

# 서울교통공사가 관리하는 역들의 지하상가 임대정보 데이터 분석

- 훈련 과정 명 : 머신러닝(Machine Learning)을 활용한 데이터 분석 과정
- 훈련 기간 : 2020-10-27 ~ 2021-03-23 (800시간/100일)
- 이름 : 전 병 우

# INDEX

✓ 주제 및 목적(기획의도)

✓ 프로젝트 일정 계획

## 주제 : 프로젝트 주제

서울교통공사가 관리하는 역들의 지하상가 임대정보 데이터 분석

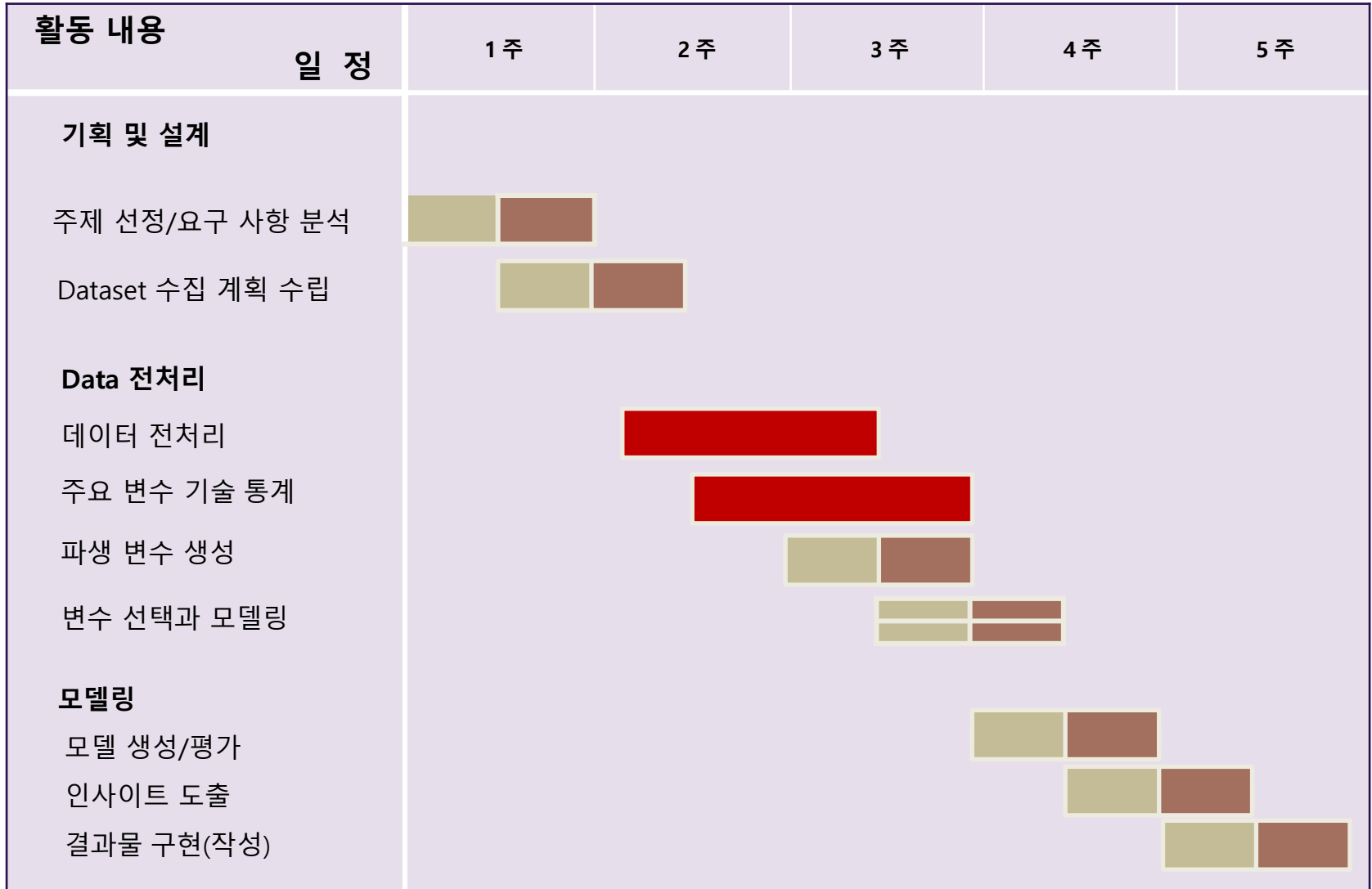
## 목적(비즈니스 관점에서의 기획의도) :

무엇을 목적으로 프로젝트를 진행하는가?

서울교통공사가 운영하는 상가 임대사업이 계속해서 고전을 하고 있는 이유를 분석

이를 통해 앞으로의 서울교통공사 임대사업 개선점 고민

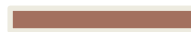
# 프로젝트 일정 계획



계획기간



완료기간



중요기간



## 주제 선정 이유

‘코로나19’ 팬데믹

“목 좋다는 것도 옛말”...코로나에 폐업 줄이는 지하철역  
상가들

“공사, 지하철 상가 개별입찰  
” “조 무더기 유찰

6·7호선 상가 돌아보니 ... 만나지 않는 노선  
처럼 합의점도 없었다

서울교통공사는 매년 지속적인 적자상황에 처해있다.

2020년 상반기에만 약 4800억원의 적자를 기록했고 2020년에 약 9700억의 적자가 예상되고 있다.  
이런 상황에서 서울교통공사의 임대상가들 역시 폐업하거나 비어있는 채로 방치되고 있다.

거기에 2019년 서울교통공사와 GS라테일과의 계약이 종료되면서 약406개의 상가가 폐점했고, 2020년  
코로나 사태가 터지면서 임대상가들이 줄줄이 폐업하고 있고 있어 공실률이 30%를 넘어섰다.

수도권 지하철 대부분을 담당하고 있는 서울교통공사의 지속적인 적자는 나라의 입장에서조차 결코 좋은 상  
황은 아니다.

이런 서울교통공사의 적자를 ‘임대상가의 공실률을 줄임으로써 어느정도 해결할 수 있지 않을까’라는 생각  
에 이번 주제를 선정했다.

# 데이터 준비

## 데이터

· 서울교통공사의 상가임대 현황 (면적,업종,계약일,임대료 등) 데이터

NO	상가 유형	호선	역명	상가 번호	면 적...	영업업종	보호 업종	계약시 작	계약종 료	임대 료
1	1 개별(일반)	1호선	서울역_1	150-107	33.00	사무실		2019-05-08	2024-06-06	527100
2	2 소송상가	1호선	서울역_1	150-109	12.00	커피		2012-05-21	2017-07-10	5178550
3	3 네트워크	1호선	서울역_1	150-110	41.30	화장품		2015-09-05	2020-11-03	17414416
4	4 개별(일반)	1호선	시정_1	151-101	29.73	의류		2017-04-04	2022-05-03	3858954
5	5 개별(일반)	1호선	시정_1	151-103	57.60	장애인생산물		2020-02-01	2025-01-31	1858300
6	6 공실	1호선	시정_1	151-104	25.00					NA
7	7 공실	1호선	시정_1	151-105	25.00					NA
8	8 개별(일반)	1호선	시정_1	151-106	14.00	액세서리		2017-09-19	2022-11-17	1801800
9	9 공실	1호선	시정_1	151-107	22.00					NA
10	10 네트워크	1호선	종각	152-101	36.85	화장품		2015-09-05	2020-11-03	20734494
11	11 개별(일반)	1호선	종각	152-105	29.30	편의점		2017-04-18	2022-04-17	6549400
12	12 개별(일반)	1호선	종각	152-106	30.70	의류		2019-04-30	2024-05-29	3618600
13	13 개별(일반)	1호선	종각	152-109	42.00	제과		2020-02-24	2025-03-25	7020000
14	14 개별(일반)	1호선	종각	152-111	16.15	액세서리		2020-03-13	2025-04-12	2900000
15	15 개별(일반)	1호선	종각	152-115	29.40	제과		2020-02-24	2025-03-25	8130000
16	16 개별(일반)	1호선	종로3가_1	153-101	23.40	편의점		2018-10-23	2023-11-21	2260000

## 변수 해석

### 1) 상가 유형 : chr

- 임대료를 받고 있는 유형 : 개별(일반), 네트워크, 개별(대형), 복합, GS
- 임대료를 받고 있지 않는 유형 : 소송상가, 공실

### 2) 호선 : chr

- 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선
- 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

### 3) 역명 : chr

- 서울교통공사가 관리하고 있는 역명
- 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다.

