

- 데이터 전처리하기
- 출근 시간대 사람들이 가장 많이 타고 내리는 역은?
- 밤 11시에 사람들이 가장 많이 타는 역은?
- 시간대별로 사람들이 가장 많이 타고 내리는 역은?
- 모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

데이터 전처리하기

- 지하철 시간대별 이용현황 데이터

2019년 01월 교통카드 통계자료.xls [호환 모드] - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 추가 기능 팀 수행할 작업을 알려 주세요. 김 윤지 공유

블러블기 클립보드 글꼴 맞춤 표시 형식 스타일 셀

일반 조건부 서식 표 스타일 삽입 삭제 서식 정렬 및 필터 찾기 및 선택 편집

P29 36,450

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	사용월	호선명	역ID	지하철역	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						
2					승차	하차	승차	하차	승차	하차	승차	하차	승차	하차	승차	하차
3	2019-01	1호선	0150	서울역	927	22	10,682	8,774	13,504	45,545	39,110	103,863	80,492	236,747	61,108	163
4	2019-01	1호선	0151	시청	74	0	2,409	4,049	3,532	19,241	7,703	66,876	9,978	209,349	12,818	88
5	2019-01	1호선	0152	종각	100	8	4,409	5,105	4,593	26,969	6,733	104,259	11,028	300,081	14,563	167
6	2019-01	1호선	0153	종로3가	122	4	5,061	2,520	3,876	12,832	6,066	29,381	9,735	84,155	14,261	100
7	2019-01	1호선	0154	종로5가	43	2	1,934	3,590	2,935	14,991	5,143	40,376	8,509	132,808	12,652	78
8	2019-01	1호선	0155	동대문	621	7	15,328	2,164	10,315	6,673	14,893	12,227	24,080	27,692	23,381	28
9	2019-01	1호선	0156	신설동	287	18	8,847	2,242	9,805	8,507	20,077	23,708	30,505	77,052	21,746	42
10	2019-01	1호선	0157	제기동	316	8	4,874	2,284	8,388	8,242	22,615	18,537	34,862	50,424	24,786	43
11	2019-01	1호선	0158	청량리(서)	1,321	11	10,844	2,912	16,911	12,013	46,015	20,676	65,427	42,910	42,110	41
12	2019-01	1호선	0159	동묘앞	93	3	3,326	832	3,418	3,433	6,088	8,945	9,415	22,873	8,843	17
13	2019-01	2호선	0201	시청	87	0	1,197	1,808	2,206	15,659	5,822	66,277	8,873	254,330	12,013	95
14	2019-01	2호선	0202	율지로입	60	0	3,044	3,128	3,503	25,964	8,459	127,983	15,748	370,776	20,905	168
15	2019-01	2호선	0203	율지로3가	20	1	1,133	1,883	2,446	17,469	5,822	68,791	10,508	176,943	13,616	88
16	2019-01	2호선	0204	율지로4가	7	0	833	840	1,536	13,828	3,225	22,942	5,851	67,532	8,888	43
17	2019-01	2호선	0205	동대문역사	387	4	10,806	869	5,926	8,238	7,973	19,653	14,495	60,112	14,614	42
18	2019-01	2호선	0206	신당	33	2	6,309	1,181	11,451	8,459	27,835	16,020	47,992	40,276	28,853	27
19	2019-01	2호선	0207	상왕십리	28	3	5,806	716	12,635	5,440	38,199	11,057	68,229	40,639	36,040	22
20	2019-01	2호선	0208	왕십리(성	667	7	7,630	1,123	11,377	6,148	31,012	11,411	60,304	31,248	36,056	24
21	2019-01	2호선	0209	한양대	11	1	1,295	346	2,424	5,635	6,313	10,512	10,105	37,533	7,567	45
22	2019-01	2호선	0210	독섬	18	0	4,142	1,797	9,130	13,242	23,263	27,099	35,147	133,396	22,016	65
23	2019-01	2호선	0211	상수	66	12	5,882	4,037	10,844	25,840	25,458	66,913	36,540	280,949	27,967	104
24	2019-01	2호선	0212	건대입구	298	11	20,178	2,206	25,927	15,658	59,821	26,327	113,196	65,846	73,042	53
25	2019-01	2호선	0213	구의(광진	62	4	13,416	1,341	25,792	10,991	70,814	15,815	130,380	44,549	74,902	27
26	2019-01	2호선	0214	강변(동서	18	0	9,803	3,838	29,345	28,033	99,368	33,030	148,781	58,847	112,708	58
27	2019-01	2호선	0215	잠실나루	6	2	3,069	1,839	12,422	10,501	45,639	22,499	62,872	43,182	36,720	32
28	2019-01	2호선	0216	잠실(송파	100	3	12,409	5,576	47,262	38,078	130,638	91,126	194,415	204,093	135,416	190
29	2019-01	2호선	0217	잠실새내	18	0	5,401	900	17,709	8,553	59,461	17,391	89,078	31,935	55,152	36

