

day3

공공 데이터 분석

2019 경북고등학교

#1. 공공 데이터 분석 1 : 우리 동네 인구구조 분석

#2. 공공 데이터 분석 2 : 지하철 데이터 분석

#3. 나의 공공 데이터 분석 프로젝트 팀구성 및 주제 선정
(플튀 금지!!! feat.이로운 선생님 적극 활용!!)

#1. 공공데이터 분석

- 우리 동네 인구 구조 분석

[step1. 호기심 발견하기]

우리 동네에는 내 또래가 많지 않아 보이는데,
우리 동네의 연령별 인구분포는 어떻게 될까?

‘대치동’ Vs. ‘중계동’ 두 동네의 인구분포는 같을까, 다를까?

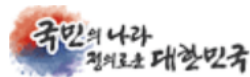
제주도에는 진짜 여자가 남자보다 많을까?

[step2. 데이터 모으기]

행정안전부(<https://www.mois.go.kr/>) : 정책자료-통계-주민등록 인구통계



국가상징
알아보기



누리집 안내지도 | 어린이 | ENGLISH



정보공개	참여·민원	뉴스·소식	정책자료	업무안내	기관소개	
정보공개 홈 정보공개안내 정보공개청구 사전정보공개 사전정보공표 공공데이터 개방 지방물가정보	참여·민원 홈 이용안내 온라인 민원신청 국민참여 적극행정 규제혁신 신고센터	뉴스·소식 홈 새소식 보도자료 지자체소식 홍보자료	정책자료 홈 주요업무계획 법령정보 전자관보 통계 통계연보·주제별 통계 승인통계 e-나라지표 주민등록 인구통계 간행물 정책연구보고서 참고자료	업무안내 홈 기획조정실 정부혁신조직실 전자정부국 지방자치분권실 지방재정경제실 안전정책실 재난관리실 재난협력실 비상대비정책국 장차관직속	기관소개 홈 행정안전부 안내 조직안내 열린장관실 찾아오시는 길	

주민등록 인구통계

> 주민등록 인구 및 세대현황

> **연령별 인구현황**

> 주민등록 인구 기타현황

연령별 인구현황

통계표

그래프

행정구역 전국 ▼ 시·군·구 ▼ i

등록구분 전체 ▼ i

조회기간 ☒ 월간 ☐ 연간 2019년 ▼ 12월 ▼ ~ 2019년 ▼ 12월 ▼ i

※매월 말일 작성 / 공표일시: 매월 1일 12시 이후(공표일이 주말, 공휴일인 경우에는 다음 평일에 공표)

구분 ☒ 계 ☐ 남·여 구분

정렬순서 행정기관코드 ▼ 오름차순 ▼

연령 구분 단위 1세 ▼

만 연령구분 0 ▼ 100이상 ▼

검색 🔍

초기화 ↺

● 현재화면

● 전체시군구현황

● 전체읍면동현황

CSV 파일 다운로드

XLSX 파일 다운로드

✓ 연령별 인구현황

행정기관									
	총 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세
전국	51,849,861	51,849,861	295,132	330,970	362,900	412,018	444,102	440,483	440,483
서울특별시	9,729,107	9,729,107	51,145	54,779	59,128	65,734	70,013	69,139	69,139
부산광역시	3,413,841	3,413,841	16,478	19,173	21,401	24,411	26,500	26,130	26,130
대구광역시	2,438,031	2,438,031	12,896	14,564	16,408	18,791	20,148	20,044	20,044
인천광역시	2,957,026	2,957,026	17,616	19,351	20,869	23,781	25,940	25,667	25,667
광주광역시	1,456,468	1,456,468	8,189	9,440	10,660	12,333	13,310	13,487	13,487
대전광역시	1,474,870	1,474,870	8,221	9,361	10,659	12,064	13,119	12,880	12,880
울산광역시	1,148,019	1,148,019	7,324	8,129	9,158	10,584	11,322	11,305	11,305

[step2. 데이터 모으기 : 내려받은 파일 확인하기]

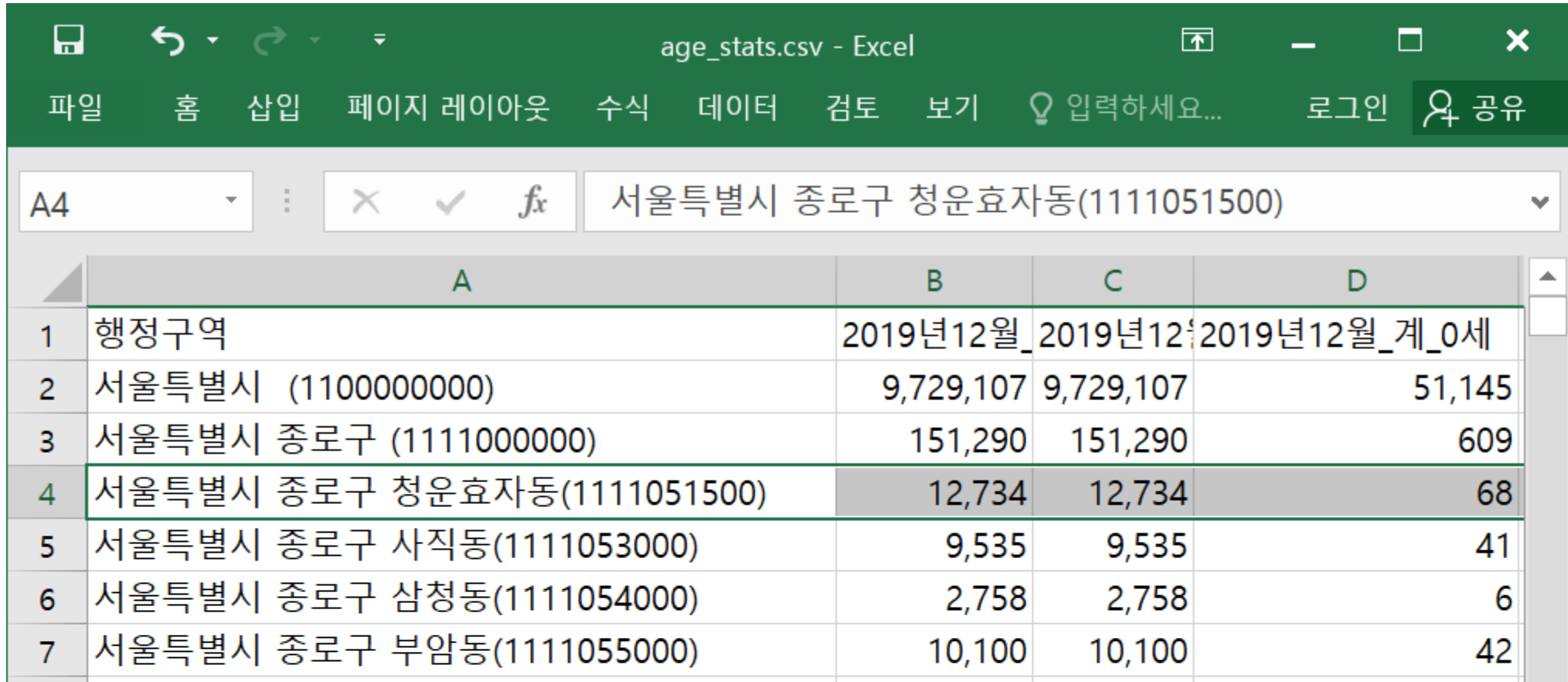
201912_201912_연령별인구현황_월간.csv - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 입력하세요... 로그인 공유

