

# 부동산 전/월세 예측 및 B2B

Ai-X APT TEAM

2025.08.29.Fri

01 프로젝트 개요

02 데이터 설명

03 수요 예측 모델의 정확도 평가

04 예측 모델 웹 플랫폼

05 B2B 활용 방안

06 자체 평가 의견

목차

CATALOGUE

01

프로젝트 개요

# 1 프로젝트 개요

## 목표

01

부동산 실거래 데이터로 수요를 예측하고,  
B2B 재고 전략 수립에 활용 가능한 인사이트를 제공

## 데이터 범위

02

기간 : 2024.01.01 ~ 2025.08.18

범위 : 서울특별시 아파트 전, 월세

## Django, Git을 활용한 프로젝트 구현

03

장고를 활용해서 웹페이지를 제작하고  
프로젝트의 예측 프로그램을 제공

02

## 데이터 설명

# 2 데이터 설명

## 1) 데이터를 수집해 2개의 예측 목표 설정

실거래 부동산 데이터를 수집  
출처 :

<https://rt.molit.go.kr/pt/xls/xls.do?mobileAt=>

## 3) 결과값 도출 및 시각화

예측 시스템의 결과값을 기반으로,  
시각적으로 이해하기 쉬운 그래프로 제공



## 2) 데이터 전처리

수집한 CSV 데이터를  
파이썬을 활용하여 전처리

## 4) 예측 시스템

Django를 활용하여,  
데이터와 예측 결과를 제공하는  
웹페이지를 제작

NO	시군구	번지	본번	부번	단지명	전월세구분	전용면적(m)	계약년월	계약일	보증금(만원)	월세금(만원)	총	건축년도	도로명	계약기간	계약구분	갱신요구권사용	총전계약보증금(만원)	종전계약월세(만원)	주택유형
0 1	서울특별시 용산구 효창동	287	287	0	용산롯데캐슬센터포레	전세	59.7900	202412	31	83,000	0	13	2019	백범로 313	202412~202612	신규	-	Nan	Nan	아파트
1 2	서울특별시 종로구 명륜2가	237	237	0	명륜동주상복합아남아파트	전세	50.4900	202412	31	42,000	0	2	1999	혜화로3길 5	202501~202701	신규	-	Nan	Nan	아파트
2 3	서울특별시 중구 신당동	844	844	0	남산타운	월세	84.8800	202412	31	10,000	215	9	2002	다산로 32	202502~202702	신규	-	Nan	Nan	아파트
3 4	서울특별시 용산구 신창동	102	102	0	세방리버하이빌	전세	84.5400	202412	31	62,000	0	15	2005	효창원로15길 16	202505~202705	신규	-	Nan	Nan	아파트
4 5	서울특별시 성동구 상왕십리동	811	811	0	텐즈힐(2단지)	월세	84.9000	202412	31	50,000	160	4	2014	마장로 137	202502~202702	신규	-	Nan	Nan	아파트
5 6	서울특별시 중구 신당동	432-1957	432	1957	신신(2차)	전세	59.9800	202412	31	23,500	0	3	2001	동호로 11자길 22	202502~202702	갱신	-	22,000	0	아파트
6 7	서울특별시 중구 만리동2가	273	273	0	서울역센트럴자이	전세	72.8899	202412	31	63,000	0	2	2017	만리재로 175	202502~202702	갱신	사용	60,000	0	아파트
7 8	서울특별시 용산구 원효로1가	134	134	0	용산원효루미니	월세	43.9656	202412	31	17,344	78	26	2022	원효로 97길 15	202502~202702	갱신	-	17,344	73	아파트
8 9	서울특별시 중구 흥인동	119	119	0	동대문와이즈캐슬	월세	15.5500	202412	31	1,000	75	10	2014	퇴계로 75길 7	202502~202702	신규	-	Nan	Nan	아파트
9 10	서울특별시 종로구 평동	233	233	0	경희궁자이(3단지)	월세	59.8547	202412	31	70,000	80	12	2017	경교장길 35	202502~202702	신규	-	Nan	Nan	아파트