준비

데이터 전처리하기

- 헤더 제외하고 map 함수를 활용한 데이터 타입 바꾸기(문자열 -> 정수)

```
next(data)
next(data)
for row in data :
    row[4:] = map(int, row[4:])
    print(row)
```

```
['2018-08', '1호선', '0150', '서울역', 838, 42, 11692, 8631, 14030, 49163, 39205, 109430,
79162, 234939, 60617, 159840, 65512, 87504, 83998, 77992, 94356, 84074, 95114,
84591, 79701, 80728, 108804, 81198, 113754, 86239, 159234, 92065, 244115, 103327,
138351, 77226, 97955, 55318, 93604, 46929, 70655, 35907, 35980, 19417, 3847, 7378,
3, 421, 0, 0, 0, 0]
(생략)
```

출근 시간대 사람들이 가장 많이 타고 내리는 역은?

- 아침 7시 승차 데이터 인덱스 확인

인덱스

0	1	2	3	4	5	6	7	8	9	10	11
사용월	호선명	역ID	역이름	4시		5시		6시		7시	
				승차	하차	승차	하차	승차	하차	승차	하차
2019-01	1호선	0150	서울역	927	22	10682	8774	13504	45545	399110	103863
2019-01	1호선	0151	시청	74	0	2409	1019	3532	19241	7702	66876

출근 시간대 사람들이 가장 많이 타고 내리는 역은?

- 아침 7시 승차 데이터 개수 및 인원수 출력하기

```
import csv
f = open('subwaytime.csv')
data = csv.reader(f)
next(data)
next(data)
result = []
for row in data :
    row[4:] = map(int, row[4:])
    result.append(row[10])
print(len(result))
print(result)
```

598

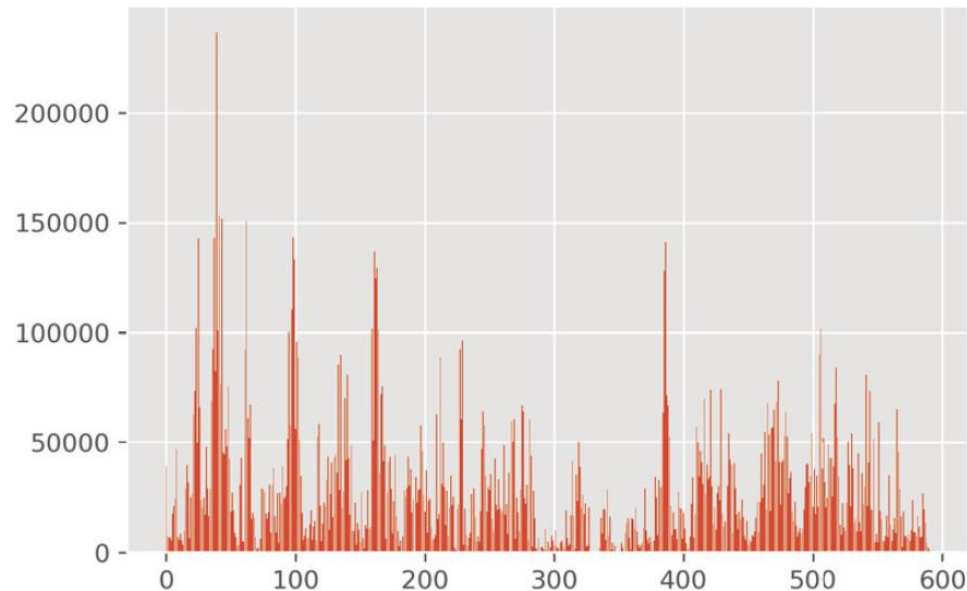
[39110, 7703, 6733, 6066, 5143, 14893, 20077, 22615, 46015, 6088, 5822, 845
3225, 7973, 27835, 38199, 31012, 6313, 23263, 25458, 59821, 70814, 99368, 4
130638, 59461, 18787, 17974, 23502, 16391, 48487, 27365, 14946, 26301, 7841
146181, 84974, 243083, 99608, 154641, 77272, 153133, 46364, 45277, 56578, 4
75744, 40444, 18646, 26197, 18343, 7965,