	A	B	C	D	E	F	G	H	I
1	행정구역	2019년12월	2019년12월	2019년12월	2019년12월	2019년12월	2019년12월	2019년12월	2019년12월
2	서울특별시	9,729,107	9,729,107	51,145	54,779	59,128	65,734	70,013	69,139
3	서울특별시	151,290	151,290	609	590	702	765	874	877
4	서울특별시	12,734	12,734	68	64	79	78	96	96
5	서울특별시	9,535	9,535	41	44	48	68	67	75
6	서울특별시	2,758	2,758	6	9	14	4	15	9
7	서울특별시	10,100	10,100	42	37	49	44	63	59
8	서울특별시	18,659	18,659	79	83	104	100	132	156
9	서울특별시	8,695	8,695	46	32	71	75	74	85
10	서울특별시	10,562	10,562	87	81	89	98	114	74
11	서울특별시	4,383	4,383	11	15	16	18	20	18

[step3. 데이터 살펴보기 및 정제하기]

- 파일명 변경 : age_stats.csv
- 우리동네 데이터 살펴보기 : “청운효자동”



	A	B	C	D
1	행정구역	2019년12월_	2019년12월_	2019년12월_계_0세
2	서울특별시 (1100000000)	9,729,107	9,729,107	51,145
3	서울특별시 종로구 (1111000000)	151,290	151,290	609
4	서울특별시 종로구 청운효자동(1111051500)	12,734	12,734	68
5	서울특별시 종로구 사직동(1111053000)	9,535	9,535	41
6	서울특별시 종로구 삼청동(1111054000)	2,758	2,758	6
7	서울특별시 종로구 부암동(1111055000)	10,100	10,100	42

[step4. 데이터 분석으로 호기심 해결하기]

우리 동네에는 내 또래가 많지 않아 보이는데,
우리 동네의 연령별 인구분포는 어떻게 될까?

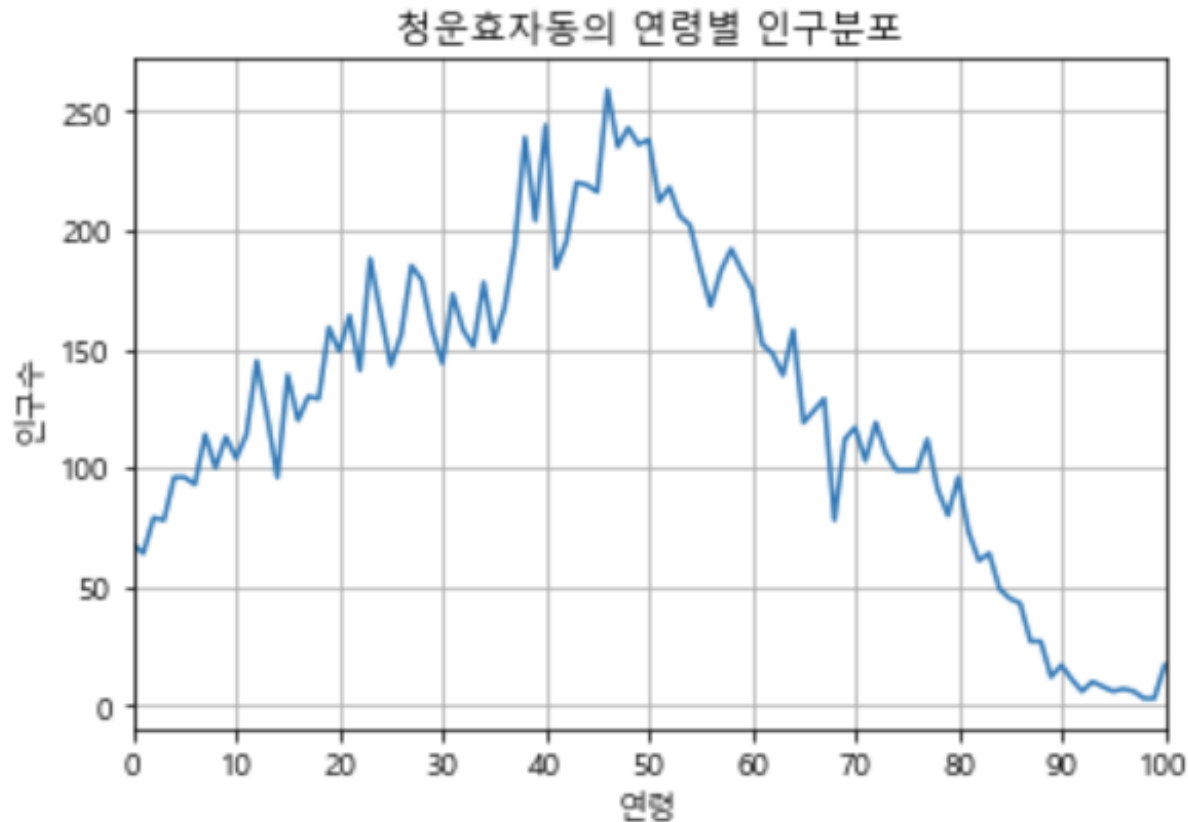
‘대치동’ Vs. ‘중계동’ 두 동네의 인구분포는 같을까, 다를까?

제주도에는 진짜 여자가 남자보다 많을까?

[step4. 데이터 분석으로 호기심 해결하기]

- 우리 동네의 연령별 인구분포는 어떻게 될까?

인구 구조가 궁금한 동(읍면동 단위) 입력 : 청운효자동



[step4. 데이터 분석으로 호기심 해결하기]

우리 동네에는 내 또래가 많지 않아 보이는데,
우리 동네의 연령별 인구분포는 어떻게 될까?

‘대치동’ Vs. ‘중계동’ 두 동네의 인구분포는 같을까, 다를까?