### 4) 면적 : num

- 계약면적

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
6.30 22.41 32.00 50.08 44.46 7475.19 59

### 5) 영업업종 : chr

- 84개의 영업업종, 공실일 경우 공백

### 6) 임대료 : num

- 단위(원)

Min. 1st Qu. Median Mean 3rd Qu. Max.  
0 0 893000 3741404 4560000 279648140

# 데이터 준비

## 데이터

· 서울교통공사의 월별 승하차 인원 데이터

소... 계.1.8 호선.	NA.	NA..1	X1941814283	X8005099	X271351777	X223795684	X185092194	X195853780	X214845496
1 1호선	150	서울역_1	19271145	78980	3198290	2271168	1601396	1803550	2112629
2 NA	151	시정_1	9305144	38136	1477362	1149641	888317	874072	962697
3 NA	152	종각	16183400	66325	2462560	1945843	1475933	1586710	1745454
4 NA	153	종로3가_1	12199520	49998	1935106	1421415	1095035	1224282	1374997
5 NA	154	종로5가	10298572	42207	1527217	1149021	947157	1040487	1182917
6 NA	155	동대문_1	5679643	23277	853074	618961	507532	578736	674203
7 NA	156	신설동_1	6485186	26579	874966	743203	629681	664350	735967
8 NA	157	제기동	8153216	33415	1218944	856496	738627	847495	926402
9 NA	158	청량리	9956457	40805	1453835	1068751	912434	1014531	1134388
10 NA	159	동묘앞_1	4505312	18464	618741	454685	420816	491397	555608
11 2호선	201	시정_2	9849634	40367	1410239	1206303	988688	979300	1028747
12 NA	202	을지로입구	17750938	72750	2913336	2175839	1626671	1692730	1819400
13 NA	203	을지로3가_2	9122400	37387	1309227	1116090	877009	909261	963709
14 NA	204	을지로4가_2	5564505	22805	761484	637805	540873	568481	611620
15 NA	205	동대문역사문화공원_2	6466676	26503	1066000	799054	579695	616463	700892
16 NA	206	신당_2	6456262	26460	867095	735019	632937	664287	723991

## 변수 해석

1) 소...계.1.8호선.: chr

- 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선
- 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

2) NA.: num

- 번호

3) NA..1 : chr

- 서울교통공사가 관리하고 있는 역명
- 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다.

4) X1941814283: num

- 1월부터 8월까지의 승 하차 인원의 총 합계

5) X8005099: num

- 일 평균 승하차 인원

6) X271351777~X0.2: num

- 각 행의 1월부터 8까지 월별 승하차 인원의 합계

# 데이터 준비

## 데이터

· 서울교통공사의 월별 환승 유입 인원 데이터

▲ 소... 계.1.8 호선.	NA.	NA..1	X492776800	X1803143	X69096966	X56677447	X46676388	X49502832	X54453873
1 1호선	150	서울역_1	6485880	23671	1037363	752705	543186	613677	711985
2 NA	151	시정_1	3648125	13314	559786	444016	354856	345914	378667
3 NA	152	종각	6856180	25023	1017744	818706	632163	672448	740600
4 NA	153	종로3가_1	3574259	13045	571505	421216	322929	357447	398811
5 NA	154	종로5가	4199277	15326	609162	467317	391997	425560	481792
6 NA	155	동대문_1	1825052	6661	268339	191142	160336	187184	221873
7 NA	156	신설동_1	2084549	7608	269288	233467	203023	214118	239407
8 NA	157	계기동	3054487	11148	463293	317758	272804	313655	344854
9 NA	158	청량리	3534447	12899	520853	372940	318474	358239	402087
10 NA	159	동묘앞_1	1474444	5381	202040	146935	135809	161493	183350
11 2호선	201	시정_2	2623705	9576	350292	313790	272715	265699	274677
12 NA	202	을지로입구	4684440	17096	724035	559195	440808	453195	481958
13 NA	203	을지로3가_2	2350128	8577	317055	280804	231427	237412	249267
14 NA	204	을지로4가_2	1171718	4276	151089	130529	115135	120829	129574
15 NA	205	동대문역사문화공원_2	1019569	3721	160050	122968	92425	98111	112080
16 NA	206	신당_2	1463320	5341	195143	166896	144869	150261	165357

## 변수 해석

### 1) 소...계.1.8호선.: chr

- 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선
- 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

### 2) NA.: chr

- 번호

### 3) NA..1 : chr

- 서울교통공사가 관리하고 있는 역명
- 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다.

### 4) X492776800: num

- 1월부터 8월까지의 환승유입 인원의 총 합계

### 5) X1803143: num

- 일 평균 환승유입 인원

### 6) X69096966~NA.. X0.2 : num

- 각 행의 1월부터 8까지 월별 환승유입 인원의 합계



# 데이터 준비

## 데이터

· 서울교통공사의 시간별 혼잡도 데이터

조사일자	호선	역번호	역명	구분	구분.1	X05.30.....	X06.00.....	X06.30.....	X07.00.....	X07.30.....	X08.00.....	X08.30.....	X09.00.....
1 평일	1호선	150	서울역_1	상선	혼잡도	8.2	12.9	21.5	28.4	46.7	70.0	74.5	52.2
2 평일	1호선	150	서울역_1	하선	혼잡도	11.8	13.5	12.2	17.8	26.4	34.8	25.2	21.0
3 평일	1호선	151	시정_1	상선	혼잡도	7.7	11.2	19.6	24.4	36.2	53.7	51.7	47.1
4 평일	1호선	151	시정_1	하선	혼잡도	9.5	11.3	11.6	16.7	23.2	34.1	26.4	21.0
5 평일	1호선	152	종각	상선	혼잡도	6.3	8.1	15.5	16.6	22.7	34.8	28.7	32.4
6 평일	1호선	152	종각	하선	혼잡도	11.1	12.3	12.8	19.3	28.6	44.2	41.2	26.7
7 평일	1호선	153	종로3가_1	상선	혼잡도	6.3	8.1	14.4	16.3	22.9	36.2	34.8	31.1
8 평일	1호선	153	종로3가_1	하선	혼잡도	12.1	12.2	15.9	23.1	39.9	59.0	69.1	37.4
9 평일	1호선	154	종로5가	상선	혼잡도	5.6	6.5	12.4	12.4	16.5	28.8	23.6	21.3
10 평일	1호선	154	종로5가	하선	혼잡도	13.4	17.2	17.1	27.4	45.1	70.8	68.2	37.2
11 평일	1호선	155	동대문_1	상선	혼잡도	8.5	6.7	12.0	11.0	16.2	26.1	23.6	20.5
12 평일	1호선	155	동대문_1	하선	혼잡도	14.0	18.2	18.1	30.8	46.6	77.4	77.0	40.5
13 평일	1호선	156	신설동_1	상선	혼잡도	7.2	7.9	9.1	8.5	10.5	17.6	17.6	18.5
14 평일	1호선	156	신설동_1	하선	혼잡도	17.1	20.3	21.6	32.4	48.5	72.8	66.8	40.0
15 평일	1호선	157	제기동	상선	혼잡도	6.9	7.4	7.2	8.7	8.6	14.3	16.0	15.0

## 변수 해석

1) 조사일자 : chr

· 평일, 토요일, 일요일

2) 호선: chr

· 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선  
· 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

3) 역번호 : int

· 번호

4) 역명: chr

· 서울교통공사가 관리하고 있는 역명  
· 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다.

