## Term Paper Presentation

# Anticipating Gentrification Through Data Similarity Analysis

Sojung Noh

Urban Studies_by Prof. YoungChul Kim

Contents

# 1 Introduction

- Gentrification in its early stages activates the district ;

  - Influx of capital, increase in consumption, employment, real estate prices ..

- The anticipation of gentrification in early stage is necessary :

  - Real Estate Developers searching for potential market
  - Entrepreneurs to-be's search on blue ocean market
  - Local municipality developing/operating touristic destinations

→ Anticipating model can be an accelerating vehicle for city-makers

**마켓 스타디움, 카카오벤처스 등서 10억원 유치**

미국 상업용 부동산 지역 분석 플랫폼, 서울 성수동처럼 젠트리피케이션이 발생할 지역을 미리 예측

Market Stadium

# 2 Literature Review

Part 1. Phenomenon of the Gentrification

Part 2. Commercial Activeness Prediction

# 2 Literature Review

## Part 1. Phenomenon of the Gentrification

**Definition of Gentrification**

- The process whereby the character of a poor urban area is changed by wealthier people moving in, improving housing, attracting new business, typically displacing current inhabitants in the process (American Heritage, 1982).

**Characteristic**

- Commercial Gentrification facilitates external investments on the region. Thus, development of the amenities and services are entailed, Stabilizing the region with population decrease. (Wang, 2011)
- In the case of Seoul, the phenomenon appears as residential / industrial districts gradually transformed into commercial amenities. (Heo, 2015)
- It is evident that these street level gentrifications tend to locate in the vicinity of sub-centers of Seoul, sharing the local neighborhood facilities while enjoying relatively low real-estate cost; Yeonnam-dong near Hongdae, Seongsu-dong near Wangshimni, Samcheong-dong near Pyeongchang-dong. (Lee, 2017)

**Cause / Correlation**

- In the case study of Gyeongridan-gil, the main cause of gentrification is rent fee. Rent fee have correlation with number of cafe, distance to subway station, low gradient of the land. (Park, 2016)

# 2 Literature Review

## Part 2. Commercial Activeness Prediction

**Index development of Gentrification**

- Influx of college graduates, high-income professionals as indicator of gentrification during 1990-2000 Seoul administrative dong. (Kim, 2007)
- Influx of college graduates, high-income professionals plus land prices as indicator. (Oh, Kim, 2017)
- To sense negative effects, displacement, the change of living population, floating population, open/closing of store, operation period, count of franchise enterprises, sales as indicator. (Lee, 2019)

**Anticipating Commercial Activeness**

- In the course anticipating price movement of real-estate, machine learning methods ; LSTM, ARIMA, Random Forest with sequential data are used to predict apartment prices. (Cho et al, 2020)
- Future expectation of commercial district chance using store opening data with LSTM. (Kang, 2022)
- Anticipating commercial activeness of city with satellite image with CNN. (Zhiyan H., 2018)
- DNN has excellent in learning the weights of each parameters, but its drawback is that it is a black-box model.

# 2 Literature Review

**Part 1. Phenomenon of the Gentrification**

**Part 2. Commercial Activeness Prediction**

**Findings**

(i) Gentrification in Seoul is lead by 2030's expenditure at street level

(ii) Attempts to utilize Deep Learning Frameworks to predict district change

**Forwards**

The objective of this project is to *propose an index for an investor* to enter the market.

In order for one to plan their investment action, the *prediction must demonstrate causal factors*.

To overcome, this project proposes ***Collaborative Filtering*** method used in ***recommendation system***,

which learns the *similarity* between each entity. (Sarwar, B., etal, 2001)

# 3 Research Question

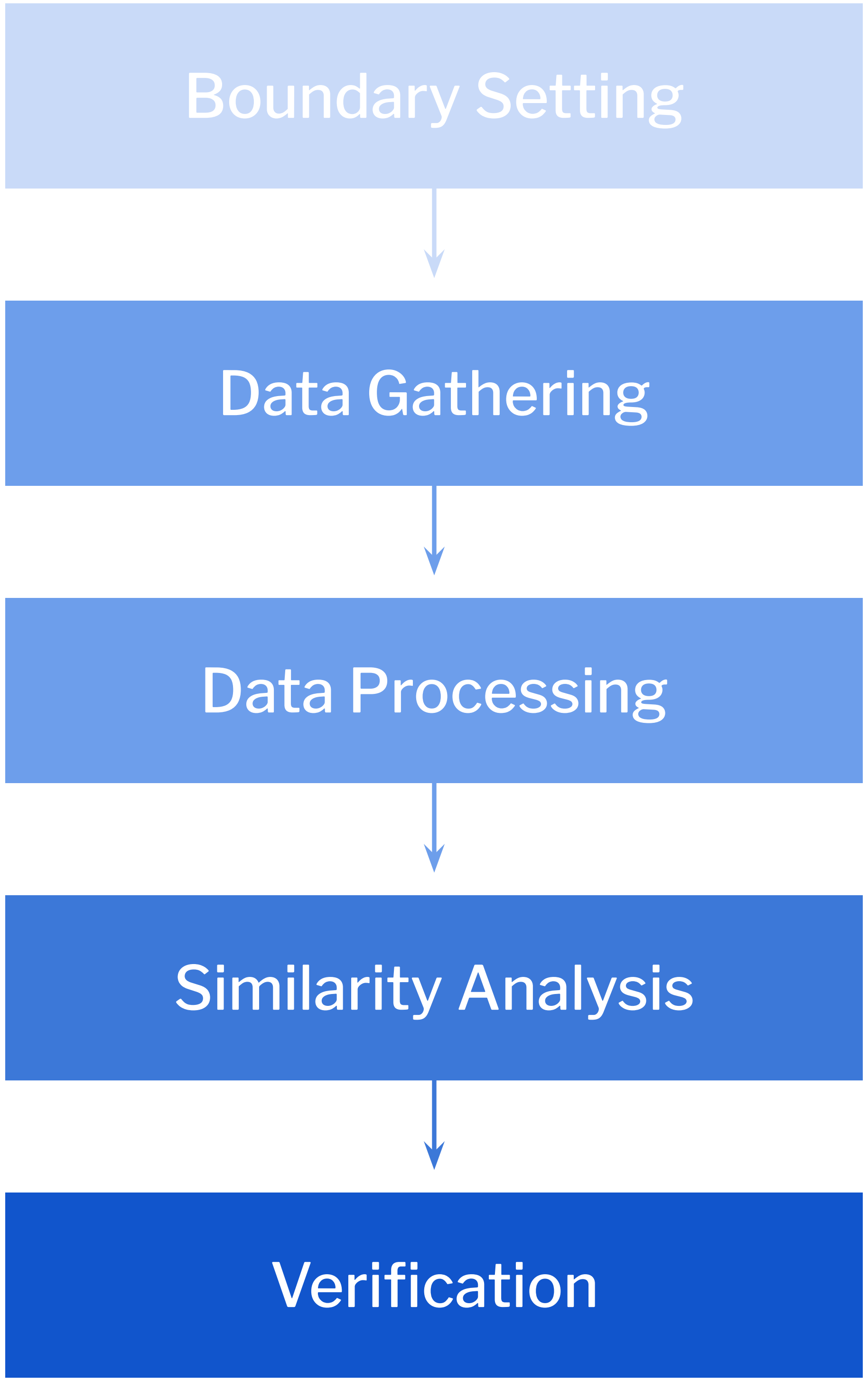
Q

In the investor(including entrepreneurs) perspective,

Is *similarity analysis* relevant methodology

for *predicting tentative commercial districts*?