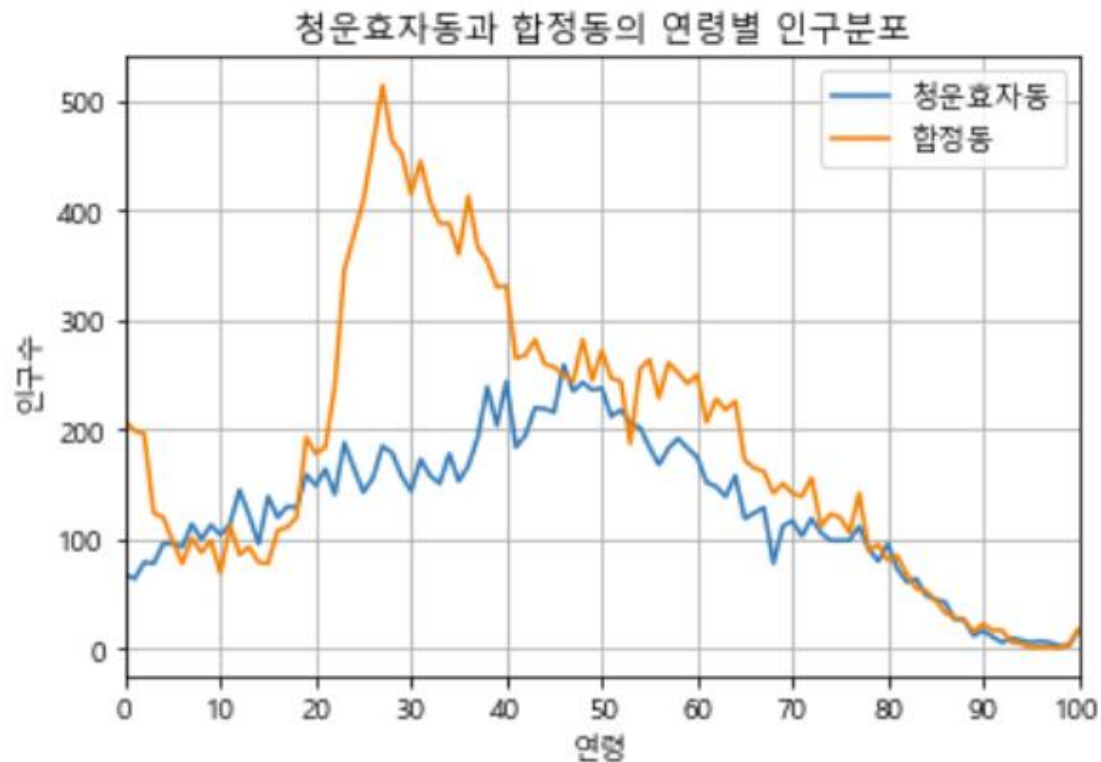
제주도에는 진짜 여자가 남자보다 많을까?

[step4. 데이터 표현하기]

- 대치동 vs. 중계동? 비교하고 싶은 2개 동네의 인구 분포를 그리자

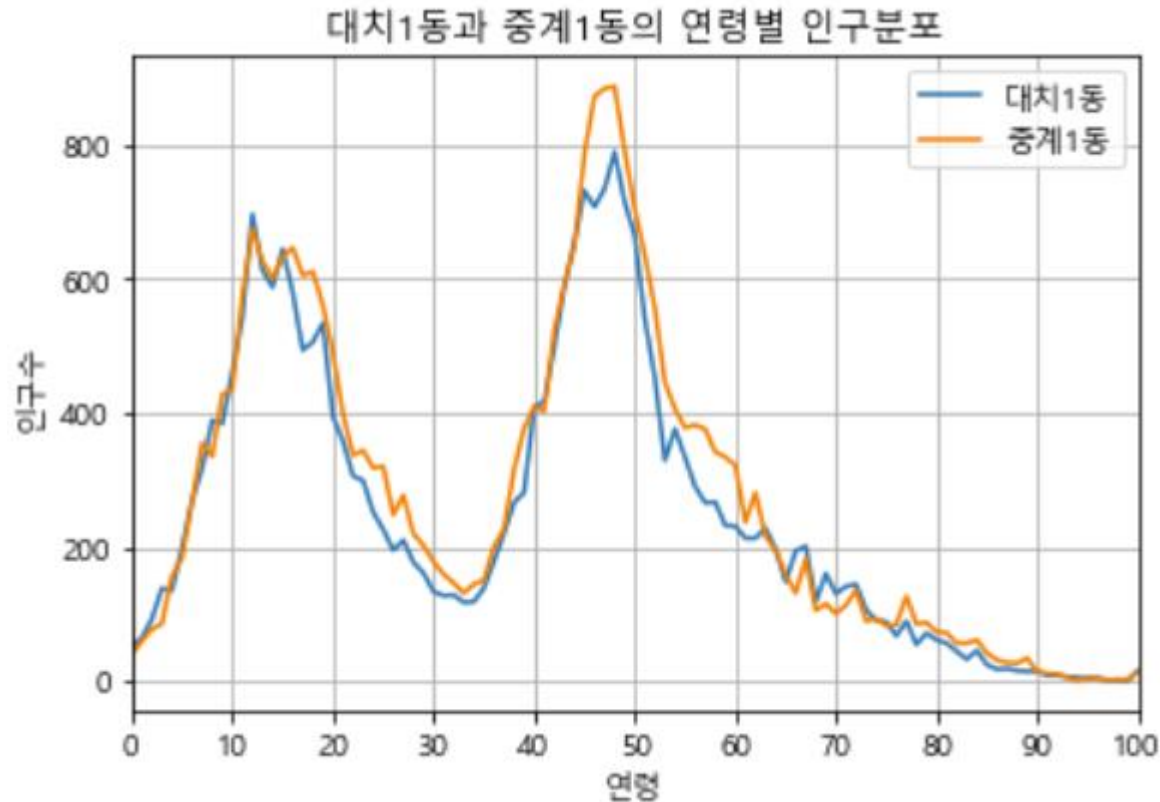
인구 구조가 궁금한 첫번째 동(읍면동 단위) 입력 : 청운효자동

인구 구조가 궁금한 두번째 동(읍면동 단위) 입력 : 합정동



[step4. 데이터 표현하기]

- 대치동 vs. 중계동? 비교하고 싶은 2개 동네의 인구 분포를 그리자



[step4. 데이터 분석으로 호기심 해결하기]

우리 동네에는 내 또래가 많지 않아 보이는데,
우리 동네의 연령별 인구분포는 어떻게 될까?

‘대치동’ Vs. ‘중계동’ 두 동네의 인구분포는 같을까, 다를까?

제주도에는 진짜 여자가 남자보다 많을까?