# 03

## 수요 예측모델의 정확도 평가

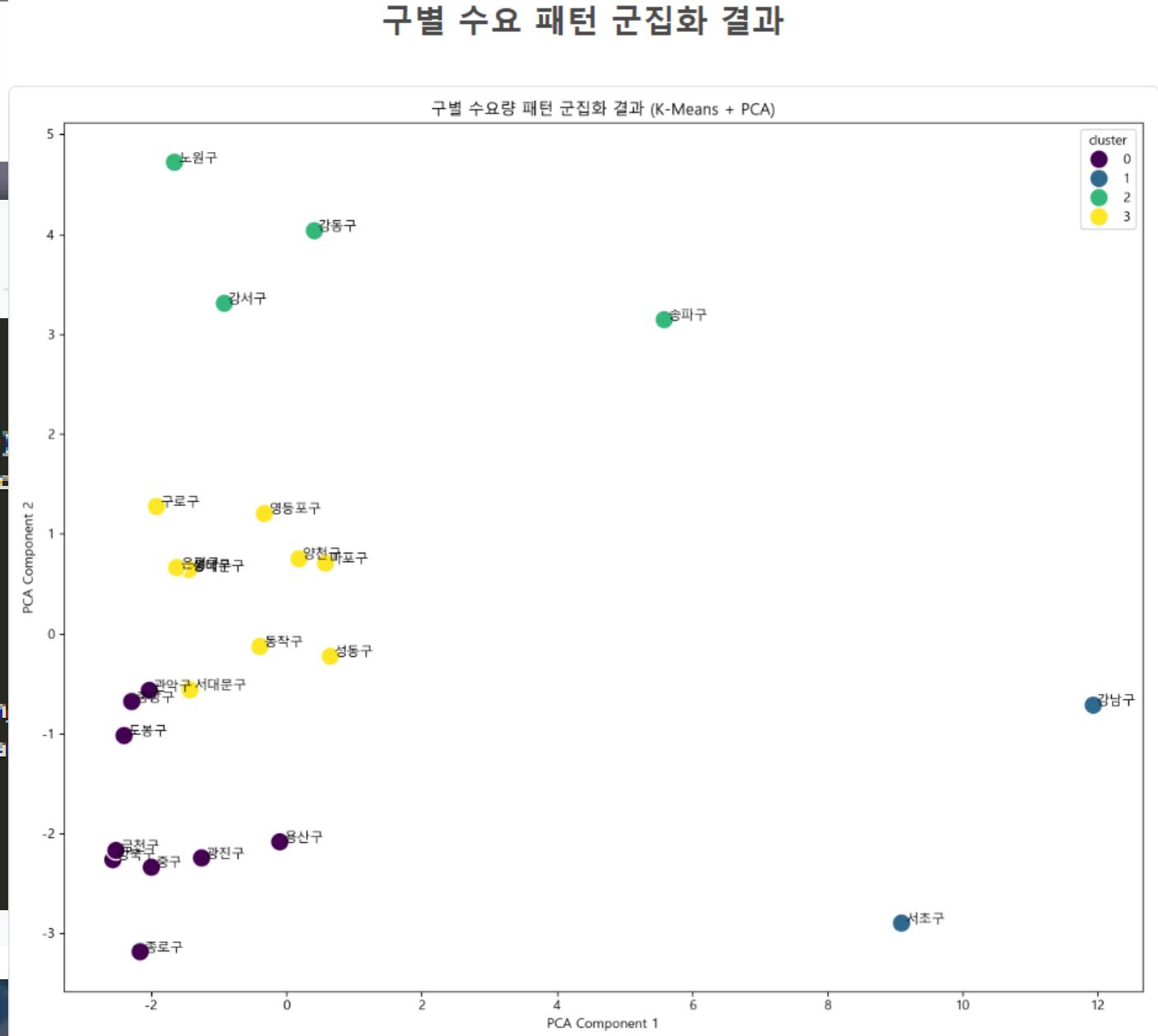
- ① 지역구 금액대별 수요량
- ② 면적 가격대별 수요량



# 3-1 지역구별 수요량 군집화

## 구별 수요 패턴 군집화 (K-Means + PCA)

```
# 구별 수요 패턴 데이터 준비  
df['계약년월'] = pd.to_datetime(df['계약년월'].astype(str))  
demand_by_gu = df.groupby(['구', '계약대']).size().unstack()  
  
# K-Means 및 PCA를 사용한 군집화  
scaler = StandardScaler()  
scaled_demand = scaler.fit_transform(demand_by_gu)  
optimal_k = 4  
kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)  
demand_by_gu['cluster'] = kmeans.fit_predict(scaled_demand)  
pca = PCA(n_components=2)  
reduced_data = pca.fit_transform(scaled_demand)
```



### 3-1-2 지역구 전, 월세 금액별 수요량 예측 및 결과

모델 : Prophet, linear\_regression, random\_forest

```
# 랜덤 포레스트 회귀 모델 예측
```

```
try:
```

```
    gu_data['time_idx'] = (gu_data['ds'] - gu_data['ds'].min()).dt.days
```

```
    X_train = gu_data[['time_idx']].iloc[:-3]
```

```
    y_train = gu_data['y'].iloc[:-3]
```

```
    X_test = gu_data[['time_idx']].iloc[-3:]
```

```
model_rf = RandomForestRegressor(n_estimators=100, random_state=42)
```

```
model_rf.fit(X_train, y_train)
```

```
y_pred = model_rf.predict(X_test)
```

```
y_true = gu_data['y'].iloc[-3:].values
```

```
mae = mean_absolute_error(y_true, y_pred)
```

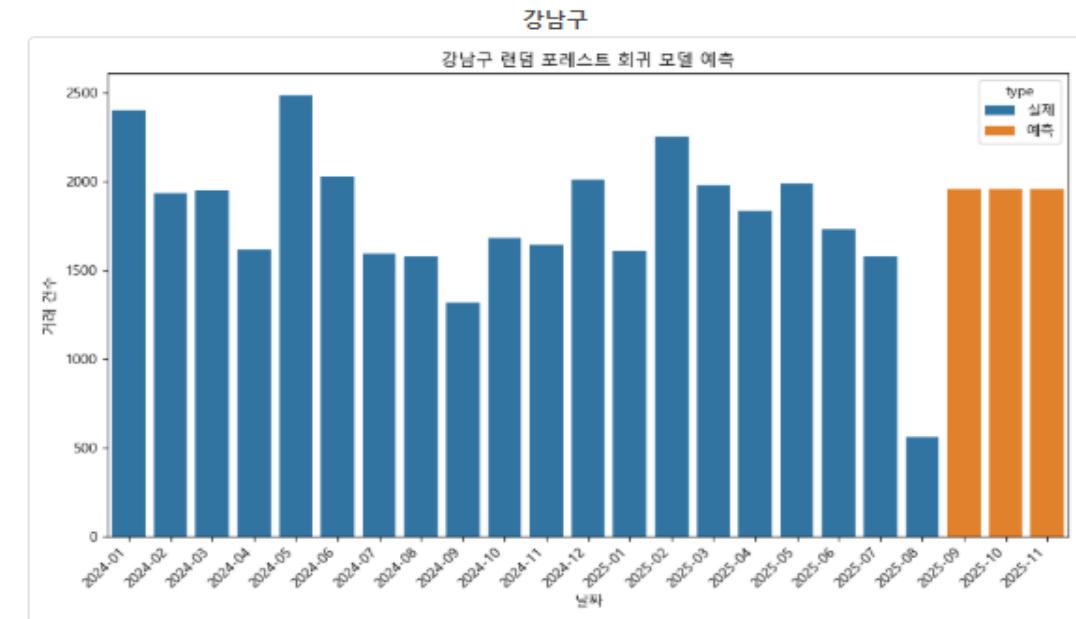
```
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
```

```
mape = np.mean(np.abs((y_true - y_pred) / y_true)) * 100 if np.all(y_true != 0) else float('inf')
```

```
random_forest_results['accuracy'][gu] = {'MAE': mae, 'RMSE': rmse, 'MAPE': mape}
```

Prophet linear regression random forest

강남구 강동구 강북구 강서구 관악구 광진구 구로구 금천구 노원구  
도봉구 동대문구 동작구 마포구 서대문구 서초구 성동구 성북구 송파구  
양천구 영등포구 용산구 은평구 종로구 종구 종량구



MAE (평균 절대 오차) 669.25

RMSE (평균 제곱근 오차) 847.00

MAPE (%) (평균 절대 백분율 오차) 95.68%

### 3-2-1 면적 금액별 전 모델 : linear\_

```
# 전/월세 및 면적 구간별로 반복
for rent_type in ['전세', '월세']:
    print(f"\n--- {rent_type} ---")
    for area_label in labels:
        subset_df = monthly_demand[(monthly_demand['전월세구분'] == rent_type) & (monthly_demand['면적구간'] == area_label)]
        if len(subset_df) < 2:
            print(f"> {area_label} : 데이터 부족으로 예측 불가")
            continue
        X = subset_df[['month_index']]
        y = subset_df['수요량']

        # GradientBoostingRegressor 모델 학습
        model = GradientBoostingRegressor(n_estimators=100)
        model.fit(X, y)

        # 2025년 10월 예측
        predicted_demand = model.predict([[predict_date]])

        # 예측 성능 평가
        y_pred = model.predict(X)
        rmse = np.sqrt(mean_squared_error(y, y_pred))
        r2 = r2_score(y, y_pred)
        mae = mean_absolute_error(y, y_pred)

        # 특성 중요도
        feature_importance = model.feature_importances_
```



선택항목

Decision Tree Regression

조회

## Ai-X APT 부동산 전/월세 예측 서비스

면적: 60~85m<sup>2</sup>

전세

월세

구분	값
훈련성능	RMSE=0.00, R <sup>2</sup> =1.00, MAE=0.00
테스트 성능	RMSE=2411.42, R <sup>2</sup> =-2.19, MAE=1997.40
과적합 여부	과적합 의심

구분	값
훈련성능	RMSE=0.00, R <sup>2</sup> =1.00, MAE=0.00
테스트 성능	RMSE=928.90, R <sup>2</sup> =-0.78, MAE=615.60
과적합 여부	과적합 의심

년-월	실제값	예측값
2024-01	5,851	5,851
2024-02	5,092	5,092
2024-03	5,680	5,680
2024-04	4,741	4,741
2024-05	5,026	5,026
2024-06	4,692	4,692
2024-07	4,966	4,966
2024-08	5,032	5,032
2024-09	3,875	3,875
2024-10	4,742	4,742
2024-11	4,700	4,700
2024-12	5,270	5,270
2025-01	5,159	5,159
2025-02	6,063	6,063
2025-03	6,462	6,462
2025-04	5,333	6,462
2025-05	5,259	6,462
2025-06	5,279	6,462
2025-07	4,642	6,462
2025-08	1,810	6,462
2025-09	0	6,462
2025-10	0	6,462

년-월	실제값	예측값
2024-01	2,667	2,667
2024-02	2,242	2,242
2024-03	2,379	2,379
2024-04	2,094	2,094
2024-05	2,264	2,264
2024-06	2,184	2,184
2024-07	2,316	2,316
2024-08	2,288	2,288
2024-09	1,941	1,941
2024-10	2,436	2,436
2024-11	2,404	2,404
2024-12	2,745	2,745
2025-01	2,407	2,407
2025-02	2,885	2,885
2025-03	2,700	2,700
2025-04	2,554	2,700
2025-05	2,480	2,700
2025-06	2,467	2,700
2025-07	2,208	2,700
2025-08	713	2,700
2025-09	0	2,700
2025-10	0	2,700



선택항목

Gradient Boosting Regression

조회

## Ai-X APT 부동산 전/월세 예측 서비스

면적: 85~100m<sup>2</sup>

전세

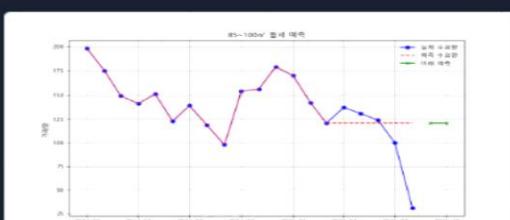
월세

구분	값
훈련성능	RMSE=0.47, R <sup>2</sup> =1.00, MAE=0.38
테스트 성능	RMSE=79.13, R <sup>2</sup> =-0.37, MAE=47.25
과적합 여부	과적합 의심

구분	값
훈련성능	RMSE=0.22, R <sup>2</sup> =1.00, MAE=0.18
테스트 성능	RMSE=42.22, R <sup>2</sup> =-0.18, MAE=27.99
과적합 여부	과적합 의심

년-월	실제값	예측값
2024-01	304	303
2024-02	263	262
2024-03	250	249
2024-04	236	236
2024-05	247	246
2024-06	230	230
2024-07	233	233
2024-08	246	245
2024-09	146	146
2024-10	239	238
2024-11	256	255
2024-12	236	236
2025-01	225	225
2025-02	237	236
2025-03	261	260
2025-04	276	260
2025-05	246	260
2025-06	259	260
2025-07	229	260
2025-08	88	260
2025-09	0	260
2025-10	0	260

년-월	실제값	예측값
2024-01	198	197
2024-02	175	174
2024-03	149	148
2024-04	141	141
2024-05	151	150
2024-06	123	123
2024-07	139	138
2024-08	119	119
2024-09	98	98
2024-10	154	153
2024-11	156	156
2024-12	179	178
2025-01	170	169
2025-02	142	141
2025-03	121	121
2025-04	137	121
2025-05	131	121
2025-06	124	121
2025-07	100	121
2025-08	31	121
2025-09	0	121
2025-10	0	121





04

## 예측 모델 웹 플랫폼



## 4 예측 모델 웹 플랫폼

Ai-X APT 부동산 전/월세별 수요량 예측 서비스

지역구  
금액대별  
거래량

면적  
가격대별  
거래량

05

## B2B 활용 방안

# ▶ 데이터 예측으로 B2B 활용 방안

01

## 시장에 맞는 물량 확보

수요 분석을 통해 적절한 재고, 물량 확보 가능,  
맞춤형 공급 전략 제공 [고객에게 맞춤형 공급 전략 제공]

02

## 지역 맞춤형 수요 예측

지역별 수요 기반으로 [인력, 광고비 배분 최적화] 가능

03

## 가격 및 거래 전략 최적화

데이터를 기반으로 [가격 조정, 재고량 관리] 가능

04

## 데이터 기반 의사 결정 지원

예측 모델을 시각화해 쉽게 인사이트 도출 가능,  
[B2B 운영 효율 극대화]



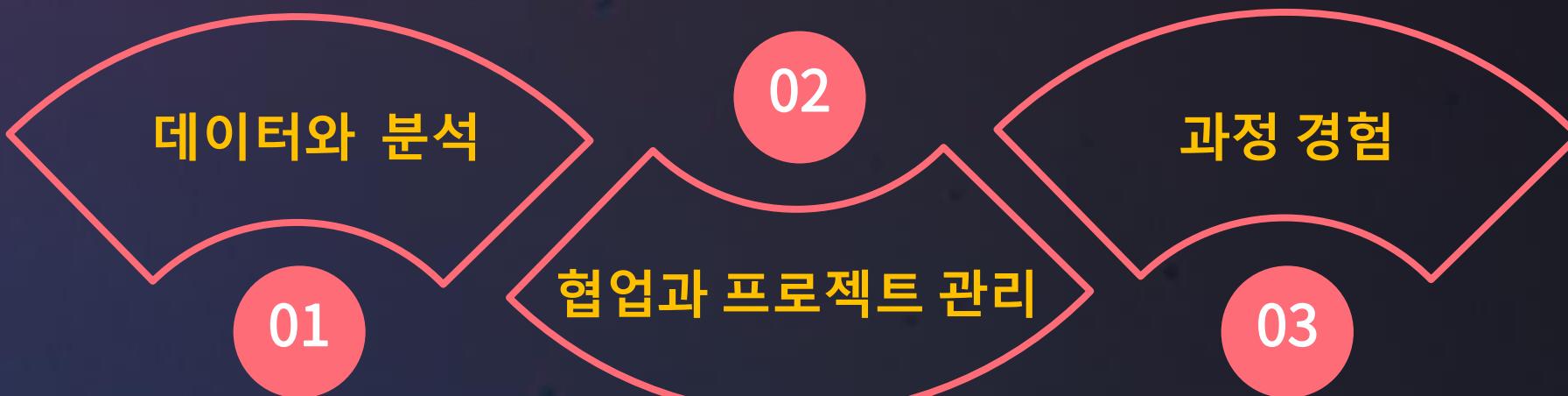
06

## 자체 평가 의견

# 자체 평가 의견

1. 프로젝트 준비기간이 짧아 데이터 양 대비 성능의 제한으로 제약을 받음
2. 경제,정책,계절 등의 다양한 변수를 충분히 반영하지 못함

- 실제 데이터 분석 과정에서 발생할 수 있는 제약과 개선 필요성을 직접 느낄 수 있었음.
- 향후 업데이트 관리까지 연결하지 못한 아쉬움
- 한정된 기간과 자료 안에서도 데이터 수집, 머신러닝, Django 웹 구현, PPT제작까지 전과정을 경험함



- ① Git을 활용해 팀원 간 소스를 공유하여 원활한 협업 진행
- ② Webex의 사용으로 프로젝트 진행과 소통이 효율적인 방향으로 흘러 팀워크를 체감함



대영



경주



창도



재준



세희

17 진행 상황

◀ [작업 관리 DB]

&lt; ...

Aa 작업명	☰ 담당자	▣ 마감일	✿ 상태	☰ 태그	+ ...
데이터 수집 & 전처리	재준 세희 대영 경주 창도	2025년 8월 20일	● 완료	데이터/EDA	
EDA 시각화	재준 세희 대영 경주 창도	2025년 8월 21일	● 완료	데이터/EDA	
모델 학습 (1차)	재준 세희 대영 경주	2025년 8월 22일	● 완료	모델링	
모델 성능 비교	재준 세희 대영 경주	2025년 8월 25일	● 완료	모델링	
Django 프로젝트 세팅	창도	2025년 8월 26일	● 완료	Django	
모델 API 연동 (Django)	재준 경주	2025년 8월 27일	● 진행 중	Django	
발표자료 제작	세희	2025년 8월 28일	● 진행 중	발표자료	
발표 준비	재준 세희 대영 경주 창도	2025년 8월 28일	● 진행 중	발표준비	
D-DAY	재준 세희 대영 경주 창도	2025년 8월 29일	● 시작 전	레츠고 발표	

+ 새 페이지

감사합니다

