

# Homework 1

## Big Data and Knowledge Management System

2022-24790 Sungwoo PARK  
Graduate School of Data Science

March 28, 2022

## 1 Database Normalization

### Used Data

The COVID-19, one of the most severe pandemics in the human history, has brought a lot of change to our lives and is still not eradicated. The biggest one is that it has kept people from gathering offline because of its contagiousness and caused huge damage to countless self-employed workers, especially who provide face-to-face services to consumers, such as restaurant owners.

So I wanted to see the relationship between the pandemic and sales of self-employed workers, especially who works at commercial district, and see how big the impact was. Thus, I used the "골목상권 분석자료" provided by the Seoul open data portal(<https://data.seoul.go.kr/>). The dataset contains not only the basic information such as the location of every commercial district in Seoul, but also the floating population in each area or quarterly estimated sales. Below is the list of files I used for this assignment.

- 서울시 우리마을가게 상권분석서비스(상권-상주인구)
- 서울시 우리마을가게 상권분석서비스(상권-생활인구)
- 서울시 우리마을가게 상권분석서비스(상권-상권변화지표)
- 서울시 우리마을가게 상권분석서비스(상권-직장인구)

- 서울시 우리마을가게 상권분석서비스(상권-추정매출)
- 서울시 우리마을가게 상권분석서비스(상권영역)
- 서울특별시 코로나19 자치구별 확진자 발생동향

Moreover, I constructed the data containing information of the district, a unit of local administration in Seoul, to get and use proper identifier of each. Finally, I also made a table named "Calendar" which contains information of each date from 1/1/2010 to 12/31/2022, such as to which quarter each date belongs or whether the date is holiday or not.

## Data Normalization

First, before creating tables, I used `Python` and `Pandas` to preprocess the data. Using the given package, I re-arranged the structure of COVID-19 confirmed cases data. Specifically, in the original data the confirmed cases of each district is arranged column-by-column. In other words, each column denotes the cases confirmed in each district. However, I re-arranged it row-by-row, consolidating all districts in one column, and the daily confirmed cases in each district in another. Therefore, the columns of the final data are: `date`, `district`, and `cases`.

For normalization, I created a table named `md_info`, which contains the basic information of each commercial district, such as code as the identifier, its name, and which district it belongs to. Then I made every table with the code of the commercial districts references this `md_info` table. Next, I also created a table containing information of units of local administration, such as the table `district_info` to satisfy the 3NF, and I made the `md_info` table reference it. Furthermore, for the `change_info` table, the index that indicates the status of each commercial district determines the specific name of each status. Therefore, to remove transitive functional dependency, I separated the index and name of each status, and made a table named `change_codes`. In the same way, I separated the class to which each commercial district is classified(`md_class_code`) and the name of each class(`md_class_name`), then created a table named `class_info`. Finally, both sales tables, `gender_sales` and `weekday_sales`, contain the code and name of types of business, `dept_code` and `dept_name`. Therefore, separated them and created a table with the name `dept`.

The ERD(Entity Relationship Diagram) created after normalization is like below. The diagram shows the functional dependency among tables. Solid lines are generated when one of the primary keys of a table is set as a foreign key. For example, the column `dept_code` in the table `gender_sales` references the same column in the table `dept`. Since the `dept_code` is

one of the primary keys in the `gender_sales`, the solid line is generated between the tables. On the other hand, dotted lines are generated when non-primary key is set as a foreign key. For example, the table `md_info` references the table `district_info`, while the foreign key, `district_code`, is not included in the primary key of `md_info`, thus the dotted line is generated between two tables.