주민등록 인구통계

> 주민등록 인구 및 세대현황

> **연령별 인구현황**

> 주민등록 인구 기타현황

연령별 인구현황

통계표

그래프

행정구역

전국 ▼

시·군·구 ▼



등록구분

전체 ▼



조회기간

☒ 월간 ☐ 연간

2019년 ▼

12월 ▼

~ 2019년 ▼

12월 ▼



※매월 말일 작성 / 공표일시 : 매월 1일 12시 이후(공표일이 주말, 공휴일인 경우에는 다음 평일에 공표)

구분

☐ 계 ☒ 남·여 구분

정렬순서

행정기관코드 ▼

오름차순 ▼

연령 구분 단위

1세 ▼

만 연령구분

0 ▼

100이상 ▼

검색 🔍

초기화 ↺

● 현재화면

● 전체시군구현황

● 전체읍면동현황

csv 파일 다운로드

xlsx 파일 다운로드

연령별 인구현황

행정기관									
	남 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세
전국	25,864,816	25,864,816	151,494	169,815	186,836	210,812	227,583	225,632	221,494
서울특별시	4,744,059	4,744,059	26,523	28,466	30,328	33,582	35,843	35,417	38,417
부산광역시	1,675,417	1,675,417	8,431	9,816	11,042	12,422	13,651	13,530	14,417
대구광역시	1,205,286	1,205,286	6,726	7,389	8,457	9,611	10,329	10,226	10,417
인천광역시	1,482,249	1,482,249	9,118	9,922	10,804	12,058	13,329	13,115	13,417
광주광역시	720,740	720,740	4,131	4,888	5,499	6,284	6,865	6,924	7,417
대전광역시	736,607	736,607	4,161	4,799	5,515	6,212	6,746	6,674	6,917
울산광역시	589,712	589,712	3,725	4,152	4,738	5,512	5,829	5,816	5,917

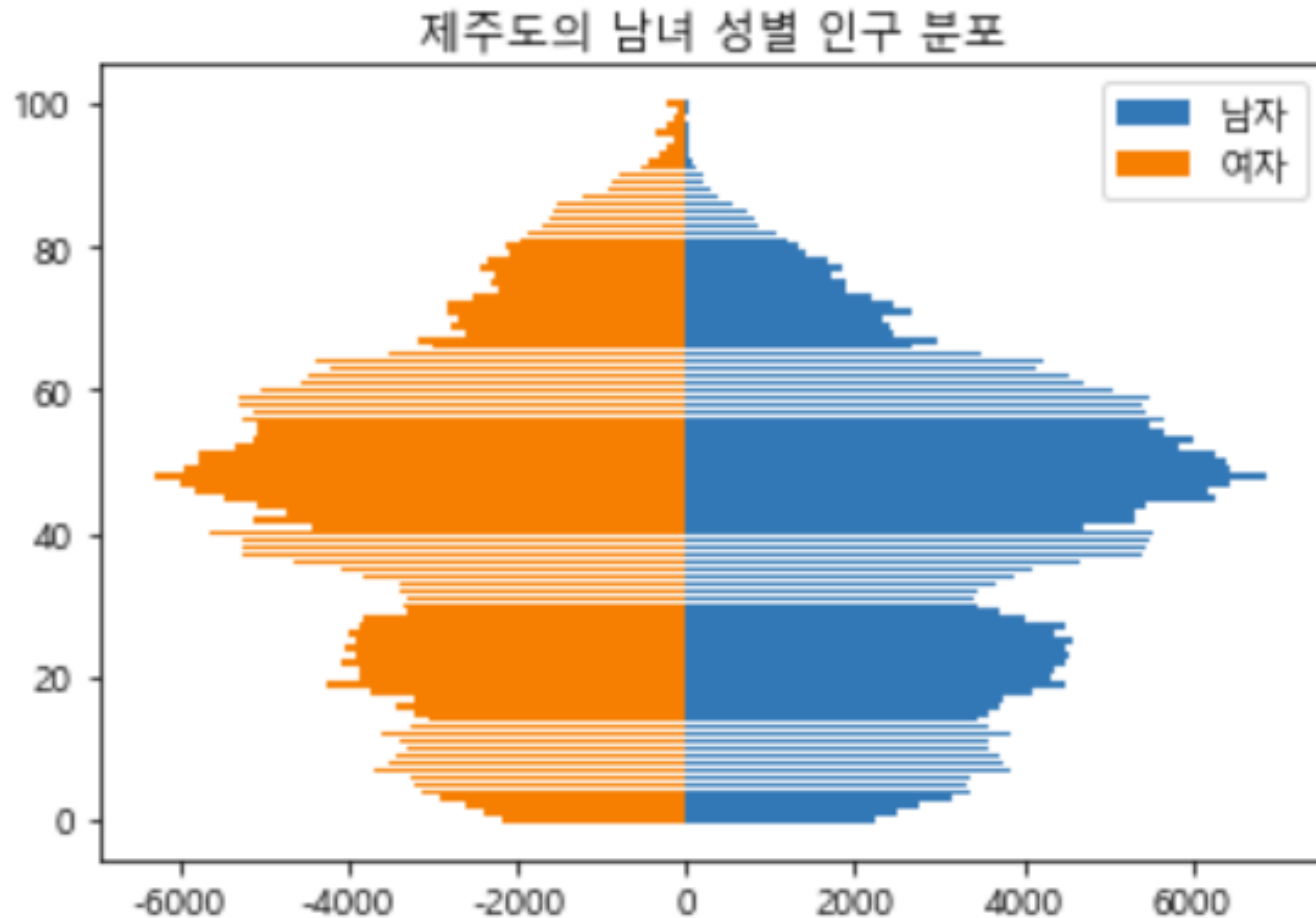
[step3. 데이터 살펴보기]

- 파일명 변경 : gender_age_stats.csv
- 데이터의 구조

	남성 연령별 인구수								여성 연령별 인구수						
열이름	지역명	총인구1	총인구2	0세	1세	...	99세	100세 이상	총인구1	총인구2	0세	1세	...	99세	100세 이상
인덱스	0	1	2	3	4	...	102	103	104	105	106	107	...	205	206

[step4. 데이터 표현하기-시도1 : *barh* 그래프]

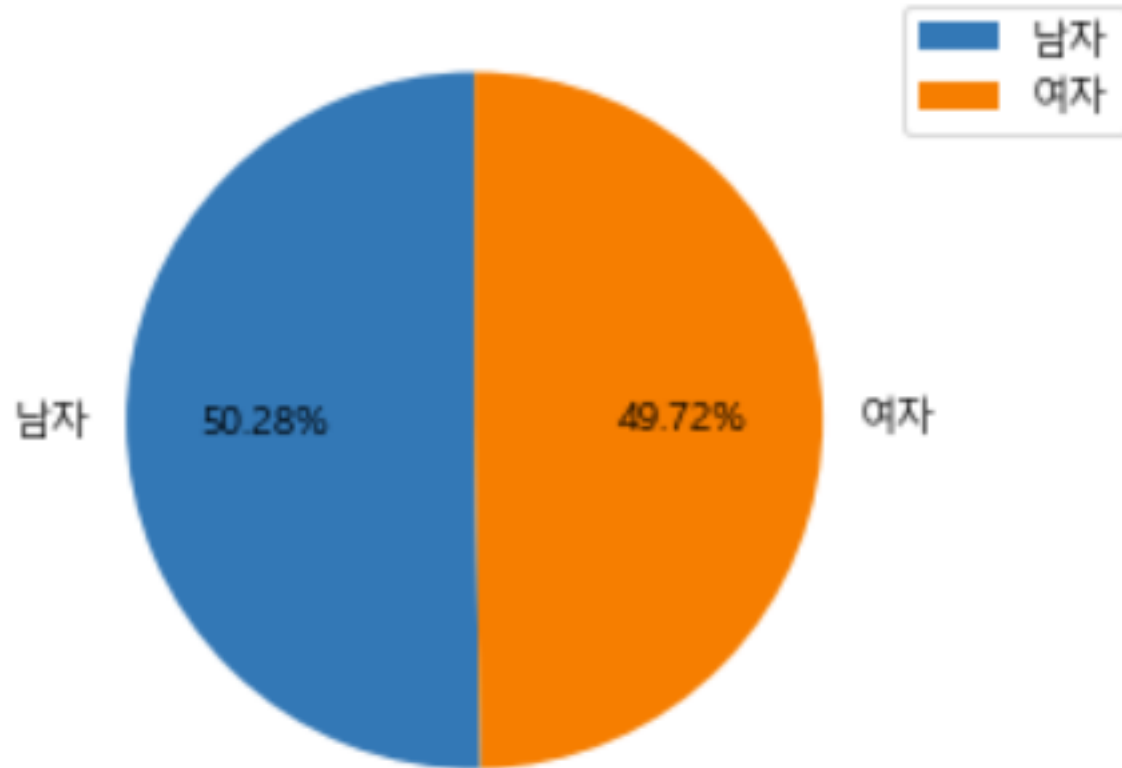
- 제주도에는 진짜 여자가 남자보다 많을까?