# 4 Method


Boundary Setting
Data Gathering
Data Processing
Similarity Analysis
Verification

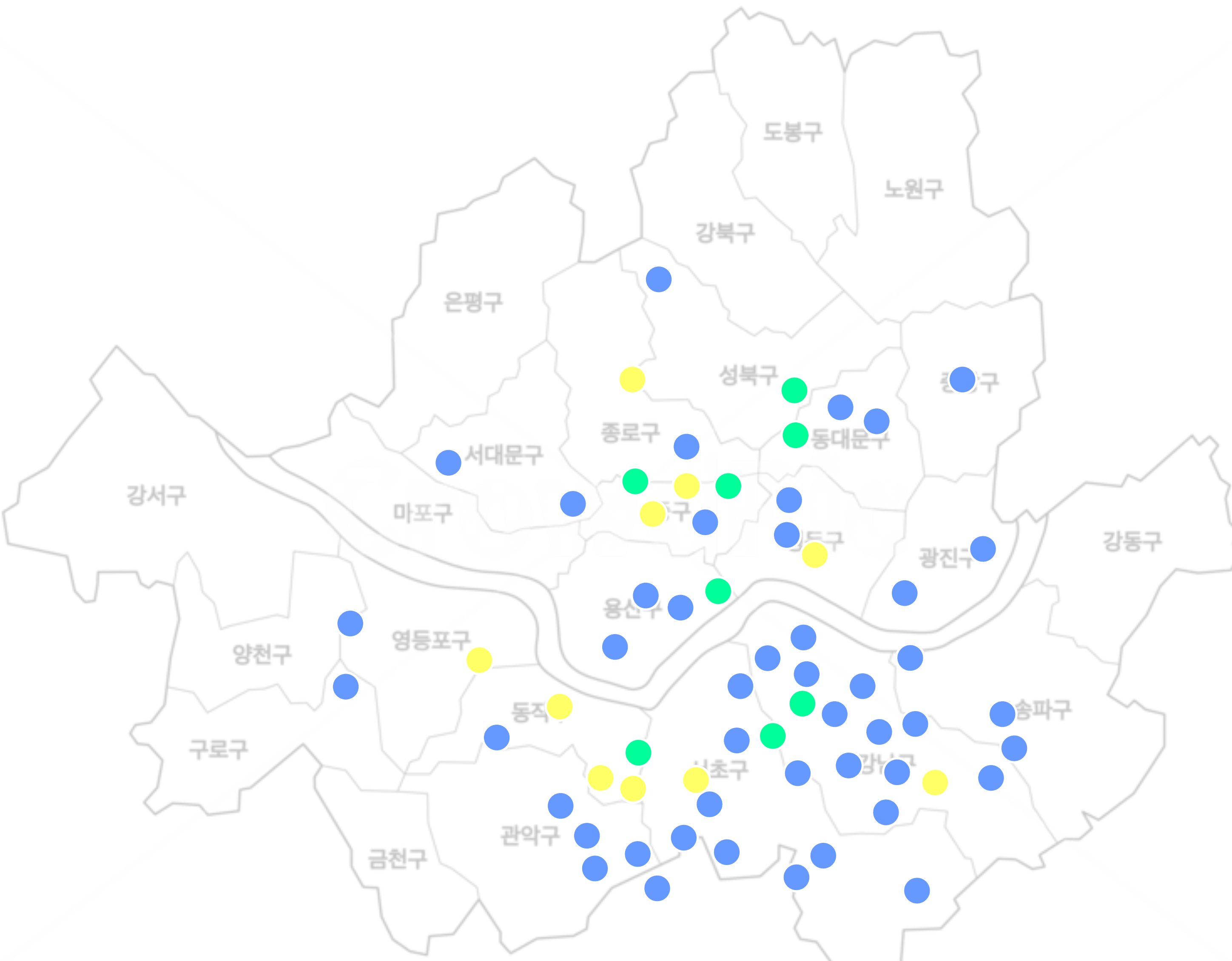
# 4 Method

Boundary Setting | Data Gathering | Data Processing | Similarity Analysis | Verification

## The Scale and Boundary of the Research

Commercial Districts in Seoul

- City of Seoul provides most up-to-date data storage on city.
- It provides datasets based on 'Commercial Streets' rather than dong/gu.




[Conceptual Diagram of the Dataset]

우리마을가게 상권분석서비스

**01 골목상권** x 1,090 EA

대로변이 아닌 거주지 안의 좁은 도로를
따라 형성되는 상업 세력 범위

□ 서울시 조사, 거주지 배후의 '길' 기준 블록 단위 데이터

**02 발달상권** x 249 EA

2000sqm 이내 50개 상점이 분포하는 상점가로,
도보이동이 가능한 범위내의 상가업소밀집지역

□ 통계청 조사, 집계구, 도로망 반영 블록 단위 데이터

**03 전통시장** x 326 EA

오랜 기간에 걸쳐 일정한 지역에서 자연발생적으로
형성된 상설시장이나 정기시장

□ 정부 조사, 상가업소DB 기반 블록 단위 데이터

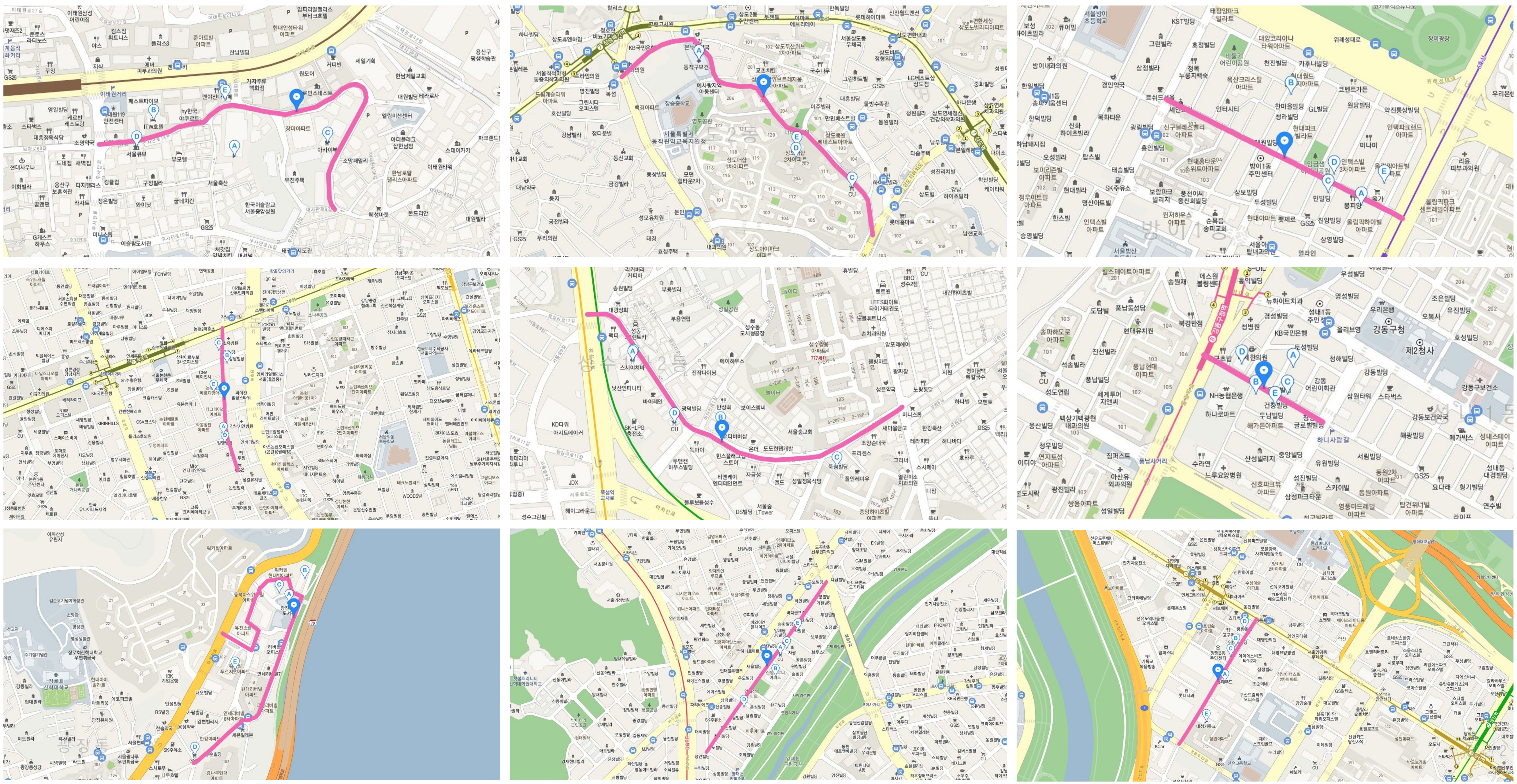
1,665 Items

# 4 Method

Boundary Setting | Data Gathering | Data Processing | Similarity Analysis | Verification

## Checking Data Shape

The data provided by Seoul City, is based on street.