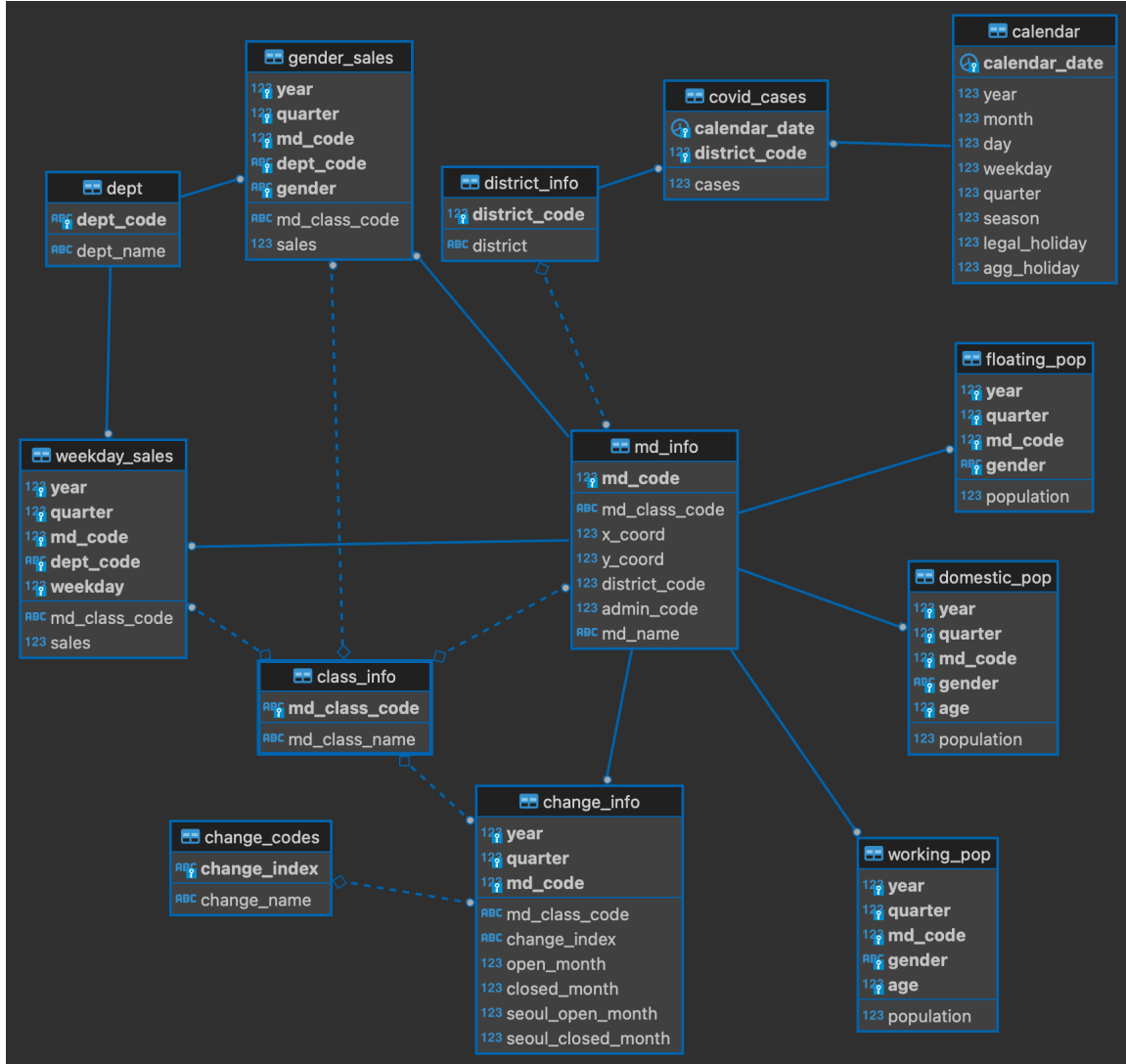


Figure 1: ERD plotted using DBeaver

A simple description of each table is like below.

- `md_info` : A table containing the basic information of each commercial district.
  - Primary Key: `md_code`, the code given to each commercial district.