5) 구분 : chr

· 상선, 하선 (2호선은 내선, 외선)

6) 구분.1: chr

· 혼잡도

7) X5.30...~X.24.30...: num

· 5시 30분부터 24시 30분 까지 30분간격으로 역별 혼잡도  
· 100% 혼잡도는 1량당 160명 탑승을 기준  
· 80%이하 여유 / 80~130% 보통 / 130~150% 주의 / 150%이상 혼잡

# 데이터 준비

## 데이터

### · 서울교통공사 역사 심도정보

▲	호선	정거장명	정거장층수	형식	㉔지반고.해발고도.100.3m.	㉔레일면고.해발고도.100.3m.	정거장.깊이.선로기준.㉔.㉔.㉔.1.1..m.	정거장.깊이.승강장기준.㉔.㉔.㉔.1.1..m.	비고
1	1	서울역_1	B2	섬식	129.99	117.040	12.950	11.850	4호선환승
2	1	시정_1	B2	상대식	129.97	118.820	11.150	10.050	2호선환승
3	1	종각	B2	상대식	128.77	116.240	12.530	11.430	
4	1	종로3가_1	B2	상대식	124.38	112.040	12.340	11.240	3호선/5호선환승
5	1	종로5가	B2	상대식	121.75	109.260	12.490	11.390	
6	1	동대문_1	B2	상대식	120.52	108.930	11.590	10.490	4호선환승
7	1	동묘앞_1	B1	상대식	119.91	109.970	9.940	8.840	6호선환승
8	1	신설동_1	B2	상대식	118.00	108.460	9.540	8.440	2호선환승
9	1	제기동	B2	상대식	115.91	105.490	10.420	9.320	
10	1	청량리	B2	섬식	114.41	107.970	6.440	5.340	중앙선환승
11	2	시정_2	B3	섬식	132.01	111.020	20.990	19.890	1호선환승
12	2	율지로입구	B2	상대식	127.09	112.530	14.560	13.460	
13	2	율지로3가_2	B2	상대식	125.47	112.770	12.700	11.600	3호선환승
14	2	율지로4가_2	B2	상대식	125.61	110.280	15.330	14.230	5호선환승
15	2	동대문역사문화공원_2	B2	상대식	123.95	108.850	15.100	14.000	4호선/5호선환승
16	2	신당_2	B2	상대식	122.43	109.040	13.390	12.290	6호선환승

## 변수 해석

### 1) 호선: int

- 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선
- 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

### 2) 정거장명: chr

- 서울교통공사가 관리하고 있는 역명
- 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다

### 3) 정거장층수: chr

- B1, B2, B3, B4, B5, B6, B8, 고가, 지상, 지상역사
- 정거장이 위치에 있는 정거장 층수

### 4) 형식: chr

- 단선, 상대식, 섬식, 섬식(복합)

### 5) ㉔지반고.해발고도.100.3m.: num

- 지반고도 해발고도

### 6) ㉔레일면고.해발고도.100.3m.: num

- 레일면고도 해발고도

### 7) 정거장.깊이.선로기준..㉔.㉔..m.: num

- t선로기준 정거장의 깊이

### 8) 정거장.깊이.승강장기준..㉔.㉔.1.1..m.: num

- 승강장기준 정거장의 깊이

### 9) 비고: chr

- 환승역 여부

# 데이터 준비

## 데이터

- 서울교통공사 지하역사공기질 측정결과 데이터

호선명	역명	미세먼지 _PM.10.140..... 이하.	이산화탄소 _CO2..1000.ppm. 이하.	포름알데히드 _HCHO.100..... 이하.	일산화탄소 _CO.9.ppm 이하.
1 1호선	서울역_1	70.5	606	4.0	0.8
2 1호선	시정_1	54.7	581	4.4	0.7
3 1호선	종각	82.9	471	11.5	0.7
4 1호선	종로3가_1	70.4	481	11.0	1.0
5 1호선	종로5가	70.2	597	5.7	0.8
6 1호선	동대문_1	49.4	624	5.4	1.0
7 1호선	동묘안_1	62.4	556	3.4	0.7
8 1호선	신설동_1	75.9	645	3.2	0.6
9 1호선	제기동	71.3	533	3.7	0.8
10 1호선	청량리	65.6	572	4.3	1.1
11 2호선	시정_2	81.4	638	7.6	0.7
12 2호선	을지로입구	72.4	614	7.2	0.6
13 2호선	을지로3가_2	44.4	470	3.6	1.3
14 2호선	을지로4가_2	71.4	521	3.5	1.0
15 2호선	동대문역사문화공원_2	49.2	458	3.1	0.5
16 2호선	신당_2	52.6	582	3.6	1.4

## 변수 해석

### 1) 호선명 : chr

- 서울교통공사가 관리하고 있는 역이 1역이라도 있는 지하철 호선
- 총 8개 호선(1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선)

### 2) 역명: chr

- 서울교통공사가 관리하고 있는 역명
- 9호선은 별도의 사내독립법인으로 운영하기 때문에 제외한다

### 3) 미세먼지\_PM.10.140.....이하.:num

- SPM : 미세먼지 (기준치 140 $\mu$ g/m<sup>3</sup>)

### 4) 이산화탄소\_CO2..1000.ppm.이하.:int

- CO2. : 이산화탄소 (기준치 1000ppm)

### 5) 포름알데히드\_HCHO.100.....이하.: num

- HCHO : 포름알데히드 (기준치 100ppm)

### 6) 일산화탄소\_CO.9.ppm이하. : num

- CO : 일산화탄소 (기준치 9ppm)

# 데이터 전처리

## 데이터

· 서울교통공사의 상가임대 현황 (면적,업종,계약일,임대료 등) 데이터

### 1. 행 이름 변경

- 면적... 행을 면적으로 이름 변경

### 2. 필요없는 행 삭제

- 데이터 분석에 사용하지 않을 행을 삭제한다.
- No, 보호업종, 상가번호, 계약시작, 계약종료는 사용하지 않을 행이므로 삭제
- 원본을 훼손하지 않기 위해 subway\_de라는 새로운 데이터를 생성

### 3. 새로운 데이터 생성

- 같은 역명끼리 묶어 임대료의 평균(mean\_i)과 임대료의 합(sum\_i) 행을 생성
- subway\_ms 라는 새로운 데이터를 생성

# 데이터 전처리

## 데이터

· 서울교통공사의 월별 승하차 인원 데이터

### 1. 필요없는 행 삭제

- 9호선은 분석하지 않을 것이기 때문에 삭제한다.

### 2. 필요없는 열 삭제

- 데이터 분석에 사용하지 않을 행을 삭제한다.
- NA., X0, X0.1, X0.2는 사용하지 않을 행이므로 삭제
- 원본을 훼손하지 않기 위해 passenger라는 새로운 데이터를 생성

### 3. 열 이름 변경

- 열의 성격에 맞게 호선, 역명, 합계, 일평균, 1월,...,9월로 행이름을 변경

### 3. 결측치 처리

- 호선행의 결측치를 해당 역에 맞는 호선으로 채워넣는다.

# 데이터 전처리

## 데이터

· 서울교통공사의 월별 환승 유입 인원 데이터

### 1. 필요없는 행 삭제

- 9호선은 분석하지 않을 것이기 때문에 삭제한다.

### 2. 필요없는 열 삭제

- 데이터 분석에 사용하지 않을 행을 삭제한다.
- NA., X0, X0.1, X0.2는 사용하지 않을 행이므로 삭제
- 원본을 훼손하지 않기 위해 transfer라는 새로운 데이터를 생성

### 3. 열 이름 변경

- 열의 성격에 맞게 호선, 역명, 합계, 일평균, 1월,...,9월로 행이름을 변경

### 3. 결측치 처리

- 호선행의 결측치를 해당 역에 맞는 호선으로 채워넣는다.