(생략)

출근 시간대 사람들이 가장 많이 타고 내리는 역은?

- 아침 7시 승차 데이터 막대 그래프로 표현하기

```
import matplotlib.pyplot as plt  
plt.bar(range(len(result)), result)  
plt.show()
```



출근 시간대 사람들이 가장 많이 타고 내리는 역은?

- 아침 7시 승차 데이터 오름차순 정렬하기

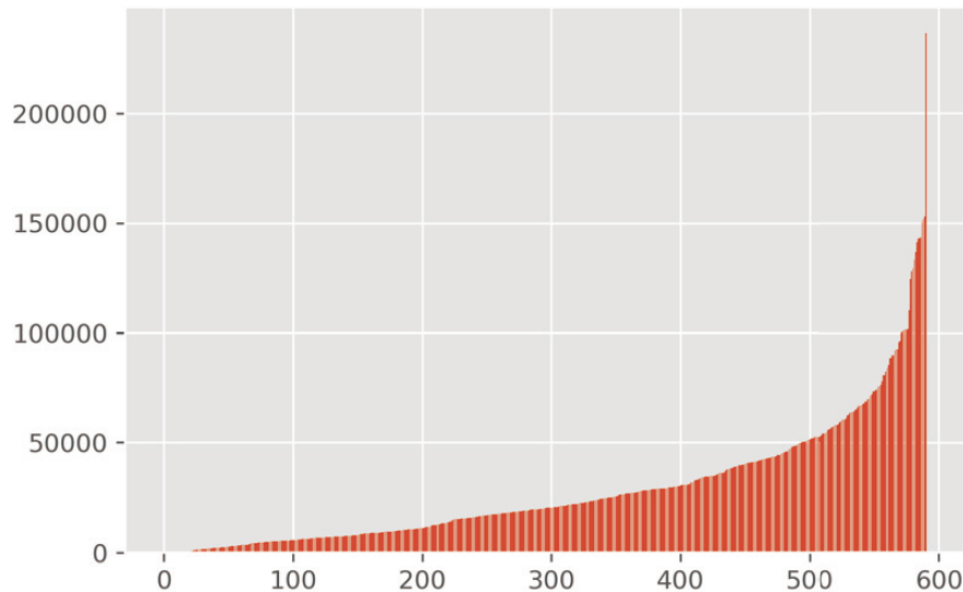
```
import matplotlib.pyplot as plt
```

```
result.sort() # 오름차순으로 정렬
```

```
plt.bar(range(len(result)), result)
```

```
plt.show()
```

**** 가장 많은 역이름과 인원을 출력하기**



출근 시간대 사람들이 가장 많이 타고 내리는 역은?

- 아침 7~9시 승차 인원 최대 역 찾기

```
import csv
f = open('subwaytime.csv')
data = csv.reader(f)
next(data)
next(data)

mx = 0 # 최댓값을 저장할 변수 초기화
mx_station = '' # 최댓값을 갖는 역 이름 저장 변수 초기화
for row in data : # 최댓값 찾기(전부 탐색하여 최댓값을 갱신하는 방식)
    row[4:] = map(int, row[4:])
    if sum(row[10:15:2]) > mx :
        mx = sum(row[10:15:2])
        mx_station = row[3]+'('+row[1]+')'
print(mx_station, mx)
```

신림(2호선) 809541

- format(value, ',') 천단위콤마넣기

- 아침 7~9시 하차 인원 최대 역 찾기

밤 11시에 사람들이 가장 많이 타는 역은?

승차 시작 시각

- 승차 시간 패턴 찾기

0	1	2	3	4	5	6	7	8	9	10	11
사용월	호선명	역ID	역이름	4시		5시		6시		7시	
				승차	하차	승차	하차	승차	하차	승차	하차
2019-01	1호선	0150	서울역	927	22	10682	8774	13504	45545	399110	103863
2019-01	1호선	0151	시청	74	0	2409	1019	3532	19241	7703	66876

- 승차 시간 패턴 찾기

승차 시각과 인덱스 사이의 패턴 찾기		
승차 시각(t)	인덱스(i)	패턴
4	4	식계산해 보기 $i = 4 + (t - 4) * 2$
5	6	
6	8	
7	10	
...	...	
23	?	

밤 11시에 사람들이 가장 많이 타는 역은?

- 밤 11시에 사람들이 가장 많이 타고 내리는 역 찾기

```
import csv
f = open('subwaytime.csv')
data = csv.reader(f)
next(data)
next(data)

mx = 0
mx_station = ''

t = int(input('몇 시의 승차 인원이 가장 많은 역이 궁금하세요? : '))

for row in data :
    row[4:] = map(int, row[4:])
    a = row[4+(t-4)*2]          # 입력 받은 시각의 승차 인원 값 추출하기
    if a > mx :                 # 모든 데이터 탐색
        mx = a
        mx_station = row[3]+'('+ row[1]+'+'
print(mx_station, mx)         # 승차 인원이 가장 큰 역과 인원 값 출력
```

몇 시의 승차 인원이 가장 많은 역이 궁금하세요? **23** → 입력
강남(2호선) 145504

** 월에따라 차이가 날 수있다는 것알기

시간대별로 사람들이 가장 많이 타고 내리는 역은?

- 시간대별(새벽4시 ~ 새벽 3시) 최대 승차 역 이름 및 승차 인원 출력하기

```
import csv
f = open('subwaytime.csv')
data = csv.reader(f)
next(data)
next(data)
mx = [0] * 24 # 시간대별 최대 승차 인원 저장 리스트 초기화
mx_station = [''] * 24 # 시간대별 최대 승차 인원 역 이름 저장 리스트 초기화
for row in data :
    row[4:] = map(int, row[4:])
    for j in range(24) :
        a = row[j*2+4] # j와 인덱스 번호 사이의 관계식 사용
        if a > mx[j] :
            mx[j] = a
            mx_station[j] = row[3]
print(mx_station)
print(mx)
```

시간대별로 사람들이 가장 많이 타고 내리는 역은?