- `district_info` : A table containing the basic information of each district(or borough) in Seoul.
  - Primary Key: `district_code`, the code given to each district.
- `covid_cases` : A table containing the number of daily COVID-19 confirmed cases in each district.
  - Primary Key: `calendar_date`, `district_code`
- `calendar`
  - Primary Key: `calendar_date`
- `working_pop` : A table of the number of population who works around each commercial district.
  - Primary Key: `year`, `quarter`, `md_code`, `gender`, `age`
- `domestic_pop` : A table of the number of population who resides around each commercial district.
  - Primary Key: `year`, `quarter`, `md_code`, `gender`, `age`
- `floating_pop` : A table of the number of population who floats around each commercial district.
  - Primary Key: `year`, `quarter`, `md_code`, `gender`
- `change_info` : A table containing information the status change of each commercial district.
  - Primary Key: `year`, `quarter`, `md_code`
- `change_codes` : A table containing the status index and the status name.
  - Primary Key: `change_index` : The status index of each commercial district.
- `class_info` : A table containing the class of each commercial district: {골목상권, 발달상권, 전통시장, 관광특구}.
  - Primary Key: `md_class_code`
- `dept` : A table containing the code and name of each type of business.

- Primary Key: dept\_code
- weekday\_sales : A table containing the revenue earned on each weekday.
  - Primary Key: year, quarter, md\_code, dept\_code, weekday
- gender\_sales : A table containing the revenue earned from each gender: male or female.
  - Primary Key: year, quarter, md\_code, dept\_code, gender

## 2 Questions

1. Which commercial district showed the biggest reduction of floating population?

Query

```
select bc.md_code, bc.before_avg, ac.after_avg,
(ac.after_avg - bc.before_avg) * 100 / before_avg as rate
from (select md_code, sum(population) as before_avg
from (select fp.*, di.district
from floating_pop fp join md_info mi on fp.md_code = mi.md_code
join district_info di on mi.district_code = di.district_code) popsum
where year >= 2018 and year <= 2019
group by md_code) bc
join
(select md_code, sum(population) / 2 as after_avg
from (select fp.*, di.district
from floating_pop fp join md_info mi on fp.md_code = mi.md_code
join district_info di on mi.district_code = di.district_code) popsum2
where year >= 2020
group by md_code) ac
on bc.md_code = ac.md_code
--and bc.gender = ac.gender
order by rate asc;
```

Figure 2: Question 1 query

Result

I compared the last two years before 2020, when the COVID-19 first spread out, and the two years right after 2020. The result looks like below.

	123 md_code T↑	123 before_avg T↑	123 after_avg T↑	123 rate T↑
1	1,000,764	3,240,988	587,964	-81.8584950021
2	1,001,169	907,966	218,716.5	-75.9113777388
3	1,001,191	863,931	215,860	-75.0142083106
4	1,001,310	210,863	52,945	-74.8912801203
5	1,001,215	4,011,304	1,012,637	-74.7554161938

Figure 3: Question 1 result

Interesting thing is that the 3 districts that showed the biggest reduction are all related to universities. The first one, whose area code equals to 1000764 that showed the largest amount of reduction of floating population, is in the Heukseok-dong, the district Dongjak near the CAU(Chung-Ang University, 중앙대학교). The second and third area are near the Sookmyung Women's University(숙명여대) and Hanyang University(한양대학교), respectively. I think this attributes to the online course policy that each university adopted after the pandemic. In addition, the last two districts are Gu-ui market(구의시장) and the station Gwanghwamun(광화문역).

## 2. Which district showed the biggest and the lowest reduction of floating population?

Query

```
select before_covid.district, before_covid.sumpop, after_covid.sumpop,
(after_covid.sumpop - before_covid.sumpop) * 100 / before_covid.sumpop as rate
from (select district, sum(population) sumpop
from (
select fp.year, fp.quarter, fp.md_code, di.district, fp.gender, fp.population
from floating_pop fp
join md_info mi on fp.md_code = mi.md_code
join district_info di on mi.district_code = di.district_code
order by district, year desc, quarter
) distpop
where year = 2019
group by district) before_covid
join
(select district, sum(population) / 2 as sumpop
from (
select fp.year, fp.quarter, fp.md_code, di.district, fp.gender, fp.population
from floating_pop fp
join md_info mi on fp.md_code = mi.md_code
join district_info di on mi.district_code = di.district_code
order by district, year desc, quarter
) distpop
where year >= 2020
group by district) after_covid
on before_covid.district = after_covid.district
order by rate
limit 5;
```