# 데이터 전처리

## 데이터

- 서울교통공사의 시간별 혼잡도 데이터

### 1. 필요없는 열 삭제

- 역번호, 호선, 구분, 구분.1 열 삭제

### 2. 결측치 처리

- 혼잡도가 결측치인 경우 0으로 설정

### 3. 열 데이터 변경

- 역명을 기준으로 역별 30분 단위에 평균 혼잡도를 계산한다.
- 원본을 훼손하지 않기 위해 congestion 새로운 데이터를 생성

# 데이터 전처리

## 데이터

· 서울교통공사 역사 심도정보

### 1. 열 이름 변경

- 다른 데이터들과의 통일성을 위해 정거장명을 역명으로 변경

### 2. 필요없는 열 삭제

- 비고 열 삭제



# 데이터 전처리

## 데이터

· 서울교통공사 지하역사공기질 측정결과 데이터

### 1. 결측치 처리

- 공기질 측정결과 데이터 내에 결측치를 0으로 변경

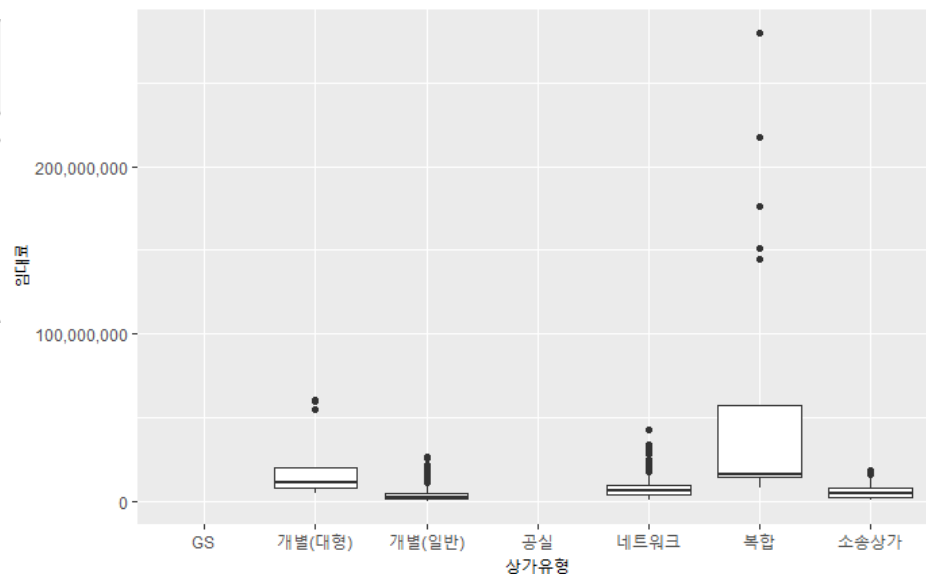
### 2. 열 이름 변경

- 호선명을 호선으로 변경
- 미세먼지\_PM.10.140.....이하.를 SPM으로 변경
- 이산화탄소\_CO2..1000.ppm.이하.를 CO2로 변경
- 일산화탄소\_CO.9.ppm이하.를 CO로 변경
- 포름알데히드\_HCHO.100.....이하.를 HCHO로 변경

# 1. 상가유형 여부에 따른 임대료의 차이

```
table(subway$상가유형)
```

##	GS 개별(대형)	개별(일반)	공실	네트워크	복합	소송상가
##	192	23	522	350	313	207
##						74



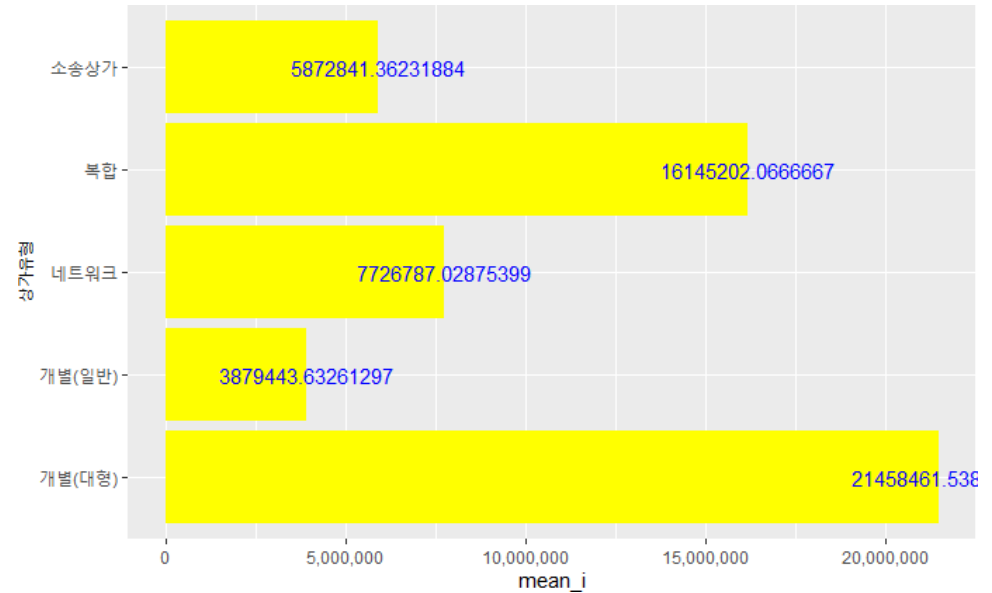
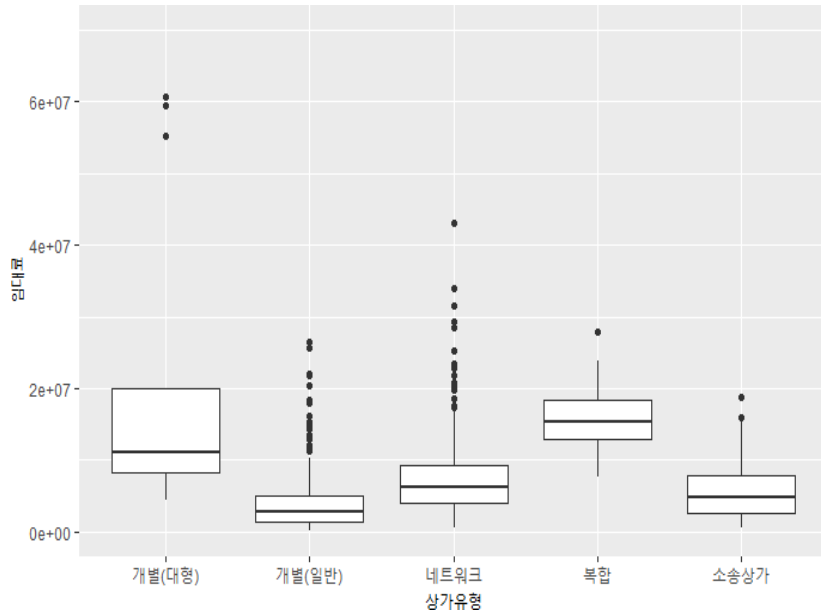
GS유형은 서울교통공사가 GS리테일과 맺은 6, 7호선 유희공간 임대계약이지만 2019년 10월 계약 종료 후 사실상 공실이 된 상태이다.

상가유형별로 어떤 상가유형이 임대료를 많이 내는지를 분석한다.

공실유형과 GS유형은 임대료가 없기 때문에 결측치 처리가 필요하다.

복합유형의 경우에 다른 유형에는 없는 1억이상의 임대료를 내는 곳들이 있어 1억이 넘는 임대료를 이상치로 처리해야 할지 고민해야한다.

# 1. 상가유형 여부에 따른 임대료의 차이



GS 유형과 공실유형은 결측치 처리하였고, 고민했던 1억이상의 복합유형 역시 이상치로 처리하였다.

**결과 :**

평균적으로 가장 많은 임대료를 내는 상가유형은 개별(대형)이며, 가장 적은 임대료를 내는 상가유형은 개별(일반)이다.

shapiro.test 결과로 봤을 때 개별(대형)을 제외하고는 서로 다 차이가 있다.

## 2. 상가면적에 따른 임대료의 차이

면적변수는 계약면적을 기준으로 한다.