## Checking Data Availability

The institution provided 10 features of the commercial streets.

The data used in the research are the following ones

| | LIST | | DATA ITEM | PERIOD | AVAIL. |
|---|---|---|---|---|---|
| 1 | 생활인구 | 상권 | 상권코드 | 2017 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 2 | 상주인구 | 상권 | 상권코드 | 2014 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 3 | 직장인구 | 상권 | 상권코드 | 2014 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 4 | 점포 | 상권 | 상권코드 | 2014 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 5 | 집객시설 | 상권 | 상권코드 | 2015 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2017 - 2021 | O |
| 6 | 아파트 | 상권 | 상권코드 | 2014 - 2021 | O |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 7 | 추정매출 | 상권 | 상권코드 | 2017 - 2021 | O |
| | | ~~상권배후지~~ | - | - | - |
| 8 | 소득소비 | ~~상권~~ | - | - | - |
| | | 상권배후지 | 상권코드 | 2014 - 2021 | O |
| 9 | 상권변화지표 | 상권 | 상권코드 | 2014 - 2021 | O |
| | | 자치구별 | 자치구코드 | 2014 - 2021 | X |
| | | 행정동별 | 행정동코드 | 2014 - 2021 | △ |
| 10 | 상권영역 | - | SHP | - | △ |
| DATASET TO BE | | | 상권코드 | 2017-2021 | |

## Preprocessing Data

Data are updated quarterly. Therefore Quarterly Average is the Yearly value.

Also unnecessary features are deleted, or standardized.

| | LIST | DATA ITEM | DATA PROCESSING |
|---|---|---|---|
| 1 | 상주인구 | 분기별 상주 인구 | 2017 – 2021, Yearly Average |
| 2 | 직장인구 | 분기별 직장 인구 | 2017 - 2021, Yearly Average |
| 3 | 집객시설 | 분기별 상업시설 개수 | 2017 - 2021, Yearly Average |
| | | 분기별 지하철 개수 | 2017 - 2021, Yearly Average |
| | | 분기별 버스정류장 개수 | 2017 - 2021, Yearly Average |
| 4 | 추정매출 | 분기별 전체 매출 | 2017 - 2021, Yearly Average |
| | | 분기별 20/30/40 매출 | 2017 - 2021, Yearly Average |
| 5 | 상권변화지표 | 분기별 {영업기간 / 서울평균영업기간} | 2017 - 2021, Yearly Average |
| | | 분기별 {폐업기간 / 서울평균폐업기간} | 2017 - 2021, Yearly Average |

# 4 Method

Boundary Setting | Data Gathering | **Data Processing** | Similarity Analysis | Verification

## Final Output of the *Commercial District x Feature* Data

| | Unnamed: 0 | T_remainpop | T_commuterpop | commercial_facility_count | subwayst_count | busst_count | 분기당_매출_금액 | 2340대_매출금액 | 영업기간/서울평균 | 폐업기간/서울평균 |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1000001 | 1919.0 | 809.0 | 14.0 | 0.0 | 4.0 | 299.028533 | 0.071191 | 0.92 | 1.11 |
| **1** | 1000003 | 1150.0 | 1079.0 | 23.0 | 0.0 | 3.0 | 264.983656 | 0.043465 | 1.07 | 1.04 |
| **2** | 1000004 | 1497.0 | 20.0 | 10.0 | 0.0 | 5.0 | 168.833421 | 0.039035 | 0.84 | 1.10 |
| **3** | 1000005 | 1772.0 | 119.0 | 6.0 | 0.0 | 3.0 | 419.180139 | 0.088284 | 0.83 | 0.96 |
| **4** | 1000006 | 682.0 | 18.0 | 8.0 | 0.0 | 2.0 | 627.819759 | 0.135100 | 0.91 | 1.21 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **1205** | 1001492 | 1299.0 | 168933.0 | 402.0 | 2.0 | 27.0 | 14409.481200 | 2.985283 | 1.38 | 1.14 |
| **1206** | 1001493 | 2509.0 | 9790.0 | 118.0 | 3.0 | 14.0 | 2325.866252 | 0.594730 | 1.22 | 1.03 |
| **1207** | 1001494 | 2750.0 | 30331.0 | 241.0 | 5.0 | 34.0 | 3807.258200 | 0.881768 | 1.45 | 1.25 |
| **1208** | 1001495 | 8977.0 | 27304.0 | 168.0 | 0.0 | 20.0 | 4789.094944 | 1.023787 | 0.85 | 1.03 |
| **1209** | 1001496 | 19.0 | 22217.0 | 39.0 | 1.0 | 8.0 | 4653.147944 | 1.053172 | 0.83 | 1.21 |

Commercial District Code

[계동길, 난계로2길, 돈화문로11가길, 명륜길, 백석동길 — 종로청계관광특구, 잠실관광특구, 강남마이스관광특구]

# 4 Method



## Recommendation System > Collaborative Filtering > Similarity Analysis

This experiment is based upon *Recommendation System, Collaborative Filtering* framework. The objective of this framework is *predicting the rating* of a user on specific item by obtaining neighbors by *similarity analysis.*

| item / user | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 290 | 3 | | | | | 4 | | 2 | |
| 291 | | 4 | | 4 | 4 | | | 4 | 4 |
| 292 | | | | 3 | | | | | |
| 293 | 4 | | 3 | | 4 | 4 | 3 | 2 | |
| 294 | | | | | | | | | |
| 295 | | | 5 | | 5 | 5 | 4 | 5 | |
| 296 | 4 | | | | | | | | |
| 297 | 4 | | 3 | | 2 | 4 | | 3 | |
| 298 | 5 | | 3 | | 5 | | | | |
| 299 | 4 | 4 | 5 | | | 5 | | | |

