



지하철 유무임 승하차 현황



8조 : 이준형 / 최성우 / 홍소민

크롤링 파트

```
import urllib.request
import json
import pandas as pd
from pandas.io.json import json_normalize
from datetime import date, timedelta
from dateutil.relativedelta import relativedelta
```

```
import matplotlib.pyplot as plt
from matplotlib import rc
rc('font', family = 'Malgun Gothic')
```

```
%matplotlib inline
plt.rcParams['axes.unicode_minus'] = False
```

```
key = '70564c7369726172313135416c615451'
start_date = date(2019,1,1)
columns_name = ([ 'USE_MON', 'LINE_NUM', 'SUB_STA_NM', 'PAY_RIDE_NUM', 'FREE_RIDE_NUM',
                  'PAY_ALIGHT_NUM', 'FREE_ALIGHT_NUM', 'WORK_DT' ])
final_data1 = pd.DataFrame(columns = columns_name ) # 데이터프레임 생성 - 컬럼명 맞추기

for i in range(1, 25):
    target_date = start_date.strftime('%Y%m')
    url = 'http://openapi.seoul.go.kr:8088/%s/json/CardSubwayPayFree/1/1000/%s/'%(key, target_date)
    result = urllib.request.urlopen(url)
    json_raw_data = result.read()
    json_utf8 = json_raw_data.decode('utf-8')
    json_complete = json.loads(json_utf8)
    monthly_plus_data = pd.DataFrame(json_complete['CardSubwayPayFree']['row'])
    final_data1 = pd.concat([final_data1, monthly_plus_data], ignore_index = True)
    plus_month = relativedelta(months=1) # 월 더하기
    start_date = start_date + plus_month
final_data1|
```

크롤링 파트

[4]:

	USE_MON	LINE_NUM	SUB_STA_NM	PAY_RIDE_NUM	FREE_RIDE_NUM	PAY_ALIGHT_NUM	FREE_ALIGHT_NUM	WORK_DT
0	201901	경의선	극산	11758.0	4209.0	11719.0	4073.0	20190203
1	201901	1호선	종로5가	571824.0	275447.0	583955.0	263887.0	20190203
2	201901	1호선	종각	1240810.0	162410.0	1177643.0	152062.0	20190203
3	201901	1호선	청량리(서울시립대입구)	463456.0	352121.0	482960.0	357571.0	20190203
4	201901	1호선	서울역	1516452.0	221180.0	1400464.0	211764.0	20190203
...
15169	202012	중앙선	도심	39346.0	11941.0	29495.0	10917.0	20210103
15170	202012	중앙선	회기	424237.0	86386.0	405759.0	85103.0	20210103
15171	202012	중앙선	도농	168265.0	39617.0	154109.0	38579.0	20210103
15172	202012	중앙선	팔당	15173.0	5036.0	15661.0	4852.0	20210103
15173	202012	중앙선	구리	232951.0	56618.0	238234.0	57343.0	20210103

15174 rows × 8 columns

데이터 전처리 - 무임승차 합

```
final_data1.drop(columns = ['WORK_DT'], inplace = True)
final_data1.columns = ['년/월', '노선번호', '역명', '유임승차', '무임승차', '유임하차', '무임하차']
```

```
final_data2019 = final_data1[:7978]
final_data2020 = final_data1[7978:]
```

```
final_data2019 = final_data2019.groupby('노선번호')[['무임승차']].sum()
final_data2019.sort_values(by = '무임승차', axis=0, ascending=True, inplace = True)
```

```
final_data2020 = final_data2020.groupby('노선번호')[['무임승차']].sum()
final_data2020.sort_values(by = '무임승차', axis=0, ascending=True, inplace = True)
```

```
final_data_merge = pd.merge(final_data2019, final_data2020, on = '노선번호')
final_data_merge.columns = ['무임승차2019', '무임승차2020']
final_data_merge
```

데이터 전처리 - 무임승차 합

무임승차2019 무임승차2020

노선번호		
장항선	1445968.0	826879.0
경강선	1860149.0	1151354.0
공항철도 1호선	3456247.0	2235697.0
수인선	3626764.0	2800040.0
경춘선	3741592.0	2880457.0
9호선2~3단계	5052687.0	3356551.0
우이신설선	5319215.0	4198999.0
과천선	6515703.0	4064471.0
안산선	6701519.0	4808470.0
경의선	7586454.0	5439583.0
중앙선	8492499.0	6258689.0
일산선	9697932.0	6672367.0
8호선	15022137.0	9505945.0
9호선	16092380.0	9913095.0
경인선	22517569.0	15951776.0
분당선	22898197.0	15723308.0
경원선	24291740.0	15461434.0

6호선	25591122.0	15965586.0
경부선	30137396.0	17892958.0
1호선	31616447.0	18937898.0
4호선	39955039.0	24417915.0
3호선	40747826.0	24242466.0
5호선	45004240.0	28223519.0
7호선	46319178.0	29198081.0
2호선	75513850.0	45255901.0

데이터 전처리 - 총 승차 합

```
ride2019 = final_data2[:7171]
ride2019 = ride2019.groupby(['LINE_NUM']) [list(final_data2.loc[:, 'FOUR_RIDE_NUM': :2].columns)].sum()
ride2019 = pd.DataFrame(ride2019.sum(axis=1), columns = ['RIDE'])
ride2019.sort_values(by = 'RIDE', axis=0, ascending=True, inplace = True)

ride2019
```

```
ride2020 = final_data2[7171:]
ride2020 = ride2020.groupby(['LINE_NUM']) [list(final_data2.loc[:, 'FOUR_RIDE_NUM': :2].columns)].sum()
ride2020 = pd.DataFrame(ride2020.sum(axis=1), columns = ['RIDE'])
ride2020.sort_values(by = 'RIDE', axis=0, ascending=True, inplace = True)

ride2020
```

데이터 전처리 - 총 승차 합

무임승차2019 무임승차2020

노선번호

장항선	1445968.0	826879.0
경강선	1860149.0	1151354.0
공항철도 1호선	3456247.0	2235697.0
수인선	3626764.0	2800040.0
경춘선	3741592.0	2880457.0
9호선2~3단계	5052687.0	3356551.0
우이신설선	5319215.0	4198999.0
과천선	6515703.0	4064471.0
안산선	6701519.0	4808470.0
경의선	7586454.0	5439583.0
중앙선	8492499.0	6258689.0
일산선	9697932.0	6672367.0
8호선	15022137.0	9505945.0
9호선	16092380.0	9913095.0
경인선	22517569.0	15951776.0
분당선	22898197.0	15723308.0
경원선	24291740.0	15461434.0

6호선	25591122.0	15965586.0
경부선	30137396.0	17892958.0
1호선	31616447.0	18937898.0
4호선	39955039.0	24417915.0
3호선	40747826.0	24242466.0
5호선	45004240.0	28223519.0
7호선	46319178.0	29198081.0
2호선	75513850.0	45255901.0

데이터 시각화-무임승차 합 / 총 승차합

```
final_data_merge.plot(ylim=(500000, 100000000),kind = 'bar', figsize = (14, 8))  
plt.title('각 호선별 무임승차')  
plt.xlabel('노선')  
plt.ylabel('무임승차건')
```

Text(0, 0.5, '무임승차건')

```
final_data_merge2.plot(ylim=(500, 800000000),kind = 'bar', figsize = (14, 8))  
plt.title('각 호선별 총승차')  
plt.xlabel('노선')  
plt.ylabel('총승차건')  
plt.show()
```