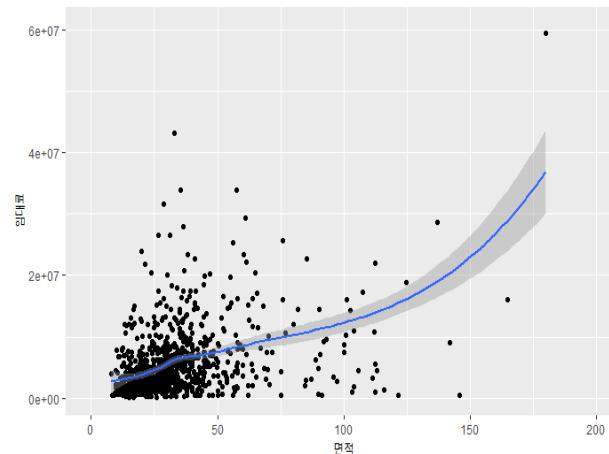
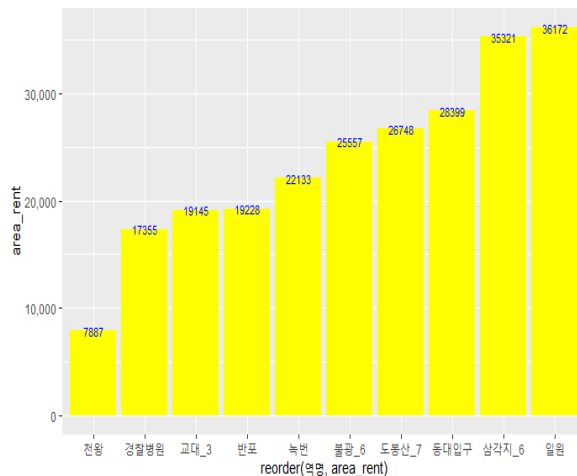
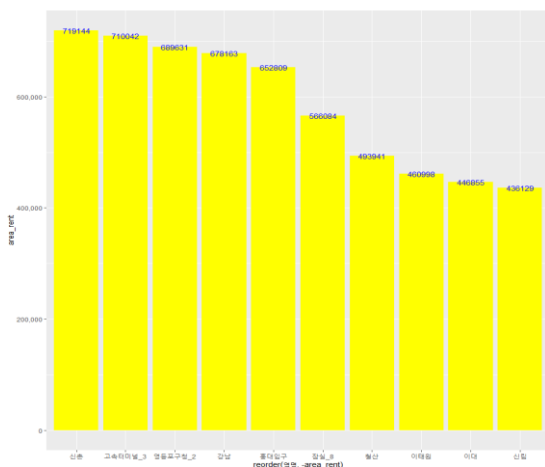
면적변수가 임대료에 영향을 미치는 지를 분석한다.

면적 변수는 결측치나 이상치에 큰 문제가 없는 것으로 확인되었다.

면적 대비 임대료는 역으로 나열할 것이기 때문에 역끼리 묶어  
임대료의 합과 면적의 합을 구해 나눈다.

그 중 상위10개와 하위 10개의 역을 뽑는다.

## 2. 면적에 따른 임대료의 차이



**결과 : 면적대비 임대료를 많이 내는 곳은 신촌역, 고속터미널(3호선)역, 영등포구청(2호선)역 순이다.**

**면적대비 임대료를 적게 내는 곳은 천왕역, 경찰병원역, 교대(3호선)역 순이다.**

**면적과 임대료의 상관관계 분석 결과 양의 상관관계를 나타내고 있다. 따라서 면적이 크면 임대료도 올라간다.**

**하지만 그 수치가 그렇게 크지 않기 때문에 강한 관계라고는 얘기하기는 힘들다.**

### 3. 영업업종에 따른 임대료의 차이

영업업종 변수는 약 50여개의 영업업종으로 구분되어 있다.

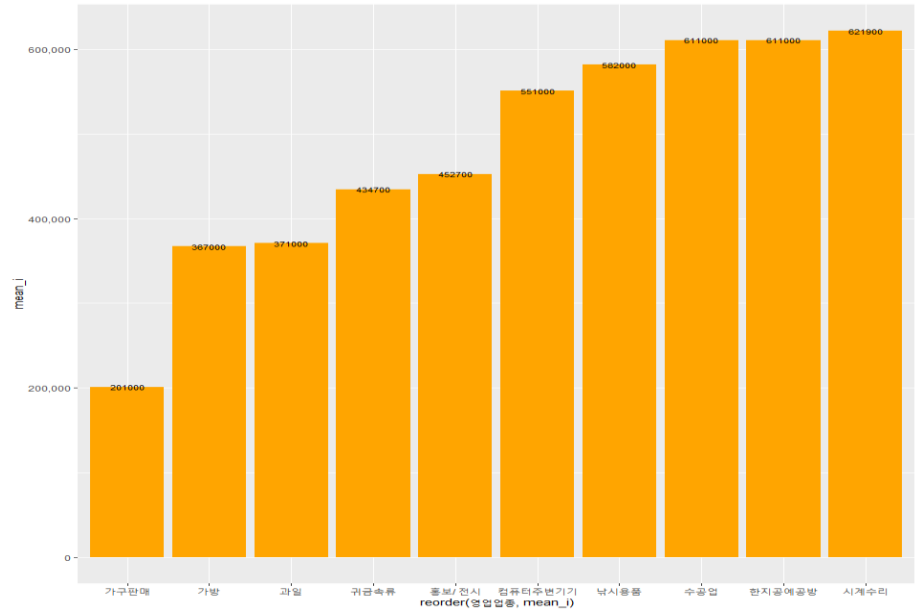
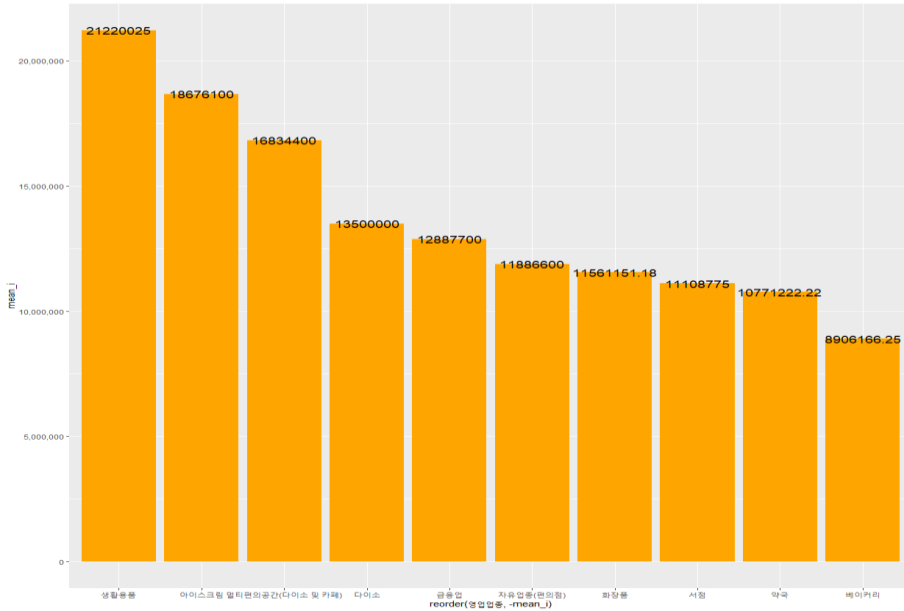
어떤 영업업종이 평균적으로 가장 많이 또는 적게 임대료를 내는지를 분석

영업업종대비 임대료는 영업업종변수 끼리 묶어 평균 임대료를 계산한다.

평균 임대료가 3천만원이 넘는 업종은 다른 영업업종과의 차이가 크기 때문에 이상치로 처리한다.

영업업종내에 임대료는 있으나 영업업종이 공백으로 표시되어 어떤 업종인지 모르는 경우에도 이상치로 처리한다.

### 3. 영업업종에 따른 임대료의 차이



상가 유형 중에서 많은 임대료를 내는 상가유형은 다이소와 같은 생활용품을 판매하는 곳을 나타냈다.

상가 유형중에서 적은 임대료를 내는 상가유형은 가구판매, 가방, 과일과 같이 지하상가에서는 잘 판매하지 않는 곳들이 낮은 것으로 나타났다.

## 4. 호선에 따른 임대료의 차이

서울 교통공사가 운영하는 지하철 노선은 1호선, 2호선, 3호선, 4호선, 5호선, 6호선, 7호선, 8호선으로 총 8개의 호선을 운영하고 있다.