[step4. 데이터 표현하기-시도1 : pie 그래프]

- 제주도에는 진짜 여자가 남자보다 많을까?

제주도의 남녀 성별 인구 분포



[step4. 데이터 표현하기-시도1 : 선 그래프]

- 제주도에는 진짜 여자가 남자보다 많을까?



#2. 공공데이터 분석

- 지하철 데이터 분석

[step1. 호기심 발견하기]

출근 시간대에 사람들이 가장 많이 타고 내리는 역은 어디일까?

시간대 별로 가장 많은 사람들이 승하차 하는 역은 어디일까?

갑자기 승하차 인원이 늘어난 역은 어디일까?

[step2. 데이터 모으기]

티머니(<https://pay.tmoney.co.kr/>)

티머니카드&페이 홈

이용안내

이용안내 홈

교통카드 통계자료

월간 교통카드 통계자료를 확인할 수 있습니다.

제목	2019년 12월 교통카드 통계자료		
첨부파일	2019년 12월 교통카드 통계자료.xls	첨부파일 다운로드 >	등록일
			2020.01.03

이용안내

그 동안 행정기관 및 연구기관이 공공의 목적으로 이용하는 경우에만 제한적으로 제공되던 교통카드 자료를 서울시의 '열린시정 2.0'에 따라 공개하고자 합니다.

누구든지 서울시 홈페이지(<http://data.seoul.go.kr/index.jsp>)와 티머니홈페이지(<http://www.t-money.co.kr>)를 통해 쉽고 편리하게 이용할 수 있습니다.

제공 자료는 '지하철 노선별/역별 승하차인원'과 버스 노선별/정류장별 승하차인원'으로 매월 업데이트 할 예정입니다.

<세부 내용>

- 대상 : 지하철(서울교통공사, 한국철도공사, 공항철도, 9호선) 버스 (서울버스, 서울마을버스)
- 산출 기준 : 선후불교통카드 및 1회용 교통카드 사용 기준

[step3. 데이터 살펴보기 및 정제하기]

2019년 12월 교통카드 통계자료.xls [호환 모드] - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 ? 수행할 작업을 알려 주세요. 로그인 공유

A1 : 사용월

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	사용월	노선ID	노선번호	노선명	버스정류장	버스정류장	승차승객수	하차승객수	작업일시					
2	2019-12	11110897	0017	0017번(청	8501292	청암자이아	581	8	2020-01-03 12:38:45					
3	2019-12	11110897	0017	0017번(청	9003185	청암동강변	1,762	32	2020-01-03 12:38:45					
4	2019-12	11110897	0017	0017번(청	0072983	청심경로당	3,279	45	2020-01-03 12:38:45					
5	2019-12	11110897	0017	0017번(청	0072984	원호2동주	5,627	186	2020-01-03 12:38:45					
6	2019-12	11110897	0017	0017번(청	0072985	산천동	7,544	345	2020-01-03 12:38:45					
7	2019-12	11110897	0017	0017번(청	0072986	신창동세병	6,945	235	2020-01-03 12:38:45					
8	2019-12	11110897	0017	0017번(청	0072987	남이장군사	1,717	1,703	2020-01-03 12:38:45					
9	2019-12	11110897	0017	0017번(청	0072988	새마을금고	4,616	3,557	2020-01-03 12:38:45					
10	2019-12	11110897	0017	0017번(청	0033066	용문시장	5,962	1,795	2020-01-03 12:38:45					
11	2019-12	11110897	0017	0017번(청	0072990	용산전자상	2,779	1,118	2020-01-03 12:38:45					
12	2019-12	11110897	0017	0017번(청	0072991	신용산지하	1,126	2,767	2020-01-03 12:38:45					
13	2019-12	11110897	0017	0017번(청	0072992	용산역	34,097	23,759	2020-01-03 12:38:45					
14	2019-12	11110897	0017	0017번(청	0072314	용산푸르지	871	2,075	2020-01-03 12:38:45					
15	2019-12	11110897	0017	0017번(청	0072993	한강대교북	6,958	5,105	2020-01-03 12:38:45					
16	2019-12	11110897	0017	0017번(청	0072994	서부이촌동	2,232	13,199	2020-01-03 12:38:45					
17	2019-12	11110897	0017	0017번(청	0072995	이촌2동대	1,211	11,854	2020-01-03 12:38:45					
18	2019-12	11110897	0017	0017번(청	0072996	이촌2동주	981	6,353	2020-01-03 12:38:45					
19	2019-12	11110897	0017	0017번(청	0072997	성춘공원3	429	2,210	2020-01-03 12:38:45					
20	2019-12	11110897	0017	0017번(청	0073000	이촌119와	311	2,408	2020-01-03 12:38:45					

버스정류장별 이용현황 | 지하철 노선별 역별 이용현황 | 지하철 유무임별 이용현황 | 지하철 시간대별 이용현황

준비

[step3. 데이터 살펴보기 및 정제하기]

- 파일명 변경 : 지하철시간대별 이용현황 시트 선택 -> subway.csv
- 데이터의 구조

사용월	지역명	호선명	역 ID	역 이름	04:00:00 ~04:59:59		05:00:00 ~05:59:59		06:00:00 ~06:59:59		07:00:00 ~07:59:59		08:00:00 ~08:59:59		09:00:00 ~09:59:59		...
					승차	하차	승차	하차	승차	하차	승차	하차	승차	하차	승차	하차	..
인덱스	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	..

[step4. 데이터 분석으로 호기심 해결하기]

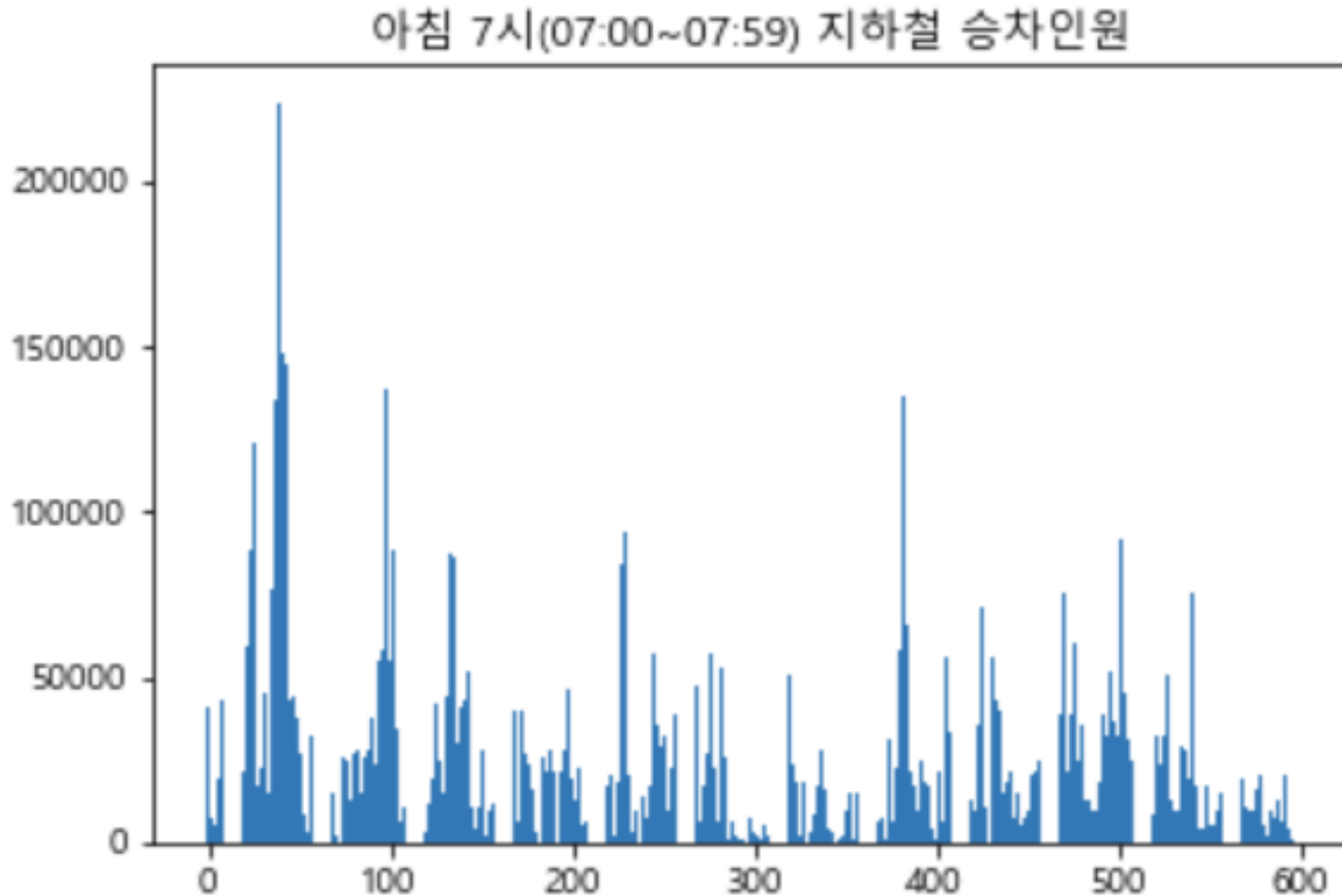
출근 시간대에 사람들이 가장 많이 타고 내리는 역은 어디일까?

시간대 별로 가장 많은 사람들이 승하차 하는 역은 어디일까?

갑자기 승하차 인원이 늘어난 역은 어디일까?

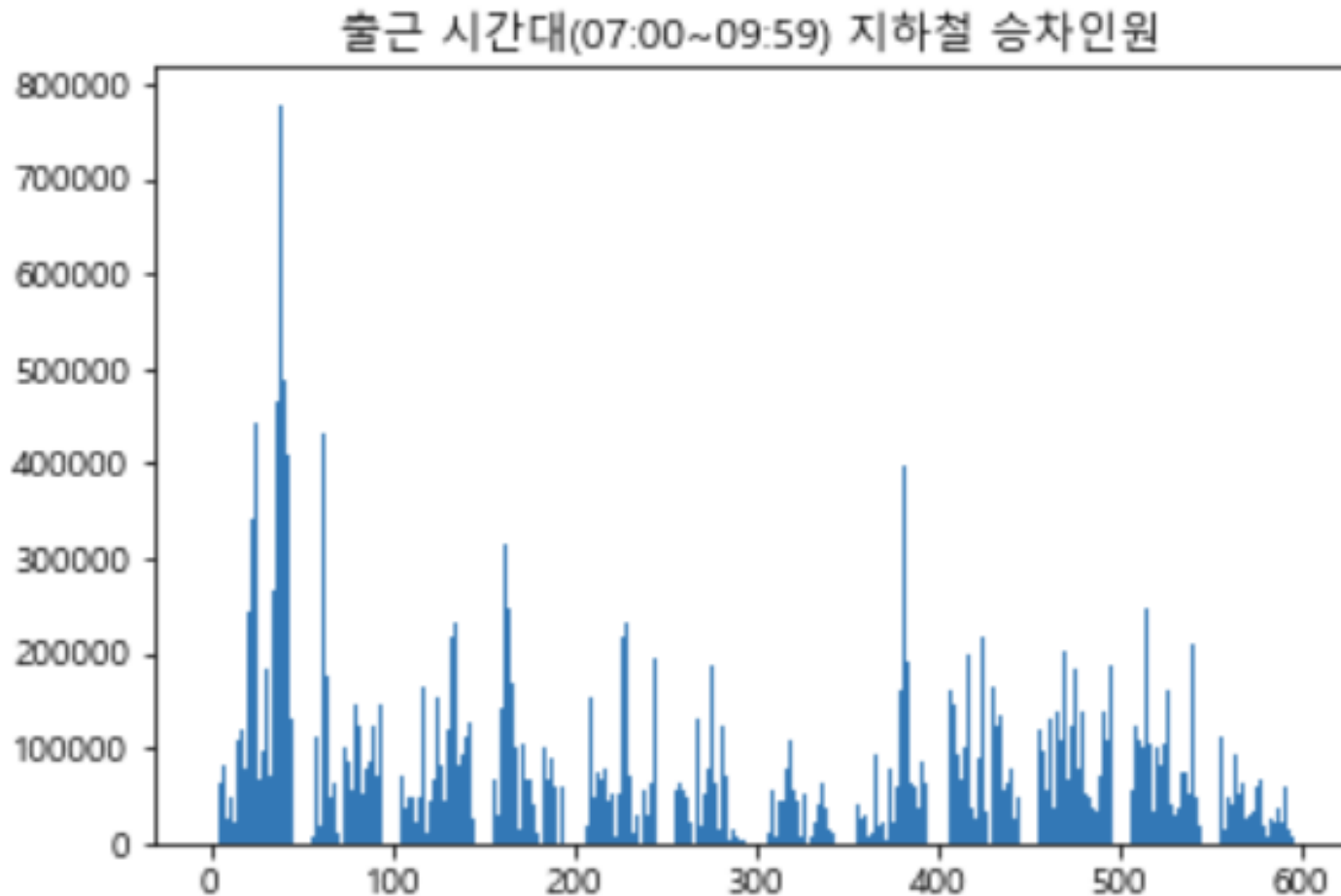
[step4. 데이터 표현하기-시도1 : 아침 7시의 승차인원그래프]

- 출근 시간대에 사람들이 가장 많이 타는 역은 어디일까?



[step4. 데이터 표현하기-시도2 : 아침 7시~9시의 그래프]

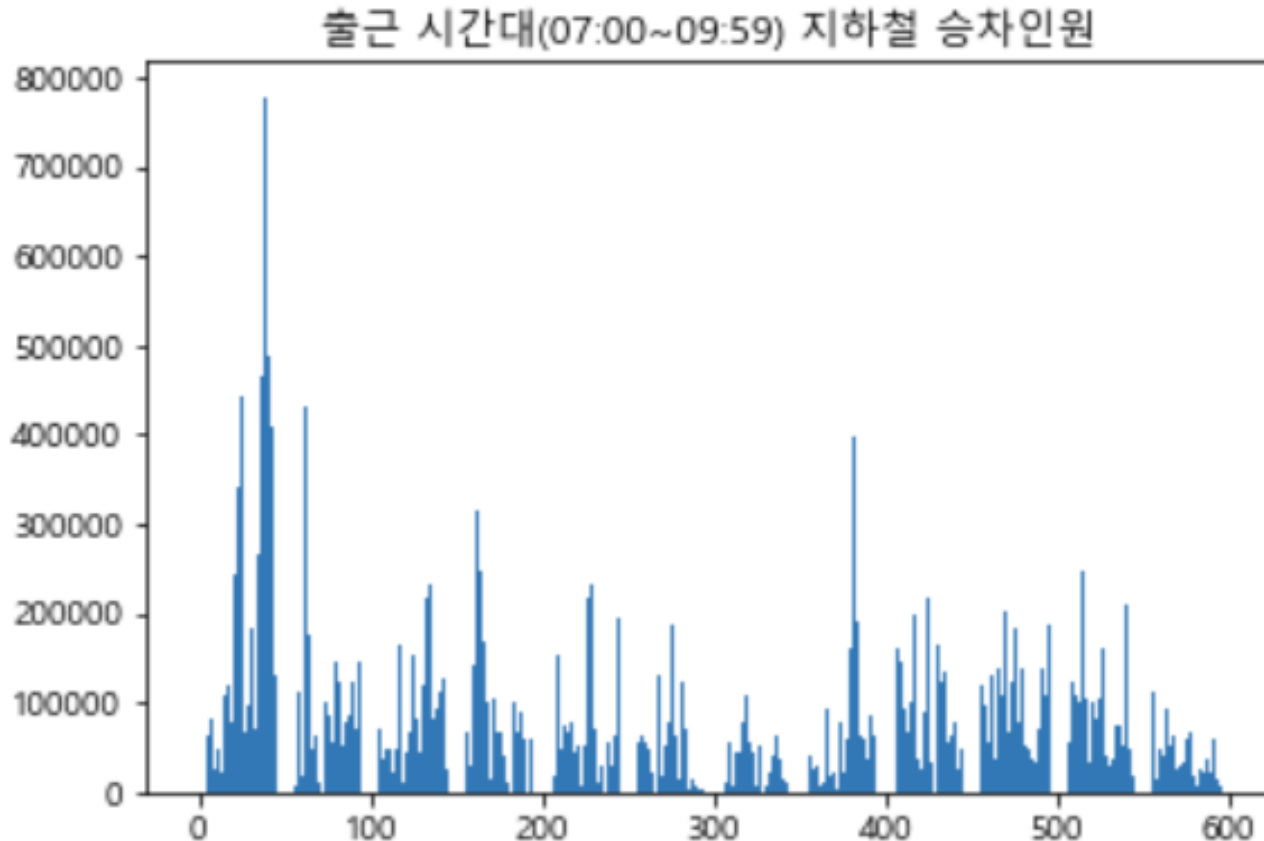
- 출근 시간대에 사람들이 가장 많이 타는 역은 어디일까?



[step4. 데이터 표현하기-시/도4: 아침 7시~9시의 **승차 최대값**]

- 출근 시간대에 사람들이 가장 많이 타는 역은 어디일까?

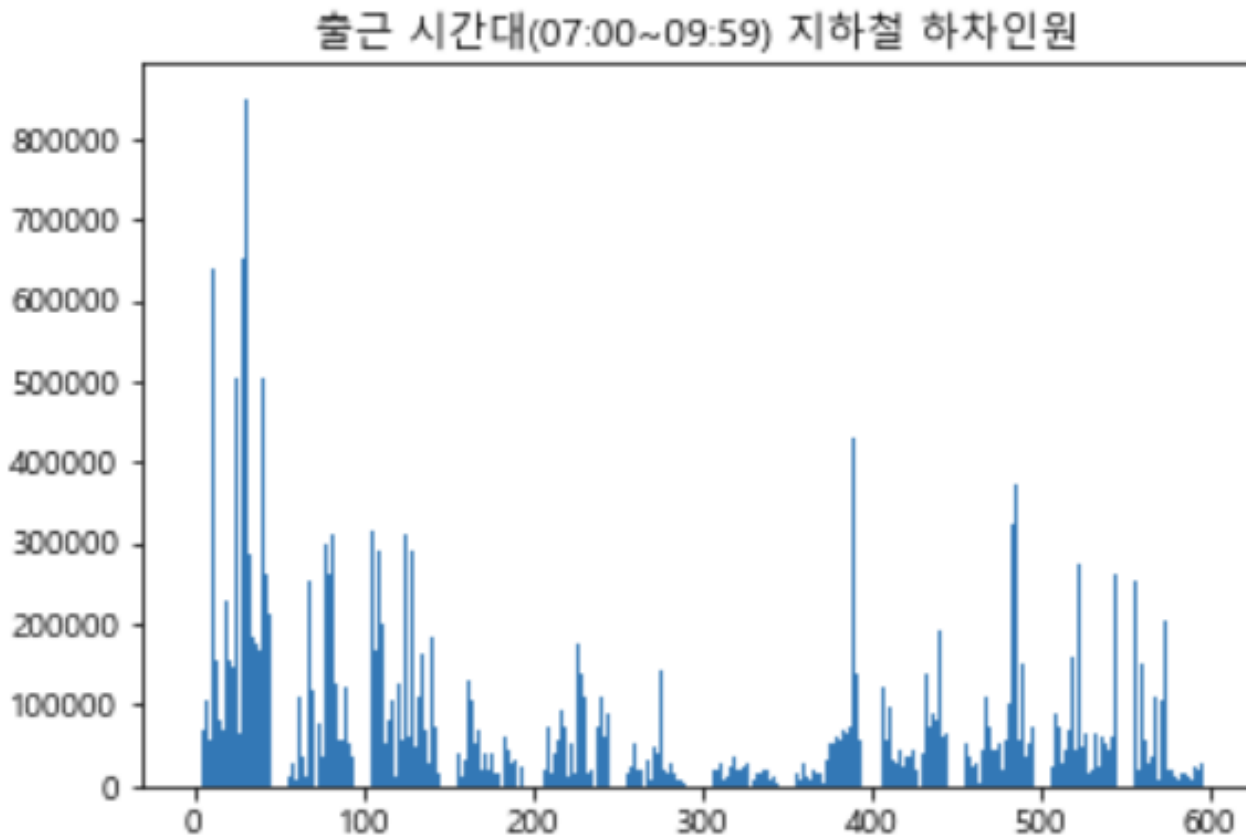
7~9시 최대 승차인원역 : 신림 (승차인원 : 778970)