- 시간대별(새벽4시 ~ 새벽 3시) 최대 승차 역 이름 및 승차 인원 출력하기

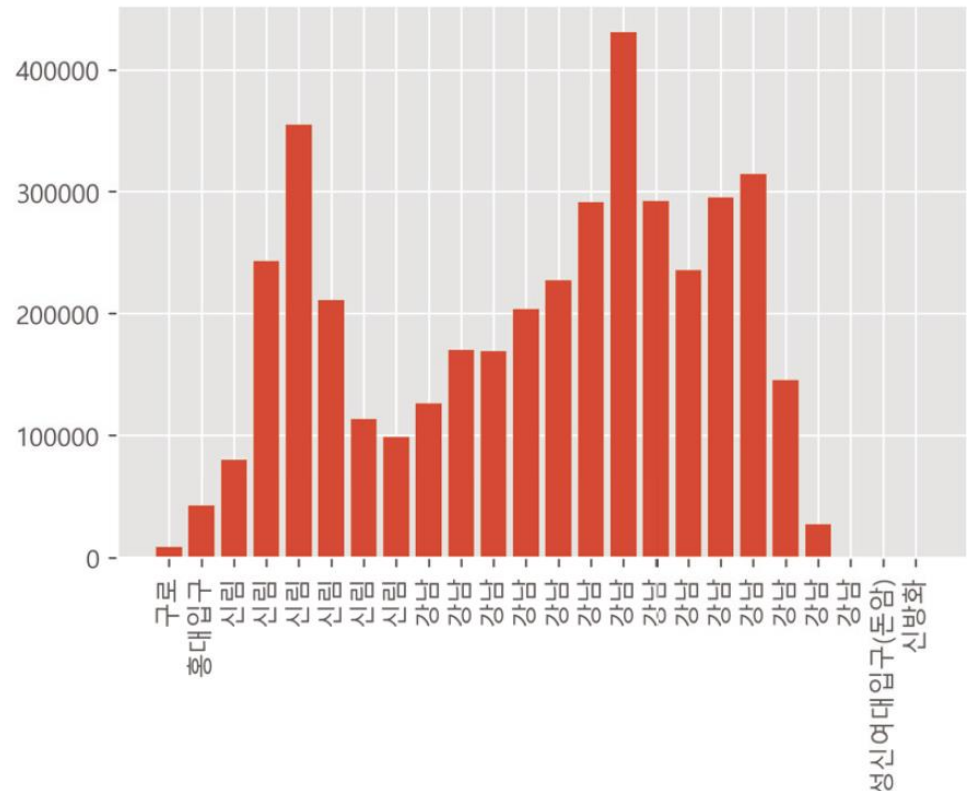
```
mx, max_station=[0]*24, ['']*24
for row in data:
    row[4:] = map(int, row[4:])
    for j in range(24):
        a=row[j*2+4]
        if a > mx[j]:
            mx[j] = a
            max_station[j] = row[3]
for i in range(24):
    print(str((i+4)%24),max_station[i], format(mx[i], ','))
```

```
4 구로 8,418
5 홍대입구 42,966
6 신림 80,407
7 신림 243,083
8 신림 355,172
9 신림 211,286
10 신림 113,830
11 신림 98,765
12 강남 126,159
13 강남 170,216
14 강남 169,097
15 강남 203,483
16 강남 227,268
17 강남 291,623
```

시간대별로 사람들이 가장 많이 타고 내리는 역은?

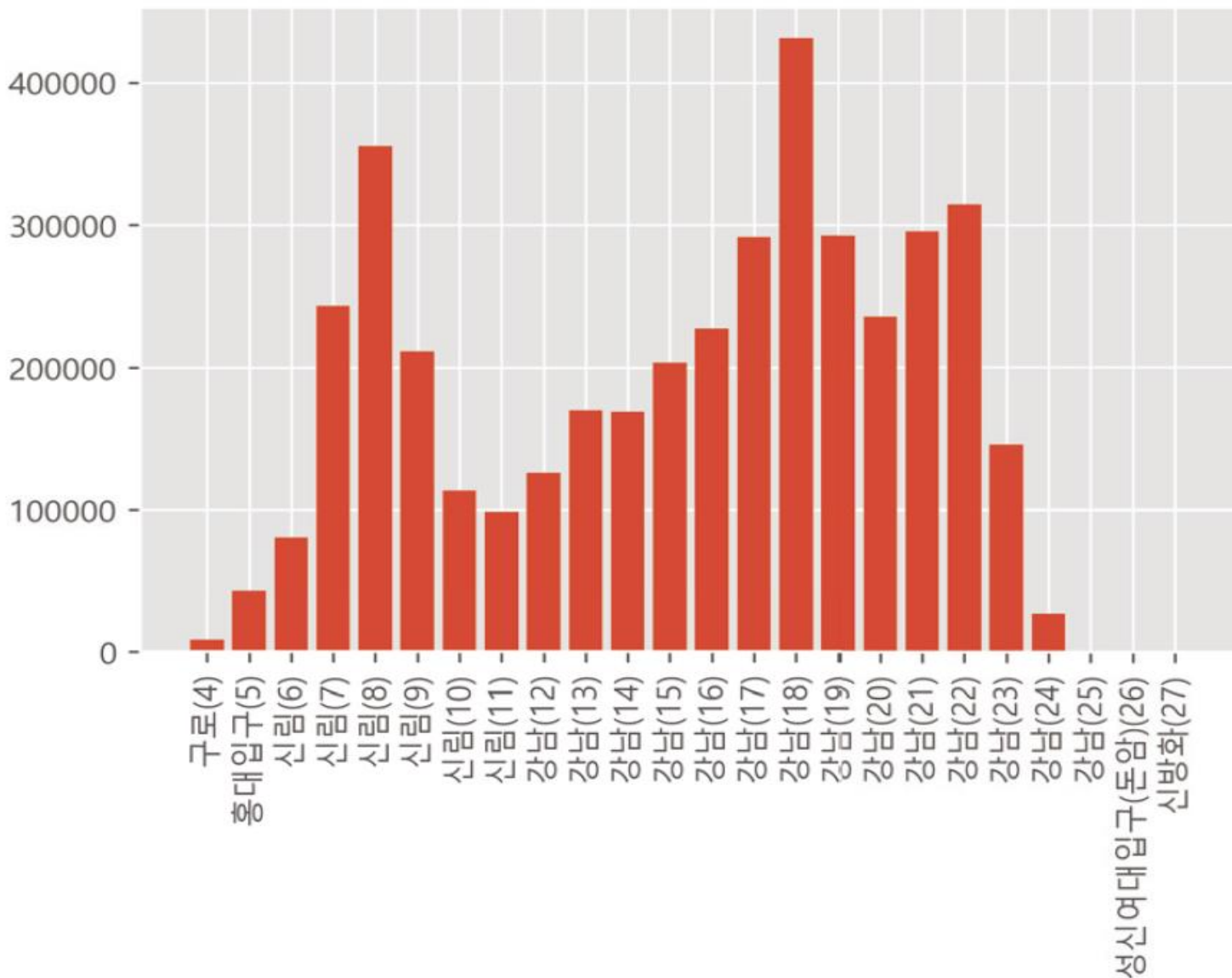
- 시간대별 최대 승차 역 이름 및 승차 인원 막대 그래프로 표현하기

```
import matplotlib.pyplot as plt
plt.rc('font', family='Malgun Gothic')
plt.bar(range(24), mx)
plt.xticks(range(24), mx_station, rotation=90)
plt.show()
```



