

## 서울 지하철 노선 사용량

- <https://www.data.go.kr/data/15071921/fileData.do#>
- <https://data.seoul.go.kr/dataList/OA-12914/S/1/datasetView.do>

```
In [2]: import pandas as pd
file_path = "./ref/CARD_SUBWAY_MONTH_202501.csv"
data = pd.read_csv(file_path, encoding="utf-8")
data.head() # 데이터 5개만 가져오기
```

```
Out[2]:
```

	사용일자	노선명	역명	승차총승객수	하차총승객수	등록일자	Unnamed: 6
0	20250101	수인선	송도	1453	1321	20250104	NaN
1	20250101	4호선	창동	12477	13408	20250104	NaN
2	20250101	4호선	쌍문	12792	12199	20250104	NaN
3	20250101	4호선	수유(강북구청)	17606	17442	20250104	NaN
4	20250101	4호선	미아(서울사이버대학)	6819	6532	20250104	NaN

```
In [3]: columns = ['date', 'state_name', 'name', 'get_person', 'set_person', 'update', 'N']
data.columns = columns
data['total']=(data['get_person']).astype(int)+(data['set_person']).astype(int)
data = data[['date', 'state_name', 'name', 'get_person', 'set_person', 'total']]
data
```

Out[3]:

	date	state_name	name	get_person	set_person	total
0	20250101	수인선	송도	1453	1321	2774
1	20250101	4호선	창동	12477	13408	25885
2	20250101	4호선	쌍문	12792	12199	24991
3	20250101	4호선	수유(강북구청)	17606	17442	35048
4	20250101	4호선	미아(서울사이버대학)	6819	6532	13351
...	...	...	...	...	...	...
19124	20250131	수인선	남동인더스파크	1622	1748	3370
19125	20250131	수인선	원인재	3834	3942	7776
19126	20250131	수인선	연수	4170	4016	8186
19127	20250131	수인선	송도	2076	1848	3924
19128	20250131	수인선	인하대	5245	4854	10099

19129 rows × 6 columns

```
In [4]: data = data.groupby(['date', 'state_name', 'name', 'get_person', 'set_person', 'total'])
data.head()
```

Out[4]:

	date	state_name	name	get_person	set_person	total
0	20250101	1호선	동대문	8630	8397	17027
1	20250101	1호선	동묘앞	8327	8754	17081
2	20250101	1호선	서울역	41901	35522	77423
3	20250101	1호선	시청	11904	11425	23329
4	20250101	1호선	신설동	6341	6152	12493

```
In [5]: gdata = data[['state_name', 'name', 'date', 'total']]
gdata = gdata.sort_values(by=['name', 'date'], ascending=[True, True]) #name, date, t
gdata.head(10)
```

Out[5]:

	state_name	name	date	total
565	우이신설선	4.19민주묘지	20250101	3220
1182	우이신설선	4.19민주묘지	20250102	5983
1798	우이신설선	4.19민주묘지	20250103	5756
2414	우이신설선	4.19민주묘지	20250104	5986
3031	우이신설선	4.19민주묘지	20250105	3250
3649	우이신설선	4.19민주묘지	20250106	5857
4266	우이신설선	4.19민주묘지	20250107	5872
4884	우이신설선	4.19민주묘지	20250108	5710
5502	우이신설선	4.19민주묘지	20250109	5150
6120	우이신설선	4.19민주묘지	20250110	5485

```
In [7]: mdata=gdata.groupby(["state_name","name"]).sum().reset_index()
# mdata=gdata.groupby(by="name").sum().reset_index()
tdata=mdata[["state_name","name","total"]]
tdata
```

Out[7]:

	state_name	name	total
0	1호선	동대문	669467
1	1호선	동묘앞	564871
2	1호선	서울역	3641701
3	1호선	시청	1403574
4	1호선	신설동	728346
...	...	...	...
616	중앙선	원덕	17402
617	중앙선	중랑	287249
618	중앙선	지평	2695
619	중앙선	팔당	41050
620	중앙선	회기	1241543

621 rows × 3 columns

```
In [8]: sdata=tdata.sort_values(by='total', ascending=False)#.reset_index()
sdata.head(10)
total_data = sdata[:20]
total_data
```

Out[8]:

	state_name	name	total
10	2호선	강남	4264556
59	2호선	홍대입구	4254846
52	2호선	잠실(송파구청)	4211104
2	1호선	서울역	3641701
14	2호선	구로디지털단지	2955955
37	2호선	신림	2952566
63	3호선	고속터미널	2803602
26	2호선	삼성(무역센터)	2739577
36	2호선	신도림	2702968
30	2호선	선릉	2633247
43	2호선	역삼	2618006
50	2호선	을지로입구	2605643
31	2호선	성수	2545701
357	경부선	용산	2520648
25	2호선	사당	2471592
28	2호선	서울대입구(관악구청)	2392684
354	경부선	영등포	2361598
350	경부선	수원	2280975
215	7호선	가산디지털단지	2053854
12	2호선	건대입구	2050035

```
In [14]: import matplotlib
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

d2_path = "./ref/D2Coding-Ver1.3.2-20180524.ttf"
fm.fontManager.addfont(d2_path)

plt.rcParams["font.family"] = "D2coding"
```

```
In [15]: import squarify
plt.figure(figsize=(14,7))
squarify.plot(sizes=total_data['total'], label=[f'{x[0]}({x[1]}명)' for x in zip
# squarify.plot(sizes=total_data['total'], label=total_data.name, alpha=0.7)
plt.rcParams['font.size']=18
# plt.rcParams['text.color']='blue'
plt.gca().invert_yaxis() #축 뒤집기
plt.axis("off") # 인덱스 없애기
# plt.savefig("saving_subway_seoul.png")
plt.show()
```



## 서울 지하철 노선 사용량 리포트

- 서울 지하철의 노선을 뽑아 사용량이 많은 역이 큰 직사각형을 띄는 차트이다.
- 강남, 홍대입구, 잠실(송파구청)순으로 많은 이용객들이 사용하고 있다.

In [ ]: