Convenience Insights

Multicampus Data Analysis & Engineer 34th Course
2024-03-07 ~ 2024-03-29

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

1. INTRODUCTION

TEAM MEMBERS

HYERIN CHOI

- TEAM LEADER
- DATA PREPROCESSING
- EDA
- MAP VISUALIZATION
- STREAMLIT IMPLEMENTATION

JINA KIM

- DATA PREPROCESSING
- EDA
- STATISTICAL ANALYSIS
- MACHINE LEARNING

MIN SONG

- EDA
- STATISTICAL ANALYSIS
- MACHINE LEARNING

JUNHO SONG

- EDA
- STATISTICAL ANALYSIS
- MACHINE LEARNING
- MAP VISUALIZATION
- STREAMLIT IMPLEMENTATION

DAEHEE HAN

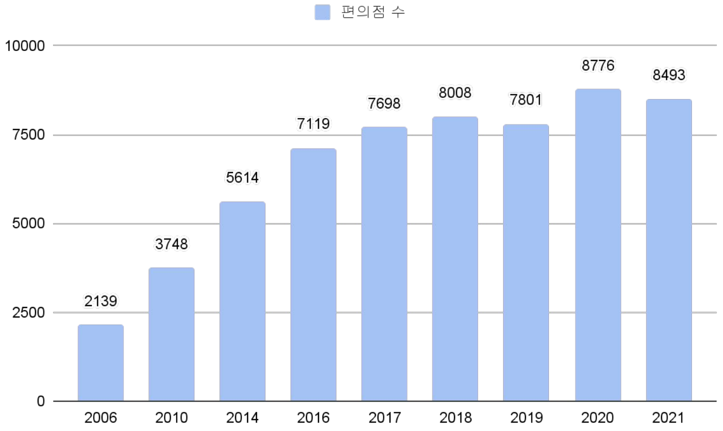
- STATISTICAL ANALYSIS
- MACHINE LEARNING

1. INTRODUCTION

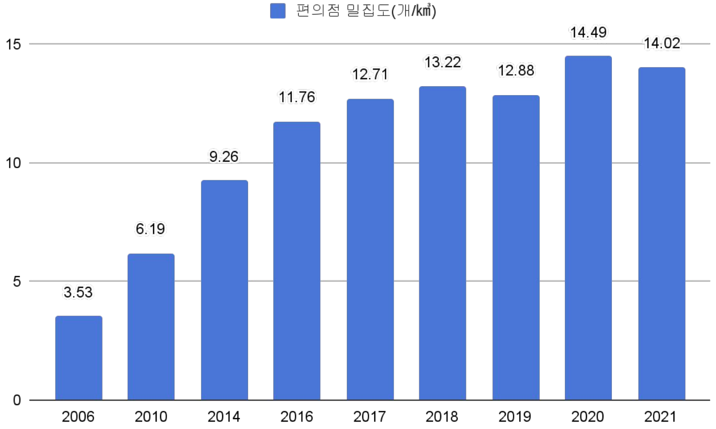
BACKGROUND

- As of the end of 2021, the number of convenience stores in Seoul totaled **8,493**, which is approximately **four times** the number compared to **2,139** in 2006.
- The convenience store density, measured by the number of stores per ㎢, also increased from 3.5 stores in 2006 to 14 stores in 2021.

| Convenience Store Count in Seoul |



| Convenience Store Density in Seoul |



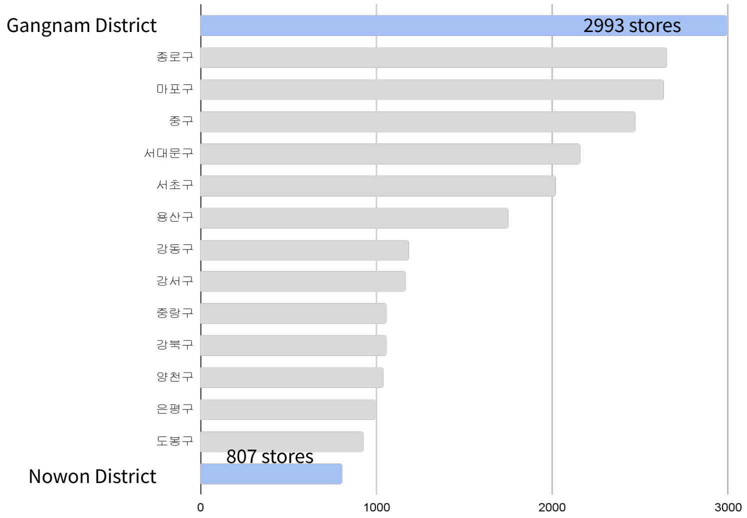
Source: Seoul Metropolitan Government, Operation Status and Current Situation Analysis Data of Convenience Stores in Seoul

1. INTRODUCTION

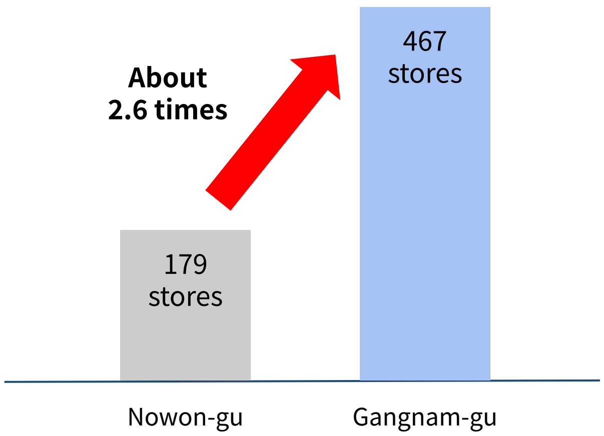
BACKGROUND

- **Gangnam District** - 467 convenience stores, **Nowon District** - 179 convenience stores indicating a difference of approximately **2.6 times**.
- Suggesting that if one were to open a convenience store in Gangnam-gu, the competition would be more intense. Also implying the necessity of conducting a market analysis.

| Number of Convenience Facilities by District |



| Number of convenience stores in Nowon-gu and Gangnam-gu |



Source: Cookie News, With 1,950 Cafés and 470 Convenience Stores, Gangnam-gu Emerges as the Largest "Premium Location" in Seoul

1. INTRODUCTION

BACKGROUND

- When opening a convenience store franchise, there exists a **law that requires the franchisor to inform the franchisee about the expected sales figures**. However, **discrepancies between the projected and actual sales figures** can occur after the establishment, leading to difficulties for the franchisees.

| The Problems in Opening Convenience Stores |

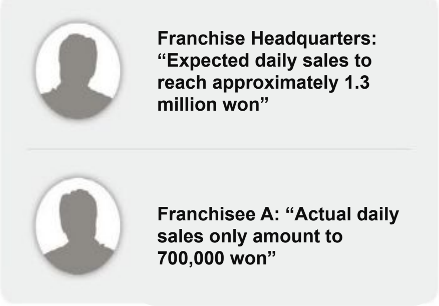
홈 > COVER STORY > 편의점의 역설

If the "projected sales figures" provided by the convenience store headquarters are exaggerated

HOME > 사회 > 사건/사고

Homeplus Inflating Projected Sales Figures for Prospective Convenience Store Owners

[Convenience Store Dispute Cases]



Franchise Headquarters: “Expected daily sales to reach approximately 1.3 million won”

Franchisee A: “Actual daily sales only amount to 700,000 won”

[Planned Closure Turns Complicated...]

[Source | Gyeonggi Fair Trade Support Center]

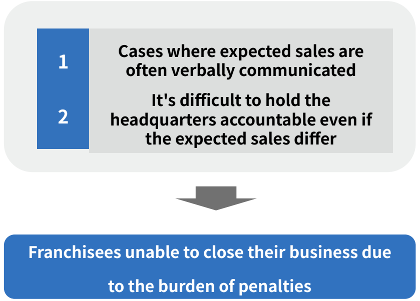
Facing a penalty of 81 million won for contract termination

Facility and interior design costs borne by the headquarters

Franchisee, application for dispute resolution

Penalty reduced by 12 million won, Must pay 69 million won

[Hidden Issues]



1 Cases where expected sales are often verbally communicated

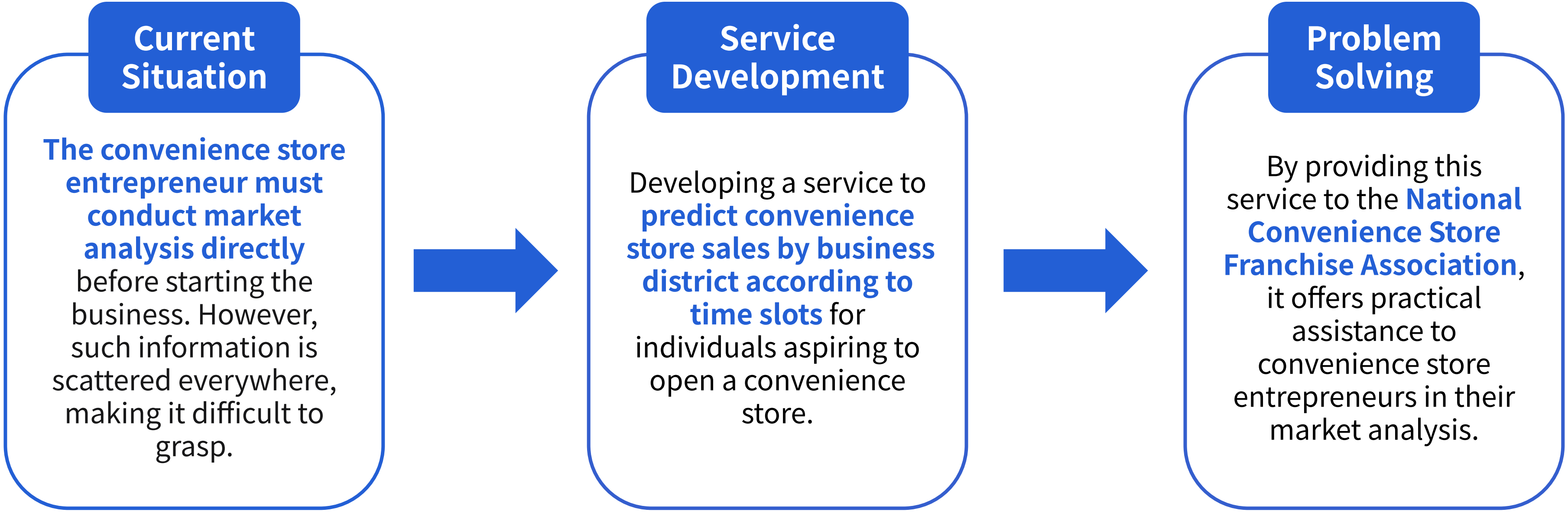
2 It's difficult to hold the headquarters accountable even if the expected sales differ

Franchisees unable to close their business due to the burden of penalties

Source: The Scoop, If the Projected Sales Figures Provided by the Convenience Store Headquarters Are Exaggerated…, Citizen Daily, Homeplus Inflating Projected Sales Figures for Prospective Convenience Store Owners

1. INTRODUCTION

BACKGROUND



Current Situation

The convenience store entrepreneur must conduct market analysis directly before starting the business. However, such information is scattered everywhere, making it difficult to grasp.

Service Development

Developing a service to **predict convenience store sales by business district according to time slots** for individuals aspiring to open a convenience store.

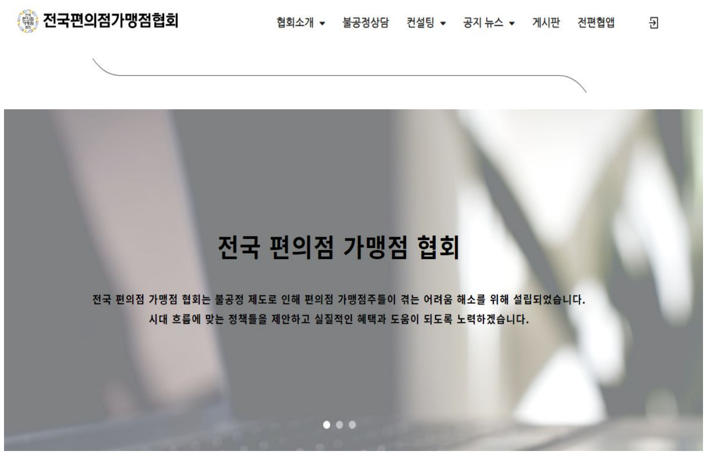
Problem Solving

By providing this service to the **National Convenience Store Franchise Association**, it offers practical assistance to convenience store entrepreneurs in their market analysis.

1. INTRODUCTION

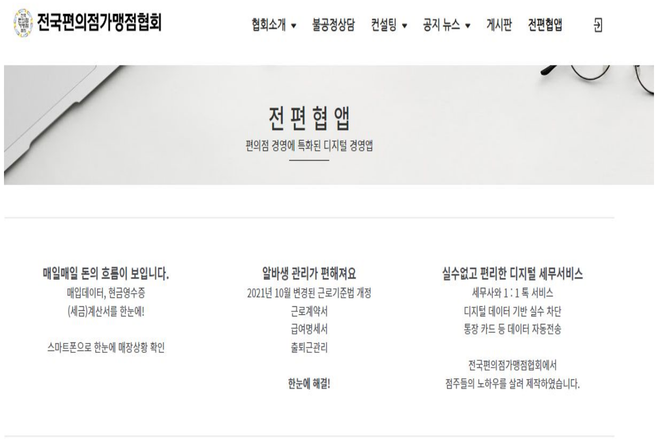
BACKGROUND

- After accessing the website of the National Convenience Store Franchise Association, it has been confirmed that services such as management reporting, labor management, and tax information exist. However, there is no **market analysis service specifically designed for convenience store entrepreneurs**.



전국 편의점 가맹점 협회

전국 편의점 가맹점 협회는 불공정 제도로 인해 편의점 가맹점주들이 겪는 어려움 해소를 위해 설립되었습니다.
시대 흐름에 맞는 정책들을 제안하고 실질적인 혜택과 도움이 되도록 노력하겠습니다.



전편협앱

편의점 경영에 특화된 디지털 경영앱

매일매일 돈의 흐름이 보입니다.
매입데이터, 현금영수증
(세금)계산서를 한눈에!

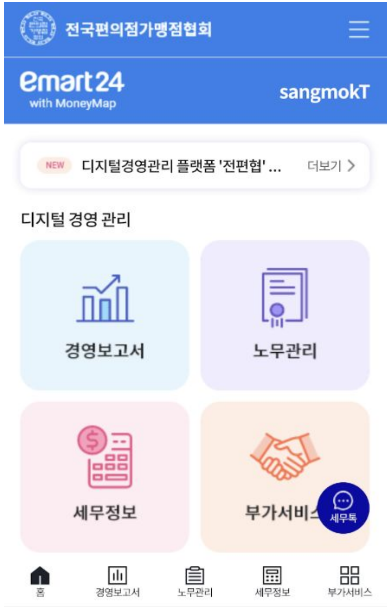
스마트폰으로 한눈에 매장상황 확인

알바생 관리가 편해져요
2021년 10월 변경된 근로기준법 개정
근로계약서
급여명세서
출퇴근관리

한눈에 해결!

실수없고 편리한 디지털 세무서비스
세무사와 1 : 1 톡 서비스
디지털 데이터 기반 실수 차단
통장 카드 등 데이터 자동전송

전국편의점가맹점협회에서
점주들의 노하우를 살려 제작하였습니다.

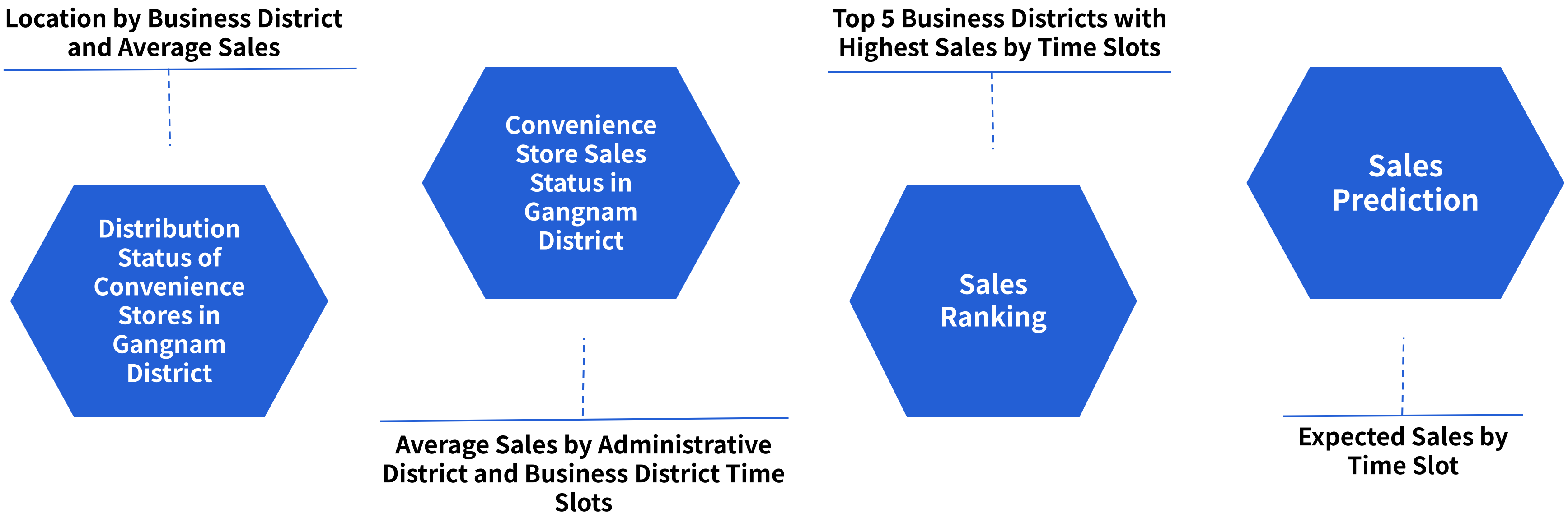


NEW 디지털경영관리 플랫폼 '전편협' ... 더보기 >

디지털 경영 관리

1. INTRODUCTION

WEB SERVICE



Location by Business District and Average Sales

Distribution Status of Convenience Stores in Gangnam District

Convenience Store Sales Status in Gangnam District

Average Sales by Administrative District and Business District Time Slots

Top 5 Business Districts with Highest Sales by Time Slots

Sales Ranking

Sales Prediction

Expected Sales by Time Slot

1. INTRODUCTION

DEVELOPMENT ENVIRONMENT

Development Environment

Visual Studio Code

jupyterlab

Deployment

GitHub

Streamlit

Language

python

Open API

서울 열린데이터 광장

1. INTRODUCTION

DEVELOPMENT ENVIRONMENT

Tool

- Visual Studio
- Jupyter lab

Deployment

- Github
- Streamlit

Language

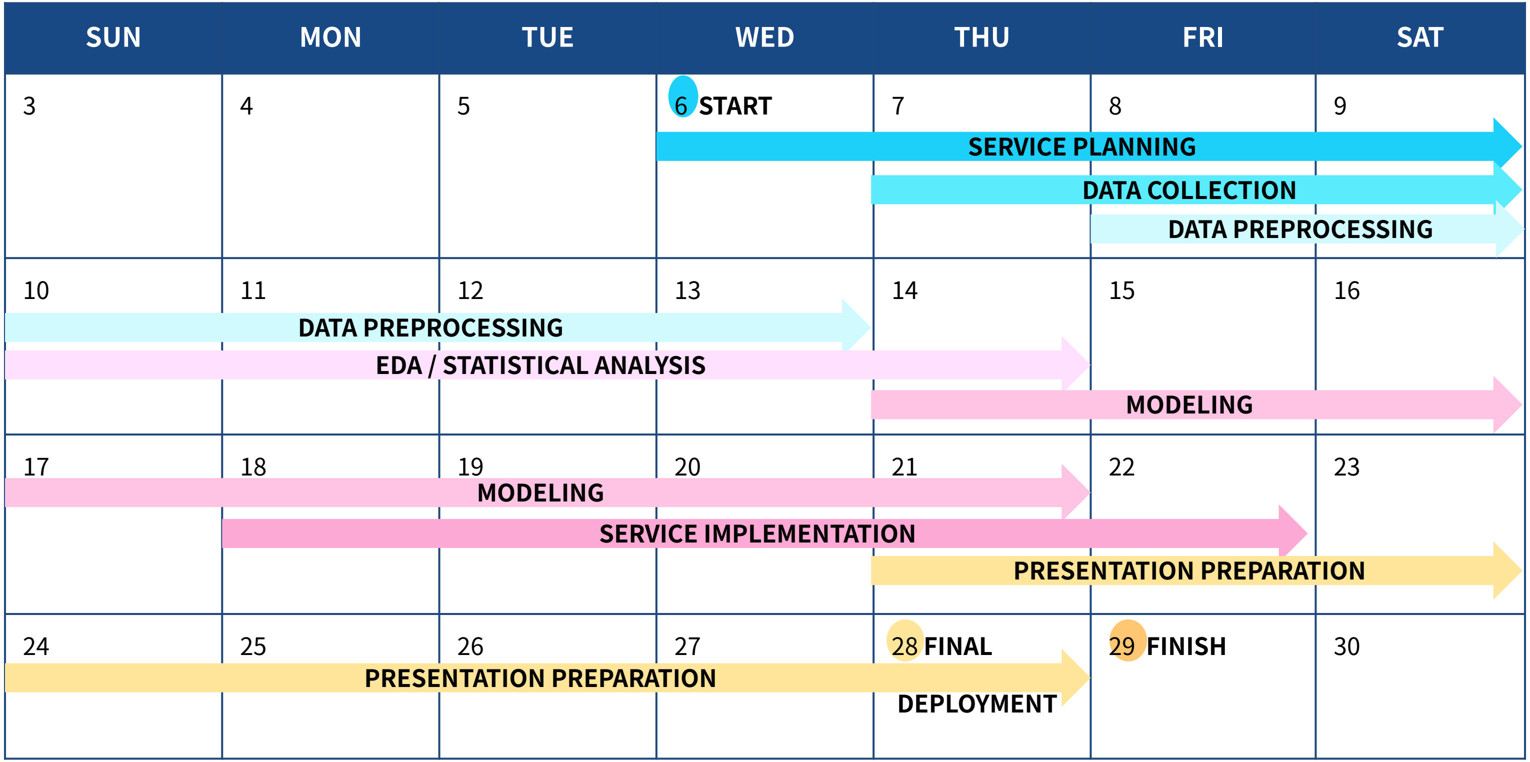
- Python

Library

- pandas
- numpy
- plotly
- matplotlib
- seaborn
- streamlt
- streamlit_option_menu
- streamlit_folium
- folium
- geopandas
- os
- time
- sklearn
- xgboost
- statsmodel
- scipy
- lgbm
- re
- jolib