시간대별로 사람들이 가장 많이 타고 내리는 역은?

- 역 이름 + 시간대 추가해서 나타내기



시간대별로 사람들이 가장 많이 타고 내리는 역은?

- 24시간기준 승차 하차 최대 역 그래프 나란히 나타내기

```
result = []
mx, max_station, mx2, max_station2 = [0]*24, ['']*24, [0]*24, ['']*24
for row in data:
    row[4:] = map(int, row[4:])
    for j in range(24):
        a=row[j*2+4]
        if a > mx[j]:
            mx[j] = a
            max_station[j] = row[3]
    for j in range(24):
        a=row[j*2+5]
        if a > mx2[j]:
            mx2[j] = a
            max_station2[j] = row[3]

# for i in range(24):
#     print(str((i+4)%24),max_station[i], format(mx[i],','))
import matplotlib.pyplot as plt
plt.rc('font',family='Malgun Gothic')
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.bar(range(24), mx)
plt.xticks(range(24), max_station, rotation=90)
plt.style.use('ggplot')
plt.subplot(1,2,2)
plt.bar(range(24), mx2)
plt.xticks(range(24), max_station2, rotation=90)
plt.show()
```

모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

5 | 모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

- 알고리즘 설계하기

- 1 | 데이터를 읽어온다.

- 2 | 모든 역에 대해 시간대별 승차 인원과 하차 인원을 누적해서 더한다.

- 3 | 시간대별 승차 인원과 하차 인원을 그래프로 표현한다.

모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

- 시간대별 승하차 인원 저장할 리스트 만들기

```
import csv
f = open('subwaytime.csv')
data = csv.reader(f)
next(data)
next(data)
s_in = [0] * 24      # 승차 인원을 저장할 리스트 초기화
s_out = [0] * 24     # 하차 인원을 저장할 리스트 초기화
```

모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

- 시간대별 승하차 인원 저장할 리스트 만들기

```
for row in data :  
    row[4:] = map(int, row[4:])  
    for i in range(24) :  
        s_in[i] += row[4+i*2]  
        s_out[i] += row[5+i*2]
```

모든 지하철역에서 시간대별 승하차 인원을 모두 더하면