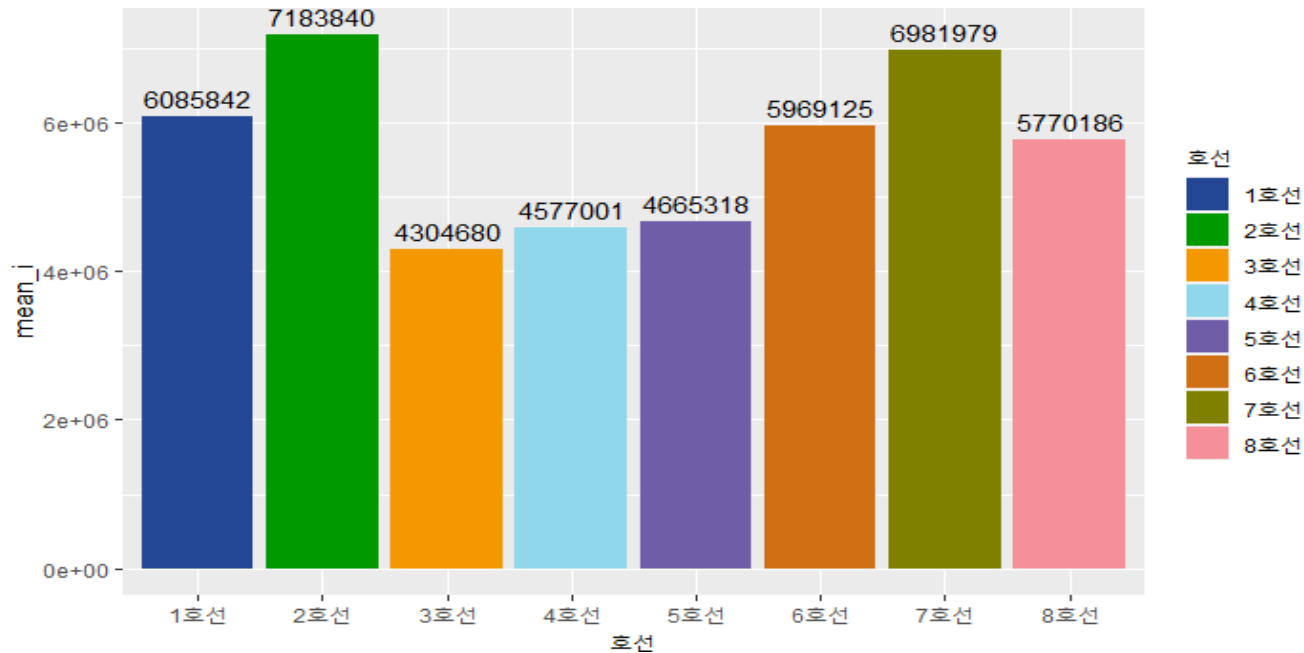
단, 일부 호선은 서울교통공사가 전체 관리하는 것이 아니라 코레일등과 분담하여 관리하고 있기 때문에 서울교통공사가 관리하는 역에 한해서만 분석한다.

이중 평균임대료가 높은 지역과 낮은 지역을 찾아 이유를 분석한다.

호선과 관련하여 더 이상의 이상치와 결측치는 문제 없는 것으로 확인했다.



## 4. 호선에 따른 임대료의 차이



승하차 인원이 많은 2호선과 7호선이 다른 호선들에 비해 임대료가 비싼 것으로 예상된다.

1호선의 경우 서울교통공사가 관리하고 있는 역들이 서울역~동묘앞역 까지로 승하차인원이 많은 역들이므로 임대료가 비싼것으로 예상된다.

하지만 6호선과 8호선의 경우 3, 4, 5호선에 비해 승하차 인원은 적으나 임대료가 비싼 것으로 나타났다.

shapiro.test 결과 1호선, 6호선, 8호선은 어떤 호선과도 임대료의 차이가 없는 것으로 나타나지만 2호선과 7호선은 다른 호선과 임대료의 차이가 있는 것으로 나타난다.

## 5. 역별 임대료의 차이

호선이 아닌 개별 역에 따라서 임대료의 차이가 있는지를 분석한다.

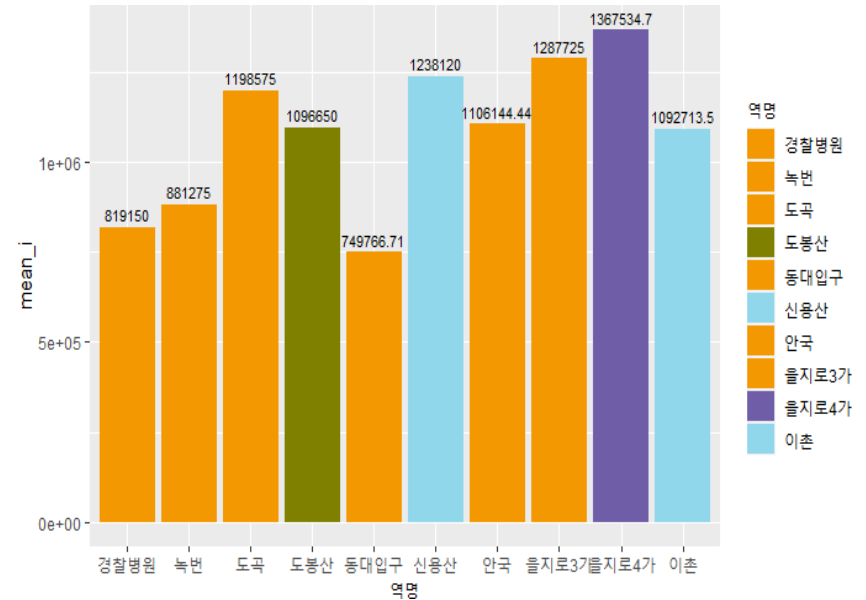
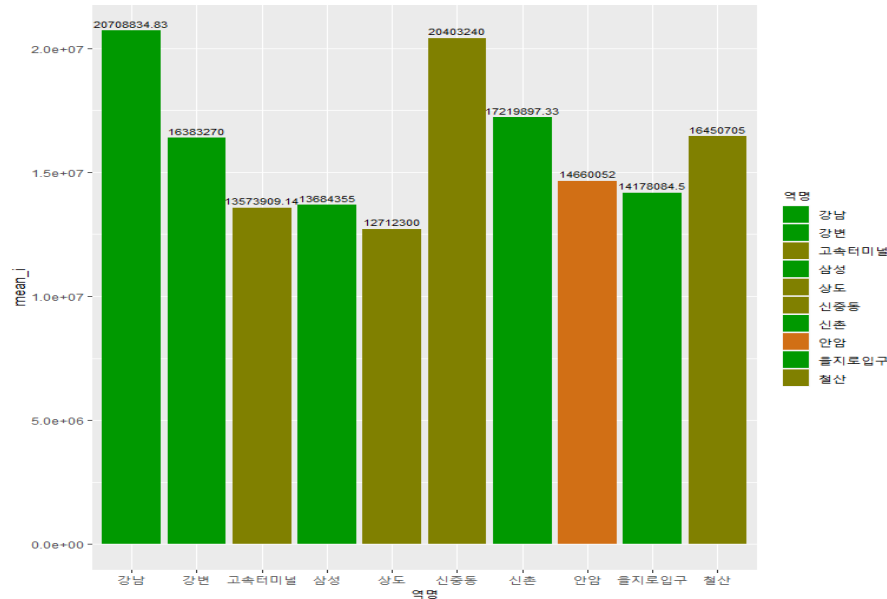
역끼리 묶어 임대료의 합계와 평균을 구하여 분석한다.

역명에 “\_[0-9]” 는 환승역이 있는 경우이며 숫자는 각 호선을 의미한다.

따라서 환승역들을 따로 구분하여 분석할지, 같은 역으로 묶어 분석할지 고민해야 한다.

더 이상 역명과 임대료의 이상치와 결측치는 없는 것으로 확인했다.

## 5. 역별 임대료의 차이



역명은 환승 구분없이 하나의 역으로 묶어 분석하기로 결정하였다.

평균 임대료가 높은 역들은 안암역을 제외한 대부분의 역들이 2호선과 7호선인 것을 알 수 있다.

평균 임대료가 낮은 역들은 3호선, 4호선, 5호선, 7호선이다.

특히 3호선에 해당하는 역들이 주를 이루고 있다.

을지로 3가역과 을지로 4가역은 평균임대료가 높은 2호선과 환승역임에도 하위권에 있다.

## 6. 역별 승하차 인원에 따른 임대료의 차이

승하차 인원이 많은 곳일수록 임대료를 많이 내는지 분석한다.

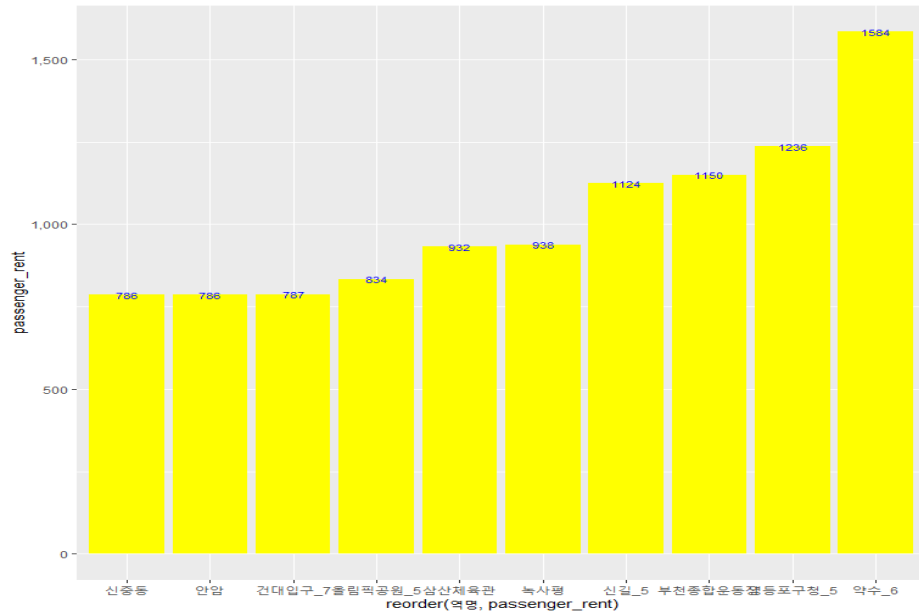
계획 때부터 가장 많은 관계가 있을 것이라고 예상했다.

승하차 인원은 일평균을 기준으로 한다.

승하차 인원 데이터에도 결측치나 이상치는 없는것 으로 확인했다.

임대료는 평균 임대료에서 일평균을 나눈것으로 한다.

## 6. 역별 승하차 인원에 따른 임대료의 차이



결과 약수 역이 일평균 인원대비 임대료가 가장 비싼것으로 나타났다.

전체적으로 봤을 때 일평균인원이 증가하면 평균 임대료도 증가하는 것처럼 보인다.

하지만 상관분석 결과 0.3이 제일 큰 값으로 나오는 것을 보아 큰 상관이 있지는 않다.

## 7. 혼잡도에 따른 임대료의 차이

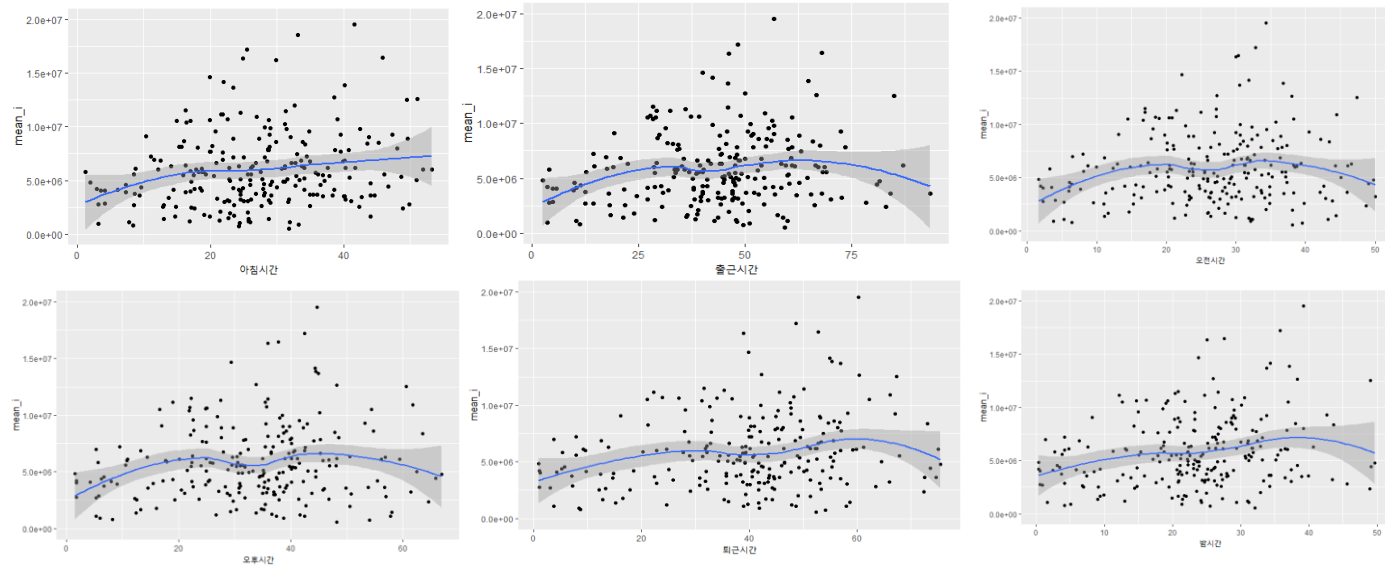
혼잡도가 높은 역일수록 임대료를 많이 내는가를 분석한다.

혼잡도는 지하철 차량 내부의 혼잡도를 말한다.

따라서 승하차 인원이 많다고 해서 무조건 혼잡도가 높지는 않고, 반대로 승하차 인원이 적다고 해서 무조건 혼잡도가 낮지도 않다.

혼잡도 데이터는 30분간격으로 되어있으며  
5시30분 부터 7시 30분까지를 아침시간,  
8시 부터 9시 30분까지 출근시간,  
10시부터 12시 30분까지 오전시간  
13시부터 15시 30분까지 오후시간  
16시부터 20시 30분까지 퇴근시간  
그 이후는 밤시간으로 임의로 설정했다.

## 7. 혼잡도에 따른 임대료의 차이



혼잡도와 평균 임대료와의 관계 그래프이다.

그래프로만 봤을 때에도 6개의 시간대 모두 혼잡도와 임대료는 관계가 없는 것으로 나타났다.

## 8. 역사 심도에 따른 임대료의 차이

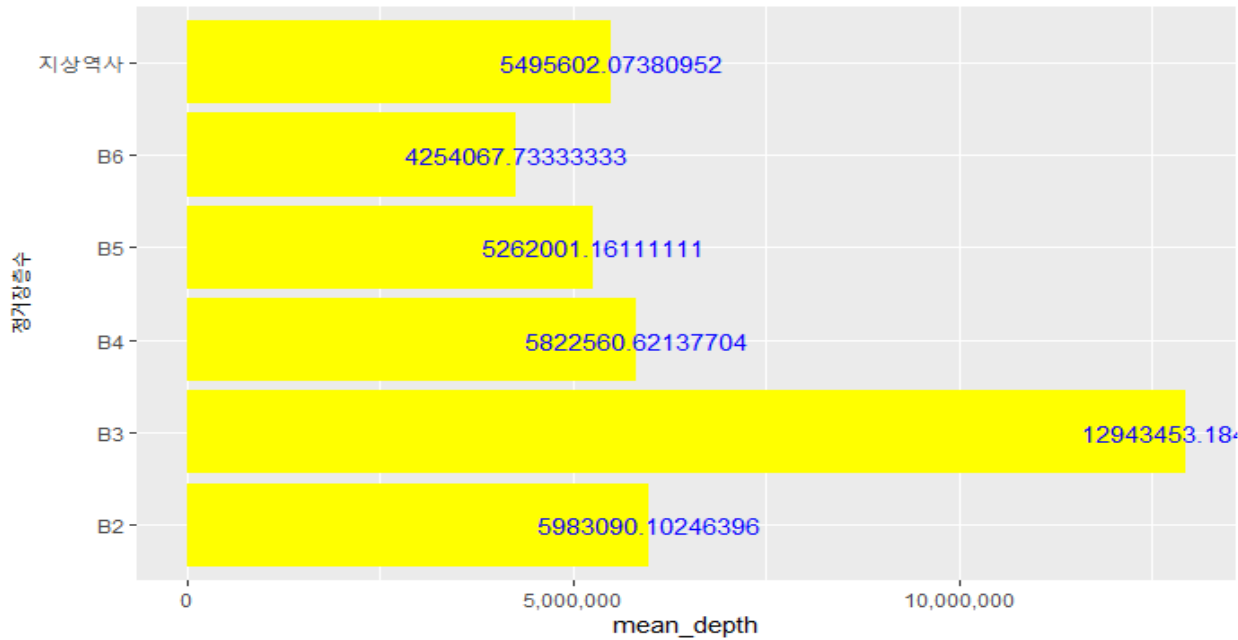
정거장이 몇 층인지에 따라 임대료가 차이가 있는지 분석한다.