1. INTRODUCTION

PROJECT PERIOD



SUN	MON	TUE	WED	THU	FRI	SAT
3	4	5	6 **START**	7 **SERVICE PLANNING**	8 **DATA COLLECTION** **DATA PREPROCESSING**	9
10	11 **DATA PREPROCESSING** **EDA / STATISTICAL ANALYSIS**	12	13	14	15 **MODELING**	16
17	18	19 **MODELING** **SERVICE IMPLEMENTATION**	20	21 **PRESENTATION PREPARATION**	22	23
24	25 **PRESENTATION PREPARATION**	26	27	28 **FINAL** **DEPLOYMENT**	29 **FINISH**	30

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

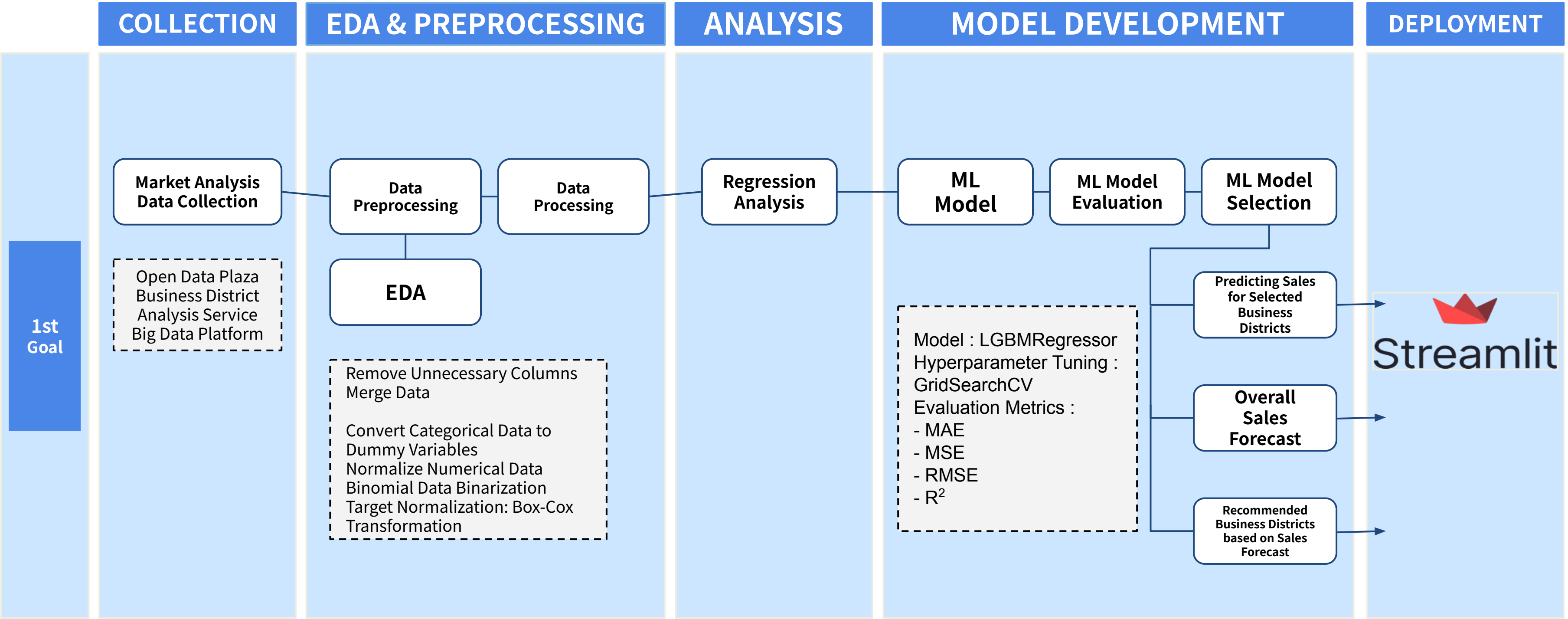
- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

2. DOCUMENTATION

FLOW CHART



COLLECTION

EDA & PREPROCESSING

ANALYSIS

MODEL DEVELOPMENT

DEPLOYMENT

1st Goal

Market Analysis Data Collection

Open Data Plaza Business District Analysis Service Big Data Platform

Data Preprocessing

Data Processing

EDA

Remove Unnecessary Columns
Merge Data

Convert Categorical Data to Dummy Variables
Normalize Numerical Data
Binomial Data Binarization
Target Normalization: Box-Cox Transformation

Regression Analysis

ML Model

ML Model Evaluation

ML Model Selection

Model : LGBMRegressor
Hyperparameter Tuning : GridSearchCV
Evaluation Metrics :
- MAE
- MSE
- RMSE
- R^2

Predicting Sales for Selected Business Districts

Overall Sales Forecast

Recommended Business Districts based on Sales Forecast

2. DOCUMENTATION

WBS

CATEGORY	MAIN TASK	1ST WEEK (3/06 ~ 3/09)			2ND WEEK (3/10 ~ 3/16)			3RD WEEK (3/17 ~ 3/23)			4TH WEEK (3/24 ~ 3/29)		
PLANNING	TOPIC SELECTION	■	■	■									
DATA COLLECTION	API DATA COLLECTION		■	■	■								
DATA PREPROCESSING	DATA FORMAT STANDARDIZATION			■	■	■							
DATA CHECK	EDA				■	■	■						
VISUALIZATION	SPACE / MAP VISUALIZATION					■	■	■	■				
STATISTICAL ANALYSIS	REGRESSION ANALYSIS				■	■	■	■					
MODELING	MODEL GENERATION / EVALUATION					■	■	■	■	■			
SERVICE IMPLEMENTATION	STREAMLIT DEPLOYMENT					■	■	■	■	■	■	■	
FINAL	FINAL PRESENTATION												■

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA EXPLORATION

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

3. DATA EXPLORATION

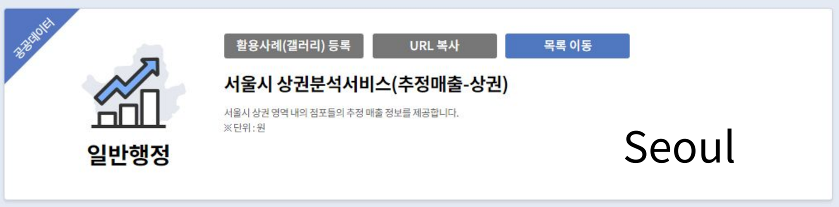
DATA COLLECTION

Major Category	Commercial Area Analysis Service Data							
Classification	Estimated Sales	Population by Road	Resident Population	Working Population	Income and Expenditure	Areas	Customer Attraction Facilities	Stores
Details	- Quarterly sales amount - Number of sales transactions per quarter - Sales amount by time slot - Number of sales transactions by time slot - Sales amount by age group - Number of sales transactions by age group - Sales amount by gender - Number of sales transactions by gender - Sales amount by day of the week - Number of sales transactions by day of the week	- Total floating population - Floating population by gender - Floating population by age - Floating population by time slot - Floating population by day of the week	- Total resident population - Resident population by gender - Resident population by age group - Total number of households - Number of apartment households - Number of non-apartment households	- Total working population - Working population by gender - Working population by age group	- Average monthly income - Income level - Total expenditure - Type of expenditure	- X coordinate - Y coordinate - District code - Neighborhood code - Area size	- Total number of customer attraction facilities - Number of customer attraction facilities by type	- Store opening/closing rate - Number of stores opened/closed - Total number of stores - Number of stores by similar business type - Number of franchise stores
Source	Seoul Open Data Plaza							

3. DATA EXPLORATION

DATA PREPROCESSING

1. Extract only convenience store data in Gangnam District

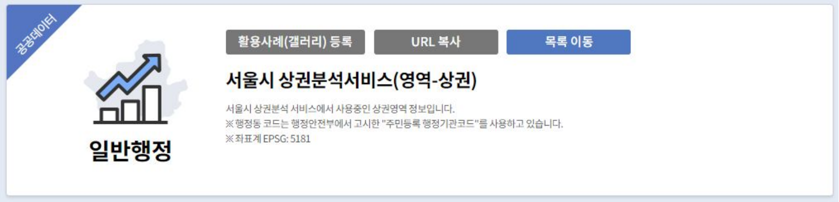


서울시 상권분석서비스(추정매출-상권)

서울시 상권 영역 내의 점포들의 추정 매출 정보를 제공합니다.
※단위 : 원

Seoul

Extract data where "Service_Industry_Code_Name" is convenience store



서울시 상권분석서비스(영역-상권)

서울시 상권분석 서비스에서 사용중인 상권영역 정보입니다.
※행정동 코드는 행정안전부에서 고시한 "주민등록 행정기관코드"를 사용하고 있습니다.
※좌표계 EPSG: 5181

Extract data where "District_Code_Name" is convenience store.

3. DATA EXPLORATION

DATA PREPROCESSING

2. Checking for Missing Values Before Data Merge

Category	Estimated Sales	Population by Road	Resident Population	Working Population	Income and Expenditure	Areas	Customer Attraction Facilities	Stores
Null Values	No Null Values	No Null Values	No Null Values	No Null Values	- Monthly Average Income Amount - Income Bracket Code - Total Expenditure Amount - Total Grocery Expenditure Amount - Total Clothing and Footwear Expenditure Amount - Total Household Goods Expenditure Amount - Total Medical Expenses Amount - Total Transportation Expenditure Amount - Total Leisure Expenditure Amount - Total Cultural Expenditure Amount - Total Education Expenditure Amount - Total Entertainment Expenditure Amount	No Null Values	- Number of Government Offices - Number of Banks - Number of General Hospitals - Number of Hospitals - Number of Pharmacies - Number of Kindergartens - Number of Elementary Schools - Number of High Schools - Number of Universities - Number of Department Stores - Number of Supermarkets - Number of Theaters - Number of Accommodation Facilities - Number of Airports - Number of Railway Stations - Number of Bus Terminals - Number of Subway Stations - Number of Bus Stops	No Null Values

3. DATA EXPLORATION

DATA PREPROCESSING

2. Checking for Missing Values Before Data Merge

Replace with the average value of a different quarter in the same commercial district

Category	Estimated Sales	Population by Road	Resident Population	Working Population	Income and Expenditure	Areas	Customer Attraction Facilities	Stores
Null Values	No Null Values	No Null Values	No Null Values	No Null Values	- Monthly Average Income Amount - Income Bracket Code - Total Expenditure Amount - Total Grocery Expenditure Amount - Total Clothing and Footwear Expenditure Amount - Total Household Goods Expenditure Amount - Total Medical Expenses Amount - Total Transportation Expenditure Amount - Total Leisure Expenditure Amount - Total Cultural Expenditure Amount - Total Education Expenditure Amount - Total Entertainment Expenditure Amount	No Null Values	- Number of Government Offices - Number of Banks - Number of General Hospitals - Number of Hospitals - Number of Pharmacies - Number of Kindergartens - Number of Elementary Schools - Number of High Schools - Number of Universities - Number of Department Stores - Number of Supermarkets - Number of Theaters - Number of Accommodation Facilities - Number of Airports - Number of Railway Stations - Number of Bus Terminals - Number of Subway Stations - Number of Bus Stops	No Null Values

3. DATA EXPLORATION

Replace Null Values with 0

DATA PREPROCESSING

2. Checking for Missing Values Before Data Merge

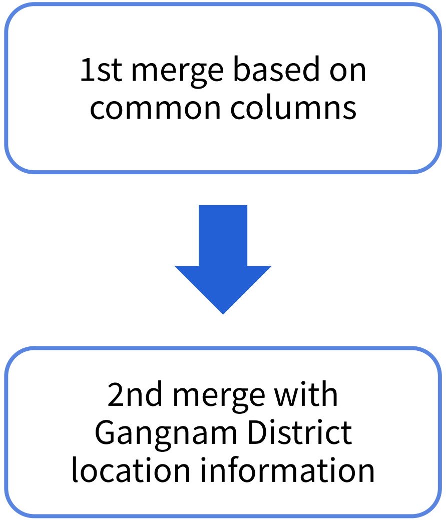
Category	Estimated Sales	Population by Road	Resident Population	Working Population	Income and Expenditure	Areas	Customer Attraction Facilities	Stores
Null Values	No Null Values	No Null Values	No Null Values	No Null Values	- Monthly Average Income Amount - Income Bracket Code - Total Expenditure Amount - Total Grocery Expenditure Amount - Total Clothing and Footwear Expenditure Amount - Total Household Goods Expenditure Amount - Total Medical Expenses Amount - Total Transportation Expenditure Amount - Total Leisure Expenditure Amount - Total Cultural Expenditure Amount - Total Education Expenditure Amount - Total Entertainment Expenditure Amount	No Null Values	- Number of Government Offices - Number of Banks - Number of General Hospitals - Number of Hospitals - Number of Pharmacies - Number of Kindergartens - Number of Elementary Schools - Number of High Schools - Number of Universities - Number of Department Stores - Number of Supermarkets - Number of Theaters - Number of Accommodation Facilities - Number of Airports - Number of Railway Stations - Number of Bus Terminals - Number of Subway Stations - Number of Bus Stops	No Null Values

3. DATA EXPLORATION

DATA PREPROCESSING

3. Data Merge

Category	Common Column
Estimated Sales	- Base Year and Quarter Code - Commercial District Division Code - Commercial District Division Code Name - Commercial District Code - Commercial District Code Name
Population by Road	
Resident Population	
Working Population	
Income and Expenditure	
Customer Attraction Facilities	
Stores	

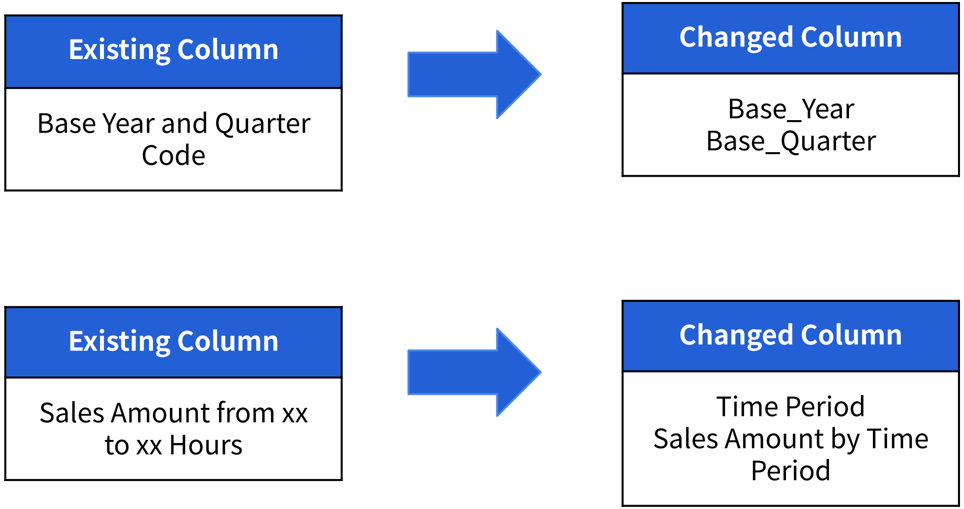

1st merge based on common columns

2nd merge with Gangnam District location information

3. DATA EXPLORATION

DATA PREPROCESSING

4. Changing the Columns



Existing Column

Base Year and Quarter Code

Changed Column

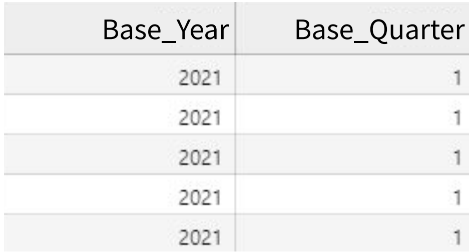
Base_Year
Base_Quarter

Existing Column

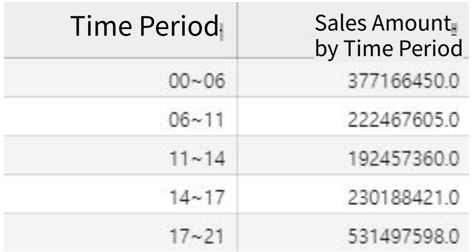
Sales Amount from xx to xx Hours

Changed Column

Time Period
Sales Amount by Time Period



Base_Year	Base_Quarter
2021	1
2021	1
2021	1
2021	1
2021	1

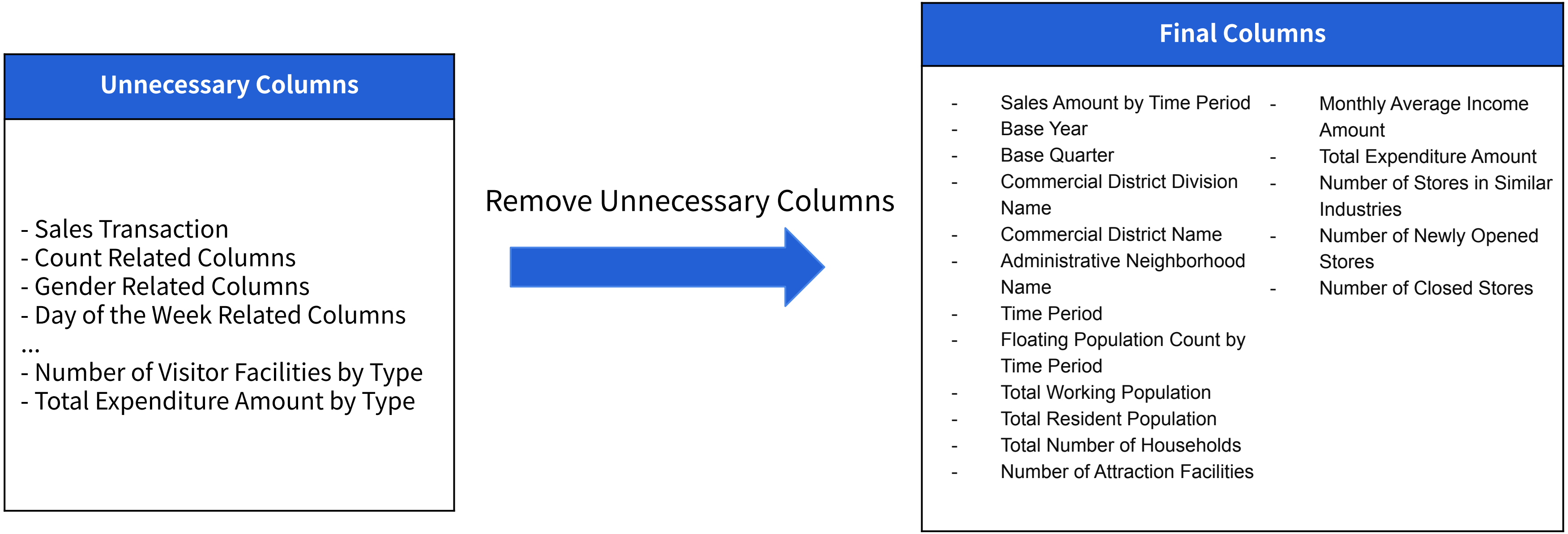


Time Period	Sales Amount by Time Period
00~06	377166450.0
06~11	222467605.0
11~14	192457360.0
14~17	230188421.0
17~21	531497598.0

3. DATA EXPLORATION

DATA PREPROCESSING

5. Removing Unnecessary Columns



Unnecessary Columns

- Sales Transaction
- Count Related Columns
- Gender Related Columns
- Day of the Week Related Columns

...

- Number of Visitor Facilities by Type
- Total Expenditure Amount by Type

Remove Unnecessary Columns

Final Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

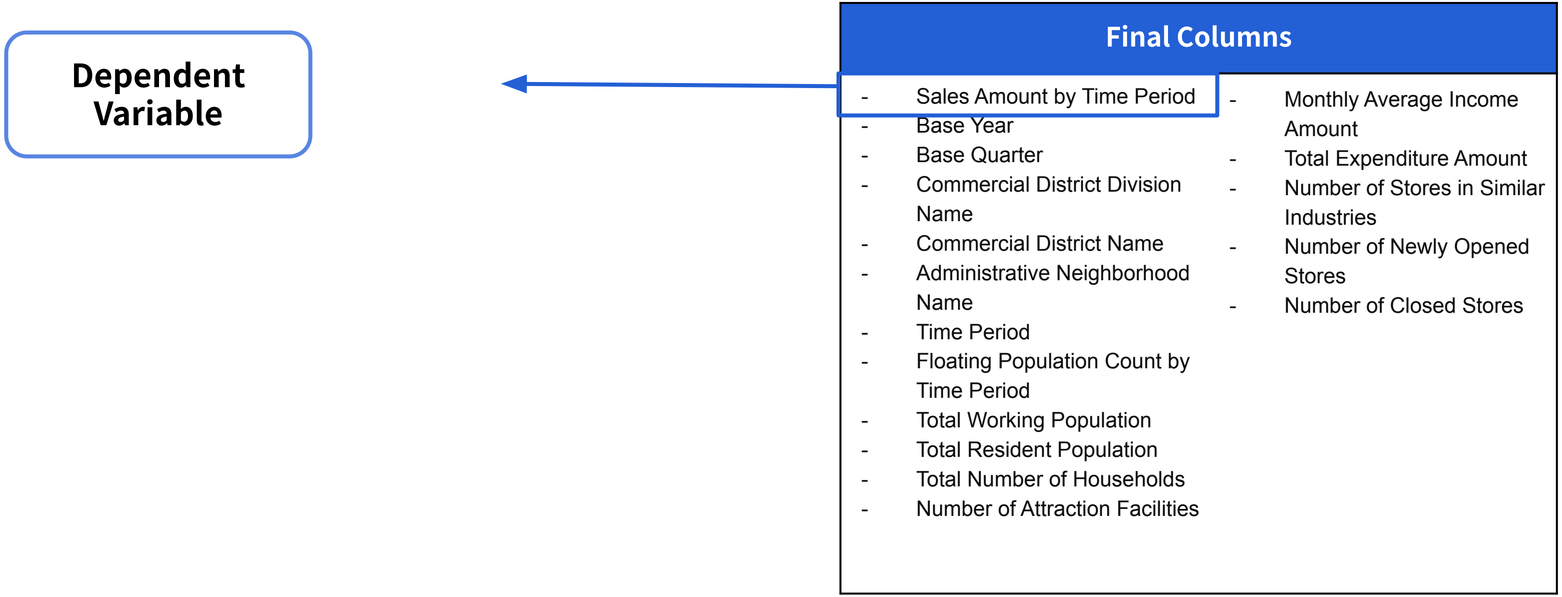
- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

4. EDA

FINAL COLUMNS CHECK



Dependent Variable

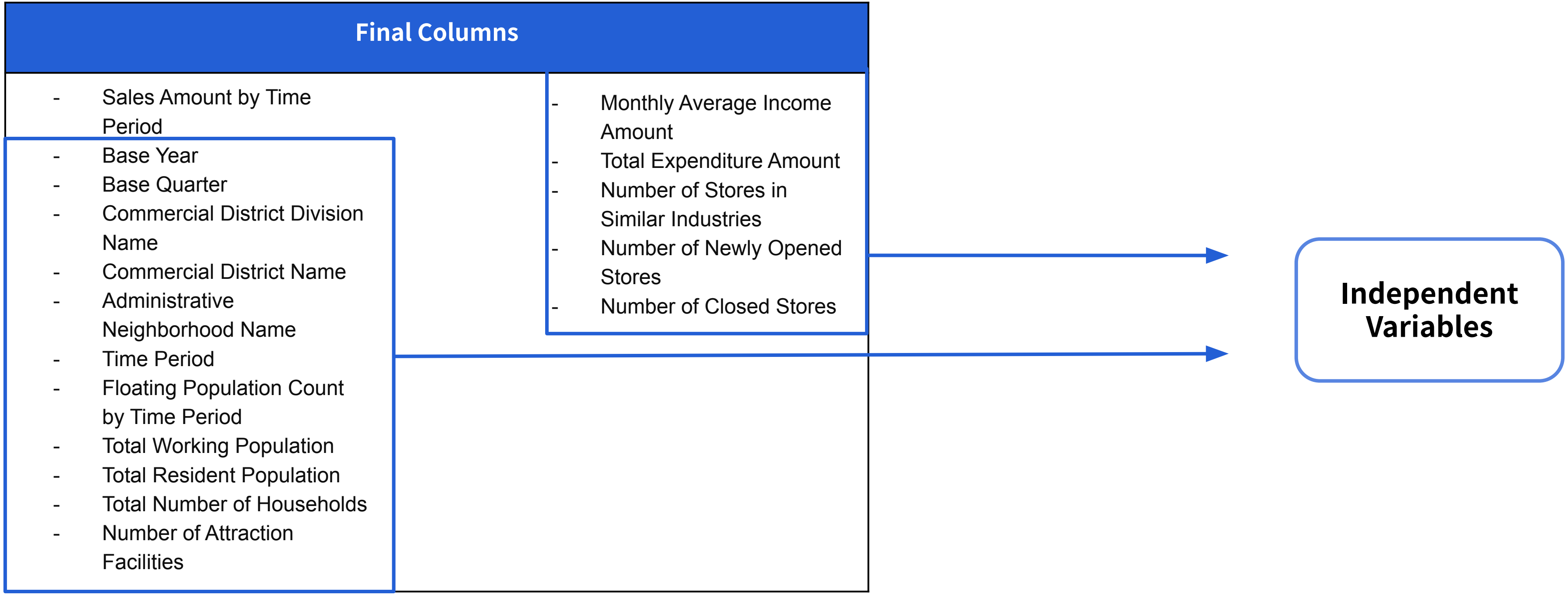
Final Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

The EDA for each variable can be found in the appendix.

4. EDA

FINAL COLUMNS CHECK



Final Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities

- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

Independent Variables

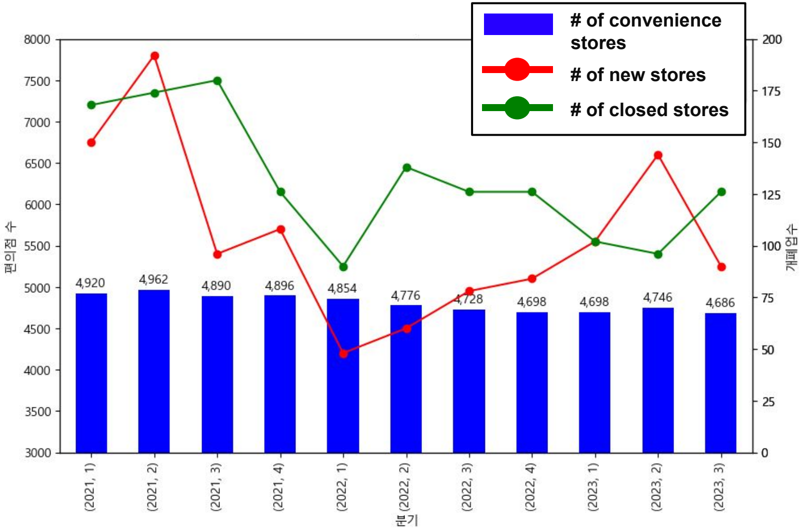
The EDA for each variable can be found in the appendix.

4. EDA

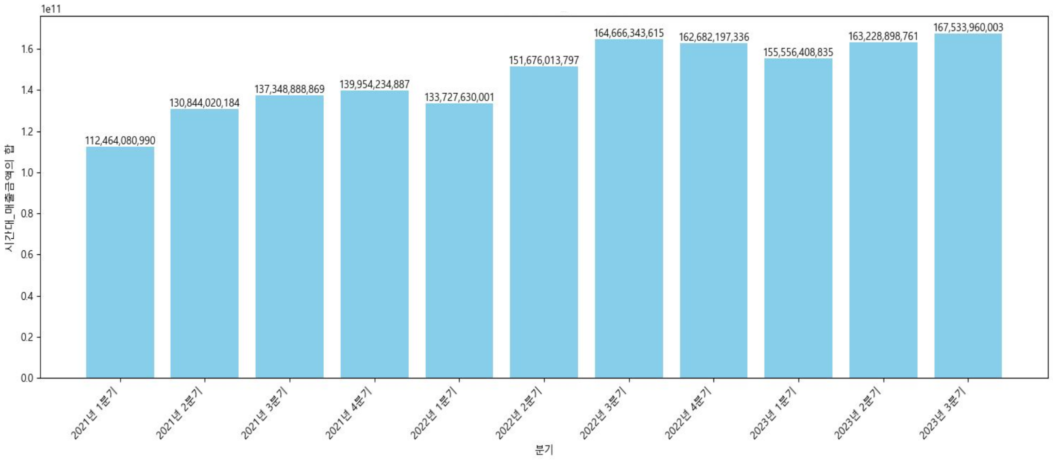
EDA

- In Gangnam District, the number of convenience stores closing is generally higher than the number of new openings.
- Although the number of convenience stores is slightly decreasing, it is recording levels that are almost the same.
- On the other hand, sales amounts are showing a trend of gradual increase.

| Quarterly Convenience Store Numbers and Openings/Closures |



| Year and Quarterly Sales Amount |

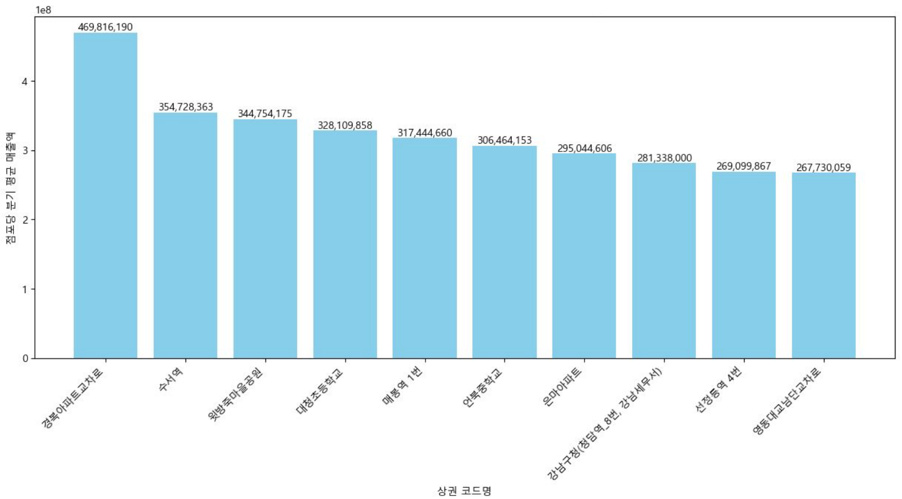


4. EDA

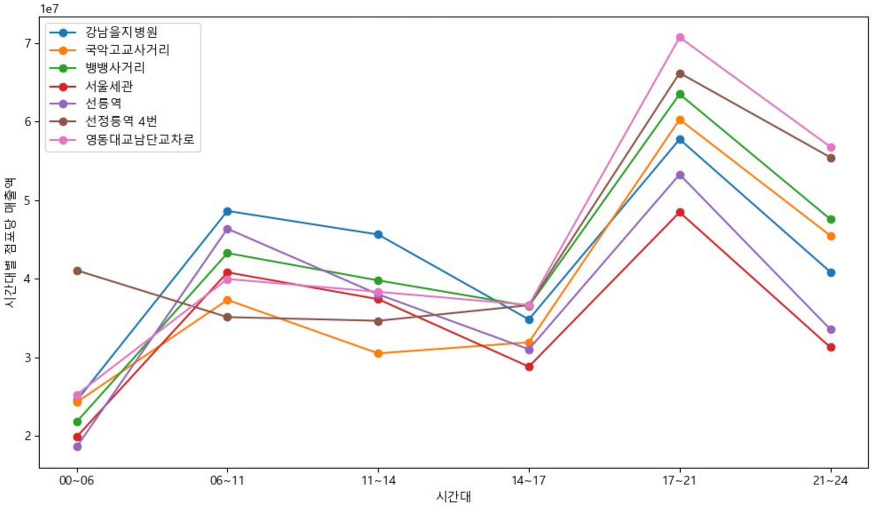
EDA

- Sales vary by commercial district.
- When extracting a random commercial district, the sales trends by time of day are similar, but each district shows slightly different sales trends.
- It can be expected that the factors affecting sales by time of day vary depending on the commercial district.

| Average Sales by Commercial District from 2021 to Q3 2023 |



| Average Sales per Store by Time of Day According to Commercial District |

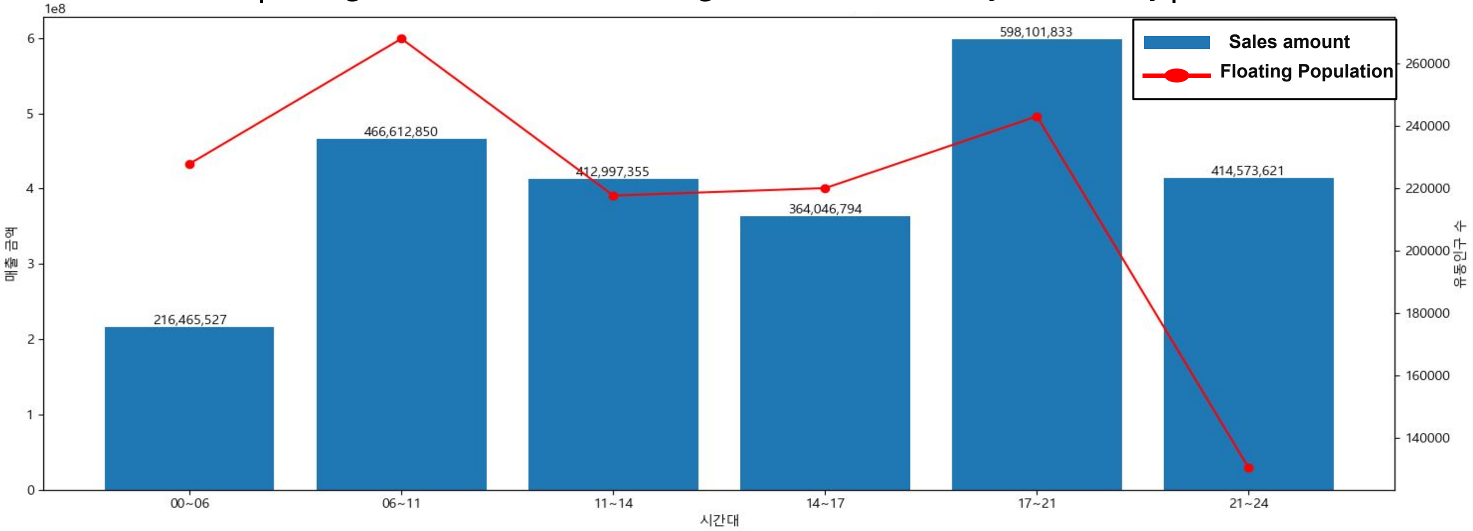


4. EDA

EDA

- High sales are recorded during commuting hours.
- Sales tend to increase when there is a high foot traffic, but despite a significant decrease in foot traffic from 21:00 to 24:00, sales during this time are the next highest after commuting hours.
- This suggests that, in addition to the number of pedestrians, other variables are also expected to influence sales by time of day.

| Average Sales Amount and Average Pedestrian Traffic by Time of Day |



4. EDA

ADDING COLUMNS

- Based on prior research, add the columns 'Convenience Store Density', 'Proportion of Resident Population by Age Group', and 'Proportion of Working Population by Age Group'.

Existing Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

Final Columns

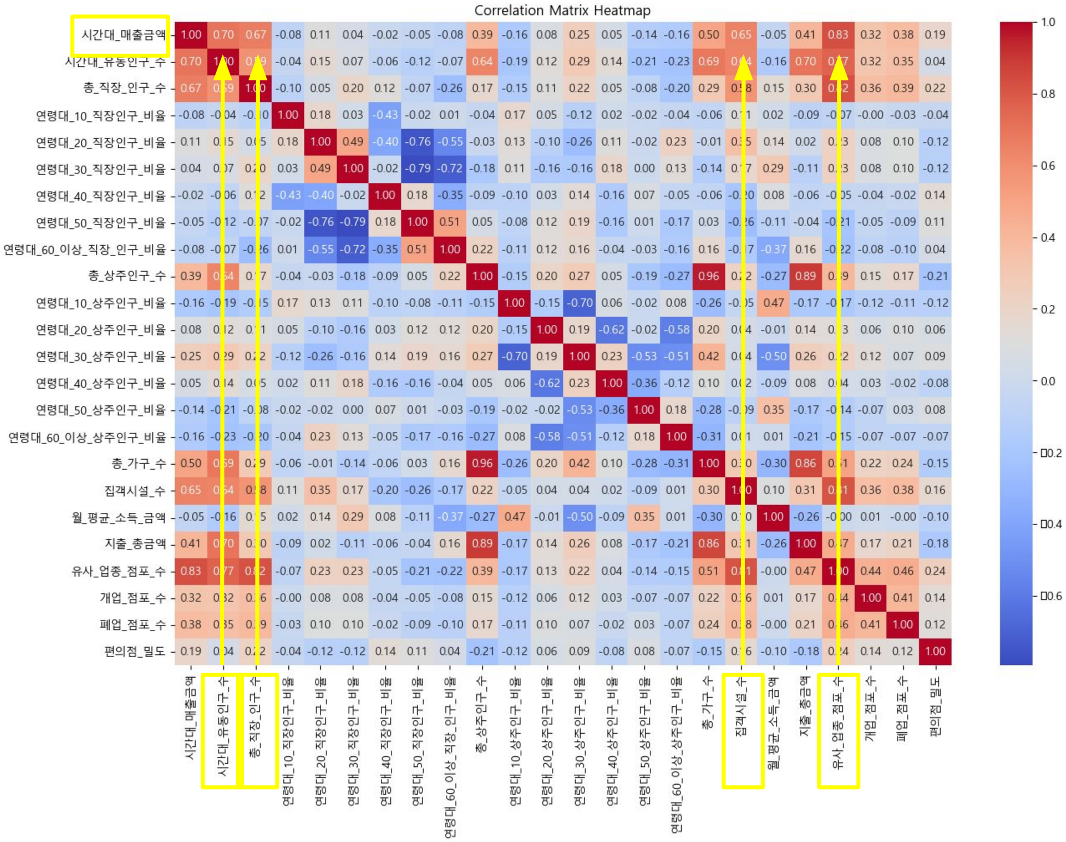
- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores
- Convenience Store Density
- Proportion of Resident Population by Age
- Proportion of Working Population by Age

4. EDA

CORRELATION ANALYSIS

- Number of Stores in Similar Industries - 0.83
- Floating Population Count by Time Period - 0.70
- Total Working Population - 0.67
- Number of Attraction Facilities - 0.65

=> are expected to have a relatively significant impact on "Sales Amount by Time Segment".


Correlation Matrix Heatmap

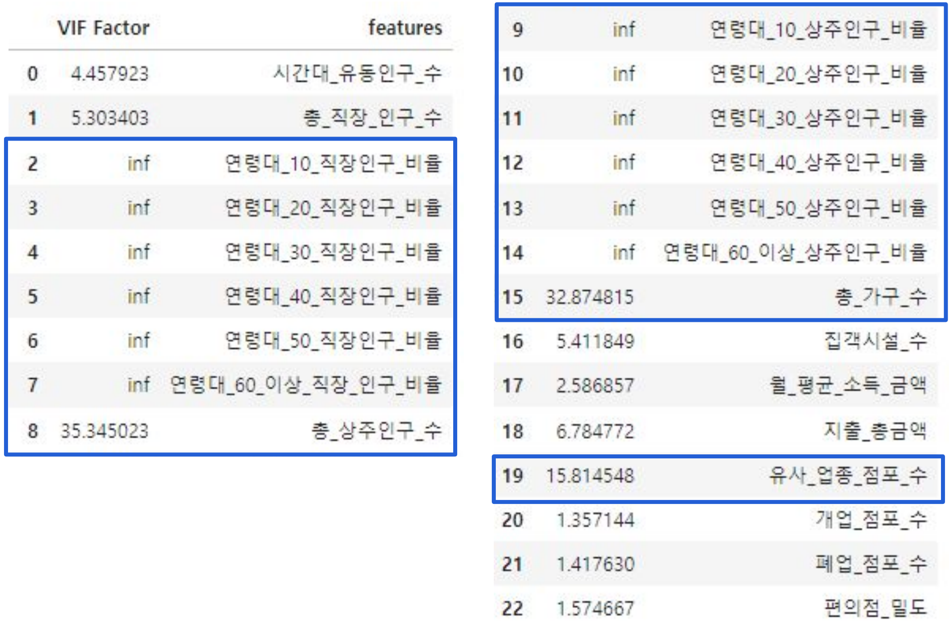
	시간대_매출금액	시간대_유동인구_수	총_직장_인구_수	연령대_10_직장인구_비율	연령대_20_직장인구_비율	연령대_30_직장인구_비율	연령대_40_직장인구_비율	연령대_50_직장인구_비율	연령대_60_이상_직장_인구_비율	총_상주인구_수	연령대_10_상주인구_비율	연령대_20_상주인구_비율	연령대_30_상주인구_비율	연령대_40_상주인구_비율	연령대_50_상주인구_비율	연령대_60_이상_상주인구_비율	총_가구_수	집객시설_수	월_평균_소득_금액	지출_총금액	유사_업종_점포_수	개업_점포_수	폐업_점포_수	편의점_밀도
시간대_매출금액	1.00	0.70	0.67	-0.08	0.11	0.04	-0.02	-0.05	-0.08	0.39	-0.16	0.08	0.25	0.05	-0.14	-0.16	0.50	0.65	-0.05	0.41	0.83	0.32	0.38	0.19
시간대_유동인구_수	0.70	1.00	0.59	-0.04	0.15	0.07	-0.06	-0.12	-0.07	0.64	-0.19	0.12	0.29	0.14	-0.21	-0.23	0.69	0.64	-0.16	0.70	0.77	0.32	0.35	0.04
총_직장_인구_수	0.67	0.59	1.00	-0.10	0.05	0.20	0.12	-0.07	-0.26	0.17	-0.15	0.11	0.22	0.05	-0.08	-0.20	0.29	0.58	0.15	0.30	0.62	0.36	0.39	0.22
연령대_10_직장인구_비율	-0.08	-0.04	-0.10	1.00	0.18	0.03	-0.43	-0.02	0.01	-0.04	0.17	0.05	-0.12	0.02	-0.02	-0.04	-0.06	0.11	0.02	-0.09	-0.07	-0.00	-0.03	-0.04
연령대_20_직장인구_비율	0.11	0.15	0.05	0.18	1.00	0.49	-0.40	-0.76	-0.55	-0.03	0.13	-0.10	-0.26	0.11	-0.02	0.23	-0.01	0.35	0.14	0.02	0.23	0.08	0.10	-0.12
연령대_30_직장인구_비율	0.04	0.07	0.20	0.03	0.49	1.00	-0.02	-0.79	-0.72	-0.18	0.11	-0.16	-0.16	0.18	0.00	0.13	-0.14	0.17	0.29	-0.11	0.23	0.08	0.10	-0.12
연령대_40_직장인구_비율	-0.02	-0.06	0.12	-0.43	-0.40	-0.02	1.00	0.18	-0.35	-0.09	-0.10	0.03	0.14	-0.16	0.07	-0.05	-0.06	-0.20	0.08	-0.06	-0.05	-0.04	-0.02	0.14
연령대_50_직장인구_비율	-0.05	-0.12	-0.07	-0.02	-0.76	-0.79	0.18	1.00	0.51	0.05	-0.08	0.12	0.19	-0.16	0.01	-0.17	0.03	-0.26	-0.11	-0.04	-0.21	-0.05	-0.09	0.11
연령대_60_이상_직장_인구_비율	-0.08	-0.07	-0.26	0.01	-0.55	-0.72	-0.35	0.51	1.00	0.22	-0.11	0.12	0.16	-0.04	-0.03	-0.16	0.16	-0.17	-0.37	0.16	-0.22	-0.08	-0.10	0.04
총_상주인구_수	0.39	0.64	0.17	-0.04	-0.03	-0.18	-0.09	0.05	0.22	1.00	-0.15	0.20	0.27	0.05	-0.19	-0.27	0.96	0.22	-0.27	0.89	0.39	0.15	0.17	-0.21
연령대_10_상주인구_비율	-0.16	-0.19	-0.15	0.17	0.13	0.11	-0.10	-0.08	-0.11	-0.15	1.00	-0.15	-0.70	0.06	-0.02	0.08	-0.26	-0.05	0.47	-0.17	-0.17	-0.12	-0.11	-0.12
연령대_20_상주인구_비율	0.08	0.12	0.11	0.05	-0.10	-0.16	0.03	0.12	0.12	0.20	-0.15	1.00	0.19	-0.62	-0.02	-0.58	0.20	0.04	-0.01	0.14	0.13	0.06	0.10	0.06
연령대_30_상주인구_비율	0.25	0.29	0.22	-0.12	-0.26	-0.16	0.14	0.19	0.16	0.27	-0.70	0.19	1.00	0.23	-0.53	-0.51	0.42	0.04	-0.50	0.26	0.22	0.12	0.07	0.09
연령대_40_상주인구_비율	0.05	0.14	0.05	0.02	0.11	0.18	-0.16	-0.16	-0.04	0.05	0.06	-0.62	0.23	1.00	-0.36	-0.12	0.10	0.02	-0.09	0.08	0.04	0.03	-0.02	-0.08
연령대_50_상주인구_비율	-0.14	-0.21	-0.08	-0.02	-0.02	0.00	0.07	0.01	-0.03	-0.19	-0.02	-0.02	-0.53	-0.36	1.00	0.18	-0.28	-0.09	0.35	-0.17	-0.14	-0.07	0.03	0.08
연령대_60_이상_상주인구_비율	-0.16	-0.23	-0.20	-0.04	0.23	0.13	-0.05	-0.17	-0.16	-0.27	0.08	-0.58	-0.51	-0.12	0.18	1.00	-0.31	0.01	0.01	-0.21	-0.15	-0.07	-0.07	-0.07
총_가구_수	0.50	0.69	0.29	-0.06	-0.01	-0.14	-0.06	0.03	0.16	0.96	-0.26	0.20	0.42	0.10	-0.28	-0.31	1.00	0.30	-0.30	0.86	0.51	0.22	0.24	-0.15
집객시설_수	0.65	0.64	0.58	0.11	0.35	0.17	-0.20	-0.26	-0.17	0.22	-0.05	0.04	0.04	0.02	-0.09	0.01	0.30	1.00	0.10	0.31	0.81	0.36	0.38	0.16
월_평균_소득_금액	-0.05	-0.16	0.15	0.02	0.14	0.29	0.08	-0.11	-0.37	-0.27	0.47	-0.01	-0.50	-0.09	0.35	0.01	-0.30	0.10	1.00	-0.26	-0.00	0.01	-0.00	-0.10
지출_총금액	0.41	0.70	0.30	-0.09	0.02	-0.11	-0.06	-0.04	0.16	0.89	-0.17	0.14	0.26	0.08	-0.17	-0.21	0.86	0.31	-0.26	1.00	0.47	0.17	0.21	-0.18
유사_업종_점포_수	0.83	0.77	0.62	-0.07	0.23	0.23	-0.05	-0.21	-0.22	0.39	-0.17	0.13	0.22	0.04	-0.14	-0.15	0.51	0.81	-0.00	0.47	1.00	0.44	0.46	0.24
개업_점포_수	0.32	0.32	0.36	-0.00	0.08	0.08	-0.04	-0.05	-0.08	0.15	-0.12	0.06	0.12	0.03	-0.07	-0.07	0.22	0.36	0.01	0.17	0.44	1.00	0.41	0.14
폐업_점포_수	0.38	0.35	0.39	-0.03	0.10	0.10	-0.02	-0.09	-0.10	0.17	-0.11	0.10	0.07	-0.02	0.03	-0.07	0.24	0.38	-0.00	0.21	0.46	0.41	1.00	0.12
편의점_밀도	0.19	0.04	0.22	-0.04	-0.12	-0.12	0.14	0.11	0.04	-0.21	-0.12	0.06	0.09	-0.08	0.08	-0.07	-0.15	0.16	-0.10	-0.18	0.24	0.14	0.12	1.00

4. EDA

MULTICOLLINEARITY

- Checking the Variance Inflation Factor (VIF) for the independent variables
- Proportion of Working Population by Age, Proportion of Resident Population by Age, Total Resident Population, Total Households, and Number of Similar Business Stores—is high, exceeding 10.

=> This suggests the presence of multicollinearity among these variables. It is expected that utilizing tree-based models will help to address the issue of multicollinearity.



	VIF Factor	features
0	4.457923	시간대_유동인구_수
1	5.303403	총_직장_인구_수
2	inf	연령대_10_직장인구_비율
3	inf	연령대_20_직장인구_비율
4	inf	연령대_30_직장인구_비율
5	inf	연령대_40_직장인구_비율
6	inf	연령대_50_직장인구_비율
7	inf	연령대_60_이상_직장_인구_비율
8	35.345023	총_상주인구_수
9	inf	연령대_10_상주인구_비율
10	inf	연령대_20_상주인구_비율
11	inf	연령대_30_상주인구_비율
12	inf	연령대_40_상주인구_비율
13	inf	연령대_50_상주인구_비율
14	inf	연령대_60_이상_상주인구_비율
15	32.874815	총_가구_수
16	5.411849	집객시설_수
17	2.586857	월_평균_소득_금액
18	6.784772	지출_총금액
19	15.814548	유사_업종_점포_수
20	1.357144	개업_점포_수
21	1.417630	폐업_점포_수
22	1.574667	편의점_밀도

CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

5. MODELING

TIME SERIES ANALYSIS

- Analyzing **the change in data y over time** e.g.) predicting stock prices, sales, or temperatures

- Typically involves examining how it varies according to trends, cycles, seasonal components, and irregular or random elements. It's assumed that these components cause fluctuations in the data.

- Observing patterns of data variation → Segmenting them into trend, seasonal, and irregular components → Applying forecasting techniques like exponential modeling or ARIMA (AutoRegressive Integrated Moving Average) methods

5. MODELING

TIME SERIES ANALYSIS

Trend: A component that represents the overall upward or downward direction of observations.

Cycle: A component indicating changes that are periodic but not seasonal, characterized by longer cycles.

Seasonal: A component that represents factors explained by regular variations according to specific periods.

Irregular: A component representing errors that cannot be explained by specific patterns.

Random: A component representing random causes that are independent of regular movements over time.

5. MODELING

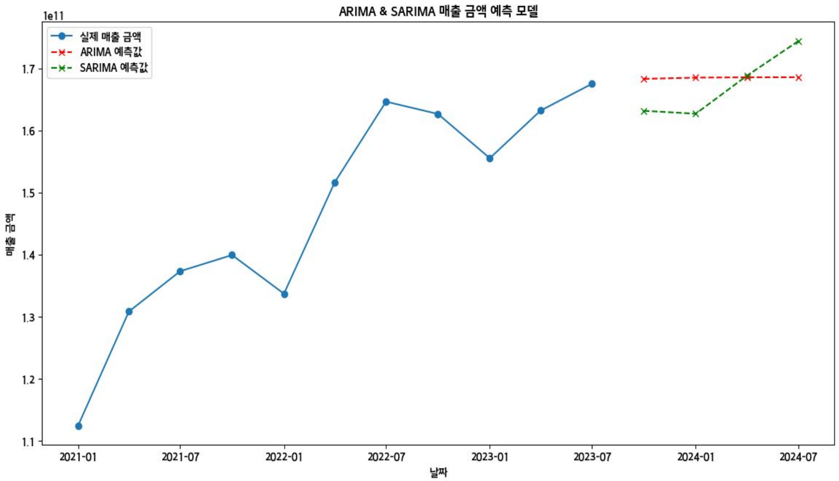
ARIMA & SARIMA

ARIMA (AutoRegressive Integrated Moving Average)

- Suitable for modeling the trend and volatility of non-seasonal data.
- **Differencing** is used to address the non-stationarity of time series.

SARIMA (Seasonal ARIMA)

- Extended version of ARIMA that can additionally model seasonal patterns.
- **Seasonal differencing** is used to handle the seasonality of time series.



Model Performance

ARIMA - MAE 21,367,509,120.59 / **RMSE** 43,353,668,254.71

SARIMA - MAE 43,774,410,318.39 / **RMSE** 77,763,589,777.93

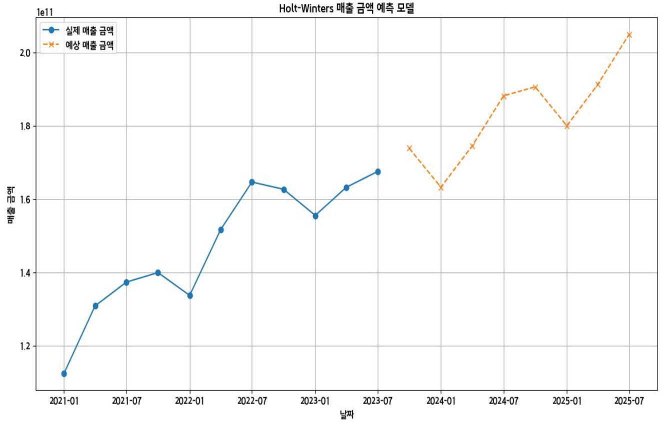
The ARIMA model shows a lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to the SARIMA model.

Since the error range is quite wide, considering other models is preferred.

5. MODELING

HOLT-WINTERS MODEL TIME SERIES PREDICTION

- Statistical techniques used for analyzing and forecasting time series data
- Considers the three main components that may exist in time series data: **level**, **trend**, and **seasonality**.
- Predict future data points by accounting for all these components, useful for short-term forecasts of time series data



- For convenience, assuming the first month of each quarter
- Fitting the Holt-Winters model/future predictions

Model Performance

Mean Absolute Error (MAE) : 7,317,570,194.14

Root Mean Squared Error (RMSE) : 8,237,463,143.96

Since the error range is quite wide, considering other models is preferred.

5. MODELING

FINAL MODEL SELECTION

- After applying various machine learning algorithms such as time series, RandomForest, XGBoost, and LightGBM, the **LightGBM** model showed the best performance.

Algorithm	Data Reprocessing	Feature Engineering	Cross validation	Hyperparameter Tuning	MAE	MSE	RMSE	R-Squared
LGBM	Categorical Variable : Dummy Encoding / Numeric Variable Standard Scaler Dependent Variable : Box-Cox Transformation	Proportion of Resident Population by Age Proportion of Working Population by Age Convenience Store Density		num_leaves': [25, 30, 35], 'learning_rate': [0.12, 0.13, 0.14], 'n_estimators': [375, 400, 425]	2,136,328,813,899,580	26930386.26	46220437.19	**0.9885977983**

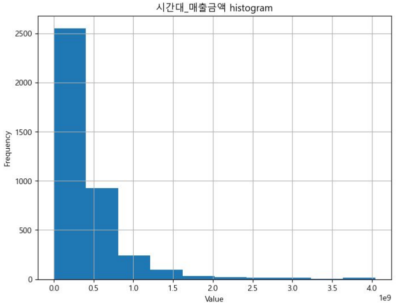
5. MODELING

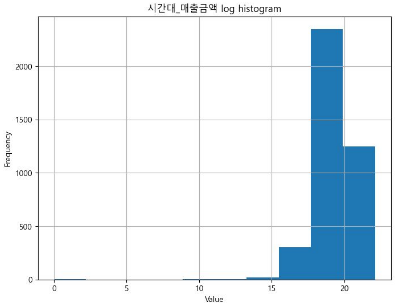
DEPENDENT VARIABLE SCALING

- Box-Cox transformation is a method of adjusting the distribution of data to control skewness.
- Commonly used to normalize data distributions in models where the data distribution is not normal.

Log Transformation

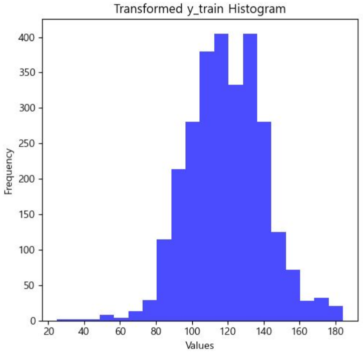
- The dependent variable is left-skewed, so attempted log transformation to convert it to a normal distribution.
- However, after log transformation, it became right-skewed and the distribution issue was not resolved.

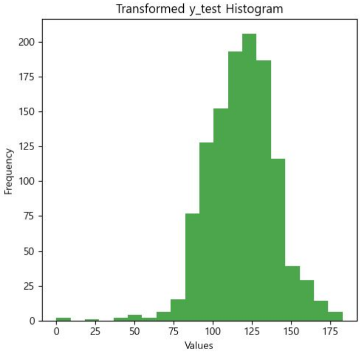




Box-Cox Transformation

- Attempted Box-Cox transformation to address the skewness in the distribution.
- The dependent variable in both the train and test datasets transformed into a shape similar to a normal distribution.

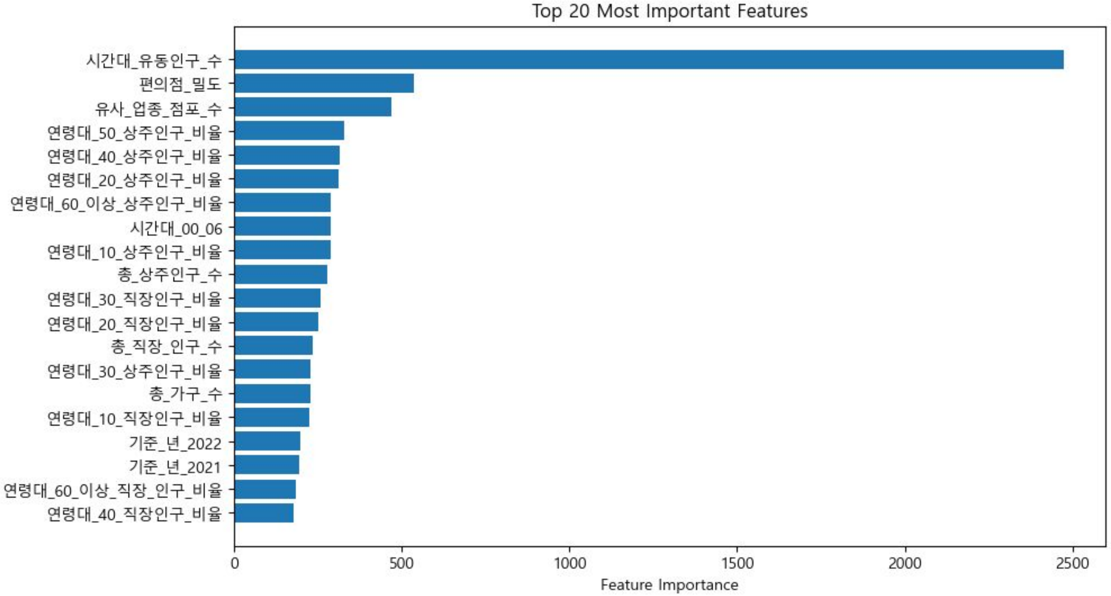




5. MODELING

FEATURE IMPORTANCE

- Confirmed that the feature **[Hourly Floating Population]** which showed a high correlation in the correlation analysis is the most important feature.

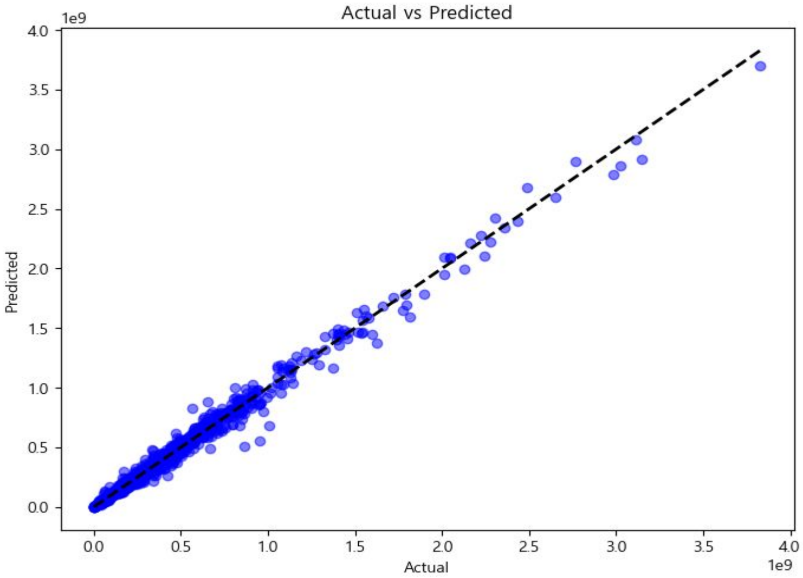


5. MODELING

Actual Values vs Predicted Values

- When visualizing actual values versus predicted values, it is observed that they are distributed almost evenly around the y=x line.

  => This indicates that the model predicts the data well.



CONTENT

INTRODUCTION

- TEAM MEMBERS
- BACKGROUND
- WEB SERVICE
- DEVELOPMENT ENVIRONMENT
- PROJECT PERIOD

PROJECT

- FLOW CHART
- WBS

DATA

- DATA COLLECTION
- DATA PREPROCESSING

EDA

- EDA
- CORRELATION ANALYSIS
- MULTICOLLINEARITY

MODELING

- MODEL
- MODEL TRAINING
- MODEL PERFORMANCE VALIDATION
- CONVENIENCE STORE SALES PREDICTION

APP & DOCUMENT

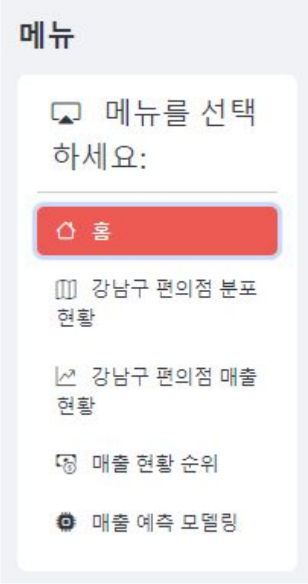
- STREAMLIT
- LIMITATIONS IMPROVEMENTS
- REFERENCES
- APPENDIX

자세한 사용법은 부록 참조

6. DOCUMENTS

SERVICE (STREAMLIT)

Menu Composition

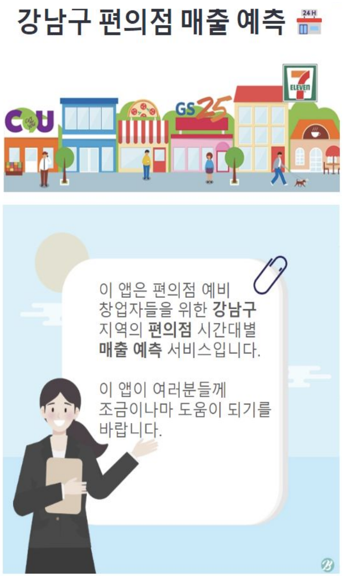


메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

Streamlit - HOME

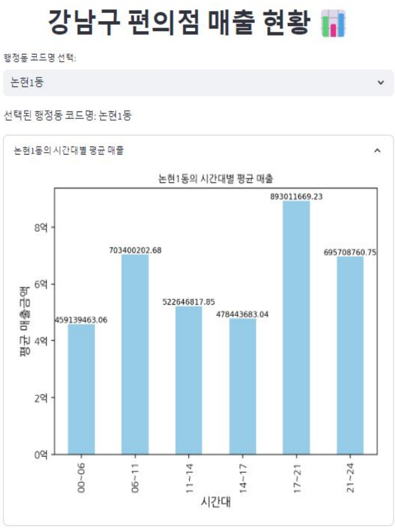


강남구 편의점 매출 예측

이 앱은 편의점 예비 창업자들을 위한 **강남구** 지역의 **편의점** 시간대별 **매출 예측** 서비스입니다.

이 앱이 여러분들께 조금이나마 도움이 되기를 바랍니다.

Streamlit - EDA



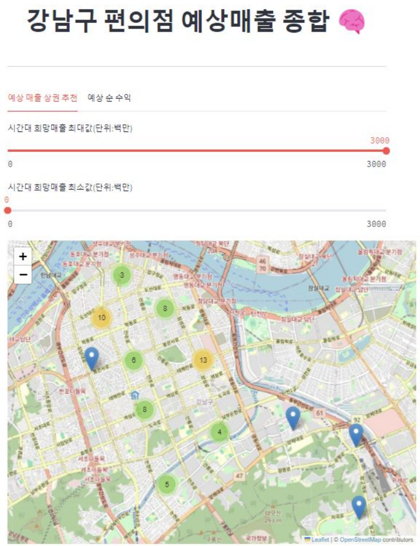
강남구 편의점 매출 현황

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출



강남구 편의점 예상매출 종합

예상 매출 상권 추천 | 예상 순 수익

시간대 희망매출 최대값(단위:백만): 3000

0 3000

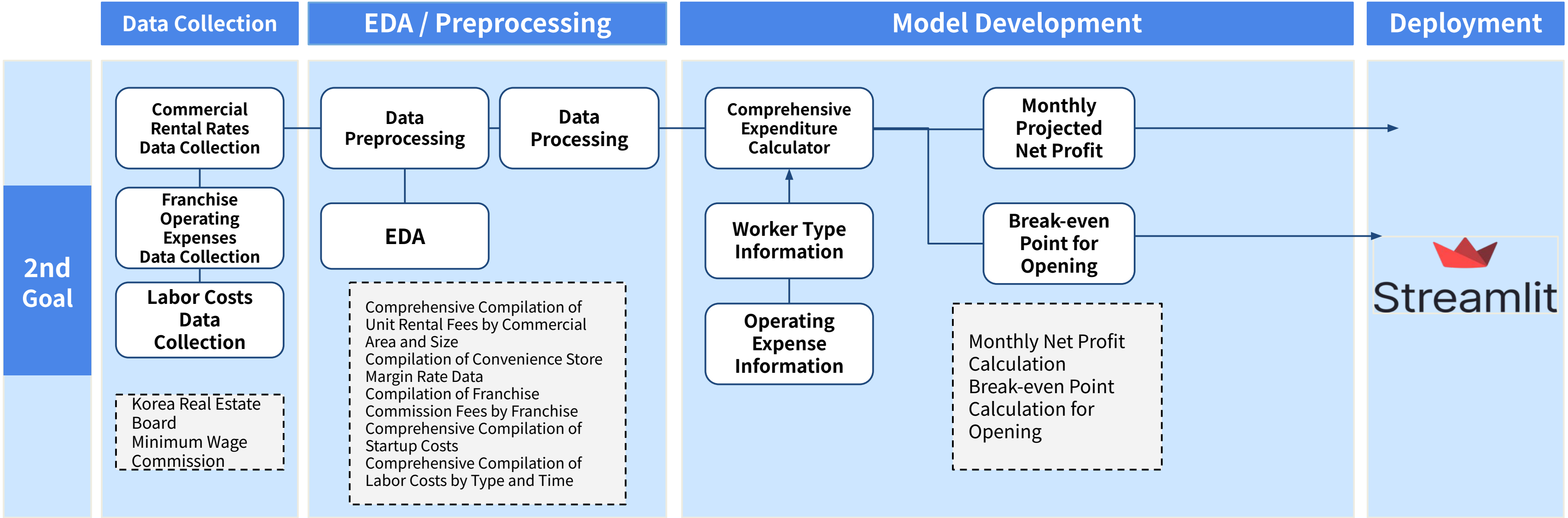
시간대 희망매출 최소값(단위:백만): 0

0 3000

6. DOCUMENTS

LIMITATIONS / IMPROVEMENTS

- Intended to implement a monthly net profit prediction model by calculating labor costs and franchise operating expenses, but unable to do so due to time constraints.



Data Collection
EDA / Preprocessing
Model Development
Deployment
2nd Goal
Commercial Rental Rates Data Collection
Franchise Operating Expenses Data Collection
Labor Costs Data Collection
Data Preprocessing
Data Processing
EDA
Comprehensive Expenditure Calculator
Worker Type Information
Operating Expense Information
Monthly Projected Net Profit
Break-even Point for Opening
Streamlit

Korea Real Estate Board
Minimum Wage Commission

Comprehensive Compilation of Unit Rental Fees by Commercial Area and Size
Compilation of Convenience Store Margin Rate Data
Compilation of Franchise Commission Fees by Franchise
Comprehensive Compilation of Startup Costs
Comprehensive Compilation of Labor Costs by Type and Time

Monthly Net Profit Calculation
Break-even Point Calculation for Opening

6. DOCUMENTS

REFERENCES

- **PAPERS**

1) 김현철, 이승일, 2019, “서울시 골목상권 매출액에 영향을 미치는 요인에 관한 연구”, 「서울도시연구」, 제 20권 제 1호
2) 황규성, 2014, “편의점 입지선정시 매출에 영향을 미치는 요인분석”
3) 김미성, 2020, “서울시 상권 데이터의 시각화에 기반한 매출액 예측”
4) 이철환, 2012, “편의점의 상권 추정과 매출 예측에 관한 연구”
5) 김동명, 2020, “시스템 다이내믹스를 활용한 편의점 특정 상품 매출 분석 및 예측”
6) 이임동, 이찬호, 강상목, 2010, “편의점 매출에 영향을 미치는 입지요인에 대한 실증연구”

- **NEWS ARTICLES**

1) 송금종, 카페 1950개·편의점 470개…강남구, 서울 최대 ‘슬세권’, 쿠키뉴스, 2023.12.24
https://www.kukinews.com/newsView/kuk202312140078

2) 이지원, 편의점 본사가 제시한 ‘예상 매출액’이 과장이라면…, 더스쿠프, 2023.02.14
https://www.thescoop.co.kr/news/articleView.html?idxno=56799

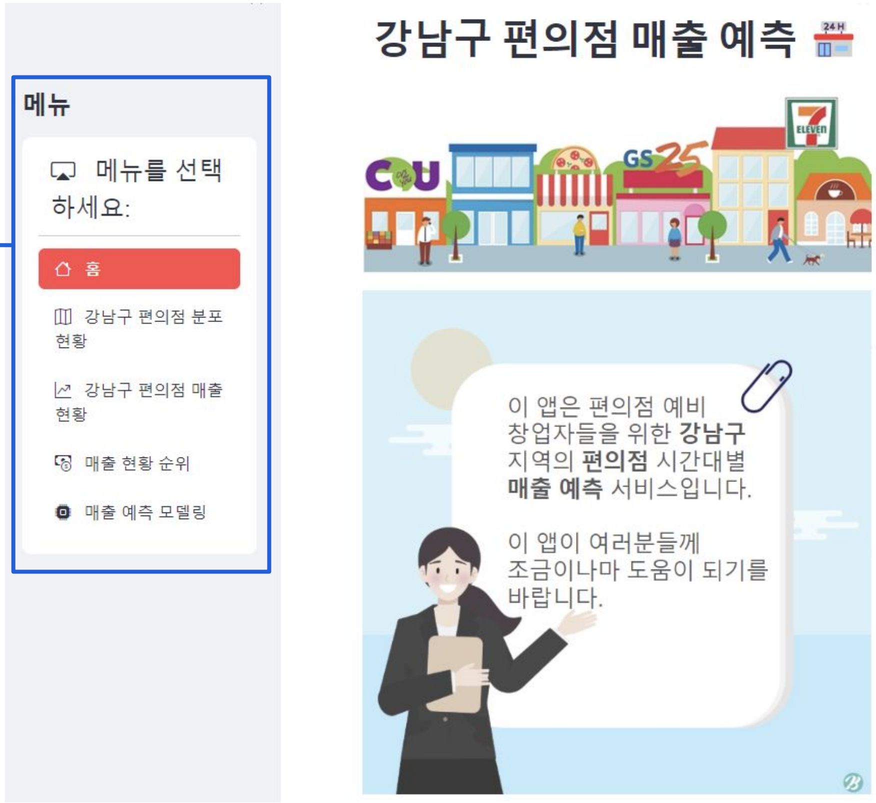
3) 이진원, 홈플러스, 예비 편의점주에 예상매출 뻥튀기, 시민일보, 2017.11.05
https://www.siminilbo.co.kr/news/articleView.html?idxno=537797

6. DOCUMENTS

APPENDIX - SERVICE

Page Navigation Sidebar

- Select a menu to navigate to the desired page



강남구 편의점 매출 예측

이 앱은 편의점 예비 창업자들을 위한 **강남구** 지역의 **편의점** 시간대별 **매출 예측** 서비스입니다.

이 앱이 여러분들께 조금이나마 도움이 되기를 바랍니다.

6. DOCUMENTS

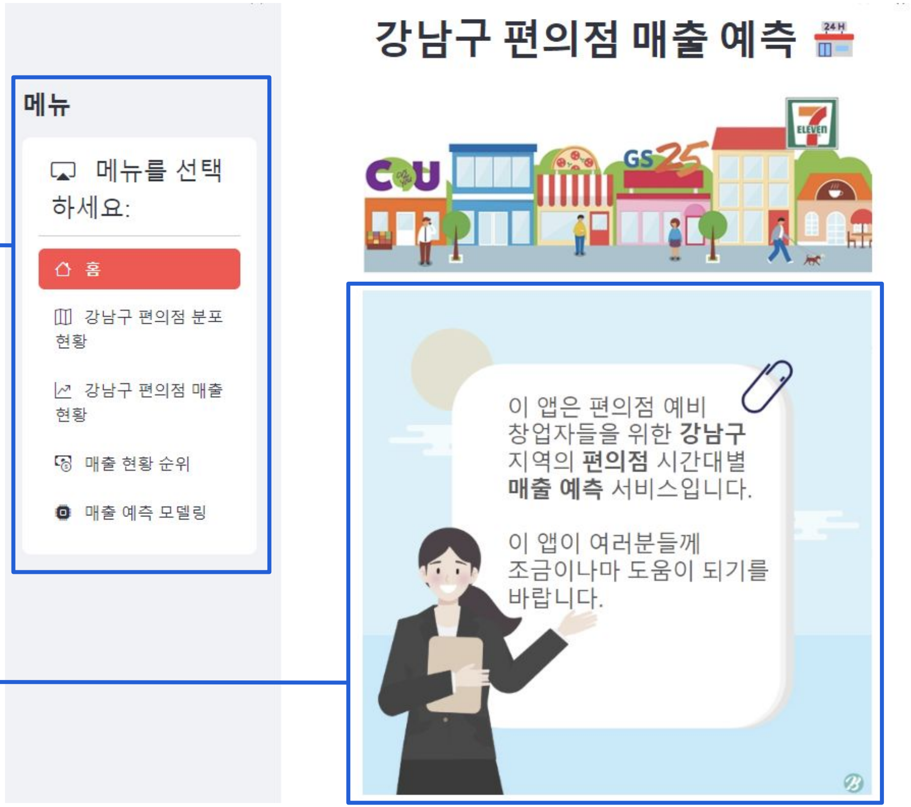
APPENDIX - SERVICE

Page Navigation Sidebar

- Select a menu to navigate to the desired page

Home Screen

- Purpose of the service



강남구 편의점 매출 예측

이 앱은 편의점 예비 창업자들을 위한 **강남구** 지역의 **편의점** 시간대별 **매출 예측** 서비스입니다.

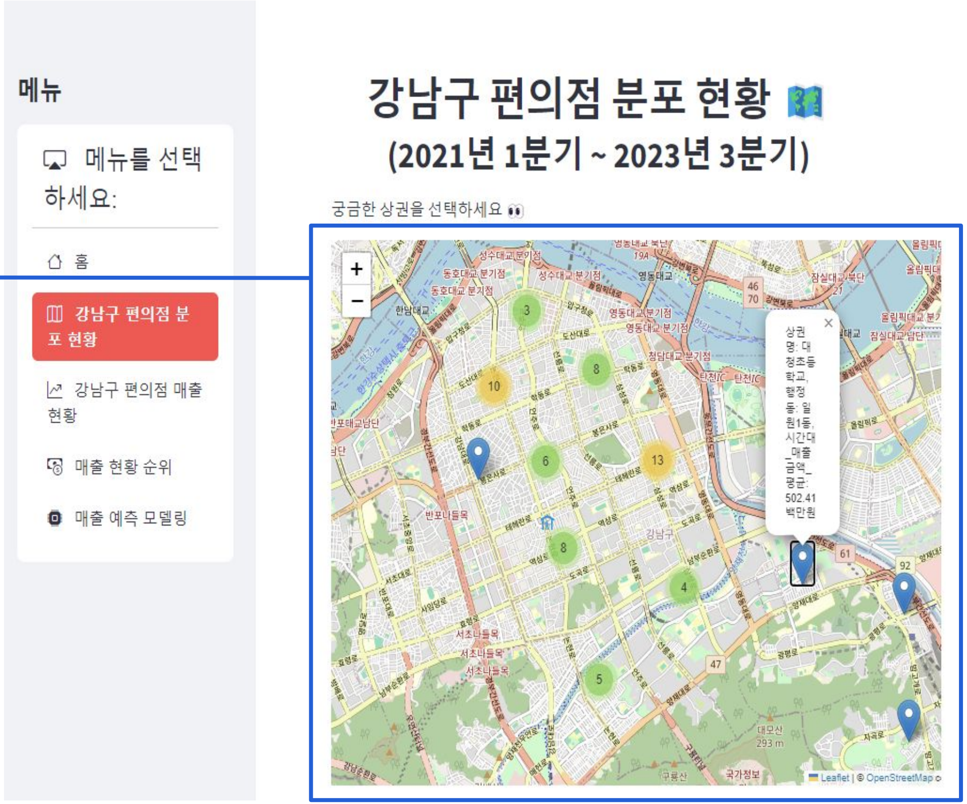
이 앱이 여러분들께 조금이나마 도움이 되기를 바랍니다.