**Predict Blank, of item 215**

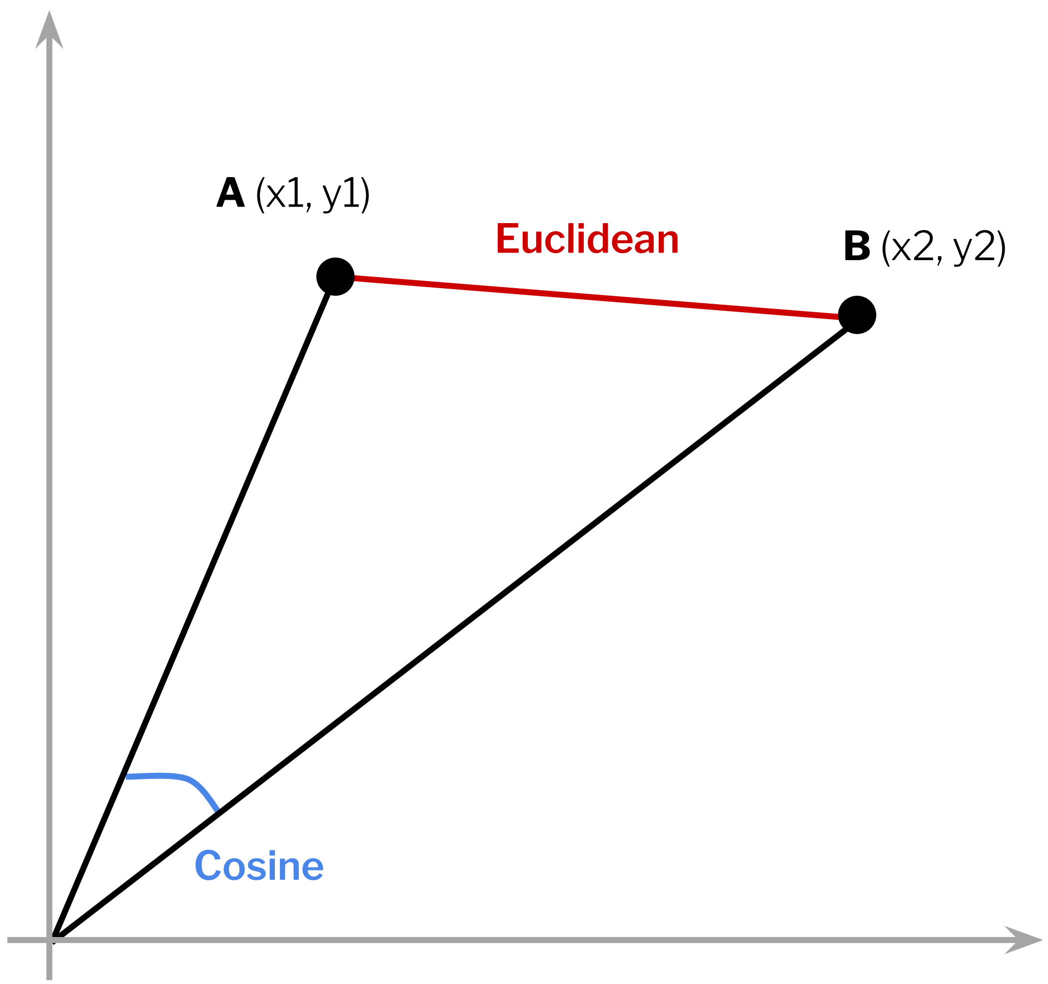
1. Find K neighbors
   a. Mean squared Difference Similarity
   b. Euclidean Similarity
   c. Cosine Similarity
   d. Pearson Similarity
   e. Pearson-Baseline Similarity
2. Average the neighbors' ratings on item

(Sarwar, B., et al, 2001)

## Cosine Similarity

Similarity between two commercial districts is evaluated through *Cosine similarity*.

*Cosine similarity* is implemented in *sklearn package*.




**Example :**

**Distance Between [6.6, 6.2] ←→ [9.7, 9.9]**

**Euclidean Distance = 4.82**

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

where $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n)$

**Cosine Similarity = 0.99**

$$sim(x,y) = \frac{\langle x,y\rangle}{\|x\|\|y\|}$$

where $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n)$

$$\|x\| = \sqrt{\sum_{i=1}^{n}(x_i)^2}\,,\langle x,y\rangle = \sum_{i=1}^{n} x_i y_i$$

## Identifying Base Commercial Districts

Filter 1) Among the commercial districts with 203040 sales bigger than 0.10,

Filter 2) Among the commercial districts with 2017-2020 growth rate bigger than 10%

| Type | Neighbor_name | Neighbor_code | 203040_Sales | Growth_rate |
|---|---|---|---|---|
| Type I | 아차산로 15길 (성수동 북측) | 1000114 | 0.107 | 82.4% |
| Type II | 도봉로 114길 (쌍문역) | 10000360 | 0.147 | 33.4% |
| Type III | 녹사평대로 32길 (이태원 서측) | 10000052 | 0.143 | 16.0% |
| Type IV | 동교로 38길 (연남동) | 1000470 | 0.260 | 10.5% |

Filter 1
203040_Sales > 0.1

Sales magnitude
bigger than **'Myeongdong'**,
Since Myeongdong is
downturned traditional
commercial street with
*COVID19(1), China Issue(2)*

Filter 2
Growth_rate > 10%

10% may seem small,
but considering the
lockdown during COVID19,
major commercial streets
suffered negative growth,
**10% is still significant value.**

## Cosine Similarity

Similarity between two commercial districts is evaluated through *Cosine similarity*.

*Cosine similarity* is implemented in *sklearn package*.

**2017 DATA OF**
**2021 CURRENT HOTPLACE**

| |
|---|
| 1000114<br>아차산로15길 |
| 1000360<br>도봉로114길 |
| 1000052<br>녹사평대로32길 |
| 1000470<br>동교로38길 |

Selection based on Top 5
Districts with sales from
20/30/40's expenditure

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

**(i) VERIFICATION** (2018–2020 DATA) | **(ii) PREDICTION** (2021 DATA)

| 2018 DATA | 2019 DATA | 2020 DATA | 2021 DATA |
|---|---|---|---|
| 1000001 | 1000001 | 1000001 | 1000001 |
| 1000002 | 1000002 | 1000002 | 1000002 |
| | | | |
| | | | |
| | | | |

| Top K Similar | Top K Similar | Top K Similar | Top K Similar |
|---|---|---|---|
| 아차산로15길 | 충정로4길 | 아차산로11길 | 아차산로11길 |
| 아차산로11길 | 삼일대로4길 | 아차산로15길 | 논현로26길 |
| 성수일로6길 | 아차산로11길 | 논현로26길 | 종로24길 |
| 성수이로18길 | 아차산로15길 | 종로24길 | 논현로28길 |
| 논현로26길 | 충정로6길 | 논현로28길 | 성수이로18길 |
| 경인로80길 | 경인로80길 | 성수이로18길 | 명동길 |
| 충정로6길 | 종로24길 | 명동길 | 한강대로52길 |
| 남부순환로339길 | 성수이로18길 | 한강대로52길 | 경인로80길 |
| 율곡로10길 | 논현로28길 | 경인로80길 | 서초중앙로8길 |
| 강남대로23길 | 당산로37길 | 서초중앙로8길 | 종암로19길 |

## Cosine Similarity

Similarity between two commercial districts is evaluated through *Cosine similarity*.

*Cosine similarity* is implemented in *sklearn package*.

**2017 DATA OF 2021 CURRENT HOTPLACE**

| |
|---|
| 1000114 아차산로15길 |
| **1000360 도봉로114길** |
| 1000052 녹사평대로32길 |
| 1000470 동교로38길 |

Selection based on Top 5 Districts with sales from 20/30/40's expenditure

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

**(i) VERIFICATION** | **(ii) PREDICTION**

| 2018 DATA | 2019 DATA | 2020 DATA | 2021 DATA |
|---|---|---|---|
| 1000001 | 1000001 | 1000001 | 1000001 |
| 1000002 | 1000002 | 1000002 | 1000002 |
| | | | |
| | | | |
| | | | |
| **Top K Similar** | **Top K Similar** | **Top K Similar** | **Top K Similar** |
| 왕십리로14길 | 논현로27길 | 당산로31길 | 원효로89길 |
| 아차산로78길 | 사임당로17길 | 봉은사로29길 | 상원길 |
| 영동대로65길 | 시흥대로63길 | 남부순환로317길 | 남부순환로317길 |
| 양평로19길 | 당산로31길 | 상원길 | 당산로31길 |
| 장한로25길 | 한강대로43길 | 원효로89길 | 양재대로71길 |
| 양화로1길 | 양재대로71길 | 학동로38길 | 학동로38길 |
| 방배로35길 | 중앙로1길 | 아차산로5길 | 아차산로5길 |
| 창경궁로35길 | 동교로25길 | 논현로27길 | 올림픽로48길 |
| 동교로27길 | 장승배기로10길 | 양재대로71길 | 우사단로14길 |
| 사임당로17길 | 한강대로62길 | 한강대로62길 | 효령로31길 |

# 4 Method

Boundary Setting | Data Gathering | Data Processing | **Similarity Analysis** | Verification

## Cosine Similarity

Similarity between two commercial districts is evaluated through *Cosine similarity*.

*Cosine similarity* is implemented in *sklearn package*.

**2017 DATA OF 2021 CURRENT HOTPLACE**

- 1000114 아차산로15길
- 1000360 도봉로114길
- **1000052 녹사평대로32길**
- 1000470 동교로38길

Selection based on Top 5 Districts with sales from 20/30/40's expenditure

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

**(i) VERIFICATION** | **(ii) PREDICTION**

| 2018 DATA | 2019 DATA | 2020 DATA | 2021 DATA |
|---|---|---|---|
| 1000001 | 1000001 | 1000001 | 1000001 |
| 1000002 | 1000002 | 1000002 | 1000002 |
| | | | |
| | | | |
| | | | |
| **Top K Similar** | **Top K Similar** | **Top K Similar** | **Top K Similar** |
| 북촌로5길 | 디지털로74길 | 개포로82길 | 개포로82길 |
| 디지털로74길 | 북촌로5길 | 디지털로74길 | 녹사평대로32길 |
| 이태원로27길 | 흑석로13길 | 북촌로5길 | 디지털로74길 |
| 와우산로29가길 | 녹사평대로32길 | 이태원로27길 | 북촌로5길 |
| 녹사평대로32길 | 이태원로27길 | 녹사평대로32길 | 이태원로 27길 |
| 사평대로26길 | 청파로47길 | 동교로38길 | 자하문로7길 |
| 인촌로24길 | 천호대로12길 | 동소문로6길 | 한남대로20길 |
| 테헤란로81길 | 동소문로6길 | 이태원로54길 | 동교로38길 |
| 마포대로12길 | 녹사평대로40나길 | 한남대로20길 | 이태원로54길 |
| 녹사평대로40나길 | 인촌로24길 | 청파로47길 | 마포대로12길 |

# 4 Method

Boundary Setting | Data Gathering | Data Processing | **Similarity Analysis** | Verification

## Cosine Similarity

Similarity between two commercial districts is evaluated through <u>*Cosine similarity*</u>.

<u>*Cosine similarity*</u> is implemented in *sklearn package*.

**2017 DATA OF**
**2021 CURRENT HOTPLACE**

1000114 아차산로15길

1000360 도봉로114길

1000052 녹사평대로32길

**1000470 동교로38길**

Selection based on Top 5 Districts with sales from 20/30/40's expenditure

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

**(i) VERIFICATION** | **(ii) PREDICTION**

| 2018 DATA | 2019 DATA | 2020 DATA | 2021 DATA |
|---|---|---|---|
| 1000001 | 1000001 | 1000001 | 1000001 |
| 1000002 | 1000002 | 1000002 | 1000002 |
| | | | |
| | | | |
| | | | |
| Top K Similar | Top K Similar | Top K Similar | Top K Similar |
| 상암로51길 | 삼성로115길 | 월드컵로11길 | 월드컵로11길 |
| 남현3길 | 목동동로12길 | 조원로10길 | 양평로19길 |
| 망우로18길 | 오패산로4길 | 양평로19길 | 남부순환로151길 |
| 선릉로18길 | 화곡로4길 | 아차산로51길 | 화곡로4길 |
| 상동대로53길 | 성균관로5길 | 공항대로61길 | 마포대로20길 |
| 아차산로65길 | 동일로192길 | 영등포로35길 | 연서로5길 |
| 용마산로118길 | 세검정로1길 | 화곡로4길 | 남현3길 |
| 자양로45길 | 천호대로109길 | 천호대로109길 | 삼양로8길 |
| 서울시립대로4길 | 백제고분로21길 | 남부순환로151길 | 영등포로35길 |
| 연서로29길 | 강동대로53길 | 삼양로8길 | 답십리로69길 |

## Filtering Similar Items

Similarity : Similarity values imply the ‘distance’ between two vectors, in this case commercial districts.

**Results per year**

Similar 2018
이태원로54길
사평대로22길
디지털로32길
. . .

Similar 2019
동교로38길
사평대로22길
논현로159길
. . .

Similar 2020
사평대로22길
인촌로1길
동광로39길
. . .

Similar 2021
사평대로22길
동광로39길
인촌로1길
. . .

Sim > 0.96

**First Filter**

이태원로54길
사평대로22길
디지털로32길
삼청로5길
청파로5길
공항대로38길
양화로11길
자하문로7길
원효로89길
보문로32길
동교로38길
사평대로22길
논현로159길
이태원로54길
자하문로7길
도산대로15길
송파대로30길
북촌로5나길
망우로21길
디지털로32길
. . .

More than 2yrs in Top10

**Second Filter**

성수이로18길
아차산로11길
경인로80길
논현로26길
아차산로15길
충정로6길
논현로28길
종로24길

If they show *positive correlation* with respect to sales levels,

It can be said that the *predicted result are plausible.*

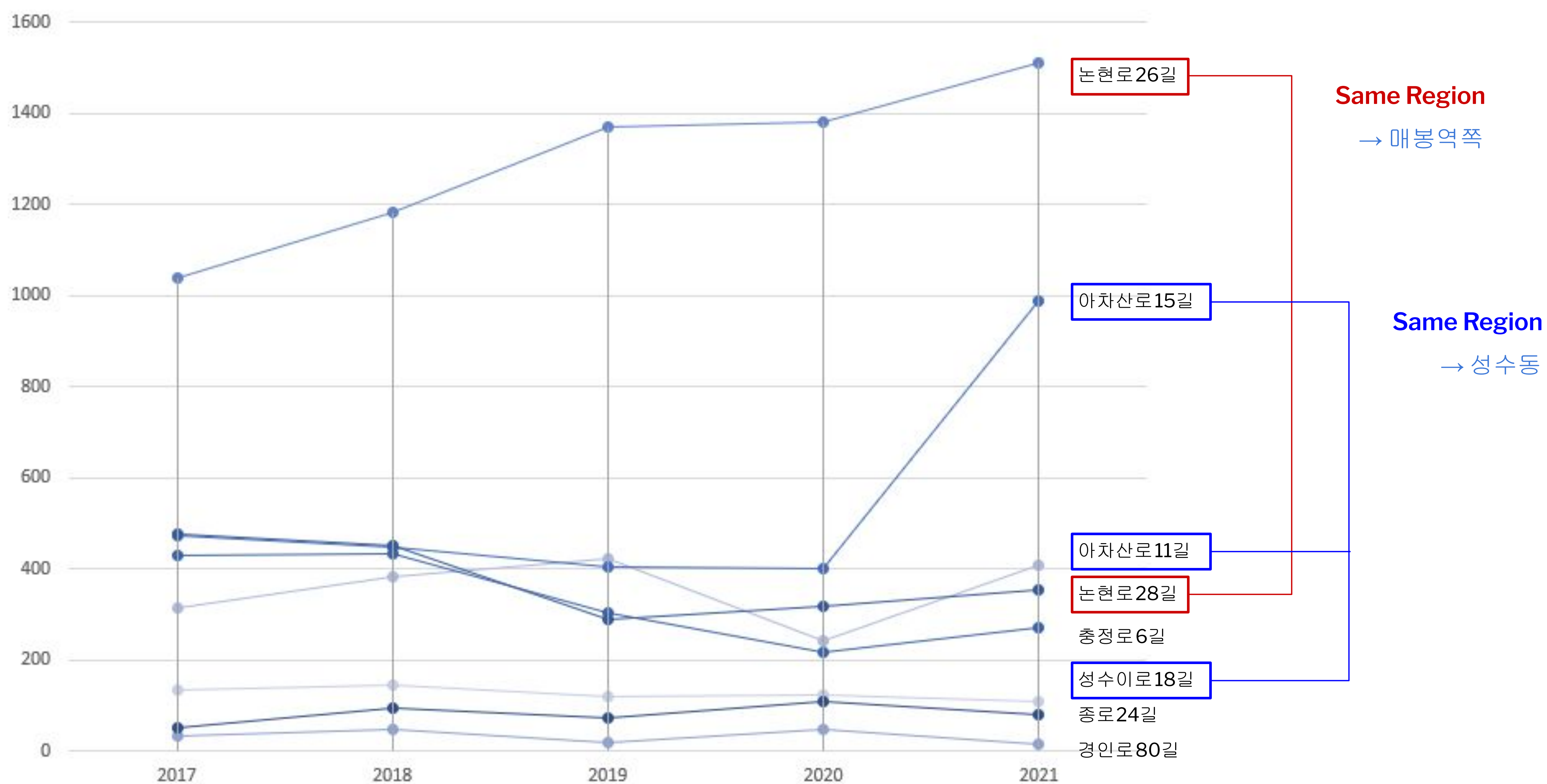
# Prediction Result

| Type | Neighbor_name | Neighbor_code |
|---|---|---|
| Type I | 아차산로 15길 (성수동 북측) | 논현로26길 / 아차산로15길 / 아차산로11길 / 논현로28길 / 충정로6길 / 성수이로18길 / 종로24길 / 경인로80길 |
| Type II | 도봉로 114길 (쌍문역) | 이태원로54길 / 청파로47길 / 동교로38길 / 자하문로7길 / 녹사평대로32길 / 성지3길 / 북촌로5나길 / 인촌로24길 / 와우산로29길 |
| Type III | 녹사평대로 32길 (이태원 서측) | 상도로62길 / 신흥로20길 / 서오릉로8길 / 청룡길 / 상도로61길 / 와우산로3길 |
| Type IV | 동교로 38길 (연남동) | 강동대로52길 / 남현3길 / 경인로80길 / 화곡로4길 / 상암로51길 / 천호대로109길 |

| Type | Neighbor_name | 203040_Sales | Growth_rate |
|---|---|---|---|
| Type I | 아차산로 15길 (성수동 북측) | 0.107 | 82.4% |
| Type II | 도봉로 114길 (쌍문역) | 0.147 | 33.4% |
| Type III | 녹사평대로 32길 (이태원 서측) | 0.143 | 16.0% |
| Type IV | 동교로 38길 (연남동) | 0.260 | 10.5% |

## Sales Plot(1)

Exhibits similar regions, with fairly inclining growth in 203040 sales.



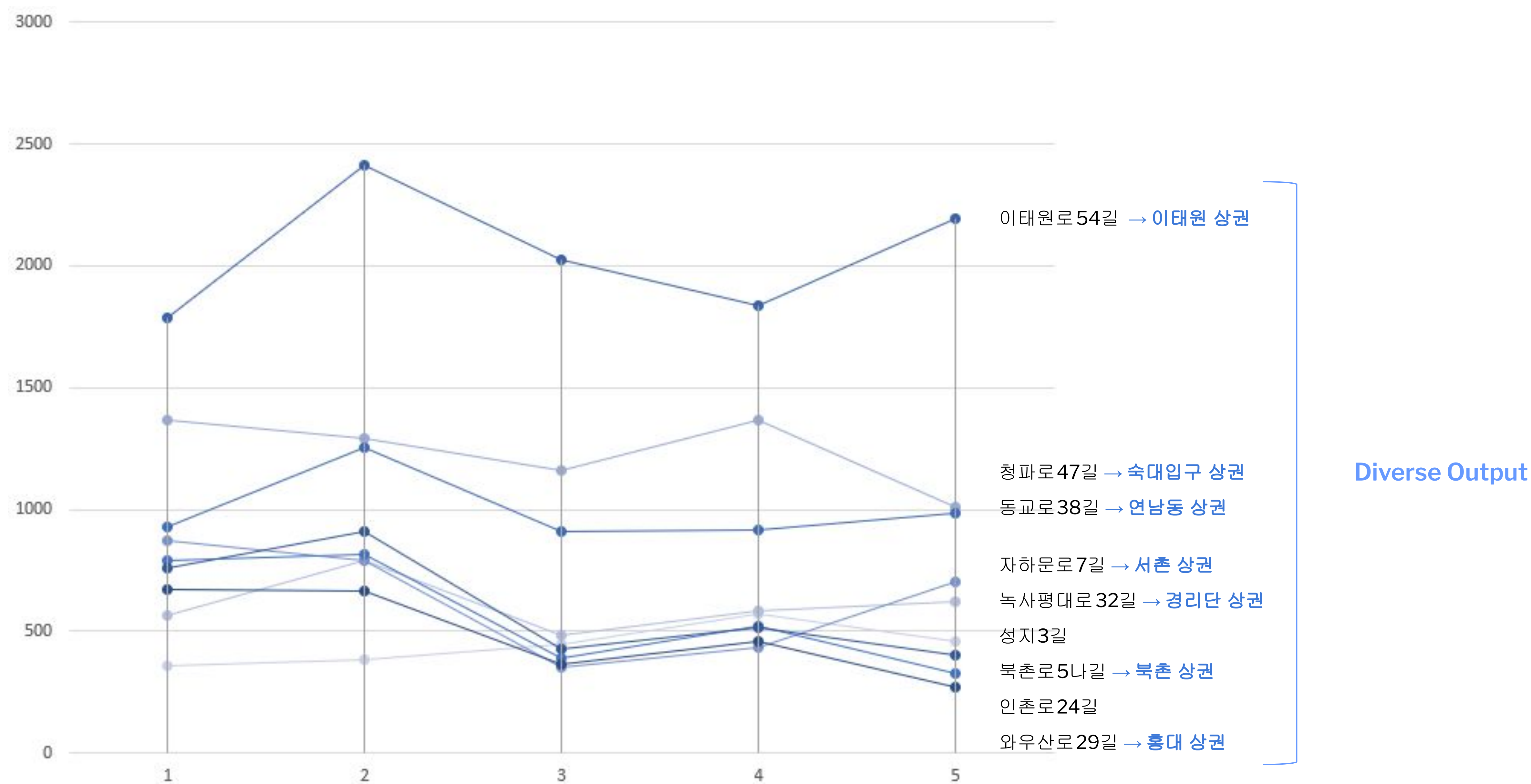
# 4 Method

Boundary Setting | Data Gathering | Data Processing | Similarity Analysis | **Verification**

## Sales Plot(2)

Exhibits fairly distinct type of commercial streets. Sign of incline is not clear

| Type | Neighbor_name | 203040_Sales | Growth_rate |
|---|---|---|---|
| Type I | 아차산로 15길 (성수동 북측) | 0.107 | 82.4% |
| Type II | 도봉로 114길 (쌍문역) | 0.147 | 33.4% |
| Type III | 녹사평대로 32길 (이태원 서측) | 0.143 | 16.0% |
| Type IV | 동교로 38길 (연남동) | 0.260 | 10.5% |



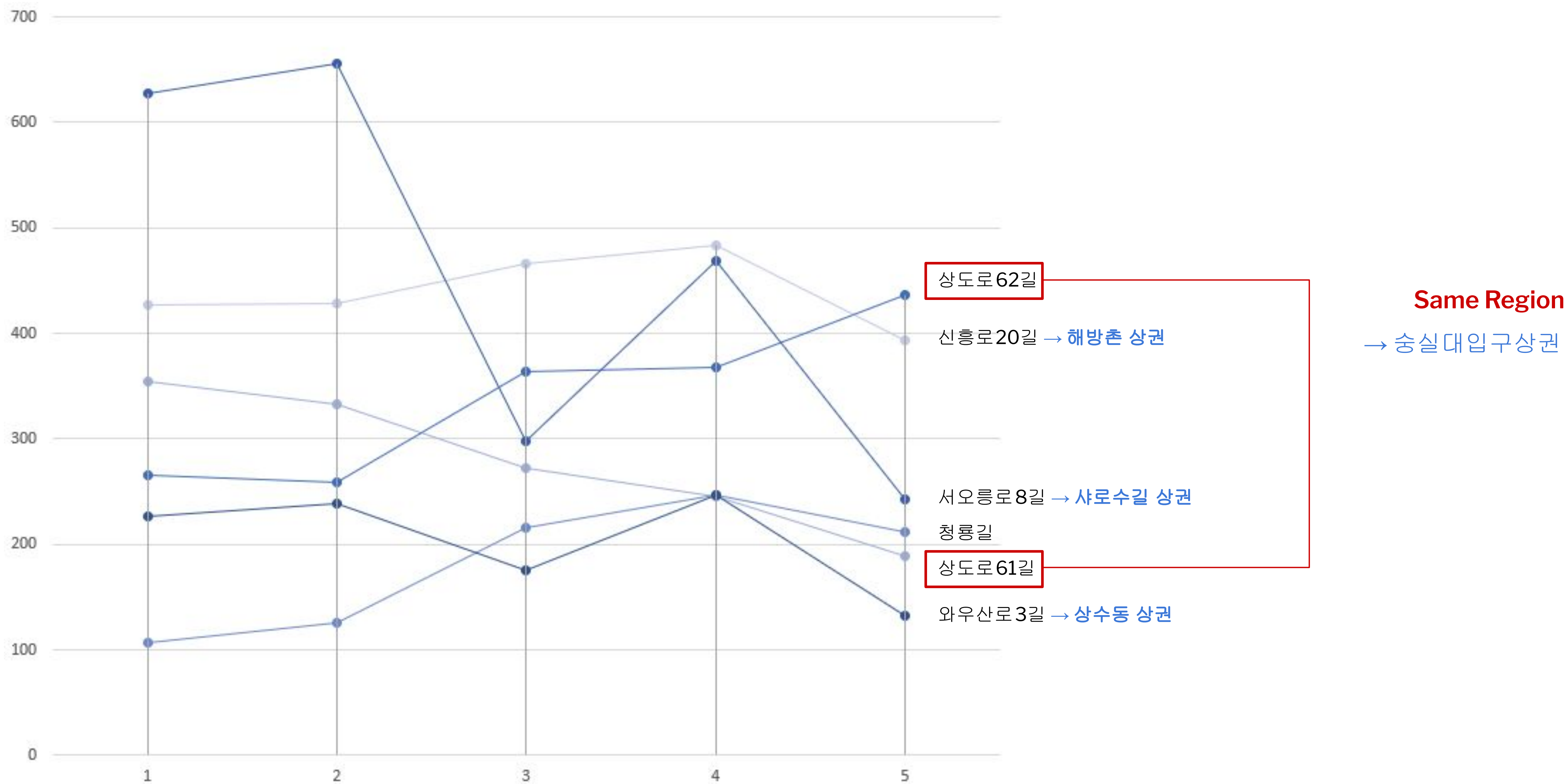
# 4 Method

Boundary Setting | Data Gathering | Data Processing | Similarity Analysis | **Verification**

## Sales Plot(3)

Declining outputs of regions. Still Diverse Output.

| Type | Neighbor_name | 203040_Sales | Growth_rate |
| --- | --- | --- | --- |
| Type I | 아차산로 15길 (성수동 북측) | 0.107 | 82.4% |
| Type II | 도봉로 114길 (쌍문역) | 0.147 | 33.4% |
| Type III | 녹사평대로 32길 (이태원 서측) | 0.143 | 16.0% |
| Type IV | 동교로 38길 (연남동) | 0.260 | 10.5% |



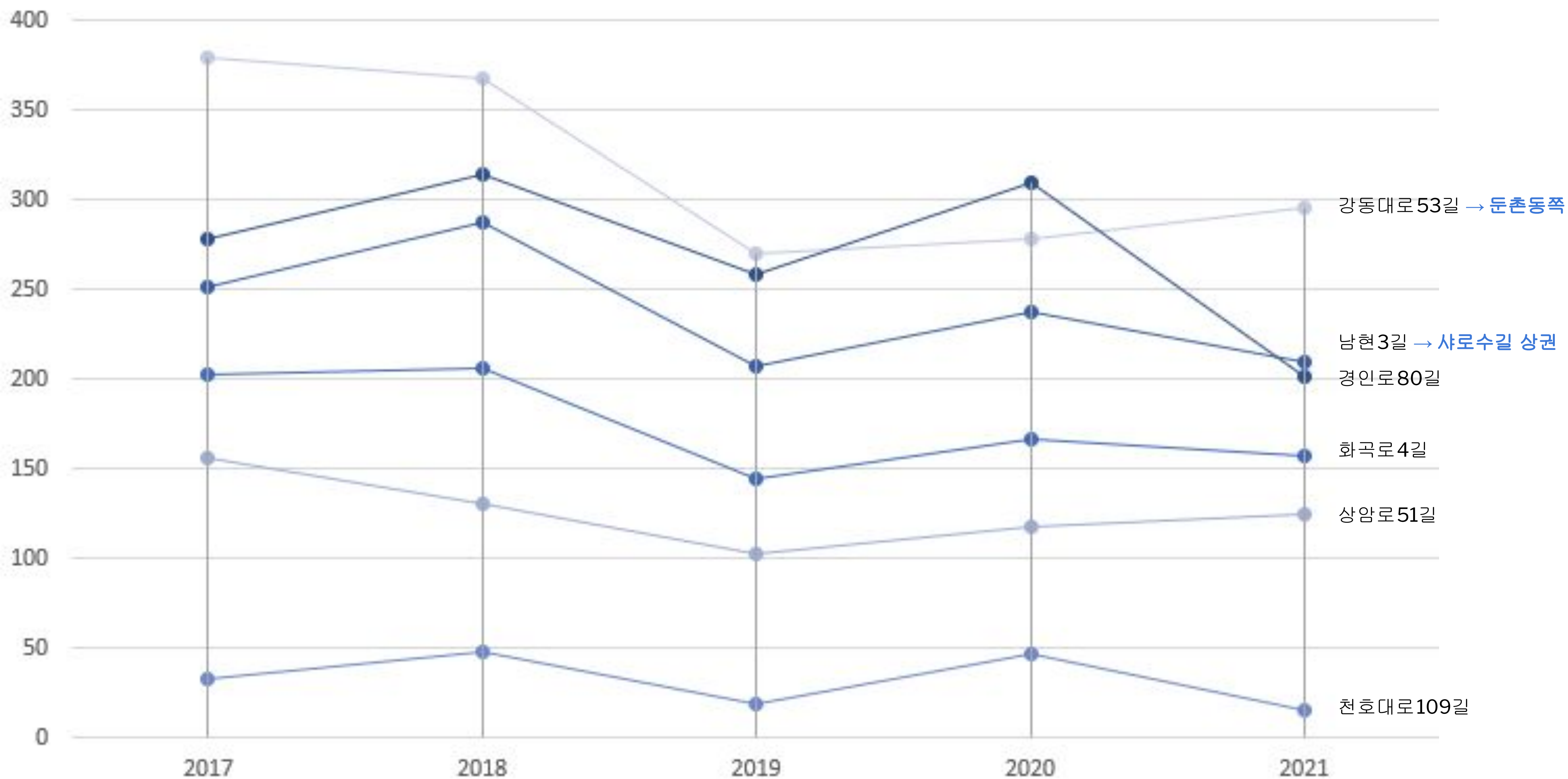
# 4 Method

Boundary Setting | Data Gathering | Data Processing | Similarity Analysis | **Verification**

## Sales Plot(4)

203040

| Type | Neighbor_name | 203040_Sales | Growth_rate |
|---|---|---|---|
| Type I | 아차산로 15길 (성수동 북측) | 0.107 | 82.4% |
| Type II | 도봉로 114길 (쌍문역) | 0.147 | 33.4% |
| Type III | 녹사평대로 32길 (이태원 서측) | 0.143 | 16.0% |
| Type IV | 동교로 38길 (연남동) | 0.260 | 10.5% |