정거장 층수끼리 묶어 평균임대료를 구하여 분석한다.

B1, B8, 고가, 지상의 경우에는 아예 없거나 한 곳밖에 없기 때문에 제외한다.



## 8. 역사 심도에 따른 임대료의 차이



정거장 층수에 따라 전부 비슷한 평균임대료 값을 보여주고 있다.

하지만 B3층은 다른 정거장 층수에 비해 평균임대료가 높은 것으로 나타나고 있다.

## 9. 환승여부에 따른 임대료의 차이

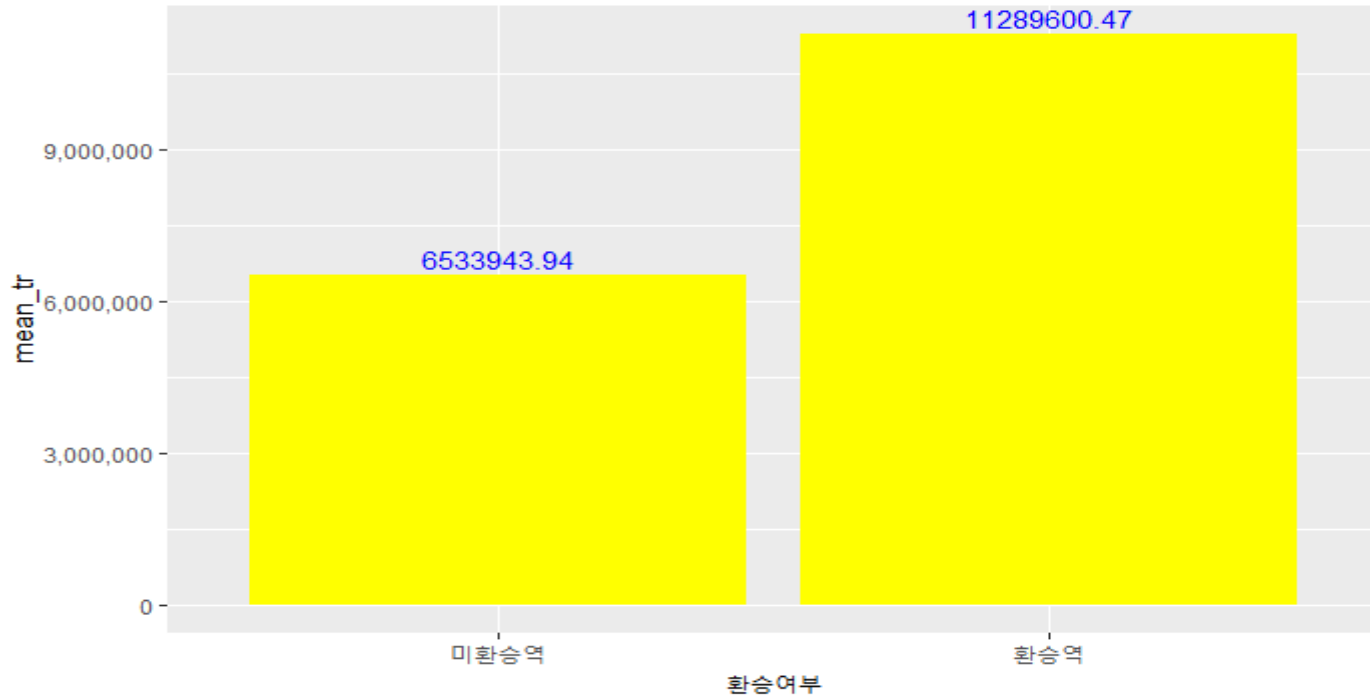
### 환승역 여부에 따른 임대료의 차이

환승역 여부에 따른 임대료의 차이가 있는지를 분석한다.

역명 중에 “\_[0-9]”가 들어가 있으면 환승역 아니면 미환승역으로 구분한다.

환승역과 미환승역 중에 이상치나 결측치는 없는 것으로 확인했다.

## 9. 환승여부에 따른 임대료의 차이



환승역이 미환승역보다 약 2배정도의 평균 임대료를 내고 있다.

정규분포이고 등분산이기 때문에 t.test결과 환승여부에 따라 차이는 없다고 볼 수 없다.

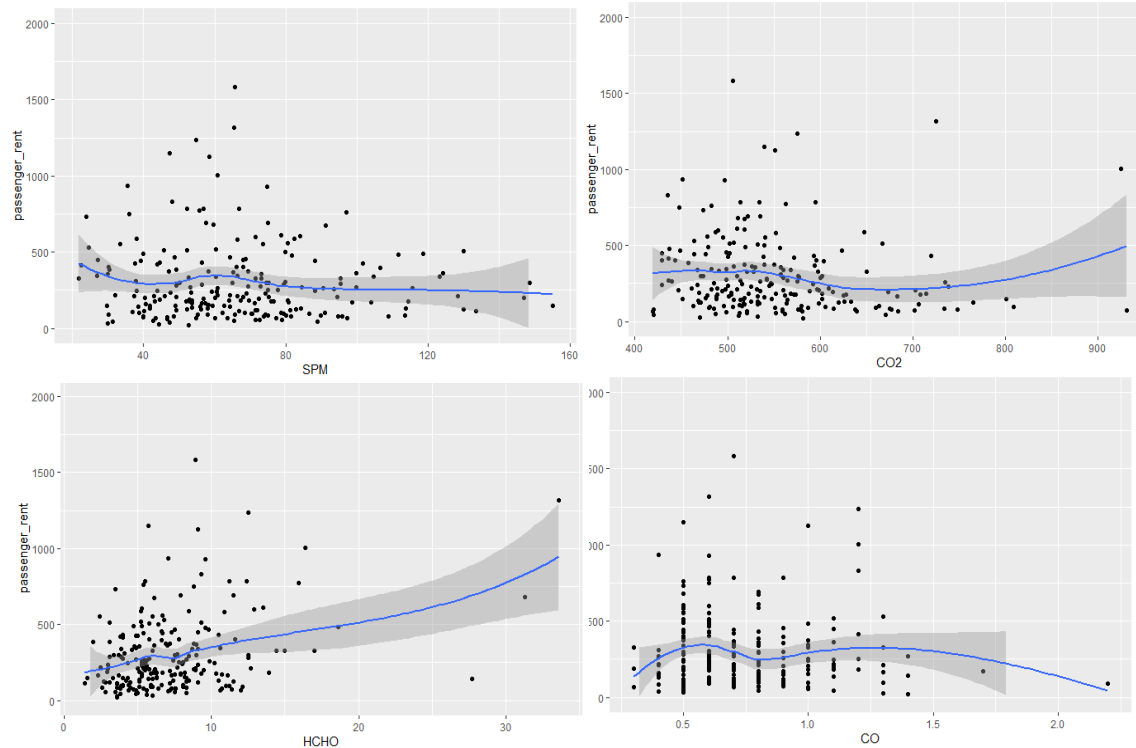
## 10. 공기질에 따른 임대료의 차이

공기질이 임대료의 영향을 미치는지를 분석한다.

공기질은 미세먼지(SPM), 이산화탄소(CO<sub>2</sub>), 일산화탄소(CO), HCHO(포름알데히드) 총 4가지 종류를 분석한다.

공기질과 관련한 이상치나 결측치는 없는것으로 확인했다.

## 10. 공기질에 따른 임대료의 차이



포름알데히드(HCHO)는 약간의 양의 관계가 있는 것으로 보인다.

하지만 나머지 공기질과 관련된 것은 관계가 없는 것으로 나타난다.

## 느낀점

1. 데이터를 수집하고 전처리하는 데 까지 걸리는 시간을 넉넉히 잡았다고 생각했음에도 시간이 부족했다.
2. 이로 인해 다 마무리 하지 못 한 것이 아쉬웠고, 더 추가하고 싶었던 것들이 있었는데 미처 하지 못해 아쉬웠다.
3. 다음 프로젝트 때에는 데이터 수집과 전처리 시간을 예상 시간보다도 더 길게 잡아 시간에 쫓기지 않도록 해야겠다고 느꼈다.