6. DOCUMENTS

APPENDIX - SERVICE

Map Visualization

- Checking the distribution status of the desired commercial area



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

강남구 편의점 분포 현황 🗺️
(2021년 1분기 ~ 2023년 3분기)

궁금한 상권을 선택하세요 👀

상권명: 대청초등학교, 행정동: 일원1동, 시간대_매출금액_평균: 502.41 백만원

6. DOCUMENTS

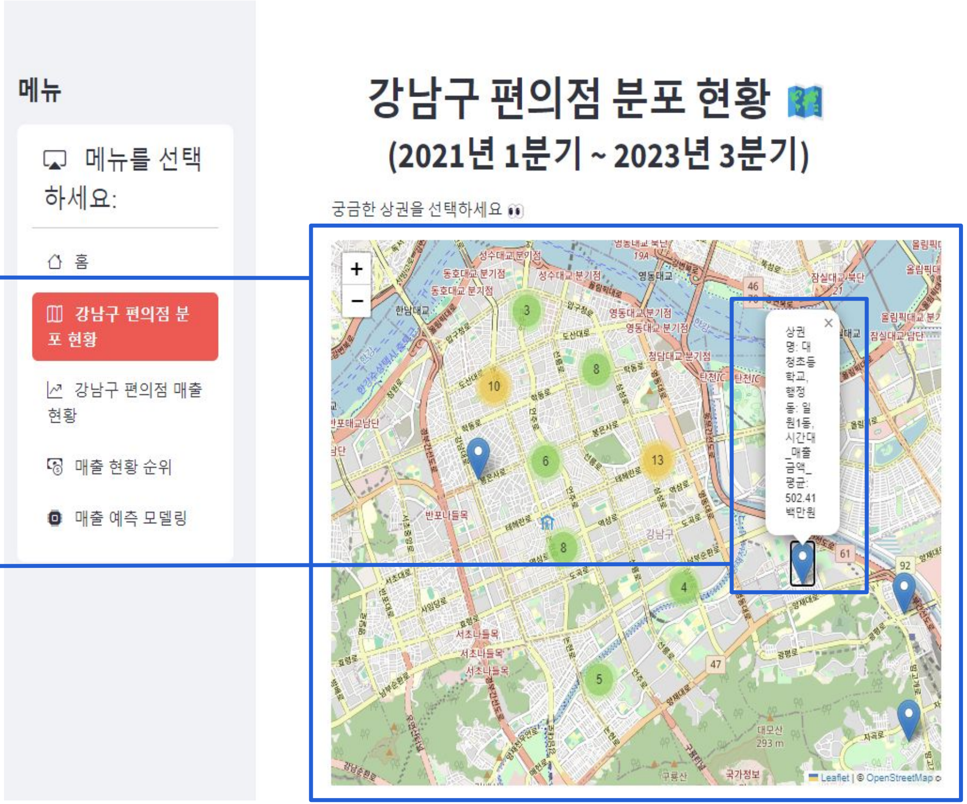
APPENDIX - SERVICE

Map Visualization

- Checking the distribution status of the desired commercial area

Detailed Information

- Average sales amount for the selected commercial area



강남구 편의점 분포 현황 🗺️

(2021년 1분기 ~ 2023년 3분기)

궁금한 상권을 선택하세요 👀

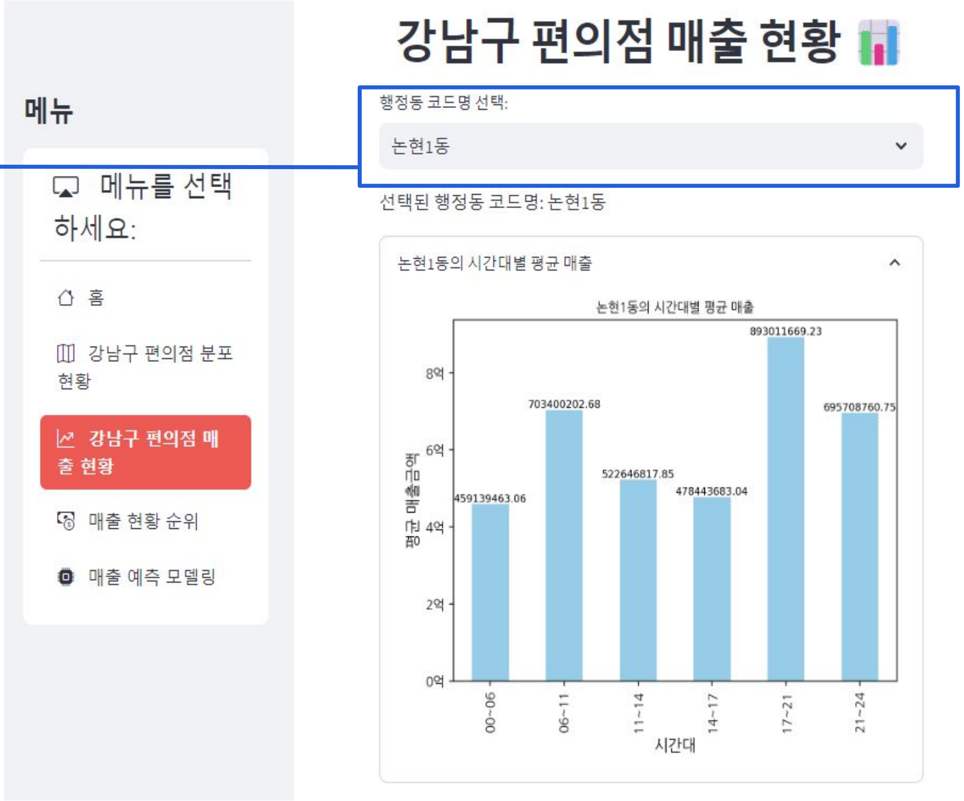
상권명: 대청초등학교, 행정동: 일원1동, 시간대_매출금액_평균: 502.41 백만원

6. DOCUMENTS

APPENDIX - SERVICE

Selectbox

- Select the desired administrative district



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

강남구 편의점 매출 현황 📊

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출

6. DOCUMENTS

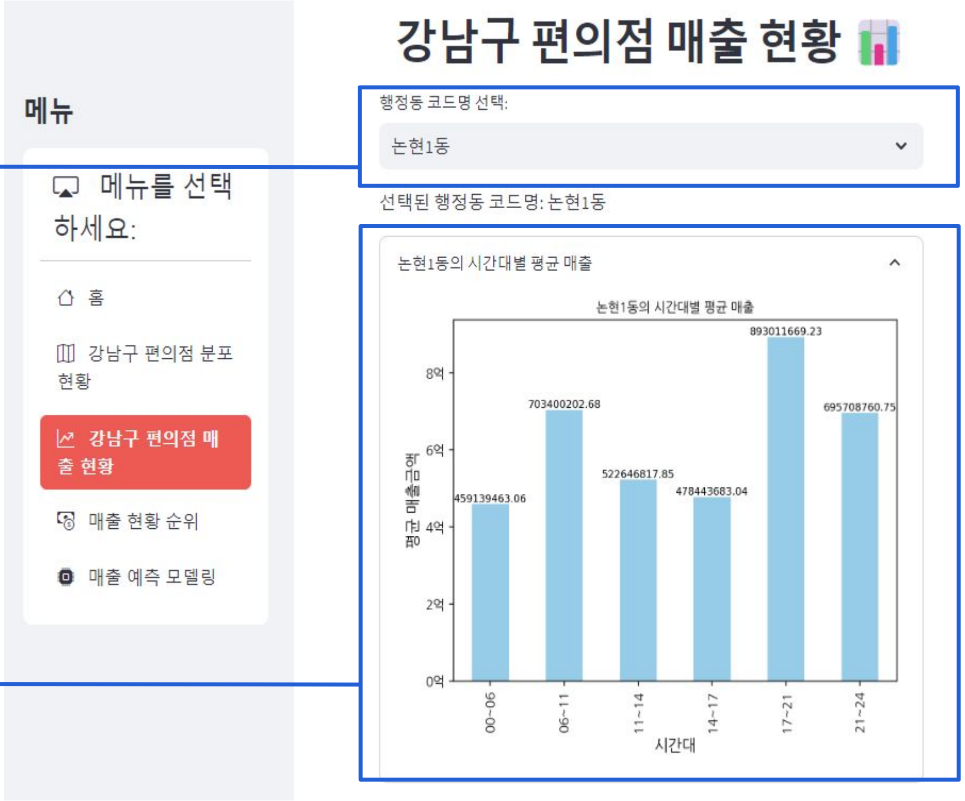
APPENDIX - SERVICE

Selectbox

- Select the desired administrative district

Bar Graph

- The hourly average sales for the administrative district selected by the user



강남구 편의점 매출 현황 📊

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

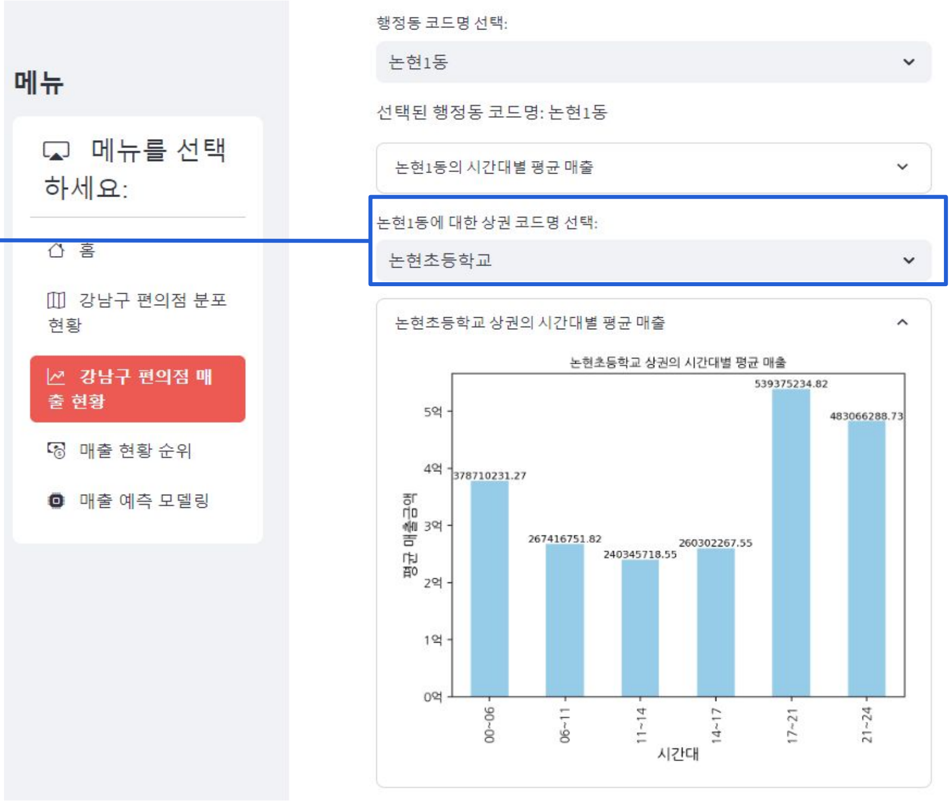
논현1동의 시간대별 평균 매출

6. DOCUMENTS

APPENDIX - SERVICE

Selectbox

- Select the commercial area corresponding to the administrative district chosen above



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출

논현1동에 대한 상권 코드명 선택:

논현초등학교

논현초등학교 상권의 시간대별 평균 매출

6. DOCUMENTS

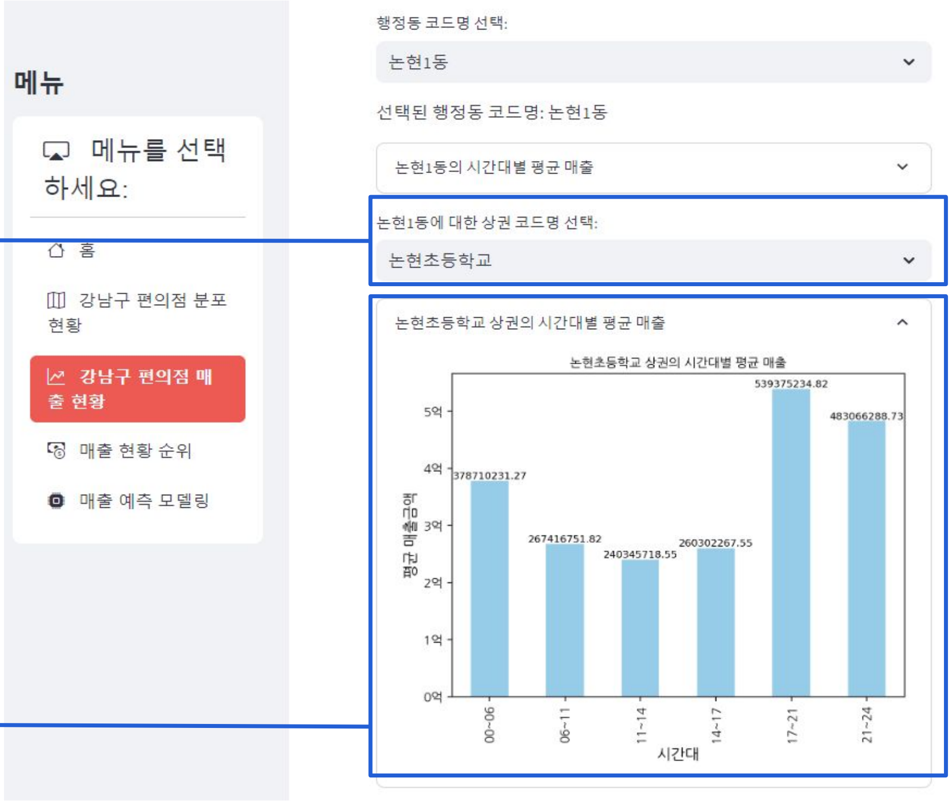
APPENDIX - SERVICE

Selectbox

- Select the commercial area corresponding to the administrative district chosen above

Bar Graph

- The hourly average sales for the commercial area selected by the user



행정동 코드명 선택:

논현1동

선택된 행정동 코드명: 논현1동

논현1동의 시간대별 평균 매출

논현1동에 대한 상권 코드명 선택:

논현초등학교

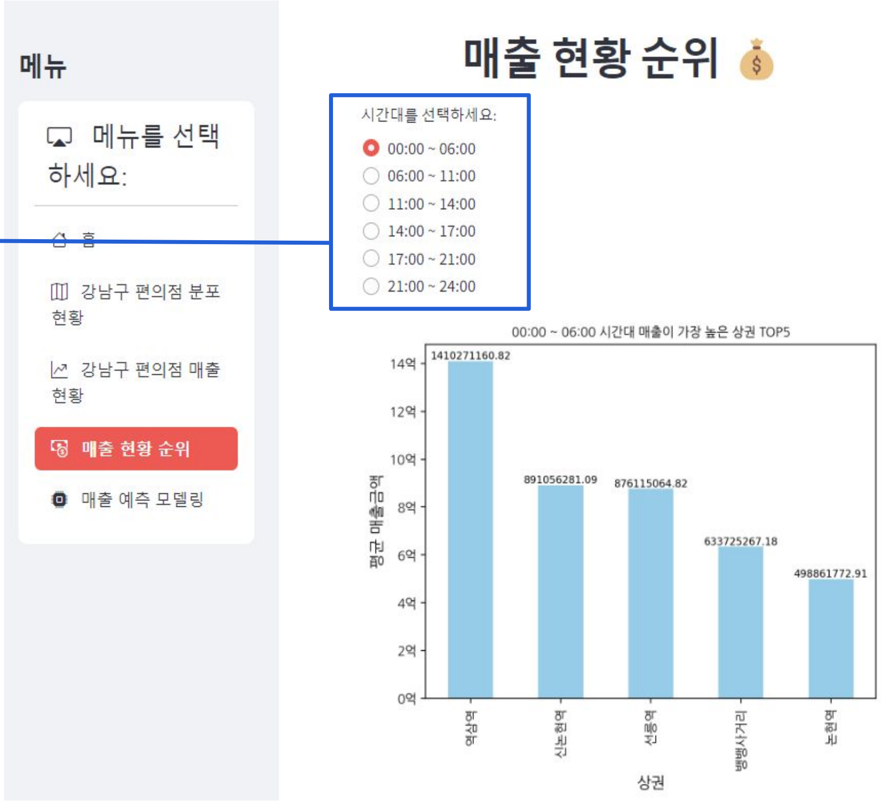
논현초등학교 상권의 시간대별 평균 매출

6. DOCUMENTS

APPENDIX - SERVICE

Radio Button

- Select the desired time slot



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

매출 현황 순위 💰

시간대를 선택하세요:

● 00:00 ~ 06:00
○ 06:00 ~ 11:00
○ 11:00 ~ 14:00
○ 14:00 ~ 17:00
○ 17:00 ~ 21:00
○ 21:00 ~ 24:00

00:00 ~ 06:00 시간대 매출이 가장 높은 상권 TOP5

6. DOCUMENTS

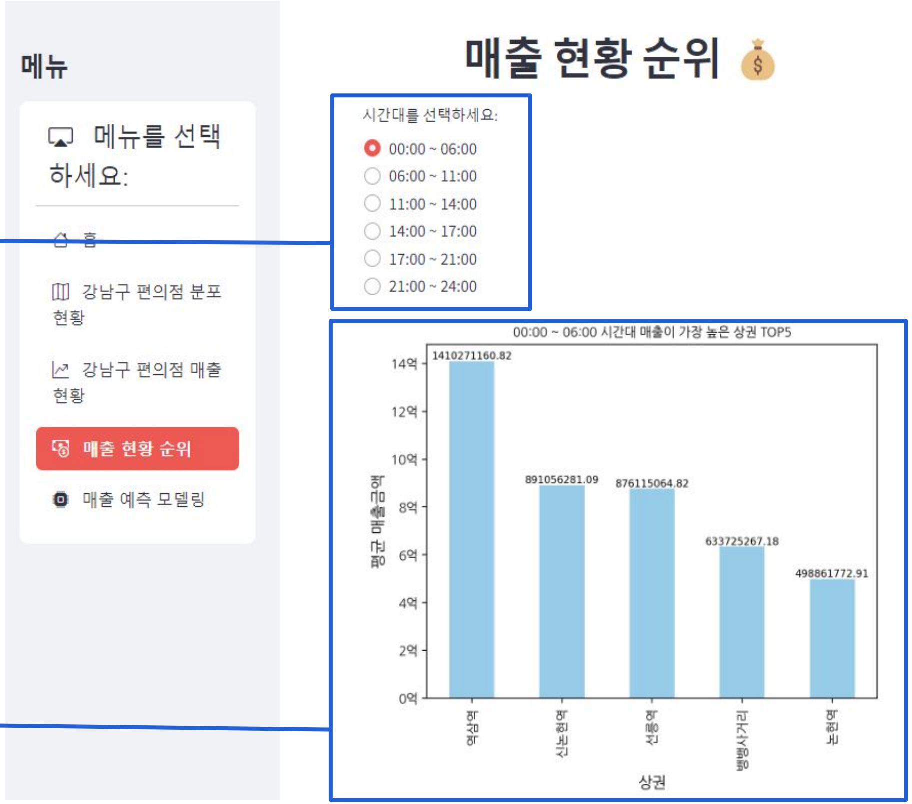
APPENDIX - SERVICE

Radio Button

- Select the desired time slot

Bar Graph

- The sales of the top 5 commercial areas that recorded the highest sales during the time slot selected by the user



시간대를 선택하세요:

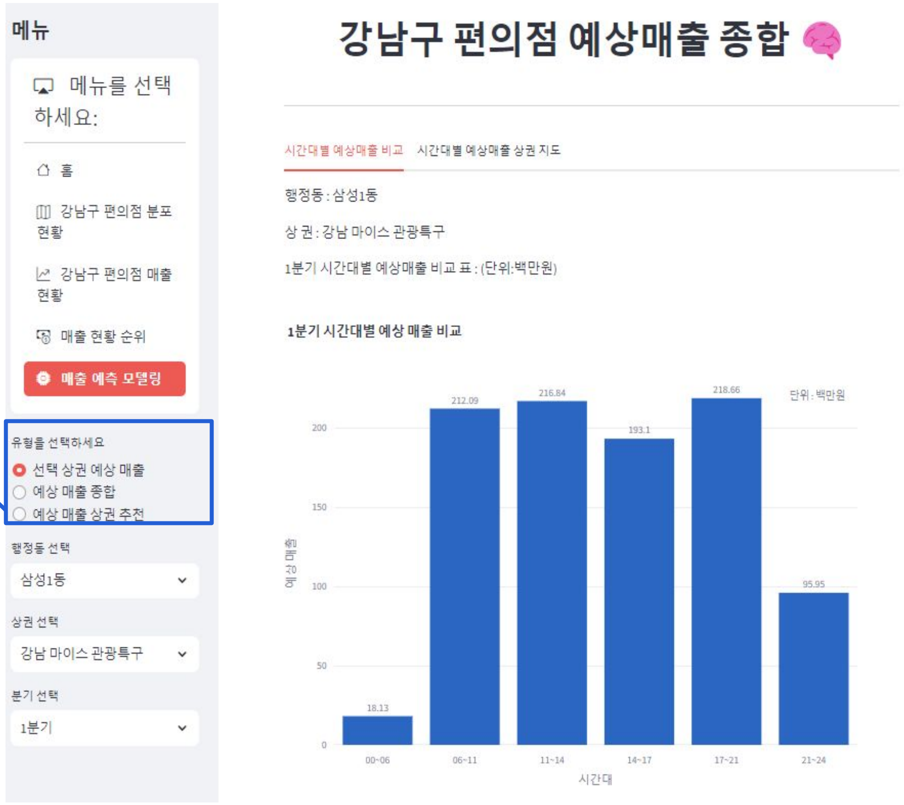
- ● 00:00 ~ 06:00
- ○ 06:00 ~ 11:00
- ○ 11:00 ~ 14:00
- ○ 14:00 ~ 17:00
- ○ 17:00 ~ 21:00
- ○ 21:00 ~ 24:00