Figure 4: Question 2 query

## Result

	ABC district ↑↓	123 sumpop ↑↓	123 sumpop ↑↓	123 rate ↑↓
1	종로구	130,370,101	108,299,564	-16.9291400641
2	중구	149,921,061	128,350,808.5	-14.3877400254
3	마포구	191,046,219	178,182,155	-6.7334826448
4	서초구	185,757,672	173,729,323.5	-6.4752902911
5	용산구	99,513,992	93,790,938	-5.7510043412

Figure 5: Question 2 result - Biggest reduction

	ABC district ↑↓	123 sumpop ↑↓	123 sumpop ↑↓	123 rate ↑↓
1	강서구	188,958,311	191,925,678	1.5703818394
2	양천구	127,780,764	129,559,329.5	1.3918882971
3	중랑구	142,366,499	142,597,794.5	0.1624648366
4	금천구	66,333,251	66,255,370.5	-0.1174079347
5	강동구	206,642,782	205,554,266.5	-0.5267619268

Figure 6: Question 2 result - Lowest reduction

The districts with the biggest reduction has one thing in common: there are well-known places (or so-called "hot places") in each district. For example, in the district Jongno(종로구), there are many historical sites such as the palace Gyeongbok(경복궁), or popular sites such as Ikseon-dong(익선동). In addition, there is Hongdae street(홍대거리), which is also one of the most popular places in Seoul, in the district Mapo(마포구).

On the other hand, the districts with the lowest reduction share a common feature that each district is well-known as a so-called "bedtown", where the most people floating around the district actually reside there.



### 3. The relationship between the number of confirmed cases and sales by types of business

Query

```
with covid as (
  select
    covid_dist.calendar_date, c.year, c.quarter,
    covid_dist.district_code, covid_dist.district,
    covid_dist.cases
  from (select cc.*, di.district
        from covid_cases cc join district_info di
        on cc.district_code = di.district_code) covid_dist
  join calendar c on covid_dist.calendar_date = c.calendar_date
  where c.year = 2021
  order by calendar_date, district_code
), qsales as (
  select
    gs.year, gs.quarter, gs.md_code, gs.dept_code, gs.gender, gs.sales,
    mi.district_code
  from gender_sales gs join md_info mi on mi.md_code = gs.md_code
), casesum as (
  select
    year, quarter, sum(cases) as casesum
  from covid group by year, quarter
), qsum as(
  select
    year, quarter, dept_code, sum(sales) as salesum
  from qsales group by year, quarter, dept_code
)
select
  c.year, c.quarter, q.dept_code, c.casesum, q.salesum
from casesum c join qsum q
on c.year = q.year and c.quarter = q.quarter
order by dept_code, quarter
```

Figure 7: Question 3 query

## Result

	123 year	123 quarter	845 dept_code	123 casesum	123 salesum
1	2,021	1	CS100001	11,621	1,448,203,383,622
2	2,021	2	CS100001	17,137	1,756,560,345,056
3	2,021	3	CS100001	48,035	1,469,715,299,110
4	2,021	4	CS100001	114,933	1,873,692,575,459
5	2,021	1	CS100002	11,621	141,909,609,878

Figure 8: Part of the question 3 result

In every type of business, the total sales reduced in the 3rd quarter in 2021 and the confirmed cases increased during the same period. This makes sense. However, in the last quarter, the confirmed cases almost "sky-rocketed" while the sales increased. This may attribute to the change of the quarantine policy. As the Omicron variant, which is known to be less fatal than the other variants before, the Korean government has changed the quarantine policy in an easing way. This might have contributed to increase in sales. Furthermore, the type that showed the biggest increase is the Karaoke room(노래방).

#### 4. The relationship between the number of confirmed cases and sales by commercial district

Query

```
with covid as (
  select
    covid_dist.calendar_date, c.year, c.quarter,
    covid_dist.district_code, covid_dist.district,
    covid_dist.cases
  from (select cc.*, di.district
        from covid_cases cc join district_info di
        on cc.district_code = di.district_code) covid_dist
  join calendar c on covid_dist.calendar_date = c.calendar_date
  where c.year = 2021
  order by calendar_date, district_code
), qsales as (
  select
    gs.year, gs.quarter, gs.md_class_code, ci.md_class_name, gs.gender, gs.sales
  from gender_sales gs join class_info ci on gs.md_class_code = ci.md_class_code
), casesum as (
  select
    year, quarter, sum(cases) as casesum
  from covid group by year, quarter
), qsum as(
  select
    year, quarter, md_class_code, md_class_name, sum(sales) as salesum
  from qsales group by year, quarter, md_class_code, md_class_name
)
select
  c.year, c.quarter, q.md_class_code, q.md_class_name, c.casesum, q.salesum
from casesum c join qsum q
on c.year = q.year and c.quarter = q.quarter
order by md_class_code, quarter
```

Figure 9: Question 4 query

Result

	123 year T	123 quarter T	ABC md_class_code T	ABC md_class_name T	123 casesum T	123 salesum T
1	2,021	1	A	골목상권	11,621	4,528,184,213,020
2	2,021	2	A	골목상권	17,137	5,060,920,763,670
3	2,021	3	A	골목상권	48,035	4,891,379,341,013
4	2,021	4	A	골목상권	114,933	5,035,104,336,613
5	2,021	1	D	발달상권	11,621	7,221,927,348,817

Figure 10: Part of the question 4 result

Like above, most of commercial district classes showed increase in sales in the last quarter of 2021. The class showed the largest increase is the class code D, which is defined as the

”shopping street”, where many kinds of shops are concentrated. However, one class showed less increase than in the 3rd quarter, which is the class code R, defined as the ”traditional market”.

## 5. The relationship between the number of confirmed cases and sales by gender

Query

```
with covid as (
  select
    covid_dist.calendar_date, c.year, c.quarter,
    covid_dist.district_code, covid_dist.district,
    covid_dist.cases
  from (select cc.*, di.district
        from covid_cases cc join district_info di
        on cc.district_code = di.district_code) covid_dist
  join calendar c on covid_dist.calendar_date = c.calendar_date
  where c.year = 2021
  order by calendar_date, district_code
), gsales as (
  select
    gs.year, gs.quarter, gs.gender, sum(gs.sales) as gendersum
  from gender_sales gs group by year, quarter, gender
), casesum as (
  select
    year, quarter, sum(cases) as casesum
  from covid group by year, quarter
)
select
  amt.year, amt.quarter, amt.gender, amt.gendersum, pop.popsum,
  (amt.gendersum / pop.popsum) as sale_per_capita
from(
  select
    c.year, c.quarter, g.gender, g.gendersum
  from casesum c join gsales g
  on c.year = g.year and c.quarter = g.quarter
  order by gender, quarter
) amt join (
  select year, quarter, gender, sum(population) as popsum
  from floating_pop
  where year = 2021
  group by year, quarter, gender
) pop
on amt.year = pop.year and amt.quarter = pop.quarter and amt.gender = pop.gender
order by quarter, gender
```

Figure 11: Question 5 query

## Result

	123 year T↑	123 quarter T↑	123 gender T↑	123 gendersum T↑	123 popsum T↑	123 sale_per_capita T↑
1	2,021	1	남성	7,831,749,416,932	436,803,378	17,929.6905916602
2	2,021	1	여성	7,248,804,312,006	477,255,488	15,188.5195545536
3	2,021	2	남성	8,406,812,721,419	435,100,785	19,321.5296575919
4	2,021	2	여성	8,029,843,470,911	478,191,346	16,792.1137387355
5	2,021	3	남성	8,140,002,359,862	424,293,381	19,184.8440828305
6	2,021	3	여성	7,687,249,767,600	465,492,252	16,514.2378515894
7	2,021	4	남성	8,781,465,630,736	430,056,153	20,419.3465655077
8	2,021	4	여성	8,341,339,012,738	471,767,803	17,681.0264704266

Figure 12: Question 5 result

Like above, regardless of gender, sales decreased in the 3rd quarter but increased in the last quarter of 2021. Interesting thing is that the sales per capita earned from the male consumers are larger than that earned from the female, in every quarter. This might attribute to the difference in economic power between gender. In other words, the female consumers visit commercial districts more often, while they tend to spend less than the male.

## 6. The status change in commercial districts before and after the pandemic

Query

```
with change_table as
(
    select
        change.*, di.district, ci2.md_class_name
    from (select ci.*, cc.change_name
    from change_info ci join change_codes cc
    on ci.change_index = cc.change_index) change
    join md_info mi on change.md_code = mi.md_code
    join district_info di on mi.district_code = di.district_code
    join class_info ci2 on ci2.md_class_code = change.md_class_code
)
select bc.district, bc.change_index, bc.cnt, ac.cnt
from (
    select district, change_index, count(*) as cnt
    from change_table
    where year < 2020
    group by district, change_index
    order by district, cnt desc
) bc join (
    select district, change_index, count(*) as cnt
    from change_table
    where year >= 2020
    group by district, change_index
    order by district, cnt desc
) ac
on bc.district = ac.district and bc.change_index = ac.change_index
```

Figure 13: Question 6 query

Result

21	관악구	↗ HH	764	199
22	관악구	↗ LL	621	265
23	관악구	↗ LH	388	128
24	관악구	↗ HL	339	112

Figure 14: Part of the question 6 result

In almost districts, the order of frequency of the status index did not change. However, there are some districts that showed change in the order, such as the district Gwanak(관악구) or the district Guro(구로구). The status is determined by comparison of average lasting period(평균생존기간) and average closure period(평균폐업기간) with the overall average of Seoul. For example, in the district Gwanak, frequency of the status HH showed less than the status LL. This can be interpreted as the commercial districts where the shops are "last short, re-enter fast" have increased while the districts where the shops are "last long, re-enter slow" have decreased after COVID-19. This might attribute to the fact that the pandemic caused severe damage on self-employed workers(thus last short), and the fact that policy has changed in favor of new start-ups(thus re-enter fast)<sup>1</sup>.

---

<sup>1</sup>[구청장이 간다] "관악구, 스타트업 1,000개 유치...창업메카 될것"

## 7. The differences in floating populations between gender

### Query

```

with md_dist as (
  select
    mi.md_code, mi.md_name, mi.district_code, di.district
  from md_info mi join district_info di on mi.district_code = di.district_code
), pop as (
  select
    fp.year, fp.md_code, fp.gender, sum(fp.population) as genpop
  from floating_pop fp
  group by fp.year, fp.md_code, fp.gender
), mdpop as (
  select
    md_dist.md_code, md_dist.md_name, md_dist.district_code, md_dist.district,
    pop.year, pop.gender, pop.genpop
  from md_dist join pop on md_dist.md_code = pop.md_code
)
(select female.*, male.gender, male.genpop, (female.genpop - male.genpop) as difference
from
  (select * from mdpop
  where gender = '여성') female
join (
  select * from mdpop
  where gender = '남성'
) male
on female.md_code = male.md_code and female.md_name = male.md_name and female.district_code = male.district_code
and female.district = male.district and female.year = male.year
where female.year = 2021
order by difference desc
limit 5)
union all
(select female.*, male.gender, male.genpop, (female.genpop - male.genpop) as difference
from
  (select * from mdpop
  where gender = '여성') female
join (
  select * from mdpop
  where gender = '남성'
) male
on female.md_code = male.md_code and female.md_name = male.md_name and female.district_code = male.district_code
and female.district = male.district and female.year = male.year
where female.year = 2021
order by difference asc
limit 5)

```

Figure 15: Question 7 query

### Result

	123 md_code	123 md_name	123 district_code	123 district	123 year	123 gender	123 genpop	123 gender	123 genpop	123 difference
1	1,001,186	서울 서대문구 이대역	11,410	서대문구	2,021	여성	3,164,587	남성	1,639,580	1,525,007
2	1,001,182	홍익대학교 주변	11,440	마포구	2,021	여성	9,221,394	남성	8,143,586	1,077,808
3	1,001,135	신사동 가로수길	11,680	강남구	2,021	여성	7,487,619	남성	6,427,959	1,059,660
4	1,001,141	압구정 로데오거리_2	11,680	강남구	2,021	여성	5,495,090	남성	4,436,049	1,059,041
5	1,001,185	서울 서대문구 신촌역	11,410	서대문구	2,021	여성	7,503,093	남성	6,511,346	991,747
6	1,000,838	호암로20길	11,620	관악구	2,021	여성	2,253,254	남성	3,674,208	-1,420,954
7	1,001,494	종로?청계 관광특구	11,110	종로구	2,021	여성	6,702,631	남성	8,104,292	-1,401,661
8	1,000,839	호암로24길	11,620	관악구	2,021	여성	2,877,117	남성	4,226,826	-1,349,709
9	1,000,813	서림7길	11,620	관악구	2,021	여성	1,651,613	남성	2,447,033	-795,420
10	1,001,050	교대역_1	11,650	서초구	2,021	여성	3,391,425	남성	4,165,350	-773,925

Figure 16: Part of the question 7 result

The top 5 places where the female visitors are more than the male visitors have one thing in common: they are one of the most well-known places for entertainment in Seoul. However, the



top 5 places where the male visitors are more than the female also have one thing in common: there are places to study and prepare for the national exams, such as the public officer exam.

## 8. Which type of business earned the most near the Seoul National University?

Query

```
with gwanak as (
  select mi.*, di.district
  from md_info mi join district_info di
  on mi.district_code = di.district_code
  where di.district like '%관악%'
), sales as (
  select gs.year, gs.quarter, gs.md_code, gs.dept_code, d.dept_name, sum(sales) as sale
  from gender_sales gs join dept d on gs.dept_code = d.dept_code
  group by gs.year, gs.quarter, gs.md_code, gs.dept_code, d.dept_name
), gwanak_sales as (
  select g.*, s.year, s.quarter, s.dept_code, s.dept_name, s.sale
  from gwanak g join sales s on g.md_code = s.md_code
)
select
  md_class_code, md_code, md_name, dept_name, sum(sale) as salesum
from gwanak_sales
group by md_class_code, md_code, md_name, dept_name
order by salesum desc
```

Figure 17: Question 8 query

Result

	ABC md_class_code	123 md_code	ABC md_name	ABC dept_name	123 salesum
1	D	1,001,025	서울 관악구 서울대입구역_1	일반의류	178,695,650,238
2	D	1,001,025	서울 관악구 서울대입구역_1	일반의원	58,194,831,660
3	A	1,000,829	조원로10길	반찬가게	56,816,091,624
4	D	1,001,025	서울 관악구 서울대입구역_1	한식음식점	42,432,747,475
5	D	1,001,031	서울 관악구 신림역_1	한식음식점	24,985,159,689
6	D	1,001,025	서울 관악구 서울대입구역_1	편의점	18,237,045,515
7	D	1,001,046	서울 관악구 신림역_4	편의점	15,572,409,684

Figure 18: Question 8 result

As we can see, amenities including restaurants, hospitals, or convenient stores recorded the largest sales around the Seoul National University. This might attribute to the fact that there are many residential facilities around the SNU. These days, students, households, and even workers who reside around the SNU have increased due to the relatively cheap rent. Therefore, what this result says is quite reasonable.

## 9. Sales by gender around the district Gwanak

### Query

```
with gwanak as (
  select mi.*, di.district
  from md_info mi join district_info di
  on mi.district_code = di.district_code
  where di.district like '%관악%'
), gwanak_gsales as (
  select g.*, gs.year, gs.quarter, gs.md_code, gs.gender, gs.sales
  from gwanak g join gender_sales gs on g.md_code = gs.md_code
), gwanak_wsales as (
  select g.*, ws.year, ws.quarter, ws.md_code, ws.weekday, ws.sales
  from gwanak g join weekday_sales ws on g.md_code = ws.md_code
)
(select female.md_name, female.fsum, male.msum, (female.fsum - male.msum) as difference
from
(select md_class_code, md_name, gender, sum(sales) as fsum
from gwanak_gsales
where gender = '여성'
group by md_class_code, md_name, gender ) female
join
(select md_class_code, md_name, gender, sum(sales) as msum
from gwanak_gsales
where gender = '남성'
group by md_class_code, md_name, gender ) male
on female.md_class_code = male.md_class_code and female.md_name = male.md_name
order by difference desc
limit 5)
union all
(select female.md_name, female.fsum, male.msum, (female.fsum - male.msum) as difference
from
(select md_class_code, md_name, gender, sum(sales) as fsum
from gwanak_gsales
where gender = '여성'
group by md_class_code, md_name, gender ) female
join
(select md_class_code, md_name, gender, sum(sales) as msum
from gwanak_gsales
where gender = '남성'
group by md_class_code, md_name, gender ) male
on female.md_class_code = male.md_class_code and female.md_name = male.md_name
order by difference asc
limit 5);
```