# 5 Conclusion

Q

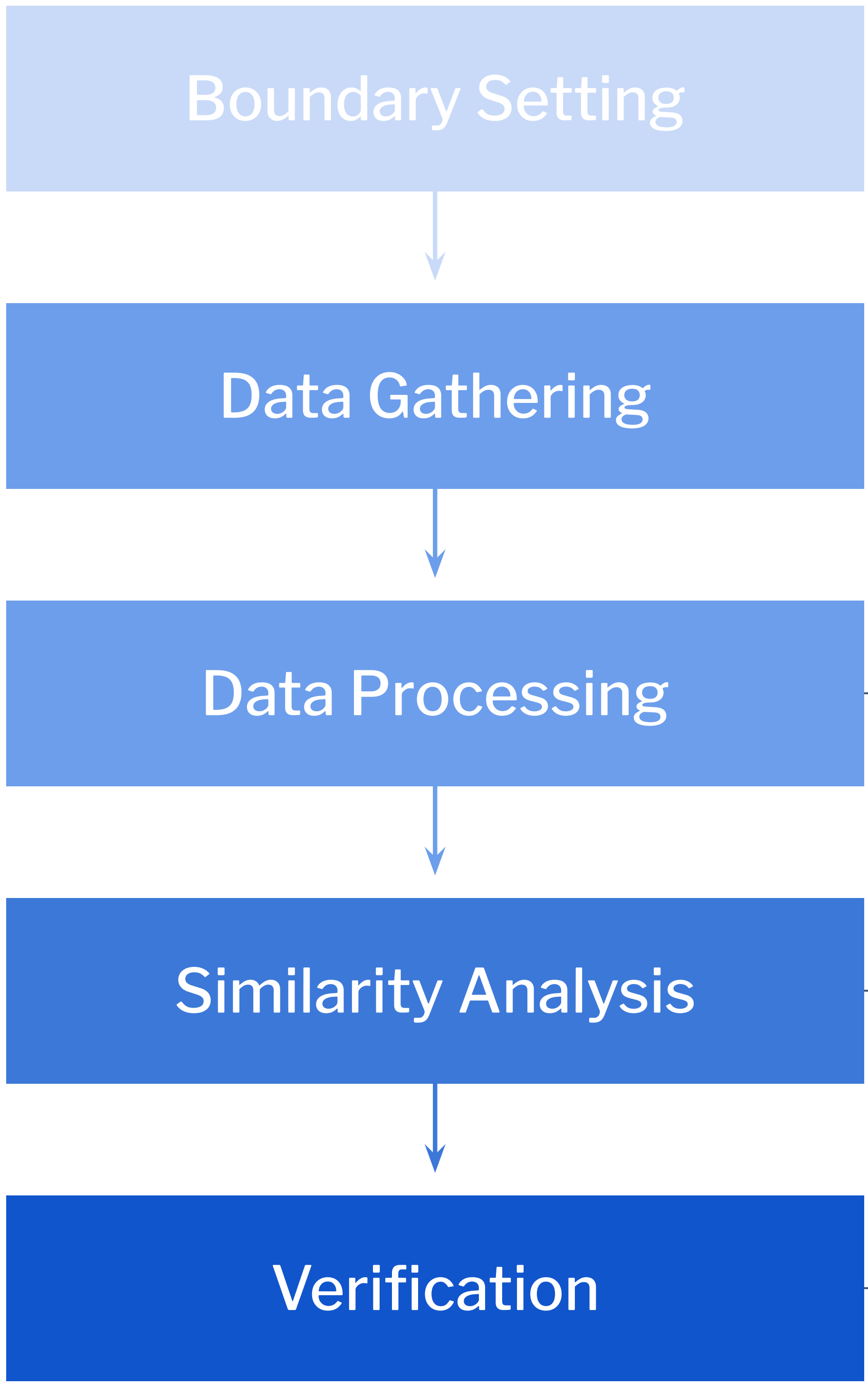
In the investor(including entrepreneurs) perspective,
Is similarity analysis relevant methodology for predicting tentative commercial districts?

a

→ With the similarity analysis that this paper suggests, it does show seemingly positive correlations, but it needs refining to get absolute result. Which will be explained next page.

# 5 Further Improvements

Method




- **Preprocessing of the Data**

  **(i) Per meter data**

  The data is based on 'Streets' with different lengths.

  It needs to be calculated per meter.

  **(ii) Sequential data format**

  The data input shouldn't be single time period, but a sequential data.

  i.e. the delta value of all the years or quarters it entails.

- **Refined categorization of base commercial streets**

  Make sure to select *'rising market'* model as baseline.

  In order to deduce meaningful result, base cases should be carefully selected.

  Some sort of *clustering analysis* needs to be done.

  Or a *qualitative approach* on categorization.

  ※ Data driven results should match with cognitive concepts in this investigation!!

- **Considering COVID19, alternative proving method is needed.**

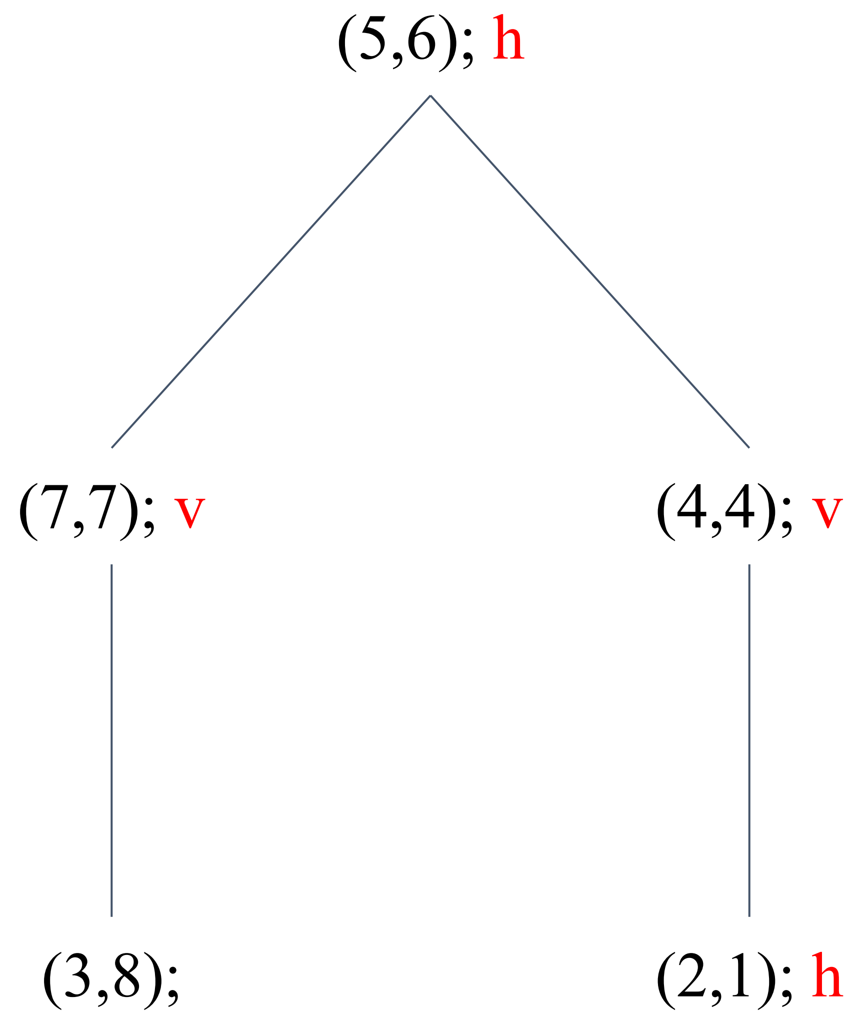
  The '203040 sales plot' here, I intended positive correlation as a proof.

  Since the data is distorted due to COVID19,

  Another metric is needed as a proof, i.e. SNS mention counts, Population … etc

😃

Thank you.


(5,6); h
(7,7); v
(4,4); v
(3,8);
(2,1); h