6. DOCUMENTS

APPENDIX - SERVICE

Radio Button

- Select the desired type



유형을 선택하세요

● 선택 상권 예상 매출

○ 예상 매출 종합

○ 예상 매출 상권 추천

행정동 선택

삼성1동

상권 선택

강남 마이스 관광특구

분기 선택

1분기

강남구 편의점 예상매출 종합

행정동 : 삼성1동

상 권 : 강남 마이스 관광특구

1분기 시간대별 예상매출 비교 표 : (단위:백만원)

1분기 시간대별 예상 매출 비교

6. DOCUMENTS

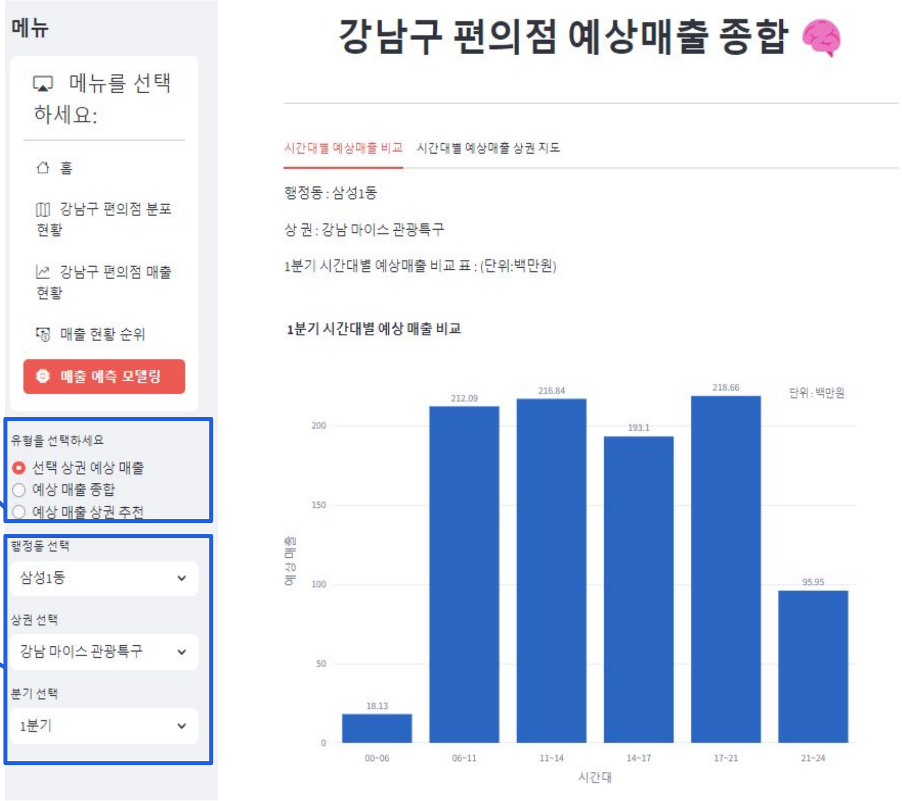
APPENDIX - SERVICE

Radio Button

- Select the desired type

Selectbox

- Select the desired administrative district, commercial area, and quarter



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

유형을 선택하세요

● 선택 상권 예상 매출
○ 예상 매출 종합
○ 예상 매출 상권 추천

행정동 선택: 삼성1동

상권 선택: 강남 마이스 관광특구

분기 선택: 1분기

강남구 편의점 예상매출 종합 🧠

시간대별 예상매출 비교 | 시간대별 예상매출 상권 지도

행정동 : 삼성1동

상 권 : 강남 마이스 관광특구

1분기 시간대별 예상매출 비교 표 : (단위:백만원)

1분기 시간대별 예상 매출 비교

6. DOCUMENTS

APPENDIX - SERVICE

Radio Button

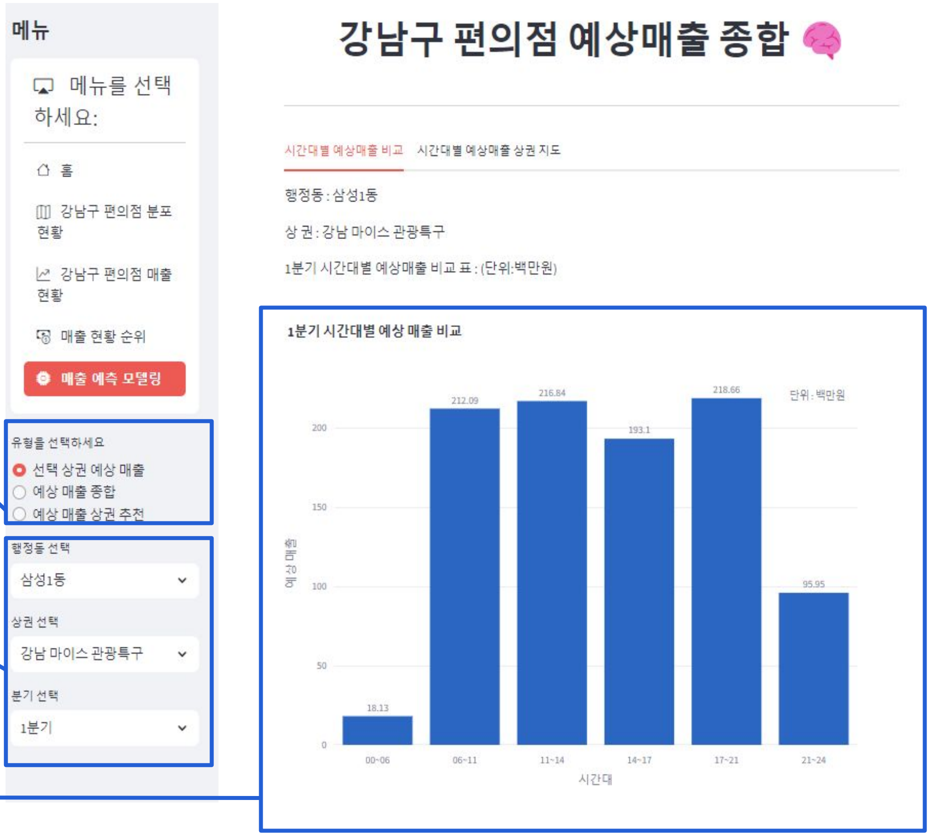
- Select the desired type

Selectbox

- Select the desired administrative district, commercial area, and quarter

Bar Graph

- The hourly projected sales based on user-selected conditions



유형을 선택하세요

● 선택 상권 예상 매출
○ 예상 매출 종합
○ 예상 매출 상권 추천

행정동 선택: 삼성1동

상권 선택: 강남 마이스 관광특구

분기 선택: 1분기

강남구 편의점 예상매출 종합

행정동 : 삼성1동

상 권 : 강남 마이스 관광특구

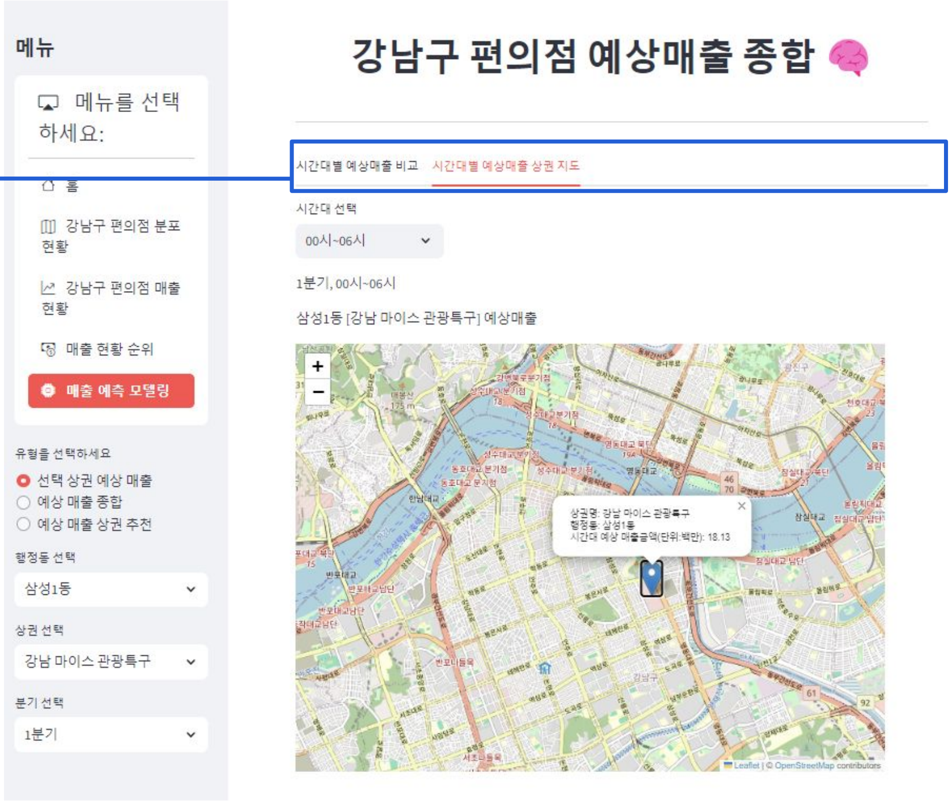
1분기 시간대별 예상매출 비교 표 : (단위:백만원)

6. DOCUMENTS

APPENDIX - SERVICE

Tab

- Move to the desired tab



유형을 선택하세요

● 선택 상권 예상 매출
○ 예상 매출 종합
○ 예상 매출 상권 추천

행정동 선택: 삼성1동

상권 선택: 강남 마이스 관광특구

분기 선택: 1분기

강남구 편의점 예상매출 종합

시간대별 예상매출 비교 | 시간대별 예상매출 상권 지도

시간대 선택: 00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스 관광특구] 예상매출

상권명: 강남 마이스 관광특구
행정동: 삼성1동
시간대 예상 매출금액(단위:백만): 18.13

6. DOCUMENTS

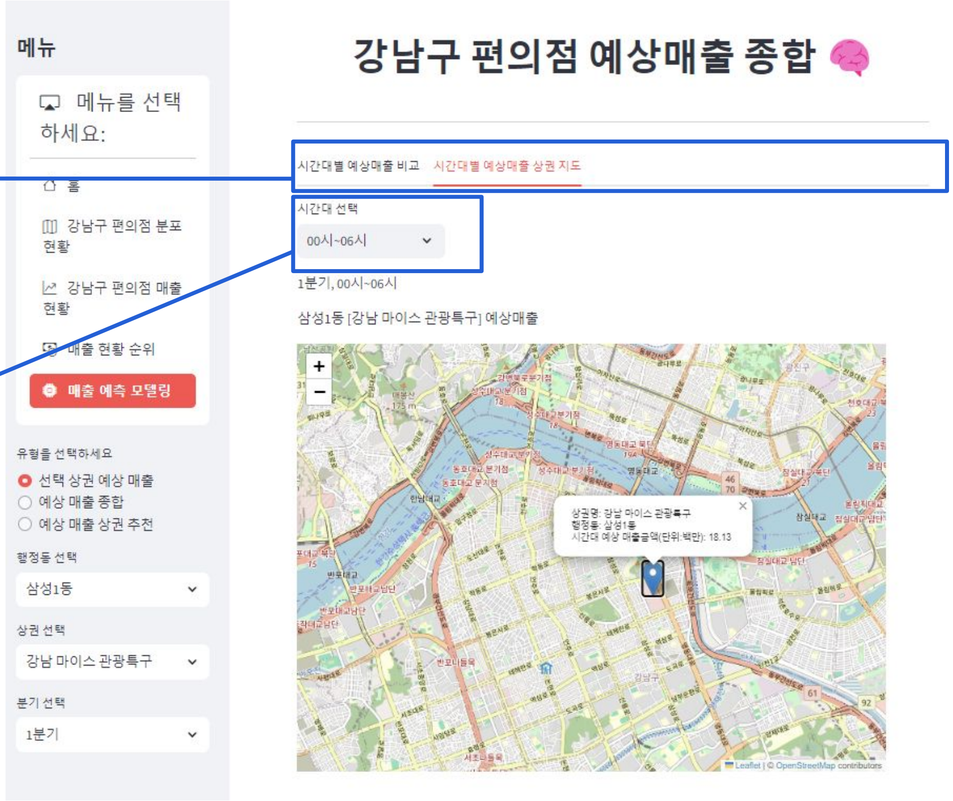
APPENDIX - SERVICE

Tab

- Move to the desired tab

Selectbox

- Select the desired time slot



메뉴

메뉴를 선택하세요:

유형을 선택하세요

● 선택 상권 예상 매출
○ 예상 매출 종합
○ 예상 매출 상권 추천

행정동 선택: 삼성1동

상권 선택: 강남 마이스 관광특구

분기 선택: 1분기

강남구 편의점 예상매출 종합 🧠

시간대 선택: 00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스 관광특구] 예상매출

상권명: 강남 마이스 관광특구
행정동: 삼성1동
시간대 예상 매출금액(단위:백만): 18.13

6. DOCUMENTS

APPENDIX - SERVICE

Tab

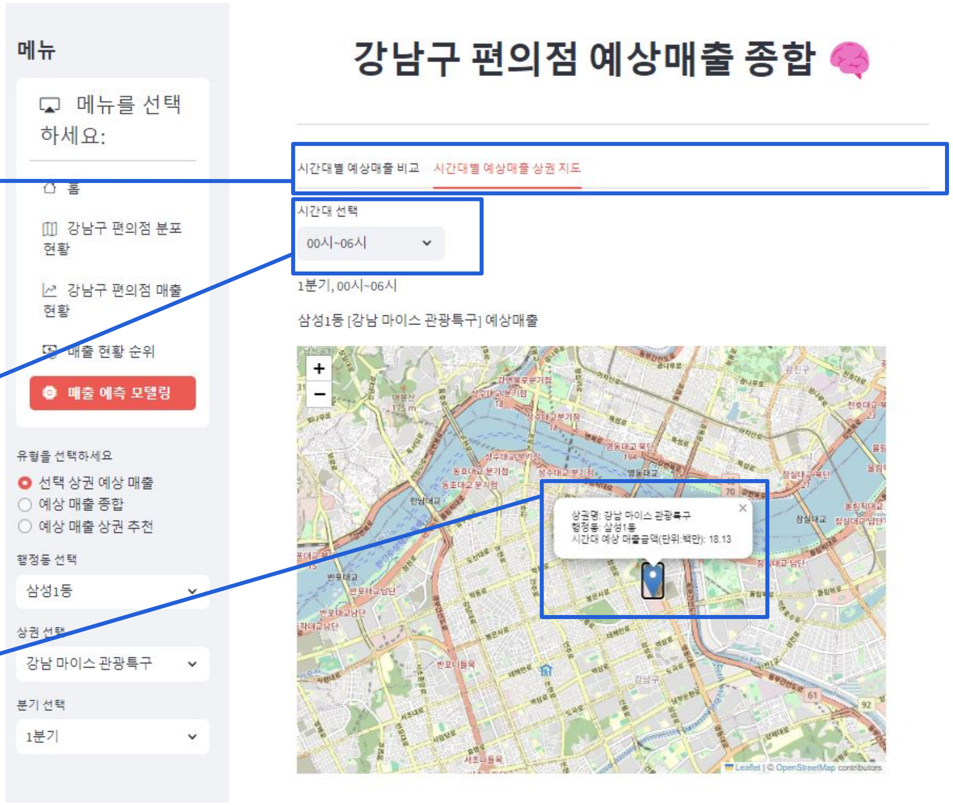
- Move to the desired tab

Selectbox

- Select the desired time slot

Detailed Information

- The projected sales amount for the time slot selected by the user



메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

유형을 선택하세요

● 선택 상권 예상 매출
○ 예상 매출 종합
○ 예상 매출 상권 추천

행정동 선택: 삼성1동

상권 선택: 강남 마이스 관광특구

분기 선택: 1분기

강남구 편의점 예상매출 종합 🧠

시간대별 예상매출 비교 | 시간대별 예상매출 상권 지도

시간대 선택: 00시~06시

1분기, 00시~06시

삼성1동 [강남 마이스 관광특구] 예상매출

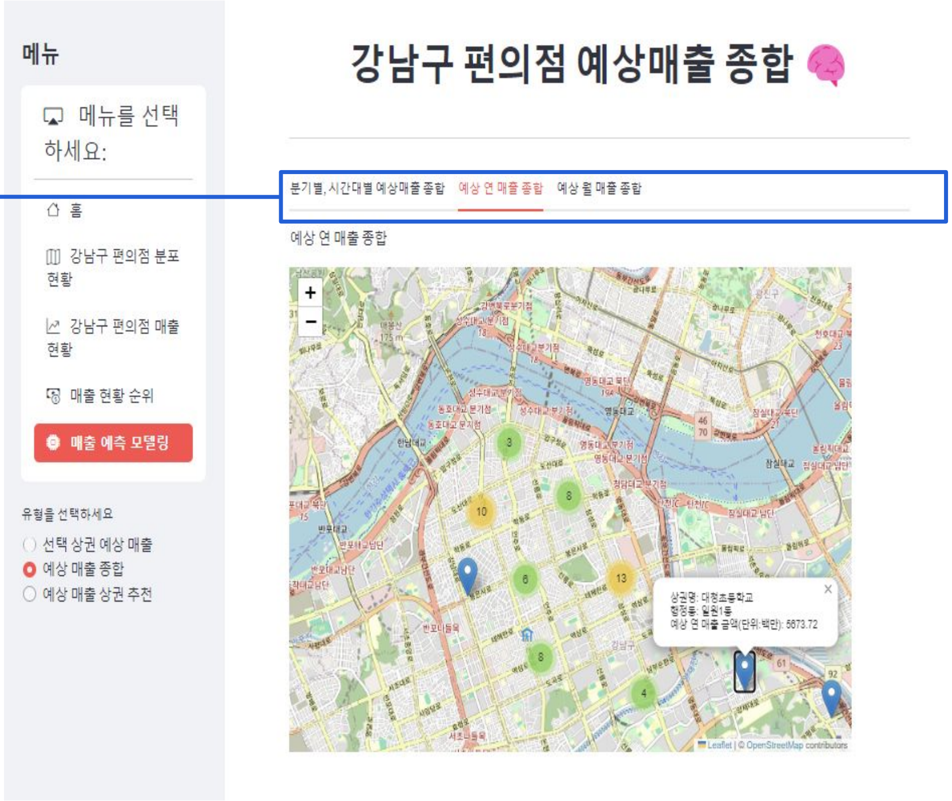
상권명: 강남 마이스 관광특구
행정동: 삼성1동
시간대 예상 매출금액(단위:백만): 18.13

6. DOCUMENTS

APPENDIX - SERVICE

Tab

- Check the projected sales for each commercial area by time slot, year, and month individually