Figure 19: Question 9 query

## Result

	ABC md_name	123 fsum	123 msum	123 difference
1	서울 관악구 서울대입구역_1	230,043,660,768	190,655,409,050	39,388,251,718
2	인현시장	31,209,900,389	23,265,516,652	7,944,383,737
3	장군봉2길	12,522,511,089	5,383,449,233	7,139,061,856
4	신원시장	23,582,347,145	17,482,408,675	6,099,938,470
5	중부시장	21,479,718,960	16,123,256,606	5,356,462,354
6	조원로10길	25,944,195,856	53,303,280,270	-27,359,084,414
7	호암로24길	23,628,054,609	37,952,467,524	-14,324,412,915
8	서울 관악구 신림역_4	40,108,239,870	48,246,650,988	-8,138,411,118
9	서울 관악구 신림역_1	40,817,945,231	48,417,434,855	-7,599,489,624
10	대학길	4,618,358,635	12,166,459,366	-7,548,100,731

Figure 20: Question 9 result

The first 5 rows show the top 5 commercial districts where the female consumers spent more than the male, while the last 5 rows show the opposite result, where the male consumers spent more. Here, female in the district Gwanak spent a lot in the traditional market. This expenditure might attribute to the living of a family, spent by housewives.

However, we can see that the results from the male consumers are the places where the male floating population is much larger than the female. Thus, it can be inferred that this result reflects the result of the query 7.

## 10. Which district showed the biggest reduction in residing and working population?

Query

```

with pop as (
  select
    d.year, d.quarter, d.md_code, d.gender,
    d.age, d.population + w.population as float_pop
  from domestic_pop d join working_pop w
  on d.year = w.year and d.quarter = w.quarter and d.gender = w.gender and d.age = w.age and d.md_code = w.md_code
  order by year desc, quarter desc, md_code, age, gender
), region_info as (
  select m.md_code, m.md_name, di.district
  from md_info m join district_info di
  on m.district_code = di.district_code
), district_pop as (
  select p.*, r.md_name, r.district
  from pop p join region_info r
  on p.md_code = r.md_code
)
select
  before_covid.district, before_covid.average_float,
  after_covid.average_float,
  (after_covid.average_float - before_covid.average_float) * 100 / after_covid.average_float as float_growth
from
  (select district, avg(float_pop) as average_float
  from district_pop
  where year < 2020
  group by district) before_covid
join
  (select district, avg(float_pop) as average_float
  from district_pop
  where year >= 2020
  group by district) after_covid
on before_covid.district = after_covid.district
order by float_growth
limit 5;

```

Figure 21: Question 10 query

Result

	ABC district	123 average_float	123 average_float	123 float_growth
1	중구	894.4108108108	510.9641726049	-75.0437425487
2	금천구	465.2809792844	336.7970020325	-38.1487888777
3	서초구	560.4391551762	415.5160024155	-34.8778752005
4	마포구	291.7428539275	224.6811594203	-29.8474935238
5	노원구	226.0063492063	181.0398284314	-24.8379161451

Figure 22: Question 10 result

This result is mainly influenced by the working population. The reason why the working population has plummeted in the district Jung(중구) can be inferred by the environment around the district. There is a well-known commercial district, called Myeong-dong(명동) in

the district, and the major consumer of the commercial district is foreigners, especially Chinese. However, overseas travel has been prevented because of the pandemic, lots of shops in the commercial district closed, thus the unemployment has increased. This might be one of the reasons why the district Jung showed this dramatic reduction.