[step4. 데이터 표현하기-시/도4: 아침 7시~9시의 **하차 최대값**]

- 출근 시간대에 사람들이 가장 많이 내리는 역은 어디일까?

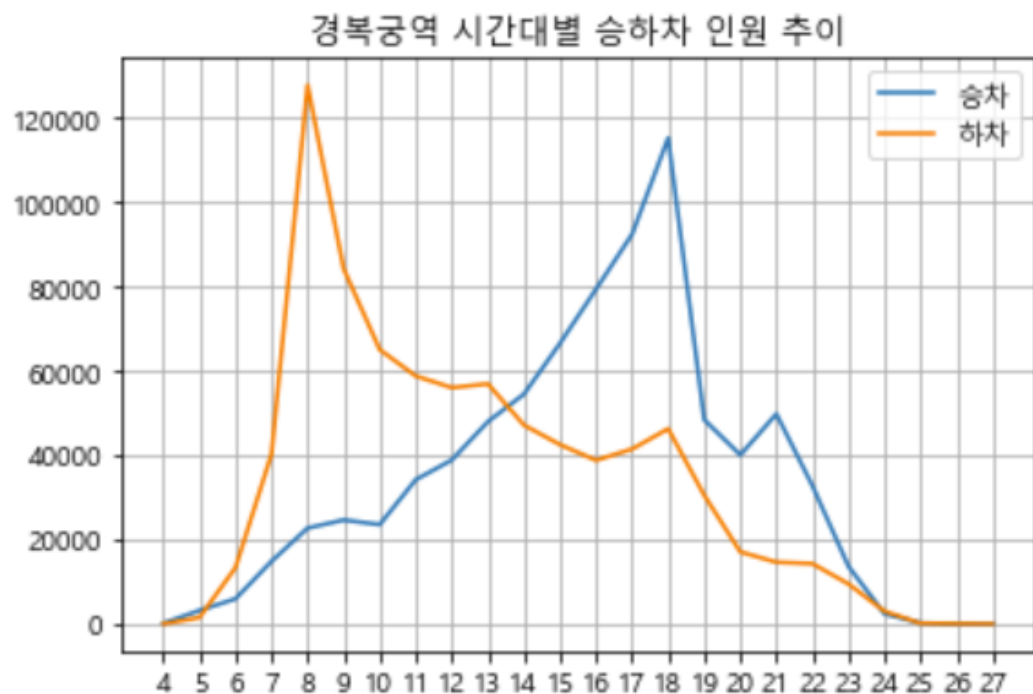
7~9시 최대 하차인원역 : 강남 (하차인원 : 850106)



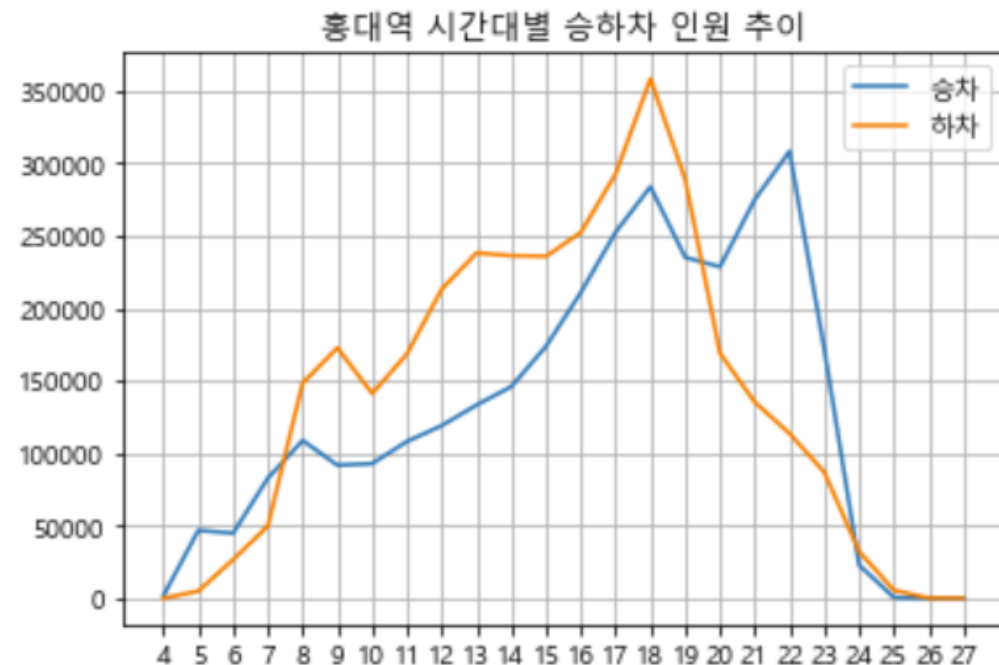
[step4. 데이터 표현하기]

- 지하철역의 시간대별 승하차 인원 추이

지하철역명을 입력하세요 : 경복궁



지하철역명을 입력하세요 : 홍대



#3. 나의 공공 데이터 분석 프로젝트 팀구성 및 주제 선정 (플튀 금지!!! feat.이로운 선생님 적극 활용!!)

1/13(월) 발표

개인 혹은 팀(최대 2명)

제출물 : 발표자료(ppt),

데이터분석 소스(학번_이름_주제명.ipynb 또는 .py)

리소스(코드를 실행시키는 데 필요한 모든 자료)

프로젝트 발표에 반드시 포함시킬 내용

1. 주제 선정 과정

우리 팀의 호기심, 질문했던 것을 모두 나열
선정된 주제 및 주제 선정의 이유

2. 데이터 처리 과정

3. 데이터 시각화 후 분석 내용

4. 프로젝트하면서 느낀점

#3. 나의 공공 데이터 분석 프로젝트 팀구성 및 주제 선정 (플튀 금지!!! feat.이로운 선생님 적극 활용!!)

구글 form에 선정한 주제를 적어주세요.

<https://forms.gle/8hqwgzYAbBFTCnTa9>