상권명: 대청초등학교
행정동: 일원1동
예상 연 매출 금액(단위:백만): 5673.72

6. DOCUMENTS

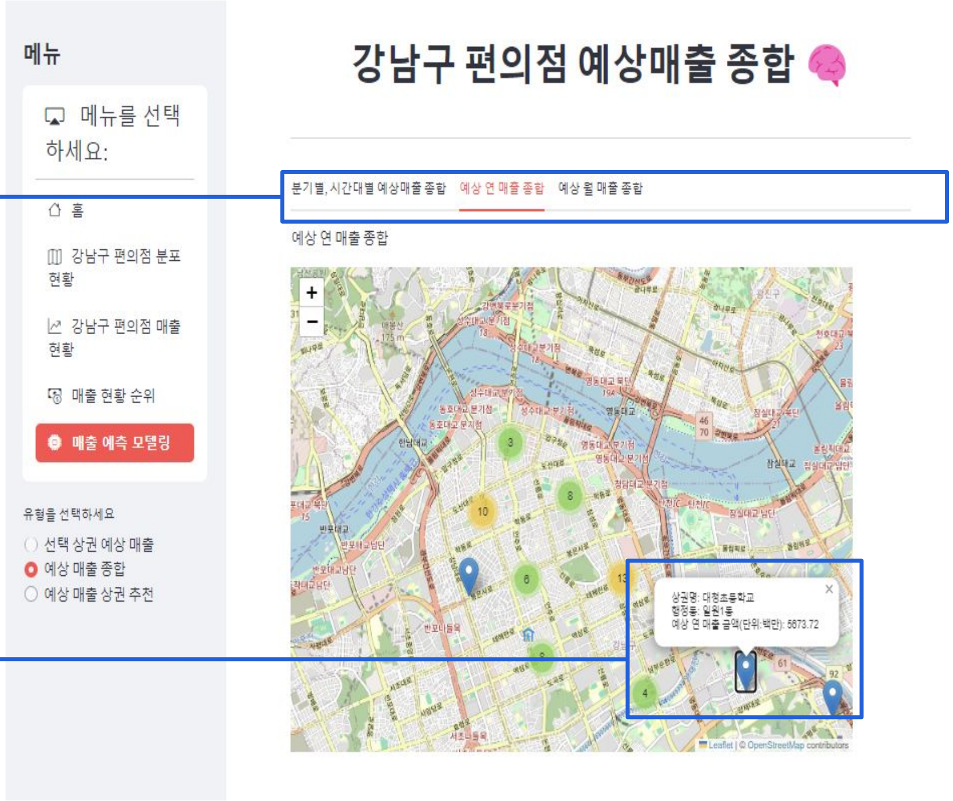
APPENDIX - SERVICE

Tab

- Check the projected sales for each commercial area by time slot, year, and month individually

Detailed Information

- Projected sales by time slot, year, and month for each commercial area can be visualized on a map



강남구 편의점 예상매출 종합

유형을 선택하세요

○ 선택 상권 예상 매출
● 예상 매출 종합
○ 예상 매출 상권 추천

예상 연 매출 종합

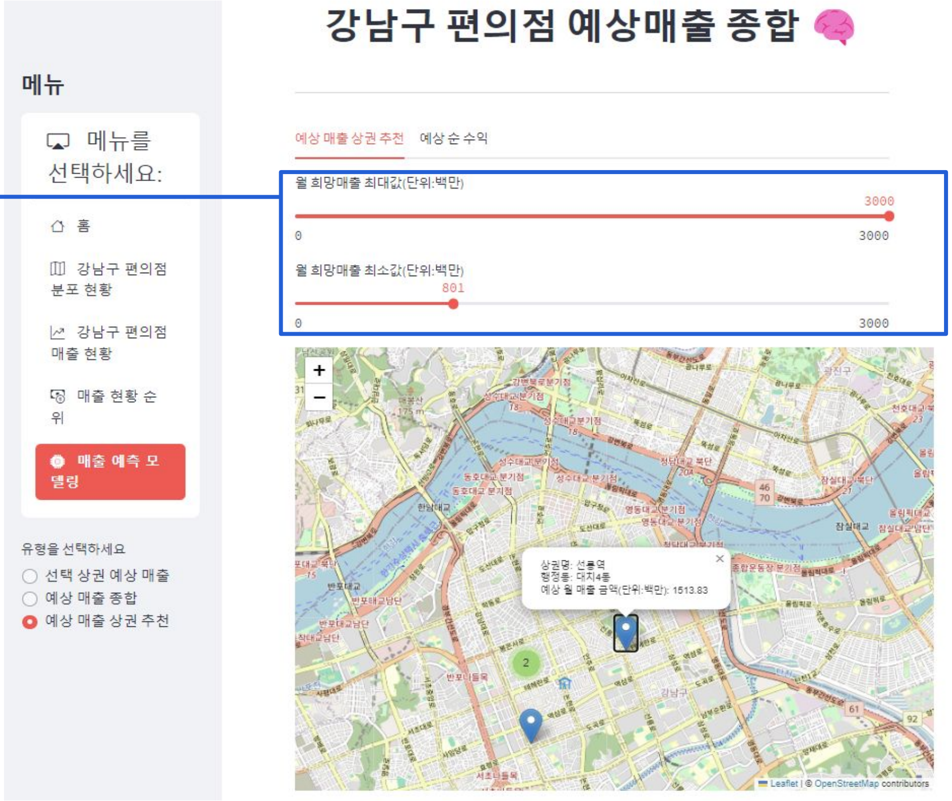
상권명: 대청초등학교
행정동: 일원1동
예상 연 매출 금액(단위:백만): 5673.72

6. DOCUMENTS

APPENDIX - SERVICE

Slider

- The user inputs the desired minimum and maximum values for sales



강남구 편의점 예상매출 종합 🧠

메뉴

메뉴를 선택하세요:

- 홈
- 강남구 편의점 분포 현황
- 강남구 편의점 매출 현황
- 매출 현황 순위
- 매출 예측 모델링

유형을 선택하세요

○ 선택 상권 예상 매출
○ 예상 매출 종합
● 예상 매출 상권 추천

예상 매출 상권 추천 | 예상 순수익

월 희망매출 최대값(단위:백만): 3000 (0 – 3000)

월 희망매출 최소값(단위:백만): 801 (0 – 3000)

상권명: 선릉역
행정동: 대치4동
예상 월 매출 금액(단위:백만): 1513.83

6. DOCUMENTS

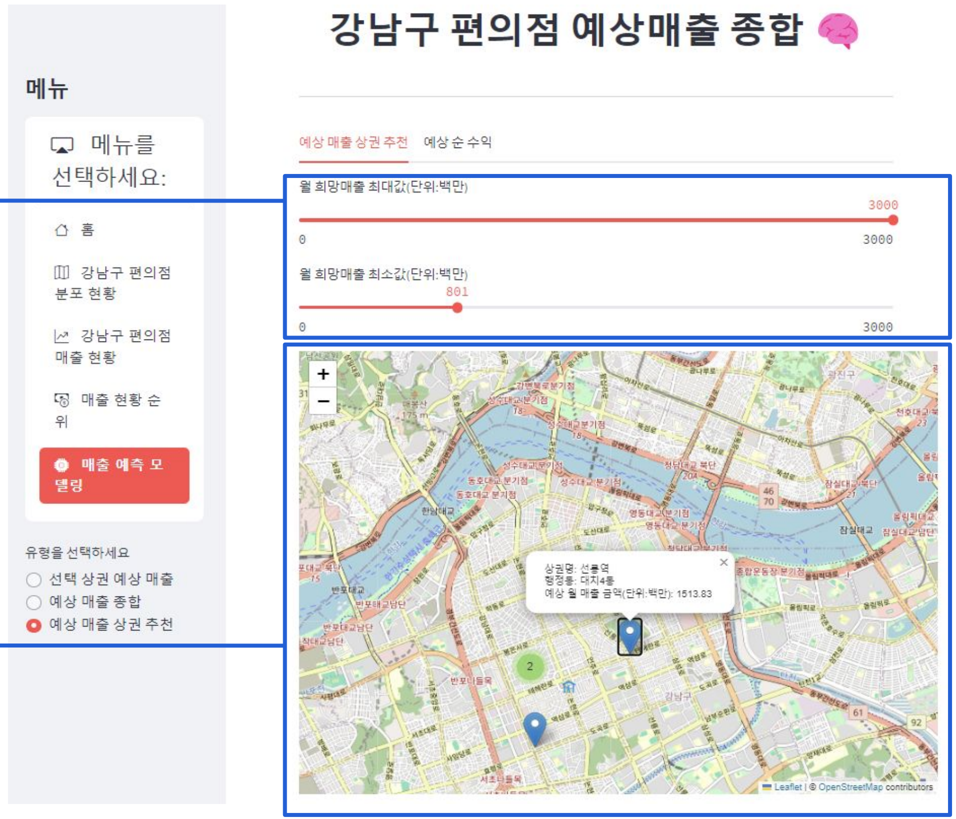
APPENDIX - SERVICE

Slider

- The user inputs the desired minimum and maximum values for sales

Detailed Information

- Information about commercial areas that meet the user's input values on the map



강남구 편의점 예상매출 종합

유형을 선택하세요

○ 선택 상권 예상 매출
○ 예상 매출 종합
● 예상 매출 상권 추천

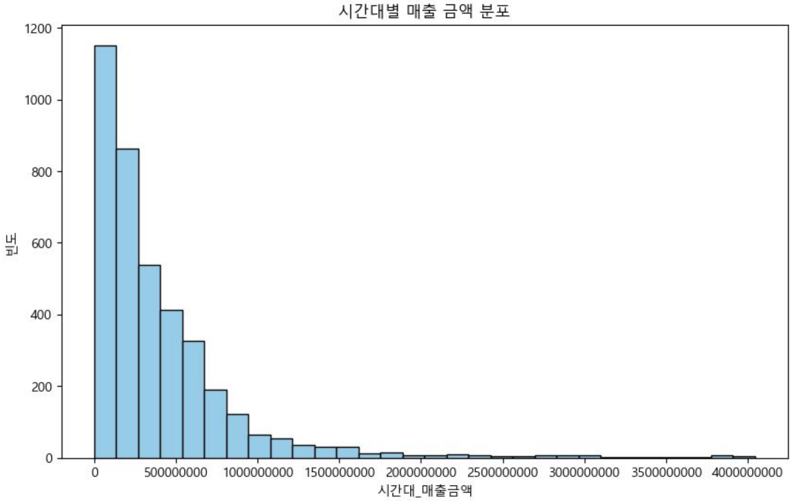
월 희망매출 최대값(단위:백만) 3000 (0 – 3000)

월 희망매출 최소값(단위:백만) 801 (0 – 3000)

상권명: 선릉역
행정동: 대치4동
예상 월 매출 금액(단위:백만): 1513.83

6. DOCUMENTS

APPENDIX - EDA



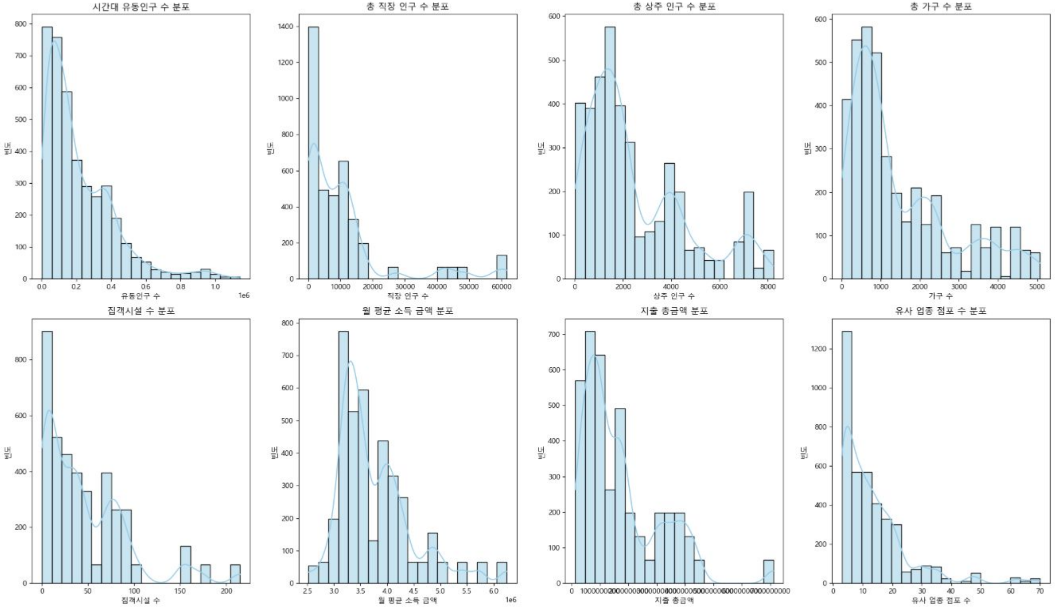
Continuous Dependent Variable

Final Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

6. DOCUMENTS

APPENDIX - EDA



Continuous Independent Variables

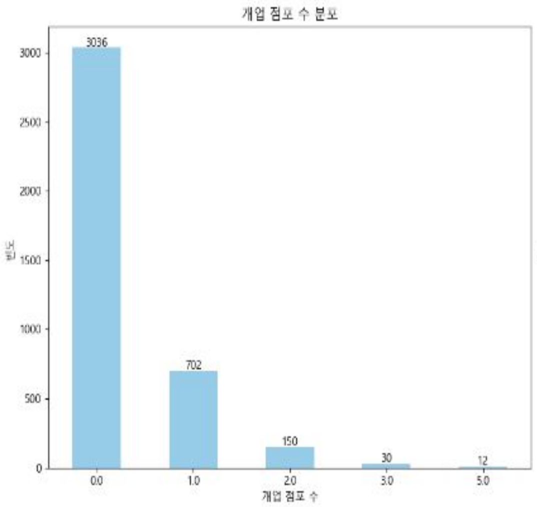
Final Columns

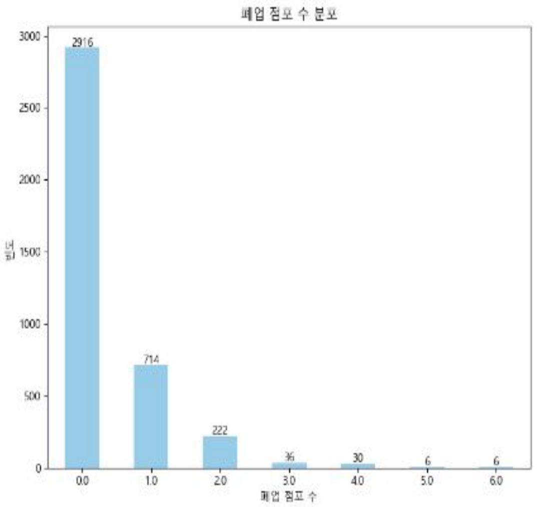
- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

6. DOCUMENTS

APPENDIX - EDA

Continuous Independent Variables
But, Similar to Categorical Data Distribution



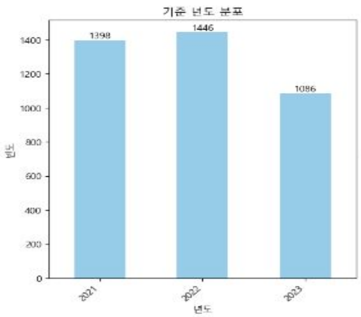


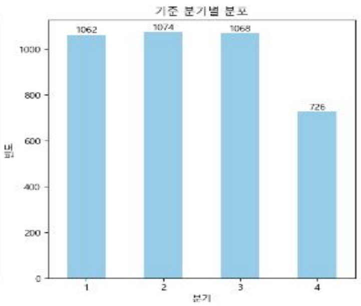
Final Columns

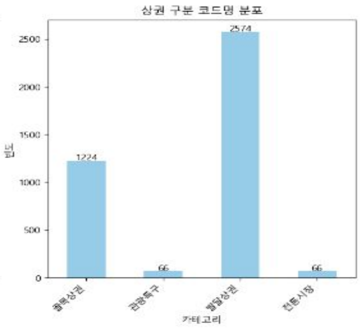
- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

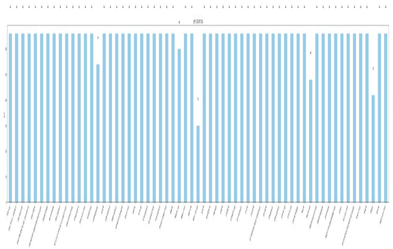
6. DOCUMENTS

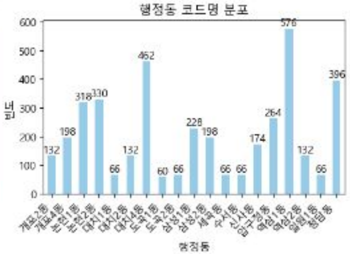
APPENDIX - EDA

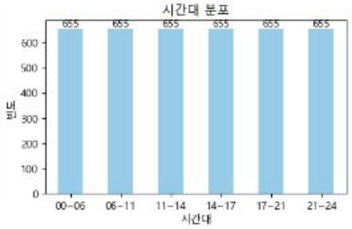












Categorical Independent Variables

Final Columns

- Sales Amount by Time Period
- Base Year
- Base Quarter
- Commercial District Division Name
- Commercial District Name
- Administrative Neighborhood Name
- Time Period
- Floating Population Count by Time Period
- Total Working Population
- Total Resident Population
- Total Number of Households
- Number of Attraction Facilities
- Monthly Average Income Amount
- Total Expenditure Amount
- Number of Stores in Similar Industries
- Number of Newly Opened Stores
- Number of Closed Stores

6. DOCUMENTS

APPENDIX - MODELING

Algorithm	Data Reprocessing	Feature Engineering	Cross validation	Hyperparameter Tuning	MAE	MSE	RMSE	R-Squared
RandomForest	StandardScaler	Employment Population Ratio by Age Group					67898424.45	**0.9648454212**
RandomForest	StandardScaler	Ratio of Household Count to Population Count					66123873.23	**0.9754612381**
RandomForest	OneHotEncoder	Employment Population Ratio by Age Group					66731521.15	**0.9648153123**
RandomForest	OneHotEncoder	Ratio of Household Count to Population Count					66197315.94	**0.9618132176**

6. DOCUMENTS

APPENDIX - MODELING

Algorithm	Data Reprocessing	Feature Engineering	Cross validation	Hyperparameter Tuning	MAE	MSE	RMSE	R-Squared
RandomForest	StandardScaler: Numeric Data OneHotEncoder: Categorical Data	Ratio of Household Count to Population Count	KFold				69963979.84	**0.9798526552**
RandomForest	StandardScaler: Numeric Data OneHotEncoder: Categorical Data	Ratio of Household Count to Population Count	KFold	{'n_estimators': 500, 'min_samples_split': 2, 'min_samples_leaf: 1, 'max_depth': 50, 'bootstrap': True}			69628037.42	**0.9800303186**
RandomForest + GradientBoosting	StandardScaler: Numeric Data OneHotEncoder: Categorical Data	StandardScaler: Numerical Data OneHotEncoder: Categorical Data						**0.9773915991**

6. DOCUMENTS

APPENDIX - MODELING

Algorithm	Data Reprocessing	Feature Engineering	Cross validation	Hyperparameter Tuning	MAE	MSE	RMSE	R-Squared
XGBoost	StandardScaler: Numeric Data OneHotEncoder: Categorical Data	Ratio of Household Count to Population Count, Ratio of Floating Population on Weekdays to Weekends	KFold				60577427.09	**0.9849553596**
XGBoost	StandardScaler: Numeric Data OneHotEncoder: Categorical Data	Ratio of Household Count to Population Count, Ratio of Floating Population on Weekdays to Weekends	KFold	{'subsample': 0.6, 'n_estimators': 500, 'min_child_weight': 1, 'max_depth': 7, 'learning_rate': 0.05, 'colsample_bytree': 0.8}			57367659.29	**0.9864766996**
XGBoost	Removing 'Hourly Sales Amount', 'Average Weekday Floating Population Count', 'Average Weekend Floating Population Count', 'Total Resident Population Count', 'Commercial Area Code', 'Administrative District Code Name'	Ratio of Household Count to Population Count					75272170.99	**0.9812817025**

6. DOCUMENTS

APPENDIX - MODELING

Algorithm	Data Reprocessing	Feature Engineering	Cross validation	Hyperparameter Tuning	MAE	MSE	RMSE	R-Squared
LGBM	Categorical Variable One-Hot Encoding / Numeric Variable Standard Scaling Dependent Variable Box-Cox Transformation	Proportion of Resident Population by Age Proportion of Working Population by Age Convenience Store Density		num_leaves=31, learning_rate=0.1, n_estimators=100	3,518,915,356,168,770	33206245.03	59617671	**0.9812185361**
LGBM	Categorical Variable One-Hot Encoding / Numeric Variable Standard Scaling Dependent Variable Box-Cox Transformation	Proportion of Resident Population by Age Proportion of Working Population by Age Convenience Store Density		GridSearchCV 'num_leaves': [15, 31, 50], 'learning_rate': [0.05, 0.1, 0.2], 'n_estimators': [50, 100, 200]	2,283,010,610,939,760	27001065.95	47780860	**0.9878149154**
LGBM	Categorical Variable One-Hot Encoding / Numeric Variable Standard Scaling Dependent Variable Box-Cox Transformation	Proportion of Resident Population by Age Proportion of Working Population by Age Convenience Store Density		num_leaves': [15, 31, 50], 'learning_rate': [0.1, 0.15, 0.2], 'n_estimators': [200, 300, 400]	2,289,286,910,788,070	27355461.16	47846493	**0.987781417**

THANK YOU