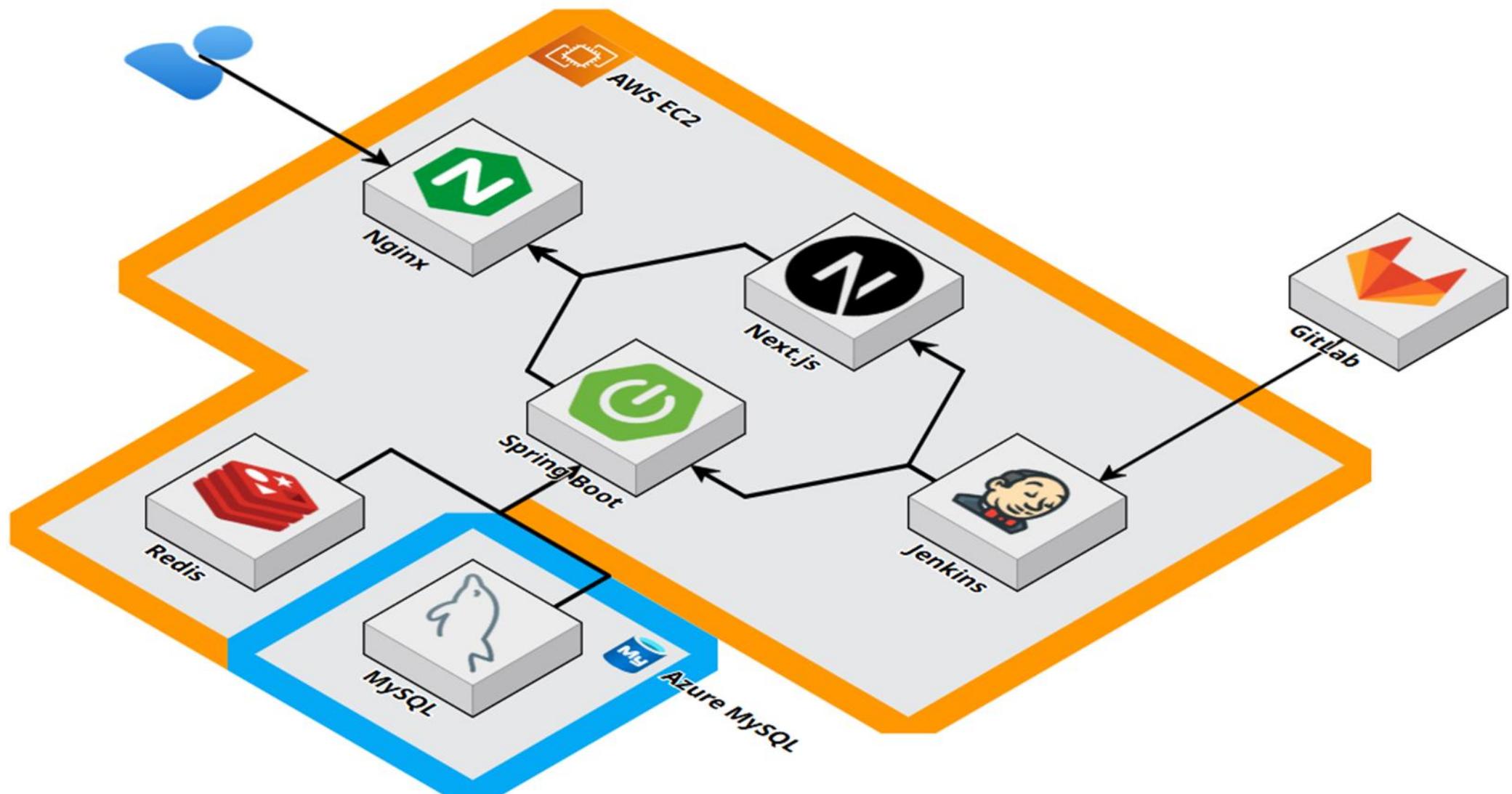


## 2. 캐시(Redis) 및 DB(MySQL)



### 3. 코드 업데이트

# in/site Architecture



FROM INSIGHT  
TO SITE  
POWERED BY  
DATA.

---

in/site